Table of Contents

Cover 1
Preface: Harrison's Principles of Internal Medicine, 18e 2
Copyright: Harrison's Principles of Internal Medicine, 18e 4
Contributors: Harrison's Principles of Internal Medicine, 18e 5
Part 1. Introduction to Clinical Medicine 35
Chapter 1. The Practice of Medicine 35
Chapter 2. Global Issues in Medicine 45
Chapter 3. Decision-Making in Clinical Medicine 60
Chapter 4. Screening and Prevention of Disease 75
Chapter 5. Principles of Clinical Pharmacology 82
Chapter 6. Women's Health 107
Chapter 7. Medical Disorders During Pregnancy 113
Chapter 8. Medical Evaluation of the Surgical Patient 122
Chapter 9. Palliative and End-of-Life Care 130
Chapter 10. The Safety and Quality of Health Care 157
Chapter e1. Primary Care in Low- and Middle-Income Countries 165
Chapter e2. Complementary, Alternative, and Integrative Medicine 176
Chapter e3. The Economics of Medical Care 184
Chapter e4. Racial and Ethnic Disparities in Health Care 189
Chapter e5. Ethical Issues in Clinical Medicine 199
Chapter e6. Neoplasia During Pregnancy 205
Part 2. Cardinal Manifestations and Presentation of Diseases 211
Section 1. Pain 211
Chapter 11. Pain: Pathophysiology and Management 211
Chapter 12. Chest Discomfort 225
Chapter 13. Abdominal Pain 233
Chapter 14. Headache 240
Chapter 15. Back and Neck Pain 265
Section 2. Alterations in Body Temperature 287
Chapter 16. Fever and Hyperthermia 287
Chapter 17. Fever and Rash 294
Chapter e7. Atlas of Rashes Associated With Fever 307
Chapter 18. Fever of Unknown Origin 336
Chapter 19. Hypothermia and Frostbite 348
Section 3. Nervous System Dysfunction 355
Chapter 20. Syncope 355
Chapter 21. Dizziness and Vertigo 363
Chapter 22. Weakness and Paralysis 369
Chapter 23. Numbness, Tingling, and Sensory Loss 377
Chapter 24. Gait and Balance Disorders 384
Chapter e8. Video Library of Gait Disorders 391
Chapter 25. Confusion and Delirium 392
Chapter 26. Aphasia, Memory Loss, and Other Focal Cerebral Disorders 400
Chapter e9. Memory Loss 417
Chapter e10. Primary Progressive Aphasia, Memory Loss, and Other Focal Cerebral Disorders 423
Chapter 27. Sleep Disorders 424
Section 4. Disorders of Eyes, Ears, Nose, and Throat 440
Chapter 28. Disorders of the Eye 440
Chapter e11. Video Library of Neuro-Ophthalmology 470
Chapter 29. Disorders of Smell and Taste 527

Chapter 30. Disorders of Hearing 537
Chapter 31. Pharyngitis, Sinusitis, Otitis, and Other Upper Respiratory Tract Infections 546
Chapter 32. Oral Manifestations of Disease 564
Chapter e12. Atlas of Oral Manifestations of Disease 576
Section 5. Alterations in Circulatory and Respiratory Functions 590
Chapter 33. Dyspnea 590
Chapter 34. Cough and Hemoptysis 597
Chapter 35. Hypoxia and Cyanosis 603
Chapter 36. Edema 609
Chapter e13. Approach to the Patient with a Heart Murmur 617
Chapter 37. Palpitations 634
Section 6. Alterations in Gastrointestinal Function 636
Chapter 38. Dysphagia 636
Chapter 39. Nausea, Vomiting, and Indigestion 640
Chapter 40. Diarrhea and Constipation 650
Chapter 41. Gastrointestinal Bleeding 668
Chapter 42. Jaundice 674
Chapter 43. Abdominal Swelling and Ascites 684
Section 7. Alterations in Renal and Urinary Tract Function 690
Chapter 44. Azotemia and Urinary Abnormalities 690
Chapter e14. Atlas of Urinary Sediments and Renal Biopsies 701
Chapter 45. Fluid and Electrolyte Disturbances 727
Chapter e15. Fluid and Electrolyte Imbalances and Acid-Base Disturbances: Case Examples 749
Chapter 46. Hypercalcemia and Hypocalcemia 766
Chapter 47. Acidosis and Alkalosis 771
Section 8. Alterations in Sexual Function and Reproduction 788
Chapter 48. Sexual Dysfunction 788
Chapter 49. Hirsutism and Virilization 799
Chapter 50. Menstrual Disorders and Pelvic Pain 805
Section 9. Alterations in the Skin 811
Chapter 51. Approach to the Patient with a Skin Disorder 811
Chapter 52. Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders 824
Chapter 53. Skin Manifestations of Internal Disease 839
Chapter 54. Immunologically Mediated Skin Diseases 869
Chapter 55. Cutaneous Drug Reactions 882
Chapter 56. Photosensitivity and Other Reactions to Light 895
Chapter e16. Atlas of Skin Manifestations of Internal Disease 907
Section 10. Hematologic Alterations 949
Chapter 57. Anemia and Polycythemia 949
Chapter 58. Bleeding and Thrombosis 967
Chapter 59. Enlargement of Lymph Nodes and Spleen 980
Chapter 60. Disorders of Granulocytes and Monocytes 991
Chapter e17. Atlas of Hematology and Analysis of Peripheral Blood Smears 1010
Part 3. Genes, the Environment, and Disease 1040
Chapter 61. Principles of Human Genetics 1040
Chapter 62. Chromosome Disorders 1077
Chapter 63. The Practice of Genetics in Clinical Medicine 1091
Chapter e18. Mitochondrial DNA and Heritable Traits and Diseases 1102
Chapter 64. The Human Microbiome 1119
Chapter e19. Systems Biology in Health and Disease 1132
Part 4. Regenerative Medicine 1140
Chapter 65. Stem Cell Biology 1140
Chapter 66. Hematopoietic Stem Cells 1145
Chapter 67. Applications of Stem Cell Biology in Clinical Medicine 1150

Chapter 68. Gene Therapy in Clinical Medicine 1156
Chapter 69. Tissue Engineering 1161
Part 5. Aging 1165
Chapter 70. World Demography of Aging 1165
Chapter 71. The Biology of Aging 1175
Chapter 72. Clinical Problems of Aging 1187
Part 6. Nutrition 1212
Chapter 73. Nutrient Requirements and Dietary Assessment 1212
Chapter 74. Vitamin and Trace Mineral Deficiency and Excess 1217
Chapter 75. Malnutrition and Nutritional Assessment 1233
Chapter 76. Enteral and Parenteral Nutrition Therapy 1243
Chapter 77. Biology of Obesity 1257
Chapter 78. Evaluation and Management of Obesity 1268
Chapter 79. Eating Disorders 1279
Chapter 80. Involuntary Weight Loss 1287
Part 7. Oncology and Hematology 1292
Section 1. Neoplastic Disorders 1292
Chapter 81. Approach to the Patient with Cancer 1292
Chapter 82. Prevention and Early Detection of Cancer 1306
Chapter 83. Cancer Genetics 1318
Chapter 84. Cancer Cell Biology and Angiogenesis 1328
Chapter 85. Principles of Cancer Treatment 1351
Chapter 86. Infections in Patients with Cancer 1386
Chapter 87. Cancer of the Skin 1402
Chapter 88. Head and Neck Cancer 1409
Chapter 89. Neoplasms of the Lung 1416
Chapter 90. Breast Cancer 1434
Chapter 91. Gastrointestinal Tract Cancer 1449
Chapter 92. Tumors of the Liver and Biliary Tree 1467
Chapter 93. Pancreatic Cancer 1479
Chapter 94. Bladder and Renal Cell Carcinomas 1485
Chapter 95. Benign and Malignant Diseases of the Prostate 1489
Chapter 96. Testicular Cancer 1496
Chapter 97. Gynecologic Malignancies 1499
Chapter 98. Soft Tissue and Bone Sarcomas and Bone Metastases 1508
Chapter 99. Carcinoma of Unknown Primary 1515
Chapter 100. Paraneoplastic Syndromes- Endocrinologic-Hematologic 1521
Chapter 101. Paraneoplastic Neurologic Syndromes 1530
Chapter e20. Thymoma 1540
Chapter 102. Late Consequences of Cancer and Its Treatment 1544
Section 2. Hematopoietic Disorders 1552
Chapter 103. Iron Deficiency and Other Hypoproliferative Anemias 1552
Chapter 104. Disorders of Hemoglobin 1564
Chapter 105. Megaloblastic Anemias 1579
Chapter 106. Hemolytic Anemias and Anemia Due to Acute Blood Loss 1592
Chapter 107. Aplastic Anemia, Myelodysplasia, and Related Bone Marrow Failure Syndromes 1607
Chapter 108. Polycythemia Vera and Other Myeloproliferative Diseases 1626
Chapter 109. Acute and Chronic Myeloid Leukemia 1636
Chapter 110. Malignancies of Lymphoid Cells 1651
Chapter e21. Less Common Hematologic Malignancies 1674
Chapter 111. Plasma Cell Disorders 1684
Chapter 112. Amyloidosis 1696
Chapter 113. Transfusion Biology and Therapy 1706
Chapter 114. Hematopoietic Cell Transplantation 1716

Section 3. Disorders of Hemostasis ... 1726
Chapter 115. Disorders of Platelets and Vessel Wall ... 1726
Chapter 116. Coagulation Disorders ... 1739
Chapter 117. Arterial and Venous Thrombosis ... 1753
Chapter 118. Antiplatelet, Anticoagulant, and Fibrinolytic Drugs ... 1761
Part 8. Infectious Diseases ... 1786
Section 1. Basic Considerations in Infectious Diseases ... 1786
Chapter 119. Introduction to Infectious Diseases; Host-Pathogen Interactions ... 1786
Chapter 120. Molecular Mechanisms of Microbial Pathogenesis ... 1787
Chapter 121. Approach to the Acutely Ill Infected Febrile Patient ... 1800
Chapter 122. Immunization Principles and Vaccine Use ... 1810
Chapter 123. Health Recommendations for International Travel ... 1827
Chapter e22. Laboratory Diagnosis of Infectious Diseases ... 1838
Section 2. Clinical Syndromes: Community-Acquired Infections ... 1848
Chapter 124. Infective Endocarditis ... 1848
Chapter 125. Infections of the Skin, Muscles, and Soft Tissues ... 1864
Chapter 126. Osteomyelitis ... 1874
Chapter 127. Intraabdominal Infections and Abscesses ... 1884
Chapter 128. Acute Infectious Diarrheal Diseases and Bacterial Food Poisoning ... 1895
Chapter 129. Clostridium Difficile Infection, Including Pseudomembranous Colitis ... 1906
Chapter 130. Sexually Transmitted Infections; Overview and Clinical Approach ... 1913
Chapter e23. Infectious Complications of Burns ... 1939
Chapter e24. Infectious Complications of Bites ... 1945
Section 3. Clinical Syndromes: Health Care-Associated Infections ... 1951
Chapter 131. Health Care-Associated Infections ... 1951
Chapter 132. Infections in Transplant Recipients ... 1952
Section 4. Approach to Therapy for Bacterial Diseases ... 1970
Chapter 134. Pneumococcal Infections ... 1970
Section 5. Diseases Caused by Gram-Positive Bacteria ... 1970
Chapter 134. Pneumococcal Infections ... 1970
Chapter 135. Staphylococcal Infections ... 1981
Chapter 136. Streptococcal Infections ... 1999
Chapter 137. Enterococcal Infections ... 2013
Chapter 138. Diphtheria and Other Infections Caused by Corynebacteria and Related Species ... 2024
Chapter 139. Listeria Monocytogenes Infections ... 2033
Chapter 140. Tetanus ... 2038
Chapter 141. Botulism ... 2042
Chapter 142. Gas Gangrene and Other Clostridial Infections ... 2047
Section 6. Diseases Caused by Gram-Negative Bacteria ... 2057
Chapter 143. Meningococcal Infections ... 2057
Chapter 144. Gonococcal Infections ... 2070
Chapter 145. Haemophilus and Moraxella Infections ... 2084
Chapter 146. Infections Due to the HACEK Group and Miscellaneous Gram-Negative Bacteria ... 2092
Chapter 147. Legionella Infections ... 2097
Chapter 148. Pertussis and Other Bordetella Infections ... 2104
Chapter 149. Diseases Caused by Gram-Negative Enteric Bacilli ... 2111
Chapter 150. Acinetobacter Infections ... 2127
Chapter 151. Helicobacter pylori Infections ... 2132
Chapter 152. Infections Due to Pseudomonas Species and Related Organisms ... 2139
Chapter 153. Salmonellosis ... 2150
Chapter 154. Shigellosis ... 2161
Chapter 155. Infections Due to Campylobacter and Related Organisms ... 2168
Chapter 156. Cholera and Other Vibrioses ... 2173
Chapter 157. Brucellosis ... 2183

Chapter 158. Tularemia 2190
Chapter 159. Plague and Other Yersinia Infections 2197
Chapter 160. Bartonella Infections, Including Cat-Scratch Disease 2210
Chapter 161. Donovanosis 2220
Section 7. Miscellaneous Bacterial Infections 2224
Chapter 162. Nocardiosis 2224
Chapter 163. Actinomycosis 2232
Chapter 164. Infections Due to Mixed Anaerobic Organisms 2239
Section 8. Mycobacterial Diseases 2252
Chapter 165. Tuberculosis 2252
Chapter 166. Leprosy 2281
Chapter 167. Nontuberculous Mycobacterial Infections 2292
Chapter 168. Antimycobacterial Agents 2299
Section 9. Spirochetal Diseases 2311
Chapter 169. Syphilis 2311
Chapter 170. Endemic Treponematoses 2324
Chapter 171. Leptospirosis 2328
Chapter 172. Relapsing Fever 2337
Chapter 173. Lyme Borreliosis 2344
Section 10. Diseases Caused by Rickettsiae, Mycoplasmas, and Chlamydiae 2352
Chapter 174. Rickettsial Diseases 2352
Chapter 175. Infections Due to Mycoplasmas 2362
Chapter 176. Chlamydial Infections 2367
Section 11. Viral Diseases: General Considerations 2382
Chapter 177. Medical Virology 2382
Chapter 178. Antiviral Chemotherapy, Excluding Antiretroviral Drugs 2396
Section 12. Infections Due to DNA Viruses 2412
Chapter 179. Herpes Simplex Virus Infections 2412
Chapter 180. Varicella-Zoster Virus Infections 2426
Chapter 181. Epstein-Barr Virus Infections, Including Infectious Mononucleosis 2434
Chapter 182. Cytomegalovirus and Human Herpesvirus Types 6, 7, and 8 2441
Chapter 183. Molluscum Contagiosum, Monkeypox, and Other Poxvirus Infections 2448
Chapter 184. Parvovirus Infections 2451
Chapter 185. Human Papillomavirus Infections 2456
Section 13. Infections Due to DNA and RNA Respiratory Viruses 2461
Chapter 186. Common Viral Respiratory Infections 2461
Chapter 187. Influenza 2472
Section 14. Infections Due to Human Immunodeficiency Virus and Other Human Retroviruses 2482
Chapter 188. The Human Retroviruses 2482
Chapter 189. Human Immunodeficiency Virus Disease: AIDS and Related Disorders 2491
Section 15. Infections Due to RNA Viruses 2578
Chapter 190. Viral Gastroenteritis 2578
Chapter 191. Enteroviruses and Reoviruses 2586
Chapter 192. Measles (Rubeola) 2596
Chapter 193. Rubella (German Measles) 2604
Chapter 194. Mumps 2609
Chapter 195. Rabies and Other Rhabdovirus Infections 2614
Chapter 196. Infections Caused by Arthropod- and Rodent-Borne Viruses 2624
Chapter 197. Ebola and Marburg Viruses 2644
Section 16. Fungal Infections 2650
Chapter 198. Diagnosis and Treatment of Fungal Infections 2650
Chapter 199. Histoplasmosis 2655
Chapter 200. Coccidioidomycosis 2660
Chapter 201. Blastomycosis 2666

Chapter 202. Cryptococcosis 2669
Chapter 203. Candidiasis 2674
Chapter 204. Aspergillosis 2680
Chapter 205. Mucormycosis 2688
Chapter 206. Superficial Mycoses and Less Common Systemic Mycoses 2694
Chapter 207. Pneumocystis Infections 2703
Section 17. Protozoal and Helminthic Infections: General Considerations 2708
Chapter e25. Laboratory Diagnosis of Parasitic Infections 2708
Chapter 208. Agents Used to Treat Parasitic Infections 2716
Chapter e26. Pharmacology of Agents Used to Treat Parasitic Infections 2722
Section 18. Protozoal Infections 2734
Chapter 209. Amebiasis and Infection with Free-Living Amebas 2734
Chapter 210. Malaria 2743
Chapter 211. Babesiosis 2763
Chapter e27. Atlas of Blood Smears of Malaria and Babesiosis 2767
Chapter 212. Leishmaniasis 2784
Chapter 213. Chagas' Disease and Trypanosomiasis 2794
Chapter 214. Toxoplasma Infections 2802
Chapter 215. Protozoal Intestinal Infections and Trichomoniasis 2814
Section 19. Helminthic Infections 2824
Chapter 216. Trichinellosis and Other Tissue Nematode Infections 2824
Chapter 217. Intestinal Nematode Infections 2831
Chapter 218. Filarial and Related Infections 2839
Chapter 219. Schistosomiasis and Other Trematode Infections 2850
Chapter 220. Cestode Infections 2863
Part 9. Terrorism and Clinical Medicine 2870
Chapter 221. Microbial Bioterrorism 2870
Chapter 222. Chemical Terrorism 2882
Chapter 223. Radiation Terrorism 2896
Part 10. Disorders of the Cardiovascular System 2907
Section 1. Introduction to Cardiovascular Disorders 2907
Chapter 224. Basic Biology of the Cardiovascular System 2907
Chapter 225. Epidemiology of Cardiovascular Disease 2929
Chapter 226. Approach to the Patient with Possible Cardiovascular Disease 2937
Section 2. Diagnosis of Cardiovascular Disorders 2942
Chapter 227. Physical Examination of the Cardiovascular System 2942
Chapter 228. Electrocardiography 2959
Chapter e28. Atlas of Electrocardiography 2973
Chapter 229. Noninvasive Cardiac Imaging: Echocardiography, Nuclear Cardiology, and MRI/CT Imaging 2985
Chapter e29. Atlas of Noninvasive Cardiac Imaging 3008
Chapter 230. Diagnostic Cardiac Catheterization and Coronary Angiography 3020
Section 3. Disorders of Rhythm 3031
Chapter 231. Principles of Electrophysiology 3031
Chapter 232. The Bradyarrhythmias 3044
Chapter 233. The Tachyarrhythmias 3060
Chapter e30. Atlas of Cardiac Arrhythmias 3089
Section 4. Disorders of the Heart 3098
Chapter 234. Heart Failure and Cor Pulmonale 3098
Chapter 235. Cardiac Transplantation and Prolonged Assisted Circulation 3121
Chapter 236. Congenital Heart Disease in the Adult 3127
Chapter 237. Valvular Heart Disease 3141
Chapter 238. Cardiomyopathy and Myocarditis 3165
Chapter 239. Pericardial Disease 3187
Chapter 240. Tumors and Trauma of the Heart 3200

Chapter e31. Cardiac Manifestations of Systemic Disease 3206
Section 5. Vascular Disease 3211
Chapter 241. The Pathogenesis, Prevention, and Treatment of Atherosclerosis 3211
Chapter e32. Atlas of Atherosclerosis 3225
Chapter 242. The Metabolic Syndrome 3227
Chapter 243. Ischemic Heart Disease 3235
Chapter 244. Unstable Angina and Non-ST-Segment Elevation Myocardial Infarction 3254
Chapter 245. ST-Segment Elevation Myocardial Infarction 3263
Chapter 246. Percutaneous Coronary Interventions and Other Interventional Procedures 3281
Chapter e33. Atlas of Percutaneous Revascularization 3296
Chapter 247. Hypertensive Vascular Disease 3313
Chapter 248. Diseases of the Aorta 3335
Chapter 249. Vascular Diseases of the Extremities 3346
Chapter 250. Pulmonary Hypertension 3361
Part 11. Disorders of the Respiratory System 3371
Section 1. Diagnosis of Respiratory Disorders 3371
Chapter 251. Approach to the Patient with Disease of the Respiratory System 3371
Chapter 252. Disturbances of Respiratory Function 3376
Chapter 253. Diagnostic Procedures in Respiratory Disease 3387
Chapter e34. Atlas of Chest Imaging 3399
Section 2. Diseases of the Respiratory System 3415
Chapter 254. Asthma 3415
Chapter 255. Hypersensitivity Pneumonitis and Pulmonary Infiltrates with Eosinophilia 3436
Chapter 256. Occupational and Environmental Lung Disease 3445
Chapter 257. Pneumonia 3460
Chapter 258. Bronchiectasis and Lung Abscess 3477
Chapter 259. Cystic Fibrosis 3486
Chapter 260. Chronic Obstructive Pulmonary Disease 3493
Chapter 261. Interstitial Lung Diseases 3508
Chapter 262. Deep Venous Thrombosis and Pulmonary Thromboembolism 3523
Chapter 263. Disorders of the Pleura and Mediastinum 3537
Chapter 264. Disorders of Ventilation 3544
Chapter 265. Sleep Apnea 3550
Chapter 266. Lung Transplantation 3555
Part 12. Critical Care Medicine 3562
Section 1. Respiratory Critical Care 3562
Chapter 267. Approach to the Patient with Critical Illness 3562
Chapter 268. Acute Respiratory Distress Syndrome 3576
Chapter 269. Mechanical Ventilatory Support 3584
Section 2. Shock and Cardiac Arrest 3592
Chapter 270. Approach to the Patient with Shock 3592
Chapter 271. Severe Sepsis and Septic Shock 3602
Chapter 272. Cardiogenic Shock and Pulmonary Edema 3616
Chapter 273. Cardiovascular Collapse, Cardiac Arrest, and Sudden Cardiac Death 3624
Section 3. Neurologic Critical Care 3635
Chapter 274. Coma 3635
Chapter 275. Neurologic Critical Care, Including Hypoxic-Ischemic Encephalopathy, and Subarachnoid Hemorrhage 3647
Section 4. Oncologic Emergencies 3667
Chapter 276. Oncologic Emergencies 3667
Part 13. Disorders of the Kidney and Urinary Tract 3683
Chapter 277. Cellular and Molecular Biology of the Kidney 3683
Chapter 278. Adaption of the Kidney to Renal Injury 3697
Chapter 279. Acute Kidney Injury 3703
Chapter 280. Chronic Kidney Disease 3719

Chapter 281. Dialysis in the Treatment of Renal Failure 3739
Chapter 282. Transplantation in the Treatment of Renal Failure 3745
Chapter 283. Glomerular Diseases 3756
Chapter 284. Polycystic Kidney Disease and Other Inherited Tubular Disorders 3784
Chapter 285. Tubulointerstitial Diseases of the Kidney 3798
Chapter 286. Vascular Injury to the Kidney 3811
Chapter 287. Nephrolithiasis 3822
Chapter 288. Urinary Tract Infections, Pyelonephritis, and Prostatitis 3830
Chapter e35. Interstitial Cystitis-Painful Bladder Syndrome 3842
Chapter 289. Urinary Tract Obstruction 3847
Part 14. Disorders of the Gastrointestinal System 3852
Section 1. Disorders of the Alimentary Tract 3852
Section 2. Liver and Biliary Tract Disease 4054
Section 3. Disorders of the Pancreas 4212
Part 15. Disorders of the Joints and Adjacent Tissues 4244

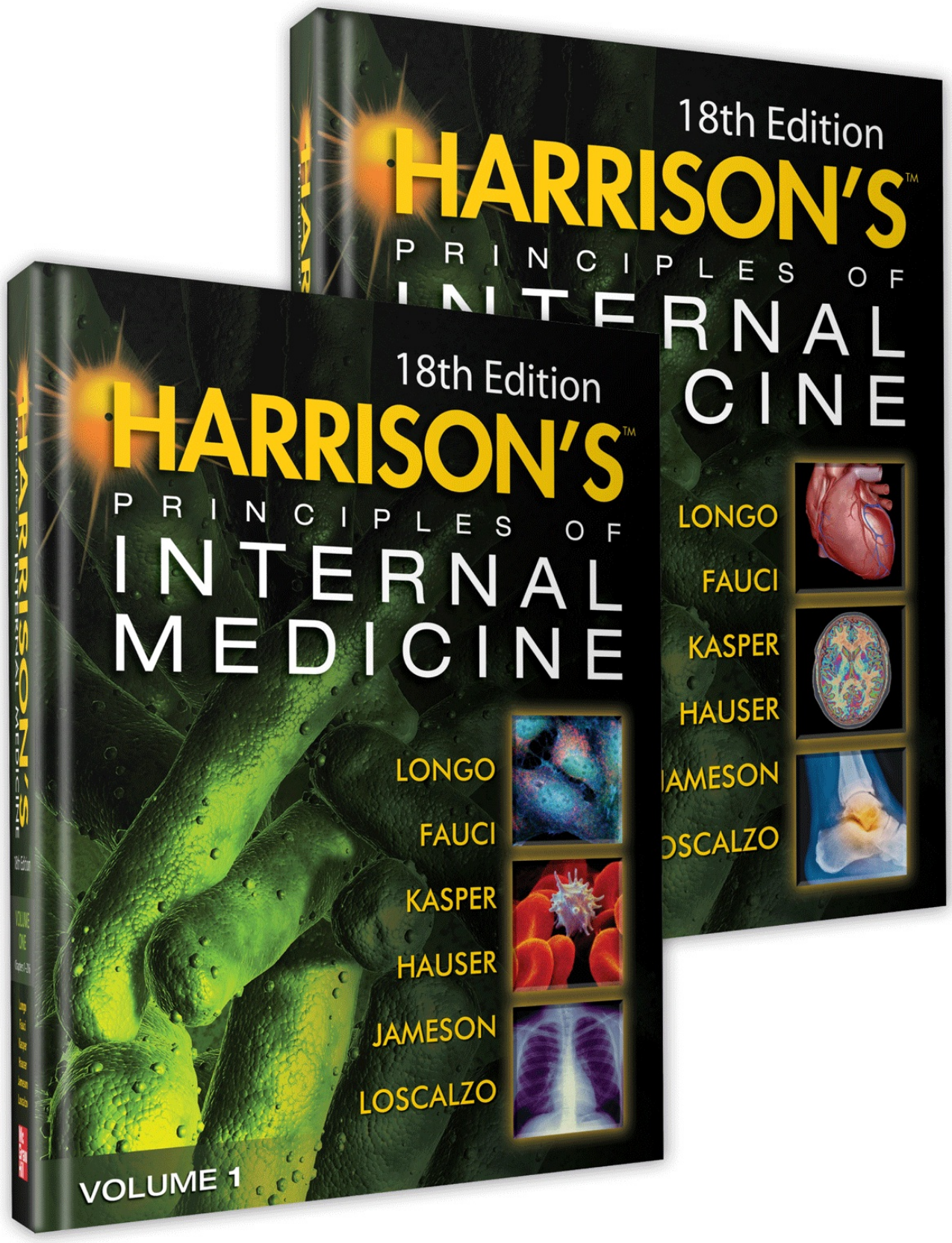
18th Edition
HARRISON'S
PRINCIPLES OF
INTERNAL
MEDICINE
LONGO
FAUCI
KASPER
HAUSER
JAMESON
LOSCALZO
VOLUME 1

ACCESS Medicine from McGraw-Hill

Close Window

Preface

Welcome to the 18th edition of *Harrison's Principles of Internal Medicine.* In the 62 years since the first edition of this textbook was published, virtually every area of medicine has evolved substantially and many new areas have emerged. In 1949, when the first edition appeared, peptic ulcer disease was thought to be caused by stress, nearly every tumor that was not resected resulted in death, rheumatic heart disease was widely prevalent, and hepatitis B and HIV infection were unknown. In the intervening years, both the infectious cause of and the cure for peptic ulcer disease were identified; advances in diagnosis and treatment made it possible to cure two-thirds of cancers; rheumatic heart disease virtually disappeared; atherosclerotic coronary artery disease waxed and then—at least in part through management of modifiable risk factors—began to wane; hepatitis B and its consequences, cirrhosis and hepatocellular carcinoma, became preventable by a vaccine; and HIV, first viewed as a uniformly fatal worldwide scourge, became a treatable chronic disease. During this same period, the amount of information required for the effective practice of medicine grew unabated, and learning options for students, residents, and practicing physicians also burgeoned to include multiple sources of information in print and electronic formats.

While retaining the founding goals of *Harrison's*, this edition has been modified extensively in light of the varied needs of the book's readers and the diverse methods and formats by which information is now acquired. The print version of the 18th edition is more reader-friendly in several respects: the book is printed in type that is easier to read than prior editions, the graphics and tables have been enhanced for ease of interpretation, and more than 300 new figures are included. This improved format requires publication of the print edition in two volumes conveniently divided by subject matter. A DVD accompanies the book and contains additional e-chapters, videos, and atlases; its image bank includes figures and photographs from the book that can be incorporated into slide presentations. All chapters have been extensively updated by experts in the field. In addition, this edition includes 25 new chapters and more than 100 new authors. The pathophysiologic approach to evaluating patients on the basis of their presentation continues to receive emphasis in an enriched section on the cardinal manifestations of disease. A new section focuses on aging, its demographics and biology, and distinctive clinical issues affecting older patients. The e-chapters have increased in number from 39 to 57 and include a new video atlas of neuro-ophthalmology, an audio-enhanced chapter on the approach to a patient with a heart murmur, a case-based teaching exercise in fluid and electrolyte imbalances and acid-base disturbances, and explorations of infectious complications of burns and bites. New videos demonstrate the neurologic examination and several commonly performed medical procedures. A new chapter focuses on neuropsychiatric problems among war veterans. E-chapters on altitude sickness and hyperbaric and diving medicine form a new section on medical effects of changes in environmental pressure.

For readers who wish to continue using *Harrison's* in a single-volume format, we are pleased to offer two new eBook versions of the 18th edition: a traditional eBook, with text and illustrations from the new edition included for reading on a portable e-reader or on a desktop, and an enhanced eBook developed especially for new tablet devices (e.g., iPad, Galaxy, Playbook, Nook) that offer high-definition resolution of multimedia content and interactive features. The *Harrison's* 18th edition enhanced eBook will contain extensive embedded video footage, including all of the new clinical procedural videos; the wonderful neurologic examination videos from Samuels and Lowenstein; examples of cardiovascular imaging and assessment; and high-resolution versions of more than 2000 color images from the book and the *Harrison's* atlases on the companion DVD. Along with other social media features, the enhanced eBook offers users the opportunity to take and share notes from lectures and their own reading. Additional resources include *Harrison's Online*, a continuously updated electronic resource that highlights and summarizes newly published articles on significant medical findings and advances. *Harrison's Self-Assessment and Board Review* , a useful study guide for board review based on information in the 18th edition, will soon be produced. *Harrison's Manual of Medicine*, a pocket version of *Harrison's Principles of Internal Medicine* , is available in both print and electronic formats.

We have many people to thank for their efforts in producing this book. First, the authors have done a superb job of producing authoritative chapters that synthesize vast amounts of scientific and clinical data to create state-of-the-art descriptions of medical disorders encompassed by internal medicine. In today's information-rich, rapidly evolving environment, they have ensured that this information is current. Helpful suggestions and critical input have been provided by a number of colleagues; particularly notable was the advice of Chung Owyang on the Gastroenterology Section. We are most grateful to our colleagues in each of our editorial offices who have kept track of the work in its various phases and facilitated communication with the authors, with the McGraw-Hill staff, and among the editors: Patricia Conrad, Emily Cowan, Patricia L. Duffey, Gregory K. Folkers, Julie B. McCoy, Elizabeth Robbins, Kristine Shontz, and Stephanie Tribuna.

The staff at McGraw-Hill has been a constant source of support and expertise. James Shanahan, Editor-in-Chief, Internal Medicine, for McGraw-Hill's Professional Publishing Division, has been a superb and insightful partner to the editors, guiding the development of the book and its related products in new formats. Kim Davis seamlessly stepped into the position of Associate Managing Editor, ensuring that the complex production of this multi-authored textbook proceeded in an efficient fashion. Paula Torres, Dominik Pucek, and Michael Crumsho oversaw the production of the new procedural and neurology videos. Phil Galea again served as Production Director on this, his final edition, and did so with a peak performance. Mary A. Murray, Director, International Rights, is retiring from McGraw Hill in 2012, after 50 years with the company. Mary joined the Blakiston Division of McGraw Hill in 1961.

in 2012, after 50 years with this company. Mary joined the Blakiston Division of McGraw-Hill in 1961, when Tinsley Harrison was still the editor of the book. Her first assignment was to distribute reprints of *Harrison's* chapters to the editors and contributors. For the next 23 years, Mary continued to be involved in the editorial process of *Harrison's.* In the early 1990s, she was given responsibility for licensing McGraw-Hill's medical titles; making use of her many cordial connections in global medical publishing, she licensed translations of *Harrison's* into 19 languages. We are extremely grateful to Mary for her many accomplishments in support of the book through 13 editions.

We are privileged to have compiled this 18th edition and are enthusiastic about all that it offers our readers. We learned much in the process of editing *Harrison's* and hope that you will find this edition a uniquely valuable educational resource.

The Editors

ACCESS Medicine from McGraw-Hill

Close Window

Copyright Information

Harrison's™
PRINCIPLES OF INTERNAL MEDICINE
Eighteenth Edition

Note: Dr. Fauci's work as editor and author was performed outside the scope of his employment as a U.S. government employee. This work represents his personal and professional views and not necessarily those of the U.S. government.

Copyright © 2012, 2008, 2005, 2001, 1998, 1994, 1991, 1987, 1983, 1980, 1977, 1974, 1970, 1966, 1962, 1958 by The McGraw-Hill Companies, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

Harrison's™ is a trademark of The McGraw-Hill Companies, Inc.

Two Volume Set ISBN 978-0-07174889-6; MHID 0-07-174889-X
Volume 1 ISBN 978-0-07-163244-7; MHID 0-07-163244-1
Volume 2 ISBN 978-0-07-174887-2; MHID 0-07-174887-3
DVD ISBN 978-0-07-174888-9; MHID 0-07-174888-1

Notice

Medicine is an ever-changing science. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy are required. The authors and the publisher of this work have checked with sources believed to be reliable in their efforts to provide information that is complete and generally in accord with the standards accepted at the time of publication. However, in view of the possibility of human error or changes in medical sciences, neither the authors nor the publisher nor any other party who has been involved in the preparation or publication of this work warrants that the information contained herein is in every respect accurate or complete, and they disclaim all responsibility for any errors or omissions or for the results obtained from use of the information contained in this work. Readers are encouraged to confirm the information contained herein with other sources. For example and in particular, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this work is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

ACCESS Medicine from McGraw-Hill

Close Window

Contributors

Editors

Dan L. Longo, MD
Lecturer, Department of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital; Deputy Editor, New England Journal of Medicine, Boston, Massachusetts; Adjunct Investigator, National Institute on Aging, National Institutes of Health, Baltimore, Maryland

Dennis L. Kasper, MD
William Ellery Channing Professor of Medicine, Professor of Microbiology and Molecular Genetics, Harvard Medical School; Director, Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts

J. Larry Jameson, MD, PhD
Robert G. Dunlop Professor of Medicine; Dean, University of Pennsylvania School of Medicine; Executive Vice-President of the University of Pennsylvania for the Health System, Philadelphia, Pennsylvania

Anthony S. Fauci, MD
Chief, Laboratory of Immunoregulation; Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland

Stephen L. Hauser, MD
Robert A. Fishman Distinguished Professor and Chairman, Department of Neurology, University of California, San Francisco, San Francisco, California

Joseph Loscalzo, MD, PhD
Hersey Professor of the Theory and Practice of Medicine, Harvard Medical School; Chairman, Department of Medicine; Physician-in-Chief, Brigham and Women's Hospital, Boston, Massachusetts

Contributors

James L. Abbruzzese, MD
Professor and Chair, Department of GI Medical Oncology; M.G. and Lillie Johnson Chair for Cancer Treatment and Research, University of Texas, MD Anderson Cancer Center, Houston, Texas [99]

Jamil Aboulhosn, MD
Assistant Professor, Departments of Medicine and Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California [236]

John C. Achermann, MD, PhD
Wellcome Trust Senior Fellow, UCL Institute of Child Health, University College London, London, United Kingdom [349]

John W. Adamson, MD
Clinical Professor of Medicine, Department of Hematology/Oncology, University of California, San Diego, San Diego, California [57, 103]

Anthony A. Amato, MD
Professor of Neurology, Harvard Medical School; Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts [384, 385, 387]

Michael J. Aminoff, MD, DSc
Professor of Neurology, University of California, San Francisco School of Medicine, San Francisco, California [22, 23, e45]

Neil M. Ampel, MD
Professor of Medicine, University of Arizona, Tucson, Arizona [200]

Kenneth C. Anderson, MD
Kraft Family Professor of Medicine, Harvard Medical School; Chief, Jerome Lipper Multiple Myeloma Center, Dana-Farber Cancer Institute, Boston, Massachusetts [111, 113]

Elliott M. Antman, MD
Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital; Boston, Massachusetts [243, 245]

Frederick R. Appelbaum, MD
Director, Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington [114]

Gordon L. Archer, MD
Professor of Medicine and Microbiology/Immunology; Senior Associate Dean for Research and Research Training, Virginia Commonwealth University School of Medicine, Richmond, Virginia [133]

Research Training, Virginia Commonwealth University School of Medicine, Richmond, Virginia [196]

Cesar A. Arias, MD, PhD
Assistant Professor, University of Texas Medical School, Houston, Texas; Director, Molecular Genetics and Antimicrobial Resistance Unit, Universidad El Bosque, Bogotá, Colombia [137]

Wiebke Arlt, MD, DSc, FRCP, FMedSci
Professor of Medicine, Centre for Endocrinology, Diabetes and Metabolism, School of Clinical and Experimental Medicine, University of Birmingham; Consultant Endocrinologist, University Hospital Birmingham, Birmingham, United Kingdom [342]

Valder R. Arruda, MD, PhD
Associate Professor of Pediatrics, University of Pennsylvania School of Medicine; Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania [116]

Arthur K. Asbury, MD, FRCP
Van Meter Professor Emeritus of Neurology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania [23]

John R. Asplin, MD
Medical Director, Litholink Corporation; Chicago, Illinois [287]

John C. Atherton, MD, FRCP
Nottingham Digestive Diseases Centre Biomedical Research Unit (NDDC BRU), University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom [151]

Evelyn Attia, MD
Professor of Clinical Psychiatry, Columbia College of Physicians and Surgeons; Weill Cornell Medical College, New York, New York [79]

Paul S. Auerbach, MD, MS
Redlich Family Professor, Department of Surgery, Division of Emergency Medicine, Stanford University School of Medicine, Palo Alto, California [396]

K. Frank Austen, MD
AstraZeneca Professor of Respiratory and Inflammatory Diseases; Director, Inflammation and Allergic Diseases Research Section, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [317]

Eric H. Awtry, MD
Assistant Professor of Medicine, Boston University School of Medicine; Inpatient Clinical Director, Section of Cardiology, Boston Medical Center Boston, Massachusetts [240, e31]

Bruce R. Bacon, MD
James F. King, MD Endowed Chair in Gastroenterology; Professor of Internal Medicine, St. Louis University Liver Center, St. Louis University School of Medicine, St. Louis, Missouri [308, 309]

Lindsey R. Baden, MD
Associate Professor of Medicine, Harvard Medical School; Dana-Farber Cancer Institute, Brigham and Women's Hospital, Boston, Massachusetts [178]

John R. Balmes, MD
Professor of Medicine, San Francisco General Hospital, San Francisco, California [256]

Manisha Balwani, MD, MS
Assistant Professor, Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine of New York University, New York, New York [358]

Peter A. Banks, MD
Professor of Medicine, Harvard Medical School; Senior Physician, Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts [312, 313]

Robert L. Barbieri, MD
Kate Macy Ladd Professor of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School; Chairperson, Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, Massachusetts [7]

Joanne M. Bargman, MD, FRCPC
Professor of Medicine, University of Toronto; Staff Nephrologist, University Health Network; Director, Home Peritoneal Dialysis Unit and Co-Director, Renal Rheumatology Lupus Clinic, University Health Network, Toronto, Ontario, Canada [280]

Tamar F. Barlam, MD
Associate Professor of Medicine, Boston University School of Medicine, Boston, Massachusetts [121, 146]

Peter J. Barnes, DM, DSc, FMedSci, FRS
Head of Respiratory Medicine, Imperial College, London, United Kingdom [254]

Richard J. Barohn, MD
Chairman, Department of Neurology; Gertrude and Dewey Ziegler Professor of Neurology, University

Chairman, Department of Neurology, Conrad and Dewey Ziegler Professor of Neurology, University of Kansas Medical Center, Kansas City, Kansas [384]

Miriam J. Baron, MD
Assistant Professor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [127]

Rebecca M. Baron, MD
Assistant Professor, Harvard Medical School; Associate Physician, Department of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [258]

John G. Bartlett, MD
Professor of Medicine and Chief, Division of Infectious Diseases, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland [258]

Robert C. Basner, MD
Professor of Clinical Medicine, Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University College of Physicians and Surgeons, New York, New York [Appendix]

Buddha Basnyat, MD, MSc, FACP, FRCP(E)
Principal Investigator, Oxford University Clinical Research Unit-Patan Academy of Health Sciences; Medical Director, Nepal International Clinic, Kathmandu, Nepal [e51]

Shari S. Bassuk, ScD
Epidemiologist, Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts [348]

John F. Bateman, PhD
Director, Cell Biology, Development and Disease, Murdoch Childrens Research Institute, Parkville, Victoria, Australia [363]

David W. Bates, MD, MSc
Professor of Medicine, Harvard Medical School; Chief, General Internal Medicine and Primary Care Division, Brigham and Women's Hospital; Medical Director, Clinical and Quality Analysis, Partners HealthCare System, Inc., Boston, Massachusetts [10]

Robert P. Baughman, MD
Department of Internal Medicine, University of Cincinnati Medical Center, Cincinnati, Ohio [329]

M. Flint Beal, MD
Chairman of Neurology and Neuroscience; Neurologist-in-Chief, New York Presbyterian Hospital; Weill Cornell Medical College, New York, New York [366, 376]

Laurence H. Beck, Jr., MD, PhD
Assistant Professor of Medicine, Boston University School of Medicine, Boston, Massachusetts [285]

Nicholas J. Beeching, MA, BM BCh, FRCP, FRACP, FFTM RCPS (Glasg), DCH, DTM&H
Senior Lecturer (Clinical) in Infectious Diseases, Liverpool School of Tropical Medicine; Clinical Lead, Tropical and Infectious Disease Unit, Royal Liverpool University Hospital; Honorary Consultant, Health Protection Agency; Honorary Civilian Consultant in Infectious Diseases, Army Medical Directorate, Liverpool, United Kingdom [157]

Robert S. Benjamin, MD
P.H. and Fay E. Robinson Distinguished Professor and Chair, Department of Sarcoma Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas [98]

Michael H. Bennett, MD, MBBS
Conjoint Associate Professor in Anesthesia and Hyperbaric Medicine; Faculty of Medicine, University of New South Wales; Senior Staff Specialist, Department of Diving and Hyperbaric Medicine, Prince of Wales Hospital, Sydney, Australia [e52]

Edward J. Benz, Jr., MD
Richard and Susan Smith Professor of Medicine, Professor of Pediatrics, Professor of Genetics, Harvard Medical School; President and CEO, Dana-Farber Cancer Institute; Director, Dana-Farber/Harvard Cancer Center (DF/HCC), Boston, Massachusetts [104]

Jean Bergounioux, MD, PhD
Pediatric Intensive Care Unit, Hôpital Necker-Enfants Malades, Paris, France [154]

Joseph R. Betancourt, MD, MPH
Associate Professor of Medicine, Harvard Medical School; Director, The Disparities Solutions Center, Massachusetts General Hospital, Boston, Massachusetts [e4]

Atul K. Bhan, MBBS, MD
Professor of Pathology, Harvard Medical School; Director of Immunopathology, Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts [e38]

Shalender Bhasin, MD
Professor of Medicine; Section Chief, Division of Endocrinology, Diabetes and Nutrition, Boston University School of Medicine, Boston, Massachusetts [346]

Deepak L. Bhatt, MD, MPH
Associate Professor of Medicine, Harvard Medical School; Chief of Cardiology, VA Boston Healthcare System; Director, Integrated Interventional Cardiovascular Program, Brigham and Women's Hospital and VA Boston Healthcare System; Senior Investigator, TIMI Study Group, Boston, Massachusetts [246, e33]

David R. Bickers, MD
Carl Truman Nelson Professor and Chair, Department of Dermatology, College of Physicians and Surgeons, Columbia University Medical Center; Dermatologist-in-Chief, New York Presbyterian Hospital, New York, New York [56]

Henry J. Binder, MD
Professor Emeritus of Medicine; Senior Research Scientist, Yale University, New Haven, Connecticut [294, e37]

William R. Bishai, MD, PhD
Professor and Co-Director, Center for Tuberculosis Research, Department of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland [138]

Bruce R. Bistrian, MD, PhD
Professor of Medicine, Harvard Medical School; Chief, Clinical Nutrition, Beth Israel Deaconess Medical Center, Boston, Massachusetts [76]

Martin J. Blaser, MD
Frederick H. King Professor of Internal Medicine; Chair, Department of Medicine; Professor of Microbiology, New York University School of Medicine, New York, New York [151, 155]

Gijs Bleijenberg, PhD
Professor; Head, Expert Centre for Chronic Fatigue, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands [389]

Clara D. Bloomfield, MD
Distinguished University Professor; William G. Pace, III Professor of Cancer Research; Cancer Scholar and Senior Advisor, The Ohio State University Comprehensive Cancer Center; Arthur G. James Cancer Hospital and Richard J. Solove Research Institute, Columbus, Ohio [109]

Richard S. Blumberg, MD
Chief, Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts [295]

Jean L. Bolognia, MD
Professor of Dermatology, Yale University School of Medicine, New Haven, Connecticut [53]

Joseph V. Bonventre, MD, PhD
Samuel A. Levine Professor of Medicine, Harvard Medical School; Chief, Renal Division; Chief, BWH HST Division of Bioengineering, Brigham and Women's Hospital, Boston, Massachusetts [279]

George J. Bosl, MD
Professor of Medicine, Weill Cornell Medical College; Chair, Department of Medicine; Patrick M. Byrne Chair in Clinical Oncology, Memorial Sloan-Kettering Cancer Center, New York, New York [96]

Richard C. Boucher, MD
Kenan Professor of Medicine, Pulmonary and Critical Care Medicine; Director, Cystic Fibrosis/Pulmonary Reseach and Treatment Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina [259]

Eugene Braunwald, MD, MA (Hon), ScD (Hon) FRCP
Distinguished Hersey Professor of Medicine, Harvard Medical School; Founding Chairman, TIMI Study Group, Brigham and Women's Hospital, Boston, Massachusetts [36, 239, 244]

Irwin M. Braverman, MD
Professor of Dermatology, Yale University School of Medicine, New Haven, Connecticut [53]

Otis W. Brawley, MD
Chief Medical Officer, American Cancer Society Professor of Hematology, Oncology, Medicine, and Epidemiology, Emory University, Atlanta, Georgia [82]

Joel G. Breman, MD, DTPH
Scientist Emeritus, Fogarty International Center, National Institutes of Health, Bethesda, Maryland [210, e27]

George J. Brewer, MD
Morton S. and Henrietta K. Sellner Professor Emeritus of Human Genetics; Emeritus Professor of Internal Medicine, University of Michigan Medical School, Senior Vice President for Research and Development, Adeona Pharmaceuticals, Inc., Ann Arbor, Michigan [360]

Josephine P. Briggs, MD
Director, National Center for Complementary and Alternative Medicine, National Institutes of Health, Bethesda, Maryland [e2]

F. Richard Bringhurst, MD

[illegible]
Associate Professor of Medicine, Harvard Medical School; Physician, Massachusetts General Hospital, Boston, Massachusetts [352]

Steven M. Bromley, MD
Clinical Assistant Professor of Neurology, Department of Medicine, New Jersey School of Medicine and Dentistry–Robert Wood Johnson Medical School, Camden, New Jersey [29]

Kevin E. Brown, MD, MRCP, FRCPath
Consultant Medical Virologist, Virus Reference Department, Health Protection Agency, London, United Kingdom [184]

Robert H. Brown, Jr., MD, PhD
Chairman, Department of Neurology, University of Massachusetts Medical School, Worchester, Massachusetts [374, 387]

Amy E. Bryant, PhD
Research Scientist, Veterans Affairs Medical Center, Boise, Idaho; Affiliate Assistant Professor, University of Washington School of Medicine, Seattle, Washington [142]

Christopher M. Burns, MD
Assistant Professor, Department of Medicine, Section of Rheumatology, Dartmouth Medical School; Dartmouth Hitchcock Medical Center, Lebanon, New Hampshire [359]

David M. Burns, MD
Professor Emeritus, Department of Family and Preventive Medicine, University of California, San Diego School of Medicine, San Diego, California [395]

Stephen B. Calderwood, MD
Morton Swartz MD Academy Professor of Medicine (Microbiology and Molecular Genetics), Harvard Medical School; Chief, Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts [128]

Michael V. Callahan, MD, DTM&H (UK), MSPH
Clinical Associate Physician, Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts; Program Manager, Biodefense, Defense Advanced Research Project Agency (DARPA), United States Department of Defense, Washington, DC [18]

Michael Camilleri, MD
Atherton and Winifred W. Bean Professor; Professor of Medicine and Physiology, Mayo Clinic College of Medicine, Rochester, Minnesota [40]

Christopher P. Cannon, MD
Associate Professor of Medicine, Harvard Medical School; Senior Investigator, TIMI Study Group, Brigham and Women's Hospital, Boston, Massachusetts [244]

Jonathan Carapetis, PhD, MBBS, FRACP, FAFPHM
Director, Menzies School of Health Research, Charles Darwin University, Darwin, Australia [322]

Kathryn M. Carbone, MD
Deputy Scientific Director, Division of Intramural Research, National Institute of Dental and Craniofacial Research, Bethesda, Maryland [194]

Brian I. Carr, MD, PhD, FRCP
Professor of Oncology and Hepatology, IRCCS De Bellis Medical Research Institute, Castellana Grotte, Italy [92]

Arturo Casadevall, MD, PhD
Chair, Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York [202]

Agustin Castellanos, MD
Professor of Medicine, and Director, Clinical Electrophysiology, Division of Cardiology, University of Miami Miller School of Medicine, Miami, Florida [273]

Bartolome R. Celli, MD
Lecturer on Medicine, Harvard Medical School; Staff Physician, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [269]

Murali Chakinala, MD
Associate Professor of Medicine, Division of Pulmonary and Critical Care Medicine, Washington University School of Medicine, St. Louis, Missouri [234]

Anil Chandraker, MD, FASN, FRCP
Associate Professor of Medicine, Harvard Medical School; Medical Director of Kidney and Pancreas Transplantation; Assistant Director, Schuster Family Transplantation Research Center, Brigham and Women's Hospital; Children's Hospital, Boston, Massachusetts [282]

Stanley W. Chapman, MD
Professor of Medicine, University of Mississippi Medical Center, Jackson, Mississippi [201]

Panithaya Chareonthaitawee, MD
Associate Professor of Medicine, Mayo Clinic College of Medicine, Rochester, Minnesota [229, e29]

Lan X. Chen, MD, PhD
Penn Presbyterian Medical Center, Philadelphia, Pennsylvania [333]

Yuan-Tsong Chen, MD, PhD
Distinguished Research Fellow, Institute of Biomedical Sciences, Academia Sinica, Taiwan [362]

Glenn M. Chertow, MD, MPH
Norman S. Coplon/Satellite Healthcare Professor of Medicine; Chief, Division of Nephrology, Stanford University School of Medicine, Palo Alto, California [281]

John S. Child, MD, FACC, FAHA, FASE
Streisand Professor of Medicine and Cardiology, Geffen School of Medicine, University of California, Los Angeles (UCLA); Director, Ahmanson-UCLA Adult Congenital Heart Disease Center; Director, UCLA Adult Noninvasive Cardiodiagnostics Laboratory, Ronald Reagan-UCLA Medical Center; Los Angeles, California [236]

Augustine M. K. Choi, MD
Parker B. Francis Professor of Medicine, Harvard Medical School; Chief, Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [251, 253, 268]

Irene Chong, MRCP, FRCR
Clinical Research Fellow, Royal Marsden NHS Foundation Trust, London and Sutton, United Kingdom [93]

Raymond T. Chung, MD
Associate Professor of Medicine, Harvard Medical School; Director of Hepatology; Vice Chief, Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts [310]

Fredric L. Coe, MD
Professor of Medicine, University of Chicago, Chicago, Illinois [287]

Jeffrey I. Cohen, MD
Chief, Medical Virology Section, Laboratory of Clinical Infectious Diseases, National Institutes of Health, Bethesda, Maryland [181, 191]

Ronit Cohen-Poradosu, MD
Senior Physician, Department of Clinical Microbiology and Infectious Diseases, Hadassah Hebrew Medical Center, Jerusalem, Israel [164]

Francis S. Collins, MD, PhD
Director, National Institutes of Health, Bethesda, Maryland [83]

Wilson S. Colucci, MD
Thomas J. Ryan Professor of Medicine, Boston University School of Medicine; Chief of Cardiovascular Medicine, Boston Medical Center, Boston, Massachusetts [240, e31]

Darwin L. Conwell, MD
Associate Professor of Medicine, Harvard Medical School; Associate Physician, Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts [312, 313]

Michael J. Corbel, PhD, DSc, FRCPath
Head, Division of Bacteriology, National Institute for Biological Standards and Control, Hertfordshire, United Kingdom [157]

William Edward Corcoran, V, MD
Clinical Instructor, Harvard Medical School; Cardiothoracic Fellow, Department of Anesthesiology, Perioperative, and Pain Medicine, Brigham and Women's Hospital, Boston, Massachusetts [e54]

Kathleen E. Corey, MD, MPH
Clinical and Research Fellow, Harvard Medical School; Fellow, Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts [43]

Lawrence Corey, MD
Professor of Medicine and Laboratory Medicine and Head, Virology Division, Department of Laboratory Medicine, University of Washington; Head, Program in Infectious Diseases, Fred Hutchinson Cancer Research Center, Seattle, Washington [179]

Felicia Cosman, MD
Professor of Clinical Medicine, Columbia University College of Physicians and Surgeons, New York [354]

Mark A. Creager, MD
Professor of Medicine, Harvard Medical School; Simon C. Fireman Scholar in Cardiovascular Medicine; Director, Vascular Center, Brigham and Women's Hospital, Boston, Massachusetts [248, 249]

Leslie J. Crofford, MD

Gloria W. Singletary Professor of Internal Medicine; Chief, Division of Rheumatology, University of Kentucky, Lexington, Kentucky [335]

Jennifer M. Croswell, MD, MPH
Acting Director, Office of Medical Applications of Research, National Institutes of Health, Bethesda, Maryland [82]

Philip E. Cryer, MD
Irene E. and Michael M. Karl Professor of Endocrinology and Metabolism in Medicine, Washington University School of Medicine; Physician, Barnes-Jewish Hospital, St. Louis, Missouri [345]

David Cunningham, MD, FRCP
Professor of Cancer Medicine, Royal Marsden NHS Foundation Trust, London and Sutton, United Kingdom [93]

John J. Cush, MD
Director of Clinical Rheumatology, Baylor Research Institute, Dallas, Texas [331]

Charles A. Czeisler, MD, PhD, FRCP
Baldino Professor of Sleep Medicine; Director, Division of Sleep Medicine, Harvard Medical School; Chief, Division of Sleep Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [27]

Marinos C. Dalakas, MD, FAAN
Professor of Neurology, Department of Pathophysiology, National University of Athens Medical School, Athens, Greece [388]

Josep Dalmau, MD, PhD
ICREA Research Professor, Institute for Biomedical Investigations, August Pi i Sunyer (IDIBAPS)/Hospital Clinic, Department of Neurology, University of Barcelona, Barcelona, Spain; Adjunct Professor of Neurology University of Pennsylvania, Philadelphia, Pennsylvania [101]

Daniel F. Danzl, MD
University of Louisville, Department of Emergency Medicine, Louisville, Kentucky [19]

Robert B. Daroff, MD
Professor and Chair Emeritus, Department of Neurology, Case Western Reserve University School of Medicine; University Hospitals–Case Medical Center, Cleveland, Ohio [21]

Charles E. Davis, MD
Professor of Pathology and Medicine, Emeritus, University of California, San Diego School of Medicine; Director Emeritus, Microbiology, University of California, San Diego Medical Center, San Diego, California [e25]

Stephen N. Davis, MBBS, FRCP
Theodore E. Woodward Professor and Chairman, Department of Medicine, University of Maryland School of Medicine; Physician-in-Chief, University of Maryland Medical Center, Baltimore, Maryland [345]

Lisa M. DeAngelis, MD
Professor of Neurology, Weill Cornell Medical College; Chair, Department of Neurology, Memorial Sloan-Kettering Cancer Center, New York, New York [379]

John Del Valle, MD
Professor and Senior Associate Chair of Medicine, Department of Internal Medicine, University of Michigan School of Medicine, Ann Arbor, Michigan [293]

Marie B. Demay, MD
Professor of Medicine, Harvard Medical School; Physician, Massachusetts General Hospital, Boston, Massachusetts [352]

Bradley M. Denker, MD
Associate Professor, Harvard Medical School; Physician, Department of Medicine, Brigham and Women's Hospital; Chief of Nephrology, Harvard Vanguard Medical Associates, Boston, Massachusetts [44]

David W. Denning, MB BS, FRCP, FRCPath
Professor of Medicine and Medical Mycology; Director, National Aspergillosis Centre, The University of Manchester and Wythenshawe Hospital, Manchester, United Kingdom [204]

Robert J. Desnick, MD, PhD
Dean for Genetics and Genomics; Professor and Chairman, Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine of New York University, New York, New York [358]

Richard A. Deyo, MD, MPH
Kaiser Permanente Professor of Evidence-Based Family Medicine, Department of Family Medicine, Department of Medicine, Department of Public Health and Preventive Medicine, Center for Research in Occupational and Environmental Toxicology, Oregon Health and Science University; Clinical Investigator, Kaiser Permanente Center for Health Research, Portland, Oregon [15]

Betty Diamond, MD
The Feinstein Institute for Medical Research, North Shore LIJ Health System; Center for Autoimmunity and Musculoskeletal Diseases, Manhasset, New York [318]

Jules L. Dienstag, MD
Carl W. Walter Professor of Medicine and Dean for Medical Education, Harvard Medical School; Physician, Gastrointestinal Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts [304, 305, 306, 310, e38]

William P. Dillon, MD
Elizabeth Guillaumin Professor of Radiology, Neurology and Neurosurgery; Executive Vice-Chair, Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, California [368, e44]

Charles A. Dinarello, MD
Professor of Medicine, Division of Infectious Diseases, University of Colorado School of Medicine, Aurora, Colorado [16]

Raphael Dolin, MD
Maxwell Finland Professor of Medicine (Microbiology and Molecular Genetics), Harvard Medical School; Beth Israel Deaconess Medical Center; Brigham and Women's Hospital, Boston, Massachusetts [178, 186, 187]

Richard L. Doty, PhD
Professor, Department of Otorhinolaryngology: Head and Neck Surgery; Director, Smell and Taste Center, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania [29]

Neil J. Douglas, MD, MB ChB, DSc, Hon MD, FRCPE
Professor of Respiratory and Sleep Medicine, University of Edinburgh, Edinburgh, Scotland, United Kingdom [265]

Daniel B. Drachman, MD
Professor of Neurology and Neuroscience, W. W. Smith Charitable Trust Professor of Neuroimmunology, Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland [386]

David F. Driscoll, PhD
Associate Professor of Medicine, University of Massachusetts Medical School, Worchester, Massachusetts [76]

Thomas D. DuBose, Jr., MD, MACP
Tinsley R. Harrison Professor and Chair, Internal Medicine; Professor of Physiology and Pharmacology, Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina [47, e15]

J. Stephen Dumler, MD
Professor, Division of Medical Microbiology, Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland [174]

Andrea Dunaif, MD
Charles F. Kettering Professor of Endocrinology and Metabolism; Vice-Chair for Research, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [6]

Samuel C. Durso, MD, MBA
Mason F. Lord Professor of Medicine; Director, Division of Geriatric Medicine and Gerontology, Johns Hopkins University School of Medicine, Baltimore, Maryland [32, e12]

Janice Dutcher, MD
Department of Oncology, New York Medical College, Montefiore, Bronx, New York [276]

Mark S. Dworkin, MD, MPH&TM
Associate Professor, Division of Epidemiology and Biostatistics, University of Illinois at Chicago School of Public Health, Chicago, Illinois [172]

Johanna Dwyer, DSc, RD
Professor of Medicine (Nutrition), Friedman School of Nutrition Science and Policy, Tufts University School of Medicine; Director, Frances Stern Nutrition Center, Tufts Medical Center, Boston, Massachusetts [73]

Jeffery S. Dzieczkowski, MD
Physician, St. Alphonsus Regional Medical Center; Medical Director, Coagulation Clinic, Saint Alphonsus Medical Group, International Medicine and Travel Medicine, Boise, Idaho [113]

Kim A. Eagle, MD
Albion Walter Hewlett Professor of Internal Medicine; Director, Cardiovascular Center, University of Michigan Health System, Ann Arbor, Michigan [8]

Robert H. Eckel, MD
Professor of Medicine, Division of Endocrinology, Metabolism and Diabetes, Division of Cardiology; Professor of Physiology and Biophysics, Charles A. Boettcher, II Chair in Atherosclerosis, University of

Professor of Physiology and Biophysics, Charles A. Boucher II Chair in Atherosclerosis, University of Colorado School of Medicine, Anschutz Medical Campus, Director Lipid Clinic, University of Colorado Hospital, Aurora, Colorado [242]

John E. Edwards, Jr,. MD
Chief, Division of Infectious Diseases, Harbor/University of California, Los Angeles (UCLA) Medical Center, Torrance, California; Professor of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California [198, 203]

David A. Ehrmann, MD
Professor of Medicine, The University of Chicago, Chicago, Illinois [49]

Andrew J. Einstein, MD, PhD
Assistant Professor of Clinical Medicine, Columbia University College of Physicians and Surgeons; Department of Medicine, Division of Cardiology, Department of Radiology, Columbia University Medical Center and New York-Presbyterian Hospital, New York, New York [Appendix]

Ezekiel J. Emanuel, MD, PhD
Chief, Department of Clinical Bioethics, National Institutes of Health, Bethesda, Maryland [9]

Joey D. English, MD
Assistant Clinical Professor, Department of Neurology, Univeristy of California, San Francisco, San Francisco, California [370]

John W. Engstrom, MD
Betty Anker Fife Distinguished Professor of Neurology; Neurology Residency Program Director; Clinical Chief of Service, University of California, San Francisco, San Francisco, California [15, 375]

Moshe Ephros, MD
Senior Lecturer, Faculty of Medicine, Technion—Israel Institute of Technology; Pediatric Infectious Disease Unit, Carmel Medical Center; Haifa, Israel [160]

Jonathan A. Epstein, MD, DTMH
William Wikoff Smith Professor of Medicine; Chairman, Department of Cell and Developmental Biology; Scientific Director, Cardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania [224]

Tim Evans, MD, PhD
Assistant Director-General, Information, Evidence, and Research, World Health Organization, Geneva, Switzerland [e1]

Christopher Fanta, MD
Associate Professor of Medicine, Harvard Medical School; Member, Pulmonary and Critical Care Division, Brigham and Women's Hospital, Boston, Massachusetts [34]

Paul Farmer, MD, PhD
Kolokotrones University Professor, Harvard University; Chair, Department of Global Health and Social Medicine, Harvard Medical School; Chief, Division of Global Health Equity, Brigham and Women's Hospital; Co-Founder, Partners in Health, Boston, Massachusetts [2]

Anthony S. Fauci, MD, DSc (Hon), DM&S (Hon), DHL (Hon), DPS (Hon), DLM (Hon), DMS (Hon)
Chief, Laboratory of Immunoregulation; Director, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [1, 188, 189, 221, 314, 326, e40]

Murray J. Favus, MD
Professor, Department of Medicine, Section of Endocrinology, Diabetes and Metabolism, Director Bone Program, University of Chicago Pritzker School of Medicine, Chicago, Illinois [287, 355]

David P. Faxon, MD
Senior Lecturer, Harvard Medical School; Vice Chair of Medicine for Strategic Planning, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [230, 246, e33]

David T. Felson, MD, MPH
Professor of Medicine and Epidemiology; Chair, Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Massachusetts [332]

Luigi Ferrucci, MD, PhD
Director, Baltimore Longitudinal Study of Aging National Institute of Health, Baltimore, Maryland [72]

Howard L. Fields, MD, PhD
Professor of Neurology, University of California, San Francisco, San Francisco, California [11]

Gregory A. Filice, MD
Professor of Medicine, University of Minnesota; Chief, Infectious Disease Section, Veterans Affairs Medical Center, Minneapolis, Minnesota [162]

Robert Finberg, MD
Chair, Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts [86, 132]

Joyce Fingeroth, MD

Joyce Fingeroth, MD
Associate Professor of Medicine, Harvard Medical School, Boston, Massachusetts [132]

Alain Fischer, MD, PhD
University Paris Descartes, Inserm Unit 768; Immunology and Pediatric Hematology Unit, Necker Children's Hospital, Paris, France [316, e39]

Jeffrey S. Flier, MD
Caroline Shields Walker Professor of Medicine and Dean, Harvard Medical School, Boston, Massachusetts [77]

Agnes B. Fogo, MD
John L. Shapiro Professor of Pathology; Professor of Medicine and Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee [e14]

Larry C. Ford, MD
Associate Researcher, Divisions of Clinical Epidemiology and Infectious Diseases, University of Utah, Salt Lake City, Utah [31]

Jane E. Freedman, MD
Professor, Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts [117]

Roy Freeman, MBCHB
Professor of Neurology, Harvard Medical School, Boston, Massachusetts [20]

Gyorgy Frendl, MD, PhD
Assistant Professor of Anesthesia and Critical Care, Harvard Medical School; Director of Research, Surgical Critical Care, Brigham and Women's Hospital, Boston, Massachusetts [e54]

Carl E. Freter, MD, PhD
Professor, Department of Internal Medicine, Division of Hematology/Medical Oncology, University of Missouri; Ellis Fischel Cancer Center, Columbia, Missouri [102]

Lawrence S. Friedman, MD
Professor of Medicine, Harvard Medical School; Professor of Medicine, Tufts University School of Medicine; Assistant Chief of Medicine, Massachusetts General Hospital, Boston, Massachusetts; Chair, Department of Medicine, Newton-Wellesley Hospital, Newton, Massachusetts [43]

Sonia Friedman, MD
Assistant Professor of Medicine, Harvard Medical School, Boston, Massachusetts [295]

Anne L. Fuhlbrigge, MD, MS
Assistant Professor, Harvard Medical School; Pulmonary and Critical Care Division, Brigham and Women's Hospital, Boston, Massachusetts [253]

Andre Furtado, MD
Associate Specialist at the Department of Radiology, Neuroradiology Section, University of California, San Francisco, San Francisco, California [e44]

Robert F. Gagel, MD
Professor of Medicine and Head, Division of Internal Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas [351]

Nicholas B. Galifianakis, MD, MPH
Assistant Clinical Professor, Surgical Movement Disorders Center, Department of Neurology, University of California, San Francisco, San Francisco, California [e8]

John I. Gallin, MD
Director, Clinical Center, National Institutes of Health, Bethesda, Maryland [60]

Charlotte A. Gaydos, DrPh, MPH, MS
Professor of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland [176]

J. Michael Gaziano, MD, MPH
Professor of Medicine, Harvard Medical School; Chief, Division of Aging, Brigham and Women's Hospital; Director, Massachusetts Veterans Epidemiology Center, Boston VA Healthcare System, Boston, Massachusetts [225]

Thomas A. Gaziano, MD, MSc
Assistant Professor, Harvard Medical School; Assistant Professor, Health Policy and Management, Center for Health Decision Sciences, Harvard School of Public Health; Associate Physician in Cardiovascular Medicine, Department of Cardiology, Brigham and Women's Hospital, Boston, Massachusetts [225]

Susan L. Gearhart, MD
Assistant Professor of Colorectal Surgery and Oncology, The Johns Hopkins University School of Medicine, Baltimore, Maryland [297, 298]

Robert H. Gelber, MD
Clinical Professor of Medicine and Dermatology, University of California, San Francisco, San

Clinical Professor of Medicine and Dermatology, University of California, San Francisco, San Francisco, California [166]

Jeffrey A. Gelfand, MD
Clinical Professor of Medicine, Harvard Medical School; Physician, Massachusetts General Hospital, Boston, Massachusetts [18, 211]

Alfred L. George, Jr., MD
Professor of Medicine and Pharmacology; Chief, Division of Genetic Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee [277]

Dale N. Gerding, MD
Professor of Medicine, Loyola University Chicago Stritch School of Medicine; Associate Chief of Staff for Research and Development, Edward Hines, Jr. VA Hospital, Hines, Illinois [129]

Alicia K. Gerke, MD
Associate, Division of Pulmonary and Critical Care Medicine, University of Iowa, Iowa City, Iowa [255]

Michael Geschwind, MD, PhD
Associate Professor of Neurology, Memory and Aging Center, University of California, San Francisco, School of Medicine, San Francisco, California [e8]

Marc G. Ghany, MD, MHSc
Staff Physician, Liver Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [301]

Michael Giladi, MD, MSc
Associate Professor of Medicine, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel [160]

Bruce C. Gilliland,† MD
Professor of Medicine and Laboratory Medicine, University of Washington School of Medicine, Seattle, Washington [337]

Roger I. Glass, MD, PhD
Director, Fogarty International Center, Bethesda, Maryland [190]

Eli Glatstein, MD
Professor and Vice Chairman, Department of Radiation Oncology, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania [223]

Peter J. Goadsby, MD, PhD, DSc, FRACP FRCP
Professor of Neurology, University of California, San Francisco, California; Honorary Consultant Neurologist, Hospital for Sick Children, London, United Kingdom [14]

Ary L. Goldberger, MD
Professor of Medicine, Harvard Medical School; Wyss Institute for Biologically Inspired Engineering, Harvard University; Beth Israel Deaconess Medical Center, Boston, Massachusetts [228, e28, e30]

David Goldblatt, PhD, MBChB, FRCP, FRCPCH
Professor of Vaccinology and Immunology; Consultant in Paediatric Immunology; Director of Clinical Research and Development; Director, NIHR Biomedical Research Centre, Institute of Child Health; University College London; Great Ormond Street Hospital for Children NHS Trust, London, United Kingdom [134]

Samuel Z. Goldhaber, MD
Professor of Medicine, Harvard Medical School; Director, Venous Thromboembolism Research Group, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts [262]

Ralph Gonzales, MD, MSPH
Professor of Medicine, University of California, San Francisco, San Francisco, California [31]

Douglas S. Goodin, MD
Professor of Neurology, University of California, San Francisco School of Medicine, San Francisco, California [380]

Craig E. Gordon, MD, MS
Assistant Professor of Medicine, Boston University School of Medicine; Attending, Section of Nephrology, Boston Medical Center, Boston, Massachusetts [284]

Jeffrey I. Gordon, MD
Dr. Robert J. Glaser Distinguished University Professor; Director, Center for Genome Sciences, Washington University School of Medicine, St. Louis, Missouri [64]

Maria Luisa Gorno-Tempini, MD, PhD
Associate Professor of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [e10]

Gregory A. Grabowski, MD
Professor, Departments of Pediatrics, and Molecular Genetics, Biochemistry, and Microbiology; University of Cincinnati College of Medicine, A. Graeme Mitchell Chair in Human Genetics; Director,

University of Cincinnati College of Medicine; A. Graham Mitchell Chair in Human Genetics; Director, Division of Human Genetics, Cincinnati Children's Hosptial Medical Center, Cincinnati, Ohio [361]

Alexander R. Green, MD, MPH
Assistant Professor of Medicine, Harvard Medical School; Associate Director, The Disparities Solutions Center, Massachusetts General Hospital, Boston, Massachusetts [e4]

Norton J. Greenberger, MD
Clinical Professor of Medicine, Harvard Medical School; Senior Physician, Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts [311, 312, 313]

Daryl R. Gress, MD, FAAN, FCCM
Professor of Neurocritical Care and Stroke; Professor of Neurology, University of California, San Francisco, San Francisco, California [275]

Rasim Gucalp, MD
Professor of Clinical Medicine, Albert Einstein College of Medicine; Associate Chairman for Educational Programs, Department of Oncology; Director, Hematology/Oncology Fellowship, Montefiore Medical Center, Bronx, New York [276]

Kalpana Gupta, MD, MPH
Associate Professor, Department of Medicine, Boston University School of Medicine; Chief, Section of Infectious Diseases, VA Boston Healthcare System, Boston, Massachusetts [288]

John G. Haaga, MD
Deputy Associate Director, Behavioral and Social Research Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland [70]

Chadi A. Hage, MD
Assistant Professor of Medicine, Pulmonary–Critical Care and Infectious Diseases, Roudebush VA Medical Center; Indiana University, Indianapolis, Indiana [199]

Bevra Hannahs Hahn, MD
Professor of Medicine, University of California, Los Angeles, David Geffen School of Medicine, Los Angeles, California [319]

Janet E. Hall, MD, MSc
Professor of Medicine, Harvard Medical School; Associate Physician, Massachusetts General Hospital, Boston, Massachusetts [50, 347]

Jesse B. Hall, MD, FCCP
Professor of Medicine, Anesthesia and Critical Care; Chief, Section of Pulmonary and Critical Care Medicine, University of Chicago, Chicago, Illinois [267]

Scott A. Halperin, MD
Professor of Pediatrics and Microbiology and Immunology; CIHR/Wyeth Chair in Clinical Vaccine Research; Head, Pediatric Infectious Diseases; Director, Canadian Center for Vaccinology, Dalhousie University, Halifax, Nova Scotia, Canada [148]

R. Doug Hardy, MD
Associate Professor of Internal Medicine and Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas [175]

Raymond C. Harris, MD
Ann and Roscoe R. Robinson Professor of Medicine; Chief, Division of Nephrology, Vanderbilt University School of Medicine, Nashville, Tennessee [278]

William L. Hasler, MD
Professor of Internal Medicine, Division of Gastroenterology, University of Michigan Health System, Ann Arbor, Michigan [39, 290]

Terry Hassold, PhD
Eastlick Distinguished Professor; Director, Center for Reproductive Biology, Washington State University School of Molecular Biosciences, Pullman, Washington [62]

Stephen L. Hauser, MD
Robert A. Fishman Distinguished Professor and Chairman, Department of Neurology, University of California, San Francisco, San Francisco, California [1, 366, 367, 376, 377, 380, 385, e46]

Barton F. Haynes, MD
Frederic M. Hanes Professor of Medicine and Immunology, Departments of Medicine and Immunology; Director, Duke Human Vaccine Institute, Duke University School of Medicine, Durham, North Carolina [314]

Douglas C. Heimburger, MD, MS
Professor of Medicine; Associate Director for Education and Training, Vanderbilt Institute for Global Health, Vanderbilt University School of Medicine, Nashville, Tennessee [75]

J. Claude Hemphill, III, MD, MAS
Professor of Clinical Neurology and Neurological Surgery, Department of Neurology, University of California, San Francisco; Director of Neurocritical Care, San Francisco General Hospital, San

California, San Francisco; Director of Neurocritical Care, San Francisco General Hospital, San Francisco, California [275]

Patrick H. Henry, MD
Clinical Adjunct Professor of Medicine, University of Iowa, Iowa City, Iowa [59]

Katherine A. High, MD
Investigator, Howard Hughes Medical Institute; William H. Bennett Professor of Pediatrics, University of Pennsylvania School of Medicine; Director, Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania [68, 116]

Ikuo Hirano, MD
Professor of Medicine, Division of Gastroenterology and Hepatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [38, 292]

Martin S. Hirsch, MD
Professor of Medicine, Harvard Medical School; Professor of Immunology and Infectious Diseases, Harvard School of Public Health; Physician, Massachusetts General Hospital, Cambridge, Massachusetts [182]

Helen H. Hobbs, MD
Professor of Internal Medicine and Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, Texas; Investigator, Howard Hughes Medical Institute, Chevy Chase, Maryland [356]

Judith S. Hochman, MD
Harold Snyder Family Professor of Cardiology; Clinical Chief, Leon Charney Division of Cardiology; Co-Director, NYU-HHC Clinical and Translational Science Institute; Director, Cardiovascular Clinical Research Center, New York University School of Medicine, New York, New York [272]

A. Victor Hoffbrand, DM
Professor Emeritus of Haematology, University College, London; Honorary Consultant Haematologist, Royal Free Hospital, London, United Kingdom [105]

David M. Hoganson, MD
Laboratory for Tissue Engineering and Organ Fabrication Center for Regenerative Medicine, Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts [69]

Charles W. Hoge, MD
Senior Scientist and Staff Psychiatrist, Center for Psychiatry and Neuroscience, Walter Reed Army Institute of Research and Water Reed Army Medical Center, Silver Spring, Maryland [e48]

Elizabeth L. Hohmann, MD
Associate Professor of Medicine and Infectious Diseases, Harvard Medical School; Massachusetts General Hospital, Boston, Massachusetts [139]

Steven M. Holland, MD
Chief, Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [60, 167]

King K. Holmes, MD, PhD
Chair, Global Health; Professor of Medicine and Global Health; Adjunct Professor, Epidemiology; Director, Center for AIDS and STD; University of Washington School of Medicine; Head, Infectious Diseases Section, Harborview Medical Center, Seattle, Washington [130]

Jay H. Hoofnagle, MD
Director, Liver Diseases Research Branch, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [301]

Robert Hopkin, MD
Associate Professor of Clinical Pediatrics, University of Cincinnati College of Medicine; Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio [361]

Leora Horn, MD, MSc
Division of Hematology and Medical Oncology, Vanderbilt University School of Medicine, Nashville, Tennessee [89]

Jonathan C. Horton, MD, PhD
William F. Hoyt Professor of Neuro-ophthalmology, Professor of Ophthalmology, Neurology and Physiology, University of California, San Francisco School of Medicine, San Francisco, California [28]

Howard Hu, MD
Environmental Health Sciences, University of Michigan Schools of Public Health and Medicine, Ann Arbor, Michigan [e49]

Gary W. Hunninghake, MD
Professor, Division of Pulmonary and Critical Care Medicine, University of Iowa, Iowa City, Iowa [255]

Sharon A. Hunt, MD, FACC
Professor, Division of Cardiovascular Medicine, Stanford University, Palo Alto, California [235]

Charles G. Hurst, MD

Charles G. Hurst, MD
Chief, Chemical Casualty Care Division, United States Medical Research Institute of Chemical Defense, APG-Edgewood Area, Maryland [222]

Ashraf S. Ibrahim, PhD
Associate Professor of Medicine, Geffen School of Medicine, University of California, Los Angeles (UCLA); Division of Infectious Diseases, Los Angeles Biomedical Research Institute at Harbor–UCLA Medical Center, Torrance, California [205]

David H. Ingbar, MD
Professor of Medicine, Pediatrics, and Physiology; Director, Pulmonary Allergy, Critical Care and Sleep Division, University of Minnesota School of Medicine, Minneapolis, Minnesota [272]

Alan C. Jackson, MD, FRCPC
Professor of Medicine (Neurology) and Medical Microbiology, University of Manitoba; Section Head of Neurology, Winnipeg Regional Health Authority, Winnipeg, Manitoba, Canada [195]

Lisa A. Jackson, MD, MPH
Senior Investigator, Group Health Research Institute; Research Professor, Department of Epidemiology; Adjunct Professor, Department of Medicine, University of Washington, Seattle, Washington [122]

Richard F. Jacobs, MD
Robert H. Fiser, Jr., MD Endowed Chair in Pediatrics; Professor and Chairman, Department of Pediatrics, University of Arkansas for Medical Sciences; President, Arkansas Children's Hospital Research Institute, Little Rock, Arkansas [158]

J. Larry Jameson, MD, PhD
Robert G. Dunlop Professor of Medicine; Dean, University of Pennsylvania School of Medicine; Executive Vice President of the University of Pennsylvania for the Health System, Philadelphia, Pennsylvania [1, 61, 63, 80, 100, 338, 339, 341, 346, 349, e41]

Robert T. Jensen, MD
Digestive Diseases Branch, National Institute of Diabetes; Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [350]

David H. Johnson, MD, FACP
Donald W. Seldin Distinguished Chair in Internal Medicine; Professor and Chairman, Department of Internal Medicine, University of Texas Southwestern Medical School, Dallas, Texas [89]

James R. Johnson, MD
Professor of Medicine, University of Minnesota, Minneapolis, Minnesota [149]

Stuart Johnson, MD
Associate Professor of Medicine, Loyola University Chicago Stritch School of Medicine; Staff Physician, Edward Hines, Jr. VA Hospital, Hines, Illinois [129]

S. Claiborne Johnston, MD, PhD
Professor of Neurology and Epidemiology, University of California, San Francisco School of Medicine, San Francisco, California [370]

S. Andrew Josephson, MD
Associate Professor, Department of Neurology; Director, Neurohospitalist Program, University of California, San Francisco, San Francisco, California [25, e47]

Harald Jüppner, MD
Professor of Pediatrics, Endocrine Unit and Pediatric Nephrology Unit, Massachusetts General Hospital, Boston, Massachusetts [353]

Peter J. Kahrilas, MD
Gilbert H. Marquardt Professor in Medicine, Division of Gastroenterology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [38, 292]

Gail Kang, MD
Assistant Clinical Professor of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [e8]

Marshall M. Kaplan, MD
Professor of Medicine, Tufts University School of Medicine, Boston, Massachusetts [42, 302]

Adolf W. Karchmer, MD
Professor of Medicine, Harvard Medical School; Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Boston, Massachusetts [124]

Dennis L. Kasper, MD, MA (Hon)
William Ellery Channing Professor of Medicine and Professor of Microbiology and Molecular Genetics, Harvard Medical School; Director, Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [1, 119, 121, 127, 146, 164]

Lloyd H. Kasper, MD
Professor of Medicine (Neurology) and Microbiology and Immunology, Dartmouth Medical School,

Professor of Medicine (Neurology) and Microbiology and Immunology, Dartmouth Medical School, Lebanon, New Hampshire [214]

Daniel L. Kastner, MD, PhD
Scientific Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland [330]

Carol A. Kauffman, MD
Professor of Internal Medicine, University of Michigan Medical School; Chief, Infectious Diseases Section, Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, Michigan [206]

Elaine T. Kaye, MD
Assistant Clinical Professor of Dermatology, Harvard Medical School, Boston, Massachusetts [17, e7]

Kenneth M. Kaye, MD
Associate Professor of Medicine, Harvard Medical School, Boston, Massachusetts [17, e7]

John A. Kessler, MD
Professor and Chair, Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, Illinois [67]

Jay S. Keystone, MD, FRCPC, MSc (CTM)
Professor of Medicine, University of Toronto, Toronto, Ontario, Canada [123]

Sundeep Khosla, MD
Professor of Medicine and Physiology, College of Medicine, Mayo Clinic, Rochester, Minnesota [46]

Elliott Kieff, MD, PhD
Harriet Ryan Albee Professor, Harvard Medical School; Chief, Infectious Diseases Division, Brigham and Women's Hospital, Boston, Massachusetts [177]

Anthony A. Killeen, MD, PhD
Associate Professor; Director of Clinical Laboratories, University of Minnesota Medical Center, Minneapolis, Minnesota [e53]

Jim Yong Kim, MD, PhD
Chair, Department of Global Health and Social Medicine, Harvard Medical School; Director, François-Xavier Bagnoud Center for Health and Human Rights, Harvard School of Public Health; Chief, Division of Global Health Equity, Brigham and Women's Hospital, Boston, Massachusetts [2]

Kami Kim, MD
Professor of Medicine (Infectious Diseases) and of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York [214]

Lindsay King, MD
Clinical and Research Fellow, Department of Medicine, Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts [e56]

Talmadge E. King, Jr., MD
Julius R. Krevans Distinguished Professor in Internal Medicine; Chair, Department of Medicine, University of California, San Francisco, San Francisco, California [261]

Louis V. Kirchhoff, MD, MPH
Professor of Internal Medicine (Infectious Diseases) and Epidemiology, Department of Internal Medicine, The University of Iowa, Iowa City, Iowa [213]

Priya S. Kishnani, MD
Professor of Pediatrics, Duke University Medical Center, Durham, North Carolina [362]

Rob Knight, PhD
Assistant Professor, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado [64]

Minoru S. H. Ko, MD, PhD
Senior Investigator and Chief, Developmental Genomics and Aging Section, Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, Maryland [65]

Barbara Konkle, MD
Professor of Medicine, Hematology, University of Washington; Director, Translational Research, Puget Sound Blood Center, Seattle, Washington [58, 115]

Peter Kopp, MD
Associate Professor, Division of Endocrinology, Metabolism and Molecular Science, Northwestern University Feinberg School of Medicine, Chicago, Illinois [61]

Walter J. Koroshetz, MD
National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland [382]

Thomas R. Kosten, MD
Baylor College of Medicine; Veteran's Administration Medical Center, Houston, Texas [393]

Baylor College of Medicine; Veterans Administration Medical Center, Houston, Texas [306]

Theodore A. Kotchen, MD
Professor Emeritus, Department of Medicine; Associate Dean for Clinical Research, Medical College of Wisconsin, Milwaukee, Wisconsin [247]

Phyllis E. Kozarsky, MD
Professor of Medicine and Infectious Diseases, Emory University School of Medicine, Atlanta, Georgia [123]

Barnett S. Kramer, MD, MPH
Associate Director for Disease Prevention, Office of Disease Prevention, National Institutes of Health, Bethesda, Maryland [82]

Joel Kramer, PsyD
Clinical Professor of Neuropsychology in Neurology; Director of Neuropsychology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [e10]

Stephen M. Krane, MD
Persis, Cyrus and Marlow B. Harrison Distinguished Professor of Medicine, Harvard Medical School; Massachusetts General Hospital, Boston, Massachusetts [352]

Alexander Kratz, MD, PhD, MPH
Associate Professor of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons; Director, Core Laboratory, Columbia University Medical Center, New York, New York [Appendix]

John P. Kress, MD
Associate Professor of Medicine, Section of Pulmonary and Critical Care, University of Chicago, Chicago, Illinois [267]

Patricia Kritek, MD, EdM
Associate Professor, Division of Pulmonary and Critical Care Medicine, University of Washington, Seattle, Washington [34, 251, e34]

Henry M. Kronenberg, MD
Professor of Medicine, Harvard Medical School; Chief, Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts [352]

Robert F. Kushner, MD, MS
Professor of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [78]

Loren Laine, MD
Professor of Medicine, University of Southern California Keck School of Medicine, Los Angeles, California [41]

Anil K. Lalwani, MD
Professor, Departments of Otolaryngology, Pediatrics, and Physiology and Neuroscience, New York University School of Medicine, New York, New York [30]

H. Clifford Lane, MD
Clinical Director; Director, Division of Clinical Research; Deputy Director, Clinical Research and Special Projects; Chief, Clinical and Molecular Retrovirology Section, Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [189, 221]

Carol A. Langford, MD, MHS
Director, Center for Vasculitis Care and Research, Department of Rheumatic and Immunologic Diseases, Cleveland Clinic, Cleveland, Ohio [326, 328, 336, 337, e40]

Regina C. LaRocque, MD
Assistant Professor of Medicine, Harvard Medical School; Assistant Physician, Massachusetts General Hospital, Boston, Massachusetts [128]

Wei C. Lau, MD
Associate Professor, Medical Director, Cardiovascular Center Operating Rooms; Director, Adult Cardiovascular and Thoracic Anesthesiology, University of Michigan Health System, Ann Arbor, Michigan [8]

Leslie P. Lawley, MD
Assistant Professor, Department of Dermatology, School of Medicine, Emory University, Atlanta, Georgia [52]

Thomas J. Lawley, MD
William P. Timmie Professor of Dermatology, Dean, Emory University School of Medicine, Atlanta, Georgia [51, 52, 54, e16]

Thomas H. Lee, MD, MSc
Professor of Medicine, Harvard Medical School; Network President, Partners Healthcare System, Boston, Massachusetts [12]

Jane A. Leopold, MD
Associate Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [230, e33]

Nelson Leung, MD
Associate Professor of Medicine, Department of Nephrology and Hypertension, Division of Hematology, Mayo Clinic, Rochester, Minnesota [286]

Bruce D. Levy, MD
Associate Professor of Medicine, Harvard Medical School; Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts [268]

Julia B. Lewis, MD
Professor, Department of Medicine, Division of Nephrology, Vanderbilt University Medical Center, Nashville, Tennessee [283]

Peter Libby, MD
Mallinckrodt Professor of Medicine, Harvard Medical School; Chief, Cardiovascular Medicine, Brigham and Women's Hospital, Boston, Massachusetts [224, 241, e32]

Richard W. Light, MD
Professor of Medicine, Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University, Nashville, Tennessee [263]

Julie Lin, MD, MPH
Assistant Professor of Medicine, Harvard Medical School, Boston, Massachusetts [44]

Robert Lindsay, MD, PhD
Chief, Internal Medicine; Professor of Clinical Medicine, Helen Hayes Hospital, West Haverstraw, New York [354]

Marc E. Lippman, MD, MACP
Kathleen and Stanley Glaser Professor; Chairman, Department of Medicine, Deputy Director, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, Florida [90]

Peter E. Lipsky, MD
Charlottesville, Virginia [318, 331]

Kathleen D. Liu, MD, PhD, MAS
Assistant Professor, Divisions of Nephrology and Critical Care Medicine, Departments of Medicine and Anesthesia, University of California, San Francisco, San Francisco, California [281]

Bernard Lo, MD
Professor of Medicine; Director, Program in Medical Ethics, University of California, San Francisco, San Francisco, California [e5]

Dan L. Longo, MD
Lecturer, Department of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital; Deputy Editor, New England Journal of Medicine, Boston, Massachusetts; Adjunct Investigator, National Institute on Aging, National Institutes of Health, Baltimore, Maryland [1, 57, 59, 66, 81, 84, 85, 100, 102, 110, 111, 188, e6, e17, e20, e21]

Nicola Longo, MD, PhD, MACP
Professor of Pediatrics; Chief, Division of Medical Genetics, Department of Pediatrics, University of Utah, Salt Lake City, Utah [364, 365]

Joseph Loscalzo, MD, PhD
Hersey Professor of the Theory and Practice of Medicine, Harvard Medical School; Chairman, Department of Medicine; Physician-in-Chief, Brigham and Women's Hospital, Boston, Massachusetts [1, 35, 36, 37, 117, 224, 226, 227, 237, 238, 243, 245, 248, 249, e13, e19]

Phillip A. Low, MD
Robert D. and Patricia E. Kern Professor of Neurology, Mayo Clinic College of Medicine, Rochester, Minnesota [375]

Daniel H. Lowenstein, MD
Dr. Robert B. and Mrs. Ellinor Aird Professor of Neurology; Director, Epilepsy Center, University of California, San Francisco, San Francisco, California [367, 369, e42]

Elyse E. Lower, MD
Medical Oncology and Hematology, University of Cincinnati; Oncology Hematology Care, Inc., Cincinnati, Ohio [329]

Franklin D. Lowy, MD
Professor of Medicine and Pathology, Columbia University College of Physicians and Surgeons, New York, New York [135]

Sheila A. Lukehart, PhD
Professor, Departments of Medicine and Global Health, University of Washington, Seattle, Washington [169, 170]

Lucio Luzzatto, MD, FRCP, FRCPath
Professor of Haematology, University of Genova, Scientific Director Istituto Toscano Tumori, Italy [106]

Lawrence C. Madoff, MD
Professor of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts; Director, Division of Epidemiology and Immunization, Massachusetts Department of Public Health, Jamaica Plain, Massachusetts [119, 334, e23, e24]

Emily Nelson Maher, MD
Clinical Instructor, Department of Anesthesiology, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [e57]

Adel A. F. Mahmoud, MD, PhD
Professor, Department of Molecular Biology and the Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey [219]

Ronald V. Maier, MD
Jane and Donald D. Trunkey Professor and Vice-Chair, Surgery, University of Washington; Surgeon-in-Chief, Harborview Medical Center, Seattle, Washington [270]

Mark E. Mailliard, MD
Frederick F. Paustian Professor; Chief, Division of Gastroenterology and Hepatology, Department of Internal Medicine, University of Nebraska College of Medicine, Omaha, Nebraska [307]

Hari R. Mallidi, MD
Assistant Professor of Cardiothoracic Surgery; Director of Mechanical Circulatory Support, Stanford University Medical Center, Stanford, California [235]

Hanna Mandel, MD
Director, Pediatric Metabolic Disorders, Rambam Health Care Campus, Haifa, Israel [e18]

Brian F. Mandell, MD, PhD, MACP, FACR
Professor and Chairman of Medicine, Cleveland Clinic Lerner College of Medicine; Department of Rheumatic and Immunologic Disease, Cleveland Clinic, Cleveland, Ohio [336]

Lionel A. Mandell, MD, FRCP(C), FRCP(LOND)
Professor of Medicine, McMaster University, Hamilton, Ontario, Canada [257]

Douglas L. Mann, MD
Lewin Chair and Chief, Cardiovascular Division; Professor of Medicine, Cell Biology and Physiology, Washington University School of Medicine, St. Louis, Missouri [234]

JoAnn E. Manson, MD, DrPH
Professor of Medicine and the Michael and Lee Bell Professor of Women's Health, Harvard Medical School; Chief, Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts [348]

Eleftheria Maratos-Flier, MD
Associate Professor of Medicine, Harvard Medical School; Division of Endocrinology, Beth Israel Deaconess Medical Center, Boston, Massachusetts [77]

Francis Marchlinski, MD
Professor of Medicine; Director, Cardiac Electrophysiology, University of Pennsylvania Health System, Philadelphia, Pennsylvania [233]

Guido Marcucci, MD
Professor of Medicine; John B. and Jane T. McCoy Chair in Cancer Research; Associate Director of Translational Research, Comprehensive Cancer Center, The Ohio State University College of Medicine, Columbus, Ohio [109]

Daniel B. Mark, MD, MPH
Professor of Medicine, Duke University Medical Center; Director, Outcomes Research, Duke Clinical Research Institute, Durham, North Carolina [3]

Alexander G. Marneros, MD, PhD
Assistant Professor, Department of Dermatology, Harvard Medical School Boston, Massachusetts; Cutaneous Biology Research Center, Massachusetts General Hospital, Charlestown, Massachusetts [56]

Jeanne M. Marrazzo, MD, MPH
Associate Professor of Medicine, Division of Infectious Diseases, Harborview Medical Center, Seattle, Washington [130]

Thomas Marrie, MD
Dean, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada [174]

Gary J. Martin, MD
Raymond J. Langenbach, MD Professor of Medicine; Vice Chairman for Faculty Affairs, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [4]

Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [4]

George M. Martin, MD
Professor of Pathology Emeritus, Adjunct Professor of Genome Sciences (Retired), University of Washington, Seattle, Washington; Visiting Scholar, Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California [71]

Joseph B. Martin, MD, PhD
Edward R. and Anne G. Lefler Professor, Department of Neurobiology, Harvard Medical School, Boston, Massachusetts [367]

Matthew Martinez, MD
Lehigh Valley Physician Group, Lehigh Valley Heart Specialists, Allentown, Pennsylvania [229, e29]

Susan Maslanka, PhD
Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, Georgia [141]

Robert J. Mayer, MD
Stephen B. Kay Family Professor of Medicine, Harvard Medical School, Boston, Massachusetts [91]

Alexander J. McAdam, MD, PhD
Assistant Professor of Pathology, Harvard Medical School, Children's Hospital, Boston, Massachusetts [e22]

Calvin O. McCall, MD
Associate Professor, Department of Dermatology, Virginia Commonwealth University Medical Center; Chief, Dermatology Section, Hunter Holmes McGuire Veterans Affairs Medical Center, Richmond, Virginia [52]

John F. McConville, MD
Assistant Professor of Medicine, University of Chicago, Chicago, Illinois [264]

Kevin T. McVary, MD, FACS
Professor of Urology, Department of Urology, Northwestern University Feinberg School of Medicine, Chicago, Illinois [48]

Nancy K. Mello, PhD
Professor of Psychology (Neuroscience), Harvard Medical School, Boston, Massachusetts; Director, Alcohol and Drug Abuse Research Center, McLean Hospital, Belmont, Massachusetts [394]

Shlomo Melmed, MD
Senior Vice President and Dean of the Medical Faculty, Cedars-Sinai Medical Center, Los Angeles, California [339]

Jack H. Mendelson,† MD
Professor of Psychiatry (Neuroscience), Harvard Medical School, Belmont, Massachusetts [394]

Robert O. Messing, MD
Professor, Department of Neurology; Senior Associate Director, Ernest Gallo Clinic and Research Center, University of California, San Francisco, San Francisco, California [390]

M.-Marsel Mesulam, MD
Professor of Neurology, Psychiatry and Psychology, Cognitive Neurology and Alzheimer's Disease Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois [26]

Susan Miesfeldt, MD
Mercy Hospital, Maine Centers for Cancer Medicine, Scarbrough, Maine [63]

Edgar L. Milford, MD
Associate Professor of Medicine, Harvard Medical School; Director, Tissue Typing Laboratory, Brigham and Women's Hospital, Boston, Massachusetts [282]

Bruce L. Miller, MD
AW and Mary Margaret Clausen Distinguished Professor of Neurology, University of California, San Francisco School of Medicine, San Francisco, California [25, 371, 383, e9, e10]

Samuel I. Miller, MD
Professor of Genome Sciences, Medicine, and Microbiology, University of Washington, Seattle, Washington [153]

Simon J. Mitchell, MB ChB, PhD
Associate Professor in Anesthesiology, Diving and Hyperbaric Medicine, Faculty of Medical and Health Sciences, University of Auckland; Consultant Anesthetist, Auckland City Hospital, Auckland, New Zealand [e52]

Thomas A. Moore, MD, FACP, FIDSA
Chairman, Department of Infectious Diseases, Ochsner Health System, New Orleans, Louisiana [208, e26]

Pat J. Morin, PhD
Senior Investigator, Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, Maryland [83]

Charles A. Morris, MD, MPH
Instructor in Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts [e55, e57]

William J. Moss, MD, MPH
Associate Professor, Departments of Epidemiology, International Health, and Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland [192]

Robert J. Motzer, MD
Professor of Medicine, Weill Cornell Medical College; Attending Physician, Genitourinary Oncology Service, Memorial Sloan-Kettering Cancer Center, New York, New York [94, 96]

David B. Mount, MD, FRCPC
Assistant Professor of Medicine, Harvard Medical School, Renal Division, VA Boston Healthcare System; Brigham and Women's Hospital, Boston, Massachusetts [45, e15]

Haralampos M. Moutsopoulos, MD, FACP, FRCP, Master ACR
Professor and Director, Department of Pathophysiology, Medical School, National University of Athens, Athens, Greece [320, 324, 327]

Robert S. Munford, MD
Bethesda, Maryland [271]

Nikhil C. Munshi, MD
Associate Professor of Medicine, Harvard Medical School; Associate Director, Jerome Lipper Multiple Myeloma Center, Dana Farber Cancer Institute, Boston, Massachusetts [111]

John R. Murphy, PhD
Professor of Medicine and Microbiology, Boston University School of Medicine, Boston, Massachusetts [138]

Timothy F. Murphy, MD
UB Distinguished Professor of Medicine and Microbiology, University of Buffalo, State University of New York, Buffalo, New York [145]

Barbara E. Murray, MD
J. Ralph Meadows Professor and Director, Division of Infectious Diseases, University of Texas Medical School, Houston, Texas [137]

Joseph A. Murray, MD
Professor of Medicine, Departments of Internal Medicine and Immunology, Mayo Clinic, Rochester, Minnesota [40]

Mark B. Mycyk, MD
Associate Professor, Department of Emergency Medicine, Boston University School of Medicine; Associate Professor, Department of Emergency Medicine; Rush University School of Medicine, Research Director, Division of Toxicology, Cook County Hospital, Chicago, Illinois [e50]

Robert J. Myerburg, MD
Professor, Departments of Medicine and Physiology, Division of Cardiology; AHA Chair in Cardiovascular Research, University of Miami Miller School of Medicine, Miami, Florida [273]

Hari Nadiminti, MD
Clinical Instructor, Department of Dermatology, Emory University School of Medicine, Atlanta, Georgia [87]

Edward T. Naureckas, MD
Associate Professor of Medicine, Section of Pulmonary and Critical Care Medicine, University of Chicago, Chicago, Illinois [252]

Eric G. Neilson, MD
Thomas Fearn Frist Senior Professor of Medicine and Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, Tennessee [277, 278, 283, e14]

Gerald T. Nepom, MD, PhD
Director, Benaroya Research Institute at Virginia Mason; Director, Immune Tolerance Network; Professor, University of Washington School of Medicine, Seattle, Washington [315]

Eric J. Nestler, MD, PhD
Nash Family Professor and Chair, Department of Neuroscience; Director, Friedman Brain Institute, Mount Sinai School of Medicine, New York, New York [390]

Hartmut P. H. Neumann, MD
Head, Section Preventative Medicine, Department of Nephrology and General Medicine, Albert-Ludwigs-University of Freiburg, Germany [343]

Joseph P. Newhouse, PhD
John D. MacArthur Professor of Health Policy and Management, Department of Health Care Policy, Harvard Medical School; Department of Health Policy and Management, Harvard School of Public Health, Harvard Kennedy School; Faculty of Arts and Sciences, Harvard University, Boston, Massachusetts [e3]

Jonathan Newmark, MD
Colonel, Medical Corps, US Army; Deputy Joint Program Executive Officer, Medical Systems, Joint Program Executive Office for Chemical/Biological Defense, US Department of Defense, Falls Church, Virginia; Chemical Casualty Care Consultant to the US Army Surgeon General; Adjunct Professor of Neurology, F. Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland [222]

Rick A. Nishimura, MD, FACC, FACP
Judd and Mary Morris Leighton Professor of Cardiovascular Diseases; Professor of Medicine; Consultant, Division of Cardiovascular Diseases and Internal Medicine, Mayo Clinic College of Medicine, Rochester, Minnesota [229, e29]

Robert L. Norris, MD
Professor, Department of Surgery, Division of Emergency Medicine, Stanford University School of Medicine, Palo Alto, California [396]

Thomas B. Nutman, MD
Head, Helminth Immunology Section; Head, Clinical Parasitology Unit, Laboratory of Parasitic Diseases, National Institutes of Health, Bethesda, Maryland [217, 218]

Katherine L. O'Brien, MDCM, MPH, FRCPC
Associate Professor, Center for American Indian Health; Departments of International Health and Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland [134]

Richard J. O'Brien, MD
Head, Product Evaluation and Demonstration, Foundation for Innovative and New Diagnostics (FIND), Geneva, Switzerland [165]

Max R. O'Donnell, MD
Assistant Professor of Medicine, Albert Einstein College of Medicine, Bronx, New York [168]

Nigel O'Farrell, MSc, MD, FRCP
Ealing Hospital, London, United Kingdom [161]

Jennifer Ogar, MS
Speech Pathologist, Memory and Aging Center, University of California, San Francisco, San Francisco, California; Acting Chief of Speech Pathology at the Department of Veterans Affairs, Martinez, California [e10]

Patrick T. O'Gara, MD
Professor of Medicine, Harvard Medical School; Director, Clinical Cardiology, Brigham and Women's Hospital, Boston, Massachusetts [227, 237, e13]

C. Warren Olanow MD, FRCPC
Department of Neurology and Neuroscience, Mount Sinai School of Medicine, New York, New York [372]

Andrew B. Onderdonk, PhD
Professor of Pathology, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [e22]

Chung Owyang, MD
H. Marvin Pollard Professor of Internal Medicine; Chief, Division of Gastroenterology, University of Michigan Health System, Ann Arbor, Michigan [290, 296]

William Pao, MD, PhD
Associate Professor of Medicine, Cancer Biology, and Pathology, Division of Hematology and Medical Oncology, Vanderbilt University School of Medicine, Nashville, Tennessee [89]

Umesh D. Parashar, MBBS, MPH
Lead, Viral Gastroenteritis Epidemiology Team, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [190]

Shreyaskumar R. Patel, MD
Center Medical Director, Sarcoma Center; Professor of Medicine; Deputy Chairman, Department of Sarcoma Medical Oncology, MD Anderson Cancer Center, Houston, Texas [98]

David L. Paterson, MD, PhD
Professor of Medicine, University of Queensland Centre for Clinical Research; Royal Brisbane and Women's Hospital, Brisbane, Australia [150]

Gustav Paumgartner, MD
Professor Emeritus of Medicine, University of Munich, Munich, Germany [311]

David A. Pegues, MD
Hospital Epidemiologist, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California [153]

Anton Y. Peleg, MBBS, PhD, MPH, FRACP
Infectious Diseases Physician, Senior Lecturer, and NHMRC Biomedical Fellow, Department of Infectious Diseases and Microbiology, The Alfred Hospital and Monash University, Melbourne, Victoria, Australia [150]

Florencia Pereyra, MD
Assistant Professor of Medicine, Harvard Medical School; Associate Physician, Infectious Disease Division, Brigham and Women's Hospital, Boston, Massachusetts [e23, e24]

Michael A. Pesce, PhD
Professor Emeritus of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons; Columbia University Medical Center, New York, New York [Appendix]

Clarence J. Peters, MD
John Sealy Distinguished University Chair in Tropical and Emerging Virology; Professor, Department of Mirobiology and Immunology; Department of Pathology; Director for Biodefense, Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, Texas [196, 197]

Gerald B. Pier, PhD
Professor of Medicine (Microbiology and Molecular Genetics), Harvard Medical School; Microbiologist, Brigham and Women's Hospital, Boston, Massachusetts [120]

Ronald E. Polk, PharmD
Professor of Pharmacy and Medicine; Chairman, Department of Pharmacy, School of Pharmacy, Virginia Commonwealth University/Medical College of Virginia Campus, Richmond, Virginia [133]

Richard J. Pollack, PhD
Research Associate Professor, Department of Biology, Boston University; Research Associate, Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts [397]

Andrew J. Pollard, PhD, FRCPCH
Professor of Pediatric Infection and Immunity; Director of the Oxford Vaccine Group, Department of Pediatrics, University of Oxford, Oxford, United Kingdom [143]

Reuven Porat, MD
Internal Medicine Department, Tel-Aviv Sourasky Medical Centre; Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel [16]

Daniel A. Portnoy, PhD
Professor of Biochemistry and Molecular Biology, Department of Molecular and Cell Biology, The School of Public Health, University of California, Berkeley, Berkeley, California [139]

John T. Potts, Jr., MD
Director of Research, Massachusetts General Hospital, Boston, Massachusetts [353]

Lawrie W. Powell, MD, PhD
Professor of Medicine; Director, Centre for the Advancement of Clinical Research, Royal Brisbane and Women's Hospital, Brisbane, Australia [357]

Alvin C. Powers, MD
Joe C. Davis Chair in Biomedical Science; Professor of Medicine, Molecular Physiology, and Biophysics; Director, Vanderbilt Diabetes Center; Chief, Division of Diabetes, Endocrinology, and Metabolism, Vanderbilt University School of Medicine, Nashville, Tennessee [344]

Daniel S. Pratt, MD
Assistant Professor of Medicine, Harvard Medical School; Massachusetts General Hospital, Boston, Massachusetts [42, 302]

Michael B. Prentice, MB ChB, PhD, MRCP(UK), FRCPath, FFPRCPI
Professor of Medical Microbiology, Department of Microbiology, University College Cork, Cork, Ireland [159]

Darwin J. Prockop, MD, PhD
Director and Professor, Institute for Regenerative Medicine, Texas A&M Health Science Center College of Medicine at Scott & White, Temple, Texas [363]

Stanley B. Prusiner, MD
Director, Institute for Neurodegenerative Diseases; Professor, Department of Neurology, University of California, San Francisco, San Francisco, California [383]

Howard I. Pryor, II, MD
Laboratory for Tissue Engineering and Organ Fabrication, Center for Regenerative Medicine, Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts [69]

Thomas C. Quinn, MD
Professor of Medicine, Johns Hopkins University, Baltimore, Maryland; Senior Investigator, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland [176]

Gil Rabinovici, MD
Attending Neurologist, Memory and Aging Center, University of California, San Francisco, San Francisco, California [e10]

Daniel J. Rader, MD
Cooper-McClure Professor of Medicine and Pharmacology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania [356]

Sanjay Ram, MD
Associate Professor of Medicine, Division of Infectious Diseases and Immunology, University of Massachusetts Medical School, Worcester, Massachusetts [144]

Reuben Ramphal, MD
Professor of Medicine, Molecular Genetics and Microbiology, University of Florida College of Medicine, Gainesville, Florida [152]

Kumanan Rasanathan, MBChB, MPH, FAFPHM
Technical Officer, Department of Ethics, Equity, Trade, and Human Rights, World Health Organization, Geneva, Switzerland [e1]

Neil H. Raskin, MD
Department of Neurology, University of California, San Francisco, San Francisco, San Francisco, California [14]

Anis Rassi, Jr., MD, PhD, FACC, FACP, FAHA
Scientific Director, Anis Rassi Hospital, Goiânia, Brazil [213]

James P. Rathmell, MD
Associate Professor of Anesthesia, Harvard Medical School; Chief, Division of Pain Medicine, Massachusetts General Hospital, Boston, Massachusetts [11]

Mario C. Raviglione, MD
Director, Stop TB Department, World Health Organization, Geneva, Switzerland [165]

Sharon L. Reed, MD
Professor of Pathology and Medicine; Director, Microbiology and Virology Laboratories, University of California, San Diego Medical Center, San Diego, California [e25]

Susan E. Reef, MD
Medical Epidemiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [193]

Richard C. Reichman, MD
Professor of Medicine and of Microbiology and Immunology, University of Rochester School of Medicine and Dentistry, Rochester, New York [185]

John J. Reilly, Jr., MD
Executive Vice Chairman; Department of Medicine; Professor of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania [260, e34]

John T. Repke, MD
University Professor and Chairman, Department of Obstetrics and Gynecology, Pennsylvania State University College of Medicine, Obstetrician-Gynecologist-in-Chief, The Milton S. Hershey Medical Center, Hershey, Pennsylvania [7]

Victor I. Reus, MD, DFAPA, FACP
Department of Psychiatry, University of California, San Francisco School of Medicine; Langley Porter Neuropsychiatric Institute, San Francisco, San Francisco, California [391]

Joseph Rhatigan, MD
Assistant Professor of Medicine, Harvard Medical School; Assistant Professor, Harvard School of Public Health; Brigham and Women's Hospital, Boston, Massachusetts [2]

Peter A. Rice, MD
Professor of Medicine, Division of Infectious Diseases and Immunology, University of Massachusetts Medical School, Worcester, Massachusetts [144]

Stuart Rich, MD
Professor of Medicine, Department of Medicine, Section of Cardiology, University of Chicago, Chicago, Illinois [250]

Gary S. Richardson, MD
Senior Research Scientist and Staff Physician, Henry Ford Hospital, Detroit, Michigan [27]

Elizabeth Robbins, MD
Clinical Professor of Pediatrics, University of California, San Francisco, San Francisco, California [e46]

Gary L. Robertson, MD
Emeritus Professor of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [340]

Russell G. Robertson, MD
Vice President for Medical Affairs, Rosalind Franklin University of Medicine and Science; Dean, Chicago Medical School, Chicago, Illinois [80]

Dan M. Roden, MD
William Stokes Professor of Experimental Therapeutics; Assistant Vice-Chancellor for Personalized Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee [5]

James A. Romano, Jr., PhD, DABT
Senior Principal Life Scientist and Technical Fellow, Science Applications International Corporation, Frederick, Maryland [222]

Karen L. Roos, MD
John and Nancy Nelson Professor of Neurology and Professor of Neurological Surgery, Indiana University School of Medicine, Indianapolis, Indiana [381]

Allan H. Ropper, MD
Professor of Neurology, Harvard Medical School; Executive Vice Chair of Neurology, Raymond D. Adams Distinguished Clinician, Brigham and Women's Hospital, Boston, Massachusetts [274, 377, 378]

Roger N. Rosenberg, MD
Zale Distinguished Chair and Professor of Neurology, Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas [373]

Myrna R. Rosenfeld, MD, PhD
Professor of Neurology and Chief, Division of Neuro-oncology, University of Pennsylvania, Philadelphia, Pennsylvania [101]

John H. Rubenstein, MD, PhD
Nina Ireland Distinguished Professor in Child Psychiatry, Center for Neurobiology and Psychiatry, Department of Psychiatry, University of California, San Francisco, San Francisco, California [390]

Michael A. Rubin, MD, PhD
Assistant Professor of Medicine, University of Utah School of Medicine, Salt Lake City, Utah [31]

Steven Rubin, MS
Acting Principal Investigator, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland [194]

Robert M. Russell, MD
Professor Emeritus of Medicine and Nutrition, Tufts University, Boston, Massachusetts; Office of Dietary Supplements, National Institutes of Health, Bethesda, Maryland [74]

Thomas A. Russo, MD, CM, FIDSA
Professor of Medicine and Microbiology and Immunology; Chief, Division of Infectious Diseases, University at Buffalo, State University of New York, Buffalo, New York [149, 163]

Anna Rutherford, MD, MPH
Instructor in Medicine, Harvard Medical School; Associate Physician, Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital, Boston, Massachusetts [e56]

Edward T. Ryan, MD, DTM&H
Associate Professor of Medicine, Harvard Medical School; Associate Professor of Immunology and Infectious Diseases, Harvard School of Public Health; Director, Tropical and Geographic Medicine, Massachusetts General Hospital, Boston, Massachusetts [128, 156]

Miguel Sabria, MD
Professor of Medicine, Autonomous University of Barcelona; Chief, Infectious Diseases Section, Germans Trias I Pujl Hospital, Barcelona, Spain [147]

David J. Salant, MD
Professor of Medicine, Boston University School of Medicine; Chief, Section of Nephrology, Boston Medical Center, Boston, Massachusetts [284, 285]

Martin A. Samuels, MD, DSc(hon), FAAN, MACP, FRCP
Professor of Neurology, Harvard Medical School; Chairman, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts [e43, e47]

Philippe Sansonetti, MD, MS
Professor, Collège de France; Institut Pasteur, Paris, France [154]

Jussi J. Saukkonen, MD
Associate Professor of Medicine, Section of Pulmonary, Allergy, and Critical Care Medicine, Boston University School of Medicine, Boston, Massachusetts [168]

Edward A. Sausville, MD, PhD
Professor, Department of Medicine, University of Maryland School of Medicine; Deputy Director and Associate Director for Clinical Research, University of Maryland Marlene and Stewart Greenebaum Cancer Center, Baltimore, Maryland [85]

Mohamed H. Sayegh, MD
Raja N. Khuri Dean, Faculty of Medicine; Professor of Medicine and Immunology; Vice President of Medical Affairs, American University of Beirut, Beirut, Lebanon; Visiting Professor of Medicine and Pediatrics, Harvard Medical School; Director, Schuster Family Transplantation Research Center, Brigham and Women's Hospital; Children's Hospital, Boston, Massachusetts [282]

David T. Scadden, MD
Gerald and Darlene Jordan Professor of Medicine, Harvard Stem Cell Institute, Harvard Medical School; Department of Stem Cell and Regenerative Biology, Massachusetts General Hospital, Boston, Massachusetts [66]

Anthony H. V. Schapira, DSc, MD, FRCP, FMedSci
University Department of Clinical Neurosciences, University College London; National Hospital for Neurology and Neurosurgery, Queen's Square, London, United Kingdom [372]

Howard I. Scher, MD
Professor of Medicine, Weill Cornell Medical College; D. Wayne Calloway Chair in Urologic Oncology; Chief, Genitourinary Oncology Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York [94, 95]

Anne Schuchat, MD
Director, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia [122]

Marc A. Schuckit, MD
Distinguished Professor of Psychiatry, University of California, San Diego School of Medicine, La Jolla, California [392]

H. Ralph Schumacher, MD
Professor of Medicine, Division of Rheumatology, University of Pennsylvania, School of Medicine, Philadelphia, Pennsylvania [333]

Gordon E. Schutze, MD
Professor of Pediatrics, Section of Retrovirology; Vice President, Baylor International Pediatric AIDS Initiative at Texas Children's Hospital, Baylor College of Medicine, Houston, Texas [158]

Stuart Schwartz, PhD
Professor of Human Genetics, Medicine and Pathology, University of Chicago, Chicago, Illinois [62]

Richard M. Schwartzstein, MD
Ellen and Melvin Gordon Professor of Medicine and Medical Education; Associate Chief, Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts [33]

William W. Seeley, MD
Associate Professor of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [371]

Michael V. Seiden, MD, PhD
Professor of Medicine; President and CEO, Fox Chase Cancer Center, Philadelphia, Pennsylvania [97]

Julian L. Seifter, MD
Associate Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [289]

David C. Seldin, MD, PhD
Chief, Section of Hematology-Oncology, Department of Medicine; Director, Amyloid Treatment and Research Program, Boston University School of Medicine; Boston Medical Center, Boston, Massachusetts [112]

Andrew P. Selwyn, MD, MBCHB
Professor of Medicine Brigham and Women's Hospital, Boston, Massachusetts [243]

Ankoor Shah, MD
Department of Medicine, Division of Rheumatology and Immunology, Duke University Medical Center, Durham, North Carolina [321]

Steven D. Shapiro, MD
Jack D. Myers Professor and Chair, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania [260]

Kanade Shinkai, MD, PhD
Assistant Professor, Department of Dermatology, University of California, San Francisco, San Francisco, California [55]

William Silen, MD
Johnson and Johnson Professor Emeritus of Surgery, Harvard Medical School, Auburndale, Massachusetts [13, 299, 300]

Edwin K. Silverman, MD, PhD
Associate Professor of Medicine, Harvard Medical School; Channing Laboratory, Pulmonary and Critical Care Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts [260]

Martha Skinner, MD
Professor, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts [112]

Karl Skorecki, MD, FRCP(C), FASN
Annie Chutick Professor in Medicine (Nephrology); Director, Rappaport Research Institute, Technion – Israel Institute of Technology; Director, Medical and Research Development, Rambam Health Care Campus, Haifa, Israel [280, e18]

Wade S. Smith, MD, PhD
Professor of Neurology, Daryl R. Gress Endowed Chair of Neurocritical Care and Stroke; Director, University of California, San Francisco Neurovascular Service, San Francisco, San Francisco, California [275, 370]

A. George Smulian, MBBCh
Associate Professor of Medicine, University of Cincinnati College of Medicine; Chief, Infectious Disease Section, Cincinnati VA Medical Center, Cincinnati, Ohio [207]

Jeremy Sobel, MD, MPH
Medical Officer, Office of Global Health, Centers for Disease Control and Prevention, Atlanta, Georgia [141]

Kelly A. Soderberg, PhD, MPH
Director, Program Management, Duke Human Vaccine Institute, Duke University School of Medicine, Durham, North Carolina [314]

Julian Solway, MD
Walter L. Palmer Distinguished Service Professor of Medicine and Pediatrics; Associate Dean for Translational Medicine, Biological Sciences Division; Vice Chair for Research, Department of Medicine; Chair, Committee on Molecular Medicine, University of Chicago, Chicago, Illinois [252, 264]

Michael F. Sorrell, MD
Robert L. Grissom Professor of Medicine, University of Nebraska Medical Center, Omaha, Nebraska [307]

Frank E. Speizer, MD
E. H. Kass Distinguished Professor of Medicine, Channing Laboratory, Harvard Medical School; Professor of Environmental Science, Harvard School of Public Health, Boston, Massachusetts [256]

Brad Spellberg, MD
Associate Professor of Medicine, Geffen School of Medicine, University of California, Los Angeles (UCLA); Divisions of General Internal Medicine and Infectious Diseases, Los Angeles Biomedical Research Institute at Harbor–UCLA Medical Center, Torrance, California [205]

Jerry L. Spivak, MD
Professor of Medicine and Oncology, Hematology Division, Johns Hopkins University School of Medicine, Baltimore, Maryland [108]

David D. Spragg, MD
Assistant Professor of Medicine, Johns Hopkins University, Baltimore, Maryland [231, 232]

Samuel L. Stanley, Jr., MD
President, Stony Brook University, Stony Brook, New York [209]

E. William St. Clair, MD
Department of Medicine, Division of Rheumatology and Immunology, Duke University Medical Center, Durham, North Carolina [321]

Allen C. Steere, MD
Professor of Medicine, Harvard Medical School; Massachusetts General Hospital, Boston, Massachusetts [173]

Robert S. Stern, MD
Carl J. Herzog Professor of Dermatology, Harvard Medical School; Chair, Department of Dermatology, Beth Israel Deaconess Medical Center, Boston, Massachusetts [55]

Dennis L. Stevens, MD, PhD
Professor of Medicine, University of Washington School of Medicine, Seattle, Washington; Chief, Infectious Disease Section, Veterans Affairs Medical Center, Boise, Idaho [125, 142]

Lynne Warner Stevenson, MD

Eugene Harvey Strickland, MD
Professor of Medicine, Harvard Medical School; Director, Heart Failure Program, Brigham and Women's Hospital, Boston, Massachusetts [238]

Stephen E. Straus,† MD
National Institute of Allergy and Infectious Diseases, Bethesda, Maryland [e2]

Stephanie Studenski, MD, MPH
Professor of Geriatric Medicine, Department of Medicine, University of Pittsburgh School of Medicine; Staff Physician, VA Pittsburgh Geriatric Research Education and Clinical Center, Pittsburgh, Pennsylvania [72]

Lewis Sudarsky, MD
Associate Professor of Neurology, Harvard Medical School; Director of Movement Disorders, Brigham and Women's Hospital, Boston, Massachusetts [24]

Donna C. Sullivan, PhD
Professor, Department of Medicine, Division of Infectious Diseases, University of Mississippi Medical School, Jackson, Mississippi [201]

Shyam Sundar, MD
Professor of Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India [212]

Paolo M. Suter, MD, MS
Professor, Clinic and Policlinic of Internal Medicine, University Hospital, Zurich, Switzerland [74]

Richard Suzman, PhD
Director, Behavioral and Social Research Program, National Institute on Aging, National Institutes of Health, Chevy Chase, Maryland [70]

Morton N. Swartz, MD
Professor of Medicine, Harvard Medical School; Chief, Jackson Firm Medical Service and Infectious Disease Unit, Massachusetts General Hospital, Boston, Massachusetts [382]

Robert A. Swerlick, MD
Alicia Leizman Stonecipher Professor and Chair of Dermatology, Emory University School of Medicine, Atlanta, Georgia [e16]

Geoffrey Tabin, MD
Professor of Ophthalmology and Visual Sciences, University of Utah School of Medicine; Director, International Ophthalmology Division, John A. Moran Eye Center; Director, Himalayan Cataract Project, Salt Lake City, Utah [e51]

Maria Carmela Tartaglia, MD, FRCPC
Clinical Instructor of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, California [e10]

Joel D. Taurog, MD
Professor of Internal Medicine, Rheumatic Diseases Division, University of Texas Southwestern Medical Center, Dallas, Texas [325]

Stephen C. Textor, MD
Professor of Medicine, Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota [286]

C. Louise Thwaites, MD, MBBS
Musculoskeletal Physician, Horsham, West Sussex; Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam [140]

Alan D. Tice, MD, FACP
Infections Limited Hawaii; John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii [126]

Zelig A. Tochner, MD
Professor of Radiation Oncology, University of Pennsylvania School of Medicine; Medical Director, Proton Therapy Center, Philadelphia, Pennsylvania [223]

Gordon F. Tomaselli, MD
Michel Mirowski, MD Professor of Cardiology; Professor of Medicine and Cellular and Molecular Medicine; Chief, Division of Cardiology, Johns Hopkins University, Baltimore, Maryland [231, 232]

Mark Topazian, MD
Professor of Medicine, Mayo Clinic, Rochester, Minnesota [291, e36]

Barbara W. Trautner, MD, PhD
Assistant Professor, Section of Infectious Diseases, Baylor College of Medicine; The Michael E. DeBakey Veterans Affairs Medical Center, Houston VA Health Services Research and Development Center of Excellence, Houston, Texas [288]

Jeffrey M. Trent, PhD, FACMG

President and Research Director, Translational Genomics Research Institute, Phoenix, Arizona; Van Andel Research Institute, Grand Rapids, Michigan [83]

Elbert P. Trulock, MD
Rosemary and I. Jerome Flance Professor in Pulmonary Medicine, Washington University School of Medicine, St. Louis, Missouri [266]

Kenneth L. Tyler, MD
Reuler-Lewin Family Professor and Chair, Department of Neurology; Professor of Medicine and Microbiology, University of Colorado School of Medicine, Denver, Colorado; Chief of Neurology, University of Colorado Hospital, Aurora, Colorado [381]

Athanasios G. Tzioufas, MD
Professor, Department of Pathophysiology, National University of Athens School of Medicine, Athens, Greece [324]

Walter J. Urba, MD, PhD
Director of Cancer Research, Robert W. Franz Cancer Research Center, Providence Portland Medical Center, Portland, Oregon [87]

Joseph P. Vacanti, MD
John Homans Professor of Surgery, Harvard Medical School; Surgeon-in-Chief, Massachusetts General Hospital for Children; Deputy Director, Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts [69]

Jos W. M. van der Meer, MD, PhD
Professor of Medicine; Head, Department of General Internal Medicine, Radboud University, Nijmegen Medical Centre, Nijmegen, Netherlands [389]

Edouard Vannier, PhD, PharmD
Assistant Professor, Division of Geographic Medicine and Infectious Diseases, Tufts University School of Medicine; Tufts Medical Center, Boston, Massachusetts [211]

Gauri R. Varadhachary, MD
Associate Professor, Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas [99]

John Varga, MD
John Hughes Professor of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois [323]

Camilo Jimenez Vasquez, MD
Assistant Professor, Department of Endocrine Neoplasia and Hormonal Disorders, Division of Internal Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas [351]

Joseph M. Vinetz, MD
Professor of Medicine, Division of Infectious Diseases, Department of Medicine, University of California, San Diego, San Diego, California [171]

Indre V. Viskontas, PhD
Visiting Scholar, Memory and Aging Center, University of California, San Francisco, San Francisco, California [e9]

Panayiotis G. Vlachoyiannopoulos, MD
Associate Professor of Medicine-Immunology, Department of Pathophysiology, Medical School, National University of Athens, Athens, Greece [320]

Bert Vogelstein, MD
Professor of Oncology and Pathology; Investigator, Howard Hughes Medical Institute; Sidney Kimmel Comprehensive Cancer Center; Johns Hopkins University School of Medicine, Baltimore, Maryland [83]

Everett E. Vokes, MD
John E. Ultmann Professor and Chairman, Department of Medicine; Physician-in-Chief, University of Chicago Medical Center, Chicago, Illinois [88]

Tamara J. Vokes, MD, FACP
Professor, Department of Medicine, Section of Endocrinology, University of Chicago, Chicago, Illinois [355]

Sushrut S. Waikar, MD, MPH
Assistant Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [279]

Matthew K. Waldor, MD, PhD
Edward H. Kass Professor of Medicine, Channing Laboratory, Brigham and Women's Hospital; Harvard Medical School and Howard Hughes Medical Institute, Boston, Massachusetts [156]

David H. Walker, MD
The Carmage and Martha Walls Distinguished University Chair in Tropical Diseases; Professor and

The Carriage and Martha Wallis Distinguished University Chair in Tropical Diseases; Professor and Chairman, Department of Pathology; Executive Director, Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, Texas [174]

Mark F. Walker, MD
Associate Professor, Department of Neurology, Case Western Reserve University School of Medicine; Daroff-Dell' Osso Ocular Motility Laboratory, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, Ohio [21]

B. Timothy Walsh, MD
Professor, Department of Psychiatry, College of Physicians and Surgeons, Columbia University; New York State Psychiatric Institute, New York, New York [79]

Peter D. Walzer, MD, MSc
Professor of Medicine, University of Cincinnati College of Medicine; Associate Chief of Staff for Research, Cincinnati VA Medical Center, Cincinnati, Ohio [207]

Fred Wang, MD
Professor of Medicine, Harvard Medical School; Brigham and Women's Hospital, Boston, Massachusetts [177, 183]

John W. Warren, MD
Professor of Medicine, University of Maryland School of Medicine, Baltimore, Maryland [e35]

Carl V. Washington, MD
Associate Professor of Dermatology, Winship Cancer Center, Emory University School of Medicine, Atlanta, Georgia [87]

Anthony P. Weetman, MD
University of Sheffield School of Medicine, Sheffield, United Kingdom [341]

Robert A. Weinstein, MD
The C Anderson Hedberg MD Professor of Internal Medicine, Rush Medical College; Interim Chairman, Department of Medicine, John Stroger Hospital, Chicago, Illinois [131]

Jeffrey I. Weitz, MD, FRCP(C), FACP
Professor of Medicine and Biochemistry; Executive Director, Thrombosis and Atherosclerosis Research Institute; HSFO/J. F. Mustard Chair in Cardiovascular Research, Canada Research Chair (Tier 1) in Thrombosis, McMaster University, Hamilton, Ontario, Canada [118]

Peter F. Weller, MD
Chief, Infectious Disease Division; Chief, Allergy and Inflammation Division, Beth Israel Deaconess Medical Center, Boston, Massachusetts [215-218, 220]

Patrick Y. Wen, MD
Professor of Neurology, Harvard Medical School; Dana-Farber Cancer Institute, Boston, Massachusetts [379]

Michael R. Wessels, MD
John F. Enders Professor of Pediatrics; Professor of Medicine, Harvard Medical School; Chief, Division of Infectious Diseases, Children's Hospital, Boston, Massachusetts [136]

Meir Wetzler, MD, FACP
Professor of Medicine, Roswell Park Cancer Institute, Buffalo, New York [109]

L. Joseph Wheat, MD
MiraVista Diagnostics and MiraBella Technologies, Indianapolis, Indiana [199]

A. Clinton White, Jr., MD
Director, Infectious Disease Division, Department of Internal Medicine, University of Texas Medical Branch, Galveston, Texas [220]

Nicholas J. White, MD, DSc, FRCP, F Med Sci, FRS
Professor of Tropical Medicine, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand [210, e27]

Richard J. Whitley, MD
Distinguished Professor of Pediatrics, Loeb Eminent Scholar Chair in Pediatrics; Professor of Pediatrics, Microbiology, Medicine, and Neurosurgery, University of Alabama at Birmingham, Birmingham, Alabama [180]

John W. Winkelman, MD, PhD
Associate Professor of Psychiatry, Harvard Medical School; Medical Director, Sleep Health Centers, Brigham and Women's Hospital, Boston, Massachusetts [27]

Bruce U. Wintroub, MD
Professor and Chair, Department of Dermatology, University of California, San Francisco, San Francisco, California [55]

Andrea Wolf, MD, MPH
Instructor in Surgery, Harvard Medical School; Chief Resident in Cardiothoracic Surgery, Division of

Instructor in Surgery, Harvard Medical School; Chief Resident in Cardiothoracic Surgery, Division of Thoracic Surgery, Brigham and Women's Hospital, Boston, Massachusetts [e55]

Allan W. Wolkoff, MD
Professor of Medicine and Anatomy and Structural Biology; Associate Chair of Medicine for Research; Chief, Division of Gastroenterology and Liver Diseases, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York [303]

John B. Wong, MD
Professor of Medicine, Tufts University School of Medicine; Chief, Division of Clinical Decision Making, Department of Medicine, Tufts Medical Center, Boston, Massachusetts [3]

Louis Michel Wong Kee Song, MD
Associate Professor, Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota [291, e36]

Robert L. Wortman, MD, FACP, MACR
Professor, Department of Medicine, Dartmouth Medical School and Dartmouth Hitchcock Medical Center, Lebanon, New Hampshire [359]

Shirley H. Wray, MB, ChB, PhD, FRCP
Professor of Neurology, Harvard Medical School; Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts [e11]

Bechien U. Wu, MD
Instructor of Medicine, Harvard Medical School; Associate Physician, Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts [313]

Richard Wunderink, MD
Professor of Medicine, Division of Pulmonary and Critical Care, Northwestern University Feinberg School of Medicine, Chicago, Illinois [257]

Kim B. Yancey, MD
Professor and Chair, Department of Dermatology, University of Texas Southwestern Medical Center, Dallas, Texas [51, 54]

Janet A. Yellowitz, DMD, MPH
Associate Professor; Director, Geriatric Dentistry, University of Maryland Dental School, Baltimore, Maryland [e12]

Lam Minh Yen, MD
Director, Tetanus Intensive Care Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam [140]

Maria A. Yialamas, MD
Instructor, Harvard Medical School; Associate Program Director, Internal Medicine Residency, Brigham and Women's Hospital, Boston, Massachusetts [e54, e56]

Neal S. Young, MD
Chief, Hematology Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland [107]

Victor L. Yu, MD
Professor of Medicine, Department of Medicine, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania [147]

Laura A. Zimmerman, MPH
Epidemiologist, Centers for Disease Control and Prevention, Atlanta, Georgia [193]

†Deceased

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE MODERN-DAY PHYSICIAN

> *No greater opportunity, responsibility, or obligation can fall to the lot of a human being than to become a physician. In the care of the suffering, [the physician] needs technical skill, scientific knowledge, and human understanding. . .. Tact, sympathy, and understanding are expected of the physician, for the patient is no mere collection of symptoms, signs, disordered functions, damaged organs, and disturbed emotions. [The patient] is human, fearful, and hopeful, seeking relief, help, and reassurance.*
>
> —*Harrison's Principles of Internal Medicine* , 1950

The practice of medicine has changed in significant ways since the first edition of this book appeared more than 60 years ago. The advent of molecular genetics, molecular biology, and molecular pathophysiology, sophisticated new imaging techniques, and advances in bioinformatics and information technology have contributed to an explosion of scientific information that has fundamentally changed the way physicians define, diagnose, treat, and prevent disease. This growth of scientific knowledge is ongoing and accelerating.

The widespread use of electronic medical records and the Internet have altered the way doctors practice medicine and exchange information. As today's physician struggles to integrate copious amounts of scientific knowledge into everyday practice, it is important to remember that the ultimate goal of medicine is to prevent disease and treat sick patients. Despite more than 60 years of scientific advances since the first edition of this text, it is critical to underscore that cultivating the intimate relationship between physician and patient still lies at the heart of successful patient care.

The Science and Art of Medicine

Deductive reasoning and applied technology form the foundation for the solution to many clinical problems. Spectacular advances in biochemistry, cell biology, and genomics, coupled with newly developed imaging techniques, allow access to the innermost parts of the cell and provide a window to the most remote recesses of the body. Revelations about the nature of genes and single cells have opened the portal for formulating a new molecular basis for the physiology of systems. Increasingly, physicians are learning how subtle changes in many different genes can affect the function of cells and organisms. Researchers are beginning to decipher the complex mechanisms by which genes are regulated. Doctors have developed a new appreciation of the role of stem cells in normal tissue function and in the development of cancer, degenerative disease, and other disorders, as well as their emerging role in the treatment of certain diseases. The knowledge gleaned from the science of medicine has already improved and undoubtedly will further improve physicians' understanding of complex disease processes and provide new approaches to disease treatment and prevention. Yet, skill in the most sophisticated application of laboratory technology and in the use of the latest therapeutic modality alone does not make a good physician.

When a patient poses challenging clinical problems, an effective physician must be able to identify the crucial elements in a complex history and physical examination; order the appropriate laboratory, imaging, and diagnostic tests; and extract the key results from the crowded computer printouts of data to determine whether to "treat" or to "watch." Deciding whether a clinical clue is worth pursuing or should be dismissed as a "red herring" and weighing whether a proposed test, preventive measure, or treatment entails a greater risk than the disease itself are essential judgments that a skilled clinician must make many times each day. This combination of medical knowledge, intuition, experience, and judgment defines the ***art of medicine*** , which is as necessary to the practice of medicine as is a sound scientific base.

CLINICAL SKILLS

History-Taking

The written history of an illness should include all the facts of medical significance in the life of the patient. Recent events should be given the most attention. The patient should, at some early point, have the opportunity to tell his or her own story of the illness without frequent interruption and, when appropriate, receive expressions of interest, encouragement, and empathy from the physician. Any event related by the patient, however trivial or seemingly irrelevant, may provide the key to solving the medical problem. In general, only patients who feel comfortable with the physician will offer complete information, and thus putting the patient at ease to the greatest extent possible contributes

substantially to obtaining an adequate history.

An informative history is more than an orderly listing of symptoms; by listening to patients and noting the way in which they describe their symptoms, physicians can gain valuable insight into the problem. Inflections of voice, facial expression, gestures, and attitude, i.e., "body language," may reveal important clues to the meaning of the symptoms to the patient. Because patients vary in their medical sophistication and ability to recall facts, the reported medical history should be corroborated whenever possible. The social history also can provide important insights into the types of diseases that should be considered. The family history not only identifies rare Mendelian disorders within a family but often reveals risk factors for common disorders, such as coronary heart disease, hypertension, and asthma. A thorough family history may require input from multiple relatives to ensure completeness and accuracy, and once recorded, it can be updated readily. The process of history-taking provides an opportunity to observe the patient's behavior and watch for features to be pursued more thoroughly during the physical examination.

The very act of eliciting the history provides the physician with an opportunity to establish or enhance the unique bond that forms the basis for the ideal patient-physician relationship. This process helps the physician develop an appreciation of the patient's perception of the illness, the patient's expectations of the physician and the health care system, and the financial and social implications of the illness to the patient. Although current health care settings may impose time constraints on patient visits, it is important not to rush the history-taking since this may lead the patient to believe that what he or she is relating is not of importance to the physician and, therefore, may withhold relevant information. The confidentiality of the patient-physician relationship cannot be overemphasized.

Physical Examination

The purpose of the physical examination is to identify the physical signs of disease. The significance of these objective indications of disease is enhanced when they confirm a functional or structural change already suggested by the patient's history. At times, however, the physical signs may be the only evidence of disease.

The physical examination should be performed methodically and thoroughly, with consideration for the patient's comfort and modesty. Although attention is often directed by the history to the diseased organ or part of the body, the examination of a new patient must extend from head to toe in an objective search for abnormalities. Unless the physical examination is systematic and is performed in a consistent manner from patient to patient, important segments may be omitted inadvertently. The results of the examination, like the details of the history, should be recorded at the time they are elicited, not hours later, when they are subject to the distortions of memory. Skill in physical diagnosis is acquired with experience, but it is not merely technique that determines success in eliciting signs of disease. The detection of a few scattered petechiae, a faint diastolic murmur, or a small mass in the abdomen is not a question of keener eyes and ears or more sensitive fingers but of a mind alert to those findings. Because physical findings can change with time, the physical examination should be repeated as frequently as the clinical situation warrants. Because a large number of highly sensitive diagnostic tests are available, particularly imaging techniques, it may be tempting to put less emphasis on the physical examination. Indeed, many patients are seen for the first time after a series of diagnostic tests have been performed and the results are known. This fact should not deter the physician from performing a thorough physical examination since clinical findings are often present that have "escaped" the barrage of preexamination diagnostic tests. The act of examining (touching) the patient also offers an opportunity for communication and may have reassuring effects that foster the patient-physician relationship.

Diagnostic Studies

Physicians have become increasingly reliant on a wide array of laboratory tests to solve clinical problems. However, accumulated laboratory data do not relieve the physician from the responsibility of carefully observing, examining, and studying the patient. It is also essential to appreciate the limitations of diagnostic tests. By virtue of their impersonal quality, complexity, and apparent precision, they often gain an aura of authority regardless of the fallibility of the tests, the instruments used in the tests, and the individuals performing or interpreting them. Physicians must weigh the expense involved in the laboratory procedures against the value of the information they are likely to provide.

Single laboratory tests are rarely ordered. Instead, physicians generally request "batteries" of multiple tests, which often prove useful. For example, abnormalities of hepatic function may provide the clue to nonspecific symptoms such as generalized weakness and increased fatigability, suggesting the diagnosis of chronic liver disease. Sometimes a single abnormality, such as an elevated serum calcium level, points to a particular disease, such as hyperparathyroidism or an underlying malignancy.

The thoughtful use of screening tests such as low-density lipoprotein cholesterol may be quite useful. A group of laboratory determinations can be carried out conveniently on a single specimen at relatively low cost. Screening tests are most informative when directed toward common diseases or disorders and when their results indicate the need for other useful tests or interventions that may be costly to perform. On the one hand, biochemical measurements, together with simple laboratory examinations

such as blood count, urinalysis, and sedimentation rate, often provide a major clue to the presence of a pathologic process. On the other hand, the physician must learn to evaluate occasional abnormalities among the screening tests that may not necessarily connote significant disease. An in-depth workup after a report of an isolated laboratory abnormality in a person who is otherwise well is almost invariably wasteful and unproductive. Because so many tests are performed routinely as screening, it would not be unusual for one or two of them to be slightly abnormal. If there is no suspicion of an underlying illness, these tests ordinarily are repeated to ensure that the abnormality does not represent a laboratory error. If an abnormality is confirmed, it is important to consider its potential significance in the context of the patient's condition and other test results.

The development of technically improved imaging studies with greater sensitivity and specificity is one of the most rapidly advancing areas of medicine. These tests provide remarkably detailed anatomic information that can be a pivotal factor in medical decision-making. Ultrasonography, a variety of isotopic scans, CT, MRI, and positron emission tomography have benefited patients by supplanting older, more invasive approaches and opening new diagnostic vistas. In light of their capabilities and the rapidity with which they can lead to a diagnosis, it is tempting to order a battery of imaging studies. All physicians have had experiences in which imaging studies turned up findings that led to an unexpected diagnosis. Nonetheless, patients must endure each of these tests, and the added cost of unnecessary testing is substantial. Furthermore, investigation of an unexpected abnormal finding may be associated with risk and/or expense and may lead to the diagnosis of an irrelevant or incidental problem. A skilled physician must learn to use these powerful diagnostic tools judiciously, always considering whether the results will alter management and benefit the patient.

PRINCIPLES OF PATIENT CARE

Evidence-Based Medicine

Evidence-based medicine refers to the concept that clinical decisions are formally supported by data, preferably data that are derived from prospectively designed, randomized, controlled clinical trials. This approach is in sharp contrast to anecdotal experience, which often may be biased. Unless they are attuned to the importance of using larger, more objective studies for making decisions, even the most experienced physicians can be influenced by recent encounters with selected patients. Evidence-based medicine has become an increasingly important part of the routine practice of medicine and has led to the publication of a number of practice guidelines.

Practice Guidelines

Professional organizations and government agencies are developing formal clinical-practice guidelines to aid physicians and other caregivers in making diagnostic and therapeutic decisions that are evidence-based, cost-effective, and most appropriate to a particular patient and clinical situation. As the evidence base of medicine increases, guidelines can provide a useful framework for managing patients with particular diagnoses or symptoms. They can protect patients—particularly those with inadequate health care benefits—from receiving substandard care. Guidelines also can protect conscientious caregivers from inappropriate charges of malpractice and society from the excessive costs associated with the overuse of medical resources. There are, however, caveats associated with clinical-practice guidelines since they tend to oversimplify the complexities of medicine. Furthermore, groups with differing perspectives may develop divergent recommendations regarding issues as basic as the need for mammographic screening of women in their forties or a prostate-specific antigen (PSA) assay in the serum of men over age 50. Finally, guidelines do not—and cannot be expected to—account for the uniqueness of each individual and his or her illness. The physician's challenge is to integrate into clinical practice the useful recommendations offered by experts without accepting them blindly or being inappropriately constrained by them.

Medical Decision-Making

Medical decision-making is an important responsibility of the physician and occurs at each stage of the diagnostic and treatment process. It involves the ordering of additional tests, requests for consults, and decisions regarding treatment and prognosis. This process requires an in-depth understanding of the pathophysiology and natural history of disease. As described above, medical decision-making should be evidence-based so that patients derive the full benefit of the scientific knowledge available to physicians. Formulating a differential diagnosis requires not only a broad knowledge base but also the ability to assess the relative probabilities of various diseases. Application of the scientific method, including hypothesis formation and data collection, is essential to the process of accepting or rejecting a particular diagnosis. Analysis of the differential diagnosis is an iterative process. As new information or test results are acquired, the group of disease processes being considered can be contracted or expanded appropriately.

Despite the importance of evidence-based medicine, much of medical decision-making relies on good clinical judgment, a process that is difficult to quantify or even to assess qualitatively. Physicians must use their knowledge and experience as a basis for weighing known factors along with the inevitable uncertainties and the need to use sound judgment; this synthesis of information is particularly important when a relevant evidence base is not available. Several quantitative tools may be invaluable

in synthesizing the available information, including diagnostic tests, *Bayes' theorem*, and multivariate statistical models. *Diagnostic tests* serve to reduce uncertainty about a diagnosis or prognosis in a particular individual and help the physician decide how best to manage that individual's condition. The battery of diagnostic tests complements the history and the physical examination. The accuracy of a particular test is ascertained by determining its sensitivity (true-positive rate) and specificity (true-negative rate) as well as the predictive value of a positive and a negative result. Bayes' theorem uses information on a test's sensitivity and specificity, in conjunction with the pretest probability of a diagnosis, to determine mathematically the posttest probability of the diagnosis. More complex clinical problems can be approached with *multivariable statistical models*, which generate highly accurate information even when multiple factors are acting individually or together to affect disease risk, progression, or response to treatment. Studies comparing the performance of statistical models with that of expert clinicians have documented equivalent accuracy, although the models tend to be more consistent. Thus, multivariate statistical models may be particularly helpful to less experienced clinicians. **See Chap. 3 for a more thorough discussion of decision-making in clinical medicine.**

Electronic Medical Records

Growing reliance on computers and the strength of information technology are playing an increasingly important role in medicine. Laboratory data are accessed almost universally through computers. Many medical centers now have electronic medical records, computerized order entry, and bar-coded tracking of medications. Some of these systems are interactive and provide reminders or warn of potential medical errors. In many ways, the health care system has lagged behind other industries in the adoption of information technology. Electronic medical records have extraordinary potential for providing rapid access to clinical information, imaging studies, laboratory results, and medications. This type of information is invaluable for ongoing efforts to enhance quality and improve patient safety. Ideally, patient records should be easily transferred across the health care system, providing reliable access to relevant data and historic information. However, technology limitations and concerns about privacy and cost continue to limit a broad-based utilization of electronic health records in most clinical settings. It also should be emphasized that information technology is merely a tool and can never replace the clinical decisions that are best made by the physician. In this regard, clinical knowledge and an understanding of the patient's needs, supplemented by quantitative tools, still seem to represent the best approach to decision-making in the practice of medicine.

Evaluation of Outcomes

Clinicians generally use *objective* and readily measurable parameters to judge the outcome of a therapeutic intervention. For example, findings on physical or laboratory examination—such as the blood pressure level, the patency of a coronary artery on an angiogram, or the size of a mass on a radiologic examination—can provide critically important information. However, patients usually seek medical attention for *subjective* reasons; they wish to obtain relief from pain, preserve or regain function, and enjoy life. The components of a patient's health status or quality of life can include bodily comfort, capacity for physical activity, personal and professional function, sexual function, cognitive function, and overall perception of health. Each of these important areas can be assessed by means of structured interviews or specially designed questionnaires. Such assessments also provide useful parameters by which the physician can judge the patient's subjective view of his or her disability and the response to treatment, particularly in chronic illness. The practice of medicine requires consideration and integration of both objective and subjective outcomes.

Women's Health and Disease

Although past epidemiologic studies and clinical trials often focused predominantly on men, more recent studies have included more women, and some, like the Women's Health Initiative, have exclusively addressed women's health issues. Significant gender differences exist in diseases that afflict both men and women. Much is still to be learned in this arena, and ongoing studies should enhance physicians' understanding of the mechanisms of gender differences in the course and outcome of certain diseases. **For a more complete discussion of women's health, see Chap. 6.**

Care of the Elderly

The relative proportion of elderly individuals in the populations of developed nations has been growing considerably over the last few decades and will continue to grow. In this regard, the practice of medicine will continue to be greatly influenced by the health care needs of this growing elderly population. The physician must understand and appreciate the decline in physiologic reserve associated with aging; the diminished responses of the elderly to vaccinations such as those against influenza; the different responses of the elderly to common diseases; and disorders that occur commonly with aging, such as depression, dementia, frailty, urinary incontinence, and fractures. **For a more complete discussion of medical care for the elderly, see Part 5, Chaps. 70, 71, and 72.**

Errors in the Delivery of Health Care

A report from the Institute of Medicine called for an ambitious agenda to reduce medical error rates and improve patient safety by designing and implementing fundamental changes in health care systems.

Adverse drug reactions occur in at least 5% of hospitalized patients, and the incidence increases with the use of a large number of drugs. No matter what the clinical situation is, it is the responsibility of the physician to use powerful therapeutic measures wisely, with due regard for their beneficial action, potential dangers, and cost. It is also the responsibility of hospitals and health care organizations to develop systems to reduce risk and ensure patient safety. Medication errors can be reduced through the use of ordering systems that eliminate misreading of handwriting. Implementation of infection control systems, enforcement of hand washing protocols, and careful oversight of antibiotic use can minimize the complications of nosocomial infections.

The Role of the Physician in the Informed Consent of the Patient

The fundamental principles of medical ethics require physicians to act in the patient's best interest and respect the patient's autonomy. This is particularly relevant to the issue of informed consent. Most patients possess only limited medical knowledge and must rely on their physicians for advice. Physicians must respect their patients' autonomy, fully discussing the alternatives for care and the risks, benefits, and likely consequences of each alternative. Special care should be taken to ensure that a physician seeking a patient's informed consent does not have a real or apparent conflict of interest involving personal gain.

Patients are required to sign a consent form for essentially any diagnostic or therapeutic procedure. In such cases, it is particularly important for the patient to understand clearly the risks and benefits of these procedures; this is the definition of *informed consent*. It is incumbent on the physician to explain the procedures in a clear and understandable manner and to ascertain that the patient comprehends both the nature of the procedure and the attendant risks and benefits. The dread of the unknown, inherent in hospitalization, can be mitigated by such explanations.

The Approach to Grave Prognoses and Death

No problem is more distressing than the diagnosis of an incurable disease, particularly when premature death is inevitable. What should the patient and family be told? What measures should be taken to maintain life? What can be done to maintain the quality of life?

Although some would argue otherwise, there is no ironclad rule that the patient must immediately be told "everything" even if the patient is an adult with substantial family responsibilities. Nevertheless, openness and honesty with the patient is a must. A patient must know the expected course of disease to make appropriate plans and preparations. The patient should participate in decision-making with an understanding of the treatment goals (cure or palliation), the disease effects, and the likely treatment effects. A wise and insightful physician often is guided by an understanding of what a patient wants to know and when he or she wants to know it. The patient's religious beliefs also may be taken into consideration. The patient must be given an opportunity to talk with the physician and ask questions. Patients may find it easier to share their feelings about death with their physician, who is likely to be more objective and less emotional, than with family members. As William Osler wrote, "One thing is certain; it is not for you to don the black cap and, assuming the judicial function, take hope away from any patient." Even when the patient directly inquires, "Am I dying?" the physician must attempt to determine whether this is a request for information or for reassurance. Only open communication between the patient and the physician can resolve this question and guide the physician in what to say and how to say it.

The physician should provide or arrange for emotional, physical, and spiritual support and must be compassionate, unhurried, and open. There is much to be gained by the laying on of hands. Pain should be controlled adequately, human dignity maintained, and isolation from family and close friends avoided. These aspects of care tend to be overlooked in hospitals, where the intrusion of life-sustaining apparatus can detract from attention to the whole person and encourage concentration instead on the life-threatening disease, against which the battle ultimately will be lost in any case. In the face of terminal illness, the goal of medicine must shift from *cure* to *care* in the broadest sense of the term. *Primum succurrere*, first hasten to provide help, is a guiding principle. In offering care to a dying patient, a physician must be prepared to provide information to family members and deal with their grief and sometimes their feelings of guilt. It is important for the doctor to assure the family that everything possible has been done. **For a more complete discussion of end-of-life care, see Chap. 9**.

THE PATIENT-PHYSICIAN RELATIONSHIP

The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both the diagnosis and treatment are directly dependent on it. One of the essential qualities of the clinician is interest in humanity, ***for the secret of the care of the patient is in caring for the patient*** *.*

—Francis W. Peabody, 1881–1927

Physicians must never forget that patients are individual human beings with problems that all too often transcend their physical complaints. They are not "cases" or "admissions" or "diseases." Patients do not fail treatments; treatments fail to benefit patients. This point is particularly important in this era of

high technology in clinical medicine. Most patients are anxious and fearful. Physicians should instill confidence and should be reassuring but should never be arrogant. A professional attitude, coupled with warmth and openness, can do much to alleviate anxiety and to encourage patients to share all aspects of their medical history. Empathy and compassion are the essential features of a caring physician. Whatever the patient's attitude is, the physician needs to consider the setting in which an illness occurs—in terms not only of the patients themselves but also of their familial, social, and cultural backgrounds. The ideal patient-physician relationship is based on thorough knowledge of the patient, mutual trust, and the ability to communicate.

The Dichotomy of Inpatient and Outpatient Internal Medicine

The hospital environment has changed dramatically over the last few decades. In more recent times, emergency departments and critical care units have evolved to identify and manage critically ill patients, allowing them to survive formerly fatal diseases. There is increasing pressure to reduce the length of stay in the hospital and to manage complex disorders in the outpatient setting. This transition has been driven not only by efforts to reduce costs but also by the availability of new outpatient technologies, such as imaging and percutaneous infusion catheters for long-term antibiotics or nutrition, minimally invasive surgical procedures, and evidence that outcomes often are improved by minimizing inpatient hospitalization. Hospitals now consist of multiple distinct levels of care, such as the emergency department, procedure rooms, overnight observation units, critical care units, and palliative care units, in addition to traditional medical beds. A consequence of this differentiation has been the emergence of new specialties such as emergency medicine, intensivists, hospitalists, and end-of-life care. Moreover, these systems frequently involve "handoffs" from the outpatient to the inpatient environment, from the critical care unit to a general medicine floor, and from the hospital to the outpatient environment. Clearly, one of the important challenges in internal medicine is to maintain continuity of care and information flow during these transitions, which threaten the traditional one-to-one relationship between patient and physician. In the current environment, teams of physicians, specialists, and other health care professionals often replace the personal interaction between doctor and patient. The patient can benefit greatly from effective collaboration among a number of health care professionals; however, *it is the duty of the patient's principal or primary physician to provide cohesive guidance through an illness*. To meet this challenge, the primary physician must be familiar with the techniques, skills, and objectives of specialist physicians and allied health professionals. The primary physician must ensure that the patient will benefit from scientific advances and from the expertise of specialists when they are needed while retaining responsibility for the major decisions concerning diagnosis and treatment.

Appreciation of the Patient's Hospital Experience

The hospital is an intimidating environment for most individuals. Hospitalized patients find themselves surrounded by air jets, buttons, and glaring lights; invaded by tubes and wires; and beset by the numerous members of the health care team—nurses, nurses' aides, physicians' assistants, social workers, technologists, physical therapists, medical students, house officers, attending and consulting physicians, and many others. They may be transported to special laboratories and imaging facilities replete with blinking lights, strange sounds, and unfamiliar personnel; they may be left unattended for periods of time; they may be obliged to share a room with other patients, who have their own health problems. It is little wonder that patients may lose their sense of reality. Physicians who can appreciate the hospital experience from the patient's perspective and make an effort to develop a strong personal relationship with the patient in which they may guide the patient through this experience can make a stressful situation more tolerable.

Trends in the Delivery of Health Care: A Challenge to the Humane Physician

Many trends in the delivery of health care tend to make medical care impersonal. These trends, some of which have been mentioned already, include (1) vigorous efforts to reduce the escalating costs of health care; (2) the growing number of managed-care programs, which are intended to reduce costs but in which the patient may have little choice in selecting a physician or in seeing that physician consistently; (3) increasing reliance on technological advances and computerization for many aspects of diagnosis and treatment; (4) the need for numerous physicians to be involved in the care of most patients who are seriously ill; and (5) an increased number of malpractice suits, some of which are justifiable because of medical errors but others of which reflect an unrealistic expectation on the part of many patients that their disease will be cured or that complications will not occur during the course of complex illnesses or procedures.

In light of these changes in the medical care system, it is a major challenge for physicians to maintain the *humane* aspects of medical care. The American Board of Internal Medicine, working together with the American College of Physicians–American Society of Internal Medicine and the European Federation of Internal Medicine, has published a *Charter on Medical Professionalism* that underscores three main principles in physicians' contract with society: (1) the primacy of patient welfare, (2) patient autonomy, and (3) social justice. Medical schools appropriately place substantial emphasis on physician professionalism (Fig. 1-1). The humanistic qualities of a physician must encompass integrity, respect, and compassion. Availability, the expression of sincere concern, the

willingness to take the time to explain all aspects of the illness, and a nonjudgmental attitude when dealing with patients whose cultures, lifestyles, attitudes, and values differ from those of the physician are just a few of the characteristics of a humane physician. Every physician will, at times, be challenged by patients who evoke strongly negative or positive emotional responses. Physicians should be alert to their own reactions to such patients and situations and consciously monitor and control their behavior so that the patient's best interest remains the principal motivation for their actions at all times.

Figure 1-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A typical "white coat" ceremony in medical school in which students are introduced to the responsibilities of patient care. *(Photo courtesy of Suzanne Camarata Photography; used with permission.)*

An important aspect of patient care involves an appreciation of the patient's "quality of life," a subjective assessment of what each patient values most. This assessment requires detailed, sometimes intimate knowledge of the patient, which usually can be obtained only through deliberate, unhurried, and often repeated conversations. Time pressures will always threaten these interactions, but they should not diminish the importance of understanding and seeking to fulfill the priorities of the patient.

THE TWENTY-FIRST-CENTURY PHYSICIAN: EXPANDING FRONTIERS

The era of "omics": genomics, epigenomics, proteomics, microbiomics, metagenomics, metabolomics . . .

In the spring of 2003, the complete sequencing of the human genome was announced, officially ushering in the genomic era. However, even before that landmark accomplishment, the practice of medicine had been evolving as a result of the insights gained from an understanding of the human genome as well as the genomes of a wide variety of microbes, whose genetic sequences were becoming widely available as a result of breathtaking advances in sequencing techniques and informatics. An example is the rapid identification of H1N1 influenza as a potentially fatal pandemic illness and the rapid development and dissemination of an effective protective vaccine. Today, gene expression profiles are being used to guide therapy and inform prognosis for a number of diseases, the use of genotyping is providing a new means to assess the risk of certain diseases as well as variation in response to a number of drugs, and physicians are beginning to understand better the role of certain genes in the causality of common conditions such as obesity and allergies. Despite these advances, scientists are still in the infancy of understanding and utilizing the complexities of genomics in the diagnosis, prevention, and treatment of disease. The task of physicians is complicated by the fact that phenotypes generally are determined not by genes alone but by the interplay of genetic and environmental factors. Indeed, researchers have just begun to scratch the surface of possibilities that the era of genomics will provide to the practice of medicine.

Rapid progress also is being made in other areas of molecular medicine. *Epigenomics* is the study of alterations in chromatin and histone proteins and methylation of DNA sequences that influence gene expression. Epigenetic alterations are associated with a number of cancers and other diseases. The study of the entire library of proteins made in a cell or organ and its relationship to disease is called *proteomics*. Proteomics is now recognized as far more complex than originally considered, enhancing the repertoire of the 30,000 genes in the human genome by alternate splicing and posttranslational processing as well as by an increasing number of posttranslational modifications, many with unique

functional consequences. The presence or absence of particular proteins in the circulation or in cells is being explored for diagnostic and disease-screening uses. *Microbiomics* is the study of the bacterial flora of a person. Interesting research is suggesting that the composition of colonic flora may play a role in obesity and in other diseases. *Metagenomics*, of which microbiomics is a part, is the genomic study of environmental species that have the potential to influence human biology directly or indirectly. An example is the study of exposures to microorganisms in farm environments that might be responsible for the lower incidence of asthma among farm-raised children. *Metabolomics* is the study of the range of metabolites in cells or organs and the ways they are altered in disease states. The aging process itself may leave telltale metabolic footprints that allow the prediction (and possibly the prevention) of dysfunction and disease. It seems likely that disease-associated patterns will be sought in lipids, carbohydrates, membranes, mitochondria, and other vital components of cells and tissues. All this new information represents a challenge to the traditional reductionist approach to medical thinking. The variability of results in different patients, together with the large number of variables that can be assessed, creates difficulties in identifying preclinical disease and defining disease states unequivocally. Accordingly, the tools of systems biology are being applied to the myriad information now obtainable from every patient and may provide new approaches to classifying disease. **For a more complete discussion of a complex systems approach to human disease, see Chap. e19.**

The rapidity of these advances may seem overwhelming to the practicing physician. However, he or she has an important role to play in ensuring that these powerful technologies and sources of new information are applied with sensitivity and intelligence to the patient. Since "omics" is such a rapidly evolving field, physicians and other health care professionals must continue to educate themselves so that they can apply this new knowledge to the benefit of their patients' health and well-being. Genetic testing requires wise counsel based on an understanding of the value and limitations of the tests as well as the implications of their results for specific individuals. **For a more complete discussion of genetic testing, see Chap. 63.**

The Globalization of Medicine

Physicians should be cognizant of diseases and health care services beyond local boundaries. Global travel has implications for disease spread, and it is not uncommon for diseases endemic to certain regions to be seen in other regions after a patient has traveled to and returned from those regions. Patients have broader access to unique expertise or clinical trials at distant medical centers, and the cost of travel may be offset by the quality of care at those distant locations. As much as any other factor influencing global aspects of medicine, the Internet has transformed the transfer of medical information throughout the world. This change has been accompanied by the transfer of technological skills through telemedicine and international consultation for radiologic images and pathologic specimens. **For a complete discussion of global issues, see Chap. 2.**

Medicine on the Internet

On the whole, the Internet has had a very positive effect on the practice of medicine; a wide range of information is available to physicians and patients through personal computers almost instantaneously at any time and from anywhere in the world. This medium holds enormous potential for delivering current information, practice guidelines, state-of-the-art conferences, journal contents, textbooks (including this text), and direct communications with other physicians and specialists, expanding the depth and breadth of information available to the physician about the diagnosis and care of patients. Medical journals are now accessible online, providing rapid sources of new information. This medium also serves to lessen the information gap felt by physicians and health care providers in remote areas by bringing them into direct and timely contact with the latest developments in medical care.

Patients, too, are turning to the Internet in increasing numbers to acquire information about their illnesses and therapies and to join Internet-based support groups. Physicians increasingly are faced with the prospect of dealing with patients who arrive with sophisticated information about their illnesses. In this regard, physicians are challenged in a positive way to keep abreast of the latest relevant information while serving as an "editor" for the patients as they navigate this seemingly endless source of information, the accuracy and validity of which are not uniform.

A critically important caveat is that virtually anything can be published on the Internet, with easy circumvention of the peer-review process that is an essential feature of academic publications. Physicians or patients who search the Internet for medical information must be aware of this danger. Notwithstanding this limitation, appropriate use of the Internet is revolutionizing information access for physicians and patients and in this regard is a great benefit that was not available to earlier practitioners.

Public Expectations and Accountability

The level of knowledge and sophistication regarding health issues on the part of the general public has grown rapidly over the last few decades. As a result, expectations of the health care system in general and of physicians in particular have risen. Physicians are expected to master rapidly advancing fields (the *science* of medicine) while considering their patients' unique needs (the *art* of medicine). Thus, physicians are held accountable not only for the technical aspects of the care that

they provide but also for their patients' satisfaction with the delivery and costs of care.

In many parts of the world, physicians increasingly are expected to account for the way in which they practice medicine by meeting certain standards prescribed by federal and local governments. The hospitalization of patients whose health care costs are reimbursed by the government and other third parties is subjected to utilization review. Thus, a physician must defend the cause for and duration of a patient's hospitalization if it falls outside certain "average" standards. Authorization for reimbursement increasingly is based on documentation of the nature and complexity of an illness, as reflected by recorded elements of the history and physical examination. There is a growing "pay for performance" movement that seeks to link reimbursement to quality of care. The goal of this movement is to improve standards of health care and contain spiraling health care costs. Physicians also are expected to give evidence of their continuing competence through mandatory continuing education, patient record audits, maintenance of certification, and relicensing.

Medical Ethics and New Technologies

The rapid pace of technological advances has profound implications for medical applications far beyond their traditional roles to prevent, treat, and cure disease. Cloning, genetic engineering, gene therapy, human-computer interfaces, nanotechnology, and designer drugs have the potential to modify inherited predispositions to disease, select desired characteristics in embryos, augment "normal" human performance, replace failing tissues, and substantially prolong life span. Because of their unique training, physicians have a responsibility to help shape the debate concerning the appropriate uses of and limits that should be placed on these new techniques.

The Physician as Perpetual Student

It becomes all too apparent from the time doctors graduate from medical school that as physicians their lot is that of the "perpetual student" and the mosaic of their knowledge and experiences is eternally unfinished. This concept can be at the same time exhilarating and anxiety-provoking. It is exhilarating because doctors will continue to expand knowledge that can be applied to their patients; it is anxiety-provoking because doctors realize that they will never know as much as they want or need to know. At best, doctors will translate this latter feeling into energy to continue to improve themselves and realize their potential as physicians. In this regard, it is the responsibility of a physician to pursue new knowledge continually by reading, attending conferences and courses, and consulting colleagues and the Internet. This is often a difficult task for a busy practitioner; however, such a commitment to continued learning is an integral part of being a physician and must be given the highest priority.

The Physician as Citizen

Being a physician is a privilege. The capacity to apply one's skills for the benefit of one's fellow human beings is a noble calling. The doctor-patient relationship is inherently unbalanced in the distribution of power. In light of a doctor's influence, he or she must always be aware of the potential impact of what he or she does and says and must always strive to strip away individual biases and preferences to find what is best for the patient. To the extent possible, a physician also should try to act within his or her community to promote health and alleviate suffering. Meeting these goals begins by setting a healthy example and continues in actions that may be taken to deliver needed care even when personal financial compensation may not be available.

G. H. T. Kimble wrote: "It is bad enough that a [person] should be ignorant, for this cuts him [or her] off from the commerce of [people's] minds. It is perhaps worse that a [person] should be poor, for this condemns him [or her] to a life of stint and scheming in which there is no time for dreams and no respite from weariness. But what surely is worse is that a [person] should be unwell, for this prevents his [or her] doing anything much about either his [or her] poverty or his [or her] ignorance." A goal for medicine and its practitioners is to strive to provide the means by which the poor can cease to be unwell.

Learning Medicine

It has been about 100 years since the publication of the Flexner Report, a seminal study that transformed medical education and emphasized the scientific foundations of medicine as well as the acquisition of clinical skills. In an era of burgeoning information and access to medical simulation and informatics, many schools are implementing new curricula that emphasize lifelong learning and the acquisition of competencies in teamwork, communication skills, system-based practice, and professionalism. These and other features of the medical school curriculum provide the foundation for many of the themes highlighted in this chapter and are expected to allow physicians to progress from competency to proficiency to mastery with progressive experience and learning.

At a time when the amount of information that one must master to practice medicine continues to expand, increasing pressures both within and outside of medicine have produced strict restrictions on the amount of time a physician in training can spend in the hospital. It was felt that the benefits associated with the continuity of medical care and observation of the patient's progress over time were outstripped by the stresses of long hours on the trainees and the fatigue-related errors they made in caring for patients. Accordingly, physicians in training had limits set on the number of patients they

could carry at a time, the number of new patients they could evaluate in a day on call, and the number of hours they could spend in the hospital. In 1980, residents in medicine worked in the hospital more than 90 hours a week on average. In 1989, their hours were restricted to no more than 80 a week. Resident physicians' hours further decreased by about 10% between 1996 and 2008, and in 2010, the Accreditation Council for Graduate Medical Education (ACGME) placed further restrictions on continuing in-hospital duty hours for first year residents (16 hours/shift). The impact of these changes is continuing to be assessed, but the evidence that medical errors have decreased as a consequence is sparse. An unavoidable by-product of fewer hours at work is an increase in the number of "handoffs" of patient responsibility from one physician to another. These transfers often involve a transition from a physician who knows the patient well, having evaluated the patient on admission, to a physician who knows the patient less well. It is imperative that these transitions of responsibility be handled with care and thoroughness with all the relevant information exchanged and acknowledged. The issue of coverage is not limited to physicians in graduate training. The average practicing physician worked 54 hours per week in 1996–1998 and 51 hours per week in 2006–2008.

Research, Teaching, and the Practice of Medicine

The title *doctor* is derived from the Latin *docere*, "to teach," and physicians should share information and medical knowledge with colleagues, students of medicine and related professions, and their patients. The practice of medicine is dependent on the sum total of medical knowledge, which in turn is based on an unending chain of scientific discovery, clinical observation, analysis, and interpretation. Advances in medicine depend on the acquisition of new information through research, and improved medical care requires the transmission of that information. As part of broader societal responsibilities, the physician should encourage patients to participate in ethical and properly approved clinical investigations if they do not impose undue hazard, discomfort, or inconvenience. However, physicians engaged in clinical research must be alert to potential conflicts of interest between their research goals and their obligations to individual patients; the best interests of the patient must always take priority.

To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.

—William Osler, 1849–1919

FURTHER READINGS

Charap MH et al: Internal medicine residency training in the 21st century: Aligning requirements with professional needs. Am J Med 118:1042, 2005 [PMID: 16164893]

Cooke M et al: American medical education 100 years after the Flexner Report. N Engl J Med 355:1339, 2006 [PMID: 17005951]

Council on Graduate Medical Education: *Thirteenth Report: Physician Education for a Changing Health Care Environment*. U.S. Department of Health and Human Services, Washington, DC

Hunter DJ et al: From Darwin's finches to canaries in the coal mine—mining the genome for new biology. N Engl J Med 358:2760, 2008 [PMID: 18579810]

Kinghorn WA: Medical education as moral formation: An Aristotelian account of medical professionalism. Perspect Biol Med 53:87, 2009

Loscalzo J et al: Human disease classification in the postgenomic era: A complex systems approach. Mol Syst Biol 3:124, 2007 [PMID: 17625512]

Staiger DO et al: Trends in the hours of physicians in the United States. JAMA 303:747, 2010 [PMID: 20179284]

Straus SE et al: Teaching evidence-based medicine skills can change practice in a community hospital. J Gen Intern Med 20:340, 2005 [PMID: 15857491]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

WHY GLOBAL HEALTH?

Global health, it has been noted, is not a discipline; it is, rather, a collection of problems. A leading group of scholars have defined global health as the study and practice concerned with improving the health of all people and achieving health equity worldwide, with an emphasis on addressing problems that are transnational. No single review can do much more than identify the leading problems in applying evidence-based medicine in settings of great poverty or across national boundaries. This chapter introduces the major international bodies that address these problems; identifies the more significant barriers to improving the health of people who to date have not, by and large, had access to modern medicines; and summarizes population-based data on the most common health problems faced by people living in poverty. Examining specific problems—notably AIDS (Chap. 189) but also tuberculosis (TB, Chap. 165), malaria (Chap. 210), and key noncommunicable diseases—helps sharpen the discussion of barriers to prevention, diagnosis, and care as well as the means of overcoming them. The chapter then discusses the role of health systems and the problem of "brain drain" on those systems. It closes by discussing global health equity, drawing on notions of social justice that once were central to international public health but have fallen out of favor over the last several decades.

BRIEF HISTORY OF GLOBAL HEALTH INSTITUTIONS

Concern about health across national boundaries dates back many centuries, predating the Black Plague and other pandemics. The first organization founded explicitly to tackle cross-border health issues was the Pan American Sanitary Bureau, which was formed by 11 countries in the Americas in 1902. The primary goal of what later became the Pan American Health Organization was the control of infectious diseases across the Americas. Of special concern was yellow fever, which had been running a deadly course through much of South and Central America and posed a threat to the construction of the Panama Canal. In 1948, the United Nations formed the first truly global health institution: the World Health Organization (WHO). In 1958, under the aegis of the WHO and in line with a long-standing focus on communicable diseases that cross borders, leaders in global health initiated the effort that led to what some see as the greatest success in international health: the eradication of smallpox. Naysayers were surprised when the smallpox eradication campaign, which engaged public health officials throughout the world, proved successful in 1979 during the Cold War. The influence of the WHO waned during the 1980s. In the early 1990s, many observers argued that with its vastly superior financial resources and close if unequal relationships with the governments of poor countries, the World Bank had eclipsed the WHO as the most important multilateral institution working in the area of health. One of the stated goals of the World Bank was to help poor countries identify "cost-effective" interventions worthy of international public support. At the same time, the World Bank encouraged many of those nations to reduce public expenditures in health and education as part of later discredited structural adjustment programs that were imposed as a condition for access to credit and assistance through international financial institutions such as the World Bank and the International Monetary Fund (IMF). At the same time, there was a resurgence of many diseases, including malaria, trypanosomiasis, are schistosomiasis, in Africa.Tuberculosis, an eminently curable disease, remained the world's leading infectious killer of adults. Half a million women per year died in childbirth during the last decade of the twentieth century, and few of the world's largest philanthropic or funding institutions focused on global health.

AIDS, first described in 1981, precipitated a change. In the United States, the advent of this newly described infectious killer marked the culmination of a series of events that discredited talk of "closing the book" on infectious diseases. In Africa, which would emerge as the global epicenter of the pandemic, HIV disease strained TB control programs, and malaria continued to take as many lives as ever. At the dawn of the twenty-first century, these three diseases alone killed an estimated 6 million people each year. New research, new policies, and new funding mechanisms were called for. The last decade has seen the rise of important multilateral global health institutions such as the Global Fund to Fight AIDS, Tuberculosis, and Malaria (GFATM) and the Joint United Nations Programme on HIV/AIDS (UNAIDS); bilateral efforts such as the U.S. President's Emergency Plan for AIDS Relief (PEPFAR); and private philanthropic organizations such as the Bill & Melinda Gates Foundation. Yet with its 193 member states and 147 country offices, the WHO remains preeminent in matters relating to the cross-border spread of infectious diseases and other health threats. In the aftermath of the severe acute respiratory syndrome (SARS) epidemic of 2003, the International Health Regulations—which provide a

legal foundation for the WHO's direct investigation of a wide range of global health problems, including pandemic influenza, in any member state—were strengthened and brought into force in May 2007.

Even as attention to and resources for health problems in poor countries grow, the lack of coherence in and among global health institutions may undermine efforts to forge a more comprehensive and effective response. The WHO is still woefully underfunded despite the ever-growing need to engage a wider and more complex range of health issues. In another instance of the paradoxical impact of success, the rapid growth of the Gates Foundation, which is clearly one of the most important developments in the history of global health, has led other foundations to question the wisdom of continuing to invest their more modest resources in this field. This indeed may be what some have called "the golden age of global health," but leaders of major organizations such as the WHO, the GFATM, the United Nations Children's Fund (UNICEF), UNAIDS, PEPFAR, and the Gates Foundation must work together to design an effective architecture that will make the most of the opportunities that now exist. To this end, new and old players in global health must invest heavily in *discovery* (relevant basic science), the *development* of new tools (preventive, diagnostic, and therapeutic), and *delivery* to ensure the equitable provision of health products and services to all who need them.

THE ECONOMICS OF GLOBAL HEALTH

Political and economic concerns have often guided global health interventions. As mentioned, early efforts to control yellow fever were tied to the completion of the Panama Canal. However, the precise nature of the link between economics and health remains a matter for debate. Some economists and demographers argue that improving the health status of populations must begin with economic development; others maintain that addressing ill health is the starting point for development in poor countries. In either case, investment in health care, especially the control of communicable diseases, should lead to increased productivity. The question is where to find the necessary resources to start the predicted "virtuous cycle."

Since 1999, spurred by the leadership of the Gates Foundation and the growing interest in addressing novel and persistent challenges such as AIDS, spending on health in poor countries has increased, with over $80 billion in new funds earmarked for the discovery and development of drugs and diagnostics targeting diseases of the poor; comprehensive responses to the AIDS, TB, and malaria epidemics; vaccine development and delivery; and even improved methods of data collection in resource-poor settings. Nevertheless, to reach the United Nations Millennium Development Goals, that include targets for poverty reduction, universal primary education, and gender equality, spending in the health sector must be increased further. To determine by how much and for how long, it is imperative to improve the ability to assess the global burden of disease (GBD) and plan interventions that more precisely match the need. Refining metrics is an important task for global health: only recently have there been solid assessments of the GBD.

MORTALITY AND THE GLOBAL BURDEN OF DISEASE

Since the late 1980s, serious efforts have been made to calculate the GBD. The first GBD study, conducted in 1990, laid the foundation for the first report on *Disease Control Priorities in Developing Countries* (DCP1) and for the World Bank's 1993 World Development Report *Investing in Health*. Those efforts represented a major advance in the understanding of health status in developing countries. *Investing in Health* has been especially influential: it familiarized a broad audience with cost-effectiveness analysis for specific health interventions and with the notion of disability-adjusted life years (DALYs). The DALY, which has become a standard measure of the impact of a specific health condition on a population, combines in a single measure both absolute years of life lost and years lost due to disability for incident cases of a condition. (See Fig. 2-1 and Table 2-1 for an analysis of the GBD by DALYs.)

Figure 2-1

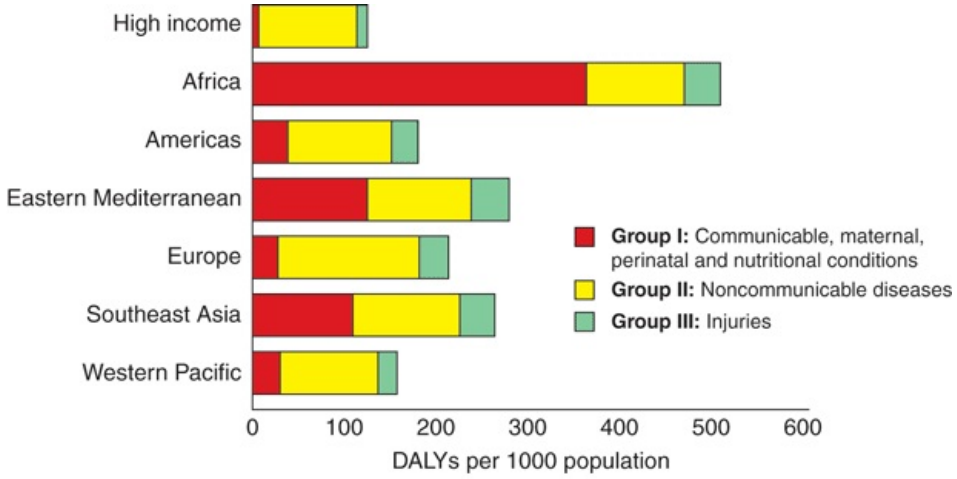

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Burden of disease, by broad cause and region, 2004. *[Source: World Health Organization. Reprinted with permission (http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf; accessed January 10, 2010).]*

Table 2-1 Leading Causes of Burden of Disease (DALYs), with Countries Grouped by Income, 2004

Disease or Injury	DALYs (millions)	Percent of total DALYs	Disease or Injury	DALYs (millions)	Percent of total DALYs
World			**Middle-income countries**		
1 Lower respiratory infections	94.5	6.2	1 Unipolar depressive disorders	29.0	5.1
2 Diarrheal diseases	72.8	4.8	2 Ischemic heart disease	28.9	5.0
3 Unipolar depressive disorders	65.5	4.3	3 Cerebrovascular disease	27.5	4.8
4 Ischemic heart disease	62.6	4.1	4 Road traffic accidents	21.4	3.7
5 HIV/AIDS	58.5	3.8	5 Lower respiratory infections	16.3	2.8
6 Cerebrovascular disease	46.6	3.1	6 COPD	16.1	2.8
7 Prematurity and low birth weight	44.3	2.9	7 HIV/AIDS	15.0	2.6
8 Birth asphyxia and birth trauma	41.7	2.7	8 Alcohol use disorders	14.9	2.6
9 Road traffic accidents	41.2	2.7	9 Refractive errors	13.7	2.4
10 Neonatal infections and other[b]	40.4	2.7	10 Diarrheal diseases	13.1	2.3
High-income countries			**Low-income countries[a]**		
1 Unipolar depressive disorders	10.0	8.2	1 Lower respiratory infections	76.9	9.3
2 Ischemic heart disease	7.7	6.3	2 Diarrheal diseases	59.2	7.2

3 Cerebrovascular disease	4.8	3.9	3 HIV/AIDS	42.9	5.2
4 Alzheimer's and other dementias	4.4	3.6	4 Malaria	32.8	4.0
5 Alcohol use disorders	4.2	3.4	5 Prematurity and low birth weight	32.1	3.9
6 Hearing loss, adult onset	4.2	3.4	6 Neonatal infections and other[b]	31.4	3.8
7 COPD	3.7	3.0	7 Birth asphyxia and birth trauma	29.8	3.6
8 Diabetes mellitus	3.6	3.0	8 Unipolar depressive disorders	26.5	3.2
9 Trachea, bronchus, lung cancers	3.6	3.0	9 Ischemic heart disease	26.0	3.1
10 Road traffic accidents	3.1	2.6	10 Tuberculosis	22.4	2.7

[a]Countries grouped by gross national income per capita.

[b]This category also includes other noninfectious causes arising in the perinatal period apart from prematurity, low birth weight, birth trauma, and asphyxia. These noninfectious causes are responsible for about 20% of the DALYs shown in this category.

Abbreviation: COPD, chronic obstructive pulmonary disease.

Source: World Health Organization. Reprinted with permission. http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf.

The most recent WHO analysis of the GBD was based on health data from 2004. This report reflects growth in the available data on health in the poorest countries and in the capacity to measure the impact of specific conditions on a population. Yet even in 2004, only 112 of 192 nations surveyed had reliable information on the causes of deaths within their borders. It is essential to expand efforts to collect the most basic health data; this task falls to the WHO, national governments, and certain academic institutions. The lack of complete data has led to considerable uncertainty in estimates of overall mortality rates. The level of uncertainty ranges from as low as ± 1% for estimates of all-cause mortality in developed countries to ± 20% for all-cause mortality in the WHO's African Region. The level of uncertainty in regional prevalence estimates ranges from ± 10% to ± 90%, with a median value of ± 41%. As analytic methods and data quality have improved, however, important trends can be identified in a comparison of GBD estimates from 1990 and 2004.

Of the 58.8 million deaths worldwide in 2004, 30% were due to communicable diseases, maternal and perinatal conditions, and nutritional deficiencies. Although the proportion of all deaths attributable to these causes has decreased marginally since 1990, the share of all deaths due to HIV/AIDS grew from just 2% to >3.5%. Among the fraction of all deaths related to communicable diseases, maternal and perinatal conditions, and nutritional deficiencies, 97% occurred in middle- and low-income countries. The leading cause of death among adults in 2004 was ischemic heart disease, accounting for 16.3% of all deaths in high-income countries, 13.9% in middle-income countries, and 9.4% in low-income countries **(Table 2-2)**. In second place was cerebrovascular disease, that accounted for 9.3% of deaths in high-income countries, 14.2% in middle-income countries, and 5.6% in low-income countries. Although the third leading cause of death in high-income countries was tracheal, bronchial, and lung cancers (that accounted for 5.9% of all deaths), those conditions did not even register in the top 10 places in low-income countries. Among the 10 leading causes of death in low-income countries, 6 were communicable diseases; in high-income countries, however, only one communicable disease—lower respiratory infection—ranked among the top 10 causes of death.

Table 2-2 Leading Causes of Death Worldwide, by Income Group, 2004

Disease or Injury	Deaths (millions)	Percent of Total Deaths	Disease or Injury	Deaths (millions)	Percent of Total Deaths
World			**Middle-income countries**		
1 Ischemic heart disease	7.2	12.2	1 Cerebrovascular disease	3.5	14.2

2 Cerebrovascular disease	5.7	9.7	2 Ischemic heart disease	3.4	13.9
3 Lower respiratory infections	4.2	7.1	3 COPD	1.8	7.4
4 COPD	3.0	5.1	4 Lower respiratory infections	0.9	3.8
5 Diarrheal diseases	2.2	3.7	5 Trachea, bronchus, lung cancers	0.7	2.9
6 HIV/AIDS	2.0	3.5	6 Road traffic accidents	0.7	2.8
7 Tuberculosis	1.5	2.5	7 Hypertensive heart disease	0.6	2.5
8 Trachea, bronchus, lung cancers	1.3	2.3	8 Stomach cancer	0.5	2.2
9 Road traffic accidents	1.3	2.2	9 Tuberculosis	0.5	2.2
10 Prematurity and low birth weight	1.2	2.0	10 Diabetes mellitus	0.5	2.1
High-income countries			**Low-income countries**[a]		
1 Ischemic heart disease	1.3	16.3	1 Lower respiratory infections	2.9	11.2
2 Cerebrovascular disease	0.8	9.3	2 Ischemic heart disease	2.5	9.4
3 Trachea, bronchus, lung cancers	0.5	5.9	3 Diarrheal diseases	1.8	6.9
4 Lower respiratory infections	0.3	3.8	4 HIV/AIDS	1.5	5.7
5 COPD	0.3	3.5	5 Cerebrovascular disease	1.5	5.6
6 Alzheimer's and other dementias	0.3	3.4	6 COPD	0.9	3.6
7 Colon and rectum cancers	0.3	3.3	7 Tuberculosis	0.9	3.5
8 Diabetes mellitus	0.2	2.8	8 Neonatal infections[b]	0.9	3.4
9 Breast cancer	0.2	2.0	9 Malaria	0.9	3.3
10 Stomach cancer	0.1	1.8	10 Prematurity and low birth weight	0.8	3.2

[a]Countries grouped by gross national income per capita: low income ($825 or less), high income ($10,066 or more). Note that these high-income groups differ slightly from those used in the Disease Control Priorities Project.

[b]This category also includes other noninfectious causes arising in the perinatal period, that are responsible for about 20% of the deaths shown in this category.

Abbreviation: COPD, chronic obstructive pulmonary disease.

Source: World Health Organization. Reprinted with permission. http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf.

A recent study found that the worldwide mortality figure among children <5 years of age dropped from 11.9 million deaths in 1990 to 7.7 million in 2010. Of the deaths in 2010, 3.1 million (40%) occurred in the neonatal period. About one-third of these deaths among children <5 years old occurred in southern Asia and almost one-half in sub-Saharan Africa; <1% occurred in high-income countries.

Among persons 15–59 years of age, noncommunicable diseases accounted for more than one-half of all deaths in all regions except sub-Saharan Africa, where communicable diseases, maternal and perinatal conditions, and nutritional deficiencies together accounted for two-thirds of all deaths. Indeed, the HIVmortality rate among 15- to 59-year-olds in sub-Saharan Africa was higher than the mortality rate due to all causes among adults in high-income countries. In this age group, injuries accounted for 23% of all deaths worldwide. Overall, death rates in this age group declined between 1990 and 2004 in all areas except Europe and Central Asia, where cardiovascular diseases and injuries caused increased mortality rates, and sub-Saharan Africa, where the impact of HIV/AIDS in this age cohort was particularly devastating.

There is greater uncertainty in calculating years of life lived with disability for specific conditions than in calculating years of life lost. Best estimates from 2004 reveal that although the prevalence of diseases common in older populations (e.g., dementia and musculoskeletal disease) was higher in high-income countries, the disability experienced as a result of cardiovascular diseases, chronic respiratory diseases, and the long-term impact of communicable diseases was greater in low- and middle-income countries. In most low- and middle-income countries, people lived shorter lives and experienced disability and poor health for a greater proportion of their lives. Indeed, >50% of the GBD occurred in southern Asia and sub-Saharan Africa, which together account for only one-third of the world's population.

Noncommunicable diseases accounted for almost 60% of all deaths in 2004 but, because of the later onset of those diseases, accounted for only 48% of years of life lost. In contrast, because they more often involve younger people, injuries accounted for 12% of years of life lost but for only 10% of deaths. Notably, 45% of the disease burden in middle-income countries in 2004 resulted from noncommunicable conditions; in 1990, the figure was 35%.

Poverty remains one of the most important root causes of poor health worldwide, and the global burden of poverty continues to be high. Among the 6.8 billion people alive today, 43% (~2.7 billion) live on <$2 per day and 17% (~1.1 billion) live on <$1 per day. Comparison of national health indicators with gross domestic product per capita among nations shows a clear relationship between higher gross domestic product and better health, with only a few outliers. Numerous studies also have documented the link between poverty and health within countries.

The Global Burden of Disease Study found that undernutrition was the leading risk factor for poor health. In an era that has seen obesity become a major health concern in many developed countries, the persistence of undernutrition is surely cause for great consternation. Inability to feed the hungry provides evidence of many years of failed development projects and must be addressed as a problem of the highest priority. Indeed, no health care initiative, however generously funded and scientifically justified, will be effective without adequate nutrition.

The second edition of *Disease Control Priorities in Developing Countries* (DCP2), published in 2006, is a document of stunning breadth and ambition, providing cost-effectiveness analyses for >100 interventions and including 21 chapters focused on strategies for strengthening health systems. Cost-effectiveness analyses that compare two relatively equal interventions and facilitate the best choices under constraint are important; however, as both resources and ambitions for global health grow, cost-effectiveness analyses (particularly those based on past conditions) must not hobble the increased worldwide commitment to provide resources and accessible services to all who need them. To illustrate this point, it is instructive to look to AIDS, that in the course of the last three decades has become the world's leading infectious cause of adult death.

AIDS

Chapter 189 provides an overview of the AIDSepidemic in the world today. Here the discussion will be limited to AIDS in the developing world. Lessons learned in tackling AIDS in resource-constrained settings are highly relevant to discussions of other chronic diseases, including noncommunicable diseases, for which effective therapies have been developed. Several of these lessons are highlighted below.

In the United States, the availability of highly active antiretroviral therapy (ART) for AIDS has transformed this disease from an inescapably fatal destruction of cell-mediated immunity into a manageable chronic illness. In developing countries, treatment has been offered more broadly only since 2003, and only in the fall of 2008 did the number of patients receiving treatment exceed 40% of the number who need it. (It remains to be seen how many of these fortunate few are receiving ART regularly and with the requisite social support.) Before 2003, many arguments were raised to justify not moving forward rapidly with ART programs for people living with HIV/AIDS in resource-limited settings. The standard litany included the price of therapy compared with the poverty of the patient, the complexity of the intervention, the lack of infrastructure for laboratory monitoring, and the lack of trained health care providers. Narrow cost-effectiveness arguments that created false dichotomies —prevention *or* treatment rather than both—too often went unchallenged. The greatest obstacle at the time was the ambivalence, if not outright silence, of political leaders and experts in public health. The cumulative effect of these factors was to condemn to death tens of millions of poor people in

developing countries who had become ill as a result of HIV infection.

The inequity between rich and poor countries in access to HIV treatment has given rise to widespread moral indignation. In several middle-income countries, including Brazil, visionary programs have bridged the access gap. Other innovative projects pioneered by international nongovernmental organizations (NGOs) in diverse settings have clearly established that a very simple approach to ART that is based on intensive community engagement and support can achieve remarkable results. In 2000, the United Nations Accelerating Access Initiative finally brought the research-based and generic pharmaceutical industries into play, and prices of AIDS drugs have fallen significantly. At the same time, fixed-dose combination drugs that are easier to administer have become more widely available.

Building on these lessons, the WHO advocated a public health approach to the treatment of people with AIDS in resource-limited settings. This approach, derived from models of care pioneered by the NGO Partners In Health and other groups, proposed standard first-line treatment regimens based on a simple five-drug formulary, with a more complex (and far more expensive) set of second-line options in reserve. Common clinical protocols were standardized, and intensive training packages for health and community health workers were developed and implemented in many countries. Those efforts were supported by new funding from the World Bank, the GFATM, and PEPFAR. In 2003, the lack of access to ART was declared a global public health emergency by the WHO and UNAIDS, and those two agencies launched the "3 by 5 initiative," setting an ambitious target: to have 3 million people in developing countries in treatment by the end of 2005. Many countries have since set corresponding national targets and have worked to integrate ART into their national AIDS programs and health systems and to harness the synergies between HIV/AIDS treatment and prevention activities. The efficacy of ART is well documented: in the United States, such therapy has prolonged life by an estimated average of 13 years per patient—a better success rate than that obtained with almost any treatment for cancer or for complications of coronary artery disease. Further lessons with implications for policy and action have come from efforts now underway in the developing world. During the last decade, through experiences in >50 countries thus far, the world has seen that ambitious policy goals, adequate funding, and knowledge about implementation can dramatically transform the prospects of people living with HIV infection in developing nations.

Tuberculosis

Chapter 165 provides a concise overview of the pathophysiology and treatment of TB, which is closely linked to HIV infection in much of the world. Indeed, a substantial proportion of the resurgence of TB registered in southern Africa may be attributed to HIV co-infection. Even before the advent of HIV, however, it was estimated that fewer than one-half of all cases of TB in developing countries were ever diagnosed, much less treated.

Primarily because of the common failure to diagnose and treat TB, international authorities devised a single strategy to reduce the burden of disease. The DOTS strategy (*d*irectly *o*bserved *t*herapy using *s*hort-course isoniazid- and rifampin-based regimens) was promoted in the early 1990s as highly cost-effective by the World Bank, the WHO, and other international bodies. Passive case finding of smear-positive patients was central to the strategy, and an uninterrupted drug supply was, of course, deemed necessary for cure. DOTS was clearly effective for most uncomplicated cases of drug-susceptible TB, but a number of shortcomings soon were identified. First, the diagnosis of TB based solely on smear microscopy—a method dating from the late nineteenth century—is not sensitive. Many patients with pulmonary TB and all patients with exclusively extrapulmonary TB are missed by smear microscopy, as are most children with active disease. Second, passive case finding relies on the availability of health care services, that is uneven in the settings where TB is most prevalent. Third, patients with multidrug-resistant TB (MDR-TB) are by definition infected with strains of *Mycobacterium tuberculosis* resistant to isoniazid and rifampin; thus, exclusive reliance on these drugs is ineffective in settings in which drug resistance is an established problem.

The crisis of antibiotic resistance registered in U.S. hospitals is not confined to the industrialized world or to bacterial infections. In some settings, a substantial minority of patients with TB are infected with strains resistant to at least one first-line anti-TB drug. As an effective DOTS-based response to MDR-TB, global health authorities adopted DOTS-Plus, that adds the diagnostics and drugs necessary to manage drug-resistant disease. Even before DOTS-Plus could be brought to scale in resource-constrained settings, however, new strains of extensively drug-resistant (XDR) *M. tuberculosis* began to threaten the success of TB control programs in already beleaguered South Africa, for example, where high rates of HIV infection have led to a doubling of TB incidence over the last decade.

Tuberculosis and AIDS as Chronic Diseases: Lessons Learned

Strategies effective against MDR-TB have implications for the management of drug-resistant HIV infection and even drug-resistant malaria, which, through repeated infections and a lack of effective therapy, has become a chronic disease in parts of Africa. Indeed, examining AIDS and TB together as chronic diseases makes it possible to draw a number of conclusions, many of them pertinent to global health in general (Fig. 2-2).

Figure 2-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An HIV/TB co-infected patient in Rwanda, before (*above*) and after (*below*) 6 months of treatment.

First, charging fees for AIDS prevention and care will pose insurmountable problems for people living in poverty, many of whom will always be unable to pay even modest amounts for services or medications. Like efforts to battle airborne TB, such services might best be seen as a public good for public health. Initially, this approach will require sustained donor contributions, but many African countries have set targets for increased national investments in health, a pledge that could render ambitious programs sustainable in the long run. Meanwhile, as local investments increase, the price of AIDS care is decreasing. The development of generic medications means that ART can now cost <$0.25 per day, and costs continue to decrease to affordable levels in developing countries.

Second, the effective scale-up of pilot projects will require strengthening and sometimes rebuilding health care systems, including those charged with delivering primary care. In the past, the lack of

health care infrastructure has been cited as a barrier to providing ART in the world's poorest regions; however, AIDS resources, which are at last considerable, may be marshaled to rebuild public health systems in sub-Saharan Africa and other HIV-burdened regions—precisely the settings in which TB is resurgent.

Third, a lack of trained health care personnel, most notably doctors, is invoked as a reason for the failure to treat AIDS in poor countries. The lack is real, and the brain drain (discussed below) continues. However, one reason doctors leave Africa is that they lack the tools to practice their trade there. AIDS funding provides an opportunity not only to recruit physicians and nurses to underserved regions but also to strengthen health systems by building infrastructure, providing diagnostic and therapeutic resources, and training community health workers to supervise care for AIDS and many other diseases within their communities. Such training should be undertaken even in places where physicians are abundant, since community-based, closely supervised care represents the highest standard of care for chronic disease, whether in the First World or the Third.

Fourth, extreme poverty makes it difficult for many patients to comply with therapy for chronic diseases, whether communicable or not. Indeed, poverty in its many dimensions is far and away the greatest barrier to the scale-up of treatment and prevention programs. It is possible to remove many of the social and economic barriers to adherence, but only with what sometimes are termed "wrap-around services": food supplements for the hungry, help with transportation to clinics, child care, and housing. In many rural regions of Africa, hunger is the major coexisting condition in patients with AIDS or TB, and those consumptive diseases cannot be treated effectively without adequate caloric intake.

Finally, there is a need for a renewed basic-science commitment to the discovery and development of vaccines; more reliable, less expensive diagnostic tools; and new classes of therapeutic agents. This need applies not only to the three leading infectious killers—against none of which is there an effective vaccine—but also to many other neglected diseases of poverty.

Malaria

Chapter 210 reviews the etiology, pathogenesis, and clinical treatment of malaria, the world's third-ranking infectious killer. Malaria's human cost is enormous, with the greatest toll among children, especially African children, living in poverty. An estimated 250 million people have malarial disease each year, and the disease annually kills 1 million people, mostly children under age 5. The poor disproportionately experience the consequences of malaria: 58% of malaria deaths occur in the poorest 20% of the world's population, and 90% are registered in sub-Saharan Africa. The differential magnitude of this mortality burden is greater than that associated with any other disease. Likewise, the morbidity differential is greater for malaria than for diseases caused by other pathogens, as documented in a study from Zambia that revealed a 40% greater prevalence of parasitemia among children under age 5 in the poorest quintile than in the richest.

Despite experiencing the greatest consequences of malaria, the poor are those least able to access effective prevention and treatment tools. Economists describe the complex interactions between malaria and poverty from an opposite but complementary perspective: they delineate ways in which malaria arrests economic development both for individuals and for whole nations. Microeconomic analyses focusing on direct and indirect costs estimate that malaria may consume up to 10% of a household's annual income. A Ghanaian study that categorized the population by income group highlighted the regressive nature of this cost: the burden of malaria represents only 1% of a wealthy family's income but 34% of a poor household's income.

At the national level, macroeconomic analyses estimate that malaria may reduce the per capita gross national product of a disease-endemic country by 50% relative to that of a nonmalarial country. The causes of this drag include high fertility rates, impaired cognitive development of children, decreased schooling, decreased saving, decreased foreign investment, and restriction of worker mobility. In light of this enormous cost, it is little wonder that an important review by the economists Sachs and Malaney concludes that "where malaria prospers most, human societies have prospered least."

ROLLING BACK MALARIA

In part because of differences in vector distribution and climate, resource-rich countries offer few blueprints for malaria control and treatment that are applicable in tropical (and resource-poor) settings. In 2001, African heads of state endorsed the WHO Roll Back Malaria (RBM) campaign, which prescribes strategies appropriate for sub-Saharan African countries. In 2008, the RBM partnership launched the Global Malaria Action Plan (GMAP). This global strategy has set out a coordinated approach to control and eliminate the disease and to ensure that gains in one nation are not lost because of failed control measures in neighboring countries. The GMAP recommends a number of key tools to reduce malaria-related morbidity and mortality rates: the use of insecticide-treated bed nets (ITNs), indoor residual spraying (IRS), and artemisinin-based combination therapy (ACT) as well as intermittent preventive treatment during pregnancy, prompt diagnosis, and other vector control measures such as larviciding and environmental management.

Insecticide-Treated Bed Nets

ITNs are an efficacious and cost-effective public health intervention. A meta-analysis of controlled trials indicates that malaria incidence is reduced by 50% among persons who sleep under ITNs compared with the incidence among those who do not use nets at all. Even untreated nets reduce malaria incidence by one-quarter. On an individual level, the utility of ITNs extends beyond protection from malaria. Several studies suggest that all-cause mortality is reduced among children under age 5 to a greater degree than can be attributed to the reduction in malarial disease alone. Morbidity (specifically that due to anemia), which predisposes children to diarrheal and respiratory illnesses and pregnant women to the delivery of low-birth-weight infants, is also reduced in populations using ITNs. In some areas, ITNs offer a supplemental benefit by preventing transmission of lymphatic filariasis, cutaneous leishmaniasis, Chagas' disease, and tick-borne relapsing fever. At the community level, investigators suggest that the use of an ITN in just one household may reduce the number of mosquito bites in households up to several hundred meters away. The cost of ITNs per DALY saved is estimated at $10–$38, which qualifies ITNs as a "very efficient use of resources and [a] good candidate for public subsidy."

The WHO recommends that all individuals living in malaria-endemic areas sleep under protective ITNs. About 140 million long-lasting ITNs were distributed in high-burden African countries in 2006–2008, and rates of household ownership of ITNs in high-burden countries increased to 31%. Although the RBM partnership has seen modest success, the WHO's 2009 World Malaria Report states that the percentage of children <5 years of age using an ITN (24%) remains well below the World Health Assembly's target of 80%. Limited success in scaling up ITN coverage reflects the inadequately acknowledged economic barriers that prevent the destitute sick from accessing critical preventive technologies and the challenges in designing and implementing effective delivery platforms for these products.

Indoor Residual Spraying

IRS is one of the most common interventions for preventing the transmission of malaria in endemic areas. Vector control using insecticides approved by the WHO, including DDT, can effectively reduce or even interrupt malaria transmission. However, studies have indicated that spraying is effective in controlling malaria transmission only if most (~80%) of the structures in the targeted community are treated. Moreover, since a successful program is dependent on well-trained spraying teams as well as on effective monitoring and planning, IRS is difficult to employ and is often reliant on health systems with a strong infrastructure that renders the approach feasible. Regardless of the limitations of IRS, the WHO recommends its use in combination with ITNs. Neither intervention alone is sufficient to prevent transmission of malaria entirely.

Artemisinin-Based Combination Therapy

The emergence and spread of chloroquine resistance have increased the necessity for antimalarial combination therapy. To limit the spread of resistance, the WHO now recommends only the use of ACT for uncomplicated falciparum malaria. Like that of other antimalarial interventions, the use of ACT has increased in the last few years, but coverage remains very low in several countries in sub-Saharan Africa. In a 2007–2008 study, fewer than 15% of children under the age of 5 with a fever were receiving ACT in 11 of 13 countries surveyed, although the World Health Assembly's target is 80%. In response, the RBM partnership has focused significant investment on enhancing access to ACT by facilitating its delivery through the public-health sector and developing innovative funding mechanisms (e.g., the Affordable Medicines Facility—malaria) through which consumer prices for ACT can be reduced significantly and ineffective artemisinin monotherapies can be eliminated from the market.

In the last several years, resistance to antimalarial medicines and insecticides has become an even larger problem. In 2009, confirmation of artemisinin resistance was reported. Although the WHO has called for an end to the use of artemisinin monotherapy, the marketing of such therapies continues in many countries. Ongoing use of artemisinin monotherapy increases the likelihood of drug resistance, a deadly prospect that will make malaria far more difficult to treat.

Meeting the challenge of malaria control will continue to require careful study of appropriate preventive and therapeutic strategies in the context of an increasingly sophisticated molecular understanding of the pathogen, vector, and host. However, an appreciation of the economic and structural devastation wrought by malaria—like that inflicted by diarrhea, AIDS, and TB—on the most vulnerable populations should heighten the commitment to the critical analysis of ways to implement proven strategies for the prevention and treatment of these diseases.

Noncommunicable Chronic Diseases

Although the burden of communicable diseases—especially HIV infection, tuberculosis, and malaria—still accounts for the majority of deaths in resource-poor regions such as sub-Saharan Africa, 60% of all deaths worldwide in 2004 were due to noncommunicable chronic diseases (NCDs). Moreover, 80% of deaths attributable to NCDs occurred in low- and middle-income countries, where 86% of the global population lives. In 2005, 8.5 million people in the world died of an NCD before their 60th birthday, a figure exceeding the total number of deaths due to AIDS, TB, and malaria combined. By 2020, NCDs will account for 80% of the GBD and for 7 out of every 10 deaths in developing countries. The recent

rise in resources for and attention to communicable diseases is both welcome and long overdue, but developing countries already are carrying a "double burden" of communicable and noncommunicable diseases.

CARDIOVASCULAR DISEASE

Unlike TB, HIV infection, and malaria—diseases caused by single pathogens that damage multiple organs—cardiovascular diseases reflect injury to a single organ system downstream of a variety of insults. The burden of chronic cardiovascular disease in low-income countries represents one consequence of decades of health system neglect; furthermore, cardiovascular research and investment have long focused on the ischemic conditions that are increasingly common in high- and middle-income countries. Meanwhile, despite awareness of its health impact during the early twentieth century, cardiovascular damage in response to infection and malnutrition has fallen out of view until recently.

The perception of cardiovascular diseases as a problem of elderly populations in middle- and high-income countries has contributed to their neglect by global health institutions. Even in Eastern Europe and Central Asia, where the collapse of the Soviet Union was followed by a catastrophic surge in cardiovascular disease deaths (mortality rates from ischemic heart disease nearly doubled between 1991 and 1994 in Russia, for example), the modest flow of overseas development assistance to the health sector focused on the communicable causes that accounted for <1 in 20 excess deaths during that period.

Predictions of an imminent rise in the share of deaths and disabilities due to NCDs in developing countries have led to calls for preventive policies to restrict tobacco use, improve diet, and increase exercise along with the prescription of multidrug regimens for persons with high levels of vascular risk. Although this agenda could do much to prevent pandemic NCD, it will do little to help those with established heart disease stemming from nonatherogenic pathologies.

The epidemiology of heart failure reflects inequalities in risk factorprevalence and treatment. Heart failure as a consequence of pericardial, myocardial, endocardial, or valvular injury accounts for as many as 1 in 10 admissions to hospitals around the world. Countries have reported a remarkably similar burden of this condition at the health system level since the 1950s, but the causes of heart failure and the age of the people affected vary with resources and ecology. In populations with a high human-development index, coronary artery disease and hypertension among the elderly account for most cases of heart failure. Among the world's poorest billion people, however, heart failure reflects poverty-driven exposure of children and young adults to rheumatogenic strains of streptococci and cardiotropic microorganisms (e.g., HIV, *Trypanosoma cruzi*, enteroviruses, *M. tuberculosis*), untreated high blood pressure, and nutrient deficiencies. The mechanisms of other causes of heart failure common in these populations—such as idiopathic dilated cardiomyopathy, peripartum cardiomyopathy, and endomyocardial fibrosis—remain unclear.

Among the 2.4 million annual cases of pediatric rheumatic heart disease, more than 40% occur in sub-Saharan Africa. This disease leads to more than 33,000 cases of endocarditis, 252,000 strokes, and 680,000 deaths per year—almost all in developing countries. Researchers in Ethiopia have reported annual death rates as high as 12.5% in rural areas. In part because the prevention of rheumatic heart disease has not advanced since the disappearance of this disease in wealthy countries, no part of sub-Saharan Africa has eradicated rheumatic heart disease despite examples of success in Costa Rica, Cuba, and some Caribbean nations.

Strategies to eliminate rheumatic heart disease may depend on active case finding confirmed by echocardiography among high-risk groups as well as efforts to extend access to surgical interventions among children with advanced valvular damage. Partnerships between established surgical programs and areas with limited or nonexistent facilities may help develop capacity and provide care to patients who otherwise would have an early and painful death. A long-term goal is the establishment of regional centers of excellence equipped to provide consistent, accessible, high-quality services.

In stark contrast to the extraordinary lengths to which patients in wealthy countries will go to treat ischemic cardiomyopathy, young patients with nonischemic cardiomyopathies in resource-poor settings have received little attention. These conditions account for as many as 25–30% of admissions for heart failure in sub-Saharan Africa and include poorly understood entities such as peripartum cardiomyopathy (which has an incidence in rural Haiti of 1 per 300 live births) and HIV cardiomyopathy. Multidrug regimens that include beta blockers, angiotensin-converting enzyme (ACE) inhibitors, and other neurohormonal antagonists can dramatically reduce mortality risk and improve quality of life for these patients. Lessons learned in the scale-up of chronic care for HIV infection and TB may be illustrative as progress is made in establishing means to deliver heart failure therapies.

Because systemic investigation of the causes of stroke and heart failure in sub-Saharan Africa has begun only recently, little is known about the impact of elevated blood pressure in this portion of the continent. Modestly elevated blood pressure in the absence of tobacco use in populations with low rates of obesity may confer little risk of adverse events in the short run. In contrast, persistently elevated blood pressure above 180/110 goes largely undetected, untreated, and uncontrolled in this

setting. In the Framingham cohort of men 45–74 years old, the prevalence of blood pressures above 210/120 declined from 1.8% in the 1950s to 0.1% in the 1990s with the introduction of effective antihypertensive agents. Although debate continues about appropriate screening strategies and treatment thresholds, rural health centers staffed by nonphysicians must quickly gain access to essential antihypertensive medications.

In 1960, Paul Dudley White and colleagues reported on the prevalence of cardiovascular disease in the region near the Albert Schweitzer Hospital in Lambaréné, Gabon. Although the group found little evidence of myocardial infarction, they concluded that "*the high prevalence of mitral stenosis* [sic] is astonishing. . . . We believe strongly that it is a duty to help bring to these sufferers the benefits of better penicillin prophylaxis and of cardiac surgery when indicated. The same responsibility exists for those with correctable congenital cardiovascular defects." Leaders from tertiary centers in sub-Saharan Africa and elsewhere have continued to call for prevention and treatment of the cardiovascular conditions of the poor. The reconstruction of health services in response to pandemic infectious disease offers an opportunity to identify and treat patients with organ damage and to undertake the prevention of cardiovascular and other chronic conditions of poverty.

CANCER

Low- and middle-income countries accounted for 54% and 60%, respectively, of the 12.4 million cases and 7.6 million deaths due to cancer in 2008. By 2020, the total number of new cancer cases will rise by 29% in developed countries and by 73% in developing countries. Also by 2020, overall mortality from cancer will increase by 104%, and the increase will be fivefold higher in developing than in developed countries. "Western" lifestyle changes will be responsible for the increased incidence of cancers of the breast, colon, and prostate, but historic realities, sociocultural and behavioral factors, genetics, and poverty itself also will have a profound impact on cancer-related mortality and morbidity rates. Whereas infectious causes are responsible for <10% of cancers in developed countries, they account for 25% of all malignancies in low- and middle-income countries. Infectious causes of cancer such as human papillomavirus (cervical cancer), hepatitis B virus (liver cancer), and *Helicobacter pylori* (stomach cancer) will continue to have a much larger impact in developing countries. Environmental and dietary factors such as indoor air pollution and high-salt diets also help account for increased rates of certain cancers (e.g., lung and stomach cancers). Tobacco use (both smoking and chewing) is the most important source of increased mortality from lung and oral cancers. In contrast to decreasing tobacco use in many developed countries, the number of smokers is growing in developing countries, especially among women and young people.

For many reasons, outcomes of malignancies are far worse in developing countries than in developed nations. Overstretched health systems in poor countries simply are not capable of early detection; 80% of patients already have incurable malignancies at diagnosis. Treatment of cancers is available for only a very small number of mostly wealthy citizens in the majority of poor countries, and even when treatment is available, the range and quality of services are often substandard. Yet this need not be the future. Only a decade ago, MDR-TB and HIV infection were considered untreatable for all but the wealthiest among the world's population. The last decade has made clear the feasibility of creating innovative programs that reduce technical and financial barriers to the provision of care for complex diseases for the world's poorest populations.

DIABETES

The International Diabetes Federation reports that the number of diabetic patients in the world is expected to increase from 285 million in 2010 to 438 million by 2030. Already, more than 70% of diabetic patients live in developing countries where, because those affected are far more frequently under age 65, the complications of micro- and macrovascular disease take a far greater toll. In 2009, the Federation projected an estimated 4 million deaths from diabetes-related illnesses in 2010, with almost 80% of those deaths occurring in low- and middle-income countries.

OBESITY AND TOBACCO USE

In 2004, the WHO released its Global Strategy on Diet, Physical Activity and Health, which focused on the populationwide promotion of healthy diet and regular physical activity in an effort to reduce the growing global problem of overweight and obesity. Passing this strategy at the World Health Assembly proved difficult because of strong opposition from the food industry and from a number of WHO member states, including the United States. Although globalization has had many positive effects, one negative aspect has been the growth in both developed and developing countries of well-financed lobbies that have aggressively promoted unhealthy dietary changes and increased consumption of alcohol and tobacco. Foreign direct investment in tobacco, beverage, and food products in developing countries reached $327 million in 2002—a figure nearly five times greater than the amount spent during that year to address NCDs by bilateral funding agencies, the WHO, and the World Bank combined.

THE THREE PILLARS OF PREVENTION

The WHO estimates that 80% of all cases of cardiovascular disease and type 2 diabetes as well as 40% of all cancers can be prevented through the three pillars of healthy diet, physical activity, and

avoidance of tobacco. Although there is some evidence that population-based measures can have some impact on these behaviors, it is sobering to note that increasing obesity levels have not been reversed in any population, including the populations of high-income countries with robust diet industries. Nonetheless, in Mauritius, for example, a single policy measure that changed the type of cooking oil available to the population led to a fall in mean serum cholesterol levels. Tobacco avoidance may be the most important and most difficult behavioral modification of all. In the twentieth century, 100 million people worldwide died of tobacco-related diseases; it is projected that >1 billion people will die of these diseases in the twenty-first century, with the vast majority of those deaths in developing countries. Today, 80% of the world's 1.2 billion smokers live in low- and middle-income countries, and although tobacco consumption is falling in most developed countries, it continues to rise at a rate of ~3.4% per year in developing countries. The WHO's 2003 Framework Convention on Tobacco Control represented a major advance, committing all of its signatories to a set of policy measures that have been shown to reduce tobacco consumption. However, most developing countries have continued to take a passive approach to the control of smoking.

Environmental Health

In a recent publication that examined how specific diseases and injuries are affected by environmental risk, the WHO determined that ~24% of the total GBD, one-third of the GBD among children, and 23% of all deaths are due to modifiable environmental factors. Many of these factors lead to deaths from infectious diseases; others lead to deaths from malignancies. Increasingly, etiology and nosology are difficult to parse. As much as 94% of diarrheal disease, which is linked to unsafe drinking water and poor sanitation, can be attributed to environmental factors. Risk factors such as indoor air pollution due to use of solid fuels, exposure to secondhand tobacco smoke, and outdoor air pollution account for 20% of lower respiratory infections in developed countries and as many as 42% of such infections in developing countries. Various forms of unintentional injury and malaria top the list of health problems to which environmental factors contribute. Some 4 million children die every year from causes related to unhealthy environments, and the number of infant deaths due to environmental factors in developing countries is 12 times that in developed countries.

Mental Health

The WHO reports that some 450 million people worldwide are affected by mental, neurologic, or behavioral problems at any given time and that ~873,000 people die by suicide every year. Major depression is the leading cause of years lost to disability in the world today. One in four patients visiting a health service has at least one mental, neurologic, or behavioral disorder, but most of these disorders are neither diagnosed nor treated. Most low- and middle-income countries devote <1% of their health expenditures to mental health.

Increasingly effective therapies exist for many of the major causes of mental disorders. Effective treatments for many neurologic diseases, including seizure disorders, have long been available. One of the greatest barriers to delivery of such therapies is the paucity of skilled personnel. Most sub-Saharan African countries have only a handful of psychiatrists, for example; most practice in cities and are unavailable within the public sector or to patients living in poverty. Among the few patients who are fortunate enough to see a psychiatrist or neurologist, fewer still are able to adhere to treatment regimens: several surveys of already diagnosed patients ostensibly receiving daily therapy have revealed that among the poor, few can take their medications as prescribed. The same barriers that prevent the poor from having reliable access to insulin or ART prevent them from benefiting from antidepressant, antipsychotic, and antiepileptic agents. To alleviate this problem, some authorities are proposing the training of health workers to provide community-based adherence support, counseling services, and referrals for patients in need of mental health services.

World Mental Health: Problems and Priorities in Low-Income Countries offers a comprehensive analysis of the burden of mental, behavioral, and social problems in low-income countries and relates the mental health consequences of social forces such as violence, dislocation, poverty, and disenfranchisement of women to current economic, political, and environmental concerns.

Primary Care

At the International Conference on Primary Health Care in Alma-Ata (in what is now Kazakhstan) in 1978, public health officials from around the world agreed on a commitment to "Health for All by 2000," a goal to be achieved by providing universal access to primary health care worldwide. Critics argued that the attainment of this goal by the proposed date was impossible. In the ensuing years, a strategy of selective primary health care emerged that included four inexpensive interventions collectively known as GOBI: *g*rowth monitoring, *o*ral rehydation, *b*reast-feeding, and *i*mmunizations for diphtheria, whooping cough, tetanus, polio, TB, and measles. GOBI later was expanded to GOBI-FFF, which also included *f*emale education, *f*ood, and *f*amily planning. Some public health figures saw this as an interim strategy to achieve "health for all," but others criticized it as a retreat from the commitments of Alma-Ata. Similar debates still rage, with "vertical" disease-specific programs for HIV, TB, and malaria often seen as competing with primary health care efforts for critical economic, human, and political resources. Global primary care is examined in detail in Chap. e1.

HEALTH SYSTEMS AND THE BRAIN DRAIN

A significant and frequently invoked barrier to effective health care in resource-poor settings is the lack of medical personnel. In what is termed the *brain drain*, many physicians and nurses emigrate from their home countries to pursue opportunities abroad, leaving behind health systems that are understaffed and ill equipped to deal with the epidemic diseases that ravage local populations. The WHO recommends a minimum of 20 physicians and 100 nurses per 100,000 persons, but recent reports from that organization and others confirm that many countries, especially in sub-Saharan Africa, fall far short of those target numbers. More than one-half of those countries register fewer than 10 physicians per 100,000 population. In contrast, the United States and Cuba register 279 and 596 doctors per 100,000 population, respectively. Similarly, the majority of sub-Saharan African countries do not have even half of the WHO-recommended minimum number of nurses. In addition to these appalling national aggregates, further inequalities in health care staffing exist *within* countries. Rural-urban disparities in health care personnel mirror disparities of both wealth and health. For instance, nearly 90% of Malawi's population is rural, but more than 95% of clinical officers work at urban facilities, and 47% of nurses work at tertiary care facilities. Even community health workers trained to provide first-line services to rural populations often transfer to urban districts. In addition to inter- and intranational transfer of personnel, the AIDS epidemic contributes to personnel shortages across Africa. Although data on the prevalence of HIV infection among health professionals are scarce, the available numbers suggest substantial and adverse impacts on an already overburdened health sector. A study that examined the fates of a small cohort of Ugandan physicians found that at least 22 of the 77 doctors who graduated from Makerere University Medical School in 1984 had died by 2004—most, presumably, of AIDS. The shortage of medical personnel in the areas hardest hit by HIV has profound implications for prevention and treatment efforts in those regions. The cycle of health-sector impoverishment, brain drain, and lack of personnel to fill positions when they are available conspires against ambitious programs to bring ART to persons living with both AIDS and poverty. Furthermore, the education of medical trainees is jeopardized as the ranks of the health and academic communities continue to shrink as a result of migration or disease. The long-term implications are sobering.

A proper biosocial analysis of the brain drain confirms that the flight of health personnel—almost always, as most reviews suggest, from poor to less poor regions—is not simply a question of desire for more equitable remuneration. Epidemiologic trends and access to the tools of the trade are also relevant, as are working conditions in general. In many settings now losing skilled health personnel, the advent of HIV has led to a sharp rise in TB incidence; in the eyes of health care providers, other opportunistic infections also have become insuperable challenges. Together, these forces have conspired to render the provision of proper care almost impossible. One Kenyan medical resident noted, "Before training we thought of doctors as supermen. . . . [Now] we are only mortuary attendants."

In light of the difficult conditions under which these health care personnel work, is it any surprise that the U.S. government's Global AIDS Coordinator remarked in 2004 that there were more Ethiopian physicians practicing in Chicago than in all of Ethiopia? When providing care for the sick becomes a nightmare for those at the beginning of clinical training, physician burnout soon follows among those who carry on in settings of impoverishment. In the public-sector institutions put in place to care for the poorest people, the confluence of epidemic disease, lack of resources with which to respond, and unrealistically high user fees has led to widespread burnout among health workers. Patients and their families are those who pay most dearly for provider burnout, just as they bear the burden of disease and—with the introduction of user fees—much of the cost of responding, however inadequately, to new epidemics and persistent plagues.

CONCLUSION: TOWARD A SCIENCE OF IMPLEMENTATION

Public-health strategies draw largely on quantitative methods—epidemiology, biostatistics, and economics. Clinical practice, including internal medicine, draws on a rapidly expanding knowledge base but remains focused on individual patient care; clinical interventions are rarely population-based. In fact, neither public health nor clinical approaches alone will prove adequate in addressing the problems of global health. There is a long way to go before evidence-based internal medicine is applied effectively among the world's poor. Complex infectious diseases such as AIDS and TB have proved difficult but not impossible to manage; drug resistance and a lack of effective health systems have complicated such work. Beyond communicable disease, in the arena of chronic diseases (e.g., cardiovascular disease), global health is a nascent endeavor. Efforts to address any one of these problems in settings of great scarcity need to be integrated into broader efforts to strengthen failing health systems and alleviate the growing personnel crisis within these systems.

For these reasons, scholarly work and practice in the field once known as international health and now often designated *global health equity* are changing rapidly. That work is still informed by the tension between clinical practice and population-based interventions, between analysis and action. Once metrics are refined, how might they inform efforts to lessen premature morbidity and mortality rates among the world's poor? As in the nineteenth century, human rights perspectives have proved helpful in turning attention to the problems of the destitute sick; such perspectives also may inform strategies

of delivering care equitably. A number of university hospitals are developing training programs for physicians with an interest in global health. In medical schools across the United States and in other wealthy countries, interest in global health has been exploding. An informal survey at Harvard Medical School in 2006 revealed that nearly one-quarter of the 160 entering students either had significant global health experience or were planning a career in global health. A similar sea change among trainees has been reported at other medical schools. Half a century or even a decade ago, such high levels of interest would have been unimaginable.

Persistent epidemics, improved metrics, and growing interest have only recently been matched by an unprecedented investment in addressing the health problems of poor people in the developing world. This is a moment of opportunity. To ensure that the opportunity is not wasted, the facts need to be laid out for specialists and laypeople alike. More than 12 million people die each year simply because they live in poverty. An absolute majority of these premature deaths occur in Africa, with the poorer regions of Asia not far behind. Most of these deaths occur because the world's poorest do not have access to the fruits of science. They include deaths from vaccine-preventable illness, deaths during childbirth, deaths from infectious diseases that might be cured with access to antibiotics and other essential medicines, deaths from malaria that would have been prevented by bed nets and access to therapy, and deaths from waterborne illnesses. Other excess mortality is attributable to the inadequacy of efforts to develop new preventive, diagnostic, and therapeutic tools. Those funding the discovery and development of new tools typically neglect the concurrent need for strategies to make them available to the poor. Indeed, some would argue that the biggest challenge facing those who seek to address this outcome gap is the lack of practical means of distribution in the most heavily affected regions.

The development of tools must be followed quickly by their equitable distribution. When new preventive and therapeutic tools are developed without concurrent attention to delivery or implementation, one encounters what sometimes are termed *perverse effects:* even as new tools are developed, inequalities of outcome—lower morbidity and mortality rates among those who can afford access, with sustained high morbidity and mortality among those who cannot—will grow in the absence of an equity plan to deliver the tools to those most at risk. Preventing such a future is the most important goal of global health.

FURTHER READINGS

Cohen J: The new world of global health. Science 311:162, 2006[PMID: 16410496]

Desjarlais R et al (eds): *World Mental Health: Problems and Priorities in Low-Income Countries*. New York, Oxford University Press, 1995

Farmer PE: *Infections and Inequalities: The Modern Plagues*, 2nd ed. Berkeley, University of California Press, 2001

Fauci AS et al: Emerging infectious diseases: A 10-year perspective from the National Institute of Allergy and Infectious Diseases. Emerg Infect Dis 11:519, 2005[PMID: 15829188]

Frenk J: Reinventing primary health care: The need for systems integration. Lancet 374:170, 2009[PMID: 19439350]

Gaziano TA et al: Scaling up interventions for chronic disease prevention: The evidence. Lancet 370:1939, 2007[PMID: 18063028]

Jamison DT et al (eds): *Disease Control Priorities in Developing Countries*, 2nd ed. Washington, DC, Oxford University Press and The World Bank, 2006

Maximizing Positive Synergies Collaborative Group: An assessment of interactions between global health initiatives and country health systems. Lancet 373:2137, 2009

Sachs J, Malaney O: The economic and social burden of malaria. Nature 415:680, 2002[PMID: 11832956]

World Health Organization: The Global Burden of Disease: 2004 Update. Geneva, World Health Organization, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DECISION-MAKING IN CLINICAL MEDICINE: INTRODUCTION

To a medical student who requires 2 hours to collect a patient's history and perform a physical examination and several additional hours to organize that information into a coherent presentation, an experienced clinician's ability to decide on a diagnosis and management plan in a fraction of the time seems extraordinary. What separates the experienced clinician's performance from that of the novice is an elusive quality called "expertise." The first part of this chapter will provide a brief introduction to what is known about the development of expertise in clinical reasoning.

Equally bewildering to the student are the proper use of diagnostic tests and the integration of the results into the patient's clinical assessment. A novice medical practitioner typically uses a "shotgun" approach to testing, hoping to hit a target without knowing exactly what that target is. The expert, in contrast, usually has a specific target in mind and adjusts the testing strategy to it. The second part of the chapter will review briefly some of the crucial basic statistical concepts useful in the interpretation of diagnostic tests. Quantitative tools available to assist in clinical decision-making also will be discussed.

Evidence-based medicine (EBM) is the term used to describe the integration of the best available research evidence with clinical judgment and experience as applied to the care of individual patients. The third part of the chapter will provide a brief overview of some of the tools of EBM.

Brief Introduction to Clinical Reasoning

CLINICAL EXPERTISE

It is surprisingly difficult to define clearly what is meant by "clinical expertise." Chess has its masters, music its virtuoso performers, and athletics its Olympians. But in medicine, once training is complete and the boards are passed, there are no further tests or benchmarks of performance or ability that can be used to identify those who have attained the highest level of abilities in their clinical roles. Of course, there are always a few clinicians who are believed by their colleagues to have special problem-solving abilities: the "elite" who are consulted when particularly difficult or obscure cases have baffled everyone else. But for all their expertise, such doctors typically cannot explain what processes and methods they use to achieve their impressive results. Furthermore, it is not clear that their diagnostic virtuosity can be generalized. In other words, an expert on hypertrophic cardiomyopathy may be no better (and possibly worse) than a first-year resident at diagnosing and managing a patient with neutropenia, fever, and hypotension.

Broadly construed, clinical expertise includes not only cognitive dimensions and the integration of verbal and visual cues or information but also complex motor skills that are required in the performance of various invasive and noninvasive procedures and tests. In addition, the ability to communicate effectively with patients and work effectively with members of the medical team could be included as important aspects of "the complete package" of expertise in medicine. In this chapter, however, the focus will be on the cognitive elements (clinical reasoning), particularly as they relate to diagnosis. This focus is driven by two factors. First, the most important "actions" in clinical medicine are not procedures or prescriptions but the judgments (both diagnoses and treatment choices) from which all other aspects of medical care flow. Second, although the research on medical expertise is relatively sparse overall, it is best developed in the area of diagnostic decision-making. Much less work has been done on treatment decisions or the technical skills involved in the performance of procedures.

The obvious difficulty involved in the study of clinical reasoning is that it takes place in the heads of doctors and is therefore not readily observable. Further, doctors may not even be aware of how they reason in many cases, and so they may be unable to describe the processes they use. To overcome this difficulty, one line of research has focused on how doctors *should* reason diagnostically rather than on how they actually *do* reason. In addition, because of the difficulties of empirical research in this area, much of what is known about clinical reasoning comes from empirical studies of nonmedical problem-solving behavior. The field has been influenced by important work from cognitive psychology, sociology, medical education, economics, informatics, and decision sciences. However, because of this diversity of perspectives, no single integrated model of clinical reasoning exists, and not infrequently, different terms and models are proposed for similar phenomena.

INTUITIVE VERSUS ANALYTIC REASONING

One useful contemporary model of reasoning (dual-process theory) distinguishes two general systems

of cognitive processes. *Intuition* (System 1) provides rapid effortless judgments from memorized associations—for example, African-American women and hilar adenopathy equals sarcoid—or from the reduction of complex data by means of pattern recognition and other heuristics. Typically, the clinician is unable to say how those judgments were formulated. *Analysis* (System 2), the other form of reasoning in the dual-process model, is slow, methodical, and effortful. These are, of course, idealized extremes of the cognitive continuum. The way these systems interact in different decision problems and the way they differ between experts and novices remain the subject of considerable debate. Much work has also been done to identify how each of these systems can lead to errors in judgment.

Pattern recognition is a complex cognitive process that appears largely intuitive. One can recognize people's faces, the breed of a dog, or the model of an automobile without necessarily being able to say what specific features prompted the recognition. Analogously, an experienced clinician often can recognize the pattern of a diagnosis she or he is very familiar with after a very short amount of time with the patient. The student, who does not have that stored repertoire of diagnostic patterns, must use a more laborious analytic approach along with much more intensive data collection to reach the diagnosis.

The following three brief scenarios of a patient with hemoptysis present three distinct patterns:

- A 46-year-old man presents to his internist with a chief complaint of hemoptysis. He is otherwise healthy, is a nonsmoker, and is recovering from an apparent viral bronchitis. For this patient, the pattern would suggest that the acute bronchitis is responsible for the small amount of blood-streaked sputum the patient has observed. In this case, a chest x-ray may provide sufficient reassurance that a more serious disorder is not present.
- In the second scenario, a 46-year-old patient with the same chief complaint who has a 100-pack-year smoking history, a productive morning cough, and episodes of blood-streaked sputum fits the pattern of carcinoma of the lung. Consequently, along with the chest x-ray, the physician obtains a sputum cytology examination and refers this patient for a chest CT scan.
- In the third scenario, a 46-year-old patient with hemoptysis who is from a developing country is evaluated with an echocardiogram as well, because the physician thinks she hears a soft diastolic rumbling murmur at the apex on cardiac auscultation, suggesting rheumatic mitral stenosis.

The primary mistake that can result from relying on the free use of pattern recognition in diagnosis is *premature closure*: concluding that one already knows the correct diagnosis and therefore failing to complete the data collection that would demonstrate the lack of fit of the initial pattern selected. Consider the following hypothetical example. A 45-year-old male patient with a 3-week history of a "flulike" upper respiratory infection (URI) presented to his physician with symptoms of dyspnea and a productive cough. On the basis of the presenting complaint, the clinician pulled out a "URI assessment form" to obtain patient information that could be beneficial in improving the quality and efficiency of care. The physician quickly completed the examination components outlined on this structured form, noting in particular the absence of fever and a clear chest examination. He then prescribed an antibiotic for presumed bronchitis, showed the patient how to breathe into a paper bag to relieve his "hyperventilation," and sent him home with the reassurance that his illness was not serious. After a sleepless night with significant dyspnea unrelieved by breathing into a bag, the patient developed nausea and vomiting and collapsed. He was brought into the emergency department in cardiac arrest and could not be resuscitated. Autopsy showed a posterior wall myocardial infarction (MI) and a fresh thrombus in an atherosclerotic right coronary artery. What went wrong? The clinician decided, on the basis of the patient's appearance, even before starting the history, that the patient's complaints were not serious. He therefore felt confident that he could perform an abbreviated and focused examination by using the URI assessment protocol rather than considering the broader range of possibilities and performing appropriate tests to confirm or refute his initial hypotheses. In particular, by concentrating on the URI, the clinician failed to elicit the full dyspnea history, which would have suggested a far more serious disorder, and he neglected to search for other symptoms that could have directed him to the correct diagnosis.

Cognitive shortcuts or rules of thumb, sometimes referred to as *heuristics*, are another type of intuitive mental process that can be invoked to understand how experts solve complex problems of the sort encountered daily in clinical medicine with great efficiency. The original work on the use of heuristics in problem solving was done largely in laboratory experiments on psychology undergraduates. The objective of the research program was to test the statistical intuition of those subjects against the rules of statistics to understand how such intuitions could be biased. Hence, discussions of the use of heuristics in decision-making tend to focus more on ways in which their use can lead to errors in judgment than on their successes. Although there are many heuristics of possible relevance to clinical reasoning, only four will be mentioned here.

When assessing a particular patient, clinicians often weigh the probability that the patient's clinical features match those of the class of patients with the leading diagnostic hypotheses being considered. In other words, the clinician is searching for the diagnosis for which the patient appears to be a representative example; this cognitive shortcut is called the *representativeness heuristic*. This heuristic is analogous to pattern recognition. However, if there are two (or more) competing diagnoses

that could explain the patient's symptoms, physicians who use the representativeness heuristic can reach erroneous conclusions if they fail to consider the underlying prevalence of the two competing diagnoses (i.e., the prior, or pretest, probabilities). Consider a patient with pleuritic chest pain, dyspnea, and a low-grade fever. A clinician might consider acute pneumonia and acute pulmonary embolism to be the two leading diagnostic alternatives. Using the representativeness heuristic, the clinician might judge both diagnostic candidates to be equally likely, although doing so would be wrong if pneumonia was much more prevalent in the underlying population. Mistakes also may result from a failure to consider that a pattern based on experience with a small number of prior cases probably will be less reliable than one based on greater experience.

A second commonly used cognitive shortcut, the *availability heuristic*, involves judgments made on the basis of how easily prior similar cases or outcomes can be brought to mind. For example, an experienced clinician may recall 20 elderly patients seen over the last few years who presented with painless dyspnea of acute onset and were found to have acute MI. A novice clinician may spend valuable time seeking a pulmonary cause for the symptoms before considering and then confirming the cardiac diagnosis. In this situation, the patient's clinical pattern does not fit the expected pattern of acute MI, but experience with this atypical presentation, along with the ability to recall it, can help direct the physician to the diagnosis.

Errors with the availability heuristic can come from several sources of recall bias. For example, rare catastrophes are likely to be remembered with a clarity and force out of proportion to their value—for example, a patient with a sore throat eventually found to have leukemia or a young athlete with leg pain eventually found to have a sarcoma—and recent experience is, of course, easier to recall and therefore more influential on clinical judgments.

The third commonly used cognitive shortcut, the *anchoring heuristic*, involves estimating a probability by starting from a familiar point (the anchor) and adjusting to the new case from that perspective. Anchoring can be a powerful tool for diagnosis but may be used incorrectly. For example, a clinician may judge the probability of coronary artery disease (CAD) to be very high after a positive exercise thallium test because the prediction has been anchored to the test result ("positive test = high probability of CAD"). Yet, as discussed below, this prediction would be inaccurate if the clinical (pretest) picture of the patient being tested indicated a low probability of disease (e.g., a 30-year-old woman with no risk factors). As illustrated in this example, anchors are not necessarily the same as the pretest probability (see "Measures of Disease Probability and Bayes' Theorem," below).

The fourth heuristic, which might be termed the *simplicity heuristic,* states that clinicians should use the simplest explanation possible that will account adequately for the patient's symptoms or findings (Occam's razor). Although this is an attractive and often useful principle, it is important to remember that there is no biologic basis for it.

Experienced physicians use analytic reasoning processes (System 2) much more often when the problem they face is recognized to be complex or to involve important unfamiliar elements or features. In such situations, the clinician proceeds much more methodically in what has been referred to as the hypothetico-deductive model of reasoning. From the outset, the expert clinician is generating, refining, and discarding diagnostic hypotheses. The questions she asks in the history are driven by the hypotheses she is working with at the moment. Even the physical examination is focused on specific questions. Is the spleen enlarged? How big is the liver? Is it tender? Does it have any palpable masses or nodules? Each question focuses the attention of the examiner on the exclusion of all other inputs until it is answered, allowing the examiner to move on to the next specific question. Each diagnostic hypothesis sets a context for the diagnostic steps to follow and provides testable predictions. For example, if the enlarged and quite tender liver felt on physical examination is due to acute hepatitis (the hypothesis), certain specific liver function tests should be markedly elevated (the prediction). If the tests come back normal, the hypothesis may have to be discarded or substantially modified.

Negative findings often are neglected but are as important as positive ones in establishing and refining diagnostic hypotheses. Chest discomfort that is not provoked or worsened by exertion in an active patient reduces the likelihood that chronic ischemic heart disease is the underlying cause. The absence of a resting tachycardia and thyroid gland enlargement reduces the likelihood of hyperthyroidism in a patient with paroxysmal atrial fibrillation.

The acuity of a patient's illness can play an important role in overriding considerations of prevalence and the other issues described above. For example, clinicians are taught to consider aortic dissection routinely as a possible cause of acute severe chest discomfort along with MI, even though the typical history of dissection is different from that of MI and dissection is far less prevalent (Chap. 248). This recommendation is based on the recognition that a relatively rare but catastrophic diagnosis such as aortic dissection is very difficult to make unless it is explicitly considered as a diagnostic imperative. If the clinician fails to elicit any of the characteristic features of dissection by history and finds equivalent blood pressures in both arms and no pulse deficits, he may feel comfortable discarding the aortic dissection hypothesis. If, however, the chest x-ray shows a possible widened mediastinum, the hypothesis may be reinstated and a diagnostic test ordered [e.g., thoracic computed tomography (CT) scan, transesophageal echocardiogram] to evaluate it more fully. In nonacute situations, the

prevalence of potential alternative diagnoses should play a much more prominent role in diagnostic hypothesis generation.

Cognitive scientists studying the thought processes of expert clinicians have observed that clinicians group data into packets, or "chunks," that are stored in their memories and manipulated to generate diagnostic hypotheses. Because short-term memory typically can hold only 7–10 items at a time, the number of packets that can be actively integrated into hypothesis-generating activities is similarly limited. For this reason, the cognitive shortcuts discussed above can play a key role in the generation of diagnostic hypotheses, many of which are discarded as rapidly as they are formed (thus demonstrating that the distinction between analytic and intuitive reasoning is arbitrary and simplistic but useful nonetheless).

Research into the hypothetico-deductive model of reasoning has had surprising difficulty identifying the elements that distinguish experts from novices. This has led to a shift from examining the problem-solving process of experts to analyzing the organization of their knowledge. For example, diagnosis may be based on the resemblance of a new case to prior individual instances (exemplars). Experts have a larger store of recalled cases, for example, visual memory in radiology. In the more abstract prototype model of knowledge, clinicians do not simply rely on specific cases but have constructed elaborate conceptual networks or models of disease to arrive at their conclusions. That is, expertise involves an increased ability to connect symptoms, signs, and risk factors to one another; relate those findings to possible diagnoses; and identify the additional information necessary to confirm the diagnosis. More recently, fuzzy trace theory has placed greater emphasis on intuitionism in which expertise involves the ability to distill the "gist" or essence of diagnosis by processing less information and discarding extraneous data with an emphasis on memory and meaning or recognition and retrieval.

Although no single theory has emerged to account for the key features of expertise in medical diagnosis, experts have more knowledge about more things and a larger repertoire of cognitive tools to employ in problem solving than do novices. One definition of expertise highlights the ability to make powerful distinctions. Memorization alone is insufficient; instead, expertise involves a working knowledge of the diagnostic possibilities and what features distinguish one from another. What remains less clear is whether there is any didactic program that would allow the accelerated development of a novice into an expert or ensure the same high level of expertise among more experienced physicians. Some current recommendations include using a combined approach to clinical reasoning, that is, emphasizing to students the importance of both conscious deliberative analytic and intuitive pattern recognition nonanalytic reasoning strategies and thus giving students flexibility in applying any particular reasoning strategy to overcome case-specific weaknesses.

Important Modifiers of Clinical Decision-Making

More than a decade of research on variations in clinician practice patterns has shed much light on the forces that shape clinical decisions. These factors can be grouped conceptually into three overlapping categories: (1) factors related to physicians' personal characteristics and practice style, (2) factors related to the practice setting, and (3) factors related to economic incentives.

FACTORS RELATED TO PRACTICE STYLE

One of the key roles of the physician in medical care is to serve as the patient's agent to ensure that necessary care is provided at a high level of quality. Factors that influence this role include the physician's knowledge, training, and experience. It is obvious that physicians cannot practice evidence-based medicine (described later in the chapter) if they are unfamiliar with the evidence. As would be expected, specialists generally know the evidence in their field better than do generalists. Surgeons may be more enthusiastic about recommending surgery than are medical doctors because their belief in the beneficial effects of surgery is stronger. For the same reason, invasive cardiologists are much more likely to refer chest pain patients for diagnostic catheterization than are noninvasive cardiologists or generalists. The physician beliefs that drive these different practice styles are based on personal experience, recollection, and interpretation of the available medical evidence. For example, heart failure specialists are much more likely than generalists to achieve target angiotensin-converting enzyme (ACE) inhibitor therapy in their heart failure patients because they are more familiar with what the targets are (as defined by large clinical trials), have more familiarity with the specific drugs (including doses and side effects), and are less likely to overreact to foreseeable problems in therapy such as a rise in creatinine levels or symptomatic hypotension. Other intriguing research has shown a wide distribution of acceptance times of antibiotic therapy for peptic ulcer disease after widespread dissemination of the "evidence" in the medical literature. Some gastroenterologists accepted this new therapy before the evidence was clear (reflecting, perhaps, an aggressive practice style), and some lagged behind (a conservative practice style, associated in this case with older physicians). As a group, internists lagged several years behind gastroenterologists.

An example of the mixed effects on patient outcomes associated with rapid acceptance of new evidence involves the case of adding spironolactone (an aldosterone receptor antagonist) to the drug regimen for patients with systolic heart failure. In a large, well-done clinical trial (Randomized Aldactone Evaluation Study, RALES) published in 1999, this therapy produced a significant reduction in all-cause

mortality rates. Over the next 2 years the use of spironolactone increased fivefold in the province of Ontario, Canada. That rapid uptake was associated with a significant increase in the rates of hospital admission for both hyperkalemia- and hyperkalemia-associated deaths. At least some of these adverse effects of using this "evidence-based medicine" appeared to be related to treatment of patients who would not have been eligible for the RALES trial and who had contraindications to the use of the drug.

The opinion of influential leaders also can have an important effect on practice patterns. That influence can occur at both the national level (e.g., expert physicians teaching at national meetings) and the local level (e.g., local educational programs, "curbside consultations"). Opinion leaders do not have to be physicians. When conducting rounds with clinical pharmacists, physicians are less likely to make medication errors and more likely to use target levels of evidence-based therapies.

The patient's welfare is not the only concern that drives clinical decisions. The physician's perception about the risk of a malpractice suit resulting from either an erroneous decision or a bad outcome creates a style of practice referred to as *defensive medicine*. This practice involves using tests and therapies with very small marginal returns to preclude future criticism if there is an adverse outcome. For example, a 40-year-old woman who presents with a long-standing history of intermittent headache and a new severe headache along with a normal neurologic examination has a very low likelihood of having structural intracranial pathology. Performance of a head CT or magnetic resonance imaging (MRI) scan in this situation would constitute defensive medicine. However, the results of the test could provide reassurance to an anxious patient.

PRACTICE SETTING FACTORS

Factors in this category relate to the physical resources available to the physician's practice and the practice environment. *Physician-induced demand* is a term that refers to the repeated observation that physicians have a remarkable ability to accommodate to and employ the medical facilities available to them. One of the foundational studies in outcomes research showed that physicians in Boston, where the ratio of hospital beds to patients was higher, had an almost 50% higher hospital admission rate than did physicians in New Haven, despite there being no obvious differences in the resulting health or mortality rate of the cities' inhabitants. The physicians in New Haven were not aware of using fewer hospital beds for their patients, nor were the Boston physicians aware of using less stringent criteria to admit patients. In both cities, physicians unconsciously adopted their practice styles to the available level of hospital beds.

Other environmental factors that can influence decision-making include the local availability of specialists for consultations and procedures; "high-tech" facilities such as angiography suites, a heart surgery program, and MRI machines; and fragmentation of care.

ECONOMIC INCENTIVES

Economic incentives are closely related to the other two categories of practice-modifying factors. Financial issues can exert both stimulatory and inhibitory influences on clinical practice. In general, physicians are paid on a fee-for-service, capitation, or salary basis. In fee-for-service, the more the physician does, the more he gets paid. The economic incentive in this case is to do more. When fees are reduced (discounted fee-for-service), doctors tend to increase the number of services provided. Capitation, in contrast, provides a fixed payment per patient per year, encouraging physicians to take on more patients but to provide each patient with fewer services. Expensive services are more likely to be affected by this type of incentive than are inexpensive preventive services. Salary compensation plans pay physicians the same regardless of the amount of clinical work performed. The incentive here is to see fewer patients.

In summary, expert clinical decision-making can be appreciated as a complex interplay between cognitive processes used to simplify and organize large amounts of complex information and physician biases reflecting education, training, and experience, all of which are shaped by powerful, sometimes perverse, external forces. In the next section, we will review a set of statistical tools and concepts that can be useful in making clinical decisions in the presence of uncertainty.

INTERPRETATION OF DIAGNOSTIC TESTS IN THE CONTEXT OF DECISION-MAKING

Despite the great technological advances in medicine over the last century, uncertainty remains a key challenge in all aspects of medical decision-making. Compounding this challenge is the massive information overload that characterizes modern medicine. Today's experienced clinician needs access to close to 2 million pieces of information to practice medicine. According to one estimate, doctors subscribe to an average of seven journals, representing over 2500 new articles each year. Of course, to be useful, this information must be integrated with the specific data collected on each patient being cared for. Although computers appear to offer the obvious solution both for management of information and for better quantitation and management of the daily uncertainties of medical care, many practical problems must be solved before computers can be integrated into the clinician's reasoning process in a way that demonstrably improves the quality of care.

Although a fully-integrated computed-based system of diagnosis and management remains a distant possibility, there are tools available now that can assist in aspects of patient management. In addition,

understanding the nature of diagnostic test information can help make a clinician a more efficient user of such data. This section of the chapter will review some important concepts related to diagnostic testing.

DIAGNOSTIC TESTING: MEASURES OF TEST ACCURACY

The purpose of performing a test on a patient is to reduce uncertainty about the patient's diagnosis or prognosis and to aid the clinician in making management decisions. Although diagnostic tests commonly are thought of as laboratory tests (e.g., measurement of serum amylase level) or procedures (e.g., colonoscopy or bronchoscopy), any technology that changes a physician's understanding of the patient's problem qualifies as a diagnostic test. Thus, even the history and physical examination can be considered a form of diagnostic test. In clinical medicine, it is common to reduce the results of a test to a dichotomous outcome, such as positive or negative, normal or abnormal. In many cases, this simplification results in the waste of useful information. However, such simplification makes it easier to demonstrate some of the quantitative ways in which test results data can be used.

The accuracy of diagnostic tests is defined in relation to an accepted "gold standard," which is presumed to reflect the true state of the patient (Table 3-1). To define the diagnostic performance of a new test, an appropriate population must be identified (ideally, patients on whom the new test would be used), and both the new and the gold standard tests are applied to all subjects (use of an inappropriate population or incomplete application of the gold standard test may lead to biased estimates of test performance). The results of the two tests are then compared. The *sensitivity,* or *true-positive, rate* of the new test is the proportion of patients with disease (defined by the gold standard) who have a positive (new) test. This measure reflects how well the test identifies patients with disease. The proportion of patients with disease who have a negative test is the *false-negative rate* and is calculated as 1 – sensitivity. The proportion of patients without disease who have a negative test is the *specificity,* or *true-negative, rate*. This measure reflects how well the test correctly identifies patients without disease. The proportion of patients without disease who have a positive test is the *false-positive rate*, calculated as 1 – specificity. A perfect test would have a sensitivity of 100% and a specificity of 100% and would completely separate patients with disease from those without it.

Table 3-1 Measures of Diagnostic Test Accuracy

	Disease Status	
Test Result	**Present**	**Absent**
Positive	True-positive (TP)	False-positive (FP)
negative	false-negative (fn)	true-negative (tn)
Identification of patients with disease		
True-positive rate (sensitivity) = TP/(TP + FN)		
False-negative rate = FN/(TP + FN)		
True-positive rate = 1 – false-negative rate		
Identification of patients without disease		
True-negative rate (specificity) = TN/(TN + FP) False-positive rate = FP/(TN + FP)		
True-negative rate = 1 – false-positive rate		

Calculating sensitivity and specificity requires selection of a threshold value or cut point at or above which the test is considered "positive." For any specific test, as this cut point is moved to improve sensitivity, specificity falls and vice versa. This dynamic trade-off between more accurate identification of subjects with disease versus those without disease is often displayed graphically as a receiver operating characteristic (ROC) curve (Fig. 3-1). An ROC curve plots sensitivity (*y* axis) versus 1 – specificity (*x* axis). Each point on the curve represents a potential cut point with an associated sensitivity and specificity value. The area under the ROC curve often is used as a quantitative measure of the information content of a test. Values range from 0.5 (no diagnostic information from testing at all; the test is equivalent to flipping a coin) to 1.0 (perfect test).

Figure 3-1

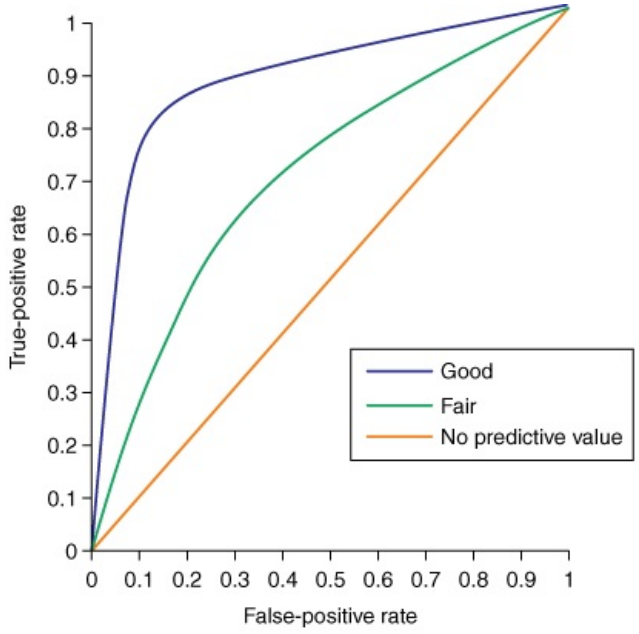

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Each ROC curve illustrates a trade-off that occurs between improved test sensitivity (accurate detection of patients with disease) and improved test specificity (accurate detection of patients without disease), as the test value defining when the test turns from "negative" to "positive" is varied. A 45° line would indicate a test with no predictive value (sensitivity = specificity at every test value). The area under each ROC curve is a measure of the informa-tion content of the test. Thus, a larger ROC area signifies increased diagnostic accuracy.

In the testing literature, ROC areas often are used to compare alternative tests that can be employed for a particular diagnostic problem. The test with the highest area (i.e., closest to 1.0) is presumed to be the most accurate. However, ROC curves are not a panacea for evaluation of diagnostic test utility. Like Bayes' theorem (discussed below), they typically are focused on only one possible test parameter (e.g., ST-segment response in a treadmill exercise test) to the exclusion of other potentially relevant data. In addition, ROC area comparisons do not simulate the way test information actually is used in clinical practice. Finally, biases in the underlying population used to generate the ROC curves (e.g., related to assessment of the test in individuals unrepresentative of those in whom the test will be used clinically) can bias the ROC area and the validity of a comparison among tests.

MEASURES OF DISEASE PROBABILITY AND BAYES' THEOREM

Unfortunately, there are no perfect tests. After every test is completed, the true disease state of the patient remains uncertain. Quantifying this residual uncertainty can be done with Bayes' theorem. This theorem provides a simple mathematical way to calculate the posttest probability of disease from three parameters: the pretest probability of disease, the test sensitivity, and the test specificity (Table 3-2). The pretest probability is a quantitative expression of the confidence in a diagnosis before the test is performed. In the absence of more relevant information, it is usually estimated from the prevalence of the disease in the underlying population. For some common conditions, such as CAD, nomograms and statistical models have been created to generate better estimates of pretest probability from elements of the history and physical examination. The posttest probability, then, is a revised statement of the confidence in the diagnosis, taking into account what was known both before and after the test.

Table 3-2 Measures of Disease Probability

Pretest probability of disease = probability of disease before test is performed. May use population prevalence of disease or more patient-specific data to generate this probability estimate.
Posttest probability of disease = probability of disease accounting for both pretest probability and test results. Also called pre-dictive value of the test.
Bayes' theorem computational version: $\text{Posttest probability} = \frac{\text{Pretest probability} \times \text{test sensitivity}}{\text{Pretest probability} \times \text{test sensitivity} + (1 - \text{pretest probability}) \times \text{test false-positive rate}}$
Bayes' theorem example: With a pretest probability of 0.50 and a "positive" diagnostic test result (test sensitivity = 0.90, test specificity = 0.90):

$$\text{Posttest probability} = \frac{(0.50)(0.90)}{(0.50)(0.90) + (0.50)(0.10)} = 0.90$$

The term *predictive value* often is used as a synonym for the posttest probability. Unfortunately, clinicians commonly misinterpret reported predictive values as intrinsic measures of test accuracy. Studies of diagnostic tests compound the confusion by calculating predictive values on the same sample used to measure sensitivity and specificity. Since all posttest probabilities are a function of the prevalence of disease in the tested population, such calculations may be misleading unless the test is applied subsequently to populations with the same disease prevalence. For these reasons, the term *predictive value* is best avoided in favor of the more informative *posttest probability*.

To understand conceptually how Bayes' theorem estimates the posttest probability of disease, it is useful to examine a nomogram version of Bayes' theorem (Fig. 3-2). In this nomogram, the accuracy of the diagnostic test in question is summarized by the *likelihood ratio*, which is defined as the ratio of the probability of a given test result (e.g., "positive" or "negative") in a patient with disease to the probability of that result in a patient without disease.

Figure 3-2

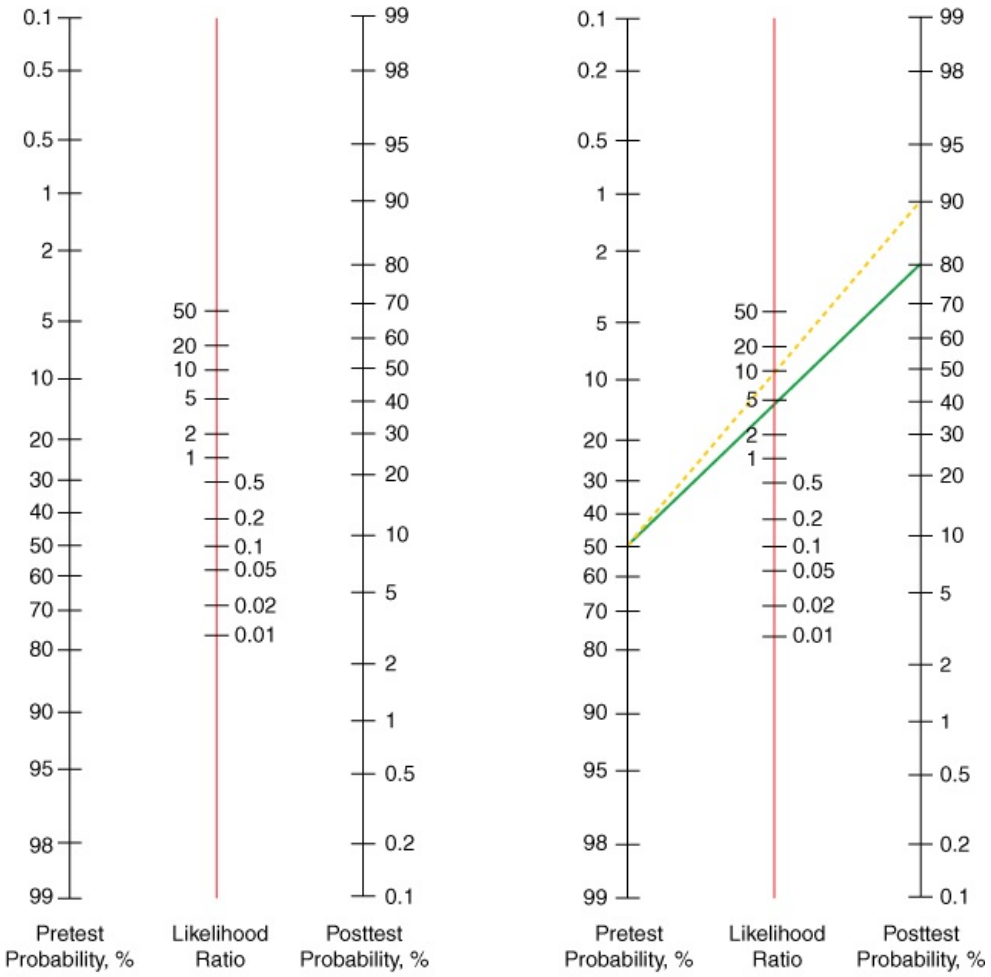

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nomogram version of Bayes' theorem used to predict the posttest probability of disease (right-hand scale) using the pretest probability of disease (left-hand scale) and the likelihood ratio for a positive test (middle scale). See text for information on calculation of likelihood ratios. To use, place a straightedge connecting the pretest probability and the likelihood ratio and read off the posttest probability. The right-hand part of the figure illustrates the value of a positive exercise treadmill test (likelihood ratio 4, green line) and a positive exercise thallium single-photon emission CT perfusion study (likelihood ratio 9, broken yellow line) in a patient with a pretest probability of coronary artery disease of 50%. *(Adapted from Centre for Evidence-Based Medicine: Likelihood ratios. Available at http://www.cebm.net/index.aspx?o=1043.)*

For a positive test, the likelihood ratio positive is calculated as the ratio of the true-positive rate to the false-positive rate [or sensitivity/(1 – specificity)]. For example, a test with a sensitivity of 0.90 and a

specificity of 0.90 has a likelihood ratio of 0.90/(1 – 0.90), or 9. Thus, for this hypothetical test, a "positive" result is 9 times more likely in a patient with the disease than in a patient without it. Most tests in medicine have likelihood ratios for a positive result between 1.5 and 20. Higher values are associated with tests that are more accurate at identifying patients with disease, with values of 10 or greater are of particular note. If sensitivity is excellent but specificity is less so, the likelihood ratio will be reduced substantially (e.g., with a 90% sensitivity but a 60% specificity, the likelihood ratio is 2.25).

For a negative test, the corresponding likelihood ratio negative is the ratio of the false-negative rate to the true-negative rate [or (1 – sensitivity)/specificity]. The smaller the likelihood ratio (i.e., the closer to 0) is, the better the test performs at ruling out disease. The hypothetical test considered above with a sensitivity of 0.9 and a specificity of 0.9 would have a likelihood ratio for a negative test result of (1 – 0.9)/0.9, or 0.11, meaning that a negative result is almost 10 times more likely if the patient is disease-free than if the patient has disease.

APPLICATIONS TO DIAGNOSTIC TESTING IN CAD

Consider two tests commonly used in the diagnosis of CAD: an exercise treadmill and an exercise single-photon emission CT (SPECT) myocardial perfusion imaging test (Chap. 229). Meta-analysis has shown that a positive treadmill ST-segment response has an average sensitivity of 66% and an average specificity of 84%, yielding a likelihood ratio of 4.1 [0.66/(1 – 0.84)]. If this test is used on a patient with a pretest probability of CAD of 10%, the posttest probability of disease after a positive result rises to only about 30%. If a patient with a pretest probability of CAD of 80% has a positive test result, the posttest probability of disease is about 95%.

The exercise SPECT myocardial perfusion test is a more accurate test for the diagnosis of CAD. For our purposes, assume that the finding of a reversible exercise-induced perfusion defect has both a sensitivity and a specificity of 90%, yielding a likelihood ratio for a positive test of 9.0 [0.90/(1 – 0.90)]. If we again test the low pretest probability patient and that patient has a positive test, by using Fig. 3-2 it can be demonstrated that the posttest probability of CAD rises from 10 to 50%. However, from a decision-making point of view, the more accurate test may not improve diagnostic confidence enough to change management. In fact, the test has moved the physician from being fairly certain that the patient did not have CAD to being completely undecided (a 50:50 chance of disease). In a patient with a pretest probability of 80%, using the more accurate exercise SPECT test raises the posttest probability to 97% (compared with 95% for the exercise treadmill). Again, the more accurate test does not provide enough improvement in posttest confidence to alter management, and neither test has improved much on what was known from clinical data alone.

Although it depends on the sensitivity and specificity, in general, if the pretest probability is low (e.g., 20%), even a positive result on a very accurate test will not move the posttest probability to a range high enough to rule in disease (e.g., 80%). Pretest probabilities are often particularly low in screening situations in which patients are asymptomatic. In such cases, specificity becomes particularly important. For example, in screening first-time female blood donors without risk factors for HIV, a positive test raised the likelihood of HIV to only 67% despite a specificity of 99.995% because the prevalence was 0.01%. One useful mnemonic is *positive SpPin*: a positive test with high specificity rules in disease (keeping in mind the caveats just noted about pretest probability). Conversely, with a high pretest probability, a negative test may not rule out disease adequately if it is not sufficiently sensitive. The other mnemonic is *negative SnNout*: a negative test with high sensitivity rules out disease. Thus, the largest gain in diagnostic confidence from a test occurs when the clinician is most uncertain before performing it (e.g., pretest probability between 30% and 70%). For example, if a patient has a pretest probability for CAD of 50%, a positive exercise treadmill test will move the posttest probability to 80% and a positive exercise SPECT perfusion test will move it to 90% (Fig. 3-2).

Bayes' theorem, as presented above, employs a number of important simplifications that should be considered. First, few tests have only two useful outcomes, positive and negative, and many tests provide numerous pieces of data about the patient. Even if these data can be integrated into a summary result, multiple levels of useful information may be present (e.g., strongly positive, positive, indeterminate, negative, strongly negative). Although Bayes' theorem can be adapted to this more detailed test result format, it is computationally complex to do so. Similarly, when multiple tests are performed, the posttest probability may be used as the pretest probability to interpret the second test. However, this simplification assumes conditional independence—that is, that the results of the first test do not affect the likelihood of the second test result—and this is often not true.

Finally, it has long been asserted that sensitivity and specificity are prevalence-independent parameters of test accuracy, and many texts still make this statement. This statistically useful assumption, however, is clinically simplistic. A treadmill exercise test, for example, has a sensitivity in a population of patients with one-vessel CAD of around 30%, whereas its sensitivity in patients with severe three-vessel CAD approaches 80%. Thus, the best estimate of sensitivity to use in a particular decision often varies, depending on the distribution of disease stages present in the tested population. A hospitalized, symptomatic, or referral population typically has a higher prevalence of disease and, in particular, a higher prevalence of more advanced disease than does an outpatient population. As a consequence, test sensitivity will tend to be higher in hospitalized patients, whereas test specificity will

be higher in outpatients.

STATISTICAL PREDICTION MODELS

Bayes' theorem, as presented above, deals with a clinical prediction problem that is unrealistically simple relative to most problems a clinician faces. Prediction models that are based on multivariable statistical models can handle much more complex problems and substantially enhance predictive accuracy for specific situations. Their particular advantage is the ability to take into account many overlapping pieces of information and assign a relative weight to each on the basis of its unique contribution to the prediction in question. For example, a logistic regression model to predict the probability of CAD considers all the relevant independent factors from the clinical examination and diagnostic testing and their significance instead of the small handful of data that clinicians can manage in their heads or with Bayes' theorem. However, despite this strength, the models are too complex computationally to use without a calculator or computer (although this limitation may be overcome once medicine is practiced from a fully computerized platform).

To date, only a handful of prediction models have been validated properly. The importance of independent validation in a population separate from the one used to develop the model cannot be overstated. An unvalidated prediction model should be viewed with the skepticism appropriate for a new drug or medical device that has not been through rigorous clinical trial testing.

When statistical models have been compared directly with expert clinicians, they have been found to be more consistent, as would be expected, but not significantly more accurate. Their biggest promise, then, would seem to be to help less-experienced clinicians become more accurate in their predictions.

FORMAL DECISION SUPPORT TOOLS

Decision Support Systems

Over the last 40 years, many attempts have been made to develop computer systems to aid clinical decision-making and patient management. Conceptually, computers offer a very attractive way to handle the vast information load that today's physicians face. The computer can help by making accurate predictions of outcome, simulating the whole decision process, or providing algorithmic guidance. Computer-based predictions using Bayesian or statistical regression models inform a clinical decision but do not actually reach a "conclusion" or "recommendation." Artificial intelligence systems attempt to simulate or replace human reasoning with a computer-based analogue. To date, such approaches have achieved only limited success. Reminder or protocol-directed systems do not make predictions but use existing algorithms, such as guidelines, to guide clinical practice. In general, however, decision support systems have had little impact on practice. Reminder systems, although not yet in widespread use, have shown the most promise, particularly in correcting drug dosing and promoting adherence to guidelines. The full impact of these approaches will be evaluable only when computers are fully integrated into medical practice.

Decision Analysis

Compared with the methods discussed above, decision analysis represents a completely different approach to decision support. Its principal application is in decision problems that are complex and involve a substantial risk, a high degree of uncertainty in some key area, or an idiosyncratic feature that does not "fit" the available evidence. An example decision tree created to evaluate strategies for screening for HIV infection is shown in Fig. 3-3. Infected individuals who are unaware of their illness may cause up to 20,000 new cases of HIV infection annually in the United States. In addition, because of delayed diagnosis, about 40% of HIV-positive patients progress to AIDS within a year of the initial diagnosis. Early identification offers the opportunity both to prevent progression to AIDS through the use of serial CD4 counts and measurements of viral load linked to selective use of combination antiretroviral therapy and to encourage reduction of risky sexual behavior.

Figure 3-3

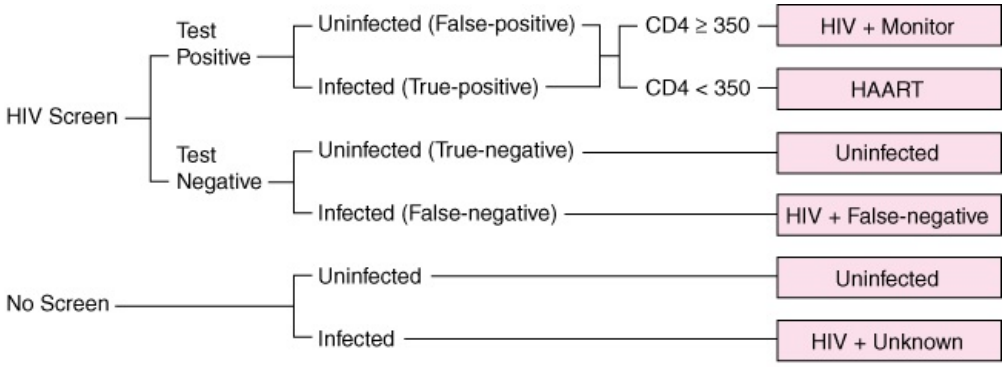

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Basic structure of decision model used to evaluate strategies for screening for HIV in the general population. HIV, human immunodeficiency virus; HAART, highly active antiretroviral therapy. *(Provided courtesy of G. Sanders, with permission.)*

The Centers for Disease Control and Prevention (CDC) proposed in 2003 that routine HIV testing should be a part of standard medical care. In a decision-model exploration of this proposed strategy compared with usual care, assuming a 1% prevalence of unidentified HIV infection in the population, routine screening of a cohort of 43-year-old men and women increased life expectancy by 5.5 days and cost $194 per subject screened. The cost-effectiveness ratio for screening relative to usual care was $15,078 per quality-adjusted life year (the additional cost to society to increase population health by 1 year of perfect health). Results were sensitive to assumptions about the effectiveness of behavior modification on subsequent sexual behavior, the benefits of early therapy for HIV infection, and the prevalence and incidence of HIV infection in the population targeted. This model, which required over 75 separate data points, provides novel insights into a clinical management problem that has not been subjected to a randomized clinical trial.

Although such models have been developed and used to estimate short- and long-term survival for alternative choices, the process of building and evaluating decision models is generally too complex for use in real-time clinical management. The potential for this tool therefore lies in the development of a set of published or online models addressing a particular decision or policy area that can serve to highlight key pressure points in the problem.

EVIDENCE-BASED MEDICINE

The "art of medicine" is defined traditionally as a practice combining medical knowledge (including scientific evidence), intuition, and judgment in the care of patients (Chap. 1). EBM updates this construct by placing much greater emphasis on the processes by which clinicians gain knowledge of the most up-to-date and relevant clinical research to determine for themselves whether medical interventions alter the disease course and improve the length or quality of life. The meaning of practicing EBM becomes clearer through an examination of its four key steps:

1. Formulating the management question to be answered
2. Searching the literature and online databases for applicable research data
3. Appraising the evidence gathered with regard to its validity and relevance
4. Integrating this appraisal with knowledge about the unique aspects of the patient (including the patient's preferences about the possible outcomes)

Step 1 involves generating well-formulated questions that involve four or five components—PICOD: patient or population, intervention, comparator, outcome, and, sometimes, D for study design, (e.g., does routine percutaneous coronary intervention improve survival compared with initial medical management in 60-year-old men with stable angina and known CAD?) Steps 2 and 3 are the heart of EBM as it is currently used in practice and relate to the underlying fundamental principle that the strength of medical evidence supporting a therapy or strategy is hierarchical. The process of searching the world's research literature and appraising the quality and relevance of studies thus identified can be quite time-consuming and requires skills and training that most clinicians do not possess. Thus, the best starting point for most EBM searches is the identification of recent systematic overviews of the problem in question (Table 3-3).

Table 3-3 Selected Tools for Finding the Evidence in Evidence-Based Medicine

Name	Description	Web Address	Availability
Evidence-	Comprehensive electronic database that	www.ovid.com	Subscription

Based Medicine Reviews	combines and integrates: 1. The Cochrane Database of Systematic Reviews 2. ACP Journal Club 3. The Database of Abstracts of Reviews of Effectiveness		required. Available through medical center libraries and other institutions.
Cochrane Library	Collection of EBM databases, including The Cochrane Database of Systematic Reviews—full text articles reviewing specific health care topics.	www.cochrane.org	Subscription required. Abstracts of systematic reviews available free online. Some countries have funding to provide free access to all residents.
ACP Journal Club	Collection of summaries of original studies and systematic reviews. Published bimonthly. All data since 1991 available on Web site, updated yearly.	www.acpjc.org	Subscription required.
Clinical Evidence	Monthly updated directory of concise overviews of common clinical interventions.	www.clinicalevidence.com	Subscription required. Free access for United Kingdom and developing countries.
MEDLINE	National Library of Medicine database with citations back to 1966.	www.nlm.nih.gov	Free via Internet.

Generally, the EBM tools listed in Table 3-3 provide access to research information in one of two forms. The first, primary research reports, is the original peer-reviewed research work that is published in medical journals. Initial access to this information in an EBM search may be gained through MEDLINE, which provides access to a huge amount of data in abstract form. However, in using MEDLINE it is often difficult to locate reports that are on point in a sea of irrelevant or unhelpful information and be reasonably certain that important reports have not been overlooked. The second form, systematic reviews, comprehensively summarizes the available evidence on a particular topic up to a certain date and provides the interpretation of the reviewer and thus is the highest level of evidence in the hierarchy. Explicit criteria are used to find all the relevant scientific research and grade its quality. The prototype for this kind of resource is the Cochrane Database of Systematic Reviews. One of the key components of a systematic review is a meta-analysis. The next two sections will review some of the major types of clinical research reports available in the literature and the process of aggregating those data into meta-analyses.

SOURCES OF EVIDENCE: CLINICAL TRIALS AND REGISTRIES

The notion of learning from observation of patients is as old as medicine itself. Over the last 50 years, physicians' understanding of how best to turn raw observation into useful evidence has evolved considerably. Case reports, personal anecdotal experience, and small single-center case series are now recognized as having severe limitations in validity and generalizability, and although they may generate hypotheses or be the first reports of adverse events, they have no role in formulating modern standards of practice. The major tools used to develop reliable evidence consist of the randomized clinical trial and the large observational registry. A registry or database typically is focused on a disease or syndrome (e.g., cancer, CAD, heart failure), a clinical procedure (e.g., bone marrow transplantation, coronary revascularization), or an administrative process (e.g., claims data used for billing and reimbursement).

By definition, in observational data, the care of the patient is not controlled by the investigator. Carefully collected *prospective* observational data can achieve a level of quality approaching that of major clinical trial data. At the other end of the spectrum, data collected retrospectively (e.g., chart review) are limited in form and content to what previous observers thought was important to record, which may not serve the research question under study particularly well. Data not specifically collected for research (e.g., claims data) often have important limitations that cannot be overcome in the analysis phase of the research. Advantages of observational data include the ability to capture a broader population than is typically represented in clinical trials because of inclusion and exclusion criteria. In addition, observational data are the primary source of evidence for questions for which a randomized trial cannot or will not be performed. For example, it may be difficult or unethical to randomize patients to test diagnostic or therapeutic strategies that are unproven but widely accepted in practice. In addition, patients cannot be randomized to a sex, racial/ethnic group, socioeconomic status, or country of residence. Physicians are also not willing to randomize patients to a potentially harmful intervention, such as smoking or overeating to develop obesity.

The major difference between a well-done randomized clinical trial and a well-done prospective

observational study of a particular management strategy is the lack of protection from treatment selection bias in the latter. The use of observational data to compare diagnostic or therapeutic strategies assumes that there is sufficient uncertainty in practice to ensure that similar patients will be managed differently by different physicians. In short, the analysis assumes that there is an element of randomness (in the sense of disorder rather than in the formal statistical sense) to clinical management. In such cases, statistical models attempt to adjust for important imbalances and "level the playing field" so that a fair comparison among treatment options can be made. When management is clearly not random (e.g., all eligible left main coronary artery disease patients are referred for coronary bypass surgery), the problem may be too confounded (biased) for statistical correction, and observational data may not provide reliable evidence.

In general, the use of concurrent controls is vastly preferable to that of historical controls. For example, comparison of current surgical management of left main CAD with left main CAD patients treated medically during the 1970s (the last time these patients were routinely treated with medicine alone) would be extremely misleading since the quality of "medical therapy" has made substantial improvements in the interval.

Randomized controlled clinical trials include the careful prospective design features of the best observational data studies but also include the use of random allocation of treatment. This design provides the best protection against confounding due to treatment selection bias (a major aspect of internal validity). However, the randomized trial may not have good external validity (generalizability) if the process of recruitment into the trial resulted in the exclusion of many potentially eligible subjects.

Consumers of medical evidence need to be aware that randomized trials vary widely in their quality and applicability to practice. The process of designing such a trial often involves a great many compromises. For example, trials designed to gain U.S. Food and Drug Administration (FDA) approval for an investigational drug or device have to address certain regulatory requirements that may result in a trial design different from what practicing clinicians would find useful.

META-ANALYSIS

The Greek prefix *meta* signifies something at a later or higher stage of development. Meta-analysis is research done on research data for the purpose of combining and summarizing the available evidence quantitatively. Although it can be used to combine nonrandomized studies, meta-analysis is used most typically to summarize all the randomized trials on a particular therapeutic problem. Ideally, unpublished trials should be identified and included to avoid publication bias (i.e., "negative" trials may not be published). Furthermore, some of the best meta-analyses obtain and analyze the raw individual patient-level data from all trials rather than working only with what is available in the published reports of each trial. Not all published meta-analyses are reliable sources of evidence on a particular problem. Their methodology must be scrutinized carefully to ensure proper study design and analysis. The results of a well-done meta-analysis are likely to be most persuasive if they include at least several large-scale, properly performed randomized trials. Although meta-analysis can help detect benefits when individual trials are inadequately powered (e.g., the benefits of streptokinase thrombolytic therapy in acute MI demonstrated by ISIS-2 in 1988 were evident by the early 1970s through meta-analysis), in cases in which the available trials are small or poorly done, meta-analysis should not be viewed as a remedy for the deficiency in primary trial data.

Meta-analyses typically focus on summary measures of relative treatment benefit, such as odds ratios or relative risks. Clinicians also should examine what absolute risk reduction (ARR) can be expected from the therapy. A useful summary metric of absolute treatment benefit is the number needed to treat (NNT) to prevent one adverse outcome event (e.g., death, stroke). NNT is simply 1/ARR. For example, if a hypothetical therapy reduced mortality rates over a 5-year follow-up by 33% (the relative treatment benefit) from 12% (control arm) to 8% (treatment arm), the absolute risk reduction would be 12% – 8% = 4% and the NNT would be 1/.04, or 25. Thus, it would be necessary to treat 25 patients for 5 years to prevent 1 death. If the hypothetical treatment was applied to a lower-risk population, say, with a 6% 5-year mortality, the 33% relative treatment benefit would reduce absolute mortality by 2% (from 6 to 4%), and the NNT for the same therapy in this lower-risk group of patients would be 50. Although not always made explicit, comparisons of NNT estimates from different studies should account for the duration of follow-up used to create each estimate.

CLINICAL PRACTICE GUIDELINES

According to the 1990 Institute of Medicine definition, clinical practice guidelines are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances." This definition emphasizes several crucial features of modern guideline development. First, guidelines are created by using the tools of EBM. In particular, the core of the development process is a systematic literature search followed by a review of the relevant peer-reviewed literature. Second, guidelines usually are focused on a clinical disorder (e.g., adult diabetes, stable angina pectoris) or a health care intervention (e.g., cancer screening). Third, the primary objective of guidelines is to improve the quality of medical care by identifying areas where care should be standardized, based on compelling evidence. Guidelines are intended to "assist" decision-making, not to define explicitly what decisions should be made in a particular situation, in part because

evidence alone is never enough for clinical decision-making (e.g., deciding whether to intubate and administer antibiotics for pneumonia in a terminally ill individual, in an individual with dementia, or in an otherwise healthy 30-year-old mother).

Guidelines are narrative documents constructed by an expert panel whose composition often is determined by interested professional organizations. These panels vary in the degree to which they represent all relevant stakeholders. The guideline documents consist of a series of specific management recommendations, a summary indication of the quantity and quality of evidence supporting each recommendation, and a narrative discussion of the recommendations. Many recommendations have little or no supporting evidence and thus reflect the expert consensus of the guideline panel. In part to protect against errors by individual panels, the final step in guideline construction is peer review, followed by a final revision in response to the critiques provided.

Guidelines are closely tied to the process of quality improvement in medicine through their identification of evidence-based best practices. Such practices can be used as quality indicators. Examples include the proportion of acute MI patients who receive aspirin upon admission to a hospital and the proportion of heart failure patients with a depressed ejection fraction treated with an ACE inhibitor. Routine measurement and reporting of such quality indicators can produce selective improvements in quality, since many physicians prefer not to be outliers.

CONCLUSIONS

In this era of EBM, it is tempting to think that all the difficult decisions practitioners face have been or soon will be solved and digested into practice guidelines and computerized reminders. However, EBM provides practitioners with an ideal rather than a finished set of tools with which to manage patients. The significant contribution of EBM has been to promote the development of more powerful and user-friendly EBM tools that can be accessed by busy practitioners. This is an enormously important contribution that is slowly changing the way medicine is practiced. One of the repeated admonitions of EBM pioneers has been to replace reliance on the local "gray-haired expert" (who may be wrong but is rarely in doubt) with a systematic search for and evaluation of the evidence. But EBM has not eliminated the need for subjective judgments. Each systematic review or clinical practice guideline presents the interpretation of "experts" whose biases remain largely invisible to the review's consumers. Moreover, even with such evidence, it is always worth remembering that the response to therapy of the "average" patient represented by the summary clinical trial outcomes may not be what can be expected for the patient sitting in front of a physician in the clinic or hospital. In addition, meta-analyses cannot generate evidence when there are no adequate randomized trials, and most of what clinicians confront in practice will never be thoroughly tested in a randomized trial. For the foreseeable future, excellent clinical reasoning skills and experience supplemented by well-designed quantitative tools and a keen appreciation for individual patient preferences will continue to be of paramount importance in the professional life of medical practitioners.

FURTHER READINGS

Del Mar C et al: *Clinical Thinking: Evidence, Communication and Decision Making* , Malden, MA, Blackwell, 2006

Graber ML et al: Diagnostic error in internal medicine. Arch Intern Med 65:1493, 2005

Grimes DA et al: Refining clinical diagnosis with likelihood ratios. Lancet 365:1500, 2005[PMID: 15850636]

Kassirer JP et al. *Learning Clinical Reasoning* . Baltimore, Lippincott Williams & Wilkins, 2009

Norman G et al: Non-analytical models of clinical reasoning: The role of experience. Med Educ 41:1140, 2007[PMID: 18004990]

Peterson ED et al: Association between hospital process performance and outcomes among patients with acute coronary syndromes. JAMA 295:1912, 2006[PMID: 16639050]

Reilly BM et al: Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. Ann Intern Med 144:201, 2006[PMID: 16461965]

Sanders GD et al: Cost-effectiveness of screening for HIV in the era of highly active antiretroviral therapy. N Engl J Med 352:570, 2005[PMID: 15703422]

Tricoci P et al: Scientific evidence underlying the AHA/ACC clinical practice guidelines. JAMA 301:831, 2009; erratum in JAMA 301:1544, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SCREENING AND PREVENTION OF DISEASE: INTRODUCTION

A primary goal of health care is to prevent disease or detect it early enough that intervention will be more effective. Strategies for disease screening and prevention are driven by evidence that testing and intervention are practical and effective. Currently, most screening tests are readily available and inexpensive. Examples include tests that are biochemical (e.g., cholesterol, glucose), physiologic (e.g., blood pressure, growth curves), or radiologic (e.g., mammogram, bone densitometry) or that involve tissue specimens (e.g., Pap smear). In the future, it is anticipated that genetic testing will play an increasingly important role in predicting disease risk (Chap. 63). However, such tests are not widely used except for individuals at risk for high-penetrancegenes on the basis of family or ethnic history (e.g., *BRCA1*, *BRCA2*). The identification of low-penetrance but high-frequency genes that cause common disorders such as diabetes, hypertension, and macular degeneration offers the possibility of new genetic tests. However, any new screening test, whether based on genetic or other methods, must be subjected to rigorous evaluation of its sensitivity, specificity, impact on disease, and cost-effectiveness. Physicians and patients are introduced continually to new screening tests, often in advance of complete evaluation. For example, the use of whole-body CT imaging has been advocated as a means to screen for a variety of disorders. Though it is appealing in concept, there is currently no evidence to justify this approach, which is associated with high cost and a substantial risk of false-positive results.

This chapter will review the basic principles of screening and prevention in the primary care setting. Recommendations for specific disorders such as cardiovascular disease, diabetes, and cancer are provided in the chapters dedicated to those topics.

Basic Principles of Screening

In general, screening is most effective when applied to relatively common disorders that carry a large disease burden (Table 4-1). The five leading causes of mortality in the United States are heart diseases, malignantneoplasms, accidents, cerebrovascular diseases, and chronic obstructive pulmonary disease. Thus, many prevention strategies are targeted at these conditions. From a global health perspective, these conditions are priorities, but malaria, malnutrition, AIDS, tuberculosis, and violence also carry a heavy disease burden (Chap. 2).

Table 4-1 Lifetime Cumulative Risk

Breast cancer for women	10%
Colon cancer	6%
Cancer of the cervix for women[a]	2%
Domestic violence for women	Up to 15%
Hip fracture for white women	16%

[a]Assuming an unscreened population.

A primary goal of screening is the early detection of a risk factor or disease at a stage at which it can be corrected or cured. For example, most cancers have a better prognosis when identified as premalignant lesions or when they are still resectable. Similarly, early identification of hypertension or hyperlipidemia allows therapeutic interventions that reduce the long-term risk of cardiovascular or cerebrovascular events. However, early detection does not necessarily influence survival. For example, in some studies of lung cancer screening, tumors are identified at an earlier stage but the overall mortality rate does not differ between screened and unscreened populations. The apparent improvement in 5-year survival rates can be attributed to the detection of smaller tumors rather than to a real change in clinical course after diagnosis. Similarly, the detection of prostate cancer may not lead to a difference in the mortality rate because the disease is often indolent and competing morbidities, such as coronary artery disease, may ultimately cause mortality (Chap. 82).

Disorders with a long latency period increase the potential gains associated with detection. For example, cancer of the cervix has a long latency between dysplasia and invasive carcinoma, providing an opportunity for detection by routine screening. It is hoped that the introduction of new

papillomavirus vaccines will provide additional disease prevention, ultimately reducing reliance on screening for cervical cancer. For colon cancer, an adenomatous polyp progresses to invasive cancer over 4–12 years, providing an opportunity to detect early lesions by fecal occult blood testing (FOBT) or endoscopy. In contrast, breast cancer screening in premenopausal women is more challenging —and controversial—because of the relatively short interval between development of a localized breast cancer and metastasis to regional nodes (estimated to be –12 months).

Methods of Measuring Health Benefits

It is not practical to perform all possible screening procedures. For example, screening for laryngeal cancer in smokers is not currently recommended. It is necessary to examine the strength of evidence in favor of screening measures relative to the cost and risk of false-positive tests. For example, should ultrasound be used to screen for ovarian cancer in average-risk women? It is currently estimated that the unnecessary laparotomies triggered by finding benign ovarian masses would cause more harm than the benefit derived from detecting the occasional curable ovarian cancer.

A variety of endpoints are used to assess the potential gain from screening and prevention interventions:

1. *The number of subjects screened to alter the outcome in one individual* . It is estimated, for example, that 731 women ages 65–69 would need to be screened by dual-energy x-ray absorptiometry (DEXA) and then treated appropriately to prevent one hip fracture from osteoporosis.
2. *The absolute and relative impact of screening on disease outcome* . A meta-analysis of Swedish mammography trials (ages 40–70) found that –1.2 fewer women per thousand would die from breast cancer if they were screened over a 12-year period. By comparison, –3 lives per 1000 might be saved from colon cancer in a population (ages 50–75) screened with annual FOBT over a 13-year period. Based on this analysis, colon cancer screening may actually save more womens lives than does mammography. The impact of FOBT (8.8/1000 versus 5.9/1000) might be stated either as 3 lives per 1000 or as a 30% reduction in colon cancer death; thus, it is important to consider both the relative impact and absolute impact on numbers of lives saved.
3. *The cost per year of life saved* is used to assess the effectiveness of many screening and prevention strategies. Typically, strategies that cost <$30,000–50,000 per year of life saved are considered "cost-effective" (Chap. 3). For example, using alendronate to treat 65-year-old women with osteoporosis approaches this threshold of approximately $30,000 per year of life saved.
4. *Increase in average life expectancy for a population* . Predicted increases in life expectancy for various screening procedures are listed in Table 4-2. It should be noted, however, that the increase in life expectancy is an average that applies to a population, not to an individual. In reality, the vast majority of the screened population does not derive any benefit and possibly incurs a slight risk from false-positive results. A small subset of patients, however, will benefit greatly from being screened. For example, Pap smears do not benefit the 98% of women who never develop cancer of the cervix. However, for the 2% who would develop localized cervical cancer, Pap smears may add as much as 25 years to their lives. Some studies suggest that a 1-month gain of life expectancy is a reasonable goal for a population-based preventive strategy.

Table 4-2 Estimated Average Increase in Life Expectancy for a Population

Screening Procedure	Average Increase
Mammography:	
Women, 40–50 years	0–5 days
Women, 50–70 years	1 month
Pap smears, age 18–65	2–3 months
Screening treadmill for a 50-year-old (asymptomatic) man	8 days
PSA and digital rectal exam for a man >50 years	Up to 2 weeks
Getting a 35-year-old smoker to quit	3–5 years
Beginning regular exercise for a 40-year-old man (30 min 3 times a week)	9 months– 2 years

Abbreviation: PSA, prostate-specific antigen.

The U.S. Preventive Services Task Force (USPSTF) provides recommendations for evidence-based screening (Table 4-3). In addition to these population-based guidelines, it is reasonable to consider family and social history to identify individuals with special risk (*www.ahrq.gov/clinic/uspstfix.htm*). For example, when there is a significant family history of breast, colon, or prostate cancer, it is prudent to initiate screening about 10 years before the age at which the youngest family member developed cancer. Screening also should be considered for many other common disorders pending the development of further evidence. Three examples are screening for diabetes (using fasting blood glucose), domestic violence, and coronary artery disease in intermediate-risk asymptomatic

individuals.

Table 4-3 Clinical Preventive Services for Normal-Risk Adults Recommended by the U.S. Preventive Services Task Force

Test or Disorder	Population,[a] Years	Frequency	Chapter Reference
Blood pressure, height and weight	>18	Periodically	77
Cholesterol	Men >35	Every 5 years	241
	Women >45	Every 5 years	
Depression	>18	Periodically[b]	
Diabetes	>45 or earlier, if there are additional risk factors	Every 3 years	344
Pap smear[c]	Within 3 years of onset of sexual activity or 21–65	Every 1–3 years	82
Chlamydia	Women 18–24	Every 1–2 years	176
Mammography[a]	Women >50	Every 2 years	82, 90
Colorectal cancer[a]	>50		82, 91
fecal occult blood and/or		Every year	
sigmoidoscopy or		Every 5 years	
colonoscopy		Every 10 years	
Osteoporosis	Women >65; >60 at risk	Periodically	354
Abdominal aortic aneurysm (ultrasound)	Men 65–75 who have ever smoked	Once	
Alcohol use	>18	Periodically	392
Vision, hearing	>65	Periodically	28, 30
Adult immunization			122, 123
Tetanus-diphtheria (Td)	>18	Every 10 years	
Varicella (VZV)	Susceptibles only, >18	Two doses	
Measles mumps rubella (MMR)	Women, childbearing age	One dose	
Pneumococcal	>65	One dose	
Influenza	>50	Yearly	
Human papillomavirus (HPV)	Up to age 26	If not done prior	

[a]Screening is performed earlier and more frequently when there is a strong family history. Randomized, controlled trials have documented that fecal occult blood testing (FOBT) confers a 15–30% reduction in colon cancer mortality. Although large randomized trials have not been performed for sigmoidoscopy or colonoscopy, well-designed case-control studies suggest similar or greater efficacy relative to FOBT.

[b]If staff support are available.

[c]In the future, Pap smear frequency may be influenced by HPV testing and the HPV vaccine.

Note: Prostate-specific antigen (PSA) testing is capable of enhancing the detection of early-stage prostate cancer, but evidence is inconclusive that it improves health outcomes. PSA testing is recommended by several professional organizations and is widely used in clinical practice, but it is not currently recommended by the U.S. Preventive Services Task Force (Chap. 85).

Source: Adapted from the U.S. Preventive Services Task Force, *Guide to Clinical Prevention*

Services, 2009. http://www.ahrq.gov/clinic/uspstfix.htm.

COST-EFFECTIVENESS

Screening techniques must be cost-effective if they are to be applied to large populations. Costs include not only the expense of testing but also time away from work, downstream costs from false-positive results, and other potential risks. When the risk-versus-benefit ratio is less favorable, it is useful to provide information to patients and factor their perspectives into the decision-making process. For example, many expert groups, including the USPSTF, recommend an individualized discussion about prostate cancer screening, as the decision-making process is complex and relies heavily on personal issues. Although the early detection of prostate cancer may seem desirable intuitively, risks include false-positive results that can lead to anxiety and unnecessary surgery. Randomized trials for prostate cancer screening have yielded mixed and relatively modest results. Potential complications from surgery and radiation treatment include erectile dysfunction, urinary incontinence, and bowel dysfunction. Some men may decline screening, whereas others may be more willing to accept the risks of an early-detection strategy. Another example of shared decision-making involves the choice of techniques for colon cancer screening (Chap. 82). In controlled studies, the use of annual FOBT reduces colon cancer deaths by 15–30%. Flexible sigmoidoscopy reduces colon cancer deaths by –60%. Colonoscopy offers the same benefit as or greater benefit than flexible sigmoidoscopy, but its use incurs additional costs and risks. These screening procedures have not been compared directly in the same population, but the estimated cost to society is similar: $10,000–25,000 per year of life saved. Thus, although one patient may prefer the ease of preparation, less time disruption, and the lower risk of flexible sigmoidoscopy, others may prefer the sedation and thoroughness of colonoscopy.

In considering the impact of screening tests, it is important to recognize that tobacco and alcohol use, diet, and exercise constitute the vast majority of factors that influence preventable deaths in developed countries. Perhaps the single greatest preventive health care measure is to help patients quit smoking (Chap. 395).

Commonly Encountered Issues

Despite compelling evidence that prevention strategies can have major health care benefits, implementation of these services is challenging because of competing demands on physician and patient time and because of gaps in health care reimbursement. Moreover, efforts to reduce disease risk frequently involve behavior changes (e.g., weight loss, exercise, seat belts) or the management of addictive conditions (e.g., tobacco and alcohol use) that are often recalcitrant to intervention. Public education and economic incentives are often useful, in addition to counseling by health care providers (Table 4-4).

Table 4-4 Counseling to Prevent Disease

Topic	Chapter Reference
Tobacco cessation	395
Drug and alcohol use	392, 393
Nutrition to maintain caloric balance and vitamin intake	73
Calcium intake in women >18 years	354
Folic acid: Women of childbearing age	74
Oral health	32
Aspirin use to prevent cardiovascular disease in selected men >45 years and women >55 years	241
Chemoprevention of breast cancer in women at high risk	90
STDs and HIV prevention	130, 189
Physical activity	
Sun exposure	56
Injury prevention (loaded handgun, seat belts, bicycle helmet)	
Issues in the elderly	9
Polypharmacy	
Fall prevention	
Hot water heater <120°	
Vision, hearing, dental evaluations	
Immunizations (pneumococcal, influenza)	

Abbreviation: STDs, sexually transmitted diseases.

A number of techniques can assist physicians with the growing number of recommended screening tests. An appropriately configured electronic health record can provide reminder systems that make it easier for physicians to track and meet guidelines. Some systems give patients secure access to their medical records, providing an additional means to enhance adherence to routine screening. Systems that provide nurses and other staff with standing orders are effective for smoking prevention and immunizations. The Agency for Healthcare Research and Quality and the Centers for Disease Control and Prevention have developed flow sheets and electronic tools as part of their "Put Prevention into Practice" program (*http://www.ahcpr.gov/clinic/ppipix.htm*). Age-specific recommendations for screening and counseling are summarized in Table 4-5.

Table 4-5 Age-Specific Causes of Mortality and Corresponding Preventive Options

Age Group	Leading Causes of Age-Specific Mortality	Screening Prevention Interventions to Consider for Each Specific Population
15–24	1. Accident 2. Homicide 3. Suicide 4. Malignancy 5. Heart disease	▪ Counseling on routine seat belt use, bicycle/motorcycle/ATV helmets (1) ▪ Counseling on diet and exercise (5) ▪ Discuss dangers of alcohol use while driving, swimming, boating (1) ▪ Ask about vaccination status (tetanus, diphtheria, hepatitis B, MMR, rubella, varicella, meningitis, HPV) ▪ Ask about gun use and/or gun possession (2,3) ▪ Assess for substance abuse history including alcohol (2,3) ▪ Screen for domestic violence (2,3) ▪ Screen for depression and/or suicidal/homicidal ideation (2,3) ▪ Pap smear for cervical cancer screening, discuss STD prevention (4) ▪ Discuss skin, breast awareness, and testicular self-exams (4) ▪ Recommend UV light avoidance and regular sunscreen use (4) ▪ Measurement of blood pressure, height, weight, and body mass index (5) ▪ Discuss health risks of tobacco use, consider emphasis on cosmetic and economic issues to improve quit rates for younger smokers (4,5) ▪ *Chlamydia* screening and contraceptive counseling for sexually active females ▪ HIV, hepatitis B, and syphilis testing if there is high-risk sexual behavior(s) or any prior history of sexually transmitted disease
25–44	1. Accident 2. Malignancy 3. Heart disease 4. Suicide 5. Homicide 6. HIV	*As above plus consider the following:* ▪ Readdress smoking status, encourage cessation at every visit (2,3) ▪ Obtain detailed family history of malignancies and begin early screening/prevention program if patient is at significant increased risk (2) ▪ Assess all cardiac risk factors (including screening for diabetes and hyperlipidemia) and consider primary prevention with aspirin for patients at >3% 5-year risk of a vascular event (3) ▪ Assess for chronic alcohol abuse, risk factors for viral hepatitis, or other risks for development of chronic liver disease ▪ Consider individualized breast cancer screening with mammography at age 40 (2)
45–64	1. Malignancy 2. Heart disease 3. Accident 4. Diabetes mellitus 5. Cerebrovascular disease 6. Chronic lower respiratory	▪ Consider prostate cancer screen with annual PSA and digital rectal exam at age 50 (or possibly earlier in African Americans or patients with family history) (1) ▪ Begin colorectal cancer screening at age 50 with fecal occult blood testing, flexible sigmoidoscopy, or colonoscopy (1) ▪ Reassess vaccination status at age 50 and give special consideration to vaccines against *Streptococcus pneumoniae,* influenza, tetanus, and viral hepatitis ▪ Consider screening for coronary disease in higher-risk patients (2,5)

	disease 7. Chronic liver disease and cirrhosis 8. Suicide	
≥65	1. Heart disease 2. Malignancy 3. Cerebrovascular disease 4. Chronic lower respiratory disease 5. Alzheimers disease 6. Influenza and pneumonia 7. Diabetes mellitus 8. Kidney disease 9. Accidents 10. Septicemia	*As above plus consider the following:* ■ Readdress smoking status, encourage cessation at every visit (1,2,3,4) ■ One-time ultrasound for AAA in men 65–75 who have ever smoked ■ Consider pulmonary function testing for all long-term smokers to assess for development of chronic obstructive pulmonary disease (4,6) ■ Vaccinate all smokers against influenza and *S. pneumoniae* at age 50 (6) ■ Screen all postmenopausal women (and all men with risk factors) for osteoporosis ■ Reassess vaccination status at age 65, emphasis on influenza and *S. pneumoniae* (4,6) ■ Screen for dementia and depression (5) ■ Screen for visual and hearing problems, home safety issues, and elder abuse (9)

Note: The numbers in parentheses refer to areas of risk in the mortality column affected by the specified intervention.

Abbreviations: AAA, abdominal aortic aneurysm. ATV, all-terrain vehicle; HPV, human papillomavirus; MMR, measles-mumps-rubella; PSA, prostate-specific antigen; STD, sexually transmitted disease; UV, ultraviolet.

A routine health care examination should be performed every 1–3 years before age 50 and every year thereafter. History should include medication use (prescription and nonprescription), allergies, dietary history, use of alcohol and tobacco, sexual practices, and a thorough family history, if not obtained previously. Routine measurements should include assessments of height, weight (body mass index), and blood pressure, in addition to the relevant physical examination. The increasing incidence of skin cancer underscores the importance of screening for suspicious skin lesions. Hearing and vision should be tested after age 65, or earlier if the patient describes difficulties. Other sex- and age-specific examinations are listed in Table 4-3. Counseling and instruction about self-examination (e.g., skin, breast "awareness") can be provided during the routine examination.

Many patients see a physician for ongoing care of chronic illnesses, and this visit provides an opportunity to include a "measure of prevention" for other health problems. For example, a patient seen for management of hypertension or diabetes can have breast cancer screening incorporated into one visit and a discussion about colon cancer screening at the next visit. Other patients may respond more favorably to a clearly defined visit that addresses all relevant screening and prevention interventions. Because of age or comorbidities, it may be appropriate with some patients to abandon certain screening and prevention activities, although there are fewer data about when to "sunset" these services. The risk of certain cancers, such as cancer of the cervix, ultimately declines, and it is reasonable to cease Pap smears after about age 65 if recent Pap smears have been negative. For breast, colon, and prostate cancer, it is reasonable to reevaluate the need for screening after about age 75. For some older patients with advanced diseases such as severe chronic obstructive pulmonary disease and congestive heart failure and for those who are immobile, the benefit of some screening procedures is low, and other priorities emerge when life expectancy is <10 years. This shift in focus needs to be done tactfully and allows greater focus on the conditions likely to affect quality and length of life.

ACKNOWLEDGMENTS

The author is grateful to Dan Evans, MD, for contributions to this topic in Harrisons Manual of Medicine.

FURTHER READINGS

Barry MJ: Screening for prostate cancer—the controversy that refuses to die. N Engl J Med 360:1351, 2009[PMID: 19297564]

Fenton JJ et al: Delivery of cancer screening: How important is the preventive health examination? Arch Intern Med 167:580, 2007[PMID: 17389289]

Greenland P et al: Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. JAMA 291:210, 2004[PMID: 14722147]

Kerlikowske K: Evidence-based breast cancer prevention: The importance of individualized risk. Ann Intern Med 151:750, 2009[PMID: 19920276]

Ransohoff DF, Sandler RS: Clinical practice: Screening for colorectal cancer. N Engl J Med 346:40, 2002[PMID: 11778002]

U.S. Preventive Services Task Force: The guide to clinical preventive services, 2009. Agency for Healthcare Research and Quality, Rockville, MD, March 2009. Available at http://www.ahrq.gov/clinic/pocketgd.htm

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRINCIPLES OF CLINICAL PHARMACOLOGY: INTRODUCTION

Drugs are the cornerstone of modern therapeutics. Nevertheless, it is well recognized among physicians and in the lay community that the outcome of drug therapy varies widely among individuals. While this variability has been perceived as an unpredictable, and therefore inevitable, accompaniment of drug therapy, this is not the case. The goal of this chapter is to describe the principles of clinical pharmacology that can be used for the safe and optimal use of available and new drugs.

Drugs interact with specific target molecules to produce their beneficial and adverse effects. The chain of events between administration of a drug and production of these effects in the body can be divided into two components, both of which contribute to variability in drug actions. The first component comprises the processes that determine drug delivery to, and removal from, molecular taets. The resulting description of the relationship between drug concentration and time is termed *pharmacokinetics*. The second component of variability in drug action comprises the processes that determine variability in drug actions despite equivalent drug delivery to effector drug sites. This description of the relationship between drug concentration and effect is termed *pharmacodynamics*. As discussed further below, pharmacodynamic variability can arise as a result of variability in function of the target molecule itself or of variability in the broad biologic context in which the drug-target interaction occurs to achieve drug effects.

Two important goals of the discipline of clinical pharmacology are (1) to provide a description of conditions under which drug actions vary among human subjects; and (2) to determine mechanisms underlying this variability, with the goal of improving therapy with available drugs as well as pointing to new drug mechanisms that may be effective in the treatment of human disease. The first steps in the discipline were empirical descriptions of the influence of disease X on drug action Y or of individuals or families with unusual sensitivities to adverse drug effects. These important descriptive findings are now being replaced by an understanding of the molecular mechanisms underlying variability in drug actions. Thus, the effects of disease, drug coadministration, or familial factors in modulating drug action can now be reinterpreted as variability in expression or function of specific genes whose products determine pharmacokinetics and pharmacodynamics. Nevertheless, it is often the personal interaction of the patient with the physician or other health care provider that first identifies unusual variability in drug actions; maintained alertness to unusual drug responses continues to be a key component of improving drug safety.

Unusual drug responses, segregating in families, have been recognized for decades and initially defined the field of *pharmacogenetics*. Now, with an increasing appreciation of common polymorphisms across the human genome, comes the opportunity to reinterpret descriptive mechanisms of variability in drug action as a consequence of specific DNA variants, or sets of variants, among individuals. This approach defines the field of *pharmacogenomics*, which may hold the opportunity of allowing practitioners to integrate a molecular understanding of the basis of disease with an individual's genomic makeup to prescribe personalized, highly effective, and safe therapies.

Indications for Drug Therapy: Risk versus Benefit

It is self-evident that the benefits of drug therapy should outweigh the risks. Benefits fall into two broad categories: those designed to alleviate a symptom and those designed to prolong useful life. An increasing emphasis on the principles of evidence-based medicine and techniques such as large clinical trials and meta-analyses have defined benefits of drug therapy in broad patient populations. Establishing the balance between risk and benefit is not always simple. An increasing body of evidence supports the idea, with which practitioners are very familiar, that individual patients may display responses that are not expected from large population studies and often have comorbidities that typically exclude them from large clinical trials. In addition, therapies that provide symptomatic benefits but shorten life may be entertained in patients with serious and highly symptomatic diseases such as heart failure or cancer. These considerations illustrate the continuing, highly personal nature of the relationship between the prescriber and the patient.

Some adverse effects are so common and so readily associated with drug therapy that they are identified very early during clinical use of a drug. By contrast, serious adverse effects may be sufficiently uncommon that they escape detection for many years after a drug begins to be widely used. The issue of how to identify rare but serious adverse effects (that can profoundly affect the benefit-risk perception in an individual patient) has not been satisfactorily resolved. Potential

approaches range from an increased understanding of the molecular and genetic basis of variability in drug actions to expanded postmarketing surveillance mechanisms. None of these have been completely effective, so practitioners must be continuously vigilant to the possibility that unusual symptoms may be related to specific drugs, or combinations of drugs, that their patients receive.

Beneficial and adverse reactions to drug therapy can be described by a series of dose-response relations (Fig. 5-1). Well-tolerated drugs demonstrate a wide margin, termed the *therapeutic ratio*, *therapeutic index*, or *therapeutic window*, between the doses required to produce a therapeutic effect and those producing toxicity. In cases where there is a similar relationship between plasma drug concentration and effects, monitoring plasma concentrations can be a highly effective aid in managing drug therapy by enabling concentrations to be maintained above the minimum required to produce an effect and below the concentration range likely to produce toxicity. Such monitoring has been widely used to guide therapy with specific agents, such as certain antiarrhythmics, anticonvulsants, and antibiotics. Many of the principles in clinical pharmacology and examples outlined below, which can be applied broadly to therapeutics, have been developed in these arenas.

Figure 5-1

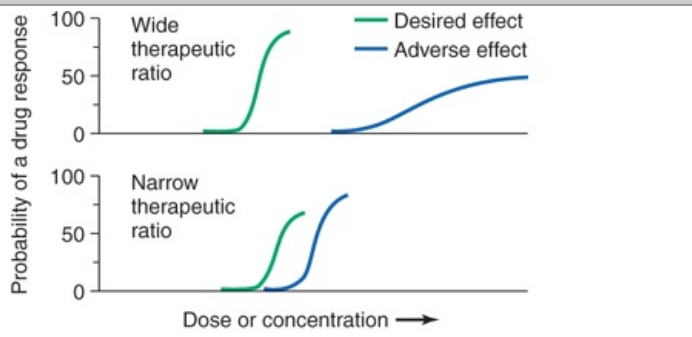

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The concept of a therapeutic ratio. Each panel illustrates the relationship between increasing dose and cumulative probability of a desired or adverse drug effect. ***Top.*** A drug with a wide therapeutic ratio, i.e., a wide separation of the two curves. ***Bottom.*** A drug with a narrow therapeutic ratio; here, the likelihood of adverse effects at therapeutic doses is increased because the curves are not well separated. Further, a steep dose-response curve for adverse effects is especially undesirable, as it implies that even small dosage increments may sharply increase the likelihood of toxicity. When there is a definable relationship between drug concentration (usually measured in plasma) and desirable and adverse effect curves, concentration may be substituted on the abscissa. Note that not all patients necessarily demonstrate a therapeutic response (or adverse effect) at any dose, and that some effects (notably some adverse effects) may occur in a dose-independent fashion.

PRINCIPLES OF PHARMACOKINETICS

The processes of absorption, distribution, metabolism, and excretion—collectively termed *drug disposition*—determine the concentration of drug delivered to target effector molecules.

Absorption

BIOAVAILABILITY

When a drug is administered orally, subcutaneously, intramuscularly, rectally, sublingually, or directly into desired sites of action, the amount of drug actually entering the systemic circulation may be less than with the intravenous route (Fig. 5-2*A*). The fraction of drug available to the systemic circulation by other routes is termed *bioavailability*. Bioavailability may be <100% for two reasons: (1) absorption is reduced, or (2) the drug undergoes metabolism or elimination prior to entering the systemic circulation.

Figure 5-2

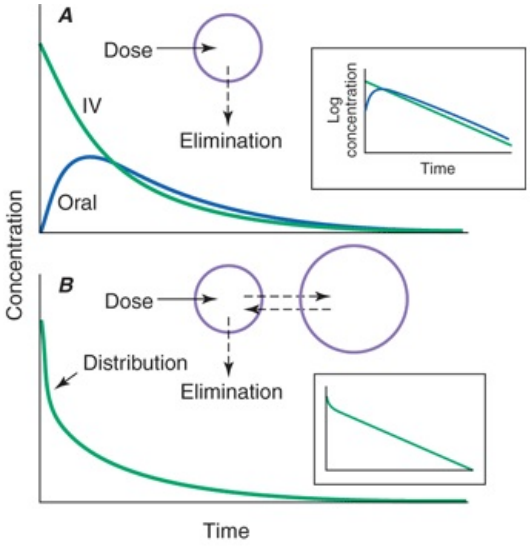

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Idealized time-plasma concentration curves after a single dose of drug.*A.* The time course of drug concentration after an instantaneous IV bolus or an oral dose in the one-compartment model shown. The area under the time-concentration curve is clearly less with the oral drug than the IV, indicating incomplete bioavailability. Note that despite this incomplete bioavailability, concentration after the oral dose can be higher than after the IV dose at some time points. The inset shows that the decline of concentrations over time is linear on a log-linear plot, characteristic of first-order elimination, and that oral and IV drug have the same elimination (parallel) time course. ***B.*** The decline of central compartment concentration when drug is distributed both to and from a peripheral compartment and eliminated from the central compartment. The rapid initial decline of concentration reflects not drug elimination but distribution.

When a drug is administered by a nonintravenous route, the peak concentration occurs later and is lower than after the same dose given by rapid intravenous injection, reflecting absorption from the site of administration (Fig. 5-2). The extent of absorption may be reduced because a drug is incompletely released from its dosage form, undergoes destruction at its site of administration, or has physicochemical properties such as insolubility that prevent complete absorption from its site of administration. Slow absorption rates are deliberately designed into "slow-release" or "sustained-release" drug formulations in order to minimize variation in plasma concentrations during the interval between doses.

"FIRST-PASS" EFFECT

When a drug is administered orally, it must traverse the intestinal epithelium, the portal venous system, and the liver prior to entering the systemic circulation (Fig. 5-3). Once a drug enters the enterocyte, it may undergo metabolism, be transported into the portal vein, or undergo excretion back into the intestinal lumen. Both excretion into the intestinal lumen and metabolism decrease systemic bioavailability. Once a drug passes this enterocyte barrier, it may also be taken up into the hepatocyte, where bioavailability can be further limited by metabolism or excretion into the bile. This elimination in intestine and liver, which reduces the amount of drug delivered to the systemic circulation, is termed *presystemic elimination*, *presystemic extraction*, or *first-pass elimination*.

Figure 5-3

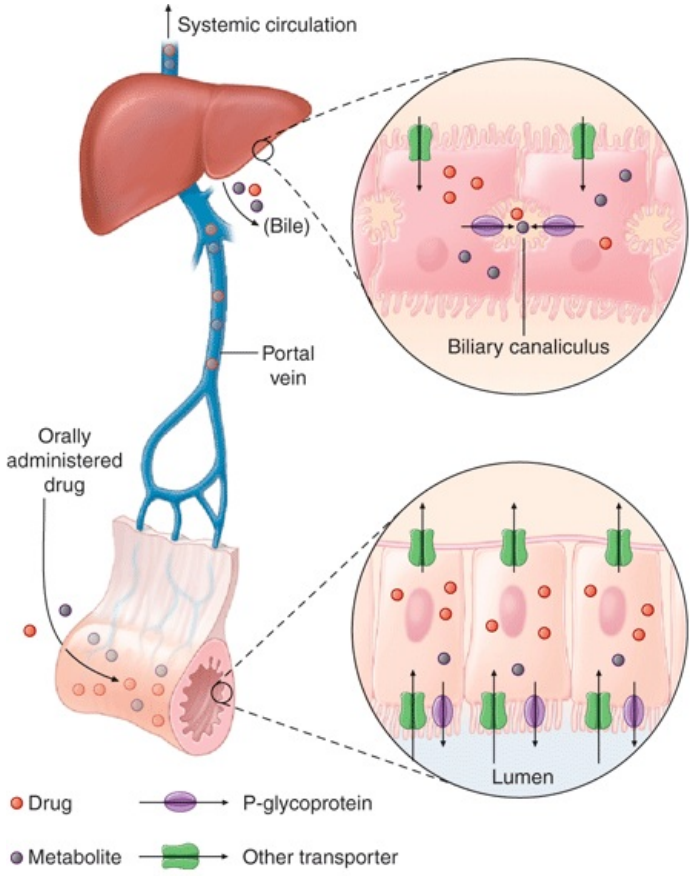

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanism of presystemic clearance. After drug enters the enterocyte, it can undergo metabolism, excretion into the intestinal lumen, or transport into the portal vein. Similarly, the hepatocyte may accomplish metabolism and biliary excretion prior to the entry of drug and metabolites to the systemic circulation. *[Adapted by permission from DM Roden, in DP Zipes, J Jalife (eds): Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.]*

Drug movement across the membrane of any cell, including enterocytes and hepatocytes, is a combination of passive diffusion and active transport, mediated by specific drug uptake and efflux molecules. The drug transport molecule that has been most widely studied is P-glycoprotein, the product of the normal expression of the *MDR1* gene. P-glycoprotein is expressed on the apical aspect of the enterocyte and on the canalicular aspect of the hepatocyte (Fig. 5-3); in both locations, it serves as an efflux pump, thus limiting availability of drug to the systemic circulation. P-glycoprotein is also an important component of the blood-brain barrier, discussed further below.

Drug metabolism generates compounds that are usually more polar and, hence, more readily excreted than parent drug. Metabolism takes place predominantly in the liver but can occur at other sites such as kidney, intestinal epithelium, lung, and plasma. "Phase I" metabolism involves chemical modification, most often oxidation accomplished by members of the cytochrome P450 (CYP) monooxygenase superfamily. CYPs that are especially important for drug metabolism (Table 5–1) include CYP3A4, CYP3A5, CYP2D6, CYP2C9, CYP2C19, CYP1A2, and CYP2E1, and each drug may be a substrate for one or more of these enzymes. "Phase II" metabolism involves conjugation of specific endogenous compounds to drugs or their metabolites. The enzymes that accomplish phase II reactions include glucuronyl-, acetyl-, sulfo-, and methyltransferases. Drug metabolites may exert important pharmacologic activity, as discussed further below.

Table 5–1 Molecular Pathways Mediating Drug Disposition

Molecule	Substrates[a]	Inhibitors[a]
CYP3A	Calcium channel blockers Antiarrhythmics (lidocaine, quinidine, mexiletine) HMG-CoA reductase inhibitors ("statins"; see text)	Amiodarone Ketoconazole, itraconazole Erythromycin, clarithromycin Ritonavir

	Cyclosporine, tacrolimus Indinavir, saquinavir, ritonavir	
CYP2D6[b]	Timolol, metoprolol, carvedilol Phenformin Codeine Propafenone, flecainide Tricyclic antidepressants Fluoxetine, paroxetine	Quinidine (even at ultra-low doses) Tricyclic antidepressants Fluoxetine, paroxetine
CYP2C9[b]	Warfarin Phenytoin Glipizide Losartan	Amiodarone Fluconazole Phenytoin
CYP2C19[b]	Omeprazole Mephenytoin Clopidogrel	Omeprazole
Thiopurine S-methyltransferase[b]	6-Mercaptopurine, azathioprine	
N-acetyltransferase[b]	Isoniazid Procainamide Hydralazine Some sulfonamides	
UGT1A1[b]	Irinotecan	
Pseudocholinesterase[b]	Succinylcholine	
P-glycoprotein	Digoxin HIV-protease inhibitors Many CYP3A substrates	Quinidine Amiodarone Verapamil Cyclosporine Itraconazole Erythromycin

[a]Inhibitors affect the molecular pathway, and thus may affect substrate.

[b]Clinically important genetic variants described; see Table 5–2.

Note: A listing of CYP substrates, inhibitors, and inducers is maintained at http://medicine.iupui.edu/flockhart/table.htm.

CLINICAL IMPLICATIONS OF ALTERED BIOAVAILABILITY

Some drugs undergo near-complete presystemic metabolism and, thus, cannot be administered orally. Nitroglycerin cannot be used orally because it is completely extracted prior to reaching the systemic circulation. The drug is, therefore, used by the sublingual or transdermal routes, which bypass presystemic metabolism.

Some drugs with very extensive presystemic metabolism can still be administered by the oral route, using much higher doses than those required intravenously. Thus, a typical intravenous dose of verapamil is 1–5 mg, compared to the usual single oral dose of 40–120 mg. Administration of low-dose aspirin can result in exposure of cyclooxygenase in platelets in the portal vein to the drug, but systemic sparing because of first-pass aspirin deacylation in the liver. This is an example of presystemic metabolism being exploited to therapeutic advantage.

Distribution and Elimination

Most pharmacokinetic processes are first-order; that is, the rate of the process depends on the amount

of drug present. Clinically important exceptions are discussed below (see "Principles of Dose Selection"). In the simplest pharmacokinetic model (Fig. 5-2*A*), a drug bolus (D) is administered instantaneously to a central compartment, from which drug elimination occurs as a first-order process. Occasionally, central and other compartments correspond to physiologic spaces (e.g., plasma volume), whereas in others they are simply mathematical functions used to describe drug disposition. The first-order nature of drug elimination leads directly to the relationship describing drug concentration (*C*) at any time (*t*) following the bolus:

$$C = \frac{D}{V_c} \; e^{(-0.69t/t_{1/2})}$$

where V_C is the volume of the compartment into which drug is delivered and $t_{1/2}$ is elimination half-life. As a consequence of this relationship, a plot of the logarithm of concentration vs time is a straight line (Fig. 5-2*A*, inset). *Half-life* is the time required for 50% of a first-order process to be complete. Thus, 50% of drug elimination is achieved after one drug-elimination half-life, 75% after two, 87.5% after three, etc. In practice, first-order processes such as elimination are near-complete after four–five half-lives.

In some cases, drug is removed from the central compartment not only by elimination but also by distribution into peripheral compartments. In this case, the plot of plasma concentration vs time after a bolus may demonstrate two (or more) exponential components (Fig. 5-2*B*). In general, the initial rapid drop in drug concentration represents not elimination but drug distribution into and out of peripheral tissues (also first-order processes), while the slower component represents drug elimination; the initial precipitous decline is usually evident with administration by intravenous but not other routes. Drug concentrations at peripheral sites are determined by a balance between drug distribution to and redistribution from those sites, as well as by elimination. Once distribution is near-complete (four–five distribution half-lives), plasma and tissue concentrations decline in parallel.

CLINICAL IMPLICATIONS OF HALF-LIFE MEASUREMENTS

The elimination half-life not only determines the time required for drug concentrations to fall to near-immeasurable levels after a single bolus, it is also the key determinant of the time required for steady-state plasma concentrations to be achieved after any change in drug dosing (Fig. 5-4). This applies to the initiation of chronic drug therapy (whether by multiple oral doses or by continuous intravenous infusion), a change in chronic drug dose or dosing interval, or discontinuation of drug.

Figure 5-4

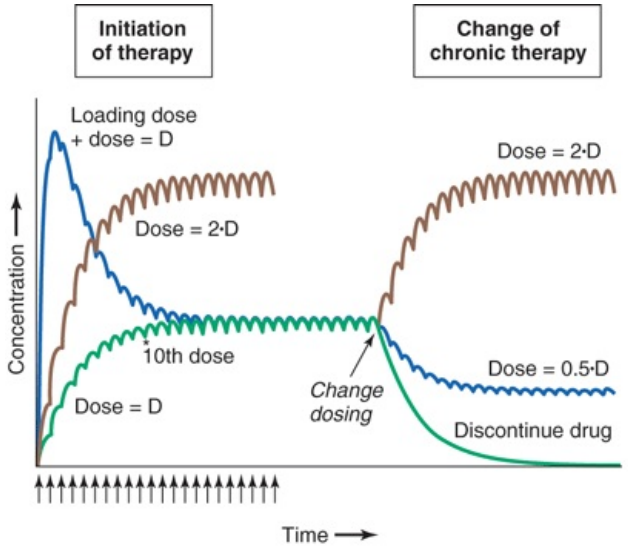

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Drug accumulation to steady state. In this simulation, drug was administered (arrows) at intervals = 50% of the elimination half-life. Steady state is achieved during initiation of therapy after ~5 elimination half-lives, or 10 doses. A loading dose did not alter the eventual steady state achieved. A doubling of the dose resulted in a doubling of the steady state but the same time course of accumulation. Once steady state is achieved, a change in dose (increase, decrease, or drug discontinuation) results in a new steady state in ~5 elimination half-lives. *[Adapted by permission from DM Roden, in DP Zipes, J Jalife (eds): Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.]*

Steady state describes the situation during chronic drug administration when the amount of drug administered per unit time equals drug eliminated per unit time. With a continuous intravenous infusion,

plasma concentrations at steady state are stable, while with chronic oral drug administration, plasma concentrations vary during the dosing interval but the time-concentration profile between dosing intervals is stable (Fig. 5-4).

Drug Distribution

In a typical 70-kg human, plasma volume is ~3 L, blood volume is ~5.5 L, and extracellular water outside the vasculature is ~20 L. The volume of distribution of drugs extensively bound to plasma proteins but not to tissue components approaches plasma volume; warfarin is one such example. By contrast, for drugs highly bound to tissues, the volume of distribution can be far greater than any physiologic space. For example, the volume of distribution of digoxin and tricyclic antidepressants is hundreds of liters, obviously exceeding total-body volume. Such drugs are not readily removed by dialysis, an important consideration in overdose.

CLINICAL IMPLICATIONS OF DRUG DISTRIBUTION

In some cases, pharmacologic effects require drug distribution to peripheral sites. In this instance, the time course of drug delivery to and removal from these sites determines the time course of drug effects. Digoxin accesses its cardiac site of action slowly, over a distribution phase of several hours. Thus, after an intravenous dose, plasma levels fall, but those at the site of action increase over hours. Only when distribution is near-complete does the concentration of digoxin in plasma reflect the pharmacologic effect. For this reason, there should be a 6–8 h wait after administration before plasma levels of digoxin are measured as a guide to therapy. Similarly, anesthetic drug penetration into and removal from the central nervous system (CNS) determines the time course of anesthesia.

Animal models have suggested, and clinical studies are confirming, that limited drug penetration into the brain, the "blood-brain barrier," often represents a robust P-glycoprotein–mediated efflux process from capillary endothelial cells in the cerebral circulation. Thus, drug distribution into the brain may be modulated by changes in P-glycoprotein function.

Loading Doses

For some drugs, the indication may be so urgent that administration of "loading" dosages is required to achieve rapid elevations of drug concentration and therapeutic effects earlier than with chronic maintenance therapy (Fig. 5-4). Nevertheless, the time required for true steady state to be achieved is still determined only by the elimination half-life.

Disease can alter loading requirements: in congestive heart failure, the central volume of distribution of lidocaine is reduced. Therefore, lower-than-normal loading regimens are required to achieve equivalent plasma drug concentrations and to avoid toxicity.

Rate of Intravenous Administration

Although the simulations in Fig. 5-2 use a single intravenous bolus, this is very rarely appropriate in practice because side effects related to transiently very high concentrations can result. Rather, drugs are more usually administered orally or as a slower intravenous infusion. Some drugs are so predictably lethal when infused too rapidly that special precautions should be taken to prevent accidental boluses. For example, solutions of potassium for intravenous administration >20 meq/L should be avoided in all but the most exceptional and carefully monitored circumstances. This minimizes the possibility of cardiac arrest due to accidental increases in infusion rates of more concentrated solutions.

While excessively rapid intravenous drug administration can lead to catastrophic consequences, transiently high drug concentrations after intravenous administration can occasionally be used to advantage. The use of midazolam for intravenous sedation, for example, depends upon its rapid uptake by the brain during the distribution phase to produce sedation quickly, with subsequent egress from the brain during the redistribution of the drug as equilibrium is achieved.

Similarly, adenosine must be administered as a rapid bolus in the treatment of reentrant supraventricular tachycardias (Chap. 233) to prevent elimination by very rapid ($t_{1/2}$ of seconds) uptake into erythrocytes and endothelial cells before the drug can reach its clinical site of action, the atrioventricular node.

Plasma Protein Binding

Many drugs circulate in the plasma partly bound to plasma proteins. Since only unbound (free) drug can distribute to sites of pharmacologic action, drug response is related to the free rather than the total circulating plasma drug concentration.

CLINICAL IMPLICATIONS OF ALTERED PROTEIN BINDING

For drugs that are normally highly bound to plasma proteins (>90%), small changes in the extent of binding (e.g., due to disease) can produce a large change in the amount of unbound drug, and hence drug effect. The acute-phase reactant α_1-acid glycoprotein binds to basic drugs, such as lidocaine or quinidine, and is increased in a range of common conditions, including myocardial infarction, surgery, neoplastic disease, rheumatoid arthritis, and burns. This increased binding can lead to reduced

pharmacologic effects at therapeutic concentrations of total drug. Conversely, conditions such as hypoalbuminemia, liver disease, and renal disease can decrease the extent of drug binding, particularly of acidic and neutral drugs, such as phenytoin. Here, plasma concentration of free drug is increased, so drug efficacy and toxicity are enhanced if total (free + bound) drug concentration is used to monitor therapy.

Clearance

Drug elimination reduces the amount of drug in the body over time. An important approach to quantifying this reduction is to consider that drug concentration at the beginning and end of a time period are unchanged and that a specific volume of the body has been "cleared" of the drug during that time period. This defines clearance as volume/time. Clearance includes both drug metabolism and excretion.

CLINICAL IMPLICATIONS OF ALTERED CLEARANCE

While elimination half-life determines the time required to achieve steady-state plasma concentrations (C_{SS}), the *magnitude* of that steady state is determined by clearance (Cl) and dose alone. For a drug administered as an intravenous infusion, this relationship is:

$$C_{ss} = \text{dosing rate}/Cl \quad \text{or} \quad \text{dosing rate} = Cl \cdot C_{ss}$$

When drug is administered orally, the average plasma concentration within a dosing interval ($C_{avg,ss}$) replaces C_{SS}, and the dosage (dose per unit time) must be increased if bioavailability (F) is less than 1:

$$\text{Dose/time} = Cl \cdot C_{avg,ss}/F$$

Genetic variants, drug interactions, or diseases that reduce the activity of drug-metabolizing enzymes or excretory mechanisms may lead to decreased clearance and, hence, a requirement for downward dose adjustment to avoid toxicity. Conversely, some drug interactions and genetic variants increase the function of drug elimination pathways, and, hence, increased drug dosage may be necessary to maintain a therapeutic effect.

Active Drug Metabolites

From an evolutionary point of view, drug metabolism may have developed as a defense against noxious xenobiotics (foreign substances, e.g., from plants) to which our ancestors inadvertently exposed themselves. The organization of the drug uptake and efflux pumps and the location of drug metabolism in the intestine and liver prior to drug entry to the systemic circulation (Fig. 5-3) support this idea of a primitive protective function.

However, drug metabolites are not necessarily pharmacologically inactive. Metabolites may produce effects similar to, overlapping with, or distinct from those of the parent drug. For example, *N*-acetylprocainamide (NAPA) is a major metabolite of the antiarrhythmic procainamide. While it exerts antiarrhythmic effects, its electrophysiologic properties differ from those of the parent drug. Indeed, NAPA accumulation is the usual explanation for marked QT prolongation and torsades des pointes ventricular tachycardia (Chap. 233) during therapy with procainamide. Thus, the common laboratory practice of adding procainamide to NAPA concentrations to estimate a total therapeutic effect is inappropriate.

Prodrugs are inactive compounds that require metabolism to generate active metabolites that mediate the drug effects. Examples include many angiotensin-converting enzyme (ACE) inhibitors, the angiotensin receptor blocker losartan, the antineoplastic irinotecan, the anti-estrogen tamoxifen, the analgesic codeine (whose active metabolite morphine probably underlies the opioid effect during codeine administration), and the antiplatelet drug clopidogrel. Drug metabolism has also been implicated in bioactivation of procarcinogens and in generation of reactive metabolites that mediate certain adverse drug effects (e.g., acetaminophen hepatotoxicity, discussed below).

The Concept of High-Risk Pharmacokinetics

When plasma concentrations of active drug depend exclusively on a single metabolic pathway, any condition that inhibits that pathway (be it disease-related, genetic, or due to a drug interaction) can lead to dramatic changes in drug concentrations and marked variability in drug action. This problem of high-risk pharmacokinetics is especially pronounced in two settings. *First*, variability in bioactivation of a prodrug can lead to striking variability in drug action; examples include decreased CYP2D6 activity, which prevents analgesia by codeine, and decreased CYP2C19 activity, which reduces the antiplatelet effects of clopidogrel. The s*econd* setting is drug elimination that relies on a single pathway. In this case, inhibition of the elimination pathway leads to striking elevation of drug concentration. For drugs with a narrow therapeutic window, this leads to an increased likelihood of dose-related toxicity. An example is digoxin, whose elimination is dependent on P-glycoprotein; many drugs inhibit P-glycoprotein activity (amiodarone, quinidine, erythromycin, cyclosporine, itraconazole) and coadministration of these with digoxin reduces digoxin clearance, and increases toxicity unless maintenance doses are lowered. When drugs undergo elimination by multiple-drug metabolizing or

excretory pathways, absence of one pathway (due to a genetic variant or drug interaction) is much less likely to have a large impact on drug concentrations or drug actions.

PRINCIPLES OF PHARMACODYNAMICS

Once a drug accesses a molecular site of action, it alters the function of that molecular target, with the ultimate result of a drug effect that the patient or health care provider can perceive. For drugs used in the urgent treatment of acute symptoms, little or no delay is anticipated (or desired) between the drug-target interaction and the development of a clinical effect. Examples of such acute situations include vascular thrombosis, shock, malignanthypertension, or status epilepticus.

For many conditions, however, the indication for therapy is less urgent, and a delay between the interaction of a drug with its pharmacologic target(s) and a clinical effect is clinically acceptable. Pharmacokinetic mechanisms that can contribute to such a delay include slow elimination (resulting in slow accumulation to steady state), uptake into peripheral compartments, or accumulation of active metabolites. Another common explanation for such a delay is that the clinical effect develops as a downstream consequence of the initial molecular effect the drug produces. Thus, administration of a proton-pump inhibitor or an H_2-receptor blocker produces an immediate increase in gastric pH but ulcer healing that is delayed. Cancer chemotherapy very frequently produces delayed therapeutic effects, often long after drug is undetectable in plasma and tissue. Translation of a molecular drug action to a clinical effect can thus be highly complex and dependent on the details of the pathologic state being treated. These complexities have made pharmacodynamics and its variability less amenable than pharmacokinetics to rigorous mathematical analysis. Nevertheless, some clinically important principles can be described.

A drug effect often depends on the presence of underlying pathophysiology. Thus, a drug may produce no action or a different spectrum of actions in unaffected individuals compared to patients. Further, concomitant disease can complicate interpretation of response to drug therapy, especially adverse effects. For example, high doses of anticonvulsants such as phenytoin may cause neurologic symptoms, which may be confused with the underlying neurologic disease. Similarly, increasing dyspnea in a patient with chronic lung disease receiving amiodarone therapy could be due to drug, underlying disease, or an intercurrent cardiopulmonary problem. Thus, the presence of chronic lung disease may alter the risk-benefit ratio in a specific patient to argue against the use of amiodarone.

While drugs interact with specific molecular receptors, drug effects may vary over time, even if stable drug and metabolite concentrations are maintained. The drug-receptor interaction occurs in a complex biologic milieu that can vary to modulate the drug effect. For example, ion channel blockade by drugs, an important anticonvulsant and antiarrhythmic effect, is often modulated by membrane potential, itself a function of factors such as extracellular potassium or local ischemia. Receptors may be up- or downregulated by disease or by the drug itself. For example, ß-adrenergic blockers upregulate ß-receptor density during chronic therapy. While this effect does not usually result in resistance to the therapeutic effect of the drugs, it may produce severe agonist–mediated effects (such as hypertension or tachycardia) if the blocking drug is abruptly withdrawn.

PRINCIPLES OF DOSE SELECTION

The desired goal of therapy with any drug is to maximize the likelihood of a beneficial effect while minimizing the risk of adverse effects. Previous experience with the drug, in controlled clinical trials or in postmarketing use, defines the relationships between dose (or plasma concentration) and these dual effects and provides a starting point for initiation of drug therapy.

Figure 5-1 illustrates the relationships among dose, plasma concentrations, efficacy, and adverse effects and carries with it several important implications:

1. *The target drug effect should be defined when drug treatment is started* . With some drugs, the desired effect may be difficult to measure objectively, or the onset of efficacy can be delayed for weeks or months; drugs used in the treatment of cancer and psychiatric disease are examples. Sometimes a drug is used to treat a symptom, such as pain or palpitations, and here it is the patient who will report whether the selected dose is effective. In yet other settings, such as anticoagulation or hypertension, the desired response can be repeatedly and objectively assessed by simple clinical or laboratory tests.
2. *The nature of anticipated toxicity often dictates the starting dose* . If side effects are minor, it may be acceptable to start at a dose highly likely to achieve efficacy and downtitrate if side effects occur. However, this approach is rarely, if ever, justified if the anticipated toxicity is serious or life-threatening; in this circumstance, it is more appropriate to initiate therapy with the lowest dose that may produce a desired effect.
3. *The above considerations do not apply if these relationships between dose and effects cannot be defined*. This is especially relevant to some adverse drug effects (discussed in further detail below) whose development are not readily related to drug dose.
4. *If a drug dose does not achieve its desired effect, a dosage increase is justified only if toxicity is absent and the likelihood of serious toxicity is small*. For example, some patients with

seizures require plasma levels of phenytoin >20 μg/mL for optimal anticonvulsant activity. Dosages to achieve this effect may be appropriate, if tolerated. Conversely, clinical experience with flecainide suggests that high dosages (e.g., >400 mg/d) may be associated with an increased risk of sudden death; thus dosage increases beyond this limit are ordinarily not appropriate, even if a higher dosage might seem tolerated.

Other mechanisms that can lead to failure of drug effect should also be considered; drug interactions and noncompliance are common examples. These are situations in which measurement of plasma drug concentrations, if available, can be especially useful. Noncompliance is an especially frequent problem in the long-term treatment of diseases such as hypertension and epilepsy, occurring in ≥25% of patients in therapeutic environments in which no special effort is made to involve patients in the responsibility for their own health. Multidrug regimens with multiple doses per day are especially prone to noncompliance.

Monitoring response to therapy, by physiologic measures or by plasma concentration measurements, requires an understanding of the relationships between plasma concentration and anticipated effects. For example, measurement of QT interval is used during treatment with sotalol or dofetilide to avoid marked QT prolongation that can herald serious arrhythmias. In this setting, evaluating the electrocardiogram at the time of anticipated peak plasma concentration and effect (e.g., 1–2 h postdose at steady state) is most appropriate. Maintained high aminoglycoside levels carry a risk of nephrotoxicity, so dosages should be adjusted on the basis of plasma concentrations measured at trough (predose). On the other hand, ensuring aminoglycoside efficacy is accomplished by adjusting dosage so that peak drug concentrations are above a minimal antibacterial concentration. For dose adjustment of other drugs (e.g., anticonvulsants), concentration should be measured at its lowest during the dosing interval, just prior to a dose at steady state (Fig. 5-4), to ensure a maintained therapeutic effect.

Concentration of Drugs in Plasma as a Guide to Therapy

Factors such as interactions with other drugs, disease-induced alterations in elimination and distribution, and genetic variation in drug disposition combine to yield a wide range of plasma levels in patients given the same dose. Hence, if a predictable relationship can be established between plasma drug concentration and beneficial or adverse drug effect, measurement of plasma levels can provide a valuable tool to guide selection of an optimal dose. This is particularly true when there is a narrow range between the plasma levels yielding therapeutic and adverse effects, as with digoxin, theophylline, some antiarrhythmics, aminoglycosides, cyclosporine, and anticonvulsants. By contrast, if no such relationship can be established (e.g., if drug access to important sites of action outside plasma is highly variable), monitoring plasma concentration may not provide an accurate guide to therapy (Fig. 5-5A).

Figure 5-5

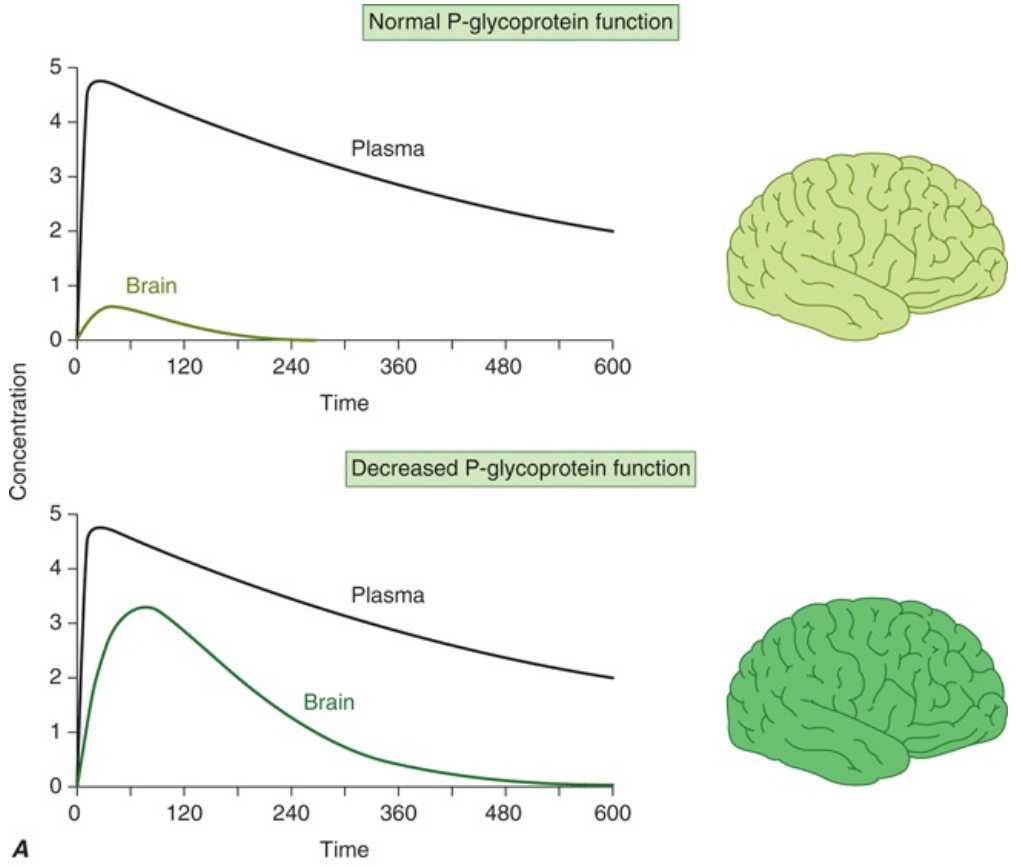

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

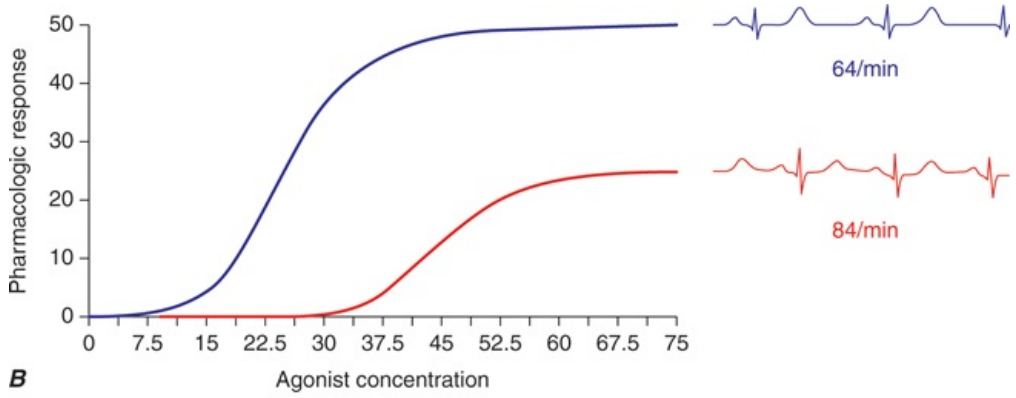

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. The efflux pump P-glycoprotein excludes drugs from the endothelium of capillaries in the brain, and so constitutes a key element of the blood-brain barrier. Thus, reduced P-glycoprotein function (e.g., due to drug interactions or genetically determined variability in gene transcription) increases penetration of substrate drugs into the brain, even when plasma concentrations are unchanged. ***B***. The graph shows an effect of a β_1-receptor polymorphism on receptor function in vitro. Patients with the hypofunctional variant (red) may display lesser heart-rate slowing or blood pressure lowering on exposure to a receptor blocking agent.

The common situation of first-order elimination implies that average, maximum, and minimum steady-state concentrations are related linearly to the dosing rate. Accordingly, the maintenance dose may be adjusted on the basis of the ratio between the desired and measured concentrations *at steady state*; for example, if a doubling of the steady-state plasma concentration is desired, the dose should be doubled. In some cases, elimination becomes saturated at high doses, and the process then occurs at a fixed amount per unit time (zero order). For drugs with this property (e.g., phenytoin and theophylline), plasma concentrations change disproportionately more than the alteration in the dosing rate. In this situation, changes in dose should be small to minimize the degree of unpredictability, and plasma concentration monitoring should be used when available to ensure that dose modification achieves the desired level.

An increase in dosage is usually best achieved by changing the drug dose but not the dosing interval,

e.g., by giving 200 mg every 8 h instead of 100 mg every 8 h. However, this approach is acceptable only if the resulting maximum concentration is not toxic and the trough value does not fall below the minimum effective concentration for an undesirable period of time. Alternatively, the steady state may be changed by altering the frequency of intermittent dosing but not the size of each dose. In this case, the magnitude of the fluctuations around the average steady-state level will change—the shorter the dosing interval, the smaller the difference between peak and trough levels.

EFFECTS OF DISEASE ON DRUG CONCENTRATION AND RESPONSE

Renal Disease

Renal excretion of parent drug and metabolites is generally accomplished by glomerular filtration and by specific drug transporters, only now being identified. If a drug or its metabolites are primarily excreted through the kidneys and increased drug levels are associated with adverse effects, drug dosages must be reduced in patients with renal dysfunction to avoid toxicity. The antiarrhythmics dofetilide and sotalol undergo predominant renal excretion and carry a risk of QT prolongation and arrhythmias if doses are not reduced in renal disease. Thus, in end-stage renal disease, sotalol can be given as 40 mg after dialysis (every second day), compared to the usual daily dose, 80–120 mg every 12 h. The narcotic analgesic meperidine undergoes extensive hepatic metabolism, so that renal failure has little effect on its plasma concentration. However, its metabolite, normeperidine, does undergo renal excretion, accumulates in renal failure, and probably accounts for the signs of CNS excitation, such as irritability, twitching, and seizures, that appear when multiple doses of meperidine are administered to patients with renal disease. Protein binding of some drugs (e.g., phenytoin) may be altered in uremia, so measuring free drug concentration may be desirable.

In non-end-stage renal disease, changes in renal drug clearance are generally proportional to those in creatinine clearance, which may be measured directly or estimated from the serum creatinine (Chap. 278). This estimate, coupled with the knowledge of how much drug is normally excreted renally vs nonrenally, allows an estimate of the dose adjustment required. In practice, most decisions involving dosing adjustment in patients with renal failure use published recommended adjustments in dosage or dosing interval based on the severity of renal dysfunction indicated by creatinine clearance. Any such modification of dose is a first approximation and should be followed by plasma concentration data (if available) and clinical observation to further optimize therapy for the individual patient.

Liver Disease

In contrast to the predictable decline in renal clearance of drugs in renal insufficiency, the effects of diseases like hepatitis or cirrhosis on drug disposition range from impaired to increased drug clearance in an unpredictable fashion. Standard tests of liver function are not useful in adjusting doses. First-pass metabolism may decrease, leading to increased oral bioavailability as a consequence of disrupted hepatocyte function, altered liver architecture, and portacaval shunts. The oral bioavailability for high first-pass drugs such as morphine, meperidine, midazolam, and nifedipine is almost doubled in patients with cirrhosis, compared to those with normal liver function. Therefore, the size of the oral dose of such drugs should be reduced in this setting.

Heart Failure and Shock

Under conditions of decreased tissue perfusion, the cardiac output is redistributed to preserve blood flow to the heart and brain at the expense of other tissues (Chap. 234). As a result, drugs may be distributed into a smaller volume of distribution, higher drug concentrations will be present in the plasma, and the tissues that are best perfused (the brain and heart) will be exposed to these higher concentrations. If either the brain or heart is sensitive to the drug, an alteration in response will occur. As well, decreased perfusion of the kidney and liver may impair drug clearance. Another consequence of severe heart failure is decreased gut perfusion, which may reduce drug absorption and, thus, lead to reduced or absent effects of orally administered therapies.

Drug Use in the Elderly

In the elderly, multiple pathologies and medications used to treat them result in more drug interactions and adverse effects. Aging also results in changes in organ function, especially of the organs involved in drug disposition. Initial doses should be less than the usual adult dosage and should be increased slowly. The number of medications, and doses per day, should be kept as low as possible.

Even in the absence of kidney disease, renal clearance may be reduced by 35–50% in elderly patients. Dosage adjustments are therefore necessary for drugs that are eliminated mainly by the kidneys. Because muscle mass and therefore creatinine production are reduced in older individuals, a normal serum creatinine concentration can be present even though creatinine clearance is impaired; dosages should be adjusted on the basis of creatinine clearance, as discussed above. Aging also results in a decrease in the size of, and blood flow to, the liver and possibly in the activity of hepatic drug-metabolizing enzymes; accordingly, the hepatic clearance of some drugs is impaired in the elderly. As with liver disease, these changes are not readily predicted.

Elderly patients may display altered drug sensitivity. Examples include increased analgesic effects of

opioids, increased sedation from benzodiazepines and other CNS depressants, and increased risk of bleeding while receiving anticoagulant therapy, even when clotting parameters are well controlled. Exaggerated responses to cardiovascular drugs are also common because of the impaired responsiveness of normal homeostatic mechanisms. Conversely, the elderly display decreased sensitivity to β-adrenergic receptor blockers.

Adverse drug reactions are especially common in the elderly because of altered pharmacokinetics and pharmacodynamics, the frequent use of multidrug regimens, and concomitant disease. For example, use of long half-life benzodiazepines is linked to the occurrence of hip fractures in elderly patients, perhaps reflecting both a risk of falls from these drugs (due to increased sedation) and the increased incidence of osteoporosis in elderly patients. In population surveys of the noninstitutionalized elderly, as many as 10% had at least one adverse drug reaction in the previous year.

GENETIC DETERMINANTS OF THE RESPONSE TO DRUGS

Principles of Genetic Variation and Human Traits

(See also Chaps. 61 and 63) The concept that genetically determined variations in drug metabolism might be associated with variable drug levels and hence, effect, was advanced at the end of the nineteenth century, and the examples of familial clustering of unusual drug responses were noted in the mid-twentieth century. Variants in the human genome resulting in variation in level of expression or function of molecules important for pharmacokinetics and pharmacodynamics are increasingly recognized. These may be mutations (very rare variants, often associated with disease) or polymorphisms, variants that are much more common in a population. Variants may occur at a single nucleotide [single nucleotide polymorphisms (SNPs)] or involve insertion or deletion of one or more nucleotides, occasionally up to thousands. They may be in the exons (coding regions), introns (noncoding intervening sequences), or intergenic regions. Exonic polymorphisms may or may not alter the encoded protein, and variant proteins may or may not display altered function. Similarly, polymorphisms in noncoding regions may or may not alter gene expression and protein level.

As variation in the human genome is increasingly well documented, associations are being described between polymorphisms and various traits (including response to drug therapy). Some of these rely on well-developed chains of evidence, including in vitro studies demonstrating variant protein function, familial aggregation of the variant allele with the trait, and association studies in large populations. In other cases, the associations are less compelling. Identifying replicated associations with important clinical consequences is a challenge that must be overcome before the concept of genotyping to identify optimal drugs (or dosages) in individual patients prior to prescribing can be considered for widespread clinical practice.

Rates of drug efficacy and adverse effects often vary among ethnic groups. Many explanations for such differences are plausible; genomic approaches have now established one mechanism that functionally important variants determining differences in drug response often display differing distributions among ethnic groups. This finding may have importance for drug use among ethnic groups, as well as in drug development.

APPROACHES TO IDENTIFYING GENETIC VARIANTS MODULATING DRUG ACTION

A goal of traditional Mendelian genetics is to identify DNA variants associated with a distinct phenotype in multiple related family members (Chap. 63). The usual approach, linkage analysis, does not generally lend itself to identifying genetic variants contributing to variable drug actions, because it is unusual for a drug response phenotype to be accurately measured in more than one family member, let alone across a kindred. Thus, alternate approaches are used to identify and validate DNA variants contributing to variable drug actions.

Most studies to date have used an understanding of the molecular mechanisms modulating drug action to identify candidate genes in which variants could explain variable drug responses. One very common scenario is that variable drug actions can be attributed to variability in plasma drug concentrations. When plasma drug concentrations vary widely (e.g., more than an order of magnitude), especially if their distribution is nonunimodal as in Fig. 5-6, variants in single genes controlling drug concentrations often contribute. In this case, the most obvious candidate genes are those responsible for drug metabolism and elimination. Other candidate genes are those encoding the target molecules with which drugs interact to produce their effects or molecules modulating that response, including those involved in disease pathogenesis.

Figure 5-6

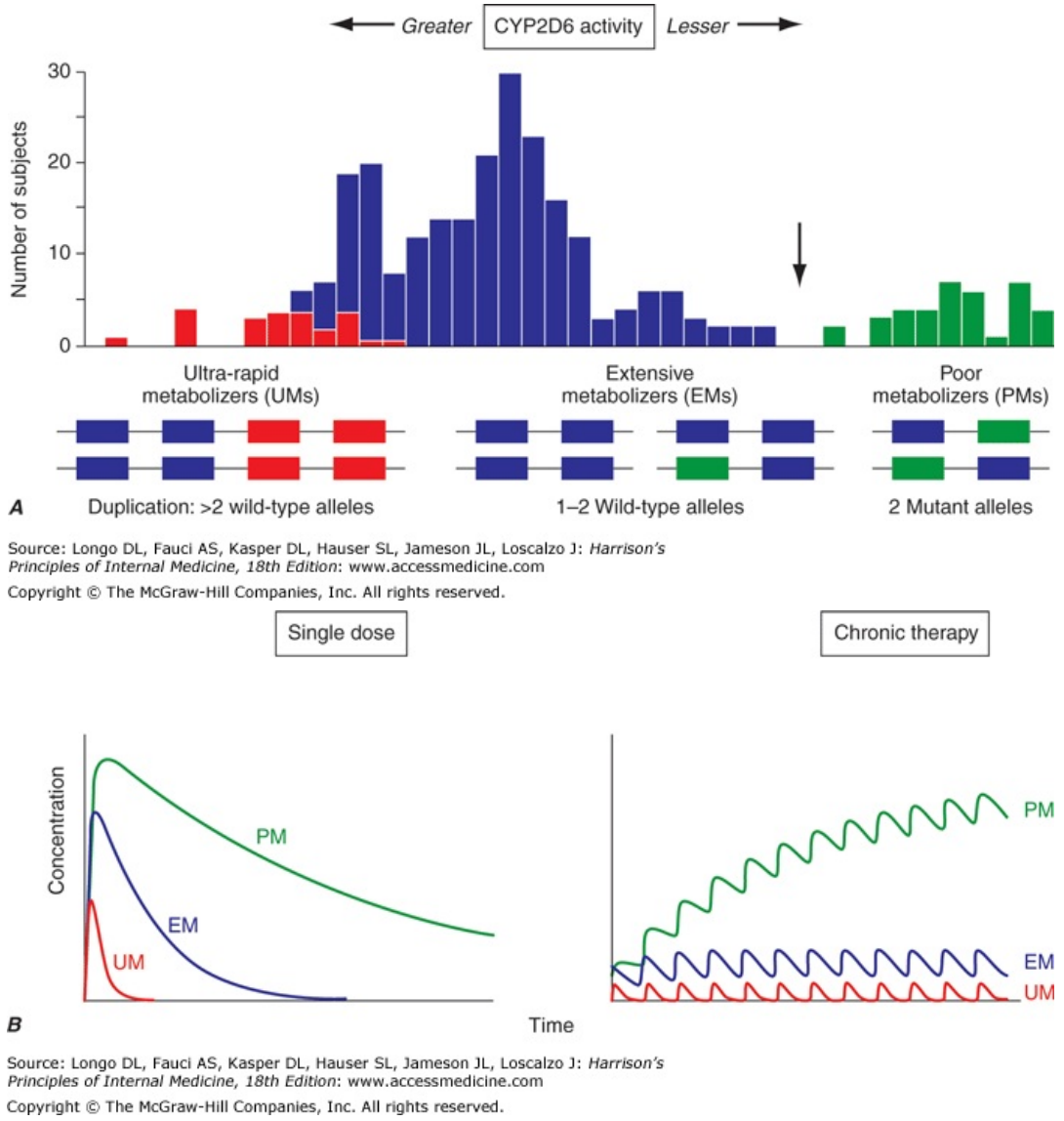

A. CYP2D6 metabolic activity was assessed in 290 subjects by administration of a test dose of a probe substrate and measurement of urinary formation of the CYP2D6-generated metabolite. The heavy arrow indicates a clear antimode, separating poor metabolizer subjects (PMs, green), with two loss-of-function CYP2D6 alleles, indicated by the intron-exon structures below the bar chart. Individuals with one or two functional alleles are grouped together as extensive metabolizers (EMs, blue). Also shown are ultra-rapid metabolizers (UMs), with 2–12 functional copies of the gene (red), displaying the greatest enzyme activity. *(Adapted by permission from M-L Dahl et al: J Pharmacol Exp Ther 274:516, 1995.)* ***B.*** These simulations show the predicted effects of CYP2D6 genotype on disposi-tion of a substrate drug. With a single dose (*left*), there is an inverse "gene-dose" relationship between the number of active alleles and the areas under the time-concentration curves (smallest in UM subjects; highest in PM subjects); this indicates that clearance is greatest in UM subjects. In addition, elimination half-life is longest in PM subjects. The right panel shows that these single dose differences are exaggerated during chronic therapy: steady-state concentration is much higher in PM subjects (decreased clearance), as is the time required to achieve steady state (longer elimination half-life).

The field has also had some success with "unbiased" approaches such as genome-wide association (GWA) (Chap. 61). GWA makes no *a priori* assumptions about the genetic loci modulating variable drug response and, instead, searches across the whole genome in an "unbiased fashion" to identify loci linked to variable drug response.

Genetically Determined Drug Disposition and Variable Effects

Clinically important genetic variants have been described in multiple molecular pathways of drug disposition (Table 5–2). A distinct multimodal distribution of drug disposition (as shown in Fig. 5-6) argues for a predominant effect of variants in a single gene in the metabolism of that substrate. Individuals with two alleles (variants) encoding for nonfunctional protein make up one group, often termed *poor metabolizers* (PM phenotype); many variants can produce such a loss of function, complicating the use of genotyping in clinical practice. Individuals with one functional allele make up a second (*intermediate metabolizers*) and may or may not be distinguishable from those with two functional alleles (*extensive metabolizers*, EMs). *Ultra-rapid metabolizers* with especially high

enzymatic activity (occasionally due to gene duplication; Fig. 5-6) have also been described for some traits. Many drugs in widespread use can inhibit specific drug disposition pathways (Table 5–1), and so EM individuals receiving such inhibitors can respond like PM patients (*phenocopying*). Polymorphisms in genes encoding drug uptake or drug efflux transporters may be other contributors to variability in drug delivery to target sites and, hence, in drug effects.

Table 5–2 Genetic Variants and Drug Responses

Gene	Drugs	Effect of genetic variants*
Variants in drug metabolism pathways		
CYP2C9	Losartan	Decreased bioactivation and effects (PMs)
	Warfarin	Decreased dose requirements; possible increased bleeding risk (PMs)
CYP2C19	Omeprazole, voriconazole	Decreased effect in extensive metabolizers (EMs)
	Celecoxib	Exaggerated effect in PMs
	Clopidogrel	Decreased effect in PMs
CYP2D6	Codeine, tamoxifen	Decreased bioactivation and drug effects in PMs
	Codeine	Morphine-like adverse effects in UMs
	Tricyclic antidepressants	Increased adverse effects in PMs; decreased therapeutic effects in UMs
	Metoprolol, carvedilol, timolol, propafenone	Increased beta blockade in PMs
Dihydropyrimidine dehydrogenase	Capecitabine, fluorouracil	Possible severe toxicity (PMs)
NAT2	Rifampin, isoniazid, pyrazinamide, hydralazine, procainamide	Increased risk of toxicity in PMs
Thiopurine S-methyltransferase (*TPMT*)	Azathioprine, 6-mercaptopurine	*3A/*3A (PMs): increased risk of bone marrow aplasia; wild-type homozygote: possible decreased drug action at usual dosages
Uridine diphosphate glucuronosyltransferase (*UGT1A1*)	Irinotecan	*28/*28 PM homozygotes: increased risk of severe adverse effects (diarrhea, bone marrow aplasia)
Variants in other genes		
Glucose 6-phosphate dehydrogenase (G6PD)	Rasburicase, primaquine, chloroquine	Increased risk of hemolytic anemia in G6PD-deficient subjects
HLA-B*1501	Carbamazepine	Carriers (1 or 2 alleles) at increased risk of severe skin toxicity
HLA-B*5701	Abacavir	Carriers (1 or 2 alleles) at increased risk of severe skin toxicity
IL28B	Interferon	Variable response in hepatitis C therapy
IL15	Childhood leukemia therapy	Variability in response
SLCO1B1	Simvastatin	Variant non-synonymous single nucleotide polymorphism increases myopathy risk
VKORC1	Warfarin	Decreased dose requirements with variant promoter haplotype
Variants in other genomes (infectious agents, tumors)		
Chemokine C-C motif receptor (CCR5)	Maraviroc	Drug effective only in HIV strains with CCR5 detectible
C-KIT	Imatinib	In gastrointestinal stromal tumors, drug indicated only with c-kit– positive cases
Epidermal Growth Factor Receptor (EGFR)	Cetuximab	Clinical trials conducted in patients with EGFR-positive tumors
Her2/neu overexpression	Trastuzumab, lapatinib	Drugs indicated only with tumor overexpression
K-ras mutation	Panitumumab, cetuximab	Lack of efficacy with KRAS mutation
Philadelphia	Busulfan, dasatinib,	Decreased efficacy in Philadelphia chromosome

chromosome	nilotinib, imatinib	–negative chronic myelogenous leukemia

*Drug effect in homozygotes unless otherwise specified.

Note: PM, poor metabolizer (homozygote for reduced or loss of function allele); EM, extensive metabolizer: normal enzymatic activity; UM, ultra-rapid metabolizer (enzymatic activity much greater than normal, e.g., with gene duplication, Fig. 5-6). Further data at U.S. Food and Drug Administration:

http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm or Pharmacogenetics Research Network/Knowledge Base:

http://www.pharmgkb.org.

CYP VARIANTS

CYP3A4 is the most abundant hepatic and intestinal CYP and is also the enzyme responsible for metabolism of the greatest number of drugs in therapeutic use. CYP3A4 activity is highly variable (up to an order of magnitude) among individuals, but the underlying mechanisms are not yet well understood. A closely related gene, encoding CYP3A5 (which shares substrates with CYP3A4), does display loss-of-function variants, especially in African populations. CYP3A refers to both enzymes.

CYP2D6 is second to CYP3A4 in the number of commonly used drugs that it metabolizes. CYP2D6 activity is polymorphically distributed, with about 7% of European- and African-derived populations (but very few Asians) displaying the PM phenotype (Fig. 5-6). Dozens of loss-of-function variants in the CYP2D6 gene have been described; the PM phenotype arises in individuals with two such alleles. In addition, ultra-rapid metabolizers with multiple functional copies of the CYP2D6 gene have been identified, particularly in Ethiopian, Eritrean, and Saudi individuals.

Codeine is biotransformed by CYP2D6 to the potent active metabolite morphine, so its effects are blunted in PMs and exaggerated in ultra-rapid metabolizers. In the case of drugs with beta-blocking properties metabolized by CYP2D6, greater signs of beta blockade (e.g., bradycardia) are seen in PM subjects than in EMs. This can be seen not only with orally administered beta blockers such as metoprolol and carvedilol, but also with ophthalmic timolol and with the sodium channel–blocking antiarrhythmic propafenone, a CYP2D6 substrate with beta-blocking properties. Further, in EM subjects, propafenone elimination becomes zero-order at higher doses; so, for example, a tripling of the dose may lead to a tenfold increase in drug concentration. Ultra-rapid metabolizers may require very high dosages of tricyclic antidepressants to achieve a therapeutic effect and, with codeine, may display transient euphoria and nausea due to very rapid generation of morphine. Tamoxifen undergoes CYP2D6-mediated biotransformation to an active metabolite, so its efficacy may be in part related to this polymorphism. In addition, the widespread use of selective serotonin reuptake inhibitors (SSRIs) to treat tamoxifen-related hot flashes may also alter the drug's effects because many SSRIs, notably fluoxetine and paroxetine, are also CYP2D6 inhibitors.

The PM phenotype for CYP2C19 is common (20%) among Asians and rarer (3–5%) in European-derived populations. The impact of polymorphic CYP2C19-mediated metabolism has been demonstrated with the proton pump inhibitor omeprazole, where ulcer cure rates with "standard" dosages were markedly lower in EM patients (29%) than in PMs (100%). Thus, understanding the importance of this polymorphism would have been important in developing the drug, and knowing a patient's *CYP2C19* genotype should improve therapy. CYP2C19 is responsible for bioactivation of the antiplatelet drug clopidogrel, and several large studies have documented decreased efficacy (e.g., increased myocardial infarction after placement of coronary stents) among Caucasian subjects with reduction of function alleles. In addition, some studies suggest that omeprazole and possibly other proton inhibitors phenocopy this effect.

There are common allelic variants of *CYP2C9* that encode proteins with loss of catalytic function. These variant alleles are associated with increased rates of neurologic complications with phenytoin, hypoglycemia with glipizide, and reduced warfarin dose required to maintain stable anticoagulation (discussed further below). The angiotensin-receptor blocker losartan is a prodrug that is bioactivated by CYP2C9; as a result, PMs and those receiving inhibitor drugs may display little response to therapy.

TRANSFERASE VARIANTS

One of the most extensively studied phase II polymorphisms is the PM trait for thiopurine S-methyltransferase (TPMT). TPMT bioinactivates the antileukemic drug 6-mercaptopurine. Further, 6-mercaptopurine is itself an active metabolite of the immunosuppressive azathioprine. Homozygotes for alleles encoding the inactive TPMT (1 in 300 individuals) predictably exhibit severe and potentially fatal pancytopenia on standard doses of azathioprine or 6-mercaptopurine. On the other hand, homozygotes for fully functional alleles may display less anti-inflammatory or antileukemic effect with the drugs.

N-acetylation is catalyzed by hepatic *N*-acetyl transferase (NAT), which represents the activity of two genes, *NAT-1* and *NAT-2*. Both enzymes transfer an acetyl group from acetyl coenzyme A to the drug; NAT-1 activity is generally constant, while polymorphisms in *NAT-2* result in individual differences in the rate at which drugs are acetylated and thus define "rapid acetylators" and "slow acetylators." Slow

acetylators make up ~50% of European- and African-derived populations but are less common among Asians.

Slow acetylators have an increased incidence of the drug-induced lupus syndrome during procainamide and hydralazine therapy and of hepatitis with isoniazid. Induction of CYPs (e.g., by rifampin) also increases the risk of isoniazid-related hepatitis, likely reflecting generation of reactive metabolites of acetylhydrazine, itself an isoniazid metabolite.

Individuals homozygous for a common promoter polymorphism that reduces transcription of uridine diphosphate glucuronosyltransferase (*UGT1A1*) have benign hyperbilirubinemia (Gilbert's syndrome; Chap. 302). This variant has also been associated with diarrhea and increased bone marrow depression with the antineoplastic prodrug irinotecan, whose active metabolite is normally detoxified by UGT1A1-mediated glucuronidation.

Variability in the Molecular Targets with Which Drugs Interact

Multiple polymorphisms identified in the β_2-adrenergic receptor appear to be linked to specific phenotypes in asthma and congestive heart failure, diseases in which β_2-receptor function might be expected to determine prognosis. Polymorphisms in the β_2-receptor gene have also been associated with response to inhaled β_2-receptor agonists, while those in the β_1-adrenergic receptor gene have been associated with variability in heart rate slowing and blood pressure lowering (Fig. 5-5*B*). In addition, in heart failure, a common polymorphism in the β_1-adrenergic receptor gene has been implicated in variable clinical outcome during therapy with the investigational beta blocker bucindolol. Response to the 5-lipoxygenase inhibitor zileuton in asthma has been linked to polymorphisms that determine the expression level of the 5-lipoxygenase gene.

Drugs may also interact with genetic pathways of disease to elicit or exacerbate symptoms of the underlying conditions. In the porphyrias, CYP inducers are thought to increase the activity of enzymes proximal to the deficient enzyme, exacerbating or triggering attacks (Chap. 358). Deficiency of glucose-6-phosphate dehydrogenase (G6PD), most often in individuals of African, Mediterranean, or South Asian descent, increases risk of hemolytic anemia in response to primaquine and a number of other drugs that do not cause hemolysis in patients with normal amounts of the enzyme (Chap. 106). Patients with mutations in the ryanodine receptor, which controls intracellular calcium in skeletal muscle and other tissues, may be asymptomatic until exposed to certain general anesthetics, which trigger the syndrome of malignant hyperthermia. Certain antiarrhythmics and other drugs can produce marked QT prolongation and torsades des pointes (Chap. 233), and in some patients this adverse effect represents unmasking of previously subclinical congenital long QT syndrome.

TUMOR AND INFECTIOUS AGENT GENOMES

The actions of drugs used to treat infectious or neoplastic disease may be modulated by variants in these non-human germline genomes. Genotyping tumors is a rapidly evolving approach to target therapies to underlying mechanisms and to avoid potentially toxic therapy in patients who would derive no benefit (Chap. 83). Trastuzumab, which potentiates anthracycline-related cardiotoxicity, is ineffective in breast cancers that do not express the herceptin receptor. Imatinib targets a specific tyrosine kinase, BCR-Abl1, that is generated by the translocation that creates the Philadelphia chromosome typical of chronic myelogenous leukemia (CML). BCR-Abl1 is not only active but may be central to the pathogenesis of CML; its use in BCR-Abl1-positive tumors has resulted in remarkable antitumor efficacy. Similarly, the antiepidermal growth factor (EGFR) antibodies cetuximab and panitumumab appear especially effective in colon cancers in which K-ras, a G-protein in the EGFR pathway, is not mutated.

Polymorphisms that Modulate the Biologic Context Within Which the Drug-Target Interactions Occur

The interaction of a drug with its molecular target is translated into a clinical action in a complex biologic milieu that is itself often perturbed by disease. Thus, polymorphisms that determine variability in this biology may profoundly influence drug response, although the genes involved are not themselves directly targets of drug action. Polymorphisms in genes important for lipid homeostasis (such as the ABCA1 transporter and the cholesterol ester transport protein) modulate response to 3-hydroxymethylglutaryl-CoA (HMG-CoA) reductase inhibitors, "statins." In one large study, the combination of diuretic use combined with a variant in the adducin gene (encoding a cytoskeletal protein important for renal tubular sodium absorption) decreased stroke or myocardial infarction risk, while neither factor alone had an effect. Common polymorphisms in ion channel genes that are not themselves the target of QT-prolonging drugs may nevertheless influence the extent to which those drugs affect the electrocardiogram and produce arrhythmias. Severe skin rashes during treatment with the anticonvulsant carbamazepine and the antiretroviral abacavir have been associated with variants in the HLA gene cluster (Table 5–2).

Multiple Variants Modulating Drug Effects

Polymorphisms in multiple candidate genes can be associated with variability in the effect of a single

drug. CYP2C9 loss-of-function variants are associated with a requirement for lower maintenance doses of the vitamin K antagonist anticoagulant warfarin. In rarer (<2%) individuals homozygous for these variant alleles, maintenance warfarin dosages may be difficult to establish, and the risk of bleeding complications appears increased. In addition to *CYP2C9*, multiple variants in the promoter region of *VKORC1*, encoding a vitamin K epoxide reductase (the warfarin target), predict warfarin dosages; these promoter variants are in tight *linkage disequilibrium*, i.e., genotyping at one polymorphic site within this *haplotype block* provides reliable information on the identity of genotypes at other linked sites (Chap. 61).

Genome-Wide Association and Variable Drug Response

A GWA study was used to compare patients with simvastatin-associated myopathy to control tolerating the drugs and identified a single noncoding SNP in *SLCO1B1*, encoding OATP1B1, a drug transporter known to modulate simvastatin uptake into the liver. The SNP was in linkage disequilibrium with a known nonsynonymous SNP modulating OATP1B1 function and was estimated to account for 60% of myopathy risk. GWA approaches have also implicated interferon variants in antileukemic responses and in response to therapy in hepatitis C (Table 5–2).

Prospects for Incorporating Pharmacogenetic Information into Clinical Practice

The description of genetic variants linked to variable drug responses naturally raises the question of if and how to use this information in practice. Indeed, the U.S. Food and Drug Administration has begun incorporating pharmacogenetic data into information ("package inserts") meant to guide prescribing. A decision to adopt pharmacogenetically guided dosing for a given drug depends on multiple factors. The most important are the magnitude and clinical importance of the genetic effect and the strength of evidence linking genetic variation to variable drug effects (e.g., anecdote versus post-hoc analysis of clinical trial data versus randomized prospective clinical trial). The evidence can be strengthened if statistical arguments from clinical trial data are complemented by an understanding of underlying physiologic mechanisms. Cost versus expected benefit may also be a factor.

When the evidence is compelling and alternate therapies are not available, there is a strong argument for deploying genetic testing as a guide to prescribing. Examples include associations between *UGT1A1* variants and irinotecan toxicity, or between HLA-B*5701 and severe skin toxicity with abacavir. In other situations, the arguments are less compelling: the magnitude of the genetic effect may be smaller, the consequences may be less serious, alternate therapies may be available, or the drug effect may be amenable to monitoring by other approaches. Ongoing clinical trials are addressing the utility of preprescription genotyping in large populations exposed to drugs with known pharmacogenetic variants (e.g., warfarin). Importantly, technological advances are now raising the possibility of inexpensive whole genome sequencing. Incorporating a patient's whole genome sequence into their electronic medical record would allow the information to be accessed as needed for many genetic and pharmacogenetic applications. There are multiple issues (e.g., economic, technological, and ethical) that need to be addressed if such a paradigm is to be adopted (Chap. 61). While barriers to bringing genomic and pharmacogenomic information to the bedside seem daunting, the field is very young and evolving rapidly. Indeed, one major result of understanding the role of genetics in drug action has been improved screening of drugs during the development process to reduce the likelihood of highly variable metabolism or unanticipated toxicity.

INTERACTIONS BETWEEN DRUGS

Drug interactions can complicate therapy by increasing or decreasing the action of a drug; interactions may be based on changes in drug disposition or in drug response in the absence of changes in drug levels. *Interactions must be considered in the differential diagnosis of any unusual response occurring during drug therapy*. Prescribers should recognize that patients often come to them with a legacy of drugs acquired during previous medical experiences, often with multiple physicians who may not be aware of all the patient's medications. A meticulous drug history should include examination of the patient's medications and, if necessary, calls to the pharmacist to identify prescriptions. It should also address the use of agents not often volunteered during questioning, such as over-the-counter (OTC) drugs, health food supplements, and topical agents such as eye drops. Lists of interactions are available from a number of electronic sources. While it is unrealistic to expect the practicing physician to memorize these, certain drugs consistently run the risk of generating interactions, often by inhibiting or inducing specific drug elimination pathways. Examples are presented below and in Table 5–3. Accordingly, when these drugs are started or stopped, prescribers must be especially alert to the possibility of interactions.

Table 5–3 Drugs with a High Risk of Generating Pharmacokinetic Interactions

Drug	Mechanism	Examples
Antacids	Reduced absorption	Antacids/tetracyclines
Bile acid		Cholestyramine/digoxin

sequestrants		
Proton pump inhibitors H_2-receptor blockers	Altered gastric pH	Ketoconazole absorption decreased
Rifampin Carbamazepine Barbiturates Phenytoin St. John's wort Glutethimide	Induction of hepatic metabolism	Decreased concentration and effects of warfarin quinidine cyclosporine losartan oral contraceptives methadone
Tricyclic antidepressants Fluoxetine Quinidine	Inhibitors of CYP2D6	Increased effect of many β blockers Decreased codeine effect
Cimetidine	Inhibitor of multiple CYPs	Increased concentration and effects of warfarin theophylline phenytoin
Ketoconazole, itraconazole Erythromycin, clarithromycin Calcium channel blockers Ritonavir	Inhibitor of CYP3A	Increased concentration and toxicity of some HMG-CoA reductase inhibitors cyclosporine, cisapride, terfenadine (now withdrawn) Increased concentration and effects of indinavir (with ritonavir) Decreased clearance and dose requirement for cyclosporine (with calcium channel blockers)
Allopurinol	Xanthine oxidase inhibitor	Azathioprine and 6-mercaptopurine toxicity
Amiodarone	Inhibitor of many CYPs and of P-glycoprotein	Decreased clearance (risk of toxicity) for warfarin digoxin quinidine
Gemfibrazol (and other fibrates)	CYP3A inhibition	Rhabdomyolysis when co-prescribed with some HMG-CoA reductase inhibitors
Quinidine Amiodarone Verapamil Cyclosporine Itraconazole Erythromycin	P-glycoprotein inhibition	Risk of digoxin toxicity
Phenylbutazone Probenecid Salicylates	Inhibition of renal tubular transport	Increased risk of meth-otrexate toxicity with salicylates

Pharmacokinetic Interactions Causing Decreased Drug Effects

Gastrointestinal absorption can be reduced if a drug interaction results in drug binding in the gut, as with aluminum-containing antacids, kaolin-pectin suspensions, or bile acid sequestrants. Drugs such as histamine H_2 receptor antagonists or proton pump inhibitors that alter gastric pH may decrease the

solubility and hence absorption of weak bases such as ketoconazole.

Expression of some genes responsible for drug elimination, notably *CYP3A* and *MDR1*, can be markedly increased by "inducing" drugs, such as rifampin, carbamazepine, phenytoin, St. John's wort, and glutethimide, and by smoking, exposure to chlorinated insecticides such as DDT (CYP1A2), and chronic alcohol ingestion. Administration of inducing agents lowers plasma levels over 2–3 weeks as gene expression is increased. If a drug dose is stabilized in the presence of an inducer that is subsequently stopped, major toxicity can occur as clearance returns to preinduction levels and drug concentrations rise. Individuals vary in the extent to which drug metabolism can be induced, likely through genetic mechanisms.

Interactions that inhibit the bioactivation of prodrugs will similarly decrease drug effects. The analgesic effect of codeine depends on its metabolism to morphine via CYP2D6. Thus, CYP2D6 inhibitors (Table 5–1) are predicted to reduce the analgesic efficacy of codeine in EMs. Similarly, omeprazole, and perhaps other proton pump inhibitors, reduce CYP2C19 activity and have been associated with reduced clopidogrel efficacy.

Interactions that decrease drug delivery to intracellular sites of action can decrease drug effects: tricyclic antidepressants can blunt the antihypertensive effect of clonidine by decreasing its uptake into adrenergic neurons. Reduced CNS penetration of multiple HIVprotease inhibitors (with the attendant risk of facilitating viral replication in a sanctuary site) appears attributable to P-glycoprotein-mediated exclusion of the drug from the CNS; indeed, inhibition of P-glycoprotein has been proposed as a therapeutic approach to enhance drug entry to the CNS (Fig. 5-5*A*).

Pharmacokinetic Interactions Causing Increased Drug Effects

The most common mechanism here is inhibition of drug elimination. In contrast to induction, new protein synthesis is not involved, and the effect develops as drug and any inhibitor metabolites accumulate (a function of their elimination half-lives). Since shared substrates of a single enzyme can compete for access to the active site of the protein, many CYP substrates can also be considered inhibitors. However, some drugs are especially potent as inhibitors (and occasionally may not even be substrates) of specific drug elimination pathways, and so it is in the use of these agents that clinicians must be most alert to the potential for interactions (Table 5–3). Commonly implicated interacting drugs of this type include amiodarone, cimetidine, erythromycin and some other macrolide antibiotics (clarithromycin but not azithromycin), ketoconazole and other azole antifungals, the antiretroviral agent ritonavir, and high concentrations of grapefruit juice (Table 5–3). The consequences of such interactions will depend on the drug whose elimination is being inhibited; high-risk drugs are those for which alternate pathways of elimination are not available and for which drug accumulation increases the risk of serious toxicity (see "The Concept of High-Risk Pharmacokinetics," above). Examples include CYP3A inhibitors increasing the risk of cyclosporine toxicity or of rhabdomyolysis with some HMG-CoA reductase inhibitors (lovastatin, simvastatin, atorvastatin), and P-glycoprotein inhibitors increasing risk of digoxin toxicity.

These interactions can be exploited to therapeutic benefit. The antiviral ritonavir is a very potent CYP3A4 inhibitor that is sometimes added to anti-HIV regimens, not because of its antiviral effects but because it decreases clearance, and hence increases efficacy, of other anti-HIV agents. Similarly, calcium channel blockers have been deliberately coadministered with cyclosporine to reduce its clearance and thus its maintenance dosage and cost.

Phenytoin, an inducer of many systems, including CYP3A, inhibits CYP2C9. CYP2C9 metabolism of losartan to its active metabolite is inhibited by phenytoin, with potential loss of antihypertensive effect.

Grapefruit (but not orange) juice inhibits CYP3A, especially at high doses; patients receiving drugs where even modest CYP3A inhibition may increase the risk of adverse effects (e.g., cyclosporine, some HMG-CoA reductase inhibitors) should therefore avoid grapefruit juice.

CYP2D6 is markedly inhibited by quinidine, a number of neuroleptic drugs (chlorpromazine and haloperidol), and the SSRIs fluoxetine and paroxetine. Clinical consequences of fluoxetine's interaction with CYP2D6 substrates may not be apparent for weeks after the drug is started, because of its very long half-life and slow generation of a CYP2D6-inhibiting metabolite.

6-Mercaptopurine is metabolized not only by TPMT but also by xanthine oxidase. When allopurinol, a potent inhibitor of xanthine oxidase, is administered with standard doses of azathioprine or 6-mercaptopurine, life-threatening toxicity (bone marrow suppression) can result.

A number of drugs are secreted by the renal tubular transport systems for organic anions. Inhibition of these systems can cause excessive drug accumulation. Salicylate, for example, reduces the renal clearance of methotrexate, an interaction that may lead to methotrexate toxicity. Renal tubular secretion contributes substantially to the elimination of penicillin, which can be inhibited (to increase its therapeutic effect) by probenecid. Similarly, inhibition of the tubular cation transport system by cimetidine decreases the renal clearance of dofetilide and of procainamide and its active metabolite NAPA.

Drug Interactions Not Mediated by Changes in Drug Disposition

Drugs may act on separate components of a common process to generate effects greater than either has alone. Antithrombotic therapy with combinations of antiplatelet agents (glycoprotein IIb/IIIa inhibitors, aspirin, clopidogrel) and anticoagulants (warfarin, heparins) are often used in the treatment of vascular disease, although such combinations carry an increased risk of bleeding.

Nonsteroidal anti-inflammatory drugs (NSAIDs) cause gastric ulcers, and in patients treated with warfarin, the risk of bleeding from a peptic ulcer is increased almost threefold by concomitant use of an NSAID.

Indomethacin, piroxicam, and probably other NSAIDs antagonize the antihypertensive effects of β-adrenergic receptor blockers, diuretics, ACE inhibitors, and other drugs. The resulting elevation in blood pressure ranges from trivial to severe. This effect is not seen with aspirin and sulindac but has been found with the cyclooxygenase 2 (COX-2) inhibitor celecoxib.

Torsades des pointes ventricular tachycardia during administration of QT-prolonging antiarrhythmics (quinidine, sotalol, dofetilide) occurs much more frequently in patients receiving diuretics, probably reflecting hypokalemia. In vitro, hypokalemia not only prolongs the QT interval in the absence of drug but also potentiates drug block of ion channels that results in QT prolongation. Also, some diuretics have direct electrophysiologic actions that prolong QT.

The administration of supplemental potassium leads to more frequent and more severe hyperkalemia when potassium elimination is reduced by concurrent treatment with ACE inhibitors, spironolactone, amiloride, or triamterene.

The pharmacologic effects of sildenafil result from inhibition of the phosphodiesterase type 5 isoform that inactivates cyclic GMP in the vasculature. Nitroglycerin and related nitrates used to treat angina produce vasodilation by elevating cyclic GMP. Thus, coadministration of these nitrates with sildenafil can cause profound hypotension, which can be catastrophic in patients with coronary disease.

Sometimes, combining drugs can increase overall efficacy and/or reduce drug-specific toxicity. Such therapeutically useful interactions are described in chapters dealing with specific disease entities.

ADVERSE REACTIONS TO DRUGS

The beneficial effects of drugs are coupled with the inescapable risk of untoward effects. The morbidity and mortality from these adverse effects often present diagnostic problems because they can involve every organ and system of the body and may be mistaken for signs of underlying disease. As well, some surveys have suggested that drug therapy for a range of chronic conditions such as psychiatric disease or hypertension does not achieve its desired goal in up to half of treated patients; thus, the most common "adverse" drug effect may be failure of efficacy.

Adverse reactions can be classified in two broad groups. One type results from exaggeration of an intended pharmacologic action of the drug, such as increased bleeding with anticoagulants or bone marrow suppression with antineoplastics. The second type of adverse reaction ensues from toxic effects unrelated to the intended pharmacologic actions. The latter effects are often unanticipated (especially with new drugs) and frequently severe and may result from recognized as well as previously undescribed mechanisms.

Drugs may increase the frequency of an event that is common in a general population, and this may be especially difficult to recognize; an excellent example is the increase in myocardial infarctions with the COX-2 inhibitor rofecoxib. Drugs can also cause rare and serious adverse effects, such as hematologic abnormalities, arrhythmias, severe skin reactions, or hepatic or renal dysfunction. Prior to regulatory approval and marketing, new drugs are tested in relatively few patients who tend to be less sick and to have fewer concomitant diseases than those patients who subsequently receive the drug therapeutically. Because of the relatively small number of patients studied in clinical trials and the selected nature of these patients, rare adverse effects are generally not detected prior to a drug's approval; indeed, if they are detected, the new drugs are generally not approved. Therefore, physicians need to be cautious in the prescription of new drugs and alert for the appearance of previously unrecognized adverse events.

Elucidating mechanisms underlying adverse drug effects can assist development of safer compounds or allow a patient subset at especially high risk to be excluded from drug exposure. National adverse reaction reporting systems, such as those operated by the FDA (suspected adverse reactions can be reported online at *http://www.fda.gov/medwatch/report/hcp.htm*) and the Committee on Safety of Medicines in Great Britain, can prove useful. The publication or reporting of a newly recognized adverse reaction can in a short time stimulate many similar such reports of reactions that previously had gone unrecognized.

Occasionally, "adverse" effects may be exploited to develop an entirely new indication for a drug. Unwanted hair growth during minoxidil treatment of severely hypertensive patients led to development of the drug for hair growth. Sildenafil was initially developed as an antianginal, but its effects to alleviate erectile dysfunction not only led to a new drug indication but also to increased understanding of the role of type 5 phosphodiesterase in erectile tissue. These examples further reinforce the

concept that prescribers must remain vigilant to the possibility that unusual symptoms may reflect unappreciated drug effects.

Some 25–50% of patients make errors in self-administration of prescribed medicines, and these errors can be responsible for adverse drug effects. Similarly, patients commit errors in taking OTC drugs by not reading or following the directions on the containers. Physicians must recognize that providing directions with prescriptions does not always guarantee compliance.

In hospitals, drugs are administered in a controlled setting, and patient compliance is, in general, ensured. Errors may occur nevertheless—the wrong drug or dose may be given or the drug may be given to the wrong patient—and improved drug distribution and administration systems are addressing this problem.

Scope of the Problem

Patients receive, on average, 10 different drugs during each hospitalization. The sicker the patient, the more drugs are given, and there is a corresponding increase in the likelihood of adverse drug reactions. When <6 different drugs are given to hospitalized patients, the probability of an adverse reaction is ~5%, but if >15 drugs are given, the probability is >40%. Retrospective analyses of ambulatory patients have revealed adverse drug effects in 20%. Serious adverse reactions are also well-recognized with "herbal" remedies and OTC compounds: examples include kava-associated hepatotoxicity, L-tryptophan-associated eosinophilia-myalgia, and phenylpropanolamine-associated stroke, each of which has caused fatalities.

A small group of widely used drugs accounts for a disproportionate number of reactions. Aspirin and other NSAIDs, analgesics, digoxin, anticoagulants, diuretics, antimicrobials, glucocorticoids, antineoplastics, and hypoglycemic agents account for 90% of reactions, although the drugs involved differ between ambulatory and hospitalized patients.

Toxicity Unrelated to a Drug's Primary Pharmacologic Activity

CYTOTOXIC REACTIONS

Drugs or more commonly reactive metabolites generated by CYPs can covalently bind to tissue macromolecules (such as proteins or DNA) to cause tissue toxicity. Because of the reactive nature of these metabolites, covalent binding often occurs close to the site of production, typically the liver.

The most common cause of drug-induced hepatotoxicity is acetaminophen overdosage. Normally, reactive metabolites are detoxified by combining with hepatic glutathione. When glutathione becomes depleted, the metabolites bind instead to hepatic protein, with resultant hepatocyte damage. The hepatic necrosis produced by the ingestion of acetaminophen can be prevented or attenuated by the administration of substances such as *N*-acetylcysteine that reduce the binding of electrophilic metabolites to hepatic proteins. The risk of acetaminophen-related hepatic necrosis is increased in patients receiving drugs such as phenobarbital or phenytoin that increase the rate of drug metabolism or ethanol that exhaust glutathione stores. Such toxicity has even occurred with therapeutic dosages, so patients at risk through these mechanisms should be warned.

IMMUNOLOGIC MECHANISMS

Most pharmacologic agents are small molecules with low molecular weights (<2,000) and thus are poor immunogens. Generation of an immune response to a drug therefore usually requires in vivo activation and covalent linkage to protein, carbohydrate, or nucleic acid.

Drug stimulation of antibody production may mediate tissue injury by several mechanisms. The antibody may attack the drug when the drug is covalently attached to a cell and thereby destroy the cell. This occurs in penicillin-induced hemolytic anemia. Antibody-drug-antigen complexes may be passively adsorbed by a bystander cell, which is then destroyed by activation of complement; this occurs in quinine- and quinidine-induced thrombocytopenia. Heparin-induced thrombocytopenia arises when antibodies against complexes of platelet factor 4 peptide and heparin generate immune complexes that activate platelets; thus, the thrombocytopenia is accompanied by "paradoxical" thrombosis and is treated with thrombin inhibitors. Drugs or their reactive metabolites may alter a host tissue, rendering it antigenic and eliciting autoantibodies. For example, hydralazine and procainamide (or their reactive metabolites) can chemically alter nuclear material, stimulating the formation of antinuclear antibodies and occasionally causing lupus erythematosus. Drug-induced pure red cell aplasia (Chap. 107) is due to an immune-based drug reaction.

Serum sickness (Chap. 317) results from the deposition of circulating drug-antibody complexes on endothelial surfaces. Complement activation occurs, chemotactic factors are generated locally, and an inflammatory response develops at the site of complex entrapment. Arthralgias, urticaria, lymphadenopathy, glomerulonephritis, or cerebritis may result. Foreign proteins (vaccines, streptokinase, therapeutic antibodies) and antibiotics are common causes. Many drugs, particularly antimicrobial agents, ACE inhibitors, and aspirin, can elicit anaphylaxis with production of IgE, which binds to mast cell membranes. Contact with a drug antigen initiates a series of biochemical events in the mast cell and results in the release of mediators that can produce the characteristic urticaria, wheezing, flushing, rhinorrhea, and (occasionally) hypotension.

Drugs may also elicit cell-mediated immune responses. Topically administered substances may interact with sulfhydryl or amino groups in the skin and react with sensitized lymphocytes to produce the rash characteristic of contact dermatitis. Other types of rashes may also result from the interaction of serum factors, drugs, and sensitized lymphocytes.

Diagnosis and Treatment of Adverse Drug Reactions

The manifestations of drug-induced diseases frequently resemble those of other diseases, and a given set of manifestations may be produced by different and dissimilar drugs. Recognition of the role of a drug or drugs in an illness depends on appreciation of the possible adverse reactions to drugs in any disease, on identification of the temporal relationship between drug administration and development of the illness, and on familiarity with the common manifestations of the drugs. A suspected adverse drug reaction developing after introduction of a new drug naturally implicates that drug; however, it is also important to remember that a drug interaction may be responsible. Thus, for example, a patient on a chronic stable warfarin dose may develop a bleeding complication after introduction of amiodarone; this does not reflect a direct reaction to amiodarone but rather its effect to inhibit warfarin metabolism. Many associations between particular drugs and specific reactions have been described, but there is always a "first time" for a novel association, and any drug should be suspected of causing an adverse effect if the clinical setting is appropriate.

Illness related to a drug's intended pharmacologic action is often more easily recognized than illness attributable to immune or other mechanisms. For example, side effects such as cardiac arrhythmias in patients receiving digitalis, hypoglycemia in patients given insulin, or bleeding in patients receiving anticoagulants are more readily related to a specific drug than are symptoms such as fever or rash, which may be caused by many drugs or by other factors.

Electronic listings of adverse drug reactions can be useful. However, exhaustive compilations often provide little sense of perspective in terms of frequency and seriousness, which can vary considerably among patients.

Eliciting a drug history from each patient is important for diagnosis. Attention must be directed to OTC drugs and herbal preparations as well as to prescription drugs. Each type can be responsible for adverse drug effects, and adverse interactions may occur between OTC drugs and prescribed drugs. Loss of efficacy of oral contraceptives or cyclosporine by concurrent use of St. John's wort are examples. In addition, it is common for patients to be cared for by several physicians, and duplicative, additive, antagonistic, or synergistic drug combinations may therefore be administered if the physicians are not aware of the patients' drug histories. Every physician should determine what drugs a patient has been taking, for the previous month or two ideally, before prescribing any medications. Medications stopped for inefficacy or adverse effects should be documented to avoid pointless and potentially dangerous reexposure. A frequently overlooked source of additional drug exposure is topical therapy; for example, a patient complaining of bronchospasm may not mention that an ophthalmic beta blocker is being used unless specifically asked. A history of previous adverse drug effects in patients is common. Since these patients have shown a predisposition to drug-induced illnesses, such a history should dictate added caution in prescribing new drugs.

Laboratory studies may include demonstration of serum antibody in some persons with drug allergies involving cellular blood elements, as in agranulocytosis, hemolytic anemia, and thrombocytopenia. For example, both quinine and quinidine can produce platelet agglutination in vitro in the presence of complement and the serum from a patient who has developed thrombocytopenia following use of this drug. Biochemical abnormalities such as G6PD deficiency, serum pseudocholinesterase level, or genotyping may also be useful in diagnosis, often after an adverse effect has occurred in the patient or a family member.

Once an adverse reaction is suspected, discontinuation of the suspected drug followed by disappearance of the reaction is presumptive evidence of a drug-induced illness. Confirming evidence may be sought by cautiously reintroducing the drug and seeing if the reaction reappears. However, that should be done only if confirmation would be useful in the future management of the patient and if the attempt would not entail undue risk. With concentration-dependent adverse reactions, lowering the dosage may cause the reaction to disappear, and raising it may cause the reaction to reappear. When the reaction is thought to be allergic, however, readministration of the drug may be hazardous, since anaphylaxis may develop.

If the patient is receiving many drugs when an adverse reaction is suspected, the drugs likeliest to be responsible can usually be identified; this should include both potential culprit agents as well as drugs that alter their elimination. All drugs may be discontinued at once or, if this is not practical, discontinued one at a time, starting with the ones most suspect, and the patient observed for signs of improvement. The time needed for a concentration-dependent adverse effect to disappear depends on the time required for the concentration to fall below the range associated with the adverse effect; that, in turn, depends on the initial blood level and on the rate of elimination or metabolism of the drug. Adverse effects of drugs with long half-lives or those not directly related to serum concentration may take a considerable time to disappear.

SUMMARY

Modern clinical pharmacology aims to replace empiricism in the use of drugs with therapy based on in-depth understanding of factors that determine an individual's response to drug treatment. Molecular pharmacology, pharmacokinetics, genetics, clinical trials, and the educated prescriber all contribute to this process. No drug response should ever be termed *idiosyncratic*; all responses have a mechanism whose understanding will help guide further therapy with that drug or successors. This rapidly expanding understanding of variability in drug actions makes the process of prescribing drugs increasingly daunting for the practitioner. However, fundamental principles should guide this process:

- The benefits of drug therapy, however defined, should always outweigh the risk.
- The smallest dosage necessary to produce the desired effect should be used.
- The number of medications and doses per day should be minimized.
- Although the literature is rapidly expanding, accessing it is becoming easier; electronic tools to search databases of literature and unbiased opinion will become increasingly commonplace.
- Genetics play a role in determining variability in drug response and may become a part of clinical practice.
- Electronic medical record and pharmacy systems will increasingly incorporate prescribing advice, such as indicated medications not used; unindicated medications being prescribed; and potential dosing errors, drug interactions, or genetically determined drug responses.
- Prescribers should be particularly wary when adding or stopping specific drugs that are especially liable to provoke interactions and adverse reactions.
- Prescribers should use only a limited number of drugs, with which they are thoroughly familiar.

FURTHER READINGS

COLLET JP et al: Cytochrome P450 2C19 polymorphism in young patients treated with clopidogrel after myocardial infarction: a cohort study. Lancet 373:309, 2009[PMID: 19108880]

GE D et al: Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461:399, 2009[PMID: 19684573]

GIACOMINI KM et al: The pharmacogenetics research network: from SNP discovery to clinical drug response. Clin Pharmacol Ther 81:328, 2007[PMID: 17339863]

HO PM et al: Risk of adverse outcomes associated with concomitant use of clopidogrel and proton pump inhibitors following acute coronary syndrome. JAMA 301:937, 2009[PMID: 19258584]

THE INTERNATIONAL WARFARIN PHARMACOGENETICS CONSORTIUM: Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med 360:753, 2009

KARAPETIS CS et al: K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med 359:1757, 2008[PMID: 18946061]

LIGGETT SB et al: A polymorphism within a conserved ß1-adrenergic receptor motif alters cardiac function and ß-blocker response in human heart failure. Proc Natl Acad Sci (USA) 103:11288, 2006[PMID: 16844790]

LINK E et al: SLCO1B1 variants and statin-induced myopathy—a genomewide study. N Engl J Med 359:789, 2008[PMID: 18650507]

MALLAL S et al: HLA-B*5701 Screening for hypersensitivity to abacavir. N Engl J Med 358:568, 2008[PMID: 18256392]

MEGA JL et al: Cytochrome P-450 polymorphisms and response to clopidogrel. N Engl J Med 360:354, 2009[PMID: 19106084]

RODEN DM, Stein CM: Clopidogrel and the concept of high risk pharmacokinetics. Circulation 119:2127, 2009[PMID: 19398674]

SIMON T et al: Genetic determinants of response to clopidogrel and cardiovascular events. N Engl J Med 360:363, 2009[PMID: 19106083]

WEISS ST et al: Creating and evaluating genetic tests predictive of drug response. Nat Rev Drug Discov 7:568, 2008[PMID: 18587383]

WILKE RA et al: Identifying genetic risk factors for serious adverse drug reactions: current progress

and challenges. Nat Rev Drug Discov 6:904, 2007[PMID: 17971785]

WOODCOCK J, LESKO LJ: Pharmacogenetics—tailoring treatment for the outliers. N Engl J Med 360:811, 2009[PMID: 19228625]

YANG JJ et al: Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. JAMA 301:393, 2009[PMID: 19176441]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

WOMENS HEALTH: INTRODUCTION

The study of biologic differences between sexes has emerged as a distinct scientific discipline. A report from the Institute of Medicine (IOM) found that sex has a broad impact on biologic and disease processes and succinctly concluded that sex matters. The National Institutes of Health established the Office of Research on Womens Health in 1990 to develop an agenda for future research in the field. In parallel, womens health has become a distinct clinical discipline with a focus on disorders that occur disproportionately in women. The integration of womens health into internal medicine and other specialties has been accompanied by novel approaches to health care delivery, including greater attention to patient education and involvement in disease prevention and medical decision-making.

The IOM report recommended the term *sex difference* to describe biologic processes that differ between males and females and *gender difference* for features related to social influences. Disorders highlighted here are reviewed in detail in other chapters.

DISEASE RISK: REALITY AND PERCEPTION

The leading causes of death are the same in women and men: (1) heart disease, (2) cancer (Table 6-1; Fig. 6-1). The leading cause of cancer death, lung cancer, is the same in both sexes. Breast cancer is the second leading cause of cancer death in women, but it causes about 60% fewer deaths than does lung cancer. Men are substantially more likely to die from suicide, homicide, and accidents than are women.

Table 6-1 Deaths and Percentage of Total Deaths for the Leading Causes of Death by Sex in the United States in 2006

Cause of Death	Female			Male		
	Rank	Deaths	Deaths,%	Rank	Deaths	Deaths,%
Diseases of heart	1	315,930	25.8	1	315,706	26.3
Malignantneoplasms	2	269,819	22.0	2	290,069	24.1
Cerebrovascular diseases	3	82,595	6.7	5	54,524	4.5
Chronic lower respiratory diseases	4	65,323	5.3	4	59,260	4.9
Alzheimers disease	5	51,281	4.2	9	21,151	1.8
Accidents (unintentional injuries)	6	42,658	3.5	3	78,941	6.6
Diabetes mellitus	7	36,443	3.0	6	36,006	3.0
Pneumonia	8	30,189	2.5	8	25,288	2.1
Renal failure	9	22,229	1.8	10	21,115	1.8
Septicemia	10	18,712	1.5	12	15,522	1.3
Essential hypertension and hypertensive renal disease	11	14,440	1.2	16	9,415	0.8
Other diseases of respiratory system	12	13,916	1.1	14	13,728	1.1
Chronic liver disease and cirrhosis	13	9,689	0.8	11	17,866	1.5
Parkinsons disease	14	8,266	0.7	15	11,300	0.9
Pneumonitis due to solids and liquids	15	7,971	0.7	17	8,916	0.7
Intentional self-harm (suicide)	17	6,992	0.6	7	26,308	2.2
Assault (homicide)	24	3,856	0.3	13	14,717	1.2

Source: Data from Centers for Disease Control and Prevention: National Vital Statistics Reports, Vol. 57, No. 14, April 17, 2009, Table 12, http://www.cdc.gov/NCHS/data/nvsr/nvsr57/nvsr57_14.pdf.

Figure 6-1 Death Rates Per 100,000 Population for 2006 by 5-Year Age Groups in U.S. Women.

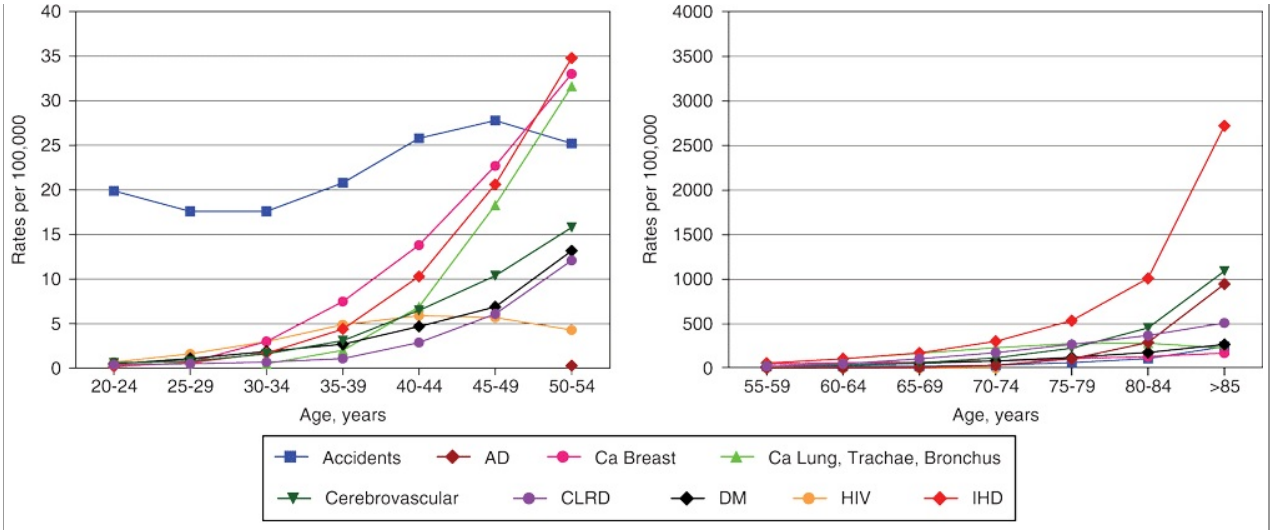

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Note that the scale of the *y* axis is increased in the graph on the right compared with that on the left. Accidents and HIV/AIDS are the leading causes of death in young women 20–34 years of age. Accidents, breast cancer, and ischemic heart disease (IHD) are the leading causes of death in women 35–49 years of age. IHD becomes the leading cause of death in women beginning at age 50 years. In older women, IHD remains the leading cause of death, cerebrovascular disease becomes the second leading cause of death, and lung cancer is the leading cause of cancer-related deaths. At age 85 years and beyond, AD becomes the third leading cause of death. AD, Alzheimers disease; Ca, cancer; CLRD, chronic lower respiratory disease; DM, diabetes mellitus. *(Data adapted from Centers for Disease Control and Prevention, http://www.cdc.gov/nchs/data/dvs/MortFinal2006_WorkTable210R.pdf.)*

Womens risk for many diseases increases at menopause, which occurs at a median age of 51.4 years. In the industrialized world, women spend one-third of their lives in the postmenopausal period. Estrogen levels fall abruptly at menopause, inducing a variety of physiologic and metabolic responses. Rates of cardiovascular disease increase and bone density begins to decrease rapidly after menopause. In the United States, women live on average about 5 years longer than men, with a life expectancy at birth in 2007 of 80.4 years compared with 75.3 years in men. Elderly women outnumber elderly men, so that age-related conditions such as hypertension have a female preponderance. However, the difference in life expectancy between men and women has decreased an average of 0.1 year every year since 1980, and if this convergence in mortality figures continues, it is projected that mortality rates will be similar by 2054.

Womens perception of disease risk is often inaccurate (Fig. 6-2). Public awareness campaigns have resulted in almost 60% of U.S. women knowing that cardiovascular disease is the leading cause of death in women. Nevertheless, the condition they fear most is breast cancer despite the fact that death rates from breast cancer have been falling since the 1990s. In any specific decade of life, a womans risk for breast cancer never exceeds 1 in 34. Although a womans lifetime risk of developing breast cancer if she lives past 85 years is about 1 in 9, it is much more likely that she will die from cardiovascular disease than from breast cancer. In other words, many elderly women have breast cancer but die from other causes. Similarly, a minority of women are aware that lung cancer is the leading cause of cancer death in women. Physicians are also less likely to recognize womens risk for cardiovascular disease. These misconceptions are unfortunate as they perpetuate inadequate attention to modifiable risk factors such as dyslipidemia, hypertension, and cigarette smoking.

Figure 6-2 Trends in Perceived Greatest Health Problem among Women by Survey Year.

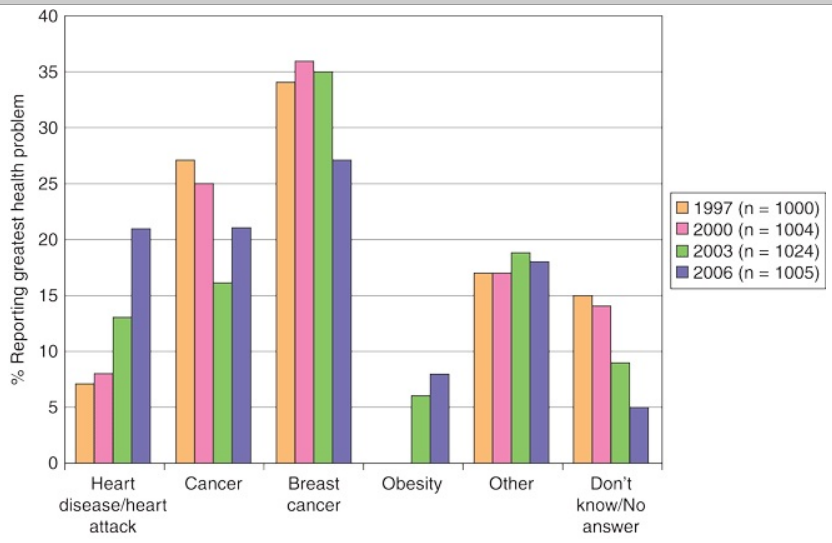

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Data for obesity were not available for 1997 and 2000. Significantly more women cited heart disease/heart attack as the greatest health problem for women in 2006 compared with previous survey years. Breast cancer remained the most commonly identified greatest health risk in all survey years. *(From Christian et al; with permission.)*

SEX DIFFERENCES IN HEALTH AND DISEASE

Alzheimers Disease

(See also Chap. 371) Alzheimers disease (AD) affects approximately twice as many women as men. Because the risk for AD increases with age, part of this sex difference is accounted for by the fact that women live longer than men. However, additional factors probably contribute to the increased risk for AD in women, including sex differences in brain size, structure, and functional organization. There is emerging evidence for sex-specific differences in gene expression, not only for genes on the X and Y chromosomes but also for some autosomal genes. Estrogens have pleiotropic genomic and nongenomic effects on the central nervous system, including neurotrophic actions in key areas involved in cognition and memory. Women with AD have lower endogenous estrogen levels than do women without AD. These observations have led to the hypothesis that estrogen is neuroprotective.

Some studies have suggested that estrogen administration improves cognitive function in nondemented postmenopausal women as well as in women with AD, and several observational studies have suggested that postmenopausal hormone therapy (HT) may decrease the risk of AD. However, HT placebo-controlled trials have found no improvement in disease progression or cognitive function in women with AD. Further, the Womens Health Initiative Memory Study (WHIMS), an ancillary study in the Womens Health Initiative (WHI), found no benefit compared with placebo of estrogen alone [combined continuous equine estrogen (CEE), 0.625 mg qd] or estrogen with progestin [CEE, 0.625 mg qd, and medroxyprogesterone acetate (MPA), 2.5 mg qd] on cognitive function or the development of dementia in women ≥65 years. Indeed, there was a significantly increased risk for both dementia and mild cognitive impairment in women receiving hormone therapy. The possible explanations for the discrepant results between the observational studies and the randomized clinical trials remain unclear (Chap. 348).

Cardiovascular Disease

(See also Chap. 243) There are major sex differences in cardiovascular disease, the leading cause of death in men and women in developed countries. Since 1984, more women than men have died of cardiovascular disease. Gonadal steroids have major effects on the cardiovascular system and lipid metabolism. Estrogen increases high-density lipoprotein (HDL) and lowers low-density lipoprotein (LDL), whereas androgens have the opposite effect. Estrogen has direct vasodilatory effects on the vascular endothelium, enhances insulin sensitivity, and has antioxidant as well as anti-inflammatory properties. There is a striking increase in coronary heart disease (CHD) after both natural and surgical menopause, suggesting that endogenous estrogens are cardioprotective. Women also have longer QT intervals on electrocardiograms, and this increases their susceptibility to certain arrhythmias. Animal studies suggest that the sex difference in the duration of the QT interval is caused by the effects of sex steroids on cardiac repolarization, in part related to their effects on cardiac voltage-gated potassium channels; there is a lower density of the rapid component (I_{Kr}) of the delayed rectifier potassium current (I_K) in females.

CHD presents differently in women, who are usually 10–15 years older than their male counterparts and are more likely to have comorbidities such as hypertension, congestive heart failure, and diabetes mellitus (DM). In the Framingham study, angina was the most common initial symptom of CHD in women, whereas myocardial infarction was the most common initial presentation in men. Women more often have atypical symptoms such as nausea, vomiting, indigestion, and upper back pain.

Women with myocardial infarction are more likely to present with cardiac arrest or cardiogenic shock, whereas men are more likely to present with ventricular tachycardia. Further, younger women with myocardial infarction are more likely to die than are men of similar age. However, this mortality gap has decreased substantially in recent years because younger women have experienced greater improvements in survival after myocardial infarction than men (Fig. 6-3). The improvement in survival is due largely to a reduction in comorbidities, suggesting a greater attention to modifiable risk factors in women.

Figure 6-3 Hospital Mortality Rates in Men and Women for Acute Myocardial Infarction in 1994–95 Compared with 2004–06.

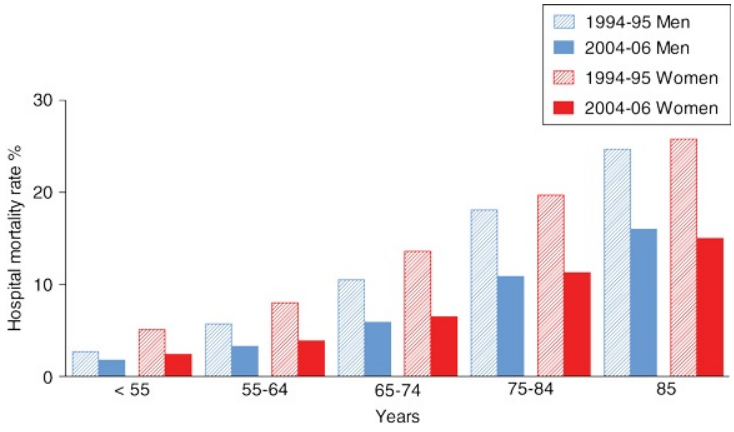

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Women younger than age 65 years had substantially greater mortality than men of similar age in 1994–95. Mortality rates declined markedly for both sexes across all age groups in 2004–06 compared with 1994–95. However, there was a more striking decrease in mortality in women younger than age 75 years compared with men of similar age. The mortality rate reduction was largest in women less than age 55 years (52.9%) and lowest in men of similar age (33.3%). *(Data adapted from Vaccarino et al.)*

Nevertheless, physicians are less likely to suspect heart disease in women with chest pain and less likely to perform diagnostic and therapeutic cardiac procedures in women. In addition, there are sex differences in the accuracy of certain diagnostic procedures. The exercise electrocardiogram has substantial false-positive as well as false-negative rates in women compared with men. Women are less likely to receive therapies such as angioplasty, thrombolytic therapy, coronary artery bypass grafts (CABGs), beta blockers, and aspirin. There are also sex differences in outcomes when women with CHD do receive therapeutic interventions. Women undergoing CABG surgery have more advanced disease, a higher perioperative mortality rate, less relief of angina, and less graft patency; however, 5- and 10-year survival rates are similar. Women undergoing percutaneous transluminal coronary angioplasty have lower rates of initial angiographic and clinical success than men, but they also have a lower rate of restenosis and a better long-term outcome. Women may benefit less and have more frequent serious bleeding complications from thrombolytic therapy compared with men. Factors such as older age, more comorbid conditions, and more

severe CHD in women at the time of events or procedures appear to account in part for the observed sex differences.

Elevated cholesterol levels, hypertension, smoking, obesity, low HDL cholesterol levels, DM, and lack of physical activity are important risk factors for CHD in both men and women. Total triglyceride levels are an independent risk factor for CHD in women but not in men. Low HDL cholesterol and DM are more important risk factors for CHD in women than in men. Smoking is an important risk factor for CHD in women—it accelerates atherosclerosis, exerts direct negative effects on cardiac function, and is associated with an earlier age of menopause. Cholesterol-lowering drugs are equally effective in men and women for primary and secondary prevention of CHD. However, because of perceptions that women are at lower risk for CHD, they receive fewer interventions for modifiable risk factors than do men.

In contrast to men, randomized trials showed that aspirin was not effective in the primary prevention of CHD in women; it did significantly reduce the risk of ischemic stroke. Secondary prevention in women with known CHD is also suboptimal. At baseline, only about 30% of women enrolled in the Heart and Estrogen/progestin Replacement Study (HERS), a secondary prevention trial in women with established CHD, were taking beta blockers, and only 45% received lipid-lowering medications.

The sex differences in CHD prevalence, beneficial biologic effects of estradiol on the cardiovascular system, and reduced risk for CHD in observational studies of women receiving HT led to the widespread use of HT for the prevention of CHD. However, the WHI, which studied more than 16,000 women on CEE plus MPA or placebo and more than 10,000 women with hysterectomy on CEE alone or placebo, did not demonstrate a benefit of HT for the primary or secondary prevention of CHD. In addition, CEE plus MPA was associated with an increased risk for CHD, particularly in the first year of therapy, whereas CEE alone neither increased nor decreased CHD risk. Both CEE plus MPA and CEE alone were associated with an increased risk for ischemic stroke. There was no evidence for cardioprotective effects of estrogens in smaller randomized trials that used either oral or transdermal estradiol, arguing against the hypothesis that the type of estrogen or its route of administration accounted for the lack of CHD risk reduction.

In the WHI, there was a suggestion of a reduction in CHD risk in women who initiated HT closer to menopause. This finding suggests that the time at which HT is initiated is critical for cardioprotection and is consistent with the "timing hypothesis." According to this hypothesis, HT has differential effects, depending on the stage of atherosclerosis; adverse effects are seen with advanced, unstable lesions. This hypothesis is under investigation in randomized clinical trials, for example, the Kronos Early Estrogen Prevention Study (KEEPS). It is noteworthy that there is no reduction in the risk for ischemic stroke when HT is initiated closer to menopause. HT is discussed further in Chap. 348.

Diabetes Mellitus

(See also Chap. 344) Women are more sensitive to insulin than men are. Despite this, the prevalence of type 2 DM is similar in men and women. There is a sex difference in the relationship between endogenous androgen levels and DM risk: Higher bioavailable testosterone levels are associated with increased risk in women, whereas lower bioavailable testosterone levels are associated with increased risk in men. Polycystic ovary syndrome and gestational DM—common conditions in premenopausal women—are associated with a significantly increased risk for type 2 DM. Premenopausal women with DM lose the cardioprotective effect of female sex and have rates of CHD identical to those in males. These women have impaired endothelial function and reduced coronary vasodilatory responses, which may predispose to cardiovascular complications. Among individuals with DM, women have a greater risk for myocardial infarction than do men. Women with DM are more likely to have left ventricular hypertrophy. In the WHI, CEE plus MPA significantly reduced the incidence of DM, whereas with CEE alone there was only a trend toward decreased DM incidence.

Hypertension

(See also Chap. 247) After age 60, hypertension is more common in U.S. women than in men, largely because of the high prevalence of hypertension in older age groups and the longer survival of women. Isolated systolic hypertension is present in 30% of women >60 years. Sex hormones affect blood pressure. Both normotensive and hypertensive women have higher blood pressure levels during the follicular phase than during the luteal phase. In the Nurses Health Study, the relative risk of hypertension was 1.8 in current users of oral contraceptives, but this risk is lower with the newer low-dose contraceptive preparations. HT is not associated with hypertension. Among secondary causes of hypertension, there is a female preponderance of renal artery fibromuscular dysplasia.

The benefits of treatment for hypertension have been dramatic in both women and men. A meta-analysis of the effects of hypertension treatment, the Individual Data Analysis of Antihypertensive Intervention Trial, found a reduction of risk for stroke and for major cardiovascular events in women. The effectiveness of various antihypertensive drugs appears to be comparable in women and men; however, women may experience more side effects. For example, women are more likely to develop cough with angiotensin-converting enzyme inhibitors.

Autoimmune Disorders

(See also Chap. 318) Most autoimmune disorders occur more commonly in women than in men; they include autoimmune thyroid and liver diseases, lupus, rheumatoid arthritis (RA), scleroderma, multiple sclerosis (MS), and idiopathic thrombocytopenic purpura. However, there is no sex difference in the incidence of type 1 DM, and ankylosing spondylitis occurs more commonly in men. There are relatively few differences in bacterial disease infection rates between men and women. In general, sex differences in viral diseases can be accounted for by differences in behaviors, such as exposures or rates of immunization. Sex differences in both immune responses and adverse reactions to vaccines have been reported. For example, there is a female preponderance of postvaccination arthritis.

The mechanisms for these sex differences remain obscure. Adaptive immune responses are more robust in women than in men; this may be explained by the stimulatory actions of estrogens and the inhibitory actions of androgens on the cellular mediators of immunity. Consistent with an important role for gonadal hormones, there is variation in immune responses during the menstrual cycle, and the activity of certain autoimmune disorders is altered by castration or pregnancy (e.g., RA and MS may remit during pregnancy). Nevertheless, the majority of studies show that exogenous estrogens and progestins in the form of HT or oral contraceptives do not alter autoimmune disease incidence or activity. Exposure to fetal antigens, including circulating fetal cells that persist in certain tissues, has been speculated to increase the risk of autoimmune responses. There is clearly an important genetic component to autoimmunity, as indicated by the familial clustering and HLA association of many such disorders. However, HLA types are not sexually dimorphic.

HIV Infection

(See also Chap. 189) Women account for almost 50% of the 40 million persons infected with HIV-1 worldwide. AIDS is an important cause of death in younger women (Fig. 6-1). Heterosexual contact with an at-risk partner is the fastest-growing transmission category, and women are more susceptible to HIV infection than are men. This increased susceptibility is accounted for in part by an increased prevalence of sexually transmitted diseases in women. Some studies have suggested that hormonal contraceptives may increase the risk of HIV transmission. Progesterone has been shown to increase susceptibility to infection in nonhuman primate models of HIV. Women are also more likely to be infected by multiple variants of the virus than are men. Women with HIV have more rapid decreases in their CD4 cell counts than do men. Compared with men, HIV-infected women more frequently develop candidiasis, but Kaposis sarcoma is less common than it is in men. Women have more adverse

reactions, such as lipodystrophy, dyslipidemia, and rash, with antiretroviral therapy than do men. This observation is explained in part by sex differences in the pharmacokinetics of certain antiretroviral drugs, resulting in higher plasma concentrations in women.

Obesity

(See also Chap. 78) The prevalence of obesity is higher in women than in men. Further, >80% of patients who undergo bariatric surgery are women. Pregnancy and menopause are risk factors for obesity. There are major sex differences in body fat distribution. Women characteristically have gluteal and femoral or gynoid pattern of fat distribution, whereas men typically have a central or android pattern. Women have more subcutaneous fat than men. Gonadal steroids appear to be the major regulators of fat distribution through a number of direct effects on adipose tissue. Studies in humans also suggest that gonadal steroids play a role in modulating food intake and energy expenditure.

In men and women, upper-body obesity characterized by increased visceral fat is associated with an increased risk for cardiovascular disease and DM. In women, endogenous androgen levels are positively associated with upper-body obesity, and androgen administration increases visceral fat. In contrast, there is an inverse relationship between endogenous androgen levels and central obesity in men. Further, androgen administration decreases visceral fat in centrally obese men. The reasons for these sex differences in the relationship between visceral fat and androgens are unknown. Obesity increases a womans risk for certain cancers, in particular postmenopausal breast and endometrial cancer, in part because adipose tissue provides an extragonadal source of estrogen through aromatization of circulating adrenal and ovarian androgens, especially the conversion of androstenedione to estrone. Obesity increases the risk of infertility, miscarriage, and complications of pregnancy.

Osteoporosis

(See also Chap. 354) Osteoporosis is about five times more common in postmenopausal women than in age-matched men, and osteoporotic hip fractures are a major cause of morbidity in elderly women. Men accumulate more bone mass and lose bone more slowly than do women. Sex differences in bone mass are found as early as infancy. Calcium intake, vitamin D, and estrogen all play important roles in bone formation and bone loss. Particularly during adolescence, calcium intake is an important determinant of peak bone mass. Vitamin D deficiency is surprisingly common in elderly women, occurring in >40% of women living in northern latitudes. Receptors for estrogens and androgens have been identified in bone. Estrogen deficiency is associated with increased osteoclast activity and a decreased number of bone-forming units, leading to net bone loss. The aromatase enzyme, which converts androgens to estrogens, is also present in bone. Estrogen is an important determinant of bone mass in men (derived from the aromatization of androgens) as well as in women.

Pharmacology

On average, women have lower body weights, smaller organs, a higher percentage of body fat, and lower total-body water than men. There are also important sex differences in drug action and metabolism that are not accounted for by these differences in body size and composition. Gonadal steroids alter the binding and metabolism of a number of drugs. Further, menstrual cycle phase and pregnancy can alter drug action. Two-thirds of cases of drug-induced torsades des pointes, a rare, life-threatening ventricular arrhythmia, occur in women because they have a longer, more vulnerable QT interval. These drugs, which include certain antihistamines, antibiotics, antiarrhythmics, and antipsychotics, can prolong cardiac repolarization by blocking cardiac voltage-gated potassium channels, particularly I_{Kr}. Women require lower doses of neuroleptics to control schizophrenia. Women awaken from anesthesia faster than do men given the same doses of anesthetics. Women also take more medications than men, including over-the-counter formulations and supplements. The greater use of medications combined with these biologic differences may account for the reported higher frequency of adverse drug reactions in women than in men.

Psychological Disorders

(See also Chap. 391) Depression, anxiety, and affective and eating disorders (bulimia and anorexia nervosa) are more common in women than in men. Epidemiologic studies from both developed and developing nations consistently find major depression to be twice as common in women as in men, with the sex difference becoming evident in early adolescence. Depression occurs in 10% of women during pregnancy and in 10–15% of women during the postpartum period. There is a high likelihood of recurrence of postpartum depression with subsequent pregnancies. The incidence of major depression diminishes after age 45 years and does not increase with the onset of menopause. Depression in women appears to have a worse prognosis than does depression in men; episodes last longer, and there is a lower rate of spontaneous remission. Schizophrenia and bipolar disorders occur at equal rates in men and women, although there may be sex differences in symptoms.

Both biologic and social factors account for the greater prevalence of depressive disorders in women. Men have higher levels of the neurotransmitter serotonin. Gonadal steroids also affect mood, and fluctuations during the menstrual cycle have been linked to symptoms of premenstrual syndrome. Sex hormones differentially affect the hypothalamic-pituitary-adrenal responses to stress. Testosterone appears to blunt cortisol responses to corticotropin-releasing hormone. Both low and high levels of estrogen can activate the hypothalamic-pituitary-adrenal axis.

Sleep Disorders

(See also Chap. 27) There are striking sex differences in sleep and its disorders. During sleep, women have an increased amount of slow-wave activity, differences in timing of delta activity, and an increase in the number of sleep spindles. Testosterone modulates neural control of breathing and upper airway mechanics. Men have a higher prevalence of sleep apnea. Testosterone administration to hypogonadal men as well as to women increases apneic episodes during sleep. Women with the hyperandrogenic disorder polycystic ovary syndrome have an increased prevalence of obstructive sleep apnea, and apneic episodes are positively correlated with their circulating testosterone levels. In contrast, progesterone accelerates breathing, and in the past, progestins were used for treatment of sleep apnea.

Substance Abuse and Tobacco

(See also Chaps. 392 and 395) Substance abuse is more common in men than in women. However, one-third of Americans who suffer from alcoholism are women. Women alcoholics are less likely to be diagnosed than men. A greater proportion of men than women seek help for alcohol and drug abuse. Men are more likely to go to an alcohol or drug treatment facility, whereas women tend to approach a primary care physician or mental health professional for help under the guise of a psychosocial problem. Late-life alcoholism is more common in women than in men. On average, alcoholic women drink less than alcoholic men but exhibit the same degree of impairment. Blood alcohol levels are higher in women than in men after drinking equivalent amounts of alcohol, adjusted for body weight. This greater bioavailability of alcohol in women is due to both the smaller volume of distribution and the slower gastric metabolism of alcohol secondary to lower activity of gastric alcohol dehydrogenase than is the case in men. In addition, alcoholic women are more likely to abuse tranquilizers, sedatives, and amphetamines. Women alcoholics have a higher mortality rate than do nonalcoholic women and alcoholic men. Women also appear to develop alcoholic liver disease and other alcohol-related diseases with shorter drinking histories and lower levels of alcohol consumption. Alcohol abuse also poses special risks to a woman, adversely affecting fertility and the health of the baby (fetal alcohol syndrome). Even moderate alcohol use increases the risk of breast cancer, hypertension, and stroke in women.

More men than women smoke tobacco, but the prevalence of smoking is declining faster in men than in women. Smoking

markedly increases the risk of cardiovascular disease in premenopausal women and is also associated with a decrease in the age of menopause. Women who smoke are more likely to develop chronic obstructive pulmonary disease and lung cancer than men and at lower levels of tobacco exposure.

Violence Against Women

Domestic violence is the most common cause of physical injury in women, exceeding the combined incidence of all other types of injury (such as from rape, mugging, and auto accidents). Sexual assault is one of the most common crimes against women. One in five adult women in the United States reports having experienced sexual assault during her lifetime. Adult women are much more likely to be raped by a spouse, ex-spouse, or acquaintance than by a stranger. Domestic violence may be an unrecognized feature of certain clinical presentations, such as chronic abdominal pain, headaches, substance abuse, and eating disorders, in addition to more obvious manifestations such as trauma.

SUMMARY

Womens health is now a mature discipline, and the importance of sex differences in biologic processes is well recognized. There has been a striking reduction in the excess mortality rate from myocardial infarction in younger women. Nevertheless, ongoing misperceptions about disease risk not only among women but also among their physicians result in inadequate attention to modifiable risk factors. Research into the fundamental mechanisms of sex differences will provide important biologic insights. Further, those insights will have an impact on both womens and mens health.

FURTHER READINGS

Christian AH et al: Nine-year trends and racial and ethnic disparities in womens awareness of heart disease and stroke: An American Heart Association national study. J Womens Health (Larchmt) 16:68, 2007[PMID: 17274739]

Mosca L et al: National study of physician awareness and adherence to cardiovascular disease prevention guidelines. Circulation 111:499, 2005[PMID: 15687140]

Prentice RL et al: Benefits and risks of postmenopausal hormone therapy when it is initiated soon after menopause. Am J Epidemiol 170:12, 2009[PMID: 19468079]

Rossouw JE et al: Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. JAMA 297:1465, 2007[PMID: 17405972]

Utian WH et al: Estrogen and progestogen use in postmenopausal women: July 2008 position statement of the North American Menopause Society. Menopause 15:584, 2008[PMID: 18580541]

Vaccarino V et al: Sex differences in mortality after acute myocardial infarction: Changes from 1994 to 2006. Arch Intern Med 169:1767, 2009[PMID: 19858434]

Wizemann TM, Pardue M-L (eds): *Exploring the Biological Contributions to Human Health: Does Sex Matter?* Washington, DC, National Academy of Sciences, 2001

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MEDICAL DISORDERS DURING PREGNANCY: INTRODUCTION

Each year, approximately 4 million births occur in the United States, and more than 130 million births occur worldwide. A significant proportion of these are complicated by medical disorders. In the past, many medical disorders were contraindications to pregnancy. Advances in obstetrics, neonatology, obstetric anesthesiology, and medicine have increased the expectation that pregnancy will result in a positive outcome for both mother and fetus despite most of these conditions. Successful pregnancy requires important physiologic adaptations, such as a marked increase in cardiac output. Medical problems that interfere with the physiologic adaptations of pregnancy increase the risk for poor pregnancy outcome; conversely, in some instances, pregnancy may adversely impact an underlying medical disorder.

HYPERTENSION

(See also Chap. 247)

In pregnancy, cardiac output increases by 40%, most of which is due to an increase in stroke volume. Heart rate increases by ~10 beats/min during the third trimester. In the second trimester, systemic vascular resistance decreases and this is associated with a fall in blood pressure. During pregnancy, a blood pressure of 140/90 mmHg is considered to be abnormally elevated and is associated with an increase in perinatal morbidity and mortality. In all pregnant women, the measurement of blood pressure should be performed in the sitting position, because the lateral recumbent position may result in a blood pressure lower than that recorded in the sitting position. The diagnosis of hypertension requires the measurement of two elevated blood pressures, at least 6 h apart. Hypertension during pregnancy is usually caused by preeclampsia, chronic hypertension, gestational hypertension, or renal disease.

Preeclampsia

Approximately 5–7% of all pregnant women develop *preeclampsia*, the new onset of hypertension (blood pressure >140/90 mmHg) and proteinuria (>300 mg/24 h) after 20 weeks of gestation. Although the precise pathophysiology of preeclampsia remains unknown, recent studies show excessive placental production of antagonists to both vascular epithelial growth factor (VEGF) (soluble fms-like tyrosine kinase 1 and sflt-1) and transforming growth factor ß (TGF-ß) (endoglin). These antagonists to VEGF and TGF-ß disrupt endothelial and renal glomerular function resulting in edema, hypertension, and proteinuria. The renal histological feature of preeclampsia is glomerular endotheliosis. Glomerular endothelial cells are swollen and encroach on the vascular lumen. Preeclampsia is associated with abnormalities of cerebral circulatory autoregulation, which increase the risk of stroke at near-normal blood pressures. Risk factors for the development of preeclampsia include nulliparity, diabetes mellitus, a history of renal disease or chronic hypertension, a prior history of preeclampsia, extremes of maternal age (>35 years or <15 years), obesity, antiphospholipid antibody syndrome, and multiple gestation.

Severe preeclampsia is the presence of new-onset hypertension and proteinuria accompanied by end organ damage. Features may include marked elevation of blood pressure (>160/110 mmHg), severe proteinuria (>5 g/24 h), or evidence of central nervous system (CNS) dysfunction (headaches, blurred vision, seizures, coma), renal dysfunction (oliguria or creatinine > 1.5 mg/dL), pulmonary edema, hepatocellular injury (ALT > 2-fold the upper limits of normal), hematologic dysfunction (platelet count < 100,000/L or disseminated intravascular coagulation), or placental dysfunction (oligohydramnios or severe intrauterine growth restriction). The HELLP (*h*emolysis, *e*levated *l*iver enzymes, *l*ow *p*latelets) syndrome is a special subgroup of severe preeclampsia and is a major cause of morbidity and mortality in this disease. The presence of platelet dysfunction and coagulation disorders further increases the risk of stroke.

TREATMENT: PREECLAMPSIA

Preeclampsia resolves within a few weeks after delivery. For pregnant women with preeclampsia prior to 37 weeks gestation, delivery reduces the mothers morbidity but exposes the fetus to the risk of premature birth. The management of preeclampsia is challenging because it requires the clinician to balance the health of the mother and fetus simultaneously. In general, prior to term, women with *mild* preeclampsia may be managed conservatively with bed rest, close monitoring of blood pressure and renal function, and careful fetal surveillance. For women with *severe* preeclampsia, delivery is

recommended unless the patient is eligible for expectant management in a tertiary hospital setting. Expectant management of severe preeclampsia remote from term affords some benefits for the fetus but significant risks for the mother.

The definitive treatment of preeclampsia is delivery of the fetus and placenta. For women with severe preeclampsia, aggressive management of blood pressures > 160/110 mmHg reduces the risk of cerebrovascular accidents. Intravenous labetalol or hydralazine are the drugs most commonly used to manage preeclampsia, but labetalol is associated with fewer episodes of maternal hypotension. Nifedipine is also commonly used in pregnancy. Elevated arterial pressure should be reduced slowly to avoid hypotension and a decrease in blood flow to the fetus. Angiotensin-converting enzyme (ACE) inhibitors as well as angiotensin-receptor blockers should be avoided in the second and third trimesters of pregnancy because of their adverse effects on fetal development.

Magnesium sulfate is the treatment of choice for the prevention and treatment of eclamptic seizures. Large, randomized clinical trials have demonstrated the superiority of magnesium sulfate over phenytoin and diazepam, and the efficacy of magnesium sulfate in reducing the risk of seizure and, possibly, the risk of maternal death. Magnesium may prevent seizures by interacting with *N*-methyl-d-aspartate (NMDA) receptors in the CNS. Given the difficulty of predicting eclamptic seizures on the basis of disease severity, once the decision to proceed with delivery is made, most patients carrying a diagnosis of preeclampsia should be treated with magnesium sulfate. Women who have had preeclampsia appear to be at increased risk of cardiovascular and renal disease later in life.

Chronic Essential Hypertension

Pregnancy complicated by chronic essential hypertension is associated with intrauterine growth restriction and increased perinatal mortality. Pregnant women with chronic hypertension are at increased risk for superimposed preeclampsia and abruptio placenta. Women with chronic hypertension should have a thorough prepregnancy evaluation, both to identify remediable causes of hypertension and to ensure that the prescribed antihypertensive agents are not associated with an adverse outcome of pregnancy (e.g., ACE inhibitors, angiotensin-receptor blockers). α-Methyldopa, labetalol, and nifedipine are the most commonly used medications for the treatment of chronic hypertension in pregnancy. α-Methyldopa is a relatively poor antihypertensive drug, but it has a long record of safe use in pregnancy. With the development of newer antihypertensives, also with long records of safe use in pregnancy, and with improved methods of fetal surveillance, we no longer recommend α-methyldopa as first-line therapy for the management of chronic hypertension. Baseline evaluation of renal function is necessary to help differentiate the effects of chronic hypertension versus superimposed preeclampsia, should the hypertension worsen during pregnancy. There are no convincing data that demonstrate that treatment of mild chronic hypertension improves perinatal outcome.

Gestational Hypertension

This is the development of elevated blood pressure during pregnancy or in the first 24 h post-partum in the absence of preexisting chronic hypertension or proteinuria. Mild gestational hypertension that does not progress to preeclampsia has not been associated with adverse pregnancy outcome or adverse long-term prognosis.

RENAL DISEASE

(See also Chaps. 278 and 286)

Normal pregnancy is characterized by an increase in glomerular filtration rate and creatinine clearance. This occurs secondary to a rise in renal plasma flow and increased glomerular filtration pressures. Patients with underlying renal disease and hypertension may expect a worsening of hypertension during pregnancy. If superimposed preeclampsia develops, the additional endothelial injury results in a capillary leak syndrome that may make the management challenging. In general, patients with underlying renal disease and hypertension benefit from aggressive management of blood pressure. Preconception counseling is also essential for these patients so that accurate risk assessment and medication changes can occur prior to pregnancy. In general, a prepregnancy serum creatinine level <133 μmol/L (<1.5 mg/dL) is associated with a favorable prognosis. When renal disease worsens during pregnancy, close collaboration between the internist and the maternal-fetal medicine specialist is essential so that decisions regarding delivery can be weighed in the context of sequelae of prematurity for the neonate versus long-term sequelae for the mother with respect to future renal function.

CARDIAC DISEASE

Valvular Heart Disease

(See also Chap. 237) This is the most common cardiac problem complicating pregnancy.

MITRAL STENOSIS

This is the valvular disease most likely to cause death during pregnancy. The pregnancy-induced

increase in blood volume, cardiac output, and tachycardia can increase the transmitral pressure gradient and cause pulmonary edema in women with mitral stenosis. Pregnancy associated with long-standing mitral stenosis may result in pulmonary hypertension. Sudden death has been reported when hypovolemia occurs. Careful control of heart rate, especially during labor and delivery, minimizes the impact of tachycardia and reduced ventricular filling times on cardiac function. Pregnant women with mitral stenosis are at increased risk for the development of atrial fibrillation and other tachyarrhythmias. Medical management of severe mitral stenosis and atrial fibrillation with digoxin and beta blockers is recommended. Balloon valvulotomy can be carried out during pregnancy. The immediate postpartum period is a time of particular concern secondary to rapid volume shifts. Careful monitoring of cardiac and fluid status should be observed.

MITRAL REGURGITATION AND AORTIC REGURGITATION AND STENOSIS

These are generally well tolerated during pregnancy. The pregnancy-induced decrease in systemic vascular resistance reduces the risk of cardiac failure with these conditions. As a rule, mitral valve prolapse does not present problems for the pregnant patient, and aortic stenosis, unless very severe, is well tolerated. In the most severe cases of aortic stenosis, limitation of activity or balloon valvuloplasty may be indicated.

Congenital Heart Disease

(See also Chap. 236) The presence of a congenital cardiac lesion in the mother increases the risk of congenital cardiac disease in the newborn. Prenatal screening of the fetus for congenital cardiac disease with ultrasound is recommended. Atrial or ventricular septal defect is usually well tolerated during pregnancy in the absence of pulmonary hypertension, provided that the womans prepregnancy cardiac status is favorable. Use of air filters on IV sets during labor and delivery in patients with intracardiac shunts is recommended.

Other Cardiac Disorders

Supraventricular tachycardia (Chap. 233) is a common cardiac complication of pregnancy. Treatment is the same as in the nonpregnant patient, and fetal tolerance of medications such as adenosine and calcium channel blockers is acceptable. When necessary, pharmacologic or electric cardioversion may be performed to improve cardiac performance and reduce symptoms. This is generally well tolerated by mother and fetus.

Peripartum cardiomyopathy (Chap. 238) is an uncommon disorder of pregnancy associated with myocarditis, and its etiology remains unknown. Treatment is directed toward symptomatic relief and improvement of cardiac function. Many patients recover completely; others are left with a progressive dilated cardiomyopathy. Recurrence in a subsequent pregnancy has been reported, and women who do not have normal baseline left ventricular function after an episode of peripartum cardiomyopathy should be counseled to avoid pregnancy.

Specific High-Risk Cardiac Lesions

MARFAN SYNDROME

(See also Chap. 363) This is an autosomal dominant disease, associated with a high risk of maternal morbidity. Approximately 15% of pregnant women with Marfan syndrome develop a major cardiovascular manifestation during pregnancy, with almost all women surviving. An aortic root diameter <40 mm is considered to be associated with a favorable outcome of pregnancy. Prophylactic therapy with beta blockers has been advocated, although large-scale clinical trials in pregnancy have not been performed. Ehlers-Danlos syndrome (EDS) may be associated with premature labor, and in type IV EDS, there is increased risk of uterine rupture.

PULMONARY HYPERTENSION

(See also Chap. 250) Maternal mortality in the setting of severe pulmonary hypertension is high, and primary pulmonary hypertension is a contraindication to pregnancy. Termination of pregnancy may be advisable in these circumstances to preserve the life of the mother. In the Eisenmenger syndrome, i.e., the combination of pulmonary hypertension with right-to-left shunting due to congenital abnormalities (Chap. 236), maternal and fetal death occur frequently. Systemic hypotension may occur after blood loss, prolonged Valsalva maneuver, or regional anesthesia; sudden death secondary to hypotension is a dreaded complication. Management of these patients is challenging, and invasive hemodynamic monitoring during labor and delivery is recommended in severe cases.

In patients with pulmonary hypertension, vaginal delivery is less stressful hemodynamically than cesarean section, which should be reserved for accepted obstetric indications.

DEEP VENOUS THROMBOSIS AND PULMONARY EMBOLISM

(See also Chap. 262)

A hypercoagulable state is characteristic of pregnancy, and deep venous thrombosis (DVT) occurs in about 1 in 2000 pregnancies. Pregnancy is associated with an increase in procoagulants such as factors V and VII, and a decrease in anticoagulant activity, including proteins C and S. Pulmonary

embolism is one of the most common causes of maternal death in the United States. In pregnant women, DVT occurs much more commonly in the left leg than in the right leg, due to compression of the left iliac vein by the iliac artery and the uterus. Activated protein C resistance caused by the factor V Leiden mutation increases the risk for DVT and pulmonary embolism during pregnancy. Approximately 25% of women with DVT during pregnancy carry the factor V Leiden allele. Additional genetic mutations associated with DVT during pregnancy include the prothrombin G20210A mutation (heterozygotes and homozygotes) and the methylenetetrahydrofolate reductase C677T mutation (homozygotes).

TREATMENT: DEEP VENOUS THROMBOSIS

Aggressive diagnosis and management of DVT and suspected pulmonary embolism optimize the outcome for mother and fetus. In general, all diagnostic and therapeutic modalities afforded the nonpregnant patient should be utilized in pregnancy. Anticoagulant therapy with low-molecular weight heparin (LMWH) or unfractionated heparin is indicated in pregnant women with DVT. LMWH may be associated with an increased risk of epidural hematoma in women receiving an epidural anesthetic in labor. Four weeks prior to anticipated delivery, LMWH should be switched to unfractionated heparin. Warfarin therapy is contraindicated in the first trimester due to its association with fetal chondrodysplasia punctata. In the second and third trimesters, warfarin may cause fetal optic atrophy and mental retardation. When DVT occurs in the postpartum period, LMWH therapy for 7–10 days may be followed by warfarin therapy for 3–6 months. Warfarin is not contraindicated in breast-feeding women. For women at moderate or high risk of DVT, mechanical or pharmacologic prophylaxis is warranted if they have a cesarean delivery.

ENDOCRINE DISORDERS

Diabetes Mellitus

(See also Chap. 344) In pregnancy, the fetoplacental unit induces major metabolic changes, the purpose of which is to shunt glucose and amino acids to the fetus while the mother uses ketones and triglycerides to fuel her metabolic needs. These metabolic changes are accompanied by maternal insulin resistance, caused in part by placental production of steroids, a growth hormone variant, and placental lactogen. Although pregnancy has been referred to as a state of "accelerated starvation," it is better characterized as "accelerated ketosis." In pregnancy, after an overnight fast, plasma glucose is lower by 0.8–1.1 mmol/L (15–20 mg/dL) than in the nonpregnant state. This is due to the use of glucose by the fetus. In early pregnancy, fasting may result in circulating glucose concentrations in the range of 2.2 mmol/L (40 mg/dL) and may be associated with symptoms of hypoglycemia. In contrast to the decrease in maternal glucose concentration, plasma hydroxybutyrate and acetoacetate levels rise to two to four times normal after a fast.

TREATMENT: DIABETES MELLITUS IN PREGNANCY

Pregnancy complicated by diabetes mellitus is associated with higher maternal and perinatal morbidity and mortality rates. Preconception counseling and treatment are important for the diabetic patient contemplating pregnancy and can reduce the risk of congenital malformations and improve pregnancy outcome. Folate supplementation reduces the incidence of fetal neural tube defects, which occur with greater frequency in fetuses of diabetic mothers. In addition, optimizing glucose control during key periods of organogenesis reduces other congenital anomalies including sacral agenesis, caudal dysplasia, renal agenesis, and ventricular septal defect.

Once pregnancy is established, glucose control should be managed more aggressively than in the nonpregnant state. In addition to dietary changes, this requires more frequent blood glucose monitoring and often involves additional injections of insulin or conversion to an insulin pump. Fasting blood glucose levels should be maintained at <5.8 mmol/L (<105 mg/dL) with no values >7.8 mmol/L (140 mg/dL). Commencing in the third trimester, regular surveillance of maternal glucose control as well as assessment of fetal growth (obstetric sonography) and fetoplacental oxygenation (fetal heart rate monitoring or biophysical profile) optimizes pregnancy outcome. Pregnant diabetic patients without vascular disease are at greater risk for delivering a macrosomic fetus, and attention to fetal growth via clinical and ultrasound examination is important. Fetal macrosomia is associated with an increased risk of maternal and fetal birth trauma, including permanent newborn Erbs palsy. Pregnant women with diabetes have an increased risk of developing preeclampsia, and those with vascular disease are at greater risk for developing intrauterine growth restriction, which is associated with an increased risk of fetal and neonatal death. Excellent pregnancy outcomes in patients with diabetic nephropathy and proliferative retinopathy have been reported with aggressive glucose control and intensive maternal and fetal surveillance.

Glycemic control may become more difficult to achieve as pregnancy progresses due to an increase in insulin resistance. Because of delayed pulmonary maturation of the fetuses of diabetic mothers, early delivery should be avoided unless there is biochemical evidence of fetal lung maturity. In general, efforts to control glucose and avoid preterm delivery result in the best overall outcome for both mother and newborn. Preterm delivery is generally performed only for the usual obstetric indications (e.g., preeclampsia, fetal growth restriction, non-reassuring fetal testing) or for worsening maternal renal or

active proliferative retinopathy.

Gestational Diabetes

Gestational diabetes occurs in approximately 4% of pregnancies. All pregnant women should be screened for gestational diabetes unless they are in a low-risk group. Women at low risk for gestational diabetes are those <25 years of age; those with a body mass index < 25 kg/m^2, no maternal history of macrosomia or gestational diabetes, and no diabetes in a first-degree relative; and those not members of a high-risk ethnic group (African American, Hispanic, Native American). A typical two-step strategy for establishing the diagnosis of gestational diabetes involves administration of a 50-g oral glucose challenge with a single serum glucose measurement at 60 min. If the plasma glucose is <7.8 mmol/L (<140 mg/dL), the test is considered normal. Plasma glucose > 7.8 mmol/L (>140 mg/dL) warrants administration of a 100-g oral glucose challenge with plasma glucose measurements obtained in the fasting state and at 1, 2, and 3 h. Normal values are plasma glucose concentrations <5.8 mmol/L (<105 mg/dL), 10.5 mmol/L (190 mg/dL), 9.1 mmol/L (165 mg/dL), and 8.0 mmol/L (145 mg/dL), respectively. Some centers have adopted more sensitive criteria, using <7.5 mmol/L (<130 mg/dL) as the screening threshold and values of <5.3 mmol/L (<95 mg/dL), <10 mmol/L (<180 mg/dL), <8.6 mmol/L (<155 mg/dL), and <7.8 mmol/L (<140 mg/dL) as the upper norms for a 3-h glucose tolerance test. Adverse pregnancy outcomes for mother and fetus appear to increase with glucose as a continuous variable, making it challenging to define the optimal threshold for establishing the diagnosis of gestational diabetes.

Pregnant women with gestational diabetes are at increased risk of stillbirth, preeclampsia, and delivering infants who are large for their gestational age with resulting birth lacerations, shoulder dystocia, and birth trauma including brachial plexus injury. Their fetuses are at risk of hypoglycemia, hyperbilirubinemia, and polycythemia. Tight control of blood sugar during pregnancy and labor can reduce these risks.

TREATMENT: GESTATIONAL DIABETES

Treatment of gestational diabetes with a two-step strategy of dietary intervention followed by insulin injections if diet alone does not adequately control blood sugar [fasting glucose < 5.6 mmol/L (<100 mg/dL) and 2-h post-prandial <7.0 mmol/L (<126 mg/dL)] is associated with a decreased risk of birth trauma for the fetus. Oral hypoglycemic agents such as glyburide and metformin have become more commonly utilized for managing gestational diabetes refractory to nutritional management. For women with gestational diabetes, within the 10 years after the index pregnancy there is a 40% risk of being diagnosed with diabetes. In women with a history of gestational diabetes, exercise, weight loss, and treatment with metformin reduce the risk of developing diabetes. All women with a history of gestational diabetes should be counseled about prevention strategies and evaluated regularly for diabetes.

OBESITY

(See also Chap. 78) Pregnant women who are obese have an increased risk of stillbirth, congenital fetal malformations, gestational diabetes, preeclampsia, urinary tract infections, and post-date deliveries. Women contemplating pregnancy should attempt to achieve a healthy weight prior to conception. For morbidly obese women who have not been able to achieve weight loss with lifestyle changes, bariatric surgery may result in weight loss and improve pregnancy outcomes. Following bariatric surgery, women should delay conception for one year to avoid pregnancy during an interval of rapid metabolic changes.

Thyroid Disease

(See also Chap. 341) In pregnancy, the estrogen-induced increase in thyroxine-binding globulin increases circulating levels of total T_3 and total T4. The normal range of circulating levels of free T_4, free T_3, and thyroid-stimulating hormone (TSH) remain unaltered by pregnancy.

The thyroid gland normally enlarges during pregnancy. Maternal hyperthyroidism occurs at a rate of ~2 per 1000 pregnancies and is generally well tolerated by pregnant women. Clinical signs and symptoms should alert the physician to the occurrence of this disease. Many of the physiologic adaptations to pregnancy may mimic subtle signs of hyperthyroidism. Although pregnant women are able to tolerate mild hyperthyroidism without adverse sequelae, more severe hyperthyroidism can cause spontaneous abortion or premature labor, and thyroid storm is associated with a significant risk of maternal mortality.

TREATMENT: HYPERTHYROIDISM IN PREGNANCY

Hyperthyroidism in pregnancy should be aggressively evaluated and treated. It is most commonly caused by Graves disease, but autonomously functioning nodules, gestational trophoblastic disease, thyroiditis, and hyperemesis gravidarum should also be considered. Methimazole crosses the placenta to a greater degree than propylthiouracil and has been associated with fetal aplasia cutis. However, propylthiouracil can be associated with liver failure. Some experts recommend propylthiouracil in the first trimester and methimazole thereafter. Radioiodine should not be used during pregnancy, either for scanning or treatment, because of effects on the fetal thyroid. In emergent circumstances, additional

treatment with beta blockers and a saturated solution of potassium iodide may be necessary. Hyperthyroidism is most difficult to control in the first trimester of pregnancy and easiest to control in the third trimester.

Testing for *hypothyroidism* using TSH measurements before or early in pregnancy may be warranted in symptomatic women and women with a personal or family history of thyroid disease. Using this case-finding approach, about 30% of pregnant women with mild hypothyroidism will remain undiagnosed, leading some to recommend universal screening. Children born to women with an elevated serum TSH (and a normal total thyroxine) during pregnancy have impaired performance on neuropsychologic tests. The goal of therapy for hypothyroidism is to maintain the serum TSH in the normal range, and thyroxine is the drug of choice. During pregnancy, the dose of thyroxine required to keep the TSH in the normal range rises. In one study, the mean replacement dose of thyroxine required to maintain the TSH in the normal range was 0.1 mg daily before pregnancy, and it increased to 0.15 mg daily during pregnancy. Since the increased thyroxine requirement occurs as early as the fifth week of pregnancy, one approach is to increase the thyroxine dose by 30% as soon as pregnancy is diagnosed and then adjust the dose by serial measurements of TSH.

HEMATOLOGIC DISORDERS

Pregnancy has been described as a state of physiologic anemia. Part of the reduction in hemoglobin concentration is dilutional, but iron and folate deficiencies are the major causes of correctable anemia during pregnancy.

In populations at high risk for hemoglobinopathies (Chap. 104), hemoglobin electrophoresis should be performed as part of the prenatal screen. Hemoglobinopathies can be associated with increased maternal and fetal morbidity and mortality. Management is tailored to the specific hemoglobinopathy and is generally the same for both pregnant and nonpregnant women. Prenatal diagnosis of hemoglobinopathies in the fetus is readily available and should be discussed with prospective parents either prior to or early in pregnancy.

Thrombocytopenia occurs commonly during pregnancy. The majority of cases are benign gestational thrombocytopenias, but the differential diagnosis should include immune thrombocytopenia (Chap. 115), thrombotic thrombocytopenic purpura, and preeclampsia. Maternal thrombocytopenia may also be caused by disseminated intravascular coagulation (DIC), which is a consumptive coagulopathy characterized by thrombocytopenia, prolonged prothrombin time (PT) and activated partial thromboplastin time (aPTT), elevated fibrin degradation products, and a low fibrinogen concentration. Several catastrophic obstetric events are associated with the development of DIC, including retention of a dead fetus, sepsis, abruptio placenta, and amniotic fluid embolism.

NEUROLOGIC DISORDERS

Headache appearing during pregnancy is usually due to migraine (Chap. 14), a condition that may worsen, improve, or be unaffected by pregnancy. A new or worsening headache, particularly if associated with visual blurring, may signal eclampsia (above) or pseudotumor cerebri (benign intracranial hypertension) (Chap. 28); diplopia due to a sixth-nerve palsy suggests pseudotumor cerebri. The risk of seizures in patients with epilepsy increases in the postpartum period but not consistently during pregnancy; management is discussed in Chap. 369. The risk of stroke is generally thought to increase during pregnancy because of a hypercoagulable state; however, studies suggest that the period of risk occurs primarily in the postpartum period and that both ischemic and hemorrhagic strokes may occur at this time. Guidelines for use of heparin therapy are summarized above (see "Deep Venous Thrombosis and Pulmonary Embolism"); warfarin is teratogenic and should be avoided.

The onset of a new movement disorder during pregnancy suggests chorea gravidarum, a variant of Sydenhams chorea associated with rheumatic fever and streptococcal infection (Chap. 322); the chorea may recur with subsequent pregnancies. Patients with preexisting multiple sclerosis (MS) (Chap. 380) experience a gradual decrease in the risk of relapses as pregnancy progresses and, conversely, an increase in attack risk during the postpartum period. Beta interferons should *not* be administered to pregnant MS patients, but moderate or severe relapses can be safely treated with pulse glucocorticoid therapy. Finally, certain tumors, particularly pituitary adenoma and meningioma (Chap. 379), may manifest during pregnancy because of accelerated growth, possibly driven by hormonal factors.

Peripheral nerve disorders associated with pregnancy include Bells palsy (idiopathic facial paralysis) (Chap. 384), which is approximately threefold more likely to occur during the third trimester and immediate postpartum period than in the general population. Therapy with glucocorticoids should follow the guidelines established for nonpregnant patients. Entrapment neuropathies are common in the later stages of pregnancy, presumably as a result of fluid retention. Carpal tunnel syndrome (median nerve) presents as pain and paresthesia in the hand, often worse at night, and later with weakness in the thenar muscles. Treatment is generally conservative; wrist splints may be helpful, and glucocorticoid injections or surgical section of the carpal tunnel can usually be postponed. Meralgia

paresthetica (lateral femoral cutaneous nerve) consists of pain and numbness in the lateral aspect of the thigh without weakness. Patients are usually reassured to learn that these symptoms are benign and can be expected to remit spontaneously after the pregnancy has been completed. Restless leg syndrome is the most common peripheral nerve and movement disorder in pregnancy. Disordered iron metabolism is the suspected etiology. Management is expectant in most cases.

GASTROINTESTINAL AND LIVER DISEASE

Up to 90% of pregnant women experience nausea and vomiting during the first trimester of pregnancy. Hyperemesis gravidarum is a severe form that prevents adequate fluid and nutritional intake and may require hospitalization to prevent dehydration and malnutrition.

Crohns disease may be associated with exacerbations in the second and third trimesters. Ulcerative colitis is associated with disease exacerbations in the first trimester and during the early postpartum period. Medical management of these diseases during pregnancy is identical to the management in the nonpregnant state (Chap. 295).

Exacerbation of gall bladder disease is commonly observed during pregnancy. In part, this may be due to pregnancy-induced alteration in the metabolism of bile and fatty acids. Intrahepatic cholestasis of pregnancy is generally a third-trimester event. Profound pruritus may accompany this condition, and it may be associated with increased fetal mortality. Placental bile salt deposition may contribute to progressive uteroplacental insufficiency. Therefore, regular fetal surveillance should be undertaken once the diagnosis of intrahepatic cholestasis is made. Favorable results with ursodiol have been reported.

Acute fatty liver is a rare complication of pregnancy. Frequently confused with the HELLP syndrome (see "Preeclampsia" above) and severe preeclampsia, the diagnosis of acute fatty liver of pregnancy may be facilitated by imaging studies and laboratory evaluation. Acute fatty liver of pregnancy is generally characterized by markedly increased levels of bilirubin and ammonia and by hypoglycemia. Management of acute fatty liver of pregnancy is supportive; recurrence in subsequent pregnancies has been reported.

All pregnant women should be screened for hepatitis B. This information is important for pediatricians after delivery of the infant. All infants receive hepatitis B vaccine. Infants born to mothers who are carriers of hepatitis B surface antigen should also receive hepatitis B immune globulin as soon after birth as possible and preferably within the first 72 h. Screening for hepatitis C is recommended for individuals at high risk for exposure.

INFECTIONS

Bacterial Infections

Other than bacterial vaginosis, the most common bacterial infections during pregnancy involve the urinary tract (Chap. 288). Many pregnant women have asymptomatic bacteriuria, most likely due to stasis caused by progestational effects on ureteral and bladder smooth muscle and later in pregnancy due to compression effects of the enlarging uterus. In itself, this condition is not associated with an adverse outcome of pregnancy. However, if asymptomatic bacteriuria is left untreated, symptomatic pyelonephritis may occur. Indeed, ~75% of cases of pregnancy-associated pyelonephritis are the result of untreated asymptomatic bacteriuria. All pregnant women should be screened with a urine culture for asymptomatic bacteriuria at the first prenatal visit. Subsequent screening with nitrite/leukocyte esterase strips is indicated for high-risk women, such as those with sickle cell trait or a history of urinary tract infections. All women with positive screens should be treated. Pregnant women who develop pyelonephritis need careful monitoring, including inpatient IV antibiotic administration due to the elevated risk of urosepsis and acute respiratory distress syndrome in pregnancy.

Abdominal pain and fever during pregnancy create a clinical dilemma. The diagnosis of greatest concern is intrauterine amniotic infection. While amniotic infection most commonly follows rupture of the membranes, this is not always the case. In general, antibiotic therapy is not recommended as a temporizing measure in these circumstances. If intrauterine infection is suspected, induced delivery with concomitant antibiotic therapy is generally indicated. Intrauterine amniotic infection is most often caused by pathogens such as *Escherichia coli* and group B streptococcus. In high-risk patients at term or in preterm patients, routine intrapartum prophylaxis of group B streptococcal (GBS) disease is recommended. Penicillin G and ampicillin are the drugs of choice. In penicillin-allergic patients, with a low risk of anaphylaxis, cefazolin is recommended. If the patient is at high risk of anaphylaxis, vancomycin is recommended. If the organism is known to be sensitive to clindamycin, this antibiotic may be used. For the reduction of neonatal morbidity due to GBS, universal screening of pregnant women for GBS between 35 and 37 weeks gestation with intrapartum antibiotic treatment of infected women is recommended.

Postpartum infection is a significant cause of maternal morbidity and mortality. Postpartum endomyometritis is more common after cesarean delivery than vaginal delivery and develops in 5% of women after elective repeat cesarean section and in up to 25% after emergency cesarean section,

following prolonged labor. Prophylactic antibiotics should be given to all patients undergoing cesarean section, and administration 30–60 minutes prior to skin incision may be preferable to administration at the time of umbilical cord clamping. As most cases of postpartum endomyometritis are polymicrobial, broad-spectrum antibiotic coverage with a penicillin, aminoglycoside, and metronidazole is recommended (Chap. 164). Most cases resolve within 72 h. Women who do not respond to antibiotic treatment for postpartum endomyometritis should be evaluated for septic pelvic thrombophlebitis. Imaging studies may be helpful in establishing the diagnosis, which is primarily a clinical diagnosis of exclusion. Patients with septic pelvic thrombophlebitis generally have tachycardia out of proportion to their fever and respond rapidly to intravenous administration of heparin.

All patients are screened prenatally for gonorrhea and chlamydial infections, and the detection of either should result in prompt treatment. Ceftriaxone and azithromycin are the agents of choice (Chaps. 144 and 176).

Viral Infections

CYTOMEGALOVIRUS INFECTION

Viral infection in pregnancy presents a significant challenge. The most common cause of congenital viral infection in the United States is cytomegalovirus (CMV) (Chap. 182). As many as 50–90% of women of childbearing age have antibodies to CMV, but only rarely does CMV reactivation result in neonatal infection. More commonly, primary CMV infection during pregnancy creates a risk of congenital CMV. No currently accepted treatment of CMV during pregnancy has been demonstrated to protect the fetus effectively. Moreover, it is difficult to predict which fetus will sustain a life-threatening CMV infection. Severe CMV disease in the newborn is characterized most often by petechiae, hepatosplenomegaly, and jaundice. Chorioretinitis, microcephaly, intracranial calcifications, hepatitis, hemolytic anemia, and purpura may also develop. CNS involvement, resulting in the development of psychomotor, ocular, auditory, and dental abnormalities over time, has been described.

RUBELLA

(See also Chap. 193) Rubella virus is a known teratogen; first-trimester rubella carries a high risk of fetal anomalies, though the risk decreases significantly later in pregnancy. Congenital rubella may be diagnosed by percutaneous umbilical blood sampling with the detection of IgM antibodies in fetal blood. All pregnant women should be screened for their immune status to rubella. Indeed, all women of childbearing age, regardless of pregnancy status, should have their immune status for rubella verified and be immunized if necessary. The incidence of congenital rubella in the United States is extremely low.

HERPESVIRUS

(See also Chap. 179) The acquisition of genital herpes during pregnancy is associated with spontaneous abortion, prematurity, and congenital and neonatal herpes. A recent cohort study of pregnant women without evidence of previous herpes infection demonstrated that ~2% of the women acquired a new herpes infection during the pregnancy. Approximately 60% of the newly infected women had no clinical symptoms. Infection occurred equally in all three trimesters. If herpes seroconversion occurred early in pregnancy, the risk of transmission to the newborn was very low. In women who acquired genital herpes shortly before delivery, the risk of transmission was high. The risk of active genital herpes lesions at term can be reduced by prescribing acyclovir for the last four weeks of pregnancy to women who have had their first episode of genital herpes during the pregnancy.

Herpesvirus infection in the newborn can be devastating. Disseminated neonatal herpes carries with it high mortality and morbidity rates from CNS involvement. It is recommended that pregnant women with active genital herpes lesions at the time of presentation in labor be delivered by cesarean section.

PARVOVIRUS

(See also Chap. 184) Parvovirus infection (human parvovirus B19) may occur during pregnancy. It rarely causes sequelae, but susceptible women infected during pregnancy may be at risk for fetal hydrops secondary to erythroid aplasia and profound anemia.

HIV INFECTION

(See also Chap. 189) The predominant cause of HIV infection in children is transmission of the virus from the mother to the newborn during the perinatal period. HIV infection with high maternal viral load; low maternal CD4+ T cell count; prolonged labor; prolonged length of membrane rupture; and the presence of other genital tract infections, such as syphilis or herpes, increase the risk of transmission. Advances in antiretroviral therapy now make it possible for many pregnant women to have no detectable viral load as they approach term. These women may elect to attempt a vaginal birth. To reduce the risk of mother-to-newborn transmission, women should be treated during the intrapartum interval with zidovudine and the newborn should also be treated after birth. Breast-feeding may also transmit HIV to the newborn and is therefore contraindicated in most developed countries for HIV-infected mothers.

SUMMARY

Maternal death is defined as the death occurring during pregnancy or within 42 days of completion of the pregnancy from a cause related to or aggravated by the pregnancy, but not from accident or incidental causes. Maternal mortality has decreased steadily during the past 70 years. The maternal death rate has decreased from nearly 600/100,000 live births in 1935 to 13/100,000 live births in 2006. The most common causes of maternal death in the United States today are pulmonary embolism, obstetric hemorrhage, hypertension, sepsis, cardiovascular conditions including peripartum cardiomyopathy, and ectopic pregnancy. With improved diagnostic and therapeutic modalities as well as with advances in the treatment of infertility, more patients with medical complications will be seeking, and be in need of, complex obstetric care. Improving outcome of pregnancy in these women will be best obtained by assembling a team of internists, specialists in maternal-fetal medicine (high-risk obstetrics), and anesthesiologists to counsel these patients about the risks of pregnancy and to plan their treatment prior to conception. The importance of preconception counseling cannot be overstated. It is the responsibility of all physicians caring for women in the reproductive age group to assess their patients reproductive plans as part of their overall health evaluation.

Global Considerations

The maternal mortality ratio in the United States is about 1 in 7500 live births. In some countries in Sub-Saharan Africa and Southeast Asia, the maternal mortality ratio is about 1 in 100 to 1 in 200 live births. The most common cause of maternal mortality in these countries is maternal hemorrhage. The high maternal death rate is due, in part, to inadequate resources, including lack of contraceptive and family-planning services, insufficient number of skilled birth attendants, limited availability of tocolytics, poor nutrition, and a high burden of infectious disease. The high rate of death from peripartum hemorrhage is due, in part, to inadequate access to anesthesia resources, blood-transfusion services, antibiotics, and skilled pelvic surgeons. Maternal death is a global public-health tragedy that could be improved with the application of modest resources.

FURTHER READINGS

Busnell CD et al: Migraines during pregnancy linked to stroke and vascular diseases: US population based case-control study. BMJ 338: b664, 2009

Hogan MC et al: Maternal mortality for 181 countries, 1980-2008: A systematic analysis of progress towards Millennium Development Goal 5. Lancet 375:1609, 2010[PMID: 20382417]

Maggard MA et al: Pregnancy and fertility following bariatric surgery. JAMA 300:2286, 2008[PMID: 19017915]

Marik PA, Plante LA: Venous thromboembolic disease and pregnancy. N Engl J Med 359:2025, 2008[PMID: 18987370]

Ratner RE et al: Prevention of diabetes in women with a history of gestational diabetes: effects of metformin and lifestyle interventions. J Clin Endocrinol Metab 93:4774, 2008[PMID: 18826999]

Rowan JA et al: Metformin versus insulin for the treatment of gestational diabetes. N Engl J Med 358:2003, 2009

Schatz M, Dombrowksi MP: Asthma in pregnancy. N Engl J Med 360:1862, 2009[PMID: 19403904]

Vaidya B et al: Detection of thyroid dysfunction in early pregnancy: universal screening or targeted high-risk case finding? J Clin Endocrinol Metab 92:203, 2007[PMID: 17032713]

Viske BE et al: Preeclampsia and the risk of end-stage renal disease. N Engl J Med 359: 800, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MEDICAL EVALUATION OF THE SURGICAL PATIENT: INTRODUCTION

Cardiovascular and pulmonary complications continue to account for major morbidity and mortality in patients undergoing noncardiac surgery. Emerging evidence-based practices dictate that the internist should perform an individualized evaluation of the surgical patient to provide an accurate preoperative risk assessment and stratification to guide optimum perioperative risk-reduction strategies. This chapter reviews cardiovascular and pulmonary preoperative risk assessment, targeting intermediate- and high-risk patients to strategically guide perioperative therapies to improve outcome. It also reviews perioperative management and prophylaxis of diabetes mellitus, endocarditis, and venous thromboembolism.

ANESTHETICS

Mortality is low with safe delivery of modern anesthesia, especially in low-risk patients undergoing low-risk surgery (Table 8-1). Inhaled anesthetics have predictable circulatory and respiratory effects; all decrease arterial pressure in a dose-dependent manner by reducing sympathetic tone, causing systemic vasodilation, myocardial depression, and decreased cardiac output. Inhaled anesthetics also cause respiratory depression with diminished responses to both hypercapnia and hypoxemia in a dose-dependent manner, and they have a variable effect on heart rate. In combination with neuromuscular blockade, inhaled anesthetic agents also cause reduction in functional residual lung capacity due to loss of diaphragmatic and intercostal muscle function. This decreases lung volume, which may lead to atelectasis in the dependent lung regions and, in turn, may result in arterial hypoxemia from ventilation-perfusion mismatch as well as an increased risk of postoperative pulmonary complications.

Table 8-1 Surgery: Gradation of Risk of Common Noncardiac Surgical Procedures

Higher	■ Emergent major operations, especially elderly ■ Aortic and other noncarotid major vascular surgery (endovascular and nonendovascular) ■ Prolonged surgery associated with large fluid shift and/or blood loss
Intermediate	■ Major thoracic surgery ■ Major abdominal surgery ■ Carotid endarterectomy surgery ■ Head/neck surgery ■ Orthopedic surgery ■ Prostate surgery
Lower	■ Eye, skin, and superficial surgery ■ Endoscopic procedures

Source: From LA Fleisher et al.

Several meta-analyses have shown that overall mortality was lower in patients receiving neuroaxial anesthesia (epidural or spinal) as compared to general (inhaled) anesthesia. Lower rates of venous thrombosis, pulmonary embolism, pneumonia, and respiratory depression were also observed in patients who were provided neuroaxial anesthesia; however, there were no significant differences in cardiac events between the two approaches. A combination of neuroaxial blockade and general anesthesia is useful when it is desired to reduce general anesthesia requirements. Evidence from a meta-analysis of randomized controlled trials also supports postoperative epidural analgesia for the purpose of pain relief for more than 24 h.

EVALUATION OF INTERMEDIATE- TO HIGH-RISK PATIENTS

Simple, standardized preoperative screening questionnaires, such as the one shown in Table 8-2, have been developed for the purpose of identifying patients at intermediate or high risk who may benefit from a more detailed clinical evaluation. Evaluation of such patients for operation should always begin with a thorough history and physical examination and with a 12-lead resting ECG, in accordance with the American College of Cardiology/American Heart Association (ACC/AHA) guideline recommendations. The

history should focus on symptoms of occult cardiac or pulmonary disease. The urgency of the surgery should be determined, as true emergency procedures are associated with an unavoidably higher morbidity and mortality. Preoperative laboratory testing should be carried out only for specific clinical conditions based on the clinical examination. Thus, healthy patients of any age undergoing elective surgical procedures without coexisting medical conditions should not require any testing unless the degree of surgical stress may result in unusual changes from the baseline state.

Table 8-2 Standardized Preoperative Questionnaires[a]

1. Age, weight, height
2. Are you: Female and 55 years of age or older or male and 45 years of age of older?
If yes, are you also 70 years of age or older?
3. Do you take anticoagulant ("blood thinners") medications?
4. Do you have or have you had any of the following heart-related conditions?
Heart disease Heart attack within the last six months
Angina (chest pain)
Irregular heartbeat
Heart failure
5. Do you have or have you ever had any of the following? Rheumatoid arthritis
Kidney disease
Liver disease
Diabetes
6. Do you get short of breath when you lie flat? 7. Are you currently on oxygen treatment?
8. Do you have a chronic cough that produces any discharge or fluid?
9. Do you have lung problems or diseases?
10. Have you or any blood member of your family ever had a problem with any anesthesia other than nausea?
If yes, describe:
11. If female, is it possible that you could be pregnant? Pregnancy test:
Please list date of last menstrual period:

[a]University of Michigan Health System patient information report. Patients who answer yes to any of questions 2–9 should receive a more detailed clinical evaluation.

Source: Adapted from KK Tremper, P Benedict: Anesthesiology 92:1212, 2000; with permission.

Preoperative Cardiac Risk Assessment

Assessment of exercise tolerance in the prediction of in-hospital perioperative risk is most helpful in patients who self-report worsening, exercise-induced cardiopulmonary symptoms; those who may benefit from noninvasive or invasive cardiac testing regardless of scheduled surgical procedure; and those with known coronary artery disease (CAD) or with multiple risk factors who are able to exercise. For predicting perioperative events, poor exercise tolerance has been defined as the inability to walk four blocks or climb two flights of stairs at a normal pace or to meet a metabolic equivalent (MET) level of four (e.g., carrying objects of 15–20 lb or playing golf or doubles tennis) because of the development of dyspnea, angina, or excessive fatigue (Table 8-3).

Table 8-3 Functional Status

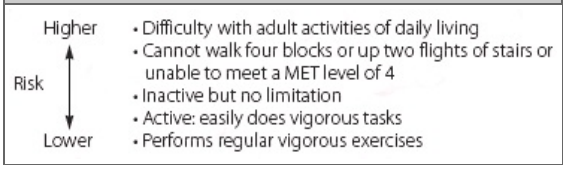

Source: From LA Fleisher et al.

Previous studies have prospectively compared several cardiac risk indices. Given its accuracy and

simplicity, the revised cardiac risk index (RCRI) (Table 8-4) is favored. The RCRI relies on the presence or absence of six identifiable predictive factors, which include high-risk surgery, ischemic heart disease, congestive heart failure, cerebrovascular disease, diabetes mellitus, and renal dysfunction. Each of these predictors is assigned one point. The risk of major cardiac events—defined as myocardial infarction, pulmonary edema, ventricular fibrillation or primary cardiac arrest, and complete heart block—can then be predicted. Based on the presence of none, one, two, three, or more of these clinical predictors, the rate of development of one of these major cardiac events is estimated to be 0.5, 1, 5, and 10%, respectively (Fig. 8-1). RCRI 0 has 0.4–0.5% risk of cardiac events; RCRI 1 has 0.9–1.3%; RCRI 2 has 4–6.6%; and RCRI ≥ 3 has 9–11% risk of cardiac events. The clinical utility of the RCRI is to identify patients with three or more predictors who are at higher risk (≥10%) for cardiac complications and who may benefit from further risk stratification with noninvasive cardiac testing or initiation of preoperative preventive medical management.

Table 8-4 Revised Cardiac Risk Index Clinical Markers

High-risk surgical procedures
Vascular surgery
Major intraperitoneal or intrathoracic procedures
Ischemic heart disease
History of myocardial infarction
Current angina considered to be ischemic
Requiring sublingual nitroglycerin
Positive exercise test
Pathological Q-waves on ECG
History of PTCA and/or CABG with current angina considered to be ischemic
Congestive heart failure
Left ventricular failure by physical examination
History of paroxysmal nocturnal dyspnea
History of pulmonary edema
S_3 gallop on cardiac auscultation
Bilateral rales on pulmonary auscultation
Pulmonary edema on chest x-ray
Cerebrovascular disease
History of transient ischemic attack
History of cerebrovascular accident
Diabetes mellitus
Treatment with insulin
Chronic renal insufficiency
Serum creatinine >2 mg/dL

Abbreviations: ECG, electrocardiogram; PTCA, per-cutaneous transluminal coronary angioplasty; CABG, coronary artery bypass grafting.

Source: Adapted from TH Lee et al, with permission.

Figure 8-1

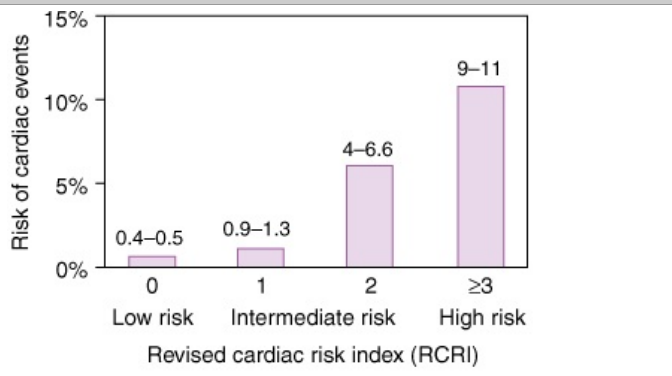

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Risk stratification based on the RCRI. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Cardiac events include myocardial infarction, pulmonary edema, ventricular fibrillation, cardiac asystole, and complete heart block. *(Adapted from TH Lee et al, with permission.)*

Preoperative Noninvasive Cardiac Testing for Risk Stratification

There is little evidence to support widespread application of preoperative noninvasive cardiac testing for all patients undergoing major surgery. Rather, a discriminate approach based on clinical risk categorization appears to be both useful clinically and cost-effective. There is potential benefit in identifying asymptomatic but high-risk patients, such as those with left main or left main–equivalent CAD or those with three-vessel CAD with poor left ventricular function who may benefit from coronary revascularization (Chap. 243). However, evidence does not support aggressive attempts to identify patients at intermediate risk with asymptomatic but advanced coronary artery disease, because coronary revascularization appears to offer little advantage over medical therapy.

An RCRI score ≥3 in patients with severe myocardial ischemia on stress testing should lead to consideration of coronary revascularization prior to noncardiac surgery. Noninvasive cardiac testing is most appropriate if it is anticipated that in the event of a strongly positive test a patient will meet guidelines for coronary angiography and coronary revascularization. Pharmacologic stress tests are more useful than exercise testing in patients with functional limitations. Dobutamine echocardiography and persantine, adenosine, or dobutamine nuclear perfusion testing (Chap. 229) have excellent negative predictive values (near 100%) but poor positive predictive values (< 20%) for identification of patients at risk for perioperative myocardial infarction or death. Thus, a negative study is reassuring, but a positive study is a relatively weak predictor of a "hard" perioperative cardiac event. A stepwise approach is illustrated in Fig. 8-2.

Figure 8-2

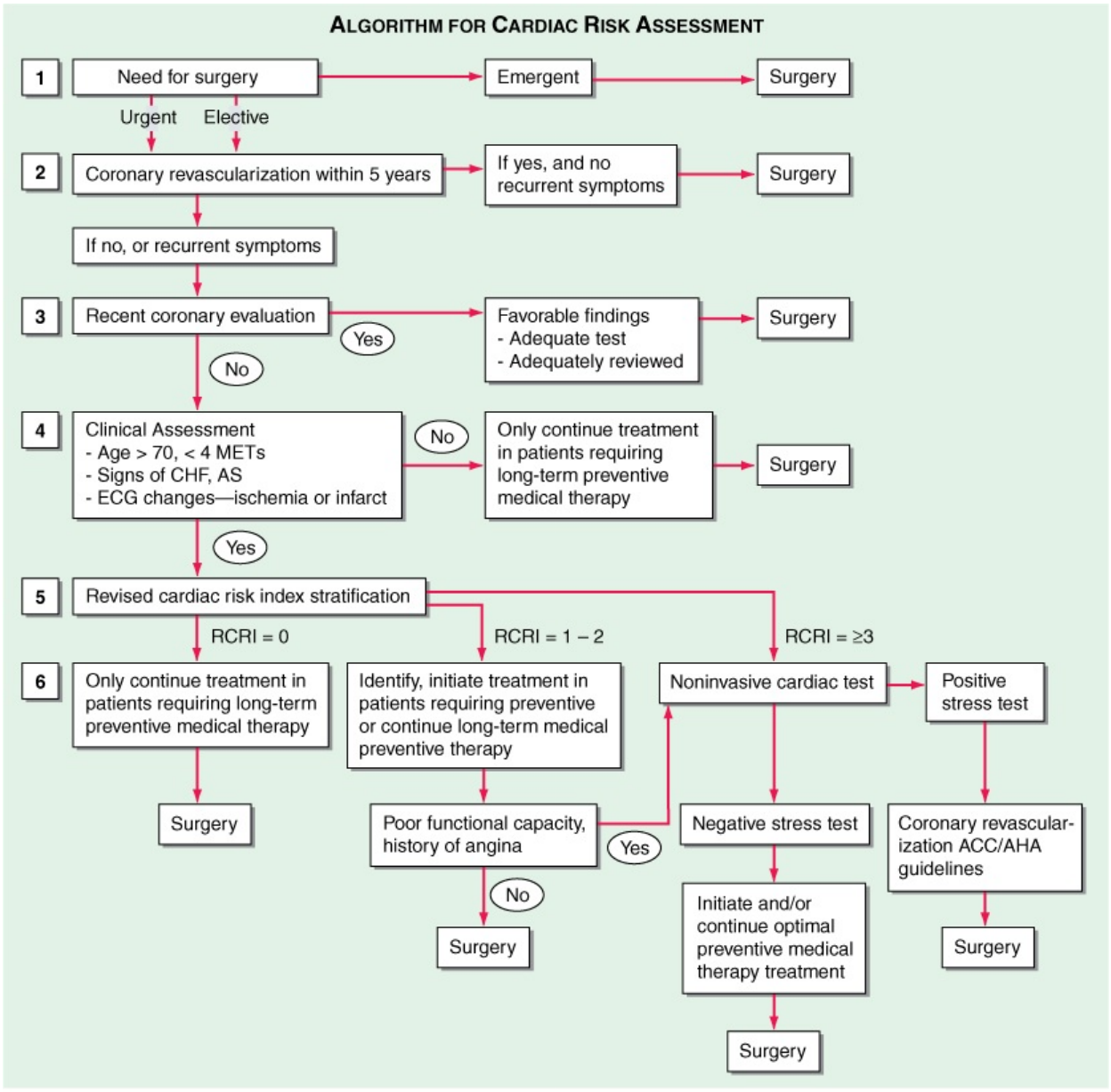

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Composite algorithm for cardiac risk assessment and stratification in patients undergoing noncardiac surgery. Stepwise clinical evaluation: [1] Emergency surgery; [2] Prior coronary revascularization; [3] Prior coronary evaluation; [4] Clinical assessment; [5] RCRI; [6] Risk modification strategies. Preventative medical therapy = beta blocker and statin therapy. RCRI, revised cardiac risk index. *(Adapted from LA Fleisher et al and TH Lee et al.)*

Risk Modification Using Preventive Strategies to Reduce Cardiac Risk

PERIOPERATIVE CORONARY REVASCULARIZATION

Currently, potential options for reducing perioperative cardiovascular risk include coronary artery revascularization and/or perioperative medical preventive therapies (Chap. 243). *Prophylactic* coronary revascularization with either coronary artery bypass grafting (CABG) or percutaneous coronary intervention (PCI) provides no short- or midterm survival benefit for patients *without* left main CAD or three-vessel CAD in the presence of poor left ventricular systolic function. Although PCI is associated with lower procedural risk than is CABG in the perioperative setting, the placement of a coronary artery stent in a short period of time prior to noncardiac surgery may increase the risk of bleeding during surgery if dual antiplatelet therapy (aspirin and thienopyridine) is administered or it increases the perioperative risk of myocardial infarction and cardiac death due to stent thrombosis if such therapy is withdrawn prematurely (Chap. 246). It is recommended that, if possible, noncardiac surgery be delayed 30–45 days after placement of a bare metal coronary stent and 365 days after a drug-eluting stent. For patients who *must* undergo noncardiac surgery early (> 14 days) after PCI, balloon angioplasty without stent placement appears to be a reasonable alternative because dual antiplatelet therapy is not necessary in such patients.

Perioperative Medical Preventive Therapies

Perioperative preventive medical therapy with β-adrenergic antagonists, HMG-CoA reductase inhibitors (statins), and aspirin have the goal of reducing perioperative adrenergic stimulation, ischemia, and inflammation, which are triggered during the perioperative period.

β-Adrenergic Antagonists

The use of perioperative beta blockade should be based on a thorough assessment of a patient's perioperative cardiac risk (RCRI ≥ 2). For patients with or without mild to moderate reactive airway disease, the cardioselective beta blocker of choice should be used and titrated to a target resting heart rate of 60–65 beats/minute. In intermediate- to high-risk patients without a long-term indication for beta blockers, the medications can be administered intravenously as a preoperative medication on the day of surgery, with a targeted heart rate of 60–65 beats/minute and continued for > 7 days (preferably 30 days) postoperatively. Intravenous preparations should be substituted for oral medication if patients are unable to take or absorb pills in the perioperative period. The results from the recent Perioperative Ischemic Evaluation (POISE) trial showed that although cardiac death, nonfatal myocardial infarction, or cardiac arrest was reduced in the patients who received metoprolol as compared to placebo, there was an increased incidence of mortality and stroke in the patients receiving metoprolol because of a high and rapid loading dose of metoprolol. The POISE trial highlights the importance for a clear risk and benefit assessment with careful initiation and titration to therapeutic efficacy of preoperative beta blockers in patients undergoing noncardiac surgery.

The ACC/AHA guidelines recommend the following: (1) Beta blockers *should be continued* in high-risk patients who previously received these drugs and undergo vascular surgery, and they *should be administered* to high-risk patients identified by myocardial ischemia on preoperative assessment who are scheduled to undergo vascular surgery. (2) Beta blockers are *probably recommended* for high-risk patients defined by multiple clinical predictors who undergo intermediate- or high-risk procedures. They *may be considered* for intermediate-risk patients who undergo intermediate- or high-risk procedures and for low-risk patients who undergo vascular surgery.

HMG-CoA Reductase Inhibitors (Statins)

A number of prospective and retrospective studies support the perioperative prophylactic use of statins for reduction of cardiac complications in patients with established atherosclerosis. The Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography (DECREASE)-IV study, in addition to confirming the benefit of perioperative bisoprolol in lowering the risk of cardiac death or myocardial infarction at 30 days, demonstrated a reduced trend in cardiac events in intermediate-risk patients undergoing noncardiac surgery who receive fluvastatin therapy. The use of perioperative statin therapy to reduce perioperative cardiac risk should, therefore, be considered in intermediate- or high-risk patients with atherosclerotic cardiovascular disease who are undergoing major noncardiac surgery.

Angiotensin-Converting Enzyme (ACE) Inhibitors

Evidence supports the discontinuation of ACE inhibitors and angiotensin receptor blockers for 24 hours prior to noncardiac surgery due to adverse circulatory effects after induction of anesthesia.

Oral Antiplatelet Agents

Evidence-based recommendations regarding perioperative use of aspirin and/or thienopyridine to reduce cardiac risk currently lack clarity. A substantial increase in perioperative bleeding and transfusion requirement in patients receiving dual antiplatelet therapy has been observed. The discontinuation of thienopyridine and aspirin for 5–7 days prior to major surgery to minimize the risk of perioperative bleeding and transfusion must be balanced with the potential increased risk of an acute coronary syndrome and of subacute stent thrombosis in patients with recent coronary stent implantation. If clinicians elect to withhold antiplatelet agents prior to surgery, they should be restarted as soon as possible postoperatively.

Calcium Channel Blockers

Evidence is lacking to support the use of calcium channel blockers as a prophylactic strategy to decrease perioperative risk in major noncardiac surgery.

Preoperative Pulmonary Assessment

Perioperative pulmonary complications occur frequently and lead to significant morbidity and mortality. The guidelines from the American College of Physicians recommend the following:

1. All patients undergoing noncardiac surgery should be assessed for risk of pulmonary complications (Table 8-5).
2. Patients undergoing emergency or prolonged (>3 h) surgery; aortic aneurysm repair; vascular surgery; major abdominal, thoracic, neuro, head, or neck surgery; and general anesthesia should be considered to be at higher risk for postoperative pulmonary complications.
3. Patients at higher risk of pulmonary complications should receive deep breathing exercises and/or incentive spirometry as well as selective use of a nasogastric tube for postoperative nausea, vomiting, or symptomatic abdominal distention to reduce postoperative risk (Table 8-6).
4. Routine preoperative spirometry and chest radiography are less helpful for predicting risk of

postoperative pulmonary complications, but may be appropriate for patients with chronic obstructive pulmonary disease (COPD) or asthma.

5. Pulmonary artery catheterization, total parenteral nutrition, and total enteral nutrition are not encouraged for postoperative pulmonary risk reduction.

Table 8-5 Predisposing Risk Factors for Pulmonary Complications

1. Upper respiratory tract infection: cough, dyspnea
2. Age >60 years
3. COPD
4. American Society of Anesthesiologists Class ≥2
5. Functionally dependent
6. Congestive heart failure
7. Serum albumin <3.5 g/dL
8. FEV_1<2 L
9. MVV <50% of predicted
10. PEF <100 L or 50% predicted value
11. PCO_2≥45 mmHg
12. PO_2≤50 mmHg

Abbreviations: COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in one second; MVV, maximum voluntary ventilation; PEF, peak expiratory flow rate; PCO_2, partial pressure of carbon dioxide; PO_2, partial pressure of oxygen.

Source: Modified from GW Smetana et al and from DN Mohr et al: Postgrad Med 100:247, 1996, with permission.

Table 8-6 Risk Modification to Reduce Perioperative Pulmonary Complications

Preoperatively ▪ Cessation of smoking ▪ Training in proper breathing (incentive spirometry) ▪ Inhalation bronchodilator therapy ▪ Control of infection and secretion, when indicated ▪ Weight reduction, when appropriate
Intraoperatively ▪ Limited duration of anesthesia ▪ Select shorter acting neuromuscular blocking drugs when indicated ▪ Prevention of aspiration ▪ Maintenance of optimal bronchodilation Postoperatively ▪ Continuation of preoperative measures, with particular attention to inspiratory capacity maneuvers mobilization of secretions early ambulation encouragement of coughing selective use of a nasogastric tube adequate pain control without excessive narcotics

Source: From VA Lawrence et al and from WF Dunn, PD Scanlon, Mayo Clin Proc 68:371, 1993, with permission.

OTHER PREOPERATIVE PULMONARY RISK-MODIFICATION STRATEGIES

Risk-modification strategies to reduce postoperative pulmonary complications should be implemented, particularly in higher-risk patients. Patients with cough or dyspnea preoperatively should be evaluated to

determine the underlying cause of these symptoms. Patients who smoke should be counseled to quit for at least eight weeks prior to elective surgery. Patients with asthma or COPD can be given steroids and bronchodilators pre- and postoperatively to optimize pulmonary function. Bacterial pulmonary infection should be treated preoperatively.

Diabetes Mellitus

(See also Chap. 344) Many patients with diabetes mellitus have significant symptomatic or asymptomatic CAD and may have silent myocardial ischemia due to autonomic dysfunction. Evidence supports intensive perioperative glycemic control to achieve near-normal glucose levels (90–110 mg/dL) versus moderate glycemic control (120–200 mg/dL), using insulin infusion. This practice must be balanced against the risk of hypoglycemic complications. Oral hypoglycemic agonists should be held on the morning of operation. Perioperative hyperglycemia should be treated with intravenous infusion of short-acting insulin or subcutaneous sliding-scale insulin. Patients who are diet-controlled may proceed to surgery with close postoperative monitoring.

Prophylaxis for Infective Endocarditis

(See also Chap. 124) Perioperative prophylactic antibiotics should be administered to patients with congenital or valvular heart disease, prosthetic valves, mitral valve prolapse, or other cardiac abnormalities in accordance with ACC/AHA practice guidelines.

Prophylaxis of Venous Thromboembolism

(See also Chap. 262) Perioperative prophylaxis of venous thromboembolism should follow established guidelines of the American College of Chest Physicians. Aspirin is not supported as a single agent for thromboprophylaxis. Low-dose unfractionated heparin (≤5000 units subcutaneous bid), low-molecular weight heparin (e.g., enoxaparin 30 mg bid or 40 mg qd) or a pentasaccharide (fondaparinux 2.5 mg qd) for patients at moderate risk, and unfractionated heparin (5000 units subcutaneous tid) for patients at high risk. Graduated compression stockings and pneumatic compression devices are useful supplements to anticoagulant therapy.

FURTHER READINGS

Devereaux PJ et al: Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomized controlled trial. Lancet 371:1839, 2008[PMID: 18479744]

Dunkelgruen M et al: Bisoprolol and fluvastatin for the reduction of perioperative cardiac mortality and myocardial infarction in intermediate-risk patients undergoing noncardiovascular surgery: a randomized controlled trial (DECREASE-IV). Ann Surg 249:921, 2009

Fleisher LA et al: ACC/AHA guideline for perioperative cardiovascular evaluation for noncardiac surgery: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery). Circulation 116:1971, 2007[PMID: 17901356]

Geerts WH et al: Prevention of venous thromboembolism: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 126:338S, 2004

Lipshutz AK, Gropper MA: Perioperative glycemic control: An evidence-based review. Anesthesiology 110:408, 2009[PMID: 19194167]

Lee TH et al: Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 100:1043, 1999[PMID: 10477528]

Lindenauer PK et al: Perioperative beta-blocker therapy and mortality after major noncardiac surgery. N Engl J Med 353:349, 2005[PMID: 16049209]

Mcfalls EO et al: Coronary-artery revascularization before elective major vascular surgery. N Engl J Med 351:2795, 2004[PMID: 15625331]

Qaseem A et al: Risk assessment for and strategies to reduce perioperative pulmonary complications for patients undergoing noncardiothoracic surgery: A guideline from the American College of Physicians. Ann Intern Med 144:575, 2006[PMID: 16618955]

Smetana GW et al: Preoperative pulmonary risk stratification for noncardiothoracic surgery: Systematic review for the American College of Physicians. Ann Intern Med 144:581, 2006[PMID: 16618956]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EPIDEMIOLOGY

In 2007, 2,423,712 individuals died in the United States (Table 9-1). Approximately 72% of all deaths occur in those >65 years of age. The epidemiology of mortality is similar in most developed countries; cardiovascular diseases and cancer are the predominant causes of death, a marked change since 1900, when heart disease caused ~8% of all deaths and cancer accounted for <4% of all deaths. In 2006, the year with the most recent available data, AIDS accounted for <1% of all U.S. deaths, although among those age 35–44, it remained one of the top five causes.

Table 9-1 Ten Leading Causes of Death in the United States and Britain

	United States			Britain	
Cause of Death	Number of Deaths	Percent of Total	Number of Deaths Among People ≥65 Years of Age	Number of Deaths	Percent of Total
All deaths	2,423,712	100	1,759,423	538,254	100
Heart disease	616,067	25.4	510,542	129,009	24
Malignant neoplasms	562,875	23.2	387,515	135,955	25.3
Cerebrovascular diseases	135,952	5.6	117,010	57,808	10.7
Chronic lower respiratory diseases	127,924	5.1	106,845	27,905	5.2
Accidents	123,706	5.1	36,689	10,979	2
Alzheimer's disease	74,632	2.2	71,660	6316	1.2
Diabetes mellitus	71,382	2.9	52,351	34,477	6.4
Influenza and pneumonia	52,717	2.2	49,346	5055	0.9
Nephritis, nephritic syndrome, nephrosis	46,448	1.9	37,377	3287	0.6
Septicemia	34,828	1.4	26,201	2206	0.4

Source: National Center for Health Statistics (data for all age groups from 2007; for age >65, from 2006), http://www.cdc.gov/nchs; National Statistics (Great Britain, 2003), http://www.statistics.gov.uk.

It is estimated that in developed countries ~70% of all deaths are preceded by a disease or condition, making it reasonable to plan for dying in the foreseeable future. Cancer has served as the paradigm for terminal care, but it is not the only type of illness with a recognizable and predictable terminal phase. Since heart failure, chronic obstructive pulmonary disease (COPD), chronic liver failure, dementia, and many other conditions have recognizable terminal phases, a systematic approach to end-of-life care should be part of all medical specialties. Many patients with illness-related suffering also can benefit from palliative care regardless of prognosis. Ideally, palliative care should be considered part of comprehensive care for all patients. Reviews of the recent literature have found strong evidence that palliative care can be improved by coordination between caregivers, doctors, and patients for advance care planning, as well as dedicated teams of physicians, nurses, and other providers.

The rapid increases in life expectancy in the United States over the last century have been accompanied by new difficulties facing individuals, families, and society as a whole in addressing the needs of an aging population. These challenges include both more complicated conditions and technologies to address them at the end of life. The development of technologies that can prolong life without restoring full health has led many Americans to seek out alternative end-of-life care settings and approaches that relieve suffering for those with terminal diseases. Over the last few decades in the United States, a significant change in the site of death has occurred that coincides with patient and family preferences. Nearly 60% of Americans died as inpatients in hospitals in 1980. By 2000, the trend was reversing, with ~40% of Americans dying as hospital inpatients (Fig. 9-1). This shift has been most dramatic for those dying from cancer and COPD and for younger and very old individuals. In the last decade, it has been associated with the increased use of hospice care; in 2008,

approximately 39% of all decedents in the United States received such care. Cancer patients currently constitute ~38.3% of hospice users. About 79% of patients receiving hospice care die out of the hospital, and around 41% of those receiving hospice care die in a private residence. In addition, in 2008, for the first time, the American Board of Medical Specialties (ABMS) offered certification in hospice and palliative medicine. With shortening of hospital stays, many serious conditions are being treated at home or on an outpatient basis. Consequently, providing optimal palliative and end-of-life care requires ensuring that appropriate services are available in a variety of settings, including noninstitutional settings.

Figure 9-1

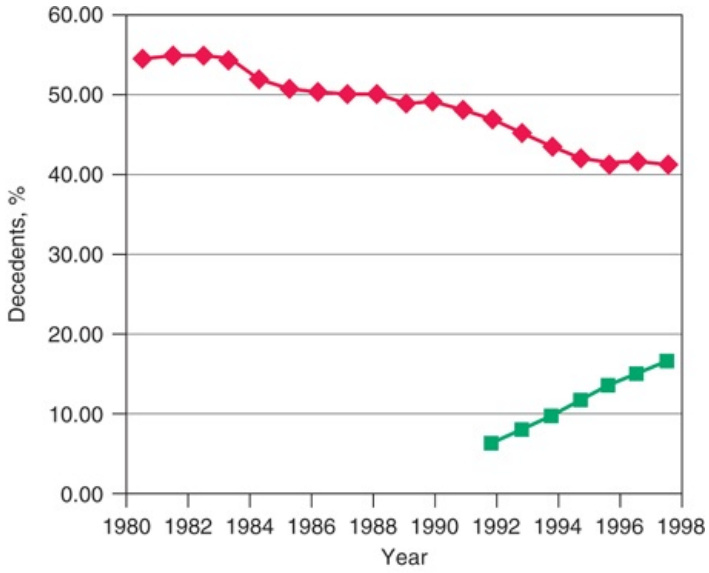

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Graph showing trends in the site of death in the last two decades. ✚, percentage of hospital inpatient deaths; ■, percentage of decedents enrolled in a hospice.

HOSPICE AND THE PALLIATIVE CARE FRAMEWORK

Central to this type of care is an interdisciplinary team approach that typically encompasses pain and symptom management, spiritual and psychological care for the patient, and support for family caregivers during the patients illness and the bereavement period.

Terminally ill patients have a wide variety of advanced diseases, often with multiple symptoms that demand relief, and require noninvasive therapeutic regimens to be delivered in flexible care settings. Fundamental to ensuring quality palliative and end-of-life care is a focus on four broad domains: (1) physical symptoms; (2) psychological symptoms; (3) social needs that include interpersonal relationships, caregiving, and economic concerns; and (4) existential or spiritual needs.

A comprehensive assessment screens for and evaluates needs in each of these four domains. Goals for care are established in discussions with the patient and/or family, based on the assessment in each of the domains. Interventions then are aimed at improving or managing symptoms and needs. Although physicians are responsible for certain interventions, especially technical ones, and for coordinating the interventions, they cannot be responsible for providing all of them. Since failing to address any one of the domains is likely to preclude a good death, a well-coordinated, effectively communicating interdisciplinary team takes on special importance in end-of-life care. Depending on the setting, critical members of the interdisciplinary team will include physicians, nurses, social workers, chaplains, nurses aides, physical therapists, bereavement counselors, and volunteers.

Assessment and Care Planning

COMPREHENSIVE ASSESSMENT

Standardized methods for conducting a comprehensive assessment focus on evaluating the patients condition in all four domains affected by illness: physical, psychological, social, and spiritual. The assessment of physical and mental symptoms should follow a modified version of the traditional medical history and physical examination that emphasizes symptoms. Questions should aim at elucidating symptoms and discerning sources of suffering and gauging how much those symptoms interfere with the patients quality of life. Standardized assessment is critical. Currently, there are 21 symptom assessment instruments for cancer alone. Further research on and validation of these assessment tools, especially taking into account patient perspectives, could improve their

effectiveness. Instruments with good psychometric properties that assess a wide range of symptoms include the Memorial Symptom Assessment Scale (MSAS), the Rotterdam Symptom Checklist, the Worthing Chemotherapy Questionnaire, and the Computerized Symptom Assessment Instrument. These instruments are long and may be useful for initial clinical or for research assessments. Shorter instruments are useful for patients whose performance status does not permit comprehensive assessments. Suitable shorter instruments include the Condensed Memorial Symptom Assessment Scale, the Edmonton Symptom Assessment System, the M.D. Anderson Symptom Assessment Inventory, and the Symptom Distress Scale. Using such instruments ensures that the assessment is comprehensive and does not focus only on pain and a few other physical symptoms. Invasive tests are best avoided in end-of-life care, and even minimally invasive tests should be evaluated carefully for their benefit-to-burden ratio for the patient. Aspects of the physical examination that are uncomfortable and unlikely to yield useful information can be omitted.

Regarding social needs, health care providers should assess the status of important relationships, financial burdens, caregiving needs, and access to medical care. Relevant questions will include the following: *How often is there someone to feel close to? How has this illness been for your family? How has it affected your relationships? How much help do you need with things like getting meals and getting around? How much trouble do you have getting the medical care you need?* In the area of existential needs, providers should assess distress and the patients sense of being emotionally and existentially settled and of finding purpose or meaning. Helpful assessment questions can include the following: *How much are you able to find meaning since your illness began? What things are most important to you at this stage?* In addition, it can be helpful to ask how the patient perceives his or her care: *How much do you feel your doctors and nurses respect you? How clear is the information from us about what to expect regarding your illness? How much do you feel that the medical care you are getting fits with your goals?* If concern is detected in any of these areas, deeper evaluative questions are warranted.

COMMUNICATION

Especially when an illness is life-threatening, there are many emotionally charged and potentially conflict-creating moments, collectively called "bad news" situations, in which empathic and effective communication skills are essential. Those moments include communicating with the patient and/or family about a terminal diagnosis, the patients prognosis, any treatment failures, deemphasizing efforts to cure and prolong life while focusing more on symptom management and palliation, advance care planning, and the patients death. Although these conversations can be difficult and lead to tension, research indicates that end-of-life discussions can lead to earlier hospice referrals rather than overly aggressive treatment, benefiting quality of life for patients and improving the bereavement process for families.

Just as surgeons plan and prepare for major operations and investigators rehearse a presentation of research results, physicians and health care providers caring for patients with significant or advanced illness can develop a practiced approach to sharing important information and planning interventions. In addition, families identify as important both how well the physician was prepared to deliver bad news and the setting in which it was delivered. For instance, 27% of families making critical decisions for patients in an intensive care unit (ICU) desired better and more private physical space to communicate with physicians, and 48% found having clergy present reassuring.

An organized and effective seven-step procedure for communicating bad news goes by the acronym P-SPIKES: (1) ***p***repare for the discussion, (2) ***s***et up a suitable environment, (3) begin the discussion by finding out what the ***p***atient and/or family understand, (4) determine how they will comprehend new ***i***nformation best and how much they want to know, (5) provide needed new ***k***nowledge accordingly, (6) allow for ***e***motional responses, and (7) ***s***hare plans for the next steps in care. Table 9-2 provides a summary of these steps along with suggested phrases and underlying rationales for each one. Additional research that further considers the response of patients to systematic methods of delivering bad news could build the evidence base for even more effective communication procedures.

Table 9-2 Elements of Communicating Bad News—the P-SPIKES Approach

Acronym	Steps	Aim of the Interaction	Preparations, Questions, or Phrases
P	Preparation	Mentally prepare for the interaction with the patient and/or family.	Review what information needs to be communicated. Plan how you will provide emotional support. Rehearse key steps and phrases in the interaction.
S	Setting of the	Ensure the appropriate setting for a serious and potentially emotionally	Ensure that patient, family, and appropriate social supports are

	interaction	charged discussion.	present. Devote sufficient time. Ensure privacy and prevent interruptions by people or beeper. Bring a box of tissues.
P	Patient's perception and preparation	Begin the discussion by establishing the baseline and whether the patient and family can grasp the information. Ease tension by having the patient and family contribute.	Start with open-ended questions to encourage participation. Possible phrases to use: *What do you understand about your illness?* *When you first had symptom X, what did you think it might be?* *What did Dr. X tell you when he or she sent you here?* *What do you think is going to happen?*
I	Invitation and information needs	Discover what information needs the patient and/or family have and what limits they want regarding the bad information.	Possible phrases to use: *If this condition turns out to be something serious, do you want to know?* *Would you like me to tell you all the details of your condition? If not, who would you like me to talk to?*
K	Knowledge of the condition	Provide the bad news or other information to the patient and/or family sensitively.	Do not just dump the information on the patient and family. Check for patient and family understanding. Possible phrases to use: *I feel badly to have to tell you this, but...* *Unfortunately, the tests showed...* *I'm afraid the news is not good...*
E	Empathy and exploration	Identify the cause of the emotions—e.g., poor prognosis. Empathize with the patient and/or family's feelings. Explore by asking open-ended questions.	Strong feelings in reaction to bad news are normal. Acknowledge what the patient and family are feeling. Remind them such feelings are normal, even if frightening. Give them time to respond. Remind patient and family you won't abandon them. Possible phrases to use: *I imagine this is very hard for you to hear.* *You look very upset.* *Tell me how you are feeling.* *I wish the news were different.* *We'll do whatever we can to help you.*
S	Summary and planning	Delineate for the patient and the family the next steps, including additional tests or interventions.	It is the unknown and uncertain that can increase anxiety. Recommend a schedule with goals and landmarks. Provide your rationale for the patient and/or family to accept (or reject).

		If the patient and/or family are not ready to discuss the next steps, schedule a follow-up visit.

Source: Adapted from Buckman.

CONTINUOUS GOAL ASSESSMENT

Major barriers to ensuring quality palliative and end-of-life care include difficulty providing an accurate prognosis and emotional resistance of patients and their families to accepting the implications of a poor prognosis. A practical solution to these barriers is to integrate palliative care with curative care regardless of prognosis. With this approach, palliative care no longer conveys the message of failure, having no more treatments, or "giving up hope." Fundamental to integrating palliative care with curative therapy is to include continuous goal assessment as part of the routine patient reassessment that occurs at most patient-physician encounters.

Goals for care are numerous, ranging from cure of a specific disease, to prolonging life, to relief of a symptom, to delaying the course of an incurable disease, to adapting to progressive disability without disrupting the family, to finding peace of mind or personal meaning, to dying in a manner that leaves loved ones with positive memories. Discernment of goals for care can be approached through a seven-step protocol: (1) ensure that medical and other information is as complete as reasonably possible and is understood by all relevant parties (see above); (2) explore what the patient and/or family are hoping for while identifying relevant and realistic goals; (3) share all the options with the patient and family; (4) respond with empathy as they adjust to changing expectations; (5) make a plan, emphasizing what can be done toward achieving the realistic goals; (6) follow through with the plan; and (7) review and revise the plan periodically, considering at every encounter whether the goals of care should be reviewed with the patient and/or family. Each of these steps need not be followed in rote order, but together they provide a helpful framework for interactions with patients and their families about goals for care. It can be especially challenging if a patient or family member has difficulty letting go of an unrealistic goal. One strategy is to help them refocus on more realistic goals and also suggest that while hoping for the best, it is still prudent to plan for other outcomes as well.

ADVANCE CARE PLANNING

Practices

Advance care planning is a process of planning for future medical care in case the patient becomes incapable of making medical decisions. A 2010 study of adults 60 or older who died between 2000 and 2006 found that 42% required decision-making about treatment in the final days of life but 70% lacked decision-making capacity. Among those lacking decision-making capacity, around one-third did not have advance planning directives. Ideally, such planning would occur before a health care crisis or the terminal phase of an illness. Unfortunately, diverse barriers prevent this. Although 80% of Americans endorse advance care planning and completing living wills, only 47% have actually done so. Most patients expect physicians to initiate advance care planning and will wait for physicians to broach the subject. Patients also wish to discuss advance care planning with their families. Yet patients with unrealistic expectations are significantly more likely to prefer aggressive treatments. Fewer than one-third of health care providers have completed advance care planning for themselves. Hence, a good first step is for health care providers to complete their own advance care planning. This makes providers aware of the critical choices in the process and the issues that are especially charged and allows them to tell their patients truthfully that they personally have done advance planning.

Steps in effective advance care planning center on (1) introducing the topic, (2) structuring a discussion, (3) reviewing plans that have been discussed by the patient and family, (4) documenting the plans, (5) updating them periodically, and (6) implementing the advance care directives (Table 9-3). Two of the main barriers to advance care planning are problems in raising the topic and difficulty in structuring a succinct discussion. Raising the topic can be done efficiently as a routine matter, noting that it is recommended for all patients, analogous to purchasing insurance or estate planning. Many of the most difficult cases have involved unexpected, acute episodes of brain damage in young individuals.

Table 9-3 Steps in Advance Care Planning

Step	Goals to be Achieved and Measures to Cover	Useful Phrases or Points to Make
Introducing advance care planning	Ask the patient what he or she knows about advance care planning and if he or she has already completed an advance care directive.	*I'd like to talk with you about something I try to discuss with all my patients. Its called advance care planning. In fact, I feel that this is such an important topic that I have done this myself. Are you familiar with advance care planning or living wills?* *Have you thought about the type of care you would want if you ever became too sick to speak for yourself? That is the*

	Indicate that you as a physician have completed advance care planning. Indicate that you try to perform advance care planning with all patients regardless of prognosis. Explain the goals of the process as empowering the patient and ensuring that you and the proxy understand the patient's preferences. Provide the patient relevant literature, including the advance care directive that you prefer to use. Recommend the patient identify a proxy decision-maker who should attend the next meeting.	*purpose of advance care planning.* *There is no change in health that we have not discussed. I am bringing this up now because it is sensible for everyone, no matter how well or ill, old or young.* Have many copies of advance care directives available, including in the waiting room, for patients and families. Know resources for state-specific forms (available at *www.nhpco.org*).
Structured discussion of scenarios and patient	Affirm that the goal of the process is to follow the patient's wishes if the patient loses decision-making capacity. Elicit the patient's overall goals related to health care. Elicit the patient's preferences for specific interventions in a few salient and common scenarios. Help the patient define the threshold for withdrawing and withholding interventions. Define the patient's preference for the role of the proxy.	Use a structured worksheet with typical scenarios. Begin the discussion with persistent vegetative state and consider other scenarios, such as recovery from an acute event with serious disability, asking the patient about his or her preferences regarding specific interventions, such as ventilators, artificial nutrition, and CPR, and then proceeding to less invasive interventions, such as blood transfusions and antibiotics.
Review the patient's preferences	After the patient has made choices of interventions, review them to ensure they are consistent and the proxy is aware of them.	
Document the patient's preferences	Formally complete the advance care directive and have a witness sign it. Provide a copy for the patient and the proxy. Insert a copy into the patient's medical record and summarize in a progress note.	
Update the directive	Periodically, and with major changes in	

	health status, review the directive with the patient and make any modifications.	
Apply the directive	The directive goes into effect only when the patient becomes unable to make medical decisions for himself or herself. Reread the directive to be sure about its content. Discuss your proposed actions based on the directive with the proxy.	

Abbreviation: CPR, cardiopulmonary resuscitation.

Structuring a focused discussion is a central communication skill. Identify the health care proxy and recommend his or her involvement in the process of advance care planning. Select a worksheet, preferably one that has been evaluated and demonstrated to produce reliable and valid expressions of patient preferences, and orient the patient and proxy to it. Such worksheets exist for both general and disease-specific situations. Discuss with the patient and proxy one scenario as an example to demonstrate how to think about the issues. It is often helpful to begin with a scenario in which the patient is likely to have settled preferences for care, such as being in a persistent vegetative state. Once the patients preferences for interventions in this scenario are determined, suggest that the patient and proxy discuss and complete the worksheet for the others. If appropriate, suggest that they involve other family members in the discussion. On a return visit, go over the patients preferences, checking and resolving any inconsistencies. After having the patient and proxy sign the document, place it in the medical chart and be sure that copies are provided to relevant family members and care sites. Since patients preferences can change, these documents have to be reviewed periodically.

Types of Documents

Advance care planning documents are of two broad types. The first includes living wills or instructional directives; these are advisory documents that describe the types of decisions that should direct care. Some are more specific, delineating different scenarios and interventions for the patient to choose from. Among these, some are for general use and others are designed for use by patients with a specific type of disease, such as cancer or HIV. Less specific directives can be general statements of not wanting life-sustaining interventions or forms that describe the values that should guide specific discussions about terminal care. The second type of advance directive allows the designation of a health care proxy (sometimes also referred to as a durable attorney for health care), who is an individual selected by the patient to make decisions. The choice is not either/or; a combined directive that includes a living will and designates a proxy is often used, and the directive should indicate clearly whether the specified patient preferences or the proxys choice takes precedence if they conflict. Some states have begun to put into practice a "Physician Orders for Life-Sustaining Treatment (POLST)" paradigm, which builds on communication between providers and patients to include guidance for end-of-life care in a color-coordinated form that follows the patient across treatment settings. The procedures for completing advance care planning documents vary according to state law.

A potentially misleading distinction relates to statutory as opposed to advisory documents. Statutory documents are drafted to fulfill relevant state laws. Advisory documents are drafted to reflect the patients wishes. Both are legal, the first under state law and the latter under common or constitutional law.

Legal Aspects

As of 2006, 48 states and the District of Columbia had enacted living will legislation. Many states have their own statutory forms. Massachusetts and Michigan do not have living will laws, although both have health care proxy laws. In 25 states, the laws state that the living will is not valid if a woman is pregnant. However, like all other states except Alaska, these states have enacted durable power of attorney for health care laws that permit patients to designate a proxy decision-maker with authority to terminate life-sustaining treatments. Only in Alaska does the law prohibit proxies from terminating life-sustaining treatments. The health reform legislation, the Affordable Care Act of 2010, raised substantial controversy when early versions of the law included Medicare reimbursement for advance care planning consultations. These provisions were withdrawn over accusations that they would lead to the rationing of care for the elderly.

The U.S. Supreme Court has ruled that patients have a constitutional right to decide about refusing

and terminating medical interventions, including life-sustaining interventions, and that mentally incompetent patients can exercise this right by providing "clear and convincing evidence" of their preferences. Since advance care directives permit patients to provide such evidence, commentators agree that they are constitutionally protected. Most commentators believe that a state is required to honor any clear advance care directive whether or not it is written on an "official" form. Many states have enacted laws explicitly to honor out-of-state directives. If a patient is not using a statutory form, it may be advisable to attach a statutory form to the advance care directive being used. State-specific forms are readily available free of charge for health care providers and patients and families through the website of the National Hospice and Palliative Care Organization (http://www.nhpco.org).

INTERVENTIONS

Physical Symptoms and Their Management

Great emphasis has been placed on addressing dying patients pain. Some institutions have made pain assessment a fifth vital sign to emphasize its importance. This also has been advocated by large health care systems such as the Veterans Administration and accrediting bodies such as the Joint Commission on the Accreditation of Health Care Organizations (JCAHO). Although this embrace of pain as the fifth vital sign has been symbolically important, no data document that it has improved pain management practices. Although good palliative care requires good pain management, it also requires more. The frequency of symptoms varies by disease and other factors. The most common physical and psychological symptoms among all terminally ill patients include pain, fatigue, insomnia, anorexia, dyspnea, depression, anxiety, and nausea and vomiting. In the last days of life, terminal delirium is also common. Assessments of patients with advanced cancer have shown that patients experienced an average of 11.5 different physical and psychological symptoms (Table 9-4).

Table 9-4 Common Physical and Psychological Symptoms of Terminally Ill Patients

Physical Symptoms	Psychological Symptoms
Pain	Anxiety
Fatigue and weakness	Depression
DyspneaInsomnia	Hopelessness
Dry mouth	Meaninglessness
Anorexia	Irritability
Nausea and vomiting	Impaired concentration
Constipation	Confusion
Cough	Delirium
Swelling of arms or legs	Loss of libido
Itching	
Diarrhea	
Dysphagia	
Dizziness	
Fecal and urinary incontinence	
Numbness/tingling in hands/feet	

Evaluations to determine the etiology of these symptoms usually can be limited to the history and physical examination. In some cases, radiologic or other diagnostic examinations will provide sufficient benefit in directing optimal palliative care to warrant the risks, potential discomfort, and inconvenience, especially to a seriously ill patient. Only a few of the common symptoms that present difficult management issues will be addressed in this chapter. **Additional information on the management of other symptoms, such as nausea and vomiting, insomnia, and diarrhea, can be found in Chaps. 39 and 81, Chap. 27, and Chap. 40, respectively.**

PAIN

Frequency

The frequency of pain among terminally ill patients varies widely. Substantial pain occurs in 36–90% of patients with advanced cancer. In the SUPPORT study of hospitalized patients with diverse conditions and an estimated survival ≤6 months, 22% reported moderate to severe pain, and caregivers of those patients noted that 50% had similar levels of pain during the last few days of life. A meta-analysis found pain prevalence of 58–69% in studies that included patients characterized as having advanced, metastatic, or terminal cancer; 44–73% in studies that included patients characterized as undergoing cancer treatment; and 21–46% in studies that included posttreatment individuals.

Etiology

Nociceptive pain is the result of direct mechanical or chemical stimulation of nociceptors and normal neural signaling to the brain. It tends to be localized, aching, throbbing, and cramping. The classic example is bone metastases. *Visceral pain* is caused by nociceptors in gastrointestinal, respiratory, and other organ systems. It is a deep or colicky type of pain classically associated with pancreatitis, myocardial infarction, or tumor invasion of viscera. *Neuropathic pain* arises from disordered nerve signals. It is described by patients as burning, electrical, or shocklike pain. Classic examples are poststroke pain, tumor invasion of the brachial plexus, and herpetic neuralgia.

Assessment

Pain is a subjective experience. Depending on the patients circumstances, perspective, and physiologic condition, the same physical lesion or disease state can produce different levels of reported pain and need for pain relief. Systematic assessment includes eliciting the following: (1) type: throbbing, cramping, burning, etc.; (2) periodicity: continuous, with or without exacerbations, or incident; (3) location; (4) intensity; (5) modifying factors; (6) effects of treatments; (7) functional impact; and (8) impact on patient. Several validated pain assessment measures may be used, such as the Visual Analogue Scale, the Brief Pain Inventory, and the pain component of one of the more comprehensive symptom assessment instruments. Frequent reassessments are essential to assess the effects of interventions.

Interventions

Interventions for pain must be tailored to each individual, with the goal of preempting chronic pain and relieving breakthrough pain. At the end of life, there is rarely reason to doubt a patients report of pain. Pain medications are the cornerstone of management. If they are failing and nonpharmacologic interventions—including radiotherapy and anesthetic or neurosurgical procedures such as peripheral nerve blocks or epidural medications—are required, a pain consultation is appropriate.

Pharmacologic interventions follow the World Health Organization three-step approach involving nonopioid analgesics, mild opioids, and strong opioids, with or without adjuvants (Chap. 11). Nonopioid analgesics, especially nonsteroidal anti-inflammatory drugs (NSAIDs), are the initial treatments for mild pain. They work primarily by inhibiting peripheral prostaglandins and reducing inflammation but also may have central nervous system (CNS) effects. They have a ceiling effect. Ibuprofen, up to 1600 mg/d qid, has a minimal risk of causing bleeding and renal impairment and is a good initial choice. In patients with a history of severe gastrointestinal (GI) or other bleeding, it should be avoided. In patients with a history of mild gastritis or gastroesophageal reflux disease (GERD), acid-lowering therapy such as a proton pump inhibitor should be used. Acetaminophen is an alternative in patients with a history of GI bleeding and can be used safely at up to 4 g/d qid. In patients with liver dysfunction due to metastases or other causes and in patients with heavy alcohol use, doses should be reduced.

If nonopioid analgesics are insufficient, opioids should be introduced. They work by interacting with mu opioid receptors in the CNS to activate pain-inhibitory neurons; most are receptor antagonists. The mixed agonist/antagonist opioids useful for post-acute pain should not be used for the chronic pain in end-of-life care. Weak opioids such as codeine can be used initially. However, if they are escalated and fail to relieve pain, strong opioids such as morphine, 5–10 mg every 4 h, should be used. Nonopioid analgesics should be combined with opioids because they potentiate the effect of opioids.

For continuous pain, opioids should be administered on a regular, around-the-clock basis consistent with their duration of analgesia. They should not be provided only when the patient experiences pain; the goal is to prevent patients from experiencing pain. Patients also should be provided rescue medication, such as liquid morphine, for breakthrough pain, generally at 20% of the baseline dose. Patients should be informed that using the rescue medication does not obviate the need to take the next standard dose of pain medication. If after 24 h the patients pain remains uncontrolled and recurs before the next dose, requiring the patient to utilize the rescue medication, the daily opioid dose can be increased by the total dose of rescue medications used by the patient, or by 50% for moderate pain and 100% for severe pain of the standing opioid daily dose.

It is inappropriate to start with extended-release preparations. Instead, an initial focus on using short-acting preparations to determine how much is required in the first 24–48 h will allow clinicians to determine opioid needs. Once pain relief is obtained with short-acting preparations, one should switch to extended-release preparations. Even with a stable extended-release preparation regimen, the patient may have incident pain, such as during movement or dressing changes. Short-acting preparations should be taken before such predictable episodes. Although less common, patients may have "end-of-dose failure" with long-acting opioids, meaning that they develop pain after 8 h in the case of an every-12-h medication. In these cases, a trial of giving an every-12-h medication every 8 h is appropriate.

Because of differences in opioid receptors, cross-tolerance among opioids is incomplete, and patients may experience different side effects with different opioids. Therefore, if a patient is not experiencing pain relief or is experiencing too many side effects, a change to another opioid preparation is appropriate. When switching, one should begin with 50–75% of the published equianalgesic dose of

the new opioid.

Unlike NSAIDs, opioids have no ceiling effect; therefore, there is no maximum dose no matter how many milligrams the patient is receiving. The appropriate dose is the dose needed to achieve pain relief. This is an important point for clinicians to explain to patients and families. Addiction or excessive respiratory depression is extremely unlikely in the terminally ill; fear of these side effects should neither prevent escalating opioid medications when the patient is experiencing insufficient pain relief nor justify using opioid antagonists.

Opioid side effects should be anticipated and treated preemptively. Nearly all patients experience constipation that can be debilitating (see below). Failure to prevent constipation often results in noncompliance with opioid therapy. Methylnaltrexone is a drug that targets opioid-induced constipation by blocking peripheral opioid receptors but not central receptors for analgesia. In placebo-controlled trials, it has been shown to cause laxation within 24 h of administration. As with the use of opioids, about a third of patients using methylnaltrexone experience nausea and vomiting, but unlike constipation, tolerance develops, usually within a week. Therefore, when one is beginning opioids, an antiemetic such as metoclopramide or a serotonin antagonist often is prescribed prophylactically and stopped after 1 week. Olanzapine also has been shown to have antinausea properties and can be effective in countering delirium or anxiety, with the advantage of some weight gain.

Drowsiness, a common side effect of opioids, also usually abates within a week. During this period, drowsiness can be treated with psychostimulants such as dextroamphetamine, methylphenidate, and modafinil. Modafinil has the advantage of everyday dosing. Pilot reports suggest that donepezil may also be helpful for opiate-induced drowsiness as well as relieving fatigue and anxiety. Metabolites of morphine and most opioids are cleared renally; doses may have to be adjusted for patients with renal failure.

Seriously ill patients who require chronic pain relief rarely if ever become addicted. Suspicion of addiction should not be a reason to withhold pain medications from terminally ill patients. Patients and families may withhold prescribed opioids for fear of addiction or dependence. Physicians and health care providers should reassure patients and families that the patient will not become addicted to opioids if they are used as prescribed for pain relief; this fear should not prevent the patient from taking the medications around the clock. However, diversion of drugs for use by other family members or illicit sale may occur. It may be necessary to advise the patient and caregiver about secure storage of opioids. Contract writing with the patient and family can help. If that fails, transfer to a safe facility may be necessary.

Tolerance is the need to increase medication dosage for the same pain relief without a change in disease. In the case of patients with advanced disease, the need for increasing opioid dosage for pain relief usually is caused by disease progression rather than tolerance. Physical dependence is indicated by symptoms from the abrupt withdrawal of opioids and should not be confused with addiction.

Adjuvant analgesic medications are nonopioids that potentiate the analgesic effects of opioids. They are especially important in the management of neuropathic pain. Gabapentin, an anticonvulsant initially studied in the setting of herpetic neuralgia, is now the first-line treatment for neuropathic pain from a variety of causes. It is begun at 100–300 mg bid or tid, with 50–100% dose increments every 3 days. Usually 900–3600 mg/d in two or three doses is effective. The combination of gabapentin and nortriptyline may be more effective than gabapentin alone. One potential side effect of gabapentin to be aware of is confusion and drowsiness, especially in the elderly. Other effective adjuvant medications include pregabalin, which has the same mechanism of action as gabapentin but is absorbed more efficiently from the GI tract. Lamotrigine is a novel agent whose mechanism of action is unknown, but it has shown effectiveness. It is recommended to begin at 25–50 mg/d, increasing to 100 mg/d. Carbamazepine, a first-generation agent, has been proved effective in randomized trials for neuropathic pain. Other potentially effective anticonvulsant adjuvants include topiramate (25–50 mg qd or bid, rising to 100–300 mg/d) and oxcarbazepine (75–300 mg bid, rising to 1200 mg bid). Glucocorticoids, preferably dexamethasone given once a day, can be useful in reducing inflammation that causes pain while elevating mood, energy, and appetite. Its main side effects include confusion, sleep difficulties, and fluid retention. Glucocorticoids are especially effective for bone pain and abdominal pain from distention of the GI tract or liver. Other drugs, including clonidine and baclofen, can be effective in pain relief. These drugs are adjuvants and generally should be used in conjunction with—not instead of—opioids. Methadone, carefully dosed because of its unpredictable half-life in many patients, has activity at the *N*-methyl-d-aspartamate (NMDA) receptor and is useful for complex pain syndromes and neuropathic pain. It generally is reserved for cases in which first-line opioids (morphine, oxycodone, hydromorphone) are either ineffective or unavailable.

Radiation therapy can treat bone pain from single metastatic lesions. Bone pain from multiple metastases can be amenable to radiopharmaceuticals such as strontium 89 and samarium 153. Bisphosphonates [such as pamidronate (90 mg every 4 weeks)] and calcitonin (200 IU intranasally once or twice a day) also provide relief from bone pain but have onset of action of days.

CONSTIPATION

Frequency

Constipation is reported in up to 87% of patients requiring palliative care.

Etiology

Although hypercalcemia and other factors can cause constipation, it is most frequently a predictable consequence of the use of opioids for the relief of pain and dyspnea and of tricyclic antidepressants, from their anticholinergic effects, as well as of the inactivity and poor diet that are common among seriously ill patients. If untreated, constipation can cause substantial pain and vomiting and also is associated with confusion and delirium. Whenever opioids and other medications known to cause constipation are used, preemptive treatment for constipation should be instituted.

Assessment

The physician should establish the patients previous bowel habits, including the frequency, consistency, and volume. Abdominal and rectal examinations should be performed to exclude impaction or acute abdomen. A number of constipation assessment scales are available, although guidelines issued in the *Journal of Palliative Medicine* did not recommend them for routine practice. Four commonly used assessment scales are the Bristol Stool Form Scale, the Constipation Assessment Scale, the Constipation Visual Analogue Scale, and the Eton Scale Risk Assessment for Constipation. Radiographic assessments beyond a simple flat plate of the abdomen in cases in which obstruction is suspected are rarely necessary.

Intervention

Intervention to reestablish comfortable bowel habits and relieve pain and discomfort should be the goals of any measures to address constipation during end-of-life care. Although physical activity, adequate hydration, and dietary treatments with fiber can be helpful, each is limited in its effectiveness for most seriously ill patients, and fiber may exacerbate problems in the setting of dehydration and if impaired motility is the etiology. Fiber is contraindicated in the presence of opioid use. Stimulant and osmotic laxatives, stool softeners, fluids, and enemas are the mainstays of therapy (Table 9-5). In preventing constipation from opioids and other medications, a combination of a laxative and a stool softener (such as senna and docusate) should be used. If after several days of treatment a bowel movement has not occurred, a rectal examination to remove impacted stool and place a suppository is necessary. For patients with impending bowel obstruction or gastric stasis, octreotide to reduce secretions can be helpful. For patients in whom the suspected mechanism is dysmotility, metoclopramide can be helpful.

Table 9-5 Medications for the Management of Constipation

Intervention	Dose	Comment
Stimulant laxatives	120–240 mL/d	These agents directly stimulate peristalsis and may reduce colonic absorption of water.
Prune juice	2–8 tablets PO bid	Work in 6–12 h.
Senna (Senokot)	5–15 mg/d PO, PR	
Bisacodyl		
Osmotic laxatives	15–30 mL PO q4–8h	These agents are not absorbed. They attract and retain water in the gastrointestinal tract.
Lactulose	15–30 mL/d PO	Lactulose may cause flatulence and bloating.
Magnesium hydroxide (Milk of Magnesia)	125–250 mL/d PO	Lactulose works in 1 day, magnesium products in 6 h.
Magnesium citrate		
Stool softeners	300–600 mg/d PO	These agents work by increasing water secretion and as detergents, increasing water penetration into the stool.
Sodium docusate (Colace)	300–600 mg/d PO	Work in 1–3 days.
Calcium docusate		
Suppositories and enemas	10–15 PR qd	Fixed dose, 4.5 oz, Fleet's.
Bisacodyl	PR qd	
Sodium phosphate enema		

NAUSEA

Frequency

Up to 70% of patients with advanced cancer have nausea, defined as the subjective sensation of wanting to vomit.

Etiology

Nausea and vomiting are both caused by stimulation at one of four sites: the GI tract, the vestibular system, the chemoreceptor trigger zone (CTZ), and the cerebral cortex. Medical treatments for nausea are aimed at receptors at each of these sites: The GI tract contains mechanoreceptors, chemoreceptors, and 5-hydroxytryptamine type 3 (5-HT3) receptors; the vestibular system probably contains histamine and acetylcholine receptors; and the CTZ contains chemoreceptors, dopamine type 2 receptors, and 5-HT3 receptors. An example of nausea that most likely is mediated by the cortex is anticipatory nausea before a dose of chemotherapy or other noxious stimuli.

Specific causes of nausea include metabolic changes (liver failure, uremia from renal failure, hypercalcemia), bowel obstruction, constipation, infection, GERD, vestibular disease, brain metastases, medications (including antibiotics, NSAIDs, proton pump inhibitors, opioids, and chemotherapy), and radiation therapy. Anxiety can also contribute to nausea.

Intervention

Medical treatment of nausea is directed at the anatomic and receptor-mediated cause that a careful history and physical examination reveals. When a single specific cause is not found, many advocate beginning treatment with a dopamine antagonist such as haloperidol or prochlorperazine. Prochlorperazine is usually more sedating than haloperidol. When decreased motility is suspected, metoclopramide can be an effective treatment. When inflammation of the GI tract is suspected, glucocorticoids such as dexamethasone are an appropriate treatment. For nausea that follows chemotherapy and radiation therapy, one of the 5-HT3 receptor antagonists (ondansetron, granisetron, dolasetron) is recommended. Clinicians should attempt prevention of postchemotherapy nausea rather than provide treatment after the fact. Current clinical guidelines recommend tailoring the strength of treatments to the specific emetic risk posed by a specific chemotherapy drug. When a vestibular cause (such as "motion sickness" or labyrinthitis) is suspected, antihistamines such as meclizine (whose primary side effect is drowsiness) or anticholinergics such as scopolamine can be effective. In anticipatory nausea, a benzodiazepine such as lorazepam is indicated. As with antihistamines, drowsiness and confusion are the main side effects.

DYSPNEA

Frequency

Dyspnea is a subjective experience of being short of breath. Nearly 75% of dying patients experience dyspnea at some point in their illness. Dyspnea is among the most distressing physical symptoms and can be even more distressing than pain.

Assessment

As with pain, dyspnea is a subjective experience that may not correlate with objective measures of P_{O2}, P_{CO2}, or respiratory rate. Consequently, measurements of oxygen saturation through pulse oximetry or blood gases are rarely helpful in guiding therapy. Despite the limitations of existing assessment methods, physicians should regularly assess and document patients experience of dyspnea and its intensity. Guidelines recommend visual or analogue dyspnea scales to assess the severity of symptoms and the effects of treatment. Potentially reversible or treatable causes of dyspnea include infection, pleural effusions, pulmonary emboli, pulmonary edema, asthma, and tumor encroachment on the airway. However, the risk-versus-benefit ratio of the diagnostic and therapeutic interventions for patients with little time left to live must be considered carefully before one undertakes diagnostic steps. Frequently, the specific etiology cannot be identified, and dyspnea is the consequence of progression of the underlying disease that cannot be treated. The anxiety caused by dyspnea and the choking sensation can significantly exacerbate the underlying dyspnea in a negatively reinforcing cycle.

Interventions

When reversible or treatable etiologies are diagnosed, they should be treated as long as the side effects of treatment, such as repeated drainage of effusions or anticoagulants, are less burdensome than the dyspnea itself. More aggressive treatments such as stenting a bronchial lesion may be warranted if it is clear that the dyspnea is due to tumor invasion at that site and if the patient and family understand the risks of such a procedure. Usually, treatment will be symptomatic (Table 9-6). A dyspnea scale and careful monitoring should guide dose adjustment. Low-dose opioids reduce the sensitivity of the central respiratory center and the sensation of dyspnea. If patients are not receiving opioids, weak opioids can be initiated; if patients are already receiving opioids, morphine or other strong opioids should be used. Controlled trials do not support the use of nebulized opioids for dyspnea at the end of life. Phenothiazines and chlorpromazine may be helpful when combined with opioids. Benzodiazepines can be helpful if anxiety is present but should be neither used as first-line therapy nor used alone in the treatment of dyspnea. If the patient has a history of COPD or asthma,

inhaled bronchodilators and glucocorticoids may be helpful. If the patient has pulmonary edema due to heart failure, diuresis with a medication such as furosemide is indicated. Excess secretions can be dried with scopolamine, transdermally or intravenously. Oxygen can be used, although it may only be an expensive placebo. For some families and patients, oxygen is distressing; for others, it is reassuring. More general interventions that medical staff can do include sitting the patient upright, removing smoke or other irritants such as perfume, ensuring a supply of fresh air with sufficient humidity, and minimizing other factors that can increase anxiety.

Table 9-6 Medications for the Management of Dyspnea

Intervention	Dose	Comments
Weak opioids		For patients with mild dyspnea
Codeine (or codeine with 325 mg acetaminophen)	30 mg PO q4h	For opioid-naïve patients
Hydrocodone	5 mg PO q4h	
Strong opioids		For opioid-naïve patients with moderate to severe dyspnea
Morphine	5–10 mg PO q4h	For patients already taking opioids for pain or other symptoms
	30–50% of baseline opioid dose q4h	
Oxycodone	5–10 mg PO q4h	
Hydromorphone	1–2 mg PO q4h	
Anxiolytics		Give a dose every hour until the patient is relaxed, then provide a dose for maintenance
Lorazepam	0.5–2.0 mg PO/SL/IV qh then q4–6h	
Clonazepam	0.25–2.0 mg PO q12h	
Midazolam	0.5 mg IV q15min	

FATIGUE

Frequency

More than 90% of terminally ill patients experience fatigue and/or weakness. Fatigue is one of the most commonly reported symptoms of cancer treatment as well as in the palliative care of multiple sclerosis, COPD, heart failure, and HIV. Fatigue frequently is cited as among the most distressing symptoms.

Etiology

The multiple causes of fatigue in the terminally ill can be categorized as resulting from the underlying disease; from disease-induced factors such as tumor necrosis factor and other cytokines; and from secondary factors such as dehydration, anemia, infection, hypothyroidism, and drug side effects. Apart from low caloric intake, loss of muscle mass and changes in muscle enzymes may play an important role in fatigue of terminal illness. The importance of changes in the CNS, especially the reticular activating system, have been hypothesized based on reports of fatigue in patients receiving cranial radiation, experiencing depression, or having chronic pain in the absence of cachexia or other physiologic changes. Finally, depression and other causes of psychological distress can contribute to fatigue.

Assessment

Fatigue is subjective; objective changes, even in body mass, may be absent. Consequently, assessment must rely on patient self-reporting. Scales used to measure fatigue, such as the Edmonton Functional Assessment Tool, the Fatigue Self-Report Scales, and the Rhoten Fatigue Scale, are usually appropriate for research rather than clinical purposes. In clinical practice, a simple performance assessment such as the Karnofsky Performance Status or the Eastern Cooperative Oncology Groups question "How much of the day does the patient spend in bed?" may be the best measure. In this 0–4 performance status assessment, 0 = normal activity; 1 = symptomatic without being bedridden; 2 = requiring some, but <50%, bed time; 3 = bedbound more than half the day; and 4 = bedbound all the time. Such a scale allows for assessment over time and correlates with overall disease severity and prognosis. A 2008 review by the European Association of Palliative Care also described several longer assessment tools with 9–20 items, including the Piper Fatigue Inventory, the Multidimensional Fatigue Inventory, and the Brief Fatigue Inventory (BFI).

Interventions

At the end of life, fatigue will not be "cured." The goal is to ameliorate it and help patients and families adjust expectations. Behavioral interventions should be utilized to avoid blaming the patient for inactivity and to educate both the family and the patient that the underlying disease causes physiologic changes that produce low energy levels. Understanding that the problem is physiologic and not psychological can help alter expectations regarding the patients level of physical activity. Practically, this may mean reducing routine activities such as housework and cooking or social events outside the house and making it acceptable to receive guests lying on a couch. At the same time, institution of exercise regimens and physical therapy can raise endorphins, reduce muscle wasting, and reduce the risk of depression. In addition, ensuring good hydration without worsening edema may help reduce fatigue. Discontinuing medications that worsen fatigue may help, including cardiac medications, benzodiazepines, certain antidepressants, or opioids if pain is well-controlled. As end-of-life care proceeds into its final stages, fatigue may protect patients from further suffering, and continued treatment could be detrimental.

Only a few pharmacologic interventions target fatigue and weakness. Glucocorticoids can increase energy and enhance mood. Dexamethasone is preferred for its once-a-day dosing and minimal mineralocorticoid activity. Benefit, if any, usually is seen within the first month. Psychostimulants such as dextroamphetamine (5–10 mg PO) and methylphenidate (2.5–5 mg PO) may also enhance energy levels, although a randomized trial did not show methylphenidate beneficial compared with placebo in cancer fatigue. Doses should be given in the morning and at noon to minimize the risk of counterproductive insomnia. Modafinil, developed for narcolepsy, has shown some promise in the treatment of fatigue and has the advantage of once-daily dosing. Its precise role in fatigue at the end of life has not been determined. Anecdotal evidence suggests that l-carnitine may improve fatigue, depression, and sleep disruption.

Psychological Symptoms and Their Management

DEPRESSION

Frequency

Depression at the end of life presents an apparently paradoxical situation. Many people believe that depression is normal among seriously ill patients because they are dying. People frequently say, "Wouldnt you be depressed?" However, depression is not a necessary part of terminal illness and can contribute to needless suffering. Although sadness, anxiety, anger, and irritability are normal responses to a serious condition, they are typically of modest intensity and transient. Persistent sadness and anxiety and the physically disabling symptoms that they can lead to are abnormal and suggestive of major depression. Although as many as 75% of terminally ill patients experience depressive symptoms, <25% of terminally ill patients have major depression.

Etiology

Previous history of depression, family history of depression or bipolar disorder, and prior suicide attempts are associated with increased risk for depression among terminally ill patients. Other symptoms, such as pain and fatigue, are associated with higher rates of depression; uncontrolled pain can exacerbate depression, and depression can cause patients to be more distressed by pain. Many medications used in the terminal stages, including glucocorticoids, and some anticancer agents, such as tamoxifen, interleukin 2, interferon α, and vincristine, also are associated with depression. Some terminal conditions, such as pancreatic cancer, certain strokes, and heart failure, have been reported to be associated with higher rates of depression, although this is controversial. Finally, depression may be attributable to grief over the loss of a role or function, social isolation, or loneliness.

Assessment

Diagnosing depression among seriously ill patients is complicated because many of the vegetative symptoms in the DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*) criteria for clinical depression—insomnia, anorexia and weight loss, fatigue, decreased libido, and difficulty concentrating —are associated with the dying process itself. The assessment of depression in seriously ill patients therefore should focus on the dysphoric mood, helplessness, hopelessness, and lack of interest and enjoyment and concentration in normal activities. The single questions "How often do you feel downhearted and blue?" (more than a good bit of the time or similar responses) and "Do you feel depressed most of the time?" are appropriate for screening.

Certain conditions may be confused with depression. Endocrinopathies such as hypothyroidism and Cushings syndrome, electrolyte abnormalities such as hypercalcemia, and akathisia, especially from dopamine-blocking antiemetics such as metoclopramide and prochlorperazine, can mimic depression and should be excluded.

Interventions

Physicians must treat any physical symptom, such as pain, that may be causing or exacerbating depression. Fostering adaptation to the many losses that the patient is experiencing can also be helpful. Nonpharmacologic interventions, including group or individual psychological counseling, and

behavioral therapies such as relaxation and imagery can be helpful, especially in combination with drug therapy.

Pharmacologic interventions remain the core of therapy. The same medications are used to treat depression in terminally ill as in non-terminally ill patients. Psychostimulants may be preferred for patients with a poor prognosis or for those with fatigue or opioid-induced somnolence. Psychostimulants are comparatively fast acting, working within a few days instead of the weeks required for selective serotonin reuptake inhibitors (SSRIs). Dextroamphetamine or methylphenidate should be started at 2.5–5.0 mg in the morning and at noon, the same starting doses used for treating fatigue. The dose can be escalated up to 15 mg bid. Modafinil is started at 100 mg qd and can be increased to 200 mg if there is no effect at the lower dose. Pemoline is a nonamphetamine psychostimulant with minimal abuse potential. It is also effective as an antidepressant beginning at 18.75 mg in the morning and at noon. Because it can be absorbed through the buccal mucosa, it is preferred for patients with intestinal obstruction or dysphagia. If it is used for prolonged periods, liver function must be monitored. The psychostimulants can also be combined with more traditional antidepressants while waiting for the antidepressants to become effective and then tapered after a few weeks if necessary. Psychostimulants have side effects, particularly initial anxiety, insomnia, and rarely paranoia, which may necessitate lowering the dose or discontinuing treatment.

Mirtazapine, an antagonist at the postsynaptic serotonin receptors, is a promising psychostimulant. It should be started at 7.5 mg before bed. It has sedating, antiemetic, and anxiolytic properties with few drug interactions. Its side effect of weight gain may be beneficial for seriously ill patients; it is available in orally disintegrating tablets.

For patients with a prognosis of several months or longer, SSRIs, including fluoxetine, sertraline, paroxetine and citalopram, and serotonin-noradrenaline reuptake inhibitors such as venlafaxine, are the preferred treatment because of their efficacy and comparatively few side effects. Because low doses of these medications may be effective for seriously ill patients, one should use half the usual starting dose for healthy adults. The starting dose for fluoxetine is 10 mg once a day. In most cases, once-a-day dosing is possible. The choice of which SSRI to use should be driven by (1) the patients past success or failure with the specific medication and (2) the most favorable side-effect profile for that specific agent. For instance, for a patient in whom fatigue is a major symptom, a more activating SSRI (fluoxetine) would be appropriate. For a patient in whom anxiety and sleeplessness are major symptoms, a more sedating SSRI (paroxetine) would be appropriate.

Atypical antidepressants are recommended only in selected circumstances, usually with the assistance of a specialty consultation. Trazodone can be an effective antidepressant but is sedating and can cause orthostatic hypotension and, rarely, priapism. Therefore, it should be used only when a sedating effect is desired and is often used for patients with insomnia, at a dose starting at 25 mg. In addition to its antidepressant effects, bupropion is energizing, making it useful for depressed patients who experience fatigue. However, it can cause seizures, preventing its use for patients with a risk of CNS neoplasms or terminal delirium. Finally, alprazolam, a benzodiazepine, starting at 0.25–1.0 mg tid, can be effective in seriously ill patients who have a combination of anxiety and depression. Although it is potent and works quickly, it has many drug interactions and may cause delirium, especially among very ill patients, because of its strong binding to the benzodiazepine–γ-aminobutyric acid (GABA) receptor complex.

Unless used as adjuvants for the treatment of pain, tricyclic antidepressants are not recommended. Similarly, monoamine oxidase (MAO) inhibitors are not recommended because of their side effects and dangerous drug interactions.

DELIRIUM (SEE CHAP. 25)

Frequency

In the weeks or months before death, delirium is uncommon, although it may be significantly underdiagnosed. However, delirium becomes relatively common in the hours and days immediately before death. Up to 85% of patients dying from cancer may experience terminal delirium.

Etiology

Delirium is a global cerebral dysfunction characterized by alterations in cognition and consciousness. It frequently is preceded by anxiety, changes in sleep patterns (especially reversal of day and night), and decreased attention. In contrast to dementia, delirium has an acute onset, is characterized by fluctuating consciousness and inattention, and is reversible, although reversibility may be more theoretical than real for patients near death. Delirium may occur in a patient with dementia; indeed, patients with dementia are more vulnerable to delirium.

Causes of delirium include metabolic encephalopathy arising from liver or renal failure, hypoxemia, or infection; electrolyte imbalances such as hypercalcemia; paraneoplastic syndromes; dehydration; and primary brain tumors, brain metastases, or leptomeningeal spread of tumor. Commonly, among dying patients, delirium can be caused by side effects of treatments, including radiation for brain metastases, and medications, including opioids, glucocorticoids, anticholinergic drugs, antihistamines, antiemetics, benzodiazepines, and chemotherapeutic agents. The etiology may be multifactorial; e.g., dehydration

may exacerbate opioid-induced delirium.

Assessment

Delirium should be recognized in any terminally ill patient with new onset of disorientation, impaired cognition, somnolence, fluctuating levels of consciousness, or delusions with or without agitation. Delirium must be distinguished from acute anxiety and depression as well as dementia. The central distinguishing feature is altered consciousness, which usually is not noted in anxiety, depression, and dementia. Although "hyperactive" delirium characterized by overt confusion and agitation is probably more common, patients also should be assessed for "hypoactive" delirium characterized by sleep-wake reversal and decreased alertness.

In some cases, use of formal assessment tools such as the Mini-Mental Status Examination (which does not distinguish delirium from dementia) and the Delirium Rating Scale (which does distinguish delirium from dementia) may be helpful in distinguishing delirium from other processes. The patients list of medications must be evaluated carefully. Nonetheless, a reversible etiologic factor for delirium is found in fewer than half of terminally ill patients. Because most terminally ill patients experiencing delirium will be very close to death and may be at home, extensive diagnostic evaluations such as lumbar punctures and neuroradiologic examinations are usually inappropriate.

Interventions

One of the most important objectives of terminal care is to provide terminally ill patients the lucidity to say goodbye to the people they love. Delirium, especially with agitation during the final days, is distressing to family and caregivers. A strong determinant of bereavement difficulties is witnessing a difficult death. Thus, terminal delirium should be treated aggressively.

At the first sign of delirium, such as day-night reversal with slight changes in mentation, the physician should let the family members know that it is time to be sure that everything they want to say has been said. The family should be informed that delirium is common just before death.

If medications are suspected of being a cause of the delirium, unnecessary agents should be discontinued. Other potentially reversible causes, such as constipation, urinary retention, and metabolic abnormalities, should be treated. Supportive measures aimed at providing a familiar environment should be instituted, including restricting visits only to individuals with whom the patient is familiar and eliminating new experiences; orienting the patient, if possible, by providing a clock and calendar; and gently correcting the patients hallucinations or cognitive mistakes.

Pharmacologic management focuses on the use of neuroleptics and, in the extreme, anesthetics (Table 9-7). Haloperidol remains first-line therapy. Usually, patients can be controlled with a low dose (1–3 mg/d), usually given every 6 h, although some may require as much as 20 mg/d. It can be administered PO, SC, or IV. IM injections should not be used, except when this is the only way to get a patient under control. Olanzapine, an atypical neuroleptic, has shown significant effectiveness in completely resolving delirium in cancer patients. It has other beneficial effects for terminally ill patients, including antinausea, antianxiety, and weight gain. It is useful for patients with longer anticipated life expectancy because it is less likely to cause dysphoria and has a lower risk of dystonic reactions. Also, because it is metabolized through multiple pathways, it can be used in patients with hepatic and renal dysfunction. Olanzapine has the disadvantage that it is available only orally and that it takes a week to reach steady state. The usual dose is 2.5–5 mg PO bid. Chlorpromazine (10–25 mg every 4–6 h) can be useful if sedation is desired and can be administered IV or PR in addition to PO. Dystonic reactions resulting from dopamine blockade are a side effect of neuroleptics, although they are reported to be rare when these drugs are used to treat terminal delirium. If patients develop dystonic reactions, benztropine should be administered. Neuroleptics may be combined with lorazepam to reduce agitation when the delirium is the result of alcohol or sedative withdrawal.

Table 9-7 Medications for the Management of Delirium

Interventions	Dose
Neuroleptics	
Haloperidol	0.5–5 mg q2–12h, PO/IV/SC/IM
Thioridazine	10–75 mg q4–8h, PO
Chlorpromazine	12.5–50 mg q4–12h, PO/IV/IM
Atypical neuroleptics	
Olanzapine	2.5–5 mg qd or bid, PO
Risperidone	1–3 mg q12h, PO
Anxiolytics	
Lorazepam	0.5–2 mg q1–4h, PO/IV/IM

Midazolam	1–5 mg/h continuous infusion, IV/SC
Anesthetics	
Propofol	0.3–2.0 mg/h continuous infusion, IV

If no response to first-line therapy is seen, a specialty consultation should be obtained with a change to a different medication. If patients fail to improve after a second neuroleptic, sedation with an anesthetic such as propofol or continuous-infusion midazolam may be necessary. By some estimates, at the very end of life as many as 25% of patients experiencing delirium, especially restless delirium with myoclonus or convulsions, may require sedation.

Physical restraints should be used with great reluctance only when the patients violence is threatening to self or others. If they are used, their appropriateness should be reevaluated frequently.

INSOMNIA

Frequency

Sleep disorders, defined as difficulty initiating sleep or maintaining sleep, sleep difficulty at least 3 nights a week, or sleep difficulty that causes impairment of daytime functioning, occurs in 19–63% of patients with advanced cancer. Some 30–74% of patients with other end-stage conditions, including AIDS, heart disease, COPD, and renal disease, experience insomnia.

Etiology

Patients with cancer may have changes in sleep efficiency such as an increase in stage I sleep. Other etiologies of insomnia are coexisting physical illness such as thyroid disease and coexisting psychological illnesses such as depression and anxiety. Medications, including antidepressants, psychostimulants, steroids, and β agonists, are significant contributors to sleep disorders, as are caffeine and alcohol. Multiple over-the-counter medications contain caffeine and antihistamines, which can contribute to sleep disorders.

Assessment

Assessment should include specific questions concerning sleep onset, sleep maintenance, and early-morning wakening as these will provide clues to the causative agents and to management. Patients should be asked about previous sleep problems, screened for depression and anxiety, and asked about symptoms of thyroid disease. Caffeine and alcohol are prominent causes of sleep problems, and a careful history of the use of these substances should be obtained. Both excessive use and withdrawal from alcohol can be causes of sleep problems.

Interventions

The mainstays of intervention include improvement of sleep hygiene (encouragement of regular time for sleep, decreased nighttime distractions, elimination of caffeine and other stimulants and alcohol), intervention to treat anxiety and depression, and treatment for the insomnia itself. For patients with depression who have insomnia and anxiety, a sedating antidepressant such as mirtazapine can be helpful. In the elderly, trazodone, beginning at 25 mg at nighttime, is an effective sleep aid at doses lower than those which cause its antidepressant effect. Zolpidem may have a decreased incidence of delirium in patients compared with traditional benzodiazepines, but this has not been clearly established. When benzodiazepines are prescribed, short-acting ones (such as lorazepam) are favored over longer-acting (such as diazepam). Patients who receive these medications should be observed for signs of increased confusion and delirium.

Social Needs and Their Management

FINANCIAL BURDENS

Frequency

Dying can impose substantial economic strains on patients and families, causing distress. In the United States, with one of the least comprehensive health insurance systems among the developed countries, ~20% of terminally ill patients and their families spend >10% of family income on health care costs over and above health insurance premiums. Between 10 and 30% of families sell assets, use savings, or take out a mortgage to pay for the patients health care costs. Nearly 40% of terminally ill patients in the United States report that the cost of their illness is a moderate or great economic hardship for their family.

The patient is likely to reduce and eventually stop working. In 20% of cases, a family member of the terminally ill patient also stops working to provide care. The major underlying causes of economic burden are related to poor physical functioning and care needs, such as the need for housekeeping, nursing, and personal care. More debilitated patients and poor patients experience greater economic burdens.

Intervention

This economic burden should not be ignored as a private matter. It has been associated with a number of adverse health outcomes, including preferring comfort care over life-prolonging care as well as consideration of euthanasia or physician-assisted suicide. Economic burdens increase the psychological distress of families and caregivers of terminally ill patients, and poverty is associated with many adverse health outcomes. Importantly, recent studies found that "patients with advanced cancer who reported having end-of-life conversations with physicians had significantly lower health care costs in their final week of life. Higher costs were associated with worse quality of death." Assistance from a social worker, early on if possible, to ensure access to all available benefits may be helpful. Many patients, families, and health care providers are unaware of options for long-term care insurance, respite care, the Family Medical Leave Act (FMLA), and other sources of assistance. Some of these options (such as respite care) may be part of a formal hospice program but others (such as the FMLA) do not require enrollment in a hospice program.

RELATIONSHIPS

Frequency

Settling personal issues and closing the narrative of lived relationships are universal needs. When asked if sudden death or death after an illness is preferable, respondents often initially select the former but soon change to the latter as they reflect on the importance of saying goodbye. Bereaved family members who have not had the chance to say goodbye often have a more difficult grief process.

Interventions

Care of seriously ill patients requires efforts to facilitate the types of encounters and time spent with family and friends that are necessary to meet those needs. Family and close friends may need to be accommodated with unrestricted visiting hours, which may include sleeping near the patient even in otherwise regimented institutional settings. Physicians and other health care providers may be able to facilitate and resolve strained interactions between the patient and other family members. Assistance for patients and family members who are unsure about how to create or help preserve memories, whether by providing materials such as a scrapbook or memory box or by offering them suggestions and informational resources, can be deeply appreciated. Taking photographs and creating videos can be especially helpful to terminally ill patients who have younger children or grandchildren.

FAMILY CAREGIVERS

Frequency

Caring for seriously ill patients places a heavy burden on families. Families frequently are required to provide transportation and homemaking as well as other services. Typically, paid professionals such as home health nurses and hospice workers supplement family care; only about a quarter of all caregiving consists of exclusively paid professional assistance. The trend toward more out-of-hospital deaths will increase reliance on families for end-of-life care. Increasingly, family members are being called upon to provide physical care (such as moving and bathing patients) and medical care (such as assessing symptoms and giving medications) in addition to emotional care and support.

Three-quarters of family caregivers of terminally ill patients are women—wives, daughters, sisters, and even daughters-in-law. Since many are widowed, women tend to be able to rely less on family for caregiving assistance and may need more paid assistance. About 20% of terminally ill patients report substantial unmet needs for nursing and personal care. The impact of caregiving on family caregivers is substantial: both bereaved and current caregivers have a higher mortality rate than that of non-caregiving controls.

Interventions

It is imperative to inquire about unmet needs and to try to ensure that those needs are met either through the family or by paid professional services when possible. Community assistance through houses of worship or other community groups often can be mobilized by telephone calls from the medical team to someone the patient or family identifies. Sources of support specifically for family caregivers should be identified through local sources or nationally through groups such as the National Family Caregivers Association (www.nfcacares.org), the American Cancer Society (www.cancer.org), and the Alzheimers Association (www.alz.org).

Existential Needs and Their Management

Frequency

Religion and spirituality are often important to dying patients. Nearly 70% of patients report becoming more religious or spiritual when they became terminally ill, and many find comfort in religious or spiritual practices such as prayer. However, ~20% of terminally ill patients become less religious, frequently feeling cheated or betrayed by becoming terminally ill. For other patients, the need is for existential meaning and purpose that is distinct from and may even be antithetical to religion or spirituality. When asked, patients and family caregivers frequently report wanting their professional caregivers to be more attentive to religion and spirituality.

Assessment

Health care providers are often hesitant about involving themselves in the religious, spiritual, and existential experiences of their patients because it may seem private or not relevant to the current illness. But physicians and other members of the care team should be able at least to detect spiritual and existential needs. Screening questions have been developed for a physicians spiritual history taking. Spiritual distress can amplify other types of suffering and even masquerade as intractable physical pain, anxiety, or depression. The screening questions in the comprehensive assessment are usually sufficient. Deeper evaluation and intervention are rarely appropriate for the physician unless no other member of a care team is available or suitable. Pastoral care providers may be helpful, whether from the medical institution or from the patients own community.

Interventions

Precisely how religious practices, spirituality, and existential explorations can be facilitated and improve end-of-life care is not well established. What is clear is that for physicians, one main intervention is to inquire about the role and importance of spirituality and religion in a patients life. This will help a patient feel heard and help physicians identify specific needs. In one study, only 36% of respondents indicated that a clergy member would be comforting. Nevertheless, the increase in religious and spiritual interest among a substantial fraction of dying patients suggests inquiring of individual patients how this need can be addressed. Some evidence supports specific methods of addressing existential needs in patients, ranging from establishing a supportive group environment for terminal patients to individual treatments emphasizing a patients dignity and sources of meaning.

MANAGING THE LAST STAGES

Withdrawing and Withholding Life-Sustaining Treatment

Legal Aspects

For centuries, it has been deemed ethical to withhold or withdraw life-sustaining interventions. The current legal consensus in the United States and most developed countries is that patients have a moral as well as constitutional or common law right to refuse medical interventions. American courts also have held that incompetent patients have a right to refuse medical interventions. For patients who are incompetent and terminally ill and who have not completed an advance care directive, next of kin can exercise that right, although this may be restricted in some states, depending how clear and convincing the evidence is of the patients preferences. Courts have limited families ability to terminate life-sustaining treatments in patients who are conscious, incompetent, but not terminally ill. In theory, patients right to refuse medical therapy can be limited by four countervailing interests: (1) preservation of life, (2) prevention of suicide, (3) protection of third parties such as children, and (4) preservation of the integrity of the medical profession. In practice, these interests almost never override the right of competent patients and incompetent patients who have left explicit and advance care directives.

For incompetent patients who either appointed a proxy without specific indications of their wishes or never completed an advance care directive, three criteria have been suggested to guide the decision to terminate medical interventions. First, some commentators suggest that ordinary care should be administered but extraordinary care could be terminated. Because the ordinary/extraordinary distinction is too vague, courts and commentators widely agree that it should not be used to justify decisions about stopping treatment. Second, many courts have advocated the use of the substituted-judgment criterion, which holds that the proxy decision-makers should try to imagine what the incompetent patient would do if he or she were competent. However, multiple studies indicate that many proxies, even close family members, cannot accurately predict what the patient would have wanted. Therefore, substituted judgment becomes more of a guessing game than a way of fulfilling the patients wishes. Finally, the best-interests criterion holds that proxies should evaluate treatments by balancing their benefits and risks and select those treatments in which the benefits maximally outweigh the burdens of treatment. Clinicians have a clear and crucial role in this by carefully and dispassionately explaining the known benefits and burdens of specific treatments. Yet even when that information is as clear as possible, different individuals can have very different views of what is in the patients best interests, and families may have disagreements or even overt conflicts. This criterion has been criticized because there is no single way to determine the balance between benefits and burdens; it depends on a patients personal values. For instance, for some people being alive even if mentally incapacitated is a benefit, whereas for others it may be the worst possible existence. As a matter of practice, physicians rely on family members to make decisions that they feel are best and object only if those decisions seem to demand treatments that the physicians consider not beneficial.

Practices

Withholding and withdrawing acutely life-sustaining medical interventions from terminally ill patients are now standard practice. More than 90% of American patients die without cardiopulmonary resuscitation (CPR), and just as many forgo other potentially life-sustaining interventions. For instance, in ICUs in the period 1987–1988, CPR was performed 49% of the time, but it was performed only 10% of the time in 1992–1993. On average, 3.8 interventions, such as vasopressors and transfusions, were stopped for each dying ICU patient. However, up to 19% of decedents in hospitals received interventions such as extubation, ventilation, and surgery in the 48 h preceding death. However, practices vary widely

among hospitals and ICUs, suggesting an important element of physician preferences rather than objective data.

Mechanical ventilation may be the most challenging intervention to withdraw. The two approaches are *terminal extubation*, which is the removal of the endotracheal tube, and *terminal weaning*, which is the gradual reduction of the Fi_{O2} or ventilator rate. One-third of ICU physicians prefer to use the terminal weaning technique, and 13% extubate; the majority of physicians utilize both techniques. The American Thoracic Societys 2008 clinical policy guidelines note that there is no single correct process of ventilator withdrawal and that physicians use and should be proficient in both methods but that the chosen approach should carefully balance benefits and burdens as well as patient and caregiver preferences. Physicians assessment of patients likelihood of survival, their prediction of possible cognitive damage, and patients preferences about the use of life support are primary factors in determining the likelihood of withdrawal of mechanical ventilation. Some recommend terminal weaning because patients do not develop upper airway obstruction and the distress caused by secretions or stridor; however, terminal weaning can prolong the dying process and not allow a patients family to be with him or her unencumbered by an endotracheal tube. To ensure comfort for conscious or semiconscious patients before withdrawal of the ventilator, neuromuscular blocking agents should be terminated and sedatives and analgesics administered. Removing the neuromuscular blocking agents permits patients to show discomfort, facilitating the titration of sedatives and analgesics; it also permits interactions between patients and their families. A common practice is to inject a bolus of midazolam (2–4 mg) or lorazepam (2–4 mg) before withdrawal, followed by 5–10 mg of morphine and continuous infusion of morphine (50% of the bolus dose per hour) during weaning. In patients who have significant upper airway secretions, IV scopolamine at a rate of 100 μg/h can be administered. Additional boluses of morphine or increases in the infusion rate should be administered for respiratory distress or signs of pain. Higher doses will be needed for patients already receiving sedatives and opioids. Families need to be reassured about treatments for common symptoms after withdrawal of ventilatory support, such as dyspnea and agitation, and warned about the uncertainty of length of survival after withdrawal of ventilatory support: up to 10% of patients unexpectedly survive for 1 day or more after mechanical ventilation is stopped.

Futile Care

Beginning in the late 1980s, some commentators argued that physicians could terminate futile treatments demanded by the families of terminally ill patients. Although no objective definition or standard of futility exists, several categories have been proposed. Physiologic futility means that an intervention will have no physiologic effect. Some have defined qualitative futility as applying to procedures that "fail to end a patients total dependence on intensive medical care." Quantitative futility occurs "when physicians conclude (through personal experience, experiences shared with colleagues, or consideration of reported empiric data) that in the last 100 cases, a medical treatment has been useless." The term conceals subjective value judgments about when a treatment is "not beneficial." Deciding whether a treatment that obtains an additional 6 weeks of life or a 1% survival advantage confers benefit depends on patients preferences and goals. Furthermore, physicians predictions of when treatments were futile deviated markedly from the quantitative definition. When residents thought CPR was quantitatively futile, more than one in five patients had a >10% chance of survival to hospital discharge. Most studies that purport to guide determinations of futility are based on insufficient data to provide statistical confidence for clinical decision-making. Quantitative futility rarely applies in ICU settings. Many commentators reject using futility as a criterion for withdrawing care, preferring instead to consider futility situations as ones that represent conflict that calls for careful negotiation between families and health care providers.

In the wake of a lack of consensus over quantitative measures of futility, many hospitals adopted process-based approaches to resolve disputes over futility and enhance communication with patients and surrogates, including focusing on interests and alternatives rather than opposing positions and generating a wide range of options. Some hospitals have enacted "unilateral DNR" policies to allow clinicians to provide a do-not-resuscitate order in cases in which consensus cannot be reached with families and medical opinion is that resuscitation would be futile if attempted. This type of a policy is not a replacement for careful and patient communication and negotiation but recognizes that agreement cannot always be reached. Over the last 15 years, many states, such as Texas, Virginia, Maryland, and California, have enacted so-called medical futility laws that provide physicians a "safe harbor" from liability if they refuse a patient or familys request for life-sustaining interventions. For instance, in Texas when a disagreement about terminating interventions between the medical team and the family has not been resolved by an ethics consultation, the hospital is supposed to try to facilitate transfer of the patient to an institution willing to provide treatment. If this fails after 10 days, the hospital and physician may unilaterally withdraw treatments determined to be futile. The family may appeal to a state court. Early data suggest that the law increases futility consultations for the ethics committee and that although most families concur with withdrawal, about 10–15% of families refuse to withdraw treatment. Approximately 12 cases have gone to court in Texas in the 7 years since the adoption of the law. As of 2007, there had been 974 ethics committee consultations on medical futility cases and 65 in which committees ruled against families and gave notice that treatment would be

terminated. Treatment was withdrawn for 27 of those patients, and the remainder transferred to other facilities or died while awaiting transfer.

Euthanasia and Physician-Assisted Suicide

Euthanasia and physician-assisted suicide are defined in Table 9-8. Terminating life-sustaining care and providing opioid medications to manage symptoms have long been considered ethical by the medical profession and legal by courts and should not be confused with euthanasia or physician-assisted suicide.

Table 9-8 Definitions of Assisted Suicide and Euthanasia

Term	Definition	Legal Status
Voluntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death with the patient's informed consent	Netherlands Belgium
Involuntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death when the patient was competent to consent but did not —e.g., the patient may not have been asked	Nowhere
Passive euthanasia	Withholding or withdrawing life-sustaining medical treatments from a patient to let him or her die (terminating life-sustaining treatments)	Everywhere
Physician-assisted suicide	A physician provides medications or other interventions to a patient with the understanding that the patient can use them to commit suicide	Oregon Netherlands Belgium Switzerland

Legal Aspects

Euthanasia is legal in the Netherlands, Belgium, and Luxembourg. It was legalized in the Northern Territory of Australia in 1995, but that legislation was repealed in 1997. Euthanasia is not legal in any state in the United States. With certain conditions, in Switzerland, a layperson can legally assist suicide. In the United States, physician-assisted suicide is legal in Oregon and Washington State if multiple criteria are met and then only after a process that includes a 15-day waiting period. In 2009, the state supreme court of Montana ruled that state law permits physician-assisted suicide for terminally ill patients. In all other countries and all other states in the United States, physician-assisted suicide and euthanasia are illegal explicitly or by common law.

Practices

Fewer than 10–20% of terminally ill patients actually consider euthanasia and/or physician-assisted suicide for themselves. In the Netherlands and Oregon, >70% of patients utilizing these interventions are dying of cancer; <10% of deaths by euthanasia or physician-assisted suicide involve patients with AIDS or amyotrophic lateral sclerosis. In the Netherlands, the share of deaths attributable to euthanasia or physician-assisted suicide declined from around 2.8% of all deaths in 2001 to around 1.8% in 2005. In 2009, the last year with complete data, around 60 patients in Oregon (~0.2% of all deaths) died by physician-assisted suicide, although this may be an underestimate. In Washington State, between March 2009 (when the law allowing physician-assisted suicide went into force) and December 2009, 36 individuals died from prescribed lethal doses.

Pain is not a primary motivator for patients requests for or interest in euthanasia and/or physician-assisted suicide. Among the first patients to receive physician-assisted suicide in Oregon, only 1 patient of 15 had inadequate pain control compared with 15 of 43 patients in a control group experiencing inadequate pain relief. Depression, hopelessness, and, more profoundly, concerns about loss of dignity or autonomy or being a burden on family members appear to be primary factors motivating a desire for euthanasia or physician-assisted suicide. In Oregon, fewer than 25% of patients cite pain as the reason for desiring physician-assisted suicide. Most cite losing autonomy, dignity, or enjoyable activities. Over a third note being a burden on family. A study from the Netherlands showed that depressed terminally ill cancer patients were four times more likely to request euthanasia and confirmed that uncontrolled pain was not associated with greater interest in euthanasia.

Euthanasia and physician-assisted suicide are no guarantee of a painless, quick death. Data from the Netherlands indicate that in as many as 20% of cases technical and other problems arose, including patients waking from coma, not becoming comatose, regurgitating medications, and experiencing a prolonged time to death. Data from Oregon indicate that between 1997 and 2009, 20 patients (around 5%) regurgitated after taking prescribed medication, 1 patient awaked, and none experienced seizures. Problems were significantly more common in physician-assisted suicide, sometimes requiring the physician to intervene and provide euthanasia.

Whether practicing in a setting where euthanasia is legal or not, over a career, 12–54% of physicians receive a request for euthanasia or physician-assisted suicide from a patient. Competency in dealing with such a request is crucial. Although challenging, the request can also provide a chance to address intense suffering. After receiving a request for euthanasia and/or physician-assisted suicide, health care providers should carefully clarify the request with empathic, open-ended questions to help elucidate the underlying cause for the request, such as: "What makes you want to consider this option?" Endorsing either moral opposition or moral support for the act tends to be counterproductive, giving an impression of being judgmental or of endorsing the idea that the patients life is worthless. Health care providers must reassure the patient of continued care and commitment. The patient should be educated about alternative, less controversial options, such as symptom management and withdrawing any unwanted treatments and the reality of euthanasia and/or physician-assisted suicide, since the patient may have misconceptions about their effectiveness as well as the legal implications of the choice. Depression, hopelessness, and other symptoms of psychological distress as well as physical suffering and economic burdens are likely factors motivating the request, and such factors should be assessed and treated aggressively. After these interventions and clarification of options, most patients proceed with another approach, declining life-sustaining interventions, possibly including refusal of nutrition and hydration.

Care during the Last Hours

Most laypersons have limited experiences with the actual dying process and death. They frequently do not know what to expect of the final hours and afterward. The family and other caregivers must be prepared, especially if the plan is for the patient to die at home.

Patients in the last days of life typically experience extreme weakness and fatigue and become bedbound; this can lead to pressure sores. The issue of turning patients who are near the end of life, however, must be balanced against the potential discomfort that movement may cause. Patients stop eating and drinking with drying of mucosal membranes and dysphagia. Careful attention to oral swabbing, lubricants for lips, and use of artificial tears can provide a form of care to substitute for attempts at feeding the patient. With loss of the gag reflex and dysphagia, patients may also experience accumulation of oral secretions, producing noises during respiration sometimes called "the death rattle." Scopolamine can reduce the secretions. Patients also experience changes in respiration with periods of apnea or Cheyne-Stokes breathing. Decreased intravascular volume and cardiac output cause tachycardia, hypotension, peripheral coolness, and livedo reticularis (skin mottling). Patients can have urinary and, less frequently, fecal incontinence. Changes in consciousness and neurologic function generally lead to two different paths to death (Fig. 9-2).

Figure 9-2

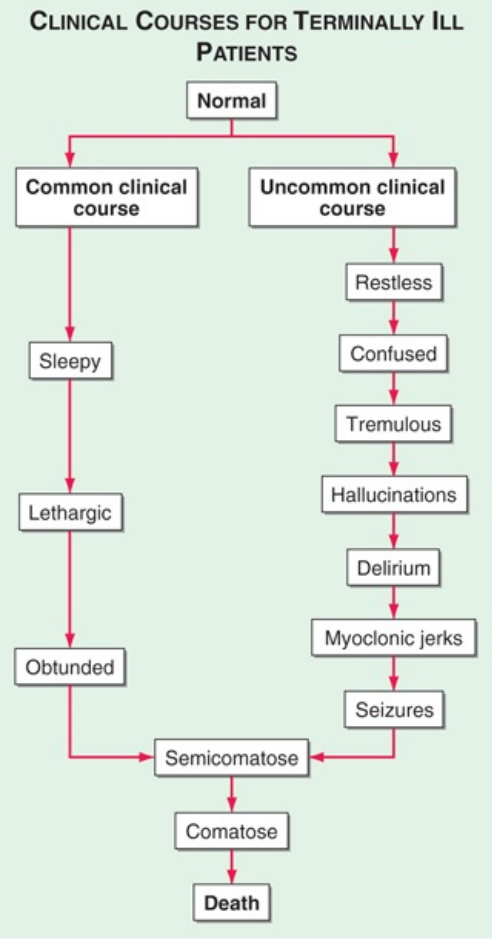

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Common and uncommon clinical courses in the last days of terminally ill patients. (*Adapted from FD Ferris et al: Module 4: Palliative care, in Comprehensive Guide for the Care of Persons with HIV Disease. Toronto: Mt. Sinai Hospital and Casey Hospice, 1995, http://www.cpsonline.info/content/resources/hivmodule/module4complete.pdf.*)

Each of these terminal changes can cause patients and families distress, requiring reassurance and targeted interventions (Table 9-9). Informing families that these changes might occur and providing them with an information sheet can help preempt problems and minimize distress. Understanding that patients stop eating because they are dying, not dying because they have stopped eating, can reduce family and caregiver anxiety. Similarly, informing the family and caregivers that the "death rattle" may occur and that it is not indicative of suffocation, choking, or pain can reduce their worry from the breathing sounds.

Table 9-9 Managing Changes in the Patient's Condition during the Final Days and Hours

Changes in the Patient's Condition	Potential Complication	Family's Possible Reaction and Concern	Advice and Intervention
Profound fatigue	Bedbound with development of pressure ulcers that are prone to infection, malodor, and pain, and joint pain	Patient is lazy and giving up.	Reassure family and caregivers that terminal fatigue will not respond to interventions and should not be resisted. Use an air mattress if necessary.
Anorexia	None	Patient is giving up; patient will suffer from hunger and will starve to death.	Reassure family and caregivers that the patient is not eating because he or she is dying; not eating at the end of life does not cause suffering or death. Forced feeding, whether oral, parenteral, or enteral, does not reduce symptoms or prolong life.
Dehydration	Dry mucosal membranes	Patient will suffer	Reassure family and caregivers that

	(see below)	from thirst and die of dehydration.	dehydration at the end of life does not cause suffering because patients lose consciousness before any symptom distress. Intravenous hydration can worsen symptoms of dyspnea by pulmonary edema and peripheral edema as well as prolong dying process.
Dysphagia	Inability to swallow oral medications needed for palliative care		Do not force oral intake. Discontinue unnecessary medications that may have been continued, including antibiotics, diuretics, antidepressants, and laxatives. If swallowing pills is difficult, convert essential medications (analgesics, antiemetics, anxiolytics, and psychotropics) to oral solutions, buccal, sublingual, or rectal administration.
"Death rattle" —noisy breathing		Patient is choking and suffocating.	Reassure the family and caregivers that this is caused by secretions in the oropharynx and the patient is not choking. Reduce secretions with scopolamine (0.2–0.4 mg SC q4h or 1–3 patches q3d). Reposition patient to permit drainage of secretions. Do not suction. Suction can cause patient and family discomfort and is usually ineffective.
Apnea, Cheyne-Stokes respirations, dyspnea		Patient is suffocating.	Reassure family and caregivers that unconscious patients do not experience suffocation or air hunger. Apneic episodes are frequently a premorbid change. Opioids or anxiolytics may be used for dyspnea. Oxygen is unlikely to relieve dyspneic symptoms and may prolong the dying process.
Urinary or fecal incontinence	Skin breakdown if days until death Potential transmission of infectious agents to caregivers	Patient is dirty, malodorous, and physically repellent.	Remind family and caregivers to use universal precautions. Frequent changes of bedclothes and bedding. Use diapers, urinary catheter, or rectal tube if diarrhea or high urine output.
Agitation or delirium	Day/night reversal Hurt self or caregivers	Patient is in horrible pain and going to have a horrible death.	Reassure family and caregivers that agitation and delirium do not necessarily connote physical pain. Depending on the prognosis and goals of treatment, consider evaluating for causes of delirium and modify medications. Manage symptoms with haloperidol, chlorpromazine, diazepam, or midazolam.
Dry mucosal membranes	Cracked lips, mouth sores, and candidiasis can also cause pain. Odor	Patient may be malodorous, physically repellent.	Use baking soda mouthwash or saliva preparation q15–30min. Use topical nystatin for candidiasis. Coat lips and nasal mucosa with petroleum jelly q60–90min. Use ophthalmic lubricants q4h or artificial tears q30min.

Families and caregivers may also feel guilty about stopping treatments, fearing that they are "killing" the patient. This may lead to demands for interventions, such as feeding tubes, that may be ineffective. In such cases, the physician should remind the family and caregivers about the inevitability of events and the palliative goals. Interventions may prolong the dying process and cause discomfort. Physicians also should emphasize that withholding treatments is both legal and ethical and that the family members are not the cause of the patients death. This reassurance may have to be provided multiple times.

Hearing and touch are said to be the last senses to stop functioning. Whether this is the case or not, families and caregivers can be encouraged to communicate with the dying patient. Encouraging them to talk directly to the patient, even if he or she is unconscious, and hold the patients hand or demonstrate affection in other ways can be an effective way to channel their urge "to do something" for the patient.

When the plan is for the patient to die at home, the physician must inform the family and caregivers how to determine that the patient has died. The cardinal signs are cessation of cardiac function and respiration; the pupils become fixed; the body becomes cool; muscles relax; and incontinence may occur. Remind the family and caregivers that the eyes may remain open even after the patient has died because the retroorbital fat pad may be depleted, permitting the orbit to fall posteriorly, which makes it difficult for the eyelids to cover the eyeball.

The physician should establish a plan for who the family or caregivers will contact when the patient is dying or has died. Without a plan, they may panic and call 911, unleashing a cascade of unwanted events, from arrival of emergency personnel and resuscitation to hospital admission. The family and caregivers should be instructed to contact the hospice (if one is involved), the covering physician, or the on-call member of the palliative care team. They should also be told that the medical examiner need not be called unless the state requires it for all deaths. Unless foul play is suspected, the health care team need not contact the medical examiner either.

Just after the patient dies, even the best-prepared family may experience shock and loss and be emotionally distraught. They need time to assimilate the event and be comforted. Health care providers are likely to find it meaningful to write a bereavement card or letter to the family. The purpose is to communicate about the patient, perhaps emphasizing the patients virtues and the honor it was to care for the patient, and to express concern for the familys hardship. Some physicians attend the funerals of their patients. Although this is beyond any medical obligation, the presence of the physician can be a source of support to the grieving family and provides an opportunity for closure for the physician.

Death of a spouse is a strong predictor of poor health, and even mortality, for the surviving spouse. It may be important to alert the spouses physician about the death so that he or she is aware of symptoms that might require professional attention.

PALLIATIVE CARE SERVICES: HOW AND WHERE

Determining the best approach to providing palliative care to patients will depend on patient preferences, the availability of caregivers and specialized services in close proximity, institutional resources, and reimbursement. Hospice is a leading, but not the only, model of palliative care services. In the United States, a plurality—40.7%—of hospice care is provided in residential homes. In 2008, just over 20% of hospice care was provided in nursing homes. In the United States, Medicare pays for hospice services under Part A, the hospital insurance part of reimbursement. Two physicians must certify that the patient has a prognosis of ≤6 months if the disease runs its usual course. Prognoses are probabilistic by their nature; patients are not required to die within 6 months but rather to have a condition from which half the individuals with it would not be alive within 6 months. Patients sign a hospice enrollment form that states their intent to forgo curative services related to their terminal illness, but they can still receive medical services for other comorbid conditions. Patients also can withdraw enrollment and reenroll later; the hospice Medicare benefit can be revoked later to secure traditional Medicare benefits. Payments to the hospice are per diem (or capitated), not fee-for-service. Payments are intended to cover physician services for the medical direction of the care team; regular home care visits by registered nurses and licensed practical nurses; home health aid and homemaker services; chaplain services; social work services; bereavement counseling; and medical equipment, supplies, and medications. No specific therapy is excluded, and the goal is for each therapy to be considered for its symptomatic (as opposed to disease-modifying) effect. Additional clinical care, including services of the primary physician, is covered by Medicare Part B even while the hospice Medicare benefit is in place. The health reform legislation signed into law in March 2010—the Affordable Care Act—directs the Secretary of Health and Human Services to gather data on Medicare hospice reimbursement with the goal of reforming payment rates to account for resource use over an entire episode of care. The legislation also requires additional evaluations and reviews of eligibility for hospice care by hospice physicians or nurses. Finally, the legislation establishes a demonstration project for concurrent hospice care in Medicare, which would test and evaluate allowing patients to remain eligible for regular Medicare during hospice care.

By 2008, the mean length of enrollment in a hospice was around 70 days, with the median being 21

days. Such short stays create barriers to establishing high-quality palliative services in patients' homes and also place financial strains on hospice providers since the initial assessments are resource intensive. Physicians should initiate early referrals to the hospice to allow more time for patients to receive palliative care.

Hospice care has been the main method for securing palliative services for terminally ill patients. However, efforts are being made to ensure continuity of palliative care across settings and through time. Palliative care services are becoming available as consultative services and more rarely as palliative care units in hospitals, in day care and other outpatient settings, and in nursing homes. Palliative care consultations for nonhospice patients can be billed as for other consultations under Medicare Part B, the physician reimbursement part. Many believe palliative care should be offered to patients regardless of their prognosis. A patient, his or her family, and physicians should not have to make a "curative vs. palliative care" decision because it is rarely possible to make such a decisive switch to embracing mortality.

FUTURE DIRECTIONS

Outcome Measures

Care near the end of life cannot be measured by most of the available validated outcome measures since palliative care does not consider death a bad outcome. Similarly, the family and patients receiving end-of-life care may not desire the elements elicited in current quality-of-life measurements. Symptom control, enhanced family relationships, and quality of bereavement are difficult to measure and are rarely the primary focus of carefully developed or widely used outcome measures. Nevertheless, outcomes are as important in end-of-life care as in any other field of medical care. Specific end-of-life care instruments are being developed both for assessment, such as The Brief Hospice Inventory and NEST (*n*eeds near the *e*nd of life *s*creening *t*ool), and for outcome measures, such as the Palliative Care Outcomes Scale, as well as for prognosis, such as the Palliative Prognostic Index. The field of end-of-life care is entering an era of evidence-based practice and continuous improvement through clinical trials.

FURTHER READINGS

Web Sites

American Academy of Hospice and Palliative Medicine: www.aahpm.org

Center to Advance Palliative Care: http://www.capc.org

Education in Palliative and End of Life Care (EPEC): http://www.epec.net

End of Life—Palliative Education Resource Center: http://www.eperc.mcw.edu

Family Caregiver Alliance: http://www.caregiver.org

The Medical Directive: http://www.medicaldirective.org

National Family Caregivers Association: http://www.nfcacares.org/

National Hospice and Palliative Care Organization (including state-specific advance directives): http://www.nhpco.org

NCCN: The National Comprehensive Cancer Network palliative care guidelines: http://www.nccn.org

Books

American Society of Clinical Oncology: *Optimizing Cancer Care—The Importance of Symptom Management,* vols 1 and 2. Alexandria, VA, ASCO, 2001

Buckman R: *How to Break Bad News: A Guide for Health Care Professionals.* Baltimore, Johns Hopkins University Press, 1992

Kuebler KK: *Palliative and End-of-Life Care: Clinical Practice Guide.* Philadelphia, Saunders, 2006

Meier DE et al (eds): *Palliative Care: Transforming the Care of Serious Illness.* San Francisco, Jossey-Bass, 2010

Articles

Christakis NA, Allison PD: Mortality after the hospitalization of a spouse. N Engl J Med 354:719, 2006[PMID: 16481639]

Emanuel L et al: Integrating palliative care into disease management guidelines. J Palliat Med 7:774, 2004[PMID: 15684844]

Gabbay E et al: The empirical basis for determinations of medical futility. J Gen Intern Med 2010

Hugel H et al: The prevalence, key causes and management of insomnia in palliative care patients. J Pain Symptom Manage 27:316, 2004[PMID: 15050659]

Kapo J et al: Palliative care for the older adult. J Palliat Med 10:185, 2007[PMID: 17298269]

Lanken PN et al: An official American Thoracic Society clinical policy statement: Palliative care for patients with respiratory diseases and critical illnesses. Am J Respir Crit Care Med 177:912 2008

Larken PJ: The management of constipation in palliative care: Clinical practice recommendations. J Palliat Med 22:796, 2008

Lorenz K et al: Evidence for improving palliative care at the end of life: A systematic review. Ann Intern Med 148:147 2008

Lynn J: Serving patients who may die soon and their families: The role of hospice and other services. JAMA 285:925, 2001[PMID: 11180736]

Meisel A et al: Seven legal barriers to end-of-life care: Myths, realities, and grains of truth. JAMA 284:2495, 2000[PMID: 11074780]

Morrison RS, Meier DE: Palliative care. N Engl J Med 350:2582, 2004[PMID: 15201415]

Morrow GR: Guidelines for the treatment of chemotherapy-induced nausea and vomiting. Clin Adv Hematol Oncol 8:4, 2010

Murray SA et al: Illness trajectories and palliative care. BMJ 330:1007, 2005[PMID: 15860828]

Qaseem A et al: Evidence-based interventions to improve the palliative care of pain, dyspnea, and depression at the end of life: A clinical practice guideline from the American College of Physicians. Ann Intern Med 148:141, 2008[PMID: 18195338]

Radbruch L et al: Fatigue in palliative care patients—an EAPC approach. Palliat Med 22:13, 2008[PMID: 18216074]

Silveira M et al: Advance directives and outcomes of surrogate decision making before death. N Engl J Med 362:1211, 2010[PMID: 20357283]

Van den Beuken-van Everdingen MH et al: Prevalence of pain in patients with cancer: A systematic review of the past 40 years. Ann Oncol 18:1437, 2007

Wright AA et al: Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA 300:1665, 2008[PMID: 18840840]

Zhang B et al: Health care costs in the last week of life: Associations with end-of-life conversations. Arch Intern Med 169:480, 2009[PMID: 19273778]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE SAFETY AND QUALITY OF HEALTH CARE: INTRODUCTION

The safety and quality of care are two of the central dimensions of health care. It is increasingly clear that both could be much better, and in recent years it has become easier to measure safety and quality. In addition, the public is—with good justification—demanding measurement and accountability, and payment for services increasingly will be based on performance in these areas. Thus, physicians must learn about these two domains, how they can be improved, and the relative strengths and limitations of the current ability to measure them.

Safety and quality are closely related but do not completely overlap. The Institute of Medicine has suggested in a seminal series of reports that safety is the first part of quality and that health care first must guarantee that it will deliver safe care, although quality is also pivotal. In the end, it is likely that more net clinical benefit will be derived from improving quality than from improving safety, though both are important and safety is in many ways more tangible to the public. Accordingly, the first section of this chapter will address issues relating to the safety of care and the second will cover quality of care.

Safety in Health Care

SAFETY THEORY

Safety theory clearly points out that individuals make errors all the time. Think of driving home from the hospital; you intend to stop and pick up a quart of milk on the way home but find yourself entering your driveway without realizing how you got there. Everybody uses low-level, semiautomatic behavior for many activities in daily life; this kind of error is called a "slip." Slips occur often during care delivery, e.g., when people intend to write an order but forget because they have to complete another action first. "Mistakes," by contrast, are errors of a higher level; they occur in new or nonstereotypic situations in which conscious decisions are being made. An example would be dosing a medication with which a physician is not familiar. The strategies used to prevent slips and mistakes are often different.

Systems theory suggests that most accidents occur as the result of a series of small failures that happen to line up in an individual instance so that an accident can occur **(Fig. 10-1)**. It also suggests that most individuals in an industry such as health care are trying to do the right thing (e.g., deliver safe care), and most accidents thus can be seen as resulting from defects in the systems. Correspondingly, systems should be designed both to make errors less likely and to identify those which do occur, as some inevitably will.

Figure 10-1

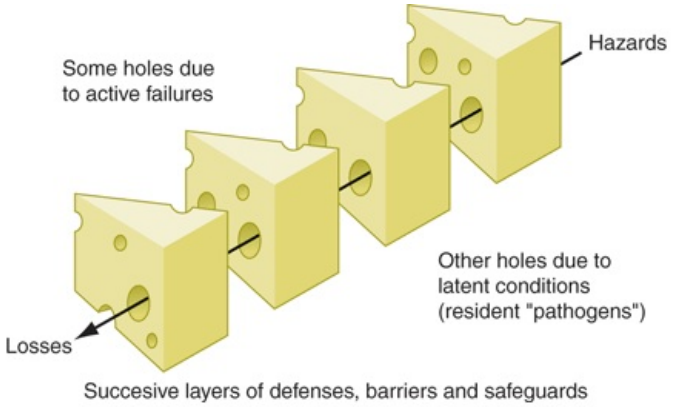

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

"Swiss cheese" diagram. Reason has argued that most accidents occur when a series of "latent failures" in a system are present and happen to line up in a given instance, resulting in an accident. Examples of latent failures in the case of a fall might be that the unit was unusually busy that day and that the floor happened to be wet. *(Adapted from J Reason: Human error: Models and management. BMJ 320:768–770, 2000; with permission.)*

FACTORS THAT INCREASE THE LIKELIHOOD OF ERRORS

Many factors ubiquitous in health care systems can increase the likelihood of errors, including fatigue, stress, interruptions, complexity, and transitions. The effects of fatigue in other industries are clear, but its effects in health care have been more controversial until recently. For example, the accident rate in truck drivers increases dramatically if they work over a certain number of hours in a week, especially with prolonged shifts. A recent study of house officers in the intensive care unit demonstrated that they were about one-third more likely to make errors when they were on a 24-h shift than when they were on a schedule that allowed them to sleep 8 h the previous night. The American College of Graduate Medical Education (ACGME) has moved to address this issue by putting in place the 80-h workweek. Although this is a step forward, it does not address the most important cause of fatigue-related errors: extended-duty shifts. High levels of stress and workload also can increase error rates. Thus, in extremely high-pressure situations, such as cardiac arrests, errors are more likely to occur. Strategies such as using protocols in these settings can be helpful, as can simply recognizing that the situation is stressful.

Interruptions also increase the likelihood of error and occur frequently in health care delivery. It is common to forget to complete an action when one is interrupted partway through it by a page, for example. Approaches that may be helpful in this area include minimizing the use of interruptions and setting up tools that help define the urgency of an interruption.

In addition, complexity represents a key issue that contributes to errors. Providers are confronted by streams of data, such as laboratory tests and vital signs, many of which provide little useful information but some of which are important and require action or suggest a specific diagnosis. Tools that emphasize specific abnormalities or combinations of abnormalities may be helpful in this area.

Transitions between providers and settings are also common in health care, especially with the advent of the 80-h workweek, and generally represent vulnerabilities. Tools that provide structure in exchanging information, e.g., when transferring care between providers, may be helpful.

THE FREQUENCY OF ADVERSE EVENTS IN HEALTH CARE

Most large studies focusing on the frequency and consequences of adverse events have been performed in the inpatient setting; some data are available for nursing homes, and much less information is available about the outpatient setting. The Harvard Medical Practice Study, one of the largest studies to address this issue, was performed with hospitalized patients in New York. The primary outcome was the adverse event, which is an injury caused by medical management rather than the patients underlying disease. In this study, an event either resulted in death or disability at discharge or prolonged the length of stay by at least 2 days. Key findings were that the adverse event rate was 3.7%, and 58% of the adverse events were considered preventable. Although New York is not representative of the rest of the country, the study was replicated later in Colorado and Utah, where the rates were essentially similar. Since then, other studies have been performed in a variety of developed nations using analogous methodologies, and the rates in developed countries appear to be ~10%. The World Health Organization has focused on this area, forming the World Alliance for Patient Safety, and rates of safety issues in developing and transitional countries appear to be even higher; thus, this is clearly an issue of global proportions.

In the Medical Practice Study, adverse drug events (ADEs) were the most common type, accounting for 19% of adverse events, followed by wound infections (14%) and technical complications (13%). Almost half the adverse events were associated with a surgical procedure. Among nonoperative events, 37% were ADEs, 15% were diagnostic mishaps, 14% were therapeutic mishaps, 13% were procedure-related, and 5% were falls.

ADEs have been studied more than any other category. Studies focusing specifically on ADEs have found that they appear to be much more common than was suggested by the Medical Practice Study, although most other studies use more inclusive criteria. Detection approaches in the research setting include chart review and the use of a computerized ADE monitor, a tool that explores the database and identifies signals that suggest an ADE may have occurred. Studies that use multiple approaches find more ADEs than does any individual approach, suggesting that the true underlying rate in the population is higher than would be identified by an individual approach. About 6–10% of patients admitted to U.S. hospitals experience an ADE.

Injuries caused by drugs are also common in the outpatient setting. One study found a rate of 21 ADEs per every 100 patients per year when patients were called to assess whether they had had a problem with one of their medications. The severity level was lower than in the inpatient setting, but approximately one-third of these ADEs were preventable.

Another area that appears to be very risky is the period immediately after a patient is discharged from the hospital. A recent study of patients hospitalized on a medical service found an adverse event rate of 19%; about a third of those events were preventable, and another third were ameliorable in that they could have been made less severe. ADEs were the single leading category.

PREVENTION STRATEGIES

Most of the work on prevention strategies for adverse events has targeted specific types of adverse events in the inpatient setting, with nosocomial infections and ADEs having received the most

attention. Nosocomial infection rates have been reduced greatly in intensive care settings, especially through the use of checklists. For ADEs, several strategies have been found to reduce the medication error rate, although it has been harder to demonstrate that they reduce the ADE rate, and studies with adequate power to demonstrate a clinically meaningful reduction have not been published.

Implementation of checklists to ensure that specific actions are carried out has had a major impact on rates of catheter-associated bloodstream infections and ventilator-associated pneumonia, two of the most serious complications occurring in intensive care. The concept is that there are several specific actions that can reduce the frequency of these issues; when these actions are all carried out for every patient, the result has been an extreme reduction in the frequency of the associated complication. Furthermore, these practices have been disseminated across wide areas, in particular in the state of Michigan.

Computerized physician order entry (CPOE) linked with clinical decision support has been found to reduce the serious medication error rate; serious medication errors are those which harm someone or have the potential to do so. In one study, CPOE, even with limited decision support, decreased the serious medication error rate by 55%. CPOE can prevent medication errors by suggesting a default dose, ensuring that all orders are complete (e.g., include a dose, route, and frequency), and checking orders for allergies, drug-drug interactions, and drug-laboratory issues. In addition, clinical decision support can suggest the right dose for a patient, tailoring it to the patients level of renal function and age. In one study, without decision support patients with renal insufficiency received the appropriate dose only one-third of the time, whereas that fraction increased to approximately two-thirds with decision support, and patients with renal insufficiency were discharged from the hospital one-half day earlier. As of 2009, only about 15% of U.S. hospitals had implemented CPOE, but many plan to do so and will receive major financial incentives for achieving this goal.

Another technology that can improve medication safety is bar coding linked with an electronic medication administration record. Bar coding can help ensure that the right patient gets the right medication at the right time. Electronic medication administration records can make it much easier to determine what medications a patient has received. Studies to assess the impact of bar coding on medication safety are under way, and the early results are promising. Another technology that can be used to improve the safety of medication administration is "smart pumps." These are pumps that can be instructed in which medication is being given and at what dose; if the nurse tries to administer too high a dose, he or she will receive a warning.

THE NATIONAL PICTURE AROUND SAFETY

Several organizations, including the National Quality Forum (NQF) and the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), have made recommendations about how to improve safety. In particular, the NQF has released recommendations to the countrys hospitals about what practices will most improve the safety of care, which all hospitals are expected to implement **(Table 10-1)**. Many of these practices arise frequently in routine care. One example is "readback," the practice of recording all verbal orders and immediately reading them back to the physician to verify the accuracy of what was heard. Another is to use only standard abbreviations and dose designations, since some abbreviations and dose designations are particularly prone to error; for example, 7U may be read as 70.

Table 10-1 Safe Practices for Better Health Care[a]

1. Create a health care culture of safety.
2. For designated high-risk, elective surgical procedures or other specified care, patients should be clearly informed of the likely reduced risk of an adverse outcome at treatment facilities that have demonstrated superior outcomes and should be referred to such facilities in accordance with the patients stated preference.
3. Specify an explicit protocol to be used to ensure an adequate level of nursing based on the institutions usual patient mix and the experience and training of its nursing staff.
4. All patients in general intensive care units (both adult and pediatric) should be managed by physicians who have specific training and certification in critical care medicine ("critical care certified").
5. Pharmacists should participate actively in the medication-use process, including, at a minimum, being available for consultation with prescribers on medication ordering, interpretation and review of medication orders; preparation of medications; dispensing of medications; and administration and monitoring of medications.
6. Verbal orders should be recorded whenever possible and immediately read back to the prescriber; i.e., a health care provider receiving a verbal order should read or repeat back the information that the prescriber conveys to verify the accuracy of what was heard.
7. Use only standardized abbreviations and dose designations.
8. Patient care summaries or other similar records should not be prepared from memory.
9. Ensure that care information, especially changes in orders and new diagnostic information, is

transmitted in a timely and clearly understandable form to all of the patients current health care providers who need that information to provide care.
10. Ask each patient or legal surrogate to recount what he or she has been told during the informed consent discussion.
11. Ensure that written documentation of the patients preference for life-sustaining treatments is displayed prominently displayed in his or her chart.
12. Implement a computerized prescriber order entry system.
13. Implement a standardized protocol to prevent the mislabeling of radiographs.
14. Implement standardized protocols to prevent the occurrence of wrong-site procedures or wrong-patient procedures.
15. Evaluate each patient undergoing elective surgery for risk of an acute ischemic cardiac event during surgery and provide prophylactic treatment of high-risk patients with beta blockers.
16. Evaluate each patient upon admission, and regularly thereafter, for the risk of developing pressure ulcers. This evaluation should be repeated at regular intervals during care. Clinically appropriate preventive methods should be implemented consequent to the evaluation.
17. Evaluate each patient upon admission, and regularly thereafter, for the risk of developing deep vein thrombosis (DVT)/venous thromboembolism (VTE). Utilize clinically appropriate methods to prevent DVT/VTE.
18. Utilize dedicated antithrombotic (anticoagulation) services that facilitate coordinated care management.
19. Upon admission, and regularly thereafter, evaluate each patient for the risk of aspiration.
20. Adhere to effective methods of preventing central venous catheter–associated bloodstream infections.
21. Evaluate each preoperative patient in light of his or her planned surgical procedure for the risk of surgical site infection and implement appropriate antibiotic prophylaxis and other preventive measures based on that evaluation.
22. Utilize validated protocols to evaluate patients who are at risk for contrast media-induced renal failure and utilize a clinically appropriate method for reducing risk of renal injury based on the patients kidney function evaluation.
23. Evaluate each patient upon admission, and regularly thereafter, for risk of malnutrition. Employ clinically appropriate strategies to prevent malnutrition.
24. Whenever a pneumatic tourniquet is used, evaluate the patient for the risk of an ischemic and/or thrombotic complication and utilize appropriate prophylactic measures.
25. Decontaminate hands with a hygienic hand rub or by washing with a disinfectant soap before and after direct contact with the patient or objects immediately around the patient.
26. Vaccinate health care workers against influenza to protect both them and patients from influenza.
27. Keep workspaces where medications are prepared clean, orderly, well lit, and free of clutter, distraction, and noise.
28. Standardize the methods for labeling, packaging, and storing medications.
29. Identify all "high alert" drugs (e.g., intravenous adrenergic agonists and antagonists, chemotherapy agents, anticoagulants and antithrombotics, concentrated parenteral electrolytes, general anesthetics, neuromuscular blockers, insulin and oral hypoglycemics, and narcotics and opiates).
30. Dispense medications in unit-dose or, when appropriate, unit-of-use form, whenever possible.

[a] These 30 practices are the recommendations from the National Quality Forum (NQF) for improving the safety of health care; the NQF believes they should be implemented universally in applicable care settings to reduce the risk of patient harm. The practices all have strong supporting evidence and are likely to have a significant benefit.

MEASUREMENT OF SAFETY

Measuring the safety of care is quite difficult and expensive, since adverse events are, fortunately, rare. Most hospitals rely on spontaneous reporting to identify errors and adverse events, but this approach has very low sensitivity, with only ~1 in 20 ADEs reported. There are promising research techniques that involve searching the electronic record for signals suggesting that an adverse event has occurred, which probably will be routine in the future but are not yet in wide use. Claims data have been used to identify the frequency of adverse events; this approach works much better for surgical care than for medical care and still requires additional validation. The net result is that except for a few specific types of events, such as falls and nosocomial infections, hospitals have little idea about the true frequency of safety issues.

Nonetheless, all providers have the responsibility to report problems with safety as they are identified. All hospitals have spontaneous reporting systems, and if providers report events as they occur, those events can be used as lessons for subsequent improvement.

CONCLUSIONS ABOUT SAFETY

It is now abundantly clear that the safety of health care can be improved substantially; as more areas are studied closely, more problems are identified. Compared with the outpatient setting, much more is known about the epidemiology of safety in the inpatient setting, and a number of effective strategies for improving safety have been identified and are being used increasingly. Some effective strategies are also available in the outpatient setting. Transitions appear to be especially risky. The solutions to improving care often will involve systematic techniques such as checklists and often will involve leveraging information technology, but they also will include many other domains, such as use of human factors techniques, team training, and building a culture of safety.

Quality in Health Care

Quality of care has remained somewhat elusive, although the tools for measuring it have increasingly improved. Selecting health care and measuring its quality is a complex process.

QUALITY THEORY

Donabedian has suggested that quality of care can be categorized by type of measurement into structure, process, and outcome. *Structure* refers to whether a particular characteristic is present, e.g., whether a hospital has a catheterization laboratory or whether a clinic uses an electronic health record. *Process* refers to the way care is delivered, and examples of process measures are whether a Pap smear was performed at the recommended interval or whether an aspirin was given to a patient with a suspected myocardial infarction. *Outcomes* refer to what actually happens, e.g., the mortality rate in myocardial infarction. It is important to note that good structure and process do not always result in good outcomes. For instance, a patient may present with a suspected myocardial infarction to an institution with a catheterization laboratory and receive recommended care, including aspirin, but still die because of the infarction.

Quality theory also suggests that overall quality will be improved more in the aggregate by raising the level of performance of all providers rather than finding a few poor performers and punishing them. This view suggests that systems changes are especially likely to be helpful in improving quality, since large numbers of providers may be affected simultaneously.

The theory of continuous quality improvement suggests that organizations should be evaluating the care they deliver on an ongoing basis and continually making small changes to improve their individual processes. This approach can be very powerful if embraced over time.

A number of specific tools have been developed to help improve process performance. One of the most important is the Plan-Do-Check-Act cycle **(Fig. 10-2)**. This approach can be used to perform what is called rapid cycle improvement for a process, e.g., the time for a patient with pneumonia to receive antibiotics after diagnosis. Often, specific statistical tools such as control charts are used in conjunction to determine whether progress is being made. Most medical care includes one or many processes, making this tool especially important for improvement.

Figure 10-2

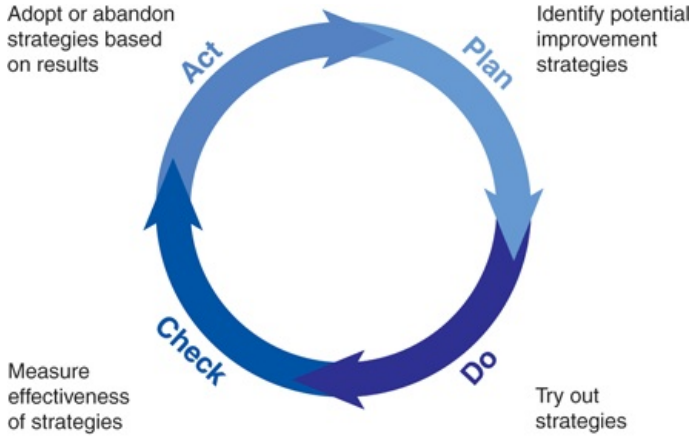

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Plan-Do-Check-Act (PDCA) cycle. The PDCA cycle approach can be used to improve a specific process rapidly. First, planning is performed, and several potential improvement strategies are identified. Next, these strategies are trialed in small "tests of change." "Checking" entails measuring whether they appeared to make a difference, and "act" refers to acting on the results.

FACTORS RELATING TO QUALITY

Many factors can decrease the level of quality, including stress to providers, high or low levels of production pressure, and poor systems, to name but a few examples. Stress can have an adverse effect on quality because it can lead providers to omit important steps, as can a high level of production pressure. Low levels of production pressure sometimes can result in worse quality, as providers may be bored or have little experience with a specific problem. Poor systems can have a tremendous impact on quality, and even extremely dedicated providers typically cannot achieve high levels of performance if they are operating within a poor system.

DATA ABOUT THE CURRENT STATE OF QUALITY

A recent RAND study has provided the most complete picture of quality of care delivered in the United States to date. The results were sobering. The authors found that across a wide range of quality parameters, patients in the United States received only 55% of recommended care overall; there was little variation by subtype, with scores of 54% for preventive care, 54% for acute care, and 56% for care of chronic conditions, leading the authors to conclude that the chances of getting high-quality care in the United States broadly were little better than those of winning a coin flip.

Work from the Dartmouth Atlas evaluating geographic variation in utilization and quality of care demonstrates that despite large variations in utilization, there is no positive correlation between the two variables at the regional level. An array of data demonstrate, however, that providers with larger volumes for specific conditions, especially for surgical conditions, do have better outcomes.

STRATEGIES FOR IMPROVING QUALITY AND PERFORMANCE

A number of specific strategies can be used to improve quality at the individual level, including rationing, education, feedback, incentives, and penalties. *Rationing* has been effective in some specific areas, such as persuading physicians to prescribe within a formulary, but it generally has been resisted. *Education* is effective in the short run and is necessary for changing opinions, but its effect decays fairly rapidly with time. *Feedback* on performance can be given either at the group or the individual level. Feedback is most effective if it is individualized and is given in close temporal proximity to the original events. *Incentives* can be effective, and many believe that this will be a key to improving quality, especially if pay-for-performance with sufficient incentives is broadly implemented (see below). *Penalties* produce provider resentment and rarely are used in health care.

Another set of strategies for improving quality involves changing the systems of care. An example would be introducing reminders about which specific actions need to be taken at a visit for a specific patient, a strategy that has been demonstrated to improve performance in certain situations, e.g., the delivery of preventive services. Another approach that has been effective is the development of "bundles" or groups of quality measures that can be implemented together with a high degree of fidelity. A number of hospitals have implemented a bundle for ventilator-associated pneumonia in the intensive care unit, which includes five measures, including, for example, ensuring that the head of the bed is elevated. The hospitals have found that they were able to improve performance substantially.

Perhaps the most pressing need is to improve the quality of care for chronic diseases. The Chronic Care Model has been developed by Wagner and colleagues **(Fig. 10-3)**; it suggests that a combination of strategies will be necessary, including self-management support, changes in delivery system design, decision support, and information systems, and that these strategies must be delivered by a practice team composed of several providers, not just a physician.

Figure 10-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The chronic care model. The chronic care model, which focuses on improving care for chronic diseases, suggests that delivery of high-quality care requires a range of strategies that must closely involve and engage the patient and, in addition, that team care is essential. *(From EH Wagner et al: Eff Clin Pract 1:2, 1998.)*

Recent evidence about the relative efficacy of strategies in reducing hemoglobin A_{1c} (HbA_{1c}) in outpatient diabetes care supports this general premise. It is especially notable that the outcome was HbA_{1c}, as it has generally been much more difficult to improve outcome measures than process measures (such as whether an HbA_{1c} was performed). In this meta-analysis, a variety of strategies were effective, but the most effective ones were the use of team changes and the use of a case manager. When cost-effectiveness is considered in addition, it appears likely that an amalgam of strategies will be needed. However, the more expensive strategies, such as the use of case managers, probably will be implemented widely only if pay-for-performance takes hold.

NATIONAL STATE OF QUALITY MEASUREMENT

In the inpatient setting, quality measurement is now being performed by a very large proportion of hospitals for several conditions, including myocardial infarction, congestive heart failure, pneumonia, and surgical infection prevention; 20 measures are included in all. This is the result of the Hospital Quality Initiative, which represents a collaboration among many entities, including the Hospital Quality Alliance, the JCAHO, the NQF, and the Agency for Healthcare Research and Quality, among others. The data are housed at the Center for Medicare and Medicaid Services, which publicly releases performance on the measures on a website called *Hospital Compare*. These data are reported voluntarily and are available for a very high proportion of the nations hospitals. Analyses demonstrate that there is substantial regional variation in quality and that there are important differences among hospitals. Analyses by the Joint Commission for very similar indicators demonstrate that performance on measures by hospitals did improve over time and that, as might be hoped, lower performers improved more than did higher performers.

PUBLIC REPORTING

Overall, public reporting of quality data is becoming increasingly common. There are now commercial websites that have quality-related data for most regions of the country that can be accessed for a fee. Similarly, national data for hospitals are available. The evidence to date is that patients have not used such data very much but that the data have had an important effect on provider and organization behavior. Instead, patients have relied on provider reputation to make choices. Part of the reason for this choice basis is that until very recently little information was available, and it was not necessarily presented in ways that were easy for patients to access. Many believe that as more information about quality becomes available, it will become increasingly central to patient choices about where to access care.

PAY-FOR-PERFORMANCE

Currently, providers in the United States get paid exactly the same amount for a specific service regardless of what quality of care is delivered. The theory of pay-for-performance suggests that if providers are paid more for higher-quality care, they will invest in strategies that enable them to deliver that care. The current key issues in the pay-for-performance debate relate to (1) how effective it is, (2)

what levels of incentives are needed, and (3) what perverse consequences are produced. The evidence about effectiveness is fairly limited, although a number of studies are ongoing. With respect to levels, most performance incentives around quality have accounted for merely 1–2% of total payment in this country to date, but in the United Kingdom, 40% of general practitioners salaries have recently been placed at risk based on performance across a wide array of parameters. This has been associated with large improvements in reported quality performance, although it is still unclear to what extent this represents better performance versus better reporting. The potential for perverse consequences exists with any incentive scheme. One problem is that if incentives are tied to outcomes, this introduces the incentive to transfer the sickest patients to other providers and systems. Another concern is that providers will pay too much attention to quality measures with incentives and ignore the rest of the quality picture. The validity of these concerns remains to be determined.

Conclusions

The safety and quality of care in the United States could be improved substantially. A number of interventions are available today that have been demonstrated to improve the safety of care and should be used more widely; others are undergoing evaluation or will be evaluated. Quality also could be dramatically better, and the science of quality improvement is increasingly mature. Implementation of pay-for-performance should make it much easier for organizations to justify investments in improving these parameters, including health information technology; however, many will also require changing the structure of care, e.g., moving to a more team-oriented approach, and ensuring that the patients are more involved in their own care. Measures of safety are still relatively immature and could be made much more robust; it would be particularly useful if organizations had measures they could use in routine operations to assess safety at a reasonable cost. Although the quality measures available are more robust than those for safety, they still cover a relatively small proportion of the entire domain of quality, and more need to be developed. The public and payers are demanding better information about safety and quality as well as better performance in these areas. The clear implication is that these domains will have to be addressed directly by providers.

FURTHER READINGS

Bates DW et al: Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA 280:1311, 1998

Brennan TA et al: Incidence of adverse events and negligence in hospitalized patients: Results from the Harvard Medical Practice Study I. N Engl J Med 324:370, 1991

Jha AK et al: Patient safety research: An overview of the global evidence. Qual Saf Health Care 19:42, 2010

McGlynn EA et al: The quality of health care delivered to adults in the United States. N Engl J Med 348:2635, 2003

Shojania KG et al: Effects of quality improvement strategies for type 2 diabetes on glycemic control: A meta-regression analysis. JAMA 296:427, 2006

Wagner EH et al: Improving chronic illness care: Translating evidence into action. Health Aff (Millwood) 20:64, 2001

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRIMARY CARE IN LOW- AND MIDDLE-INCOME COUNTRIES: INTRODUCTION

The twentieth century witnessed the rise of an unprecedented global health divide. Industrialized or high-income countries experienced rapid improvement in standards of living, nutrition, health, and health care. Meanwhile, in low- and middle-income countries with much less favorable conditions, health and health care progressed much more slowly. The scale of this divide is reflected in the current extremes of life expectancy at birth, with Japan at the high end (82 years) and Sierra Leone at the low end (32 years). This 50-year difference reflects the daunting range of health challenges faced by low- and middle-income countries. These nations are faced not only with a complex mixture of diseases (both infectious and chronic) and illness-promoting conditions but, more fundamentally, with the fragility of the foundations underlying good health (e.g., sufficient food, water, sanitation, and education) and of the systems necessary for universal access to good-quality health care. In the last decades of the twentieth century, the need to bridge this global health divide and establish health equity was increasingly recognized. The Declaration of Alma Ata in 1978 crystallized a vision of justice in health, regardless of income, gender, ethnicity, or education, and called for "health for all by the year 2000" through primary health care. While much progress has been made since the declaration, at the end of the first decade of the twenty-first century, much remains to be done to achieve global health equity.

This chapter looks first at the nature of the health challenges in low- and middle-income countries that underlie the health divide. It then outlines the values and principles of a primary health care approach with a focus on primary care services. Next, the chapter reviews the experience of low- and middle-income countries in addressing health challenges through primary care and a primary health care approach. Finally, the chapter identifies how current challenges and global context provide an agenda and opportunities for the renewal of primary health care and primary care.

PRIMARY CARE AND PRIMARY HEALTH CARE

The term *primary care* has been used in many different ways: to describe a level of care or setting of the health system, a set of treatment and prevention activities carried out by specific personnel, a set of attributes for the way care is delivered, or an approach to organizing health systems that is synonymous with the term *primary health care*. In 1996, the U.S. Institute of Medicine encompassed many of these different usages, defining primary care as "the provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of family and community."[1] We use this definition of *primary care* in this chapter. Primary care performs an essential function for health systems, providing the first point of contact when people seek health care, dealing with most problems, and referring patients onward to other services when necessary. As is increasingly evident in countries of all income levels, without strong primary care, health systems cannot function properly or address the health challenges of the communities they serve.

[1] Institute of Medicine. Primary Care: America's Health in a New Era (1996).

Primary care is only one part of a primary health care approach. The Declaration of Alma Ata, drafted in 1978 at the International Conference on Primary Health Care in Alma Ata (now Almaty in Kazakhstan), identified many features of primary care as being essential to achieving the goal of "health for all by the year 2000." However, it also identified the need to work across different sectors, address the social and economic factors that determine health, mobilize the participation of communities in health systems, and ensure the use and development of technology that was appropriate in terms of setting and cost. The declaration drew from the experiences of low- and middle-income countries in trying to improve the health of their people following independence. Commonly, these countries had built hospital-based systems similar to those in high-income countries. This effort had resulted in the development of high-technology services in urban areas while leaving the bulk of the population without access to health care unless they traveled great distances to these urban facilities. Furthermore, much of the population lacked access to basic public health measures. Primary health care efforts aimed to move care closer to where people lived, to ensure their involvement in decisions about their own health care, and to address key aspects of the physical and social environment essential to health such as water, sanitation, and education.

After the Declaration of Alma Ata, many countries implemented reforms of their health systems based on primary health care. Most progress involved strengthening of primary care services; unexpectedly, however, much of this progress was seen in high-income countries, most of which constructed systems that made primary care available at low or no cost to their entire populations and that delivered the bulk of services in primary care settings. This endeavor also saw the reinforcement of family medicine as a specialty to provide primary care services. Even in the United States (an obvious exception to this trend), it became clear that the populations of states with more primary care physicians and services were healthier than those with fewer such resources.

Progress was also made in many low- and middle-income countries. However, the target of "health for all by the year 2000" was missed by a large margin. The reasons were complex but partly entailed a general failure to implement all aspects of the primary health care approach, particularly work across sectors to address social and economic factors that affect health and provision of sufficient human and other resources to make possible the access to primary care attained in high-income countries. Furthermore, despite the consensus in Alma Ata in 1978, the global health community rapidly became fractured in its commitment to the far-reaching measures called for by the declaration. Economic recession tempered enthusiasm for primary health care, and momentum shifted to programs concentrating on a few priority measures such as immunization, oral rehydration, breast-feeding, and growth monitoring for child survival. Success with these initiatives supported the continued movement of health development efforts away from the comprehensive approach of primary health care and toward programs that targeted specific public health priorities. This approach was reinforced by the need to address the HIV/AIDSepidemic. By the 1990s, primary health care had fallen out of favor in many global-health policy circles, and low- and middle-income countries were being encouraged to reduce public sector spending on health and to focus on cost-effectiveness analysis to provide a package of health care measures thought to offer the greatest health benefits.

HEALTH CHALLENGES IN LOW- AND MIDDLE-INCOME COUNTRIES

Low- and middle-income countries, defined by a per capita gross national income of <$12,000 (U.S.) per person per year, account for >80% of the world's population. Average life expectancy in these countries lags far behind that in high-income countries: whereas the average life expectancy at birth in high-income countries is 74 years, it is only 68 years in middle-income countries and 58 years in low-income countries. This discrepancy has received growing attention over the past 40 years. Initially, the situation in poor countries was characterized primarily in terms of high fertility and high infant, child, and maternal mortality rates, with most deaths and illnesses attributable to infectious or tropical diseases among remote, largely rural populations. With growing adult (and especially elderly) populations and changing lifestyles linked to global forces of urbanization, a new set of health challenges characterized by chronic diseases, environmental overcrowding, and road traffic injuries has emerged rapidly **(Fig. e1-1)**. The majority of tobacco-related deaths globally now occur in low- and middle-income countries, and the risk of a child's dying from a road traffic injury in Africa is more than twice that in Europe. Hence, low- and middle-income countries in the twenty-first century face a full spectrum of health challenges —infectious, chronic, and injury-related—at much higher incidences and prevalences than are documented in high-income countries and with many fewer resources to address these challenges.

Figure e1-1

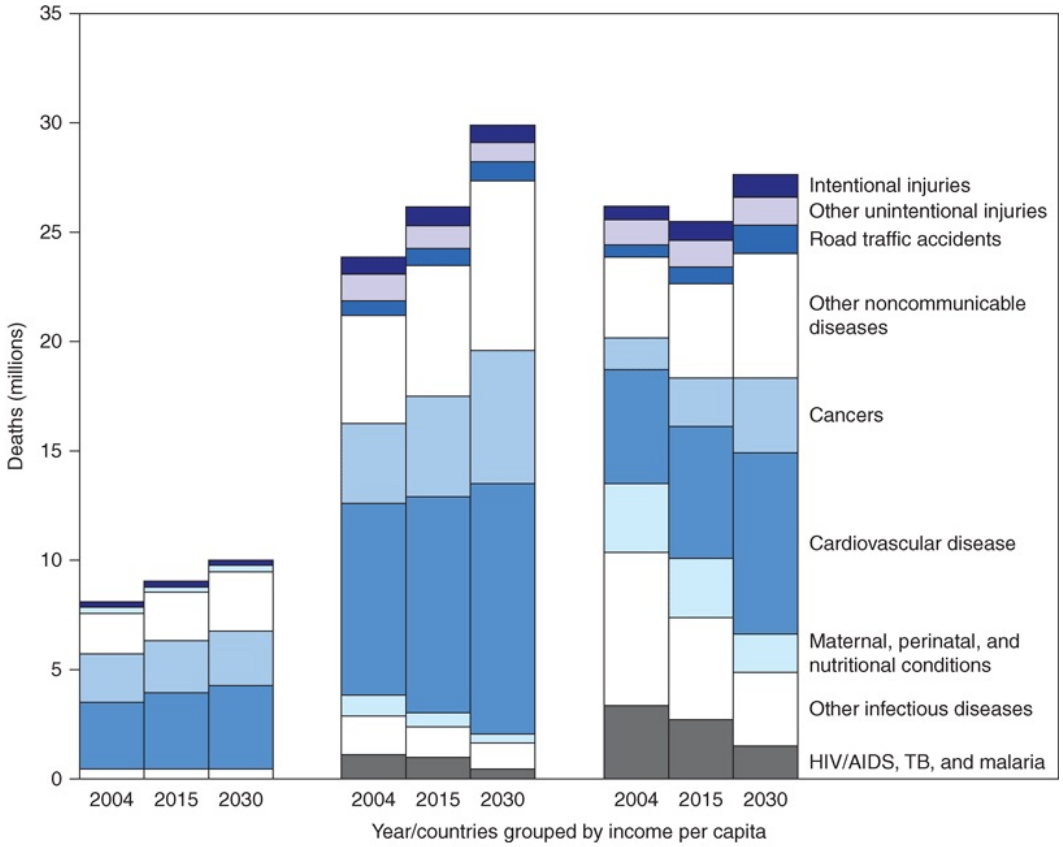

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Projections of disease burden to 2030 for high-, middle-, and low-income countries (*left*, *center*, and *right*, respectively). *(Source: World Health Organization, 2008b.)*

Addressing these challenges, however, does not mean simply waiting for economic growth. Analysis of the association between wealth and health across countries reveals that, for any given level of wealth, there is a substantial variation in life expectancy at birth that has persisted despite overall global progress in life expectancy during the past 30 years **(Fig. e1-2)**. Health status in low- and middle-income countries varies enormously. Nations such as Cuba and Costa Rica have life expectancies and childhood mortality rates similar to or even better than those in high-income countries; in contrast, countries in sub-Saharan Africa and the former Soviet bloc have experienced significant reverses in these health markers in the past 20 years.

Figure e1-2

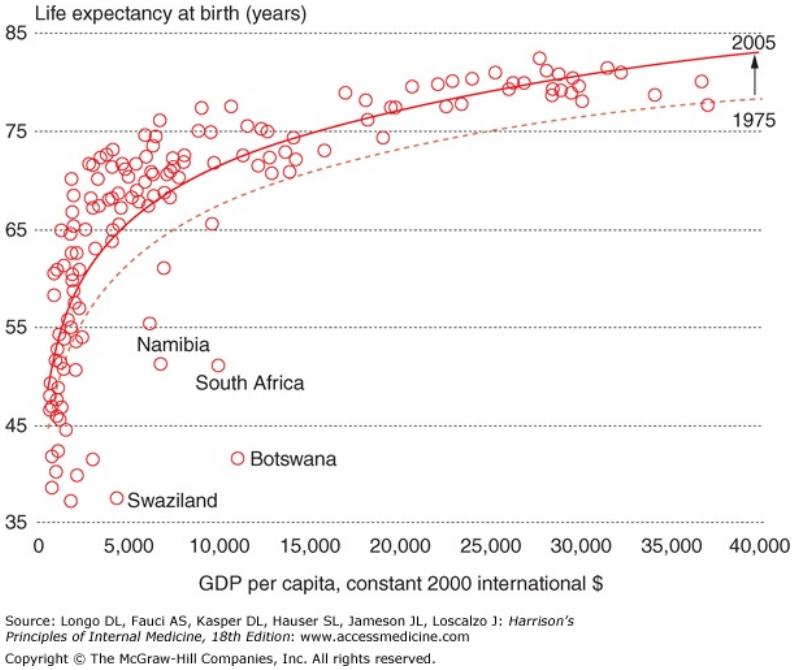

Gross domestic product (GDP) per capita and life expectancy at birth in 169 countries, 1975 and 2005. Only outlying countries are named. *(Source: World Health Organization, 2008a.)*

As Angus Deaton stated in the WIDER annual lecture on September 29, 2006, "People in poor countries are sick not primarily because they are poor but because of other social organizational failures including health delivery, which are not automatically ameliorated by higher income." This analysis concurs with classic studies of the array of societal factors explaining good health in poor settings such as Cuba and Kerala State in India. Analyses conducted over the past three decades indeed show that rapid health improvement is possible in very different contexts. That some countries continue to lag far behind can be understood through a comparison of regional differences in progress in terms of life expectancy over this period **(Fig. e1-3)**. While most regions have made impressive progress, sub-Saharan Africa and the former Soviet states have seen stagnation and even reversals.

Figure e1-3

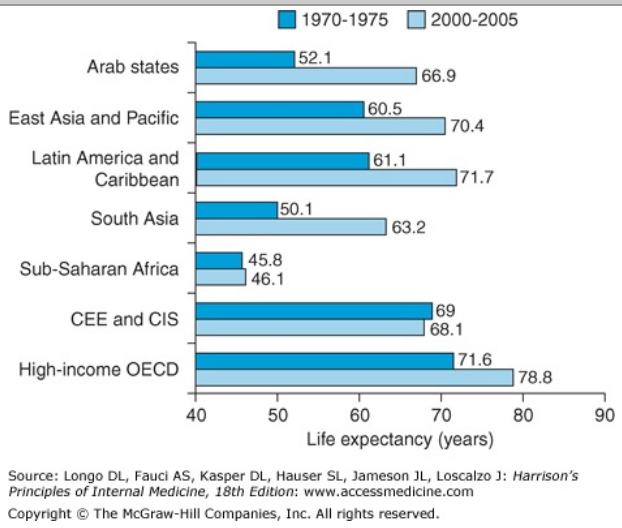

Regional trends in life expectancy. CEE and CIS, Central and Eastern Europe and the Commonwealth of Independent States; OECD, Organization for Economic Co-operation and Development. *(Source: Commission on Social Determinants of Health, 2008.)*

As average levels of health vary across regions and countries, so too do they vary within countries **(Fig. e1-4)**. Indeed, disparities within countries are often greater than those between high-income and low-income countries. For example, if low- and middle-income countries could reduce their overall childhood mortality rate to that of the richest one-fifth of their populations, global childhood mortality could be decreased by 40%. Disparities in health are mostly a result of social and economic factors such as daily living conditions, access to resources, and ability to participate in life-affecting decisions. In most countries, the health care sector actually tends to exacerbate health inequalities (the "inverse-care law"); because of neglect and discrimination, poor and marginalized communities are much less likely to benefit from public health services than those that are better off. Reforming health systems toward people-centered primary care provides an opportunity to reverse these negative trends.

Figure e1-4

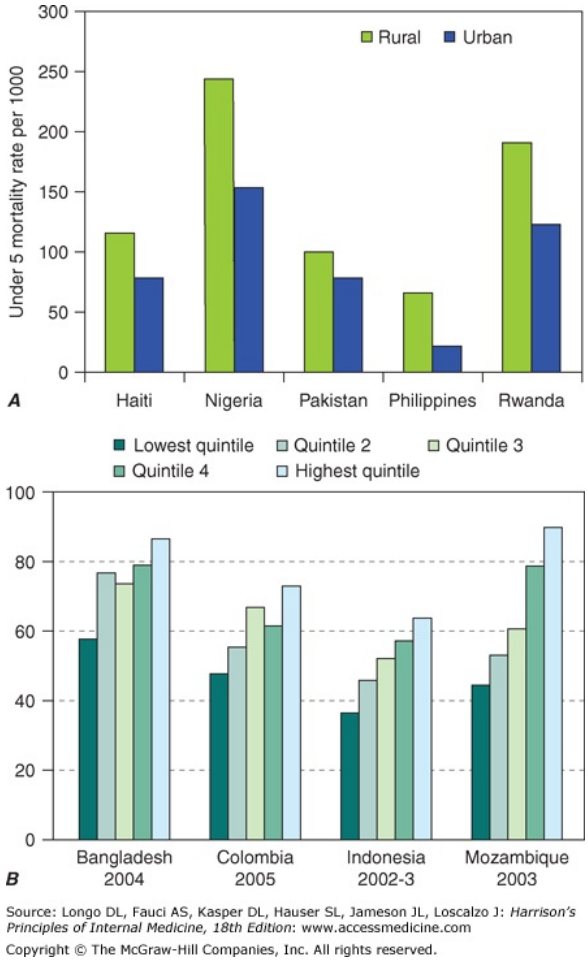

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Under-5 childhood mortality, by place of residence, in five countries. (Source: Data from the World Health Organization.) ***B.*** Full basic immunization coverage (%), by income group. *(Source: Data from the World Health Organization, 2008a.)*

Health services have failed to make their contribution to reducing these pervasive social inequalities by ensuring universal access to existing, scientifically validated, low-cost interventions such as insecticide-treated bed nets for malaria, taxes on cigarettes, short-course chemotherapy for tuberculosis, antibiotic treatment for pneumonia, dietary modification and secondary prevention measures for high blood pressure and high cholesterol levels, and water treatment and oral rehydration therapy for diarrhea. Despite decades of "essential packages" and "basic" health campaigns, the effective implementation of what is already known to work appears (deceptively) to be difficult.

Recent analyses have begun to focus on "the how" (as opposed to "the what") of health care delivery, exploring why health progress is slow and sluggish despite the abundant availability of proven interventions for health conditions in low- and middle-income countries. Three general categories of reasons are being identified: (1) shortfalls in performance of health systems; (2) stratifying social conditions; and (3) skews in science.

Shortfalls in Performance of Health Systems

Specific health problems often require the development of specific health intervention (e.g., tuberculosis requires short-course chemotherapy). However, the delivery of different interventions is often facilitated by a common set of resources or functions: money or financing, trained health workers, and facilities with reliable supplies fit for multiple purposes. Unfortunately, the current state of health systems in most low- and middle-income countries is largely dysfunctional.

In the large majority of low- and middle-income countries, the level of public financing for health is woefully insufficient: whereas high-income countries spend, on average, 7% of the gross domestic product on health, middle-income countries spend <4% and low-income countries <3%. External financing for health through various donor channels has grown significantly over time. While these funds for health are significant [~$20 billion (U.S.) in 2008 for low- and middle-income countries] and have been growing in the past decade, they represent <2% of total health expenditures in low- and middle-income countries and hence are neither a sufficient nor a long-term solution to chronic underfinancing. In Africa, 70% of health expenditures come from domestic sources. The predominant form of health care financing—charging patients at the point of service—is the least efficient and the most inequitable, tipping millions of households into poverty annually.

Health workers, who represent another critical resource, are often inadequately trained and supported in their work. Recent estimates indicate a shortage of >4 million health workers, constituting a crisis that is greatly exacerbated by the migration of health workers from low- and middle-income countries to high-income countries. Sub-Saharan Africa carries 24% of the global disease burden but has only 3% of the health workforce **(Fig. e1-5)**. The International Organization for Migration estimated in 2006 that there were more Ethiopian physicians practicing in Chicago than in Ethiopia itself.

Figure e1-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Global burden of disease and health workforce. *(Source: World Health Organization, 2006.)*

Critical diagnostics and drugs often do not reach patients in need because of supply chain failures. Moreover, facilities fail to provide safe care: new evidence suggests much higher rates of adverse events among hospitalized patients in low- and middle-income countries than in high-income countries. Weak government planning, regulatory, monitoring, and evaluation capacities are associated with rampant, unregulated commercialization of health services and chaotic fragmentation of these services as donors "push" their respective priority programs. With such fragile foundations, it is not surprising that low-cost, affordable, validated interventions are not reaching those who need them.

Stratifying Social Conditions

Health care delivery systems do not exist in a vacuum but rather are embedded in a complex of social and economic forces that often stratify opportunities for health unfairly. Most worrisome are the pervasive forces of social inequality that serve to marginalize populations with disproportionately large health needs (e.g., the urban poor; illiterate mothers). Why should a poor slum dweller with no income be expected to come up with the money for a bus fare needed to travel to a clinic to learn the results of a sputum test for tuberculosis? How can a mother living in a remote rural village and caring for an infant with febrile convulsions find the means to get her child to appropriate care? Shaky or nonexistent social security systems, dangerous work environments, isolated communities with little or no infrastructure, and systematic discrimination against minorities are among the myriad forces with which efforts for more equitable health care delivery must contend.

Skews in Science

While science has yielded enormous breakthroughs in health in high-income countries, with some spillover to low- and middle-income countries, many important health problems continue to affect primarily low- and middle-income countries whose research and development investments are woefully insufficient. The past decade has seen growing efforts to right this imbalance with research and development investment for new drugs, vaccines, and diagnostics that effectively cater to the specific health needs of populations in low- and middle-income countries. For example, the Medicines for Malaria Venture has revitalized a previously "dry" pipeline for new malaria drugs. This is but one of many such efforts, but much more needs to be done.

As discussed above, the primary constraint on better health in low- and middle-income countries is related less to the availability of health technologies and more to their effective delivery. Underlying these systems and social challenges to greater equity in health is a major bias regarding what constitutes legitimate "science" to improve health equity. The lion's share of health research financing is channeled toward the development of new technologies—drugs, vaccines, and diagnostics; in contrast, virtually no resources are directed toward research on how health care delivery systems can become more reliable and overcome adverse social conditions. The complexity of systems and social context is such that this issue of delivery requires an enormous investment in terms not only of money but also of scientific rigor, with the development of new research methods and measures and the attainment of greater legitimacy in the mainstream scientific establishment.

These common challenges to low- and middle-income countries partly explain the resurgence of interest in the primary health care approach. In some countries (mostly middle-income), significant progress has been made in expanding coverage by health systems based on primary care and even in improving indicators of population health. More countries are embarking on the creation of primary care services despite the challenges that exist, particularly in low-income countries. Even when these challenges are acknowledged, there are many reasons for optimism that low- and middle-income countries can accelerate progress in building primary care.

PRIMARY HEALTH CARE IN THE TWENTY-FIRST CENTURY

The past decade has seen a resurgence of interest in primary health care as a means of addressing the global health challenges of the new millennium. This interest has been driven by many of the same issues that led to the Declaration of Alma Ata: rapidly increasing disparities in health between and within countries, spiraling costs of health care at a time when many people lack quality care, dissatisfaction of communities with the care they are able to access, and failure to address changes in health threats, especially noncommunicable disease epidemics. These challenges require a comprehensive approach and strong health systems with effective primary care. Global health development agencies have recognized that sustaining gains in public health priorities such as HIV/AIDS requires not only robust health systems but also the tackling of social and economic factors related to disease incidence and progression. Weak health systems have proved a major obstacle to delivering new technologies such as antiretroviral therapy, to all who need them. Changing disease patterns have led to a demand for health systems that can treat people as individuals whether or not they present to a health facility with the public health "priority" (e.g., HIV/AIDS or tuberculosis) to which that facility is targeted. We discuss experiences in low- and middle-income countries in relation to primary care in greater detail below. First, we consider the features of primary health care and primary care as currently understood.

Revitalization of Primary Health Care

At the 2009 World Health Assembly [an annual meeting of all countries to discuss the work of the World Health

Organization (WHO)], a resolution was passed reaffirming the principles of the Declaration of Alma Ata and the need for national health systems to be based on primary health care. This resolution did not suggest that nothing had changed in the intervening 30 years since the declaration nor that its prescription did not need reframing in light of changing public health needs. The 2008 WHO World Health Report describes how a primary health care approach is necessary "now more than ever" to address global health priorities, especially in terms of disparities and new health challenges. It highlights four broad areas where reform is required **(Fig. e1-6)**, as discussed below. One of these areas—the need to organize health care so that it places the needs of people first—essentially relates to the necessity for strong primary care in health systems and what this requirement entails. The other three areas also relate to primary care. All four areas require action to move health systems in a direction that will reduce disparities and increase the satisfaction of those they serve. The World Health Report's recommendations present a vision of primary health care that is based on the principles of Alma Ata but that differs from many attempts to implement primary health care in the 1970s and 1980s.

Figure e1-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The four reforms of primary health care renewal. *(Source: World Health Organization, 2008a.)*

SERVICE DELIVERY REFORMS TO MAKE HEALTH SYSTEMS PEOPLE-CENTERED

Health systems have often been organized around the needs of those who provide health care services such as clinicians and policy makers. The result is a centralization of services or the provision of vertical programs that target single diseases. The principles of primary health care, including the development of primary care, reorient care around the needs of the people to whom services cater. This "people-centered" approach aims to provide health care that is both more effective and appropriate.

The increase in noncommunicable diseases in low- and middle-income countries offers a further stimulus for urgent reform of service delivery to improve chronic disease care. As discussed above, large numbers of people currently fail to receive relatively low-cost interventions that have reduced the incidence of these diseases in high-income countries. Delivery of these interventions requires health systems that can address multiple problems and manage people over a long period within their own communities, yet many low- and middle-income countries are only now starting to adapt and build primary care services that can address noncommunicable diseases and communicable diseases requiring chronic care. Even some countries (e.g., Iran) that have had significant success in reducing communicable diseases and improving child survival have been slow to adapt their health systems to rapidly accelerating noncommunicable disease epidemics.

People-centered care requires a safe, comprehensive, and integrated response to the needs of those presenting to health systems, with treatment at the first point of contact or referral to appropriate services. Because no discrete boundary separates people's needs for health promotion, curative interventions, and rehabilitation services across different diseases, primary care services must address all presenting problems in a unified way. Meeting people's needs also involves improved communication between patients and their clinicians, who must take the time to understand the impact of the patients' social context on the problems they present with. This enhanced understanding is made possible by improvements in the continuity of care so that responsibility transcends the limited time people spend in health care facilities. Primary care plays a vital role in navigating people through the health system; when people are referred elsewhere for services, primary care providers must monitor the resulting consultations and perform follow-up. All too often, people do not receive the benefit of complex interventions undertaken in hospitals because they lose contact with the health care system once discharged. Comprehensiveness and continuity of care are best achieved by ensuring that people have an ongoing personal relationship with a care team.

UNIVERSAL COVERAGE REFORMS TO IMPROVE HEALTH EQUITY

Despite progress in many countries, most people in the world can receive health care services only if they can pay at the point of service. Disparities in health are caused not only by a lack of access to necessary health services but also by the impact of expenditure on health. More than 100 million people are driven into poverty each year by health care costs. Universal coverage is therefore a major priority in low- and middle-income countries. Increasing coverage of health services can be considered in terms of three axes: the proportion of the population covered, the range of services underwritten, and the percentage of costs paid. Moving toward universal coverage requires ensuring the availability of health care services to all, eliminating barriers to access, and organizing pooled financing mechanisms such as taxation or

insurance, to remove user fees at the point of service. It also requires measures beyond financing, including expansion of health services in poorly served areas, improvement in the quality of services provided to marginalized communities, and increased coverage of other social services that significantly affect health (e.g., education).

PUBLIC POLICY REFORMS TO PROMOTE AND PROTECT THE HEALTH OF COMMUNITIES

Public policies in sectors other than health care are essential to reduce disparities in health and to make progress toward global public health targets. The 2008 final report of the WHO Commission on Social Determinants of Health provides an exhaustive review of the intersectoral policies required to address health inequities at the local, national, and global levels. Advances against major challenges such as HIV/AIDS, tuberculosis, emerging pandemics, cardiovascular disease, cancers, and injuries require effective collaboration with sectors such as transport, housing, labor, agriculture, urban planning, trade, and energy. While tobacco control provides a striking example of what is possible if different sectors work together toward health goals, the lack of implementation of many evidence-based tobacco control measures in most countries just as clearly illustrates the difficulties encountered in such intersectoral work and the unrealized potential of public policies to improve health. At the local level, primary care services can help enact health-promoting public policies in other sectors.

LEADERSHIP REFORMS TO MAKE HEALTH AUTHORITIES MORE RELIABLE

The Declaration of Alma Ata emphasized the importance of participation by people in their own health care. In fact, participation is important at all levels of decision making. Contemporary health challenges require new models of leadership that acknowledge the role of government in reducing disparities in health but that also recognize the many types of organizations that provide health care services. Governments need to guide and negotiate among these different groups, including nongovernmental organizations (NGOs) and the private sector, and to provide strong regulation where necessary. This difficult task requires a massive reinvestment in leadership and governance capacity, especially if action by different sectors is to be effectively implemented. Moreover, disadvantaged groups must be able to voice their health needs in a way that actively influences decision making.

Experiences with Primary Care in Low- and Middle-Income Countries

Aspects of the primary health care approach described above, with an emphasis on primary care services, have been implemented to varying degrees in many low- and middle-income countries over the past half-century. As discussed above, some of these experiences inspired and informed the Declaration of Alma Ata, which itself led many more countries to attempt to implement primary health care. This section describes the experiences of a selection of low- and middle-income countries in improving primary care services that have enhanced the health of their populations.

Before Alma Ata, few countries had attempted to develop primary health care on a national level. Rather, most focused on expanding primary care services to specific communities (often rural villages), making use of community volunteers to compensate for the absence of facility-based care. In contrast, in the post–World War II period, China invested in primary care on a national scale, and life expectancy doubled within roughly 20 years. The Chinese expansion of primary care services included a massive investment in infrastructure for public health (e.g., water and sanitation systems) linked to innovative use of community health workers. These "barefoot doctors" lived in and expanded care to rural villages. They received a basic level of training that enabled them to provide immunizations, maternal care, and basic medical interventions, including the use of antibiotics. Through the work of the barefoot doctors, China brought low-cost universal basic health care coverage to its entire population, most of which had previously had no access to these services.

In 1982, the Rockefeller Foundation convened a conference to review the experiences of China along with those of Costa Rica, Sri Lanka, and the state of Kerala in India. In all of these locations, good health care at low cost appeared to have been achieved. Despite lower levels of economic development and health spending, all of these jurisdictions, along with Cuba, had health indicators approaching—or in some cases exceeding—those of developed countries. Analysis of these experiences revealed a common emphasis on primary care services, with expansion of care to the entire population free of charge or at low cost, combined with community participation in decision making about health services and coordinated work in different sectors (especially education) toward health goals. During the three decades since the Rockefeller meeting, some of these countries have built on this progress, while others have experienced setbacks. Recent experiences in developing primary care services show that the same combination of features is necessary for success. For example, Brazil—a large country with a dispersed population—has made major strides in increasing the availability of health care in the past 20 years. In the past decade, the Brazilian Family Health Program has expanded progressively across the country, with almost all areas now covered. This program provides communities with free access to primary care teams made up of primary care physicians, community health workers, nurses, dentists, obstetricians, and pediatricians. These teams have responsibility for the health of people in a specified geographic area—not only those who access health clinics. Moreover, individual community health workers are responsible for a named list of people within the area covered by the primary care team. Problems with access to health care persist in Brazil, especially in isolated areas and urban slums. However, solid evidence indicates that the Family Health Program has already contributed to impressive gains in population health, particularly in terms of childhood mortality and health inequities. In fact, this program has already had an especially marked impact on childhood mortality reduction in less-developed areas **(Fig. e1-7)**.

Figure e1-7

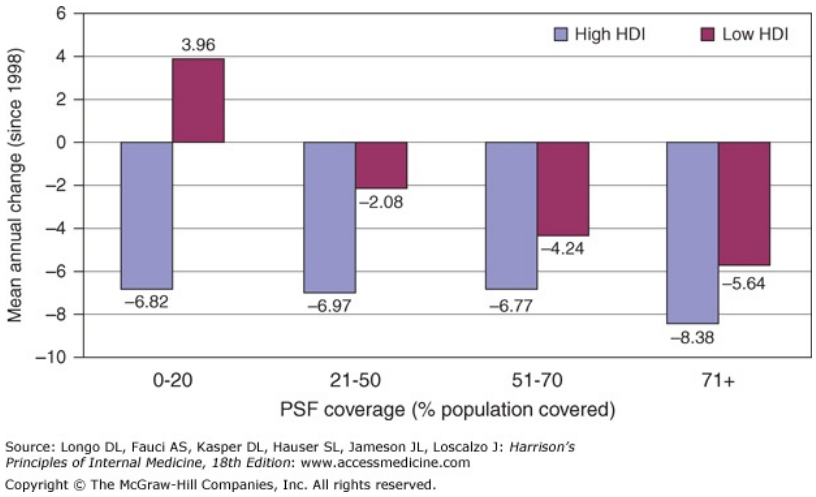

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Improvements in childhood mortality following the Family Health Program in Brazil. HDI, Human Development Index; PSF, Program Saúde da Família (Family Health Program). *(Source: Ministry of Health, Brazil.)*

Chile has also built on its existing primary care services in the past decade, aiming to improve the quality of care and the extent of coverage in remote areas, above all for disadvantaged populations. This effort has been made in concert with measures aimed at reducing social inequalities and fostering development, including social welfare benefits for families and disadvantaged groups and increased access to early-childhood educational facilities. As in Brazil, these steps have improved maternal and child health and have reduced health inequities. In addition to directly enhancing primary care services, Brazil and Chile have instituted measures to increase both the accountability of health providers and the participation of communities in decision making. In Brazil, national and regional health assemblies with high levels of public participation are integral parts of the health policy–making process. Chile has instituted a patient's charter that explicitly specifies the rights of patients in terms of the range of services to which they are entitled.

Other countries that have made recent progress with primary health care include Bangladesh, one of the poorest countries in the world. Since achieving its independence from Pakistan in 1971, Bangladesh has seen a dramatic increase in life expectancy, and childhood mortality rates are now lower than those in neighboring nations such as India and Pakistan. The expansion of access to primary health care services has played a major role in these achievements. This progress has been spearheaded by a vibrant NGO community that has focused its attention on improving the lives and livelihoods of poor women and their families through innovative and integrated microcredit, education, and primary care programs.

The above examples, along with others from the past 30 years in countries such as Thailand, Malaysia, Portugal, and Oman, illustrate how the implementation of a primary health care approach, with a greater emphasis on primary care, has led to better access to health care services—a trend that has not been seen in many other low- and middle-income countries. This trend, in turn, has contributed to improvements in population health and reductions in health inequities. However, as these nations have progressed, other countries have shown how previous gains with primary care can easily be eroded. In sub-Saharan Africa, undermining of primary care services has contributed to catastrophic reverses in health outcomes catalyzed by the HIV/AIDSepidemic. Countries such as Botswana and Zimbabwe implemented primary health care strategies in the 1980s, increasing access to care and making impressive gains in child health. Both countries have since been severely affected by HIV/AIDS, with pronounced decreases in life expectancy. However, Zimbabwe has also seen political turmoil, a decline of health and other social services, and the flight of health personnel, whereas Botswana has maintained primary care services to a greater extent and has managed to organize widespread access to antiretroviral therapy for people living with HIV/AIDS. Zimbabwe's health situation has therefore become more desperate than that in Botswana.

China provides a particularly striking example of how changes in health policy relevant to the organization of health systems **(Fig. e1-8)** can have rapid, far-reaching consequences for population health. Even as the 1982 Rockefeller conference was celebrating China's achievements in primary care, its health system was unraveling. The decision to open up the economy in the early 1980s led to rapid privatization of the health sector and the breakdown of universal health coverage. As a result, by the end of the 1980s, most people, especially the poorer segments of the population, were paying directly out of pocket for health care, and almost no Chinese had insurance—a dramatic transformation. The "barefoot doctor" schemes collapsed, and the population either turned to care paid for at hospitals or simply became unable to access care. This undermining of access to primary care services in the Chinese system and the resulting increase in impoverishment due to illness contributed to the stagnation of progress in health in China at the same time that incomes in that country increased at an unprecedented rate. Reversals in primary care have meant that China now increasingly faces health care issues similar to those faced by India. In both countries, rapid economic growth has been linked to lifestyle changes and noncommunicable disease epidemics. The health care systems of the two nations share two negative features that are common when primary care is weak: a disproportionate focus on specialty services provided in hospitals and unregulated commercialization of health services. China and India have both seen expansion of private hospital services that cater to middle-class and urban populations who can afford care; at the same time, hundreds of millions of people in rural areas now struggle to access the most basic services. Even in the former groups, a lack of primary care services has been associated with late presentation with illness and with insufficient investment in primary prevention approaches. This neglect of prevention poses a risk of large-scale epidemics of cardiovascular disease, which could endanger continued economic growth. In addition, the health systems of both countries now depend for the majority of their funding on out-of-pocket payments by people when they use services. Thus substantial proportions of the population must sacrifice other essential goods as a result of health expenditure or are even driven into poverty by this cost. The commercial nature of health services with inadequate or no regulation has also led to the proliferation of charlatan providers, inappropriate care, and pressure for people to pay for expensive and sometimes unnecessary care. Commercial providers have limited incentives to use interventions (including public health measures) that cannot be charged for or that are limited to the person paying.

Figure e1-8

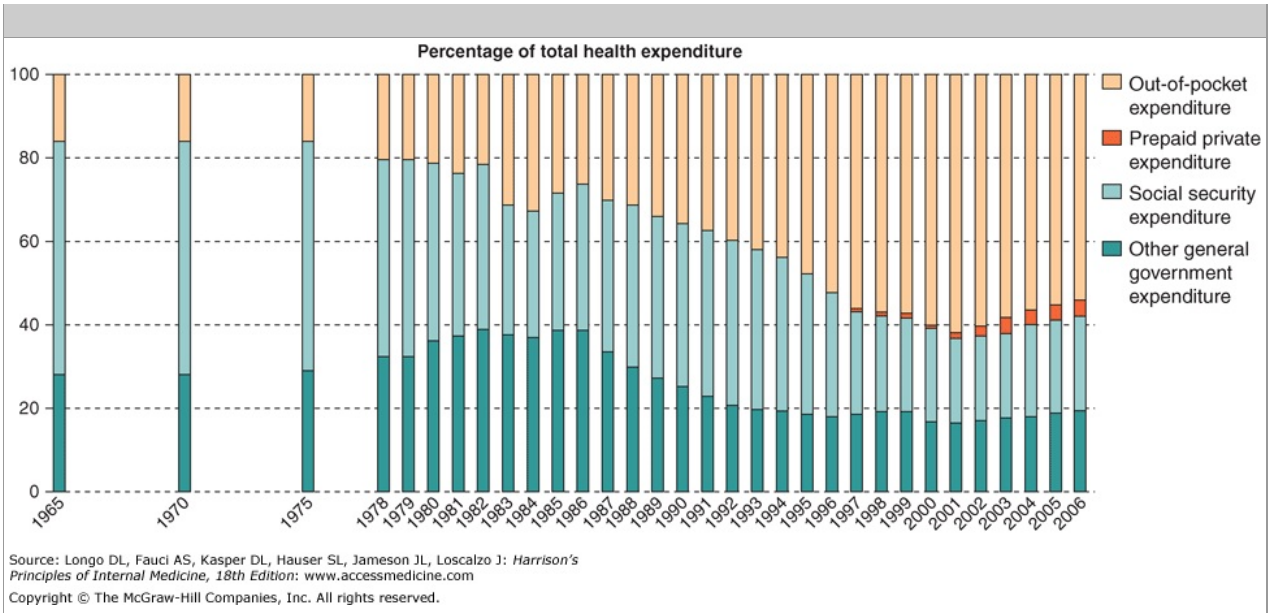

Changes in source of health expenditure in China over the past 40 years. *(Source: World Health Organization, 2008a.)*

Faced with these problems, China and India have recently implemented measures to strengthen primary health care. China has increased government funding of health care, has taken steps toward restoring health insurance, and has enacted a target of universal access to primary care services. India has similarly mobilized funding to greatly expand primary care services in rural areas and is now duplicating this process in urban settings. Both countries are increasingly using public resources from their growing economies to fund primary care services. These encouraging trends are illustrative of new opportunities to implement a primary health care approach and strengthen primary care services in low- and middle-income countries.

Opportunities to Build Primary Care in Low- and Middle-Income Countries

Global public health targets will not be met unless health systems are significantly strengthened. More money is currently being spent on health than ever before. In 2005, global health spending totaled $5.1 trillion (U.S.)—double the amount spent a decade earlier. Although most expenditure occurs in high-income countries, spending in many emerging middle-income countries has rapidly accelerated, as has the allocation of monies for this purpose by both governments in and donors to low-income countries. These twin trends—greater emphasis on building health systems based on primary care and allotment of more money for health care—provide opportunities to address many of the challenges discussed above in low- and middle-income countries.

Accelerating progress requires a better understanding of how global health initiatives can more effectively facilitate the development of primary care in low-income countries. A recent review by the WHO Maximizing Positive Synergies Collaborative Group looked at programs funded by the Global Fund to Fight AIDS, Tuberculosis and Malaria; the Global Alliance for Vaccines and Immunisation (GAVI); the U.S. President's Emergency Plan for AIDS Relief (PEPFAR); and the World Bank (on HIV/AIDS). This group found that global health initiatives had improved access to and quality of the targeted health services and had led to better information systems and more adequate financing. The review also identified the need for better alignment of global health initiatives with other national health priorities and systematic exploitation of potential synergies. If global health initiatives implement programs that work in tandem with other components of national health systems without undermining staffing and procurement of supplies, they have the potential to contribute substantially to the capacity of health systems to provide comprehensive primary care services.

Global health initiatives continue to draw increased funding. In 2009, for example, U.S. President Barack Obama announced increasing development assistance from the United States for global health, earmarking $63 billion over the period 2009–2014 for a Global Health Initiative. New funding is also promised through a range of other initiatives focusing particularly on maternal and child health in low-income countries. The general trend is to coordinate this funding to reduce fragmentation of national health systems and to concentrate more on strengthening these systems. Comprehensive primary care in low-income countries must inevitably deal with the rapid emergence of chronic diseases and the growing prominence of injury-related health problems; thus, international health development assistance must become more responsive to these needs.

Beyond the new streams of funding for health services, other opportunities exist. Increased social participation in health systems can help build primary care services. In many countries, political pressure from community advocates for more holistic and accountable care as well as entrepreneurial initiatives to scale up community-based services through NGOs have accelerated progress in primary care without major increases in funding. Participation of the population in the provision of health care services and in relevant decision making often drives services to cater to people's needs as a whole rather than to narrow public health priorities.

Participation and innovation can help address critical issues related to the health workforce in low- and middle-income countries by establishing effective people-centered primary care services. Many primary care services do not need to be delivered by a physician or a nurse. Multidisciplinary teams can include paid community workers who have access to a physician if necessary but who can provide a range of health services on their own. In Ethiopia, more than 30,000 community health workers have been trained and deployed to improve access to primary care services, and there is increasing evidence that this measure is contributing to better health outcomes. In India, more than 600,000 community health advocates have been recruited as part of expanded rural primary care services. After the Declaration of Alma Ata, experiences with community health workers were mixed, with particular problems about levels of training and lack of payment. Current endeavors are not immune from these concerns. However, with access to physician support and the deployment of teams, some of these concerns may be addressed. Growing evidence from many countries indicates that shifting appropriate tasks to primary care workers who have had shorter, less expensive training than physicians will be

essential to address the human resources crisis.

Finally, recent improvements in information and communication technologies, particularly mobile phone and Internet systems, have created the potential to systematically implement e-health, telemedicine, and improved health data initiatives in low- and middle-income countries. These developments raise the tantalizing possibility that health systems in these countries, which have long lagged behind those in high-income countries but are less encumbered by legacy systems that have proved hard to modernize in many settings, could leapfrog their wealthier counterparts in exploiting these technologies. Although the challenges posed by poor or absent infrastructure and investment in many low- and middle-income countries cannot be underestimated and will need to be addressed to make this possibility a reality, the rapid rollout of mobile networks and their use for health and other social services in many low-income countries where access to fixed telephone lines was previously very limited offer great promise in building primary care services in low- and middle-income countries.

CONCLUSION

As concern continues to mount about glaring inequities in global health, there is a growing commitment to redress these egregious shortfalls, as exemplified by global mobilization around the United Nations' Millennium Development Goals. This commitment begins first and foremost with a clear vision of the fundamental importance of health in all countries, regardless of income. The values of health and health equity are shared across all borders, and primary health care provides a framework for their effective translation across all contexts.

The translation of these fundamental values has its roots in four types of reforms that reflect the distinct but interlinked challenges of (re-)orienting a society's resources on the basis of its citizens' health needs: (1) organizing health care services around the needs of people and communities; (2) harnessing services and sectors beyond health care to promote and protect health more effectively; (3) establishing sustainable and equitable financing mechanisms for universal coverage; and (4) investing in effective leadership of the whole of society. This common primary health care agenda highlights the striking similarity, despite enormous differences in context, in the nature and direction of the reforms that national health systems must undertake to promote greater equity in health. This shared agenda is complemented by the growing reality of global health interconnectedness due, for example, to shared microbial threats, bridging of ethnolinguistic diversity, flows in migrant health workers, and mobilization of global funds to support the neediest populations. Embracing solidarity in global health, while strengthening health systems using a primary health care approach, is fundamental to sustained progress in global health.

FURTHER READINGS

Abegunde DO et al: The burden and costs of chronic diseases in low-income and middle-income countries. Lancet 370:1929, 2007[PMID: 18063029]

Aquino R et al: Impact of the family health program on infant mortality in Brazilian municipalities. Am J Public Health 99:87, 2009[PMID: 19008516]

Commission on Social Determinants of Health: *Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health: Commission on Social Determinants of Health Final Report*. Geneva, World Health Organization, 2008

Kruk ME et al: The contribution of primary care to health and health systems in low- and middle-income countries: A critical review of major primary care initiatives. Soc Sci Med 70:904, 2010[PMID: 20089341]

Macinko J et al: The impact of primary healthcare on population health in low- and middle-income countries. J Ambul Care Manage 32:150, 2009[PMID: 19305227]

Rasanathan K et al: Primary health care and the social determinants of health: Essential and complementary approaches for reducing inequities in health. J Epidemiol Community Health (epub ahead of print, May 27, 2010)

Samb B et al: Prevention and management of chronic disease: A litmus test for health-systems strengthening in low-income and middle-income countries. Lancet 376:1785, 2010[PMID: 21074253]

Smillie I: *Freedom from Want: The Remarkable Success Story of BRAC, the Global Grassroots Organization That's Winning the Fight Against Poverty*. Sterling, VA, Kumarian Press, 2009

Starfield B et al: Contribution of primary care to health systems and health. Milbank Q 83:457, 2005[PMID: 16202000]

World Health Organization: *Health Systems Financing: The Path to Universal Coverage.* World Health Report 2010. Geneva, World Health Organization, 2010

———: *Primary Health Care: Now More Than Ever* . World Health Report 2008. Geneva, World Health Organization, 2008a

———: *The Global Burden of Disease 2004 Update* . Geneva, World Health Organization, 2008b

———: *Working Together for Health* . Geneva, World Health Organization, 2006

World Health Organization Maximizing Positive Synergies Collaborative Group: An assessment of interactions between global health initiatives and country health systems. Lancet 373:2137, 2009

World Health Organization Regional Office for the Western Pacific: *People-Centred Health Care: A Policy Framework* . Geneva, World Health Organization, 2007

Yip W, Mahal A: The health care systems of China and India: Performance and future challenges. Health Aff (Millwood) 27:921, 2008[PMID: 18607024]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BACKGROUND

Medicine, once the domain of solitary generalists and their nurse assistants, now engages scores of specialists and allied professionals—radiation physicists, cytologists, nurse practitioners, psychiatric social workers, dental hygienists, and many more—who wield tools of unprecedented ability to extend life and sustain its quality. This evolution of the health care system has been achieved in part by a formidable enterprise of critical observation and formal investigation that disproves some accepted practices and stimulates the emergence of new approaches. One need only peruse the serial editions of this textbook to comprehend the scope of these changes.

Other factors also have affected evolutionary changes in medicine. The U.S. health care system has always been pluralistic, including many practices that are outside mainstream medicine. The public's expectations of health and the nature of the health care system have been altered by unprecedented access to sources of information, goods, and services; the disposable income to afford them; and a patchwork quilt of regulations and laws that constrain medical practice on the one hand and facilitate increased choice in health care on the other. Immigration and related demographic changes have created diverse communities that value their own health traditions. The emergence of complementary and alternative health practices and the approach called integrative medicine are manifestations of these changes in health care.

DEFINITIONS

Complementary and alternative medicine (CAM) refers to a group of diverse medical and health care systems, practices, and products that are not considered part of conventional or allopathic medicine or that have historic origins outside mainstream medicine. Most of these practices are used together with conventional therapies and therefore have been called *complementary* to distinguish them from *alternative* practices, which are those used instead of standard care. Use of dietary supplements, mind-body practices such as hypnosis, and care from a traditional healer all fall under the umbrella of CAM. Although some CAM practices are directed by an alternative health care provider such as a chiropractor, acupuncturist, or naturopathic practitioner, much of CAM is undertaken as "self-care" and paid for out of pocket. CAM does not encompass practices that have not been translated fully from the laboratory to the clinic or practices that have been well studied and disproved but still have some public appeal. Rather, CAM entails approaches with surprising pervasiveness, many of which can claim at least some evidentiary support. Until a few years ago, CAM also could be defined as practices that are neither widely taught in medical schools nor reimbursed, but this definition is no longer useful, since medical students increasingly seek and receive some instruction about CAM and some CAM practices are reimbursed by third-party payers. Definitions of common CAM practices are provided in **Table e2-1**.

Table e2-1 Terminology of Complementary and Alternative Medical Practices

Mind-Body and Manipulative Practices	
Acupuncture and acupressure	A family of procedures involving stimulation of defined anatomic points, a component of the major Asian medical traditions. Most common application involves the insertion and manipulation of thin metallic needles
Alexander technique	A movement therapy that uses guidance and education to improve posture, movement, and efficient use of muscles for improvement of overall body functioning
Guided imagery	The use of relaxation techniques followed by the visualization of images, usually calm and peaceful in nature, to invoke specific images to alter neurologic function or physiologic states
Hypnosis	The induction of an altered state of consciousness characterized by increased responsiveness to suggestion
Massage	Manual therapies that manipulate muscle and connective tissues to promote muscle relaxation, healing, and sense of well-being
Meditation	A group of practices, largely based in Eastern spiritual traditions, intended to focus or control attention and obtain greater awareness of the present moment, or mindfulness
Reflexology	Manual stimulation of points on hands or feet that are believed to affect organ

[illegible]	function
Rolfing/structural integration	A manual therapy that attempts to realign the body by deep tissue manipulation of fascia
Spinal manipulation	A range of manual techniques, employed by chiropractors and osteopaths, for adjustments of the spine to affect neuromuscular function and other health outcomes
Tai chi	A mind-body practice originating in China that involves slow, gentle movements and sometimes is described as "moving meditation"
Therapeutic touch	Secular version of the laying on of hands, described as "healing meditation"
Yoga	An exercise practice, originally east Indian, that combines breathing exercises, physical postures, and meditation
Traditional Medical Systems	
Ayurvedic medicine	The major East Indian traditional medicine system. Treatment includes meditation, diet, exercise, herbs, and elimination regimens (using emetics and diarrheals)
Curanderismo	A spiritual healing tradition common in Latin American communities that uses ritual cleansing, herbs, and incantations
Native American medicine	Diverse traditional systems that incorporate chanting, shaman healing ceremonies, herbs, laying on of hands, and smudging (ritual cleansing with smoke from sacred plants)
Siddha medicine	An East Indian medical system (prevalent among Tamil-speaking people)
Tibetan medicine	A medical system that uses diagnosis by pulse and urine examination; therapies include herbs, diet, and massage
Traditional Chinese medicine	A medical system that uses acupuncture, herbal mixtures, massage, exercise, and diet
Unani medicine	An East Indian medical system, derived from Persian medicine, practiced primarily in the Muslim community; also called "hikmat"
"Modern" Medical Systems	
Anthroposophic medicine	A spiritually based system of medicine that incorporates herbs, homeopathy, diet, and a movement therapy called eurythmy
Chiropractic	Chiropractic care involves the adjustment of the spine and joints to alleviate pain and improve general health; primarily used to treat back problems, musculoskeletal complaints, and headaches
Homeopathy	A medical system with origins in Germany that is based on a core belief in the theory of "like cures like"—compounds that produce certain syndromes, if administered in very diluted solutions, will be curative
Naturopathy	A clinical discipline that emphasizes a holistic approach to the patient, herbal medications, diet, and exercise. Practitioners have degrees as doctors of naturopathy
Osteopathy	A clinical discipline, now incorporated into mainstream medicine, that historically emphasized spinal manipulative techniques to relieve pain, restore function, and promote overall health

In the last decade, the term *integrative medicin*e has entered this dialogue.

Integrative medicine refers to a style of practice that places strong emphasis on a holistic approach to patient care, focusing on preventive strategies for maintenance of health and reduced use of technology. Physicians advocating this approach generally include selected CAM practices, particularly mind-body practices and dietary supplements, in the care they offer patients, and some have established practice settings that include CAM practitioners. Although this approach appears to be attractive to many patients, the weaknesses in the evidence base for some of the interventions offered in integrative practices continue to attract substantial concern and controversy.

PATTERNS OF USE

The first large survey of CAM use by Eisenberg and associates in 1993 surprised the medical community by showing that more than 30% of Americans use CAM approaches. Many studies since that time have extended those conclusions. Subsequently, the National Health Interview Survey (NHIS), a large national survey conducted by the National Center for Health Statistics, a component of the Centers for Disease Control and Prevention, has addressed the use of CAM and largely confirmed those results. The NHIS is a household survey of many kinds of health practices in the civilian population; it uses methods that create a nationally representative sample and has a sample size large enough to permit valid estimates about some subgroups. In 2002 and again in 2007, the survey included a set of questions that addressed CAM use. Information was obtained from 31,000 adults in

2002 and 23,300 adults and 9400 children in 2007. In both surveys, approximately 40% of adults were using some form of CAM. In the 2007 study, 38% of adults and 12% of children had used one or more modalities, with nonvitamin, nonmineral dietary supplements; relaxation techniques and meditation; chiropractic; and massage being the most prevalent. Over 1% reported having received acupuncture treatment. Americans are willing to pay for these services; the estimated out-of pocket expenditure for CAM in 2007 was $34 billion, representing 1.5% of total health expenditures and 11% of out-of pocket costs.

The appeal of unproven CAM approaches continues to surprise many physicians. Many factors contribute to these choices. Some patients seek out CAM practitioners because they offer optimism or greater personal attention. For others, alternative approaches reflect a "self-help" approach to health and wellness or satisfy a search for "natural" or less invasive alternatives, since dietary supplements and other natural products are believed, correctly or not, to be inherently healthier and safer than synthetic ones. In NHIS surveys, the most common health conditions cited by patients for CAM use involve management of symptoms often poorly controlled by conventional care, particularly back pain and other painful musculoskeletal complaints.

FIELDS OF PRACTICE AND LICENSURE

At present, six fields of CAM practice—osteopathic manipulation, chiropractic, acupuncture and traditional Asian medicine, massage therapy, naturopathy, and homeopathy—are subject to licensure requirements and some form of educational accreditation. Mind-body practices such as meditation and yoga are not licensed in any state, and training in those practices is not subject to national accreditation.

OSTEOPATHIC MEDICINE

Founded in 1892 in the American heartland by the physician Andrew Taylor Still, osteopathic medicine was based originally on the belief that manipulation of soft tissue and bone can correct a wide range of diseases of the musculoskeletal and other organ systems. Over the ensuing century, osteopathy evolved progressively toward conventional (allopathic) medicine. Today, the training, practice, credentialing, licensure, and reimbursement of osteopathic physicians are virtually indistinguishable from those of allopathic physicians, with 4 years of osteopathic medical school followed by specialty and subspecialty training and certification by organizations such as the American Board of Internal Medicine. Some osteopathic physicians continue to practice spinal manipulation, primarily as a tool to address specific musculoskeletal complaints.

CHIROPRACTIC

In 1895, Daniel David Palmer founded in Missouri the first school of chiropractic medicine to teach manipulation of the spine. Palmer believed that subluxations, or partial dislocations of vertebrae, cause disease by impinging on key nerve roots. Today, chiropractors undertake 5 years of training in basic and relevant clinical sciences. Increasingly, they complete additional postgraduate training in radiology and outpatient therapeutics, primarily of musculoskeletal conditions, although within the discipline there are factions that continue to perform manipulation for many other pathologic entities. Chiropractors also advise on nutrition, exercise, and other health maintenance approaches. Over 50,000 doctors of chiropractic medicine are licensed to practice in the United States.

ACUPUNCTURE AND TRADITIONAL ASIAN MEDICINE

A venerable component of traditional Chinese medicine, acupuncture emerged in recent decades as a free-standing clinical discipline. Over 3000 American physicians have acquired targeted postgraduate training that permits them to practice acupuncture in over 40 states and the District of Columbia. Over 4000 non-MDs have taken far more extended training that leads to licensure to practice independently or under the supervision of a physician.

MASSAGE THERAPY

Drawing on millennia of empirical knowledge, some 80 schools in the United States instruct students in an array of the soft tissue manipulative approaches that constitute massage. Thirty-one states and the District of Columbia license trainees to perform therapeutic massage.

NATUROPATHY

Fifteen states license practitioners of naturopathy, a discipline that emerged in central Europe in the late eighteenth century. The fact that conventional treatments of the day were usually ineffective, if not overtly harmful, stimulated the search for safer and more "natural" approaches—naturopathy is one of them. The concept underlying this discipline is that the body has powerful mechanisms for self-healing that a properly instructed practitioner can harness. About 1400 naturopathic physicians have completed 4 years of education in basic and clinical sciences and are licensed to manage a predominantly outpatient population. Conventional and unconventional diagnostic tests and medications are prescribed, with an emphasis on relatively low doses of drugs, herbal medicines, special diets, and exercises.

HOMEOPATHY

The late eighteenth century also witnessed the emergence of homeopathy, another discipline that arose at least in part as a reaction to the toxicity inherent in many of the allopathic approaches of the day. Homeopathy was developed by Samuel Hahnemann, a German physician, who postulated that substances that cause particular side effects in a well person may be used to treat or prevent such symptoms in an ill person if administered in minuscule amounts—what is known as "the doctrine of similars." For example, contact with poison ivy *(Rhus toxicodendron)* causes susceptible persons to experience an itchy, blistering rash. Homeopathy espouses the administration of highly diluted extracts of poison ivy to treat other blistering, pruritic eruptions, such as varicella. During the early nineteenth century, the then-nascent field of homeopathy used blinded tests on volunteers, presaging wider use of placebo-controlled trials, to "prove" which materials were the most able to induce or relieve symptoms. By the mid-nineteenth century homeopathy had gained considerable presence in the American medical establishment and may, in fact, have facilitated the development of immunization and allergen desensitization, both of which utilize very small quantities of materials to elicit measurable biologic outcomes. Today, however, homeopathy is accepted less fully in the United States than in some other countries. It is the largest of all CAM practices in the United Kingdom, Germany, and France and is widely used in India. In the United States, only three states currently license the practice of homeopathy. The relative decline of homeopathy relates, at least in part, to the field's inability to articulate a rational mechanism that explains why products that are diluted more than 10^{60}-fold, greater than Avogadro's number, could incite biologic effects. Nonetheless, homeopathic remedies are readily available and commonly recommended by naturopathic physicians and other licensed and unlicensed practitioners.

REGULATION OF DIETARY SUPPLEMENTS

Herbal medicines, and dietary supplements more generally, have a unique regulatory status that gives the public remarkable freedom of choice but also many undesired challenges. An element of traditional healing approaches, herbal medicines were presumed safe long before the implementation of drug regulations by the U.S. Food and Drug Administration (FDA). In 1994, the U.S. Congress passed the Dietary Supplements Health and Education Act (DSHEA), which permits sale of dietary supplements "over the counter," as it were, but without the requirement imposed on manufacturers of prescription or conventional over-the-counter drugs to prove that their products are safe and effective before marketing. Supplements can be removed by the FDA from the market only if they are proved to be hazardous. Purveyors of dietary supplements cannot claim that they prevent or treat any disease. They can, however, claim that they maintain "normal structure and function" of body systems. For example, a product cannot claim to treat arthritis, but it can claim to maintain "normal joint health." Homeopathic products predate FDA drug regulations and are sold with no requirement that they be proved effective. Although homeopathic products are widely believed to be safe because they are highly dilute, one product, a nasal spray called Zicam, was withdrawn from the market when it was found to produce anosmia, probably because of a significant zinc content. Homeopathic products, and indeed other CAM products and practices, also convey the very significant risk that individuals will use them instead of effective conventional modalities.

Under the current regulatory framework, members of the public have considerable freedom to determine what is in their own best interest, even if those decisions deny them effective treatment; however, the courts have ruled that the rights of parents to withhold treatment of their children is limited in instances of life-threatening illnesses. Investigators have a broad ethical obligation not to withhold proven treatments for serious illnesses for the sake of testing unproven ones.

SAFETY

Risks imposed by the use of CAM approaches include injuries inflicted by a practice, inherent toxicities of the modality, and interference by the modality with more conventional treatments.

INJURY

Physical and manipulative interventions can harm patients. In past decades, reused acupuncture needles transmitted hepatitis B virus infection; today, the standard of care requires disposable needles. Aggressive massage can cause soft tissue injuries. Spinal manipulation of patients with unrecognized vertebral lesions has been associated with cord injuries, and cervical manipulation has been associated with stroke. These appear to be rare events.

INHERENT TOXICITY

Although the public may believe that "natural" equates with "safe," it is abundantly clear that natural products can be toxic. Misidentification of medicinal mushrooms has led to liver failure. Contamination of tryptophan supplements caused the eosinophilia-myalgia syndrome. Herbal products containing particular species of *Aristolochia* were associated with genitourinary malignancies. In 2001, extracts of kava, long used by Pacific Islanders for its mild anxiolytic and sedative properties, were associated with fulminant liver failure. A number of products, including the popular *Ginkgo biloba*, are known to prolong bleeding times and have been associated with postoperative hemorrhage. Among the most controversial dietary supplements is *Ephedra sinica*, or ma huang, a product used in traditional

Chinese medicine for short-term treatment of asthma and bronchial congestion. The scientific basis for these indications was revealed when ephedra was shown to contain the ephedrine alkaloids, especially ephedrine and pseudoephedrine. With the promulgation of the DSHEA regulations, supplements containing ephedra and herbs rich in caffeine sold widely in the U.S. marketplace because of their claims to promote weight loss and enhance athletic performance. Reports of severe and fatal adverse events associated with use of ephedra-containing products led to an evidence-based review of the data surrounding them, and in 2004 the FDA banned their sale in the United States.

ADULTERATION

A major current concern with dietary supplements is adulteration with pharmacologic active compounds. Multi-ingredient products marketed for weight loss, body building, "sexual health," and athletic performance are of particular concern. Recent FDA recalls have involved adulteration with steroids, diuretics, stimulants, and phosphodiesterase type 5 inhibitors.

HERB-DRUG INTERACTIONS

The constituents of a few natural products are toxic; others are known to interfere with the metabolism of life-saving drugs. This effect was illustrated most compellingly with the demonstration in 2000 that consumption of St. John's wort interferes with the bioavailability of the HIVprotease inhibitor indinavir. Later studies showed its similar interference with metabolism of topoisomerase inhibitors such as irinotecan, with cyclosporine, and with many other drugs. The breadth of interference stems from the ability of hyperforin in St. John's wort to upregulate expression of the pregnane X receptor, a promiscuous nuclear regulatory factor that promotes the expression of many hepatic oxidative, conjugative, and efflux enzymes involved in drug and food metabolism.

Because of the large number of compounds that alter drug metabolism and the large number of agents some patients are taking, identification of all potential interactions can be a daunting task. Several useful Web resources can ameliorate this problem **(Table e2-2)**. Clearly, attention to this problem is particularly important with drugs with a narrow therapeutic index, such as anticoagulants, antiseizure medications, immunosuppressants, and cancer chemotherapeutic agents.

Table e2-2 Resources for Dietary Supplement Drug Interactions

Medscape
http://www.medscape.com/druginfo/druginterchecker?cid=med
This Web site is maintained by WebMD and includes a free drug interaction checker tool that provides information on interactions between two or more drugs, herbals, and/or dietary supplements.
Natural Medicine Comprehensive Database
http://naturaldatabase.therapeuticresearch.com
This Web site provides an interactive natural product/drug interaction checker tool that identifies interactions between drugs and natural products, including herbals and dietary supplements. This service is available by subscription. A PDA version is available.
Natural Standard
http://www.naturalstandard.com/tools
This Web site provides an interactive tool for checking drug and herb/supplement interactions. This service is available by subscription. A PDA version is available.

THE EVIDENCE BASE FOR CAM

Alternative health practices have evolved through an epistemologic framework completely different from that of contemporary biomedicine. Empirical observations of individual patients constitute the primary evidentiary base on which CAM practices are guided and taught. Nonetheless, over the last few decades, thousands of studies have been performed of various CAM approaches, including hundreds of trials involving herbs, acupuncture, and homeopathy. To date, very few CAM approaches have been proved definitively to be effective. Several factors contribute to this lack of convincing evidence. The vast majority of CAM studies have been seriously flawed by lack of appropriate controls, bias on the part of the investigators, small sample sizes, reliance on highly subjective and nonvalidated measures of benefit, and inappropriate statistical tests.

Methodologic Challenges

In addition, a series of methodologic issues challenge even the better-designed CAM studies. By and large, uniform practice guidelines do not exist, and the herbal products marketed in the United States are highly variable in quality and composition. Many CAM practices are not amenable to blinding. For example, both the patient and the practitioner would know if spinal manipulation had been performed, and conventional research paradigms cannot be used to test these approaches. These problems are not unique to CAM, as they also complicate attempts to study conventional practices such as

psychotherapy and surgery.

Even with ongoing improvements in study design and conduct, issues of belief also stand in the way of comprehending and accepting the results of some CAM studies. Many physicians are reluctant to believe positive outcomes of clinical approaches that have not emerged through the classic experimental paradigm by which drugs and biologic agents are developed: the orderly progression from preclinical testing through serial phases of clinical trials. More important, it is difficult to accept results that are counterintuitive or the underlying mechanism of which defies rational explanation. As suggested above, an example of this dilemma involves studies of homeopathy. Some clinical trials of homeopathy for asthma, infantile diarrhea, and other common conditions reported positive results. Two systematic reviews of homeopathy trials reported an overall favorable impression of the clinical trials data, concluding that the treatments were more beneficial than placebo. Even the best trials and those reviews have been criticized on methodologic grounds. It remains unclear what evidence could compel a change in belief about the benefits of homeopathy when there are no cogent explanations for how substances diluted beyond the point at which only solute remains could exert physiologic effects.

Scientific Evidence and Public Acceptance

By contrast, although methodologic problems continue to plagueacupuncture trials, belief has been growing even in academic centers that acupuncture may be effective. The emerging acceptance of acupuncture results in part from its widespread availability and use in the United States today, even within the walls of major medical centers where it is used as an ancillary approach to pain management. Yet its acceptance appears to stem from more than just its communal appeal. Since the mid-1970s, biochemical and imaging studies have begun to yield evidence that needling can alter central pain-processing pathways, possibly by triggering release of neural mediators that bind to specific receptors in the brain regions that modulate pain perception.

Although it is difficult to conclude decisively that a particular CAM approach lacks any merit, it is quite feasible to discern that its effect size, or degree of benefit, is too small or inconsistent to be worth pursuing further. Over the last century, many once unconventional medical approaches failed—one need only think back to the exotic electrical devices, procedures, and tonics that fell out of fashion. Two questions often are asked: (1) Do any of the more contemporary CAM modalities deserve to be rejected? (2) Would data showing them to be ineffective change anyone's mind about using them?

The case of laetrile is instructive. This extract of apricot seeds was touted in the 1970s as a cure for solid tumors. Thousands crossed the Mexican border to secure laetrile for their personal use. The lack of any positive preclinical data discouraged oncologists from agreeing to study laetrile until public pressure required that an answer be obtained. Two studies in the 1980s showed no benefit of laetrile treatment. Today, some continue to seek the product, but the numbers are vastly smaller than they were before meaningful data were obtained. A similar fate befell a cocktail of drugs used for cancer patients through the 1970s and 1980s by Dr. Luigi Di Bella in Italy once large studies revealed that it had no detectable impact on the course of a variety of advanced cancers.

In contrast, some modalities that have been well tested and found ineffective are still in fairly common use. For example, the renowned biochemist Linus Pauling proclaimed that vitamin C can treat and even prevent the common cold. Several high-quality studies failed to demonstrate clinically important effects of vitamin C in preventing or treating viral colds. Nonetheless, ingestion of extra vitamin C remains a common habit of individuals who perceive the onset of cold symptoms. For most people, this practice is wasteful but not harmful; however, people with iron overload (either hemochromatosis or chronic transfusion requirement) can be damaged by vitamin C, which generates free radicals in the setting of iron excess.

The Impact of Emerging Evidence

Nevertheless, emerging evidence on CAM approaches is having some impact on public practice. In the last decade the National Institutes of Health has sponsored a number of major randomized clinical trials of widely used dietary supplements. By and large, these studies have failed to confirm the benefits expected from the smaller studies that preceded them. *Ginkgo biloba* was not found to prevent cognitive decline or dementia. No clear-cut benefits of glucosamine for most patients with osteoarthritis of the knee could be demonstrated. Vitamin E and or selenium did not change the rates of development of prostate cancer. St. John's wort *(Hypericum perforatum)* did not have an impact on major depression. Although controversy exists about a number of these studies, they have received major media coverage; the research results have affected determinations by the FDA and the U.S. Federal Trade Commission about permitted marketing claims; and analysis by the supplement industry indicates that the research has had substantial effect on the sale of these compounds.

INFORMATION SOURCES

One of the difficult problems facing practicing physicians is providing patients with good advice and education about CAM practices. Irresponsible and misleading marketing claims abound, particularly on the Internet. Claims about antiaging regimens, enhancement of sexual function, weight loss, and increased athletic capacity are particularly common. A number of valuable Internet resources exist that

provide good sources for patient education and can help counter some of this misinformation. Some particularly useful sites are summarized in **Table e2-3**.

Table e2-3 Internet Resources on Complementary and Alternative Medicine (CAM)
The Cochrane Collaboration Complementary Medicine Reviews
This Web site offers rigorous systematic reviews of mainstream and CAM health interventions using standardized methods. It includes more than 300 reviews of CAM therapies. Complete reviews require institutional or individual subscription, but summaries are available to the public.
http://www.cochrane.org/reviews/en/topics/22_reviews.html
MedlinePlus All Herbs and Supplements, A–Z List
MedlinePlus Complementary and Alternative Medicine
NLM FAQ: Dietary Supplements, Complementary or Alternative Medicines
These National Library of Medicine Web pages provide an A–Z database of science-based information on herbal and dietary supplements; basic facts about CAM therapies; and federal government sources on information about using natural products, dietary supplements, medicinal plants, and other CAM modalities.
http://www.nlm.nih.gov/medlineplus/druginfo/herb_All.html
http://www.nlm.nih.gov/medlineplus/complementaryandalternativemedicine.html
http://www.nlm.nih.gov/medlineplus/dietarysupplements.html
NIH National Center for Complementary and Alternative Medicine (NCCAM)
This National Institutes of Health NCCAM Web site contains information for consumers and health care providers on many aspects of CAM. Downloadable information sheets include short summaries of CAM therapies, uses and risks of herbal therapies, and advice on wise use of dietary supplements.
http://www.nccam.nih.gov

In the last decade, it has become possible for scientists to perform the kind of rigorous systematic reviews of CAM that are the cornerstone of evidence-based medicine. A particularly valuable resource in this respect is the Cochrane Collaboration, which has performed more than 300 systematic reviews of CAM practices. Practitioners will find this a valuable resource to answer patient questions. By and large, most of these summaries do not conclude there is definite benefit; any beneficial effects that have been documented are modest but frequently are balanced by little indication of risk.

SUMMARY

An array of unproven modalities will always be used by the patients under physicians' care. Physicians must approach each encounter as an opportunity to understand their patients' beliefs and expectations and use those insights to help guide their personal health care practices in a constructive way. Many of these choices are entirely innocuous and can be accommodated in the context of the established diagnostic and therapeutic interventions. Some of these choices should be actively discouraged. Along the way, scientific evidence will drive many CAM approaches out of favor. Some modalities will garner sufficient support to become part of mainstream care and the next generation of physicians will never know they were once controversial.

FURTHER READINGS

BARNES PM et al: Complementary and alternative medicine use among adults and children: United States, 2007. Natl Health Stat Rep 12:1, 2008[PMID: 19361005]

EISENBERG DM et al: Trends in alternative medicine use in the United States, 1990–1997: Results of a follow-up national survey. JAMA 280:1569, 1998[PMID: 9820257]

KAPTCHUK TJ: Acupuncture: Theory, efficacy and practice. Ann Intern Med 136:374, 2002[PMID: 11874310]

KINSEL JF, STRAUS SE: Complementary and alternative therapeutics: Rigorous research is needed to support claims. Annu Rev Pharmacol Toxicol 43:463, 2003[PMID: 12540748]

MACPHERSON H et al: Closing the evidence gap in integrative medicine. BMJ 339:b3335, 2009

MANEK NJ et al: What rheumatologists in the United States think of complementary and alternative medicine: Results of a national survey. BMC Complement Altern Med 10:5, 2010[PMID: 20109215]

SNYDERMAN R, WEIL AT: Integrative medicine: Bringing medicine back to its roots. Arch Intern Med 162:395, 2002[PMID: 11863470]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE ECONOMICS OF MEDICAL CARE: INTRODUCTION

This chapter attempts to explain to physicians how economists think about physicians and medical care. Economists' mode of thinking has shaped health care policy and institutions and thus the environment in which physicians practice. As a result, it may be useful for physicians to understand some aspects of this way of thinking even if at times it may seem foreign or uncongenial.

Physicians see themselves as professionals and as healers, assisting their patients with their health care needs. When economists are patients, they probably see physicians the same way, but when they view doctors through the lens of economics as a discipline, they see them as economic agents. In other words, economists are interested in the degree to which physicians respond to various incentives, both those that they face and those facing their patients, in deciding how to deploy the resources they control. Examples include how much of their own time to devote to seeing a patient; which tests to order; what drugs, if any, to prescribe; whether to recommend a procedure; whether to refer a patient; and whether to admit a patient.

This interest stems from fundamental economic questions: What goods and services are produced and consumed? In particular, how much medical care is available, and how much of other goods and services? How is that medical care produced? For example, what mix of specific services is used to treat a particular episode of illness? Who receives particular treatments?

Physicians in all societies live and function in economic markets, although those markets differ greatly from the simple competitive markets depicted in introductory economic textbooks and also differ from country to country, depending on an individual country's institutions. Many of the differences between actual medical markets and textbook competitive markets cause what economists term *market failure*, a condition in which some individuals can be made better off without making anyone else worse off.

This chapter explains two features of health care financing that cause market failure: selection and moral hazard. A common response to market failure in medical care is what economists term administered prices, which is another concept this chapter describes. Administered prices also exact an economic cost, leading to what economists call *regulatory failure*. All developed societies seek a balance between market failure and regulatory failure, a topic addressed in this chapter's conclusion.

Selection

In the idealized competitive market found in economic textbooks, buyers and sellers know the same amount about the good or service they are buying and selling. When one party knows more—or when goods of different quality are being sold at the same price, which is analytically similar—markets can break down in the following sense: There may be a price at which an equally well informed buyer and seller could make a transaction that would make them both better off, but the transaction does not occur because one party knows more than the other. Hence, both the potential buyer and the potential seller are worse off.

The used car market is a classic example of differential information. Owners of used cars (potential sellers) know more about the quality of their cars than do potential buyers. At any specific price for a certain make and model of car, the only used cars offered will be those whose sellers value them at less than that price. Such cars will differentially be of low quality ("lemons") relative to the given price; in fact, assuming a continuum of quality, the average cars offered will be those which are valued at or less than the price in the market. However, that means that any potential buyer, lacking information about the quality of a car, would potentially pay (much) more than that car is worth. Because buyers know that the sellers know more about the quality of the car, transactions that would occur if a buyer and seller had equally good information about the quality of the car may not occur. (It is for this reason that sellers may offer warranties and guarantees.)

The same thing happens when goods of different quality are sold at the supermarket at the same price. Shoppers are happy to take quickly any box of a particular brand of breakfast cereal or bottle of soft drink on the shelf because the quality of any box or bottle is the same, but they will take their time inspecting produce to make sure that the apple they pick up and put in the cart is not bruised or the banana is not overly ripe. At the end of the day, it is the bruised apples and overly ripe bananas that are left in the store. In effect, the seller has not used all the information in pricing the produce, and buyers exploit that information differential.

Selection affects markets for individual and, to some degree, small group health insurance in a fashion

similar to the used car market and the produce stand, but in this case it is the buyer of insurance who has more knowledge than the seller. Individuals who use above-average amounts of care—for example those with a chronic disease or a strong proclivity to seek care for a symptom—will value health insurance more than will those who are healthy or who for various reasons shun medical care even if they are symptomatic. However, the insurer does not necessarily know the risk of those it insures, and so it gears insurance premiums to an average risk or sometimes an average risk conditional on certain observable characteristics, such as age. Just as shoppers do not want the bruised apples and used car buyers do not want lemons, many healthy people will not want to buy insurance voluntarily if its price mainly reflects the use of those who are sick. (Healthy but very risk averse individuals still may be willing to pay premiums well above their expected use.) In an extreme case, healthy people drop out of the insurance pool, premiums rise (the average person left in the pool is sicker), and that rise causes still more people to drop out of the pool, causing premiums to rise further, and so forth, until few people are left buying insurance.

For this reason, no developed country relies primarily on voluntary individual insurance to finance health care, although many countries use it in the supplemental insurance market, and selection is often a feature of that market. Instead, governments and/or employers provide or heavily subsidize the purchase of either mandated or voluntary health insurance (e.g., the American Medicare and Medicaid programs, Canada, Germany, the state of Massachusetts) or provide health services directly (e.g., the United Kingdom, the United States Veterans Health Administration, and the Indian Health Service).

Moreover, countries that rely on employment-based health insurance, such as the United States and Germany, either mandate taxes to finance that insurance or provide large tax subsidies for its purchase; otherwise, many healthy employees would prefer that the employer give them the money the employer pays toward the insurance in the form of cash wages. Because an employer who offers health insurance will pay lower cash wages than an otherwise equivalent employer that does not, many American employers, who are not required to offer insurance, may not do so in low-wage industries; if they did, the cash wage they could afford to pay could be below the minimum wage. (Nor, typically, do they offer a pension benefit for the same reason.) These low-wage employers are often, but not always, small businesses. Whereas 18% of the American labor force worked in firms with 3 to 24 employees in 2008, only 13.8% of workers with health insurance were employed in those firms. Some self-employed individuals or those who work at small firms may belong to a trade association or a professional society through which they can purchase insurance (including some physicians), but because that purchase is voluntary, it is subject to selection. Those who do not have access to such insurance must use the individual insurance market.

How does this affect the practice of medicine? Individual and small group insurance policies typically have preexisting condition clauses that protect the insurer against selection, that is, provide protection against a person's purchasing insurance (or more complete insurance) after that person has been diagnosed with a disease that is expensive to treat. Even so, elements of selection remain in the individual insurance market, and thus premiums tend to be high. As a result, Americans who are not eligible for employer-subsidized insurance or who cannot obtain such insurance through a spouse may not voluntarily purchase insurance, yet many of those persons will seek care even without insurance. Caring for such "self-pay" patients may give the physician a choice between making do with a less than clinically optimal treatment and proceeding in a clinically optimal way but leaving the patient with a large bill and possible bankruptcy (and leaving the physician with bill collection issues or unpaid bills). Those who do purchase insurance at high premiums may buy a bare bones policy or a policy that covers only a few physician visits and some drugs, again leaving the physician the problem of determining a realistic treatment plan that the patient can afford.

Selection can arise in a different guise when physicians are reimbursed a fixed amount per patient (i.e., capitation) rather than receiving fee-for-service payments. Depending on the adequacy of any adjustments in the capitated amount for the resources that a specific patient will require—frequently, there are no such adjustments—physicians who receive a fixed amount have a financial incentive to avoid caring for sicker patients. Similarly, physicians who receive a capitated amount for their own services but are not financially responsible for the services of other physicians may make an excessive number of referrals, just as a physician reimbursed in a fee-for-service arrangement may make too few.

Moral Hazard

The term *moral hazard* comes from the actuarial literature; it originally referred to the weaker incentives of an insured individual to prevent the loss against which he or she is being insured. A classic example is failure of homeowners in areas prone to brush fires to cut brush around their houses or possibly install fire-resistant shingles on their roofs because of their expectation that insurance will compensate them if their houses burn down. In some lines of insurance, however, moral hazard is not a large issue. Persons who buy life insurance on their own lives are not likely to commit suicide so that the policy will pay off. (Because of moral hazard, however, the law prohibits buying life insurance on another person with whom one does not have a relationship or a substantial economic interest.) Also, despite the brush fire example, homeowners insurance probably has little moral hazard associated with it because individuals often cannot replace lost belongings such as a photo album

when a house burns down or property that is destroyed when there is a burglary. In short, if moral hazard is negligible, insured persons take appropriate precautions against the potential loss.

In the context of health insurance, this classic form of moral hazard means potentially reduced incentives to prevent illness and is probably not a large issue. Sickness and disease generally imply some pain and suffering, not to mention possibly shortened life expectancy. Because there is no insurance for the pain and suffering, individuals have strong incentives to try to remain healthy regardless of how much health insurance they have. Put another way, having better health insurance probably does not strengthen those incentives much.

Instead of weakened incentives to prevent illness, in the health insurance context, moral hazard typically refers to the incentives for better-insured individuals to use more medical services. For instance, a patient with back pain or shoulder pain might request an expensive imaging test such as an MRI if it is free or of low cost to him or her even if the physician feels the clinical value of the test is negligible. Conversely, the physician may be more cautious in ordering a test that seems likely to produce little information if there are severe financial consequences for the patient.

The strongest evidence on this point comes from a randomized experiment done in the late 1970s and early 1980s: the RAND Health Insurance Experiment. Families whose members were <65 years of age were randomized to insurance plans in which the amount they had to pay when using services ("cost sharing") varied from nothing (fully insured care) to a large deductible (catastrophic insurance). All the plans capped families' annual out-of-pocket payments, with a reduced cap for low-income families. Families with complete insurance used about 40% more services in a year than did families with catastrophic insurance, a figure that did not vary much across the six geographically dispersed sites in which the experiment was run. Although these data come from the era before managed care in the United States, subsequent observational studies in the United States and elsewhere have largely confirmed its findings with respect to the relationship between variations in care use and variations in patient payment at the point of service.

Those in the RAND experiment for whom medical care was free were more likely to seek care for problems for which care could be efficacious. One would assume that this would have resulted in improved outcomes, but by and large it did not. In fact, there was little or no difference in average health outcomes among those on different health plans with the important exception that people with hypertension, especially those with low incomes, were better controlled when care was free. Moreover, those in the experiment for whom medical care was free were also more likely to seek care for problems for which care was not likely to be useful.

A possible explanation for the paucity of beneficial effects from the additional medical services those on the free plan used is that the population in the experiment, which consisted of nonelderly community-dwelling individuals, was mostly healthy. Those with a large deductible made about two physician visits each year on average; those whose care was free made about twice that number and were hospitalized roughly 25% more often. It is possible that the additional two visits and the greater number of hospitalizations were as likely to lead to poor outcomes as good outcomes in that population. Certainly, the subsequent literature on quality of care and medical error rates implies that a good deal of inappropriate care was—and is—provided to patients. For example, over half the antibiotics prescribed to the experiment's participants were for viral conditions. Moreover, about a quarter of those hospitalized (in all plans) were admitted for procedures that could have been done equally well outside the hospital, consistent with the large fall in hospital use over the last three decades. In short, the additional inappropriate care when care was free was not necessarily innocuous; if a mainly healthy person saw a physician, he or she could have been made worse off. The literature on inappropriate care is mostly American in origin, but the finding probably holds elsewhere as well.

Insurance is certainly desirable to protect families against the financial risk of large medical expenses and in some instances to address underuse of valuable services, such as a patient with diabetes or hypertension failing to take medications for financial reasons. Thus, the remedy for moral hazard is not to abolish insurance but instead to strike the right balance between financial protection and incentives to seek care. Moreover, it is probably useful to vary the amounts patients pay out of pocket depending on the specific service and the patient's clinical condition. It may be better, for example, to have small or no copayments for antidiabetics or antihypertensives or statin drugs for patients who have had a myocardial infarction and have higher copayments for drugs whose use is more discretionary.

Administered Prices

Because insurers, whether public or private, cannot pay any price a physician sets, prices in medical markets with widespread insurance are either set administratively or negotiated. In the simple textbook model, competitive market prices approximate the cost of production, but this does not necessarily happen when prices are administered. In the American Medicare program, for example, the government sets a take-it-or-leave-it price. Because of the market share represented by the Medicare program, virtually all physicians choose to take the government's price rather than leave the program. In some countries (e.g., Canada and Germany), medical societies negotiate fees for all physicians in the nation or in a subnational area. In the United States, commercial insurers negotiate fees with

individual physicians or physician groups.

The principal problem with administered prices is that someone must set them. If the price that is set departs markedly from incremental cost, distortions inevitably result; the price setter typically has very little information about incremental cost, and so the price could be (and often is) far from the cost. If the regulator sets the price below cost, the service may not be available, or if it is, it will have to be cross-subsidized from a profitable service. If the price is set above cost, there may well be excess entry and too many services being offered on too small a scale. Moreover, cost probably varies across physicians and hospitals, meaning that a single price will not fit all. A beneficent regulator in theory could approximate an equilibrium price and the cost of production by trial and error if technology did not change, but clearly, that condition does not hold in medical care. Not only do new goods and services appear continually, physicians often become more skilled at delivering a service that already is available or developing new tools to deliver that service at a different and frequently lower cost. For example, cataract surgery, which took upward of 8 hours when first introduced, can now be completed in <30 minutes.

The distortions between price and cost when prices are administered have consequences for the way medical care is produced. There may well be too much capacity in "profitable" areas of medicine, such as cardiac services and sports medicine, and too little in less profitable areas, such as primary care. A fee above cost for a procedure encourages more procedures.

Conversely, payment methods that attempt to pay for many services with one fixed payment, such as capitation and a per-admission payment, pay nothing for doing more and therefore may result in too few services or in selection by providers to reduce the number of unprofitable patients under their care, much as a hospital may shutter an emergency room if it becomes a magnet for the uninsured. As was noted above, these phenomena also reflect the asymmetry of information between patients and physicians and, in the case of fee-for-service payment, the incentive for insured patients to go along with a recommendation for additional services ("I am pretty sure I know what the problem is, but let's just carry out this additional test to be sure").

There is good evidence that physicians as a group respond to the prices that are set. For example, if there is a general reduction in fees that, other things equal, would lower practice income, physicians order more services; conversely, they do the opposite if there is an increase in fees. This behavior is sufficiently well established empirically that the U.S. Medicare program's actuaries account for it in their cost estimates of what various changes in fees will cost or save.

Negotiated prices may get closer to cost than administered prices that are set, but they are not a panacea. First, the relevant cost for patients is the total cost of treating the entire medical problem and the outcome obtained, both of which probably are not reflected in the negotiated physician's fee. Second, in the United States, commercial insurers often negotiate fees as a multiple of the Medicare fee schedule, and so any distortion in the administratively determined relative fees is carried over into the negotiated fees. For example, in the Medicare fee schedule, procedures generally are more profitable than cognitive services known as "evaluation and management," and this probably plays a role in the United States having too few primary care physicians. Third, both in the United States and elsewhere, the negotiation may be for a large number of physicians who have different costs, in which case fees will depart from costs for some or all the physicians in the group. Fourth, negotiated prices may well exceed cost when there is no effective competition among similar physicians in a particular market. Because medical care markets are typically local, there may only be one group in any particular specialty in a smaller market, in which case that group will have considerable market power to obtain more favorable reimbursement. Related to this point, physicians may seek to negotiate together to increase their market power. Finally, many, probably most, patients are reluctant to change physicians because their current physician knows their medical history and because they are uncertain whether a new physician would be an improvement and also because insurance shields them from most of the cost differences among physicians.

Conclusion

One branch of economics—positive economics—seeks to explain actual phenomena without making a judgment about the desirability of those phenomena. Another branch—normative economics—seeks to prescribe what should happen and, in particular, what public policy should be to ensure that it happens. Its main result is that under certain very special assumptions, competitive markets will lead to a result in which no one can be made better off without another person being made worse off. These assumptions do not hold in medical care, in part because of selection and moral hazard; economists term the result a market failure. By contrast, even a beneficent regulator will introduce distortions from lack of sufficient information, as the discussion of administered prices in this chapter indicated, and there is no guarantee that a regulator will be beneficent, as periodic corruption scandals indicate. Economists term this phenomenon regulatory or government failure. Economists see decisions about the proper form and amount of public intervention and regulation in medical care as a matter of finding the right balance between various types of market failures and various types of regulatory failures, a balance that different societies may choose to strike differently.

FURTHER READINGS

Brook RH: Assessing the appropriateness of care—its time has come. JAMA 302:997, 2009[PMID: 19724048]

Chernew ME et al: Value-based insurance design. Health Affairs 26:w195, 2007

Fuchs VR: Health care for the elderly: How much? Who will pay for it? Health Affairs 18:11, 1999[PMID: 9926642]

Institute of Medicine: *To Err Is Human*. Washington, DC, National Academy Press, 1999

———: *Crossing the Quality Chasm.* Washington, DC, National Academy Press, 2001

McGuire TG: Physician fees and behavior, in *Incentives and Choice in Health Care*, FA Sloan, H Kasper (eds). Cambridge, MA, MIT Press, 2008

Newhouse JP: *Pricing the Priceless: A Health Care Conundrum.* Cambridge, MA, MIT Press, 2002

———, Sinaiko AD: What we know and don't know about the effects of cost sharing on demand for medical care—and so what? in *Incentives and Choice in Health Care,* FA Sloan, H Kasper (eds). Cambridge: MIT Press, 2008

Rice TH: The impact of changing Medicare reimbursement rates on physician-induced demand. Medical Care 21:803, 1983[PMID: 6350746]

Siu A et al: Inappropriate use of hospitals in a randomized trial of health insurance plans. N Engl J Med 315:1259, 1986[PMID: 3773939]

Zuckerman S et al: Price controls and Medicare spending. Med Care Res Rev 55:457, 1998[PMID: 9844351]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

RACIAL AND ETHNIC DISPARITIES IN HEALTH CARE: INTRODUCTION

Over the course of its history, the United States has experienced dramatic improvements in overall health and life expectancy due largely to initiatives in public health, health promotion, disease prevention, and chronic care management. Our ability to prevent, detect, and treat diseases in their early stages has allowed us to target and reduce morbidity and mortality. Despite interventions that have improved the overall health of the majority of Americans, racial and ethnic minorities (blacks, Hispanics/Latinos, Native Americans/Alaskan Natives, Asian/Pacific Islanders) have benefited less from these advances and suffer poorer health outcomes than whites from many major diseases (e.g., cardiovascular disease, cancer, diabetes) in the United States. Research has highlighted that minorities may receive less care and lower quality care than whites, even when confounders such as stage of presentation, comorbidities, and health insurance are controlled. These differences in quality are called *racial and ethnic disparities in health care*. This chapter will provide an overview of racial and ethnic disparities in health and health care, identify root causes, and provide key recommendations to address them at both the clinical and health system levels.

Nature and Extent of Racial and Ethnic Disparities in Health and Health Care

Minority Americans have poorer health outcomes (compared with whites) from preventable and treatable conditions such as cardiovascular disease, diabetes, asthma, cancer, and HIV/AIDS, among others **(Fig. e4-1)**. Multiple factors contribute to these racial and ethnic disparities in health. First and foremost, there is little doubt that social determinants—such as lower levels of education, lower socioeconomic status, inadequate and unsafe housing, racism, and living in close proximity to environmental hazards—disproportionately impact minority populations and thus contribute to poorer health outcomes. For example, three of the five largest landfills in the country are found in black and Latino communities; these environmental hazards have contributed to some of the highest rates of pediatric asthma among these populations. Second, lack of access to care also takes a significant toll, as uninsured individuals are less likely to have a regular source of care, and are more likely to delay seeking care and go without needed care—all resulting in avoidable hospitalizations, emergency hospital care, and adverse health outcomes.

Figure e4-1

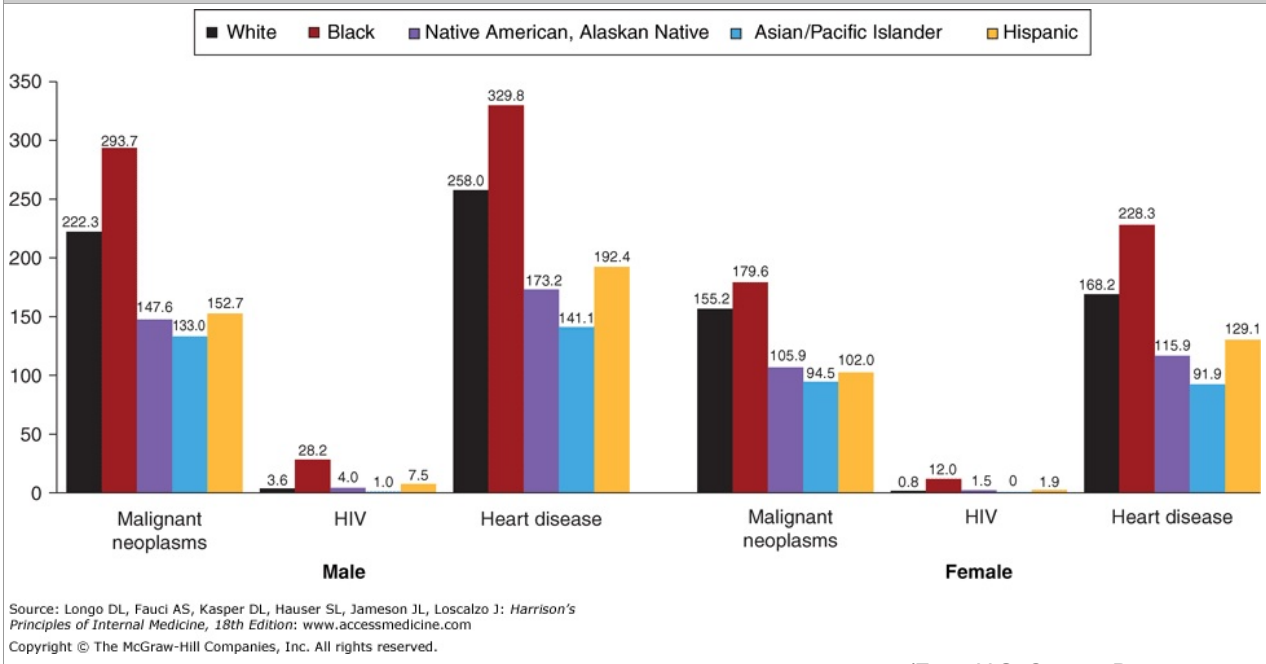

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Age-adjusted death rates for selected causes by race and Hispanic origin, 2005. *(From U.S. Census Bureau, 2009.)*

In addition to the existence of racial and ethnic disparities in *health*, there are racial/ethnic disparities in the *quality of care* for those with access to the health care system. For instance, disparities have been found in the treatment of pneumonia **(Fig. e4-2)** and congestive heart failure (blacks receiving less optimal care than whites when hospitalized for these conditions) and referral to renal transplantation (blacks with end-stage renal disease being referred less often to the transplant list than whites) **(Fig. e4-3)**. Disparities have also been found in the utilization of cardiac diagnostic and therapeutic procedures (blacks being referred less often than whites for cardiac catheterization and bypass grafting), prescription of analgesia for pain control (blacks and Latinos receiving less pain medication than whites for long bone fractures and cancer), and surgical treatment of lung cancer (blacks receiving less curative surgery than whites for non-small-cell lung cancer), among others. Again, many of these disparities occurred even when variations in factors such as insurance status, income, age, comorbid conditions, and symptom expression are taken into account.

Figure e4-2

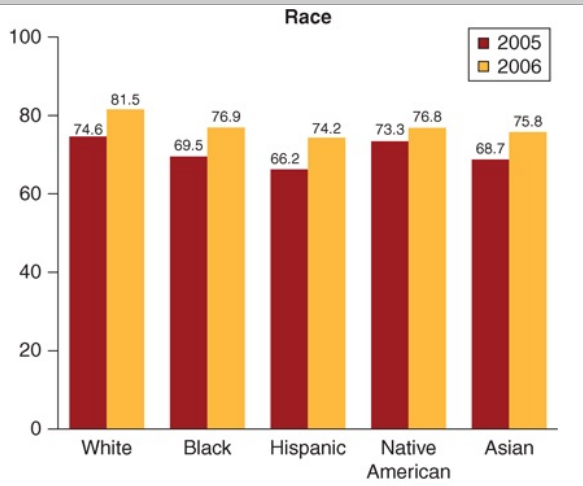

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Recommended hospital care received by Medicare patients with pneumonia, by race/ethnicity, 2006. Reference population is Medicare beneficiaries with pneumonia who are hospitalized. Composite is calculated by averaging the percentage of the population that received each of the five incorporated components of care. *(Adapted from Agency for Health Care Research and Quality: The 2008 National Health Care Disparities Report.)*

Figure e4-3

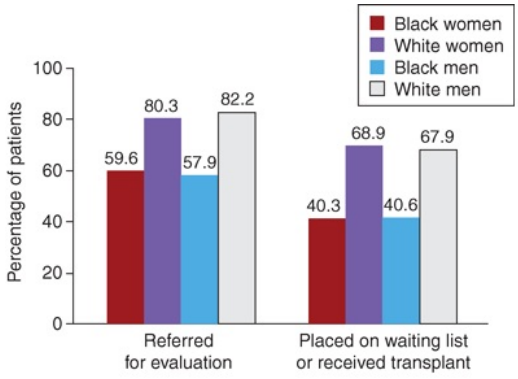

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Referral for evaluation at a transplantation center or placement on a waiting list or receipt of a renal transplantation within 18 months after the start of dialysis among patients who wanted a transplant, according to race and sex. Reference population is 239 black women, 280 white women, 271 black men, and 271 white men. Racial differences were statistically significant among the women and the men (p<.0001 for each comparison). *(From JZ Ayanian et al: N Engl J Med 341:1661, 1999.)*

Little progress has been made in addressing racial/ethnic disparities in cardiovascular procedures and other advanced surgical procedures, while some progress has been made in eliminating disparities in primary care process measures. Data from the National Registry of Myocardial Infarction found evidence of continued disparities in guideline-based admission, procedural, and discharge therapy use from 1994 to 2006. Compared to whites, black patients were less likely to receive percutaneous coronary intervention/coronary artery bypass grafting (PCI/CABG), a disparity that has shown little improvement since 1994. Further, compared to whites, black patients were less likely to receive lipid-lowering medications at discharge, with a gap that has widened since 1998 **(Fig. e4-4)**. The Centers for Disease Control and Prevention (CDC) analyzed national and state rates of total knee replacement (TKR) for Medicare enrollees for the period 2000 to 2006, stratified by sex, age group, and black or white race. TKR rates overall in the United States increased 58%, with similar increases among whites (61%) and blacks (56%). However, the TKR rate for blacks was 37% lower than the rate for whites in 2000 and 39% lower in 2006—no improvement and even a slight worsening of the disparity **(Fig. e4-5)**. Using data from enrollees in Medicare managed care plans, there is evidence for a narrowing in racial disparities between 1997 and 2003 in several "report card" preventive care measures such as mammography and glucose and cholesterol testing. However, racial disparities in more complex items such as glucose control in diabetics and cholesterol levels in patients after a heart attack had actually worsened.

Figure e4-4

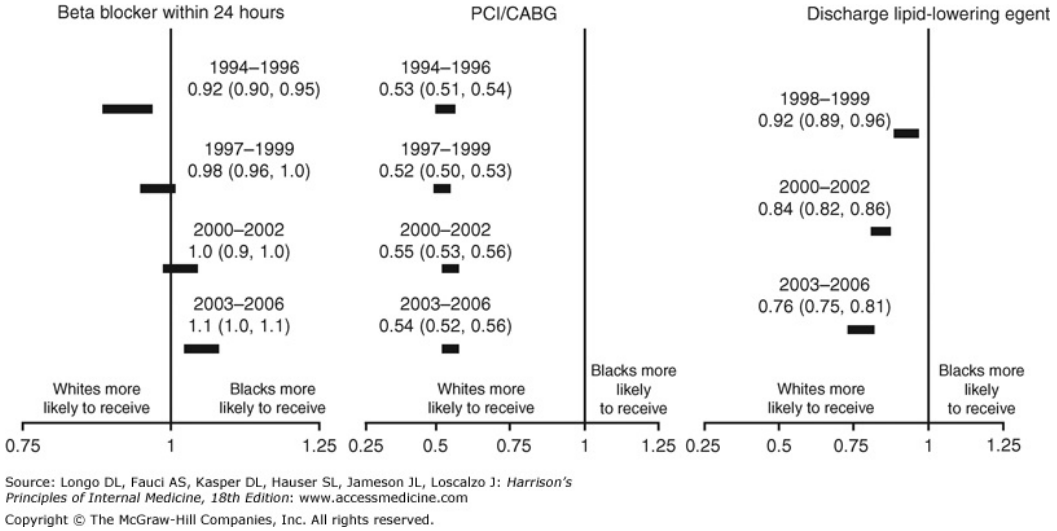

Racial differences in guideline-based treatments for acute myocardial infarction (AMI). Reference population is 2,515,106 patients with AMI admitted to U.S. hospitals between July 1990 and December 2006. CABG, coronary artery bypass grafting; PCI, percutaneous coronary intervention. (*From ED Peterson et al: Am Heart J 156:1045, 2008.)*

Figure e4-5

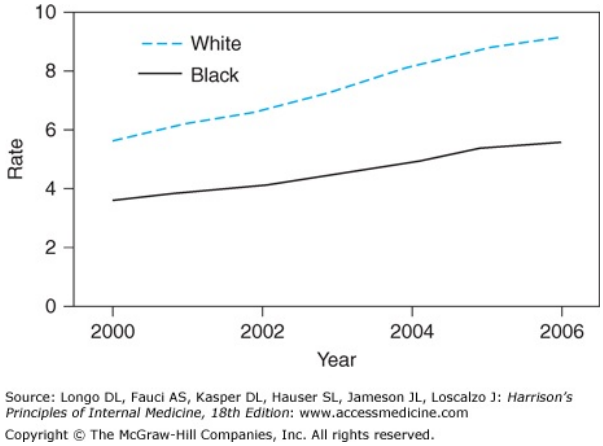

Racial trends in age-adjusted total knee replacement in Medicaid enrollees from 2000 to 2006. Reference population are Medicaid part A enrollees who are aged 65 or older and are not members of a managed care plan. *(From Centers for Disease Control and Prevention, 2009.)*

The sixth National Healthcare Disparities Report (NHDR), released by the Agency for Healthcare Research and Quality in January 2008, found that disparities in core measures of quality have improved little between 2001 and 2006. The report shows that for blacks, Asians, Native Americans/Alaskan Natives, Hispanics, and poor people, 60–80% of core quality measures, including measures of effectiveness, patient safety, and timeliness of care either stayed the same or worsened. While a smaller number of these measures have improved, in none of the measured areas have disparities been eliminated.

Root Causes for Racial/Ethnic Disparities in Health Care

The Institute of Medicine (IOM) report *Unequal Treatment*, released in March 2002, remains the preeminent study of the issue of racial and ethnic disparities in health care in the United States. The IOM was charged to assess the extent of racial/ethnic differences in health care that are not otherwise attributable to known factors such as access to care. To provide recommendations regarding interventions to eliminate health care disparities, the IOM studied health system, provider, and patient factors. The report found the following:

- Racial and ethnic disparities in health care exist and, because they are associated with worse health outcomes, are unacceptable.
- Racial and ethnic disparities in health care occur in the context of (1) broader historic and contemporary social and economic inequality and (2) evidence of persistent racial and ethnic discrimination in many sectors of American life.
- Many sources—including health systems, health care providers, patients, and utilization managers—may contribute to racial and ethnic disparities in health care.
- Bias, stereotyping, prejudice, and clinical uncertainty on the part of health care providers may contribute to racial and ethnic disparities in health care.
- A small number of studies suggest that certain patients may be more likely to refuse treatments, yet these refusal rates are generally small and do not fully explain health care disparities.

Unequal Treatment went on to identify a set of root causes that included the following, among others:

- Health system factors: These include issues related to the complexity of the health care system, the difficulty that minority patients may have in navigating this complex health system, and the lack of availability of interpreter services to assist patients with limited English proficiency. In addition, health care systems are generally ill-

prepared to identify and address disparities.

- Provider-level factors: These include issues related to the health care provider, including stereotyping, the impact of race/ethnicity on clinical decision making, and clinical uncertainty due to poor communication.
- Patient-level factors: These include patient's refusal of services, poor adherence to treatment, and delay in seeking care.

A more detailed analysis of these root causes is presented here.

HEALTH SYSTEM FACTORS

Health System Complexity

Even among those who are insured and educated, and who have a high degree of health literacy, navigating the health care system can be complicated and confusing. Some individuals, however, may be at higher risk for receiving substandard care because of their difficulty navigating the complexities of the U.S. health care system. These individuals may include those from cultures unfamiliar with the Western model of health care delivery, those with limited English proficiency, those with low health literacy, and those who are mistrustful of the health care system. These individuals may have difficulty knowing how and where to go for a referral to a specialist; how to prepare for a procedure such as a colonoscopy; or how to follow up on an abnormal test result such as a mammogram, for example. Since people of color in the United States tend to be overrepresented among the groups listed above, the inherent complexity of navigating our health care system has been seen as a root cause for racial/ethnic disparities in health care.

Other Health System Factors

Racial/ethnic disparities are due not only to differences in care provided within hospitals, but also result from where and from whom minorities receive their care (i.e., specific providers, geographic regions, or hospitals that are lower-performing on certain aspects of quality).

For example, one study showed that 25% of hospitals cared for 90% of black Medicare patients in the United States and these hospitals tended to have lower performance scores on certain quality measures than other hospitals. This being said, health systems are generally not well prepared to measure, report, and intervene to reduce disparities in care. Few hospitals or health plans stratify their quality data by race/ethnicity or language to measure disparities, and even fewer use data of this type to develop disparities-targeted interventions.

PROVIDER-LEVEL FACTORS

Provider-Patient Communication

Significant evidence highlights the impact of sociocultural factors, race, ethnicity, and limited English proficiency on health and clinical care. Health care professionals frequently care for diverse patient populations who present varied perspectives, values, beliefs, and behaviors regarding health and well-being. These include variations in recognition of symptoms, thresholds for seeking care, comprehension of management strategies, expectations of care (including preferences for or against diagnostic and therapeutic procedures), and adherence to preventive measures and medications. In addition, sociocultural differences between patient and provider influence communication and clinical decision making and are especially pertinent given evidence that clearly links provider-patient communication to improved patient satisfaction, adherence, and, subsequently, better health outcomes **(Fig. e4-6)**. Thus, when sociocultural differences between patient and provider are not appreciated, explored, understood, or communicated effectively in the medical encounter, patient dissatisfaction, poor adherence, poorer health outcomes, and racial/ethnic disparities in care may result.

Figure e4-6

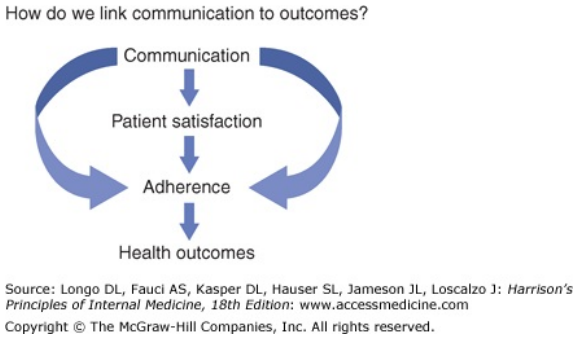

The link between effective communication, patient satisfaction, adherence, and health outcomes. *(From Institute of Medicine: Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC, National Academy Press, 2002.)*

A survey of 6722 Americans age 18 and older is particularly relevant given the important link between provider-patient communication and health outcomes. Whites, blacks, Hispanics/Latinos, and Asian Americans who had a medical visit in the past 2 years were asked whether they had trouble understanding their doctors; whether they felt the doctors did not listen; and whether they had medical questions they were afraid to ask. The survey found that 19% of all patients experienced one or more of these problems, yet whites experienced them 16% of the time, compared with 23% of the time for blacks, 33% for Hispanics/Latinos, and 27% for Asian Americans **(Fig. e4-7)**.

Figure e4-7

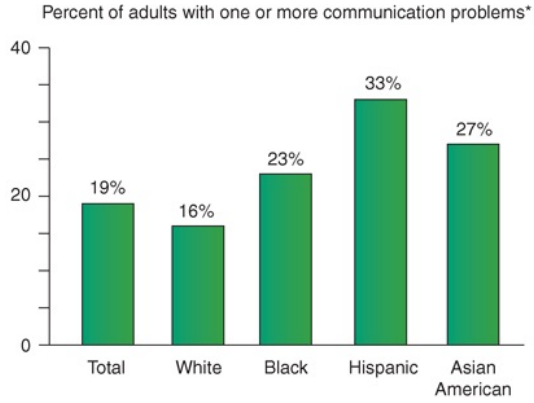

Communication difficulties with physicians, by race/ethnicity. Reference population is 6722 Americans age 18 and older who had a medical visit in the last 2 years and were asked whether they had trouble understanding their doctors, whether they felt the doctors did not listen, and whether they had medical questions they were afraid to ask. *(From Commonwealth Fund Health Care Quality Survey, 2001.)*

In addition, provider-patient communication without an interpreter, in the setting of even a minimal language barrier, is recognized as a major challenge to effective health care delivery. Spanish-speaking patients discharged from the emergency room are less likely than their English-speaking counterparts to understand their diagnosis, prescribed medications, special instructions, and plans for follow-up care; less likely to be satisfied with their care or willing to return if they have a problem; more likely to report problems with their care; and less satisfied with the patient-provider relationship. In addition, physicians who have access to trained interpreters report a significantly higher quality of patient-physician communication than physicians who used other methods. Hispanic patients with language-discordant physicians are more likely to omit medication, miss office appointments, and visit the emergency department for care. Communication issues related to discordant language disproportionately affect minorities and others with limited English proficiency and likely contribute to racial/ethnic disparities in health care.

Clinical Decision Making

Theory and research suggest that variations in clinical decision making may contribute to racial and ethnic disparities in health care. Two factors are central to this process: clinical uncertainty and stereotyping.

First, a doctor's decision-making process is nested in *clinical uncertainty*. Doctors depend on inferences about severity based on what they understand about illness and the information obtained from the patient. If the doctor is caring for a patient for whom he or she has difficulty understanding the symptoms and is less sure of the "signal"—the set of clues and indications that physicians rely on to make clinical decisions—the decision may be different than for another patient who presents with the exact same clinical condition. Given that the expression of symptoms may differ among and between cultural and racial groups, doctors—the overwhelming majority of whom are white—may understand symptoms best from patients of their own racial/ethnic groups. The consequence is that white patients may be treated differently from minority patients. Differences in clinical decisions from this mechanism can arise even when the doctor has the same regard for each patient (i.e., no prejudice).

Second, the literature on social cognitive theory highlights how natural tendencies to stereotype may influence clinical decision making. *Stereotyping* can be defined as the process by which people use social categories (e.g., race, gender, age) in acquiring, processing, and recalling information about others. Faced with enormous information loads and the need to make many decisions, people frequently subconsciously simplify the decision-making process and lessen cognitive effort by using "categories" or "stereotypes" that bundle information into groups or types that can be more quickly processed. Although functional, stereotyping can be systematically biased as people are automatically classified into social categories relating to dimensions such as *race*, *gender*, and *age*. Interestingly, people may not be aware of their attitudes, may not consciously endorse specific stereotypes, and paradoxically may consider themselves egalitarian and not prejudiced.

Stereotypes may be strongly influenced by the messages presented consciously and subconsciously in society. For instance, if the media and our social/professional contacts tend to present images of minorities as being less educated, violent, and non-adherent to health care recommendations, these impressions may generate stereotypes that unnaturally and unjustly impact clinical decision making. Thus, as signs of racism, classism, gender bias, and ageism are experienced —consciously or unconsciously in our society—stereotypes may be created that impact the way doctors manage patients from these groups. Based on training or practice location, doctors may develop certain perceptions about race/ethnicity, culture, and class that may evolve into stereotypes. For example, many medical students and residents are often trained —and minorities cared for—in academic health centers or public hospitals located in socioeconomically disadvantaged areas. As a result, doctors may begin to equate certain races and ethnicities with specific health beliefs and behaviors (e.g., "these patients" engage in risky behaviors, or "those patients" tend to be noncompliant) that are more associated with the social environment (e.g., poverty) than a patient's racial/ethnic background or cultural traditions. This "conditioning" phenomenon may also occur if doctors are faced with certain racial/ethnic patient groups who don't frequently choose aggressive forms of diagnostic or therapeutic interventions. The result over time may be that doctors begin to believe that "these patients" don't like invasive procedures, and thus they may not offer them as options. A wide range of studies have documented the potential for provider biases to contribute to racial/ethnic disparities in health care. For example, one study measured physicians' unconscious (or implicit) biases and showed that these were related to differences in decisions to provide thrombolysis for a hypothetical black or white patient with a myocardial infarction. (See Green in "Further Readings.")

To further compound this issue, doctors are commonly taught that their own personal characteristics (race, ethnicity,

socioeconomic status), the personal characteristics of the patient, and the clinical setting, should be excluded in the formulation of clinical decisions. Many nonmedical factors, however, ranging from the patient's physical appearance and background (race/ethnicity, gender, age, socioeconomic status, insurance status) to the doctor's background (specialty, level of training, clinical experience, age, gender, race/ethnicity) to the organizational setting in which medical care is delivered (location, form of compensation, performance expectations, incentives), may have as much influence on clinical decisions as the actual signs and symptoms of disease.

It is important to differentiate stereotyping from prejudice and discrimination. Prejudice is a conscious prejudgment of individuals that may lead to disparate treatment, and discrimination is conscious and intentional disparate treatment. All individuals stereotype subconsciously, yet if left unchecked, this may lead to lower quality of care for certain groups—such as minorities—who may be deemed less worthy of diagnostic or therapeutic procedures or resources. What is particularly salient is that stereotypes tend to be activated most in environments where the individual is stressed, multitasking, and under the time pressure—the hallmarks of the clinical encounter.

PATIENT-LEVEL FACTORS

Mistrust

Lack of trust has become a major concern for many health care institutions today. For example, an Institute of Medicine Report, *To Err Is Human: Building a Safer Health System* , documented alarming rates of medical errors and made patients feel vulnerable and less trustful of the U.S. health care system. The increased media and academic attention to problems of quality of care (and even disparities themselves) have clearly diminished trust in doctors and nurses.

Trust is a crucial element in the therapeutic alliance between patient and health care provider. It facilitates open communication and is directly correlated with adherence to physician recommendations and patient satisfaction. Patients who mistrust their health care providers are less satisfied with the care they receive, and mistrust of the health care system greatly affects patients' use of services. It can also result in inconsistent care, doctor-shopping, self-medicating, and an increased demand for referrals and diagnostic tests by patients.

Based on historic factors of discrimination, segregation, and medical experimentation, Black may be especially mistrustful of providers. The exploitation of blacks by the U.S. Public Health Service during the Tuskegee syphilis study from 1932 to 1972 left a legacy of mistrust that persists even today among this population. Other populations including Native Americans/Alaskan Natives, Hispanics/Latinos, and Asian Americans also harbor significant mistrust of the health care system. A national Kaiser Family Foundation survey of 3884 individuals found that 36% of Hispanics and 35% of black (compared with 15% of whites) felt they were treated unfairly in the health care system in the past based on their race and ethnicity. Perhaps even more alarming, 65% of blacks and 58% of Hispanics (compared with 22% of whites) were afraid of being treated unfairly in the future based on their race/ethnicity **(Fig. e4-8)**.

Figure e4-8

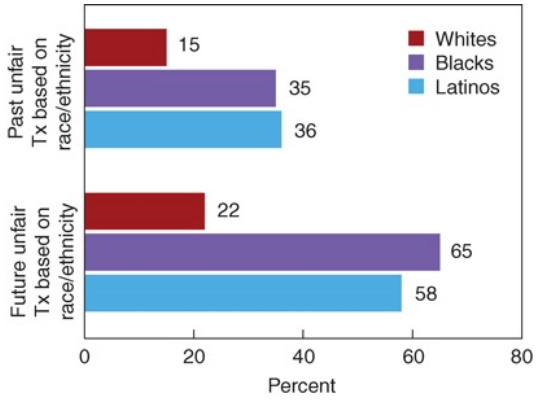

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Patient perspectives regarding unfair treatment, based on race/ethnicity. Reference population is 3884 individuals surveyed about how fairly they have been treated in the health care system in the past, and how fairly they feel they will be treated in the future based on their race/ethnicity. *(From Race, Ethnicity & Medical Care: A Survey of Public Perceptions and Experiences. Kaiser Family Foundation, 2005.)*

This mistrust may contribute to wariness in accepting or following recommendations, undergoing invasive procedures, or participating in clinical research, which may lead to misunderstanding and stereotypes by health professionals.

Key Recommendations to Address Racial/Ethnic Disparities in Health Care

The publication *Unequal Treatment* provides a series of recommendations to address racial and ethnic disparities in health care, focusing on a broad set of stakeholders. These include *health systems interventions*, *provider interventions*, *patient interventions*, and *general recommendations*, described in more detail below.

HEALTH SYSTEM INTERVENTIONS

1. *Collect and report health care access and utilization data by patient's race/ethnicity*

   Unequal Treatment found that the appropriate systems to track and monitor racial and ethnic disparities in health care are lacking, and there is less known about the disparities for minority groups (Hispanics, Asian Americans, Pacific Islanders, Native Americans, and Alaskan Natives) other than African Americans. For instance, only in the mid-1980s did the Medicare database begin to collect data on patient groups outside the standard categories of "white," "black," and "other." Federal, private, and state-supported data collection efforts are scattered and unsystematic, and many health care systems and hospitals still do not collect data on the race, ethnicity, or primary language of patients or enrollees. A survey of 501 U.S. hospitals by Regenstein and Sickler in 2006 found that 78% collected race information, 50% collected data on patient ethnicity, and 50% collected data on primary language. However, the information was not collected using standard categories or collection methods, so it is difficult to assess the accuracy of the data. Surveys by America's Health Insurance Plans (AHIP) in 2003 and 2006 showed the number of enrollees

who were in plans that collected race/ethnicity data of some type increased from 54 to 67%.

2. *Encourage the use of evidence-based guidelines and quality improvement*

   Unequal Treatment highlights the subjectivity of clinical decision making as a potential cause of racial and ethnic disparities in health care by describing how clinicians may offer different diagnostic and treatment options to different patients (consciously and unconsciously) based on their race or ethnicity, despite the existence of well-delineated practice guidelines. Therefore, the adoption and implementation of evidence-based guidelines broadly is a major recommendation to eliminate disparities. For instance, there now exist evidence-based guidelines for the management of diabetes, HIV/AIDS, cardiovascular diseases, cancer screening and management, and asthma—all areas where significant disparities exist. As part of ongoing quality improvement efforts, particular attention should be paid to the implementation of evidence-based guidelines for all patients, regardless of their race and ethnicity.

3. *Support the use of language interpretation services in the clinical setting*

   As described previously, health care systems that lack efficient and effective interpreter services can lead to patient dissatisfaction, poor comprehension and adherence, and ineffective/lower-quality care for patients with limited English proficiency. *Unequal Treatment*'s recommendation to support the use of interpretation services has clear implications for delivery of quality health care by improving doctors' ability to communicate effectively with patients with limited English proficiency.

4. *Increase the proportion of underrepresented minorities in the health care workforce*

   Data from the Association of American Medical Colleges (AAMC) indicate that in 2004, of the 72.4% of U.S. physicians whose race and ethnicity is known, Hispanics make up 2.8%, blacks 3.3%, and Native American and Alaskan Natives 0.3%. Data regarding the racial/ethnic composition of medical school faculty are no different, with minorities, excluding Asians, composing 7.5% nationally. Additionally, minority faculty in 2007 were more likely to be at or below the rank of assistant professor, while whites composed the highest proportion of full professors. Despite representing about 26% of the U.S. population (a number projected to almost double by 2050), minority students are still underrepresented in medical schools. In 2007, matriculants to U.S. medical school were 7.2% Latino, 6.4% African American, 0.2% Native Hawaiian or Other Pacific Islander, and 0.3% Native American or Alaskann Native. These numbers have decreased or maintained the same since 2006. It will be difficult to develop a diverse health care workforce that can meet the needs of an increasingly diverse population without dramatic change in the racial and ethnic composition of medical student bodies.

PROVIDER INTERVENTIONS

Integrate Cross-Cultural Education into the Training of All Health Care Professionals

The goal of cross-cultural education is to improve providers' ability to understand, communicate with, and care for patients from diverse backgrounds. Such education focuses on enhancing awareness of sociocultural influences on health beliefs and behaviors, and on building skills to understand and manage these factors in the medical encounter. Cross-cultural education includes curricula on health care disparities, how to use an interpreter, and how to effectively communicate and negotiate across cultures. These curricula can be incorporated into health professions training in medical schools, residency programs, and nursing schools, and as part of continuing education. Despite the importance of this area of education, as well as the attention it has attracted from medical education accreditation bodies, a national survey of senior resident physicians by Weissman and colleagues found that up to 28% felt unprepared to deal with cross-cultural issues, including caring for patients who have religious beliefs that may affect treatment, patients who use complementary medicine, patients with health beliefs at odds with Western medicine, patients with mistrust of the health care system, and new immigrants. Efforts to incorporate cross-cultural education into medical education will contribute to improving communication and to better quality of care for all patients.

Incorporate Teaching on the Impact of Race, Ethnicity, and Culture on Clinical Decision-Making

Unequal Treatment and more recent studies found that stereotyping by health care providers can lead to disparate treatment based on a patient's race or ethnicity. The Liaison Committee on Medical Education (LCME), which accredits medical schools, now has a directive that medical education should teach how a patient's race, ethnicity, and culture might unconsciously impact communication and clinical decision making.

PATIENT INTERVENTIONS

Educate Patients on How to Navigate the Health Care System and How to Be More Active in the Medical Encounter

Difficulty navigating the health care system and obtaining access to care can be a hindrance to all populations, particularly to minorities. Similarly, lack of empowerment or involvement in the medical encounter by minorities can be a barrier to care as well. Interventions should be used to increase patients' knowledge of how to best access care and participate in treatment decisions.

GENERAL RECOMMENDATIONS

Increase Awareness of Racial/Ethnic Disparities in Health Care

Efforts to raise awareness of racial/ethnic health care disparities have done little for the general public but have been fairly successful among physicians according to a Kaiser Family Foundation report. In 2006, nearly 6 in 10 people surveyed believed blacks received the same quality of care as whites, and 5 in 10 believed Latinos received the same quality care as whites. These estimates are similar to findings in a 1999 survey. Despite this lack of awareness, most believed that all Americans deserve quality care, regardless of their background. In contrast, the level of awareness among physicians has risen sharply. In 2002, the majority (69%) of physicians said that the health care system "rarely or never" treated people unfairly based on an individual's racial/ethnic background. In 2005, less than a quarter (24%) of physicians disagreed with the statement "minority patients generally receive lower quality care than white patients." Increasing awareness of racial and ethnic disparities among health care professionals and the public is an important first step in addressing disparities in health care. The ultimate goal is to generate discourse and mobilize action to address disparities in multiple areas, including at the level of health policy, health systems, and the community.

Conduct Further Research to Identify Sources of Disparities and Promising Interventions

While the literature that formed the basis of the findings and recommendations of *Unequal Treatment* provided significant evidence for racial and ethnic disparities, additional research is needed in several areas. First, most of the literature on disparities focuses on black-versus-white differences; much less is known about the experiences of other minority groups. Improving the ability to collect racial and ethnic patient data should facilitate this process, but in instances where those systems are not yet in place, racial and ethnic patient data may be collected prospectively in the setting of clinical or health services research to better understand disparities for other populations. Second, much of the literature on disparities to

date has focused on defining areas where they exist, but less has been done to identify the multiple factors that contribute to disparities, or to test interventions to address them. There is clearly a need for research that identifies promising practices and solutions to disparities.

Implications for Clinical Practice

Individual health care providers can do several things in the clinical encounter to address racial and ethnic disparities in health care. These approaches are discussed here.

BE AWARE THAT DISPARITIES EXIST

Increasing awareness of racial and ethnic disparities among health care professionals is an important first step in addressing disparities in health care. Only then can they be attuned to monitoring their behavior and clinical practice so as to ensure that all patients receive the highest quality of care, regardless of their race, ethnicity, or culture.

PRACTICE CULTURALLY COMPETENT CARE

Previous efforts in cultural competence have aimed to teach clinicians about the attitudes, values, beliefs, and behaviors of certain cultural groups—the key practice "do's and don'ts" for caring for "the Hispanic patient," or the "Asian Patient" for example. In certain situations, learning about a particular local community or cultural group can be helpful (following the principles of community-oriented primary care), but, when broadly and uncritically applied, this approach can also lead to stereotyping and oversimplification of culture without respect for its complexity.

Cultural competence has thus evolved from learning information and making assumptions about patients based on their background to focusing on the development of skills that follow the principles of patient-centered care. Patient centeredness encompasses the qualities of compassion, empathy, and responsiveness to the needs, values, and expressed preferences of the individual patient. Cultural competence aims to take this a step further, by expanding the repertoire of knowledge and skills classically defined as patient-centered to include those that are especially useful in cross-cultural interactions (but remain vital to all clinical encounters). This includes effectively using interpreter services, eliciting the patient's understanding of his or her condition, assessing decision-making preferences and the role of family, determining the patient's views about biomedicine versus complementary and alternative medicine, recognizing sexual and gender issues, and building trust. For example, while it is important to understand all patients' health beliefs, it may be particularly crucial to understand the health beliefs of those who come from a different culture or have a different health care experience. With the individual patient as teacher, one can adjust his or her practice style accordingly to meet the patient's specific needs.

AVOID STEREOTYPING

Several strategies can allow us to counteract, both systemically and individually, our normal tendency to stereotype. For example, when racially/ethnically/culturally/socially diverse teams are assembled (in which each member is given equal power) and are tasked to achieve a common goal, a sense of camaraderie develops and prevents the future development of stereotypes based on race/ethnicity, gender, culture, or class. Thus, we should aim to gain experiences working with, and learning from, a diverse set of colleagues. In addition, simply being aware of the operation of social cognitive factors allows one to actively "check" or "monitor" behavior. For instance, physicians can constantly ensure that they are offering the same things, in the same ways, to all patients. Understanding how we are susceptible to stereotyping—and how this may lead to disparities—is essential if we are to provide equitable, high-quality care to all patients.

WORK TO BUILD TRUST

Patient mistrust of the health care system and health care providers impacts multiple facets of the medical encounter, from decreased patient satisfaction to the delay of care. Although the historic legacy of discrimination can never be erased, several steps can be taken to build trust with patients and address disparities. First, providers must be aware that mistrust exists and is more prevalent among minority populations given the history of discrimination in the United States and other countries. Second, providers must reassure patients that they come first, that we will do everything in our power to ensure that they always get the best care possible, and that we will serve as their advocates. Third, interpersonal skills and communication techniques that demonstrate honesty, openness, compassion, and respect on the part of the health care provider are essential tools in dismantling mistrust. Finally, patients indicate that trust is built when there is shared, participatory decision making and the provider makes a concerted effort to understand the patient's background. By reframing the doctor-patient relationship as one of solidarity, the patient's sense of vulnerability can be transformed into one of trust. For the process of eliminating disparities to be successful, we must utilize trust-building interventions and strengthen the doctor-patient relationship.

CONCLUSION

The issue of racial and ethnic disparities in health care has gained national prominence, both with the release of the IOM report *Unequal Treatment* and with more recent articles that have confirmed their persistence and explored their root causes. Furthermore, another influential IOM report, *Crossing the Quality Chasm*, highlights the importance of equity —that there be no variations in quality of care by personal characteristics including race and ethnicity—as a central principle of quality. There are many obvious opportunities for interventions to eliminate racial and ethnic disparities in health care. Greater attention to addressing the root causes of disparities will improve the care provided to all patients, not just those who are racial and ethnic minorities.

ACKNOWLEDGMENT

The authors thank Marina Cervantes for her contributions to this chapter.

FURTHER READINGS

Agency for Health Care Research and Quality: *The 2008 National Health Care Disparities Report*. Rockville, MD, Agency for Health Care Research and Quality, 2009

America's Health Insurance Plans (AHIP): Collection and Use of Race and Ethnicity Data for Quality Improvement. AHIP-RWJF Survey of Health Insurance Plans, 2006

Andrulis DP: Access to care is the centerpiece in the elimination of socioeconomic disparities in health. Ann Intern Med 129:412, 1998[PMID: 9735070]

Association of American Medical Colleges (AAMC): Diversity in the Physician Workforce: Facts & Figures 2006, 2006

———: Diversity in Medical Education: Facts and Figures 2008, 2008

Ayanian JZ et al: Quality of care by race and gender for congestive heart failure and pneumonia. Med Care 37:1260, 1999[PMID: 10599607]

———: The effect of patients' preferences on racial differences in access to renal transplantation. N Engl J Med 341:1661, 1999

Baicker K et al: Who you are and where you live: How race and geography affect the treatment of medicare beneficiaries. Health Aff (Millwood) Suppl Web Exclusive: VAR 33, 2004

Berger JT: Culture and ethnicity in clinical care. Arch Intern Med 158:2085, 1998[PMID: 9801174]

Betancourt JR: Cultural competence—Marginal or mainstream movement? N Engl J Med 351:953, 2004[PMID: 15342800]

——— et al: Hypertension in multicultural and minority populations: Linking communication to compliance. Curr Hypertens Rep 1:482, 1999

Carrasquillo O et al: Impact of language barriers on patient satisfaction in an emergency department. J Gen Intern Med 14:82, 1999[PMID: 10051778]

Carrillo JE et al: Cross-cultural primary care: A patient-based approach. Ann Intern Med 130:829, 1999[PMID: 10366373]

Centers for Disease Control and Prevention (CDC): Racial disparities in total knee replacement among Medicare enrollees —United States, 2000–2006. MMWR Morb Mortal Wkly Rep 20:133, 2009

Flores G et al: The health of Latino children: Urgent priorities, unanswered questions, and a research agenda. J Am Med Assoc 288:82, 2002[PMID: 12090866]

Gamble V: Under the shadow of Tuskegee: African Americans and health care. Am J Public Health 87:1773, 1997[PMID: 9366634]

Green AR et al: Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. J Gen Intern Med 22:1231, 2007[PMID: 17594129]

Institute of Medicine: *Health Literacy: A Prescription to End Confusion*. Washington, DC, National Academy Press, 2004

———: *To Err Is Human: Building a Safer Health System* . Washington, DC, National Academy Press, 2000

———: *Crossing the Quality Chasm: A New Health System for the 21st Century* . Washington, DC, National Academy Press, 2001

———: *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care* . Washington, DC, National Academy Press, 2002

———: *In the Nation's Compelling Interest: Ensuring Diversity in the Health Care Workforce* . Washington, DC, National Academy Press, 2004

Jha AK et al: Concentration and quality of hospitals that care for elderly black patients. Arch Intern Med 167:1177, 2007[PMID: 17563027]

Kaiser Family Foundation: Race Ethnicity & Medical Care: Survey of Public Perceptions and Expectations, 1999

———: National Survey of Physicians, Part 1: Doctors on Disparities in Medical Care, 2002

Perez-Stable EJ et al: The effects of ethnicity and language on medical outcomes of patients with hypertension or diabetes. Med Care 35:1212, 1997[PMID: 9413309]

Peterson ED et al: Trends in quality of care for patients with acute myocardial infarction in the National Registry of Myocardial Infarction from 1990 to 2006. Am Heart J 156:1045, 2008[PMID: 19032998]

Rathore SS et al: The effects of patient sex and race on medical students' ratings of quality of life. Am J Med 108:561, 2000[PMID: 10806285]

Regenstein M, Sickler D: *Race, Ethnicity and Language of Patients: Hospital Practices Regarding Collection of Information to Address Disparities in Health Care*. Washington, DC: National Public Health and Hospital Institute, 2006

Schulman KA et al: The effect of race and sex on physicians' recommendations for cardiac catheterization. N Engl J Med 340:618, 1999[PMID: 10029647]

Skinner J et al: Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 12:2634, 2005

Todd KH et al: Ethnicity as a risk factor for inadequate emergency department analgesia. J Am Med Assoc 269:1537, 1993[PMID: 8445817]

Trivedi AN et al: Trends in the quality of care and racial disparities in Medicare managed care. N Engl J Med 353:692, 2005[PMID: 16107622]

U.S. Census Bureau: *Statistical Abstract of the United States: 2010,* 129th ed. Washington, DC, 2009; Available at *http://www.census.gov/statab*

Vaccarino V et al: Sex and race differences in the management of acute myocardial infarction, 1994–2002. N Engl J Med 353:671, 2005[PMID: 16107620]

van Ryn M, Burke J: The effect of patient race and socio-economic status on physician's perceptions of patients. Soc Sci Med 50:813, 2000

———, Fu SS: Paved with good intentions: Do public health and human service providers contribute to racial/ethnic disparities in health? Am J Public Health 93:248, 2003

Weissman JS et al: Resident physicians' preparedness to provide cross-cultural care. J Am Med Assoc 294:1058, 2005[PMID: 16145026]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ETHICAL ISSUES IN CLINICAL MEDICINE: INTRODUCTION

Physicians frequently confront ethical issues in clinical practice that are perplexing, time-consuming, and emotionally draining. Experience, common sense, and simply being a good person do not guarantee that physicians can identify or resolve ethical dilemmas. Knowledge about common ethical dilemmas is also essential.

FUNDAMENTAL ETHICAL GUIDELINES

Physicians should follow two fundamental but frequently conflicting ethical guidelines: respect patient autonomy and act in the patient's best interests.

Respecting Patient Autonomy

Treating patients with respect requires doctors to accept the medical decisions of persons who are informed and acting freely. Individuals place different values on health, medical care, and risk. In most clinical settings, different goals and approaches are possible, outcomes are uncertain, and an intervention may cause both benefits and harms. Thus competent, informed patients may refuse recommended interventions and choose among reasonable alternatives.

INFORMED CONSENT

For patients to make informed decisions, physicians need to discuss with them the nature of the proposed care, the alternatives, the risks and benefits of each, and the likely consequences, and obtain the patient's agreement to care. Informed consent involves more than obtaining signatures on consent forms. Physicians need to educate patients, answer questions, make recommendations, and help them deliberate. Patients can be overwhelmed with medical jargon, needlessly complicated explanations, or too much information at once.

NONDISCLOSURE OF INFORMATION

Physicians may consider withholding a serious diagnosis, misrepresenting it, or limiting discussions of prognosis or risks because they fear that a patient will develop severe anxiety or depression or refuse needed care. Generally, physicians should provide relevant information, while adjusting the pace of disclosure, offering empathy and hope, and helping patients cope with bad news.

In many cultures, patients traditionally are not told of a diagnosis of cancer or of other serious illness because such disclosure is believed to cause patients to suffer, while withholding information is believed to promote serenity, security, and hope. Patients should not be forced to receive information against their will, even in the name of promoting informed decisions. However, many individuals want to know their diagnosis and prognosis, even if they are terminally ill. Physicians, therefore, should ask patients how they want health care decisions to be made, adding that they usually provide information and make decisions together with patients, while offering patients the option not to receive information or to turn decision-making over to someone else.

EMERGENCY CARE

Informed consent is not required when patients cannot give consent and when delay of treatment would place their lives or health in peril. People are presumed to want such emergency care, unless they have previously indicated otherwise.

FUTILE INTERVENTIONS

Autonomy does not entitle patients to insist on whatever care they want. Physicians are not obligated to provide futile interventions that have no physiologic rationale or have already failed. For example, cardiopulmonary resuscitation would be futile in a patient with progressive hypotension despite maximal therapy. But physicians should be wary of using the term *futile* in looser senses to justify unilateral decisions to forego interventions when they believe that the probability of success is too low, no worthwhile goals can be achieved, the patient's quality of life is unacceptable, or the costs are too high. Such looser usages of the term are problematic because they may be inconsistent and mask important value judgments. Several states have laws allowing physicians not to provide interventions that they deem futile or not medically indicated, provided they follow procedures such as obtaining the concurrence of the hospital ethics committee and giving the patient an opportunity to be transferred to another hospital.

MAINTAINING CONFIDENTIALITY

Confidentiality respects patients' autonomy and privacy, encourages them to seek treatment and discuss their problems candidly, and prevents discrimination. However, maintaining confidentiality is not an absolute rule. Confidentiality may be overridden in certain situations to prevent serious harm to third parties or to the patient. The law may require physicians to override confidentiality in order to protect third parties, as with reporting of tuberculosis and syphilis. In other situations, medical providers have a legal duty to report victims of elder abuse, child abuse, and domestic violence. These exceptions to confidentiality are justified because the risk is serious and probable, there are no less-restrictive measures to avert risk, the adverse effects of overriding confidentiality are minimized, and these adverse effects are deemed acceptable by society.

The Health Insurance Portability and Accountability Act (HIPAA) health privacy regulations have heightened awareness of the importance of confidentiality. These regulations are not meant to inhibit transmission of information needed for treatment: disclosure of patient information to other health care providers for the purposes of treatment without having the patient sign an authorization form is permissible.

AVOIDING DECEPTION

Health care providers sometimes consider using lies or deception in order to protect the patient from bad news or to obtain benefits for the patient. Lying refers to statements that the speaker knows are false and that are intended to mislead the listener. Deception, which is broader, includes statements and actions that are intended to mislead the listener, whether or not they are literally true. For example, the health care provider may tell a patient that she has a "small growth" so that she does not think she has cancer. Or the provider may complete and sign a form for a patient to get a bus pass, even though he does not meet criteria for physical disability. Although such deception may be motivated by a desire to help the patient, it is ethically problematic. The person who is deceived cannot make informed decisions if he or she receives misleading information. Furthermore, deception undermines physicians' credibility and trustworthiness.

Acting in the Best Interests of Patients

The guideline of beneficence requires physicians to act for the patient's benefit. Laypeople do not possess medical expertise and may be vulnerable because of their illness. They justifiably rely on physicians to provide sound advice and to promote their well-being. Physicians encourage such trust. Hence, physicians have a fiduciary duty to act in the best interests of the patient, which should prevail over physicians' self-interest or the interests of third parties, such as hospitals or insurers. These fiduciary obligations of physicians contrast sharply with business relationships, which are characterized by "let the buyer beware," not by trust and reliance. The guideline of "do no harm" forbids physicians from providing ineffective interventions or acting without due care. This precept, while often cited, provides only limited guidance, because many beneficial interventions also have serious risks.

Conflicts between Beneficence and Autonomy

Patients' refusals of care may thwart their own goals or cause them serious harm. For example, a young man with asthma may refuse mechanical ventilation for reversible respiratory failure. Simply to accept such refusals, in the name of respecting autonomy, seems morally constricted. Physicians can elicit patients' expectations and concerns, correct misunderstandings, and try to persuade them to accept beneficial therapies. If disagreements persist after discussions, the patient's informed choices and view of his or her best interests should prevail. While refusing recommended care does not render a patient incompetent, it may lead the physician to probe further to ensure that the patient is able to make informed decisions.

Justice

The term *justice* is used in a general sense to mean fairness: people should receive what they deserve. In addition, it is important to act consistently in cases that are similar in ethically relevant ways. Otherwise, decisions would be arbitrary, biased, and unfair. Justice forbids discrimination in health care based on race, religion, or gender and supports a moral right to health care, with access based on medical need rather than ability to pay.

PATIENTS WHO LACK DECISION-MAKING CAPACITY

Patients may not be able to make informed decisions because of unconsciousness, dementia, delirium, or other conditions. Physicians should ask two questions regarding such patients: Who is the appropriate surrogate? What would the patient want done?

Assessing Capacity to Make Medical Decisions

All adults are considered legally competent unless declared incompetent by a court. In practice, physicians usually determine that patients lack the capacity to make health care decisions and arrange for surrogates to make them, without involving the courts. By definition, competent patients can express a choice and appreciate the medical situation; the nature of the proposed care; the alternatives; and the risks, benefits, and consequences of each. Their choices should be consistent

with their values and should not result from delusions or hallucinations. Psychiatrists may help in difficult cases because they are skilled at interviewing mentally impaired patients and can identify treatable depression or psychosis. When impairments are fluctuating or reversible, decisions should be postponed if possible until the patient recovers decision-making capacity.

Choice of Surrogate

If a patient lacks decision-making capacity, physicians routinely ask family members to serve as surrogates. Most patients want their family members to be surrogates, and family members generally know the patient's preferences and have the patient's best interests at heart. Patients may designate a particular individual to serve as proxy; such choices should be respected. Some states have established a prioritized list of which relative may serve as surrogate if the patient has not designated a proxy.

Standards for Surrogate Decision-Making

ADVANCE DIRECTIVES

These are statements by competent patients to direct care if they lose decision-making capacity. They may indicate (1) what interventions they would refuse or accept or (2) who should serve as surrogate. Following the patient's advance directives, the surrogate respects the patient's autonomy.

Oral conversations are the most frequent form of advance directives. While such conversations are customarily followed in clinical practice, casual or vague comments may not be trustworthy. Living wills direct physicians to forego or provide life-sustaining interventions if the patient develops a terminal condition or persistent vegetative state. Generally patients may refuse only interventions that "merely prolong the process of dying."

A health care proxy is someone appointed by the patient to make health care decisions if he or she loses decision-making capacity. It is more flexible and comprehensive than the living will, applying whenever the patient is unable to make decisions.

Physicians can encourage patients to carry out advance care planning, which might include clarifying values and priorities, identifying unfinished goals, and addressing spiritual concerns. Such discussions often facilitate completion of advance directives. In discussions with patients, physicians can ensure that advance directives are informed, up-to-date, and address likely clinical scenarios. Such discussions are best carried out in the ambulatory setting. The federal Patient Self-Determination Act requires hospitals and health maintenance organizations to inform patients of their right to make health care decisions and to provide advance directives.

SUBSTITUTED JUDGMENT

In the absence of clear advance directives, surrogates and physicians should try to decide as the patient would under the circumstances, using all information that they know about the patient. While such substituted judgments try to respect the patient's values, they may be speculative or inaccurate. A surrogate may be mistaken about the patient's preferences, particularly when they have not been discussed explicitly.

BEST INTERESTS

When the patient's preferences are unclear or unknown, decisions should be based on the patient's best interests. Patients generally take into account the quality of life as well as the duration of life when making decisions for themselves. It is understandable that surrogates would also consider quality of life of patients who lack decision-making capacity. Judgments about quality of life are appropriate if they reflect the patient's own values. Bias or discrimination may occur, however, if others project their values onto the patient or weigh the perceived social worth of the patient. Most patients with chronic illness rate their quality of life higher than their family members and physicians do.

LEGAL ISSUES

Physicians need to know pertinent state laws regarding patients who lack decision-making capacity. A few state courts allow doctors to forego life-sustaining interventions only if patients have provided written advance directives or very specific oral ones.

DISAGREEMENTS

Disagreements may occur among potential surrogates or between the physician and surrogate. Physicians can remind everyone to base decisions on what the patient would want, not what they would want for themselves. Consultation with the hospital ethics committee or with another physician often helps resolve disputes. Such consultation is also helpful when patients have no surrogate and no advance directives. The courts should be used only as a last resort when disagreements cannot be resolved in the clinical setting.

DECISIONS ABOUT LIFE-SUSTAINING INTERVENTIONS

Although medical technology can save lives, it can also prolong the process of dying. Competent, informed patients may refuse life-sustaining interventions. When patients lack decision-making

capacity, such interventions may also be withheld on the basis of advance directives or decisions by appropriate surrogates. Courts have ruled that foregoing life-sustaining interventions is neither suicide nor murder.

Misleading Distinctions

People commonly draw distinctions that are intuitively plausible but prove untenable on closer analysis.

EXTRAORDINARY AND ORDINARY CARE

Some physicians are willing to forego "extraordinary" or "heroic" interventions, such as surgery, mechanical ventilation, or renal dialysis but insist on providing "ordinary" ones, such as antibiotics, IV fluids, or feeding tubes. However, this distinction is not logical because all medical interventions have both risks and benefits. Any intervention may be withheld, if the burdens for the individual patient outweigh the benefits.

WITHDRAWING AND WITHHOLDING INTERVENTIONS

Many health care providers find it more difficult to discontinue interventions than to withhold them in the first place. Although such emotions need to be acknowledged, there is no logical distinction between the two acts. Reasons that justify withholding interventions, such as refusal by patients or surrogates, also justify withdrawing them. In addition, after an intervention has been started, new data may indicate that it is no longer appropriate. The intervention may prove unsuccessful, or it may be learned that the patient did not want the intervention. If interventions could never be discontinued, patients and surrogates might not even attempt treatments that might prove beneficial.

Do Not Attempt Resuscitation (Dnar) Orders

When a patient suffers a cardiopulmonary arrest, cardiopulmonary resuscitation (CPR) is initiated unless a DNAR order has been made. Although CPR can restore people to vigorous health, it can also disrupt a peaceful death. After CPR is attempted on a general hospital service, only 14% of patients survive to discharge, and even fewer in certain subgroups. DNAR orders are appropriate if the patient or surrogate requests them or if CPR would be futile. To prevent misunderstandings, physicians should write DNAR orders and the reasons for them in the medical record. "Slow" or "show" codes that merely appear to provide CPR are deceptive and therefore unacceptable. Although a DNAR order signifies only that CPR will be withheld, the reasons that justify DNAR orders may lead to a reconsideration of other plans for care. Many hospitals have standardized order forms that specify whether other life-sustaining interventions, such as intubation and mechanical ventilation, are to be administered or withheld.

Assisted Suicide and Active Euthanasia

Proponents of these controversial acts believe that competent, terminally ill patients should have control over the end of life and that physicians should relieve refractory suffering. Opponents assert that such actions violate the sanctity of life, that suffering can generally be relieved, that abuses are inevitable, and that such actions are outside the physician's proper role. These actions are illegal throughout the United States, except that physician-assisted suicide is legal in Oregon and Washington under restricted circumstances. Whatever their personal views, physicians should respond to patients' inquiries about these actions with compassion and concern. Physicians should elicit and address any underlying problems, such as physical symptoms, loss of control, or depression. Often, additional efforts to relieve distress are successful, and after this is done patients generally withdraw their requests for these acts.

Care of Dying Patients

Patients often suffer unrelieved pain and other symptoms during their final days of life. Physicians may hesitate to order high doses of narcotics and sedatives, fearing they will hasten death. Relieving distressing symptoms in terminal illness and when patients forego life-sustaining interventions such as mechanical ventilation enhances patient comfort and dignity. If lower doses of narcotics and sedatives have failed to relieve suffering, increasing the dose to levels that might suppress respiratory drive or lower blood pressure is ethically appropriate because the physician's intention is to relieve suffering, not hasten death. Such palliative sedation is distinguished ethically and legally from active euthanasia, which is administering a lethal dose with the intention of ending the patient's life. Physicians can also relieve suffering by spending time with dying patients, listening to them, and attending to their psychological distress.

CONFLICTS OF INTEREST

Acting in the patient's best interests may conflict with the physician's self-interest or the interests of third parties such as insurers or hospitals. The ethical ideal is to keep the patient's interests paramount. Even the appearance of a conflict of interest may undermine trust in the profession.

Financial Incentives

The 2010 health care reform law includes pilot programs that offer physicians financial incentives to

improve the quality and efficiency of care. There are concerns that pay-for-performance incentives may lead physicians to avoid sicker, more complicated patients or to focus on benchmarked interventions even if they are not in the best interests of an individual patient. In contrast, physicians have incentives to provide interventions regardless of the probability or magnitude of benefit when they receive fee-for-service reimbursement or when they refer patients to laboratory or imaging facilities in which they have a financial stake. Regardless of financial incentives, physicians should recommend available care that is in the patient's best interests, no more and no less.

Relationships with Pharmaceutical and Device Companies

Financial relationships between physicians and industry are under increasing scrutiny. Gifts from drug and device companies may create an inappropriate risk of undue influence, induce subconscious feelings of reciprocity, impair public trust, and increase the cost of health care. Policies at many academic medical centers and companies have eliminated pens, notepads, and meals to physicians. Under new federal sunshine requirements, companies must disclose publicly the names of physicians to whom they have made payments together with the amount of payment. It will be a challenge to structure such disclosure to distinguish between payments for scientific consulting, contracts, and grants—which are consistent with professional and academic missions and need to be encouraged—from promotional speaking and consulting whose goal is to increase sales of company products.

Occupational Risks

Some health care workers, fearing fatal occupational infections, refuse to care for persons with HIV infection or multidrug-resistant tuberculosis. Such fears about personal safety need to be acknowledged, and health care institutions should reduce occupational risk by providing proper training, protective equipment, and supervision. To fulfill their mission of helping patients, physicians should provide appropriate care within their clinical expertise, despite some personal risk.

Medical Errors

Errors are inevitable in clinical medicine. They may cause serious harm to patients or result in substantial changes in management. Physicians and students may fear that disclosing such errors could damage their careers. Without disclosure, however, patients cannot understand their clinical situation or make informed choices about subsequent care. Furthermore, patients are often outraged when physicians do not acknowledge and apologize for errors. Several states have enacted laws that allow physicians to say they are sorry for errors without increasing legal liability. Similarly, unless attending physicians are informed of trainees' errors, they cannot provide optimal care and help trainees learn from mistakes.

Learning Clinical Skills

Learning clinical medicine, particularly learning to perform invasive procedures, may present inconvenience or risk to patients. To ensure patient cooperation, students may be introduced as physicians or patients may not be told that trainees will be performing procedures. Such misrepresentation undermines trust, may lead to more elaborate deception, and makes it difficult for patients to make informed choices about their care. Patients should be told who is providing care, what benefits and burdens can be attributed to trainees, and how trainees are supervised. Most patients, when informed, allow trainees to play an active role in their care.

Impaired Physicians

Physicians may hesitate to intervene when colleagues impaired by alcohol abuse, drug abuse, or psychiatric or medical illness place patients at risk. However, society relies on physicians to regulate themselves. If colleagues of an impaired physician do not take steps to protect patients, no one else may be in a position to do so.

Conflicts for Trainees

Medical students and residents may fear that they will receive poor grades or evaluations if they act on the patient's behalf by disclosing mistakes, avoiding misrepresentation of their role, and reporting impaired colleagues. Discussing such dilemmas with more senior physicians can help trainees check their interpretation of the situation and obtain advice and assistance.

ALLOCATING RESOURCES JUSTLY

Access to needed care still remains a moral aspiration in the United States, even after the passage of health care reform laws in 2010. Physicians caring for underserved populations must act ethically in a health care system that has serious ethical shortcomings in access to and quality of care. Some patients with a clear need for medical care cannot pay for medications, tests, or hospitalizations, or the insurer may deny coverage. If this occurs, physicians should advocate for patients, trying to help them obtain needed care. Doctors might consider, or patients might request, using lying or deception to help them gain such benefits. However, avoiding deception is a basic ethical guideline that sets limits on advocating for patients. Allocation of health care resources is unavoidable because resources are limited. Ideally, allocation decisions should be made as public policy, with physician input. Allocation of

resources at the bedside on an ad hoc basis is problematic because it may be inconsistent, unfair, and ineffective. At the bedside, physicians generally should act as patient advocates within constraints set by society, reasonable insurance coverage, and evidence-based practice. For example, a patient's insurer may have a higher copayment for nonformulary drugs. It is reasonable for physicians to advocate for nonformulary drugs only if there are compelling reasons for an exception, as when the formulary drugs are ineffective or not tolerated.

ASSISTANCE WITH ETHICAL ISSUES

Discussing perplexing ethical issues with other members of the health care team, colleagues, or the hospital ethics committee often clarifies issues and suggests ways to improve communication and to deal with strong emotions. When struggling with difficult ethical issues, physicians may need to reevaluate their basic convictions, tolerate uncertainty, and maintain their integrity while respecting the opinions of others.

FURTHER READINGS

Beauchamp TL, Childress JF: *Principles of Biomedical Ethics,* 6th ed. New York, Oxford University Press, 2008

Chiong W: Justifying patient risks associated with medical education. JAMA 298:1046, 2007[PMID: 17785649]

Curtis JR, Vincent JL: Ethics and end-of-life care for adults in the intensive care unit. Lancet 376:1347, 2010[PMID: 20934213]

Gallagher TH: A 62-year-old woman with skin cancer who experienced wrong-site surgery: Review of medical error. JAMA 302:669, 2009[PMID: 19584321]

Lo B: *Resolving Ethical Dilemmas: A Guide for Clinicians,* 4th ed. Baltimore, Lippincott Williams & Wilkins, 2009

———Field MJ: *Conflicts of Interest in Medical Research, Education, and Clinical Practice* . Washington, DC, National Academies Press, 2009

Snyder L, Leffler C: Ethics manual: Fifth edition. Ann Intern Med 142:560, 2005[PMID: 15809467]

Truog RD: Counterpoint: The Texas Advance Directives Act is ethically flawed: Medical futility disputes must be resolved by a fair process. Chest 136:968, 2009[PMID: 19809046]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NEOPLASIA DURING PREGNANCY: INTRODUCTION

Cancer develops during ~1 in every 1000 pregnancies. Of all the cancers that occur in women, less than 1% occur in pregnant women. The four cancers most commonly developing during pregnancy are cervical cancer, breast cancer, melanoma, and lymphomas (particularly Hodgkin's lymphoma); however, virtually every form of cancer has been reported in pregnant women (Table e6-1). In addition to cancers developing in other organs of the mother, gestational trophoblastic tumors can arise from the placenta. The problem of cancer in a pregnant woman is complex. One must take into account the possible influence of the pregnancy on the natural history of the cancer, the effects of the diagnostic and staging procedures, and the treatments of the cancer on both the mother and the developing fetus. Theese issues may lead to dilemmas: what is best for the mother may be harmful to the fetus, and what is best for the fetus may be harmful to the mother.

Table e6-1 Incidence of Malignant Tumors During Gestation

Tumor Type	Incidence per 10,000 Pregnancies[a]	% of Cases[b]
Breast cancer	1–3	25%
Cervical cancer	1.2–4.5	25%
Thyroid cancer	1.2	15%
Hodgkin's disease	1.6	10%
Melanoma	1–2.6	8%
Ovarian cancer	0.8	2%
All sites	10	100%

[a] These are estimates based on extrapolations from a review of more than 3 million pregnancies (Smith LH et al: Am J Obstet Gynecol 184: 1504, 2001).

[b] Based on accumulating case reports from the literature; the precision of these data is not high.

Another complicating issue in women who develop cancer during pregnancy is that many of the early symptoms of cancer are ignored in pregnant women. The many changes in a woman's body during pregnancy dull one's senses to changes that may be related to an underlying disease rather than the pregnancy. Thus, many cancers that occur in pregnancy present in advanced stages.

As a general rule, one should assume that no diagnostic or therapeutic intervention is safe in the first trimester of pregnancy other than surgery. If the mother develops life-threatening complications during the first trimester that require radiation therapy or systemic chemotherapy, and these interventions cannot be safely delayed, a recommendation should be made for an abortion. Indeed, radiation, even in the form of diagnostic radiography, should be avoided throughout pregnancy. No exposure to radiation is safe, and efforts to shield the fetus with barriers placed on the abdomen cannot block internal scatter radiation. It is safest to omit radiation exposure of any kind. Fortunately, its use is seldom an essential component of treatment before delivery.

Chemotherapy exposure is also to be avoided, if at all possible. It should never be given in the first trimester; a variety of single agents and combinations have been given in the second and third trimesters, without a high frequency of catastrophic effects to the pregnancy or the fetus, but data on safety are sparse. Maternal factors that may influence the pharmacology of chemotherapeutic agents include the 50% increase in plasma volume, altered absorption and protein binding, increased glomerular filtration rate, increased hepatic mixed function oxidase activity, and third space created by amniotic fluid. The fetus is protected from some agents by placental expression of drug efflux pumps, but decreased fetal hepatic mixed function oxidase and glucuronidation activity may prolong the half-life of agents that do cross the placenta. A database on the risks associated with individual chemotherapy agents is available on the Internet (www.motherisk.org).

The optimal management strategies have not been developed based on prospective clinical trials. Instead, a guiding principle has been to delay therapeutic interventions until as late as possible during the pregnancy. Delivery is recommended at 32 weeks. By and large, this approach minimizes exposure

of the child to noxious cancer treatments, spares the mother complications of pregnancy, and is generally accomplished without an adverse impact on treatment outcome. Pregnancy appears to have little or no impact on the natural history of malignancies, despite the hormonal influences. Spread of the mother's cancer to the fetus (so-called vertical transmission) is exceedingly rare.

CERVICAL CANCER DURING PREGNANCY

The incidence of cervical cancer in pregnant women is roughly comparable to that of age-matched controls who are not pregnant. Invasive cervical cancer develops at a rate of about 0.45 in 1000 live births and carcinoma in situ is seen in 1 in 750 pregnancies. About 1% of women diagnosed with cervical cancer are pregnant at the time of diagnosis. Early signs of cervical cancer include vaginal spotting or discharge, pain, and postcoital bleeding that are also common features of pregnancy. Early visual changes in the cervix related to invasive cancer can be mistaken for cervical decidualization or ectropion (columnar epithelium on the cervix) due to pregnancy. Women diagnosed with cervical cancer during pregnancy report having had symptoms for 4.5 months on average.

Human papillomavirus (HPV) types 16 and 18 account for about 70% of cervical cancer. The rate of carriage of these serotypes can be reduced with the use of vaccination before exposure. Screening is recommended at the first prenatal visit and 6 weeks postpartum. The rate of cytologic abnormalities on Pap smear in pregnant women is about 5–8% and is not much different than the rate in nonpregnant women of the same age. Consensus guidelines dictate that specific tests are indicated based on the level of atypia seen on Pap smear. Atypical squamous cells of unknown significance (ASCUS) generally trigger HPV testing with colposcopy reserved for the subset of women with a high-risk HPV-type infection. By contrast, the presence of dysplasia is considered an indication for colposcopy regardless of HPV type. Women with either low- or high-grade squamous intraepithelial lesions (LSIL or HSIL) and HIV-infected women with ASCUS are recommended for colposcopy.

At colposcopy, any areas suspicious for invasive disease are biopsied. However, endocervical curettage is contraindicated in pregnant patients. The only indication for therapy of cervical neoplasia in pregnant women is the documentation of invasive cancer. Accordingly, some physicians defer colposcopy in pregnant women until six weeks post-partum unless they are at high risk for invasive disease. Cervical intraepithelial neoplasia has a low risk of progression to invasive cancer during pregnancy (~0.4%) and many such lesions (36–70%) regress spontaneously postpartum. If invasive disease is suspected at colposcopy and the pregnancy is between 16 and 20 weeks, a cone biopsy may be performed to make the diagnosis; however, the procedure is associated with bleeding because of the increased vasculature in the gravid cervix and increases the risk of premature rupture of membranes and preterm labor two- to threefold. Cone biopsy should not be done within 4 weeks of delivery.

Management of invasive disease is guided by the stage of disease, the gestational age of the fetus, and the desire of the mother to have the baby. If the disease is in early stage and the pregnancy is desired, it is safe to delay treatment regardless of gestational age until fetal maturity allows for safe delivery. If the disease is in advanced stage and the pregnancy is desired, the safety of delaying therapy is unproven. Abortion followed by definitive therapy is recommended for women with advanced cancer in the first or second trimester (see Chap. 97). In women in the third trimester with advanced disease, the baby should be delivered at the earliest possible time and followed immediately by stage-appropriate therapy. Most women with invasive cancer have early-stage disease. If the disease is microinvasive, vaginal delivery can take place and be followed by definitive treatment, usually conization. If a lesion is visible on the cervix, delivery is best done by caesarian section and followed by radical hysterectomy.

BREAST CANCER DURING PREGNANCY

Breast cancer occurs once in 3000 to 10,000 live births. About 5% of all breast cancers occur in women 40 years or younger. Among all premenopausal women with breast cancer, 25–30% were pregnant at the time of diagnosis. While early pregnancy is a protective factor against breast cancer in women as a whole, the breast cancers diagnosed during pregnancy are often diagnosed at a later stage of disease and so have a poorer outcome. The late diagnosis has at least two contributions. One is the more aggressive behavior of the cancer possibly related to the hormonal milieu (estrogen increases 100-fold; progesterone increases 1000-fold) of the pregnancy. However, about 70% of the breast cancers found in pregnancy are estrogen receptor–negative. Another factor is that early physical signs of the disease are often attributed to the changes that occur in the breast normally as a part of pregnancy. However, a breast mass in a pregnant woman is never normal. Younger women with breast cancer have a higher likelihood of having mutations in *BRCA1* or *BRCA2*. Pregnancy retains its protective effects in carriers of *BRCA1* mutations; such women with four or more children had a 38% reduction in breast cancer risk compared to nulliparous carriers. However, pregnancy seems to increase the risk of breast cancer among carriers of *BRCA2* mutations, particularly in the first 2 years after pregnancy. About 28–58% of the tumors express HER-2.

Primary tumors in pregnant women are 3.5 cm on average, compared to <2 cm in nonpregnant women. A dominant mass and a nipple discharge are the most common presenting signs and they should

prompt ultrasonography and breast MRI exam (if available) followed by lumpectomy if the mass is solid and aspiration if the mass is cystic. Mammography is less reliable in pregnancy due to the increased breast density. Needle aspirates of breast masses in pregnant women are often nondiagnostic or falsely positive. Even in pregnancy, most breast masses are benign (~80% are adenoma, lobular hyperplasia, milk retention cyst, fibrocystic disease, fibroadenoma, and other rarer entities).

Differences between pregnancy-associated breast cancer (often defined as cancer detected during the pregnancy and up to 1 year after delivery) are shown in Table e6-2. About 20% of breast cancers are detected in the first trimester, 45% in the second trimester, and 35% in the third trimester. Some argue that stage for stage, the outcome is the same for breast cancer diagnosed in pregnant and nonpregnant women.

Table e6-2 Differences in Breast Cancers in Pregnant and Nonpregnant Women

	Pregnant	Nonpregnant
Tumor size	3.5 cm	2 cm
Estrogen receptor +	30%[a]	67%
HER-2 +	up to 58%	10–25%
Stage II, III	65–90%	45–66%
Lymph node +	56–89%	38–54%

[a] Lower measured levels could be in part artifactual due to the increased levels of estrogen in the milieu

Staging the axillary lymph nodes is currently somewhat controversial. Sentinel lymph node sampling is not straightforward in pregnant women. Blue dye has been carcinogenic in rats, and shielding the fetus from administered radionuclides is of unproven efficacy. For this reason, many surgeons favor axillary node dissection to stage the nodes. Largely due to the typical delay in diagnosis, axillary nodes are more often positive in pregnant than in nonpregnant women.

As with other types of cancer in pregnant women, diagnosis in the first trimester often triggers a recommendation for an abortion to allow definitive therapeutic intervention at the earliest possible time. While definitive local surgery is applicable in the first trimester, radiation therapy and chemotherapy are considerably more risky. Delay in administration of systemic therapy can increase the risk of axillary spread. In the second and third trimesters, chemotherapy (particularly anthracycline-based combinations) are both safe and effective (see Chap. 90). Lumpectomy followed by adjuvant chemotherapy is frequently used; fluorouracil and cyclophosphamide with either doxorubicin or epirubicin have been given without major risk to the fetus. Taxanes and gemcitabine are also beginning to be used; however, safety data are sparse. Methotrexate and other folate antagonists are to be avoided because of effects on the fetal nervous system. Myelotoxic therapy is generally not administered after the 33rd or 34th week of gestation to allow 3 weeks off therapy before delivery for recovery of blood counts. Endocrine therapy and trastuzumab are unsafe during pregnancy. Experience with lapatinib is anecdotal, but no fetal malformations have been reported. Antiemetics and colony-stimulating factors are also considered safe. Women being treated into the postpartum period should not nurse their babies because of excretion of cancer chemotherapy agents, particularly alkylating agents, in milk.

Subsequent pregnancies following gestational breast cancer do not appear to influence relapse rate or overall survival. Indeed, a meta-analysis has suggested that pregnancy in breast cancer survivors may reduce the risk of dying from breast cancer by as much as 42%.

MELANOMA DURING PREGNANCY

Speculation about melanoma occurring during pregnancy based largely on anecdotal evidence and small case series concluded that it occurred with increased frequency, was more aggressive in its natural history, and was caused in part by the hormonal changes that also produced hyperpigmentation (so-called melasma) during pregnancy. However, more complete epidemiologic data suggest that melanoma is no more frequent in pregnant women than in nonpregnant women in the same age group, melanoma is not more aggressive during pregnancy, and that hormones seem to have little or nothing to do with the etiology. Pregnant and nonpregnant women do not differ in the location of primary tumor, depth of primary tumor, tumor ulceration, or vascular invasion.

Suspicious lesions should be looked for and managed definitively with excisional biopsy during pregnancy. Wide excision with sampling of regional lymph nodes is warranted. If lymph nodes are involved, the course of action is less clear. Several agents have demonstrated some activity in melanoma, but none have been used during pregnancy. Adjuvant interferon-alfa is toxic and its safety in pregnancy has not been documented. Agents active in advanced disease include dacarbazine,

interleukin-2, ipilumimab (antibody to CTLA-4), and in those with BRAF mutation V600E, a BRAF kinase inhibitor. In the setting of metastatic disease, abortion may be indicated so that systemic therapy can be initiated as soon as possible (see Chap. 87).

Pregnancy subsequent to the diagnosis and treatment of melanoma is also not associated with an increased risk of melanoma recurrence.

HODGKIN'S DISEASE AND NON-HODGKIN'S LYMPHOMA DURING PREGNANCY

(See Chap. 110) Hodgkin's disease occurs mainly in the age range of people who are of child-bearing age. However, Hodgkin's disease is not more common in pregnant than nonpregnant women. Hodgkin's disease is diagnosed in approximately 1 in 6,000 pregnancies. It generally presents as a nontender lymph node swelling, most often in the left supraclavicular region. It may be accompanied by B symptoms (fever, night sweats, unexplained weight loss). Excisional biopsy is the preferred diagnostic procedure as fine-needle aspiration cannot reveal the architectural framework that is an essential component of Hodgkin's disease diagnosis. The stage at presentation appears to be unaffected by pregnancy. Women diagnosed in the second and third trimester can be treated safely with combination chemotherapy, usually doxorubicin, bleomycin, vinblastine, and dacarbazine (ABVD). In general, the patient in the first trimester is asymptomatic and a woman with a desired pregnancy can be followed until the second or third trimester when definitive multiagent chemotherapy, can be safely given. Radiation therapy is not given during pregnancy. If symptoms requiring treatment appear during the first trimester, anecdotal evidence suggests that Hodgkin's disease symptoms can be controlled with weekly low-dose vinblastine. Such an approach has been safely used to avoid termination of pregnancy. Pregnancy does not have an adverse effect on treatment outcome.

Non-Hodgkin's lymphomas are more unusual in pregnancy (approximately 0.8 per 100,000 pregnancies), but are usually tumors with an aggressive natural history like diffuse large B-cell lymphoma, Burkitt's lymphoma, or peripheral T-cell lymphoma. Diagnosis relies on an excisional biopsy of a tumor mass, not fine-needle aspiration. Staging evaluation is generally limited to ultrasound or MRI examinations. Diagnosis in the first trimester should prompt termination of the pregnancy followed by definitive treatment with combination chemotherapy, as aggressive lymphomas are not likely to be held at bay with single-agent chemotherapy. Women diagnosed in the second or third trimesters can be treated with standard chemotherapy, such as with cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP). The experience with rituximab in this setting is anecdotal. However, infants born of mothers who have received rituximab may have transient delay in B-cell development that typically normalizes by 6 months. The treatment outcome is similar in lymphomas diagnosed in pregnant and nonpregnant women of the same clinical stage.

THYROID CANCER DURING PREGNANCY

(See Chap. 341) Thyroid cancer along with melanomas, brain tumors, and lymphomas are cancers that are increasing in incidence in the general population. Thyroid cancers are rising faster among women in North America than the other increasing tumor types. The Endocrine Society has developed practice guidelines to inform the management of patients with thyroid disease during pregnancy (http://www.endo-society.org/guidelines/final/upload/Clinical-Guideline-Executive-Summary-Management-of-Thyroid-Dysfunction-during-Pregnancy-Postpartum.pdf). Thyroid nodules 1 cm or larger are approached by fine-needle aspiration. If a malignancy is diagnosed, surgery is generally recommended in the second and third trimesters. However, surgical complications appear to be twice as common when the patient is pregnant. Because the growth of thyroid tumors is often indolent, surgery is not recommended in the first trimester. Patients with follicular cancer or early papillary cancer can be observed until the postpartum period. Radioactive iodine can be safely administered after delivery. Patients with a history of thyroid cancer who become pregnant should be maintained on thyroid hormone replacement during pregnancy because of the adverse impact of maternal hypothyroidism on the fetus. Women who are breast-feeding should not be treated with radioactive iodine and women treated with radioactive iodine should not become pregnant for 6–12 months after treatment.

The assessment of thyroid function during pregnancy is challenging because of the physiologic changes that occur during pregnancy. Women who have previously been treated for thyroid cancer are at risk of hypothyroidism. The demand for thyroid hormone increases during pregnancy, and doses to maintain normal function may increase by 30–50%. Total T4 levels are higher during pregnancy but target therapeutic levels also increase (Table e6-3). It is recommended that the upper and lower limits of the laboratory range be multiplied by 1.5 in the second and third trimester to establish a pregnancy-specific normal range. The target TSH level is lower than 2.5 mIU/L.

Table e6-3 Thyroid Function Test During Pregnancy (Mean Levels)

	Nonpregnant	First Trimester	Second Trimester	Third Trimester
TSH (mIU/L)	1.38	0.91	1.03	1.32

Total thyroxine (μg/dL)	7.35	10.98	11.88	11.08

Source: Based on the National Health and Nutrition Examination Survey III (NHANES III) (OP Soldin et al: Ther Drug Monit 17:303, 2007).

GESTATIONAL TROPHOBLASTIC DISEASE

(See Chap. 97) Gestational Trophoblastic Disease encompasses hydatidiform mole, choriocarcinoma, placental site trophoblastic tumor, and assorted miscellaneous and unclassifiable trophoblastic tumors. Moles are the most common, occurring in 1 in 1500 preg-nancies in the United States. The incidence is higher in Asia. In general, if the serum level of beta-human chorionic gonadotropin (HCG) returns to normal after surgical removal (evacuation) of the mole, the illness is considered gestational trophoblastic disease. By contrast, if the HCG level remains elevated after mole evacuation, the patient is considered to have gestational trophoblastic neoplasia. Choriocarcinoma occurs in 1 in 25,000 pregnancies. Maternal age >45 years and prior history of molar pregnancy are risk factors. A previous molar pregnancy makes choriocarcinoma about 1000 times more likely to occur (incidence 1 –2%).

Hydatidiform moles are characterized by clusters of villi with hydropic changes, trophoblastic hyperplasia, and absence of fetal blood vessels. Invasive moles are distinguished by invasion of the myometrium. Placental site trophoblastic tumors are composed mainly of cytotrophoblast cells arising at the site of placental origin. Choriocarcinomas contain anaplastic trophoblastic tissue with both cytotrophoblast and syncytiotrophoblast features and no identifiable villi.

Moles can be partial or complete. Partial moles have a distinct molecular origin and usually are smaller tumors with less hydropic villi. Partial moles result from fertilization of an egg by two sperm, resulting in diandric triploidy. Complete moles usually have a 46,XX genotype; 95% develop by a single male sperm fertilizing an empty egg and undergoing gene duplication (diandric diploidy); 5% develop from dispermic fertilization of an empty egg (diandric dispermy).

Women with gestational trophoblastic disease often present with first-trimester bleeding and unusually large uterine size. Ultrasound shows absence of fetal parts or heart sounds. Patients are monitored by chest radiograph, pelvic examination, and weekly measurement of HCG levels.

Patients with molar pregnancies require suction curettage with postoperative HCG monitoring. In 80% of cases, HCG declines within 8–10 days. Patients should not become pregnant for at least 12 months. Women with invasive moles generally undergo hysterectomy followed by chemotherapy. About half of choriocarcinomas develop after a molar pregnancy and half develop after ectopic pregnancy or rarely, after a normal full-term pregnancy. Disease is classified as stage I if it is confined to the uterus, stage II if disease is limited to genital structures (~30% have vaginal involvement), stage III if disease has spread to the lungs but no other organs, and stage IV if disease has spread to liver, brain, or other organs.

Specific criteria have been developed to aid the decision about when disease becomes neoplasia:

1. Four consecutive increased HCG levels over the 3 weeks after evacuation surgery.
2. A rise in HCG of 10% or more on three consecutive values over 2 or more weeks.
3. The presence of choriocarcinoma.
4. Persistent HCG elevations 6 months after evacuation.

Patients without widely metastatic disease are generally managed with single-agent methotrexate (either 30 mg/m^2 IM weekly until HCG normalizes or 1 mg/kg IM every other day for 4 doses followed by leukovorin 0.1 mg/kg IV 24 hours after methotrexate), which cures >90% of patients. Patients with very high HCG levels, presenting >4 months after a pregnancy, with brain or liver metastases, or failing to be cured by single agent methotrexate are treated with combination chemotherapy. Etoposide, methotrexate, and dactinomycin alternating with cyclophosphamide and vincristine (EMA-CO) is the most commonly used regimen producing long-term survival in >80% of patients. Brain metastases can usually be controlled with brain radiation therapy. Women cured of trophoblastic disease who have not undergone hysterectomy do not appear to have increased risk of fetal abnormalities or maternal complications with subsequent pregnancies.

FURTHER READINGS

Azim HA Jr et al: Treatment of the pregnant mother with cancer: A systematic review on the use of cytotoxic, endocrine, targeted agents, and immunotherapy during pregnancy. Part I: Solid tumors. Cancer Treat Rev 36:101, 2010[PMID: 20015593]

——— et al: Treatment of the pregnant mother with cancer: A systematic review on the use of cytotoxic, endocrine, targeted agents, and immunotherapy during pregnancy. Part II: Hematological tumors. Cancer Treat Rev 36:110, 2010

Leslie KK (ed): Cancer complicating pregnancy. Obstet Gynecol Clin N Am 32: 533, 2005

Pentheroudakis G et al: Cancer, fertility and pregnancy: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 21:v266, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PAIN: PATHOPHYSIOLOGY AND MANAGEMENT: INTRODUCTION

The task of medicine is to preserve and restore health and to relieve suffering. Understanding pain is essential to both of these goals. Because pain is universally understood as a signal of disease, it is the most common symptom that brings a patient to a physician's attention. The function of the pain sensory system is to protect the body and maintain homeostasis. It does this by detecting, localizing, and identifying potential or actual tissue-damaging processes. Because different diseases produce characteristic patterns of tissue damage, the quality, time course, and location of a patient's pain complaint provide important diagnostic clues. It is the physician's responsibility to provide rapid and effective pain relief.

THE PAIN SENSORY SYSTEM

Pain is an unpleasant sensation localized to a part of the body. It is often described in terms of a penetrating or tissue-destructive process (e.g., stabbing, burning, twisting, tearing, squeezing) and/or of a bodily or emotional reaction (e.g., terrifying, nauseating, sickening). Furthermore, any pain of moderate or higher intensity is accompanied by anxiety and the urge to escape or terminate the feeling. These properties illustrate the duality of pain: it is both sensation and emotion. When it is acute, pain is characteristically associated with behavioral arousal and a stress response consisting of increased blood pressure, heart rate, pupil diameter, and plasma cortisol levels. In addition, local muscle contraction (e.g., limb flexion, abdominal wall rigidity) is often present.

Peripheral Mechanisms

THE PRIMARY AFFERENT NOCICEPTOR

A peripheral nerve consists of the axons of three different types of neurons: primary sensory afferents, motor neurons, and sympathetic postganglionic neurons **(Fig. 11-1)**. The cell bodies of primary sensory afferents are located in the dorsal root ganglia in the vertebral foramina. The primary afferent axon has two branches: one projects centrally into the spinal cord and the other projects peripherally to innervate tissues. Primary afferents are classified by their diameter, degree of myelination, and conduction velocity. The largest-diameter afferent fibers, A-beta (Aβ), respond maximally to light touch and/or moving stimuli; they are present primarily in nerves that innervate the skin. In normal individuals, the activity of these fibers does not produce pain. There are two other classes of primary afferents: the small-diameter myelinated A-delta (Aδ) and the unmyelinated (C fiber) axons (Fig. 11-1). These fibers are present in nerves to the skin and to deep somatic and visceral structures. Some tissues, such as the cornea, are innervated only by Aδ and C fiber afferents. Most Aδ and C fiber afferents respond maximally only to intense (painful) stimuli and produce the subjective experience of pain when they are electrically stimulated; this defines them as *primary afferent nociceptors* (*pain receptors*). The ability to detect painful stimuli is completely abolished when conduction in Aδ and C fiber axons is blocked.

Figure 11-1

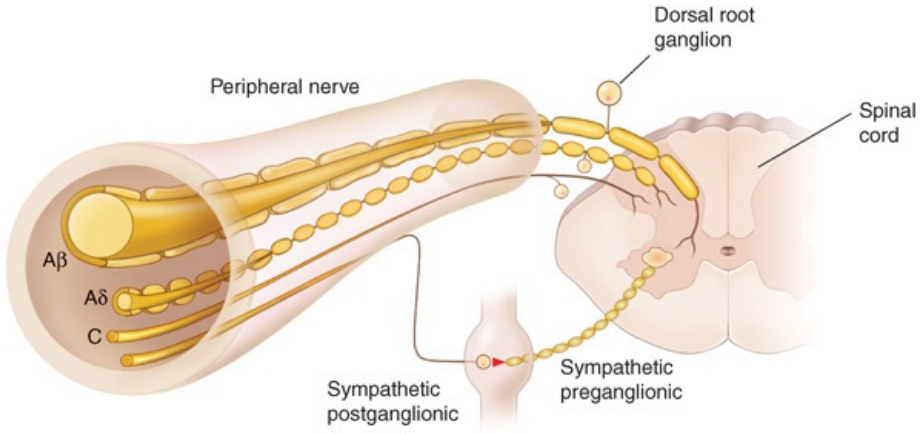

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Components of a typical cutaneous nerve. There are two distinct functional categories of axons: primary afferents with cell bodies in the dorsal root ganglion, and sympathetic postganglionic fibers with cell bodies in the sympathetic ganglion. Primary afferents include those with large-diameter myelinated (Aβ), small-diameter myelinated (Aδ), and unmyelinated (C) axons. All sympathetic postganglionic fibers are unmyelinated.

Individual primary afferent nociceptors can respond to several different types of noxious stimuli. For example, most nociceptors respond to heat; intense cold; intense mechanical stimuli, such as a pinch; changes in pH, particularly an acidic environment; and application of chemical irritants including adenosine triphosphate (ATP), serotonin, bradykinin, and histamine.

SENSITIZATION

When intense, repeated, or prolonged stimuli are applied to damaged or inflamed tissues, the threshold for activating primary afferent nociceptors is lowered, and the frequency of firing is higher for all stimulus intensities. Inflammatory mediators such as bradykinin, nerve-growth factor, some prostaglandins, and leukotrienes contribute to this process, which is called *sensitization*. Sensitization occurs at the level of the peripheral nerve terminal (*peripheral sensitization)* as well as at the level of the dorsal horn of the spinal cord (*central sensitization).* Peripheral sensitization occurs in damaged or inflamed tissues, when inflammatory mediators activate intracellular signal transduction in nociceptors, prompting an increase in the production, transport, and membrane insertion of chemically gated and voltage-gated ion channels. These changes increase the excitability of nociceptor terminals and lower their threshold for activation by mechanical, thermal, and chemical stimuli. Central sensitization occurs when activity, generated by nociceptors during inflammation, enhances the excitability of nerve cells in the dorsal horn of the spinal cord. Following injury and resultant sensitization, normally innocuous stimuli can produce pain. Sensitization is a clinically important process that contributes to tenderness, soreness, and hyperalgesia (increased pain intensity in response to the same noxious stimulus; e.g. moderate pressure causes severe pain). A striking example of sensitization is sunburned skin, in which severe pain can be produced by a gentle slap on the back or a warm shower.

Sensitization is of particular importance for pain and tenderness in deep tissues. Viscera are normally relatively insensitive to noxious mechanical and thermal stimuli, although hollow viscera do generate significant discomfort when distended. In contrast, when affected by a disease process with an inflammatory component, deep structures such as joints or hollow viscera characteristically become exquisitely sensitive to mechanical stimulation.

A large proportion of Aδ and C fiber afferents innervating viscera are completely insensitive in normal noninjured, noninflamed tissue. That is, they cannot be activated by known mechanical or thermal stimuli and are not spontaneously active. However, in the presence of inflammatory mediators, these afferents become sensitive to mechanical stimuli. Such afferents have been termed *silent nociceptors*, and their characteristic properties may explain how, under pathologic conditions, the relatively insensitive deep structures can become the source of severe and debilitating pain and tenderness. Low pH, prostaglandins, leukotrienes, and other inflammatory mediators such as bradykinin play a significant role in sensitization.

NOCICEPTOR-INDUCED INFLAMMATION

Primary afferent nociceptors also have a neuroeffector function. Most nociceptors contain polypeptide mediators that are released from their peripheral terminals when they are activated **(Fig. 11-2)**. An example is substance P, an 11-amino-acid peptide. Substance P is released from primary afferent nociceptors and has multiple biologic activities. It is a potent vasodilator, degranulates mast cells, is a chemoattractant for leukocytes, and increases the production and release of inflammatory mediators. Interestingly, depletion of substance P from joints reduces the severity of experimental arthritis.

Primary afferent nociceptors are not simply passive messengers of threats to tissue injury but also play an active role in tissue protection through these neuroeffector functions.

Figure 11-2

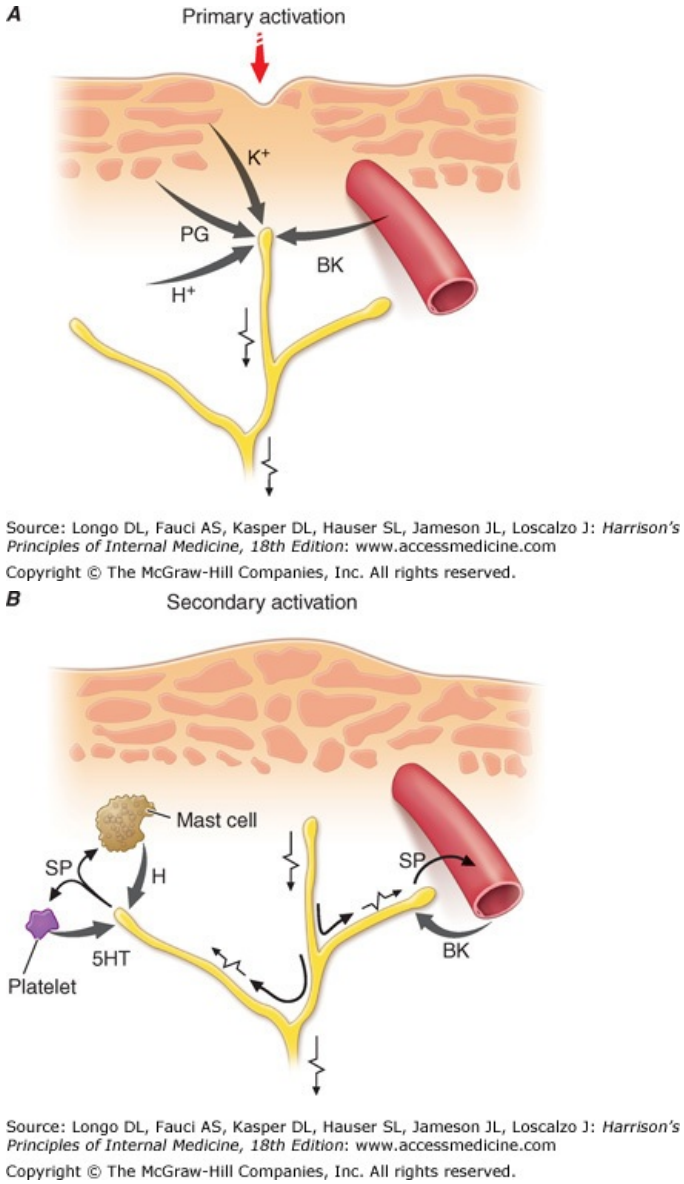

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Events leading to activation, sensitization, and spread of sensitization of primary afferent nociceptor terminals. ***A.*** Direct activation by intense pressure and consequent cell damage. Cell damage induces lower pH (H^+) and leads to release of potassium (K^+) and to synthesis of prostaglandins (PG) and bradykinin (BK). Prostaglandins increase the sensitivity of the terminal to bradykinin and other pain-producing substances. ***B.*** Secondary activation. Impulses generated in the stimulated terminal propagate not only to the spinal cord but also into other terminal branches where they induce the release of peptides, including substance P (SP). Substance P causes vasodilation and neurogenic edema with further accumulation of bradykinin (BK). Substance P also causes the release of histamine (H) from mast cells and serotonin (5HT) from platelets.

Central Mechanisms

THE SPINAL CORD AND REFERRED PAIN

The axons of primary afferent nociceptors enter the spinal cord via the dorsal root. They terminate in the dorsal horn of the spinal gray matter **(Fig. 11-3)**. The terminals of primary afferent axons contact spinal neurons that transmit the pain signal to brain sites involved in pain perception. When primary afferents are activated by noxious stimuli, they release neurotransmitters from their terminals that excite the spinal cord neurons. The major neurotransmitter released is glutamate, which rapidly excites dorsal horn neurons. Primary afferent nociceptor terminals also release peptides, including substance P and calcitoningene-related peptide, which produce a slower and longer-lasting excitation of the dorsal horn neurons. The axon of each primary afferent contacts many spinal neurons, and each spinal neuron receives convergent inputs from many primary afferents.

Figure 11-3

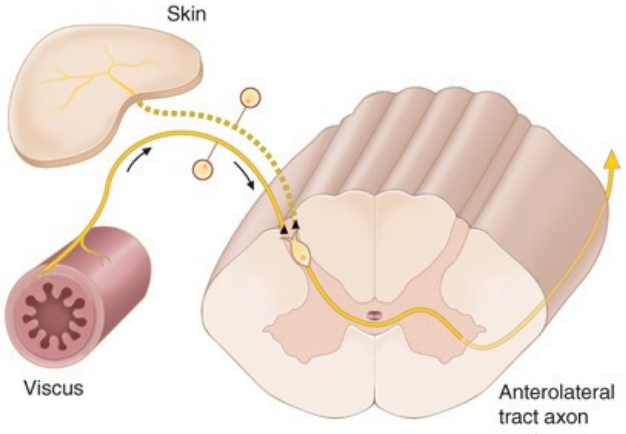

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The convergence-projection hypothesis of referred pain. According to this hypothesis, visceral afferent nociceptors converge on the same pain-projection neurons as the afferents from the somatic structures in which the pain is perceived. The brain has no way of knowing the actual source of input and mistakenly "projects" the sensation to the somatic structure.

The convergence of sensory inputs to a single spinal pain-transmission neuron is of great importance because it underlies the phenomenon of referred pain. All spinal neurons that receive input from the viscera and deep musculoskeletal structures also receive input from the skin. The convergence patterns are determined by the spinal segment of the dorsal root ganglion that supplies the afferent innervation of a structure. For example, the afferents that supply the central diaphragm are derived from the third and fourth cervical dorsal root ganglia. Primary afferents with cell bodies in these same ganglia supply the skin of the shoulder and lower neck. Thus, sensory inputs from both the shoulder skin and the central diaphragm converge on pain-transmission neurons in the third and fourth cervical spinal segments. *Because of this convergence and the fact that the spinal neurons are most often activated by inputs from the skin, activity evoked in spinal neurons by input from deep structures is mislocalized by the patient to a place that roughly corresponds with the region of skin innervated by the same spinal segment*. Thus, inflammation near the central diaphragm is usually reported as shoulder discomfort. This spatial displacement of pain sensation from the site of the injury that produces it is known as *referred pain*.

ASCENDING PATHWAYS FOR PAIN

A majority of spinal neurons contacted by primary afferent nociceptors send their axons to the contralateral thalamus. These axons form the contralateral spinothalamic tract, which lies in the anterolateral white matter of the spinal cord, the lateral edge of the medulla, and the lateral pons and midbrain. The spinothalamic pathway is crucial for pain sensation in humans. Interruption of this pathway produces permanent deficits in pain and temperature discrimination.

Spinothalamic tract axons ascend to several regions of the thalamus. There is tremendous divergence of the pain signal from these thalamic sites to broad areas of the cerebral cortex that subserve different aspects of the pain experience **(Fig. 11-4)**. One of the thalamic projections is to the somatosensory cortex. This projection mediates the purely sensory aspects of pain, i.e., its location, intensity, and quality. Other thalamic neurons project to cortical regions that are linked to emotional responses, such as the cingulate gyrus and other areas of the frontal lobes, including the insular cortex. These pathways to the frontal cortex subserve the affective or unpleasant emotional dimension of pain. This affective dimension of pain produces suffering and exerts potent control of behavior. Because of this dimension, fear is a constant companion of pain. As a consequence, injury or surgical lesions to areas of the frontal cortex activated by painful stimuli diminish the emotional impact of pain while largely preserving the individual's ability to recognize noxious stimuli as painful.

Figure 11-4

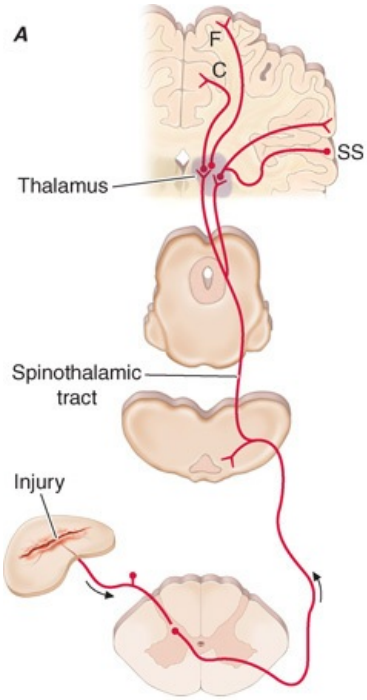

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

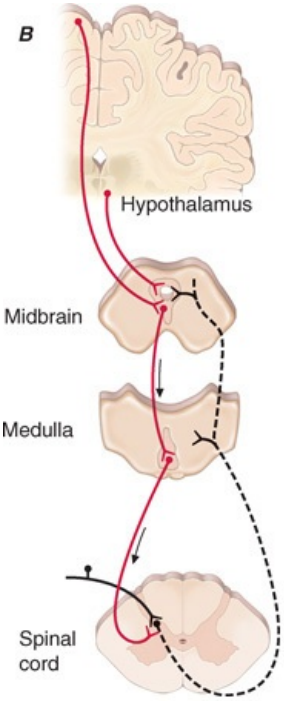

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pain transmission and modulatory pathways. *A.* Transmission system for nociceptive messages. Noxious stimuli activate the sensitive peripheral ending of the primary afferent nociceptor by the process of transduction. The message is then transmitted over the peripheral nerve to the spinal cord, where it synapses with cells of origin of the major ascending pain pathway, the spinothalamic tract. The message is relayed in the thalamus to the anterior cingulate (C), frontal insular (F), and somatosensory cortex (SS). ***B.*** Pain-modulation network. Inputs from frontal cortex and hypothalamus activate cells in the midbrain that control spinal pain-transmission cells via cells in the medulla.

Pain Modulation

The pain produced by injuries of similar magnitude is remarkably variable in different situations and in different individuals. For example, athletes have been known to sustain serious fractures with only minor pain, and Beecher's classic World War II survey revealed that many soldiers in battle were unbothered by injuries that would have produced agonizing pain in civilian patients. Furthermore, even the suggestion that a treatment will relieve pain can have a significant analgesic effect (the placebo

effect). On the other hand, many patients find even minor injuries (such as venipuncture) frightening and unbearable, and the expectation of pain can induce pain even without a noxious stimulus. The suggestion that pain will worsen following administration of an inert substance can increase its perceived intensity (the nocebo effect).

The powerful effect of expectation and other psychological variables on the perceived intensity of pain is explained by brain circuits that modulate the activity of the pain-transmission pathways. One of these circuits has links to the hypothalamus, midbrain, and medulla, and it selectively controls spinal pain-transmission neurons through a descending pathway (Fig. 11-4).

Human brain–imaging studies have implicated this pain-modulating circuit in the pain-relieving effect of attention, suggestion, and opioid analgesic medications **(Fig. 11-5)**. Furthermore, each of the component structures of the pathway contains opioid receptors and is sensitive to the direct application of opioid drugs. In animals, lesions of this descending modulatory system reduce the analgesic effect of systemically administered opioids such as morphine. Along with the opioid receptor, the component nuclei of this pain-modulating circuit contain endogenous opioid peptides such as the enkephalins and β-endorphin.

Figure 11-5

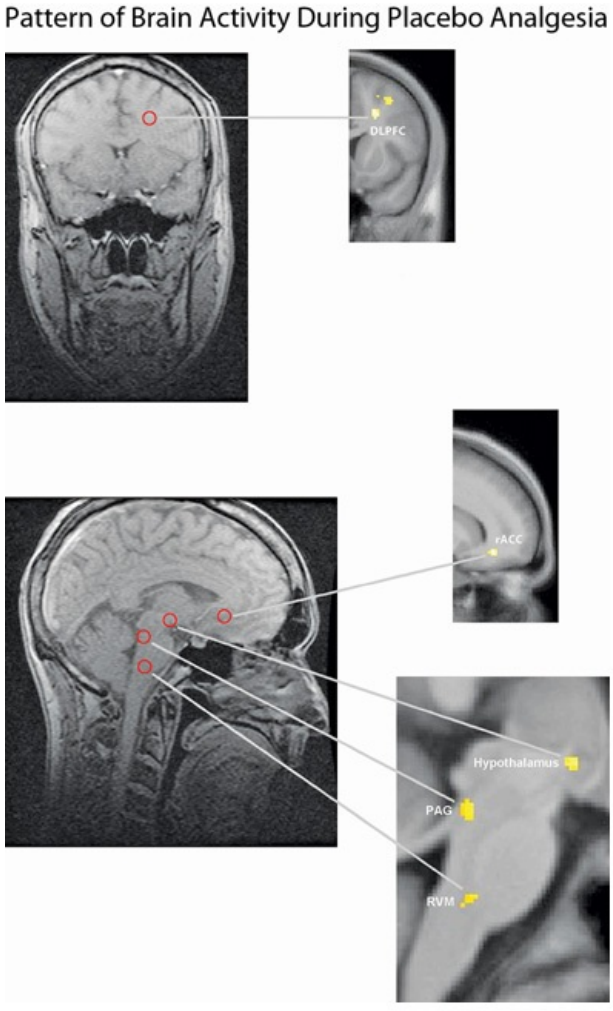

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Functional magnetic resonance imaging (fMRI) demonstrates placebo-enhanced brain activity in anatomic regions correlating with the opioidergic descending pain control system. Top panel, Frontal fMRI image shows placebo-enhanced brain activity in the dorsal lateral prefrontal cortex (DLPFC). Bottom panel, Sagittal fMRI images show placebo-enhanced responses in the rostral anterior cingulate cortex (rACC), the rostral ventral medullae (RVM), the periaqueductal gray (PAG) area, and the hypothalamus. The placebo-enhanced activity in all areas was reduced by naloxone, demonstrating the link between the descending opioidergic system and the placebo analgesic response. *(Adapted with permission from Eippert et al.)*

The most reliable way to activate this endogenous opioid-mediated modulating system is by suggestion of pain relief or by intense emotion directed away from the pain-causing injury (e.g., during severe threat or an athletic competition). In fact, pain-relieving endogenous opioids are released

following surgical procedures and in patients given a placebo for pain relief.

Pain-modulating circuits can enhance as well as suppress pain. Both pain-inhibiting and pain-facilitating neurons in the medulla project to and control spinal pain-transmission neurons. Because pain-transmission neurons can be activated by modulatory neurons, it is theoretically possible to generate a pain signal with no peripheral noxious stimulus. In fact, human functional imaging studies have demonstrated increased activity in this circuit during migraine headaches. A central circuit that facilitates pain could account for the finding that pain can be induced by suggestion or enhanced by expectation and provides a framework for understanding how psychological factors can contribute to chronic pain.

Neuropathic Pain

Lesions of the peripheral or central nociceptive pathways typically result in a loss or impairment of pain sensation. Paradoxically, damage to or dysfunction of these pathways can also produce pain. For example, damage to peripheral nerves, as occurs in diabetic neuropathy, or to primary afferents, as in herpes zoster, can result in pain that is referred to the body region innervated by the damaged nerves. Pain may also be produced by damage to the central nervous system (CNS), for example, in some patients following trauma or cerebrovascular injury to spinal cord, brainstem, or thalamic areas that contain central nociceptive pathways. Such neuropathic pains are often severe and are typically resistant to standard treatments for pain.

Neuropathic pain typically has an unusual burning, tingling, or electric shock–like quality and may be triggered by very light touch. These features are rare in other types of pain. On examination, a sensory deficit is characteristically present in the area of the patient's pain. Hyperpathia, a greatly exaggerated pain sensation to innocuous or mild nociceptive stimuli, is also characteristic of neuropathic pain; patients often complain that the very lightest moving stimulus evokes exquisite pain (allodynia). In this regard, it is of clinical interest that a topical preparation of 5% lidocaine in patch form is effective for patients with postherpetic neuralgia who have prominent allodynia.

A variety of mechanisms contribute to neuropathic pain. As with sensitized primary afferent nociceptors, damaged primary afferents, including nociceptors, become highly sensitive to mechanical stimulation and may generate impulses in the absence of stimulation. Increased sensitivity and spontaneous activity are due, in part, to an increased concentration of sodium channels. Damaged primary afferents may also develop sensitivity to norepinephrine. Interestingly, spinal cord pain-transmission neurons cut off from their normal input may also become spontaneously active. Thus, both CNS and peripheral nervous system hyperactivity contribute to neuropathic pain.

SYMPATHETICALLY MAINTAINED PAIN

Patients with peripheral nerve injury occasionally develop spontaneous pain in the region innervated by the nerve. This pain is often described as having a burning quality. The pain typically begins after a delay of hours to days or even weeks and is accompanied by swelling of the extremity, periarticular bone loss, and arthritic changes in the distal joints. The pain may be relieved by a local anesthetic block of the sympathetic innervation to the affected extremity. Damaged primary afferent nociceptors acquire adrenergic sensitivity and can be activated by stimulation of the sympathetic outflow. This constellation of spontaneous pain and signs of sympathetic dysfunction following injury has been termed complex regional pain syndrome (CRPS). When this occurs after an identifiable nerve injury, it is termed CRPS type II (also known as posttraumatic neuralgia or, if severe, *causalgia*). When a similar clinical picture appears without obvious nerve injury, it is termed CRPS type I (also known as *reflex sympathetic dystrophy*). CRPS can be produced by a variety of injuries, including fractures of bone, soft tissue trauma, myocardial infarction, and stroke (Chap. 375). CRPS type I typically resolves with symptomatic treatment; however, when it persists, detailed examination often reveals evidence of peripheral nerve injury. Although the pathophysiology of CRPS is poorly understood, the pain and the signs of inflammation, when acute, can be rapidly relieved by blocking the sympathetic nervous system. This implies that sympathetic activity can activate undamaged nociceptors when inflammation is present. Signs of sympathetic hyperactivity should be sought in patients with posttraumatic pain and inflammation and no other obvious explanation.

TREATMENT: ACUTE PAIN

The ideal treatment for any pain is to remove the cause; thus, while treatment can be initiated immediately, efforts to establish the underlying etiology should always proceed as treatment begins. Sometimes, treating the underlying condition does not immediately relieve pain. Furthermore, some conditions are so painful that rapid and effective analgesia is essential (e.g., the postoperative state, burns, trauma, cancer, or sickle cell crisis). Analgesic medications are a first line of treatment in these cases, and all practitioners should be familiar with their use.

Aspirin, Acetaminophen, and Nonsteroidal Anti-Inflammatory Agents (NSAIDs)

These drugs are considered together because they are used for similar problems and may have a similar mechanism of action **(Table 11-1)**. All these compounds inhibit cyclooxygenase (COX), and, except for acetaminophen, all have anti-inflammatory actions, especially at higher dosages. They are

particularly effective for mild to moderate headache and for pain of musculoskeletal origin.

Table 11-1 Drugs for Relief of Pain

Generic Name	Dose, mg	Interval	Comments
Nonnarcotic analgesics: usual doses and intervals			
Acetylsalicylic acid	650 PO	q 4 h	Enteric-coated preparations available
Acetaminophen	650 PO	q 4 h	Side effects uncommon
Ibuprofen	400 PO	q 4–6 h	Available without prescription
Naproxen	250–500 PO	q 12 h	Delayed effects may be due to long half-life
Fenoprofen	200 PO	q 4–6 h	Contraindicated in renal disease
Indomethacin	25–50 PO	q 8 h	Gastrointestinal side effects common
Ketorolac	15–60 IM/IV	q 4–6 h	Available for parenteral use
Celecoxib	100–200 PO	q 12–24 h	Useful for arthritis
Valdecoxib	10–20 PO	q12–24 h	Removed from U.S. market in 2005

Generic Name	Parenteral Dose, mg	PO Dose, mg	Comments
Narcotic analgesics: usual doses and intervals			
Codeine	30–60 q 4 h	30–60 q 4 h	Nausea common
Oxycodone	—	5–10 q 4–6 h	Usually available with acetaminophen or aspirin
Morphine	5 q 4 h	30 q 4 h	
Morphine sustained release	—	15–60 bid to tid	Oral slow-release preparation
Hydromorphone	1–2 q 4 h	2–4 q 4 h	Shorter acting than morphine sulfate
Levorphanol	2 q 6–8 h	4 q 6–8 h	Longer acting than morphine sulfate; absorbed well PO
Methadone	5-10 q 6–8 h	5-20 q 6–8 h	Delayed sedation due to long half-life; therapy should not be initiated with greater than 40 mg/day and dose escalation should be made no more frequently than every 3 days
Meperidine	50–100 q 3–4 h	300 q 4 h	Poorly absorbed PO; normeperidine a toxic metabolite; routine use of this agent is not recommended
Butorphanol	—	1–2 q 4 h	Intranasal spray
Fentanyl	25–100 μg/h	—	72-h transdermal patch
Tramadol	—	50–100 q 4–6 h	Mixed opioid/adrenergic action

Generic Name	Uptake Blockade		Sedative Potency	Anticholinergic Potency	Orthostatic Hypotension	Cardiac Arrhythmia	Ave. Dose, mg/d	Range, mg/d
	5-HT	NE						
Antidepressants[a]								
Doxepin	++	+	High	Moderate	Moderate	Less	200	75–400
Amitriptyline	++++	++	High	Highest	Moderate	Yes	150	25–300
Imipramine	++++	++	Moderate	Moderate	High	Yes	200	75–400
Nortriptyline	+++	++	Moderate	Moderate	Low	Yes	100	40–150
Desipramine	+++	++++	Low	Low	Low	Yes	150	50–300
Venlafaxine	+++	++	Low	None	None	No	150	75–400

Duloxetine	+++	+++	Low	None	None	No	40	30–60

Generic Name	PO Dose, mg	Interval	Generic Name	PO Dose, mg	Interval
Anticonvulsants and antiarrhythmics[a]					
Phenytoin	300	daily/qhs	Clonazepam	1	q 6 h
Carbamazepine	200–300	q 6 h	Gabapentin[b]	600–1200	q 8 h
Oxcarbazepine	300	bid	Pregabalin	150–600	bid

[a]Antidepressants, anticonvulsants, and antiarrhythmics have not been approved by the U.S. Food and Drug Administration (FDA) for the treatment of pain.

[b]Gabapentin in doses up to 1800 mg/d is FDA approved for postherpetic neuralgia.

Note: 5-HT, serotonin; NE, norepinephrine.

Because they are effective for these common types of pain and are available without prescription, COX inhibitors are by far the most commonly used analgesics. They are absorbed well from the gastrointestinal tract and, with occasional use, have only minimal side effects. With chronic use, gastric irritation is a common side effect of aspirin and NSAIDs and is the problem that most frequently limits the dose that can be given. Gastric irritation is most severe with aspirin, which may cause erosion and ulceration of the gastric mucosa leading to bleeding or perforation. Because aspirin irreversibly acetylates platelet cyclooxygenase and thereby interferes with coagulation of the blood, gastrointestinal bleeding is a particular risk. Older age and history of gastrointestinal disease increase the risks of aspirin and NSAIDs. In addition to the well-known gastrointestinal toxicity of NSAIDs, nephrotoxicity is a significant problem for patients using these drugs on a chronic basis. Patients at risk for renal insufficiency, particularly those with significant contraction of their intravascular volume as occurs with chronic diuretic use or acute hypovolemia, should be monitored closely. NSAIDs can also increase blood pressure in some individuals. Long-term treatment with NSAIDs requires regular blood pressure monitoring and treatment if necessary. Although toxic to the liver when taken in high doses, acetaminophen rarely produces gastric irritation and does not interfere with platelet function.

The introduction of a parenteral form of NSAID, ketorolac, extends the usefulness of this class of compounds in the management of acute severe pain. Ketorolac is sufficiently potent and rapid in onset to supplant opioids for many patients with acute severe headache and musculoskeletal pain.

There are two major classes of COX: COX-1 is constitutively expressed, and COX-2 is induced in the inflammatory state. COX-2–selective drugs have similar analgesic potency and produce less gastric irritation than the nonselective COX inhibitors. The use of COX-2–selective drugs does not appear to lower the risk of nephrotoxicity compared to nonselective NSAIDs. On the other hand, COX-2 –selective drugs offer a significant benefit in the management of acute postoperative pain because they do not affect blood coagulation. Nonselective COX inhibitors are usually contraindicated postoperatively because they impair platelet-mediated blood clotting and are thus associated with increased bleeding at the operative site. COX-2 inhibitors, including celecoxib (Celebrex) are associated with increased cardiovascular risk. It is possible that this is a class effect of NSAIDs, excluding aspirin. These drugs are contraindicated in patients in the immediate period after coronary artery bypass surgery and should be used with caution in patients with a history of or significant risk factors for cardiovascular disease.

Opioid Analgesics

Opioids are the most potent pain-relieving drugs currently available. Of all analgesics, they have the broadest range of efficacy and provide the most reliable and effective method for rapid pain relief. Although side effects are common, most are reversible: nausea, vomiting, pruritus, and constipation are the most frequent and bothersome side effects. Respiratory depression is uncommon at standard analgesic doses, but can be life-threatening. Opioid-related side effects can be reversed rapidly with the narcotic antagonist naloxone. The physician should not hesitate to use opioid analgesics in patients with acute severe pain. **Table 11-1** lists the most commonly used opioid analgesics.

Opioids produce analgesia by actions in the CNS. They activate pain-inhibitory neurons and directly inhibit pain-transmission neurons. Most of the commercially available opioid analgesics act at the same opioid receptor (μ-receptor), differing mainly in potency, speed of onset, duration of action, and optimal route of administration. Some side effects are due to accumulation of nonopioid metabolites that are unique to individual drugs. One striking example of this is normeperidine, a metabolite of meperidine. Normeperidine produces hyperexcitability and seizures that are not reversible with naloxone. Normeperidine accumulation is increased in patients with renal failure.

The most rapid relief with opioids is obtained by intravenous administration; relief with oral

administration is significantly slower. Common side effects include nausea, vomiting, constipation, and sedation. The most serious side effect is respiratory depression. Patients with any form of respiratory compromise must be kept under close observation following opioid administration; an oxygen-saturation monitor may be useful. Opioid-induced respiratory depression is typically accompanied by significant sedation and a reduction in respiratory rate. A fall in oxygen saturation represents a critical level of respiratory depression and the need for immediate intervention to prevent life-threatening hypoxemia. Ventilatory assistance should be maintained until the opioid-induced respiratory depression has resolved. The opioid antagonist naloxone should be readily available whenever opioids are used at high doses or in patients with compromised pulmonary function. Opioid effects are dose-related, and there is great variability among patients in the doses that relieve pain and produce side effects. Because of this, initiation of therapy requires titration to optimal dose and interval. The most important principle is to provide adequate pain relief. This requires determining whether the drug has adequately relieved the pain and frequent reassessment to determine the optimal interval for dosing. *The most common error made by physicians in managing severe pain with opioids is to prescribe an inadequate dose. Because many patients are reluctant to complain, this practice leads to needless suffering*. In the absence of sedation at the expected time of peak effect, a physician should not hesitate to repeat the initial dose to achieve satisfactory pain relief.

An innovative approach to the problem of achieving adequate pain relief is the use of patient-controlled analgesia (PCA). PCA uses a microprocessor-controlled infusion device that can deliver a baseline continuous dose of an opioid drug as well as preprogrammed additional doses whenever the patient pushes a button. The patient can then titrate the dose to the optimal level. This approach is used most extensively for the management of postoperative pain, but there is no reason why it should not be used for any hospitalized patient with persistent severe pain. PCA is also used for short-term home care of patients with intractable pain, such as that caused by metastatic cancer.

It is important to understand that the PCA device delivers small, repeated doses to maintain pain relief; in patients with severe pain, the pain must first be brought under control with a loading dose before transitioning to the PCA device. The bolus dose of the drug (typically 1 mg morphine, 0.2 mg of hydromorphone, or 10 μg fentanyl) can then be delivered repeatedly as needed. To prevent overdosing, PCA devices are programmed with a lockout period after each demand dose is delivered (5–10 min) and a limit on the total dose delivered per hour. While some have advocated the use of a simultaneous continuous or basal infusion of the PCA drug, this increases the risk of respiratory depression and has not been shown to increase the overall efficacy of the technique.

Many physicians, nurses, and patients have a certain trepidation about using opioids that is based on an exaggerated fear of addiction. In fact, there is a vanishingly small chance of patients becoming addicted to narcotics as a result of their appropriate medical use.

The availability of new routes of administration has extended the usefulness of opioid analgesics. Most important is the availability of spinal administration. Opioids can be infused through a spinal catheter placed either intrathecally or epidurally. By applying opioids directly to the spinal or epidural space adjacent to the spinal cord, regional analgesia can be obtained using relatively low total doses. Indeed, the dose required to produce effective localized analgesia when using morphine intrathecally (0.1–0.3 mg) is a fraction of that required to produce similar analgesia when administered intravenously (5–10 mg). In this way, side effects such as sedation, nausea, and respiratory depression can be minimized. This approach has been used extensively in obstetric procedures and for postoperative pain relief following surgical procedures on the lower extremities. Continuous intrathecal delivery via implanted spinal drug-delivery systems is now commonly used, particularly for the treatment of cancer-related pain that would require sedating doses for adequate pain control if given systemically. Opioids can also be given intranasally (butorphanol), rectally, and transdermally (fentanyl), thus avoiding the discomfort of frequent injections in patients who cannot be given oral medication. The fentanyl transdermal patch has the advantage of providing fairly steady plasma levels, which maximizes patient comfort.

Recent additions to the armamentarium for treating opioid-induced side effects are the peripherally acting opioid antagonists alvimopan (Entereg) and methylnaltrexone (Rellistor). Alvimopan is available as an orally administered agent that is restricted to the intestinal lumen by limited absorption; methylnaltrexone is available in a subcutaneously administered form that has virtually no penetration into the CNS. Both agents act by binding to peripheral μ-receptors, thereby inhibiting or reversing the effects of opioids at these peripheral sites. The action of both agents is restricted to receptor sites outside of the CNS; thus, these drugs can reverse the adverse effects of opioid analgesics that are mediated through their peripheral receptors without reversing their analgesic effects. Both agents are effective for persistent ileus following abdominal surgery to the extent that opioid analgesics used for postoperative pain control contribute to this serious problem. Likewise, both agents have been tested for their effectiveness in treating opioid-induced bowel dysfunction (constipation) in patients taking opioid analgesics on a chronic basis. Although contradictory, the weight of evidence indicates that alvimopan can reduce the incidence and duration of ileus following major abdominal surgery and methylnaltrexone can produce rapid reversal of constipation in many patients receiving opioids on a chronic basis.

Opioid and COX Inhibitor Combinations

When used in combination, opioids and COX inhibitors have additive effects. Because a lower dose of each can be used to achieve the same degree of pain relief and their side effects are nonadditive, such combinations are used to lower the severity of dose-related side effects. However, fixed-ratio combinations of an opioid with acetaminophen carry a special risk. Dose escalation as a result of increased severity of pain or decreased opioid effect as a result of tolerance may lead to levels of acetaminophen that are toxic to the liver. Although acetaminophen-related hepatotoxicity is uncommon, it remains a leading cause for liver failure. Thus, many practitioners have moved away from the use of opioid-acetaminophen combination analgesics to avoid the risk of excessive acetaminophen exposure as the dose of the analgesic is escalated.

CHRONIC PAIN

Managing patients with chronic pain is intellectually and emotionally challenging. The patient's problem is often difficult or impossible to diagnose with certainty; such patients are demanding of the physician's time and often appear emotionally distraught. The traditional medical approach of seeking an obscure organic pathology is usually unhelpful. On the other hand, psychological evaluation and behaviorally based treatment paradigms are frequently helpful, particularly in the setting of a multidisciplinary pain-management center. Unfortunately, this approach, while effective, remains largely underused in current medical practice.

There are several factors that can cause, perpetuate, or exacerbate chronic pain. First, of course, the patient may simply have a disease that is characteristically painful for which there is presently no cure. Arthritis, cancer, chronic daily headaches, fibromyalgia, and diabetic neuropathy are examples of this. Second, there may be secondary perpetuating factors that are initiated by disease and persist after that disease has resolved. Examples include damaged sensory nerves, sympathetic efferent activity, and painful reflex muscle contraction. Finally, a variety of psychological conditions can exacerbate or even cause pain.

There are certain areas to which special attention should be paid in a patient's medical history. Because depression is the most common emotional disturbance in patients with chronic pain, patients should be questioned about their mood, appetite, sleep patterns, and daily activity. A simple standardized questionnaire, such as the Beck Depression Inventory, can be a useful screening device. It is important to remember that major depression is a common, treatable, and potentially fatal illness.

Other clues that a significant emotional disturbance is contributing to a patient's chronic pain complaint include pain that occurs in multiple, unrelated sites; a pattern of recurrent, but separate, pain problems beginning in childhood or adolescence; pain beginning at a time of emotional trauma, such as the loss of a parent or spouse; a history of physical or sexual abuse; and past or present substance abuse.

On examination, special attention should be paid to whether the patient guards the painful area and whether certain movements or postures are avoided because of pain. Discovering a mechanical component to the pain can be useful both diagnostically and therapeutically. Painful areas should be examined for deep tenderness, noting whether this is localized to muscle, ligamentous structures, or joints. Chronic myofascial pain is very common, and, in these patients, deep palpation may reveal highly localized trigger points that are firm bands or knots in muscle. Relief of the pain following injection of local anesthetic into these trigger points supports the diagnosis. A neuropathic component to the pain is indicated by evidence of nerve damage, such as sensory impairment, exquisitely sensitive skin, weakness, and muscle atrophy, or loss of deep tendon reflexes. Evidence suggesting sympathetic nervous system involvement includes the presence of diffuse swelling, changes in skin color and temperature, and hypersensitive skin and joint tenderness compared with the normal side. Relief of the pain with a sympathetic block is diagnostic.

A guiding principle in evaluating patients with chronic pain is to assess both emotional and organic factors before initiating therapy. Addressing these issues together, rather than waiting to address emotional issues after organic causes of pain have been ruled out, improves compliance in part because it assures patients that a psychological evaluation does not mean that the physician is questioning the validity of their complaint. Even when an organic cause for a patient's pain can be found, it is still wise to look for other factors. For example, a cancer patient with painful bony metastases may have additional pain due to nerve damage and may also be depressed. Optimal therapy requires that each of these factors be looked for and treated.

Treatment: Chronic Pain

Once the evaluation process has been completed and the likely causative and exacerbating factors identified, an explicit treatment plan should be developed. An important part of this process is to identify specific and realistic functional goals for therapy, such as getting a good night's sleep, being able to go shopping, or returning to work. A multidisciplinary approach that uses medications, counseling, physical therapy, nerve blocks, and even surgery may be required to improve the patient's quality of life. There are also some newer, relatively invasive procedures that can be helpful for some patients with intractable pain. These include image-guided interventions such as epidural injection of

glucocorticoids for acute radicular pain, radiofrequency treatment of the facet joints for chronic facet-related pain, percutaneous intradiscal treatments for both axial and radicular pain, and placement of implanted intraspinal electrodes and implantation of intrathecal drug-delivery systems for severe and persistent pain that is unresponsive to more conservative treatments. There are no set criteria for predicting which patients will respond to these procedures. They are generally reserved for patients who have not responded to conventional pharmacologic approaches. Referral to a multidisciplinary pain clinic for a full evaluation should precede any invasive procedures. Such referrals are clearly not necessary for all chronic pain patients. For some, pharmacologic management alone can provide adequate relief.

ANTIDEPRESSANT MEDICATIONS

The tricyclic antidepressants (TCAs), particularly nortriptyline and desipramine (Table 11-1) are useful for the management of chronic pain. Although developed for the treatment of depression, the TCAs have a spectrum of dose-related biologic activities that include analgesia in a variety of chronic clinical conditions. Although the mechanism is unknown, the analgesic effect of TCAs has a more rapid onset and occurs at a lower dose than is typically required for the treatment of depression. Furthermore, patients with chronic pain who are not depressed obtain pain relief with antidepressants. There is evidence that TCAs potentiate opioid analgesia, so they may be useful adjuncts for the treatment of severe persistent pain such as occurs with malignant tumors. **Table 11-2** lists some of the painful conditions that respond to TCAs. TCAs are of particular value in the management of neuropathic pain such as occurs in diabetic neuropathy and postherpetic neuralgia, for which there are few other therapeutic options.

Table 11-2 Painful Conditions that Respond to Tricyclic Antidepressants

Postherpetic neuralgia[a]
Diabetic neuropathy[a]
Tension headache[a]
Migraine headache[a]
Rheumatoid arthritis[a,b]
Chronic low back pain[b]
Cancer Central post-stroke pain

[a]Controlled trials demonstrate analgesia.

[b]Controlled studies indicate benefit but not analgesia.

The TCAs that have been shown to relieve pain have significant side effects (Table 11-1; Chap. 390). Some of these side effects, such as orthostatic hypotension, drowsiness, cardiac-conduction delay, memory impairment, constipation, and urinary retention, are particularly problematic in elderly patients, and several are additive to the side effects of opioid analgesics. The selective serotonin reuptake inhibitors such as fluoxetine (Prozac) have fewer and less serious side effects than TCAs, but they are much less effective for relieving pain. It is of interest that venlafaxine (Effexor) and duloxetine (Cymbalta), which are nontricyclic antidepressants that block both serotonin and norepinephrine reuptake, appear to retain most of the pain-relieving effect of TCAs with a side-effect profile more like that of the selective serotonin reuptake inhibitors. These drugs may be particularly useful in patients who cannot tolerate the side effects of TCAs.

ANTICONVULSANTS AND ANTIARRHYTHMICS

These drugs are useful primarily for patients with neuropathic pain. Phenytoin (Dilantin) and carbamazepine (Tegretol) were first shown to relieve the pain of trigeminal neuralgia. This pain has a characteristic brief, shooting, electric shock–like quality. In fact, anticonvulsants seem to be particularly helpful for pains that have such a lancinating quality. Newer anticonvulsants, gabapentin (Neurontin) and pregabalin (Lyrica), are effective for a broad range of neuropathic pains. Furthermore, because of their favorable side-effect profile, these newer anticonvulsants are often used as first-line agents.

CHRONIC OPIOID MEDICATION

The long-term use of opioids is accepted for patients with pain due to malignant disease. Although opioid use for chronic pain of nonmalignant origin is controversial, it is clear that for many such

patients, opioid analgesics are the best available option. This is understandable because opioids are the most potent and have the broadest range of efficacy of any analgesic medications. Although addiction is rare in patients who first use opioids for pain relief, some degree of tolerance and physical dependence is likely with long-term use. Therefore, before embarking on opioid therapy, other options should be explored, and the limitations and risks of opioids should be explained to the patient. It is also important to point out that some opioid analgesic medications have mixed agonist-antagonist properties (e.g., pentazocine and butorphanol). From a practical standpoint, this means that they may worsen pain by inducing an abstinence syndrome in patients who are physically dependent on other opioid analgesics.

With long-term outpatient use of orally administered opioids, it is desirable to use long-acting compounds such as levorphanol, methadone, or sustained-release morphine (Table 11-1). Transdermal fentanyl is another excellent option. The pharmacokinetic profile of these drug preparations enables prolonged pain relief, minimizes side effects such as sedation that are associated with high peak plasma levels, and reduces the likelihood of rebound pain associated with a rapid fall in plasma opioid concentration. While long-acting opioid preparations may provide superior pain relief in patients with a continuous pattern of ongoing pain, others suffer from intermittent severe episodic pain and experience superior pain control and fewer side effects with the periodic use of short-acting opioid analgesics. Constipation is a virtually universal side effect of opioid use and should be treated expectantly. A recent advance for patients with chronic debilitating conditions is the development of methylnaltrexone, a peripherally acting mu opioid antagonist that blocks the constipation and itching associated with chronic opioid use without interfering with analgesia; the usual dose is 0.15 mg/kg of body weight given subcutaneously no more often than once daily.

TREATMENT OF NEUROPATHIC PAIN

It is important to individualize treatment for patients with neuropathic pain. Several general principles should guide therapy: the first is to move quickly to provide relief, and the second is to minimize drug side effects. For example, in patients with postherpetic neuralgia and significant cutaneous hypersensitivity, topical lidocaine (Lidoderm patches) can provide immediate relief without side effects. Anticonvulsants (gabapentin or pregabalin, see above) or antidepressants (nortriptyline, desipramine, duloxetine, or venlafaxine) can be used as first-line drugs for patients with neuropathic pain. Systemically administered antiarrhythmic drugs such as lidocaine and mexiletene are less likely to be effective; although intravenous infusion of lidocaine predictably provides analgesia in those with many forms of neuropathic pain, the relief is usually transient, typically lasting just hours after the cessation of the infusion. The oral lidocaine congener mexiletene is poorly tolerated, producing frequent gastrointestinal adverse effects. There is no consensus on which class of drug should be used as a first-line treatment for any chronically painful condition. However, because relatively high doses of anticonvulsants are required for pain relief, sedation is very common. Sedation is also a problem with TCAs but is much less of a problem with serotonin/norepinephrine reuptake inhibitors (SNRIs, e.g., venlafaxine and duloxetine). Thus, in the elderly or in those patients whose daily activities require high-level mental activity, these drugs should be considered the first line. In contrast, opioid medications should be used as a second- or third-line drug class. While highly effective for many painful conditions, opioids are sedating, and their effect tends to lessen over time, leading to dose escalation and, occasionally, a worsening of pain due to physical dependence. Drugs of different classes can be used in combination to optimize pain control.

It is worth emphasizing that many patients, especially those with chronic pain, seek medical attention primarily because they are suffering and because only physicians can provide the medications required for pain relief. A primary responsibility of all physicians is to minimize the physical and emotional discomfort of their patients. Familiarity with pain mechanisms and analgesic medications is an important step toward accomplishing this aim.

FURTHER READINGS

Costigan M: Neuropathic pain: A maladaptive response of the nervous system to damage. Annu Rev Neurosci 32:1, 2009 [PMID: 19400724]

Craig AD: How do you feel? Interoception: The sense of the physiological condition of the body. Nat Rev Neurosci 8:655, 2002

Dworkin RH: Pharmacologic management of neuropathic pain: Evidence-based recommendations. Pain 123:237, 2007

Eippert F et al: Activation of the opioidergic descending pain control system underlies placebo analgesia. Neuron 63:533, 2009 [PMID: 19709634]

Fields HL: Should we be reluctant to prescribe opioids for chronic nonmalignant pain? Pain 129:233, 2007 [PMID: 17449177]

Macintyre PE: Safety and efficacy of patient-controlled analgesia. Br J Anaesth 87:36, 2001 [PMID: 11460812]

Oaklander AL: Is reflex sympathetic dystrophy/complex regional pain syndrome type I a small-fiber neuropathy? Ann Neurol 64:629, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHEST DISCOMFORT: INTRODUCTION

Chest discomfort is a common challenge for clinicians in the office or emergency department. The differential diagnosis includes conditions affecting organs throughout the thorax and abdomen, with prognostic implications that vary from benign to life-threatening **(Table 12-1)**. Failure to recognize potentially serious conditions such as acute ischemic heart disease, aortic dissection, tension pneumothorax, or pulmonary embolism can lead to serious complications, including death. Conversely, overly conservative management of low-risk patients leads to unnecessary hospital admissions, tests, procedures, and anxiety.

Table 12-1 Diagnoses among Chest Pain Patients Without Myocardial Infarction

Diagnosis	Percent
Gastroesophageal disease[a]	42
Gastroesophageal reflux	
Esophageal motility disorders	
Peptic ulcer	
Gallstones	
Ischemic heart disease	31
Chest wall syndromes	28
Pericarditis	4
Pleuritis/pneumonia	2
Pulmonary embolism	2
Lung cancer	1.5
Aortic aneurysm	1
Aortic stenosis	1
Herpes zoster	1

[a]In order of frequency.

Source: P Fruergaard et al: Eur Heart J 17:1028, 1996.

Causes of Chest Discomfort

MYOCARDIAL ISCHEMIA AND INJURY

Myocardial ischemia occurs when the oxygen supply to the heart is insufficient to meet metabolic needs. This mismatch can result from a decrease in oxygen supply, a rise in demand, or both. The most common underlying cause of myocardial ischemia is obstruction of coronary arteries by atherosclerosis; in the presence of such obstruction, transient ischemic episodes are usually precipitated by an increase in oxygen demand as a result of physical exertion. However, ischemia can also result from psychological stress, fever, or large meals or from compromised oxygen delivery due to anemia, hypoxia, or hypotension. Ventricular hypertrophy due to valvular heart disease, hypertrophic cardiomyopathy, or hypertension can predispose the myocardium to ischemia because of impaired penetration of blood flow from epicardial coronary arteries to the endocardium.

Angina Pectoris

(See also Chap. 243) The chest discomfort of myocardial ischemia is a visceral discomfort that is usually described as a heaviness, pressure, or squeezing **(Table 12-2)**. Other common adjectives for anginal pain are burning and aching. Some patients deny any "pain" but may admit to dyspnea or a vague sense of anxiety. The word "sharp" is sometimes used by patients to describe intensity rather than quality.

Table 12-2 Typical Clinical Features of Major Causes of Acute Chest Discomfort

Condition	Duration	Quality	Location	Associated Features
Angina	More than 2 and less	Pressure, tightness,	Retrosternal, often with radiation to or isolated	Precipitated by exertion, exposure to cold,

	than 10 min	squeezing, heaviness, burning	discomfort in neck, jaw, shoulders, or arms —frequently on left	psychologic stress S_4 gallop or mitral regurgitationmurmur during pain
Unstable angina	10–20 min	Similar to angina but often more severe	Similar to angina	Similar to angina, but occurs with low levels of exertion or even at rest
Acute myocardial infarction	Variable; often more than 30 min	Similar to angina but often more severe	Similar to angina	Unrelieved by nitroglycerin May be associated with evidence of heart failure or arrhythmia
Aortic stenosis	Recurrent episodes as described for angina	As described for angina	As described for angina	Late-peaking systolic murmurradiating to carotid arteries
Pericarditis	Hours to days; may be episodic	Sharp	Retrosternal or toward cardiac apex; may radiate to left shoulder	May be relieved by sitting up andleaning forward Pericardial friction rub
Aortic dissection	Abrupt onset of unrelenting pain	Tearing or rippingsensation; knifelike	Anterior chest, often radiating to back, between shoulder blades	Associated with hypertension and/or underlying connective tissue disorder, e.g., Marfan syndrome Murmur of aortic insufficiency, pericardial rub, pericardial tamponade, or loss of peripheral pulses
Pulmonary embolism	Abrupt onset; several minutes to a few hours	Pleuritic	Often lateral, on the side of the embolism	Dyspnea, tachypnea, tachycardia, and hypotension
Pulmonary hypertension	Variable	Pressure	Substernal	Dyspnea, signs of increased venous pressure including edema and jugular venous distention
Pneumonia or pleuritis	Variable	Pleuritic	Unilateral, often localized	Dyspnea, cough, fever, rales, occasional rub
Spontaneous pneumothorax	Sudden onset; several hours	Pleuritic	Lateral to side of pneumothorax	Dyspnea, decreased breath sounds on side of pneumothorax
Esophageal reflux	10–60 min	Burning	Substernal, epigastric	Worsened by postprandial recumbency Relieved by antacids
Esophageal spasm	2–30 min	Pressure, tightness, burning	Retrosternal	Can closely mimic angina
Peptic ulcer	Prolonged	Burning	Epigastric, substernal	Relieved with food or antacids
Gallbladder disease	Prolonged	Burning, pressure	Epigastric, right upper quadrant, substernal	May follow meal
Musculoskeletal disease	Variable	Aching	Variable	Aggravated by movement May be reproduced by localized pressure on examination
Herpes zoster	Variable	Sharp or burning	Dermatomal distribution	Vesicular rash in area of discomfort
Emotional and psychiatricconditions	Variable; may be fleeting	Variable	Variable; may be retrosternal	Situational factors may precipitate symptoms Anxiety or depression often detectable with careful history

The location of angina pectoris is usually retrosternal; most patients do not localize the pain to any small area. The discomfort may radiate to the neck, jaw, teeth, arms, or shoulders, reflecting the common origin in the posterior horn of the spinal cord of sensory neurons supplying the heart and these areas. Some patients present

with aching in sites of radiated pain as their only symptoms of ischemia. Occasional patients report epigastric distress with ischemic episodes. Less common is radiation to below the umbilicus or to the back.

Stable angina pectoris usually develops gradually with exertion, emotional excitement, or after heavy meals. Rest or treatment with sublingual nitroglycerin typically leads to relief within several minutes. In contrast, pain that is fleeting (lasting only a few seconds) is rarely ischemic in origin. Similarly, pain that lasts for several hours is unlikely to represent angina, particularly if the patient's electrocardiogram (ECG) does not show evidence of ischemia.

Anginal episodes can be precipitated by any physiologic or psychological stress that induces tachycardia. Most myocardial perfusion occurs during diastole, when there is minimal pressure opposing coronary artery flow from within the left ventricle. Since tachycardia decreases the percentage of the time in which the heart is in diastole, it decreases myocardial perfusion.

Unstable Angina and Myocardial Infarction

(See also Chaps. 244 and 245) Patients with these acute ischemic syndromes usually complain of symptoms similar in quality to angina pectoris, but more prolonged and severe. The onset of these syndromes may occur with the patient at rest, or awakened from sleep, and sublingual nitroglycerin may lead to transient or no relief. Accompanying symptoms may include diaphoresis, dyspnea, nausea, and light-headedness.

The physical examination may be completely normal in patients with chest discomfort due to ischemic heart disease. Careful auscultation during ischemic episodes may reveal a third or fourth heart sound, reflecting myocardial systolic or diastolic dysfunction. A transient murmur of mitral regurgitation suggests ischemic papillary muscle dysfunction. Severe episodes of ischemia can lead to pulmonary congestion and even pulmonary edema.

Other Cardiac Causes

Myocardial ischemia caused by hypertrophic cardiomyopathy or aortic stenosis leads to angina pectoris similar to that caused by coronary atherosclerosis. In such cases, a loud systolic murmur or other findings usually suggest that abnormalities other than coronary atherosclerosis may be contributing to the patient's symptoms. Some patients with chest pain and normal coronary angiograms have functional abnormalities of the coronary circulation, ranging from coronary spasm visible on coronary angiography to abnormal vasodilator responses and heightened vasoconstrictor responses. The term "cardiac syndrome X" is used to describe patients with angina-like chest pain and ischemic-appearing ST-segment depression during stress despite normal coronary arteriograms. Some data indicate that many such patients have limited changes in coronary flow in response to pacing stress or coronary vasodilators.

PERICARDITIS

(See also Chap. 239) The pain in pericarditis is believed to be due to inflammation of the adjacent parietal pleura, since most of the pericardium is believed to be insensitive to pain. Thus, infectious pericarditis, which usually involves adjoining pleural surfaces, tends to be associated with pain, while conditions that cause only local inflammation (e.g., myocardial infarction or uremia) and cardiac tamponade tend to result in mild or no chest pain.

The adjacent parietal pleura receives its sensory supply from several sources, so the pain of pericarditis can be experienced in areas ranging from the shoulder and neck to the abdomen and back. Most typically, the pain is retrosternal and is aggravated by coughing, deep breaths, or changes in position—all of which lead to movements of pleural surfaces. The pain is often worse in the supine position and relieved by sitting upright and leaning forward. Less common is a steady aching discomfort that mimics acute myocardial infarction.

DISEASES OF THE AORTA

(See also Chap. 248) *Aortic dissection* is a potentially catastrophic condition that is due to spread of a subintimal hematoma within the wall of the aorta. The hematoma may begin with a tear in the intima of the aorta or with rupture of the vasa vasorum within the aortic media. This syndrome can occur with trauma to the aorta, including motor vehicle accidents or medical procedures in which catheters or intraaortic balloon pumps damage the intima of the aorta. Nontraumatic aortic dissections are rare in the absence of hypertension and/or conditions associated with deterioration of the elastic or muscular components of the media within the aorta's wall. Cystic medial degeneration is a feature of several inherited connective tissue diseases, including Marfan and Ehlers-Danlos syndromes. About half of all aortic dissections in women under 40 years of age occur during pregnancy.

Almost all patients with acute dissections present with severe chest pain, although some patients with chronic dissections are identified without associated symptoms. Unlike the pain of ischemic heart disease, symptoms of aortic dissection tend to reach peak severity immediately, often causing the patient to collapse from its intensity. The classic teaching is that the adjectives used to describe the pain reflect the process occurring within the wall of the aorta—"ripping" and "tearing"—but more recent data suggest that the most common presenting complaint is sudden onset of severe, sharp pain. The location often correlates with the site and extent of the dissection. Thus, dissections that begin in the ascending aorta and extend to the descending aorta tend to cause pain in the front of the chest that extends into the back, between the shoulder blades.

Physical findings may also reflect extension of the aortic dissection that compromises flow into arteries branching off the aorta. Thus, loss of a pulse in one or both arms, cerebrovascular accident, or paraplegia can all be catastrophic consequences of aortic dissection. Hematomas that extend proximally and undermine the coronary arteries or aortic valve apparatus may lead to acute myocardial infarction or acute aortic insufficiency. Rupture of the hematoma into the pericardial space leads to pericardial tamponade.

Another abnormality of the aorta that can cause chest pain is a *thoracic aortic aneurysm*. Aortic aneurysms are frequently asymptomatic but can cause chest pain and other symptoms by compressing adjacent structures. This pain tends to be steady, deep, and sometimes severe.

PULMONARY EMBOLISM

(See also Chap. 262) Chest pain due to pulmonary embolism is believed to be due to distention of the pulmonary

artery or infarction of a segment of the lung adjacent to the pleura. Massive pulmonary emboli may lead to substernal pain that is suggestive of acute myocardial infarction. More commonly, smaller emboli lead to focal pulmonary infarctions that cause pain that is lateral and pleuritic. Associated symptoms include dyspnea and, occasionally, hemoptysis. Tachycardia is usually present. Although not always present, certain characteristic ECG changes can support the diagnosis.

PNEUMOTHORAX

(See also Chap. 263) Sudden onset of pleuritic chest pain and respiratory distress should lead to consideration of spontaneous pneumothorax, as well as pulmonary embolism. Such events may occur without a precipitating event in persons without lung disease, or as a consequence of underlying lung disorders.

PNEUMONIA OR PLEURITIS

(See also Chaps. 257 and 263) Lung diseases that damage and cause inflammation of the pleura of the lung usually cause a sharp, knifelike pain that is aggravated by inspiration or coughing.

GASTROINTESTINAL CONDITIONS

(See also Chap. 292) Esophageal pain from acid reflux from the stomach, spasm, obstruction, or injury can be difficult to differentiate from myocardial syndromes. Acid reflux typically causes a deep burning discomfort that may be exacerbated by alcohol, aspirin, or some foods; this discomfort is often relieved by antacid or other acid-reducing therapies. Acid reflux tends to be exacerbated by lying down and may be worse in early morning when the stomach is empty of food that might otherwise absorb gastric acid.

Esophageal spasm may occur in the presence or absence of acid reflux and leads to a squeezing pain indistinguishable from angina. Prompt relief of esophageal spasm is often provided by antianginal therapies such as sublingual nifedipine, further promoting confusion between these syndromes. Chest pain can also result from injury to the esophagus, such as a Mallory-Weiss tear caused by severe vomiting.

Chest pain can result from diseases of the gastrointestinal tract below the diaphragm, including *peptic ulcer disease*, *biliary disease*, and *pancreatitis*. These conditions usually cause abdominal pain as well as chest discomfort; symptoms are not likely to be associated with exertion. The pain of ulcer disease typically occurs 60 to 90 min after meals, when postprandial acid production is no longer neutralized by food in the stomach. Cholecystitis usually causes a pain that is described as aching, occurring an hour or more after meals.

NEUROMUSCULOSKELETAL CONDITIONS

Cervical disk disease can cause chest pain by compression of nerve roots. Pain in a dermatomal distribution can also be caused by *intercostal muscle cramps* or by *herpes zoster*. Chest pain symptoms due to herpes zoster may occur before skin lesions are apparent.

Costochondral and *chondrosternal syndromes* are the most common causes of anterior chest musculoskeletal pain. Only occasionally are physical signs of costochondritis such as swelling, redness, and warmth (Tietze's syndrome) present. The pain of such syndromes is usually fleeting and sharp, but some patients experience a dull ache that lasts for hours. Direct pressure on the chondrosternal and costochondral junctions may reproduce the pain from these and other musculoskeletal syndromes. Arthritis of the shoulder and spine and bursitis may also cause chest pain. Some patients who have these conditions and myocardial ischemia blur and confuse symptoms of these syndromes.

EMOTIONAL AND PSYCHIATRIC CONDITIONS

As many as 10% of patients who present to emergency departments with acute chest discomfort have panic disorder or other emotional conditions. The symptoms in these populations are highly variable, but frequently the discomfort is described as visceral tightness or aching that lasts more than 30 min. Some patients offer other atypical descriptions, such as pain that is fleeting, sharp, and/or localized to a small region. The ECG in patients with emotional conditions may be difficult to interpret if hyperventilation causes ST-T-wave abnormalities. A careful history may elicit clues of depression, prior panic attacks, somatization, agoraphobia, or other phobias.

APPROACH TO THE PATIENT: CHEST DISCOMFORT

The evaluation of the patient with chest discomfort must accommodate two goals—determining the diagnosis and assessing the safety of the immediate management plan. The latter issue is often dominant when the patient has acute chest discomfort, such as patients seen in the emergency department. In such settings, the clinician must focus first on identifying patients who require aggressive interventions to diagnose or manage potentially life-threatening conditions, including acute ischemic heart disease, acute aortic dissection, pulmonary embolism, and tension pneumothorax. If such conditions are unlikely, the clinician must address questions such as the safety of discharge to home, admission to a non-coronary care unit facility, or immediate exercise testing. **Table 12-3** displays a sequence of questions that can be used in the evaluation of the patient with chest discomfort, with the diagnostic entities that are most important for consideration at each stage of the evaluation.

Table 12-3 Considerations in the Assessmentof the Patient with Chest Discomfort

1. Could the chest discomfort be due to an acute, potentially life-threatening condition that warrants immediate hospitalization and aggressive evaluation?	
Acute ischemic heart disease	Pulmonary embolism
Aortic dissection	Spontaneous pneumothorax
2. If not, could the discomfort be due to a chronic condition likely to lead to serious complications?	
Stable angina Aortic stenosis	
Pulmonary hypertension	

3. If not, could the discomfort be due to an acute condition that warrants specific treatment?	
Pericarditis Pneumonia/pleuritis Herpes zoster	
4. If not, could the discomfort be due to another treatable chronic condition?	
Esophageal reflux Esophageal spasm	Cervical disk disease Arthritis of the shoulderor spine
Peptic ulcer disease Gallbladder disease Other gastrointestinal conditions	Costochondritis Other musculoskeletal disorders Anxiety state

Acute Chest Discomfort

In patients with acute chest discomfort, the clinician must first assess the patient's respiratory and hemodynamic status. If either is compromised, initial management should focus on stabilizing the patient before the diagnostic evaluation is pursued. If, however, the patient does not require emergent interventions, then a focused history, physical examination, and laboratory evaluation should be performed to assess the patient's risk of life-threatening conditions.

Clinicians who are seeing patients in the office setting should not assume that they do not have acute ischemic heart disease, even if the prevalence may be lower. Malpractice litigation related to myocardial infarctions that were missed during office evaluations is becoming increasingly common, and ECGs were not performed in many such cases. The prevalence of high-risk patients seen in office settings may be increasing due to congestion in emergency departments.

In either setting, the *history* should include questions about the quality and location of the chest discomfort (Table 12-2). The patient should also be asked about the nature of onset of the pain and its duration. Myocardial ischemia is usually associated with a gradual intensification of symptoms over a period of minutes. Pain that is fleeting or that lasts hours without being associated with electrocardiographic changes is not likely to be ischemic in origin. Although the presence of risk factors for coronary artery disease may heighten concern for this diagnosis, the absence of such risk factors does not lower the risk for myocardial ischemia enough to be used to justify a decision to discharge a patient.

Wide radiation of chest pain increases probability that pain is due to myocardial infarction. Radiation of chest pain to the left arm is common with acute ischemic heart disease, but radiation to the right arm is also highly consistent with this diagnosis. **Figure 12-1** shows estimates derived from several studies of the impact of various clinical features from the history on the probability that a patient has an acute myocardial infarction.

Figure 12-1

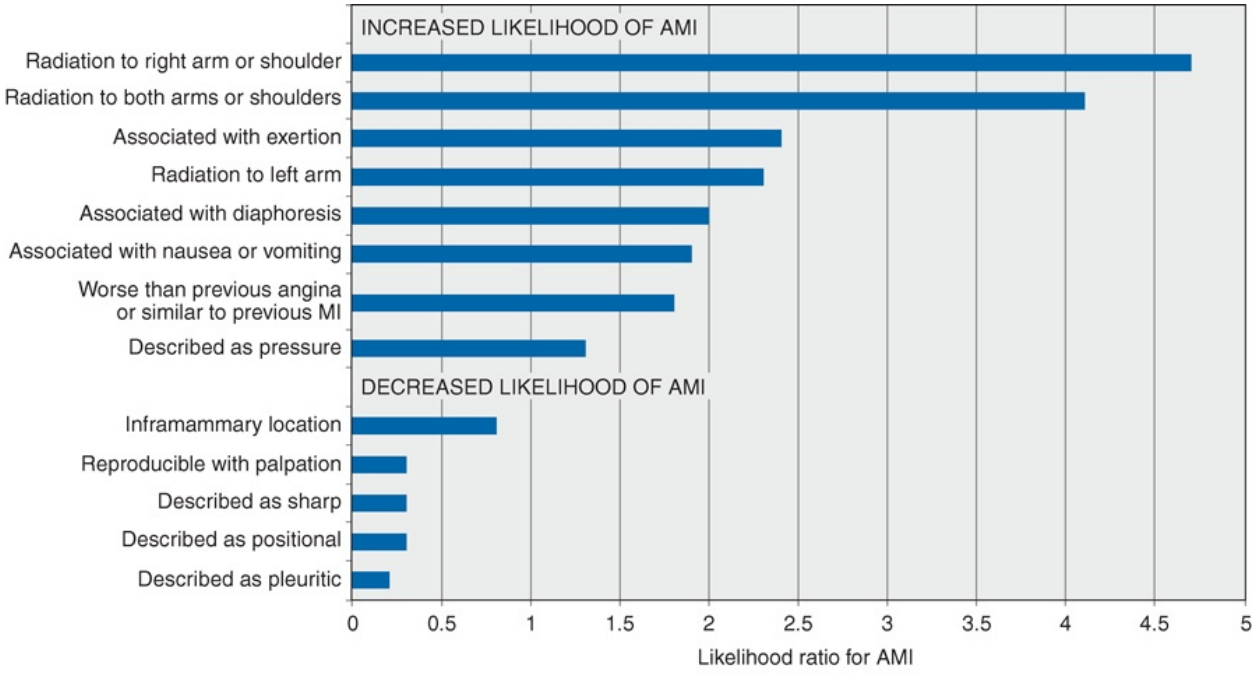

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Impact of chest pain characteristics on odds of acute myocardial infarction (AMI). *(Figure prepared from data in Swap and Nagurney.)*

Right shoulder pain is also common with acute cholecystitis, but this syndrome is usually accompanied by pain that is located in the abdomen rather than chest. Chest pain that radiates between the scapulae raises the

question of aortic dissection.

The *physical examination* should include evaluation of blood pressure in both arms and of pulses in both legs. Poor perfusion of a limb may be due to an aortic dissection that has compromised flow to an artery branching from the aorta. Chest auscultation may reveal diminished breath sounds; a pleural rub; or evidence of pneumothorax, pulmonary embolism, pneumonia, or pleurisy. Tension pneumothorax may lead to a shift in the trachea from the midline, away from the side of the pneumothorax. The cardiac examination should seek pericardial rubs, systolic and diastolic murmurs, and third or fourth heart sounds. Pressure on the chest wall may reproduce symptoms in patients with musculoskeletal causes of chest pain; it is important that the clinician ask the patient if the chest pain syndrome is being completely reproduced before drawing too much reassurance that more serious underlying conditions are not present.

An *ECG* is an essential test for adults with chest discomfort that is not due to an obvious traumatic cause. In such patients, the presence of electrocardiographic changes consistent with ischemia or infarction (Chap. 228) is associated with high risks of acute myocardial infarction or unstable angina **(Table 12-4)**; such patients should be admitted to a unit with electrocardiographic monitoring and the capacity to respond to a cardiac arrest. The absence of such changes does not exclude acute ischemic heart disease, but the risk of life-threatening complications is low for patients with normal electrocardiograms or only nonspecific ST-T-wave changes. If these patients are not considered appropriate for immediate discharge, they are often candidates for early or immediate exercise testing.

Table 12-4 Prevalence of Acute Ischemic Heart Disease Syndromes among Subsets of Emergency Department Patients with Chest Pain

Finding	Prevalence Myocardial Infarction, %	Unstable Angina, %
ST elevation (≥1 mm) or Q waves on ECG not known to be old	79	12
Ischemia or strain on ECG not known to be old (ST depression ≥1 mm or ischemic T waves)	20	41
None of the preceding ECG changes but a prior history of angina or myocardial infarction (history of heart attack or nitroglycerin use)	4	51
None of the preceding ECG changes and no prior history of angina or myocardial infarction (history of heart attack or nitroglycerin use)	2	14

Abbreviation: ECG, electrocardiogram.

Unpublished data from Brigham and Women's Hospital Chest Pain Study, 1997–1999.

Markers of myocardial injury are often obtained in the emergency department evaluation of acute chest discomfort. In recent years, the cardiac troponins (I and T) have superceded creatine kinase (CK) and CK-MB as the markers of choice for detecting myocardial injury. Some data support the use of other markers, such as myeloperoxidase and B-type natriuretic peptide (BNP), but their roles in routine care have not been established. Single values of any of these markers do not have high sensitivity for acute myocardial infarction or for prediction of complications. Hence, decisions to discharge patients home should not be made on the basis of single negative values of these tests, including the cardiac troponins.

Provocative tests for coronary artery disease are not appropriate for patients with ongoing chest pain. In such patients, rest myocardial perfusion scans can be considered; a normal scan reduces the likelihood of coronary artery disease and can help avoid admission of low-risk patients to the hospital. Computerized tomographic angiography (CTA) is emerging as an alternative diagnostic strategy for patients in whom the likelihood of coronary disease is not clear.

Clinicians frequently employ therapeutic trials with sublingual nitroglycerin or antacids or, in the stable patient seen in the office setting, a proton pump inhibitor. A common error is to assume that a response to any of these interventions clarifies the diagnosis. While such information is often helpful, the patient's response may be due to the placebo effect. Hence, myocardial ischemia should never be considered excluded solely because of a response to antacid therapy. Similarly, failure of nitroglycerin to relieve pain does not exclude the diagnosis of coronary disease.

If the patient's history or examination is consistent with aortic dissection, imaging studies to evaluate the aorta must be pursued promptly because of the high risk of catastrophic complications with this condition. Appropriate tests include a chest CT scan with contrast, MRI, or transesophageal echocardiography. Current data indicate that elevated d-dimer levels should raise clinicians' suspicion of aortic dissection.

Acute pulmonary embolism should be considered in patients with respiratory symptoms, pleuritic chest pain, hemoptysis, or a history of venous thromboembolism or coagulation abnormalities. Initial tests usually include CT angiography or a lung scan, which are sometimes combined with lower extremity venous ultrasound or D-dimer testing.

If patients with acute chest discomfort show no evidence of life-threatening conditions, the clinician should then focus on serious chronic conditions with the potential to cause major complications, the most common of which is stable angina. Early use of exercise electrocardiography, stress echocardiography, or stress perfusion imaging for such patients, whether in the office or the emergency department, is now an accepted management strategy for low-risk patients. Exercise testing is not appropriate, however, for patients who (1) report pain that is believed to be ischemic occurring at rest or (2) have electrocardiographic changes not known to be old that are consistent with ischemia.

Patients with sustained chest discomfort who do not have evidence for life-threatening conditions should be

evaluated for evidence of conditions likely to benefit from acute treatment (Table 12-3). Pericarditis may be suggested by the history, physical examination, and ECG (Table 12-2). Clinicians should carefully assess blood pressure patterns and consider echocardiography in such patients to detect evidence of impending pericardial tamponade. Chest x-rays can be used to evaluate the possibility of pulmonary disease.

Guidelines and Critical Pathwaysfor Acute Chest Discomfort

Guidelines for the initial evaluation for patients with acute chest pain have been developed by the American College of Cardiology, American Heart Association, and other organizations. These guidelines recommend performance of an ECG for virtually all patients with chest pain who do not have an obvious noncardiac cause of their pain, and performance of a chest x-ray for patients with signs or symptoms consistent with congestive heart failure, valvular heart disease, pericardial disease, or aortic dissection or aneurysm.

The American College of Cardiology/American Heart Association guidelines on exercise testing support its use in low-risk patients presenting to the emergency department, as well as in selected intermediate-risk patients. However, these guidelines emphasize that exercise tests should be performed only after patients have been screened for high-risk features or other indicators for hospital admission.

Many medical centers have adopted critical pathways and other forms of guidelines to increase efficiency and to expedite the treatment of patients with high-risk acute ischemic heart disease syndromes. These guidelines emphasize the following strategies:

- Rapid identification and treatment of patients for whom emergent reperfusion therapy, either via percutaneous coronary interventions or thrombolytic agents, is likely to lead to improved outcomes.
- Triage to non-coronary care unit monitored facilities such as intermediate-care units or chest pain units of patients with a low risk for complications, such as patients without new ischemic changes on their ECGs and without ongoing chest pain. Such patients can usually be safely observed in non-coronary care unit settings, undergo early exercise testing, or be discharged home. Risk stratification can be assisted through use of prospectively validated multivariate algorithms that have been published for acute ischemic heart disease and its complications.
- Shortening lengths of stay in the coronary care unit and hospital. Recommendations regarding the minimum length of stay in a monitored bed for a patient who has no further symptoms have decreased in recent years to 12 h or less if exercise testing or other risk stratification technologies are available.

NONACUTE CHEST DISCOMFORT

The management of patients who do not require admission to the hospital or who no longer require inpatient observation should seek to identify the cause of the symptoms and the likelihood of major complications. Noninvasive tests for coronary disease serve both to diagnose this condition and to identify patients with high-risk forms of coronary disease who may benefit from revascularization. Gastrointestinal causes of chest pain can be evaluated via endoscopy or radiology studies, or with trials of medical therapy. Emotional and psychiatric conditions warrant appropriate evaluation and treatment; randomized trial data indicate that cognitive therapy and group interventions lead to decreases in symptoms for such patients.

FURTHER READINGS

Gibbons RJ et al: ACC/AHA 2002 guideline update for exercise testing. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). Available at *www.acc.org/ qualityandscience/clinical/guidelines/exercise/dirindex.htm*. Accessed on May 8, 2010

Leber AW et al: Quantification of obstructive and nonobstructive coronary lesions of 64-slice computed tomography. J Am Coll Cardiol 46:147, 2005 [PMID: 15992649]

Miller JM et al: Diagnostic performance of coronary angiography by 64-row CT. N Engl J Med 359:2324, 2008 [PMID: 19038879]

Reichlin T et al: Early diagnosis of myocardial infarction with sensitive cardiac troponin assays. N Engl J Med 361:858, 2009 [PMID: 19710484]

Sequist T, Lee TH: Prediction of missed myocardial infarction among symptomatic outpatients without coronary heart disease. Am Heart J 149:74, 2005 [PMID: 15660037]

Sequist TD et al: Missed opportunities in the primary care management of early acute ischemic heart disease. Arch Intern Med 166: 2237, 2006 [PMID: 17101942]

Suzuki T et al: Diagnosis of acute aortic dissection by d-dimer. Circulation 119:2702, 2009 [PMID: 19433758]

Swap CJ, Nagurney JT: Value and limitations of chest pain history in the evaluation of patients with suspected acute coronary syndromes. JAMA 294:2623, 2005 [PMID: 16304077]

Tong KL et al: Myocardial contrast echocardiography versus Thrombolysis in Myocardial Infarction score in patients presenting to the emergency department with chest pain and a nondiagnostic electrocardiogram. J Am Coll Cardiol 46:928, 2005

Tsai TT et al: Acute aortic syndromes. Circulation 205:3802, 2005

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ABDOMINAL PAIN: INTRODUCTION

The correct interpretation of acute abdominal pain is challenging. Few other clinical situations demand greater judgment, because the most catastrophic of events may be forecast by the subtlest of symptoms and signs. A meticulously executed, detailed history and physical examination are of the greatest importance. The etiologic classification in **Table 13-1**, although not complete, forms a useful basis for the evaluation of patients with abdominal pain.

Table 13-1 Some Important Causes of Abdominal Pain

Pain Originating in the Abdomen
Parietal peritoneal inflammation
Bacterial contamination
Perforated appendix or other perforated viscus
Pelvic inflammatory disease
Chemical irritation
Perforated ulcer
Pancreatitis
Mittelschmerz
Mechanical obstruction of hollow viscera
Obstruction of the small or large intestine
Obstruction of the biliary tree
Obstruction of the ureter
Vascular disturbances
Embolism or thrombosis
Vascular rupture
Pressure or torsional occlusion
Sickle cell anemia
Abdominal wall
Distortion or traction of mesentery
Trauma or infection of muscles
Distention of visceral surfaces, e.g., by hemorrhage
Hepatic or renal capsules
Inflammation of a viscus
Appendicitis
Typhoid fever
Typhlitis
Pain Referred from Extraabdominal Source
Cardiothoracic
Acute myocardial infarction
Myocarditis, endocarditis, pericarditis
Congestive heart failure
Pneumonia
Pulmonary embolus
Pleurodynia

Pneumothorax
Empyema
Esophageal disease, spasm, rupture, inflammation
Genitalia
Torsion of the testis
Metabolic Causes
Diabetes
Uremia
Hyperlipidemia
Hyperparathyroidism
Acute adrenal insufficiency
Familial Mediterranean fever
Porphyria
C'1 esterase inhibitor deficiency (angioneurotic edema)
Neurologic/Psychiatric Causes
Herpes zoster
Tabes dorsalis
Causalgia
Radiculitis from infection or arthritis
Spinal cord or nerve root compression
Functional disorders
Psychiatric disorders
Toxic Causes
Lead poisoning
Insect or animal envenomations
Black widow spiders
Snake bites
Uncertain Mechanisms
Narcotic withdrawal
Heat stroke

The diagnosis of "acute or surgical abdomen" is not an acceptable one because of its often misleading and erroneous connotation. The most obvious of "acute abdomens" may not require operative intervention, and the mildest of abdominal pains may herald an urgently correctable lesion. Any patient with abdominal pain of recent onset requires early and thorough evaluation and accurate diagnosis.

Some Mechanisms of Pain Originating in the Abdomen

INFLAMMATION OF THE PARIETAL PERITONEUM

The pain of parietal peritoneal inflammation is steady and aching in character and is located directly over the inflamed area, its exact reference being possible because it is transmitted by somatic nerves supplying the parietal peritoneum. The intensity of the pain is dependent on the type and amount of material to which the peritoneal surfaces are exposed in a given time period. For example, the sudden release into the peritoneal cavity of a small quantity of *sterile* acid gastric juice causes much more pain than the same amount of grossly contaminated neutral feces. Enzymatically active pancreatic juice incites more pain and inflammation than does the same amount of sterile bile containing no potent enzymes. Blood and urine are often so bland as to go undetected if their contact with the peritoneum has not been sudden and massive. In the case of bacterial contamination, such as in pelvic inflammatory disease, the pain is frequently of low intensity early in the illness until bacterial multiplication has caused the elaboration of irritating substances.

The rate at which the irritating material is applied to the peritoneum is important. Perforated peptic ulcer may be associated with entirely different clinical pictures dependent only on the rapidity with which the gastric juice enters the peritoneal cavity.

The pain of peritoneal inflammation is invariably accentuated by pressure or changes in tension of

the peritoneum, whether produced by palpation or by movement, as in coughing or sneezing. The patient with peritonitis lies quietly in bed, preferring to avoid motion, in contrast to the patient with colic, who may writhe incessantly.

Another characteristic feature of peritoneal irritation is tonic reflex spasm of the abdominal musculature, localized to the involved body segment. The intensity of the tonic muscle spasm accompanying peritoneal inflammation is dependent on the location of the inflammatory process, the rate at which it develops, and the integrity of the nervous system. Spasm over a perforated retrocecal appendix or perforated ulcer into the lesser peritoneal sac may be minimal or absent because of the protective effect of overlying viscera. A slowly developing process often greatly attenuates the degree of muscle spasm. Catastrophic abdominal emergencies such as a perforated ulcer may be associated with minimal or no detectable pain or muscle spasm in obtunded, seriously ill, debilitated elderly patients or in psychotic patients.

OBSTRUCTION OF HOLLOW VISCERA

The pain of obstruction of hollow abdominal viscera is classically described as intermittent, or colicky. Yet the lack of a truly cramping character should not be misleading, because distention of a hollow viscus may produce steady pain with only very occasional exacerbations. It is not nearly as well localized as the pain of parietal peritoneal inflammation.

The colicky pain of obstruction of the small intestine is usually periumbilical or supraumbilical and is poorly localized. As the intestine becomes progressively dilated with loss of muscular tone, the colicky nature of the pain may diminish. With superimposed strangulating obstruction, pain may spread to the lower lumbar region if there is traction on the root of the mesentery. The colicky pain of colonic obstruction is of lesser intensity than that of the small intestine and is often located in the infraumbilical area. Lumbar radiation of pain is common in colonic obstruction.

Sudden distention of the biliary tree produces a steady rather than colicky type of pain; hence, the term *biliary colic* is misleading. Acute distention of the gallbladder usually causes pain in the right upper quadrant with radiation to the right posterior region of the thorax or to the tip of the right scapula, but is not uncommonly midline. Distention of the common bile duct is often associated with pain in the epigastrium radiating to the upper part of the lumbar region. Considerable variation is common, however, so that differentiation between these may be impossible. The typical subscapular pain or lumbar radiation is frequently absent. Gradual dilatation of the biliary tree, as in carcinoma of the head of the pancreas, may cause no pain or only a mild aching sensation in the epigastrium or right upper quadrant. The pain of distention of the pancreatic ducts is similar to that described for distention of the common bile duct but, in addition, is very frequently accentuated by recumbency and relieved by the upright position.

Obstruction of the urinary bladder results in dull suprapubic pain, usually low in intensity. Restlessness without specific complaint of pain may be the only sign of a distended bladder in an obtunded patient. In contrast, acute obstruction of the intravesicular portion of the ureter is characterized by severe suprapubic and flank pain that radiates to the penis, scrotum, or inner aspect of the upper thigh. Obstruction of the ureteropelvic junction is felt as pain in the costovertebral angle, whereas obstruction of the remainder of the ureter is associated with flank pain that often extends into the same side of the abdomen.

VASCULAR DISTURBANCES

A frequent misconception, despite abundant experience to the contrary, is that pain associated with intraabdominal vascular disturbances is sudden and catastrophic in nature. The pain of embolism or thrombosis of the superior mesenteric artery or that of impending rupture of an abdominal aortic aneurysm certainly may be severe and diffuse. Yet, just as frequently, the patient with occlusion of the superior mesenteric artery has only mild continuous or cramping diffuse pain for two or three days before vascular collapse or findings of peritoneal inflammation appear. The early, seemingly insignificant discomfort is caused by hyperperistalsis rather than peritoneal inflammation. Indeed, absence of tenderness and rigidity in the presence of continuous, diffuse pain in a patient likely to have vascular disease is quite characteristic of occlusion of the superior mesenteric artery. Abdominal pain with radiation to the sacral region, flank, or genitalia should always signal the possible presence of a rupturing abdominal aortic aneurysm. This pain may persist over a period of several days before rupture and collapse occur.

ABDOMINAL WALL

Pain arising from the abdominal wall is usually constant and aching. Movement, prolonged standing, and pressure accentuate the discomfort and muscle spasm. In the case of hematoma of the rectus sheath, now most frequently encountered in association with anticoagulant therapy, a mass may be present in the lower quadrants of the abdomen. Simultaneous involvement of muscles in other parts of the body usually serves to differentiate myositis of the abdominal wall from an intraabdominal process that might cause pain in the same region.

Referred Pain in Abdominal Diseases

Pain referred to the abdomen from the thorax, spine, or genitalia may prove a vexing diagnostic problem, because diseases of the upper part of the abdominal cavity such as acute cholecystitis or perforated ulcer are frequently associated with intrathoracic complications. A most important, yet often forgotten, dictum is that the possibility of intrathoracic disease must be considered in every patient with abdominal pain, especially if the pain is in the upper part of the abdomen. Systematic questioning and examination directed toward detecting myocardial or pulmonary infarction, pneumonia, pericarditis, or esophageal disease (the intrathoracic diseases that most often masquerade as abdominal emergencies) will often provide sufficient clues to establish the proper diagnosis. Diaphragmatic pleuritis resulting from pneumonia or pulmonary infarction may cause pain in the right upper quadrant and pain in the supraclavicular area, the latter radiation to be distinguished from the referred subscapular pain caused by acute distention of the extrahepatic biliary tree. The ultimate decision as to the origin of abdominal pain may require deliberate and planned observation over a period of several hours, during which repeated questioning and examination will provide the diagnosis or suggest the appropriate studies.

Referred pain of thoracic origin is often accompanied by splinting of the involved hemithorax with respiratory lag and decrease in excursion more marked than that seen in the presence of intraabdominal disease. In addition, apparent abdominal muscle spasm caused by referred pain will diminish during the inspiratory phase of respiration, whereas it is persistent throughout both respiratory phases if it is of abdominal origin. Palpation over the area of referred pain in the abdomen also does not usually accentuate the pain and in many instances actually seems to relieve it. Thoracic disease and abdominal disease frequently coexist and may be difficult or impossible to differentiate. For example, the patient with known biliary tract disease often has epigastric pain during myocardial infarction, or biliary colic may be referred to the precordium or left shoulder in a patient who has suffered previously from angina pectoris. **For an explanation of the radiation of pain to a previously diseased area, see Chap. 11.**

Referred pain from the spine, which usually involves compression or irritation of nerve roots, is characteristically intensified by certain motions such as cough, sneeze, or strain and is associated with hyperesthesia over the involved dermatomes. Pain referred to the abdomen from the testes or seminal vesicles is generally accentuated by the slightest pressure on either of these organs. The abdominal discomfort is of dull, aching character and is poorly localized.

Metabolic Abdominal Crises

Pain of metabolic origin may simulate almost any other type of intraabdominal disease. Several mechanisms may be at work. In certain instances, such as hyperlipidemia, the metabolic disease itself may be accompanied by an intraabdominal process such as pancreatitis, which can lead to unnecessary laparotomy unless recognized. C'1 esterase deficiency associated with angioneurotic edema is often associated with episodes of severe abdominal pain. Whenever the cause of abdominal pain is obscure, a metabolic origin always must be considered. Abdominal pain is also the hallmark of familial Mediterranean fever (Chap. 330).

The problem of differential diagnosis is often not readily resolved. The pain of porphyria and of lead colic is usually difficult to distinguish from that of intestinal obstruction, because severe hyperperistalsis is a prominent feature of both. The pain of uremia or diabetes is nonspecific, and the pain and tenderness frequently shift in location and intensity. Diabetic acidosis may be precipitated by acute appendicitis or intestinal obstruction, so if prompt resolution of the abdominal pain does not result from correction of the metabolic abnormalities, an underlying organic problem should be suspected. Black widow spider bites produce intense pain and rigidity of the abdominal muscles and back, an area infrequently involved in intraabdominal disease.

Neurogenic Causes

Causalgic pain may occur in diseases that injure sensory nerves. It has a burning character and is usually limited to the distribution of a given peripheral nerve. Normal stimuli such as touch or change in temperature may be transformed into this type of pain, which is frequently present in a patient at rest. The demonstration of irregularly spaced cutaneous pain spots may be the only indication of an old nerve lesion underlying causalgic pain. Even though the pain may be precipitated by gentle palpation, rigidity of the abdominal muscles is absent, and the respirations are not disturbed. Distention of the abdomen is uncommon, and the pain has no relationship to the intake of food.

Pain arising from spinal nerves or roots comes and goes suddenly and is of a lancinating type (Chap. 15). It may be caused by herpes zoster, impingement by arthritis, tumors, herniated nucleus pulposus, diabetes, or syphilis. It is not associated with food intake, abdominal distention, or changes in respiration. Severe muscle spasm, as in the gastric crises of tabes dorsalis, is common but is either relieved or is not accentuated by abdominal palpation. The pain is made worse by movement of the spine and is usually confined to a few dermatomes. Hyperesthesia is very common.

Pain due to functional causes conforms to none of the aforementioned patterns. Mechanism is hard to define. Irritable bowel syndrome (IBS) is a functional gastrointestinal disorder characterized by abdominal pain and altered bowel habits. The diagnosis is made on the basis of clinical criteria (Chap.

296) and after exclusion of demonstrable structural abnormalities. The episodes of abdominal pain are often brought on by stress, and the pain varies considerably in type and location. Nausea and vomiting are rare. Localized tenderness and muscle spasm are inconsistent or absent. The causes of IBS or related functional disorders are not known.

APPROACH TO THE PATIENT: ABDOMINAL PAIN

Few abdominal conditions require such urgent operative intervention that an orderly approach need be abandoned, no matter how ill the patient. Only those patients with exsanguinating intraabdominal hemorrhage (e.g., ruptured aneurysm) must be rushed to the operating room immediately, but in such instances only a few minutes are required to assess the critical nature of the problem. Under these circumstances, all obstacles must be swept aside, adequate venous access for fluid replacement obtained, and the operation begun. Many patients of this type have died in the radiology department or the emergency room while awaiting such unnecessary examinations as electrocardiograms or CT scans. *There are no contraindications to operation when massive intraabdominal hemorrhage is present*. Fortunately, this situation is relatively rare. These comments do not pertain to gastrointestinal hemorrhage, which can often be managed by other means (Chap. 41).

Nothing will supplant an orderly, painstakingly *detailed history*, which is far more valuable than any laboratory or radiographic examination. This kind of history is laborious and time-consuming, making it not especially popular, even though a reasonably accurate diagnosis can be made on the basis of the history alone in the majority of cases. Computer-aided diagnosis of abdominal pain provides no advantage over clinical assessment alone. In cases of *acute* abdominal pain, a diagnosis is readily established in most instances, whereas success is not so frequent in patients with *chronic* pain. IBS is one of the most common causes of abdominal pain and must always be kept in mind (Chap. 296). The location of the pain can assist in narrowing the differential diagnosis (**Table 13-2**); however, the *chronological sequence of events* in the patient's history is often more important than emphasis on the location of pain. If the examiner is sufficiently open-minded and unhurried, asks the proper questions, and listens, the patient will usually provide the diagnosis. Careful attention should be paid to the extra-abdominal regions that may be responsible for abdominal pain. An accurate menstrual history in a female patient is essential. Narcotics or analgesics should *not* be withheld until a definitive diagnosis or a definitive plan has been formulated; obfuscation of the diagnosis by adequate analgesia is unlikely.

Table 13-2 Differential Diagnoses of Abdominal Pain by Location

Right Upper Quadrant	**Epigastric**	**Left Upper Quadrant**
Cholecystitis	Peptic ulcer disease	Splenic infarct
Cholangitis	Gastritis	Splenic rupture
Pancreatitis	GERD	Splenic abscess
Pneumonia/empyema	Pancreatitis	Gastritis
Pleurisy/pleurodynia	Myocardial infarction	Gastric ulcer
Subdiaphragmatic abscess	Pericarditis	Pancreatitis
Hepatitis	Ruptured aortic aneurysm	Subdiaphragmatic abscess
Budd-Chiari syndrome	Esophagitis	
Right Lower Quadrant	**Periumbilical**	**Left Lower Quadrant**
Appendicitis	Early appendicitis	Diverticulitis
Salpingitis	Gastroenteritis	Salpingitis
Inguinal hernia	Bowel obstruction	Inguinal hernia
Ectopic pregnancy	Ruptured aortic aneurysm	Ectopic pregnancy
Nephrolithiasis		Nephrolithiasis
Inflammatory bowel disease		Irritable bowel syndrome
Mesenteric lymphadenitis		Inflammatory bowel disease
Typhlitis		
Diffuse Nonlocalized Pain		
Gastroenteritis		
Mesenteric ischemia		
Bowel obstruction		
Irritable bowel syndrome		

Peritonitis
Diabetes
Malaria
Familial Mediterranean fever
Metabolic diseases
Psychiatric disease

Abbreviation: GERD, gastroesophageal reflux disease.

In the examination, simple critical inspection of the patient, e.g., of facies, position in bed, and respiratory activity, provides valuable clues. The amount of information to be gleaned is directly proportional to the *gentleness* and thoroughness of the examiner. Once a patient with peritoneal inflammation has been examined brusquely, accurate assessment by the next examiner becomes almost impossible. Eliciting rebound tenderness by sudden release of a deeply palpating hand in a patient with suspected peritonitis is cruel and unnecessary. The same information can be obtained by gentle percussion of the abdomen (rebound tenderness on a miniature scale), a maneuver that can be far more precise and localizing. Asking the patient to cough will elicit true rebound tenderness without the need for placing a hand on the abdomen. Furthermore, the forceful demonstration of rebound tenderness will startle and induce protective spasm in a nervous or worried patient in whom true rebound tenderness is not present. A palpable gallbladder will be missed if palpation is so brusque that voluntary muscle spasm becomes superimposed on involuntary muscular rigidity.

As with history taking, sufficient time should be spent in the examination. Abdominal signs may be minimal but nevertheless, if accompanied by consistent symptoms, may be exceptionally meaningful. Abdominal signs may be virtually or totally absent in cases of pelvic peritonitis, so careful *pelvic and rectal examinations are mandatory in every patient with abdominal pain*. Tenderness on pelvic or rectal examination in the absence of other abdominal signs can be caused by operative indications such as perforated appendicitis, diverticulitis, twisted ovarian cyst, and many others.

Much attention has been paid to the presence or absence of peristaltic sounds, their quality, and their frequency. Auscultation of the abdomen is one of the least revealing aspects of the physical examination of a patient with abdominal pain. Catastrophes such as strangulating small intestinal obstruction or perforated appendicitis may occur in the presence of normal peristaltic sounds. Conversely, when the proximal part of the intestine above an obstruction becomes markedly distended and edematous, peristaltic sounds may lose the characteristics of borborygmi and become weak or absent, even when peritonitis is not present. It is usually the severe chemical peritonitis of sudden onset that is associated with the truly silent abdomen. Assessment of the patient's state of hydration is important.

Laboratory examinations may be valuable in assessing the patient with abdominal pain, yet, with few exceptions, they rarely establish a diagnosis. Leukocytosis should never be the single deciding factor as to whether or not operation is indicated. A white blood cell count >20,000/μL may be observed with perforation of a viscus, but pancreatitis, acute cholecystitis, pelvic inflammatory disease, and intestinal infarction may be associated with marked leukocytosis. A normal white blood cell count is not rare in cases of perforation of abdominal viscera. The diagnosis of anemia may be more helpful than the white blood cell count, especially when combined with the history.

The urinalysis may reveal the state of hydration or rule out severe renal disease, diabetes, or urinary infection. Blood urea nitrogen, glucose, and serum bilirubin levels may be helpful. Serum amylase levels may be increased by many diseases other than pancreatitis, e.g., perforated ulcer, strangulating intestinal obstruction, and acute cholecystitis; thus, elevations of serum amylase do not rule out the need for an operation. The determination of the serum lipase may have greater accuracy than that of the serum amylase.

Plain and upright or lateral decubitus radiographs of the abdomen may be of value in cases of intestinal obstruction, perforated ulcer, and a variety of other conditions. They are usually unnecessary in patients with acute appendicitis or strangulated external hernias. In rare instances, barium or water-soluble contrast study of the upper part of the gastrointestinal tract may demonstrate partial intestinal obstruction that may elude diagnosis by other means. If there is any question of obstruction of the colon, oral administration of barium sulfate should be avoided. On the other hand, in cases of suspected colonic obstruction (without perforation), contrast enema may be diagnostic.

In the absence of trauma, peritoneal lavage has been replaced as a diagnostic tool by ultrasound, CT, and laparoscopy. Ultrasonography has proved to be useful in detecting an enlarged gallbladder or pancreas, the presence of gallstones, an enlarged ovary, or a tubal pregnancy. Laparoscopy is especially helpful in diagnosing pelvic conditions, such as ovarian cysts, tubal pregnancies, salpingitis, and acute appendicitis. Radioisotopic hepatobiliary iminodiacetic acid scans (HIDAs) may help differentiate acute cholecystitis from acute pancreatitis. A CT scan may demonstrate an enlarged

pancreas, ruptured spleen, or thickened colonic or appendiceal wall and streaking of the mesocolon or mesoappendix characteristic of diverticulitis or appendicitis.

Sometimes, even under the best circumstances with all available aids and with the greatest of clinical skill, a definitive diagnosis cannot be established at the time of the initial examination. Nevertheless, despite lack of a clear anatomic diagnosis, it may be abundantly clear to an experienced and thoughtful physician and surgeon that on clinical grounds alone operation is indicated. Should that decision be questionable, watchful waiting with repeated questioning and examination will often elucidate the true nature of the illness and indicate the proper course of action.

FURTHER READINGS

Jones PF: Suspected acute appendicitis: Trends in management over 30 years. Br J Surg 88:1570, 2001 [PMID: 11736966]

Lyon C, Clark DC: Diagnosis of acute abdominal pain in older patients. Am Fam Physician 74:1537, 2006 [PMID: 17111893]

Merlin MA et al: Evidence-based appendicitis: The initial work-up. Postgrad Med 122:189, 2010 [PMID: 20463429]

Silen W: Cope's Early Diagnosis of the Acute Abdomen, 21st ed, New York and Oxford: Oxford University Press, 2005

Thomas SH, Silen W: Effect on diagnostic efficiency of analgesia for undifferentiated abdominal pain. Br J Surg 90:5, 2003 [PMID: 12520567]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEADACHE: INTRODUCTION

Headache is among the most common reasons patients seek medical attention. Diagnosis and management is based on a careful clinical approach augmented by an understanding of the anatomy, physiology, and pharmacology of the nervous system pathways that mediate the various headache syndromes.

GENERAL PRINCIPLES

A classification system developed by the International Headache Society characterizes headache as primary or secondary (**Table 14–1)**. *Primary headaches* are those in which headache and its associated features are the disorder in itself, whereas *secondary headaches* are those caused by exogenous disorders. Primary headache often results in considerable disability and a decrease in the patient's quality of life. Mild secondary headache, such as that seen in association with upper respiratory tract infections, is common but rarely worrisome. Life-threatening headache is relatively uncommon, but vigilance is required in order to recognize and appropriately treat such patients.

Table 14-1 Common Causes of Headache

Primary Headache		**Secondary Headache**	
Type	**%**	**Type**	**%**
Tension-type	69	Systemic infection	63
Migraine	16	Head injury	4
Idiopathic stabbing	2	Vascular disorders	1
Exertional	1	Subarachnoid hemorrhage	<1
Cluster	0.1	Brain tumor	0.1

Source: After J Olesen et al: *The Headaches.* Philadelphia, Lippincott, Williams & Wilkins, 2005.

Anatomy and Physiology of Headache

Pain usually occurs when peripheral nociceptors are stimulated in response to tissue injury, visceral distension, or other factors (Chap. 11). In such situations, pain perception is a normal physiologic response mediated by a healthy nervous system. Pain can also result when pain-producing pathways of the peripheral or central nervous system (CNS) are damaged or activated inappropriately. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain-producing; these include the scalp, middle meningeal artery, dural sinuses, falx cerebri, and proximal segments of the large pial arteries. The ventricular ependyma, choroid plexus, pial veins, and much of the brain parenchyma are not pain-producing.

The key structures involved in primary headache appear to be

- the large intracranial vessels and dura mater and the peripheral terminals of the trigeminal nerve that innervate these structures
- the caudal portion of the trigeminal nucleus, which extends into the dorsal horns of the upper cervical spinal cord and receives input from the first and second cervical nerve roots (the trigeminocervical complex)
- rostral pain-processing regions, such as the ventroposteromedial thalamus and the cortex
- the pain-modulatory systems in the brain that modulate input from trigeminal nociceptors at all levels of the pain-processing pathways

The innervation of the large intracranial vessels and dura mater by the trigeminal nerve is known as the *trigeminovascular system*. Cranial autonomic symptoms, such as *lacrimation* and *nasal congestion*, are prominent in the trigeminal autonomic cephalalgias, including cluster headache and paroxysmal hemicrania, and may also be seen in migraine. These autonomic symptoms reflect

activation of cranial parasympathetic pathways, and functional imaging studies indicate that vascular changes in migraine and cluster headache, when present, are similarly driven by these cranial autonomic systems. Migraine and other primary headache types are not "vascular headaches"; these disorders do not reliably manifest vascular changes, and treatment outcomes cannot be predicted by vascular effects. Migraine is a brain disorder, and best understood and managed as such.

Clinical Evaluation of Acute, New-Onset Headache

The patient who presents with a new, severe headache has a differential diagnosis that is quite different from the patient with recurrent headaches over many years. In new-onset and severe headache, the probability of finding a potentially serious cause is considerably greater than in recurrent headache. Patients with recent onset of pain require prompt evaluation and appropriate treatment. Serious causes to be considered include meningitis, subarachnoid hemorrhage, epidural or subdural hematoma, glaucoma, tumor, and purulent sinusitis. When worrisome symptoms and signs are present (**Table 14–2)**, rapid diagnosis and management is critical.

Table 14-2 Headache Symptoms that Suggest a Serious Underlying Disorder
"Worst" headache ever First severe headache
Subacute worsening over days or weeks
Abnormal neurologic examination Fever or unexplained systemic signs
Vomiting that precedes headache
Pain induced by bending, lifting, cough
Pain that disturbs sleep or presents immediately upon awakening Known systemic illness
Onset after age 55
Pain associated with local tenderness, e.g., region of temporal artery

A complete neurologic examination is an essential first step in the evaluation. In most cases, patients with an abnormal examination or a history of recent-onset headache should be evaluated by a CT or MRI study. As an initial screening procedure for intracranial pathology in this setting, CT and MRI methods appear to be equally sensitive. In some circumstances, a lumbar puncture (LP) is also required, unless a benign etiology can be otherwise established. A general evaluation of acute headache might include the investigation of cardiovascular and renal status by blood pressure monitoring and urine examination; eyes by funduscopy, intraocular pressure measurement, and refraction; cranial arteries by palpation; and cervical spine by the effect of passive movement of the head and by imaging.

The psychological state of the patient should also be evaluated since a relationship exists between head pain and depression. Many patients in chronic daily pain cycles become depressed, although depression itself is rarely a cause of headache. Drugs with antidepressant actions are also effective in the prophylactic treatment of both tension-type headache and migraine.

Underlying recurrent headache disorders may be activated by pain that follows otologic or endodontic surgical procedures. Thus, pain about the head as the result of diseased tissue or trauma may reawaken an otherwise quiescent migrainous syndrome. Treatment of the headache is largely ineffective until the cause of the primary problem is addressed.

Serious underlying conditions that are associated with headache are described below. Brain tumor is a rare cause of headache and even less commonly a cause of severe pain. The vast majority of patients presenting with severe headache have a benign cause.

SECONDARY HEADACHE

The management of secondary headache focuses on diagnosis and treatment of the underlying condition.

Meningitis

Acute, severe headache with stiff neck and fever suggests meningitis. Lumbar puncture is mandatory. Often there is striking accentuation of pain with eye movement. Meningitis can be easily mistaken for

migraine in that the cardinal symptoms of pounding headache, photophobia, nausea, and vomiting are frequently present, perhaps reflecting the underlying biology of some of the patients.

Meningitis is discussed in Chaps. 381 and 382.

Intracranial Hemorrhage

Acute, severe headache with stiff neck but without fever suggests subarachnoid hemorrhage. A ruptured aneurysm, arteriovenous malformation, or intraparenchymal hemorrhage may also present with headache alone. Rarely, if the hemorrhage is small or below the foramen magnum, the head CT scan can be normal. Therefore, lumbar puncture may be required to diagnose definitively subarachnoid hemorrhage.

Intracranial hemorrhage is discussed in Chap. 275.

Brain Tumor

Approximately 30% of patients with brain tumors consider headache to be their chief complaint. The head pain is usually nondescript—an intermittent deep, dull aching of moderate intensity, which may worsen with exertion or change in position and may be associated with nausea and vomiting. This pattern of symptoms results from migraine far more often than from brain tumor. The headache of brain tumor disturbs sleep in about 10% of patients. Vomiting that precedes the appearance of headache by weeks is highly characteristic of posterior fossa brain tumors. A history of amenorrhea or galactorrhea should lead one to question whether a prolactin-secreting pituitary adenoma (or the polycystic ovary syndrome) is the source of headache. Headache arising de novo in a patient with known malignancy suggests either cerebral metastases or carcinomatous meningitis, or both. Head pain appearing abruptly after bending, lifting, or coughing can be due to a posterior fossa mass, a Chiari malformation, or low CSF volume.

Brain tumors are discussed in Chap. 379.

Temporal Arteritis

(See also Chaps. 28 and 326) Temporal (giant cell) arteritis is an inflammatory disorder of arteries that frequently involves the extracranial carotid circulation. It is a common disorder of the elderly; its annual incidence is 77 per 100,000 individuals age 50 and older. The average age of onset is 70 years, and women account for 65% of cases. About half of patients with untreated temporal arteritis develop blindness due to involvement of the ophthalmic artery and its branches; indeed, the ischemic optic neuropathy induced by giant cell arteritis is the major cause of rapidly developing bilateral blindness in patients >60 years. Because treatment with glucocorticoids is effective in preventing this complication, prompt recognition of the disorder is important.

Typical presenting symptoms include headache, polymyalgia rheumatica (Chap. 326), jaw claudication, fever, and weight loss. Headache is the dominant symptom and often appears in association with malaise and muscle aches. Head pain may be unilateral or bilateral and is located temporally in 50% of patients but may involve any and all aspects of the cranium. Pain usually appears gradually over a few hours before peak intensity is reached; occasionally, it is explosive in onset. The quality of pain is only seldom throbbing; it is almost invariably described as dull and boring, with superimposed episodic stabbing pains similar to the sharp pains that appear in migraine. Most patients can recognize that the origin of their head pain is superficial, external to the skull, rather than originating deep within the cranium (the pain site for migraineurs). Scalp tenderness is present, often to a marked degree; brushing the hair or resting the head on a pillow may be impossible because of pain. Headache is usually worse at night and often aggravated by exposure to cold. Additional findings may include reddened, tender nodules or red streaking of the skin overlying the temporal arteries, and tenderness of the temporal or, less commonly, the occipital arteries.

The erythrocyte sedimentation rate (ESR) is often, though not always, elevated; a normal ESR does not exclude giant cell arteritis. A temporal artery biopsy followed by immediate treatment with prednisone 80 mg daily for the first 4–6 weeks should be initiated when clinical suspicion is high. The prevalence of migraine among the elderly is substantial, considerably higher than that of giant cell arteritis. Migraineurs often report amelioration of their headaches with prednisone; thus, caution must be used when interpreting the therapeutic response.

Glaucoma

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The headache often starts with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.

Glaucoma is discussed in Chap. 28.

PRIMARY HEADACHE SYNDROMES

Primary headaches are disorders in which headache and associated features occur in the absence of any exogenous cause (Table 14–1). The most common are migraine, tension-type headache, and cluster headache.

Migraine

Migraine, the second most common cause of headache, afflicts approximately 15% of women and 6% of men over a one year period. It is usually an episodic headache associated with certain features such as sensitivity to light, sound, or movement; nausea and vomiting often accompany the headache. A useful description of migraine is a benign and recurring syndrome of headache associated with other symptoms of neurologic dysfunction in varying admixtures (**Table 14–3)**. Migraine can often be recognized by its activators, referred to as *triggers*.

Table 14-3 Symptoms Accompanying Severe Migraine Attacks in 500 Patients

Symptom	Patients Affected, %
Nausea	87
Photophobia	82
Lightheadedness	72
Scalp tenderness	65
Vomiting	56
Visual disturbances	36
Paresthesias	33
Vertigo	33
Photopsia	26
Alteration of consciousness	18
Diarrhea	16
Fortification spectra	10
Syncope	10
Seizure	4
Confusional state	4

Source: From NH Raskin, *Headache*, 2nd ed. New York, Churchill Livingston, 1988; with permission.

The brain of the migraineur is particularly sensitive to environmental and sensory stimuli; migraine-prone patients do not habituate easily to sensory stimuli. This sensitivity is amplified in females during the menstrual cycle. Headache can be initiated or amplified by various triggers, including glare, bright lights, sounds, or other afferent stimulation; hunger; excess stress; physical exertion; stormy weather or barometric pressure changes; hormonal fluctuations during menses; lack of or excess sleep; and alcohol or other chemical stimulation. Knowledge of a patient's susceptibility to specific triggers can be useful in management strategies involving lifestyle adjustments.

PATHOGENESIS

The sensory sensitivity that is characteristic of migraine is probably due to dysfunction of monoaminergic sensory control systems located in the brainstem and thalamus (**Fig. 14-1)**.

Figure 14-1

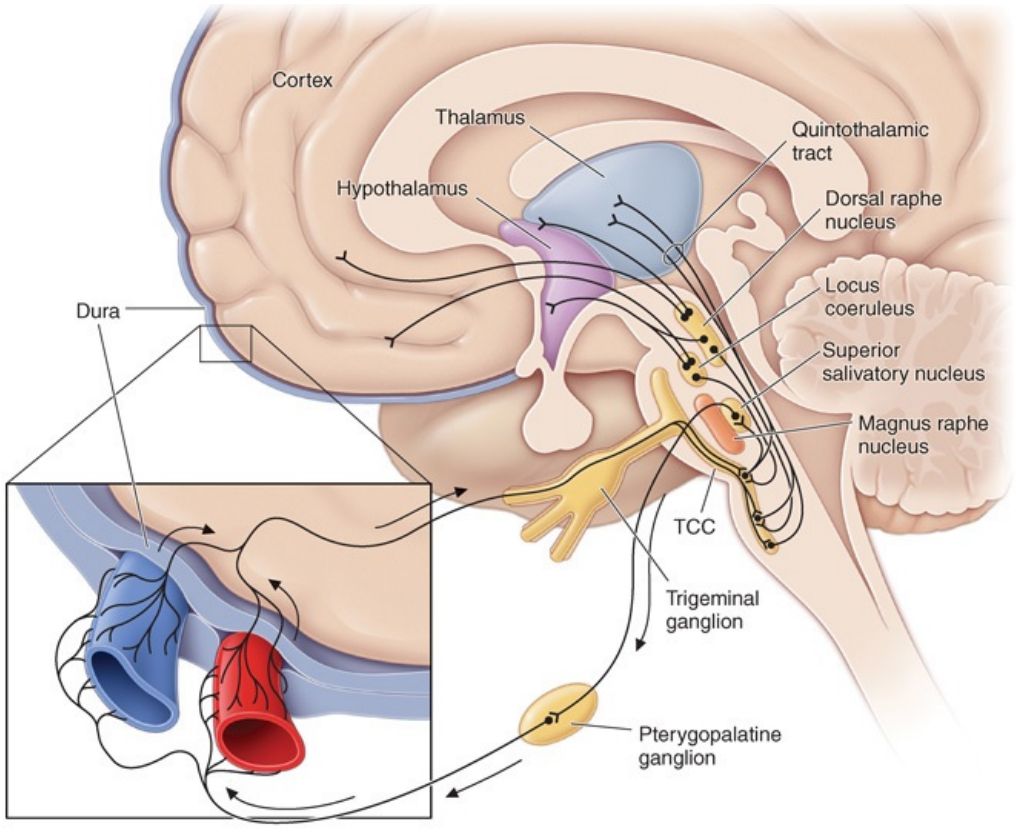

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Brainstem pathways that modulate sensory input. The key pathway for pain in migraine is the trigeminovascular input from the meningeal vessels, which passes through the trigeminal ganglion and synapses on second-order neurons in the trigeminocervical complex (TCC). These neurons in turn project in the quintothalamic tract and, after decussating in the brainstem, synapse on neurons in the thalamus. Important modulation of the trigeminovascular nociceptive input comes from the dorsal raphe nucleus, locus coeruleus, and nucleus raphe magnus.

Activation of cells in the trigeminal nucleus results in the release of vasoactive neuropeptides, particularly calcitonin gene–related peptide (CGRP), at vascular terminations of the trigeminal nerve and within the trigeminal nucleus. CGRP receptor antagonists have now been shown to be effective in the acute treatment of migraine. Centrally, the second-order trigeminal neurons cross the midline and project to ventrobasal and posterior nuclei of the thalamus for further processing. Additionally, there are projections to the periaqueductal gray and hypothalamus, from which reciprocal descending systems have established antinociceptive effects. Other brainstem regions likely to be involved in descending modulation of trigeminal pain include the nucleus locus coeruleus in the pons and the rostroventromedial medulla.

Pharmacologic and other data point to the involvement of the neurotransmitter 5-hydroxytryptamine (5-HT; also known as serotonin) in migraine. Approximately 60 years ago, methysergide was found to antagonize certain peripheral actions of 5-HT and was introduced as the first drug capable of preventing migraine attacks. The triptans are designed to selectively stimulate subpopulations of 5-HT receptors; at least 14 different 5-HT receptors exist in humans. The triptans are potent agonists of 5-HT_{1B}, 5-HT_{1D}, and 5-HT_{1F} receptors and are less potent at the 5-HT_{1A} receptor. A growing body of data indicates that the antimigraine efficacy of the triptans relates to their ability to stimulate 5-$HT_{1B/1D}$ receptors, which are located on both blood vessels and nerve terminals. Separately, it has now been shown that selective 5-HT_{1F} receptor activation, which has a purely neural effect, can terminate acute migraine.

Data also support a role for dopamine in the pathophysiology of migraine. Most migraine symptoms can be induced by dopaminergic stimulation. Moreover, there is dopamine receptor hypersensitivity in migraineurs, as demonstrated by the induction of yawning, nausea, vomiting, hypotension, and other symptoms of a migraine attack by dopaminergic agonists at doses that do not affect nonmigraineurs. Dopamine receptor antagonists are effective therapeutic agents in migraine, especially when given parenterally or concurrently with other antimigraine agents.

Migraine genes identified by studying families with familial hemiplegic migraine (FHM) reveal involvement of ion channels, suggesting that alterations in membrane excitability can predispose to

migraine. Mutations involving the $Ca_V2.1$ (P/Q)–type voltage-gated calcium channel *CACNA1A* gene are now known to cause FHM 1; this mutation is responsible for about 50% of FHM. Mutations in the Na^+-K^+ATPase *ATP1A2* gene, designated FHM 2, are responsible for about 20% of FHM. Mutations in the neuronal voltage-gated sodium channel *SCN1A* cause FHM 3. Functional neuroimaging has suggested that brainstem regions in migraine (**Fig. 14-2)** and the posterior hypothalamic gray matter region close to the human circadian pacemaker cells of the suprachiasmatic nucleus in cluster headache (**Fig. 14-3)** are good candidates for specific involvement in primary headache.

Figure 14-2

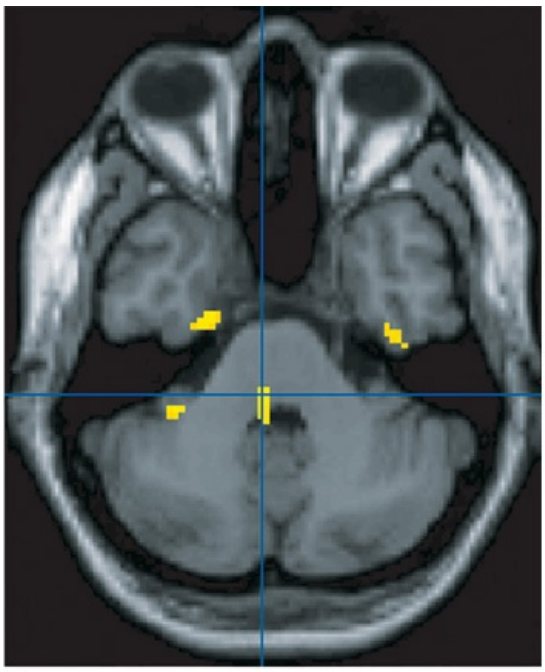

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

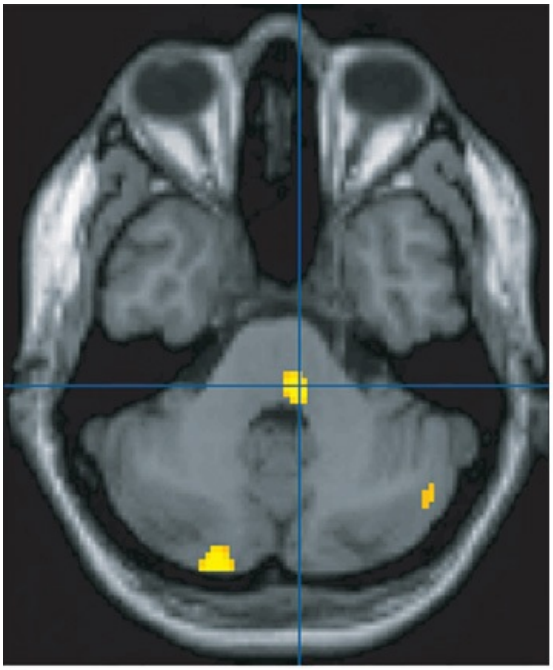

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Positron emission tomography (PET) activation in migraine. In spontaneous attacks of episodic migraine there is activation of the region of the dorsolateral pons; an identical pattern is found in chronic migraine (not shown). This area, which includes the noradrenergic locus coeruleus, is

fundamental to the expression of migraine. Moreover, lateralization of changes in this region of the brainstem correlates with lateralization of the head pain in hemicranial migraine; the scans shown in panels ***A*** and ***B*** are of patients with acute migraine headache on the right and left side, respectively. *(From S Afridi et al: Brain 128:932, 2005.)*

Figure 14-3

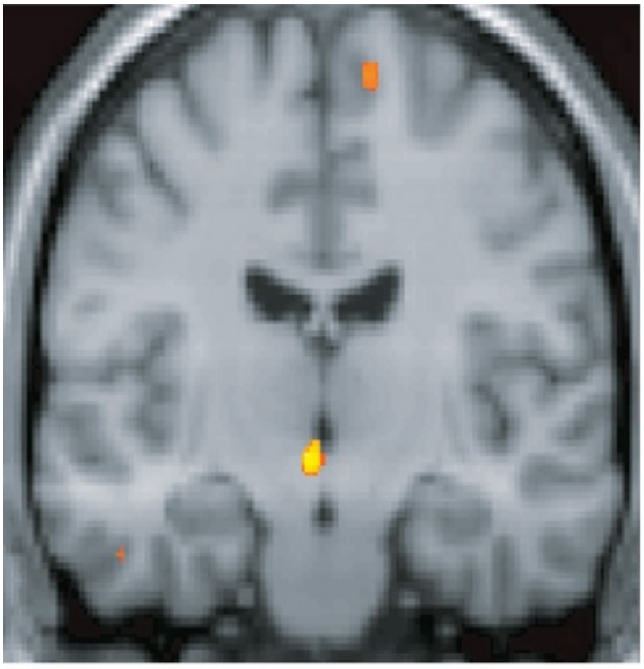

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

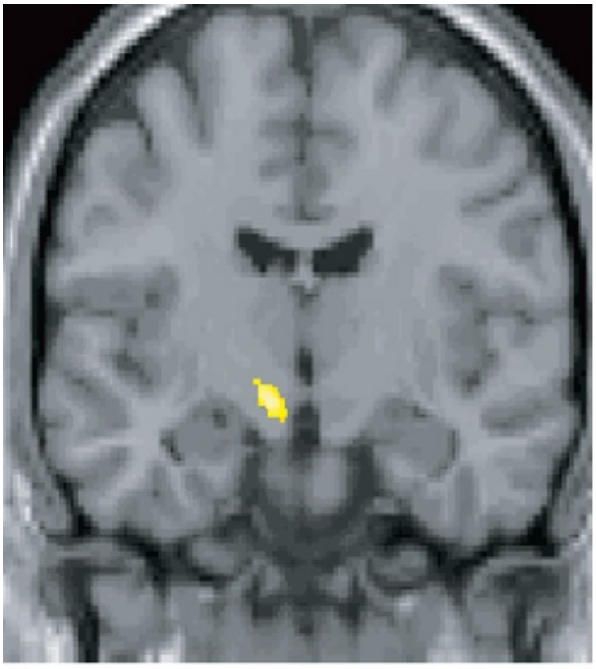

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Posterior hypothalamic gray matter activation on positron emission tomography (PET) in a patient with acute cluster headache (***A***). *(From A May et al: Lancet 352:275, 1998.)* High-resolution T1 weighted MRI obtained using voxel-based morphometry demonstrates increased gray matter activity, lateralized to the side of pain in a patient with cluster headache (***B***). *(From A May et al: Nat Med 5:836, 1999.)*

DIAGNOSIS AND CLINICAL FEATURES

Diagnostic criteria for migraine headache are listed in **Table 14–4**. A high index of suspicion is required to diagnose migraine: the migraine aura, consisting of visual disturbances with flashing lights or zigzag lines moving across the visual field or of other neurologic symptoms, is reported in only 20–25% of patients. A headache diary can often be helpful in making the diagnosis; this is also helpful in assessing disability and the frequency of treatment for acute attacks. Patients with episodes of migraine that occur daily or near-daily are considered to have chronic migraine (see "Chronic Daily Headache," below). Migraine must be differentiated from tension-type headache (discussed below), the most common primary headache syndrome seen in clinical practice. *Migraine at its most basic level is headache with associated features, and tension-type headache is headache that is featureless. Most patients with disabling headache probably have migraine.*

Table 14-4 Simplified Diagnostic Criteria for Migraine

Repeated attacks of headache lasting 4–72 h in patients with a normal physical examination, no other reasonable cause for the headache, and:

At Least 2 of the Following Features:	Plus at Least 1 of the Following Features:
Unilateral pain	Nausea/vomiting
Throbbing pain	Photophobia and phonophobia
Aggravation by movement	
Moderate or severe intensity	

Source: Adapted from the International Headache Society Classification (Headache Classification Committee of the International Headache Society, 2004).

Patients with acephalgic migraine experience recurrent neurologic symptoms, often with nausea or vomiting, but with little or no headache. Vertigo can be prominent; it has been estimated that one-third of patients referred for vertigo or dizziness have a primary diagnosis of migraine.

TREATMENT: MIGRAINE HEADACHES

Once a diagnosis of migraine has been established, it is important to assess the extent of a patient's disease and disability. The Migraine Disability Assessment Score (MIDAS) is a well-validated, easy-to-use tool (**Fig. 14-4**).

Figure 14-4

***MIDAS Questionnaire**

INSTRUCTIONS: Please answer the following questions about ALL headaches you have had over the last 3 months. Write zero if you did not do the activity in the last 3 months.

1. On how many days in the last 3 months did you miss work or school because of your headaches? ..... ____ days
2. How many days in the last 3 months was your productivity at work or school reduced by half or more because of your headaches *(do not include days you counted in question 1 where you missed work or school)*? ..... ____ days
3. On how many days in the last 3 months did you **not** do household work because of your headaches? ..... ____ days
4. How many days in the last 3 months was your productivity in household work reduced by half or more because of your headaches *(do not include days you counted in question 3 where you did not do household work)*? ..... ____ days
5. On how many days in the last 3 months did you miss family, social, or leisure activities because of your headaches? ..... ____ days

A. On how many days in the last 3 months did you have a headache? *(If a headache lasted more than one day, count each day.)* ..... ____ days

B. On a scale of 0–10, on average how painful were these headaches? *(Where 0 = no pain at all, and 10 = pain as bad as it can be.)* ..... ____

*Migraine Disability Assessment Score
(Questions 1–5 are used to calculate the MIDAS score.)
Grade I—Minimal or Infrequent Disability: 0–5
Grade II—Mild or Infrequent Disability: 6–10
Grade III—Moderate Disability: 11–20
Grade IV—Severe Disability: > 20

© Innovative Medical Research 1997

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MIDAS Questionnaire.

Patient education is an important aspect of migraine management. Information for patients is available at *www.achenet.org*, the website of the American Council for Headache Education (ACHE). It is helpful for patients to understand that migraine is an inherited tendency to headache; that migraine can be modified and controlled by lifestyle adjustments and medications, but it cannot be eradicated; and that, except in some occasions in women on oral estrogens or contraceptives, migraine is not associated with serious or life-threatening illnesses.

Nonpharmacologic Management

Migraine can often be managed to some degree by a variety of nonpharmacologic approaches. Most patients benefit by the identification and avoidance of specific headache triggers. A regulated lifestyle is helpful, including a healthful diet, regular exercise, regular sleep patterns, avoidance of excess caffeine and alcohol, and avoidance of acute changes in stress levels.

The measures that benefit a given individual should be used routinely since they provide a simple, cost-effective approach to migraine management. Patients with migraine do not encounter more stress than headache-free individuals; overresponsiveness to stress appears to be the issue. Since the stresses of everyday living cannot be eliminated, lessening one's response to stress by various techniques is helpful for many patients. These may include yoga, transcendental meditation, hypnosis, and conditioning techniques such as biofeedback. For most patients, this approach is, at best, an adjunct to pharmacotherapy. Nonpharmacologic measures are unlikely to prevent all migraine attacks. If these measures fail to prevent an attack, pharmacologic approaches are then needed to abort an attack.

Acute Attack Therapies for Migraine

The mainstay of pharmacologic therapy is the judicious use of one or more of the many drugs that are effective in migraine **(Table 14–5)**. The selection of the optimal regimen for a given patient depends on a number of factors, the most important of which is the severity of the attack. Mild migraine attacks can usually be managed by oral agents; the average efficacy rate is 50–70%. Severe migraine attacks may require parenteral therapy. Most drugs effective in the treatment of migraine are members of one of three major pharmacologic classes: anti-inflammatory agents, 5-$HT_{1B/1D}$ receptor agonists, and dopamine receptor antagonists.

Table 14-5 Treatment of Acute Migraine

Drug	Trade Name	Dosage
Simple Analgesics		

Acetaminophen, aspirin, caffeine	Excedrin Migraine	Two tablets or caplets q6h (max 8 per day)
NSAIDs		
Naproxen	Aleve, Anaprox, generic	220–550 mg PO bid
Ibuprofen	Advil, Motrin, Nuprin, generic	400 mg PO q3–4h
Tolfenamic acid	Clotam Rapid	200 mg PO. May repeat x1 after 1–2 h
5-HT$_1$ Agonists		
Oral		
Ergotamine	Ergomar	One 2 mg sublingual tablet at onset and q$^1/_2$h (max 3 per day, 5 per week)
Ergotamine 1 mg, caffeine 100 mg	Ercaf, Wigraine	One or two tablets at onset, then one tablet q$^1/_2$h (max 6 per day, 10 per week)
Naratriptan	Amerge	2.5 mg tablet at onset; may repeat once after 4 h
Rizatriptan	Maxalt	5–10 mg tablet at onset; may repeat after 2 h (max 30 mg/d)
	Maxalt-MLT	
Sumatriptan	Imitrex	50–100 mg tablet at onset; may repeat after 2 h (max 200 mg/d)
Frovatriptan	Frova	2.5 mg tablet at onset, may repeat after 2 h (max 5 mg/d)
Almotriptan	Axert	12.5 mg tablet at onset, may repeat after 2 h (max 25 mg/d)
Eletriptan	Relpax	40 or 80 mg
Zolmitriptan	Zomig Zomig Rapimelt	2.5 mg tablet at onset; may repeat after 2 h (max 10 mg/d)
Nasal		
Dihydroergotamine	Migranal Nasal Spray	Prior to nasal spray, the pump must be primed 4 times; 1 spray (0.5 mg) is administered, followed in 15 min by a second spray
Sumatriptan	Imitrex Nasal Spray	5–20 mg intranasal spray as 4 sprays of 5 mg or a single 20 mg spray (may repeat once after 2 h, not to exceed a dose of 40 mg/d)
Zolmitriptan	Zomig	5 mg intranasal spray as one spray (may repeat once after 2 h, not to exceed a dose of 10 mg/d)
Parenteral		
Dihydroergotamine	DHE-45	1 mg IV, IM, or SC at onset and q1h (max 3 mg/d, 6 mg per week)
Sumatriptan	Imitrex Injection	6 mg SC at onset (may repeat once after 1 h for max of 2 doses in 24 h)
Dopamine Antagonists		
Oral		
Metoclopramide	Reglan,[a] generic[a]	5–10 mg/d
Prochlorperazine	Compazine,[a] generic[a]	1–25 mg/d
Parenteral		
Chlorpromazine	Generic[a]	0.1 mg/kg IV at 2 mg/min; max 35 mg/d
Metoclopramide	Reglan,[a]	10 mg IV

	generic	
Prochlorperazine	Compazine,[a] generic[a]	10 mg IV
Other		
Oral		
Acetaminophen, 325 mg, *plus* dichloralphenazone, 100 mg, *plus* isometheptene, 65 mg	Midrin, Duradrin, generic	Two capsules at onset followed by 1 capsule q1h (max 5 capsules)
Nasal		
Butorphanol	Stadol[a]	1 mg (1 spray in 1 nostril), may repeat if necessary in 1–2 h
Parenteral		
Narcotics	Generic[a]	Multiple preparations and dosages; see Table 11-1

[a]Not all drugs are specifically indicated by the FDA for migraine. Local regulations and guidelines should be consulted.

Note: Antiemetics (e.g., domperidone 10 mg or ondansetron 4 or 8 mg) or prokinetics (e.g., metoclopramide 10 mg) are sometimes useful adjuncts.

Abbreviations: NSAIDs, nonsteroidal anti-inflammatory drugs; 5-HT, 5-hydroxytryptamine.

In general, an adequate dose of whichever agent is chosen should be used as soon as possible after the onset of an attack. If additional medication is required within 60 min because symptoms return or have not abated, the initial dose should be increased for subsequent attacks. Migraine therapy must be individualized; a standard approach for all patients is not possible. A therapeutic regimen may need to be constantly refined until one is identified that provides the patient with rapid, complete, and consistent relief with minimal side effects **(Table 14–6)**.

Table 14-6 Clinical Stratification of Acute Specific Migraine Treatments

Clinical Situation	Treatment Options
Failed NSAIDS/analgesics	**First tier** Sumatriptan 50 mg or 100 mg PO
	Almotriptan 12.5 mg PO
	Rizatriptan 10 mg PO
	Eletriptan 40 mg PO
	Zolmitriptan 2.5 mg PO
	Slower effect/better tolerability
	Naratriptan 2.5 mg PO Frovatriptan 2.5 mg PO
	Infrequent headache
	Ergotamine 1–2 mg PO
	Dihydroergotamine nasal spray 2 mg
Early nausea or difficulties taking tablets	Zolmitriptan 5 mg nasal spray
	Sumatriptan 20 mg nasal spray Rizatriptan 10 mg MLT wafer
Headache recurrence	Ergotamine 2 mg (most effective PR/usually with caffeine)
	Naratriptan 2.5 mg PO Almotriptan 12.5 mg PO
	Eletriptan 40 mg
Tolerating acute treatments poorly	Naratriptan 2.5 mg
Early vomiting	Almotriptan 12.5 mg Zolmitriptan 5 mg nasal spray

	Sumatriptan 25 mg PR
Menses-related headache	Sumatriptan 6 mg SC **Prevention**
	Ergotamine PO at night
	Estrogen patches
	Treatment
	Triptans Dihydroergotamine nasal spray
Very rapidly developing symptoms	Zolmitriptan 5 mg nasal spray
	Sumatriptan 6 mg SC
	Dihydroergotamine 1 mg IM

Nonsteroidal Anti-Inflammatory Drugs (NSAIDs)

Both the severity and duration of a migraine attack can be reduced significantly by nonsteroidal anti-inflammatory agents **(Table 14–5)**. Indeed, many undiagnosed migraineurs are self-treated with nonprescription NSAIDs. A general consensus is that NSAIDs are most effective when taken early in the migraine attack. However, the effectiveness of anti-inflammatory agents in migraine is usually less than optimal in moderate or severe migraine attacks. The combination of acetaminophen, aspirin, and caffeine has been approved for use by the U.S. Food and Drug Administration (FDA) for the treatment of mild to moderate migraine. The combination of aspirin and metoclopramide has been shown to be comparable to a single dose of sumatriptan. Important side effects of NSAIDs include dyspepsia and gastrointestinal irritation.

5-HT$_1$ Receptor Agonists

Oral

Stimulation of 5-HT$_{1B/1D}$ receptors can stop an acute migraine attack. Ergotamine and dihydroergotamine are nonselective receptor agonists, while the triptans are selective 5-HT$_{1B/1D}$ receptor agonists. A variety of triptans, 5-HT$_{1B/1D}$ receptor agonists—naratriptan, rizatriptan, eletriptan, sumatriptan, zolmitriptan, almotriptan, and frovatriptan—are now available for the treatment of migraine.

Each drug in the triptan class has similar pharmacologic properties but varies slightly in terms of clinical efficacy. Rizatriptan and eletriptan are the most efficacious of the triptans currently available in the United States. Sumatriptan and zolmitriptan have similar rates of efficacy as well as time to onset, with an advantage of having multiple formulations, whereas almotriptan, frovatriptan, and naratriptan are somewhat slower in onset and are better tolerated. Clinical efficacy appears to be related more to the t_{max} (time to peak plasma level) than to the potency, half-life, or bioavailability. This observation is consistent with a large body of data indicating that faster-acting analgesics are more effective than slower-acting agents.

Unfortunately, monotherapy with a selective oral 5-HT$_{1B/1D}$ agonist does not result in rapid, consistent, and complete relief of migraine in all patients. Triptans are not effective in migraine with aura unless given after the aura is completed and the headache initiated. Side effects are common though often mild and transient. Moreover, 5-HT$_{1B/1D}$ agonists are contraindicated in individuals with a history of cardiovascular and cerebrovascular disease. Recurrence of headache is another important limitation of triptan use and occurs at least occasionally in most patients. Evidence from randomized controlled trials show that coadministration of a longer-acting NSAID, naproxen 500 mg, with sumatriptan will augment the initial effect of sumatriptan and, importantly, reduce rates of headache recurrence.

Ergotamine preparations offer a nonselective means of stimulating 5-HT$_1$ receptors. A nonnauseating dose of ergotamine should be sought since a dose that provokes nausea is too high and may intensify head pain. Except for a sublingual formulation of ergotamine, oral formulations of ergotamine also contain 100 mg caffeine (theoretically to enhance ergotamine absorption and possibly to add additional analgesic activity). The average oral ergotamine dose for a migraine attack is 2 mg. Since the clinical studies demonstrating the efficacy of ergotamine in migraine predated the clinical trial methodologies used with the triptans, it is difficult to assess the clinical efficacy of ergotamine versus the triptans. In general, ergotamine appears to have a much higher incidence of nausea than triptans, but less headache recurrence.

Nasal

The fastest-acting nonparenteral antimigraine therapies that can be self-administered include nasal formulations of dihydroergotamine (Migranal), zolmitriptan (Zomig nasal), or sumatriptan. The nasal sprays result in substantial blood levels within 30–60 min. Although in theory nasal sprays might provide faster and more effective relief of a migraine attack than oral formulations, their reported efficacy is only approximately 50–60%. Studies with an inhalational formulation of dihydroergotamine indicate that its absorption problems can be overcome to produce rapid onset of action with good tolerability.

Parenteral

Parenteral administration of drugs such as dihydroergotamine and sumatriptan is approved by the FDA for the rapid relief of a migraine attack. Peak plasma levels of dihydroergotamine are achieved 3 min after IV dosing, 30 min after IM dosing, and 45 min after SC dosing. If an attack has not already peaked, SC or IM administration of 1 mg dihydroergotamine suffices for about 80–90% of patients. Sumatriptan, 6 mg SC, is effective in ~70–80% of patients.

Dopamine Antagonists

Oral

Oral dopamine antagonists should be considered as adjunctive therapy in migraine. Drug absorption is impaired during migraine because of reduced gastrointestinal motility. Delayed absorption occurs even in the absence of nausea and is related to the severity of the attack and not its duration. Therefore, when oral NSAIDs and/or triptan agents fail, the addition of a dopamine antagonist such as metoclopramide 10 mg should be considered to enhance gastric absorption. In addition, dopamine antagonists decrease nausea/vomiting and restore normal gastric motility.

Parenteral

Parenteral dopamine antagonists (e.g., chlorpromazine, prochlorperazine, metoclopramide) can also provide significant acute relief of migraine; they can be used in combination with parenteral 5-$HT_{1B/1D}$ agonists. A common IV protocol used for the treatment of severe migraine is the administration over 2 min of a mixture of 5 mg of prochlorperazine and 0.5 mg of dihydroergotamine.

Other Medications for Acute Migraine

Oral

The combination of acetaminophen, dichloralphenazone, and isometheptene, one to two capsules, has been classified by the FDA as "possibly" effective in the treatment of migraine. Since the clinical studies demonstrating the efficacy of this combination analgesic in migraine predated the clinical trial methodologies used with the triptans, it is difficult to compare the efficacy of this sympathomimetic compound to other agents.

Nasal

A nasal preparation of butorphanol is available for the treatment of acute pain. As with all narcotics, the use of nasal butorphanol should be limited to a select group of migraineurs, as described below.

Parenteral

Narcotics are effective in the acute treatment of migraine. For example, IV meperidine (50–100 mg) is given frequently in the emergency room. This regimen "works" in the sense that the pain of migraine is eliminated. However, this regimen is clearly suboptimal for patients with recurrent headache. Narcotics do not treat the underlying headache mechanism; rather, they act to alter the pain sensation. Moreover, in patients taking oral narcotics such as oxycodone or hydrocodone, narcotic addiction can greatly confuse the treatment of migraine. Narcotic craving and/or withdrawal can aggravate and accentuate migraine. Therefore, it is recommended that narcotic use in migraine be limited to patients with severe, but infrequent, headaches that are unresponsive to other pharmacologic approaches.

Medication-Overuse Headache

Acute attack medications, particularly codeine or barbiturate-containing compound analgesics, have a propensity to aggravate headache frequency and induce a state of refractory daily or near-daily headache called *medication-overuse headache*. This condition is likely not a separate headache entity but a reaction of the migraine patient to a particular medicine. Migraine patients who have two or more headache days a week should be cautioned about frequent analgesic use (see "Chronic Daily Headache," below).

Preventive Treatments for Migraine

Patients with an increasing frequency of migraine attacks, or with attacks that are either unresponsive or poorly responsive to abortive treatments, are good candidates for preventive agents. In general, a preventive medication should be considered in the subset of patients with five or more attacks a month. Significant side effects are associated with the use of many of these agents; furthermore, determination of dose can be difficult since the recommended doses have been derived for conditions other than migraine. The mechanism of action of these drugs is unclear; it seems likely that the brain sensitivity that underlies migraine is modified. Patients are usually started on a low dose of a chosen

treatment; the dose is then gradually increased, up to a reasonable maximum to achieve clinical benefit.

Drugs that have the capacity to stabilize migraine are listed in **Table 14–7**. Drugs must be taken daily, and there is usually a lag of at least 2–12 weeks before an effect is seen. The drugs that have been approved by the FDA for the prophylactic treatment of migraine include propranolol, timolol, sodium valproate, topiramate, and methysergide (not available in the United States). In addition, a number of other drugs appear to display prophylactic efficacy. This group includes amitriptyline, nortriptyline, flunarizine, phenelzine, gabapentin, and cyproheptadine. Placebo-controlled trials of onabotulinum toxin type A in episodic migraine were negative, while, overall, placebo-controlled trials in chronic migraine were positive. Phenelzine and methysergide are usually reserved for recalcitrant cases because of their serious potential side effects. Phenelzine is a monoamine oxidase inhibitor (MAOI); therefore, tyramine-containing foods, decongestants, and meperidine are contraindicated. Methysergide may cause retroperitoneal or cardiac valvular fibrosis when it is used for >6 months, and thus monitoring is required for patients using this drug; the risk of fibrosis is about 1:1500 and is likely to reverse after the drug is stopped.

Table 14-7 Preventive Treatments in Migraine[a]

Drug	Dose	Selected Side Effects
Pizotifen[b]	0.5–2 mg qd	Weight gain
		Drowsiness
Beta blocker		
Propranolol	40–120 mg bid	Reduced energy
		Tiredness
		Postural symptoms
		Contraindicated in asthma
Tricyclics		
Amitriptyline	10–75 mg at night	Drowsiness
Dothiepin	25–75 mg at night	
Nortriptyline	25–75 mg at night	***Note:*** Some patients may only need a total dose of 10 mg, although generally 1–1.5 mg/kg body weight is required
Anticonvulsants		
Topiramate	25–200 mg/d	Paresthesias
		Cognitive symptoms
		Weight loss
		Glaucoma
		Caution with nephrolithiasis
Valproate	400–600 mg bid	Drowsiness
		Weight gain
		Tremor
		Hair loss
		Fetal abnormalities

		Hematologic or liver abnormalities
Gabapentin	900–3600 mg qd	Dizziness
		Sedation
Serotonergic drugs		
Methysergide	1–4 mg qd	Drowsiness
		Leg cramps
		Hair loss
		Retroperitoneal fibrosis (1-month drug holiday is required every 6 months)
Flunarizine[b]	5–15 mg qd	Drowsiness
		Weight gain
		Depression
		Parkinsonism
No convincing evidence from controlled trials Verapamil		
Controlled trials demonstrate *no effect*		
Nimodipine Clonidine		
SSRIs: fluoxetine		

[a]Commonly used preventives are listed with typical doses and common side effects. Not all listed medicines are approved by the FDA; local regulations and guidelines should be consulted.

[b]Not available in the United States.

The probability of success with any one of the antimigraine drugs is 50–75%. Many patients are managed adequately with low-dose amitriptyline, propranolol, topiramate, gabapentin, or valproate. If these agents fail or lead to unacceptable side effects, second-line agents such as methysergide or phenelzine can be used. Once effective stabilization is achieved, the drug is continued for ~6 months and then slowly tapered to assess the continued need. Many patients are able to discontinue medication and experience fewer and milder attacks for long periods, suggesting that these drugs may alter the natural history of migraine.

Tension-Type Headache

CLINICAL FEATURES

The term *tension-type headache* (TTH) is commonly used to describe a chronic head-pain syndrome characterized by bilateral tight, bandlike discomfort. The pain typically builds slowly, fluctuates in severity, and may persist more or less continuously for many days. The headache may be episodic or chronic (present >15 days per month).

A useful clinical approach is to diagnose TTH in patients whose headaches are completely without accompanying features such as nausea, vomiting, photophobia, phonophobia, osmophobia, throbbing, and aggravation with movement. Such an approach neatly separates migraine, which has one or more of these features and is the main differential diagnosis, from TTH. The International Headache Society's main definition of TTH allows an admixture of nausea, photophobia, or phonophobia in various combinations, although the appendix definition does not; this illustrates the difficulty in distinguishing these two clinical entities. In clinical practice, dichotomizing patients on the basis of the presence of associated features (migraine) and the absence of associated features (TTH) is highly recommended. Indeed patients whose headaches fit the TTH phenotype and who have migraine at

other times, along with a family history of migraine, migrainous illnesses of childhood, or typical migraine triggers to their migraine attacks, may be biologically different from those who have TTH headache with none of the features.

PATHOPHYSIOLOGY

The pathophysiology of TTH is incompletely understood. It seems likely that TTH is due to a primary disorder of CNS pain modulation alone, unlike migraine, which involves a more generalized disturbance of sensory modulation. Data suggest a genetic contribution to TTH, but this may not be a valid finding: given the current diagnostic criteria, the studies undoubtedly included many migraine patients. The name *tension-type headache* implies that pain is a product of *nervous tension*, but there is no clear evidence for tension as an etiology. Muscle contraction has been considered to be a feature that distinguishes TTH from migraine, but there appear to be no differences in contraction between the two headache types.

TREATMENT: TENSION-TYPE HEADACHE

The pain of TTH can generally be managed with simple analgesics such as acetaminophen, aspirin, or NSAIDs. Behavioral approaches including relaxation can also be effective. Clinical studies have demonstrated that triptans in pure TTH are not helpful, although triptans are effective in TTH when the patient also has migraine. For chronic TTH, amitriptyline is the only proven treatment **(Table 14–7)**; other tricyclics, selective serotonin reuptake inhibitors, and the benzodiazepines have not been shown to be effective. There is no evidence for the efficacy of acupuncture. Placebo-controlled trials of onabotulinum toxin type A in chronic TTH have not shown benefit.

Trigeminal Autonomic Cephalalgias, Including Cluster Headache

The trigeminal autonomic cephalalgias (TACs) describe a grouping of primary headaches including cluster headache, paroxysmal hemicrania, and SUNCT (*s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing)/SUNA (*s*hort-lasting *u*nilateral *n*euralgiform headache attacks with cranial *a*utonomic symptoms). TACs are characterized by relatively short-lasting attacks of head pain associated with cranial autonomic symptoms, such as lacrimation, conjunctival injection, or nasal congestion (**Table 14–8**). Pain is usually severe and may occur more than once a day. Because of the associated nasal congestion or rhinorrhea, patients are often misdiagnosed with "sinus headache" and treated with decongestants, which are ineffective.

Table 14-8 Clinical Features of the Trigeminal Autonomic Cephalalgias

	Cluster Headache	Paroxysmal Hemicrania	SUNCT
Gender Pain	M > F	F = M	F ~ M
Type	Stabbing, boring	Throbbing, boring, stabbing	Burning, stabbing, sharp
Severity	Excruciating	Excruciating	Severe to excruciating
Site	Orbit, temple	Orbit, temple	Periorbital
Attack frequency	1/alternate day–8/d	1–40/d (>5/d for more than half the time)	3–200/d
Duration of attack	15–180 min	2–30 min	5–240 s
Autonomic features	Yes	Yes	Yes (prominent conjunctival injection and lacrimation)[a]
Migrainous features[b]	Yes	Yes	Yes
Alcohol trigger	Yes	No	No
Cutaneous triggers	No	No	Yes
Indomethacin effect	—	Yes[c]	—
Abortive treatment	Sumatriptan injection or nasal spray	No effective treatment	Lidocaine (IV)

	Oxygen		
Prophylactic treatment	Verapamil Methysergide	Indomethacin	Lamotrigine Topiramate
	Lithium		Gabapentin

[a] If conjunctival injection and tearing not present, consider SUNA.

[b] Nausea, photophobia, or phonophobia; photophobia and phonophobia are typically unilateral on the side of the pain.

[c] Indicates complete response to indomethacin.

Abbreviation: SUNCT, *s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing.

TACs must be differentiated from short-lasting headaches that do not have prominent cranial autonomic syndromes, notably trigeminal neuralgia, primary stabbing headache, and hypnic headache. The cycling pattern and length, frequency, and timing of attacks are useful in classifying patients. Patients with TACs should undergo pituitary imaging and pituitary function tests as there is an excess of TAC presentations in patients with pituitary tumor–related headache.

CLUSTER HEADACHE

Cluster headache is a rare form of primary headache with a population frequency of approximately 0.1%. The pain is deep, usually retroorbital, often excruciating in intensity, nonfluctuating, and explosive in quality. A core feature of cluster headache is periodicity. At least one of the daily attacks of pain recurs at about the same hour each day for the duration of a cluster bout. The typical cluster headache patient has daily bouts of one to two attacks of relatively short-duration unilateral pain for 8 to 10 weeks a year; this is usually followed by a pain-free interval that averages a little less than 1 year. Cluster headache is characterized as chronic when there is no significant period of sustained remission. Patients are generally perfectly well between episodes. Onset is nocturnal in about 50% of patients, and men are affected three times more often than women. Patients with cluster headache tend to move about during attacks, pacing, rocking, or rubbing their head for relief; some may even become aggressive during attacks. This is in sharp contrast to patients with migraine, who prefer to remain motionless during attacks.

Cluster headache is associated with ipsilateral symptoms of cranial parasympathetic autonomic activation: conjunctival injection or lacrimation, rhinorrhea or nasal congestion, or cranial sympathetic dysfunction such as ptosis. The sympathetic deficit is peripheral and likely to be due to parasympathetic activation with injury to ascending sympathetic fibers surrounding a dilated carotid artery as it passes into the cranial cavity. When present, photophobia and phonophobia are far more likely to be unilateral and on the same side of the pain, rather than bilateral, as is seen in migraine. This phenomenon of unilateral photophobia/phonophobia is characteristic of TACs. Cluster headache is likely to be a disorder involving central pacemaker neurons in the region of the posterior hypothalamus (Fig. 14-3).

Treatment: Cluster Headache

The most satisfactory treatment is the administration of drugs to prevent cluster attacks until the bout is over. However, treatment of acute attacks is required for all cluster headache patients at some time.

Acute Attack Treatment

Cluster headache attacks peak rapidly, and thus a treatment with quick onset is required. Many patients with acute cluster headache respond very well to oxygen inhalation. This should be given as 100% oxygen at 10–12 L/min for 15–20 min. It appears that high flow and high oxygen content are important. Sumatriptan 6 mg SC is rapid in onset and will usually shorten an attack to 10–15 min; there is no evidence of tachyphylaxis. Sumatriptan (20 mg) and zolmitriptan (5 mg) nasal sprays are both effective in acute cluster headache, offering a useful option for patients who may not wish to self-inject daily. Oral sumatriptan is not effective for prevention or for acute treatment of cluster headache.

Preventive Treatments

(**Table 14–9**) The choice of a preventive treatment in cluster headache depends in part on the length of the bout. Patients with long bouts or those with chronic cluster headache require medicines that are safe when taken for long periods. For patients with relatively short bouts, limited courses of oral glucocorticoids or methysergide (not available in the United States) can be very useful. A 10-day course of prednisone, beginning at 60 mg daily for 7 days and followed by a rapid taper, may interrupt the pain bout for many patients. When ergotamine (1–2 mg) is used, it is most effective when given 1–2 h before an expected attack. Patients who use ergotamine daily must be educated regarding the

early symptoms of ergotism, which may include vomiting, numbness, tingling, pain, and cyanosis of the limbs; a weekly limit of 14 mg should be adhered to. Lithium (600–900 mg qd) appears to be particularly useful for the chronic form of the disorder.

Table 14-9 Preventive Management of Cluster Headache

Short-Term Prevention Episodic Cluster Headache	Long-Term Prevention Episodic Cluster Headache & Prolonged Chronic Cluster Headache
Prednisone 1 mg/kg up to 60 mg qd, tapering over 21 days	Verapamil 160–960 mg/d
	Lithium 400–800 mg/d
Methysergide 3–12 mg/d	Methysergide 3–12 mg/d
Verapamil 160–960 mg/d	Topiramate[a] 100–400 mg/d
Greater occipital nerve injection	Gabapentin[a] 1200–3600 mg/d
	Melatonin[a] 9–12 mg/d

[a] Unproven but of potential benefit.

Many experts favor verapamil as the first-line preventive treatment for patients with chronic cluster headache or prolonged bouts. While verapamil compares favorably with lithium in practice, some patients require verapamil doses far in excess of those administered for cardiac disorders. The initial dose range is 40–80 mg twice daily; effective doses may be as high as 960 mg/d. Side effects such as constipation and leg swelling can be problematic. Of paramount concern, however, is the cardiovascular safety of verapamil, particularly at high doses. Verapamil can cause heart block by slowing conduction in the atrioventricular node, a condition that can be monitored by following the PR interval on a standard ECG. Approximately 20% of patients treated with verapamil develop ECG abnormalities, which can be observed with doses as low as 240 mg/d; these abnormalities can worsen over time in patients on stable doses. A baseline ECG is recommended for all patients. The ECG is repeated 10 days after a dose change in those patients whose dose is being increased above 240 mg daily. Dose increases are usually made in 80-mg increments. For patients on long-term verapamil, ECG monitoring every 6 months is advised.

Neurostimulation Therapy

When medical therapies fail in chronic cluster headache, neurostimulation strategies can be employed. Deep-brain stimulation of the region of the posterior hypothalamic gray matter has proven successful in a substantial proportion of patients. Favorable results have also been reported with the less-invasive approach of occipital nerve stimulation.

Paroxysmal Hemicrania

Paroxysmal hemicrania (PH) is characterized by frequent unilateral, severe, short-lasting episodes of headache. Like cluster headache, the pain tends to be retroorbital but may be experienced all over the head and is associated with autonomic phenomena such as lacrimation and nasal congestion. Patients with remissions are said to have episodic PH, whereas those with the nonremitting form are said to have chronic PH. The essential features of PH are unilateral, very severe pain; short-lasting attacks (2–45 min); very frequent attacks (usually more than five a day); marked autonomic features ipsilateral to the pain; rapid course (<72 h); and excellent response to indomethacin. In contrast to cluster headache, which predominantly affects males, the male:female ratio in PH is close to 1:1.

Indomethacin (25–75 mg tid), which can completely suppress attacks of PH, is the treatment of choice. Although therapy may be complicated by indomethacin-induced gastrointestinal side effects, currently there are no consistently effective alternatives. Topiramate is helpful in some cases. Piroxicam has been used, although it is not as effective as indomethacin. Verapamil, an effective treatment for cluster headache, does not appear to be useful for PH. In occasional patients, PH can coexist with trigeminal neuralgia (PH-tic syndrome); similar to cluster-tic syndrome, each component may require separate treatment.

Secondary PH has been reported with lesions in the region of the sella turcica, including arteriovenous malformation, cavernous sinus meningioma, and epidermoid tumors. Secondary PH is more likely if the patient requires high doses (>200 mg/d) of indomethacin. In patients with apparent bilateral PH, raised CSF pressure should be suspected. It is important to note that indomethacin reduces CSF pressure. When a diagnosis of PH is considered, MRI is indicated to exclude a pituitary lesion.

SUNCT/SUNA

SUNCT (*s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing) is a rare primary headache syndrome characterized by severe, unilateral orbital or temporal pain that is stabbing or throbbing in quality. Diagnosis requires at least 20 attacks, lasting for 5–240 s; ipsilateral conjunctival injection and lacrimation should be present. In some patients conjunctival injection or lacrimation is missing, and the diagnosis of SUNA (*s*hort-lasting *u*nilateral *n*euralgiform headache attacks with cranial *a*utonomic symptoms) can be made.

Diagnosis

The pain of SUNCT/SUNA is unilateral and may be located anywhere in the head. Three basic patterns can be seen: single stabs, which are usually short-lived; groups of stabs; or a longer attack comprising many stabs between which the pain does not completely resolve, thus giving a "saw-tooth" phenomenon with attacks lasting many minutes. Each pattern may be seen in the context of an underlying continuous head pain. Characteristics that lead to a suspected diagnosis of SUNCT are the cutaneous (or other) triggerability of attacks, a lack of refractory period to triggering between attacks, and the lack of a response to indomethacin. Apart from trigeminal sensory disturbance, the neurologic examination is normal in primary SUNCT.

The diagnosis of SUNCT is often confused with trigeminal neuralgia (TN) particularly in first-division TN (Chap. 376). Minimal or no cranial autonomic symptoms and a clear refractory period to triggering indicate a diagnosis of TN.

Secondary (Symptomatic) Sunct

SUNCT can be seen with posterior fossa or pituitary lesions. All patients with SUNCT/SUNA should be evaluated with pituitary function tests and a brain MRI with pituitary views.

TREATMENT: SUNCT/SUNA

Abortive Therapy

Therapy of acute attacks is not a useful concept in SUNCT/SUNA since the attacks are of such short duration. However, IV lidocaine, which arrests the symptoms, can be used in hospitalized patients.

Preventive Therapy

Long-term prevention to minimize disability and hospitalization is the goal of treatment. The most effective treatment for prevention is lamotrigine, 200–400 mg/d. Topiramate and gabapentin may also be effective. Carbamazepine, 400–500 mg/d, has been reported by patients to offer modest benefit.

Surgical approaches such as microvascular decompression or destructive trigeminal procedures are seldom useful and often produce long-term complications. Greater occipital nerve injection has produced limited benefit in some patients. Occipital nerve stimulation is probably helpful in an important subgroup of these patients. Complete control with deep-brain stimulation of the posterior hypothalamic region was reported in a single patient. For intractable cases, short-term prevention with IV lidocaine can be effective, as can occipital nerve stimulation.

Chronic Daily Headache

The broad diagnosis of chronic daily headache (CDH) can be applied when a patient experiences headache on 15 days or more per month. CDH is not a single entity; it encompasses a number of different headache syndromes, including chronic TTH as well as headache secondary to trauma, inflammation, infection, medication overuse, and other causes (**Table 14–10**). Population-based estimates suggest that about 4% of adults have daily or near-daily headache. Daily headache may be primary or secondary, an important consideration in guiding management of this complaint.

Table 14-10 Classification of Chronic Daily Headache

Primary		
>4 h Daily	**<4 h Daily**	**Secondary**
Chronic migraine[a]	Chronic cluster headache[b]	Posttraumatic Head injury Iatrogenic Postinfectious
Chronic tension-type headache[a]	Chronic paroxysmal hemicrania	Inflammatory, such as Giant cell arteritis Sarcoidosis Behçet's syndrome

Hemicrania continua[a]	SUNCT/SUNA	Chronic CNS infection
New daily persistent headache[a]	Hypnic headache	Medication-overuse headache[a]

[a] May be complicated by analgesic overuse.

[b] Some patients may have headache >4 h/d.

Abbreviations: SUNA, *s*hort-lasting *u*nilateral *n*euralgiform headache *a*ttacks with cranial autonomic symptoms; SUNCT, *s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing.

APPROACH TO THE PATIENT: CHRONIC DAILY HEADACHE

The first step in the management of patients with CDH is to diagnose any underlying condition (Table 14–10). For patients with primary headaches, diagnosis of the headache type will guide therapy. Preventive treatments such as tricyclics, either amitriptyline or nortriptyline at doses up to 1 mg/kg, are very useful in patients with CDH arising from migraine or tension-type headache. Tricyclics are started in low doses (10–25 mg) daily and may be given 12 h before the expected time of awakening in order to avoid excess morning sleepiness. Anticonvulsants, such as topiramate, valproate, and gabapentin, are also useful in migraineurs. Flunarizine can also be very effective for some patients, as can methysergide or phenelzine.

Management of Medically Intractable Disabling Chronic Daily Headache

The management of medically intractable headache is difficult. At this time, the only promising approach is occipital nerve stimulation, which appears to modulate thalamic processing in migraine and has also shown promise in chronic cluster headache, SUNCT/SUNA, and hemicrania continua (see below).

Medication-Overuse Headache

Overuse of analgesic medication for headache can aggravate headache frequency and induce a state of refractory daily or near-daily headache called *medication-overuse headache*. A proportion of patients who stop taking analgesics will experience substantial improvement in the severity and frequency of their headache. However, even after cessation of analgesic use, many patients continue to have headache, although they may feel clinically improved in some way, especially if they have been using codeine or barbiturates regularly. The residual symptoms probably represent the underlying headache disorder.

Management of Medication Overuse: Outpatients

For patients who overuse medications, it is essential that analgesic use be reduced and eliminated. One approach is to reduce the medication dose by 10% every 1–2 weeks. Immediate cessation of analgesic use is possible for some patients, provided there is no contraindication. Both approaches are facilitated by the use of a medication diary maintained during the month or two before cessation; this helps to identify the scope of the problem. A small dose of an NSAID such as naproxen, 500 mg bid, if tolerated, will help relieve residual pain as analgesic use is reduced. NSAID overuse is not usually a problem for patients with daily headache when the dose is taken once or twice daily; however, overuse problems may develop with more frequent dosing schedules. Once the patient has substantially reduced analgesic use, a preventive medication should be introduced. It must be emphasized that *preventives generally do not work in the presence of analgesic overuse* . The most common cause of unresponsiveness to treatment is the use of a preventive when analgesics continue to be used regularly. For some patients, discontinuing analgesics is very difficult; often the best approach is to directly inform the patient that some degree of pain is inevitable during this initial period.

Management of Medication Overuse: Inpatients

Some patients will require hospitalization for detoxification. Such patients have typically failed efforts at outpatient withdrawal or have a significant medical condition, such as diabetes mellitus, which would complicate withdrawal as an outpatient. Following admission to the hospital, acute medications are withdrawn completely on the first day, in the absence of a contraindication. Antiemetics and fluids are administered as required; clonidine is used for opiate withdrawal symptoms. For acute intolerable pain during the waking hours aspirin, 1 g IV (not approved in United States), is useful. IM chlorpromazine can be helpful at night; patients must be adequately hydrated. Three to five days into the admission as the effect of the withdrawn substance settles a course of IV dihydroergotamine (DHE) can be employed. DHE, administered every 8 h for 5 consecutive days, can induce a significant remission that allows a preventive treatment to be established. 5-HT_3 antagonists, such as ondansetron or granisetron, are often required with DHE to prevent significant nausea, and domperidone (not approved in the United States) orally or by suppository can be very helpful.

New Daily Persistent Headache

New daily persistent headache (NDPH) is a clinically distinct syndrome; its causes are listed in **Table**

14–11.

Table 14-11 Differential Diagnosis of New Daily Persistent Headache	
Primary	**Secondary**
Migrainous-type	Subarachnoid hemorrhage
Featureless (tension-type)	Low CSF volume headache
	Raised CSF pressure headache
	Posttraumatic headache[a]
	Chronic meningitis

[a] Includes postinfectious forms.

Clinical Presentation

The patient with NDPH presents with headache on most if not all days and the patient can clearly, and often vividly, recall the moment of onset. The headache usually begins abruptly, but onset may be more gradual; evolution over 3 days has been proposed as the upper limit for this syndrome. Patients typically recall the exact day and circumstances of the onset of headache; the new, persistent head pain does not remit. The first priority is to distinguish between a primary and a secondary cause of this syndrome. Subarachnoid hemorrhage is the most serious of the secondary causes and must be excluded either by history or appropriate investigation (Chap. 275).

Secondary Ndph

Low CSF Volume Headache

In these syndromes, head pain is positional: it begins when the patient sits or stands upright and resolves upon reclining. The pain, which is occipitofrontal, is usually a dull ache but may be throbbing. Patients with chronic low CSF volume headache typically present with a history of headache from one day to the next that is generally not present on waking but worsens during the day. Recumbency usually improves the headache within minutes, but it takes only minutes to an hour for the pain to return when the patient resumes an upright position.

The most common cause of headache due to persistent low CSF volume is CSF leak following lumbar puncture (LP). Post-LP headache usually begins within 48 h but may be delayed for up to 12 days. Its incidence is between 10 and 30%. Beverages with caffeine may provide temporary relief. Besides LP, index events may include epidural injection or a vigorous Valsalva maneuver, such as from lifting, straining, coughing, clearing the eustachian tubes in an airplane, or multiple orgasms. Spontaneous CSF leaks are well recognized, and the diagnosis should be considered whenever the headache history is typical, even when there is no obvious index event. As time passes from the index event, the postural nature may become less apparent; cases in which the index event occurred several years before the eventual diagnosis have been recognized. Symptoms appear to result from low volume rather than low pressure: although low CSF pressures, typically 0–50 mmH_2O, are usually identified, a pressure as high as 140 mmH_2O has been noted with a documented leak.

Postural orthostatic tachycardia syndrome [POTS (Chap. 375)] can present with orthostatic headache similar to low CSF volume headache and is a diagnosis that needs consideration here.

When imaging is indicated to identify the source of a presumed leak, an MRI with gadolinium is the initial study of choice **(Fig. 14-5)**. A striking pattern of diffuse meningeal enhancement is so typical that in the appropriate clinical context the diagnosis is established. Chiari malformations may sometimes be noted on MRI; in such cases, surgery to decompress the posterior fossa usually worsens the headache. Spinal MRI with T2 weighting may reveal a leak and spinal MRI may demonstrate spinal meningeal cysts whose role in these syndromes is yet to be elucidated. The source of CSF leakage may be identified by spinal MRI, by CT, or increasingly with MR myelography, or with ^{111}In-DTPA CSF studies; in the absence of a directly identified site of leakage, early emptying of ^{111}In-DTPA tracer into the bladder or slow progress of tracer across the brain suggests a CSF leak.

Figure 14-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Magnetic resonance image showing diffuse meningeal enhancement after gadolinium administration in a patient with low CSF volume headache.

Initial treatment for low CSF volume headache is bed rest. For patients with persistent pain, IV caffeine (500 mg in 500 mL saline administered over 2 h) can be very effective. An ECG to screen for arrhythmia should be performed before administration. It is reasonable to administer at least two infusions of caffeine before embarking on additional tests to identify the source of the CSF leak. Since IV caffeine is safe and can be curative, it spares many patients the need for further investigations. If unsuccessful, an abdominal binder may be helpful. If a leak can be identified, an autologous blood patch is usually curative. A blood patch is also effective for post-LP headache; in this setting, the location is empirically determined to be the site of the LP. In patients with intractable pain, oral theophylline is a useful alternative; however, its effect is less rapid than caffeine.

Raised CSF Pressure Headache

Raised CSF pressure is well recognized as a cause of headache. Brain imaging can often reveal the cause, such as a space-occupying lesion. NDPH due to raised CSF pressure can be the presenting symptom for patients with idiopathicintracranial hypertension (pseudotumor cerebri) without visual problems, particularly when the fundi are normal. Persistently raised intracranial pressure can trigger chronic migraine. These patients typically present with a history of generalized headache that is present on waking and improves as the day goes on. It is generally worse with recumbency. Visual obscurations are frequent. The diagnosis is relatively straightforward when papilledema is present, but the possibility must be considered even in patients without funduscopic changes. Formal visual field testing should be performed even in the absence of overt ophthalmic involvement. Headache on rising in the morning or nocturnal headache is also characteristic of obstructive sleep apnea or poorly controlled hypertension.

Evaluation of patients suspected to have raised CSF pressure requires brain imaging. It is most efficient to obtain an MRI, including an MR venogram, as the initial study. If there are no contraindications, the CSF pressure should be measured by LP; this should be done when the patient is symptomatic so that both the pressure and the response to removal of 20–30 mL of CSF can be determined. An elevated opening pressure and improvement in headache following removal of CSF is diagnostic.

Initial treatment is with acetazolamide (250–500 mg bid); the headache may improve within weeks. If ineffective, topiramate is the next treatment of choice; it has many actions that may be useful in this setting, including carbonic anhydrase inhibition, weight loss, and neuronal membrane stabilization, likely mediated via effects on phosphorylation pathways. Severely disabled patients who do not respond to medical treatment require intracranial pressure monitoring and may require shunting.

Post-Traumatic Headache

A traumatic event can trigger a headache process that lasts for many months or years after the event. The term *trauma* is used in a very broad sense: headache can develop following an injury to the head, but it can also develop after an infectious episode, typically viral meningitis, a flulike illness, or a

parasitic infection. Complaints of dizziness, vertigo, and impaired memory can accompany the headache. Symptoms may remit after several weeks or persist for months and even years after the injury. Typically the neurologic examination is normal and CT or MRI studies are unrevealing. Chronic subdural hematoma may on occasion mimic this disorder. In one series, one-third of patients with NDPH reported headache beginning after a transient flulike illness characterized by fever, neck stiffness, photophobia, and marked malaise. Evaluation reveals no apparent cause for the headache. There is no convincing evidence that persistent Epstein-Barr infection plays a role in this syndrome. A complicating factor is that many patients undergo LP during the acute illness; iatrogenic low CSF volume headache must be considered in these cases. Posttraumatic headache may also be seen after carotid dissection and subarachnoid hemorrhage, and following intracranial surgery. The underlying theme appears to be that a traumatic event involving the pain-producing meninges can trigger a headache process that lasts for many years.

Treatment is largely empirical. Tricyclic antidepressants, notably amitriptyline, and anticonvulsants such as topiramate, valproate, and gabapentin, have been used with reported benefit. The MAOI phenelzine may also be useful in carefully selected patients. The headache usually resolves within 3–5 years, but it can be quite disabling.

Primary NDPH

Primary NDPH occurs in both males and females. It can be of the migrainous type, with features of migraine, or it can be featureless, appearing as new-onset TTH (Table 14–11). Migrainous features are common and include unilateral headache and throbbing pain; each feature is present in about one-third of patients. Nausea, photophobia, and/or phonophobia occur in about half of patients. Some patients have a previous history of migraine; however, the proportion of NDPH sufferers with preexisting migraine is no greater than the frequency of migraine in the general population. At 24 months, ~86% of patients are headache-free. Treatment of migrainous-type primary NDPH consists of using the preventive therapies effective in migraine (Table 14–7). Featureless NDPH is one of the primary headache forms most refractory to treatment. Standard preventive therapies can be offered but are often ineffective.

Other Primary Headaches

HEMICRANIA CONTINUA

The essential features of hemicrania continua are moderate and continuous unilateral pain associated with fluctuations of severe pain; complete resolution of pain with indomethacin; and exacerbations that may be associated with autonomic features, including conjunctival injection, lacrimation, and photophobia on the affected side. The age of onset ranges from 11 to 58 years; women are affected twice as often as men. The cause is unknown.

Treatment: Hemicrania Continua

Treatment consists of indomethacin; other NSAIDs appear to be of little or no benefit. The IM injection of 100 mg indomethacin has been proposed as a diagnostic tool and administration with a placebo injection in a blinded fashion can be very useful diagnostically. Alternatively, a trial of oral indomethacin, starting with 25 mg tid, then 50 mg tid, and then 75 mg tid, can be given. Up to two weeks at the maximal dose may be necessary to assess whether a dose has a useful effect. Topiramate can be helpful in some patients. Occipital nerve stimulation may have a role in patients with hemicrania continua who are unable to tolerate indomethacin.

Primary Stabbing Headache

The essential features of primary stabbing headache are stabbing pain confined to the head or, rarely, the face, lasting from 1 to many seconds or minutes and occurring as a single stab or a series of stabs; absence of associated cranial autonomic features; absence of cutaneous triggering of attacks; and a pattern of recurrence at irregular intervals (hours to days). The pains have been variously described as "ice-pick pains" or "jabs and jolts." They are more common in patients with other primary headaches, such as migraine, the TACs, and hemicrania continua.

TREATMENT: PRIMARY STABBING HEADACHE

The response of primary stabbing headache to indomethacin (25–50 mg two to three times daily) is usually excellent. As a general rule, the symptoms wax and wane, and after a period of control on indomethacin, it is appropriate to withdraw treatment and observe the outcome.

PRIMARY COUGH HEADACHE

Primary cough headache is a generalized headache that begins suddenly, lasts for several minutes, and is precipitated by coughing; it is preventable by avoiding coughing or other precipitating events, which can include sneezing, straining, laughing, or stooping. In all patients with this syndrome, serious etiologies must be excluded before a diagnosis of "benign" primary cough headache can be established. A Chiari malformation or any lesion causing obstruction of CSF pathways or displacing cerebral structures can be the cause of the head pain. Other conditions that can present with cough or exertional headache as the initial symptom include cerebral aneurysm, carotid stenosis, and

vertebrobasilar disease. Benign cough headache can resemble benign exertional headache (below), but patients with the former condition are typically older.

Treatment: Primary Cough Headache

Indomethacin 25–50 mg two to three times daily is the treatment of choice. Some patients with cough headache obtain pain relief with LP; this is a simple option when compared to prolonged use of indomethacin, and it is effective in about one-third of patients. The mechanism of this response is unclear.

PRIMARY EXERTIONAL HEADACHE

Primary exertional headache has features resembling both cough headache and migraine. It may be precipitated by any form of exercise; it often has the pulsatile quality of migraine. The pain, which can last from 5 min to 24 h, is bilateral and throbbing at onset; migrainous features may develop in patients susceptible to migraine. Primary exertional headache can be prevented by avoiding excessive exertion, particularly in hot weather or at high altitude.

The mechanism of primary exertional headache is unclear. Acute venous distension likely explains one syndrome, the acute onset of headache with straining and breath holding, as in weightlifter's headache. As exertion can result in headache in a number of serious underlying conditions, these must be considered in patients with exertional headache. Pain from angina may be referred to the head, probably by central connections of vagal afferents, and may present as exertional headache (cardiac cephalgia). The link to exercise is the main clinical clue that headache is of cardiac origin. Pheochromocytoma may occasionally cause exertional headache. Intracranial lesions and stenosis of the carotid arteries are other possible etiologies.

Treatment: Primary Exertional Headache

Exercise regimens should begin modestly and progress gradually to higher levels of intensity. Indomethacin at daily doses from 25 to 150 mg is generally effective in benign exertional headache. Indomethacin (50 mg), ergotamine (1 mg orally), dihydroergotamine (2 mg by nasal spray), or methysergide (1–2 mg orally given 30–45 min before exercise) are useful prophylactic measures.

PRIMARY SEX HEADACHE

Sex headache is precipitated by sexual excitement. The pain usually begins as a dull bilateral headache that suddenly becomes intense at orgasm. The headache can be prevented or eased by ceasing sexual activity before orgasm. Three types of sex headache are reported: a dull ache in the head and neck that intensifies as sexual excitement increases; a sudden, severe, explosive headache occurring at orgasm; and a postural headache developing after coitus that resembles the headache of low CSF pressure. The latter arises from vigorous sexual activity and is a form of low CSF pressure headache. Headaches developing at the time of orgasm are not always benign; 5–12% of cases of subarachnoid hemorrhage are precipitated by sexual intercourse. Sex headache is reported by men more often than women and may occur at any time during the years of sexual activity. It may develop on several occasions in succession and then not trouble the patient again, even without an obvious change in sexual activity. In patients who stop sexual activity when headache is first noticed, the pain may subside within a period of 5 min to 2 h. In about half of patients, sex headache will subside within 6 months. About half of patients with sex headache have a history of exertional headaches, but there is no excess of cough headache. Migraine is probably more common in patients with sex headache.

Treatment: Primary Sex Headache

Benign sex headaches recur irregularly and infrequently. Management can often be limited to reassurance and advice about ceasing sexual activity if a mild, warning headache develops. Propranolol can be used to prevent headache that recurs regularly or frequently, but the dosage required varies from 40 to 200 mg/d. An alternative is the calcium channel–blocking agent diltiazem, 60 mg tid. Ergotamine (1 mg) or indomethacin (25–50 mg) taken about 30–45 min prior to sexual activity can also be helpful.

PRIMARY THUNDERCLAP HEADACHE

Sudden onset of severe headache may occur in the absence of any known provocation. The differential diagnosis includes the sentinel bleed of an intracranial aneurysm, cervicocephalic arterial dissection, and cerebral venous thrombosis. Headaches of explosive onset may also be caused by the ingestion of sympathomimetic drugs or of tyramine-containing foods in a patient who is taking MAOIs, or they may be a symptom of pheochromocytoma. Whether thunderclap headache can be the presentation of an unruptured cerebral aneurysm is uncertain. When neuroimaging studies and LP exclude subarachnoid hemorrhage, patients with thunderclap headache usually do very well over the long term. In one study of patients whose CT scans and CSF findings were negative, ~15% had recurrent episodes of thunderclap headache, and nearly half subsequently developed migraine or tension-type headache.

The first presentation of any sudden-onset severe headache should be vigorously investigated with neuroimaging (CT or, when possible, MRI with MR angiography) and CSF examination. Formal

cerebral angiography should be reserved for those cases in which no primary diagnosis is forthcoming and for clinical situations that are particularly suggestive of intracranial aneurysm. Reversible segmental cerebral vasoconstriction may be seen in primary thunderclap headache without an intracranial aneurysm. In the presence of posterior leukoencephalopathy, the differential diagnosis includes cerebral angiitis, drug toxicity (cyclosporine, intrathecal methotrexate/cytarabine, pseudoephedrine, or cocaine), posttransfusion effects, and postpartum angiopathy. Treatment with nimodipine may be helpful, although by definition the vasoconstriction of primary thunderclap headache resolves spontaneously.

HYPNIC HEADACHE

This headache syndrome typically begins a few hours after sleep onset. The headaches last from 15 to 30 min and are typically moderately severe and generalized, although they may be unilateral and can be throbbing. Patients may report falling back to sleep only to be awakened by a further attack a few hours later; up to three repetitions of this pattern occur through the night. Daytime naps can also precipitate head pain. Most patients are female, and the onset is usually after age 60 years. Headaches are bilateral in most, but may be unilateral. Photophobia or phonophobia and nausea are usually absent. The major secondary consideration in this headache type is poorly controlled hypertension; 24-h blood pressure monitoring is recommended to detect this treatable condition.

Treatment: Hypnic Headache

Patients with hypnic headache generally respond to a bedtime dose of lithium carbonate (200–600 mg). For those intolerant of lithium, verapamil (160 mg) or methysergide (1–4 mg at bedtime) may be alternative strategies. One to two cups of coffee or caffeine, 60 mg orally, at bedtime may be effective in approximately one-third of patients. Case reports suggest that flunarizine, 5 mg nightly, can be effective.

FURTHER READINGS

Cittadini E, Goadsby PJ: Hemicrania continua: A clinical study of 39 patients with diagnostic implications. Brain 133:1973, 2010 [PMID: 20558416]

Cohen AS et al: High-flow oxygen for treatment of cluster headache: A randomized trial. JAMA 302:2451, 2009 [PMID: 19996400]

Goadsby PJ et al: Chronic Daily Headache for Clinicians. Hamilton, Ontario, Canada, BC Decker, 2005

Goadsby PJ et al: Trigeminal autonomic cephalalgias: Paroxysmal hemicrania, SUNCT/SUNA and hemicrania continua. Semin Neurol 30:186, 2010 [PMID: 20352588]

Headache Classification Committee of The International Headache Society: *The International Classification of Headache Disorders*, 2nd ed. Cephalalgia 24:1, 2004

Lance JW, Goadsby PJ: *Mechanism and Management of Headache* , 7th ed. Philadelphia, Elsevier, 2005

Lipton RB, Bigal M: *Migraine and Other Headache Disorders.* New York, Marcel Dekker, Taylor & Francis, 2006

Loder E: Triptan therapy in migraine. N Engl J Med 363:63, 2010 [PMID: 20592298]

Olesen J et al: *The Headaches*. Philadelphia, Lippincott, Williams & Wilkins, 2005

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BACK AND NECK PAIN: INTRODUCTION

The importance of back and neck pain in our society is underscored by the following: (1) the cost of back pain in the United States exceeds $100 billion annually; approximately one-third of these costs are direct health care expenses, and two-thirds are indirect costs resulting from loss of wages and productivity; (2) back symptoms are the most common cause of disability in those <45 years; (3) low back pain is the second most common reason for visiting a physician in the United States; and (4) ~1% of the U.S. population is chronically disabled because of back pain.

ANATOMY OF THE SPINE

The anterior portion of the spine consists of cylindrical vertebral bodies separated by intervertebral disks and held together by the anterior and posterior longitudinal ligaments. The intervertebral disks are composed of a central gelatinous nucleus pulposus surrounded by a tough cartilaginous ring, the annulus fibrosis. Disks are responsible for 25% of spinal column length and allow the bony vertebrae to move easily upon each other (Figs. 15-1 and 15-2). Desiccation of the nucleus pulposus and degeneration of the annulus fibrosus increase with age and results in loss of height. The disks are largest in the cervical and lumbar regions where movements of the spine are greatest. The functions of the anterior spine are to absorb the shock of body movements such as walking and running, and to protect the contents of the spinal canal.

Figure 15-1

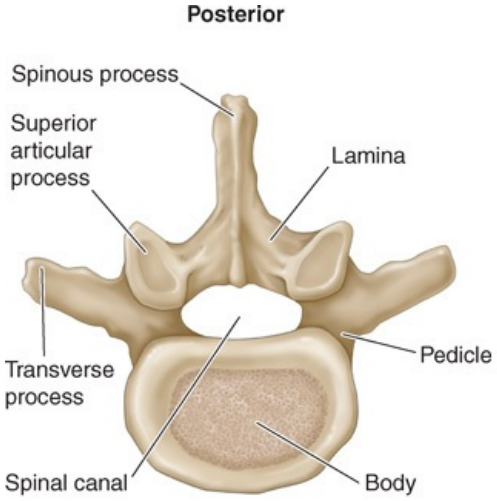

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

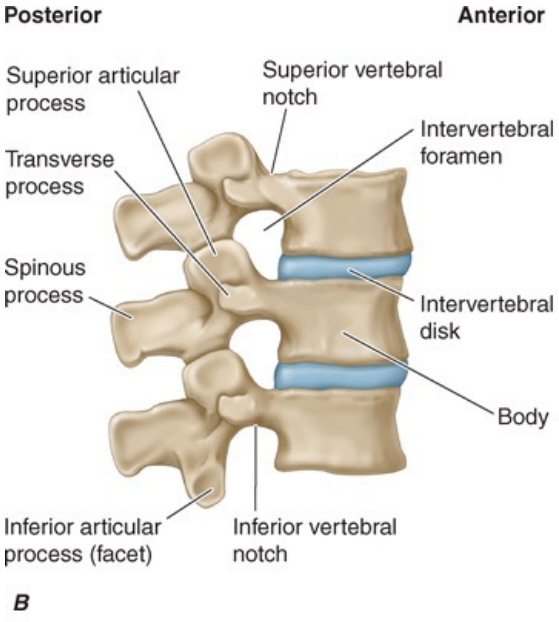

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vertebral anatomy. *(From A Gauthier Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998; with permission.)*

Figure 15-2

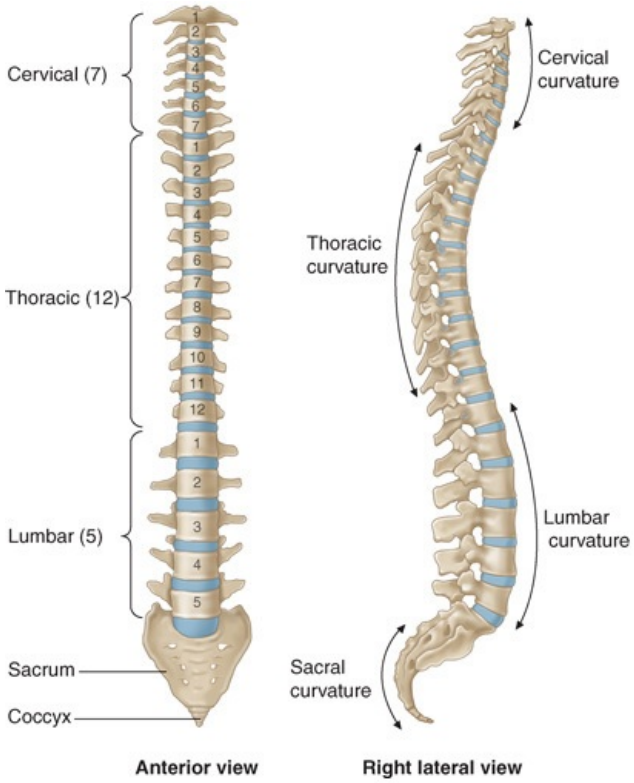

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spinal column. *(From A Gauthier Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998; with permission.)*

The posterior portion of the spine consists of the vertebral arches and processes. Each arch consists of paired cylindrical pedicles anteriorly and paired laminae posteriorly. The vertebral arch also gives rise to two transverse processes laterally, one spinous process posteriorly, plus two superior and two

inferior articular facets. The apposition of a superior and inferior facet constitutes a *facet joint*. The functions of the posterior spine are to protect the spinal cord and nerves within the spinal canal and to provide an anchor for the attachment of muscles and ligaments. The contraction of muscles attached to the spinous and transverse processes and laminae works like a system of pulleys and levers that results in flexion, extension, and lateral bending movements of the spine.

Nerve root injury (*radiculopathy*) is a common cause of neck, arm, low back, buttock, and leg pain (Figs. 23-2 and 23-3). The nerve roots exit at a level above their respective vertebral bodies in the cervical region (e.g., the C7 nerve root exits at the C6-C7 level) and below their respective vertebral bodies in the thoracic and lumbar regions (e.g., the T1 nerve root exits at the T1-T2 level). The cervical nerve roots follow a short intraspinal course before exiting. By contrast, because the spinal cord ends at the vertebral L1 or L2 level, the lumbar nerve roots follow a long intraspinal course and can be injured anywhere from the upper lumbar spine to their exit at the intervertebral foramen. For example, disk herniation at the L4-L5 level can produce not only L5 root compression, but also compression of the traversing S1 nerve root **(Fig. 15-3)**.

Figure 15-3

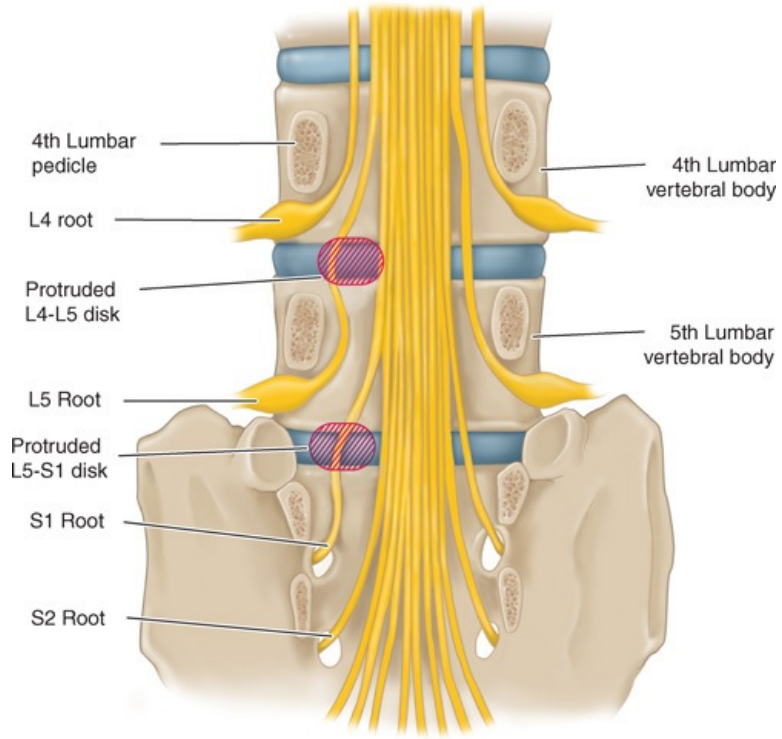

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Compression of L5 and S1 roots by herniated disks. (*From Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009; with permission.*)

Pain-sensitive structures of the spine include the periosteum of the vertebrae, dura, facet joints, annulus fibrosus of the intervertebral disk, epidural veins and arteries, and the posterior longitudinal ligament. Disease of these diverse structures may explain many cases of back pain without nerve root compression. The nucleus pulposus of the intervertebral disk is not pain-sensitive under normal circumstances. Pain sensation from within the spinal canal is conveyed partially by the sinuvertebral nerve that arises from the spinal nerve at each spine segment and reenters the spinal canal through the intervertebral foramen at the same level.

Approach to the Patient: Back Pain

TYPES OF BACK PAIN

Understanding the types of pain reported by patients is the essential first step. Attention is also focused on identification of risk factors for serious underlying diseases; the majority of these are due to radiculopathy, fracture, tumor, infection, or referred pain from visceral structures **(Table 15-1)**.

Table 15-1 Acute Low Back Pain: Risk Factors for an Important Structural Cause

History

Pain worse at rest or at night
Prior history of cancer History of chronic infection (esp. lung, urinary tract, skin)
History of trauma
Incontinence Age >70 years
Intravenous drug use
Glucocorticoid use History of a rapidly progressive neurologic deficit
Examination
Unexplained fever
Unexplained weight loss Percussion tenderness over the spine
Abdominal, rectal, or pelvic mass
Patrick's sign or heel percussion sign Straight leg or reverse straight leg–raising signs
Progressive focal neurologic deficit

Local pain is caused by injury to pain-sensitive structures that compress or irritate sensory nerve endings. The site of the pain is near the affected part of the back.

Pain referred to the back may arise from abdominal or pelvic viscera. The pain is usually described as primarily abdominal or pelvic but is accompanied by back pain and usually unaffected by posture. The patient may occasionally complain of back pain only.

Pain of spine origin may be located in the back or referred to the buttocks or legs. Diseases affecting the upper lumbar spine tend to refer pain to the lumbar region, groin, or anterior thighs. Diseases affecting the lower lumbar spine tend to produce pain referred to the buttocks, posterior thighs, or rarely the calves or feet. Referred or "sclerotomal" pain may explain instances where the pain crosses multiple dermatomes without evidence of nerve root compression.

Radicular back pain is typically sharp and radiates from the low back to a leg within the territory of a nerve root (see "Lumbar Disk Disease," below). Coughing, sneezing, or voluntary contraction of abdominal muscles (lifting heavy objects or straining at stool) may elicit the radiating pain. The pain may increase in postures that stretch the nerves and nerve roots. Sitting with the leg outstretched places traction on the sciatic nerve and L5 and S1 roots because the nerve passes posterior to the hip. The femoral nerve (L2, L3, and L4 roots) passes anterior to the hip and is not stretched by sitting. The description of the pain alone often fails to distinguish between sclerotomal pain and radiculopathy.

Pain associated with muscle spasm, although of obscure origin, is commonly associated with many spine disorders. The spasms are accompanied by abnormal posture, tense paraspinal muscles, and dull or achy pain in the paraspinal region.

Knowledge of the circumstances associated with the onset of back pain is important when weighing possible serious underlying causes for the pain. Some patients involved in accidents or work-related injuries may exaggerate their pain for the purpose of compensation or for psychological reasons.

EXAMINATION OF THE BACK

A physical examination that includes the abdomen and rectum is advisable. Back pain referred from visceral organs may be reproduced during palpation of the abdomen [pancreatitis, abdominal aortic aneurysm (AAA)] or percussion over the costovertebral angles (pyelonephritis).

The normal spine has a cervical and lumbar lordosis, and a thoracic kyphosis. Exaggeration of these normal alignments may result in hyperkyphosis of the thoracic spine or hyperlordosis of the lumbar spine. Inspection may reveal a lateral curvature of the spine (scoliosis) or an asymmetry in the prominence of the paraspinal muscles, suggesting muscle spasm. Back pain of bony spine origin is often reproduced by palpation or percussion over the spinous process of the affected vertebrae.

Forward bending is often limited by paraspinal muscle spasm; the latter may flatten the usual lumbar lordosis. Flexion at the hips is normal in patients with lumbar spine disease, but flexion of the lumbar spine is limited and sometimes painful. Lateral bending to the side opposite the injured spinal element may stretch the damaged tissues, worsen pain, and limit motion. Hyperextension of the spine (with the patient prone or standing) is limited when nerve root compression, facet joint pathology, or other bony spine disease is present.

Pain from hip disease may mimic the pain of lumbar spine disease. Hip pain can be reproduced by internal and external rotation at the hip with the knee and hip in flexion (Patrick's sign) and by tapping the heel with the examiner's palm while the leg is extended (heel percussion sign).

With the patient supine, passive flexion of the extended leg at the hip stretches the L5 and S1 nerve roots and the sciatic nerve (straight leg–raising maneuver). Passive dorsiflexion of the foot during the maneuver adds to the stretch. While flexion to at least 80° is normally possible without causing pain, many patients normally report a tight, stretching sensation in the hamstring muscles unrelated to back pain. The *straight leg–raising (SLR)* test is positive if the maneuver reproduces the patient's usual back or limb pain. Eliciting the SLR sign in the sitting position can help determine if the finding is reproducible. The patient may describe pain in the low back, buttocks, posterior thigh, or lower leg, but the *key feature is reproduction of the patient's usual pain* . The *crossed SLR sign* is positive when flexion of one leg reproduces the usual pain in the opposite leg or buttocks. The crossed SLR sign is less sensitive but more specific for disk herniation than the SLR sign. The nerve or nerve root lesion is always on the side of the pain. The *reverse SLR sign* is elicited by standing the patient next to the examination table and passively extending each leg with the knee fully extended. This maneuver, which stretches the L2-L4 nerve roots, lumbosacral plexus, and femoral nerve, is considered positive if the patient's usual back or limb pain is reproduced.

The neurologic examination includes a search for focal weakness or muscle atrophy, focal reflex changes, diminished sensation in the legs, or signs of spinal cord injury. The examiner should be alert to the possibility of breakaway weakness, defined as fluctuating strength during muscle testing. Breakaway weakness may be due to pain or a combination of pain and underlying true weakness. Breakaway weakness without pain is almost always due to a lack of effort. In uncertain cases, electromyography (EMG) can determine whether or not true weakness due to nerve tissue injury is present. Findings with specific nerve lumbosacral nerve root lesions are shown in **Table 15-2** and are discussed below.

Table 15-2 Lumbosacral Radiculopathy—Neurologic Features

	Examination Findings			
Lumbosacral Nerve Roots	**Reflex**	**Sensory**	**Motor**	**Pain Distribution**
L2[a]	—	Upper anterior thigh	Psoas (hip flexion)	Anterior thigh
L3[a]	—	Lower anterior thigh Anterior knee	Psoas (hip flexion) Quadriceps (knee extension) Thigh adduction	Anterior thigh, knee
L4[a]	Quadriceps (knee)	Medial calf	Quadriceps (knee extension)[b] Thigh adduction Tibialis anterior (foot dorsiflexion)	Knee, medial calf Anterolateral thigh
L5[c]	—	Dorsal surface —foot Lateral calf	Peroneii (foot eversion)[b] Tibialis anterior (foot dorsiflexion) Gluteus medius (hip abduction) Toe dorsiflexors	Lateral calf, dorsal foot, posterolateral thigh, buttocks
S1[c]	Gastrocnemius/soleus (ankle)	Plantar surface —foot Lateral aspect —foot	Gastrocnemius/soleus (foot plantar flexion)[b] Abductor hallucis (toe flexors)[b] Gluteus maximus (hip extension)	Bottom foot, posterior calf, posterior thigh, buttocks

[a]Reverse straight leg–raising sign present—see "Examination of the Back."

[b]These muscles receive the majority of innervation from this root.

[c]Straight leg–raising sign present—see "Examination of the Back."

LABORATORY, IMAGING, AND EMG STUDIES

Routine laboratory studies are rarely needed for the initial evaluation of nonspecific acute (<3 months duration) low back pain (ALBP). If risk factors for a serious underlying cause are present (Table 15-1), then laboratory studies [complete blood count (CBC), erythrocyte sedimentation rate (ESR), urinalysis] are indicated.

CT scanning is superior to routine x-rays for the detection of fractures involving posterior spine structures, craniocervical and craniothoracic junctions, C1 and C2 vertebrae, bone fragments within the spinal canal, or misalignment; CT scans are increasingly used as a primary screening modality for moderate to severe trauma. In the absence of risk factors, these imaging studies are rarely helpful in nonspecific ALBP. MRI and CT-myelography are the radiologic tests of choice for evaluation of most serious diseases involving the spine. MRI is superior for the definition of soft tissue structures, whereas CT-myelography provides optimal imaging of the lateral recess of the spinal canal, and is better tolerated by claustrophobic patients. While the added diagnostic value of modern neuroimaging is significant, there is concern that these studies may be overutilized in patients with benign ALBP.

Electrodiagnostic studies can be used to assess the functional integrity of the peripheral nervous system (Chap. e45). Sensory nerve conduction studies are normal when focal sensory loss is due to nerve root damage because the nerve roots are proximal to the nerve cell bodies in the dorsal root ganglia. Injury to nerve tissue distal to the dorsal root ganglion (e.g., plexus or peripheral nerve) results in reduced sensory nerve signals. Needle EMG complements nerve conduction studies by detecting denervation or reinnervation changes in a myotomal (segmental) distribution. Multiple muscles supplied by different nerve roots and nerves are sampled; the pattern of muscle involvement indicates the nerve root(s) responsible for the injury. Needle EMG provides objective information about motor nerve fiber injury when clinical evaluation of weakness is limited by pain or poor effort. EMG and nerve conduction studies will be normal when sensory nerve root injury or irritation is the source of the pain.

CAUSES OF BACK PAIN

(Table 15-3)

Table 15-3 Causes of Back or Neck Pain
Congenital/Developmental
Spondylolysis and spondylolisthesis
Kyphoscoliosis
Spina bifida occulta
Tethered spinal cord
Minor Trauma
Strain or sprain
Whiplash injury
Fractures
Trauma—falls, motor vehicle accidents
Atraumatic fractures—osteoporosis, neoplastic infiltration, exogenous steroids, osteomyelitis
Intervertebral Disk Herniation
Degenerative
Intervertebral foraminal narrowing
Disk-osteophyte complex
Internal disk disruption
LSS with neurogenic claudication
Uncovertebral joint disease
Atlantoaxial joint disease (e.g., rheumatoid arthritis)

Arthritis
Spondylosis
Facet or sacroiliac arthropathy
Neoplasms—metastatic, hematologic, primary bone tumors
Infection/Inflammation
Vertebral osteomyelitis
Spinal epidural abscess
Septic disk (discitis)
Meningitis
Lumbar arachnoiditis
Autoimmune [e.g., ankylosing spondylitis, reactive arthritis (formerly known as Reiter's syndrome)]
Metabolic
Osteoporosis—hyperparathyroidism, immobility
Osteosclerosis (e.g., Paget's disease)
Vascular
Abdominal aortic aneurysm
Vertebral artery dissection
Other
Referred pain from visceral disease
Postural
Psychiatric, malingering, chronic pain syndromes

Congenital Anomalies of the Lumbar Spine

Spondylolysis is a bony defect in the vertebral pars interarticularis (a segment near the junction of the pedicle with the lamina); the cause is usually a stress microfracture in a congenitally abnormal segment. It occurs in up to 6% of adolescents. The defect (usually bilateral) is best visualized on plain x-rays, CT scan, or bone scan and is frequently asymptomatic. Symptoms may occur in the setting of a single injury, repeated minor injuries, or growth. Spondylolysis is the most common cause of persistent low back pain in adolescents and is often associated with sports-related activities.

Spondylolisthesis is the anterior slippage of the vertebral body, pedicles, and superior articular facets, leaving the posterior elements behind. Spondylolisthesis can be associated with spondylolysis, congenital anomalies, degenerative spine disease, or other causes of mechanical weakness of the pars (e.g., infection, osteoporosis, tumor, trauma, prior surgery). The slippage may be asymptomatic or may cause low back pain and hamstring tightness, nerve root injury (the L5 root most frequently), symptomatic spinal stenosis, or cauda equina syndrome (CES) in severe cases. Tenderness may be elicited near the segment that has "slipped" forward (most often L4 on L5 or occasionally L5 on S1). A "step" may be present on deep palpation of the posterior elements of the segment above the spondylolisthetic joint. The trunk may be shortened and the abdomen protuberant as a result. Anterolisthesis or retrolisthesis can also occur at other cervical or lumbar levels in adults and be the source of neck or low back pain. Plain x-rays with the neck or low back in flexion and extension will reveal the movement at the abnormal spinal segment. Surgery is considered for pain symptoms that do not respond to conservative measures (e.g., rest, physical therapy), and in cases with progressive neurologic deficit, postural deformity, slippage >50%, or scoliosis.

Spina bifida occulta is a failure of closure of one or several vertebral arches posteriorly; the meninges and spinal cord are normal. A dimple or small lipoma may overlie the defect. Most cases are asymptomatic and discovered incidentally during an evaluation for back pain.

Tethered cord syndrome usually presents as a progressive cauda equina disorder (see below), although myelopathy may also be the initial manifestation. The patient is often a young adult who complains of perineal or perianal pain, sometimes following minor trauma. MRI studies reveal a low-

lying conus (below L1-L2) and a short and thickened filum terminale.

Trauma

A patient complaining of back pain and an inability to move the legs may have a spinal fracture or dislocation, and, with fractures above L1, spinal cord compression. Care must be taken to avoid further damage to the spinal cord or nerve roots by immobilizing the back pending the results of radiologic studies.

SPRAINS AND STRAINS

The terms *low back sprain*, *strain*, and *mechanically induced muscle spasm* refer to minor, self-limited injuries associated with lifting a heavy object, a fall, or a sudden deceleration such as in an automobile accident. These terms are used loosely and do not clearly describe a specific anatomic lesion. The pain is usually confined to the lower back, and there is no radiation to the buttocks or legs. Patients with paraspinal muscle spasm often assume unusual postures.

TRAUMATIC VERTEBRAL FRACTURES

Most traumatic fractures of the lumbar vertebral bodies result from injuries producing anterior wedging or compression. With severe trauma, the patient may sustain a fracture-dislocation or a "burst" fracture involving the vertebral body and posterior elements. Traumatic vertebral fractures are caused by falls from a height (a pars interarticularis fracture of the L5 vertebra is common), sudden deceleration in an automobile accident, or direct injury. Neurologic impairment is common, and early surgical treatment is indicated. In victims of blunt trauma, CT scans of the chest, abdomen, or pelvis can be reformatted to detect associated vertebral fractures.

Lumbar Disk Disease

This is a common cause of chronic or recurrent low back and leg pain (Figs. 15-3 and **15-4)**. Disk disease is most likely to occur at the L4-L5 or L5-S1 levels, but upper lumbar levels are involved occasionally. The cause is often unknown; the risk is increased in overweight individuals. Disk herniation is unusual prior to age 20 years and is rare in the fibrotic disks of the elderly. Genetic factors may play a role in predisposing some patients to disk disease. The pain may be located in the low back only or referred to a leg, buttock, or hip. A sneeze, cough, or trivial movement may cause the nucleus pulposus to prolapse, pushing the frayed and weakened annulus posteriorly. With severe disk disease, the nucleus may protrude through the annulus (herniation) or become extruded to lie as a free fragment in the spinal canal.

Figure 15-4

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

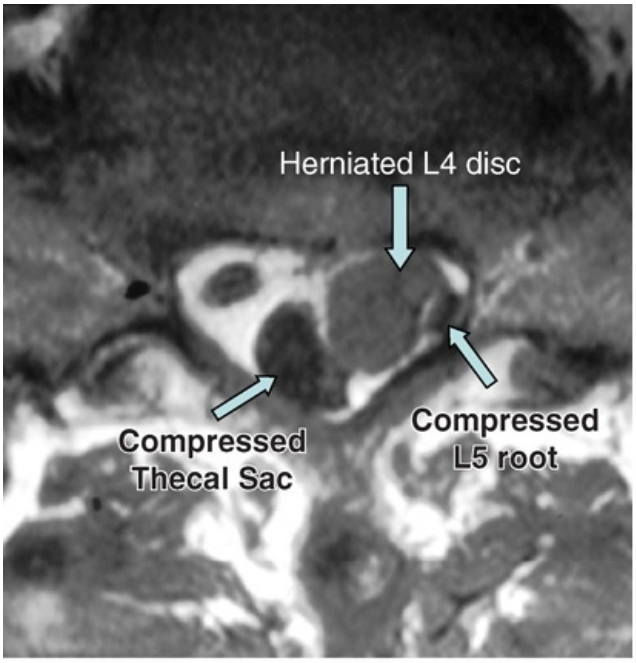

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left L5 radiculopathy. *A.* Sagittal T2-weighted image on the left reveals disk herniation at the L4-5 level. ***B.*** Axial T1-weighted image shows paracentral disk herniation with displacement of the thecal sac medially and the left L5 nerve root posteriorly in the left lateral recess

The mechanism by which intervertebral disk injury causes back pain is controversial. The inner annulus fibrosus and nucleus pulposus are normally devoid of innervation. Inflammation and production of proinflammatory cytokines within the protruding or ruptured disk may trigger or perpetuate back pain. Ingrowth of nociceptive (pain) nerve fibers into inner portions of a diseased disk may be responsible for chronic "diskogenic" pain. Nerve root injury (radiculopathy) from disk herniation may be due to compression, inflammation, or both; pathologically, demyelination and axonal loss are usually present.

A ruptured disk may be asymptomatic or cause back pain, abnormal posture, limitation of spine motion (particularly flexion), a focal neurologic deficit or radicular pain. A dermatomal pattern of sensory loss or a reduced or absent deep tendon reflex is more suggestive of a specific root lesion than is the pattern of pain. Motor findings (focal weakness, muscle atrophy, or fasciculations) occur less frequently than focal sensory or reflex changes. Symptoms and signs are usually unilateral, but bilateral involvement does occur with large central disk herniations that compress multiple roots or cause inflammation of nerve roots within the spinal canal. Clinical manifestations of specific nerve root lesions are summarized in Table 15-2. There is suggestive evidence that lumbar disk herniation with a nonprogressive nerve root deficit can be managed nonsurgically.

The differential diagnosis covers a variety of serious and treatable conditions, including epidural abscess, hematoma, fracture, or tumor. Fever, constant pain uninfluenced by position, sphincter abnormalities, or signs of spinal cord disease suggest an etiology other than lumbar disk disease. Absence of ankle reflexes can be a normal finding in persons older than age 60 years or a sign of bilateral S1 radiculopathy. An absent deep tendon reflex or focal sensory loss may indicate injury to a nerve root, but other sites of injury along the nerve must also be considered. For example, an absent knee reflex may be due to a femoral neuropathy or an L4 nerve root injury. A loss of sensation over the foot and lateral lower calf may result from a peroneal or lateral sciatic neuropathy or an L5 nerve root injury. Focal muscle atrophy may reflect a nerve root, peripheral nerve, anterior horn cell disease, or disuse.

A lumbar spine MRI scan or CT-myelogram is necessary to establish the location and type of pathology. Spine MRIs yield exquisite views of intraspinal and adjacent soft tissue anatomy. Bony lesions of the lateral recess or intervertebral foramen are optimally visualized by CT-myelography. The correlation of neuroradiologic findings to symptoms, particularly pain, is not simple. Contrast-enhancing tears in the annulus fibrosus or disk protrusions are widely accepted as common sources of back pain; however, studies have found that many asymptomatic adults have similar findings. Asymptomatic disk protrusions are also common and may enhance with contrast. *Furthermore, in patients with known disk herniation treated either medically or surgically, persistence of the herniation 10 years later had no relationship to the clinical outcome.* In summary, MRI findings of disk

protrusion, tears in the annulus fibrosus, or contrast enhancement are common incidental findings that, by themselves, should not dictate management decisions for patients with back pain.

The diagnosis of nerve root injury is most secure when the history, examination, results of imaging studies, and the EMG are concordant. The correlation between CT and EMG for localization of nerve root injury is between 65 and 73%. Up to one-third of asymptomatic adults have a lumbar disk protrusion detected by CT or MRI scans.

Management of lumbar disc disease is discussed below.

Cauda equina syndrome (CES) signifies an injury of multiple lumbosacral nerve roots within the spinal canal distal to the termination of the spinal cord at L1-2. Low back pain, weakness and areflexia in the legs, saddle anesthesia, or loss of bladder function may occur. The problem must be distinguishedfrom disorders of the lower spinal cord (conus medullaris syndrome), acute transverse myelitis (Chap. 377), and Guillain-Barré syndrome (Chap. 385). Combined involvement of the conus medullaris and cauda equina can occur. CES is commonly due to a ruptured lumbosacral intervertebral disk, lumbosacral spine fracture, hematoma within the spinal canal (e.g., following lumbar puncture in patients with coagulopathy), compressive tumor, or other mass lesion. Treatment options include surgical decompression, sometimes urgently, in an attempt to restore or preserve motor or sphincter function, or radiotherapy for metastatic tumors (Chap. 379).

Degenerative Conditions

Lumbar spinal stenosis (LSS) describes a narrowed lumbar spinal canal and is frequently asymptomatic. *Neurogenic claudication* is the usual symptom, consisting of back and buttock or leg pain induced by walking or standing and relieved by sitting. Symptoms in the legs are usually bilateral. Lumbar stenosis, by itself, is frequently asymptomatic, and the correlation between the severity of symptoms and degree of stenosis of the spinal canal is poor. Unlike vascular claudication, symptoms are often provoked by standing without walking. Unlike lumbar disk disease, symptoms are usually relieved by sitting. Patients with neurogenic claudication can often walk much farther when leaning over a shopping cart and can pedal a stationary bike while sitting with ease. These flexed positions increase the anteroposterior spinal canal diameter and reduce intraspinal venous hypertension, resulting in pain relief. Focal weakness, sensory loss, or reflex changes may occur when spinal stenosis is associated with neural foraminal narrowing and radiculopathy. Severe neurologic deficits, including paralysis and urinary incontinence, occur only rarely.

LSS can be acquired (75%), congenital, or due to a combination of these factors. Congenital forms (achondroplasia, idiopathic) are characterized by short, thick pedicles that produce both spinal canal and lateral recess stenosis. Acquired factors that contribute to spinal stenosis include degenerative diseases (spondylosis, spondylolisthesis, scoliosis), trauma, spine surgery, metabolic or endocrine disorders (epidural lipomatosis, osteoporosis, acromegaly, renal osteodystrophy, hypoparathyroidism), and Paget's disease. MRI provides the best definition of the abnormal anatomy **(Fig. 15-5)**.

Figure 15-5

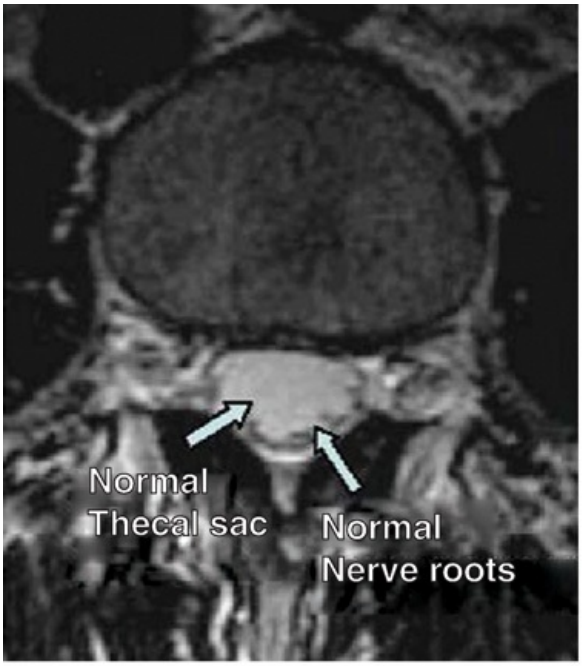

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

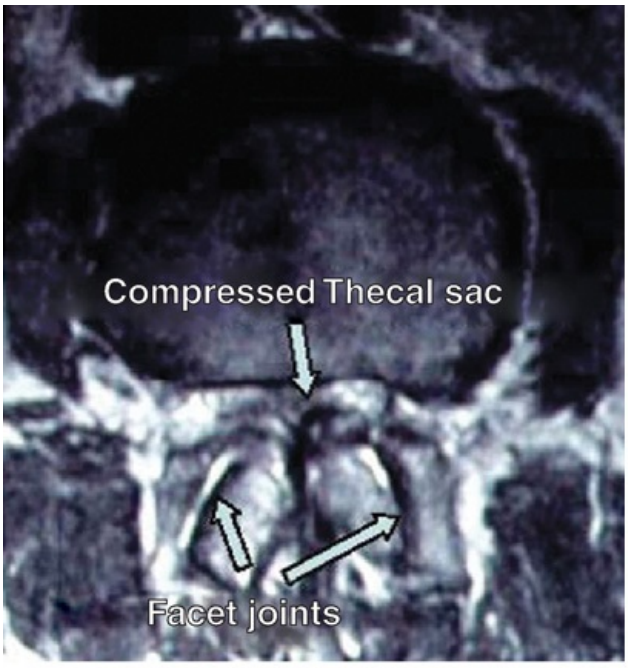

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial T2-weighted images of the lumbar spine. *A*. The image shows a normal thecal sac within the lumbar spinal canal. The thecal sac is bright. The lumbar roots are dark punctuate dots in the posterior thecal sac with the patient supine. ***B.*** The thecal sac is not well visualized due to severe lumbar spinal canal stenosis, partially the result of hypertrophic facet joints.

Conservative treatment of symptomatic LSS includes nonsteroidal anti-inflammatory drugs (NSAIDs), acetaminophen, exercise programs, and symptomatic treatment of acute pain episodes. There is insufficient evidence to support the use of epidural glucocorticoid injections. Surgical therapy is considered when medical therapy does not relieve symptoms sufficiently to allow for activities of daily living or when significant focal neurologic signs are present. Most patients with neurogenic claudication treated surgically experience significant relief of back and leg pain within 6 weeks post-operation, and pain relief persists for at least 2 years. Patients treated nonoperatively improve uncommonly. Up to one-quarter develop recurrent stenosis at the same spinal level or an adjacent

level 7–10 years after the initial surgery; recurrent symptoms usually respond to a second surgical decompression.

Neural foraminal narrowing with radiculopathy is a common degenerative disorder most often caused by the same processes that cause lumbar spinal stenosis (Figs. 15-1 and 15-6), including osteophytes, lateral disc protrusion, calcified disk-osteophytes, facet joint hypertrophy, uncovertebral joint hypertrophy (cervical spine), congenitally shortened pedicles, or, frequently, a combination of these processes. Neoplasms (primary or metastatic), fractures, infections (epidural abscess), or hematomas are other considerations. These conditions can produce unilateralnerve root symptoms or signs due to bony compression at the intervertebral foramen or lateral recess; symptoms are indistinguishable from disk-related radiculopathy, but treatment may differ depending upon the specific etiology. The history and neurologic examination alone cannot distinguish between these possibilities, and a spinal neuroimaging (CT or MRI) procedure is required to identify the underlying cause. Neurologic findings from the examination and EMG can help direct the attention of the radiologist to specific nerve or root structures that are best visualized on axial images. For *facet joint hypertrophy,* surgical foraminotomy produces long-term relief of leg and back pain in 80–90% of patients. The usefulness of therapeutic facet joint blocks for pain has not been rigorously studied.

Figure 15-6

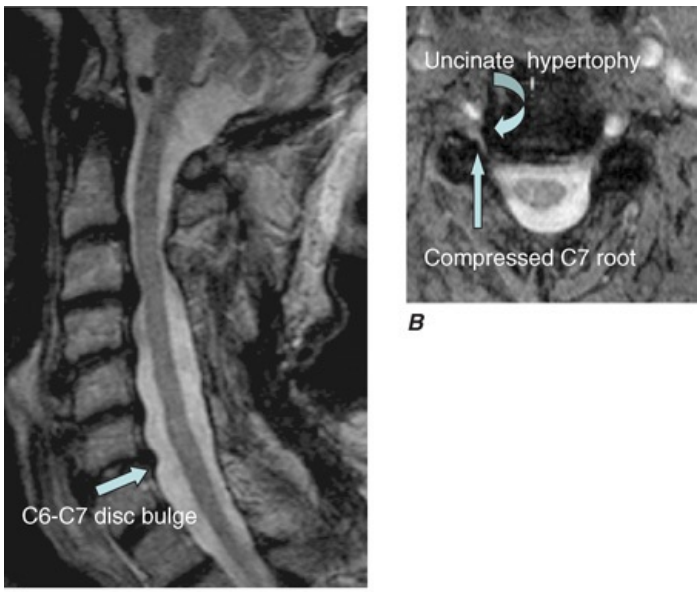

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Right C7 radiculopathy. *A.* Sagittal T2-weighted image shows mild disk bulging at C6-C7 and a mildly narrowed spinal canal, but no visible nerve root compression. ***B.*** Axial T2-weighted image. The combination of uncinate hypertrophy and facet hypertrophy (ovoid dark space just lateral to the C7 root) narrows the right C6-C7 intervertebral foramen resulting in right C7 nerve root compression.

Arthritis

Spondylosis, or osteoarthritic spine disease, typically occurs in later life and primarily involves the cervical and lumbosacral spine. Patients often complain of back pain that increases with movement and is associated with stiffness. The relationship between clinical symptoms and radiologic findings is usually not straightforward. Pain may be prominent when x-ray, CT, or MRI findings are minimal, and prominent degenerative spine disease can be seen in asymptomatic patients. Osteophytes or combined disk-osteophytes may cause or contribute to central spinal canal stenosis, lateral recess stenosis, or neural foraminal narrowing.

ANKYLOSING SPONDYLITIS

(See also Chap. 325) This distinctive arthritic spine disease typically presents with the insidious onset of low back and buttock pain. Patients are often males below age 40. Associated features include morning back stiffness, nocturnal pain, pain unrelieved by rest, an elevated ESR, and the histocompatibility antigen HLA-B27. Onset at a young age and back pain improving with exercise are characteristic. Loss of the normal lumbar lordosis and exaggeration of thoracic kyphosis develop as the disease progresses. Inflammation and erosion of the outer fibers of the annulus fibrosus at the point of contact with the vertebral body are followed by ossification and bony growth that bridges adjacent vertebral bodies and reduces spine mobility in all planes. MRI has been used to assess the

presence of inflammation in joints as well as response to treatment and is more sensitive than plain x-rays. In later stages, plain x-rays reveal bridging of vertebral bodies to produce the fused "bamboo spine."

Stress fractures after minimal or no trauma can occur through the spontaneously ankylosed posterior bony elements of the rigid, osteoporotic spine and can produce focal pain, spinal instability, spinal cord compression, or CES. Atlantoaxial subluxation with spinal cord compression can occur in up to 20% of patients over time. Ankylosis of the ribs to the spine and a decrease in the height of the thoracic spine may compromise respiratory function. Therapy with anti–tumor necrosis factor agents is effective in reducing disease activity and improving function. Similar to ankylosing spondylitis, restricted movements may accompany reactive arthritis (formerly known as Reiter's syndrome), psoriatic arthritis, and chronic inflammatory bowel disease.

Neoplasms

Back pain is the most common neurologic symptom in patients with systemic cancer and is the presenting symptom in 20%. The cause is usually vertebral body metastasis but can also result from spread of cancer through the intervertebral foramen (especially with lymphoma) or from carcinomatous meningitis. Cancer-related back pain tends to be constant, dull, unrelieved by rest, and worse at night. By contrast, mechanical low back pain usually improves with rest. MRI, CT, and CT-myelography are the studies of choice when spinal metastasis is suspected. Once a metastasis is found, imaging of the entire spine reveals additional tumor deposits in 1/3 of patients. MRI is preferred for soft tissue definition, but the most rapidly available imaging modality is best because the patient's condition may worsen quickly without intervention. Fewer than 5% of patients who are nonambulatory at the time of diagnosis ever regain the ability to walk; thus, early diagnosis is crucial. **The management of spinal metastasis is discussed in detail in Chap. 379.**

Infections/Inflammation

Vertebral osteomyelitis is often caused by staphylococci, but other bacteria or tuberculosis (Pott's disease) may be responsible. The primary source of infection is usually the urinary tract, skin, or lungs. Intravenous drug use is a well-recognized risk factor. Whenever pyogenic osteomyelitis is found, the possibility of bacterial endocarditis should be considered. Back pain unrelieved by rest, spine tenderness over the involved spine segment, and an elevated ESR are the most common findings in vertebral osteomyelitis. Fever or an elevated white blood cell count is found in a minority of patients. MRI and CT are sensitive and specific for early detection of osteomyelitis; CT may be more readily available in emergency settings and better tolerated by some patients with severe back pain. The intervertebral disk can also be affected by infection (discitis), and very rarely by tumor.

Spinal epidural abscess (Chap. 377) presents with back pain (aggravated by movement or palpation), fever, radiculopathy, or signs of spinal cord compression. The subacute development of two or more of these findings should increase the index of suspicion for spinal epidural abscess. The abscess may track over multiple spinal levels and is best delineated by spine MRI.

Lumbar adhesive arachnoiditis with radiculopathy is due to fibrosis following inflammation within the subarachnoid space. The fibrosis results in nerve root adhesions, and presents as back and leg pain associated with motor, sensory, or reflex changes. Causes of arachnoiditis include multiple lumbar operations, chronic spinal infections (especially tuberculosis in the developing world), spinal cord injury, intrathecal hemorrhage, myelography (rare), intrathecal injections (glucocorticoids, anesthetics, or other agents), and foreign bodies. The MRI shows clumped nerve roots or loculations of cerebrospinal fluid within the thecal sac. Clumped nerve roots may also occur with demyelinating polyneuropathy or neoplastic infiltration. Treatment is usually unsatisfactory. Microsurgical lysis of adhesions, dorsal rhizotomy, dorsal root ganglionectomy, and epidural steroids have been tried, but outcomes have been poor. Dorsal column stimulation for pain relief has produced varying results.

Metabolic Causes

OSTEOPOROSIS AND OSTEOSCLEROSIS

Immobilization or underlying conditions such as osteomalacia, the postmenopausal state, renal disease, multiple myeloma, hyperparathyroidism, hyperthyroidism, metastatic carcinoma, or glucocorticoid use may accelerate osteoporosis and weaken the vertebral body, leading to compression fractures and pain. Up to two-thirds of compression fractures seen on radiologic imaging are asymptomatic. The most common nontraumatic vertebral body fractures are due to postmenopausal or senile osteoporosis (Chap. 354). The risk of an additional vertebral fracture at 1 year following a first vertebral fracture is 20%. The presence of fever, weight loss, fracture at a level above T4, or other conditions described above should increase the suspicion for a cause other than senile osteoporosis. If tumor is suspected, a bone biopsy or diagnostic search for a primary tumor is indicated. The sole manifestation of a compression fracture may be localized back pain or radicular pain exacerbated by movement and often reproduced by palpation over the spinous process of the affected vertebra. The clinical context, neurologic signs, and radiologic appearance of the spine establish the diagnosis.

Relief of acute pain can often be achieved with acetaminophen or a combination of opioids and acetaminophen. The role of NSAIDs is controversial. Both pain and disability are improved with bracing. Antiresorptive drugs, especially bisphosphonates (e.g., alendronate), have been shown to reduce the risk of osteoporotic fractures and are the preferred treatment to prevent additional fractures. Less than one-third of patients with prior compression fractures are adequately treated for osteoporosis despite the increased risk for future fractures; even fewer at-risk patients without a history of fracture are adequately treated. Interventions [percutaneous vertebroplasty (PVP), kyphoplasty] exist for osteoporotic compression fractures associated with debilitating pain. Controlled studies suggest a benefit for pain reduction acutely, but not at 2 months, when compared with conservative care. Relief of pain following PVP has also been reported in patients with vertebral metastases, myeloma, or hemangiomas.

Osteosclerosis, an abnormally increased bone density often due to Paget's disease, is readily identifiable on routine x-ray studies and can sometimes be a source of back pain. It may be associated with an isolated increase in alkaline phosphatase in an otherwise healthy older person. Spinal cord or nerve root compression can result from bony encroachment. It should not be assumed that Paget's disease is the cause of a patient's back pain until other etiologies have been carefully considered.

For further discussion of these bone disorders, see Chaps. 353, 354, and 355.

Referred Pain from Visceral Disease

Diseases of the thorax, abdomen, or pelvis may refer pain to the posterior portion of the spinal segment that innervates the diseased organ. Occasionally, back pain may be the first and only manifestation. Upper abdominal diseases generally refer pain to the lower thoracic or upper lumbar region (eighth thoracic to the first and second lumbar vertebrae), lower abdominal diseases to the midlumbar region (second to fourth lumbar vertebrae), and pelvic diseases to the sacral region. Local signs (pain with spine palpation, paraspinal muscle spasm) are absent, and little or no pain accompanies routine movements of the spine.

LOW THORACIC OR LUMBAR PAIN WITH ABDOMINAL DISEASE

Tumors of the posterior wall of the stomach or duodenum typically produce epigastric pain (Chaps. 91 and 293), but midline back or paraspinal pain may occur if retroperitoneal extension is present. Fatty foods occaionally induce back pain associated with biliary disease. Diseases of the pancreas can produce right paraspinal back pain (head of the pancreas involved) or left paraspinal pain (body or tail involved). Pathology in retroperitoneal structures (hemorrhage, tumors, pyelonephritis) can produce paraspinal pain that radiates to the lower abdomen, groin, or anterior thighs. A mass in the iliopsoas region can produce unilateral lumbar pain with radiation toward the groin, labia, or testicle. The sudden appearance of lumbar pain in a patient receiving anticoagulants suggests retroperitoneal hemorrhage.

Isolated low back pain occurs in some patients with a contained rupture of an abdominal aortic aneurysm (AAA). The classic clinical triad of abdominal pain, shock, and back pain occurs in <20% of patients. The typical patient at risk is an elderly male smoker with back pain. Frequently, the diagnosis is initially missed because the symptoms and signs can be nonspecific. Misdiagnoses include nonspecific back pain, diverticulitis, renal colic, sepsis, and myocardial infarction. A careful abdominal examination revealing a pulsatile mass (present in 50–75% of patients) is an important physical finding. Patients with suspected AAA should be evaluated with abdominal ultrasound, CT, or MRI (Chap. 248).

SACRAL PAIN WITH GYNECOLOGIC AND UROLOGIC DISEASE

Pelvic organs rarely cause low back pain, except for gynecologic disorders involving the uterosacral ligaments. The pain is referred to the sacral region. Endometriosis or uterine cancers may invade the uterosacral ligaments. Pain associated with endometriosis is typically premenstrual and often continues until it merges with menstrual pain. Uterine malposition may cause uterosacral ligament traction (retroversion, descensus, and prolapse) or produce sacral pain after prolonged standing.

Menstrual pain may be felt in the sacral region. Poorly localized, cramping pain can radiate down the legs. Pain due to neoplastic infiltration of nerves is typically continuous, progressive in severity, and unrelieved by rest at night. Less commonly, radiation therapy of pelvic tumors may produce sacral pain from late radiation necrosis of tissue. Low back pain that radiates into one or both thighs is common in the last weeks of pregnancy.

Urologic sources of lumbosacral back pain include chronic prostatitis, prostate cancer with spinal metastasis (Chap. 95), and diseases of the kidney or ureter. Lesions of the bladder and testes do not often produce back pain. Infectious, inflammatory, or neoplastic renal diseases may produce ipsilateral lumbosacral pain, as can renal artery or vein thrombosis. Paraspinal lumbar pain may be a symptom of ureteral obstruction due to nephrolithiasis.

Other Causes of Back Pain

POSTURAL BACK PAIN

There is a group of patients with nonspecific chronic low back pain (CLBP) in whom no specific

anatomic lesion can be found despite exhaustive investigation. These individuals complain of vague, diffuse back pain with prolonged sitting or standing that is relieved by rest. Exercises to strengthen the paraspinal and abdominal muscles are sometimes helpful.

PSYCHIATRIC DISEASE

CLBP may be encountered in patients who seek financial compensation; in malingerers; or in those with concurrent substance abuse. Many patients with CLBP have a history of psychiatric illness (depression, anxiety states), or childhood trauma (physical or sexual abuse) that antedates the onset of back pain. Preoperative psychological assessment has been used to exclude patients with marked psychological impairments that predict a poor surgical outcome from spine surgery.

Idiopathic

The cause of low back pain occasionally remains unclear. Some patients have had multiple operations for disk disease but have persistent pain and disability. The original indications for surgery may have been questionable, with back pain only, no definite neurologic signs, or a minor disk bulge noted on CT or MRI. Scoring systems based upon neurologic signs, psychological factors, physiologic studies, and imaging studies have been devised to minimize the likelihood of unsuccessful surgery.

Treatment: Back Pain

ACUTE LOW BACK PAIN (ALBP) WITHOUT RADICULOPATHY

ALBP is defined as pain of <3 months' duration. Full recovery can be expected in 85% of adults with ALBP without leg pain. Most have purely "mechanical" symptoms (i.e., pain that is aggravated by motion and relieved by rest).

The initial assessment excludes serious causes of spine pathology that require urgent intervention, including infection, cancer, or trauma. Risk factors for a serious cause of ALBP are shown in Table 15-1. Laboratory and imaging studies are unnecessary if risk factors are absent. CT or plain spine films are rarely indicated in the first month of symptoms unless a spine fracture is suspected.

The prognosis is generally excellent. Many patients do not seek medical care and apparently improve on their own. Even among those seen in primary care, two-thirds report being substantially improved after seven weeks. This spontaneous improvement can mislead clinicians and researchers about the efficacy of treatment interventions. Perhaps as a result, many ineffective treatments have become widespread in the past, such as bed rest, lumbar traction, sacroiliac fusion, and coccygectomy.

Clinicians should reassure patients that improvement is very likely, and instruct them in self-care. Education is an important part of treatment. Satisfaction and the likelihood of follow-up increase when patients are educated about prognosis, treatment methods, activity modifications, and strategies to prevent future exacerbations. In one study, patients who felt they did not receive an adequate explanation for their symptoms wanted further diagnostic tests. In general, bed rest should be avoided, or kept to a day or two at most, for relief of severe symptoms. Several randomized trials suggest that bed rest does not accelerate the pace of recovery. In general, the best activity recommendation is for walking and early resumption of normal physical activity, avoiding only strenuous manual labor. Possible advantages of early ambulation for acute back pain include maintenance of cardiovascular conditioning, improved disk and cartilage nutrition, improved bone and muscle strength, and increased endorphin levels. Specific back exercises or early vigorous exercise have not shown benefits for acute back pain, but may be useful for chronic pain. Application of heat by heating pads or heated blankets is sometimes helpful.

Evidence-based guidelines suggest that over-the-counter medicines such as acetaminophen and NSAIDs are first-line options for the treatment of ALBP. Skeletal muscle relaxants, such as cyclobenzaprine or methocarbamol, may be useful, but sedation is a common side effect. Limiting the use of muscle relaxants to nighttime only may be an option for some patients. Because of the risk of abuse of some drugs in this category, including benzodiazepines and carisoprodol, short courses are generally recommended.

It is unclear whether opioid analgesics and tramadol are more effective than NSAIDs or acetaminophen for treating ALBP; most of the available efficacy data are for treatment of chronic back pain. Their use is best reserved for patients who cannot tolerate acetaminophen or NSAIDs, or for those with severe refractory pain. As with muscle relaxants, these drugs are often sedating, so it may be useful to prescribe them at nighttime only, Side effects of short-term opioid use include nausea, constipation and pruritis; risks of long-term opioid use include hypersensitivity to pain, hypogonadism, and dependency.

There is no evidence to support use of oral or injected glucocorticoids for acute low back pain without radiculopathy. Antiepileptic drugs, such as gabapentin, are not FDA approved for treating low back pain, and there is insufficient evidence to support their use in this setting.

Nonpharmacologic treatments for acute low back pain include spinal manipulation, physical therapy, massage, acupuncture, transcutaneous electrical nerve stimulation, ultrasound, diathermy, and magnets. Spinal manipulation appears to be roughly equivalent to conventional medical treatments and

may be a useful alternative for patients who wish to avoid or who cannot tolerate drug therapy. There is little evidence to support the use of physical therapy, massage, acupuncture, laser therapy, therapeutic ultrasound, magnets, corsets, or lumbar traction. Though important for chronic pain, back exercises for acute back pain are generally not supported by clinical evidence. There is no useful evidence regarding the value of ice or heat applications for ABLP; many patients report temporary symptomatic relief from ice, and heat may produce a short-term reduction in pain after the first week.

CHRONIC LOW BACK PAIN WITHOUT RADICULOPATHY

Chronic low back pain is defined as pain lasting >12 weeks; it accounts for 50% of total back pain costs. Risk factors include obesity, female gender, older age, prior history of back pain, restricted spinal mobility, pain radiating into a leg, high levels of psychological distress, poor self-rated health, minimal physical activity, smoking, job dissatisfaction, and widespread pain. In general, the same treatments that are recommended for acute low back pain can be useful for patients with chronic low back pain. In this setting, however, the benefit of opioid therapy or muscle relaxants is less clear.

Evidence supports the use of exercise therapy, and this can be one of the mainstays of treatment for chronic back pain. Effective regimens have generally included a combination of gradually increasing aerobic exercise, strengthening exercises, and stretching exercises. Motivating patients is sometimes challenging, and supervised exercise is best, for example, with a supportive physical therapist. In general, activity tolerance is the primary goal, while pain relief is secondary. Exercise programs can reverse atrophy in paraspinal muscles and strengthen extensors of the trunk. Supervised intensive physical exercise or "work hardening" regimens have been effective in returning some patients to work, improving walking distance, and reducing pain. In addition, some forms of yoga have been evaluated in randomized trials and may be helpful for patients who are interested.

Medications for chronic low back pain may include acetaminophen, NSAIDs, and tricyclic antidepressants. Trials of the latter suggest some benefit even for patients without evidence of depression. Trials do not support the efficacy of selective serotonin reuptake inhibitors for back pain. However, depression is common among patients with chronic pain and should be appropriately treated.

Cognitive-behavioral therapy is based on evidence that psychological and social factors, as well as somatic pathology, are important in the genesis of chronic pain and disability. The patient's attitudes and beliefs, psychological distress, and patterns of illness behavior may all influence responses to chronic pain. Thus, in addition to addressing pathophysiologic mechanisms, psychological treatments are aimed at reducing disability by modifying cognitive processes and environmental contingencies. Cognitive-behavioral therapy includes efforts to identify and modify patients' thinking about their pain and disability by strategies that may involve imagery, attention diversion, or modifying maladaptive thoughts, feelings, and beliefs. This approach includes educating patients about a multidimensional view of pain, identifying pain-eliciting or pain-aggravating thoughts and feelings, using coping strategies and relaxation techniques, and even hypnosis. A systematic review concluded that such treatments are more effective than a waiting list control group for short-term pain relief; however, long-term results remain unclear. Behavioral treatments may have effects similar in magnitude to exercise therapy.

Back pain is the most common reason for seeking complementary and alternative treatments. The most common of these for back pain are spinal manipulation, acupuncture, and massage. The role of complementary and alternative medicine approaches, aside from spinal manipulation, remains unclear. Biofeedback has not been studied rigorously. As with acute back pain, spinal manipulation may on average offer benefits similar to conventional care. Rigorous recent trials of acupuncture suggest that true acupuncture is not superior to sham acupuncture, but that both may offer an advantage over routine care. Whether this is due entirely to placebo effects or to stimulation provided even by sham acupuncture is uncertain. Some trials of massage therapy have been encouraging, but this has been less well studied than manipulation or acupuncture.

Studies of transcutaneous electrical nerve stimulation (TENS) have reached conflicting conclusions, but a recent evidence-based guideline suggested that there was no convincing evidence for its efficacy in treating chronic back pain.

Various injections, including epidural glucocorticoid injections, facet joint injections, and trigger point injections have been used for treating chronic low back pain. However, in the absence of radiculopathy, there is no evidence that epidural glucocorticoids are effective for treating chronic back pain. Several randomized trials suggest that facet joint injections are not more effective than saline injections, and recent evidence-based guidelines recommend against their use. Similarly, there is little evidence to support the use of trigger point injections. Injection studies are sometimes used diagnostically to help determine the anatomic source of back pain. Reproduction of the patient's typical pain with discography has been used as evidence that a specific disk is the pain generator. Pain relief following a foraminal nerve root block or glucocorticoid injection into a facet has been similarly used as evidence that the facet joint or nerve root is the source. However, the possibility that the injection response was a placebo effect or due to systemic absorption of the glucocorticoids is often not excluded.

Another category of intervention for chronic back pain includes electrothermal and radiofrequency therapies. Intradiscal therapy has been proposed using both types of energy to thermocoagulate and destroy nerves in the intervertebral disc, using specially designed catheters or electrodes. A systematic review has suggested that current evidence does not support the use of these intradiscal therapies.

Radiofrequency denervation is sometimes used to destroy nerves that are thought to mediate pain, and this technique has been used for facet joint pain (with the target nerve being the medial branch of the primary dorsal ramus), for back pain thought to arise from the intervertebral disc (ramus communicans), and radicular back pain (dorsal root ganglia). A few small trials have resulted in conflicting results for facet joint pain. The evidence for presumed discogenic pain and for radicular pain is similarly meager. A trial for patients with chronic radicular pain found no difference between radiofrequency denervation of the dorsal root ganglia and sham treatment. Recent systematic reviews have thus concluded that there is insufficient evidence to reliably evaluate these interventional therapies.

Surgical intervention for chronic low back pain in the absence of radiculopathy has been evaluated in a small number of randomized trials, all conducted in Europe. Each of these studies included patients with back pain and a degenerative disc, but no sciatica. Three of the four trials concluded that lumbar fusion surgery was no more effective than highly structured, rigorous rehabilitation combined with cognitive-behavioral therapy. The fourth trial found an advantage of fusion surgery over haphazard "usual care," which appeared to be less effective than the structured rehabilitation in other trials. Given conflicting evidence, indications for surgery for chronic back pain alone have remained controversial. Both U.S. and British guidelines suggest considering referral for an opinion on spinal fusion for people who have completed an optimal nonsurgical treatment program (including combined physical and psychological treatment) and who have persistent severe back pain for which they would consider surgery.

The newest surgical treatment for degenerated discs with back pain is disc replacement with prosthetic disks. These are generally designed as metal plates with a polyethylene cushion sandwiched in between. The trials that led to approval of these devices compared them to spine fusion, and concluded that the artificial discs were "not inferior." Serious complications appeared to be somewhat more likely with the artificial disc. This treatment remains controversial for low back pain.

Intensive multidisciplinary rehabilitation programs may involve daily or frequent care involving physical therapy, exercise, cognitive-behavioral therapy, a workplace evaluation, and other interventions. For patients who have not responded to other interventions, such programs appear to offer some benefit. Systematic reviews suggest that the evidence is limited and effects are moderate.

Some observers have raised concern that chronic back pain may often be overtreated. The use of opioids, epidural glucocorticoid injections, facet joint injections, and surgical intervention has increased rapidly in the past decade, without corresponding population-level improvements in pain or functioning among patients with back pain. In each case, randomized trials provide only minimal support for these treatments in the setting of chronic back pain without radiculopathy. For low back pain without radiculopathy, new British guidelines explicitly recommend against use of selective serotonin reuptake inhibitors (SSRIs), any type of injection, TENS, lumbar supports, traction, radiofrequency facet joint denervation, intradiscal electrothermal therapy, or intradiscal radio frequency thermocoagulation. Similarly, these treatments are not recommended in guidelines from the American College of Physicians and the American Pain Society. On the other hand, exercise therapy and treatment of depression appear to be underused.

LOW BACK PAIN WITH RADICULOPATHY

A common cause of back pain with radiculopathy is a herniated disc with nerve root impingement, resulting in back pain with radiation down the leg. The prognosis for acute low back pain with radiculopathy due to disk herniation (sciatica) is generally favorable, with most patients demonstrating substantial improvement over a matter of months. Serial imaging studies suggest spontaneous regression of the herniated portion of the disc in two-thirds of patients over 6 months. Nonetheless, there are several important treatment options for providing symptom relief while this natural healing process unfolds.

Resumption of normal activity as much as possible is usually the best activity recommendation. Randomized trial evidence suggests that bed rest is ineffective for treating sciatica as well as for back pain alone. Acetaminophen and NSAIDs are appropriate for pain relief, although severe pain may require short courses of opioid analgesics.

Epidural glucocorticoid injections have a role in providing temporary symptom relief for sciatica due to a herniated disc. Although randomized trial evidence is conflicting, there appears to be some overall short-term benefit for pain relief of sciatica. However, there does not appear to be a benefit in terms of reducing subsequent surgical interventions. Diagnostic nerve root blocks have been advocated to determine if pain originates from a specific nerve root. However, improvement may result even when the nerve root is not responsible for the pain; this may occur as a placebo effect, from a pain-generating lesion located distally along the peripheral nerve, or from anesthesia of the sinuvertebral

nerve. The utility of diagnostic nerve root blocks remains a subject of debate.

Surgical intervention is indicated for patients who have progressive motor weakness, demonstrated on clinical examination or EMG, as a result of nerve root injury. Urgent surgery is recommended for patients who have evidence of the cauda equina syndrome or spinal cord compression, generally suggested by bowel or bladder dysfunction, diminished sensation in a saddle distribution, a sensory level, bilateral leg weakness, or bilateral leg spasticity.

Surgery is also an important option for patients who have disabling radicular pain despite optimal conservative treatment. Sciatica is perhaps the most common reason for recommending spine surgery. Because patients with a herniated disc and sciatica generally experience rapid improvement over a matter of weeks, most experts do not recommend considering surgery unless the patient has failed to respond to 6–8 weeks of appropriate nonsurgical management. For patients who have not improved, randomized trials indicate that, compared to nonsurgical treatment, surgery results in more rapid pain relief. However, after the first year or two of follow-up, patients with sciatica appear to have much the same level of pain relief and functional improvement with or without surgery. Thus, both treatment approaches are reasonable, and patient preferences should play a major role in decision making. Some patients will want the fastest possible relief and find surgical risks acceptable. Others will be more risk-averse and more tolerant of symptoms, and will choose watchful waiting if they understand that improvement is likely in the end.

The usual surgical procedure is a partial hemilaminectomy with excision of the prolapsed disk. Fusion of the involved lumbar segments should be considered only if significant spinal instability is present (i.e., degenerative spondylolisthesis). The costs associated with lumbar interbody fusion have increased dramatically in recent years. There are no large prospective, randomized trials comparing fusion to other types of surgical intervention. In one study, patients with persistent low back pain despite an initial diskectomy fared no better with spine fusion than with a conservative regimen of cognitive intervention and exercise. Artificial disks have been in use in Europe for the past decade; their utility remains controversial in the United States.

PAIN IN THE NECK AND SHOULDER

(Table 15-4) Neck pain, which usually arises from diseases of the cervical spine and soft tissues of the neck, is common. Neck pain arising from the cervical spine is typically precipitated by movement and may be accompanied by focal tenderness and limitation of motion. Pain arising from the brachial plexus, shoulder, or peripheral nerves can be confused with cervical spine disease, but the history and examination usually identify a more distal origin for the pain. Cervical spine trauma, disk disease, or spondylosis with intervertebral foraminal narrowing may be asymptomatic or painful and can produce a myelopathy, radiculopathy, or both. The same risk factors for a serious cause of low back pain are thought to apply to neck pain with the addition that neurologic signs of myelopathy (incontinence, sensory level, spastic legs) may also occur. Lhermitte's sign, an electrical shock down the spine with neck flexion, suggests cervical spinal cord involvement from any cause.

Table 15-4 Cervical Radiculopathy—Neurologic Features

	Examination Findings			
Cervical Nerve Roots	**Reflex**	**Sensory**	**Motor**	**Pain Distribution**
C5	Biceps	Over lateral deltoid	Supraspinatus* (initial arm abduction) Infraspinatus* (arm external rotation) Deltoid* (arm abduction) Biceps (arm flexion)	Lateral arm, medial scapula
C6	Biceps	Thumb, index fingers Radial hand/forearm	Biceps (arm flexion) Pronator teres (internal forearm rotation)	Lateral forearm, thumb, index finger
C7	Triceps	Middle fingers Dorsum forearm	Triceps* (arm extension) Wrist extensors* Extensor digitorum* (finger extension)	Posterior arm, dorsal forearm, lateral hand

C8	Finger flexors	Little finger Medial hand and forearm	Abductor pollicis brevis (abduction D1) First dorsal interosseous (abduction D2) Abductor digiti minimi (abduction D5)	4th and 5th fingers, medial forearm
T1	Finger flexors	Axilla and medial arm	Abductor pollicis brevis (abduction D1) First dorsal interosseous (abduction D2) Abductor digiti minimi (abduction D5)	Medial arm, axilla

* These muscles receive the majority of innervation from this root.

Trauma to the Cervical Spine

Trauma to the cervical spine (fractures, subluxation) places the spinal cord at risk for compression. Motor vehicle accidents, violent crimes, or falls account for 87% of cervical spinal cord injuries (Chap. 377). Immediate immobilization of the neck is essential to minimize further spinal cord injury from movement of unstable cervical spine segments. The decision to obtain imaging should be based upon the nature of the injury. The NEXUS low-risk criteria established that normally alert patients without palpation tenderness in the midline; intoxication; neurologic deficits; and painful distracting injuries had a very low likelihood of a clinically significant traumatic injury to the cervical spine. The Canadian C-spine rule recommends that imaging should be obtained following neck region trauma if the patient is >65 years old, has limb paresthesias, or a dangerous mechanism for the injury (e.g., bicycle collision with tree or parked car, fall from height >3 feet or 5 stairs, diving accident). A CT scan is the diagnostic procedure of choice for detection of acute fractures. When traumatic injury to the vertebral arteries or cervical spinal cord is suspected, visualization by MRI with MR angiography is preferred.

Whiplash injury is due to rapid flexion and extension of the neck, usually in automobile accidents, and causes cervical musculoligamental injury. This diagnosis should not be applied to patients with fractures, disk herniation, head injury, focal neurologic findings, or altered consciousness. Up to 50% of persons reporting whiplash injury acutely have persistent neck pain 1 year later. Once personal compensation for pain and suffering was removed from the Australian health care system, the prognosis for recovery at 1 year from whiplash injury improved also. Imaging of the cervical spine is not cost-effective acutely but is useful to detect disk herniations when symptoms persist for >6 weeks following the injury. Severe initial symptoms have been associated with a poor long-term outcome.

Cervical Disk Disease

Herniation of a lower cervical disk is a common cause of neck, shoulder, arm, or hand pain or tingling. Neck pain, stiffness, and a range of motion limited by pain are the usual manifestations. A herniated cervical disk is responsible for ~25% of cervical radiculopathies. Extension and lateral rotation of the neck narrows the ipsilateral intervertebral foramen and may reproduce radicular symptoms (Spurling's sign). In young persons, acute nerve root compression from a ruptured cervical disk is often due to trauma. Cervical disk herniations are usually posterolateral near the lateral recess. The cervical nerve roots most commonly affected are C7 and C6. Typical patterns of reflex, sensory, and motor changes that accompany specific cervical nerve root lesions are summarized in Table 15-4. While the classic patterns are clinically helpful, there are numerous exceptions because (1) there is overlap in function between adjacent nerve roots, (2) symptoms and signs may be evident in only part of the injured nerve root territory, and (3) the location of pain is the most variable of the clinical features.

Cervical Spondylosis

Osteoarthritis of the cervical spine may produce neck pain that radiates into the back of the head, shoulders, or arms, or may be the source of headaches in the posterior occipital region (supplied by the C2-C4 nerve roots). Osteophytes, disk protrusions, or hypertrophic facet or uncovertebral joints may alone or in combination compress one or several nerve roots at the intervertebral foramina (Fig. 15-6); this compression accounts for 75% of cervical radiculopathies. The roots most commonly affected are C7 and C6. Narrowing of the spinal canal by osteophytes, ossification of the posterior longitudinal ligament (OPLL), or a large central disk may compress the cervical spinal cord. Combinations of radiculopathy and myelopathy may be present. When little or no neck pain accompanies cord compression, the diagnosis may be confused with amyotrophic lateral sclerosis (Chap. 374), multiple sclerosis (Chap. 380), spinal cord tumors, or syringomyelia (Chap. 377). The possibility of cervical spondylosis should be considered even when the patient presents with symptoms or signs in the legs only. MRI is the study of choice to define the anatomic abnormalities, but plain CT is adequate to assess bony spurs, foraminal narrowing, lateral recess stenosis, or OPLL.

EMG and nerve conduction studies can localize and assess the severity of the nerve root injury.

Other Causes of Neck Pain

Rheumatoid arthritis (RA) (Chap. 321) of the cervical apophyseal joints produces neck pain, stiffness, and limitation of motion. In advanced RA, synovitis of the atlantoaxial joint (C1-C2; Fig. 15-2) may damage the transverse ligament of the atlas, producing forward displacement of the atlas on the axis (atlantoaxial subluxation). Radiologic evidence of atlantoaxial subluxation occurs in 30% of patients with RA. Not surprisingly, the degree of subluxation correlates with the severity of erosive disease. When subluxation is present, careful assessment is important to identify early signs of myelopathy. Occasional patients develop high spinal cord compression leading to quadriparesis, respiratory insufficiency, and death. Surgery should be considered when myelopathy or spinal instability is present. MRI is the imaging modality of choice.

Ankylosing spondylitis can cause neck pain and less commonly atlantoaxial subluxation; surgery may be required to prevent spinal cord compression. Acute *herpes zoster* presents as acute posterior occipital or neck pain prior to the outbreak of vesicles. *Neoplasms* metastatic to the cervical spine, *infections* (osteomyelitis and epidural abscess), and *metabolic bone diseases* may be the cause of neck pain. Neck pain may also be referred from the heart with coronary artery ischemia (cervical angina syndrome).

Thoracic Outlet

The thoracic outlet contains the first rib, the subclavian artery and vein, the brachial plexus, the clavicle, and the lung apex. Injury to these structures may result in postural or movement-induced pain around the shoulder and supraclavicular region. *True neurogenicthoracic outlet syndrome* (TOS) is an uncommon disorder resulting from compression of the lower trunk of the brachial plexus or ventral rami of the C8 or T1 nerve roots most often by an anomalous band of tissue connecting an elongate transverse process at C7 with the first rib. Pain is mild or absent. Signs include weakness and wasting of intrinsic muscles of the hand and diminished sensation on the palmar aspect of the fifth digit. An anteroposterior cervical spine x-ray will show the elongate C7 transverse process, and EMG and nerve conduction studies confirm the diagnosis. Treatment consists of surgical resection of the anomalous band. The weakness and wasting of intrinsic hand muscles typically does not improve, but surgery halts the insidious progression of weakness. *Arterial TOS* results from compression of the subclavian artery by a cervical rib resulting in poststenotic dilatation of the artery and thrombus formation. Blood pressure is reduced in the affected limb, and signs of emboli may be present in the hand. Neurologic signs are absent. Ultrasound can confirm the diagnosis noninvasively. Treatment is with thrombolysis or anticoagulation (with or without embolectomy) and surgical excision of the cervical rib compressing the subclavian artery. *Venous TOS* is due to subclavian vein thrombosis resulting in swelling of the arm and pain. The vein may be compressed by a cervical rib or anomalous scalene muscle. Venography is the diagnostic test of choice. *Disputed TOS* includes a large number of patients with chronic arm and shoulder pain of unclear cause. The lack of sensitive and specific findings on physical examination or laboratory markers for this condition frequently results in diagnostic uncertainty. The role of surgery in disputed TOS is controversial. Multidisciplinary pain management is a conservative approach, although treatment is often unsuccessful.

Brachial Plexus and Nerves

Pain from injury to the brachial plexus or peripheral nerves of the arm can occasionally mimic pain of cervical spine origin. Neoplastic infiltration of the lower trunk of the brachial plexus may produce shoulder or supraclavicular pain radiating down the arm, numbness of the fourth and fifth fingers or medial forearm, and weakness of intrinsic hand muscles innervated by the ulnar and median nerves. Delayed radiation injury may produce similar findings, although pain is less often present and almost always less severe. A Pancoast tumor of the lung (Chap. 89) is another cause and should be considered, especially when a Horner's syndrome is present. *Suprascapular neuropathy* may produce severe shoulder pain, weakness, and wasting of the supraspinatus and infraspinatus muscles. *Acute brachial neuritis* is often confused with radiculopathy; the acute onset of severe shoulder or scapular pain is followed typically over days by weakness of the proximal arm and shoulder girdle muscles innervated by the upper brachial plexus. The onset is often preceded by an infection. The long thoracic nerve may be affected; the latter results in a winged scapula. Brachial neuritis may also present as an isolated paralysis of the diaphragm or with involvement of other nerves of the upper limb. Recovery is generally good but may take up to 3 years to be complete.

Occasional cases of carpal tunnel syndrome produce pain and paresthesias extending into the forearm, arm, and shoulder resembling a C5 or C6 root lesion. Lesions of the radial or ulnar nerve can mimic a radiculopathy at C7 or C8, respectively. EMG and nerve conduction studies can accurately localize lesions to the nerve roots, brachial plexus, or peripheral nerves.

For further discussion of peripheral nerve disorders, see Chap. 384.

Shoulder

Pain arising from the shoulder can on occasion mimic pain from the spine. If symptoms and signs of

radiculopathy are absent, then the differential diagnosis includes mechanical shoulder pain (tendonitis, bursitis, rotator cuff tear, dislocation, adhesive capsulitis, and cuff impingement under the acromion) and referred pain (subdiaphragmatic irritation, angina, Pancoast tumor). Mechanical pain is often worse at night, associated with local shoulder tenderness and aggravated by abduction, internal rotation, or extension of the arm. Pain from shoulder disease may radiate into the arm or hand, but sensory, motor, and reflex changes are absent.

Treatment: Neck Pain Without Radiculopathy

The evidence regarding treatment for neck pain is less complete than that for low back pain. As with low back pain, spontaneous improvement is the norm for acute neck pain, and the usual goal of therapy is to provide symptom relief while natural healing processes proceed.

The evidence in support of nonsurgical treatments for whiplash-associated disorders is generally of poor quality and neither supports nor refutes the effectiveness of common treatments used for symptom relief. Gentle mobilization of the cervical spine combined with exercise programs may be more beneficial than usual care. Evidence is insufficient to recommend for or against the use of cervical traction, neck collars, TENS, ultrasound, diathermy, or massage. The role of acupuncture for neck pain also remains ambiguous, with poor-quality studies and conflicting results.

For patients with neck pain unassociated with trauma, supervised exercise, with or without mobilization, appears to be effective. Exercises often include shoulder rolls and neck stretches. Although there is relatively little evidence about the use of muscle relaxants, analgesics, and NSAIDs in neck pain, many clinicians use these medications in much the same way as for low back pain.

Low-level laser therapy directed at areas of tenderness, local acupuncture points, or a grid of predetermined points is a controversial approach to the treatment of neck pain. The putative benefits might be mediated by anti-inflammatory effects, reduction of skeletal muscle fatigue, or inhibition of transmission at neuromuscular junctions. A 2009 meta-analysis suggested that this treatment may provide greater pain relief than sham therapy for both acute and chronic neck pain. Comparison to other conservative treatment measures is needed.

Although some surgical studies have proposed a role for anterior diskectomy and fusion in patients with neck pain, these studies generally have not been rigorously conducted. A systematic review suggested that there was no valid clinical evidence to support either cervical fusion or cervical disc arthroplasty in patients with neck pain without radiculopathy. Similarly, there is no evidence to support radiofrequency neurotomy or cervical facet injections for neck pain without radiculopathy.

Treatment: Neck Pain with Radiculopathy

The natural history of neck pain even with radiculopathy is favorable, and many patients will improve without specific therapy. Although there are no randomized trials of NSAIDs for neck pain, a course of NSAIDs, with or without muscle relaxants, may be appropriate initial therapy. Other nonsurgical treatments are commonly used, including opioid analgesics, oral glucocorticoids, cervical traction, and immobilization with a hard or soft cervical collar. However, there are no randomized trials to establish the effectiveness of these treatments in comparison to natural history alone. Soft cervical collars can be modestly helpful by limiting spontaneous and reflex neck movements that exacerbate pain.

As for lumbar radiculopathy, epidural glucocorticoids may provide short-term symptom relief in cervical radiculopathy. If cervical radiculopathy is due to bony compression from cervical spondylosis with foraminal narrowing, then surgical decompression is generally indicated to forestall progression of neurologic signs.

Surgical treatment can produce rapid and substantial symptom relief, although it is unclear whether long-term outcomes are improved over nonsurgical therapy. Reasonable indications for cervical disk surgery include a progressive radicular motor deficit, functionally limiting pain that fails to respond to conservative management, or spinal cord compression.

Surgical treatments include anterior cervical diskectomy alone, laminectomy with discectomy, discectomy with fusion, and disk arthroplasty (implanting an artificial cervical disk). Fusions can be performed with a variety of techniques. The risk of subsequent radiculopathy or myelopathy at cervical segments adjacent to a fusion is ~3% per year and 26% per decade. Although this risk is sometimes portrayed as a late complication of surgery, it may also reflect the natural history of degenerative cervical disk disease. The durability of disk prostheses is uncertain. Available data do not strongly support one surgical technique over another.

FURTHER READINGS

Bagley LJ: Imaging of spinal trauma. Radiol Clin North Am 44:1, 2006 [PMID: 16297679]

Bhangle SD et al: Back pain made simple: An approach based on principles and evidence. Cleve Clin J Med 76:393, 2009 [PMID: 19570970]

Cassidy JD et al: Effect of eliminating compensation for pain and suffering on the outcome of

insurance claims for whiplash injury. N Engl J Med 342:1179, 2000 [PMID: 10770984]

Cavalier R et al: Spondylolysis and spondylolisthesis in children and adolescents: Diagnosis, natural history, and non-surgical management. J Am Acad Orthop Surg 14:417, 2006 [PMID: 16822889]

Cowan JA JR et al: Changes in the utilization of spinal fusion in the United States. Neurosurgery 59:1, 2006

Datta S et al: Systematic assessment of diagnostic accuracy and therapeutic utility of lumbar facet joint interventions. Pain Physician 12:437, 2009 [PMID: 19305489]

Lamb SE et al: Group cognitive behavioural treatment for low-back pain in primary care: A randomised controlled trial and cost-effectiveness analysis. Lancet 375:916, 2010 Mummaneni PV et al: Clinical and radiographic analysis of cervical disk arthroplasty compared with allograft fusion: A randomized controlled clinical trial. J Neurosurg Spine 6:198, 2007

Peul WC et al: Surgery versus prolonged conservative treatment for sciatica. N Engl J Med 356:2245, 2007 [PMID: 17538084]

Weinstein JN et al: Surgical versus nonsurgical therapy for lumbar spinal stenosis. N Engl J Med 358:794, 2008 [PMID: 18287602]

Weinstein JN et al: Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. N Engl J Med 356:2257, 2007 [PMID: 17538085]

Weinstein JN et al: Surgical vs nonoperative treatment for lumbar disc herniation. The spine patient outcomes research trial (SPORT): A randomized trial. JAMA 296:2441, 2006 [PMID: 17119140]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

FEVER AND HYPERTHERMIA: INTRODUCTION

Body temperature is controlled by the hypothalamus. Neurons in both the preoptic anterior hypothalamus and the posterior hypothalamus receive two kinds of signals: one from peripheral nerves that transmit information from warmth/cold receptors in the skin and the other from the temperature of the blood bathing the region. These two types of signals are integrated by the thermoregulatory center of the hypothalamus to maintain normal temperature. In a neutral temperature environment, the metabolic rate of humans produces more heat than is necessary to maintain the core body temperature in the range of 36.5–37.5°C (97.7–99.5°F).

A normal body temperature is maintained ordinarily, despite environmental variations, because the hypothalamic thermoregulatory center balances the excess heat production derived from metabolic activity in muscle and the liver with heat dissipation from the skin and lungs. According to studies of healthy individuals 18–40 years of age, the mean oral temperature is 36.8° ± 0.4°C (98.2° ± 0.7°F), with low levels at 6 a.m. and higher levels at 4–6 p.m. The maximum normal oral temperature is 37.2°C (98.9°F) at 6 a.m. and 37.7°C (99.9°F) at 4 p.m.; these values define the 99th percentile for healthy individuals. In light of these studies, *an a.m. temperature of >37.2°C (>98.9°F) or a p.m. temperature of >37.7°C (>99.9°F) defines a fever*. The normal daily temperature variation is typically 0.5°C (0.9°F). However, in some individuals recovering from a febrile illness, this daily variation can be as great as 1.0°C. During a febrile illness, the diurnal variation usually is maintained, but at higher, febrile levels. The daily temperature variation appears to be fixed in early childhood; in contrast, elderly individuals can exhibit a reduced ability to develop fever, with only a modest fever even in severe infections.

Rectal temperatures are generally 0.4°C (0.7°F) higher than oral readings. The lower oral readings are probably attributable to mouth breathing, which is a factor in patients with respiratory infections and rapid breathing. Lower-esophageal temperatures closely reflect core temperature. Tympanic membrane (TM) thermometers measure radiant heat from the tympanic membrane and nearby ear canal and display that absolute value (unadjusted mode) or a value automatically calculated from the absolute reading on the basis of nomograms relating the radiant temperature measured to actual core temperatures obtained in clinical studies (adjusted mode). These measurements, although convenient, may be more variable than directly determined oral or rectal values. Studies in adults show that readings are lower with unadjusted-mode than with adjusted-mode TM thermometers and that unadjusted-mode TM values are 0.8°C (1.6°F) lower than rectal temperatures.

In women who menstruate, the a.m. temperature is generally lower in the 2 weeks before ovulation; it then rises by ~0.6°C (1°F) with ovulation and remains at that level until menses occur. Body temperature can be elevated in the postprandial state. Pregnancy and endocrinologic dysfunction also affect body temperature.

FEVER VERSUS HYPERTHERMIA

Fever

Fever is an elevation of body temperature that exceeds the normal daily variation and occurs *in conjunction with an increase in the hypothalamic set point* [e.g., from 37°C to 39°C (98.6°F to 102.2°F)]. This shift of the set point from "normothermic" to febrile levels very much resembles the resetting of the home thermostat to a higher level to raise the ambient temperature in a room. Once the hypothalamic set point is raised, neurons in the vasomotor center are activated and vasoconstriction commences. The individual first notices vasoconstriction in the hands and feet. Shunting of blood away from the periphery to the internal organs essentially decreases heat loss from the skin, and the person feels cold. For most fevers, body temperature increases by 1°–2°C (33.8°–35.6°F). Shivering, which increases heat production from the muscles, may begin at this time; however, shivering is not required if heat conservation mechanisms raise blood temperature sufficiently. Nonshivering heat production from the liver also contributes to increasing core temperature. In humans, behavioral adjustments (e.g., putting on more clothing or bedding) help raise body temperature by decreasing heat loss.

The processes of heat conservation (vasoconstriction) and heat production (shivering and increased nonshivering thermogenesis) continue until the temperature of the blood bathing the hypothalamic neurons matches the new thermostat setting. Once that point is reached, the hypothalamus maintains the temperature at the febrile level by the same mechanisms of heat balance that function in the afebrile state. When the hypothalamic set point is again reset downward (in response to either a

reduction in the concentration of pyrogens or the use of antipyretics), the processes of heat loss through vasodilation and sweating are initiated. Loss of heat by sweating and vasodilation continues until the blood temperature at the hypothalamic level matches the lower setting. Behavioral changes (e.g., removal of clothing) facilitate heat loss.

A fever of >41.5°C (>106.7°F) is called *hyperpyrexia*. This extraordinarily high fever can develop in patients with severe infections but most commonly occurs in patients with central nervous system (CNS) hemorrhages. In the preantibiotic era, fever due to a variety of infectious diseases rarely exceeded 41.1°C (106°F), and there has been speculation that this natural "thermal ceiling" is mediated by neuropeptides that function as central antipyretics.

In rare cases, the hypothalamic set point is elevated as a result of local trauma, hemorrhage, tumor, or intrinsic hypothalamic malfunction. The term *hypothalamic fever* sometimes is used to describe elevated temperature caused by abnormal hypothalamic function. However, most patients with hypothalamic damage have *sub*normal, not *supra*normal, body temperatures.

Hyperthermia

Although most patients with elevated body temperature have fever, there are circumstances in which elevated temperature represents not fever but hyperthermia (also called heat stroke; Table 16-1). Hyperthermia is characterized by an uncontrolled increase in body temperature that exceeds the body's ability to lose heat. The setting of the hypothalamic thermoregulatory center is unchanged. In contrast to fever in infections, hyperthermia does not involve pyrogenic molecules (see "Pyrogens," below). Exogenous heat exposure and endogenous heat production are two mechanisms by which hyperthermia can result in dangerously high internal temperatures. Excessive heat production can easily cause hyperthermia despite physiologic and behavioral control of body temperature. For example, work or exercise in hot environments can produce heat faster than peripheral mechanisms can lose it.

Table 16-1 Causes of Hyperthermia Syndromes

Heat Stroke
Exertional: Exercise in higher than normal heat and/or humidity
Nonexertional: Anticholinergics, including antihistamines; antiparkinsonian drugs; diuretics; phenothiazines
Drug-Induced Hyperthermia
Amphetamines, cocaine, phencyclidine (PCP), methylenedioxymethamphetamine (MDMA; "ecstasy"), lysergic acid diethylamide (LSD), salicylates, lithium, anticholinergics, sympathomimetics
Neuroleptic Malignant Syndrome
Phenothiazines; butyrophenones, including haloperidol and bromperidol; fluoxetine; loxapine; tricyclic dibenzodiazepines; metoclopramide; domperidone; thiothixene; molindone; withdrawal of dopaminergic agents
Serotonin Syndrome
Selective serotonin reuptake inhibitors (SSRIs), monoamine oxidase inhibitors (MAOIs), tricyclic antidepressants
Malignant Hyperthermia
Inhalational anesthetics, succinylcholine
Endocrinopathy
Thyrotoxicosis, pheochromocytoma
Central Nervous System Damage
Cerebral hemorrhage, status epilepticus, hypothalamic injury

Source: After FJ Curley et al (eds): *Intensive Care Medicine*, 3rd ed. Boston, Little, Brown, 1996.

Heat stroke in association with a warm environment may be categorized as exertional or nonexertional. *Exertional heat stroke* typically occurs in individuals exercising at elevated ambient temperatures and/or humidity. In a dry environment and at maximal efficiency, sweating can dissipate ~600 kcal/h, requiring the production of >1 L of sweat. Even in healthy individuals, dehydration or the use of common medications (e.g., over-the-counter antihistamines with anticholinergic side effects) may precipitate exertional heat stroke. *Nonexertional heat stroke* typically occurs in either very young or elderly individuals, particularly during heat waves. According to the Centers for Disease Control and Prevention, there were 7000 deaths attributed to heat injury in the United States from 1979 to 1997. The elderly, the bedridden, persons taking anticholinergic or antiparkinsonian drugs or diuretics, and individuals confined to poorly ventilated and non-air-conditioned environments are most susceptible.

Drug-induced hyperthermia has become increasingly common as a result of the increased use of

prescription psychotropic drugs and illicit drugs. Drug-induced hyperthermia may be caused by monoamine oxidase inhibitors (MAOIs), tricyclic antidepressants, and amphetamines and by the illicit use of phencyclidine (PCP), lysergic acid diethylamide (LSD), methylenedioxymethamphetamine (MDMA, "ecstasy"), or cocaine.

Malignant hyperthermia occurs in individuals with an inherited abnormality of skeletal-muscle sarcoplasmic reticulum that causes a rapid increase in intracellular calcium levels in response to halothane and other inhalational anesthetics or to succinylcholine. Elevated temperature, increased muscle metabolism, muscle rigidity, rhabdomyolysis, acidosis, and cardiovascular instability develop within minutes. This rare condition is often fatal. The *neuroleptic malignant syndrome* occurs in the setting of the use of neuroleptic agents (antipsychotic phenothiazines, haloperidol, prochlorperazine, metoclopramide) or the withdrawal of dopaminergic drugs and is characterized by "lead-pipe" muscle rigidity, extrapyramidal side effects, autonomic dysregulation, and hyperthermia. This disorder appears to be caused by the inhibition of central dopamine receptors in the hypothalamus, which results in increased heat generation and decreased heat dissipation. The *serotonin syndrome*, seen with selective serotonin uptake inhibitors (SSRIs), MAOIs, and other serotonergic medications, has many features that overlap with those of the neuroleptic malignant syndrome (including hyperthermia) but may be distinguished by the presence of diarrhea, tremor, and myoclonus rather than lead-pipe rigidity. Thyrotoxicosis and pheochromocytoma also can cause increased thermogenesis.

It is important to distinguish between fever and hyperthermia since hyperthermia can be rapidly fatal and characteristically does not respond to antipyretics. In an emergency situation, however, making this distinction can be difficult. For example, in systemic sepsis, fever (hyperpyrexia) can be rapid in onset, and temperatures can exceed 40.5°C (104.9°F). Hyperthermia often is diagnosed on the basis of the events immediately preceding the elevation of core temperature—e.g., heat exposure or treatment with drugs that interfere with thermoregulation. In patients with heat stroke syndromes and in those taking drugs that block sweating, the skin is hot but dry, whereas in fever the skin can be cold as a consequence of vasoconstriction. Antipyretics do not reduce the elevated temperature in hyperthermia, whereas in fever—and even in hyperpyrexia—adequate doses of either aspirin or acetaminophen usually result in some decrease in body temperature.

PATHOGENESIS OF FEVER

Pyrogens

The term *pyrogen* is used to describe any substance that causes fever. *Exogenous* pyrogens are derived from outside the patient; most are microbial products, microbial toxins, or whole microorganisms. The classic example of an exogenous pyrogen is the lipopolysaccharide (endotoxin) produced by all gram-negative bacteria. Pyrogenic products of gram-positive organisms include the enterotoxins of *Staphylococcus aureus* and the group A and B streptococcal toxins, also called *superantigens*. One staphylococcal toxin of clinical importance is that associated with isolates of *S. aureus* from patients with toxic shock syndrome. These products of staphylococci and streptococci cause fever in experimental animals when injected intravenously at concentrations of 1–10 μg/kg. Endotoxin is a highly pyrogenic molecule in humans: When it is injected intravenously into volunteers, a dose of 2–3 ng/kg produces fever, leukocytosis, acute-phase proteins, and generalized symptoms of malaise.

Pyrogenic Cytokines

Cytokines are small proteins (molecular mass, 10,000–20,000 Da) that regulate immune, inflammatory, and hematopoietic processes. For example, the elevated leukocytosis seen in several infections with an absolute neutrophilia is the result of the cytokines interleukin (IL) 1 and IL-6. Some cytokines also cause fever; formerly referred to as *endogenous pyrogens*, they are now called *pyrogenic cytokines*. The pyrogenic cytokines include IL-1, IL-6, tumor necrosis factor (TNF), ciliary neurotropic factor (CNTF), and interferon (IFN) α. (IL-18, a member of the IL-1 family, does not appear to be a pyrogenic cytokine.) Other pyrogenic cytokines probably exist. Each cytokine is encoded by a separate gene, and each pyrogenic cytokine has been shown to cause fever in laboratory animals and in humans. When injected into humans, IL-1 and TNF produce fever at low doses (10–100 ng/kg); in contrast, for IL-6, a dose of 1–10 μg/kg is required for fever production.

A wide spectrum of bacterial and fungal products induce the synthesis and release of pyrogenic cytokines, as do viruses. However, fever can be a manifestation of disease in the absence of microbial infection. For example, inflammatory processes, trauma, tissue necrosis, and antigen-antibody complexes can induce the production of IL-1, TNF, and/or IL-6, which—individually or in combination—trigger the hypothalamus to raise the set point to febrile levels.

Elevation of the Hypothalamic Set Point by Cytokines

During fever, levels of prostaglandin E2 (PGE2) are elevated in hypothalamic tissue and the third cerebral ventricle. The concentrations of PGE_2 are highest near the circumventricular vascular organs (organum vasculosum of lamina terminalis)—networks of enlarged capillaries surrounding the hypothalamic regulatory centers. Destruction of these organs reduces the ability of pyrogens to

produce fever. Most studies in animals have failed to show, however, that pyrogenic cytokines pass from the circulation into the brain itself. Thus, it appears that both exogenous and endogenous pyrogens interact with the endothelium of these capillaries and that this interaction is the first step in initiating fever—i.e., in raising the set point to febrile levels.

The key events in the production of fever are illustrated in **Fig. 16-1**. As has been mentioned, several cell types can produce pyrogenic cytokines. Pyrogenic cytokines such as IL-1, IL-6, and TNF are released from the cells and enter the systemic circulation. Although the systemic effects of these circulating cytokines lead to fever by inducing the synthesis of PGE2 , they also induce PGE_2 in peripheral tissues. The increase in PGE_2 in the periphery accounts for the nonspecific myalgias and arthralgias that often accompany fever. It is thought that some systemic PGE_2 escapes destruction by the lung and gains access to the hypothalamus via the internal carotid. However, it is the elevation of PGE_2 in the brain that starts the process of raising the hypothalamic set point for core temperature.

Figure 16-1

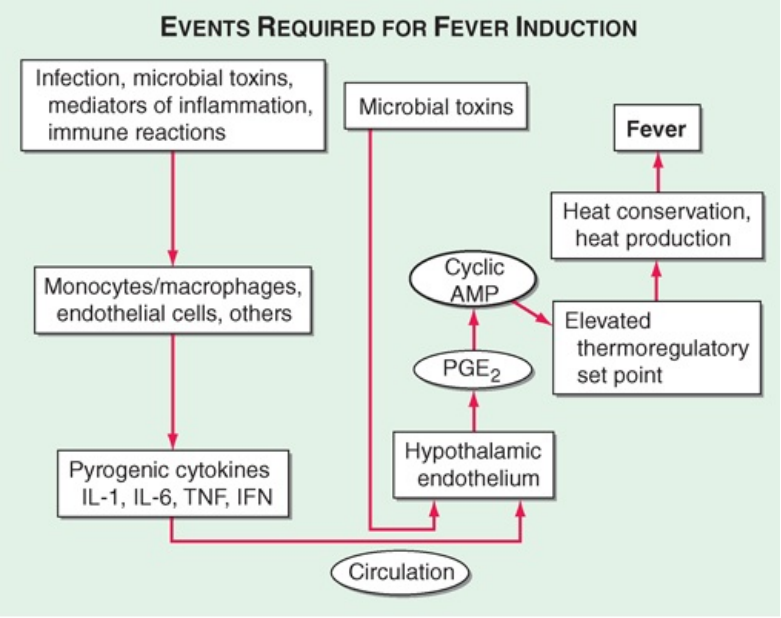

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronology of events required for the induction of fever. AMP, adenosine 5'-monophosphate; IFN, interferon; IL, interleukin; PGE2 , prostaglandin E2 ; TNF, tumor necrosis factor.

There are four receptors for PGE2 , and each signals the cell in different ways. Of the four receptors, the third (EP-3) is essential for fever: when the gene for this receptor is deleted in mice, no fever follows the injection of IL-1 or endotoxin. Deletion of the other PGE_2 receptor genes leaves the fever mechanism intact. Although PGE_2 is essential for fever, it is not a neurotransmitter. Rather, the release of PGE_2 from the brain side of the hypothalamic endothelium triggers the PGE_2 receptor on glial cells, and this stimulation results in the rapid release of cyclic adenosine 5'-monophosphate (cyclic AMP), which is a neurotransmitter. As shown in **Fig. 16-1**, the release of cyclic AMP from the glial cells activates neuronal endings from the thermoregulatory center that extend into the area. The elevation of cyclic AMP is thought to account for changes in the hypothalamic set point either directly or indirectly (by inducing the release of neurotransmitters). Distinct receptors for microbial products are located on the hypothalamic endothelium. These receptors are called *Toll-like receptors* and are similar in many ways to IL-1 receptors. The direct activation of Toll-like receptors also results in PGE_2 production and fever.

Production of Cytokines in the CNS

Cytokines produced in the brain may account for the hyperpyrexia of CNS hemorrhage, trauma, or infection. Viral infections of the CNS induce microglial and possibly neuronal production of IL-1, TNF, and IL-6. In experimental animals, the concentration of a cytokine required to cause fever is several orders of magnitude lower with direct injection into the brain substance or brain ventricles than with systemic injection. Therefore, cytokines produced in the CNS can raise the hypothalamic set point, bypassing the circumventricular organs. CNS cytokines probably account for the hyperpyrexia of CNS hemorrhage, trauma, or infection.

Approach to the Patient: Fever or Hyperthermia

PHYSICAL EXAMINATION

The chronology of events preceding the fever (e.g., exposure to other infected individuals or to vectors of disease) should be ascertained. Electronic devices for measuring oral, tympanic membrane, and rectal temperatures are reliable, but the same site should be used consistently to monitor a febrile disease. Moreover, physicians should be aware that newborns, elderly patients, patients with chronic hepatic or renal failure, and patients taking glucocorticoids may have infections in the absence of fever.

LABORATORY TESTS

The workup should include a complete blood count; a differential count should be performed manually or with an instrument sensitive to the identification of juvenile or band forms, toxic granulations, and Döhle bodies, which are suggestive of bacterial infection. Neutropenia may be present with some viral diseases.

Measurement of circulating cytokines in patients with fever is of little use since levels of pyrogenic cytokines in the circulation often are below the detection limit of the assay or do not coincide with fever. In patients with low-grade fevers, the most valuable measurements are C-reactive protein level and erythrocyte sedimentation rate. These markers of inflammatory processes are particularly helpful in detecting the possible presence of occult disease. Acute-phase reactants are discussed in Chap. 271.

FEVER IN RECIPIENTS OF ANTICYTOKINE THERAPY

With the increasing use of anticytokines to reduce the activity of IL-1, IL-6, IL-12, or TNF in Crohn's disease, rheumatoid arthritis, or psoriasis, the potential of these therapies to blunt the febrile response must be considered. Chronic administration of anticytokines to block cytokine activity has the distinct clinical drawback of lowering the level of host defenses against both routine bacterial and opportunistic infections. The opportunistic infections reported in patients treated with agents that neutralize TNF-α are similar to those reported in the HIV-1-infected population (e.g., new infection with or reactivation of *Mycobacterium tuberculosis*, with dissemination). In nearly all reported cases of infection associated with anticytokine therapy, fever is among the presenting signs. However, the extent to which the febrile response is blunted in these patients remains unknown. This situation is similar to that in patients receiving high-dose glucocorticoid therapy or anti-inflammatory agents such as ibuprofen. Therefore, low-grade fever is of considerable concern in patients receiving anticytokine therapies. The physician must undertake early and rigorous diagnostic evaluation of these patients.

Treatment: Fever or Hyperthermia

THE DECISION TO TREAT FEVER

Most fevers are associated with self-limited infections, such as common viral diseases. The use of antipyretics is not contraindicated in these infections: there is no significant clinical evidence that antipyretics delay the resolution of viral or bacterial infections, nor is there evidence that fever facilitates recovery from infection or acts as an adjuvant to the immune system. In short, routine treatment of fever and its symptoms with antipyretics does no harm and does not slow the resolution of common viral and bacterial infections.

However, with bacterial infections, withholding antipyretic therapy can be helpful in evaluating the effectiveness of a particular antibiotic, particularly in the absence of positive cultures of the infecting organism. Therefore, the routine use of antipyretics can mask an inadequately treated bacterial infection. Withholding antipyretics in some cases may facilitate the diagnosis of an unusual febrile disease. Temperature-pulse dissociation (relative bradycardia) occurs in typhoid fever, brucellosis, leptospirosis, some drug-induced fevers, and factitious fever. In newborns, the elderly, patients with chronic hepatic or renal failure, and patients taking glucocorticoids, fever may not be present despite infection. Hypothermia can be observed in patients with septic shock.

Some infections have characteristic patterns in which febrile episodes are separated by intervals of normal temperature. For example, *Plasmodium vivax* causes fever every third day, whereas fever occurs every fourth day with *P. malariae*. Another relapsing fever is related to *Borrelia* infection, with days of fever followed by a several-day afebrile period and then a relapse of days of fever. In the Pel-Ebstein pattern, fever lasting 3–10 days is followed by afebrile periods of 3–10 days; this pattern can be classic for Hodgkin's disease and other lymphomas. In cyclic neutropenia, fevers occur every 21 days and accompany the neutropenia. There is no periodicity of fever in patients with familial Mediterranean fever. However, these patterns have limited or no diagnostic value compared with specific and rapid laboratory tests.

ANTICYTOKINE THERAPY TO REDUCE FEVER IN AUTOIMMUNE AND AUTOINFLAMMATORY DISEASES

Recurrent fever is documented at some point in most autoimmune diseases but in all autoinflammatory diseases. Although fever also can be a manifestation of autoimmune diseases, recurrent fevers are

characteristic of autoinflammatory diseases. The autoinflammatory diseases (Table 16-2) include adult and juvenile Still's disease, familial Mediterranean fever, and hyper-IgD syndrome. In addition to recurrent fevers, neutrophilia and serosal inflammation characterize autoinflammatory diseases. The fevers associated with these illnesses are reduced dramatically by blocking of IL-1β activity. Anticytokines therefore reduce fever in autoimmune and autoinflammatory diseases. Although fevers in autoinflammatory diseases are mediated by IL-1β, these patients also respond to antipyretics.

Table 16-2 Autoinflammatory Diseases
Adult and juvenile Still's disease
Cryopyrin-associated periodic syndromes (CAPS)
Familial Mediterranean fever
Hyper-IgD syndrome
Behçet's syndrome
Macrophage activation syndrome
Normocomplementemic urticarial vasculitis Antisynthetase myositis
PAPA[a] syndrome
Blau syndrome
Gouty arthritis

[a]Pyogenic arthritis, pyoderma gangrenosum, and acne.

MECHANISMS OF ANTIPYRETIC AGENTS

The reduction of fever by lowering of the elevated hypothalamic set point is a direct function of reducing the level of PGE2 in the thermoregulatory center. The synthesis of PGE_2 depends on the constitutively expressed enzyme cyclooxygenase. The substrate for cyclooxygenase is arachidonic acid released from the cell membrane, and this release is the rate-limiting step in the synthesis of PGE_2. Therefore, inhibitors of cyclooxygenase are potent antipyretics. The antipyretic potency of various drugs is directly correlated with the inhibition of brain cyclooxygenase. Acetaminophen is a poor cyclooxygenase inhibitor in peripheral tissue and lacks noteworthy anti-inflammatory activity; in the brain, however, acetaminophen is oxidized by the p450 cytochrome system, and the oxidized form inhibits cyclooxygenase activity. Moreover, in the brain, the inhibition of another enzyme, COX-3, by acetaminophen may account for the antipyretic effect of this agent. However, COX-3 is not found outside the CNS.

Oral aspirin and acetaminophen are equally effective in reducing fever in humans. Nonsteroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen and specific inhibitors of COX-2 are also excellent antipyretics. Chronic high-dose therapy with antipyretics such as aspirin or any NSAID does not reduce normal core body temperature. Thus, PGE2 appears to play no role in normal thermoregulation.

As effective antipyretics, glucocorticoids act at two levels. First, similar to the cyclooxygenase inhibitors, glucocorticoids reduce PGE2 synthesis by inhibiting the activity of phospholipase A_2, which is needed to release arachidonic acid from the cell membrane. Second, glucocorticoids block the transcription of the mRNA for the pyrogenic cytokines. Limited experimental evidence indicates that ibuprofen and COX-2 inhibitors reduce IL-1-induced IL-6 production and may contribute to the antipyretic activity of NSAIDs.

REGIMENS FOR THE TREATMENT OF FEVER

The objectives in treating fever are first to reduce the elevated hypothalamic set point and second to facilitate heat loss. Reducing fever with antipyretics also reduces systemic symptoms of headache, myalgias, and arthralgias.

Oral aspirin and NSAIDs effectively reduce fever but can adversely affect platelets and the gastrointestinal tract. Therefore, use of acetaminophen is preferred as an antipyretic. In children, acetaminophen or oral ibuprofen must be used because aspirin increases the risk of Reye's syndrome. If the patient cannot take oral antipyretics, parenteral preparations of NSAIDs and rectal suppositories of various antipyretics can be used.

Treatment of fever in some patients is highly recommended. Fever increases the demand for oxygen (i.e., for every increase of 1°C over 37°C, there is a 13% increase in oxygen consumption) and can aggravate the condition of patients with preexisting impairment of cardiac, pulmonary, or CNS function. Children with a history of febrile or nonfebrile seizure should be treated aggressively to reduce fever.

However, it is unclear what triggers the febrile seizure, and there is no correlation between absolute temperature elevation and onset of a febrile seizure in susceptible children.

In hyperpyrexia, the use of cooling blankets facilitates the reduction of temperature; however, cooling blankets should not be used without oral antipyretics. In hyperpyretic patients with CNS disease or trauma (CNS bleeding), reducing core temperature mitigates the detrimental effects of high temperature on the brain.

TREATING HYPERTHERMIA

A high core temperature in a patient with an appropriate history (e.g., environmental heat exposure or treatment with anticholinergic or neuroleptic drugs, tricyclic antidepressants, succinylcholine, or halothane) along with appropriate clinical findings (dry skin, hallucinations, delirium, pupil dilation, muscle rigidity, and/or elevated levels of creatine phosphokinase) suggests hyperthermia. Antipyretics are of no use in treating hyperthermia. Physical cooling with sponging, fans, cooling blankets, and even ice baths should be initiated immediately in conjunction with the administration of IV fluids and appropriate pharmacologic agents (see below). If sufficient cooling is not achieved by external means, internal cooling can be achieved by gastric or peritoneal lavage with iced saline. In extreme circumstances, hemodialysis or even cardiopulmonary bypass with cooling of blood may be performed.

Malignant hyperthermia should be treated immediately with cessation of anesthesia and IV administration of dantrolene sodium. The recommended dose of dantrolene is 1–2.5 mg/kg given intravenously every 6 h for at least 24–48 h—until oral dantrolene can be administered, if needed. Dantrolene at similar doses is indicated in the neuroleptic malignant syndrome and in drug-induced hyperthermia and may even be useful in the hyperthermia of the serotonin syndrome and thyrotoxicosis. The neuroleptic malignant syndrome also may be treated with bromocriptine, levodopa, amantadine, or nifedipine or by induction of muscle paralysis with curare and pancuronium. Tricyclic antidepressant overdose may be treated with physostigmine.

FURTHER READINGS

De Koning HD et al: Beneficial response to anakinra and thalidomide in Schnitzler's syndrome. Ann Rheum Dis 65:542, 2006

Dinarello CA: Immunological and inflammatory functions of the interleukin-1 family. Annu Rev Immunol 27:519, 2009 [PMID: 19302047]

———: Infection, fever, and exogenous and endogenous pyrogens: Some concepts have changed. J Endotoxin Res 10:202, 2004

Hawkins PN et al: Spectrum of clinical features in Muckle-Wells syndrome and response to anakinra. Arthritis Rheum 50:607, 2004 [PMID: 14872505]

Hoffman HM et al: Prevention of cold-associated acute inflammation in familial cold autoinflammatory syndrome by interleukin-1 receptor antagonist. Lancet 364:1779, 2004 [PMID: 15541451]

Kastner DL et al: Autoinflammatory disease reloaded: A clinical perspective. Cell 140:784; 2010

Keane J et al: Tuberculosis associated with infliximab, a tumor necrosis factor-α-neutralizing agent. N Engl J Med 345:1098, 2001 [PMID: 11596589]

Pascual V et al: Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade. J Exp Med 201:1479, 2005 [PMID: 15851489]

Simon A, van der Meer JW: Pathogenesis of familial periodic fever syndromes or hereditary autoinflammatory syndromes. Am J Physiol Regul Integr Comp Physiol 292:R86, 2007

——— et al: Beneficial response to interleukin-1 receptor antagonist in TRAPS. Am J Med 117:208, 2004

Wallis RS et al: Differential effects of TNF blockers on TB immunity. Ann Rheum Dis 64(Suppl 3):132, 2005

——— et al: Granulomatous infectious diseases associated with tumor necrosis factor antagonists. Clin Infect Dis 38:1261, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

FEVER AND RASH: INTRODUCTION

The acutely ill patient with fever and rash often presents a diagnostic challenge for physicians. The distinctive appearance of an eruption in concert with a clinical syndrome may facilitate a prompt diagnosis and the institution of life-saving therapy or critical infection-control interventions. **Representative images of many of the rashes discussed in this chapter are included in Chap. e7.**

Approach to the Patient: Fever and Rash

A thorough history of patients with fever and rash includes the following relevant information: immune status, medications taken within the previous month, specific travel history, immunization status, exposure to domestic pets and other animals, history of animal (including arthropod) bites, existence of cardiac abnormalities, presence of prosthetic material, recent exposure to ill individuals, and exposure to sexually transmitted diseases. The history should also include the site of the onset of the rash and its direction and rate of spread.

A thorough physical examination entails close attention to the rash, with an assessment and precise definition of its salient features. First, it is critical to determine the *type* of lesions that make up the eruption. *Macules* are flat lesions defined by an area of changed color (i.e., a blanchable erythema). *Papules* are raised, solid lesions <5 mm in diameter; *plaques* are lesions >5 mm in diameter with a flat, plateaulike surface; and *nodules* are lesions >5 mm in diameter with a more rounded configuration. *Wheals* (urticaria, hives) are papules or plaques that are pale pink and may appear annular (ringlike) as they enlarge; classic (nonvasculitic) wheals are transient, lasting only 24 h in any defined area. *Vesicles* (<5 mm) and *bullae* (>5 mm) are circumscribed, elevated lesions containing fluid. *Pustules* are raised lesions containing purulent exudate; vesicular processes such as varicella or herpes simplex may evolve to pustules. *Nonpalpable purpura* is a flat lesion that is due to bleeding into the skin. If <3 mm in diameter, the purpuric lesions are termed *petechiae*; if >3 mm, they are termed *ecchymoses. Palpable purpura* is a raised lesion that is due to inflammation of the vessel wall (vasculitis) with subsequent hemorrhage. An *ulcer* is a defect in the skin extending at least into the upper layer of the dermis, and an *eschar* (tâche noire) is a necrotic lesion covered with a black crust.

Other pertinent features of rashes include their *configuration* (i.e., annular or target), the *arrangement* of their lesions, and their *distribution* (i.e., central or peripheral).

For further discussion, see Chaps. 51, 53, and 121.

CLASSIFICATION OF RASH

This chapter reviews rashes that reflect systemic disease, but it does not include localized skin eruptions (i.e., cellulitis, impetigo) that may also be associated with fever (Chap. 125). This chapter does not intend to be all-inclusive, but it covers the most important and most common diseases associated with fever and rash. Rashes are classified herein on the basis of the morphology and distribution of lesions. For practical purposes, this classification system is based on the most typical disease presentations. However, morphology may vary as rashes evolve, and the presentation of diseases with rashes is subject to many variations (Chap. 53). For instance, the classic petechial rash of Rocky Mountain spotted fever (RMSF) (Chap. 174) may initially consist of blanchable erythematous macules distributed peripherally; at times, however, the rash associated with RMSF may not be predominantly acral or no rash may develop at all.

Diseases with fever and rash may be classified by type of eruption: centrally distributed maculopapular, peripheral, confluent desquamative erythematous, vesiculobullous, urticaria-like, nodular, purpuric, ulcerated, or eschar. Diseases are listed by these categories in **Table 17-1**, and many are highlighted in the text. However, for a more detailed discussion of each disease associated with a rash, the reader is referred to the chapter dealing with that specific disease. (Reference chapters are cited in the text and listed in Table 17-1.)

Table 17-1 Diseases Associated with Fever and Rash

Disease	Etiology	Description	Group Affected/Epidemiologic Factors	Clinical Syndrome	Chapter
Centrally distributed maculopapular eruptions					
Acute meningococcemia[a]	—	—	—	—	143
Drug-induced hypersensitivity syndrome/drug reaction with eosinophilia and systemic symptoms (DIHS/DRESS)[b]	—	—	—	—	55
Rubeola (measles, first disease)	Paramyxovirus	Discrete lesions that become confluent as rash spreads from hairline downward, sparing palms and soles; lasts ≥3 days; Koplik's spots	Nonimmune individuals	Cough, conjunctivitis, coryza, severe prostration	192
Rubella (German measles, third disease)	Togavirus	Spreads from hairline downward, clearing as it spreads; Forschheimer spots	Nonimmune individuals	Adenopathy, arthritis	193

Erythema infectiosum (fifth disease)	Human parvovirus B19	Brightred "slapped-cheeks" appearance followed by lacy reticular rash that waxes and wanes over 3 weeks; rarely, papular-purpuric "gloves-and-socks" syndrome on hands and feet	Most common among children 3–12 years old; occurs in winter and spring	Mild fever; arthritis in adults; rash following resolution of fever	184
Exanthem subitum (roseola, sixth disease)	Human herpesvirus 6	Diffuse maculopapular eruption over trunk and neck; resolves within 2 days	Usually affects children <3 years old	Rash following resolution of fever; similar to Boston exanthem (echovirus 16); febrile seizures may occur	182
Primary HIV infection	HIV	Nonspecific diffuse macules and papules; less commonly, urticarial or vesicular oral or genital ulcers	Individuals recently infected with HIV	Pharyngitis, adenopathy, arthralgias	189
Infectious mononucleosis	Epstein-Barr virus	Diffuse maculopapular eruption (5% of cases; 90% if ampicillin is given); urticaria, petechiae in some cases; periorbital edema (50%); palatal petechiae (25%)	Adolescents, young adults	Hepatosplenomegaly, pharyngitis, cervical lymphadenopathy, atypical lymphocytosis, heterophile antibody	181
Other viral exanthems	Echoviruses 2, 4, 9, 11, 16, 19, 25; coxsackieviruses A9, B1, B5; etc.	Wide range of skin findings that may mimic rubella or measles	Affect children more commonly than adults	Nonspecific viral syndromes	191
Exanthematous drug-induced eruption	Drugs (antibiotics, anticonvulsants, diuretics, etc.)	Intensely pruritic, brightred macules and papules, symmetric on trunk and extremities; may become confluent	Occurs 2–3 days after exposure in previously sensitized individuals; otherwise, after 2–3 weeks (but can occur anytime, even shortly after drug is discontinued)	Variable findings: fever and eosinophilia	55
Epidemic typhus	*Rickettsia prowazekii*	Maculopapular eruption appearing in axillae, spreading to trunk and later to extremities; usually spares face, palms, soles; evolves from blanchable macules to confluent eruption with petechiae; rash evanescent in recrudescent typhus (Brill-Zinsser disease)	Exposure to body lice; occurrence of recrudescent typhus as relapse after 30–50 years	Headache, myalgias; 10–40% mortality if untreated; milder clinical presentation in recrudescent form	174
Endemic (murine) typhus	*Rickettsia typhi*	Maculopapular eruption, usually sparing palms, soles	Exposure to rat or cat fleas	Headache, myalgias	174
Scrub typhus	*Orientia tsutsugamushi*	Diffuse macular rash starting on trunk; eschar at site of mite bite	Endemic in South Pacific, Australia, Asia; transmitted by mites	Headache, myalgias, regional adenopathy; mortality up to 30% if untreated	174
Rickettsial spotted fevers	*Rickettsia conorii* (boutonneuse fever), *Rickettsia australis* (North Queensland tick typhus), *Rickettsia sibirica* (Siberian tick typhus), and others	Eschar common at bite site; maculopapular (rarely, vesicular and petechial) eruption on proximal extremities, spreading to trunk and face	Exposure to ticks; *R. conorii* in Mediterranean region, India, Africa; *R. australis* in Australia; *R. sibirica* in Siberia, Mongolia	Headache, myalgias, regional adenopathy	174

Human monocytotropic ehrlichiosis[c]	*Ehrlichia chaffeensis*	Maculopapular eruption (40% of cases), involves trunk and extremities; may be petechial	Tickborne; most common in U.S. Southeast, southern Midwest, and mid-Atlantic regions	Headache, myalgias, leukopenia	174
Leptospirosis	*Leptospira interrogans*	Maculopapular eruption; conjunctivitis; scleral hemorrhage in some cases	Exposure to water contaminated with animal urine	Myalgias; aseptic meningitis; *fulminant form?*: icterohemorrhagic fever (Weil's disease)	171
Lyme disease	*Borrelia burgdorferi*	Papule expanding to erythematous annular lesion with central clearing (erythema migrans; average diameter, 15 cm), sometimes with concentric rings, sometimes with indurated or vesicular center; multiple secondary erythema migrans lesions in some cases	Bite of tick vector	Headache, myalgias, chills, photophobia occurring acutely; CNS disease, myocardial disease, arthritis weeks to months later in some cases	173
Southern tick-associated rash illness (STARI, Master's disease)	*Borrelia lonestari*	Similar to erythema migrans of Lyme disease with several differences, including: multiple secondary lesions less likely; lesions tending to be smaller (average diameter, ~8 cm); central clearing more likely	Bite of tick vector *Amblyomma americanum* (Lone Star tick); often found in regions where Lyme disease is uncommon, including southern United States	Compared with Lyme disease: fewer constitutional symptoms, tick bite more likely to be recalled; other Lyme disease sequelae lacking	173
Typhoid fever	*Salmonella typhi*	Transient, blanchable erythematous macules and papules, 2–4 mm, usually on trunk (rose spots)	Ingestion of contaminated food or water (rare in U.S.)	Variable abdominal pain and diarrhea; headache, myalgias, hepatosplenomegaly	153
Dengue fever[d]	Dengue virus (4 serotypes; flaviviruses)	Rash in 50% of cases; initially diffuse flushing; midway through illness, onset of maculopapular rash, which begins on trunk and spreads centrifugally to extremities and face; pruritus, hyperesthesia in some cases; after defervescence, petechiae on extremities in some cases	Occurs in tropics and subtropics; transmitted by mosquito	Headache, musculoskeletal pain ("breakbone fever"); leukopenia; occasionally biphasic ("saddleback") fever	196
Ratbite fever (sodoku)	*Spirillum minus*	Eschar at bite site; then blotchy violaceous or redbrown rash involving trunk and extremities	Rat bite; primarily found in Asia; rare in U.S.	Regional adenopathy, recurrent fevers if untreated	e24
Relapsing fever	*Borrelia* species	Central rash at end of febrile episode; petechiae in some cases	Exposure to ticks or body lice	Recurrent fever, headache, myalgias, hepatosplenomegaly	172
Erythema marginatum (rheumatic fever)	Group A *Streptococcus*	Erythematous annular papules	Patients with rheumatic fever	Pharyngitis preceding	322

		and plaques occurring as polycyclic lesions in waves over trunk, proximal extremities; evolving and resolving within hours		polyarthritis, carditis, subcuta-neous nodules, chorea	
Systemic lupus erythematosus	Autoimmune disease	Macular and papular erythema, often in sun-exposed areas; discoid lupus lesions (local atrophy, scale, pigmentary changes); periungual telangiectasis; malar rash; vasculitis sometimes causing urticaria, palpable purpura; oral erosions in some cases	Most common in young to middle-aged women; flares precipitated by sun exposure	Arthritis; cardiac, pulmonary, renal, hematologic, and vasculitic disease	319
Still's disease	Autoimmune disease	Transient 2- to 5-mm erythematous papules appearing at height of fever on trunk, proximal extremities; lesions evanescent	Children and young adults	High spiking fever, polyarthritis, splenomegaly; erythrocyte sedimentation rate, >100 mm/h	337
African trypanosomiasis	*Trypanosoma brucei rhodesiense/gambiense*	Blotchy or annular erythematous macular and papular rash (trypanid), primarily on trunk; pruritus; chancre at site of tsetse fly bite may precede rash by several weeks	Tsetse fly bite in East (*T. brucei rho-desiense)* or West *(T. brucei gambiense)* Africa	Hemolymphatic disease followed by meningoencephali-tis; Winterbottom's sign (posterior cervical lymphadenopathy) *(T. brucei gambiense)*	213
Arcanobacterial pharyngitis	*Arcanobacterium (Corynebacterium) haemolyticum*	Diffuse, erythematous, maculopapular eruption involving trunk and proximal extremities; may desquamate	Children and young adults	Exudative pharyngitis, lymphadenopathy	138
Peripheral eruptions					
Chronic meningococcemia, disseminated gonococcal infection,[a] human parvovirus B19 infection[e]	—	—	—	—	143, 144, 184
Rocky Mountain spotted fever	*Rickettsia rickettsii*	Rash beginning on wrists and ankles and spreading centripetally; appears on palms and soles later in disease; lesion evolution from blanchable macules to petechiae	Tick vector; wide-spread but more common in south-eastern and southwest-central U.S.	Headache, myalgias, abdominal pain; mortality up to 40% if un-treated	174
Secondary syphilis	*Treponema pallidum*	Coincident primary chancre in 10% of cases; copper-colored, scaly papular eruption, diffuse but prominent on palms and soles; rash never vesicular	Sexually transmitted	Fever, constitutional symptoms	169

		in adults; condyloma latum, mucous patches, and alopecia in some cases			
Chikungunya fever	Chikungunya virus	Maculopapular eruption; prominent on upper extremities and face, but can also occur on trunk and lower extremities	*Aedes aegypti* and *A. albopictus* mosquito bites; primarily in Africa and Indian Ocean region	Severe polyarticular, migratory arthralgias, especially involving small joints (e.g., hands, wrists, ankles)	196
Hand-foot-and-mouth disease	Coxsackievirus A16 most common cause	Tender vesicles, erosions in mouth; 0.25-cm papules on hands and feet with rim of erythema evolving into tender vesicles	Summer and fall; primarily children <10 years old; multiple family members	Transient fever	191
Erythema multiforme (EM)	Infection, drugs, idiopathic causes	Target lesions (central erythema surrounded by area of clearing and another rim of erythema) up to 2 cm; symmetric on knees, elbows, palms, soles; spreads centripetally; papular, sometimes vesicular; when extensive and involving mucous membranes, termed *EM major*	Herpes simplex virus or *Mycoplasma pneumoniae* infection; drug intake (i.e., sulfa, phenytoin, penicillin)	50% younger than 20 years old; fever more common in most severe form, EM major, which can be confused with Stevens-Johnson syndrome (but EM major lacks prominent skin sloughing)	*f*
Rat-bite fever (Haverhill fever)	*Streptobacillus moniliformis*	Maculopapular eruption over palms, soles, and extremities; tends to be more severe at joints; eruption sometimes becoming generalized; may be purpuric; may desquamate	Rat bite, ingestion of contaminated food	Myalgias; arthritis (50%); fever recurrence in some cases	e24
Bacterial endocarditis	*Streptococcus, Staphylococcus*, etc.	*Subacute course*: Osler's nodes (tender pink nodules on finger or toe pads); petechiae on skin and mucosa; splinter hemorrhages. *Acute course* (*Staphylococcus aureus*): Janeway lesions (painless erythematous or hemorrhagic macules, usually on palms and soles)	Abnormal heart valve *(Streptococcus)*, intravenous drug use	New or changing heart murmur	124
Confluent desquamative erythemas					
Scarlet fever (second disease)	Group A *Streptococcus* (pyrogenic exotoxins A, B, C)	Diffuse blanchable erythema beginning on face and spreading to trunk and extremities; circumoral pallor; "sandpaper"	Most common among children 2–10 years old; usually follows group A streptococcal pharyngitis	Fever, pharyngitis, headache	136

		texture to skin; accentuation of linear erythema in skin folds (Pastia's lines); enanthem of white evolving into red "strawberry" tongue; desquamation in second week			
Kawasaki disease	Idiopathic causes	Rash similar to scarlet fever (scarlatiniform) or erythema multiforme; fissuring of lips, strawberry tongue; conjunctivitis; edema of hands, feet; desquamation later in disease	Children <8 years old	Cervical adenopathy, pharyngitis, coronary artery vasculitis	53, 326
Streptococcal toxic shock syndrome	Group A *Streptococcus* (associated with pyrogenic exotoxin A and/or B or certain M types)	When present, rash often scarlatiniform	May occur in setting of severe group A streptococcal infections (e.g., necrotizing fasciitis, bacteremia, pneumonia)	Multiorgan failure, hypotension; 30% mortality rate	136
Staphylococcal toxic shock syndrome	*S. aureus* (toxic shock syndrome toxin 1, enterotoxin B or C)	Diffuse erythema involving palms; pronounced erythema of mucosal surfaces; conjunctivitis; desquamation 7–10 days into illness	Colonization with toxin-producing *S. aureus*	Fever >39°C (>102°F), hypotension, multiorgan dysfunction	135
Staphylococcal scaldedskin syndrome	*S. aureus*, phage group II	Diffuse tender erythema, often with bullae and desquamation; Nikolsky's sign	Colonization with toxin-producing *S. aureus*; occurs in children <10 years old (termed "Ritter's disease" in neonates) or adults with renal dysfunction	Irritability; nasal or conjunctival secretions	135
Exfoliative erythroderma syndrome	Underlying psoriasis, eczema, drug eruption, mycosis fungoides	Diffuse erythema (often scaling) interspersed with lesions of underlying condition	Usually occurs in adults over age 50; more common among men	Fever, chills (i.e., difficulty with thermoregulation); lymphadenopathy	52, 55
DIHS/DRESS	Aromatic anticonvulsants; other drugs, including sulfonamides, minocycline	Maculopapular eruption (mimicking exanthematous drug rash) sometimes progressing to exfoliative erythroderma; profound edema, especially facial; pustules may occur	Individuals genetically unable to detoxify arene oxides (anticonvulsants), patients with slow *N*-acetylating capacity (sulfonamides)	Lymphadenopathy, multiorgan failure (especially hepatic), eosinophilia, atypical lymphocytes; mimics sepsis	55
Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN)	Drugs (80% of cases; often allopurinol, anticonvulsants, antibiotics), infection, idiopathic	Erythematous and purpuric macules, sometimes targetoid, or diffuse erythema progressing to bullae, with sloughing and necrosis of entire epidermis; Nikolsky's sign; involves mucosal surfaces; TEN (>30% epidermal necrosis) is maximal form; SJS involves <10%; SJS/TEN overlap involves	Uncommon among children; more common among patients with HIV infection, SLE, certain HLA types, or slow acetylators	Dehydration, sepsis sometimes resulting from lack of normal skin integrity; up to 30% mortality	55

		10–30% of epidermis			
Vesiculobullous or pustular eruptions					
Hand-foot-and-mouth syndrome[g]; staphylococcal scalded-skin syndrome; toxic epidermal necrolysis[b];DIHS/DRESS[b]	—	—	—	—	—[f]
Varicella (chicken-pox)	Varicella-zoster virus	Macules (2–3 mm) evolving into papules, then vesicles (sometimes umbilicated), on an erythematous base ("dewdrops on a rose petal"); pustules then forming and crusting; lesions appearing in crops; may involve scalp, mouth; intensely pruritic	Usually affects children; 10% of adults susceptible; most common in late winter and spring	Malaise; generally mild disease in healthy children; more severe disease with complications in adults and immunocompromised children	180
Pseudomonas "hot-tub" folliculitis	*Pseudomonas aeruginosa*	Pruritic erythematous follicular, papular, vesicular, or pustular lesions that may involve axillae, buttocks, abdomen, and especially areas occluded by bathing suits; can mani-fest as tender isolated nodules on palmar or plantar surfaces (the latter designated "*Pseudomonas* hot-foot syndrome")	Bathers in hot tubs or swimming pools; occurs in out-breaks	Earache, sore eyes and/or throat; generally self-limited	152
Variola (smallpox)	Variola major virus	Red macules on tongue, palate evolving to papules and vesicles; skin macules evolving to papules, then vesicles, then pustules over 1 week, with subsequent lesion crusting; lesions initially appearing on face and spreading centrifugally from trunk to extremities; differs from varicella in that (1) skin lesions in any given area are at same stage of development and (2) there is a prominent distribution of lesions on face and extremities (including palms, soles)	Nonimmune individuals exposed to smallpox	Prodrome of fever, headache, backache, myalgias; vomiting in 50% of cases	221
Primary herpes simplex virus (HSV) infection	HSV	Erythema rapidly followed by hallmark painful *grouped vesicles* that may evolve into pustules that ulcerate, especially on	Primary infection most common among children and young adults for HSV-1 and among sexually active young adults for HSV-2; no fever in recurrent infection	Regional lymphadenopathy	179

		mucosal surfaces; lesions at site of inoculation: commonly gingivostomatitis for HSV-1 and genital lesions for HSV-2; recurrent disease milder (e.g., herpes labialis does not involve oral mucosa)			
Disseminated herpesvirus infection	Varicella-zoster virus or HSV	Generalized vesicles that can evolve to pustules and ulcerations; individual lesions similar for varicella-zoster and HSV. *Zoster cutaneous dissemination*: >25 lesions extending outside involved dermatome. *HSV*: extensive, progressive mucocutaneous lesions that may occur in absence of dissemination, sometimes disseminate in eczematous skin (eczema herpeticum); HSV visceral dissemination may occur with only localized mucocutaneous disease; in disseminated neonatal disease, skin lesions diagnostically helpful when present, but rash absent in a substantial minority of cases	Patients with immunosuppression, eczema; neonates	Visceral organ involvement (especially liver) in some cases; neonatal disease particularly severe	179, 180, 381
Rickettsialpox	*Rickettsia akari*	Eschar found at site of mite bite; generalized rash involving face, trunk, extremities; may involve palms and soles; <100 papules and plaques (2–10 mm); tops of lesions developing vesicles that may evolve into pustules	Seen in urban settings; transmitted by mouse mites	Headache, myalgias, regional adenopathy; mild disease	174
Acute generalized eruptive pustulosis (AGEP)	Drugs (mostly anticonvulsants or antimicrobials); also viral	Tiny sterile nonfollicular pustules on erythematous, edema-tous skin; begins on face and in body folds, then becomes general-ized	Appears 2–21 days after start of drug therapy, depend-ing on whether previously sensitized	Acute fever, pruritus, leukocytosis	55
Disseminated *Vibrio vulnificus* infection	*V. vulnificus*	Erythematous lesions evolving into hemorrhagic bullae and then into necrotic ulcers	Patients with cirrhosis, diabetes, renal failure; ex-posure by ingestion of contaminated saltwater, seafood	Hypotension; 50% mortality	156

Ecthyma gangrenosum	*P. aeruginosa*, other gram-negative rods, fungi	Indurated plaque evolving into hemorrhagic bulla or pustule that sloughs, resulting in eschar formation; erythema-tous halo; most common in axillary, groin, perianal regions	Usually affects neutropenic patients; occurs in up to 28% of individuals with *Pseudomonas* bacteremia	Clinical signs of sepsis	152
Urticaria-like eruptions					
Urticarial vasculitis	Serum sickness, often due to infection (including hepatitis B, enteroviral, para-sitic), drugs; connective tissue disease	Erythematous, edematous "urticaria-like" plaques, pruritic or burning; unlike urticaria: typical lesion duration >24 h (up to 5 days) and lack of complete lesion blanching with compression due to hemorrhage	Patients with serum sickness (including hepatitis B), connective tissue disease	Fever variable; arthralgias/arthritis	326[f]
Nodular eruptions					
Disseminated infection	Fungi (e.g., candidiasis, histoplasmosis, cryptococcosis, sporotrichosis, coccidioidomycosis); mycobacteria	Subcutaneous nodules (up to 3 cm); fluctuance, draining common with mycobacteria; necrotic nodules (extremities, periorbital or nasal regions) common with *Aspergillus*, *Mucor*	Immunocompromised hosts (i.e., bone marrow transplant recipients, patients undergoing chemotherapy, HIV-infected patients, alcoholics)	Features vary with organism	—[f]
Erythema nodosum (septal panniculitis)	Infections (e.g., streptococcal, fungal, mycobacterial, yersinial); drugs (e.g., sulfas, penicillins, oral contraceptives); sarcoidosis; idiopathic causes	Large, violaceous, nonulcerative, subcutaneous nodules; exquisitely tender; usually on lower legs but also on upper extremities	More common among girls and women 15–30 years old	Arthralgias (50%); features vary with associated condi-tion	—[f]
Sweet's syndrome (acute febrile neutrophilic dermatosis)	Yersinial infection; lymphoproliferative disorders; idiopathic causes	Tender red or blue edematous nodules giving impression of vesiculation; usually on face, neck, upper extremities; when on lower extremities, may mimic erythema nodosum	More common among women and among persons 30–60 years old; 20% of cases associated with malignancy (men and women equally affected in this group)	Headache, arthralgias, leukocytosis	53
Bacillary angiomatosis	*Bartonella henselae*, *B. quintana*	Many forms, including erythematous, smooth vascular nodules; friable, exophytic lesions; erythematous plaques (may be dry, scaly); subcutaneous nodules (may be erythematous)	Usually patients with HIV infection	Peliosis of liver and spleen in some cases; lesions sometimes involving multiple organs; bacteremia	160
Purpuric eruptions					
Rocky Mountain spotted fever, rat-bite fever, endocarditis[g]; epidemic typhus[e]; dengue fever[d]; human parvovirus B19 infection[e]	—	—	—	—	—[f]
Acute meningococcemia	*Neisseria meningitidis*	Initially pink maculopapular lesions evolving	Most common among children, individuals with asplenia or terminal	Hypotension, meningitis (sometimes preceded	143

		into petechiae; petechiae rapidly becoming numerous, sometimes enlarging and becoming vesicular; trunk, extremities most commonly involved; may appear on face, hands, feet; may include purpura fulminans (see below) reflecting disseminated intravascular coagulation	complement component deficiency (C5–C8)	by upper respiratory infection)	
Purpura fulminans	Severe disseminated intravascular coagulation	Large ecchymoses with sharply irregular shapes evolving into hemorrhagic bullae and then into black necrotic lesions	Individuals with sepsis (e.g., involving *N. meningitidis*), malignancy, or massive trauma; asplenic patients at high risk for sepsis	Hypotension	143, 271
Chronic meningococcemia	*N. meningitidis*	Variety of recurrent eruptions, including pink maculopapular; nodular (usually on lower extremities); petechial (sometimes developing vesicular centers); purpuric areas with pale bluegray centers	Individuals with complement deficiencies	Fevers, sometimes intermittent; arthritis, myalgias, headache	143
Disseminated gonococcal infection	*Neisseria gonorrhoeae*	Papules (1–5 mm) evolving over 1–2 days into hemorrhagic pustules with gray necrotic centers; hemorrhagic bullae occurring rarely; lesions (usually <40) distributed peripherally near joints (more commonly on upper extremities)	Sexually active individuals (more often females), some with complement deficiency	Low-grade fever, tenosynovitis, arthritis	144
Enteroviral petechial rash	Usually echovirus 9 or coxsackievirus A9	Disseminated petechial lesions (may also be maculopapular, vesicular, or urticarial)	Often occurs in outbreaks	Pharyngitis, headache; aseptic meningitis with echovirus 9	191
Viral hemorrhagic fever	Arboviruses (including dengue) and arenaviruses	Petechial rash	Residence in or travel to endemic areas, other virus exposure	Triad of fever, shock, hemorrhage from mucosa or gastrointestinal tract	196, 197
Thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome	Idiopathic, *Escherichia coli* O157:H7 (Shiga toxin), drugs	Petechiae	Individuals with *E. coli* O157:H7 gastroenteritis (especially children), cancer chemotherapy, HIV infection, autoimmune diseases; pregnant/postpartum women	Fever (not always present), hemolytic anemia, thrombocytopenia, renal dysfunction, neurologic dysfunction; coagulation studies normal	53, 106, 115, 149, 154
Cutaneous small-vessel vasculitis (leukocytoclastic vasculitis)	Infections (including group A *Streptococcus*, viral hepatitis), drugs, chemicals, food allergens, idio-pathic causes	Palpable purpuric lesions appearing in crops on legs or other dependent areas; may become vesicular or ulcerative; usually resolve over 3–4 weeks	Occurs in a wide spectrum of diseases, including connective tissue disease, cryoglobulinemia, malignancy, Henoch-Schönlein purpura (HSP); more common among children	Fever, malaise, arthralgias, myalgias; systemic vasculitis in some cases; renal, joint, and gastrointestinal involvement commonly seen in HSP	53

Eruptions with ulcers and/or eschars					
Scrub typhus, rickettsial spotted fevers, rat-bite fever[e]; rickettsialpox, ecthyma gangrenosum[h]	—	—	—	—	—[f]
Tularemia	*Francisella tularensis*	Ulceroglandular form: erythematous, tender papule evolves into necrotic, tender ulcer with raised borders; in 35% of cases, eruptions (maculo-papular, vesiculopapular, acneiform, urticarial, erythema nodosum, or erythema multiforme) may occur	Exposure to ticks, biting flies, infected animals	Fever, headache, lymphadenopathy	158
Anthrax	*Bacillus anthracis*	Pruritic papule enlarging and evolving into a 1- by 3-cm painless ulcer surrounded by vesicles and then developing a central eschar with edema; residual scar	Exposure to infected animals or animal products, other exposure to anthrax spores	Lymphadenopathy, headache	221

[a] See "Purpuric erup-tions."

[b] See "Confluent desquamative erythemas."

[c] In human granulocytotropic ehrlichiosis or anaplasmosis (caused by *Anaplasma phagocytophila;* most common in the upper midwestern and northeastern regions of the United States), rash is rare.

[d] See "Viral hemorrhagic fever" under "Purpuric erup-tions" for dengue hemorrhagic fever/dengue shock syndrome.

[e] See "Centrally distributed maculopapular eruptions."

[f] See etiology-specific chapters.

[g] See "Peripheral eruptions."

[h] See "Vesiculobullous or pustular eruptions."

Centrally Distributed Maculopapular Eruptions

Centrally distributed rashes, in which lesions are primarily truncal, are the most common type of eruption. The rash of *rubeola* (measles) starts at the hairline 2–3 days into the illness and moves down the body, sparing the palms and soles (Chap. 192). It begins as discrete erythematous lesions, which become confluent as the rash spreads. Koplik's spots (1- to 2-mm white or bluish lesions with an erythematous halo on the buccal mucosa) are pathognomonic for measles and are generally seen during the first 2 days of symptoms. They should not be confused with Fordyce's spots (ectopic sebaceous glands), which have no erythematous halos and are found in the mouth of healthy individuals. Koplik's spots may briefly overlap with the measles exanthem.

Rubella (German measles) also spreads from the hairline downward; unlike that of measles, however, the rash of rubella tends to clear from originally affected areas as it migrates, and it may be pruritic (Chap. 193). Forchheimer spots (palatal petechiae) may develop but are nonspecific because they also develop in mononucleosis (Chap. 181) and scarlet fever (Chap. 136). Postauricular and suboccipital adenopathy and arthritis are common among adults with rubella. Exposure of pregnant women to ill individuals should be avoided, as rubella causes severe congenital abnormalities. Numerous strains of enteroviruses (Chap. 191), primarily echoviruses and coxsackieviruses, cause nonspecific syndromes of fever and eruptions that may mimic rubella or measles. Patients with infectious mononucleosis caused by Epstein-Barr virus (Chap. 181) or with primary infection caused by HIV (Chap. 189) may exhibit pharyngitis, lymphadenopathy, and a nonspecific maculopapular exanthem.

The rash of *erythema infectiosum* (fifth disease), which is caused by human parvovirus B19, primarily affects children 3–12 years old; it develops after fever has resolved as a bright blanchable erythema on the cheeks ("slapped cheeks") with perioral pallor (Chap. 184). A more diffuse rash (often pruritic) appears the next day on the trunk and extremities and then rapidly develops into a lacy reticular eruption that may wax and wane (especially with temperature change) over 3 weeks. Adults with fifth disease often have arthritis, and fetal hydrops can develop in association with this condition in pregnant women.

Exanthem subitum (roseola) is caused by human herpesvirus 6 and is most common among children <3 years of age (Chap. 182). As in erythema infectiosum, the rash usually appears after fever has subsided. It consists of 2- to 3-mm rose-pink macules and papules that rarely coalesce, which occur initially on the trunk and sometimes on the extremities (sparing the face) and fade within 2 days.

Although drug reactions have many manifestations, including urticaria, exanthematous *drug-induced eruptions* (Chap. 55) are most common and are often difficult to distinguish from viral exanthems. Eruptions elicited by drugs are usually more intensely erythematous and pruritic than viral exanthems, but this distinction is not reliable. A history of new medications and an absence of prostration may help to distinguish a drug-related rash from an eruption of another etiology. Rashes may persist for up to two weeks after administration of the offending agent is discontinued. Certain populations are more prone than others to drug rashes. Of HIV-infected patients, 50–60% develop a rash in response to sulfa drugs; 90% of patients with mononucleosis due to Epstein-Barr virus develop a rash when given ampicillin.

Rickettsial illnesses (Chap. 174) should be considered in the evaluation of individuals with centrally distributed maculopapular

eruptions. The usual setting for *epidemic typhus* is a site of war or natural disaster in which people are exposed to body lice. *Endemic typhus* or *leptospirosis* (the latter caused by a spirochete) (Chap. 171) may be seen in urban environments where rodents proliferate. Outside the United States, other rickettsial diseases cause a spotted-fever syndrome and should be considered in residents of or travelers to endemic areas. Similarly, *typhoid fever* , a nonrickettsial disease caused by *Salmonella typhi* (Chap. 153), is usually acquired during travel outside the United States. Dengue fever, caused by a mosquito-transmitted flavivirus, occurs in tropical and subtropical regions of the world (Chap. 196).

Some centrally distributed maculopapular eruptions have distinctive features. Erythema migrans, the rash of Lyme disease (Chap. 173), typically manifests as singular or multiple annular plaques. Untreated erythema migrans lesions usually fade within a month but may persist for more than a year. Southern tick-associated rash illness (STARI) has an erythema migrans–like rash but is less severe than Lyme disease and often occurs in regions where Lyme is not endemic. *Erythema marginatum*, the rash of acute rheumatic fever (Chap. 322), has a distinctive pattern of enlarging and shifting transient annular lesions.

Collagen vascular diseases may cause fever and rash. Patients with *systemic lupus erythematosus* (Chap. 319) typically develop a sharply defined, erythematous eruption in a butterfly distribution on the cheeks (malar rash) as well as many other skin manifestations. *Still's disease* (Chap. 337) presents as an evanescent, salmon-colored rash on the trunk and proximal extremities that coincides with fever spikes.

Peripheral Eruptions

These rashes are alike in that they are most prominent peripherally or begin in peripheral (acral) areas before spreading centripetally. Early diagnosis and therapy are critical in RMSF (Chap. 174) because of its grave prognosis if untreated. Lesions evolve from macular to petechial, start on the wrists and ankles, spread centripetally, and appear on the palms and soles only later in the disease. The rash of *secondary syphilis* (Chap. 169), which may be generalized but is prominent on the palms and soles, should be considered in the differential diagnosis of pityriasis rosea, especially in sexually active patients. Chikungunya fever (Chap. 196), which is transmitted by mosquito bite in Africa and the Indian Ocean region, is associated with a maculopapular eruption and severe polyarticular small-joint arthralgias. *Hand-foot-and-mouth disease* (Chap. 191), most commonly caused by coxsackievirus A16, is distinguished by tender vesicles distributed peripherally and in the mouth; outbreaks commonly occur within families. The classic target lesions of *erythema multiforme* appear symmetrically on the elbows, knees, palms, soles, and face. In severe cases, these lesions spread diffusely and involve mucosal surfaces. Lesions may develop on the hands and feet in *endocarditis* (Chap. 124).

Confluent Desquamative Erythemas

These eruptions consist of diffuse erythema frequently followed by desquamation. The eruptions caused by group A *Streptococcus* or *Staphylococcus aureus* are toxin-mediated. *Scarlet fever* (Chap. 136) usually follows pharyngitis; patients have a facial flush, a "strawberry" tongue, and accentuated petechiae in body folds (Pastia's lines). *Kawasaki disease* (Chaps. 53 and 326) presents in the pediatric population as fissuring of the lips, a strawberry tongue, conjunctivitis, adenopathy, and sometimes cardiac abnormalities. *Streptococcal toxic shock syndrome* (Chap. 136) manifests with hypotension, multiorgan failure, and, often, a severe group A streptococcal infection (e.g., necrotizing fasciitis). *Staphylococcal toxic shock syndrome* (Chap. 135) also presents with hypotension, and multiorgan failure, but usually only *S. aureus* colonization—not a severe *S. aureus* infection—is documented. *Staphylococcal scalded-skin syndrome* (Chap. 135) is seen primarily in children and in immunocompromised adults. Generalized erythema is often evident during the prodrome of fever and malaise; profound tenderness of the skin is distinctive. In the exfoliative stage, the skin can be induced to form bullae with light lateral pressure (Nikolsky's sign). In a mild form, a scarlatiniform eruption mimics scarlet fever, but the patient does not exhibit a strawberry tongue or circumoral pallor. In contrast to the staphylococcal scalded-skin syndrome, in which the cleavage plane is superficial in the epidermis, *toxic epidermal necrolysis* (Chap. 55), a maximal variant of Stevens-Johnson syndrome, involves sloughing of the entire epidermis, resulting in severe disease. *Exfoliative erythroderma syndrome* (Chaps. 52 and 55) is a serious reaction associated with systemic toxicity that is often due to eczema, psoriasis, a drug reaction, or mycosis fungoides. Drug-induced hypersensitivity syndrome (DIHS) due to antiepileptic and antibiotic agents (Chap. 55) initially appears similar to an exanthematous drug reaction but may progress to exfoliative erythroderma; it is accompanied by multiorgan failure and has an associated mortality rate of ~10%.

Vesiculobullous or Pustular Eruptions

Varicella (Chap. 180) is highly contagious, often occurring in winter or spring. At any point in time, within a given region of the body, varicella lesions are in different stages of development. In immunocompromised hosts, varicella vesicles may lack the characteristic erythematous base or may appear hemorrhagic. Lesions of *Pseudomonas* "hot-tub" folliculitis (Chap. 152) are also pruritic and may appear similar to those of varicella. However, hot-tub folliculitis generally occurs in outbreaks after bathing in hot tubs or swimming pools, and lesions occur in regions occluded by bathing suits. Lesions of *variola* (smallpox) (Chap. 221) also appear similar to those of varicella but are all at the same stage of development in a given region of the body. Variola lesions are most prominent on the face and extremities, while varicella lesions are most prominent on the trunk. Herpes simplex virus infection (Chap. 179) is characterized by hallmark grouped vesicles on an erythematous base. Primary herpes infection is accompanied by fever and toxicity, while recurrent disease is milder. *Rickettsialpox* (Chap. 174) is often documented in urban settings and is characterized by vesicles followed by pustules. It can be distinguished from varicella by an eschar at the site of the mouse-mite bite and the papule/plaque base of each vesicle. Acute generalized eruptive pustulosis (AGEP) should be considered in individuals who are acutely febrile and are taking new medications, especially anticonvulsant or antimicrobial agents (Chap. 55). Disseminated *Vibrio vulnificus* infection (Chap. 156) or *ecthyma gangrenosum* due to *Pseudomonas aeruginosa* (Chap. 152) should be considered in immunosuppressed individuals with sepsis and hemorrhagic bullae.

Urticaria-Like Eruptions

Individuals with classic urticaria ("hives") usually have a hypersensitivity reaction without associated fever. In the presence of fever, urticaria-like eruptions are usually due to *urticarial vasculitis*(Chap. 326). Unlike individual lesions of classic urticaria, which last up to 24 h, these lesions may last 3–5 days. Etiologies include serum sickness (often induced by drugs such as penicillins, sulfas, salicylates, or barbiturates), connective-tissue disease (e.g., systemic lupus erythematosus or Sjögren's syndrome), and infection (e.g., with hepatitis B virus, enteroviruses, or parasites). Malignancy, especially lymphoma, may be associated with fever and chronic urticaria (Chap. 53).

Nodular Eruptions

In immunocompromised hosts, nodular lesions often represent disseminated infection. Patients with disseminated *candidiasis* (often due to *Candida tropicalis*) may have a triad of fever, myalgias, and eruptive nodules (Chap. 203). Disseminated *cryptococcosis* lesions (Chap. 202) may resemble molluscum contagiosum (Chap. 183). Necrosis of nodules should raise the suspicion of *aspergillosis* (Chap. 204) or *mucormycosis* (Chap. 205). *Erythema nodosum* presents with exquisitely tender nodules on the lower extremities. *Sweet's syndrome* (Chap. 53) should be considered in individuals with multiple nodules and plaques, often so edematous that they give the appearance of vesicles or bullae. Sweet's syndrome may affect either healthy individuals or persons with lymphoproliferative disease.

Purpuric Eruptions

Acute meningococcemia (Chap. 143) classically presents in children as a petechial eruption, but initial lesions may appear as

blanchable macules or urticaria. RMSF should be considered in the differential diagnosis of acute meningococcemia. *Echovirus 9 infection* (Chap. 191) may mimic acute meningococcemia; patients should be treated as if they have bacterial sepsis because prompt differentiation of these conditions may be impossible. Large ecchymotic areas of *purpura fulminans* (Chaps. 143 and 271) reflect severe underlying disseminated intravascular coagulation, which may be due to infectious or noninfectious causes. The lesions of *chronic meningococcemia* (Chap. 143) may have a variety of morphologies, including petechial. Purpuric nodules may develop on the legs and resemble erythema nodosum but lack its exquisite tenderness. Lesions of *disseminated gonococcemia* (Chap. 144) are distinctive, sparse, countable hemorrhagic pustules, usually located near joints. The lesions of chronic meningococcemia and those of gonococcemia may be indistinguishable in terms of appearance and distribution. *Viral hemorrhagic fever* (Chaps. 196 and 197) should be considered in patients with an appropriate travel history and a petechial rash. *Thrombotic thrombocytopenic purpura* (Chaps. 53,106, and 115) and *hemolytic-uremic syndrome* (Chaps. 115,149, and 154) are closely related and are noninfectious causes of fever and petechiae. *Cutaneous small-vessel vasculitis* (*leukocytoclastic vasculitis*) typically manifests as palpable purpura and has a wide variety of causes (Chap. 53).

Eruptions with Ulcers or Eschars

The presence of an ulcer or eschar in the setting of a more widespread eruption can provide an important diagnostic clue. For example, the presence of an eschar may suggest the diagnosis of scrub typhus or rickettsialpox (Chap. 174) in the appropriate setting. In other illnesses (e.g., anthrax) (Chap. 221), an ulcer or eschar may be the only skin manifestation.

FURTHER READINGS

Cherry JD: Contemporary infectious exanthems. Clin Infect Dis 16:199, 1993 [PMID: 8443297]

———: Cutaneous manifestations of systemic infections, in *Textbook of Pediatric Infectious Diseases* , vol. 1, 4th ed, RD Feigin, JD Cherry (eds). Philadelphia, Saunders, 1998, pp 713–737

Eichenfield LF et al (eds): *Neonatal Dermatology*, 2nd ed. Philadelphia, Saunders, 2008

Levin S, Goodman LJ: An approach to acute fever and rash (AFR) in the adult. Curr Clin Top Infect Dis 15:19, 1995 [PMID: 7546368]

Paller AS, Mancini AJ (eds): *Hurwitz Clinical Pediatric Dermatology*, 3rd ed. Philadelphia, Elsevier Saunders, 2006

Schlossberg D: Fever and rash. Infect Dis Clin North Am 10:101, 1996 [PMID: 8698985]

Weber DJ et al: The acutely ill patient with fever and rash, in *Principles and Practice of Infectious Diseases* , vol 1, 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 791–807

Wenner HA: Virus diseases associated with cutaneous eruptions. Prog Med Virol 16:269, 1973 [PMID: 4356897]

Wolff K, Johnson Raj: *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology* , 6th ed. New York, McGraw-Hill, 2009

——— et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 7th ed. New York, McGraw-Hill, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF RASHES ASSOCIATED WITH FEVER

Given the extremely broad differential diagnosis, the presentation of a patient with fever and rash often poses a thorny diagnostic challenge for even the most astute and experienced clinician. Rapid narrowing of the differential by prompt recognition of a rash's key features can result in appropriate and sometimes life-saving therapy. This atlas presents high-quality images of a variety of rashes that have an infectious etiology and are commonly associated with fever.

Figure e7-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lacy reticular rash of **erythema infectiosum** (fifth disease) caused by parvovirus B19.

Figure e7-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Koplik's spots, which manifest as white or bluish lesions with an erythematous halo on the buccal mucosa, usually occur in the first two days of measles symptoms and may briefly overlap the measles exanthem. The presence of the erythematous halo differentiates Koplik's spots from Fordyce's spots (ectopic sebaceous glands), which occur in the mouths of healthy individuals. *(Source: Centers for Disease Control and Prevention.)*

Figure e7-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

In **measles,** discrete erythematous lesions become confluent on the face and neck over 2–3 days as the rash spreads downward to the trunk and arms, where lesions remain discrete. *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, p 788.)*

Figure e7-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

In **rubella,** an erythematous exanthem spreads from the hairline downward and clears as it spreads. *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Figure e7-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Exanthem subitum (roseola) occurs most commonly in young children. A diffuse maculopapular exanthem follows resolution of fever. *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Figure e7-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythematous macules and papules are apparent on the trunk and arm of this patient with **primary HIV infection.***(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005.)*

Figure e7-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

This **exanthematous, drug-induced eruption** consists of brightly erythematous macules and papules, some of which are confluent, distributed symmetrically on the trunk and extremities. Ampicillin caused this rash. *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005.)*

Figure e7-8

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema migrans is the early cutaneous manifestation of Lyme disease and is characterized by erythematous annular patches, often with a central erythematous papule at the tick-bite site. *(Courtesy of Yale Resident's Slide Collection; with permission.)*

Figure e7-9

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rose spots are evident as erythematous macules on the trunk of this patient with **typhoid fever.** *(Source: Centers for Disease Control and Prevention.)*

Figure e7-10

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Systemic lupus erythematosus showing prominent, scaly, malar erythema. Involvement of other sun-exposed sites is also common.

Figure e7-11

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute lupus erythematosus on the upper chest, with brightly erythematous and slightly edematous coalescent papules and plaques. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e7-12

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Discoid lupus erythematosus. Violaceous, hyperpigmented, atrophic plaques, often with evidence of follicular plugging (which may result in scarring), are characteristic of this cutaneous form of lupus. *(Courtesy of Marilynne McKay, MD; with permission.)*

Figure e7-13

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The rash of **Still's disease** typically exhibits evanescent, erythematous papules that appear at the height of fever on the trunk and proximal extremities. *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Figure e7-14

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Impetigo is a superficial group A streptococcal or *Staphylococcus aureus* infection consisting of honey-colored crusts and erythematous weeping erosions. Occasionally, bullous lesions may be seen. *(Courtesy of Mary Spraker, MD; with permission.)*

Figure e7-15

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erysipelas is a group A streptococcal infection of the superficial dermis and consists of well-demarcated, erythematous, edematous, warm plaques.

Figure e7-16

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Top: Petechial lesions of **Rocky Mountain spotted fever** on the lower legs and soles of a young, otherwise healthy patient. ***Bottom:*** Close-up of lesions from the same patient. *(Courtesy of Lindsey Baden, MD; with permission.)*

Figure e7-17

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary syphilis with a firm, nontender chancre.

Figure e7-18

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Secondary syphilis, demonstrating the papulosquamous truncal eruption.

Figure e7-19

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Secondary syphilis commonly affects the palms and soles with scaling, firm, red-brown papules.

Figure e7-20

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Condylomata lata are moist, somewhat verrucous intertriginous plaques seen in secondary syphilis.

Figure e7-21

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mucous patches on the tongue of a patient with **secondary syphilis.** *(Courtesy of Ron Roddy; with permission.)*

Figure e7-22

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Petechial lesions in a patient with **atypical measles.** *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Figure e7-23

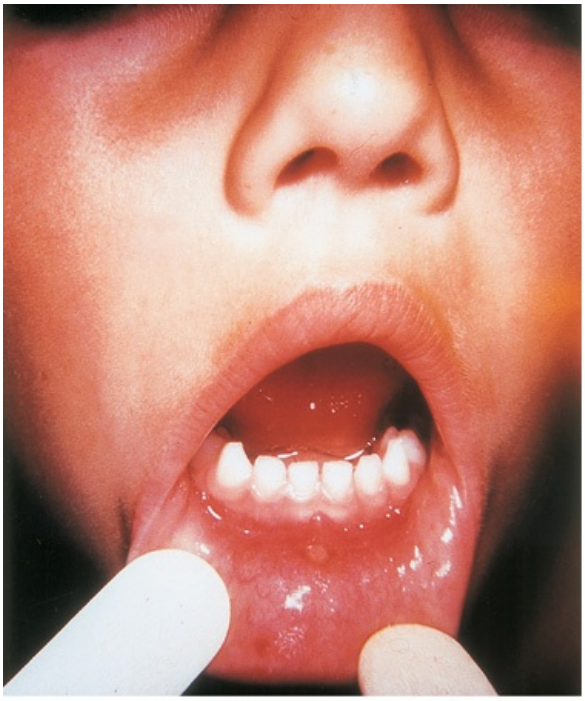

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tender vesicles and erosions in the mouth of a patient with **hand-foot-and-mouth disease.** *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Figure e7-24

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Septic emboli with hemorrhage and infarction due to acute *Staphylococcus aureus* **endocarditis.** *(Courtesy of Lindsey Baden, MD; with permission.)*

Figure e7-25

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema multiforme is characterized by multiple erythematous plaques with a target or iris morphology and usually represents a hypersensitive reaction to drugs or infections (especially herpes simplex virus). *(Courtesy of the Yale Resident's Slide Collection; with permission.)*

Figure e7-26

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scarlet fever exanthem. Finely punctuated erythema has become confluent (scarlatiniform); accentuation of linear erythema in body folds (Pastia's lines) is seen here. *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.)*

Figure e7-27

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema progressing to bullae with resulting sloughing of the entire thickness of the epidermis occurs in **toxic epidermal necrolysis.** This reaction was due to a sulfonamide. *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005.)*

Figure e7-28

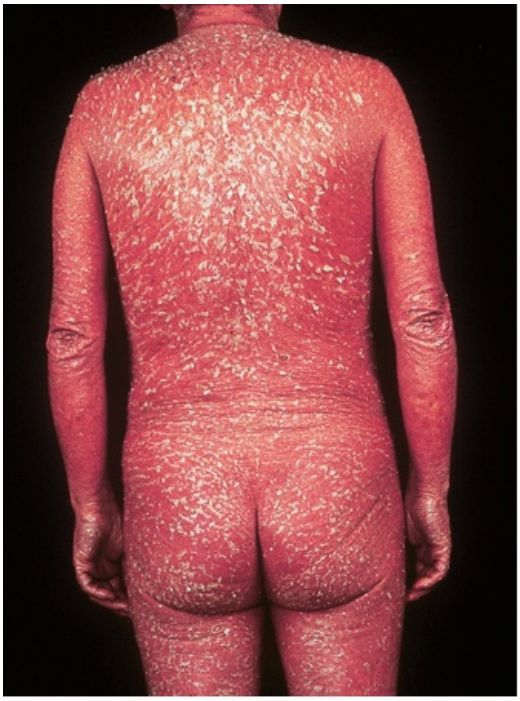

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diffuse erythema and scaling are present in this patient with psoriasis and the **exfoliative erythroderma syndrome.** *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.)*

Figure e7-29

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

This infant with **staphylococcal scalded skin syndrome** demonstrates generalized desquamation. *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009.)*

Figure e7-30

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fissuring of the lips and an erythematous exanthem are evident in this patient with **Kawasaki's disease.** *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Figure e7-31

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Numerous **varicella** lesions at various stages of evolution: vesicles on an erythematous base, umbilicated vesicles, and crusting lesions. *(Courtesy of R. Hartman; with permission.)*

Figure e7-32

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Close-up of lesions of **disseminated zoster.** Note lesions at different stages of evolution, including pustules and crusting. *(Courtesy of Lindsey Baden, MD; with permission.)*

Figure e7-33

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Herpes zoster is seen in this HIV-infected patient as hemorrhagic vesicles and pustules on an erythematous base grouped in a dermatomal distribution.

Figure e7-34

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Top: Eschar at the site of the mite bite in a patient with **rickettsialpox.** ***Middle:*** Papulovesicular lesions on the trunk of the same patient. ***Bottom:*** Close-up of lesions from the same patient. *(Reprinted from A Krusell et al: Emerg Infect Dis 8:727, 2002.)*

Figure e7-35

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ecthyma gangrenosum in a neutropenic patient with *Pseudomonas aeruginosa* bacteremia.

Figure e7-36

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Urticaria showing characteristic discrete and confluent, edematous, erythematous papules and plaques.

Figure e7-37

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Disseminated cryptococcal infection. A liver-transplant recipient developed six cutaneous lesions similar to the one shown. Biopsy and serum-antigen testing demonstrated *Cryptococcus.* Important features of the lesion include a benign-appearing fleshy papule with central umbilication resembling molluscum contagiosum. *(Courtesy of Lindsey Baden, MD; with permission.)*

Figure e7-38

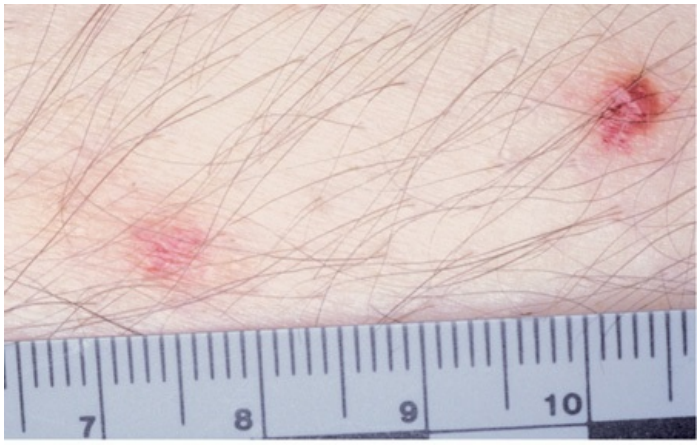

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Disseminated candidiasis. Tender, erythematous, nodular lesions developed in a neutropenic patient with leukemia who was undergoing induction chemotherapy. *(Courtesy of Lindsey Baden, MD; with permission.)*

Figure e7-39

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Disseminated *Aspergillus* infection. Multiple necrotic lesions developed in this neutropenic patient undergoing hematopoietic stem cell transplantation. The lesion in the photograph is on the inner thigh and is several centimeters in diameter. Biopsy demonstrated infarction caused by *Aspergillus fumigatus*. *(Courtesy of Lindsey Baden, MD; with permission.)*

Figure e7-40

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema nodosum is a panniculitis characterized by tender, deep-seated nodules and plaques usually located on the lower extremities. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e7-41

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sweet's syndrome: an erythematous indurated plaque with a pseudovesicular border. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e7-42

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fulminant meningococcemia with extensive angular purpuric patches. *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Figure e7-43

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythematous papular lesions are seen on the leg of this patient with **chronic meningococcemia.**

Figure e7-44

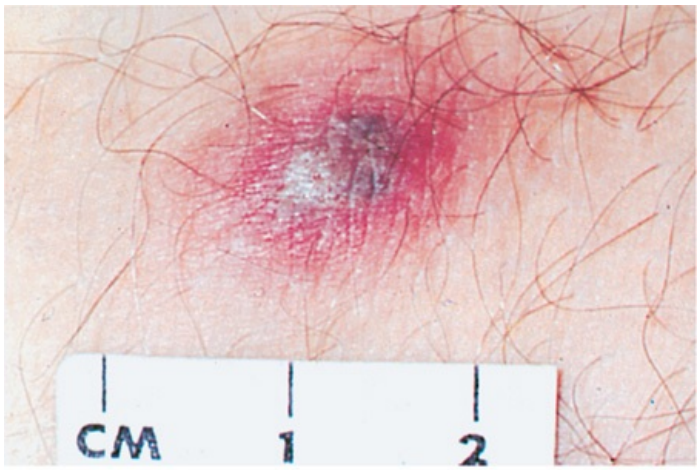

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Disseminated gonococcemia in the skin is seen as hemorrhagic papules and pustules with purpuric centers in a centrifugal distribution. *(Courtesy of Daniel M. Musher, MD; with permission.)*

Figure e7-45

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Palpable purpuric papules on the lower legs are seen in this patient with **cutaneous small-vessel vasculitis.** *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e7-46

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The thumb of a patient with a necrotic ulcer of **tularemia.** *(Source: Centers for Disease Control and Prevention.)*

Figure e7-47

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

This 50-year-old man developed high fever and massive inguinal lymphadenopathy after a small ulcer healed on his foot. **Tularemia** was diagnosed. *(Courtesy of Lindsey Baden, MD; with permission.)*

Figure e7-48

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

This painful **trypanosomal chancre** developed at the site of a tsetse-fly bite on the dorsum of the foot. *Trypanosoma brucei* was diagnosed from an aspirate of the ulcer. *(Courtesy of Edward T. Ryan, MD. N Engl J Med 346: 2069, 2002; with permission.)*

Figure e7-49

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Drug-induced hypersensitivity syndrome (DIHS/DRESS): This patient developed a progressive eruption exhibiting early desquamation after taking phenobarbital. There was also associated lymphadenopathy and hepatomegaly. *(Courtesy of Peter Lio, MD; with permission.)*

Figure e7-50

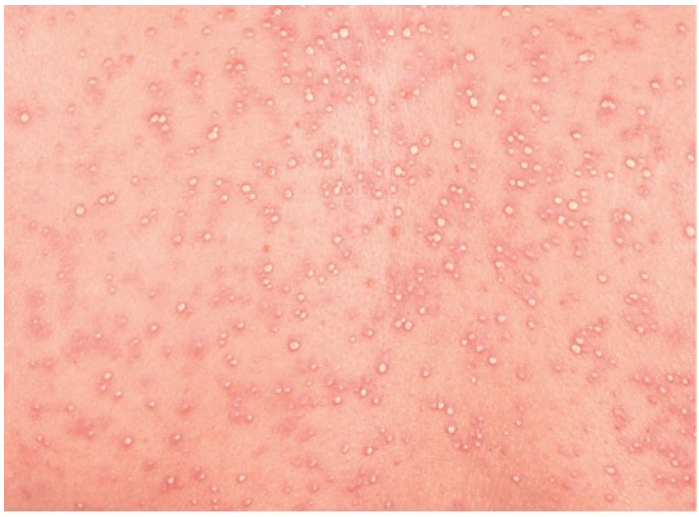

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Many small, nonfollicular pustules are seen against a background of erythema in this patient with **acute generalized eruptive pustulosis (AGEP)**. The rash began in body folds and progressed to cover the trunk and face. *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009, p. 561.)*

Figure e7-51

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Smallpox is shown with many pustules on the face, becoming confluent (A), and on the trunk (B). Pustules are all in the same stage of development. In (C), crusting, healing lesions are noted on the trunk, arms, and hands. *(Reprinted from K Wolff, RA Johnson: Color Atlas & Synopsis of Clinical Dermatology, 6th ed. New York, McGraw-Hill, 2009, p. 780.)*

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

FEVER OF UNKNOWN ORIGIN: INTRODUCTION

Definition and Classification

Fever of unknown origin (FUO) was defined by Petersdorf and Beeson in 1961 as (1) temperatures of >38.3°C (>101°F) on several occasions; (2) a duration of fever of >3 weeks; and (3) failure to reach a diagnosis despite 1 week of inpatient investigation. While this classification has stood for more than 30 years, Durack and Street have proposed a revised system for classification of FUO that better accounts for nonendemic and emerging diseases, improved diagnostic technologies, and adverse reactions to new therapeutic interventions. This updated classification includes (1) classic FUO, (2) nosocomial FUO, (3) neutropenic FUO, and (4) FUO associated with HIV infection.

Classic FUO corresponds closely to the earlier definition of FUO, differing only with regard to the prior requirement for 1 week's study in the hospital. The newer definition is broader, stipulating three outpatient visits or 3 days in the hospital without elucidation of a cause or 1 week of "intelligent and invasive" ambulatory investigation. In *nosocomial FUO*, a temperature of ≥38.3°C (≥101°F) develops on several occasions in a hospitalized patient who is receiving acute care and in whom infection was not manifest or incubating on admission. Three days of investigation, including at least 2 days' incubation of cultures, is the minimum requirement for this diagnosis. *Neutropenic FUO* is defined as a temperature of ≥38.3°C (≥101°F) on several occasions in a patient whose neutrophil count is <500/μL or is expected to fall to that level in 1–2 days. The diagnosis of neutropenic FUO is invoked if a specific cause is not identified after 3 days of investigation, including at least 2 days' incubation of cultures. *HIV-associated FUO* is defined by a temperature of ≥38.3°C (≥101°F) on several occasions over a period of >4 weeks for outpatients or >3 days for hospitalized patients with HIV infection. This diagnosis is invoked if appropriate investigation over 3 days, including 2 days' incubation of cultures, reveals no source.

Adoption of these categories of FUO in the literature has allowed a more rational compilation of data regarding these disparate groups. In the remainder of this chapter, the discussion will focus on classic FUO in the adult patient unless otherwise specified.

Causes of Classic FUO

Table 18-1 summarizes the findings of several large studies of FUO carried out since the advent of the antibiotic era, including a prospective study of 167 adult patients with FUO encompassing all eight university hospitals in the Netherlands and using a standardized protocol in which the first author reviewed every patient's case. Coincident with the widespread use of antibiotics, increasingly useful diagnostic technologies—both noninvasive and invasive—have been developed. Newer studies reflect not only changing patterns of disease but also the impact of diagnostic techniques that make it possible to eliminate many patients with specific illness from the FUO category. The ubiquitous use of potent broad-spectrum antibiotics may have decreased the number of infections causing FUO. The wide availability of ultrasonography, CT, MRI, radionuclide scanning, and positron emission tomography (PET) scanning has enhanced the detection of localized infections and of occult neoplasms and lymphomas in patients previously thought to have FUO. Likewise, the widespread availability of highly specific and sensitive immunologic testing has reduced the number of undetected cases of adult Still's disease, systemic lupus erythematosus, and polyarteritis nodosa.

Table 18-1 Classic FUO in Adults

Authors (Year of Publication)	Years of Study	No. of Cases	Infections (%)	Neoplasms (%)	Noninfectious Inflammatory Diseases (%)	Miscellaneous Causes (%)	Undiagnosed Causes (%)
Petersdorf and Beeson (1961)	1952–1957	100	36	19	19[a]	19[a]	7
Larson and Featherstone (1982)	1970–1980	105	30	31	16[a]	11[a]	12
Knockaert and Vanneste	1980–1989	199	22.5	7	23[a]	21.5[a]	25.5

(1992)						
de Kleijn et al. (1997, Part I)	1992–1994	167	26	12.5	24	8	30
Bleeker-Rovers et al. (2007)	2003–2005	73	16	7	22[b]	4	51

[a]Authors' raw data retabulated to conform to altered diagnostic categories.

[b]Connective tissue diseases.

Source: Modified from de Kleijn et al., 1997 (Part I).

Infections such as extrapulmonary tuberculosis and—in endemic areas—typhoid fever and malaria remain a leading diagnosable cause of FUO. Prolonged mononucleosis syndromes caused by Epstein-Barr virus, cytomegalovirus (CMV), or HIV are conditions whose consideration as a cause of FUO are sometimes confounded by delayed antibody responses. Intraabdominal abscesses (sometimes poorly localized) and renal, retroperitoneal, and paraspinal abscesses continue to be difficult to diagnose. Renal malacoplakia, with submucosal plaques or nodules involving the urinary tract, may cause fatal FUO if untreated; it is associated with intracellular bacterial infection, is seen in patients with defects of intracellular bacterial killing, and is treated with fluoroquinolones or trimethoprim-sulfamethoxazole. Occasionally, other organs may be involved. Osteomyelitis, especially where prosthetic devices have been implanted, must be considered. Although true culture-negative infective endocarditis is rare, one may be misled by cryptic endocarditis caused by indolent, slow-growing microorganisms of the HACEK group (*Haemophilus aphrophilus*, *Aggregatibacter (formerly Actinobacillus) actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella kingae*), *Bartonella* spp. (previously *Rochalimaea*), *Legionella* spp., *Coxiella burnetii*, *Chlamydophila psittaci*, and fungi. Prostatitis, dental abscesses, sinusitis, and cholangitis continue to be sources of occult fever.

Fungal diseases, most notably histoplasmosis involving the reticuloendothelial system, may cause FUO, particularly outside of the endemic regions where these diseases may be more readily recognized. FUO following travel to neotropical regions and the desert southwest of the United States, even for very limited periods, should prompt evaluation for paracoccidioidomycosis and coccidioidomycosis, respectively. The rising popularity of adventure travel among citizens of Western countries has increased the incidence in these nations of presentation for FUO due to otherwise uncommon endemic vector-borne infections, notably Chikungunya fever and scrub typhus. FUO with headache should prompt examination of spinal fluid for *Cryptococcus neoformans, Mycobacterium tuberculosis*, and travel-acquired trypanosomes. Malaria (which may result from transfusion, failure to take a prescribed prophylactic agent, or infection with a drug-resistant *Plasmodium* strain) continues to be a cause of FUO, particularly of the asynchronous variety. A related protozoan infection, babesiosis, may cause FUO and is increasing in geographic distribution and in incidence, especially among the elderly and the immunosuppressed.

In most earlier series, neoplasms were the next most common cause of FUO after infections (Table 18-1). In more recent series, a decrease in the percentage of FUO cases due to malignancy was attributed to improvement in diagnostic technologies—in particular, high-resolution tomography, MRI, PET scanning, and tumor antigen assays. This observation does not diminish the importance of considering neoplasia in the initial diagnostic evaluation of a patient with fever. A number of patients in these series had temporal arteritis, adult Still's disease, drug-related fever, and factitious fever. In recent series, ~25–50% of cases of FUO have remained undiagnosed. The general term *noninfectious inflammatory diseases* applies to systemic rheumatologic or vasculitic diseases such as polymyalgia rheumatica, lupus, and adult Still's disease as well as to granulomatous diseases such as sarcoidosis, Crohn's disease, and granulomatous hepatitis.

In the elderly, multisystem disease is the most frequent cause of FUO, giant-cell arteritis being the leading etiologic entity in this category. In patients >50 years of age, this disease accounts for 15–20% of FUO cases. Tuberculosis is the most common infection causing FUO in the elderly, and colon cancer is an important cause of FUO with malignancy in this age group.

Many diseases have been grouped in the various studies as "miscellaneous." On this list are drug fever, pulmonary embolism, factitious fever, the hereditary periodic fever syndromes [familial Mediterranean fever, hyper-IgD syndrome, tumor necrosis factor (TNF) receptor–associated periodic syndrome (also known as TRAPS or familial Hibernian fever), familial cold urticaria, and the Muckle-Wells syndrome)], and congenital lysosomal storage diseases such as Gaucher's and Fabry's disease.

A drug-related etiology must be considered in any case of prolonged fever. Any febrile pattern may be elicited by a drug. Virtually all classes of drugs can cause fever, but antimicrobial agents (especially β-lactam antibiotics), cardiovascular drugs (e.g., quinidine), antineoplastic drugs, and drugs acting on the

central nervous system (e.g., phenytoin) are particularly common causes. The use of TNF inhibitors for treatment of inflammatory diseases has led to atypical presentations of tuberculosis, histoplasmosis, coccidioidomycosis, and JC virus infection associated with FUO.

It is axiomatic that, as the duration of fever increases, the likelihood of an infectious cause decreases, even for the more indolent infectious etiologies (e.g., brucellosis, paracoccidioidomycosis, malaria due to *Plasmodium malariae*). In a series of 347 patients referred to the National Institutes of Health from 1961 to 1977, only 6% had an infection **(Table 18-2)**. A significant proportion (9%) had factitious fevers—i.e., fevers due either to false elevations of temperature or to self-induced disease. A substantial number of these factitious cases were in young women in the health professions. It is worth noting that 8% of the patients with prolonged fevers (some of whom had completely normal liver function studies) had granulomatous hepatitis, and 6% had adult Still's disease. After prolonged investigation, 19% of cases still had no specific diagnosis. A total of 27% of patients had no actual fever during inpatient observation or had an exaggerated circadian temperature rhythm without chills, elevated pulse, or other abnormalities.

Table 18-2 Causes of FUO Lasting >6 Months

Cause	Cases, %
None identified	19
Miscellaneous causes	13
Factitious causes	9
Granulomatous hepatitis	8
Neoplasm	7
Still's disease	6
Infection	6
Collagen vascular disease	4
Familial Mediterranean fever	3
No fever[a]	27

[a]No actual fever observed during 2–3 weeks of inpatient observation. Includes patients with exaggerated circadian rhythm.

Source: From a study of 347 patients referred to the National Institutes of Health from 1961–1977 with a presumptive diagnosis of FUO of >6 months' duration (Data from R Aduan et al: Prolonged fever of unknown origin. Clin Res 26:558A, 1978).

Global Considerations

More than 200 conditions may be considered in the differential diagnosis of classic FUO in adults; the most common of these are listed in **Table 18-3**. This list applies predominantly to Western nations such as the United States. The workup of FUO must take into careful consideration the patient's country of origin, recent and remote travel (including past service in foreign wars), unusual environmental exposures associated with travel or hobbies (e.g., caving, hunting, and safaris), and pets. The increasing number of returning sojourners with exotic travel itineraries underscores the need for a detailed history of travel and associated activities in the setting of undiagnosed fever, as do the changing demographics of the travelers themselves. For example, increasing numbers of travelers are immunosuppressed, are undergoing disease-modifying interventions such as TNF-α suppression, or have recently reconstituted immunity. Immigrants with unexplained fever, including naturalized citizens who have left their countries of origin decades previously, should be carefully interviewed with regard to childhood exposures, including immunization with nonstandard or unidentified live vaccines. In both foreign-born individuals and veterans of foreign wars, subclinical infections may be unmasked decades after exposure by new malignancies or immunosuppressive conditions. The differential diagnosis of FUO must also take into account changes in the range of arthropod vectors or the possibility that local permissive vectors have become infected with previously nonendemic pathogens. Evaluation of FUO in under-resourced medical settings requires increased reliance on history and clinical examination. Patients, family members, and close occupational contacts may need to be interviewed. If specialized laboratory and imaging studies cannot be conducted, diagnosis may be facilitated by maximizing the quality and precision of locally available approaches (e.g., culture of lysed, centrifuged blood cultures, and microscopic examination by an experienced technician). Emerging infectious diseases may include FUO first presenting as clusters of cases in remote regions; insight may be gained from contacting local epidemiologists.

Table 18-3 Causes of FUO in Adults in the United States

Infections

Localized pyogenic infections
- Appendicitis
- Cat-scratch disease
- Cholangitis
- Cholecystitis
- Dental abscess
- Diverticulitis/abscess
- Lesser sac abscess
- Liver abscess
- Mesenteric lymphadenitis
- Osteomyelitis
- Pancreatic abscess
- Pelvic inflammatory disease
- Perinephric/intrarenal abscess
- Prostatic abscess
- Renal malacoplakia
- Sinusitis
- Subphrenic abscess
- Suppurative thrombophlebitis
- Tuboovarian abscess

Intravascular infections
- Bacterial aortitis
- Bacterial endocarditis
- Vascular catheter infection

Systemic bacterial infections
- Bartonellosis
- Brucellosis
- *Campylobacter* infection
- Cat-scratch disease/bacillary angiomatosis (*B. henselae*)
- Gonococcemia
- Legionnaires' disease
- Leptospirosis
- Listeriosis
- Lyme disease
- Melioidosis
- Meningococcemia
- Rat-bite fever
- Relapsing fever
- Salmonellosis
- Syphilis
- Tularemia
- Typhoid fever
- Vibriosis
- *Yersinia* infection

Mycobacterial infections
- *M. avium/M. intracellulare* infections
- Other atypical mycobacterial infections
- Tuberculosis

Other bacterial infections

- Actinomycosis
- Bacillary angiomatosis
- Nocardiosis
- Whipple's disease

Rickettsial infections

- Anaplasmosis
- Ehrlichiosis
- Murine typhus
- Q fever
- Rickettsialpox
- Rocky Mountain spotted fever
- Scrub typhus

Mycoplasmal infections

Chlamydial infections

- Lymphogranuloma venereum
- Psittacosis
- TWAR (*C. pneumoniae*) infection

Viral infections

- Chikungunya fever
- Colorado tick fever
- Coxsackievirus group B infection
- Cytomegalovirus infection
- Dengue
- Epstein-Barr virus infection
- Hepatitis A, B, C, D, and E
- HIV infection
- Human herpesvirus 6 infection
- Lymphocytic choriomeningitis
- Parvovirus B19 infection
- Picornavirus infection

Fungal infections

- Aspergillosis
- Blastomycosis
- Candidiasis
- Coccidioidomycosis
- Cryptococcosis
- Histoplasmosis
- Mucormycosis

Paracoccidioidomycosis

- *Pneumocystis* infection
- Sporotrichosis

Parasitic infections

- Amebiasis
- Babesiosis
- Chagas' disease
- Leishmaniasis
- Malaria
- Strongyloidiasis
- Toxocariasis

Toxoplasmosis
Trichinellosis

Presumed infections, agent undetermined

Kawasaki's disease (mucocutaneous lymph node syndrome)
Kikuchi's necrotizing lymphadenitis

Neoplasms

Malignant

Colon cancer
Gall bladder carcinoma
Hepatoma
Hodgkin's lymphoma
Immunoblastic T-cell lymphoma
Leukemia
Lymphomatoid granulomatosis
Malignant histiocytosis
Non-Hodgkin's lymphoma
Pancreatic cancer
Renal cell carcinoma
Sarcoma

Benign

Atrial myxoma
Castleman's disease
Renal angiomyolipoma

Habitual Hyperthermia

(Exaggerated circadian rhythm)

Collagen Vascular/Hypersensitivity Diseases

Adult Still's disease
Behçet's disease
Erythema multiforme
Erythema nodosum
Giant-cell arteritis/polymyalgia rheumatica
Hypersensitivity pneumonitis
Hypersensitivity vasculitis
Mixed connective-tissue disease
Polyarteritis nodosa
Relapsing polychondritis
Rheumatic fever
Rheumatoid arthritis
Schnitzler's syndrome
Systemic lupus erythematosus
Takayasu's aortitis
Weber-Christian disease
Granulomatosis with polyangiitis (Wegener's)

Granulomatous Diseases

Crohn's disease
Granulomatous hepatitis
Midline granuloma
Sarcoidosis

Miscellaneous Conditions
Aortic dissection
Drug fever
Gout
Hematomas
Hemoglobinopathies
Laennec's cirrhosis
PFPA syndrome: periodic fever, adenitis, pharyngitis, aphthae
Postmyocardial infarction syndrome
Recurrent pulmonary emboli
Subacute thyroiditis (de Quervain's)
Tissue infarction/necrosis
Inherited and Metabolic Diseases
Adrenal insufficiency
Cyclic neutropenia
Deafness, urticaria, and amyloidosis
Fabry disease
Familial cold urticaria
Familial Mediterranean fever
Hyperimmunoglobulinemia D and periodic fever
Muckle-Wells syndrome
Tumor necrosis factor receptor–associated periodic syndrome (familial Hibernian fever)
Type V hypertriglyceridemia
Thermoregulatory Disorders
Central
Brain tumor
Cerebrovascular accident
Encephalitis
Hypothalamic dysfunction
Peripheral
Hyperthyroidism
Pheochromocytoma
Factitious Fevers
"Afebrile" FUO [<38.3°C (100.94°F)]

Source: Modified from RK Root, RG Petersdorf, in JD Wilson et al (eds): *Harrison's Principles of Internal Medicine*, 12th ed. New York, McGraw-Hill, 1991.

The possibility of international and domestic terrorist activity involving the intentional release of infectious agents, many of which cause illnesses presenting with prolonged fever, underscores the need for obtaining an insightful environmental, occupational, and professional history, with early notification of public health authorities in cases of suspicious etiology (Chap. 221). Moreover, the global spread of genetic engineering technologies raises the possibility that traditional agents —including Centers for Disease Control Categories A, B, and C agents; see Table 221-2)—that circumvent vaccine-acquired immunity could be developed or that novel recombinant organisms could be engineered to produce clinical or laboratory responses that defy current diagnostic approaches.

Specialized Diagnostic Studies

CLASSIC FUO

A stepwise flow chart depicting the diagnostic workup and therapeutic management of FUO is provided in **Fig. 18-1**. In this flow chart, reference is made to "potentially diagnostic clues," as outlined by de Kleijn and colleagues; these clues may be key findings in the history (e.g., travel), localizing signs, or key symptoms. Certain specific diagnostic maneuvers become critical in dealing with

prolonged fevers. If factitious fever is suspected, temperature-taking should be supervised, and simultaneous urine and body temperatures should be measured. Thick blood smears should be examined for *Plasmodium*; thin blood smears, prepared with proper technique and quality stains and subjected to expert microscopy, should be used to speciate *Plasmodium* and to identify *Babesia*, *Trypanosoma*, *Leishmania*, *Leptospira*, *Rickettsia*, and *Borrelia*. Specialized staining of mononuclear cells and granulocytes can help to identify intracellular bacteria, protozoal amastigotes, and the inclusion bodies of ehrlichiosis and anaplasmosis. Any tissue removed during prior relevant surgery should be reexamined; slides should be requested, and, if necessary, paraffin blocks of fixed pathologic material should be reexamined and additional special studies performed. Relevant x-rays should be reexamined; review of prior radiologic reports may be insufficient. Serum should be set aside in the laboratory as soon as possible and retained for future examination for rising antibody titers.

Figure 18-1

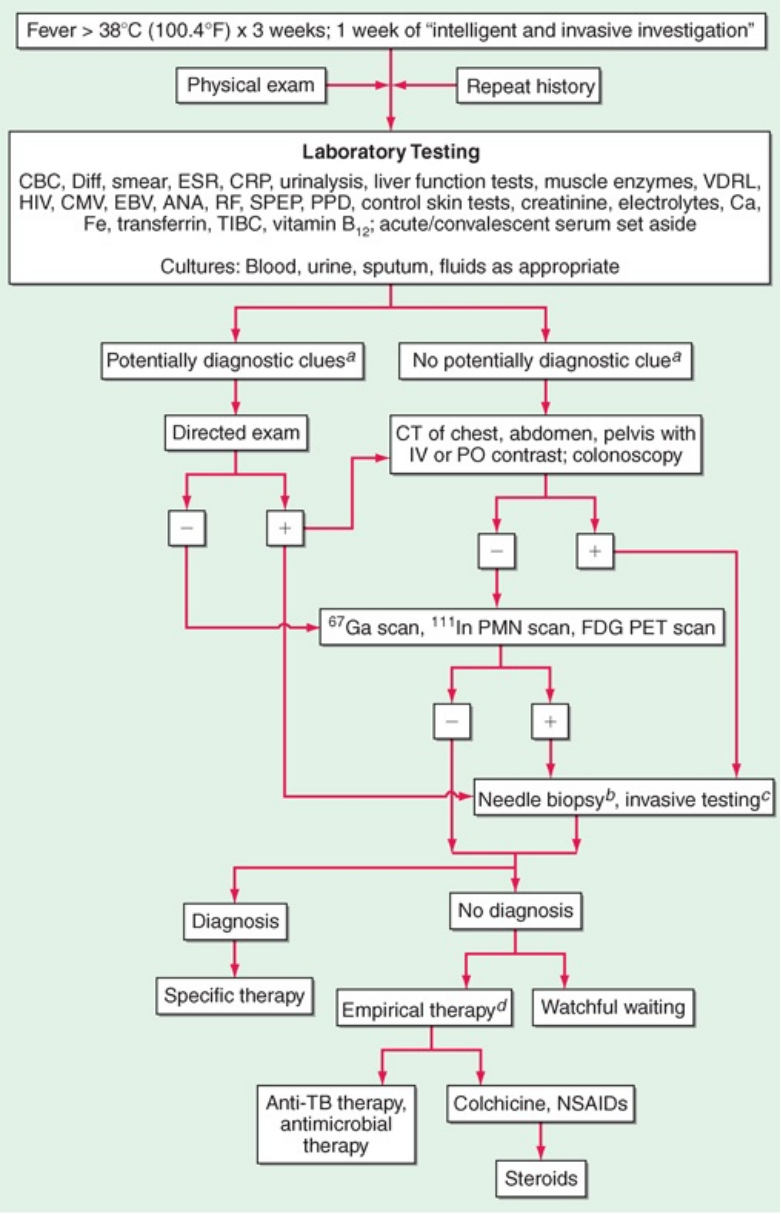

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the patient with classic FUO. [a] "Potentially diagnostic clues," as outlined by de Kleijn and colleagues (1997, Part II), may be key findings in the history, localizing signs, or key symptoms. [b] Needle biopsy of liver as well as any other tissue indicated by "potentially diagnostic clues." [c] Invasive testing could involve laparoscopy. [d] Empirical therapy is a last resort, given the good prognosis of most patients with FUO persisting without a diagnosis. *Abbreviations:* ANA, antinuclear antibody; CBC, complete blood count; CMV, cytomegalovirus; CRP, C-reactive protein; CT, computed tomography; Diff, differential; EBV, Epstein-Barr virus; ESR, erythrocyte sedimentation rate; FDG,

tomography; Dif, differential; EBV, Epstein-Barr virus; ESR, erythrocyte sedimentation rate; FDG, fluorodeoxyglucose F18; NSAIDs, nonsteroidal anti-inflammatory drugs; PET, positron emission tomography; PMN, polymorphonuclear leukocyte; PPD, purified protein derivative; RF, rheumatoid factor; SPEP, serum protein electrophoresis; TB, tuberculosis; TIBC, total iron-binding capacity; VDRL, Venereal Disease Research Laboratory test.

Febrile agglutinins is a vague term that, in most laboratories, refers to serologic studies for salmonellosis, brucellosis, and rickettsial diseases. These studies are seldom useful, having low sensitivity and variable specificity. Multiple blood samples (no fewer than three and rarely more than six, including samples for anaerobic culture) should be cultured in the laboratory—with and without increased CO_2—for 2 to 3 weeks to ensure ample growth time for any HACEK organisms (Chap. 146). It is critical to inform the laboratory of the intent to test for unusual organisms. Specialized media should be used if an exposure or travel history suggests uncommon causes of endocarditis, such as *Histoplasma*, *Chlamydophila*, *Mycoplasma*, *Bartonella*, *Coxiella*, or *Tropheryma whipplei*. Blood culture media should be supplemented with l-cysteine or pyridoxal to assist in the isolation of nutritionally variant streptococci. Lysis-centrifugation blood culture techniques should be employed when prior antimicrobial therapy or fungal or atypical mycobacterial infection is suspected. It should be noted that sequential cultures positive for multiple organisms may reflect self-injection of contaminated substances. Cultures of sinus fluid and pulmonary secretions on multiple permissive cell lines may prove helpful in identifying new respiratory viruses implicated in FUO. Urine cultures, including cultures for mycobacteria, fungi, and CMV, are indicated. In the setting of recurrent fevers with lymphocytic meningitis (Mollaret's meningitis), cerebrospinal fluid can be tested for herpesvirus, with use of the polymerase chain reaction (PCR) to amplify and detect viral nucleic acid (Chap. 179). A highly multiplexed oligonucleotide microarray using PCR amplification and containing probes for all recognized virus species hosted by vertebrates and up to 135 bacterial, 73 fungal, and 63 parasitic genera and species has been developed but has not yet been approved for clinical use. The continued clinical validation of such microarrays will further diminish rates of undiagnosed FUO of infectious etiology.

In any FUO workup, the erythrocyte sedimentation rate (ESR) should be determined. Striking elevation of the ESR and anemia of chronic disease are frequently seen in association with giant cell arteritis or polymyalgia rheumatica—common causes of FUO in patients >50 years of age. Still's disease is suggested by elevations of ESR, leukocytosis, and anemia and is often accompanied by arthralgias, polyserositis (pleuritis, pericarditis), lymphadenopathy, splenomegaly, and rash. The C-reactive protein level may be a useful cross-reference for the ESR and is a more sensitive and specific indicator of an "acute-phase" inflammatory metabolic response. Antinuclear antibody, antineutrophil cytoplasmic antibody, rheumatoid factor, and serum cryoglobulins should be measured to rule out other collagen vascular diseases and vasculitis. Elevated levels of angiotensin-converting enzyme in serum may point to sarcoidosis. With rare exceptions, the intermediate-strength purified protein derivative (PPD) skin test should be used to screen patients with classic FUO for tuberculosis. Concurrent control tests, such as the mumps skin test antigen (Aventis-Pasteur, Swiftwater, PA), should be employed. It should be kept in mind that both the PPD tuberculin skin test (TST) and control tests may yield false-negative results in patients with miliary tuberculosis, sarcoidosis, Hodgkin's disease, malnutrition, or AIDS. Two interferon γ–release assays have been approved by the U.S. Food and Drug Administration for the diagnosis of tuberculosis. These tests—the QuantiFERON-TB Gold In-Tube (QFT-GIT) assay and the T-SPOT TB assay—measure the production of interferon γ by T lymphocytes upon exposure to antigens of M. tuberculosis. In direct comparisons, the sensitivity of the QFT-GIT test was statistically similar to that of the TST for detecting infection in persons with untreated, culture-confirmed tuberculosis. The QFT-GIT test is more specific, is less influenced by previous infection with nontuberculous mycobacteria, and is not affected by prior vaccination with bacille Calmette-Guérin (BCG); TSTs are variably affected by these factors. Repeating the QFT-GIT test does not boost the in vitro response, while injection of PPD for the TST can boost subsequent TST responses, primarily in persons who have been infected with nontuberculous mycobacteria or vaccinated with BCG. Negative results in the QFT-GIT test—as in the TST—do not definitively exclude a diagnosis of tuberculosis.

Noninvasive procedures should include an upper gastrointestinal contrast study with small-bowel follow-through and colonoscopy to examine the terminal ileum and cecum for early evidence of lymphoma or subclinical Crohn's disease. Colonoscopy is especially strongly indicated in the elderly. Chest x-rays should be repeated if new symptoms arise. Sputum should be induced with an ultrasonic nebulizer for cultures, cytology, and molecular diagnostic testing. If there are pulmonary signs or symptoms, bronchoscopy with bronchoalveolar lavage for cultures, PCR, and cytology should be considered. High-resolution spiral CT of the chest and abdomen should be performed with both IV and oral contrast. If a spinal or paraspinal lesion is suspected, however, MRI is preferred. MRI may be superior to CT in demonstrating intraabdominal abscesses and aortic dissection, but the comparative utility of MRI and CT in the diagnosis of FUO is unknown. At present, abdominal CT with contrast should be used unless MRI is specifically indicated. Arteriography may be useful for patients in whom systemic necrotizing vasculitis is suspected. Saccular aneurysms may be seen, most commonly in renal or hepatic vessels, and may permit diagnosis of arteritis when biopsy is difficult. Ultrasonography of the abdomen is useful for investigation of the hepatobiliary tract, kidneys, spleen, and pelvis. Echocardiography may be helpful in an evaluation for bacterial endocarditis, pericarditis, nonbacterial

thrombotic endocarditis, and atrial myxomas. Transesophageal echocardiography is preferred for these lesions.

Radionuclide scanning procedures using technetium (Tc) 99m sulfur colloid, gallium (Ga) 67 citrate, or indium (In) 111–labeled leukocytes may be useful in identifying and/or localizing inflammatory processes such as aortitis or abscess. In one study, Ga scintigraphy yielded useful diagnostic information in almost one-third of cases, and it was suggested that this procedure might actually be used before other imaging techniques if no specific organ is suspected of being abnormal. It is likely that PET scanning, which provides quicker results (hours vs days), will prove even more sensitive and specific than ^{67}Ga scanning in FUO. ^{99m}Tc bone scan should be undertaken to look for osteomyelitis or bony metastases; ^{67}Ga scan may be used to identify sarcoidosis (Chap. 329) or *Pneumocystis* infection (Chap. 207) in the lungs or Crohn's disease (Chap. 295) in the abdomen. ^{111}In-labeled white blood cell (WBC) scan may be used to locate abscesses. With these scans, false-positive and false-negative findings are common. Fluorodeoxyglucose F18 (FDG) PET scanning appears to be superior to other forms of nuclear imaging. The FDG used in PET scans accumulates in tumors and at sites of inflammation and has even been shown to accumulate reliably at sites of vasculitis. Where available, FDG PET scanning should therefore be chosen over ^{67}Ga scanning in the diagnosis of FUO.

Biopsy of the liver and bone marrow should be considered in the workup of FUO if the studies mentioned above are unrevealing and if fever is prolonged. Granulomatous hepatitis has been diagnosed by liver biopsy, even when liver enzymes are normal and no other diagnostic clues point to liver disease. All biopsy specimens should be cultured for bacteria, mycobacteria, and fungi. Likewise, in the absence of clues pointing to the bone marrow, bone marrow biopsy (not simple aspiration) for histology and culture has yielded diagnoses late in the workup. When possible, a section of the tissue block should be retained for further sections or stains. At some research centers, PCR technology makes it possible in some cases to identify and speciate mycobacterial DNA in paraffin-embedded, fixed tissues. Thus, a retrospective diagnosis can sometimes be made on the basis of studies of long-fixed pathologic tissues. In a patient over age 50 (or occasionally in a younger patient) with the appropriate symptoms and laboratory findings, "blind biopsy" of one or both temporal arteries may yield a diagnosis of arteritis. Tenderness or decreased pulsation, if noted, should guide the selection of a site for biopsy. Lymph node biopsy may be helpful if nodes are enlarged, but inguinal nodes are often palpable and are seldom diagnostically useful.

Exploratory laparotomy has been performed when all other diagnostic procedures fail but has largely been replaced by imaging and guided-biopsy techniques. Peritoneal lavage may be used as a minimally invasive approach to peritoneal cytology studies. Laparoscopic biopsy may provide more adequate guided sampling of lymph nodes or liver, with less invasive morbidity.

NOSOCOMIAL FUO

(See also Chap. 131) The primary considerations in diagnosing nosocomial FUO are the underlying susceptibility of the patient coupled with the potential complications of hospitalization. The original surgical or procedural field is the place to begin a directed physical and laboratory examination for abscesses, hematomas, or infected foreign bodies. More than 50% of patients with nosocomial FUO are infected. Intravascular lines, septic phlebitis, and prostheses are all suspect. In this setting, the best approach is to focus on sites where occult infections may be sequestered, such as the sinuses of intubated patients or a prostatic abscess in a man with a urinary catheter. *Clostridium difficile* colitis may be associated with fever and leukocytosis before the onset of diarrhea. In ~25% of patients with nosocomial FUO, the fever has a noninfectious cause. Among these causes are acalculous cholecystitis, deep-vein thrombophlebitis, and pulmonary embolism. Drug fever, transfusion reactions, alcohol/drug withdrawal, adrenal insufficiency, thyroiditis, pancreatitis, gout, and pseudogout are among the many possible causes to consider. As in classic FUO, repeated meticulous physical examinations, coupled with focused diagnostic techniques, are imperative. Multiple blood, wound, and fluid cultures are mandatory. The pace of diagnostic tests is accelerated, and the threshold for procedures—CT scans, ultrasonography, ^{111}In WBC scans, noninvasive venous studies—is low. Even so, 20% of cases of nosocomial FUO may go undiagnosed.

Like diagnostic measures, therapeutic maneuvers must be swift and decisive, as many patients are already critically ill. IV lines must be changed (and cultured), drugs stopped for 72 hours, and empirical therapy started if bacteremia, fungemia, or persistently high virus loads are a threat. In many hospital settings, empirical antibiotic therapy for nosocomial FUO now includes vancomycin for coverage of methicillin-resistant *Staphylococcus aureus* as well as broad-spectrum gram-negative coverage with piperacillin/tazobactam, ticarcillin/clavulanate, imipenem, or meropenem. Practice guidelines covering many of these issues have been published jointly by the Infectious Diseases Society of America (IDSA) and the American College of Critical Care Medicine and can be accessed on the IDSA website (*www.journals.uchicago.edu/IDSA/guidelines*).

NEUTROPENIC FUO

(See also Chap. 86) Neutropenic patients are susceptible to focal bacterial and fungal infections, to

bacteremic infections, to infections involving catheters (including septic thrombophlebitis), and to perianal infections. *Candida* and *Aspergillus* infections are common. Infections due to herpes simplex virus or CMV are sometimes causes of FUO in this group. While the duration of illness may be short in these patients, the consequences of untreated infection may be catastrophic; 50–60% of febrile neutropenic patients are infected, and 20% are bacteremic. The IDSA has published extensive practice guidelines covering these critically ill neutropenic patients (*www.journals.uchicago.edu/IDSA/guidelines*). In these patients, severe mucositis, quinolone prophylaxis, colonization with methicillin-resistant *S. aureus*, obvious catheter-related infection, or hypotension dictates the use of vancomycin plus ceftazidime, cefepime, or a carbapenem with or without an aminoglycoside to provide empirical coverage for bacterial sepsis.

HIV-ASSOCIATED FUO

HIV infection alone may be a cause of fever. The infectious etiology varies with the extent of immunosuppression and the geographic region. Infection due to *Mycobacterium avium* or *M. intracellulare*, tuberculosis, toxoplasmosis, CMV infection, *Pneumocystis* infection, salmonellosis, cryptococcosis, histoplasmosis, strongyloidiasis, non-Hodgkin's lymphoma, and (of particular importance) drug fever are all possible causes of FUO. Mycobacterial infection can be diagnosed by blood cultures and by liver, bone marrow, and lymph node biopsies. Chest CT should be performed to identify enlarged mediastinal nodes. Serologic studies may reveal cryptococcal antigen, and ^{67}Ga scan may help identify *Pneumocystis* pulmonary infection. FUO has an infectious etiology in >80% of HIV-infected patients, but drug fever and lymphoma remain important considerations. Treatment of HIV-associated FUO depends on many factors and is discussed in Chap. 189.

Treatment: Fever of Unknown Origin

The focus here is on classic FUO. Other modifiers of FUO—neutropenia, HIV infection, a nosocomial setting—all vastly affect the risk equation and dictate therapy based on the probability of various causes of fever and on the calculated risks and benefits of a guided empirical approach. The age and physical state of the patient are factors as well: the frail, elderly patient may merit a trial of empirical therapy earlier than the robust young adult.

The emphasis in patients with classic FUO is on continued observation and examination, with the avoidance of "shotgun" empirical therapy. Antibiotic therapy (even that for tuberculosis) may irrevocably alter the ability to culture fastidious bacteria or mycobacteria and delineate ultimate cause. However, vital-sign instability or neutropenia is an indication for empirical therapy with a fluoroquinolone plus piperacillin or the regimen mentioned above (see "Nosocomial FUO"), for example. Cirrhosis, asplenia, disease-modifying biologic therapy, intercurrent immunosuppressive drug use, or exotic travel or environmental exposures (e.g., cave interiors) may all tip the balance toward earlier empirical anti-infective therapy. If the TST is positive or if granulomatous hepatitis or other granulomatous disease is present with anergy (and sarcoid seems unlikely), then a therapeutic trial for tuberculosis should be undertaken, with treatment usually continued for up to 6 weeks. A failure of the fever to respond over this period suggests an alternative diagnosis.

The response of rheumatic fever and Still's disease to aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs) may be dramatic. The effects of glucocorticoids on temporal arteritis, polymyalgia rheumatica, and granulomatous hepatitis are equally dramatic. Colchicine is highly effective in preventing attacks of familial Mediterranean fever but is of little use once an attack is well under way. The ability of glucocorticoids and NSAIDs to mask fever while permitting the spread of infection dictates that their use be avoided unless infection has been largely ruled out and unless inflammatory disease is both probable and debilitating or threatening.

When no underlying source of FUO is identified after prolonged observation (>6 months), the prognosis is generally good, however vexing the fever may be to the patient. Under such circumstances, debilitating symptoms are treated with NSAIDs, and glucocorticoids are the last resort. The initiation of empirical therapy does not mark the end of the diagnostic workup; rather, it commits the physician to continued thoughtful reexamination and evaluation. Patience, compassion, equanimity, vigilance, and intellectual flexibility are indispensable attributes for the clinician in dealing successfully with FUO.

ACKNOWLEDGMENTS

Sheldon M. Wolff, MD, now deceased, was an author of a previous version of this chapter. It is to his memory that the chapter is dedicated. The substantial contributions of Charles A. Dinarello, MD, to this chapter in previous editions are gratefully acknowledged.

FURTHER READINGS

Bleeker-Rovers CP et al: A prospective multicenter study on fever of unknown origin: The yield of a structured diagnostic protocol. Medicine 86:26, 2007 [PMID: 17220753]

——— et al: Fever of unknown origin. Semin Nucl Med 39:81, 2009

De Kleijn EM et al: Fever of unknown origin (FUO): I. A prospective multicenter study of 167 patients with FUO, using fixed epidemiologic entry criteria. Medicine 76:392, 1997

——— et al: Fever of unknown origin (FUO): II. Diagnostic procedures in a prospective multicenter study of 167 patients. Medicine 76:401, 1997

Goto M et al: A retrospective review of 226 hospitalized patients with fever. Intern Med 46:17, 2007 [PMID: 17202728]

High KP et al: IDSA guidelines: Clinical practice guideline for the evaluation of fever and infection in older adult residents of long-term care facilities: 2008 update by the Infectious Diseases Society of America. Clin Infect Dis 48:149, 2009 [PMID: 19072244]

Hot A et al: Yield of bone marrow examination in diagnosing the source of fever of unknown origin. Arch Intern Med 169:2018, 2009 [PMID: 19933965]

Knockaert DC et al: Fever of unknown origin in adults: 40 years on. J Intern Med 253:263, 2003 [PMID: 12603493]

Mourad O et al: A comprehensive evidence-based approach to fever of unknown origin. Arch Intern Med 163:545, 2003 [PMID: 12622601]

O'Grady NP et al: Guidelines for evaluation of new fever in critically ill adult patients: 2008 update from the American College of Critical Care Medicine and the Infectious Diseases Society of America. Crit Care Med 36:1330, 2008 [PMID: 21598908]

Simons KS et al: F-18-fluorodeoxyglucose positron emission tomography combined with CT in critically ill patients with suspected infection. Intensive Care Med 36:504, 2010 [PMID: 19847397]

Zenone T: Fever of unknown origin in adults: Evaluation of 144 cases in a non-university hospital. Scand J Infect Dis 38:632, 2006 [PMID: 16857607]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HYPOTHERMIA

Accidental hypothermia occurs when there is an unintentional drop in the body's core temperature below 35°C (95°F). At this temperature, many of the compensatory physiologic mechanisms that conserve heat begin to fail. *Primary accidental hypothermia* is a result of the direct exposure of a previously healthy individual to the cold. The mortality rate is much higher for patients who develop *secondary hypothermia* as a complication of a serious systemic disorder.

Causes

Primary accidental hypothermia is geographically and seasonally pervasive. Although most cases occur in the winter months and in colder climates, it is surprisingly common in warmer regions as well. Multiple variables make individuals at the extremes of age, the elderly and neonates, particularly vulnerable to hypothermia **(Table 19-1)**. The elderly have diminished thermal perception and are more susceptible to immobility, malnutrition, and systemic illnesses that interfere with heat generation or conservation. Dementia, psychiatric illness, and socioeconomic factors often compound these problems by impeding adequate measures to prevent hypothermia. Neonates have high rates of heat loss because of their increased surface-to-mass ratio and their lack of effective shivering and adaptive behavioral responses. At all ages, malnutrition can contribute to heat loss because of diminished subcutaneous fat and as a result of depleted energy stores used for thermogenesis.

Table 19-1 Risk Factors for Hypothermia

Age extremes	Endocrine-related
Elderly	Diabetes mellitus
Neonates	Hypoglycemia
Environmental exposure	Hypothyroidism
Occupational	Adrenal insufficiency
Sports-related	Hypopituitarism
Inadequate clothing	Neurologic-related
Immersion	Cerebrovascular accident
Toxicologic and pharmacologic	Hypothalamic disorders
Ethanol	Parkinson's disease
Phenothiazines	Spinal cord injury
Barbiturates	Multisystem
Carcinomatosis	Trauma
Anesthetics	Sepsis
Neuromuscular blockers	Shock
Antidepressants	Hepatic or renal failure
Insufficient fuel	Burns and exfoliative dermatologic disorders
Malnutrition	Immobility or debilitation
Marasmus	
Kwashiorkor	

Individuals whose occupations or hobbies entail extensive exposure to cold weather are at increased risk for hypothermia. Military history is replete with hypothermic tragedies. Hunters, sailors, skiers, and climbers also are at great risk of exposure, whether it involves injury, changes in weather, or lack of preparedness.

Ethanol causes vasodilation (which increases heat loss), reduces thermogenesis and gluconeogenesis, and may impair judgment or lead to obtundation. Phenothiazines, barbiturates, benzodiazepines, cyclic antidepressants, and many other medications reduce centrally mediated vasoconstriction. Up to 25% of patients admitted to an intensive care unit because of drug overdose are hypothermic. Anesthetics can block the shivering responses; their effects are compounded when patients are not insulated adequately in the operating or recovery rooms.

Several types of endocrine dysfunction can lead to hypothermia. Hypothyroidism—particularly when extreme, as in myxedema coma—reduces the metabolic rate and impairs thermogenesis and behavioral responses. Adrenal insufficiency and hypopituitarism also increase susceptibility to hypothermia. Hypoglycemia, most commonly caused by insulin or oral hypoglycemic drugs, is associated with hypothermia, in part a result of neuroglycopenic effects on hypothalamic function. Increased osmolality and metabolic derangements associated with uremia, diabetic ketoacidosis, and lactic acidosis can lead to altered hypothalamic thermoregulation.

Neurologic injury from trauma, cerebrovascular accident, subarachnoid hemorrhage, and hypothalamic lesion increases susceptibility to hypothermia. Agenesis of the corpus callosum, or Shapiro syndrome, is one cause of episodic hypothermia, characterized by profuse perspiration followed by a rapid fall in temperature. Acute spinal cord injury disrupts the autonomic pathways that lead to shivering and prevents cold-induced reflex vasoconstrictive responses.

Hypothermia associated with sepsis is a poor prognostic sign. Hepatic failure causes decreased glycogen stores and gluconeogenesis, as well as a diminished shivering response. In acute myocardial infarction associated with low cardiac output, hypothermia may be reversed after adequate resuscitation. With extensive burns, psoriasis, erythrodermas, and other skin diseases, increased peripheral blood flow leads to excessive heat loss.

Thermoregulation

Heat loss occurs through five mechanisms: radiation (55–65% of heat loss), conduction (10–15% of heat loss but much greater in cold water), convection (increased in the wind), respiration, and evaporation (which are affected by the ambient temperature and the relative humidity).

The preoptic anterior hypothalamus normally orchestrates thermoregulation (Chap. 16). The immediate defense of thermoneutrality is via the autonomic nervous system, whereas delayed control is mediated by the endocrine system. Autonomic nervous system responses include the release of norepinephrine, increased muscle tone, and shivering, leading to thermogenesis and an increase in the basal metabolic rate. Cutaneous cold thermoreception causes direct reflex vasoconstriction to conserve heat. Prolonged exposure to cold also stimulates the thyroid axis, leading to an increased metabolic rate.

Clinical Presentation

In most cases of hypothermia, the history of exposure to environmental factors, such as prolonged exposure to the outdoors without adequate clothing, makes the diagnosis straightforward. In urban settings, however, the presentation is often more subtle and other disease processes, toxin exposures, or psychiatric diagnoses should be considered.

After initial stimulation by hypothermia, there is progressive depression of all organ systems. The timing of the appearance of these clinical manifestations varies widely **(Table 19-2)**. Without knowing the core temperature, it can be difficult to interpret other vital signs. For example, a tachycardia disproportionate to the core temperature suggests secondary hypothermia resulting from hypoglycemia, hypovolemia, or a toxin overdose. Because carbon dioxide production declines progressively, the respiratory rate should be low; persistent hyperventilation suggests a central nervous system (CNS) lesion or one of the organic acidoses. A markedly depressed level of consciousness in a patient with mild hypothermia should raise suspicion of an overdose or CNS dysfunction due to infection or trauma.

Table 19-2 Physiologic Changes Associated with Accidental Hypothermia

Severity	Body Temperature	Central Nervous System	Cardiovascular	Respiratory	Renal and Endocrine	Neuromuscular
Mild	35°C (95°F) –32.2°C (90°F)	Linear depression of cerebral metabolism; amnesia; apathy; dysarthria; impaired judgment; maladaptive behavior	Tachycardia, then progressive bradycardia; cardiac cycle prolongation; vasoconstriction; increase in cardiac output and blood pressure	Tachypnea, then progressive decrease in respiratory minute volume; declining oxygen consumption; bronchorrhea; bronchospasm	Diuresis; increase in catecholamines, adrenal steroids, triiodothyronine and thyroxine; increase in metabolism with shivering	Increased preshivering muscle tone, then fatiguing
Moderate	<32.2°C (90°F)–28°C (82.4°F)	EEG abnormalities; progressive depression of level of consciousness; pupillary dilation; paradoxical undressing; hallucinations	Progressive decrease in pulse and cardiac output; increased atrial and ventricular arrhythmias; suggestive (J-wave) ECG changes	Hypoventilation; 50% decrease in carbon dioxide production per 8°C (46°F)drop in temperature; absence of protective airway reflexes	50% increase in renal blood flow; renal autoregulation intact; impaired insulin action	Hyporeflexia; diminishing shivering-induced thermogenesis; rigidity
Severe	<28°C (82.4°F)	Loss of cerebrovascular autoregulation; decline in cerebral blood flow; coma; loss of ocular reflexes; progressive decrease in EEG	Progressive decrease in blood pressure, heart rate, and cardiac output; re-entrant dysrhythmias; maximum risk of ventricular fibrillation; asystole	Pulmonic congestion and edema; 75% decrease in oxygen consumption; apnea	Decrease in renal blood flow parallels decrease in cardiac output; extreme oliguria; poikilothermia; 80% decrease in basal metabolism	No motion; decreased nerve-conduction velocity; peripheral areflexia; no corneal or oculocephalic reflexes

Source: Modified from DF Danzl, RS Pozos: N Engl J Med 331:1756, 1994.

Physical examination findings can also be altered by hypothermia. For instance, the assumption that areflexia is solely attributable to hypothermia can obscure and delay the diagnosis of a spinal cord lesion. Patients with hypothermia may be confused or combative; these symptoms abate more rapidly with rewarming than with chemical or physical restraint. A classic example of maladaptive behavior in patients with hypothermia is paradoxical undressing, which involves the inappropriate removal of clothing in response to a cold stress. The cold-induced ileus and abdominal rectus spasm can mimic, or mask, the presentation of an acute abdomen (Chap. 13).

When a patient in hypothermic cardiac arrest is first discovered, cardiopulmonary resuscitation is indicated unless (1) a do-not-resuscitate status is verified, (2) obviously lethal injuries are identified, or (3) the depression of a frozen chest wall is not possible. As the resuscitation proceeds, the prognosis is grave if there is evidence of widespread cell lysis, as reflected by potassium levels >10 mmol/L (10 meq/L). Other findings that may preclude continuing resuscitation include a core temperature <10–12°C (50–54°F), a pH <6.5, and evidence of intravascular thrombosis with a fibrinogen value <0.5 g/L (<50 mg/dL). The decision to terminate resuscitation before rewarming the patient past 33°C (91°F) should be predicated on the type and severity of the precipitants of hypothermia. There are no validated prognostic indicators for recovery from hypothermia. A history of asphyxia with secondary cooling is the most important negative predictor of survival.

Diagnosis and Stabilization

Hypothermia is confirmed by measuring the core temperature, preferably at two sites. Rectal probes should be placed to a depth of 15 cm and not adjacent to cold feces. A simultaneous esophageal probe should be placed 24 cm below the larynx; it may read falsely high during heated inhalation therapy. Relying solely on infrared tympanic thermography is not advisable.

After a diagnosis of hypothermia is established, cardiac monitoring should be instituted, along with attempts to limit further heat loss. If the patient is in ventricular fibrillation, one defibrillation attempt (2 J/kg) should be administered. If the rhythm does not convert, the patient should be rewarmed to 30°C (86°F) before defibrillation attempts are repeated. Supplemental oxygenation is always warranted, since tissue oxygenation is affected adversely by the leftward shift of the oxyhemoglobin dissociation curve. Pulse oximetry may be unreliable in patients with vasoconstriction. If protective airway reflexes are absent, gentle endotracheal intubation should be performed. Adequate preoxygenation will prevent ventricular arrhythmias. Although cardiac pacing for hypothermic bradydysrhythmias is rarely indicated, the transthoracic technique is preferable.

Insertion of a gastric tube prevents dilation secondary to decreased bowel motility. Indwelling bladder catheters facilitate monitoring of cold-induced diuresis. Dehydration is encountered commonly with chronic hypothermia, and most patients benefit from a bolus of crystalloid. Normal saline is preferable to lactated Ringer's solution, as the liver in hypothermic patients inefficiently metabolizes lactate. The placement of a pulmonary artery catheter can cause perforation of the less compliant pulmonary artery. Insertion of a central venous catheter into the cold right atrium should be avoided, since this can precipitate arrhythmias.

Arterial blood gases should not be corrected for temperature (Chap. 47). An uncorrected pH of 7.42 and a P_{CO2} of 40 mmHg reflect appropriate alveolar ventilation and acid-base balance at any core temperature. Acid-base imbalances should be corrected gradually, since the bicarbonate buffering system is inefficient. A common error is overzealous hyperventilation in the setting of depressed CO_2 production. When the P_{CO2} decreases 10 mmHg at 28°C (82°F), it doubles the pH increase of 0.08 that occurs at 37°C (99°F).

The severity of anemia may be underestimated because the hematocrit increases 2% for each 1°C drop in temperature. White blood cell sequestration and bone marrow suppression are common, potentially masking an infection. Although hypokalemia is more common in chronic hypothermia, hyperkalemia also occurs; the expected electrocardiographic changes can be obscured by hypothermia. Patients with renal insufficiency, metabolic acidoses, or rhabdomyolysis are at greatest risk for electrolyte disturbances.

Coagulopathies are common because cold inhibits the enzymatic reactions required for activation of the intrinsic cascade. In addition, thromboxane B_2 production by platelets is temperature-dependent, and platelet function is impaired. The administration of platelets and fresh frozen plasma is therefore not effective. The prothrombin or partial thromboplastin times or INR (international normalized ratio) can be deceptively normal and contrast with the observed in vivo coagulopathy. This contradiction occurs because all coagulation tests are routinely performed at 37°C (99°F), and the enzymes are thus rewarmed.

Rewarming Strategies

The key initial decision is whether to rewarm the patient passively or actively. *Passive external rewarming* simply involves covering and insulating the patient in a warm environment. With the head also covered, the rate of rewarming is usually 0.5° to 2°C (33° to 36°F) per hour. This technique is ideal for previously healthy patients who develop acute, mild primary accidental hypothermia. The patient must have sufficient glycogen to support endogenous thermogenesis.

The application of heat directly to the extremities of patients with chronic severe hypothermia should be avoided because it can induce peripheral vasodilation and precipitate core temperature "afterdrop," a response characterized by a continual decline in the core temperature after removal of the patient from the cold. Truncal heat application reduces the risk of afterdrop.

Active rewarming is necessary under the following circumstances: core temperature <32°C (90°F) (poikilothermia), cardiovascular instability, age extremes, CNS dysfunction, hormone insufficiency, and suspicion of secondary hypothermia. *Active external rewarming* is best accomplished with forced-air heating blankets. Other options include devices that circulate water through external heat exchange pads, radiant heat sources,

and hot packs. Monitoring a patient with hypothermia in a heated tub is extremely difficult. Electric blankets should be avoided because vasoconstricted skin is easily burned.

There are numerous widely available options for *active core rewarming*. Airway rewarming with heated humidified oxygen [40°–45°C (104°–113°F)] is a convenient option via mask or endotracheal tube. Although airway rewarming provides less heat than do some other forms of active core rewarming, it eliminates respiratory heat loss and adds 1°–2°C (34°–36°F) to the overall rewarming rate. Crystalloids should be heated to 40°–42°C (104°–108°F), but the quantity of heat provided is significant only during massive volume resuscitation. The most efficient method for heating and delivering fluid or blood is with a countercurrent in-line heat exchanger. Heated irrigation of the gastrointestinal tract or bladder transfers minimal heat because of the limited available surface area. These methods should be reserved for patients in cardiac arrest and then used in combination with all available active rewarming techniques. Closed thoracic lavage is far more efficient in severely hypothermic patients with cardiac arrest. The hemithoraxes are irrigated through two large-bore thoracostomy tubes that are inserted into the hemithoraxes. Thoracostomy tubes should not be placed in the left chest of a spontaneously perfusing patient for purposes of rewarming. Peritoneal lavage with the dialysate at 40°–45°C (104°–113°F) efficiently transfers heat when delivered through two catheters with outflow suction. Like peritoneal dialysis, standard hemodialysis is especially useful for patients with electrolyte abnormalities, rhabdomyolysis, or toxin ingestions. Another option involves the central venous insertion of a rapid endovascular warming device.

Extracorporeal blood rewarming options **(Table 19-3)** should be considered in severely hypothermic patients, especially those with *primary accidental hypothermia*. Cardiopulmonary bypass should be considered in nonperfusing patients without documented contraindications to resuscitation. Circulatory support may be the only effective option in patients with completely frozen extremities or those with significant tissue destruction coupled with rhabdomyolysis. There is no evidence that extremely rapid rewarming improves survival in perfusing patients. The best strategy is usually a combination of passive, truncal active, and active core rewarming techniques.

Table 19-3 Options for Extracorporeal Blood Rewarming

Extracorporeal Rewarming (ECR) Technique	Considerations
Continuous venovenous (CVV)	Circuit—CV catheter to CV or dual lumen CV or peripheral catheter
	No oxygenator/circulatory support
	Flow rates 150–400 mL/min
Hemodialysis (HD)	ROR 2°–3°C (36°–37°F)/h Circuit—single- or dual-vessel cannulation
	Stabilizes electrolyte or toxicologic abnormalities
	Exchange cycle volumes 200–500 mL/min
	ROR 2°–3°C (36°–37°F)/h
Continuous arteriovenous rewarming (CAVR)	Circuit—percutaneous 8.5 Fr femoral catheters
	Requires BP 60 mmHg systolic
	No perfusionist/pump/anticoagulation
	Flow rates 225–375 mL/min
	ROR 3°–4°C (37°–39°F)/h
Cardiopulmonary bypass (CPB)	Circuit—full circulatory support with pump and oxygenator
	Perfusate-temperature gradient [5°–10°C (41°–50°F)]
	Flow rates 2–7 L/min (ave. 3–4)
	ROR up to 9.5°C (49°F)/h

Abbreviations: BP, blood pressure; CV, central venous; ROR, rate of rewarming.

Treatment: Hypothermia

When a patient is hypothermic, target organs and the cardiovascular system respond minimally to most medications. Generally, IV medications are withheld below 30°C (86°F). If they are given, an increased interval between doses can prevent cumulative toxicity during rewarming because of increased binding of drugs to proteins as well as impaired metabolism and excretion. As an example, the administration of repeated doses of digoxin or insulin would be ineffective while the patient is hypothermic, and the residual drugs are potentially toxic during rewarming.

Achieving a mean arterial pressure of at least 60 mmHg should be an early objective. If the hypotension does not respond to crystalloid/colloid infusion and rewarming, low-dose dopamine (2–5 μg/kg per min) support should be considered. Perfusion of the vasoconstricted cardiovascular system also may be improved with low-dose IV nitroglycerin.

Atrial arrhythmias should be monitored initially without intervention, as the ventricular response will be slow, and unless preexistent, most will convert spontaneously during rewarming. The role of prophylaxis and treatment of ventricular arrhythmias is problematic. Preexisting ventricular ectopy may be suppressed by hypothermia and reappear during rewarming. None of the class I agents has proved to be safe and efficacious. There is also no

evidence that the class III ventricular antiarrhythmic amiodarone is safe. Initiating empirical therapy for adrenal insufficiency usually is not warranted unless there is a history suggesting steroid dependence or hypoadrenalism or a failure to rewarm with standard therapy. The administration of parenteral levothyroxine to euthyroid patients with hypothermia, however, is potentially hazardous. Because laboratory results can be delayed and confounded by the presence of the sick euthyroid syndrome (Chap. 341), historic clues or physical findings suggestive of hypothyroidism should be sought. When myxedema is the cause of hypothermia, the relaxation phase of the Achilles reflex is prolonged more than is the contraction phase.

Hypothermia obscures most of the symptoms and signs of infection, notably fever and leukocytosis. Shaking rigors from infection may be mistaken for shivering. Except in mild cases, extensive cultures and repeated physical examinations are essential. Unless an infectious source is identified, empirical antibiotic prophylaxis is most warranted in the elderly, neonates, and immunocompromised patients.

Preventive measures should be discussed with high-risk individuals, such as the elderly and people whose work frequently exposes them to extreme cold. The importance of layered clothing and headgear, adequate shelter, increased caloric intake, and the avoidance of ethanol should be emphasized, along with access to rescue services.

FROSTBITE

Peripheral cold injuries include both freezing and nonfreezing injuries to tissue. Tissue freezes quickly when in contact with thermal conductors such as metal and volatile solutions. Other predisposing factors include constrictive clothing or boots, immobility, and vasoconstrictive medications. Frostbite occurs when the tissue temperature drops below 0°C (32°F). Ice crystal formation subsequently distorts and destroys the cellular architecture. Once the vascular endothelium is damaged, stasis progresses rapidly to microvascular thrombosis. After the tissue thaws, there is progressive dermal ischemia. The microvasculature begins to collapse, arteriovenous shunting increases tissue pressures, and edema forms. Finally, thrombosis, ischemia, and superficial necrosis appear. The development of mummification and demarcation may take weeks to months.

Clinical Presentation

The initial presentation of frostbite can be deceptively benign. The symptoms always include a sensory deficiency affecting light touch, pain, and temperature perception. The acral areas and distal extremities are the most common insensate areas. Some patients complain of a clumsy or "chunk of wood" sensation in the extremity.

Deep frostbitten tissue can appear waxy, mottled, yellow, or violaceous-white. Favorable presenting signs include some warmth or sensation with normal color. The injury is often superficial if the subcutaneous tissue is pliable or if the dermis can be rolled over bony prominences.

Clinically, it is most practical to classify frostbite as superficial or deep. Superficial does not entail tissue loss. Superficial frostbite causes only anesthesia and erythema. The appearance of vesiculation surrounded by edema and erythema implies deeper involvement **(Fig. 19-1)**. Hemorrhagic vesicles reflect a serious injury to the microvasculature and indicate severe frostbite. Damages in subcuticular, muscular, or osseous tissues may result in amputation.

Figure 19-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Frostbite with vesiculation, surrounded by edema and erythema.

The two most common nonfreezing peripheral cold injuries are *chilblain (pernio)* and *immersion (trench) foot*. Chilblain results from neuronal and endothelial damage induced by repetitive exposure to dry cold. Young females, particularly those with a history of Raynaud's phenomenon, are at greatest risk. Persistent vasospasticity and vasculitis can cause erythema, mild edema, and pruritus. Eventually plaques, blue nodules, and ulcerations develop. These lesions typically involve the dorsa of the hands and feet. In contrast, immersion (trench) foot results from repetitive exposure to wet cold above the freezing point. The feet initially appear cyanotic, cold, and edematous. The subsequent development of bullae is often indistinguishable from frostbite. This vesiculation rapidly progresses to ulceration and liquefaction gangrene. Patients with milder cases complain of hyperhidrosis, cold sensitivity, and painful ambulation for many years.

Treatment: Frostbite

Frozen tissue should be thawed rapidly and completely by immersion in circulating water at 37°–40°C (99°–104°F). Rapid rewarming often produces an initial hyperemia. The early formation of large clear distal blebs is more favorable than that of smaller proximal dark hemorrhagic blebs. A common error is the premature termination of thawing, since the reestablishment of perfusion is intensely painful. Parenteral narcotics will be necessary with deep frostbite. If cyanosis persists after rewarming, the tissue compartment pressures should be monitored carefully.

Numerous experimental antithrombotic and vasodilatory treatment regimens have been evaluated. There is no conclusive evidence that dextran, heparin, steroids, calcium channel blockers, hyperbaric oxygen, or prostaglandin inhibitors salvage tissue. Intraarterial thrombolysis may reduce the need for digital and more proximal amputations when administered within 24 h of severe injuries. A treatment protocol for frostbite is summarized in **Table 19-4**.

Table 19-4 Treatment for Frostbite

Before Thawing	During Thawing	After Thawing
Remove from environment	Consider parenteral analgesia and ketorolac	Gently dry and protect part; elevate; pledgets between toes, if macerated
Prevent partial thawing and refreezing	Administer ibuprofen, 400 mg PO	If clear vesicles are intact, aspirate sterilely; if broken, debride and dress with antibiotic or sterile aloe vera ointment

Stabilize core temperature and treat hypothermia	Immerse part in 37°–40°C (99°–104°F) (thermometer-monitored) circulating water containing an antiseptic soap until distal flush (10–45 min)	Leave hemorrhagic vesicles intact to prevent dessication and infection
Protect frozen part—no friction or massage	Encourage patient to gently move part	Continue ibuprofen 400 mg PO (12 mg/kg per day) q8–12h
Address medical or surgical conditions	If pain is refractory, reduce water temperature to 35°–37°C (95°–99°F) and administer parenteral narcotics	Consider tetanus and streptococcal prophylaxis; elevate part
		Hydrotherapy at 37°C (99°F)
		Consider phenoxybenzamine or thrombolysis in severe cases

Unless infection develops, any decision regarding debridement or amputation should be deferred until there is clear evidence of demarcation, mummification, and sloughing. Magnetic resonance angiography may demonstrate the line of demarcation earlier than does clinical demarcation. The most common symptomatic sequelae reflect neuronal injury and the persistently abnormal sympathetic tone, including paresthesias, thermal misperception, and hyperhidrosis. Delayed findings include nail deformities, cutaneous carcinomas, and epiphyseal damage in children.

Management of the chilblain syndrome is usually supportive. With refractory perniosis, alternatives include nifedipine, steroids, and limaprost, a prostaglandin E1 analogue.

FURTHER READINGS

Alfonzo A et al: Survival after 5-h resuscitation attempt for hypothermic cardiac arrest using CVVG for extracorporeal rewarming. Nephrol Dial Transplant 24:1054, 2009 [PMID: 19059934]

Bruen KJ et al: Reduction of the incidence of amputation in frostbite injury with thrombolytic therapy. Arch Surg 142:546, 2007 [PMID: 17576891]

Danzl DF: Accidental hypothermia, in *Rosen's Emergency Medicine: Concepts and Clinical Practice,* 7th ed, J Marx et al (eds). Philadelphia, Mosby, 2010, p. 1861

Jurkovich GJ: Environmental cold-induced injury. Surg Clin North Am 87:247, 2007 [PMID: 17127131]

Kempainen RR, Brunette DD: The evaluation and management of accidental hypothermia. Respir Care 49:192, 2004 [PMID: 14744270]

Mulcahy AR, Watts MR: Accidental hypothermia: An evidence-based approach. Emerg Med Pract 11:1, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SYNCOPE: INTRODUCTION

Syncope is a transient, self-limited loss of consciousness due to acute global impairment of cerebral blood flow. The onset is rapid, duration brief, and recovery spontaneous and complete. Other causes of transient loss of consciousness need to be distinguished from syncope; these include seizures, vertebrobasilar ischemia, hypoxemia, and hypoglycemia. A syncopal prodrome (*presyncope*) is common, although loss of consciousness may occur without any warning symptoms. Typical presyncopal symptoms include dizziness, lightheadedness or faintness, weakness, fatigue, and visual and auditory disturbances. The causes of syncope can be divided into three general categories: (1) neurally mediated syncope (also called *reflex syncope*), (2) orthostatic hypotension, and (3) cardiac syncope.

Neurally mediated syncope comprises a heterogeneous group of functional disorders that are characterized by a transient change in the reflexes responsible for maintaining cardiovascular homeostasis. Episodic vasodilation and bradycardia occur in varying combinations, resulting in temporary failure of blood pressure control. In contrast, in patients with orthostatic hypotension due to autonomic failure, these cardiovascular homeostatic reflexes are chronically impaired. Cardiac syncope may be due to arrhythmias or structural cardiac diseases that cause a decrease in cardiac output. The clinical features, underlying pathophysiologic mechanisms, therapeutic interventions, and prognoses differ markedly among these three causes.

EPIDEMIOLOGY AND NATURAL HISTORY

Syncope is a common presenting problem, accounting for approximately 3% of all emergency room visits and 1% of all hospital admissions. The annual cost for syncope-related hospitalization in the United States is ~ $2 billion. Syncope has a lifetime cumulative incidence of up to 35% in the general population. The peak incidence in the young occurs between ages 10 and 30 years, with a median peak around 15 years. Neurally mediated syncope is the etiology in the vast majority of these cases. In elderly adults, there is a sharp rise in the incidence of syncope after 70 years.

In population-based studies, neurally mediated syncope is the most common cause of syncope. The incidence is slightly higher in females than males. In young subjects there is often a family history in first-degree relatives. Cardiovascular disease due to structural disease or arrhythmias is the next most common cause in most series, particularly in emergency room settings and in older patients. Orthostatic hypotension also increases in prevalence with age because of the reduced baroreflex responsiveness, decreased cardiac compliance, and attenuation of the vestibulosympathetic reflex associated with aging. In the elderly, orthostatic hypotension is substantially more common in institutionalized (54–68%) than community dwelling (6%) individuals, an observation most likely explained by the greater prevalence of predisposing neurologic disorders, physiologic impairment, and vasoactive medication use among institutionalized patients.

The prognosis after a single syncopal event for all age groups is generally benign. In particular, syncope of noncardiac and unexplained origin in younger individuals has an excellent prognosis; life expectancy is unaffected. By contrast, syncope due to a cardiac cause, either structural heart disease or primary arrhythmic disease, is associated with an increased risk of sudden cardiac death and mortality from other causes. Similarly, mortality rate is increased in individuals with syncope due to orthostatic hypotension related to age and the associated comorbid conditions (Table 20-1).

Table 20-1 High-Risk Features Indicating Hospitalization or Intensive Evaluation of Syncope

Chest pain suggesting coronary ischemia
Features of congestive heart failure
Moderate or severe valvular disease
Moderate or severe structural cardiac disease
Electrocardiographic features of ischemia
History of ventricular arrhythmias
Prolonged QT interval (>500 msec)
Repetitive sinoatrial block or sinus pauses
Persistent sinus bradycardia
Trifascicular block
Atrial fibrillation
Nonsustained ventricular tachycardia
Family history of sudden death
Preexcitation syndromes
Brugada pattern on ECG

PATHOPHYSIOLOGY

The upright posture imposes a unique physiologic stress upon humans; most, although not all, syncopal episodes occur from a standing position. Standing results in pooling of 500–1000 mL of blood in the lower extremities and splanchnic circulation. There is a decrease in venous return to the heart and reduced ventricular filling that result in diminished cardiac output and blood pressure. These hemodynamic changes provoke a compensatory reflex response, initiated by the baroreceptors in the carotid sinus and aortic arch, resulting in increased sympathetic outflow and decreased vagal nerve activity **(Fig. 20-1)**. The reflex increases peripheral resistance, venous return to the heart, and cardiac output and thus limits the fall in blood pressure. If this response fails, as is the case chronically in orthostatic hypotension and transiently in neurally mediated syncope, cerebral hypoperfusion occurs.

Figure 20-1

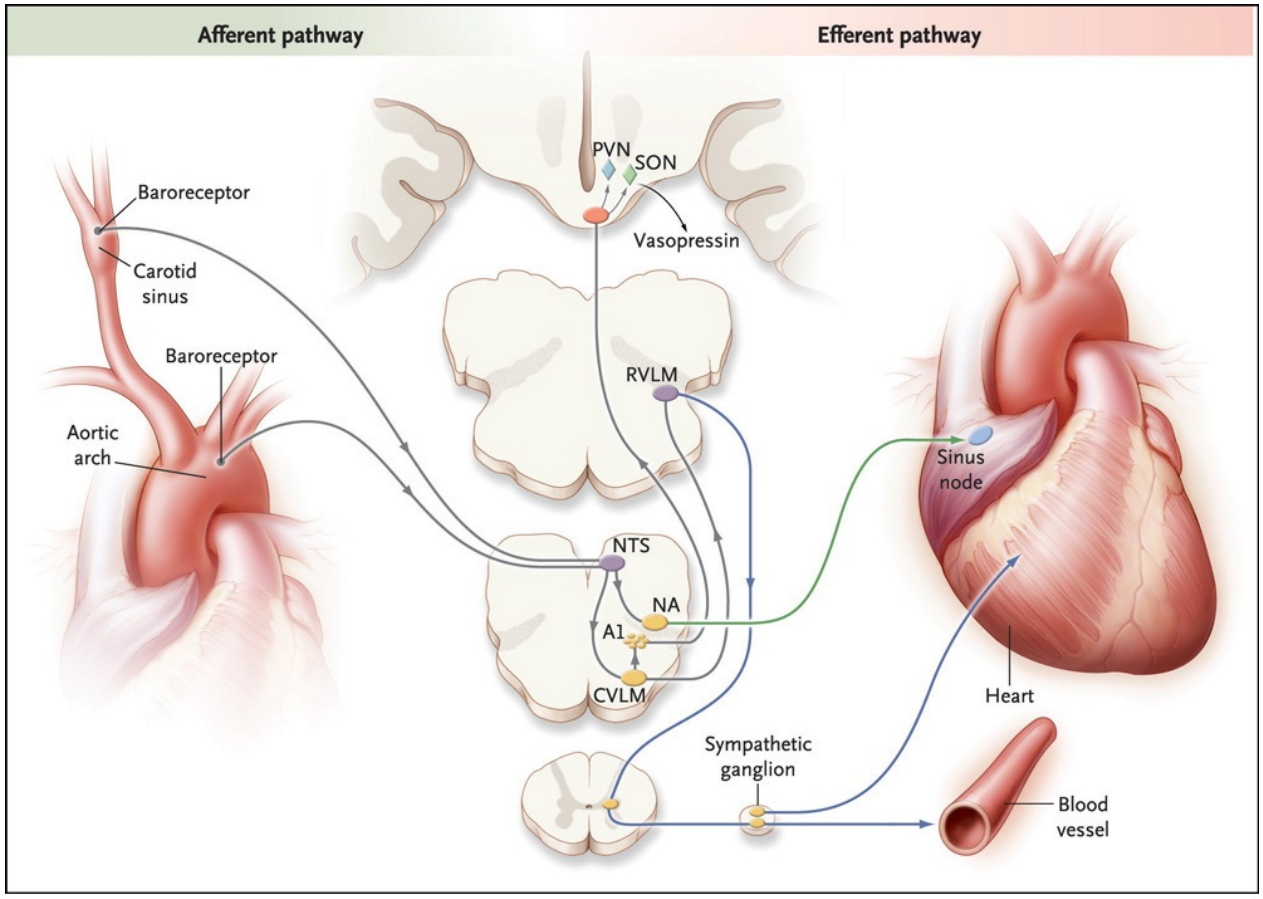

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The Baroreflex. A decrease in arterial pressure unloads the baroreceptors—the terminals of afferent fibers of the glossopharyngeal and vagus nerves—that are situated in the carotid sinus and aortic arch. This leads to a reduction in the afferent impulses that are relayed from these mechanoreceptors through the glossopharyngeal and vagus nerves to the nucleus of the tractus solitarius (NTS) in the dorsomedial medulla. The reduced baroreceptor afferent activity produces a decrease in vagal nerve input to the sinus node that is mediated by the neuroanatomical connections of the NTS to the nucleus ambiguus (NA). There is an increase in sympathetic efferent activity that is mediated by the NTS projections to the caudal ventrolateral medulla (CVLM) (an excitatory pathway) and from there to the rostral ventrolateral medulla (RVLM) (an inhibitory pathway). The activation of RVLM presympathetic neurons in response to hypotension is thus predominantly due to disinhibition. In response to a sustained fall in blood pressure, vasopressin release is mediated by projections from the A1 noradrenergic cell group in the ventrolateral medulla. This projection activates vasopressin-synthesizing neurons in the magnocellular portion of the paraventricular nucleus (PVN) and the supraoptic nucleus (SON) of the hypothalamus. Blue denotes sympathetic neurons and green parasympathetic neurons.) *(From R Freeman: N Engl J Med 358:615, 2008.)*

Syncope is a consequence of global cerebral hypoperfusion and thus represents a failure of cerebral blood flow autoregulatory mechanisms. Myogenic factors, local metabolites, and to a lesser extent autonomic neurovascular control are responsible for the autoregulation of cerebral blood flow (Chap. 275). Typically cerebral blood flow ranges from 50 to 60 mL/min per 100 g brain tissue and remains relatively constant over perfusion pressures ranging from 50 to 150 mmHg. Cessation of blood flow for 6–8 seconds will result in loss of consciousness, while impairment of consciousness ensues when blood flow decreases to 25 mL/min per 100 g brain tissue.

From the clinical standpoint, a fall in systemic systolic blood pressure to ~ 50 mmHg or lower will result in syncope. A decrease in cardiac output and/or systemic vascular resistance—the determinants of blood pressure—thus underlies the pathophysiology of syncope. Common causes of impaired cardiac output include decreased effective circulating blood volume; increased thoracic pressure; massive pulmonary embolus; cardiac brady- and tachyarrhythmias; valvular heart disease; and myocardial dysfunction. Systemic vascular resistance may be decreased by central and peripheral autonomic nervous system diseases, sympatholytic medications, and transiently during neurally mediated syncope. Increased cerebral vascular resistance, most frequently due to hypocarbia induced by hyperventilation, may also contribute to the pathophysiology of syncope.

The sequence of changes on the electroencephalogram of syncopal subjects during syncope comprises background slowing (often of high amplitude), followed by attenuation or cessation of cortical activity prior to return of slow waves, and then normal activity. Despite the presence of myoclonic movements and other motor activity, electroencephalographic seizure discharges are not present in syncopal subjects.

CLASSIFICATION

Neurally Mediated Syncope

Neurally mediated syncope is the final pathway of a complex central and peripheral nervous system reflex arc. There is a sudden, transient change in autonomic efferent activity characterized by increased parasympathetic outflow causing bradycardia and sympathoinhibition causing vasodilation. The change in autonomic efferent activity leads to a decrease in blood pressure and a subsequent fall in cerebral blood flow to below the limits of autoregulation **(Fig. 20-2)**. In order to elicit this reflex, a normal or functioning autonomic nervous system is necessary; this is in contrast to the situation in autonomic failure. The triggers of the afferent limb of the reflex arc vary and may be clearly defined, e.g., the carotid sinus, the gastrointestinal tract, or the bladder. In many cases, however, the afferent arc is less easily recognized and, under many circumstances, the cause is multifactorial. Under these circumstances it is likely that multiple afferent pathways converge on

the central autonomic network within the medulla that integrates the neural impulses and mediates the vasodepressor-bradycardic response.

Figure 20-2

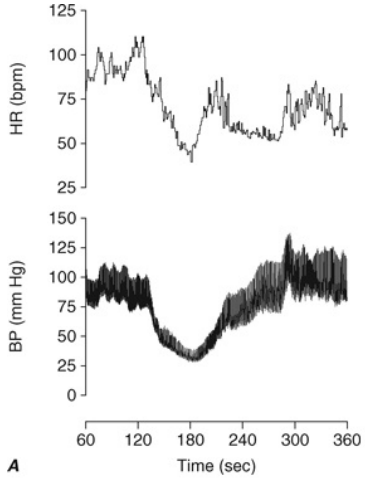

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

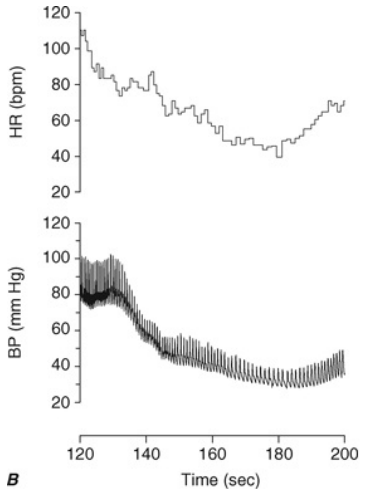

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***A*. The paroxysmal hypotensive-bradycardic response** that is characteristic of neurally mediated syncope. Noninvasive beat-to-beat blood pressure and heart rate are shown over 5 minutes (from 60 to 360 seconds) of an upright tilt on a tilt table. ***B*.** The same tracing expanded to show 80 seconds of the episode (from 80 to 200 seconds). BP, blood pressure; bpm, beats per minute; HR, heart rate.

CLASSIFICATION OF NEURALLY MEDIATED SYNCOPE

Neurally mediated syncope may be subdivided based on the afferent pathway and provocative trigger. Vasovagal syncope (the common faint) is provoked by intense emotion, pain, and/or orthostatic stress, whereas the situational reflex syncopes have specific localized stimuli that provoke the reflex vasodilation and bradycardia that leads to syncope. The underlying mechanisms have been identified and pathophysiology delineated for most of these situational reflex syncopes. The afferent trigger may originate in the pulmonary system, gastrointestinal system, urogenital system, heart, and carotid artery **(Table 20-2)**. Hyperventilation leading to hypocarbia and cerebral vasoconstriction, and raised intrathoracic pressure that impairs venous return to the heart, play a central role in many of the situational reflex syncopes. The afferent pathway of the reflex arc differs among these disorders but the efferent response via the vagus and sympathetic pathways is similar.

Table 20-2 Causes of Syncope

A. NEURALLY MEDIATED SYNCOPE
Vasovagal syncope
Provoked fear, pain, anxiety, intense emotion, sight of blood, unpleasant sights and odors, orthostatic stress
Situational reflex syncope
Pulmonary
Cough syncope, wind instrument player's syncope, weightlifter's syncope, "mess trick"[a] and "fainting lark,"[b] sneeze syncope, airway instrumentation
Urogenital
Postmicturition syncope, urogenital tract instrumentation, prostatic massage
Gastrointestinal

Swallow syncope, glossopharyngeal neuralgia, esophageal stimulation, gastrointestinal tract instrumentation, rectal examination, defecation syncope
Cardiac
Swallow syncope, glossopharyngeal neuralgia, esophageal stimulation, gastrointestinal tract instrumentation, rectal examination, defecation syncope
Carotid sinus
Carotid sinus sensitivity, carotid sinus massage
Ocular
Ocular pressure, ocular examination, ocular surgery
B. ORTHOSTATIC HYPOTENSION
Primary autonomic failure due to idiopathic central and peripheral neurodegenerative diseases—the "synucleinopathies"
Lewy body diseases
Parkinson's disease
Lewy body dementia
Pure autonomic failure
Multiple system atrophy (the Shy-Drager syndrome)
Secondary autonomic failure due to autonomic peripheral neuropathies
Diabetes
Hereditary amyloidosis (familial amyloid polyneuropathy)
Primary amyloidosis (AL amyloidosis; immunoglobulin light chain associated)
Hereditary sensory and autonomic neuropathies (HSAN) (especially type III—familial dysautonomia)
Idiopathic immune-mediated autonomic neuropathy
Autoimmune autonomic ganglionopathy
Sjögren's syndrome
Paraneoplastic autonomic neuropathy
HIV neuropathy
Postprandial hypotension
Iatrogenic (drug-induced)
Volume depletion
C. CARDIAC SYNCOPE
Arrhythmias
Sinus node dysfunction
Atrioventricular dysfunction
Supraventricular tachycardias
Ventricular tachycardias
Inherited channelopathies
Cardiac structural disease
Valvular disease
Myocardial ischemia
Obstructive and other cardiomyopathies
Atrial myxoma
Pericardial effusions and tamponade

[a]Hyperventilation for 1 minute, followed by sudden chest compression.

[b]Hyperventilation (20 breaths) in a squatting position, rapid rise to standing, then Valsalva.

Alternately, neurally mediated syncope may be subdivided based on the predominant efferent pathway. Vasodepressor syncope describes syncope predominantly due to efferent, sympathetic, vasoconstrictor failure; cardioinhibitory syncope describes syncope predominantly associated with bradycardia or asystole due to increased vagal outflow; while mixed syncope describes syncope in which there are both vagal and sympathetic reflex changes.

FEATURES OF NEURALLY MEDIATED SYNCOPE

In addition to symptoms of orthostatic intolerance such as dizziness, lightheadedness, and fatigue, premonitory features of autonomic activation may be present in patients with neurally mediated syncope. These include diaphoresis, pallor, palpitations, nausea, hyperventilation, and yawning. During the syncopal event proximal and distal myoclonus (typically arrhythmic and multifocal) may occur, raising the possibility of epilepsy. The eyes typically remain open and usually deviate upward. Urinary but not fecal incontinence may occur. Postictal confusion is rare, although visual and auditory hallucinations are sometimes reported.

While some predisposing factors and provocative stimuli are well established (for example, motionless upright posture, warm ambient temperature, intravascular volume depletion, alcohol ingestion, hypoxemia, anemia, pain, the sight of blood, venipuncture, and intense emotion), the underlying basis for the widely different thresholds for syncope among individuals exposed to the same provocative stimulus is not known. A genetic basis for neurally mediated syncope may exist; several studies have reported an increased incidence of syncope in first-degree relatives of fainters, but no gene or genetic marker has been identified, and environmental, social, and cultural factors have not been excluded by these studies.

TREATMENT: NEURALLY MEDIATED SYNCOPE

Reassurance, avoidance of provocative stimuli, and plasma volume expansion with fluid and salt are the cornerstones of the management of neurally mediated syncope. Isometric counterpressure maneuvers of the limbs (leg crossing or handgrip and arm tensing) may raise blood pressure and, by maintaining pressure in the autoregulatory zone, avoid or delay the onset of syncope. Randomized controlled trials support this intervention.

Fludrocortisone, vasoconstricting agents, and beta-adrenoreceptor antagonists are widely used by experts to treat refractory

patients, although there is no consistent evidence from randomized, controlled trials for any pharmacotherapy to treat neurally mediated syncope. Because vasodilation is the dominant pathophysiologic syncopal mechanism in most patients, use of a cardiac pacemaker is rarely beneficial. Possible exceptions are older patients in whom syncope is associated with asystole or severe bradycardia, and patients with prominent cardioinhibition due to carotid sinus syndrome. In these patients, dual-chamber pacing may be helpful.

Orthostatic Hypotension

Orthostatic hypotension, defined as a reduction in systolic blood pressure of at least 20 mmHg or diastolic blood pressure of at least 10 mmHg within 3 minutes of standing or head-up tilt on a tilt table, is a manifestation of sympathetic vasoconstrictor (autonomic) failure **(Fig. 20-3)**. In many (but not all) cases, there is no compensatory increase in heart rate despite hypotension; with partial autonomic failure, heart rate may increase to some degree but is insufficient to maintain cardiac output. A variant of orthostatic hypotension is "delayed" orthostatic hypotension which occurs beyond 3 minutes of standing; this may reflect a mild or early form of sympathetic adrenergic dysfunction. In some cases, orthostatic hypotension occurs within 15 seconds of standing (so-called "initial" orthostatic hypotension), a finding that may represent a transient mismatch between cardiac output and peripheral vascular resistance and does not represent autonomic failure.

Figure 20-3

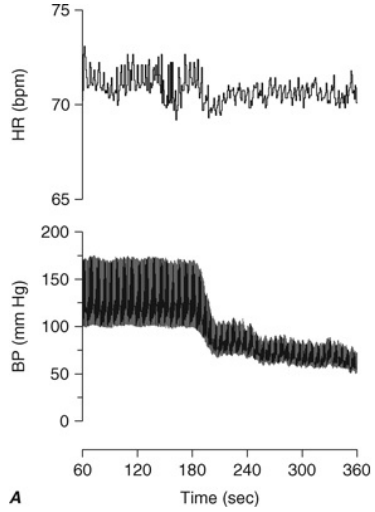

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

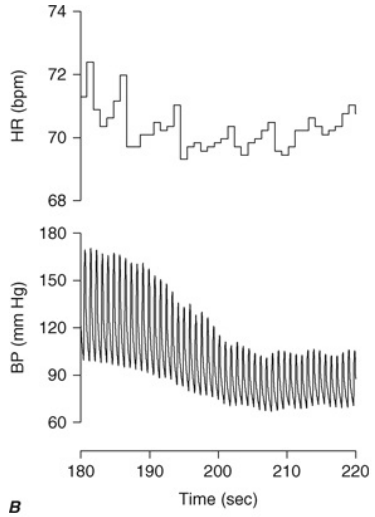

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A*. The gradual fall in blood pressure** without a compensatory heart rate increase that is characteristic of orthostatic hypotension due to autonomic failure. Blood pressure and heart rate are shown over 5 minutes (from 60 to 360 seconds) of an upright tilt on a tilt table. ***B. The same tracing expanded to show 40 seconds of the episode (from 180 to 220 seconds). BP, blood pressure; bpm, beats per minute; HR, heart rate.

Characteristic symptoms of orthostatic hypotension include light-headedness, dizziness, and presyncope (near-faintness) occurring in response to sudden postural change. However, symptoms may be absent or non-specific, such as generalized weakness, fatigue, cognitive slowing, leg buckling, or headache. Visual blurring may occur, likely due to retinal or occipital lobe ischemia. Neck pain—typically in the suboccipital, posterior cervical, and shoulder region (the "coat-hanger headache"), most likely due to neck muscle ischemia, may be the only symptom. Patients may report orthostatic dyspnea (thought to reflect ventilation-perfusion mismatch due to inadequate perfusion of ventilated lung apices) or angina (attributed to impaired myocardial perfusion even with normal coronary arteries). Symptoms may be exacerbated by exertion, prolonged standing, increased ambient temperature, or meals. Syncope is usually preceded by warning symptoms, but may occur suddenly, suggesting the possibility of a seizure or cardiac cause.

Supine hypertension is common in patients with orthostatic hypotension due to autonomic failure, affecting over 50% of patients in some series. Orthostatic hypotension may present after initiation of therapy for hypertension, and supine

hypertension may follow treatment of orthostatic hypotension. However, in other cases, the association of the two conditions is unrelated to therapy; it may in part be explained by baroreflex dysfunction in the presence of residual sympathetic outflow, particularly in patients with central autonomic degeneration.

CAUSES OF NEUROGENIC ORTHOSTATIC HYPOTENSION

Causes of neurogenic orthostatic hypotension include central and peripheral autonomic nervous system dysfunction (Chap. 375). Autonomic dysfunction of other organ systems (including the bladder, bowels, sexual organs, and sudomotor system) of varying severity frequently accompanies orthostatic hypotension in these disorders **(Table 20-2)**.

The primary autonomic degenerative disorders are multiple system atrophy (the Shy-Drager syndrome; Chap. 375), Parkinson's disease (Chap. 372), dementia with Lewy bodies (Chap. 371), and pure autonomic failure (Chap. 375). These are often grouped together as "synucleinopathies" due to the presence of alpha-synuclein, a small protein that precipitates predominantly in the cytoplasm of neurons in the Lewy body disorders (Parkinson's disease, dementia with Lewy bodies, and pure autonomic failure) and in the glia in multiple system atrophy.

Peripheral autonomic dysfunction may also accompany small fiber peripheral neuropathies such as those seen in diabetes, amyloid, immune-mediated neuropathies, hereditary sensory and autonomic neuropathies (HSAN; particularly HSAN type III; familial dysautonomia), and inflammatory neuropathies (Chaps. 385 and 386). Less frequently, orthostatic hypotension is associated with the peripheral neuropathies that accompany vitamin B_{12} deficiency, neurotoxic exposure, HIV and other infections, and porphyria.

Patients with autonomic failure and the elderly are susceptible to falls in blood pressure associated with meals. The magnitude of the blood pressure fall is exacerbated by large meals, meals high in carbohydrate, and alcohol intake. The mechanism of postprandial syncope is not fully elucidated.

Orthostatic hypotension is often iatrogenic. Drugs from several classes may lower peripheral resistance (e.g., alpha-adrenoreceptor antagonists used to treat hypertension and prostatic hypertrophy; antihypertensive agents of several classes; nitrates and other vasodilators; tricyclic agents and phenothiazines). Iatrogenic volume depletion due to diuresis and volume depletion due to medical causes (hemorrhage, vomiting, diarrhea, or decreased fluid intake) may also result in decreased effective circulatory volume, orthostatic hypotension, and syncope.

TREATMENT: ORTHOSTATIC HYPOTENSION

The first step is to remove reversible causes—usually vasoactive medications (Table 375-6). Next, nonpharmacologic interventions should be introduced. These interventions include patient education regarding staged moves from supine to upright; warnings about the hypotensive effects of meal ingestion; instructions about the isometric counterpressure maneuvers that increase intravascular pressure (see above); and raising the head of the bed to reduce supine hypertension. Intravascular volume should be expanded by increasing dietary fluid and salt. If these nonpharmacologic measures fail, pharmacologic intervention with fludrocortisone acetate and vasoconstricting agents such as midodrine and pseudoephedrine should be introduced. Some patients with intractable symptoms require additional therapy with supplementary agents that include pyridostigmine, yohimbine, desmopressin acetate (DDAVP), and erythropoietin (Chap. 375).

Cardiac Syncope

Cardiac (or cardiovascular) syncope is caused by arrhythmias and structural heart disease. These may occur in combination because structural disease renders the heart more vulnerable to abnormal electrical activity.

ARRHYTHMIAS

Bradyarrhythmias that cause syncope include those due to severe sinus node dysfunction (e.g., sinus arrest or sinoatrial block) and atrioventricular block (e.g., Mobitz type II, high-grade, and complete AV block). The bradyarrhythmias due to sinus node dysfunction are often associated with an atrial tachyarrhythmia, a disorder known as the tachycardia-bradycardia syndrome. A prolonged pause following the termination of a tachycardic episode is a frequent cause of syncope in patients with the tachycardia-bradycardia syndrome. Medications of several classes may also cause bradyarrhythmias of sufficient severity to cause syncope. Syncope due to bradycardia or asystole is referred to as a Stokes-Adams attack.

Ventricular tachyarrhythmias frequently cause syncope. The likelihood of syncope with ventricular tachycardia is in part dependent on the ventricular rate; rates below 200 beats per minute are less likely to cause syncope. The compromised hemodynamic function during ventricular tachycardia is caused by ineffective ventricular contraction, reduced diastolic filling due to abbreviated filling periods, loss of atrioventricular synchrony, and concurrent myocardial ischemia.

Several disorders associated with cardiac electrophysiologic instability and arrhythmogenesis are due to mutations in ion channel subunit genes. These include the long QT syndrome, Brugada syndrome, and catecholaminergic polymorphic ventricular tachycardia. The long QT syndrome is a genetically heterogeneous disorder associated with prolonged cardiac repolarization and a predisposition to ventricular arrhythmias. Syncope and sudden death in patients with long QT syndrome result from a unique polymorphic ventricular tachycardia called torsades des pointes that degenerates into ventricular fibrillation. The long QT syndrome has been linked to genes encoding K^+ channel α-subunits, K^+ channel β-subunits, voltage-gated Na^+ channel, and a scaffolding protein, ankyrin B (ANK2). Brugada syndrome is characterized by idiopathic ventricular fibrillation in association with right ventricular electrocardiogram (ECG) abnormalities without structural heart disease. This disorder is also genetically heterogeneous, although it is most frequently linked to mutations in the Na+ channel α-subunit, SCN5A. Catecholaminergic polymorphic tachycardia is an inherited, genetically heterogeneous disorder associated with exercise-or stress-induced ventricular arrhythmias, syncope, or sudden death. Acquired QT interval prolongation, most commonly due to drugs, may also result in ventricular arrhythmias and syncope. **These disorders are discussed in detail in Chap. 233.**

STRUCTURAL DISEASE

Structural heart disease, (e.g., valvular disease, myocardial ischemia, hypertrophic and other cardiomyopathies, cardiac masses such as atrial myxoma, and pericardial effusions) may lead to syncope by compromising cardiac output. Structural disease may also contribute to other pathophysiologic mechanisms of syncope. For example, cardiac structural disease may predispose to arrhythmogenesis; aggressive treatment of cardiac failure with diuretics and/or vasodilators may lead to orthostatic hypotension; and inappropriate reflex vasodilation may occur with structural disorders such as aortic stenosis and hypertrophic cardiomyopathy, possibly provoked by increased ventricular contractility.

TREATMENT: CARDIAC SYNCOPE

Treatment of cardiac disease depends upon the underlying disorder. Therapies for arrhythmias include cardiac pacing for sinus node disease and AV block, and ablation, anti-arrhythmic drugs, and cardioverter-defibrillators for atrial and ventricular tachyarrhythmias. These disorders are best managed by physicians with specialized skills in this area.

APPROACH TO THE PATIENT

Syncope

Differential Diagnosis

Syncope is easily diagnosed when the characteristic features are present; however, several disorders with transient real or apparent loss of consciousness may create diagnostic confusion.

Generalized and partial seizures may be confused with syncope; however, there are a number of differentiating features. Whereas tonic-clonic movements are the hallmark of a generalized seizure, myoclonic and other movements also may occur in up to 90% of syncopal episodes. Myoclonic jerks associated with syncope may be multifocal or generalized. They are typically arrhythmic and of short duration (<30 s). Mild flexor and extensor posturing also may occur. Partial- or partial-complex seizures with secondary generalization are usually preceded by an aura, commonly an unpleasant smell; fear anxiety; abdominal discomfort or other visceral sensations. These phenomena should be differentiated from the premonitory features of syncope.

Autonomic manifestations of seizures (autonomic epilepsy) may provide a more difficult diagnostic challenge. Autonomic seizures have cardiovascular, gastrointestinal, pulmonary, urogenital, pupillary, and cutaneous manifestations that are similar to the premonitory features of syncope. Furthermore, the cardiovascular manifestations of autonomic epilepsy include clinically significant tachycardias and bradycardias that may be of sufficient magnitude to cause loss of consciousness. The presence of accompanying non-autonomic auras may help differentiate these episodes from syncope.

Loss of consciousness associated with a seizure usually lasts longer than 5 minutes and is associated with prolonged postictal drowsiness and disorientation, whereas reorientation occurs almost immediately after a syncopal event. Muscle aches may occur after both syncope and seizures, although they tend to last longer following a seizure. Seizures, unlike syncope, are rarely provoked by emotions or pain. Incontinence of urine may occur with both seizures and syncope; however, fecal incontinence does not occur with syncope.

Hypoglycemia may cause transient loss of consciousness, typically in individuals with Type 1 or Type 2 diabetes treated with insulin. The clinical features associated with impending or actual hypoglycemia include tremor, palpitations, anxiety, diaphoresis, hunger, and paresthesias. These symptoms are due to autonomic activation to counter the falling blood glucose. Hunger, in particular, is not a typical premonitory feature of syncope. Hypoglycemia also impairs neuronal function, leading to fatigue, weakness, dizziness, and cognitive and behavioral symptoms. Diagnostic difficulties may occur in individuals in strict glycemic control; repeated hypoglycemia impairs the counterregulatory response and leads to a loss of the characteristic warning symptoms that are the hallmark of hypoglycemia.

Patients with cataplexy experience an abrupt partial or complete loss of muscular tone triggered by strong emotions, typically anger or laughter. Unlike syncope, consciousness is maintained throughout the attacks, which typically last between 30 seconds and 2 minutes. There are no premonitory symptoms. Cataplexy occurs in 60–75% of patients with narcolepsy.

The clinical interview and interrogation of eyewitnesses usually allows differentiation of syncope from falls due to vestibular dysfunction, cerebellar disease, extrapyramidal system dysfunction, and other gait disorders. If the fall is accompanied by head trauma, a postconcussive syndrome, amnesia for the precipitating events, and/or the presence of loss of consciousness may contribute to diagnostic difficulty.

Apparent loss of consciousness can be a manifestation of psychiatric disorders such as generalized anxiety, panic disorders, major depression, and somatization disorder. These possibilities should be considered in individuals who faint frequently without prodromal symptoms. Such patients are rarely injured despite numerous falls. There are no clinically significant hemodynamic changes concurrent with these episodes. In contrast, transient loss of consciousness due to vasovagal syncope precipitated by fear, stress, anxiety, and emotional distress is accompanied by hypotension, bradycardia, or both.

Initial Evaluation

The goals of the initial evaluation are to determine whether the transient loss of consciousness was due to syncope; to identify the cause; and to assess risk for future episodes and serious harm (Table 20-1). The initial evaluation should include a detailed history, thorough questioning of eyewitnesses, and a complete physical and neurologic examination. Blood pressure and heart rate should be measured in the supine position and after 3 minutes of standing to determine whether orthostatic hypotension is present. An ECG should be performed if there is suspicion of syncope due to an arrhythmia or underlying cardiac disease. Relevant electrocardiographic abnormalities include bradyarrhythmias or tachyarrhythmias, atrioventricular block, ischemia, old myocardial infarction, long QT syndrome, and bundle branch block. This initial assessment will lead to the identification of a cause of syncope in approximately 50% of patients and also allows stratification of patients at risk for cardiac mortality.

Laboratory Tests

Baseline laboratory blood tests are rarely helpful in identifying the cause of syncope. Blood tests should be performed when specific disorders, e.g., myocardial infarction, anemia, and secondary autonomic failure are suspected (Table 20-2).

Autonomic Nervous System Testing (Chap. 375)

Autonomic testing including tilt table testing can be performed in specialized centers. Autonomic testing is helpful to uncover objective evidence of autonomic failure and also to demonstrate a predisposition to neurally mediated syncope. Autonomic testing includes assessments of parasympathetic autonomic nervous system function (e.g., heart rate variability to deep respiration and a Valsalva maneuver), sympathetic cholinergic function (e.g., thermoregulatory sweat response and quantitative sudomotor axon reflex test), and sympathetic adrenergic function (e.g., blood pressure response to a Valsalva maneuver and a tilt table test with beat-to-beat blood pressure measurement). The hemodynamic abnormalities demonstrated on tilt table test (Figs. 20-2 and 20-3) may be useful in distinguishing orthostatic hypotension due to autonomic failure from the hypotensive bradycardic response of neurally mediated syncope. Similarly, the tilt table test may help identify patients with syncope due to delayed or initial orthostatic hypotension.

Carotid sinus massage should be considered in patients with symptoms suggestive of carotid sinus syncope and in patients over age 50 years with recurrent syncope of unknown etiology. This test should only be carried out under continuous ECG and blood pressure monitoring and should be avoided in patients with carotid bruits, plaques, or stenosis.

Cardiac Evaluation

ECG monitoring is indicated for patients with a high pretest probability of arrhythmia causing syncope. Patients should be monitored in hospital if the likelihood of a life-threatening arrhythmia is high, e.g., patients with severe structural or coronary artery disease, nonsustained ventricular tachycardia, trifascicular heart block, prolonged QT interval, Brugada's syndrome ECG pattern, and family history of sudden cardiac death. Outpatient Holter monitoring is recommended for patients who experience frequent syncopal episodes (one or more per week), whereas loop recorders, which continually record and erase cardiac rhythm, are indicated for patients with suspected arrhythmias with low risk of sudden cardiac death. Loop recorders may be external (recommended for evaluation of episodes that occur at a frequency of greater than one per month) or implantable (if syncope occurs less frequently).

Echocardiography should be performed in patients with a history of cardiac disease or if abnormalities are found on physical

examination or the electrocardiogram. Echocardiographic diagnoses that may be responsible for syncope include aortic stenosis, hypertrophic cardiomyopathy, cardiac tumors, aortic dissection, and pericardial tamponade. Echocardiography also has a role in risk stratification based on the left ventricular ejection fraction.

Treadmill exercise testing with ECG and blood pressure monitoring should be performed in patients who have experienced syncope during or shortly after exercise. Treadmill testing may help identify exercise-induced arrhythmias (e.g., tachycardia-related AV block) and exercise-induced exaggerated vasodilation.

Electrophysiologic studies are indicated in patients with structural heart disease and ECG abnormalities in whom noninvasive investigations have failed to yield a diagnosis. Electrophysiologic studies have low sensitivity and specificity and should only be performed when a high pretest probability exists. Currently, this test is rarely performed to evaluate patients with syncope.

Psychiatric Evaluation

Screening for psychiatric disorders may be appropriate in patients with recurrent unexplained syncope episodes. Tilt table testing, with demonstration of symptoms in the absence of hemodynamic change, may be useful in reproducing syncope in patients with suspected psychogenic syncope.

FURTHER READINGS

COLEMAN N et al: Syncope, in *Course and Treatment of Neurological Diseases,* 2nd ed, T Brandt et al (eds). San Diego Academic Press, 2003

FREEMAN R: Autonomic peripheral neuropathy. Lancet 365:1259, 2005 [PMID: 15811460]

———: Clinical practice. Neurogenic orthostatic hypotension. N Engl J Med 358:615, 2008

GANZEBOOM KS et al: Lifetime cumulative incidence of syncope in the general population: A study of 549 Dutch subjects aged 35–60 years. J Cardiovasc Electrophysiol 17:1172, 2006 [PMID: 17074006]

KAUFMANN H, BIAGGIONI I: Autonomic failure in neurodegenerative disorders. Semin Neurol 23:351, 2003 [PMID: 15088256]

LEMPERT T: Recognizing syncope: Pitfalls and surprises. J R Soc Med 89:372, 1996 [PMID: 8774533]

MOYA A et al: Guidelines for the diagnosis and management of syncope (version 2009): The Task Force for the Diagnosis and Management of Syncope of the European Society of Cardiology (ESC). Eur Heart J 30:2631, 2009 [PMID: 19713422]

SHAH M et al: Molecular basis of arrhythmias. Circulation 112:2517, 2005 [PMID: 16230503]

SULE S et al: Etiology of syncope in patients hospitalized with syncope and predictors of mortality and rehospitalization for syncope at 27-month follow-up. Clin Cardiol 34:35, 2011 [PMID: 21259276]

WIELING W et al: Symptoms and signs of syncope: A review of the link between physiology and clinical clues. Brain 132:2630, 2009 [PMID: 19587129]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIZZINESS AND VERTIGO: INTRODUCTION

Dizziness is a common, vexing symptom, and epidemiologic data indicate that more than 20% of adults experience dizziness within a given year. The diagnosis is frequently challenging, in part because patients use the term to refer to a variety of different sensations, including feelings of faintness, spinning, and other illusions of motion, imbalance, and anxiety. Other descriptive words, such as *light-headedness*, are equally ambiguous, referring in some cases to a presyncopal sensation due to hypoperfusion of the brain and in others to disequilibrium and imbalance. Patients often have difficulty distinguishing among these various symptoms, and the words they choose do not describe the underlying etiology reliably.

Vascular disorders cause presyncopal dizziness as a result of cardiac dysrhythmia, orthostatic hypotension, medication effects, or another cause. Such presyncopal sensations vary in duration; they may increase in severity until loss of consciousness occurs, or they may resolve before loss of consciousness if the cerebral ischemia is corrected. Faintness and syncope, which are discussed in detail in Chap. 20, should always be considered when one is evaluating patients with brief episodes of dizziness or dizziness that occurs with upright posture.

Vestibular causes of dizziness may be due to peripheral lesions that affect the labyrinths or vestibular nerves or to involvement of the central vestibular pathways. They may be paroxysmal or due to a fixed unilateral or bilateral vestibular deficit. Acute unilateral lesions cause vertigo due to a sudden imbalance in vestibular inputs from the two labyrinths. Bilateral lesions cause imbalance and instability of vision when the head moves (*oscillopsia*). Other causes of dizziness include nonvestibular imbalance and gait disorders (e.g., loss of proprioception from sensory neuropathy, parkinsonism) and anxiety.

In evaluating patients with dizziness, questions to consider include the following: (1) is it dangerous (e.g., arrhythmia, transient ischemic attack/stroke)? (2) is it vestibular? and (3) if vestibular, is it peripheral or central? A careful history and examination often provide enough information to answer these questions and determine whether additional studies or referral to a specialist is necessary.

Approach to the Patient: Dizziness

HISTORY

When a patient presents with dizziness, the first step is to delineate more precisely the nature of the symptom. In the case of vestibular disorders, the physical symptoms depend on whether the lesion is unilateral or bilateral and whether it is acute or chronic and progressive. Vertigo, an illusion of self or environmental motion, implies asymmetry of vestibular inputs from the two labyrinths or in their central pathways and is usually acute. Symmetric bilateral vestibular hypofunction causes imbalance but no vertigo. Because of the ambiguity in patients' descriptions of their symptoms, diagnosis based simply on symptom character is typically unreliable. The history should focus closely on other features, including whether dizziness is paroxysmal or has occurred only once, the duration of each episode, any provoking factors, and the symptoms that accompany the dizziness.

Causes of dizziness can be divided into episodes that last for seconds, minutes, hours, or days. Common causes of brief dizziness (seconds) include benign paroxysmal positional vertigo (BPPV) and orthostatic hypotension, both of which typically are provoked by changes in position. Attacks of migrainous vertigo and Ménière's disease often last hours. When episodes are of intermediate duration (minutes), transient ischemic attacks of the posterior circulation should be considered, although these episodes also could be due to migraine or a number of other causes.

Symptoms that accompany vertigo may be helpful in distinguishing peripheral vestibular lesions from central causes. Unilateral hearing loss and other aural symptoms (ear pain, pressure, fullness) typically point to a peripheral cause. Because the auditory pathways quickly become bilateral upon entering the brainstem, central lesions are unlikely to cause unilateral hearing loss (unless the lesion lies near the root entry zone of the auditory nerve). Symptoms such as double vision, numbness, and limb ataxia suggest a brainstem or cerebellar lesion.

EXAMINATION

Because dizziness and imbalance can be a manifestation of a variety of neurologic disorders, the neurologic examination is important in the evaluation of these patients. Particular focus should be given to assessment of eye movements, vestibular function, and hearing. The range of eye

movements and whether they are equal in each eye should be observed. Peripheral eye movement disorders (e.g., cranial neuropathies, eye muscle weakness) are usually *disconjugate* (different in the two eyes). One should check pursuit (the ability to follow a smoothly moving target) and saccades (the ability to look back and forth accurately between two targets). Poor pursuit or inaccurate (*dysmetric*) saccades usually indicates central pathology, often involving the cerebellum. Finally, one should look for spontaneous nystagmus, an involuntary back-and-forth movement of the eyes. Most often nystagmus is of the jerk type, in which a slow drift (*slow phase*) in one direction alternates with a rapid saccadic movement (*quick phase* or *fast phase*) in the opposite direction that resets the position of the eyes in the orbits. Table 21-1 lists features that help distinguish peripheral vestibular nystagmus from central nystagmus. Except in the case of acute vestibulopathy (e.g., vestibular neuritis), if primary position nystagmus is easily seen in the light, it is probably due to a central cause. Two forms of nystagmus that are characteristic of lesions of the cerebellar pathways are vertical nystagmus with downward fast phases (downbeat nystagmus) and horizontal nystagmus that changes direction with gaze (gaze-evoked nystagmus).

Table 21-1 Features of Peripheral and Central Vertigo

Sign or Symptom	Peripheral (Labyrinth or Vestibular Nerve)	Central (Brainstem or Cerebellum)
Direction of associated nystagmus	Unidirectional; fast phase opposite lesion[a]	Bidirectional (direction-changing) or unidirectional
Purely horizontal nystagmus without torsional component	Uncommon	May be present
Purely vertical or purely torsional nystagmus	Never present[b]	May be present
Visual fixation	Inhibits nystagmus	No inhibition
Tinnitus and/or deafness	Often present	Usually absent
Associated central nervous system abnormalities	None	Extremely common (e.g., diplopia, hiccups, cranial neuropathies, dysarthria)
Common causes	Benign paroxysmal positional vertigo, infection (labyrinthitis), vestibular neuritis, Ménière's disease, labyrinthine ischemia, trauma, toxin	Vascular, demyelinating, neoplasm

[a] In Ménière's disease, the direction of the fast phase is variable.

[b] Combined vertical-torsional nystagmus suggests BPPV.

Specialists find that the most useful bedside test of peripheral vestibular function is the *head impulse test*, in which the vestibuloocular reflex (VOR) is assessed with small-amplitude (approximately 20 degrees) rapid head rotations; beginning in the primary position, the head is rotated to the left or right while the patient is instructed to fixate on the examiner's face. If the VOR is deficient, a catch-up saccade is seen at the end of the rotation. This test can identify both unilateral (deficient VOR when the head is rotated toward the weak side) and bilateral vestibular hypofunction.

All patients with episodic dizziness, especially if it is provoked by positional change, should be tested with the Dix-Hallpike maneuver. The patient begins in a sitting position with the head turned 45 degrees; holding the back of the head, the examiner then gently lowers the patient into a supine position with the head extended backward by about 20 degrees, and observes for nystagmus; after 30 s the patient is raised to the sitting position and after a 1-min rest the procedure is repeated with the head turned to the other side. Use of Frenzel eyeglasses (self-illuminated goggles with convex lenses that blur the patient's vision but allow the examiner to see the eyes greatly magnified) can improve the sensitivity of the test. If transient upbeating and torsional nystagmus are elicited in the supine position, posterior canal BPPV can be diagnosed confidently and treated with a repositioning maneuver, and additional testing can be avoided.

Dynamic visual acuity is a functional test that can be useful in assessing vestibular function. Visual acuity is measured with the head still and when the head is rotated back and forth by the examiner (about 1–2 Hz). A drop in visual acuity during head motion of more than one line on a near card or Snellen chart is abnormal.

The choice of ancillary tests should be guided by the history and examination findings. Audiometry should be performed whenever a vestibular disorder is suspected. Unilateral sensorineural hearing loss supports a peripheral disorder (e.g., vestibular schwannoma). Predominantly low-frequency

hearing loss is characteristic of Ménière's disease. Electro- or videonystagmography includes recordings of spontaneous nystagmus (if present), pursuit, and saccades; caloric testing to assess the responses of the two horizontal semicircular canals; and measurement of positional nystagmus. Patients with unexplained unilateral hearing loss or vestibular hypofunction should undergo magnetic resonance imaging of the internal auditory canals, including administration of gadolinium, to rule out a schwannoma.

Treatment: Dizziness

Treatment of vestibular symptoms should be driven by the underlying diagnosis. Simply treating dizziness with vestibular suppressant medications is often not helpful and may make the symptoms worse. The diagnostic and specific treatment approaches for the most commonly encountered vestibular disorders are discussed below.

ACUTE PROLONGED VERTIGO

An acute unilateral vestibular lesion causes constant vertigo, nausea, vomiting, oscillopsia (motion of the visual scene), and imbalance. These symptoms are due to a sudden asymmetry of inputs from the two labyrinths or in their central connections, simulating a continuous rotation of the head. Unlike BPPV, the vertigo persists even when the head is not moving.

When a patient presents with an acute vestibular syndrome, the most important question is whether the lesion is central (e.g., a cerebellar or brainstem infarct or hemorrhage), which may be life-threatening, or peripheral, affecting the vestibular nerve or labyrinth. Attention should be given to any symptoms or signs that point to central dysfunction (diplopia, weakness or numbness, dysarthria). The pattern of spontaneous nystagmus, if present, may be helpful (Table 21-1). If the head impulse test is normal, an acute peripheral vestibular lesion is unlikely. However, a central lesion cannot always be excluded with certainly on the basis of symptoms and examination alone; thus, older patients with vascular risk factors who present with an acute vestibular syndrome generally should be evaluated for the possibility of stroke even when there are no specific findings that indicate a central lesion.

Most patients with vestibular neuritis recover spontaneously, but glucocorticoids can improve outcome if administered within 3 days of symptom onset. Antiviral medications are of no proven benefit unless there is evidence to suggest herpes zoster oticus (Ramsay Hunt syndrome). Vestibular suppressant medications may reduce acute symptoms but should be avoided after the first several days as they may impede central compensation and recovery. Patients should be encouraged to resume a normal level of activity as soon as possible, and directed vestibular rehabilitation therapy may accelerate improvement.

BENIGN PAROXYSMAL POSITIONAL VERTIGO

BPPV is a common cause of recurrent vertigo. Episodes are brief (<1 min and typically 15–20 s) and are always provoked by changes in head position relative to gravity, such as lying down, rolling over in bed, rising from a supine position, and extending the head to look upward. The attacks are caused by free-floating otoconia (calcium carbonate crystals) that have been dislodged from the utricular macula and have moved into one of the semicircular canals, usually the posterior canal. When head position changes, gravity causes the otoconia to move within the canal, producing vertigo and nystagmus. With posterior canal BPPV, the nystagmus beats upward and torsionally (the upper poles of the eyes beat toward the affected ear). Less commonly, the otoconia enter the horizontal canal, resulting in a horizontal nystagmus when the patient is lying with either ear down. Superior (also called anterior) canal involvement is rare. BPPV is treated with repositioning maneuvers that utilize gravity to remove the otoconia from the semicircular canal. For posterior canal BPPV, the Epley maneuver is the most commonly used procedure. For more refractory cases of BPPV, patients can be taught a variant of this maneuver that they can perform alone at home.

VESTIBULAR MIGRAINE

Vestibular symptoms occur frequently in migraine, sometimes as a headache aura but often independent of headache. The duration of vertigo may be from minutes to hours, and some patients also experience more prolonged periods of disequilibrium (lasting days to weeks). Motion sensitivity and sensitivity to visual motion (e.g., movies) are common in patients with vestibular migraine. Although data from controlled studies are generally lacking, vestibular migraine typically is treated with medications that are used for prophylaxis of migraine headaches. Antiemetics may be helpful to relieve symptoms at the time of an attack.

MÉNIÈRE'S DISEASE

Attacks of Ménière's disease consist of vertigo, hearing loss, and pain, pressure, or fullness in the affected ear. The hearing loss and aural symptoms are key features that distinguish Ménière's disease from other peripheral vestibulopathies. Audiometry at the time of an attack shows a characteristic asymmetric low-frequency hearing loss; hearing commonly improves between attacks, although permanent hearing loss may occur eventually. Ménière's disease is thought to be due to excess fluid (endolymph) in the inner ear, hence the term *endolymphatic hydrops*. Patients suspected of having Ménière's disease should be referred to an otolaryngologist for further evaluation. Diuretics and

sodium restriction are the initial treatments. If attacks persist, injections of gentamicin into the middle ear are typically the next line of therapy. Full ablative procedures (vestibular nerve section, labyrinthectomy) seldom are required.

VESTIBULAR SCHWANNOMA

Vestibular schwannomas (sometimes less correctly termed *acoustic neuromas*) and other tumors at the cerebellopontine angle cause slowly progressive unilateral sensorineural hearing loss and vestibular hypofunction. These patients typically do not have vertigo, because the gradual vestibular deficit is compensated centrally as it develops. The diagnosis often is not made until there is sufficient hearing loss to be noticed. The examination will show a deficient Halmagyi-Curthoys head impulse response when the head is rotated toward the affected side. Any patient with unexplained asymmetric vestibular function (e.g., no prior history of vestibular neuritis) or asymmetric sensorineural hearing loss (documented on audiometry) should undergo MRI of the internal auditory canals, including gadolinium administration, to rule out a schwannoma.

BILATERAL VESTIBULAR HYPOFUNCTION

Patients with bilateral loss of vestibular function also typically do not have vertigo, since vestibular function is lost on both sides simultaneously, thus there is no asymmetry of vestibular input. Symptoms include loss of balance, particularly in the dark, where vestibular input is most critical, and oscillopsia during head movement, such as while walking or riding in a car. Bilateral vestibular hypofunction may be (1) idiopathic and progressive, (2) part of a neurodegenerative disorder, or (3) iatrogenic, due to medication ototoxicity (most commonly gentamicin or other aminoglycoside antibiotics). Other causes include bilateral vestibular schwannomas (neurofibromatosis type 2), autoimmune disease, meningeal-based infection or tumor, and other toxins. It also may occur in patients with peripheral polyneuropathy; in these patients, both vestibular loss and impaired proprioception may contribute to poor balance. Finally, unilateral processes such as vestibular neuritis and Ménière's disease may involve both ears sequentially, resulting in bilateral vestibulopathy.

Examination findings include diminished dynamic visual acuity (see above) due to loss of stable vision when the head is moving, abnormal head impulse responses in both directions, and a Romberg sign. In the laboratory, responses to caloric testing are reduced. Patients with bilateral vestibular hypofunction should be referred for vestibular rehabilitation therapy. Vestibular suppressant medications should not be used, as they will increase the imbalance. Evaluation by a neurologist is important not only to confirm the diagnosis but also to consider any other associated neurologic abnormalities that may clarify the etiology.

PSYCHOSOMATIC DIZZINESS

Psychological factors play an important role in chronic dizziness. First, dizziness may be a somatic manifestation of a psychiatric condition such as major depression, anxiety, or panic disorder. Second, patients may develop anxiety and autonomic symptoms as a consequence or comorbidity of an independent vestibular disorder. One particular form of this has been termed variously *phobic postural vertigo*, *psychophysiologic vertigo*, or *chronic subjective dizziness*. These patients have a chronic feeling (months or longer) of dizziness and disequilibrium, an increased sensitivity to self-motion and visual motion (e.g., movies), and a particular intensification of symptoms when moving through complex visual environments such as supermarkets (*visual vertigo*). Although there may be a past history of an acute vestibular disorder (e.g., vestibular neuritis), the neurootologic examination and vestibular testing are normal or indicative of a compensated vestibular deficit, indicating that the ongoing subjective dizziness cannot be explained by a primary vestibular disorder. Anxiety disorders are common in patients with chronic dizziness and contribute substantially to the morbidity. Thus, treatment with antianxiety medications [selective serotonin reuptake inhibitors (SSRIs)] and cognitive/behavioral therapy may be helpful. Vestibular rehabilitation therapy is also sometimes beneficial. Vestibular suppressant medications generally should be avoided. This condition should be suspected when the patient states, "My dizziness is so bad, I'm afraid to leave my house" (agoraphobia). General treatment of vertigo consists of vestibular suppressant medications and vestibular rehabilitation therapy.

Treatment: Vertigo

Table 21-2 provides a list of commonly used medications for suppression of vertigo. As noted, these medications should be reserved for short-term control of active vertigo, such as during the first few days of acute vestibular neuritis, or for acute attacks of Ménière's disease. They are less helpful for chronic dizziness and, as previously stated, may hinder central compensation. An exception is that benzodiazepines may attenuate psychosomatic dizziness and the associated anxiety, although SSRIs are generally preferable in such patients.

Table 21-2 Treatment of Vertigo

Agent[a]	Dose[b]

Antihistamines	
Meclizine	25–50 mg 3 times daily
Dimenhydrinate	50 mg 1–2 times daily
Promethazine	25 mg 2–3 times daily (also can be given rectally and IM)
Benzodiazepines	
Diazepam	2.5 mg 1–3 times daily
Clonazepam	0.25 mg 1–3 times daily
Anticholinergic	
Scopolamine transdermal[c]	Patch
Physical therapy	
Repositioning maneuvers[d]	
Vestibular rehabilitation	
Other	
Diuretics and/or low-sodium (1 g/d) diet[e]	
Antimigrainous drugs[f]	
Methylprednisolone[g]	100 mg daily days 1–3; 80 mg daily days 4–6; 60 mg daily days 7–9; 40 mg daily days 10–12; 20 mg daily days 13–15; 10 mg daily days 16–18, 20, 22
Selective serotonin reuptake inhibitors[h]	

[a] All listed drugs are approved by the U.S. Food and Drug Administration, but most are not approved for the treatment of vertigo.

[b] Usual oral (unless otherwise stated) starting dose in adults; a higher maintenance dose can be reached by a gradual increase.

[c] For motion sickness only.

[d] For benign paroxysmal positional vertigo.

[e] For Ménière's disease.

[f] For vestibular migraine.

[g] For acute vestibular neuritis (started within three days of onset).

[h] For psychosomatic vertigo.

Vestibular rehabilitation therapy promotes central adaptation processes that compensate for vestibular loss and also may help habituate motion sensitivity and other symptoms of psychosomatic dizziness. The general approach is to use a graded series of exercises that progressively challenge gaze stabilization and balance.

FURTHER READINGS

BRONSTEIN AM et al: Chronic dizziness: A practical approach. Pract Neurol 10:129, 2010 [PMID: 20498184]

HALMAGYI GM: Diagnosis and management of vertigo. Clin Med 5:159, 2005 [PMID: 15847010]

JEN JC: Bilateral vestibulopathy: Clinical, diagnostic, and genetic considerations. Semin Neurol 29:528, 2009 [PMID: 19834864]

LEMPERT T et al: Vertigo as a symptom of migraine. Ann NY Acad Sci 1164:242, 2009 [PMID: 19645907]

RUCKENSTEIN MJ, STAAB JP: Chronic subjective dizziness. Otolaryngol Clin North Am 42:71, 2009 [PMID: 19134491]

STRUPP M, BRANDT T: Vestibular neuritis. Semin Neurol 29:509, 2009 [PMID: 19834862]

WALKER MF, ZEE DS: Bedside vestibular examination. Otolaryngol Clin North Am 33: 495, 2000 [PMID: 10815033]

ZINGLER VC et al: Causative factors and epidemiology of bilateral vestibulopathy in 255 patients. Ann Neurol 61:524, 2007 [PMID: 17393465]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

WEAKNESS AND PARALYSIS: INTRODUCTION

Normal motor function involves integrated muscle activity that is modulated by the activity of the cerebral cortex, basal ganglia, cerebellum, and spinal cord. Motor system dysfunction leads to weakness or paralysis, which is discussed in this chapter, or to ataxia (Chap. 373) or abnormal movements (Chap. 372). The mode of onset, distribution, and accompaniments of weakness help suggest its cause.

Weakness is a reduction in the power that can be exerted by one or more muscles. Increased fatigability or limitation in function due to pain or articular stiffness often is confused with weakness by patients. *Increased fatigability* is the inability to sustain the performance of an activity that should be normal for a person of the same age, sex, and size. Increased time is required sometimes for full power to be exerted, and this *bradykinesia* may be misinterpreted as weakness. Severe proprioceptive sensory loss also may lead to complaints of weakness because adequate feedback information about the direction and power of movements is lacking. Finally, *apraxia,* a disorder of planning and initiating a skilled or learned movement unrelated to a significant motor or sensory deficit (Chap. 26), sometimes is mistaken for weakness.

Paralysis indicates weakness that is so severe that a muscle cannot be contracted at all, whereas *paresis* refers to weakness that is mild or moderate. The prefix "hemi-" refers to one-half of the body, "para-" to both legs, and "quadri-" to all four limbs. The suffix "-plegia" signifies severe weakness or paralysis.

The distribution of weakness helps to indicate the site of the underlying lesion. Weakness from involvement of upper motor neurons occurs particularly in the extensors and abductors of the upper limb and the flexors of the lower limb. Lower motor neuron weakness does not have this selectivity but depends on whether involvement is at the level of the anterior horn cells, nerve root, limb plexus, or peripheral nerve—only muscles supplied by the affected structure are weak. Myopathic weakness is generally most marked in proximal muscles, whereas weakness from impaired neuromuscular transmission has no specific pattern of involvement. Weakness often is accompanied by other neurologic abnormalities that help indicate the site of the responsible lesion. These abnormalities include changes in tone, muscle bulk, muscle stretch reflexes, and cutaneous reflexes **(Table 22-1)**.

Table 22-1 Signs that Distinguish the Origin of Weakness

Sign	Upper Motor Neuron	Lower Motor Neuron	Myopathic
Atrophy	None	Severe	Mild
Fasciculations	None	Common	None
Tone	Spastic	Decreased	Normal/decreased
Distribution of weakness	Pyramidal/regional	Distal/segmental	Proximal
Tendon reflexes	Hyperactive	Hypoactive/absent	Normal/hypoactive
Babinski sign	Present	Absent	Absent

Tone is the resistance of a muscle to passive stretch. Central nervous system (CNS) abnormalities that cause weakness generally produce *spasticity*, an increase in tone associated with disease of upper motor neurons. Spasticity is velocity-dependent, has a sudden release after reaching a maximum (the "clasp-knife" phenomenon), and predominantly affects the antigravity muscles (i.e., upper-limb flexors and lower-limb extensors). Spasticity is distinct from rigidity and paratonia, two other types of hypertonia. *Rigidity* is increased tone that is present throughout the range of motion (a "lead pipe" or "plastic" stiffness) and affects flexors and extensors equally; it sometimes has a cogwheel quality that is enhanced by voluntary movement of the contralateral limb (reinforcement). Rigidity occurs with certain extrapyramidal disorders, such as Parkinson's disease. *Paratonia* (or *gegenhalten*) is increased tone that varies irregularly in a manner that may seem related to the degree of relaxation, is present throughout the range of motion, and affects flexors and extensors equally; it usually results from disease of the frontal lobes. Weakness with *decreased tone* (*flaccidity*) or normal tone occurs with disorders of *motor units.* A motor unit consists of a single lower motor neuron and all the muscle fibers that it innervates.

Muscle bulk generally is not affected in patients with upper motor neuron lesions, although mild disuse atrophy eventually may occur. By contrast, atrophy is often conspicuous when a lower motor neuron lesion is responsible for weakness and also may occur with advanced muscle disease.

Muscle stretch (tendon) reflexes are usually increased with upper motor neuron lesions, although they may be decreased or absent for a variable period immediately after onset of an acute lesion. This is usually—but not invariably—accompanied by abnormalities of *cutaneous reflexes* (such as superficial abdominals; Chap. 367) and, in particular, by an extensor plantar (Babinski) response. The muscle stretch reflexes are depressed in patients with lower motor neuron lesions when there is direct involvement of specific reflex arcs. The stretch reflexes generally are preserved in patients with myopathic weakness except in advanced stages, when they sometimes are attenuated. In disorders of the neuromuscular junction, the intensity of the reflex responses may be affected by preceding voluntary activity of affected muscles; that activity may lead to enhancement of initially depressed reflexes in Lambert-Eaton myasthenic syndrome and, conversely, to depression of initially normal reflexes in myasthenia gravis (Chap. 386).

The distinction of *neuropathic* (lower motor neuron) from *myopathic* weakness is sometimes difficult clinically, although distal weakness is likely to be neuropathic, and symmetric proximal weakness myopathic. *Fasciculations* (visible or palpable twitch within a muscle due to the spontaneous discharge of a motor unit) and early atrophy indicate that weakness is neuropathic.

Pathogenesis

UPPER MOTOR NEURON WEAKNESS

This pattern of weakness results from disorders that affect the upper motor neurons or their axons in the cerebral cortex, subcortical white matter, internal capsule, brainstem, or spinal cord **(Fig. 22-1)**. These lesions produce weakness through decreased activation of the lower motor neurons. In general, distal muscle groups are affected more severely than are proximal ones, and axial movements are spared unless the lesion is severe and bilateral. With corticobulbar involvement, weakness usually is observed only in the lower face and tongue; extraocular, upper facial, pharyngeal, and jaw muscles almost always are spared. With bilateral corticobulbar lesions, *pseudobulbar palsy* often develops: dysarthria, dysphagia, dysphonia, and emotional lability accompany bilateral facial weakness and a brisk jaw jerk. Spasticity accompanies upper motor neuron weakness but may not be present in the acute phase. Upper motor neuron lesions also affect the ability to perform rapid repetitive movements. Such movements are slow and coarse, but normal rhythmicity is maintained. Finger-nose-finger and heel-knee-shin maneuvers are performed slowly but adequately.

Figure 22-1

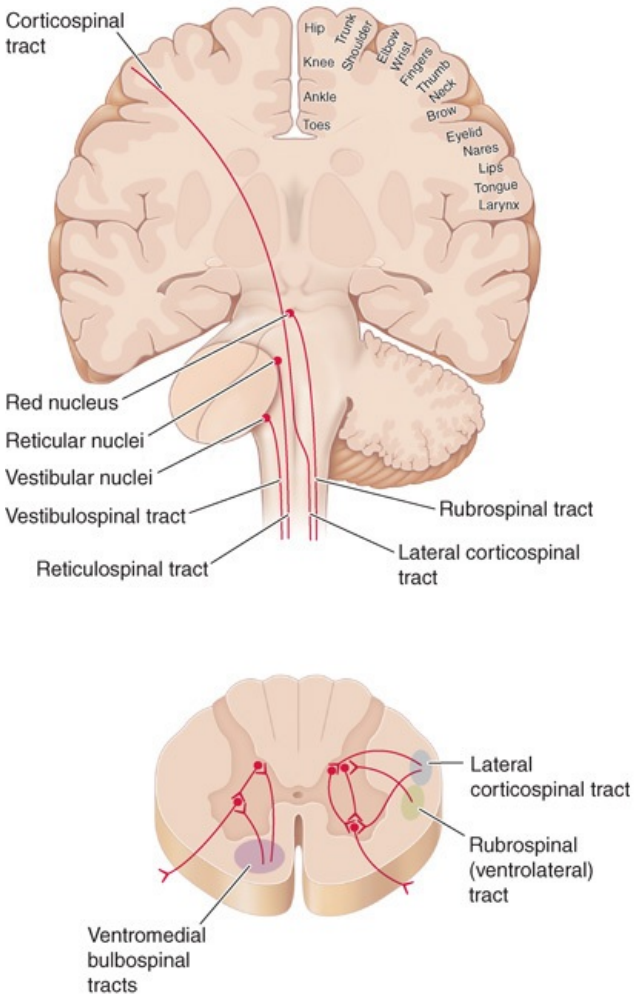

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The corticospinal and bulbospinal upper motor neuron pathways. Upper motor neurons have their cell bodies in layer V of the primary motor cortex (the precentral gyrus, or Brodmann's area 4) and in the premotor and supplemental motor cortex (area 6). The upper motor neurons in the primary motor cortex are somatotopically organized, as illustrated on the right side of the figure.

Axons of the upper motor neurons descend through the subcortical white matter and the posterior limb of the internal capsule. Axons of the *pyramidal* or *corticospinal system* descend through the brainstem in the cerebral peduncle of the midbrain, the basis pontis, and the medullary pyramids. At the cervicomedullary junction, most pyramidal axons decussate into the contralateral corticospinal tract of the lateral spinal cord, but 10–30% remain ipsilateral in the anterior spinal cord. Pyramidal neurons make direct monosynaptic connections with lower motor neurons. They innervate most densely the lower motor neurons of hand muscles and are involved in the execution of learned, fine movements. Corticobulbar neurons are similar to corticospinal neurons but innervate brainstem motor nuclei.

Bulbospinal upper motor neurons influence strength and tone but are not part of the pyramidal system. The descending *ventromedial bulbospinal pathways* originate in the tectum of the midbrain (tectospinal pathway), the vestibular nuclei (vestibulospinal pathway), and the reticular formation (reticulospinal pathway). These pathways influence axial and proximal muscles and are involved in the maintenance of posture and integrated movements of the limbs and trunk. The descending *ventrolateral bulbospinal pathways*, which originate predominantly in the red nucleus (rubrospinal pathway), facilitate distal limb muscles. The bulbospinal system sometimes is referred to as the *extrapyramidal upper motor neuron system* . In all figures, nerve cell bodies and axon terminals are shown, respectively, as closed circles and forks.

LOWER MOTOR NEURON WEAKNESS

This pattern results from disorders of cell bodies of lower motor neurons in the brainstem motor nuclei and the anterior horn of the spinal cord or from dysfunction of the axons of these neurons as they pass to skeletal muscle **(Fig. 22-2)**. Weakness is due to a decrease in the number of muscle fibers that can be activated through a loss of α motor neurons or disruption of their connections to muscle. Loss of γ motor neurons does not cause weakness but decreases tension on the muscle spindles, which decreases muscle tone and attenuates the stretch reflexes elicited on examination. An absent stretch reflex suggests involvement of spindle afferent fibers.

Figure 22-2

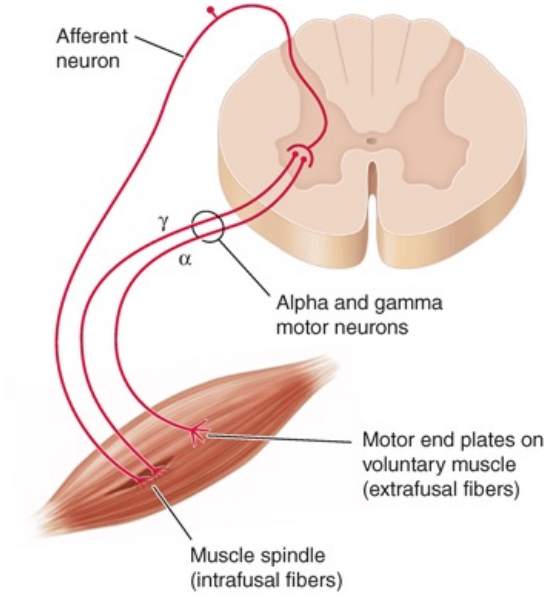

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lower motor neurons are divided into α and γ types. The larger α motor neurons are more numerous and innervate the extrafusal muscle fibers of the motor unit. Loss of α motor neurons or disruption of their axons produces lower motor neuron weakness. The smaller, less numerous γ motor neurons innervate the intrafusal muscle fibers of the muscle spindle and contribute to normal tone and stretch reflexes. The α motor neuron receives direct excitatory input from corticomotoneurons and primary muscle spindle afferents. The α and γ motor neurons also receive excitatory input from other descending upper motor neuron pathways, segmental sensory inputs, and interneurons. The α motor neurons receive direct inhibition from Renshaw cell interneurons, and other interneurons indirectly inhibit the α and γ motor neurons.

A tendon reflex requires the function of all the illustrated structures. A tap on a tendon stretches muscle spindles (which are tonically activated by γ motor neurons) and activates the primary spindle afferent neurons. These neurons stimulate the α motor neurons in the spinal cord, producing a brief muscle contraction, which is the familiar tendon reflex.

When a motor unit becomes diseased, especially in anterior horn cell diseases, it may discharge spontaneously, producing *fasciculations* that may be seen or felt clinically or recorded by electromyography (EMG). When α motor neurons or their axons degenerate, the denervated muscle fibers also may discharge spontaneously. These single muscle fiber discharges, or *fibrillation potentials,* cannot be seen or felt but can be recorded with EMG. If lower motor neuron weakness is present, recruitment of motor units is delayed or reduced, with fewer than normal activated at a particular discharge frequency. This contrasts with weakness of the upper motor neuron type, in which a normal number of motor units is activated at a given frequency but with a diminished maximal discharge frequency.

MYOPATHIC WEAKNESS

Myopathic weakness is produced by disorders of the muscle fibers. Disorders of the neuromuscular junctions also produce weakness, but this is variable in degree and distribution and is influenced by preceding activity of the affected muscle. At a muscle fiber, if the nerve terminal releases a normal number of acetylcholine molecules presynaptically and a sufficient number of postsynaptic acetylcholine receptors are opened, the end plate reaches threshold and thereby generates an action potential that spreads across the muscle fiber membrane and into the transverse tubular system. This electrical excitation activates intracellular events that produce an energy-dependent contraction of the muscle fiber (excitation-contraction coupling).

Myopathic weakness is produced by a decrease in the number or contractile force of muscle fibers activated within motor units. With muscular dystrophies, inflammatory myopathies, or myopathies with muscle fiber necrosis, the number of muscle fibers is reduced within many motor units. On EMG, the size of each motor unit action potential is decreased, and motor units must be recruited more rapidly than normal to produce the desired power. Some myopathies produce weakness through loss of contractile force of muscle fibers or through relatively selective involvement of type II (fast) fibers. These myopathies may not affect the size of individual motor unit action potentials and are detected by a discrepancy between the electrical activity and force of a muscle.

Diseases of the neuromuscular junction, such as myasthenia gravis, produce weakness in a similar

manner, but the loss of muscle fibers is functional (due to inability to activate them) rather than related to muscle fiber loss. The number of muscle fibers that are activated varies over time, depending on the state of rest of the neuromuscular junctions. Thus, fatigable weakness is suggestive of myasthenia gravis or other disorders of the neuromuscular junction.

HEMIPARESIS

Hemiparesis results from an upper motor neuron lesion above the midcervical spinal cord; most such lesions are above the foramen magnum. The presence of other neurologic deficits helps localize the lesion. Thus, language disorders, cortical sensory disturbances, cognitive abnormalities, disorders of visual-spatial integration, apraxia, or seizures point to a cortical lesion. Homonymous visual field defects reflect either a cortical or a subcortical hemispheric lesion. A "pure motor" hemiparesis of the face, arm, and leg often is due to a small, discrete lesion in the posterior limb of the internal capsule, cerebral peduncle, or upper pons. Some brainstem lesions produce "crossed paralyses," consisting of ipsilateral cranial nerve signs and contralateral hemiparesis (Chap. 370). The absence of cranial nerve signs or facial weakness suggests that a hemiparesis is due to a lesion in the high cervical spinal cord, especially if it is associated with ipsilateral loss of proprioception and contralateral loss of pain and temperature sense (the Brown-Séquard syndrome).

Acute or episodic hemiparesis usually results from ischemic or hemorrhagic stroke but also may relate to hemorrhage occurring into brain tumors or may be a result of trauma; other causes include a focal structural lesion or an inflammatory process as in multiple sclerosis, abscess, or sarcoidosis. Evaluation **(Fig. 22-3)** begins immediately with a CT scan of the brain and laboratory studies. If the CT is normal and an ischemic stroke is unlikely, MRI of the brain or cervical spine is performed.

Figure 22-3

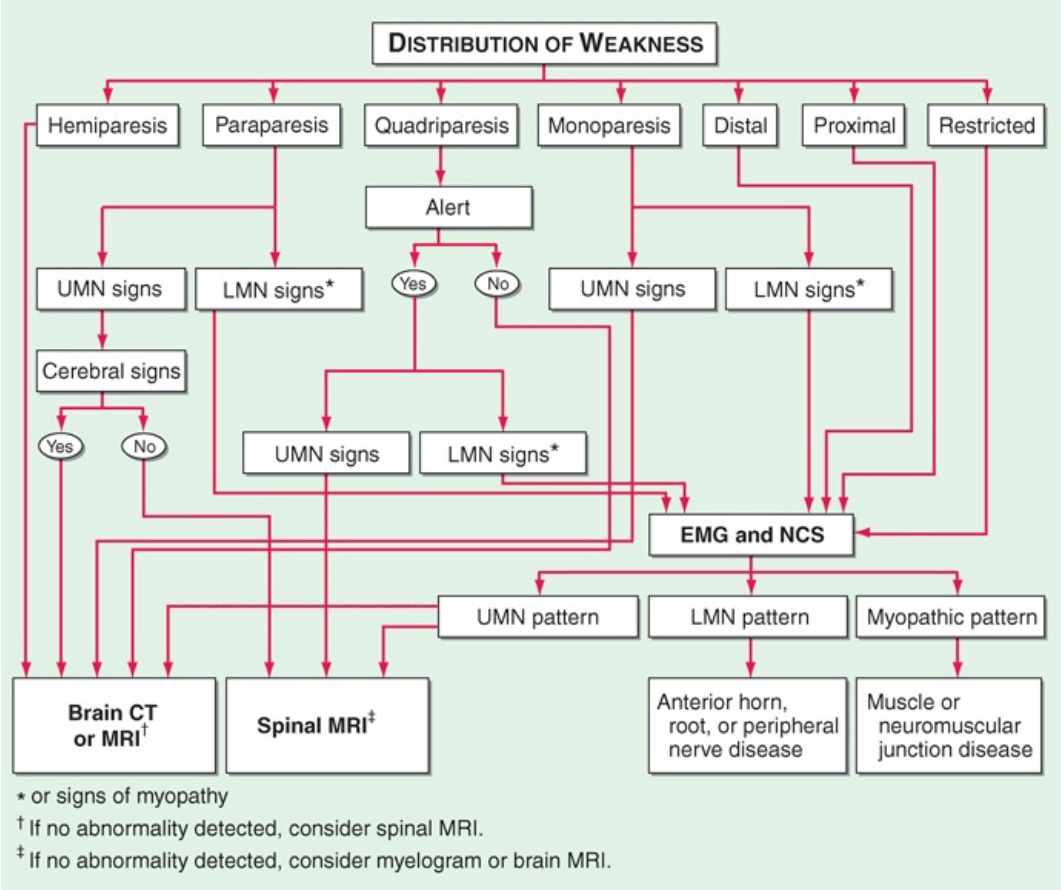

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An algorithm for the initial workup of a patient with weakness. CT, computed tomography; EMG, electromyography; LMN, lower motor neuron; MRI, magnetic resonance imaging; NCS, nerve conduction studies; UMN, upper motor neuron.

Subacute hemiparesis that evolves over days or weeks has an extensive differential diagnosis. A common cause is subdural hematoma, especially in elderly or anticoagulated patients, even when there is no history of trauma. Infectious possibilities include cerebral abscess, fungal granuloma or meningitis, and parasitic infection. Weakness from primary and metastatic neoplasms may evolve over

days to weeks. AIDS may present with subacute hemiparesis due to toxoplasmosis or primary CNS lymphoma. Noninfectious inflammatory processes such as multiple sclerosis or, less commonly, sarcoidosis merit consideration. If the brain MRI is normal and there are no cortical and hemispheric signs, MRI of the cervical spine should be undertaken.

Chronic hemiparesis that evolves over months usually is due to a neoplasm or vascular malformation, a chronic subdural hematoma, or a degenerative disease. If an MRI of the brain is normal, the possibility of a foramen magnum or high cervical spinal cord lesion should be considered.

PARAPARESIS

An intraspinal lesion at or below the upper thoracic spinal cord level is most commonly responsible, but a paraparesis also may result from lesions at other locations that disturb upper motor neurons (especially parasagittal intracranial lesions) and lower motor neurons [anterior horn cell disorders, cauda equina syndromes due to involvement of nerve roots derived from the lower spinal cord (Chap. 377), and peripheral neuropathies].

Acute paraparesis may not be recognized as due to spinal cord disease at an early stage if the legs are flaccid and areflexic. Usually, however, there is sensory loss in the legs with an upper level on the trunk, a dissociated sensory loss suggestive of a central cord syndrome, or exaggerated stretch reflexes in the legs with normal reflexes in the arms. It is important to image the spinal cord (Fig. 22-3). Compressive lesions (particularly epidural tumor, abscess, and hematoma but also a prolapsed intervertebral disk and vertebral involvement by malignancy or infection), spinal cord infarction (proprioception usually is spared), an arteriovenous fistula or other vascular anomaly, and transverse myelitis are among the possible causes (Chap. 377).

Diseases of the cerebral hemispheres that produce acute paraparesis include anterior cerebral artery ischemia (shoulder shrug also is affected), superior sagittal sinus or cortical venous thrombosis, and acute hydrocephalus. If upper motor neuron signs are associated with drowsiness, confusion, seizures, or other hemispheric signs, MRI of the brain should be undertaken.

Paraparesis may result from a cauda equina syndrome, for example, after trauma to the low back, a midline disk herniation, or an intraspinal tumor; although sphincters are affected, hip flexion often is spared, as is sensation over the anterolateral thighs. Rarely, paraparesis is caused by a rapidly evolving anterior horn cell disease (such as poliovirus or West Nile virus infection), peripheral neuropathy (such as Guillain-Barré syndrome; Chap. 385), or myopathy (Chap. 387). In such cases, electrophysiologic studies are diagnostically helpful and refocus the subsequent evaluation.

Subacute or chronic paraparesis with spasticity is caused by upper motor neuron disease. When there is associated lower-limb sensory loss and sphincter involvement, a chronic spinal cord disorder is likely (Chap. 377). If an MRI of the spinal cord is normal, MRI of the brain may be indicated. If hemispheric signs are present, a parasagittal meningioma or chronic hydrocephalus is likely and MRI of the brain is the initial test. In the rare situation in which a long-standing paraparesis has a lower motor neuron or myopathic etiology, the localization usually is suspected on clinical grounds by the absence of spasticity and confirmed by EMG and nerve conduction tests.

QUADRIPARESIS OR GENERALIZED WEAKNESS

Generalized weakness may be due to disorders of the CNS or the motor unit. Although the terms *quadriparesis* and *generalized weakness* often are used interchangeably, quadriparesis is commonly used when an upper motor neuron cause is suspected, and generalized weakness when a disease of the motor unit is likely. Weakness from CNS disorders usually is associated with changes in consciousness or cognition, with spasticity and brisk stretch reflexes, and with alterations of sensation. Most neuromuscular causes of generalized weakness are associated with normal mental function, hypotonia, and hypoactive muscle stretch reflexes. The major causes of intermittent weakness are listed in **Table 22-2**. A patient with generalized fatigability without objective weakness may have the chronic fatigue syndrome (Chap. 389).

Table 22-2 Causes of Episodic Generalized Weakness

1. Electrolyte disturbances, e.g., hypokalemia, hyperkalemia, hypercalcemia, hypernatremia, hyponatremia, hypophosphatemia, hypermagnesemia
2. Muscle disorders
a. Channelopathies (periodic paralyses)
b. Metabolic defects of muscle (impaired carbohydrate or fatty acid utilization; abnormal mitochondrial function)
3. Neuromuscular junction disorders
a. Myasthenia gravis
b. Lambert-Eaton myasthenic syndrome
4. Central nervous system disorders

a. Transient ischemic attacks of the brainstem
b. Transient global cerebral ischemia
c. Multiple sclerosis

Acute Quadriparesis

Acute quadriparesis with onset over minutes may result from disorders of upper motor neurons (e.g., anoxia, hypotension, brainstem or cervical cord ischemia, trauma, and systemic metabolic abnormalities) or muscle (electrolyte disturbances, certain inborn errors of muscle energy metabolism, toxins, and periodic paralyses). Onset over hours to weeks may, in addition to these disorders, be due to lower motor neuron disorders. Guillain-Barré syndrome (Chap. 385) is the most common lower motor neuron weakness that progresses over days to 4 weeks; the finding of an elevated protein level in the cerebrospinal fluid is helpful but may be absent early in the course.

In obtunded patients, evaluation begins with a CT scan of the brain. If upper motor neuron signs are present but the patient is alert, the initial test is usually an MRI of the cervical cord. If weakness is lower motor neuron, myopathic, or uncertain in origin, the clinical approach begins with blood studies to determine the level of muscle enzymes and electrolytes and an EMG and nerve conduction study.

Subacute or Chronic Quadriparesis

When quadriparesis due to upper motor neuron disease develops over weeks, months, or years, the distinction between disorders of the cerebral hemispheres, brainstem, and cervical spinal cord is usually possible clinically. An MRI is obtained of the clinically suspected site of pathology. EMG and nerve conduction studies help distinguish lower motor neuron disease (which usually presents with weakness that is most profound distally) from myopathic weakness, which is typically proximal.

MONOPARESIS

Monoparesis usually is due to lower motor neuron disease, with or without associated sensory involvement. Upper motor neuron weakness occasionally presents as a monoparesis of distal and nonantigravity muscles. Myopathic weakness rarely is limited to one limb.

Acute Monoparesis

If the weakness is predominantly in distal and nonantigravity muscles and is not associated with sensory impairment or pain, focal cortical ischemia is likely (Chap. 370); diagnostic possibilities are similar to those for acute hemiparesis. Sensory loss and pain usually accompany acute lower motor neuron weakness; the weakness commonly is localized to a single nerve root or peripheral nerve within the limb but occasionally reflects plexus involvement. If lower motor neuron weakness is suspected or the pattern of weakness is uncertain, the clinical approach begins with an EMG and a nerve conduction study.

Subacute or Chronic Monoparesis

Weakness and atrophy that develop over weeks or months are usually of lower motor neuron origin. If they are associated with sensory symptoms, a peripheral cause (nerve, root, or plexus) is likely; in the absence of such symptoms, anterior horn cell disease should be considered. In either case, an electrodiagnostic study is indicated. If weakness is of the upper motor neuron type, a discrete cortical (precentral gyrus) or cord lesion may be responsible, and an imaging study of the appropriate site is performed.

DISTAL WEAKNESS

Involvement of two or more limbs distally suggests lower motor neuron or peripheral nerve disease. Acute distal lower limb weakness results occasionally from an acute toxic polyneuropathy or cauda equina syndrome. Distal symmetric weakness usually develops over weeks, months, or years and, when associated with numbness, is due to diseases of peripheral nerves (Chap. 384). Anterior horn cell disease may begin distally but is typically asymmetric and without accompanying numbness (Chap. 374). Rarely, myopathies present with distal weakness (Chap. 387). Electrodiagnostic studies help localize the disorder (Fig. 22-3).

PROXIMAL WEAKNESS

Myopathy often produces symmetric weakness of the pelvic or shoulder girdle muscles (Chap. 387). Diseases of the neuromuscular junction [such as myasthenia gravis (Chap. 386)], may present with symmetric proximal weakness often associated with ptosis, diplopia, or bulbar weakness and fluctuating in severity during the day. The extreme fatigability present in some cases of myasthenia gravis may even suggest episodic weakness, but strength rarely returns fully to normal. In anterior horn cell disease, proximal weakness is usually asymmetric, but it may be symmetric if familial. Numbness does not occur with any of these diseases. The evaluation usually begins with determination of the serum creatine kinase level and electrophysiologic studies.

WEAKNESS IN A RESTRICTED DISTRIBUTION

Weakness may not fit any of these patterns, being limited, for example, to the extraocular, hemifacial,

bulbar, or respiratory muscles. If it is unilateral, restricted weakness usually is due to lower motor neuron or peripheral nerve disease, such as in a facial palsy or an isolated superior oblique muscle paresis. Weakness of part of a limb usually is due to a peripheral nerve lesion such as carpal tunnel syndrome or another entrapment neuropathy. Relatively symmetric weakness of extraocular or bulbar muscles usually is due to a myopathy (Chap. 387) or neuromuscular junction disorder (Chap. 386). Bilateral facial palsy with areflexia suggests Guillain-Barré syndrome (Chap. 385). Worsening of relatively symmetric weakness with fatigue is characteristic of neuromuscular junction disorders. Asymmetric bulbar weakness usually is due to motor neuron disease. Weakness limited to respiratory muscles is uncommon and usually is due to motor neuron disease, myasthenia gravis, or polymyositis/dermatomyositis (Chap. 388).

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NUMBNESS, TINGLING, AND SENSORY LOSS: INTRODUCTION

Normal somatic sensation reflects a continuous monitoring process, little of which reaches consciousness under ordinary conditions. By contrast, disordered sensation, particularly when experienced as painful, is alarming and dominates the patient's attention. Physicians should be able to recognize abnormal sensations by how they are described, know their type and likely site of origin, and understand their implications. **Pain is considered separately in Chap. 11.**

Positive and Negative Symptoms

Abnormal sensory symptoms can be divided into two categories: positive and negative. The prototypical positive symptom is tingling (pins and needles); other positive sensory phenomena include altered sensations that are described as pricking, bandlike, lightning-like shooting feelings (lancinations), aching, knifelike, twisting, drawing, pulling, tightening, burning, searing, electrical, or raw feelings. Such symptoms are often painful.

Positive phenomena usually result from trains of impulses generated at sites of lowered threshold or heightened excitability along a peripheral or central sensory pathway. The nature and severity of the abnormal sensation depend on the number, rate, timing, and distribution of ectopic impulses and the type and function of nervous tissue in which they arise. Because positive phenomena represent excessive activity in sensory pathways, they are not necessarily associated with a sensory deficit (loss) on examination.

Negative phenomena represent loss of sensory function and are characterized by diminished or absent feeling that often is experienced as numbness and by abnormal findings on sensory examination. In disorders affecting peripheral sensation, it is estimated that at least one-half the afferent axons innervating a particular site are lost or functionless before a sensory deficit can be demonstrated by clinical examination. This threshold varies in accordance with how rapidly function is lost in sensory nerve fibers. If the rate of loss is slow, lack of cutaneous feeling may be unnoticed by the patient and difficult to demonstrate on examination, even though few sensory fibers are functioning; if it is rapid, both positive and negative phenomena are usually conspicuous. Subclinical degrees of sensory dysfunction may be revealed by sensory nerve conduction studies or somatosensory evoked potentials (**Chap. e45**).

Whereas sensory symptoms may be either positive or negative, sensory signs on examination are always a measure of negative phenomena.

Terminology

Words used to characterize sensory disturbance are descriptive and based on convention. Paresthesias and dysesthesias are general terms used to denote positive sensory symptoms. The term *paresthesias* typically refers to tingling or pins-and-needles sensations but may include a wide variety of other abnormal sensations, except pain; it sometimes implies that the abnormal sensations are perceived spontaneously. The more general term *dysesthesia*s denotes all types of abnormal sensations, including painful ones, regardless of whether a stimulus is evident.

Another set of terms refers to sensory abnormalities found on examination. *Hypesthesia* or *hypoesthesia* refers to a reduction of cutaneous sensation to a specific type of testing such as pressure, light touch, and warm or cold stimuli; *anesthesia,* to a complete absence of skin sensation to the same stimuli plus pinprick; and *hypalgesia* or *analgesia,* to reduced or absent pain perception (nociception), such as perception of the pricking quality elicited by a pin. *Hyperesthesia* means pain or increased sensitivity in response to touch. Similarly, *allodynia* describes the situation in which a nonpainful stimulus, once perceived, is experienced as painful, even excruciating. An example is elicitation of a painful sensation by application of a vibrating tuning fork. *Hyperalgesia* denotes severe pain in response to a mildly noxious stimulus, and *hyperpathia*, a broad term, encompasses all the phenomena described by hyperesthesia, allodynia, and hyperalgesia. With hyperpathia, the threshold for a sensory stimulus is increased and perception is delayed, but once felt, it is unduly painful.

Disorders of deep sensation arising from muscle spindles, tendons, and joints affect proprioception (position sense). Manifestations include imbalance (particularly with eyes closed or in the dark), clumsiness of precision movements, and unsteadiness of gait, which are referred to collectively as *sensory ataxia*. Other findings on examination usually, but not invariably, include reduced or absent joint position and vibratory sensibility and absent deep tendon reflexes in the affected limbs. The Romberg sign is positive, which means that the patient sways markedly or topples when asked to stand with feet close together and eyes closed. In severe states of deafferentation involving deep sensation, the patient cannot walk or stand unaided or even sit unsupported. Continuous involuntary movements (*pseudoathetosis*) of the outstretched hands and fingers occur, particularly with eyes closed.

Anatomy of Sensation

Cutaneous afferent innervation is conveyed by a rich variety of receptors, both naked nerve endings (nociceptors and thermoreceptors) and encapsulated terminals (mechanoreceptors). Each type of receptor has its own set of sensitivities to specific stimuli, size and distinctness of receptive fields, and adaptational qualities. Much of the knowledge about these receptors has come from the development of techniques to study single intact nerve fibers intraneurally in awake, unanesthetized human subjects. It is possible not only to record from but also to stimulate single fibers in isolation. A single impulse, whether elicited by a natural stimulus or evoked by electrical microstimulation in a large myelinated afferent fiber, may be both perceived and localized.

Afferent fibers of all sizes in peripheral nerve trunks traverse the dorsal roots and enter the dorsal horn of the spinal cord **(Fig. 23-1)**. From there the smaller fibers take a route to the parietal cortex different from that of the larger fibers. The polysynaptic projections of the smaller fibers (unmyelinated and small myelinated), which subserve mainly nociception, temperature sensibility, and touch, cross and ascend in the opposite anterior and lateral columns of the spinal cord, through the brainstem, to the ventral posterolateral (VPL) nucleus of the thalamus and ultimately project to the postcentral gyrus of the parietal cortex (Chap. 11). This is the *spinothalamic pathway* or *anterolateral system*. The larger fibers, which subserve tactile and position sense and kinesthesia, project rostrally in the posterior column on the same side of the spinal cord and make their first synapse in the gracile or cuneate nucleus of the lower medulla. Axons of second-order neurons decussate and ascend in the medial lemniscus located medially in the medulla and in the tegmentum of the pons and midbrain and synapse in the VPL nucleus; third-order neurons project to parietal cortex. This large-fiber system is referred to as the *posterior column–medial lemniscal pathway* (lemniscal, for short). Note that although the lemniscal and the anterolateral pathways both project up the spinal cord to the thalamus, it is the (crossed) anterolateral pathway that is referred to as the *spinothalamic tract* by convention.

Figure 23-1

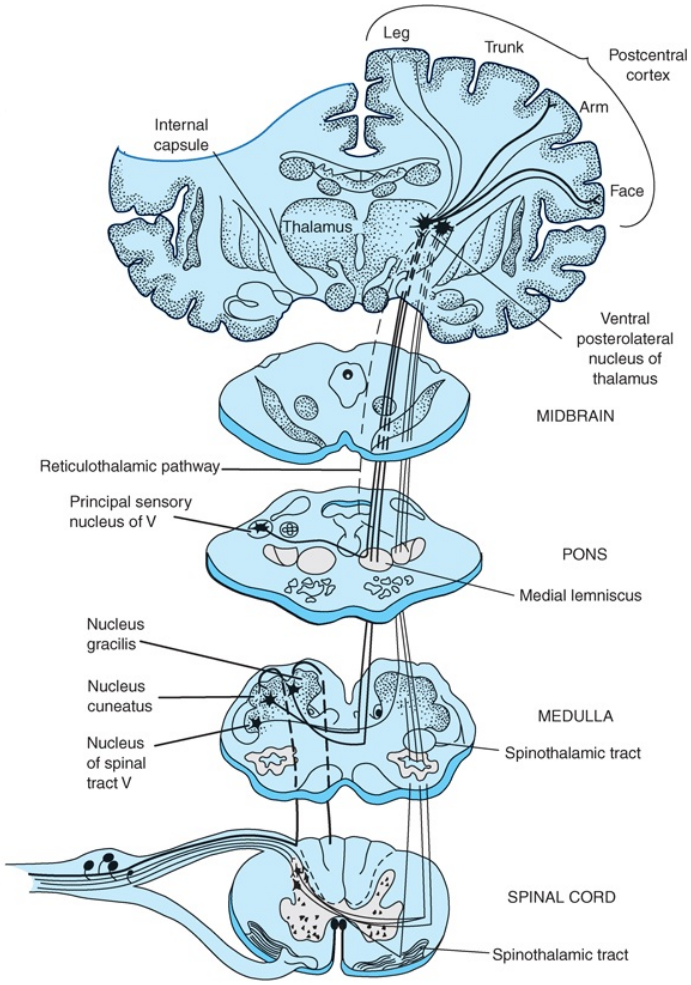

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The main somatosensory pathways. The spinothalamic tract (pain, thermal sense) and the posterior column–lemniscal system (touch, pressure, joint position) are shown. Offshoots from the ascending anterolateral fasciculus (spinothalamic tract) to nuclei in the medulla, pons, and mesencephalon and nuclear terminations of the tract are indicated. *(From AH Ropper, RH Brown, in Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009.)*

Although the fiber types and functions that make up the spinothalamic and lemniscal systems are relatively well known, many other fibers, particularly those associated with touch, pressure, and position sense, ascend in a diffusely distributed pattern both ipsilaterally and contralaterally in the anterolateral quadrants of the spinal cord. This explains why a complete lesion of the posterior columns of the spinal cord may be associated with little sensory deficit on examination.

EXAMINATION OF SENSATION

The main components of the sensory examination are tests of primary sensation (pain, touch, vibration, joint position, and thermal sensation; (**Table 23-1**).

Table 23-1 Testing Primary Sensation

Sense	Test Device	Endings Activated	Fiber Size Mediating	Central Pathway
Pain	Pinprick	Cutaneous nociceptors	Small	SpTh, also D
Temperature, heat	Warm metal object	Cutaneous thermoreceptors for hot	Small	SpTh
Temperature, cold	Cold metal object	Cutaneous thermoreceptors for cold	Small	SpTh
Touch	Cotton wisp, fine brush	Cutaneous mechanoreceptors, also naked endings	Large and small	Lem, also D and SpTh
Vibration	Tuning fork, 128 Hz	Mechanoreceptors, especially pacinian corpuscles	Large	Lem, also D
Joint position	Passive movement of specific joints	Joint capsule and tendon endings, muscle spindles	Large	Lem, also D

Abbreviations: D, diffuse ascending projections in ipsilateral and contralateral anterolateral columns; SpTh, spinothalamic projection, contralateral; Lem, posterior column and lemniscal projection, ipsilateral.

Some general principles pertain. The examiner must depend on patient responses, particularly when testing cutaneous sensation (pin, touch, warm, or cold), and this complicates interpretation. Further, examination may be limited in some patients. In a stuporous patient, for example, sensory examination is reduced to observing the briskness of withdrawal in response to a pinch or another noxious stimulus. Comparison of response on one side of the body to that on the other is essential. In an alert but uncooperative patient, it may not be possible to examine cutaneous sensation, but some idea of proprioceptive function may be gained by noting the patient's best performance of movements requiring balance and precision. Frequently, patients present with sensory symptoms that do not fit an anatomic localization and that are accompanied by either no abnormalities or gross inconsistencies on examination. The

examiner should consider whether the sensory symptoms are a disguised request for help with psychological or situational problems. Discretion must be used in pursuing this possibility. Finally, sensory examination of a patient who has no neurologic complaints can be brief and consist of pinprick, touch, and vibration testing in the hands and feet plus evaluation of stance and gait, including the Romberg maneuver. Evaluation of stance and gait also tests the integrity of motor and cerebellar systems.

PRIMARY SENSATION

(See Table 23-1) The sense of pain usually is tested with a clean pin, with the patient asked to focus on the pricking or unpleasant quality of the stimulus, not just the pressure or touch sensation elicited. Areas of hypalgesia should be mapped by proceeding radially from the most hypalgesic site **(Figs. 23-2, 23-3 and 23-4)**.

Figure 23-2

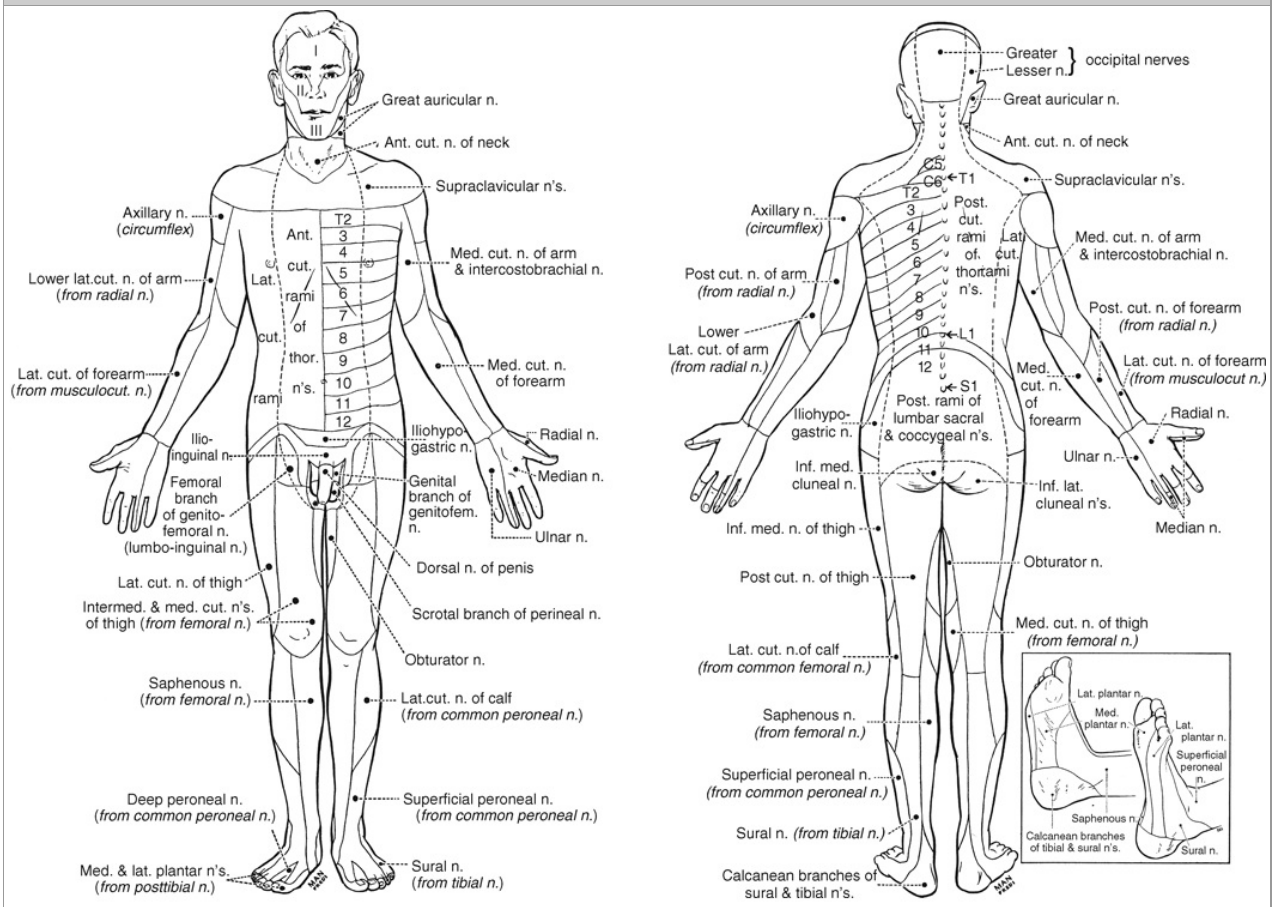

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The cutaneous fields of peripheral nerves. *(Reproduced by permission from W Haymaker, B Woodhall: Peripheral Nerve Injuries, 2nd ed. Philadelphia, Saunders, 1953.)*

Figure 23-3

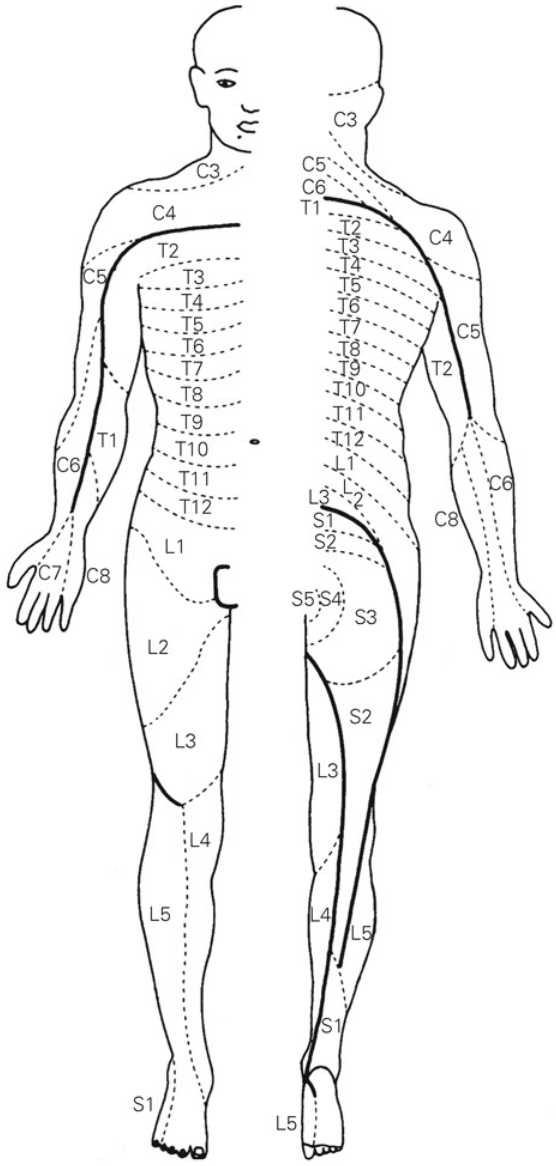

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Distribution of the sensory spinal roots on the surface of the body (dermatomes). (From D Sinclair: Mechanisms of Cutaneous Sensation. Oxford, UK, Oxford University Press, 1981; with permission from Dr. David Sinclair.)

Figure 23-4

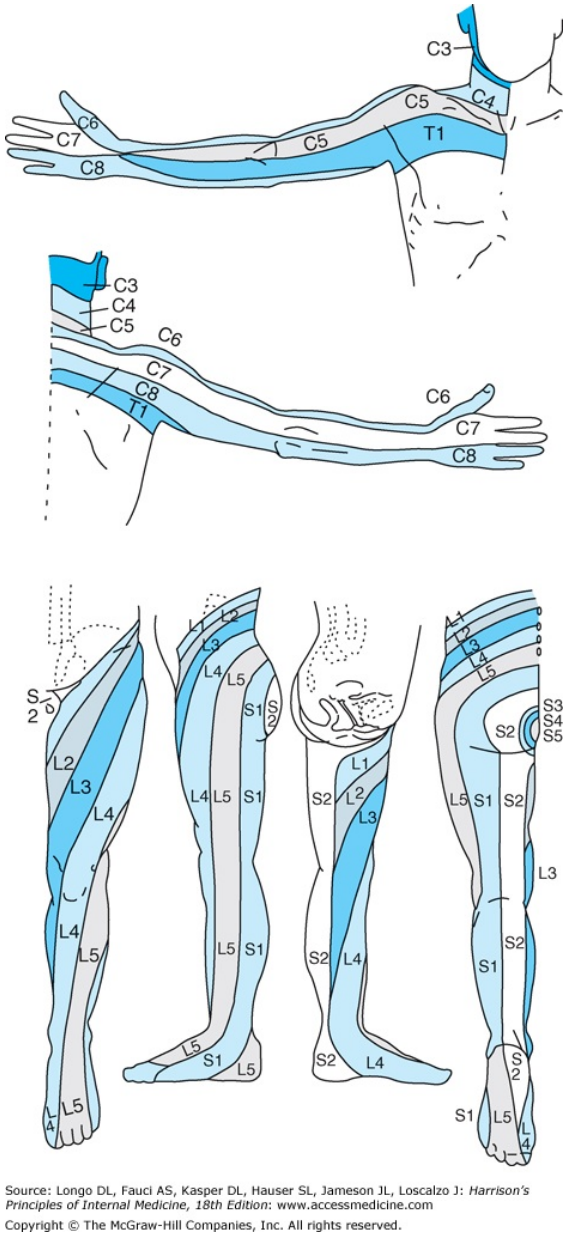

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dermatomes of the upper and lower extremities, outlined by the pattern of sensory loss following lesions of single nerve roots. (*From JJ Keegan, FD Garrett: Anat Rec 102:409, 1948.*)

Temperature sensation to both hot and cold is best tested with small containers filled with water of the desired temperature. This is impractical in most settings. An alternative way to test cold sensation is to touch a metal object, such as a tuning fork at room temperature, to the skin. For testing warm temperatures, the tuning fork or another metal object may be held under warm water of the desired temperature and then used. The appreciation of both cold and warmth should be tested because different receptors respond to each.

Touch usually is tested with a wisp of cotton or a fine camel hair brush. In general, it is better to avoid testing touch on hairy skin because of the profusion of the sensory endings that surround each hair follicle.

Joint position testing is a measure of proprioception, one of the most important functions of the sensory system. With the patient's eyes closed, joint position is tested in the distal interphalangeal joint of the great toe and fingers. If errors are made in recognizing the direction of passive movements, more proximal joints are tested. A test of proximal joint position sense, primarily at the shoulder, is performed by asking the patient to bring the two index fingers together with arms extended and eyes closed. Normal individuals can do this accurately, with errors of 1 cm or less.

The sense of vibration is tested with a tuning fork that vibrates at 128 Hz. Vibration usually is tested over bony points, beginning distally; in the feet it is tested over the dorsal surface of the distal phalanx of the big toes and at the malleoli of the ankles, and in the hands dorsally at the distal phalanx of the fingers. If abnormalities are found, more proximal sites can be examined. Vibratory thresholds at the same site in the patient and the examiner may be compared for control purposes.

QUANTITATIVE SENSORY TESTING

Effective sensory testing devices are now available commercially. Quantitative sensory testing is particularly useful for serial evaluation of cutaneous sensation in clinical trials. Threshold testing for touch and vibratory and thermal sensation is the most widely used application.

CORTICAL SENSATION

The most commonly used tests of cortical function are two-point discrimination, touch localization, and bilateral simultaneous stimulation and tests for graphesthesia and stereognosis. Abnormalities of these sensory tests, in the presence of normal primary sensation in an alert cooperative patient, signify a lesion of the parietal cortex or thalamocortical projections to the parietal lobe. If primary sensation is altered, these cortical discriminative functions usually will be abnormal also. Comparisons should always be made between analogous sites on the two sides of the body because the deficit with a specific parietal lesion is likely to be unilateral. Interside comparisons are important for all cortical sensory testing.

Two-point discrimination is tested with special calipers, the points of which may be set from 2 mm to several centimeters apart and then applied simultaneously to the site to be tested. The pulp of the fingertips is a common site to test; a normal individual can distinguish about 3-mm separation of points there.

Touch localization is performed by light pressure for an instant with the examiner's fingertip or a wisp of cotton wool; the patient, whose eyes are closed, is required to identify the site of touch with the fingertip. *Bilateral simultaneous stimulation* at analogous sites (e.g., the dorsum of both hands) can be carried out to determine whether the perception of touch is extinguished consistently on one side or the other. The phenomenon is referred to as *extinction or neglect* . *Graphesthesia* refers to the capacity to recognize with eyes closed letters or numbers drawn by the examiner's fingertip on the palm of the hand. Once again, interside comparison is of prime importance. Inability to recognize numbers or letters is termed *agraphesthesia*.

Stereognosis refers to the ability to identify common objects by palpation, recognizing their shape, texture, and size. Common standard objects such as keys, paper clips, and coins are best used. Patients with normal stereognosis should be able to distinguish a dime from a penny and a nickel from a quarter without looking. Patients should be allowed to feel the object with only one hand at a time. If they are unable to identify it in one hand, it should be placed in the other for comparison. Individuals who are unable to identify common objects and coins in one hand but can do so in the other are said to have *astereognosis* of the abnormal hand.

LOCALIZATION OF SENSORY ABNORMALITIES

Sensory symptoms and signs can result from lesions at almost any level of the nervous system from the parietal cortex to the peripheral sensory receptor. Noting the distribution and nature of sensory symptoms and signs is the most important way to localize their source. Their extent, configuration, symmetry, quality, and severity are the key observations.

Dysesthesias without sensory findings by examination may be difficult to interpret. To illustrate, tingling dysesthesias in an acral distribution (hands and feet) can be systemic in origin, e.g., secondary to hyperventilation, or induced by a medication such as acetazolamide. Distal dysesthesias also can be an early event in an evolving polyneuropathy or may herald a myelopathy, such as from vitamin B12 deficiency. Sometimes distal dysesthesias have no definable basis. In contrast, dysesthesias that correspond in distribution to a particular peripheral nerve territory denote a lesion of that nerve trunk. For instance, dysesthesias restricted to the fifth digit and the adjacent one-half of the fourth finger on one hand reliably point to disorder of the ulnar nerve, most commonly at the elbow.

NERVE AND ROOT

In focal nerve trunk lesions severe enough to cause a deficit, sensory abnormalities are readily mapped and generally have discrete boundaries (Figs. 23-2, 23-3 and 23-4). Root ("radicular") lesions frequently are accompanied by deep, aching pain along the course of the related nerve trunk. With compression of a fifth lumbar (L5) or first sacral (S1) root, as from a ruptured intervertebral disk, sciatica (radicular pain relating to the sciatic nerve trunk) is a common manifestation (Chap. 15). With a lesion affecting a single root, sensory deficits may be minimal or absent because adjacent root territories overlap extensively.

Isolated mononeuropathies may cause symptoms beyond the territory supplied by the affected nerve, but abnormalities on examination typically are confined to appropriate anatomic boundaries. In multiple mononeuropathies, symptoms and signs occur in discrete territories supplied by different individual nerves and—as more nerves are affected—may simulate a polyneuropathy if deficits become confluent. With polyneuropathies, sensory deficits are generally graded, distal, and symmetric in distribution (Chap. 384). Dysesthesias, followed by numbness, begin in the toes and ascend symmetrically. When dysesthesias reach the knees, they usually also have appeared in the fingertips. The process appears to be nerve length–dependent, and the deficit is often described as "stocking-glove" in type. Involvement of both hands and feet also occurs with lesions of the upper cervical cord or the brainstem, but an upper level of the sensory disturbance may then be found on the trunk and other evidence of a central lesion may be present, such as sphincter involvement or signs of an upper motor neuron lesion (Chap. 22). Although most polyneuropathies are pansensory and affect all modalities of sensation, selective sensory dysfunction according to nerve fiber size may occur. Small-fiber polyneuropathies are characterized by burning, painful dysesthesias with reduced pinprick and thermal sensation but with sparing of proprioception, motor function, and deep tendon reflexes. Touch is involved variably; when it is spared, the sensory pattern is referred to as exhibiting *sensory dissociation*. Sensory dissociation may occur with spinal cord lesions as well as small-fiber neuropathies. Large-fiber polyneuropathies are characterized by vibration and position sense deficits, imbalance, absent tendon reflexes, and variable motor dysfunction but preservation of most cutaneous sensation. Dysesthesias, if present at all, tend to be tingling or bandlike in quality.

Sensory neuronopathy is characterized by widespread but asymmetric sensory loss occurring in a non-length-dependent manner so that it may occur proximally or distally and in the arms, legs, or both. Pain and numbness progress to sensory ataxia and impairment of all sensory modalities with time. This condition is usually paraneoplastic or idiopathic in origin (Chaps. 101 and 384).

SPINAL CORD

(See also Chap. 377) If the spinal cord is transected, all sensation is lost below the level of transection. Bladder and bowel function also are lost, as is motor function. Hemisection of the spinal cord produces the Brown-Séquard syndrome, with absent pain and temperature sensation contralaterally and loss of proprioceptive sensation and power ipsilaterally below the lesion (see Figs. 23-1 and 377-1).

Numbness or paresthesias in both feet may arise from a spinal cord lesion; this is especially likely when the upper level of the sensory loss extends to the trunk. When all extremities are affected, the lesion is probably in the cervical region or brainstem unless a peripheral neuropathy is responsible. The presence of upper motor neuron signs (Chap. 22) supports a central lesion; a hyperesthetic band on the trunk may suggest the level of involvement.

A dissociated sensory loss can reflect spinothalamic tract involvement in the spinal cord, especially if the deficit is unilateral and has an upper level on the torso. Bilateral spinothalamic tract involvement occurs with lesions affecting the center of the spinal cord, such as in syringomyelia. There is a dissociated sensory loss with impairment of pinprick and temperature appreciation but relative preservation of light touch, position sense, and vibration appreciation.

Dysfunction of the posterior columns in the spinal cord or of the posterior root entry zone may lead to a bandlike sensation around the trunk or a feeling of tight pressure in one or more limbs. Flexion of the neck sometimes leads to an electric shock–like sensation that radiates down the back and into the legs (Lhermitte's sign) in patients with a cervical lesion affecting the posterior columns, such as from multiple sclerosis, cervical spondylosis, or recent irradiation to the cervical region.

BRAINSTEM

Crossed patterns of sensory disturbance, in which one side of the face and the opposite side of the body are affected, localize to the lateral medulla. Here a small lesion may damage both the ipsilateral descending trigeminal tract and the ascending spinothalamic fibers subserving the opposite arm, leg, and hemitorso (see "Lateral medullary syndrome" in Fig. 370-10). A lesion in the tegmentum of the pons and midbrain, where the lemniscal and spinothalamic tracts merge, causes pansensory loss contralaterally.

THALAMUS

Hemisensory disturbance with tingling numbness from head to foot is often thalamic in origin but also can arise from the anterior parietal region. If abrupt in onset, the lesion is likely to be due to a small stroke (lacunar infarction), particularly if localized to the thalamus. Occasionally, with lesions affecting the VPL nucleus or adjacent white matter, a syndrome of thalamic pain, also called *Déjerine-Roussy syndrome*, may ensue. The persistent, unrelenting unilateral pain often is described in dramatic terms.

CORTEX

With lesions of the parietal lobe involving either the cortex or the subjacent white matter, the most prominent symptoms are contralateral hemineglect, hemi-inattention, and a tendency not to use the affected hand and arm. On cortical sensory testing (e.g., two-point discrimination, graphesthesia), abnormalities are often found but primary sensation is usually intact. Anterior parietal infarction may present as a pseudothalamic syndrome with contralateral loss of primary sensation from head to toe. Dysesthesias or a sense of numbness may also occur and, rarely, a painful state.

FOCAL SENSORY SEIZURES

These seizures generally are due to lesions in the area of the postcentral or precentral gyrus. The principal symptom of focal sensory seizures is tingling, but additional, more complex sensations may occur, such as a rushing feeling, a sense of warmth, or a sense of movement without detectable motion. Symptoms typically are unilateral; commonly begin in the arm or hand, face, or foot; and often spread in a manner that reflects the cortical representation of different bodily parts, as in a Jacksonian march. Duration of seizures is variable; seizures may be transient, lasting only for seconds, or persist for an hour or more. Focal motor features may supervene, often becoming generalized with loss of consciousness and tonic-clonic jerking.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GAIT AND BALANCE DISORDERS: INTRODUCTION

Prevalence, Morbidity, and Mortality

Gait and balance problems are common in the elderly and contribute to the risk of falls and injury. Gait disorders have been described in 15% of individuals older than age 65 years. By age 80 years, one person in four will use a mechanical aid to assist ambulation. Among those 85 and older, the prevalence of gait abnormality approaches 40%. In epidemiologic studies, gait disorders are consistently identified as a major risk factor for falls and injury.

A substantial number of older persons report insecure balance and experience falls and fear of falling. Prospective studies indicate that 30% of those age >65 years fall each year; the proportion is even higher in frail elderly and nursing home patients. Each year, 8% of individuals age >75 years suffer a serious fall-related injury. Hip fractures often result in hospitalization and nursing home admission. For each person who is physically disabled, there are others whose functional independence is constrained by anxiety and fear of falling. Nearly one in five elderly individuals voluntarily limits activity because of fear of falling. With loss of ambulation, there is a diminished quality of life and increased morbidity and mortality rates.

Anatomy and Physiology

Upright bipedal gait depends on the successful integration of postural control and locomotion. These functions are widely distributed in the central nervous system. The biomechanics of bipedal walking are complex, and the performance is easily compromised by neurologic deficit at any level. Command and control centers in the brainstem, cerebellum, and forebrain modify the action of spinal pattern generators to promote stepping. While a form of "fictive locomotion" can be elicited from quadrupedal animals after spinal transection, this capacity is limited in primates. Step generation in primates is dependent on locomotor centers in the pontine tegmentum, midbrain, and subthalamic region. Locomotor synergies are executed through the reticular formation and descending pathways in the ventromedial spinal cord. Cerebral control provides a goal and purpose for walking and is involved in avoidance of obstacles and adaptation of locomotor programs to context and terrain.

Postural control requires the maintenance of the center of mass over the base of support through the gait cycle. Unconscious postural adjustments maintain standing balance: long latency responses are measurable in the leg muscles, beginning 110 milliseconds after a perturbation. Forward motion of the center of mass provides propulsive force for stepping, but failure to maintain the center of mass within stability limits results in falls. The anatomic substrate for dynamic balance has not been well defined, but the vestibular nucleus and midline cerebellum contribute to balance control in animals. Human patients with damage to these structures have impaired balance with standing and walking.

Standing balance depends on good-quality sensory information about the position of the body center with respect to the environment, support surface, and gravitational forces. Sensory information for postural control is primarily generated by the visual system, the vestibular system, and by proprioceptive receptors in the muscle spindles and joints. A healthy redundancy of sensory afferent information is generally available, but loss of two of the three pathways is sufficient to compromise standing balance. Balance disorders in older individuals sometimes result from multiple insults in the peripheral sensory systems (e.g., visual loss, vestibular deficit, peripheral neuropathy), critically degrading the quality of afferent information needed for balance stability.

Older patients with cognitive impairment from neurodegenerative diseases appear to be particularly prone to falls and injury. Frailty, muscle weakness, and deconditioning also contribute to the risk. It has been shown that older people who continue walking while talking are at increased risk for falls. There is a growing literature on the use of attentional resources to manage gait and balance. Walking is generally considered to be unconscious and automatic, but the ability to walk while attending a cognitive task (dual-task walking) may be compromised in frail elderly with a history of falls. Older patients with deficits in executive function may have particular difficulty in managing the attentional resources needed for dynamic balance when distracted.

Disorders of Gait

The heterogeneity of gait disorders observed in clinical practice reflects the large network of neural systems involved in the task. Walking is vulnerable to neurologic disease at every level. Gait disorders have been classified descriptively, based on the abnormal physiology and biomechanics. One problem

with this approach is that many failing gaits look fundamentally similar. This overlap reflects common patterns of adaptation to threatened balance stability and declining performance. *The gait disorder observed clinically must be viewed as the product of a neurologic deficit and a functional adaptation*. Unique features of the failing gait are often overwhelmed by the adaptive response. Some of the common patterns of abnormal gait are summarized below. Gait disorders can also be classified by etiology, as listed in **Table 24-1**.

Table 24-1 Etiology of Gait Disorders

	Cases	Percent
Sensory deficits	22	18.3
Myelopathy	20	16.7
Multiple infarcts	18	15.0
Parkinsonism	14	11.7
Cerebellar degeneration	8	6.7
Hydrocephalus	8	6.7
Toxic/metabolic	3	2.5
Psychogenic	4	3.3
Other	6	5.0
Unknown cause	17	14.2
Total	120	100%

Source: Reproduced with permission from J Masdeu et al: *Gait Disorders of Aging: With Special Reference to Falls*. Boston, Little Brown, 1995.

CAUTIOUS GAIT

The term *cautious gait* is used to describe the patient who walks with an abbreviated stride and lowered center of mass, as if walking on a slippery surface. This disorder is both common and nonspecific. It is, in essence, an adaptation to a perceived postural threat. There may be an associated fear of falling. In one study, this disorder was observed in more than one-third of older patients with a higher level gait disturbance. Physical therapy often improves walking to the degree that follow-up observation may reveal a more specific underlying disorder.

STIFF-LEGGED GAIT

Spastic gait is characterized by stiffness in the legs, an imbalance of muscle tone, and a tendency to circumduct and scuff the feet. The disorder reflects compromise of corticospinal command and overactivity of spinal reflexes. The patient may walk on his or her toes. In extreme instances, the legs cross due to increased tone in the adductors. Upper motor neuron signs are present on physical examination. Shoes often reflect an uneven pattern of wear across the outside. The disorder may be cerebral or spinal in origin.

Myelopathy from cervical spondylosis is a common cause of spastic or spastic-ataxic gait. Demyelinating disease and trauma are the leading causes of myelopathy in younger patients. In a chronic progressive myelopathy of unknown cause, workup with laboratory and imaging tests may establish a diagnosis. A family history should suggest hereditary spastic paraplegia (HSP; Chap. 374). Genetic testing is now available for some of the common HSP mutations. Tropical spastic paraparesis related to the retrovirus HTLV-I is endemic in parts of the Caribbean and South America. A structural lesion, such as tumor or spinal vascular malformation, should be excluded with appropriate testing. **Spinal cord disorders are discussed in detail in Chap. 377.**

With cerebral spasticity, asymmetry is common, involvement of the upper extremities is usually observed, and dysarthria is often an associated feature. Common causes include vascular disease (stroke), multiple sclerosis, and perinatal injury to the nervous system (cerebral palsy).

Other stiff-legged gaits include dystonia (Chap. 387) and stiff-person syndrome. Dystonia is a disorder characterized by sustained muscle contractions, resulting in repetitive twisting movements and abnormal posture. It often has a genetic basis. Dystonic spasms produce plantar flexion and inversion of the feet, sometimes with torsion of the trunk. In autoimmune stiff-person syndrome (Chap. 101), there is exaggerated lordosis of the lumbar spine and overactivation of antagonist muscles, which restricts trunk and lower limb movement and results in a wooden or fixed posture.

PARKINSONISM AND FREEZING GAIT

Parkinson's disease (Chap. 372) is common, affecting 1% of the population age >55 years. The stooped posture and shuffling gait are characteristic and distinctive features. Patients sometimes accelerate (festinate) with walking or display retropulsion. There may be difficulty with gait initiation (freezing) and a tendency to turn en bloc. Imbalance and falls may develop as the disease progresses

over years. Gait freezing is described in 7% of Parkinson's patients within 2 years of onset, 26% by the end of 5 years. Freezing of gait is even more common in some of the Parkinson's-related neurodegenerative disorders, such as progressive supranuclear palsy, multiple-system atrophy, and corticobasal degeneration. These patients frequently present with axial stiffness, postural instability, and a shuffling gait while lacking the characteristic pill-rolling tremor of Parkinson's disease. Falls within the first year suggest the possibility of progressive supranuclear palsy.

Hyperkinetic movement disorders also produce characteristic and recognizable disturbances in gait. In Huntington's disease (Chap. 371), the unpredictable occurrence of choreic movements gives the gait a dancing quality. Tardivedyskinesia is the cause of many odd, stereotypic gait disorders seen in patients chronically exposed to antipsychotics and other drugs that block the D_2 dopamine receptor.

FRONTAL GAIT DISORDER

Frontal gait disorder, sometimes known as "gait apraxia," is common in the elderly and has a variety of causes. The term is used to describe a shuffling, freezing gait with imbalance and other signs of higher cerebral dysfunction. Typical features include a wide base of support, short stride, shuffling along the floor, and difficulty with starts and turns. Many patients exhibit difficulty with gait initiation, descriptively characterized as the "slipping clutch" syndrome. The term *lower body parkinsonism* is also used to describe such patients. Strength is generally preserved, and patients are able to make stepping movements when not standing and maintaining balance at the same time. This disorder is best considered a higher level motor control disorder, as opposed to an apraxia (Chap. 26).

The most common cause of frontal gait disorder is vascular disease, particularly subcortical small-vessel disease. Lesions are frequently found in the deep frontal white matter and centrum ovale. Gait disorder may be the salient feature in hypertensive patients with ischemic lesions of the deep hemisphere white matter (Binswanger's disease). The clinical syndrome includes mental change (variable in degree), dysarthria, pseudobulbar affect (emotional disinhibition), increased tone, and hyperreflexia in the lower limbs.

Communicating hydrocephalus in adults also presents with a gait disorder of this type. Other features of the diagnostic triad (mental change, incontinence) may be absent in the initial stages. MRI demonstrates ventricular enlargement, an enlarged flow void about the aqueduct, and a variable degree of periventricular white matter change. A lumbar puncture or dynamic test is necessary to confirm the presence of hydrocephalus.

CEREBELLAR GAIT ATAXIA

Disorders of the cerebellum have a dramatic impact on gait and balance. Cerebellar gait ataxia is characterized by a wide base of support, lateral instability of the trunk, erratic foot placement, and decompensation of balance when attempting to walk tandem. Difficulty maintaining balance when turning is often an early feature. Patients are unable to walk tandem heel to toe, and display truncal sway in narrow-based or tandem stance. They show considerable variation in their tendency to fall in daily life.

Causes of cerebellar ataxia in older patients include stroke, trauma, tumor, and neurodegenerative disease, including multiple-system atrophy (Chaps. 372 and 375) and various forms of hereditary cerebellar degeneration (Chap. 373). A short expansion at the site of the fragile X mutation (fragile X pre-mutation) has been associated with gait ataxia in older men. Alcoholic cerebellar degeneration can be screened by history and often confirmed by MRI. In patients with ataxia, MRI demonstrates the extent and topography of cerebellar atrophy.

SENSORY ATAXIA

As reviewed above, balance depends on high-quality afferent information from the visual and the vestibular systems and proprioception. When this information is lost or degraded, balance during locomotion is impaired and instability results. The sensory ataxia of tabetic neurosyphilis is a classic example. The contemporary equivalent is the patient with neuropathy affecting large fibers. Vitamin B12 deficiency is a treatable cause of large-fiber sensory loss in the spinal cord and peripheral nervous system. Joint position and vibration sense are diminished in the lower limbs. The stance in such patients is destabilized by eye closure; they often look down at their feet when walking and do poorly in the dark. Patients have been described with imbalance from bilateral vestibular loss, caused by disease or by exposure to ototoxic drugs. **Table 24-2** compares sensory ataxia with cerebellar ataxia and frontal gait disorder. Some frail older patients exhibit a syndrome of imbalance from the combined effect of multiple sensory deficits. Such patients have disturbances in proprioception, vision, and vestibular sense that impair postural support.

Table 24-2 Features of Cerebellar Ataxia, Sensory Ataxia, and Frontal Gait Disorders

	Cerebellar Ataxia	Sensory Ataxia	Frontal Gait
Base of support	Wide-based	Narrow base, looks	Wide-based

		down	
Velocity	Variable	Slow	Very slow
Stride	Irregular, lurching	Regular with path deviation	Short, shuffling
Romberg	+/–	Unsteady, falls	+/–
Heel → shin	Abnormal	+/–	Normal
Initiation	Normal	Normal	Hesitant
Turns	Unsteady	+/–	Hesitant, multistep
Postural instability	+	+++	++++
			Poor postural synergies getting up from a chair
Falls	Late event	Frequent	Frequent

NEUROMUSCULAR DISEASE

Patients with neuromuscular disease often have an abnormal gait, occasionally as a presenting feature. With distal weakness (peripheral neuropathy) the step height is increased to compensate for footdrop, and the sole of the foot may slap on the floor during weight acceptance. Neuropathy may be associated with a degree of sensory imbalance, as described above. Patients with myopathy or muscular dystrophy more typically exhibit proximal weakness. Weakness of the hip girdle may result in a degree of excess pelvic sway during locomotion.

TOXIC AND METABOLIC DISORDERS

Alcohol intoxication is the most common cause of acute walking difficulty. Chronic toxicity from medications and metabolic disturbances can impair motor function and gait. Mental status changes may be present, and examination may reveal asterixis or myoclonus. Static equilibrium is disturbed, and such patients are easily thrown off balance. Disequilibrium is particularly evident in patients with chronic renal disease and those with hepatic failure, in whom asterixis may impair postural support. Sedative drugs, especially neuroleptics and long-acting benzodiazepines, affect postural control and increase the risk for falls. These disorders are important to recognize because they are often treatable.

PSYCHOGENIC GAIT DISORDER

Psychogenic disorders are common in neurologic practice, and the presentation often involves gait. Some patients with extreme anxiety or phobia walk with exaggerated caution with abduction of the arms, as if walking on ice. This inappropriately overcautious gait differs in degree from the gait of the patient who is insecure and making adjustments for imbalance. Depressed patients exhibit primarily slowness, a manifestation of psychomotor retardation, and lack of purpose in their stride. Hysterical gait disorders are among the most spectacular encountered. Odd gyrations of posture with wastage of muscular energy (astasia-abasia), extreme slow motion, and dramatic fluctuations over time may be observed in patients with somatoform disorders and conversion reaction.

Approach to the Patient: Slowly Progressive Disorder of Gait

When reviewing the history, it is helpful to inquire about the onset and progression of disability. Initial awareness of an unsteady gait often follows a fall. Stepwise evolution or sudden progression suggests vascular disease. Gait disorder may be associated with urinary urgency and incontinence, particularly in patients with cervical spine disease or hydrocephalus. It is always important to review the use of alcohol and medications that affect gait and balance. Information on localization derived from the neurologic examination can be helpful to narrow the list of possible diagnoses.

Gait observation provides an immediate sense of the patient's degree of disability. Characteristic patterns of abnormality are sometimes observed, though failing gaits often look fundamentally similar. Cadence (steps/min), velocity, and stride length can be recorded by timing a patient over a fixed distance. Watching the patient get out of a chair provides a good functional assessment of balance.

Brain imaging studies may be informative in patients with an undiagnosed disorder of gait. MRI is sensitive for cerebral lesions of vascular or demyelinating disease and is a good screening test for occult hydrocephalus. Patients with recurrent falls are at risk for subdural hematoma. Many elderly patients with gait and balance difficulty have white matter abnormalities in the periventricular region and centrum semiovale. While these lesions may be an incidental finding, a substantial burden of white matter disease will ultimately impact cerebral control of locomotion.

Disorders of Balance

Balance is the ability to maintain equilibrium: a state in which opposing physical forces cancel. In physiology, this is taken to mean the ability to control the center of mass with respect to gravity and the support surface. In reality, we are not consciously aware of what or where our center of mass is, but everyone, including gymnasts, figure skaters, and platform divers, moves so as to manage it.

Imbalance implies a disturbance of equilibrium. Disorders of balance present with difficulty maintaining posture standing and walking and with a subjective sense of disequilibrium, a form of dizziness.

The cerebellum and vestibular system organize antigravity responses needed to maintain the upright posture. As reviewed above, these responses are physiologically complex, and the anatomic representation is not well understood. Failure, resulting in disequilibrium, can occur at several levels: cerebellar, vestibular, somatosensory, and higher level disequilibrium. Patients with hereditary ataxia or alcoholic cerebellar degeneration do not generally complain of dizziness, but balance is visibly impaired. Neurologic examination will reveal a variety of cerebellar signs. Postural compensation may prevent falls early on, but falls inevitably occur with disease progression. The progression of a neurodegenerative ataxia is often measured by the number of years to loss of stable ambulation. Vestibular disorders (Chap. 21) have symptoms and signs in three categories: (1) vertigo, the subjective appreciation or illusion of movement; (2) nystagmus, a vestibulo-oculomotor sign; and (3) poor standing balance, an impairment of vestibulospinal function. Not every patient has all manifestations. Patients with vestibular deficits related to ototoxic drugs may lack vertigo or obvious nystagmus, but balance is impaired on standing and walking, and the patient cannot navigate in the dark. Laboratory testing is available to explore vestibulo-oculomotor and vestibulospinal deficits.

Somatosensory deficits also produce imbalance and falls. There is often a subjective sense of insecure balance and fear of falling. Postural control is compromised by eye closure (Romberg's sign); these patients also have difficulty navigating in the dark. A dramatic example is the patient with autoimmune subacute sensory neuropathy, sometimes a paraneoplastic disorder (Chap. 101). Compensatory strategies enable such patients to walk in the virtual absence of proprioception, but the task requires active visual monitoring. Patients with higher level disorders of equilibrium have difficulty maintaining balance in daily life and may present with falls. There may be reduced awareness of balance impairment. Classic examples include patients with progressive supranuclear palsy and normal pressure hydrocephalus. Patients on sedating medications are also in this category. In prospective studies, cognitive impairment and the use of sedative medications substantially increase the risk for falls.

Falls

Falls are common in the elderly; 30% of people older than age 65 years living in the community fall each year. Modest changes in balance function have been described in fit older subjects as a result of normal aging. Subtle deficits in sensory systems, attention, and motor reaction time contribute to the risk, and environmental hazards abound. Epidemiologic studies have identified a number of risk factors for falls, summarized in **Table 24-3**. A fall is not a neurologic problem, nor reason for referral to a specialist, but there are circumstances in which neurologic evaluation is appropriate. In a classic study, 90% of fall events occurred among 10% of individuals, a group known as *recurrent fallers*. Some of these are frail older persons with chronic diseases. Recurrent falls sometimes indicate the presence of serious balance impairment. Syncope, seizure, or falls related to loss of consciousness require appropriate evaluation and treatment (Chaps. 20 and 369).

Table 24-3 Risk Factors for Falls, a Meta-Analysis: Summary of Sixteen Controlled Studies

Risk Factor	Mean RR (OR)	Range
Weakness	4.9	1.9–10.3
Balance deficit	3.2	1.6–5.4
Gait disorder	3.0	1.7–4.8
Visual deficit	2.8	1.1–7.4
Mobility limitation	2.5	1.0–5.3
Cognitive impairment	2.4	2.0–4.7
Impaired functional status	2.0	1.0–3.1
Postural hypotension	1.9	1.0–3.4

Abbreviations: OR, odds ratios from retrospective studies; RR, relative risks from prospective studies.

Source: Reproduced with permission from J Masdeu et al: *Gait Disorders of Aging: With Special Reference to Falls*. Boston, Little Brown, 1995.

The descriptive classification of falls is as difficult as the classification of gait disorders, for many of the same reasons. Postural control systems are widely distributed, and a number of disease-related abnormalities occur. Unlike gait problems that are apparent on observation, falls are rarely observed in the office. The patient and family may have limited information about what triggered the fall. Injuries can complicate the physical examination. While there is no standard nosology of falls, common patterns can be identified.

SLIPPING, TRIPPING, AND "MECHANICAL FALLS"

Slipping on icy pavement, tripping on obstacles, and falls related to obvious environmental factors are often termed *mechanical falls*. They occasionally occur in healthy individuals with good balance compensation. Frequent tripping falls raise suspicion about an underlying neurologic deficit. Patients with spasticity, leg weakness, or footdrop experience tripping falls.

WEAKNESS AND FRAILTY

Patients who lack strength in antigravity muscles have difficulty rising from a chair, fatigue easily when walking, and have difficulty maintaining their balance after a perturbation. These patients are often unable to get up after a fall and may be on the floor for an hour or more before help arrives. Deconditioning of this sort is often treatable. Resistance strength training can increase muscle mass and leg strength in people in their eighties and nineties.

DROP ATTACKS AND COLLAPSING FALLS

Drop attacks are sudden collapsing falls without loss of consciousness. Patients who collapse from lack of postural tone present a diagnostic challenge. The patient may report that his or her legs just gave out underneath; the family may describe the patient as "collapsing in a heap." Orthostatic hypotension may be a factor in some such falls. Asterixis or epilepsy may impair postural support. A colloid cyst of the third ventricle can present with intermittent obstruction of the foramen of Monroe, resulting in a drop attack. While collapsing falls are more common in older patients with vascular risk factors, they should not be confused with vertebrobasilar ischemic attacks.

TOPPLING FALLS

Some patients maintain tone in antigravity muscles but fall over like a tree trunk, as if postural defenses had disengaged. There may be a consistent direction to such falls. The patient with cerebellar pathology may lean and topple over toward the side of the lesion. Patients with lesions of the vestibular system or its central pathways may experience lateral pulsion and toppling falls. Patients with progressive supranuclear palsy often fall over backward. Falls of this nature occur in patients with advanced Parkinson's disease once postural instability has developed.

GAIT FREEZING

Another fall pattern in Parkinson's disease and related disorders is the fall due to freezing of gait. The feet stick to the floor and the center of mass keeps moving, resulting in a disequilibrium from which the patient has difficulty recovering. This sequence of events can result in a forward fall. Gait freezing can also occur as the patient attempts to turn and change direction. Similarly, the patient with Parkinson's disease and festinating gait may find his feet unable to keep up, resulting in a forward fall.

FALLS RELATED TO SENSORY DEFICIT

Patients with somatosensory, visual, or vestibular deficits are prone to falls. These patients have particular difficulty dealing with poor illumination or walking on uneven ground. These patients often express subjective imbalance, apprehension, and fear of falling. Deficits in joint position and vibration sense are apparent on physical examination.

Treatment: Interventions to Reduce the Risk of Falls and Injury

Efforts should be made to define the etiology of the gait disorder and mechanism of the falls. Standing blood pressure should be recorded. Specific treatment may be possible, once a diagnosis is established. Therapeutic intervention is often recommended for older patients at substantial risk for falls, even if no neurologic disease is identified. A home visit to look for environmental hazards can be helpful. A variety of modifications may be recommended to improve safety, including improved lighting and the installation of grab bars and nonslip surfaces.

Rehabilitation interventions attempt to improve muscle strength and balance stability and to make the patient more resistant to injury. High-intensity resistance strength training with weights and machines is useful to improve muscle mass, even in frail older patients. Improvements are realized in posture and gait, which should translate to reduced risk of falls and injury. Sensory balance training is another approach to improve balance stability. Measurable gains can be achieved in a few weeks of training, and benefits can be maintained over 6 months by a 10- to 20-min home exercise program. This strategy is particularly successful in patients with vestibular and somatosensory balance disorders. The Yale Health and Aging study used a strategy of targeted, multiple risk factor abatement to reduce falls in the elderly. Prescription medications were adjusted, and home-based exercise programs were tailored to the patients' needs, based on an initial geriatric assessment. The program realized a 44% reduction in falls, compared with a control group of patients who had periodic social visits.

FURTHER READINGS

BEAUCHET O et al: Stops walking when talking: A predictor of falls in older adults? Eur J Neurol 16:786, 2009 [PMID: 19473368]

BRONSTEIN A et al: *Clinical Disorders of Balance, Posture and Gait* . London, Arnold Press, 2003

DE LAAT KF et al: Loss of white matter integrity is associated with gait disorders in cerebral small vessel disease. Brain 134:73, 2011

GILLESPIE LD et al: Interventions for preventing falls in older people living in the community. Cochrane Rev CD007146, 2009

NUTT JG et al: Human walking and higher-level gait disorders. Neurol 43:268, 1993 [PMID: 8437689]

SNIJDERS AH et al: Neurological gait disorders in elderly people: Clinical approach and classification. Lancet Neurol 6:63, 2007 [PMID: 17166803]

SPRINGER S et al: Dual-tasking effects on gait variability: The role of aging, falls, and executive function. Mov Disord 21:950, 2006 [PMID: 16541455]

SUDARSKY L: Gait disorders in the elderly. N Engl J Med 322:1441, 1990 [PMID: 2184358]

———: Psychogenic gait disorders. Semin Neurol 26:351, 2006

TINETTI ME, KUMAR C: The patient who falls: "It's always a trade-off." JAMA 20:258, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VIDEO LIBRARY OF GAIT DISORDERS

Problems with gait and balance are major causes of falls, accidents, and resulting disability, especially in later life, and are often harbingers of neurologic disease. Early diagnosis is essential, especially for treatable conditions, as it may permit the institution of prophylactic measures to prevent dangerous falls, and also to reverse or ameliorate the underlying cause. In this video, examples of gait disorders due to Parkinson's disease, other extrapyramidal disorders, and ataxias, as well as other common gait disorders, are presented.

Video e8-1
Video Library of Gait Disorders

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CONFUSION AND DELIRIUM: INTRODUCTION

Confusion, a mental and behavioral state of reduced comprehension, coherence, and capacity to reason, is one of the most common problems encountered in medicine, accounting for a large number of emergency department visits, hospital admissions, and inpatient consultations. *Delirium*, a term used to describe an acute confusional state, remains a major cause of morbidity and mortality rates, costing billions of dollars yearly in health care costs in the United States alone. Delirium often goes unrecognized despite clear evidence that it is usually the cognitive manifestation of serious underlying medical or neurologic illness.

Clinical Features of Delirium

A multitude of terms are used to describe delirium, including *encephalopathy, acute brain failure, acute confusional state,* and *postoperative or intensive care unit (ICU) psychosis* . Delirium has many clinical manifestations, but essentially it is defined as a relatively acute decline in cognition that fluctuates over hours or days. The hallmark of delirium is a deficit of attention, although all cognitive domains—including memory, executive function, visuospatial tasks, and language—are variably involved. Associated symptoms may include altered sleep-wake cycles, perceptual disturbances such as hallucinations or delusions, affect changes, and autonomic findings that include heart rate and blood pressure instability.

Delirium is a clinical diagnosis that can be made only at the bedside. Two broad clinical categories have been described—the hyperactive and hypoactive subtypes—that are based on differential psychomotor features. The cognitive syndrome associated with severe alcohol withdrawal remains the classic example of the hyperactive subtype, featuring prominent hallucinations, agitation, and hyperarousal, often accompanied by life-threatening autonomic instability. In striking contrast is the hypoactive subtype, exemplified by opiate intoxication, in which patients are withdrawn and quiet, with prominent apathy and psychomotor slowing.

This dichotomy between subtypes of delirium is a useful construct, but patients often fall somewhere along a spectrum between the hyperactive and hypoactive extremes, sometimes fluctuating from one to the other within minutes. Therefore, clinicians must recognize the broad range of presentations of delirium to identify all patients with this potentially reversible cognitive disturbance. Hyperactive patients, such as those with delirium tremens, are easily recognized by their characteristic severe agitation, tremor, hallucinations, and autonomic instability. Patients who are quietly disturbed are overlooked more often on the medical wards and in the ICU, yet multiple studies suggest that this underrecognized hypoactive subtype is associated with worse outcomes.

The reversibility of delirium is emphasized because many etiologies, such as systemic infection and medication effects, can be treated easily. However, the long-term cognitive effects of delirium remain largely unknown and understudied. Some episodes of delirium continue for weeks, months, or even years. The persistence of delirium in some patients and its high recurrence rate may be due to inadequate treatment of the underlying etiology of the syndrome. In some instances, delirium does not disappear because there is underlying permanent neuronal damage. Even after an episode of delirium resolves, there may be lingering effects of the disorder. A patient's recall of events after delirium varies widely, ranging from complete amnesia to repeated reexperiencing of the frightening period of confusion in a disturbing manner, similar to what is seen in patients with posttraumatic stress disorder.

Risk Factors

An effective primary prevention strategy for delirium begins with identification of patients at highest risk, including those preparing for elective surgery or being admitted to the hospital. Although no single validated scoring system has been widely accepted as a screen for asymptomatic patients, there are multiple well-established risk factors for delirium.

The two most consistently identified risks are older age and baseline cognitive dysfunction. Individuals who are over age 65 or exhibit low scores on standardized tests of cognition develop delirium upon hospitalization at a rate approaching 50%. Whether age and baseline cognitive dysfunction are truly independent risk factors is uncertain. Other predisposing factors include sensory deprivation, such as preexisting hearing and visual impairment, as well as indices for poor overall health, including baseline immobility, malnutrition, and underlying medical or neurologic illness.

In-hospital risks for delirium include the use of bladder catheterization, physical restraints, sleep and

sensory deprivation, and the addition of three or more new medications. Avoiding such risks remains a key component of delirium prevention as well as treatment. Surgical and anesthetic risk factors for the development of postoperative delirium include specific procedures such as those involving cardiopulmonary bypass and inadequate or excessive treatment of pain in the immediate postoperative period.

The relationship between delirium and dementia (Chap. 371) is complicated by significant overlap between the two conditions, and it is not always simple to distinguish between them. Dementia and preexisting cognitive dysfunction serve as major risk factors for delirium, and at least two-thirds of cases of delirium occur in patients with coexisting underlying dementia. A form of dementia with parkinsonism, termed *dementia with Lewy bodies*, is characterized by a fluctuating course, prominent visual hallucinations, parkinsonism, and an attentional deficit that clinically resembles hyperactive delirium. Delirium in the elderly often reflects an insult to the brain that is vulnerable due to an underlying neurodegenerative condition. Therefore, the development of delirium sometimes heralds the onset of a previously unrecognized brain disorder.

Epidemiology

Delirium is a common disease, but its reported incidence has varied widely with the criteria used to define the disorder. Estimates of delirium in hospitalized patients range from 14 to 56%, with higher rates reported for elderly patients and patients undergoing hip surgery. Older patients in the ICU have especially high rates of delirium that range from 70 to 87%. The condition is not recognized in up to one-third of delirious inpatients, and the diagnosis is especially problematic in the ICU environment, where cognitive dysfunction is often difficult to appreciate in the setting of serious systemic illness and sedation. Delirium in the ICU should be viewed as an important manifestation of organ dysfunction not unlike liver, kidney, or heart failure. Outside the acute hospital setting, delirium occurs in nearly two-thirds of patients in nursing homes and in over 80% of those at the end of life. These estimates emphasize the remarkably high frequency of this cognitive syndrome in older patients, a population expected to grow in the upcoming decade with the aging of the "baby boom" generation.

In previous decades an episode of delirium was viewed as a transient condition that carried a benign prognosis. Delirium now has been clearly associated with substantial morbidity rate and increased mortality rate and increasingly is recognized as a sign of serious underlying illness. Recent estimates of in-hospital mortality rates among delirious patients have ranged from 25 to 33%, a rate similar to that of patients with sepsis. Patients with an in-hospital episode of delirium have a higher mortality rate in the months and years after their illness compared with age-matched nondelirious hospitalized patients. Delirious hospitalized patients have a longer length of stay, are more likely to be discharged to a nursing home, and are more likely to experience subsequent episodes of delirium; as a result, this condition has enormous economic implications.

Pathogenesis

The pathogenesis and anatomy of delirium are incompletely understood. The attentional deficit that serves as the neuropsychological hallmark of delirium appears to have a diffuse localization with the brainstem, thalamus, prefrontal cortex, and parietal lobes. Rarely, focal lesions such as ischemic strokes have led to delirium in otherwise healthy persons; right parietal and medial dorsal thalamic lesions have been reported most commonly, pointing to the relevance of these areas to delirium pathogenesis. In most cases, delirium results from widespread disturbances in cortical and subcortical regions rather than a focal neuroanatomic cause. Electroencephalogram (EEG) data in persons with delirium usually show symmetric slowing, a nonspecific finding that supports diffuse cerebral dysfunction.

Deficiency of acetylcholine often plays a key role in delirium pathogenesis. Medications with anticholinergic properties can precipitate delirium in susceptible individuals, and therapies designed to boost cholinergic tone such as cholinesterase inhibitors have, in small trials, been shown to relieve symptoms of delirium. Dementia patients are susceptible to episodes of delirium, and those with Alzheimer's pathology are known to have a chronic cholinergic deficiency state due to degeneration of acetylcholine-producing neurons in the basal forebrain. Another common dementia associated with decreased acetylcholine levels, dementia with Lewy bodies, clinically mimics delirium in some patients. Other neurotransmitters are also likely to be involved in this diffuse cerebral disorder. For example, increases in dopamine can also lead to delirium. Patients with Parkinson's disease treated with dopaminergic medications can develop a delirium-like state that features visual hallucinations, fluctuations, and confusion. In contrast, reducing dopaminergic tone with dopamine antagonists such as typical and atypical antipsychotic medications has long been recognized as effective symptomatic treatment in patients with delirium.

Not all individuals exposed to the same insult will develop signs of delirium. A low dose of an anticholinergic medication may have no cognitive effects on a healthy young adult but may produce a florid delirium in an elderly person with known underlying dementia. However, an extremely high dose of the same anticholinergic medication may lead to delirium even in healthy young persons. This concept of delirium developing as the result of an insult in predisposed individuals is currently the most widely accepted pathogenic construct. Therefore, if a previously healthy individual with no known

history of cognitive illness develops delirium in the setting of a relatively minor insult such as elective surgery or hospitalization, an unrecognized underlying neurologic illness such as a neurodegenerative disease, multiple previous strokes, or another diffuse cerebral cause should be considered. In this context, delirium can be viewed as the symptom resulting from a "stress test for the brain" induced by the insult. Exposure to known inciting factors such as systemic infection and offending drugs can unmask a decreased cerebral reserve and herald a serious underlying and potentially treatable illness.

Approach to the Patient: Delirium

As the diagnosis of delirium is clinical and is made at the bedside, a careful history and physical examination is necessary in evaluating patients with possible confusional states. Screening tools can aid physicians and nurses in identifying patients with delirium, including the Confusion Assessment Method (CAM); the Organic Brain Syndrome Scale; the Delirium Rating Scale; and, in the ICU, the Delirium Detection Score and the ICU version of the CAM. Using the CAM, a diagnosis of delirium is made if there is (1) an acute onset and fluctuating course and (2) inattention accompanied by either (3) disorganized thinking or (4) an altered level of consciousness. These scales are based on criteria from the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM) or the World Health Organization's International Classification of Diseases (ICD). Unfortunately, these scales do not identify the full spectrum of patients with delirium. All patients who are acutely confused should be presumed delirious regardless of their presentation due to the wide variety of possible clinical features. A course that fluctuates over hours or days and may worsen at night (termed *sundowning*) is typical but not essential for the diagnosis. Observation of the patient usually will reveal an altered level of consciousness or a deficit of attention. Other hallmark features that may be present in a delirious patient include alteration of sleep-wake cycles, thought disturbances such as hallucinations or delusions, autonomic instability, and changes in affect.

HISTORY

It may be difficult to elicit an accurate history in delirious patients who have altered levels of consciousness or impaired attention. Information from a collateral source such as a spouse or another family member is therefore invaluable. The three most important pieces of history are the patient's baseline cognitive function, the time course of the present illness, and current medications.

Premorbid cognitive function can be assessed through the collateral source or, if needed, via a review of outpatient records. Delirium by definition represents a change that is relatively acute, usually over hours to days, from a cognitive baseline. As a result, an acute confusional state is nearly impossible to diagnose without some knowledge of baseline cognitive function. Without this information, many patients with dementia or depression may be mistaken as delirious during a single initial evaluation. Patients with a more hypoactive, apathetic presentation with psychomotor slowing may be identified as being different from baseline only through conversations with family members. A number of validated instruments have been shown to diagnose cognitive dysfunction accurately by using a collateral source, including the modified Blessed Dementia Rating Scale and the Clinical Dementia Rating (CDR). Baseline cognitive impairment is common in patients with delirium. Even when no such history of cognitive impairment is elicited, there should still be a high suspicion for a previously unrecognized underlying neurologic disorder.

Establishing the time course of cognitive change is important not only to make a diagnosis of delirium but also to correlate the onset of the illness with potentially treatable etiologies such as recent medication changes or symptoms of systemic infection.

Medications remain a common cause of delirium, especially compounds with anticholinergic or sedative properties. It is estimated that nearly one-third of all cases of delirium are secondary to medications, especially in the elderly. Medication histories should include all prescription as well as over-the-counter and herbal substances taken by the patient and any recent changes in dosing or formulation, including substitution of generics for brand-name medications.

Other important elements of the history include screening for symptoms of organ failure or systemic infection, which often contributes to delirium in the elderly. A history of illicit drug use, alcoholism, or toxin exposure is common in younger delirious patients. Finally, asking the patient and collateral source about other symptoms that may accompany delirium, such as depression and hallucinations, may help identify potential therapeutic targets.

PHYSICAL EXAMINATION

The general physical examination in a delirious patient should include a careful screening for signs of infection such as fever, tachypnea, pulmonary consolidation, heart murmur, and stiff neck. The patient's fluid status should be assessed; both dehydration and fluid overload with resultant hypoxemia have been associated with delirium, and each is usually easily rectified. The appearance of the skin can be helpful, showing jaundice in hepatic encephalopathy, cyanosis in hypoxemia, or needle tracks in patients using intravenous drugs.

The neurologic examination requires a careful assessment of mental status. Patients with delirium often present with a fluctuating course; therefore, the diagnosis can be missed when one relies on a

single time point of evaluation. Some but not all patients exhibit the characteristic pattern of sundowning, a worsening of their condition in the evening. In these cases, assessment only during morning rounds may be falsely reassuring.

An altered level of consciousness ranging from hyperarousal to lethargy to coma is present in most patients with delirium and can be assessed easily at the bedside. In a patient with a relatively normal level of consciousness, a screen for an attentional deficit is in order, as this deficit is the classic neuropsychological hallmark of delirium. Attention can be assessed while taking a history from the patient. Tangential speech, a fragmentary flow of ideas, or inability to follow complex commands often signifies an attentional problem. There are formal neuropsychological tests to assess attention, but a simple bedside test of digit span forward is quick and fairly sensitive. In this task, patients are asked to repeat successively longer random strings of digits beginning with two digits in a row. Average adults can repeat a string of five to seven digits before faltering; a digit span of four or less usually indicates an attentional deficit unless hearing or language barriers are present.

More formal neuropsychological testing can be extraordinarily helpful in assessing a delirious patient, but it is usually too cumbersome and time-consuming in the inpatient setting. A simple Mini Mental Status Examination (MMSE) (see Table 371-5) can provide some information regarding orientation, language, and visuospatial skills; however, performance of some tasks on the MMSE such as spelling "world" backward and serial subtraction of digits will be impaired by delirious patients' attentional deficits alone and are therefore unreliable.

The remainder of the screening neurologic examination should focus on identifying new focal neurologic deficits. Focal strokes or mass lesions in isolation are rarely the cause of delirium, but patients with underlying extensive cerebrovascular disease or neurodegenerative conditions may not be able to cognitively tolerate even relatively small new insults. Patients also should be screened for additional signs of neurodegenerative conditions such as parkinsonism, which is seen not only in idiopathic Parkinson's disease but also in other dementing conditions such as Alzheimer's disease, dementia with Lewy bodies, and progressive supranuclear palsy. The presence of multifocal myoclonus or asterixis on the motor examination is nonspecific but usually indicates a metabolic or toxic etiology of the delirium.

ETIOLOGY

Some etiologies can be easily discerned through a careful history and physical examination, whereas others require confirmation with laboratory studies, imaging, or other ancillary tests. A large, diverse group of insults can lead to delirium, and the cause in many patients is often multifactorial. Common etiologies are listed in Table 25-1.

Table 25-1 Common Etiologies of Delirium

Toxins
Prescription medications: especially those with anticholinergic properties, narcotics, and benzodiazepines
Drugs of abuse: alcohol intoxication and alcohol withdrawal, opiates, ecstasy, LSD, GHB, PCP, ketamine, cocaine
Poisons: inhalants, carbon monoxide, ethylene glycol, pesticides
Metabolic conditions
Electrolyte disturbances: hypoglycemia, hyperglycemia, hyponatremia, hypernatremia, hypercalcemia, hypocalcemia, hypomagnesemia
Hypothermia and hyperthermia
Pulmonary failure: hypoxemia and hypercarbia
Liver failure/hepatic encephalopathy
Renal failure/uremia
Cardiac failure
Vitamin deficiencies: B_{12}, thiamine, folate, niacin
Dehydration and malnutrition
Anemia
Infections
Systemic infections: urinary tract infections, pneumonia, skin and soft tissue infections, sepsis
CNS infections: meningitis, encephalitis, brain abscess
Endocrinologic conditions
Hyperthyroidism, hypothyroidism
Hyperparathyroidism

Adrenal insufficiency
Cerebrovascular disorders
Global hypoperfusion states
Hypertensive encephalopathy
Focal ischemic strokes and hemorrhages: especially nondominant parietal and thalamic lesions
Autoimmune disorders
CNS vasculitis
Cerebral lupus
Seizure-related disorders
Nonconvulsive status epilepticus
Intermittent seizures with prolonged postictal states
Neoplastic disorders
Diffuse metastases to the brain
Gliomatosis cerebri
Carcinomatous meningitis
Hospitalization
Terminal end-of-life delirium

Abbreviations: LSD, lysergic acid diethylamide; GHB, γ-hydroxybutyrate; PCP, phencyclidine; CNS, central nervous system.

Prescribed, over-the-counter, and herbal medications are common precipitants of delirium. Drugs with anticholinergic properties, narcotics, and benzodiazepines are especially common offenders, but nearly any compound can lead to cognitive dysfunction in a predisposed patient. Whereas an elderly patient with baseline dementia may become delirious upon exposure to a relatively low dose of a medication, less susceptible individuals may become delirious only with very high doses of the same medication. This observation emphasizes the importance of correlating the timing of recent medication changes, including dose and formulation, with the onset of cognitive dysfunction.

In younger patients especially, illicit drugs and toxins are common causes of delirium. In addition to more classic drugs of abuse, the recent rise in availability of so-called club drugs, such as methylenedioxymethamphetamine (MDMA, ecstasy), γ-hydroxybutyrate (GHB), and the phencyclidine (PCP)-like agent ketamine, has led to an increase in delirious young persons presenting to acute care settings. Many common prescription drugs such as oral narcotics and benzodiazepines are often abused and readily available on the street. Alcohol intoxication with high serum levels can cause confusion, but more commonly it is withdrawal from alcohol that leads to a classic hyperactive delirium. Alcohol and benzodiazepine withdrawal should be considered in all cases of delirium as even patients who drink only a few servings of alcohol every day can experience relatively severe withdrawal symptoms upon hospitalization.

Metabolic abnormalities such as electrolyte disturbances of sodium, calcium, magnesium, or glucose can cause delirium, and mild derangements can lead to substantial cognitive disturbances in susceptible individuals. Other common metabolic etiologies include liver and renal failure, hypercarbia and hypoxemia, vitamin deficiencies of thiamine and B_{12}, autoimmune disorders including central nervous system (CNS) vasculitis, and endocrinopathies such as thyroid and adrenal disorders.

Systemic infections often cause delirium, especially in the elderly. A common scenario involves the development of an acute cognitive decline in the setting of a urinary tract infection in a patient with baseline dementia. Pneumonia, skin infections such as cellulitis, and frank sepsis also can lead to delirium. This so-called septic encephalopathy, often seen in the ICU, is probably due to the release of proinflammatory cytokines and their diffuse cerebral effects. CNS infections such as meningitis, encephalitis, and abscess are less common etiologies of delirium; however, in light of the high mortality rates associated with these conditions when they are not treated quickly, clinicians must always maintain a high index of suspicion.

In some susceptible individuals, exposure to the unfamiliar environment of a hospital can lead to delirium. This etiology usually occurs as part of a multifactorial delirium and should be considered a diagnosis of exclusion after all other causes have been thoroughly investigated. Many primary prevention and treatment strategies for delirium involve relatively simple methods to address the aspects of the inpatient setting that are most confusing.

Cerebrovascular etiologies are usually due to global hypoperfusion in the setting of systemic hypotension from heart failure, septic shock, dehydration, or anemia. Focal strokes in the right parietal lobe and medial dorsal thalamus rarely can lead to a delirious state. A more common scenario involves a new focal stroke or hemorrhage causing confusion in a patient who has decreased cerebral reserve. In these individuals, it is sometimes difficult to distinguish between cognitive dysfunction resulting from

the new neurovascular insult itself and delirium due to the infectious, metabolic, and pharmacologic complications that can accompany hospitalization after stroke.

Because a fluctuating course often is seen in delirium, intermittent seizures may be overlooked when one is considering potential etiologies. Both nonconvulsive status epilepticus and recurrent focal or generalized seizures followed by postictal confusion can cause delirium; EEG remains essential for this diagnosis. Seizure activity spreading from an electrical focus in a mass or infarct can explain global cognitive dysfunction caused by relatively small lesions.

It is very common for patients to experience delirium at the end of life in palliative care settings. This condition, sometimes described as *terminal restlessness*, must be identified and treated aggressively as it is an important cause of patient and family discomfort at the end of life. It should be remembered that these patients also may be suffering from more common etiologies of delirium such as systemic infection.

LABORATORY AND DIAGNOSTIC EVALUATION

A cost-effective approach to the diagnostic evaluation of delirium allows the history and physical examination to guide tests. No established algorithm for workup will fit all delirious patients due to the staggering number of potential etiologies, but one stepwise approach is detailed in Table 25-2. If a clear precipitant is identified early, such as an offending medication, little further workup is required. If, however, no likely etiology is uncovered with initial evaluation, an aggressive search for an underlying cause should be initiated.

Table 25-2 Stepwise Evaluation of a Patient with Delirium

Initial evaluation
History with special attention to medications (including over-the-counter and herbals)
General physical examination and neurologic examination
Complete blood count Electrolyte panel including calcium, magnesium, phosphorus
Liver function tests, including albumin
Renal function tests
First-tier further evaluation guided by initial evaluation
Systemic infection screen
Urinalysis and culture
Chest radiograph
Blood cultures Electrocardiogram
Arterial blood gas
Serum and/or urine toxicology screen (perform earlier in young persons)
Brain imaging with MRI with diffusion and gadolinium (preferred) or CT
Suspected CNS infection: lumbar puncture after brain imaging
Suspected seizure-related etiology: electroencephalogram (EEG) (if high suspicion, should be performed immediately)
Second-tier further evaluation
Vitamin levels: B_{12}, folate, thiamine
Endocrinologic laboratories: thyroid-stimulating hormone (TSH) and free T4 ; cortisol
Serum ammonia
Sedimentation rate Autoimmune serologies: antinuclear antibodies (ANA), complement levels; p-ANCA, c-ANCA
Infectious serologies: rapid plasmin reagin (RPR); fungal and viral serologies if high suspicion; HIV antibody
Lumbar puncture (if not already performed)
Brain MRI with and without gadolinium (if not already performed)

Abbreviations: p-ANCA, perinuclear antineutrophil cytoplasmic antibody; c-ANCA, cytoplasmic antineutrophil cytoplasmic antibody.

Basic screening labs, including a complete blood count, electrolyte panel, and tests of liver and renal function, should be obtained in all patients with delirium. In elderly patients, screening for systemic infection, including chest radiography, urinalysis and culture, and possibly blood cultures, is important.

In younger individuals, serum and urine drug and toxicology screening may be appropriate early in the workup. Additional laboratory tests addressing other autoimmune, endocrinologic, metabolic, and infectious etiologies should be reserved for patients in whom the diagnosis remains unclear after initial testing.

Multiple studies have demonstrated that brain imaging in patients with delirium is often unhelpful. However, if the initial workup is unrevealing, most clinicians quickly move toward imaging of the brain to exclude structural causes. A noncontrast CT scan can identify large masses and hemorrhages but is otherwise relatively insensitive for discovering an etiology of delirium. The ability of MRI to identify most acute ischemic strokes as well as to provide neuroanatomic detail that gives clues to possible infectious, inflammatory, neurodegenerative, and neoplastic conditions makes it the test of choice. Since MRI techniques are limited by availability, speed of imaging, patient cooperation, and contraindications to magnetic exposure, many clinicians begin with CT scanning and proceed to MRI if the etiology of delirium remains elusive.

Lumbar puncture (LP) must be obtained immediately after appropriate neuroimaging in all patients in whom CNS infection is suspected. Spinal fluid examination can also be useful in identifying inflammatory and neoplastic conditions as well as in the diagnosis of hepatic encephalopathy through elevated cerebrospinal fluid (CSF) glutamine levels. As a result, LP should be considered in any delirious patient with a negative workup. EEG does not have a routine role in the workup of delirium, but it remains invaluable if seizure-related etiologies are considered.

Treatment: Delirium

Management of delirium begins with treatment of the underlying inciting factor (e.g., patients with systemic infections should be given appropriate antibiotics, and underlying electrolyte disturbances judiciously corrected). These treatments often lead to prompt resolution of delirium. Blindly targeting the symptoms of delirium pharmacologically only serves to prolong the time patients remain in the confused state and may mask important diagnostic information. Recent trials of medications used to boost cholinergic tone in delirious patients have led to mixed results, and this strategy is not currently recommended.

Relatively simple methods of supportive care can be highly effective in treating patients with delirium. Reorientation by the nursing staff and family combined with visible clocks, calendars, and outside-facing windows can reduce confusion. Sensory isolation should be prevented by providing glasses and hearing aids to patients who need them. Sundowning can be addressed to a large extent through vigilance to appropriate sleep-wake cycles. During the day, a well-lit room should be accompanied by activities or exercises to prevent napping. At night, a quiet, dark environment with limited interruptions by staff can assure proper rest. These sleep-wake cycle interventions are especially important in the ICU setting as the usual constant 24-h activity commonly provokes delirium. Attempting to mimic the home environment as much as possible also has been shown to help treat and even prevent delirium. Visits from friends and family throughout the day minimize the anxiety associated with the constant flow of new faces of staff and physicians. Allowing hospitalized patients to have access to home bedding, clothing, and nightstand objects makes the hospital environment less foreign and therefore less confusing. Simple standard nursing practices such as maintaining proper nutrition and volume status as well as managing incontinence and skin breakdown also help alleviate discomfort and resulting confusion.

In some instances, patients pose a threat to their own safety or to the safety of staff members, and acute management is required. Bed alarms and personal sitters are more effective and much less disorienting than physical restraints. Chemical restraints should be avoided, but when necessary, very low dose typical or atypical antipsychotic medications administered on an as-needed basis are effective. The recent association of antipsychotic use in the elderly with increased mortality rates underscores the importance of using these medications judiciously and only as a last resort. Benzodiazepines are not as effective as antipsychotics and often worsen confusion through their sedative properties. Although many clinicians still use benzodiazepines to treat acute confusion, their use should be limited to cases in which delirium is caused by alcohol or benzodiazepine withdrawal.

Prevention

In light of the high morbidity associated with delirium and the tremendously increased health care costs that accompany it, development of an effective strategy to prevent delirium in hospitalized patients is extremely important. Successful identification of high-risk patients is the first step, followed by initiation of appropriate interventions. One trial randomized more than 850 elderly inpatients to simple standardized protocols used to manage risk factors for delirium, including cognitive impairment, immobility, visual impairment, hearing impairment, sleep deprivation, and dehydration. Significant reductions in the number and duration of episodes of delirium were observed in the treatment group, but unfortunately, delirium recurrence rates were unchanged. Recent trials in the ICU have focused on identifying sedatives, such as dexmedetomidine, that are less likely to lead to delirium in critically ill patients. All hospitals and health care systems should work toward developing standardized protocols to address common risk factors with the goal of decreasing the incidence of delirium.

ACKNOWLEDGMENT

In the 16th edition, Allan H. Ropper contributed to a section on acute confusional states that was incorporated into this current chapter.

FURTHER READINGS

FONG TG et al: Delirium in elderly adults: Diagnosis, prevention, and treatment. Nat Rev Neurol 5:210, 2009 [PMID: 19347026]

GIRARD TD et al: Delirium as a predictor of long-term cognitive impairment in survivors of critical illness. Crit Care Med 38:1513,2010 [PMID: 20473145]

INOUYE SK et al: Clarifying confusion: The confusion assessment method. A new method for detection of delirium. Ann Intern Med 113:941, 1990 [PMID: 2240918]

———: A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med 340:669, 1999

LAT I et al: The impact of delirium on clinical outcomes in mechanically ventilated surgical and trauma patients. Crit Care Med 37:1898, 2009 [PMID: 19384221]

RIKER RR et al: Dexmedetomidine vs midazolam for sedation of critically ill patients: A randomized trial. JAMA 301:489, 2009 [PMID: 19188334]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APHASIA, MEMORY LOSS, AND OTHER FOCAL CEREBRAL DISORDERS: INTRODUCTION

The cerebral cortex of the human brain contains approximately 20 billion neurons spread over an area of 2.5 m^2. The *primary sensory* areas provide an obligatory portal for the entry of sensory information into cortical circuitry, and the *primary motor* areas provide a final common pathway for coordinating complex motor acts. The primary sensory and motor areas constitute 10% of the cerebral cortex. The rest is subsumed by modality-selective, heteromodal, paralimbic, and limbic areas collectively known as the *association cortex* (**Fig. 26-1**). The association cortex mediates the integrative processes that subserve cognition, emotion, and behavior. A systematic testing of these mental functions is necessary for the effective clinical assessment of the association cortex and its diseases.

Figure 26-1

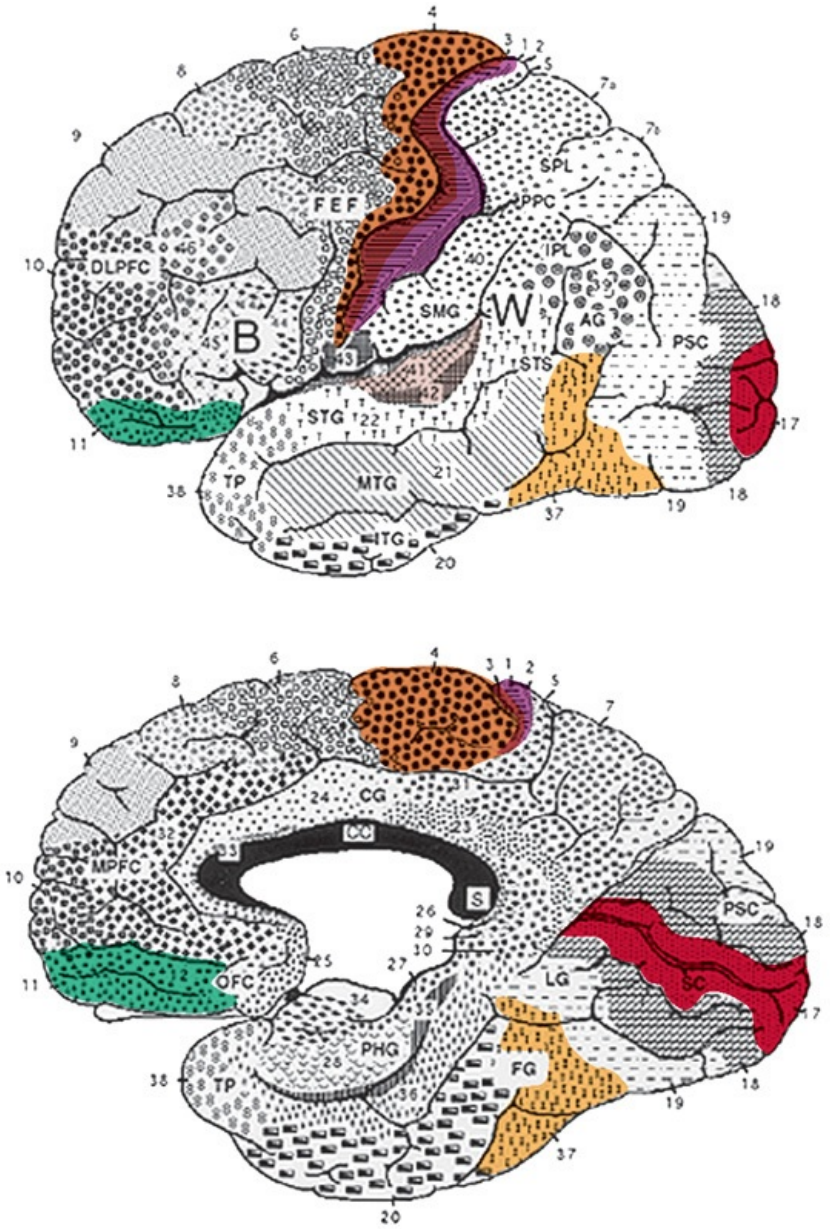

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lateral (*top*) and medial (*bottom*) views of the cerebral hemispheres. The numbers refer to the Brodmann cytoarchitectonic designations. Area 17 corresponds to the primary visual cortex, 41–42 to the primary auditory cortex, 1–3 to the primary somatosensory cortex, and 4 to the primary motor cortex. The rest of the cerebral cortex contains association areas. AG, angular gyrus; B, Broca's area; CC, corpus callosum; CG, cingulate gyrus; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye fields (premotor cortex); FG, fusiform gyrus; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; LG, lingual gyrus; MPFC, medial prefrontal cortex; MTG, middle temporal gyrus; OFC, orbitofrontal cortex; PHG, parahippocampal gyrus; PPC, posterior parietal cortex; PSC, peristriate cortex; SC, striate cortex; SMG, supramarginal gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus; STS, superior temporal sulcus; TP, temporopolar cortex; W, Wernicke's area.

According to current thinking, there are no centers for "hearing words," "perceiving space," or "storing memories." Cognitive and behavioral functions (domains) are coordinated by intersecting *large-scale neural networks* that contain interconnected cortical and subcortical components. The network approach to higher cerebral function has at least four implications of clinical relevance: (1) A single domain such as language or memory can be disrupted by damage to any one of several areas as long as those areas belong to the same network, (2) damage confined to a single area can give rise to multiple deficits involving the functions of all the networks that intersect in that region, (3) damage to a network component may give rise to minimal or transient deficits if other parts of the network undergo compensatory reorganization, and (4) individual anatomic sites within a network display a relative (but not absolute) specialization for different behavioral aspects of the relevant function. Five anatomically defined large-scale networks are most relevant to clinical practice: (1) a perisylvian network for language, (2) a parietofrontal network for spatial cognition, (3) an occipitotemporal network for face

and object recognition, (4) a limbic network for retentive memory, and (5) a prefrontal network for cognitive and behavioral control.

THE LEFT PERISYLVIAN NETWORK FOR LANGUAGE: APHASIAS AND RELATED CONDITIONS

Language allows the communication and elaboration of thoughts and experiences by linking them to arbitrary symbols known as words. The neural substrate of language is composed of a distributed network centered in the perisylvian region of the *left* hemisphere. The posterior pole of this network is located at the temporoparietal junction and includes a region known as *Wernicke's area*. An essential function of Wernicke's area is to transform sensory inputs into their neural word representations so that they can establish the distributed associations that give a word its meaning. The anterior pole of the language network is located in the inferior frontal gyrus and includes a region known as *Broca's area*. An essential function of this area is to transform neural word representations into their articulatory sequences so that the words can be uttered in the form of spoken language. The sequencing function of Broca's area also appears to involve the ordering of words into sentences that contain a meaning-appropriate *syntax* (grammar). Wernicke's and Broca's areas are interconnected with each other and with additional perisylvian, temporal, prefrontal, and posterior parietal regions, making up a neural network that subserves the various aspects of language function. Damage to any one of these components or to their interconnections can give rise to language disturbances (*aphasia*). Aphasia should be diagnosed only when there are deficits in the formal aspects of language, such as naming, word choice, comprehension, spelling, and syntax. Dysarthria and mutism do not by themselves lead to a diagnosis of aphasia. The language network shows a left hemisphere dominance pattern in the vast majority of the population. In approximately 90% of right-handers and 60% of left-handers, aphasia occurs only after lesions of the left hemisphere. In some individuals no hemispheric dominance for language can be discerned, and in some others (including a small minority of right handers) there is a right hemisphere dominance for language. A language disturbance that occurs after a right hemisphere lesion in a right hander is called *crossed aphasia*.

Clinical Examination

The clinical examination of language should include the assessment of naming, spontaneous speech, comprehension, repetition, reading, and writing. A deficit of naming (*anomia*) is the single most common finding in aphasic patients. When asked to name a common object (pencil or wristwatch), the patient may fail to come up with the appropriate word, may provide a circumlocutious description of the object ("the thing for writing"), or may come up with the wrong word (*paraphasia*). If the patient offers an incorrect but related word ("pen" for "pencil"), the naming error is known as a *semantic paraphasia*; if the word approximates the correct answer but is phonetically inaccurate ("plentil" for "pencil"), it is known as a *phonemic paraphasia*. Asking the patient to name body parts, geometric shapes, and component parts of objects (lapel of coat, cap of pen) can elicit mild forms of anomia in patients who otherwise can name common objects. In most anomias, the patient cannot retrieve the appropriate name when shown an object but can point to the appropriate object when the name is provided by the examiner. This is known as a one-way (or retrieval-based) naming deficit. A two-way naming deficit exists if the patient can neither provide nor recognize the correct name, indicating the likely presence of a comprehension impairment for the word. *Spontaneous speech* is described as "fluent" if it maintains appropriate output volume, phrase length, and melody or as "nonfluent" if it is sparse and halting and average utterance length is below four words. The examiner also should note if the speech is paraphasic or circumlocutious; if it shows a relative paucity of substantive nouns and action verbs versus function words (prepositions, conjunctions); and if word order, tenses, suffixes, prefixes, plurals, and possessives are appropriate. *Comprehension* can be tested by assessing the patient's ability to follow conversation, asking yes-no questions ("Can a dog fly?", "Does it snow in summer?") or asking the patient to point to appropriate objects ("Where is the source of illumination in this room?"). Statements with embedded clauses or a passive voice construction ("If a tiger is eaten by a lion, which animal stays alive?") help assess the ability to comprehend complex syntactic structure. Commands to close or open the eyes, stand up, sit down, or roll over should not be used to assess overall comprehension since appropriate responses aimed at such axial movements can be preserved in patients who otherwise have profound comprehension deficits.

Repetition is assessed by asking the patient to repeat single words, short sentences, or strings of words such as "No ifs, ands, or buts." The testing of repetition with tongue twisters such as "hippopotamus" and "Irish constabulary" provides a better assessment of dysarthria and pallilalia than of aphasia. Aphasic patients who have little difficulty with tongue twisters may have a particularly hard time repeating a string of function words. It is important to make sure that the number of words does not exceed the patient's attention span. Otherwise, the failure of repetition becomes a reflection of the narrowed attention span rather than an indication of an aphasic deficit. *Reading* should be assessed for deficits in reading aloud as well as comprehension. *Writing* is assessed for spelling errors, word order, and grammar. *Alexia* describes an inability to either read aloud or comprehend single words and simple sentences; *agraphia* (or dysgraphia) is used to describe an acquired deficit in the spelling or grammar of written language.

The correspondence between individual deficits of language function and lesion location does not display a rigid one-to-one relationship and should be conceptualized within the context of the distributed network model. Nonetheless, the classification of aphasias into specific clinical syndromes helps determine the most likely anatomic distribution of the underlying neurologic disease and has implications for etiology and prognosis (**Table 26-1**). The syndromes listed in Table 26-1 are most applicable to aphasias caused by cerebrovascular accidents (CVAs). They can be divided into "central" syndromes, which result from damage to the two epicenters of the language network (Broca's and Wernicke's areas), and "disconnection" syndromes, which arise from lesions that interrupt the functional connectivity of those centers with each other and with the other components of the language network. The syndromes outlined below are idealizations; pure syndromes occur rarely.

Table 26-1 Clinical Features of Aphasias and Related Conditions

	Comprehension	Repetition of Spoken Language	Naming	Fluency
Wernicke's	Impaired	Impaired	Impaired	Preserved or increased
Broca's	Preserved (except grammar)	Impaired	Impaired	Decreased
Global	Impaired	Impaired	Impaired	Decreased
Conduction	Preserved	Impaired	Impaired	Preserved
Nonfluent (motor) transcortical	Preserved	Preserved	Impaired	Impaired
Fluent (sensory) transcortical	Impaired	Preserved	Impaired	Preserved
Isolation	Impaired	Echolalia	Impaired	No purposeful speech
Anomic	Preserved	Preserved	Impaired	Preserved except for word-finding pauses
Pure word deafness	Impaired only for spoken language	Impaired	Preserved	Preserved
Pure alexia	Impaired only for reading	Preserved	Preserved	Preserved

WERNICKE'S APHASIA

Comprehension is impaired for spoken and written language, for single words as well as sentences. Language output is fluent but is highly paraphasic and circumlocutious. The tendency for paraphasic errors may be so pronounced that it leads to strings of neologisms, which form the basis of what is known as "jargon aphasia." Speech contains large numbers of function words (e.g., prepositions, conjunctions) but few substantive nouns or verbs that refer to specific actions. The output is therefore voluminous but uninformative. For example, a patient attempts to describe how his wife accidentally threw away something important, perhaps his dentures: "We don't need it anymore, she says. And with it when that was downstairs was my teeth-tick. . .a. . .den. . .dentith. . .my dentist. And they happened to be in that bag. . .see? How could this have happened? How could a thing like this happen. . .So she says we won't need it anymore. . .I didn't think we'd use it. And now if I have any problems anybody coming a month from now, 4 months from now, or 6 months from now, I have a new dentist. Where my two. . .two little pieces of dentist that I use. . .that I. . .all gone. If she throws the whole thing away. . .visit some friends of hers and she can't throw them away."

Gestures and pantomime do not improve communication. The patient does not seem to realize that his or her language is incomprehensible and may appear angry and impatient when the examiner fails to decipher the meaning of a severely paraphasic statement. In some patients this type of aphasia can be associated with severe agitation and paranoid behaviors. One area of comprehension that may be preserved is the ability to follow commands aimed at axial musculature. The dissociation between the failure to understand simple questions ("What is your name?") in a patient who rapidly closes his or her eyes, sits up, or rolls over when asked to do so is characteristic of Wernicke's aphasia and helps differentiate it from deafness, psychiatric disease, or malingering. Patients with Wernicke's aphasia cannot express their thoughts in meaning-appropriate words and cannot decode the meaning of words in any modality of input. This aphasia therefore has expressive as well as receptive components. Repetition, naming, reading, and writing also are impaired.

The lesion site most commonly associated with Wernicke's aphasia is the posterior portion of the language network and tends to involve at least parts of Wernicke's area. An embolus to the inferior division of the middle cerebral artery, to the posterior temporal or angular branches in particular, is the most common etiology (Chap. 370). Intracerebral hemorrhage, severe head trauma, and neoplasm are other causes. A coexisting right hemianopia or superior quadrantanopia is common and mild right nasolabial flattening may be found, but otherwise the examination is often unrevealing. The

paraphasic, neologistic speech in an agitated patient with an otherwise unremarkable neurologic examination may lead to the suspicion of a primary psychiatric disorder such as schizophrenia or mania, but the other components characteristic of acquired aphasia and the absence of prior psychiatric disease usually settle the issue. Some patients with Wernicke's aphasia due to intracerebral hemorrhage or head trauma may improve as the hemorrhage or the injury heals. In most other patients, prognosis for recovery of language function is guarded.

BROCA'S APHASIA

Speech is nonfluent, labored, interrupted by many word-finding pauses, and usually dysarthric. It is impoverished in function words but enriched in meaning-appropriate nouns and verbs. Abnormal word order and the inappropriate deployment of *bound morphemes* (word endings used to denote tenses, possessives, or plurals) lead to a characteristic agrammatism. Speech is telegraphic and pithy but quite informative. In the following passage, a patient with Broca's aphasia describes his medical history: "I see. . .the dotor, dotor sent me. . .Bosson. Go to hospital. Dotor. . .kept me beside. Two, tee days, doctor send me home."

Output may be reduced to a grunt or single word ("yes" or "no"), which is emitted with different intonations in an attempt to express approval or disapproval. In addition to fluency, naming and repetition are impaired. Comprehension of spoken language is intact except for syntactically difficult sentences with a passive voice structure or embedded clauses. Reading comprehension also is preserved with the occasional exception of a specific inability to read small grammatical words such as conjunctions and pronouns. The last two features indicate that Broca's aphasia is not just an "expressive" or "motor" disorder and that it also may involve a comprehension deficit for function words and syntax. Patients with Broca's aphasia can be tearful, easily frustrated, and profoundly depressed. Insight into their condition is preserved, in contrast to Wernicke's aphasia. Even when spontaneous speech is severely dysarthric, the patient may be able to display a relatively normal articulation of words when singing. This dissociation has been used to develop specific therapeutic approaches (melodic intonation therapy) for Broca's aphasia. Additional neurologic deficits usually include right facial weakness, hemiparesis or hemiplegia, and a buccofacial apraxia characterized by an inability to carry out motor commands involving oropharyngeal and facial musculature (e.g., patients are unable to demonstrate how to blow out a match or suck through a straw). Visual fields are intact. The cause is most often infarction of Broca's area (the inferior frontal convolution; "B" in Fig. 26-1) and surrounding anterior perisylvian and insular cortex due to occlusion of the superior division of the middle cerebral artery (Chap. 370). Mass lesions, including tumor, intracerebral hemorrhage, and abscess, also may be responsible. Small lesions confined to the posterior part of Broca's area may lead to a nonaphasic and often reversible deficit of speech articulation that usually is accompanied by mild right facial weakness. When the cause of Broca's aphasia is stroke, recovery of language function generally peaks within 2 to 6 months, after which time further progress is limited.

GLOBAL APHASIA

Speech output is nonfluent, and comprehension of spoken language is severely impaired. Naming, repetition, reading, and writing also are impaired. This syndrome represents the combined dysfunction of Broca's and Wernicke's areas and usually results from strokes that involve the entire middle cerebral artery distribution in the left hemisphere. Most patients are initially mute or say a few words, such as "hi" or "yes." Related signs include right hemiplegia, hemisensory loss, and homonymous hemianopia. Occasionally, a patient with a lesion in Wernicke's area will present with a global aphasia that soon resolves into Wernicke's aphasia.

CONDUCTION APHASIA

Speech output is fluent but paraphasic, comprehension of spoken language is intact, and repetition is severely impaired. Naming and writing also are impaired. Reading aloud is impaired, but reading comprehension is preserved. The lesion sites spare Broca's and Wernicke's areas but may induce a functional disconnection between the two so that neural word representations formed in Wernicke's area and adjacent regions cannot be conveyed to Broca's area for assembly into corresponding articulatory patterns. Occasionally, a Wernicke's area lesion gives rise to a transient Wernicke's aphasia that rapidly resolves into a conduction aphasia. The paraphasic output in conduction aphasia interferes with the ability to express meaning, but this deficit is not nearly as severe as the one displayed by patients with Wernicke's aphasia. Associated neurologic signs in conduction aphasia vary according to the primary lesion site.

NONFLUENT TRANSCORTICAL APHASIA (TRANSCORTICAL MOTOR APHASIA)

The features are similar to those of Broca's aphasia, but repetition is intact and agrammatism may be less pronounced. The neurologic examination may be otherwise intact, but a right hemiparesis also can exist. The lesion site disconnects the intact language network from prefrontal areas of the brain and usually involves the anterior watershed zone between anterior and middle cerebral artery territories or the supplementary motor cortex in the territory of the anterior cerebral artery.

FLUENT TRANSCORTICAL APHASIA (TRANSCORTICAL SENSORY APHASIA)

Clinical features are similar to those of Wernicke's aphasia, but repetition is intact. The lesion site

disconnects the intact core of the language network from other temporoparietal association areas. Associated neurologic findings may include hemianopia. Cerebrovascular lesions (e.g., infarctions in the posterior watershed zone) and neoplasms that involve the temporoparietal cortex posterior to Wernicke's area are the most common causes.

ISOLATION APHASIA

This rare syndrome represents a combination of the two transcortical aphasias. Comprehension is severely impaired, and there is no purposeful speech output. The patient may parrot fragments of heard conversations (*echolalia*), indicating that the neural mechanisms for repetition are at least partially intact. This condition represents the pathologic function of the language network when it is isolated from other regions of the brain. Broca's and Wernicke's areas tend to be spared, but there is damage to the surrounding frontal, parietal, and temporal cortex. Lesions are patchy and can be associated with anoxia, carbon monoxide poisoning, or complete watershed zone infarctions.

ANOMIC APHASIA

This form of aphasia may be considered the "minimal dysfunction" syndrome of the language network. Articulation, comprehension, and repetition are intact, but confrontation naming, word finding, and spelling are impaired. Speech is enriched in function words but impoverished in substantive nouns and verbs denoting specific actions. Language output is fluent but paraphasic, circumlocutious, and uninformative. Fluency may be interrupted by word-finding hesitations. The lesion sites can be anywhere within the left hemisphere language network, including the middle and inferior temporal gyri. *Anomic aphasia is the single most common language disturbance seen in head trauma, metabolic encephalopathy, and Alzheimer's disease*.

PURE WORD DEAFNESS

The most common causes are either bilateral or left-sided middle cerebral artery (MCA) strokes affecting the superior temporal gyrus. The net effect of the underlying lesion is to interrupt the flow of information from the auditory association cortex to Wernicke's area. Patients have no difficulty understanding written language and can express themselves well in spoken or written language. They have no difficulty interpreting and reacting to environmental sounds since primary auditory cortex and subcortical auditory relays are intact. Since auditory information cannot be conveyed to the language network, however, it cannot be decoded into neural word representations, and the patient reacts to speech as if it were in an alien tongue that cannot be deciphered. Patients cannot repeat spoken language but have no difficulty naming objects. In time, patients with pure word deafness teach themselves lipreading and may appear to have improved. There may be no additional neurologic findings, but agitated paranoid reactions are common in the acute stages. Cerebrovascular lesions are the most common cause.

PURE ALEXIA WITHOUT AGRAPHIA

This is the visual equivalent of pure word deafness. The lesions (usually a combination of damage to the left occipital cortex and to a posterior sector of the corpus callosum—the splenium) interrupt the flow of visual input into the language network. There is usually a right hemianopia, but the core language network remains unaffected. The patient can understand and produce spoken language, name objects in the left visual hemifield, repeat, and write. However, the patient acts as if illiterate when asked to read even the simplest sentence because the visual information from the written words (presented to the intact left visual hemifield) cannot reach the language network. Objects in the left hemifield may be named accurately because they activate nonvisual associations in the right hemisphere, which in turn can access the language network through transcallosal pathways anterior to the splenium. Patients with this syndrome also may lose the ability to name colors, although they can match colors. This is known as a *color anomia*. The most common etiology of pure alexia is a vascular lesion in the territory of the posterior cerebral artery or an infiltrating neoplasm in the left occipital cortex that involves the optic radiations as well as the crossing fibers of the splenium. Since the posterior cerebral artery also supplies medial temporal components of the limbic system, a patient with pure alexia also may experience an amnesia, but this is usually transient because the limbic lesion is unilateral.

APHEMIA

There is an acute onset of severely impaired fluency (often mutism), which cannot be accounted for by corticobulbar, cerebellar, or extrapyramidal dysfunction. Recovery is the rule and involves an intermediate stage of hoarse whispering. Writing, reading, and comprehension are intact, and so this is not a true aphasic syndrome. Partial lesions of Broca's area or subcortical lesions that undercut its connections with other parts of the brain may be present. Occasionally, the lesion site is on the medial aspects of the frontal lobes and may involve the supplementary motor cortex of the left hemisphere.

APRAXIA

This generic term designates a complex motor deficit that cannot be attributed to pyramidal, extrapyramidal, cerebellar, or sensory dysfunction and that does not arise from the patient's failure to understand the nature of the task. The form that is encountered most frequently in clinical practice is

known as *ideomotor apraxia*. Commands to perform a specific motor act ("cough," "blow out a match") or pantomime the use of a common tool (a comb, hammer, straw, or toothbrush) in the absence of the real object cannot be followed. The patient's ability to comprehend the command is ascertained by demonstrating multiple movements and establishing that the correct one can be recognized. Some patients with this type of apraxia can imitate the appropriate movement (when it is demonstrated by the examiner) and show no impairment when handed the real object, indicating that the sensorimotor mechanisms necessary for the movement are intact. Some forms of ideomotor apraxia represent a disconnection of the language network from pyramidal motor systems: commands to execute complex movements are understood but cannot be conveyed to the appropriate motor areas even though the relevant motor mechanisms are intact. *Buccofacial apraxia* involves apraxic deficits in movements of the face and mouth. *Limb apraxia* encompasses apraxic deficits in movements of the arms and legs. Ideomotor apraxia almost always is caused by lesions in the left hemisphere and is commonly associated with aphasic syndromes, especially Broca's aphasia and conduction aphasia. Its presence cannot be ascertained in patients with language comprehension deficits. The ability to follow commands aimed at axial musculature ("close the eyes," "stand up") is subserved by different pathways and may be intact in otherwise severely aphasic and apraxic patients. Since the handling of real objects is not impaired, ideomotor apraxia by itself causes no major limitation of daily living activities. Patients with lesions of the anterior corpus callosum can display ideomotor apraxia confined to the left side of the body, a sign known as *sympathetic dyspraxia*. A severe form of sympathetic dyspraxia known as the *alien hand* syndrome is characterized by additional features of motor disinhibition on the left hand.

Ideational apraxia refers to a deficit in the execution of a goal-directed sequence of movements in patients who have no difficulty executing the individual components of the sequence. For example, when the patient is asked to pick up a pen and write, the sequence of uncapping the pen, placing the cap at the opposite end, turning the point toward the writing surface, and writing may be disrupted, and the patient may be seen trying to write with the wrong end of the pen or even with the removed cap. These motor sequencing problems usually are seen in the context of confusional states and dementias rather than focal lesions associated with aphasic conditions. *Limb-kinetic apraxia* involves a clumsiness in the actual use of tools that cannot be attributed to sensory, pyramidal, extrapyramidal, or cerebellar dysfunction. This condition can emerge in the context of focal premotor cortex lesions or *corticobasal degeneration*.

GERSTMANN'S SYNDROME

The combination of *acalculia* (impairment of simple arithmetic), *dysgraphia* (impaired writing), *finger anomia* (an inability to name individual fingers such as the index and thumb), and *right-left confusion* (an inability to tell whether a hand, foot, or arm of the patient or examiner is on the right or left side of the body) is known as Gerstmann's syndrome. In making this diagnosis it is important to establish that the finger and left-right naming deficits are not part of a more generalized anomia and that the patient is not otherwise aphasic. When Gerstmann's syndrome is seen in isolation, it is commonly associated with damage to the inferior parietal lobule (especially the angular gyrus) in the left hemisphere.

APROSODIA

Variations of melodic stress and intonation influence the meaning and impact of spoken language. For example, the two statements "He *is* clever." and "He is *clever*?" contain an identical word choice and syntax but convey vastly different messages because of differences in the intonation and stress with which the statements are uttered. This aspect of language is known as *prosody*. Damage to perisylvian areas in the right hemisphere can interfere with speech prosody and can lead to syndromes of aprosodia. Damage to right hemisphere regions corresponding to Wernicke's area can selectively impair decoding of speech prosody, whereas damage to right hemisphere regions corresponding to Broca's area yields a greater impairment in the ability to introduce meaning-appropriate prosody into spoken language. The latter deficit is the most common type of aprosodia identified in clinical practice; the patient produces grammatically correct language with accurate word choice, but the statements are uttered in a monotone that interferes with the ability to convey the intended stress and affect. Patients with this type of aprosodia give the mistaken impression of being depressed or indifferent.

SUBCORTICAL APHASIA

Damage to subcortical components of the language network (e.g., the striatum and thalamus of the left hemisphere) also can lead to aphasia. The resulting syndromes contain combinations of deficits in the various aspects of language but rarely fit the specific patterns described in Table 26-1. In a patient with a CVA, an anomic aphasia accompanied by dysarthria or a fluent aphasia with hemiparesis should raise the suspicion of a subcortical lesion site.

PROGRESSIVE APHASIAS

Aphasias caused by cerebrovascular accidents start suddenly and display maximal deficits at the onset. The underlying lesion is relatively circumscribed and is associated with a total loss of neural function in at least part of the lesion site. These are the "classic" aphasias described above. Aphasias caused by neurodegenerative diseases have an insidious onset and a relentless progression so that

the symptomatology changes over time. Since the neuronal loss within the areas encompassed by the neurodegeneration is partial and since it tends to include multiple components of the language network, the clinico-anatomic patterns are different from those described in Table 26-1.

Clinical Presentation and Diagnosis of Primary Progressive Aphasia (PPA)

When a neurodegenerative disease selectively undermines language function, a clinical diagnosis of PPA is made. A patient with PPA comes to medical attention because of word-finding difficulties, abnormal speech patterns, word-comprehension impairments, or spelling errors of recent onset. PPA is diagnosed when other mental faculties, such as memory for daily events, visuospatial skills (assessed by tests of drawing and face recognition), and comportment (assessed by history obtained from a third party), remain relatively intact; when language is the major area of dysfunction for the first few years of the disease; and when structural brain imaging does not reveal a specific lesion, other than atrophy, that accounts for the language deficit. Impairments in other cognitive functions may emerge eventually, but the language dysfunction remains the most salient feature and deteriorates most rapidly throughout the illness.

Language in PPA

The language impairment in PPA varies from patient to patient. Some patients cannot find the right words to express thoughts; others cannot understand the meaning of heard or seen words; still others cannot name objects in the environment. The language impairment can be fluent (that is, with normal articulation, flow, and number of words per utterance) or nonfluent. The single most common sign of primary progressive aphasia is an inability to come up with the right word during conversation and/or an inability to name objects shown by the examiner (anomia). Distinct forms of agrammatism and/or word comprehension deficits also can arise. The agrammatism consists of inappropriate word order and misuse of small grammatical words. Comprehension deficits, if present, start with an occasional inability to understand single low-frequency words and gradually progress to encompass the comprehension of conversational speech.

The impairments of syntax, comprehension, naming, or writing in PPA form slightly different patterns from those seen in CVA-caused aphasias. Three subtypes of PPA can be recognized: an agrammatic variant characterized by poor fluency and impaired grammar, a semantic variant characterized by preserved fluency and syntax but poor single word comprehension, and a logopenic variant characterized by preserved syntax and comprehension but frequent word-finding pauses during spontaneous speech. The agrammatic variant also is known as *progressive nonfluent aphasia* and displays similarities to Broca's aphasia. However, dysarthria is usually absent. The semantic variant of PPA displays similarities to Wernicke's aphasia, but the comprehension difficulty tends to be most profound for single words denoting concrete objects.

Pathophysiology

The three variants of PPA display overlapping distributions of neuronal loss, but the agrammatic variant is most closely associated with atrophy in the anterior parts of the language network (where Broca's area is located), the semantic variant with atrophy in the anterior temporal components of the language network, and the logopenic variant with atrophy in the temporoparietal component of the language network. The abnormalities may remain confined to the left hemisphere perisylvian and anterior temporal cortices initially, but gradual deterioration in PPA leads to a loss of syndromic specificity as the disease progresses.

Neuropathology

In the majority of PPA cases, the neuropathology falls within the family of frontotemporal lobar degenerations (FTLDs) and displays various combinations of focal neuronal loss, gliosis, tau-positive inclusions including Pick bodies, and tau-negative TDP-43 inclusions. Familial forms of PPA with TDP-43 inclusions recently were linked to mutations of the progranulin gene on chromosome 17. The agrammatic variant most frequently is associated with tauopathy, whereas the semantic variant is most closely associated with TDP-43 inclusions. Alzheimer's pathology is seen most frequently in the logopenic variant. The clinical subtyping of PPA thus may help predict the nature of the underlying neuropathology. The intriguing possibility has been raised that a personal or family history of dyslexia may be a risk factor for primary progressive aphasia, at least in some patients, suggesting that this disease may arise on a background of genetic or developmental vulnerability that affects language-related areas of the brain.

THE PARIETOFRONTAL NETWORK FOR SPATIAL ORIENTATION: NEGLECT AND RELATED CONDITIONS

Hemispatial Neglect

Adaptive orientation to significant events within the extrapersonal space is subserved by a large-scale network containing three major cortical components. The *cingulate cortex* provides access to a motivational mapping of the extrapersonal space, the *posterior parietal cortex* to a sensorimotor representation of salient extrapersonal events, and the *frontal eye fields* to motor strategies for

attentional behaviors (**Fig. 26-2**). Subcortical components of this network include the striatum and the thalamus. Contralesional hemispatial neglect represents one outcome of damage to any of the cortical or subcortical components of this network. *The traditional view that hemispatial neglect always denotes a parietal lobe lesion is inaccurate*. In keeping with this anatomic organization, the clinical manifestations of neglect display three behavioral components: sensory events (or their mental representations) within the neglected hemispace have a lesser impact on overall awareness, there is a paucity of exploratory and orienting acts directed toward the neglected hemispace, and the patient behaves as if the neglected hemispace were motivationally devalued.

Figure 26-2

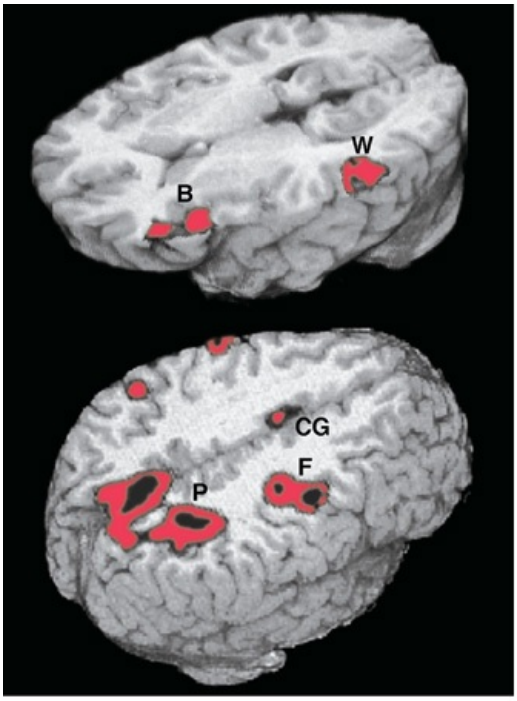

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Functional magnetic resonance imaging of language and spatial attention in neurologically intact subjects. The red and black areas show regions of task-related significant activation. (*Top*) The subjects were asked to determine if two words were synonymous. This language task led to the simultaneous activation of the two epicenters of the language network, Broca's area (B) and Wernicke's area (W). The activations are exclusively in the left hemisphere. (*Bottom*) The subjects were asked to shift spatial attention to a peripheral target. This task led to the simultaneous activation of the three epicenters of the attentional network: the posterior parietal cortex (P), the frontal eye fields (F), and the cingulate gyrus (CG). The activations are predominantly in the right hemisphere. (*Courtesy of Darren Gitelman, MD;* with permission.)

According to one model of spatial cognition, the right hemisphere directs attention within the *entire* extrapersonal space, whereas the left hemisphere directs attention mostly within the contralateral right hemispace. Consequently, unilateral left hemisphere lesions do not give rise to much contralesional neglect since the global attentional mechanisms of the right hemisphere can compensate for the loss of the *contralaterally* directed attentional functions of the left hemisphere. Unilateral right hemisphere lesions, however, give rise to severe contralesional left hemispatial neglect because the unaffected left hemisphere does not contain ipsilateral attentional mechanisms. This model is consistent with clinical experience, which shows that contralesional neglect is more common, severe, and lasting after damage to the right hemisphere than after damage to the left hemisphere. Severe neglect for the right hemispace is rare, even in left-handers with left hemisphere lesions.

CLINICAL EXAMINATION

Patients with severe neglect may fail to dress, shave, or groom the left side of the body; fail to eat food placed on the left side of the tray; and fail to read the left half of sentences. When the examiner draws a large circle [12 to 15 cm (5 to 6 in.) in diameter] and asks the patient to place the numbers 1 to 12 as if the circle represented the face of a clock, there is a tendency to crowd the numbers on the right side and leave the left side empty. When asked to copy a simple line drawing, the patient fails to copy detail on the left, and when the patient is asked to write, there is a tendency to leave an unusually wide margin on the left.

Two bedside tests that are useful in assessing neglect are *simultaneous bilateral stimulation* and

visual target cancellation. In the former, the examiner provides either unilateral or simultaneous bilateral stimulation in the visual, auditory, and tactile modalities. After right hemisphere injury, patients who have no difficulty detecting unilateral stimuli on either side experience the bilaterally presented stimulus as coming only from the right. This phenomenon is known as *extinction* and is a manifestation of the sensory-representational aspect of hemispatial neglect. In the target detection task, targets (e.g., A's) are interspersed with foils (e.g., other letters of the alphabet) on a 21.5- to 28.0-cm (8.5 to 11 in.) sheet of paper, and the patient is asked to circle all the targets. A failure to detect targets on the left is a manifestation of the exploratory deficit in hemispatial neglect (**Fig. 26-3*A***). Hemianopia is not by itself sufficient to cause the target detection failure since the patient is free to turn the head and eyes to the left. Target detection failures therefore reflect a distortion of spatial attention, not just of sensory input. The normal tendency in target detection tasks is to start from the left upper quadrant and move systematically in horizontal or vertical sweeps. Some patients show a tendency to start the process from the right and proceed in a haphazard fashion. This represents a subtle manifestation of left neglect even if the patient eventually manages to detect all the appropriate targets. Some patients with neglect also may deny the existence of hemiparesis and may even deny ownership of the paralyzed limb, a condition known as *anosognosia*.

Figure 26-3

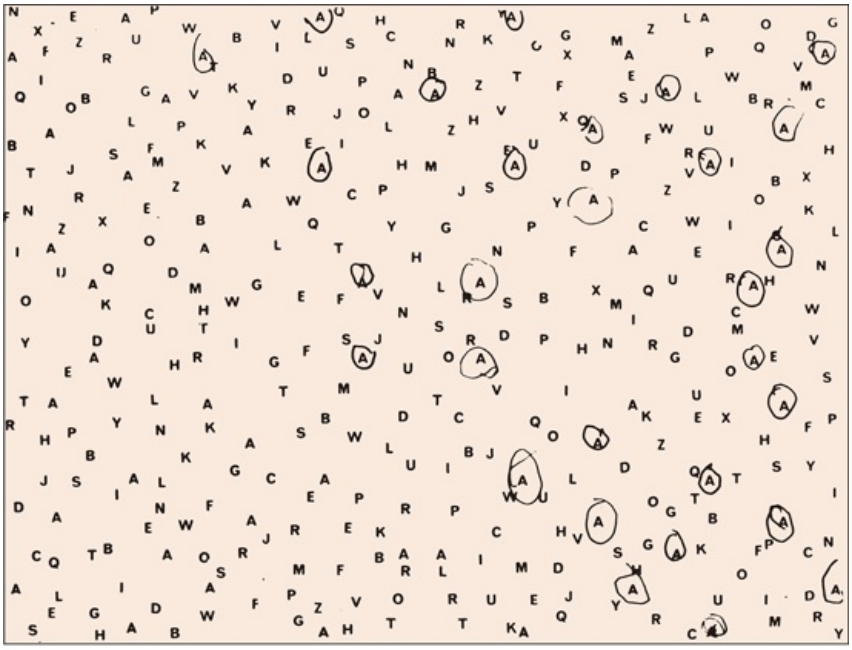

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

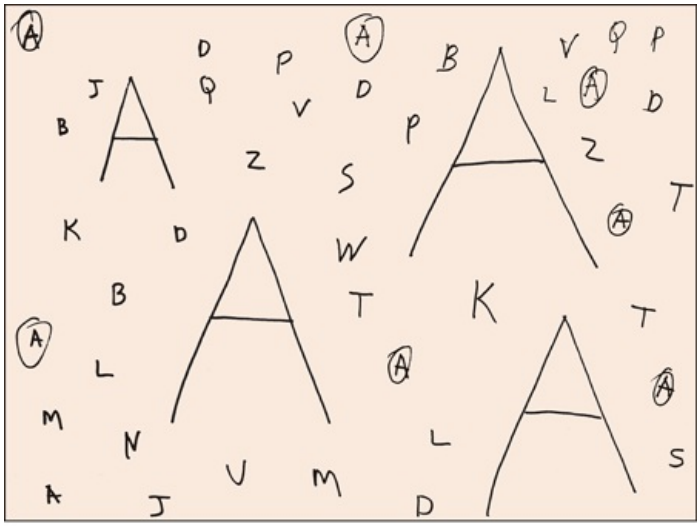

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. A 47-year-old man with a large frontoparietal lesion in the right hemisphere was asked to circle all the A's. Only targets on the right are circled. This is a manifestation of left hemispatial neglect. ***B.*** A 70-year-old woman with a 2-year history of degenerative dementia was able to circle most of the small targets but ignored the larger ones. This is a manifestation of simultanagnosia.

Bálint's Syndrome, Simultanagnosia, Dressing Apraxia, and Construction Apraxia

Bilateral involvement of the network for spatial attention, especially its parietal components, leads to a state of severe spatial disorientation known as *Bálint's syndrome*. Bálint's syndrome involves deficits in the orderly visuomotor scanning of the environment (*oculomotor apraxia*) and in accurate manual reaching toward visual targets (*optic ataxia*). The third and most dramatic component of Bálint's syndrome is known as *simultanagnosia* and reflects an inability to integrate visual information in the center of gaze with more peripheral information. The patient gets stuck on the detail that falls in the center of gaze without attempting to scan the visual environment for additional information. A patient with simultanagnosia "misses the forest for the trees." Complex visual scenes cannot be grasped in their entirety, leading to severe limitations in the visual identification of objects and scenes. For example, a patient who is shown a table lamp and asked to name the object may look at its circular base and call it an ashtray. Some patients with simultanagnosia report that objects they look at may vanish suddenly, probably indicating an inability to look back at the original point of gaze after brief saccadic displacements. Movement and distracting stimuli greatly exacerbate the difficulties of visual perception. Simultanagnosia sometimes can occur without the other two components of Bálint's syndrome.

A modification of the letter cancellation task described above can be used for the bedside diagnosis of simultanagnosia. In this modification, some of the targets (e.g., A's) are made to be much larger than the others [7.5 to 10 cm vs. 2.5 cm (3 to 4 in. vs. 1 in.) in height], and all targets are embedded among foils. Patients with simultanagnosia display a counterintuitive but characteristic tendency to miss the larger targets (Fig. 26-3*B*). This occurs because the information needed for the identification of the larger targets cannot be confined to the immediate line of gaze and requires the integration of visual information across a more extensive field of view. The greater difficulty in the detection of the larger targets also indicates that poor acuity is not responsible for the impairment of visual function and that the problem is central rather than peripheral.

Another manifestation of bilateral (or right-sided) dorsal parietal lobe lesions is *dressing apraxia*. A patient with this condition is unable to align the body axis with the axis of the garment and can be seen struggling as he or she holds a coat from its bottom or extends his or her arm into a fold of the garment rather than into its sleeve. Lesions that involve the posterior parietal cortex also lead to severe difficulties in copying simple line drawings. This is known as a *construction apraxia* and is much more severe if the lesion is in the right hemisphere. In some patients with right hemisphere lesions, the drawing difficulties are confined to the left side of the figure and represent a manifestation of hemispatial neglect; in others, there is a more universal deficit in reproducing contours and three-dimensional perspective. Dressing apraxia and construction apraxia represent special instances of a more general disturbance in spatial orientation.

CAUSES OF SPATIAL DISORIENTATION

Cerebrovascular lesions and neoplasms in the right hemisphere are the most common causes of hemispatial neglect. Depending on the site of the lesion, a patient with neglect also may have hemiparesis, hemihypesthesia, and hemianopia on the left, but these are not invariant findings. The majority of these patients display considerable improvement of hemispatial neglect, usually within the first several weeks. Bálint's syndrome results from bilateral dorsal parietal lesions; common settings include watershed infarction between the middle and posterior cerebral artery territories, hypoglycemia, and sagittal sinus thrombosis.

A progressive form of spatial disorientation known as the *posterior cortical atrophy* syndrome most commonly represents a variant of Alzheimer's disease with unusual concentrations of neurofibrillary degeneration in the parieto-occipital cortex and the superior colliculus. The patient displays a progressive Bálint's syndrome, usually accompanied by dressing and construction apraxia. Corticobasal degeneration, a type of FTLD with abnormal tau inclusions, can have an asymmetric distribution. When the atrophy shows a predilection for the right cerebral hemisphere, a progressive left hemineglect syndrome emerges on a background of left-sided extrapyramidal dysfunction.

THE OCCIPITOTEMPORAL NETWORK FOR FACE AND OBJECT RECOGNITION: PROSOPAGNOSIA AND OBJECT AGNOSIA

Perceptual information about faces and objects initially is encoded in primary (striate) visual cortex and adjacent (upstream) peristriate visual association areas. This information subsequently is relayed first to the downstream visual association areas of occipitotemporal cortex and then to other heteromodal and paralimbic areas of the cerebral cortex. Bilateral lesions in the fusiform and lingual gyri of the occipitotemporal cortex disrupt this process and interfere with the ability of otherwise intact perceptual information to activate the distributed multimodal associations that lead to the recognition of faces and

objects. The resultant face and object recognition deficits are known as *associative prosopagnosia* and *visual object agnosia*.

A patient with prosopagnosia cannot recognize familiar faces, including, sometimes, the reflection of his or her own face in the mirror. This is not a perceptual deficit since prosopagnosic patients easily can tell whether two faces are identical. Furthermore, a prosopagnosic patient who cannot recognize a familiar face by visual inspection alone can use auditory cues to reach appropriate recognition if allowed to listen to the person's voice. The deficit in prosopagnosia is therefore modality-specific and reflects the existence of a lesion that prevents the activation of otherwise intact multimodal templates by relevant visual input. The deficit in prosopagnosia is not limited to the recognition of faces but also can extend to the recognition of individual members of larger generic object groups. For example, prosopagnosic patients characteristically have no difficulty with the generic identification of a face as a face or a car as a car, but they cannot recognize the identity of an individual face or the make of an individual car. This reflects a visual recognition deficit for proprietary features that characterize individual members of an object class. When recognition problems become more generalized and extend to the generic identification of common objects, the condition is known as *visual object agnosia*. In contrast to prosopagnosic patients, those with object agnosia cannot recognize a face as a face or a car as a car. It is important to distinguish visual object agnosia from anomia. A patient with anomia cannot name the object but can describe its use. In contrast, a patient with visual agnosia is unable either to name a visually presented object or to describe its use. Face and object recognition disorders also can result from the simultanagnosia of Bálint's syndrome, in which case they are known as *apperceptive* agnosias as opposed to the *associative* agnosias that result from inferior temporal lobe lesions.

Causes

The characteristic lesions in prosopagnosia and visual object agnosia consist of bilateral infarctions in the territory of the posterior cerebral arteries. Associated deficits can include visual field defects (especially superior quadrantanopias) and a centrally based color blindness known as achromatopsia. Rarely, the responsible lesion is unilateral. In such cases, prosopagnosia is associated with lesions in the right hemisphere, and object agnosia with lesions in the left. Degenerative diseases of anterior and inferior temporal cortex can cause progressive associative prosopagnosia and object agnosia. The combination of progressive associative agnosia and a fluent aphasia is known as *semantic dementia* and usually is caused by FTLD with TDP-43 inclusions. Patients with semantic dementia fail to recognize faces and objects and cannot understand the meaning of words denoting objects.

THE LIMBIC NETWORK FOR EPISODIC MEMORY: AMNESIAS

Limbic and paralimbic areas (such as the hippocampus, amygdala, and entorhinal cortex), the anterior and medial nuclei of the thalamus, the medial and basal parts of the striatum, and the hypothalamus collectively constitute a distributed network known as the *limbic system*. The behavioral affiliations of this network include the coordination of emotion, motivation, autonomic tone, and endocrine function. An additional area of specialization for the limbic network and the one that is of most relevance to clinical practice is that of declarative (conscious) memory for recent episodes and experiences. A disturbance in this function is known as an *amnestic state*. In the absence of deficits in motivation, attention, language, or visuospatial function, the clinical diagnosis of a persistent global amnestic state is always associated with bilateral damage to the limbic network, usually within the hippocampo-entorhinal complex or the thalamus.

Although the limbic network is the site of damage for amnestic states, it is almost certainly not the storage site for memories. Memories are stored in widely distributed form throughout the cerebral cortex. The role attributed to the limbic network is to bind these distributed fragments into coherent events and experiences that can sustain conscious recall. Damage to the limbic network does not necessarily destroy memories but interferes with their conscious (declarative) recall in coherent form. The individual fragments of information remain preserved despite the limbic lesions and can sustain what is known as *implicit memory*. For example, patients with amnestic states can acquire new motor or perceptual skills even though they may have no conscious knowledge of the experiences that led to the acquisition of these skills.

The memory disturbance in the amnestic state is multimodal and includes retrograde and anterograde components. The *retrograde amnesia* involves an inability to recall experiences that occurred before the onset of the amnestic state. Relatively recent events are more vulnerable to retrograde amnesia than are more remote and more extensively consolidated events. A patient who comes to the emergency room complaining that he cannot remember his or her identity but can remember the events of the previous day almost certainly does not have a neurologic cause of memory disturbance. The second and most important component of the amnestic state is the *anterograde amnesia*, which indicates an inability to store, retain, and recall new knowledge. Patients with amnestic states cannot remember what they ate a few minutes ago or the details of an important event they may have experienced a few hours ago. In the acute stages, there also may be a tendency to fill in memory gaps with inaccurate, fabricated, and often implausible information. This is known as *confabulation*. Patients

with the amnestic syndrome forget that they forget and tend to deny the existence of a memory problem when questioned.

Clinical Examination

A patient with an amnestic state is almost always disoriented, especially to time. Accurate temporal orientation and accurate knowledge of current news rule out a major amnestic state. The anterograde component of an amnestic state can be tested with a list of four to five words read aloud by the examiner up to five times or until the patient can immediately repeat the entire list without an intervening delay. In the next phase of testing, the patient is allowed to concentrate on the words and rehearse them internally for 1 min before being asked to recall them. Accurate performance in this phase indicates that the patient is motivated and sufficiently attentive to hold the words online for at least 1 min. The final phase of the testing involves a retention period of 5 to 10 min during which the patient is engaged in other tasks. Adequate recall at the end of this interval requires offline storage, retention, and retrieval. Amnestic patients fail this phase of the task and may even forget that they were given a list of words to remember. Accurate recognition of the words by multiple choice in a patient who cannot recall them indicates a less severe memory disturbance that affects mostly the retrieval stage of memory. The retrograde component of an amnesia can be assessed with questions related to autobiographical or historic events. The anterograde component of amnestic states is usually much more prominent than the retrograde component. In rare instances, usually associated with temporal lobe epilepsy or benzodiazepine intake, the retrograde component may dominate.

The assessment of memory can be quite challenging. Bedside evaluations may detect only the most severe impairments. Less severe memory impairments, as in the case of patients with temporal lobe epilepsy, mild head injury, or early dementia, require quantitative evaluations by neuropsychologists. Confusional states caused by toxic-metabolic encephalopathies and some types of frontal lobe damage interfere with attentional capacity and lead to secondary memory impairments, even in the absence of any limbic lesions. This sort of memory impairment can be differentiated from the amnestic state by the presence of additional impairments in the attention-related tasks described below in the section on the frontal lobes.

Causes, Including Alzheimer's Disease

Many neurologic diseases can give rise to an amnestic state. They include tumors (of the sphenoid wing, posterior corpus callosum, thalamus, or medial temporal lobe), infarctions (in the territories of the anterior or posterior cerebral arteries), head trauma, herpes simplex encephalitis, Wernicke-Korsakoff encephalopathy, paraneoplastic limbic encephalitis, and degenerative dementias such as Alzheimer's disease and Pick's disease. The one common denominator of all these diseases is the presence of bilateral lesions within one or more components in the limbic network. Occasionally, unilateral left-sided hippocampal lesions can give rise to an amnestic state, but the memory disorder tends to be transient. Depending on the nature and distribution of the underlying neurologic disease, the patient also may have visual field deficits, eye movement limitations, or cerebellar findings.

Alzheimer's disease (AD) and its prodromal state of mild cognitive impairment (MCI) are the most common causes of progressive memory impairments. Temporal disorientation and poor recall of recent conversations are early manifestations. The predilection of the entorhinal cortex and hippocampus for early neurofibrillary degeneration in the MCI-AD spectrum is responsible for the initially selective impairment of episodic memory. In time, a full amnestic state emerges, but usually with additional impairments in language, attention, and visuospatial skills as the neurofibrillary degeneration spreads to additional neocortical areas.

Transient global amnesia is a distinctive syndrome usually seen in late middle age. Patients become acutely disoriented and repeatedly ask who they are, where they are, and what they are doing. The spell is characterized by anterograde amnesia (inability to retain new information) and a retrograde amnesia for relatively recent events that occurred before the onset. The syndrome usually resolves within 24 to 48 h and is followed by the filling in of the period affected by the retrograde amnesia, although there is persistent loss of memory for the events that occurred during the ictus. Recurrences are noted in approximately 20% of patients. Migraine, temporal lobe seizures, and perfusion abnormalities in the posterior cerebral territory have been postulated as causes of transient global amnesia. The absence of associated neurologic findings occasionally may lead to the incorrect diagnosis of a psychiatric disorder.

THE PREFRONTAL NETWORK FOR ATTENTION AND BEHAVIOR

Approximately one-third of all the cerebral cortex in the human brain is situated in the frontal lobes. The frontal lobes can be subdivided into motor-premotor, dorsolateral prefrontal, medial prefrontal, and orbitofrontal components. The terms *frontal lobe syndrome* and *prefrontal cortex* refer only to the last three of these four components. These are the parts of the cerebral cortex that show the greatest phylogenetic expansion in primates, especially in humans. The dorsolateral prefrontal, medial prefrontal, and orbitofrontal areas, along with the subcortical structures with which they are interconnected (i.e., the head of the caudate and the dorsomedial nucleus of the thalamus), collectively make up a large-scale network that coordinates exceedingly complex aspects of human

cognition and behavior.

The prefrontal network plays an important role in behaviors that require multitasking and the integration of thought with emotion. Its integrity appears important for the simultaneous awareness of context, options, consequences, relevance, and emotional impact that allows the formulation of adaptive inferences, decisions, and actions. Damage to this part of the brain impairs mental flexibility, reasoning, hypothesis formation, abstract thinking, foresight, judgment, the online (attentive) holding of information, and the ability to inhibit inappropriate responses. Cognitive operations impaired by prefrontal cortex lesions often are referred to as "executive functions."

Even very large bilateral prefrontal lesions may leave all sensory, motor, and basic cognitive functions intact while leading to isolated but dramatic alterations of personality and behavior. The most common clinical manifestations of damage to the prefrontal network take the form of two relatively distinct syndromes. In the *frontal abulic syndrome*, the patient shows a loss of initiative, creativity, and curiosity and displays a pervasive emotional blandness and apathy. In the *frontal disinhibition syndrome*, the patient becomes socially disinhibited and shows severe impairments of judgment, insight, and foresight. The dissociation between intact intellectual function and a total lack of even rudimentary common sense is striking. Despite the preservation of all essential memory functions, the patient cannot learn from experience and continues to display inappropriate behaviors without appearing to feel emotional pain, guilt, or regret when those behaviors repeatedly lead to disastrous consequences. The impairments may emerge only in real-life situations when behavior is under minimal external control and may not be apparent within the structured environment of the medical office. Testing judgment by asking patients what they would do if they detected a fire in a theater or found a stamped and addressed envelope on the road is not very informative since patients who answer these questions wisely in the office may still act very foolishly in the more complex real-life setting. The physician must therefore be prepared to make a diagnosis of frontal lobe disease on the basis of historic information alone even when the mental state is quite intact in the office examination.

Clinical Examination

The emergence of developmentally primitive reflexes, also known as frontal release signs, such as grasping (elicited by stroking the palm) and sucking (elicited by stroking the lips) are seen primarily in patients with large structural lesions that extend into the premotor components of the frontal lobes or in the context of metabolic encephalopathies. The vast majority of patients with prefrontal lesions and frontal lobe behavioral syndromes do not display these reflexes.

Damage to the frontal lobe disrupts a variety of attention-related functions, including working memory (the transient online holding of information), concentration span, the scanning and retrieval of stored information, the inhibition of immediate but inappropriate responses, and mental flexibility. The capacity for focusing on a trend of thought and the ability to shift the focus of attention voluntarily from one thought or stimulus to another can become impaired. Digit span (which should be seven forward and five reverse) is decreased; the recitation of the months of the year in reverse order (which should take less than 15 s) is slowed; and the fluency in producing words starting with the letter a, f, or s that can be generated in 1 min (normally ≥ 12 per letter) is diminished even in nonaphasic patients. Characteristically, there is a progressive slowing of performance as the task proceeds; e.g., a patient asked to count backward by threes may say "100, 97, 94,. . .91,. . .88," etc., and may not complete the task. In "go–no go" tasks (where the instruction is to raise the finger upon hearing one tap but keep it still upon hearing two taps), the patient shows a characteristic inability to keep still in response to the "no go" stimulus. Mental flexibility (tested by the ability to shift from one criterion to another in sorting or matching tasks) is impoverished; distractibility by irrelevant stimuli is increased; and there is a pronounced tendency for impersistence and perseveration.

These attentional deficits disrupt the orderly registration and retrieval of new information and lead to *secondary* memory deficits. Those memory deficits can be differentiated from the *primary* memory impairments of the amnestic state by showing that they improve when the attentional load of the task is decreased. Working memory (also known as immediate memory) is an attentional function based on the temporary online holding of information. It is closely associated with the integrity of the prefrontal network and the ascending reticular activating system. Retentive memory, in contrast, depends on the stable (offline) storage of information and is associated with the integrity of the limbic network. The distinction of the underlying neural mechanisms is illustrated by the observation that severely amnestic patients who cannot remember events that occurred a few minutes ago may have intact if not superior working memory capacity as shown in tests of digit span.

Causes: Trauma, Neoplasm, and Frontotemporal Dementia

The abulic syndrome tends to be associated with damage in dorsal prefrontal cortex, and the disinhibition syndrome with damage in ventral prefrontal cortex. These syndromes tend to arise almost exclusively after bilateral lesions. Unilateral lesions confined to the prefrontal cortex may remain silent until the pathology spreads to the other side; this explains why thromboembolic CVA is an unusual cause of the frontal lobe syndrome. Common settings for frontal lobe syndromes include head trauma, ruptured aneurysms, hydrocephalus, tumors (including metastases, glioblastoma, and falx or olfactory

groove meningiomas), and focal degenerative diseases. A major clinical form of FTLD known as the behavioral variant of frontotemporal dementia (bvFTD) causes a progressive frontal lobe syndrome that can start as early as the fifth decade of life. In these patients, the anterior temporal lobe and caudate nucleus are also atrophic. The behavioral changes can include shoplifting, compulsive gambling, sexual indiscretions, and obsessive-compulsive preoccupations, arising on a background of indifference. In many patients with Alzheimer's disease, neurofibrillary degeneration eventually spreads to prefrontal cortex and gives rise to components of the frontal lobe syndrome, but almost always on a background of severe memory impairment.

Lesions in the caudate nucleus or in the dorsomedial nucleus of the thalamus (subcortical components of the prefrontal network) also can produce a frontal lobe syndrome. This is one reason why the changes in mental state associated with degenerative basal ganglia diseases such as Parkinson's disease and Huntington's disease may take the form of a frontal lobe syndrome. Because of its widespread connections with other regions of association cortex, one essential computational role of the prefrontal network is to function as an integrator, or "orchestrator," for other networks. Bilateral multifocal lesions of the cerebral hemispheres, none of which are individually large enough to cause specific cognitive deficits such as aphasia and neglect, can collectively interfere with the connectivity and integrating function of the prefrontal cortex. A frontal lobe syndrome is the single most common behavioral profile associated with a variety of bilateral multifocal brain diseases, including metabolic encephalopathy, multiple sclerosis, and vitamin B12 deficiency, among others. Many patients with the clinical diagnosis of a frontal lobe syndrome tend to have lesions that do not involve prefrontal cortex but involve either the subcortical components of the prefrontal network or its connections with other parts of the brain. To avoid making a diagnosis of "frontal lobe syndrome" in a patient with no evidence of frontal cortex disease, it is advisable to use the diagnostic term *frontal network syndrome*, with the understanding that the responsible lesions can lie anywhere within this distributed network.

A patient with frontal lobe disease raises potential dilemmas in differential diagnosis: the abulia and blandness may be misinterpreted as depression, and the disinhibition as idiopathicmania or acting out. Appropriate intervention may be delayed while a treatable tumor keeps expanding. An informed approach to frontal lobe disease and its behavioral manifestations may help prevent such errors.

CARING FOR PATIENTS WITH DEFICITS OF HIGHER CEREBRAL FUNCTION

Some of the deficits described in this chapter are so complex that they may bewilder not only the patient and family but also the physician. It is imperative to carry out a systematic clinical evaluation to characterize the nature of the deficits and explain them in lay terms to the patient and family. Such an explanation can allay at least some of the anxieties, address the mistaken impression that the deficit (e.g., social disinhibition or inability to recognize family members) is psychologically motivated, and lead to practical suggestions for daily living activities. The consultation of a skilled neuropsychologist may aid in the formulation of diagnosis and management. Patients with simultanagnosia, for example, may benefit from the counterintuitive instruction to stand back when they cannot find an item so that a greater search area falls within the immediate field of gaze. Some patients with frontal lobe disease can be extremely irritable and abusive to spouses yet display all the appropriate social graces during a visit to the medical office. In such cases, the history may be more important than the bedside examination in charting a course of treatment.

Reactive depression is common in patients with higher cerebral dysfunction and should be treated. These patients may be overly sensitive to the usual doses of antidepressants or anxiolytics and require a careful titration of dosage. Brain damage may cause a dissociation between feeling states and their expression so that a patient who may superficially appear jocular could still be suffering from an underlying depression that needs to be treated. In many cases, agitation may be controlled with reassurance. In other cases, treatment with benzodiazepines, antiepilectics, or sedating antidepressants may become necessary. If neuroleptics become absolutely necessary for the control of agitation, atypical neuroleptics are preferable because of their lower extrapyramidal side effects. Treatment with neuroleptics in elderly patients with dementia requires weighing the potential benefits against the potentially serious side effects.

Spontaneous improvement of cognitive deficits due to acute neurologic lesions is common. It is most rapid in the first few weeks but may continue for up to 2 years, especially in young individuals with single brain lesions. The mechanisms for this recovery are incompletely understood. Some of the initial deficits appear to arise from remote dysfunction (diaschisis) in parts of the brain that are interconnected with the site of initial injury. Improvement in these patients may reflect, at least in part, a normalization of the remote dysfunction. Other mechanisms may involve functional reorganization in surviving neurons adjacent to the injury or the compensatory use of homologous structures, e.g., the right superior temporal gyrus with recovery from Wernicke's aphasia. In some patients with large lesions involving Broca's and Wernicke's areas, only Wernicke's area may show contralateral compensatory reorganization (or bilateral functionality), giving rise to a situation in which a lesion that should have caused a global aphasia becomes associated with a residual Broca's aphasia. Prognosis for recovery from aphasia is best when Wernicke's area is spared. Cognitive rehabilitation procedures have been used in the treatment of higher cortical deficits. There are few controlled studies, but some

show a benefit of rehabilitation in the recovery from hemispatial neglect and aphasia. Some types of deficits may be more prone to recovery than others. For example, patients with CVA and nonfluent aphasias are more likely to benefit from speech therapy than are patients with fluent aphasias and comprehension deficits. In general, lesions that lead to a denial of illness (e.g., anosognosia) are associated with cognitive deficits that are more resistant to rehabilitation. Periodic neuropsychological assessment is necessary for quantifying the pace of the improvement (or of the progression in the case of dementias) and for generating specific recommendations for cognitive rehabilitation, modifications in the home environment, the timetable for returning to work in patients recovering from acute lesions, and the scheduling of retirement or disability status in patients with degenerative diseases. Determining driving competence is challenging, especially in the early stages of dementing diseases. The diagnosis of a neurodegenerative disease is not by itself sufficient for asking the patient to stop driving. An on-the-road driving test and reports from family members may help time decisions related to this very important activity.

There is a mistaken belief that dementias are anatomically diffuse and that they cause global cognitive impairments. This is true only at the terminal stages. During most of the clinical course, dementias are exquisitely selective with respect to anatomy and cognitive pattern. Alzheimer's disease, for example, causes the greatest destruction in medial temporal areas belonging to the memory network and is clinically characterized by a correspondingly severe amnesia. There are other dementias in which memory is intact. Selective degeneration of the frontal lobes in FTLD leads to a gradual dissolution of behavior and executive functions. Primary progressive aphasia is characterized by a gradual atrophy of the left perisylvian language network and a selective dissolution of language that can remain isolated for up to 10 years. An enlightened approach to the differential diagnosis and to the individualized care of patients with acute and progressive damage to the cerebral cortex requires an understanding of the principles that link neural networks to higher cerebral functions.

FURTHER READINGS

BARTSCH T el al: Selective affection of hippocampal CA-1 neurons in patients with transient global amnesia without long-term sequelae. *Brain,* 129**:**2874, 2006 [PMID: 17003071]

CATANI M et al: Perisylvian language networks of the human brain. Ann Neurol 57:8, 2006

DELEON J et al: Neural regions essential for distinct cognitive processes underlying picture naming. Brain, 130:1408, 2007 [PMID: 17337482]

DORICCI F et al: White matter (dis)connections and gray matter (dys)functions in visual neglect: Gaining insights into the brain networks of spatial awareness. Cortex 44:983, 2008

HEISS W-D et al: Differential capacity of left and right hemispheric areas for compensation of poststroke aphasia. Ann Neurol 45:430, 1999 [PMID: 10211466]

LEIGUARDA RC, MARSDEN CD: Limb apraxias: Higher-order disorders of sensorimotor integration. Brain 123:860, 2000 [PMID: 10775533]

MESULAM M: Representation, inference and transcendent encoding in neurocognitive networks of the human brain. Ann Neurol 64:367, 2008 [PMID: 18991346]

——— et al: Quantitative template for subtyping primary progressive aphasia. Arch Neurol 66:1545, 2009

RUFF CC et al: Hemispheric differences in frontal and parietal influences on human occipital cortex: Direct confirmation with concurrent TMS-fMRI. J Cogn Neurosci 21:1146, 2008

RUSCONI E et al: A disconnection account of Gerstmann syndrome: Functional neuroanatomy evidence. Ann Neurol 66:654, 2009 [PMID: 19938150]

SCHMAHMANN JD, PANDYA DN: Disconnection syndromes of basal ganglia, thalamus, and cerebrocerebellar systems. Cortex 44:1037, 2008 [PMID: 18614161]

SEELEY W et al: Early frontotemporal dementia targets neurons unique to apes and humans. Ann Neurol 60:660, 2006 [PMID: 17187353]

WEINTRAUB S, MESULAM M: With or without FUS, it is the anatomy that dictates the dementia phenotype. Brain 132:2906, 2009 [PMID: 19861505]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MEMORY LOSS: INTRODUCTION

Memory allows us to store, retain, and retrieve information. These three processes influence and are modified by the type of information that is to be remembered, the duration of time over which it must be retained, and the way in which the brain will use the information in the future. The neural circuits underlying these processes are dynamic, reflecting the flexibility of memory itself. To delineate the neural circuitry underlying it, it is helpful to break down memory into simpler components. This categorization, however, need not lead to the assumption that memory is not a unitary phenomenon.

LONG-TERM MEMORY

In an effort to explain why focal brain damage affects some aspects of memory but not others, a fundamental distinction has been made between *declarative memory*, which refers to the conscious memory for facts and events, and *nondeclarative memory*, which refers to memory for skills, habits, or other manifestations of learning that can be expressed without awareness of what was learned **(Fig. e9-1)**. Patients with bilateral medial temporal lobe (MTL) damage show declarative memory deficits in the face of intact nondeclarative memory. For example, such a patient may learn to play the piano over time without remembering a single practice session or even recognizing the teacher who patiently works with him everyday.

Figure e9-1

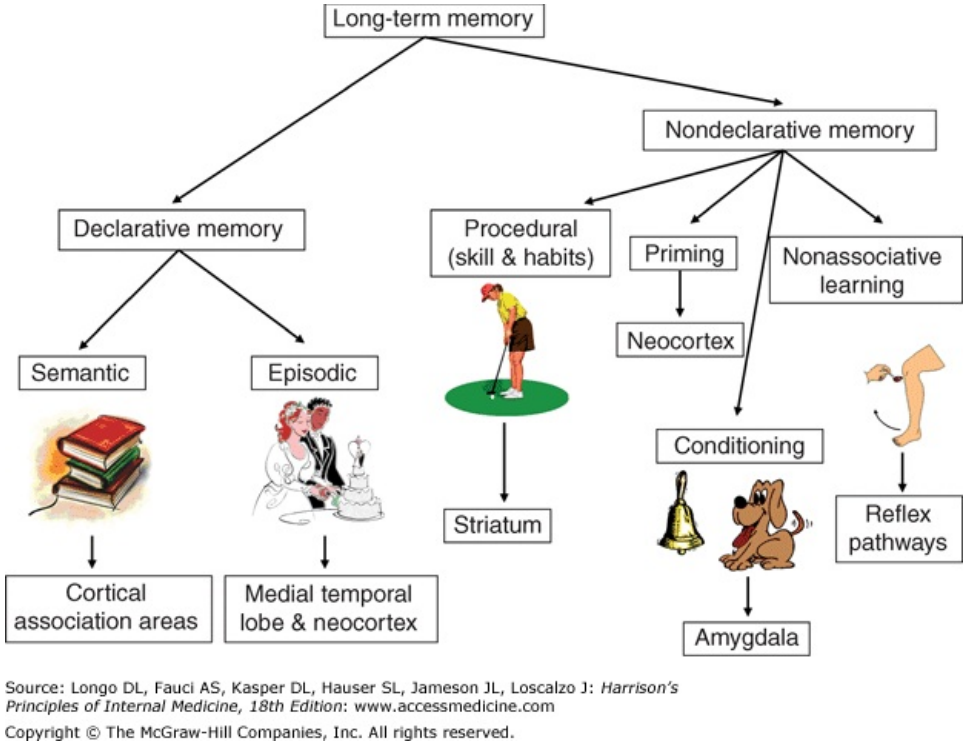

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fractionation of long-term memory. *(Adapted from LR Squire, SM Zola: Proc Natl Acad Sci U S A, 24: 13515, 1996.)*

Declarative Memory

Within the declarative memory system, *episodic* and *semantic* memory can be distinguished. *Episodic memory* allows the recollection of unique personal experiences. With episodic memory, the person reexperiences the sights, sounds, smells, and other details of a specific event. Many episodic memories are kept for minutes and hours but soon discarded. Others remain for the course of a lifetime. This temporal difference in storage probably reflects different physiologic processes at work (see below). *Semantic memory*, in contrast, refers to knowledge about the world; generic information that is acquired across many different contexts and accessed without accompanying details of the time

when the words or facts were remembered. One's vocabulary and knowledge of the associations between verbal concepts make up the bulk of semantic memory. This fractionation of declarative memory is supported by evidence that episodic and semantic memory have distinctive anatomic substrates.

EPISODIC MEMORY

In the MTL, the hippocampal formation receives processed sensory information from association areas in the frontal, parietal, and occipital lobes via the parahippocampal cortex. Given these multiple cortical neuroanatomic connections, the hippocampus is well placed to create associations between simultaneously occurring stimuli in our sensory world. Key structures involved with episodic memory include the hippocampus, entorhinal cortex, mammillary bodies, and thalamus. Alterations of episodic memory can be devastating. Overly intense or painful episodic memories can result in posttraumatic stress disorder, a devastating illness in which patients repeatedly reexperience unpleasant episodes from their lives. By contrast, loss of episodic memories, as in Alzheimer's disease (AD), will prevent the individual from learning new things about the world and will eventually strip away the old memories that constitute a life biography.

Given its anatomic placement and architecture, the hippocampus has the unique ability to bind together "what happened," "when it happened," and "where it happened." The architecture of the hippocampus includes a circular pathway of neurons from the entorhinal cortex to the dentate gyrus and CA3 and CA1 neurons of the hippocampus to the subiculum and back to the entorhinal cortex. This pathway is heavily damaged in AD. Individual elements of episodic memories are permanently stored within the same neocortical regions that are involved in initial processing and analyzing of sensory information (neocortex). Each different cortical region makes a unique contribution to the storage of a given memory, and all regions participate together in the creation of a complete memory representation. The hippocampal formation, then, is saddled with the task of binding together these different regional contributions into a coherent memory trace. The connections within the hippocampal formation and between the MTL and neocortical regions are formed more rapidly than are the connections between disparate cortical regions. Therefore, when a particular cue in the environment or the mental state of the person activates cells in the cortical regions, the MTL network that is associated with that cue is reactivated and the entire neocortical representation is strengthened. As multiple reactivations occur, the connections between the relevant neocortical regions are slowly strengthened until the memory trace no longer depends on the MTL's activity but is instead entirely represented in the cortex.

Recent work in humans and rodents has shown that the hippocampus and associated cortical structures are involved not only in the remembrance of things past but also in the imagination and/or prediction of the future. FMRI studies have shown that this region is active when people imagine the future, and patients with bilateral hippocampal damage are unable to conceive of future events. In rats, hippocampal ensembles preplay and replay event sequences even in the absence of overt behavior. These event sequences may then be recombined to form new events as a means of predicting the future. Interestingly, the hippocampal system remains active during sleep and "task-free" thought, suggesting that it is likely involved in imagination and free-form thinking.

While the MTL learns quickly and has a limited capacity for storage, the neocortex learns slowly and has a very large storage capacity. In both regions, learning occurs via Hebbian synapses, whereby "cells that fire together, wire together." With repeated activations, memories become "consolidated" in the neocortex and, therefore, independent of the MTL. This process, by which the burden of long-term (permanent) memory storage is gradually assumed by the neocortex, ensures that the MTL system is constantly available for the acquisition of new information. Recent evidence, however, points to the hippocampus as serving a critical function in the retrieval of detailed episodic memories, regardless of the age of the memory.

Injury anywhere along this septohippocampal pathway can lead to severe loss of episodic memory. Patients with injury to this system will exhibit anterograde amnesia, an inability to commit new information to memory. Memories that were established before the injury (remote memories) tend to be relatively preserved, although a retrograde amnesia, going back anywhere from minutes to years, is usually evident. Larger lesions cause a more extensive retrograde memory deficit. Also, as brain injury improves over time, the retrograde memory impairment tends to diminish, and the temporal gradient for memory loss shrinks.

The most common cause for entorhinal dysfunction is AD, which begins in the entorhinal cortex and then spreads to the hippocampus. A common mechanism for hippocampal dysfunction is traumatic injury because the hippocampi sit adjacent to, and are easily pushed against, bone in the middle cranial fossa. The hippocampus, particularly the CA1 and subicular region, is very sensitive to metabolic insults, including hypoxia, hypoglycemia, and prolonged seizures. Vascular infarction can occur with occlusion of the hippocampal branches off the posterior cerebral arteries. Infections, in particular herpes simplex encephalitis, can selectively attack the medial temporal regions, leading to severe and permanent deficits in episodic memory.

Additionally, severe loss of episodic memory can also be due to dysfunction in the mammillothalamic memory system. The amnesia of Korsakoff's syndrome is due to injury from hemorrhage into the mammillary bodies and dorsomedial nuclei of the thalamus. Furthermore, recent studies of patients with stroke in the left dorsomedial nucleus of the thalamus suggest that injury here alone will precipitate a severe amnesia. Finally, tumors can injure the septum, fornix, or MTLs, leading to amnesia.

Emotion plays a key role in enhancing the ability to remember personal episodes and other information encoded in a particular affective state. Emotionally charged events are more easily remembered than emotionally neutral episodes, and highly vivid "flashbulb" memories are often laid down during traumatic or emotional events, sometimes during positive but often during negative events. In humans, the amygdala modulates memory processes during emotional experiences.

One simple way to test episodic memory is to ask the patient to recall recent events such as what he or she did on the last big holiday, what is going on in the news, or what he or she had for breakfast. With regard to personal episodic memories, it is always necessary to have a historian who can verify that the recent memories are correct (not confabulated).

SEMANTIC MEMORY

Unlike episodic memory, the recall of semantic memory does not lead to the retrieval of details of when, or where, the information was acquired. For example, we remember that a fork is a utensil that is used for eating food without remembering when we learned the word *fork* or when we discovered its use. Semantic memory is composed of a complex hierarchy of knowledge about the world. Knowing that a fork is generally used for eating depends on understanding that in certain social situations, eating with only our hands is inappropriate, and that some foods are more easily eaten with a fork than another available utensil, such as a spoon. While a fork may be useful in many different situations, our semantic hierarchy reminds us that its main function is to facilitate eating. These ideas are held together in the semantic memory system, which spans across the association areas of the neocortex. Therefore, if we are in a situation that requires using a fork as a tool in a novel manner, we can still call upon our semantic memory system to aid us in solving the problem.

Evidence that semantic memories are independent of the septohippocampal and mamillothalamic memory systems comes from humans with injury to these systems who maintain access to semantic knowledge despite profound deficits in episodic memory. In contrast, patients with primarily anterior and lateral temporal lobe damage show intact episodic memory but impaired semantic memory. The finding that children born with hippocampal sclerosis and lifelong episodic memory impairments can still function fairly well in school suggests that semantic memories are not wholly dependent upon intact episodic memory.

In semantic dementia, a syndrome associated with neurodegenerative disease that begins in the anterior temporal lobes, both the simple labeling process (naming) and knowledge about the identity of people and objects are lost. Patients with semantic dementia classify objects into increasingly superordinate categories, having lost access to specific exemplars. Hence, a hawk becomes a "hunting bird," then a "bird," then an "animal," and then a "thing" as the disease worsens. Eventually all objects are classified with a series of simple stereotyped phrases. Bilateral anterior temporal dysfunction is the anatomic substrate of semantic dementia, a subtype of the frontotemporal lobar degenerations. Aside from semantic dementia, the other disorders that lead to this syndrome include limbic encephalitis, associated with viral or paraneoplastic processes, and herpes simplex encephalitis.

Bedside assessment of semantic memory is difficult, but thegravest deficits may be seen if the patient is unable to name common objects such as a pen or watch or less common objects such as a stethoscope or a fluorescent bulb.

Nondeclarative Memory

Nondeclarative memory is an umbrella term for a heterogeneous collection of nonconscious memory abilities that involve multiple distinct neural regions, including the amygdala, basal ganglia, cerebellum, and sensory cortex (Fig. e9-1). *Procedural memory* is one type of nondeclarative memory. The difference between declarative memory and procedural memory is the difference between "knowing that" and "knowing how." *Procedural learning* describes the formation of skills and habits. Because it requires extensive practice, it is a slow and inflexible learning system that eventually takes on an automatic or reflexive quality. It is, however, long-lasting and reliable: even after years of absence from a bicycle, a bike rider does not lose the skill entirely.

Procedural memory involves motor, perceptual, and cognitive processes. For example, flipping pancakes is a motor skill, a parent's attentiveness to his or her baby's cry in a distant room involves perceptual learning, and increasing alacrity in solving Sudoku puzzles with practice requires cognitive skills. While declarative memory can, in some cases, enhance or hasten the acquisition of skills and habits, conscious awareness of learning is not necessary; once the information is acquired, it often becomes difficult to verbalize how it was learned. Cognitive psychologists have shown that in some cases, declarative memory processes can hinder nondeclarative learning, suggesting that there are

times when the two memory "systems" may compete for cognitive resources.

The forms of perceptual and motor learning that can occur without conscious recollections are mediated in part by contractions and expansions of representations in the sensory and motor cortex. One study, for example, has shown that the cortical representation of the fingers of the left hand of musical string players is larger than that in nonmusicians, suggesting that the representation of different parts of the body in the primary somatosensory cortex of humans depends on use and changes to conform to the current needs and experiences of the individual. Discrete cortical regions exist in the anterior temporal lobes in which object knowledge (such as words related to color, animals, tools, or action) is organized as a distributed system. Here the attributes of an object are stored close to the regions of the cortex that mediate perception of those attributes.

Recent research now points to the basal ganglia as fundamental in motor skill learning, while the cerebellum is involved in the association of a visual cue with a motor action. Parkinson's disease (PD) causes damage to the basal ganglia and is associated with impairments in habit learning but spares declarative memory. The basal ganglia project to and receive projections from the frontal cortex, and this corticostriatal loop has been implicated in the learning of skills and habits. Furthermore, recent functional MRI work suggests that the MTL-based declarative memory and the corticostriatal procedural memory systems operate independently from each other and may in fact compete for cognitive resources. That is, basal ganglia activity is negatively correlated with MTL activity when both systems are engaged by a particular task.

Bedside testing of nondeclarative memory is outside the realm of the generalist, but deficits may be reported by patients or their families.

Molecular and Neurochemical Basis of Long-Term Memory

Long-term potentiation (LTP), which refers to a long-lasting enhancement of synaptic transmission resulting from repetitive stimulation of excitatory synapses, is presumed to be involved in episodic memory acquisition and storage. LTP occurs in the hippocampus and is mediated by *N*-methyl-d-aspartate (NMDA) receptors as well as the cyclic AMP–responsive element-binding (CREB) protein. Animal experiments have shown that the formation of new episodic memories leads to physiologic changes in the synapse, while longer-term memory requires new protein synthesis and leads to physical changes at neuronal synapses. Recent work has also begun to suggest mechanisms by which memories are allocated to specific neural circuits within a neural network at different time scales. For example, neurogenesis in the dentate gyrus may be a mechanism by which memories are allocated to specific cells: young neurons are more excitable and their synapses are more plastic than more mature neurons, giving them a "competitive advantage" during encoding, and thereby creating a "timestamp" of when a particular event that is to be remembered occurred. Whereas neurogenesis may be one mechanism by which memories are time-stamped on the scale of days to weeks, LTP at a specific synapse in a given CA1 pyramidal neuron lowers the threshold for the induction of changes in nearby synapses within a 10-μm radius for up to 10 minutes, suggesting that events clustering closely in time in the order of minutes are more likely to be allocated to neighboring neurons in the CA1.

The cholinergic system also plays an important role in memory, and anticholinergic agents such as atropine and scopolamine interfere with memory. Choline acetyltransferase (the enzyme catalyzing the formation of acetylcholine) is known to be deficient in the cortex of patients with AD. The brains of AD patients show severe neuronal loss in the nucleus basalis of Meynert, the major source of cholinergic input to the cerebral cortex. These findings form the basis for the use of cholinesterase inhibitors in the treatment of AD, with benefits thought to arise from increased levels of available acetylcholine. Behavior and mood are modulated by noradrenergic, serotonergic, and dopaminergic pathways, and norepinephrine has been shown to be reduced in the brainstem locus coeruleus in AD. Also, neurotrophins are postulated to play a role in memory in part by preserving cholinergic neurons. GABA agonists including the benzodiazepines are associated with reversible but sometimes severe episodes of amnesia. Working memory (see below) is strongly modulated by dopamine.

SHORT-TERM MEMORY

Working Memory

While the fractionation of memory into declarative and nondeclarative systems has provided a reasonable framework for understanding many aspects of memory's neurologic underpinnings, another major division of memory has used time as the distinguishing characteristic. While some information is retained for only a few seconds—enough time to hear, remember, and dial a phone number—other memories are seemingly remembered throughout a life span. This brief type of memory differs from long-term memory, not only in terms of duration of retention but also with regard to its function and neuroanatomy.

Working memory stores items only as long as the information is in consciousness and is either being rehearsed (subvocally) or manipulated in some other fashion (i.e., rotated or integrated with existing information in semantic memory). The capacity of working memory is limited by attention. Normal individuals can hold about seven (plus or minus two) "bits" of information in working memory; these bits

can be manipulated and either discarded or associated and transferred into long-term memory. Working memory is highly vulnerable to distraction and sometimes is even called *working attention* to emphasize the conscious and effortful processes that it entails. In the most widely accepted conceptualization of working memory, there are four main components: (1) a central executive that keeps track of and gathers information; (2) a visual system called the *visuospatial scratchpad*, which holds visual representations of objects; (3) a phonologic "system" that holds verbal information; and (4) an episodic buffer that is capable of binding together information from different modalities into a coherent trace.

Lesions that disrupt the structure or function of the dorsolateral frontal or posterior parietal regions decimate working memory. These deficits in working memory have a profound effect on the organism by disrupting the learning process downstream to working memory, or by affecting activities that directly depend on an intact working memory. In the classic amnesic syndrome, patients have intact working memory but cannot transfer information from working memory into long-term store.

Single-cell recordings have uncovered a network of neurons in the posterior parietal and dorsolateral frontal lobes where activity is high only during periods when information is being held in memory for use over just a few seconds. These neurons appear to provide an important functional basis for working memory. Similarly, functional imaging studies from humans show that the dorsolateral frontal lobes, particularly Brodmann area 46, are critical for working memory. A recent meta-analysis of working memory studies in patients with Alzheimer's disease suggests that whereas the phonological loop remains intact early in the disease, some deficits attributed to dysfunction in the central executive system may present even in preclinical stages, perhaps related to impairments in the episodic buffer.

Testing Memory at the Bedside

Testing of memory should be performed in anyone in whom memory deficits are a concern, whether these concerns are raised by the patient, family, or health care workers. If the deficits are subtle, the testing may require a comprehensive consultation with a neuropsychologist, neuropsychiatrist, or behavioral neurologist. However, memory testing can be an extremely valuable component of the neurologic examination and performed effectively at the bedside. There are a wide variety of brief standardized screens of cognition, but the most commonly used test is the Mini-Mental Status Examination (Table 371-5), a 30-point test that is strongly dependent on working (spell "world" backwards) and episodic memory (orientation and three-word recall). Testing semantic and procedural memory is usually outside the realm of the generalist, but if deficits in these systems are suspected, further tests are warranted.

Of all the memory processes, working memory is perhaps the easiest to assess at the bedside. The most common bedside test of working memory involves asking patients to repeat a series of digits orally, with the clinician gradually increasing the number of to-be-retained digits. There are two ways of administering the test. Asking the patient to repeat the digits in the same order as they were delivered is called *digit span forward*. In contrast, the clinician may also ask the patient to repeat the digits in reverse order, called *digit span backward*. Digit span forward is a test of attention, while digit span backward is a simple probe of working memory. The capacity for digit span forward is typically six numbers, while normal adults can generally repeat five digits backward.

FURTHER READINGS

Baddeley A: Working memory: Looking back and looking forward. Nat Rev Neurosci 4:829, 2003[PMID: 14523382]

Buckner RL: The role of the hippocampus in prediction and imagination. Annu Rev Psych 61:27, 2010[PMID: 19958178]

Eichenbaum H et al: The medial temporal lobe and recognition memory. Ann Rev Neurosci 30:123, 2007[PMID: 17417939]

Gilboa A et al: Retrieval of autobiographical memory in Alzheimer's disease: Relation to volumes of medial temporal lobe and other structures. Hippocampus 15:535, 2005[PMID: 15884035]

Nadel L, Moscovitch M: Memory consolidation, retrograde amnesia and the hippocampal complex. Curr Opin Neurobiol 7:217, 1997[PMID: 9142752]

Packard MG, Knowlton BJ: Learning and memory functions of the basal ganglia. Annu Rev Neurosci 25:563, 2002[PMID: 12052921]

Perry RJ, Hodges JR: Spectrum of memory dysfunction in degenerative disease. Curr Opin Neurol 9:281, 1996[PMID: 8858186]

Poldrack RA et al: Interactive memory systems in the human brain. Nature 414:546, 2001[PMID:

11734855]

Redondo RL, Morris RG: Making memories last: The synaptic tagging and capture hypothesis. Nat Rev Neurosci 12:17, 2011[PMID: 21170072]

Silva AJ et al: Molecular and cellular approaches to memory allocation in neural circuits. Science 326:391, 2009[PMID: 19833959]

Squire LR, Zola-Morgan S: The medial temporal lobe memory system. Science 253:1380, 1991[PMID: 1896849]

——— et al: The medial temporal lobe. Annu Rev Neurosci 27:279, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRIMARY PROGRESSIVE APHASIA, MEMORY LOSS, AND OTHER FOCAL CEREBRAL DISORDERS

Language and memory are essential human functions. For the experienced clinician, the recognition of different types of language and memory disturbances often provides essential clues to the anatomic localization and diagnosis of neurologic disorders. This video illustrates classic disorders of language and speech (including the aphasias), memory (the amnesias), and other disorders of cognition that are commonly encountered in clinical practice.

Video e10-1
Primary Progressive Aphasia, Memory Loss, and Other Focal Cerebral Disorders

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SLEEP DISORDERS: INTRODUCTION

Disturbed sleep is among the most frequent health complaints physicians encounter. More than one-half of adults in the United States experience at least intermittent sleep disturbance. For most, it is an occasional night of poor sleep or daytime sleepiness. However, the Institute of Medicine has estimated that 50–70 million Americans suffer from a chronic disorder of sleep and wakefulness, which can lead to serious impairment of daytime functioning. In addition, such problems may contribute to or exacerbate medical or psychiatric conditions. Thirty years ago, many such complaints were treated with hypnotic medications without further diagnostic evaluation. Since then, a distinct class of sleep and arousal disorders has been identified.

PHYSIOLOGY OF SLEEP AND WAKEFULNESS

Given the opportunity, most adults will sleep 7–8 h per night, although the timing, duration, and internal structure of sleep vary among healthy individuals and as a function of age. At the extremes, infants and the elderly have frequent interruptions of sleep. In the United States, adults tend to have one consolidated sleep episode per day, although in some cultures sleep may be divided into a mid-afternoon nap and a shortened night sleep. Two principal neural systems govern the expression of the sleep and wakefulness states within the daily cycle. The first potentiates sleep in proportion to the duration of wakefulness (the "sleep homeostat"), while the second rhythmically modulates sleep and wakefulness tendencies at appropriate phases of the 24-h day (the circadian clock). Intrinsic abnormalities in the function of either of these systems, or extrinsic disturbances (environmental, drug- or illness-related) that supersede their normal expression, can lead to clinically recognizable sleep disorders.

States and Stages of Sleep

States and stages of human sleep are defined on the basis of characteristic patterns in the electroencephalogram (EEG), the electrooculogram (EOG—a measure of eye-movement activity), and the surface electromyogram (EMG) measured on the chin and neck. The continuous recording of this array of electrophysiologic parameters to define sleep and wakefulness is termed *polysomnography*.

Polysomnographic profiles define two states of sleep: (1) rapid-eye-movement (REM) sleep and (2) non-rapid-eye-movement (NREM) sleep. NREM sleep is further subdivided into three stages, characterized by increasing arousal threshold and slowing of the cortical EEG. REM sleep is characterized by a low-amplitude, mixed-frequency EEG similar to that of NREM stage N1 sleep. The EOG shows bursts of REM similar to those seen during eyes-open wakefulness. Chin EMG activity is absent, reflecting the brainstem-mediated muscle atonia that is characteristic of that state.

Organization of Human Sleep

Normal nocturnal sleep in adults displays a consistent organization from night to night **(Fig. 27-1)**. After sleep onset, sleep usually progresses through NREM stages N1–N3 sleep within 45–60 min. Slow-wave sleep (NREM stage N3 sleep) predominates in the first third of the night and comprises 15–25% of total nocturnal sleep time in young adults. The percentage of slow-wave sleep is influenced by several factors, most notably age (see below). Prior sleep deprivation increases the rapidity of sleep onset and both the intensity and amount of slow-wave sleep.

Figure 27-1

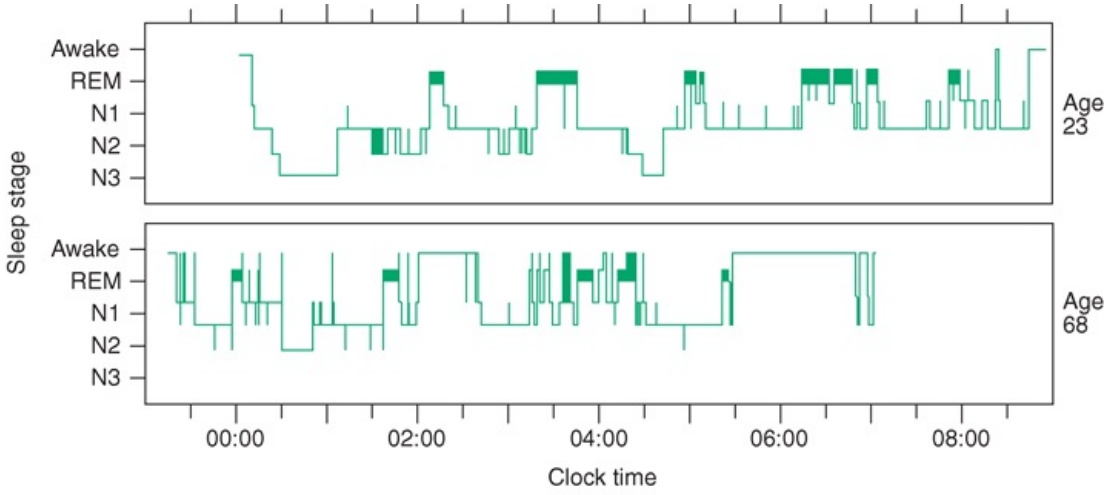

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stages of REM sleep (solid bars), the three stages of NREM sleep, and wakefulness over the course of the entire night for representative young and older adult men. Characteristic features of sleep in older people include reduction of slow-wave sleep, frequent spontaneous awakenings, early sleep onset, and early morning awakening. *(From the Division of Sleep Medicine, Brigham and Women's Hospital.)*

The first REM sleep episode usually occurs in the second hour of sleep. More rapid onset of REM sleep in an adult (particularly if <30 min) may suggest pathology such as endogenous depression, narcolepsy, circadian rhythm disorders, or drug withdrawal. NREM and REM alternate through the night with an average period of 90–110 min (the "ultradian" sleep cycle). Overall, REM sleep constitutes 20–25% of total sleep, and NREM stages N1 and N2 are 50–60%.

Age has a profound impact on sleep state organization (Fig. 27-1). Slow-wave sleep is most intense and prominent during childhood, decreasing sharply coincident with puberty and across the second and third decades of life. After age 30, there is a continued decline in the amount of slow-wave sleep, and the amplitude of delta EEG activity comprising slow-wave sleep is profoundly reduced. The depth of slow-wave sleep, as measured by the arousal threshold to auditory stimulation, also decreases with age. In the otherwise healthy older person, slow-wave sleep may be completely absent, particularly in males. Paradoxically, older people are better able to tolerate acute sleep deprivation than young adults, maintaining reaction time and sustaining vigilance with fewer lapses of attention.

A different age profile exists for REM sleep than for slow-wave sleep. In infancy, REM sleep may comprise 50% of total sleep time, and the percentage is inversely proportional to developmental age. The amount of REM sleep falls off sharply over the first postnatal year as a mature REM-NREM cycle develops; thereafter, REM sleep occupies a relatively constant percentage of total sleep time.

Neuroanatomy of Sleep

Experimental studies in animals have variously implicated the medullary reticular formation, the thalamus, and the basal forebrain in the generation of sleep, while the brainstem reticular formation, the midbrain, the subthalamus, the thalamus, and the basal forebrain have all been suggested to play a role in the generation of wakefulness or EEG arousal.

Current models suggest that the capacity for sleep and wakefulness generation is distributed along an axial "core" of neurons extending from the brainstem rostrally to the basal forebrain. A cluster of γ-aminobutyric acid (GABA) and galaninergic neurons in the ventrolateral preoptic (VLPO) hypothalamus is selectively activated coincident with sleep onset. These neurons project to and inhibit the multiple neural wakefulness centers that comprise the ascending arousal system, and selective cell-specific lesions of VLPO substantially reduce sleep time, indicating that the hypothalamic VLPO neurons play an executive role in sleep regulation. More recent data have identified another sleep center, the median preoptic nucleus (MnPOn) of the hypothalamus with similar activation patterns and projections, suggesting that, like that of wakefulness, executive control of sleep may also be multicentric.

Specific regions in the pons are associated with the neurophysiologic correlates of REM sleep. Small lesions in the dorsal pons result in the loss of the descending muscle inhibition normally associated with REM sleep; microinjections of the cholinergic agonist carbachol into the pontine reticular formation produce a state with all of the features of REM sleep. These experimental manipulations are mimicked by pathologic conditions in humans and animals. A prominent feature of narcolepsy, for example, is abrupt, complete, or partial paralysis (cataplexy) in response to a variety of stimuli, a pathologic activation of neural systems mediating the atonia of normal REM sleep. In narcoleptic dogs, physostigmine, a central cholinesterase inhibitor, increases the frequency of cataplexy episodes, while atropine decreases their frequency. Conversely, in REM sleep behavior disorder (see below), patients

suffer from a failure of normal motor inhibition during REM sleep, resulting in involuntary, occasionally violent, movement arising out of dream episodes.

Neurochemistry of Sleep

Early experimental studies that focused on the raphe nuclei of the brainstem appeared to implicate serotonin as the primary sleep-promoting neurotransmitter, while catecholamines were considered to be responsible for wakefulness. Simple neurochemical models have given way to more complex formulations involving multiple parallel waking systems. Pharmacologic studies suggest that histamine, acetylcholine, dopamine, serotonin, and noradrenaline are all involved in wake promotion. A novel neuropeptide, orexin (also known as hypocretin), localized to a cluster of neurons in the lateral hypothalamus and originally identified based on its pathogenic role in narcolepsy (see below), also appears to be involved in the control of wakefulness.

In the basal forebrain (BF), adenosine receptors on cholinergic neurons are thought to play a role in assessing homeostatic sleep need by providing an index of cellular energy status. Projections from these BF neurons to executive sleep centers such as VLPO thus allow incorporation of homeostatic sleep need into the control of sleep state expression. At a practical level, this model suggests that the alerting effects of caffeine, an adenosine receptor antagonist, reflect the attenuation of the homeostatic signal from the basal forebrain.

The prominent hypnotic effects of benzodiazepine receptor agonists suggest that endogenous ligands of this receptor may be involved in normal sleep physiology. While neurosteroids with activity at this receptor have been identified, their role in normal sleep-wake control remains unclear. In addition, a broad array of endogenous sleep- and wake-promoting substances have been identified, the role of which in normal sleep-wake control remains unclear. These include corticotropin-releasing hormone (CRH), prostaglandin D_2, delta sleep–inducing peptide, muramyl dipeptide, interleukin 1, fatty acid primary amides, and melatonin.

Physiology of Circadian Rhythmicity

The sleep-wake cycle is the most evident of the many 24-h rhythms in humans. Prominent daily variations also occur in endocrine, thermoregulatory, cardiac, pulmonary, renal, gastrointestinal, and neurobehavioral functions. At the molecular level, endogenous circadian rhythmicity is driven by self-sustaining transcriptional/translational feedback loops **(Fig. 27-2)**. In evaluating a daily variation in humans, it is important to distinguish between those rhythmic components passively evoked by periodic environmental or behavioral changes (e.g., the increase in blood pressure and heart rate that occurs upon assumption of the upright posture) and those actively driven by an endogenous oscillatory process (e.g., the circadian variation in plasma cortisol that persists under a variety of environmental and behavioral conditions).

Figure 27-2

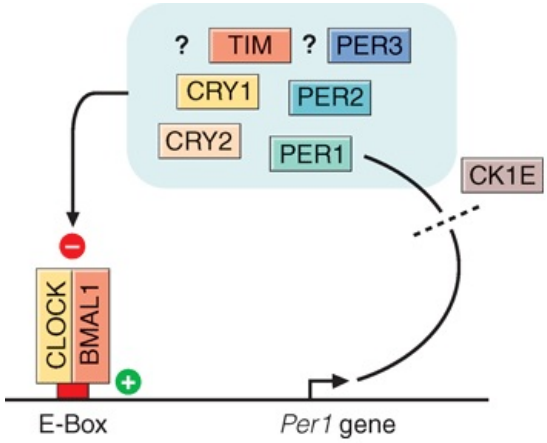

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Model of the molecular feedback loop at the core of the mammalian circadian clock. The positive element of the feedback loop (+) is the transcriptional activation of the *Per1* gene (and probably other clock genes) by a heterodimer of the transcription factors CLOCK and BMAL1 (also called MOP3) bound to an E-box DNA regulatory element. The *Per1* transcript and its product, the clock component PER1 protein, accumulate in the cell cytoplasm. As it accumulates, the PER1 protein is recruited into a multiprotein complex thought to contain other circadian clock component proteins such as cryptochromes (CRYs), Period proteins (PERs), and others. This complex is then transported into the cell nucleus (across the dotted line), where it functions as the negative element in the feedback loop (–) by inhibiting the activity of the CLOCK-BMAL1 transcription factor heterodimer. As a consequence of this action, the concentration of PER1 and other clock proteins in the inhibitory

As a consequence of this action, the concentration of PER and other clock proteins in the inhibitory complex falls, allowing CLOCK-BMAL1 to activate transcription of *Per1* and other genes and begin another cycle. The dynamics of the 24-h molecular cycle are controlled at several levels, including regulation of the rate of PER protein degradation by casein kinase-1 epsilon (CK1E). Additional limbs of this genetic regulatory network, omitted for the sake of clarity, are thought to contribute stability. Question marks denote putative clock proteins, such as Timeless (TIM), as yet lacking genetic proof of a role in the mammalian clock mechanism. *(Copyright Charles J. Weitz, PhD, Department of Neurobiology, Harvard Medical School.)*

While it is now recognized that many peripheral tissues in mammals have circadian clocks that regulate diverse physiologic processes, these independent tissue-specific oscillations are coordinated by a central neural pacemaker located in the suprachiasmatic nuclei (SCN) of the hypothalamus. Bilateral destruction of these nuclei results in a loss of the endogenous circadian rhythm of locomotor activity, which can be restored only by transplantation of the same structure from a donor animal. The genetically determined period of this endogenous neural oscillator, which averages ~24.2 h in humans, is normally synchronized to the 24-h period of the environmental light-dark cycle. Small differences in circadian period underlie variations in diurnal preference, with the circadian period shorter in individuals who typically rise early compared to those who typically go to bed late. Entrainment of mammalian circadian rhythms by the light-dark cycle is mediated via the retinohypothalamic tract, a monosynaptic pathway that links specialized, photoreceptive retinal ganglion cells directly to the SCN. Humans are exquisitely sensitive to the resetting effects of light, particularly the shorter wavelengths (~460–500 nm) of the visible spectrum.

The timing and internal architecture of sleep are directly coupled to the output of the endogenous circadian pacemaker. Paradoxically, the endogenous circadian rhythms of sleep tendency, sleepiness, and REM sleep propensity all peak near the habitual wake time, just after the nadir of the endogenous circadian temperature cycle, whereas the circadian wake propensity rhythm peaks 1–3 h before the habitual bedtime. These rhythms are thus timed to oppose the homeostatic decline of sleep tendency during the habitual sleep episode and the rise of sleep tendency throughout the usual waking day, respectively. Misalignment of the output of the endogenous circadian pacemaker with the desired sleep-wake cycle can, therefore, induce insomnia, decreased alertness, and impaired performance evident in night-shift workers and airline travelers.

Behavioral Correlates of Sleep States and Stages

Polysomnographic staging of sleep correlates with behavioral changes during specific states and stages. During the transitional state between wakefulness and sleep (stage N1 sleep), subjects may respond to faint auditory or visual signals without "awakening." Short-term memory incorporation is inhibited at the onset of NREM stage N1 sleep, which may explain why individuals aroused from that transitional sleep stage frequently deny having been asleep. Such transitions may intrude upon behavioral wakefulness after sleep deprivation, notwithstanding attempts to remain continuously awake (see "Shift-Work Disorder," below).

Awakenings from REM sleep are associated with recall of vivid dream imagery >80% of the time. The reliability of dream recall increases with REM sleep episodes occurring later in the night. Imagery may also be reported after NREM sleep interruptions, though these typically lack the detail and vividness of REM sleep dreams. The incidence of NREM sleep dream recall can be increased by selective REM sleep deprivation, suggesting that REM sleep and dreaming per se are not inexorably linked.

Physiologic Correlates of Sleep States and Stages

All major physiologic systems are influenced by sleep. Changes in cardiovascular function include a decrease in blood pressure and heart rate during NREM and particularly during slow-wave sleep. During REM sleep, phasic activity (bursts of eye movements) is associated with variability in both blood pressure and heart rate mediated principally by the vagus nerve. Cardiac dysrhythmias may occur selectively during REM sleep. Respiratory function also changes. In comparison to relaxed wakefulness, respiratory rate becomes more regular during NREM sleep (especially slow-wave sleep) and tonic REM sleep and becomes very irregular during phasic REM sleep. Minute ventilation decreases in NREM sleep out of proportion to the decrease in metabolic rate at sleep onset, resulting in a higher PCO_2.

Endocrine function also varies with sleep. Slow-wave sleep is associated with secretion of growth hormone in men, while sleep in general is associated with augmented secretion of prolactin in both men and women. Sleep has a complex effect on the secretion of luteinizing hormone (LH): during puberty, sleep is associated with increased LH secretion, whereas sleep in the postpubertal female inhibits LH secretion in the early follicular phase of the menstrual cycle. Sleep onset (and probably slow-wave sleep) is associated with inhibition of thyroid-stimulating hormone and of the adrenocorticotropic hormone–cortisol axis, an effect that is superimposed on the prominent circadian rhythms in the two systems.

The pineal hormonemelatonin is secreted predominantly at night in both day- and night-active species, reflecting the direct modulation of pineal activity by the circadian pacemaker through a circuitous neural pathway from the SCN to the pineal gland. Melatonin secretion is not dependent upon the occurrence of sleep, persisting in individuals kept awake at night. Secretion is inhibited by ambient

light, an effect mediated by a neural connection from the retina via the SCN. The role of endogenous melatonin in normal sleep-wake regulation is unclear, but administration of exogenous melatonin can potentiate sleepiness and facilitate sleep onset when administered in the afternoon or evening, at a time when endogenous melatonin levels are low. The efficacy of melatonin as a sleep-promoting therapy for patients with insomnia is currently not known.

Sleep is also accompanied by alterations of thermoregulatory function. NREM sleep is associated with an attenuation of thermoregulatory responses to either heat or cold stress, and animal studies of thermosensitive neurons in the hypothalamus document an NREM-sleep-dependent reduction of the thermoregulatory set-point. REM sleep is associated with complete absence of thermoregulatory responsiveness, resulting in functional poikilothermy. However, the potential adverse impact of this failure of thermoregulation is blunted by inhibition of REM sleep by extreme ambient temperatures.

DISORDERS OF SLEEP AND WAKEFULNESS

Approach to the Patient: Sleep Disorders

Patients may seek help from a physician because of one of several symptoms: (1) an acute or chronic inability to initiate or maintain sleep adequately at night (insomnia); (2) chronic fatigue, sleepiness, or tiredness during the day; or (3) a behavioral manifestation associated with sleep itself. The specific approach to an insomnia complaint will depend on the nature of comorbid medical or psychiatric disease, if present. In general, however, the insomnia complaint should be specifically addressed as soon as it is recognized. This more aggressive approach reflects growing evidence that chronic insomnia may contribute to comorbid disease processes. For example, specific management of symptomatic insomnia at the time of diagnosis of major depressive disorder (MDD) has been shown to positively impact the response to antidepressants. Evidence that insomnia and sleep loss affect the perception of pain suggests that a similar approach is warranted in acute and chronic pain management. In general, at least for chronic insomnia, there is little evidence justifying an expectant approach in which specific insomnia therapy is deferred while comorbid disease is first addressed.

Table 27-1 outlines the diagnostic and therapeutic approach to the patient with a complaint of excessive daytime sleepiness.

Table 27-1 Evaluation of the Patient with the Complaint of Excessive Daytime Somnolence

Findings on History and Physical Examination	Diagnostic Evaluation	Diagnosis	Therapy
Obesity, snoring, hypertension	Polysomnography with respiratory monitoring	Obstructive sleep apnea	Continuous positive airway pressure; ENT surgery (e.g., uvulopalatopharyngoplasty); dental appliance; pharmacologic therapy (e.g., protriptyline); weight loss
Cataplexy, hypnogogic hallucinations, sleep paralysis, family history	Polysomnography with multiple sleep latency testing	Narcolepsy-cataplexy syndrome	Stimulants (e.g., modafinil, methylphenidate); REM-suppressant antidepressants (e.g., protriptyline); genetic counseling
Restless legs, disturbed sleep, predisposing medical condition (e.g., iron deficiency or renal failure)	Assessment for predisposing medical conditions	Restless legs syndrome	Treatment of predisposing condition, if possible; dopamine agonists (e.g., pramipexole, ropinirole)
Disturbed sleep, predisposing medical conditions (e.g., asthma), and/or predisposing medical therapies (e.g., theophylline)	Sleep-wake diary recording	Insomnias (see text)	Treatment of predisposing condition and/or change in therapy, if possible; behavioral therapy; short-acting benzodiazepine receptor agonist (e.g., zolpidem)

Abbreviations: EMG, electromyogram; ENT, ears, nose, throat; REM, rapid eye movement.

A careful history is essential. In particular, the duration, severity, and consistency of the symptoms are important, along with the patient's estimate of the consequences of the sleep disorder on waking function. Information from a friend or family member can be invaluable; some patients may be unaware of, or will underreport, such potentially embarrassing symptoms as heavy snoring or falling asleep while driving.

Patients with excessive sleepiness should be advised to avoid all driving until effective therapy has been achieved.

Completion by the patient of a day-by-day sleep-work-drug log for at least 2 weeks can help the physician better understand the nature of the complaint. Work times and sleep times (including daytime naps and nocturnal awakenings) as well as drug and alcohol use, including caffeine and hypnotics, should be noted each day.

Polysomnography is necessary for the diagnosis of specific disorders such as narcolepsy and sleep apnea and may be of utility in other settings as well.

Evaluation of Insomnia

Insomnia is the complaint of inadequate sleep; it can be classified according to the nature of sleep disruption and the duration of the complaint. Insomnia is subdivided into difficulty falling asleep (*sleep onset insomnia*), frequent or sustained awakenings (*sleep maintenance insomnia*), or early morning awakenings (*sleep offset insomnia*), though most insomnia patients present with two or more of these symptoms. Other insomnia patients present with persistent sleepiness/fatigue despite sleep of adequate duration (*nonrestorative sleep*). Similarly, the duration of the symptom influences diagnostic and therapeutic considerations. An insomnia complaint lasting one to several nights (within a single episode) is termed *transient insomnia* and is typically the result of situational stress or a change in sleep schedule or environment (e.g., jet lag disorder). *Short-term insomnia* lasts from a few days to 3 weeks. Disruption of this duration is usually associated with more protracted stress, such as recovery from surgery or short-term illness. *Long-term insomnia*, or *chronic insomnia*, lasts for months or years and, in contrast with short-term insomnia, requires a thorough evaluation of underlying causes (see below). Chronic insomnia is often a waxing and waning disorder, with spontaneous or stressor-induced exacerbations.

An occasional night of poor sleep, typically in the setting of stress or excitement about external events, is both common and without lasting consequences. However, persistent insomnia can lead to impaired daytime function, injury due to accidents, and the development of major depression. In addition, there is emerging evidence that individuals with chronic insomnia have increased utilization of health care resources, even after controlling for comorbid medical and psychiatric disorders.

All insomnias can be exacerbated and perpetuated by behaviors that are not conducive to initiating or maintaining sleep. *Inadequate sleep hygiene* is characterized by a behavior pattern prior to sleep or a bedroom environment that is not conducive to sleep, or irregularity in the timing or duration of the nightly sleep episode. Noise, light, or technology (e.g., television, radio, cell phone, mobile email device or computer) in the bedroom can interfere with sleep, as can a bed partner with periodic limb movements during sleep or one who snores loudly. Clocks can heighten the anxiety about the time it has taken to fall asleep. Drugs that act on the central nervous system, large meals, vigorous exercise, or hot showers just before sleep may all interfere with sleep onset. Many individuals participate in stressful work-related activities in the evening, producing a state incompatible with sleep onset. In preference to hypnotic medications, patients should be counseled to avoid stressful activities before bed, develop a soporific bedtime ritual, and prepare and reserve the bedroom environment for sleeping. Consistent, regular bedtimes and rising times should be maintained daily, including weekends.

Primary Insomnia

Many patients with chronic insomnia have no clear, single identifiable underlying cause for their difficulties with sleep. Rather, such patients often have multiple etiologies for their insomnia, which may evolve over the years. In addition, the chief sleep complaint may change over time, with initial insomnia predominating at one point, and multiple awakenings or nonrestorative sleep occurring at other times. Subsyndromal psychiatric disorders (e.g., anxiety and mood complaints), negative conditioning to the sleep environment (psychophysiologic insomnia, see below), amplification of the time spent awake (paradoxical insomnia), physiologic hyperarousal, and poor sleep hygiene (see above) may all be present. As these processes may be both causes and consequences of chronic insomnia, many individuals will have a progressive course to their symptoms in which the severity is proportional to the chronicity, and much of the complaint may persist even after effective treatment of the initial inciting etiology. Treatment of insomnia is often directed to each of the putative contributing factors: behavior therapies for anxiety and negative conditioning (see below), pharmacotherapy and/or psychotherapy for mood/anxiety disorders, and an emphasis on maintenance of good sleep hygiene.

If insomnia persists after treatment of these contributing factors, empirical pharmacotherapy is often used on a nightly or intermittent basis. A variety of sedative compounds are used for this purpose. Alcohol and antihistamines are the most commonly used nonprescription sleep aids. The former may help with sleep onset but is associated with sleep disruption during the night and can escalate into abuse, dependence, and withdrawal in the predisposed individual. Antihistamines, which are the primary active ingredient in most over-the-counter sleep aids, may be of benefit when used intermittently but often produce rapid tolerance and may have multiple side effects (especially anticholinergic), which limit their use, particularly in the elderly. Benzodiazepine-receptor agonists are the most effective and well-tolerated class of medications for insomnia. The broad range of half-lives allows flexibility in the duration of sedative action. The most commonly prescribed agents in this family are zaleplon (5–20 mg), with a half-life of 1–2 h; zolpidem (5–10 mg) and triazolam (0.125–0.25 mg), with half-lives of 2–3 h; eszopiclone (1–3 mg), with a half-life of 5.5–8 h; and temazepam (15–30 mg) and lorazepam (0.5–2 mg), with half-lives of 6–12 h. Generally, side effects are minimal when the dose is kept low and the serum concentration is minimized during the waking hours (by using the shortest-acting effective agent). At least one benzodiazepine receptor agonist (eszopiclone) continues to be

effective for 6 months of nightly use. However, longer durations of use have not been evaluated, and it is unclear whether this is true of other agents in this class. Moreover, with even brief continuous use of benzodiazepine-receptor agonists, rebound insomnia can occur upon discontinuation. The likelihood of rebound insomnia and tolerance can be minimized by short durations of treatment, intermittent use, or gradual tapering of the dose. For acute insomnia, nightly use of a benzodiazepine receptor agonist for a maximum of 2–4 weeks is advisable. For chronic insomnia, intermittent use is recommended, unless the consequences of untreated insomnia outweigh concerns regarding chronic use. Benzodiazepine receptor agonists should be avoided, or used very judiciously, in patients with a history of substance or alcohol abuse. The heterocyclic antidepressants (trazodone, amitriptyline, and doxepin) are the most commonly prescribed alternatives to benzodiazepine-receptor agonists due to their lack of abuse potential and lower cost. Trazodone (25–100 mg) is used more commonly than the tricyclic antidepressants as it has a much shorter half-life (5–9 h), has much less anticholinergic activity (sparing patients, particularly the elderly, constipation, urinary retention, and tachycardia), is associated with less weight gain, and is much safer in overdose. The risk of priapism is small (~1 in 10,000).

PSYCHOPHYSIOLOGIC INSOMNIA

Persistent *psychophysiologic insomnia* is a behavioral disorder in which patients are preoccupied with a perceived inability to sleep adequately at night. This sleep disorder begins like any other acute insomnia; however, the poor sleep habits and sleep-related anxiety ("insomnia phobia") persist long after the initial incident. Such patients become hyperaroused by their own efforts to sleep or by the sleep environment, and the insomnia becomes a conditioned or learned response. Patients may be able to fall asleep more easily at unscheduled times (when not trying) or outside the home environment. Polysomnographic recording in patients with psychophysiologic insomnia reveals an objective sleep disturbance, often with an abnormally long sleep latency; frequent nocturnal awakenings; and an increased amount of stage N1 transitional sleep. Rigorous attention should be paid to improving sleep hygiene, correction of counterproductive, arousing behaviors before bedtime, and minimizing exaggerated beliefs regarding the negative consequences of insomnia. Behavioral therapies are the treatment modality of choice, with intermittent use of medications. When patients are awake for >20 min, they should read or perform other relaxing activities to distract themselves from insomnia-related anxiety. In addition, bedtime and wake time should be scheduled to restrict time in bed to be equal to their perceived total sleep time. This will generally produce sleep deprivation, greater sleep drive, and, eventually, better sleep. Time in bed can then be gradually expanded. In addition, methods directed toward producing relaxation in the sleep setting (e.g., meditation, muscle relaxation) are encouraged.

ADJUSTMENT INSOMNIA (ACUTE INSOMNIA)

This typically develops after a change in the sleeping environment (e.g., in an unfamiliar hotel or hospital bed) or before or after a significant life event, such as a change of occupation, loss of a loved one, illness, or anxiety over a deadline or examination. Increased sleep latency, frequent awakenings from sleep, and early morning awakening can all occur. Recovery is generally rapid, usually within a few weeks. Treatment is symptomatic, with intermittent use of hypnotics and resolution of the underlying stress. *Altitude insomnia* describes a sleep disturbance that is a common consequence of exposure to high altitude. Periodic breathing of the Cheyne-Stokes type occurs during NREM sleep about half the time at high altitude, with restoration of a regular breathing pattern during REM sleep. Both hypoxia and hypocapnia are thought to be involved in the development of periodic breathing. Frequent awakenings and poor quality sleep characterize altitude insomnia, which is generally worse on the first few nights at high altitude but may persist. Treatment with acetazolamide can decrease time spent in periodic breathing and substantially reduce hypoxia during sleep.

Comorbid Insomnia

INSOMNIA ASSOCIATED WITH MENTAL DISORDERS

Approximately 80% of patients with psychiatric disorders describe sleep complaints. There is considerable heterogeneity, however, in the nature of the sleep disturbance both between conditions and among patients with the same condition. *Depression* can be associated with sleep onset insomnia, sleep maintenance insomnia, or early morning wakefulness. However, hypersomnia occurs in some depressed patients, especially adolescents and those with either bipolar or seasonal (fall/winter) depression (Chap. 391). Indeed, sleep disturbance is an important vegetative sign of depression and may commence before any mood changes are perceived by the patient. Consistent polysomnographic findings in depression include decreased REM sleep latency, lengthened first REM sleep episode, and shortened first NREM sleep episode; however, these findings are not specific for depression, and the extent of these changes varies with age and symptomatology. Depressed patients also show decreased slow-wave sleep and reduced sleep continuity.

In *mania* and *hypomania*, sleep latency is increased and total sleep time can be reduced. Patients with *anxiety disorders* tend not to show the changes in REM sleep and slow-wave sleep seen in endogenously depressed patients. *Chronic alcoholics* lack slow-wave sleep, have decreased amounts

of REM sleep (as an acute response to alcohol), and have frequent arousals throughout the night. This is associated with impaired daytime alertness. The sleep of chronic alcoholics may remain disturbed for years after discontinuance of alcohol usage. Sleep architecture and physiology are disturbed in *schizophrenia*, with a decreased amount of slow-wave sleep (NREM stage N3 sleep) and a lack of augmentation of REM sleep following REM sleep deprivation; chronic schizophrenics often show day-night reversal, sleep fragmentation, and insomnia.

INSOMNIA ASSOCIATED WITH NEUROLOGIC DISORDERS

A variety of neurologic diseases result in sleep disruption through both indirect, nonspecific mechanisms (e.g., pain in cervical spondylosis or low back pain) or by impairment of central neural structures involved in the generation and control of sleep itself. For example, *dementia* from any cause has long been associated with disturbances in the timing of the sleep-wake cycle, often characterized by nocturnal wandering and an exacerbation of symptomatology at night (so-called sundowning).

Epilepsy may rarely present as a sleep complaint (Chap. 369). Often the history is of abnormal behavior, at times with convulsive movements during sleep. The differential diagnosis includes REM sleep behavior disorder, sleep apnea syndrome, and periodic movements of sleep (see above). Diagnosis requires nocturnal polysomnography with a full EEG montage. Other neurologic diseases associated with abnormal movements, such as *Parkinson's disease*, *hemiballismus*, *Huntington's chorea*, and *Tourette's syndrome* (Chap. 372), are also associated with disrupted sleep, presumably through secondary mechanisms. However, the abnormal movements themselves are greatly reduced during sleep. Headache syndromes (*migraine* or *cluster headache*) may show sleep-associated exacerbations (Chap. 14) by unknown mechanisms.

Fatal familial insomnia is a rare hereditary disorder caused by degeneration of anterior and dorsomedial nuclei of the thalamus. Insomnia is a prominent early symptom. Patients develop progressive autonomic dysfunction, followed by dysarthria, myoclonus, coma, and death. The pathogenesis is a mutation in the prion gene (Chap. 383).

INSOMNIA ASSOCIATED WITH OTHER MEDICAL DISORDERS

A number of medical conditions are associated with disruptions of sleep. The association is frequently nonspecific, e.g., sleep disruption due to chronic pain from rheumatologic disorders. Attention to this association is important in that sleep-associated symptoms are often the presenting or most bothersome complaint. Treatment of the underlying medical problem is the most useful approach. Sleep disruption can also result from the use of medications such as glucocorticoids (see below).

One prominent association is between sleep disruption and *asthma*. In many asthmatics there is a prominent daily variation in airway resistance that results in marked increases in asthmatic symptoms at night, especially during sleep. In addition, treatment of asthma with theophylline-based compounds, adrenergic agonists, or glucocorticoids can independently disrupt sleep. When sleep disruption is a side effect of asthma treatment, inhaled glucocorticoids (e.g., beclomethasone) that do not disrupt sleep may provide a useful alternative.

Cardiac ischemia may also be associated with sleep disruption. The ischemia itself may result from increases in sympathetic tone as a result of sleep apnea. Patients may present with complaints of nightmares or vivid, disturbing dreams, with or without awareness of the more classic symptoms of angina or of the sleep-disordered breathing. Treatment of the sleep apnea may substantially improve the angina and the nocturnal sleep quality. *Paroxysmal nocturnal dyspnea* can also occur as a consequence of sleep-associated cardiac ischemia that causes pulmonary congestion exacerbated by the recumbent posture.

Chronic obstructive pulmonary disease is also associated with sleep disruption, as is *cystic fibrosis*, *menopause*, *hyperthyroidism*, *gastroesophageal reflux*, *chronic renal failure*, and *liver failure*.

Medication-, Drug-, or Alcohol-Dependent Insomnia

Disturbed sleep can result from ingestion of a wide variety of agents. Caffeine is perhaps the most common pharmacologic cause of insomnia. It produces increased latency to sleep onset, more frequent arousals during sleep, and a reduction in total sleep time for up to 8–14 h after ingestion. Even small amounts of coffee can significantly disturb sleep in some patients; therefore, a 1- to 2-month trial without caffeine should be attempted in patients with these symptoms. Similarly, alcohol and nicotine can interfere with sleep, despite the fact that many patients use them to relax and promote sleep. Although alcohol can increase drowsiness and shorten sleep latency, even moderate amounts of alcohol increase awakenings in the second half of the night. In addition, alcohol ingestion prior to sleep is contraindicated in patients with sleep apnea because of the inhibitory effects of alcohol on upper airway muscle tone. Acutely, amphetamines and cocaine suppress both REM sleep and total sleep time, which return to normal with chronic use. Withdrawal leads to an REM sleep rebound. A number of prescribed medications can produce insomnia. Antidepressants, sympathomimetics, and glucocorticoids are common causes. In addition, severe rebound insomnia can result from the acute withdrawal of hypnotics, especially following the use of high doses of benzodiazepines with a short half-life. For this reason, hypnotic doses should be low to moderate, and prolonged drug tapering is

encouraged.

Restless Legs Syndrome (RLS)

Patients with this sensorimotor disorder report an irresistible urge to move the legs, or sometimes the upper extremities, that is often associated with creepy-crawling or aching dysesthesias deep within the affected limbs. For most patients with RLS, the dysesthesias and restlessness are much worse in the evening or night compared to the daytime and frequently interfere with the ability to fall asleep. The symptoms appear with inactivity and are temporarily relieved by movement. In contrast, paresthesias secondary to peripheral neuropathy persist with activity. The severity of this chronic disorder may wax and wane over time and can be exacerbated by sleep deprivation, caffeine, alcohol, serotonergic antidepressants, and pregnancy. The prevalence is 1–5% of young to middle-age adults and 10–20% of those aged >60 years. There appear to be important differences in RLS prevalence among racial groups, with higher prevalence in those of Northern European ancestry. Roughly one-third of patients (particularly those with an early age of onset) will have multiple affected family members. At least three separate chromosomal loci have been identified in familial RLS, though no gene has been identified to date. Iron deficiency and renal failure may cause RLS, which is then considered secondary RLS. The symptoms of RLS are exquisitely sensitive to dopaminergic drugs (e.g., pramipexole 0.25–0.5 mg q8PM or ropinirole 0.5–4 mg q8PM), which are the treatments of choice. Opioids, benzodiazepines, and gabapentin may also be of therapeutic value. Most patients with restless legs also experience periodic limb movements of sleep, although the reverse is not the case.

Periodic Limb Movement Disorder (PLMD)

Periodic limb movements of sleep (PLMS), previously known as *nocturnal myoclonus*, consists of stereotyped, 0.5- to 5.0-s extensions of the great toe and dorsiflexion of the foot, which recur every 20–40 s during NREM sleep, in episodes lasting from minutes to hours, as documented by bilateral surface EMG recordings of the anterior tibialis on polysomnography. PLMS is the principal objective polysomnographic finding in 17% of patients with insomnia and 11% of those with excessive daytime somnolence **(Fig. 27-3)**. It is often unclear whether it is an incidental finding or the cause of disturbed sleep. When deemed to be the latter, PLMS is called PLMD. PLMS occurs in a wide variety of sleep disorders (including narcolepsy, sleep apnea, REM sleep behavior disorder, and various forms of insomnia) and may be associated with frequent arousals and an increased number of sleep-stage transitions. The pathophysiology is not well understood, though individuals with high spinal transections can exhibit periodic leg movements during sleep, suggesting the existence of a spinal generator. Treatment options include dopaminergic medications or benzodiazepines.

Figure 27-3

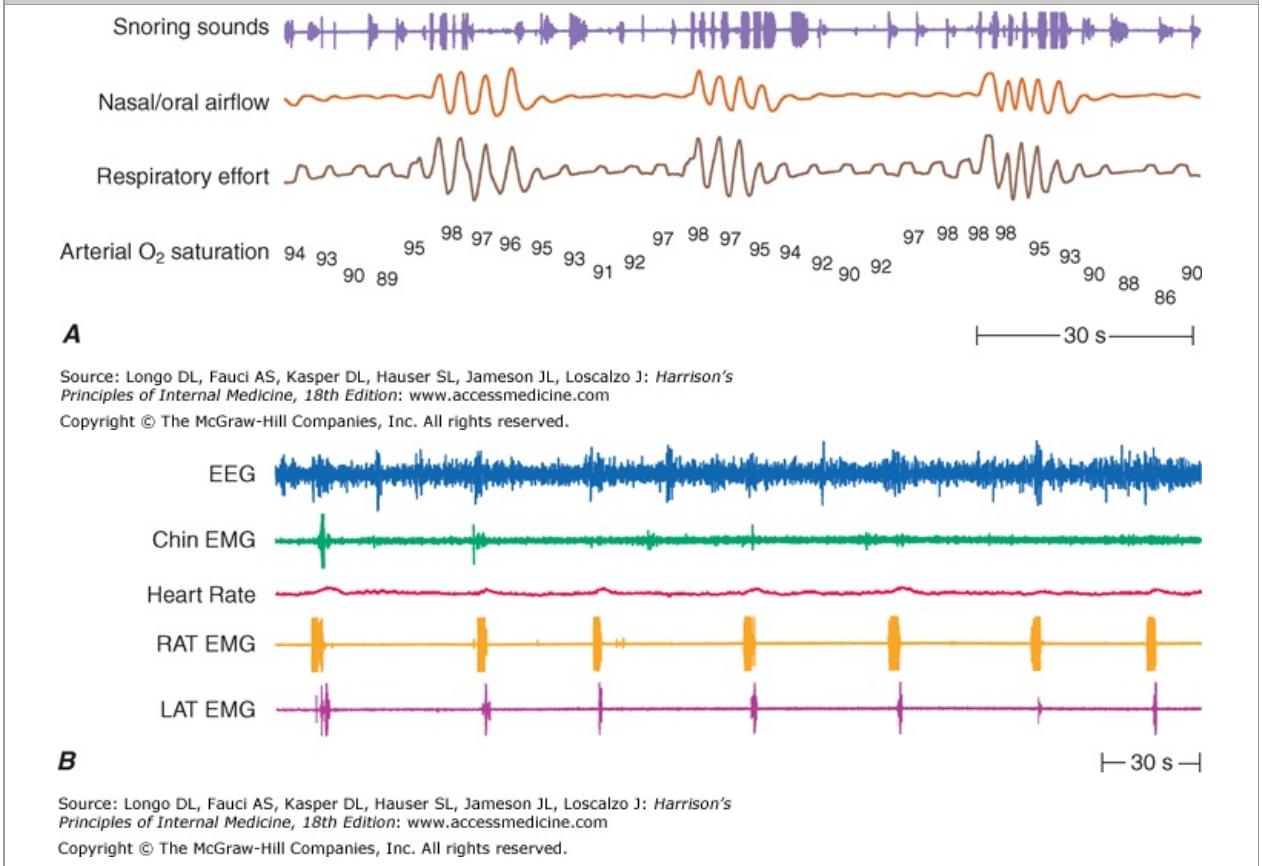

Polysomnographic recordings of (*A*) obstructive sleep apnea and (*B*) periodic limb

movement of sleep. Note the snoring and reduction in air flow in the presence of continued respiratory effort, associated with the subsequent oxygen desaturation (upper panel). Periodic limb movements occur with a relatively constant intermovement interval and are associated with changes in the EEG and heart rate acceleration (lower panel). RAT, right anterior tibialis; LAT, left anterior tibialis. (From the Division of Sleep Medicine, Brigham and Women's Hospital.)

Evaluation of Daytime Sleepiness

Daytime impairment due to sleep loss may be difficult to quantify for several reasons. First, patients may be unaware of the extent of sleep deprivation. In obstructive sleep apnea, for example, the brief arousals from sleep associated with respiratory recovery after each apneic episode results in daytime sleepiness, despite the fact that the patient may be unaware of the sleep fragmentation. Second, subjective descriptions of waking impairment vary from patient to patient. Patients may describe themselves as "sleepy," "fatigued," or "tired" and may have a clear sense of the meaning of those terms, while others may use the same terms to describe a completely different condition. Third, sleepiness, particularly when profound, may affect judgment in a manner analogous to ethanol, such that subjective awareness of the condition and the consequent cognitive and motor impairment is reduced. Finally, patients may be reluctant to admit that sleepiness is a problem, both because they are generally unaware of what constitutes normal alertness and because sleepiness is generally viewed pejoratively, ascribed more often to a deficit in motivation than to an inadequately addressed physiologic sleep need.

Specific questioning about the occurrence of sleep episodes during normal waking hours, both intentional and unintentional, is necessary to determine the extent of the adverse effects of sleepiness on a patient's daytime function. Specific areas to be addressed include the occurrence of inadvertent sleep episodes while driving or in other safety-related settings, sleepiness while at work or school (and the relationship of sleepiness to work and school performance), and the effect of sleepiness on social and family life. Standardized questionnaires, e.g., the Epworth Sleepiness Scale, are now commonly used in clinical and research settings to quantify daytime sleep tendency and screen for excessive sleepiness.

Driving is particularly hazardous for patients with increased sleepiness. Reaction time is equally impaired by 24 h of sleep loss as by a blood alcohol level of 0.10 g/dL. More than half of Americans admit to having fallen asleep while driving. An estimated 250,000 motor vehicle crashes per year are due to drowsy drivers, causing about 20% of all serious crash injuries and deaths. Drowsy driving legislation, aimed at improving education of all drivers about the hazards of driving drowsy and establishing sanctions comparable to those for drunk driving, has been enacted in New Jersey and is pending in several other states. Screening for sleep disorders, provision of an adequate number of safe highway rest areas, maintenance of unobstructed shoulder rumble strips, and strict enforcement and compliance monitoring of hours-of-service policies are needed to reduce the risk of sleep-related transportation crashes. Evidence for significant daytime impairment in association either with the diagnosis of a primary sleep disorder, such as narcolepsy or sleep apnea, or with imposed or self-selected sleep-wake schedules (see "Shift-Work Disorder," below) raises the issue of the physician's responsibility to notify motor vehicle licensing authorities of the increased risk of sleepiness-related motor vehicle crashes. As with epilepsy, legal requirements vary from state to state, and existing legal precedents do not provide a consistent interpretation of the balance between the physician's responsibility and the patient's right to privacy. At a minimum, physicians should inform patients who report a history of nodding off or falling asleep at the wheel or who have excessive daytime sleepiness about the increased risk of operating a motor vehicle, advise such patients not to drive a motor vehicle until the cause of the excessive sleepiness has been diagnosed and successful treatment has been implemented, and reevaluate the patient to determine when it is safe for the patient to resume driving. Each of those steps should be documented in the patient's medical record.

The distinction between fatigue and sleepiness can be useful in the differentiation of patients with complaints of fatigue or tiredness in the setting of disorders such as fibromyalgia (Chap. 335), chronic fatigue syndrome (Chap. 389), or endocrine deficiencies such as hypothyroidism (Chap. 341) or Addison's disease (Chap. 342). While patients with these disorders can typically distinguish their daytime symptoms from the sleepiness that occurs with sleep deprivation, substantial overlap can occur. This is particularly true when the primary disorder also results in chronic sleep disruption (e.g., sleep apnea in hypothyroidism) or in abnormal sleep (e.g., fibromyalgia).

While clinical evaluation of the complaint of excessive sleepiness is usually adequate, objective quantification is sometimes necessary. Assessment of daytime functioning as an index of the adequacy of sleep can be made with the multiple sleep latency test (MSLT), which utilizes repeated measurement of sleep latency (time to onset of sleep) under standardized conditions during a day following quantified nocturnal sleep. The average latency across four to six tests (administered every 2 h across the waking day) provides an objective measure of daytime sleep tendency. Disorders of sleep that result in pathologic daytime somnolence can be reliably distinguished with the MSLT. In addition, the multiple measurements of sleep onset may identify direct transitions from wakefulness to REM sleep that are suggestive of specific pathologic conditions (e.g., narcolepsy).

Narcolepsy

Narcolepsy is both a disorder of the ability to sustain wakefulness voluntarily and a disorder of REM sleep regulation **(Table 27-2)**. The classic "narcolepsy tetrad" consists of excessive daytime somnolence plus three specific symptoms related to an intrusion of REM sleep characteristics (e.g., muscle atonia, vivid dream imagery) into the transition between wakefulness and sleep: (1) sudden weakness or loss of muscle tone without loss of consciousness, often elicited by emotion (cataplexy); (2) hallucinations at sleep onset (hypnagogic hallucinations) or upon awakening (hypnopompic hallucinations); and (3) muscle paralysis upon awakening (sleep paralysis). The severity of cataplexy varies, as patients may have two to three attacks per day or per decade. Some patients with objectively confirmed narcolepsy (see below) may show no evidence of cataplexy. In those with cataplexy, the extent and duration of an attack may also vary, from a transient sagging of the jaw lasting a few seconds to rare cases of flaccid paralysis of the entire voluntary musculature for up to 20 –30 min. Symptoms of narcolepsy typically begin in the second decade, although the onset ranges from ages 5–50 years. Once established, the disease is chronic without remissions. Secondary forms of narcolepsy have been described (e.g., after head trauma).

Table 27-2 Prevalence of Symptoms in Narcolepsy

Symptom	Prevalence, %
Excessive daytime somnolence	100
Disturbed sleep	87
Cataplexy	76
Hypnagogic hallucinations	68
Sleep paralysis	64
Memory problems	50

Source: Modified from TA Roth, L Merlotti, in SA Burton et al (eds): *Narcolepsy 3rd International Symposium: Selected Symposium Proceedings.* Chicago, Matrix Communications, 1989.

Narcolepsy affects about 1 in 4000 people in the United States and appears to have a genetic basis. Recently, several convergent lines of evidence suggest that the hypothalamic neuropeptide hypocretin (orexin) is involved in the pathogenesis of narcolepsy: (1) a mutation in the hypocretin receptor 2 gene has been associated with canine narcolepsy; (2) hypocretin "knockout" mice that are genetically unable to produce this neuropeptide exhibit behavioral and electrophysiologic features resembling human narcolepsy; and (3) cerebrospinal fluid levels of hypocretin are reduced in most patients who have narcolepsy with cataplexy. The inheritance pattern of narcolepsy in humans is more complex than in the canine model. However, almost all narcoleptics with cataplexy are positive for HLA DQB1*0602 (Chap. 315), suggesting that an autoimmune process may be responsible.

DIAGNOSIS

The diagnostic criteria continue to be a matter of debate. Certainly, objective verification of excessive daytime somnolence, typically with MSLT mean sleep latencies <8 min, is an essential if nonspecific diagnostic feature. Other conditions that cause excessive sleepiness, such as sleep apnea or chronic sleep deprivation, must be rigorously excluded. The other objective diagnostic feature of narcolepsy is the presence of REM sleep in at least two of the naps during the MSLT. Abnormal regulation of REM sleep is also manifested by the appearance of REM sleep immediately or within minutes after sleep onset in 50% of narcoleptic patients, a rarity in unaffected individuals maintaining a conventional sleep-wake schedule. The REM-related symptoms of the classic narcolepsy tetrad are variably present. There is increasing evidence that narcoleptics with cataplexy (one-half to two-thirds of patients) may represent a more homogeneous group than those without this symptom. However, a history of cataplexy can be difficult to establish reliably. Hypnagogic and hypnopompic hallucinations and sleep paralysis are often found in nonnarcoleptic individuals and may be present in only one-half of narcoleptics. Nocturnal sleep disruption is commonly observed in narcolepsy but is also a nonspecific symptom. Similarly, a history of "automatic behavior" during wakefulness (a trancelike state during which simple motor behaviors persist) is not specific for narcolepsy and serves principally to corroborate the presence of daytime somnolence.

Treatment: Narcolepsy

The treatment of narcolepsy is symptomatic. Somnolence is treated with wake-promoting therapeutics. Modafinil is now the drug of choice, principally because it is associated with fewer side effects than older stimulants and has a long half-life; 200–400 mg is given as a single daily dose. Older drugs such as methylphenidate (10 mg bid to 20 mg qid) or dextroamphetamine (10 mg bid) are still used as alternatives, particularly in refractory patients.

These latter medications are now available in slow-release formulations, extending their duration of action and allowing once-daily dosing.

Treatment of the REM-related phenomena of cataplexy, hypnagogic hallucinations, and sleep paralysis requires the potent REM sleep suppression produced by antidepressant medications. The tricyclic antidepressants [e.g., protriptyline (10–40 mg/d) and clomipramine (25–50 mg/d)] and the selective serotonin reuptake inhibitors (SSRIs) [e.g., fluoxetine (10–20 mg/d)] are commonly used for this purpose. Efficacy of the antidepressants is limited largely by anticholinergic side effects (tricyclics) and by sleep disturbance and sexual dysfunction (SSRIs). Alternately, gamma hydroxybutyrate (GHB), given at bedtime, and 4 h later, is effective in reducing daytime cataplectic episodes. Adequate nocturnal sleep time and planned daytime naps (when possible) are important preventive measures.

Sleep Apnea Syndromes

Respiratory dysfunction during sleep is a common, serious cause of excessive daytime somnolence as well as of disturbed nocturnal sleep. An estimated 2–5 million individuals in the United States have a reduction or cessation of breathing for 10–150 s from thirty to several hundred times every night during sleep. These episodes may be due to either an occlusion of the airway (*obstructive sleep apnea*), absence of respiratory effort (*central sleep apnea*), or a combination of these factors (*mixed sleep apnea*) (Fig. 27-3). Failure to recognize and treat these conditions appropriately may lead to impairment of daytime alertness, increased risk of sleep-related motor vehicle accidents, hypertension and other serious cardiovascular complications, and increased mortality. Sleep apnea is particularly prevalent in overweight men and in the elderly, yet it is estimated to remain undiagnosed in 80–90% of affected individuals. This is unfortunate since effective treatments are available. **Readers are referred to Chap. 265 for a comprehensive review of the diagnosis and treatment of patients with these conditions.**

Parasomnias

The term *parasomnia* refers to abnormal behaviors or experiences that arise from or occur during sleep. A continuum of parasomnias arise from NREM sleep, from brief confusional arousals to sleepwalking and night terrors. The presenting complaint is usually related to the behavior itself, but the parasomnias can disturb sleep continuity or lead to mild impairments in daytime alertness. Two main parasomnias occur in REM sleep: REM sleep behavior disorder (RBD), which will be described below, and nightmare disorder.

SLEEPWALKING (SOMNAMBULISM)

Patients affected by this disorder carry out automatic motor activities that range from simple to complex. Individuals may walk, urinate inappropriately, eat, or exit from the house while remaining only partially aware. Full arousal may be difficult, and individuals may rarely respond to attempted awakening with agitation or even violence. Sleepwalking arises from slow-wave sleep (NREM stage N3 sleep), usually in the first 2 h of the night, and is most common in children and adolescents, when these sleep stages are most robust. Episodes are usually isolated but may be recurrent in 1–6% of patients. The cause is unknown, though it has a familial basis in roughly one-third of cases.

SLEEP TERRORS

This disorder, also called *pavor nocturnus*, occurs primarily in young children during the first several hours after sleep onset, in slow-wave sleep (NREM stage N3 sleep). The child suddenly screams, exhibiting autonomic arousal with sweating, tachycardia, and hyperventilation. The individual may be difficult to arouse and rarely recalls the episode on awakening in the morning. Parents are usually reassured to learn that the condition is self-limited and benign and that no specific therapy is indicated. Both sleep terrors and sleepwalking represent abnormalities of arousal. In contrast, *nightmares* occur during REM sleep and cause full arousal, with intact memory for the unpleasant episode.

SLEEP BRUXISM

Bruxism is an involuntary, forceful grinding of teeth during sleep that affects 10–20% of the population. The patient is usually unaware of the problem. The typical age of onset is 17–20 years, and spontaneous remission usually occurs by age 40. Sex distribution appears to be equal. In many cases, the diagnosis is made during dental examination, damage is minor, and no treatment is indicated. In more severe cases, treatment with a rubber tooth guard is necessary to prevent disfiguring tooth injury. Stress management or, in some cases, biofeedback can be useful when bruxism is a manifestation of psychological stress. There are anecdotal reports of benefit using benzodiazepines.

SLEEP ENURESIS

Bedwetting, like sleepwalking and night terrors, is another parasomnia that occurs during sleep in the young. Before age 5 or 6 years, nocturnal enuresis should probably be considered a normal feature of development. The condition usually improves spontaneously by puberty, has a prevalence in late adolescence of 1–3%, and is rare in adulthood. In older patients with enuresis, a distinction must be made between primary and secondary enuresis, the latter being defined as bedwetting in patients who have previously been fully continent for 6–12 months. Treatment of primary enuresis is reserved for patients of appropriate age (>5 or 6 years) and consists of bladder training exercises and behavioral therapy. Urologic abnormalities are more common in primary enuresis and must be assessed by urologic examination. Important causes of secondary enuresis include emotional disturbances, urinary

tract infections or malformations, cauda equina lesions, epilepsy, sleep apnea, and certain medications. Symptomatic pharmacotherapy is usually accomplished with desmopressin (0.2 mg qhs), oxybutynin chloride (5–10 mg qhs), or imipramine (10–50 mg qhs).

MISCELLANEOUS PARASOMNIAS

Other clinical entities may be characterized as a parasomnia or a sleep-related movement disorder in that they occur selectively during sleep and are associated with some degree of sleep disruption. Examples include *jactatio capitis nocturna* (nocturnal headbanging, rhythmic movement disorder), confusional arousals, sleep-related eating disorder, and nocturnal leg cramps.

REM SLEEP BEHAVIOR DISORDER (RBD)

RBD is a rare condition that is distinct from other parasomnias in that it occurs during REM sleep. It primarily afflicts men of middle age or older, many of whom have an existing, or developing, neurologic disease. Approximately one-half of patients with RBD will develop Parkinson's disease (Chap. 372) within 10–20 years. Presenting symptoms consist of agitated or violent behavior during sleep, as reported by a bed partner. In contrast to typical somnambulism, injury to the patient or bed partner is not uncommon, and, upon awakening, the patient reports vivid, often unpleasant, dream imagery. The principal differential diagnosis is nocturnal seizures, which can be excluded with polysomnography. In RBD, seizure activity is absent on the EEG, and disinhibition of the usual motor atonia is observed in the EMG during REM sleep, at times associated with complex motor behaviors. The pathogenesis is unclear, but damage to brainstem areas mediating descending motor inhibition during REM sleep may be responsible. In support of this hypothesis are the remarkable similarities between RBD and the sleep of animals with bilateral lesions of the pontine tegmentum in areas controlling REM sleep motor inhibition. Treatment with clonazepam (0.5–1.0 mg qhs) provides sustained improvement in almost all reported cases.

CIRCADIAN RHYTHM SLEEP DISORDERS

A subset of patients presenting with either insomnia or hypersomnia may have a disorder of sleep *timing* rather than sleep *generation*. Disorders of sleep timing can be either organic [i.e., due to an abnormality of circadian pacemaker(s) or its input from entraining stimuli] or environmental (i.e., due to a disruption of exposure to entraining stimuli from the environment). Regardless of etiology, the symptoms reflect the influence of the underlying circadian pacemaker on sleep-wake function. Thus, effective therapeutic approaches should aim to entrain the oscillator at an appropriate phase.

JET LAG DISORDER

More than 60 million persons experience transmeridian air travel annually, which is often associated with excessive daytime sleepiness, sleep onset insomnia, and frequent arousals from sleep, particularly in the latter half of the night. Gastrointestinal discomfort is common. The syndrome is transient, typically lasting 2–14 d depending on the number of time zones crossed, the direction of travel, and the traveler's age and phase-shifting capacity. Travelers who spend more time outdoors reportedly adapt more quickly than those who remain in hotel rooms, presumably due to brighter (outdoor) light exposure. Avoidance of antecedent sleep loss and obtaining nap sleep on the afternoon prior to overnight travel greatly reduce the difficulty of extended wakefulness. Laboratory studies suggest that sub milligram doses of the pineal hormonemelatonin can enhance sleep efficiency, but only if taken when endogenous melatonin concentrations are low (i.e., during biologic daytime), and that melatonin may induce phase shifts in human rhythms. A large-scale clinical trial evaluating the safety and efficacy of melatonin as a treatment for jet lag disorder and other circadian sleep disorders is needed.

In addition to jet lag associated with travel across time zones, many patients report a behavioral pattern that has been termed *social jet lag*, in which their bedtimes and wake times on weekends or days off occur 4–8 h later than they do during the week. This recurrent displacement of the timing of the sleep-wake cycle is common in adolescents and young adults and is associated with sleep onset insomnia, poorer academic performance, increased risk of depressive symptoms, and excessive daytime sleepiness.

SHIFT-WORK DISORDER

More than 7 million workers in the United States regularly work at night, either on a permanent or rotating schedule. Many more begin work between 4 A.M. and 7 A.M., requiring them to commute and then work during the time of day that they would otherwise be asleep. In addition, each week millions more elect to remain awake at night to meet deadlines, drive long distances, or participate in recreational activities. This results in both sleep loss and misalignment of the circadian rhythm with respect to the sleep-wake cycle.

Studies of regular night-shift workers indicate that the circadian timing system usually fails to adapt successfully to such inverted schedules. This leads to a misalignment between the desired work-rest schedule and the output of the pacemaker and in disturbed daytime sleep in most individuals. Sleep deprivation, increased length of time awake prior to work, and misalignment of circadian phase produce decreased alertness and performance, increased reaction time, and increased risk of

performance lapses, thereby resulting in greater safety hazards among night workers and other sleep-deprived individuals. Sleep disturbance nearly doubles the risk of a fatal work accident. Additional problems include higher rates of breast, colorectal, and prostate cancer and of cardiac, gastrointestinal, and reproductive disorders in long-term night-shift workers. Recently, the World Health Organization has added night-shift work to its list of probable carcinogens.

Sleep onset is associated with marked attenuation in perception of both auditory and visual stimuli and lapses of consciousness. The sleepy individual may thus attempt to perform routine and familiar motor tasks during the transition state between wakefulness and sleep (stage N1 sleep) in the absence of adequate processing of sensory input from the environment. Motor vehicle operators are especially vulnerable to sleep-related accidents since the sleep-deprived driver or operator often fails to heed the warning signs of fatigue. Such attempts to override the powerful biologic drive for sleep by the sheer force of will can yield a catastrophic outcome when sleep processes intrude involuntarily upon the waking brain. Such sleep-related attentional failures typically last only seconds but are known on occasion to persist for longer durations. These frequent brief intrusions of stage N1 sleep into behavioral wakefulness are a major component of the impaired psychomotor performance seen with sleepiness. There is a significant increase in the risk of sleep-related, fatal-to-the-driver highway crashes in the early morning and late afternoon hours, coincident with bimodal peaks in the daily rhythm of sleep tendency.

Resident physicians constitute another group of workers at risk for accidents and other adverse consequences of lack of sleep and misalignment of the circadian rhythm. Recurrent scheduling of resident physicians to work shifts of 24 h or more consecutive hours impairs psychomotor performance to a degree that is comparable to alcohol intoxication, doubles the risk of attentional failures among intensive care unit interns working at night, and significantly increases the risk of serious medical errors in intensive care units, including a fivefold increase in the risk of serious diagnostic mistakes. Some 20% of hospital interns report making a fatigue-related mistake that injured a patient, and 5% admit making a fatigue-related mistake that results in the death of a patient. Moreover, working for >24 h consecutively increases the risk of percutaneous injuries and more than doubles the risk of motor vehicle crashes on the commute home. For these reasons, in 2008 the Institute of Medicine concluded that the practice of scheduling resident physicians to work for more than 16 consecutive hours without sleep is hazardous for both resident physicians and their patients.

From 5 to 10% of individuals scheduled to work at night or in the early morning hours have much greater than average difficulties remaining awake during night work and sleeping during the day; these individuals are diagnosed with chronic and severe shift-work disorder (SWD). Patients with this disorder have a level of excessive sleepiness during night work and insomnia during day sleep that the physician judges to be clinically significant; the condition is associated with an increased risk of sleep-related accidents and with some of the illnesses associated with night-shift work. Patients with chronic and severe SWD are profoundly sleepy at night. In fact, their sleep latencies during night work average just 2 min, comparable to mean sleep latency durations of patients with narcolepsy or severe daytime sleep apnea.

Treatment: Shift-Work Disorder

Caffeine is frequently used to promote wakefulness. However, it cannot forestall sleep indefinitely, and it does not shield users from sleep-related performance lapses. Postural changes, exercise, and strategic placement of nap opportunities can sometimes temporarily reduce the risk of fatigue-related performance lapses. Properly timed exposure to bright light can facilitate rapid adaptation to night-shift work.

While many techniques (e.g., light treatment) used to facilitate adaptation to night-shift work may help patients with this disorder, modafinil is the only therapeutic intervention that has ever been evaluated as a treatment for this specific patient population. Modafinil (200 mg, taken 30–60 min before the start of each night shift) is approved by the U.S. Food and Drug Administration as a treatment for the excessive sleepiness during night work in patients with SWD. Although treatment with modafinil significantly increases sleep latency and reduces the risk of lapses of attention during night work, SWD patients remain excessively sleepy at night, even while being treated with modafinil.

Safety programs should promote education about sleep and increase awareness of the hazards associated with night work. The goal should be to minimize both sleep deprivation and circadian disruption. Work schedules should be designed to minimize (1) exposure to night work, (2) the frequency of shift rotation so that shifts do not rotate more than once every 2–3 weeks, (3) the number of consecutive night shifts, and (4) the duration of night shifts. Shift durations of >16 h should be universally recognized as increasing the risk of sleep-related errors and performance lapses to a level that is unacceptable in nonemergency circumstances. At least 11 h off duty should be provided between work shifts, with at least one day off every week and two consecutive days off every month. Additional off duty time should be allocated after night work, since sleep efficiency is much lower during daytime hours.

DELAYED SLEEP PHASE DISORDER

Delayed sleep phase disorder is characterized by (1) reported sleep onset and wake times intractably later than desired, (2) actual sleep times at nearly the same clock hours daily, and (3) essentially normal all-night polysomnography except for delayed sleep onset. Patients exhibit an abnormally delayed endogenous circadian phase, with the temperature minimum during the constant routine occurring later than normal. This delayed phase could be due to (1) an abnormally long, genetically determined intrinsic period of the endogenous circadian pacemaker; (2) an abnormally reduced phase-advancing capacity of the pacemaker; (3) a slower rate of buildup of homeostatic sleep drive during wakefulness; or (4) an irregular prior sleep-wake schedule, characterized by frequent nights when the patient chooses to remain awake well past midnight (for social, school, or work reasons). In most cases, it is difficult to distinguish among these factors, since patients with an abnormally long intrinsic period are more likely to "choose" such late-night activities because they are unable to sleep at that time. Patients tend to be young adults. This self-perpetuating condition can persist for years and does not usually respond to attempts to reestablish normal bedtime hours. Treatment methods involving bright-light phototherapy during the morning hours or melatonin administration in the evening hours show promise in these patients, although the relapse rate is high.

ADVANCED SLEEP PHASE DISORDER

Advanced sleep phase disorder (ASPD) is the converse of the delayed sleep phase syndrome. Most commonly, this syndrome occurs in older people, 15% of whom report that they cannot sleep past 5 A.M., with twice that number complaining that they wake up too early at least several times per week. Patients with ASPD experience excessive daytime sleepiness during the evening hours, when they have great difficulty remaining awake, even in social settings. Typically, patients awaken from 3 to 5 A.M. each day, often several hours before their desired wake times. In addition to age-related ASPD, an early-onset familial variant of this condition has also been reported. In one such family, autosomal dominant ASPD was due to a missense mutation in a circadian clock component (PER2, as shown in Fig. 27-2) that altered the circadian period. Patients with ASPD may benefit from bright-light phototherapy during the evening hours, designed to reset the circadian pacemaker to a later hour.

NON-24-H SLEEP-WAKE DISORDER

This condition can occur when the synchronizing input (i.e., the light-dark cycle) from the environment to the circadian pacemaker is compromised (as in many blind people with no light perception) or when the maximal phase-advancing capacity of the circadian pacemaker is not adequate to accommodate the difference between the 24-h geophysical day and the intrinsic period of the pacemaker in the patient. Alternatively, patients' self-selected exposure to artificial light may drive the circadian pacemaker to a >24-h schedule. Affected patients are not able to maintain a stable phase relationship between the output of the pacemaker and the 24-h day. Such patients typically present with an incremental pattern of successive delays in sleep propensity, progressing in and out of phase with local time. When the patient's endogenous circadian rhythms are out of phase with the local environment, insomnia coexists with excessive daytime sleepiness. Conversely, when the endogenous circadian rhythms are in phase with the local environment, symptoms remit. The intervals between symptomatic periods may last several weeks to several months. Blind individuals unable to perceive light are particularly susceptible to this disorder, although it can occur in sighted patients. Nightly low-dose (0.5 mg) melatonin administration has been reported to improve sleep and, in some cases, to induce synchronization of the circadian pacemaker.

Medical Implications of Circadian Rhythmicity

Prominent circadian variations have been reported in the incidence of acute myocardial infarction, sudden cardiac death, and stroke, the leading causes of death in the United States. Platelet aggregability is increased in the early morning hours, coincident with the peak incidence of these cardiovascular events. Misalignment of circadian phase, such as occurs during night-shift work, induces insulin resistance and higher glucose levels in response to a standard meal. Blood pressure of night workers with sleep apnea is higher than that of day workers. A better understanding of the possible role of circadian rhythmicity in the acute destabilization of a chronic condition such as atherosclerotic disease could improve the understanding of its pathophysiology.

Diagnostic and therapeutic procedures may also be affected by the time of day at which data are collected. Examples include blood pressure, body temperature, the dexamethasone suppression test, and plasma cortisol levels. The timing of chemotherapy administration has been reported to have an effect on the outcome of treatment. In addition, both the toxicity and effectiveness of drugs can vary during the day. For example, more than a fivefold difference has been observed in mortality rates following administration of toxic agents to experimental animals at different times of day. Anesthetic agents are particularly sensitive to time-of-day effects. Finally, the physician must be increasingly aware of the public health risks associated with the ever-increasing demands made by the duty-rest-recreation schedules in our round-the-clock society.

FURTHER READINGS

AHMED I, THORPY M: Clinical features, diagnosis and treatment of narcolepsy. Clin Chest Med 31:371, 2010 [PMID: 20488294]

CZEISLER CA: Medical and genetic differences in the adverse impact of sleep loss on performance: Ethical considerations for the medical profession. Trans Am Clin Climatol Assoc 120:249, 2009 [PMID: 19768182]

DODSON ER et al: Therapeutics for circadian rhythm sleep disorder. Sleep Med Clin 5:701, 2010 [PMID: 21243069]

DOGHRAMJI K: The evaluation and management of insomnia. Clin Chest Med 31:327, 2010 [PMID: 20488291]

GAY PC: Sleep and sleep-disordered breathing in the hospitalized patient. Respir Care 55:1240, 2010 [PMID: 20800004]

MONJAN AA: Perspective on sleep and aging. Front Neurol 1:124, 2010 [PMID: 21173893]

ROMERO-CORRAL A et al: Interactions between obesity and obstructive sleep apnea: Implications for treatment. Chest 137:711, 2010 [PMID: 20202954]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE HUMAN VISUAL SYSTEM

The visual system provides a supremely efficient means for the rapid assimilation of information from the environment to aid in the guidance of behavior. The act of seeing begins with the capture of images focused by the cornea and lens on a light-sensitive membrane in the back of the eye called the *retina*. The retina is actually part of the brain, banished to the periphery to serve as a transducer for the conversion of patterns of light energy into neuronal signals. Light is absorbed by photopigment in two types of receptors: rods and cones. In the human retina there are 100 million rods and 5 million cones. The rods operate in dim (scotopic) illumination. The cones function under daylight (photopic) conditions. The cone system is specialized for color perception and high spatial resolution. The majority of cones are within the macula, the portion of the retina that serves the central 10° of vision. In the middle of the macula a small pit termed the *fovea*, packed exclusively with cones, provides the best visual acuity.

Photoreceptors hyperpolarize in response to light, activating bipolar, amacrine, and horizontal cells in the inner nuclear layer. After processing of photoreceptor responses by this complex retinal circuit, the flow of sensory information ultimately converges on a final common pathway: the ganglion cells. These cells translate the visual image impinging on the retina into a continuously varying barrage of action potentials that propagates along the primary optic pathway to visual centers within the brain. There are a million ganglion cells in each retina and hence a million fibers in each optic nerve.

Ganglion cell axons sweep along the inner surface of the retina in the nerve fiber layer, exit the eye at the optic disc, and travel through the optic nerve, optic chiasm, and optic tract to reach targets in the brain. The majority of fibers synapse on cells in the lateral geniculate body, a thalamic relay station. Cells in the lateral geniculate body project in turn to the primary visual cortex. This massive afferent retinogeniculocortical sensory pathway provides the neural substrate for visual perception. Although the lateral geniculate body is the main target of the retina, separate classes of ganglion cells project to other subcortical visual nuclei involved in different functions. Ganglion cells that mediate pupillary constriction and circadian rhythms are light sensitive owing to a novel visual pigment, melanopsin. Pupil responses are mediated by input to the pretectal olivary nuclei in the midbrain. The pretectal nuclei send their output to the Edinger-Westphal nuclei, which in turn provide parasympathetic innervation to the iris sphincter via an interneuron in the ciliary ganglion. Circadian rhythms are timed by a retinal projection to the suprachiasmatic nucleus. Visual orientation and eye movements are served by retinal input to the superior colliculus. Gaze stabilization and optokinetic reflexes are governed by a group of small retinal targets known collectively as the *brainstem accessory optic system*.

The eyes must be rotated constantly within their orbits to place and maintain targets of visual interest on the fovea. This activity, called *foveation*, or looking, is governed by an elaborate efferent motor system. Each eye is moved by six extraocular muscles that are supplied by cranial nerves from the oculomotor (III), trochlear (IV), and abducens (VI) nuclei. Activity in these ocular motor nuclei is coordinated by pontine and midbrain mechanisms for smooth pursuit, saccades, and gaze stabilization during head and body movements. Large regions of the frontal and parietooccipital cortex control these brainstem eye movement centers by providing descending supranuclear input.

CLINICAL ASSESSMENT OF VISUAL FUNCTION

Refractive State

In approaching a patient with reduced vision, the first step is to decide whether refractive error is responsible. In *emmetropia*, parallel rays from infinity are focused perfectly on the retina. Sadly, this condition is enjoyed by only a minority of the population. In *myopia*, the globe is too long, and light rays come to a focal point in front of the retina. Near objects can be seen clearly, but distant objects require a diverging lens in front of the eye. In *hyperopia*, the globe is too short, and hence a converging lens is used to supplement the refractive power of the eye. In *astigmatism*, the corneal surface is not perfectly spherical, necessitating a cylindrical corrective lens. In recent years it has become possible to correct refractive error with the excimer laser by performing LASIK (laser in situ keratomileusis) to alter the curvature of the cornea.

With the onset of middle age, *presbyopia* develops as the lens within the eye becomes unable to increase its refractive power to accommodate on near objects. To compensate for presbyopia, an

emmetropic patient must use reading glasses. A patient already wearing glasses for distance correction usually switches to bifocals. The only exception is a myopic patient, who may achieve clear vision at near simply by removing glasses containing the distance prescription.

Refractive errors usually develop slowly and remain stable after adolescence, except in unusual circumstances. For example, the acute onset of diabetes mellitus can produce sudden myopia because of lens edema induced by hyperglycemia. Testing vision through a pinhole aperture is a useful way to screen quickly for refractive error. If visual acuity is better through a pinhole than it is with the unaided eye, the patient needs refraction to obtain best corrected visual acuity.

Visual Acuity

The Snellen chart is used to test acuity at a distance of 6 m (20 ft). For convenience, a scale version of the Snellen chart called the Rosenbaum card is held at 36 cm (14 in.) from the patient (Fig. 28-1). All subjects should be able to read the 6/6 m (20/20 ft) line with each eye using their refractive correction, if any. Patients who need reading glasses because of presbyopia must wear them for accurate testing with the Rosenbaum card. If 6/6 (20/20) acuity is not present in each eye, the deficiency in vision must be explained. If it is worse than 6/240 (20/800), acuity should be recorded in terms of counting fingers, hand motions, light perception, or no light perception. Legal blindness is defined by the Internal Revenue Service as a best corrected acuity of 6/60 (20/200) or less in the better eye or a binocular visual field subtending 20° or less. For driving the laws vary by state, but most states require a corrected acuity of 6/12 (20/40) in at least one eye for unrestricted privileges. Patients with a homonymous hemianopia should not drive.

Figure 28-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The Rosenbaum card is a miniature, scale version of the Snellen chart for testing visual acuity at near. When the visual acuity is recorded, the Snellen distance equivalent should bear a notation indicating that vision was tested at near, not at 6 m (20 ft), or else the Jaeger number system should be used to report the acuity.

Pupils

The pupils should be tested individually in dim light with the patient fixating on a distant target. If the pupils respond briskly to light, there is no need to check the near response, because isolated loss of constriction (miosis) to accommodation does not occur. For this reason, the ubiquitous abbreviation PERRLA (pupils equal, round, and reactive to light and accommodation) implies a wasted effort with the last step. However, it is important to test the near response if the light response is poor or absent. Light-near dissociation occurs with neurosyphilis (Argyll Robertson pupil), with lesions of the dorsal midbrain (obstructive hydrocephalus, pineal region tumors), and after aberrant regeneration (oculomotor nerve palsy, Adie's tonic pupil).

An eye with no light perception has no pupillary response to direct light stimulation. If the retina or optic nerve is only partially injured, the direct pupillary response will be weaker than the consensual pupillary response evoked by shining a light into the other eye. This *relative afferent pupillary defect* (Marcus Gunn pupil) can be elicited with the swinging flashlight test (Fig. 28-2). It is an extremely

useful sign in retrobulbar optic neuritis and other optic nerve diseases, in which it may be the sole objective evidence for disease.

Figure 28-2

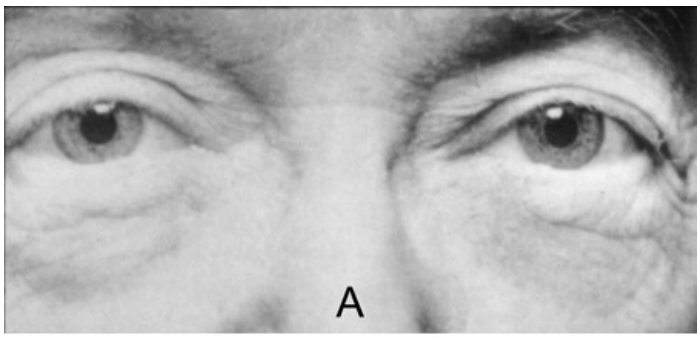

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

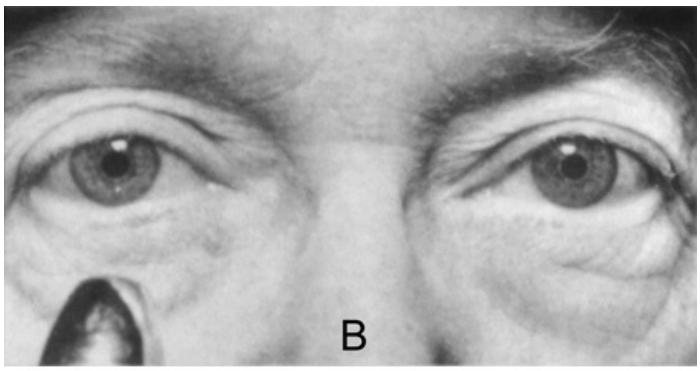

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

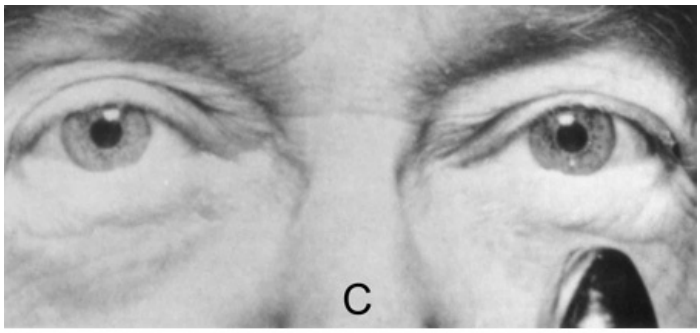

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Demonstration of a relative afferent pupil defect (Marcus Gunn pupil) in the left eye, done with the patient fixating on a distant target. ***A.*** With dim background lighting, the pupils are equal and relatively large. ***B.*** Shining a flashlight into the right eye evokes equal, strong constriction of both pupils. ***C.*** Swinging the flashlight over to the damaged left eye causes dilation of both pupils, although they remain smaller than in ***A.*** Swinging the flashlight back over to the healthy right eye would result in symmetric constriction back to the appearance shown in ***B.*** Note that the pupils always remain equal; the damage to the left retina/optic nerve is revealed by weaker bilateral pupil constriction to a flashlight in the left eye compared with the right eye. *(From P Levatin: Arch Ophthalmol 62:768, 1959. Copyright © 1959 American Medical Association. All rights reserved.*)

Subtle inequality in pupil size, up to 0.5 mm, is a fairly common finding in normal persons. The diagnosis of essential or physiologic anisocoria is secure as long as the relative pupil asymmetry remains constant as ambient lighting varies. Anisocoria that increases in dim light indicates a sympathetic paresis of the iris dilator muscle. The triad of miosis with ipsilateral ptosis and anhidrosis constitutes *Horner's syndrome*, although anhidrosis is an inconstant feature. Brainstem stroke, carotid dissection, and neoplasm impinging on the sympathetic chain occasionally are identified as the cause of Horner's syndrome, but most cases are idiopathic.

Anisocoria that increases in bright light suggests a parasympathetic palsy. The first concern is an oculomotor nerve paresis. This possibility is excluded if the eye movements are full and the patient has

no ptosis or diplopia. Acute pupillary dilation (mydriasis) can result from damage to the ciliary ganglion in the orbit. Common mechanisms are infection (herpes zoster, influenza), trauma (blunt, penetrating, surgical), and ischemia (diabetes, temporal arteritis). After denervation of the iris sphincter the pupil does not respond well to light, but the response to near is often relatively intact. When the near stimulus is removed, the pupil redilates very slowly compared with the normal pupil, hence the term *tonic pupil*. In *Adie's syndrome*, a tonic pupil occurs in conjunction with weak or absent tendon reflexes in the lower extremities. This benign disorder, which occurs predominantly in healthy young women, is assumed to represent a mild dysautonomia. Tonic pupils are also associated with Shy-Drager syndrome, segmental hypohidrosis, diabetes, and amyloidosis. Occasionally, a tonic pupil is discovered incidentally in an otherwise completely normal, asymptomatic individual. The diagnosis is confirmed by placing a drop of dilute (0.125%) pilocarpine into each eye. Denervation hypersensitivity produces pupillary constriction in a tonic pupil, whereas the normal pupil shows no response. Pharmacologic dilatation from accidental or deliberate instillation of anticholinergic agents (atropine, scopolamine drops) into the eye also can produce pupillary mydriasis. In this situation, normal strength (1%) pilocarpine causes no constriction.

Both pupils are affected equally by systemic medications. They are small with narcotic use (morphine, heroin) and large with anticholinergics (scopolamine). Parasympathetic agents (pilocarpine, demecarium bromide) used to treat glaucoma produce miosis. In any patient with an unexplained pupillary abnormality, a slit-lamp examination is helpful to exclude surgical trauma to the iris, an occult foreign body, perforating injury, intraocular inflammation, adhesions (synechia), angle-closure glaucoma, and iris sphincter rupture from blunt trauma.

Eye Movements and Alignment

Eye movements are tested by asking the patient, with both eyes open, to pursue a small target such as a penlight into the cardinal fields of gaze. Normal ocular versions are smooth, symmetric, full, and maintained in all directions without nystagmus. Saccades, or quick refixation eye movements, are assessed by having the patient look back and forth between two stationary targets. The eyes should move rapidly and accurately in a single jump to their target. Ocular alignment can be judged by holding a penlight directly in front of the patient at about 1 m. If the eyes are straight, the corneal light reflex will be centered in the middle of each pupil. To test eye alignment more precisely, the cover test is useful. The patient is instructed to gaze upon a small fixation target in the distance. One eye is covered suddenly while the second eye is observed. If the second eye shifts to fixate on the target, it was misaligned. If it does not move, the first eye is uncovered and the test is repeated on the second eye. If neither eye moves, the eyes are aligned orthotropically. If the eyes are orthotropic in primary gaze but the patient complains of diplopia, the cover test should be performed with the head tilted or turned in whatever direction elicits diplopia. With practice the examiner can detect an ocular deviation (heterotropia) as small as 1–2° with the cover test. Deviations can be measured by placing prisms in front of the misaligned eye to determine the power required to neutralize the fixation shift evoked by covering the other eye.

Stereopsis

Stereoacuity is determined by presenting targets with retinal disparity separately to each eye by using polarized images. The most popular office tests measure a range of thresholds from 800–40 seconds of arc. Normal stereoacuity is 40 seconds of arc. If a patient achieves this level of stereoacuity, one is assured that the eyes are aligned orthotropically and that vision is intact in each eye. Random dot stereograms have no monocular depth cues and provide an excellent screening test for strabismus and amblyopia in children.

Color Vision

The retina contains three classes of cones, with visual pigments of differing peak spectral sensitivity: red (560 nm), green (530 nm), and blue (430 nm). The red and green cone pigments are encoded on the X chromosome, and the blue cone pigment on chromosome 7. Mutations of the blue cone pigment are exceedingly rare. Mutations of the red and green pigments cause congenital X-linked color blindness in 8% of males. Affected individuals are not truly color blind; rather, they differ from normal subjects in the way they perceive color and the way they combine primary monochromatic lights to match a particular color. Anomalous trichromats have three cone types, but a mutation in one cone pigment (usually red or green) causes a shift in peak spectral sensitivity, altering the proportion of primary colors required to achieve a color match. Dichromats have only two cone types and therefore will accept a color match based on only two primary colors. Anomalous trichromats and dichromats have 6/6 (20/20) visual acuity, but their hue discrimination is impaired. Ishihara color plates can be used to detect red-green color blindness. The test plates contain a hidden number that is visible only to subjects with color confusion from red-green color blindness. Because color blindness is almost exclusively X-linked, it is worth screening only male children.

The Ishihara plates often are used to detect acquired defects in color vision, although they are intended as a screening test for congenital color blindness. Acquired defects in color vision frequently result from disease of the macula or optic nerve. For example, patients with a history of optic neuritis

often complain of color desaturation long after their visual acuity has returned to normal. Color blindness also can result from bilateral strokes involving the ventral portion of the occipital lobe (cerebral achromatopsia). Such patients can perceive only shades of gray and also may have difficulty recognizing faces (prosopagnosia). Infarcts of the dominant occipital lobe sometimes give rise to color anomia. Affected patients can discriminate colors but cannot name them.

Visual Fields

Vision can be impaired by damage to the visual system anywhere from the eyes to the occipital lobes. One can localize the site of the lesion with considerable accuracy by mapping the visual field deficit by finger confrontation and then correlating it with the topographic anatomy of the visual pathway (Fig. 28-3). Quantitative visual field mapping is performed by computer-driven perimeters (Humphrey, Octopus) that present a target of variable intensity at fixed positions in the visual field (Fig. 28-3*A*). By generating an automated printout of light thresholds, these static perimeters provide a sensitive means of detecting scotomas in the visual field. They are exceedingly useful for serial assessment of visual function in chronic diseases such as glaucoma and pseudotumor cerebri.

Figure 28-3

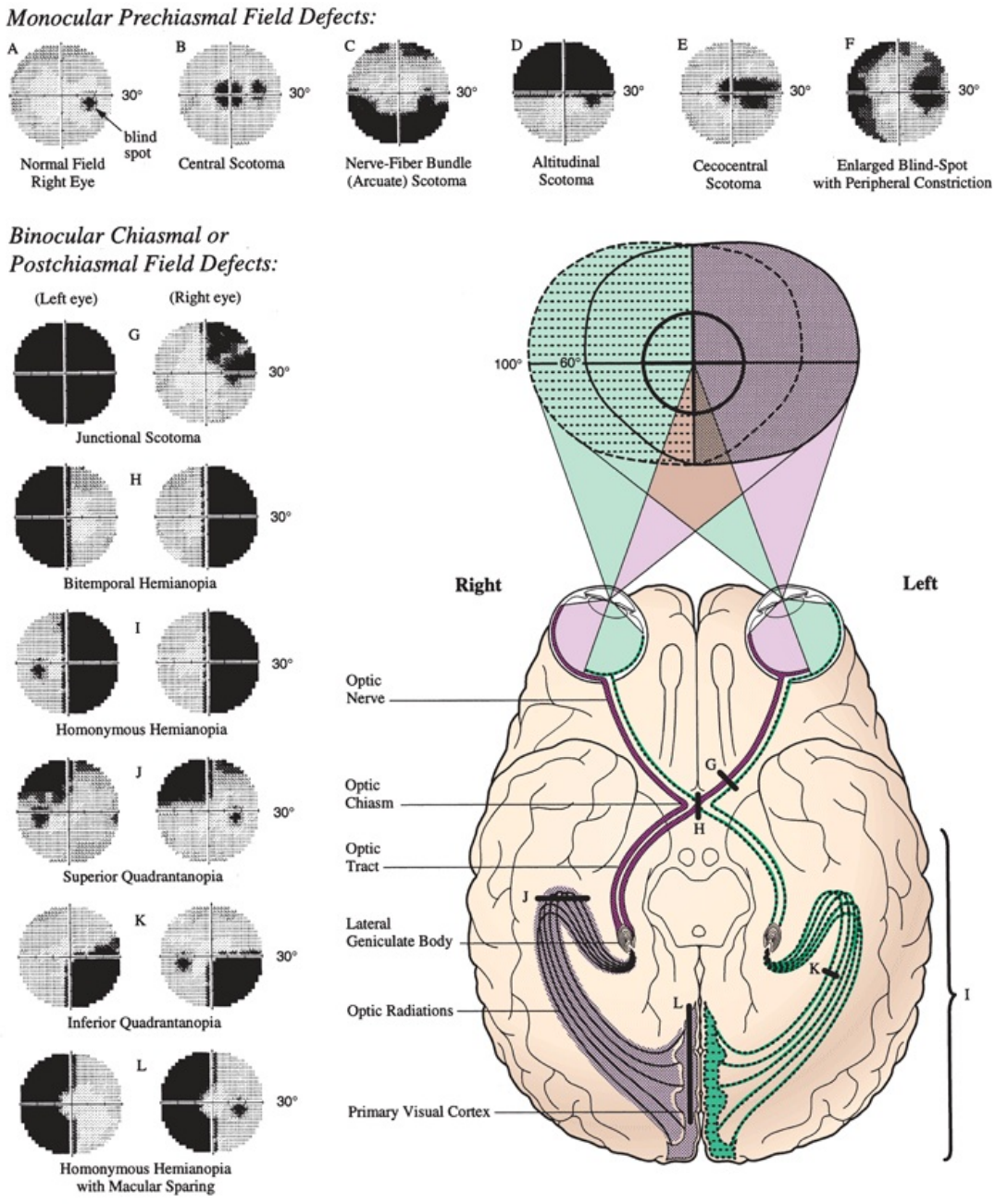

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ventral view of the brain, correlating patterns of visual field loss with the sites of lesions in

the visual pathway. The visual fields overlap partially, creating 120° of central binocular field flanked by a 40° monocular crescent on either side. The visual field maps in this figure were done with a computer-driven perimeter (Humphrey Instruments, Carl Zeiss, Inc.). It plots the retinal sensitivity to light in the central 30° by using a gray scale format. Areas of visual field loss are shown in black. The examples of common monocular, prechiasmal field defects are all shown for the right eye. By convention, the visual fields are always recorded with the left eye's field on the left and the right eye's field on the right, just as the patient sees the world.

The crux of visual field analysis is to decide whether a lesion is before, at, or behind the optic chiasm. If a scotoma is confined to one eye, it must be due to a lesion anterior to the chiasm, involving either the optic nerve or the retina. Retinal lesions produce scotomas that correspond optically to their location in the fundus. For example, a superior-nasal retinal detachment results in an inferior-temporal field cut. Damage to the macula causes a central scotoma (Fig. 28-3*B*).

Optic nerve disease produces characteristic patterns of visual field loss. Glaucoma selectively destroys axons that enter the superotemporal or inferotemporal poles of the optic disc, resulting in arcuate scotomas shaped like a Turkish scimitar, which emanate from the blind spot and curve around fixation to end flat against the horizontal meridian (Fig. 28-3*C*). This type of field defectmirrors the arrangement of the nerve fiber layer in the temporal retina. Arcuate or nerve fiber layer scotomas also result from optic neuritis, ischemic optic neuropathy, optic disc drusen, and branch retinal artery or vein occlusion.

Damage to the entire upper or lower pole of the optic disc causes an altitudinal field cut that follows the horizontal meridian (Fig. 28-3*D*). This pattern of visual field loss is typical of ischemic optic neuropathy but also results from retinal vascular occlusion, advanced glaucoma, and optic neuritis.

About half the fibers in the optic nerve originate from ganglion cells serving the macula. Damage to papillomacular fibers causes a cecocentral scotoma that encompasses the blind spot and macula (Fig. 28-3*E*). If the damage is irreversible, pallor eventually appears in the temporal portion of the optic disc. Temporal pallor from a cecocentral scotoma may develop in optic neuritis, nutritional optic neuropathy, toxic optic neuropathy, Leber's hereditary optic neuropathy, and compressive optic neuropathy. It is worth mentioning that the temporal side of the optic disc is slightly more pale than the nasal side in most normal individuals. Therefore, it sometimes can be difficult to decide whether the temporal pallor visible on fundus examination represents a pathologic change. Pallor of the nasal rim of the optic disc is a less equivocal sign of optic atrophy.

At the optic chiasm, fibers from nasal ganglion cells decussate into the contralateral optic tract. Crossed fibers are damaged more by compression than are uncrossed fibers. As a result, mass lesions of the sellar region cause a temporal hemianopia in each eye. Tumors anterior to the optic chiasm, such as meningiomas of the tuberculum sella, produce a junctional scotoma characterized by an optic neuropathy in one eye and a superior-temporal field cut in the other eye (Fig. 28-3*G*). More symmetric compression of the optic chiasm by a pituitary adenoma (Fig. 339-4), meningioma, craniopharyngioma, glioma, or aneurysm results in a bitemporal hemianopia (Fig. 28-3*H*). The insidious development of a bitemporal hemianopia often goes unnoticed by the patient and will escape detection by the physician unless each eye is tested separately.

Figure 28-4

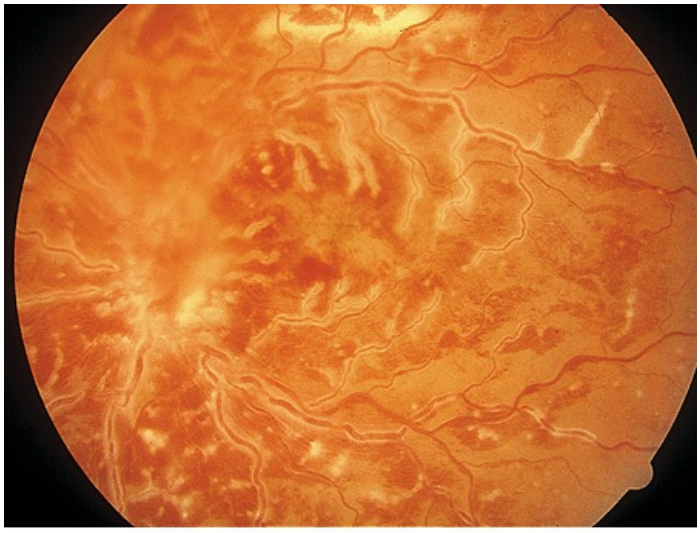

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Retinal vasculitis, uveitis, and hemorrhage in a 32-year-old woman with Crohn's disease. Note that the veins are frosted with a white exudate. Visual acuity improved from 20/400 to 20/20 after treatment with intravenous methylprednisolone.

It is difficult to localize a postchiasmal lesion accurately, because injury anywhere in the optic tract, lateral geniculate body, optic radiations, or visual cortex can produce a homonymous hemianopia (i.e., a temporal hemifield defect in the contralateral eye and a matching nasal hemifield defect in the ipsilateral eye) (Fig. 28-3*I*). A unilateral postchiasmal lesion leaves the visual acuity in each eye unaffected, although the patient may read the letters on only the left or right half of the eye chart. Lesions of the optic radiations tend to cause poorly matched or incongruous field defects in each eye. Damage to the optic radiations in the temporal lobe (Meyer's loop) produces a superior quadrantic homonymous hemianopia (Fig. 28-3*J*), whereas injury to the optic radiations in the parietal lobe results in an inferior quadrantic homonymous hemianopia (Fig. 28-3*K*). Lesions of the primary visual cortex give rise to dense, congruous hemianopic field defects. Occlusion of the posterior cerebral artery supplying the occipital lobe is a common cause of total homonymous hemianopia. Some patients with hemianopia after occipital stroke have macular sparing, because the macular representation at the tip of the occipital lobe is supplied by collaterals from the middle cerebral artery (Fig. 28-3*L*). Destruction of both occipital lobes produces cortical blindness. This condition can be distinguished from bilateral prechiasmal visual loss by noting that the pupil responses and optic fundi remain normal.

DISORDERS

Red or Painful Eye

CORNEAL ABRASIONS

Corneal abrasions are seen best by placing a drop of fluorescein in the eye and looking with the slit lamp, using a cobalt-blue light. A penlight with a blue filter will suffice if a slit lamp is not available. Damage to the corneal epithelium is revealed by yellow fluorescence of the exposed basement membrane underlying the epithelium. It is important to check for foreign bodies. To search the conjunctival fornices, the lower lid should be pulled down and the upper lid everted. A foreign body can be removed with a moistened cotton-tipped applicator after a drop of a topical anesthetic such as proparacaine has been placed in the eye. Alternatively, it may be possible to flush the foreign body from the eye by irrigating copiously with saline or artificial tears. If the corneal epithelium has been abraded, antibiotic ointment and a patch should be applied to the eye. A drop of an intermediate-acting cycloplegic such as cyclopentolate hydrochloride 1% helps reduce pain by relaxing the ciliary body. The eye should be reexamined the next day. Minor abrasions may not require patching and cycloplegia.

SUBCONJUNCTIVAL HEMORRHAGE

This results from rupture of small vessels bridging the potential space between the episclera and the conjunctiva. Blood dissecting into this space can produce a spectacular red eye, but vision is not affected and the hemorrhage resolves without treatment. Subconjunctival hemorrhage is usually spontaneous but can result from blunt trauma, eye rubbing, or vigorous coughing. Occasionally it is a clue to an underlying bleeding disorder.

PINGUECULA

Pinguecula is a small, raised conjunctival nodule at the temporal or nasal limbus. In adults such lesions are extremely common and have little significance unless they become inflamed (pingueculitis). A *pterygium* resembles a pinguecula but has crossed the limbus to encroach on the corneal surface. Removal is justified when symptoms of irritation or blurring develop, but recurrence is a common problem.

BLEPHARITIS

This refers to inflammation of the eyelids. The most common form occurs in association with acne rosacea or seborrheic dermatitis. The eyelid margins usually are colonized heavily by staphylococci. Upon close inspection, they appear greasy, ulcerated, and crusted with scaling debris that clings to the lashes. Treatment consists of warm compresses, strict eyelid hygiene, and topical antibiotics such as bacitracin/polymyxin B ophthalmic ointment. An external *hordeolum* (sty) is caused by staphylococcal infection of the superficial accessory glands of Zeis or Moll located in the eyelid margins. An internal hordeolum occurs after suppurative infection of the oil-secreting meibomian glands within the tarsal plate of the eyelid. Systemic antibiotics, usually tetracyclines or azithromycin, sometimes are necessary for treatment of meibomian gland inflammation (meibomitis) or chronic, severe blepharitis. A *chalazion* is a painless, granulomatous inflammation of a meibomian gland that produces a pealike nodule within the eyelid. It can be incised and drained or injected with glucocorticoids. Basal cell, squamous cell, or meibomian gland carcinoma should be suspected with any nonhealing ulcerative lesion of the eyelids.

DACRYOCYSTITIS

An inflammation of the lacrimal drainage system, dacryocystitis can produce epiphora (tearing) and

ocular injection. Gentle pressure over the lacrimal sac evokes pain and reflux of mucus or pus from the tear puncta. Dacryocystitis usually occurs after obstruction of the lacrimal system. It is treated with topical and systemic antibiotics, followed by probing or surgery to reestablish patency. *Entropion* (inversion of the eyelid) or *ectropion* (sagging or eversion of the eyelid) can also lead to epiphora and ocular irritation.

CONJUNCTIVITIS

Conjunctivitis is the most common cause of a red, irritated eye. Pain is minimal, and visual acuity is reduced only slightly. The most common viral etiology is adenovirus infection. It causes a watery discharge, a mild foreign-body sensation, and photophobia. Bacterial infection tends to produce a more mucopurulent exudate. Mild cases of infectious conjunctivitis usually are treated empirically with broad-spectrum topical ocular antibiotics such as sulfacetamide 10%, polymyxin-bacitracin, or a trimethoprim-polymyxin combination. Smears and cultures usually are reserved for severe, resistant, or recurrent cases of conjunctivitis. To prevent contagion, patients should be admonished to wash their hands frequently, not to touch their eyes, and to avoid direct contact with others.

ALLERGIC CONJUNCTIVITIS

This condition is extremely common and often is mistaken for infectious conjunctivitis. Itching, redness, and epiphora are typical. The palpebral conjunctiva may become hypertropic with giant excrescences called cobblestone papillae. Irritation from contact lenses or any chronic foreign body also can induce formation of cobblestone papillae. *Atopic conjunctivitis* occurs in subjects with atopic dermatitis or asthma. Symptoms caused by allergic conjunctivitis can be alleviated with cold compresses, topical vasoconstrictors, antihistamines, and mast cell stabilizers such as cromolyn sodium. Topical glucocorticoid solutions provide dramatic relief of immune-mediated forms of conjunctivitis, but their long-term use is ill advised because of the complications of glaucoma, cataract, and secondary infection. Topical nonsteroidal anti-inflammatory drugs (NSAIDs) (e.g., ketorolac tromethamine) are better alternatives.

KERATOCONJUNCTIVITIS SICCA

Also known as dry eye, this produces a burning foreign-body sensation, injection, and photophobia. In mild cases the eye appears surprisingly normal, but tear production measured by wetting of a filter paper (Schirmer strip) is deficient. A variety of systemic drugs, including antihistaminic, anticholinergic, and psychotropic medications, result in dry eye by reducing lacrimal secretion. Disorders that involve the lacrimal gland directly, such as sarcoidosis and Sjögren's syndrome, also cause dry eye. Patients may develop dry eye after radiation therapy if the treatment field includes the orbits. Problems with ocular drying are also common after lesions affecting cranial nerve V or VII. Corneal anesthesia is particularly dangerous, because the absence of a normal blink reflex exposes the cornea to injury without pain to warn the patient. Dry eye is managed by frequent and liberal application of artificial tears and ocular lubricants. In severe cases the tear puncta can be plugged or cauterized to reduce lacrimal outflow.

KERATITIS

Keratitis is a threat to vision because of the risk of corneal clouding, scarring, and perforation. Worldwide, the two leading causes of blindness from keratitis are trachoma from chlamydial infection and vitamin A deficiency related to malnutrition. In the United States, contact lenses play a major role in corneal infection and ulceration. They should not be worn by anyone with an active eye infection. In evaluating the cornea, it is important to differentiate between a superficial infection (*keratoconjunctivitis*) and a deeper, more serious ulcerative process. The latter is accompanied by greater visual loss, pain, photophobia, redness, and discharge. Slit-lamp examination shows disruption of the corneal epithelium, a cloudy infiltrate or abscess in the stroma, and an inflammatory cellular reaction in the anterior chamber. In severe cases, pus settles at the bottom of the anterior chamber, giving rise to a hypopyon. Immediate empirical antibiotic therapy should be initiated after corneal scrapings are obtained for Gram's stain, Giemsa stain, and cultures. Fortified topical antibiotics are most effective, supplemented with subconjunctival antibiotics as required. A fungal etiology should always be considered in a patient with keratitis. Fungal infection is common in warm humid climates, especially after penetration of the cornea by plant or vegetable material.

HERPES SIMPLEX

The *herpesviruses* are a major cause of blindness from keratitis. Most adults in the United States have serum antibodies to herpes simplex, indicating prior viral infection (Chap. 179). Primary ocular infection generally is caused by herpes simplex type 1 rather than type 2. It manifests as a unilateral follicular blepharoconjunctivitis that is easily confused with adenoviral conjunctivitis unless telltale vesicles appear on the periocular skin or conjunctiva. A dendritic pattern of corneal epithelial ulceration revealed by fluorescein staining is pathognomonic for herpes infection but is seen in only a minority of primary infections. Recurrent ocular infection arises from reactivation of the latent herpesvirus. Viral eruption in the corneal epithelium may result in the characteristic herpes dendrite. Involvement of the corneal stroma produces edema, vascularization, and iridocyclitis. Herpes keratitis is treated with topical antiviral agents, cycloplegics, and oral acyclovir. Topical glucocorticoids are effective in

mitigating corneal scarring but must be used with extreme caution because of the danger of corneal melting and perforation. Topical glucocorticoids also carry the risk of prolonging infection and inducing glaucoma.

HERPES ZOSTER

Herpes zoster from reactivation of latent varicella (chickenpox) virus causes a dermatomal pattern of painful vesicular dermatitis. Ocular symptoms can occur after zoster eruption in any branch of the trigeminal nerve but are particularly common when vesicles form on the nose, reflecting nasociliary (V1) nerve involvement (Hutchinson's sign). Herpes zoster ophthalmicus produces corneal dendrites, which can be difficult to distinguish from those seen in herpes simplex. Stromal keratitis, anterior uveitis, raised intraocular pressure, ocular motor nerve palsies, acute retinal necrosis, and postherpetic scarring and neuralgia are other common sequelae. Herpes zoster ophthalmicus is treated with antiviral agents and cycloplegics. In severe cases, glucocorticoids may be added to prevent permanent visual loss from corneal scarring.

EPISCLERITIS

This is an inflammation of the episclera, a thin layer of connective tissue between the conjunctiva and the sclera. Episcleritis resembles conjunctivitis, but it is a more localized process and discharge is absent. Most cases of episcleritis are idiopathic, but some occur in the setting of an autoimmune disease. *Scleritis* refers to a deeper, more severe inflammatory process that frequently is associated with a connective tissue disease such as rheumatoid arthritis, lupus erythematosus, polyarteritis nodosa, granulomatosis with polyangiitis (Wegener's) or relapsing polychondritis. The inflammation and thickening of the sclera can be diffuse or nodular. In anterior forms of scleritis, the globe assumes a violet hue and the patient complains of severe ocular tenderness and pain. With posterior scleritis the pain and redness may be less marked, but there is often proptosis, choroidal effusion, reduced motility, and visual loss. Episcleritis and scleritis should be treated with NSAIDs. If these agents fail, topical or even systemic glucocorticoid therapy may be necessary, especially if an underlying autoimmune process is active.

UVEITIS

Involving the anterior structures of the eye, uveitis also is called *iritis* or *iridocyclitis*. The diagnosis requires slit-lamp examination to identify inflammatory cells floating in the aqueous humor or deposited on the corneal endothelium (keratic precipitates). Anterior uveitis develops in sarcoidosis, ankylosing spondylitis, juvenile rheumatoid arthritis, inflammatory bowel disease, psoriasis, reactive arthritis (formerly known as Reiter's syndrome), and Behçet's disease. It also is associated with herpes infections, syphilis, Lyme disease, onchocerciasis, tuberculosis, and leprosy. Although anterior uveitis can occur in conjunction with many diseases, no cause is found to explain the majority of cases. For this reason, laboratory evaluation usually is reserved for patients with recurrent or severe anterior uveitis. Treatment is aimed at reducing inflammation and scarring by judicious use of topical glucocorticoids. Dilatation of the pupil reduces pain and prevents the formation of synechiae.

POSTERIOR UVEITIS

This is diagnosed by observing inflammation of the vitreous, retina, or choroid on fundus examination. It is more likely than anterior uveitis to be associated with an identifiable systemic disease. Some patients have panuveitis, or inflammation of both the anterior and posterior segments of the eye. Posterior uveitis is a manifestation of autoimmune diseases such as sarcoidosis, Behçet's disease, Vogt-Koyanagi-Harada syndrome, and inflammatory bowel disease **(Fig. 28-4)**. It also accompanies diseases such as toxoplasmosis, onchocerciasis, cysticercosis, coccidioidomycosis, toxocariasis, and histoplasmosis; infections caused by organisms such as *Candida*, *Pneumocystis carinii*, *Cryptococcus*, *Aspergillus*, herpes, and cytomegalovirus (see Fig. 182-1); and other diseases, such as syphilis, Lyme disease, tuberculosis, cat-scratch disease, Whipple's disease, and brucellosis. In multiple sclerosis, chronic inflammatory changes can develop in the extreme periphery of the retina (pars planitis or intermediate uveitis).

ACUTE ANGLE-CLOSURE GLAUCOMA

This is a rare and frequently misdiagnosed cause of a red, painful eye. Susceptible eyes have a shallow anterior chamber because the eye has either a short axial length (hyperopia) or a lens enlarged by the gradual development of cataract. When the pupil becomes mid-dilated, the peripheral iris blocks aqueous outflow via the anterior chamber angle and the intraocular pressure rises abruptly, producing pain, injection, corneal edema, obscurations, and blurred vision. In some patients, ocular symptoms are overshadowed by nausea, vomiting, or headache, prompting a fruitless workup for abdominal or neurologic disease. The diagnosis is made by measuring the intraocular pressure during an acute attack or by observing a narrow chamber angle by means of a specially mirrored contact lens. Acute angle closure is treated with acetazolamide (PO or IV), topical beta blockers, prostaglandin analogues, α_2-adrenergic agonists, and pilocarpine to induce miosis. If these measures fail, a laser can be used to create a hole in the peripheral iris to relieve pupillary block. Many physicians are reluctant to dilate patients routinely for fundus examination because they fear precipitating an angle-closure glaucoma. The risk is actually remote and more than outweighed by the potential benefit to

patients of discovering a hidden fundus lesion visible only through a fully dilated pupil. Moreover, a single attack of angle closure after pharmacologic dilatation rarely causes any permanent damage to the eye and serves as an inadvertent provocative test to identify patients with narrow angles who would benefit from prophylactic laser iridectomy.

ENDOPHTHALMITIS

This results from bacterial, viral, fungal, or parasitic infection of the internal structures of the eye. It usually is acquired by hematogenous seeding from a remote site. Chronically ill, diabetic, or immunosuppressed patients, especially those with a history of indwelling IV catheters or positive blood cultures, are at greatest risk for endogenous endophthalmitis. Although most patients have ocular pain and injection, visual loss is sometimes the only symptom. Septic emboli from a diseased heart valve or a dental abscess that lodge in the retinal circulation can give rise to endophthalmitis. White-centered retinal hemorrhages (Roth's spots) are considered pathognomonic for subacute bacterial endocarditis, but they also appear in leukemia, diabetes, and many other conditions. Endophthalmitis also occurs as a complication of ocular surgery, occasionally months or even years after the operation. An occult penetrating foreign body or unrecognized trauma to the globe should be considered in any patient with unexplained intraocular infection or inflammation.

Transient or Sudden Visual Loss

AMAUROSIS FUGAX

This term refers to a transient ischemic attack of the retina (Chap. 370). Because neural tissue has a high rate of metabolism, interruption of blood flow to the retina for more than a few seconds results in *transient monocular blindness* , a term used interchangeably with amaurosis fugax. Patients describe a rapid fading of vision like a curtain descending, sometimes affecting only a portion of the visual field. Amaurosis fugax usually results from an embolus that becomes stuck within a retinal arteriole **(Fig. 28-5)**. If the embolus breaks up or passes, flow is restored and vision returns quickly to normal without permanent damage. With prolonged interruption of blood flow, the inner retina suffers infarction. Ophthalmoscopy reveals zones of whitened, edematous retina following the distribution of branch retinal arterioles. Complete occlusion of the central retinal artery produces arrest of blood flow and a milky retina with a cherry-red fovea **(Fig. 28-6)**. Emboli are composed of cholesterol (Hollenhorst plaque), calcium, or platelet-fibrin debris. The most common source is an atherosclerotic plaque in the carotid artery or aorta, although emboli also can arise from the heart, especially in patients with diseased valves, atrial fibrillation, or wall motion abnormalities.

Figure 28-5

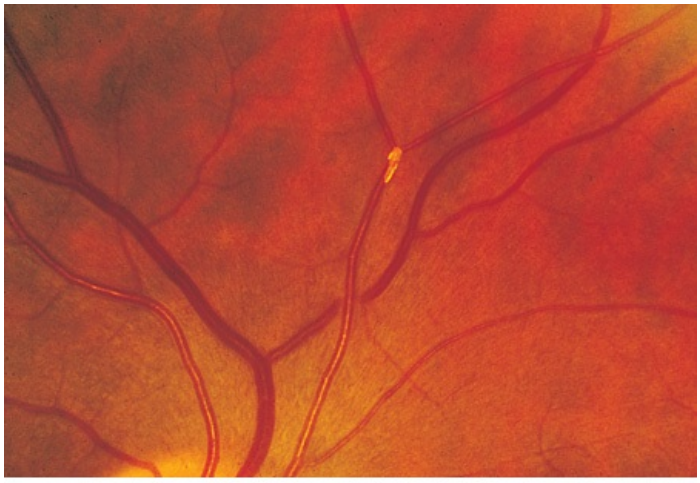

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hollenhorst plaque lodged at the bifurcation of a retinal arteriole proves that a patient is shedding emboli from the carotid artery, great vessels, or heart.

Figure 28-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Central retinal artery occlusion combined with ischemic optic neuropathy in a 19-year-old woman with an elevated titer of anticardiolipin antibodies. Note the orange dot (rather than cherry red) corresponding to the fovea and the spared patch of retina just temporal to the optic disc.

In rare instances, amaurosis fugax results from low central retinal artery perfusion pressure in a patient with a critical stenosis of the ipsilateral carotid artery and poor collateral flow via the circle of Willis. In this situation, amaurosis fugax develops when there is a dip in systemic blood pressure or a slight worsening of the carotid stenosis. Sometimes there is contralateral motor or sensory loss, indicating concomitant hemispheric cerebral ischemia.

Retinal arterial occlusion also occurs rarely in association with retinal migraine, lupus erythematosus, anticardiolipin antibodies (Fig. 28-6), anticoagulant deficiency states (protein S, protein C, and antithrombin deficiency), pregnancy, IV drug abuse, blood dyscrasias, dysproteinemias, and temporal arteritis.

Marked *systemic hypertension* causes sclerosis of retinal arterioles, splinter hemorrhages, focal infarcts of the nerve fiber layer (cotton-wool spots), and leakage of lipid and fluid (hard exudate) into the macula **(Fig. 28-7)**. In hypertensive crisis, sudden visual loss can result from vasospasm of retinal arterioles and retinal ischemia. In addition, acute hypertension may produce visual loss from ischemic swelling of the optic disc. Patients with acute hypertensive retinopathy should be treated by lowering the blood pressure. However, the blood pressure should not be reduced precipitously, because there is a danger of optic disc infarction from sudden hypoperfusion.

Figure 28-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypertensive retinopathy with scattered flame (splinter) hemorrhages and cotton-wool spots (nerve fiber layer infarcts) in a patient with headache and a blood pressure of 234/120.

Impending *branch* or *central retinal vein occlusion* can produce prolonged visual obscurations that resemble those described by patients with amaurosis fugax. The veins appear engorged and phlebitic, with numerous retinal hemorrhages **(Fig. 28-8)**. In some patients venous blood flow recovers spontaneously, whereas others evolve a frank obstruction with extensive retinal bleeding ("blood and thunder" appearance), infarction, and visual loss. Venous occlusion of the retina is often idiopathic, but hypertension, diabetes, and glaucoma are prominent risk factors. Polycythemia, thrombocythemia, or other factors leading to an underlying hypercoagulable state should be corrected; aspirin treatment may be beneficial.

Figure 28-8

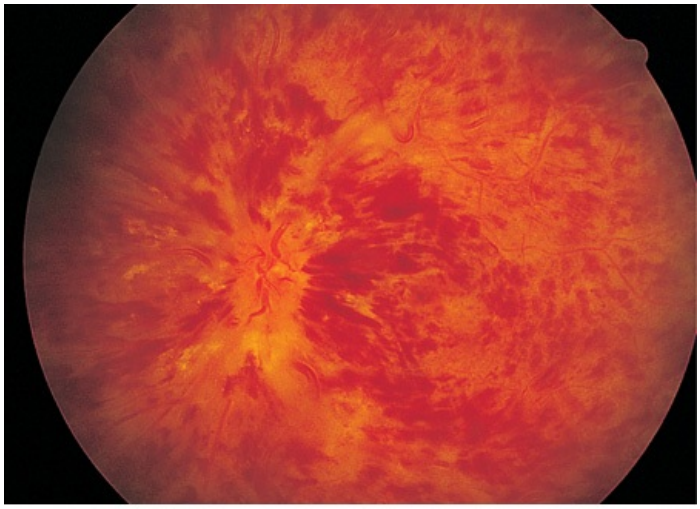

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Central retinal vein occlusion can produce massive retinal hemorrhage ("blood and thunder"), ischemia, and vision loss.

ANTERIOR ISCHEMIC OPTIC NEUROPATHY (AION)

This is caused by insufficient blood flow through the posterior ciliary arteries that supply the optic disc. It produces painless, monocular visual loss that is usually sudden, although some patients have progressive worsening. The optic disc appears swollen and surrounded by nerve fiber layer splinter hemorrhages **(Fig. 28-9)**. AION is divided into two forms: arteritic and nonarteritic. The nonarteritic form is most common. No specific cause can be identified, although diabetes and hypertension are common risk factors. No treatment is available. About 5% of patients, especially those >age 60, develop the arteritic form of AION in conjunction with giant cell (temporal) arteritis (Chap. 326). It is urgent to recognize arteritic AION so that high doses of glucocorticoids can be instituted immediately to prevent

blindness in the second eye. Symptoms of polymyalgia rheumatica may be present; the sedimentation rate and C-reactive protein level are usually elevated. In a patient with visual loss from suspected arteritic AION, temporal artery biopsy is mandatory to confirm the diagnosis. Glucocorticoids should be started immediately, without waiting for the biopsy to be completed. The diagnosis of arteritic AION is difficult to sustain in the face of a negative temporal artery biopsy, but such cases do occur rarely.

Figure 28-9

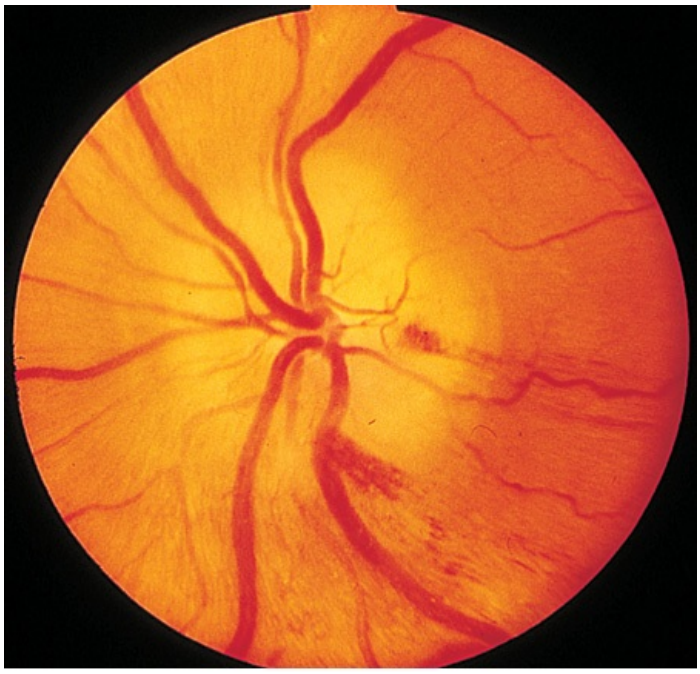

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anterior ischemic optic neuropathy from temporal arteritis in a 78-year-old woman with pallid disc swelling, hemorrhage, visual loss, myalgia, and an erythrocyte sedimentation rate of 86 mm/h.

POSTERIOR ISCHEMIC OPTIC NEUROPATHY

This is an uncommon cause of acute visual loss, induced by the combination of severe anemia and hypotension. Cases have been reported after major blood loss during surgery, exsanguinating trauma, gastrointestinal bleeding, and renal dialysis. The fundus usually appears normal, although optic disc swelling develops if the process extends far enough anteriorly. Vision can be salvaged in some patients by prompt blood transfusion and reversal of hypotension.

OPTIC NEURITIS

This is a common inflammatory disease of the optic nerve. In the Optic Neuritis Treatment Trial (ONTT), the mean age of patients was 32 years, 77% were female, 92% had ocular pain (especially with eye movements), and 35% had optic disc swelling. In most patients, the demyelinating event was retrobulbar and the ocular fundus appeared normal on initial examination **(Fig. 28-10)**, although optic disc pallor slowly developed over subsequent months.

Figure 28-10

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Retrobulbar optic neuritis is characterized by a normal fundus examination initially, hence the rubric "the doctor sees nothing, and the patient sees nothing." Optic atrophy develops after severe or repeated attacks.

Virtually all patients experience a gradual recovery of vision after a single episode of optic neuritis, even without treatment. This rule is so reliable that failure of vision to improve after a first attack of optic neuritis casts doubt on the original diagnosis. Treatment with high-dose IV methylprednisolone (250 mg every 6 h for 3 days) followed by oral prednisone (1 mg/kg per d for 11 days) makes no difference in final acuity (measured 6 months after the attack), but the recovery of visual function occurs more rapidly.

For some patients, optic neuritis remains an isolated event. However, the ONTT showed that the 15-year cumulative probability of developing clinically definite multiple sclerosis after optic neuritis is 50%. In patients with two or more demyelinating plaques on brain magnetic resonance (MR) imaging, treatment with interferon β-1a can retard the development of more lesions. In summary, an MR scan is recommended in every patient with a first attack of optic neuritis. When visual loss is severe (worse than 20/100), treatment with IV followed by oral glucocorticoids hastens recovery. If multiple lesions are present on the MR scan, treatment with interferon β-1a should be considered.

LEBER'S HEREDITARY OPTIC NEUROPATHY

This disease usually affects young men, causing gradual, painless, severe central visual loss in one eye, followed weeks or months later by the same process in the other eye. Acutely, the optic disc appears mildly plethoric with surface capillary telangiectases but no vascular leakage on fluorescein angiography. Eventually optic atrophy ensues. Leber's optic neuropathy is caused by a point mutation at codon 11778 in the mitochondrial gene encoding nicotinamide adenine dinucleotide dehydrogenase (NADH) subunit 4. Additional mutations responsible for the disease have been identified, most in mitochondrial genes that encode proteins involved in electron transport. Mitochondrial mutations that cause Leber's neuropathy are inherited from the mother by all her children, but usually only sons develop symptoms.

TOXIC OPTIC NEUROPATHY

This can result in acute visual loss with bilateral optic disc swelling and central or cecocentral scotomas. Such cases have been reported to result from exposure to ethambutol, methyl alcohol (moonshine), ethylene glycol (antifreeze), or carbon monoxide. In toxic optic neuropathy, visual loss also can develop gradually and produce optic atrophy **(Fig. 28-11)** without a phase of acute optic disc edema. Many agents have been implicated as a cause of toxic optic neuropathy, but the evidence supporting the association for many is weak. The following is a partial list of potential offending drugs or toxins: disulfiram, ethchlorvynol, chloramphenicol, amiodarone, monoclonal anti-CD3 antibody, ciprofloxacin, digitalis, streptomycin, lead, arsenic, thallium, d-penicillamine, isoniazid, emetine, and sulfonamides. Deficiency states induced by starvation, malabsorption, or alcoholism can lead to insidious visual loss. Thiamine, vitamin B12 , and folate levels should be checked in any patient with unexplained bilateral central scotomas and optic pallor.

Figure 28-11

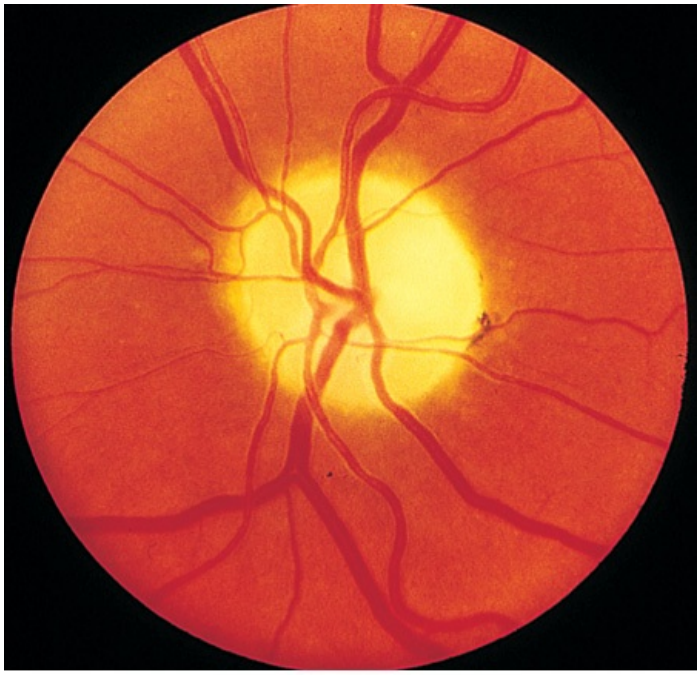

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Optic atrophy is not a specific diagnosis but refers to the combination of optic disc pallor, arteriolar narrowing, and nerve fiber layer destruction produced by a host of eye diseases, especially optic neuropathies.

PAPILLEDEMA

This connotes bilateral optic disc swelling from raised intracranial pressure **(Fig. 28-12)**. Headache is a common but not invariable accompaniment. All other forms of optic disc swelling (e.g., from optic neuritis or ischemic optic neuropathy, should be called "optic disc edema"). This convention is arbitrary but serves to avoid confusion. Often it is difficult to differentiate papilledema from other forms of optic disc edema by fundus examination alone. Transient visual obscurations are a classic symptom of papilledema. They can occur in only one eye or simultaneously in both eyes. They usually last seconds but can persist longer. Obscurations follow abrupt shifts in posture or happen spontaneously. When obscurations are prolonged or spontaneous, the papilledema is more threatening. Visual acuity is not affected by papilledema unless the papilledema is severe, long-standing, or accompanied by macular edema and hemorrhage. Visual field testing shows enlarged blind spots and peripheral constriction (Fig. 28-3*F*). With unremitting papilledema, peripheral visual field loss progresses in an insidious fashion while the optic nerve develops atrophy. In this setting, reduction of optic disc swelling is an ominous sign of a dying nerve rather than an encouraging indication of resolving papilledema.

Figure 28-12

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Papilledema means optic disc edema from raised intracranial pressure. This obese young woman with pseudotumor cerebri was misdiagnosed as a migraineur until fundus examination was performed, showing optic disc elevation, hemorrhages, and cotton-wool spots.

Evaluation of papilledema requires neuroimaging to exclude an intracranial lesion. MR angiography is appropriate in selected cases to search for a dural venous sinus occlusion or an arteriovenous shunt. If neuroradiologic studies are negative, the subarachnoid opening pressure should be measured by lumbar puncture. An elevated pressure, with normal cerebrospinal fluid, points by exclusion to the diagnosis of *pseudotumor cerebri* (idiopathicintracranial hypertension). The majority of patients are young, female, and obese. Treatment with a carbonic anhydrase inhibitor such as acetazolamide lowers intracranial pressure by reducing the production of cerebrospinal fluid. Weight reduction is vital but often unsuccessful. If acetazolamide and weight loss fail and visual field loss is progressive, a shunt should be performed without delay to prevent blindness. Occasionally, emergency surgery is required for sudden blindness caused by fulminant papilledema.

OPTIC DISC DRUSEN

These are refractile deposits within the substance of the optic nerve head **(Fig. 28-13)**. They are unrelated to drusen of the retina, which occur in age-related macular degeneration. Optic disc drusen are most common in people of northern European descent. Their diagnosis is obvious when they are visible as glittering particles on the surface of the optic disc. However, in many patients they are hidden beneath the surface, producing pseudopapilledema. It is important to recognize optic disc drusen to avoid an unnecessary evaluation for papilledema. Ultrasound or CT scanning is sensitive for detection of buried optic disc drusen because they contain calcium. In most patients, optic disc drusen are an incidental, innocuous finding, but they can produce visual obscurations. On perimetry they give rise to enlarged blind spots and arcuate scotomas from damage to the optic disc. With increasing age, drusen tend to become more exposed on the disc surface as optic atrophy develops. Hemorrhage, choroidal neovascular membrane, and AION are more likely to occur in patients with optic disc drusen. No treatment is available.

Figure 28-13

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Optic disc drusen are calcified deposits of unknown etiology within the optic disc. They sometimes are confused with papilledema.

VITREOUS DEGENERATION

This occurs in all individuals with advancing age, leading to visual symptoms. Opacities develop in the vitreous, casting annoying shadows on the retina. As the eye moves, these distracting "floaters" move synchronously, with a slight lag caused by inertia of the vitreous gel. Vitreous traction on the retina causes mechanical stimulation, resulting in perception of flashing lights. This photopsia is brief and is confined to one eye, in contrast to the bilateral, prolonged scintillations of cortical migraine. Contraction of the vitreous can result in sudden separation from the retina, heralded by an alarming shower of floaters and photopsia. This process, known as *vitreous detachment*, is a common involutional event in the elderly. It is not harmful unless it damages the retina. A careful examination of the dilated fundus is important in any patient complaining of floaters or photopsia to search for peripheral tears or holes. If such a lesion is found, laser application can forestall a retinal detachment. Occasionally a tear ruptures a retinal blood vessel, causing vitreous hemorrhage and sudden loss of vision. On attempted ophthalmoscopy the fundus is hidden by a dark red haze of blood. Ultrasound is required to examine the interior of the eye for a retinal tear or detachment. If the hemorrhage does not resolve spontaneously, the vitreous can be removed surgically. Vitreous hemorrhage also results from the fragile neovascular vessels that proliferate on the surface of the retina in diabetes, sickle cell anemia, and other ischemic ocular diseases.

RETINAL DETACHMENT

This produces symptoms of floaters, flashing lights, and a scotoma in the peripheral visual field corresponding to the detachment **(Fig. 28-14)**. If the detachment includes the fovea, there is an afferent pupil defect and the visual acuity is reduced. In most eyes, retinal detachment starts with a hole, flap, or tear in the peripheral retina (rhegmatogenous retinal detachment). Patients with peripheral retinal thinning (lattice degeneration) are particularly vulnerable to this process. Once a break has developed in the retina, liquefied vitreous is free to enter the subretinal space, separating the retina from the pigment epithelium. The combination of vitreous traction on the retinal surface and passage of fluid behind the retina leads inexorably to detachment. Patients with a history of myopia, trauma, or prior cataract extraction are at greatest risk for retinal detachment. The diagnosis is confirmed by ophthalmoscopic examination of the dilated eye.

Figure 28-14

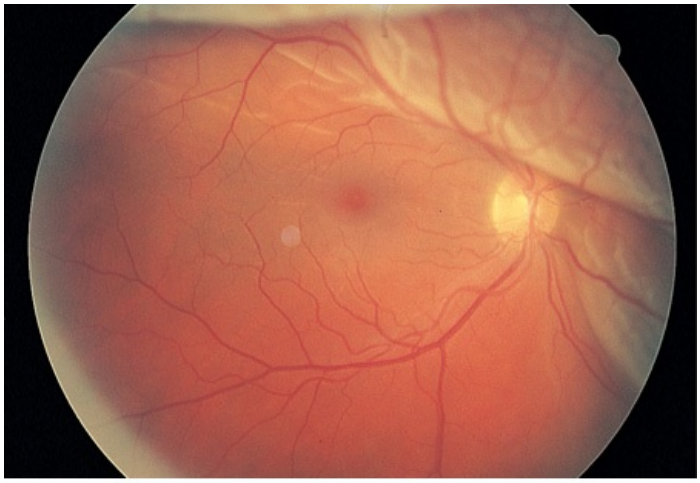

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Retinal detachment appears as an elevated sheet of retinal tissue with folds. In this patient the fovea was spared, so acuity was normal, but a superior detachment produced an inferior scotoma.

CLASSIC MIGRAINE

(See also Chap. 14) This usually occurs with a visual aura lasting about 20 min. In a typical attack, a small central disturbance in the field of vision marches toward the periphery, leaving a transient scotoma in its wake. The expanding border of migraine scotoma has a scintillating, dancing, or zigzag edge, resembling the bastions of a fortified city, hence the term *fortification spectra*. Patients' descriptions of fortification spectra vary widely and can be confused with amaurosis fugax. Migraine patterns usually last longer and are perceived in both eyes, whereas amaurosis fugax is briefer and occurs in only one eye. Migraine phenomena also remain visible in the dark or with the eyes closed. Generally they are confined to either the right or the left visual hemifield, but sometimes both fields are involved simultaneously. Patients often have a long history of stereotypic attacks. After the visual symptoms recede, headache develops in most patients.

TRANSIENT ISCHEMIC ATTACKS

Vertebrobasilar insufficiency may result in acute homonymous visual symptoms. Many patients mistakenly describe symptoms in the left or right eye when in fact the symptoms are occurring in the left or right hemifield of both eyes. Interruption of blood supply to the visual cortex causes a sudden fogging or graying of vision, occasionally with flashing lights or other positive phenomena that mimic migraine. Cortical ischemic attacks are briefer in duration than migraine, occur in older patients, and are not followed by headache. There may be associated signs of brainstem ischemia, such as diplopia, vertigo, numbness, weakness, and dysarthria.

STROKE

Stroke occurs when interruption of blood supply from the posterior cerebral artery to the visual cortex is prolonged. The only finding on examination is a homonymous visual field defect that stops abruptly at the vertical meridian. Occipital lobe stroke usually is due to thrombotic occlusion of the vertebrobasilar system, embolus, or dissection. Lobar hemorrhage, tumor, abscess, and arteriovenous malformation are other common causes of hemianopic cortical visual loss.

FACTITIOUS (FUNCTIONAL, NONORGANIC) VISUAL LOSS

This is claimed by hysterics or malingerers. The latter account for the vast majority, seeking sympathy, special treatment, or financial gain by feigning loss of sight. The diagnosis is suspected when the history is atypical, physical findings are lacking or contradictory, inconsistencies emerge on testing, and a secondary motive can be identified. In our litigious society, the fraudulent pursuit of recompense has spawned an epidemic of factitious visual loss.

Chronic Visual Loss

CATARACT

Cataract is a clouding of the lens sufficient to reduce vision. Most cataracts develop slowly as a result of aging, leading to gradual impairment of vision. The formation of cataract occurs more rapidly in patients with a history of ocular trauma, uveitis, or diabetes mellitus. Cataracts are acquired in a variety of genetic diseases, such as myotonic dystrophy, neurofibromatosis type 2, and galactosemia. Radiation therapy and glucocorticoid treatment can induce cataract as a side effect. The cataracts associated with radiation or glucocorticoids have a typical posterior subcapsular location. Cataract can be detected by noting an impaired red reflex when viewing light reflected from the fundus with an ophthalmoscope or by examining the dilated eye with the slit lamp.

The only treatment for cataract is surgical extraction of the opacified lens. Over a million cataract operations are performed each year in the United States. The operation generally is done under local anesthesia on an outpatient basis. A plastic or silicone intraocular lens is placed within the empty lens capsule in the posterior chamber, substituting for the natural lens and leading to rapid recovery of sight. More than 95% of patients who undergo cataract extraction can expect an improvement in vision. In some patients, the lens capsule remaining in the eye after cataract extraction eventually turns cloudy, causing secondary loss of vision. A small opening is made in the lens capsule with a laser to restore clarity.

GLAUCOMA

Glaucoma is a slowly progressive, insidious optic neuropathy that usually is associated with chronic elevation of intraocular pressure. In African Americans it is the leading cause of blindness. The mechanism by which raised intraocular pressure injures the optic nerve is not understood. Axons entering the inferotemporal and superotemporal aspects of the optic disc are damaged first, producing typical nerve fiber bundle or arcuate scotomas on perimetric testing. As fibers are destroyed, the neural rim of the optic disc shrinks and the physiologic cup within the optic disc enlarges **(Fig. 28-15)**. This process is referred to as pathologic "cupping." The cup-to-disc diameter is expressed as a ratio (e.g., 0.2/1). The cup-to-disc ratio ranges widely in normal individuals, making it difficult to diagnose glaucoma reliably simply by observing an unusually large or deep optic cup. Careful documentation of serial examinations is helpful. In a patient with physiologic cupping the large cup remains stable, whereas in a patient with glaucoma it expands relentlessly over the years. Detection of visual field loss by computerized perimetry also contributes to the diagnosis. Finally, most patients with glaucoma have raised intraocular pressure. However, many patients with typical glaucomatous cupping and visual field loss have intraocular pressures that apparently never exceed the normal limit of 20 mmHg (so-called low-tension glaucoma).

Figure 28-15

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Glaucoma results in "cupping" as the neural rim is destroyed and the central cup becomes enlarged and excavated. The cup-to-disc ratio is about 0.7/1.0 in this patient.

In acute angle-closure glaucoma, the eye is red and painful due to abrupt, severe elevation of intraocular pressure. Such cases account for only a minority of glaucoma cases: most patients have open, anterior chamber angles. The cause of raised intraocular pressure in open angle glaucoma is unknown, but it is associated with gene mutations in the heritable forms.

Glaucoma is usually painless (except in angle-closure glaucoma). Foveal acuity is spared until end-stage disease is reached. For these reasons, severe and irreversible damage can occur before either the patient or the physician recognizes the diagnosis. Screening of patients for glaucoma by noting the cup-to-disc ratio on ophthalmoscopy and by measuring intraocular pressure is vital. Glaucoma is treated with topical adrenergic agonists, cholinergic agonists, beta blockers, and prostaglandin analogues. Occasionally, systemic absorption of beta blocker from eyedrops can be sufficient to cause side effects of bradycardia, hypotension, heart block, bronchospasm, or depression. Topical or oral carbonic anhydrase inhibitors are used to lower intraocular pressure by reducing aqueous production. Laser treatment of the trabecular meshwork in the anterior chamber angle improves aqueous outflow

from the eye. If medical or laser treatments fail to halt optic nerve damage from glaucoma, a filter must be constructed surgically (trabeculectomy) or a valve placed to release aqueous from the eye in a controlled fashion.

MACULAR DEGENERATION

This is a major cause of gradual, painless, bilateral central visual loss in the elderly. The old term, "senile macular degeneration," misinterpreted by many patients as an unflattering reference, has been replaced with "age-related macular degeneration." It occurs in a nonexudative (dry) form and an exudative (wet) form. Inflammation may be important in both forms of macular degeneration; recent genetic data indicate that susceptibility is associated with variants in the gene for complement factor H, an inhibitor of the alternative complement pathway. The nonexudative process begins with the accumulation of extracellular deposits called drusen underneath the retinal pigment epithelium. On ophthalmoscopy, they are pleomorphic but generally appear as small discrete yellow lesions clustered in the macula **(Fig. 28-16)**. With time they become larger, more numerous, and confluent. The retinal pigment epithelium becomes focally detached and atrophic, causing visual loss by interfering with photoreceptor function. Treatment with vitamins C and E, beta-carotene, and zinc may retard dry macular degeneration.

Figure 28-16

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Age-related macular degeneration begins with the accumulation of drusen within the macula. They appear as scattered yellow subretinal deposits.

Exudative macular degeneration, which develops in only a minority of patients, occurs when neovascular vessels from the choroid grow through defects in Bruch's membrane and proliferate underneath the retinal pigment epithelium or the retina. Leakage from these vessels produces elevation of the retina, with distortion (metamorphopsia) and blurring of vision. Although the onset of these symptoms is usually gradual, bleeding from a subretinal choroidal neovascular membrane sometimes causes acute visual loss. Neovascular membranes can be difficult to see on fundus examination because they are located beneath the retina. Fluorescein angiography and optical coherence tomography, a new technique for acquiring images of the retina in cross-section, are extremely useful for their detection. Major or repeated hemorrhage under the retina from neovascular membranes results in fibrosis, development of a round (disciform) macular scar, and permanent loss of central vision.

A major therapeutic advance has occurred recently with the discovery that exudative macular degeneration can be treated with intraocular injection of a vascular endothelial growth factor antagonist. Either bevacizumab or ranibizumab is administered by direct injection into the vitreous cavity, beginning on a monthly basis. These antibodies cause the regression of neovascular membranes by blocking the action of vascular endothelial growth factor, thereby improving visual acuity.

CENTRAL SEROUS CHORIORETINOPATHY

This primarily affects males between the ages of 20 and 50. Leakage of serous fluid from the choroid causes small, localized detachment of the retinal pigment epithelium and the neurosensory retina.

These detachments produce acute or chronic symptoms of metamorphopsia and blurred vision when the macula is involved. They are difficult to visualize with a direct ophthalmoscope because the detached retina is transparent and only slightly elevated. Diagnosis of central serous chorioretinopathy is made easily by fluorescein angiography, which shows dye streaming into the subretinal space. The cause of central serous chorioretinopathy is unknown. Symptoms may resolve spontaneously if the retina reattaches, but recurrent detachment is common. Laser photocoagulation has benefited some patients with this condition.

DIABETIC RETINOPATHY

A rare disease until 1921, when the discovery of insulin resulted in a dramatic improvement in life expectancy for patients with diabetes mellitus, diabetic retinopathy is now a leading cause of blindness in the United States. The retinopathy takes years to develop but eventually appears in nearly all cases. Regular surveillance of the dilated fundus is crucial for any patient with diabetes. In advanced diabetic retinopathy, the proliferation of neovascular vessels leads to blindness from vitreous hemorrhage, retinal detachment, and glaucoma (see Fig. 344-9). These complications can be avoided in most patients by administration of panretinal laser photocoagulation at the appropriate point in the evolution of the disease. **For further discussion of the manifestations and management of diabetic retinopathy, see Chap. 344.**

RETINITIS PIGMENTOSA

This is a general term for a disparate group of rod-cone dystrophies characterized by progressive night blindness, visual field constriction with a ring scotoma, loss of acuity, and an abnormal electroretinogram (ERG). It occurs sporadically or in an autosomal recessive, dominant, or X-linked pattern. Irregular black deposits of clumped pigment in the peripheral retina, called *bone spicules* because of their vague resemblance to the spicules of cancellous bone, give the disease its name **(Fig. 28-17)**. The name is actually a misnomer because retinitis pigmentosa is not an inflammatory process. Most cases are due to a mutation in the gene for rhodopsin, the rod photopigment, or in the gene for peripherin, a glycoprotein located in photoreceptor outer segments. Vitamin A (15,000 IU/d) slightly retards the deterioration of the ERG in patients with retinitis pigmentosa but has no beneficial effect on visual acuity or fields.

Figure 28-17

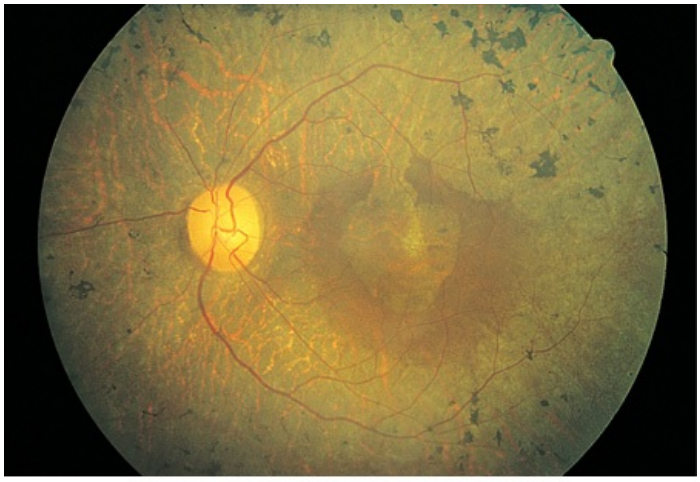

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Retinitis pigmentosa with black clumps of pigment in the retinal periphery known as "bone spicules." There is also atrophy of the retinal pigment epithelium, making the vasculature of the choroid easily visible.

Leber's congenital amaurosis, a rare cone dystrophy, has been treated by replacement of the missing RPE65 protein through gene therapy, resulting in modest improvement in visual function. Some forms of retinitis pigmentosa occur in association with rare, hereditary systemic diseases (olivopontocerebellar degeneration, Bassen-Kornzweig disease, Kearns-Sayre syndrome, Refsum's disease). Chronic treatment with chloroquine, hydroxychloroquine, and phenothiazines (especially thioridazine) can produce visual loss from a toxic retinopathy that resembles retinitis pigmentosa.

EPIRETINAL MEMBRANE

This is a fibrocellular tissue that grows across the inner surface of the retina, causing metamorphopsia and reduced visual acuity from distortion of the macula. A crinkled, cellophane-like membrane is visible on the retinal examination. Epiretinal membrane is most common in patients over 50 years of age and is usually unilateral. Most cases are idiopathic, but some occur as a result of hypertensive retinopathy,

diabetes, retinal detachment, or trauma. When visual acuity is reduced to the level of about 6/24 (20/80), vitrectomy and surgical peeling of the membrane to relieve macular puckering are recommended. Contraction of an epiretinal membrane sometimes gives rise to a *macular hole*. Most macular holes, however, are caused by local vitreous traction within the fovea. Vitrectomy can improve acuity in selected cases.

MELANOMA AND OTHER TUMORS

Melanoma is the most common primary tumor of the eye **(Fig. 28-18)**. It causes photopsia, an enlarging scotoma, and loss of vision. A small melanoma is often difficult to differentiate from a benign choroidal nevus. Serial examinations are required to document a malignant pattern of growth. Treatment of melanoma is controversial. Options include enucleation, local resection, and irradiation. *Metastatic tumors* to the eye outnumber primary tumors. Breast and lung carcinomas have a special propensity to spread to the choroid or iris. Leukemia and lymphoma also commonly invade ocular tissues. Sometimes their only sign on eye examination is cellular debris in the vitreous, which can masquerade as a chronic posterior uveitis. *Retrobulbar tumor* of the optic nerve (meningioma, glioma) or *chiasmal tumor* (pituitary adenoma, meningioma) produces gradual visual loss with few objective findings except for optic disc pallor. Rarely, sudden expansion of a pituitary adenoma from infarction and bleeding (*pituitary apoplexy*) causes acute retrobulbar visual loss, with headache, nausea, and ocular motor nerve palsies. In any patient with visual field loss or optic atrophy, CT or MR scanning should be considered if the cause remains unknown after careful review of the history and thorough examination of the eye.

Figure 28-18

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Melanoma of the choroid, appearing as an elevated dark mass in the inferior temporal fundus, just encroaching upon the fovea.

Proptosis

When the globes appear asymmetric, the clinician must first decide which eye is abnormal. Is one eye recessed within the orbit (*enophthalmos*), or is the other eye protuberant (*exophthalmos*, or *proptosis*)? A small globe or a Horner's syndrome can give the appearance of enophthalmos. True enophthalmos occurs commonly after trauma, from atrophy of retrobulbar fat, or from fracture of the orbital floor. The position of the eyes within the orbits is measured by using a Hertel exophthalmometer, a handheld instrument that records the position of the anterior corneal surface relative to the lateral orbital rim. If this instrument is not available, relative eye position can be judged by bending the patient's head forward and looking down upon the orbits. A proptosis of only 2 mm in one eye is detectable from this perspective. The development of proptosis implies a space-occupying lesion in the orbit and usually warrants CT or MR imaging.

GRAVES' OPHTHALMOPATHY

This is the leading cause of proptosis in adults (Chap. 341). The proptosis is often asymmetric and can even appear to be unilateral. Orbital inflammation and engorgement of the extraocular muscles, particularly the medial rectus and the inferior rectus, account for the protrusion of the globe. Corneal

exposure, lid retraction, conjunctival injection, restriction of gaze, diplopia, and visual loss from optic nerve compression are cardinal symptoms. Graves' ophthalmopathy is treated with oral prednisone (60 mg/d) for 1 month, followed by a taper over several months, topical lubricants, eyelid surgery, eye muscle surgery, or orbital decompression. Radiation therapy is not effective.

ORBITAL PSEUDOTUMOR

This is an idiopathic, inflammatory orbital syndrome that frequently is confused with Graves' ophthalmopathy. Symptoms are pain, limited eye movements, proptosis, and congestion. Evaluation for sarcoidosis, granulomatosis with polyangiitis (Wegener's), and other types of orbital vasculitis or collagen-vascular disease is negative. Imaging often shows swollen eye muscles (orbital myositis) with enlarged tendons. By contrast, in Graves' ophthalmopathy the tendons of the eye muscles usually are spared. The Tolosa-Hunt syndrome may be regarded as an extension of orbital pseudotumor through the superior orbital fissure into the cavernous sinus. The diagnosis of orbital pseudotumor is difficult. Biopsy of the orbit frequently yields nonspecific evidence of fat infiltration by lymphocytes, plasma cells, and eosinophils. A dramatic response to a therapeutic trial of systemic glucocorticoids indirectly provides the best confirmation of the diagnosis.

ORBITAL CELLULITIS

This causes pain, lid erythema, proptosis, conjunctival chemosis, restricted motility, decreased acuity, afferent pupillary defect, fever, and leukocytosis. It often arises from the paranasal sinuses, especially by contiguous spread of infection from the ethmoid sinus through the lamina papyracea of the medial orbit. A history of recent upper respiratory tract infection, chronic sinusitis, thick mucus secretions, or dental disease is significant in any patient with suspected orbital cellulitis. Blood cultures should be obtained, but they are usually negative. Most patients respond to empirical therapy with broad-spectrum IV antibiotics. Occasionally, orbital cellulitis follows an overwhelming course, with massive proptosis, blindness, septic cavernous sinus thrombosis, and meningitis. To avert this disaster, orbital cellulitis should be managed aggressively in the early stages, with immediate imaging of the orbits and antibiotic therapy that includes coverage of methicillin-resistant *Staphylococcus aureus* (MRSA). Prompt surgical drainage of an orbital abscess or paranasal sinusitis is indicated if optic nerve function deteriorates despite antibiotics.

TUMORS

Tumors of the orbit cause painless, progressive proptosis. The most common primary tumors are hemangioma, lymphangioma, neurofibroma, dermoid cyst, adenoid cystic carcinoma, optic nerve glioma, optic nerve meningioma, and benign mixed tumor of the lacrimal gland. Metastatic tumor to the orbit occurs frequently in breast carcinoma, lung carcinoma, and lymphoma. Diagnosis by fine-needle aspiration followed by urgent radiation therapy sometimes can preserve vision.

CAROTID CAVERNOUS FISTULAS

With anterior drainage through the orbit these fistulas produce proptosis, diplopia, glaucoma, and corkscrew, arterialized conjunctival vessels. Direct fistulas usually result from trauma. They are easily diagnosed because of the prominent signs produced by high-flow, high-pressure shunting. Indirect fistulas, or dural arteriovenous malformations, are more likely to occur spontaneously, especially in older women. The signs are more subtle, and the diagnosis frequently is missed. The combination of slight proptosis, diplopia, enlarged muscles, and an injected eye often is mistaken for thyroid ophthalmopathy. A bruit heard upon auscultation of the head or reported by the patient is a valuable diagnostic clue. Imaging shows an enlarged superior ophthalmic vein in the orbits. Carotid cavernous shunts can be eliminated by intravascular embolization.

Ptosis

BLEPHAROPTOSIS

This is an abnormal drooping of the eyelid. Unilateral or bilateral ptosis can be congenital, from dysgenesis of the levator palpebrae superioris, or from abnormal insertion of its aponeurosis into the eyelid. Acquired ptosis can develop so gradually that the patient is unaware of the problem. Inspection of old photographs is helpful in dating the onset. A history of prior trauma, eye surgery, contact lens use, diplopia, systemic symptoms (e.g., dysphagia or peripheral muscle weakness), or a family history of ptosis should be sought. Fluctuating ptosis that worsens late in the day is typical of myasthenia gravis. Examination should focus on evidence for proptosis, eyelid masses or deformities, inflammation, pupil inequality, or limitation of motility. The width of the palpebral fissures is measured in primary gaze to quantitate the degree of ptosis. The ptosis will be underestimated if the patient compensates by lifting the brow with the frontalis muscle.

MECHANICAL PTOSIS

This occurs in many elderly patients from stretching and redundancy of eyelid skin and subcutaneous fat (dermatochalasis). The extra weight of these sagging tissues causes the lid to droop. Enlargement or deformation of the eyelid from infection, tumor, trauma, or inflammation also results in ptosis on a purely mechanical basis.

APONEUROTIC PTOSIS

This is an acquired dehiscence or stretching of the aponeurotic tendon, which connects the levator muscle to the tarsal plate of the eyelid. It occurs commonly in older patients, presumably from loss of connective tissue elasticity. Aponeurotic ptosis is also a common sequela of eyelid swelling from infection or blunt trauma to the orbit, cataract surgery, or hard contact lens use.

MYOGENIC PTOSIS

The causes of *myogenic ptosis* include myasthenia gravis (Chap. 386) and a number of rare myopathies that manifest with ptosis. The term *chronic progressive external ophthalmoplegia* refers to a spectrum of systemic diseases caused by mutations of mitochondrial DNA. As the name implies, the most prominent findings are symmetric, slowly progressive ptosis and limitation of eye movements. In general, diplopia is a late symptom because all eye movements are reduced equally. In the *Kearns-Sayre* variant, retinal pigmentary changes and abnormalities of cardiac conduction develop. Peripheral muscle biopsy shows characteristic "ragged-red fibers." *Oculopharyngeal dystrophy* is a distinct autosomal dominant disease with onset in middle age, characterized by ptosis, limited eye movements, and trouble swallowing. *Myotonic dystrophy*, another autosomal dominant disorder, causes ptosis, ophthalmoparesis, cataract, and pigmentary retinopathy. Patients have muscle wasting, myotonia, frontal balding, and cardiac abnormalities.

NEUROGENIC PTOSIS

This results from a lesion affecting the innervation to either of the two muscles that open the eyelid: Müller's muscle or the levator palpebrae superioris. Examination of the pupil helps distinguish between these two possibilities. In Horner's syndrome, the eye with ptosis has a smaller pupil and the eye movements are full. In an oculomotor nerve palsy, the eye with the ptosis has a larger or a normal pupil. If the pupil is normal but there is limitation of adduction, elevation, and depression, a pupil-sparing oculomotor nerve palsy is likely (see next section). Rarely, a lesion affecting the small, central subnucleus of the oculomotor complex will cause bilateral ptosis with normal eye movements and pupils.

Double Vision (Diplopia)

The first point to clarify is whether diplopia persists in either eye after the opposite eye is covered. If it does, the diagnosis is monocular diplopia. The cause is usually intrinsic to the eye and therefore has no dire implications for the patient. Corneal aberrations (e.g., keratoconus, pterygium), uncorrected refractive error, cataract, or foveal traction may give rise to monocular diplopia. Occasionally it is a symptom of malingering or psychiatric disease. Diplopia alleviated by covering one eye is binocular diplopia and is caused by disruption of ocular alignment. Inquiry should be made into the nature of the double vision (purely side-by-side versus partial vertical displacement of images), mode of onset, duration, intermittency, diurnal variation, and associated neurologic or systemic symptoms. If the patient has diplopia while being examined, motility testing should reveal a deficiency corresponding to the patient's symptoms. However, subtle limitation of ocular excursions is often difficult to detect. For example, a patient with a slight left abducens nerve paresis may appear to have full eye movements despite a complaint of horizontal diplopia upon looking to the left. In this situation, the cover test provides a more sensitive method for demonstrating the ocular misalignment. It should be conducted in primary gaze and then with the head turned and tilted in each direction. In the above example, a cover test with the head turned to the right will maximize the fixation shift evoked by the cover test.

Occasionally, a cover test performed in an asymptomatic patient during a routine examination will reveal an ocular deviation. If the eye movements are full and the ocular misalignment is equal in all directions of gaze (concomitant deviation), the diagnosis is strabismus. In this condition, which affects about 1% of the population, fusion is disrupted in infancy or early childhood. To avoid diplopia, vision is suppressed from the nonfixating eye. In some children, this leads to impaired vision (amblyopia, or "lazy" eye) in the deviated eye.

Binocular diplopia results from a wide range of processes: infectious, neoplastic, metabolic, degenerative, inflammatory, and vascular. One must decide whether the diplopia is neurogenic in origin or is due to restriction of globe rotation by local disease in the orbit. Orbital pseudotumor, myositis, infection, tumor, thyroid disease, and muscle entrapment (e.g., from a blowout fracture) cause restrictive diplopia. The diagnosis of restriction is usually made by recognizing other associated signs and symptoms of local orbital disease in conjunction with imaging.

MYASTHENIA GRAVIS

(See also Chap. 386) This is a major cause of diplopia. The diplopia is often intermittent, variable, and not confined to any single ocular motor nerve distribution. The pupils are always normal. Fluctuating ptosis may be present. Many patients have a purely ocular form of the disease, with no evidence of systemic muscular weakness. The diagnosis can be confirmed by an IV edrophonium injection or by an assay for antiacetylcholine receptor antibodies. Negative results from these tests do not exclude the diagnosis. *Botulism* from food or wound poisoning can mimic ocular myasthenia.

After restrictive orbital disease and myasthenia gravis are excluded, a lesion of a cranial nerve

supplying innervation to the extraocular muscles is the most likely cause of binocular diplopia.

OCULOMOTOR NERVE

The third cranial nerve innervates the medial, inferior, and superior recti; inferior oblique; levator palpebrae superioris; and the iris sphincter. Total palsy of the oculomotor nerve causes ptosis, results in a dilated pupil, and leaves the eye "down and out" because of the unopposed action of the lateral rectus and superior oblique. This combination of findings is obvious. More challenging is the diagnosis of early or partial oculomotor nerve palsy. In this setting, any combination of ptosis, pupil dilation, and weakness of the eye muscles supplied by the oculomotor nerve may be encountered. Frequent serial examinations during the evolving phase of the palsy help ensure that the diagnosis is not missed. The advent of an oculomotor nerve palsy with a pupil involvement, especially when accompanied by pain, suggests a compressive lesion, such as a tumor or circle of Willis aneurysm. Neuroimaging should be obtained, along with a CT or MR angiogram. Occasionally, a catheter arteriogram must be done to exclude an aneurysm.

A lesion of the oculomotor nucleus in the rostral midbrain produces signs that differ from those caused by a lesion of the nerve itself. There is bilateral ptosis because the levator muscle is innervated by a single central subnucleus. There is also weakness of the contralateral superior rectus, because it is supplied by the oculomotor nucleus on the other side. Occasionally both superior recti are weak. Isolated nuclear oculomotor palsy is rare. Usually neurologic examination reveals additional signs that suggest brainstem damage from infarction, hemorrhage, tumor, or infection.

Injury to structures surrounding fascicles of the oculomotor nerve descending through the midbrain has given rise to a number of classic eponymic designations. In *Nothnagel's syndrome*, injury to the superior cerebellar peduncle causes ipsilateral oculomotor palsy and contralateral cerebellar ataxia. In *Benedikt's syndrome*, injury to the red nucleus results in ipsilateral oculomotor palsy and contralateral tremor, chorea, and athetosis. *Claude's syndrome* incorporates features of both of these syndromes, by injury to both the red nucleus and the superior cerebellar peduncle. Finally, in *Weber's syndrome*, injury to the cerebral peduncle causes ipsilateral oculomotor palsy with contralateral hemiparesis.

In the subarachnoid space the oculomotor nerve is vulnerable to aneurysm, meningitis, tumor, infarction, and compression. In cerebral herniation the nerve becomes trapped between the edge of the tentorium and the uncus of the temporal lobe. Oculomotor palsy also can result from midbrain torsion and hemorrhages during herniation. In the cavernous sinus, oculomotor palsy arises from carotid aneurysm, carotid cavernous fistula, cavernous sinus thrombosis, tumor (pituitary adenoma, meningioma, metastasis), herpes zoster infection, and the Tolosa-Hunt syndrome.

The etiology of an isolated, pupil-sparing oculomotor palsy often remains an enigma even after neuroimaging and extensive laboratory testing. Most cases are thought to result from microvascular infarction of the nerve somewhere along its course from the brainstem to the orbit. Usually the patient complains of pain. Diabetes, hypertension, and vascular disease are major risk factors. Spontaneous recovery over a period of months is the rule. If this fails to occur or if new findings develop, the diagnosis of microvascular oculomotor nerve palsy should be reconsidered. Aberrant regeneration is common when the oculomotor nerve is injured by trauma or compression (tumor, aneurysm). Miswiring of sprouting fibers to the levator muscle and the rectus muscles results in elevation of the eyelid upon downgaze or adduction. The pupil also constricts upon attempted adduction, elevation, or depression of the globe. Aberrant regeneration is not seen after oculomotor palsy from microvascular infarct and hence vitiates that diagnosis.

TROCHLEAR NERVE

The fourth cranial nerve originates in the midbrain, just caudal to the oculomotor nerve complex. Fibers exit the brainstem dorsally and cross to innervate the contralateral superior oblique. The principal actions of this muscle are to depress and intort the globe. A palsy therefore results in hypertropia and excyclotorsion. The cyclotorsion seldom is noticed by patients. Instead, they complain of vertical diplopia, especially upon reading or looking down. The vertical diplopia also is exacerbated by tilting the head toward the side with the muscle palsy and alleviated by tilting it away. This "head tilt test" is a cardinal diagnostic feature.

Isolated trochlear nerve palsy results from all the causes listed above for the oculomotor nerve except aneurysm. The trochlear nerve is particularly apt to suffer injury after closed head trauma. The free edge of the tentorium is thought to impinge on the nerve during a concussive blow. Most isolated trochlear nerve palsies are idiopathic and hence are diagnosed by exclusion as "microvascular." Spontaneous improvement occurs over a period of months in most patients. A base-down prism (conveniently applied to the patient's glasses as a stick-on Fresnel lens) may serve as a temporary measure to alleviate diplopia. If the palsy does not resolve, the eyes can be realigned by weakening the inferior oblique muscle.

ABDUCENS NERVE

The sixth cranial nerve innervates the lateral rectus muscle. A palsy produces horizontal diplopia, worse on gaze to the side of the lesion. A nuclear lesion has different consequences, because the

abducens nucleus contains interneurons that project via the medial longitudinal fasciculus to the medial rectus subnucleus of the contralateral oculomotor complex. Therefore, an abducens nuclear lesion produces a complete lateral gaze palsy from weakness of both the ipsilateral lateral rectus and the contralateral medial rectus. *Foville's syndrome* after dorsal pontine injury includes lateral gaze palsy, ipsilateral facial palsy, and contralateral hemiparesis incurred by damage to descending corticospinal fibers. *Millard-Gubler syndrome* from ventral pontine injury is similar except for the eye findings. There is lateral rectus weakness only, instead of gaze palsy, because the abducens fascicle is injured rather than the nucleus. Infarct, tumor, hemorrhage, vascular malformation, and multiple sclerosis are the most common etiologies of brainstem abducens palsy.

After leaving the ventral pons, the abducens nerve runs forward along the clivus to pierce the dura at the petrous apex, where it enters the cavernous sinus. Along its subarachnoid course it is susceptible to meningitis, tumor (meningioma, chordoma, carcinomatous meningitis), subarachnoid hemorrhage, trauma, and compression by aneurysm or dolichoectatic vessels. At the petrous apex, mastoiditis can produce deafness, pain, and ipsilateral abducens palsy (*Gradenigo's syndrome*). In the cavernous sinus, the nerve can be affected by carotid aneurysm, carotid cavernous fistula, tumor (pituitary adenoma, meningioma, nasopharyngeal carcinoma), herpes infection, and Tolosa-Hunt syndrome.

Unilateral or bilateral abducens palsy is a classic sign of raised intracranial pressure. The diagnosis can be confirmed if papilledema is observed on fundus examination. The mechanism is still debated but probably is related to rostral-caudal displacement of the brainstem. The same phenomenon accounts for abducens palsy from low intracranial pressure (e.g., after lumbar puncture, spinal anesthesia, or spontaneous dural cerebrospinal fluid leak).

Treatment of abducens palsy is aimed at prompt correction of the underlying cause. However, the cause remains obscure in many instances despite diligent evaluation. As was mentioned above for isolated trochlear or oculomotor palsy, most cases are assumed to represent microvascular infarcts because they often occur in the setting of diabetes or other vascular risk factors. Some cases may develop as a postinfectious mononeuritis (e.g., after a viral flu). Patching one eye or applying a temporary prism will provide relief of diplopia until the palsy resolves. If recovery is incomplete, eye muscle surgery nearly always can realign the eyes, at least in primary position. A patient with an abducens palsy that fails to improve should be reevaluated for an occult etiology (e.g., chordoma, carcinomatous meningitis, carotid cavernous fistula, myasthenia gravis). Skull base tumors are easily missed even on contrast-enhanced neuroimaging studies.

MULTIPLE OCULAR MOTOR NERVE PALSIES

These should not be attributed to spontaneous microvascular events affecting more than one cranial nerve at a time. This remarkable coincidence does occur, especially in diabetic patients, but the diagnosis is made only in retrospect after all other diagnostic alternatives have been exhausted. Neuroimaging should focus on the cavernous sinus, superior orbital fissure, and orbital apex, where all three ocular motor nerves are in close proximity. In a diabetic or immunocompromised host, fungal infection (*Aspergillus*, Mucorales, *Cryptococcus*) is a common cause of multiple nerve palsies. In a patient with systemic malignancy, carcinomatous meningitis is a likely diagnosis. Cytologic examination may be negative despite repeated sampling of the cerebrospinal fluid. The cancer-associated Lambert-Eaton myasthenic syndrome also can produce ophthalmoplegia. Giant cell (temporal) arteritis occasionally manifests as diplopia from ischemic palsies of extraocular muscles. Fisher's syndrome, an ocular variant of Guillain-Barré, produces ophthalmoplegia with areflexia and ataxia. Often the ataxia is mild, and the reflexes are normal. Antiganglioside antibodies (GQ1b) can be detected in about 50% of cases.

SUPRANUCLEAR DISORDERS OF GAZE

These are often mistaken for multiple ocular motor nerve palsies. For example, Wernicke's encephalopathy can produce nystagmus and a partial deficit of horizontal and vertical gaze that mimics a combined abducens and oculomotor nerve palsy. The disorder occurs in malnourished or alcoholic patients and can be reversed by thiamine. Infarct, hemorrhage, tumor, multiple sclerosis, encephalitis, vasculitis, and Whipple's disease are other important causes of supranuclear gaze palsy. Disorders of vertical gaze, especially downward saccades, are an early feature of progressive supranuclear palsy. Smooth pursuit is affected later in the course of the disease. Parkinson's disease, Huntington's disease, and olivopontocerebellar degeneration also can affect vertical gaze.

The *frontal eye field* of the cerebral cortex is involved in generation of saccades to the contralateral side. After hemispheric stroke, the eyes usually deviate toward the lesioned side because of the unopposed action of the frontal eye field in the normal hemisphere. With time, this deficit resolves. Seizures generally have the opposite effect: the eyes deviate conjugately away from the irritative focus. *Parietal lesions* disrupt smooth pursuit of targets moving toward the side of the lesion. Bilateral parietal lesions produce *Bálint's syndrome*, which is characterized by impaired eye-hand coordination (optic ataxia), difficulty initiating voluntary eye movements (ocular apraxia), and visuospatial disorientation (simultanagnosia).

HORIZONTAL GAZE

Descending cortical inputs mediating horizontal gaze ultimately converge at the level of the pons. Neurons in the paramedian pontine reticular formation are responsible for controlling conjugate gaze toward the same side. They project directly to the ipsilateral abducens nucleus. A lesion of either the paramedian pontine reticular formation or the abducens nucleus causes an ipsilateral conjugate gaze palsy. Lesions at either locus produce nearly identical clinical syndromes, with the following exception: vestibular stimulation (oculocephalic maneuver or caloric irrigation) will succeed in driving the eyes conjugately to the side in a patient with a lesion of the paramedian pontine reticular formation but not in a patient with a lesion of the abducens nucleus.

Internuclear Ophthalmoplegia

This results from damage to the medial longitudinal fasciculus ascending from the abducens nucleus in the pons to the oculomotor nucleus in the midbrain (hence, "internuclear"). Damage to fibers carrying the conjugate signal from abducens interneurons to the contralateral medial rectus motoneurons results in a failure of adduction on attempted lateral gaze. For example, a patient with a left internuclear ophthalmoplegia (INO) will have slowed or absent adducting movements of the left eye **(Fig. 28-19)**. A patient with bilateral injury to the medial longitudinal fasciculus will have bilateral INO. Multiple sclerosis is the most common cause, although tumor, stroke, trauma, or any brainstem process may be responsible. *One-and-a-half syndrome* is due to a combined lesion of the medial longitudinal fasciculus and the abducens nucleus on the same side. The patient's only horizontal eye movement is abduction of the eye on the other side.

Figure 28-19

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

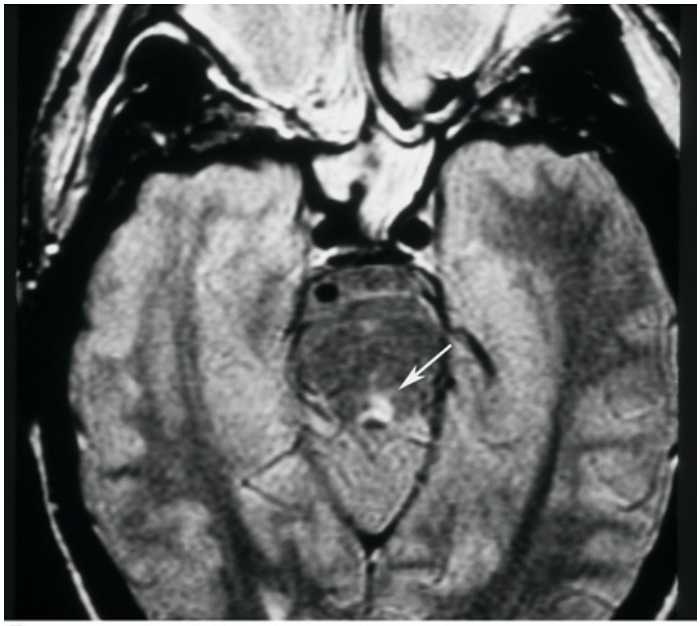

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left internuclear ophthalmoplegia (INO). ***A.*** In primary position of gaze the eyes appear normal. ***B.*** Horizontal gaze to the left is intact. ***C.*** On attempted horizontal gaze to the right, the left eye fails to adduct. In mildly affected patients the eye may adduct partially or more slowly than normal. Nystagmus is usually present in the abducted eye. ***D.*** T2-weighted axial MRI image through the pons showing a demyelinating plaque in the left medial longitudinal fasciculus (*arrow*).

VERTICAL GAZE

This is controlled at the level of the midbrain. The neuronal circuits affected in disorders of vertical gaze are not fully elucidated, but lesions of the rostral interstitial nucleus of the medial longitudinal fasciculus and the interstitial nucleus of Cajal cause supranuclear paresis of upgaze, downgaze, or all vertical eye movements. Distal basilar artery ischemia is the most common etiology. *Skew deviation* refers to a vertical misalignment of the eyes, usually constant in all positions of gaze. The finding has poor localizing value because skew deviation has been reported after lesions in widespread regions of the brainstem and cerebellum.

Parinaud's Syndrome

Also known as dorsal midbrain syndrome, this is a distinct supranuclear vertical gaze disorder caused by damage to the posterior commissure. It is a classic sign of hydrocephalus from aqueductal stenosis. Pineal region tumors, cysticercosis, and stroke also cause Parinaud's syndrome. Features include loss of upgaze (and sometimes downgaze), convergence-retraction nystagmus on attempted upgaze, downward ocular deviation ("setting sun" sign), lid retraction (Collier's sign), skew deviation, pseudoabducens palsy, and light-near dissociation of the pupils.

NYSTAGMUS

This is a rhythmic oscillation of the eyes, occurring physiologically from vestibular and optokinetic stimulation or pathologically in a wide variety of diseases (Chap. 21). Abnormalities of the eyes or optic nerves, present at birth or acquired in childhood, can produce a complex, searching nystagmus with irregular pendular (sinusoidal) and jerk features. This nystagmus is commonly referred to as *congenital sensory nystagmus*. This is a poor term because even in children with congenital lesions, the nystagmus does not appear until several months of age. *Congenital motor nystagmus* , which looks similar to congenital sensory nystagmus, develops in the absence of any abnormality of the sensory visual system. Visual acuity also is reduced in congenital motor nystagmus, probably by the nystagmus itself, but seldom below a level of 20/200.

Jerk Nystagmus

This is characterized by a slow drift off the target, followed by a fast corrective saccade. By convention, the nystagmus is named after the quick phase. Jerk nystagmus can be downbeat, upbeat, horizontal (left or right), and torsional. The pattern of nystagmus may vary with gaze position. Some patients will be oblivious to their nystagmus. Others will complain of blurred vision or a subjective to-and-fro movement of the environment (oscillopsia) corresponding to the nystagmus. Fine nystagmus may be difficult to see on gross examination of the eyes. Observation of nystagmoid movements of the optic disc on ophthalmoscopy is a sensitive way to detect subtle nystagmus.

Gaze-Evoked Nystagmus

This is the most common form of jerk nystagmus. When the eyes are held eccentrically in the orbits, they have a natural tendency to drift back to primary position. The subject compensates by making a corrective saccade to maintain the deviated eye position. Many normal patients have mild gaze-evoked nystagmus. Exaggerated gaze-evoked nystagmus can be induced by drugs (sedatives, anticonvulsants, alcohol); muscle paresis; myasthenia gravis; demyelinating disease; and cerebellopontine angle, brainstem, and cerebellar lesions.

Vestibular Nystagmus

Vestibular nystagmus results from dysfunction of the labyrinth (Ménière's disease), vestibular nerve, or vestibular nucleus in the brainstem. Peripheral vestibular nystagmus often occurs in discrete attacks, with symptoms of nausea and vertigo. There may be associated tinnitus and hearing loss. Sudden shifts in head position may provoke or exacerbate symptoms.

Downbeat Nystagmus

Downbeat nystagmus results from lesions near the craniocervical junction (Chiari malformation, basilar invagination). It also has been reported in brainstem or cerebellar stroke, lithium or anticonvulsant intoxication, alcoholism, and multiple sclerosis. *Upbeat nystagmus* is associated with damage to the pontine tegmentum from stroke, demyelination, or tumor.

OPSOCLONUS

This rare, dramatic disorder of eye movements consists of bursts of consecutive saccades (saccadomania). When the saccades are confined to the horizontal plane, the term *ocular flutter* is preferred. It can result from viral encephalitis, trauma, or a paraneoplastic effect of neuroblastoma, breast carcinoma, and other malignancies. It has also been reported as a benign, transient phenomenon in otherwise healthy patients.

FURTHER READINGS

Albert DM et al (eds): *Albert and Jakobiec's Principles and Practice of Ophthalmology* , 3rd ed. Philadelphia, Saunders, 2008

Cheung N et al: Diabetic retinopathy. Lancet 376:124, 2010 [PMID: 20580421]

D'amico DJ: Clinical practice: Primary retinal detachment. N Engl J Med 359:2346, 2008 [PMID: 21607619]

Jager RD et al: Age-related macular degeneration. N Engl J Med 358:2606, 2008 [PMID: 18550876]

Maguire AM et al: Safety and efficacy of gene transfer for Leber's congenital amaurosis. N Engl J Med 358:2240, 2008 [PMID: 18441370]

Mancuso K et al: Gene therapy for red-green colour blindness in adult primates. Nature 461:784, 2009 [PMID: 19759534]

Optic Neuritis Study Group: Multiple sclerosis risk after optic neuritis: Final optic neuritis treatment trial follow-up. Arch Neurol 65:1545, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VIDEO LIBRARY OF NEURO-OPHTHALMOLOGY

The proper control of eye movements requires the coordinated activity of many different anatomic structures in the peripheral and central nervous system, and in turn manifestations of a diverse array of neurological and medical disorders are revealed as disorders of eye movement. In this remarkable video collection, an introduction to distinctive eye movement disorders encountered in the context of neuromuscular, paraneoplastic, demyelinating, neurovascular and neurodegenerative disorders is presented.

Cases with Multiple Sclerosis

Video e11-1
Fisher's One and a Half Syndrome (ID164-2)
Video e11-2
A Case of Ocular Flutter (ID166-2)
Video e11-3
Downbeat Nystagmus and Periodic Alternating Nystagmus (ID168-6)
Video e11-4
Bilateral Internuclear Ophthalmoplegia (ID933-1)

Video e11-1	
Metadata Element Field	**Metadata**
Identifier	164-2
Title	Fisher's One-and-a-Half Syndrome
Ocular Movements	Unilateral Internuclear Ophthalmoplegia; Unilateral Horizontal Gaze Palsy; Upbeat Nystagmus on Upgaze; Primary Position Upbeat Nystagmus; Convergence Normal;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Unilateral Internuclear Ophthalmoplegia; Unilateral Horizontal Gaze Palsy; Upbeat Nystagmus on Upgaze; Primary Position Upbeat Nystagmus; Convergence Normal; Fisher's One-and-a-Half Syndrome; Multiple Sclerosis;
Presenting Symptom	Diplopia
History	This young patient presented with double vision and was found to have on examination the classic **findings of Fisher's one-and-a-half syndrome, which are:**

	• Right internuclear ophthalmoplegia on gaze left with adduction weakness OD • Right horizontal gaze paresis with gaze-evoked nystagmus • Full vertical gaze
	• In primary position, vertical upbeat nystagmus • Large-amplitude upbeat nystagmus on upgaze • Convergence normal **Neurologic Examination:** Normal **Diagnosis:** Multiple Sclerosis
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2008
Date Original	1973
Resource.Type Format.Creation	Video
Source	16-mm Film
Language	En
Relation-Is Part Of	166-14, 166-20, 937-8, 941-4, 941-6
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist
Reviewer	David Zee, MD, Johns Hopkins Hospital, 2006
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	This MS patient **had Fisher's One-and-a-Half Syndrome** characterized by: • A left horizontal gaze palsy with gaze-evoked nystagmus • A left internuclear ophthalmoplegia on gaze right with adduction weakness OS • Abducting nystagmus OD • Full vertical gaze • Upbeat nystagmus on upgaze (ill sustained) • Convergence normal The **"One-and-a-Half" syndrome** was first described by **Fisher** in **1967**. The disorder is characterized by a lateral gaze palsy in one direction with an internuclear ophthalmoplegia (INO) in the other direction. In the complete form of the syndrome, the ipsilateral eye lies fixed at the midline for all lateral movements; the other eye can only abduct and exhibits horizontal jerk nystagmus in abduction. The syndrome is usually due to a unilateral lesion in the lower part of the dorsal pontine tegmentum affecting the ipsilateral paramedian pontine reticular formation (PPRF), the abducens nucleus, and internuclear fibers of the ipsilateral medial longitudinal fasciculus (MLF). These internuclear fibers originate in the contralateral abducens nucleus and terminate in the ipsilateral medial rectus subnucleus of the third nerve nucleus. In **1983 Michael Wall**, who was a Fellow of mine at that time, reviewed all the patients whom I had seen with the one-and-a-half syndrome between 1968 and 1982. Films or video tapes of eye movements were available for review in 9 cases. In 12 of 20 patients, the lesion was on the right side. The patients ranged in age from 16 to 78 years. Prior to our publication in Neurology there had been only 29 reported cases of the

	one-and-a-half syndrome in the literature: 6 proven and 6 probable infarctions, 8 hemorrhages (1 traumatic), 5 gliomas, 1 arteriovenous malformation, 1 metastatic melanoma, and 2 patients with MS. We added a further 20 patients. **Brainstem infarction** was the most common cause in elderly adults (patients 17 to 20). Four patients were age 61 to 78 years (mean, 70 years). Two of them had hypertension. Three patients presented with neurologic complaints and one with diplopia. **MS** was the most common cause in 14 young adults age 18 to 52 years (mean, 32 years). Twelve had definite MS and 2 had possible MS.
	In 4 patients the one-and-a-half syndrome represented the initial presentation of MS. Two of these patients complained first of diplopia and one of blurred vision. Only 7 of 14 MS patients had visual complaints as the major symptom. Nine had diplopia, 5 had blurred vision, 3 had oscillopsia, and 2 had difficulty looking to one side. **Associated ocular motility signs:** Gaze-evoked upbeat nystagmus Skew deviation Horizontal ipsilateral gaze nystagmus Rotary component to horizontal ipsilateral gaze nystagmus Spontaneous nystagmus to the contralateral side Absent or impaired convergence Saccadic vertical pursuit Gaze-evoked downbeat nystagmus Impaired upward gaze Exotropia Esotropia Orthotropic In 10 MS patients whose eye position was documented in straight-ahead gaze, 4 had exotropia, 2 had an esotropia, and 4 were orthotropic. **Associated Neurologic Signs:** The most common associated cranial nerve lesions were an ipsilateral trigeminal sensory loss and a peripheral-type facial palsy.
	Seven patients had impaired coordination of the limbs. Only one patient had no neurologic abnormality other than the one-and-a-half syndrome. Of the 14 MS patients, 10 had recovery of eye movements: Recovery was complete in 8 patients between 4 and 16 weeks after onset, and partial recovery occurred in 2 other patients followed for 4 weeks and 9 weeks.
Neuroimaging	No neuroimaging studies were available on this patient.
Anatomy	The two constituents of the one-and-a-half syndrome (ipsilateral horizontal gaze palsy and INO) can be analyzed anatomically to provide the basis for topographic localization of the lesion at the bedside. **Horizontal Gaze Palsy:** There are four theoretical possibilities to account for the ipsilateral horizontal gaze palsy. It may be due to a single unilateral lesion affecting: 1. The ipsilateral **paramedial pontine reticular formation(PPRF)** only; 2. The ipsilateral abducens nucleus alone; 3. Both the ipsilateral paramedial pontine reticular formation **(PPRF)** and the **abducens nucleus**; or, when two lesions are involved 4. The motoneuron root fibers of the ipsilateral **abducens nucleus** to the lateral rectus and the contralateral **medial longitudinal fasciculus (MLF)**. **Paramedial Pontine Reticular Formation:** The medial portions of the nucleus reticularis magnocellularis (or nucleus centralis pontis oralis and caudalis) have been designated the "paramedian pontine reticular formation" **(PPRF)**, rostral to the abducens nucleus. The region extends from the abducens nucleus in a rostral direction toward the brachium conjunctivum and trochlear nucleus. It has been defined functionally because there are no

distinct histologic boundaries. Anatomically, **Graybiel, Büttner-Ennever, and Grantyn and colleagues** all showed inputs from discrete areas.

Two major oculomotor pathways originate from the PPRF. One pathway ascends rostrally, close to but outside of the MLF, to the rostral mesencephalon to coordinate horizontal and vertical gaze. The other pathway descends caudally and sends off a direct projection to the ipsilateral abducens nucleus.

This excitatory pathway controls horizontal gaze to the ipsilateral side by stimulating motoneurons to the lateral rectus and internuclear abducens neurons, which project via the MLF to the medial rectus motoneurons of the contralateral oculomotor nucleus. Other major projections likely to be involved in gaze are projections from the PPRF to the nucleus prepositus hypoglossi and PPRF cerebellar pathways.

No direct inhibitory pathway from the PPRF projects to the contralateral abducens nucleus. The inhibitory pathway from inhibitory burst neurons travels from the dorsomedial gigantocellular tegmental field just caudal to the abducens nucleus, crosses the brainstem, and terminates in the contralateral abducens nucleus.

The importance of this pathway has been demonstrated for saccades and quick phases of vestibular nystagmus, but it is still unclear whether an analogous pathway exists to control smooth pursuit movements. In any case, the PPRF generates all ipsilateral-directed saccades (including the quick phases of vestibular nystagmus).

Abducens nucleus. The abducens nucleus contains typical motoneurons that give rise to root fibers that innervate the lateral rectus as well as internuclear neurons whose axons cross the midline and ascend via the contralateral MLF to the medial rectus subnucleus in the contralateral oculomotor (third nerve) nucleus. This projection is the main excitatory input to medial rectus motoneurons in lateral gaze.

An old clinicopathologic case published by **Bennett and Savill** described a unilateral gaze palsy with an associated "patch of softening" in the abducens nucleus without apparent involvement of neighboring structures.

Clinical distinction PPRF:

Abducens nucleus. At the bedside distinction can be made between the manifestations of gaze palsies in lesions of the PPRF in the upper pons from those of the PPRF in the lower pons at the level of the abducens nucleus.

With **PPRF lesions rostral to the abducens**, there is ipsilateral paralysis of saccades and pursuit, but the eyes can be driven to the side of the gaze palsy with vestibular stimulation.

At the level of the abducens nucleus, lesions of the PPRF are associated with ipsilateral gaze palsy and loss of reflex vestibular (and tonic neck) movements. This presumes that there is a critical synapse within the caudal PPRF for the vestibulo-ocular pathways or that, at the very least, the functional integrity of the PPRF at that level is necessary for vestibulo-ocular eye movements.

An associated sign with **abducens nuclear lesions**, however, is an **ipsilateral "peripheral" facial palsy** because the genu of the seventh nerve wraps around the abducens nucleus. The gustatory fibers are spared because these fibers are carried in the intermediate branch of the facial nerve to the nucleus solitarius of the medulla. This sign was present in four of our MS patients.

An identical ipsilateral gaze palsy can be produced by damage to axons of abducens neurons as they course through the brainstem—namely, the ipsilateral sixth nerve fascicle and those axons that ascend the contralateral MLF. Separation of these two anatomic sites suggests two lesions. A lateral gaze palsy that always remains conjugate is consistent with one lesion of the abducens nucleus, whereas one that is not conjugate at any time would better fit two lesions. Two examples, in our patient group, were both patients with definite MS.

Internuclear Ophthalmoplegia:

Both human and experimental data have established the localization of the lesion that causes INO. A unilateral INO is due to the interruption of the ipsilateral MLF after it has crossed the midline caudally in the pons from its site of origin in the contralateral abducens nucleus.

Clinically, the syndrome is characterized by:

1. Paresis or paralysis of adduction of the ipsilateral eye on attempted horizontal gaze to the contralateral side;

2. Horizontal jerk nystagmus in the contralateral abducting eye; and

3. Typically convergence is intact if the lesion does not extend to the mesencephalon.

	Other associated findings are abnormalities in vertical smooth pursuit, OKN, the vertical VOR with normal vertical saccades if the INO is bilateral, and gaze-evoked vertical nystagmus on upward gaze more frequent than downgaze if the lesion is bilateral and skew deviation. The one-had-a-half syndrome is characterized by, on horizontal gaze: 1. **An ipsilateral gaze paresis or palsy;** 2. An INO on contralateral gaze; and 3. At rest, an **exotropia of the eye contralateral to the lesion** in the acute phase, or no deviation at rest, or, less commonly, an esotropia of the eye ipsilateral to the lesion resulting from paresis of the sixth nerve. Occasionally, ocular **bobbing occurs**; this may be periodic or limited to the ipsilateral eye. Anisocoria is also reported. **Personal Observation:** All of our patients with the one-and-a-half syndrome showed a complete or partial ipsilateral horizontal gaze palsy and a clinical or subclinical INO on contralateral gaze. Slow and incomplete adduction of the ipsilateral eye was attributed to deficient medial rectus excitation, and horizontal jerk nystagmus in the abducting eye was attributed to deficient tonic inhibition of the medial rectus. This interpretation is consistent with the results of oculographic studies. The association of **exotropia in the one-and-a-half syndrome** was observed by **Fisher** and later termed "**paralytic pontine exotropia**" by **Sharpe and colleagues**. In this distinctive supranuclear syndrome, the deviated, exotropic eye shows abduction nystagmus during attempts to move it further laterally, and there is extreme slowness of adduction saccades when the eye is used to fixate to move it to the midline. **Paralytic pontine exotropia** is attributed to tonic contralateral ocular deviation of the eyes, which implies acute ipsilateral involvement of the PPRF. Failure of the ipsilateral eye to deviate medially is explained by the INO. Three autopsy cases of paralytic pontine exotropia confirm the lesion site. **Esotropia** of the ipsilateral eye may be due to a lesion of the fasciculus of the **ipsilateral sixth nerve.**
Pathology	The anatomic localization of the lesion in the one-and-a-half syndrome has been confirmed at autopsy in seven patients. Six of them had a single unilateral lesion in the pontine tegmentum ipsilateral to the gaze palsy involving the PPRF and the ipsilateral MLF. The abducens nucleus was spared by discrete lesions and involved in extensive lesions resulting from infarction, hemorrhage, or glioma. In **Fisher's** case, extensive pontine infarction involved both the PPRF and the abducens nucleus. **Crevits and colleagues** correlated the gaze palsy with a single discrete infarct 3 by 2 mm in diameter in the ipsilateral PPRF and MLF. The lower fascicles of the ipsilateral sixth nerve passed through the necrotic area. This was probably the smallest lesion associated with the one-and-a-half syndrome. **Newman and colleagues** reported a similar clinical case, but they found an ipsilateral PPRF lesion and "ischemic necrosis in the region of the abducens nucleus, although individual neurons could be identified." Partial damage to the contralateral PPRF was also found. In another pathologically confirmed case evaluated clinically by electro-oculography, a hypertensive hemorrhage in the rostral pontine tegmentum had spread into the right basis pontis to destroy the ipsilateral PPRF and abducens nucleus.
Etiology	Review (Wall, Wray, 1983)
Treatment	
Disease/Diagnosis	Multiple Sclerosis
Further Readings	Bennett H, Savill TH: A case of permanent conjugate deviation of the eyes and head, the result of a lesion limited to the sixth nucleus, with remarks on associated lateral movements of the eyeballs, and rotation of the head and neck. Brain 12:102, 1889 Bogousslavsky J et al: One-and-a-half syndrome in ischemic locked-in state. J Neurol Neurosurg Psychiatry 47:927, 1984 Carter JE, Rauch RA: One-and-a-half syndrome type II. Arch Neurol 51:87, 1994 Crevits L et al: Paralytic pontine exotropia in subarachnoid hemorrhage: A

clinocopathological correlation. Clin Neurol Neurosurg 78:269, 1975

Fisher CM: Some neuro-ophthalmological observations. J Neurol Neurosurg Psychiatry 30:383, 1967

———: Neuroanatomic evidence to explain why bilateral internuclear ophthalmoplegia may result from occlusion of a unilateral pontine branch artery. J Neuroophthalmol 24:39, 2004

Jackel RA et al: Metastatic adenocarcinoma presenting as a one-and-a-half syndrome. J Clin Neuroopthalmol 6:116, 1986

Kataoka S et al: Paramedian pontine infarction. Neurological/topographical correlation. Stroke 28:809, 1997

Minagar A et al: Case report: one-and-a-half syndrome and tuberculosis of the pons in a patient with AIDS. AIDS Patient Care STDS 14:461, 2000

Newman NM et al: Paralytic pontine exotropia a case report with clinicopathologic confirmation. Augenbewegungastörungen Neurophysiologie und Klinik, München JF Bergman Verlag, 1978.

Newton HB, Miner ME: One-and-a-half syndrome after resection of a midline cerebellar astrocytoma: Case report and discussion of the literature. Neurosurgery 29:768, 1991

Oommen KJ et al: Pontine hemorrhage causing Fisher one-and-a-half syndrome with facial paralysis. J Clin Neuroophthalmol 2:129, 1982

Raps EC et al: Isolated one-and-a-half syndrome with pontine cavernous angioma; successful surgical removal. J Clin Neuroophthalmol 10:287, 1990

Sharpe JA et al: Paralytic pontine exotropia. A sign of acute unilateral pontine gaze palsy and internuclear ophthalmoplegia. Neurology 24:1076, 1974

Smith MS et al: One-and-a-half syndrome. Occurrence after trauma with computerized tomographic correlation. Arch Neurol 37:251, 1980

Wall M, Wray SH: The one-and-a-half syndrome: A unilateral lesion of the pontine tegmentum. A study of 20 cases and review of the literature. Neurology 33:971, 1983

Yigit A et al: The one-and-a-half syndrome in systemic lupus erythematosus. J Neuroophthalmol 16:274, 1996

Video e11-2

Metadata Element Field	Metadata
Identifier	166-2
Title	Ocular Flutter
Ocular Movements	Ocular Flutter;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Ocular Flutter; Multiple Sclerosis;
Presenting Symptom	Blurred Vision
History	This patient was seen in the Neurovisual Clinic for evaluation of a monocular attack of optic neuritis that completely resolved. Six months later, she became unsteady walking and was found to have bilateral ataxia. She was given a diagnosis of multiple sclerosis (MS). **Neuro-Ophthalmologic Examination:** The examination was normal apart from the eye movements, which showed, in central gaze, episodic bursts of horizontal back-to-back saccades without an intersaccadic interval characteristic of ocular flutter. The patient was seen prior to the availability of neuroimaging by MRI. **In 1954 Cogan** first used the term *ocular flutter* to describe a rare disorder of horizontal eye movements characterized by rapid bursts of synchronous back-to-back horizontal oscillatory movements usually seen in the primary position of gaze. Since then, there have been more than 50 reports, usually single cases or small series, linking the phenomenon to a wide variety of brainstem and cerebellar

	conditions, e.g., post–enteroviral infection, cerebral malaria, cyclosporine treatment, and meningitis, but perhaps most frequently associated with parainfectious states or, with opsoclonus, as a paraneoplastic manifestation of occult malignancy.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2008
Date Original	1978
Resource.Type	Video
Format.Creation Source	16-mm Film
Language	En
Relation-Is Part Of	162-4, 166-1,931-1,936-7, 936-8
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist
Reviewer	
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	• Horizontal saccadic oscillations are present without an intersaccadic interval; • The frequency of oscillations is usually high, typically 10–25 cycles per second, as evident in this case; • Ocular flutter is intermittent and mainly associated with voluntary saccades: ***flutter dysmetria;*** and • When the amplitude is very small, the oscillation can be detected only with an ophthalmoscope or on eye movement recordings, and then the term ***microflutter*** is used to describe the disorder. **Ocular flutter** occurs in: 1. Paraneoplastic opsoclonus/flutter due to an occult neoplasm; 2. MS; 3. Side effects of drugs: lithium, amitriptyline, cocaine, and phenytoin with diazepam; 4. Toxins: chlordecone, thallium, strychnine, toluene, and organophosphates; or 5. Complication of pregnancy In three reported cases from Japan of the opsoclonus-myoclonus syndrome (OMS) during pregnancy, the neurologic symptoms were entirely similar to those of OMS unrelated to pregnancy with opsoclonus/flutter, trunkal ataxia, and myoclonic jerks of the neck and limbs. The condition of the fetuses were good, except in the one case that miscarried. In all three cases the OMS occurred in the middle to late stages of pregnancy. Whether the symptoms improved because the pregnancy ended or because of corticosteroid therapy remains unclear. However, the OMS gradually improved after spontaneous miscarriage in one case. Taken together, the authors suggested that these results raise the possibility that pregnancy influences the appearance of OMS. When agents with a variety of potential effects on neurotransmitters have been excluded, a patient with ocular flutter warrants a careful evaluation for an occult neoplasm and long-term follow-up. In filming this short clip, the patient was asked to try to hold her gaze steady fixating on the camera. You will see that she is unable to do this because of bursts of spontaneous back-to-back horizontal saccades without a saccadic interval; this abnormality is characteristic of ocular flutter. Ocular flutter in the absence of other neurologic signs implies a paraneoplastic syndrome due to an occult neoplasm, most frequently cancer of the breast.

Neuroimaging	Ocular flutter has been hypothesized to be caused by loss of "pause" neuronal inhibition of burst neuron function in the paramedian pontine reticular formation (PPRF). However, there has only been one imaging study confirming this anatomic localization. The report is of a young woman with a definite relapse of her MS who developed prominent ocular flutter without any obvious ophthalmoplegia or nystagmus. An axial MRI FLAIR sequence through the pons and medulla showed a single prominent midline high signal lesion in the region of the PPRF. A sagittal midline section showing the same lesion demonstrated its craniocaudal distribution and its subventricular localization. Repeat axial and sagittal images through the same region months later showed the disappearance of the midline pontine lesion. The disappearance of the lesion followed acute treatment with a 3-day course of IV methylprednisolone, and the ocular flutter dramatically improved. The authors of this case referred to the MRI atlas by Kretschman and Weinrich and confirmed that the brainstem lesion was at the level of the sixth nerve nuclei and was virtually exactly occupying the position of the PPRF. There were other far less obvious small areas of abnormal signal in the posterior fossa, including two in the left cerebellar hemisphere and one each in the right cerebellar hemisphere and the superior medulla. As she recovered from the MS attack, both the midline pontine lesion and the ocular flutter dramatically improved. This case is the first clear evidence that at least some cases of ocular flutter are due to lesions involving the PPRF. **(Schon F et al: Ocular flutter associated with a localized lesion in the paramedian pontine reticular formation. Ann Neurol 50:413, 2001)**
Anatomy	In 1979, Zee and colleagues studied a single patient with ocular flutter and related their findings to what was known about the anatomy and physiology of saccade generation in the monkey. They proposed that "pause" neurons normally prevent saccadic oscillations during fixation by inhibiting "burst" neuron firing and that this mechanism is disturbed in ocular flutter. They later proposed a similar disturbance in both voluntary and blink-induced saccadic oscillations. They hypothesized that the anatomic site involved in this group of eye movement disorders, which also includes opsoclonus, would be in the medial region of the PPRF, which is the human equivalent of the pontine raphe interpositus nucleus (RIN), in which pause neurons are located in the monkey. In humans, it lies adjacent to the midline and in the upper pons at the level of the sixth nerve nucleus but slightly ventral to it. However, experimental lesions of the omnipause region with excitotoxins caused slow saccades rather than oscillations. One possible explanation is that burst neurons may also have been affected. The only case comparable to the MS patient reported by Schon and colleagues is one reported by Averbuch-Heller and colleagues. Their patient developed macrosaccadic oscillations 5 years after a head injury. MRI in this patient showed a lesion in the right side of the pons extending upward from the level of the sixth nerve nucleus into the tegmentum and basis pontis. It was proposed that the eye movement abnormality was caused by damage to the adjacent omnipause neuron projections.
Pathology	Multiple sclerosis
Etiology	
Treatment	Ocular flutter ceased when the acute MS attack went into remission
Disease/Diagnosis	Multiple Sclerosis Ocular Flutter
Further Readings	Apsner R et al: Cyclosporin A induced ocular flutter after marrow transplantation. Bone Marrow Transplant 20:255, 1997 Averbuch-Heller L et al: Dysfunction of pontine ominipause neurons causes impaired fixation: Macrosaccadic oscillations with a unilateral pontine lesion. Neuroophthalmology 16:99, 1996 Buttner U et al: Opsoclonus and ocular flutter. Nervenarzt 68:633, 1997 Buttner-Ennever JA, Pause M: Neuroanatomic identification of the raphe nucleus in the pons associated with omnipause neurons of the oculomotor system in the monkey. J Comp Neurol 267:307, 1988 Cogan DG: Ocular dysmetria: Flutter like oscillations of the eyes, and opsoclonus. Arch Ophthalmol 51:318, 1954

Francis DA, Heron JR: Ocular flutter in suspected multiple sclerosis: A presenting paroxysmal manifestation. Postgrad Med J 61:333, 1985

Gotot T et al: Case of opsoclonus-myoclonus syndrome (OMS) developing during pregnancy. Nippon Naika Gakkai Zasshi 88:344, 1999

Gresty MA et al: Mechanism of rotary eye movements in opsoclonus. Br J Ophthalmol 62:533, 1980

Idris MNA, Sokrab TEO: Post malaria cerebellar ataxia and ocular flutter: Report of two cases. East Afr Med J 7:417, 1999

Kaneko CRS: Effect of ibotinic acid lesions of the omnipause neurons on saccadic eye movements in rhesus macaques. J Neurophysiol 75:2229, 1996

Kretschmann HJ, Weinrich W: Cranial neuroimaging and clinical neuroanatomy. Stuttgart: Thieme Verlag, 1992

Leigh RJ, Zee DS: Diagnosis of nystagmus and saccadic intrusion, in *The Neurology of Eye Movements*, 4th ed. New York, Oxford University Press, 2006, pp 475-558

Orimo S et al: An autopsied case of purulent meningitis associated with ocular flutter. Brain Nerve 50:469, 1998

Reiji K et al: Opsoclonus-myoclonus syndrome during pregnancy. J Neuroophthalmol 24:273, 2004

Schon F et al: Ocular flutter associated with a localized lesion in the paramedian pontine reticular formation. Ann Neurol 50:413, 2001

Wiest G et al: Ocular flutter and trunkal ataxia may be associated with enterovirus infection. J Neurol 244:288, 1997

Zee DS, Robinson DA: A hypothetical explanation of saccadic oscillations. Ann Neurol 5:405, 1979

Video e11-3

Metadata Element Field	Metadata
Identifier	168-6
Title	Downbeat Nystagmus Periodic Alternating Nystagmus
Ocular Movements	Downbeat Nystagmus Periodic Alternating Nystagmus
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Downbeat Nystagmus Periodic Alternating Nystagmus Multiple Sclerosis
Presenting Symptom	Blurred vision
History	This patient carries a diagnosis of multiple sclerosis (MS).
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2008
Date Original	1973
Resource.Type	Video
Format.Creation	
Source	16-mm Film
Language	En
Relation-Is Part Of	932-4, 946-8
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890

Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist
Reviewer	David S. Zee, MD, Johns Hopkins Hospital 2009
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	This is a unique patient with MS who has periodic downbeat nystagmus and periodic alternating nystagmus (PAN). The combination of downbeat nystagmus and PAN was first reported in a case of multiple sclerosis by Keane in 1974. It is also reported to occur in the setting of severe hypomagnesemia possibly associated with thiamine deficiency (Du Pasquier and colleagues, 1998). The eye movements show: 1. No nystagmus initially in primary gaze; 2. A period of downbeat nystagmus in central gaze; 3. A period of PAN; and 4. No nystagmus on upgaze. Partway through the period of downbeat nystagmus, the direction of the nystagmus changes to a spontaneous horizontal jerk nystagmus present in central gaze, which reverses direction approximately every 2 min; this is diagnostic of PAN. **Periodic Alternating Nystagmus:** Because the period of oscillation in one direction is long, about 4 min, the diagnosis of PAN may be missed unless the examiner observes the nystagmus for several minutes. As the nystagmus finishes one cycle of right beating nystagmus, a brief transition period occurs during which there may be brief beats of downbeat nystagmus before the next half cycle starts of left beating nystagmus. Although rare, acquired PAN is perhaps the best understood of all forms of nystagmus and was the first for which an effective treatment was identified with the drug baclofen. Convergence can be used to suppress PAN in some patients. The pathogenesis of PAN is due to a lesion of the cerebellar nodulus and uvula. Acquired PAN has been reported in association with a number of conditions.
Neuroimaging	Neuroimages were not available in this patient.
Anatomy	Lesions of the cerebellar nodulus and uvula
Pathology	Demyelination
Etiology	
Treatment	The GABAergic drug baclofen (30 mg/d) abolishes acquired PAN in most patients but helps only occasional patients with the congenital form of PAN. This patient was seen prior to the availability of baclofen to treat PAN.
Disease/Diagnosis	Multiple Sclerosis Downbeat Nystagmus Periodic Alternating Nystagmus
Further Readings	Du Pasquier R et al: Periodic downbeat nystagmus. Neurology 51:1478, 1998 Furman JM et al: Vestibular function in periodic alternating nystagmus. Brain 113:1425, 1990 Garbutt S et al: Effects of visual fixation and convergence in periodic alternating nystagmus due to MS. Neuroophthalmol 28:221, 2004 Halmagyi GM et al: Treatment of periodic alternating nystagmus. Ann Neurol 8:609, 1980 Jay WM et al: Periodic alternating nystagmus clearing after cataract surgery. J Clin Neuroophthalmol 5:149, 1985

Keane JR: Periodic alternating nystagmus with downward beating nystagmus. Arch Neurol 30:399, 1974

Leigh RJ et al: A hypothetical explanation for periodic alternating nystagmus: Instability in the optokinetic-vestibular system. Ann NY Acad Sci 374:619, 1981

Leigh RJ, Zee DS: Diagnosis of nystagmus and saccadic intrusion, in *The Neurology of Eye Movements*, 4th ed. New York, Oxford University Press, 2006, pp :475-558

Matsumoto S et al: Periodic alternating nystagmus in a patient with MS. Neurology 56:276, 2001

Video e11-4

Metadata Element	Field Metadata
Identifier	933-1
Title	Bilateral Internuclear Ophthalmoplegia
Ocular Movements	Bilateral Internuclear Ophthalmoplegia; Abducting Nystagmus; Normal Convergence; Gaze-Evoked Upbeat Nystagmus; Gaze-Evoked Downbeat Nystagmus; Saccadic Dysmetria;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Bilateral Internuclear Ophthalmoplegia; Abducting Nystagmus; Normal Convergence; Gaze-Evoked Upbeat Nystagmus; Gaze-Evoked Downbeat Nystagmus; Saccadic Dysmetria; Multiple Sclerosis;
Presenting Symptom	Blurred vision
History	The patient is a 25-year-old woman who was in excellent health until 4 days prior to admission, when she noted blurred vision and horizontal double vision on lateral gaze to right and left. **Past History:** Negative for strabismus as a child. No previous episodes of transient neurologic symptoms. **Family History:** Negative for neurologic diseases. **Neuro-Ophthalmologic Examination:** Visual acuity 20/20 OU Visual fields, pupils and fundus examination normal **Ocular Motility:** Paresis of adduction of the right eye on gaze left Abducting nystagmus of the left eye on gaze left Paresis of adduction of the left eye on gaze right Abducting nystagmus of the right eye on gaze right Normal convergence Upbeat nystagmus on upgaze Downbeat nystagmus on downgaze

	Saccadic dysmetria Hypermetria of the adducting eye Hypometria of the abducting eye **Diagnosis:** Bilateral internuclear ophthalmoplegia (INO) Saccadic dysmetria **Brain CT with and without Contrast:** Normal. **Prognosis:** On follow-up 6 weeks later, the eye movements were normal. **Diagnosis:** Multiple Sclerosis (MS) The diagnosis of MS was suspected and discussed with the patient and her parents. The rapid recovery of her motility dis-order was consistent with the diagnosis. MS is the most common cause of *bilateral* internuclear ophthalmoplegia in a young adult. Brainstem infarction is the commonest cause of *unilateral* internuclear ophthalmoplegia in middle-aged and elderly adults. A bilateral internuclear ophthalmoplegia in a child raises the possibility of a fourth-ventricle tumor.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2006
Date Original	1990
Resource.Type	Video
Format.Creation	
Source	3/4″ Umatic master videotape
Language	En
Relation-Is Part Of	163-6, 163-15,168-6, 906-4, 937-8, 941-2, 941-3
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video compressionist Steve Smith, Videographer
Reviewer	David Zee, MD, Johns Hopkins Hospital, 2006
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	This patient with MS had a bilateral INO with: ■ Paresis of adduction of the right eye on gaze left; ■ Abducting nystagmus of the left eye on gaze left; ■ Paresis of adduction of the left eye on gaze right; ■ Abducting nystagmus of the right eye on gaze right; ■ Normal convergence; ■ Upbeat nystagmus on upgaze; ■ Downbeat nystagmus on down gaze; and ■ Saccadic dysmetria Hypermetria (overshoot) of the adducting eye Hypometria (undershoot) of the abducting eye

	The Clinical Features of an INO are: 1. Medial rectus muscle weakness ipsilateral to the side of the lesion with **paresis of adduc-tion** or **adduction lag;** 2. Abducting nystagmus of the eye contralateral to the lesion—**dissociated nystag-mus;** 3. Normal convergence; 4. **Skew deviation**—hypertropia on the side of the lesion; and 5. **Dissociated** vertical nystagmus—downbeat with greater torsional component in the contralateral eye. **Bilateral INO** with bilateral lesions of the MLF may also have: Gaze-evoked vertical nystagmus; Impaired vertical pursuit; Decreased vertical vestibular response; and Small-amplitude saccadic intrusions suggesting involvement of the brainstem adjacent to the MLF. **Weakness of adduction** is due to impaired conduction in axons from the abducens internuclear neurons, which project to the medial rectus motor neurons in the contralateral oculomotor (third nerve) nucleus. Adduction weakness is most evident during saccades, and **adduction lag** is brought out clinically by asking the patient to look all the way to the right and all the way to the left (i.e., to make large saccades) back and forth across the midline. The speed of the adducting eye depends on a strong agonist contraction. The adducting saccade may be slow and hypometric. In the abducting eye, abducting saccades are hypometric with centripetal drifts of the eye and slowing. A series of small sac-cades and drifts have the clinical appearance of **abducting nystagmus—dissociated nys-tagmus.** **Dissociated nystagmus may be due to:** 1. Impaired ability to inhibit the affected medial rectus; or 2. The brain's attempts to compensate for the adduction weakness. (For further discussion, review Leigh JR, Zee DS: Diagnosis and central disorders of ocular motility, in The Neurology of Eye Movements, 4th ed. New York, Oxford University Press, 2006, pp 620-627.) **Skew deviation,** commonly seen in unilateral INO, is due to inter-ruption of central projections in the otolithic pathway ascending in the MLF to the midbrain. The higher eye (hypertrophic) is usually on the side of the MLF lesion. Interruption of pathways mediating the vertical vestibulo-ocular reflex (VOR) may cause **downbeat nystagmus** with a greater torsional component in the eye contralat-eral to a unilateral INO. **Oscillopsia,** an illusion of movement of the visual world, is a com-mon presenting symptom of INO. Horizontal oscillopsia usually occurs from either the adduction lag or the abducting nystagmus. Vertical oscillopsia occurs during head movements and is caused by a deficient vertical VOR or, as in this case, by pendu-lar vertical oscillations.
Neuroimaging	Neuroimaging studies were not available in this patient.
Anatomy	**The medial longitudinal fasciculus (MLF)** is a major pathway in the brainstem extending from the pons up to the midbrain. The MLF carries signals for the control of horizontal eye movements. For horizontal gaze: 1. The MLF contains axons from the abducens internuclear neurons and carries signals for horizontal saccades, the vestibulo-ocular reflex (VOR), and smooth pursuit. 2. These axons project to the medial rectus motor neurons in the contralateral oculomotor (third nerve) nucleus. For vertical gaze: 1. The MLF contains axons from the rostral interstitial MLF (riMLF), which carry vertical saccadic signals. 2. The MLF also contains ascending axons from the vestibular nuclei, which carry

	signals for the vertical VOR, smooth pur-suit, gaze holding, and otolith-ocular reflex. 3. Axons project to the oculomotor and trochlear (fourth nerve) nuclei, as well as to the interstitial nucleus of Cajal.
Pathology	Review (7)
Etiology	
Treatment	Review (3)
Disease/Diagnosis	Multiple Sclerosis
Further Readings	Gordon RM, Bender MB: Visual phenomenon in lesions of the medial longitudinal fasciculus. Arch Neurol 15:238, 1966 Gresty MA et al: Disorders of the vestibuloocular reflex producing oscillopsia and mechanisms compensating for loss of labyrinthine function. Brain 100:693, 1977 Hauser SL, Goodin DE: Multiple Sclerosis and other demyelinating diseases, in Harrison's Principals of Internal Medicine, 16th ed, DL Kasper et al (eds). New York, McGraw-Hill, 2005 Keane JR: Internuclear ophthalmoplegia: Unusual causes in 114 of 410 Patients. Arch Neurol 62:714, 2005 Leigh JR, Zee DS: Diagnosis of central disorders of ocular motility, in *The Neurology of Eye Move-ments*, 4th ed. New York, Oxford University Press, 2006. pp 598-718 Muri RM, Meienberg O: The clinical spectrum of internuclear ophthalmoplegia in multiple sclerosis. Arch Neurol 42: 851, 1985 Strominger MB et al: Bilateral internuclear ophthalmoplegia with absence of convergence eye movements. Clinico-pathological correlations. J Clin Neuro-ophthalmol 6:57, 1986

Cases with Myasthenia Gravis or Mitochondrial Myopathy

Video e11-5
Unilateral Ptosis: Myasthenia Gravis (Thymic Tumor) (ID163-1)

Video e11-6
Progressive External Ophthalmoplegia (PEO: Mitochondrial Cytopathy) (ID906-2)

Video e11-5

Metadata Element Field	**Metadata**
Identifier	163-1
Title	Unilateral Ptosis
Ocular Movements	Unilateral Ptosis; Unilateral Lid Retraction; Myasthenic Lid Twitch; External Ophthalmoplegia;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Unilateral Ptosis; Unilateral Lid Retraction; Myasthenic Lid Twitch; External Ophthalmoplegia; Myasthenia Gravis; Tensilon Test; Thymolipoma; Generalized Myasthenia Gravis
Presenting	Transient double vision

Symptom	
History	The patient is a 46-year-old woman who presented in July 1977 with horizontal double vision lasting 2 weeks; 3 weeks later the left upper eyelid started to droop and by the end of the day the eye was closed. She had no ptosis of the right eye and no generalized fatigue. She consulted an internist; a glucose tolerance test was normal. She was referred to a neurosurgeon, who noted weakness of the medial rectus muscle—third nerve palsy. **CT Brain Scan:** Normal. She was referred to the Neurovisual Clinic, Massachusetts General Hospital. **Past History:** Negative for previous attacks of diplopia, ptosis, or fatigue. **Neuro-Ophthalmologic examination:** Visual acuity: 20/30, J1 OU Visual fields, pupils, and fundus examination normal **Eyelids:** ■ Partial ptosis left eye (OS) ■ Lid retraction right (OD) ■ Bilateral overaction of the frontalis muscle ■ Myasthenic lid twitch OS ■ Slight increase in ptosis OS on fatigue ■ No recovery of ptosis on gentle eye closure ■ Impaired ability to bury eyelashes fully **Ocular Motility:** ■ Mild weakness of the medial rectus muscle bilaterally, left > right ■ Poor convergence ■ Vertical gaze normal **Intravenous Tensilon Test (Edrophonium Chloride):** The test dose of 0.2 mL was adequate to produce a positive response with elevation of the ptotic left eyelid and correction of lid retraction OD. The response lasted 30 s and then the left eyelid drooped. A further 0.3 mL of edrophonium again resulted in correction of ptosis OS. The full 1-mL (10-mg) dose of edrophonium was not given. **Diagnosis:** Ocular myasthenia gravis. **Hematologic Tests:** Thyroid studies normal Anti–skeletal muscle antibodies positive (The presence of antibodies to striatal muscle suggests that the patient harbors a thymoma.) Tests for antibodies to the nicotinic acetylcholine receptor (AChR) were not available at that time. **Chest X-Ray PA and Lateral:** A large anterior mediastinal mass was found consistent with an enlarged thymus gland. **CBT of Thorax:** Multiple transverse sections through the mid-thorax showed the presence of a softly demarcated rounded mass in the anterior mediastinum directly contiguous and anterior to the inferior portion of the transverse aortic arch. The mass measured approximately 4 cm and was surrounded by fat; it had diminished attenuation in the center, indicating the presence of a moderate amount of fat or liquid within the tumor. Tomograms revealed the mediastinal mass to be homogeneous without evidence

of calcification or lobulation. There was no hilar adenopathy.

Diagnosis:

Thymoma

Thymectomy:

On **August 8, 1977** a thymectomy was performed and encapsulated tumor completely excised.

Pathology:

The specimen contained a partially cystic mass and a solid portion that constituted approximately one-half of the tumor. The cystic mass was 5 x 4 x 1 cm. and the overall dimensions of the tumor 11 x 5 x 1 cm, weighing 45 g.

Diagnosis: Thymolipoma

Postoperative Status:

On day one, the patient complained of diplopia in mid-afternoon and marked drooping of her eyelids. At that time she had:

- Bilateral asymmetrical ptosis, OS > OD;
- Palpebral fissure OD 9, OS 7;
- Increased ptosis bilaterally on prolonged upgaze;
- Full horizontal and vertical gaze;
- No facial weakness;
- Normal bulbar muscles;
- Neck flexion 3/4 mild weakness;
- Good proximal strength in the limbs; and
- Normal ventilatory capacity.

Patient made an excellent recovery and was discharged home without any medication.

Second Admission:

In **September 1977,** 6 weeks post-op, she was readmitted with increasing ptosis, diplopia and generalized fatigue. At the end of the day she had fatigue chewing and weakness of the jaw and neck.

Importantly, she had no difficulty swallowing or breathing and no change in the quality of her voice. She had become depressed and anxious.

Ocular Motility:

- Bilateral symmetrical ptosis with weakness of the orbicularis oculi, and an inability to bury her eyelashes
- Increased ptosis on fatigue
- Myasthenic lid twitch OS
- Fatigue of horizontal saccades after rapid gaze right and left, to the point where the eyes came to a standstill

Neurologic Examination:

- A mild bulbar palsy
- Bilateral facial weakness with difficulty pursing her lips
- Inability to sustain the arms elevated for long periods
- Normal vital capacity

Electrophysiologic Studies:

The technique for **repetitive stimulation** studies is similar to motor nerve conduction studies. Rather than a single supramaximal stimulus, trains of repetitive stimuli are delivered at a rate of 3 stimuli per second, with 6–10 stimuli in a train. The compound muscle action potential of the first response is compared with the fifth response and the percentage decrement measured. A decrement of greater than 10% represents a positive test for myasthenia gravis (MG).

In this patient the study revealed decrements in the right deltoid, right biceps, and inferior orbicularis oculi muscle, confirming a diagnosis of generalized myasthenia gravis.

Stimulated Single-Fiber EMG (SF-EMG) was performed in the right digitorum communis. Seven pairs were recorded. Jitter was abnormal in one pair. MCD ranged from 15.7 to 134.8 μs.

	Blocking was present. SF-EMG is the more sensitive electrophysiologic method for the diagnosis of myasthenia. This is a special technique for the recording of single-muscle-fiber action potentials and is used to measure fiber density and so-called jitter. Jitter is the variability of the interpotential interval of successive discharges of two single muscle fibers belonging to the same motor unit. This phenomenon is due largely to the very slight variability of delay at the branch points in the distal axon and by synaptic delay at the neuromuscular junction. **Diagnosis:** Myasthenia gravis **Treatment:** Mestinon (pyridostigmine bromide) 60 mg q3h Prednisone 40 mg daily **Hospital Course:** Twenty-four h after starting medication there was striking improvement. By day four, she had fully recovered. In **July 1980** she was admitted for the third time with acute difficulty swallowing, shortness of breath, ptosis, diplopia, and limb weakness. Her relapse was due to her starting to taper her own prednisone dose down from 5 mg/d to 2.5 mg/d because she was mildly cushingoid. In the emergency department she had on examination: ■ A mild bulbar palsy; ■ Moderate weakness of proximal muscles; ■ Bilateral ptosis; ■ Easy fatigue of her eye movements; and ■ Normal vital capacity. **Medication:** Prednisone was increased to 10 mg bid Pyridostigmine 60 mg 4x/d She rapidly recovered back to her normal baseline and was discharged home. In **May 1998** she was admitted as an emergency with occlusion of the right dorsalis pedis and anterior tibialis arteries. A **thrombectomy** was performed. Hematologic studies revealed a hypercoagulable state: Prothrombin time 15.6 Partial thromboplastin time 24.1 CEA level 120 Hematocrit 26.7 The presence of a hypercoagulable state and elevated CEA level led to a workup for an occult malignancy. **Colonoscopy** revealed a sigmoid mass approximately 2 x 2 cm obstructing the colon. A left colectomy and colorectal anastomosis were performed. Observations during surgery revealed metastatic spread with four nodules located in the right lobe of the liver. **Pathology:** Adenocarcinoma of the colon with 18/18 positive mesenteric lymph nodes. Patient died in 1999.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2007
Date Original	1977
Resource.Type Format.Creation	Video
Source	16-mm Film
Language	En
Relation-Is Part Of	163-10, 166-25, 920-1,

Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist
Reviewer	
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	The video of this patient illustrates eyelid signs of ocular myasthenia gravis. ■ Partial ptosis OS ■ Retraction of the upper eyelid OD ■ Myasthenic lid twitch OS (overshoot of the upper lid on looking up after full gaze down) ■ Full eye movements A **positive IV Tensilon test** showed: ■ Full recovery of ptosis OS ■ Loss of lid retraction OD ■ Watering of the eyes ■ Frequent blinking **Lid Twitch—Cogan's Sign:** In **1965 Cogan** described a transient eyelid retraction occurring during refixation from downgaze to straight ahead gaze. The twitch is an "overshoot" of the eyelid. Cogan's lid twitch sign is not pathognomonic for MG. It may occur with brainstem or peripheral ocular motor disorders. **Hering's Law of Equal Eyelid Innervation:** Unilateral ptosis and contralateral lid retraction demonstrate Hering's law of equal eyelid innervation. Thus, when the ptotic lid is manually raised, the contralateral lid falls to a normal position since a large innervation is no longer required. **Ptosis:** Ptosis is defined as the lid covering more than 2 mm of the cornea. Ptosis is measured by documenting the width of the palpebral fissure in millimeters with the eyes in primary gaze and the eyebrows held down straight. Approximately 50% of patients with MG present with ptosis. More than 90% eventually develop eye movement abnormalities and typical ocular MG. Of those patients who present with only ocular symptoms, half persist with purely ocular myasthenia and half go on to develop generalized MG. Of those who develop generalized MG, most do so within 2 years of the onset of the ocular symptoms as in this patient.
Neuroimaging	CT and MRI of the mediastinum are the most sensitive radiologic techniques for detecting a thymoma. CT is superior in screening for thymoma. MRI may, however, offer better resolution in evaluating the extent or spread of a thymoma to the pleural cavity.
Anatomy	Both **thymic hyperplasia** and thymoma are associated with MG. **Thymic hyperplasia** occurs in as many as 65–70% of all myasthenic patients, particularly younger patients. It is characterized by infiltration of the thymus with lymphocytes and plasma cells and the formation of lymphoid follicles (germinal centers). **Thymoma** occurs in 5–20% of myasthenic patients. The incidence of this tumor increases with age. Patients with thymoma tend to have more severe disease, higher serum titers of AChR antibodies, and more severe abnormalities on EMG than patients without a thymoma.

	Associated Autoimmune Diseases: There is a 23% incidence of associated autoimmune disease in patients with thymoma, although no sex predisposition or HLA antigen has been found.
Pathophysiology	MG is an autoimmune disease caused by sensitized T helper cells and an IgG-directed attack on the nicotinic acetylcholine receptor of the neuromuscular junction (NMJ). The mechanism of antibody damage to the receptor and motor endplate probably involves several steps. 1. There is a complement-directed attack with the destruction of acetylcholine receptor and the junctional folds. 2. Binding of the antibody to the receptor can cause receptor blockade. 3. The abnormal and reduced numbers of acetylcholine receptors lead to impaired NMJ transmission. 4. In postsynaptic disorders such as MG, the number of quanta of acetylcholine released by each nerve stimulus is normal, but the effect of each quantum on its receptor is reduced. 5. The net result is a lower endplate potential and a reduced safety factor of transmission at the NMJ. Clinically this manifests as pathologic fatigability, i.e., progressive muscle weakness with use—the hallmark of MG. Patients typically improve after rest or upon arising in the morning, with worsening as the day passes. 6. In MG, fatigue is limited to muscular fatigue alone and often progresses to frank muscle weakness.
Etiology	Autoimmune
Treatment	See above
Disease/Diagnosis	Generalized myasthenia gravis Thymolipoma
Further Readings	Averbuch-Heller L et al: Hering's law for eyelids: Still valid. Neurology 45:1781, 1995 Bever CT et al: Prognosis of ocular myasthenia. Ann Neurol 14:516, 1983 Cogan DG: Myasthenia gravis. A review of the disease and a description of lid twitch as a characteristic sign. Arch Ophthalmol 74:217, 1965 ——— et al: Rapid eye movements in myasthenia gravis. I. Clinical observations. Arch Ophthalmol 94:1083, 1976 Daroff RB: The office Tensilon test for ocular myasthenia gravis. Arch Neurol 43:843, 1986 Elrod RD, Weinberg DA: Ocular myasthenia gravis. Ophthalmol Clin N Am 17:275, 2004 Golnik KC et al: An ice test for the diagnosis of myasthenia gravis. Ophthal 106:1282, 1999 Hanisch F et al: MuSK-antibody positive pure ocular myasthenia gravis. J Neurol 253:659, 2006 Kaminski HJ et al: Susceptibility of ocular tissues to Autoimmune Diseases. Ann NY Acad Sci 998:362, 2003 Kupersmith MJ et al: Development of generalized disease at 2 years in patients with ocular myasthenia gravis. Arch Neurol 60:243, 2003 Leigh JR, Zee DS: Diagnosis of peripheral ocular motor palsies and strabismus, in *The Neurology of Eye Movements* , 4th ed, New York, Oxford University Press, 2006, pp 385-474 Meriggioli MN, Sanders DB: Myasthenia gravis: Diagnosis. Semin Neurol 24:31, 2004 Moorthy G et al: Ocular pseudomyasthenia or ocular myasthenia "plus": A warning to clinicians. Neurology 39:1150, 1989 Pelak VS, Quan D: Ocular Myasthenia Gravis, in *UpToDat*e, BD Rose BD (ed). Wellesley, MA, 2006 Seybold ME. The office Tensilon test for ocular myasthenia gravis. Arch Neurol 43:842, 1986 Sommer N et al: Ocular myasthenia gravis. A critical review of clinical and

pathophysiological aspects. Doc Ophthalmol 84:309, 1993

Valls-Canals J et al: Stimulated single-fiber EMG of the frontalis and orbicularis oculi muscles in ocular myasthenia gravis. Muscle Nerve 28:501, 2003

Vincent A, Newsom-Davis J: Anti-acetylcholine receptor antibodies. J Neurol Neurosurg Psychiatry 43:590, 1980

Wittbrodt ET: Drugs and myasthenia gravis an update. Arch Intern Med 157:399, 1997

Video e11-6

Metadata Element Field	Metadata
Identifier	906-2
Title	Progressive External Ophthalmoplegia
Ocular Movements	Bilateral Ptosis; Facial Weakness; Complete External Ophthalmoplegia; Normal Pupils;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Bilateral Ptosis; Facial Weakness; Complete External Ophthalmoplegia; Normal Pupils; Progressive External Ophthalmoplegia (PEO); Mitochondrial Myopathy;
Presenting Symptom	Droopy eyelids
History	The patient is an 18-year-old girl first seen in **1990** with a 6-year history of progressive ptosis. In **1986**, at **age 14**, she was seen by an ophthalmologist and pediatric neurologist and investigated. Myasthenia gravis was ruled out by a negative Tensilon test, negative anti –acetylcholine receptor antibodies, and normal electromyography and repetitive nerve stimulation studies of the right upper extremity. She was given a trial on pyridostigmine 50 mg tid and 180 mg time span overnight for 1 month without any improvement. In **1987**, at **age 15**, she had bilateral ptosis surgery repeated on the right eye in August 1990. By this time she had developed limitation of all conjugate eye movements due to progressive external ophthalmoplegia (PEO). In **1990,** at **age 18,** a **muscle biopsy** performed at Johns Hopkins showed ragged-red fibers and scattered atrophic myofibers. **Diagnosis: Mitochondrial Myopathy** The fundus exam showed no abnormality. Cardiac and endocrine evaluations were normal. The patient moved out of Boston and was lost to follow-up. The term **mitochondrial cytopathy** has been used to emphasize multisystem involvement in PEO.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2004
Date Original	1990
Resource.Type	Video
Format.Creation Source	3/4″ Umatic master videotape

Language	En
Relation-Is Part Of	926-2, 926-3, 945-3, 946-1
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist Steve Smith, Videographer
Reviewer	
Context URL	http://library.med.utah.edu/NOVEL/Wray
Clinical	This 18-year-old girl with PEO presented with the insidious onset of bilateral ptosis. Muscle involvement is diagnostic. The myopathic signs illustrated are: 1. Bilateral ptosis with overaction of the frontalis muscle; 2. Weakness of the orbicularis oculi muscle with impaired eye closure; 3. No recovery of ptosis on gentle eye closure (which is a sign seen in ptosis due to myasthenia gravis); 4. No fatigue of the eyelids with increased ptosis on sustained up gaze; 5. No overaction of the eyelid on suddenly looking up (a technique used to detect a myasthenic lid twitch, a sign described by Dr. David Cogan); 6. A complete external ophthalmoplegia with gaze fixed in primary position (noteworthy that the patient was unaware that she could not move her eyes fully until the time of her neurovisual examination); and 7. Weakness of the lower face, impairing the ability to grip the lips tightly together.
Neuroimaging	Neuroimaging studies were not done in this case. Wray and colleagues (1995) provided MR images in Kearns-Sayre syndrome (KSS) and PEO. The figures included: A 61-year-old woman (patient 1) with KSS, moderately severe truncal and appendicular ataxia, and a documented mtDNA deletion. A. T1-weighted sagittal image demonstrates severe cerebellar vermian atrophy (arrow). A 23-year-old man (patient 2) with KSS, cognitive impairment, ataxia, and an mtDNA deletion. A. T2-weighted image demonstrates regions of hyperintense signal (arrows) in the subcortical white matter. The periventricular regions were spared. B. T2-weighted image shows foci of hyperintense signal (arrows) in the dorsal midbrain. A 37-year-old woman (patient 8) with CPEO manifested by external ophthalmoplegia, ataxia, and sensorineural hearing loss. A. Long-repetition-time/short-echo-time (proton density) axial image. In the frontal lobes, abnormal hyperintense signal predominates in the subcortical white matter (arrows), whereas in the posterior temporal and parietal lobes the abnormal signal extended from the subcortical regions to the ventricular surface (curved arrows). B. T2-weighted axial MR image demonstrates bilateral hyperintense signal abnormalities in the globus pallidus (arrows). Hyperintense white matter abnormalities and ventricular dilatation are also present. C. T1-weighted sagittal image demonstrates cerebral cortical and cerebellar vermian atrophy (arrow) and thinning of the corpus callosum. Other PEO patients are reported show predominantly white matter damage that correlated with **spongiform degeneration** of the brain verified by autopsy

	examinations.
Anatomy	
Pathology	A skeletal muscle biopsy is diagnostic in mitochondrial myopathy due to an mtDNA deletion. In mitochondrial myopathy, defective oxidative phosphorylation results in mitochondrial proliferation in type 1 and 2A muscle fibers. Fibers with the most severe biochemical defects may degenerate, and adjacent fibers with less severe or no defects may appear normal. The combination of a patchy moth-eaten appearance in individual muscle fibers along with mitochondrial proliferation gives rise to the ragged-red fiber seen on modified Gomori trichrome staining. NADH staining shows abnormal subsarcolemmal mitochondria in the muscle fibers. The electron microscopic sections of skeletal muscle show abnormal mitochondria.
Etiology	Mutations in mtDNA are maternally inherited in a graded fashion. A single mtDNA mutation can lead to dramatically different clinical phenotypes, creating a very large spectrum of expressivity. For example, the A3243G mutation associated with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes **(MELAS)** can also cause cardiomyopathy, diabetes and deafness, or external ophthalmoplegia. Deletions of mtDNA in skeletal muscle, ranging in size from 3.8 to 9.1 kb, were found in an identical location on muscle biopsy in 5 of 11 personal cases (3 KSS, 8 PEO). The deletion encompasses structural genes for the mitochondrial respiratory chain and is associated with impaired mitochondrial function. The variable involvement of multiple organs, (e.g. heart, brain and retina in PEO and KSS) may be attributable to a mixed population of mutant and normal genomes in varying amounts in different tissues. Both muscle and brain are also involved in patients with the MELAS syndrome, which is characterized by mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes; and **MERRF**, characterized by myoclonic epilepsy associated with ragged-red fibers. In **MELAS**, dysfunction of the central nervous system dominates the clinical picture. While there is considerable overlap of symptoms and signs between **PEO, KSS, MELAS,** and **MERRF,** there is general agreement that cases of mitochondrial myopathy, **PEO** and **KSS,** with or without clinical involvement of the brain, should be considered separately. The term **mitochondrial encephalomyopathy** or **cytopathy** has been applied to the multisystem diseases involving brain, skeletal muscle, and other organs. These disorders and the clinical phenotypes of mtDNA disease span the spectrum of all known oxidative phosphorylation disorders and include **PEO,** deafness, cardiomyopathy, **MELAS,** and **MERRF.**
Treatment	Coenzyme Q (ubiquinone) deficiency is present in KSS; treatment strategies for KSS are based on supplying electron transport chain cofactors and substrates, as well as antioxidants, in an attempt to protect against mtDNA free-radical damage. The literature best supports the efficacy of coenzyme Q10 (ubiquinone) 4 mg/kg per day in mitochondrial disease.
Disease/Diagnosis	Progressive External Ophthalmoplegia; Mitochondrial Cytopathy
References	DiMauro S et al: Mitochondrial myopathies. Ann Neurol 17:521, 1985 Evans OB, et al: Clinical and Laboratory Features of Mitochondrial Encephalomyopathy Syndromes. In Inborn Errors of Metabolism of the Nervous System, in *Neurology in Clinical Practice*, 3rd ed. Philadelphia, Butterworth Henemann 2000, 1595-1662. Gallastegui J et al: Cardiac involvement in the Kearns-Sayre syndrome. Am J Cardiol 60: 385, 1987 Holt IJ et al: Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. Nature 331:717, 1988 ———: Mitochondrial myopathies: Clinical and biochemical features of 30 patients with major deletions of muscle mitochondrial DNA. Ann Neurol 26:699, 1989 Kearns TP, Savre GP: Retinitis pigmentosa, external ophthalmoplegia and

complete heart block: Unusual syndrome with histologic study in one of two cases. AMA Arch Ophthalmol 60:280, 1958

Kosmorsky G, Johns DR: Neuro-ophthalmologic manifestations of mitochondrial DNA disorders: Chronic progressive external ophthalmoplegia, Kearns-Sayre syndrome, and Leber's hereditary optic neuropathy. Neurol Clin 9:147, 1991

Mitsumoto H et al: Progressive external ophthalmoplegia (PEO): Clinical, morphologic and biochemical studies. Neurology 33:452, 1983

Moraes CT et al: Mitochondrial DNA deletions in progressive external ophthalmoplegia and Kearns-Sayre Syndrome. N Eng J Med 320:1293, 1989

Naviauz RK: Mitochondrial DNA disorders. Eur J Pediatr 159:S219, 2000

Van Goethem G, et al: Progressive external ophthalmoplegia characterized by multiple deletions of mitochondrial DNA: Unraveling the pathogenesis of human mitochondrial DNA instability and the initiation of a genetic classification. Neuromolecular Med 3:129, 2003

Wallace DC: Mitochondrial genetics: A paradigm for aging and degenerative diseases? Science 256:628, 1992

——— et al: Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. Science 242:1427, 1988

Wray SH et al: MR of the brain in mitochondrial myopathy. Am J Neuroradiol 16:1167, 1995

Zeviani M et al: Deletions of mitochondrial DNA in Kearns-Sayre syndrome. Neurology 38:1339, 1988

Cases with Paraneoplastic Disease

Video e11-7

Paraneoplastic Upbeat Nystagmus. Ca Pancreas. Positive Anti-Hu Antibody (ID212-3)

Video e11-8

Paraneoplastic Ocular Flutter, Small Cell Adenocarcinoma of the Lung, Negative Marker (ID936-7)

Video e11-9

Opsoclonus/Flutter. Bilateral Sixth Nerve Palsy, Adenocarcinoma of the Breast, Negative Marker (ID939-8)

Video e11-7

Metadata Element Field	Metadata
Identifier	212-3
Title	Paraneoplastic Upbeat Nystagmus
Ocular Movements	Upbeat Nystagmus; Lid Nystagmus; Opsoclonus; Saccadic Dysmetria; Saccadic Pursuit;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Upbeat Nystagmus; Lid Nystagmus; Saccadic Dysmetria; Oscillopsia; Saccadic Pursuit; Paraneoplastic Cerebellar Syndrome; Pancreatic Endocrine Carcinoma;

	Anti-Hu–Associated Paraneoplastic Encephalitis Paraneoplastic Opsoclonus;
Presenting Symptom	Dizziness
History	**This case was presented to the Clinical Eye Movement Society at the American Neurological Association Meeting in October 2009.** The patient is a 65-year-old woman who was in good health until 7 weeks prior to admission. On **June 22, 2009,** on the return flight from her daughter's wedding in Oregon, she began to feel "dizziness" that she characterized as an "inability to sense herself in space." This progressed insidiously over the course of hours and became intense enough to cause difficulty in standing and an inability to walk unassisted off the plane on arrival. Fully upright she felt as though "there is a sensation of backwards motion, with someone trying to push me off my heels." She also reported difficulty with short-term memory, intermittent blurring of vision, and "eyes bobbing up and down," a prominent feature that "caused quite a stir among physicians." She had no impairment in her speech or swallowing, no motor or sensory changes, and no hearing loss, tinnitus, or headache. On return home, she consulted an ENT specialist on Cape Cod who diagnosed an inner ear problem and prescribed meclizine. Her PCP diagnosed vestibular neuritis and prescribed a short course of prednisone. Her symptoms progressed and she was referred to the Massachusetts Eye and Ear Infirmary and then admitted to the Massachusetts General Hospital. **Past Medical History:** Hypertension **Family History:** Hypertension and alcoholism in both parents **Social History:** Retired but working in the family business Smoked 1 to 2 packs per day for 25 years, quit 4 years ago Alcohol: At least 2 glasses of wine per night for many years, occasionally "the better part of a bottle of wine on weekends" **Symptomatic Inquiry:** Weight loss of 10 pounds in the last 3 months Appetite good No GI symptoms **Neurologic Examination:** Alert, appropriately interactive, normal affect Orientation: Oriented to person, states "MEEI" for place and "Cambridge" for city Oriented to 2009 but states "June" for month (August) Attention: WORLD backwards without error Speech: Fluent Followed simple and complex commands Repetition, naming, comprehension intact Memory: 3/3 at registration and 0/3 at 5 min. Normal fund of knowledge Calculations: Intact Praxis: Normal **Cranial Nerves:** Normal apart from ocular motility **Motor System:** 5/5 throughout 2+ symmetric reflexes

Plantar responses flexor

Sensory System:

Impaired vibration sense in the toes

All other modalities normal

Coordination:

No titubation

Prominent trunkal ataxia

Ataxic gait with tendency to sway backwards

Neurovisual Consult:

Patient complained of marked oscillopsia and difficulty reading.

Neuro-Ophthalmologic Examination:

Visual acuity: 20/60 with difficulty

Confrontation fields, pupil reflexes, and fundoscopy normal.

Ocular Motility:

Upbeat nystagmus in primary position

Lid nystagmus

Full horizontal and vertical eye movements

Normal convergence

Upbeat nystagmus suppressed on convergence

Horizontal and vertical saccadic dysmetria

Saccadic pursuit in all directions

Horizontal optokinetic nystagmus (OKN) present

Absent vertical OKN

Horizontal and vertical oculocephalic reflex normal

No skew deviation

Blood Studies:

WBC 12,100/mm^3 (4.5–11.0)

Polys 94%

Neuts 11,340/mm^3 (1.8–7.7)

Serum electrolytes normal

Brain MRI without Gadolinium:

Nonspecific white matter foci representing chronic small vessel ischemic change

Brain MRI Thin Slices through the Brainstem/Cerebellum:

Stable scattered T2/FLAIR hyperintensities in the periventricular and deep white matter that are nonspecific.

A demyelinating process cannot be fully excluded.

Head and Neck MRI with 3D Reformatting:

Normal studies

CT Scan of the Chest with IV Contrast:

Normal study

Bone Scan:

Skeletal degenerative changes

No suspicious lytic or blastic lesions

Lumbar Puncture:

Cerebrospinal fluid protein 69 mg/dL (elevated)

Sugar 60 mg/dL

WBC 7

97% lymphs

3% monos

Elevated IgG 22.5 mg/dL (0–8.0)

CSF albumin 33.2 mg/dL (normal)

Paraneoplastic Markers:

Serum was sent to the Mayo Clinic for a paraneoplastic panel of antibodies including anti-Ma1, anti-Ma2, anti-Ri, anti-Yo, anti-Hu, anti-Zic4, anti-CV2.

Result: Anti-Hu antibody positive—titer 1:15,360

Transabdominal and Transvaginal Ultrasound:

Heterogeneous, thickened endometrial stripe measuring 17 mm

Endometrial tissue biopsy negative

CT of Abdomen and Pelvis with IV Contrast:

Solid-appearing 3.8 x 2.9 x 3.5 cm well-defined heterogeneous mass (measuring 80 Hounsfield units post-contrast) arising from the tail of the pancreas

Biopsy:

Pancreatic mass core biopsy positive for **pancreatic endocrine carcinoma**

Pathology:

The tumor was relatively well circumscribed but extended focally into the peripancreatic soft tissue and focally involved the edge of the specimen at its inferior aspect (confirmed by synaptophysin stain). The proximal resection margin was free of tumor.

Immunohistochemical study revealed that ki-67 proliferative index was <2%, and no lymphovascular invasion was identified on D2-40-stained sections.

Based on its size and the presence of lymph node metastasis

(1 of 23 lymph nodes) the lesion was best classified into well-differentiated endocrine carcinoma.

The diagnosis was confirmed with histochemical stains that showed strong tumor cell positivity for cytokeratin cocktail, chromogranin, and synaptophysin, and no significant tumor cell staining with trypsin.

Diagnosis:

Anti-Hu–associated paraneoplastic encephalomyelitis with limbic encephalitis

Brainstem encephalitis causing upbeat nystagmus

8/20/1990 Surgery:

Distal pancreatectomy with splenic preservation.

A 3.2 x 3.2 x 2.5 cm heterogeneous mass was palpable in the tail of the pancreas.

Treatment:

1. Appropriate therapy for carcinoma
   - Distal pancreatectomy
   - Chemotherapy with cyclophosphamide (one dose of 600 mg/m^2 IV)
2. Immune modulation
   - IV methylprednisolone (total 5 g)
   - IV immunoglobulin for 5 days
3. Symptomatic treatment for oscillopsia
   - Trial of clonazepam
4. Other medication
   - IV thiamine

Hospital Course:

The patient became very anxious. Impairment in cognition progressed rapidly, leading so severe confusion and inability to hold a conversation and answer questions.

There was no significant response to IV IgG.

She was discharged to rehab on **9/14/09.**

	Follow-Up: The progressive deterioration in her mental function and gait ataxia stabilized and in fact improved 5 days prior to her follow-up visit on **9/28/09:** 6 weeks after IV steroid therapy, 5 weeks post–tumor resection, and 3 weeks following IV IgG/ cyclophosphamide. On examination, although she still had significant memory impairment, trunkal ataxia, opsoclonus and upbeat nystagmus, she was more attentive, oriented x3, and conversational. She asked specific questions and had considerable insight into her illness. She was able to walk with a walker. **Blood Studies:** Serum for repeat anti-Hu antibody titer **Pathology Studies:** Pathologic paraffin sections of the pancreatic tumor and serum samples were sent to **Dr. Dalmau, Dept of Neurology, University of Pennsylvania, Philadelphia, PA, who kindly agreed to study Hu immunoreactivity.** **I also consulted Dr. Dalmau regarding long-term therapy.** He responded: "I usually treat my patients with metronomic doses of cytoxan (1mg/kg per day) for a very long time (months). They usually tolerate the treatment very well. The treatment should focus on the cytotoxic T cell immune response, not the antibodies. The antibodies often remain detectable for several years. **Rituximab** may help, not because of antibody reduction, but because of B cell depletion. (B cells are very good antigen presenters to the T cells.) I have not personally used both rituximab and cytoxan for this disorder, although I used it in other disorders and they were usually well tolerated."
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2009
Date Original	2009
Resource.Type	Video
Format.Creation	
Source	3/4' Umatic master videotape
Language	En
Relation-Is Part Of	166-1, 931-1, 936-8, 941-5
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist
Reviewer	
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	This 65-year-old patient with pancreatic cancer and a paraneoplastic cerebellar syndrome had oscillopsia. **Ocular motility** signs: ■ Upbeat nystagmus in primary position ■ Lid nystagmus ■ Upbeat nystagmus and opsoclonus under closed eyelids ■ Full horizontal and vertical eye movements ■ Normal convergence ■ Upbeat nystagmus suppressed on convergence ■ Horizontal and vertical saccadic dysmetria ■ Saccadic pursuit in all directions

	■ Horizontal optokinetic nystagmus (OKN) present ■ Absent vertical OKN ■ Horizontal and vertical oculocephalic reflex normal ■ No skew deviation The clinical features of **upbeat nystagmus** illustrated by this case are: ■ Nystagmus is present in central gaze and increases on looking up; ■ Convergence is almost completely suppressed the nystagmus; ■ It is associated with abnormal smooth pursuit; and ■ Opsoclonus is present. View additional cases of **upbeat nystagmus** alongside this case: **ID917-5** Post–resection of a cerebellar astrocytoma **ID942-3** Post–infectious brainstem encephalitis **ID906-4** Multiple sclerosis **ID208-1** Wernicke's encephalopathy post–stomach stapling for morbid obesity **Saccadic Hypermetria:** Lesions of the **dorsal cerebellar vermis and fastigial nuclei** cause **saccadic dysmetria**—typically **hypometria** if the **vermis** alone is involved and typically **hypermetria** if the **fastigial nuclei** are involved. View additional cases of paraneoplastic syndromes alongside this case: **ID931-1 Paraneoplastic opsoclonus, CA breast** (the index case of the anti-Ri antibody) **ID936-7** Paraneoplastic ocular flutter, CA lung **ID936-8** Paraneoplastic opsoclonus, downbeat nystagmus, neuroblastoma
Neuroimaging	See text
Anatomy	**Upbeat nystagmus** is reported most commonly with lesions in the caudal medulla, which variably involve the **perihyperglossal group of nuclei, including the nucleus intercalates, the nucleus of Roller, and the nucleus pararaphales.**
Pathophysiology	**Graus and colleagues (2001)** analyzed 200 patients with **anti-Hu –associated paraneoplastic encephalomyelitis (PEM).** The mean age of the 200 patients was 63 years (range, 28–82 years), and 75% were men. The predominant neurologic syndromes were sensory neuropathy (54%), cerebellar ataxia (10%), limbic encephalitis (9%), and multifocal involvement (11%). Pathologic or x-ray evidence of a tumor was obtained in 167 patients (83%) and was a small-cell lung cancer in 74% of those with histologic diagnosis. PEM preceded the diagnosis of the tumor in 71% of patients (mean delay, 7 months; range, 0.1–47 months). The location and histologic diagnosis of neoplasms in 167 patients in this group included 3 patients with carcinoma of the breast, 6 with carcinoma of the prostate, 1 with carcinoma of the ovary, and 2 with carcinoma of the pancreas (1 small cell). The predominant neurologic symptom of PEM in this study was a sensory neuropathy. The sensory symptoms in PEM patients are usually due to a lesion in the dorsal root ganglia rather than in the peripheral nerves, and in our patient the only finding of a sensory neuropathy was loss of vibration sense in the toes. Previous work suggested that when a tumor other than a small cell lung carcinoma (SCLC) was discovered in PEM patients the possibility of a concomitant SCLC was high and appropriate studies should be made to rule out this tumor, as in our case. The authors hypothesize that a **positive Hu immunoreactivity** by the extrathoracic tumor would support that the tumor was responsible for the PEM. Further studies currently conducted by Dr. Dalmau are underway in our patient. **Prognosis:**

The prognosis of PEM remains poor in terms of survival and neurologic disability. In the **Graus and colleagues** series, up to 53% of patients were severely disabled at the time of diagnosis and only a minority (5%) had a benign indolent course.

Regarding therapy, although immunotherapy alone is probably not effective in the majority of patients, four patients in the series improved with different immunotherapies. No specific immunotherapy treatment can be recommended, however; ideally, all PEM patients should be offered the possibility of inclusion in therapeutic protocols.

Autopsy Studies:

Dalmau and Graus and colleagues, in a clinical analysis of **anti-Ma2 –associated encephalitis,** obtained pathologic studies in 12 patients. In 11 patients the microscopic studies showed perivascular lymphocytic cuffing and interstitial infiltrates of lymphocytes with variable gliosis and neuronal degeneration. The inflammatory infiltrates were mainly composed of T cells with a smaller number of B cells, macrophages, and microglial activation.

The autopsy of four patients showed a correlation between the clinical MRI findings and the degree of pathologic involvement. The areas with major pathologic abnormalities were always the most symptomatic, but all four patients had abnormal findings in areas that were clinically asymptomatic. In a patient whose symptoms appeared restricted to limbic and hypothalamic dysfunction, the autopsy showed severe involvement of the hippocampus and entorhinal cortex, and mild chronic inflammation in numerous areas including frontal and parietal cortex, midbrain, substantia nigra, pontine nuclei, dorsal grey matter of the medulla, olivary nuclei, and occasional dropout of Purkinje cells with Bergmann's gliosis.

The spinal cord was spared, but the dorsal root ganglia showed several small aggregates of reactive lymphocytes.

Etiology

Treatment

The approach to treatment in paraneoplastic opsoclonus syndromes is:

1. Appropriate therapy for the cancer
   - Surgery
   - Chemotherapy and/or radiation
2. Immune modulation
   - IV immunoglobulin
   - Plasmapheresis
   - Immunoadsorption therapy
   - Steroids
3. Symptomatic treatment for vertigo

Disease/Diagnosis

Paraneoplastic Upbeat Nystagmus

Further Readings

Bataller L, Dalmau J: Paraneoplastic neurologic syndromes: Approaches to diagnosis and treatment. Semin Neurol 23:215, 2003

Budde-Steffen C et al: An anti-neuronal autoantibody in paraneoplastic opsoclonus. Ann Neurol 23:528, 1988

Byrne T et al: Spontaneous neurological improvement in anti-Hu associated encephalomyelitis. J Neurol Neurosurg Psychiatry 62:276, 1997

Dalmau J et al: Clinical analysis of anti-Ma2-associated encephalitis. Brain 127:1831, 2004

Dropcho E, Payne R: Paraneoplastic opsoclonus-myoclonus. Association with medullary thyroid carcinoma and review of the literature. Arch Neurol 43:410, 1986

Graus F et al: Anti-Hu-associated paraneoplastic encephalomyelitis: Analysis

of 200 patients. Brain 124:1138, 2001

Hierons R et al: The late effects of necrotizing encephalitis of the temporal lobes and limbic areas: A clinico-pathological study of 10 cases. Psychol Med 8:21, 1978

Leigh RJ, Zee DS: Diagnosis of nystagmus and saccadic intrusion, in *The Neurology of Eye Movements*, 4th ed.

New York, Oxford University Press, 2006, pp 521-534.

Pierrot-Deseilligny C, Milea D: Vertical nystagmus: Clinical facts and hypotheses. Brain 128:1237, 2005

Pierrot-Deseilligny C et al: Upbeat nystagmus due to a caudal medullary lesion and influenced by gravity. J Neurol 254:120, 2007

Sanders MD: Lid nystagmus evoked by ocular convergence: An ocular electromyographic study. J Neurol Neurosurg Psychiatry 1968;31:368, 1968

Tilikete C et al: Saccadic lateropulsion and upbeat nystagmus: Disorders of caudal medulla. Ann Neurol 52:658, 2002

Voltz R: Intravenous immunoglobulin therapy in paraneoplastic neurological syndromes. J Neurol 253:33, 2006

Wong AM et al: Opsoclonus in three dimensions: Oculographic, neuropathologic and modeling correlates. J Neurol Sci 2001;189:71, 2001

Video e11-8

Metadata Element Field	Metadata
Identifier	936-7
Title	Paraneoplastic Ocular Flutter
Ocular Movements	Ocular Flutter;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Ocular Flutter; Oscillopsia; Trunkal Ataxia; Paraneoplastic Cerebellar Syndrome; Small Cell Carcinoma of the Lung;
Presenting Symptom	Unsteady walking
History	The patient is a 58-year-old woman with known hypertension. In **1994,** 2 weeks prior to admission she had a dramatic change in behavior with insomnia, agitation and depression. This was accompanied by "wringing of hands and anxiety for no apparent reason." She became anorexic, lost 15 lb, and developed increasing headache, dizziness, head shaking, unsteadiness walking, and a feeling of generalized weakness. She denied nausea, vomiting, lethargy, or syncope. After consulting her doctor she came to the emergency department. She was seen by Medicine, Psychiatry and Neurology, and admitted to Psychiatry with the chief complaint, "I cannot stand up." Psychiatry considered her unsteady gait to be histrionic and related to depression and her perception that she was unable to walk. **On examination** she was extremely agitated and cooperated poorly with the examiner. She was able to walk only with the use of a walker and two nursing assistants. At that time, she also complained of difficulty focusing and an abnormality of her eye movements were noted. A neuro-ophthalmology consultation led to her transfer from Psychiatry to Neurology. **Past History:** Significant for a previous episode of depression treated with antidepressant medication and alcohol abuse in the past.

She was a heavy smoker, smoking a pack per day for more than 30 years.

Family History:

Significant for alcohol abuse in both parents.

Social History:

Eldest of six siblings, unmarried and living alone. Worked as a secretary until 10 years prior to admission when she lost her job.

Neurologic Examination:

She was oriented in time, place, and person but failed to obey three-step commands

She had:

- Fluctuating level of alertness and attentiveness;
- Motor impersistence and perseveration; and
- Impaired short-term memory.

A constellation of cerebellar signs:

- Dysarthria
- Titubation
- Fine tremor of the face and chin
- Ocular flutter with occasional spontaneous obtrusive saccades (no continuous opsoclonus)
- Marked trunkal ataxia sitting and standing
- Gait ataxia with inability to walk without support

The motor system was intact.

Reflexes symmetric, slightly brisk, with flexor plantar responses

Coordination intact in the limbs

Sensory examination showed no abnormality

The electronystagmogram showed:

1. Spontaneous and gaze nystagmus—intermittent ocular flutter epochs (5 Hz), 1- to 2-cycle sinusoids in duration throughout the record.

Ocular flutter occurred without any rhythmic pattern, with highest frequency on saccades to the left and on eye closure.

She had normal vertical excursions and normal vertical vestibulo-ocular reflex.

2. Saccades: normal velocity and latency; gaze-evoked nystagmus with large saccades in either direction.

3. Pursuit (tracking)

0.2 Hz: Sinusoidal. Ocular flutter intrusions.

0.3 Hz: Saccadic breakdown to the right and left.

Vertical pursuit gain was negligible.

4. Optokinetic nystagmus: Poor gain in each direction.

5. Positional testing:

a. Hallpike's; eyes open in the light: Dizziness with brief burst of right beating nystagmus on the first positioning.

b. Static; eyes open in the dark: No nystagmus or dizziness.

c. Static; eyes closed: No nystagmus or dizziness.

The study showed:

- Ocular flutter epochs;
- Loss of vertical pursuit tracking;
- Lateral gaze-evoked nystagmus;
- Saccadic breakdown of pursuit tracking at higher frequencies; and
- A brief burst of right beating nystagmus with accompanying dizziness upon Hallpike positioning to a head hanging right position.

These findings were consistent with brainstem and cerebellar dysfunction affecting fixation and saccadic and pursuit pathways.

Diagnosis: Paraneoplastic Cerebellar Syndrome

Ocular Flutter

Investigations:

The workup was directed toward detection of an occult neoplasm in breast, ovary, or lung.

Serum electrolytes were abnormal:

Sodium 125 mmol/L (135–145)

Potassium 2.8 mmol/L (3.4–4.8)

Osmolality 262 mOsm/kg (280–296)

Hyponatremia was secondary to the syndrome of inappropriate ADH Secretion (SIADH).

The diagnosis of **SIADH** is made by finding a concentrated urine sodium (>20 mmol/L) in the presence of hyponatremia or low plasma osmolality (<260 mmol/kg).

The causes of SIADH include malignancy such as small cell carcinoma of the lung, as in this patient.

Electroencephalogram normal.

Lumbar puncture showed normal cerebrospinal fluid (CSF) with 4 WBC, negative cytology, and no oligoclonal bands

Mammogram normal.

Abdominal and pelvic CT normal.

Bone scan negative for metastases.

Brain MRI focal nonenhancing periventricular T1 hypointensity and T2 hyperintensity noted adjacent to the left lateral ventricle.

Differential Diagnosis:

Nonspecific demyelination versus gliosis.

Chest X-Ray:

X-ray revealed a well-circumcised 6-cm mass consistent with adenocarcinoma of the lung without significant hilar lymphadenopathy.

Chest CT:

In the region of the superior segment and basilar segments of left lower lung, a large, well-rounded, smooth, homogeneously soft-tissue, dense, noncalcified mass was found: approximately 5 x 3 x 5 cm with a single pleural tag.

(Figures 1 and 2)

Diagnosis:

Primary bronchogenic carcinoma.

Surgery:

The patient had a left lower lobe lobectomy and excision of a number of hilar lymph nodes.

Pathology:

Small cell undifferentiated carcinoma of the lung with metastatic carcinoma in 2/5 hilar lymph nodes.

(Figures 3 and 4)

Paraneoplastic Markers:

A fresh-frozen segment of the lung tumor, a sample of CSF, and three samples of serum were sent to **Dr. Posner at the Memorial Sloan Kettering Cancer Center** for antibody studies (Anti-Ri, Anti-Yo, Anti-Hu). No antibodies were detected.

Yo Autoantibody Test: Negative

Antibodies to the Yo antigen are the most common paraneoplastic antibodies to Purkinje cells.

A negative result for the Yo autoantibody test proved that the cause of the disease was not due to this paraneoplastic antibody but did not rule out paraneoplastic cerebellar degeneration as the cause of the symptoms (Clouston et al, 1992; Hammack et al, 1992).

Ri Autoantibody Test: Negative

	Anti-Ri is a highly specific antineuronal antibody that reacts with nuclei of neurons in the central nervous system. The presence of anti-Ri antibody identifies the subset of patients with paraneoplastic ataxia and opsoclonus who suffer from breast or gynecologic cancer. The antibody when present is a useful marker for this type of underlying malignancy. The relative amount of anti-Ri was found always to be higher in CSF than in serum. (Anderson et al, 1988; Budde-Steffen et al, 1988; Luque et al, 1989) **Protein-A Immunoadsorption:** In spite of the absence of known paraneoplastic antibodies, the patient was treated twice weekly for three weeks with protein-A immunoadsorption: 250 mL of plasma was passed through a disposable column of staphylococcal protein-A covalently bound to a silica matrix that bound both the Fc portion of IgG molecules and immune complexes. The plasma was then returned to the patient. The first treatment was given in hospital because of the possible risk of cutaneous immune complex vasculitis. In addition she received prednisone 80 mg daily and haloperidol 2.5 mg PO qhs to minimize the risk of steroid psychosis. The mechanism of action of the immunoadsorption therapy is not clear, but it may have removed small quantities of immune complexes or antineuronal antibodies or facilitated the formation of anti-idiotype antibodies. **Oncology:** She had a limited stage small cell lung cancer, stage T2, N1, MO of the left lower lobe, treated initially with a left lower lobectomy and six cycles of postoperative chemotherapy and then with subsequent radiation to the chest (50.4 Gy in 28 fractions). **Prognosis:** She made a full recovery on completion of all the treatments. She has had no reoccurrence of ocular flutter. The patient has been followed up for more than 10 years. When last seen in 2006, she was under treatment for chronic obstructive pulmonary disease and anxiety. She was still smoking one pack of cigarettes per day.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2004
Date Original	1994
Resource.Type	Video
Format.Creation	
Source	¾ Umatic master videotape
Language	En
Relation-Is Part Of	166-1, 931-1, 936-8,
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002-2004. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neuro-ophthalmology
Catalogue	http://medstat.med.utah.edu/neuroophth/Wray/
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist Steve Smith, Videographer
Reviewer	David Zee, MD, The Johns Hopkins University Hospital, 2007. Frederick Hochberg, MD, Massachusetts General Hospital, 2007.
Context URL	http://medstat.med.utah.edu/neuroophth/Wray/
Clinical	This 58-year-old woman with paraneoplastic ocular flutter is slow to obey commands to move her eyes. Her speech is slurred and she has very little spontaneous conversation. Titubation is best seen at the start of the video.

	Ocular flutter and **flutter dysmetria** are the major diagnostic signs. ■ Intermittent epochs of **ocular flutter**—horizontal back-to-back saccadic oscillations without an intersaccadic interval. ■ **Flutter dysmetria** most evident when the eyes make voluntary saccadic movements to the left Occasionally the amplitude of flutter is very small—**microflutter**—and the oscillations detected only with an ophthalmoscope or eye movement recordings (Cogan, 1954). Ocular flutter is often associated with paraneoplastic opsoclonus. **Comment Dr. Zee (DZ),** March 2007 **SHW:** David, I neglected to test for ocular flutter under closed lids by asking the patient to look to the right and back to center and to the left and back to center. Would I expect to see flutter with eyes closed? **DZ:** Good question. This would seem to be very likely since eye closure shuts off the pause cells, and flutter itself can often be triggered by making a saccade, even in the light.
Neuroimaging	Neuroimaging studies are unavailable in this case.
Pathophysiology	Paraneoplastic opsoclonus/ocular flutter is thought to be humorally mediated, and antibodies to diverse autoantigens have been reported, but most patients are seronegative as in this case. The antineural antibodies associated with opsoclonus include Anti-Ri, Anti-Hu, Anti-Yo, Anti-Mal, and Anti-amphiphysin antibodies.
Treatment	The approach to treatment in paraneoplastic syndromes is: 1. Appropriate therapy for the cancer ■ Surgery ■ Chemotherapy and/or radiation 2. Immune modulation ■ Plasmapheresis ■ Immunoadsorption therapy ■ Intravenous immunoglobulin ■ Steroids 3. Symptomatic treatment for vertigo, etc.
Disease/Diagnosis	Small Cell Carcinoma of the Lung
Further Readings	Anderson NE et al: Paraneoplastic cerebellar degeneration: Clinical-immunological correlations. Ann Neurol 24:559, 1988 Bataller L, Dalmau J: Paraneoplastic neurologic syndromes: Approaches to diagnosis and treatment. Semin Neurol 23:215, 2003 Budde-Steffen C et al: An anti-neuronal autoantibody in paraneoplastic opsoclonus. Ann Neurol 23:528, 1988 Ciavarella D: The use of protein A columns in treatment of cancer and allied diseases. Int J Clin Lab Res 21:210, 1992 Cher LM et al: Therapy for paraneoplastic neurologic syndromes in six patients with protein A column immunoadsorption. Cancer 75:1678, 1995 Clouston PD et al. Paraneoplastic cerebellar degeneration. III. Cerebellar degeneration, cancer, and the Lambert-Eaton myasthenic syndrome. Neurology 42:1944, 1992 Cogan DG: Ocular dysmetria, flutter-like oscillations of the eyes and opsoclonus. Arch Ophthalmol 51:318, 1954 Ellenberger C et al: Opsoclonus and parenchymatous degeneration of the cerebellum. The cerebellar origin of an abnormal eye movement. Neurology 18:1041, 1968 Hammack J et al: Paraneoplastic cerebellar degeneration. II. Clinical and immunologic findings in 21 patients with Hodgkin's disease. Neurology 42:1938, 1992 Helmchen C et al: Cerebellar activation in opsoclonus: An fMRI study. Neurology 61:412, 2003 Leigh RJ, Zee DS: Diagnosis of nystagmus and saccadic intrusion, in *The Neurology of Eye Movements*, 4th ed. New York, Oxford University Press, 2006,.

pp 521-534

Luque AF et al:. Anti-Ri: An antibody associated with paraneoplastic opsoclonus and breast cancer. Ann Neurol 29:241, 1991

Luque A et al: Ann Neurol 26:178, 1989 (Abstract)

Nitschke M et al: Opsoclonus-myoclonus in small cell lung cancer: Detection of an autoantibody and improvement after protein A Column immunoadsorption therapy. N Engl J Med, 1994 (Letter)

Ridley A et al: Omnipause neurons in two cases of opsoclonus associated with oat cell carcinoma of the Lung. Brain 110:1699, 1987

Wong AM et al: Opsoclonus in three dimensions: Oculographic, neuropathologic and modeling correlates. J Neurol Sci 189:71, 2001

Zee D: Personal communication, 2007.

Video e11-9

Metadata Element Field	Metadata
Identifier	939-8
Title	Bilateral Sixth Nerve Palsy
Ocular Movements	Bilateral Sixth Nerve Palsy; Bilateral Esotropia; Paraneoplastic Opsoclonus/Flutter;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Bilateral Sixth Nerve Palsy; Bilateral Esotropia; Paraneoplastic Opsoclonus/Flutter; Paraneoplastic Cerebellar Syndrome; Carcinoma of the Breast;
Presenting Symptom	Double vision and ataxia
History	This patient is a 62-year-old woman with a 6-month history of double vision and difficulty walking. In **August 1996,** she first noted her right upper eyelid twitching followed by dizziness, nausea and vomiting. Soon after her voice became "shaky" and she experienced mild difficulty walking. She consulted her primary care physician and was prescribed sertraline for depression. In **November 1996,** she developed double vision and increasing unsteadiness walking. In **December 1996,** she consulted a neurologist. Brain MRI at an outside hospital reported to be normal In **January 1997,** she was admitted to Massachusetts General Hospital. **Past History:** Significant for right breast cancer in 1986, status post lumpectomy and radiation therapy **Family History:** Negative for neurologic disease **Neurologic Examination:** Normal cognitive function Mild dysarthria and tremulous voice Spontaneous myoclonus and bilateral hand tremors Mild left hemiparesis with hyperreflexia and extensor plantar response Right flexor plantar response Bilateral limb ataxia and gait ataxia **Neuro-Ophthalmologic Examination:** Visual acuity 20/20 OU Visual fields, pupils, and fundus examination no abnormality **Ocular Motility:** Intermittent **opsoclonus** with multivectorial conjugate saccades

without an intersaccadic interval and

Ocular flutter with horizontal back to back saccades in central gaze

- Esotropia OS > OD
- Weakness of abduction OS—left sixth nerve palsy
- Weakness of adduction OS
- Weakness of abduction OD—right sixth nerve palsy
- Full upgaze with upbeat nystagmus
- Full downgaze with no nystagmus
- Pursuit movements normal

Testing for ocular dysmetria show slow refixation saccades. No dysmetria.

Diagnosis:

Paraneoplastic cerebellar syndrome with opsoclonus/flutter.

1/27/97 Brain MRI: The following sequences were obtained:

Axial FLAIR, T1, FSE T2, diffusion-weighted images, and post gadolinium.

The study showed areas of T2 and FLAIR hyperintensity involving the tegmen and extending cranially to the midbrain adjacent to the midline and just inferior to the cerebellar peduncle.

Also noted were areas of FLAIR and T2 hyperintensity within the corona radiata and centrum semiovale bilaterally, left greater than right. Similar signal abnormalities were seen within the right putamen and globus pallidus.

Impression:

Diffuse white matter process consistent with a paraneoplastic syndrome.

Lumbar Puncture: Cerebrospinal fluid protein 36 mg/dL

Sugar 68 mg/dL

136 RBC

12 WBC

95% lymphs

4% monocytes

Positive for oligoclonal bands

Cytology negative for malignant cells twice

Chest/Abdomen/Pelvic CT

Normal

Bone Scan and Bone Marrow Biopsy: Normal

Mammogram: Bilateral mammogram

Cluster of microcalcifications in the subareolar region of the lateral aspect of the left breast suspicious for intraductal carcinoma

Procedure: Needle-guided left breast biopsy

Pathology:

Intraductal adenocarcinoma in situ in the left breast, the primary breast cancer being in the right breast.

Paraneoplastic Markers: CSF and serum were sent to **Dr. Posner at the Memorial Sloane Kettering Cancer Center** for antibody studies (Anti-Ri, Anti-Yo). No antibodies were detected.

Ri Autoantibody Test: Negative

Anti-Ri is a highly specific antineuronal antibody that reacts with nuclei of neurons in the central nervous system. The presence of anti-Ri antibody identifies the subset of patients with paraneoplastic ataxia and opsoclonus who suffer from breast or gynecologic cancer. The antibody when present is a useful marker for this type of underlying malignancy. The relative amount of anti-Ri was found always to be higher in CSF than in serum. (Anderson et al, 1988; Budde-Steffen et al, 1988; Luque et al, 1989.)

All diagnostic antibody tests should be considered within the context of clinical findings (Dalmau et al, 1992).

Diagnosis: 1. Paraneoplastic opsoclonus/ocular flutter

	2. Paraneoplastic cerebellar syndrome 3. Paraneoplastic brainstem encephalitis 4. Bilateral sixth nerve palsy 5. Intraductal adenocarcinoma in situ left breast Surgical resection of the intraductal adenocarcinoma in situ was done. **Therapy:** Ten days of IV solumedrol 1 g/d Five-day course of IV IgG **Hospital Course:** Over a period of 10 days the opsoclonus and flutter completely resolved and there was significant improvement in her ophthalmoplegia. The left eye regained almost full abduction and the right eye was able to cross the midline looking right. The spontaneous myoclonus stopped, hand tremors waxed and waned but overall improved, and the gait instability was markedly improved. She was able to walk in her room without a walker and in the hallway with a wheeled walker. She was hospitalized for 3 weeks and then discharged to a rehabilitation hospital. **Discharge Medication:** Prednisone 10 mg daily, three times a day and on a tapering dose to stop after 2 weeks. The patient was followed by the neuro-oncology team.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2008
Date Original	1997
Resource.Type	Video
Format.Creation	
Source	3/4" Umatic master videotape
Language	En
Relation-Is Part Of	163-3, 169-34, 944-5, 946-2
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist Steve Smith, Videographer
Reviewer	
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	This 62-year-old woman with paraneoplastic opsoclonus/flutter and diplopia due to bilateral sixth nerve palsy describes the visual disturbance of oscillopsia. At the time the video was made she had no opsoclonus/flutter. **The eyes show:** • Esotropia OS > OD • Weakness of abduction OS—left sixth nerve palsy • Weakness of abduction OD—right sixth nerve palsy • Impaired adduction OS • Full upgaze with upbeat nystagmus • Full downgaze with no nystagmus • Pursuit movements normal Testing for ocular dysmetria showed slow refixation saccades. No dysmetria. **In addition this patient had:** • Bilateral ataxia on finger/nose test and heel-knee-shin

	• Loss of rhythm hand tapping • Gait ataxia **Ocular flutter** and **flutter dysmetria** are the major diagnostic signs. • Intermittent epochs of **ocular flutter**—horizontal back-to-back saccadic oscillations without an intersaccadic interval • **Flutter dysmetria** most evident when the eyes make voluntary saccadic movements to the left Occasionally the amplitude of flutter is very small—**microflutter**—and the oscillations detected only with an ophthalmoscope or eye movement recordings (Cogan, 1954). Ocular flutter is often associated with paraneoplastic opsoclonus. **Comment Dr. Zee (DZ),** March 2007 **SHW:** David, I neglected to test for ocular flutter under closed lids by asking the patient to look to the right and back to center and to the left and back to center. Would I expect to see flutter with eyes closed? **DZ:** Good question. This would seem to be very likely since eye closure shuts off the pause cells, and flutter itself can often be triggered by making a saccade, even in the light.
Neuroimaging	Neuroimaging studies are unavailable in this case.
Anatomy	
Pathology	Paraneoplastic opsoclonus/ocular flutter is thought to be humorally mediated, and antibodies to diverse autoantigens have been reported, but most patients are seronegative as in this case. The antineural antibodies associated with opsoclonus include Anti-Ri, Anti-Hu, Anti-Yo, Anti-Mal, and Anti-amphiphysin antibodies. It is not known whether opsoclonus is a cerebellar or brainstem disorder. **Cerebellum:** Wong and colleagues (2001) have suggested, on theoretical grounds, that the deep cerebellar nuclei should be activated in patients with opsoclonus. **Helmchen and colleagues (2003)** assessed this hypothesis with fMRI in two patients with opsoclonus and compared them with healthy subjects. They used three-dimensional scleral search coil recordings to characterize the pathologic eye oscillations. Fortuitously, both patients showed a decrease of or no OC with the eyes closed, so fMRI signals under two conditions (open eyes with OC and closed eyes without OC) could be compared. A comparison of these two states revealed neither cerebellar vermal nor brainstem activation but showed, for the first time, bilateral functional activation of the deep cerebellar fastigial nuclei. This result supports the recent hypothesis of Wong and colleagues that OC results from a disinhibition of the fastigial ocular motor region (FOR). FOR contains saccade-related neurons that augment the ongoing discharge of pontine excitatory and inhibitory burst neurons. Because the cerebellar ocular motor vermis physiologically inhibits FOR, the authors concluded that the absence of vermal activation during OC may reflect a cause of OC. **Brainstem:** Glycine has been identified as the neurotransmitter of omnipause neurons, and poisoning with a glycine antagonist, strychnine, is reported to produce opsoclonus and myoclonus. Zee postulates that an immune-mediated **glycine channelopathy affecting the membrane of omnipause cells** may be the underlying mechanism for ocular flutter and opsoclonus and that it may be possible that membrane-stabilizing drugs may have a therapeutic role for flutter and opsoclonus in the future. **(Personal communication, 2007.)** (For further discussion see Leigh and Zee, 2006.)
Etiology	Intraductal adenocarcinoma of the breast
Treatment	Immune modulation—IV IgG
Disease/Diagnosis	Paraneoplastic opsoclonus/flutter Paraneoplastic cerebellar syndrome Paraneoplastic brainstem encephalitis

	Bilateral sixth nerve palsy Intraductal adenocarcinoma of the breast
Further Readings	Alderson LM: Neurologic complications of breast cancer, in *Office Practice of Neurology*, 2nd ed, MA Samuels, SK Feske (eds). Philadelphia, Churchill Livingstone, 2003, pp 1181-1186 Anderson NE et al: Paraneoplastic cerebellar degeneration: Clinical-immunological correlations. Ann Neurol 24:559, 1988 Bataller L, Dalmau J: Paraneoplastic neurologic syndromes: Approaches to diagnosis and treatment. Semin Neurol 23:215, 2003 Budde-Steffen C et al: An anti-neuronal autoantibody in paraneoplastic opsoclonus. Ann Neurol 23:528, 1988 Cogan DG: Ocular dysmetria, flutter-like oscillations of the eyes and opsoclonus. Arch Ophthalmol 51:318, 1954 Dalmau J et al: Anti-Hu–associated paraneoplastic encephalomyelitis/sensory neuronopathy. A clinical study of 71 patients. Medicine (Baltimore) 71:59, 1992 Helmchen C et al: Cerebellar activation in opsoclonus: An fMRI study. Neurology 61:412, 2003 Leigh RJ, Zee DS: Diagnosis of peripheral ocular motor palsies and strabismus, in *The Neurology of Eye Movements* , 4th ed. New York, Oxford University Press, 2006, pp. 385-474 Luque A et al: Ann Neurol 26:178, 1989 (Abstract) Wong AM et al: Opsoclonus in three dimensions: Oculographic, neuropathologic and modeling correlates. J Neurol Sci 189:71, 2001

Cases with Fisher Syndrome

Video e11-10

Bilateral Ptosis: Facial Diplegia, Total External Ophthalmoplegia, Positive Anti-GQ1b Antibody (ID944-1)

Video e11-10

Metadata Element Field	**Metadata**
Identifier	944-1
Title	Bilateral Ptosis Facial Diplegia
Ocular Movements	Bilateral Ptosis; Total External Ophthalmoplegia; Facial Diplegia;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP
Subject	Bilateral Ptosis; Total External Ophthalmoplegia; Facial Diplegia; Areflexia; Acute Inflammatory Demyelinating Neuropathy; Guillain-Barré Syndrome—Miller Fisher Syndrome;
Presenting Symptom	Double vision
History	The patient is a 47-year-old attorney who was transferred from an outside hospital to the Massachusetts General Hospital (MGH) for treatment of the Miller Fisher variant of Guillain-Barré syndrome (GBS). On the morning of **September 14, 1993,** the patient awoke feeling dizzy and he was unsteady walking. He went back to sleep and felt better when he got up and went to work.

The next day, he noted horizontal double vision driving to work and he returned home to bed. Later that day, he developed slurred speech and tingling in the right arm and both hands. He went next door to speak to his neighbor, a nurse, and she told him to go to the hospital.

He was admitted to a local hospital.

Head and Neck MRA:

Normal.

Lumbar Puncture: Cerebrospinal fluid protein 102 mg/mL

No white blood cells.

Hospital Course: The patient got progressively worse.

On **day 5,** he had the sensation that "someone injected my whole mouth with Novocain" and his speech was profoundly worse. He also reported that his reflexes, which were normal on admission, became decreased.

On **day 10,** he was transferred to the MGH. At this time he had mild headache and dizziness and he could no longer hold his eyes open. He recalled that he had had a period of diarrhea one week prior to the onset of these symptoms.

On admission he complained of generalized weakness and marked weakness of the face and arms, less in the legs. He was able to chew using the left side of his mouth but needed to drink with a straw. He had no difficulty swallowing or breathing.

Family History: Negative for neurologic disease

Neurological/Neuro-Ophthalmic Examination: Speech dysarthric

Cranial Nerves:

Complete bilateral ptosis

Total external ophthalmoplegia

Only a trace of vertical movements up and down

Pupils equal, minimally reactive to light and near

Facial sensation—slight decrement of pin sensation bilaterally in all three divisions of the trigeminal nerve

Severe facial diplegia

Motor System: He felt generally weak and fatigued, muscle strength 4+ to 5/5 bilaterally

Deep tendon reflexes 2-4+ in the upper extremities and 1+ in the lower extremities

Plantar responses flexor

Coordination: Impaired in all four limbs and marked gait ataxia

Sensation: Light touch and temperature decreased distally in the hands and feet

Brain CT Noncontrast: Normal

Brain MRI: Normal

Chest X-Ray: Normal

ECG: Normal sinus rhythm, normal trace

Diagnosis:

Miller Fisher syndrome (MFS)

The constellation of signs:

1. Bilateral ptosis
2. External ophthalmoplegia
3. Facial diplegia
4. Limb and gait ataxia
5. Pupils sluggishly reacting to light

Investigations: The most important investigations in suspected cases of GBS or its variant, MFS, are:

1. Electrodiagnostic studies; and
2. Examination of the cerebrospinal fluid (CSF).

Met the criteria for MFS.

	The most frequent early electrodiagnostic findings in GBS and MFS are: 1. A reduction in the amplitude of muscle action potentials; 2. Slow nerve conduction velocity; 3. Conduction block in motor nerves singly or in combination; 4. Prolonged distal latencies (reflecting distal conduction block); 5. Prolonged or absent F-responses (indicating involvement of proximal parts of nerves and roots) and reflecting focal demyelination; and 6. Delayed or absent H-reflex (which merely confirms the loss of ankle jerks). A limited electrodiagnostic examination may be normal early in the illness, as in this case. A more thorough study, which includes measurement of late responses, almost invariably shows disordered conduction in an affected limb within days of the first symptom. **Electrophysiologic Studies:** Nerve conduction in the lower extremities normal on **day 14.** **Lumbar Puncture:** Protein 108 mg/mL No white cells Usually, the CSF protein is normal in the first few days of the illness. Then the protein level begins to rise, reaching a peak in 4–6 weeks and persisting elevated for several weeks. In **Dr. Ropper's** experience patients with MFS have a higher incidence of normal or only slightly elevated CSF protein during the course of their illness. **Stool Cultures:** Positive for ***Campylobacter jejuni*** **Serum protein electrophoresis** revealed IgG kappa M components in the slow gamma region and there were oligoclonal bands present as well. **Antibody Studies:** To check for Anti-GQ_{1b} antibody, I sent the serum of this patient and the serum from a second case of MFS, a young man presenting with bilateral sixth nerve palsy, (ID944-5) to **Professor Newsom-Davis** in Oxford to study. **Both of these patients were found to have the Anti-GQ_{1b} IgG antibody.** **Treatment:** The patient received a total of six sessions of plasmapheresis with replacement with albumin. Ciprofloxacin 500 mg PO bid **Hospital Course:** Initially, the patient was too weak to get out of bed and became areflexic. He started to improve while receiving plasmapheresis and physical therapy to the point that at 5 weeks post onset of his illness he regained a few degrees of eye movement in all directions, the ptosis resolved, and he was able to open his eyes. He also gained more control of his mouth and lips, resulting in improvement in his speech, but he was unable to smile or whistle. **Follow-Up:** The patient was seen in the clinic at 5 weeks, 12 weeks, and 4 months after the onset of his illness. By 4 months, the ptosis and ophthalmoplegia had resolved completely and he had minimal facial weakness and speech was 90% back to normal. He had 5/5 muscle strength and no ataxia.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2004
Date Original	1993
Resource.Type	Video
Format.Creation	
Source	3/4 Umatic master videotape
Language	En
Relation-Is Part Of	925-2, 944-5
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002-2004. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No

Collection	Neuro-ophthalmology
Catalogue	http://medstat.med.utah.edu/neuroophth/Wray/
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist Steve Smith, Videographer
Reviewer	
Context URL	http://medstat.med.utah.edu/neuroophth/Wray/
Clinical	This patient with a severe attack of **MFS** was followed over a period of 4 months. I had the opportunity to film him on four occasions. In the first clip, made on **day 15,** he had: Complete ptosis Total external ophthalmoplegia Inability to close his eyes tightly Inability to raise his eyebrows Severe facial diplegia with inability to show his teeth or smile Dysarthric speech due to inability to articulate his words The second clip was made at **5 weeks.** At this time: Ptosis had started to improve Patient was able to open and close his eyes Patient was not able to blink—definite paucity of blinking Facial diplegia still pronounced and unable to show his teeth External ophthalmoplegia beginning to improve Partial recovery of horizontal gaze to right and left and vertically Gait ataxia much improved and no limb ataxia The third clip was made at **12 weeks.** At this time he reported: General strength recovering Full use of his hands and arms Speech a good deal clearer Ptosis had recovered and he was aware of diplopia Able to close his eyes fairly tightly Able to raise his eyebrows up Able to grip his lips together but not whistle The fourth clip was made at **4 months.** At this time, there was striking improvement and the patient was a much happier man. Speech clearer but still not precise External ophthalmoplegia has completely recovered with full horizontal and vertical eye movements Lip movements improved Able to grip his lips together Eye closure now tight and able to open and close his eyes normally Able to raise his eyebrows up and frown Lower face weakness improved, able to show his teeth and smile No gait ataxia Reflexes 1+ throughout **In 1956 Miller Fisher** published a paper in the **New England Journal of Medicine** describing an unusual acute idiopathic polyneuritis characterized by: 1. Total external ophthalmoplegia; 2. Severe ataxia; and 3. Loss of the deep tendon reflexes. The nature of the illness was not recognized until he saw the third case, when, in

	association with a mild peripheral neuropathy, the cerebrospinal fluid showed an albuminocytologic dissociation with a total protein of 348 mg/100 mL and no cells. The syndrome Miller Fisher described proved to be a variant of acute idiopathic polyneuritis (GBS) in which limb involvement was minimal or absent. In two of Miller Fisher's cases (cases 1 and 2), external ophthalmoplegia was complete and the eyes fixed in primary gaze. Case 3 had bilateral sixth nerve paralysis and slight rotary nystagmus on attempted lateral gaze. Bilateral ptosis was added to the picture 2 days later. **GBS** is now viewed as a group of distinct disorders including the following variants. *Regional* Fisher syndrome of ophthalmoplegia, ataxia, and areflexia Cervico-brachial-pharyngeal, often with ptosis Oculopharyngeal weakness Predominant paraparesis Bilateral facial or abducens weakness with distal paresthesias Ophthalmoplegia with GQ_{1b} autoantibodies *Functional* Generalized ataxia without dysarthria or nystagmus Pure sensory Pure motor Pandysautonomia Axonal The degree of ophthalmoparesis is variable but certain patterns suggest involvement of either the peripheral or central nervous system. The ophthalmoplegia may resemble: Horizontal or vertical gaze palsy Internuclear ophthalmoplegia Ptosis is often absent even in the presence of significant ophthalmoparesis. Bell's phenomenon is often preserved even when vertical eye movements are absent. Signs pointing to cerebellar dysfunction are: Rebound nystagmus Impairment of smooth pursuit Suppression of the vestibular ocular reflex **Dr. Fisher** was himself impressed by the presence of ataxia unaccompanied by sensory loss and "reluctantly interpreted" the clinical signs as "manifestations of an unusual and unique disturbance of peripheral neurons."
Neuroimaging	No imaging studies are available in this patient. The MRI findings in another patient with MFS have been reported. The patient was a 25-year-old woman with complete external ophthalmoplegia, ptosis, limb ataxia, and areflexia. **Brain MRI:** T1-weighted images with Gd-DTPA on day 15 of her illness demonstrated enhancement of the posterior nerve roots of the cauda equina. MRI on **day 32** revealed swelling and enhancement of the bilateral ocular motor nerves as well as the facial nerves and the abducen nerves. The patient received high-dose IV immunoglobulin therapy and had marked improvement in her ophthalmoplegia. Repeat MRI with gadolinium, after recovery, showed no enhancement of the cauda equina nor of the cranial nerves. In this patient IgG anti-GQ1b and GD_{1b} antibodies were detected.
Anatomy	
Pathology	In **1993, Chiba and colleagues** reported the presence of serum IgG antibody against **ganglioside GQ_{1b}** in patients in the acute phase of MFS and pointed out

	that this immunologic feature was common to both MFS and GBS. To check for Anti-GQ_{1b} antibody, I sent the serum of this patient and the serum from a second case of MFS, a young man presenting with bilateral sixth nerve palsy (ID944-5), to **Professor Newsom-Davis** in Oxford to study. **Both these patients were found to have the Anti-GQ_{1b} IgG antibody.** Both **GBS** and the **MFS** are cell-mediated immunologic diseases directed at peripheral nerve resulting in acute inflammatory demyelinating neuropathy. **Autoantibodies:** A number of autoantibodies directed at components of nerve ganglioside are detected inconsistently in patients with GBS. Anti-GQ_{1b} IgG is the most important and is the autoantibody found in almost all patients with ophthalmoplegia. Antibodies against the ganglioside GQ_{1b} have also been detected in patients with Bickerstaff's brainstem encephalitis. **Bickerstaff's encephalitis** is characterized by ophthalmoplegia and ataxia but is also accompanied by pyramidal and sensory tract findings and cerebrospinal fluid pleocytosis. GM1 autoantibody may be found in approximately one-third of patients with GBS early in their course, corresponding in most instances to a predominantly motor presentation and to axonal damage. The highest titers of anti-GM1 antibody are usually associated with cases that follow *Campylobacter* infections. Autoantibodies directed against GD1a or GT1b have been associated in some cases with the pharyngeal-brachial-cervical variant. Pathologic studies in cases of GBS have failed to demonstrate any changes within the neuroaxis, and there is thus no morbid anatomic basis for attributing the clinical picture to a disturbance within the brainstem. Virtually all cases have shown perivascular (mainly perivenous) lymphocytic infiltrates scattered throughout the cranial nerves, ventral and dorsal nerve roots, and dorsal root ganglia and along the entire length of the peripheral nerves. Swelling of nerve roots at the site of their dural exit has been emphasized by some authors and theorized to cause root damage. In patient's whose electrophysiologic tests display severe axonal damage early in the illness, the pathologic findings corroborate the predominately axonal nature of the disease with secondary myelin damage and little inflammatory response.
Etiology	***Campylobacter jejuni*** may be the responsible trigger in GBS and MFS since anti-GQ1b antibodies bind to surface epitopes on this organism, and its lipopolysaccharide fraction may molecularly mimic the ganglioside.
Treatment	Specific treatment of the presumed immune disorder that underlies GBS and MFS include plasma exchange and IV immunoglobulin (0.4 g/kg per day for 5 consecutive days). Approximately 3–5% of patients do not survive an attack of GBS.
Disease/Diagnosis	Miller Fisher syndrome Acute inflammatory demyelinating neuropathy
Further Readings	Al-Din ASN et al: Brainstem encephalitis and the syndrome of Miller Fisher. Brain 105:481, 1982 Ang CW et al: Structure of *Campylobacter jejuni* lipopolysaccharides determines antiganglioside specificity and clinical features of Guillain-Barré and Miller Fisher patients. Infect Immun 70:1202, 2002 Landry de Thézillat JBO. Paralysies. Gazette hebdomadaire de médecine et de chirurgie 6: 472, 1859 Berlit P, Rakicky J: The Miller Fisher Syndrome. Review of the Literature. J Clin Neuroophthalmol 12:57, 1992 Caccavale A, Mignemi L: Acute onset of a bilateral areflexical mydriasis in Miller Fisher syndrome: A rare neuron-ophthalmologic disease. J Neuroophthalmol 20:61, 2000 Chiba A et al: Serum anti-GQ_{1b} IgG antibody is associated with ophthalmoplegia in Miller Fisher syndrome and Guillain-Barré syndrome: Clinical and immunohistochemical studies. Neurology 43:1911, 1993 Fisher CM: An unusual variant of acute idiopathic polyneuritis (syndrome of ophthalmoplegia, ataxia and areflexia). N Engl J Med 225:57, 1956 Guillain G et al: Le réflexe médico-plantaire: Étude de ses caracteres graphiques

et de son temps perdu. Bulletins et mémoires de la Société des Médecins des Hôpitaux de Paris 40:1459, 1916

——— et al: Sur un syndrome de radiculonévrite avec hyperalbuminose du liquide céphalo-rachidien sans réaction cellulaire. Remarques sur les caractères cliniques et graphiques des réflexes tendineux. Bulletins et mémoires de la Société desMédecins des Hôpitaux de Paris 40:1462, 1916

Hattori M et al: A case of Miller Fisher syndrome with gadolinium-enhancing lesions in the cranial nerves and the cauda equine on magnetic resonance imaging. Rinsho Shinkeigaku 39:1054, 1999

Jacobs BC et al: Immunoglobulins inhibit pathophysiological effects of anti-GQ1b-positive sera at motor nerve terminals through inhibition of antibody binding. Brain 126:2220, 2003

Keane JR: Bilateral seventh nerve palsy. Analysis of 43 cases and review of the literature. Neurology 44:1198, 1994

Leigh RJ, Zee DS: Diagnosis of peripheral ocular motor palsies and strabismus, in *The Neurology of Eye Movements* , 4th ed. New York, Oxford University Press, 2006, pp 385-474

Meienberg O, Ryffel E: Supranuclear eye movement disorders in Fisher's syndrome of ophthalmoplegia, ataxia, and areflexia. Report of a case and literature review. Arch Neurol 40:402, 1983

Mori M et al: Clinical features and prognosis of Miller Fisher syndrome. Neurology 56:1104, 2001

Newsom-Davis J: Myasthenia gravis and the Miller-Fisher variant of GBS. Curr Opin Neurol 10:18, 1997

Ohtsuka K et al: Fisher Syndrome associated with IgG Anti-GQ_{1b} antibody following infection by a specific serotype of *Campylobacter jejuni* . Ophthalmology 105:1281, 1998

Ropper, AH, Brown RH: Disease of the peripheral nerves, in *Adams and Victor's Principles of Neurology*, 8th ed, New York, McGraw Hill, 2005, pp 1110-1177

Sato K, Yoshikawa H: Bilateral abducens nerve paresis associated with anti-GQ_{1b} IgG antibody. Am J Ophthalmol 131:816, 2001

Shimamura H et al: Clinical, electrophysiological, and serological overlap between Miller Fisher syndrome and acute sensory ataxic neuropathy. Acta Neurol Scand 105:411, 2002

Shuaib A, Becker WJ: Variants of Guillain-Barré syndrome: Miller Fisher syndrome, facial diplegia and multiple cranial nerve palsies. Can J Neurol Sci 14:611, 1987

Susuki K et al: Acute facial diplegia and hyperreflexia: A Guillain-Barré syndrome variant. Neurology 62:825, 2004

Swick HM: Pseudointernuclear ophthalmoplegia in acute idiopathic polyneuritis (Fisher's syndrome). Am J Ophthalmol 77:725, 1974

Takahashi M et al: Epidemiology of Campylobacter jejuni isolated from patients with Guillain-Barré and Fisher syndrome in Japan. J Clin Microbiol 2005;43:335, 2005

van Doorn PA et al: Clinical features, pathogenesis, and treatment of Guillain-Barré syndrome. Lancet Neurol 7:939, 2008

Willison HJ and Plomp JJ: Anti-ganglioside antibodies and the presynaptic motor nerve terminal. Annals of the NY Acad Sci 2008;1132:114, 2008

Yuki N et al: Acute ophthalmoparesis (without ataxia) associated with anti-GQ_{1b} IgG antibody: clinical features. Ophthalmology 108:196, 2001

——— et al: An immunological abnormality common to Bickerstaff's brain stem encephalitis and Fisher's syndrome. J Neurol Sci 131:108, 1993

Cases with Vascular Disease

Video e11-11
Retinal Emboli (Film or Fundus) (ID16-1)
Video e11-12

Third Nerve Palsy (Microinfarct) (ID939-2)

Video e11-11

Metadata Element Field	Metadata
Identifier	16-1
Title	Retinal Emboli
Ocular Movements	
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Retinal Emboli Transient Monocular Blindness Ocular Stroke
Presenting Symptom	Transient monocular blindness
History	Temporary loss of vision in one eye, termed *transient monocular blindness* (TMB), is the most important visual symptom of arteriosclerotic vascular disease, arteritis, and states of altered coagulability and thrombocytosis. In most patients, the visual disturbance during each individual attack of TMB is stereotypic. It may recur over a period of months or over a much briefer span of hours, days, or weeks. A meticulous history of the attack and duration of the visual disturbance will permit classification of the TMB occurrence into one of four types. **Type I** is due to transient retinal ischemia. **Type II** is due to retinal vascular insufficiency. **Type III** is due to vasospasm. **Type IV** occurs in association with antiphospholipid antibodies but includes cases of unknown cause. (Wray, 1988, Table 7-1; Review ID937-2.)
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2008
Date Original	
Resource.Type	Video
Format.Creation Source	16-mm Film
Language	En
Relation-Is Part Of	937-2
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology / Neuro-ophthalmology
Catalogue	http://library.med.utah.edu/NOVEL/Wray
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist
Reviewer	
Context URL	http://library.med.utah.edu/NOVEL/Wray/
Clinical	A unique film of retinal emboli passing through the microcirculation of the retina recorded by Roger Lancaster (photographer) in 1967 with Dr. Cogan and his two fellows, Dr. Philip Zweifach and Dr. David Haining, is shown here. I believe this is the only film in existence showing moving emboli in branches of the central retinal artery.

	A retinal embolus is virtually diagnostic of localized disease of the ipsilateral internal carotid artery (ICA) when a typical carotid bruit is present, when aortic or cardiac disease is absent, and when there is no exogenous source of emboli, as, for example, in IV drug use (talc and microcrystalline cellulose), severe trauma, or injection (air). **Cholesterol emboli (Hollenhorst plaques)** appear in the branches of the central retinal artery as bright or shiny bodies whose diameter seems to exceed the intraluminal diameter of the arteriole. These emboli tend to lodge at arterial bifurcations. They may be invisible except on ocular compression or by varying the incidence angle of the ophthalmoscope light. They may be permanent or quite transient, moving on to the next bifurcation or disappearing before the next examiner can verify them. The presence of a cholesterol embolus is a poor prognostic sign: 93% of such patients have vascular disease at presentation; 15% die within the first year and 55% within 7 years. The cause of death is usually heart disease, 6:1 compared with stroke. Pale white platelet plugs can also be seen transiently within retinal arteries. In a hypertensive patient, the caliber of the retinal arteries on the side of an ICA stenosis may be reduced and will show fewer hypertensive changes than the retinal vessels of the opposite eye. Focal cotton-wool spots (cystoid bodies), in the absence of hypertensive retinopathy, are due to embolic microinfarction and may be seen when no emboli are visible.
Neuroimaging	
Anatomy	
Pathology	
Etiology	
Treatment	
Disease/Diagnosis	Transient Monocular Blindness Retinal Emboli
Further Readings	Beal MF et al: Cholesterol embolism as a cause of transient ischemic attacks and cerebral infarction. Neurology 31:860, 1981 Burde RM: Amaurosis fugax, an overview. J Clin Neuroophthalmol 9:185, 1989 Cogan DG, Wray SH: Vascular occlusions in the eye from cardiac myxomas. Am J Ophthalmol 80:396, 1975 Fisher CM: Transient monocular blindness associated with hemiplegia. Am Arch Ophthalmol 47:167, 1952 ———: Observations of the fundus oculi in transient monocular blindness. Neurology 9:333, 1959 Hollenhorst RW: The ocular manifestations of internal carotid arterial thrombosis. Med Clin North Am 44:897, 1960 ———: Significance of bright plaques in the retinal arterioles. JAMA 178:123, 1961 Savino PJ et al: Retinal stroke: Is the patient at risk? Arch Ophthalmol 95:1185, 1977 Wray SH: Extracranial internal carotid artery disease, in *Amaurosis Fugax*, EF Bernstein (ed). New York, Springer-Verlag 1988, pp 72-80 ———: Visual Symptoms, in *Stroke Syndromes,* 2nd ed, J Bogousslavsky, L Caplan (eds). Cambridge, United Kingdom, Cambridge University Press, 2001, pp 111-128.

Video e11-12

Metadata Element Field	**Metadata**
Identifier	939-2
Title	Third Nerve Palsy
Ocular Movements	Ptosis; Third Nerve Palsy;
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Ptosis;

	Third Nerve Palsy; Microinfarct Third Nerve; Oculomotor Nerve;
Presenting Symptom	Double vision
History	The patient is a 57-year-old man who carried a diagnosis of atrial fibrillation and coronary artery disease post CABG. He was seen in the Massachusetts General Hospital ER with acute double vision and headache and was admitted. Four days prior to admission (PTA) he developed a bifrontal headache accompanied by double vision looking down. He reported that the double vision involved "diagonal images" and was worse on looking to the left. He had no nausea, vomiting, or eye pain. One day PTA he developed ptosis of the left eye (OS). **Past History:** Notable for coronary artery disease with multivessel coronary artery bypass grafts in 1991. Hypercholesterolemia Atrial fibrillation (not on anticoagulation) A history of previous episodes of diplopia Ten years PTA, he developed diplopia on left lateral gaze lasting three days. He was seen by an ophthalmologist but no diagnosis was made. His vision returned to normal. One year PTA, he had another attack of double vision and was seen by an ophthalmologist who diagnosed a left sixth nerve palsy. Workup included a brain MRI, which was normal. The sixth nerve palsy recovered completely. **Social History:** Heavy cigarette smoker in the past. **General Examination:** Normal. BP 120/70 Temporal artery pulses normal No carotid or orbital bruits **Neuro-Ophthalmologic Examination:** Visual acuity 20/20 OU Pupils equal reacting briskly to light and near Visual fields and fundus examination normal **Ocular Motility:** Ptosis OS Marked paresis of superior rectus and inferior oblique Able to elevate the eye only 10 degrees above the horizontal meridian. Paresis of medial rectus, able to adduct only 20 degrees past the midline. Inferior rectus paretic, able to depress the eye only 20 degrees. Cranial nerves 4 and 6 intact No proptosis or ocular pulsation Ocular motility OD normal **Motor System:** Normal **Sensory System:** Normal **Brain MRI:** Normal **MR Angiogram:** Normal **Blood Studies:** Complete blood count, differential, and platelet count normal Erythrocyte sedimentation rate normal C-reactive protein and fibrinogen normal Tests for diabetes negative Test for syphilis negative **Chest X-Ray:** Normal

	Lumbar Puncture: Cerebrospinal fluid clear Protein 39 mg/mL Sugar 77 mg/dL No cells **Diagnosis:** Left third nerve palsy involving the nerve trunk and sparing the pupil **Etiology:** Microinfarction Close observation of all patients with a third nerve palsy, particularly those with a progressive history, as in this patient, should be watched carefully with pupil examinations checked regularly. Anisocoria greater than 2 mm may be considered grounds for an arteriogram. **Prognosis for Recovery:** Good The patient was discharged on the third hospital day. He returned 5 weeks later 95% recovered with full eye movements and only partial left ptosis. One month later he was fully recovered.
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2004
Date Original	1996
Resource.Type	Video
Format.Creation	
Source	3/4 Umatic master videotape
Language	En
Relation-Is Part Of	5-1, 932-6, 940-3, 946-6
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002-2004. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neuro-ophthalmology
Catalogue	http://medstat.med.utah.edu/neuroophth/Wray/
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Ray Balhorn, Video Compressionist Steve Smith, Videographer
Reviewer	
Context URL	http://medstat.med.utah.edu/neuroophth/Wray/
Clinical	This patient with a microinfarct of the trunk of the left third nerve had: • Ptosis OS • Paresis of all the muscles innervated by the third nerve, with marked paresis of the inferior oblique and superior rectus In primary gaze, left eye deviating down and out due to the unopposed action of the intact lateral rectus and superior oblique muscles Pupil normal A second video clip taken 5 weeks after the onset of the palsy showed 95% recovery with: Minimal ptosis OS Minimal paresis of elevation of the eye Cover/uncover test showed an alternating exophoria **Review alongside this case:** **ID163-21** Nuclear third nerve palsy **ID919-2** Nuclear third nerve palsy with isolated bilateral ptosis **ID166-25** Fascicular third nerve palsy—Claude's syndrome
Neuroimaging	Normal studies

Anatomy	The diagnosis of a third nerve palsy is straight forward but it is important to consider whether it is: 1. A nuclear lesion; 2. A complete or partial lesion of the nerve trunk; or a 3. Superior division of the third nerve; or an 4. Inferior division of the third nerve. A lesion involving the superior division of the third nerve results in paresis of the levator palpebrae muscle and the superior rectus so that the patient will have partial ptosis and paresis of elevation in the line of action of the superior rectus. A lesion involving the inferior division of the third nerve involves all the extraocular muscles innervated by the third nerve, except the levator palpebrae and the superior rectus, with or without pupil involvement.
Pathology	
Etiology	In adults, the most common cause of a progressive painful pupil sparing the third nerve palsy is microinfarction of the nerve in association with diabetes, hypertension, temporal arteritis, or syphilis. The most common cause of a progressive painful third nerve palsy **involving the pupil** is an aneurysm of the posterior communicating or posterior cerebral artery, until proved otherwise and an emergency CT angiogram and/or MR angiogram is indicated. Patients with partial involvement of the pupil and complete paresis of the extraocular muscles and eyelid should also undergo MRI/MRA study and be closely observed. MRA will reveal some but not all aneurysms compressing the third nerve.
Treatment	Control of risk factors for stroke
Disease/Diagnosis	Third nerve palsy Microinfarction of the nerve
Further Readings	Blake PY et al: MR of oculomotor nerve. Am J Neuroradiol 16:1665, 1995 Chou KL et al: Acute ocular motor mononeuropathies: Prospective study of the roles of neuroimaging and clinical assessment. J Neurol Sci 219:35, 2004 Ettl A, Salomonowitz E: Visualization of the oculomotor cranial nerves by magnetic resonance imaging. Strabismus 12:85, 2004 Eustace P: Partial nuclear third nerve palsies. Neuro-ophthalmology 1985;5:259, 1985 Leigh JR, Zee DS: Diagnosis of peripheral ocular motor palsy and strabismus, in *The Neurology of Eye Movements* , 4th ed. New York, Oxford University Press, 2006, pp 385-474

Case with Neurodegenerative Disease

Video e11-13
Apraxia of Eyelid Opening (Progressive Supranuclear Palsy) (ID932-3)

Video e11-13	
Metadata Element Field	**Metadata**
Identifier	932-3
Title	Apraxia of Eyelid Opening
Ocular Movements	Apraxia of Eyelid Opening Impaired Initiation of Horizontal Saccades Slow Hypometric Horizontal Saccades Supranuclear Paralysis of Up and Downgaze Square Wave Jerks
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital

Subject	Apraxia of Eyelid Opening Impaired Initiation of Horizontal Saccades Slow Hypometric Horizontal Saccades Supranuclear Paralysis of Up and Downgaze Square wave Jerks Progressive Supranuclear Palsy—Tauopthay Steele Richardson Olszewski Syndrome CNS—Degeneration
Presenting symptom	Frequent falls
History	**In January 1997,** this 73-year-old patient was referred to the Neurovisual Clinic. At that time his speech was slurred and he stated that his eyes were his "biggest" complaint because: 1. He had impaired focusing "close up"; 2. His eyes shut spontaneously much of the time; 3. Bright sunlight provoked eye closure; 4. The frequent closure of his eyes made it difficult for him to hold a conversation with anybody; and 5. With his eyes open his vision was clear. **Neuro-Ophthalmologic Examination** documented: Visual acuity: 20/40 OU **Eyelids:** Age-related bilateral ptosis Infrequent blinking Positive glabella tap Able to close his eyes but very slow opening his eyes Mild blepharoclonus **Ocular Motility:** Slow hypometric horizontal saccades Slow vertical saccades Absent convergence Saccadic pursuit in all directions Convergence insufficiency with exophoria at near Within 6 months his wife noted that his condition had deteriorated in that he: Seemed to have lost his motivation Rarely spoke unless spoken to Tended to sit longer without moving Did not appear to wish to read anything Had more difficulty feeding himself and dressing Was much slower in all his movements **Neurologic Examination** at that time showed: Striking paucity of movements Rigidity of the neck Difficulty getting up out of a chair, toppling backwards Slowness walking, needing to make several small steps in order to turn Speech that had become dysphonic more than dysarthric Tongue that moved well Absence of jaw jerk and facial jerks **Diagnosis:**

	Progressive supranuclear palsy
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2004
Date Original	1996
Resource.Type	Video
Format.Creation	
Source	3/4" Umatic master videotape
Language	En
Relation-Is Part of	166-18,168-3, 924-2, 936-5, 939-3
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002-2004. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/neuroophth/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neurology/Neuro-ophthalmology
Catalogue	http://medstat.med.utah.edu/neuroophth/Wray/
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributor Secondary	Tessa Hedley-Whyte, MD, Massachusetts General Hospital, Pathology; Anne Osborn, MD, University of Utah, Neuroimaging; Ray Balhorn, Video Compressionist Steve Smith, Videographer
Reviewers	David Zee, MD, Johns Hopkins University, 2005
Context URL	http://medstat.med.utah.edu/neuroophth/Wray/
Clinical	This patient with progressive supranuclear palsy (PSP) has: • Supranuclear saccadic and pursuit paralysis of upward gaze • Square wave jerks looking up • Slow hypometric horizontal saccades • Slow eyelid closure and impaired voluntary eyelid opening with striking inability to open his eyes on command (apraxia) • Dysphonic and dysarthric speech **Comment:** The inability to initiate eyelid opening is a dysfunction of voluntary lid control due, in this patient, to disease of the extrapyramidal system. The patient can close his eyes slowly on command and reopen them normally after blinking. However, the major problem was a difficulty in voluntarily opening his eyes on command (eyelid apraxia). The term *eyelid apraxia* is a misnomer and should not be used to describe the eyelid dysfunction in this case of PSP because the motor system is affected. Confirmation that the vertical gaze disorder is supranuclear and localized to the rostral interstitial nucleus of the MLF (riMLF) in the midbrain is the presence of: 1. Upward deviation of the eyes on forced eye closure (intact Bell's) and 2. Full upward eye movements when the head is bent forward—the oculocephalic or doll's eye reflex. At the onset of PSP: • Vertical saccades are slow • Vertical saccadic range is progressively reduced • Initiation of vertical saccades is impaired • Vertical smooth pursuit is impaired (reduced range) • Vertical optokinetic stimulation can cause gaze to tonically deviate in the direction of the stripe movement (Personal communication, Zee 2005.) Additional PSP signs are:

	1. Positive glabella tap (inability to inhibit a blink when the forehead is tapped) 2. Myerson's sign (inability to inhibit a blink to a bright pen light shown in the eyes) 3. Blepharoclonus (tremor of the lids on gentle eye closure) 4. Square wave jerks 5. Bilateral ophthalmoparesis in the late stages of the disease
Neuroimaging	PSP has characteristic changes on neuroimaging. In another case a sagittal T2-weighted MR scan shows the tectal plate is markedly thinned and atrophic **(Figure 1).** **Functional MRI** reveals global metabolic reduction most pronounced in the frontal lobes, anterior cingulate gyrus, the basal ganglia, the ventrolateral and dorsomedial nuclei of thalamus, and the upper brainstem. **PET** scans using fluorodopa demonstrate diminished striatal dopamine formation and storage. In addition to hypometabolism in the putamen, severe caudate involvement on PET scanning distinguishes PSP from Parkinson's disease **(Figures 2 and 3)**.
Anatomy	Supranuclear paralysis of vertical gaze localizes to the midbrain and to the rostral interstitial nucleus of the medial longitudinal fasciculus (riMLF). The riMLF is a wing-shaped structure that lies dorsomedial to the red nucleus and rostral to the interstitial nucleus of Cajal. The riMLF contains burst neurons for vertical and torsional saccades. The riMLF projects predominantly to the ipsilateral oculomotor (third nerve) and trochlear (fourth nerve) nuclei.
Pathology	In an autopsy case of PSP, a pale locus coeruleus and substantia nigra are two typical gross features on inspection of the brainstem. Neuronal loss, granulovacuolar degeneration, and fibrillary gliosis are present in areas of neuronal change. There is widespread neuronal and glial tau accumulation in the cortex, basal ganglia, periaqueductal gray matter, subthalamic nucleus, red nucleus, substantia nigra, pedunculopontine nucleus, superior colliculus, and the dentate nucleus of the cerebellum. Microscopically, globose neurofibrillary tangles are observed on LHE stain. The neuronal cytoplasmic inclusions are strongly immunoreactive for tau. Tufted astrocytes are a specific finding in PSP. They are often binucleate and have long tau reactive processes. The intracellular aggregation of tau in PSP may be sufficient to cause nerve cell degeneration.
Etiology	PSP is a **"tauopathy."** Studies suggest that it is a recessive disorder in linkage disequilibrium with the tau gene. Rare familial forms of PSP exist including an autosomal dominant transmission with incomplete penetrance. The relationship to the tau gene further suggests a relationship to frontotemporal dementia/Pick's disease (FTDP), and some families carrying the FTDP-17 mutation (chromosome 17) have affected members with PSP-like phenotypes.
Treatment	There is no cure for PSP. Once the disease has begun, its course is relentlessly progressive.
Disease/Diagnosis	Progressive Supranuclear Palsy—tauopathy
Further Readings	Buttner-Ennever JA, Horn AK: Pathways from cell groups of the paramedian tracts to the floccular region. Ann NY Acad Sci 781:532, 1996 Daniel SE et al: The clinical and pathological spectrum of Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy): A reappraisal. Brain 118:759, 1995 Friedman DI et al: Neuro-ophthalmic findings in progressive supranuclear palsy. J Clin Neuroophthalmol 12:104, 1992 Growdon JH, Rossor MN: *The Dementias. Blue Books of Practical Neurology* , Vol 19. Philadelphia, Butterworth-Heinemann, 1998 Leigh RJ, Zee DS: Diagnosis of central disorders of ocular motility, in *The Neurology of Eye Movements*, 4th ed. New York, Oxford University Press, 2006, pp 598-718

Mendez MG, Cummings JL: *Dementia: A Clinical Approach* . 3rd ed. Philadelphia, Butterworth Heinemann, 2003

Richardson JC et al: Supranuclear ophthalmoplegia, pseudobulbar palsy, nuchal dystonia and dementia. A clinical report on eight cases of heterogenous system degeneration. Trans Am Neurol Assoc 88:25, 1963

Stanford PM et al: Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene: Expansion of the disease phenotype caused by tau gene mutations. Brain 123:880, 2000

Sir Charles Bell (*http://www.whonamedit.com/doctor.cfm/2103.html?*)

Case of Thyroid Associated Ophthalmopathy

Video e11-14
Restrictive Orbitopathy of Graves Disease, Bilateral Exophthalmos (ID925-4)

Video e11-14	
Metadata Element Field	**Metadata**
Identifier	925-4
Title	Restrictive Orbitopathy
Ocular Movements	Bilateral Lid Retraction Lid Lag Restricted Upgaze
Creator	Shirley H. Wray, MB, ChB, PhD, FRCP, Professor of Neurology, Harvard Medical School; Director, Unit for Neurovisual Disorders, Massachusetts General Hospital
Subject	Bilateral Lid Retraction Lid Lag Bilateral Exophthalmos Restrictive Orbitopathy of Graves' disease
Presenting Symptom	Prominent eyes
History	This 71-year-old woman was referred with bilateral optic neuropathy and thyroid-associated ophthalmopathy (TAO) of Graves' disease. She had been treated for primary hyperthyroidism on three occasions with radioactive iodine and was taking methimazole 5 mg daily. **Neuro-Ophthalmologic Examination:** Vision was reduced to 20/200 in each eye with bilateral central scotoma and mild disc hyperemia. She had the classical signs of Graves' disease: A prominent stare. Retraction of all four eyelids Bilateral exophthalmos Hertel exophthalmometer 25 OD, 28 OS, base 108. Tight orbits/reduced orbital resilience Restricted horizontal eye movements Marked limitation of upward gaze Mild limitation of downgaze Lid lag (persistent elevation of the upper eyelid in downgaze)—von Graefe sign Positive forced duction test Prominent congested scleral blood vessels A visible rim of sclera on gentle eye closure Bell's reflex absent, eyes fail to move up under closed lids)

	Convergence normal **Investigations:** Thyroid tests showed TT3 elevated to 243 (6/7/90) and 324 (7/3/90). **CT Orbits:** Greatly enlarged extraocular muscles crowding the optic nerves at the apex Considerable bilateral proptosis right > left. The inferior rectus and medial rectus muscles were especially enlarged with fusiform dilatation of the midposition of the muscle. **Diagnosis:** Advanced Graves' disease Thyroid-associated orbitopathy Bilateral compressive optic neuropathy **Therapy:** Patient received a course of oral steroids **Surgery:** Bilateral orbital decompression and ethmoidectomy. **CT Orbits Post-Op:** The scan confirmed adequate removal of the medial orbital walls and the orbital floors over the maxillary sinuses. The surgeon also partly divided the levator muscles because of severe upper lid retraction. Her vision recovered to 20/40 OD and 20/30 OS. **Comment:** **Compressive optic neuropathy** is the most serious complication of TAO. The incidence of visual loss is between 2% and 9% in all patients with TAO. However, in patients with "severe" TAO, requiring orbital decompression, as in this case, optic neuropathy occurs in up to 50% of patients. **Vision loss from optic nerve compression requires immediate management.**
Publisher	Spencer S. Eccles Health Sciences Library, University of Utah
Date Digital	2004
Date Original	1990
Resource.Type	Video
Format.Use Format.Creation	Video/RealMedia/0:0:00 minutes; Video/QuickTime/0:0:00 minutes;
Source	¾" Umatic master videotape
Language	En
Relation-Is Part of	4-3
Rights Management	Shirley H. Wray, MB, ChB, PhD, FRCP, Copyright 2002. For further information regarding the rights to this collection, please visit http://medlib.med.utah.edu/NOVEL/Wray/copyright.htm
Holding.Institution	Spencer S. Eccles Health Sciences Library, University of Utah, 10 N 1900 E, Salt Lake City, UT 84112-5890
Annotation	No
Collection	Neuro-ophthalmology
Catalogue	http://medstat.med.utah.edu/NOVEL/Wray/
Contributor Primary	Shirley H. Wray, MB, ChB, PhD, FRCP
Contributory Secondary	Ray Balhorn, Video Compressionist Hugh Curtin, M.D., Massachusetts Eye and Ear Infirmary, Boston, MA Steve Smith, Videographer
Reviewer	Arthur Grove, MD, Massachusetts Eye and Ear Infirmary, Boston, MA.
Context URL	http://medstat.med.utah.edu/NOVEL/Wray/
Clinical	This 71-year-old woman with TAO of Graves' disease has: A prominent stare Retraction of all four eyelids Bilateral exophthalmos

	Hertel exophthalmometer 25 OD, 28 OS, base 108. Tight orbits/reduced orbital resilience Restricted horizontal eye movements Marked limitation of upgaze Mild limitation of downgaze Lid lag (persistent elevation of the upper eyelid in downgaze)—von Graefe sign Positive forced duction test Prominent congested scleral blood vessels A visible rim of sclera on gentle eye closure Also illustrated is: • The use of a Hertel exophthalmometer to measure the forward protrusion of the proptotic eye • How to evaluate reduced orbital resilience by digital pressure on the globe • The absent Bell's reflex **Comment:** TAO can be differentiated from ocular myasthenia gravis by the lack of ptosis and the presence of proptosis, lid retraction, lid-lag, and periorbital edema. TAO can, however, co-exist with ocular myasthenia gravis. As a result, screening thyroid studies are essential prior to treatment even when the clinical diagnosis of TAO together with ocular myasthenia gravis seems clear. **Forced Duction Test:** In TAO the limitation of upgaze is due to tethering of the eyeball in the floor of the orbit by soft tissue changes. Tethering of the eyeball inferiorly can be confirmed by performing a forced duction test. The test requires anesthetizing the eyeball with topical anesthesia. Inability to move the eye up despite pushing on the globe with a cotton tip swab or pulling with a small pair of blunt tweezers, suggests mechanical restriction—interpreted as a positive forced duction test.
Neuroimaging	CT of the orbit is the gold standard for the diagnosis of TAO. The classic finding is enlargement of the extraocular muscle belly with relative sparing of the tendon. Proptosis may be recognized without extraocular muscle enlargement, presumably resulting from an increased volume of intraorbital fat. To order an MRI of the orbit, the recommended sequence is short tau inversion recovery (STIR, to highlight the extraocular muscles. **CT Orbit:** Images in another case of TAO showed on **Figure 1:** Axial CT through the orbit without contrast shows enlargement of the medial rectus muscle bilaterally. The tendinous insertion is spared. **Figure 2:** The coronal CT (reformatted from axial data set) without contrast showed enlargement of the medial rectus muscle, inferior rectus muscle, and upper muscle complex on both sides.
Anatomy	Orbit—enlargement of extraocular muscles
Pathology	Graves' disease is an autoimmune condition. For unknown reasons, the extraocular muscles develop lymphocytic and plasmacytic infiltration with secondary production of acid mucopolysaccharides. In the acute stages, the changes are largely inflammatory. In the chronic inactive stage, there is often fatty infiltration of muscles.
Etiology	Autoimmune disorder
Treatment	Treatment of TAO associated with Graves' disease is extremely successful. Irritation and swelling can be treated with a short (1- to 2-month) course of systemic corticosteroids or with low-dose (1500 to 2000 cGy) orbital radiation therapy. Proptosis can be treated with orbital decompression using a variety of techniques. (For a full discussion of therapy, see reference Galetta et al, 1996; Jacobson, 1995; and Riddick, 1991.)
Disease/Diagnosis	Restrictive orbitopathy of Graves' Disease
Further Readings	Bahn RS, Heufelder AE: Pathogenesis of Graves' ophthalmopathy. N Engl J Med 329:1468, 1993

Galetta SL et al: Pretectal eyelid retraction and lag. Ann Neurol 33:554, 1993

——— et al: Eyelid lag without eyelid retraction in pretectal disease. J Neuroophthalmol 16:96, 1996

Garrity JA et al: Results of transantral orbital decompression in 428 patients with severe Graves' ophthalmopathy. Am J Ophthalmol 116:533, 1993

Hoffman PN: in *Walsh and Hoyt's Clinical Neuro-Ophthalmology*, 6th ed, NR Miller, NJ Newman (eds). Philadelphia, Lippincott Williams & Wilkins, 2005

Jacobson DM. Acetylcholine receptor antibodies in patients with Graves' ophthalmopathy. J Neuroophthalmol 15:166, 1995

Kennerdell JS et al: Apical optic nerve compression of dysthyroid optic neuropathy on computed tomography. Arch Ophthalmol 99:807, 1981

Liu GT et al: Orbit disease in neuro-ophthalmology, in *Neuro-ophthalmology Diagnosis and Management*. Philadelphia, WB Saunders Company, 2001

Neigel JM et al: Dysthyroid optic neuropathy. The crowded orbital apex syndrome. Ophthalmology 95:1515, 1988

Riddick FA Jr: Update on thyroid diseases. Ophthalmology 88:467, 1981

Trobe JD et al: Dysthyroid optic neuropathy: Clinical profile and rationale for management. Arch Ophthalmol 96:1199, 1978

Wiersinga WM: Advances in medical therapy of thyroid-associated ophthalmopathy. Orbit 15:177, 1996

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF SMELL AND TASTE: INTRODUCTION

All environmental chemicals necessary for life enter the body by the nose and mouth. The senses of smell (olfaction) and taste (gustation) monitor those chemicals, determine the flavor and palatability of foods and beverages, and warn of dangerous environmental conditions, including fire, air pollution, leaking natural gas, and bacteria-laden foodstuffs. These senses contribute significantly to quality of life and, when dysfunctional, can have untoward physical and psychological consequences. A basic understanding of these senses in health and disease is critical for the physician, since thousands of patients present to doctors' offices each year with complaints of chemosensory dysfunction. Among the more important developments in neurology has been the discovery that decreased smell function is perhaps the first sign of neurodegenerative diseases such as Alzheimer's disease (AD) and Parkinson's disease (PD), signifying their "presymptomatic" phase.

Anatomy and Physiology

OLFACTORY SYSTEM

Odorous chemicals enter the nose during inhalation and active sniffing as well as during deglutition. After reaching the highest recesses of the nasal cavity, they dissolve in the olfactory mucus and diffuse or are actively transported to receptors on the cilia of olfactory receptor cells. The cilia, dendrites, cell bodies, and proximal axonal segments of these bipolar cells are situated within a specialized neuroepithelium that covers the cribriform plate, the superior nasal septum, the superior turbinate, and sectors of the middle turbinate (**Fig. 29-1**). Each of the ~6 million bipolar receptor cells expresses only one of ~450 receptor protein types, most of which respond to more than a single chemical. When damaged, the receptor cells can be replaced by stem cells near the basement membrane. Unfortunately, such replacement is often incomplete.

Figure 29-1

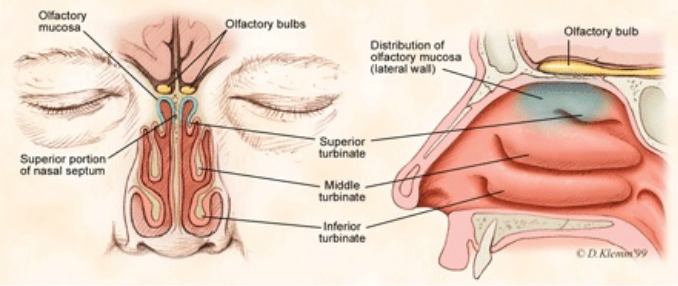

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anatomy of the olfactory neural pathways, showing the distribution of olfactory receptors in the roof of the nasal cavity. *[Copyright David Klemm, Faculty and Curriculum Support (FACS), Georgetown University Medical Center; used with permission.]*

After coalescing into bundles surrounded by glia-like ensheathing cells (termed fila), the receptor cell axons pass through the cribriform plate to the olfactory bulbs, where they synapse with dendrites of other cell types within the glomeruli (**Fig. 29-2**). These spherical structures, which make up a distinct layer of the olfactory bulb, are a site of convergence of information, since many more fibers enter than leave them. Receptor cells that express the same type of receptor project to the same glomeruli, effectively making each glomerulus a functional unit. The major projection neurons of the olfactory system—the mitral and tufted cells—send primary dendrites into the glomeruli, connecting not only with the incoming receptor cell axons but with dendrites of periglomerular cells. The activity of the mitral/tufted cells is modulated by the periglomerular cells, secondary dendrites from other mitral/tufted cells, and granule cells, the most numerous cells of the bulb. The latter cells, which are largely GABAergic, receive inputs from central brain structures and modulate the output of the mitral/tufted cells. Interestingly, like the olfactory receptor cells, some cells within the bulb undergo replacement. Thus, neuroblasts formed within the anterior subventricular zone of the brain migrate along the rostral

migratory stream, ultimately becoming granule and periglomerular cells.

Figure 29-2

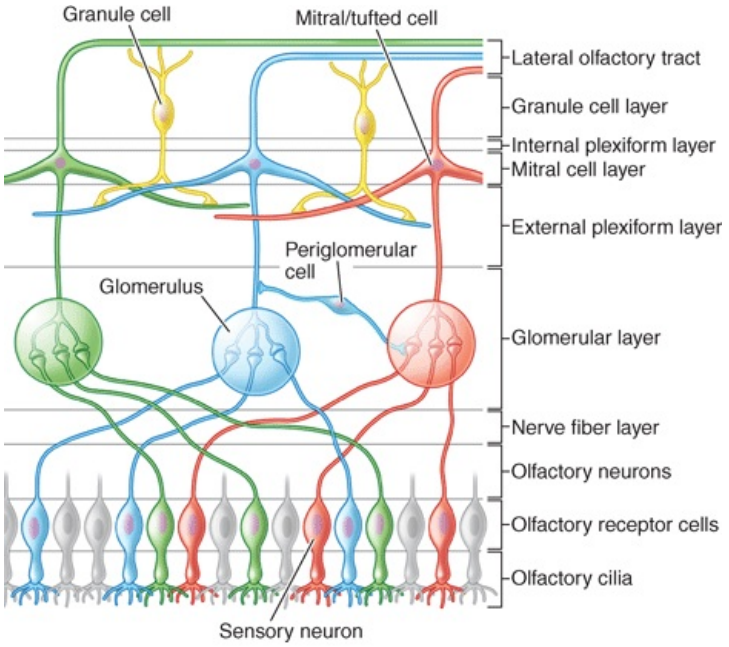

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic of the layers and wiring of the olfactory bulb. Each receptor type (red, green, blue) projects to a common glomerulus. The neural activity within each glomerulus is modulated by periglomerular cells. The activity of the primary projection cells, the mitral and tufted cells, is modulated by granule cells, periglomerular cells, and secondary dendrites from other mitral and tufted cells. *(From www.med.yale.edu/neurosurg/treloar/index.html.)*

The axons of the mitral and tufted cells synapse within the primary olfactory cortex (POC) (**Fig. 29-3**). The POC is defined as the cortical structures that receive direct projections from the olfactory bulb, most notably the piriform and entorhinal cortices. Although olfaction is unique in that its initial afferent projections bypass the thalamus, persons with damage to the thalamus can exhibit olfactory deficits, particularly ones of odor identification. Those deficits probably reflect the involvement of thalamic connections between the primary olfactory cortex and the orbitofrontal cortex (OFC), where odor identification occurs. The close anatomic ties between the olfactory system and the amygdala, hippocampus, and hypothalamus help explain the intimate associations between odor perception and cognitive functions such as memory, motivation, arousal, autonomic activity, digestion, and sex.

Figure 29-3

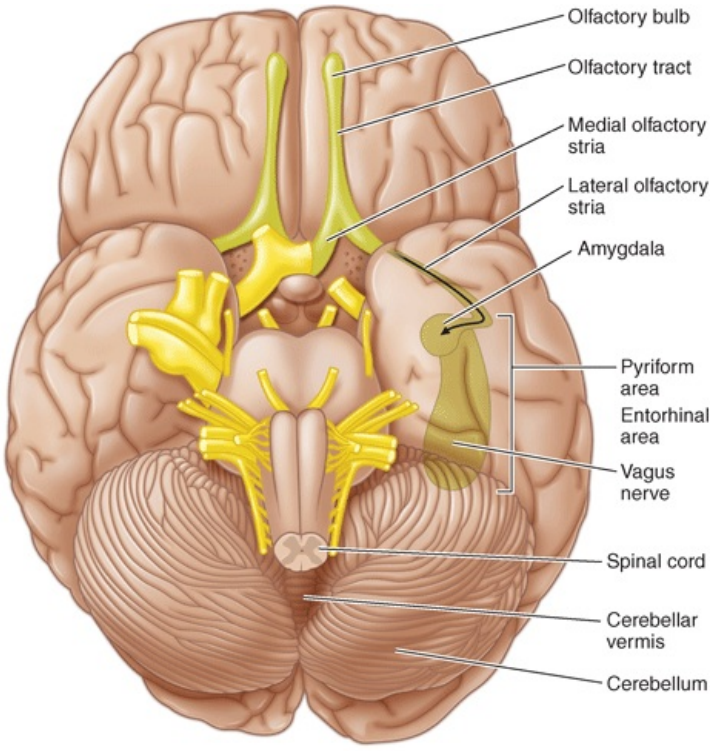

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anatomy of the base of the brain showing the primary olfactory cortex.

TASTE SYSTEM

Tastants are sensed by specialized receptor cells present within taste buds: small grapefruit-like segmented structures on the lateral margins and dorsum of the tongue, the roof of the mouth, the pharynx, the larynx, and the superior esophagus (**Fig. 29-4**). Lingual taste buds are embedded in well-defined protuberances termed fungiform, foliate, and circumvallate papillae. After dissolving in a liquid, tastants enter the opening of the taste bud—the taste pore—and bind to receptors on microvilli, small extensions of receptor cells within each taste bud. Such binding changes the electrical potential across the taste cell, resulting in neurotransmitter release onto the first-order taste neurons. Although humans have ~7500 taste buds, not all harbor taste-sensitive cells; some contain only one class of receptor (e.g., cells responsive only to sugars), whereas others contain cells sensitive to more than one class. The number of taste receptor cells per taste bud ranges from zero to well over 100. A small family of three G-protein-coupled receptors (GPCRs)—T1R1, T1R2, and T1R3—mediate sweet and umami taste sensations. Umami ("savory") refers to the flavors of meat, cheese, and broth due to glutamate and related compounds. Bitter sensations, in contrast, depend on T2R receptors, a family of ~30 GPCRs expressed on cells different from those which express the sweet and umami receptors. T2Rs sense a wide range of bitter substances but do not distinguish among them. Sour tastants are sensed by the PKD2L1 receptor, a member of the transient receptor potential protein (TRP) family. Perception of salty sensations, such as those induced by sodium chloride, arises from the entry of Na^+ ions into the cells via specialized membrane channels such as the amiloride-sensitive Na^+ channel.

Figure 29-4

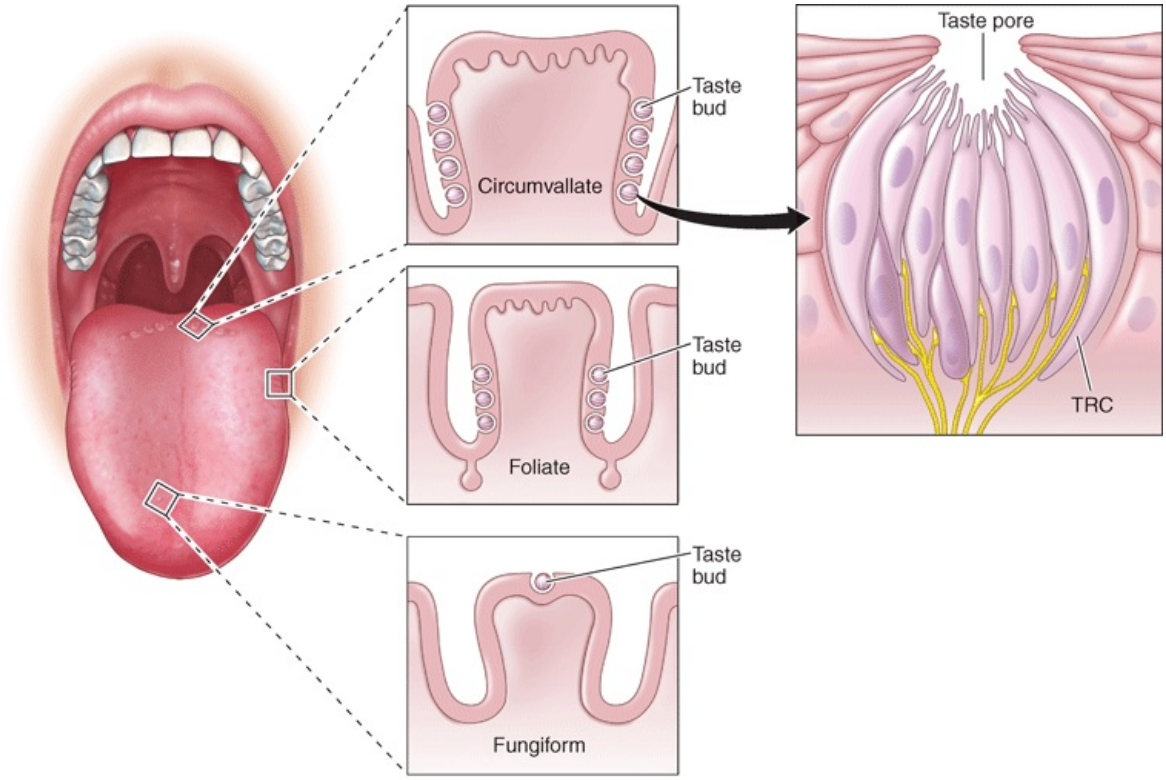

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic of the taste bud and its opening (pore), as well as the location of buds on the three major types of papillae: fungiform (anterior), foliate (lateral), and circumvallate (posterior). TRC, taste receptor cell.

Taste information is sent to the brain via three cranial nerves (CNs): CN VII (the *facial nerve,* which involves the intermediate nerve with its branches, the greater petrosal and chorda tympani nerves); CN IX (the *glossopharyngeal nerve*); and CN X (the *vagus nerve*) (**Fig. 29-5**). CN VII innervates the anterior tongue and all of the soft palate, CN IX innervates the posterior tongue, and CN X innervates the laryngeal surface of the epiglottis, the larynx, and the proximal portion of the esophagus. The mandibular branch of CN V (V_3) conveys somatosensory information (e.g., touch, burning, cooling, irritation) to the brain. Although not technically a gustatory nerve, CN V shares primary nerve routes with many of the gustatory nerve fibers and adds temperature, texture, pungency, and spiciness to the taste experience. The chorda tympani nerve is notable for taking a recurrent course through the facial canal in the petrosal portion of the temporal bone, passing through the middle ear, then exiting the skull via the petrotympanic fissure, where it joins the lingual nerve (a division of CN V) near the tongue. This nerve also carries parasympathetic fibers to the submandibular and sublingual glands, whereas the greater petrosal nerve supplies the palatine glands, thereby influencing saliva production.

Figure 29-5

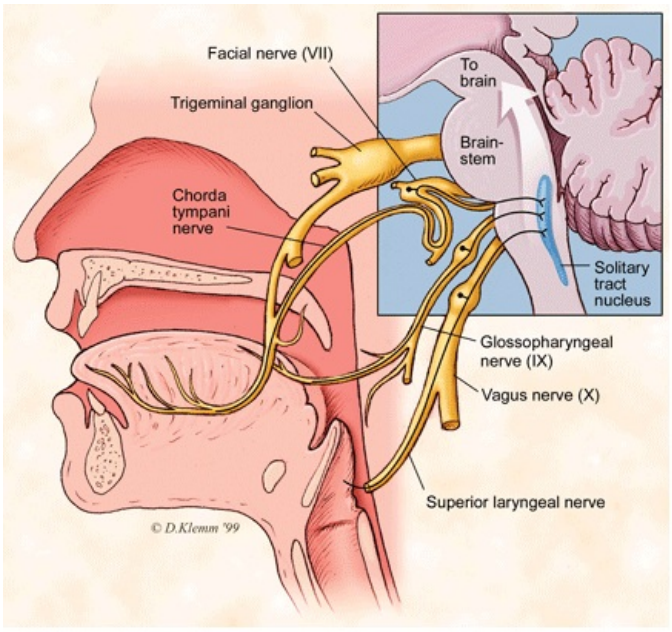

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic of the cranial nerves that mediate taste function, including the chorda tympani nerve (CN VII), the glossopharyngeal nerve (CN IX), and the vagus nerve (CN X). *[Copyright David Klemm, Faculty and Curriculum Support (FACS), Georgetown University Medical Center; used with permission.]*

The axons of the projection cells that synapse with taste buds enter the rostral portion of the nucleus of the solitary tract (NTS) within the medulla of the brainstem (Fig. 29-5). From the NTS, neurons then project to a division of the ventroposteromedial thalamic nucleus (VPM) via the medial lemniscus. From there projections are made to the rostral part of the frontal operculum and adjoining insula, a brain region considered the *primary taste cortex* (PTC). Projections from the primary taste cortex then go to the *secondary taste cortex*, namely, the caudolateral OFC. This brain region is involved in the conscious recognition of taste qualities. Moreover, since it contains cells that are activated by several sensory modalities, it is probably a center for establishing "flavor."

Disorders of Olfaction

The ability to smell is influenced by factors such as age, sex, general health, nutrition, smoking, and reproductive state. Women typically outperform men on tests of olfactory function and retain normal smell function to a later age. Significant decrements in the ability to smell are present in over 50% of the population between 65 and 80 years of age and in 75% of those 80 years and older (**Fig. 29-6**). Such *presbyosmia* helps explain why many elderly persons report that food has little flavor, a problem that can result in nutritional disturbances. It also helps explain why a disproportionate number of the elderly die in accidental gas poisonings. A relatively complete listing of conditions and disorders that have been associated with olfactory dysfunction is presented in **Table 29-1**.

Figure 29-6

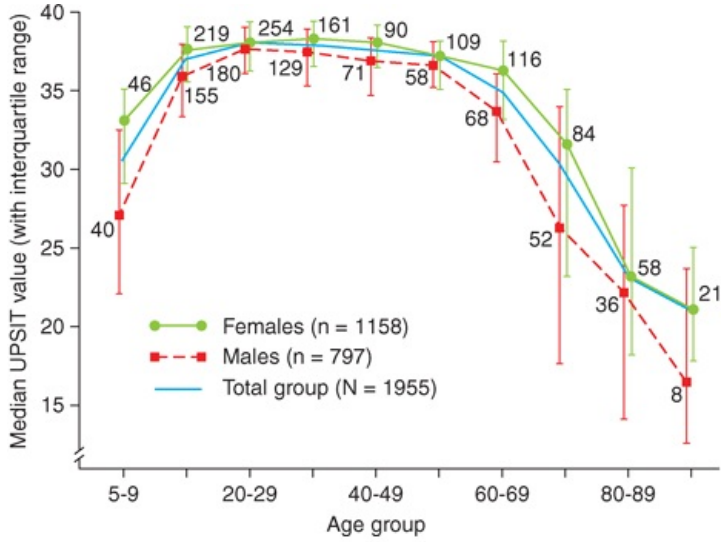

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scores on the University of Pennsylvania Smell Identification Test (UPSIT) as a function of subject age and sex. Numbers by each data point indicate sample sizes. Note that women identify odorants better than men at all ages. *(From Doty et al: Science 226:1421, 1984. Copyright 1984 American Association for the Advancement of Science.)*

Table 29-1 Disorders and Conditions Associated with Compromised Olfactory Function as Measured by Olfactory Testing

22q11 deletion syndrome	Liver disease
AIDS/HIV infection	Lubag disease
Adenoid hypertrophy	Medications
Adrenal cortical insufficiency	Migraine
Age	Multiple sclerosis
Alcoholism	Multi-infarct dementia
Allergies	Narcolepsy with cataplexy
Amyotrophic lateral sclerosis	Nutritional deficiencies
Anorexia nervosa	Obstructive pulmonary disease
Asperger's syndrome	Obesity
Ataxias	Obsessive-compulsive disorder
Attention deficit/hyperactivity disorder	Orthostatic tremor
Bardet-Biedl syndrome	Panic disorder
Chemical exposure	Parkinson's disease
Chronic obstructive pulmonary disease	Pick's disease
Congenital	Posttraumatic stress disorder
Cushing's syndrome	Pregnancy
Cystic fibrosis	Pseudohypoparathyroidism
Degenerative ataxias	Psychopathy
Diabetes	Radiation (therapeutic, cranial)
Down syndrome	REM behavior disorder
Epilepsy	Refsum disease
Facial paralysis	Renal failure/end-stage kidney disease
Frontotemporal lobe degeneration	Restless leg syndrome
Gonadal dysgenesis (Turner syndrome)	Rhinosinusitis/polyposis
Guamanian ALS/PD/dementia syndrome	Schizophrenia
Head trauma	Seasonal affective disorder
Herpes simplex encephalitis	Sjögren's syndrome
Hypothyroidism	Stroke
Huntington's disease	Tobacco smoking

Iatrogenesis	Toxic chemical exposure
Kallmann's syndrome	Upper respiratory infections
Korsakoff's psychosis	Usher syndrome
Leprosy	Vitamin B12 deficiency

Aside from aging, the three most common identifiable causes of long-lasting or permanent smell loss seen in the clinic are, in order of frequency, severe upper respiratory infections, head trauma, and chronic rhinosinusitis. The physiologic basis for most head trauma–related losses is the shearing and subsequent scarring of the olfactory fila as they pass from the nasal cavity into the brain cavity. The cribriform plate does not have to be fractured or show pathology for smell loss to be present. Severity of trauma, as indexed by a poor Glasgow Coma Rating on presentation and the length of posttraumatic amnesia, is associated with higher risk of olfactory impairment. Fewer than 10% of posttraumatic anosmic patients recover age-related normal function over time. Upper respiratory infections, such as those associated with the common cold, influenza, pneumonia, or HIV, can directly and permanently harm the olfactory epithelium by decreasing receptor cell numbers, damaging cilia on remaining receptor cells, and inducing the replacement of sensory epithelium with respiratory epithelium. The smell loss associated with chronic rhinosinusitis is related to disease severity, with most loss occurring in cases in which rhinosinusitis and polyposis are both present. Although systemic glucocorticoid therapy usually can induce short-term functional improvement, it does not, on average, return smell test scores to normal, implying that chronic permanent neural loss is present and/or that short-term administration of systemic glucocorticoids does not mitigate the inflammation completely. It is well established that microinflammation in an otherwise seemingly normal epithelium can influence smell function.

A number of neurodegenerative diseases are accompanied by olfactory impairment, including AD, PD, Huntington's disease, Down syndrome, parkinsonism-dementia complex of Guam, dementia with Lewy bodies (DLB), multiple system atrophy, vascular parkinsonism, corticobasal syndrome, frontotemporal dementia, multiple sclerosis (MS), and idiopathic rapid eye movement (REM) behavioral sleep disorder (iRBD). The olfactory disturbance of MS varies as a function of the plaque activity within the frontal and temporal lobes. In postmortem studies of patients with very mild "presymptomatic" signs of AD, poorer smell function has been associated with higher levels of AD-related pathology even after controlling for apolipoprotein E4 alleles and the level of episodic memory function present at the time of olfactory testing. Olfactory impairment in PD often predates the clinical diagnosis by at least 4 years. Studies of the sequence of Lewy body and abnormal α-synuclein development in staged PD cases, along with evidence that the smell loss presents early, is stable over time, and is not affected by PD medications, suggest that the olfactory bulbs may be, along with the dorsomotor nucleus of the vagus, the site of first neural damage in PD. Smell loss is more marked in patients with early clinical manifestations of DLB than in those with mild AD. Interestingly, smell loss is minimal or nonexistent in progressive supranuclear palsy and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-induced parkinsonism.

The smell loss seen in iRBD is of the same magnitude as that found in PD. This is of particular interest to clinicians since patients with iRBD frequently develop PD and hyposmia. iRBD may actually represent an early associated condition of PD. REM behavior disorder not only is seen in its idiopathic form but also can be associated with narcolepsy. This led to a study of narcoleptic patients with and without REM behavior disorder that demonstrated that narcolepsy, independent of REM behavior disorder, was associated with significant impairments in olfactory function. Orexin A, also known as hypocretin-1, is dramatically diminished or undetectable in the cerebrospinal fluid of patients with narcolepsy and cataplexy. The orexin-containing neurons in the hypothalamus project throughout the olfactory system (from the olfactory epithelium to the olfactory cortex), and damage to these orexin-containing projections may be one underlying mechanism for impaired olfactory performance in narcoleptic patients. The administration of intranasal orexin A (hypocretin-1) appears to result in improved olfactory function relative to a placebo, supporting the notion that mild olfactory impairment is not only a primary feature of narcolepsy with cataplexies but that CNS orexin deficiency may be a fundamental part of the mechanism for this loss.

Disorders of Taste

The majority of patients who present with complaints of taste dysfunction exhibit olfactory, not taste, loss. This is the case because most flavors attributed to taste actually depend on retronasal stimulation of the olfactory receptors during deglutition. As noted earlier, taste buds only mediate basic tastes such as sweet, sour, bitter, salty, and umami. Significant impairment of whole-mouth gustatory function is rare outside of generalized metabolic disturbances or systemic use of some medications, since taste bud regeneration occurs and peripheral damage alone would require the involvement of multiple cranial nerve pathways. Nonetheless, taste can be influenced by (1) the release of foul-tasting materials from the oral cavity from oral medical conditions and appliances (e.g. gingivitis, purulent sialadenitis), (2) transport problems of tastants to the taste buds (e.g., drying of the orolingual mucosa, infections, inflammatory conditions), (3) damage to the taste buds themselves (e.g., local trauma,

invasive carcinomas), (4) damage to the neural pathways innervating the taste buds (e.g., middle ear infections), (5) damage to central structures (e.g., multiple sclerosis, tumor, epilepsy, stroke), and(6) systemic disturbances of metabolism (e.g., diabetes, thyroid disease, medications). Bell's palsy is among the most common causes of CN VII injury that results in taste disturbance. Unlike CN VII, CN IX is relatively protected along its path, although iatrogenic interventions such as tonsillectomy, bronchoscopy, laryngoscopy, and radiation therapy can result in selective injury. Migraine is associated on rare occasions with a gustatory prodrome or aura, and certain tastes may trigger a migraine. Although a number of disorders can affect CN IX, including tumors, trauma, vascular lesions, and infection, it remains unclear if noticeable taste disturbance can result from such factors.

Although both taste and smell can be adversely influenced by pharmacologic agents, drug-related taste alterations are more common. Indeed, over 250 medications have been reported to alter the ability to taste. Major offenders include antineoplastic agents, antirheumatic drugs, antibiotics, and blood pressure medications. Terbinafine, a commonly used antifungal, has been linked to taste disturbance lasting up to 3 years. In a controlled trial, nearly two-thirds of individuals taking eszopiclone (Lunesta) experienced a bitter dysgeusia which was stronger in women, systematically related to the time since drug administration, and positively correlated with both blood and saliva levels of the drug. Intranasal use of nasal gels and sprays containing zinc—common over-the-counter prophylactics for upper respiratory viral infections—has been implicated in loss of smell function. Whether their efficacy in preventing such infections, which are the most common cause of anosmia and hyposmia, outweighs their potential detriment to smell function requires study.

As with olfaction, a number of systemic disorders can affect taste. They include chronic renal failure, end-stage liver disease, vitamin and mineral deficiencies, diabetes, and hypothyroidism, to name a few. Psychiatric conditions can be associated with chemosensory alterations (e.g., depression, schizophrenia, bulimia). A review of tactile, gustatory, and olfactory hallucinations demonstrated that no one type of hallucinatory experience is pathognomonic to any specific diagnosis.

Clinical Evaluation

In most cases, a careful clinical history will establish the probable etiology of a chemosensory problem, including questions about its nature, onset, duration, and pattern of fluctuations. *Sudden loss* suggests the possibility of head trauma, ischemia, infection, or a psychiatric condition. *Gradual loss* can reflect the development of a progressive obstructive lesion. *Intermittent loss* suggests the likelihood of an inflammatory process. The patient should be asked about potential precipitating events, such as cold or flu infections before symptom onset, as they often are underappreciated. Information regarding head trauma, smoking habits, drug and alcohol abuse (e.g., intranasal cocaine, chronic alcoholism in the context of Wernicke's and Korsakoff's syndromes), exposures to pesticides and other toxic agents, and medical interventions are also informative. A determination of all the medications the patient was taking before and at the time of symptom onset is important, since many can cause chemosensory disturbances. Comorbid medical conditions associated with smell impairment, such as renal failure, liver disease, hypothyroidism, diabetes, and dementia, should be assessed. Delayed puberty in association with anosmia (with or without midline craniofacial abnormalities, deafness, and renal anomalies) suggests the possibility of Kallmann syndrome. Recollection of epistaxis, discharge (clear, purulent, or bloody), nasal obstruction, allergies, and somatic symptoms, including headache or irritation, may have localizing value. Questions related to memory, parkinsonian signs, and seizure activity (e.g., automatisms, occurrence of blackouts, auras, and déjàvu) should be posed. Pending litigation and the possibility of malingering should be considered.

Neurologic and otorhinolaryngologic (ORL) examinations, along with appropriate brain and nasosinus imaging, aid in the evaluation of patients with olfactory or gustatory complaints. The neural evaluation should focus on cranial nerve function, with particular attention to possible skull base and intracranial lesions. Visual acuity, field, and optic disc examinations aid in the detection of intracranial mass lesions that induce elevations in intracranial pressure (papilledema) and optic atrophy, especially when one is considering Foster Kennedy syndrome (ipsilateral optic nerve atrophy and contralateral papilledema usually due to a meningioma near the olfactory bulb or tract). The ORL examination should thoroughly assess the intranasal architecture and mucosal surfaces. Polyps, masses, and adhesions of the turbinates to the septum may compromise the flow of air to the olfactory receptors, since less than a fifth of the inspired air traverses the olfactory cleft in the unobstructed state. Blood serum tests may be helpful to identify conditions such as diabetes, infection, heavy metal exposure, nutritional deficiency (e.g., vitamins B_6 and B_{12}), allergy, and thyroid, liver, and kidney disease.

As with other sensory disorders, quantitative sensory testing is advised. Self-reports of patients can be misleading, and a number who complain of chemosensory dysfunction have normal function for their age and sex. Quantitative smell and taste testing provides valid information for worker's compensation and other legal claims as well as a way to assess treatment interventions accurately. A number of standardized olfactory and taste tests are commercially available. Most evaluate the ability of patients to detect and identify odors or tastes. For example, the most widely used of these tests, the 40-item University of Pennsylvania Smell Identification Test (UPSIT), employs norms based on nearly 4000 normal subjects. A determination is made of both absolute dysfunction (i.e., mild loss, moderate loss,

severe loss, total loss, probable malingering) and relative dysfunction (percentile rank for age and sex). Although electrophysiologic testing is available at some smell and taste centers (e.g., odor event-related potentials), such tests require complex stimulus presentation and recording equipment and rarely provide additional diagnostic information. In addition to electrogustometers, commercial chemical taste tests are now available. Most employ filter paper strips impregnated with tastants, so no stimulus preparation is required. Like the UPSIT, these tests have published norms for establishing the degree of dysfunction.

Treatment and Management

Because of the various mechanisms by which olfactory and gustatory disturbance can occur, management of patients tends to be condition-specific. For example, patients with hypothyroidism, diabetes, or infections need to be given specific treatments to correct the underlying process adversely influencing chemoreception. For most patients who present primarily with obstructive/transport loss affecting the nasal and paranasal regions (e.g., allergic rhinitis, polyposis, intranasal neoplasms, nasal deviations), medical and/or surgical intervention is often beneficial. Antifungal and antibiotic treatments may reverse taste problems secondary to candidiasis or other oral infections. Chlorohexidine mouthwash mitigates some salty or bitter dysgeusias, conceivably as a result of its strong positive charge. Excessive dryness of the oral mucosa is a problem with many medications and conditions, and artificial saliva (e.g., Xerolube) or oral pilocarpine treatments may prove beneficial. Other methods to improve salivary flow include the use of mints, lozenges, or sugarless gum. Flavor enhancers may make food more palatable (e.g., monosodium glutamate), but caution is advised to avoid overusing ingredients containing sodium or sugar, particularly in circumstances in which a patient also has underlying hypertension or diabetes. Medications that induce distortions of taste often can be discontinued and replaced with other types of medications or modes of therapy. As mentioned earlier, pharmacologic agents result in taste disturbances much more frequently than smell disturbances, and over 250 medications have been reported to alter the sense of taste. Many drug-related effects are long-lasting and are not reversed by short-term drug discontinuance.

A study of endoscopic sinus surgery in patients with chronic rhinosinusitis and hyposmia revealed that patients with severe olfactory dysfunction before the surgery had a more dramatic and sustained improvement over time compared with patients with more mild olfactory dysfunction before intervention. In the case of intranasal and sinus-related inflammatory conditions such as those seen with allergy, viruses, and traumas, the use of intranasal or systemic glucocorticoids may be helpful. One common approach is a short course of oral prednisone, typically 60 mg daily for 4 days and then tapered by 10 mg daily. The utility of restoring olfaction with either topical or systemic glucocorticoids has been studied. Topical intranasal glucocorticoids were less effective in general than systemic glucocorticoids; however, nasal steroid administration techniques were not analyzed. Intranasal glucocorticoids are more effective if administered in Moffett's position (head in the inverted position such as over the edge of the bed with the bridge of the nose perpendicular to the floor). After head trauma, an initial trial of glucocorticoids may help reduce local edema and the potential deleterious deposition of scar tissue around olfactory fila at the level of the cribriform plate.

Treatments are limited for patients with chemosensory loss or primary injury to neural pathways. Nonetheless, spontaneous recovery can occur. In a follow-up study of 542 patients presenting with smell loss from a variety of causes, modest improvement occurred over an average period of 4 years in about half the participants. However, only 11% of the anosmic and 23% of the hyposmic patients regained normal age-related function. Interestingly, the amount of dysfunction present at the time of presentation, not etiology, was the best predictor of prognosis. Other predictors were the patient's age and the time between the onset of dysfunction and initial testing.

A nonblinded study reported that patients with hyposmia may benefit from smelling strong odors (e.g., eucalyptol, citronella, eugenol, and phenyl ethyl alcohol) before going to bed and immediately upon awaking each day over the course of several months. The rationale for this approach comes from animal studies demonstrating that prolonged exposure to odorants can induce increased neural activity within the olfactory bulb. α-Lipoic acid (200 mg two or three times daily), an essential cofactor for many enzyme complexes with possible antioxidant effects, has been reported to be beneficial in mitigating smell loss after viral infection of the upper respiratory tract, although double-blind studies are needed to confirm this observation. This agent has also been suggested to be useful in some cases of hypogeusia and burning mouth syndrome.

The use of zinc and vitamin A in treating olfactory disturbances is controversial; not much benefit is obtained beyond replenishing established deficiencies. However, zinc improves taste function secondary to hepatic deficiencies, and retinoids (bioactive vitamin A derivatives) are known to play an essential role in the survival of olfactory neurons. One protocol in which zinc was infused with chemotherapy treatments suggested a possible protective effect against developing taste impairment. Diseases of the alimentary tract can not only influence chemoreceptive function but occasionally influence B_{12} absorption. This can result in a relative deficiency of B_{12}, theoretically contributing to olfactory nerve disturbance. B_2 (riboflavin) and magnesium supplements are reported in the alternative medicine literature to aid in the management of migraine headaches that may be associated with smell

dysfunction.

A number of medicines have been reported to ameliorate olfactory symptoms, although strong scientific evidence for efficacy is generally lacking. A report that theophylline improved smell function was not double-blinded and lacked a control group, failing to take into account that some meaningful improvement occurs without treatment. Indeed, the percentage of patients reported to be responsive to the treatment was about the same as that noted by others to show spontaneous improvement over a similar time period (~50%). Antiepileptics and some antidepressants (e.g. amitriptyline) have been used to treat dysosmias and smell distortions, particularly after head trauma. Ironically, amitriptyline is also frequently on the list of medications that can ultimately distort smell and taste function, possibly from its anticholinergic effects. The use of donepezil (an acetylcholinesterase inhibitor) in AD may result in improvements in smell identification measures that correlate with overall clinician-based impressions of change scales [Clinician Interview Based Impression of Severity (CIBIC)-plus]. Smell identification function could become a useful measure to assess overall treatment response with this medication.

A major and often overlooked element of therapy comes from chemosensory testing itself. Confirmation or lack of confirmation of loss is beneficial to patients who come to believe, in light of unsupportive family members and medical providers, that they may be "crazy." In cases in which the loss is minor, patients can be informed of the likelihood of a more positive prognosis. Importantly, quantitative testing places the patient's problem into overall perspective. Thus, it is often therapeutic for an older person to know that although his or her smell function is not what it used to be, it still falls above the average of his or her peer group. Without testing, many such patients are simply told they are getting old and nothing can be done for them, leading in some cases to depression and decreased self-esteem.

FURTHER READINGS

Bromley SM, Doty RL: Olfaction in dentistry. Oral Dis 16:221, 2010 [PMID: 19732354]

Calderòn-Garcidueñas L et al: Urban air pollution: Influences on olfactory function and pathology in exposed children and young adults. Exp Toxicol Pathol 62:91, 2010 [PMID: 21602559]

Deems DA et al: Smell and taste disorders: A study of 750 patients from the University of Pennsylvania Smell and Taste Center. Arch Otolaryngol Head Neck Surg 117:519, 1991 [PMID: 2021470]

Doty RL: The olfactory vector hypothesis of neurodegenerative disease: Is it viable? Ann Neurol 63:7, 2008 [PMID: 18232016]

——— (ed): *Handbook of Olfaction and Gustation.* New York, Marcel Dekker, 2003

——— et al: Drug-induced taste disorders: Incidence, prevention and management. Drug Safety 31:199, 2008

Gottfried JA: Function follows form: Ecological constraints on odor codes and olfactory percepts. Curr Opin Neurobiol 19:422, 2009 [PMID: 19671493]

Hawkes CH, Doty RL. *Neurology of Olfaction* . Cambridge, Cambridge University Press, 2009

Kern RC: Chronic sinusitis and anosmia: Pathologic changes in the olfactory mucosa. Laryngoscope 110:1071, 2000 [PMID: 10892672]

London B et al: Predictors of prognosis in patients with olfactory disturbance. Ann Neurol 63:159, 2008 [PMID: 18058814]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF HEARING: INTRODUCTION

Hearing loss is one of the most common sensory disorders in humans and can present at any age. Nearly 10% of the adult population has some hearing loss, and one-third of individuals age >65 years have a hearing loss of sufficient magnitude to require a hearing aid.

Physiology of Hearing

The function of the external and middle ear is to amplify sound to facilitate conversion of the mechanical energy of the sound wave into an electrical signal by the inner ear hair cells, a process called mechanotransduction **(Fig. 30-1)**. Sound waves enter the external auditory canal and set the tympanic membrane in motion, which in turn moves the malleus, incus, and stapes of the middle ear. Movement of the footplate of the stapes causes pressure changes in the fluid-filled inner ear, eliciting a traveling wave in the basilar membrane of the cochlea. The tympanic membrane and the ossicular chain in the middle ear serve as an impedance-matching mechanism, improving the efficiency of energy transfer from air to the fluid-filled inner ear.

Figure 30-1

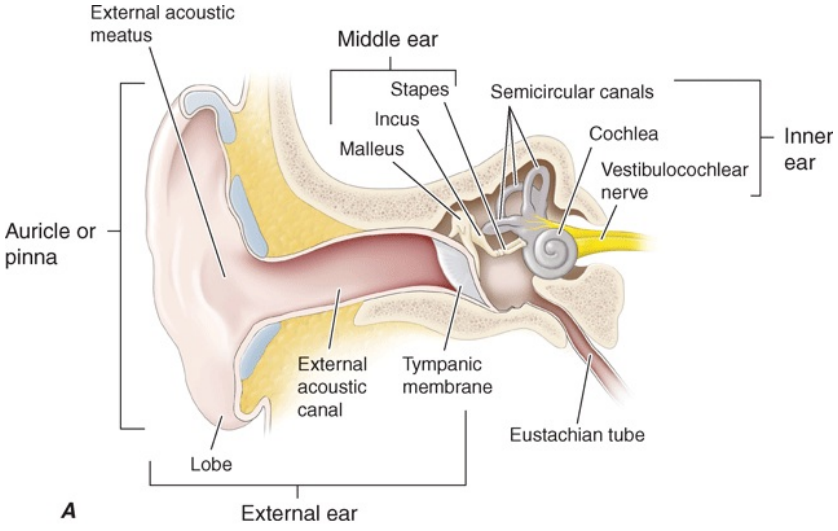

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

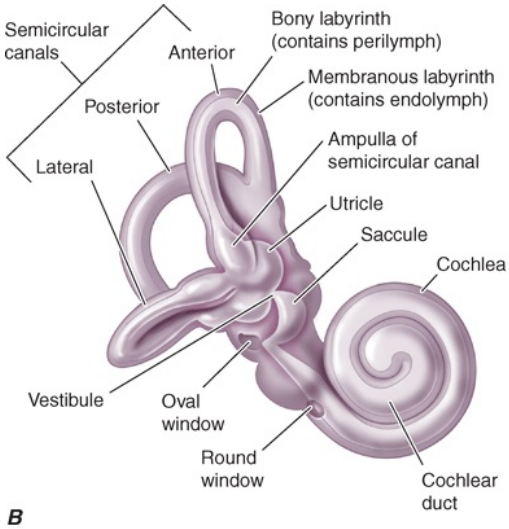

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ear anatomy. ***A.*** Drawing of modified coronal section through external ear and temporal bone, with structures of the middle and inner ear demonstrated. ***B.*** High-resolution view of inner ear.

Stereocilia of the hair cells of the organ of Corti, which rests on the basilar membrane, are in contact with the tectorial membrane and are deformed by the traveling wave. A point of maximal displacement of the basilar membrane is determined by the frequency of the stimulating tone. High-frequency tones cause maximal displacement of the basilar

membrane near the base of the cochlea, whereas for low-frequency sounds, the point of maximal displacement is toward the apex of the cochlea.

The inner and outer hair cells of the organ of Corti have different innervation patterns, but both are mechanoreceptors. The afferent innervation relates principally to the inner hair cells, and the efferent innervation relates principally to outer hair cells. The motility of the outer hair cells alters the micromechanics of the inner hair cells, creating a cochlear amplifier, which explains the exquisite sensitivity and frequency selectivity of the cochlea.

Beginning in the cochlea, the frequency specificity is maintained at each point of the central auditory pathway: dorsal and ventral cochlear nuclei, trapezoid body, superior olivary complex, lateral lemniscus, inferior colliculus, medial geniculate body, and auditory cortex. At low frequencies, individual auditory nerve fibers can respond more or less synchronously with the stimulating tone. At higher frequencies, phase-locking occurs so that neurons alternate in response to particular phases of the cycle of the sound wave. Intensity is encoded by the amount of neural activity in individual neurons, the number of neurons that are active, and the specific neurons that are activated.

Disorders of the Sense of Hearing

Hearing loss can result from disorders of the auricle, external auditory canal, middle ear, inner ear, or central auditory pathways **(Fig. 30-2)**. *In general, lesions in the auricle, external auditory canal, or middle ear that impede the transmission of sound from the external environment to the inner ear cause conductive hearing loss, whereas lesions that impair mechanotransduction in the inner ear or transmission of the electrical signal along the eighth nerve to the brain cause sensorineural hearing loss.*

Figure 30-2

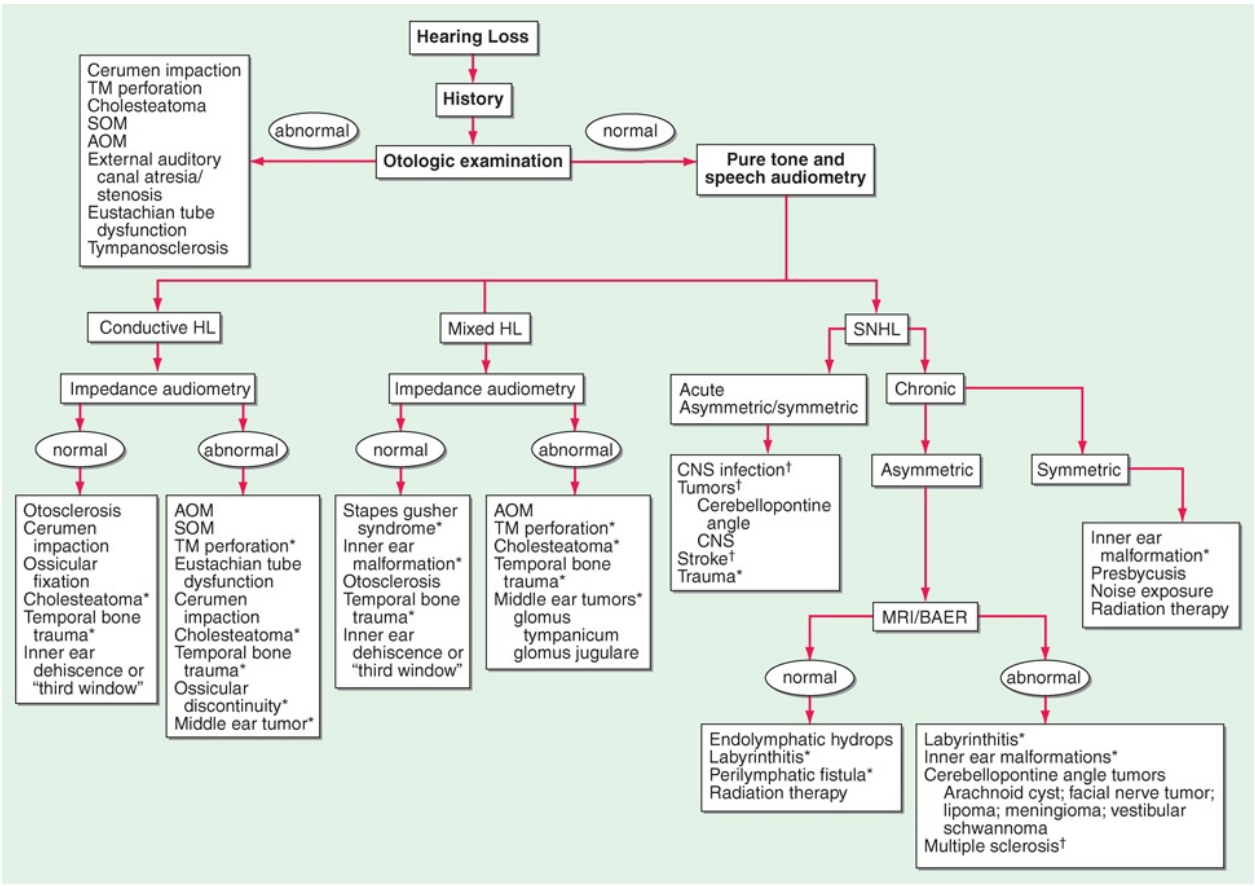

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An algorithm for the approach to hearing loss. HL, hearing loss; SNHL, sensorineural hearing loss; TM, tympanic membrane; SOM, serous otitis media; AOM, acute otitis media; BAER, brainstem auditory evoked response; *, CT scan of temporal bone; †, MRI scan.

CONDUCTIVE HEARING LOSS

The external ear, the external auditory canal, and the middle ear apparatus is designed to collect and amplify sound and efficiently transfer the mechanical energy of the sound wave to the fluid-filled cochlea. Factors that obstruct the transmission of sound or serve to dampen the acoustical energy result in conductive hearing loss. Conductive hearing loss can occur from obstruction of the external auditory canal by cerumen, debris, and foreign bodies; swelling of the lining of the canal; atresia or neoplasms of the canal; perforations of the tympanic membrane; disruption of the ossicular chain, as occurs with necrosis of the long process of the incus in trauma or infection; otosclerosis; or fluid, scarring, or neoplasms in the middle ear. Rarely, inner ear malformations or pathologies may also be associated with conductive hearing loss.

Eustachian tube dysfunction is extremely common in adults and may predispose to acute otitis media (AOM) or serous otitis media (SOM). Trauma, AOM, or chronic otitis media are the usual factors responsible for tympanic membrane perforation. While small perforations often heal spontaneously, larger defects usually require surgical intervention. Tympanoplasty is highly effective (>90%) in the repair of tympanic membrane perforations. Otoscopy is usually sufficient to diagnose AOM, SOM, chronic otitis media, cerumen impaction, tympanic membrane perforation, and eustachian tube dysfunction;

tympanometry can be useful to confirm the clinical suspicion of these conditions.

Cholesteatoma, a benign tumor composed of stratified squamous epithelium in the middle ear or mastoid, occurs frequently in adults. This is a slowly growing lesion that destroys bone and normal ear tissue. Theories of pathogenesis include traumatic immigration and invasion of squamous epithelium through a retraction pocket, implantation of squamous epithelia in the middle ear through a perforation or surgery, and metaplasia following chronic infection and irritation. On examination, there is often a perforation of the tympanic membrane filled with cheesy white squamous debris. A chronically draining ear that fails to respond to appropriate antibiotic therapy should raise suspicion of a cholesteatoma. Conductive hearing loss secondary to ossicular erosion is common. Surgery is required to remove this destructive process.

Conductive hearing loss with a normal ear canal and intact tympanic membrane suggests either ossicular pathology or the presence of "third window" in the inner ear (see below). Fixation of the stapes from *otosclerosis* is a common cause of low-frequency conductive hearing loss. It occurs equally in men and women and is inherited as an autosomal dominant trait with incomplete penetrance; in some cases, it may be a manifestation of osteogenesis imperfecta. Hearing impairment usually presents between the late teens and the forties. In women, the otosclerotic process is accelerated during pregnancy, and the hearing loss is often first noticeable at this time. A hearing aid or a simple outpatient surgical procedure (stapedectomy) can provide adequate auditory rehabilitation. Extension of otosclerosis beyond the stapes footplate to involve the cochlea (cochlear otosclerosis) can lead to mixed or sensorineural hearing loss. Fluoride therapy to prevent hearing loss from cochlear otosclerosis is of uncertain value.

Disorders that lead to the formation of a pathologic "third window" in the inner ear can be associated with conductive hearing loss. There are normally two major openings, or windows, that connect the inner ear with the middle ear and serve as conduits for transmission of sound; these are, respectively, the oval and round windows. A third window is formed where the normally hard otic bone surrounding the inner ear is eroded; dissipation of the acoustic energy at the third window is responsible for the "inner ear conductive hearing loss." The superior semicircular canal dehiscence syndrome resulting from erosion of the otic bone over the superior circular canal can present with conductive hearing loss that mimics otosclerosis. A common symptom is vertigo evoked by loud sounds (Tullio phenomenon), by Valsalva maneuvers that change middle ear pressure, or by applying positive pressure on the tragus (the cartilage anterior to the external opening of the ear canal),. Patients with this syndrome also complain of being able to hear the movement of their eyes and neck. A large jugular bulb or jugular bulb diverticulum can create a "third window" by eroding into the vestibular aqueduct or posterior semicircular canal; the symptoms are similar to those of the superior semicircular canal dehiscence syndrome.

SENSORINEURAL HEARING LOSS

Sensorineural hearing loss results from either damage to the mechanotransduction apparatus of the cochlea or disruption of the electrical conduction pathway from the inner ear to the brain. Thus, injury to hair cells, supporting cells, auditory neurons, or the central auditory pathway can cause sensorineural hearing loss. Damage to the hair cells of the organ of Corti may be caused by intense noise, viral infections, ototoxic drugs (e.g., salicylates, quinine and its synthetic analogues, aminoglycoside antibiotics, loop diuretics such as furosemide and ethacrynic acid, and cancer chemotherapeutic agents such as cisplatin), fractures of the temporal bone, meningitis, cochlear otosclerosis (see above), Ménière's disease, and aging. Congenital malformations of the inner ear may be the cause of hearing loss in some adults. Genetic predisposition alone or in concert with environmental exposures may also be responsible (see below).

Presbycusis (age-associated hearing loss) is the most common cause of sensorineural hearing loss in adults. In the early stages, it is characterized by symmetric, gentle to sharply sloping high-frequency hearing loss. With progression, the hearing loss involves all frequencies. More importantly, the hearing impairment is associated with significant loss in clarity. There is a loss of discrimination for phonemes, recruitment (abnormal growth of loudness), and particular difficulty in understanding speech in noisy environments such as at restaurants and social events. Hearing aids are helpful in enhancing the signal-to-noise ratio by amplifying sounds that are close to the listener. *Although hearing aids are able to amplify sounds, they cannot restore the clarity of hearing.* Thus, amplification with hearing aids may provide only limited rehabilitation once the word recognition score deteriorates below 50%. Cochlear implants are the treatment of choice when hearing aids prove inadequate, even when hearing loss is incomplete (see below).

Ménière's disease is characterized by episodic vertigo, fluctuating sensorineural hearing loss, tinnitus, and aural fullness. Tinnitus and/or deafness may be absent during the initial attacks of vertigo, but it invariably appears as the disease progresses and increases in severity during acute attacks. The annual incidence of Ménière's disease is 0.5–7.5 per 1000; onset is most frequently in the fifth decade of life but may also occur in young adults or the elderly. Histologically, there is distention of the endolymphatic system (endolymphatic hydrops) leading to degeneration of vestibular and cochlear hair cells. This may result from endolymphatic sac dysfunction secondary to infection, trauma, autoimmune disease, inflammatory causes, or tumor; an idiopathic etiology constitutes the largest category and is most accurately referred to as Ménière's disease. Although any pattern of hearing loss can be observed, typically, low-frequency, unilateral sensorineural hearing impairment is present. MRI should be obtained to exclude retrocochlear pathology such as a cerebellopontine angle tumor or demyelinating disorder. Therapy is directed toward the control of vertigo. A 2-g/d low-salt diet is the mainstay of treatment for control of rotatory vertigo. Diuretics, a short course of glucocorticoids, and intratympanic gentamicin may also be useful adjuncts in recalcitrant cases. Surgical therapy of vertigo is reserved for unresponsive cases and includes endolymphatic sac decompression, labyrinthectomy, and vestibular nerve section. Both labyrinthectomy and vestibular nerve section abolish rotatory vertigo in >90% of cases. Unfortunately, there is no effective therapy for hearing loss, tinnitus, or aural fullness from Ménière's disease.

Sensorineural hearing loss may also result from any neoplastic, vascular, demyelinating, infectious, or degenerative disease or trauma affecting the central auditory pathways. HIV leads to both peripheral and central auditory system pathology and is associated with sensorineural hearing impairment.

Primary diseases of the central nervous system can also present with hearing impairment. Characteristically, a reduction in clarity of hearing and speech comprehension is much greater than the loss of the ability to hear pure tone. Auditory testing is consistent with an auditory neuropathy; normal otoacoustic emissions (OAE) and an abnormal auditory brainstem response (ABR) is typical (see below). Hearing loss can accompany hereditary sensorimotor neuropathies and inherited disorders of myelin. Tumors of the cerebellopontine angle such as vestibular schwannoma and meningioma usually present with asymmetric sensorineural hearing loss with greater deterioration of speech understanding than pure tone hearing. Multiple sclerosis may present with acute unilateral or bilateral hearing loss; typically, pure tone testing remains relatively stable while speech understanding fluctuates. Isolated labyrinthine infarction can present with acute hearing loss and vertigo due to a cerebrovascular accident involving the posterior circulation, usually the anterior inferior cerebellar artery; it may also be the heralding sign of impending catastrophic basilar artery infarction (Chap. 370).

A finding of conductive and sensory hearing loss in combination is termed *mixed hearing loss*. Mixed hearing losses are

due to pathology of both the middle and inner ear, as can occur in otosclerosis involving the ossicles and the cochlea, head trauma, chronic otitis media, cholesteatoma, middle ear tumors, and some inner ear malformations.

Trauma resulting in temporal bone fractures may be associated with conductive, sensorineural, or mixed hearing loss. If the fracture spares the inner ear, there may simply be conductive hearing loss due to rupture of the tympanic membrane or disruption of the ossicular chain. These abnormalities can be surgically corrected. Profound hearing loss and severe vertigo are associated with temporal bone fractures involving the inner ear. A perilymphatic fistula associated with leakage of inner ear fluid into the middle ear can occur and may require surgical repair. An associated facial nerve injury is not uncommon. CT is best suited to assess fracture of the traumatized temporal bone, evaluate the ear canal, and determine the integrity of the ossicular chain and the involvement of the inner ear. CSF leaks that accompany temporal bone fractures are usually self-limited; the value of prophylactic antibiotics is uncertain.

Tinnitus is defined as the perception of a sound when there is no sound in the environment. It may have a buzzing, roaring, or ringing quality and may be pulsatile (synchronous with the heartbeat). Tinnitus is often associated with either a conductive or sensorineural hearing loss. The pathophysiology of tinnitus is not well understood. The cause of the tinnitus can usually be determined by finding the cause of the associated hearing loss. Tinnitus may be the first symptom of a serious condition such as a vestibular schwannoma. Pulsatile tinnitus requires evaluation of the vascular system of the head to exclude vascular tumors such as glomus jugulare tumors, aneurysms, dural arteriovenous fistulas, and stenotic arterial lesions; it may also occur with SOM. It is most commonly associated with some abnormality of the jugular bulb such as a large jugular bulb or jugular bulb diverticulum.

Genetic Causes of Hearing Loss

More than half of childhood hearing impairment is thought to be hereditary; hereditary hearing impairment (HHI) can also manifest later in life. HHI may be classified as either nonsyndromic, when hearing loss is the only clinical abnormality, or syndromic, when hearing loss is associated with anomalies in other organ systems. Nearly two-thirds of HHIs are nonsyndromic, and the remaining one-third are syndromic. Between 70 and 80% of nonsyndromic HHI is inherited in an autosomal recessive manner and designated DFNB; another 15–20% is autosomal dominant (DFNA). Less than 5% is X-linked or maternally inherited via the mitochondria.

Nearly 100 loci harboring genes for nonsyndromic HHI have been mapped, with equal numbers of dominant and recessive modes of inheritance; numerous genes have now been cloned **(Table 30-1)**. The hearing genes fall into the categories of structural proteins (MYH9, MYO7A, MYO15, TECTA, DIAPH1), transcription factors (POU3F4, POU4F3), ion channels (KCNQ4, SLC26A4), and gap junction proteins (GJB2, GJB3, GJB6). Several of these genes, including GJB2, TECTA, and TMC1, cause both autosomal dominant and recessive forms of nonsyndromic HHI. In general, the hearing loss associated with dominant genes has its onset in adolescence or adulthood and varies in severity, whereas the hearing loss associated with recessive inheritance is congenital and profound. Connexin 26, product of the GJB2 gene, is particularly important because it is responsible for nearly 20% of all cases of childhood deafness; half of genetic deafness in children is GJB2-related. Two frameshift mutations, 35delG and 167delT, account for >50% of the cases; however, screening for these two mutations alone is insufficient and sequencing of the entire gene is required to diagnose GJB2-related recessive deafness. The 167delT mutation is highly prevalent in Ashkenazi Jews; ~1 in 1765 individuals in this population are homozygous and affected. The hearing loss can also vary among the members of the same family, suggesting that other genes or factors influence the auditory phenotype.

Table 30-1 Hereditary Hearing Impairment Genes

Designation	Gene	Function
Autosomal Dominant		
	CRYM	Thyroid hormoneñbinding protein
DFNA1	*DIAPH1*	Cytoskeletal protein
DFNA2A	*KCNQ4*	Potassium channel
DFNA2B	*GJB3 (Cx31)*	Gap junction
DFNA3A	*GJB2 (Cx26)*	Gap junction
DFNA3B	*GJB6 (Cx30)*	Gap junction
DFNA4	*MYH14*	Class II nonmuscle myosin
DFNA5	*DFNA5*	Unknown
DFNA6/14/38	*WFS1*	Transmembrane protein
DFNA8/12	*TECTA*	Tectorial membrane protein
DFNA9	*COCH*	Unknown
DFNA10	*EYA4*	Developmental gene
DFNA11	*MYO7A*	Cytoskeletal protein
DFNA13	*COL11A2*	Cytoskeletal protein
DFNA15	*POU4F3*	Transcription factor
DFNA17	*MYH9*	Cytoskeletal protein
DFNA20/26	*ACTG1*	Cytoskeletal protein
DFNA22	*MYO6*	Unconventional myosin
DFNA28	*TFCP2L3*	Transcription factor
DFNA36	*TMC1*	Transmembrane protein
DFNA44	*CCDC50*	Effector of EGF-mediated signaling
DFNA48	*MYO1A*	Unconventional myosin
DFNA50	*MIRN96*	MicroRNA
DFNA51	*TJP2*	Tight junction protein
Autosomal Recessive		
DFNB1A	*GJB2 (CX26)*	Gap junction

DFNB1B	*GJB6 (CX30)*	Gap junction
DFNB2	*MYO7A*	Cytoskeletal protein
DFNB3	*MYO15*	Cytoskeletal protein
DFNB4	*PDS(SLC26A4)*	Chloride/iodide transporter
DFNB6	*TMIE*	Transmembrane protein
DFNB7/B11	*TMC1*	Transmembrane protein
DFNB9	*OTOF*	Trafficking of membrane vesicles
DFNB8/10	*TMPRSS3*	Transmembrane serine protease
DFNB12	*CDH23*	Intercellular adherence protein
DFNB16	*STRC*	Stereocilia protein
DFNB18	*USH1C*	Unknown
DFNB21	*TECTA*	Tectorial membrane protein
DFNB22	*OTOA*	Gel attachment to nonsensory
DFNB23	*PCDH15*	Morphogenesis and cohesion
DFNB24	*RDX*	Cytoskeletal protein
DFNB25	*GRXCR1*	Reversible S-glutathionylation proteins
DFNB28	*TRIOBP*	Cytoskeletal-organizing protein
DFNB29	*CLDN14*	Tight junctions
DFNB30	*MYO3A*	Hybrid motor-signaling myosin
DFNB31	*WHRN*	PDZ domainñcontaining protein
DFNB35	*ESRRB*	Estrogen-related receptor beta protein
DFNB36	*ESPN*	Ca-insensitive actin-bundling protein
DFNB37	*MYO6*	Unconventional myosin
DFNB39	*HFG*	Hepatocyte growth factor
DFNB49	*MARVELD2*	Tight junction protein
DFNB53	*COL11A2*	Collagen protein
DFNB59	*PJVK*	Zn-binding protein
DFNB61	*SLC26A5*	Motor protein
DFNB63	*LRTOMT/COMT2*	Putative methyltransferase
DFNB66/67	*LHFPL5*	Tetraspan protein
DFNB77	*LOXHD1*	Stereociliary protein
DFNB79	*TPRN*	Unknown
DFNB82	*GPSM2*	G protein signaling modulator
DFNB84	*PTPRQ*	Type III receptor-like proteintyrosine phosphatase family

In addition to GJB2, several other nonsyndromic genes are associated with hearing loss that progresses with age. The contribution of genetics to presbycusis is also becoming better understood. Sensitivity to aminoglycoside ototoxicity can be maternally transmitted through a mitochondrial mutation. Susceptibility to noise-induced hearing loss may also be genetically determined.

There are >400 syndromic forms of hearing loss. These include Usher syndrome (retinitis pigmentosa and hearing loss), Waardenburg syndrome (pigmentary abnormality and hearing loss), Pendred syndrome (thyroid organification defect and hearing loss), Alport syndrome (renal disease and hearing loss), Jervell and Lange-Nielsen syndrome (prolonged QT interval and hearing loss), neurofibromatosis type 2 (bilateral acoustic schwannoma), and mitochondrial disorders [mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS); myoclonic epilepsy and ragged red fibers (MERRF); progressive external ophthalmoplegia (PEO)] **(Table 30-2)**.

Table 30-2 Syndromic Hereditary Hearing Impairment Genes

Syndrome	Gene	Function
Alport syndrome	*COL4A3-5*	Cytoskeletal protein
BOR syndrome	*EYA1*	Developmental gene
	SIX5	Developmental gene
	SIX1	Developmental gene
Jervell and Lange-Nielsen syndrome	*KCNQ1*	Delayed rectifier K+ channel
	KCNE1	Delayed rectifier K+ channel
Norrie disease	*NDP*	Cell-cell interactions
Pendred syndrome	*SLC26A4*	Chloride/iodide transporter
	FOXI1	Transcriptional activator of *SLC26A4*
Treacher Collins	*TCOF1*	Nucleolar-cytoplasmic transport
Usher syndrome	*MYO7A*	Cytoskeletal protein
	USH1C	Unknown
	CDH23	Intercellular adherence protein
	PCDH15	Cell adhesion molecule

	SANS	Harmonin-associated protein
	USH2A	Cell adhesion molecule
	VLGR1	G protein–coupled receptor
	USH3	Unknown
	WHRN	PDZ domainñcontaining protein
WS type I, III	*PAX3*	Transcription factor
WS type II	*MITF*	Transcription factor
	SNAI2	Transcription factor
WS type IV	*EDNRB*	Endothelin B receptor
	EDN3	Endothelin B receptor ligand
	SOX10	Transcription factor

Abbreviations: BOR, branchio-oto-renal syndrome; WS, Waardenburg syndrome.

Approach to the Patient: Disorders of the Sense of Hearing

The goal in the evaluation of a patient with auditory complaints is to determine (1) the nature of the hearing impairment (conductive vs. sensorineural vs. mixed), (2) the severity of the impairment (mild, moderate, severe, profound), (3) the anatomy of the impairment (external ear, middle ear, inner ear, or central auditory pathway), and (4) the etiology. The history should elicit characteristics of the hearing loss, including the duration of deafness, unilateral vs. bilateral involvement, nature of onset (sudden vs. insidious), and rate of progression (rapid vs. slow). Symptoms of tinnitus, vertigo, imbalance, aural fullness, otorrhea, headache, facial nerve dysfunction, and head and neck paresthesias should be noted. Information regarding head trauma, exposure to ototoxins, occupational or recreational noise exposure, and family history of hearing impairment may also be important. A sudden onset of unilateral hearing loss, with or without tinnitus, may represent a viral infection of the inner ear or a stroke. Patients with unilateral hearing loss (sensory or conductive) usually complain of reduced hearing, poor sound localization, and difficulty hearing clearly with background noise. Gradual progression of a hearing deficit is common with otosclerosis, noise-induced hearing loss, vestibular schwannoma, or Ménière's disease. Small vestibular schwannomas typically present with asymmetric hearing impairment, tinnitus, and imbalance (rarely vertigo); cranial neuropathy, in particular of the trigeminal or facial nerve, may accompany larger tumors. In addition to hearing loss, Ménière's disease may be associated with episodic vertigo, tinnitus, and aural fullness. Hearing loss with otorrhea is most likely due to chronic otitis media or cholesteatoma.

Examination should include the auricle, external ear canal, and tympanic membrane. The external ear canal of the elderly is often dry and fragile; it is preferable to clean cerumen with wall-mounted suction or cerumen loops and to avoid irrigation. In examining the eardrum, the topography of the tympanic membrane is more important than the presence or absence of the light reflex. In addition to the pars tensa (the lower two-thirds of the eardrum), the pars flaccida above the short process of the malleus should also be examined for retraction pockets that may be evidence of chronic eustachian tube dysfunction or cholesteatoma. Insufflation of the ear canal is necessary to assess tympanic membrane mobility and compliance. Careful inspection of the nose, nasopharynx, and upper respiratory tract is indicated. Unilateral serous effusion should prompt a fiberoptic examination of the nasopharynx to exclude neoplasms. Cranial nerves should be evaluated with special attention to facial and trigeminal nerves, which are commonly affected with tumors involving the cerebellopontine angle.

The Rinne and Weber tuning fork tests, with a 512-Hz tuning fork, are used to screen for hearing loss, differentiate conductive from sensorineural hearing losses, and to confirm the findings of audiologic evaluation. Rinne's test compares the ability to hear by air conduction with the ability to hear by bone conduction. The tines of a vibrating tuning fork are held near the opening of the external auditory canal, and then the stem is placed on the mastoid process; for direct contact, it may be placed on teeth or dentures. The patient is asked to indicate whether the tone is louder by air conduction or bone conduction. Normally, and in the presence of sensorineural hearing loss, a tone is heard louder by air conduction than by bone conduction; however, with conductive hearing loss of ≳30 dB (see "Audiologic Assessment," below), the bone-conduction stimulus is perceived as louder than the air-conduction stimulus. For the Weber test, the stem of a vibrating tuning fork is placed on the head in the midline and the patient asked whether the tone is heard in both ears or better in one ear than in the other. With a unilateral conductive hearing loss, the tone is perceived in the affected ear. With a unilateral sensorineural hearing loss, the tone is perceived in the unaffected ear. A 5-dB difference in hearing between the two ears is required for lateralization.

Laboratory Assessment of Hearing

AUDIOLOGIC ASSESSMENT

The minimum audiologic assessment for hearing loss should include the measurement of pure tone air-conduction and bone-conduction thresholds, speech reception threshold, word recognition score, tympanometry, acoustic reflexes, and acoustic-reflex decay. This test battery provides a screening evaluation of the entire auditory system and allows one to determine whether further differentiation of a sensory (cochlear) from a neural (retrocochlear) hearing loss is indicated.

Pure tone audiometry assesses hearing acuity for pure tones. The test is administered by an audiologist and is performed in a sound-attenuated chamber. The pure tone stimulus is delivered with an audiometer, an electronic device that allows the presentation of specific frequencies (generally between 250 and 8000 Hz) at specific intensities. Air- and bone-conduction thresholds are established for each ear. Air-conduction thresholds are determined by presenting the stimulus in air with the use of headphones. Bone-conduction thresholds are determined by placing the stem of a vibrating tuning fork or an oscillator of an audiometer in contact with the head. In the presence of a hearing loss, broad-spectrum noise is presented to the nontest ear for *masking* purposes so that responses are based on perception from the ear under test.

The responses are measured in decibels. An *audiogram* is a plot of intensity in decibels of hearing threshold versus frequency. A decibel (dB) is equal to 20 times the logarithm of the ratio of the sound pressure required to achieve threshold in the patient to the sound pressure required to achieve threshold in a normal hearing person. Therefore, a change of 6 dB represents doubling of sound pressure, and a change of 20 dB represents a tenfold change in sound pressure. Loudness, which depends on the frequency, intensity, and duration of a sound, doubles with approximately each 10-dB increase in sound pressure level. Pitch, on the other hand, does not directly correlate with frequency. The perception of pitch changes slowly in the low and high frequencies. In the middle tones, which are important for human speech, pitch varies more rapidly with changes in frequency.

Pure tone audiometry establishes the presence and severity of hearing impairment, unilateral vs. bilateral involvement, and

the type of hearing loss. Conductive hearing losses with a large mass component, as is often seen in middle ear effusions, produce elevation of thresholds that predominate in the higher frequencies. Conductive hearing losses with a large stiffness component, as in fixation of the footplate of the stapes in early otosclerosis, produce threshold elevations in the lower frequencies. Often, the conductive hearing loss involves all frequencies, suggesting involvement of both stiffness and mass. In general, sensorineural hearing losses such as presbycusis affect higher frequencies more than lower frequencies. An exception is Ménière's disease, which is characteristically associated with low-frequency sensorineural hearing loss. Noise-induced hearing loss has an unusual pattern of hearing impairment in which the loss at 4000 Hz is greater than at higher frequencies. Vestibular schwannomas characteristically affect the higher frequencies, but any pattern of hearing loss can be observed.

Speech recognition requires greater synchronous neural firing than is necessary for appreciation of pure tones. *Speech audiometry* tests the clarity with which one hears. The *speech reception threshold* (SRT) is defined as the intensity at which speech is recognized as a meaningful symbol and is obtained by presenting two-syllable words with an equal accent on each syllable. The intensity at which the patient can repeat 50% of the words correctly is the SRT. Once the SRT is determined, discrimination or word recognition ability is tested by presenting one-syllable words at 25–40 dB above the SRT. The words are phonetically balanced in that the phonemes (speech sounds) occur in the list of words at the same frequency that they occur in ordinary conversational English. An individual with normal hearing or conductive hearing loss can repeat 88–100% of the phonetically balanced words correctly. Patients with a sensorineural hearing loss have variable loss of discrimination. As a general rule, neural lesions produce greater deficits in discrimination than do cochlear lesions. For example, in a patient with mild asymmetric sensorineural hearing loss, a clue to the diagnosis of vestibular schwannoma is the presence of greater than expected deterioration in discrimination ability. Deterioration in discrimination ability at higher intensities above the SRT also suggests a lesion in the eighth nerve or central auditory pathways.

Tympanometry measures the impedance of the middle ear to sound and is useful in diagnosis of middle-ear effusions. A *tympanogram* is the graphic representation of change in impedance or compliance as the pressure in the ear canal is changed. Normally, the middle ear is most compliant at atmospheric pressure, and the compliance decreases as the pressure is increased or decreased (type A); this pattern is seen with normal hearing or in the presence of sensorineural hearing loss. Compliance that does not change with change in pressure suggests middle-ear effusion (type B). With a negative pressure in the middle ear, as with eustachian tube obstruction, the point of maximal compliance occurs with negative pressure in the ear canal (type C). A tympanogram in which no point of maximal compliance can be obtained is most commonly seen with discontinuity of the ossicular chain (type A_d). A reduction in the maximal compliance peak can be seen in otosclerosis (type A_S).

During tympanometry, an intense tone elicits contraction of the stapedius muscle. The change in compliance of the middle ear with contraction of the stapedius muscle can be detected. The presence or absence of this *acoustic reflex* is important in determining the etiology of hearing loss as well as in the anatomic localization of facial nerve paralysis. The acoustic reflex can help differentiate between conductive hearing loss due to otosclerosis and that caused by an inner ear "third window": it is absent in otosclerosis and present in inner ear conductive hearing loss. Normal or elevated acoustic reflex thresholds in an individual with sensorineural hearing impairment suggests a cochlear hearing loss. An absent acoustic reflex in the setting of sensorineural hearing loss is not helpful in localizing the site of lesion. Assessment of *acoustic reflex decay* helps differentiate sensory from neural hearing losses. In neural hearing loss, the reflex adapts or decays with time.

Otoacoustic emissions (OAE) generated by outer hair cells only can be measured with microphones inserted into the external auditory canal. The emissions may be spontaneous or evoked with sound stimulation. The presence of OAEs indicates that the outer hair cells of the organ of Corti are intact and can be used to assess auditory thresholds and to distinguish sensory from neural hearing losses.

EVOKED RESPONSES

Electrocochleography measures the earliest evoked potentials generated in the cochlea and the auditory nerve. Receptor potentials recorded include the cochlear microphonic, generated by the outer hair cells of the organ of Corti, and the summating potential, generated by the inner hair cells in response to sound. The whole nerve action potential representing the composite firing of the first-order neurons can also be recorded during electrocochleography. Clinically, the test is useful in the diagnosis of Ménière's disease, where an elevation of the ratio of summating potential to action potential is seen.

Brainstem auditory evoked responses (BAERs), also known as *auditory brainstem responses* (ABRs), are useful in differentiating the site of sensorineural hearing loss. In response to sound, five distinct electrical potentials arising from different stations along the peripheral and central auditory pathway can be identified using computer averaging from scalp surface electrodes. BAERs are valuable in situations in which patients cannot or will not give reliable voluntary thresholds. They are also used to assess the integrity of the auditory nerve and brainstem in various clinical situations, including intraoperative monitoring and in determination of brain death.

The *vestibular-evoked myogenic potential (VEMP) test* elicits a vestibulocollic reflex whose afferent limb arises from acoustically sensitive cells in the saccule, with signals conducted via the inferior vestibular nerve. VEMP is a biphasic, short-latency response recorded from the tonically contracted sternocleidomastoid muscle in response to loud auditory clicks or tones. VEMPs may be diminished or absent in patients with early and late Ménière's disease, vestibular neuritis, benign paroxysmal positional vertigo, and vestibular schwannoma. On the other hand, the threshold for VEMPs may be lower in cases of superior canal dehiscence, other inner ear dehiscence, and perilymphatic fistula.

IMAGING STUDIES

The choice of radiologic tests is largely determined by whether the goal is to evaluate the bony anatomy of the external, middle, and inner ear or to image the auditory nerve and brain. Axial and coronal CT of the temporal bone with fine 0.3- to 0.6-mm cuts is ideal for determining the caliber of the external auditory canal, integrity of the ossicular chain, and presence of middle-ear or mastoid disease; it can also detect inner ear malformations. CT is also ideal for the detection of bone erosion with chronic otitis media and cholesteatoma. MRI is superior to CT for imaging of retrocochlear pathology such as vestibular schwannoma, meningioma, other lesions of the cerebellopontine angle, demyelinating lesions of the brainstem, and brain tumors. Both CT and MRI are equally capable of identifying inner ear malformations and assessing cochlear patency for preoperative evaluation of patients for cochlear implantation.

Treatment: Disorders of the Sense of Hearing

In general, conductive hearing losses are amenable to surgical correction, while sensorineural hearing losses are more difficult to manage. Atresia of the ear canal can be surgically repaired, often with significant improvement in hearing. Tympanic membrane perforations due to chronic otitis media or trauma can be repaired with an outpatient tympanoplasty.

Likewise, conductive hearing loss associated with otosclerosis can be treated by stapedectomy, which is successful in 90–95% of cases. Tympanostomy tubes allow the prompt return of normal hearing in individuals with middle ear effusions. Hearing aids are effective and well tolerated in patients with conductive hearing losses.

Patients with mild, moderate, and severe sensorineural hearing losses are regularly rehabilitated with hearing aids of varying configuration and strength. Hearing aids have been improved to provide greater fidelity and have been miniaturized. The current generation of hearing aids can be placed entirely within the ear canal, thus reducing any stigma associated with their use. In general, the more severe the hearing impairment, the larger the hearing aid required for auditory rehabilitation. Digital hearing aids lend themselves to individual programming, and multiple and directional microphones at the ear level may be helpful in noisy surroundings. Since all hearing aids amplify noise as well as speech, the only absolute solution to the problem of noise is to place the microphone closer to the speaker than the noise source. This arrangement is not possible with a self-contained, cosmetically acceptable device. A significant limitation of rehabilitation with a hearing aid is that while it is able to enhance detection of sound with amplification, it cannot restore clarity of hearing that is lost with presbycusis.

Patients with unilateral deafness have difficulty with sound localization and reduced clarity of hearing in background noise. They may benefit from a CROS (contralateral routing of signal) hearing aid in which a microphone is placed on the hearing-impaired side and the sound is transmitted to the receiver placed on the contralateral ear. The same result may be obtained with a bone-anchored hearing aid (BAHA), in which a hearing aid clamps to a screw osseointegrated into the skull on the hearing-impaired side. Like the CROS hearing aid, the BAHA transfers the acoustic signal to the contralateral hearing ear, but it does so by vibrating the skull. Patients with profound deafness on one side and some hearing loss in the better ear are candidates for a BICROS hearing aid; it differs from the CROS hearing aid in that the patient wears a hearing aid, and not simply a receiver, in the better ear. Unfortunately, CROS and BAHA devices are often judged by patients to be unsatisfactory.

In many situations, including lectures and the theater, hearing-impaired persons benefit from assistive devices that are based on the principle of having the speaker closer to the microphone than any source of noise. Assistive devices include infrared and frequency-modulated (FM) transmission as well as an electromagnetic loop around the room for transmission to the individual's hearing aid. Hearing aids with telecoils can also be used with properly equipped telephones in the same way.

In the event that the hearing aid provides inadequate rehabilitation, cochlear implants may be appropriate. Criteria for implantation include severe to profound hearing loss with open-set sentence cognition of ≤40% under best aided conditions. Worldwide, nearly 200,000 hearing impaired children and adults have received cochlear implants. Cochlear implants are neural prostheses that convert sound energy to electrical energy and can be used to stimulate the auditory division of the eighth nerve directly. In most cases of profound hearing impairment, the auditory hair cells are lost but the ganglionic cells of the auditory division of the eighth nerve are preserved. Cochlear implants consist of electrodes that are inserted into the cochlea through the round window, speech processors that extract acoustical elements of speech for conversion to electrical currents, and a means of transmitting the electrical energy through the skin. Patients with implants experience sound that helps with speech reading, allows open-set word recognition, and helps in modulating the person's own voice. Usually, within the first 3–6 months after implantation, adult patients can understand speech without visual cues. With the current generation of multichannel cochlear implants, nearly 75% of patients are able to converse on the telephone. For individuals who have had both eighth nerves destroyed by trauma or bilateral vestibular schwannomas (e.g., neurofibromatosis type 2), brainstem auditory implants placed near the cochlear nucleus may provide auditory rehabilitation.

Tinnitus often accompanies hearing loss. As for background noise, tinnitus can degrade speech comprehension in individuals with hearing impairment. Therapy for tinnitus is usually directed toward minimizing the appreciation of tinnitus. Relief of the tinnitus may be obtained by masking it with background music. Hearing aids are also helpful in tinnitus suppression, as are tinnitus maskers, devices that present a sound to the affected ear that is more pleasant to listen to than the tinnitus. The use of a tinnitus masker is often followed by several hours of inhibition of the tinnitus. Antidepressants have been shown to be beneficial in helping patients cope with tinnitus.

Hard-of-hearing individuals often benefit from a reduction in unnecessary noise in the environment (e.g., radio or television) to enhance the signal-to-noise ratio. Speech comprehension is aided by lip reading; therefore, the impaired listener should be seated so that the face of the speaker is well illuminated and easily seen. Although speech should be in a loud, clear voice, one should be aware that in sensorineural hearing losses in general and in hard-of-hearing elderly in particular, recruitment (abnormal perception of loud sounds) may be troublesome. Above all, optimal communication cannot take place without both parties giving it their full and undivided attention.

Prevention

Conductive hearing losses may be prevented by prompt antibiotic therapy of adequate duration for AOM and by ventilation of the middle ear with tympanostomy tubes in middle-ear effusions lasting ≥12 weeks. Loss of vestibular function and deafness due to aminoglycoside antibiotics can largely be prevented by careful monitoring of serum peak and trough levels.

Some 10 million Americans have noise-induced hearing loss, and 20 million are exposed to hazardous noise in their employment. Noise-induced hearing loss can be prevented by avoidance of exposure to loud noise or by regular use of ear plugs or fluid-filled ear muffs to attenuate intense sound. High-risk activities for noise-induced hearing loss include wood and metal working with electrical equipment and target practice and hunting with small firearms. All internal-combustion and electric engines, including snow and leaf blowers, snowmobiles, outboard motors, and chain saws, require protection of the user with hearing protectors. Virtually all noise-induced hearing loss is preventable through education, which should begin before the teenage years. Programs of industrial conservation of hearing are required by Occupational Safety and Health Administration (OSHA) when the exposure over an 8-h period averages 85 dB. OSHA mandates that workers in such noisy environments have hearing monitoring and protection programs that include a pre-employment screen, annual audiologic assessment, as well as the mandatory use of hearing protectors. Exposure to loud sounds above 85 dB in the work environment is restricted by OSHA, with halving of allowed exposure time for each increment of 5 dB above this threshold: for example 90 dB exposure is permitted for 8 h; 95 dB for 4 h, and 100 dB for 2 h.

FURTHER READINGS

Bishop CE, Eby TL: The current status of audiologic rehabilitation for profound unilateral sensorineural hearing loss. Laryngoscope 120:552, 2010 [PMID: 20014322]

Hilgert N et al: Function and expression pattern of nonsyndromic deafness genes. Curr Mol Med 9:546, 2009 [PMID:

19601806]

Lalwani AK (ed): *Current Diagnosis and Treatment in Otolaryngology—Head & Neck Surgery* , 3rd ed. New York, McGraw-Hill, 2011

Merchant SN, Rosowski JJ: Conductive hearing loss caused by third-window lesions of the inner ear. Otol Neurotol 29:282, 2008 [PMID: 18223508]

Sprinzl GM, Riechelmann H: Current trends in treating hearing loss in elderly people: A review of the technology and treatment options—a mini-review. Gerontology 56:351, 2010 [PMID: 20090297]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PHARYNGITIS, SINUSITIS, OTITIS, AND OTHER UPPER RESPIRATORY TRACT INFECTIONS: INTRODUCTION

Infections of the upper respiratory tract (URIs) have a tremendous impact on public health. They are among the most common reasons for visits to primary care providers, and although the illnesses are typically mild, their high incidence and transmission rates place them among the leading causes of time lost from work or school. Even though a minority (~25%) of cases are caused by bacteria, URIs are the leading diagnoses for which antibiotics are prescribed on an outpatient basis in the United States. The enormous consumption of antibiotics for these illnesses has contributed to the rise in antibiotic resistance among common community-acquired pathogens such as *Streptococcus pneumoniae*—a trend that in itself has had an enormous influence on public health.

Although most URIs are caused by viruses, distinguishing patients with primary viral infection from those with primary bacterial infection is difficult. Signs and symptoms of bacterial and viral URIs are typically indistinguishable. Until consistent, inexpensive, and rapid testing becomes available and is used widely, acute infections will be diagnosed largely on clinical grounds. The judicious use and potential for misuse of antibiotics in this setting pose definite challenges.

NONSPECIFIC INFECTIONS OF THE UPPER RESPIRATORY TRACT

Nonspecific URIs are a broadly defined group of disorders that collectively constitute the leading cause of ambulatory care visits in the United States. By definition, nonspecific URIs have no prominent localizing features. They are identified by a variety of descriptive names, including *acute infective rhinitis*, *acute rhinopharyngitis/nasopharyngitis*, *acute coryza*, and *acute nasal catarrh*, as well as by the inclusive label *common cold*.

ETIOLOGY

The large assortment of URI classifications reflects the wide variety of causative infectious agents and the varied manifestations of common pathogens. Nearly all nonspecific URIs are caused by viruses spanning multiple virus families and many antigenic types. For instance, there are at least 100 immunotypes of rhinovirus (Chap. 186), the most common cause of URI (~30–40% of cases); other causes include influenza virus (three immunotypes; Chap. 187) as well as parainfluenza virus (four immunotypes), coronavirus (at least three immunotypes), and adenovirus (47 immunotypes) (Chap. 186). Respiratory syncytial virus (RSV), a well-established pathogen in pediatric populations, is also a recognized cause of significant disease in elderly and immunocompromised individuals. A host of additional viruses, including some viruses not typically associated with URIs (e.g., enteroviruses, rubella virus, and varicella-zoster virus), account for a small percentage of cases in adults each year. Although new diagnostic modalities [e.g., nasopharyngeal swab for polymerase chain reaction (PCR)] can assign a viral etiology, there are few specific treatment options, and no pathogen is identified in a substantial proportion of cases. A specific diagnostic workup beyond a clinical diagnosis is generally unnecessary in an otherwise healthy adult.

CLINICAL MANIFESTATIONS

The signs and symptoms of nonspecific URI are similar to those of other URIs but lack a pronounced localization to one particular anatomic location, such as the sinuses, pharynx, or lower airway. Nonspecific URI commonly presents as an acute, mild, and self-limited catarrhal syndrome with a median duration of ~1 week (range, 2–10 days). Signs and symptoms are diverse and frequently variable across patients. The principal signs and symptoms of nonspecific URI include rhinorrhea (with or without purulence), nasal congestion, cough, and sore throat. Other manifestations, such as fever, malaise, sneezing, lymphadenopathy, and hoarseness, are more variable, with fever more common among infants and young children. This varying presentation may reflect differences in host response as well as in infecting organisms; myalgias and fatigue, for example, sometimes are seen with influenza and parainfluenza infections, whereas conjunctivitis may suggest infection with adenovirus or enterovirus. Findings on physical examination are frequently nonspecific and unimpressive. Between 0.5% and 2% of colds are complicated by secondary bacterial infections (e.g., rhinosinusitis, otitis media, and pneumonia), particularly in higher-risk populations such as infants, elderly persons, and chronically ill individuals. Secondary bacterial infections usually are associated with a prolonged course of illness, increased severity of illness, and localization of signs and symptoms, often as a rebound after initial clinical improvement. Purulent secretions from the nares or throat often are

misinterpreted as an indication of bacterial sinusitis or pharyngitis. These secretions, however, are also seen in nonspecific URI and, in the absence of other clinical features, are poor predictors of bacterial infection.

Treatment: Upper Respiratory Infections

Antibiotics have no role in the treatment of uncomplicated nonspecific URI, and their misuse probably facilitates the emergence of antimicrobial resistance; even in healthy volunteers, a single course of azithromycin or clarithromycin can lead to macrolide resistance among oral streptococci months later. In the absence of clinical evidence of bacterial infection, treatment remains entirely symptom-based, with use of decongestants and nonsteroidal anti-inflammatory drugs. Other therapies directed at specific symptoms are often useful, including dextromethorphan for cough and lozenges with topical anesthetic for sore throat. Clinical trials of zinc, vitamin C, echinacea, and other alternative remedies have revealed no consistent benefit for the treatment of nonspecific URI.

INFECTIONS OF THE SINUS

Rhinosinusitis refers to an inflammatory condition involving the four paired structures surrounding the nasal cavities. Although most cases of sinusitis involve more than one sinus, the maxillary sinus is most commonly involved; next, in order of frequency, are the ethmoid, frontal, and sphenoid sinuses. Each sinus is lined with a respiratory epithelium that produces mucus, which is transported out by ciliary action through the sinus ostium and into the nasal cavity. Normally, mucus does not accumulate in the sinuses, which remain mostly sterile despite their adjacency to the bacterium-filled nasal passages. When the sinus ostia are obstructed, however, or when ciliary clearance is impaired or absent, the secretions can be retained, producing the typical signs and symptoms of sinusitis. As these secretions accumulate with obstruction, they become more susceptible to infection with a variety of pathogens, including viruses, bacteria, and fungi. Sinusitis affects a tremendous proportion of the population, accounts for millions of visits to primary care physicians each year, and is the fifth leading diagnosis for which antibiotics are prescribed. It typically is classified by duration of illness (acute vs. chronic); by etiology (infectious vs. noninfectious); and, when infectious, by the offending pathogen type (viral, bacterial, or fungal).

Acute Rhinosinusitis

Acute rhinosinusitis—defined as sinusitis of <4 weeks' duration—constitutes the vast majority of sinusitis cases. Most cases are diagnosed in the ambulatory care setting and occur primarily as a consequence of a preceding viral URI. Differentiating acute bacterial from viral sinusitis on clinical grounds is difficult. Therefore, it is perhaps not surprising that antibiotics are prescribed frequently (in 85–98% of all cases) for this condition.

ETIOLOGY

The ostial obstruction that results in rhinosinusitis can arise from both infectious and noninfectious causes. Noninfectious causes include allergic rhinitis (with either mucosal edema or polyp obstruction), barotrauma (e.g., from deep-sea diving or air travel), and exposure to chemical irritants. Obstruction also can occur with nasal and sinus tumors (e.g., squamous cell carcinoma) or granulomatous diseases (e.g., granulomatosis with polyangiitis (Wegener's) or rhinoscleroma), and conditions leading to altered mucus content (e.g., cystic fibrosis) can cause sinusitis through impaired mucus clearance. In ICUs, nasotracheal intubation and nasogastric tubes are major risk factors for nosocomial sinusitis.

Viral rhinosinusitis is far more common than bacterial sinusitis, although relatively few studies have sampled sinus aspirates for the presence of different viruses. In the studies that have done so, the viruses most commonly isolated—both alone and with bacteria—have been rhinovirus, parainfluenza virus, and influenza virus. Bacterial causes of sinusitis have been better described. Among community-acquired cases, *S. pneumoniae* and nontypable *Haemophilus influenzae* are the most common pathogens, accounting for 50–60% of cases. *Moraxella catarrhalis* causes disease in a significant percentage (20%) of children but a lesser percentage in adults. Other streptococcal species and *Staphylococcus aureus* cause only a small percentage of cases, although there is increasing concern about community-acquired methicillin-resistant *S. aureus* (MRSA) as an emerging cause. It is difficult to assess whether a cultured bacterium represents a true infecting organism, an insufficiently deep sample (which would not be expected to be sterile), or—especially in the case of previous sinus surgeries—a colonizing organism. Anaerobes occasionally are found in association with infections of the roots of premolar teeth that spread into the adjacent maxillary sinuses. The role of *Chlamydophila pneumoniae* and *Mycoplasma pneumoniae* in the pathogenesis of acute sinusitis is unclear. Nosocomial cases commonly are associated with bacteria found in the hospital environment, including*S. aureus*, *Pseudomonas aeruginosa*, *Serratia marcescens*, *Klebsiella pneumoniae*, and *Enterobacter* species. Often, these infections are polymicrobial and involve organisms that are highly resistant to numerous antibiotics. Fungi are also established causes of sinusitis, although most acute cases are in immunocompromised patients and represent invasive, life-threatening infections. The best-known example is rhinocerebral mucormycosis caused by fungi of the order Mucorales, which includes *Rhizopus*, *Rhizomucor*, *Mucor*, *Mycocladus (formerly Absidia)*, and *Cunninghamella* (Chap.

205). These infections classically occur in diabetic patients with ketoacidosis but also can develop in transplant recipients, patients with hematologicmalignancies, and patients receiving chronic glucocorticoid or deferoxamine therapy. Other hyaline molds, such as *Aspergillus* and *Fusarium* species, are also occasional causes of this disease.

CLINICAL MANIFESTATIONS

Most cases of acute sinusitis present after or in conjunction with a viral URI, and it can be difficult to discriminate the clinical features of one from the other. A large proportion of patients with colds have sinus inflammation, although, as previously stated, true bacterial sinusitis complicates only 0.2–2% of these viral infections. Common presenting symptoms of sinusitis include nasal drainage and congestion, facial pain or pressure, and headache. Thick, purulent or discolored nasal discharge is often thought to indicate bacterial sinusitis but also occurs early in viral infections such as the common cold and is not specific to bacterial infection. Other nonspecific manifestations include cough, sneezing, and fever. Tooth pain, most often involving the upper molars, as well as halitosis can be associated with bacterial sinusitis.

In acute sinusitis, sinus pain or pressure often localizes to the involved sinus (particularly the maxillary sinus) and can be worse when the patient bends over or is supine. Although rare, manifestations of advanced sphenoid or ethmoid sinus infection can be profound, including severe frontal or retroorbital pain radiating to the occiput, thrombosis of the cavernous sinus, and signs of orbital cellulitis. Acute focal sinusitis is uncommon but should be considered over the maxillary sinus and fever in patients with severe symptoms, regardless of illness duration. Similarly, patients with advanced frontal sinusitis can present with a condition known as *Pott's puffy tumor* , with soft tissue swelling and pitting edema over the frontal bone from a communicating subperiosteal abscess. Life-threatening complications of sinusitis include meningitis, epidural abscess, and cerebral abscess.

Patients with acute fungal rhinosinusitis (such as mucormycosis; Chap. 205) often present with symptoms related to pressure effects, particularly when the infection has spread to the orbits and cavernous sinus. Signs such as orbital swelling and cellulitis, proptosis, ptosis, and decreased extraocular movement are common, as is retroorbital or periorbital pain. Nasopharyngeal ulcerations, epistaxis, and headaches are also common, and involvement of cranial nerves V and VII has been described in more advanced cases. Bony erosion may be evident on examination. Often the patient does not appear seriously ill despite the rapidly progressive nature of these infections.

Patients with acute nosocomial sinusitis are often critically ill and thus do not manifest the typical clinical features of sinus disease. This diagnosis should be suspected, however, when hospitalized patients who have appropriate risk factors (e.g., nasotracheal intubation) develop fever without another apparent cause.

DIAGNOSIS

Distinguishing viral from bacterial rhinosinusitis in the ambulatory setting is usually difficult because of the relatively low sensitivity and specificity of the common clinical features. One clinical feature that has been used to help guide diagnostic and therapeutic decision making is illness duration. Because acute bacterial sinusitis is uncommon in patients whose symptoms have lasted <10 days, expert panels now recommend reserving this diagnosis for patients with "persistent" symptoms (i.e., symptoms lasting >10 days in adults or >10–14 days in children) accompanied by the three cardinal signs of purulent nasal discharge, nasal obstruction, and facial pain (**Table 31-1)**. Even among patients who meet these criteria, only 40–50% have true bacterial sinusitis. The use of CT or sinus radiography is not recommended for acute disease, particularly early in the course of illness (i.e., at <10 days) in light of the high prevalence of similar abnormalities among patients with acute viral rhinosinusitis. In the evaluation of persistent, recurrent, or chronic sinusitis, CT of the sinuses is the radiographic study of choice.

Table 31-1 Guidelines for the Diagnosis and Treatment of Acute Sinusitis

Age Group	Diagnostic Criteria	Treatment Recommendations[a]
Adults	Moderate symptoms (e.g., nasal purulence/congestion or cough) for >10 d *or* Severe symptoms of any duration, including unilateral/focal facial swelling or tooth pain	*Initial therapy:* Amoxicillin, 500 mg PO tid or 875 mg PO bid *Penicillin allergy:* TMP-SMX, 1 DS tablet PO bid for 10–14 d *Exposure to antibiotics within 30 d or >30% prevalence of penicillin-resistant Streptococcus pneumoniae:* Amoxicillin/clavulanate (extended release), 2000 mg PO bid; *or* Antipneumococcal fluoroquinolone (e.g., levofloxacin,

		500 mg PO qd) *Recent treatment failure:* Amoxicillin/clavulanate (extended release), 2000 mg PO bid; *or* Amoxicillin, 1500 mg bid, plus clindamycin, 300 mg PO qid; *or* Antipneumococcal fluoroquinolone (e.g., levofloxacin, 500 mg PO qd)
Children	Moderate symptoms (e.g., nasal purulence/congestion or cough) for >10–14 d *or* Severe symptoms of any duration, including fever (>102°F), unilateral/focal facial swelling or pain	*Initial therapy:* Amoxicillin, 45–90 mg/kg qd (up to 2 g) PO in divided doses (bid or tid); *or* Cefuroxime axetil, 30 mg/kg qd PO in divided doses (bid); *or* Cefdinir, 14 mg/kg PO qd *Exposure to antibiotics within 30 d, recent treatment failure, or >30% prevalence of penicillin-resistant S. pneumoniae:* Amoxicillin, 90 mg/kg qd (up to 2 g) PO in divided doses (bid), plus clavulanate, 6.4 mg/kg qd PO in divided doses (bid) (extra-strength suspension); *or* Cefuroxime axetil, 30 mg/kg qd PO in divided doses (bid); *or* Cefdinir, 14 mg/kg PO qd

[a] Unless otherwise specified, the duration of therapy is generally 10 days, with appropriate follow-up.

Abbreviations: DS, double-strength; TMP-SMX, trimethoprim-sulfamethoxazole.

Sources: American Academy of Pediatrics Subcommittee on Management of Sinusitis and Committee on Quality Improvement, 2001; Rosenfeld et al., 2007.

The clinical history and/or setting often can identify cases of acute anaerobic bacterial sinusitis, acute fungal sinusitis, or sinusitis from noninfectious causes (e.g., allergic rhinosinusitis). In the case of an immunocompromised patient with acute fungal sinus infection, immediate examination by an otolaryngologist is required. Biopsy specimens from involved areas should be examined by a pathologist for evidence of fungal hyphal elements and tissue invasion. Cases of suspected acute nosocomial sinusitis should be confirmed by sinus CT. Because therapy should target the offending organism, a sinus aspirate for culture and susceptibility testing should be obtained, if possible, before the initiation of antimicrobial therapy.

Treatment: Acute Sinusitis

Most patients with a clinical diagnosis of acute rhinosinusitis improve without antibiotic therapy. The preferred initial approach in patients with mild to moderate symptoms of short duration is therapy aimed at symptom relief and facilitation of sinus drainage, such as with oral and topical decongestants, nasal saline lavage, and—at least in patients with a history of chronic sinusitis or allergies—nasal glucocorticoids. Newer studies have cast doubt on the role of antibiotics and inhaled glucocorticoids in acute rhinosinusitis. In one notable double-blind, randomized, placebo-controlled trial, neither antibiotics nor topical glucocorticoids had a significant impact on cure in the study population of patients, the majority of whom had had symptoms for <7 days. Adult patients whose condition does not improve after 7 days, children whose condition does not improve after 10–14 days, and patients with more severe symptoms (regardless of duration) should be treated with antibiotics (Table 31-1). Empirical therapy for adults with community-acquired sinusitis should consist of the narrowest-spectrum agent active against the most common bacterial pathogens, including *S. pneumoniae* and *H. influenzae,* e.g., amoxicillin. No clinical trials support the use of broad-spectrum agents for routine cases of bacterial sinusitis, even in the current era of drug-resistant *S. pneumoniae*. Up to 10% of patients do not respond to initial antimicrobial therapy; sinus aspiration and/or lavage by an otolaryngologist should be considered in these cases. Antibiotic prophylaxis to prevent episodes of recurrent acute bacterial sinusitis is not recommended.

Surgical intervention and IV antibiotic administration usually are reserved for patients with severe disease or those with intracranial complications such as abscess and orbital involvement. Immunocompromised patients with acute invasive fungal sinusitis usually require extensive surgical debridement and treatment with IV antifungal agents active against fungal hyphal forms, such as amphotericin B. Specific therapy should be individualized according to the fungal species and its

susceptibilities as well as the individual patient's characteristics.

Treatment of nosocomial sinusitis should begin with broad-spectrum antibiotics to cover common and often resistant pathogens such as *S. aureus* and gram-negative bacilli. Therapy then should be tailored to the results of culture and susceptibility testing of sinus aspirates.

Chronic Sinusitis

Chronic sinusitis is characterized by symptoms of sinus inflammation lasting >12 weeks. This illness is most commonly associated with either bacteria or fungi, and clinical cure in most cases is very difficult. Many patients have undergone treatment with repeated courses of antibacterial agents and multiple sinus surgeries, increasing their risk of colonization with antibiotic-resistant pathogens and of surgical complications. These patients often have high rates of morbidity, sometimes over many years.

In *chronic bacterial sinusitis*, infection is thought to be due to the impairment of mucociliary clearance from repeated infections rather than to persistent bacterial infection. The pathogenesis of this condition, however, is poorly understood. Although certain conditions (e.g., cystic fibrosis) can predispose patients to chronic bacterial sinusitis, most patients with chronic rhinosinusitis do not have obvious underlying conditions that result in the obstruction of sinus drainage, the impairment of ciliary action, or immune dysfunction. Patients experience constant nasal congestion and sinus pressure, with intermittent periods of greater severity, which may persist for years. CT can be helpful in determining the extent of disease, detecting an underlying anatomic defect or obstructing process (e.g., a polyp), and assessing the response to therapy. The management team should include an otolaryngologist to conduct endoscopic examinations and obtain tissue samples for histologic examination and culture. An endoscopy-derived culture not only has a higher yield but also allows direct visualization for abnormal anatomy.

Chronic fungal sinusitis is a disease of immunocompetent hosts and is usually noninvasive, although slowly progressive invasive disease sometimes is seen. Noninvasive disease, which typically is associated with hyaline molds such as *Aspergillus* species and dematiaceous molds such as *Curvularia* or *Bipolaris* species, can present as a number of different scenarios. In mild, indolent disease, which usually occurs in the setting of repeated failures of antibacterial therapy, only nonspecific mucosal changes may be seen on sinus CT. Although there is some controversy on this point, endoscopic surgery is usually curative in these cases, with no need for antifungal therapy. Another form of disease presents as long-standing, often unilateral symptoms and opacification of a single sinus on imaging studies as a result of a mycetoma (fungus ball) within the sinus. Treatment for this condition is also surgical, although systemic antifungal therapy may be warranted in the rare case in which bony erosion occurs. A third form of disease, known as *allergic fungal sinusitis*, is seen in patients with a history of nasal polyposis and asthma, who often have had multiple sinus surgeries. Patients with this condition produce a thick, eosinophil-laden mucus with the consistency of peanut butter that contains sparse fungal hyphae on histologic examination. These patients often present with pansinusitis.

Treatment: Chronic Sinusitis

Treatment of chronic bacterial sinusitis can be challenging and consists primarily of repeated culture-guided courses of antibiotics, sometimes for 3–4 weeks at a time; administration of intranasal glucocorticoids; and mechanical irrigation of the sinus with sterile saline solution. When this management approach fails, sinus surgery may be indicated and sometimes provides significant, albeit short-term, alleviation. Treatment of chronic fungal sinusitis consists of surgical removal of impacted mucus. Recurrence, unfortunately, is common.

INFECTIONS OF THE EAR AND MASTOID

Infections of the ear and associated structures can involve both the middle and the external ear, including the skin, cartilage, periosteum, ear canal, and tympanic and mastoid cavities. Both viruses and bacteria are known causes of these infections, some of which result in significant morbidity if not treated appropriately.

Infections of the External Ear Structures

Infections involving the structures of the external ear are often difficult to differentiate from noninfectious inflammatory conditions with similar clinical manifestations. Clinicians should consider inflammatory disorders as possible causes of external ear irritation, particularly in the absence of local or regional adenopathy. Aside from the more salient causes of inflammation, such as trauma, insect bite, and overexposure to sunlight or extreme cold, the differential diagnosis should include less common conditions such as autoimmune disorders (e.g., lupus or relapsing polychondritis) and vasculitides (e.g., granulomatosis with polyangiitis [Wegener's]).

AURICULAR CELLULITIS

Auricular cellulitis is an infection of the skin overlying the external ear and typically follows minor local trauma. It presents as the typical signs and symptoms of cellulitis, with tenderness, erythema, swelling, and warmth of the external ear (particularly the lobule) but without apparent involvement of the ear

canal or inner structures. Treatment consists of warm compresses and oral antibiotics such as dicloxacillin that are active against typical skin and soft tissue pathogens (specifically, *S. aureus* and streptococci). IV antibiotics such as a first-generation cephalosporin (e.g., cefazolin) or a penicillinase-resistant penicillin (e.g., nafcillin) occasionally are needed for more severe cases, with consideration of MRSA if either risk factors or failure of therapy point to this organism.

PERICHONDRITIS

Perichondritis, an infection of the perichondrium of the auricular cartilage, typically follows local trauma (e.g., ear piercing, burns, or lacerations). Occasionally, when the infection spreads down to the cartilage of the pinna itself, patients may develop chondritis. The infection may closely resemble auricular cellulitis, with erythema, swelling, and extreme tenderness of the pinna, although the lobule is less often involved in perichondritis. The most common pathogens are *P. aeruginosa* and *S. aureus*, although other gram-negative and gram-positive organisms occasionally are involved. Treatment consists of systemic antibiotics active against both *P. aeruginosa* and *S. aureus*. An antipseudomonal penicillin (e.g., piperacillin) or a combination of a penicillinase-resistant penicillin and an antipseudomonal quinolone (e.g., nafcillin plus ciprofloxacin) is typically used. Incision and drainage may be helpful for culture and for resolution of infection, which often takes weeks. When perichondritis fails to respond to adequate antimicrobial therapy, clinicians should consider a noninfectious inflammatory etiology such as relapsing polychondritis.

OTITIS EXTERNA

The term *otitis externa* refers to a collection of diseases involving primarily the auditory meatus. Otitis externa usually results from a combination of heat and retained moisture, with desquamation and maceration of the epithelium of the outer ear canal. The disease exists in several forms: localized, diffuse, chronic, and invasive. All forms are predominantly bacterial in origin, with *P. aeruginosa* and *S. aureus* the most common pathogens.

Acute localized otitis externa (*furunculosis*) can develop in the outer third of the ear canal, where skin overlies cartilage and hair follicles are numerous. As in furunculosis elsewhere on the body, *S. aureus* is the usual pathogen, and treatment typically consists of an oral antistaphylococcal penicillin (e.g., dicloxacillin), with incision and drainage in cases of abscess formation.

Acute diffuse otitis externa is also known as *swimmer's ear*, although it can develop in patients who have not recently been swimming. Heat, humidity, and the loss of protective cerumen lead to excessive moisture and elevation of the pH in the ear canal, which in turn lead to skin maceration and irritation. Infection may then occur; the predominant pathogen is *P. aeruginosa*, although other gram-negative and gram-positive organisms—and rarely yeasts—have been recovered from patients with this condition. The illness often starts with itching and progresses to severe pain, which usually is elicited by manipulation of the pinna or tragus. The onset of pain generally is accompanied by the development of an erythematous, swollen ear canal, often with scant white, clumpy discharge. Treatment consists of cleansing the canal to remove debris and enhance the activity of topical therapeutic agents—usually hypertonic saline or mixtures of alcohol and acetic acid. Inflammation also can be decreased by adding glucocorticoids to the treatment regimen or by using Burow's solution (aluminum acetate in water). Antibiotics are most effective when given topically. Otic mixtures provide adequate pathogen coverage; these preparations usually combine neomycin with polymyxin, with or without glucocorticoids. Systemic antimicrobial agents typically are reserved for severe disease or infections in immunocompromised hosts.

Chronic otitis externa is caused primarily by repeated local irritation, most commonly arising from persistent drainage from a chronic middle-ear infection. Other causes of repeated irritation, such as insertion of cotton swabs or other foreign objects into the ear canal, can lead to this condition, as can rare chronic infections such as syphilis, tuberculosis, and leprosy. Chronic otitis externa typically presents as erythematous, scaling dermatitis in which the predominant symptom is pruritus rather than pain; this condition must be differentiated from several others that produce a similar clinical picture, such as atopic dermatitis, seborrheic dermatitis, psoriasis, and dermatomycosis. Therapy consists of identifying and treating or removing the offending process, although successful resolution is frequently difficult.

Invasive otitis externa, also known as *malignant* or *necrotizing* otitis externa, is an aggressive and potentially life-threatening disease that occurs predominantly in elderly diabetic patients and other immunocompromised persons. The disease begins in the external canal as a soft tissue infection that progresses slowly over weeks to months and often is difficult to distinguish from a severe case of chronic otitis externa because of the presence of purulent otorrhea and an erythematous swollen ear and external canal. Severe, deep-seated otalgia, frequently out of proportion to findings on examination, is often noted and can help differentiate invasive from chronic otitis externa. The characteristic finding on examination is granulation tissue in the posteroinferior wall of the external canal, near the junction of bone and cartilage. If left unchecked, the infection can migrate to the base of the skull (resulting in skull-base osteomyelitis) and onto the meninges and brain, with a high-associated mortality rate. Cranial nerve involvement is seen occasionally, with the facial nerve usually

affected first and most often. Thrombosis of the sigmoid sinus can occur if the infection extends to that area. CT, which can reveal osseous erosion of the temporal bone and skull base, can be used to help determine the extent of disease, as can gallium and technetium-99 scintigraphy studies. *P. aeruginosa* is by far the most common pathogen, although *S. aureus*, *S. epidermidis*, *Aspergillus*, *Actinomyces*, and some gram-negative bacteria have also been associated with this disease. In all cases, the external ear canal should be cleansed and a biopsy specimen of the granulation tissue within the canal (or of deeper tissues) obtained for culture of the offending organism. IV antibiotic therapy should be given for a prolonged course (6–8 weeks) and directed specifically toward the recovered pathogen. For *P. aeruginosa*, the regimen typically includes an antipseudomonal penicillin or cephalosporin (e.g., piperacillin or ceftazidime) with an aminoglycoside. A fluoroquinolone antibiotic is frequently used in place of the aminoglycoside and can even be administered orally because of the excellent bioavailability of this drug class. In addition, antibiotic drops containing an agent active against *Pseudomonas* (e.g., ciprofloxacin) usually are prescribed and are combined with glucocorticoids to reduce inflammation. Cases of invasive *Pseudomonas* otitis externa recognized in the early stages sometimes can be treated with oral and otic fluoroquinolones alone, albeit with close follow-up. Extensive surgical debridement, once an important component of the treatment approach, is now rarely indicated.

In necrotizing otitis externa, recurrence is documented up to 20% of the time. Aggressive glycemic control in diabetics is important not only for effective treatment but also for prevention of recurrence. The role of hyperbaric oxygen has not been clearly established.

Infections of Middle-Ear Structures

Otitis media is an inflammatory condition of the middle ear that results from dysfunction of the eustachian tube in association with a number of illnesses, including URIs and chronic rhinosinusitis. The inflammatory response to these conditions leads to the development of a sterile transudate within the middle ear and mastoid cavities. Infection may occur if bacteria or viruses from the nasopharynx contaminate this fluid, producing an acute (or sometimes chronic) illness.

ACUTE OTITIS MEDIA

Acute otitis media results when pathogens from the nasopharynx are introduced into the inflammatory fluid collected in the middle ear (e.g., by nose blowing during a URI). The proliferation of these pathogens in this space leads to the development of the typical signs and symptoms of acute middle-ear infection. The diagnosis of acute otitis media requires the demonstration of fluid in the middle ear [with tympanic membrane (TM) immobility] and the accompanying signs or symptoms of local or systemic illness **(Table 31-2)**.

Table 31-2 Guidelines for the Diagnosis and Treatment of Acute Otitis Media

Illness Severity	Diagnostic Criteria	Treatment Recommendations
Mild to moderate	>2 yrs *or* 6 mo to 2 yrs without middle-ear effusion	*Observation alone* (deferring antibiotic therapy for 48–72 h and limiting management to symptom relief)
	<6 mo; *or* 6 mo to 2 yrs with middle-ear effusion (fluid in the middle ear, evidenced by decreased TM mobility, air/fluid level behind TM, bulging TM, purulent otorrhea) *and* acute onset of signs and symptoms of middle-ear inflammation, including fever, otalgia, decreased hearing, tinnitus, vertigo, erythematous TM; *or* >2 yrs with bilateral disease, TM perforation, high fever, immunocompromise, emesis	*Initial therapy* [a] Amoxicillin, 80–90 mg/kg qd (up to 2 g) PO in divided doses (bid or tid); *or* Cefdinir, 14 mg/kg qd PO in 1 dose or divided doses (bid); *or* Cefuroxime, 30 mg/kg qd PO in divided doses (bid); *or* Azithromycin, 10 mg/kg qd PO on day 1 followed by 5 mg/kg qd PO for 4 d *Exposure to antibiotics within 30 d or recent treatment failure* [a,b]*:* Amoxicillin, 90 mg/kg qd (up to 2 g) PO in divided doses (bid), plus

		clavulanate, 6.4 mg/kg qd PO in divided doses (bid); *or* Ceftriaxone, 50 mg/kg IV/IM qd for 3 d; *or* Clindamycin, 30–40 mg/kg qd PO in divided doses (tid)
Severe	As above, with temperature ≥39.0°C (102°F) *or* Moderate to severe otalgia	*Initial therapy* [a] Amoxicillin, 90 mg/kg qd (up to 2 g) PO in divided doses (bid), plus clavulanate, 6.4 mg/kg qd PO in divided doses (bid); *or* Ceftriaxone, 50 mg/kg IV/IM qd for 3 d *Exposure to antibiotics within 30 d or recent treatment failure* [a,b] Ceftriaxone, 50 mg/kg IV/IM qd for 3 d; *or* Clindamycin, 30–40 mg/kg qd PO in divided doses (tid); *or* Consider tympanocentesis with culture

[a] Duration (unless otherwise specified): 10 days for patients <6 years old and patients with severe disease; 5–7 days (with consideration of observation only in previously healthy individuals with mild disease) for patients ≥6 years old.

[b] Failure to improve and/or clinical worsening after 48–72 h of observation or treatment.

Abbreviation: TM, tympanic membrane.

Source: American Academy of Pediatrics Subcommittee on Management of Acute Otitis Media, 2004.

Etiology

Acute otitis media typically follows a viral URI. The causative viruses (most commonly RSV, influenza virus, rhinovirus, and enterovirus) can themselves cause subsequent acute otitis media; more often, they predispose the patient to bacterial otitis media. Studies using tympanocentesis have consistently found *S. pneumoniae* to be the most important bacterial cause, isolated in up to 35% of cases. *H. influenzae* (nontypable strains) and *M. catarrhalis* are also common bacterial causes of acute otitis media, and concern is increasing about community strains of MRSA as an emerging etiologic agent. Viruses, such as those mentioned above, have been recovered either alone or with bacteria in 17–40% of cases.

Clinical Manifestations

Fluid in the middle ear is typically demonstrated or confirmed with pneumatic otoscopy. In the absence of fluid, the tympanic membrane moves visibly with the application of positive and negative pressure, but this movement is dampened when fluid is present. With bacterial infection, the tympanic membrane can also be erythematous, bulging, or retracted and occasionally can perforate spontaneously. The signs and symptoms accompanying infection can be local or systemic, including otalgia, otorrhea, diminished hearing, fever, and irritability. Erythema of the tympanic membrane is often evident but is nonspecific as it frequently is seen in association with inflammation of the upper respiratory mucosa (e.g., during examination of young children). Other signs and symptoms that are occasionally reported include vertigo, nystagmus, and tinnitus.

Treatment: Acute Otitis Media

There has been considerable debate on the usefulness of antibiotics for the treatment of acute otitis media. A higher proportion of treated than untreated patients are free of illness 3–5 days after diagnosis. The difficulty of predicting which patients will benefit from antibiotic therapy has led to different approaches. In the Netherlands, for instance, physicians typically manage acute otitis media

with initial observation, administering anti-inflammatory agents for aggressive pain management and reserving antibiotics for high-risk patients, patients with complicated disease, or patients whose condition does not improve after 48–72 h. In contrast, many experts in the United States continue to recommend antibiotic therapy for children <6 months old in light of the higher frequency of secondary complications in this young and functionally immunocompromised population. However, observation without antimicrobial therapy is now the recommended option in the United States for acute otitis media in children ≥2 years of age and for mild to moderate disease without middle-ear effusion in children 6 months to 2 years of age. Treatment is typically indicated for patients <6 months old; for children 6 months to 2 years old who have middle-ear effusion and signs/symptoms of middle-ear inflammation; for all patients >2 years old who have bilateral disease, tympanic membrane perforation, immunocompromise, or emesis; and for any patient who has severe symptoms, including a fever ≥39°C or moderate to severe otalgia (Table 31-2).

Because most studies of the etiologic agents of acute otitis media consistently document similar pathogen profiles, therapy is generally empirical except in those few cases in which tympanocentesis is warranted—e.g., cases in newborns, cases refractory to therapy, and cases in patients who are severely ill or immunodeficient. Despite resistance to penicillin and amoxicillin in roughly one-quarter of *S. pneumoniae* isolates, one-third of *H. influenzae* isolates, and nearly all *M. catarrhalis* isolates, outcome studies continue to find that amoxicillin is as successful as any other agent, and it remains the drug of first choice in recommendations from multiple sources (Table 31-2). Therapy for uncomplicated acute otitis media typically is administered for 5–7 days to patients ≥6 years old; longer courses (e.g., 10 days) should be reserved for children <6 years old and patients with severe disease, in whom short-course therapy may be inadequate.

A switch in regimen is recommended if there is no clinical improvement by the third day of therapy in light of the possibility of infection with a β-lactamase-producing strain of *H. influenzae* or *M. catarrhalis* or with a strain of penicillin-resistant *S. pneumoniae*. Decongestants and antihistamines are frequently used as adjunctive agents to reduce congestion and relieve obstruction of the eustachian tube, but clinical trials have yielded no significant evidence of benefit with either class of agents.

RECURRENT ACUTE OTITIS MEDIA

Recurrent acute otitis media (more than three episodes within 6 months or four episodes within 12 months) generally is due to relapse or reinfection, although data indicate that the majority of early recurrences are new infections. In general, the same pathogens responsible for acute otitis media cause recurrent disease; even so, the recommended treatment consists of antibiotics active against β-lactamase-producing organisms. Antibiotic prophylaxis [e.g., with trimethoprim-sulfamethoxazole (TMP-SMX) or amoxicillin] can reduce recurrences in patients with recurrent acute otitis media by an average of one episode per year, but this benefit is small compared with the cost of the drug and the high likelihood of colonization with antibiotic-resistant pathogens. Other approaches, including placement of tympanostomy tubes, adenoidectomy, and tonsillectomy plus adenoidectomy, are of questionable overall value in light of the relatively small benefit compared with the potential for complications.

SEROUS OTITIS MEDIA

In serous otitis media (otitis media with effusion), fluid is present in the middle ear for an extended period in the absence of signs and symptoms of infection. In general, acute effusions are self-limited; most resolve in 2–4 weeks. In some cases, however (in particular after an episode of acute otitis media), effusions can persist for months. These chronic effusions are often associated with significant hearing loss in the affected ear. In younger children, persistent effusions and decreased hearing can be associated with impairment of language acquisition skills. The great majority of cases of otitis media with effusion resolve spontaneously within 3 months without antibiotic therapy. Antibiotic therapy or myringotomy with insertion of tympanostomy tubes typically is reserved for patients in whom bilateral effusion (1) has persisted for at least 3 months and (2) is associated with significant bilateral hearing loss. With this conservative approach and the application of strict diagnostic criteria for acute otitis media and otitis media with effusion, it is estimated that 6–8 million courses of antibiotics could be avoided each year in the United States.

CHRONIC OTITIS MEDIA

Chronic suppurative otitis media is characterized by persistent or recurrent purulent otorrhea in the setting of tympanic membrane perforation. Usually, there is also some degree of conductive hearing loss. This condition can be categorized as active or inactive. Inactive disease is characterized by a central perforation of the tympanic membrane, which allows drainage of purulent fluid from the middle ear. When the perforation is more peripheral, squamous epithelium from the auditory canal may invade the middle ear through the perforation, forming a mass of keratinaceous debris (*cholesteatoma*) at the site of invasion. This mass can enlarge and has the potential to erode bone and promote further infection, which can lead to meningitis, brain abscess, or paralysis of cranial nerve VII. Treatment of chronic active otitis media is surgical; mastoidectomy, myringoplasty, and tympanoplasty can be performed as outpatient surgical procedures, with an overall success rate of ~80%. Chronic inactive otitis media is more difficult to cure, usually requiring repeated courses of topical antibiotic drops during periods of drainage. Systemic antibiotics may offer better cure rates, but their role in the

treatment of this condition remains unclear.

MASTOIDITIS

Acute mastoiditis was relatively common among children before the introduction of antibiotics. Because the mastoid air cells connect with the middle ear, the process of fluid collection and infection is usually the same in the mastoid as in the middle ear. Early and frequent treatment of acute otitis media is most likely the reason that the incidence of acute mastoiditis has declined to only 1.2–2.0 cases per 100,000 person-years in countries with high prescribing rates for acute otitis media.

In countries such as the Netherlands, where antibiotics are used sparingly for acute otitis media, the incidence rate of acute mastoiditis is roughly twice that in countries like the United States. However, neighboring Denmark has a rate of acute mastoiditis similar to that in the Netherlands but an antibiotic-prescribing rate for acute otitis media more similar to that in the United States.

In typical acute mastoiditis, purulent exudate collects in the mastoid air cells **(Fig. 31-1)**, producing pressure that may result in erosion of the surrounding bone and formation of abscess-like cavities that are usually evident on CT. Patients typically present with pain, erythema, and swelling of the mastoid process along with displacement of the pinna, usually in conjunction with the typical signs and symptoms of acute middle-ear infection. Rarely, patients can develop severe complications if the infection tracks under the periosteum of the temporal bone to cause a subperiosteal abscess, erodes through the mastoid tip to cause a deep neck abscess, or extends posteriorly to cause septic thrombosis of the lateral sinus.

Figure 31-1

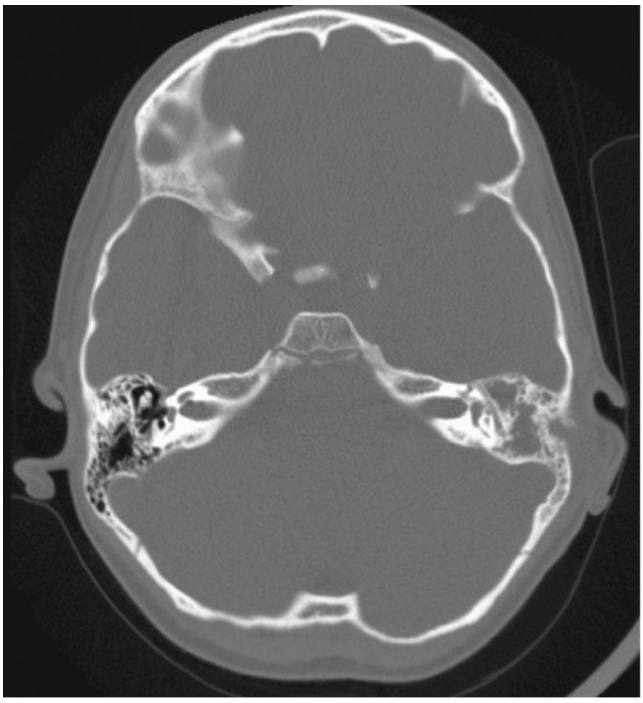

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute mastoiditis. Axial CT image shows an acute fluid collection within the mastoid air cells on the left.

Purulent fluid should be cultured whenever possible to help guide antimicrobial therapy. Initial empirical therapy usually is directed against the typical organisms associated with acute otitis media, such as *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*. Some patients with more severe or prolonged courses of illness should be treated for infection with *S. aureus* and gram-negative bacilli (including *Pseudomonas*). Broad empirical therapy is usually narrowed once culture results become available. Most patients can be treated conservatively with IV antibiotics; surgery (cortical mastoidectomy) can be reserved for complicated cases and those in which conservative treatment has failed.

INFECTIONS OF THE PHARYNX AND ORAL CAVITY

Oropharyngeal infections range from mild, self-limited viral illnesses to serious, life-threatening bacterial infections. The most common presenting symptom is sore throat—one of the most common reasons for ambulatory care visits by both adults and children. Although sore throat is a symptom in many noninfectious illnesses as well, the overwhelming majority of patients with a new sore throat

have acute pharyngitis of viral or bacterial etiology.

Acute Pharyngitis

Millions of visits to primary care providers each year are for sore throat; the majority of cases of acute pharyngitis are caused by typical respiratory viruses. The most important source of concern is infection with group A β-hemolytic *Streptococcus* (*S. pyogenes*) that is associated with acute glomerulonephritis and acute rheumatic fever. The risk of rheumatic fever can be reduced by timely penicillin therapy.

ETIOLOGY

A wide variety of organisms cause acute pharyngitis. The relative importance of the different pathogens can only be estimated, since a significant proportion of cases (~30%) have no identified cause. Together, respiratory viruses are the most common identifiable cause of acute pharyngitis, with rhinoviruses and coronaviruses accounting for large proportions of cases (~20% and at least 5%, respectively). Influenza virus, parainfluenza virus, and adenovirus also account for a measurable share of cases, the latter as part of the more clinically severe syndrome of pharyngoconjunctival fever. Other important but less common viral causes include herpes simplex virus (HSV) types 1 and 2, coxsackievirus A, cytomegalovirus (CMV), and Epstein-Barr virus (EBV). Acute HIV infection can present as acute pharyngitis and should be considered in at-risk populations.

Acute bacterial pharyngitis is typically caused by *S. pyogenes*, which accounts for ~5–15% of all cases of acute pharyngitis in adults; rates vary with the season and with utilization of the health care system. Group A streptococcal pharyngitis is primarily a disease of children 5–15 years of age; it is uncommon among children <3 years old, as is rheumatic fever. Streptococci of groups C and G account for a minority of cases, although these serogroups are nonrheumatogenic. *Fusobacterium necrophorum* has been increasingly recognized as a cause of pharyngitis in adolescents and young adults and is isolated nearly as often as group A streptococci. This information is important because of the rare but life-threatening Lemierre's disease, which is generally associated with *F. necrophorum* and is usually preceded by pharyngitis (see "Oral Infections," below). The remaining bacterial causes of acute pharyngitis are seen infrequently (<1% of cases each) but should be considered in appropriate exposure groups because of the severity of illness if left untreated; these etiologic agents include *Neisseria gonorrhoeae*, *Corynebacterium diphtheriae*, *Corynebacterium ulcerans*, *Yersinia enterocolitica*, and *Treponema pallidum* (in secondary syphilis). Anaerobic bacteria can also cause acute pharyngitis (*Vincent's angina*) and can contribute to more serious polymicrobial infections, such as peritonsillar or retropharyngeal abscess (see below). Atypical organisms such as *M. pneumoniae* and *C. pneumoniae* have been recovered from patients with acute pharyngitis; whether these agents are commensals or causes of acute infection is debatable.

CLINICAL MANIFESTATIONS

Although the signs and symptoms accompanying acute pharyngitis are not reliable predictors of the etiologic agent, the clinical presentation occasionally suggests that one etiology is more likely than another. Acute pharyngitis due to respiratory viruses such as rhinovirus or coronavirus is usually not severe and typically is associated with a constellation of coryzal symptoms better characterized as nonspecific URI. Findings on physical examination are uncommon; fever is rare, and tender cervical adenopathy and pharyngeal exudates are not seen. In contrast, acute pharyngitis from influenza virus can be severe and is much more likely to be associated with fever as well as with myalgias, headache, and cough. The presentation of pharyngoconjunctival fever due to adenovirus infection is similar. Since pharyngeal exudate may be present on examination, this condition can be difficult to differentiate from streptococcal pharyngitis. However, adenoviral pharyngitis is distinguished by the presence of conjunctivitis in one-third to one-half of patients. Acute pharyngitis from primary HSV infection can also mimic streptococcal pharyngitis in some cases, with pharyngeal inflammation and exudate, but the presence of vesicles and shallow ulcers on the palate can help differentiate the two diseases. This HSV syndrome is distinct from pharyngitis caused by coxsackievirus (*herpangina*), which is associated with small vesicles that develop on the soft palate and uvula and then rupture to form shallow white ulcers. Acute exudative pharyngitis coupled with fever, fatigue, generalized lymphadenopathy, and (on occasion) splenomegaly is characteristic of infectious mononucleosis due to EBV or CMV. Acute primary infection with HIV is frequently associated with fever and acute pharyngitis as well as with myalgias, arthralgias, malaise, and occasionally a nonpruritic maculopapular rash, which may be followed by lymphadenopathy and mucosal ulcerations without exudate.

The clinical features of acute pharyngitis caused by streptococci of groups A, C, and G are all similar, ranging from a relatively mild illness without many accompanying symptoms to clinically severe cases with profound pharyngeal pain, fever, chills, and abdominal pain. A hyperemic pharyngeal membrane with tonsillar hypertrophy and exudate is usually seen, along with tender anterior cervical adenopathy. Coryzal manifestations, including cough, are typically absent; when present, they suggest a viral etiology. Strains of *S. pyogenes* that generate erythrogenic toxin can also produce scarlet fever characterized by an erythematous rash and strawberry tongue. The other types of acute bacterial pharyngitis (e.g., gonococcal, diphtherial, and yersinial) often present as exudative pharyngitis with or without other clinical features. Their etiologies are often suggested only by the clinical history.

DIAGNOSIS

The primary goal of diagnostic testing is to separate acute streptococcal pharyngitis from pharyngitis of other etiologies (particularly viral) so that antibiotics can be prescribed more efficiently for patients to whom they may be beneficial. The most appropriate standard for the diagnosis of streptococcal pharyngitis, however, has not been established definitively. Throat swab culture is generally regarded as the most appropriate but cannot distinguish between infection and colonization and requires 24–48 h to yield results that vary with technique and culture conditions. Rapid antigen-detection tests offer good specificity (>90%) but lower sensitivity when implemented in routine practice. The sensitivity has also been shown to vary across the clinical spectrum of disease (65–90%). Several clinical prediction systems **(Table 31-3)** can increase the sensitivity of rapid antigen-detection tests to >90% in controlled settings. Since the sensitivities achieved in routine clinical practice are often lower, several medical and professional societies continue to recommend that all negative rapid antigen-detection tests in children be confirmed by a throat culture to limit transmission and complications of illness caused by group A streptococci. The Centers for Disease Control and Prevention, the Infectious Diseases Society of America, and the American Academy of Family Physicians do not recommend backup culture when adults have negative results in a highly sensitive rapid antigen-detection test, however, because of the lower prevalence and smaller benefit in this age group.

Table 31-3 Guidelines for the Diagnosis and Treatment of Acute Pharyngitis

Age Group	Diagnostic Criteria	Treatment Recommendations[a]
Adults	Clinical suspicion of streptococcal pharyngitis (e.g., fever, tonsillar swelling, exudate, enlarged/tender anterior cervical lymph nodes, absence of cough or coryza)[b] *with:* History of rheumatic fever *or* Documented household exposure *or* Positive rapid strep screen	Penicillin VK, 500 mg PO tid; *or* Amoxicillin, 500 mg PO bid; *or* Erythromycin, 250 mg PO qid; *or* Benzathine penicillin G, single dose of 1.2 million units IM
Children	Clinical suspicion of streptococcal pharyngitis (e.g., tonsillar swelling, exudate, enlarged/tender anterior cervical lymph nodes, absence of coryza) *with:* History of rheumatic fever *or* Documented household exposure *or* Positive rapid strep screen *or* Positive throat culture (for patients with negative rapid strep screen)	Amoxicillin, 45 mg/kg qd PO in divided doses (bid or tid); *or* Penicillin VK, 50 mg/kg qd PO in divided doses (bid); *or* Cephalexin, 50 mg/kg qd PO in divided doses (qid); *or* Benzathine penicillin G, single dose of 25,000 units/kg IM

[a] Unless otherwise specified, the duration of therapy is generally 10 days, with appropriate follow-up.

[b] Some organizations support treating adults who have these symptoms and signs without administering a rapid streptococcal antigen test.

Sources: Cooper et al, 2001; Schwartz et al, 1998.

Cultures and rapid diagnostic tests for other causes of acute pharyngitis, such as influenza virus, adenovirus, HSV, EBV, CMV, and *M. pneumoniae*, are available in some locations and can be used when these infections are suspected. The diagnosis of acute EBV infection depends primarily on the detection of antibodies to the virus with a heterophile agglutination assay (monospot slide test) or enzyme-linked immunosorbent assay. Testing for HIV RNA or antigen (p24) should be performed when acute primary HIV infection is suspected. If other bacterial causes are suspected (particularly *N. gonorrhoeae*, *C. diphtheriae*, or *Y. enterocolitica*), specific cultures should be requested since these organisms may be missed on routine throat swab culture.

Treatment: Pharyngitis

Antibiotic treatment of pharyngitis due to *S. pyogenes* confers numerous benefits, including a decrease in the risk of rheumatic fever. The magnitude of this benefit is fairly small, since rheumatic fever is now a rare disease, even among untreated patients. Nevertheless, when therapy is started within 48 h of

illness onset, symptom duration is decreased. An additional benefit of therapy is the potential to reduce the transmission of streptococcal pharyngitis, particularly in areas of overcrowding or close contact. Antibiotic therapy for acute pharyngitis is therefore recommended in cases in which *S. pyogenes* is confirmed as the etiologic agent by rapid antigen-detection test or throat swab culture. Otherwise, antibiotics should be given in routine cases only when another bacterial cause has been identified. Effective therapy for streptococcal pharyngitis consists of either a single dose of IM benzathine penicillin or a full 10-day course of oral penicillin (Table 31-3).

Erythromycin can be used in place of penicillin, although resistance to erythromycin among *S. pyogenes* strains in some parts of the world (particularly Europe) can prohibit the use of this drug. Newer (and more expensive) antibiotics are also active against streptococci but offer no greater efficacy than the agents mentioned above. Testing for cure is unnecessary and may reveal only chronic colonization. There is no evidence to support antibiotic treatment of group C or G streptococcal pharyngitis or pharyngitis in which mycoplasmas or chlamydiae have been recovered. Penicillin prophylaxis (benzathine penicillin G, 1.2 million units IM every 3–4 weeks) is indicated for patients at risk of recurrent rheumatic fever.

Treatment of viral pharyngitis is entirely symptom-based except in infection with influenza virus or HSV. For influenza, the repertoire of therapeutic agents includes the adamantanes amantadine and rimantadine and the neuraminidase inhibitors oseltamivir and zanamivir. Administration of all these agents needs to be started within 36–48 h of symptom onset to reduce illness duration meaningfully. Among these agents, only oseltamivir and zanamivir are active against both influenza A and influenza B and therefore can be used when local patterns of infection and antiviral resistance are unknown. Oropharyngeal HSV infection sometimes responds to treatment with antiviral agents such as acyclovir, although these drugs are often reserved for immunosuppressed patients.

COMPLICATIONS

Although rheumatic fever is the best-known complication of acute streptococcal pharyngitis, the risk of its following acute infection remains quite low. Other complications include acute glomerulonephritis and numerous suppurative conditions, such as peritonsillar abscess (*quinsy*), otitis media, mastoiditis, sinusitis, bacteremia, and pneumonia—all of which occur at low rates. Although antibiotic treatment of acute streptococcal pharyngitis can prevent the development of rheumatic fever, there is no evidence that it can prevent acute glomerulonephritis. Some evidence supports antibiotic use to prevent the suppurative complications of streptococcal pharyngitis, particularly peritonsillar abscess, which can also involve oral anaerobes such as *Fusobacterium*. Abscesses usually are accompanied by severe pharyngeal pain, dysphagia, fever, and dehydration; in addition, medial displacement of the tonsil and lateral displacement of the uvula are often evident on examination. Although early use of IV antibiotics (e.g., clindamycin, penicillin G with metronidazole) may obviate the need for surgical drainage in some cases, treatment typically involves needle aspiration or incision and drainage.

Oral Infections

Aside from periodontal disease such as gingivitis, infections of the oral cavity most commonly involve HSV or *Candida* species. In addition to causing painful cold sores on the lips, HSV can infect the tongue and buccal mucosa, causing the formation of irritating vesicles. Although topical antiviral agents (e.g., acyclovir and penciclovir) can be used externally for cold sores, oral or IV acyclovir is often needed for primary infections, extensive oral infections, and infections in immunocompromised patients. Oropharyngeal candidiasis (*thrush*) is caused by a variety of *Candida* species, most often *C. albicans*. Thrush occurs predominantly in neonates, immunocompromised patients (especially those with AIDS), and recipients of prolonged antibiotic or glucocorticoid therapy. In addition to sore throat, patients often report a burning tongue, and physical examination reveals friable white or gray plaques on the gingiva, tongue, and oral mucosa. Treatment that usually consists of an oral antifungal suspension (nystatin or clotrimazole) or oral fluconazole, is frequently successful. In the uncommon cases of fluconazole-refractory thrush that are seen in some patients with AIDS, other therapeutic options include oral formulations of itraconazole, amphotericin B, posaconazole, or voriconazole as well as an IV echinocandin (caspofungin, micafungin, or anidulafungin) or amphotericin B deoxycholate, if needed. In these cases, therapy based on culture and susceptibility test results is ideal.

Vincent's angina, also known as *acute necrotizing ulcerative gingivitis* or *trench mouth*, is a unique and dramatic form of gingivitis characterized by painful, inflamed gingiva with ulcerations of the interdental papillae that bleed easily. Since oral anaerobes are the cause, patients typically have halitosis and frequently present with fever, malaise, and lymphadenopathy. Treatment consists of debridement and oral administration of penicillin plus metronidazole, with clindamycin alone as an alternative.

Ludwig's angina is a rapidly progressive, potentially fulminant form of cellulitis that involves the bilateral sublingual andsubmandibular spaces and that typically originates from an infected or recently extracted tooth, most commonly the lower second and third molars. Improved dental care has reduced the incidence of this disorder substantially. Infection in these areas leads to dysphagia, odynophagia,

and "woody" edema in the sublingual region, forcing the tongue up and back with the potential for airway obstruction. Fever, dysarthria, and drooling also may be noted, and patients may speak in a "hot potato" voice. Intubation or tracheostomy may be necessary to secure the airway, as asphyxiation is the most common cause of death. Patients should be monitored closely and treated promptly with IV antibiotics directed against streptococci and oral anaerobes. Recommended agents include ampicillin/sulbactam and high-dose penicillin plus metronidazole.

Postanginal septicemia (*Lemierre's disease*) is a rare anaerobic oropharyngeal infection caused predominantly by *F. necrophorum*. The illness typically starts as a sore throat (most commonly in adolescents and young adults), which may present as exudative tonsillitis or peritonsillar abscess. Infection of the deep pharyngeal tissue allows organisms to drain into the lateral pharyngeal space, which contains the carotid artery and internal jugular vein. Septic thrombophlebitis of the internal jugular vein can result, with associated pain, dysphagia, and neck swelling and stiffness. Sepsis usually occurs 3–10 days after the onset of sore throat and is often coupled with metastatic infection to the lung and other distant sites. Occasionally, the infection can extend along the carotid sheath and into the posterior mediastinum, resulting in mediastinitis, or it can erode into the carotid artery, with the early sign of repeated small bleeds into the mouth. The mortality rate from these invasive infections can be as high as 50%. Treatment consists of IV antibiotics (penicillin G or clindamycin) and surgical drainage of any purulent collections. The concomitant use of anticoagulants to prevent embolization remains controversial but is often advised, with careful consideration of both the risks and the benefits.

INFECTIONS OF THE LARYNX AND EPIGLOTTIS

LARYNGITIS

Laryngitis is defined as any inflammatory process involving the larynx and can be caused by a variety of infectious and noninfectious processes. The vast majority of laryngitis cases seen in clinical practice in developed countries are acute. Acute laryngitis is a common syndrome caused predominantly by the same viruses responsible for many other URIs. In fact, most cases of acute laryngitis occur in the setting of a viral URI.

ETIOLOGY

Nearly all major respiratory viruses have been implicated in acute viral laryngitis, including rhinovirus, influenza virus, parainfluenza virus, adenovirus, coxsackievirus, coronavirus, and RSV. Acute laryngitis can also be associated with acute bacterial respiratory infections such as those caused by group A *Streptococcus* or *C. diphtheriae* (although diphtheria has been virtually eliminated in the United States). Another bacterial pathogen thought to play a role (albeit unclear) in the pathogenesis of acute laryngitis is *M. catarrhalis*, which has been recovered on nasopharyngeal culture in a significant percentage of cases.

Chronic laryngitis of infectious etiology is much less common in developed than in developing countries. Laryngitis due to *Mycobacterium tuberculosis* is often difficult to distinguish from laryngeal cancer, in part because of the frequent absence of signs, symptoms, and radiographic findings typical of pulmonary disease. *Histoplasma* and *Blastomyces* may cause laryngitis, often as a complication of systemic infection. *Candida* species can cause laryngitis as well, often in association with thrush or esophagitis and particularly in immunosuppressed patients. Rare cases of chronic laryngitis are due to *Coccidioides* and *Cryptococcus*.

CLINICAL MANIFESTATIONS

Laryngitis is characterized by hoarseness and also can be associated with reduced vocal pitch or aphonia. As acute laryngitis is caused predominantly by respiratory viruses, these symptoms usually occur in association with other symptoms and signs of URI, including rhinorrhea, nasal congestion, cough, and sore throat. Direct laryngoscopy often reveals diffuse laryngeal erythema and edema, along with vascular engorgement of the vocal folds. In addition, chronic disease (e.g., tuberculous laryngitis) often includes mucosal nodules and ulcerations visible on laryngoscopy; these lesions sometimes are mistaken for laryngeal cancer.

Treatment: Laryngitis

Acute laryngitis usually is treated with humidification and voice rest alone. Antibiotics are not recommended except when group A *Streptococcus* is cultured, in which case penicillin is the drug of choice. The choice of therapy for chronic laryngitis depends on the pathogen, whose identification usually requires biopsy with culture. Patients with laryngeal tuberculosis are highly contagious because of the large number of organisms that are easily aerosolized. These patients should be managed in the same way as patients with active pulmonary disease.

Croup

The term *croup* actually denotes a group of diseases collectively referred to as "croup syndrome," all of which are acute and predominantly viral respiratory illnesses characterized by marked swelling of the subglottic region of the larynx. Croup primarily affects children <6 years old. For a detailed discussion of this entity, the reader should consult a textbook of pediatric medicine.

Epiglottitis

Acute epiglottitis (supraglottitis) is an acute, rapidly progressive form of cellulitis of the epiglottis and adjacent structures that can result in complete—and potentially fatal—airway obstruction in both children and adults. Before the widespread use of *H. influenzae* type b (Hib) vaccine, this entity was much more common among children, with a peak incidence at ~3.5 years of age. In some countries, mass vaccination against Hib has reduced the annual incidence of acute epiglottitis in children by >90%; in contrast, the annual incidence in adults has changed little since the introduction of Hib vaccine. Because of the danger of airway obstruction, acute epiglottitis constitutes a medical emergency, particularly in children, and prompt diagnosis and airway protection are of the utmost importance.

ETIOLOGY

After the introduction of the Hib vaccine in the mid-1980s, disease incidence among children in the United States declined dramatically. Nevertheless, lack of vaccination or vaccine failure has meant that many pediatric cases seen today are still due to Hib. In adults and (more recently) in children, a variety of other bacterial pathogens have been associated with epiglottitis, the most common being group A *Streptococcus*. Other pathogens seen less frequently include *S. pneumoniae*, *Haemophilus parainfluenzae*, and *S. aureus* (including MRSA). Viruses have not been established as causes of acute epiglottitis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Epiglottitis typically presents more acutely in young children than in adolescents or adults. On presentation, most children have had symptoms for <24 h, including high fever, severe sore throat, tachycardia, systemic toxicity, and (in many cases) drooling while sitting forward. Symptoms and signs of respiratory obstruction may also be present and may progress rapidly. The somewhat milder illness in adolescents and adults often follows 1–2 days of severe sore throat and is commonly accompanied by dyspnea, drooling, and stridor. Physical examination of patients with acute epiglottitis may reveal moderate or severe respiratory distress, with inspiratory stridor and retractions of the chest wall. These findings *diminish* as the disease progresses and the patient tires. Conversely, oropharyngeal examination reveals infection that is much less severe than would be predicted from the symptoms—a finding that should alert the clinician to a cause of symptoms and obstruction that lies beyond the tonsils. The diagnosis often is made on clinical grounds, although direct fiberoptic laryngoscopy is frequently performed in a controlled environment (e.g., an operating room) to visualize and culture the typical edematous "cherry-red" epiglottis and facilitate placement of an endotracheal tube. Direct visualization in an examination room (i.e., with a tongue blade and indirect laryngoscopy) is not recommended because of the risk of immediate laryngospasm and complete airway obstruction. Lateral neck radiographs and laboratory tests can assist in the diagnosis but may delay the critical securing of the airway and cause the patient to be moved or repositioned more than is necessary, thereby increasing the risk of further airway compromise. Neck radiographs typically reveal an enlarged edematous epiglottis (the "thumbprint sign," **Fig. 31-2**), usually with a dilated hypopharynx and normal subglottic structures. Laboratory tests characteristically document mild to moderate leukocytosis with a predominance of neutrophils. Blood cultures are positive in a significant proportion of cases.

Figure 31-2

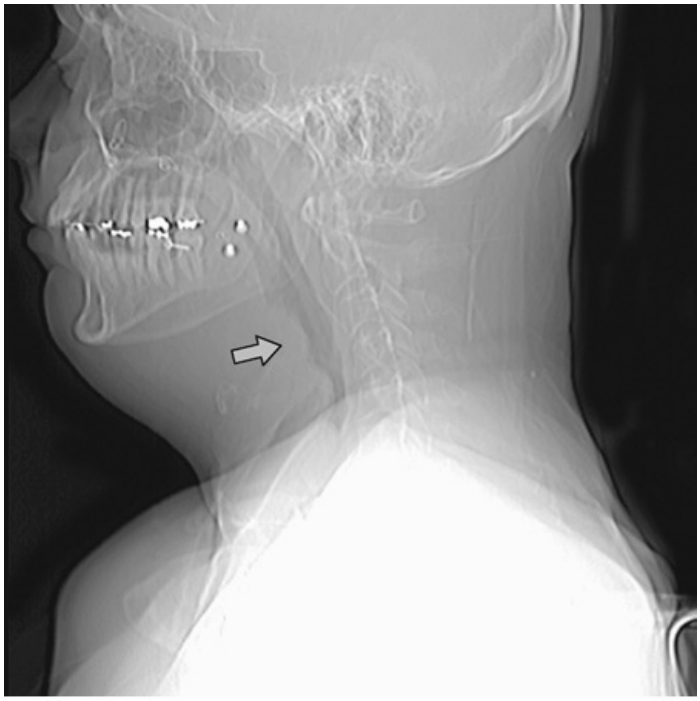

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute epiglottitis. In this lateral soft tissue radiograph of the neck, the arrow indicates the enlarged edematous epiglottis (the "thumbprint sign").

Treatment: Epiglottitis

Security of the airway is always of primary concern in acute epiglottitis, even if the diagnosis is only suspected. Mere observation for signs of impending airway obstruction is not routinely recommended, particularly in children. Many adults have been managed with observation only since the illness is perceived to be milder in this age group, but some data suggest that this approach may be risky and probably should be reserved only for adult patients who have yet to develop dyspnea or stridor. Once the airway has been secured and specimens of blood and epiglottis tissue have been obtained for culture, treatment with IV antibiotics should be given to cover the most likely organisms, particularly *H. influenzae*. Because rates of ampicillin resistance in this organism have risen significantly in recent years, therapy with a β-lactam/β-lactamase inhibitor combination or a second- or third-generation cephalosporin is recommended. Typically, ampicillin/sulbactam, cefuroxime, cefotaxime, or ceftriaxone is given, with clindamycin and TMP-SMX reserved for patients allergic to β-lactams. Antibiotic therapy should be continued for 7–10 days and should be tailored to the organism recovered in culture. If the household contacts of a patient with *H. influenzae* epiglottitis include an unvaccinated child under age 4, all members of the household (including the patient) should receive prophylactic rifampin for 4 days to eradicate carriage of *H. influenzae*.

INFECTIONS OF THE DEEP NECK STRUCTURES

Deep neck infections are usually extensions of infection from other primary sites, most often within the pharynx or oral cavity. Many of these infections are life threatening but are difficult to detect at early stages, when they may be more easily managed. Three of the most clinically relevant spaces in the neck are the submandibular (and sublingual) space, the lateral pharyngeal (or parapharyngeal) space, and the retropharyngeal space. These spaces communicate with one another and with other important structures in the head, neck, and thorax, providing pathogens with easy access to areas that include the mediastinum, carotid sheath, skull base, and meninges. Once infection reaches these sensitive areas, mortality rates can be as high as 20–50%.

Infection of the submandibular and/or sublingual space typically originates from an infected or recently extracted lower tooth. The result is the severe, life-threatening infection referred to as Ludwig's angina (see "Oral Infections," above). Infection of the lateral pharyngeal (or parapharyngeal) space is most often a complication of common infections of the oral cavity and upper respiratory tract, including tonsillitis, peritonsillar abscess, pharyngitis, mastoiditis, and periodontal infection. This space, situated deep in the lateral wall of the pharynx, contains a number of sensitive structures, including the carotid artery, internal jugular vein, cervical sympathetic chain, and portions of cranial nerves IX through XII; at its distal end, it opens into the posterior mediastinum. Involvement of this space with infection can therefore be rapidly fatal. Examination may reveal some tonsillar displacement, trismus, and neck rigidity, but swelling of the lateral pharyngeal wall can easily be missed. The diagnosis can be

confirmed by CT. Treatment consists of airway management, operative drainage of fluid collections, and at least 10 days of IV therapy with an antibiotic active against streptococci and oral anaerobes (e.g., ampicillin/sulbactam). A particularly severe form of this infection involving the components of the carotid sheath (postanginal septicemia, Lemierre's disease) is described above (see "Oral Infections"). Infection of the retropharyngeal space can also be extremely dangerous, as this space runs posterior to the pharynx from the skull base to the superior mediastinum. Infections in this space are more common among children <5 years old because of the presence of several small retropharyngeal lymph nodes that typically atrophy by age 4 years. Infection is usually a consequence of extension from another site of infection, most commonly, acute pharyngitis. Other sources include otitis media, tonsillitis, dental infections, Ludwig's angina, and anterior extension of vertebral osteomyelitis. Retropharyngeal space infection also can follow penetrating trauma to the posterior pharynx (e.g., from an endoscopic procedure). Infections are commonly polymicrobial, involving a mixture of aerobes and anaerobes; group Aß-hemolytic streptococci and *S. aureus* are the most common pathogens. *M. tuberculosis* was a common cause in the past but now is rarely involved in the United States.

Patients with retropharyngeal abscess typically present with sore throat, fever, dysphagia, and neck pain and are often drooling because of difficulty and pain with swallowing. Examination may reveal tender cervical adenopathy, neck swelling, and diffuse erythema and edema of the posterior pharynx as well as a bulge in the posterior pharyngeal wall that may not be obvious on routine inspection. A soft tissue mass is usually demonstrable by lateral neck radiography or CT. Because of the risk of airway obstruction, treatment begins with securing of the airway, followed by a combination of surgical drainage and IV antibiotic administration. Initial empirical therapy should cover streptococci, oral anaerobes, and *S. aureus*; ampicillin/sulbactam, clindamycin alone, or clindamycin plus ceftriaxone is usually effective. Complications result primarily from extension to other areas (e.g., rupture into the posterior pharynx may lead to aspiration pneumonia and empyema). Extension may also occur to the lateral pharyngeal space and mediastinum, resulting in mediastinitis and pericarditis, or into nearby major blood vessels. All these events are associated with a high mortality rate.

FURTHER READINGS

American Academy of Pediatrics Subcommittee on Management of Acute Otitis Media: Diagnosis and management of acute otitis media. Pediatrics 113:1451, 2004

American Academy of Pediatrics Subcommittee on Management of Sinusitis and Committee on Quality Improvement: Clinical practice guideline: Management of sinusitis. Pediatrics 108:798, 2001

Centor RM: Expand the pharyngitis paradigm for adolescents and young adults. Ann Intern Med 151:812, 2009 [PMID: 19949147]

Cooper RJ et al: Principles of appropriate antibiotic use for acute pharyngitis in adults: Background. Ann Intern Med 134:509, 2001 [PMID: 11255530]

Gonzales R et al: Principles of appropriate antibiotic use for treatment of nonspecific upper respiratory tract infections in adults: Background. Ann Intern Med 134:490, 2001 [PMID: 11255526]

Grijalva CG et al: Antibiotic prescription rates for acute respiratory tract infections in US ambulatory settings. JAMA 302:758, 2009 [PMID: 19690308]

Meropol SB: Valuing reduced antibiotic use for pediatric acute otitis media. Pediatrics 122:669, 2008

Rafei K et al: Airway infectious disease emergencies. Pediatr Clin North Am 53:215, 2006 [PMID: 16574523]

Rosenfeld RM et al: Clinical practice guideline: Adult sinusitis. Otolaryngol Head Neck Surg 137(3 Suppl):S1, 2007

Schwartz B et al: Pharyngitis—principles of judicious use of antimicrobial agents. Pediatrics 101:171, 1998

Thompson P et al: Effect of antibiotics for otitis media on mastoiditis in children: A retrospective cohort study using the United Kingdom general practice research database. Pediatrics 123:424, 2009 [PMID: 19171605]

Williamson IG et al: Antibiotics and topical nasal steroid for treatment of acute maxillary sinusitis. JAMA 298:2487, 2007 [PMID: 18056902]

Young J et al: Antibiotics for adults with clinically diagnosed acute rhinosinusitis: A meta-analysis of individual patient data. Lancet 371:908, 2008 [PMID: 18342685]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ORAL MANIFESTATIONS OF DISEASE: INTRODUCTION

As primary care physicians and consultants, internists are often asked to evaluate patients with disease of the oral soft tissues, teeth, and pharynx. Knowledge of the oral milieu and its unique structures is necessary to guide preventive services and recognize oral manifestations of local or systemic disease (**Chap. e12**). Furthermore, internists frequently collaborate with dentists in the care of patients who have a variety of medical conditions that affect oral health or who undergo dental procedures that increase their risk of medical complications.

DISEASES OF THE TEETH AND PERIODONTAL STRUCTURES

Tooth and Periodontal Structure

Tooth formation begins during the sixth week of embryonic life and continues through the first 17 years of age. Tooth development begins in utero and continues until after the tooth erupts. Normally, all 20 deciduous teeth have erupted by age 3 and have been shed by age 13. Permanent teeth, eventually totaling 32, begin to erupt by age 6 and have completely erupted by age 14, though third molars (wisdom teeth) may erupt later.

The erupted tooth consists of the visible crown covered with enamel and the root submerged below the gum line and covered with bonelike cementum. *Dentin*, a material that is denser than bone and exquisitely sensitive to pain, forms the majority of the tooth substance. Dentin surrounds a core of myxomatous *pulp* containing the vascular and nerve supply. The tooth is held firmly in the alveolar socket by the *periodontium*, supporting structures that consist of the gingivae, alveolar bone, cementum, and periodontal ligament. The periodontal ligament tenaciously binds the tooth's cementum to the alveolar bone. Above this ligament is a collar of attached gingiva just below the crown. A few millimeters of unattached or free gingiva (1–3 mm) overlap the base of the crown, forming a shallow sulcus along the gum-tooth margin.

DENTAL CARIES, PULPAL AND PERIAPICAL DISEASE, AND COMPLICATIONS

Dental caries begin asymptomatically as a destructive process of the hard surface of the tooth. *Streptococcus mutans*, principally, along with other bacteria colonize the organic buffering film on the tooth surface to produce *plaque*. If not removed by brushing or the natural cleaning action of saliva and oral soft tissues, bacterial acids demineralize the enamel. Fissures and pits on the occlusion surfaces are the most frequent sites of decay. Surfaces adjacent to tooth restorations and exposed roots are also vulnerable, particularly as teeth are retained in an aging population. Over time, dental caries extend to the underlying dentin, leading to cavitation of the enamel and, ultimately, penetration to the tooth pulp, producing *acute pulpitis*. At this early stage, when the pulp infection is limited, the tooth becomes sensitive to percussion and hot or cold, and pain resolves immediately when the irritating stimulus is removed. Should the infection spread throughout the pulp, *irreversible pulpitis* occurs, leading to pulp necrosis. At this late stage, pain is severe and has a sharp or throbbing visceral quality that may be worse when the patient lies down. Once pulp necrosis is complete, pain may be constant or intermittent, but cold sensitivity is lost.

Treatment of caries involves removal of the softened and infected hard tissue; sealing the exposed dentin; and restoration of the tooth structure with silver amalgam, composite resin, gold, or porcelain. Once irreversible pulpitis occurs, root canal therapy is necessary, and the contents of the pulp chamber and root canals are removed, followed by thorough cleaning, antisepsis, and filling with an inert material. Alternatively, the tooth may be extracted.

Pulpal infection, if it does not egress through the decayed enamel, leads to *periapical abscess* formation, which produces pain on chewing. If the infection is mild and chronic, a *periapical granuloma* or eventually a *periapical cyst* forms, either of which produces radiolucency at the root apex. When unchecked, a periapical abscess can erode into the alveolar bone producing osteomyelitis, penetrate and drain through the gingivae (parulis or gumboil), or track along deep fascial planes, producing a virulent cellulitis (Ludwig's angina) involving the submandibular space and floor of the mouth (Chap. 164). Elderly patients, those with diabetes mellitus, and patients taking glucocorticoids may experience little or no pain and fever as these complications develop.

PERIODONTAL DISEASE

Periodontal disease accounts for more tooth loss than caries, particularly in the elderly. Like dental

caries, chronic infection of the gingiva and anchoring structures of the tooth begins with formation of bacterial plaque. The process begins invisibly above the gum line and in the gingival sulcus. Plaque, including mineralized plaque (calculus), is preventable by appropriate dental hygiene, including periodic professional cleaning. Left undisturbed, chronic inflammation ensues and produces a painless hyperemia of the free and attached gingivae (*gingivitis*) that typically bleeds with brushing. If ignored, severe *periodontitis* occurs, leading to deepening of the physiologic sulcus and destruction of the periodontal ligament. Pockets develop around the teeth and become filled with pus and debris. As the periodontium is destroyed, teeth loosen and exfoliate. Eventually, there is resorption of the alveolar bone. A role for the chronic inflammation resulting from chronic periodontal disease in promoting coronary heart disease and stroke has been proposed. Epidemiologic studies demonstrate a moderate but significant association between chronic periodontal inflammation and atherogenesis, though a causal role remains unproven.

Acute and aggressive forms of periodontal disease are less common than the chronic forms described above. However, if the host is stressed or exposed to a new pathogen, rapidly progressive and destructive disease of the periodontal tissue can occur. A virulent example is *acute necrotizing ulcerative gingivitis* (ANUG) or *Vincent's infection*. Stress, poor oral hygiene, and tobacco and alcohol use are risk factors. The presentation includes sudden gingival inflammation, ulceration, bleeding, interdental gingival necrosis, and fetid halitosis. *Localized juvenile periodontitis*, seen in adolescents, is particularly destructive and appears to be associated with impaired neutrophil chemotaxis. *AIDS-related periodontitis* resembles ANUG in some patients or a more destructive form of adult chronic periodontitis in others. It may also produce a gangrene-like destructive process of the oral soft tissues and bone that resembles *noma*, seen in severely malnourished children in developing nations.

PREVENTION OF TOOTH DECAY AND PERIODONTAL INFECTION

Despite the reduced prevalence of dental caries and periodontal disease in the United States due in large part to water fluoridation and improved dental care, respectively, both diseases constitute a major public health problem worldwide and for certain groups. The internist should promote preventive dental care and hygiene as part of health maintenance. Special populations at high risk for dental caries and periodontal disease include those with xerostomia, diabetics, alcoholics, tobacco users, those with Down's syndrome, and those with gingival hyperplasia. Furthermore, patients lacking dental-care access (low socioeconomic status) and those with reduced ability to provide self-care (e.g., nursing home residents and those with dementia or upper-extremity disability) suffer at a disproportionate rate. It is important to provide counseling regarding regular dental hygiene and professional cleaning, use of fluoride-containing toothpaste, professional fluoride treatments, and use of electric toothbrushes for patients with limited dexterity and to give instruction to caregivers for those unable to perform self-care. Internists caring for international students studying in the United States should be aware of the high prevalence of dental decay in this population. Cost, fear of dental care, and language and cultural differences may create barriers that prevent some from seeking preventive dental services.

DEVELOPMENTAL AND SYSTEMIC DISEASE AFFECTING THE TEETHAND PERIODONTIUM

Malocclusion is the most common developmental problem, which, in addition to a problem with cosmesis, can interfere with mastication unless corrected through orthodontic techniques. Impacted third molars are common and occasionally become infected. Acquired prognathism due to *acromegaly* may also lead to malocclusion, as may deformity of the maxilla and mandible due to *Paget's disease* of the bone. Delayed tooth eruption, receding chin, and a protruding tongue are occasional features of *cretinism* and *hypopituitarism*. Congenital syphilis produces tapering, notched (Hutchinson's) incisors and finely nodular (mulberry) molar crowns.

Enamel hypoplasia results in crown defects ranging from pits to deep fissures of primary or permanent teeth. Intrauterine infection (syphilis, rubella), vitamin deficiency (A, C, or D), disorders of calcium metabolism (malabsorption, vitamin D–resistant rickets, hypoparathyroidism), prematurity, high fever, or rare inherited defects (*amelogenesis imperfecta*) are all causes. Tetracycline, given in sufficiently high doses during the first eight years, may produce enamel hypoplasia and discoloration. Exposure to endogenous pigments can discolor developing teeth: *erythroblastosis fetalis* (green or bluish-black), congenital liver disease (green or yellow-brown), and porphyria (red or brown that fluoresces with ultraviolet light). *Mottled enamel* occurs if excessive fluoride is ingested during development. Worn enamel is seen with age, bruxism, or excessive acid exposure (e.g., chronic gastric reflux or bulimia).

Premature tooth loss resulting from periodontitis is seen with cyclic neutropenia, Papillon-Lefèvre syndrome, Chédiak-Higashi syndrome, and leukemia. Rapid focal tooth loosening is most often due to infection, but rarer causes include Langerhans cell histiocytosis, Ewing's sarcoma, osteosarcoma, or Burkitt's lymphoma. Early loss of primary teeth is a feature of *hypophosphatasia*, a rare inborn error of metabolism.

Pregnancy may produce severe gingivitis and localized *pyogenic granulomas*. Severe periodontal disease occurs with Down's syndrome and diabetes mellitus. *Gingival hyperplasia* may be caused by phenytoin, calcium channel blockers (e.g., nifedipine), and cyclosporine. *Idiopathic familial gingival fibromatosis* and several syndrome-related disorders appear similar. Removal of the medication often

reverses the drug-induced form, though surgery may be needed to control both. *Linear gingival erythema* is variably seen in patients with advanced HIV infection and probably represents immune deficiency and decreased neutrophil activity. Diffuse or focal gingival swelling may be a feature of early or late acute myelomonocytic leukemia (AMML) as well as of other lymphoproliferative disorders. A rare, but pathognomonic, sign of Wegener's granulomatosis is a red-purplish, granular gingivitis (strawberry gums).

DISEASES OF THE ORAL MUCOSA

INFECTION

Most oral mucosal diseases involve microorganisms **(Table 32-1)**.

Table 32-1 Vesicular, Bullous, or Ulcerative Lesions of the Oral Mucosa

Condition	Usual Location	Clinical Features	Course
Viral Diseases			
Primary acute herpetic gingivostomatitis [herpes simplex virus (HSV) type 1, rarely type 2]	Lip and oral mucosa (buccal, gingival, lingual mucosa)	Labial vesicles that rupture and crust, and intraoral vesicles that quickly ulcerate; extremely painful; acute gingivitis, fever, malaise, foul odor, and cervical lymphadenopathy; occurs primarily in infants, children, and young adults	Heals spontaneously in 10 –14 days. Unless secondarily infected, lesions lasting >3 weeks are not due to primary HSV infection
Recurrent herpes labialis	Mucocutaneous junction of lip, perioral skin	Eruption of groups of vesicles that may coalesce, then rupture and crust; painful to pressure or spicy foods	Lasts about 1 week, but condition may be prolonged if secondarily infected. If severe, topical or oral antiviral may reduce healing time
Recurrent intraoral herpes simplex	Palate and gingiva	Small vesicles on keratinizedepithelium that rupture and coalesce; painful	Heals spontaneously in about 1 week. If severe, topical or oral antiviral may reduce healing time.
Chickenpox (varicella-zoster virus)	Gingiva and oral mucosa	Skin lesions may be accompanied by small vesicles on oral mucosa that rupture to form shallow ulcers; may coalesce to form large bullous lesions that ulcerate; mucosa may have generalized erythema	Lesions heal spontaneously within 2 weeks
Herpes zoster (reactivation of varicella-zoster virus)	Cheek, tongue, gingiva, or palate	Unilateral vesicular eruptions and ulceration in linear pattern following sensory distribution of trigeminal nerve or one of its branches	Gradual healing without scarring unless secondarily infected; postherpetic neuralgia is common. Oral acyclovir, famciclovir, or valacyclovir reduce healing time and postherpetic neuralgia
Infectious mononucleosis (Epstein-Barr virus)	Oral mucosa	Fatigue, sore throat, malaise, fever, and cervical lymphadenopathy; numerous small ulcers usually appear several days before lymphadenopathy; gingival bleeding and multiple petechiae at junction of hard and soft palates	Oral lesions disappear during convalescence; no treatment though glucocorticoids indicated if tonsillar swelling compromises airway
Herpangina (coxsackievirus A; also possibly coxsackie B and echovirus)	Oral mucosa, pharynx, tongue	Sudden onset of fever, sore throat, and oropharyngeal vesicles, usually in children under 4 years, during summer months; diffuse pharyngeal congestion and vesicles (1–2 mm), grayish-white surrounded by red areola; vesicles enlarge and ulcerate	Incubation period 2–9 days; fever for 1–4 days; recovery uneventful
Hand, foot, and mouth disease (coxsackievirus A16 most common)	Oral mucosa, pharynx, palms, and soles	Fever, malaise, headache with oropharyngeal vesicles that become painful, shallow ulcers; highly infectious; usually affects children under age 10	Incubation period 2–18 days; lesions heal spontaneously in 2–4 weeks
Primary HIV	Gingiva, palate,	Acute gingivitis and oropharyngeal	Followed by HIV

infection	and pharynx	ulceration, associated with febrile illness resembling mononucleosis and including lymphadenopathy	seroconversion, asymptomatic HIV infection, and usually ultimately by HIV disease
Bacterial or Fungal Diseases			
Acute necrotizing ulcerative gingivitis ("trench mouth," Vincent's infection)	Gingiva	Painful, bleeding gingiva characterized by necrosis and ulceration of gingival papillae and margins plus lymphadenopathy and foul odor	Debridement and diluted (1:3) peroxide lavage provide relief within 24 h; antibiotics in acutely ill patients; relapse may occur
Prenatal (congenital) syphilis	Palate, jaws, tongue, and teeth	Gummatous involvement of palate, jaws, and facial bones; Hutchinson's incisors, mulberry molars, glossitis, mucous patches, and fissures on corner of mouth	Tooth deformities in permanent dentition irreversible
Primary syphilis (chancre)	Lesion appears where organism enters body; may occur on lips, tongue, or tonsillar area	Small papule developing rapidly into a large, painless ulcer with indurated border; unilateral lymphadenopathy; chancre and lymph nodes containing spirochetes; serologic tests positive by third to fourth weeks	Healing of chancre in 1–2 months, followed by secondary syphilis in 6–8 weeks
Secondary syphilis	Oral mucosa frequently involved with mucous patches, primarily on palate, also at commissures of mouth	Maculopapular lesions of oral mucosa, 5–10 mm in diameter with central ulceration covered by grayish membrane; eruptions occurring on various mucosal surfaces and skin accompanied by fever, malaise, and sore throat	Lesions may persist from several weeks to a year
Tertiary syphilis	Palate and tongue	Gummatous infiltration of palate or tongue followed by ulceration and fibrosis; atrophy of tongue papillae produces characteristic bald tongue and glossitis	Gumma may destroy palate, causing complete perforation
Gonorrhea	Lesions may occur in mouth at site of inoculation or secondarily by hematogenous spread from a primary focus elsewhere	Most pharyngeal infection is asymptomatic; may produce burning or itching sensation; oropharynx and tonsils may be ulcerated and erythematous; saliva viscous and fetid	More difficult to eradicate than urogenital infection, though pharyngitis usually resolves with appropriate antimicrobial treatment
Tuberculosis	Tongue, tonsillar area, soft palate	A painless, solitary, 1–5 cm, irregular ulcer covered with a persistent exudate; ulcer has a firm undermined border	Autoinoculation from pulmonary infection usual; lesions resolve with appropriate antimicrobial therapy
Cervicofacial actinomycosis	Swellings in region of face, neck, and floor of mouth	Infection may be associated with an extraction, jaw fracture, or eruption of molar tooth; in acute form resembles an acute pyogenic abscess, but contains yellow "sulfur granules" (gram-positive mycelia and their hyphae)	Typically, swelling is hard and grows painlessly; multiple abscesses with draining tracts develop; penicillin first choice; surgery usually necessary
Histoplasmosis	Any area of the mouth, particularly tongue, gingiva, or palate	Nodular, verrucous, or granulomatous lesions; ulcers are indurated and painful; usual source hematogenous or pulmonary, but may be primary	Systemic antifungal therapy necessary to treat
Candidiasis (Table 32-3)			
Dermatologic Diseases			
Mucous membrane pemphigoid	Typically produces marked gingival erythema and ulceration; other areas of	Painful, grayish-white collapsed vesicles or bullae of full-thickness epithelium with peripheral erythematous zone; gingival lesions desquamate, leaving	Protracted course with remissions and exacerbations; involvement of different sites occurs slowly; glucocorticoids may

	oral cavity, esophagus, and vagina may be affected	ulcerated area	temporarily reduce symptoms but do not control the disease
Erythema multiforme minor and major (Stevens-Johnson syndrome)	Primarily the oral mucosa and the skin of hands and feet	Intraoral ruptured bullae surrounded by an inflammatory area; lips may show hemorrhagic crusts; the "iris," or "target," lesion on the skin is pathognomonic; patient may have severe signs of toxicity	Onset very rapid; usually idiopathic, but may be associated with trigger such as drug reaction; condition may last 3–6 weeks; mortality with EM major 5–15% if untreated
Pemphigus vulgaris	Oral mucosa and skin; sites of mechanical trauma (soft/hard palate, frenulum, lips, buccal mucosa)	Usually (>70%) presents with oral lesions; fragile, ruptured bullae and ulcerated oral areas; mostly in older adults	With repeated occurrence of bullae, toxicity may lead to cachexia, infection, and death within 2 years; often controllable with oral glucocorticoids
Lichen planus	Oral mucosa and skin	White striae in mouth; purplish nodules on skin at sites of friction; occasionally causes oral mucosal ulcers and erosive gingivitis	White striae alone usually asymptomatic; erosive lesions often difficult to treat, but may respond to glucocorticoids
Other Conditions			
Recurrent aphthous ulcers	Usually on nonkeratinized oral mucosa (buccal and labial mucosa, floor of mouth, soft palate, lateral and ventral tongue)	Single or clusters of painful ulcers with surrounding erythematous border; lesions may be 1–2 mm in diameter in crops (herpetiform), 1–5 mm (minor), or 5–15 mm (major)	Lesions heal in 1–2 weeks but may recur monthly or several times a year; protective barrier with orabase and topical steroids give symptomatic relief; systemic glucocorticoids may be needed in severe cases
Behçet's syndrome	Oral mucosa, eyes, genitalia, gut, and CNS	Multiple aphthous ulcers in mouth; inflammatory ocular changes, ulcerative lesions on genitalia; inflammatory bowel disease and CNS disease	Oral lesions often first manifestation; persist several weeks and heal without scarring
Traumatic ulcers	Anywhere on oral mucosa; dentures frequently responsible for ulcers in vestibule	Localized, discrete ulcerated lesions with red border; produced by accidental biting of mucosa, penetration by a foreign object, or chronic irritation by a denture	Lesions usually heal in 7–10 days whenirritant is removed, unless secondarily infected
Squamous cell carcinoma	Any area in the mouth, most commonly on lower lip, tongue, and floor of mouth	Ulcer with elevated, indurated border; failure to heal, pain not prominent; lesions tend to arise in areas of erythro/leukoplakia or in smooth atrophic tongue	Invades and destroys underlying tissues; frequently metastasizes to regional lymph nodes
Acute myeloid leukemia (usually monocytic)	Gingiva	Gingival swelling and superficial ulceration followed by hyperplasia of gingiva with extensive necrosis and hemorrhage; deep ulcers may occur elsewhere on the mucosa complicated by secondary infection	Usually responds to systemic treatmentof leukemia; occasionally requires local radiation therapy
Lymphoma	Gingiva, tongue, palate and tonsillar area	Elevated, ulcerated area that may proliferate rapidly, giving the appearance of traumatic inflammation	Fatal if untreated; may indicate underlying HIV infection
Chemical or thermal burns	Any area in mouth	White slough due to contact with corrosive agents (e.g., aspirin, hot cheese) applied locally; removal of slough leaves raw, painful surface	Lesion heals in several weeks if not secondarily infected

Note: CNS, central nervous system.

PIGMENTED LESIONS

See **Table 32-2**.

Table 32-2 Pigmented Lesions of the Oral Mucosa

Condition	Usual Location	Clinical Features	Course
Oral melanotic macule	Any area of the mouth	Discrete or diffuse localized, brown to black macule	Remains indefinitely; no growth
Diffuse melanin pigmentation	Any area of the mouth	Diffuse pale to dark-brown pigmentation; may be physiologic ("racial") or due to smoking	Remains indefinitely
Nevi	Any area of the mouth	Discrete, localized, brown to black pigmentation	Remains indefinitely
Malignant melanoma	Any area of the mouth	Can be flat and diffuse, painless, brown to black, or can be raised and nodular	Expands and invades early; metastasis leads to death
Addison's disease	Any area of the mouth, but mostly buccal mucosa	Blotches or spots of bluish-black to dark-brownpigmentation occurring early in the disease, accompanied by diffuse pigmentation of skin; other symptoms of adrenal insufficiency	Condition controlled by adrenal steroid replacement
Peutz-Jeghers syndrome	Any area of the mouth	Dark-brown spots on lips, buccal mucosa, with characteristic distribution of pigment around lips, nose, eyes, and on hands; concomitant intestinal polyposis	Oral pigmented lesions remain indefinitely; gastrointestinal polyps may become malignant
Drug ingestion (neuroleptics, oral contraceptives, minocycline, zidovudine, quinine derivatives)	Any area of the mouth	Brown, black, or gray areas of pigmentation	Gradually disappears following cessation of drug
Amalgam tattoo	Gingiva and alveolar mucosa	Small blue-black pigmented areas associated with embedded amalgam particles in soft tissues; these may show up on radiographs as radiopaque particles in some cases	Remains indefinitely
Heavy metal pigmentation (bismuth, mercury, lead)	Gingival margin	Thin blue-black pigmented line along gingivalmargin; rarely seen except for children exposedto lead-based paint	Indicative of systemic absorption; no significance for oral health
Black hairy tongue	Dorsum of tongue	Elongation of filiform papillae of tongue, which become stained by coffee, tea, tobacco, or pigmented bacteria	Improves within 1–2 weeks with gentle brushing of tongue or discontinuation of antibiotic if due to bacterial overgrowth
Fordyce "spots"	Buccal and labial mucosa	Numerous small yellowish spots just beneath mucosal surface; no symptoms; due to hyperplasia of sebaceous glands	Benign; remains without apparent change
Kaposi's sarcoma	Palate most common, but may occur in any other site	Red or blue plaques of variable size and shape; often enlarge, become nodular and may ulcerate	Usually indicative of HIV infection or non-Hodgkin's lymphoma; rarely fatal, but may require treatment for comfort or cosmesis
Mucous retention cysts	Buccal and labial mucosa	Bluish-clear fluid-filled cyst due to extravasated mucous from injured minor salivary gland	Benign; painless unless traumatized; may be removed surgically

DERMATOLOGIC DISEASES

See Tables 32-1, 32-2, and **32-3** and Chaps. 51, 52, 53, 54, and 55.

Table 32-3 White Lesions of Oral Mucosa

Condition	Usual Location	Clinical Features	Course
Lichen planus	Buccal mucosa, tongue, gingiva, and lips; skin	Striae, white plaques, red areas, ulcers in mouth;purplish papules on skin; may be asymptomatic, sore, or painful; lichenoid drug reactions may look similar	Protracted; responds to topicalglucocorticoids
White sponge nevus	Oral mucosa, vagina, anal mucosa	Painless white thickening of epithelium; adolescent/early adult onset; familial	Benign and permanent
Smoker's leukoplakia and smokeless tobacco lesions	Any area of oral mucosa, sometimes related to location of habit	White patch that may become firm, rough, orred-fissured and ulcerated; may become sore andpainful but usually painless	May or may not resolve with cessation of habit; 2% develop squamous cell carcinoma; early biopsy essential
Erythroplakia with or without white patches	Floor of mouth common in men; tongue and buccal mucosa in women	Velvety, reddish plaque; occasionally mixed with white patches or smooth red areas	High risk of squamous cell cancer; early biopsy essential
Candidiasis	Any area in mouth	*Pseudomembranous type* ("thrush"): creamy white curdlike patches that reveal a raw, bleeding surface when scraped; found in sick infants, debilitated elderly patients receiving high doses of glucocorticoids or broad-spectrum antibiotics, or in patients with AIDS	Responds favorably to antifungal therapy and correction of predisposing causes where possible
		Erythematous type: flat, red, sometimes sore areas in same groups of patients	Course same as for pseudomembranous type
		Candidal leukoplakia: nonremovable white thickening of epithelium due to *Candida*	Responds to prolonged antifungal therapy
		Angular cheilitis: sore fissures at corner of mouth	Responds to topical antifungal therapy
Hairy leukoplakia	Usually lateral tongue, rarely elsewhere on oral mucosa	White areas ranging from small and flat to extensive accentuation of vertical folds; found in HIV carriers in all risk groups for AIDS	Due to EBV; responds to high-dose acyclovir but recurs; rarely causes discomfort unless secondarily infected with *Candida*
Warts (papillomavirus)	Anywhere on skin and oral mucosa	Single or multiple papillary lesions, with thick, white keratinized surfaces containing many pointed projections; cauliflower lesions covered with normal-colored mucosa or multiple pink or pale bumps (focal epithelial hyperplasia)	Lesions grow rapidly and spread; consider squamous cell carcinoma and rule out with biopsy; excision or laser therapy; may regress in HIV infected patients on antiretroviral therapy

Note: EBV, Epstein-Barr virus.

DISEASES OF THE TONGUE

See **Table 32-4**.

Table 32-4 Alterations of the Tongue

Type of Change	Clinical Features
Size or Morphology Changes	
Macroglossia	Enlarged tongue that may be part of a syndrome found in developmental conditions

	such as Down syndrome, Simpson-Golabi-Behmel syndrome, or Beckwith-Wiedemann syndrome may be due to tumor (hemangioma or lymphangioma), metabolic disease (such as primary amyloidosis), or endocrine disturbance (such as acromegaly or cretinism)
Fissured ("scrotal") tongue	Dorsal surface and sides of tongue covered by painless shallow or deep fissures that may collect debris and become irritated
Median rhomboid glossitis	Congenital abnormality of tongue with ovoid, denuded area in median posterior portion of the tongue; may be associated with candidiasis and may respond to antifungals
Color Changes	
"Geographic" tongue (benign migratory glossitis)	Asymptomatic inflammatory condition of the tongue, with rapid loss and regrowth of filiform papillae, leading to appearance of denuded red patches "wandering" across the surface of the tongue
Hairy tongue	Elongation of filiform papillae of the medial dorsal surface area due to failure of keratin layer of the papillae to desquamate normally; brownish-black coloration may be due to staining by tobacco, food, or chromogenic organisms
"Strawberry" and "raspberry" tongue	Appearance of tongue during scarlet fever due to the hypertrophy of fungiform papillae plus changes in the filiform papillae
"Bald" tongue	Atrophy may be associated with xerostomia, pernicious anemia, iron-deficiency anemia, pellagra, or syphilis; may be accompanied by painful burning sensation; may be an expression of erythematous candidiasis and respond to antifungals

HIV DISEASE AND AIDS

See Tables 32-1, 32-2, 32-3, and **32-5**; Chap. 189; and Figures 181-1 and 203-1.

Table 32-5 Oral Lesions Associated with HIV Infection

Lesion Morphology	Etiologies
Papules, nodules, plaques	Candidiasis (hyperplastic and pseudomembranous)[a] Condyloma acuminatum (human papillomavirus infection) Squamous cell carcinoma (preinvasive and invasive) Non-Hodgkin's lymphoma[a] Hairy leukoplakia[a]
Ulcers	Recurrent aphthous ulcers[a] Angular cheilitis Squamous cell carcinoma Acute necrotizing ulcerative gingivitis[a] Necrotizing ulcerative periodontitis[a] Necrotizing ulcerative stomatitis Non-Hodgkin's lymphoma[a] Viral infection (herpes simplex, herpes zoster, cytomegalovirus) *Mycobacterium tuberculosis, Mycobacterium avium-intracellulare* Fungal infection (histoplasmosis, cryptococcosis, candidiasis, geotrichosis, aspergillosis) Bacterial infection (*Escherichia coli, Enterobacter cloacae, Klebsiellapneumoniae, Pseudomonas aeruginosa*) Drug reactions (single or multiple ulcers)
Pigmented lesions	Kaposi's sarcoma[a] Bacillary angiomatosis (skin and visceral lesions more common than oral)

	Zidovudine pigmentation (skin, nails, and occasionally oral mucosa) Addison's disease
Miscellaneous	Linear gingival erythema[a]

[a] Strongly associated with HIV infection.

ULCERS

Ulceration is the most common oral mucosal lesion. Although there are many causes, the host and pattern of lesions, including the presence of systemic features, narrow the differential diagnosis (Table 32-1). Most acute ulcers are painful and self-limited. Recurrent aphthous ulcers and herpes simplex infection constitute the majority. Persistent and deep aphthous ulcers can be idiopathic or seen with HIV/AIDS. Aphthous lesions are often the presenting symptom in *Behçet's syndrome* (Chap. 327). Similar-appearing, though less painful, lesions may occur with reactive arthritis (formerly known as Reiter's syndrome), and aphthous ulcers are occasionally present during phases of discoid or *systemic lupus erythematosus* (Chap. 323). Aphthous-like ulcers are seen in Crohn's disease (Chap. 295), but unlike the common aphthous variety, they may exhibit granulomatous inflammation histologically. Recurrent aphthae in some patients with *celiac disease* have been reported to remit with elimination of gluten.

Of major concern are chronic, relatively painless ulcers and mixed red/white patches (erythroplakia and leukoplakia) of more than two weeks' duration. Squamous cell carcinoma and premalignant dysplasia should be considered early and a diagnostic biopsy obtained. The importance is underscored because early-stage malignancy is vastly more treatable than late-stage disease. High-risk sites include the lower lip, floor of the mouth, ventral and lateral tongue, and soft palate–tonsillar pillar complex. Significant risk factors for oral cancer in Western countries include sun exposure (lower lip) and tobacco and alcohol use. In India and some other Asian countries, smokeless tobacco mixed with betel nut, slaked lime, and spices is a common cause of oral cancer. Less common etiologies include syphilis and Plummer-Vinson syndrome (iron deficiency).

Rarer causes of chronic oral ulcer such as tuberculosis, fungal infection, granulomatosis with polyangiitis (Wegener's), and midline granuloma may look identical to carcinoma. Making the correct diagnosis depends on recognizing other clinical features and biopsy of the lesion. The syphilitic chancre is typically painless and therefore easily missed. Regional lymphadenopathy is invariably present. Confirmation is achieved using appropriate bacterial and serologic tests.

Disorders of mucosal fragility often produce painful oral ulcers that fail to heal within two weeks. *Mucous membrane pemphigoid* and *pemphigus vulgaris* are the major acquired disorders. While clinical features are often distinctive, immunohistochemical examination should be performed for diagnosis and to distinguish these entities from *lichen planus* and drug reactions.

HEMATOLOGIC AND NUTRITIONAL DISEASE

Internists are more likely to encounter patients with acquired, rather than congenital, bleeding disorders. Bleeding after minor trauma should stop after 15 min and within an hour of tooth extraction if local pressure is applied. More prolonged bleeding, if not due to continued injury or rupture of a large vessel, should lead to investigation for a clotting abnormality. In addition to bleeding, petechiae and ecchymoses are prone to occur at the line of vibration between the soft and hard palates in patients with platelet dysfunction or thrombocytopenia.

All forms of leukemia, but particularly acute myelomonocyticleukemia, can produce gingival bleeding, ulcers, and gingival enlargement. Oral ulcers are a feature of agranulocytosis, and ulcers and mucositis are often severe complications of chemotherapy and radiation therapy for hematologic and other malignancies. Plummer-Vinson syndrome (iron deficiency, angular stomatitis, glossitis, and dysphagia) raises the risk of oral squamous cell cancer and esophageal cancer at the postcricoidal tissue web. Atrophic papillae and a red, burning tongue may occur with pernicious anemia. B-group vitamin deficiencies produce many of these same symptoms as well as oral ulceration and cheilosis. Swollen, bleeding gums, ulcers, and loosening of the teeth are a consequence of scurvy.

NONDENTAL CAUSES OF ORAL PAIN

Most, but not all, oral pain emanates from inflamed or injured tooth pulp or periodontal tissues. Nonodontogenic causes may be overlooked. In most instances, toothache is predictable and proportional to the stimulus applied, and an identifiable condition (e.g., caries, abscess) is found. Local anesthesia eliminates pain originating from dental or periodontal structures, but not referred pains. The most common nondental origin is myofascial pain referred from muscles of mastication, which become tender and ache with increased use. Many sufferers exhibit bruxism (the grinding of teeth, often during sleep) that is secondary to stress and anxiety. *Temporomandibular disorder* is closely related. It affects both sexes with a higher prevalence in women. Features include pain, limited mandibular

movement, and temporomandibular joint sounds. The etiologies are complex, and malocclusion does not play the primary role once attributed to it. *Osteoarthritis* is a common cause of masticatory pain. Anti-inflammatory medication, jaw rest, soft foods, and heat provide relief. The temporomandibular joint is involved in 50% of patients with *rheumatoid arthritis* and is usually a late feature of severe disease. Bilateral preauricular pain, particularly in the morning, limits range of motion.

Migrainous neuralgia may be localized to the mouth. Episodes of pain and remission without identifiable cause and absence of relief with local anesthesia are important clues. *Trigeminal neuralgia* (*tic douloureux*) may involve the entire branch or part of the mandibular or maxillary branches of the fifth cranial nerve and produce pain in one or a few teeth. Pain may occur spontaneously or may be triggered by touching the lip or gingiva, brushing the teeth, or chewing. *Glossopharyngeal neuralgia* produces similar acute neuropathic symptoms in the distribution of the ninth cranial nerve. Swallowing, sneezing, coughing, or pressure on the tragus of the ear triggers pain that is felt in the base of the tongue, pharynx, and soft palate and may be referred to the temporomandibular joint. *Neuritis* involving the maxillary and mandibular divisions of the trigeminal nerve (e.g., maxillary sinusitis, neuroma, and leukemic infiltrate) is distinguished from ordinary toothache by the neuropathic quality of the pain. Occasionally, *phantom pain* follows tooth extraction. Often the earliest symptom of Bell's palsy in the day or so before facial weakness develops is pain and hyperalgesia behind the ear and side of the face. Likewise, similar symptoms may precede visible lesions of herpes zoster infecting the seventh nerve (Ramsey-Hunt syndrome) or trigeminal nerve. *Postherpetic neuralgia* may follow either condition. *Coronary ischemia* may produce pain exclusively in the face and jaw and, like typical angina pectoris, is usually reproducible with increased myocardial demand. Aching in several upper molar or premolar teeth that is unrelieved by anesthetizing the teeth may point to *maxillary sinusitis*.

Giant cell arteritis is notorious for producing headache, but it may also produce facial pain or sore throat without headache. Jaw and tongue claudication with chewing or talking is relatively common. Tongue infarction is rare. Patients with subacute thyroiditis often experience pain referred to the face or jaw before the tender thyroid gland and transient hyperthyroidism are appreciated.

Burning mouth syndrome (glossodynia) is present in the absence of an identifiable cause (e.g., vitamin B12 deficiency, iron deficiency, diabetes mellitus, low-grade *Candida* infection, food sensitivity, or subtle xerostomia) and predominantly affects postmenopausal women. The etiology may be neuropathic. Clonazepam, alpha-lipoic acid, and cognitive behavioral therapy have benefited some. Some cases associated with ACE inhibitors have remitted when the drug was discontinued.

DISEASES OF THE SALIVARY GLANDS

Saliva is essential to oral health. Its absence leads to tooth decay and loss. Its major components, water and mucin, serve as a cleansing solvent and lubricating fluid. In addition, it contains antimicrobial factors (e.g., lysozyme, lactoperoxidase, secretory IgA), epidermal growth factor, minerals, and buffering systems. The major salivary glands secrete intermittently in response to autonomic stimulation, which is high during a meal but low otherwise. Hundreds of minor glands in the lips and cheeks secrete mucus continuously. Consequently, oral function becomes impaired when salivary function is reduced. Dry mouth (*xerostomia*) is perceived when salivary flow is reduced by 50%. The most common etiology is medication, especially drugs with anticholinergic properties, but also alpha and beta blockers, calcium channel blockers, and diuretics. Other causes include Sjögren's syndrome, chronic parotitis, salivary duct obstruction, diabetes mellitus, HIV/AIDS, and radiation therapy that includes the salivary glands in the field (Hodgkin's disease and head and neck cancer). Management involves eliminating or limiting drying medications, preventive dental care, and supplementing oral liquid. Sugarless mints or chewing gum may stimulate salivary secretion if dysfunction is mild. When sufficient exocrine tissue remains, pilocarpine or cevimeline has been shown to increase secretions. Commercial saliva substitutes or gels relieve dryness but must be supplemented with fluoride applications to prevent caries.

Sialolithiasis presents most often as painful swelling but in some instances as just swelling or pain. Conservative therapy consists of local heat, massage, and hydration. Promotion of salivary secretion with mints or lemon drops may flush out small stones. Antibiotic treatment is necessary when bacterial infection in suspected. In adults, *acute bacterial parotitis* is typically unilateral and most commonly affects postoperative, dehydrated, and debilitated patients. *Staphylococcus aureus* including methicillin-resistant forms and anaerobic bacteria are the most common pathogens. Chronic bacterial sialadenitis results from lowered salivary secretion and recurrent bacterial infection. When suspected bacterial infection is not responsive to therapy, the differential diagnosis should be expanded to include benign and malignantneoplasms, lymphoproliferative disorders, Sjögren's syndrome, sarcoidosis, tuberculosis, lymphadenitis, actinomycosis, and granulomatosis with polyangiitis (Wegener's). Bilateral nontender parotid enlargement occurs with diabetes mellitus, cirrhosis, bulimia, HIV/AIDS, and drugs (e.g., iodide, propylthiouracil).

Pleomorphic adenoma comprises two-thirds of all salivary neoplasms. The parotid is the principal salivary gland affected, and the tumor presents as a firm, slow-growing mass. Though benign,

recurrence is common if resection is incomplete. Malignant tumors such as mucoepidermoid carcinoma, adenoid cystic carcinoma, and adenocarcinoma tend to grow relatively fast, depending upon grade. They may ulcerate and invade nerves, producing numbness and facial paralysis. Surgical resection is the primary treatment. Radiation therapy (particularly neutron-beam therapy) is used when surgery is not feasible and it is used post-resection for certain histological types with a high risk of recurrence. Malignant salivary gland tumors have a 5-year survival rate of about 68%.

DENTAL CARE OF MEDICALLY COMPLEX PATIENTS

Routine dental care (e.g., extraction, scaling and cleaning, tooth restoration, and root canal) is remarkably safe. The most common concerns regarding care of dental patients with medical disease are fear of excessive bleeding for patients on anticoagulants, infection of the heart valves and prosthetic devices from hematogenous seeding of oral flora, and cardiovascular complications resulting from vasopressors used with local anesthetics during dental treatment. Experience confirms that the risks of any of these complications are very low.

Patients undergoing tooth extraction or alveolar and gingival surgery rarely experience uncontrolled bleeding when warfarin anticoagulation is maintained within the therapeutic range currently recommended for prevention of venous thrombosis, atrial fibrillation, or mechanical heart valve. Embolic complications and death, however, have been reported during subtherapeutic anticoagulation. Therapeutic anticoagulation should be confirmed before and continued through the procedure. Likewise, low-dose aspirin (e.g., 81–325 mg) can be safely continued. For patients on aspirin and another antiplatelet medication (e.g., clopidogrel), the decision to continue the second antiplatelet medication should be based on individual consideration of the risks of thrombosis and bleeding.

Patients at risk for bacterial endocarditis (Chap. 124) should maintain optimal oral hygiene, including flossing, and have regular professional cleaning. Currently, guidelines recommend that prophylactic antibiotics be restricted to those patients at high risk of bacterial endocarditis who undergo dental and oral procedures that involve significant manipulation of gingival or periapical tissue or penetration of the oral mucosa. If unexpected bleeding occurs, antibiotics given within 2 h following the procedure provide effective prophylaxis.

Hematogenous bacterial seeding from oral infection can undoubtedly produce late prosthetic joint infection and therefore requires removal of the infected tissue (e.g., drainage, extraction, root canal) and appropriate antibiotic therapy. However, evidence that late prosthetic joint infection occurs following routine dental procedures is lacking. For this reason, antibiotic prophylaxis is not recommended before dental surgery in patients with orthopedic pins, screws, and plates. It is, however, advised within the first2 years after joint replacement for patients who have inflammatory arthropathies, immunosuppression, type 1 diabetes mellitus, previous prosthetic joint infection, hemophilia, or malnourishment.

Concern often arises regarding the use of vasoconstrictors in patients with hypertension and heart disease. Vasoconstrictors enhance the depth and duration of local anesthesia, thus reducing the anesthetic dose and potential toxicity. If intravascular injection is avoided, 2% lidocaine with 1:100,000 epinephrine (limited to a total of 0.036 mg epinephrine) can be used safely in those with controlled hypertension and stable coronary heart disease, arrhythmia, or congestive heart failure. Precaution should be taken with patients taking tricyclic antidepressants and nonselective beta blockers because these drugs may potentiate the effect of epinephrine.

Elective dental treatments should be postponed for at least one month after myocardial infarction, after which the risk of reinfarction is low provided the patient is medically stable (e.g., stable rhythm, stable angina, and free of heart failure). Patients who have suffered a stroke should have elective dental care deferred for six months. In both situations, effective stress reduction requires good pain control, including the use of the minimal amount of vasoconstrictor necessary to provide good hemostasis and local anesthesia.

Bisphosphonate therapy is associated with *osteonecrosis* of the jaw. However, the risk with oral bisphosphonate therapy is very low. Most patients affected have received high-dose aminobisphosphonate therapy for multiple myeloma or metastatic breast cancer and have undergone tooth extraction or dental surgery. Intra-oral lesions appear as exposed yellow-white hard bone involving the mandible or maxilla. Two-thirds are painful. Screening tests for determining risk of osteonecrosis are unreliable. Patients slated for aminobisphosphonate therapy should receive preventive dental care that reduces the risk of infection and need for future dentoalveolar surgery.

HALITOSIS

Halitosis typically emanates from the oral cavity or nasal passages. Volatile sulfur compounds resulting from bacterial decay of food and cellular debris account for the malodor. Periodontal disease, caries, acute forms of gingivitis, poorly fitting dentures, oral abscess, and tongue coating are usual causes. Treatment includes correcting poor hygiene, treating infection, and tongue brushing. Xerostomia can produce and exacerbate halitosis. Pockets of decay in the tonsillar crypts, esophageal diverticulum, esophageal stasis (e.g., achalasia, stricture), sinusitis, and lung abscess account for some instances.

A few systemic diseases produce distinctive odors: renal failure (ammoniacal), hepatic (fishy), and ketoacidosis (fruity). *Helicobacter pylori* gastritis can also produce ammoniac breath. If no odor is detectable, then pseudohalitosis or even halitophobia must be considered. These conditions represent varying degrees of psychiatric illness.

AGING AND ORAL HEALTH

While tooth loss and dental disease are not normal consequences of aging, a complex array of structural and functional changes occurs with age that can affect oral health. Subtle changes in tooth structure (e.g., diminished pulp space and volume, sclerosis of dentinal tubules, and altered proportions of nerve and vascular pulp content) result in diminished or altered pain sensitivity, reduced reparative capacity, and increased tooth brittleness. In addition, age-associated fatty replacement of salivary acini may reduce physiologic reserve, thus increasing the risk of xerostomia.

Poor oral hygiene often results when vision fails or when patients lose manual dexterity and upper-extremity flexibility. This is particularly common for nursing home residents and must be emphasized because regular oral cleaning and dental care have been shown to reduce the incidence of pneumonia and mortality in this population. Other risks for dental decay include limited lifetime fluoride exposure and preference by some older adults for intensely sweet foods when taste and olfaction wane. These factors occur in an increasing proportion of persons over age 75 who retain teeth that have extensive restorations and exposed roots. Without assiduous care, decay can become quite advanced yet remain asymptomatic. Consequently, much or the entire tooth can be destroyed before the process is detected.

Periodontal disease, a leading cause of tooth loss, is indicated by loss of alveolar bone height. Over 90% of Americans have some degree of periodontal disease by age 50. Healthy adults who have not experienced significant alveolar bone loss by the sixth decade do not typically develop significant worsening with advancing age.

Complete edentulousness with advanced age, though less common than in previous decades, is still present in approximately 50% of Americans age ≥85. Speech, mastication, and facial contours are dramatically affected. Edentulousness may also worsen obstructive sleep apnea, particularly in those without symptoms while wearing dentures. Dentures can improve speech articulation and restore diminished facial contours. Mastication is restored less predictably, and those expecting dentures to improve oral intake are often disappointed. Dentures require periodic adjustment to accommodate inevitable remodeling that leads to a diminished volume of the alveolar ridge. Pain can result from friction or traumatic lesions produced by loose dentures. Poor fit and poor oral hygiene may permit candidiasis to develop. This may be asymptomatic or painful and is indicated by erythematous smooth or granular tissue conforming to an area covered by the appliance.

FURTHER READINGS

Durso SC: Interaction with other health team members in caring for elderly patients. Dent Clin N Am 49:377, 2005 [PMID: 15755411]

Edwards EJ et al: Updated recommendations for managing the care of patients receiving oral bisphosphonate therapy: an advisory statement from the American Dental Association Council on Scientific Affairs. J Am Dent Assoc 139: 1674, 2008 [PMID: 19047674]

Little JW. Periodontal disease and heart disease: Are they related? Gen Dent. 56:733, 2008 [PMID: 19014036]

——— et al (eds): *Dental Management of the Medically Compromised Patient* , 7th ed. St. Louis, Mosby, 2008

Logan RM: Links between oral and gastrointestinal health. Curr Opin Support Palliat Care 4:31, 2010 [PMID: 19952929]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF ORAL MANIFESTATIONS OF DISEASE: INTRODUCTION

The health status of the oral cavity is linked to cardiovascular disease, diabetes, and other systemic illnesses. Thus, examining the oral cavity for signs of disease is a key part of the physical exam. This chapter presents numerous outstanding clinical photographs illustrating many of the conditions discussed in Chap. 32, Oral Manifestations of Disease. Conditions affecting the teeth, periodontal tissues, and oral mucosa are all represented.

Figure e12-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gingival overgrowth secondary to calcium channel blocker use.

Figure e12-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oral lichen planus.

Figure e12-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erosive lichen planus.

Figure e12-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stevens-Johnson syndrome—reaction to nevirapine.

Figure e12-5

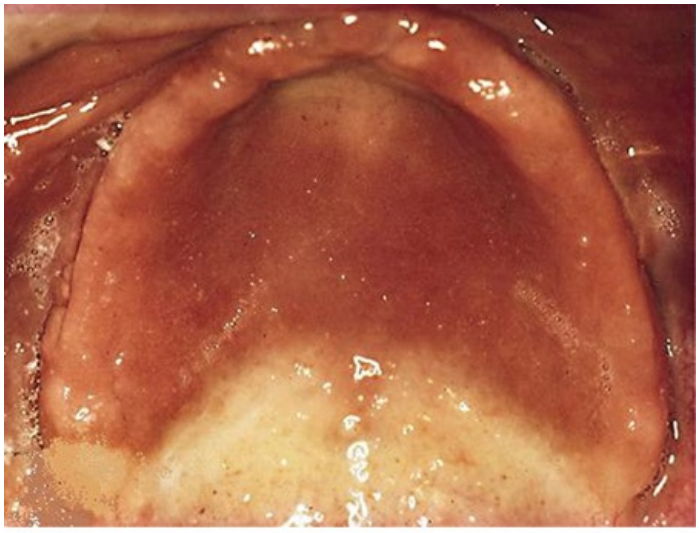

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Inflamed palate.

Figure e12-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe periodontitis.

Figure e12-7

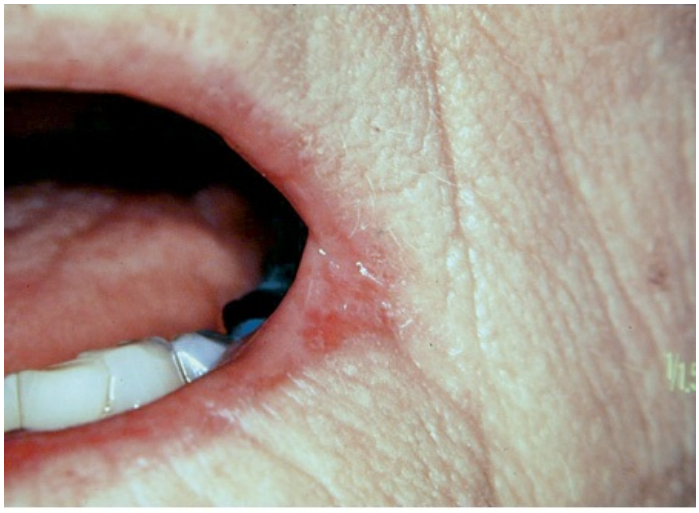

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Angular cheilitis.

Figure e12-8

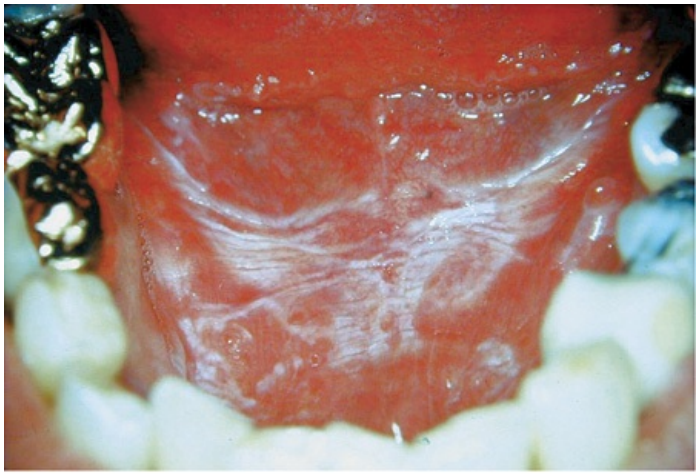

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sublingual leukoplakia.

Figure e12-9

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Epulis (gingival hypertrophy) under denture. ***B.*** Epulis fissuratum.

Figure e12-10

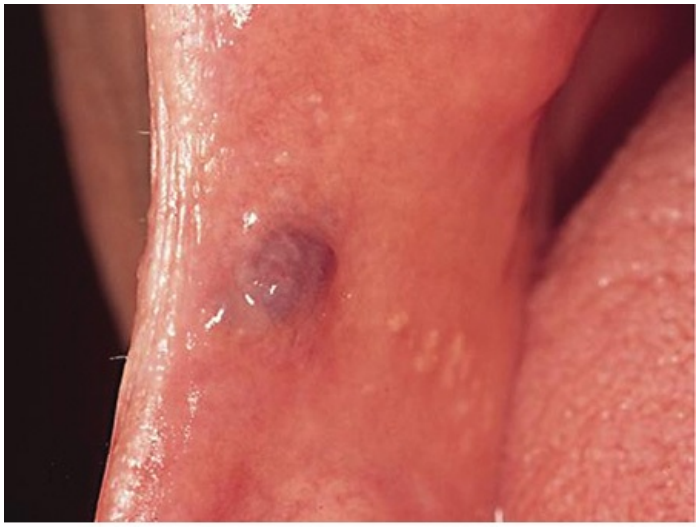

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Traumatic lesion inside of cheek.

Figure e12-11

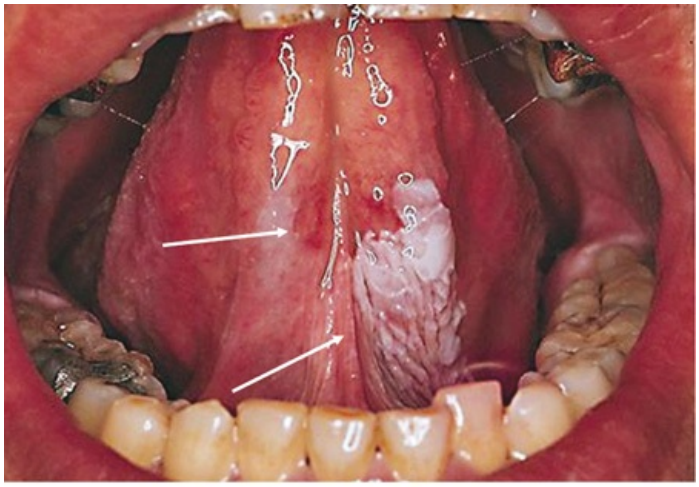

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sublingual keratosis.

Figure e12-12

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oral carcinoma.

Figure e12-13

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Healthy mouth.

Figure e12-14

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Geographic tongue.

Figure e12-15

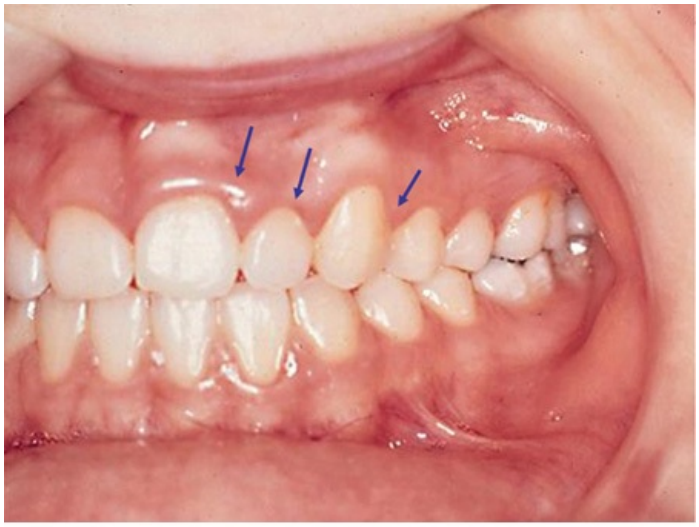

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Moderate gingivitis.

Figure e12-16

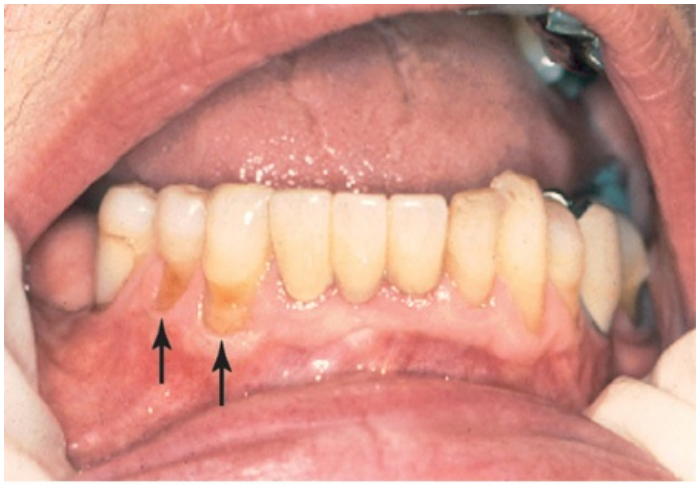

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gingival recession.

Figure e12-17

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Heavy calculus and gingival inflammation.

Figure e12-18

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe gingival inflammation and heavy calculus.

Figure e12-19

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Heavy plaque and gingival inflammation.

Figure e7-20

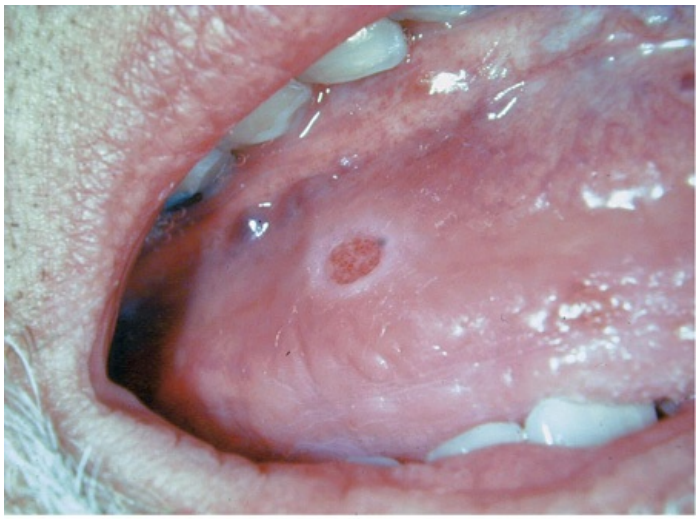

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ulcer on lateral border of tongue—potential carcinoma.

Figure e12-21

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Osteonecrosis.

Figure e12-22

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe periodontal disease, missing tooth, very mobile teeth.

Figure e12-23

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Salivary stone.

Figure e12-24

A ***B***

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Calculus. ***B.*** Teeth cleaned.

Figure e12-25

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Traumatic ulcer.

Figure e12-26

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fissured tongue.

Figure e12-27

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

White coated tongue—likely candidiasis.

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DYSPNEA

The American Thoracic Society defines *dyspnea* as a "subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity. The experience derives from interactions among multiple physiological, psychological, social, and environmental factors and may induce secondary physiological and behavioral responses." Dyspnea, a symptom, must be distinguished from the signs of increased work of breathing.

Mechanisms of Dyspnea

Respiratory sensations are the consequence of interactions between the *efferent*, or outgoing, motor output from the brain to the ventilatory muscles (feed-forward) and the *afferent*, or incoming, sensory input from receptors throughout the body (feedback), as well as the integrative processing of this information that we infer must be occurring in the brain **(Fig. 33-1)**. In contrast to painful sensations, which can often be attributed to the stimulation of a single nerve ending, dyspnea sensations are more commonly viewed as holistic, more akin to hunger or thirst. A given disease state may lead to dyspnea by one or more mechanisms, some of which may be operative under some circumstances, e.g., exercise, but not others, e.g., a change in position.

Figure 33-1

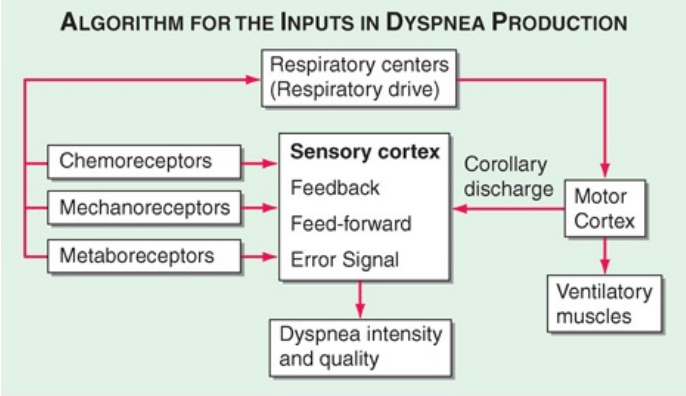

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypothetical model for integration of sensory inputs in the production of dyspnea. Afferent information from the receptors throughout the respiratory system projects directly to the sensory cortex to contribute to primary qualitative sensory experiences and provide feedback on the action of the ventilatory pump. Afferents also project to the areas of the brain responsible for control of ventilation. The motor cortex, responding to input from the control centers, sends neural messages to the ventilatory muscles and a corollary discharge to the sensory cortex (feed-forward with respect to the instructions sent to the muscles). If the feed-forward and feedback messages do not match, an error signal is generated and the intensity of dyspnea increases. *(Adapted from Gillette and Schwartzstein, 2005.)*

MOTOR EFFERENTS

Disorders of the ventilatory pump, most commonly increase airway resistance or stiffness (decreased compliance) of the respiratory system, are associated with increased work of breathing or a sense of an increased effort to breathe. When the muscles are weak or fatigued, greater effort is required, even though the mechanics of the system are normal. The increased neural output from the motor cortex is sensed via a corollary discharge, a neural signal that is sent to the sensory cortex at the same time that motor output is directed to the ventilatory muscles.

SENSORY AFFERENTS

Chemoreceptors in the carotid bodies and medulla are activated by hypoxemia, acute hypercapnia, and acidemia. Stimulation of these receptors, as well as others that lead to an increase in ventilation, produce a sensation of air hunger. Mechanoreceptors in the lungs, when stimulated by bronchospasm, lead to a sensation of chest tightness. J-receptors, sensitive to interstitial edema, and pulmonary vascular receptors, activated by acute changes in pulmonary artery pressure, appear to contribute to air hunger. Hyperinflation is associated with the sensation of increased work of breathing and an inability to get a deep breath or of an unsatisfying breath. Metaboreceptors, located in skeletal muscle, are believed to be activated by

changes in the local biochemical milieu of the tissue active during exercise and, when stimulated, contribute to the breathing discomfort.

INTEGRATION: EFFERENT-REAFFERENT MISMATCH

A discrepancy or mismatch between the feed-forward message to the ventilatory muscles and the feedback from receptors that monitor the response of the ventilatory pump increases the intensity of dyspnea. This is particularly important when there is a mechanical derangement of the ventilatory pump, such as in asthma or chronic obstructive pulmonary disease (COPD).

ANXIETY

Acute anxiety may increase the severity of dyspnea either by altering the interpretation of sensory data or by leading to patterns of breathing that heighten physiologic abnormalities in the respiratory system. In patients with expiratory flow limitation, for example, the increased respiratory rate that accompanies acute anxiety leads to hyperinflation, increased work and effort of breathing, and a sense of an unsatisfying breath.

Assessing Dyspnea

QUALITY OF SENSATION

As with pain, dyspnea assessment begins with a determination of the quality of the discomfort **(Table 33-1)**. Dyspnea questionnaires, or lists of phrases commonly used by patients, assist those who have difficulty describing their breathing sensations.

Table 33-1 Association of Qualitative Descriptors and Pathophysiologic Mechanisms of Shortness of Breath

Descriptor	Pathophysiology
Chest tightness or constriction	Bronchoconstriction, interstitial edema (asthma, myocardial ischemia)
Increased work or effort of breathing	Airway obstruction, neuromuscular disease (COPD, moderate to severe asthma, myopathy, kyphoscoliosis)
Air hunger, need to breathe, urge to breathe	Increased drive to breathe (CHF, pulmonary embolism, moderate to severe airflow obstruction)
Cannot get a deep breath, unsatisfying breath	Hyperinflation (asthma, COPD) and restricted tidal volume (pulmonary fibrosis, chest wall restriction)
Heavy breathing, rapid breathing, breathing more	Deconditioning

Abbreviations: CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease.

Source: From Schwartzstein and Feller-Kopman.

SENSORY INTENSITY

A modified Borg scale or visual analogue scale can be utilized to measure dyspnea at rest, immediately following exercise, or on recall of a reproducible physical task, e.g., climbing the stairs at home. An alternative approach is to inquire about the activities a patient can do, i.e., to gain a sense of the patient's disability. The Baseline Dyspnea Index and the Chronic Respiratory Disease Questionnaire are commonly used tools for this purpose.

AFFECTIVE DIMENSION

For a sensation to be reported as a symptom, it must be perceived as unpleasant and interpreted as abnormal. Laboratory studies have demonstrated that air hunger evokes a stronger affective response than does increased effort or work of breathing. Some therapies for dyspnea, such as pulmonary rehabilitation, may reduce breathing discomfort, in part, by altering this dimension.

Differential Diagnosis

Dyspnea is the consequence of deviations from normal function in the cardiopulmonary systems. These deviations produce breathlessness as a consequence of increased drive to breathe; increased effort or work of breathing; and/or stimulation of receptors in the heart, lungs, or vascular system. Most diseases of the respiratory system are associated with alterations in the mechanical properties of the lungs and/or chest wall, frequently as a consequence of disease of the airways or lung parenchyma. In contrast, disorders of the cardiovascular system more commonly lead to dyspnea by causing gas exchange abnormalities or stimulating pulmonary and/or vascular receptors (**Table 33-2**).

Table 33-2 Mechanisms of Dyspnea in Common Diseases

Disease	↑ Work of breathing	↑ Drive to breathe	Hypoxemia[a]	Acute Hypercapniaa	Stimulation of pulmonary receptors	Stimulation of vascular receptors	Metaboreceptors
COPD	•		•	•			

Asthma	•	•	•	•	•		
ILD	•	•	•	•	•		
PVD		•	•			•	
CPE	•	•	•		•	•	•
NCPE	•	•	•		•		
Anemia							•
Decond							•

[a] Hypoxemia and hypercapnia are not always present in these conditions. When hypoxemia is present, dyspnea usually persists, albeit at a reduced intensity, with correction of hypoxemia by the administration of supplemental oxygen.

Abbreviations: COPD, chronic obstructive pulmonary disease; CPE, cardiogenic pulmonary edema; Decond, deconditioning; ILD, interstitial lung disease; NCPE, noncardiogenic pulmonary edema; PVD, pulmonary vascular disease.

RESPIRATORY SYSTEM DYSPNEA

Diseases of the Airways

Asthma and COPD, the most common obstructive lung diseases, are characterized by expiratory airflow obstruction, which typically leads to dynamic hyperinflation of the lungs and chest wall. Patients with moderate to severe disease have increased resistive and elastic loads (a term that relates to the stiffness of the system) on the ventilatory muscles and increased work of breathing. Patients with acute bronchoconstriction also complain of a sense of tightness, which can exist even when lung function is still within the normal range. These patients commonly hyperventilate. Both the chest tightness and hyperventilation are probably due to stimulation of pulmonary receptors. Both asthma and COPD may lead to hypoxemia and hypercapnia from ventilation-perfusion (V̇/Q) mismatch (and diffusion limitation during exercise with emphysema); hypoxemia is much more common than hypercapnia as a consequence of the different ways in which oxygen and carbon dioxide bind to hemoglobin.

Diseases of the Chest Wall

Conditions that stiffen the chest wall, such as kyphoscoliosis, or that weaken ventilatory muscles, such as myasthenia gravis or the Guillain-Barré syndrome, are also associated with an increased effort to breathe. Large pleural effusions may contribute to dyspnea, both by increasing the work of breathing and by stimulating pulmonary receptors if there is associated atelectasis.

Diseases of the Lung Parenchyma

Interstitial lung diseases, which may arise from infections, occupational exposures, or autoimmune disorders, are associated with increased stiffness (decreased compliance) of the lungs and increased work of breathing. In addition, V̇/Q mismatch, and destruction and/or thickening of the alveolar-capillary interface may lead to hypoxemia and an increased drive to breathe. Stimulation of pulmonary receptors may further enhance the hyperventilation characteristic of mild to moderate interstitial disease.

CARDIOVASCULAR SYSTEM DYSPNEA

Diseases of the Left Heart

Diseases of the myocardium resulting from coronary artery disease and nonischemic cardiomyopathies result in a greater left-ventricular end-diastolic volume and an elevation of the left-ventricular end-diastolic, as well as pulmonary capillary pressures. These elevated pressures lead to interstitial edema and stimulation of pulmonary receptors, thereby causing dyspnea; hypoxemia due to V̇/Q mismatch may also contribute to breathlessness. Diastolic dysfunction, characterized by a very stiff left ventricle, may lead to severe dyspnea with relatively mild degrees of physical activity, particularly if it is associated with mitral regurgitation.

Diseases of the Pulmonary Vasculature

Pulmonary thromboemoblic disease and primary diseases of the pulmonary circulation (primary pulmonary hypertension, pulmonary vasculitis) cause dyspnea via increased pulmonary-artery pressure and stimulation of pulmonary receptors. Hyperventilation is common, and hypoxemia may be present. However, in most cases, use of supplemental oxygen has minimal effect on the severity of dyspnea and hyperventilation.

Diseases of the Pericardium

Constrictive pericarditis and cardiac tamponade are both associated with increased intracardiac and pulmonary vascular pressures, which are the likely cause of dyspnea in these conditions. To the extent that cardiac output is limited, at rest or with exercise, stimulation of metaboreceptors and chemoreceptors (if lactic acidosis develops) contribute as well.

DYSPNEA WITH NORMAL RESPIRATORY AND CARDIOVASCULAR SYSTEMS

Mild to moderate anemia is associated with breathing discomfort during exercise. This is thought to be related to stimulation of metaboreceptors; oxygen saturation is normal in patients with anemia. The breathlessness associated with obesity is probably due to multiple mechanisms, including high cardiac

output and impaired ventilatory pump function (decreased compliance of the chest wall). Cardiovascular deconditioning (poor fitness) is characterized by the early development of anaerobic metabolism and the stimulation of chemoreceptors and metaboreceptors.

Approach to the Patient: Dyspnea

(Fig. 33-2) In obtaining a *history*, the patient should be asked to describe in his/her own words what the discomfort feels like, as well as the effect of position, infections, and environmental stimuli on the dyspnea. Orthopnea is a common indicator of congestive heart failure (CHF), mechanical impairment of the diaphragm associated with obesity, or asthma triggered by esophageal reflux. Nocturnal dyspnea suggests CHF or asthma. Acute, intermittent episodes of dyspnea are more likely to reflect episodes of myocardial ischemia, bronchospasm, or pulmonary embolism, while chronic persistent dyspnea is typical of COPD, interstitial lung disease, and chronic thromboembolic disease. Risk factors for occupational lung disease and for coronary artery disease should be elicited. Left atrial myxoma or hepatopulmonary syndrome should be considered when the patient complains of *platypnea*, defined as dyspnea in the upright position with relief in the supine position.

Figure 33-2

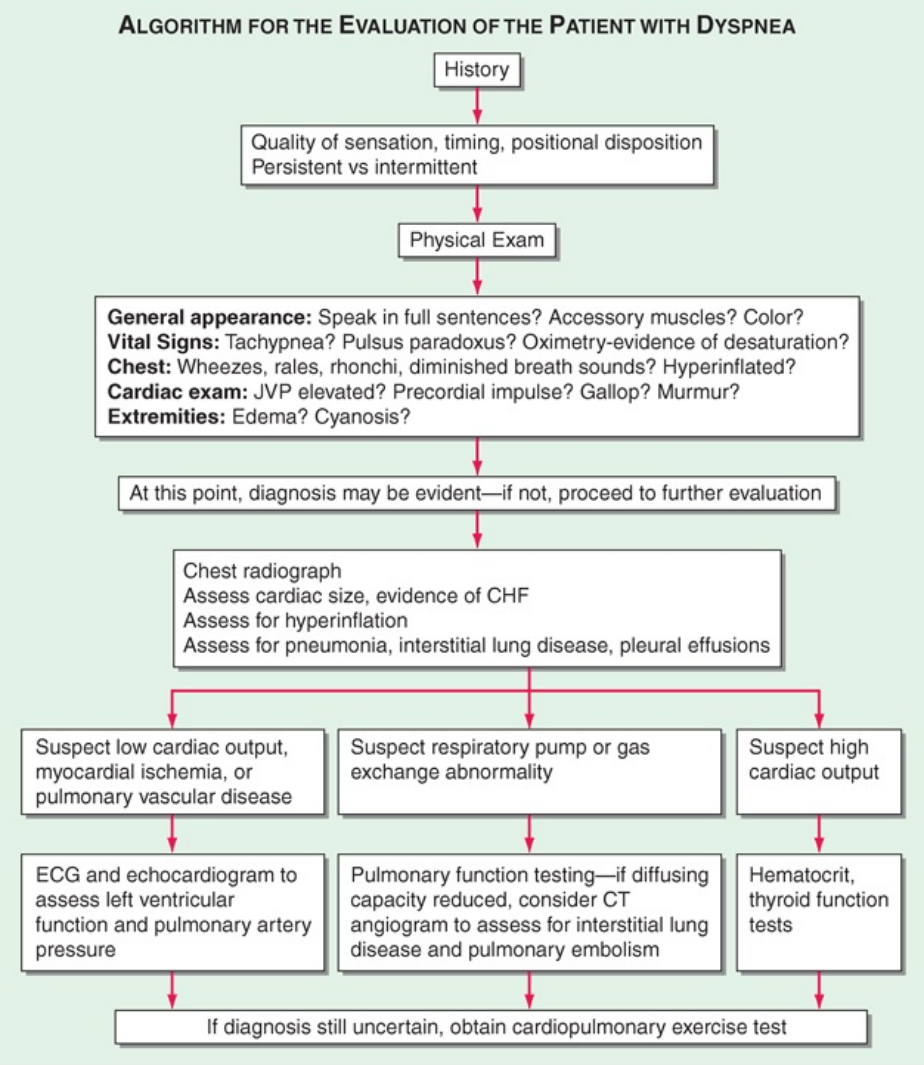

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An algorithm for the evaluation of the patient with dyspnea. JVP, jugular venous pulse; CHF, congestive heart failure; ECG, electrocardiogram; CT, computed tomography. *(Adapted from Schwartzstein and Feller-Kopman, 2003.)*

The *physical examination* should begin during the interview of the patient. Inability of the patient to speak in full sentences before stopping to get a deep breath suggests a condition that leads to stimulation of the controller or an impairment of the ventilatory pump with reduced vital capacity. Evidence for increased work of breathing (supraclavicular retractions, use of accessory muscles of ventilation, and the tripod position, characterized by sitting with one's hands braced on the knees) is indicative of increased airway resistance or stiff lungs and chest wall. When measuring the vital signs, one should accurately assess the respiratory rate and measure the pulsus paradoxus (Chap. 239); if it is >10 mmHg, consider the presence of COPD or

acute asthma. During the general examination, signs of anemia (pale conjunctivae), cyanosis, and cirrhosis (spider angiomata, gynecomastia) should be sought. Examination of the chest should focus on symmetry of movement; percussion (dullness indicative of pleural effusion, hyperresonance a sign of emphysema); and auscultation (wheezes, rales, rhonchi, prolonged expiratory phase, diminished breath sounds, which are clues to disorders of the airways, and interstitial edema or fibrosis). The cardiac examination should focus on signs of elevated right heart pressures (jugular venous distention, edema, accentuated pulmonic component to the second heart sound); left ventricular dysfunction (S3 and S4 gallops); and valvular disease (murmurs). When examining the abdomen with the patient in the supine position, it should be noted whether there is paradoxical movement of the abdomen (inward motion during inspiration), a sign of diaphragmatic weakness; rounding of the abdomen during exhalation is suggestive of pulmonary edema. Clubbing of the digits may be an indication of interstitial pulmonary fibrosis, and the presence of joint swelling or deformation as well as changes consistent with Raynaud's disease may be indicative of a collagen-vascular process that can be associated with pulmonary disease.

Patients with exertional dyspnea should be asked to walk under observation in order to reproduce the symptoms. The patient should be examined for new findings that were not present at rest and for oxygen saturation.

Following the history and physical examination, a *chest radiograph* should be obtained. The lung volumes should be assessed (hyperinflation indicates obstructive lung disease; low lung volumes suggest interstitial edema or fibrosis, diaphragmatic dysfunction, or impaired chest wall motion). The pulmonary parenchyma should be examined for evidence of interstitial disease and emphysema. Prominent pulmonary vasculature in the upper zones indicates pulmonary venous hypertension, while enlarged central pulmonary arteries suggest pulmonary artery hypertension. An enlarged cardiac silhouette suggests a dilated cardiomyopathy or valvular disease. Bilateral pleural effusions are typical of CHF and some forms of collagen vascular disease. Unilateral effusions raise the specter of carcinoma and pulmonary embolism but may also occur in heart failure. *Computed tomography* (CT) *of the chest* is generally reserved for further evaluation of the lung parenchyma (interstitial lung disease) and possible pulmonary embolism.

Laboratory studies should include an electrocardiogram to look for evidence of ventricular hypertrophy and prior myocardial infarction. Echocardiography is indicated in patients in whom systolic dysfunction, pulmonary hypertension, or valvular heart disease is suspected. Bronchoprovocation testing is useful in patients with intermittent symptoms suggestive of asthma but normal physical examination and lung function; up to one-third of patients with the clinical diagnosis of asthma do not have reactive airways disease when formally tested.

DISTINGUISHING CARDIOVASCULAR FROM RESPIRATORY SYSTEM DYSPNEA

If a patient has evidence of both pulmonary and cardiac disease, a cardiopulmonary exercise test should be carried out to determine which system is responsible for the exercise limitation. If, at peak exercise, the patient achieves predicted maximal ventilation, demonstrates an increase in dead space or hypoxemia, or develops bronchospasm, the respiratory system is probably the cause of the problem. Alternatively, if the heart rate is >85% of the predicted maximum, if anaerobic threshold occurs early, if the blood pressure becomes excessively high or decreases during exercise, if the O_2 pulse (O_2 consumption/heart rate, an indicator of stroke volume) falls, or if there are ischemic changes on the electrocardiogram, an abnormality of the cardiovascular system is likely the explanation for the breathing discomfort.

Treatment: Dyspnea

The first goal is to correct the underlying problem responsible for the symptom. If this is not possible, one attempts to lessen the intensity of the symptom and its effect on the patient's quality of life. Supplemental O_2 should be administered if the resting O_2 saturation is ≤89% or if the patient's saturation drops to these levels with activity. For patients with COPD, pulmonary rehabilitation programs have demonstrated positive effects on dyspnea, exercise capacity, and rates of hospitalization. Studies of anxiolytics and antidepressants have not demonstrated consistent benefit. Experimental interventions—e.g., cold air on the face, chest-wall vibration, and inhaled furosemide—to modulate the afferent information from receptors throughout the respiratory system are being studied.

PULMONARY EDEMA

Mechanisms of Fluid Accumulation

The extent to which fluid accumulates in the interstitium of the lung depends on the balance of hydrostatic and oncotic forces within the pulmonary capillaries and in the surrounding tissue. Hydrostatic pressure favors movement of fluid from the capillary into the interstitium. The oncotic pressure, which is determined by the protein concentration in the blood, favors movement of fluid into the vessel. Albumin, the primary protein in the plasma, may be low in conditions such as cirrhosis and nephrotic syndrome. While hypoalbuminemia favors movement of fluid into the tissue for any given hydrostatic pressure in the capillary, it is usually not sufficient by itself to cause interstitial edema. In a healthy individual, the tight junctions of the capillary endothelium are impermeable to proteins, and the lymphatics in the tissue carry away the small amounts of protein that may leak out; together, these factors result in an oncotic force that maintains fluid in the capillary. Disruption of the endothelial barrier, however, allows protein to escape the capillary bed and enhances the movement of fluid into the tissue of the lung.

CARDIOGENIC PULMONARY EDEMA

(See also Chap. 272) Cardiac abnormalities that lead to an increase in pulmonary venous pressure shift the balance of forces between the capillary and the interstitium. Hydrostatic pressure is increased and fluid exits the capillary at an increased rate, resulting in interstitial and, in more severe cases, alveolar edema. The development of pleural effusions may further compromise respiratory system function and contribute to breathing discomfort.

Early signs of pulmonary edema include exertional dyspnea and orthopnea. Chest radiographs show peribronchial thickening, prominent vascular markings in the upper lung zones, and Kerley B lines. As the pulmonary edema worsens, alveoli fill with fluid; the chest radiograph shows patchy alveolar filling, typically in a perihilar distribution, which then progresses to diffuse alveolar infiltrates. Increasing airway edema is associated with rhonchi and wheezes.

NONCARDIOGENIC PULMONARY EDEMA

In noncardiogenic pulmonary edema, lung water increases due to damage of the pulmonary capillary lining with leakage of proteins and other macromolecules into the tissue; fluid follows the protein as oncotic forces are shifted from the vessel to the surrounding lung tissue. This process is associated with dysfunction of the surfactant lining the alveoli, increased surface forces, and a propensity for the alveoli to collapse at low lung volumes. Physiologically, noncardiogenic pulmonary edema is characterized by intrapulmonary shunt with hypoxemia and decreased pulmonary compliance. Pathologically, hyaline membranes are evident in the alveoli, and inflammation leading to pulmonary fibrosis may be seen. Clinically, the picture ranges from mild dyspnea to respiratory failure. Auscultation of the lungs may be relatively normal despite chest radiographs that show diffuse alveolar infiltrates. CT scans demonstrate that the distribution of alveolar edema is more heterogeneous than was once thought. Although normal intracardiac pressures are considered by many to be part of the definition of noncardiogenic pulmonary edema, the pathology of the process, as described above, is distinctly different, and one can observe a combination of cardiogenic and noncardiogenic pulmonary edema in some patients.

It is useful to categorize the causes of noncardiogenic pulmonary edema in terms of whether the injury to the lung is likely to result from direct, indirect, or pulmonary vascular causes **(Table 33-3)**. Direct injuries are mediated via the airways (e.g., aspiration) or as the consequence of blunt chest trauma. Indirect injury is the consequence of mediators that reach the lung via the blood stream. The third category includes conditions that may be the consequence of acute changes in pulmonary vascular pressures, possibly the result of sudden autonomic discharge in the case of neurogenic and high-altitude pulmonary edema, or sudden swings of pleural pressure, as well as transient damage to the pulmonary capillaries in the case of reexpansion pulmonary edema.

Table 33-3 Common Causes of Noncardiogenic Pulmonary Edema

Direct injury to lung
Chest trauma, pulmonary contusion
Aspiration
Smoke inhalation
Pneumonia
Oxygen toxicity
Pulmonary embolism, reperfusion
Hematogenous injury to lung
Sepsis
Pancreatitis
Nonthoracic trauma
Leukoagglutination reactions
Multiple transfusions
Intravenous drug use, e.g., heroin
Cardiopulmonary bypass
Possible lung injury plus elevated hydrostatic pressures
High-altitude pulmonary edema
Neurogenic pulmonary edema
Reexpansion pulmonary edema

DISTINGUISHING CARDIOGENIC FROM NONCARDIOGENIC PULMONARY EDEMA

The *history* is essential for assessing the likelihood of underlying cardiac disease as well as for identification of one of the conditions associated with noncardiogenic pulmonary edema. The *physical examination* in cardiogenic pulmonary edema is notable for evidence of increased intracardiac pressures (S3 gallop, elevated jugular venous pulse, peripheral edema), and rales and/or wheezes on auscultation of

the chest. In contrast, the physical examination in noncardiogenic pulmonary edema is dominated by the findings of the precipitating condition; pulmonary findings may be relatively normal in the early stages. The *chest radiograph* in cardiogenic pulmonary edema typically shows an enlarged cardiac silhouette, vascular redistribution, interstitial thickening, and perihilar alveolar infiltrates; pleural effusions are common. In noncardiogenic pulmonary edema, heart size is normal, alveolar infiltrates are distributed more uniformly throughout the lungs, and pleural effusions are uncommon. Finally, the *hypoxemia* of cardiogenic pulmonary edema is due largely to V̇/Q mismatch and responds to the administration of supplemental oxygen. In contrast, hypoxemia in noncardiogenic pulmonary edema is due primarily to intrapulmonary shunting and typically persists despite high concentrations of inhaled O_2.

FURTHER READINGS

Aaron SD et al: Overdiagnosis of asthma in obese and nonobese adults. CMAJ 179:1121, 2008 [PMID: 19015563]

Abidov A et al: Prognostic significance of dyspnea in patients referred for cardiac stress testing. N Engl J Med 353:1889, 2005 [PMID: 16267320]

Banzett RB et al: The affective dimension of laboratory dyspnea: Air hunger is more unpleasant than work/effort. Am J Respir Crit Care Med 177:1384, 2008 [PMID: 18369200]

Dyspnea mechanisms, assessment, and management: A consensus statement. Am J Respir Crit Care Med 159:321, 1999

Gillette MA, Schwartzstein RM: Mechanisms of dyspnea, in *Supportive Care in Respiratory Disease*, SH Ahmedzai and MF Muer (eds). Oxford, U.K., Oxford University Press, 2005

Mahler DA et al: Descriptors of breathlessness in cardiorespiratory diseases. Am J Respir Crit Care Med 154:1357, 1996 [PMID: 8912748]

———, O'Donnell de (eds): *Dyspnea: Mechanisms, Measurement, and Management*. New York, Marcel Dekker, 2005

Schwartzstein RM: The language of dyspnea, in *Dyspnea: Mechanisms, Measurement, and Management*, DA Mahler and De O'Donnell (eds). New York, Marcel Dekker, 2005

——— Feller-Kopman D: Shortness of breath, in *Primary Cardiology*, 2nd ed, E Braunwald and L Goldman (eds). Philadelphia: WB Saunders, 2003

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

COUGH

Cough provides an essential protective function for human airways and lungs. Without an effective cough reflex, we are at risk for retained airway secretions and aspirated material, predisposing to infection, atelectasis, and respiratory compromise. At the other extreme, excessive coughing can be exhausting; can be complicated by emesis, syncope, muscular pain, or rib fractures; and can aggravate abdominal or inguinal hernias and urinary incontinence. Cough is often a clue to the presence of respiratory disease. In many instances, cough is an expected and accepted manifestation of disease, such as during an acute respiratory tract infection. However, persistent cough in the absence of other respiratory symptoms commonly causes patients to seek medical attention, accounting for as many as 10–30% of referrals to pulmonary specialists.

Cough Mechanism

Spontaneous cough is triggered by stimulation of sensory nerve endings that are thought to be primarily rapidly adapting receptors and C-fibers. Both chemical (e.g., capsaicin) and mechanical (e.g., particulates in air pollution) stimuli may initiate the cough reflex. A cationic ion channel, called the type-1 vanilloid receptor, is found on rapidly adapting receptors and C-fibers; it is the receptor for capsaicin, and its expression is increased in patients with chronic cough. Afferent nerve endings richly innervate the pharynx, larynx, and airways to the level of terminal bronchioles and into the lung parenchyma. They may also be found in the external auditory meatus (the auricular branch of the vagus nerve, called the Arnold nerve) and in the esophagus. Sensory signals travel via the vagus and superior laryngeal nerves to a region of the brainstem in the nucleus tractus solitarius, vaguely identified as the "cough center." Mechanical stimulation of bronchial mucosa in a transplanted lung (in which the vagus nerve has been severed) does not produce cough.

The cough reflex involves a highly orchestrated series of involuntary muscular actions, with the potential for input from cortical pathways as well. The vocal cords adduct, leading to transient upper-airway occlusion. Expiratory muscles contract, generating positive intrathoracic pressures as high as 300 mm Hg. With sudden release of the laryngeal contraction, rapid expiratory flows are generated, exceeding the normal "envelope" of maximal expiratory flow seen on the flow-volume curve (**Fig. 34-1**). Bronchial smooth muscle contraction together with dynamic compression of airways narrows airway lumens and maximizes the velocity of exhalation (as fast as 50 miles per hour). The kinetic energy available to dislodge mucus from the inside of airway walls is directly proportional to the square of the velocity of expiratory airflow. A deep breath preceding a cough optimizes the function of the expiratory muscles; a series of repetitive coughs at successively lower lung volumes sweeps the point of maximal expiratory velocity progressively further into the lung periphery.

Figure 34-1

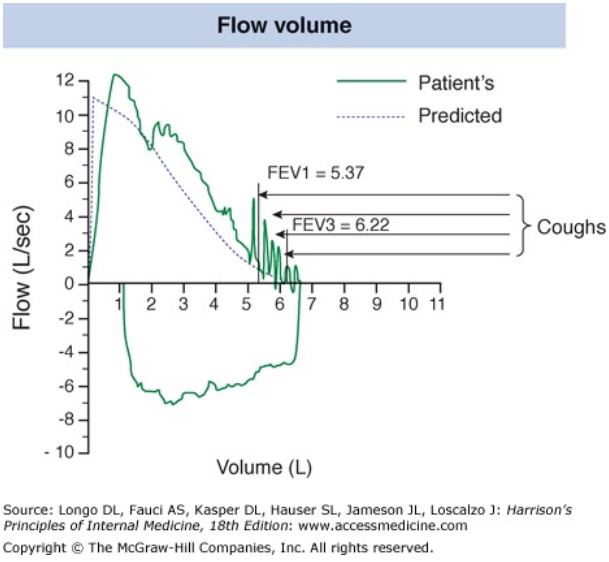

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flow-Volume Loop. Flow-volume curve with spikes of high expiratory flow achieved with cough.

Impaired Cough

Weak or ineffective cough compromises the ability to clear lower respiratory tract infections, predisposing to more serious infections and their sequelae. Weakness, paralysis, or pain of the expiratory (abdominal and intercostal) muscles is foremost on the list of causes of impaired cough (**Table 34-1**). Cough strength is generally assessed qualitatively; peak expiratory flow or maximal expiratory pressure at the mouth can be used as a surrogate marker for cough strength. A variety of assistive devices and techniques have been developed to improve cough strength, spanning the gamut from simple (splinting the abdominal muscles with a tightly-held pillow to reduce post-operative pain while coughing) to complex (a mechanical cough-assist device applied via face mask or tracheal tube that applies a cycle of positive pressure followed rapidly by negative pressure). Cough may fail to clear secretions despite a preserved ability to generate normal expiratory velocities, either due to abnormal airway secretions (e.g., bronchiectasis due to cystic fibrosis) or structural abnormalities of the airways (e.g., tracheomalacia with expiratory collapse during cough).

Table 34-1 Causes of Impaired Cough
Decreased expiratory-muscle strength
Decreased inspiratory-muscle strength
Chest-wall deformity
Impaired glottic closure or tracheostomy
Tracheomalacia
Abnormal airway secretions
Central respiratory depression (e.g., anesthesia, sedation, or coma)

Symptomatic Cough

The cough of chronic bronchitis in long-term cigarette smokers rarely leads the patient to seek medical advice. It lasts only seconds to a few minutes, is productive of benign-appearing mucoid sputum, and is not discomforting. Similarly, cough may occur in the context of other respiratory symptoms that, together, point to a diagnosis, such as when cough is accompanied by wheezing, shortness of breath, and chest tightness after exposure to a cat or other sources of allergens. At times, however, cough is the dominant or sole symptom of disease, and it may be of sufficient duration and severity that relief is sought. The duration of cough is a clue to its etiology. Acute cough (<3 weeks) is most commonly due to a respiratory tract infection, aspiration event, or inhalation of noxious chemicals or smoke. Subacute cough (3–8 weeks duration) is frequently the residuum from a tracheobronchitis, such as in pertussis or "post-viral tussive syndrome." Chronic cough (>8 weeks) may be caused by a wide variety of cardiopulmonary diseases, including those of inflammatory, infectious, neoplastic, and cardiovascular etiologies. When initial assessment with chest examination and radiograph is normal, cough-variant asthma, gastroesophageal reflux, nasopharyngeal drainage, and medications (angiotensin converting enzyme [ACE] inhibitors) are the most common causes of chronic cough. Cough of less than 8 weeks' duration may be the early manifestation of a disease causing chronic cough.

Assessment of Chronic Cough

Details as to the sound, time of occurrence during the day, and pattern of coughing infrequently provide useful etiology clues. Regardless of cause, cough often worsens when one first lies down at night or with talking or in association with the hyperpnea of exercise; it frequently improves with sleep. Exceptions might include the characteristic inspiratory whoop after a paroxysm of coughing that suggests pertussis or the cough that occurs only with certain allergic exposures or exercise in cold air, as in asthma. Useful historical questions include the circumstances surrounding the onset of cough, what makes the cough better or worse, and whether or not the cough produces sputum.

The physical examination seeks clues to the presence of cardiopulmonary disease, including findings such as wheezing or crackles on chest examination. Examination of the auditory canals and tympanic membranes (for irritation of the tympanic membrane resulting in stimulation of Arnold's nerve), the nasal passageways (for rhinitis), and nails (for clubbing) may also provide etiologic clues. Because cough can be a manifestation of a systemic disease, such as sarcoidosis or vasculitis, a thorough general examination is equally important.

In virtually all instances, evaluation of chronic cough merits a chest radiograph. The list of diseases that can cause persistent coughing without other symptoms and without detectable abnormality on physical examination is long. It includes serious illnesses such as Hodgkin's disease in young adults and lung cancer in an older population. An abnormal chest film leads to evaluation of the radiographic abnormality to explain the symptom of cough. A normal chest image provides valuable reassurance to the patient and the patient's family, who may have imagined the direst explanation for the cough.

In a patient with chronic productive cough, examination of expectorated sputum is warranted. Purulent-appearing sputum should be sent for routine bacterial culture and, in certain circumstances, mycobacterial culture as well. Cytologic examination of mucoid sputum may be useful to assess for malignancy and to distinguish neutrophilic from eosinophilic bronchitis. Expectoration of blood—whether streaks of blood, blood mixed with airway secretions, or pure blood —deserves a special approach to assessment and management, as discussed below.

Chronic Cough with a Normal Chest Radiograph

It is commonly held that use of an angiotensin-converting enzyme inhibitor; post-nasal drainage; gastroesophageal reflux; and asthma, alone or in combination, account for more than 90% of patients who have chronic cough and a normal or noncontributory chest radiograph. However, clinical experience does not support this contention, and strict adherence to this concept discourages the search for alternative explanations by both clinicians and researchers. On the one hand, chronic idiopathic cough is common and its management deserves study and discussion. On the other hand, serious pulmonary diseases, including inflammatory lung diseases, chronic infections, and neoplasms, may remain occult on plain chest imaging and require additional testing for detection.

ACE inhibitor-induced cough occurs in 5–30% of patients taking ACE inhibitors and is not dose-dependent. Any patient with chronic unexplained cough who is taking an ACE inhibitor should be given a trial period off the medication, regardless of the timing of the onset of cough relative to the initiation of ACE inhibitor therapy. In most instances, a safe alternative is available; angiotensin-receptor blockers do not cause cough. Failure to observe a decrease in cough after one month off medication argues strongly against this diagnosis. ACE metabolizes bradykinin and other tachykinins, such as substance P. The mechanism of ACE inhibitor cough may involve sensitization of sensory nerve endings due to accumulation of bradykinin. In support of this hypothesis, polymorphisms in the neurokinin-2 receptor gene are associated with ACE inhibitor–induced cough.

Post-nasal drainage of any etiology can cause cough as a response to stimulation of sensory receptors of the cough-reflex pathway in the hypopharynx or aspiration of draining secretions into the trachea. Clues to this etiology include symptoms of post-nasal drip, frequent throat clearing, and sneezing and rhinorrhea. On speculum examination of the nose, one may see excess mucoid or purulent secretions, inflamed and edematous nasal mucosa, and/or nasal polyps; in addition, one might visualize secretions or a cobblestoned appearance of the mucosa along the posterior pharyngeal wall. Unfortunately, there is no means by which to quantitate post-nasal drainage. In many instances, one is left to rely on a qualitative judgment based on subjective information provided by the patient. This assessment must also be counterbalanced by the fact that many people who have chronic post-nasal drainage do not experience cough.

Linking gastroesophageal reflux to chronic cough poses similar challenges. It is thought that reflux of gastric contents into the lower esophagus may trigger cough via reflex pathways initiated in the esophageal mucosa. Reflux to the level of the

pharynx with consequent aspiration of gastric contents causes a chemical bronchitis and possible pneumonitis that can elicit cough for days after the aspiration event. Retrosternal burning after meals or on recumbency, frequent eructation, hoarseness, and throat pain are potential clues to gastroesophageal reflux. Reflux may also elicit no or minimal symptoms. Glottic inflammation may be a clue to recurrent reflux to the level of the throat, but it is a nonspecific finding and requires direct or indirect laryngoscopy for detection. Quantification of the frequency and level of reflux requires a somewhat invasive procedure to measure esophageal pH directly (a catheter with pH probe placed nasopharyngeally in the esophagus for 24 h, or pH monitoring using a radiotransmitter capsule placed endoscopically into the esophagus). Precise interpretation of test results enabling one to link reflux and cough in a causative way remains debated. Again, assigning the cause of cough to gastroesophageal reflux must be weighed against the observation that many people with chronic reflux (such as frequently occurs during pregnancy) do not experience chronic cough.

Cough alone as a manifestation of asthma is common in children but not in adults. Cough due to asthma in the absence of wheezing, shortness of breath, and chest tightness is referred to as "cough-variant asthma." A history suggestive of cough-variant asthma ties the onset of cough to typical triggers for asthma and resolution of cough upon withdrawal from exposure to them. Objective testing can establish the diagnosis of asthma (airflow obstruction on spirometry that varies over time or reverses in response to bronchodilator) or exclude it with certainty (negative response to bronchoprovocation challenge, such as with methacholine). In a patient capable of making reliable measurements, home expiratory peak flow monitoring can be used as a cost-effective method to support or discount a diagnosis of asthma.

Chronic eosinophilic bronchitis causes chronic cough with a normal chest radiograph. This condition is characterized by sputum eosinophilia in excess of 3% without airflow obstruction or bronchial hyperresponsiveness and is successfully treated with inhaled glucocorticoids.

Treatment of chronic cough in a patient with a normal chest radiograph is often empiric and is targeted at the most likely cause or causes of cough as determined by history, physical examination, and possibly pulmonary-function testing. Therapy for post-nasal drainage depends on the presumed etiology (infection, allergy, or vasomotor rhinitis) and may include systemic antihistamines; antibiotics; nasal saline irrigation; and nasal pump sprays with corticosteroids, antihistamines, or anticholinergics. Antacids, histamine type-2 (H2) receptor antagonists, and proton-pump inhibitors are used to neutralize or decrease production of gastric acid in gastroesophageal reflux disease; dietary changes, elevation of the head and torso during sleep, and medications to improve gastric emptying are additional therapies. Cough-variant asthma typically responds well to inhaled glucocorticoids and intermittent use of inhaled beta-agonist bronchodilators.

Patients who fail to respond to treatment of the common causes of cough or who have had these causes excluded by appropriate diagnostic testing should undergo chest CT. Examples of diseases causing cough that may be missed on chest x-ray include carcinoid tumor, early interstitial lung disease, bronchiectasis, and atypical mycobacterial pulmonary infection. On the other hand, patients with chronic cough who have normal chest examination, lung function, oxygenation, and chest CT imaging can be reassured as to the absence of serious pulmonary pathology.

Symptomatic Treatment of Cough

Chronic idiopathic cough is distressingly common. It is often experienced as a tickle or sensitivity in the throat area, occurs more often in women, and is typically "dry" or at most productive of scant amounts of mucoid sputum. It can be exhausting, interfere with work, and cause social embarrassment. Once serious underlying cardiopulmonary pathology has been excluded, an attempt at cough suppression is appropriate. Most effective are narcotic cough suppressants, such as codeine or hydrocodone, which are thought to act in the "cough center" in the brainstem. The tendency of narcotic cough suppressants to cause drowsiness and constipation and their potential for addictive dependence limit their appeal for long-term use. Dextromethorphan is an over-the-counter, centrally acting cough suppressant with fewer side effects and less efficacy compared to the narcotic cough suppressants. It is thought to have a different site of action than narcotic cough suppressants and can be used in combination with them if necessary. Benzonatate is thought to inhibit neural activity of sensory nerves in the cough-reflex pathway. It is generally free of side effects; however, its effectiveness in suppressing cough is variable and unpredictable. Novel cough suppressants without the limitations of currently available therapies are greatly needed. Approaches that are being explored include development of neurokinin receptor antagonists, type-1 vanilloid receptor antagonists, and novel opioid and opioidlike receptor agonists.

HEMOPTYSIS

Hemoptysis is the expectoration of blood from the respiratory tract. It can arise from any part of the respiratory tract, from the alveoli to the glottis. It is important, however, to distinguish hemoptysis from epistaxis (i.e., bleeding from the nasopharynx) and hematemesis (i.e., bleeding from the upper gastrointestinal tract). Hemoptysis can range from blood-tinged sputum to life-threatening large volumes of bright red blood. For most patients, any degree of hemoptysis can be anxiety-producing and often prompts medical evaluation.

While precise epidemiologic data are lacking, the most common etiology of hemoptysis is infection of the medium-sized airways. In the United States, this is usually due to a viral or bacterial bronchitis. Hemoptysis can arise in the setting of either acute bronchitis or during an exacerbation of chronic bronchitis. Worldwide, the most common cause of hemoptysis is tuberculous infection presumably owing to the high prevalence of the disease and its predilection for cavity formation. While these are the most common causes, there is an extensive differential diagnosis for hemoptysis, and a step-wise approach to the evaluation of this symptom is appropriate.

Etiology

One way to approach the source of hemoptysis is systematically to assess for potential sites of bleeding from the alveolus to the mouth. Diffuse bleeding in the alveolar space, often referred to as diffuse alveolar hemorrhage (DAH), may present with hemoptysis, although this is not always the case. Causes of DAH can be divided into inflammatory and noninflammatory types. Inflammatory DAH is due to small vessel vasculitis/capillaritis from a variety of diseases, including granulomatosis with polyangiitis (Wegener's) and microscopic polyangiitis. Similarly, systemic autoimmune disease, such as systemic lupus erythematosus (SLE), can manifest as pulmonary capillaritis and result in DAH. Antibodies to the alveolar basement membrane, as are seen in Goodpasture's disease, can also result in alveolar hemorrhage. In the early time period after a bone marrow transplant (BMT), patients can also develop a form of inflammatory DAH, which can be catastrophic and life-threatening. The exact pathophysiology of this process is not well understood, but DAH should be suspected in patients with sudden-onset dyspnea and hypoxemia in the first 100 days after a BMT.

Alveoli can also bleed due to noninflammatory causes, most commonly due to direct inhalational injury. This category includes thermal injury from fires, inhalation of illicit substances (e.g., cocaine), and inhalation of toxic chemicals. If alveoli are irritated from any process, patients with thrombocytopenia, coagulopathy, or antiplatelet or anticoagulant use will have an increased risk of developing hemoptysis.

As already noted, the most common site of hemoptysis is bleeding from the small- to medium-sized airways. Irritation and injury of the bronchial mucosal can lead to small-volume bleeding. More significant hemoptysis can also occur because of the proximity of the bronchial artery and vein to the airway, running together in what is often referred to as the "bronchovascular bundle." In the smaller airways, these blood vessels are close to the airspace and, therefore, lesser degrees of inflammation or injury can result in rupture of these vessels into the airways. Of note, while alveolar hemorrhage arises from capillaries that are part of the low-pressure pulmonary circulation, bronchial bleeding is generally from bronchial arteries, which are under systemic pressure and, therefore, predisposed to larger-volume bleeding.

Any infection of the airways can result in hemoptysis, although, most commonly, acute bronchitis is caused by viral infection. In patients with a history of chronic bronchitis, bacterial super infection with organisms such as *Streptococcus pneumoniae*, *Hemophilus influenzae*, or *Moraxella catarrhalis* can also result in hemoptysis. Patients with bronchiectasis, a permanent dilation and irregularity of the airways, are particularly prone to hemoptysis due to anatomic abnormalities that bring the bronchial arteries closer to the mucosal surface and the associated chronic inflammatory state. One common presentation of patients with advanced cystic fibrosis, the prototypical bronchiectatic lung disease, is hemoptysis, which, at times, can be life-threatening.

Pneumonias of any sort can cause hemoptysis. Tuberculous infection, which can lead to bronchiectasis or cavitary pneumonia, is a very common cause of hemoptysis worldwide. Community-acquired pneumonia and lung abscess can also result in bleeding. Once again, if the infection results in cavitation, there is a greater likelihood of bleeding due to erosion into blood vessels. Infections with *Staphylococcus aureus* and gram-negative rods (e.g., *Klebsiella pneumoniae*) are more likely to cause necrotizing lung infections and, thus, are more often associated with hemoptysis. Previous severe pneumonias can cause scarring and abnormal lung architecture, which may predispose a patient to hemoptysis with subsequent infections.

While it is not commonly seen in North America, pulmonary paragonimiasis (i.e., infection with the lung fluke *Paragonimus westermani*) often presents with fever, cough, and hemoptysis. This infection is a public health issue in Southeast Asia and China and is commonly confused with active tuberculosis, because the clinical pictures can be similar. Paragonimiasis should be considered in recent immigrants from endemic areas with new or recurrent hemoptysis. In addition, there are reports of pulmonary paragonimiasis in the United States secondary to ingestion of crayfish or small crabs.

Other causes of irritation of the airways resulting in hemoptysis include inhalation of toxic chemicals, thermal injury, direct trauma from suctioning of the airways (particularly in intubated patients), and irritation from inhalation of foreign bodies. All of these etiologies should be suggested by the individual patient's history and exposures.

Perhaps the most feared cause of hemoptysis is bronchogenic lung cancer, although hemoptysis is not a particularly common presenting symptom of this disease with only approximately 10% of patients having frank hemoptysis on initial assessment. Cancers arising in the proximal airways are much more likely to cause hemoptysis, although any malignancy in the chest can do so. Because both squamous cell carcinoma and small cell carcinoma are more commonly central and large at presentation, they are more often a cause of hemoptysis. These cancers can present with large-volume and life-threatening hemoptysis because of erosion into the hilar vessels. Carcinoid tumors, which are almost exclusively found as endobronchial lesions with friable mucosa, can also present with hemoptysis.

In addition to cancers arising in the lung, metastatic disease in the pulmonary parenchyma can also bleed. Malignancies that commonly metastasize to the lungs include renal cell, breast, colon, testicular, and thyroid cancers as well as melanoma. While they are not a common way for metastatic disease to present, multiple pulmonary nodules and hemoptysis should raise the suspicion for this etiology.

Finally, disease of the pulmonary vasculature can cause hemoptysis. Perhaps most commonly, congestive heart failure with transmission of elevated left atrial pressures, if severe enough, can lead to rupture of small alveolar capillaries. These patients rarely present with bright red blood but more commonly have pink, frothy sputum or blood-tinged secretions. Patients with a focal jet of mitral regurgitation can present with an upper-lobe infiltrate on chest radiograph together with hemoptysis. This is thought to be due to focal increases in pulmonary capillary pressure due to the regurgitant jet. Pulmonary arterio-venous malformations are prone to bleeding. Pulmonary embolism can also lead to the development of hemoptysis, which is generally associated with pulmonary infarction. Pulmonary arterial hypertension from other causes rarely results in hemoptysis.

Evaluation

As with most symptoms, the initial step in the evaluation of hemoptysis is a thorough history and physical examination (**Fig. 34-2**). As already mentioned, questioning should begin with determining if the bleeding is truly from the respiratory tract and not the nasopharynx or gastrointestinal tract, because these sources of bleeding require different evaluation and treatment approaches.

Figure 34-2

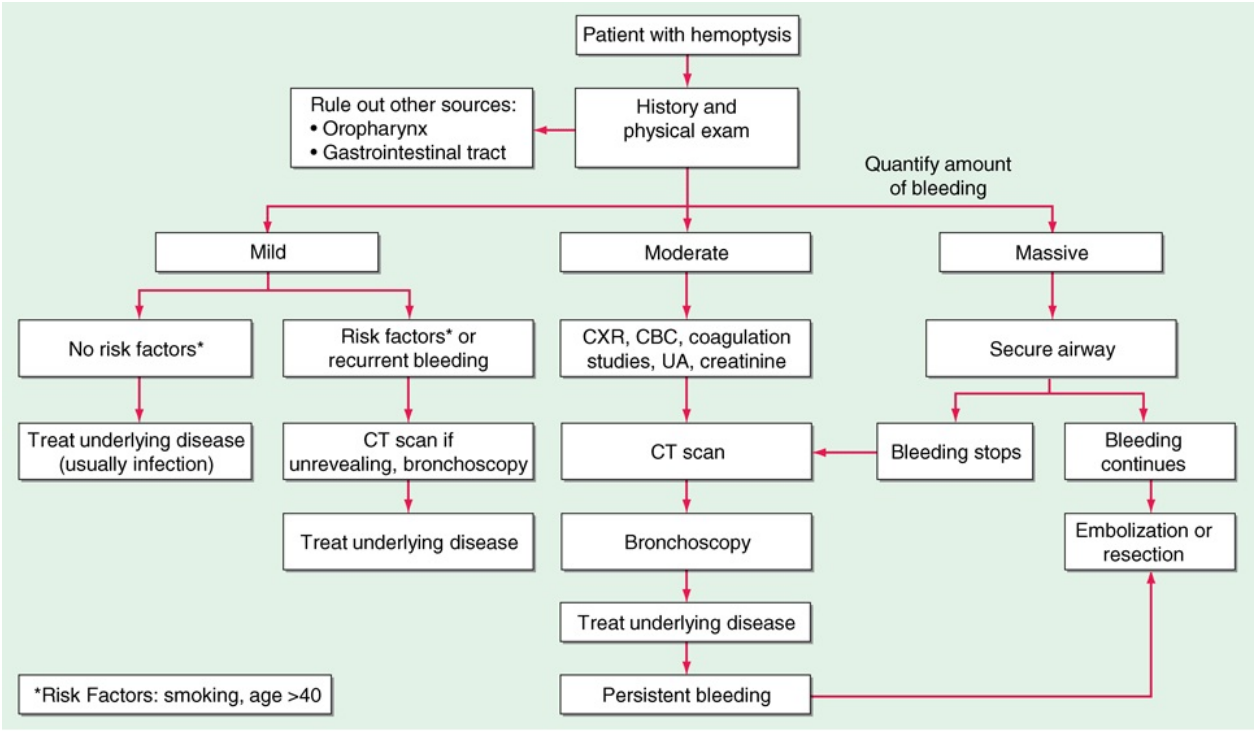

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flowchart—evaluation of hemoptysis. Decision tree for evaluation of hemoptysis. CBC, complete blood count; CT, computed tomography; CXR, chest x-ray; UA, urinalysis.

History and Physical Exam

The nature of the hemoptysis, whether they are blood-tinged, purulent secretions; pink, frothy sputum; or frank blood, may be helpful in determining an etiology. Specific triggers of the bleeding, such as recent inhalation exposures as well as any previous episodes of hemoptysis, should be elicited during history-taking. Monthly hemoptysis in a woman suggests catamenial hemoptysis from pulmonary endometriosis. The volume of the hemoptysis is also important not only in determining the cause, but in gauging the urgency for further diagnostic and therapeutic maneuvers. Patients rarely exsanguinate from hemoptysis but can effectively "drown" in aspirated blood. Large-volume hemoptysis, referred to as *massive hemoptysis*, is variably defined as hemoptysis of greater than 200–600 cc in 24 h. Massive hemoptysis should be considered a medical emergency. The medical urgency related to hemoptysis depends on both the amount of bleeding and the severity of underlying pulmonary disease.

All patients should be asked about current or former cigarette smoking; this behavior predisposes to both chronic bronchitis and increases the likelihood of bronchogenic cancer. Symptoms suggestive of respiratory tract infection—including fever, chills, and dyspnea—should be elicited. The practitioner should inquire about recent inhalation exposures or use of illicit substances as well as risk factors for venous thromboembolism.

Past medical history of malignancy or treatment thereof, rheumatologic disease, vascular disease, or underlying lung disease such as bronchiectasis may be relevant to the cause of hemoptysis. Because many of the causes of DAH can be part of a pulmonary-renal syndrome, specific inquiry into a history of renal insufficiency also is important.

The physical examination begins with an assessment of vital signs and oxygen saturation to gauge whether there is evidence of life-threatening bleeding. Tachycardia, hypotension, and decreased oxygen saturation should dictate a more expedited evaluation of hemoptysis. Specific focus on respiratory and cardiac examinations are important and should include inspection of the nares, auscultation of the lungs and heart, assessment of the lower extremities for symmetric or asymmetric edema, and evaluation for jugular venous distention. Clubbing of the digits may suggest underlying lung diseases such as bronchogenic carcinoma or bronchiectasis, which predispose to hemoptysis. Similarly, mucocutaneous telangiectasias should raise the specter of pulmonary arterial-venous malformations.

Diagnostic Evaluation

For most patients, the next step in evaluation of hemoptysis should be a standard chest radiograph. If a source of bleeding is not identified on plain film, a CT of the chest should be obtained. CT allows better delineation of bronchiectasis, alveolar filling, cavitary infiltrates, and masses than does chest x-ray; it also gives further information on mediastinal lymphadenopathy, which may support a diagnosis of thoracic malignancy. The practitioner should consider a CT protocol to assess for pulmonary embolism if the history or examination suggests venous thromboembolism as a cause of the bleeding.

Laboratory studies should include a complete blood count to assess both the hematocrit as well as platelet count and coagulation studies. Renal function and urinalysis should be assessed because of the possibility of pulmonary-renal syndromes presenting with hemoptysis. Acute renal insufficiency, or red blood cells or red blood cell casts on urinalysis should increase suspicion for small-vessel vasculitis, and studies such as antineutrophil cytoplasmic antibody (ANCA), antiglomerular basement membrane antibody (anti-GBM), and antinuclear antibody (ANA), should be considered. If a patient is producing sputum, Gram and acid-fast stains as well as culture should be obtained.

If all of these studies are unrevealing, bronchoscopy should be considered. In any patient with a history of cigarette smoking, airway inspection should be part of the evaluation of new hemoptysis. Because these patients are at increased risk of bronchogenic carcinoma, and endobronchial lesions are often not reliably visualized on computed tomogram, bronchoscopy should be seriously considered to add to the completeness of the evaluation.

Treatment: Hemoptysis

For the most part, the treatment of hemoptysis will vary based on its etiology. However, large-volume, life-threatening hemoptysis generally requires immediate intervention regardless of the cause. The first step is to establish a patent airway usually by endotracheal intubation and subsequent mechanical ventilation. As most large-volume hemoptysis arises from an airway lesion, it is ideal if the site of the bleeding can be identified either by chest imaging or bronchoscopy (more commonly rigid than flexible). The goal is then to isolate the bleeding to one lung and not allow the preserved airspaces in the other lung to be filled with blood, further impairing gas exchange. Patients should be placed with the bleeding lung in a dependent position (i.e., bleeding-side down) and, if possible, dual lumen endotracheal tubes or an airway blocker should be placed in the proximal airway of the bleeding lung. These interventions generally require the assistance of anesthesiologists, interventional pulmonologists, or thoracic surgeons.

If the bleeding does not stop with therapies of the underlying cause and passage of time, severe hemoptysis from bronchial arteries can be treated with angiographic embolization of the culprit bronchial artery. This intervention should only be entertained in the most severe and life-threatening cases of hemoptysis because there is a risk of unintentional spinal-artery embolization and consequent paraplegia with this procedure. Endobronchial lesions can be treated with a variety of bronchoscopically directed interventions, including cauterization and laser therapy. In extreme conditions, surgical resection of the affected region of lung is considered. Most cases of hemoptysis will resolve with treatment of the infection or inflammatory process or with removal of the offending stimulus.

FURTHER READINGS

Bidwell JL, Pachner RW: Hemoptysis: diagnosis and management. Am Fam Physician 72:1253, 2005[PMID: 16225028]

Chung KF, Pavord ID: Prevalence, pathogenesis, and causes of chronic cough. Lancet 37: 1364, 2008

Irwin RS et al: Diagnosis and management of cough executive summary: ACCP evidence-based clinical practice guidelines. Chest 129:1S, 2006

Jean-Baptiste E: Clinical assessment and management of massive hemoptysis. Crit Care Med 28:1642, 2000[PMID: 10834728]

Lara AR, Schwarz MI: Diffuse alveolar hemorrhage. Chest 137:1164, 2010[PMID: 20442117]

Pavord ID, Chung KF: Management of chronic cough. Lancet 371: 1375, 2008[PMID: 18424326]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HYPOXIA

The fundamental purpose of the cardiorespiratory system is to deliver O_2 and nutrients to cells and to remove CO_2 and other metabolic products from them. Proper maintenance of this function depends not only on intact cardiovascular and respiratory systems but also on an adequate number of red blood cells and hemoglobin and a supply of inspired gas containing adequate O_2.

Responses to Hypoxia

Decreased O_2 availability to cells results in an inhibition of oxidative phosphorylation and increased anaerobic glycolysis. This switch from aerobic to anaerobic metabolism, the Pasteur effect, maintains some, albeit reduced, adenosine 5'-triphosphate (ATP) production. In severe hypoxia, when ATP production is inadequate to meet the energy requirements of ionic and osmotic equilibrium, cell membrane depolarization leads to uncontrolled Ca^{2+} influx and activation of Ca^{2+}-dependent phospholipases and proteases. These events, in turn, cause cell swelling and, ultimately, cell death.

The adaptations to hypoxia are mediated, in part, by the upregulation of genes encoding a variety of proteins, including glycolytic enzymes such as phosphoglycerate kinase and phosphofructokinase, as well as the glucose transporters Glut-1 and Glut-2; and by growth factors, such as vascular endothelial growth factor (VEGF) and erythropoietin, which enhance erythrocyte production. The hypoxia-induced increase in expression of these key proteins is governed by the hypoxia-sensitive transcription factor, hypoxia-inducible factor-1 (HIF-1).

During hypoxia, systemic arterioles dilate, at least in part, by opening of K_{ATP} channels in vascular smooth-muscle cells due to the hypoxia-induced reduction in ATP concentration. By contrast, in pulmonary vascular smooth-muscle cells, inhibition of K^+ channels causes depolarization which, in turn, activates voltage-gated Ca^{2+} channels raising the cytosolic [Ca^{2+}] and causing smooth-muscle cell contraction. Hypoxia-induced pulmonary arterial constriction shunts blood away from poorly ventilated portions toward better ventilated portions of the lung; however, it also increases pulmonary vascular resistance and right ventricular afterload.

EFFECTS ON THE CENTRAL NERVOUS SYSTEM

Changes in the central nervous system (CNS), particularly the higher centers, are especially important consequences of hypoxia. Acute hypoxia causes impaired judgment, motor incoordination, and a clinical picture resembling acute alcohol intoxication. High-altitude illness is characterized by headache secondary to cerebral vasodilation, gastrointestinal symptoms, dizziness, insomnia, fatigue, or somnolence. Pulmonary arterial and sometimes venous constriction cause capillary leakage and high-altitude pulmonary edema (HAPE) (Chap. 33), which intensifies hypoxia, further promoting vasoconstriction. Rarely, high-altitude cerebral edema (HACE) develops, which is manifest by severe headache and papilledema and can cause coma. As hypoxia becomes more severe, the regulatory centers of the brainstem are affected, and death usually results from respiratory failure.

Causes of Hypoxia

RESPIRATORY HYPOXIA

When hypoxia occurs from respiratory failure, PaO_2 declines, and when respiratory failure is persistent, the hemoglobin-oxygen (Hb-O_2) dissociation curve (Fig. 104-2) is displaced to the right, with greater quantities of O_2 released at any level of tissue PO_2. Arterial hypoxemia, i.e., a reduction of O_2 saturation of arterial blood (SaO_2), and consequent cyanosis are likely to be more marked when such depression of PaO_2 results from pulmonary disease than when the depression occurs as the result of a decline in the fraction of oxygen in inspired air (FiO_2). In this latter situation, $PaCO_2$ falls secondary to anoxia-induced hyperventilation and the Hb-O_2 dissociation curve is displaced to the left, limiting the decline in SaO_2 at any level of PaO_2.

The most common cause of respiratory hypoxia is *ventilation-perfusion mismatch* resulting from perfusion of poorly ventilated alveoli. Respiratory hypoxemia may also be caused by *hypoventilation*, in which case it is then associated with an elevation of $PaCO_2$ (Chap. 252). These two forms of respiratory hypoxia are usually correctable by inspiring 100% O_2 for several minutes. A third cause of

respiratory hypoxia is shunting of blood across the lung from the pulmonary arterial to the venous bed (*intrapulmonary right-to-left shunting*) by perfusion of nonventilated portions of the lung, as in pulmonary atelectasis or through pulmonary arteriovenous connections. The low PaO_2 in this situation is only partially corrected by an FiO_2 of 100%.

HYPOXIA SECONDARY TO HIGH ALTITUDE

As one ascends rapidly to 3000 m (~10,000 ft), the reduction of the O_2 content of inspired air (FiO_2) leads to a decrease in alveolar PO_2 to approximately 60 mmHg, and a condition termed *high-altitude illness* develops (see above). At higher altitudes, arterial saturation declines rapidly and symptoms become more serious; and at 5000 m, unacclimated individuals usually cease to be able to function normally owing to the changes in CNS function described above.

HYPOXIA SECONDARY TO RIGHT-TO-LEFT EXTRAPULMONARY SHUNTING

From a physiologic viewpoint, this cause of hypoxia resembles intrapulmonary right-to-left shunting but is caused by congenital cardiac malformations, such as tetralogy of Fallot, transposition of the great arteries, and Eisenmenger's syndrome (Chap. 236). As in pulmonary right-to-left shunting, the PaO_2 cannot be restored to normal with inspiration of 100% O_2.

ANEMIC HYPOXIA

A reduction in hemoglobin concentration of the blood is accompanied by a corresponding decline in the O_2-carrying capacity of the blood. Although the PaO_2 is normal in anemic hypoxia, the absolute quantity of O_2 transported per unit volume of blood is diminished. As the anemic blood passes through the capillaries and the usual quantity of O_2 is removed from it, the PO_2 and saturation in the venous blood decline to a greater extent than normal.

CARBON MONOXIDE (CO) INTOXICATION

(See also Chap. e49) Hemoglobin that binds with CO (carboxyhemoglobin, COHb) is unavailable for O_2 transport. In addition, the presence of COHb shifts the Hb-O_2 dissociation curve to the left (Fig. 104-2) so that O_2 is unloaded only at lower tensions, contributing further to tissue hypoxia.

CIRCULATORY HYPOXIA

As in anemic hypoxia, the PaO_2 is usually normal, but venous and tissue PO_2 values are reduced as a consequence of reduced tissue perfusion and greater tissue O_2 extraction. This pathophysiology leads to an increased arterial-mixed venous O_2 difference (a-v-O_2 difference), or gradient. Generalized circulatory hypoxia occurs in heart failure (Chap. 234) and in most forms of shock (Chap. 270).

SPECIFIC ORGAN HYPOXIA

Localized circulatory hypoxia may occur as a result of decreased perfusion secondary to arterial obstruction, as in localized atherosclerosis in any vascular bed, or as a consequence of vasoconstriction, as observed in Raynaud's phenomenon (Chap. 249). Localized hypoxia may also result from venous obstruction and the resultant expansion of interstitial fluid causing arteriolar compression and, thereby, reduction of arterial inflow. Edema, which increases the distance through which O_2 must diffuse before it reaches cells, can also cause localized hypoxia. In an attempt to maintain adequate perfusion to more vital organs in patients with reduced cardiac output secondary to heart failure or hypovolemic shock, vasoconstriction may reduce perfusion in the limbs and skin, causing hypoxia of these regions.

INCREASED O_2 REQUIREMENTS

If the O_2 consumption of tissues is elevated without a corresponding increase in perfusion, tissue hypoxia ensues and the PO_2 in venous blood declines. Ordinarily, the clinical picture of patients with hypoxia due to an elevated metabolic rate, as in fever or thyrotoxicosis, is quite different from that in other types of hypoxia: the skin is warm and flushed owing to increased cutaneous blood flow that dissipates the excessive heat produced, and cyanosis is usually absent.

Exercise is a classic example of increased tissue O_2 requirements. These increased demands are normally met by several mechanisms operating simultaneously: (1) increase in the cardiac output and ventilation and, thus, O_2 delivery to the tissues; (2) a preferential shift in blood flow to the exercising muscles by changing vascular resistances in the circulatory beds of exercising tissues, directly and/or reflexly; (3) an increase in O_2 extraction from the delivered blood and a widening of the arteriovenous O_2 difference; and (4) a reduction in the pH of the tissues and capillary blood, shifting the Hb-O_2 curve to the right (Fig. 104-2), and unloading more O_2 from hemoglobin. If the capacity of these mechanisms is exceeded, then hypoxia, especially of the exercising muscles, will result.

IMPROPER OXYGEN UTILIZATION

Cyanide (Chap. e50) and several other similarly acting poisons cause cellular hypoxia. The tissues are unable to utilize O_2, and, as a consequence, the venous blood tends to have a high O_2 tension. This condition has been termed *histotoxic hypoxia*.

Adaptation to Hypoxia

An important component of the respiratory response to hypoxia originates in special chemosensitive cells in the carotid and aortic bodies and in the respiratory center in the brainstem. The stimulation of these cells by hypoxia increases ventilation, with a loss of CO_2, and can lead to respiratory alkalosis. When combined with the metabolic acidosis resulting from the production of lactic acid, the serum bicarbonate level declines (Chap. 47).

With the reduction of PaO_2, cerebrovascular resistance decreases and cerebral blood flow increases in an attempt to maintain O_2 delivery to the brain. However, when the reduction of PaO_2 is accompanied by hyperventilation and a reduction of $PaCO_2$, cerebrovascular resistance rises, cerebral blood flow falls, and tissue hypoxia intensifies.

The diffuse, systemic vasodilation that occurs in generalized hypoxia increases the cardiac output. In patients with underlying heart disease, the requirements of peripheral tissues for an increase of cardiac output with hypoxia may precipitate congestive heart failure. In patients with ischemic heart disease, a reduced PaO_2 may intensify myocardial ischemia and further impair left ventricular function.

One of the important compensatory mechanisms for chronic hypoxia is an increase in the hemoglobin concentration and in the number of red blood cells in the circulating blood, i.e., the development of polycythemia secondary to erythropoietin production (Chap. 108). In persons with chronic hypoxemia secondary to prolonged residence at a high altitude (>13,000 ft, 4200 m), a condition termed *chronic mountain sickness* develops. This disorder is characterized by a blunted respiratory drive, reduced ventilation, erythrocytosis, cyanosis, weakness, right ventricular enlargement secondary to pulmonary hypertension, and even stupor.

CYANOSIS

Cyanosis refers to a bluish color of the skin and mucous membranes resulting from an increased quantity of reduced hemoglobin (i.e., deoxygenated hemoglobin) or of hemoglobin derivatives (e.g., methemoglobin or sulfhemoglobin) in the small blood vessels of those tissues. It is usually most marked in the lips, nail beds, ears, and malar eminences. Cyanosis, especially if developed recently, is more commonly detected by a family member than the patient. The florid skin characteristic of polycythemia vera (Chap. 108) must be distinguished from the true cyanosis discussed here. A cherry-colored flush, rather than cyanosis, is caused by COHb (Chap. e50).

The degree of cyanosis is modified by the color of the cutaneous pigment and the thickness of the skin, as well as by the state of the cutaneous capillaries. The accurate clinical detection of the presence and degree of cyanosis is difficult, as proved by oximetric studies. In some instances, central cyanosis can be detected reliably when the SaO_2 has fallen to 85%; in others, particularly in dark-skinned persons, it may not be detected until it has declined to 75%. In the latter case, examination of the mucous membranes in the oral cavity and the conjunctivae rather than examination of the skin is more helpful in the detection of cyanosis.

The increase in the quantity of reduced hemoglobin in the mucocutaneous vessels that produces cyanosis may be brought about either by an increase in the quantity of venous blood as a result of dilation of the venules and venous ends of the capillaries or by a reduction in the SaO_2 in the capillary blood. In general, cyanosis becomes apparent when the concentration of reduced hemoglobin in capillary blood exceeds 40 g/L (4 g/dL).

It is the *absolute*, rather than the *relative*, quantity of reduced hemoglobin that is important in producing cyanosis. Thus, in a patient with severe anemia, the *relative* quantity of reduced hemoglobin in the venous blood may be very large when considered in relation to the total quantity of hemoglobin in the blood. However, since the concentration of the latter is markedly reduced, the *absolute* quantity of reduced hemoglobin may still be small, and, therefore, patients with severe anemia and even *marked* arterial desaturation may not display cyanosis. Conversely, the higher the total hemoglobin content, the greater the tendency toward cyanosis; thus, patients with marked polycythemia tend to be cyanotic at higher levels of SaO_2 than patients with normal hematocrit values. Likewise, local passive congestion, which causes an increase in the total quantity of reduced hemoglobin in the vessels in a given area, may cause cyanosis. Cyanosis is also observed when nonfunctional hemoglobin, such as methemoglobin or sulfhemoglobin (Chap. 104), is present in blood.

Cyanosis may be subdivided into central and peripheral types. In *central* cyanosis, the SaO_2 is reduced or an abnormal hemoglobin derivative is present, and the mucous membranes and skin are both affected. *Peripheral* cyanosis is due to a slowing of blood flow and abnormally great extraction of O_2 from normally saturated arterial blood; it results from vasoconstriction and diminished peripheral

blood flow, such as occurs in cold exposure, shock, congestive failure, and peripheral vascular disease. Often in these conditions, the mucous membranes of the oral cavity or those beneath the tongue may be spared. Clinical differentiation between central and peripheral cyanosis may not always be simple, and in conditions such as cardiogenic shock with pulmonary edema there may be a mixture of both types.

Differential Diagnosis

CENTRAL CYANOSIS

(Table 35-1) Decreased SaO_2 results from a marked reduction in the PaO_2. This reduction may be brought about by a decline in the FiO_2 without sufficient compensatory alveolar hyperventilation to maintain alveolar PO_2. Cyanosis usually becomes manifest in an ascent to an altitude of 4000 m (13,000 ft).

Table 35-1 Causes of Cyanosis

Central Cyanosis
Decreased arterial oxygen saturation
Decreased atmospheric pressure—high altitude
Impaired pulmonary function
Alveolar hypoventilation
Uneven relationships between pulmonary ventilation and perfusion (perfusion of hypoventilated alveoli)
Impaired oxygen diffusion
Anatomic shunts
Certain types of congenital heart disease
Pulmonary arteriovenous fistulas
Multiple small intrapulmonary shunts
Hemoglobin with low affinity for oxygen
Hemoglobin abnormalities
Methemoglobinemia—hereditary, acquired
Sulfhemoglobinema—acquired
Carboxyhemoglobinemia (not true cyanosis)
Peripheral Cyanosis
Reduced cardiac output
Cold exposure
Redistribution of blood flow from extremities
Arterial obstruction
Venous obstruction

Seriously *impaired pulmonary function*, through perfusion of unventilated or poorly ventilated areas of the lung or alveolar hypoventilation, is a common cause of central cyanosis (Chap. 252). This condition may occur acutely, as in extensive pneumonia or pulmonary edema, or chronically, with chronic pulmonary diseases (e.g., emphysema). In the latter situation, secondary polycythemia is generally present and clubbing of the fingers (see below) may occur. Another cause of reduced SaO_2 is *shunting of systemic venous blood into the arterial circuit* . Certain forms of congenital heart disease are associated with cyanosis on this basis (see above and Chap. 236).

Pulmonary arteriovenous fistulae may be congenital or acquired, solitary or multiple, microscopic or massive. The severity of cyanosis produced by these fistulae depends on their size and number. They occur with some frequency in hereditary hemorrhagic telangiectasia. SaO_2 reduction and cyanosis may also occur in some patients with cirrhosis, presumably as a consequence of pulmonary arteriovenous fistulae or portal vein–pulmonary vein anastomoses.

In patients with cardiac or pulmonary right-to-left shunts, the presence and severity of cyanosis depend on the size of the shunt relative to the systemic flow as well as on the Hb-O_2 saturation of the venous blood. With increased extraction of O_2 from the blood by the exercising muscles, the venous blood returning to the right side of the heart is more unsaturated than at rest, and shunting of this blood intensifies the cyanosis. Secondary polycythemia occurs frequently in patients in this setting and contributes to the cyanosis.

Cyanosis can be caused by small quantities of circulating methemoglobin (Hb Fe^{3+}) and by even smaller quantities of sulfhemoglobin (Chap. 104); both of these hemoglobin derivatives are unable to bind oxygen. Although they are uncommon causes of cyanosis, these abnormal hemoglobin species should be sought by spectroscopy when cyanosis is not readily explained by malfunction of the circulatory or respiratory systems. Generally, digital clubbing does not occur with them.

PERIPHERAL CYANOSIS

Probably the most common cause of peripheral cyanosis is the normal vasoconstriction resulting from exposure to cold air or water. When cardiac output is reduced, cutaneous vasoconstriction occurs as a compensatory mechanism so that blood is diverted from the skin to more vital areas such as the CNS and heart, and cyanosis of the extremities may result even though the arterial blood is normally saturated.

Arterial obstruction to an extremity, as with an embolus, or arteriolar constriction, as in cold-induced vasospasm (Raynaud's phenomenon) (Chap. 249), generally results in pallor and coldness, and there may be associated cyanosis. Venous obstruction, as in thrombophlebitis or deep venous thrombosis, dilates the subpapillary venous plexuses and thereby intensifies cyanosis.

Approach to the Patient: Cyanosis

Certain features are important in arriving at the cause of cyanosis:

1. It is important to ascertain the time of onset of cyanosis. Cyanosis present since birth or infancy is usually due to congenital heart disease.
2. Central and peripheral cyanosis must be differentiated. Evidence of disorders of the respiratory or cardiovascular systems are helpful. Massage or gentle warming of a cyanotic extremity will increase peripheral blood flow and abolish peripheral, but not central, cyanosis.
3. The presence or absence of clubbing of the digits (see below) should be ascertained. The combination of cyanosis and clubbing is frequent in patients with congenital heart disease and right-to-left shunting and is seen occasionally in patients with pulmonary disease, such as lung abscess or pulmonary arteriovenous fistulae. In contrast, peripheral cyanosis or acutely developing central cyanosis is *not* associated with clubbed digits.
4. PaO_2 and SaO_2 should be determined, and, in patients with cyanosis in whom the mechanism is obscure, spectroscopic examination of the blood performed to look for abnormal types of hemoglobin (critical in the differential diagnosis of cyanosis).

CLUBBING

The selective bulbous enlargement of the distal segments of the fingers and toes due to proliferation of connective tissue, particularly on the dorsal surface, is termed *clubbing*; there is also increased sponginess of the soft tissue at the base of the clubbed nail. Clubbing may be hereditary, idiopathic, or acquired and associated with a variety of disorders, including cyanotic congenital heart disease (see above), infective endocarditis, and a variety of pulmonary conditions (among them primary and metastatic lung cancer, bronchiectasis, asbestosis, sarcoidosis, lung abscess, cystic fibrosis, tuberculosis, and mesothelioma), as well as with some gastrointestinal diseases (including inflammatory bowel disease and hepatic cirrhosis). In some instances, it is occupational, e.g., in jackhammer operators.

Clubbing in patients with primary and metastatic lung cancer, mesothelioma, bronchiectasis, or hepatic cirrhosis may be associated with *hypertrophic osteoarthropathy*. In this condition, the subperiosteal formation of new bone in the distal diaphyses of the long bones of the extremities causes pain and symmetric arthritis-like changes in the shoulders, knees, ankles, wrists, and elbows. The diagnosis of hypertrophic osteoarthropathy may be confirmed by bone radiograph or MRI. Although the mechanism of clubbing is unclear, it appears to be secondary to humoral substances that cause dilation of the vessels of the distal digits as well as growth factors released from unfragmented platelet precursors in the digital circulation.

ACKNOWLEDGEMENTS

Dr. Eugene Braunwald authored this chapter in the previous edition.Some of the material from the 17th edition has been carried forward.

FURTHER READINGS

Fawcett RS et al: Nail abnormalities: Clues to systemic disease. Am Fam Physician 69:1417, 2004 [PMID: 15053406]

Giordano FJ: Oxygen, oxidative stress, hypoxia, and heart failure. J Clin Invest 115:500, 2005 [PMID: 15765131]

Griffey RT et al: Cyanosis. J Emerg Med 18:369, 2000 [PMID: 10729678]

Hackett PH, Roach RC: Current concepts: High altitude illness. N Engl J Med 345:107, 2001 [PMID: 11450659]

León-Velarde F et al: Chronic mountain sickness and the heart. Prog Cardiovasc Dis 52:540, 2010 [PMID: 21608089]

Levy MM: Pathophysiology of oxygen delivery in respiratory failure. Chest 128:547S, 2005

Michiels C: Physiological and pathological responses to hypoxia. Am J Pathol 164:1875, 2004 [PMID: 15161623]

Scherrer U et al: New insights in the pathogenesis of high-altitude pulmonary edema. Prog Cardiovasc Dis 52:485, 2010 [PMID: 20417341]

Semenza GL: Involvement of oxygen-sensing pathways in physiological and pathological erythropoiesis. Blood 114:2015, 2009 [PMID: 19494350]

Spicknall KE et al: Clubbing: an update on diagnosis, differential diagnosis, pathophysiology, and clinical relevance. J Am Acad Dermatol 52:1020, 2005 [PMID: 15928621]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EDEMA: INTRODUCTION

Edema is defined as a clinically apparent increase in the interstitial fluid volume, which may expand by several liters before the abnormality is evident. Therefore, a weight gain of several kilograms usually precedes overt manifestations of edema, and a similar weight loss from diuresis can be induced in a slightly edematous patient before "dry weight" is achieved. *Anasarca* refers to gross, generalized edema. *Ascites* (Chap. 43) and *hydrothorax* refer to accumulation of excess fluid in the peritoneal and pleural cavities, respectively, and are considered special forms of edema.

Depending on its cause and mechanism, edema may be localized or have a generalized distribution. Edema is recognized in its generalized form by puffiness of the face, which is most readily apparent in the periorbital areas, and by the persistence of an indentation of the skin after pressure; this is known as "pitting" edema. In its more subtle form, edema may be detected by noting that after the stethoscope is removed from the chest wall, the rim of the bell leaves an indentation on the skin of the chest for a few minutes. When the ring on a finger fits more snugly than in the past or when a patient complains of difficulty putting on shoes, particularly in the evening, edema may be present.

Pathogenesis

About one-third of total-body water is confined to the extracellular space. Approximately 75% of the latter is interstitial fluid, and the remainder is in the plasma compartment.

STARLING FORCES

The forces that regulate the disposition of fluid between these two components of the extracellular compartment frequently are referred to as the *Starling forces*. The hydrostatic pressure within the vascular system and the colloid oncotic pressure in the interstitial fluid tend to promote movement of fluid from the vascular to the extravascular space. By contrast, the colloid oncotic pressure contributed by plasma proteins and the hydrostatic pressure within the interstitial fluid promote the movement of fluid into the vascular compartment.

As a consequence of these forces, there is movement of water and diffusible solutes from the vascular space at the arteriolar end of the capillaries. Fluid is returned from the interstitial space into the vascular system at the venous end of the capillaries and by way of the lymphatics. Unless these channels are obstructed, lymph flow rises with increases in net movement of fluid from the vascular compartment to the interstitium. These flows are usually balanced so that there is a steady state in the sizes of the intravascular and interstitial compartments, yet a large exchange between them occurs. However, if either the hydrostatic or the oncotic pressure gradient is altered significantly, a further net movement of fluid between the two components of the extracellular space will take place. The development of edema then depends on one or more alterations in the Starling forces so that there is increased flow of fluid from the vascular system into the interstitium or into a body cavity.

Edema due to an increase in capillary pressure may result from an elevation of venous pressure caused by obstruction to venous and/or lymphatic drainage. An increase in capillary pressure may be generalized, as occurs in congestive heart failure (see below). The Starling forces also may be imbalanced when the colloid oncotic pressure of the plasma is reduced owing to any factor that may induce hypoalbuminemia, such as severe malnutrition, liver disease, loss of protein into the urine or into the gastrointestinal tract, or a severe catabolic state. Edema may be localized to one extremity when venous pressure is elevated due to unilateral thrombophlebitis (see below).

CAPILLARY DAMAGE

Edema may also result from damage to the capillary endothelium, which increases its permeability and permits the transfer of proteins into the interstitial compartment. Injury to the capillary wall can result from drugs, viral or bacterial agents, and thermal or mechanical trauma. Increased capillary permeability also may be a consequence of a hypersensitivity reaction and is characteristic of immune injury. Damage to the capillary endothelium is presumably responsible for inflammatory edema, which is usually nonpitting, localized, and accompanied by other signs of inflammation—i.e., erythema, heat, and tenderness.

REDUCTION OF EFFECTIVE ARTERIAL VOLUME

In many forms of edema, the effective arterial blood volume, a parameter that represents the filling of the arterial tree, is reduced. Underfilling of the arterial tree may be caused by a reduction of cardiac

output and/or systemic vascular resistance. As a consequence of underfilling, a series of physiologic responses designed to restore the effective arterial volume to normal are set into motion. A key element of these responses is the retention of salt and, therefore, of water, ultimately leading to edema.

RENAL FACTORS AND THE RENIN-ANGIOTENSIN-ALDOSTERONE (RAA) SYSTEM

(See also Chap. 342) In the final analysis, renal retention of Na^+ is central to the development of generalized edema (**Fig. 36-1**). The diminished renal blood flow characteristic of states in which the effective arterial blood volume is reduced is translated by the renal juxtaglomerular cells (specialized myoepithelial cells surrounding the afferent arteriole) into a signal for increased renin release. Renin is an enzyme with a molecular mass of about 40,000 Da that acts on its substrate, angiotensinogen, an α_2-globulin synthesized by the liver, to release angiotensin I, a decapeptide, which in turn is converted to angiotensin II (AII), an octapeptide. AII has generalized vasoconstrictor properties; it is especially active on the renal efferent arterioles. This action reduces the hydrostatic pressure in the peritubular capillaries, whereas the increased filtration fraction raises the colloid osmotic pressure in these vessels, thereby enhancing salt and water reabsorption in the proximal tubule as well as in the ascending limb of the loop of Henle.

Figure 36-1

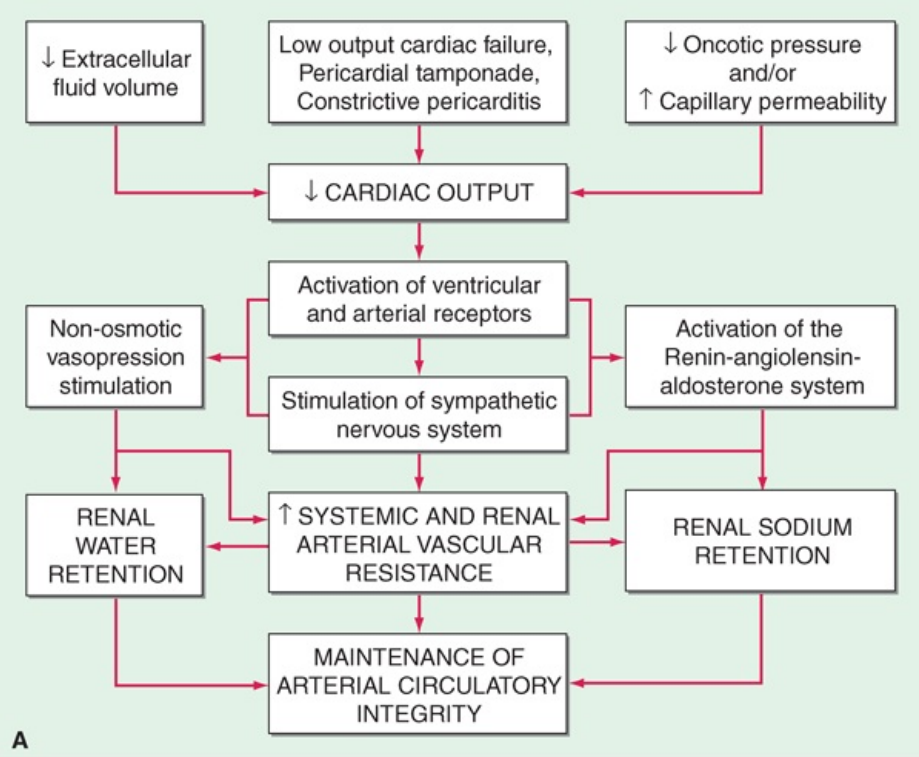

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

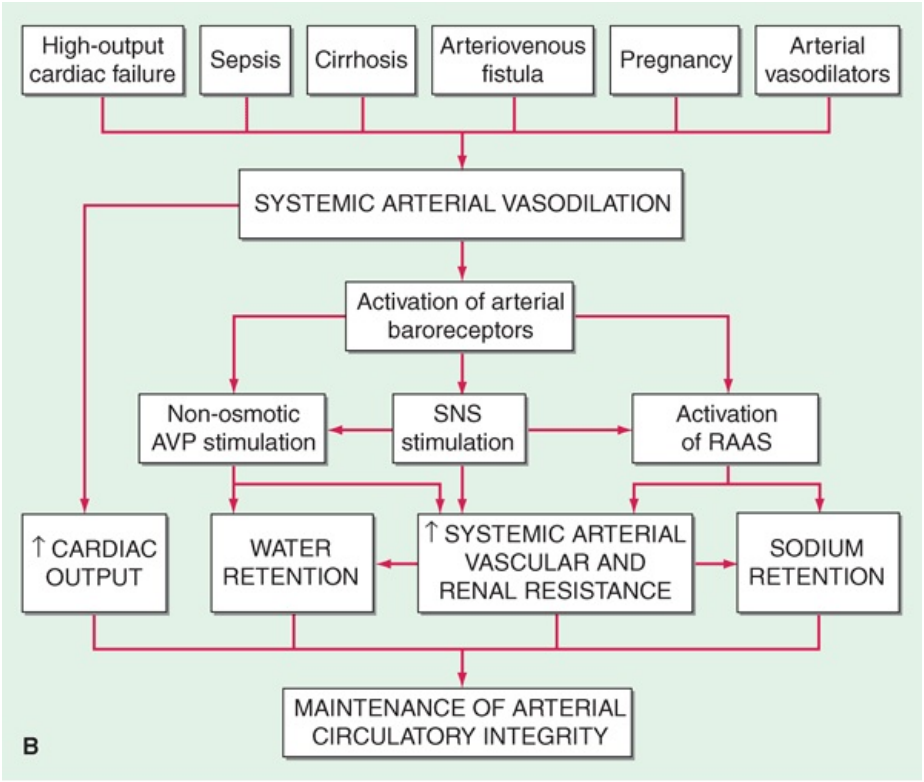

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical conditions in which a decrease in cardiac output (*A*) and systemic arterial vasodilation (*B*) cause arterial underfilling with resulting neurohumoral activation and renal sodium and water retention. In addition to activating the neurohumoral axis, adrenergic stimulation causes renal vasoconstriction and enhances sodium and fluid transport by the proximal tubuleepithelium. SNS, sympathetic nervous system; RAAS, renin-angiotensin aldosterone system. *(Reprinted from RW Schrier: Ann Intern Med 113:155, 1990.)*

The renin-angiotensin-aldosterone (RAA) system has long been recognized as a hormonal system; however, it also operates locally. Intrarenally produced AII contributes to glomerular efferent arteriolar constriction, and this "tubuloglomerular feedback" causes salt and water retention and thereby contributes to the formation of edema.

AII that enters the systemic circulation stimulates the production of aldosterone by the zona glomerulosa of the adrenal cortex. Aldosterone in turn enhances Na^+ reabsorption (and K^+ excretion) by the collecting tubule. In patients with heart failure, not only is aldosterone secretion elevated but the biologic half-life of aldosterone is prolonged, which increases further the plasma level of the hormone. A depression of hepatic blood flow, especially during exercise, is responsible for reduced hepatic catabolism of aldosterone.

Increased quantities of aldosterone are secreted in heart failure and in other edematous states, and blockade of the action of aldosterone by spironolactone or eplerenone (aldosterone antagonists) or by amiloride (a blocker of epithelial Na^+ channels) often induces a moderate diuresis in edematous states. Yet persistently augmented levels of aldosterone (or other mineralocorticoids) alone do not always promote accumulation of edema, as witnessed by the lack of significant fluid retention in most instances of primary aldosteronism (Chap. 342). Furthermore, although normal individuals retain some NaCl and water with the administration of potent mineralocorticoids, such as deoxycorticosterone acetate and fludrocortisone, this accumulation is self-limiting despite continued exposure to the steroid, a phenomenon known as *mineralocorticoid escape*. The failure of normal individuals who receive large doses of mineralocorticoids to accumulate large quantities of extracellular fluid and develop edema is probably a consequence of an increase in glomerular filtration rate (pressure natriuresis) and the action of natriuretic substance(s) (see below). The continued secretion of aldosterone may be more important in the accumulation of fluid in edematous states because patients with edema secondary to heart failure, nephrotic syndrome, and hepatic cirrhosis are generally unable to repair the deficit in effective arterial blood volume. As a consequence, they do not develop pressure natriuresis.

ARGININE VASOPRESSIN (AVP)

(See also Chap. 340) The secretion of AVP occurs in response to increased intracellular osmolar concentration, and, by stimulating V_2 receptors, AVP increases the reabsorption of free water in the renal distal tubule and collecting duct, thereby increasing total-body water. Circulating AVP is elevated

in many patients with heart failure secondary to a nonosmotic stimulus associated with decreased effective arterial volume. Such patients fail to show the normal reduction of AVP with a reduction of osmolality, contributing to edema formation and hyponatremia.

ENDOTHELIN

This potent peptide vasoconstrictor is released by endothelial cells. Its concentration is elevated in heart failure and contributes to renal vasoconstriction, Na^+ retention, and edema in heart failure.

NATRIURETIC PEPTIDES

Atrial distention and/or a Na^+ load cause release into the circulation of atrial natriuretic peptide (ANP), a polypeptide; a high-molecular-weight precursor of ANP is stored in secretory granules within atrial myocytes. Release of ANP causes (1) excretion of sodium and water by augmenting glomerular filtration rate, inhibiting sodium reabsorption in the proximal tubule, and inhibiting release of renin and aldosterone and (2) arteriolar and venous dilation by antagonizing the vasoconstrictor actions of AII, AVP, and sympathetic stimulation. Thus, ANP has the capacity to oppose Na^+ retention and arterial pressure elevation in hypervolemic states.

The closely related brain natriuretic peptide (BNP) is stored primarily in ventricular myocardium and is released when ventricular diastolic pressure rises. Its actions are similar to those of ANP, and both BNP and ANP bind to the natriuretic receptor-A, which is found in the myocardium. Yet another natriuretic peptide, C-type (CNP), is of endothelial and renal origin. CNP binds preferentially to the natriuretic peptide receptor-B, which is expressed principally in veins. Circulating levels of ANP and BNP are elevated in congestive heart failure and in cirrhosis with ascites, but obviously not sufficiently to prevent edema formation. In addition, in edematous states there is abnormal resistance to the actions of natriuretic peptides.

Clinical Causes of Edema

OBSTRUCTION OF VENOUS (AND LYMPHATIC) DRAINAGE OF A LIMB

In this condition the hydrostatic pressure in the capillary bed upstream (proximal) to the obstruction increases so that an abnormal quantity of fluid is transferred from the vascular to the interstitial space. Since the alternative route (i.e., the lymphatic channels) also may be obstructed or maximally filled, an increased volume of interstitial fluid in the limb develops (i.e., there is trapping of fluid in the interstitium of the extremity). The displacement of fluid into a limb may occur at the expense of the blood volume in the remainder of the body, thereby reducing effective arterial blood volume and leading to the retention of NaCl and H_2O until the deficit in plasma volume has been corrected.

CONGESTIVE HEART FAILURE

(See also Chap. 234) In this disorder the impaired systolic emptying of the ventricle(s) and/or the impairment of ventricular relaxation promotes an accumulation of blood in the venous circulation at the expense of the effective arterial volume, and the aforementioned sequence of events **(Fig. 36-1)** is initiated. In mild heart failure, a small increment of total blood volume may repair the deficit of arterial volume and establish a new steady state. Through the operation of Starling's law of the heart, an increase in ventricular diastolic volume promotes a more forceful contraction and may thereby maintain the cardiac output. However, if the cardiac disorder is more severe, fluid retention continues, and the increment in blood volume accumulates in the venous circulation, raising venous pressure and causing edema.

Incomplete ventricular emptying (systolic heart failure) and/or inadequate ventricular relaxation (diastolic heart failure) both lead to an elevation of ventricular diastolic pressure. If the impairment of cardiac function primarily involves the right ventricle, pressures in the systemic veins and capillaries rise, augmenting the transudation of fluid into the interstitial space and enhancing the likelihood of peripheral edema. The elevated systemic venous pressure is transmitted to the thoracic duct with consequent reduction of lymph drainage, further increasing the accumulation of edema. If the impairment of cardiac function involves the left ventricle primarily, pulmonary venous and capillary pressures rise. Pulmonary artery pressure rises, and this in turn interferes with the emptying of the right ventricle, leading to an elevation of right ventricular diastolic and central and systemic venous pressures, enhancing the likelihood of the formation of peripheral edema. The elevation of pulmonary capillary pressure may cause pulmonary edema, which impairs gas exchange. The resulting hypoxemia may impair cardiac function further, sometimes causing a vicious circle.

NEPHROTIC SYNDROME AND OTHER HYPOALBUMINEMIC STATES

(See also Chap. 283) The primary alteration in this disorder is a diminished colloid oncotic pressure due to losses of large quantities of protein into the urine. With severe hypoalbuminemia and the consequent reduced colloid osmotic pressure, the NaCl and H_2O that are retained cannot be restrained within the vascular compartment, and total and effective arterial blood volumes decline. This process initiates the edema-forming sequence of events described above, including activation of the RAA system. Impaired renal function contributes further to the formation of edema. A similar sequence

of events occurs in other conditions that lead to *severe* hypoalbuminemia, including (1) severe nutritional deficiency states, (2) severe, chronic liver disease (see below), and (3) protein-losing enteropathy.

CIRRHOSIS

(See also Chaps. 43 and 308) This condition is characterized in part by hepatic venous outflow blockade, which in turn expands the splanchnic blood volume and increases hepatic lymph formation. Intrahepatic hypertension acts as a stimulus for renal Na^+ retention and a reduction of effective arterial blood volume. These alterations frequently are complicated by hypoalbuminemia secondary to reduced hepatic synthesis, as well as systemic vasodilation. These effects reduce the effective arterial blood volume further, leading to activation of the RAA system, renal sympathetic nerves, and other NaCl- and H_2O-retaining mechanisms. The concentration of circulating aldosterone often is elevated by the failure of the liver to metabolize this hormone. Initially, the excess interstitial fluid is localized preferentially proximal (upstream) to the congested portal venous system and obstructed hepatic lymphatics, i.e., in the peritoneal cavity (ascites, Chap. 43). In later stages, particularly when there is severe hypoalbuminemia, peripheral edema may develop. The excess production of prostaglandins (PGE2 and PGI2) in cirrhosis attenuates renal Na^+ retention. When the synthesis of these substances is inhibited by nonsteroidal anti-inflammatory drugs (NSAIDs), renal function deteriorates and Na^+ retention increases.

DRUG-INDUCED EDEMA

A large number of widely used drugs can cause edema **(Table 36-1)**. Mechanisms include renal vasoconstriction (NSAIDs and cyclosporine), arteriolar dilation (vasodilators), augmented renal Na^+ reabsorption (steroid hormones), and capillary damage (interleukin 2).

Table 36-1 Drugs Associated with Edema Formation

Nonsteroidal anti-inflammatory drugs
Antihypertensive agents
Direct arterial/arteriolar vasodilators
Hydralazine
Clonidine
Methyldopa
Guanethidine
Minoxidil
Calcium channel antagonists
α-Adrenergic antagonists
Thiazolidinediones
Steroid hormones
Glucocorticoids
Anabolic steroids
Estrogens
Progestins
Cyclosporine
Growth hormone
Immunotherapies
Interleukin 2
OKT3 monoclonal antibody

Source: From Chertow.

DIFFERENTIAL DIAGNOSIS

LOCALIZED EDEMA

(See also Chap. 249) Localized edema due to venous or lymphatic obstruction may be caused by thrombophlebitis, chronic lymphangitis, resection of regional lymph nodes, filariasis, etc. Lymphedema is particularly intractable because restriction of lymphatic flow results in increased protein concentration in the interstitial fluid, a circumstance that aggravates retention of fluid.

Generalized Edema

The differences among the major causes of generalized edema are shown in **Table 36-2**.

Table 36-2 Principal Causes of Generalized Edema: History, Physical Examination, and Laboratory Findings

Organ System	History	Physical Examination	Laboratory Findings
Cardiac	Dyspnea with exertion prominent—often associated with orthopnea—or paroxysmal nocturnal dyspnea	Elevated jugular venous pressure, ventricular (S3) gallop; occasionally with displaced or dyskinetic apical pulse; peripheral cyanosis, cool extremities, small pulse pressure when severe	Elevated urea nitrogen-to-creatinine ratio common; elevated uric acid; serum sodium often diminished; liver enzymes occasionally elevated with hepatic congestion
Hepatic	Dyspnea uncommon, except if associated with significant degree of ascites; most often a history of ethanol abuse	Frequently associated with ascites; jugular venous pressure normal or low; blood pressure lower than in renal or cardiac disease; one or more additional signs of chronic liver disease (jaundice, palmar erythema, Dupuytren—s contracture, spider angiomata, male gynecomastia; asterixis and other signs of encephalopathy) may be present	If severe, reductions in serum albumin, cholesterol, other hepatic proteins (transferrin, fibrinogen); liver enzymes elevated, depending on the cause and acuity of liver injury; tendency toward hypokalemia, respiratory alkalosis; macrocytosis from folate deficiency
Renal (CRF)	Usually chronic: may be associated with uremic signs and symptoms, including decreased appetite, altered (metallic or fishy) taste, altered sleep pattern, difficulty concentrating, restless legs or myoclonus; dyspnea can be present, but generally less prominent than in heart failure	Elevated blood pressure; hypertensive retinopathy; nitrogenous fetor; pericardial friction rub in advanced cases with uremia	Albuminuria, hypoalbuminemia; sometimes, elevation of serum creatinine and urea nitrogen; hyperkalemia, metabolic acidosis, hyperphosphatemia, hypocalcemia, anemia (usually normocytic)
Renal (NS)	Childhood diabetes mellitus; plasma cell dyscrasias	Periorbital edema; hypertension	Proteinuria (3.5 g/d); hypoalbuminemia; hypercholesterolemia; microscopic hematuria

Abbreviations: CRF, chronic renal failure; NS, nephrotic syndrome.

Source: Modified from Chertow.

A majority of patients with generalized edema develop cardiac, renal, hepatic, or nutritional disorders. Consequently, the differential diagnosis of generalized edema should be directed toward identifying or excluding these several conditions.

EDEMA OF HEART FAILURE

(See also Chap. 234) The presence of heart disease, as manifested by cardiac enlargement and a gallop rhythm, together with evidence of cardiac failure, such as dyspnea, basilar rales, venous distention, and hepatomegaly, usually indicates that edema results from heart failure. Noninvasive tests such as echocardiography may be helpful in establishing the diagnosis of heart disease. The edema of heart failure typically occurs in the dependent portions of the body.

EDEMA OF ACUTE GLOMERULONEPHRITIS AND OTHER FORMS OF RENAL FAILURE

(See also Chap. 283) The edema that occurs during the acute phases of glomerulonephritis is characteristically associated with hematuria, proteinuria, and hypertension. Although some evidence supports the view that the fluid retention is due to increased capillary permeability, in most instances, the edema results from primary retention of NaCl and H_2O by the kidneys owing to renal insufficiency.

This state differs from congestive heart failure in that it is characterized by a normal (or sometimes even increased) cardiac output and a normal arterial–mixed venous oxygen difference. Patients with edema due to renal failure commonly have evidence of arterial hypertension as well as pulmonary congestion on chest roentgenogram even without cardiac enlargement, but they may not develop orthopnea. Patients with *chronic* renal failure may also develop edema due to primary renal retention of NaCl and H_2O.

EDEMA OF THE NEPHROTIC SYNDROME

(See also Chap. 283) Marked proteinuria (>3.5 g/d), hypoalbuminemia (<35 g/L), and, in some instances, hypercholesterolemia are present. This syndrome may occur during the course of a variety of kidney diseases, which include glomerulonephritis, diabetic glomerulosclerosis, and hypersensitivity reactions. A history of previous renal disease may or may not be elicited.

EDEMA OF CIRRHOSIS

(See also Chap. 308) Ascites and biochemical and clinical evidence of hepatic disease (collateral venous channels, jaundice, and spider angiomata) characterize edema of hepatic origin. The ascites (Chap. 43) is frequently refractory to treatment because it collects as a result of a combination of obstruction of hepatic lymphatic drainage, portal hypertension, and hypoalbuminemia. A sizable accumulation of ascitic fluid may increase intra-abdominal pressure and impede venous return from the lower extremities; hence, it tends to promote accumulation of edema in this region as well.

EDEMA OF NUTRITIONAL ORIGIN

A diet grossly deficient in protein over a prolonged period may produce hypoproteinemia and edema. The latter may be intensified by the development of beriberi heart disease, which also is of nutritional origin, in which multiple peripheral arteriovenous fistulas result in reduced effective systemic perfusion and effective arterial blood volume, thereby enhancing edema formation (Chap. 74). Edema may actually become intensified when famished subjects are first provided with an adequate diet. The ingestion of more food may increase the quantity of NaCl ingested, which is then retained along with H_2O. So-called refeeding edema also may be linked to increased release of insulin, which directly increases tubular Na^+ reabsorption. In addition to hypoalbuminemia, hypokalemia and caloric deficits may be involved in the edema of starvation.

OTHER CAUSES OF EDEMA

These causes include hypothyroidism (myxedema) and hyperthyroidism (pretibial myxedema secondary to Graves' disease), the edema in which is typically nonpitting and due to deposition of hyaluronic acid and, in Graves' disease, lymphocytic infiltration and inflammation; exogenous hyperadrenocortism; pregnancy; and administration of estrogens and vasodilators, particularly dihydropyridines such as nifedipine.

Distribution of Edema

The distribution of edema is an important guide to its cause. Thus, edema limited to one leg or to one or both arms is usually the result of venous and/or lymphatic obstruction. Edema resulting from hypoproteinemia characteristically is generalized, but it is especially evident in the very soft tissues of the eyelids and face and tends to be most pronounced in the morning because of the recumbent posture assumed during the night. Less common causes of facial edema include trichinosis, allergic reactions, and myxedema. Edema associated with heart failure, by contrast, tends to be more extensive in the legs and to be accentuated in the evening, a feature also determined largely by posture. When patients with heart failure have been confined to bed, edema may be most prominent in the presacral region. Paralysis reduces lymphatic and venous drainage on the affected side and may be responsible for unilateral edema.

Additional Factors in Diagnosis

The color, thickness, and sensitivity of the skin are significant. Local tenderness and warmth suggest inflammation. Local cyanosis may signify venous obstruction. In individuals who have had repeated episodes of prolonged edema, the skin over the involved areas may be thickened, indurated, and often red.

Estimation of the venous pressure is of importance in evaluating edema. Ordinarily, a significant generalized increase in venous pressure can be recognized by the level at which cervical veins collapse (Chap. 227). In patients with obstruction of the superior vena cava, edema is confined to the face, neck, and upper extremities, in which the venous pressure is elevated compared with that in the lower extremities. Severe heart failure may cause ascites that may be distinguished from the ascites caused by hepatic cirrhosis by the jugular venous pressure, which is usually elevated in heart failure and normal in cirrhosis.

Determination of the concentration of serum albuminaids importantly in identifying those patients in whom edema is due, at least in part, to diminished intravascular colloid oncotic pressure. The presence of proteinuria also affords useful clues. The absence of proteinuria excludes nephrotic syndrome but cannot exclude nonproteinuric causes of renal failure. Slight to moderate proteinuria is the rule in patients with heart failure.

Approach to the Patient: Edema

An important first question is whether the edema is localized or generalized. If it is localized, the local phenomena that may be responsible should be considered. If the edema is generalized, one should first determine if there is serious hypoalbuminemia, e.g., serum albumin <25 g/L. If so, the history, physical examination, urinalysis, and other laboratory data will help evaluate the question of cirrhosis, severe malnutrition, or the nephrotic syndrome as the underlying disorder. If hypoalbuminemia is not

present, one should determine if there is evidence of congestive heart failure severe enough to promote generalized edema. Finally, one should determine whether the patient has an adequate urine output or if there is significant oliguria or anuria. **These abnormalities are discussed in Chaps. 44, 279, and 280.**

FURTHER READINGS

Bansal S et al: Sodium retention in heart failure and cirrhosis: Potential role of natriuretic doses of mineralocorticoid antagonist? Circ Heart Fail 2:370, 2009 [PMID: 19808361]

Chertow GM: Approach to the patient with edema, in *Primary Cardiology*, 2nd ed, E Braunwald, L Goldman (eds). Philadelphia, Saunders, 2003, pp 117–128

Lee CYW, Burnett JC Jr: Natriuretic peptides and therapeutic application. Heart Fail Rev 12:131, 2007 [PMID: 17440808]

McCullough JC: Renal disorders and heart disease, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

Schrier RW: Decreased effective blood volume in edematous disorders: What does this mean? J Am Soc Nephrol 18:2028, 2007 [PMID: 17568020]

Skorecki KL et al: Extracellular fluid and edema formation, in *Brenner and Rector's The Kidney* , 8th ed. Philadelphia, Elsevier, 2008

Streeten DH: Idiopathic edema: Pathogenesis, clinical features, and treatment. Endocrinol Metab Clin North Am 24:531, 1995 [PMID: 8575408]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

The differential diagnosis of a heart murmur begins with a careful assessment of its major attributes and response to bedside maneuvers. The history, clinical context, and associated physical examination findings provide additional clues by which the significance of a heart murmur is established. Accurate bedside identification of a heart murmur can inform decisions regarding the indications for noninvasive testing and the need for referral to a cardiovascular specialist. Preliminary discussions can be held with the patient regarding antibiotic or rheumatic fever prophylaxis, the need to restrict various forms of physical activity, and the potential role for family screening.

Heart murmurs are caused by audible vibrations that are due to increased turbulence from accelerated blood flow through normal or abnormal orifices, flow through a narrowed or irregular orifice into a dilated vessel or chamber, or backward flow through an incompetent valve, ventricular septal defect, or patent ductus arteriosus. They traditionally are defined in terms of their timing within the cardiac cycle **(Fig. e13-1)**. *Systolic murmurs* begin with or after the first heart sound (S_1) and terminate at or before the component (A_2 or P_2) of the second heart sound (S_2) that corresponds to their site of origin (left or right, respectively). *Diastolic murmurs* begin with or after the associated component of S_2 and end at or before the subsequent S_1. *Continuous murmurs* are not confined to either phase of the cardiac cycle but instead begin in early systole and proceed through S_2 into all or part of diastole. The accurate timing of heart murmurs is the first step in their identification. The distinction between S_1 and S_2 and, therefore, systole and diastole is usually a straightforward process but can be difficult in the setting of a tachyarrhythmia, in which case the heart sounds can be distinguished by simultaneous palpation of the carotid upstroke, which should closely follow S_1.

Figure e13-1

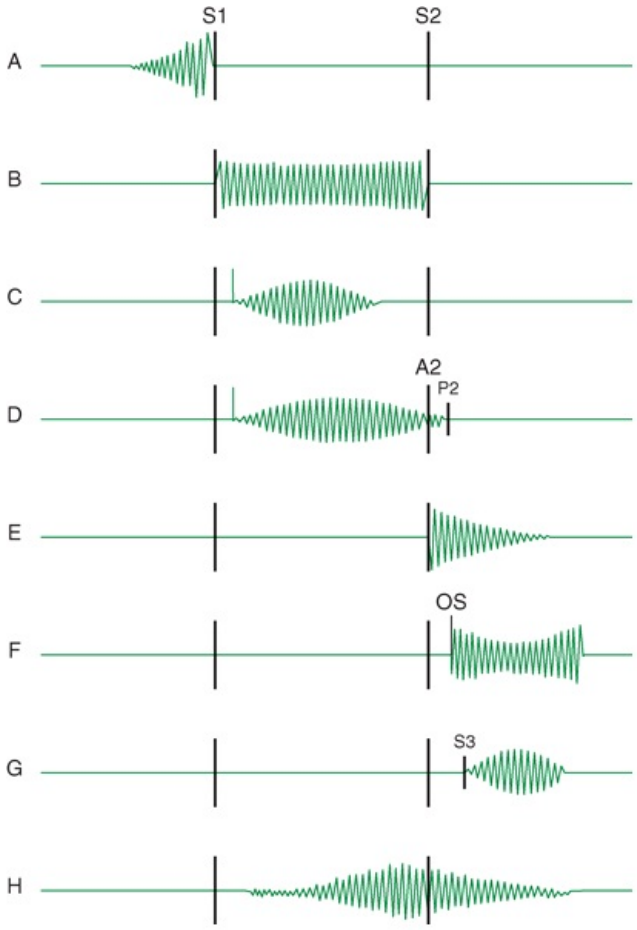

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram depicting principal heart murmurs. ***A.*** Presystolic murmur of mitral or tricuspid stenosis. ***B.*** Holosystolic (pansystolic) murmur of mitral or tricuspid regurgitation or of ventricular septal defect. ***C.*** Aortic ejection murmur beginning with an ejection click and fading before the second heart sound. ***D.*** Systolic murmur in pulmonic stenosis spilling through the aortic second sound, pulmonic valve closure being delayed. ***E.*** Aortic or pulmonary diastolic murmur. ***F.*** Long diastolic murmur of mitral stenosis after the opening snap (OS). ***G.*** Short mid-diastolic inflow murmur after a third heart sound. ***H.*** Continuous murmur of patent ductus arteriosus. *(Adapted from P Wood, Diseases of the Heart and Circulation, London, Eyre & Spottiswood, 1968. Permission granted courtesy of Antony and Julie Wood.)*

DURATION

The duration of a heart murmur depends on the length of time over which a pressure difference exists between two cardiac chambers, the left ventricle and the aorta, the right ventricle and the pulmonary artery, or the great vessels. The magnitude and variability of this pressure difference, coupled with the geometry and compliance of the involved chambers or vessels, dictate the velocity of flow; the degree of turbulence; and the resulting frequency, configuration, and intensity of the murmur. The diastolic murmur of chronic aortic regurgitation (AR) is a blowing, high-frequency event, whereas the murmur of mitral stenosis (MS), indicative of the left atrial–left ventricular diastolic pressure gradient, is a low-frequency event, heard as a rumbling sound with the bell of the stethoscope. The frequency components of a heart murmur may vary at different sites of auscultation. The coarse systolic murmur of aortic stenosis (AS) may sound higher-pitched and more acoustically pure at the apex, a phenomenon eponymously referred to as the *Gallavardin effect*. Some murmurs may have a distinct or unusual quality, such as the "honking" sound appreciated in some patients with mitral regurgitation (MR) due to mitral valve prolapse (MVP).

The configuration of a heart murmur may be described as crescendo, decrescendo, crescendo-decrescendo, or plateau. The decrescendo configuration of the murmur of chronic AR (Fig. e13-1*E*) can be understood in terms of the progressive decline in the diastolic pressure gradient between the aorta and the left ventricle. The crescendo-decrescendo configuration of the murmur of AS reflects the changes in the systolic pressure gradient between the left ventricle and the aorta as ejection occurs, whereas the plateau configuration of the murmur of chronic MR (Fig. e13-1*B*) is consistent with the large and nearly constant pressure difference between the left ventricle and the left atrium.

INTENSITY

The intensity of a heart murmur is graded on a scale of 1–6 (or I–VI). A grade 1 murmur is very soft and is heard only with great effort. A grade 2 murmur is easily heard but not particularly loud. A grade 3 murmur is loud but is not accompanied by a palpable thrill over the site of maximal intensity. A grade 4 murmur is very loud and is accompanied by a thrill. A grade 5 murmur is loud enough to be heard with only the edge of the stethoscope touching the chest, whereas a grade 6 murmur is loud enough to be heard with the stethoscope slightly off the chest. Murmurs of grade 3 or greater intensity usually signify important structural heart disease and indicate high blood flow velocity at the site of murmur production. Small ventricular septal defects (VSDs), for example, are accompanied by loud, usually grade 4 or greater, systolic murmurs as blood is ejected at high velocity from the left ventricle to the right ventricle. Low-velocity events, such as left-to-right shunting across an atrial septal defect (ASD), are usually silent. The intensity of a heart murmur also may be diminished by any process that increases the distance between the intracardiac source and the stethoscope on the chest wall, such as obesity, obstructive lung disease, and a large pericardial effusion. The intensity of a murmur also may be misleadingly soft when cardiac output is reduced significantly or when the pressure gradient between the involved cardiac structures is low.

LOCATION AND RADIATION

Recognition of the location and radiation of the murmur helps facilitate its accurate identification **(Fig. e13-2)**. Adventitious sounds, such as a systolic click or diastolic snap, or abnormalities of S_1 or S_2 may provide additional clues. Careful attention to the characteristics of the murmur and other heart sounds during the respiratory cycle and the performance of simple bedside maneuvers complete the auscultatory examination. These features, along with recommendations for further testing, are discussed below in the context of specific systolic, diastolic, and continuous heart murmurs **(Table e13-1)**.

Figure e13-2

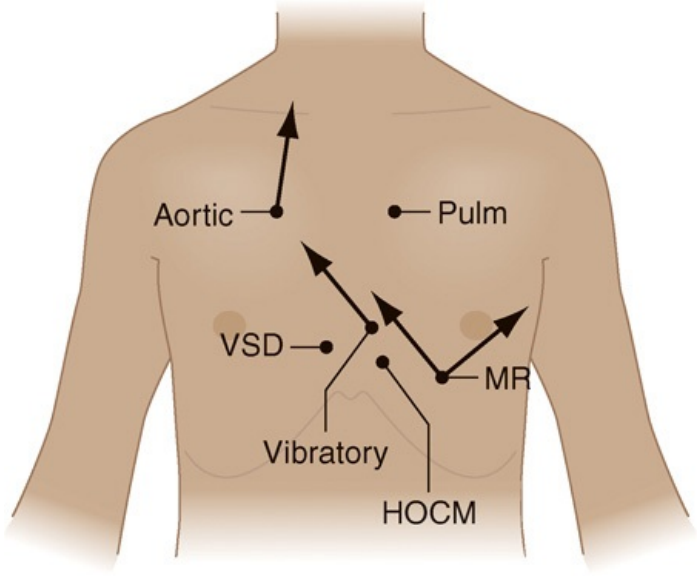

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Maximal intensity and radiation of six isolated systolic murmurs. HOCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; Pulm, pulmonary stenosis; Aortic, aortic stenosis; VSD, ventricular septal defect. (*From JB Barlow, Perspectives on the Mitral Valve. Philadelphia, FA Davis, 1987, p 140.)*

Table e13-1 Principal Causes of Heart Murmurs

Systolic Murmurs
Early systolic
Mitral
Acute MR
VSD
Muscular
Nonrestrictive with pulmonary hypertension
Tricuspid

TR with normal pulmonary artery pressure	
Mid-systolic	
Aortic	
Obstructive	
Supravalvular—supravalvular aortic stenosis, coarctation of the aorta	
Valvular—AS and aortic sclerosis	
Subvalvular—discrete, tunnel or HOCM	
Increased flow, hyperkinetic states, AR, complete heart block	
Dilation of ascending aorta, atheroma, aortitis	
Pulmonary	
Obstructive	
Supravalvular—pulmonary artery stenosis	
Valvular–pulmonic valve stenosis	
Subvalvular–infundibular stenosis (dynamic)	
Increased flow, hyperkinetic states, left-to-right shunt (e.g., ASD)	
Dilation of pulmonary artery	
Late systolic	
Mitral	
MVP, acute myocardial ischemia	
Tricuspid	
TVP	
Holosystolic	
Atrioventricular valve regurgitation (MR, TR)	
Left-to-right shunt at ventricular level (VSD)	
Early Diastolic Murmurs	
Aortic regurgitation	
Valvular: congenital (bicuspid valve), rheumatic deformity, endocarditis, prolapse, trauma, post-valvulotomy	
Dilation of valve ring: aorta dissection, annulo-aortic ectasia, cystic medial degeneration, hypertension, ankylosing spondylitis	
Widening of commissures: syphilis	
Pulmonic regurgitation	
Valvular: post-valvulotomy, endocarditis, rheumatic fever, carcinoid	
Dilation of valve ring: pulmonary hypertension; Marfan syndrome	
Congenital: isolated or associated with tetralogy of Fallot, VSD, pulmonic stenosis	
Mid-Diastolic Murmurs	
Mitral stenosis	
Mitral	
Carey-Coombs murmur (mid-diastolic apical murmur in acute rheumatic fever)	
Increased flow across nonstenotic mitral valve (e.g., MR, VSD, PDA, high-output states, and complete heart block)	
Tricuspid	
Tricuspid stenosis	
Increased flow across nonstenotic tricuspid valve (e.g., TR, ASD, and anomalous pulmonary venous return)	
Left and right atrial tumors (myxoma)	
Severe AR (Austin Flint murmur)	
Continuous Murmurs	
Patent ductus arteriosus	Proximal coronary artery stenosis
Coronary AV fistula	Mammary souffle of pregnancy
Ruptured sinus of Valsalva aneurysm	Pulmonary artery branch stenosis
Aortic septal defect	Bronchial collateral circulation
Cervical venous hum	Small (restrictive) ASD with MS
Anomalous left coronary artery	Intercostal AV fistula

Abbreviations: AR, aortic regurgitation; AS, aortic stenosis; ASD, atrial septal defect; AV, arteriovenous; HOCM, hypertrophic obstructive cardiomyopathy; MR, mitral regurgitation; MS, mitral stenosis; MVP, mitral valve prolapse; PDA, patent ductus arteriosus; TR, tricuspid regurgitation; TVP, tricuspid valve prolapse; VSD, ventricular septal defect.

Source: E Braunwald, JK Perloff, in D Zipes et al (eds): *Braunwald's Heart Disease*, 7th ed. Philadelphia, Elsevier, 2005; PJ Norton, RA O'Rourke, in E Braunwald, L Goldman (eds): *Primary Cardiology*, 2nd ed. Philadelphia, Elsevier, 2003.

Systolic Heart Murmurs

EARLY SYSTOLIC MURMURS

Early systolic murmurs begin with S_1 and extend for a variable period, ending well before S_2. Their causes are relatively few in number. *Acute, severe MR* into a normal-sized, relatively noncompliant left atrium results in an early, decrescendo systolic murmur best heard at or just medial to the apical impulse. These characteristics reflect the progressive attenuation of the pressure gradient between the left ventricle and the left atrium during systole owing to the rapid rise in left atrial pressure caused by the sudden volume load into an unprepared chamber and contrast sharply with the auscultatory features of chronic MR. Clinical settings in which acute, severe MR occur include (1) papillary muscle rupture complicating acute myocardial infarction (MI) (Chap. 245), (2) rupture of chordae tendineae in the setting of myxomatous mitral valve disease (MVP, Chap. 237), (3) infective endocarditis (Chap. 124), and (4) blunt chest wall trauma.

Acute, severe MR from papillary muscle rupture usually accompanies an inferior, posterior, or lateral MI and occurs 2–7 days after presentation. It often is signaled by chest pain, hypotension, and pulmonary edema, but a murmur may be absent in up to 50% of cases. The posteromedial papillary muscle is involved 6 to 10 times more frequently than the anterolateral papillary muscle. The murmur is to be distinguished from that associated with post-MI ventricular septal rupture, which is accompanied by a systolic thrill at the left sternal border in nearly all patients and is holosystolic in duration. A new heart murmur after an MI is an indication for transthoracic echocardiography (TTE) (Chap. 229), which allows bedside delineation of its etiology and pathophysiologic significance. The distinction between acute MR and ventricular septal rupture also can be achieved with right heart catheterization, sequential determination of oxygen saturations, and analysis of the pressure waveforms (tall *v* wave in the pulmonary artery wedge pressure in MR). Post-MI mechanical complications of this nature mandate aggressive medical stabilization and prompt referral for surgical repair.

Spontaneous chordal rupture can complicate the course of myxomatous mitral valve disease (MVP) and result in new-onset or "acute on chronic" severe MR. MVP may occur as an isolated phenomenon, or the lesion may be part of a more generalized connective tissue disorder as seen, for example, in patients with Marfan syndrome. Acute, severe MR as a consequence of infective endocarditis results from destruction of leaflet tissue, chordal rupture, or both. Blunt chest wall trauma is usually self-evident but may be disarmingly trivial; it can result in papillary muscle contusion and rupture, chordal detachment, or leaflet avulsion. TTE is indicated in all cases of suspected acute, severe MR to define its mechanism and severity, delineate left ventricular size and systolic function, and provide an assessment of suitability for primary valve repair.

A congenital, small muscular VSD (Chap. 236) may be associated with an early systolic murmur. The defect closes progressively during septal contraction, and thus, the murmur is confined to early systole. It is localized to the left sternal border (Fig. e13-2) and is usually of grade 4 or 5 intensity. Signs of pulmonary hypertension or left ventricular volume overload are absent. Anatomically large and uncorrected VSDs, which usually involve the membranous portion of the septum, may lead to pulmonary hypertension. The murmur associated with the left-to-right shunt, which earlier may have been holosystolic, becomes limited to the first portion of systole as the elevated pulmonary vascular resistance leads to an abrupt rise in right ventricular pressure and an attenuation of the interventricular pressure gradient during the remainder of the cardiac cycle. In such instances, signs of pulmonary hypertension (right ventricular lift, loud and single or closely split S_2) may predominate. The murmur is best heard along the left sternal border but is softer. Suspicion of a VSD is an indication for TTE.

Tricuspid regurgitation (TR) with normal pulmonary artery pressures, as may occur with infective endocarditis, may produce an early systolic murmur. The murmur is soft (grade 1 or 2), is best heard at the lower left sternal border, and may increase in intensity with inspiration (Carvallo's sign). Regurgitant "*c-v*" waves may be visible in the jugular venous pulse. TR in this setting is not associated with signs of right heart failure.

MID-SYSTOLIC MURMURS

Mid-systolic murmurs begin at a short interval after S_1, end before S_2 (Fig. e13-1*C*), and are usually crescendo-decrescendo in configuration. Aortic stenosis is the most common cause of a mid-systolic murmur in an adult. The murmur of AS is usually loudest to the right of the sternum in the second intercostal space (aortic area, Fig. e13-2) and radiates into the carotids. Transmission of the mid-

systolic murmur to the apex, where it becomes higher-pitched, is common (Gallavardin effect; see above).

Differentiation of this apical systolic murmur from MR can be difficult. The murmur of AS will increase in intensity, or become louder, in the beat after a premature beat, whereas the murmur of MR will have constant intensity from beat to beat. The intensity of the AS murmur also varies directly with the cardiac output. With a normal cardiac output, a systolic thrill and a grade 4 or higher murmur suggest severe AS. The murmur is softer in the setting of heart failure and low cardiac output. Other auscultatory findings of severe AS include a soft or absent A_2, paradoxical splitting of S_2, an apical S_4, and a late-peaking systolic murmur. In children, adolescents, and young adults with congenital valvular AS, an early ejection sound (click) is usually audible, more often along the left sternal border than at the base. Its presence signifies a flexible, noncalcified bicuspid valve (or one of its variants) and localizes the left ventricular outflow obstruction to the valvular (rather than sub- or supravalvular) level.

Assessment of the volume and rate of rise of the carotid pulse can provide additional information. A small and delayed upstroke (*parvus et tardus*) is consistent with severe AS. The carotid pulse examination is less discriminatory, however, in older patients with stiffened arteries. The electrocardiogram (ECG) shows signs of left ventricular hypertrophy (LVH) as the severity of the stenosis increases. TTE is indicated to assess the anatomic features of the aortic valve, the severity of the stenosis, left ventricular size, wall thickness and function, and the size and contour of the aortic root and proximal ascending aorta.

The obstructive form of hypertrophic cardiomyopathy (HOCM) is associated with a mid-systolic murmur that is usually loudest along the left sternal border or between the left lower sternal border and the apex (Chap. 238, Fig. e13-2). The murmur is produced by both dynamic left ventricular outflow tract obstruction and MR, and thus, its configuration is a hybrid between ejection and regurgitant phenomena. The intensity of the murmur may vary from beat to beat and after provocative maneuvers but usually does not exceed grade 3. The murmur classically will increase in intensity with maneuvers that result in increasing degrees of outflow tract obstruction, such as a reduction in preload or afterload (Valsalva, standing, vasodilators), or with an augmentation of contractility (inotropic stimulation). Maneuvers that increase preload (squatting, passive leg raising, volume administration) or afterload (squatting, vasopressors) or that reduce contractility (β-adrenoreceptor blockers) decrease the intensity of the murmur. In rare patients, there may be reversed splitting of S_2. A sustained left ventricular apical impulse and an S_4 may be appreciated. In contrast to AS, the carotid upstroke is rapid and of normal volume. Rarely, it is bisferiens or bifid in contour (see Fig. 227-2*D*) due to mid-systolic closure of the aortic valve. LVH is present on the ECG, and the diagnosis is confirmed by TTE. Although the systolic murmur associated with MVP behaves similarly to that due to HOCM in response to the Valsalva maneuver and to standing/squatting **(Fig. e13-3)**, these two lesions can be distinguished on the basis of their associated findings, such as the presence of LVH in HOCM or a nonejection click in MVP.

Figure e13-3

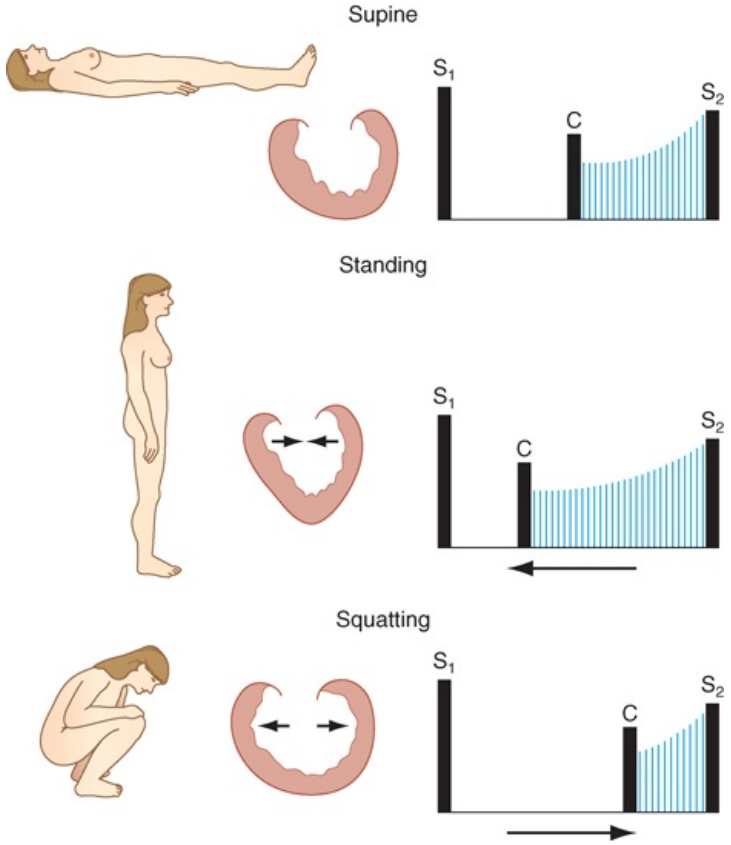

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A mid-systolic nonejection sound (C) occurs in mitral valve prolapse and is followed by a late systolic murmur that crescendos to the second heart sound (S_2). Standing decreases venous return; the heart becomes smaller; C moves closer to the first heart sound (S_1), and the mitral regurgitant murmur has an earlier onset. With prompt squatting, venous return increases; the heart becomes larger; C moves toward S_2, and the duration of the murmur shortens. *(From JA Shaver, JJ Leonard, DF Leon, Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 13. Copyright, American Heart Association.)*

The mid-systolic, crescendo-decrescendo murmur of congenital pulmonic stenosis (PS, Chap. 236) is best appreciated in the second and third left intercostal spaces (pulmonic area) (**Figs.**e13-2 and **e13-4**). The duration of the murmur lengthens and the intensity of P_2 diminishes with increasing degrees of valvular stenosis (Fig. e13-1*D*). An early ejection sound, the intensity of which decreases with inspiration, is heard in younger patients. A parasternal lift and ECG evidence of right ventricular hypertrophy indicate severe pressure overload. If obtained, the chest x-ray may show poststenotic dilation of the main pulmonary artery. TTE is recommended for complete characterization.

Figure e13-4

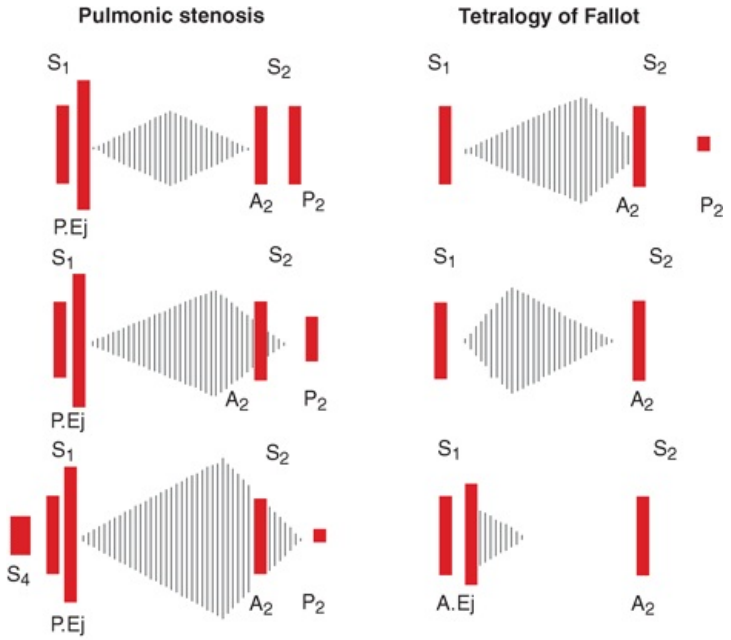

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left. In valvular pulmonic stenosis with intact ventricular septum, right ventricular systolic ejection becomes progressively longer, with increasing obstruction to flow. As a result, the murmur becomes longer and louder, enveloping the aortic component of the second heart sound (A_2). The pulmonic component (P_2) occurs later, and splitting becomes wider but more difficult to hear because A_2 is lost in the murmur and P_2 becomes progressively fainter and lower-pitched. As the pulmonic gradient increases, isometric contraction shortens until the pulmonic valve ejection sound fuses with the first heart sound (S_1). In severe pulmonic stenosis with concentric hypertrophy and decreasing right ventricular compliance, a fourth heart sound appears. ***Right.*** In tetralogy of Fallot with increasing obstruction at the pulmonic infundibular area, an increasing amount of right ventricular blood is shunted across the silent ventricular septal defect and flow across the obstructed outflow tract decreases. Therefore, with increasing obstruction the murmur becomes shorter, earlier, and fainter. P_2 is absent in severe tetralogy of Fallot. A large aortic root receives almost all cardiac output from both ventricular chambers, and the aorta dilates and is accompanied by a root ejection sound that does not vary with respiration. *(From JA Shaver, JJ Leonard, DF Leon, Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 45. Copyright, American Heart Association.)*

Significant left-to-right intracardiac shunting due to an ASD (Chap. 236) leads to an increase in pulmonary blood flow and a grade 2–3 mid-systolic murmur at the middle to upper left sternal border attributed to increased flow rates across the pulmonic valve with fixed splitting of S_2. Ostium secundum ASDs are the most common cause of these shunts in adults. Features suggestive of a primum ASD include the coexistence of MR due to a cleft anterior mitral valve leaflet and left axis deviation of the QRS complex on the ECG. With sinus venosus ASDs, the left-to-right shunt is usually not large enough to result in a systolic murmur, although the ECG may show abnormalities of sinus node function. A grade 2 or 3 mid-systolic murmur may also be heard best at the upper left sternal border in patients with idiopathic dilation of the pulmonary artery; a pulmonary ejection sound is also present in these patients. TTE is indicated to evaluate a grade 2 or 3 mid-systolic murmur when there are other signs of cardiac disease.

An isolated grade 1 or 2 midsystolic murmur, heard in the absence of symptoms or signs of heart disease, is most often a benign finding for which no further evaluation, including TTE, is necessary. The most common example of a murmur of this type in an older adult patient is the crescendo-decrescendo murmur of aortic valve sclerosis, heard at the second right interspace (Fig. e13-2). Aortic sclerosis is defined as focal thickening and calcification of the aortic valve to a degree that does not interfere with leaflet opening. The carotid upstrokes are normal, and electrocardiographic LVH is not present. A grade 1 or 2 mid-systolic murmur often can be heard at the left sternal border with pregnancy, hyperthyroidism, or anemia, physiologic states that are associated with accelerated blood flow. *Still's murmur* refers to a benign grade 2, vibratory mid-systolic murmur at the lower left sternal border in normal children and adolescents (Fig. e13-2).

LATE SYSTOLIC MURMURS

A late systolic murmur that is best heard at the left ventricular apex is usually due to MVP (Chap. 237). Often, this murmur is introduced by one or more nonejection clicks. The radiation of the murmur can

help identify the specific mitral leaflet involved in the process of prolapse or flail. The term *flail* refers to the movement made by an unsupported portion of the leaflet after loss of its chordal attachment(s). With posterior leaflet prolapse or flail, the resultant jet of MR is directed anteriorly and medially, as a result of which the murmur radiates to the base of the heart and masquerades as AS. Anterior leaflet prolapse or flail results in a posteriorly directed MR jet that radiates to the axilla or left infrascapular region. Leaflet flail is associated with a murmur of grade 3 or 4 intensity that can be heard throughout the precordium in thin-chested patients. The presence of an S_3 or a short, rumbling mid-diastolic murmur due to enhanced flow signifies severe MR.

Bedside maneuvers that decrease left ventricular preload, such as standing, will cause the click and murmur of MVP to move closer to the first heart sound, as leaflet prolapse occurs earlier in systole. Standing also causes the murmur to become louder and longer. With squatting, left ventricular preload and afterload are increased abruptly, leading to an increase in left ventricular volume, and the click and murmur move away from the first heart sound as leaflet prolapse is delayed; the murmur becomes softer and shorter in duration (Fig. e13-3). As noted above, these responses to standing and squatting are directionally similar to those observed in patients with HOCM.

A late, apical systolic murmur indicative of MR may be heard transiently in the setting of acute myocardial ischemia; it is due to apical tethering and malcoaptation of the leaflets in response to structural and functional changes of the ventricle and mitral annulus. The intensity of the murmur varies as a function of left ventricular afterload and will increase in the setting of hypertension. TTE is recommended for assessment of late systolic murmurs.

HOLOSYSTOLIC MURMURS

(**Figs. e13-1B** and **e13-5**) Holosystolic murmurs begin with S_1 and continue through systole to S_2. They are usually indicative of chronic mitral or tricuspid valve regurgitation or a VSD and warrant TTE for further characterization. The holosystolic murmur of chronic MR is best heard at the left ventricular apex and radiates to the axilla (Fig. e13-2); it is usually high-pitched and plateau in configuration because of the wide difference between left ventricular and left atrial pressure throughout systole. In contrast to acute MR, left atrial compliance is normal or even increased in chronic MR. As a result, there is only a small increase in left atrial pressure for any increase in regurgitant volume.

Figure e13-5

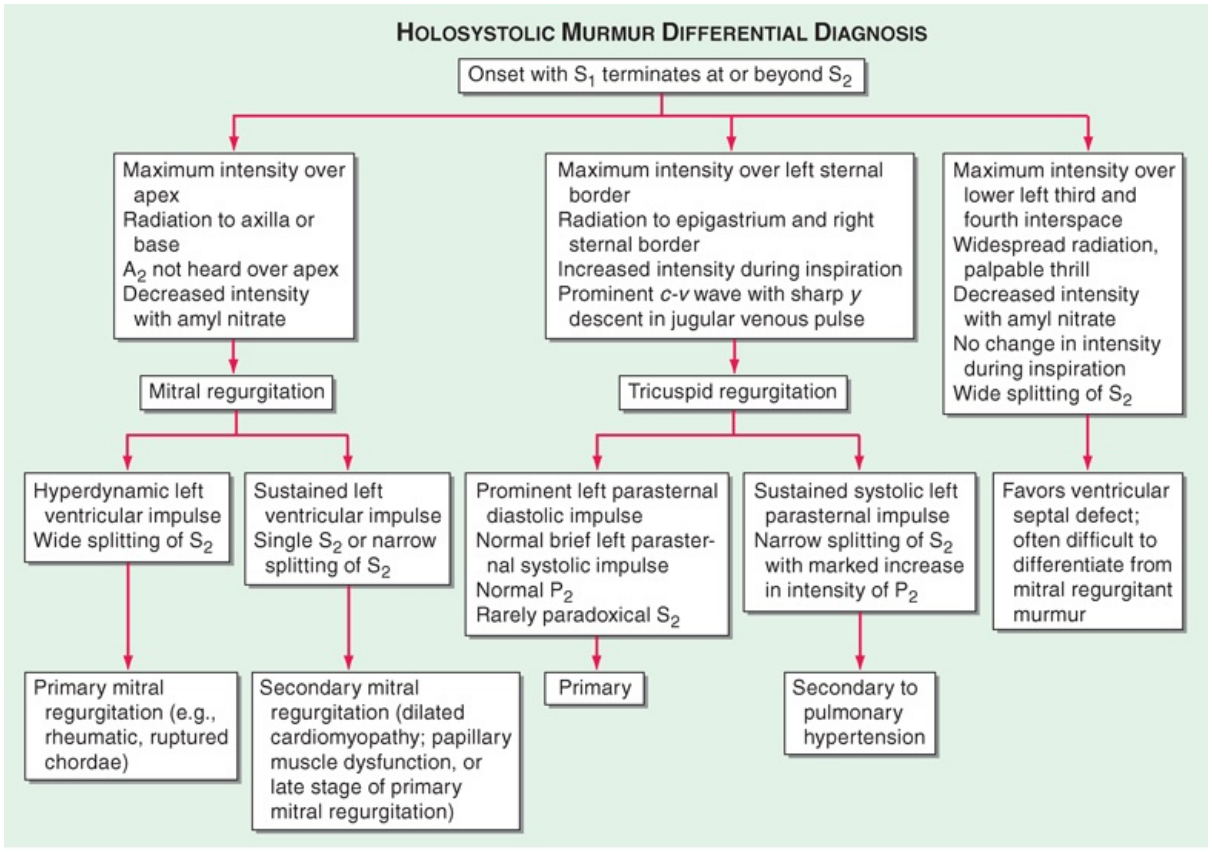

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Differential diagnosis of a holosystolic murmur.

Several conditions are associated with chronic MR and an apical holosystolic murmur, including rheumatic scarring of the leaflets, mitral annular calcification, postinfarction left ventricular remodeling, and severe left ventricular chamber enlargement. The circumference of the mitral annulus increases as the left ventricle enlarges and leads to failure of leaflet coaptation with central MR in patients with dilated cardiomyopathy (Chap. 238). The severity of the MR is worsened by any contribution from apical displacement of the papillary muscles and leaflet tethering (remodeling). Because the mitral annulus is contiguous with the left atrial endocardium, gradual enlargement of the left atrium from chronic MR will result in further stretching of the annulus and more MR; thus, "MR begets MR." Chronic severe MR results in enlargement and leftward displacement of the left ventricular apex beat and, in some patients, a diastolic filling complex, as described previously.

The holosystolic murmur of chronic TR is generally softer than that of MR, is loudest at the left lower sternal border, and usually increases in intensity with inspiration (Carvallo's sign). Associated signs include *c-v* waves in the jugular venous pulse, an enlarged and pulsatile liver, ascites, and peripheral edema. The abnormal jugular venous waveforms are the predominant finding and are seen very often in the absence of an audible murmur despite Doppler echocardiographic verification of TR. Causes of primary TR include myxomatous disease (prolapse), endocarditis, rheumatic disease, carcinoid, Ebstein's anomaly, and chordal detachment after the performance of right ventricular endomyocardial biopsy. TR is more commonly a passive process that results secondarily from chronic elevations of pulmonary artery and right ventricular pressures, leading to right ventricular enlargement, annular dilation, papillary muscle displacement, and failure of leaflet coaptation.

The holosystolic murmur of a VSD is loudest at the mid- to lower left sternal border (Fig. e13-2) and radiates widely. A thrill is present at the site of maximal intensity in the majority of patients. There is no change in the intensity of the murmur with inspiration. The intensity of the murmur varies as a function of the anatomic size of the defect. Small, restrictive VSDs, as exemplified by the *maladie de Roger*, create a very loud murmur due to the significant and sustained systolic pressure gradient between the left and right ventricles. With large defects, the ventricular pressures tend to equalize, shunt flow is balanced, and a murmur is not appreciated. The distinction between post-MI ventricular septal rupture and MR has been reviewed previously.

Diastolic Heart Murmurs

EARLY DIASTOLIC MURMURS

(Fig. e13-1*E*) Chronic AR results in a high-pitched, blowing, decrescendo, early to mid-diastolic murmur that begins after the aortic component of S_2 (A_2) and is best heard at the second right interspace **(Fig. e13-6)**. The murmur may be soft and difficult to hear unless auscultation is performed with the patient leaning forward at end expiration. This maneuver brings the aortic root closer to the anterior chest wall. Radiation of the murmur may provide a clue to the cause of the AR. With primary valve disease, such as that due to congenital bicuspid disease, prolapse, or endocarditis, the diastolic murmur tends to radiate along the left sternal border, where it is often louder than appreciated in the second right interspace. When AR is caused by aortic root disease, the diastolic murmur may radiate along the right sternal border. Diseases of the aortic root cause dilation or distortion of the aortic annulus and failure of leaflet coaptation. Causes include Marfan syndrome with aneurysm formation, annulo-aortic ectasia, ankylosing spondylitis, and aortic dissection.

Figure e13-6

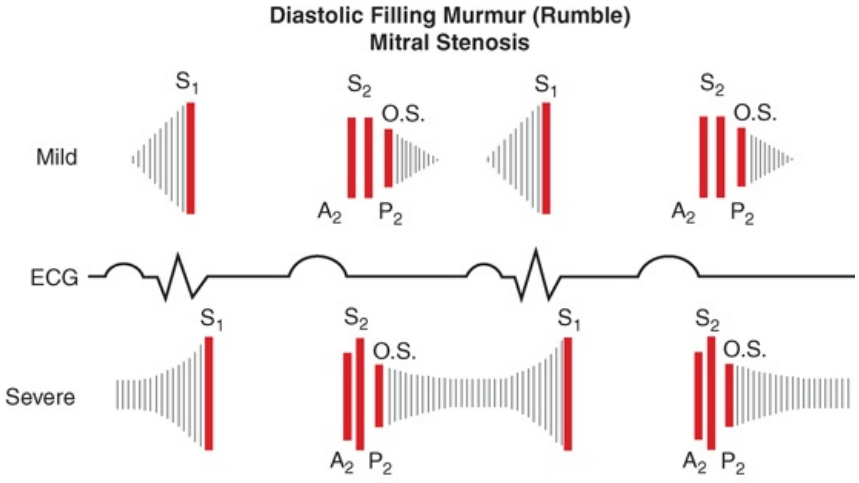

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diastolic filling murmur (rumble) in mitral stenosis. In mild mitral stenosis, the diastolic gradient

Diastolic filling murmur (rumble) in mitral stenosis. In mild mitral stenosis, the diastolic gradient across the valve is limited to the phases of rapid ventricular filling in early diastole and presystole. The rumble may occur during either or both periods. As the stenotic process becomes severe, a large pressure gradient exists across the valve during the entire diastolic filling period, and the rumble persists throughout diastole. As the left atrial pressure becomes greater, the interval between A_2 (or P_2) and the opening snap (O.S.) shortens. In severe mitral stenosis, secondary pulmonary hypertension develops and results in a loud P_2 and the splitting interval usually narrows. ECG, electrocardiogram. *(From JA Shaver, JJ Leonard, DF Leon, Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 55. Copyright, American Heart Association.)*

Chronic, severe AR also may produce a lower-pitched mid- to late, grade 1 or 2 diastolic murmur at the apex (Austin Flint murmur), which is thought to reflect turbulence at the mitral inflow area from the admixture of regurgitant (aortic) and forward (mitral) blood flow (Fig. e13-1*G*). This lower-pitched, apical diastolic murmur can be distinguished from that due to MS by the absence of an opening snap and the response of the murmur to a vasodilator challenge. Lowering afterload with an agent such as amyl nitrite will decrease the duration and magnitude of the aortic–left ventricular diastolic pressure gradient, and thus, the Austin Flint murmur of severe AR will become shorter and softer. The intensity of the diastolic murmur of mitral stenosis **(Fig. e13-6)** may either remain constant or increase with afterload reduction because of the reflex increase in cardiac output and mitral valve flow.

Although AS and AR may coexist, a grade 2 or 3 crescendo-decrescendo mid-systolic murmur frequently is heard at the base of the heart in patients with isolated, severe AR and is due to an increased volume and rate of systolic flow. Accurate bedside identification of coexistent AS can be difficult unless the carotid pulse examination is abnormal or the mid-systolic murmur is of grade 4 or greater intensity. In the absence of heart failure, chronic severe AR is accompanied by several peripheral signs of significant diastolic run-off, including a wide pulse pressure, a "water-hammer" carotid upstroke (Corrigan's pulse), and Quincke's pulsations of the nail beds. The diastolic murmur of *acute, severe AR* is notably shorter in duration and lower-pitched than the murmur of chronic AR. It can be very difficult to appreciate in the presence of a rapid heart rate. These attributes reflect the abrupt rate of rise of diastolic pressure within the unprepared and noncompliant left ventricle and the correspondingly rapid decline in the aortic–left ventricular diastolic pressure gradient. Left ventricular diastolic pressure may increase sufficiently to result in premature closure of the mitral valve and a soft first heart sound. Peripheral signs of significant diastolic run-off are not present.

Pulmonic regurgitation (PR) results in a decrescendo, early to mid-diastolic murmur (*Graham Steell murmur*) that begins after the pulmonic component of S_2 (P_2), is best heard at the second left interspace, and radiates along the left sternal border. The intensity of the murmur may increase with inspiration. PR is most commonly due to dilation of the valve annulus from chronic elevation of the pulmonary artery pressure. Signs of pulmonary hypertension, including a right ventricular lift and a loud single or narrowly split S_2, are present. These features also help distinguish PR from AR as the cause of a decrescendo diastolic murmur heard along the left sternal border. PR in the absence of pulmonary hypertension can occur with endocarditis or a congenitally deformed valve. It is usually present after repair of tetralogy of Fallot in childhood. When pulmonary hypertension is not present, the diastolic murmur is softer and lower-pitched than the classic Graham Steell murmur, and the severity of the PR can be difficult to appreciate.

TTE is indicated for the further evaluation of a patient with an early to mid-diastolic murmur. Longitudinal assessment of the severity of the valve lesion and ventricular size and systolic function help guide a potential decision for surgical management. TTE also can provide anatomic information regarding the root and proximal ascending aorta, although computed tomographic or magnetic resonance angiography may be indicated for more precise characterization (Chap. 229).

MID-DIASTOLIC MURMURS

(Figs. e13-1*G* and e13-1*H*) Mid-diastolic murmurs result from obstruction and/or augmented flow at the level of the mitral or tricuspid valve. Rheumatic fever is the most common cause of MS (Fig. e13-6). In younger patients with pliable valves, S_1 is loud and the murmur begins after an opening snap, which is a high-pitched sound that occurs shortly after S_2. The interval between the pulmonic component of the second heart sound (P_2) and the opening snap is inversely related to the magnitude of the left atrial –left ventricular pressure gradient. The murmur of MS is low-pitched and thus is best heard with the bell of the stethoscope. It is loudest at the left ventricular apex and often is appreciated only when the patient is turned in the left lateral decubitus position. It is usually of grade 1 or 2 intensity but may be absent when the cardiac output is severely reduced despite significant obstruction. The intensity of the murmur increases during maneuvers that increase cardiac output and mitral valve flow, such as exercise. The duration of the murmur reflects the length of time over which left atrial pressure exceeds left ventricular diastolic pressure. An increase in the intensity of the murmur just before S_1, a phenomenon known as *presystolic accentuation* (Figs. e13-1*A*, e13-6), occurs in patients in sinus rhythm and is due to a late increase in transmitral flow with atrial contraction. Presystolic accentuation does not occur in patients with atrial fibrillation.

The mid-diastolic murmur associated with tricuspid stenosis is best heard at the lower left sternal border and increases in intensity with inspiration. A prolonged *y* descent may be visible in the jugular venous waveform. This murmur is very difficult to hear and often is obscured by left-sided acoustical events.

There are several other causes of mid-diastolic murmurs. Large left atrial myxomas may prolapse across the mitral valve and cause variable degrees of obstruction to left ventricular inflow (Chap. 240). The murmur associated with an atrial myxoma may change in duration and intensity with changes in body position. An opening snap is not present, and there is no presystolic accentuation. Augmented mitral diastolic flow can occur with isolated severe MR or with a large left-to-right shunt at the ventricular or great vessel level and produce a soft, rapid filling sound (S_3) followed by a short, low-pitched mid-diastolic apical murmur. The Austin Flint murmur of severe, chronic AR has already been described.

A short, mid-diastolic murmur is rarely heard during an episode of acute rheumatic fever (Carey-Coombs murmur) and probably is due to flow through an edematous mitral valve. An opening snap is not present in the acute phase, and the murmur dissipates with resolution of the acute attack. Complete heart block with dyssynchronous atrial and ventricular activation may be associated with intermittent mid- to late diastolic murmurs if atrial contraction occurs when the mitral valve is partially closed. Mid-diastolic murmurs indicative of increased tricuspid valve flow can occur with severe, isolated TR and with large ASDs and significant left-to-right shunting. Other signs of an ASD are present (Chap. 236), including fixed splitting of S_2 and a mid-systolic murmur at the mid- to upper left sternal border. TTE is indicated for evaluation of a patient with a mid- to late diastolic murmur. Findings specific to the diseases discussed above will help guide management.

Continuous Murmurs

(**Figs.** e13-1*H* and e13-7) Continuous murmurs begin in systole, peak near the second heart sound, and continue into all or part of diastole. Their presence throughout the cardiac cycle implies a pressure gradient between two chambers or vessels during both systole and diastole. The continuous murmur associated with a patent ductus arteriosus is best heard at the upper left sternal border. Large, uncorrected shunts may lead to pulmonary hypertension, attenuation or obliteration of the diastolic component of the murmur, reversal of shunt flow, and differential cyanosis of the lower extremities. A ruptured sinus of Valsalva aneurysm creates a continuous murmur of abrupt onset at the upper right sternal border. Rupture typically occurs into a right heart chamber, and the murmur is indicative of a continuous pressure difference between the aorta and either the right ventricle or the right atrium. A continuous murmur also may be audible along the left sternal border with a coronary arteriovenous fistula and at the site of an arteriovenous fistula used for hemodialysis access. Enhanced flow through enlarged intercostal collateral arteries in patients with aortic coarctation may produce a continuous murmur along the course of one or more ribs. A cervical bruit with both systolic and diastolic components (a to-fro murmur, **Fig. e13-7**) usually indicates a high-grade carotid artery stenosis.

Figure e13-7

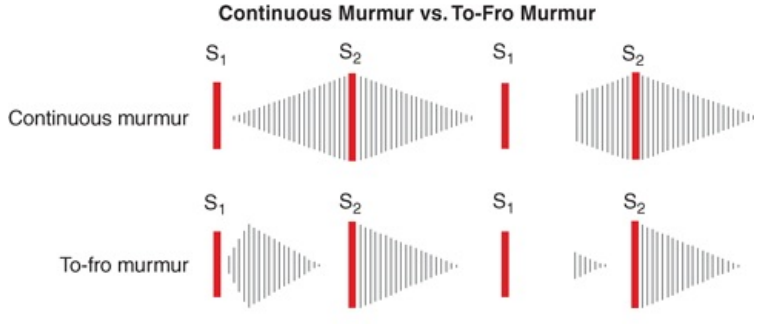

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Comparison of the continuous murmur and the to-fro murmur. During abnormal communication between high-pressure and low-pressure systems, a large pressure gradient exists throughout the cardiac cycle, producing a continuous murmur. A classic example is patent ductus arteriosus. At times, this type of murmur can be confused with a to-fro murmur, which is a combination of systolic ejection murmur and a murmur of semilunar valve incompetence. A classic example of a to-fro murmur is aortic stenosis and regurgitation. A continuous murmur crescendos to around the second heart sound (S_2), whereas a to-fro murmur has two components. The mid-systolic ejection component decrescendos and disappears as it approaches S_2. *(From JA Shaver, JJ Leonard, DF Leon, Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 55. Copyright, American Heart Association.)*

Not all continuous murmurs are pathologic. A continuous venous hum can be heard in healthy children

and young adults, especially during pregnancy; it is best appreciated in the right supraclavicular fossa and can be obliterated by pressure over the right internal jugular vein or by having the patient turn his or her head toward the examiner. The continuous mammary souffle of pregnancy is created by enhanced arterial flow through engorged breasts and usually appears during the late third trimester or early puerperium. The murmur is louder in systole. Firm pressure with the diaphragm of the stethoscope can eliminate the diastolic portion of the murmur.

Dynamic Auscultation

(**Tables e13-2** and 227-1) Careful attention to the behavior of heart murmurs during simple maneuvers that alter cardiac hemodynamics can provide important clues to their cause and significance.

Table e13-2 Dynamic Auscultation: Bedside Maneuvers that Can Be Used to Change the Intensity of Cardiac Murmurs (See Text)

1. Respiration
2. Isometric exercise (handgrip)
3. Transient arterial occlusion
4. Pharmacologic manipulation of preload and/or afterload
5. Valsalva maneuver
6. Rapid standing/squatting
7. Passive leg raising
8. Post-premature beat

RESPIRATION

Auscultation should be performed during quiet respiration or with a modest increase in inspiratory effort, as more forceful movement of the chest tends to obscure the heart sounds. Left-sided murmurs may be best heard at end expiration, when lung volumes are minimized and the heart and great vessels are brought closer to the chest wall. This phenomenon is characteristic of the murmur of AR. Murmurs of right-sided origin, such as tricuspid or pulmonic regurgitation, increase in intensity during inspiration. The intensity of left-sided murmurs either remains constant or decreases with inspiration.

Bedside assessment also should evaluate the behavior of S_2 with respiration and the dynamic relationship between the aortic and pulmonic components **(Fig. e13-8)**. Reversed splitting can be a feature of severe AS, HOCM, left bundle branch block, right ventricular apical pacing, or acute myocardial ischemia. Fixed splitting of S_2 in the presence of a grade 2 or 3 mid-systolic murmur at the mid- or upper left sternal border indicates an ASD. Physiologic but wide splitting during the respiratory cycle implies either premature aortic valve closure, as can occur with severe MR, or delayed pulmonic valve closure due to PS or right bundle branch block.

Figure e13-8

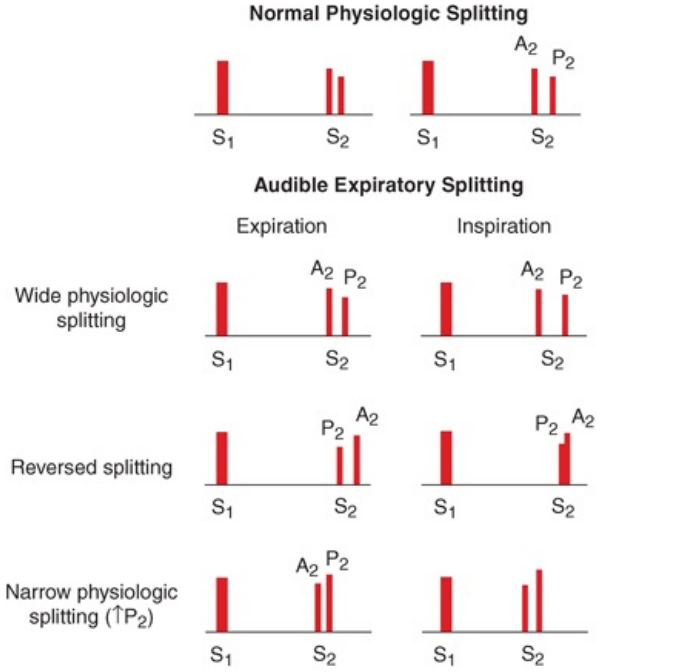

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Top. Normal physiologic splitting. During expiration, the aortic (A_2) and pulmonic (P_2) components of the second heart sound are separated by <30 ms and are appreciated as a single sound. During inspiration, the splitting interval widens, and A_2 and P_2 are clearly separated into two distinct sounds. ***Bottom.*** Audible expiratory splitting. Wide physiologic splitting is caused by a delay in P_2. Reversed splitting is caused by a delay in A_2, resulting in paradoxical movement; i.e., with inspiration P_2 moves toward A_2, and the splitting interval narrows. Narrow physiologic splitting occurs in pulmonary hypertension, and both A_2 and P_2 are heard during expiration at a narrow splitting interval because of the increased intensity and high-frequency composition of P_2. *(From JA Shaver, JJ Leonard, DF Leon, Examination of the Heart, Part IV, Auscultation of the Heart. Dallas, American Heart Association, 1990, p 17. Copyright, American Heart Association.)*

ALTERATIONS OF SYSTEMIC VASCULAR RESISTANCE

Murmurs can change characteristics after maneuvers that alter systemic vascular resistance and left ventricular afterload. The systolic murmurs of MR and VSD become louder during sustained handgrip, simultaneous inflation of blood pressure cuffs on both upper extremities to pressures 20–40 mmHg above systolic pressure for 20 s, or infusion of a vasopressor agent. The murmurs associated with AS or HOCM will become softer or remain unchanged with these maneuvers. The diastolic murmur of AR becomes louder in response to interventions that raise systemic vascular resistance.

Opposite changes in systolic and diastolic murmurs may occur with the use of pharmacologic agents that lower systemic vascular resistance. Inhaled amyl nitrite is now rarely used for this purpose but can help distinguish the murmur of AS or HOCM from that of either MR or VSD, if necessary. The former two murmurs increase in intensity, whereas the latter two become softer after exposure to amyl nitrite. As noted previously, the Austin Flint murmur of severe AR becomes softer, but the mid-diastolic rumble of MS becomes louder, in response to the abrupt lowering of systemic vascular resistance with amyl nitrite.

CHANGES IN VENOUS RETURN

The Valsalva maneuver results in an increase in intrathoracic pressure, followed by a decrease in venous return, ventricular filling, and cardiac output. The majority of murmurs decrease in intensity during the strain phase of the maneuver. Two notable exceptions are the murmurs associated with MVP and obstructive HOCM, both of which become louder during the Valsalva maneuver. The murmur of MVP may also become longer as leaflet prolapse occurs earlier in systole at smaller ventricular volumes. These murmurs behave in a similar and parallel fashion with standing. Both the click and the murmur of MVP move closer in timing to S_1 on rapid standing from a squatting position (Fig. e13-3).

The increase in the intensity of the murmur of HOCM is predicated on the augmentation of the dynamic left ventricular outflow tract gradient that occurs with reduced ventricular filling. Squatting results in abrupt increases in both venous return (preload) and left ventricular afterload that increases ventricular volume, changes that predictably cause a decrease in the intensity and duration of the murmurs associated with MVP and HOCM; the click and murmur of MVP move away from S_1 with squatting. Passive leg raising can be used to increase venous return in patients who are unable to

squat and stand. This maneuver may lead to a decrease in the intensity of the murmur associated with HOCM but has less effect in patients with MVP.

POST-PREMATURE VENTRICULAR CONTRACTION

A change in the intensity of a systolic murmur in the first beat after a premature beat, or in the beat after a long cycle length in patients with atrial fibrillation, can help distinguish AS from MR, particularly in an older patient in whom the murmur of AS is well transmitted to the apex. Systolic murmurs due to left ventricular outflow obstruction, including that due to AS, increase in intensity in the beat after a premature beat because of the combined effects of enhanced left ventricular filling and post-extrasystolic potentiation of contractile function. Forward flow accelerates, causing an increase in the gradient and a louder murmur. The intensity of the murmur of MR does not change in the post-premature beat as there is relatively little further increase in mitral valve flow or change in the left ventricular–left atrial gradient.

The Clinical Context

Additional clues to the etiology and importance of a heart murmur can be gleaned from the history and other physical examination findings. Symptoms suggestive of cardiovascular, neurologic, or pulmonary disease help focus the differential diagnosis, as do findings relevant to the jugular venous pressure and waveforms, the arterial pulses, other heart sounds, the lungs, the abdomen, the skin and the extremities. In many instances, laboratory studies, an ECG, and/or a chest x-ray may have been obtained earlier and may contain valuable information. A patient with suspected infective endocarditis, for example, may have a murmur in the setting of fever, chills, anorexia, fatigue, dyspnea, splenomegaly, petechiae, and positive blood cultures. A new systolic murmur in a patient with a marked fall in blood pressure after a recent MI suggests myocardial rupture. By contrast, an isolated grade 1 or 2 mid-systolic murmur at the left sternal border in a healthy, active, and asymptomatic young adult is most likely a benign finding for which no further evaluation is indicated. The context in which the murmur is appreciated often dictates the need for further testing.

Echocardiography

(See **Fig. e13-9**, Chaps. 227 and 229) Echocardiography with color flow and spectral Doppler is a valuable tool for the assessment of cardiac murmurs. Information regarding valve structure and function, chamber size, wall thickness, ventricular function, estimated pulmonary artery pressures, intracardiac shunt flow, pulmonary and hepatic vein flow, and aortic flow can be ascertained readily. It is important to note that Doppler signals of trace or mild valvular regurgitation of no clinical consequence can be detected with structurally normal tricuspid, pulmonic, and mitral valves. Such signals are not likely to generate enough turbulence to create an audible murmur.

Figure e13-9

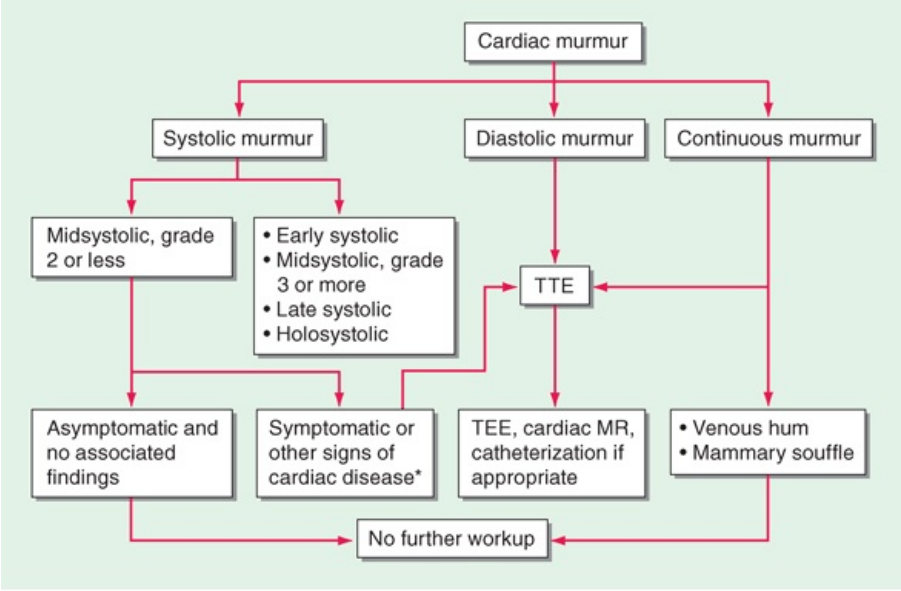

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Strategy for evaluating heart murmurs. * If an electrocardiogram or chest x-ray has been obtained and is abnormal, echocardiography is indicated. TTE, transthoracic echocardiography; TEE, transesophageal echocardiography; MR, magnetic resonance. *(Adapted from RO Bonow et al: J Am Coll Cardiol 32:1486, 1998.)*

Echocardiography is indicated for the evaluation of patients with early, late, or holosystolic murmurs and patients with grade 3 or louder mid-systolic murmurs. Patients with grade 1 or 2 mid-systolic murmurs but other symptoms or signs of cardiovascular disease, including those from ECG or chest x-ray, should also undergo echocardiography. Echocardiography is indicated for the evaluation of any patient with a diastolic murmur and for patients with continuous murmurs not due to a venous hum or mammary souffle. Echocardiography also should be considered when there is a clinical need to verify normal cardiac structure and function in a patient whose symptoms and signs are probably noncardiac in origin. The performance of serial echocardiography to follow the course of asymptomatic individuals with valvular heart disease is a central feature of their longitudinal assessment and provides valuable information that may have an impact on decisions regarding the timing of surgery. Routine echocardiography is *not* recommended for asymptomatic patients with a grade 1 or 2 mid-systolic murmur without other signs of heart disease. For this category of patients, referral to a cardiovascular specialist should be considered if there is doubt about the significance of the murmur after the initial examination.

The selective use of echocardiography outlined above has not been subjected to rigorous analysis of its cost-effectiveness. At least one study has suggested that initial referral of pediatric patients with heart murmurs to a specialist results in modest cost savings. For some clinicians, handheld or miniaturized cardiac ultrasound devices have replaced the stethoscope. Although several reports attest to the improved sensitivity of such devices for the detection of valvular heart disease, accuracy is highly operator-dependent, and incremental cost considerations have not been addressed adequately. The use of electronic or digital stethoscopes with spectral display capabilities has also been proposed as a method to improve the characterization of heart murmurs and the mentored teaching of cardiac auscultation.

Other Cardiac Testing

(Chap. 229, Fig. e13-9) In relatively few patients, clinical assessment and TTE do not adequately characterize the origin and significance of a heart murmur. Transesophageal echocardiography (TEE) can be considered for further evaluation, especially when the TTE windows are limited by body size, chest configuration, or intrathoracic pathology. TEE offers enhanced sensitivity for the detection of a wide range of structural cardiac disorders. Electrocardiographically gated cardiac magnetic resonance (CMR) imaging, although limited in its ability to display valvular morphology, can provide quantitative information regarding valvular function, stenosis severity, regurgitant fraction, regurgitant volume, shunt flow, chamber and great vessel size, ventricular function, and myocardial perfusion. CMR has greater capability than cardiac computed tomography (CCT) in this regard and has largely supplanted the need for cardiac catheterization and invasive hemodynamic assessment when there is a discrepancy between the clinical and echocardiographic findings. Invasive coronary angiography is performed routinely in most adult patients before valve surgery, especially when there is a suspicion of coronary artery disease predicated on symptoms, risk factors, and/or age. The use of computed tomography coronary angiography (CCTA) to exclude coronary artery disease in young patients with a low pretest probability of disease before valve surgery is under active investigation.

Integrated Approach

The accurate identification of a heart murmur begins with a systematic approach to cardiac auscultation. Characterization of its major attributes, as reviewed above, allows the examiner to construct a preliminary differential diagnosis, which is then refined by integration of information available from the history, associated cardiac findings, the general physical examination, and the clinical context. The need for and urgency of further testing follow sequentially. Correlation of the findings on auscultation with the noninvasive data provides an educational feedback loop and an opportunity for improving physical examination skills. Cost constraints mandate that noninvasive imaging be justified on the basis of its incremental contribution to diagnosis, treatment, and outcome. Additional study is required to assess the cost-effective application of newer imaging technology in patients with heart murmurs.

FURTHER READINGS

Babu AN et al: Eponyms and the diagnosis of aortic regurgitation: What says the evidence? Ann Intern Med 138:736, 2003[PMID: 12729428]

Barrett MJ et al: Mastering cardiac murmurs: The power of repetition. Chest 126:470, 2004[PMID: 15302733]

Bonow RO et al: ACC/AHA 2006 Guidelines for the management of patients with valvular heart disease: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients with Valvular Heart Disease). American College of Cardiology Web Site. Available at *http://www.acc.org/clinical/guidelines/valvular/index.pdf*

Choundhry NK, Etchells EE: The rational clinical examination: Does this patient have aortic

regurgitation? JAMA 281:2231, 1999

Etchells E et al: Does this patient have an abnormal systolic heart murmur? JAMA 277:564, 1997[PMID: 9032164]

Fang J, O'Gara P: *The History and Physical Examination* in P Libby et al (eds). *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 8th ed, Philadelphia, Saunders Elsevier, 2008

Hoyte H et al: Cardiac auscultation training of medical students: A comparison of electronic sensor-based and acoustic stethoscopes. BMC Med Educ 5:14, 2005[PMID: 15882458]

Kobal SL et al: Comparison of effectiveness of hand-carried ultrasound to bedside cardiovascular physical examination. Am J Cardiol 96:1002, 2005.[PMID: 16188532]

Tavel ME: Cardiac auscultation: A glorious past—and it does have a future! Circulation 113:1255, 2006[PMID: 16520426]

Vukanovic-Criley JM et al: Competency in cardiac examination skills in medical students, trainees, physicians and faculty: A multicenter study. Arch Intern Med 166:610, 2006[PMID: 16567598]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PALPITATIONS: INTRODUCTION

Palpitations are extremely common among patients who present to their internist and can best be defined as an intermittent "thumping," "pounding," or "fluttering" sensation in the chest. This sensation can be either intermittent or sustained and either regular or irregular. Most patients interpret palpitations as an unusual awareness of the heartbeat, and become especially concerned when they sense that they have had "skipped" or "missing" heartbeats. Palpitations are often noted when the patient is quietly resting, during which time other stimuli are minimal. Palpitations that are positional generally reflect a structural process within (e.g., atrial myxoma) or adjacent to (e.g., mediastinal mass) the heart.

Palpitations are brought about by cardiac (43%), psychiatric (31%), miscellaneous (10%), and unknown (16%) causes, according to one large series. Among the cardiovascular causes are premature atrial and ventricular contractions, supraventricular and ventricular arrhythmias, mitral valve prolapse (with or without associated arrhythmias), aortic insufficiency, atrial myxoma, and pulmonary embolism. Intermittent palpitations are commonly caused by premature atrial or ventricular contractions: the post-extrasystolic beat is sensed by the patient owing to the increase in ventricular end-diastolic dimension following the pause in the cardiac cycle and the increased strength of contraction (post-extrasystolic potentiation) of that beat. Regular, sustained palpitations can be caused by regular supraventricular and ventricular tachycardias. Irregular, sustained palpitations can be caused by atrial fibrillation. It is important to note that most arrhythmias are not associated with palpitations. In those that are, it is often useful either to ask the patient to "tap out" the rhythm of the palpitations or to take his/her pulse while experiencing palpitations. In general, hyperdynamic cardiovascular states caused by catecholaminergic stimulation from exercise, stress, or pheochromocytoma can lead to palpitations. Palpitations are common among athletes, especially older endurance athletes. In addition, the enlarged ventricle of aortic regurgitation and accompanying hyperdynamic precordium frequently lead to the sensation of palpitations. Other factors that enhance the strength of myocardial contraction, including tobacco, caffeine, aminophylline, atropine, thyroxine, cocaine, and amphetamines, can cause palpitations.

Psychiatric causes of palpitations include panic attacks or disorders, anxiety states, and somatization, alone or in combination. Patients with psychiatric causes for palpitations more commonly report a longer duration of the sensation (>15 min) and other accompanying symptoms than do patients with other causes. Among the miscellaneous causes of palpitations included are thyrotoxicosis, drugs (see above) and ethanol, spontaneous skeletal muscle contractions of the chest wall, pheochromocytoma, and systemic mastocytosis.

Approach to the Patient: Palpitations

The principal goal in assessing patients with palpitations is to determine if the symptom is caused by a life-threatening arrhythmia. Patients with preexisting coronary artery disease (CAD) or risk factors for CAD are at greatest risk for ventricular arrhythmias as a cause for palpitations. In addition, the association of palpitations with other symptoms suggesting hemodynamic compromise, including syncope or lightheadedness, supports this diagnosis. Palpitations caused by sustained tachyarrhythmias in patients with CAD can be accompanied by angina pectoris or dyspnea, and in patients with ventricular dysfunction (systolic or diastolic), aortic stenosis, hypertrophic cardiomyopathy, or mitral stenosis, with or without CAD, can be accompanied by dyspnea from increased left atrial and pulmonary venous pressure.

Key features of the physical examination that will help confirm or refute the presence of an arrhythmia as a cause for the palpitations and its adverse hemodynamic consequences include measurement of the vital signs, assessment of the jugular venous pressure and pulse, and auscultation of the chest and precordium. A resting electrocardiogram can be used to document the arrhythmia. If exertion is known to induce the arrhythmia and accompanying palpitations, exercise electrocardiography can be used to make the diagnosis. If the arrhythmia is sufficiently infrequent, other methods must be used, including continuous electrocardiographic (Holter) monitoring; telephonic monitoring, through which the patient can transmit an electrocardiographic tracing during a sensed episode; loop recordings (external or implantable), which can capture the electrocardiographic event for later review; and mobile cardiac outpatient telemetry. Recent data suggest that Holter monitoring is of limited clinical utility, while the implantable loop recorder and mobile cardiac outpatient telemetry are safe and possibly more cost-effective in the assessment of patients with recurrent, unexplained palpitations.

Most patients with palpitations do not have serious arrhythmias or underlying structural heart disease. Occasional benign atrial or ventricular premature contractions can often be managed with beta-blocker therapy if sufficiently troubling to the patient. Palpitations incited by alcohol, tobacco, or illicit drugs need to be managed by abstention, while those caused by pharmacologic agents should be addressed by considering alternate therapies when appropriate or possible. Psychiatric causes of palpitations may benefit from cognitive or pharmacotherapies. The physician should note that palpitations are at the very least bothersome and, on occasion, frightening to the patient. Once serious causes for the symptom have been excluded, the patient should be reassured that the palpitations will not adversely affect prognosis.

FURTHER READINGS

Abbott AV: Diagnostic approach to palpitations. Am Fam Phys 71:743, 2005[PMID: 15742913]

Giada F et al: Recurrent unexplained palpitations (RUP) study. J Am Coll Cardiol 49:1951, 2007[PMID: 17498580]

Lawless CE, Briner W: Palpitations in athletes. Sports Med 38:687, 2008[PMID: 18620468]

Olson JA et al: Utility of mobile cardiac outpatient telemetry for the diagnosis of palpitations, presyncope, syncope, and the assessment of therapy efficacy. J Cardiovasc Electrophysiol 18:473, 2007[PMID: 17343724]

Sulfi S et al: Limited clinical utility of Holter monitoring in patients with palpitations or altered consciousness: analysis of 8973 recordings in 7394 patients. Ann Noninvasive Electrocardiol 13:39, 2008[PMID: 18234005]

Weber BE, Kapoor WN: Evaluation and outcomes of patients with palpitations. Am J Med 100:138, 1996[PMID: 8629647]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DYSPHAGIA: INTRODUCTION

Dysphagia—difficulty with swallowing—refers to problems with the transit of food or liquid from the mouth to the hypopharynx or through the esophagus. Severe dysphagia can compromise nutrition, cause aspiration, and reduce quality of life. Additional terminology pertaining to swallowing dysfunction is as follows. *Aphagia* denotes complete esophageal obstruction, most commonly encountered in the acute setting of a food bolus or foreign body impaction. *Odynophagia* refers to painful swallowing, typically resulting from mucosal ulceration within the oropharynx or esophagus. It commonly is accompanied by dysphagia, but the converse is not true. *Globus pharyngeus* is a foreign body sensation localized in the neck that does not interfere with swallowing and sometimes is relieved by swallowing. *Transfer dysphagia* frequently results in nasal regurgitation and pulmonary aspiration during swallowing and is characteristic of oropharyngeal dysphagia. *Phagophobia* (fear of swallowing) and *refusal to swallow* may be psychogenic or related to anticipatory anxiety about food bolus obstruction, odynophagia, or aspiration.

Physiology of Swallowing

Swallowing begins with a voluntary (oral) phase that includes preparation during which food is masticated and mixed with saliva. This is followed by a transfer phase during which the bolus is pushed into the pharynx by the tongue. Bolus entry into the hypopharynx initiates the pharyngeal swallow response, which is centrally mediated and involves a complex series of actions, the net result of which is to propel food through the pharynx into the esophagus while preventing its entry into the airway. To accomplish this, the larynx is elevated and pulled forward, actions that also facilitate upper esophageal sphincter (UES) opening. Tongue pulsion then propels the bolus through the UES, followed by a peristaltic contraction that clears residue from the pharynx and through the esophagus. The lower esophageal sphincter (LES) relaxes as the food enters the esophagus and remains relaxed until the peristaltic contraction has delivered the bolus into the stomach. Peristaltic contractions elicited in response to a swallow are called *primary peristalsis* and involve sequenced inhibition followed by contraction of the musculature along the entire length of the esophagus. The inhibition that precedes the peristaltic contraction is called *deglutitive inhibition*. Local distention of the esophagus anywhere along its length, as may occur with gastroesophageal reflux, activates *secondary peristalsis* that begins at the point of distention and proceeds distally. Tertiary esophageal contractions are nonperistaltic, disordered esophageal contractions that may be observed to occur spontaneously during fluoroscopic observation.

The musculature of the oral cavity, pharynx, UES, and cervical esophagus is striated and directly innervated by lower motor neurons carried in cranial nerves (Fig. 38-1). Oral cavity muscles are innervated by the seventh (trigeminal) and eighth (facial) cranial nerves; the tongue, by the twelfth (hypoglossal) cranial nerve. Pharyngeal muscles are innervated by the eleventh (glossopharyngeal) and twelfth (vagus) cranial nerves.

Figure 38-1

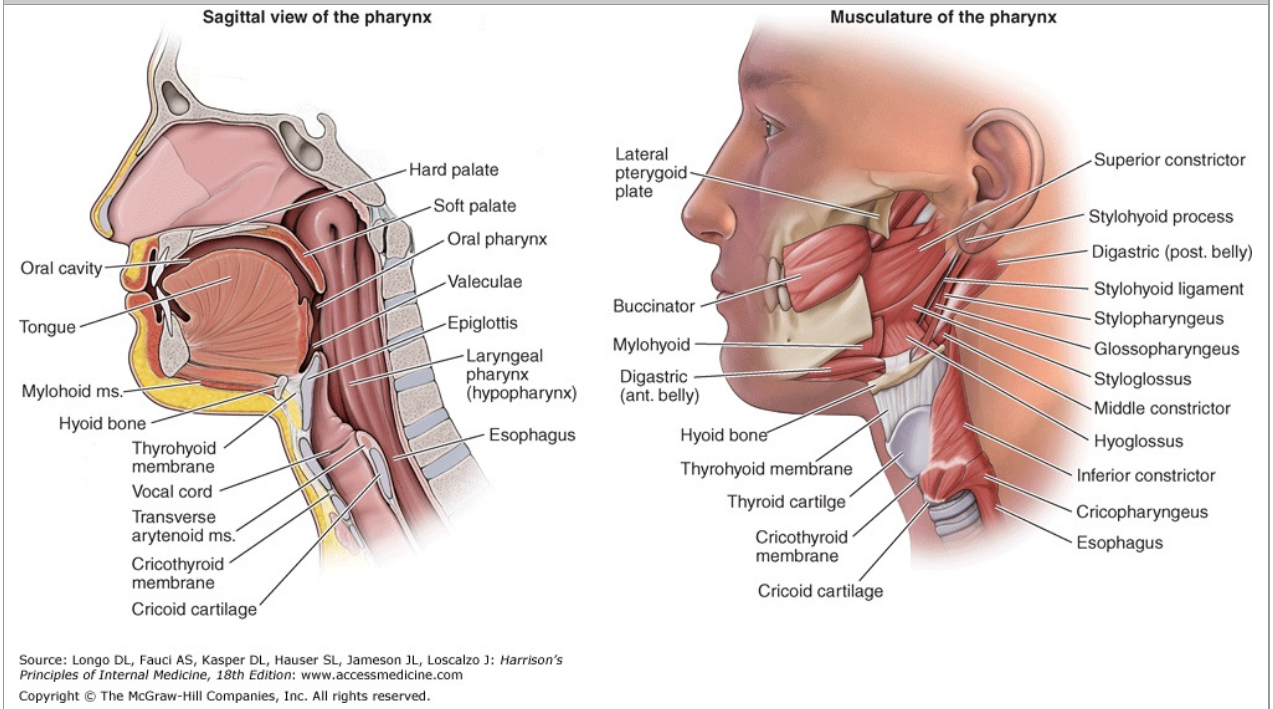

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sagittal and diagrammatic views of the musculature involved in enacting oropharyngeal swallowing. Note the dominance of the tongue in the sagittal view and the intimate relationship between the entrance to the larynx (airway) and the esophagus. In the resting configuration illustrated, the esophageal inlet is closed. This is transiently reconfigured such that the esophageal inlet is open and the laryngeal inlet closed during swallowing. *[Adapted from PJ Kahrilas, in DW Gelfand and JE Richter (eds): Dysphagia: Diagnosis and Treatment. New York: Igaku-Shoin Medical Publishers, 1989, pp. 11–28.]*

Physiologically, the UES consists of the cricopharyngeus muscle, the adjacent inferior pharyngeal constrictor, and the proximal portion of the cervical esophagus. UES innervation is derived from the vagus nerve, whereas the innervation to the musculature acting on the UES to facilitate its opening during swallowing comes from the fifth, seventh, and twelfth cranial nerves. The UES remains closed at rest owing to both its inherent elastic properties and neurogenically mediated contraction of the

cricopharyngeus muscle. UES opening during swallowing involves both cessation of vagal excitation to the cricopharyngeus and simultaneous contraction of the suprahyoid and geniohyoid muscles that pull open the UES in conjunction with the upward and forward displacement of the larynx.

The neuromuscular apparatus for peristalsis is distinct in proximal and distal parts of the esophagus. The cervical esophagus, like the pharyngeal musculature, consists of striated muscle and is directly innervated by lower motor neurons of the vagus nerve. Peristalsis in the proximal esophagus is governed by the sequential activation of the vagal motor neurons in the nucleus ambiguus. In contrast, the distal esophagus and LES are composed of smooth muscle and are controlled by excitatory and inhibitory neurons within the esophageal myenteric plexus. Medullary preganglionic neurons from the dorsal motor nucleus of the vagus trigger peristalsis via these ganglionic neurons during primary peristalsis. Neurotransmitters of the excitatory ganglionic neurons are acetylcholine and substance P; those of the inhibitory neurons are vasoactive intestinal peptide and nitric oxide. Peristalsis results from the patterned activation of inhibitory followed by excitatory ganglionic neurons, with progressive dominance of the inhibitory neurons distally. Similarly, LES relaxation occurs with the onset of deglutitive inhibition and persists until the peristaltic sequence is complete. At rest, the LES is contracted because of excitatory ganglionic stimulation and its intrinsic myogenic tone, a property that distinguishes it from the adjacent esophagus. The function of the LES is supplemented by the surrounding muscle of the right diaphragmatic crus, which acts as an external sphincter during inspiration, cough, or abdominal straining.

Pathophysiology of Dysphagia

Dysphagia can be subclassified both by location and by the circumstances in which it occurs. With respect to location, distinct considerations apply to oral, pharyngeal, or esophageal dysphagia. Normal transport of an ingested bolus depends on the consistency and size of the bolus, the caliber of the lumen, the integrity of peristaltic contraction, and deglutitive inhibition of both the UES and the LES. Dysphagia caused by an oversized bolus or a narrow lumen is called *structural dysphagia*, whereas dysphagia due to abnormalities of peristalsis or impaired sphincter relaxation after swallowing is called *propulsive* or *motor dysphagia*. More than one mechanism may be operative in a patient with dysphagia. Scleroderma commonly presents with absent peristalsis as well as a weakened LES that predisposes patients to peptic stricture formation. Likewise, radiation therapy for head and neck cancer may compound the functional deficits in the oropharyngeal swallow attributable to the tumor and cause cervical esophageal stenosis.

ORAL AND PHARYNGEAL (OROPHARYNGEAL) DYSPHAGIA

Oral-phase dysphagia is associated with poor bolus formation and control so that food has prolonged retention within the oral cavity and may seep out of the mouth. Drooling and difficulty in initiating swallowing are other characteristic signs. Poor bolus control also may lead to premature spillage of food into the hypopharynx with resultant aspiration into the trachea or regurgitation into the nasal cavity. Pharyngeal-phase dysphagia is associated with retention of food in the pharynx due to poor tongue or pharyngeal propulsion or obstruction at the UES. Signs and symptoms of concomitant hoarseness or cranial nerve dysfunction may be associated with oropharyngeal dysphagia.

Oropharyngeal dysphagia may be due to neurologic, muscular, structural, iatrogenic, infectious, and metabolic causes. Iatrogenic, neurologic, and structural pathologies are most common. Iatrogenic causes include surgery and radiation, often in the setting of head and neck cancer. Neurogenic dysphagia resulting from cerebrovascular accidents, Parkinson's disease, and amyotrophic lateral sclerosis is a major source of morbidity related to aspiration and malnutrition. Medullary nuclei directly innervate the oropharynx. Lateralization of pharyngeal dysphagia implies either a structural pharyngeal lesion or a neurologic process that selectively targeted the ipsilateral brainstem nuclei or cranial nerve. Advances in functional brain imaging have elucidated an important role of the cerebral cortex in swallow function and dysphagia. Asymmetry in the cortical representation of the pharynx provides an explanation for the dysphagia that occurs as a consequence of unilateral cortical cerebrovascular accidents.

Oropharyngeal structural lesions causing dysphagia include Zenker's diverticulum, cricopharyngeal bar, and neoplasia. Zenker's diverticulum typically is encountered in elderly patients, with an estimated prevalence between 1:1000 and 1:10,000. In addition to dysphagia, patients may present with regurgitation of particulate food debris, aspiration, and halitosis. The pathogenesis is related to stenosis of the cricopharyngeus that causes diminished opening of the UES and results in increased hypopharyngeal pressure during swallowing with development of a pulsion diverticulum immediately above the cricopharyngeus in a region of potential weakness known as Killian's dehiscence. A cricopharyngeal bar, appearing as a prominent indentation behind the lower third of the cricoid cartilage, is related to Zenker's diverticulum in that it involves limited distensibility of the cricopharyngeus and can lead to the formation of a Zenker's diverticulum. However, a cricopharyngeal bar is a common radiographic finding, and most patients with transient cricopharyngeal bars are asymptomatic, making it important to rule out alternative etiologies of dysphagia before treatment. Furthermore, cricopharyngeal bars may be secondary to other neuromuscular disorders.

Since the pharyngeal phase of swallowing occurs in less than a second, rapid-sequence fluoroscopy is necessary to evaluate for functional abnormalities. Adequate fluoroscopic examination requires that the patient be conscious and cooperative. The study incorporates recordings of swallow sequences during ingestion of food and liquids of varying consistencies. The pharynx is examined to detect bolus retention, regurgitation into the nose, or aspiration into the trachea. Timing and integrity of pharyngeal contraction and opening of the UES with a swallow are analyzed to assess both aspiration risk and the potential for swallow therapy. Structural abnormalities of the oropharynx, especially those which may require biopsies, also should be assessed by direct laryngoscopic examination.

ESOPHAGEAL DYSPHAGIA

The adult esophagus measures 18–26 cm in length and is anatomically divided into the cervical esophagus, extending from the pharyngoesophageal junction to the suprasternal notch, and the thoracic esophagus, which continues to the diaphragmatic hiatus. When distended, the esophageal lumen has internal dimensions of about 2 cm in the anteroposterior plane and 3 cm in the lateral plane. Solid food dysphagia becomes common when the lumen is narrowed to <13 mm but also can occur with larger diameters in the setting of poorly masticated food or motor dysfunction. Circumferential lesions are more likely to cause dysphagia than are lesions that involve only a partial circumference of the esophageal wall. The most common structural causes of dysphagia are Schatzki's rings, eosinophilic esophagitis, and peptic strictures. Dysphagia also occurs in the setting of gastroesophageal reflux disease without a stricture, perhaps on the basis of altered esophageal sensation, distensibility, or motor function.

Propulsive disorders leading to esophageal dysphagia result from abnormalities of peristalsis and/or deglutitive inhibition, potentially affecting the cervical or thoracic esophagus. Since striated muscle pathology usually involves both the oropharynx and the cervical esophagus, the clinical manifestations usually are dominated by oropharyngeal dysphagia. Diseases affecting smooth muscle involve both the thoracic esophagus and the LES. A dominant manifestation of this, absent peristalsis, refers to either the complete absence of swallow-induced contraction or the presence of nonperistaltic, disordered contractions. Absent peristalsis and failure of deglutitive LES relaxation are the defining features of achalasia. In diffuse esophageal spasm (DES), LES function is normal, with the disordered motility restricted to the esophageal body. Absent peristalsis combined with severe weakness of the LES is a nonspecific pattern commonly found in patients with scleroderma.

Approach to the Patient: Dysphagia

Figure 38-2 shows an algorithm for the approach to a patient with dysphagia.

Figure 38-2

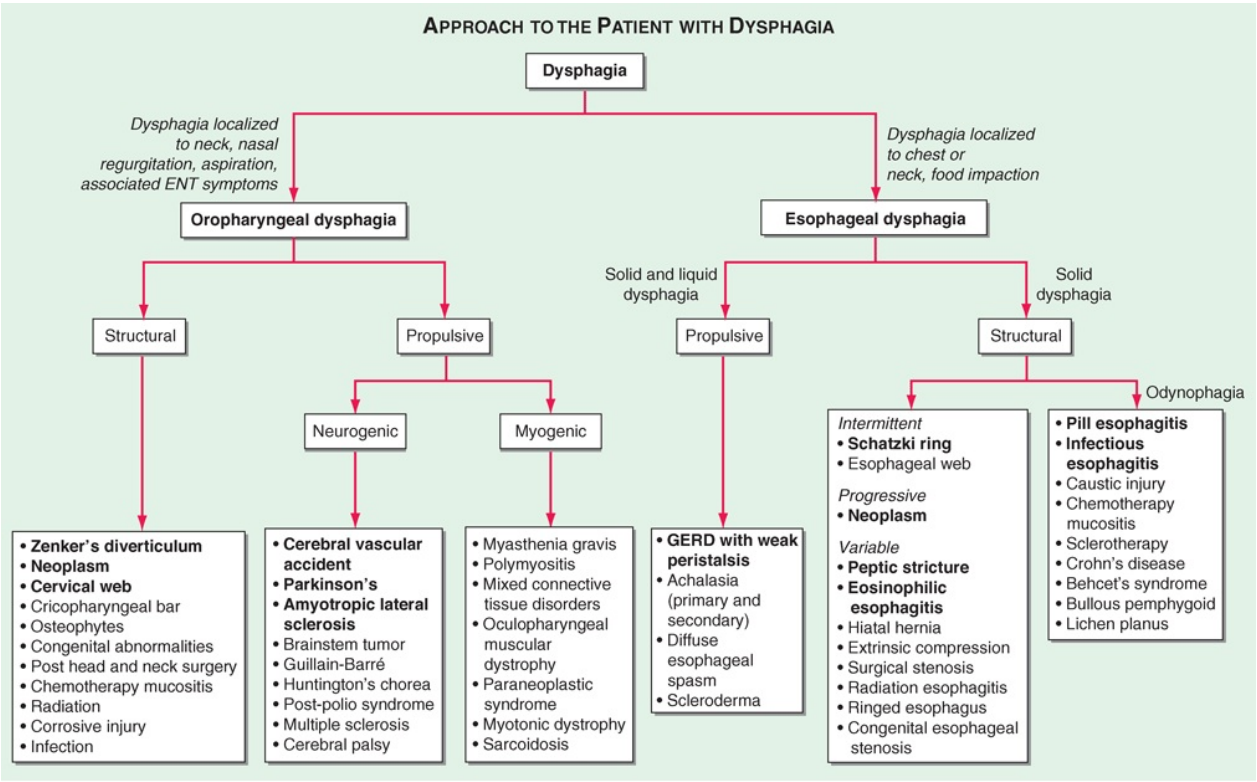

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the patient with dysphagia. Etiologies in bold print are the most common. ENT, ear, nose, and throat; GERD, gastroesophageal reflux disease.

HISTORY

The patient history is extremely valuable in making a presumptive diagnosis or at least substantially restricting the differential diagnoses in most patients. Key elements of the history are the localization of dysphagia, the circumstances in which dysphagia is experienced, other symptoms associated with dysphagia, and progression. Dysphagia that localizes to the suprasternal notch may indicate either an oropharyngeal or an esophageal etiology as distal dysphagia is referred proximally about 30% of the time. Dysphagia that localizes to the chest is esophageal in origin. Nasal regurgitation and tracheobronchial aspiration with swallowing are hallmarks of oropharyngeal dysphagia or a tracheoesophageal fistula. The presence of hoarseness may be another important diagnostic clue. When hoarseness precedes dysphagia, the primary lesion is usually laryngeal; hoarseness that occurs after the development of dysphagia may result from compromise of the recurrent laryngeal nerve by a malignancy. The type of food causing dysphagia is a crucial detail. Intermittent dysphagia that occurs only with solid food implies structural dysphagia, whereas constant dysphagia with both liquids and solids strongly suggests a motor abnormality. Two caveats to this pattern are that despite having a motor abnormality, patients with scleroderma generally develop mild dysphagia for solids only and, somewhat paradoxically, that patients with oropharyngeal dysphagia often have greater difficulty managing liquids than solids. Dysphagia that is progressive over the course of weeks to months raises concern for neoplasia. Episodic dysphagia to solids that is unchanged over years indicates a benign disease process such as a Schatzki's ring or eosinophilic esophagitis. Food impaction with a prolonged inability to pass an ingested bolus even with ingestion of liquid is typical of a structural dysphagia. Chest pain frequently accompanies dysphagia whether it is related to motor disorders, structural disorders, or reflux disease. A prolonged history of heartburn preceding the onset of dysphagia is suggestive of peptic stricture and, less commonly, esophageal adenocarcinoma. A history of prolonged nasogastric intubation, esophageal or head and neck surgery, ingestion of caustic agents or pills, previous radiation or chemotherapy, or associated mucocutaneous diseases may help isolate the cause of dysphagia. With accompanying odynophagia, which usually is indicative of ulceration, infectious or pill-induced esophagitis should be suspected. In patients with AIDS or other immunocompromised states, esophagitis due to opportunistic infections such as *Candida*, herpes simplex virus, or cytomegalovirus and to tumors such as Kaposi's sarcoma and lymphoma should be considered. A strong history of atopy increases concerns for eosinophilic esophagitis.

PHYSICAL EXAMINATION

Physical examination is important in the evaluation of oral and pharyngeal dysphagia because dysphagia is usually only one of many manifestations of a more global disease process. Signs of bulbar or pseudobulbar palsy, including dysarthria, dysphonia, ptosis, tongue atrophy, and hyperactive jaw jerk, in addition to evidence of generalized neuromuscular disease, should be elicited. The neck should be examined for thyromegaly. A careful inspection of the mouth and pharynx should disclose lesions that may interfere with passage of food. Physical examination is less helpful in the evaluation of esophageal dysphagia as most relevant pathology is restricted to the esophagus. The notable exception is skin disease. Changes in the skin may suggest a diagnosis of scleroderma or mucocutaneous diseases such as pemphigoid and epidermol-ysis bullosa, all of which can involve the esophagus.

DIAGNOSTIC PROCEDURES

Although most instances of dysphagia are attributable to benign disease processes, dysphagia is also a cardinal symptom of several malignancies, making it an important symptom to evaluate. Even when not attributable to malignancy, dysphagia is usually a manifestation of an identifiable and treatable disease entity, making its evaluation beneficial to the patient and

gratifying to the practitioner. The specific diagnostic algorithm to pursue is guided by the details of the history. If oral or pharyngeal dysphagia is suspected, a fluoroscopic swallow study, usually done by a swallow therapist, is the procedure of choice. Otolaryngoscopic and neurologic evaluation also can be important, depending on the circumstances. For suspected esophageal dysphagia, endoscopy is the single most useful test. Endoscopy allows better visualization of mucosal lesions than does barium radiography and also allows one to obtain mucosal biopsies. Furthermore, therapeutic intervention with esophageal dilatation can be done as part of the procedure if it is deemed necessary. Of note, the emergence of eosinophilic esophagitis as a common cause of dysphagia in adults has led to the recommendation that esophageal mucosal biopsies be obtained routinely in the evaluation of unexplained dysphagia even if no endoscopic lesions are evident. For cases of suspected esophageal motility disorders, esophagogastroscopy is still the primary examination as neoplastic and inflammatory conditions can secondarily produce patterns of either achalasia or esophageal spasm. Esophageal manometry is done if dysphagia is not adequately explained by endoscopy or to confirm the diagnosis of a suspected esophageal motor disorder. Barium radiography can provide useful adjunctive information in cases of subtle or complex esophageal strictures, esophageal diverticula, or paraesophageal herniation. In specific cases, CT examination and endoscopic ultrasonography may be useful.

TREATMENT

Treatment of dysphagia depends on both the locus and the specific etiology. Oropharyngeal dysphagia most commonly results from functional deficits caused by neurologic disorders. In such circumstances, the treatment focuses on utilizing postures or maneuvers devised to reduce pharyngeal residue and enhance airway protection learned under the direction of a trained swallow therapist. Aspiration risk may be reduced by altering the consistency of ingested food and liquid. Dysphagia resulting from a cerebrovascular accident usually, but not always, spontaneously improves within the first few weeks after the event. More severe and persistent cases may require gastrostomy and enteral feeding. Patients with myasthenia gravis (Chap. 386) and polymyositis (Chap. 388) may respond to medical treatment of the primary neuromuscular disease. Surgical intervention with cricopharyngeal myotomy is usually not helpful, with the exception of specific disorders such as the idiopathic cricopharyngeal bar, Zenker's diverticulum, and oculopharyngeal muscular dystrophy. Chronic neurologic disorders such as Parkinson's disease and amyotrophic lateral sclerosis may manifest with severe oropharyngeal dysphagia. Feeding by a nasogastric tube or an endoscopically placed gastrostomy tube may be considered for nutritional support; however, these maneuvers do not provide protection against aspiration of salivary secretions or refluxed gastric contents.

Treatment of esophageal dysphagia is covered in detail in Chap. 292. The majority of causes of esophageal dysphagia are effectively managed by means of esophageal dilatation using bougie or balloon dilators. Cancer and achalasia are often managed surgically, although endoscopic techniques are available for both palliation and primary therapy, respectively. Infectious etiologies respond to antimicrobial medications or treatment of the underlying immunosuppressive state. Finally, eosinophilic esophagitis has emerged as an important cause of dysphagia that is amenable to treatment by elimination of dietary allergens or topical glucocorticoids.

FURTHER READINGS

Castell DO, Richter JE: *The Esophagus*, 3rd ed. Philadelphia, Lippincott Williams and Wilkins, 2004

Cook IJ, Kahrilas PJ: AGA technical review on management of oropharyngeal dysphagia. Gastroenterology 116:455, 1999[PMID: 9922328]

Gomes CA Jr et al: Percutaneous endoscopic gastrostomy versus nasogastric tube feeding for adults with swallowing disturbances. Cochrane Database Syst Rev 11:CD008096, 2010

Goyal RK, Shaker R: GI Motility. *http://www.nature.com/gimo/index.html*

Hirano I: Esophagus: Anatomy and structural anomalies, in *Textbook of Gastroenterology* , 5th ed, T Yamada (ed). Oxford, United Kingdom, Wiley- Blackwell, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NAUSEA, VOMITING, AND INDIGESTION: INTRODUCTION

***Nausea* is the subjective feeling of a need to vomit. *Vomiting* (emesis) is the oral expulsion of gastrointestinal contents resulting from contractions of gut and thoracoabdominal wall musculature. Vomiting is contrasted with *regurgitation*, the effortless passage of gastric contents into the mouth. *Rumination* is the repeated regurgitation of stomach contents, which may be rechewed and reswallowed. In contrast to vomiting, these phenomena often exhibit volitional control. *Indigestion* is a nonspecific term that encompasses a variety of upper abdominal complaints including nausea, vomiting, heartburn, regurgitation, and dyspepsia (the presence of symptoms thought to originate in the gastroduodenal region). Some individuals with dyspepsia report predominantly epigastric burning, gnawing discomfort, or pain. Others with dyspepsia experience a constellation of symptoms including postprandial fullness, early satiety (an inability to complete a meal due to premature fullness), bloating, eructation (belching), and anorexia.**

NAUSEA AND VOMITING

Mechanisms

Vomiting is coordinated by the breainstem and is effected by responses in the gut, pharynx, and thoracoabdominal wall. The mechanisms underlying nausea are poorly understood but likely involve the cerebral cortex, because nausea requires conscious perception. This is supported by electroencephalographic studies showing activation of temporofrontal regions during nausea.

COORDINATION OF EMESIS

Brain stem nuclei—including the nucleus tractus solitarius; dorsal vagal and phrenic nuclei; medullary nuclei that regulate respiration; and nuclei that control pharyngeal, facial, and tongue movements—coordinate the initiation of emesis. Neurotransmitters involved in this coordination are uncertain, but neurokinin NK_1, serotonin5-HT_3, and vasopressin pathways may participate.

Somatic and visceral muscles exhibit stereotypic responses during emesis. Inspiratory thoracic and abdominal wall muscles contract, producing high intrathoracic and intraabdominal pressures that facilitate expulsion of gastric contents. The gastric cardia herniates across the diaphragm and the larynx moves upward to promote oral propulsion of the vomitus. Under normal conditions, distally migrating gut contractions are regulated by an electrical phenomenon, the slow wave, which cycles at 3 cycles/min in the stomach and 11 cycles/min in the duodenum. With emesis, there is slow-wave abolition and initiation of orally propagating spikes that evoke retrograde contractions that assist in oral expulsion of intestinal contents.

ACTIVATORS OF EMESIS

Emetic stimuli act at several sites. Emesis provoked by unpleasant thoughts or smells originates in the cerebral cortex, whereas cranial nerves mediate vomiting after gag reflex activation. Motion sickness and inner ear disorders act on the labyrinthine apparatus, whereas gastric irritants and cytotoxic agents such as cisplatin stimulate gastroduodenal vagal afferent nerves. Nongastric visceral afferents are activated by intestinal and colonic obstruction and mesenteric ischemia. The area postrema, a medullary nucleus, responds to bloodborne emetic stimuli and is termed the *chemoreceptor trigger zone*. Many emetogenic drugs act on the area postrema, as do bacterial toxins and metabolic factors produced during uremia, hypoxia, and ketoacidosis.

Neurotransmitters that mediate induction of vomiting are selective for these anatomic sites. Labyrinthine disorders stimulate vestibular muscarinic M_1 and histaminergic H_1 receptors, whereas vagal afferent stimuli activate serotonin5-HT_3 receptors. The area postrema is richly served by nerves acting on 5-HT_3, M_1, H_1, and dopamine D_2 subtypes. Transmitters in the cerebral cortex are poorly understood, although cannabinoid CB_1 pathways may participate. Optimal pharmacologic therapy of vomiting requires understanding of these pathways.

Differential Diagnosis

Nausea and vomiting are caused by conditions within and outside the gut as well as by drugs and circulating toxins **(Table 39–1)**.

Table 39–1 Causes of Nausea and Vomiting

Intraperitoneal	Extraperitoneal	Medications/Metabolic Disorders
Obstructing disorders	Cardiopulmonary disease	Drugs
Pyloric obstruction	Cardiomyopathy	Cancer chemotherapy
Small bowel obstruction	Myocardial infarction	Antibiotics
Colonic obstruction	Labyrinthine disease	Cardiac antiarrhythmics
Superior mesenteric artery syndrome	Motion sickness	Digoxin
Enteric infections	Labyrinthitis	Oral hypoglycemics
Viral	Malignancy	Oral contraceptives
Bacterial	Intracerebral disorders	Endocrine/metabolic disease
Inflammatory diseases	Malignancy	Pregnancy
Cholecystitis	Hemorrhage	Uremia
Pancreatitis	Abscess	Ketoacidosis
Appendicitis	Hydrocephalus	Thyroid and parathyroid disease
Hepatitis	Psychiatric illness	Adrenal insufficiency
Altered sensorimotor function	Anorexia and bulimia nervosa	Toxins
Gastroparesis	Depression	Liver failure
Intestinal pseudoobstruction	Postoperative vomiting	Ethanol
Gastroesophageal reflux		
Chronic idiopathicnausea		
Functional vomiting		
Cyclic vomiting syndrome		
Biliary colic		
Abdominal irradiation		

INTRAPERITONEAL DISORDERS

Visceral obstruction and inflammation of hollow and solid viscera may produce vomiting. Gastric obstruction results from ulcer disease and malignancy, while small-bowel and colonic obstruction occur because of adhesions, benign or malignant tumors, volvulus, intussusception, or inflammatory diseases such as Crohn's disease. The superior mesenteric artery syndrome, occurring after weight loss or prolonged bed rest, results when the duodenum is compressed by the overlying superior mesenteric artery. Abdominal irradiation impairs intestinal motor function and induces strictures. Biliary colic causes nausea via action on visceral afferent nerves. Vomiting with pancreatitis, cholecystitis, and appendicitis is due to visceral irritation and induction of ileus. Enteric infections with viruses or bacteria such as *Staphylococcus aureus* and *Bacillus cereus* commonly cause vomiting, especially in children. Opportunistic infections such as cytomegalovirus or herpes simplex virus induce emesis in immunocompromised individuals.

Disordered gut sensorimotor function commonly causes nausea and vomiting. *Gastroparesis* is defined as a delay in gastric emptying of food and occurs after vagotomy, with pancreatic adenocarcinoma, with mesenteric vascular insufficiency, or in systemic diseases such as diabetes, scleroderma, and amyloidosis. The most common form of disease, idiopathic gastroparesis, occurs in the absence of systemic illness and may follow a viral prodrome, suggesting an infectious etiology. Intestinal pseudoobstruction is characterized by disrupted intestinal and colonic motor activity and leads to retention of food residue and secretions; bacterial overgrowth; nutrient malabsorption; and symptoms of nausea, vomiting, bloating, pain, and altered defecation. *Intestinal pseudoobstruction* may be idiopathic or inherited as a familial visceral myopathy or neuropathy, or it may result from systemic disease or as a paraneoplastic complication of a malignancy such as small cell lung carcinoma. Patients with gastroesophageal reflux may report nausea and vomiting, as do some individuals with irritable bowel syndrome (IBS).

Other functional disorders without organic abnormalities have been characterized in adults. *Chronic idiopathicnausea* is defined as nausea without vomiting occurring several times weekly, whereas *functional vomiting* is defined as one or more vomiting episodes weekly in the absence of an eating disorder or psychiatric disease. *Cyclic vomiting syndrome* is a rare disorder of unknown etiology that

produces periodic discrete episodes of relentless nausea and vomiting. The syndrome shows a strong association with migraine headaches, suggesting that some cases may be migraine variants. Cyclic vomiting is most common in children, although adult cases have been described in association with rapid gastric emptying and with chronic cannabis use.

EXTRAPERITONEAL DISORDERS

Myocardial infarction and congestive heart failure may cause nausea and vomiting. Postoperative emesis occurs after 25% of surgeries, most commonly laparotomy and orthopedic surgery, and is more prevalent in women. Increased intracranial pressure from tumors, bleeding, abscess, or obstruction to cerebrospinal fluid outflow produces prominent vomiting with or without nausea. Motion sickness, labyrinthitis, and Méni`re's disease evoke emesis via labyrinthine pathways. Patients with psychiatric illnesses including anorexia nervosa, bulimia nervosa, anxiety, and depression may report significant nausea that may be associated with delayed gastric emptying.

MEDICATIONS AND METABOLIC DISORDERS

Drugs evoke vomiting by action on the stomach (analgesics, erythromycin) or area postrema (digoxin, opiates, anti-Parkinsonian drugs). Emetogenic agents include antibiotics, cardiac antiarrhythmics, antihypertensives, oral hypoglycemics, and contraceptives. Cancer chemotherapy causes vomiting that is acute (within hours of administration), delayed (after one or more days), or anticipatory. Acute emesis resulting from highly emetogenic agents such as cisplatin is mediated by 5-HT_3 pathways, whereas delayed emesis is 5-HT_3-independent. Anticipatory nausea often responds better to anxiolytic therapy than to antiemetics.

Several metabolic disorders elicit nausea and vomiting. Pregnancy is the most prevalent endocrinologic cause of nausea, which affects 70% of women in the first trimester. Hyperemesis gravidarum is a severe form of nausea of pregnancy that can produce significant fluid loss and electrolyte disturbances. Uremia, ketoacidosis, and adrenal insufficiency, as well as parathyroid and thyroid disease, are other metabolic causes of emesis.

Circulating toxins evoke emesis via effects on the area postrema. Endogenous toxins are generated in fulminant liver failure, whereas exogenous enterotoxins may be produced by enteric bacterial infection. Ethanolintoxication is a common toxic etiology of nausea and vomiting.

Approach to the Patient: Nausea and Vomiting

HISTORY AND PHYSICAL EXAMINATION

The history helps define the etiology of unexplained nausea and vomiting. Drugs, toxins, and gastrointestinal infections commonly cause acute symptoms, whereas established illnesses evoke chronic complaints. Pyloric obstruction and gastroparesis produce vomiting within one hour of eating, whereas emesis from intestinal obstruction occurs later. In severe cases of gastroparesis, the vomitus may contain food residue ingested hours or days previously. Hematemesis raises suspicion of an ulcer, malignancy, or Mallory-Weiss tear, whereas feculent emesis is noted with distal intestinal or colonic obstruction. Bilious vomiting excludes gastric obstruction, while emesis of undigested food is consistent with a Zenker's diverticulum or achalasia. Relief of abdominal pain by emesis characterizes intestinal obstruction, whereas vomiting has no effect on pancreatitis or cholecystitis pain. Pronounced weight loss raises concern about malignancy or obstruction. Fevers suggest inflammation; an intracranial source is considered if there are headaches or visual field changes. Vertigo or tinnitus indicates labyrinthine disease.

The physical examination complements information from the history. Orthostatic hypotension and reduced skin turgor indicate intravascular fluid loss. Pulmonary abnormalities raise concern for aspiration of vomitus. Abdominal auscultation may reveal absent bowel sounds with ileus. High-pitched rushes suggest bowel obstruction, while a succussion splash upon abrupt lateral movement of the patient is found with gastroparesis or pyloric obstruction. Tenderness or involuntary guarding raises suspicion of inflammation, whereas fecal blood suggests mucosal injury from ulcer, ischemia, or tumor. Neurologic disease presents with papilledema, visual field loss, or focal neural abnormalities. Neoplasm is suggested by palpation of masses or adenopathy.

DIAGNOSTIC TESTING

For intractable symptoms or an elusive diagnosis, selected screening tests can direct clinical care. Electrolyte replacement is indicated for hypokalemia or metabolic alkalosis. Detection of iron-deficiency anemia mandates a search for mucosal injury. Pancreaticobiliary disease is indicated by abnormal pancreatic or liver biochemistries, whereas endocrinologic, rheumatologic, or paraneoplastic etiologies are suggested by hormone or serologic abnormalities. If bowel obstruction is suspected, supine and upright abdominal radiographs may show intestinal air-fluid levels with reduced colonic air. Ileus is characterized by diffusely dilated air-filled bowel loops.

Anatomic studies may be indicated if initial testing is nondiagnostic. Upper endoscopy detects ulcers or malignancy, while small-bowel barium radiography diagnoses partial intestinal obstruction. Colonoscopy or contrast enema radiography can detect colonic obstruction. Ultrasound or CT defines

intraperitoneal inflammatory processes, while CT or MRI of the head can delineate intracranial disease. Advances in CT and MRI enterography have improved definition of bowel inflammation, as in Crohn's disease. Mesenteric angiography, CT, or MRI is useful for suspected ischemia.

Gastrointestinal motility testing may detect a motor disorder that contributes to symptoms when anatomic abnormalities are absent. Gastroparesis commonly is diagnosed using gastric scintigraphy, by which emptying of a radiolabeled meal is measured. Isotopic breath tests and wireless motility capsule methods have been validated and may become important alternatives to scintigraphy to define gastroparesis. The diagnosis of intestinal pseudoobstruction often is suggested by abnormal barium transit and luminal dilation on small-bowel contrast radiography. Delayed small-bowel transit also may be detected by wireless capsule techniques. Small-intestinal manometry can confirm the diagnosis and further characterize the motor abnormality as neuropathic or myopathic based on contractile patterns. Such investigation can obviate the need for open intestinal biopsy to evaluate for smooth muscle or neuronal degeneration.

Treatment: Nausea and Vomiting

GENERAL PRINCIPLES

Therapy of vomiting is tailored to correcting medically or surgically remediable abnormalities if possible. Hospitalization is considered for severe dehydration, especially if oral fluid replenishment cannot be sustained. Once oral intake is tolerated, nutrients are restarted with liquids that are low in fat, as lipids delay gastric emptying. Foods high in indigestible residues are avoided because these also prolong gastric retention.

ANTIEMETIC MEDICATIONS

The most commonly used antiemetic agents act on sites in the central nervous system **(Table 39–2)**. Antihistamines such as meclizine and dimenhydrinate and anticholinergic drugs like scopolamine act on labyrinthine pathways and are useful in motion sickness and inner ear disorders. Dopamine D_2 antagonists treat emesis evoked by area postrema stimuli and are useful for medication, toxic, and metabolic etiologies. Dopamine antagonists freely cross the blood-brain barrier and cause anxiety, dystonic reactions, hyperprolactinemic effects (galactorrhea and sexual dysfunction), and irreversible tardivedyskinesia.

Table 39–2 Treatment of Nausea and Vomiting

Treatment	Mechanism	Examples	Clinical Indications
Antiemetic agents	Antihistaminergic	Dimenhydrinate, meclizine	Motion sickness, inner ear disease
	Anticholinergic	Scopolamine	Motion sickness, inner ear disease
	Antidopaminergic 5-HT_3 antagonist	Prochlorperazine, thiethylperazine	Medication-, toxin-, or metabolic-induced emesis
	NK_1 antagonist	Ondansetron, granisetron	Chemotherapy- and radiation-induced emesis, postoperative emesis
	Tricyclic antidepressant	Aprepitant	Chemotherapy-induced nausea and vomiting
	Other antidepressant	Amitriptyline, nortriptyline	Chronic idiopathic nausea, functional vomiting, cyclic vomiting syndrome, ? gastroparesis
		Mirtazapine	?Functional vomiting, ?gastroparesis
Prokinetic agents	5-HT_4 agonist and antidopaminergic	Metoclopramide	Gastroparesis
	Motilin agonist	Erythromycin	Gastroparesis, ?intestinal pseudoobstruction
	Peripheral antidopaminergic	Domperidone	Gastroparesis
	Somatostatin analogue	Octreotide	Intestinal pseudoobstruction
	Acetylcholinesterase inhibitor	Pyridostigmine	?Small intestinal dysmotility/pseudoobstruction
Special settings	Benzodiazepines	Lorazepam	Anticipatory nausea and vomiting with chemotherapy
	Glucocorticoids	Methylprednisolone, dexamethasone	Chemotherapy-induced emesis
	Cannabinoids	Tetrahydrocannabinol	Chemotherapy-induced emesis

Note:?, indication is uncertain.

Other drug classes exhibit antiemetic properties. Serotonin5-HT_3 antagonists such as ondansetron and granisetron exhibit utility in postoperative vomiting, after radiation therapy, and for preventing cancer chemotherapy–induced emesis. The usefulness of 5-HT_3 antagonists for other causes of emesis is less well established. Low-dose tricyclic antidepressant agents provide symptomatic benefit in patients with chronic idiopathicnausea and functional vomiting as well as in diabetic patients with nausea and vomiting whose disease is of long standing. Other antidepressants such as mirtazapine also may exhibit antiemetic effects.

GASTROINTESTINAL MOTOR STIMULANTS

Drugs that stimulate gastric emptying are indicated for gastroparesis **(Table 39–2).** Metoclopramide, a combined 5-HT_4 agonist and D_2 antagonist, exhibits efficacy in gastroparesis, but antidopaminergic side effects, particularly tardivedyskinesia, limit its use in <25% of patients. Erythromycin, a macrolide antibiotic, increases gastroduodenal motility by action on receptors for motilin, an endogenous stimulant of fasting motor activity. Intravenous erythromycin is useful for inpatients with refractory gastroparesis; however, oral forms also have some utility. Domperidone, a D_2 antagonist not available in the United States, exhibits prokinetic and antiemetic effects but does not cross into most other brain regions; thus, anxiety and dystonic reactions are rare. The main side effects of domperidone relate to induction of hyperprolactinemia via effects on pituitary regions served by a porous blood-brain barrier.

Refractory upper gut motility disorders pose significant challenges. Liquid suspensions of prokinetic drugs may be beneficial, because liquids empty from the stomach more rapidly than pills do. Metoclopramide can be administered subcutaneously in patients unresponsive to oral drugs. Intestinal pseudoobstruction may respond to the somatostatin analogue octreotide, which induces propagative small intestinal motor complexes. Acetylcholinesterase inhibitors such as pyridostigmine are anecdotally observed to benefit some patients with small bowel dysmotility. Pyloric injections of botulinum toxin are reported in uncontrolled studies to benefit patients with gastroparesis. Placement of a feeding jejunostomy reduces hospitalizations and improves overall health in some patients with drug-refractory gastroparesis. Surgical options are limited for unresponsive cases, but postvagotomy gastroparesis may improve with near-total resection of the stomach. Implanted gastric electrical stimulators may reduce symptoms, enhance nutrition, improve quality of life, and decrease health care expenditures in medication-refractory gastroparesis, although small controlled trials report only modest benefits with this method.

SELECTED CLINICAL SETTINGS

Some cancer chemotherapeutic agents such as cisplatin are intensely emetogenic **(Chap. 85)**. Given prophylactically, 5-HT_3 antagonists prevent chemotherapy-induced acute vomiting in most cases **(Table 39–2).** Optimal antiemetic effects often are obtained with a 5-HT_3 antagonist combined with a glucocorticoid. Benzodiazepines such as lorazepam are useful to reduce anticipatory nausea and vomiting. Therapy of delayed emesis 1–5 days after chemotherapy is less successful. Neurokinin NK_1 antagonists (e.g., aprepitant) exhibit antiemetic and antinausea effects during both the acute and delayed periods after chemotherapy. Cannabinoids such as tetrahydrocannabinol, long advocated for cancer-associated emesis, produce significant side effects and exhibit no more efficacy than antidopaminergic agents. Most antiemetic regimens produce greater reductions in vomiting than in nausea.

The clinician should exercise caution in managing the pregnant patient with nausea. Studies of the teratogenic effects of available antiemetic agents provide conflicting results. Few controlled trials have been performed in nausea of pregnancy, although antihistamines such as meclizine and antidopaminergics such as prochlorperazine demonstrate efficacy greater than placebo. Some obstetricians offer alternative therapies such as pyridoxine, acupressure, or ginger.

Controlling emesis in cyclic vomiting syndrome is a challenge. In many patients, prophylaxis with tricyclic antidepressants, cyproheptadine, or β-adrenoceptor antagonists can reduce the frequency of attacks. Intravenous 5-HT_3 antagonists combined with the sedating effects of a benzodiazepine such as lorazepam are a mainstay of treatment of acute symptom flares. Small studies report benefits with antimigraine therapies, including the serotonin 5-HT_1 agonist sumatriptan, as well as selected anticonvulsant drugs such as zonisamide and levetiracetam.

INDIGESTION

Mechanisms

The most common causes of indigestion are gastroesophageal reflux and functional dyspepsia. Other cases are a consequence of a more serious organic illness.

GASTROESOPHAGEAL REFLUX

Gastroesophageal reflux can result from a variety of physiologic defects. Reduced lower esophageal sphincter (LES) tone is an important cause of reflux in scleroderma and pregnancy; it may also be a factor in patients without other systemic conditions. Many individuals exhibit frequent transient LES

relaxations during which acid or nonacidic fluid bathes the esophagus. Overeating and aerophagia can transiently override the barrier function of the LES, whereas impaired esophageal body motility and reduced salivary secretion prolong fluid exposure. The role of hiatal hernias is controversial—although most reflux patients exhibit hiatal hernias, most individuals with hiatal hernias do not have excess heartburn.

GASTRIC MOTOR DYSFUNCTION

Disturbed gastric motility is purported to cause gastroesophageal reflux in some cases of indigestion. Delayed gastric emptying is also found in 25–50% of functional dyspeptics. The relation of these defects to symptom induction is uncertain; studies show poor correlation between symptom severity and degrees of motor dysfunction. Impaired gastric fundus relaxation after eating may underlie selected dyspeptic symptoms like bloating, nausea, and early satiety.

VISCERAL AFFERENT HYPERSENSITIVITY

Disturbed gastric sensory function is proposed as a pathogenic factor in functional dyspepsia. Visceral afferent hypersensitivity was first demonstrated in patients with IBS who had heightened perception of rectal balloon inflation without changes in rectal compliance. Similarly, dyspeptic patients experience discomfort with fundic distention to lower pressures than healthy controls. Some patients with heartburn exhibit no increase in reflux of acid or nonacidic fluid. These individuals with functional heartburn are believed to have heightened perception of normal esophageal pH and volume.

OTHER FACTORS

Helicobacter pylori has a clear etiologic role in peptic ulcer disease, but ulcers cause a minority of cases of dyspepsia. *H. pylori* is considered to be a minor factor in the genesis of functional dyspepsia. In contrast, functional dyspepsia is associated with a reduced sense of physical and mental well-being and is exacerbated by stress, suggesting important roles for psychological factors. Analgesics cause dyspepsia, while nitrates, calcium channel blockers, theophylline, and progesterone promote gastroesophageal reflux. Other stimuli that induce reflux include ethanol, tobacco, and caffeine via LES relaxation. Genetic factors may promote development of reflux.

Differential Diagnosis

GASTROESOPHAGEAL REFLUX DISEASE

Gastroesophageal reflux disease (GERD) is prevalent in Western society. Heartburn is reported once monthly by 40% of Americans and daily by 7–10%. Most cases of heartburn occur because of excess acid reflux, although reflux of non-acidic fluid may produce similar symptoms. Alkaline reflux esophagitis produces GERD-like symptoms most often in patients who have had surgery for peptic ulcer disease. Approximately 10% of patients with heartburn of a functional nature exhibit normal degrees of esophageal acid exposure and no increase in nonacidic reflux.

FUNCTIONAL DYSPEPSIA

Nearly 25% of the populace has dyspepsia at least 6 times yearly, but only 10–20% of these individuals present to physicians. Functional dyspepsia, the cause of symptoms in 60% of dyspeptic patients, is defined as ≥3 months of bothersome postprandial fullness, early satiety, or epigastric pain or burning with symptom onset at least 6 months before diagnosis in the absence of organic cause. Most cases follow a benign course, but some patients with *H. pylori* infection or on nonsteroidal anti-inflammatory drugs (NSAIDs) develop ulcers. As with idiopathic gastroparesis, some cases of functional dyspepsia result from prior gastrointestinal infection.

ULCER DISEASE

In most cases of GERD, there is no destruction of the esophagus. However, 5% of patients develop esophageal ulcers, and some form strictures. Symptoms do not reliably distinguish nonerosive from erosive or ulcerative esophagitis. Some 15–25% of cases of dyspepsia stem from ulcers of the stomach or duodenum. The most common causes of ulcer disease are gastric infection with *H. pylori* and use of NSAIDs. Other rare causes of gastroduodenal ulcer include Crohn's disease **(Chap. 295)** and Zollinger-Ellison syndrome **(Chap. 293),** a condition resulting from gastrin overproduction by an endocrine tumor.

MALIGNANCY

Dyspeptic patients often seek care because of fear of cancer. However, <2% of cases result from gastroesophageal malignancy. Esophageal squamous cell carcinoma occurs most often in those with histories of tobacco or ethanol intake. Other risk factors include prior caustic ingestion, achalasia, and the hereditary disorder tylosis. Esophageal adenocarcinoma usually complicates long-standing acid reflux. Between 8 and 20% of GERD patients exhibit intestinal metaplasia of the esophagus, termed *Barrett's metaplasia*. This condition predisposes to esophageal adenocarcinoma **(Chap. 91).** Gastric malignancies include adenocarcinoma, which is prevalent in certain Asian societies, and lymphoma.

OTHER CAUSES

Opportunistic fungal or viral esophageal infections may produce heartburn or chest discomfort but

more often cause odynophagia. Other causes of esophageal inflammation include eosinophilic esophagitis and pill esophagitis. Biliary colic is in the differential diagnosis of dyspepsia, but most patients with true biliary colic report discrete episodes of right upper quadrant or epigastric pain rather than chronic burning discomfort, nausea, and bloating. Intestinal lactase deficiency produces gas, bloating, discomfort, and diarrhea after lactose ingestion. Lactase deficiency occurs in 15–25% of whites of northern European descent but is more common in blacks and Asians. Intolerance of other carbohydrates (e.g., fructose, sorbitol) produces similar symptoms. Small-intestinal bacterial overgrowth may produce dyspepsia, often with bowel dysfunction, distention, and malabsorption. Eosinophilic infiltration of the duodenal mucosa is described in some cases of dyspepsia. Pancreatic disease (chronic pancreatitis and malignancy), hepatocellular carcinoma, celiac disease, Ménétrier's disease, infiltrative diseases (sarcoidosis and eosinophilic gastroenteritis), mesenteric ischemia, thyroid and parathyroid disease, and abdominal wall strain cause dyspepsia. Extraperitoneal etiologies of indigestion include congestive heart failure and tuberculosis. Investigation is ongoing into genetic markers that predispose to developing functional dyspepsia.

Approach to the Patient: Indigestion

HISTORY AND PHYSICAL EXAMINATION

Care of the patient with indigestion requires a thorough interview. GERD classically produces heartburn, a substernal warmth in the epigastrium that moves toward the neck. Heartburn often is exacerbated by meals and may awaken the patient. Associated symptoms include regurgitation of acid or nonacidic fluid and water brash, the reflex release of salty salivary secretions into the mouth. Atypical symptoms include pharyngitis, asthma, cough, bronchitis, hoarseness, and chest pain that mimics angina. Some patients with acid reflux on esophageal pH testing do not report heartburn, but note abdominal pain or other symptoms.

Some patients with dyspepsia report a predominance of epigastric pain or burning that is intermittent and not generalized or localized to other regions. Others experience a postprandial distress syndrome characterized by fullness occurring after normal-sized meals and early satiety that prevents completion of regular meals, with associated bloating, belching, or nausea. Functional dyspepsia overlaps with other functional disorders such as IBS.

The physical exam with GERD and functional dyspepsia usually is normal. In atypical GERD, pharyngeal erythema and wheezing may be noted. Recurrent acid regurgitation may cause poor dentition. Functional dyspeptics may report epigastric tenderness or distention.

Discrimination between functional and organic causes of indigestion mandates exclusion of selected historic and examination features. Odynophagia suggests esophageal infection, while dysphagia is worrisome for a benign or malignant esophageal blockage. Other alarming features include unexplained weight loss, recurrent vomiting, occult or gross gastrointestinal bleeding, jaundice, a palpable mass or adenopathy, and a family history of gastrointestinal malignancy.

DIAGNOSTIC TESTING

Because indigestion is prevalent and most cases result from GERD or functional dyspepsia, a general principle is to perform only limited and directed diagnostic testing of selected individuals.

Once alarm factors are excluded **(Table 39–3)**, patients with typical GERD do not need further evaluation and are treated empirically. Upper endoscopy is indicated to exclude mucosal injury in cases with atypical symptoms, symptoms unresponsive to acid suppressing drugs, or alarm factors. For heartburn <5 years in duration, especially in patients <50 years old, endoscopy is recommended to screen for Barrett's metaplasia. However, the clinical benefits and cost-effectiveness of this approach have not been validated in controlled studies. Ambulatory esophageal pH testing using a catheter method or an implanted esophageal capsule device is considered for drug-refractory symptoms and atypical symptoms like unexplained chest pain. Esophageal manometry most commonly is ordered when surgical treatment of GERD is considered. A low LES pressure may predict failure of drug therapy and helps select patients who may require surgery. Demonstration of disordered esophageal body peristalsis may affect the decision to operate or modify the type of operation chosen. High-resolution manometric methods improve characterization of ineffective esophageal propulsion, which may contribute to impaired esophageal acid clearance in some GERD patients. Manometry with provocative testing may clarify the diagnosis in patients with atypical symptoms. Blind perfusion of saline and then acid into the esophagus, known as the *Bernstein test*, can delineate whether unexplained chest discomfort results from acid reflux. Nonacidic reflux may be suggested by nuclear medicine reflux scanning or detected by combined esophageal impedance-pH testing, which increases the diagnostic yield by 15% versus pH testing alone. Ambulatory measurement of esophageal bilirubin levels facilitates diagnosis of alkaline reflux.

Table 39–3 Alarm Symptoms in GERD
Odynophagia
Unexplained weight loss

Recurrent vomiting
Occult or gross gastrointestinal bleeding
Jaundice
Palpable mass or adenopathy
Family history of gastrointestinal malignancy

Upper endoscopy is performed as the initial diagnostic test in patients with unexplained dyspepsia who are >55 years old or who have alarm factors because of the elevated risks of malignancy and ulcer in these groups. The management approach to patients <55 years old without alarm factors is dependent on the local prevalence of *H. pylori* infection. For individuals in regions with low *H. pylori* prevalence (<10%), a 4-week trial of an acid-suppressing medication such as a proton pump inhibitor is recommended. If this fails, a "test and treat" approach is most commonly applied. *H. pylori* status is determined with urea breath testing, stool antigen measurement, or blood serology testing. Those who are *H. pylori* positive are given therapy to eradicate the infection. If symptoms resolve on either regimen, no further intervention is required. For patients in areas with high *H. pylori* prevalence (>10%), an initial test and treat approach is advocated, with a subsequent trial of an acid-suppressing regimen offered for those in whom *H. pylori* treatment fails or for those who are negative for the infection. In each of these patient subsets, upper endoscopy is reserved for those whose symptoms fail to respond to therapy.

Further testing is indicated if other factors are present. If bleeding is reported, a blood count is obtained to exclude anemia. Thyroid chemistries or calcium levels screen for metabolic disease, whereas serologies may suggest celiac disease. For possible pancreaticobiliary causes, pancreatic and liver chemistries are obtained. If abnormalities are found, ultrasound or CT may give important information. Gastric emptying measurement is considered to exclude gastroparesis in patients whose dyspeptic symptoms resemble postprandial distress when drug therapy fails. Gastric scintigraphy also assesses for gastroparesis in patients with GERD, especially if surgical intervention is being considered. Breath testing after carbohydrate ingestion may detect lactase deficiency, intolerance to other carbohydrates, or small-intestinal bacterial overgrowth.

Treatment: General Principles

For mild indigestion, reassurance that a careful evaluation revealed no serious organic disease may be the only intervention needed. Drugs that cause gastroesophageal reflux or dyspepsia should be stopped, if possible. Patients with GERD should limit ethanol, caffeine, chocolate, and tobacco use because of their effects on the LES. Other measures in GERD include ingesting a low-fat diet, avoiding snacks before bedtime, and elevating the head of the bed.

Specific therapies for organic disease should be offered when possible. Surgery is appropriate in disorders like biliary colic, while diet changes are indicated for lactase deficiency or celiac disease. Some illnesses such as peptic ulcer disease may be cured by specific medical regimens. However, because most indigestion is caused by GERD or functional dyspepsia, medications that reduce gastric acid, modulate motility, or blunt gastric sensitivity are indicated.

ACID-SUPPRESSING OR NEUTRALIZING MEDICATIONS

Drugs that reduce or neutralize gastric acid are often prescribed for GERD. Histamine H_2 antagonists such as cimetidine, ranitidine, famotidine, and nizatidine are useful in mild to moderate GERD. For severe symptoms or for many cases of erosive or ulcerative esophagitis, proton pump inhibitors such as omeprazole, lansoprazole, rabeprazole, pantoprazole, esomeprazole, or dexlansoprazole are needed. These drugs, which inhibit gastric H^+, K^+-ATPase, are more potent than H_2 antagonists. Up to one third of GERD patients do not respond to proton pump inhibitors; one third of these patients have nonacidic reflux while 10% have persistent acid-related disease. Acid suppressants may be taken continuously or on demand depending on symptom severity. Infrequent potential complications of long-term proton pump inhibitors may include infection, small intestinal bacterial overgrowth, nutrient deficiency (vitamin B12, iron, calcium), bone demineralization, and impaired medication absorption (e.g. clopidogrel). Many patients started on a proton pump inhibitor can be stepped down to an H2 antagonist. Combining a proton pump inhibitor with an H_2 antagonist is provided for some refractory cases.

Acid-suppressing drugs are also effective in appropriately selected patients with functional dyspepsia. Meta-analysis of eight controlled trials calculated a risk ratio of 0.86, with a 95% confidence interval of 0.78–0.95, favoring proton pump inhibitor therapy over placebo. The benefits of less potent acid reducing therapies such as H_2 antagonists are unproven.

Liquid antacids are useful for short-term control of mild GERD but are less effective for severe disease unless given at high doses that elicit side effects (diarrhea and constipation with magnesium- and aluminum-containing agents, respectively). Alginic acid in combination with antacids may form a floating barrier to acid reflux in individuals with upright symptoms. Sucralfate is a salt of aluminum

hydroxide and sucrose octasulfate that buffers acid and binds pepsin and bile salts. Its efficacy in GERD is felt to be comparable to that of H_2 antagonists.

HELICOBACTER PYLORI ERADICATION

H. pylori eradication is clearly indicated only for peptic ulcer and mucosa-associated lymphoid tissue gastric lymphoma. The utility of eradication therapy in functional dyspepsia is less well established, but <15% of cases relate to this infection. Meta-analysis of 13 controlled trials calculated a risk ratio of 0.91, with a 95% confidence interval of 0.87–0.96, favoring *H. pylori* eradication therapy over placebo. Several drug combinations show efficacy in eliminating the infection **(Chap. 293);** most include 10–14 days of a proton pump inhibitor or bismuth subsalicylate in concert with two antibiotics. *H. pylori* infection is associated with reduced prevalence of GERD, especially in the elderly. However, eradication of the infection does not worsen GERD symptoms. To date, no consensus recommendations regarding *H. pylori* eradication in GERD patients have been offered.

AGENTS THAT MODIFY GASTROINTESTINAL MOTOR ACTIVITY

Motor stimulants (also known as prokinetics) such as metoclopramide, erythromycin, and domperidone have limited utility in GERD. Several studies have evaluated the effectiveness of motor-stimulating drugs in functional dyspepsia; however, convincing evidence of their benefits has not been found. Some clinicians suggest that patients with symptoms resembling postprandial distress may respond preferentially to prokinetic drugs. The γ-aminobutyric acid B (GABA-B) agonist baclofen reduces esophageal exposure to acid and non-acidic fluids by inhibiting transient LES relaxations; this drug is proposed for refractory acid and non-acid reflux.

OTHER OPTIONS

Antireflux surgery (fundoplication) is most often offered to GERD patients who are young and may require lifelong therapy, have typical heartburn and regurgitation, and are responsive to proton pump inhibitors. Surgery also is effective for some cases of non-acidic reflux. Individuals who may respond less well to operative therapy include those with atypical symptoms and those who have esophageal body motor disturbances. Fundoplications are performed laparoscopically when possible and include the Nissen and Toupet procedures in which the proximal stomach is partly or completely wrapped around the distal esophagus to increase LES pressure. Dysphagia, gas-bloat syndrome, and gastroparesis may be long-term complications of these procedures. The utility and safety of endoscopic therapies for increasing the barrier function of the gastroesophageal junction, including radiofrequency energy delivery and gastroplication, have not been fully investigated for patients with refractory GERD.

Some patients with functional heartburn and functional dyspepsia refractory to standard therapies may respond to low-dose antidepressants in tricyclic and other classes. Their mechanism of action is unknown but may involve blunting of visceral pain processing in the brain. Gas and bloating are among the most troubling symptoms in some patients with indigestion and can be difficult to treat. Dietary exclusion of gas-producing foods such as legumes and use of simethicone or activated charcoal provide benefits in some cases. Therapies that modify gut flora, including antibiotics and probiotic preparations containing active bacterial cultures, are useful for cases of bacterial overgrowth and functional lower gastrointestinal disorders, but their utility in functional dyspepsia is unproven. Psychological treatments may be offered for refractory functional dyspepsia, but no convincing data suggest their efficacy.

FURTHER READINGS

Camilleri M, Tack JF: Current medical treatments of dyspepsia and irritable bowel syndrome. Gastroenterol Clin North Am 39:481, 2010 DeVault KR, Talley NJ: Insights into the future of gastric acid suppression. Nature Rev Gastroenterol Hepatol 6: 524, 2009

Hasler WL: Methods of gastric electrical stimulation and pacing: a review of their benefits and mechanisms of action in gastroparesis and obesity. Neurogastroenterol Motil 21: 229, 2009[PMID: 19254353]

Hesketh PJ: Chemotherapy-induced nausea and vomiting. N Engl J Med 358: 2482, 2008[PMID: 18525044]

Kahrilas PJ, Sifrim D: High-resolution manometry and impedance-pH/manometry: valuable tools in clinical and investigational esophagology. Gastroenterology 135: 756, 2008[PMID: 18639550]

——— et al: American Gastroenterological Association Institute technical review on the management of gastroesophageal reflux disease. Gastroenterology 135: 1392, 2008

Olden KW, Chepyala P: Functional nausea and vomiting. Nature Clin Pract Gastroenterol Hepatol 5: 202, 2008[PMID: 18301404]

Parkman HP, Jones MP: Tests of gastric neuromuscular function. Gastroenterology 136: 1526, 2009[PMID: 19293005]

Shafi MA, Bresalier RS: The gastrointestinal complications of oncologic therapy. Gastroenterol Clin North Am 39:629, 2010[PMID: 20951921]

Stapleton J, WO JM: Current treatment of nausea and vomiting associated with gastroparesis: antiemetics, prokinetics, tricyclics. Gastroenterol Clin North Am 19: 57, 2009[PMID: 19232281]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIARRHEA AND CONSTIPATION: INTRODUCTION

Diarrhea and constipation are exceedingly common and, together, exact an enormous toll in terms of mortality, morbidity, social inconvenience, loss of work productivity, and consumption of medical resources. Worldwide, >1 billion individuals suffer one or more episodes of acute diarrhea each year. Among the 100 million persons affected annually by acute diarrhea in the United States, nearly half must restrict activities, 10% consult physicians, ~250,000 require hospitalization, and ~5000 die (primarily the elderly). The annual economic burden to society may exceed $20 billion. Acute infectious diarrhea remains one of the most common causes of mortality in developing countries, particularly among children, accounting for 2–3 million deaths per year. Constipation, by contrast, is rarely associated with mortality and is exceedingly common in developed countries, leading to frequent self-medication and, in a third of those, to medical consultation. Population statistics on chronic diarrhea and constipation are more uncertain, perhaps due to variable definitions and reporting, but the frequency of these conditions is also high. United States population surveys put prevalence rates for chronic diarrhea at 2–7% and for chronic constipation at 12–19%, with women being affected twice as often as men. Diarrhea and constipation are among the most common patient complaints faced by internists and primary care physicians, and they account for nearly 50% of referrals to gastroenterologists.

Although diarrhea and constipation may present as mere nuisance symptoms at one extreme, they can be severe or life-threatening at the other. Even mild symptoms may signal a serious underlying gastrointestinal lesion, such as colorectal cancer, or systemic disorder, such as thyroid disease. Given the heterogeneous causes and potential severity of these common complaints, it is imperative for clinicians to appreciate the pathophysiology, etiologic classification, diagnostic strategies, and principles of management of diarrhea and constipation, so that rational and cost-effective care can be delivered.

NORMAL PHYSIOLOGY

While the primary function of the small intestine is the digestion and assimilation of nutrients from food, the small intestine and colon together perform important functions that regulate the secretion and absorption of water and electrolytes, the storage and subsequent transport of intraluminal contents aborally, and the salvage of some nutrients after bacterial metabolism of carbohydrate that are not absorbed in the small intestine. The main motor functions are summarized in **Table 40–1**. Alterations in fluid and electrolyte handling contribute significantly to diarrhea. Alterations in motor and sensory functions of the colon result in highly prevalent syndromes such as irritable bowel syndrome (IBS), chronic diarrhea, and chronic constipation.

Table 40–1 Normal Gastrointestinal Motility: Functions at Different Anatomic Levels

Stomach and small bowel
Synchronized MMC in fasting
Accommodation, trituration, mixing, transit
Stomach ~3 h
Small bowel ~3 h
Ileal reservoir empties boluses
Colon: irregular mixing, fermentation, absorption, transit
Ascending, transverse: reservoirs
Descending: conduit Sigmoid/rectum: volitional reservoir

Note: MMC, migrating motor complex.

Neural Control

The small intestine and colon have intrinsic and extrinsic innervation. The *intrinsic innervation*, also called the enteric nervous system, comprises myenteric, submucosal, and mucosal neuronal layers. The function of these layers is modulated by interneurons through the actions of neurotransmitter

amines or peptides, including acetylcholine, vasoactive intestinal peptide (VIP), opioids, norepinephrine, serotonin, Adenosine triphosphate (ATP), and nitric oxide (NO). The myenteric plexus regulates smooth-muscle function, and the submucosal plexus affects secretion, absorption, and mucosal blood flow.

The *extrinsic innervations* of the small intestine and colon are part of the autonomic nervous system and also modulate motor and secretory functions. The parasympathetic nerves convey visceral sensory and excitatory pathways to the colon. Parasympathetic fibers via the vagus nerve reach the small intestine and proximal colon along the branches of the superior mesenteric artery. The distal colon is supplied by sacral parasympathetic nerves ($S_{2–4}$) via the pelvic plexus; these fibers course through the wall of the colon as ascending intracolonic fibers as far as, and in some instances including, the proximal colon. The chief excitatory neurotransmitters controlling motor function are acetylcholine and the tachykinins, such as substance P. The sympathetic nerve supply modulates motor functions and reaches the small intestine and colon alongside their arterial vessels. Sympathetic input to the gut is generally excitatory to sphincters and inhibitory to nonsphincteric muscle. Visceral afferents convey sensation from the gut to the central nervous system (CNS); initially, they course along sympathetic fibers, but as they approach the spinal cord they separate, have cell bodies in the dorsal root ganglion, and enter the dorsal horn of the spinal cord. Afferent signals are conveyed to the brain along the lateral spinothalamic tract and the nociceptive dorsal column pathway and are then projected beyond the thalamus and brainstem to the insula and cerebral cortex to be perceived. Other afferent fibers synapse in the prevertebral ganglia and reflexly modulate intestinal motility.

Intestinal Fluid Absorption and Secretion

On an average day, 9 L of fluid enter the gastrointestinal (GI) tract, ~1 L of residual fluid reaches the colon, and the stool excretion of fluid constitutes about 0.2 L/d. The colon has a large capacitance and functional reserve and may recover up to four times its usual volume of 0.8 L/d, provided the rate of flow permits reabsorption to occur. Thus, the colon can partially compensate for excess fluid delivery to the colon because of intestinal absorptive or secretory disorders.

In the colon, sodium absorption is predominantly electrogenic, and uptake takes place at the apical membrane; it is compensated for by the export functions of the basolateral sodium pump. A variety of neural and non-neural mediators regulate colonic fluid and electrolyte balance, including cholinergic, adrenergic, and serotonergic mediators. Angiotensin and aldosterone also influence colonic absorption, reflecting the common embryologic development of the distal colonic epithelium and the renal tubules.

Small-Intestinal Motility

During fasting, the motility of the small intestine is characterized by a cyclical event called the migrating motor complex (MMC), which serves to clear nondigestible residue from the small intestine (the intestinal "housekeeper"). This organized, propagated series of contractions last, on average, 4 min, occurs every 60–90 min, and usually involve the entire small intestine. After food ingestion, the small intestine produces irregular, mixing contractions of relatively low amplitude, except in the distal ileum where more powerful contractions occur intermittently and empty the ileum by bolus transfers.

Ileocolonic Storage and Salvage

The distal ileum acts as a reservoir, emptying intermittently by bolus movements. This action allows time for salvage of fluids, electrolytes, and nutrients. Segmentation by haustra compartmentalizes the colon and facilitates mixing, retention of residue, and formation of solid stools. There is increased appreciation of the intimate interaction between the colonic function and the luminal ecology. The resident bacteria in the colon are necessary for the digestion of unabsorbed carbohydrates that reach the colon even in health, thereby providing a vital source of nutrients to the mucosa. Normal colonic flora also keeps pathogens at bay by a variety of mechanisms. In health, the ascending and transverse regions of colon function as reservoirs (average transit, 15 h), and the descending colon acts as a conduit (average transit, 3 h). The colon is efficient at conserving sodium and water, a function that is particularly important in sodium-depleted patients in whom the small intestine alone is unable to maintain sodium balance. Diarrhea or constipation may result from alteration in the reservoir function of the proximal colon or the propulsive function of the left colon. Constipation may also result from disturbances of the rectal or sigmoid reservoir, typically as a result of dysfunction of the pelvic floor, the anal sphincters, or the coordination of defecation.

Colonic Motility and Tone

The small intestinal MMC only rarely continues into the colon. However, short duration or phasic contractions mix colonic contents, and high-amplitude (>75 mmHg) propagated contractions (HAPCs) are sometimes associated with mass movements through the colon and normally occur approximately five times per day, usually on awakening in the morning and postprandially. Increased frequency of HAPCs may result in diarrhea or urgency. The predominant phasic contractions in the colon are irregular and nonpropagated and serve a "mixing" function.

Colonic tone refers to the background contractility upon which phasic contractile activity (typically

contractions lasting <15 s) is superimposed. It is an important cofactor in the colon's capacitance (volume accommodation) and sensation.

Colonic Motility after Meal Ingestion

After meal ingestion, colonic phasic and tonic contractility increase for a period of ~2 h. The initial phase (~10 min) is mediated by the vagus nerve in response to mechanical distention of the stomach. The subsequent response of the colon requires caloric stimulation and is mediated at least in part by hormones (e.g., gastrin and serotonin).

Defecation

Tonic contraction of the puborectalis muscle, which forms a sling around the rectoanal junction, is important to maintain continence; during defecation, sacral parasympathetic nerves relax this muscle, facilitating the straightening of the rectoanal angle **(Fig. 40-1)**. Distention of the rectum results in transient relaxation of the internal anal sphincter via intrinsic and reflex sympathetic innervation. As sigmoid and rectal contractions increase the pressure within the rectum, the rectosigmoid angle opens by >15°. Voluntary relaxation of the external anal sphincter (striated muscle innervated by the pudendal nerve) in response to the sensation produced by distention permits the evacuation of feces; this evacuation process can be augmented by an increase in intraabdominal pressure created by the Valsalva maneuver. Defecation can also be delayed voluntarily by contraction of the external anal sphincter.

Figure 40-1

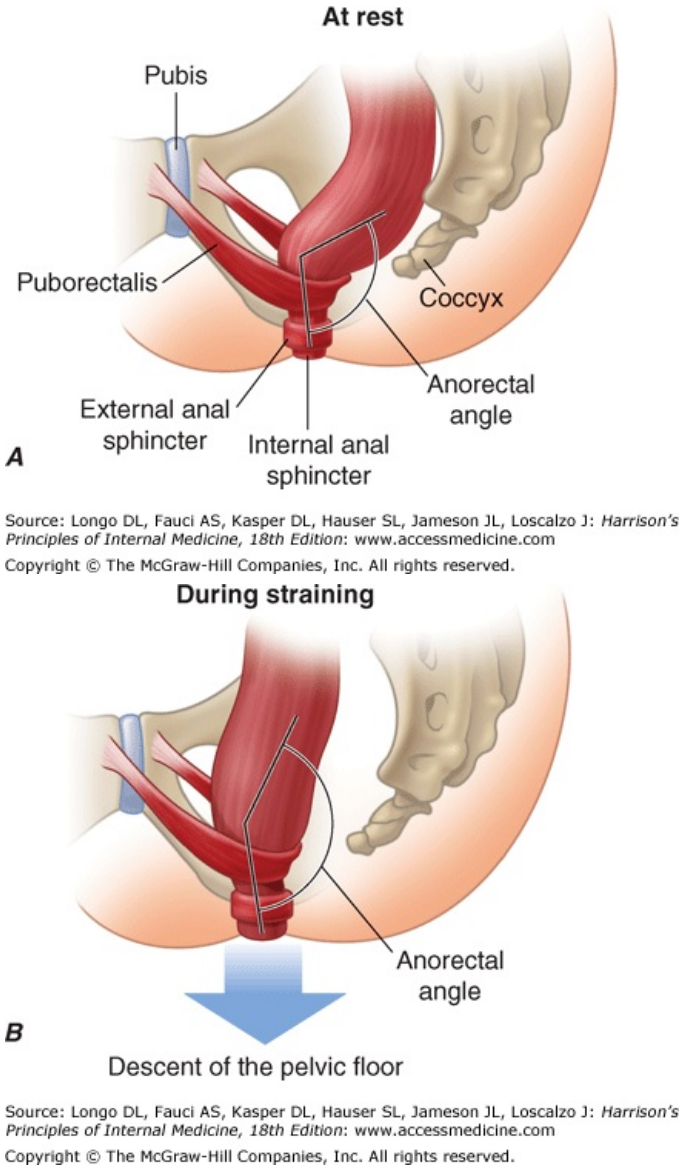

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sagittal view of the anorectum (*A*) at rest and (*B*) during straining to defecate. Continence is maintained by normal rectal sensation and tonic contraction of the internal anal sphincter and the puborectalis muscle, which wraps around the anorectum, maintaining an anorectal angle between 80° and 110°. During defecation, the pelvic floor muscles (including the puborectalis) relax, allowing the anorectal angle to straighten by at least 15°, and the perineum descends by 1–3.5 cm. The external

anorectal angle is straighten by at least 15°, and the perineum descends by 1–3 cm. The external anal sphincter also relaxes and reduces pressure on the anal canal. *(Reproduced with permission from Lembo and Camilleri.)*

DIARRHEA

Definition

Diarrhea is loosely defined as passage of abnormally liquid or unformed stools at an increased frequency. For adults on a typical Western diet, stool weight >200 g/d can generally be considered diarrheal. Diarrhea may be further defined as *acute* if <2 weeks, *persistent* if 2–4 weeks, and *chronic* if >4 weeks in duration.

Two common conditions, usually associated with the passage of stool totaling <200 g/d, must be distinguished from diarrhea, because diagnostic and therapeutic algorithms differ. *Pseudodiarrhea*, or the frequent passage of small volumes of stool, is often associated with rectal urgency and accompanies IBS or proctitis. *Fecal incontinence* is the involuntary discharge of rectal contents and is most often caused by neuromuscular disorders or structural anorectal problems. Diarrhea and urgency, especially if severe, may aggravate or cause incontinence. Pseudodiarrhea and fecal incontinence occur at prevalence rates comparable to or higher than that of chronic diarrhea and should always be considered in patients complaining of "diarrhea." Overflow diarrhea may occur in nursing home patients due to fecal impaction that is readily detectable by rectal examination. A careful history and physical examination generally allow these conditions to be discriminated from true diarrhea.

Acute Diarrhea

More than 90% of cases of acute diarrhea are caused by infectious agents; these cases are often accompanied by vomiting, fever, and abdominal pain. The remaining 10% or so are caused by medications, toxic ingestions, ischemia, and other conditions.

INFECTIOUS AGENTS

Most infectious diarrheas are acquired by fecal-oral transmission or, more commonly, via ingestion of food or water contaminated with pathogens from human or animal feces. In the immunocompetent person, the resident fecal microflora, containing >500 taxonomically distinct species, are rarely the source of diarrhea and may actually play a role in suppressing the growth of ingested pathogens. Disturbances of flora by antibiotics can lead to diarrhea by reducing the digestive function or by allowing the overgrowth of pathogens, such as *Clostridium difficile* (Chap. 129). Acute infection or injury occurs when the ingested agent overwhelms or bypasses the host's mucosal immune and nonimmune (gastric acid, digestive enzymes, mucus secretion, peristalsis, and suppressive resident flora) defenses. Established clinical associations with specific enteropathogens may offer diagnostic clues.

In the United States, five high-risk groups are recognized:

1. *Travelers*. Nearly 40% of tourists to endemic regions of Latin America, Africa, and Asia develop so-called traveler's diarrhea, most commonly due to enterotoxigenic or enteroaggregative *Escherichia coli* as well as to *Campylobacter*, *Shigella*, *Aeromonas*, norovirus, *Coronavirus,* and *Salmonella*. Visitors to Russia (especially St. Petersburg) may have increased risk of *Giardia*-associated diarrhea; visitors to Nepal may acquire *Cyclospora*. Campers, backpackers, and swimmers in wilderness areas may become infected with *Giardia*. Cruise ships may be affected by outbreaks of gastroenteritis caused by agents such as norovirus.
2. *Consumers of certain foods*. Diarrhea closely following food consumption at a picnic, banquet, or restaurant may suggest infection with *Salmonella*, *Campylobacter*, or *Shigella* from chicken; enterohemorrhagic *E. coli* (O157:H7) from undercooked hamburger; *Bacillus cereus* from fried rice or other reheated food; *Staphylococcus aureus* or *Salmonella* from mayonnaise or creams; *Salmonella* from eggs; *Listeria* from uncooked foods or soft cheeses; and *Vibrio* species, *Salmonella*, or acute hepatitis A from seafood, especially if raw.
3. *Immunodeficient persons*. Individuals at risk for diarrhea include those with either primary immunodeficiency (e.g., IgA deficiency, common variable hypogammaglobulinemia, chronic granulomatous disease) or the much more common secondary immunodeficiency states (e.g., AIDS, senescence, pharmacologic suppression). Common enteric pathogens often cause a more severe and protracted diarrheal illness, and, particularly in persons with AIDS, opportunistic infections, such as by *Mycobacterium* species, certain viruses (cytomegalovirus, adenovirus, and herpes simplex), and protozoa (*Cryptosporidium*, *Isospora belli*, Microsporida, and *Blastocystis hominis*) may also play a role (Chap. 189). In patients with AIDS, agents transmitted venereally per rectum (e.g., *Neisseria gonorrhoeae*, *Treponema pallidum*, *Chlamydia*) may contribute to proctocolitis. Persons with hemochromatosis are especially prone to invasive, even fatal, enteric infections with *Vibrio* species and *Yersinia* infections and should avoid raw fish.
4. *Daycare attendees and their family members*. Infections with *Shigella*, *Giardia*, *Cryptosporidium*, rotavirus, and other agents are very common and should be considered.
5. *Institutionalized persons*. Infectious diarrhea is one of the most frequent categories of nosocomial

infections in many hospitals and long-term care facilities; the causes are a variety of microorganisms but most commonly *C. difficile*. *C. difficile* can affect those with no history of antibiotic use and may be acquired in the community.

The pathophysiology underlying acute diarrhea by infectious agents produces specific clinical features that may also be helpful in diagnosis **(Table 40–2)**. Profuse, watery diarrhea secondary to small-bowel hypersecretion occurs with ingestion of preformed bacterial toxins, enterotoxin-producing bacteria, and enteroadherent pathogens. Diarrhea associated with marked vomiting and minimal or no fever may occur abruptly within a few hours after ingestion of the former two types; vomiting is usually less, abdominal cramping or bloating is greater, and fever is higher with the latter. Cytotoxin-producing and invasive microorganisms all cause high fever and abdominal pain. Invasive bacteria and *Entamoeba histolytica* often cause bloody diarrhea (referred to as *dysentery*). *Yersinia* invades the terminal ileal and proximal colon mucosa and may cause especially severe abdominal pain with tenderness mimicking acute appendicitis.

Table 40–2 Association between Pathobiology of Causative Agents and Clinical Features in Acute Infectious Diarrhea

Pathobiology/Agents	Incubation Period	Vomiting	Abdominal Pain	Fever	Diarrhea
Toxin producers					
Preformed toxin					
Bacillus cereus, Staphylococcus aureus, Clostridium perfringens	1–8 h 8–24 h	3–4+	1–2+	0–1+	3–4+, watery
Enterotoxin					
Vibrio cholerae, enterotoxigenic *Escherichia coli, Klebsiella pneumoniae, Aeromonas* species	8–72 h	2–4+	1–2+	0–1+	3–4+, watery
Enteroadherent					
Enteropathogenic and enteroadherent *E. coli, Giardia* organisms, cryptosporidiosis, helminths	1–8 d	0–1+	1–3+	0–2+	1–2+, watery, mushy
Cytotoxin producers					
C. difficile	1–3 d	0–1+	3–4+	1–2+	1–3+, usually watery, occasionally bloody
Hemorrhagic *E. coli*	12–72 h	0–1+	3–4+	1–2+	1–3+, initially watery, quickly bloody
Invasive organisms					
Minimal inflammation					
Rotavirus and norovirus	1–3 d	1–3+	2–3+	3–4+	1–3+, watery
Variable inflammation					
Salmonella, Campylobacter, and *Aeromonas* species, *Vibrio parahaemolyticus, Yersinia*	12 h–11 d	0–3+	2–4+	3–4+	1–4+, watery or bloody
Severe inflammation					
Shigella species, enteroinvasive *E. coli*, *Entamoeba histolytica*	12 h–8 d	0–1+	3–4+	3–4+	1–2+, bloody

Source: Adapted from DW Powell, in T Yamada (ed): *Textbook of Gastroenterology and Hepatology* , 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2003.

Finally, infectious diarrhea may be associated with systemic manifestations. Reactive arthritis (formerly known as Reiter's syndrome), arthritis, urethritis, and conjunctivitis may accompany or follow infections by *Salmonella*, *Campylobacter*, *Shigella*, and *Yersinia*. Yersiniosis may also lead to an autoimmune-type thyroiditis, pericarditis, and glomerulonephritis. Both enterohemorrhagic *E. coli* (O157:H7) and *Shigella* can lead to the *hemolytic-uremic syndrome* with an attendant high mortality rate. The syndrome of postinfectious IBS has now been recognized as a complication of infectious diarrhea.

Acute diarrhea can also be a major symptom of several systemic infections including *viral hepatitis*, *listeriosis*, *legionellosis*, and *toxic shock syndrome*.

OTHER CAUSES

Side effects from medications are probably the most common noninfectious causes of acute diarrhea, and etiology may be suggested by a temporal association between use and symptom onset. Although innumerable medications may produce diarrhea, some of the more frequently incriminated include antibiotics, cardiac antidysrhythmics, antihypertensives, nonsteroidal anti-inflammatory drugs (NSAIDs), certain antidepressants, chemotherapeutic agents, bronchodilators, antacids, and laxatives. Occlusive or nonocclusive *ischemic colitis* typically occurs in persons >50 years; often presents as acute lower abdominal pain preceding watery, then bloody diarrhea; and generally results in acute inflammatory changes in the sigmoid or left colon while sparing the rectum. Acute diarrhea may accompany colonic *diverticulitis* and *graft-versus-host disease*. Acute diarrhea, often associated with systemic compromise, can follow ingestion of toxins including organophosphate insecticides; amanita and other mushrooms; arsenic; and preformed environmental toxins in seafood, such as ciguatera and scombroid. Acute anaphylaxis to food ingestion can have a similar presentation. Conditions causing chronic diarrhea can also be confused with acute diarrhea early in their course. This confusion may occur with inflammatory bowel disease (IBD) and some of the other inflammatory chronic diarrheas that may have an abrupt rather than insidious onset and exhibit features that mimic infection.

APPROACH TO THE PATIENT: ACUTE DIARRHEA

The decision to evaluate acute diarrhea depends on its severity and duration and on various host factors **(Fig. 40-2)**. Most episodes of acute diarrhea are mild and self-limited and do not justify the cost and potential morbidity rate of diagnostic or pharmacologic interventions. Indications for evaluation include profuse diarrhea with dehydration, grossly bloody stools, fever ≥38.5°C (≥101°F), duration >48 h without improvement, recent antibiotic use, new community outbreaks, associated severe abdominal pain in patients >50 years, and elderly (≥70 years) or immunocompromised patients. In some cases of moderately severe febrile diarrhea associated with fecal leukocytes (or increased fecal levels of the leukocyte proteins) or with gross blood, a diagnostic evaluation might be avoided in favor of an empirical antibiotic trial (see below).

Figure 40-2

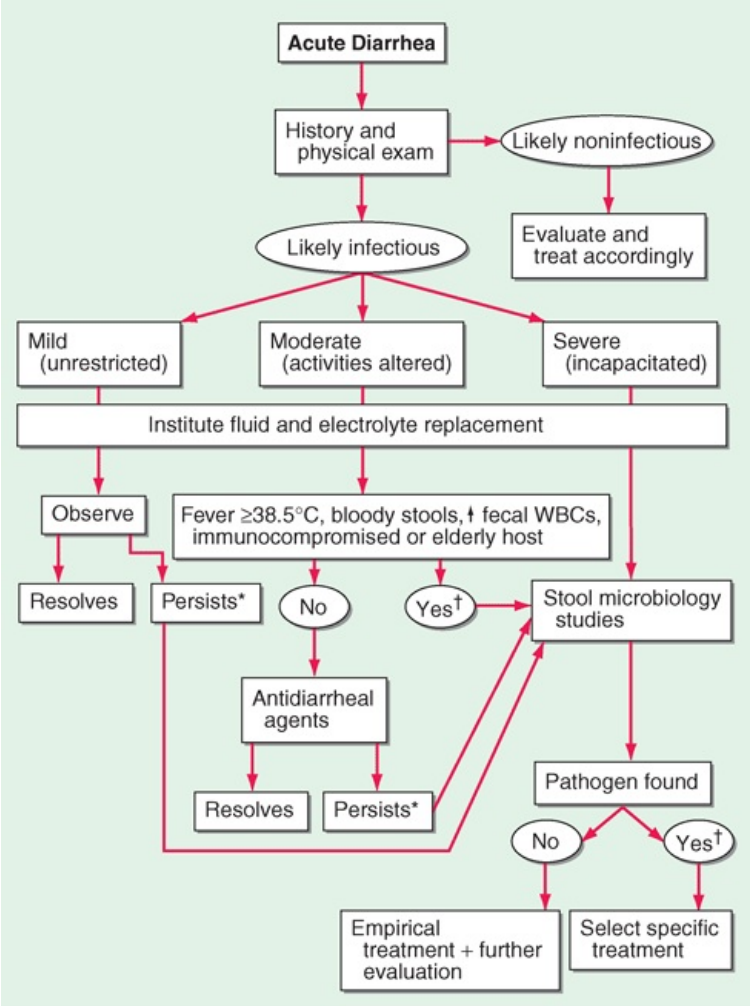

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the management of acute diarrhea. Consider empirical Rx before evaluation with (*) metronidazole and with (†) quinolone. WBCs, white blood cells.

The cornerstone of diagnosis in those suspected of severe acute infectious diarrhea is microbiologic analysis of the stool. Workup includes cultures for bacterial and viral pathogens, direct inspection for ova and parasites, and immunoassays for certain bacterial toxins (*C. difficile*), viral antigens (rotavirus), and protozoal antigens (*Giardia*, *E. histolytica*). The aforementioned clinical and epidemiologic associations may assist in focusing the evaluation. If a particular pathogen or set of possible pathogens is so implicated, then either the whole panel of routine studies may not be necessary or, in some instances, special cultures may be appropriate as for enterohemorrhagic and other types of *E. coli*, *Vibrio* species, and *Yersinia*. Molecular diagnosis of pathogens in stool can be made by identification of unique DNA sequences; and evolving microarray technologies could lead to a more rapid, sensitive, specific, and cost-effective diagnostic approach in the future.

Persistent diarrhea is commonly due to *Giardia* (Chap. 209), but additional causative organisms that should be considered include *C. difficile* (especially if antibiotics had been administered), *E. histolytica*, *Cryptosporidium*, *Campylobacter*, and others. If stool studies are unrevealing, flexible sigmoidoscopy with biopsies and upper endoscopy with duodenal aspirates and biopsies may be indicated. Brainerd diarrhea is an increasingly recognized entity characterized by an abrupt-onset diarrhea that persists for at least 4 weeks, but may last 1–3 years, and is thought to be of infectious origin. It may be associated with subtle inflammation of the distal small intestine or proximal colon.

Structural examination by sigmoidoscopy, colonoscopy, or abdominal CT scanning (or other imaging approaches) may be appropriate in patients with uncharacterized persistent diarrhea to exclude IBD or as an initial approach in patients with suspected noninfectious acute diarrhea such as might be caused by ischemic colitis, diverticulitis, or partial bowel obstruction.

TREATMENT: ACUTE DIARRHEA

Fluid and electrolyte replacement are of central importance to all forms of acute diarrhea. Fluid replacement alone may suffice for mild cases. Oral sugar-electrolyte solutions (sport drinks or designed formulations) should be instituted promptly with severe diarrhea to limit dehydration, which is

the major cause of death. Profoundly dehydrated patients, especially infants and the elderly, require IV rehydration.

In moderately severe nonfebrile and nonbloody diarrhea, antimotility and antisecretory agents such as loperamide can be useful adjuncts to control symptoms. Such agents should be avoided with febrile dysentery, which may be exacerbated or prolonged by them. Bismuth subsalicylate may reduce symptoms of vomiting and diarrhea but should not be used to treat immunocompromised patients or those with renal impairment because of the risk of bismuthencephalopathy.

Judicious use of antibiotics is appropriate in selected instances of acute diarrhea and may reduce its severity and duration **(Fig. 40-2)**. Many physicians treat moderately to severely ill patients with febrile dysentery empirically without diagnostic evaluation using a quinolone, such as ciprofloxacin (500 mg bid for 3–5 d). Empirical treatment can also be considered for suspected giardiasis with metronidazole (250 mg qid for 7 d). Selection of antibiotics and dosage regimens are otherwise dictated by specific pathogens, geographic patterns of resistance, and conditions found (Chaps. 128, 149, and 153, 154, 155, 156, 157, 158, and 159). Antibiotic coverage is indicated, whether or not a causative organism is discovered, in patients who are immunocompromised, have mechanical heart valves or recent vascular grafts, or are elderly. Bismuth subsalicylate may reduce the frequency of traveler's diarrhea. Antibiotic prophylaxis is only indicated for certain patients traveling to high-risk countries in whom the likelihood or seriousness of acquired diarrhea would be especially high, including those with immunocompromise, IBD, hemochromatosis, or gastric achlorhydria. Use ofciprofloxacin or rifaximin may reduce bacterial diarrhea in such travelers by 90%, though rifaximin is not suitable for invasive disease, but rather as treatment for uncomplicated traveler's diarrhea. Finally, physicians should be vigilant to identify if an outbreak of diarrheal illness is occurring and to alert the public health authorities promptly. This may reduce the ultimate size of the affected population.

Chronic Diarrhea

Diarrhea lasting >4 weeks warrants evaluation to exclude serious underlying pathology. In contrast to acute diarrhea, most of the causes of chronic diarrhea are noninfectious. The classification of chronic diarrhea by pathophysiologic mechanism facilitates a rational approach to management, though many diseases cause diarrhea by more than one mechanism **(Table 40–3)**.

Table 40–3 Major Causes of Chronic Diarrhea According to Predominant Pathophysiologic Mechanism

Secretory causes	**Inflammatory causes**
Exogenous stimulant laxatives	Idiopathic inflammatory bowel disease (Crohn's, chronic ulcerative colitis)
Chronic ethanol ingestion	Lymphocytic and collagenous colitis
Other drugs and toxins	Immune-related mucosal disease (1° or 2° immunodeficiencies, food allergy, eosinophilic gastroenteritis, graft-vs-host disease)
Endogenous laxatives (dihydroxy bile acids)	Infections (invasive bacteria, viruses, and parasites, Brainerd diarrhea)
Idiopathic secretory diarrhea	Radiation injury Gastrointestinal malignancies
Certain bacterial infections	**Dysmotile causes**
Bowel resection, disease, or fistula (↑ absorption)	Irritable bowel syndrome (including postinfectious IBS)
Partial bowel obstruction or fecal impaction	Visceral neuromyopathies
Hormone-producing tumors (carcinoid, VIPoma, medullary cancer of thyroid, mastocytosis, gastrinoma, colorectal villous adenoma)	Hyperthyroidism
Addison's disease	Drugs (prokinetic agents) Postvagotomy
Congenital electrolyte absorption defects	**Factitial causes**
Osmotic causes	Munchausen
Osmotic laxatives (Mg^{2+}, PO_4^{-3}, SO_4^{-2}	Eating disorders
Lactase and other disaccharide deficiencies	**Iatrogenic causes**
Nonabsorbable carbohydrates (sorbitol, lactulose, polyethylene glycol)	Cholecystectomy
Steatorrheal causes	Ileal resection
Intraluminal maldigestion (pancreatic exocrine insufficiency, bacterial overgrowth, bariatric surgery, liver disease)	Bariatric surgery
Mucosal malabsorption (celiac sprue, Whipple's disease, infections, abetalipoproteinemia, ischemia)	Vagotomy, fundoplication

Postmucosal obstruction (1° or 2° lymphatic obstruction)	

SECRETORY CAUSES

Secretory diarrheas are due to derangements in fluid and electrolyte transport across the enterocolonic mucosa. They are characterized clinically by watery, large-volume fecal outputs that are typically painless and persist with fasting. Because there is no malabsorbed solute, stool osmolality is accounted for by normal endogenous electrolytes with no fecal osmotic gap.

Medications

Side effects from regular ingestion of drugs and toxins are the most common secretory causes of chronic diarrhea. Hundreds of prescription and over-the-counter medications (see "Acute Diarrhea, Other Causes," above) may produce diarrhea. Surreptitious or habitual use of stimulant laxatives [e.g., senna, cascara, bisacodyl, ricinoleic acid (castor oil)] must also be considered. Chronic ethanol consumption may cause a secretory-type diarrhea due to enterocyte injury with impaired sodium and water absorption as well as rapid transit and other alterations. Inadvertent ingestion of certain environmental toxins (e.g., arsenic) may lead to chronic rather than acute forms of diarrhea. Certain bacterial infections may occasionally persist and be associated with a secretory-type diarrhea.

Bowel Resection, Mucosal Disease, or Enterocolic Fistula

These conditions may result in a secretory-type diarrhea because of inadequate surface for reabsorption of secreted fluids and electrolytes. Unlike other secretory diarrheas, this subset of conditions tends to worsen with eating. With disease (e.g., Crohn's ileitis) or resection of <100 cm of terminal ileum, dihydroxy bile acids may escape absorption and stimulate colonic secretion (cholorrheic diarrhea). This mechanism may contribute to so-called *idiopathic secretory diarrhea*, in which bile acids are functionally malabsorbed from a normal-appearing terminal ileum. This *idiopathic bile acid malabsorption* may account for an average of 40% of unexplained chronic diarrhea. Reduced negative feedback regulation of bile acid synthesis by fibroblast growth factor 19 produced by enterocytes results in a degree of bile-acid synthesis that exceeds the normal capacity for ileal reabsorption, producing bile acid diarrhea.

Partial bowel obstruction, ostomy stricture, or fecal impaction may paradoxically lead to increased fecal output due to fluid hypersecretion.

Hormones

Although uncommon, the classic examples of secretory diarrhea are those mediated by hormones. *Metastatic gastrointestinal carcinoid tumors* or, rarely, *primary bronchial carcinoids* may produce watery diarrhea alone or as part of the carcinoid syndrome that comprises episodic flushing, wheezing, dyspnea, and right-sided valvular heart disease. Diarrhea is due to the release into the circulation of potent intestinal secretagogues including serotonin, histamine, prostaglandins, and various kinins. Pellagra-like skin lesions may rarely occur as the result of serotonin overproduction with niacin depletion. *Gastrinoma*, one of the most common neuroendocrine tumors, most typically presents with refractory peptic ulcers, but diarrhea occurs in up to one-third of cases and may be the only clinical manifestation in 10%. While other secretagogues released with gastrin may play a role, the diarrhea most often results from fat maldigestion owing to pancreatic enzyme inactivation by low intraduodenal pH. The watery diarrhea hypokalemia achlorhydria syndrome, also called *pancreatic cholera*, is due to a non-β cell pancreatic adenoma, referred to as a *VIPoma*, that secretes VIP and a host of other peptide hormones including pancreatic polypeptide, secretin, gastrin, gastrin-inhibitory polypeptide (also called glucose-dependent insulinotropic peptide), neurotensin, calcitonin, and prostaglandins. The secretory diarrhea is often massive with stool volumes >3 L/d; daily volumes as high as 20 L have been reported. Life-threatening dehydration; neuromuscular dysfunction from associated hypokalemia, hypomagnesemia, or hypercalcemia; flushing; and hyperglycemia may accompany a VIPoma. *Medullary carcinoma of the thyroid* may present with watery diarrhea caused by calcitonin, other secretory peptides, or prostaglandins. Prominent diarrhea is often associated with metastatic disease and poor prognosis. *Systemic mastocytosis*, which may be associated with the skin lesion urticaria pigmentosa, may cause diarrhea that is either secretory and mediated by histamine or inflammatory due to intestinal infiltration by mast cells. Large *colorectal villous adenomas* may rarely be associated with a secretory diarrhea that may cause hypokalemia, can be inhibited by NSAIDs, and are apparently mediated by prostaglandins.

Congenital Defects in Ion Absorption

Rarely, defects in specific carriers associated with ion absorption cause watery diarrhea from birth.

These disorders include defective Cl^-/HCO_3^- exchange (*congenital chloridorrhea*) with alkalosis

(which results from a mutated *DRA* [down-regulated in adenoma] gene) and defective Na^+/H^+ exchange (*congential sodium diarrhea),* which results from a mutation in the *NHE3* (sodium-hydrogen exchanger) gene and results in acidosis.

Some hormone deficiencies may be associated with watery diarrhea, such as occurs with adrenocortical insufficiency (Addison's disease) that may be accompanied by skin hyperpigmentation.

OSMOTIC CAUSES

Osmotic diarrhea occurs when ingested, poorly absorbable, osmotically active solutes draw enough fluid into the lumen to exceed the reabsorptive capacity of the colon. Fecal water output increases in proportion to such a solute load. Osmotic diarrhea characteristically ceases with fasting or with discontinuation of the causative agent.

Osmotic Laxatives

Ingestion of magnesium-containing antacids, health supplements, or laxatives may induce osmotic diarrhea typified by a stool osmotic gap (>50 mosmol/L): serum osmolarity (typically 290 mosmol/kg)-[2 x (fecal sodium + potassium concentration)]. Measurement of fecal osmolarity is no longer recommended because, even when measured immediately after evacuation, it may be erroneous because carbohydrates are metabolized by colonic bacteria, causing an increase in osmolarity.

Carbohydrate Malabsorption

Carbohydrate malabsorption due to acquired or congenital defects in brush-border disaccharidases and other enzymes leads to osmotic diarrhea with a low pH. One of the most common causes of chronic diarrhea in adults is *lactase deficiency*, which affects three-fourths of non-whites worldwide and 5–30% of persons in the United States; the total lactose load at any one time influences the symptoms experienced. Most patients learn to avoid milk products without requiring treatment with enzyme supplements. Some sugars, such as sorbitol, lactulose, or fructose, are frequently malabsorbed, and diarrhea ensues with ingestion of medications, gum, or candies sweetened with these poorly or incompletely absorbed sugars.

STEATORRHEAL CAUSES

Fat malabsorption may lead to greasy, foul-smelling, difficult-to-flush diarrhea often associated with weight loss and nutritional deficiencies due to concomitant malabsorption of amino acids and vitamins. Increased fecal output is caused by the osmotic effects of fatty acids, especially after bacterial hydroxylation, and, to a lesser extent, by the neutral fat. Quantitatively, steatorrhea is defined as stool fat exceeding the normal 7 g/d; rapid-transit diarrhea may result in fecal fat up to 14 g/d; daily fecal fat averages 15–25 g with small intestinal diseases and is often >32 g with pancreatic exocrine insufficiency. Intraluminal maldigestion, mucosal malabsorption, or lymphatic obstruction may produce steatorrhea.

Intraluminal Maldigestion

This condition most commonly results from pancreatic exocrine insufficiency, which occurs when >90% of pancreatic secretory function is lost. *Chronic pancreatitis*, usually a sequel of ethanol abuse, most frequently causes pancreatic insufficiency. Other causes include *cystic fibrosis*; *pancreatic duct obstruction*; and, rarely, *somatostatinoma*. Bacterial overgrowth in the small intestine may deconjugate bile acids and alter micelle formation, impairing fat digestion; it occurs with stasis from a blind-loop, small-bowel diverticulum or dysmotility and is especially likely in the elderly. Finally, cirrhosis or biliary obstruction may lead to mild steatorrhea due to deficient intraluminal bile acid concentration.

Mucosal Malabsorption

Mucosal malabsorption occurs from a variety of enteropathies, but it most commonly occurs from *celiac disease*. This gluten-sensitive enteropathy affects all ages and is characterized by villous atrophy and crypt hyperplasia in the proximal small bowel and can present with fatty diarrhea associated with multiple nutritional deficiencies of varying severity. Celiac disease is much more frequent than previously thought; it affects ~1% of the population, frequently presents without steatorrhea, can mimic IBS, and has many other GI and extraintestinal manifestations. *Tropical sprue* may produce a similar histologic and clinical syndrome but occurs in residents of or travelers to tropical climates; abrupt onset and response to antibiotics suggest an infectious etiology. *Whipple's disease*, due to the bacillus *Tropheryma whipplei* and histiocytic infiltration of the small-bowel mucosa, is a less common cause of steatorrhea that most typically occurs in young or middle-aged men; it is frequently associated with arthralgias, fever, lymphadenopathy, and extreme fatigue, and it may affect the CNS and endocardium. A similar clinical and histologic picture results from *Mycobacterium avium-intracellulare* infection in patients with AIDS. *Abetalipoproteinemia* is a rare defect of chylomicron formation and fat malabsorption in children, associated with acanthocytic erythrocytes, ataxia, and retinitis pigmentosa. Several other conditions may cause mucosal malabsorption including infections, especially with protozoa such as *Giardia*; numerous medications (e.g., colchicine, cholestyramine, neomycin); amyloidosis; and chronic ischemia.

Postmucosal Lymphatic Obstruction

The pathophysiology of this condition, which is due to the rare *congenital intestinal lymphangiectasia* or to *acquired lymphatic obstruction* secondary to trauma, tumor, cardiac disease or infection, leads to the unique constellation of fat malabsorption with enteric losses of protein (often causing edema) and

lymphocytopenia. Carbohydrate and amino acid absorption are preserved.

INFLAMMATORY CAUSES

Inflammatory diarrheas are generally accompanied by pain, fever, bleeding, or other manifestations of inflammation. The mechanism of diarrhea may not only be exudation but, depending on lesion site, may include fat malabsorption, disrupted fluid/electrolyte absorption, and hypersecretion or hypermotility from release of cytokines and other inflammatory mediators. The unifying feature on stool analysis is the presence of leukocytes or leukocyte-derived proteins such as calprotectin. With severe inflammation, exudative protein loss can lead to anasarca (generalized edema). Any middle-aged or older person with chronic inflammatory-type diarrhea, especially with blood, should be carefully evaluated to exclude a colorectal tumor.

Idiopathic Inflammatory Bowel Disease

The illnesses in this category, which include *Crohn's disease* and *chronic ulcerative colitis*, are among the most common organic causes of chronic diarrhea in adults and range in severity from mild to fulminant and life-threatening. They may be associated with uveitis, polyarthralgias, cholestatic liver disease (primary sclerosing cholangitis), and skin lesions (erythema nodosum, pyoderma gangrenosum). *Microscopic colitis*, including both lymphocytic and *collagenous colitis*, is an increasingly recognized cause of chronic watery diarrhea, especially in middle-aged women and those on NSAIDs, statins, proton pump inhibitors (PPIs), and selective serotonin reuptake inhibitors (SSRIs); biopsy of a normal-appearing colon is required for histologic diagnosis. It may coexist with symptoms suggesting IBS or with celiac sprue. It typically responds well to anti-inflammatory drugs (e.g., bismuth), to the opioid agonist loperamide, or to budesonide.

Primary or Secondary Forms of Immunodeficiency

Immunodeficiency may lead to prolonged infectious diarrhea. With selective IgA deficiency or common variable *hypogammaglobulinemia*, diarrhea is particularly prevalent and often the result of giardiasis, bacterial overgrowth, or sprue.

Eosinophilic Gastroenteritis

Eosinophil infiltration of the mucosa, muscularis, or serosa at any level of the GI tract may cause diarrhea, pain, vomiting, or ascites. Affected patients often have an atopic history, Charcot-Leyden crystals due to extruded eosinophil contents may be seen on microscopic inspection of stool, and peripheral eosinophilia is present in 50–75% of patients. While hypersensitivity to certain foods occurs in adults, true food allergy causing chronic diarrhea is rare.

Other Causes

Chronic inflammatory diarrhea may be caused by *radiation enterocolitis*, *chronic graft-versus-host disease*, *Behçt's syndrome*, and *Cronkhite-Canada syndrome*, among others.

DYSMOTILITY CAUSES

Rapid transit may accompany many diarrheas as a secondary or contributing phenomenon, but primary dysmotility is an unusual etiology of true diarrhea. Stool features often suggest a secretory diarrhea, but mild steatorrhea of up to 14 g of fat per day can be produced by maldigestion from rapid transit alone. *Hyperthyroidism*, *carcinoid syndrome*, and certain drugs (e.g., prostaglandins, prokinetic agents) may produce hypermotility with resultant diarrhea. Primary visceral neuromyopathies or idiopathic acquired intestinal pseudoobstruction may lead to stasis with secondary bacterial overgrowth causing diarrhea. *Diabetic diarrhea*, often accompanied by peripheral and generalized autonomic neuropathies, may occur in part because of intestinal dysmotility.

The exceedingly common IBS (10% point prevalence, 1–2% per year incidence) is characterized by disturbed intestinal and colonic motor and sensory responses to various stimuli. Symptoms of stool frequency typically cease at night, alternate with periods of constipation, are accompanied by abdominal pain relieved with defecation, and rarely result in weight loss.

FACTITIAL CAUSES

Factitial diarrhea accounts for up to 15% of unexplained diarrheas referred to tertiary care centers. Either as a form of *Munchausen syndrome* (deception or self-injury for secondary gain) or *eating disorders*, some patients covertly self-administer laxatives alone or in combination with other medications (e.g., diuretics) or surreptitiously add water or urine to stool sent for analysis. Such patients are typically women, often with histories of psychiatric illness, and disproportionately from careers in health care. Hypotension and hypokalemia are common co-presenting features. The evaluation of such patients may be difficult: contamination of the stool with water or urine is suggested by very low or high stool osmolarity, respectively. Such patients often deny this possibility when confronted, but they do benefit from psychiatric counseling when they acknowledge their behavior.

APPROACH TO THE PATIENT: CHRONIC DIARRHEA

The laboratory tools available to evaluate the very common problem of chronic diarrhea are extensive, and many are costly and invasive. As such, the diagnostic evaluation must be rationally directed by a

careful history and physical examination **(Fig. 40-3*A*)**. When this strategy is unrevealing, simple triage tests are often warranted to direct the choice of more complex investigations **(Fig. 40-3*B*)**. The history, physical examination **(Table 40–4)**, and routine blood studies should attempt to characterize the mechanism of diarrhea, identify diagnostically helpful associations, and assess the patient's fluid/electrolyte and nutritional status. Patients should be questioned about the onset, duration, pattern, aggravating (especially diet) and relieving factors, and stool characteristics of their diarrhea. The presence or absence of fecal incontinence, fever, weight loss, pain, certain exposures (travel, medications, contacts with diarrhea), and common extraintestinal manifestations (skin changes, arthralgias, oral aphthous ulcers) should be noted. A family history of IBD or sprue may indicate those possibilities. Physical findings may offer clues such as a thyroid mass, wheezing, heart murmurs, edema, hepatomegaly, abdominal masses, lymphadenopathy, mucocutaneous abnormalities, perianal fistulas, or anal sphincter laxity. Peripheral blood leukocytosis, elevated sedimentation rate, or C-reactive protein suggests inflammation; anemia reflects blood loss or nutritional deficiencies; or eosinophilia may occur with parasitoses, neoplasia, collagen-vascular disease, allergy, or eosinophilic gastroenteritis. Blood chemistries may demonstrate electrolyte, hepatic, or other metabolic disturbances. Measuring tissue transglutaminase antibodies may help detect celiac disease.

Figure 40-3

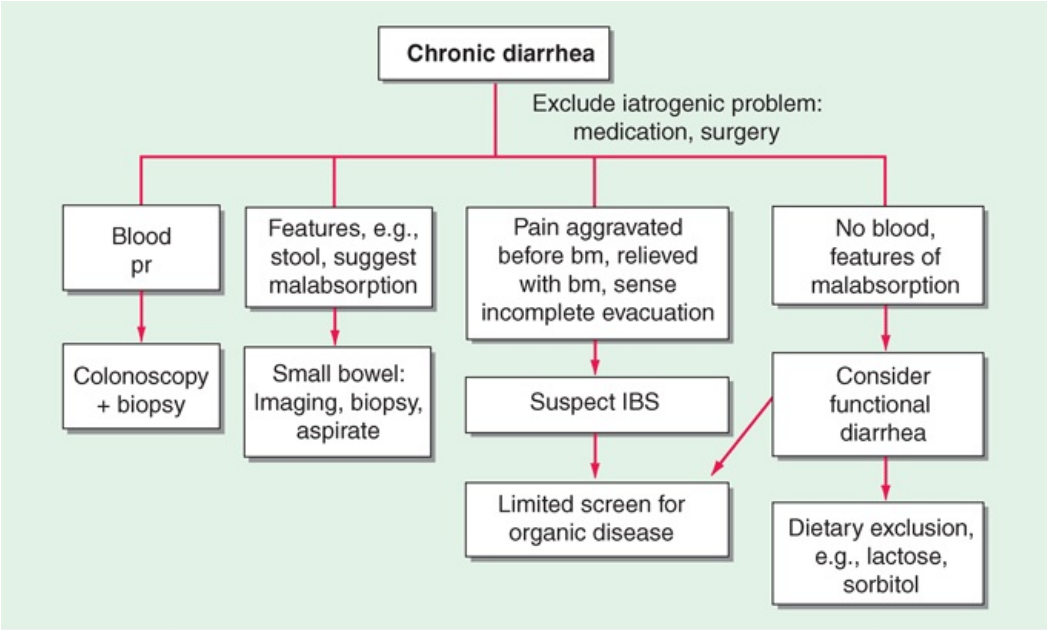

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

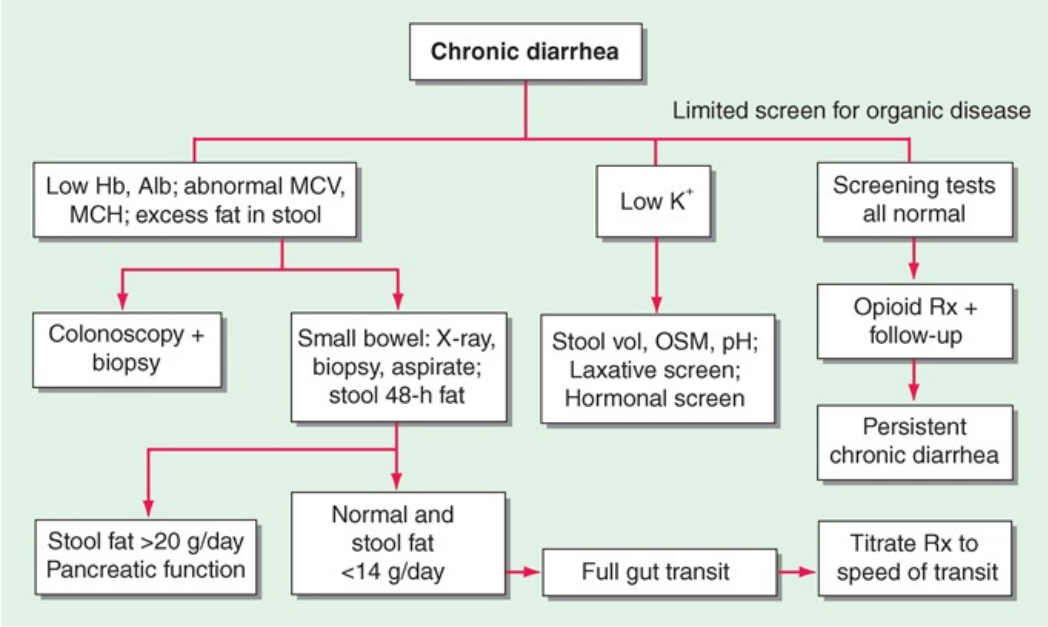

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic diarrhea. ***A.*** Initial management based on accompanying symptoms or features. ***B.*** Evaluation based on findings from a limited age-appropriate screen for organic disease. pr, per rectum; bm, bowel movement; IBS, irritable bowel syndrome; Hb, hemoglobin; Alb, albumin; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; OSM, osmolality. *(Reprinted from M Camilleri: Clin Gastroenterol Hepatol. 2:198, 2004.)*

Table 40–4 Physical Examination in Patients with Chronic Diarrhea

1. Are there general features to suggest malabsorption or inflammatory bowel disease (IBD) such as anemia, dermatitis herpetiformis, edema, or clubbing?
2. Are there features to suggest underlying autonomic neuropathy or collagen-vascular disease in the pupils, orthostasis, skin, hands, or joints?
3. Is there an abdominal mass or tenderness?
4. Are there any abnormalities of rectal mucosa, rectal defects, or altered anal sphincter functions?
5. Are there any mucocutaneous manifestations of systemic disease such as dermatitis herpetiformis (celiac disease), erythema nodosum (ulcerative colitis), flushing (carcinoid), or oral ulcers for IBD or celiac disease?

A therapeutic trial is often appropriate, definitive, and highly cost effective when a specific diagnosis is suggested on the initial physician encounter. For example, chronic watery diarrhea, which ceases with fasting in an otherwise healthy young adult, may justify a trial of a lactose-restricted diet; bloating and diarrhea persisting since a mountain backpacking trip may warrant a trial of metronidazole for likely giardiasis; and postprandial diarrhea persisting following resection of terminal ileum might be due to bile acid malabsorption and be treated with cholestyramine or colesevelam before further evaluation. Persistent symptoms require additional investigation.

Certain diagnoses may be suggested on the initial encounter (e.g., idiopathic IBD); however, additional focused evaluations may be necessary to confirm the diagnosis and characterize the severity or extent of disease so that treatment can be best guided. Patients suspected of having IBS should be initially evaluated with flexible sigmoidoscopy with colorectal biopsies; those with normal findings might be reassured and, as indicated, treated empirically with antispasmodics, antidiarrheals, bulk agents, anxiolytics, or antidepressants. Any patient who presents with chronic diarrhea and hematochezia should be evaluated with stool microbiologic studies and colonoscopy.

In an estimated two-thirds of cases, the cause for chronic diarrhea remains unclear after the initial encounter, and further testing is required. Quantitative stool collection and analyses can yield important objective data that may establish a diagnosis or characterize the type of diarrhea as a triage for focused additional studies (Fig. 40-3*B*). If stool weight is >200 g/d, additional stool analyses should be performed that might include electrolyte concentration, pH, occult blood testing, leukocyte inspection (or leukocyte protein assay), fat quantitation, and laxative screens.

For secretory diarrheas (watery, normal osmotic gap), possible medication-related side effects or

surreptitious laxative use should be reconsidered. Microbiologic studies should be done including fecal bacterial cultures (including media for *Aeromonas* and *Pleisiomonas*), inspection for ova and parasites, and *Giardia* antigen assay (the most sensitive test for giardiasis). Small-bowel bacterial overgrowth can be excluded by intestinal aspirates with quantitative cultures or with glucose or lactulose breath tests involving measurement of breath hydrogen, methane, or other metabolite (e.g., $^{14}CO_2$). However, interpretation of these breath tests may be confounded by disturbances of intestinal transit. Upper endoscopy and colonoscopy with biopsies and small-bowel barium x-rays are helpful to rule out structural or occult inflammatory disease. When suggested by history or other findings, screens for peptide hormones should be pursued (e.g., serum gastrin, VIP, calcitonin, and thyroid hormone/thyroid-stimulating hormone, or urinary 5-hydroxyindolacetic acid, and histamine).

Further evaluation of osmotic diarrhea should include tests for lactose intolerance and magnesium ingestion, the two most common causes. Low fecal pH suggests carbohydrate malabsorption; lactose malabsorption can be confirmed by lactose breath testing or by a therapeutic trial with lactose exclusion and observation of the effect of lactose challenge (e.g., a liter of milk). Lactase determination on small-bowel biopsy is not generally available. If fecal magnesium or laxative levels are elevated, inadvertent or surreptitious ingestion should be considered and psychiatric help should be sought.

For those with proven fatty diarrhea, endoscopy with small-bowel biopsy (including aspiration for *Giardia* and quantitative cultures) should be performed; if this procedure is unrevealing, a small-bowel radiograph is often an appropriate next step. If small-bowel studies are negative or if pancreatic disease is suspected, pancreatic exocrine insufficiency should be excluded with direct tests, such as the secretin-cholecystokinin stimulation test or a variation that could be performed endoscopically. In general, indirect tests such as assay of fecal elastase or chymotrypsin activity or a bentiromide test have fallen out of favor because of low sensitivity and specificity.

Chronic inflammatory-type diarrheas should be suspected by the presence of blood or leukocytes in the stool. Such findings warrant stool cultures; inspection for ova and parasites; *C. difficile* toxin assay; colonoscopy with biopsies; and, if indicated, small-bowel contrast studies.

TREATMENT: CHRONIC DIARRHEA

Treatment of chronic diarrhea depends on the specific etiology and may be curative, suppressive, or empirical. If the cause can be eradicated, treatment is curative as with resection of a colorectal cancer, antibiotic administration for Whipple's disease or tropical sprue, or discontinuation of a drug. For many chronic conditions, diarrhea can be controlled by suppression of the underlying mechanism. Examples include elimination of dietary lactose for lactase deficiency or gluten for celiac sprue, use of glucocorticoids or other anti-inflammatory agents for idiopathic IBDs, adsorptive agents such as cholestyramine for ileal bile acid malabsorption, proton pump inhibitors such as omeprazole for the gastric hypersecretion of gastrinomas, somatostatin analogues such as octreotide for malignantcarcinoid syndrome, prostaglandin inhibitors such as indomethacin for medullary carcinoma of the thyroid, and pancreatic enzyme replacement for pancreatic insufficiency. When the specific cause or mechanism of chronic diarrhea evades diagnosis, empirical therapy may be beneficial. Mild opiates, such as diphenoxylate or loperamide, are often helpful in mild or moderate watery diarrhea. For those with more severe diarrhea, codeine or tincture of opium may be beneficial. Such antimotility agents should be avoided with severe IBD, because toxic megacolon may be precipitated. Clonidine, an α_2-adrenergic agonist, may allow control of diabetic diarrhea. For all patients with chronic diarrhea, fluid and electrolyte repletion is an important component of management (see "Acute Diarrhea," above). Replacement of fat-soluble vitamins may also be necessary in patients with chronic steatorrhea.

CONSTIPATION

Definition

Constipation is a common complaint in clinical practice and usually refers to persistent, difficult, infrequent, or seemingly incomplete defecation. Because of the wide range of normal bowel habits, constipation is difficult to define precisely. Most persons have at least three bowel movements per week; however, low stool frequency alone is not the sole criterion for the diagnosis of constipation. Many constipated patients have a normal frequency of defecation but complain of excessive straining, hard stools, lower abdominal fullness, or a sense of incomplete evacuation. The individual patient's symptoms must be analyzed in detail to ascertain what is meant by "constipation" or "difficulty" with defecation.

Stool form and consistency are well correlated with the time elapsed from the preceding defecation. Hard, pellety stools occur with slow transit, while loose, watery stools are associated with rapid transit. Both small pellety or very large stools are more difficult to expel than normal stools.

The perception of hard stools or excessive straining is more difficult to assess objectively, and the need for enemas or digital disimpaction is a clinically useful way to corroborate the patient's perceptions of difficult defecation.

Psychosocial or cultural factors may also be important. A person whose parents attached great importance to daily defecation will become greatly concerned when he or she misses a daily bowel movement; some children withhold stool to gain attention or because of fear of pain from anal irritation; and some adults habitually ignore or delay the call to have a bowel movement.

Causes

Pathophysiologically, chronic constipation generally results from inadequate fiber or fluid intake or from disordered colonic transit or anorectal function. These result from neurogastroenterologic disturbance, certain drugs, advancing age, or in association with a large number of systemic diseases that affect the GI tract **(Table 40–5)**. Constipation of recent onset may be a symptom of significant organic disease such as tumor or stricture. In *idiopathic constipation*, a subset of patients exhibit delayed emptying of the ascending and transverse colon with prolongation of transit (often in the proximal colon) and a reduced frequency of propulsive HAPCs. *Outlet obstruction to defecation* (also called *evacuation disorders*) may cause delayed colonic transit, which is usually corrected by biofeedback retraining of the disordered defecation. Constipation of any cause may be exacerbated by hospitalization or chronic illnesses that lead to physical or mental impairment and result in inactivity or physical immobility.

Table 40–5 Causes of Constipation in Adults

Types of Constipation and Causes	Examples
Recent onset	
Colonic obstruction	Neoplasm; stricture: ischemic, diverticular, inflammatory
Anal sphincter spasm	Anal fissure, painful hemorrhoids
Medications	
Chronic	
Irritable bowel syndrome	Constipation-predominant, alternating
Medications	Ca^{2+} blockers, antidepressants
Colonic pseudoobstruction	Slow-transit constipation, megacolon (rare Hirschsprung's, Chagas' diseases)
Disorders of rectal evacuation	Pelvic floor dysfunction; anismus; descending perineum syndrome; rectal mucosal prolapse; rectocele
Endocrinopathies	Hypothyroidism, hypercalcemia, pregnancy
Psychiatric disorders	Depression, eating disorders, drugs
Neurologic disease	Parkinsonism, multiple sclerosis, spinal cord injury
Generalized muscle disease	Progressive systemic sclerosis

Treatment: Constipation

A careful history should explore the patient's symptoms and confirm whether he or she is indeed constipated based on frequency (e.g., fewer than three bowel movements per week), consistency (lumpy/hard), excessive straining, prolonged defecation time, or need to support the perineum or digitate the anorectum. In the vast majority of cases (probably >90%), there is no underlying cause (e.g., cancer, depression, or hypothyroidism), and constipation responds to ample hydration, exercise, and supplementation of dietary fiber (15–25 g/d). A good diet and medication history and attention to psychosocial issues are key. Physical examination and, particularly, a rectal examination should exclude fecal impaction and most of the important diseases that present with constipation and possibly indicate features suggesting an evacuation disorder (e.g., high anal sphincter tone).

The presence of weight loss, rectal bleeding, or anemia with constipation mandates either flexible sigmoidoscopy plus barium enema or colonoscopy alone, particularly in patients >40 years, to exclude structural diseases such as cancer or strictures. Colonoscopy alone is most cost-effective in this setting because it provides an opportunity to biopsy mucosal lesions, perform polypectomy, or dilate strictures. Barium enema has advantages over colonoscopy in the patient with isolated constipation because it is less costly and identifies colonic dilation and all significant mucosal lesions or strictures that are likely to present with constipation. Melanosis coli, or pigmentation of the colon mucosa, indicates the use of anthraquinone laxatives such as cascara or senna; however, this is usually apparent from a careful history. An unexpected disorder such as megacolon or cathartic colon may also be detected by colonic radiographs. Measurement of serum calcium, potassium, and thyroid-stimulating hormone levels will identify rare patients with metabolic disorders.

Patients with more troublesome constipation may not respond to fiber alone and may be helped by a

bowel-training regimen: taking an osmotic laxative (lactulose, sorbitol, polyethylene glycol) and evacuating with enema or glycerine suppository as needed. After breakfast, a distraction-free 15–20 min on the toilet without straining is encouraged. Excessive straining may lead to development of hemorrhoids, and, if there is weakness of the pelvic floor or injury to the pudendal nerve, may result in obstructed defecation from descending perineum syndrome several years later. Those few who do not benefit from the simple measures delineated above or require long-term treatment with potent laxatives, with the attendant risk of developing laxative abuse syndrome, are assumed to have severe or intractable constipation and should have further investigation **(Fig. 40-4)**. Novel agents that induce secretion (e.g., lubiprostone, a chloride channel activator) are also available.

Figure 40-4

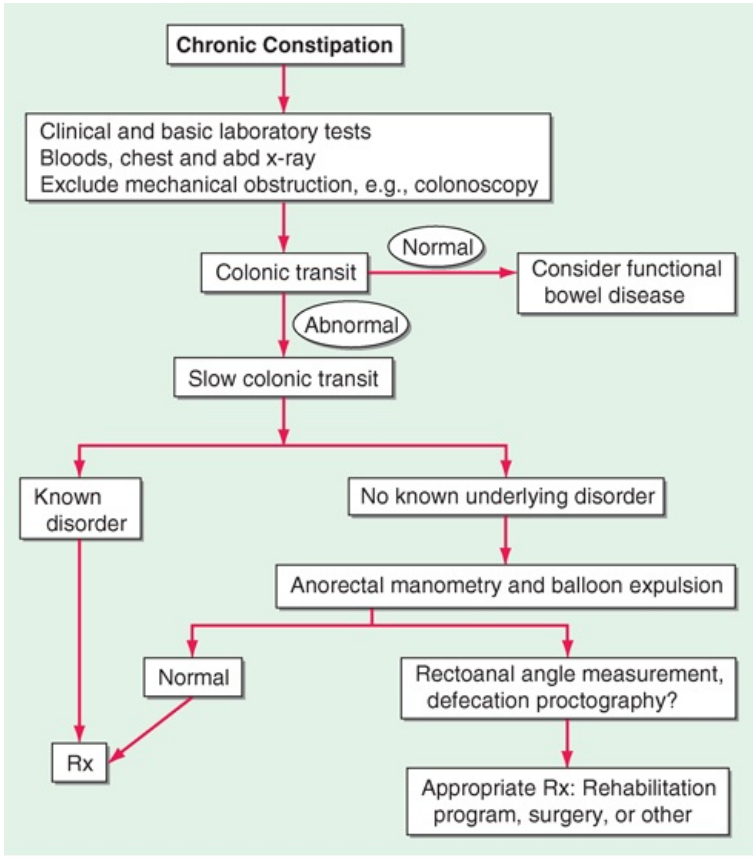

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the management of constipation.

Investigation of Severe Constipation

A small minority (probably <5%) of patients have severe or "intractable" constipation. These are the patients most likely to be seen by gastroenterologists or in referral centers. Further observation of the patient may occasionally reveal a previously unrecognized cause, such as an evacuation disorder, laxative abuse, malingering, or psychological disorder. In these patients, evaluations of the physiologic function of the colon and pelvic floor and of psychological status aid in the rational choice of treatment. Even among these highly selected patients with severe constipation, a cause can be identified in only about two-thirds of tertiary referral patients (see below).

MEASUREMENT OF COLONIC TRANSIT

Radiopaque marker transit tests are easy, repeatable, generally safe, inexpensive, reliable, and highly applicable in evaluating constipated patients in clinical practice. Several validated methods are very simple. For example, radiopaque markers are ingested; an abdominal flat film taken 5 days later should indicate passage of 80% of the markers out of the colon without the use of laxatives or enemas. This test does not provide useful information about the transit profile of the stomach and small bowel.

Radioscintigraphy with a delayed-release capsule containing radiolabeled particles has been used to noninvasively characterize normal, accelerated, or delayed colonic function over 24–48 h with low radiation exposure. This approach simultaneously assesses gastric, small bowel (which may be important in ~20% of patients with delayed colonic transit because they reflect a more generalized GI motility disorder), and colonic transit. The disadvantages are the greater cost and the need for specific

materials prepared in a nuclear medicine laboratory.

ANORECTAL AND PELVIC FLOOR TESTS

Pelvic floor dysfunction is suggested by the inability to evacuate the rectum, a feeling of persistent rectal fullness, rectal pain, the need to extract stool from the rectum digitally, application of pressure on the posterior wall of the vagina, support of the perineum during straining, and excessive straining. These significant symptoms should be contrasted with the sense of incomplete rectal evacuation, which is common in IBS.

Formal psychological evaluation may identify eating disorders, "control issues," depression, or post-traumatic stress disorders that may respond to cognitive or other intervention and may be important in restoring quality of life to patients who might present with chronic constipation.

A simple clinical test in the office to document a nonrelaxing puborectalis muscle is to have the patient strain to expel the index finger during a digital rectal examination. Motion of the puborectalis posteriorly during straining indicates proper coordination of the pelvic floor muscles.

Measurement of perineal descent is relatively easy to gauge clinically by placing the patient in the left decubitus position and watching the perineum to detect inadequate descent (<1.5 cm, a sign of pelvic floor dysfunction) or perineal ballooning during straining relative to bony landmarks (>4 cm, suggesting excessive perineal descent).

A useful overall test of evacuation is the balloon expulsion test. A balloon-tipped urinary catheter is placed and inflated with 50 mL of water. Normally, a patient can expel it while seated on a toilet or in the left lateral decubitus position. In the lateral position, the weight needed to facilitate expulsion of the balloon is determined; normally, expulsion occurs with <200 g added.

Anorectal manometry, when used in the evaluation of patients with severe constipation, may find an excessively high resting (>80 mmHg) or squeeze anal sphincter tone, suggesting anismus (anal sphincter spasm). This test also identifies rare syndromes, such as adult Hirschsprung's disease, by the absence of the rectoanal inhibitory reflex.

Defecography (a dynamic barium enema including lateral views obtained during barium expulsion) reveals "soft abnormalities" in many patients; the most relevant findings are the measured changes in rectoanal angle, anatomic defects of the rectum such as internal mucosal prolapse, and enteroceles or rectoceles. Surgically remediable conditions are identified in only a few patients. These include severe, whole-thickness intussusception with complete outlet obstruction due to funnel-shaped plugging at the anal canal or an extremely large rectocele that fills preferentially during attempts at defecation instead of expulsion of the barium through the anus. In summary, defecography requires an interested and experienced radiologist, and abnormalities are not pathognomonic for pelvic floor dysfunction. The most common cause of outlet obstruction is failure of the puborectalis muscle to relax; this is not identified by defecography but requires a dynamic study such as proctography. MRI is being developed as an alternative and provides more information about the structure and function of the pelvic floor, distal colorectum, and anal sphincters.

Dynamic imaging studies such as proctography during defecation or scintigraphic expulsion of artificial stool help measure perineal descent and the rectoanal angle during rest, squeezing, and straining, and scintigraphic expulsion quantitates the amount of "artificial stool" emptied. Lack of straightening of the rectoanal angle by at least 15° during defecation confirms pelvic floor dysfunction.

Neurologic testing (electromyography) is more helpful in the evaluation of patients with incontinence than of those with symptoms suggesting obstructed defecation. The absence of neurologic signs in the lower extremities suggests that any documented denervation of the puborectalis results from pelvic (e.g., obstetric) injury or from stretching of the pudendal nerve by chronic, long-standing straining. Constipation is common among patients with spinal cord injuries, neurologic diseases such as Parkinson's disease, multiple sclerosis, and diabetic neuropathy.

Spinal-evoked responses during electrical rectal stimulation or stimulation of external anal sphincter contraction by applying magnetic stimulation over the lumbosacral cord identify patients with limited sacral neuropathies with sufficient residual nerve conduction to attempt biofeedback training.

In summary, a balloon expulsion test is an important screening test for anorectal dysfunction. If positive, an anatomic evaluation of the rectum or anal sphincters and an assessment of pelvic floor relaxation are the tools for evaluating patients in whom obstructed defecation is suspected.

Treatment: Constipation

After the cause of constipation is characterized, a treatment decision can be made. Slow-transit constipation requires aggressive medical or surgical treatment; anismus or pelvic floor dysfunction usually responds to biofeedback management **(Fig. 40-4)**. However, only ~60% of patients with severe constipation are found to have such a physiologic disorder (half with colonic transit delay and half with evacuation disorder). Patients with spinal cord injuries or other neurologic disorders require a dedicated bowel regimen that often includes rectal stimulation, enema therapy, and carefully timed laxative therapy.

Patients with slow-transit constipation are treated with bulk, osmotic, prokinetic, secretory, and stimulant laxatives including fiber, psyllium, milk of magnesia, lactulose, polyethylene glycol (colonic lavage solution), lubiprostone, and bisacodyl. Newer treatment aimed at enhancing motility and secretion may have application in circumstances such as constipation-predominant IBS in females or severe constipation. If a three- to six-month trial of medical therapy fails and patients continue to have documented slow-transit constipation unassociated with obstructed defecation, the patients should be considered for laparoscopiccolectomy with ileorectostomy; however, this should not be undertaken if there is continued evidence of an evacuation disorder or a generalized GI dysmotility. Referral to a specialized center for further tests of colonic motor function is warranted. The decision to resort to surgery is facilitated in the presence of megacolon and megarectum. The complications after surgery include small-bowel obstruction (11%) and fecal soiling, particularly at night during the first postoperative year. Frequency of defecation is 3–8 per day during the first year, dropping to 1–3 per day from the second year after surgery.

Patients who have a combined (evacuation and transit/motility) disorder should pursue pelvic floor retraining (biofeedback and muscle relaxation), psychological counseling, and dietetic advice first, followed by colectomy and ileorectostomy if colonic transit studies do not normalize and symptoms are intractable despite biofeedback and optimized medical therapy. In patients with pelvic floor dysfunction alone, biofeedback training has a 70–80% success rate, measured by the acquisition of comfortable stool habits. Attempts to manage pelvic floor dysfunction with operations (internal anal sphincter or puborectalis muscle division) have achieved only mediocre success and have been largely abandoned.

FURTHER READINGS

Bartlett JG: Narrative review: The new epidemic of *Clostridium difficile*–associated enteric disease. Ann Intern Med 145:758, 2006

Binder HJ : Causes of chronic diarrhea. N Engl J Med 355:236, 2006

Camilleri M: Chronic diarrhea: A review on pathophysiology and management for the clinical gastroenterologist. Clin Gastroenterol Hepatol 2:198, 2004

DuPont HL: Clinical practice. Bacterial diarrhea. N Engl J Med 361:1560, 2009

Lembo A, Camilleri M: Chronic constipation. N Engl J Med 349:1360, 2003

Musher DM, Musher BL : Contagious acute gastrointestinal infections. N Engl J Med 351:2417, 2004

Navaneethan U, Giannella RA: Mechanisms of infectious diarrhea. Nat Clin Pract Gastroenterol Hepatol 5:637, 2008

Riddle MS, et al: Effect of adjunctive loperamide in combination with antibiotics on treatment outcomes in traveler's diarrhea: a systematic review and meta-analysis. Clin Infect Dis 47:1007, 2008

Rostom A, et al: American Gastroenterological Association (AGA) Institute technical review on the diagnosis and management of celiac disease. Gastroenterology 131:1981, 2006

Wald A: Clinical practice. Fecal incontinence in adults. N Engl J Med 356:1648, 2007

Walters JR, et al: A new mechanism for bile acid diarrhea: defective feedback inhibition of bile acid biosynthesis. Clin Gastroenterol Hepatol 7:1189, 2009

Whitehead WE, Bharucha AE: Diagnosis and treatment of pelvic floor disorders: What's new and what to do. Gastroenterology 138:1231, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GASTROINTESTINAL BLEEDING: INTRODUCTION

Bleeding from the gastrointestinal (GI) tract may present in five ways. *Hematemesis* is vomitus of red blood or "coffee-grounds" material. *Melena* is black, tarry, foul-smelling stool. *Hematochezia* is the passage of bright red or maroon blood from the rectum. *Occult GI bleeding* (GIB) may be identified in the absence of overt bleeding by a fecal occult blood test or the presence of iron deficiency. Finally, patients may present only with *symptoms of blood loss or anemia* such as lightheadedness, syncope, angina, or dyspnea.

Sources of Gastrointestinal Bleeding

UPPER GASTROINTESTINAL SOURCES OF BLEEDING

(Table 41-1) The annual incidence of hospital admissions for upper GIB (UGIB) in the United States and Europe is ~0.1%, with a mortality rate of ~5–10%. Patients rarely die from exsanguination; rather, they die due to decompensation from other underlying illnesses. The mortality rate for patients <60 years in the absence of major concurrent illness is <1%. Independent predictors of rebleeding and death in patients hospitalized with UGIB include increasing age, comorbidities, and hemodynamic compromise (tachycardia or hypotension).

Table 41-1 Sources of Bleeding in Patients Hospitalized for Upper GI Bleeding

Sources of Bleeding	Proportion of Patients, %
Ulcers	31–67
Varices	6–39
Mallory-Weiss tears	2–8
Gastroduodenal erosions	2–18
Erosive esophagitis	1–13
Neoplasm	2–8
Vascular ectasias	0–6
No source identified	5–14

Source: Data on hospitalizations from year 2000 onward from Am J Gastroenterol 98:1494, 2003; Gastrointest Endosc 57:AB147, 2003; 60;875, 2004; Eur J Gastroenterol Hepatol 16:177, 2004; 17:641, 2005; J Clin Gastroenterol 42:128, 2008; World J Gastroenterol 14:5046, 2008; Dig Dis Sci 54:333, 2009.

Peptic ulcers are the most common cause of UGIB, accounting for up to ~50% of cases; an increasing proportion is due to nonsteroidal anti-inflammatory drugs (NSAIDs), with the prevalence of *Helicobacter pylori* decreasing. Mallory-Weiss tears account for ~5–10% of cases. The proportion of patients bleeding from varices varies widely from ~5 to 40%, depending on the population. Hemorrhagic or erosive gastropathy (e.g., due to NSAIDs or alcohol) and erosive esophagitis often cause mild UGIB, but major bleeding is rare.

Peptic Ulcers

In addition to clinical features, characteristics of an ulcer at endoscopy provide important prognostic information. One-third of patients with active bleeding or a nonbleeding visible vessel have further bleeding that requires urgent surgery if they are treated conservatively. These patients clearly benefit from endoscopic therapy with bipolar electrocoagulation; heater probe; injection therapy (e.g., absolute alcohol, 1:10,000 epinephrine); and/or clips with reductions in bleeding, hospital stay, mortality rate, and costs. In contrast, patients with clean-based ulcers have rates of recurrent bleeding approaching zero. If there is no other reason for hospitalization, such patients may be discharged on the first hospital day, following stabilization. Patients without clean-based ulcers should usually remain in the hospital for three days because most episodes of recurrent bleeding occur within three days.

Randomized controlled trials document that a high-dose, constant-infusion IV proton pump inhibitor (PPI) (e.g., omeprazole 80-mg bolus and 8-mg/h infusion), designed to sustain intragastric pH > 6 and enhance clot stability, decreases further bleeding and mortality in patients with high-risk ulcers (active bleeding, nonbleeding visible vessel, adherent clot) when given after endoscopic therapy. Institution of PPI therapy at presentation in all patients with UGIB decreases high-risk ulcer characteristics (e.g.,

active bleeding) but does not significantly improve outcomes such as further bleeding, transfusions, or mortality as compared to initiating therapy only when high-risk ulcers are identified at the time of endoscopy.

Approximately one-third of patients with bleeding ulcers will rebleed within the next 1–2 years if no preventive strategies are employed. Prevention of recurrent bleeding focuses on the three main factors in ulcer pathogenesis, *H. pylori*, NSAIDs, and acid. Eradication of *H. pylori* in patients with bleeding ulcers decreases rates of rebleeding to <5%. If a bleeding ulcer develops in a patient taking NSAIDs, the NSAIDs should be discontinued, if possible. If NSAIDs must be continued or reinstituted, a cyclooxygenase 2 (COX-2) selective inhibitor (coxib) plus a PPI should be used. PPI co-therapy alone or a coxib alone is associated with an annual rebleeding rate of ~10% in patients with a recent bleeding ulcer, while combination of a coxib and PPI provides a further significant decrease in recurrent ulcer bleeding. Patients with cardiovascular disease who develop bleeding ulcers while taking low-dose aspirin should restart aspirin as soon as possible after their bleeding episode (e.g., ≤ 7 days). A randomized trial showed that failure to restart aspirin was associated with a nonsignificant difference in rebleeding (5% vs. 10% at 30 days), but a significant increase in mortality at 30 days (9% vs. 1%) and 8 weeks (13% vs. 1%) as compared to immediate reinstitution of aspirin. Patients with bleeding ulcers unrelated to *H. pylori* or NSAIDs should remain on full-dose antisecretory therapy indefinitely. **Peptic ulcers are discussed in Chap. 293.**

Mallory-Weiss Tears

The classic history is vomiting, retching, or coughing preceding hematemesis, especially in an alcoholic patient. Bleeding from these tears, which are usually on the gastric side of the gastroesophageal junction, stops spontaneously in 80–90% of patients and recurs in only 0–7%. Endoscopic therapy is indicated for actively bleeding Mallory-Weiss tears. Angiographic therapy with embolization and operative therapy with oversewing of the tear are rarely required. **Mallory-Weiss tears are discussed in Chap. 292.**

Esophageal Varices

Patients with variceal hemorrhage have poorer outcomes than patients with other sources of UGIB. Endoscopic therapy for acute bleeding and repeated sessions of endoscopic therapy to eradicate esophageal varices significantly reduce rebleeding and mortality. Ligation is the endoscopic therapy of choice for esophageal varices because it has less rebleeding, a lower mortality rate, fewer local complications, and it requires fewer treatment sessions to achieve variceal eradication than sclerotherapy.

Octreotide (50-μg bolus and 50-μg/h IV infusion for 2–5 days) further helps in the control of acute bleeding when used in combination with endoscopic therapy. Other vasoactive agents such as somatostatin and terlipressin, available outside the United States, are also effective. Antibiotic therapy (e.g., ceftriaxone) is also recommended for patients with cirrhosis presenting with UGIB because antibiotics decrease bacterial infections and mortality in this population. Over the long term, treatment with nonselective beta blockers decreases recurrent bleeding from esophageal varices. Chronic therapy with beta blockers plus endoscopic ligation is recommended for prevention of recurrent esophageal variceal bleeding.

In patients who have persistent or recurrent bleeding despite endoscopic and medical therapy, more invasive therapy with transjugular intrahepatic portosystemic shunt (TIPS) is recommended. Older studies indicate that most patients with TIPS developed shunt stenosis within 1–2 years and required reintervention to maintain shunt patency. The use of coated stents appears to decrease shunt dysfunction by ~50% in the first 2 years. A randomized comparison of TIPS (with uncoated stents) and distal splenorenal shunt in Child-Pugh class A or B cirrhotic patients with refractory variceal bleeding revealed no significant difference in rebleeding, encephalopathy, or survival, but had a much higher rate of reintervention with TIPS (82% vs. 11%). Therefore, decompressive surgery may be an option in patients with milder, well-compensated cirrhosis.

Portal hypertension is also responsible for bleeding from gastric varices, varices in the small and large intestine, and portal hypertensive gastropathy and enterocolopathy.

Hemorrhagic and Erosive Gastropathy ("Gastritis")

Hemorrhagic and erosive gastropathy, often labeled gastritis, refers to endoscopically visualized subepithelial hemorrhages and erosions. These are mucosal lesions and, thus, do not cause major bleeding. They develop in various clinical settings, the most important of which are NSAID use, alcohol intake, and stress. Half of patients who chronically ingest NSAIDs have erosions (15–30% have ulcers), while up to 20% of actively drinking alcoholic patients with symptoms of UGIB have evidence of subepithelial hemorrhages or erosions.

Stress-related gastric mucosal injury occurs only in extremely sick patients: those who have experienced serious trauma, major surgery, burns covering more than one-third of the body surface area, major intracranial disease, or severe medical illness (i.e., ventilator dependence, coagulopathy). Significant bleeding probably does not develop unless ulceration occurs. The mortality rate in these patients is quite high because of their serious underlying illnesses.

The incidence of bleeding from stress-related gastric mucosal injury or ulceration has decreased dramatically in recent years, most likely due to better care of critically ill patients. Pharmacologic prophylaxis for bleeding may be considered in the high-risk patients mentioned above. Multiple trials document the efficacy of intravenous H_2-receptor antagonist therapy, which is more effective than sucralfate but not superior to a PPI immediate-release suspension given via nasogastric tube. Prophylactic therapy decreases bleeding but does not lower the mortality rate.

Other Causes

Other less frequent causes of UGIB include erosive duodenitis, neoplasms, aortoenteric fistulas, vascular lesions [including hereditary hemorrhagic telangiectasias (Osler-Weber-Rendu) and gastric antral vascular ectasia ("watermelon stomach")], Dieulafoy's lesion (in which an aberrant vessel in the mucosa bleeds from a pinpoint mucosal defect), prolapse gastropathy (prolapse of proximal stomach into esophagus with retching, especially in alcoholics), and hemobilia or hemosuccus pancreaticus (bleeding from the bile duct or pancreatic duct).

SMALL-INTESTINAL SOURCES OF BLEEDING

Small-intestinal sources of bleeding (bleeding from sites beyond the reach of the standard upper endoscope) are difficult to diagnose and are responsible for the majority of cases of obscure GIB. Fortunately, small-intestinal bleeding is uncommon. The most common causes in adults are vascular ectasias, tumors (e.g., adenocarcinoma, leiomyoma, lymphoma, benign polyps, carcinoid, metastases, and lipoma), and NSAID-induced erosions and ulcers. Other less common causes in adults include Crohn's disease, infection, ischemia, vasculitis, small-bowel varices, diverticula, Meckel's diverticulum, duplication cysts, and intussusception.

Meckel's diverticulum is the most common cause of significant lower GIB (LGIB) in children, decreasing in frequency as a cause of bleeding with age. In adults <40–50 years, small-bowel tumors often account for obscure GIB; in patients >50–60 years, vascular ectasias and NSAID-induced lesions are more commonly responsible.

Vascular ectasias should be treated with endoscopic therapy if possible. Surgical therapy can be used for vascular ectasias isolated to a segment of the small intestine when endoscopic therapy is unsuccessful. Although estrogen/progesterone compounds have been used for vascular ectasias, a double-blind trial found no benefit in prevention of recurrent bleeding. Isolated lesions, such as tumors, diverticula, or duplications, are generally treated with surgical resection.

COLONIC SOURCES OF BLEEDING

The incidence of hospitalizations for LGIB is ≥20% that for UGIB. Hemorrhoids are probably the most common cause of LGIB; anal fissures also cause minor bleeding and pain. If these local anal processes, which rarely require hospitalization, are excluded, the most common causes of LGIB in adults are diverticula, vascular ectasias (especially in the proximal colon of patients >70 years), neoplasms (primarily adenocarcinoma), and colitis—most commonly infectious or idiopathic inflammatory bowel disease, but occasionally ischemic or radiation-induced. Uncommon causes include post-polypectomy bleeding, solitary rectal ulcer syndrome, NSAID-induced ulcers or colitis, trauma, varices (most commonly rectal), lymphoid nodular hyperplasia, vasculitis, and aortocolic fistulas. In children and adolescents, the most common colonic causes of significant GIB are inflammatory bowel disease and juvenile polyps.

Diverticular bleeding is abrupt in onset, usually painless, sometimes massive, and often from the right colon; minor and occult bleeding is not characteristic. Clinical reports suggest that bleeding colonic diverticula stop bleeding spontaneously in ~80% of patients and rebleed in about 20–25% of patients. Intraarterial vasopressin or embolization by superselective technique should stop bleeding in a majority of patients. If bleeding persists or recurs, segmental surgical resection is indicated.

Bleeding from right colonic vascular ectasias in the elderly may be overt or occult; it tends to be chronic and only occasionally is hemodynamically significant. Endoscopic hemostatic therapy may be useful in the treatment of vascular ectasias, as well as discrete bleeding ulcers and post-polypectomy bleeding, while endoscopic polypectomy, if possible, is used for bleeding colonic polyps. Surgical therapy is generally required for major, persistent, or recurrent bleeding from the wide variety of colonic sources of GIB that cannot be treated medically, angiographically, or endoscopically.

Approach to the Patient: Gastrointestinal Bleeding

Measurement of the heart rate and blood pressure is the best way to initially assess a patient with GIB. Clinically significant bleeding leads to postural changes in heart rate or blood pressure, tachycardia, and, finally, recumbent hypotension. In contrast, the hemoglobin does not fall immediately with acute GIB, due to proportionate reductions in plasma and red cell volumes (i.e., "people bleed whole blood"). Thus, hemoglobin may be normal or only minimally decreased at the initial presentation of a severe bleeding episode. As extravascular fluid enters the vascular space to restore volume, the hemoglobin falls, but this process may take up to 72 h. Patients with slow, chronic GIB may have very low hemoglobin values despite normal blood pressure and heart rate. With the development of iron-deficiency anemia, the mean corpuscular volume will be low and red blood cell distribution width will increase.

DIFFERENTIATION OF UPPER FROM LOWER GIB

Hematemesis indicates an upper GI source of bleeding (above the ligament of Treitz). Melena indicates that blood has been present in the GI tract for at least 14 h (and as long as 3–5 days). The more proximal the bleeding site, the more likely melena will occur. Hematochezia usually represents a lower GI source of bleeding, although an upper GI lesion may bleed so briskly that blood does not remain in the bowel long enough for melena to develop. When hematochezia is the presenting symptom of UGIB, it is associated with hemodynamic instability and dropping hemoglobin. Bleeding lesions of the small bowel may present as melena or hematochezia. Other clues to UGIB include hyperactive bowel sounds and an elevated blood urea nitrogen level (due to volume depletion and blood proteins absorbed in the small intestine).

A nonbloody nasogastric aspirate may be seen in up to 18% of patients with UGIB—usually from a duodenal source. Even a bile-stained appearance does not exclude a bleeding postpyloric lesion because reports of bile in the aspirate are incorrect in ~50% of cases. Testing of aspirates that are not grossly bloody for occult blood is not useful.

DIAGNOSTIC EVALUATION OF THE PATIENT WITH GIB

Upper Gib

(Fig. 41-1) History and physical examination are not usually diagnostic of the source of GIB. Upper endoscopy is the test of choice in patients with UGIB and should be performed urgently in patients who present with hemodynamic instability (hypotension, tachycardia, or postural changes in heart rate or blood pressure). Early endoscopy is also beneficial in cases of milder bleeding for management decisions. Patients with major bleeding and high-risk endoscopic findings (e.g., varices, ulcers with active bleeding or a visible vessel) benefit from endoscopic hemostatic therapy, while patients with low-risk lesions (e.g., clean-based ulcers, nonbleeding Mallory-Weiss tears, erosive or hemorrhagic gastropathy) who have stable vital signs and hemoglobin, and no other medical problems, can be discharged home.

Figure 41-1

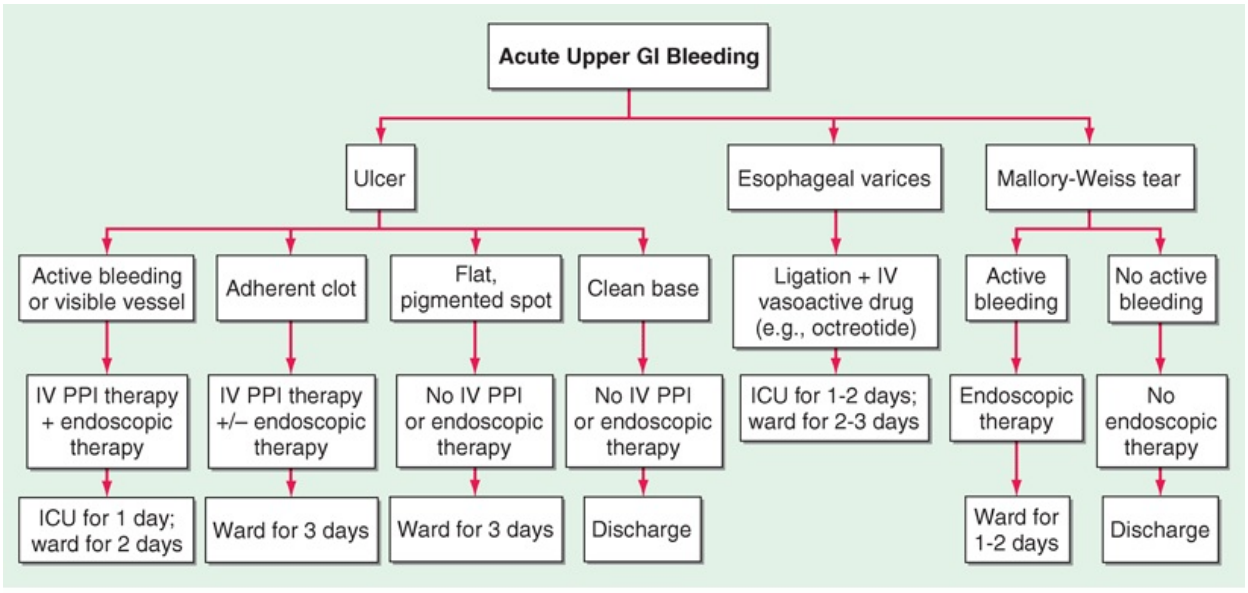

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Suggested algorithm for patients with acute upper gastrointestinal bleeding. Recommendations on level of care and time of discharge assume patient is stabilized without further bleeding or other concomitant medical problems. ICU, intensive care unit; PPI, proton pump inhibitor.

Lower Gib

(Fig. 41-2) Patients with hematochezia and hemodynamic instability should have upper endoscopy to rule out an upper GI source before evaluation of the lower GI tract. Patients with presumed LGIB may undergo early sigmoidoscopy for the detection of obvious, low-lying lesions. However, the procedure is difficult with brisk bleeding, and it is usually not possible to identify the area of bleeding. Sigmoidoscopy is useful primarily in patients <40 years with minor bleeding.

Figure 41-2

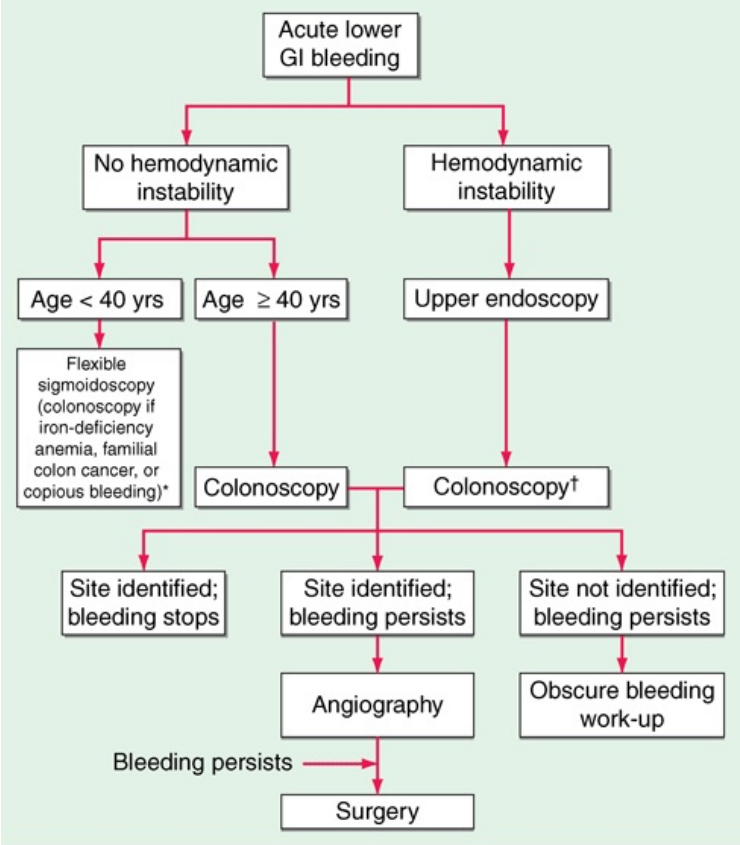

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Suggested algorithm for patients with acute lower gastrointestinal bleeding. *Some suggest colonoscopy for any degree of rectal bleeding in patients <40 years as well. †If massive bleeding does not allow time for colonic lavage, proceed to angiography.

Colonoscopy after an oral lavage solution is the procedure of choice in patients admitted with LGIB unless bleeding is too massive or unless sigmoidoscopy has disclosed an obvious actively bleeding lesion. ^{99m}Tc-labeled red cell scan allows repeated imaging for up to 24 h and may identify the general location of bleeding. However, radionuclide scans should be interpreted with caution because results, especially from later images, are highly variable. In active LGIB, angiography can detect the site of bleeding (extravasation of contrast into the gut) and permits treatment with embolization or intraarterial infusion of vasopressin. Even after bleeding has stopped, angiography may identify lesions with abnormal vasculature, such as vascular ectasias or tumors.

Gib of Obscure Origin

Obscure GIB is defined as persistent or recurrent bleeding for which no source has been identified by routine endoscopic and contrast x-ray studies; it may be overt (melena, hematochezia) or occult (iron-deficiency anemia). Current guidelines suggest angiography as the initial test for massive obscure bleeding, and video capsule endoscopy, which allows examination of the entire small intestine, for all others. Push enteroscopy, with a specially designed enteroscope or a pediatric colonoscope to inspect the entire duodenum and part of the jejunum, also may be considered as an initial evaluation. A systematic review of 14 trials comparing push enteroscopy to capsule revealed "clinically significant findings" in 26% and 56% of patients, respectively. However, in contrast to enteroscopy, lack of control of the capsule prevents its manipulation and full visualization of the intestine; in addition, tissue cannot be sampled and therapy cannot be applied.

If capsule endoscopy is positive, management (e.g., enteroscopy, laparoscopy) is dictated by the finding. If capsule is negative, current recommendations suggest patients may be either observed, or if their clinical course mandates (e.g., recurrent bleeding, need for transfusions or hospitalization), undergo further testing. Newer endoscopic techniques (e.g., double-balloon, single-balloon, and spiral enteroscopy) allow the endoscopist to examine, obtain specimens from, and provide therapy to much or all of the small intestine. Newer imaging techniques (CT and MR enterography) are now frequently being used in place of older specialized small-bowel radiographic exams (e.g., enteroclysis). Other tests include ^{99m}Tc-labeled red blood cell scintigraphy; angiography, which may be useful even if bleeding has subsided because it may disclose vascular anomalies or tumor vessels; and ^{99m}Tc-pertechnetate scintigraphy for diagnosis of Meckel's diverticulum (especially in young patients). When all tests are

unrevealing, intraoperative endoscopy is indicated in patients with severe recurrent or persistent bleeding requiring repeated transfusions.

Positive Fecal Occult Blood Test

Fecal occult blood testing is recommended only for colorectal cancer screening and may be used in average-risk adults (beginning at age 50) and in adults with a first-degree relative with colorectal neoplasm at ≥60 years or two second-degree relatives with colorectal cancer (beginning at age 40). A positive test necessitates colonoscopy. If evaluation of the colon is negative, further workup is not recommended unless iron-deficiency anemia or GI symptoms are present.

FURTHER READINGS

Barkun AN et al: International consensus recommendations on the management of patients with nonvariceal upper gastrointestinal bleeding. Ann Intern Med 152:101, 2010

Conrad SA et al: Randomized, double-blind comparison of immediate-release omeprazole oral suspension versus intravenous cimetidine for the prevention of upper gastrointestinal bleeding in critically ill patients. Crit Care Med 33:760, 2005

Garcia-Tsao G et al: Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. Am J Gastroenterol 102:2086, 2007

Henderson JM et al: Distal splenorenal shunt versus transjugular intrahepatic portal systematic shunt for variceal bleeding: A randomized trial. Gastroenterology 130:1643, 2006

Laine L, *McQuaid* KR: Endoscopic therapy for bleeding ulcers: an evidence-based approach based on meta-analyses of randomized controlled trials. Clin Gastroenterol Hepatol 7:33, 2009

Raju GS et al: American Gastroenterological Association (AGA) Institute technical review on obscure gastrointestinal bleeding. Gastroenterology 133:1697, 2007

Rockall TA et al: Risk assessment after acute upper gastrointestinal haemorrhage. Gut 38:316, 1996

Sung JJY et al: Continuation of low-dose aspirin in peptic ulcer bleeding. A randomized trial. Ann Intern Med 152:1, 2010

Triester SL et al: A meta-analysis of the yield of capsule endoscopy compared to other diagnostic modalities in patients with obscure gastrointestinal bleeding. Am J Gastroenterol 100:2407, 2005

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

JAUNDICE: INTRODUCTION

Jaundice, or icterus, is a yellowish discoloration of tissue resulting from the deposition of bilirubin. Tissue deposition of bilirubin occurs only in the presence of serum hyperbilirubinemia and is a sign of either liver disease or, less often, a hemolytic disorder. The degree of serum bilirubin elevation can be estimated by physical examination. Slight increases in serum bilirubin are best detected by examining the sclerae, which have a particular affinity for bilirubin due to their high elastin content. The presence of scleral icterus indicates a serum bilirubin of at least 51 μmol/L (3 mg/dL). The ability to detect scleral icterus is made more difficult if the examining room has fluorescent lighting. If the examiner suspects scleral icterus, a second place to examine is underneath the tongue. As serum bilirubin levels rise, the skin will eventually become yellow in light-skinned patients and even green if the process is long-standing; the green color is produced by oxidation of bilirubin to biliverdin.

The differential diagnosis for yellowing of the skin is limited. In addition to jaundice, it includes carotenoderma, the use of the drug quinacrine, and excessive exposure to phenols. Carotenoderma is the yellow color imparted to the skin by the presence of carotene; it occurs in healthy individuals who ingest excessive amounts of vegetables and fruits that contain carotene, such as carrots, leafy vegetables, squash, peaches, and oranges. Unlike jaundice, where the yellow coloration of the skin is uniformly distributed over the body, in carotenoderma, the pigment is concentrated on the palms, soles, forehead, and nasolabial folds. Carotenoderma can be distinguished from jaundice by the sparing of the sclerae. Quinacrine causes a yellow discoloration of the skin in 4–37% of patients treated with it. Unlike carotene, quinacrine can cause discoloration of the sclerae.

Another sensitive indicator of increased serum bilirubin is darkening of the urine, which is due to the renal excretion of conjugated bilirubin. Patients often describe their urine as tea- or cola-colored. Bilirubinuria indicates an elevation of the direct serum bilirubin fraction and, therefore, the presence of liver disease.

Increased serum bilirubin levels occur when an imbalance exists between bilirubin production and clearance. A logical evaluation of the patient who is jaundiced requires an understanding of bilirubin production and metabolism.

Production and Metabolism of Bilirubin

(See also Chap. 303) Bilirubin, a tetrapyrrole pigment, is a breakdown product of heme (ferroprotoporphyrin IX). About 70–80% of the 250–300 mg of bilirubin produced each day is derived from the breakdown of hemoglobin in senescent red blood cells. The remainder comes from prematurely destroyed erythroid cells in bone marrow and from the turnover of hemoproteins such as myoglobin and cytochromes found in tissues throughout the body.

The formation of bilirubin occurs in reticuloendothelial cells, primarily in the spleen and liver. The first reaction, catalyzed by the microsomal enzyme heme oxygenase, oxidatively cleaves the α bridge of the porphyrin group and opens the heme ring. The end products of this reaction are biliverdin, carbon monoxide, and iron. The second reaction, catalyzed by the cytosolic enzyme biliverdin reductase, reduces the central methylene bridge of biliverdin and converts it to bilirubin. Bilirubin formed in the reticuloendothelial cells is virtually insoluble in water. This is due to tight internal hydrogen bonding between the water-soluble moieties of bilirubin, proprionic acid carboxyl groups of one dipyrrolic half of the molecule with the imino and lactam groups of the opposite half. This configuration blocks solvent access to the polar residues of bilirubin and places the hydrophobic residues on the outside. To be transported in blood, bilirubin must be solubilized. This is accomplished by its reversible, noncovalent binding to albumin. Unconjugated bilirubin bound to albumin is transported to the liver, where it, but not the albumin, is taken up by hepatocytes via a process that at least partly involves carrier-mediated membrane transport. No specific bilirubin transporter has yet been identified (Chap. 303, Fig. 303-1).

After entering the hepatocyte, unconjugated bilirubin is bound in the cytosol to a number of proteins including proteins in the glutathione-S-transferase superfamily. These proteins serve both to reduce efflux of bilirubin back into the serum and to present the bilirubin for conjugation. In the endoplasmic reticulum, bilirubin is solubilized by conjugation to glucuronic acid, a process that disrupts the internal hydrogen bonds and yields bilirubin monoglucuronide and diglucuronide. The conjugation of glucuronic acid to bilirubin is catalyzed by bilirubin uridine diphosphate-glucuronosyl transferase (UDPGT). The now hydrophilic bilirubin conjugates diffuse from the endoplasmic reticulum to the canalicular membrane, where bilirubin monoglucuronide and diglucuronide are actively transported into

canalicular bile by an energy-dependent mechanism involving the multiple drug resistance protein 2.

The conjugated bilirubin excreted into bile drains into the duodenum and passes unchanged through the proximal small bowel. Conjugated bilirubin is not taken up by the intestinal mucosa. When the conjugated bilirubin reaches the distal ileum and colon, it is hydrolyzed to unconjugated bilirubin by bacterial β-glucuronidases. The unconjugated bilirubin is reduced by normal gut bacteria to form a group of colorless tetrapyrroles called urobilinogens. About 80–90% of these products are excreted in feces, either unchanged or oxidized to orange derivatives called urobilins. The remaining 10–20% of the urobilinogens are passively absorbed, enter the portal venous blood, and are reexcreted by the liver. A small fraction (usually <3 mg/dL) escapes hepatic uptake, filters across the renal glomerulus, and is excreted in urine.

Measurement of Serum Bilirubin

The terms direct and indirect bilirubin, conjugated and unconjugated bilirubin, respectively, are based on the original van den Bergh reaction. This assay, or a variation of it, is still used in most clinical chemistry laboratories to determine the serum bilirubin level. In this assay, bilirubin is exposed to diazotized sulfanilic acid, splitting into two relatively stable dipyrrylmethene azopigments that absorb maximally at 540 nm, allowing for photometric analysis. The direct fraction is that which reacts with diazotized sulfanilic acid in the absence of an accelerator substance such as alcohol. The direct fraction provides an approximate determination of the conjugated bilirubin in serum. The total serum bilirubin is the amount that reacts after the addition of alcohol. The indirect fraction is the difference between the total and the direct bilirubin and provides an estimate of the unconjugated bilirubin in serum.

With the van den Bergh method, the normal serum bilirubin concentration usually is 17 μmol/L (<1 mg/dL). Up to 30%, or 5.1 μmol/L (0.3 mg/dL), of the total may be direct-reacting (conjugated) bilirubin. Total serum bilirubin concentrations are between 3.4 and 15.4 μmol/L (0.2 and 0.9 mg/dL) in 95% of a normal population.

Several new techniques, although less convenient to perform, have added considerably to our understanding of bilirubin metabolism. First, they demonstrate that in normal persons or those with Gilbert's syndrome, almost 100% of the serum bilirubin is unconjugated; <3% is monoconjugated bilirubin. Second, in jaundiced patients with hepatobiliary disease, the total serum bilirubin concentration measured by these new, more accurate methods is lower than the values found with diazo methods. This suggests that there are diazo-positive compounds distinct from bilirubin in the serum of patients with hepatobiliary disease. Third, these studies indicate that, in jaundiced patients with hepatobiliary disease, monoglucuronides of bilirubin predominate over the diglucuronides. Fourth, part of the direct-reacting bilirubin fraction includes conjugated bilirubin that is covalently linked to albumin. This albumin-linked bilirubin fraction (*delta fraction,* or *biliprotein*) represents an important fraction of total serum bilirubin in patients with cholestasis and hepatobiliary disorders. Albumin-bound conjugated bilirubin is formed in serum when hepatic excretion of bilirubin glucuronides is impaired and the glucuronides are present in serum in increasing amounts. By virtue of its tight binding to albumin, the clearance rate of albumin-bound bilirubin from serum approximates the half-life of albumin, 12–14 days, rather than the short half-life of bilirubin, about 4 hours.

The prolonged half-life of albumin-bound conjugated bilirubin explains two previously unexplained enigmas in jaundiced patients with liver disease: (1) that some patients with conjugated hyperbilirubinemia do not exhibit bilirubinuria during the recovery phase of their disease because the bilirubin is covalently bound to albumin and therefore not filtered by the renal glomeruli, and (2) that the elevated serum bilirubin level declines more slowly than expected in some patients who otherwise appear to be recovering satisfactorily. Late in the recovery phase of hepatobiliary disorders, all the conjugated bilirubin may be in the albumin-linked form. Its value in serum falls slowly because of the long half-life of albumin.

Measurement of Urine Bilirubin

Unconjugated bilirubin is always bound to albumin in the serum, is not filtered by the kidney, and is not found in the urine. Conjugated bilirubin is filtered at the glomerulus and the majority is reabsorbed by the proximal tubules; a small fraction is excreted in the urine. Any bilirubin found in the urine is conjugated bilirubin. The presence of bilirubinuria implies the presence of liver disease. A urine dipstick test (Ictotest) gives the same information as fractionation of the serum bilirubin. This test is very accurate. A false-negative test is possible in patients with prolonged cholestasis due to the predominance of conjugated bilirubin covalently bound to albumin.

Approach to the Patient: Bilirubin

The bilirubin present in serum represents a balance between input from production of bilirubin and hepatic/biliary removal of the pigment. Hyperbilirubinemia may result from (1) overproduction of bilirubin; (2) impaired uptake, conjugation, or excretion of bilirubin; or (3) regurgitation of unconjugated or conjugated bilirubin from damaged hepatocytes or bile ducts. An increase in unconjugated bilirubin in serum results from either overproduction, impairment of uptake, or conjugation of bilirubin. An increase in conjugated bilirubin is due to decreased excretion into the bile ductules or backward

leakage of the pigment. The initial steps in evaluating the patient with jaundice are to determine (1) whether the hyperbilirubinemia is predominantly conjugated or unconjugated in nature, and (2) whether other biochemical liver tests are abnormal. The thoughtful interpretation of limited data will allow for a rational evaluation of the patient (Fig. 42-1). This discussion will focus solely on the evaluation of the adult patient with jaundice.

Figure 42-1

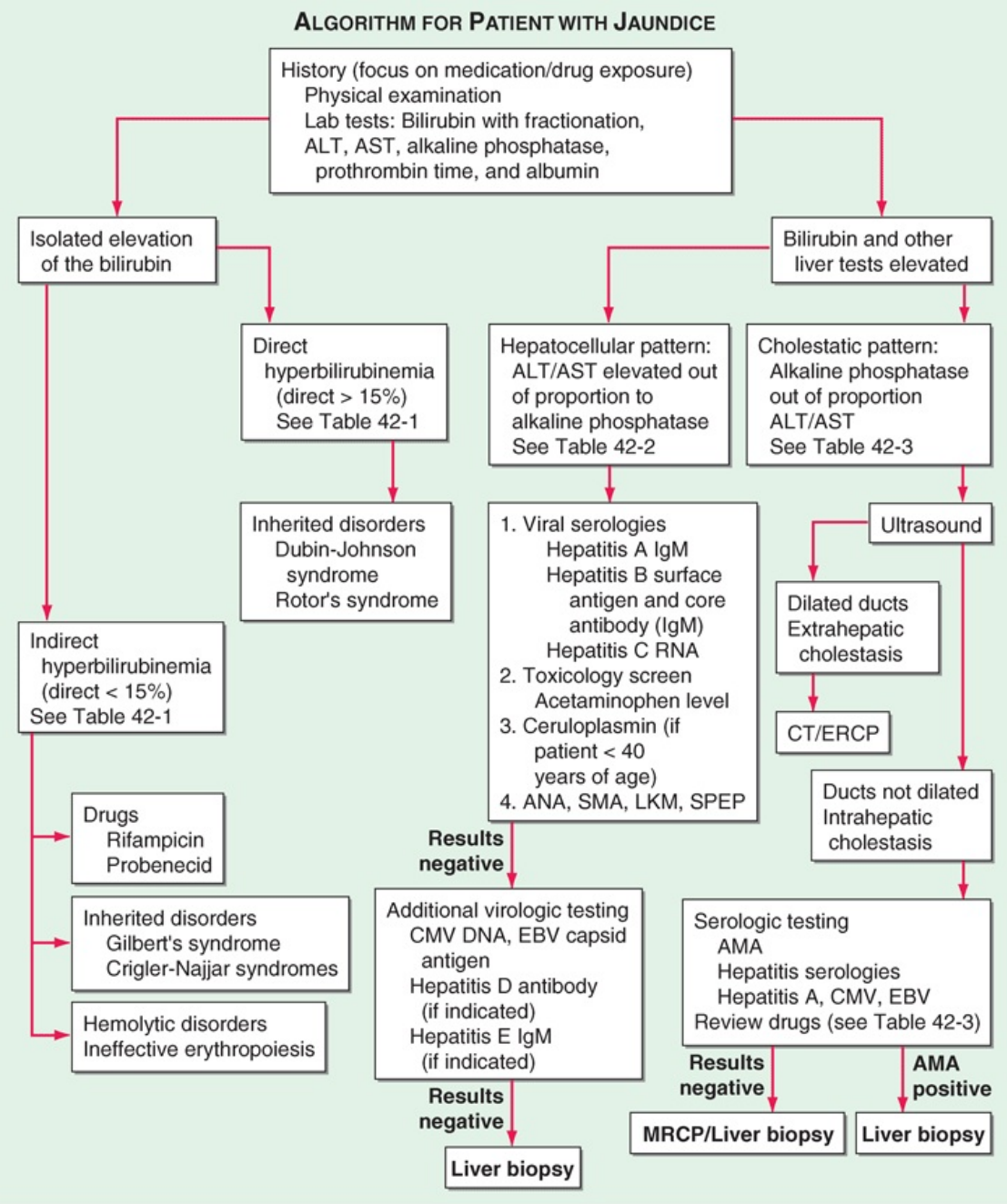

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evaluation of the patient with jaundice. ALT, alanine aminotransferase; AMA, antimitochondrial antibody; ANA, antinuclear antibody; AST, aspartate aminotransferase; CMV, cytomegalovirus; EBV, Epstein-Barr virus; LKM, liver-kidney microsomal antibody; MRCP, magnetic resonance cholangiopancreatography; SMA, smooth-muscle antibody; SPEP, serum protein electrophoresis.

ISOLATED ELEVATION OF SERUM BILIRUBIN

Unconjugated Hyperbilirubinemia

The differential diagnosis of an isolated unconjugated hyperbilirubinemia is limited (Table 42–1). The critical determination is whether the patient is suffering from a hemolytic process resulting in an overproduction of bilirubin (hemolytic disorders and ineffective erythropoiesis) or from impaired hepatic

uptake/conjugation of bilirubin (drug effect or genetic disorders).

Table 42–1 Causes of Isolated Hyperbilirubinemia

I. Indirect hyperbilirubinemia
A. Hemolytic disorders
1. Inherited
a. Spherocytosis, elliptocytosis Glucose-6-phosphate dehydrogenase and pyruvate kinase deficiencies
b. Sickle cell anemia
2. Acquired
a. Microangiopathic hemolytic anemias
b. Paroxysmal nocturnal hemoglobinuria
c. Spur cell anemia
d. Immune hemolysis
e. Parasitic infections
1. Malaria
2. Babesiosis
B. Ineffective erythropoiesis
1. Cobalamin, folate, thalassemia, and severe iron deficiencies
C. Drugs
1. Rifampicin, probenecid, ribavirin
D. Inherited conditions
1. Crigler-Najjar types I and II
2. Gilbert's syndrome
II. Direct hyperbilirubinemia
A. Inherited conditions
1. Dubin-Johnson syndrome
2. Rotor's syndrome

Hemolytic disorders that cause excessive heme production may be either inherited or acquired. Inherited disorders include spherocytosis, sickle cell anemia, thalassemia, and deficiency of red cell enzymes such as pyruvate kinase and glucose-6-phosphate dehydrogenase. In these conditions, the serum bilirubin rarely exceeds 86 μmol/L (5 mg/dL). Higher levels may occur when there is coexistent renal or hepatocellular dysfunction or in acute hemolysis such as a sickle cell crisis. In evaluating jaundice in patients with chronic hemolysis, it is important to remember the high incidence of pigmented (calcium bilirubinate) gallstones found in these patients, which increases the likelihood of choledocholithiasis as an alternative explanation for hyperbilirubinemia.

Acquired hemolytic disorders include microangiopathic hemolytic anemia (e.g., hemolytic-uremic syndrome), paroxysmal nocturnal hemoglobinuria, spur cell anemia, and immune hemolysis and parasitic infections including malaria and babesiosis. Ineffective erythropoiesis occurs in cobalamin, folate, and iron deficiencies.

In the absence of hemolysis, the physician should consider a problem with the hepatic uptake or conjugation of bilirubin. Certain drugs, including rifampicin and probenecid, may cause unconjugated hyperbilirubinemia by diminishing hepatic uptake of bilirubin. Impaired bilirubin conjugation occurs in three genetic conditions: Crigler-Najjar syndrome, types I and II, and Gilbert's syndrome. *Crigler-Najjar type I* is an exceptionally rare condition found in neonates and characterized by severe jaundice [bilirubin > 342 μmol/L (>20 mg/dL)] and neurologic impairment due to kernicterus, frequently leading to death in infancy or childhood. These patients have a complete absence of bilirubin UDPGT activity, usually due to mutations in the critical 3' domain of the *UDPGT* gene, and are totally unable to conjugate, hence cannot excrete, bilirubin. The only effective treatment is orthotopic liver transplantation. Use of gene therapy and allogeneic hepatocyte infusion are experimental approaches of future promise for this devastating disease.

Crigler-Najjar type II is somewhat more common than type I. Patients live into adulthood with serum bilirubin levels that range from 103–428 μmol/L (6–25 mg/dL). In these patients, mutations in the bilirubin *UDPGT* gene cause reduced but not completely absent activity of the enzyme. Bilirubin *UDPGT* activity can be induced by the administration of phenobarbital, which can reduce serum bilirubin levels in these patients. Despite marked jaundice, these patients usually survive into adulthood, although they may be susceptible to kernicterus under the stress of intercurrent illness or

surgery.

Gilbert's syndrome is also marked by the impaired conjugation of bilirubin due to reduced bilirubin UDPGT activity to approximately 1/3 of normal. Gilbert's syndrome is very common, with a reported incidence of 3–12%. Patients with Gilbert's syndrome have a mild unconjugated hyperbilirubinemia with serum levels almost always <103 μmol/L (6 mg/dL). The serum levels may fluctuate, and jaundice is often identified only during periods of fasting. One molecular defect that has been identified in patients with Gilbert's syndrome is in the TATAA element in the 5' promoter region of the bilirubin *UDPGT* gene upstream of exon 1. This defect alone is not necessarily sufficient for producing the clinical syndrome of Gilbert's as there are patients who are homozygous for this defect yet do not have the levels of hyperbilirubinemia typically seen in Gilbert's syndrome. An enhancer polymorphism that lowers transcriptional activity has been identified. The decrease in transcription caused by both mutations together may be critical for producing the syndrome. Unlike both Crigler-Najjar syndromes, Gilbert's syndrome is very common. The reported incidence is 3–7% of the population with males predominating over females by a ratio of 2–7:1.

Conjugated Hyperbilirubinemia

Elevated conjugated hyperbilirubinemia is found in two rare inherited conditions: *Dubin-Johnson syndrome* and *Rotor's syndrome* (Table 42–1). Patients with both conditions present with asymptomatic jaundice, typically in the second generation of life. The defect in Dubin-Johnson syndrome is mutations in the gene for multiple drug resistance protein 2. These patients have altered excretion of bilirubin into the bile ducts. Rotor's syndrome seems to be a problem with the hepatic storage of bilirubin. Differentiating between these syndromes is possible, but clinically unnecessary, due to their benign nature.

ELEVATION OF SERUM BILIRUBIN WITH OTHER LIVER TEST ABNORMALITIES

The remainder of this chapter will focus on the evaluation of the patient with a conjugated hyperbilirubinemia in the setting of other liver test abnormalities. This group of patients can be divided into those with a primary hepatocellular process and those with intra- or extrahepatic cholestasis. Being able to make this differentiation will guide the physician's evaluation (Fig. 42-1). This differentiation is made on the basis of the history and physical examination as well as the pattern of liver test abnormalities.

History

A complete medical history is perhaps the single most important part of the evaluation of the patient with unexplained jaundice. Important considerations include the use of or exposure to any chemical or medication, either physician-prescribed, over-the-counter, complementary or alternative medicines such as herbal and vitamin preparations, or other drugs such as anabolic steroids. The patient should be carefully questioned about possible parenteral exposures, including transfusions, intravenous and intranasal drug use, tattoos, and sexual activity. Other important questions include recent travel history; exposure to people with jaundice; exposure to possibly contaminated foods; occupational exposure to hepatotoxins; alcohol consumption; the duration of jaundice; and the presence of any accompanying symptoms such as arthralgias, myalgias, rash, anorexia, weight loss, abdominal pain, fever, pruritus, and changes in the urine and stool. While none of these latter symptoms are specific for any one condition, they can suggest a particular diagnosis. A history of arthralgias and myalgias predating jaundice suggests hepatitis, either viral or drug-related. Jaundice associated with the sudden onset of severe right upper quadrant pain and shaking chills suggests choledocholithiasis and ascending cholangitis.

Physical Examination

The general assessment should include assessment of the patient's nutritional status. Temporal and proximal muscle wasting suggests long-standing diseases such as pancreatic cancer or cirrhosis. Stigmata of chronic liver disease, including spider nevi, palmar erythema, gynecomastia, caput medusae, Dupuytren's contractures, parotid gland enlargement, and testicular atrophy are commonly seen in advanced alcoholic (Laennec's) cirrhosis and occasionally in other types of cirrhosis. An enlarged left supraclavicular node (Virchow's node) or periumbilical nodule (Sister Mary Joseph's nodule) suggests an abdominal malignancy. Jugular venous distention, a sign of right-sided heart failure, suggests hepatic congestion. Right pleural effusion, in the absence of clinically apparent ascites, may be seen in advanced cirrhosis.

The abdominal examination should focus on the size and consistency of the liver, whether the spleen is palpable and hence enlarged, and whether there is ascites present. Patients with cirrhosis may have an enlarged left lobe of the liver, which is felt below the xiphoid, and an enlarged spleen. A grossly enlarged nodular liver or an obvious abdominal mass suggests malignancy. An enlarged tender liver could be viral or alcoholic hepatitis; an infiltrative process such as amyloid; or, less often, an acutely congested liver secondary to right-sided heart failure. Severe right upper quadrant tenderness with respiratory arrest on inspiration (Murphy's sign) suggests cholecystitis or, occasionally, ascending cholangitis. Ascites in the presence of jaundice suggests either cirrhosis or malignancy with peritoneal spread.

Laboratory Tests

When the physician encounters a patient with unexplained jaundice, there is a battery of tests that are helpful in the initial evaluation. These include total and direct serum bilirubin with fractionation, aminotransferases, alkaline phosphatase, albumin, and prothrombin time tests. Enzyme tests [alanine aminotransferase (ALT), aspartate aminotransferase (AST), and alkaline phosphatase (ALP)] are helpful in differentiating between a hepatocellular process and a cholestatic process (Table 302-1; Fig. 42-1), a critical step in determining what additional workup is indicated. Patients with a hepatocellular process generally have a disproportionate rise in the aminotransferases compared to the ALP. Patients with a cholestatic process have a disproportionate rise in the ALP compared to the aminotransferases. The bilirubin can be prominently elevated in both hepatocellular and cholestatic conditions and, therefore, is not necessarily helpful in differentiating between the two.

In addition to the enzyme tests, all jaundiced patients should have additional blood tests, specifically an albumin level and a prothrombin time, to assess liver function. A low albumin level suggests a chronic process such as cirrhosis or cancer. A normal albumin level is suggestive of a more acute process such as viral hepatitis or choledocholithiasis. An elevated prothrombin time indicates either vitamin K deficiency due to prolonged jaundice and malabsorption of vitamin K or significant hepatocellular dysfunction. The failure of the prothrombin time to correct with parenteral administration of vitamin K indicates severe hepatocellular injury.

The results of the bilirubin, enzyme tests, albumin, and prothrombin time tests will usually indicate whether a jaundiced patient has a hepatocellular or a cholestatic disease, as well as some indication of the duration and severity of the disease. The causes and evaluation of hepatocellular and cholestatic diseases are quite different.

Hepatocellular Conditions

Hepatocellular diseases that can cause jaundice include viral hepatitis, drug or environmental toxicity, alcohol, and end-stage cirrhosis from any cause (Table 42–2). Wilson's disease, once believed to occur primarily in young adults, should be considered in all adults if no other cause of jaundice is found. Autoimmune hepatitis is typically seen in young to middle-aged women but may affect men and women of any age. Alcoholic hepatitis can be differentiated from viral and toxin-related hepatitis by the pattern of the aminotransferases. Patients with alcoholic hepatitis typically have an AST:ALT ratio of at least 2:1. The AST rarely exceeds 300 U/L. Patients with acute viral hepatitis and toxin-related injury severe enough to produce jaundice typically have aminotransferases > 500 U/L, with the ALT greater than or equal to the AST. The degree of aminotransferase elevation can occasionally help in differentiating between hepatocellular and cholestatic processes. While ALT and AST values less than 8 times normal may be seen in either hepatocellular or cholestatic liver disease, values 25 times normal or higher are seen primarily in acute hepatocellular diseases. Patients with jaundice from cirrhosis can have normal or only slight elevations of the aminotransferases.

Table 42–2 Hepatocellular Conditions that May Produce Jaundice

Viral hepatitis
Hepatitis A, B, C, D, and E
Epstein-Barr virus
Cytomegalovirus
Herpes simplex
Alcohol
Drug toxicity
Predictable, dose-dependent (e.g., acetaminophen)
Unpredictable, idiosyncratic (e.g., isoniazid)
Environmental toxins
Vinyl chloride
Jamaica bush tea—pyrrolizidine alkaloids
Kava Kava
Wild mushrooms—*Amanita phalloides* or *A. verna*
Wilson's disease
Autoimmune hepatitis

When the physician determines that the patient has a hepatocellular disease, appropriate testing for acute viral hepatitis includes a hepatitis A IgM antibody, a hepatitis B surface antigen and core IgM antibody, and a hepatitis C viral RNA test. It can take many weeks for the hepatitis C antibody to become detectable, making it an unreliable test if acute hepatitis C is suspected. Depending on

circumstances, studies for hepatitis D and E, Epstein-Barr virus (EBV), and cytomegalovirus (CMV) may be indicated. Ceruloplasmin is the initial screening test for Wilson's disease. Testing for autoimmune hepatitis usually includes an antinuclear antibody and measurement of specific immunoglobulins.

Drug-induced hepatocellular injury can be classified either as predictable or unpredictable. Predictable drug reactions are dose-dependent and affect all patients who ingest a toxic dose of the drug in question. The classic example is acetaminophen hepatotoxicity. Unpredictable or idiosyncratic drug reactions are not dose-dependent and occur in a minority of patients. A great number of drugs can cause idiosyncratic hepatic injury. Environmental toxins are also an important cause of hepatocellular injury. Examples include industrial chemicals such as vinyl chloride, herbal preparations containing pyrrolizidine alkaloids (Jamaica bush tea) and Kava Kava, and the mushrooms *Amanita phalloides* or *A. verna* that contain highly hepatotoxic amatoxins.

Cholestatic Conditions

When the pattern of the liver tests suggests a cholestatic disorder, the next step is to determine whether it is intra- or extrahepatic cholestasis (Fig. 42-1). Distinguishing intrahepatic from extrahepatic cholestasis may be difficult. History, physical examination, and laboratory tests are often not helpful. The next appropriate test is an ultrasound. The ultrasound is inexpensive, does not expose the patient to ionizing radiation, and can detect dilation of the intra- and extrahepatic biliary tree with a high degree of sensitivity and specificity. The absence of biliary dilatation suggests intrahepatic cholestasis, while the presence of biliary dilatation indicates extrahepatic cholestasis. False-negative results occur in patients with partial obstruction of the common bile duct or in patients with cirrhosis or primary sclerosing cholangitis (PSC) where scarring prevents the intrahepatic ducts from dilating.

Although ultrasonography may indicate extrahepatic cholestasis, it rarely identifies the site or cause of obstruction. The distal common bile duct is a particularly difficult area to visualize by ultrasound because of overlying bowel gas. Appropriate next tests include CT, magnetic resonance cholangiography (MRCP), and endoscopic retrograde cholangiopancreatography (ERCP). CT scanning and MRCP are better than ultrasonography for assessing the head of the pancreas and for identifying choledocholithiasis in the distal common bile duct, particularly when the ducts are not dilated. ERCP is the "gold standard" for identifying choledocholithiasis. It is performed by introducing a side-viewing endoscope perorally into the duodenum. The ampulla of Vater is visualized, and a catheter is advanced through the ampulla. Injection of dye allows for the visualization of the common bile duct and the pancreatic duct. Beyond its diagnostic capabilities, ERCP allows for therapeutic interventions, including the removal of common bile duct stones and the placement of stents. In patients in whom ERCP is unsuccessful and there is a high likelihood of the need for a therapeutic intervention, transhepatic cholangiography can provide the same information and allow for intervention. MRCP has replaced ERCP as the initial diagnostic test in cases where the need for intervention is felt to be small.

In patients with apparent *intrahepatic cholestasis*, the diagnosis is often made by serologic testing in combination with percutaneous liver biopsy. The list of possible causes of intrahepatic cholestasis is long and varied (Table 42–3). A number of conditions that typically cause a hepatocellular pattern of injury can also present as a cholestatic variant. Both hepatitis B and C can cause a cholestatic hepatitis (fibrosing cholestatic hepatitis). This disease variant has been reported in patients who have undergone solid organ transplantation. Hepatitis A, alcoholic hepatitis, EBV, and CMV may also present as cholestatic liver disease.

Table 42–3 Cholestatic Conditions that May Produce Jaundice

I. Intrahepatic
A. Viral hepatitis
1. Fibrosing cholestatic hepatitis—hepatitis B and C
2. Hepatitis A, Epstein-Barr virus, cytomegalovirus
B. Alcoholic hepatitis
C. Drug toxicity
1. Pure cholestasis—anabolic and contraceptive steroids
2. Cholestatic hepatitis—chlorpromazine, erythromycin estolate
3. Chronic cholestasis—chlorpromazine and prochlorperazine
D. Primary biliary cirrhosis
E. Primary sclerosing cholangitis
F. Vanishing bile duct syndrome
1. Chronic rejection of liver transplants
2. Sarcoidosis

3. Drugs
G. Inherited
1. Progressive familial intrahepatic cholestasis
2. Benign recurrent cholestasis
H. Cholestasis of pregnancy
I. Total parenteral nutrition
J. Nonhepatobiliary sepsis
K. Benign postoperative cholestasis
L. Paraneoplastic syndrome
M. Venoocclusive disease
N. Graft-versus-host disease
O. Infiltrative disease
1. TB
2. Lymphoma
3. Amyloid
P. Infections
1. Malaria
2. Leptospirosis
II. Extrahepatic
A. Malignant
1. Cholangiocarcinoma
2. Pancreatic cancer
3. Gallbladder cancer
4. Ampullary cancer
5. Malignant involvement of the porta hepatis lymph nodes
B. Benign
1. Choledocholithiasis
2. Postoperative biliary structures
3. Primary sclerosing cholangitis
4. Chronic pancreatitis
5. AIDS cholangiopathy
6. Mirizzi's syndrome
7. Parasitic disease (ascariasis)

Drugs may cause intrahepatic cholestasis, a variant of drug-induced hepatitis. Drug-induced cholestasis is usually reversible after eliminating the offending drug, although it may take many months for cholestasis to resolve. Drugs most commonly associated with cholestasis are the anabolic and contraceptive steroids. Cholestatic hepatitis has been reported with chlorpromazine, imipramine, tolbutamide, sulindac, cimetidine, and erythromycin estolate. It also occurs in patients taking trimethoprim; sulfamethoxazole; and penicillin-based antibiotics such as ampicillin, dicloxacillin, and clavulinic acid. Rarely, cholestasis may be chronic and associated with progressive fibrosis despite early discontinuation of the drug. Chronic cholestasis has been associated with chlorpromazine and prochlorperazine.

Primary biliary cirrhosis is an autoimmune disease predominantly of middle-aged women in which there is a progressive destruction of interlobular bile ducts. The diagnosis is made by the presence of the antimitochondrial antibody that is found in 95% of patients. *Primary sclerosing cholangitis* is characterized by the destruction and fibrosis of larger bile ducts. The disease may involve only the intrahepatic ducts and present as intrahepatic cholestasis. However, in 95% of patients with PSC, both intra- and extrahepatic ducts are involved. The diagnosis of PSC is made by imaging the biliary tree. The pathognomonic findings are multiple strictures of bile ducts with dilatations proximal to the strictures. Approximately 75% of patients with PSC have inflammatory bowel disease.

The *vanishing bile duct syndrome* and *adult bile ductopenia* are rare conditions in which there are a decreased number of bile ducts seen in liver biopsy specimens. The histologic picture is similar to that found in primary biliary cirrhosis. This picture is seen in patients who develop chronic rejection after liver transplantation and in those who develop graft-versus-host disease after bone marrow transplantation. Vanishing bile duct syndrome also occurs in rare cases of sarcoidosis, in patients taking certain drugs including chlorpromazine, and idiopathically.

There are also familial forms of intrahepatic cholestasis. The familial intrahepatic cholestatic syndromes include *progressive familial intrahepatic cholestasis* (PFIC) *types 1–3,* and *benign recurrent cholestasis* (BRC). PFIC1 and BRC are autosomal recessive diseases that result from mutations in the *ATP8B1* gene that encodes a protein belonging to the subfamily of P-type ATPases; the exact function of this protein remains poorly defined. While PFIC1 is a progressive condition that manifests in childhood, BRC presents later than PFIC1 and is marked by recurrent episodes of jaundice and pruritus; the episodes are self-limited but can be debilitating. PFIC2 is caused by mutations in the *ABCB11* gene, which encodes the bile salt export pump, and PFIC3 is caused by mutations in the multidrug-resistant P-glycoprotein 3. *Cholestasis of pregnancy* occurs in the second and third trimesters and resolves after delivery. Its cause is unknown, but the condition is probably inherited and cholestasis can be triggered by estrogen administration.

Other causes of intrahepatic cholestasis include total parenteral nutrition (TPN); nonhepatobiliary sepsis; benign postoperative cholestasis; and a paraneoplastic syndrome associated with a number of different malignancies, including Hodgkin's disease, medullary thyroid cancer, renal cell cancer, renal sarcoma, T cell lymphoma, prostate cancer, and several gastrointestinal malignancies. The term *Stauffer's syndrome* has been used for intrahepatic cholestasis specifically associated with renal cell cancer. In patients developing cholestasis in the intensive care unit, the major considerations should be sepsis, shock liver, and TPN jaundice. Jaundice occurring after bone marrow transplantation is most likely due to venoocclusive disease or graft-versus-host disease. Jaundice with associated liver dysfunction can be seen in severe cases of Plasmodium falciparum. The jaundice in these cases is a combination of indirect hyperbilirubinemia from hemolysis and both cholestatic and hepatocellular jaundice. Poor outcomes are seen in these cases when the jaundice is accompanied by encephalopathy and renal failure. Weil's disease, a severe presentation of leptospirosis, is marked by jaundice with renal failure, fever, headache, and muscle pain.

Causes of *extrahepatic cholestasis* can be split into malignant and benign (Table 42–3). Malignant causes include pancreatic, gallbladder, ampullary, and cholangiocarcinoma. The latter is most commonly associated with PSC and is exceptionally difficult to diagnose because its appearance is often identical to that of PSC. Pancreatic and gallbladder tumors, as well as cholangiocarcinoma, are rarely resectable and have poor prognoses. Ampullary carcinoma has the highest surgical cure rate of all the tumors that present as painless jaundice. Hilar lymphadenopathy due to metastases from other cancers may cause obstruction of the extrahepatic biliary tree.

Choledocholithiasis is the most common cause of extrahepatic cholestasis. The clinical presentation can range from mild right upper quadrant discomfort with only minimal elevations of the enzyme tests to ascending cholangitis with jaundice, sepsis, and circulatory collapse. PSC may occur with clinically important strictures limited to the extrahepatic biliary tree. In cases where there is a dominant stricture, patients can be effectively managed with serial endoscopic dilatations. Chronic pancreatitis rarely causes strictures of the distal common bile duct, where it passes through the head of the pancreas. AIDS cholangiopathy is a condition, usually due to infection of the bile duct epithelium with CMV or cryptosporidia, which has a cholangiographic appearance similar to that of PSC. These patients usually present with greatly elevated serum alkaline phosphatase levels (mean, 800 IU/L), but the bilirubin is often near normal. These patients do not typically present with jaundice.

Summary

The goal of this chapter is not to provide an encyclopedic review of all of the conditions that can cause jaundice. Rather, it is intended to provide a framework that helps a physician to evaluate the patient with jaundice in a logical way (Fig. 42-1).

Simply stated, the initial step is to obtain appropriate blood tests to determine if the patient has an isolated elevation of serum bilirubin. If so, is the bilirubin elevation due to an increased unconjugated or conjugated fraction? If the hyperbilirubinemia is accompanied by other liver test abnormalities, is the disorder hepatocellular or cholestatic? If cholestatic, is it intra- or extrahepatic? All of these questions can be answered with a thoughtful history, physical examination, and interpretation of laboratory and radiologic tests and procedures.

FURTHER READINGS

Bosma PJ: Inherited disorders of bilirubin metabolism. J Hepatol 38:107, 2003[PMID: 12480568]

Ferenci P: Wilson's disease. Clin Gastroenterol Hepatol 3:726, 2005[PMID: 16233999]

Glasova H, Beuers U: Extrahepatic manifestations of cholestasis. J Gastroenterol Hepatol 9:938, 2002

Pratt DS, Kaplan MM: Laboratory tests, in *Schiff's Diseases of the Liver* , 10th ed, ER Schiff et al (eds). Philadelphia, Lippincott Williams & Wilkins, 2006

Trauner M et al: Molecular pathogenesis of cholestasis. N Engl J Med 339:1217, 1998[PMID: 9780343]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ABDOMINAL SWELLING

Abdominal swelling is a manifestation of numerous diseases. Patients may complain of bloating or abdominal fullness and may note increasing abdominal girth on the basis of increased clothing or belt size. Abdominal discomfort is often reported, but pain is less frequent. When abdominal pain does accompany swelling, it is frequently the result of an intraabdominal infection, peritonitis, or pancreatitis. Patients with abdominal distention from ascites (fluid in the abdomen) may report the new onset of an inguinal or umbilical hernia. Dyspnea may result from pressure against the diaphragm and the inability to expand the lungs fully.

The causes of abdominal swelling can be remembered conveniently by the *six Fs*: flatus, fat, fluid, fetus, feces, or a "fatal growth" (often a neoplasm).

Flatus

Abdominal swelling may be the result of increased intestinal gas. The normal small intestine contains approximately 200 mL of gas made up of nitrogen, oxygen, carbon dioxide, hydrogen, and methane. Nitrogen and oxygen are consumed (swallowed), whereas carbon dioxide, hydrogen, and methane are produced intraluminally by bacterial fermentation. Increased intestinal gas can occur in a number of conditions. Aerophagia, the swallowing of air, can result in increased amounts of oxygen and nitrogen in the small intestine and lead to abdominal swelling. Aerophagia typically results from gulping food; chewing gum; smoking; or as a response to anxiety, which leads to repetitive belching. In some cases, increased intestinal gas is the result of bacterial metabolism of excess fermentable substances such as lactose and other oligosaccharides that can lead to production of hydrogen, carbon dioxide, or methane. In many cases, the precise cause of abdominal distention cannot be determined. In some persons, particularly those with irritable bowel syndrome and bloating, the subjective sense of abdominal pressure is attributable to impaired intestinal transit of gas rather than increased gas volume. Abdominal distention, an objective increase in girth, is the result of a lack of coordination between diaphragmatic contraction and anterior abdominal wall relaxation in response to an increase in intraabdominal volume loads. Occasionally, increased lumbar lordosis accounts for apparent abdominal distention.

Fat

Weight gain with an increase in abdominal fat can result in an increase in abdominal girth and can be perceived as abdominal swelling. Abdominal fat may be the result of an imbalance between calorie intake and energy expenditure associated with a poor diet and sedentary lifestyle and also can be a manifestation of certain diseases such as Cushing's syndrome. Excess abdominal fat has been associated with an increased risk of insulin resistance and cardiovascular disease.

Fluid

Fluid within the abdominal cavity, or ascites, often results in abdominal distention and is discussed in detail below.

Fetus

Pregnancy results in increased abdominal girth. Typically, an increase in abdominal size is first noted at 12 to 14 weeks of gestation, when the uterus moves from the pelvis into the abdomen. Abdominal distention may be seen before this point as a result of fluid retention and relaxation of the abdominal muscles.

Feces

Increased stool in the colon, in the setting of severe constipation or intestinal obstruction, also leads to increased abdominal girth. These conditions often are accompanied by abdominal pain, nausea, and vomiting and can be diagnosed by imaging studies.

Fatal Growth

An abdominal mass can result in abdominal swelling. Enlargement of the intraabdominal organs, specifically the liver (hepatomegaly) or spleen (splenomegaly) or an abdominal aortic aneurysm, can result in abdominal distention. Bladder distention also may result in abdominal swelling. In addition, malignancies, abscesses, or cysts can grow to sizes that lead to increased abdominal girth.

History and Physical Examination

Determining the etiology of abdominal swelling begins with history-taking and a physical examination. Patients should be questioned regarding symptoms suggestive of malignancy, including weight loss, night sweats, and anorexia. Inability to pass stool or flatus together with nausea or vomiting suggest bowel obstruction, severe constipation, or an ileus (lack of peristalsis). Increased eructation and flatus may point toward aerophagia or increased intestinal production of gas. Patients should be questioned about risk factors for or symptoms of chronic liver disease, including excessive alcohol use and jaundice, which suggest ascites. Patients should also be asked about other symptoms of medical conditions, including heart failure and tuberculosis, which may cause ascites.

Physical examination should assess for signs of systemic disease. The presence of lymphadenopathy, especially supraclavicular lymphadenopathy (Virchow's node), suggests metastatic abdominal malignancy. Care also should be taken during the cardiac examination to evaluate for elevation of jugular venous pressure (JVP); Kussmaul's sign (elevation of the JVP during inspiration); or a pericardial knock, which may be seen in heart failure or constrictive pericarditis, as well as a murmur of tricuspid regurgitation. Spider angiomas, palmar erythema, dilated superficial veins around the umbilicus (caput medusae), and gynecomastia suggest chronic liver disease.

The abdominal examination should begin with inspection for the presence of uneven distention or an obvious mass. Auscultation should follow. The absence of bowel sounds or the presence of high-pitched localized bowel sounds point toward an ileus or intestinal obstruction. An umbilical venous hum may suggest the presence of portal hypertension, and a harsh bruit over the liver is heard rarely in patients with hepatocellular carcinoma or alcoholic hepatitis. Abdominal swelling caused by intestinal gas can be differentiated from swelling caused by fluid or a solid mass by percussion; an abdomen filled with gas is tympanic, whereas an abdomen containing a mass or fluid is dull to percussion. The absence of abdominal dullness, however, does not exclude ascites, because a minimum of 1500 mL of ascites is required for detection on physical examination. Finally, the abdomen should be palpated to assess for tenderness, a mass, enlargement of the spleen or liver, or presence of a nodular liver suggesting cirrhosis or tumor. Light palpation of the liver may detect pulsations suggesting retrograde vascular flow from the heart in patients with right-sided heart failure, particularly tricuspid regurgitation.

Imaging and Laboratory Evaluation

Abdominal x-rays can be used to detect dilated loops of bowel, suggesting intestinal obstruction or ileus. An abdominal ultrasound can detect as little as 100 mL of ascites, hepatosplenomegaly, a nodular liver, or a mass. Ultrasound is often inadequate to detect retroperitoneal lymphadenopathy or a pancreatic lesion because of overlying bowel gas. If malignancy or pancreatic disease is suspected, CT can be performed. CT may also detect changes associated with advanced cirrhosis and portal hypertension (**Fig. 43-1**).

Figure 43-1

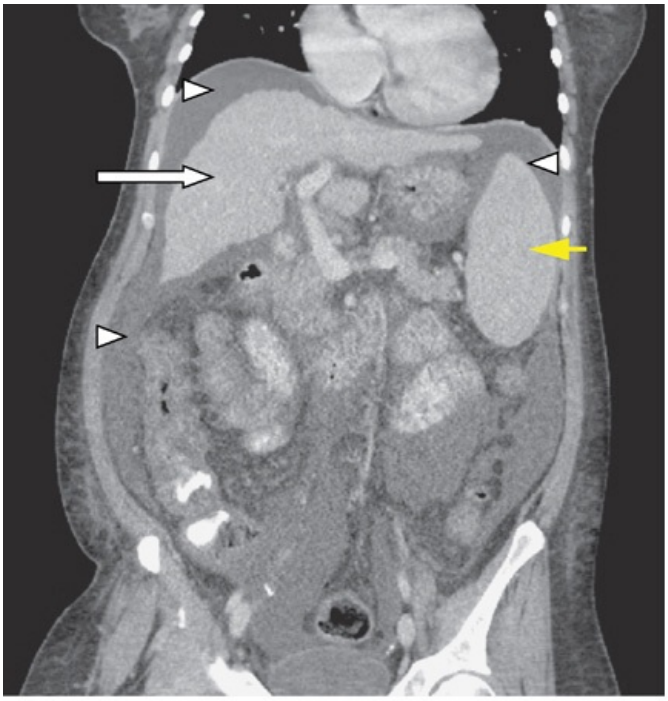

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT of a patient with a cirrhotic, nodular liver (white arrow), splenomegaly (yellow arrow), and

ascites (arrowheads).

Laboratory evaluation should include liver biochemical testing, a serum albumin level, a prothrombin time (international normalized ratio) to assess hepatic function, and a complete blood count to evaluate for the presence of cytopenias that may result from portal hypertension or leukocytosis, anemia, and thrombocytosis that may result from systemic infection. Serum amylase and lipase levels should be checked to evaluate the patient for acute pancreatitis. Urinary protein quantitation is indicated when nephrotic syndrome, which may cause ascites, is suspected.

In selected cases, measurement of the hepatic venous pressure gradient (pressure across the liver between the portal and hepatic veins) can be obtained via cannulation of the hepatic vein to confirm that ascites is caused by cirrhosis (see Chapter 308). In some cases, a liver biopsy may be necessary to confirm cirrhosis.

ASCITES

Pathogenesis in Cirrhosis

Ascites in patients with cirrhosis is the result of portal hypertension and renal salt and water retention. Portal hypertension signifies elevation of the pressure within the portal vein. According to Ohm's law, pressure is the product of resistance and flow. Increased hepatic resistance occurs by several mechanisms. First, the development of hepatic fibrosis, which defines cirrhosis, disrupts the normal architecture of the hepatic sinusoids and impedes normal blood flow through the liver. Second, activation of hepatic stellate cells, which mediate fibrogenesis, leads to smooth muscle contraction and fibrosis. Finally, cirrhosis is associated with a decrease in endothelial nitric oxide synthetase (eNOS) production, which results in decreased nitric oxide production and increased intrahepatic vasoconstriction.

The development of cirrhosis is also associated with increased systemic circulating levels of nitric oxide (contrary to the decrease seen intrahepatically) as well as increased levels of vascular endothelial growth factor and tumor necrosis factor that result in splanchnic arterial vasodilatation. Vasodilatation of the splanchnic circulation results in pooling of blood and a decrease in the effective circulating volume, which is perceived by the kidneys as hypovolemia. Compensatory vasoconstriction via release of antidiuretic hormone ensues, thereby leading to free water retention and activation of the sympathetic nervous system and renin angiotensin aldosterone system, leading in turn to renal sodium and water retention.

Pathogenesis in the Absence of Cirrhosis

Ascites in the absence of cirrhosis generally results from peritoneal carcinomatosis, peritoneal infection, or pancreatic disease. Peritoneal carcinomatosis can result from primary peritoneal malignancies such as mesothelioma or sarcoma, abdominal malignancies such as gastric or colonic adenocarcinoma, or metastatic disease from breast or lung carcinoma or melanoma (**Fig. 43-2**). The tumor cells lining the peritoneum produce a protein-rich fluid that contributes to the development of ascites. Fluid from the extracellular space is drawn into the peritoneum, further contributing to the development of ascites. Tuberculous peritonitis causes ascites via a similar mechanism; tubercles deposited on the peritoneum exude a proteinaceous fluid. Pancreatic ascites results from leakage of pancreatic enzymes into the peritoneum.

Figure 43-2

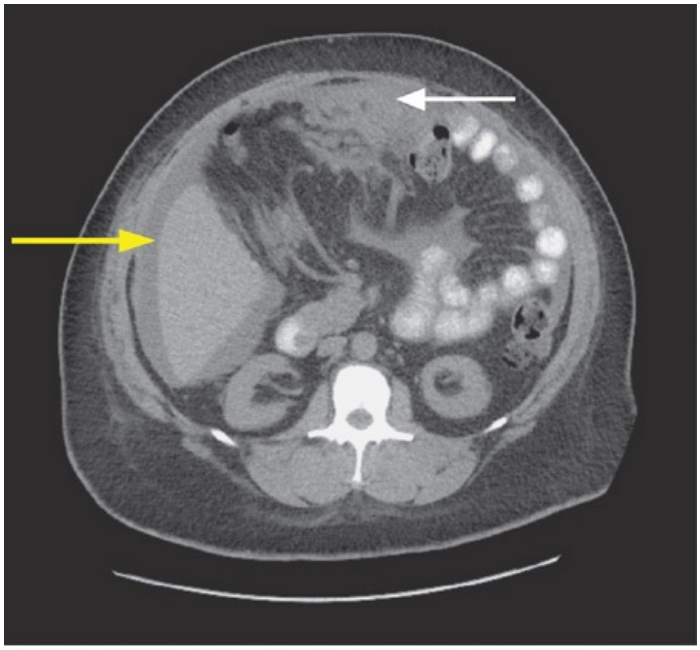

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT of a patient with peritoneal carcinomatosis (white arrow) and ascites (yellow arrow).

Causes

Cirrhosis accounts for 84% of cases of ascites. Cardiac ascites, peritoneal carcinomatosis, and "mixed" ascites resulting from cirrhosis and a second disease account for 10 to 15% of cases. Less common causes of ascites include massive hepatic metastasis, infection (tuberculosis, Chlamydia), pancreatitis, and renal disease (nephrotic syndrome). Rare causes of ascites include hypothyroidism and familial Mediterranean fever.

Evaluation

Once the presence of ascites has been confirmed, the etiology of the ascites is best determined by *paracentesis*. Paracentesis is a bedside procedure in which a needle or small catheter is passed transcutaneously to extract ascitic fluid from the peritoneum. The lower quadrants are the most frequent sites for paracentesis. Occasionally, an infraumbilical approach is used. The left lower quadrant is preferred because of the greater depth of ascites and thinner abdominal wall. Paracentesis is a safe procedure even in patients with coagulopathy; complications, including abdominal wall hematomas, hypotension, hepatorenal syndrome, and infection, are infrequent.

Once ascitic fluid has been extracted, its gross appearance should be examined. Turbid fluid can result from infection or tumor cells within the fluid. White, milky fluid indicates the presence of triglycerides in levels >200 mg/dL (and often >1000 mg/dL), which is the hallmark of *chylous ascites*. Chylous ascites results from lymphatic disruption that may occur with trauma, cirrhosis, tumor, tuberculosis, or certain congenital abnormalities. Dark brown fluid can reflect a high bilirubin concentration and indicates biliary tract perforation. Black fluid may indicate the presence of pancreatic necrosis or metastatic melanoma.

The ascitic fluid should be sent for measurement of the albumin and total protein levels, cell and differential counts, and, if infection is suspected, Gram's stain and culture, with inoculation of the fluid into blood culture bottles at the patient's bedside to maximize the yield. In addition, a serum albumin level should be sent simultaneously to permit calculation of the *serum-ascites albumin gradient* (*SAAG*).

The SAAG is useful for distinguishing ascites caused by portal hypertension from nonportal hypertensive ascites (**Fig. 43-3**). The SAAG reflects the pressure within the hepatic sinusoids and correlates with the hepatic venous pressure gradient. The SAAG is calculated by subtracting the ascitic albumin from the serum albumin and does not change with diuresis. A SAAG ≥1.1 g/dL reflects the presence of portal hypertension and indicates that the ascites is from an increased pressure in the hepatic sinusoids. According to Starling's law, a high SAAG reflects the oncotic pressure that counterbalances the portal pressure. Possible causes include cirrhosis, cardiac ascites, sinusoidal obstruction syndrome (venoocclusive disease), massive liver metastasis, or hepatic vein thrombosis (Budd-Chiari syndrome). A SAAG <1.1 g/dL indicates that the ascites is not related to portal hypertension as in tuberculous peritonitis, peritoneal carcinomatosis, or pancreatic ascites.

Figure 43-3

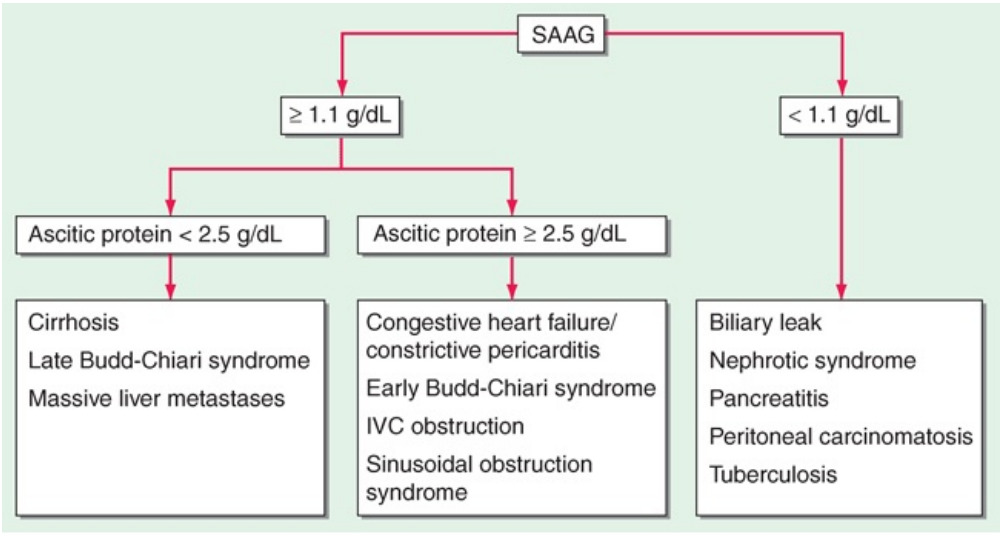

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the diagnosis of ascites according to the serum-ascites albumin gradient (SAAG). IVC, inferior vena cava.

For high-SAAG (≥ 1.1) ascites, the ascitic protein level can provide further clues to the etiology **(see Fig. 43-3)**. An ascitic protein level of ≥2.5 g/dL indicates that the hepatic sinusoids are normal and allows passage of protein into the ascites, as occurs in cardiac ascites, sinusoidal obstruction syndrome, or early Budd-Chiari syndrome. An ascitic protein level <2.5 g/dL indicates that the hepatic sinusoids have been damaged and scarred and no longer allow passage of protein, as occurs with cirrhosis, late Budd-Chiari syndrome, or massive liver metastases. Pro-brain-type natriuretic peptide (BNP) is a natriuretic hormone released by the heart as a result of increased volume and ventricular wall stretch. High levels of BNP in serum occur in heart failure and may be useful in identifying congestive heart failure as the cause of high-SAAG ascites.

Further tests are indicated only in specific clinical circumstances. When secondary peritonitis resulting from a perforated hollow viscus is suspected, ascitic glucose and lactate dehydrogenase (LDH) levels can be sent. In contrast to "spontaneous" bacterial peritonitis (SBP), which may complicate cirrhotic ascites, secondary peritonitis is suggested by an ascitic glucose level <50 mg/dL, an ascitic LDH greater than the serum LDH level, and multiple pathogens on ascitic fluid culture. When pancreatic ascites is suspected, an ascitic amylase should be measured and is typically >1000 mg/dL. Cytology can be useful in the diagnosis of peritoneal carcinomatosis. At least 50 mL of fluid should be obtained and sent for immediate processing. Tuberculous peritonitis can be difficult to diagnose by paracentesis. A smear for acid-fast bacilli has a sensitivity of only 0 to 3%, and a culture increases the sensitivity for diagnosis to 35 to 50%. In patients without cirrhosis, an elevated ascitic adenosine deaminase level has a sensitivity of more than 90% when a cut-off value of 30 to 45 U/L is used. When the cause of ascites remains uncertain, laparotomy or laparoscopy with peritoneal biopsies for histology and culture remains the gold standard.

Treatment: Ascites

The initial treatment of cirrhotic ascites is restriction of sodium intake to 2 g/d. When sodium restriction alone is inadequate to control ascites, oral diuretics, typically the combination of spironolactone and furosemide, are used. Spironolactone is an aldosterone antagonist that inhibits Na^+ resorption in the distal convoluted tubule of the kidney. Use of spironolactone may be limited by hyponatremia, hyperkalemia, and painful gynecomastia. If the gynecomastia is distressing, amiloride, 5–40 mg/d, may be substituted for spironolactone. Furosemide is a loop diuretic that is generally combined with spironolactone in a ratio of 40:100; maximal daily doses of spironolactone and furosemide are generally 400 mg and 160 mg, respectively.

Refractory cirrhotic ascites is defined by the persistence of ascites despite sodium restriction and maximal (or maximally tolerated) diuretic use. Refractory ascites can be managed by serial large volume paracentesis (LVP) or a transjugular intrahepatic peritoneal shunt (TIPS), a radiologically placed portosystemic shunt to decompress the hepatic sinusoids. TIPS is superior to LVP in reducing the reaccumulation of ascites but is associated with an increased frequency of hepatic encephalopathy with no difference in mortality rates.

Malignantascites does not respond to sodium restriction or diuretics. Patients must undergo serial LVPs, transcutaneous drainage catheter placement, or, rarely, creation of a peritovenous shunt (a

shunt from the abdominal cavity to the vena cava).

Ascites caused by tuberculous peritonitis is treated with standard antituberculosis therapy. Noncirrhotic ascites of other causes is treated by correction of the precipitating condition.

Complications

Spontaneous bacterial peritonitis (SBP) is a common and potentially lethal complication of cirrhotic ascites. SBP also can occasionally complicate ascites caused by nephrotic syndrome, heart failure, acute hepatitis, and acute liver failure but is rare in malignant ascites. Patients with SBP generally note an increase in abdominal girth; however, abdominal tenderness is found in only 40% of patients, and rebound tenderness is rare. Patients may present with fever, nausea, vomiting, or the new onset of or exacerbation of preexisting hepatic encephalopathy.

SBP is defined by a polymorphonuclear neutrophil (PMN) count of ≥250/mm^3 in the ascitic fluid. Ascitic fluid cultures typically reveal one bacterial pathogen. The presence of multiple pathogens in the setting of an elevated ascitic PMN count suggests secondary peritonitis from a ruptured viscus or abscess. The presence of multiple pathogens without an elevated PMN count suggests bowel perforation from the paracentesis needle. SBP is generally the result of enteric bacteria that have translocated across an edematous bowel wall. The most common pathogens are Gram-negative rods, including *Escherichia coli* and *Klebsiella*, as well as streptococci and enterococci.

Treatment of SBP with an antibiotic such as intravenous cefotaxime is effective against gram-negative and Gram-positive aerobes. Five days of treatment are sufficient if the patient improves clinically.

Cirrhotic patients with a history of SBP, an ascitic fluid total protein concentration <1 g/dL, or active gastrointestinal bleeding should receive prophylactic antibiotics to prevent SBP; oral norfloxacin is commonly used. Diuresis increases the activity of ascitic fluid protein opsonins and may decrease the risk of SBP.

Hepatic hydrothorax occurs when ascites, often caused by cirrhosis, migrates via fenestrae in the diaphragm into the pleural space and can result in shortness of breath, hypoxia, and infection. Treatment is similar to that of cirrhotic ascites and includes sodium restriction, diuretics, and, if needed, thoracentesis or TIPS placement. Chest tube placement should be avoided.

FURTHER READINGS

Hoefs JC: Serum protein concentration and portal pressure determine the ascitic fluid protein concentration in patients with chronic liver disease. J Lab Clin Med 102:260, 1983[PMID: 6864073]

Runyon BA: Cardiac ascites: A characterization. J Clin Gastroenterol 10:410, 1988[PMID: 3418089]

———: Management of adult patients with ascites due to cirrhosis: An update. Hepatology 49:2087, 2009

———, et al: Diuresis increases ascitic fluid opsonic activity in patients who survive spontaneous bacterial peritonitis. J Hepatol 14:249, 1992

———, et al: Ascitic fluid analysis in malignancy-related ascites. Hepatology 8:1104, 1988

Sheer TA, et al: Usefulness of serum N-terminal-ProBNP in distinguishing ascites due to cirrhosis from ascites due to heart failure. J Clin Gastroenterol 44:e23, 2010

Simrén M: Bloating and abdominal distention: Not so poorly understood anymore! Gastroenterology 136:1487, 2009[PMID: 21596195]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

AZOTEMIA AND URINARY ABNORMALITIES: INTRODUCTION

Normal kidney functions occur through numerous cellular processes to maintain body homeostasis. Disturbances in any of those functions can lead to a constellation of abnormalities that may be detrimental to survival. The clinical manifestations of those disorders depend on the pathophysiology of the renal injury and often are identified initially as a complex of symptoms, abnormal physical findings, and laboratory changes that together make possible the identification of specific syndromes. These renal syndromes **(Table 44-1)** may arise as a consequence of a systemic illness or can occur as a primary renal disease. Nephrologic syndromes usually consist of several elements that reflect the underlying pathologic processes. The duration and severity of the disease affect those findings and typically include one or more of the following: (1) reduction in glomerular filtration rate (GFR) (azotemia), (2) abnormalities of urine sediment [red blood cells (RBC), white blood cells, casts, and crystals], (3) abnormal excretion of serum proteins (proteinuria), (4) disturbances in urine volume (oliguria, anuria, polyuria), (5) presence of hypertension and/or expanded total body fluid volume (edema), (6) electrolyte abnormalities, (7) in some syndromes, fever/pain. The combination of these findings should permit identification of one of the major nephrologic syndromes (Table 44-1) and will allow differential diagnoses to be narrowed and the appropriate diagnostic evaluation and therapeutic course to be determined. All these syndromes and their associated diseases are discussed in more detail in subsequent chapters. This chapter focuses on several aspects of renal abnormalities that are critically important for distinguishing among those processes: (1) reduction in GFR leading to azotemia, (2) alterations of the urinary sediment and/or protein excretion, and (3) abnormalities of urinary volume.

Table 44-1 Initial Clinical and Laboratory Data Base for Defining Major Syndromes in Nephrology

Syndromes	Important Clues to Diagnosis	Findings That Are Common	Location of Discussion of Disease-Causing Syndrome
Acute or rapidly progressive renal failure	Anuria Oliguria Documented recent decline in GFR	Hypertension, hematuria Proteinuria, pyuria Casts, edema	Chaps. 279, 283, 285, 289
Acute nephritis	Hematuria, RBC casts Azotemia, oliguria Edema, hypertension	Proteinuria Pyuria Circulatory congestion	Chap. 283
Chronic renal failure	Azotemia for >3 months Prolonged symptoms or signs of uremia Symptoms or signs of renal osteodystrophy Kidneys reduced in size bilaterally Broad casts in urinary sediment	Proteinuria Casts Polyuria, nocturia Edema, hypertension Electrolyte disorders	Chaps. 278, 280
Nephrotic syndrome	Proteinuria >3.5 g per 1.73 m2 per 24 h Hypoalbuminemia Edema Hyperlipidemia	Casts Lipiduria	Chap. 283
Asymptomatic urinary	Hematuria		Chap. 283

abnormalities	Proteinuria (below nephrotic range) Sterile pyuria, casts		
Urinary tract infection/pyelonephritis	Bacteriuria >10^5 colonies per milliliter Other infectious agent documented in urine Pyuria, leukocyte casts Frequency, urgency Bladder tenderness, flank tenderness	Hematuria Mild azotemia Mild proteinuria Fever	Chap. 288
Renal tubule defects	Electrolyte disorders Polyuria, nocturia Renal calcification Large kidneys Renal transport defects	Hematuria "Tubular" proteinuria (<1 g/24 h) Enuresis	Chaps. 284, 285
Hypertension	Systolic/diastolic hypertension	Proteinuria Casts Azotemia	Chaps. 247, 286
Nephrolithiasis	Previous history of stone passage or removal Previous history of stone seen by x-ray Renal colic	Hematuria Pyuria Frequency, urgency	Chap. 287
Urinary tract obstruction	Azotemia, oliguria, anuria Polyuria, nocturia, urinary retention Slowing of urinary stream Large prostate, large kidneys Flank tenderness, full bladder after voiding	Hematuria Pyuria Enuresis, dysuria	Chap. 289

Abbreviations: GFR; glomerular filtration rate; RBC, red blood cell.

AZOTEMIA

Assessment of Glomerular Filtration Rate (GFR)

Monitoring the GFR is important in both the hospital and outpatient settings, and several different methodologies are available. GFR is the primary metric for kidney "function," and its direct measurement involves administration of a radioactive isotope (such as inulin or iothalamate) that is filtered at the glomerulus but neither reabsorbed nor secreted throughout the tubule. Clearance of inulin or iothalamate in milliliters per minute equals the GFR and is calculated from the rate of removal from the blood and appearance in the urine over several hours. Direct GFR measurements are frequently available through nuclear radiology departments. In most clinical circumstances direct measurement of GFR is not available, and the serum creatinine level is used as a surrogate to estimate GFR. Serum creatinine is the most widely used marker for GFR, and the GFR is related directly to the urine creatinine excretion and inversely to the serum creatinine (U_{Cr}/P_{Cr}). Based on this relationship and some important caveats (discussed below), the GFR will fall in roughly inverse proportion to the rise in P_{Cr}. Failure to account for GFR reductions in drug dosing can lead to significant morbidity and mortality from drug toxicities (e.g., digoxin, aminoglycosides). In the outpatient setting, the serum creatinine serves as an estimate for GFR (although much less accurate; see below). In patients with chronic progressive renal disease, there is an approximately linear relationship between $1/P_{Cr}$ (*y* axis) and time (*x* axis). The slope of that line will remain constant for an individual patient, and when values are obtained that do not fall on the line, an investigation for a superimposed

acute process (e.g., volume depletion, drug reaction) should be initiated. Signs and symptoms of uremia develop at significantly different levels of serum creatinine, depending on the patient (size, age, and sex), the underlying renal disease, the existence of concurrent diseases, and true GFR. In general, patients do not develop symptomatic uremia until renal insufficiency is quite severe (GFR <15 mL/min).

A significantly reduced GFR (either acute or chronic) usually is reflected in a rise in serum creatinine and leads to retention of nitrogenous waste products (azotemia) such as urea. Azotemia may result from reduced renal perfusion, intrinsic renal disease, or postrenal processes (ureteral obstruction; see below and **Fig. 44-1**). Precise determination of GFR is problematic as both commonly measured indices (urea and creatinine) have characteristics that affect their accuracy as markers of clearance. Urea clearance may underestimate GFR significantly because of urea reabsorption by the tubule. In contrast, creatinine is derived from muscle metabolism of creatine, and its generation varies little from day to day.

Figure 44-1

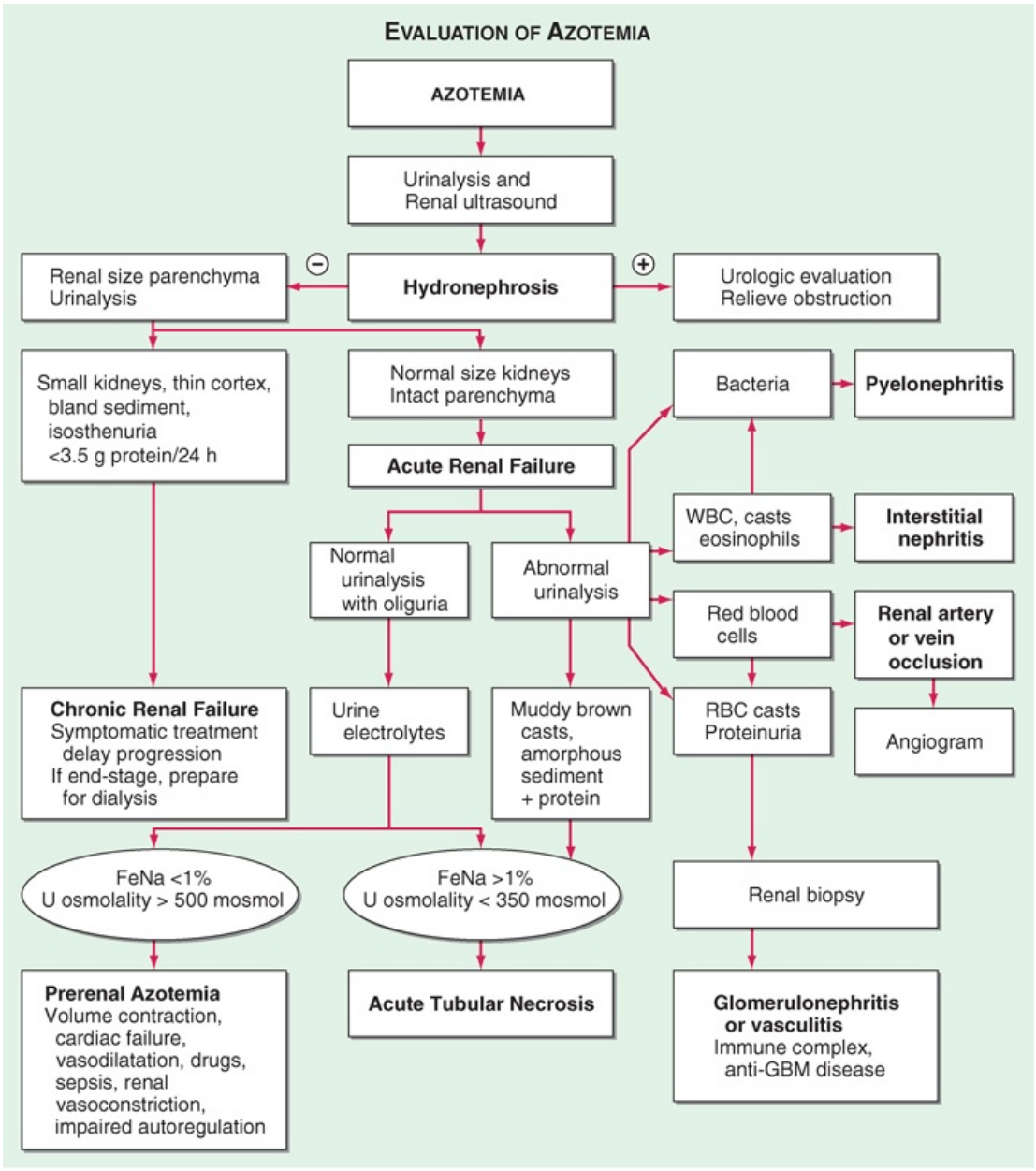

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the patient with azotemia. FeNa, fractional excretion of sodium; GBM, glomerular basement membrane; RBC, red blood cell; WBC, white blood cell.

Creatinine clearance, an approximation of GFR, is measured from plasma and urinary creatinine excretion rates for a defined time period (usually 24 h) and is expressed in milliliters per minute: CrCl =

$(U_{vol} \times U_{Cr})/(P_{Cr} \times T_{min})$. Creatinine is useful for estimating GFR because it is a small, freely filtered solute that is not reabsorbed by the tubules. Serum creatinine levels can increase acutely from dietary ingestion of cooked meat, however, and creatinine can be secreted into the proximal tubule through an organic cation pathway (especially in advanced progressive chronic kidney disease), leading to overestimation of GFR. When a timed collection for creatinine clearance is not available, decisions about drug dosing must be based on serum creatinine alone. Two formulas are used widely to estimate kidney function from serum creatinine: (1) Cockcroft-Gault and (2) four-variable MDRD (Modification of Diet in Renal Disease).

Cockcroft-Gault: CrCl (mL/min) = (140 – age (years) x weight (kg) x [0.85 if female])/(72 x sCr (mg/dL)

MDRD: eGFR (mL/min per 1.73 m^2) = 186.3 x P_{Cr} ($e^{-1.154}$) x age ($e^{-0.203}$) x (0.742 if female) x (1.21 if black).

Numerous websites are available for making these calculations (*www.kidney.org/professionals/kdoqi/gfr_calculator.cfm*). A newer CKD-EPI eGFR was developed by pooling several cohorts with and without kidney disease who had data on directly measured GFR and appears to be more accurate:

CKD-EPI: eGFR = 141 x min (Scr/k, 1)^a x max (Scr/k, 1)$^{-1.209}$ x 0.993^{Age} x 1.018 [if female] x 1.159 [if black]

where Scr is serum creatinine, k is 0.7 for females and 0.9 for males, a is –0.329 for females and –0.411 for males, min indicates the minimum of Scr/k or 1, and max indicates the maximum of Scr/k or 1 (*http://www.qxmd.com/renal/Calculate-CKD-EPI-GFR.php*).

Several limitations of using serum creatinine–based estimating equations must be acknowledged. Each equation, along with the 24-h urine collection for measurement of creatinine clearance, is based on the assumption that the patient is in steady state, without daily increases or decreases in serum creatinine levels as a result of rapidly changing GFR. The MDRD equation has poorer accuracy when GFR >60 mL/min per 1.73 m^2. The gradual loss of muscle from chronic illness, chronic use of glucocorticoids, or malnutrition can mask significant changes in GFR with small or imperceptible changes in serum creatinine concentration. Cystatin C is a member of the cystatin superfamily of cysteineprotease inhibitors and is produced at a relatively constant rate from all nucleated cells. Serum cystatin C has been proposed to be a more sensitive marker of early GFR decline than is plasma creatinine; however, like serum creatinine, cystatin C is influenced by age, race, and sex and additionally is associated with diabetes, smoking, and markers of inflammation.

Approach to the Patient: Azotemia

Once it has been established that GFR is reduced, the physician must decide if this represents acute or chronic renal injury. The clinical situation, history, and laboratory data often make this an easy distinction. However, the laboratory abnormalities characteristic of chronic renal failure, including anemia, hypocalcemia, and hyperphosphatemia, often are also present in patients presenting with acute renal failure. Radiographic evidence of renal osteodystrophy (Chap. 280) can be seen only in chronic renal failure but is a very late finding, and these patients are usually on dialysis. The urinalysis and renal ultrasound occasionally can facilitate distinguishing acute from chronic renal failure. An approach to the evaluation of azotemic patients is shown in Fig. 44-1. Patients with advanced chronic renal insufficiency often have some proteinuria, nonconcentrated urine (isosthenuria; isoosmotic with plasma), and small kidneys on ultrasound, characterized by increased echogenicity and cortical thinning. Treatment should be directed toward slowing the progression of renal disease and providing symptomatic relief for edema, acidosis, anemia, and hyperphosphatemia, as discussed in Chap. 280. Acute renal failure (Chap. 279) can result from processes that affect renal blood flow (prerenal azotemia), intrinsic renal diseases (affecting small vessels, glomeruli, or tubules), or postrenal processes (obstruction to urine flow in ureters, bladder, or urethra) (Chap. 289).

PRERENAL FAILURE

Decreased renal perfusion accounts for 40–80% of acute renal failure and, if appropriately treated, is readily reversible. The etiologies of prerenal azotemia include any cause of decreased circulating blood volume (gastrointestinal hemorrhage, burns, diarrhea, diuretics), volume sequestration (pancreatitis, peritonitis, rhabdomyolysis), or decreased effective arterial volume (cardiogenic shock, sepsis). Renal perfusion also can be affected by reductions in cardiac output from peripheral vasodilation (sepsis, drugs) or profound renal vasoconstriction [severe heart failure, hepatorenal syndrome, drugs such as nonsteroidal anti-inflammatory drugs (NSAIDs)]. True or "effective" arterial hypovolemia leads to a fall in mean arterial pressure, which in turn triggers a series of neural and humoral responses that include activation of the sympathetic nervous and renin-angiotensin-aldosterone systems and antidiuretic hormone (ADH) release. GFR is maintained by prostaglandin-mediated relaxation of afferent arterioles and angiotensin II–mediated constriction of efferent arterioles. Once the mean arterial pressure falls below 80 mmHg, there is a steep decline in GFR.

Blockade of prostaglandin production by NSAIDs can result in severe vasoconstriction and acute renal

failure. Blocking angiotensin action with angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) decreases efferent arteriolar tone and in turn decreases glomerular capillary perfusion pressure. Patients on NSAIDs and/or ACE inhibitors/ARBs are most susceptible to hemodynamically mediated acute renal failure when blood volume is reduced for any reason. Patients with bilateral renal artery stenosis (or stenosis in a solitary kidney) are dependent on efferent arteriolar vasoconstriction for maintenance of glomerular filtration pressure and are particularly susceptible to a precipitous decline in GFR when given ACE inhibitors or ARBs.

Prolonged renal hypoperfusion may lead to acute tubular necrosis (ATN), an intrinsic renal disease that is discussed below. The urinalysis and urinary electrolytes can be useful in distinguishing prerenal azotemia from ATN **(Table 44-2)**. The urine of patients with prerenal azotemia can be predicted from the stimulatory actions of norepinephrine, angiotensin II, ADH, and low tubule fluid flow rate on salt and water reabsorption. In prerenal conditions, the tubules are intact, leading to a concentrated urine (>500 mosmol), avid Na retention (urine Na concentration <20 m*M*/L, fractional excretion of Na <1%), and U_{Cr}/P_{Cr} >40 (Table 44-2). The prerenal urine sediment is usually normal or has occasional hyaline and granular casts, whereas the sediment of ATN usually is filled with cellular debris and dark (muddy brown) granular casts.

Table 44-2 Laboratory Findings in Acute Renal Failure

Index	Prerenal Azotemia	Oliguric Acute Renal Failure
BUN/P_{Cr} ratio	>20:1	10-15:1
Urine sodium (U_{Na}), meq/L	<20	>40
Urine osmolality, mosmol/L H_2O	>500	<350
Fractional excretion of sodium $FE_{Na} = \frac{U_{Na} \times P_{Cr} \times 100}{P_{Na} \times U_{Cr}}$	<1%	>2%
Urine/plasma creatinine (U_{Cr}/P_{Cr})	>40	<20

Abbreviations: BUN, blood urea nitrogen; P_{Cr}, plasma creatinine; P_{Na}, plasma sodium concentration; U_{Cr}, urine creatinine concentration; U_{Na}, urine sodium concentration.

POSTRENAL AZOTEMIA

Urinary tract obstruction accounts for <5% of cases of acute renal failure, but it is usually reversible and must be ruled out early in the evaluation (Fig. 44-1). Since a single kidney is capable of adequate clearance, obstructive acute renal failure requires obstruction at the urethra or bladder outlet, bilateral ureteral obstruction, or unilateral obstruction in a patient with a single functioning kidney. Obstruction usually is diagnosed by the presence of ureteral and renal pelvic dilation on renal ultrasound. However, early in the course of obstruction or if the ureters are unable to dilate (e.g., encasement by pelvic tumors or periureteral), the ultrasound examination may be negative. **The specific urologic conditions that cause obstruction are discussed in Chap. 289.**

INTRINSIC RENAL DISEASE

When prerenal and postrenal azotemia have been excluded as etiologies of renal failure, an intrinsic parenchymal renal disease is present. Intrinsic renal disease can arise from processes involving large renal vessels, intrarenal microvasculature and glomeruli, or the tubulointerstitium. Ischemic and toxic ATN account for ~90% of cases of acute intrinsic renal failure. As outlined in Fig. 44-1, the clinical setting and urinalysis are helpful in separating the possible etiologies of acute intrinsic renal failure. Prerenal azotemia and ATN are part of a spectrum of renal hypoperfusion; evidence of structural tubule injury is present in ATN, whereas prompt reversibility occurs with prerenal azotemia upon restoration of adequate renal perfusion. Thus, ATN often can be distinguished from prerenal azotemia by urinalysis and urine electrolyte composition (Table 44-2 and Fig. 44-1). Ischemic ATN is observed most frequently in patients who have undergone major surgery, trauma, severe hypovolemia, overwhelming sepsis, or extensive burns. Nephrotoxic ATN complicates the administration of many common medications, usually by inducing a combination of intrarenal vasoconstriction, direct tubule toxicity, and/or tubule obstruction. The kidney is vulnerable to toxic injury by virtue of its rich blood supply (25% of cardiac output) and its ability to concentrate and metabolize toxins. A diligent search for hypotension and nephrotoxins usually will uncover the specific etiology of ATN. Discontinuation of nephrotoxins and stabilization of blood pressure often will suffice without the need for dialysis while the tubules recover. **An extensive list of potential drugs and toxins implicated in ATN can be found**

in Chap. 279.

Processes that involve the tubules and interstitium can lead to acute kidney injury (AKI), a subtype of acute renal failure. These processes include drug-induced interstitial nephritis (especially antibiotics, NSAIDs, and diuretics), severe infections (both bacterial and viral), systemic diseases (e.g., systemic lupus erythematosus), and infiltrative disorders (e.g., sarcoid, lymphoma, or leukemia). A list of drugs associated with allergic interstitial nephritis can be found in Chap. 285. The urinalysis usually shows mild to moderate proteinuria, hematuria, and pyuria (~75% of cases) and occasionally shows white blood cell casts. The finding of RBC casts in interstitial nephritis has been reported but should prompt a search for glomerular diseases (Fig. 44-1). Occasionally, renal biopsy will be needed to distinguish among these possibilities. The finding of eosinophils in the urine is suggestive of allergic interstitial nephritis or atheroembolic renal disease and is optimally observed by using a Hansel stain. The absence of eosinophiluria, however, does not exclude these etiologies.

Occlusion of large renal vessels including arteries and veins is an uncommon cause of acute renal failure. A significant reduction in GFR by this mechanism suggests bilateral processes or a unilateral process in a patient with a single functioning kidney. Renal arteries can be occluded with atheroemboli, thromboemboli, in situ thrombosis, aortic dissection, or vasculitis. Atheroembolic renal failure can occur spontaneously but most often is associated with recent aortic instrumentation. The emboli are cholesterol-rich and lodge in medium and small renal arteries, leading to an eosinophil-rich inflammatory reaction. Patients with atheroembolic acute renal failure often have a normal urinalysis, but the urine may contain eosinophils and casts. The diagnosis can be confirmed by renal biopsy, but this is often unnecessary when other stigmata of atheroemboli are present (livedo reticularis, distal peripheral infarcts, eosinophilia). Renal artery thrombosis may lead to mild proteinuria and hematuria, whereas renal vein thrombosis typically induces heavy proteinuria and hematuria. **These vascular complications often require angiography for confirmation and are discussed in Chap. 286.**

Diseases of the glomeruli (glomerulonephritis and vasculitis) and the renal microvasculature (hemolytic-uremic syndromes, thrombotic thrombocytopenic purpura, and malignanthypertension) usually present with various combinations of glomerular injury: proteinuria, hematuria, reduced GFR, and alterations of sodium excretion that lead to hypertension, edema, and circulatory congestion (acute nephritic syndrome). These findings may occur as primary renal diseases or as renal manifestations of systemic diseases. The clinical setting and other laboratory data help distinguish primary renal diseases from systemic diseases. The finding of RBC casts in the urine is an indication for early renal biopsy (Fig. 44-1) as the pathologic pattern has important implications for diagnosis, prognosis, and treatment. Hematuria without RBC casts also can be an indication of glomerular disease; this evaluation is summarized in **Fig. 44-2**. **A detailed discussion of glomerulonephritis and diseases of the microvasculature can be found in Chap. 285.**

Figure 44-2

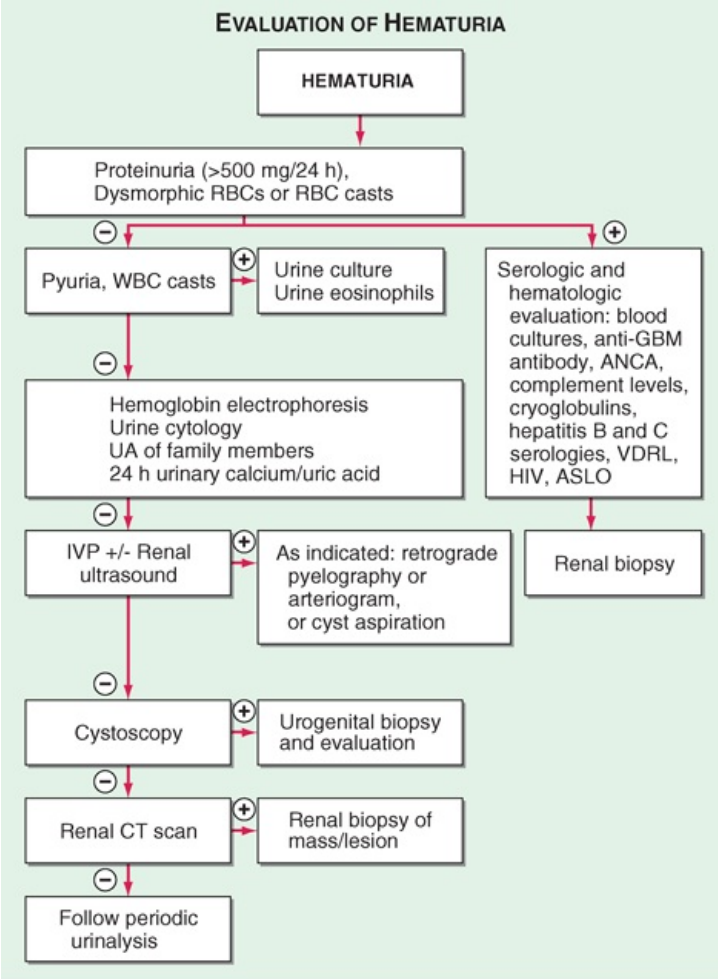

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the patient with hematuria. ANCA, antineutrophil cytoplasmic antibody; ASLO, antistreptolysin O; CT, computed tomography; GBM, glomerular basement membrane; IVP, intravenous pyelography; RBC, red blood cell; UA, urinalysis; VDRL, Venereal Disease Research Laboratory; WBC, white blood cell.

OLIGURIA AND ANURIA

Oliguria refers to a 24-h urine output <400 mL, and *anuria* is the complete absence of urine formation (<100 mL). Anuria can be caused by total urinary tract obstruction, total renal artery or vein occlusion, and shock (manifested by severe hypotension and intense renal vasoconstriction). Cortical necrosis, ATN, and rapidly progressive glomerulonephritis occasionally cause anuria. Oliguria can accompany any cause of acute renal failure and carries a more serious prognosis for renal recovery in all conditions except prerenal azotemia. *Nonoliguria* refers to urine output >400 mL/d in patients with acute or chronic azotemia. With nonoliguric ATN, disturbances of potassium and hydrogen balance are less severe than in oliguric patients, and recovery to normal renal function is usually more rapid.

ABNORMALITIES OF THE URINE

Proteinuria

The evaluation of proteinuria is shown schematically in **Fig. 44-3** and typically is initiated after detection of proteinuria by dipstick examination. The dipstick measurement detects only albumin and gives false-positive results when pH >7.0 and the urine is very concentrated or contaminated with blood. Because the dipstick relies on urinary albumin concentration, a very dilute urine may obscure significant proteinuria on dipstick examination. Quantification of urinary albumin on a spot urine sample (ideally from a first morning void) by measuring an albumin-to-creatinine ratio (ACR) is helpful in approximating a 24-h albumin excretion rate (AER) where ACR (mg/g) ≈ AER (mg/24 h). Furthermore, proteinuria that is not predominantly albumin will be missed by dipstick screening. This is particularly important for the detection of Bence Jones proteins in the urine of patients with multiple myeloma. Tests to measure total urine protein concentration accurately rely on precipitation with sulfosalicylic or trichloracetic acid (Fig. 44-3).

Figure 44-3

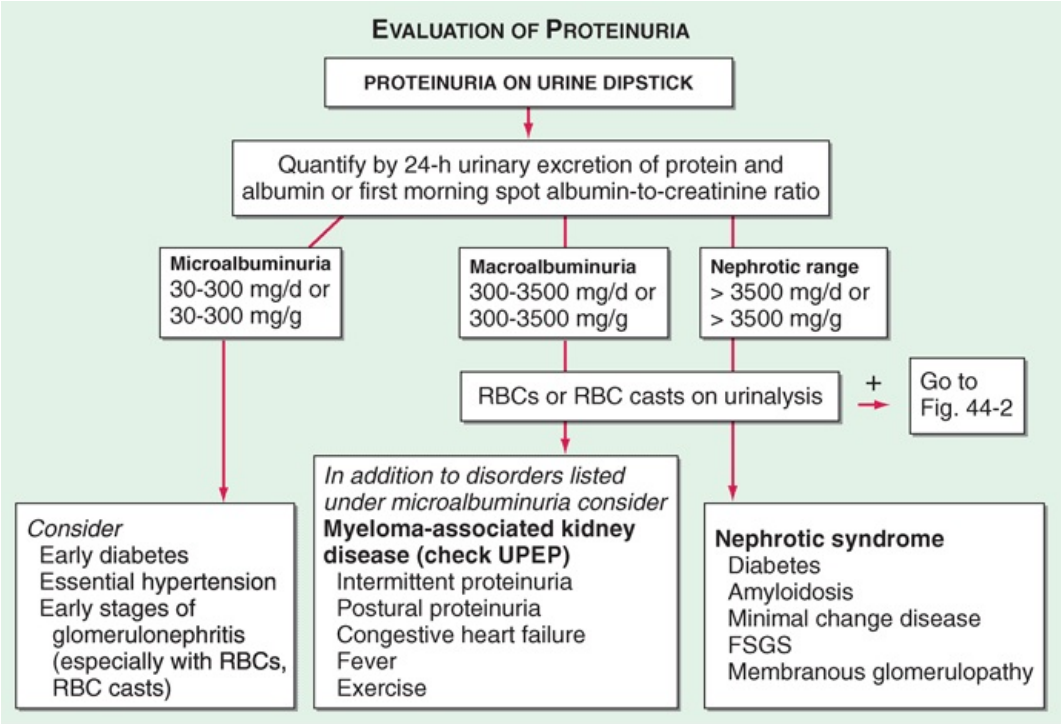

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the patient with proteinuria. Investigation of proteinuria is often initiated by a positive dipstick on routine urinalysis. Conventional dipsticks detect predominantly albumin and provide a semiquantitative assessment (trace, 1+, 2+, or 3+), which is influenced by urinary concentration as reflected by urine specific gravity (minimum <1.005, maximum 1.030). However, more exact determination of proteinuria should employ a spot morning protein/creatinine ratio (mg/g) or a 24-h urine collection (mg/24 h). FSGS, focal segmental glomerulosclerosis; MPGN, membranoproliferative glomerulonephritis; RBC, red blood cell.

The magnitude of proteinuria and the protein composition of the urine depend on the mechanism of renal injury that leads to protein losses. Both charge and size selectivity normally prevent virtually all plasma albumin, globulins, and other high-molecular-weight proteins from crossing the glomerular wall; however, if this barrier is disrupted, plasma proteins may leak into the urine (glomerular proteinuria; Fig. 44-3). Smaller proteins (<20 kDa) are freely filtered but are readily reabsorbed by the proximal tubule. Traditionally, healthy individuals excrete <150 mg/d of total protein and <30 mg/d of albumin. However, even at albuminuria levels <30 mg/d, risk for progression to overt nephropathy or subsequent cardiovascular disease is increased. The remainder of the protein in the urine is secreted by the tubules (Tamm-Horsfall, IgA, and urokinase) or represents small amounts of filtered β_2-microglobulin, apoproteins, enzymes, and peptide hormones. Another mechanism of proteinuria occurs when there is excessive production of an abnormal protein that exceeds the capacity of the tubule for reabsorption. This most commonly occurs with plasma cell dyscrasias, such as multiple myeloma, amyloidosis, and lymphomas that are associated with monoclonal production of immunoglobulin light chains.

The normal glomerular endothelial cell forms a barrier composed of pores of ~100 nm that retain blood cells but offer little impediment to passage of most proteins. The glomerular basement membrane traps most large proteins (>100 kDa), and the foot processes of epithelial cells (podocytes) cover the urinary side of the glomerular basement membrane and produce a series of narrow channels (slit diaphragms) to allow molecular passage of small solutes and water but not proteins. Some glomerular diseases, such as minimal change disease, cause fusion of glomerular epithelial cell foot processes, resulting in predominantly "selective" (Fig. 44-3) loss of albumin. Other glomerular diseases can present with disruption of the basement membrane and slit diaphragms (e.g., by immune complex deposition), resulting in losses of albumin and other plasma proteins. The fusion of foot processes causes increased pressure across the capillary basement membrane, resulting in areas with larger pore sizes. The combination of increased pressure and larger pores results in significant proteinuria ("nonselective"; Fig. 44-3).

When the total daily excretion of protein is >3.5 g, hypoalbuminemia, hyperlipidemia, and edema (nephrotic syndrome; Fig. 44-3) are often present as well. However, total daily urinary protein excretion >3.5 g can occur without the other features of the nephrotic syndrome in a variety of other renal diseases (Fig. 44-3). Plasma cell dyscrasias (multiple myeloma) can be associated with large amounts

of excreted light chains in the urine, which may not be detected by dipstick. The light chains produced from these disorders are filtered by the glomerulus and overwhelm the reabsorptive capacity of the proximal tubule.. Renal failure from these disorders occurs through a variety of mechanisms, including tubule obstruction (cast nephropathy) and light chain deposition.

Hypoalbuminemia in nephrotic syndrome occurs through excessive urinary losses and increased proximal tubule catabolism of filtered albumin. Edema forms from renal sodium retention and reduced plasma oncotic pressure, which favors fluid movement from capillaries to interstitium. To compensate for the perceived decrease in effective intravascular volume, activation of the renin-angiotensin system, stimulation of ADH, and activation of the sympathetic nervous system occur that promote continued renal salt and water reabsorption and progressive edema. The urinary loss of regulatory proteins and changes in hepatic synthesis contribute to the other manifestations of the nephrotic syndrome. A hypercoagulable state may arise from urinary losses of antithrombin III, reduced serum levels of proteins S and C, hyperfibrinogenemia, and enhanced platelet aggregation. Hypercholesterolemia may be severe and results from increased hepatic lipoprotein synthesis. Loss of immunoglobulins contributes to an increased risk of infection. Many diseases (some listed in Fig. 44-3) and drugs can cause the nephrotic syndrome; a complete list can be found in Chap. 283.

Hematuria, Pyuria, and Casts

Isolated hematuria without proteinuria, other cells, or casts is often indicative of bleeding from the urinary tract. Hematuria is defined as two to five RBCs per high-power field (HPF) and can be detected by dipstick. A false-positive dipstick for hematuria (where no RBCs are seen on urine microscopy) may occur when myoglobinuria is present, often in the setting of rhabdomyolysis. Common causes of isolated hematuria include stones, neoplasms, tuberculosis, trauma, and prostatitis. Gross hematuria with blood clots is usually not an intrinsic renal process; rather, it suggests a postrenal source in the urinary collecting system. Evaluation of patients presenting with microscopic hematuria is outlined in Fig. 44-2. A single urinalysis with hematuria is common and can result from menstruation, viral illness, allergy, exercise, or mild trauma. Persistent or significant hematuria (>3 RBCs/HPF on three urinalyses, a single urinalysis with >100 RBCs, or gross hematuria) is associated with significant renal or urologic lesions in 9.1% of cases. Even patients who are chronically anticoagulated should be investigated as outlined in Fig. 44-2. The suspicion for urogenital neoplasms in patients with isolated painless hematuria and nondysmorphic RBCs increases with age. Neoplasms are rare in the pediatric population, and isolated hematuria is more likely to be "idiopathic" or associated with a congenital anomaly. Hematuria with pyuria and bacteriuria is typical of infection and should be treated with antibiotics after appropriate cultures. Acute cystitis or urethritis in women can cause gross hematuria. Hypercalciuria and hyperuricosuria are also risk factors for unexplained isolated hematuria in both children and adults. In some of these patients (50–60%), reducing calcium and uric acid excretion through dietary interventions can eliminate the microscopic hematuria.

Isolated microscopic hematuria can be a manifestation of glomerular diseases. The RBCs of glomerular origin are often dysmorphic when examined by phase-contrast microscopy. Irregular shapes of RBCs may also result from pH and osmolarity changes produced along the distal nephron. Observer variability in detecting dysmorphic RBCs is common. The most common etiologies of isolated glomerular hematuria are IgA nephropathy, hereditary nephritis, and thin basement membrane disease. IgA nephropathy and hereditary nephritis can lead to episodic gross hematuria. A family history of renal failure is often present in patients with hereditary nephritis, and patients with thin basement membrane disease often have other family members with microscopic hematuria. A renal biopsy is needed for the definitive diagnosis of these disorders, which are discussed in more detail in Chap. 283. Hematuria with dysmorphic RBCs, RBC casts, and protein excretion >500 mg/d is virtually diagnostic of glomerulonephritis. RBC casts form as RBCs that enter the tubule fluid become trapped in a cylindrical mold of gelled Tamm-Horsfall protein. Even in the absence of azotemia, these patients should undergo serologic evaluation and renal biopsy as outlined in Fig. 44-2.

Isolated pyuria is unusual since inflammatory reactions in the kidney or collecting system also are associated with hematuria. The presence of bacteria suggests infection, and white blood cell casts with bacteria are indicative of pyelonephritis. White blood cells and/or white blood cell casts also may be seen in acute glomerulonephritis as well as in tubulointerstitial processes such as interstitial nephritis and transplant rejection. In chronic renal diseases, degenerated cellular casts called *waxy casts* can be seen in the urine. *Broad casts* are thought to arise in the dilated tubules of enlarged nephrons that have undergone compensatory hypertrophy in response to reduced renal mass (i.e., chronic renal failure). A mixture of broad casts typically seen with chronic renal failure together with cellular casts and RBCs may be seen in smoldering processes such as chronic glomerulonephritis.

ABNORMALITIES OF URINE VOLUME

The volume of urine produced varies with the fluid intake, renal function, and physiologic demands of the individual. See "Azotemia," above, for discussion of decreased (oliguria) or absent urine production (anuria). **The physiology of water formation and renal water conservation are discussed in Chap. 278.**

Polyuria

By history, it is often difficult for patients to distinguish urinary frequency (often of small volumes) from true polyuria (>3 L/d), and a quantification of volume by 24-h urine collection may be needed **(Fig. 44-4)**. Polyuria results from two potential mechanisms: (1) excretion of nonabsorbable solutes (such as glucose) or (2) excretion of water (usually from a defect in ADH production or renal responsiveness). To distinguish a solute diuresis from a water diuresis and to determine if the diuresis is appropriate for the clinical circumstances, a urine osmolality is measured. The average person excretes between 600 and 800 mosmol of solutes per day, primarily as urea and electrolytes. If the urine output is >3 L/d and the urine is dilute (<250 mosmol/L), total mosmol excretion is normal and a water diuresis is present. This circumstance could arise from polydipsia, inadequate secretion of vasopressin (central diabetes insipidus), or failure of renal tubules to respond to vasopressin (nephrogenic diabetes insipidus). If the urine volume is >3 L/d and urine osmolality is >300 mosmol/L, a solute diuresis is clearly present and a search for the responsible solute(s) is mandatory.

Figure 44-4

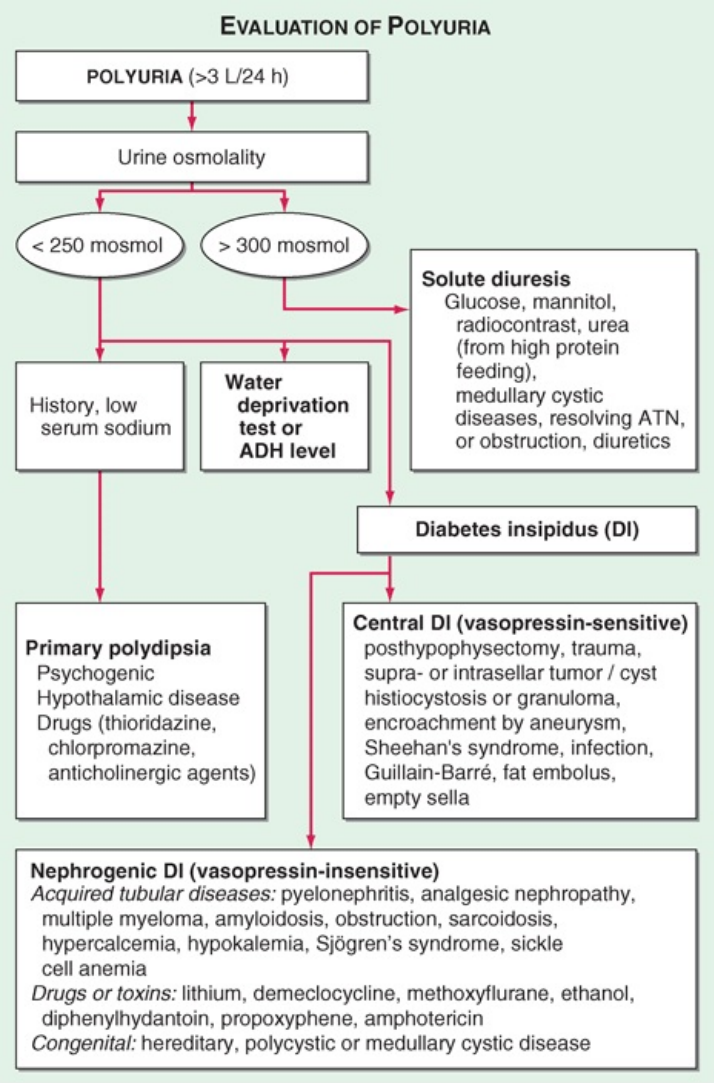

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the patient with polyuria. ADH, antidiuretic hormone; ATN, acute tubular necrosis.

Excessive filtration of a poorly reabsorbed solute such as glucose, mannitol, or urea can depress reabsorption of NaCl and water in the proximal tubule and lead to enhanced excretion in the urine. Poorly controlled diabetes mellitus with glucosuria is the most common cause of a solute diuresis, leading to volume depletion and serum hypertonicity. Since the urine sodium concentration is less than that of blood, more water than sodium is lost, causing hypernatremia and hypertonicity. Common iatrogenic solute diuresis occurs from mannitol administration, radiocontrast media, and high-protein feedings (enterally or parenterally), leading to increased urea production and excretion. Less commonly, excessive sodium loss may result from cystic renal diseases or Bartter's syndrome or during

the course of a tubulointerstitial process (such as resolving ATN). In these so-called salt-wasting disorders, the tubule damage results in direct impairment of sodium reabsorption and indirectly reduces the responsiveness of the tubule to aldosterone. Usually, the sodium losses are mild, and the obligatory urine output is <2 L/d (resolving ATN and postobstructive diuresis are exceptions and may be associated with significant natriuresis and polyuria).

Formation of large volumes of dilute urine is usually due to polydipsic states or diabetes insipidus. Primary polydipsia can result from habit, psychiatric disorders, neurologic lesions, or medications. During deliberate polydipsia, extracellular fluid volume is normal or expanded and plasma vasopressin levels are reduced because serum osmolality tends to be near the lower limits of normal. Urine osmolality should also be maximally dilute at 50 mosmol/L.

Central diabetes insipidus may be idiopathic in origin or secondary to a variety of hypothalamic conditions, including posthypophysectomy or trauma or neoplastic, inflammatory, vascular, or infectious hypothalamic diseases. Idiopathic central diabetes insipidus is associated with selective destruction of the vasopressin-secreting neurons in the supraoptic and paraventricular nuclei and can be inherited as an autosomal dominant trait or occur spontaneously. Nephrogenic diabetes insipidus can occur in a variety of clinical situations, as summarized in Fig. 44-4.

A plasma vasopressin level is recommended as the best method for distinguishing between central and nephrogenic diabetes insipidus. Alternatively, a water deprivation test plus exogenous vasopressin may distinguish primary polydipsia from central and nephrogenic diabetes insipidus. **For a detailed discussion, see Chap. 340.**

FURTHER READINGS

Anderson S et al: Renal and systemic manifestations of glomerular disease, in *Brenner & Rector's The Kidney*, 8th ed, BM Brenner (ed). Philadelphia, Saunders, 2008, pp. 820–838

Israni AK, Kasiske BL: Laboratory assessment of renal disease: Clearance, urinalysis and renal biopsy, in *Brenner & Rector's The Kidney* , 8th ed, BM Brenner (ed). Philadelphia, Saunders, 2008, pp 724 –756

Rodrigo E et al: Measurement of renal function in pre-ESRD patients. Kidney Int Suppl 80:11, 2002 [PMID: 11982806]

Sasaki S: Nephrogenic diabetes insipidus: Update of genetic and clinical aspects. Nephrol Dial Transplant 19:1351, 2004 [PMID: 15004257]

Shrier RW et al: Acute renal failure: Definitions, diagnosis, pathogenesis and therapy. J Clin Invest 114:5, 2004

Verbalis JG, Berl T: Disorders of water balance, in *Brenner & Rector's The Kidney* , 8th ed, BM Brenner (ed). Philadelphia, Saunders, 2008, pp 459–504

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF URINARY SEDIMENTS AND RENAL BIOPSIES: INTRODUCTION

Key diagnostic features of selected diseases in renal biopsy are illustrated, with light, immunofluorescence, and electron microscopic images. Common urinalysis findings are also documented.

Figure e14-1

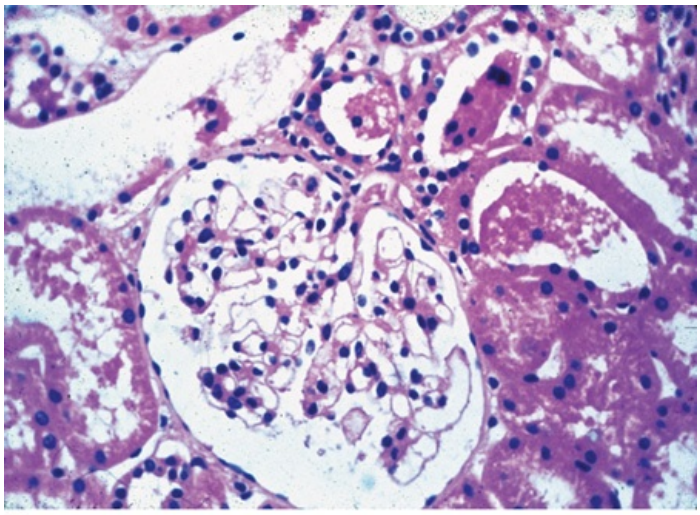

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

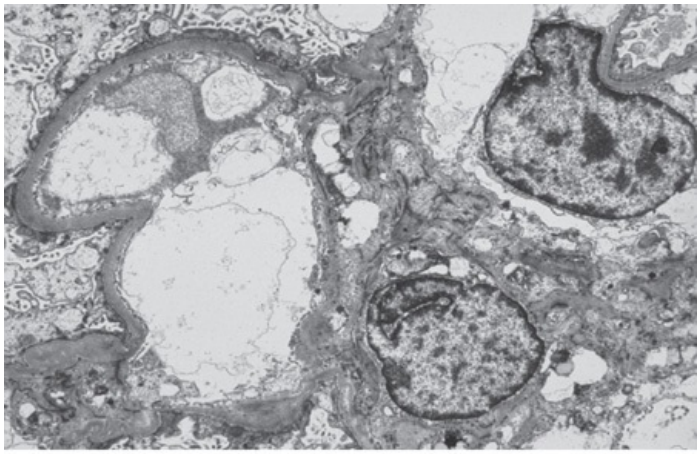

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Minimal-change disease. In minimal-change disease, light microscopy is unremarkable (A), while electron microscopy (B) reveals podocyte injury evidenced by complete foot process effacement. *(ABF/Vanderbilt Collection.)*

Figure e14-2

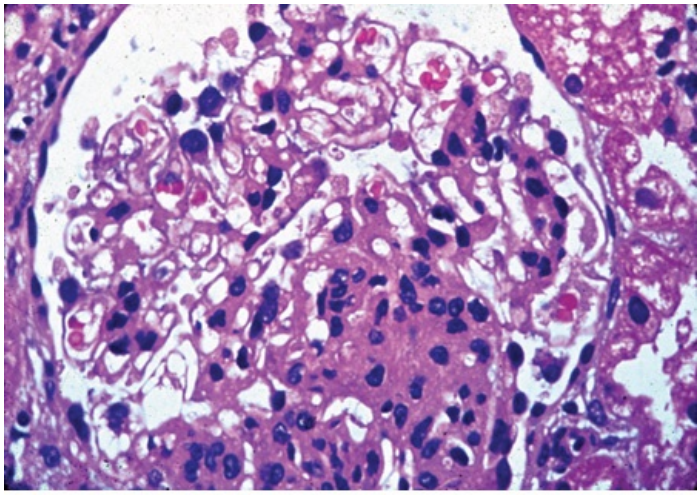

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Focal segmental glomerulosclerosis (FSGS). There is a well-defined segmental increase in matrix and obliteration of capillary loops, the sine qua non of segmental sclerosis not otherwise specified (nos) type. *(EGN/UPenn Collection.)*

Figure e14-3

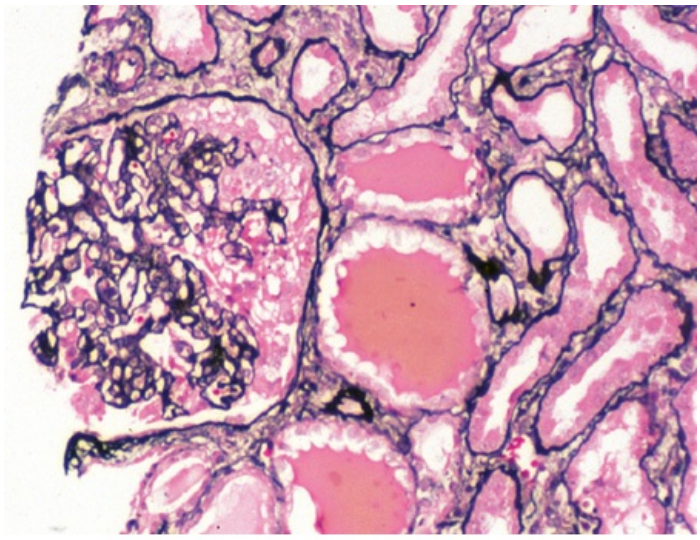

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Collapsing glomerulopathy. There is segmental collapse of the glomerular capillary loops and overlying podocyte hyperplasia. This lesion may be idiopathic or associated with HIV infection and has a particularly poor prognosis. *(ABF/Vanderbilt Collection.)*

Figure e14-4

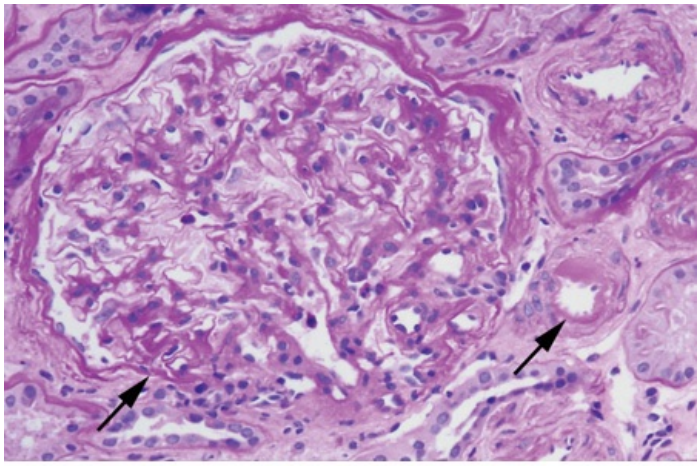

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hilar variant of FSGS. There is segmental sclerosis of the glomerular tuft at the vascular pole with associated hyalinosis, also present in the afferent arteriole (*arrows*). This lesion often occurs as a secondary response when nephron mass is lost due to, e.g., scarring from other conditions. Patients usually have less proteinuria and less steroid response than FSGS, nos type. *(ABF/Vanderbilt Collection.)*

Figure e14-5

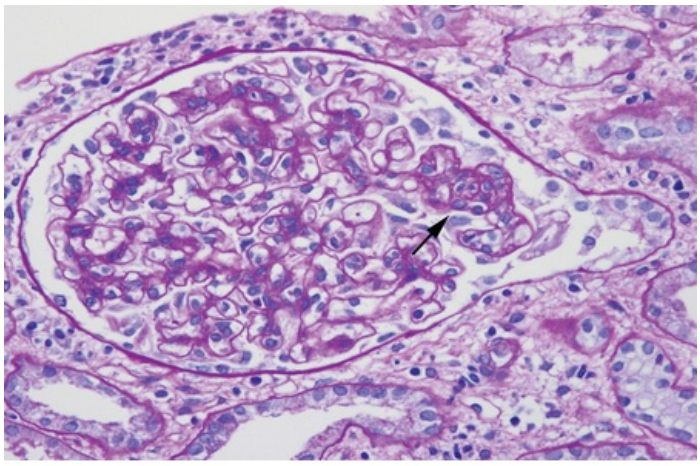

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tip lesion variant of FSGS. There is segmental sclerosis of the glomerular capillary loops at the proximal tubular outlet (*arrow*). This lesion has a better prognosis than other types of FSGS. *(ABF/Vanderbilt Collection.)*

Figure e14-6

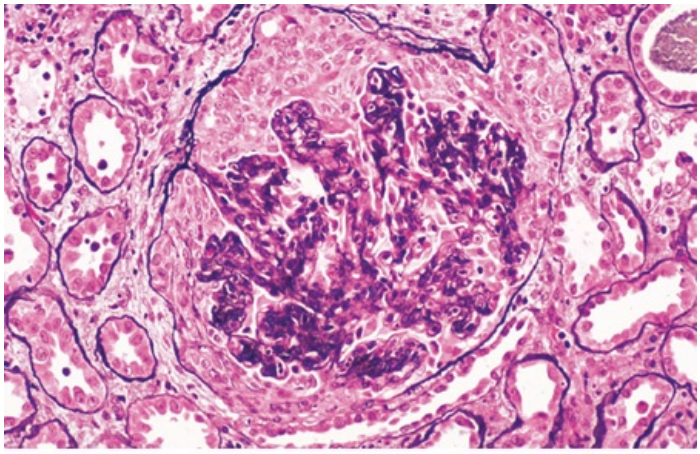

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

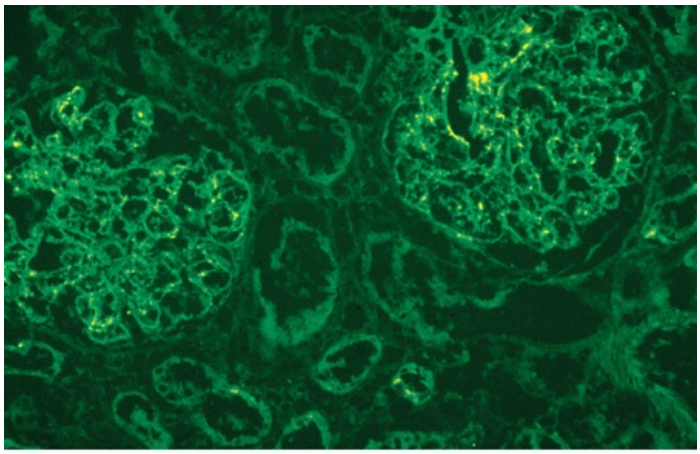

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

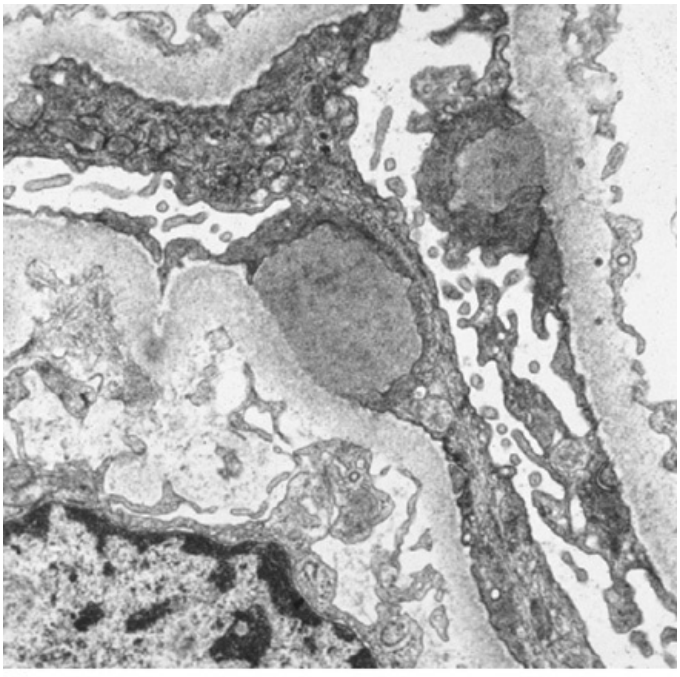

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postinfectious (poststreptococcal) glomerulonephritis. The glomerular tuft shows proliferative

changes with numerous PMNs, with a crescentic reaction in severe cases (*A*). These deposits localize in the mesangium and along the capillary wall in a subepithelial pattern and stain dominantly for C3 and to a lesser extent for IgG (*B*). Subepithelial hump-shaped deposits are seen by electron microscopy (*C*). *(ABF/Vanderbilt Collection.)*

Figure e14-7

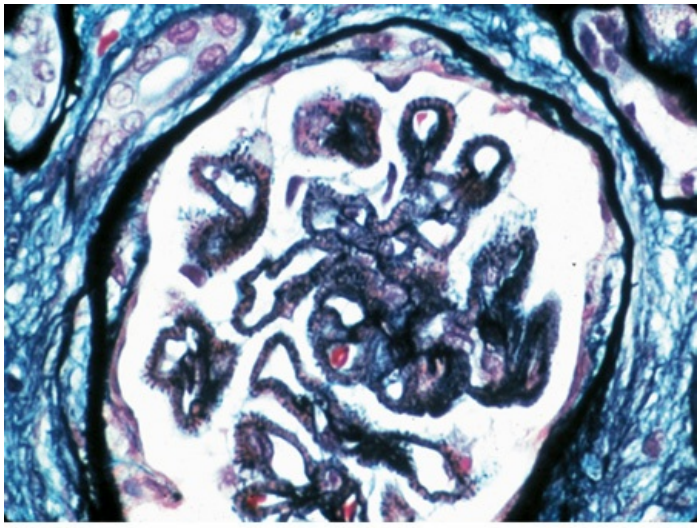

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

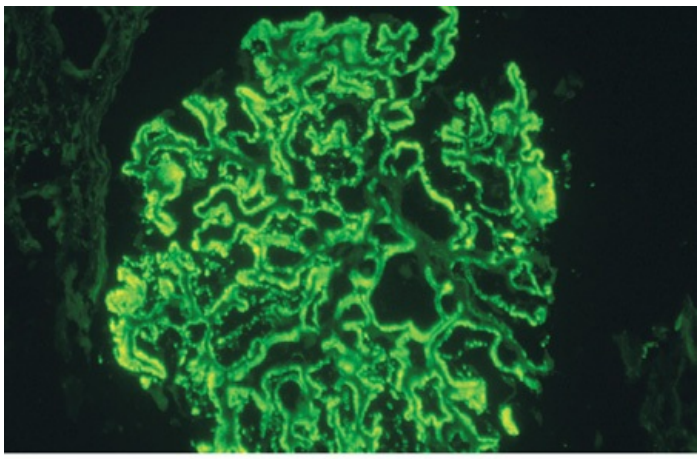

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

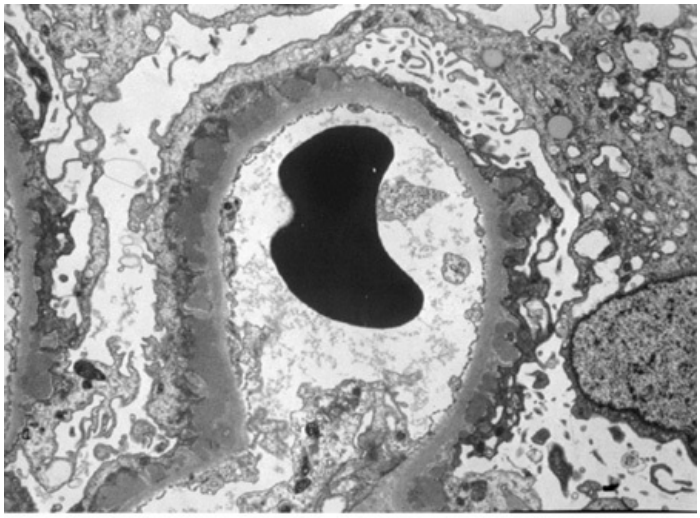

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Membranous glomerulopathy. Membranous glomerulopathy is due to subepithelial deposits, with resulting basement membrane reaction, resulting in the appearance of spike-like projections on silver stain (*A*). The deposits are directly visualized by fluorescent anti-IgG, revealing diffuse granular capillary loop staining (*B*). By electron microscopy, the subepithelial location of the deposits and early surrounding basement membrane reaction is evident, with overlying foot process effacement (*C*). *(ABF/Vanderbilt Collection.)*

Figure e14-8

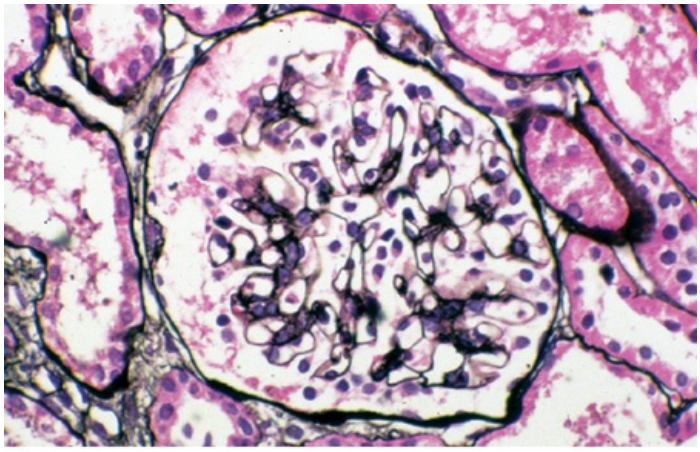

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

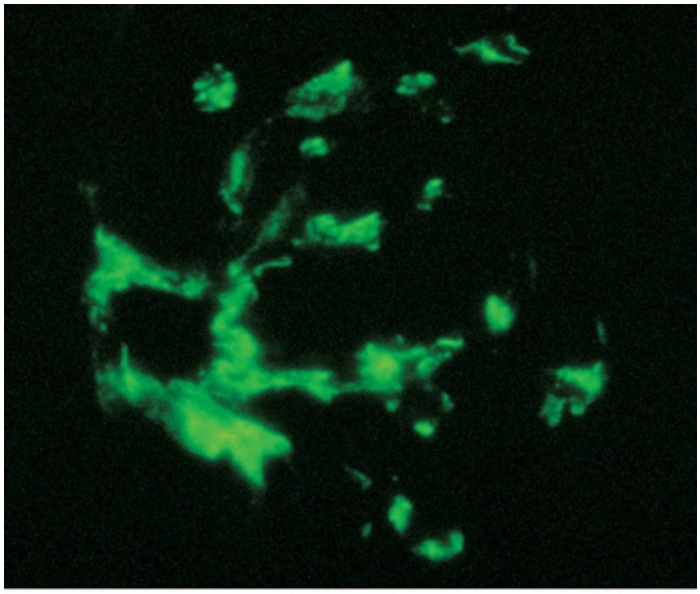

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

IgA nephropathy. There is variable mesangial expansion due to mesangial deposits, with some cases also showing endocapillary proliferation or segmental sclerosis (*A*). By immunofluorescence, mesangial IgA deposits are evident (*B*). *(ABF/Vanderbilt Collection.)*

Figure e14-9

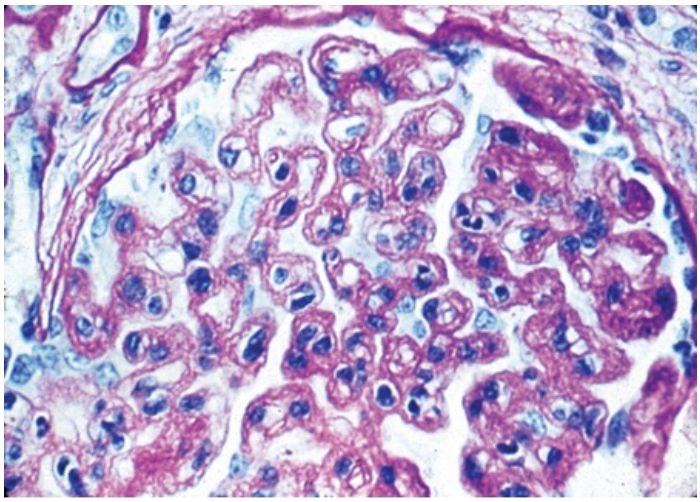

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Membranoproliferative glomerulonephritis. There is mesangial expansion and endocapillary proliferation with cellular interposition in response to subendothelial deposits, resulting in the "tram-track" of duplication of glomerular basement membrane. *(EGN/UPenn Collection.)*

Figure e14-10

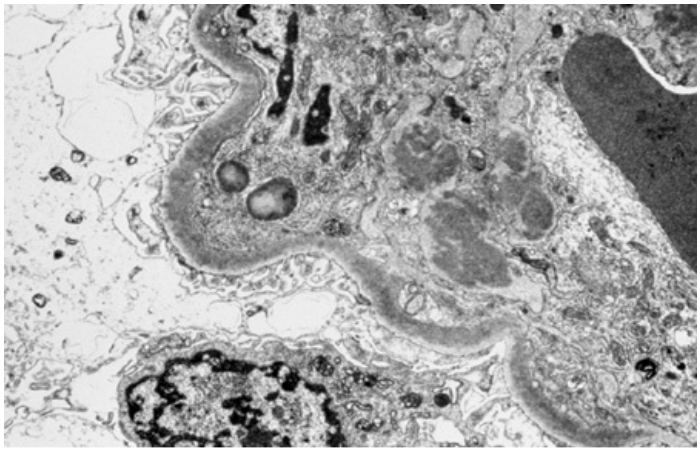

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dense deposit disease (membranoproliferative glomerulonephritis type II). By light microscopy, there is a membranoproliferative pattern. By electron microscopy, there is a dense transformation of the glomerular basement membrane with round, globular deposits within the mesangium. By immunofluorescence, only C3 staining is usually present. *(ABF/Vanderbilt Collection.)*

Figure e14-11

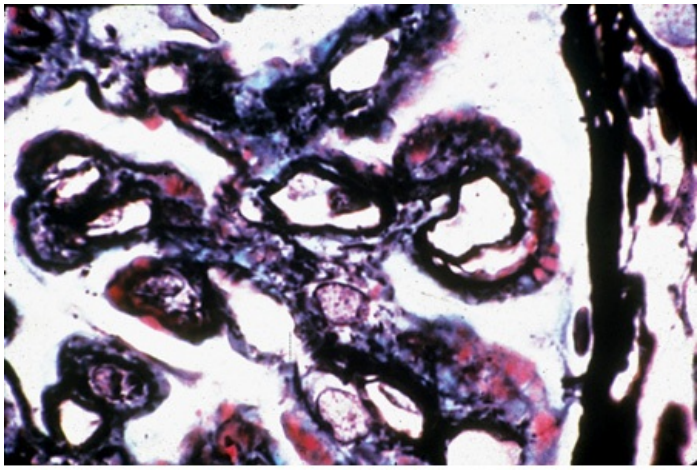

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mixed proliferative and membranous glomerulonephritis. This specimen shows pink subepithelial deposits with spike reaction, and the "tram-track" sign of reduplication of glomerular basement membrane, resulting from subendothelial deposits, as may be seen in mixed membranous and proliferative lupus nephritis (ISN/RPS class V and IV). *(EGN/UPenn Collection.)*

Figure e14-12

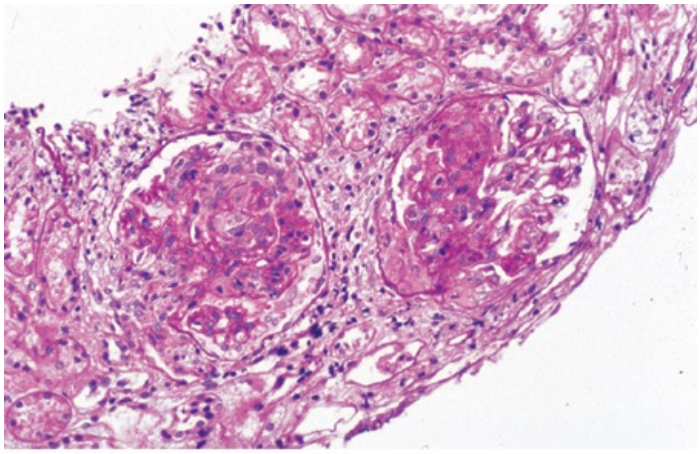

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

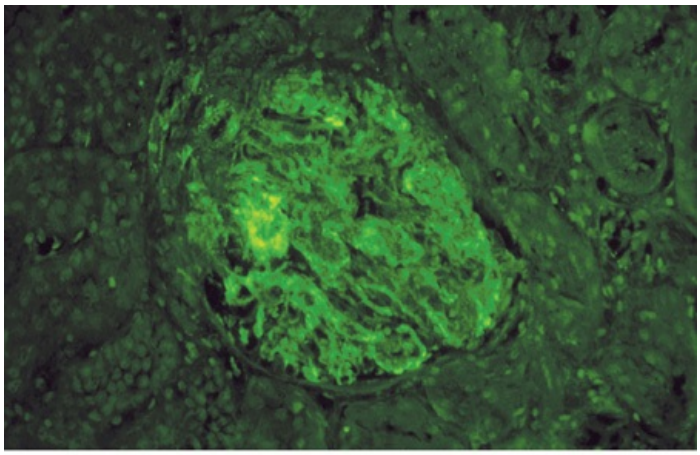

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

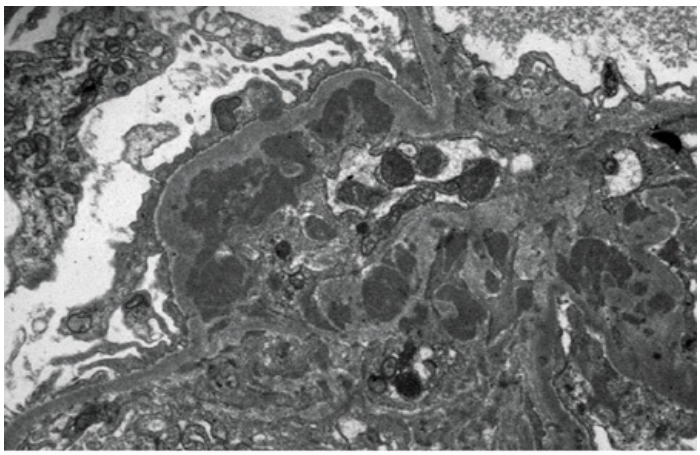

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lupus nephritis. Proliferative lupus nephritis, ISN/RPS class III (focal) or IV (diffuse), manifests as endocapillary proliferation, which may result in segmental necrosis due to deposits, particularly in the subendothelial area (A). By immunofluorescence, chunky irregular mesangial and capillary loop deposits are evident, with some of the peripheral loop deposits having a smooth, molded outer contour due to their subendothelial location. These deposits typically stain for all three immunoglobulins, IgG, IgA, IgM, and both C3 and C1q (*B*). By electron microscopy, subendothelial, mesangial, and rare subepithelial dense immune complex deposits are evident, along with extensive foot process effacement (*C*). *(ABF/Vanderbilt Collection.)*

Figure e14-13

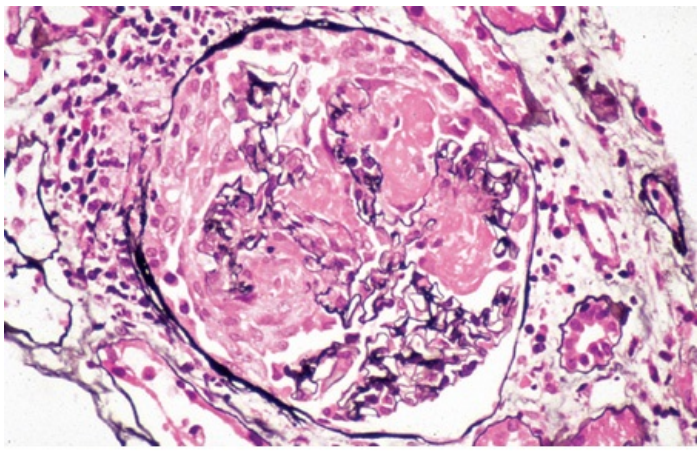

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Granulomatosis with polyangiitis (Wegener's). This pauci-immune necrotizing crescentic glomerulonephritis shows numerous breaks in the glomerular basement membrane with associated segmental fibrinoid necrosis, and a crescent formed by proliferation of the parietalepithelium. Note that the uninvolved segment of the glomerulus (at ~5 o'clock) shows no evidence of proliferation or immune complexes. *(ABF/Vanderbilt Collection.)*

Figure e14-14

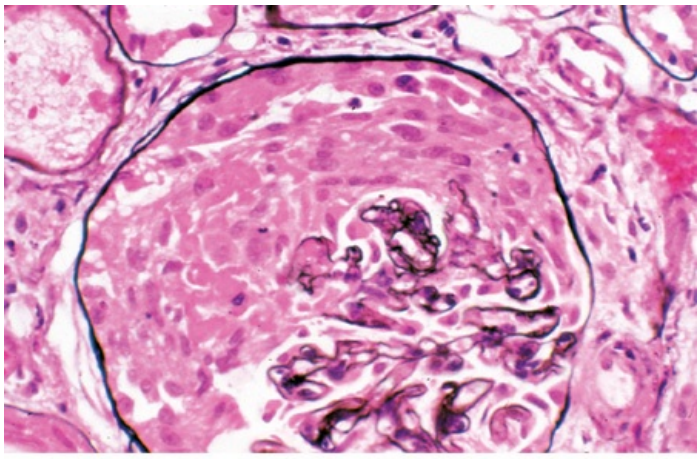

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

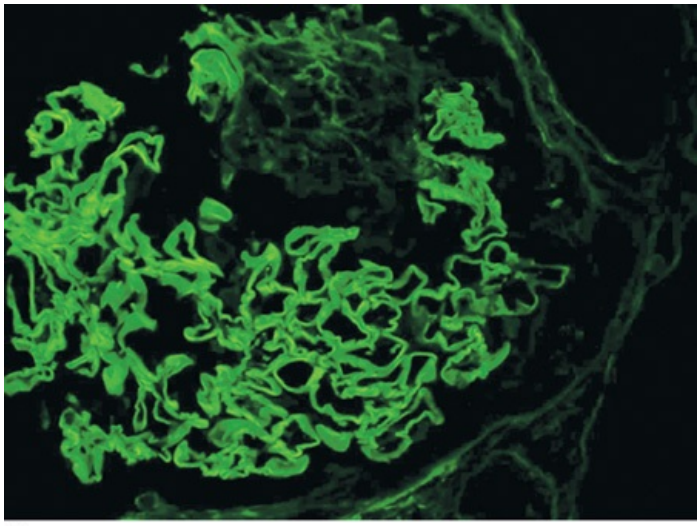

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anti-GBM antibody-mediated glomerulonephritis. There is segmental necrosis with a break of the glomerular basement membrane and a cellular crescent (*A*), and immunofluorescence for IgG shows linear staining of the glomerular basement membrane with a small crescent at ~1 o'clock (*B*). *(ABF/Vanderbilt Collection.)*

Figure e14-15

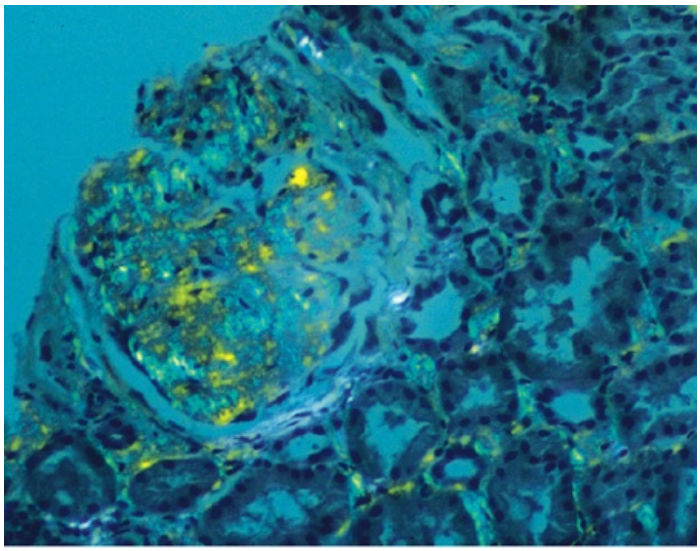

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

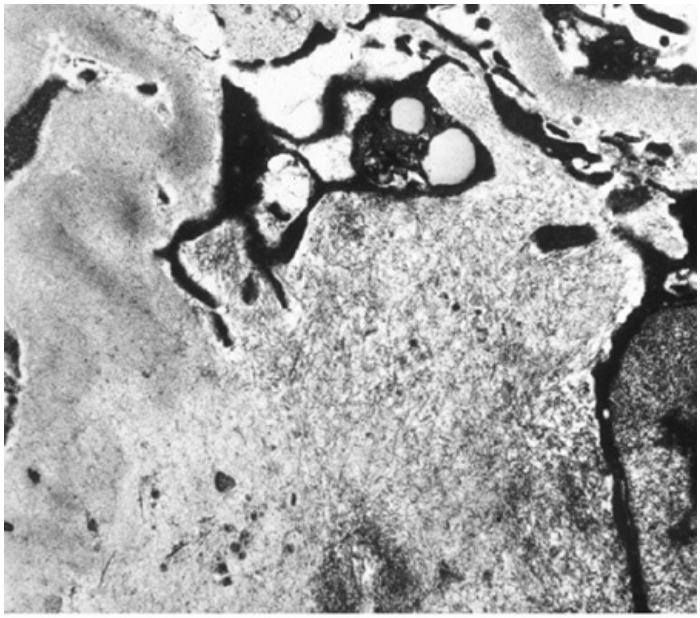

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Amyloidosis. Amyloidosis shows amorphous, acellular expansion of the mesangium, with material often also infiltrating glomerular basement membranes, vessels, and in the interstitium, with apple-green birefringence by polarized congo red stain (*A*). the deposits are composed of randomly organized 9- to 11-nm fibrils by electron microscopy (*B*). *(ABF/Vanderbilt collection.)*

Figure e14-16

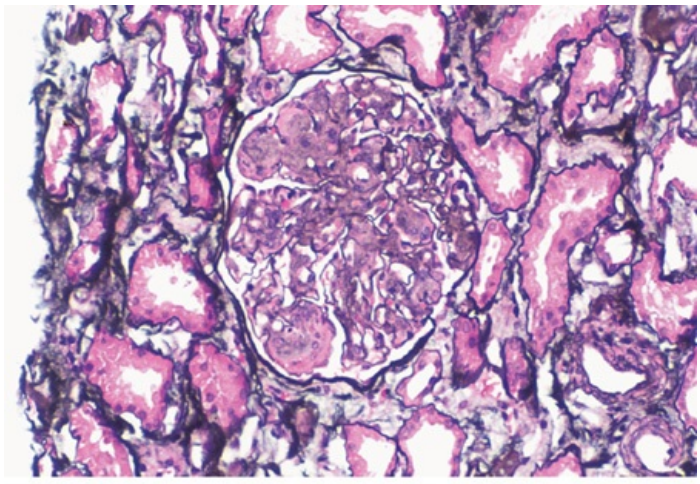

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

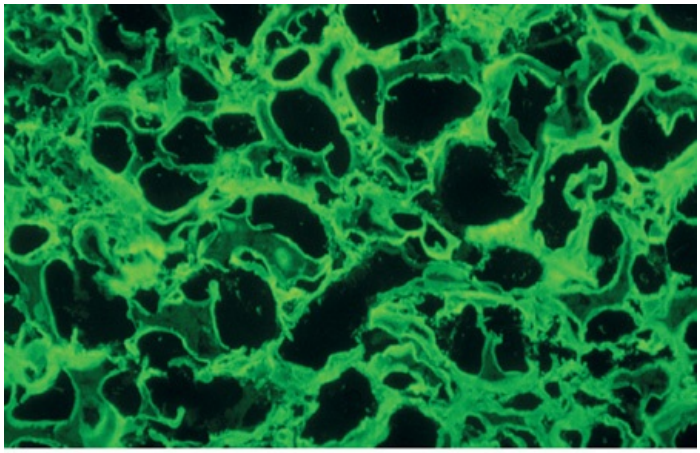

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

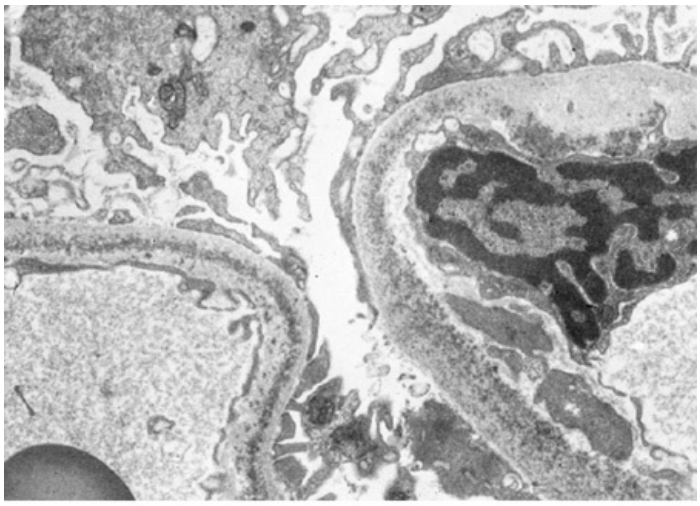

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Light chain deposition disease. There is mesangial expansion, often nodular by light microscopy (*A*), with immunofluorescence showing monoclonal staining, more commonly with kappa than lambda light chain, of tubules (*B*) and glomerular tufts. By electron microscopy (C), the deposits show an amorphous granular appearance and line the inside of the glomerular basement membrane and are also found along the tubular basement membranes. *(ABF/Vanderbilt Collection.)*

Figure e14-17

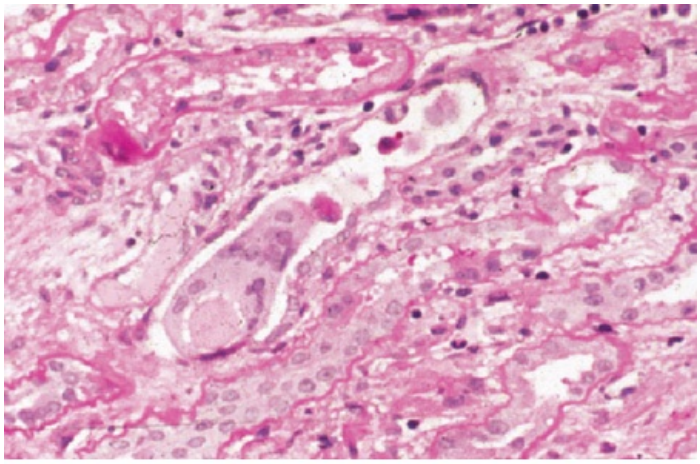

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Light chain cast nephropathy (myeloma kidney). Monoclonal light chains precipitate in tubules and result in a syncytial giant cell reaction surrounding the casts, and a surrounding chronic interstitial nephritis with tubulointerstitial fibrosis. *(ABF/Vanderbilt Collection.)*

Figure e14-18

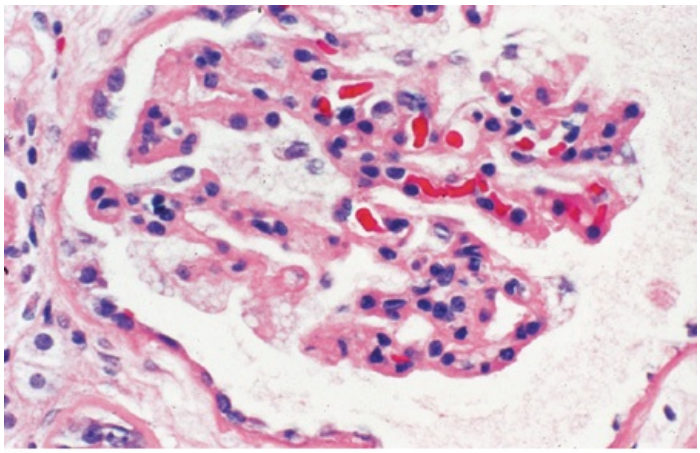

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

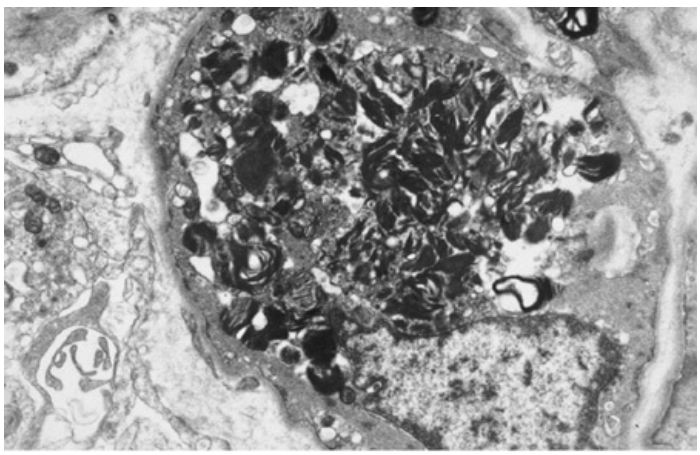

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fabry's disease. Due to deficiency of α-galactosidase, there is abnormal accumulation of glycolipids, resulting in foamy podocytes by light microscopy (*A*). These deposits can be directly visualized by

electron microscopy (*B*), where the glycosphingolipid appears as whorled so-called myeloid bodies, particularly in the podocytes. *(ABF/Vanderbilt Collection.)*

Figure e14-19

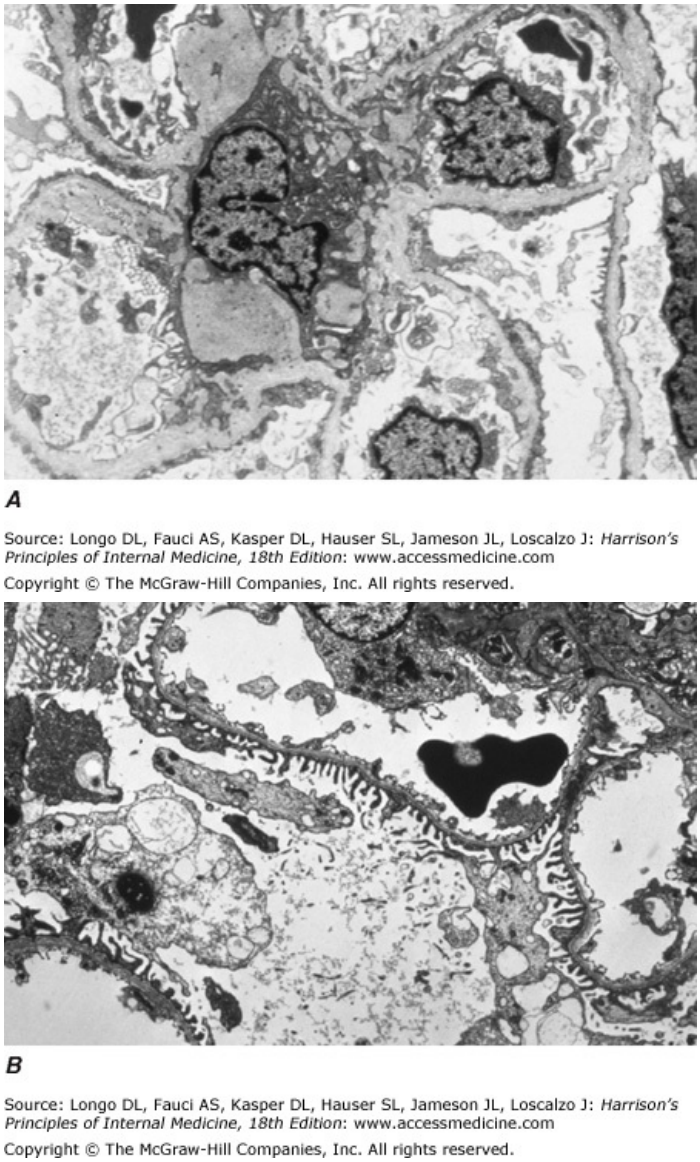

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Alport's syndrome and thin glomerular basement membrane lesion. In Alport's syndrome, there is irregular thinning alternating with thickened so-called basket-weaving abnormal organization of the glomerular basement membrane (*A*). In benign familial hematuria, or in early cases of Alport's syndrome or female carriers, only extensive thinning of the GBM is seen by electron microscopy (*B*). *(ABF/Vanderbilt Collection.)*

Figure e14-20

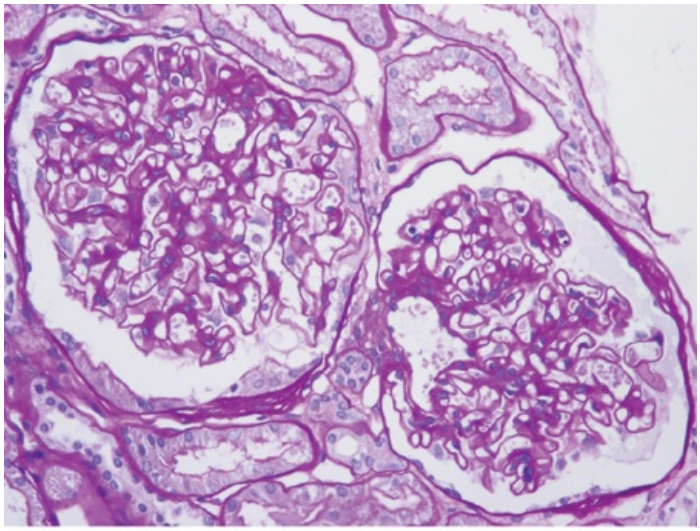

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

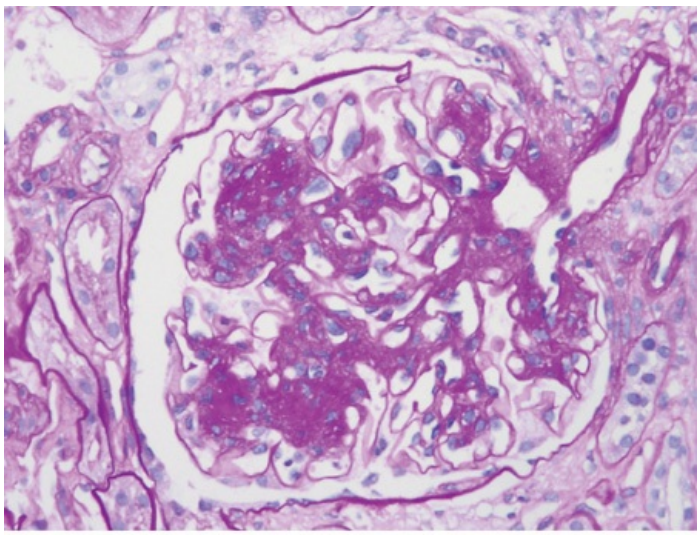

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

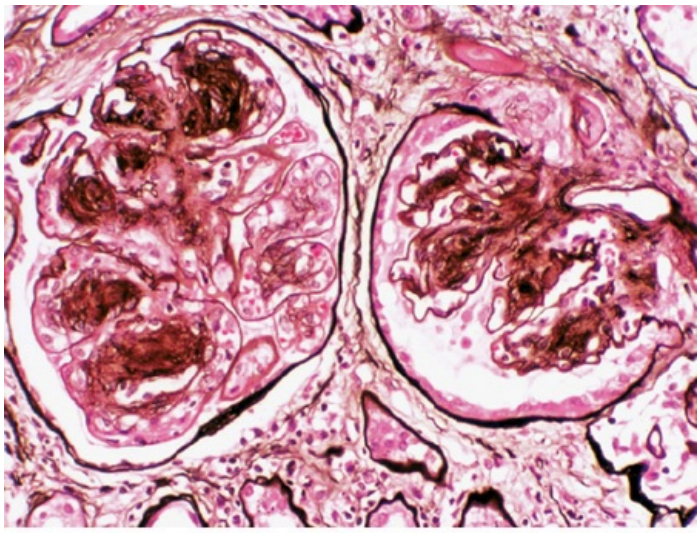

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diabetic nephropathy. In the earliest stage of diabetic nephropathy, only mild mesangial increase

Diabetic nephropathy. In the earliest stage of diabetic nephropathy, only mild mesangial increase and prominent glomerular basement membranes (confirmed to be thickened by electron microscopy) are present (*A*). In slightly more advanced stages, more marked mesangial expansion with early nodule formation develops, with evident arteriolar hyaline (*B*). In established diabetic nephropathy, there is nodular mesangial expansion, so-called Kimmelstiel-Wilson nodules, with increased mesangial matrix and cellularity, microaneurysm formation in the glomerulus on the left, and prominent glomerular basement membranes without evidence of immune deposits and arteriolar hyalinosis of both afferent and efferent arterioles (*C*). *(ABF/Vanderbilt Collection.)*

Figure e14-21

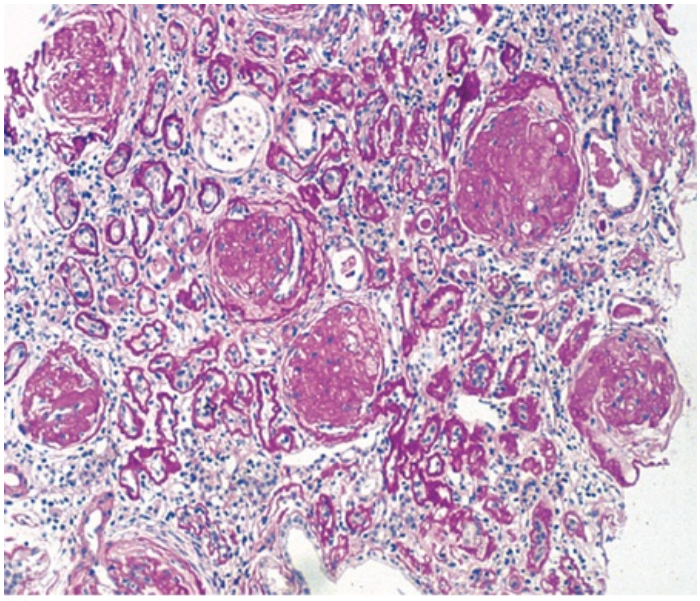

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

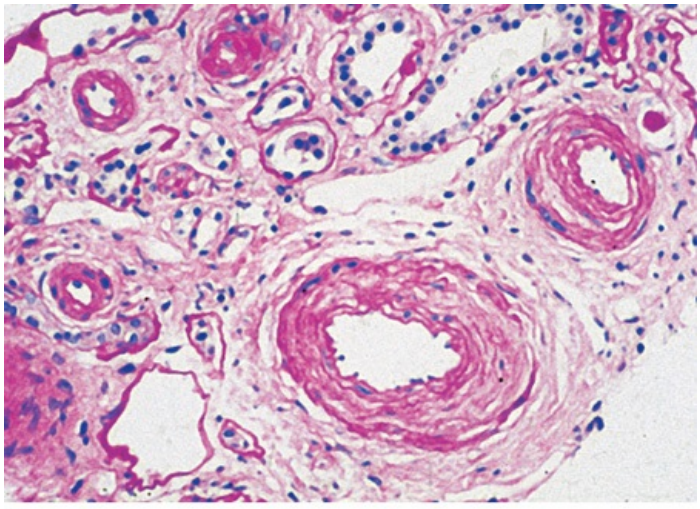

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arterionephrosclerosis. Hypertension-associated injury often manifests extensive global sclerosis of glomeruli, with accompanying and proportional tubulointerstitial fibrosis and pericapsular fibrosis, and there may be segmental sclerosis (A). The vessels show disproportionately severe changes of intimal fibrosis, medial hypertrophy, and arteriolar hyaline deposits (B). *(ABF/Vanderbilt Collection.)*

Figure e14-22

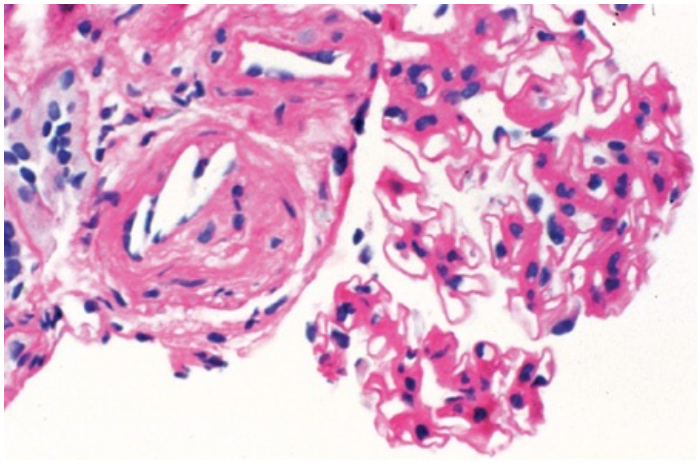

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cholesterol emboli. Cholesterol emboli cause cleft-like spaces where the lipid has been extracted during processing, with smooth outer contours, and surrounding fibrotic and mononuclear cell reaction in these arterioles. *(ABF/Vanderbilt Collection.)*

Figure e14-23

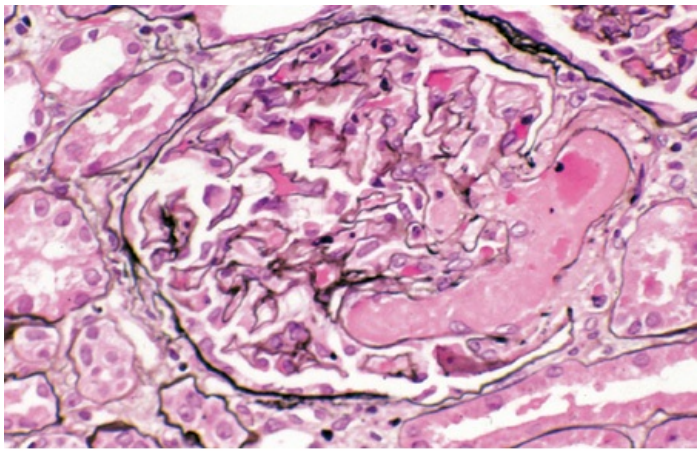

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hemolytic uremic syndrome. There are characteristic intraglomerular fibrinthrombi, with a chunky pink appearance (thrombotic microangiopathy). The remaining portion of the capillary tuft shows corrugation of the glomerular basement membrane due to ischemia. *(ABF/Vanderbilt Collection.)*

Figure e14-24

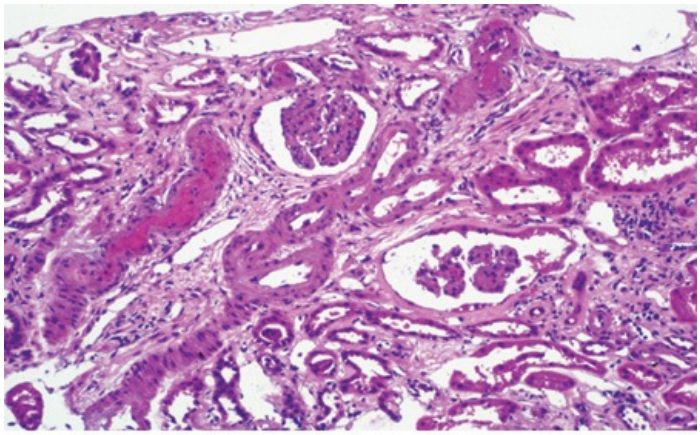

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

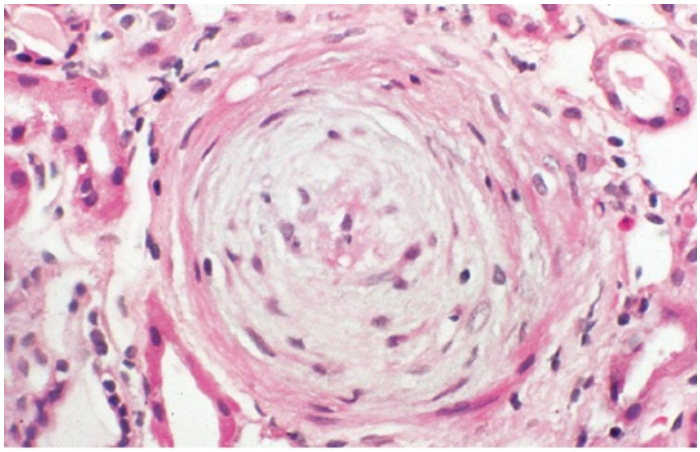

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Progressive systemic sclerosis. Acutely, there is fibrinoid necrosis of interlobular and larger vessels, with intervening normal vessels and ischemic change in the glomeruli (*A*). Chronically, this injury leads to intimal proliferation, the so-called onion-skinning appearance (*B*). *(ABF/Vanderbilt Collection.)*

Figure e14-25

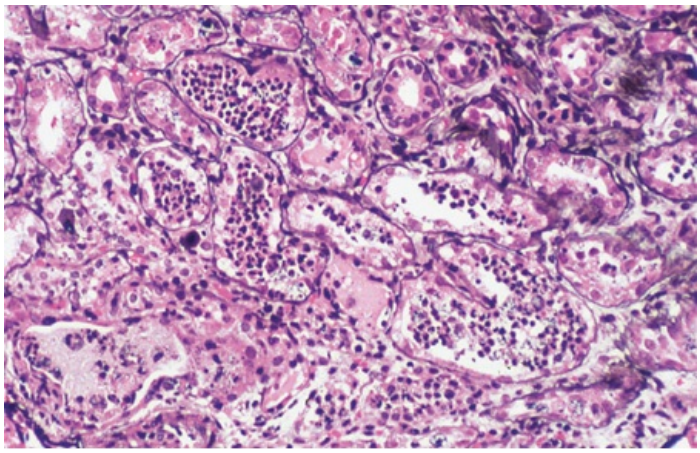

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute pyelonephritis. There are characteristic intratubular plugs and casts of PMNs with inflammation extending into the surrounding interstitium, and accompanying tubular injury. *(ABF/Vanderbilt Collection.)*

Figure e14-26

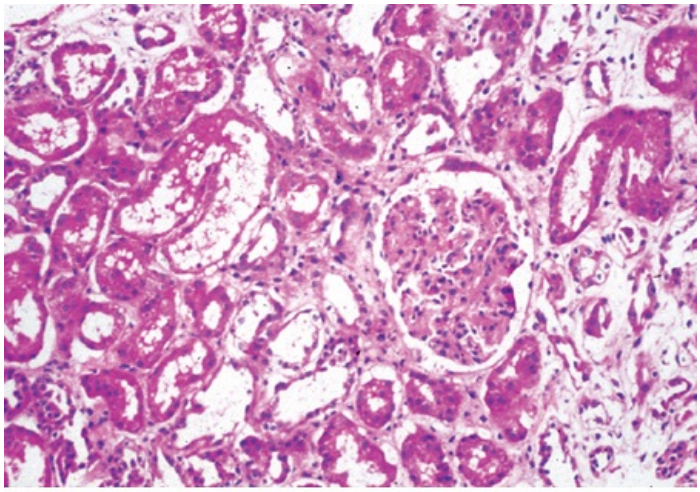

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute tubular injury. There is extensive flattening of the tubular epithelium and loss of the brush border, with mild interstitial edema, characteristic of acute tubular injury due to ischemia. *(ABF/Vanderbilt Collection.)*

Figure e14-27

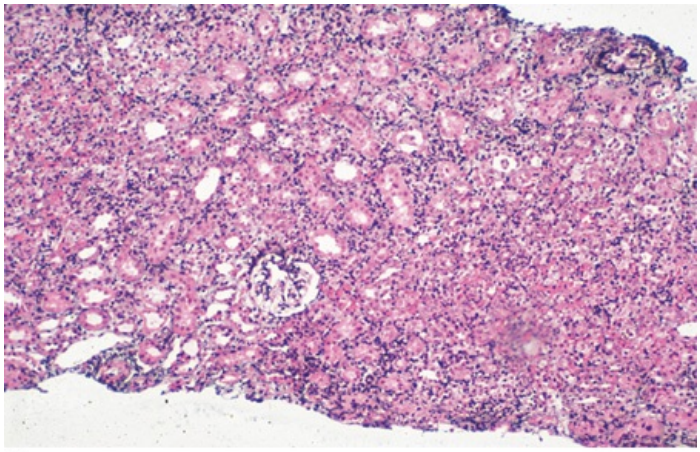

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

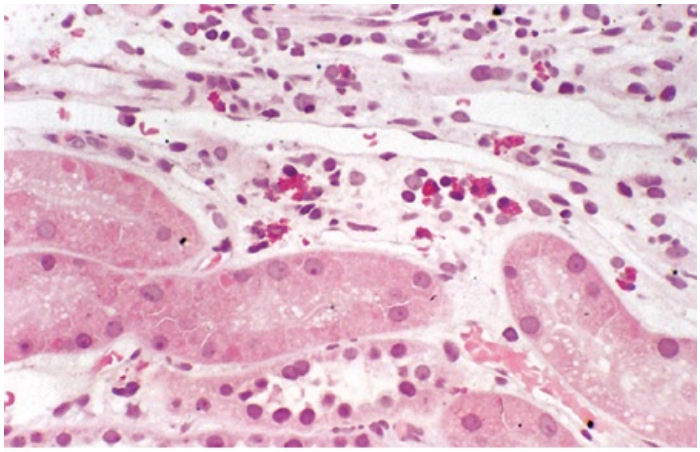

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute interstitial nephritis. There is extensive interstitial lymphoplasmocytic infiltrate with mild edema and associated tubular injury (A), which is frequently associated with interstitial eosinophils (*B*) when caused by a drug hypersensitivity reaction. *(ABF/Vanderbilt Collection.)*

Figure e14-28

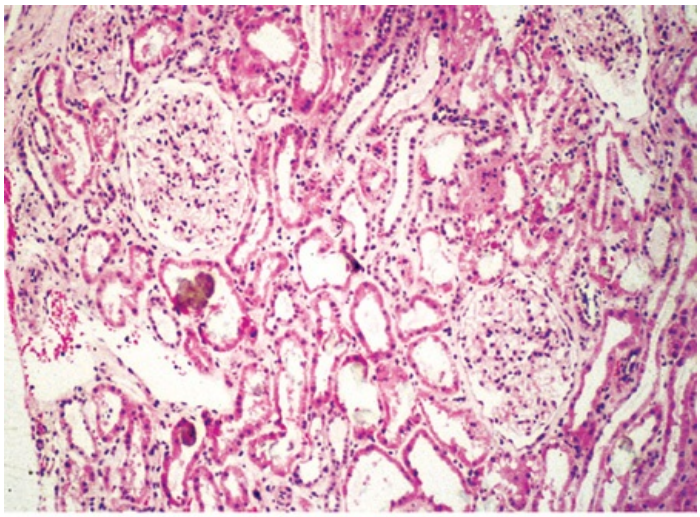

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

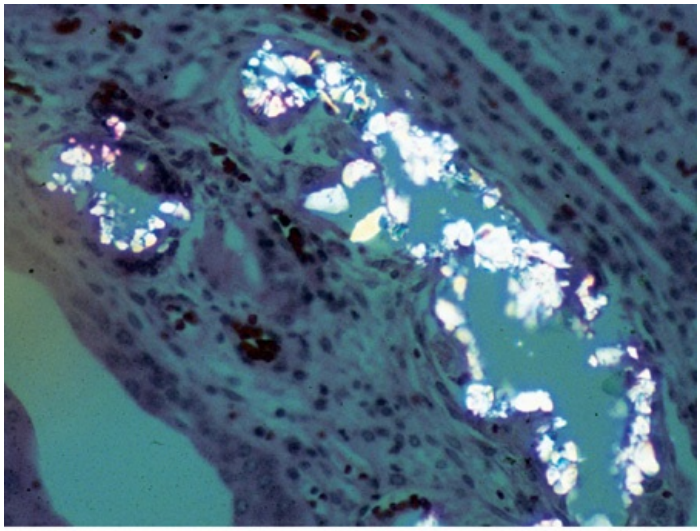

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oxalosis. Calcium oxalate crystals have caused extensive tubular injury, with flattening and regeneration of tubular epithelium (A). Crystals are well visualized as sheaves when viewed under polarized light (*B*). *(ABF/Vanderbilt Collection.)*

Figure e14-29

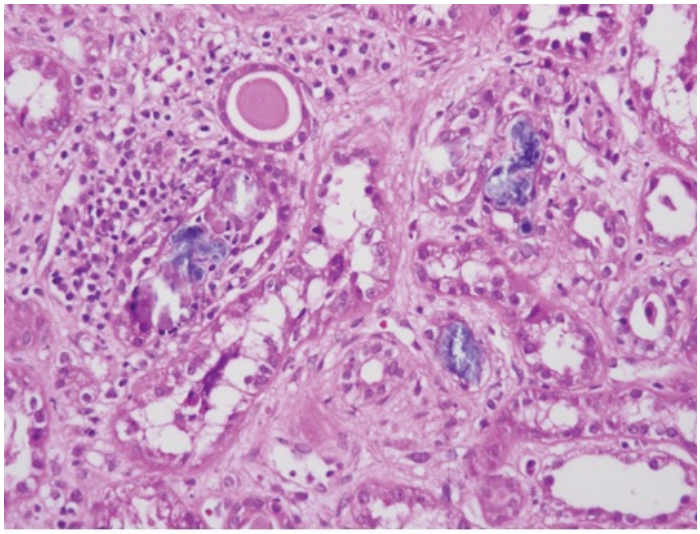

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute phosphate nephropathy. There is extensive acute tubular injury with intratubular nonpolarizable calcium phosphate crystals. *(ABF/Vanderbilt Collection.)*

Figure e14-30

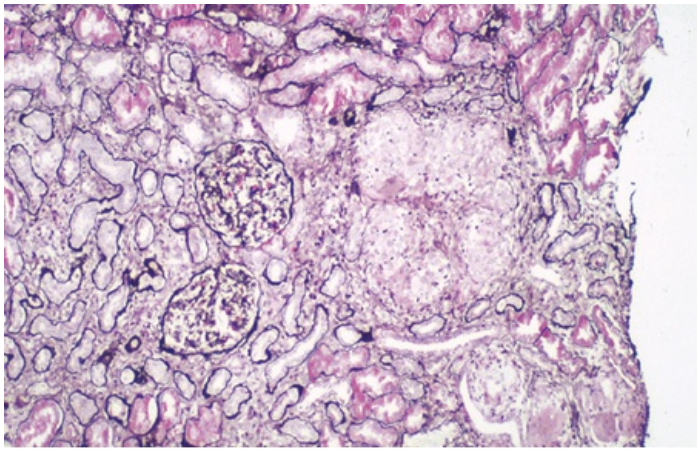

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sarcoidosis. There is chronic interstitial nephritis with numerous, confluent, non-necrotizing granulomas. The glomeruli are unremarkable, but there is moderate tubular atrophy and interstitial fibrosis. *(ABF/Vanderbilt Collection.)*

Figure e14-31

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hyaline cast.*(ABF/Vanderbilt Collection.)*

Figure e14-32

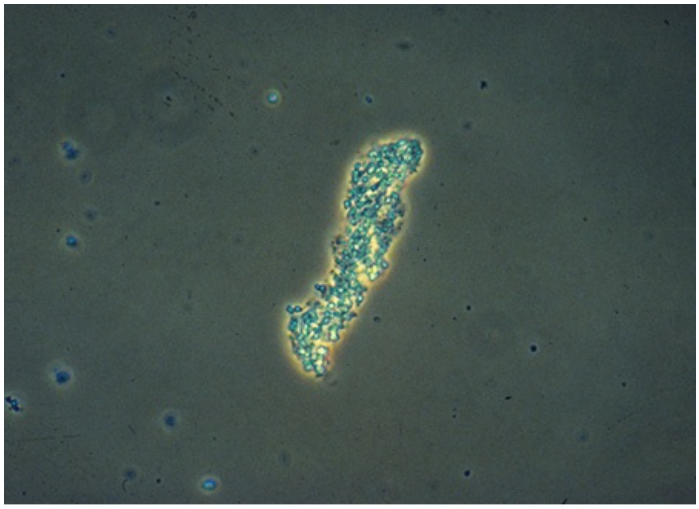

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coarse granular cast.*(ABF/Vanderbilt Collection.)*

Figure e14-33

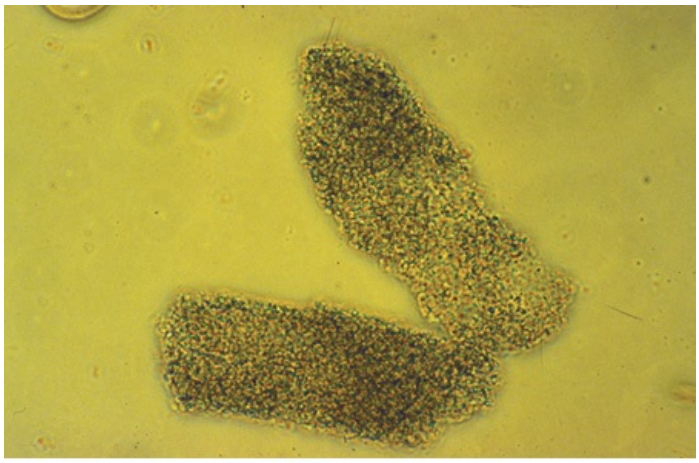

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fine granular casts.*(ABF/Vanderbilt Collection.)*

Figure e14-34

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Red blood cell cast.*(ABF/Vanderbilt Collection.)*

Figure e14-35

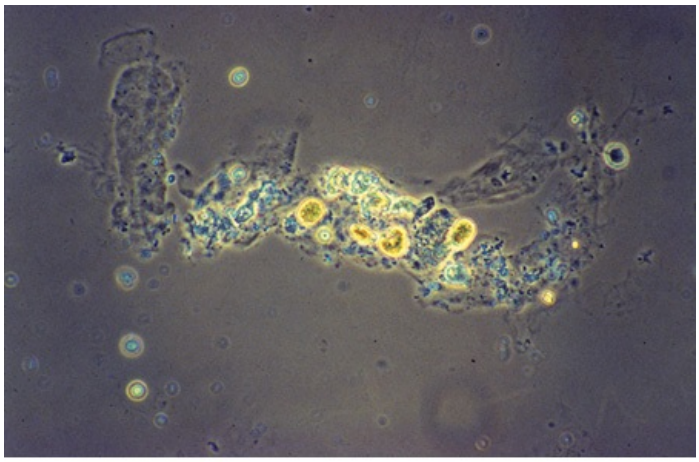

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

WBC cast.*(ABF/Vanderbilt Collection.)*

Figure e14-36

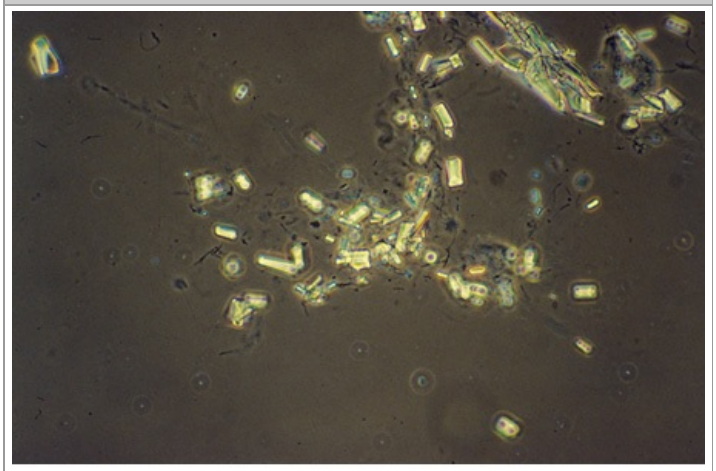

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Triple phosphate crystals.*(ABF/Vanderbilt Collection.)*

Figure e14-37

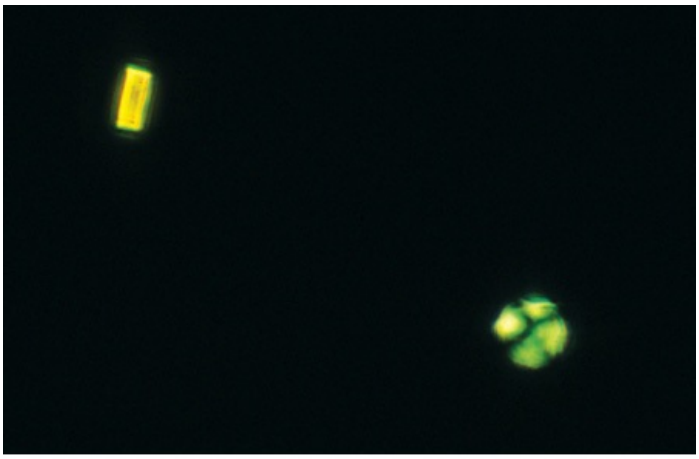

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

"Maltese cross" formation in an oval fat body.*(ABF/Vanderbilt Collection.)*

Figure e14-38

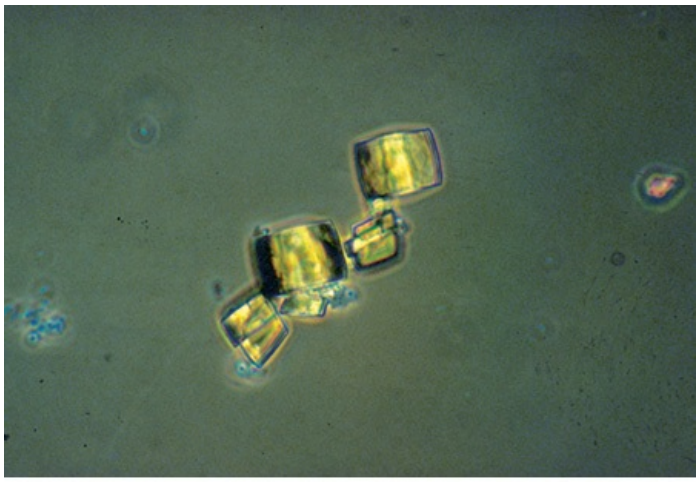

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Uric acid crystals.*(ABF/Vanderbilt Collection.)*

Copyright © The McGraw-Hill Companies. All rights reserved.
The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SODIUM AND WATER

Composition of Body Fluids

Water is the most abundant constituent in the body, accounting for ~50% of body weight in women and 60% in men. Total body water is distributed in two major compartments: 55–75% is intracellular [intracellular fluid (ICF)], and 25–45% is extracellular [extracellular fluid (ECF)]. ECF is subdivided into intravascular (plasma water) and extravascular (interstitial) spaces in a ratio of 1:3. Fluid movement between the intravascular and interstitial spaces occurs across the capillary wall and is determined by Starling forces, i.e., capillary hydraulic pressure and colloid osmotic pressure. The transcapillary hydraulic pressure gradient exceeds the corresponding oncotic pressure gradient, thus favoring the movement of plasma ultrafiltrate into the extravascular space. The return of fluid into the intravascular compartment occurs via lymphatic flow.

The solute or particle concentration of a fluid is known as its osmolality and is expressed as milliosmoles per kilogram of water (mosmol/kg). Water easily diffuses across most cell membranes to achieve osmotic equilibrium (ECF osmolality = ICF osmolality). Notably, the extracellular and intracellular solute compositions differ considerably owing to the activity of various transporters, channels, and ATP-driven membrane pumps. The major ECF particles are Na^+ and its accompanying anions Cl^- and HCO_3^-, whereas K^+ and organic phosphate esters (ATP, creatine phosphate, and phospholipids) are the predominant ICF osmoles. Solutes that are restricted to the ECF or the ICF determine the tonicity or effective osmolality of that compartment. Certain solutes, particularly urea, do not contribute to water shifts across most membranes and are thus known as *ineffective osmoles*.

WATER BALANCE

Vasopressin secretion, water ingestion, and renal water transport collaborate to maintain human body fluid osmolality between 280 and 295 mosmol/kg. Vasopressin (AVP) is synthesized in magnocellular neurons within the hypothalamus; the distal axons of those neurons project to the posterior pituitary or neurohypophysis, from which AVP is released into the circulation. A network of central osmoreceptor neurons that includes the AVP-expressing magnocellular neurons themselves sense circulating osmolality via nonselective, stretch-activated cation channels. These osmoreceptor neurons are activated or inhibited by modest increases and decreases in circulating osmolality, respectively; activation leads to AVP release and thirst.

AVP secretion is stimulated as systemic osmolality increases above a threshold level of ~285 mosmol/kg, above which there is a linear relationship between osmolality and circulating AVP (**Fig. 45-1**). Thirst and thus water ingestion also are activated at ~285 mosmol/kg, beyond which there is an equivalent linear increase in the perceived intensity of thirst as a function of circulating osmolality. Changes in blood volume and blood pressure are also direct stimuli for AVP release and thirst, albeit with a less sensitive response profile. Of perhaps greater clinical relevance to the pathophysiology of water homeostasis, ECF volume strongly modulates the relationship between circulating osmolality and AVP release so that *hypovolemia* reduces the osmotic threshold and increases the slope of the response curve to osmolality; *hypervolemia* has the opposite effect, increasing the osmotic threshold and reducing the slope of the response curve (Fig. 45-1). Notably, AVP has a half-life in the circulation of only 10–20 min; thus, changes in extracellular fluid volume and/or circulating osmolality can affect water homeostasis rapidly. In addition to volume status, a number of nonosmotic stimuli have potent activating effects on osmosensitive neurons and AVP release, including nausea, intracerebral angiotensin II, serotonin, and multiple drugs.

Figure 45-1

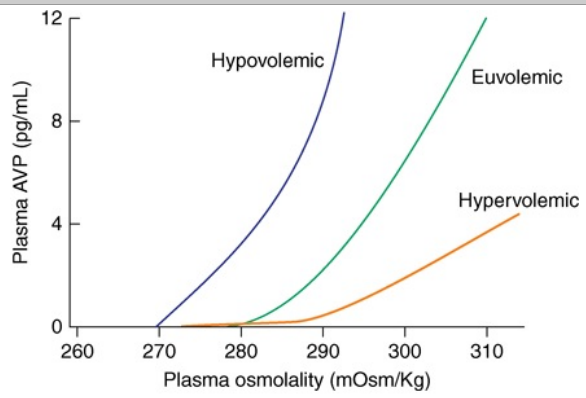

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Circulating levels of vasopressin (AVP) in response to changes in osmolality. Plasma vasopressin becomes detectable in euvolemic, healthy individuals at a threshold of ~285 mOsm/Kg, above which there is a linear relationship between osmolality and circulating AVP. The vasopressin response to osmolality is modulated strongly by volume status. The osmotic threshold is thus slightly lower in hypovolemia, with a steeper response curve; hypervolemia reduces the sensitivity of circulating AVP levels to osmolality.

The excretion or retention of electrolyte-free water by the kidney is modulated by circulating AVP. AVP acts on renal V_2-type

receptors in the thick ascending limb of Henle and principal cells of the collecting duct (CD), increasing cyclic adenosine monophosphate (AMP) and activating protein kinase A (PKA)–dependent phosphorylation of multiple transport proteins. The AVP- and PKA-dependent activation of Na^+-Cl^- and K^+ transport by the thick ascending limb of the loop of Henle (TALH) is a key participant in the countercurrent mechanism (**Fig. 45-2**). The countercurrent mechanism ultimately increases the interstitial osmolality in the inner medulla of the kidney, driving water absorption across the renal collecting duct. However, water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (Fig. 45-2). Water transport across apical and basolateral aquaporin-1 water channels in the descending thin limb of the loop of Henle is thus involved, as is passive absorption of Na^+-Cl^- by the thin ascending limb, via apical and basolateral CLC-K1 chloride channels and paracellular Na^+ transport. Renal urea transport in turn plays important roles in the generation of the medullary osmotic gradient and the ability to excrete solute-free water under conditions of both high and low protein intake (Fig. 45-2).

Figure 45-2

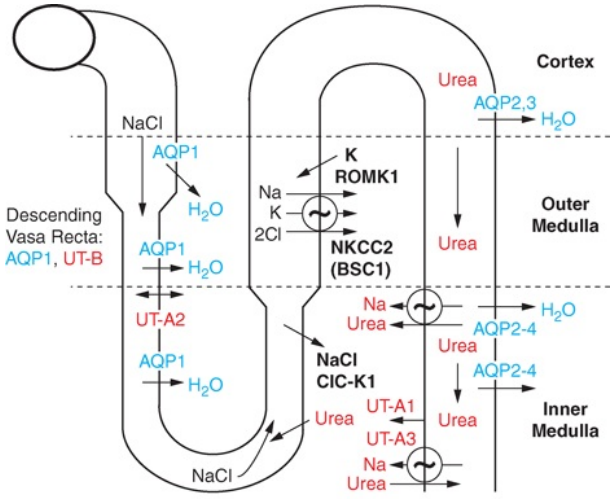

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The renal concentrating mechanism. Water, salt, and solute transport by both proximal and distal nephron segments participates in the renal concentrating mechanism (see text for details). Diagram showing the location of the major transport proteins involved; a loop of Henle is depicted on the left, a collecting duct on the right. UT, urea transporter; AQP, aquaporin; NKCC2, Na-K-2Cl cotransporter; ROMK, renal outer medullary K+ channel; CLC-K1, chloride channel. *(From JM Sands: J Am Soc Nephrol 13:2795, 2002; with permission.)*

AVP-induced, PKA-dependent phosphorylation of the aquaporin-2 water channel in principal cells stimulates the insertion of active water channels into the lumen of the collecting duct, resulting in transepithelial water absorption down the medullary osmotic gradient (**Fig. 45-3**). Under antidiuretic conditions, with increased circulating AVP, the kidney reabsorbs water filtered by the glomerulus, equilibrating the osmolality across the collecting duct epithelium to excrete a hypertonic, concentrated urine (osmolality of up to 1200 mosmol/kg). In the absence of circulating AVP, insertion of aquaporin-2 channels and water absorption across the collecting duct are essentially abolished, resulting in secretion of a hypotonic, dilute urine (osmolality as low as 30–50 mosmol/kg). Abnormalities in this final common pathway are involved in most disorders of water homeostasis, e.g., a reduced or absent insertion of active aquaporin-2 water channels into the membrane of principal cells in diabetes insipidus.

Figure 45-3

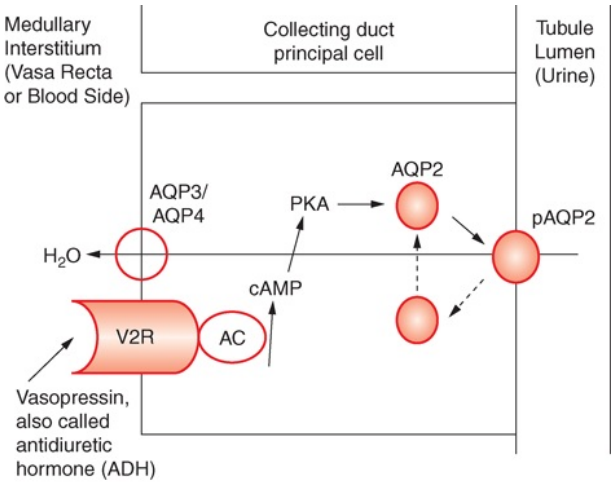

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vasopressin and the regulation of water permeability in the renal collecting duct. Vasopressin binds to the type 2 vasopressin receptor (V2R) on the basolateral membrane of principal cells, activates adenylyl cyclase (AC), increases intracellular cyclic adenosine monophosphatase (cAMP), and stimulates protein kinase A (PKA) activity. Cytoplasmic vesicles carrying aquaporin-2 (AQP) water channel proteins are inserted into the luminal membrane in response to vasopressin, increasing the water permeability of this membrane. When vasopressin stimulation ends, water channels are retrieved by an endocytic process and water permeability returns to its low basal rate. The AQP3 and AQP4 water channels are expressed on the basolateral membrane and complete the transcellular pathway for water reabsorption. pAQP2 = phosphorylated aquaporin-2. *(From JM Sands, DG Bichet: Ann Intern Med 144:186, 2006; with permission.)*

MAINTENANCE OF ARTERIAL CIRCULATORY INTEGRITY

Sodium is actively pumped out of cells by the Na^+, K^+-ATPase membrane pump. In consequence, 85–90% of body Na^+ is extracellular, and the extracellular fluid volume (ECFV) is a function of total-body Na^+ content. Arterial perfusion and circulatory integrity are, in turn, determined by renal Na^+ retention or excretion, in addition to the modulation of systemic arterial resistance. Within the kidney, Na^+ is filtered by the glomeruli and then sequentially reabsorbed by the renal tubules. The Na^+ cation typically is reabsorbed with the chloride anion (Cl^-); thus, chloride homeostasis also affects the ECFV. On a quantitative level, at a glomerular filtration rate (GFR) of 180 L/d and serum Na^+ of ~140 m*M*, the kidney filters some 25,200 mmol/d of Na^+. This is equivalent to ~1.5 kg of salt, which would occupy roughly 10 times the extracellular space; 99.6% of filtered Na^+-Cl^- must be reabsorbed to excrete 100 m*M* per day. Minute changes in renal Na^+-Cl^- excretion will thus have significant effects on the ECFV, leading to edema syndromes or hypovolemia.

Approximately two-thirds of filtered Na^+-Cl^- is reabsorbed by the renal proximal tubule via both paracellular and transcellular mechanisms. The TALH subsequently reabsorbs another 25–30% of filtered Na^+-Cl^- via the apical, furosemide-sensitive Na^+-K^+-$2Cl^-$ cotransporter. The adjacent aldosterone-sensitive distal nephron, which encompasses the distal convoluted tubule (DCT), connecting tubule (CNT), and collecting duct, accomplishes the "fine-tuning" of renal Na^+-Cl^- excretion. The thiazide-sensitive apical Na^+-Cl^- cotransporter (NCC) reabsorbs 5–10% of filtered Na^+-Cl^- in the DCT. Principal cells in the CNT and CD reabsorb Na^+ via electrogenic, amiloride-sensitive epithelial Na^+ channels (ENaC); Cl^- ions are reabsorbed primarily by adjacent intercalated cells via apical Cl^- exchange (Cl^--OH^- and Cl^--HCO_3^- exchange, mediated by the SLC26A4 anion exchanger) (**Fig. 45-4**).

Figure 45-4

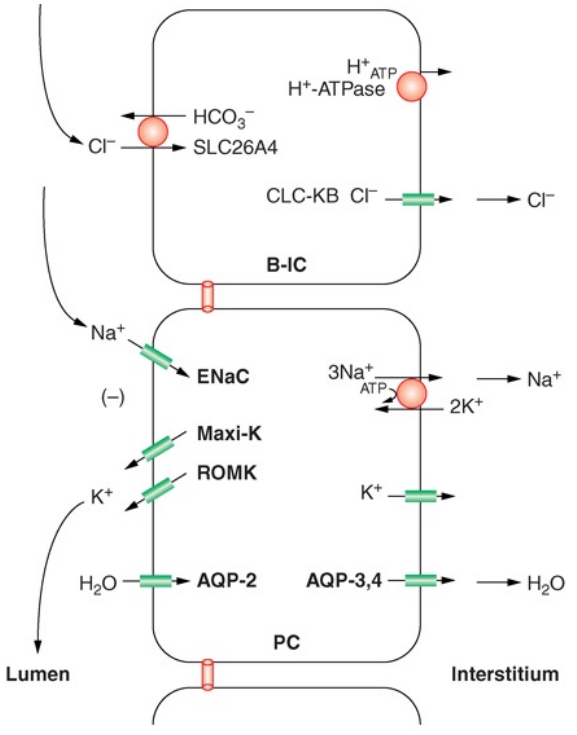

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sodium, water, and potassium transport in principal cells (PC) and adjacent β-intercalated cells (B-IC). The absorption of Na^+ via the amiloride-sensitive epithelial sodium channel (ENaC) generates a lumen-negative potential difference that drives K^+ excretion through the apical secretory K^+ channel ROMK (renal outer medullary K^+ channel) and/or the flow-dependent maxi-K channel. Transepithelial Cl^- transport occurs in adjacent β-intercalated cells via apical Cl^--HCO_3^- and Cl^--OH^- exchange (SLC26A4 anion exchanger, also known as pendrin) basolateral CLC chloride channels. Water is absorbed down the osmotic gradient by principal cells, through the apical aquaporin-2 (AQP-2) and basolateral aquaporin-3 and aquaporin-4 (Fig. 45-3).

Renal tubular reabsorption of filtered Na^+-Cl^- is regulated by multiple circulating and paracrine hormones in addition to the

activity of renal nerves. Angiotensin II activates proximal Na^+-Cl^- reabsorption, as do adrenergic receptors under the influence of renal sympathetic innervation; locally generated dopamine, in contrast, has a *natriuretic* effect. Aldosterone primarily activates Na^+-Cl^- reabsorption within the aldosterone-sensitive distal nephron. In particular, aldosterone activates the ENaC channel in principal cells, inducing Na^+ absorption and promoting K^+ excretion (see Fig. 45-4).

Figure 45-5.

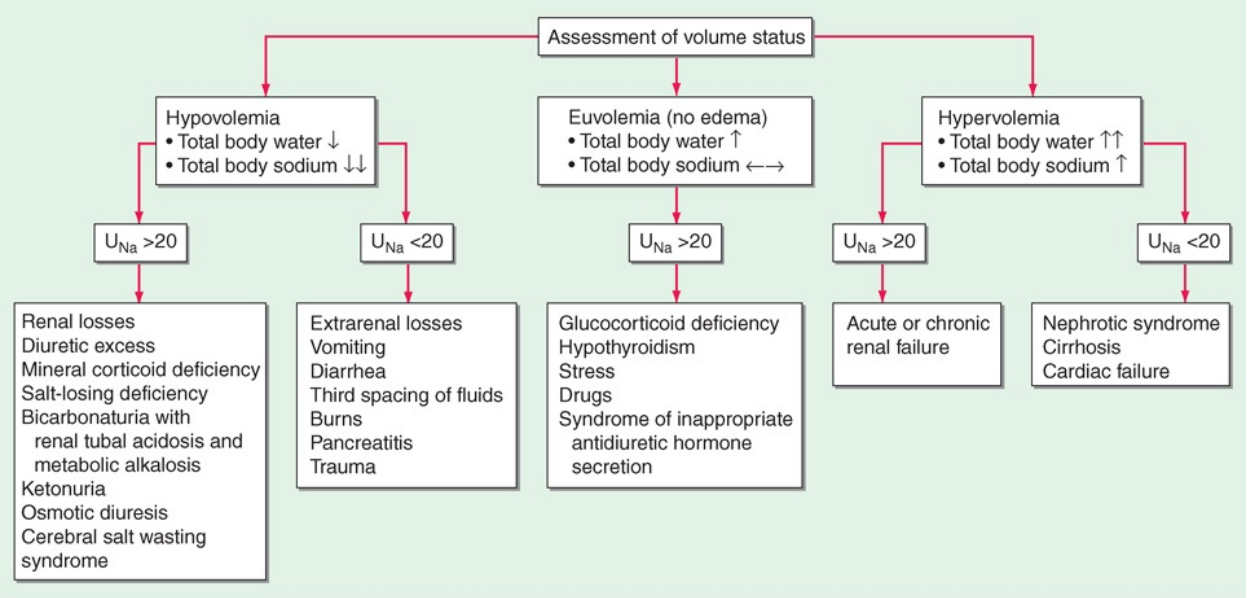

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The diagnostic approach to hyponatremia. *[From S Kumar, T Berl: Diseases of water metabolism, in Atlas of Diseases of the Kidney, RW Schrier (ed). Philadelphia, Current Medicine, Inc, 1999; with permission.]*

Circulatory integrity is critical for the perfusion and function of vital organs. Underfilling of the arterial circulation is sensed by ventricular and vascular pressure receptors, resulting in a neurohumoral activation (increased sympathetic tone, activation of the renin-angiotensin-aldosterone axis, and increased circulating AVP) that synergistically increases renal Na^+-Cl^- reabsorption, vascular resistance, and renal water reabsorption. This occurs in the context of decreased cardiac output, as occurs in hypovolemic states, low-output cardiac failure, decreased oncotic pressure, and/or increased capillary permeability. Alternatively, excessive arterial vasodilation results in *relative* arterial underfilling, leading to neurohumoral activation in the defense of tissue perfusion. These physiologic responses play important roles in many of the disorders discussed in this chapter. In particular, it is important to appreciate that AVP functions in the defense of circulatory integrity, inducing vasoconstriction, increasing sympathetic nervous system tone, increasing renal retention of both water and Na^+-Cl^-, and modulating the arterial baroreceptor reflex. Most of these responses involve activation of systemic V_{1A} AVP receptors, but concomitant activation of V_2 receptors in the kidney can result in renal water retention and hyponatremia.

Hypovolemia

ETIOLOGY

True volume depletion, or hypovolemia, generally refers to a state of combined salt and water loss that leads to contraction of the ECFV. The loss of salt and water may be renal or nonrenal in origin.

Renal Causes

Excessive urinary Na^+-Cl^- and water loss is a feature of several conditions. A high filtered load of endogenous solutes, such as glucose and urea, can impair tubular reabsorption of Na^+-Cl^- and water, leading to an osmotic diuresis. Exogenous mannitol, which often is used to decrease intracerebral pressure, is filtered by glomeruli but not reabsorbed by the proximal tubule, thus causing an osmotic diuresis. Pharmacologic diuretics selectively impair Na^+-Cl^- reabsorption at specific sites along the nephron, leading to increased urinary Na^+-Cl^- excretion. Other drugs can induce natriuresis as a side effect. For example, acetazolamide can inhibit proximal tubular Na^+-Cl^- absorption through its inhibition of carbonic anhydrase; other drugs, such as the antibiotics trimethroprim and pentamidine, inhibit distal tubular Na^+ reabsorption through the amiloride-sensitive ENaC channel, leading to urinary Na^+-Cl^- loss. Hereditary defects in renal transport proteins also are associated with reduced reabsorption of filtered Na^+-Cl^- and/or water. Alternatively, mineralocorticoid deficiency, mineralocorticoid resistance, or inhibition of the mineralocorticoid receptor (MLR) can reduce Na^+-Cl^- reabsorption by the aldosterone-sensitive distal nephron. Finally, tubulointerstitial injury, as occurs in interstitial nephritis, acute tubular injury, or obstructive uropathy, can reduce distal tubular Na^+-Cl^- and/or water absorption.

Excessive excretion of free water, i.e., water without electrolytes, also can lead to hypovolemia. However, the effect on ECFV is usually less marked in light of the fact that two-thirds of the water volume is lost from the ICF. Excessive renal water excretion occurs in the setting of decreased circulating AVP or renal resistance to AVP (central and nephrogenic diabetes insipidus, respectively).

Extrarenal Causes

Nonrenal causes of hypovolemia include fluid loss from the gastrointestinal tract, skin, and respiratory system. Accumulations of fluid within specific tissue compartments—typically the interstitium, peritoneum, or gastrointestinal tract—also can cause hypovolemia.

Approximately 9 L of fluid enters the gastrointestinal tract daily, 2 L by ingestion and 7 L by secretion; almost 98% of this volume is absorbed so that daily fecal fluid loss is only 100–200 mL. Impaired gastrointestinal reabsorption or enhanced secretion of fluid can cause hypovolemia. Since gastric secretions have a low pH (high H^+ concentration), whereas biliary, pancreatic, and intestinal secretions are alkaline (high HCO_3^- concentration), vomiting and diarrhea often are accompanied by metabolic alkalosis and acidosis, respectively.

Evaporation of water from the skin and respiratory tract (so-called insensible losses) is the major route for loss of solute-free water, which is typically 500–650 mL/d in healthy adults. This evaporative loss can increase during febrile illness or prolonged heat exposure. Hyperventilation also can increase insensible losses via the respiratory tract, particularly in ventilated patients; the humidity of inspired air is another determining factor. In addition, increased exertion and/or ambient temperature will increase insensible losses via sweat, which is hypotonic to plasma. Profuse sweating without adequate repletion of water and Na^+-Cl^- thus can lead to both hypovolemia and hypertonicity. Alternatively, replacement of these insensible losses with a surfeit of free water without adequate replacement of electrolytes may lead to hypovolemic hyponatremia.

Excessive fluid accumulation in interstitial and/or peritoneal spaces also can cause intravascular hypovolemia. Increases in vascular permeability and/or a reduction in oncotic pressure (hypoalbuminemia) alter Starling forces, resulting in excessive "third spacing" of the ECFV. This occurs in sepsis syndrome, burns, pancreatitis, nutritional hypoalbuminemia, and peritonitis. Alternatively, distributive hypovolemia can result from accumulation of fluid within specific compartments, for example, within the bowel lumen in gastrointestinal obstruction or ileus. Hypovolemia also can occur after extracorporeal hemorrhage or after significant hemorrhage into an expandable space, for example, the retroperitoneum.

DIAGNOSTIC EVALUATION

A careful history usually determines the etiologic cause of hypovolemia. Symptoms of hypovolemia are nonspecific and include fatigue, weakness, thirst, and postural dizziness; more severe symptoms and signs include oliguria, cyanosis, abdominal and chest pain, and confusion or obtundation. Associated electrolyte disorders may cause additional symptoms, for example, muscle weakness in patients with hypokalemia. On examination, diminished skin turgor and dry oral mucous membranes are less than ideal markers of a decreased ECFV in adult patients; reliable signs of hypovolemia include a decreased jugular venous pressure (JVP), orthostatic tachycardia (an increase of >15–20 beats per minute upon standing), and orthostatic hypotension (a >10–20-mmHg drop in blood pressure on standing). More severe fluid loss leads to hypovolemic shock, with hypotension, tachycardia, peripheral vasoconstriction, and peripheral hypoperfusion; these patients may exhibit peripheral cyanosis, cold extremities, oliguria, and altered mental status.

Routine chemistries may reveal an increase in blood urea nitrogen (BUN) and creatinine, reflecting a decrease in GFR. Creatinine is the more dependable measure of GFR, since BUN levels may be influenced by an increase in tubular reabsorption (prerenal azotemia), an increase in urea generation in catabolic states, hyperalimentation, or gastrointestinal bleeding and/or a decreased urea generation in decreased protein intake. In hypovolemic shock, liver function tests and cardiac biomarkers may show evidence of hepatic and cardiac ischemia, respectively. Routine chemistries and/or blood gases may reveal evidence of acid-base disorders. For example, bicarbonate loss due to diarrheal illness is a very common cause of metabolic acidosis; alternatively, patients with severe hypovolemic shock may develop lactic acidosis with an elevated anion gap.

The neurohumoral response to hypovolemia stimulates an increase in renal tubular Na^+ and water reabsorption. Therefore, the urine Na^+ concentration is typically <20 m*M* in nonrenal causes of hypovolemia, with a urine osmolality of >450 mosmol/kg. The reduction in both GFR and distal tubular Na^+ delivery may cause a defect in renal potassium excretion, with an increase in plasma K^+ concentration. Of note, patients with hypovolemia and a hypochloremic alkalosis due to vomiting, diarrhea, or diuretics typically have a urine Na^+ concentration >20 m*M* and urine pH >7.0 due to the increase in filtered HCO_3^-; the urine Cl^- concentration in this setting is a more accurate indicator of volume status, with a level <25 m*M* suggestive of hypovolemia. The urine Na^+ concentration is often >20 m*M* in patients with *renal* causes of hypovolemia, such as acute tubular necrosis; similarly, patients with diabetes insipidus will have an inappropriately dilute urine.

TREATMENT: HYPOVOLEMIA

The therapeutic goals in hypovolemia are to restore normovolemia and replace ongoing fluid losses. Mild hypovolemia usually can be treated with oral hydration and resumption of a normal maintenance diet. More severe hypovolemia requires intravenous hydration, with the choice of solution tailored to the underlying pathophysiology. Isotonic, "normal" saline (0.9% NaCl, 154 m*M* Na^+) is the most appropriate resuscitation fluid for normonatremic or hyponatremic patients with severe hypovolemia; colloid solutions such as intravenous albumin are not demonstrably superior for this purpose. Hypernatremic patients should receive a hypotonic solution: 5% dextrose if there has been only water loss (as in diabetes insipidus) or hypotonic saline (1/2 or 1/4 normal saline) if there has been water and Na^+-Cl^- loss. Patients with bicarbonate loss and metabolic acidosis, as occurs frequently in diarrhea, should receive intravenous bicarbonate, either an isotonic solution (150 meq of Na^+-HCO_3^- in 5% dextrose) or a more hypotonic bicarbonate solution in dextrose or dilute saline. Patients with severe hemorrhage or anemia should receive red cell transfusions without increasing the hematocrit beyond 35%.

SODIUM DISORDERS

Disorders of serum Na^+ concentration are caused by abnormalities in water homeostasis that lead to changes in the relative ratio of Na^+ to body water. Water intake and circulating AVP constitute the two key effectors in the defense of serum osmolality; defects in one or both of these defense mechanisms cause most cases of hyponatremia and hypernatremia. In contrast, abnormalities in sodium homeostasis per se lead to a deficit or surplus of whole-body Na^+-Cl^- content, a key determinant of the ECFV and circulatory integrity. Notably, volume status also modulates the release of AVP by the posterior pituitary so that hypovolemia is associated with higher circulating levels of the hormone at each level of serum osmolality. Similarly, in hypervolemic causes of arterial underfilling, e.g., heart failure and cirrhosis, the associated

neurohumoral activation is associated with an increase in circulating AVP, leading to water retention and hyponatremia. Therefore, a key concept in sodium disorders is that the absolute plasma Na^+ concentration tells one nothing about the volume status of a specific patient; this must be taken into account in the diagnostic and therapeutic approach.

Hyponatremia

Hyponatremia, which is defined as a plasma Na^+ concentration <135 m*M*, is a very common disorder, occurring in up to 22% of hospitalized patients. This disorder is almost always the result of an increase in circulating AVP and/or increased renal sensitivity to AVP, combined with any intake of free water; a notable exception is hyponatremia due to low solute intake (see below). The underlying pathophysiology for the exaggerated or inappropriate AVP response differs in patients with hyponatremia as a function of their ECFV. Hyponatremia thus is subdivided diagnostically into three groups, depending on clinical history and volume status: hypovolemic, euvolemic, and hypervolemic (**Fig. 45-5**).

HYPOVOLEMIC HYPONATREMIA

Hypovolemia causes a marked neurohumoral activation, increasing circulating levels of AVP. The increase in circulating AVP helps preserve blood pressure via vascular and baroreceptor V_{1A} receptors and increases water reabsorption via renal V_2 receptors; activation of V_2 receptors can lead to hyponatremia in the setting of increased free-water intake. Nonrenal causes of hypovolemic hyponatremia include gastrointestinal (GI) loss (vomiting, diarrhea, tube drainage, etc.) and insensible loss (sweating, burns) of Na^+-Cl^- and water in the absence of adequate oral replacement; urine Na^+ concentration is typically <20 m*M*. Notably, these patients may be clinically classified as euvolemic, with only the reduced urinary Na^+ concentration to indicate the cause of their hyponatremia. Indeed, a urine Na^+ concentration <20 m*M* in the absence of a cause of hypervolemic hyponatremia predicts a rapid increase in plasma Na^+ concentration in response to intravenous normal saline; saline induces a water diuresis in this setting, as circulating AVP levels plummet.

The *renal* causes of hypovolemic hyponatremia share an inappropriate loss of Na^+-Cl^- in the urine, leading to volume depletion and an increase in circulating AVP; urine Na^+ concentration is typically >20 m*M* (Fig. 45-5). A deficiency in circulating aldosterone and/or its renal effects can lead to hyponatremia in primary adrenal insufficiency and other causes of hypoaldosteronism; hyperkalemia and hyponatremia in a hypotensive and/or hypovolemic patient with high urine Na^+ concentration (much >20 m*M*) should strongly suggest this diagnosis. Salt-losing nephropathies may lead to hyponatremia when sodium intake is reduced due to impaired renal tubular function; typical causes include reflux nephropathy, interstitial nephropathies, post-obstructive uropathy, medullary cystic disease, and the recovery phase of acute tubular necrosis. Thiazide diuretics cause hyponatremia via a number of mechanisms, including polydipsia and diuretic-induced volume depletion. Notably, thiazides do not inhibit the renal concentrating mechanism so that circulating AVP has a maximal effect on renal water retention. In contrast, loop diuretics, which are associated less frequently with hyponatremia, inhibit Na^+-Cl^- and K^+ absorption by the TALH, blunting the countercurrent mechanism and reducing the ability to concentrate the urine. Increased excretion of an osmotically active nonreabsorbable or poorly reabsorbable solute also can lead to volume depletion and hyponatremia; important causes include glycosuria, ketonuria (e.g., in starvation or in diabetic or alcoholic ketoacidosis), and bicarbonaturia (e.g., in renal tubular acidosis or metabolic alkalosis, in which the associated bicarbonaturia leads to loss of Na^+).

Finally, the syndrome of "cerebral salt wasting" is a rare cause of hypovolemic hyponatremia, encompassing hyponatremia with clinical hypovolemia and inappropriate natriuresis in association with intracranial disease; associated disorders include subarachnoid hemorrhage, traumatic brain injury, craniotomy, encephalitis, and meningitis. Distinction from the more common syndrome of inappropriate antidiuresis (SIAD) is critical, since cerebral salt wasting typically responds to aggressive Na^+-Cl^- repletion.

HYPERVOLEMIC HYPONATREMIA

Patients with hypervolemic hyponatremia develop an increase in total body Na^+-Cl^- that is accompanied by a proportionately *greater* increase in total body water, leading to a reduced plasma Na^+ concentration. As in hypovolemic hyponatremia, the causative disorders can be separated by the effect on urine Na^+ concentration, with acute or chronic renal failure uniquely associated with an increase in urine Na^+ concentration (Fig. 45-5). The pathophysiology of hyponatremia in the sodium-avid edematous disorders [congestive heart failure (CHF), cirrhosis, and nephrotic syndrome] is similar to that in hypovolemic hyponatremia except that arterial filling and circulatory integrity are decreased due to the specific etiologic factors, e.g., cardiac dysfunction in CHF and peripheral vasodilation in cirrhosis. Urine Na^+ concentration is typically very low, i.e., <10 m*M*, even after hydration with normal saline; this Na^+-avid state may be obscured by diuretic therapy. The degree of hyponatremia provides an indirect index of the associated neurohumoral activation and is an important prognostic indicator in hypervolemic hyponatremia.

EUVOLEMIC HYPONATREMIA

Euvolemic hyponatremia can occur in moderate to severe hypothyroidism, with correction after the achievement of a euthyroid state. Severe hyponatremia also can be a consequence of secondary adrenal insufficiency due to pituitary disease; whereas the deficit in circulating aldosterone in primary adrenal insufficiency causes *hypovolemic* hyponatremia, the predominant glucocorticoid deficiency in secondary adrenal failure is associated with *euvolemic* hyponatremia. Glucocorticoids exert a negative feedback on AVP release by the posterior pituitary so that hydrocortisone replacement in these patients will rapidly normalize the AVP response to osmolality, reducing circulating AVP.

The syndrome of inappropriate antidiuresis is the most common cause of euvolemic hyponatremia (**Table 45-1**). The generation of hyponatremia in SIAD requires an intake of free water, with persistent intake at serum osmolalities that are lower than the usual threshold for thirst; as one would expect, the osmotic threshold and osmotic response curves for the sensation of thirst are shifted downward in patients with SIAD. Four distinct patterns of AVP secretion have been recognized in patients with SIAD, independent for the most part of the underlying cause. Unregulated, erratic AVP secretion is seen in about a third of patients, with no obvious correlation between serum osmolality and circulating AVP levels. Other patients fail to suppress AVP secretion at lower serum osmolalities, with a normal response curve to hyperosmolar conditions; others have a reset osmostat, with a lower threshold osmolality and a left-shifted osmotic response curve. The fourth subset consists of patients who have essentially no detectable circulating AVP, suggesting either a gain in function in renal water reabsorption or a circulating antidiuretic substance that is distinct from AVP. Gain-in-function mutations of a

single specific residue in the V_2 vasopressin receptor have been described in some of these patients, leading to constitutive activation of the receptor in the absence of AVP and a nephrogenic subset of SIAD.

Table 45-1 Causes of the Syndrome of Inappropriate Antidiuresis

Malignant Diseases	Pulmonary Disorders	Disorders of the Central Nervous System	Drugs	Other Causes
Carcinoma	Infections	Infection	Drugs that stimulate release of AVP or enhance its action	Hereditary (gain-of-function mutations in the vasopressin V_2 receptor)
Lung	Bacterial pneumonia	Encephalitis	Chlorpropamide	Idiopathic
Small cell	Viral pneumonia	Meningitis	SSRIs	Transient
Mesothelioma	Pulmonary abscess	Brain abscess	Tricyclic antidepressants	Endurance exercise
Oropharynx	Tuberculosis	Rocky Mountain spotted fever	Clofibrate	General anesthesia
Gastrointestinal tract	Aspergillosis	AIDS	Carbamazepine	Nausea
Stomach	Asthma	Bleeding and masses	Vincristine	Pain
Duodenum	Cystic fibrosis	Subdural hematoma	Nicotine	Stress
Pancreas	Respiratory failure associated with positive-pressure breathing	Subarachnoid hemorrhage	Narcotics	
Genitourinary tract		Cerebrovascular accident	Antipsychotic drugs	
Ureter		Brain tumors	Ifosfamide	
Bladder		Head trauma	Cyclophosphamide	
Prostate		Hydrocephalus	Nonsteroidal anti-inflammatory drugs	
Endometrium		Cavernous sinus thrombosis	MDMA (ecstasy)	
Endocrine thymoma		Other	AVP analogues	
Lymphomas		Multiple sclerosis	Desmopressin	
Sarcomas		Guillain-Barré syndrome	Oxytocin	
Ewing's sarcoma		Shy-Drager syndrome	Vasopressin	
		Delerium tremens		
		Acute intermittent polyphyria		

Abbreviations: AIDS, acquired immunodeficiency syndrome; AVP, vasopressin; MDMA, 3,4-methylenedioxymethamphetamine (ecstasy); SSRI, selective serotonin reuptake inhibitor.

Source: From Ellison and Berl.

Strictly speaking, patients with SIAD are not euvolemic but are subclinically volume expanded due to AVP-induced water and Na^+-Cl^- retention; vasopressin escape mechanisms invoked by sustained increases in AVP serve to limit distal renal tubular transport, preserving a modestly hypervolemic steady state. Serum uric acid is often low (<4 mg/dL) in patients with SIAD, consistent with suppressed proximal tubular transport in the setting of increased distal tubular Na^+-Cl^- and water transport; in contrast, patients with hypovolemic hyponatremia are often hyperuricemic due to a shared activation of proximal tubular Na^+-Cl^- and urate transport.

Common causes of SIAD include pulmonary disease (pneumonia, tuberculosis, pleural effusion, etc.) and central nervous system (CNS) diseases (tumor, subarachnoid hemorrhage, meningitis, etc.). SIAD also occurs with malignancies, most commonly with small cell lung carcinoma (75% of cases of malignancy-associated SIAD); ~10% of patients with this tumor will have a plasma Na^+ concentration <130 m*M* at presentation. SIAD is also a common complication of certain drugs, most commonly the selective serotonin reuptake inhibitors (SSRIs). Other drugs can potentiate the renal effect of AVP without exerting direct effects on circulating AVP levels (Table 45-1).

LOW SOLUTE INTAKE AND HYPONATREMIA

Hyponatremia occasionally can occur in patients with a very low intake of dietary solutes. Classically, this occurs in alcoholics whose sole nutrient is beer, hence the diagnostic label "beer potomania"; beer is very low in protein and salt content, containing only1–2 millimole per liter of Na^+. The syndrome also has been described in nonalcoholic patients with highly restricted solute intake due to nutrient-restricted diets, e.g., extreme vegetarian diets. Patients with hyponatremia due to low solute intake typically present with a very low urine osmolality, <100–200 mosmol/kg, with a urine Na^+ concentration that is <10–20 m*M*. The fundamental abnormality is the inadequate dietary intake of solutes; the reduced urinary solute excretion limits water excretion so that hyponatremia ensues after relatively modest polydipsia. The ability to excrete a free-water load is thus a function of urinary solute excretion; at a urine osmolality of 80 mosmol/kg, free-water clearance is 2.7 L daily for a solute excretion of 300 mosmol/d, 5.4 L daily at 600 mosmol/d, and 8.1 L at 900 mosmol/d. AVP levels have not been reported in patients with beer potomania but are expected to be suppressed or rapidly suppressible with saline hydration; this fits with the overly rapid correction in plasma Na^+ concentration that can be seen with saline hydration. Resumption of a normal diet and/or saline hydration also will correct the causative deficit in urinary solute excretion so that patients with beer potomania typically correct their plasma Na^+ concentration promptly after

admission to the hospital.

CLINICAL FEATURES OF HYPONATREMIA

Hyponatremia induces generalized cellular swelling, a consequence of water movement down the osmotic gradient from the hypotonic ECF to the ICF. The symptoms of hyponatremia are primarily neurologic, reflecting the development of cerebral edema within a rigid skull. The initial CNS response to acute hyponatremia is an increase in interstitial pressure, leading to shunting of ECF and solutes from the interstitial space into the cerebrospinal fluid and then into the systemic circulation. This is accompanied by an efflux of the major intracellular ions, Na^+, K^+, and Cl^-, from brain cells. Acute hyponatremic encephalopathy ensues when these volume regulatory mechanisms are overwhelmed by a rapid decrease in tonicity, resulting in acute cerebral edema. Early symptoms can include nausea, headache, and vomiting. However, severe complications can evolve rapidly, including seizure activity, brainstem herniation, coma, and death. A key complication of acute hyponatremia is normocapnic or hypercapnic respiratory failure; the associated hypoxemia may amplify the neurologic injury. Normocapnic respiratory failure in this setting typically is due to noncardiogenic, neurogenic pulmonary edema, with a normal pulmonary capillary wedge pressure.

Acute symptomatic hyponatremia is a medical emergency that occurs in a number of specific settings (**Table 45-2**). Women, particularly before menopause, are much more likely to develop encephalopathy and severe neurologic sequelae. Acute hyponatremia often has an iatrogenic component, e.g., when hypotonic intravenous fluids are given to postoperative patients with an increase in circulating AVP. Exercise-associated hyponatremia, an important clinical issue at marathons and other endurance events, similarly has been linked to both a nonosmotic increase in circulating AVP and excessive free-water intake. The recreational drug ecstasy (MDMA, 3,4-methylenedioxymethamphetamine) causes a rapid and potent induction of both thirst and AVP, leading to severe acute hyponatremia.

Table 45-2 Causes of Acute Hyponatremia

Iatrogenic
Postoperative: premenopausal women
Hypotonic fluids with cause of ↑ vasopressin
Glycine irrigation: TURP, uterine surgery
Colonoscopy preparation
Recent institution of thiazides
Polydipsia
MDMA ingestion
Exercise-induced
Multifactorial, e.g., thiazide and polydipsia

Abbreviations: MDMA, 3,4-methylenedioxymethamphetamine (ecstasy); TURP, transurethral resection of the prostate.

Persistent, chronic hyponatremia results in an efflux of organic osmolytes (creatine, betaine, glutamate, *myo*-inositol, and taurine) from brain cells; this response reduces intracellular osmolality and the osmotic gradient, favoring water entry. This reduction in intracellular osmolytes is largely complete within 48 h, the time period that clinically defines chronic hyponatremia; this temporal definition has considerable relevance for the treatment of hyponatremia (see below). The cellular response to chronic hyponatremia does not fully protect patients from symptoms, which can include vomiting, nausea, confusion, and seizures, usually at a plasma Na^+ concentration <125 m*M*. Even patients who are judged asymptomatic can manifest subtle gait and cognitive defects that reverse with correction of hyponatremia; notably, chronic asymptomatic hyponatremia increases the risk of falls. Chronic hyponatremia also increases the risk of bony fractures owing to the associated neurologic dysfunction and to a hyponatremia-associated reduction in bone density. Therefore, every attempt should be made to correct plasma Na^+ concentration safely in patients with chronic hyponatremia, even in the absence of overt symptoms (see the section on treatment of hyponatremia, below).

The management of chronic hyponatremia is complicated significantly by the asymmetry of the cellular response to correction of plasma Na^+ concentration. Specifically, the *reaccumulation* of organic osmolytes by brain cells is attenuated and delayed as osmolality increases after correction of hyponatremia, sometimes resulting in degenerative loss of oligodendrocytes and an osmotic demyelination syndrome (ODS). Overly rapid correction of hyponatremia (>8–10 m*M* in 24 h or 18 m*M* in 48 h) also is associated with a disruption in integrity of the blood-brain barrier, allowing the entry of immune mediators that may contribute to demyelination. The lesions of ODS classically affect the pons, a structure in which the delay in the reaccumulation of osmotic osmolytes is particularly pronounced; clinically, patients with central pontine myelinolysis can present one or more days after overcorrection of hyponatremia with para- or quadraparesis, dysphagia, dysarthria, diplopia, a "locked-in syndrome," and/or loss of consciousness. Other regions of the brain also can be involved in ODS, most commonly in association with lesions of the pons but occasionally in isolation; in order of frequency, the lesions of extrapontine myelinolysis can occur in the cerebellum, lateral geniculate body, thalamus, putamen, and cerebral cortex or subcortex. The clinical presentation of ODS therefore can vary as a function of the extent and localization of extrapontine myelinolysis, with the reported development of ataxia, mutism, parkinsonism, dystonia, and catatonia. Relowering of plasma Na^+ concentration after overly rapid correction can prevent or attenuate ODS (see the section on treatment of hyponatremia, below). However, even appropriately slow correction can be associated with ODS, particularly in patients with additional risk factors; these factors include alcoholism, malnutrition, hypokalemia, and liver transplantation.

DIAGNOSTIC EVALUATION OF HYPONATREMIA

Clinical assessment of hyponatremic patients should focus on the underlying cause; a detailed drug history is particularly crucial (Table 45-1). A careful clinical assessment of volume status is obligatory for the classical diagnostic approach to hyponatremia (Fig. 45-5). Hyponatremia is frequently multifactorial, particularly when severe; clinical evaluation should consider *all* the possible causes for excessive circulating AVP, including volume status, drugs, and the presence of nausea and/or pain. Radiologic imaging also may be appropriate to assess whether patients have a pulmonary or CNS cause for hyponatremia. A screening chest x-ray may fail to detect a small cell carcinoma of the lung; CT scanning of the thorax should be considered in patients at high risk for this tumor, e.g., patients with a history of smoking.

Laboratory investigation should include a measurement of serum osmolality to exclude pseudohyponatremia, which is

defined as the coexistence of hyponatremia with a normal or increased plasma tonicity. Most clinical laboratories measure plasma Na^+ concentration by testing diluted samples with automated ion-sensitive electrodes, correcting for this dilution by assuming that plasma is 93% water; this correction factor can be inaccurate in patients with pseudohyponatremia due to extreme hyperlipidemia and/or hyperproteinemia, in whom serum lipid or protein makes up a greater percentage of plasma volume. The measured osmolality also should be converted to the effective osmolality (tonicity) by subtracting the measured concentration of urea (divided by 2.8 if in mg/dL); patients with hyponatremia have an effective osmolality <275 mosmol/kg.

Elevated BUN and creatinine in routine chemistries also can indicate renal dysfunction as a potential cause of hyponatremia, whereas hyperkalemia may suggest adrenal insufficiency or hypoaldosteronism. Serum glucose also should be measured; plasma Na^+ concentration falls by ~1.6 to 2.4 m*M* for every 100-mg/dL increase in glucose due to glucose-induced water efflux from cells; this "true" hyponatremia resolves after correction of hyperglycemia. Measurement of serum uric acid also should be performed; whereas patients with SIAD-type physiology typically will be hypouricemic (serum uric acid <4 mg/dL), volume-depleted patients often will be hyperuricemic. In the appropriate clinical setting, thyroid, adrenal, and pituitary function should also be tested; hypothyroidism and secondary adrenal failure due to pituitary insufficiency are important causes of euvolemic hyponatremia, whereas primary adrenal failure causes hypovolemic hyponatremia. A cosyntropin stimulation test is necessary to assess for primary adrenal insufficiency.

Urine electrolytes and osmolality are crucial tests in the initial evaluation of hyponatremia. A urine Na^+ concentration <20–30 m*M* is consistent with hypovolemic hyponatremia in the clinical absence of a hypervolemic, Na^+-avid syndrome such as CHF (Fig. 45-5). In contrast, patients with SIAD typically excrete urine with a Na^+ concentration that is >30 m*M*. However, there can be substantial overlap in urine Na^+ concentration values in patients with SIAD and hypovolemic hyponatremia, particularly in the elderly; the ultimate "gold standard" for the diagnosis of hypovolemic hyponatremia is the demonstration that plasma Na^+ concentration corrects after hydration with normal saline. Patients with thiazide-associated hyponatremia also may present with a higher than expected urine Na^+ concentration and other findings suggestive of SIAD; one should defer making a diagnosis of SIAD in these patients until 1–2 weeks after discontinuation of the thiazide. A urine osmolality <100 mosmol/kg is suggestive of polydipsia; urine osmolality >400 mosmol/kg indicates that AVP excess is playing a more dominant role, whereas intermediate values are more consistent with multifactorial pathophysiology (e.g., AVP excess with a significant component of polydipsia). Patients with hyponatremia due to decreased solute intake (beer potomania) typically have urine Na^+ concentration <20 m*M* and urine osmolality in the range of <100 to the low 200's.

Finally, the measurement of urine K^+ concentration is required to calculate the urine:plasma electrolyte ratio, which is useful to predict the response to fluid restriction (see the section on treatment of hyponatremia, below).

TREATMENT: HYPONATREMIA

Three major considerations guide therapy for hyponatremia. First, the presence and/or severity of symptoms determine the urgency and goals of therapy. Patients with acute hyponatremia (Table 45-2) present with symptoms that can range from headache, nausea, and/or vomiting to seizures, obtundation, and central herniation; patients with chronic hyponatremia that is present for >48 h are less likely to have severe symptoms. Second, patients with chronic hyponatremia are at risk for ODS if plasma Na^+ concentration is corrected by >8–10 m*M* within the first 24 h and/or by >18 m*M* within the first 48 h. Third, the response to interventions such as hypertonic saline, isotonic saline, and vasopressin antagonists can be highly unpredictable, and so frequent monitoring of plasma Na^+ concentration during corrective therapy is imperative.

Once the urgency in correcting the plasma Na^+ concentration has been established and appropriate therapy instituted, the focus should be on treatment or withdrawal of the underlying cause. Patients with euvolemic hyponatremia due to SIAD, hypothyroidism, or secondary adrenal failure will respond to successful treatment of the underlying cause, with an increase in plasma Na^+ concentration. However, not all causes of SIAD are immediately reversible, necessitating pharmacologic therapy to increase the plasma Na^+ concentration (see below). Hypovolemic hyponatremia will respond to intravenous hydration with isotonic normal saline, with a rapid reduction in circulating AVP and a brisk water diuresis; it may be necessary to reduce the rate of correction if the history suggests that hyponatremia has been chronic, i.e., present for more than48 h (see below). Hypervolemic hyponatremia due to congestive heart failure often responds to improved therapy of the underlying cardiomyopathy, e.g., after the institution or intensification of angiotensin-converting enzyme (ACE) inhibition. Finally, patients with hyponatremia due to beer potomania and low solute intake respond very rapidly to intravenous saline and the resumption of a normal diet. Notably, patients with beer potomania have a very high risk of developing ODS due to the associated hypokalemia, alcoholism, and malnutrition and the high risk of overcorrecting the plasma Na^+ concentration.

Water deprivation has long been a cornerstone of therapy for chronic hyponatremia. However, patients who are excreting minimal electrolyte-free water will require aggressive fluid restriction; this can be very difficult for patients with SIAD to tolerate because their thirst is also inappropriately stimulated. The urine:plasma electrolyte ratio (urinary $[Na^+]+[K^+]$/plasma $[Na^+]$) can be exploited as a quick indicator of electrolyte-free water excretion (**Table 45-3**); patients with a ratio >1 should be restricted more aggressively (<500 mL/d), those with a ratio ~1 should be restricted to 500–700 mL/d, and those with a ratio <1 should be restricted to <1 L/d. In hypokalemic patients, potassium replacement will serve to increase plasma Na^+ concentration in light of the fact that the plasma Na^+ concentration is a function of both exchangeable Na^+ and exchangeable K^+ divided by total body water; a corollary is that aggressive repletion of K^+ has the potential to overcorrect the plasma Na^+ concentration even in the absence of hypertonic saline. Plasma Na^+ concentration also tends to respond to an increase in dietary solute intake, which increases the ability to excrete free water; however, the use of oral urea and/or salt tablets for this purpose is generally not practical or well tolerated.

Table 45-3 Management of Hypernatremia

Water deficit
1. Estimate total-body water (TBW): 50% of body weight in women and 60% in men
2. Calculate free-water deficit: {([Na^+]-140)/140} x TBW

3. Administer deficit over 48–72 h, without increasing the plasma Na^+ concentration by >10 m*M*/24 h
Ongoing water losses
4. Calculate electrolyte-free water clearance, C_eH_2O: $C_eH_2O = \frac{V(1 - U_{Na} + U_K)}{P_{Na}}$ where V is urinary volume, U_{Na} is urinary [Na^+], U_K is urinary [K^+], and P_{Na} is plasma [Na^+]
Insensible losses
5. ~10 mL/kg per day: less if ventilated, more if febrile
Total
6. Add components to determine water deficit and ongoing water loss; correct the water deficit over 48–72 h and replace daily water loss. Avoid correction of plasma [Na^+] by >10 m*M/d*

Patients in whom therapy with fluid restriction, potassium replacement, and/or increased solute intake fails may require pharmacologic therapy to increase their plasma Na^+ concentration. Many patients with SIAD respond to combined therapy with oral furosemide, 20 mg twice a day (higher doses may be necessary in renal insufficiency), and oral salt tablets; furosemide serves to inhibit the renal countercurrent mechanism and blunt urinary concentrating ability, whereas the salt tablets counteract diuretic-associated natriuresis. Demeclocycline is a potent inhibitor of principal cells and can be utilized in patients whose Na+ levels do not increase in response to furosemide and salt tablets. However, this agent can be associated with a reduction in GFR due to excessive natriuresis and/or direct renal toxicity; it should be avoided in cirrhotic patients in particular, who are at higher risk of nephrotoxicity due to drug accumulation.

Vasopressin antagonists (vaptans) are highly effective in treating SIAD and hypervolemic hyponatremia due to heart failure or cirrhosis, reliably increasing plasma Na^+ concentration as a result of their aquaretic effects (augmentation of free-water clearance). Most of these agents specifically antagonize theV_2 vasopressin receptor; tolvaptan is currently the only oral V_2 antagonist approved by the U.S. Food and Drug Administration. Conivaptan, the only available intravenous vaptan, is a mixed V_{1A}/V_2 antagonist with a modest risk of hypotension due to V_{1A} receptor inhibition. Therapy with vaptans must be initiated in a hospital setting, with a liberalization of fluid restriction (>2 L/d) and close monitoring of plasma Na^+ concentration. Although these agents are approved for the management of all but hypovolemic hyponatremia and acute hyponatremia, the clinical indications for them are not completely clear. Oral tolvaptan is perhaps most appropriate for the management of significant and persistent SIAD (e.g., in small cell lung carcinoma) that has not responded to water restriction and/or oral furosemide and salt tablets.

Treatment of acute symptomatic hyponatremia should include hypertonic 3% saline (513 m*M*) to acutely increase plasma Na^+ concentration by 1–2 m*M*/h to a total of 4–6 m*M*; this modest increase is typically sufficient to alleviate severe acute symptoms, after which corrective guidelines for "chronic" hyponatremia are appropriate (see below). A number of equations have been developed to estimate the required rate of hypertonic saline. The traditional approach is to calculate a Na^+ deficit, in which the Na^+ deficit = 0.6 x body weight x (target plasma Na^+ concentration – starting plasma Na^+ concentration), followed by a calculation of the required rate. Regardless of the method used to determine the rate of administration, the increase in plasma Na^+ concentration can be highly unpredictable during treatment with hypertonic saline due to rapid changes in the underlying physiology; plasma Na^+ concentration should be monitored every 2–4 h during treatment, with appropriate changes in therapy based on the observed rate of change. The administration of supplemental oxygen and ventilatory support is also critical in the management of patients with acute hyponatremia who develop acute pulmonary edema or hypercapnic respiratory failure. Intravenous loop diuretics will help treat acute pulmonary edema and also increase free-water excretion by interfering with the renal countercurrent multiplication system. Vasopressin antagonists do *not* have an approved role in the management of acute hyponatremia.

The rate of correction should be comparatively slow in *chronic* hyponatremia (<8–10 m*M* in the first 24 h and <18 m*M* in the first 48 h) to avoid ODS. Overcorrection of the plasma Na^+ concentration can occur when AVP levels rapidly normalize, for example, after treatment of patients with chronic hypovolemic hyponatremia with intravenous saline or after glucocorticoid replacement in patients with hypopituitarism and secondary adrenal failure. Approximately 10% of patients treated with vaptans will overcorrect; the risk is increased if water intake is not liberalized. If the plasma Na^+ concentration overcorrects after therapy—whether with hypertonic saline, isotonic saline, or a vaptan—hyponatremia can be reinduced safely or stabilized by the administration of the vasopressin *agonist* desmopressin acetate (DDAVP) and/or the administration of free water, typically intravenous D5W; the goal is to prevent or reverse the development of ODS.

Hypernatremia

ETIOLOGY

Hypernatremia is defined as an increase in the plasma Na^+ concentration to >145 m*M*. Considerably less common than hyponatremia, hypernatremia nonetheless is associated with mortality rates as high as 40–60%, mostly due to the severity of the associated underlying disease processes. Hypernatremia is usually the result of a combined water and electrolyte deficit, with losses of H^2O in excess of those of Na^+. Less frequently, the ingestion or iatrogenic administration of excess Na^+ can be causative, for example, after IV administration of excessive hypertonic Na^+-Cl^- or Na^+-HCO_3^- (**Fig. 45-6**).

Figure 45-6

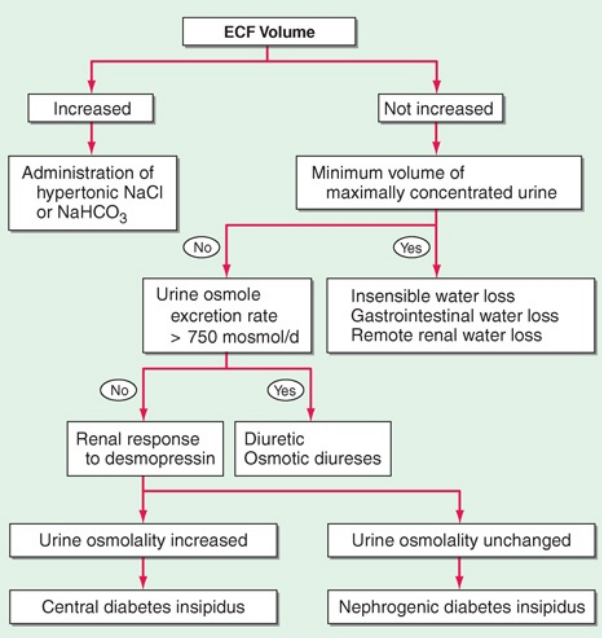

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The diagnostic approach to hypernatremia. ECF, extracellular fluid.

Elderly individuals with reduced thirst and/or diminished access to fluids are at the highest risk of developing hypernatremia. Patients with hypernatremia may rarely have a central defect in hypothalamic osmoreceptor function, with a mixture of both decreased thirst and reduced AVP secretion. Causes of this adipsic diabetes insipidus include primary or metastatic tumor, occlusion or ligation of the anterior communicating artery, trauma, hydrocephalus, and inflammation.

Hypernatremia can develop after the loss of water via both renal and nonrenal routes. Insensible losses of water may increase in the setting of fever, exercise, heat exposure, severe burns, or mechanical ventilation. Diarrhea is the most common gastrointestinal cause of hypernatremia. Notably, osmotic diarrhea and viral gastroenteritis typically generate stools with Na^+ and K^+ <100 m*M*, thus leading to water loss and hypernatremia; in contrast, secretory diarrhea typically results in isotonic stool and thus hypovolemia with or without hypovolemic hyponatremia.

Common causes of renal water loss include osmotic diuresis secondary to hyperglycemia, excess urea, postobstructive diuresis, and mannitol; these disorders share an increase in urinary solute excretion and urinary osmolality (see "Diagnostic Approach," below). Hypernatremia due to a water diuresis occurs in central or nephrogenic diabetes insipidus (DI).

Nephrogenic DI (NDI) is characterized by renal resistance to AVP, which can be partial or complete (see "Diagnostic Approach," below). Genetic causes include loss-of-function mutations in the X-linked V_2 receptor; mutations in the AVP-responsive aquaporin-2 water channel can cause autosomal recessive and autosomal dominant nephrogenic DI, whereas recessive deficiency of the aquaporin-1 water channel causes a more modest concentrating defect (Fig. 45-2). Hypercalcemia also can cause polyuria and NDI; calcium signals directly through the calcium-sensing receptor to downregulate Na^+, K^+, and Cl^- transport by the TALH and water transport in principal cells, thus reducing renal concentrating ability in hypercalcemia. Another common acquired cause of NDI is hypokalemia, which inhibits the renal response to AVP and downregulates aquaporin-2 expression. Several drugs can cause acquired NDI, in particular lithium, ifosfamide, and several antiviral agents. Lithium causes NDI by multiple mechanisms, including direct inhibition of renal glycogen synthase kinase-3 (GSK3), a kinase thought to be the pharmacologic target of lithium in bipolar disease; GSK3 is required for the response of principal cells to AVP. The entry of lithium through the amiloride-sensitive Na^+ channel ENaC (Fig. 45-4) is required for the effect of the drug on principal cells; thus, combined therapy with lithium and amiloride can mitigate lithium-associated NDI. However, lithium causes chronic tubulointerstitial scarring and chronic kidney disease after prolonged therapy so that patients may have a persistent NDI long after stopping the drug, with a reduced therapeutic benefit from amiloride.

Finally, gestational diabetes insipidus is a rare complication of late-term pregnancy in which increased activity of a circulating placental protease with vasopressinase activity leads to reduced circulating AVP and polyuria, often accompanied by hypernatremia. DDAVP is an effective therapy for this syndrome because of its resistance to the vasopressinase enzyme.

CLINICAL FEATURES

Hypernatremia increases osmolality of the ECF, generating an osmotic gradient between the ECF and the ICF, an efflux of intracellular water, and cellular shrinkage. As in hyponatremia, the symptoms of hypernatremia are predominantly neurologic. Altered mental status is the most common manifestation, ranging from mild confusion and lethargy to deep coma. The sudden shrinkage of brain cells in acute hypernatremia may lead to parenchymal or subarachnoid hemorrhages and/or subdural hematomas; however, these vascular complications are encountered primarily in pediatric and neonatal patients. Osmotic damage to muscle membranes also can lead to hypernatremic rhabdomyolysis. Brain cells accommodate to a chronic increase in ECF osmolality (>48 h) by activating membrane transporters that mediate influx and intracellular accumulation of organic osmolytes (creatine, betaine, glutamate, *myo*-inositol, and taurine); this results in an increase in ICF water and normalization of brain parenchymal volume. In consequence, patients with *chronic* hypernatremia are less likely to develop severe neurologic compromise. However, the cellular response to chronic hypernatremia predisposes these patients to the development of cerebral edema and seizures during overly rapid hydration (overcorrection of plasma Na^+ concentration by >10 *mM*/d).

DIAGNOSTIC APPROACH

The history should focus on the presence or absence of thirst, polyuria, and/or an extrarenal source for water loss, such as diarrhea. The physical examination should include a detailed neurologic exam and an assessment of the ECFV; patients with a particularly large water deficit and/or a combined deficit in electrolytes and water may be hypovolemic, with reduced JVP and orthostasis. Accurate documentation of daily fluid intake and daily urine output is also critical for the diagnosis and management of hypernatremia.

Laboratory investigation should include a measurement of serum and urine osmolality in addition to urine electrolytes. The appropriate response to hypernatremia and a serum osmolality >295 mosmol/kg is an increase in circulating AVP and the excretion of low volumes (<500 mL/d) of maximally concentrated urine, i.e., urine with osmolality >800 mosmol/kg; if this is the case, an extrarenal source of water loss is primarily responsible for the generation of hypernatremia. Many patients with hypernatremia are polyuric; if an osmotic diuresis is responsible, with excessive excretion of Na^+-Cl^-, glucose, and/or urea, solute excretion will be >750–1000 mosmol/d (>15 mosmol/kg body water per day) (Fig. 45-6). More commonly, patients with hypernatremia and polyuria will have a predominant water diuresis, with excessive excretion of hypotonic, dilute urine.

Adequate differentiation between nephrogenic and central causes of DI requires the measurement of the response in urinary osmolality to DDAVP, combined with measurement of circulating AVP in the setting of hypertonicity. By definition, patients with baseline hypernatremia are hypertonic, with an adequate stimulus for AVP by the posterior pituitary.

Therefore, in contrast to polyuric patients with a normal or reduced baseline plasma Na^+ concentration and osmolality, a water deprivation test (Chap. 44) is unnecessary in hypernatremia; indeed, water deprivation is absolutely contraindicated in this setting because of the risk for worsening the hypernatremia. Patients with NDI will fail to respond to DDAVP, with a urine osmolality that increases by <50% or <150 mosmol/kg from baseline, in combination with a normal or high circulating AVP level; patients with central DI will respond to DDAVP, with a reduced circulating AVP. Patients may exhibit a partial response to DDAVP, with a >50% rise in urine osmolality that nonetheless fails to reach 800 mosmol/kg; the level of circulating AVP will help differentiate the underlying cause, i.e., nephrogenic versus central DI. In pregnant patients, AVP assays should be drawn in tubes containing the protease inhibitor 1,10-phenanthroline to prevent in vitro degradation of AVP by placental vasopressinase.

For patients with hypernatremia due to renal loss of water it is critical to quantify *ongoing* daily losses using the calculated electrolyte-free water clearance in addition to calculation of the baseline water deficit (the relevant formulas are discussed in Table 45-3). This requires daily measurement of urine electrolytes, combined with accurate measurement of daily urine volume.

TREATMENT: HYPERNATREMIA

The underlying cause of hypernatremia should be withdrawn or corrected, whether it is drugs, hyperglycemia, hypercalcemia, hypokalemia, or diarrhea. The approach to the correction of hypernatremia is outlined in Table 45-3. It is imperative tocorrect hypernatremia slowly to avoid cerebral edema, typically replacing the calculated free-water deficit over 48 h. Notably, the plasma Na^+ concentration should be corrected by no more than 10 m*M*/d, which may take longer than 48 h in patients with severe hypernatremia (>160 m*M*). A rare exception is patients with acute hypernatremia (<48 h) due to sodium loading, who can safely be corrected rapidly at a rate of 1 m*M*/h.

Water ideally should be administered by mouth or by nasogastric tube as the most direct way to provide free water, i.e., water without electrolytes. Alternatively, patients can receive free water in dextrose-containing IV solutions such as 5% dextrose (D5W); blood glucose should be monitored to avoid hyperglycemia. Depending on the history, blood pressure, or clinical volume status, it may be appropriate to treat initially with hypotonic saline solutions (1/4 or 1/2 normal saline); normal saline is usually inappropriate in the absence of very severe hypernatremia, in which normal saline is proportionally more hypotonic relative to plasma, or frank hypotension. Calculation of urinary electrolyte-free water clearance (see Table 45-3) is required to estimate daily, ongoing loss of free water in patients with nephrogenic or central DI, which should be replenished daily.

Additional therapy may be feasible in specific cases. Patients with central DI should respond to the administration of intravenous, intranasal, or oral DDAVP. Patients with NDI due to lithium may reduce their polyuria with amiloride (2.5–10 mg/d), which decreases entry of lithium into principal cells by inhibiting ENaC (see above); in practice, however, most patients with lithium-associated DI are able to compensate for their polyuria simply by increasing their daily water intake. Thiazides may reduce polyuria due to NDI, ostensibly by inducing hypovolemia and increasing proximal tubular water reabsorption. Occasionally, nonsteroidal anti-inflammatory drugs (NSAIDs) have been used to treat polyuria associated with NDI, reducing the negative effect of intrarenal prostaglandins on urinary concentrating mechanisms; however, this creates the risk of NSAID-associated gastric and/or renal toxicity. Furthermore, it must be emphasized that thiazides, amiloride, and NSAIDs are appropriate only for *chronic* management of polyuria from NDI and have *no* role in the acute management of associated hypernatremia, in which the focus is on replacing free-water deficits and ongoing free-water loss.

POTASSIUM DISORDERS

Homeostatic mechanisms maintain plasma K^+ concentration between 3.5 and 5.0 m*M* despite marked variation in dietary K^+ intake. In a healthy individual at steady state, the entire daily intake of potassium is excreted, approximately 90% in the urine and 10% in the stool; the kidney thus plays a dominant role in potassium homeostasis. However, more than 98% of total-body potassium is intracellular, chiefly in muscle; buffering of extracellular K^+ by this large intracellular pool plays a crucial role in the regulation of plasma K^+ concentration. Changes in the exchange and distribution of intra- and extracellular K^+ thus can lead to marked hypo- or hyperkalemia. A corollary is that massive necrosis and the attendant release of tissue K^+ can cause severe hyperkalemia, particularly in the setting of acute kidney injury and reduced excretion of K^+.

Changes in whole-body K^+ content are mediated primarily by the kidney, which *reabsorbs* filtered K^+ in hypokalemic, K^+-deficient states and *secretes* K^+ in hyperkalemic, K^+-replete states. Although K^+ is transported along the entire nephron, it is the principal cells of the connecting tubule (CNT) and cortical collecting duct (CD) that play a dominant role in renal K^+ secretion, whereas alpha-intercalated cells of the outer medullary CD function in renal tubular reabsorption of filtered K^+ in

K^+-deficient states. In principal cells, apical Na^+ entry via the amiloride-sensitive ENaC generates a lumen-negative potential difference that drives passive K^+ exit through apical K^+ channels (Fig. 45-4). Two major K^+ channels mediate distal tubular K^+ secretion: the secretory K^+ channel ROMK (the renal outer medullary K^+ channel, also known as Kir1.1 or KcnJ1) and the flow-sensitive maxi-K K^+ channel (also known as the BK K+ channel). ROMK is thought to mediate the bulk of constitutive K^+ secretion, whereas increases in distal flow rate and/or genetic absence of ROMK activate K^+ secretion via the maxi-K channel.

An appreciation of the relationship between ENaC-dependent Na^+ entry and distal K^+ secretion (Fig. 45-4) is required for the bedside interpretation of potassium disorders. For example, decreased distal delivery of Na^+, as occurs in hypovolemic, prerenal states, tends to blunt the ability to excrete K^+, leading to hyperkalemia; in contrast, an *increase* in distal delivery of Na^+ and distal flow rate, as occurs after treatment with thiazide and loop diuretics, can enhance K^+ secretion and lead to hypokalemia. Hyperkalemia is also a predictable consequence of drugs that directly inhibit ENaC due to the role of this Na^+ channel in generating a lumen-negative potential difference. Aldosterone in turn has a major influence on potassium excretion, increasing the activity of ENaC channels and thus amplifying the driving force for K^+ secretion across the luminal membrane of principal cells. Abnormalities in the renin-angiotensin-aldosterone system thus can cause both hypokalemia and hyperkalemia. Notably, however, potassium excess and potassium restriction have opposing, aldosterone-independent effects on the density and activity of apical K^+ channels in the distal nephron; i.e., other factors modulate the renal capacity to secrete K^+. In addition, potassium restriction and hypokalemia activate aldosterone-independent distal *reabsorption* of filtered K^+, activating apical H^+/K^+-ATPase activity in intercalated cells within the outer medullary CD. Reflective perhaps of this physiology, changes in plasma K^+ concentration are not universal in disorders associated with changes in aldosterone activity.

Hypokalemia

Hypokalemia, defined as a plasma K^+ concentration <3.6 m*M*, occurs in up to 20% of hospitalized patients. Hypokalemia is associated with a tenfold increase in in-hospital mortality rates due to adverse effects on cardiac rhythm, blood pressure, and cardiovascular morbidity rate. Mechanistically, hypokalemia can be caused by redistribution of K^+ between tissues and the ECF or by renal and nonrenal loss of K^+ (**Table 45-4**). Systemic hypomagnesemia also can cause treatment-resistant hypokalemia due to a combination of reduced cellular uptake of K^+ and exaggerated renal secretion. Spurious hypokalemia or pseudohypokalemia occasionally can result from in vitro cellular uptake of K^+ after venipuncture, for example, due to profound leukocytosis in acute leukemia.

Table 45-4 Causes of Hypokalemia

I. Decreased intake
A. Starvation
B. Clay ingestion
II. Redistribution into cells
A. Acid-base
1. Metabolic alkalosis
B. Hormonal
1. Insulin
2. Increased β_2-adrenergic sympathetic activity: post-myocardial infarction, head injury
3. β_2-Adrenergic agonists: bronchodilators, tocolytics
4. α-Adrenergic antagonists
5. Thyrotoxic periodic paralysis
6. Downstream stimulation of Na^+/K^+-ATPase: theophylline, caffeine
C. Anabolic state
1. Vitamin B12 or folic acid administration (red blood cell production)
2. Granulocyte-macrophage colony-stimulating factor (white blood cell production)
3. Total parenteral nutrition
D. Other
1. Pseudohypokalemia
2. Hypothermia
3. Familial hypokalemic periodic paralysis
4. Barium toxicity: systemic inhibition of "leak" K^+ channels
III. Increased loss
A. Nonrenal
1. Gastrointestinal loss (diarrhea)
2. Integumentary loss (sweat)
B. Renal
1. Increased distal flow and distal Na^+ delivery: diuretics, osmotic diuresis, salt-wasting nephropathies

2. Increased secretion of potassium
a. Mineralocorticoid excess: primary hyperaldosteronism [aldosterone-producing adenomas (APAs)], primary or unilateral adrenal hyperplasia (PAH), idiopathic hyperaldosteronism (IHA) due to bilateral adrenal hyperplasia, and adrenal carcinoma], familial hyperaldosteronism (FH-I, FH-II, congenital adrenal hyperplasias), secondary hyperaldosteronism (malignanthypertension, renin-secreting tumors, renal arterystenosis, hypovolemia), Cushing's syndrome, Bartter's syndrome, Gitelman's syndrome
b. Apparent mineralocorticoid excess: genetic deficiency of 11β-dehydrogenase-2 (syndrome of apparent mineralocorticoid excess), inhibition of 11β-dehydrogenase-2 (glycyrrhetinic/glycyrrhizinic acid and/or carbenoxolone; licorice, food products, drugs), Liddle's syndrome [genetic activation of epithelial Na+ channels (ENaC)]
c. Distal delivery of nonreabsorbed anions: vomiting, nasogastric suction, proximal renal tubular acidosis, diabetic ketoacidosis, glue sniffing (toluene abuse), penicillin derivatives (penicillin, nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin)
3. Magnesium deficiency

REDISTRIBUTION AND HYPOKALEMIA

Insulin, β_2-adrenergic activity, and thyroid hormone promote Na^+,K^+-ATPase-mediated cellular uptake of K^+, leading to hypokalemia. Inhibition of passive *efflux* of K^+ also can cause hypokalemia, albeit rarely; this typically occurs in the setting of systemic inhibition of K^+ channels by toxic barium ions. Exogenous insulin can cause iatrogenic hypokalemia, particularly during the management of K^+-deficient states such as diabetic ketoacidosis. Alternatively, the stimulation of *endogenous* insulin can provoke hypokalemia, hypomagnesemia, and/or hypophosphatemia in malnourished patients who are given a carbohydrate load. Alterations in the activity of the endogenous sympathetic nervous system can cause hypokalemia in several settings, including alcohol withdrawal, hyperthyroidism, acute myocardial infarction, and severe head injury. β_2 agonists, including both bronchodilators and tocolytics (ritodrine), are powerful activators of cellular K^+ uptake; "hidden" sympathomimetics such as pseudoephedrine and ephedrine in cough syrup or dieting agents also may cause unexpected hypokalemia. Finally, xanthine-dependent activation of cyclic AMP–dependent signaling downstream of the β_2 receptor can lead to hypokalemia, usually in the setting of overdose (theophylline) or marked overingestion (dietary caffeine).

Redistributive hypokalemia also can occur in the setting of hyperthyroidism, with periodic attacks of hypokalemic paralysis[thyrotoxic periodic paralysis (TPP)]. Similar episodes of hypokalemic weakness in the absence of thyroid abnormalities occur in *familial* hypokalemic periodic paralysis, usually caused by missense mutations of voltage sensor domains within the α_1 subunit of L-type calcium channels or the skeletal Na^+ channel; these mutations generate an abnormal gating pore current activated by hyperpolarization. TPP develops more frequently in patients of Asian or Hispanic origin; this shared predisposition has been linked to genetic variation in Kir2.6, a muscle-specific, thyroid hormone –responsive K^+ channel. Patients typically present with weakness of the extremities and limb girdles, with paralytic episodes that occur most frequently between 1 and 6 a.m. Signs and symptoms of hyperthyroidism are not invariably present. Hypokalemia is usually profound and almost invariably is accompanied by hypophosphatemia and hypomagnesemia. The hypokalemia in TPP is attributed to both direct and indirect activation of Na^+,K^+-ATPase, resulting in increased uptake of K^+ by muscle and other tissues. Increases in β-adrenergic activity play an important role in that high-dose propranolol (3 mg/kg) rapidly reverses the associated hypokalemia, hypophosphatemia, and paralysis.

NONRENAL LOSS OF POTASSIUM

The loss of K^+ in sweat is typically low except under extremes of physical exertion. Direct gastric losses of K^+ due to vomiting or nasogastric suctioning are also minimal; however, the ensuing hypochloremic alkalosis results in persistent kaliuresis due to secondary hyperaldosteronism and bicarbonaturia, i.e., a *renal* loss of K^+. Intestinal loss of K^+ due to diarrhea is a globally important cause of hypokalemia in light of the worldwide prevalence of diarrheal disease. Noninfectious gastrointestinal processes such as celiac disease, ileostomy, villous adenomas, VIPomas, and chronic laxative abuse also can cause significant hypokalemia. Colonic pseudo-obstruction (Ogilvie's syndrome) can lead to hypokalemia from secretory diarrhea with an abnormally high potassium content caused by a marked activation of colonic K+ secrection.

RENAL LOSS OF POTASSIUM

Drugs can increase renal K^+ excretion by a variety of different mechanisms. Diuretics are a particularly common cause due to associated increases in distal tubular Na^+ delivery and distal tubular flow rate in addition to secondary hyperaldosteronism. Thiazides have an effect on plasma K^+ concentration greater than that of loop diuretics despite their lesser natriuretic effect. The higher propensity of thiazides to cause hypokalemia may be secondary to thiazide-associated hypocalciuria versus the *hypercalciuria* seen with loop diuretics; increases in downstream luminal calcium in response to loop diuretics will inhibit ENaC in principal cells, thus reducing the lumen-negative potential difference and attenuating distal K^+ excretion. High doses of penicillin-related antibiotics (nafcillin, dicloxacillin, ticarcillin, oxacillin, and carbenicillin) can increase obligatory K^+ excretion by acting as nonreabsorbable anions in the distal nephron. Finally, several renal tubular toxins cause renalK^+ and magnesium wasting, leading to hypokalemia andhypomagnesemia; these drugs include aminoglycosides, amphotericin, foscarnet, cisplatin, and ifosfamide (see also "Magnesium Deficiency and Hypokalemia," below).

Aldosterone activates the ENaC channel in principal cells via multiple synergistic mechanisms, thus increasing the driving force for K^+ excretion. In consequence, increases in aldosterone bioactivity and/or gains in function of aldosterone-dependent signaling pathways are associated with hypokalemia. Increases in circulating aldosterone (hyperaldosteronism) may be primary or secondary. Increased levels of circulating renin in secondary forms of hyperaldosteronism lead to increased angiotensin II (AT-II) and thus aldosterone; renal artery stenosis is perhaps the most common cause (Table 45-4). Primary hyperaldosteronism may be genetic or acquired. Hypertension and hypokalemia due to increases in circulating 11-deoxycorticosterone occur in patients with congenital adrenal hyperplasia caused by defects in either steroid 11β-hydroxylase or steroid 17α-hydroxylase; deficient 11β-hydroxylase results in associated virilization and other signs of

androgen excess, whereas reduced sex steroids in 17α-hydroxylase deficiency lead to hypogonadism. The two major forms of *isolated* primary hyperaldosteronism are familial hyperaldosteronism type I [FH-I, also known as glucocorticoid-remediable hyperaldosteronism (GRA)] and familial hyperaldosteronism type II (FH-II), in which aldosterone production is not repressible by exogenous glucocorticoids. FH-I is caused by a chimeric gene duplication between the homologous 11β-hydroxylase (*CYP11B1*) and aldosterone synthase (*CYP11B2*) genes, fusing the adrenocorticotropic hormone (ACTH)-responsive 11β-hydroxylase promoter to the coding region of aldosterone synthase; this chimeric gene is under the control of ACTH and thus is repressible by glucocorticoids.

Acquired causes of primary hyperaldosteronism include aldosterone-producing adenomas (APAs), primary or unilateral adrenal hyperplasia (PAH), idiopathic hyperaldosteronism (IHA) due to bilateral adrenal hyperplasia, and adrenal carcinoma; APA and IHA account for close to 60% and 40%, respectively, of diagnosed cases of hyperaldosteronism. Random testing of plasma renin activity (PRA) and aldosterone is a helpful screening tool in hypokalemic and/or hypertensive patients, with an aldosterone:PRA ratio >50 suggestive of primary hyperaldosteronism.

The glucocorticoid cortisol has affinity for the mineralocorticoid receptor (MLR) equal to that of aldosterone, with resultant mineralocorticoid-like activity. However, cells in the aldosterone-sensitive distal nephron are protected from this illicit activation by the enzyme 11β-hydroxysteroid dehydrogenase-2 (11βHSD-2), which converts cortisol to cortisone; cortisone has minimal affinity for the MLR. Recessive loss-of-function mutations in the *11βHSD-2* gene thus are associated with cortisol-dependent activation of the MLR and the syndrome of apparent mineralocorticoid excess (SAME), encompassing hypertension, hypokalemia, hypercalciuria, and metabolic alkalosis, with suppressed PRA and suppressed aldosterone. A similar syndrome is caused by biochemical inhibition of 11βHSD-2 by glycyrrhetinic/glycyrrhizinic acid and/or carbenoxolone. Glycyrrhizinic acid is a natural sweetener found in licorice root, typically encountered in licorice and its many guises or as a flavoring agent in tobacco and food products.

Finally, hypokalemia may occur with systemic increases in glucocorticoids. In Cushing's syndrome caused by increases in pituitary ACTH (Chap. 342) the incidence of hypokalemia is only 10%, whereas it is 60–100% in patients with ectopic secretion of ACTH despite a similar incidence of hypertension. Indirect evidence suggests that the activity of renal 11β HSD-2 is reduced in patients with ectopic ACTH compared with Cushing's syndrome, resulting in a syndrome of apparent mineralocorticoid excess.

Finally, defects in multiple renal tubular transport pathways are associated with hypokalemia. For example, loss-of-function mutations in subunits of the acidifying H^+-ATPAse in alpha-intercalated cells cause hypokalemic distal renal tubular acidosis, as do many acquired disorders of the distal nephron. Liddle's syndrome is caused by autosomal dominant gain-in-function mutations of ENaC subunits. Disease-associated mutations either activate the channel directly or abrogate aldosterone-inhibited retrieval of ENaC subunits from the plasma membrane; the end result is increased expression of activated ENaC channels at the plasma membrane of principal cells. Patients classically manifest severe hypertension with hypokalemia that is unresponsive to spironolactone yet sensitive to amiloride. Hypertension and hypokalemia are, however, variable aspects of the Liddle's phenotype; more consistent features include a blunted aldosterone response to ACTH and reduced urinary aldosterone excretion.

Loss of the transport functions of the TALH and DCT nephron segments causes two distinct subtypes of hereditary hypokalemic alkalosis, with TALH dysfunction causing Bartter's syndrome (BS) and DCT dysfunction causing Gitelman's syndrome (GS). Patients with "classic" BS typically have polyuria and polydipsia due to the reduction in renal concentrating ability. They may have an increase in urinary calcium excretion, and 20% are hypomagnesemic. Other features include marked activation of the renin-angiotensin-aldosterone axis. Patients with "antenatal" BS have a severe systemic disorder characterized by marked electrolyte wasting, polyhydramnios, and hypercalciuria with nephrocalcinosis; renal prostaglandin synthesis and excretion are increased significantly, accounting for many of the systemic symptoms. There are five disease genes for BS, all of which function in some aspect of regulated Na^+, K^+, and Cl^- transport by the TALH. Gitelman's syndrome is, in contrast, genetically homogeneous, caused almost exclusively by loss-of-function mutations in the thiazide-sensitive Na^+-Cl^- cotransporter of the DCT. Patients with GS are uniformly hypomagnesemic and exhibit marked hypocalciuria rather than the hypercalciuria typically seen in BS; urinary calcium excretion is thus a critical diagnostic test in GS. GS has a milder phenotype than BS; however, patients with GS may suffer from chondrocalcinosis, an abnormal deposition of calcium pyrophosphate dihydrate (CPPD) in joint cartilage (Chap. 284).

MAGNESIUM DEFICIENCY AND HYPOKALEMIA

Magnesium depletion has inhibitory effects on muscle Na^+,K^+-ATPase activity, reducing influx into muscle cells and causing a secondary kaliuresis. In addition, magnesium depletion causes exaggerated K^+ secretion by the distal nephron; this is attributed to a reduction in the magnesium-dependent, intracellular block of K^+ efflux through the secretory K^+ channel of principal cells (ROMK; Fig. 45-4). Regardless of the dominant mechanism(s), hypomagnesemic patients are clinically refractory to K^+ replacement in the absence of Mg^{2+} repletion. Notably, magnesium deficiency is also a common concomitant of hypokalemia, since many disorders of the distal nephron may cause both potassium and magnesium wasting (Chap. 284).

CLINICAL FEATURES

Hypokalemia has prominent effects on cardiac, skeletal, and intestinal muscle cells. In particular, it is a major risk factor for both ventricular and atrial arrhythmias. Hypokalemia predisposes to digoxin toxicity by a number of mechanisms, including reduced competition between K^+ and digoxin for shared binding sites on cardiac Na^+,K^+-ATPase subunits. Electrocardiographic changes in hypokalemia include broad flat T waves, ST depression, and QT prolongation; these are most marked when serum K^+ is <2.7 mmol/L. Hypokalemia also results in hyperpolarization of skeletal muscle, thus impairing the capacity to depolarize and contract; weakness and even paralysis may ensue. It also causes a skeletal myopathy and predisposes to rhabdomyolysis. Finally, the paralytic effects of hypokalemia on intestinal smooth muscle may cause intestinal ileus.

The functional effects of hypokalemia on the kidney include Na^+-Cl^- and HCO_3^- retention, polyuria, phosphaturia, hypocitraturia, and an activation of renal ammoniagenesis. Bicarbonate retention and other acid-base effects of hypokalemia can contribute to the generation of metabolic alkalosis. Hypokalemic polyuria is due to a combination of polydipsia and an AVP-resistant renal concentrating defect. Structural changes in the kidney due to hypokalemia include a relatively specific vacuolizing injury to proximal tubular cells, interstitial nephritis, and renal cysts. Hypokalemia also predisposes to acute kidney injury and can lead to end-stage renal disease in patients with long-standing hypokalemia due to eating disorders and/or laxative abuse.

Hypokalemia and/or reduced dietary K^+ are implicated in the pathophysiology and progression of hypertension, heart failure, and stroke. For example, short-term K^+ restriction in healthy humans and patients with essential hypertension induces Na^+-Cl^- retention and hypertension. Correction of hypokalemia is particularly important in hypertensive patients treated with diuretics, in whom blood pressure improves with the establishment of normokalemia.

DIAGNOSTIC APPROACH

The cause of hypokalemia is usually evident from history, physical examination, and/or basic laboratory tests. The history should focus on medications (e.g., laxatives, diuretics, antibiotics), diet and dietary habits (e.g., licorice), and/or symptoms that suggest a particular cause (e.g., periodic weakness, diarrhea). The physical examination should pay particular attention to blood pressure, volume status, and signs suggestive of specific hypokalemic disorders, e.g., hyperthyroidism and Cushing's syndrome. Initial laboratory evaluation should include electrolytes, BUN, creatinine, serum osmolality, Mg^{2+}, Ca^{2+}, a complete blood count, and urinary pH (**Fig. 45-7**). The presence of a non-anion-gap acidosis suggests a distal, hypokalemic renal tubular acidosis or diarrhea; calculation of the urinary anion gap can help differentiate these two diagnoses. Renal K^+ excretion can be assessed with a 24-h urine collection; a 24-h K^+ excretion of <15 m*M* is indicative of an extrarenal cause of hypokalemia(Fig. 45-7). Alternatively, serum and urine osmolality can be used to calculate the transtubular K^+ gradient (TTKG), which should be <3-4 in the presence of hypokalemia (see the section on hyperkalemia in this chapter). Urine Cl^- is usually decreased in patients with hypokalemia from a nonreabsorbable anion, such as antibiotics or HCO_3^-. Other causes of chronic, hypokalemic alkalosis are surreptitious vomiting, diuretic abuse, and GS. Hypokalemic patients with bulimia thus have a urinary Cl^- <10 mmol/L; urine Na^+, K^+, and Cl^- are persistently elevated in GS due to loss of function in the thiazide-sensitive Na^+-Cl^- cotransporter but less elevated in diuretic abuse and with greater variability. Urine diuretic screens for loop diuretics and thiazides may be necessary to further exclude diuretic abuse.

Figure 45-7

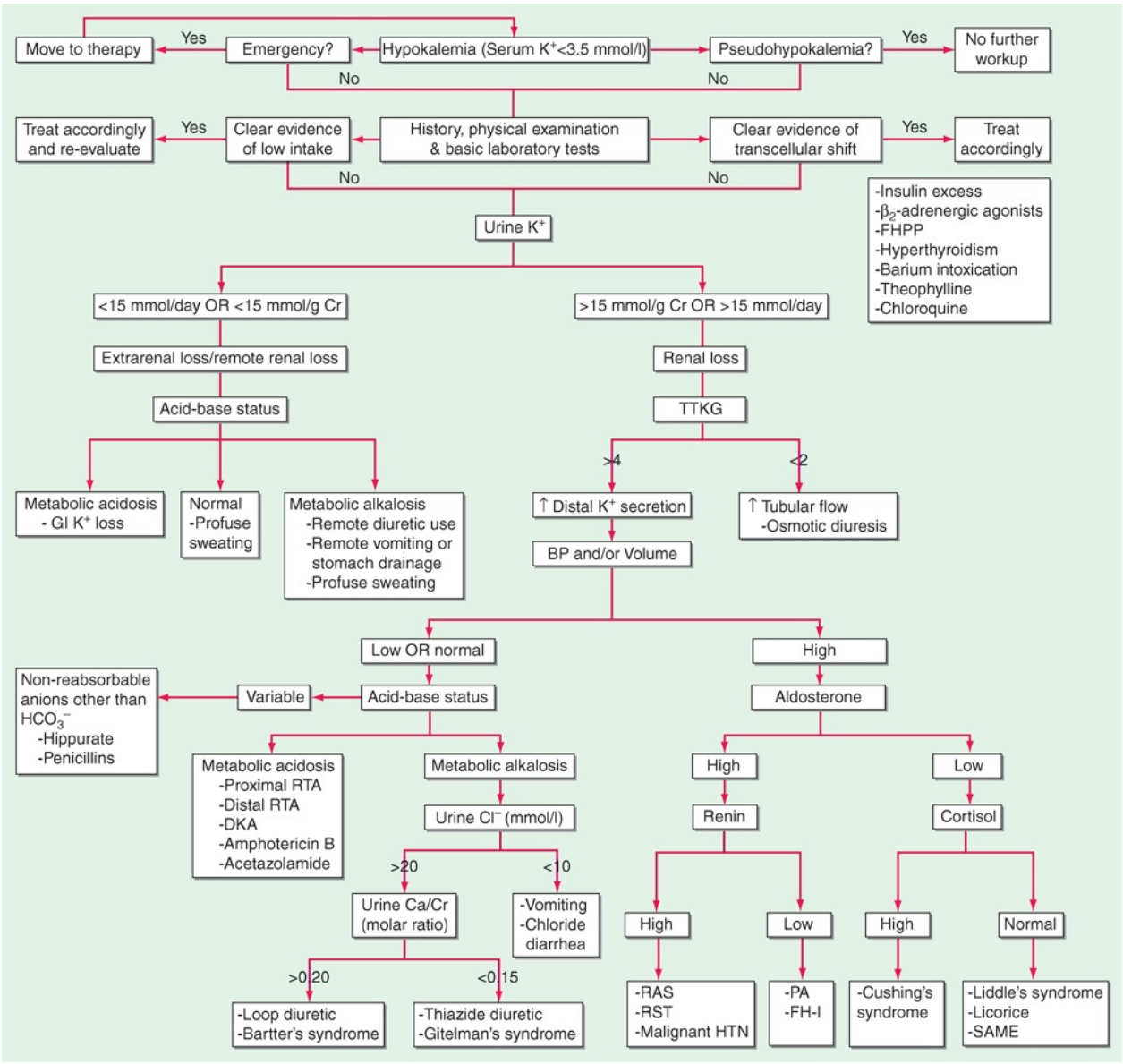

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The diagnostic approach to hypokalemia. See text for details. BP, blood pressure; DKA, diabetic ketoacidosis; FHPP, familial hypokalemic periodic paralysis; FH-I, familial hyperaldosteronism type I; GI, gastrointestinal; HTN, hypertension; PA, primary aldosteronism; RAS, renal artery stenosis; RST, renin-secreting tumor; RTA, renal tubular acidosis; SAME, syndrome of apparent mineralocorticoid excess; TTKG, transtubular potassium gradient. *(From Mount and Zandi-Nejad; with permission.)*

Other tests, such as urinary Ca^{2+}, thyroid function tests, and/or PRA and aldosterone levels, may be appropriate in specific cases. A plasma aldosterone:PRA ratio >50 is suggestive of hyperaldosteronism. Patients with hyperaldosteronism or apparent mineralocorticoid excess may require further testing, for example, adrenal vein sampling (Chap. 342) or the clinically available tests for specific genetic causes (FH-I, SAME, Liddle's syndrome, etc.). Patients with primary aldosteronism thus should be tested for the chimeric FH-I/GRA gene (see above) if they are younger than 20 years of age or have a family history of primary aldosteronism or stroke at a young age (<40 years). Preliminary differentiation of Liddle's syndrome due to mutant ENaC channels from SAME due to mutant 11βHSD-2 (see above)—both of which cause hypokalemia and hypertension with aldosterone suppression—can be made on a clinical basis; patients with Liddle's syndrome should respond to amiloride (ENaC inhibition) but not spironolactone, whereas patients with SAME will respond to spironolactone.

TREATMENT: HYPOKALEMIA

The goals of therapy for hypokalemia are to prevent life-threatening and/or chronic consequences, replace the associated K^+ deficit, and correct the underlying cause and/or mitigate future hypokalemia. The urgency of therapy depends on the severity of hypokalemia, associated clinical factors (cardiac disease, digoxin therapy, etc.), and the rate of decline in serum K^+. Urgent but cautious K^+ replacement should be considered in patients with severe redistributive hypokalemia (plasma K^+ concentration <2.5 m*M*) and/or when serious complications ensue; however, this creates a risk of rebound hyperkalemia after resolution of the underlying cause. When excessive activity of the sympathetic nervous system is thought to play a dominant role in redistributive hypokalemia, as in thyrotoxic periodic paralysis, high-dose propranolol (3 mg/kg) should be considered; this nonspecific β-adrenergic blocker will correct hypokalemia without the risk of rebound hyperkalemia.

Oral replacement with K^+-Cl^- is the mainstay of therapy for hypokalemia. Potassium phosphate, oral or IV, may be appropriate in patients with combined hypokalemia and hypophosphatemia. Potassium bicarbonate or potassium citrate should be considered in patients with concomitant metabolic acidosis. Notably, hypomagnesemic patients are refractory to K^+ replacement alone, and so concomitant Mg^{2+} deficiency should *always* be corrected with oral or intravenous repletion. The deficit of K^+ and the rate of correction should be estimated as accurately as possible; renal function, medications, and comorbid conditions such as diabetes should be considered to gauge the risk of overcorrection. In the absence of abnormal K^+ redistribution, the total deficit correlates with serum K^+ so that serum K^+ drops by approximately 0.27 m*M* for every 100-mmol reduction in total-body stores; loss of 400 to 800 mmol of total-body K^+ results in a reduction in serum K^+ of approximately 2.0 m*M*. However, because of the difficulty in assessing the deficit accurately, plasma K^+ concentration must be monitored carefully during repletion.

The use of intravenous administration should be limited to patients unable to utilize the enteral route or in the setting of severe complications (paralysis, arrhythmia, etc.). Intravenous K^+-Cl^- should always be administered in saline solutions rather than dextrose since the dextrose-induced increase in insulin can acutely exacerbate hypokalemia. The peripheral intravenous dose is usually 20–40 mmol of K^+-Cl^- per liter; higher concentrations can cause localized pain from chemical phlebitis, irritation, and sclerosis. If hypokalemia is severe (<2.5 mmol/L) and/or critically symptomatic, intravenous K^+-Cl^- can be administered through a central vein with cardiac monitoring in an intensive care setting at rates of 10–20 mmol/h; higher rates should be reserved for acutely life-threatening complications. The absolute amount of administered K^+ should be restricted (e.g., 20 mmol in 100 mL of saline solution) to prevent inadvertent infusion of a large dose. Femoral veins are preferable, since infusion through internal jugular or subclavian central lines can acutely increase the local concentration of K^+ and affect cardiac conduction.

Strategies to minimize K^+ losses also should be considered. These measures may include minimizing the dose of non-K^+-sparing diuretics, restricting Na^+ intake, and using clinically appropriate combinations of non-K^+-sparing and K^+-sparing medications (e.g., loop diuretics with ACE inhibitors).

Hyperkalemia

Hyperkalemia is defined as a plasma potassium level of 5.5 m*M*. It occurs in up to 10% of hospitalized patients; severe hyperkalemia (>6.0 m*M*) occurs in approximately 1%, with a significantly increased risk of mortality. Although redistribution and reduced tissue uptake can acutely cause hyperkalemia, a decrease in renal K^+ excretion is the most common underlying cause (**Table 45-5**). Excessive intake of K^+ is a rare cause because of the adaptive capacity to increase renal secretion; however, dietary intake can have a major effect in susceptible patients, e.g., diabetic patients with hyporeninemic hypoaldosteronism and chronic kidney disease. Drugs that have an impact on the renin-angiotensin-aldosterone axis are also a major cause of hyperkalemia.

Table 45-5 Causes of Hyperkalemia

I. "Pseudo" hyperkalemia
A. Cellular efflux: thrombocytosis, erythrocytosis, leukocytosis, in vitro hemolysis
B. Hereditary defects in red cell membrane transport
II. Intra- to extracellular shift
A. Acidosis
B. Hyperosmolality; radiocontrast, hypertonic dextrose, mannitol
C. β-adrenergic antagonists (noncardioselective agents)
D. Digoxin and related glycosides (yellow oleander, foxglove, bufadienolide)
E. Hyperkalemic periodic paralysis
F. Lysine, arginine, and ε-aminocaproic acid (structurally similar, positively charged)
G. Succinylcholine; thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization
H. Rapid tumor lysis
III. Inadequate excretion
A. Inhibition of the renin-angiotensin-aldosterone axis; ↑ risk of hyperkalemia when used in combination
1. Angiotensin-converting enzyme (ACE) inhibitors
2. Renin inhibitors: aliskiren [in combination with ACE-inhibitors or angiotensin receptor blockers (ARBs)]
3. ARBs
4. Blockade of the mineralocorticoid receptor: spironolactone, eplerenone, drospirenone
5. Blockade of ENaC: amiloride, triamterene, trimethoprim, pentamidine, nafamostat
B. Decreased distal delivery
1. Congestive heart failure
2. Volume depletion
C. Hyporeninemic hypoaldosteronism
1. Tubulointerstitial diseases: systemic lupus erythematosus (SLE), sickle cell anemia, obstructive uropathy
2. Diabetes, diabetic nephropathy
3. Drugs: nonsteroidal anti-inflammatory drugs, cyclooxygenase 2 (COX-2) inhibitors, beta blockers, cyclosporine, tacrolimus
4. Chronic kidney disease, advanced age
5. Pseudohypoaldosteronism type II: defects in WNK1 or WNK4 kinases
D. Renal resistance to mineralocorticoid
1. Tubulointerstitial diseases: SLE, amyloidosis, sickle cell anemia, obstructive uropathy, post-acute tubular necrosis

2. Hereditary: pseudohypoaldosteronism type I: defects in the mineralocorticoid receptor *or* ENaC
E. Advanced renal insufficiency
1. Chronic kidney disease
2. End-stage renal disease
3. Acute oliguric kidney injury
F. Primary adrenal insufficiency
1. Autoimmune: Addison's disease, polyglandular endocrinopathy
2. Infectious: HIV, cytomegalovirus, tuberculosis, disseminated fungal infection
3. Infiltrative: amyloidosis, malignancy, metastatic cancer
4. Drug-associated: heparin, low-molecular-weight heparin
5. Hereditary: adrenal hypoplasia congenita, congenital lipoid adrenal hyperplasia, aldosterone synthase deficiency
6. Adrenal hemorrhage or infarction, including in antiphospholipid syndrome

PSEUDOHYPERKALEMIA

Hyperkalemia should be distinguished from factitious hyperkalemia or pseudohyperkalemia, an artifactual increase in serum K^+ due to the release of K^+ during or after venipuncture. Pseudohyperkalemia can occur in the setting of excessive muscle activity during venipuncture (fist clenching, etc.), a marked increase in cellular elements (thrombocytosis, leukocytosis, and/or erythrocytosis) with in vitro efflux of K^+, and acute anxiety during venipuncture with respiratory alkalosis and redistributive hyperkalemia. Cooling of blood after venipuncture is another cause, due to reduced cellular uptake; the converse is the increased uptake of K^+ by cells at high ambient temperatures, leading to normal values for hyperkalemic patients and/or to spurious hypokalemia in normokalemic patients. Finally, there are multiple genetic subtypes of hereditary pseudohyperkalemia caused by increases in the passive K^+ permeability of erythrocytes. For example, causative mutations have been described in the red cell anion exchanger (AE1, encoded by the *SLC4A1* gene), leading to reduced red cell anion transport, hemolytic anemia, the acquisition of a novel AE1-mediated K^+ leak, and pseudohyperkalemia.

REDISTRIBUTION AND HYPERKALEMIA

Several different mechanisms can induce an efflux of intracellular K^+ and hyperkalemia. Hyperkalemia due to hypertonic mannitol, hypertonic saline, and intravenous immunoglobulin generally is attributed to a "solvent drag" effect as water moves out of cells along the osmotic gradient. Diabetic patients are also prone to osmotic hyperkalemia in response to intravenous hypertonic glucose when it is given without adequate insulin. Cationic amino acids—specifically lysine, arginine, and the structurally related drug ε-aminocaproic acid—cause efflux of K^+ and hyperkalemia through an effective cation-K^+ exchange of unknown identity and mechanism. Digoxin inhibits Na^+,K^+-ATPase and impairs the uptake of K^+ by skeletal muscle so that digoxin overdose predictably results in hyperkalemia. Structurally related glycosides are found in specific plants (yellow oleander, foxglove, etc.) and in the cane toad, *Bufo marinus* (bufadienolide); ingestion of these substances and extracts from them also can cause hyperkalemia. Finally, fluoride ions also inhibit Na^+,K^+-ATPase, and so fluoride poisoning is typically associated with hyperkalemia.

Succinylcholine depolarizes muscle cells, causing an efflux of K^+ through acetylcholine receptors (AChRs). The use of this agent is contraindicated in patients who have sustained thermal trauma, neuromuscular injury, disuse atrophy, mucositis, or prolonged immobilization. These disorders share a marked increase and redistribution of AChRs at the plasma membrane of muscle cells; depolarization of these upregulated AChRs by succinylcholine leads to an exaggerated efflux of K^+ through the receptor-associated cation channels, resulting in acute hyperkalemia.

HYPERKALEMIA DUE TO EXCESS INTAKE OR TISSUE NECROSIS

Increased intake of even small amounts of K^+ may provoke severe hyperkalemia in patients with predisposing factors; hence, an assessment of dietary intake is crucial. Foods rich in potassium include tomatoes, bananas, and citrus fruits; occult sources of K^+, particularly K^+-containing salt substitutes, also may contribute significantly. Iatrogenic causes include simple overreplacement with K^+-Cl^- and the administration of a potassium-containing medication (e.g., K^+-penicillin) to a susceptible patient. Red cell transfusion is a well-described cause of hyperkalemia, typically in the setting of massive transfusions. Finally, tissue necrosis, as in acute tumor lysis syndrome and rhabdomyolysis, predictably causes hyperkalemia from the release of intracellular K^+.

HYPOALDOSTERONISM AND HYPERKALEMIA

Aldosterone release from the adrenal gland may be reduced by hyporeninemic hypoaldosteronism, medications, or primary hypoaldosteronism or by isolated deficiency of ACTH (secondary hypoaldosteronism). Primary hypoaldosteronism may be genetic or acquired (Chap. 342) but is commonly caused by autoimmunity either in Addison's disease or in the context of a polyglandular endocrinopathy. HIV has surpassed tuberculosis as the most important infectious cause of adrenal insufficiency. The adrenal involvement in HIV disease is usually subclinical; however, adrenal insufficiency may be precipitated by stress, drugs such as ketoconazole that inhibit steroidogenesis, or the acute withdrawal of steroid agents such as megestrol.

Hyporeninemic hypoaldosteronism is a very common predisposing factor in several overlapping subsets of hyperkalemic patients: diabetic patients, the elderly, and patients with renal insufficiency. Classically, these patients should have suppressed PRA and aldosterone; approximately 50% have an associated acidosis with a reduced renal excretion of NH_4^+, a positive urinary anion gap, and urine pH <5.5. Most patients are volume expanded, with secondary increases in circulating atrial natriuretic peptide (ANP) that inhibit both renal renin release and adrenal aldosterone release.

RENAL DISEASE AND HYPERKALEMIA

Chronic kidney disease and end-stage kidney disease are very common causes of hyperkalemia because of the associated deficit or absence of functioning nephrons. Hyperkalemia is more common in oliguric acute kidney injury; distal

tubular flow rate and Na^+ delivery is less of a limiting factor in nonoliguric patients. Hyperkalemia out of proportion to GFR can also be seen in the context of tubulointerstitial disease that affects the distal nephron, such as amyloidosis, sickle cell anemia, interstitial nephritis, and obstructive uropathy.

Hereditary renal causes of hyperkalemia have overlapping clinical features with hypoaldosteronism, hence the diagnostic label *pseudohypoaldosteronism* (PHA). PHA-I has both an autosomal recessive and an autosomal dominant form. The autosomal dominant form is due to loss-of-function mutations in MLR; the recessive form is caused by various combinations of mutations in the three subunits of ENaC, resulting in impaired Na^+ channel activity in principal cells and other tissues. Patients with recessive PHA-I experience lifelong salt wasting, hypotension, and hyperkalemia, whereas the phenotype of autosomal dominant PHA-I due to MLR dysfunction improves in adulthood. Pseudohypoaldosteronism type II (PHA-II, also known as hereditary hypertension with hyperkalemia) is in every respect the mirror image of GS caused by loss of function in NCC, the thiazide-sensitive Na^+-Cl^- cotransporter (see above); the clinical phenotype includes hypertension, hyperkalemia, hyperchloremic metabolic acidosis, suppressed PRA and aldosterone, hypercalciuria, and reduced bone density. PHA-II thus behaves like a gain of function in NCC, and treatment with thiazides results in resolution of the entire clinical phenotype; however, PHA-II is caused by mutations in the WNK1 and WNK4 serine-threonine kinases, which regulate NCC activity.

MEDICATION-ASSOCIATED HYPERKALEMIA

Most medications associated with hyperkalemia cause inhibition of some component of the renin-angiotensin-aldosterone axis. ACE inhibitors, angiotensin-receptor blockers, renin inhibitors, and mineralocorticoid receptors are predictable and common causes of hyperkalemia, particularly when prescribed in combination. The oral contraceptive agent Yasmin-28 contains the progestin drospirenone, which inhibits the MLR and can cause hyperkalemia in susceptible patients. Cyclosporine, tacrolimus, NSAIDs, and cyclooxygenase 2 (COX-2) inhibitors cause hyperkalemia by multiple mechanisms but share the ability to cause hyporeninemic hypoaldosteronism. Notably, most drugs that affect the renin-angiotensin-aldosterone axis also block the local adrenal response to hyperkalemia, thus attenuating the *direct* stimulation of aldosterone release by increased plasma K^+ concentration.

Inhibition of apical ENaC activity in the distal nephron by amiloride and other K^+-sparing diuretics results in hyperkalemia, often with a voltage-dependent hyperchloremic acidosis and/or hypovolemic hyponatremia. Amiloride is structurally similar to the antibiotics trimethoprim (TMP) and pentamidine, which also block ENaC; risk factors for TMP-associated hyperkalemia include the administered dose, renal insufficiency, and hyporeninemic hypoaldosteronism. Indirect inhibition of ENaC at the plasma membrane is also a cause of hyperkalemia; nafamostat, a protease inhibitor utilized in the management of pancreatitis, inhibits aldosterone-induced proteases that activate ENaC by proteolytic cleavage.

CLINICAL FEATURES

Hyperkalemia is a medical emergency because of its effects on the heart. Cardiac arrhythmias associated with hyperkalemia include sinus bradycardia, sinus arrest, slow idioventricular rhythms, ventricular tachycardia, ventricular fibrillation, and asystole. Mild increases in extracellular K^+ affect the repolarization phase of the cardiac action potential, resulting in changes in T-wave morphology; further increase in plasma K^+ concentration depresses intracardiac conduction, with progressive prolongation of the PR and QRS intervals. Severe hyperkalemia results in loss of the P wave and a progressive widening of the QRS complex; development of a sine-wave sinoventricular rhythm suggests impending ventricular fibrillation or asystole. Classically, the electrocardiographic manifestations in hyperkalemia progress from tall peaked T waves (5.5–6.5 m*M*), to a loss of P waves (6.5–7.5 m*M*), to a widened QRS complex (7–8 m*M*), and ultimately to a sine wave pattern (8 m*M*). However, these changes are notoriously insensitive, particularly in patients with chronic kidney disease or end-stage renal disease.

Hyperkalemia from a variety of causes can also present with ascending paralysis; this is denoted secondary hyperkalemic paralysis to differentiate it from familial hyperkalemic periodic paralysis (HYPP). The presentation may include diaphragmatic paralysis and respiratory failure. Patients with familial HYPP develop myopathic weakness during hyperkalemia induced by increased K^+ intake or rest after heavy exercise. Depolarization of skeletal muscle by hyperkalemia unmasks an inactivation defect in skeletal Na^+ channels; autosomal dominant mutations in the *SCN4A* gene encoding this channel are the predominant cause.

Within the kidney, hyperkalemia has negative effects on the ability to excrete an acid load, and so hyperkalemia per se can contribute to metabolic acidosis. This defect appears to be due in part to competition between K^+ and NH^{4+} for reabsorption by the TALH and subsequent countercurrent multiplication, ultimately reducing the medullary gradient for NH^3/NH^4 excretion by the distal nephron. Regardless of the underlying mechanism, restoration of normokalemia can in many instances correct hyperkalemic metabolic acidosis.

DIAGNOSTIC APPROACH

The first priority in the management of hyperkalemia is to assess the need for emergency treatment, followed by a comprehensive workup to determine the cause (**Fig. 45-8**). History and physical examination should focus on medications, diet and dietary supplements, risk factors for kidney failure, reduction in urine output, blood pressure, and volume status. Initial laboratory tests should include electrolytes, BUN, creatinine, serum osmolality, Mg^{2+} and Ca^{2+}, a complete blood count, and urinary pH. A urine Na^+ concentration <20 m*M* indicates that distal Na^+ delivery is a limiting factor in K^+ excretion; volume repletion with 0.9% saline or treatment with furosemide may be effective in reducing plasma K^+ concentration. Serum and urine osmolality is required for calculation of the TTKG (Fig. 45-8**)**. The expected values of the TTKG are largely based on historic data and are <3-4 in the presence of hypokalemia and >6-7 in the presence of hyperkalemia. TTKG is measured as follows:

$$\text{TTKG} = \frac{[K^+]_{\text{urine}} \times \text{osmol}_{\text{serum}}}{[K^+]_{\text{serum}} \times \text{osmol}_{\text{urine}}}$$

Figure 45-8

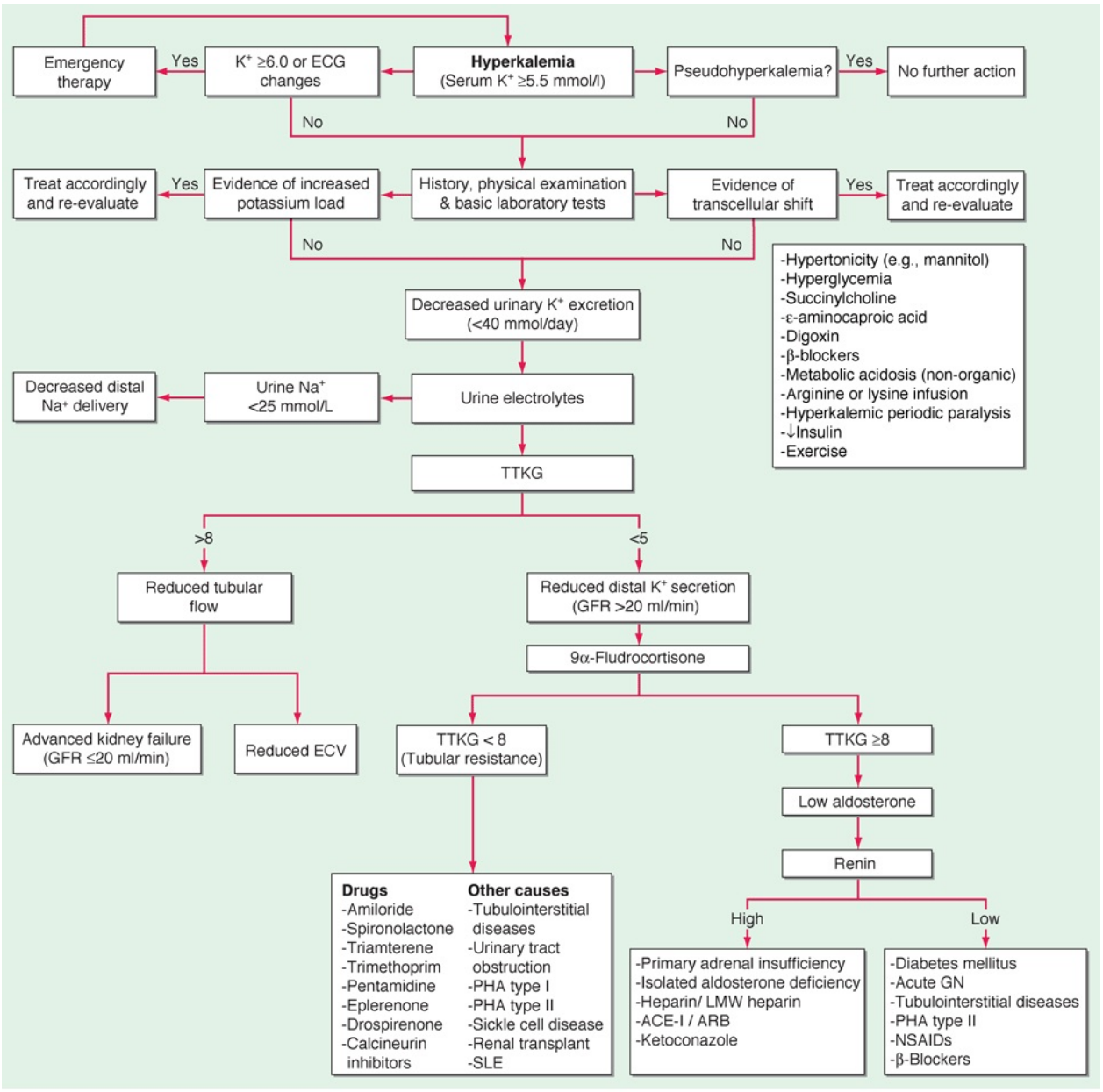

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The diagnostic approach to hyperkalemia. See text for details. ACE-I, angiotensin converting enzyme inhibitor; acute GN, acute glomerulonephritis; ARB, angiotensin II receptor blocker; ECG, electrocardiogram; ECV, effective circulatory volume; GFR, glomerular filtration rate; LMW heparin, low-molecular-weight heparin; NSAIDs, nonsteroidal anti-inflammatory drugs; PHA, pseudohypoaldosteronism; SLE, systemic lupus erythematosus; TTKG, transtubular potassium gradient. *(From Mount and Zandi-Nejad; with permission.)*

TREATMENT: HYPERKALEMIA

Electrocardiographic manifestations of hyperkalemia should be considered a medical emergency and treated urgently.

However, patients with significant hyperkalemia (plasma K^+ concentration ≥6.5–7 m*M*) in the absence of ECG changes should be aggressively managed because of the limitations of ECG changes as a predictor of cardiac toxicity. Urgent management of hyperkalemia includes admission to the hospital, continuous cardiac monitoring, and immediate treatment. The treatment of hyperkalemia is divided into three stages:

1. *Immediate antagonism of the cardiac effects of hyperkalemia* . Intravenous calcium serves to protect the heart while measures are taken to correct hyperkalemia. Calcium raises the action potential threshold and reduces excitability without changing the resting membrane potential. By restoring the difference between the resting and threshold potentials, calcium reverses the depolarization blockade caused by hyperkalemia. The recommended dose is 10 mL of 10% calcium gluconate (3–4 mL of calcium chloride), infused intravenously over 2 to 3 min with cardiac monitoring. The effect of the infusion starts in 1–3 min and lasts 30–60 min; the dose should be repeated if there is no change in ECG findings or if they recur after initial improvement. Hypercalcemia potentiates the cardiac toxicity of digoxin; hence, intravenous calcium should be used with extreme caution in patients taking this medication. If judged necessary,10 mL of 10% calcium gluconate can be added to 100 mL of5% dextrose in water and infused over 20–30 min to avoid acute hypercalcemia.
2. *Rapid reduction in plasma K^+ concentration by redistribution into cells*. Insulin lowers plasma K^+ concentration by shifting K^+ into cells. The recommended dose is 10 units of IV regular insulin followed immediately by 50 mL of 50% dextrose (D50W, 25 g of glucose total); the effect begins in 10–20 min, peaks at 30–60 min, and lasts 4 to 6 h. Bolus D50W without insulin is *never* appropriate because of the risk of acutely worsening hyperkalemia due to the osmotic

effect of hypertonic glucose. Hypoglycemia is common with insulin plus glucose; hence, this should be followed by an infusion of 10% dextrose at 50 to 75 mL/h, with close monitoring of plasma glucose concentration. In hyperkalemic patients with glucose concentrations ≥200–250 mg/dL, insulin should be administered *without* glucose, again with close monitoring of glucose concentrations.

3. β^2-agonists, most commonly albuterol, are effective but underutilized agents for the acute management of hyperkalemia. Albuterol and insulin with glucose have an additive effect on plasma K^+ concentration; however, ~20% of patients with end-stage renal disease are resistant to the effect of β^2-agonists; hence, these drugs should not be used without insulin. The recommended dose for inhaled albuterol is 10–20 mg of nebulized albuterol in 4 mL of normal saline, inhaled over 10 min; the effect starts at about 30 min, reaches its peak at about 90 min, and lasts 2–6 h. Hyperglycemia is a side effect, along with tachycardia; β^2-agonists should be used with caution in hyperkalemic patients with known cardiac disease.

Intravenous bicarbonate has no role in the routine treatment of hyperkalemia. It should be reserved for patients with hyperkalemia and concomitant metabolic acidosis, and only if judged appropriate for management of the acidosis. It should not be given as a hypertonic intravenous bolus in light of the risk of hypernatremia but should be infused in an isotonic or hypotonic fluid (e.g., 150 meq in 1 L of D5W).

Removal of potassium. This typically is accomplished by using cation exchange resins, diuretics, and/or dialysis. Sodium polystyrene sulfonate (SPS) exchanges Na^+ for K^+ in the gastrointestinal tract and increases the fecal excretion of K^+. The recommended dose of SPS is 15-30 g, typically given in a premade suspension with 33% sorbitol to avoid constipation. The effect of SPS on plasma K^+ concentration is slow; the full effect may take up to 24 hours and usually requires repeated doses every 4–6 hours. Intestinal necrosis is the most serious complication of SPS. Studies in experimental animals suggest that sorbitol is required for the intestinal injury; however, SPS crystals can often be detected in the injured human intestine, suggesting a direct role for SPS crystals in this complication. Regardless, in light of the risk of intestinal necrosis, the U.S. Food and Drug Administration has recently stated that the administration of sorbitol with SPS is no longer recommended; however, administering SPS without sorbitol might not eliminate the risk of intestinal necrosis, given the evident role for the SPS resin. Therefore, clinicians must carefully consider whether emergency treatment with SPS is necessary and appropriate for the treatment of hyperkalemia; for example, SPS is unnecessary if acute dialysis is appropriate and immediately available. If SPS is administered, the preparation should ideally not contain sorbitol. Reasonable substitutes for the laxative effect of sorbitol include lactulose and some preparations of polyethylene glycol 3350; however, data demonstrating the efficacy and safety of these laxatives with SPS are not available. SPS should not be administered in patients at higher risk for intestinal necrosis, including postoperative patients, patients with a history of bowel obstruction, patients with slow intestinal transit, patients with ischemic bowel disease, and renal transplant patients. Loop and thiazide diuretics can be utilized to reduce plasma K^+ concentration in volume-replete or hypervolemic patients with sufficient renal function for a diuretic response. Finally, hemodialysis is the most effective and reliable method to reduce plasma K^+ concentration; peritoneal dialysis is considerably less effective. The amount of K^+ removed during hemodialysis depends on the relative distribution of K^+ between ICF and ECF (potentially affected by prior therapy for hyperkalemia), the type and surface area of the dialyzer used, dialysate and blood flow rates, the dialysate flow rate, dialysis duration, and the plasma to dialysate K^+ gradient.

FURTHER READINGS

Adrogue HJ, Madias NE: Hypernatremia. N Engl J Med 342:1493, 2000 [PMID: 10816188]

Ellison DH, Berl T: Clinical practice: The syndrome of inappropriate antidiuresis. N Engl J Med 356:2064, 2007 [PMID: 17507705]

Mohmand HK et al: Hypertonic saline for hyponatremia: Risk of inadvertent overcorrection. Clin J Am Soc Nephrol 2:1110, 2007 [PMID: 17913972]

Mount DB, Zandi-Nejad K: Disorders of potassium balance, in *Brenner and Rector's The Kidney*, 8th ed, BM Brenner (ed). Philadelphia, W.B. Saunders, 2008, pp 547–587

Perianayagam A et al: DDAVP is effective in preventing and reversing inadvertent overcorrection of hyponatremia. Clin J Am Soc Nephrol 3:331, 2008 [PMID: 18235152]

Schrier RW: Decreased effective blood volume in edematous disorders: What does this mean? J Am Soc Nephrol 18:2028, 2007 [PMID: 17568020]

——— et al: Tolvaptan, a selective oral vasopressin V2-receptor antagonist, for hyponatremia. N Engl J Med 355:2099, 2006

Sterns RH et al: Ion-exchange resins for the treatment of hyperkalemia: Are they safe and effective? J Am Soc Nephrol 21:733, 2010 [PMID: 20167700]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CASE 1

A 23-year-old woman was admitted with a 3-day history of fever, cough productive of blood-tinged sputum, confusion, and orthostasis. Past medical history included Type I diabetes mellitus. A physical examination in the emergency department indicated postural hypotension, tachycardia, and Kussmaul respiration, and the breath was noted to smell of "acetone." Examination of the thorax suggested consolidation in the right lower lobe.

Laboratory Data		**Units**
Na^+	130	meq/L
K^+	5.0	meq/L
Cl^-	96	meq/L
CO_2	14	meq/L
Blood urea nitrogen (BUN)	20	mg/dL
Creatinine (Creat)	1.3	mg/dL
Glucose	450	mg/dL
Arterial Blood Gases	**On room air**	
pH	7.39	
PCO_2	24	mmHg
PaO_2	89	mmHg
HCO_3^-	14	meq/L
Anion gap	20	meq/L
Urinalysis		
Urine ketones	Positive 4+	
Glucose	Positive 4+	
Serum ketones	Strongly positive 1:8	
Chest x-ray		
Pneumonic infiltrate, right lower lobe		

Approach to Diagnosis

The diagnosis of the acid-base disorder should proceed in a stepwise fashion:

1. The normal anion gap (AG) is 10 meq/L, but in this case the anion gap is elevated (20 meq/L). Therefore, the change in AG (ΔAG) = 10 meq/L.
2. Compare the ΔAG and the $\Delta[HCO_3^-]$. In this case, the ΔAG, as noted above, is 10 and the $\Delta[HCO_3^-]$ (25 – 14) is 11. Therefore, the increment in the anion gap is approximately equal to the decrement in bicarbonate.
3. The next step is to estimate the respiratory compensatory response. In this case, the predicted $Paco_2$ for an $[HCO_3^-]$ of 14 should be approximately 29 mmHg. This value is obtained by adding 15 to the measured $[HCO_3^-]$ (15 + 14 = 29) or by calculating the predicted $Paco_2$ from the Winter equation: 1.5 x $[HCO_3^-]$ + 8. In either case, the predicted value for $Paco_2$ of 29 is significantly

higher than the measured value of 24. Therefore, the prevailing $Paco_2$ exceeds the range for compensation alone and is too low.

4. Therefore, this patient has a mixed acid-base disturbance with two components: (a) high anion-gap acidosis secondary to ketoacidosis and (b) respiratory alkalosis secondary to community-acquired pneumonia. The respiratory alkalosis resulted in an additional component of hyperventilation that exceeded the compensatory response driven by metabolic acidosis, explaining the normal pH in this case. The finding of respiratory alkalosis in the setting of a high-gap acidosis suggests another cause of the respiratory component, which in this case may be attributed to the community-acquired pneumonia.

The clinical features in this case include hyperglycemia, hypovolemia, ketoacidosis, central nervous system (CNS) signs of confusion, and superimposed pneumonia. This clinical scenario is consistent with diabetic ketoacidosis (DKA) developing in a patient with known Type I diabetes mellitus. Infections in DKA are common and may be a precipitating feature in the development of ketoacidosis.

The diagnosis of DKA is usually not challenging but should be considered in all patients with an elevated anion-gap and metabolic acidosis. Hyperglycemia and ketonemia (positive acetoacetate at a dilution of 1:8 or greater) are sufficient criteria for diagnosis in patients with Type 1 diabetes mellitus.

The $\Delta[HCO_3^-]$ should approximate the increase in the plasma anion gap (ΔAG), but this equality can be modified by several factors. For example, the ΔAG often decreases with IV hydration, as glomerular filtration increases and ketones are excreted into the urine. The decrement in plasma sodium results from hyperglycemia, which induces the movement of water into the extracellular compartment from the intracellular compartment of cells that require insulin for the transport of glucose. Additionally, a natriuresis occurs in response to an osmotic diuresis associated with hyperglycemia. Moreover, in patients with DKA, thirst is very common and water ingestion often continues. The plasma potassium concentration is usually mildly elevated, but in the face of acidosis and as a result of the ongoing osmotic diuresis, a significant total-body deficit of potassium is almost always present. Recognition of this total-body deficit is critically important. The inclusion of potassium replacement in the therapeutic regimen at the appropriate time and with the appropriate indications (see below) is essential. Volume depletion is a very common finding in diabetic ketoacidosis and is a pivotal component in the pathogenesis of the disorder.

Approach to Management

Patients with diabetic ketoacidosis often have a sustained and significant deficit of sodium, potassium, water, bicarbonate, and phosphate. The general approach to treatment requires attention to all those abnormalities. Successful treatment of DKA involves a stepwise approach, as follows:

1. *Replace extracellular fluid (ECF) volume deficits* . Since most patients present with actual or relative hypotension and, at times, impending shock, the initial fluid administered should be 0.9% NaCl infused rapidly until the systolic blood pressure is >100 mmHg or until 2–3 L cumulatively has been administered. During the initial 2–3 h of infusion of saline, the decline in blood glucose can be accounted for by dilution and increased renal excretion. Glucose should be added to the infusion as D_5 NS or D_5 0.45% NS once the plasma glucose declines to 230 mg/dL or less.
2. *Abate the production of ketoacids.* Regular insulin is required during diabetic ketoacidosis as an initial bolus of 0.1 U/kg body weight (BW) IV followed immediately by a continuous infusion of 0.1 U/kg BW per h in NS. The effectiveness of IV (not subcutaneous) insulin can be tracked by observing the decline in plasma ketones. Since the increment in the anion gap above the normal value of 10 meq/L represents accumulated ketoacids in DKA, the disappearance of ketoacid anions is reflected by the narrowing and eventual normalization of the anion gap. Typically, the plasma anion gap returns to normal within 8–12 h.
3. *Replace potassium deficits*. Although patients with DKA often have hyperkalemia due to insulin deficiency, they are usually severely K^+ depleted. KCl (20 meq/L) should be added to each liter of IV fluids when urine output is established and after insulin has been administered.
4. *Correct the metabolic acidosis*. The plasma bicarbonate concentration usually will not increase for several hours because of dilution from administered IV NaCl. The plasma $[HCO^-]$ approaches 18 meq/L once ketoacidosis disappears. Sodium bicarbonate therapy often is not recommended, especially in children. Bicarbonate is administered to adults with DKA for extreme acidemia (pH <7.1); for elderly patients (>70 years), a threshold pH of 7.20 is recommended. Sodium bicarbonate, if administered, should be given only in small amounts. Since ketoacids are metabolized after insulin therapy, bicarbonate will be added to the ECF. Overshoot alkalosis may result from the combination of exogenously administered sodium bicarbonate and metabolic production of bicarbonate.
5. *Phosphate.* In the first 6–8 h of therapy, it may be necessary to infuse potassium with phosphate because of the unmasking of phosphate depletion during combined insulin and glucose therapy. The latter drives phosphate into the cell. Therefore, in patients with DKA, the plasma phosphate level should be followed closely but phosphate should never be replaced empirically. Phosphate should be administered to patients with a declining plasma phosphate once the phosphate level declines into the low-normal level. Therapy is advisable in the form of potassium phosphate at a

rate of 6 mmol/h.

6. Always seek underlying factors such as infection, myocardial infarction, pancreatitis, cessation of insulin therapy, or other events that are responsible for initiating diabetic ketoacidosis. The case presented here is illustrative of this point.
7. Volume overexpansion with IV fluid administration is not uncommon and contributes to the development of hyperchloremic acidosis during the later stages of treatment of DKA. Volume overexpansion should be avoided.

CASE 2

A 25-year-old man with a 6-year history of HIV/AIDS complicated recently by *Pneumocystis jiroveci* pneumonia (PCP) was treated with intravenous trimethoprim-sulfamethoxazole (20 mg TMP/kg per day). On day 4 of treatment, the following laboratory data were obtained:

Laboratory Data	Units	Plasma	Urine
Na^+	meq/L	135	60
K^+	meq/L	6.5	15
Cl^-	meq/L	110	43
HCO_3^-	meq/L	15	0
pH		7.30	5.5
BUN	mg/dL	14	—
Creat	mg/dL	0.9	—
Osmolality	mOsmol/kg H_2O	268	270

Approach to Diagnosis

What caused the hyperkalemia and metabolic acidosis in this patient? What other medications may be associated with a similar presentation? How does one utilize the urine electrolyte data to determine whether the hyperkalemia is of renal origin or is due to a shift from the cell to the extracellular compartment?

Hyperkalemia occurs in 15–20% of hospitalized patients with HIV/AIDS. The usual causes are adrenal insufficiency, the syndrome of hyporeninemic hypoaldosteronism, or one of several drugs, including trimethoprim, pentamidine, nonsteroidal anti-inflammatory drugs, angiotension-converting enzyme (ACE) inhibitors, angiotensin II receptor blockers, spironolactone, and eplerenone. Trimethoprim usually is given in combination with sulfamethoxazole or dapsone for PCP and on average increases the plasma K^+ concentration by about 1 meq/L; however, the hyperkalemia may be severe. Trimethoprim is structurally and chemically related to amiloride and triamterene and in this way may function as a potassium-sparing diuretic. This effect is caused by inhibition of the epithelial sodium channel (ENaC) in the principal cell of the collecting duct. When the Na^+ channel is blocked, K^+ secretion also is inhibited; K^+ secretion is dependent on the lumen-negative potential difference generated by Na^+ entry through ENaC (**Fig. e15-1**).

Figure e15-1

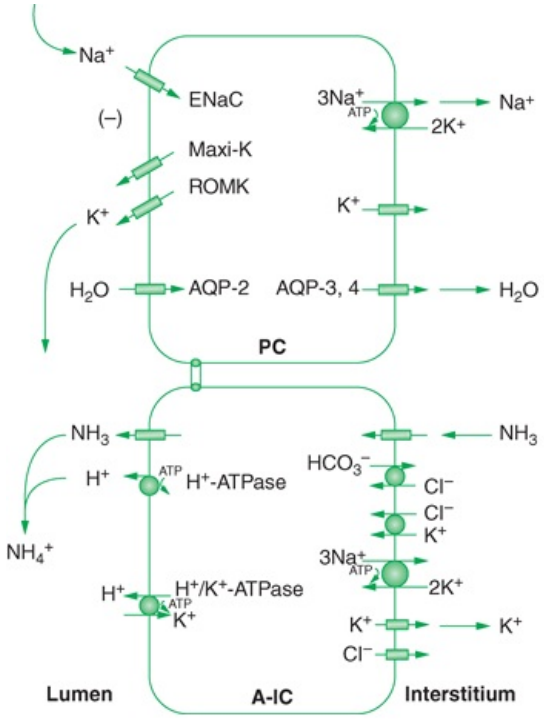

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Water, sodium, potassium, ammonia, and proton transport in principal cells (PC) and adjacent type A intercalated cells. Water is absorbed down the osmotic gradient by principal cells through the apical aquaporin-2 (AQP-2) and basolateral aquaporin-3 and aquaporin-4 channels. The absorption of Na^+ via the amiloride-sensitive epithelial sodium channel (ENaC) generates a lumen-negative potential difference that drives K^+ excretion through the apical secretory K^+ channel, ROMK (renal outer medullary K^+ channel), and/or the flow-dependent maxi-K channel. Transepithelial ammonia (NH_3) transport and proton transport occur in adjacent type A intercalated cells via apical and basolateral ammonia channels and apical H^+-ATPase pumps, respectively; NH_4^+ ultimately is excreted in the urine in the defense of systemic pH. Electrogenic proton secretion by type A intercalated cells also is affected by the lumen-negative potential difference generated by the adjacent principal cells so that reduction of this lumen-negative electrical gradient can reduce H^+ excretion. Type A intercalated cells also reabsorb filtered K^+ in potassium-deficient states via apical H^+/K^+-ATPase.

TMP is associated with a non-anion-gap acidosis that parallels the development of hyperkalemia such that the co-occurrence of hyperkalemia and metabolic acidosis is not uncommon in this setting. H^+ secretion via apical H^+-ATPase pumps in adjacent type A intercalated cells (Fig. e15-1) is also electrogenic so that the reduction in the lumen-negative potential difference due to TMP inhibits distal H^+ secretion; this often is referred to as a voltage defect form of distal renal tubular acidosis. Systemic hyperkalemia also suppresses renal ammoniagenesis, ammonium excretion, and thus acid excretion; i.e., hyperkalemia per se has multiple effects on urinary acidification.

The inhibitory effect of trimethoprim on K^+ and H^+ secretion in the cortical collecting tubule follows a dose-response relationship; therefore, the higher doses of this agent used in HIV/AIDS patients with PCP or in deep tissue infections with methicillin-resistant *Staphylococcus aureus* (MRSA) result in a higher prevalence of hyperkalemia and acidosis. Conventional doses of trimethoprim also can induce hyperkalemia and/or acidosis in predisposed patients, in particular the elderly, patients with renal insufficiency, and/or those with baseline hyporeninemic hypoaldosteronism.

One means by which to assess the role of the kidney in the development of hyperkalemia is to calculate, from a spot urine and coincident plasma sample, the transtubular potassium gradient (TTKG). The TTKG is calculated as $(P_{osmol} \times U_{Potassium})/(P_{Potassium} \times U_{osmol})$. The expected values of the TTKG are <3 in the presence of hypokalemia (see also Case 7 and Case 8) and >7–8 in the presence of hyperkalemia. In this case, the value for the TTKG of approximately 2 indicates that renal excretion of potassium is abnormally low for the prevailing hyperkalemia. Therefore, the inappropriately low TTKG indicates that the hyperkalemia is of renal tubular origin.

Approach to Management

Knowledge of the factors that control potassium secretion by the cortical collecting tubule principal cell can be helpful in understanding the basis for treatment of the hyperkalemia, especially if discontinuing the offending agent is not a reasonable clinical option. Potassium secretion is encouraged by a higher urine flow rate, increased distal delivery of sodium, distal delivery of a poorly reabsorbed anion (such as bicarbonate), and/or administration of a loop diuretic. Therefore, the approach to treatment in this patient should include intravenous 0.9% NaCl to expand the ECF and deliver more Na^+ and Cl^- to the cortical collecting tubule. In addition, as the TMP molecule must be protonated to inhibit ENaC, alkalinization of the renal tubule blunts the effect of the drug on distal tubular K^+ secretion. As an alternative to inducing bicarbonaturia to assist in potassium secretion, a carbonic anhydrase inhibitor could be administered to induce a significant kaliuresis. However, in the case presented here, for acetazolamide to be effective, the non-anion-gap metabolic acidosis in this patient would have to be corrected first; acetazolamide thus would require the coadministration of intravenous sodium bicarbonate for maximal benefit. Finally, systemic hyperkalemia directly suppresses renal ammoniagenesis, ammonium excretion, and thus acid excretion. Correcting the hyperkalemia with a potassium-binding resin (Kayexalate) sometimes is appropriate in these patients; the subsequent decline in the plasma K^+ concentration also will increase urinary ammonium excretion, helping to correct the acidosis.

CASE 3

A 63-year-old man was admitted to the intensive care unit (ICU) with a severe aspiration pneumonia. Past medical history included schizophrenia for which he required institutional care; treatment had included neuroleptics and intermittent lithium, with the lithium restarted 6 months before admission. The patient was treated with antibiotics and intubated for several days, with the development of polyuria (3–5 L/d), hypernatremia, and acute renal insufficiency; the peak plasma Na^+ concentration was 156 meq/L, and peak creatinine was 2.6 mg/dL. Urine osmolality was measured once and reported as 157 mosmol/kg, with a coincident plasma osmolality of 318 mosmol/kg. Lithium was stopped on admission to the ICU.

On physical examination, the patient was alert, extubated, and thirsty. Weight was 97.5 kg. Urine output for the last 24 h had been 3.4 L, with an IV intake of 2 L/d of D5W.

Laboratory Data					
Na^+ 150	K^+ 3.9	Cl^- 114	HCO_3^- 26	BUN 8	Creat 1.7
Glu 95	Alb 3.1	Ca^{2+} 8.1	Phos 2.6	Mg 2	Plasma osm 315
Urine:	Na^+ 34	K^+ 5.2	Osm 137		

After 3 days of intravenous hydration, a water deprivation test was performed. A single dose of 2 μg IV desmospressin (DDAVP) was given at 9 h (+9):

Laboratory Data					
Time (h)	0	+6	+8	+12	+18
Na^+	145	148	150	152	149
K^+	5.4	5.3	3.9	3.9	3.9
Cl^-	111	110	118	120	114
HCO_3^-	24	27	25	242	25
Creat	1.3	1.3	1.4	1.3	1.3
S_{osmol}	300	311	315		
U_{osmol}	132	140	201	237	257
AVP		8.4	6.3		

Approach to Diagnosis

Why did the patient develop hypernatremia, polyuria, and acute renal insufficiency? What does the water deprivation test demonstrate? What is the underlying pathophysiology of this patient's hypernatremic syndrome?

This patient became polyuric after admission to the ICU with severe pneumonia and developed significant hypernatremia and acute renal insufficiency. Polyuria can result from either an osmotic diuresis or a water diuresis. An osmotic diuresis can be caused by excessive excretion of Na^+-Cl^-, mannitol, glucose, and/or urea, with a daily solute excretion of >750–1000 mosmol/d (>15 mosmol/kg body water per day). In this case, however, the patient was excreting large volumes of very hypotonic urine, with a urine osmolality that was substantially lower than that of plasma; this, by definition, was a water diuresis, resulting in inappropriate excretion of free water and hypernatremia. The appropriate response to hypernatremia and a plasma osmolality >295 mosmol/kg is an increase in circulating vasopressin (AVP) and the excretion of low volumes (<500 mL/d) of maximally concentrated urine, i.e., urine with osmolality >800 mosmol/kg; this patient's response to hypernatremia was clearly inappropriate, due to either a loss of circulating AVP [central diabetes insipidus (CDI)] or renal resistance to AVP [nephrogenic diabetes insipidus (NDI)]. Ongoing loss of free water was sufficiently severe that absolute hypovolemia ensued despite the fact that approximately two-thirds of the excreted water was derived from the intracellular fluid (ICF) compartment rather than the ECF compartment. Hypovolemia led to an acute decrease in glomerular filtration rate (GFR), i.e., acute renal insufficiency, with gradual improvement following hydration (see below).

After correction of the hypernatremia and acute renal insufficiency with appropriate hydration (see below), the patient was subjected to a water deprivation test followed by administration of DDAVP. This test helps determine whether an inappropriate water diuresis is caused by CDI or NDI. The patient was water-restricted beginning in the early morning, with careful monitoring of vital signs and urine output; overnight water deprivation of patients with diabetes insipidus is unsafe and clinically inappropriate because of the potential for severe hypernatremia. The plasma Na^+ concentration—more accurate and more immediately available than plasma osmolality—was monitored hourly during water deprivation. A baseline AVP sample was drawn at the beginning of the test, with a second sample drawn once the plasma Na^+ reached 148–150 meq/L. At this point a single 2-μg dose of the V2 vasopressin receptor agonist DDAVP was administered. An alternative approach would have been to measure AVP and administer DDAVP when the patient was initially hypernatremic; however, it would have been less safe to administer DDAVP in the setting of renal impairment as clearance of DDAVP is renal-dependent.

The patient's water deprivation test was consistent with NDI, with an AVP level within the normal range in the setting of hypernatremia (i.e., no evidence of CDI) and an inappropriately low urine osmolality that failed to increase by >50%, or >150 mosmol/kg, after both water deprivation and the administration of DDAVP. This defect would be considered compatible with "complete" NDI; patients with "partial NDI" can achieve urine osmolalities of 500–600 mosmol/kg after DDAVP treatment but will not concentrate their urine maximally to 800 mosmol/kg or higher.

NDI has a number of genetic and acquired causes, which all share interference with some aspect of the renal concentrating mechanism. For example, loss-of-function mutations in the V2 AVP receptor cause X-linked NDI. This patient had NDI due to lithium therapy, perhaps the most common cause of NDI in adults. Lithium causes NDI via direct inhibition of renal glycogen synthase kinase-3 (GSK3), a kinase thought to be the pharmacologic target of lithium in psychiatric disease; renal GSK3 is required for the response of principal cells to AVP. Lithium also induces the expression of cyclooxygenase 2 (COX-2) in the renal medulla; COX-2-derived prostaglandins inhibit AVP-stimulated salt transport by the thick ascending limb and AVP-stimulated water transport by the collecting duct, exacerbating lithium-associated polyuria. The entry of lithium through the amiloride-sensitive Na^+ channel ENaC (Fig. e15-1) is required for the effect of the drug on principal cells, and so combined therapy with lithium and amiloride can mitigate lithium-associated NDI. However, lithium causes chronic tubulointerstitial scarring and chronic kidney disease after prolonged therapy so that patients may have a persistent NDI long after stopping the drug, with a reduced therapeutic benefit from amiloride. Notably, this patient had been treated intermittently for several years with lithium, with the development of chronic kidney disease (baseline creatinine 1.3–1.4) and NDI that persisted after the drug was stopped.

Approach to Management

How should this patient have been treated? What are the major pitfalls of therapy?

This patient developed severe hypernatremia due to a water diuresis from lithium-associated NDI. Treatment of hypernatremia must include both replacement of the existing free water deficit and daily replacement of ongoing free water loss. The first step is to estimate total-body water (TBW), typically estimated as 50% of body weight in women and 60% in men. The free water deficit is then calculated as $[(Na^+ - 140)/140] \times TBW$. In this patient, the free water deficit was 4.2 L at a weight of 97.5 kg and a plasma Na^+ concentration of 150 meq/L. This free water deficit should be replaced slowly over 48

–72 h to avoid increasing the plasma Na^+ concentration by >10 meq/L per 24 h. A common mistake is to replace this deficit while neglecting to replace ongoing losses of free water so that plasma Na^+ concentration either fails to correct or, in fact, increases.

Ongoing losses of free water can be estimated by using the equation for electrolyte-free-water clearance:

where V is urinary volume, U_{Na} is urinary $[Na^+]$, U_K is urinary $[K^+]$, and P_{Na} is plasma $[Na^+]$.

For this particular patient the C_eH_2O was 2.5 L/d when initially evaluated, i.e., with urine Na^+ and K^+ concentrations of 34 and 5.2 meq/L, plasma Na^+ concentration of 150 meq/L, and a urinary volume of 3.4 L. Therefore, the patient was given 2.5 L of D5W over the first 24 h to replace ongoing free water losses, along with 2.1 L of D_5W to replace half his free water deficit. Daily random urine electrolytes and urinary volume measurement can be utilized to monitor C_eH_2O and adjust daily fluid administration in this manner while following plasma Na^+ concentration. Physicians often calculate the free water deficit to guide therapy for hypernatremia, providing half the deficit in the first 24 h. This approach can be adequate in patients who do not have significant ongoing losses of free water, e.g., with hypernatremia due to decreased free water intake. This case illustrates how free water requirements can be grossly underestimated in hypernatremic patients if ongoing daily free water losses are not taken into account.

CASE 4

A 78-year-old man was admitted with pneumonia and hyponatremia. Plasma Na^+ concentration was initially 129 meq/L, decreasing within 3 days to 118–120 meq/L despite fluid restriction to 1 L/d. A chest CT revealed a right 2.8 x 1.6 cm infrahilar mass and postobstructive pneumonia. The patient was an active smoker. Past medical history was notable for laryngeal carcinoma treated 15 years earlier with radiation therapy, renal cell carcinoma, peripheral vascular disease, and hypothyroidism. On review of systems, he denied headache, nausea, and vomiting. He had chronic hip pain that was managed with acetaminophen with codeine. Other medications included cilostazol, amoxicillin/clavulanate, digoxin, diltiazem, and thyroxine. He was euvolemic on examination, with no lymphadenopathy and a normal chest examination.

Laboratory data:

Laboratory Data						
Na^+ 120	K^+ 4.3	Cl^- 89	HCO_3^-	BUN 8	Creat 1	Glu 93
Alb 3.1	Ca^{2+} 8.9	Phos 2.8	23 Mg 2	Plasma osm 248 mosmol/kg		
Cortisol	25 μg/dL	TSH 2.6	Uric acid 2.7 mg/dL			
Urine:	Na^+ 97	K^+ 22	Cl^- 86	Osm 597		

The patient was treated with furosemide, 20 mg PO bid, and salt tablets. The plasma Na^+ concentration increased to 129 meq/L with this therapy; however, the patient developed orthostatic hypotension and dizziness. He was started on demeclocycline, 600 mg PO in the morning and 300 mg in the evening, just before discharge from the hospital. Plasma Na^+ concentration increased to 140 meq/L with a BUN of 23 and creatinine of 1.4, at which point demeclocycline was reduced to 300 mg PO bid. Bronchoscopic biopsy eventually showed small cell lung cancer; the patient declined chemotherapy and was admitted to hospice.

Approach to Diagnosis and Management

What factors contributed to this patient's hyponatremia? What are the therapeutic options?

This patient developed hyponatremia in the context of a central lung mass and post-obstructive pneumonia. He was clinically euvolemic, with a generous urine Na^+ concentration and low plasma uric acid concentration. He was euthyroid, with no evidence of pituitary dysfunction or secondary adrenal insufficiency. The clinical presentation is consistent with the syndrome of inappropriate antidiuresis (SIAD). Although pneumonia was a potential contributor to the SIAD, it was notable that the plasma Na^+ concentration decreased despite a clinical response to antibiotics. It was suspected that this patient had SIAD due to small cell lung cancer, with a central lung mass on chest CT and a significant smoking history. There was a history of laryngeal cancer and renal cancer, but with no evidence of recurrent disease; those malignancies were not considered contributory to the SIAD. Biopsy of the

lung mass ultimately confirmed the diagnosis of small cell lung cancer, which is responsible for ~75% of malignancy-associated SIAD; ~10% of patients with this neuroendocrine tumor have a plasma Na^+ concentration <130 meq/L at presentation. The patient had no other "nonosmotic" stimuli for an increase in AVP, with no medications associated with SIAD and minimal pain or nausea.

The patient had no symptoms attributable to hyponatremia but was judged at risk for worsening hyponatremia from severe SIAD. Persistent, chronic hyponatremia (duration >48 h) results in an efflux of organic osmolytes (creatine, betaine, glutamate, myo-inositol, and taurine) from brain cells; this response reduces intracellular osmolality and the osmotic gradient favoring water entry. This cellular response does not fully protect patients from symptoms, which can include vomiting, nausea, confusion, and seizures, usually at a plasma Na^+ concentration <125 meq/L. Even patients who are judged "asymptomatic" can manifest subtle gait and cognitive defects that reverse with correction of hyponatremia. Chronic hyponatremia also increases the risk of bony fractures due to an increased risk of falls and a hyponatremia-associated reduction in bone density. Therefore, every attempt should be made to correct plasma Na^+ concentration safely in patients with chronic hyponatremia. This is particularly true in malignancy-associated SIAD, in which it can take weeks to months for a tissue diagnosis and the subsequent reduction in AVP after the initiation of chemotherapy, radiotherapy, and/or surgery.

What are the therapeutic options in SIAD? Water deprivation, a cornerstone of therapy for SIAD, had little effect on the plasma Na^+ concentration in this patient. The urine:plasma electrolyte ratio (urinary $[Na^+] + [K^+]$/plasma $[Na^+]$) can be utilized to estimate electrolyte-free water excretion and the required degree of water restriction; patients with a ratio >1 should be restricted more aggressively (<500 mL/d), those with a ratio ~1 should be restricted to 500–700 mL/d, and those with a ratio <1 should be restricted to <1 L/d. This patient had a urine:plasma electrolyte ratio of 1 and predictably did not respond to a moderate water restriction of ~1 L/day. A more aggressive water restriction theoretically would have been successful; however, this can be very difficult for patients with SIAD to tolerate because their thirst is also inappropriately stimulated.

Combined therapy with furosemide and salt tablets often can increase the plasma Na^+ concentration in SIAD; furosemide reduces maximal urinary concentrating ability by inhibiting the countercurrent mechanism, and the salt tablets mitigate diuretic-associated NaCl loss. This regimen is not always successful and requires careful titration of salt tablets to avoid volume depletion; indeed, in this patient the plasma Na^+ concentration remained <130 meq/L, and the patient became orthostatic. The principal cell toxin, demeclocycline, is an alternative oral agent for SIAD. Treatment with demeclocycline was very successful in this patient, with an increase in plasma Na^+ concentration to 140 meq/L. However, demeclocycline can be natriuretic, leading to a prerenal decrease in GFR. Demeclocycline also has been implicated in nephrotoxic injury, particularly in patients with cirrhosis and chronic liver disease, in whom the drug accumulates. Notably, this patient developed a significant but stable decrease in GFR while on demeclocycline, necessitating a reduction in the administered dose.

A major advance in the management of hyponatremia was the clinical development of vasopressin antagonists (vaptans). These agents inhibit the effect of AVP on renal V2 receptors, resulting in the excretion of electrolyte-free water and correction of hyponatremia. The specific indications for these agents are not clear despite U.S. Food and Drug Administration (FDA) approval for the management of both euvolemic and hypervolemic hyponatremia. It is, however, anticipated that the vaptans will play an increasing role in the management of SIAD and other causes of hyponatremia. Indeed, if this particular patient had continued with active therapy for his cancer, substitution of demeclocycline with oral tolvaptan (a V2-specific oral vaptan) would have been the next appropriate step in light of the development of renal insufficiency with demeclocycline. As with other measures to correct hyponatremia (hypertonic saline, demeclocycline, etc.) the vaptans have the potential to "overcorrect" plasma Na^+ concentration (a rise of >8–10 meq/L per 24 h or 18 meq/L per 18 h), thus increasing the risk for osmotic demyelination (see Case 5). Therefore, the plasma Na^+ concentration should be monitored closely during the initiation of therapy with these agents as well.

CASE 5

A 76-year-old woman presented with a several-month history of diarrhea, with marked worsening over the 2–3 weeks before admission (up to 12 stools a day). Review of systems was negative for fever, orthostatic dizziness, nausea and vomiting, and headache. Past medical history included hypertension, kidney stones, and hypercholesterolemia; medications included atenolol, spironolactone, and lovastatin. She also reliably consumed >2 L of liquid a day in management of nephrolithiasis.

The patient received 1 L of saline over the first 5 h of her hospital admission. On examination at hour 6, the heart rate was 72 sitting and 90 standing, and blood pressure was 105/50 lying and standing. Her jugular venous pressure (JVP) was indistinct with no peripheral edema. On abdominal

examination, the patient had a slight increase in bowel sounds but a nontender abdomen and no organomegaly.

The plasma Na^+ concentration on admission was 113 meq/L, with a creatinine of 2.35 (**Table e15-1**). At hospital hour 7 the plasma Na^+ concentration was 120 meq/L, potassium 5.4 meq/L, chloride 90 meq/L, bicarbonate 22 meq/L, BUN 32 mg/dL, creatinine 2.02 mg/dL, glucose 89 mg/dL, total protein 5.0, and albumin 1.9. The hematocrit was 33.9, white count 7.6, and platelets 405. A morning cortisol was 19.5, with thyroid-stimulating hormone (TSH) 1.7. The patient was treated with 1 μg of intravenous DDAVP along with 75 mL/h of intravenous half-normal saline. After the plasma Na^+ concentration dropped to 116 meq/L, intravenous fluid was switched to normal saline at the same infusion rate. The subsequent results are shown in Table e15-1.

Table e15–1 Serial Laboratory Data for Case 5

Hospital hour	Baseline	0	3	7	11	14	24	48	72
Plasma Na^+ (meq/L)	137	113	115	120	117	116	117	124	130
Creat (mg/dL)	1.2	2.35	2.10	2.02	1.97	1.79	1.53	1.20	1.13
Urine osmolality (mosmol/kg)				319		415	397		
Urine Na^+ (meq/L)				17		23	47		

Approach to Diagnosis

This patient presented with hypovolemic hyponatremia and a "prerenal" reduction in GFR, with an increase in serum creatinine. She had experienced diarrhea for some time and manifested an orthostatic tachycardia after 1 L of normal saline. As expected for hypovolemic hyponatremia, the urine Na^+ concentration was <20 meq/L in the absence of congestive heart failure or other causes of "hypervolemic" hyponatremia, and she responded to saline hydration with an increase in plasma Na^+ concentration and a decrease in creatinine.

The initial hypovolemia increased the sensitivity of this patient's AVP response to osmolality, both decreasing the osmotic threshold for AVP release and increasing the slope of the osmolality response curve. AVP has a half-life of only 10–20 min; therefore, the acute increase in intravascular volume after 1 L of intravenous saline led to a rapid reduction in circulating AVP. The ensuing water diuresis is the primary explanation for the rapid increase in plasma Na^+ concentration in the first 7 h of her hospitalization.

Approach to Management

The key concern in this case was the evident chronicity of the patient's hyponatremia, with several weeks of diarrhea followed by 2–3 d of acute exacerbation. This patient was judged to have "chronic" hyponatremia, i.e., with a suspected duration >48 h; as such, she would be predisposed to osmotic demyelination if she underwent too rapid a correction in her plasma Na^+ concentration, i.e., by more than 8–10 meq/L in 24 h or 18 meq/L in 48 h. At presentation, she had no symptoms that one typically would attribute to acute hyponatremia, and the plasma Na^+ concentration already had increased by a sufficient amount to protect her from cerebral edema; however, she had corrected by 1 meq/L per h within the first 7 h of admission, consistent with impending overcorrection. To reduce or halt the increase in plasma Na^+ concentration, the patient received 1 μg of intravenous DDAVP along with intravenous free water. In light of the hypovolemia and resolving acute renal insufficiency, a decision was made to administer half-normal saline as a source of free water rather than D5W; this was switched to normal saline when plasma Na^+ concentration acutely dropped to 117 meq/L (Table e15-1).

Overcorrection of chronic hyponatremia is a major risk factor for the development of osmotic demyelination syndrome (ODS). Animal studies show a neurologic and survival benefit in ODS of "relowering" plasma Na^+ concentration with DDAVP and free water administration; this approach is demonstrably safe in patients with hyponatremia, with no evident risk of a seizure or other sequelae. This combination can be used to prevent an overcorrection or to relower plasma Na^+ concentration in patients who already have overcorrected. DDAVP is required, since in most of these patients' endogenous AVP levels have plummeted, resulting in a free water diuresis; the administration of free water alone has minimal effect in this setting because of the relative absence of circulating AVP.

This patient's plasma Na^+ concentration remained depressed for several days after DDAVP

administration. It is conceivable that residual hypovolemic hyponatremia attenuated the recovery of the plasma Na^+ concentration. Alternatively, attenuated recovery was due to persistent effects of the single dose of DDAVP. Of note, although the plasma half-life of DDAVP is only 1–2 h, pharmacodynamic studies indicate a much more prolonged effect on urine output and/or urine osmolality. A final consideration is the effect of the patient's initial renal dysfunction on the pharmacokinetics and pharmacodynamics of the administered DDAVP, which is excreted renally; DDAVP should be administered with caution for the reinduction of hyponatremia in patients with chronic kidney disease or acute renal dysfunction.

CASE 6

A 44-year-old woman was referred from a local hospital after presenting with flaccid paralysis. Severe hypokalemia was documented (2.0 meq/L), and an infusion containing KCl was initiated.

Laboratory Data	Value	Units
Na^+	140	meq/L
K^+	2.6	meq/L
Cl^-	115	meq/L
Bicarbonate	15	meq/L
Anion gap	10	meq/L
BUN	22	mg/dL
Creat	1.4	mg/dL
Arterial Blood Gases		
pH	7.32	U
$PaCO_2$	30	mmHg
HCO_3^-	15	meq/L

Additional laboratory data were rheumatoid factor–positive, anti-Ro/SS-A–, and anti-La/SS-B–positive. Urinalysis revealed a pH of 6.0, normal sediment without white or red blood cell casts, and no bacteria. The ratio of urine protein to creatinine was 0.150 g/g. Urinary electrolyte values were Na^+ 35, K^+ 40, Cl^- 18 meq/L. Therefore, the urine anion gap was positive, indicating low urine NH_4^+ excretion.

Approach to Diagnosis

The most accurate diagnosis in this case is classic hypokalemic distal renal tubular acidosis from Sjögren's syndrome. This patient presented with a non-anion-gap metabolic acidosis. The urine anion gap was positive, indicating an abnormally low excretion of ammonium in the face of systemic acidosis. The urine pH was inappropriately alkaline, yet there was no evidence of hypercalciuria, nephrocalcinosis, or bone disease. The patient subsequently was shown to exhibit hyperglobulinemia. These findings, taken together, indicate that the cause of this patient's hypokalemia and non-anion-gap metabolic acidosis was a renal tubular abnormality. The hypokalemia and abnormally low excretion of ammonium, as estimated by the urine anion gap, in the absence of glycosuria, phosphaturia, or aminoaciduria (Fanconi syndrome), defines the entity classical distal renal tubular acidosis (dRTA), also known as type 1 RTA. Because of the hyperglobulinemia, additional serology was obtained, providing evidence for the diagnosis of primary Sjögren's syndrome. Furthermore, additional history indicated a 5-year history of xerostomia and keratoconjunctivitis sicca but without synovitis, arthritis, or rash.

Classical dRTA occurs frequently in patients with Sjögren's syndrome and is a result of an immunologic attack on the collecting tubule, causing failure of H^+-ATPase in the type A intercalated cell. Sjögren's syndrome is one of the best documented acquired causes of classical dRTA. The loss of H^+-ATPase function also occurs with certain inherited forms of classical dRTA. There was no family history in this case, and other family members were not affected. A number of autoantibodies have been associated with Sjögren's syndrome; it is likely that those autoantibodies prevent trafficking or function of the H^+-ATPase in the type A intercalated cell of the collecting tubule. Although proximal RTA has also been reported in patients with Sjögren's syndrome, it is much less common, and there were no features of proximal tubule dysfunction (Fanconi syndrome) in this patient. The hypokalemia is due to secondary hyperaldosteronism from volume depletion.

Approach to Management

The long-term renal prognosis for patients with classic dRTA due to Sjögren's syndrome has not been established. Nevertheless, the metabolic acidosis and the hypokalemia respond to alkali replacement with either sodium citrate solution (Shohl's solution) or sodium bicarbonate tablets. Obviously, potassium deficits must be replaced initially, but potassium replacement usually is not required in a dRTA patient over the long term because sodium bicarbonate (or citrate) therapy expands volume and corrects the secondary hyperaldosteronism. A consequence of the interstitial infiltrate seen in patients with Sjögren's syndrome and classic dRTA is progression of chronic kidney disease. Cytotoxic therapy plus glucocorticoids has been the mainstay of therapy in Sjögren's syndrome for many years, although infiltration of B lymphocytes in salivary gland tissue subsides and urinary acidification improves after treatment with rituximab.

CASE 7

A 32-year-old man was admitted to the hospital with weakness and hypokalemia. The patient had been very healthy until 2 months previously, when he developed intermittent leg weakness. His review of systems was otherwise negative. He denied drug or laxative abuse and was on no medications. Past medical history was unremarkable, with no history of neuromuscular disease. Family history was notable for a sister with thyroid disease. Physical examination was notable only for reduced deep tendon reflexes.

Laboratory Data	Admission	Baseline	Units
Na^+	139	143	meq/L
K^+	2.0	3.8	meq/L
Cl^-	105	107	meq/L
Bicarbonate	26	29	meq/L
BUN	11	16	mg/dL
Creat	0.6	1.0	mg/dL
Ca^{2+}	8.8	8.8	mg/dL
Phosphate	1.2		mg/dL
Albumin	3.8		mg/dL
Plasma osmolality	290		mosmol/kg
Urine osmolality	590		mosmol/kg
Urine K^+	10		meq/L
TSH 0.08 +IU/L (normal 0.2–5.39) Free T4 41 pmol/L (normal 10–27)			

Approach to Diagnosis

This patient developed hypokalemia due to a redistribution of potassium between the intracellular and extracellular compartments; this pathophysiology was readily apparent after calculation of the transtubular potassium gradient. The TTKG is calculated as $(P_{osmol} \times U_{Potassium})/(P_{Potassium} \times U_{osmol})$. The expected values for TTKG are <3 in the presence of hypokalemia and >7–8 in the presence of hyperkalemia (see Case 2 and Case 8). In this case the calculated TTKG was 2.5, consistent with appropriate renal conservation of K^+ and a nonrenal cause for hypokalemia. In the absence of significant gastrointestinal loss of K^+, the patient was diagnosed with a "redistributive" subtype of hypokalemia.

More than 98% of total-body potassium is intracellular; regulated buffering of extracellular K^+ by this large intracellular pool plays a crucial role in the maintenance of a stable plasma K^+ concentration. Clinically, changes in the exchange and distribution of intra- and extracellular K^+ can cause significant hypo- or hyperkalemia. Insulin, β_2-adrenergic activity, thyroid hormone, and alkalosis promote cellular uptake of K^+ by multiple interrelated mechanisms, leading to hypokalemia. In particular, alterations in the activity of the endogenous sympathetic nervous system can cause hypokalemia in several settings, including alcohol withdrawal, hyperthyroidism, acute myocardial infarction, and severe head injury.

Weakness is common in severe hypokalemia; hypokalemia causes hyperpolarization of muscle,

impairing the capacity to depolarize and contract. In this patient, Graves' disease caused hyperthyroidism and hypokalemic paralysis [thyrotoxic periodic paralysis (TPP)]. TPP develops more frequently in patients of Asian or Hispanic origin. This predisposition has been linked to genetic variation in Kir2.6, a muscle-specific thyroid hormone–induced K^+ channel; however, the pathophysiologic mechanisms that link dysfunction of this ion channel to TPP have not been elucidated. The hypokalemia in TPP is attributed to both direct and indirect activation of Na^+, K^+-ATPase by thyroid hormone, resulting in increased uptake of K^+ by muscle and other tissues. Thyroid hormone induces expression of multiple subunits of Na^+, K^+-ATPase in skeletal muscle, increasing the capacity for uptake of K^+; hyperthyroid increases in β-adrenergic activity are also thought to play an important role in TPP.

Clinically, patients with TPP present with weakness of the extremities and limb girdle, with paralytic episodes that occur most frequently between 1 and 6 a.m. Precipitants of weakness include high carbohydrate loads and strenuous exercise. Signs and symptoms of hyperthyroidism are not always present, often leading to delays in diagnosis. Hypokalemia is often profound and usually is accompanied by redistributive hypophosphatemia, as in this case. A TTKG <2–3 separates patients with TPP from those with hypokalemia due to renal potassium wasting, who have TTKG values >4. This distinction is of considerable importance for therapy; patients with large potassium deficits require aggressive repletion with K^+-Cl^-, which confers a significant risk of rebound hyperkalemia in TPP and related disorders.

Approach to Management

Ultimately, definitive therapy for TPP requires treatment of the associated hyperthyroidism. In the short term, however, potassium replacement is necessary to hasten muscle recovery and prevent cardiac arrhythmias. The average recovery time for an acute attack is reduced by ~50% in patients treated with intravenous K^+-Cl^- at a rate of 10 meq/h; however, this incurs a significant risk of rebound hyperkalemia, with up to 70% developing a plasma K^+ concentration >5.0 meq/L. This potential for rebound hyperkalemia is a general problem in the management of all causes of "redistributive hypokalemia," resulting in the need to distinguish these patients accurately and rapidly from those with a large K^+ deficit due to renal or extrarenal loss of K^+. An attractive alternative to K^+-Cl^- replacement in TPP is treatment with high-dose propranolol (3 mg/kg), which rapidly reverses the associated hypokalemia, hypophosphatemia, and paralysis. Notably, rebound hyperkalemia is not associated with this treatment.

CASE 8

A 66-year-old man was admitted to the hospital with a plasma K^+ concentration of 1.7 meq/L and profound weakness. The patient had noted progressive weakness over several days, to the point where he was unable to rise from bed. Past medical history was notable for small cell lung cancer with metastases to brain, liver, and adrenals. The patient had been treated with one cycle of cisplatin/etoposide 1 year before this admission, complicated by acute kidney injury (peak creatinine of 5 mg/dL, with residual chronic kidney disease), and three subsequent cycles of cyclophosphamide/doxorubicin/vincristine, in addition to 15 treatments with whole-brain radiation.

On physical examination the patient was jaundiced. Blood pressure (bp) was 130/70, increasing to 160/98 after 1 L of saline, with a JVP of 8 cm. There was generalized muscle weakness.

Laboratory Data	2 Months PTA	Admission	HD2	Units
Na^+	143	149	144	meq/L
K^+	3.7	1.7	3.5	meq/L
Cl^-	103	84	96	meq/L
Bicarbonate	26	44	34	meq/L
Venous pH		7.47		pH
Venous PCO_2		62		mm Hg
BUN	21	41	40	mg/dL
Creat	2.8	2.9	2.3	mg/dL
Mg^{2+}	1.3	1.6	2.4	mg/dL

CPK		183		U/L
ALT	8	75		U/L
Albumin	3.4	2.8	2.3	
Adjusted anion gap	15	24	18	
Total bilirubin	0.65	5.19		mg/dL
Alkaline phosphatase	93	217		U/L
Urine Na^+		35	28	meq/L
Urine K^+		25	49	meq/L
Urine chloride		48	51	meq/L
Urine osmolality		391		mosmol/kg
Plasma osmolality		312		mosmol/kg
Urine pH		5.5		
Plasma ACTH		185		pg/mL (7–50 pg/mL)
Plasma cortisol 24-h urine cortisol		1044		μg/24 h (4–50 μg/24 h)

The patient's hospital course was complicated by acute respiratory failure attributed to pulmonary embolism; he died 2 weeks after admission.

Approach to Diagnosis

Why was this patient hypokalemic? Why was he weak? Why did he have an alkalosis?

This patient had metastatic small cell lung cancer that was persistent despite several rounds of chemotherapy and radiotherapy. He presented with profound hypokalemia, alkalosis, hypertension, severe weakness, jaundice, and worsening liver function tests.

With respect to the hypokalemia, there was no evident cause of nonrenal potassium loss, e.g., diarrhea. The urinary TTKG was 11.7, with a plasma K^+ concentration of 1.7 meq/L; this TTKG value is consistent with inappropriate renal K^+ secretion despite severe hypokalemia. The TTKG is calculated as $(P_{osmol} \times U_{Potassium})/(P_{Potassium} \times U_{osmol})$. The expected values for the TTKG are <3 in the presence of hypokalemia and >7–8 in the presence of hyperkalemia (see Case 2 and Case 6).

There were several explanations for the excessive renal loss of potassium. First, the patient had had a history of cisplatin-associated acute kidney injury with residual chronic kidney disease. Cisplatin can cause persistent renal tubular defects, with prominent hypokalemia and hypomagnesemia; however, this patient had not previously required potassium or magnesium repletion, suggesting that cisplatin-associated renal tubular defects did not play a major role in this presentation with severe hypokalemia. Second, he was hypomagnesemic on presentation, suggesting total-body magnesium depletion.

Magnesium depletion has inhibitory effects on muscle Na^+,K^+-ATPase activity, reducing influx into muscle cells and causing a secondary increase in K^+ excretion. Magnesium depletion also increases K^+ secretion by the distal nephron; this is attributed to a reduction in the magnesium-dependent intracellular block of K^+ efflux through the secretory K^+ channel of principal cells [renal outer medullary K^+ channel (ROMK); Fig. e15-1]. Clinically, hypomagnesemic patients are refractory to K^+ replacement in the absence of Mg^{2+} repletion. Again, however, this patient had not previously developed significant hypokalemia despite periodic hypomagnesemia, and so other factors must have caused the severe hypokalemia.

The associated hypertension in this case suggested an increase in mineralocorticoid activity, causing increased activity of ENaC channels in principal cells, NaCl retention, hypertension, and hypokalemia. The increase an ENaC-mediated Na^+ transport in principal cells would have led to an increase in the lumen-negative potential difference in the connecting tubule and cortical collecting duct, driving an increase in K^+ secretion through apical K^+ channels (Fig. e15-1). This explanation is compatible with the very high TTKG, i.e., an increase in K^+ excretion that is inappropriate for the plasma K^+ concentration.

What caused an increase in mineralocorticoid activity in this patient? The patient had bilateral adrenal metastases, indicating that primary hyperaldosteronism was unlikely. The clinical presentation (hypokalemia, hypertension, and alkalosis) and the history of small cell lung cancer (SCLC) suggested Cushing's syndrome, with a massive increase in circulating glucocorticoids, in response to ectopic adrenocorticotropic hormone (ACTH) secretion by the SCLC tumor. Confirmation of this diagnosis was

provided by a very high plasma cortisol level, high ACTH level, and increased urinary cortisol (see the laboratory data, above).

Why would an increase in circulating cortisol cause an apparent increase in mineralocorticoid activity? Cortisol and aldosterone have equal affinity for the mineralocorticoid receptor (MLR); thus, cortisol has "mineralocorticoid-like" activity. However, cells in the aldosterone-sensitive distal nephron [the distal convoluted tubule (DCT), connecting tubule (CNT), and collecting duct] are protected from circulating cortisol by the enzyme 11β-hydroxysteroid dehydrogenase-2 (11βHSD-2), which converts cortisol to cortisone (**Fig. e15-2**); cortisone has minimal affinity for the MLR. Activation of the mineralocorticoid receptor causes activation of the basolateral Na^+, K^+-ATPase, activation of the thiazide-sensitive Na^+-Cl^- co-transporter in the DCT, and activation of apical ENaC channels in principal cells of the CNT and collecting duct (Fig. e15-2). Recessive loss-of-function mutations in the *11βHSD-2* gene lead to cortisol-dependent activation of the MLR and the syndrome of apparent mineralocorticoid excess (SAME), which consists of hypertension, hypokalemia, hypercalciuria, and metabolic alkalosis, with suppressed plasma renin activity (PRA) and suppressed aldosterone. A similar syndrome is caused by biochemical inhibition of 11βHSD-2 by glycyrrhetinic/glycyrrhizinic acid (found in licorice, for example) and/or carbenoxolone.

Figure e15-2

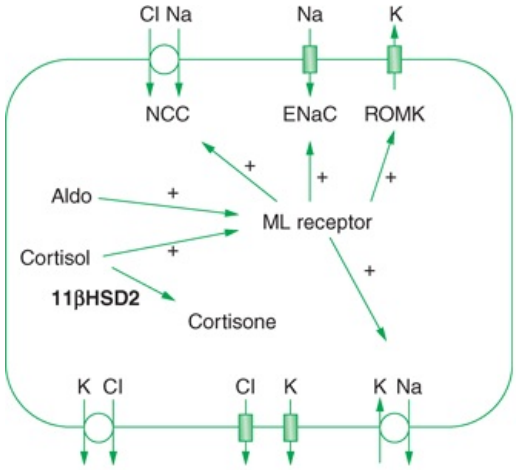

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

11β-hydroxysteroid dehydrogenase-2 and syndromes of apparent mineralocorticoid excess. The enzyme 11β-hydroxysteroid dehydrogenase-2 (11βHSD-2) protects cells in the aldosterone-sensitive distal nephron [the distal convoluted tubule (DCT), connecting tubule (CNT), and collecting duct] from the illicit activation of mineralocorticoid receptors (MLR) by cortisol. Binding of aldosterone to the MLR leads to activation of the thiazide-sensitive Na^+-Cl^- co-transporter in DCT cells and the amiloride-sensitive epithelial sodium channel (ENaC) in principal cells (CNT and collecting duct). Aldosterone also activates basolateral Na^+, K^+-ATPase and, to a lesser extent, the apical secretory K^+ channel ROMK (renal outer medullary K^+ channel). Cortisol has equivalent affinity for the MLR to that of aldosterone; metabolism of cortisol to cortisone, which has no affinity for the MLR, prevents these cells from activation by circulating cortisol. Genetic deficiency of 11βHSD-2 or inhibition of its activity causes the syndromes of apparent mineralocorticoid excess (see Case 8).

In Cushing's syndrome caused by increases in pituitary ACTH, the incidence of hypokalemia is only 10%, whereas it is ~70% in patients with ectopic secretion of ACTH despite a similar incidence of hypertension. The activity of renal 11βHSD-2 is reduced in patients with ectopic ACTH compared with Cushing's syndrome, resulting in a syndrome of apparent mineralocorticoid excess; the prevailing theory is that the much greater cortisol production in ectopic ACTH syndromes overwhelms the renal 11βHSD-2 enzyme, resulting in activation of renal MLRs by unmetabolized cortisol (Fig. e15-2).

Why was the patient so weak? The patient was profoundly weak due to the combined effect of hypokalemia and increased cortisol. Hypokalemia causes hyperpolarization of muscle, impairing the capacity to depolarize and contract. Weakness and even ascending paralysis frequently can complicate severe hypokalemia. Hypokalemia also causes a myopathy and predisposes to rhabdomyolysis; notably, however, the patient had a normal creatine phosphokinase (CPK) level. Cushing's syndrome often is accompanied by a proximal myopathy due to the protein-wasting effects of cortisol excess.

The patient presented with a mixed acid-base disorder, with a significant metabolic alkalosis and a bicarbonate concentration of 44 meq/L. A venous blood gas was drawn soon after his presentation; venous and arterial blood gases demonstrate a high level of agreement in hemodynamically stable

patients, allowing for the interpretation of acid-base disorders with venous blood gas results. In response to his metabolic alkalosis, the Pco_2 should have increased by 0.75 mmHg for each 1 meq/L increase in bicarbonate; the expected Pco_2 should have been ~55 mmHg. In light of the Pco_2 of 62 mmHg, he had an additional respiratory acidosis, probably caused by respiratory muscle weakness from the acute hypokalemia and subacute hypercortisolism.

The patient's albumin-adjusted anion gap was 21 + [(4–2.8) x 2.5] = 24; this suggests a third acid-base disorder: anion-gap acidosis. Notably, the measured anion gap can increase in alkalosis due to both increases in plasma protein concentrations (in hypovolemic alkalosis) and the alkalemia-associated increase in net negative charge of plasma proteins, with both causing an increase in "unmeasured anions." However, this patient was neither volume-depleted nor particularly alkalemic, suggesting that these effects played a minimal role in the increased anion gap. Alkalosis also stimulates an increase in lactic acid production due to activation of phosphofructokinase and accelerated glycolysis; unfortunately, however, a lactic acid level was not measured in this patient. It should be noted in this regard that alkalosis typically increases lactic acid levels by a mere 1.5–3 meq/L and that the patient was not significantly alkalemic. Regardless of the underlying pathophysiology, the increased anion gap probably was related to the metabolic alkalosis in light of the fact that the anion gap had decreased to 18 by hospital day 2, coincident with a reduction in plasma bicarbonate.

Why did the patient have a metabolic alkalosis? The activation of mineralocorticoid receptors in the distal nephron increases distal nephron acidification and net acid secretion. In consequence, mineralocorticoid excess causes a saline-resistant metabolic alkalosis, which is exacerbated significantly by the development of hypokalemia. Hypokalemia plays a key role in the generation of most forms of metabolic alkalosis, stimulating proximal tubular ammonium production, proximal tubular bicarbonate reabsorption, and distal tubular H^+,K^+-ATPase activity.

Approach to Management

The first priority in the management of this patient was to increase his plasma K^+ and magnesium concentrations rapidly; hypomagnesemic patients are refractory to K^+ replacement alone, and there is a need to correct hypomagnesemia immediately. This was accomplished via the administration of both oral and intravenous K^+-Cl^-, giving a total of 240 meq over the first 18 h; 5 g of intravenous magnesium sulfate also was administered. Multiple 100-mL "minibags" of saline containing 20 meq each were infused, with cardiac monitoring and frequent measurement of plasma electrolytes. Of note, intravenous K^+-Cl^- should always be given in saline solutions, since dextrose-containing solutions can increase insulin levels and exacerbate hypokalemia.

This case illustrates the difficulty in predicting the whole-body deficit of K^+ in hypokalemic patients. In the absence of abnormal K^+ redistribution, the total deficit correlates with plasma K^+ concentration, which drops by approximately 0.27 mM for every 100-mmol reduction in total-body stores; this would suggest a deficit of ~650 meq of K^+ in this patient at the admission plasma K^+ concentration of 1.7 meq/L. Notably, however, alkalemia induces a modest intracellular shift of circulating K^+, and so this patient's initial plasma K^+ concentration was not an ideal indicator of the total potassium deficit. Regardless of the underlying pathophysiology in this case, close monitoring of plasma K^+ concentration is always essential during the correction of severe hypokalemia to gauge the adequacy of repletion and avoid overcorrection.

Subsequent management of this patient's Cushing's syndrome and ectopic ACTH secretion was complicated by the respiratory issues. The prognosis in patients with ectopic ACTH secretion depends on the tumor histology and the presence or absence of distant metastases. This patient had an exceptionally poor prognosis with widely metastatic small cell lung cancer that had failed treatment; other patients with ectopic ACTH secretion caused by more benign, isolated tumors, most commonly bronchial carcinoid tumors, have a much better prognosis. In the absence of successful surgical resection of the causative tumor, management of this syndrome can include surgical adrenalectomy or medical therapy to block adrenal steroid production.

CASE 9

A stuporous 22-year-old man was admitted with a history of behaving strangely. His friends indicated he had experienced recent emotional problems stemming from a failed relationship and had threatened suicide. There was a history of alcohol abuse, but his friends were unaware of recent alcohol consumption. The patient was obtunded on admission, with no evident focal neurologic deficits. The remainder of the physical examination was unremarkable.

Laboratory data:

Na^+	140	meq/L

K^+	5	meq/L
Cl^-	95	meq/L
HCO_3^-	10	meq/L
Glucose	125	mg/dL
BUN	15	mg/dL
Creat	0.9	mg/dL
Ionized Ca^{2+}	4.0	mg/dL
Plasma osmolality	325	mosmol kg/H_2O

Urinalysis revealed crystalluria, with a mixture of envelope-shaped and needle-shaped crystals.

Approach to Diagnosis

This patient presented with CNS manifestations and a history of suspicious behavior, suggesting ingestion of a toxin. The anion gap was strikingly elevated at 35 meq/L. The δAG of 25 significantly exceeded the δHCO_3^- of 15. The fact that the delta values are significantly disparate indicates that the most likely acid-base diagnosis in this patient is a mixed high-anion-gap metabolic acidosis and a metabolic alkalosis. The metabolic alkalosis in this case may have been the result of vomiting. Nevertheless, the most useful finding is that the osmolar gap was elevated. The osmolar gap of 33 (difference in measured and calculated osmolality or 325–292) in the face of a high-gap metabolic acidosis is diagnostic of an osmotically active metabolite in plasma; a difference of >10 mosmol/kg indicates a significant concentration of an unmeasured osmolyte. Examples of toxic osmolytes include ethylene glycol, diethylene glycol, methanol, and propylene glycol.

Several caveats apply to the interpretation of the osmolar and anion gaps in the differential diagnosis of toxic alcohol ingestions. First, unmeasured, neutral osmolytes also can accumulate in lactic acidosis and alcoholic ketoacidosis; i.e., an elevated osmolar gap is not specific to anion-gap acidoses associated with toxic alcohol ingestions. Second, patients can present having extensively metabolized the ingested toxin, with an insignificant osmolar gap but a large anion gap; i.e., the absence of an elevated osmolar gap does not rule out toxic alcohol ingestion. Third, the converse can be seen in patients who present earlier after ingestion of the toxin, i.e., a large osmolar gap with minimal elevation of the anion gap. Finally, clinicians should be aware of the effect of coingested ethanol, which can itself elevate the osmolar gap and can reduce metabolism of the toxic alcohols via competitive inhibition of alcohol dehydrogenase (see below), thus attenuating the expected increase in the anion gap.

Ethylene glycol is commonly available as antifreeze or solvents and may be ingested accidentally or as a suicide attempt. The metabolism of ethylene glycol by alcohol dehydrogenase generates acids such as glycoaldehyde, glycolic acid, and oxalic acid. The initial effects of intoxication on the CNS in the earliest stages mimic inebriation but may quickly progress to full-blown coma. Delay in treatment is one of the most common causes of mortality with toxic alcohol poisoning. The kidney shows evidence of acute tubular injury with widespread deposition of calcium oxalate crystals within tubular epithelial cells. Cerebral edema is common, as is crystal deposition in the brain; the deposition is irreversible.

The co-occurrent crystalluria is typical of ethylene glycol intoxication; both needle-shaped monohydrate and envelope-shaped dihydrate calcium oxalate crystals can be seen in the urine as the process evolves. Circulating oxalate also can complex with plasma calcium, reducing the ionized calcium, as in this case.

Ethylene glycol intoxication should be verified by measuring ethylene glycol levels. However, therapy must be initiated immediately in this life-threatening situation. Although therapy can be initiated with confidence in cases with known or witnessed ingestions, such histories are rarely available. Therapy thus should be initiated in patients with severe metabolic acidosis and elevated anion and osmolar gaps. Other diagnostic features, such as hypocalcemia and acute renal failure with crystalluria, can provide important confirmation for urgent, empirical therapy.

Approach to Management

Since all four osmotically active toxic alcohols—ethylene glycol, diethylene glycol, methanol, and propylene glycol—are metabolized by alcohol dehydrogenase to generate toxic products, competitive inhibition of this key enzyme is common to the treatment of all four intoxications. The most potent

inhibitor of alcohol dehydrogenase, and the drug of choice in this circumstance, is fomepizole (4-methyl pyrazole). Fomepizole should be administered intravenously as a loading dose (15 mg/kg) followed by doses of 10 mg/kg every 12 h for four doses and then 15 mg/kg every 12 h thereafter until ethylene glycol levels have been reduced to <20 mg/dL and the patient is asymptomatic with a normal pH. Additional very important components of the treatment of toxic alcohol ingestion include fluid resuscitation, thiamine, pyridoxine, folate, sodium bicarbonate, and hemodialysis. Hemodialysis is used to remove both the parent compound and toxic metabolites but also removes administered fomepizole, necessitating adjustment of dose frequency. Gastric aspiration, induced emesis, or the use of activated charcoal is effective only if initiated within 30–60 min after ingestion of the toxin. When fomepizole is not available, ethanol, which has more than tenfold affinity for alcohol dehydrogenase compared with other alcohols, may be substituted and is quite effective. Ethanol must be administered IV to achieve a blood level of 22 meq/L (100 mg/dL). A disadvantage of ethanol is the obtundation that follows its administration, which is additive to the CNS effects of ethylene glycol. Furthermore, if hemodialysis is utilized, the infusion rate of ethanol must be increased because it is dialyzed rapidly. In general, hemodialysis is indicated for all patients with ethylene glycol intoxication when the arterial pH is <7.3 or the osmolar gap exceeds 20 mosmol/kg H_2O.

FURTHER READINGS

DUBOSE TD Jr: Acid-base disorders, in *Brenner and Rector's The Kidney* , 8th ed, B Brenner (ed), Philadelphia, W.B. Saunders, 2008

———: A 42-year-old woman with flaccid paralysis. Am J Kidney Dis 54:965, 2009

———: Acidosis and alkalosis, in *Harrison's Principles of Internal Medicine* , 18th ed, DL Long et al (eds). New York, McGraw-Hill, 2012

KRAUT JA, KURTZ I: Toxic alcohol ingestions: Clinical features, diagnosis, and management. Clin J Am Soc Nephrol 3:208, 2008[PMID: 18045860]

———, Madias NE: Serum anion gap: Its uses and limitations in clinical medicine. Clin J Am Soc Nephrol 2:162, 2007

MOUNT DB: Fluid and electrolyte disorders, in *Harrison's Principles of Internal Medicine* , 18th ed, DL Long et al (eds). New York, McGraw-Hill, 2012

———, Krahn ta: Hyponatremia: Case vignettes. Semin Nephrol 29:300, 2009

———, Zandi-nejad K: Disorders of potassium balance, in *Brenner and Rector's The Kidney* , 8th ed, B Brenner (ed). Philadelphia, W.B. Saunders, 2008, pp 547–558

PERIANAYAGAM A et al: DDAVP is effective in preventing and reversing inadvertent overcorrection of hyponatremia. Clin J Am Soc Nephrol 3:331, 2008[PMID: 18235152]

SCHRIER RW et al: Tolvaptan, a selective oral vasopressin V2-receptor antagonist, for hyponatremia. N Engl J Med 355:2099, 2006[PMID: 17105757]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HYPERCALCEMIA AND HYPOCALCEMIA: INTRODUCTION

The calcium ion plays a critical role in normal cellular function and signaling, regulating diverse physiologic processes such as neuromuscular signaling, cardiac contractility, hormone secretion, and blood coagulation. Thus, extracellular calcium concentrations are maintained within an exquisitely narrow range through a series of feedback mechanisms that involve parathyroid hormone (PTH) and the active vitamin D metabolite 1,25-dihydroxyvitmin D [1,25$(OH)_2$D]. These feedback mechanisms are orchestrated by integrating signals between the parathyroid glands, kidney, intestine, and bone (Fig. 46-1; Chap. 352).

Figure 46-1

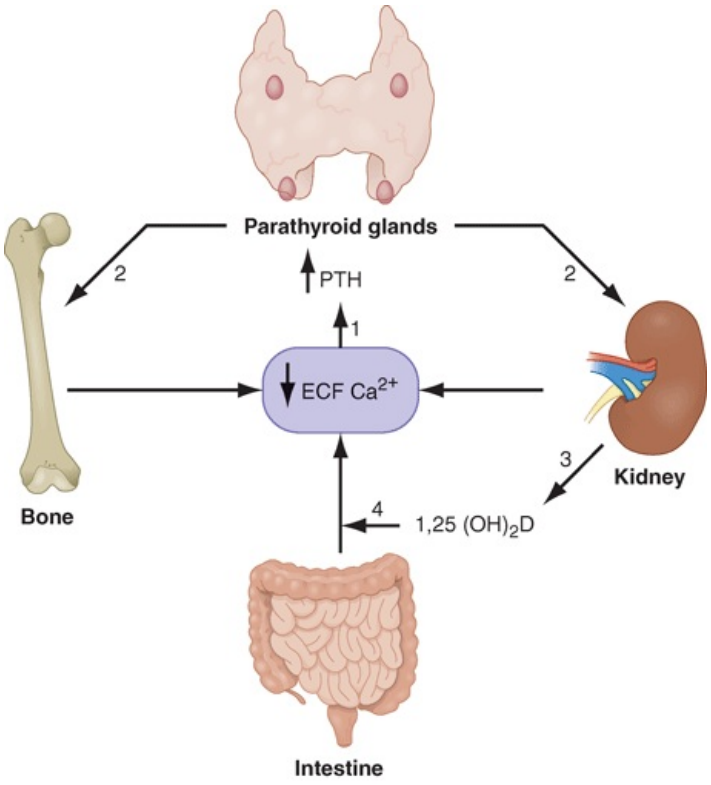

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Feedback mechanisms maintaining extracellular calcium concentrations within a narrow, physiologic range [8.9–10.1 mg/dL (2.2–2.5 m*M*)]. A decrease in extracellular (ECF) calcium (Ca^{2+}) triggers an increase in parathyroid hormone (PTH) secretion (1) via the calcium sensor receptor on parathyroid cells. PTH, in turn, results in increased tubular reabsorption of calcium by the kidney (2) and resorption of calcium from bone (2) and also stimulates renal 1,25$(OH)_2$D production (3). 1,25$(OH)_2$D, in turn, acts principally on the intestine to increase calcium absorption (4). Collectively, these homeostatic mechanisms serve to restore serum calcium levels to normal.

Disorders of serum calcium concentration are relatively common and often serve as a harbinger of underlying disease. This chapter provides a brief summary of the approach to patients with altered serum calcium levels. **See Chap. 353 for a detailed discussion of this topic.**

HYPERCALCEMIA

Etiology

The causes of hypercalcemia can be understood and classified based on derangements in the normal feedback mechanisms that regulate serum calcium (Table 46-1). Excess PTH production, which is not

appropriately suppressed by increased serum calcium concentrations, occurs in primary neoplastic disorders of the parathyroid glands (parathyroid adenomas; hyperplasia; or, rarely, carcinoma) that are associated with increased parathyroid cell mass and impaired feedback inhibition by calcium. Inappropriate PTH secretion for the ambient level of serum calcium also occurs with heterozygous inactivating calcium sensor receptor (CaSR) mutations, which impair extracellular calcium sensing by the parathyroid glands and the kidneys, resulting in familial hypocalciuric hypercalcemia (FHH). Although PTH secretion by tumors is extremely rare, many solid tumors produce PTH-related peptide (PTHrP), which shares homology with PTH in the first 13 amino acids and binds the PTH receptor, thus mimicking effects of PTH on bone and the kidney. In PTHrP-mediated hypercalcemia of malignancy, PTH levels are suppressed by the high serum calcium levels. Hypercalcemia associated with granulomatous disease (e.g., sarcoidosis) or lymphomas is caused by enhanced conversion of 25(OH)D to the potent $1,25(OH)_2D$. In these disorders, $1,25(OH)_2D$ enhances intestinal calcium absorption, resulting in hypercalcemia and suppressed PTH. Disorders that directly increase calcium mobilization from bone, such as hyperthyroidism or osteolytic metastases, also lead to hypercalcemia with suppressed PTH secretion as does exogenous calcium overload, as in milk-alkali syndrome, or total parenteral nutrition with excessive calcium supplementation.

Table 46-1 Causes of Hypercalcemia

Excessive PTH production
Primary hyperparathyroidism (adenoma, hyperplasia, rarely carcinoma)
Tertiary hyperparathyroidism (long-term stimulation of PTH secretion in renal insufficiency)
Ectopic PTH secretion (very rare)
Inactivating mutations in the CaSR (FHH)
Alterations in CaSR function (lithium therapy)
Hypercalcemia of malignancy
Overproduction of PTHrP (many solid tumors)
Lytic skeletal metastases (breast, myeloma)
Excessive 1,25(OH)2D production
Granulomatous diseases (sarcoidosis, tuberculosis, silicosis)
Lymphomas
Vitamin D intoxication
Primary increase in bone resorption
Hyperthyroidism
Immobilization
Excessive calcium intake
Milk-alkali syndrome
Total parenteral nutrition
Other causes
Endocrine disorders (adrenal insufficiency, pheochromocytoma, VIPoma)
Medications (thiazides, vitamin A, antiestrogens)

Abbreviations: CaSR, calcium sensor receptor; FHH, familial hypocalciuric hypercalcemia; PTH, parathyroid hormone; PTHrP, PTH-related peptide.

Clinical Manifestations

Mild hypercalcemia (up to 11–11.5 mg/dL) is usually asymptomatic and recognized only on routine calcium measurements. Some patients may complain of vague neuropsychiatric symptoms, including trouble concentrating, personality changes, or depression. Other presenting symptoms may include peptic ulcer disease or nephrolithiasis, and fracture risk may be increased. More severe hypercalcemia (>12–13 mg/dL), particularly if it develops acutely, may result in lethargy, stupor, or coma, as well as gastrointestinal symptoms (nausea, anorexia, constipation, or pancreatitis). Hypercalcemia decreases renal concentrating ability, which may cause polyuria and polydipsia. With long-standing hyperparathyroidism, patients may present with bone pain or pathologic fractures. Finally, hypercalcemia can result in significant electrocardiographic changes, including bradycardia, AV block, and short QT interval; changes in serum calcium can be monitored by following the QT interval (Fig. 228-16).

Diagnostic Approach

The first step in the diagnostic evaluation of hyper- or hypocalcemia is to ensure that the alteration in serum calcium levels is not due to abnormal albumin concentrations. About 50% of total calcium is

ionized, and the rest is bound principally to albumin. Although direct measurements of ionized calcium are possible, they are easily influenced by collection methods and other artifacts; thus, it is generally preferable to measure total calcium and albumin to "correct" the serum calcium. When serum albumin concentrations are reduced, a corrected calcium concentration is calculated by adding 0.2 m*M* (0.8 mg/dL) to the total calcium level for every decrement in serum albumin of 1.0 g/dL below the reference value of 4.1 g/dL for albumin, and, conversely, for elevations in serum albumin.

A detailed history may provide important clues regarding the etiology of the hypercalcemia (Table 46-1). Chronic hypercalcemia is most commonly caused by primary hyperparathyroidism, as opposed to the second most common etiology of hypercalcemia, an underlying malignancy. The history should include medication use, previous neck surgery, and systemic symptoms suggestive of sarcoidosis or lymphoma.

Once true hypercalcemia is established, the second most important laboratory test in the diagnostic evaluation is a PTH level using a two-site assay for the intact hormone. Increases in PTH are often accompanied by hypophosphatemia. In addition, serum creatinine should be measured to assess renal function; hypercalcemia may impair renal function, and renal clearance of PTH may be altered depending on the fragments detected by the assay. If the PTH level is increased (or "inappropriately normal") in the setting of elevated calcium and low phosphorus, the diagnosis is almost always primary hyperparathyroidism. Because individuals with familial hypocalciuric hypercalcemia (FHH) may also present with mildly elevated PTH levels and hypercalcemia, this diagnosis should be considered and excluded because parathyroid surgery is ineffective in this condition. A calcium/creatinine clearance ratio (calculated as urine calcium/serum calcium divided by urine creatinine/serum creatinine) of <0.01 is suggestive of FHH, particularly when there is a family history of mild, asymptomatic hypercalcemia. In addition, a number of laboratories are now offering sequence analysis of the CaSR gene for the definitive diagnosis of FHH. Ectopic PTH secretion is extremely rare.

A suppressed PTH level in the face of hypercalcemia is consistent with non-parathyroid-mediated hypercalcemia, most often due to underlying malignancy. Although a tumor that causes hypercalcemia is generally overt, a PTHrP level may be needed to establish the diagnosis of hypercalcemia of malignancy. Serum $1,25(OH)_2D$ levels are increased in granulomatous disorders, and clinical evaluation in combination with laboratory testing will generally provide a diagnosis for the various disorders listed in Table 46-1.

Treatment: Hypercalcemia

Mild, asymptomatic hypercalcemia does not require immediate therapy, and management should be dictated by the underlying diagnosis. By contrast, significant, symptomatic hypercalcemia usually requires therapeutic intervention independent of the etiology of hypercalcemia. Initial therapy of significant hypercalcemia begins with volume expansion because hypercalcemia invariably leads to dehydration; 4–6 L of intravenous saline may be required over the first 24 h, keeping in mind that underlying comorbidities (e.g., congestive heart failure) may require the use of loop diuretics to enhance sodium and calcium excretion. However, loop diuretics should not be initiated until the volume status has been restored to normal. If there is increased calcium mobilization from bone (as in malignancy or severe hyperparathyroidism), drugs that inhibit bone resorption should be considered. Zoledronic acid (e.g., 4 mg intravenously over ~30 min), pamidronate (e.g., 60–90 mg intravenously over 2–4 h), and etidronate (e.g., 7.5 mg/kg per day for 3–7 consecutive days) are approved by the U.S. Food and Drug Administration for the treatment of hypercalcemia of malignancy in adults. Onset of action is within 1–3 days, with normalization of serum calcium levels occurring in 60–90% of patients. Bisphosphonate infusions may need to be repeated if hypercalcemia relapses. Because of their effectiveness, bisphosphonates have replaced calcitonin or plicamycin, which are rarely used in current practice for the management of hypercalcemia. In rare instances, dialysis may be necessary. Finally, while intravenous phosphate chelates calcium and decreases serum calcium levels, this therapy can be toxic because calcium-phosphate complexes may deposit in tissues and cause extensive organ damage.

In patients with $1,25(OH)_2D$-mediated hypercalcemia, glucocorticoids are the preferred therapy, as they decrease $1,25(OH)_2D$ production. Intravenous hydrocortisone (100–300 mg daily) or oral prednisone (40–60 mg daily) for 3–7 days are used most often. Other drugs, such as ketoconazole, chloroquine, and hydroxychloroquine, may also decrease $1,25(OH)_2D$ production and are used occasionally.

HYPOCALCEMIA

Etiology

The causes of hypocalcemia can be differentiated according to whether serum PTH levels are low (hypoparathyroidism) or high (secondary hyperparathyroidism). Although there are many potential causes of hypocalcemia, impaired PTH or vitamin D production are the most common etiologies (Table 46-2) (Chap. 353). Because PTH is the main defense against hypocalcemia, disorders associated with deficient PTH production or secretion may be associated with profound, life-

threatening hypocalcemia. In adults, hypoparathyroidism most commonly results from inadvertent damage to all four glands during thyroid or parathyroid gland surgery. Hypoparathyroidism is a cardinal feature of autoimmune endocrinopathies (Chap. 351); rarely, it may be associated with infiltrative diseases such as sarcoidosis. Impaired PTH secretion may be secondary to magnesium deficiency or to activating mutations in the CaSR, which suppress PTH, leading to effects that are opposite to those that occur in FHH.

Table 46-2 Causes of Hypocalcemia

Low Parathyroid Hormone Levels (Hypoparathyroidism)
Parathyroid agenesis
Isolated
DiGeorge syndrome
Parathyroid destruction
Surgical
Radiation
Infiltration by metastases or systemic diseases
Autoimmune
Reduced parathyroid function
Hypomagnesemia
Activating CaSR mutations
High Parathyroid Hormone Levels (Secondary Hyperparathyroidism)
Vitamin D deficiency or impaired $1,25(OH)_2D$ production/action
Nutritional vitamin D deficiency (poor intake or absorption)
Renal insufficiency with impaired $1,25(OH)_2D$ production
Vitamin D resistance, including receptor defects
Parathyroid hormone resistance syndromes
PTH receptor mutations
Pseudohypoparathyroidism (G protein mutations)
Drugs
Calcium chelators
Inhibitors of bone resorption (bisphosphonates, plicamycin)
Altered vitamin D metabolism (phenytoin, ketoconazole)
Miscellaneous causes
Acute pancreatitis
Acute rhabdomyolysis
Hungry bone syndrome after parathyroidectomy
Osteoblastic metastases with marked stimulation of bone formation (prostate cancer)

Abbreviations: CaSR, calcium sensor receptor; PTH, parathyroid hormone.

Vitamin D deficiency, impaired $1,25(OH)_2D$ production (primarily secondary to renal insufficiency), or vitamin D resistance also cause hypocalcemia. However, the degree of hypocalcemia in these disorders is generally not as severe as that seen with hypoparathyroidism because the parathyroids are capable of mounting a compensatory increase in PTH secretion. Hypocalcemia may also occur in conditions associated with severe tissue injury such as burns, rhabdomyolysis, tumor lysis, or pancreatitis. The cause of hypocalcemia in these settings may include a combination of low albumin, hyperphosphatemia, tissue deposition of calcium, and impaired PTH secretion.

Clinical Manifestations

Patients with hypocalcemia may be asymptomatic if the decreases in serum calcium are relatively mild and chronic, or they may present with life-threatening complications. Moderate to severe hypocalcemia is associated with paresthesias, usually of the fingers, toes, and circumoral regions, and is caused by increased neuromuscular irritability. On physical examination, a Chvostek's sign (twitching of the circumoral muscles in response to gentle tapping of the facial nerve just anterior to the ear) may be elicited, although it is also present in ~10% of normal individuals. Carpal spasm may be induced by inflation of a blood pressure cuff to 20 mmHg above the patient's systolic blood pressure for 3 min (Trousseau's sign). Severe hypocalcemia can induce seizures, carpopedal spasm, bronchospasm,

laryngospasm, and prolongation of the QT interval.

Diagnostic Approach

In addition to measuring serum calcium, it is useful to determine albumin, phosphorus, and magnesium levels. As for the evaluation of hypercalcemia, determining the PTH level is central to the evaluation of hypocalcemia. A suppressed (or "inappropriately low") PTH level in the setting of hypocalcemia establishes absent or reduced PTH secretion (hypoparathyroidism) as the cause of the hypocalcemia. Further history will often elicit the underlying cause (i.e., parathyroid agenesis vs. destruction). By contrast, an elevated PTH level (secondary hyperparathyroidism) should direct attention to the vitamin D axis as the cause of the hypocalcemia. Nutritional vitamin D deficiency is best assessed by obtaining serum 25-hydroxyvitamin D levels, which reflect vitamin D stores. In the setting of renal insufficiency or suspected vitamin D resistance, serum $1,25(OH)_2D$ levels are informative.

Treatment: Hypocalcemia

The approach to treatment depends on the severity of the hypocalcemia, the rapidity with which it develops, and the accompanying complications (e.g., seizures, laryngospasm). Acute, symptomatic hypocalcemia is initially managed with calcium gluconate, 10 mL 10% wt/vol (90 mg or 2.2 mmol) intravenously, diluted in 50 mL of 5% dextrose or 0.9% sodium chloride, given intravenously over 5 min. Continuing hypocalcemia often requires a constant intravenous infusion (typically 10 ampuls of calcium gluconate or 900 mg of calcium in 1 L of 5% dextrose or 0.9% sodium chloride administered over 24 h). Accompanying hypomagnesemia, if present, should be treated with appropriate magnesium supplementation.

Chronic hypocalcemia due to hypoparathyroidism is treated with calcium supplements (1000–1500 mg/d elemental calcium in divided doses) and either vitamin D2 or D3 (25,000–100,000 U daily) or calcitriol [$1,25(OH)_2D$, 0.25–2 μg/d]. Other vitamin D metabolites (dihydrotachysterol, alfacalcidiol) are now used less frequently. Vitamin D deficiency, however, is best treated using vitamin D supplementation, with the dose depending on the severity of the deficit and the underlying cause. Thus, nutritional vitamin D deficiency generally responds to relatively low doses of vitamin D (50,000 U, 2–3 times per week for several months), while vitamin D deficiency due to malabsorption may require much higher doses (100,000 U/d or more). The treatment goal is to bring serum calcium into the low normal range and to avoid hypercalciuria, which may lead to nephrolithiasis.

FURTHER READINGS

Bilezikian JP: Primary hyperparathyroidism and hypoparathyroidism, in *Conn's Current Therapy*, RE Rakel, ET Bope (eds), Philadelphia, Elsevier, 2009

Bilezikian JP, Khan AA et al: Guidelines for the management of asymptomatic primary hyperparathyroidism: Summary statement from the third international workship. J Clin Endocrinol Metab 94:335, 2009 [PMID: 19193908]

Egbuna OI, Brown EM: Hypercalcaemic and hypocalcaemic conditions due to calcium-sensing receptor mutations. Best Pract Clin Rheumatol 22:129, 2008 [PMID: 18328986]

Shoback D: Clinical practice. Hypoparathyroidism. N Engl J Med 359:391, 2008 [PMID: 18650515]

Stewart AF: Hypercalcemia associated with cancer. N Engl J Med 352:373, 2005 [PMID: 15673803]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NORMAL ACID-BASE HOMEOSTASIS

Systemic arterial pH is maintained between 7.35 and 7.45 by extracellular and intracellular chemical buffering together with respiratory and renal regulatory mechanisms. The control of arterial CO_2 tension (Pa_{CO2}) by the central nervous system (CNS) and respiratory systems and the control of the plasma bicarbonate by the kidneys stabilize the arterial pH by excretion or retention of acid or alkali. The metabolic and respiratory components that regulate systemic pH are described by the Henderson-Hasselbalch equation:

$$pH = 6.1 + \log \frac{HCO_3^-}{Pa_{co_2} \times 0.0301}$$

Under most circumstances, CO_2 production and excretion are matched, and the usual steady-state Pa_{CO2} is maintained at 40 mmHg. Underexcretion of CO_2 produces hypercapnia, and overexcretion causes hypocapnia. Nevertheless, production and excretion are again matched at a new steady-state Pa_{CO2}. Therefore, the Pa_{CO2} is regulated primarily by neural respiratory factors and is not subject to regulation by the rate of CO_2 production. Hypercapnia is usually the result of hypoventilation rather than of increased CO_2 production. Increases or decreases in Pa_{CO2} represent derangements of neural respiratory control or are due to compensatory changes in response to a primary alteration in the plasma [HCO_3^-].

The kidneys regulate plasma [HCO_3^-] through three main processes: (1) "reabsorption" of filtered HCO_3^-, (2) formation of titratable acid, and (3) excretion of NH_4^+ in the urine. The kidney filters ~4000 mmol of HCO_3^- per day. To reabsorb the filtered load of HCO_3^-, the renal tubules must therefore secrete 4000 mmol of hydrogen ions. Between 80 and 90% of HCO_3^- is reabsorbed in the proximal tubule. The distal nephron reabsorbs the remainder and secretes H^+ to defend systemic pH. While this quantity of protons, 40–60 mmol/d, is small, it must be secreted to prevent chronic positive H^+ balance and metabolic acidosis. This quantity of secreted protons is represented in the urine as titratable acid and NH_4^+. Metabolic acidosis in the face of normal renal function increases NH_4^+ production and excretion. NH_4^+ production and excretion are impaired in chronic renal failure, hyperkalemia, and renal tubular acidosis.

DIAGNOSIS OF GENERAL TYPES OF DISTURBANCES

The most common clinical disturbances are simple acid-base disorders; i.e., metabolic acidosis or alkalosis or respiratory acidosis or alkalosis. Because compensation is not complete, the pH is abnormal in simple disturbances. More complicated clinical situations can give rise to mixed acid-base disturbances.

Simple Acid-Base Disorders

Primary respiratory disturbances (primary changes in Pa_{CO2}) invoke compensatory metabolic responses (secondary changes in [HCO_3^-]), and primary metabolic disturbances elicit predictable compensatory respiratory responses (secondary changes in Pa_{CO2}). Physiologic compensation can be predicted from the relationships displayed in **Table 47-1**. Metabolic acidosis due to an increase in endogenous acids (e.g., ketoacidosis) lowers extracellular fluid [HCO_3^-] and decreases extracellular pH. This stimulates the medullary chemoreceptors to increase ventilation and to return the ratio of [HCO_3^-] to Pa_{CO2}, and thus pH, toward, but not to, normal. The degree of respiratory compensation expected in a simple form of metabolic acidosis can be predicted from the relationship: Pa_{CO2} = (1.5 x [HCO_3^-]) + 8 ± 2. Thus, a patient with metabolic acidosis and [HCO_3^-] of 12 mmol/L would be expected to have a Pa_{CO2} between 24 and 28 mmHg. Values for Pa_{CO2} <24 or >28 mmHg define a mixed disturbance (metabolic acidosis and respiratory alkalosis or metabolic alkalosis and respiratory

acidosis, respectively). Another way to judge the appropriateness of the response in $[HCO_3^-]$ or Pa_{CO2} is to use an acid-base nomogram **(Fig. 47-1)**. While the shaded areas of the nomogram show the 95% confidence limits for normal compensation in simple disturbances, finding acid-base values within the shaded area does not necessarily rule out a mixed disturbance. Imposition of one disorder over another may result in values lying within the area of a third. Thus, the nomogram, while convenient, is not a substitute for the equations in Table 47-1.

Table 47-1 Prediction of Compensatory Responses on Simple Acid-Base Disturbances and Pattern of Changes

		Range of Values		
Disorder	**Prediction of Compensation**	**pH**	**HCO_3^-**	**Pa_{co2}**
Metabolic acidosis	$Pa_{co2} = (1.5X\ HCO_3^-) + 8 \pm 2$	Low	Low	Low
	or			
	Pa_{co2} will ↓ 1.25 mmHg per mmol/L ↓ in $[HCO_3^-]$			
	or			
	$Pa_{co2} = [HCO_3^-] + 15$			
Metabolic alkalosis	Pa_{co2} will ↑ 0.75 mmHg per mmol/L ↑ in $[HCO_3–]$	High	High	High
	or			
	Pa_{co2} will ↑ 6 mmHg per 10 mmol/L ↑ in $[HCO_3–]$			
	or			
	$Pa_{co2} = [HCO_3^-] + 15$			
Respiratory alkalosis		High	Low	Low
Acute	$[HCO_3–]$ will ↓ 0.2 mmol/L per mmHg ↓ in Pa_{co2}			
Chronic	$[HCO_3–]$ will ↓ 0.4 mmol/L per mmHg ↓ in Pa_{co2}			
Respiratory acidosis		Low	High	High
Acute	$[HCO_3–]$ will ↑ 0.1 mmol/L per mmHg ↑ in Pa_{co2}			
Chronic	$[HCO_3–]$ will ↑ 0.4 mmol/L per mmHg ↑ in Pa_{co2}			

Figure 47-1

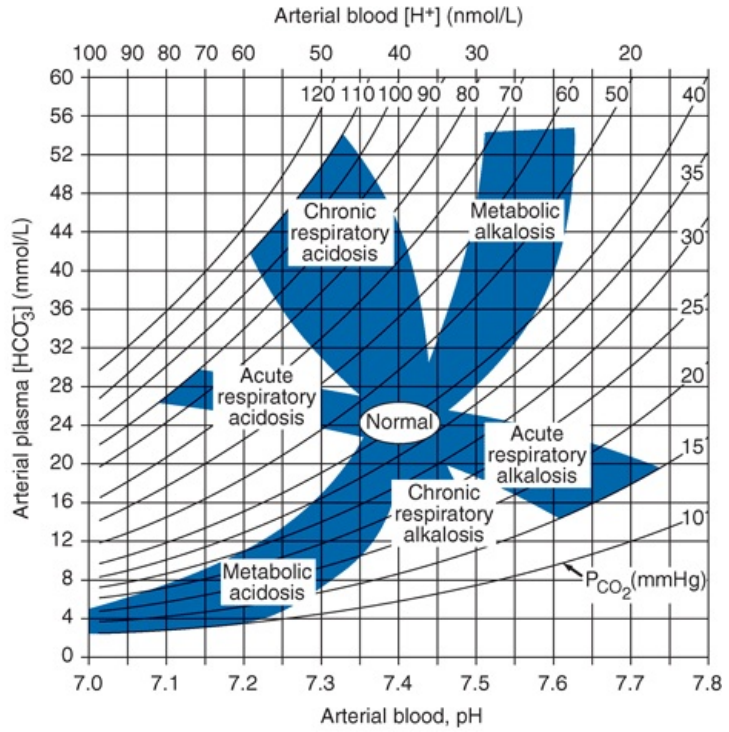

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acid-base nomogram. Shown are the 90% confidence limits (range of values) of the normal respiratory and metabolic compensations for primary acid-base disturbances. *(From DuBose, used with permission.)*

Mixed Acid-Base Disorders

Mixed acid-base disorders—defined as independently coexisting disorders, not merely compensatory responses—are often seen in patients in critical care units and can lead to dangerous extremes of pH **(Table 47-2)**. A patient with diabetic ketoacidosis (metabolic acidosis) may develop an independent respiratory problem (e.g., pneumonia) leading to respiratory acidosis or alkalosis. Patients with underlying pulmonary disease (e.g., COPD) may not respond to metabolic acidosis with an appropriate ventilatory response because of insufficient respiratory reserve. Such imposition of respiratory acidosis on metabolic acidosis can lead to severe acidemia. When metabolic acidosis and metabolic alkalosis coexist in the same patient, the pH may be normal or near normal. When the pH is normal, an elevated anion gap (AG; see below) reliably denotes the presence of an AG metabolic acidosis. A discrepancy in the ΔAG (prevailing minus normal AG) and the ΔHCO_3^- (normal minus prevailing HCO_3^-) indicates the presence of a mixed high-gap acidosis—metabolic alkalosis (see example below). A diabetic patient with ketoacidosis may have renal dysfunction resulting in simultaneous metabolic acidosis. Patients who have ingested an overdose of drug combinations such as sedatives and salicylates may have mixed disturbances as a result of the acid-base response to the individual drugs (metabolic acidosis mixed with respiratory acidosis or respiratory alkalosis, respectively). Triple acid-base disturbances are more complex. For example, patients with metabolic acidosis due to alcoholic ketoacidosis may develop metabolic alkalosis due to vomiting and superimposed respiratory alkalosis due to the hyperventilation of hepatic dysfunction or alcohol withdrawal.

Table 47-2 Examples of Mixed Acid-Base Disorders

Mixed Metabolic and Respiratory Metabolic acidosis—respiratory alkalosis
Key: High- or normal-AG metabolic acidosis; prevailing Pa_{CO2} *below* predicted value (Table 47-1)
Example: Na^+, 140; K^+, 4.0; Cl^-, 106; $HCO3^-$, 14; AG, 20; Pa_{CO2}, 24; pH, 7.39 (lactic acidosis, sepsis in ICU)
Metabolic acidosis—respiratory acidosis
Key: High$^-$ or normal-AG metabolic acidosis; prevailing Pa_{CO2} *above* predicted value (Table 47-1)

Example: Na^+, 140; K^+, 4.0; Cl^-, 102; $HCO3^-$, 18; AG, 20; Pa_{CO2}, 38; pH, 7.30 (severe pneumonia, pulmonary edema)
Metabolic alkalosis—respiratory alkalosis
Key: Pa_{CO2} does not increase as predicted; pH higher than expected
Example: Na^+, 140; K^+, 4.0; Cl^-, 91; $HCO3^-$, 33; AG, 16; Pa_{CO2}, 38; pH, 7.55 (liver disease and diuretics)
Metabolic alkalosis—respiratory acidosis
Key: Pa_{CO2} higher than predicted; pH normal
Example: Na^+, 140; K^+, 3.5; Cl^-, 88; $HCO3^-$, 42; AG, 10; Pa_{CO2}, 67; pH, 7.42 (COPD on diuretics)
Mixed Metabolic Disorders
Metabolic acidosis—metabolic alkalosis
Key: Only detectable with high-AG acidosis; ΔAG >> $\Delta HCO3^-$
Example: Na^+, 140; K^+, 3.0; Cl^-, 95; $HCO3^-$, 25; AG, 20; Pa_{CO2}, 40; pH, 7.42 (uremia with vomiting)
Metabolic acidosis—metabolic acidosis
Key: Mixed high-AG—normal-AG acidosis; $\Delta HCO3^-$ accounted for by
combined change in ΔAG and ΔCl^-
Example: Na^+, 135; K^+, 3.0; Cl^-, 110; $HCO3^-$, 10; AG, 15; Pa_{CO2}, 25; pH, 7.20 (diarrhea and lactic acidosis, toluene toxicity, treatment of diabetic ketoacidosis)

Abbreviations: AG, anion gap; COPD, chronic obstructive pulmonary disease; ICU, intensive care unit.

Approach to the Patient: Acid-Base Disorders

A stepwise approach to the diagnosis of acid-base disorders follows **(Table 47-3)**. Care should be taken when measuring blood gases to obtain the arterial blood sample without using excessive heparin. Blood for electrolytes and arterial blood gases should be drawn simultaneously prior to therapy, because an increase in [HCO_3^-] occurs with metabolic alkalosis and respiratory acidosis. Conversely, a decrease in [HCO_3^-] occurs in metabolic acidosis and respiratory alkalosis. In the determination of arterial blood gases by the clinical laboratory, both pH and Pa_{CO2} are measured, and the [HCO_3^-] is calculated from the Henderson-Hasselbalch equation. This calculated value should be compared with the measured [HCO_3^-] (total CO_2) on the electrolyte panel. These two values should agree within 2 mmol/L. If they do not, the values may not have been drawn simultaneously, a laboratory error may be present, or an error could have been made in calculating the [HCO_3^-]. After verifying the blood acid-base values, the precise acid-base disorder can then be identified.

Table 47-3 Steps in Acid-Base Diagnosis

1. Obtain arterial blood gas (ABG) and electrolytes simultaneously.
2. Compare [HCO_3^-] on ABG and electrolytes to verify accuracy.
3. Calculate anion gap (AG).
4. Know four causes of high-AG acidosis (ketoacidosis, lactic acid acidosis, renal failure, and toxins).
5. Know two causes of hyperchloremic or nongap acidosis (bicarbonate loss from GI tract, renal tubular acidosis).
6. Estimate compensatory response (Table 47-1).

7. Compare ΔAG and ΔHCO_3–.
8. Compare change in [Cl– with change in [Na^+].

Calculate the Anion Gap

All evaluations of acid-base disorders should include a simple calculation of the AG; it represents those unmeasured anions in plasma (normally 10 to 12 mmol/L) and is calculated as follows: AG = Na^+ – (Cl^- + HCO_3^-). The unmeasured anions include anionic proteins, (e.g., albumin), phosphate, sulfate, and organic anions. When acid anions, such as acetoacetate and lactate, accumulate in extracellular fluid, the AG increases, causing a high-AG acidosis. An increase in the AG is most often due to an increase in unmeasured anions and, less commonly, is due to a decrease in unmeasured cations (calcium, magnesium, potassium). In addition, the AG may increase with an increase in anionic albumin, because of either increased albumin concentration or alkalosis, which alters albumin charge. A decrease in the AG can be due to (1) an increase in unmeasured cations; (2) the addition to the blood of abnormal cations, such as lithium (lithium intoxication) or cationic immunoglobulins (plasma cell dyscrasias); (3) a reduction in the major plasma anion albumin concentration (nephrotic syndrome); (4) a decrease in the effective anionic charge on albumin by acidosis; or (5) hyperviscosity and severe hyperlipidemia, which can lead to an underestimation of sodium and chloride concentrations. A fall in serum albumin by 1 g/dL from the normal value (4.5 g/dL) decreases the AG by 2.5 meq/L. Know the common causes of a high-AG acidosis (Table 47-3).

In the face of a normal serum albumin, a high AG is usually due to non-chloride containing acids that contain inorganic (phosphate, sulfate), organic (ketoacids, lactate, uremic organic anions), exogenous (salicylate or ingested toxins with organic acid production), or unidentified anions. The high AG is significant even if an additional acid-base disorder is superimposed to modify the [HCO_3^-] independently. Simultaneous metabolic acidosis of the high-AG variety plus either chronic respiratory acidosis or metabolic alkalosis represents such a situation in which [HCO_3^-] may be normal or even high (Table 47-3). Compare the change in [HCO_3^-] (ΔHCO_3^-) and the change in the AG (ΔAG).

Similarly, normal values for [HCO_3^-], Pa_{CO2}, and pH do not ensure the absence of an acid-base disturbance. For instance, an alcoholic who has been vomiting may develop a metabolic alkalosis with a pH of 7.55, Pa_{CO2} of 47 mmHg, [HCO_3^-] of 40 mmol/L, [Na^+] of 135, [Cl^-] of 80, and [K^+] of 2.8. If such a patient were then to develop a superimposed alcoholic ketoacidosis with a ß-hydroxybutyrate concentration of 15 m*M*, arterial pH would fall to 7.40, [HCO_3^-] to 25 mmol/L, and the Pa_{CO2} to 40 mmHg. Although these blood gases are normal, the AG is elevated at 30 mmol/L, indicating a mixed metabolic alkalosis and metabolic acidosis. A mixture of high-gap acidosis and metabolic alkalosis is recognized easily by comparing the differences (Δ values) in the normal to prevailing patient values. In this example, the ΔHCO_3^- is 0 (25 – 25 mmol/L) but the ΔAG is 20 (30 – 10 mmol/L). Therefore, 20 mmol/L is unaccounted for in the Δ/Δ value (ΔAG to ΔHCO_3^-).

METABOLIC ACIDOSIS

Metabolic acidosis can occur because of an increase in endogenous acid production (such as lactate and ketoacids), loss of bicarbonate (as in diarrhea), or accumulation of endogenous acids (as in renal failure). Metabolic acidosis has profound effects on the respiratory, cardiac, and nervous systems. The fall in blood pH is accompanied by a characteristic increase in ventilation, especially the tidal volume (Kussmaul respiration). Intrinsic cardiac contractility may be depressed, but inotropic function can be normal because of catecholamine release. Both peripheral arterial vasodilation and central venoconstriction can be present; the decrease in central and pulmonary vascular compliance predisposes to pulmonary edema with even minimal volume overload. CNS function is depressed, with headache, lethargy, stupor, and, in some cases, even coma. Glucose intolerance may also occur.

There are two major categories of clinical metabolic acidosis: high-AG and normal-AG, or hyperchloremic acidosis (Table 47-3 and **Table 47-4**).

Table 47-4 Causes of High Anion–Gap Metabolic Acidosis

Lactic acidosis	Toxins
Ketoacidosis	Ethylene glycol
Diabetic	Methanol
Alcoholic	Salicylates

Starvation	Propylene glycol
	Pyroglutamic acid
	Renal failure (acute and chronic)

Treatment: Metabolic Acidosis

Treatment of metabolic acidosis with alkali should be reserved for severe acidemia except when the patient has no "potential HCO_3^-" in plasma. Potential [HCO_3^-] can be estimated from the increment (Δ) in the AG (ΔAG = patient's AG – 10). It must be determined if the acid anion in plasma is metabolizable (i.e., β-hydroxybutyrate, acetoacetate, and lactate) or nonmetabolizable (anions that accumulate in chronic renal failure and after toxin ingestion). The latter requires return of renal function to replenish the [HCO_3^-] deficit, a slow and often unpredictable process. Consequently, patients with a normal AG acidosis (hyperchloremic acidosis), a slightly elevated AG (mixed hyperchloremic and AG acidosis), or an AG attributable to a nonmetabolizable anion in the face of renal failure should receive alkali therapy, either PO (NaHCO3 or Shohl's solution) or IV ($NaHCO_3$), in an amount necessary to slowly increase the plasma [HCO_3^-] into the 20–22 mmol/L range.

Controversy exists, however, in regard to the use of alkali in patients with a pure AG acidosis owing to accumulation of a metabolizable organic acid anion (ketoacidosis or lactic acidosis). In general, severe acidosis (pH < 7.10) warrants the IV administration of 50–100 meq of NaHCO3 , over 30–45 min, during the initial 1–2 h of therapy. Provision of such modest quantities of alkali in this situation seems to provide an added measure of safety, but it is essential to monitor plasma electrolytes during the course of therapy, because the [K^+] may decline as pH rises. The goal is to increase the [HCO_3^-] to 10 meq/L and the pH to 7.20, not to increase these values to normal.

High–Anion Gap Acidoses

APPROACH TO THE PATIENT: HIGH—ANION GAP ACIDOSES

There are four principal causes of a high-AG acidosis: (1) lactic acidosis, (2) ketoacidosis, (3) ingested toxins, and (4) acute and chronic renal failure (Table 47-4). Initial screening to differentiate the high-AG acidoses should include (1) a probe of the history for evidence of drug and toxin ingestion and measurement of arterial blood gas to detect coexistent respiratory alkalosis (salicylates); (2) determination of whether diabetes mellitus is present (diabetic ketoacidosis); (3) a search for evidence of alcoholism or increased levels of β-hydroxybutyrate (alcoholic ketoacidosis); (4) observation for clinical signs of uremia and determination of the blood urea nitrogen (BUN) and creatinine (uremic acidosis); (5) inspection of the urine for oxalate crystals (ethylene glycol); and (6) recognition of the numerous clinical settings in which lactate levels may be increased (hypotension, shock, cardiac failure, leukemia, cancer, and drug or toxin ingestion).

LACTIC ACIDOSIS

An increase in plasma l-lactate may be secondary to poor tissue perfusion (type A)—circulatory insufficiency (shock, cardiac failure), severe anemia, mitochondrial enzyme defects, and inhibitors (carbon monoxide, cyanide)—or to aerobic disorders (type B)—malignancies, nucleoside analogue reverse transcriptase inhibitors in HIV, diabetes mellitus, renal or hepatic failure, thiamine deficiency, severe infections (cholera, malaria), seizures, or drugs/toxins (biguanides, ethanol, methanol, propylene glycol, isoniazid, and fructose). Propylene glycol may be used as a vehicle for IV medications including lorazepam, and toxicity has been reported in several settings. Unrecognized bowel ischemia or infarction in a patient with severe atherosclerosis or cardiac decompensation receiving vasopressors is a common cause of lactic acidosis. Pyroglutamic acidemia has been reported in critically ill patients receiving acetaminophen, which is associated with depletion of glutathione. D-Lactic acid acidosis, which may be associated with jejunoileal bypass, short bowel syndrome, or intestinal obstruction, is due to formation of d-lactate by gut bacteria.

Approach to the Patient: Lactic Acid Acidosis

The underlying condition that disrupts lactate metabolism must first be corrected; tissue perfusion must be restored when inadequate. Vasoconstrictors should be avoided, if possible, because they may worsen tissue perfusion. Alkali therapy is generally advocated for acute, severe acidemia (pH < 7.15) to improve cardiac function and lactate use. However, NaHCO3 therapy may paradoxically depress cardiac performance and exacerbate acidosis by enhancing lactate production (HCO_3^- stimulates phosphofructokinase). While the use of alkali in moderate lactic acidosis is controversial, it is generally agreed that attempts to return the pH or [HCO_3^-] to normal by administration of exogenous $NaHCO_3$ are deleterious. A reasonable approach is to infuse sufficient $NaHCO_3$ to raise the arterial pH to no more than 7.2 over 30–40 min.

NaHCO3 therapy can cause fluid overload and hypertension because the amount required can be

massive when accumulation of lactic acid is relentless. Fluid administration is poorly tolerated because of central venoconstriction, especially in the oliguric patient. When the underlying cause of the lactic acidosis can be remedied, blood lactate will be converted to HCO_3^- and may result in an overshoot alkalosis.

KETOACIDOSIS

Diabetic Ketoacidosis (DKA)

This condition is caused by increased fatty acid metabolism and the accumulation of ketoacids (acetoacetate and β-hydroxybutyrate). DKA usually occurs in insulin-dependent diabetes mellitus in association with cessation of insulin or an intercurrent illness such as an infection, gastroenteritis, pancreatitis, or myocardial infarction, which increases insulin requirements temporarily and acutely. The accumulation of ketoacids accounts for the increment in the AG and is accompanied most often by hyperglycemia [glucose > 17 mmol/L (300 mg/dL)]. The relationship between the ΔAG and ΔHCO_3^- is typically ~1:1 in DKA. It should be noted that, because insulin prevents production of ketones, bicarbonate therapy is rarely needed except with extreme acidemia (pH < 7.1), and then in only limited amounts. Patients with DKA are typically volume depleted and require fluid resuscitation with isotonic saline. Volume overexpansion with IV-fluid administration is not uncommon, however, and contributes to the development of a hyperchloremic acidosis during treatment of DKA. The mainstay for treatment of this condition is IV regular insulin and is described in Chap. 344 in more detail.

Alcoholic Ketoacidosis (AKA)

Chronic alcoholics can develop ketoacidosis when alcohol consumption is abruptly curtailed and nutrition is poor. AKA is usually associated with binge drinking, vomiting, abdominal pain, starvation, and volume depletion. The glucose concentration is variable, and acidosis may be severe because of elevated ketones, predominantly β-hydroxybutyrate. Hypoperfusion may enhance lactic acid production, chronic respiratory alkalosis may accompany liver disease, and metabolic alkalosis can result from vomiting (refer to the relationship between ΔAG and ΔHCO_3^-). Thus, mixed acid-base disorders are common in AKA. As the circulation is restored by administration of isotonic saline, the preferential accumulation of β-hydroxybutyrate is then shifted to acetoacetate. This explains the common clinical observation of an increasingly positive nitroprusside reaction as the patient improves. The nitroprusside ketone reaction (Acetest) can detect acetoacetic acid but not β-hydroxybutyrate, so that the degree of ketosis and ketonuria can not only change with therapy, but can be underestimated initially. Patients with AKA usually present with relatively normal renal function, as opposed to DKA, where renal function is often compromised because of volume depletion (osmotic diuresis) or diabetic nephropathy. The AKA patient with normal renal function may excrete relatively large quantities of ketoacids in the urine, therefore, and may have a relatively normal AG and a discrepancy in the $\Delta AG/\Delta HCO_3^-$ relationship.

Treatment: Alcoholic Ketoacidosis

Extracellular fluid deficits almost always accompany AKA and should be repleted by IV administration of saline and glucose (5% dextrose in 0.9% NaCl). Hypophosphatemia, hypokalemia, and hypomagnesemia may coexist and should be corrected. Hypophosphatemia usually emerges 12–24 h after admission, may be exacerbated by glucose infusion, and, if severe, may induce rhabdomyolysis. Upper gastrointestinal hemorrhage, pancreatitis, and pneumonia may accompany this disorder.

DRUG- AND TOXIN-INDUCED ACIDOSIS

Salicylates

(See also **Chap. e49**) Salicylateintoxication in adults usually causes respiratory alkalosis or a mixture of high-AG metabolic acidosis and respiratory alkalosis. Only a portion of the AG is due to salicylates. Lactic acid production is also often increased.

Treatment: Salicylate-Induced Acidosis

Vigorous gastric lavage with isotonic saline (not NaHCO3) should be initiated immediately, followed by administration of activated charcoal per NG tube. In the acidotic patient, to facilitate removal of salicylate, intravenous $NaHCO_3$ is administered in amounts adequate to alkalinize the urine and to maintain urine output (urine pH > 7.5). While this form of therapy is straightforward in acidotic patients, a coexisting respiratory alkalosis may make this approach hazardous. Alkalemic patients should not receive $NaHCO_3$. Acetazolamide may be administered in the face of alkalemia, when an alkaline diuresis cannot be achieved, or to ameliorate volume overload associated with $NaHCO_3$ administration, but this drug can cause systemic metabolic acidosis if HCO_3^- is not replaced. Hypokalemia should be anticipated with an alkaline diuresis and should be treated promptly and aggressively. Glucose-containing fluids should be administered because of the danger of hypoglycemia. Excessive insensible fluid losses may cause severe volume depletion and hypernatremia. If renal failure prevents rapid clearance of salicylate, hemodialysis can be performed

against a bicarbonate dialysate.

Alcohols

Under most physiologic conditions, sodium, urea, and glucose generate the osmotic pressure of blood. Plasma osmolality is calculated according to the following expression: $P_{osm} = 2Na^+ + Glu + BUN$ (all in mmol/L), or, using conventional laboratory values in which glucose and BUN are expressed in milligrams per deciliter: $P_{osm} = 2Na^+ + Glu/18 + BUN/2.8$. The calculated and determined osmolality should agree within 10–15 mmol/kg H_2O. When the measured osmolality exceeds the calculated osmolality by >15–20 mmol/kg H_2O, one of two circumstances prevails. Either the serum sodium is spuriously low, as with hyperlipidemia or hyperproteinemia (pseudohyponatremia), or osmolytes other than sodium salts, glucose, or urea have accumulated in plasma. Examples of such osmolytes include mannitol, radiocontrast media, ethanol, isopropyl alcohol, ethylene glycol, propylene glycol, methanol, and acetone. In this situation, the difference between the calculated osmolality and the measured osmolality (*osmolar gap*) is proportional to the concentration of the unmeasured solute. With an appropriate clinical history and index of suspicion, identification of an osmolar gap is helpful in identifying the presence of poison-associated AG acidosis. Three alcohols may cause fatal intoxications: ethylene glycol, methanol, and isopropyl alcohol. All cause an elevated osmolal gap, but only the first two cause a high-AG acidosis.

Ethylene Glycol

(See also **Chap. e49**) Ingestion of ethylene glycol (commonly used in antifreeze) leads to a metabolic acidosis and severe damage to the CNS, heart, lungs, and kidneys. The increased AG and osmolar gap are attributable to ethylene glycol and its metabolites, oxalic acid, glycolic acid, and other organic acids. Lactic acid production increases secondary to inhibition of the tricarboxylic acid cycle and altered intracellular redox state. Diagnosis is facilitated by recognizing oxalate crystals in the urine, the presence of an osmolar gap in serum, and a high-AG acidosis. Treatment should not be delayed while awaiting measurement of ethylene glycol levels in this setting.

Treatment: Ethylene Glycol—Induced Acidosis

This includes the prompt institution of a saline or osmotic diuresis, thiamine and pyridoxine supplements, fomepizole or ethanol, and hemodialysis. The IV administration of the alcohol dehydrogenase inhibitor fomepizole (4-methylpyrazole; 15 mg/kg as a loading dose) or ethanol IV to achieve a level of 22 mmol/L (100 mg/dL) serves to lessen toxicity because they compete with ethylene glycol for metabolism by alcohol dehydrogenase. Fomepizole, although expensive, is the agent of choice and offers the advantages of a predictable decline in ethylene glycol levels without excessive obtundation during ethyl alcohol infusion. Hemodialysis is indicated when the arterial pH is <7.3, or the osmolar gap exceeds 20 mOsm/kg.

Methanol

(See also **Chap. e49**) The ingestion of methanol (wood alcohol) causes metabolic acidosis, and its metabolites formaldehyde and formic acid cause severe optic nerve and CNS damage. Lactic acid, ketoacids, and other unidentified organic acids may contribute to the acidosis. Due to its low molecular mass (32 Da), an osmolar gap is usually present.

Treatment: Methanol-Induced Acidosis

This is similar to that for ethylene glycol intoxication, including general supportive measures, fomepizole, and hemodialysis (as above).

Isopropyl Alcohol

Ingested isopropanol is absorbed rapidly and may be fatal when as little as 150 mL of rubbing alcohol, solvent, or de-icer is consumed. A plasma level >400 mg/dL is life-threatening. Isopropyl alcohol differs from ethylene glycol and methanol in that the parent compound, not the metabolites, causes toxicity, and an AG acidosis is not present because acetone is rapidly excreted.

Treatment: Isopropyl Alcohol Toxicity

Isopropanol alcohol toxicity is treated by watchful waiting and supportive therapy; IV fluids, pressors, ventilatory support if needed, and occasionally hemodialysis for prolonged coma or levels >400 mg/dL.

RENAL FAILURE

(See also Chap. 280) The hyperchloremic acidosis of moderate renal insufficiency is eventually converted to the high-AG acidosis of advanced renal failure. Poor filtration and reabsorption of organic anions contribute to the pathogenesis. As renal disease progresses, the number of functioning nephrons eventually becomes insufficient to keep pace with net acid production. Uremic acidosis is characterized, therefore, by a reduced rate of NH_4^+ production and excretion. The acid retained in chronic renal disease is buffered by alkaline salts from bone. Despite significant retention of acid (up to 20 mmol/d), the serum [HCO_3^-] does not decrease further, indicating participation of buffers outside

the extracellular compartment. Chronic metabolic acidosis results in significant loss of bone mass due to reduction in bone calcium carbonate. Chronic acidosis also increases urinary calcium excretion, proportional to cumulative acid retention.

Treatment: Renal Failure

Because of the association of renal failure acidosis with muscle catabolism and bone disease, both uremic acidosis and the hyperchloremic acidosis of renal failure require oral alkali replacement to maintain the [HCO_3^-] between 20 and 24 mmol/L. This can be accomplished with relatively modest amounts of alkali (1.0–1.5 mmol/kg body weight per day). Sodium citrate (Shohl's solution) or NaHCO3 tablets (650-mg tablets contain 7.8 meq) are equally effective alkalinizing salts. Citrate enhances the absorption of aluminum from the gastrointestinal tract and should never be given together with aluminum-containing antacids because of the risk of aluminum intoxication. When hyperkalemia is present, furosemide (60–80 mg/d) should be added.

Non–Anion Gap Metabolic Acidoses

Alkali can be lost from the gastrointestinal tract in diarrhea or from the kidneys (renal tubular acidosis, RTA). In these disorders **(Table 47-5)**, reciprocal changes in [Cl^-] and [HCO_3^-] result in a normal AG. In pure non–AG acidosis, therefore, the increase in [Cl^-] above the normal value approximates the decrease in [HCO_3^-]. The absence of such a relationship suggests a mixed disturbance.

Table 47-5 Causes of Non–Anion Gap Acidosis

I. Gastrointestinal bicarbonate loss
- A. Diarrhea
- B. External pancreatic or small-bowel drainage
- C. Ureterosigmoidostomy, jejunal loop, ileal loop
- D. Drugs
  - 1. Calcium chloride (acidifying agent)
  - 2. Magnesium sulfate (diarrhea)
  - 3. Cholestyramine (bile acid diarrhea)

II. Renal acidosis
- A. Hypokalemia
  - 1. Proximal RTA (type 2)
  - Drug-induced: acetazolamide, topiramate
  - 2. Distal (classic) RTA (type 1)
  - Drug induced: amphotericin B, ifosfamide
- B. Hyperkalemia
  - 1. Generalized distal nephron dysfunction (type 4 RTA)
    - a. Mineralocorticoid deficiency
    - b. Mineralocorticoid resistance (autosomal dominant PHA I)
    - c. Voltage defect (autosomal dominant PHA I and PHA II)
    - d. Tubulointerstitial disease

III. Drug-induced hyperkalemia (with renal insufficiency)
- A. Potassium-sparing diuretics (amiloride, triamterene, spironolactone)
- B. Trimethoprim
- C. Pentamidine
- D. ACE-Is and ARBs
- E. Nonsteroidal anti-inflammatory drugs
- F. Cyclosporine and tacrolimus

IV. Other
- A. Acid loads (ammonium chloride, hyperalimentation)
- B. Loss of potential bicarbonate: ketosis with ketone excretion
- C. Expansion acidosis (rapid saline administration)
- D. Hippurate

E. Cation exchange resins

Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; PHA, pseudohypoaldosteronism; RTA, renal tubular acidosis.

TREATMENT: NON–ANION GAP METABOLIC ACIDOSES

In diarrhea, stools contain a higher [HCO_3^-] and decomposed HCO_3^- than plasma so that metabolic acidosis develops along with volume depletion. Instead of an acid urine pH (as anticipated with systemic acidosis), urine pH is usually around 6 because metabolic acidosis and hypokalemia increase renal synthesis and excretion of NH_4^+, thus providing a urinary buffer that increases urine pH. Metabolic acidosis due to gastrointestinal losses with a high urine pH can be differentiated from RTA because urinary NH_4^+ excretion is typically low in RTA and high with diarrhea. Urinary NH_4^+ levels can be estimated by calculating the urine anion gap (UAG): UAG = $[Na^+ + K^+]_u - [Cl^-]_u$. When $[Cl^-]_u > [Na^+ + K^+]_u$, the UAG is negative by definition. This indicates that the urine ammonium level is appropriately increased, suggesting an extrarenal cause of the acidosis. Conversely, when the UAG is positive, the urine ammonium level is low, suggesting a renal cause of the acidosis.

Loss of functioning renal parenchyma by progressive renal disease leads to hyperchloremic acidosis when the glomerular filtration rate (GFR) is between 20 and 50 mL/min and to uremic acidosis with a high AG when the GFR falls to <20 mL/min. In advanced renal failure, ammoniagenesis is reduced in proportion to the loss of functional renal mass, and ammonium accumulation and trapping in the outer medullary collecting tubule may also be impaired. Because of adaptive increases in K^+ secretion by the collecting duct and colon, the acidosis of chronic renal insufficiency is typically normokalemic.

Proximal RTA (type 2 RTA) (Chap. 284) is most often due to generalized proximal tubular dysfunction manifested by glycosuria, generalized aminoaciduria, and phosphaturia (Fanconi syndrome). With a low plasma [HCO_3^-], the urine pH is acid (pH < 5.5). The fractional excretion of [HCO_3^-] may exceed 10–15% when the serum HCO_3^- > 20 mmol/L. Because HCO_3^- is not reabsorbed normally in the proximal tubule, therapy with NaHCO3 will enhance renal potassium wasting and hypokalemia.

The typical findings in acquired or inherited forms of classic distal RTA (type 1 RTA) include hypokalemia, non-AG metabolic acidosis, low urinary NH_4^+ excretion (positive UAG, low urine [NH_4^+]), and inappropriately high urine pH (pH > 5.5). Most patients have hypocitraturia and hypercalciuria, so nephrolithiasis, nephrocalcinosis, and bone disease are common. In generalized distal nephron dysfunction (type 4 RTA), hyperkalemia is disproportionate to the reduction in GFR because of coexisting dysfunction of potassium and acid secretion. Urinary ammonium excretion is invariably depressed, and renal function may be compromised, for example, due to diabetic nephropathy, obstructive uropathy, or chronic tubulointerstitial disease.

Hyporeninemic hypoaldosteronism typically causes non-AG metabolic acidosis, most commonly in older adults with diabetes mellitus or tubulointerstitial disease and renal insufficiency. Patients usually have mild to moderate CKD (GFR, 20–50 mL/min) and acidosis, with elevation in serum [K+] (5.2–6.0 mmol/L), concurrent hypertension, and congestive heart failure. Both the metabolic acidosis and the hyperkalemia are out of proportion to impairment in GFR. Nonsteroidal anti-inflammatory drugs, trimethoprim, pentamidine, and angiotensin-converting enzyme (ACE) inhibitors can also cause non-AG metabolic acidosis in patients with renal insufficiency (Table 47-5).

METABOLIC ALKALOSIS

Metabolic alkalosis is manifested by an elevated arterial pH, an increase in the serum [HCO_3^-], and an increase in Pa_{CO2} as a result of compensatory alveolar hypoventilation (Table 47-1). It is often accompanied by hypochloremia and hypokalemia. The arterial pH establishes the diagnosis, because it is increased in metabolic alkalosis and decreased or normal in respiratory acidosis. Metabolic alkalosis frequently occurs in association with other disorders such as respiratory acidosis or alkalosis or metabolic acidosis.

Pathogenesis

Metabolic alkalosis occurs as a result of net gain of [HCO_3^-] or loss of nonvolatile acid (usually HCl by vomiting) from the extracellular fluid. For HCO_3^- to be added to the extracellular fluid, it must be administered exogenously or synthesized endogenously, in part or entirely by the kidneys. Because it is unusual for alkali to be added to the body, the disorder involves a generative stage, in which the loss of acid usually causes alkalosis, and a maintenance stage, in which the kidneys fail to compensate by

excreting HCO_3^-.

Under normal circumstances, the kidneys have an impressive capacity to excrete HCO_3^-.

Continuation of metabolic alkalosis represents a failure of the kidneys to eliminate HCO_3^- in the usual manner. The kidneys will retain, rather than excrete, the excess alkali and maintain the alkalosis if (1) volume deficiency, chloride deficiency, and K^+ deficiency exist in combination with a reduced GFR, which augments distal tubule H^+ secretion; or (2) hypokalemia exists because of autonomous hyperaldosteronism. In the first example, alkalosis is corrected by administration of NaCl and KCl, whereas, in the latter, it is necessary to repair the alkalosis by pharmacologic or surgical intervention, not with saline administration.

Differential Diagnosis

To establish the cause of metabolic alkalosis **(Table 47-6)**, it is necessary to assess the status of the extracellular fluid volume (ECFV), the recumbent and upright blood pressure, the serum [K^+], and the renin-aldosterone system. For example, the presence of chronic hypertension and chronic hypokalemia in an alkalotic patient suggests either mineralocorticoid excess or that the hypertensive patient is receiving diuretics. Low plasma renin activity and normal urine [Na^+] and [Cl^-] in a patient who is not taking diuretics indicate a primary mineralocorticoid excess syndrome. The combination of hypokalemia and alkalosis in a normotensive, nonedematous patient can be due to Bartter's or Gitelman's syndrome, magnesium deficiency, vomiting, exogenous alkali, or diuretic ingestion.

Determination of urine electrolytes (especially the urine [Cl^-]) and screening of the urine for diuretics may be helpful. If the urine is alkaline, with an elevated [Na^+] and [K^+] but low [Cl^-], the diagnosis is usually either vomiting (overt or surreptitious) or alkali ingestion. If the urine is relatively acid and has low concentrations of Na^+, K^+, and Cl^-, the most likely possibilities are prior vomiting, the posthypercapnic state, or prior diuretic ingestion. If, on the other hand, neither the urine sodium, potassium, nor chloride concentrations are depressed, magnesium deficiency, Bartter's or Gitelman's syndrome, or current diuretic ingestion should be considered. Bartter's syndrome is distinguished from Gitelman's syndrome because of hypocalciuria and hypomagnesemia in the latter disorder.

Table 47-6 Causes of Metabolic Alkalosis

I. Exogenous HCO_3^- loads
A. Acute alkali administration
B. Milk-alkali syndrome
II. Effective ECFV contraction, normotension, K^+ deficiency, and secondary hyperreninemic hyperaldosteronism
A. Gastrointestinal origin
1. Vomiting
2. Gastric aspiration
3. Congenital chloridorrhea
4. Villous adenoma
B. Renal origin
1. Diuretics
2. Posthypercapnic state
3. Hypercalcemia/hypoparathyroidism
4. Recovery from lactic acidosis or ketoacidosis
5. Nonreabsorbable anions including penicillin, carbenicillin
6. Mg^{2+} deficiency
7. K^+ depletion
8. Bartter's syndrome (loss of function mutations in TALH)
9. Gitelman's syndrome (loss of function mutation in Na^+-Cl– cotransporter in DCT)

III. ECFV expansion, hypertension, K^+ deficiency, and mineralocorticoid excess
A. High renin
1. Renal artery stenosis
2. Accelerated hypertension
3. Renin-secreting tumor
4. Estrogen therapy
B. Low renin
1. Primary aldosteronism
a. Adenoma
b. Hyperplasia
c. Carcinoma
2. Adrenal enzyme defects
a. 11 ß-Hydroxylase deficiency
b. 17 α-Hydroxylase deficiency
3. Cushing's syndrome or disease
4. Other
a. Licorice
b. Carbenoxolone
c. Chewer's tobacco
IV. Gain-of-function mutation of renal sodium channel with ECFV expansion, hypertension, K^+ deficiency, and hyporeninemic-hypoaldosteronism
A. Liddle's syndrome

Abbreviations: DCT, distal convoluted tubule; ECFV, extracellular fluid volume; TALH, thick ascending limb of Henle's loop.

ALKALI ADMINISTRATION

Chronic administration of alkali to individuals with normal renal function rarely causes alkalosis. However, in patients with coexistent hemodynamic disturbances, alkalosis can develop because the normal capacity to excrete HCO_3^- may be exceeded or there may be enhanced reabsorption of HCO_3^-. Such patients include those who receive HCO_3^- (PO or IV), acetate loads (parenteral hyperalimentation solutions), citrate loads (transfusions), or antacids plus cation-exchange resins (aluminum hydroxide and sodium polystyrene sulfonate). Nursing-home patients receiving tube feedings have a higher incidence of metabolic alkalosis than nursing-home patients receiving oral feedings.

Metabolic Alkalosis Associated with ECFV Contraction, K^+ Depletion, and Secondary Hyperreninemic Hyperaldosteronism

GASTROINTESTINAL ORIGIN

Gastrointestinal loss of H^+ from vomiting or gastric aspiration results in retention of HCO_3^-. The loss of fluid and NaCl in vomitus or nasogastric suction results in contraction of the ECFV and an increase in the secretion of renin and aldosterone. Volume contraction through a reduction in GFR results in an enhanced capacity of the renal tubule to reabsorb HCO_3^-. During active vomiting, however, the filtered load of bicarbonate is acutely increased to the point that the reabsorptive capacity of the proximal tubule for HCO_3^- is exceeded. The excess NaHCO3 issuing out of the proximal tubule reaches the distal tubule, where H^+ secretion is enhanced by an aldosterone and the delivery of the poorly reabsorbed anion, HCO_3^-. Correction of the contracted ECFV with NaCl and repair of K^+ deficits corrects the acid-base disorder, and chloride deficiency.

RENAL ORIGIN

Diuretics

(See also Chap. 234) Drugs that induce chloruresis, such as thiazides and loop diuretics (furosemide, bumetanide, torsemide, and ethacrynic acid), acutely diminish the ECFV without altering the total body bicarbonate content. The serum [HCO_3^-] increases because the reduced ECFV "contracts" the

[HCO_3^-] in the plasma (contraction alkalosis). The chronic administration of diuretics tends to generate an alkalosis by increasing distal salt delivery, so that K^+ and H^+ secretion are stimulated. The alkalosis is maintained by persistence of the contraction of the ECFV, secondary hyperaldosteronism, K^+ deficiency, and the direct effect of the diuretic (as long as diuretic administration continues). Repair of the alkalosis is achieved by providing isotonic saline to correct the ECFV deficit.

Solute Losing Disorders: Bartter's Syndrome and Gitelman's Syndrome

See Chap. 284.

Nonreabsorbable Anions and Magnesium Deficiency

Administration of large quantities of nonreabsorbable anions, such as penicillin or carbenicillin, can enhance distal acidification and K^+ secretion by increasing the transepithelial potential difference. Mg^{2+} deficiency results in hypokalemic alkalosis by enhancing distal acidification through stimulation of renin and hence aldosterone secretion.

Potassium Depletion

Chronic K^+ depletion may cause metabolic alkalosis by increasing urinary acid excretion. Both NH_4^+ production and absorption are enhanced and HCO_3^- reabsorption is stimulated. Chronic K^+ deficiency upregulates the renal H^+, K^+-ATPase to increase K^+ absorption at the expense of enhanced H^+ secretion. Alkalosis associated with severe K^+ depletion is resistant to salt administration, but repair of the K^+ deficiency corrects the alkalosis.

After Treatment of Lactic Acidosis or Ketoacidosis

When an underlying stimulus for the generation of lactic acid or ketoacid is removed rapidly, as with repair of circulatory insufficiency or with insulin therapy, the lactate or ketones are metabolized to yield an equivalent amount of HCO_3^-. Other sources of new HCO_3^- are additive with the original amount generated by organic anion metabolism to create a surfeit of HCO_3^-. Such sources include (1) new HCO_3^- added to the blood by the kidneys as a result of enhanced acid excretion during the preexisting period of acidosis, and (2) alkali therapy during the treatment phase of the acidosis. Acidosis-induced contraction of the ECFV and K^+ deficiency act to sustain the alkalosis.

Posthypercapnia

Prolonged CO_2 retention with chronic respiratory acidosis enhances renal HCO_3^- absorption and the generation of new HCO_3^- (increased net acid excretion). If the Pa_{CO2} is returned to normal, metabolic alkalosis results from the persistently elevated [HCO_3^-]. Alkalosis develops if the elevated Pa_{CO2} is abruptly returned toward normal by a change in mechanically controlled ventilation. Associated ECFV contraction does not allow complete repair of the alkalosis by correction of the Pa_{CO2} alone, and alkalosis persists until Cl^- supplementation is provided.

Metabolic Alkalosis Associated with ECFV Expansion, Hypertension, and Hyperaldosteronism

Increased aldosterone levels may be the result of autonomous primary adrenal overproduction or of secondary aldosterone release due to renal overproduction of renin. Mineralocorticoid excess increases net acid excretion and may result in metabolic alkalosis, which may be worsened by associated K^+ deficiency. ECFV expansion from salt retention causes hypertension. The kaliuresis persists because of mineralocorticoid excess and distal Na^+ absorption causing enhanced K^+ excretion, continued K^+ depletion with polydipsia, inability to concentrate the urine, and polyuria.

Liddle's syndrome (Chap. 284) results from increased activity of the collecting duct Na^+ channel (ENaC) and is a rare monogenic form of hypertension due to volume expansion manifested as hypokalemic alkalosis and normal aldosterone levels.

SYMPTOMS

With metabolic alkalosis, changes in CNS and peripheral nervous system function are similar to those of hypocalcemia (Chap. 352); symptoms include mental confusion; obtundation; and a predisposition to seizures, paresthesia, muscular cramping, tetany, aggravation of arrhythmias, and hypoxemia in chronic obstructive pulmonary disease. Related electrolyte abnormalities include hypokalemia and hypophosphatemia.

Treatment: Metabolic Alkalosis

This is primarily directed at correcting the underlying stimulus for HCO_3^- generation. If primary aldosteronism, renal artery stenosis, or Cushing's syndrome is present, correction of the underlying cause will reverse the alkalosis. [H^+] loss by the stomach or kidneys can be mitigated by the use of proton pump inhibitors or the discontinuation of diuretics. The second aspect of treatment is to remove the factors that sustain the inappropriate increase in HCO_3^- reabsorption, such as ECFV contraction or K^+ deficiency. K^+ deficits should always be repaired. Isotonic saline is usually sufficient to reverse the alkalosis if ECFV contraction is present.

If associated conditions preclude infusion of saline, renal HCO_3^- loss can be accelerated by administration of acetazolamide, a carbonic anhydrase inhibitor, which is usually effective in patients with adequate renal function but can worsen K^+ losses. Dilute hydrochloric acid (0.1 *N* HCl) is also effective but can cause hemolysis, and must be delivered centrally and slowly. Hemodialysis against a dialysate low in [HCO_3^-] and high in [Cl^-] can be effective when renal function is impaired.

RESPIRATORY ACIDOSIS

Respiratory acidosis can be due to severe pulmonary disease, respiratory muscle fatigue, or abnormalities in ventilatory control and is recognized by an increase in Pa_{CO2} and decrease in pH **(Table 47-7)**. In acute respiratory acidosis, there is an immediate compensatory elevation (due to cellular buffering mechanisms) in HCO_3^-, which increases 1 mmol/L for every 10-mmHg increase in Pa_{CO2}. In chronic respiratory acidosis (>24 h), renal adaptation increases the [HCO_3^-] by 4 mmol/L for every 10-mmHg increase in Pa_{CO2}. The serum HCO_3^- usually does not increase above 38 mmol/L.

Table 47-7 Respiratory Acid-Base Disorders

I. Alkalosis
- A. Central nervous system stimulation
  - 1. Pain
  - 2. Anxiety, psychosis
  - 3. Fever
  - 4. Cerebrovascular accident
  - 5. Meningitis, encephalitis
  - 6. Tumor
  - 7. Trauma
- B. Hypoxemia or tissue hypoxia
  - 1. High altitude
  - 2. Pneumonia, pulmonary edema
  - 3. Aspiration
  - 4. Severe anemia
- C. Drugs or hormones
  - 1. Pregnancy, progesterone
  - 2. Salicylates
  - 3. Cardiac failure
- D. Stimulation of chest receptors
  - 1. Hemothorax
  - 2. Flail chest
  - 3. Cardiac failure
  - 4. Pulmonary embolism
- E. Miscellaneous
  - 1. Septicemia
  - 2. Hepatic failure
  - 3. Mechanical hyperventilation

- 4. Heat exposure
- 5. Recovery from metabolic acidosis

II. Acidosis
- A. Central
  - 1. Drugs (anesthetics, morphine, sedatives)
  - 2. Stroke
  - 3. Infection
- B. Airway
  - 1. Obstruction
  - 2. Asthma
- C. Parenchyma
  - 1. Emphysema
  - 2. Pneumoconiosis
  - 3. Bronchitis
  - 4. Adult respiratory distress syndrome
  - 5. Barotrauma
- D. Neuromuscular
  - 1. Poliomyelitis
  - 2. Kyphoscoliosis
  - 3. Myasthenia
  - 4. Muscular dystrophies
- E. Miscellaneous
  - 1. Obesity
  - 2. Hypoventilation
  - 3. Permissive hypercapnia

The clinical features vary according to the severity and duration of the respiratory acidosis, the underlying disease, and whether there is accompanying hypoxemia. A rapid increase in Pa_{co2} may cause anxiety, dyspnea, confusion, psychosis, and hallucinations and may progress to coma. Lesser degrees of dysfunction in chronic hypercapnia include sleep disturbances; loss of memory; daytime somnolence; personality changes; impairment of coordination; and motor disturbances such as tremor, myoclonic jerks, and asterixis. Headaches and other signs that mimic raised intracranial pressure, such as papilledema, abnormal reflexes, and focal muscle weakness, are due to vasoconstriction secondary to loss of the vasodilator effects of CO_2.

Depression of the respiratory center by a variety of drugs, injury, or disease can produce respiratory acidosis. This may occur acutely with general anesthetics, sedatives, and head trauma or chronically with sedatives, alcohol, intracranial tumors, and the syndromes of sleep-disordered breathing including the primary alveolar and obesity-hypoventilation syndromes (Chaps. 264 and 265). Abnormalities or disease in the motor neurons, neuromuscular junction, and skeletal muscle can cause hypoventilation via respiratory muscle fatigue. Mechanical ventilation, when not properly adjusted and supervised, may result in respiratory acidosis, particularly if CO_2 production suddenly rises (because of fever, agitation, sepsis, or overfeeding) or alveolar ventilation falls because of worsening pulmonary function. High levels of positive end-expiratory pressure in the presence of reduced cardiac output may cause hypercapnia as a result of large increases in alveolar dead space (Chap. 252). Permissive hypercapnia is being used with increasing frequency because of studies suggesting lower mortality rates than with conventional mechanical ventilation, especially with severe CNS or heart disease. The respiratory acidosis associated with permissive hypercapnia may require administration of NaHCO3 to increase the arterial pH to 7.25 but overcorrection of the acidemia may be deleterious.

Acute hypercapnia follows sudden occlusion of the upper airway or generalized bronchospasm as in severe asthma, anaphylaxis, inhalational burn, or toxin injury. Chronic hypercapnia and respiratory acidosis occur in end-stage obstructive lung disease. Restrictive disorders involving both the chest wall and the lungs can cause respiratory acidosis because the high metabolic cost of respiration causes ventilatory muscle fatigue. Advanced stages of intrapulmonary and extrapulmonary restrictive defects present as chronic respiratory acidosis.

The diagnosis of respiratory acidosis requires the measurement of Pa_{co2} and arterial pH. A detailed

history and physical examination often indicate the cause. Pulmonary function studies (Chap. 252), including spirometry, diffusion capacity for carbon monoxide, lung volumes, and arterial Pa_{CO2} and O_2 saturation, usually make it possible to determine if respiratory acidosis is secondary to lung disease. The workup for nonpulmonary causes should include a detailed drug history, measurement of hematocrit, and assessment of upper airway, chest wall, pleura, and neuromuscular function.

Treatment: Respiratory Acidosis

The management of respiratory acidosis depends on its severity and rate of onset. Acute respiratory acidosis can be life-threatening, and measures to reverse the underlying cause should be undertaken simultaneously with restoration of adequate alveolar ventilation. This may necessitate tracheal intubation and assisted mechanical ventilation. Oxygen administration should be titrated carefully in patients with severe obstructive pulmonary disease and chronic CO_2 retention who are breathing spontaneously (Chap. 260). When oxygen is used injudiciously, these patients may experience progression of the respiratory acidosis. Aggressive and rapid correction of hypercapnia should be avoided, because the falling Pa_{CO2} may provoke the same complications noted with acute respiratory alkalosis (i.e., cardiac arrhythmias, reduced cerebral perfusion, and seizures). The Pa_{CO2} should be lowered gradually in chronic respiratory acidosis, aiming to restore the Pa_{CO2} to baseline levels and to provide sufficient Cl^- and K^+ to enhance the renal excretion of HCO_3^-.

Chronic respiratory acidosis is frequently difficult to correct, but measures aimed at improving lung function (Chap. 260) can help some patients and forestall further deterioration in most.

RESPIRATORY ALKALOSIS

Alveolar hyperventilation decreases Pa_{CO2} and increases the HCO_3^-/Pa_{CO2} ratio, thus increasing pH (Table 47-7). Nonbicarbonate cellular buffers respond by consuming HCO_3^-. Hypocapnia develops when a sufficiently strong ventilatory stimulus causes CO_2 output in the lungs to exceed its metabolic production by tissues. Plasma pH and $[HCO_3^-]$ appear to vary proportionately with Pa_{CO2} over a range from 40–15 mmHg. The relationship between arterial $[H^+]$ concentration and Pa_{CO2} is ~0.7 mmol/L per mmHg (or 0.01 pH unit/mmHg), and that for plasma $[HCO_3^-]$ is 0.2 mmol/L per mmHg. Hypocapnia sustained for >2–6 h is further compensated by a decrease in renal ammonium and titratable acid excretion and a reduction in filtered HCO_3^- reabsorption. Full renal adaptation to respiratory alkalosis may take several days and requires normal volume status and renal function. The kidneys appear to respond directly to the lowered Pa_{CO2} rather than to alkalosis per se. In chronic respiratory alkalosis a 1-mmHg fall in Pa_{CO2} causes a 0.4- to 0.5-mmol/L drop in $[HCO_3^-]$ and a 0.3-mmol/L fall (or 0.003 rise in pH) in $[H^+]$.

The effects of respiratory alkalosis vary according to duration and severity but are primarily those of the underlying disease. Reduced cerebral blood flow as a consequence of a rapid decline in Pa_{CO2} may cause dizziness, mental confusion, and seizures, even in the absence of hypoxemia. The cardiovascular effects of acute hypocapnia in the conscious human are generally minimal, but in the anesthetized or mechanically ventilated patient, cardiac output and blood pressure may fall because of the depressant effects of anesthesia and positive-pressure ventilation on heart rate, systemic resistance, and venous return. Cardiac arrhythmias may occur in patients with heart disease as a result of changes in oxygen unloading by blood from a left shift in the hemoglobin-oxygen dissociation curve (Bohr effect). Acute respiratory alkalosis causes intracellular shifts of Na^+, K^+, and PO_4^{2-} and reduces free $[Ca^{2+}]$ by increasing the protein-bound fraction. Hypocapnia-induced hypokalemia is usually minor.

Chronic respiratory alkalosis is the most common acid-base disturbance in critically ill patients and, when severe, portends a poor prognosis. Many cardiopulmonary disorders manifest respiratory alkalosis in their early to intermediate stages, and the finding of normocapnia and hypoxemia in a patient with hyperventilation may herald the onset of rapid respiratory failure and should prompt an assessment to determine if the patient is becoming fatigued. Respiratory alkalosis is common during mechanical ventilation.

The hyperventilation syndrome may be disabling. Paresthesia; circumoral numbness; chest wall tightness or pain; dizziness; inability to take an adequate breath; and, rarely, tetany may be sufficiently stressful to perpetuate the disorder. Arterial blood-gas analysis demonstrates an acute or chronic respiratory alkalosis, often with hypocapnia in the range of 15–30 mmHg and no hypoxemia. CNS diseases or injury can produce several patterns of hyperventilation and sustained Pa_{CO2} levels of 20–30 mmHg. Hyperthyroidism, high caloric loads, and exercise raise the basal metabolic rate, but

ventilation usually rises in proportion so that arterial blood gases are unchanged and respiratory alkalosis does not develop. Salicylates are the most common cause of drug-induced respiratory alkalosis as a result of direct stimulation of the medullary chemoreceptor (**Chap. e49**). The methylxanthines, theophylline, and aminophylline stimulate ventilation and increase the ventilatory response to CO_2. Progesterone increases ventilation and lowers arterial Pa_{CO2} by as much as 5–10 mmHg. Therefore, chronic respiratory alkalosis is a common feature of pregnancy. Respiratory alkalosis is also prominent in liver failure, and the severity correlates with the degree of hepatic insufficiency. Respiratory alkalosis is often an early finding in gram-negative septicemia, before fever, hypoxemia, or hypotension develops.

The diagnosis of respiratory alkalosis depends on measurement of arterial pH and Pa_{CO2}. The plasma $[K^+]$ is often reduced and the $[Cl^-]$ increased. In the acute phase, respiratory alkalosis is not associated with increased renal HCO_3^- excretion, but within hours net acid excretion is reduced. In general, the HCO_3^- concentration falls by 2.0 mmol/L for each 10-mmHg decrease in Pa_{CO2}. Chronic hypocapnia reduces the serum $[HCO_3^-]$ by 4.0 mmol/L for each 10-mmHg decrease in Pa_{CO2}. It is unusual to observe a plasma $HCO_3^- < 12$ mmol/L as a result of a pure respiratory alkalosis.

When a diagnosis of respiratory alkalosis is made, its cause should be investigated. The diagnosis of hyperventilation syndrome is made by exclusion. In difficult cases, it may be important to rule out other conditions such as pulmonary embolism, coronary artery disease, and hyperthyroidism.

Treatment: Respiratory Alkalosis

The management of respiratory alkalosis is directed toward alleviation of the underlying disorder. If respiratory alkalosis complicates ventilator management, changes in dead space, tidal volume, and frequency can minimize the hypocapnia. Patients with the hyperventilation syndrome may benefit from reassurance, rebreathing from a paper bag during symptomatic attacks, and attention to underlying psychological stress. Antidepressants and sedatives are not recommended. β-Adrenergic blockers may ameliorate peripheral manifestations of the hyperadrenergic state.

FURTHER READINGS

DuBose TD: Metabolic alkalosis, in *Primer on Kidney Diseases,* 5th ed, A Greenberg (ed). Saunders Elsevier, 2009, pp 84-90

Dubose TD Jr: Acid-base disorders, in *Brenner and Rector's The Kidney* , 8th ed, BM Brenner (ed). Philadelphia, Saunders, 2008, pp 505-546

———, Alpern RJ: Renal tubular acidosis, in *The Metabolic and Molecular Bases of Inherited Disease*, 8th ed, CR Scriver et al (eds). New York, McGraw-Hill, 2001

Kraut JA, Madias NE: Metabolic acidosis: Pathophysiology, diagnosis and management. Nat Rev Nephrol. 6:274, 2010 [PMID: 20308999]

Laski ME, Wesson DE: Lactic acidosis, in *Acid-Base and Electrolyte Disorders—A Companion to Brenner and Rector's The Kidney*, TD DuBose, LL Hamm (eds). Philadelphia, Saunders, 2002, pp 83–107

Madias NE: Respiratory alkalosis, in *Acid-Base and Electrolyte Disorders—A Companion to Brenner and Rector's The Kidney*, TD DuBose, LL Hamm (eds). Philadelphia, Saunders, 2002, pp 147–164

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SEXUAL DYSFUNCTION: INTRODUCTION

Male sexual dysfunction affects 10–25% of middle-aged and elderly men, and female sexual dysfunction occurs with a similar frequency. Demographic changes, the popularity of newer treatments, and greater awareness of sexual dysfunction by patients and society have led to increased diagnosis and associated health care expenditures for the management of this common disorder. Because many patients are reluctant to initiate discussion of their sex lives, physicians should address this topic directly to elicit a history of sexual dysfunction.

MALE SEXUAL DYSFUNCTION

Physiology of Male Sexual Response

Normal male sexual function requires (1) an intact libido, (2) the ability to achieve and maintain penile erection, (3) ejaculation, and (4) detumescence. *Libido* refers to sexual desire and is influenced by a variety of visual, olfactory, tactile, auditory, imaginative, and hormonal stimuli. Sex steroids, particularly testosterone, act to increase libido. Libido can be diminished by hormonal or psychiatric disorders and by medications.

Penile tumescence leading to erection depends on an increased flow of blood into the lacunar network accompanied by complete relaxation of the arteries and corporal smooth muscle. The microarchitecture of the corpora is composed of a mass of smooth muscle (trabecula) that contains a network of endothelial-lined vessels (lacunar spaces). Subsequent compression of the trabecular smooth muscle against the fibroelastic tunica albuginea causes a passive closure of the emissary veins and accumulation of blood in the corpora. In the presence of a full erection and a competent valve mechanism, the corpora become noncompressible cylinders from which blood does not escape.

The central nervous system (CNS) exerts an important influence by either stimulating or antagonizing spinal pathways that mediate erectile function and ejaculation. The erectile response is mediated by a combination of central (psychogenic) innervation and peripheral (reflexogenic) innervation. Sensory nerves that originate from receptors in the penile skin and glans converge to form the dorsal nerve of the penis, which travels to the S2-S4 dorsal root ganglia via the pudendal nerve. Parasympathetic nerve fibers to the penis arise from neurons in the intermediolateral columns of the S2-S4 sacral spinal segments. Sympathetic innervation originates from the T-11 to the L-2 spinal segments and descends through the hypogastric plexus.

Neural input to smooth-muscle tone is crucial to the initiation and maintenance of an erection. There is also an intricate interaction between the corporal smooth-muscle cell and its overlying endothelial cell lining **(Fig. 48-1*A*)**. Nitric oxide, which induces vascular relaxation, promotes erection and is opposed by endothelin 1 (ET-1) and Rho kinase, which mediate vascular contraction. Nitric oxide is synthesized from l-arginine by nitric oxide synthase and is released from the nonadrenergic, noncholinergic (NANC) autonomic nerve supply to act postjunctionally on smooth-muscle cells. Nitric oxide increases the production of cyclic 3',5'-guanosine monophosphate (cyclic GMP), which induces relaxation of smooth muscle (Fig. 48-1*B*). Cyclic GMP is gradually broken down by phosphodiesterase type 5 (PDE-5). Inhibitors of PDE-5 such as the oral medications sildenafil, vardenafil, and tadalafil maintain erections by reducing the breakdown of cyclic GMP. However, if nitric oxide is not produced at some level, PDE-5 inhibitors are ineffective, as these drugs facilitate, but do not initiate, the initial enzyme cascade. In addition to nitric oxide, vasoactive prostaglandins (PGE1, $PGF_{2\alpha}$) are synthesized within the cavernosal tissue and increase cyclic AMP levels, also leading to relaxation of cavernosal smooth-muscle cells.

Figure 48-1

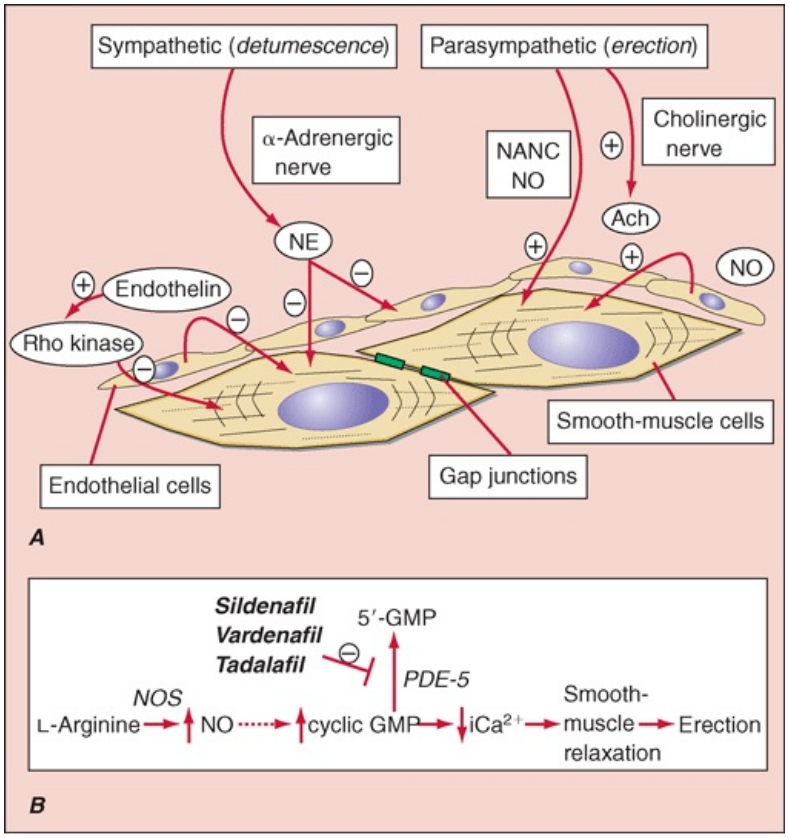

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Figure 48-1 Pathways that control erection and detumescence. *A.* Erection is mediated by cholinergic parasympathetic pathways and nonadrenergic, noncholinergic (NANC) pathways, which release nitric oxide (NO). Endothelial cells also release NO, which induces vascular smooth-muscle cell relaxation, allowing enhanced blood flow and leading to erection. Detumescence is mediated by sympathetic pathways that release norepinephrine and stimulate α-adrenergic pathways, leading to contraction of vascular smooth-muscle cells. Endothelin, released from endothelial cells, also induces contraction. Rho kinase activation via endothelin activity (among others) also contributes to detumescence by alteration of calcium signaling. ***B.*** Biochemical pathways of NO synthesis and action. Sildenafil, vardenafil, and tadalafil enhance erectile function by inhibiting phosphodiesterase type 5 (PDE-5), thereby maintaining high levels of cyclic 3',5'-guanosine monophosphate (cyclic GMP). NOS, nitric oxide synthase; iCa^{2+}, intracellular calcium.

Ejaculation is stimulated by the sympathetic nervous system; this results in contraction of the epididymis, vas deferens, seminal vesicles, and prostate, causing seminal fluid to enter the urethra. Seminal fluid emission is followed by rhythmic contractions of the bulbocavernosus and ischiocavernosus muscles, leading to ejaculation. *Premature ejaculation* usually is related to anxiety or a learned behavior and is amenable to behavioral therapy or treatment with medications such as selective serotonin reuptake inhibitors (SSRIs). *Retrograde ejaculation* results when the internal urethral sphincter does not close; it may occur in men with diabetes or after surgery involving the bladder neck.

Detumescence is mediated by norepinephrine from the sympathetic nerves, endothelin from the vascular surface, and smooth-muscle contraction induced by postsynaptic α-adrenergic receptors and activation of Rho kinase. These events increase venous outflow and restore the flaccid state. Venous leak can cause premature detumescence and is caused by insufficient relaxation of the corporal smooth muscle rather than a specific anatomic defect. *Priapism* refers to a persistent and painful erection and may be associated with sickle cell anemia, hypercoagulable states, spinal cord injury, or injection of vasodilator agents into the penis.

Erectile Dysfunction

EPIDEMIOLOGY

Erectile dysfunction (ED) is not considered a normal part of the aging process. Nonetheless, it is associated with certain physiologic and psychological changes related to age. In the Massachusetts Male Aging Study (MMAS), a community-based survey of men age 40–70, 52% of responders reported some degree of ED. Complete ED occurred in 10% of respondents, moderate ED in 25%, and minimal ED in 17%. The incidence of moderate or severe ED more than doubled between the ages of 40 and 70. In the National Health and Social Life Survey (NHSLS), which included a sample of men and

women age 18–59, 10% of men reported being unable to maintain an erection (corresponding to the proportion of men in the MMAS reporting severe ED). Incidence was highest among men in the age group 50–59 (21%) and men who were poor (14%), divorced (14%), and less educated (13%).

The incidence of ED is also higher among men with certain medical disorders, such as diabetes mellitus, obesity, lower urinary tract symptoms secondary to benign prostatic hyperplasia (BPH), heart disease, hypertension, and decreased high-density lipoprotein (HDL) levels. Cardiovascular disease and ED share etiologies as well as pathophysiology (e.g., endothelial dysfunction), and the degree of ED appears to correlate with the severity of cardiovascular disease. Consequently, ED represents a "sentinel symptom" in patients with occult cardiovascular and peripheral vascular disease.

Smoking is also a significant risk factor in the development of ED. Medications used in treating diabetes or cardiovascular disease are additional risk factors (see below). There is a higher incidence of ED among men who have undergone radiation or surgery for prostate cancer and in those with a lower spinal cord injury. Psychological causes of ED include depression, anger, stress from unemployment, and other stress-related causes.

PATHOPHYSIOLOGY

ED may result from three basic mechanisms: (1) failure to initiate (psychogenic, endocrinologic, or neurogenic), (2) failure to fill (arteriogenic), and (3) failure to store adequate blood volume within the lacunar network (venoocclusive dysfunction). These categories are not mutually exclusive, and multiple factors contribute to ED in many patients. For example, diminished filling pressure can lead secondarily to venous leak. Psychogenic factors frequently coexist with other etiologic factors and should be considered in all cases. Diabetic, atherosclerotic, and drug-related causes account for >80% of cases of ED in older men.

Vasculogenic

The most common organic cause of ED is a disturbance of blood flow to and from the penis. Atherosclerotic or traumatic arterial disease can decrease flow to the lacunar spaces, resulting in decreased rigidity and an increased time to full erection. Excessive outflow through the veins despite adequate inflow also may contribute to ED. Structural alterations to the fibroelastic components of the corpora may cause a loss of compliance and inability to compress the tunical veins. This condition may result from aging, increased cross-linking of collagen fibers induced by nonenzymatic glycosylation, hypoxemia, or altered synthesis of collagen associated with hypercholesterolemia.

Neurogenic

Disorders that affect the sacral spinal cord or the autonomic fibers to the penis preclude nervous system relaxation of penile smooth muscle, thus leading to ED. In patients with spinal cord injury, the degree of ED depends on the completeness and level of the lesion. Patients with incomplete lesions or injuries to the upper part of the spinal cord are more likely to retain erectile capabilities than are those with complete lesions or injuries to the lower part. Although 75% of patients with spinal cord injuries have some erectile capability, only 25% have erections sufficient for penetration. Other neurologic disorders commonly associated with ED include multiple sclerosis and peripheral neuropathy. The latter is often due to either diabetes or alcoholism. Pelvic surgery may cause ED through disruption of the autonomic nerve supply.

Endocrinologic

Androgens increase libido, but their exact role in erectile function is unclear. Individuals with castrate levels of testosterone can achieve erections from visual or sexual stimuli. Nonetheless, normal levels of testosterone appear to be important for erectile function, particularly in older males. Androgen replacement therapy can improve depressed erectile function when it is secondary to hypogonadism; however, it is not useful for ED when endogenous testosterone levels are normal. Increased prolactin may decrease libido by suppressing gonadotropin-releasing hormone (GnRH), and it also leads to decreased testosterone levels. Treatment of hyperprolactinemia with dopamine agonists can restore libido and testosterone.

Diabetic

ED occurs in 35–75% of men with diabetes mellitus. Pathologic mechanisms are related primarily to diabetes-associated vascular and neurologic complications. Diabetic macrovascular complications are related mainly to age, whereas microvascular complications correlate with the duration of diabetes and the degree of glycemic control (Chap. 344). Individuals with diabetes also have reduced amounts of nitric oxide synthase in both endothelial and neural tissues.

Psychogenic

Two mechanisms contribute to the inhibition of erections in psychogenic ED. First, psychogenic stimuli to the sacral cord may inhibit reflexogenic responses, thereby blocking activation of vasodilator outflow to the penis. Second, excess sympathetic stimulation in an anxious man may increase penile smooth-muscle tone. The most common causes of psychogenic ED are performance anxiety, depression, relationship conflict, loss of attraction, sexual inhibition, conflicts over sexual preference, sexual abuse

in childhood, and fear of pregnancy or sexually transmitted disease. Almost all patients with ED, even when it has a clear-cut organic basis, develop a psychogenic component as a reaction to ED.

Medication-Related

Medication-induced ED **(Table 48-1)** is estimated to occur in 25% of men seen in general medical outpatient clinics. The adverse effects related to drug therapy are additive, especially in older men. In addition to the drug itself, the disease being treated is likely to contribute to sexual dysfunction. Among the antihypertensive agents, the thiazide diuretics and beta blockers have been implicated most frequently. Calcium channel blockers and angiotensin converting-enzyme inhibitors are cited less frequently. These drugs may act directly at the corporal level (e.g., calcium channel blockers) or indirectly by reducing pelvic blood pressure, which is important in the development of penile rigidity. α-Adrenergic blockers are less likely to cause ED. Estrogens, GnRH agonists, H_2 antagonists, and spironolactone cause ED by suppressing gonadotropin production or by blocking androgen action. Antidepressant and antipsychotic agents—particularly neuroleptics, tricyclics, and SSRIs—are associated with erectile, ejaculatory, orgasmic, and sexual desire difficulties.

Table 48-1 Drugs Associated with Erectile Dysfunction

Classification	Drugs
Diuretics	Thiazides Spironolactone
Antihypertensives	Calcium channel blockers Methyldopa Clonidine Reserpine Beta blockers Guanethidine
Cardiac/antihyperlipidemics	Digoxin Gemfibrozil Clofibrate
Antidepressants	Selective serotonin reuptake inhibitors Tricyclic antidepressants Lithium Monoamine oxidase inhibitors
Tranquilizers	Butyrophenones Phenothiazines
H_2 antagonists	Ranitidine Cimetidine
Hormones	Progesterone Estrogens Corticosteroids GnRH agonists 5α-Reductase inhibitors Cyproterone acetate
Cytotoxic agents	Cyclophosphamide Methotrexate Roferon-A
Anticholinergics	Disopyramide Anticonvulsants
Recreational	Ethanol Cocaine

	Marijuana

Abbreviation: GnRH, gonadotropin-releasing hormone.

If there is a strong association between the institution of a drug and the onset of ED, alternative medications should be considered. Otherwise, it is often practical to treat the ED without attempting multiple changes in medications, as it may be difficult to establish a causal role for a drug.

Figure 48-2

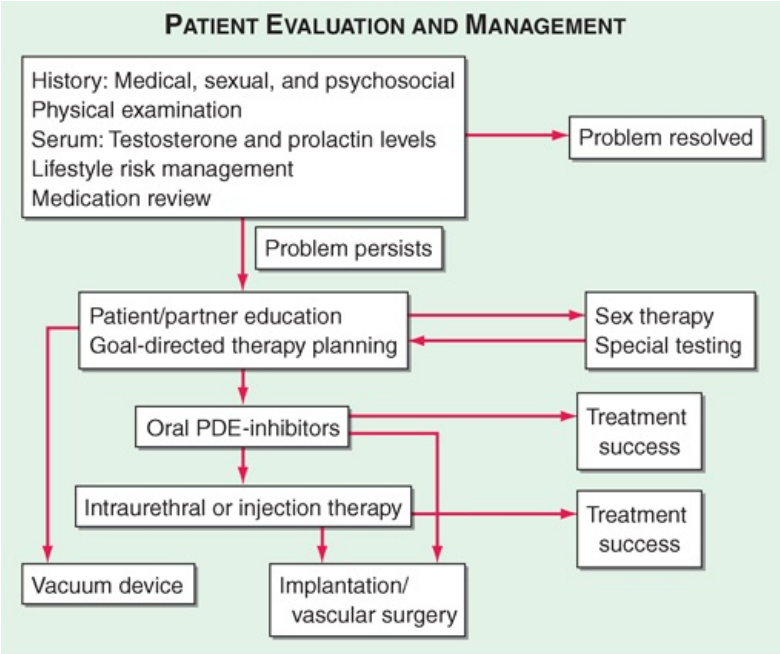

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the evaluation and management of patients with ED. PDE, phosphodiesterase.

Approach to the Patient: Erectile Dysfunction

A good physician-patient relationship helps unravel the possible causes of ED, many of which require discussion of personal and sometimes embarrassing topics. For this reason, a primary care provider is often ideally suited to initiate the evaluation. However, a significant percentage of men experience ED and remain undiagnosed unless specifically questioned about this issue. By far the most common reason for underreporting of ED is patient embarrassment. Once the topic is initiated by the physician, patients are more willing to discuss their potency issues. A complete medical and sexual history should be taken in an effort to assess whether the cause of ED is organic, psychogenic, or multifactorial (**Fig. 48-2**).

Figure 48-2

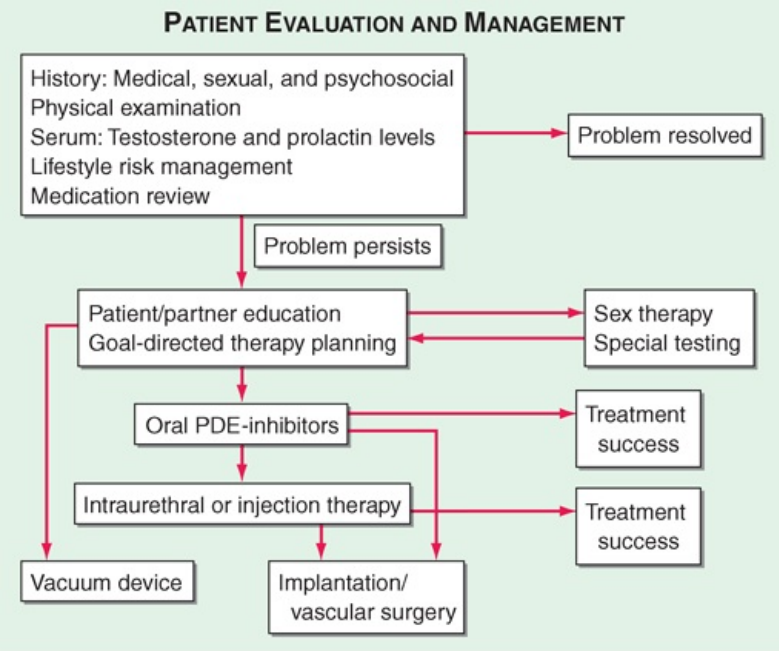

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the evaluation and differential diagnosis of hirsutism. ACTH, adrenocorticotropic hormone; CAH, congenital adrenal hyperplasia; DHEAS, sulfated form of dehydroepiandrosterone; PCOS, polycystic ovarian syndrome.

Both the patient and his sexual partner should be interviewed regarding sexual history. ED should be distinguished from other sexual problems, such as premature ejaculation. Lifestyle factors such as sexual orientation, the patient's distress from ED, performance anxiety, and details of sexual techniques should be addressed. Standardized questionnaires are available to assess ED, including the International Index of Erectile Function (IIEF) and the more easily administered Sexual Health Inventory for Men (SHIM), a validated abridged version of the IIEF.

The initial evaluation of ED begins with a review of the patient's medical, surgical, sexual, and psychosocial histories. The history should note whether the patient has experienced pelvic trauma, surgery, or radiation. In light of the increasing recognition of the relationship between lower urinary tract symptoms and ED, it is advisable to evaluate for the presence of symptoms of bladder outlet obstruction. Questions should focus on the onset of symptoms, the presence and duration of partial erections, and the progression of ED. A history of nocturnal or early morning erections is useful for distinguishing physiologic ED from psychogenic ED. Nocturnal erections occur during rapid eye movement (REM) sleep and require intact neurologic and circulatory systems. Organic causes of ED generally are characterized by a gradual and persistent change in rigidity or the inability to sustain nocturnal, coital, or self-stimulated erections. The patient should be questioned about the presence of penile curvature or pain with coitus. It is also important to address libido, as decreased sexual drive and ED are sometimes the earliest signs of endocrine abnormalities (e.g., increased prolactin, decreased testosterone levels). It is useful to ask whether the problem is confined to coitus with one partner or also involves other partners; ED not uncommonly arises in association with new or extramarital sexual relationships. Situational ED, as opposed to consistent ED, suggests psychogenic causes. Ejaculation is much less commonly affected than erection, but questions should be asked about whether ejaculation is normal, premature, delayed, or absent. Relevant risk factors should be identified, such as diabetes mellitus, coronary artery disease (CAD), and neurologic disorders. The patient's surgical history should be explored with an emphasis on bowel, bladder, prostate, and vascular procedures. A complete drug history is also important. Social changes that may precipitate ED are also crucial to the evaluation, including health worries, spousal death, divorce, relationship difficulties, and financial concerns.

Because ED commonly involves a host of endothelial cell risk factors, men with ED report higher rates of overt and silent myocardial infarction. Therefore, ED in an otherwise asymptomatic male warrants consideration of other vascular disorders, including CAD.

The physical examination is an essential element in the assessment of ED. Signs of hypertension as well as evidence of thyroid, hepatic, hematologic, cardiovascular, or renal diseases should be sought. An assessment should be made of the endocrine and vascular systems, the external genitalia, and the prostate gland. The penis should be palpated carefully along the corpora to detect fibrotic plaques. Reduced testicular size and loss of secondary sexual characteristics are suggestive of hypogonadism. Neurologic examination should include assessment of anal sphincter tone, investigation of the

bulbocavernosus reflex, and testing for peripheral neuropathy.

Although hyperprolactinemia is uncommon, a serum prolactin level should be measured, as decreased libido and/or ED may be the presenting symptoms of a prolactinoma or another mass lesion of the sella (Chap. 339). The serum testosterone level should be measured, and if it is low, gonadotropins should be measured to determine whether hypogonadism is primary (testicular) or secondary (hypothalamic-pituitary) in origin (Chap. 346). If not performed recently, serum chemistries, complete blood count (CBC), and lipid profiles may be of value, as they can yield evidence of anemia, diabetes, hyperlipidemia, or other systemic diseases associated with ED. Determination of serum prostate-specific antigen (PSA) should be conducted according to recommended clinical guidelines (Chap. 95).

Additional diagnostic testing is rarely necessary in the evaluation of ED. However, in selected patients, specialized testing may provide insight into pathologic mechanisms of ED and aid in the selection of treatment options. Optional specialized testing includes (1) studies of nocturnal penile tumescence and rigidity, (2) vascular testing (in-office injection of vasoactive substances, penile Doppler ultrasound, penile angiography, dynamic infusion cavernosography/cavernosometry), (3) neurologic testing (biothesiometry-graded vibratory perception, somatosensory evoked potentials), and (4) psychological diagnostic tests. The information potentially gained from these procedures must be balanced against their invasiveness and cost.

Treatment: Male Sexual Dysfunction

PATIENT EDUCATION

Patient and partner education is essential in the treatment of ED. In goal-directed therapy, education facilitates understanding of the disease, the results of the tests, and the selection of treatment. Discussion of treatment options helps clarify how treatment is best offered and stratify first- and second-line therapies. Patients with high-risk lifestyle issues such as obesity, smoking, alcohol abuse, and recreational drug use should be counseled on the role those factors play in the development of ED.

Therapies currently employed for the treatment of ED include oral phosphodiesterase type 5 inhibitor therapy (most commonly used), injection therapies, testosterone therapy, penile devices, and psychological therapy. In addition, limited data suggest that treatments for underlying risk factors and comorbidities—for example, weight loss, exercise, stress reduction, and smoking cessation—may improve erectile function. Decisions regarding therapy should take into account the preferences and expectations of patients and their partners.

ORAL AGENTS

Sildenafil, tadalafil, and vardenafil are the only approved and effective oral agents for the treatment of ED. These three medications have markedly improved the management of ED because they are effective for the treatment of a broad range of causes, including psychogenic, diabetic, vasculogenic, post-radical prostatectomy (nerve-sparing procedures), and spinal cord injury. They belong to a class of medications that are selective and potent inhibitors of PDE-5, the predominant phosphodiesterase isoform found in the penis. They are administered in graduated doses and enhance erections after sexual stimulation. The onset of action is approximately 60–120 min, depending on the medication used and other factors, such as recent food intake. Reduced initial doses should be considered for patients who are elderly, are taking concomitant alpha blockers, have renal insufficiency, or are taking medications that inhibit the CYP3A4 metabolic pathway in the liver (e.g., erythromycin, cimetidine, ketoconazole, and possibly itraconazole and mibefradil), as they may increase the serum concentration of the PDE-5 inhibitors (PDE-5i) or promote hypotension.

Several randomized trials have demonstrated the efficacy of this class of medications. There are no compelling data to support the superiority of one PDE-5i over another.

Patients may fail to respond to a PDE-5i for several reasons (**Table 48-2).** Some patients may not tolerate PDE-5i secondary to adverse events from vasodilation in nonpenile tissues expressing PDE-5 or from the inhibition of homologous nonpenile isozymes (i.e., PDE-6 found in the retina). Abnormal vision attributed to the effects of PDE-5i on retinal PDE-6 is of short duration, reported only with sildenafil and not thought to be clinically significant. A more serious concern is the possibility that PDE-5i may cause nonarteritic anterior ischemic optic neuropathy; although data to support that association are limited, it is prudent to avoid the use of these agents in men with a prior history of nonarteritic anterior ischemic optic neuropathy.

Table 48-2 Issues to Consider If Patients Report Failure of PDE-5i to Improve Erectile Dysfunction

- A trial of medication on at least six different days at the maximal dose should be made before declaring patient nonresponsive to PDE-5i use
- Taking medication after a high-fat meal
- Failure to include physical and psychic stimulation at the time of foreplay to induce endogenous NO

- Unrecognized hypogonadism

Abbreviations: NO, nitric oxide; PDE-5i, phosphodiesterase type 5 inhibitor.

Testosterone supplementation combined with a PDE-5i may be beneficial in improving erectile function in hypogonadal men with ED who are unresponsive to PDE-5i alone. These drugs do not affect ejaculation, orgasm, or sexual drive. Side effects associated with PDE-5i include headaches (19%), facial flushing (9%), dyspepsia (6%), and nasal congestion (4%). Approximately 7% of men using sildenafil may experience transient altered color vision (blue halo effect), and 6% of men taking tadalafil may experience loin pain. PDE-5i is contraindicated in men receiving nitrate therapy for cardiovascular disease, including agents delivered by the oral, sublingual, transnasal, and topical routes. These agents can potentiate its hypotensive effect and may result in profound shock. Likewise, amyl/butyl nitrate "poppers" may have a fatal synergistic effect on blood pressure. PDE-5i also should be avoided in patients with congestive heart failure and cardiomyopathy because of the risk of vascular collapse. Because sexual activity leads to an increase in physiologic expenditure [5–6 metabolic equivalents (METS)], physicians have been advised to exercise caution in prescribing any drug for sexual activity to those with active coronary disease, heart failure, borderline hypotension, or hypovolemia and to those on complex antihypertensive regimens.

Although the various forms of PDE-5i have a common mechanism of action, there are a few differences among the three agents. Tadalafil is unique in its longer half-life. All three drugs are effective for patients with ED of all ages, severities, and etiologies. Although there are pharmacokinetic and pharmacodynamic differences among these agents, clinically relevant differences are not clear.

ANDROGEN THERAPY

Testosterone replacement is used to treat both primary and secondary causes of hypogonadism (Chap. 346). Androgen supplementation in the setting of normal testosterone is rarely efficacious in the treatment of ED and is discouraged. Methods of androgen replacement include transdermal patches and gels, parenteral administration of long-acting testosterone esters (enanthate and cypionate), and oral preparations (17 α-alkylated derivatives) (Chap. 346). Oral androgen preparations have the potential for hepatotoxicity and should be avoided.

Men who receive testosterone should be reevaluated after 1–3 months and at least annually thereafter for testosterone levels, erectile function, and adverse effects, which may include gynecomastia, sleep apnea, development or exacerbation of lower urinary tract symptoms or benign prostatic hyperplasia, prostate cancer, lowering of HDL, erythrocytosis, elevations of liver function tests, and reduced fertility. Periodic reevaluation should include measurement of CBC and PSA and digital rectal exam. Therapy should be discontinued in patients who do not respond within 3 months.

VACUUM CONSTRICTION DEVICES

Vacuum constriction devices (VCDs) are a well-established noninvasive therapy. They are a reasonable treatment alternative for select patients who cannot take sildenafil or do not desire other interventions. VCDs draw venous blood into the penis and use a constriction ring to restrict venous return and maintain tumescence. Adverse events with VCD include pain, numbness, bruising, and altered ejaculation. Additionally, many patients complain that the devices are cumbersome and that the induced erections have a nonphysiologic appearance and feel.

INTRAURETHRAL ALPROSTADIL

If a patient fails to respond to oral agents, a reasonable next choice is intraurethral or self-injection of vasoactive substances. Intraurethral prostaglandin E1 (alprostadil), in the form of a semisolid pellet (doses of 125–1000 μg), is delivered with an applicator. Approximately 65% of men receiving intraurethral alprostadil respond with an erection when tested in the office, but only 50% achieve successful coitus at home. Intraurethral insertion is associated with a markedly reduced incidence of priapism in comparison to intracavernosal injection.

INTRACAVERNOSAL SELF-INJECTION

Injection of synthetic formulations of alprostadil is effective in 70–80% of patients with ED, but discontinuation rates are high because of the invasive nature of administration. Doses range between 1 and 40 μg. Injection therapy is contraindicated in men with a history of hypersensitivity to the drug and men at risk for priapism (hypercoagulable states, sickle cell disease). Side effects include local adverse events, prolonged erections, pain, and fibrosis with chronic use. Various combinations of alprostadil, phentolamine, and/or papaverine sometimes are used.

SURGERY

A less frequently used form of therapy for ED involves the surgical implantation of a semirigid or inflatable penile prosthesis. The choice of prosthesis is dependent on patient preference and should take into account body habitus and manual dexterity, which may affect the ability of the patient to manipulate the device. Because of the permanence of prosthetic devices, patients should be advised to first consider less invasive options for treatment. These surgical treatments are invasive, are

associated with potential complications, and generally are reserved for treatment of refractory ED. Despite their high cost and invasiveness, penile prostheses are associated with high rates of patient and partner satisfaction.

SEX THERAPY

A course of sex therapy may be useful for addressing specific interpersonal factors that may affect sexual functioning. Sex therapy generally consists of in-session discussion and at-home exercises specific to the person and the relationship. Psychosexual therapy involves techniques such as sensate focus (nongenital massage), sensory awareness exercises, correction of misconceptions about sexuality, and interpersonal difficulties therapy (e.g., open communication about sexual issues, physical intimacy scheduling, and behavioral interventions). These approaches may be useful in patients who have psychogenic or social components to their ED, although data from randomized trials are scanty and inconsistent. It is preferable if therapy includes both partners if the patient is involved in an ongoing relationship.

FEMALE SEXUAL DYSFUNCTION

Female sexual dysfunction (FSD) has traditionally included disorders of desire, arousal, pain, and muted orgasm. The associated risk factors for FSD are similar to those in males: cardiovascular disease, endocrine disorders, hypertension, neurologic disorders, and smoking **(Table 48-3)**.

Table 48-3 Risk Factors for Female Sexual Dysfunction

Neurologic disease: stroke, spinal cord injury, parkinsonism
Trauma, genital surgery, radiation
Endocrinopathies: diabetes, hyperprolactinemia
Liver and/or renal failure
Cardiovascular disease
Psychological factors and interpersonal relationship disorders: sexual abuse, life stressors
Medications
Antiandrogens: cimetidine, spironolactone
Antidepressants, alcohol, hypnotics, sedatives
Antiestrogens or GnRH antagonists
Antihistamines, sympathomimetic amines
Antihypertensives: diuretics, calcium channel blockers
Alkylating agents
Anticholinergics

Abbreviation: GnRH, gonadotropin-releasing hormone.

Epidemiology

Epidemiologic data are limited, but the available estimates suggest that as many as 43% of women complain of at least one sexual problem. Despite the recent interest in organic causes of FSD, desire and arousal phase disorders (including lubrication complaints) remain the most common presenting problems when surveyed in a community-based population.

Physiology of the Female Sexual Response

The female sexual response requires the presence of estrogens. A role for androgens is also likely but less well established. In the CNS, estrogens and androgens work synergistically to enhance sexual arousal and response. A number of studies report enhanced libido in women during preovulatory phases of the menstrual cycle, suggesting that hormones involved in the ovulatory surge (e.g., estrogens) increase desire.

Sexual motivation is heavily influenced by context, including the environment and partner factors. Once sufficient sexual desire is reached, sexual arousal is mediated by the central and autonomic nervous systems. Cerebral sympathetic outflow is thought to increase desire, and peripheral parasympathetic activity results in clitoral vasocongestion and vaginal secretion (lubrication).

The neurotransmitters for clitoral corporal engorgement are similar to those in the male, with a prominent role for neural, smooth-muscle, and endothelial released nitric oxide (NO). A fine network of vaginal nerves and arterioles promotes a vaginal transudate. The major transmitters of this complex vaginal response are not certain, but roles for NO and vasointestinal polypeptide (VIP) are suspected. Investigators studying the normal female sexual response have challenged the long-held construct of a linear and unmitigated relationship between initial desire, arousal, vasocongestion, lubrication, and

eventual orgasm. Caregivers should consider a paradigm of a positive emotional and physical outcome with one, many, or no orgasmic peak and release.

Although there are anatomic differences as well as variation in the density of vascular and neural beds in males and females, the primary effectors of sexual response are strikingly similar. Intact sensation is important for arousal. Thus, reduced levels of sexual functioning are more common in women with peripheral neuropathies (e.g., diabetes). Vaginal lubrication is a transudate of serum that results from the increased pelvic blood flow associated with arousal. Vascular insufficiency from a variety of causes may compromise adequate lubrication and result in dyspareunia. Cavernosal and arteriole smooth-muscle relaxation occurs via increased nitric oxide synthase (NOS) activity and produces engorgement in the clitoris and the surrounding vestibule. Orgasm requires an intact sympathetic outflow tract; hence, orgasmic disorders are common in female patients with spinal cord injuries.

Approach to the Patient: Female Sexual Dysfunction

Many women do not volunteer information about their sexual response. Open-ended questions in a supportive atmosphere are helpful in initiating a discussion of sexual fitness in women who are reluctant to discuss such issues. Once a complaint has been voiced, a comprehensive evaluation should be performed, including a medical history, a psychosocial history, a physical examination, and limited laboratory testing.

The history should include the usual medical, surgical, obstetric, psychological, gynecologic, sexual, and social information. Past experiences, intimacy, knowledge, and partner availability should also be ascertained. Medical disorders that may affect sexual health should be delineated. They include diabetes, cardiovascular disease, gynecologic conditions, obstetric history, depression, anxiety disorders, and neurologic disease. Medications should be reviewed as they may affect arousal, libido, and orgasm. The need for counseling and recognizing life stresses should be identified. The physical examination should assess the genitalia, including the clitoris. Pelvic floor examination may identify prolapse or other disorders. Laboratory studies are needed, especially if menopausal status is uncertain. Estradiol, follicle-stimulating hormone (FSH), and luteinizing hormone (LH) are usually obtained, and dehydroepiandrosterone (DHEA) should be considered as it reflects adrenal androgen secretion. A CBC, liver function assessment, and lipid studies may be useful, if not otherwise obtained. Complicated diagnostic evaluation such as clitoral Doppler ultrasonography and biothesiometry require expensive equipment and are of uncertain utility. It is important for the patient to identify which symptoms are most distressing.

The evaluation of FSD previously occurred mainly in a psychosocial context. However, inconsistencies between diagnostic categories based only on psychosocial considerations and the emerging recognition of organic etiologies have led to a new classification of FSD. This diagnostic scheme is based on four components that are not mutually exclusive: (1) *Hypoactive sexual desire*—the persistent or recurrent lack of sexual thoughts and/or receptivity to sexual activity, which causes personal distress. Hypoactive sexual desire may result from endocrine failure or may be associated with psychological or emotional disorders, (2) *Sexual arousal disorder*—the persistent or recurrent inability to attain or maintain sexual excitement, which causes personal distress, (3) *Orgasmic disorder*—the persistent or recurrent loss of orgasmic potential after sufficient sexual stimulation and arousal, which causes personal distress, and (4) *Sexual pain disorder*—persistent or recurrent genital pain associated with noncoital sexual stimulation, which causes personal distress. This newer classification emphasizes "personal distress" as a requirement for dysfunction and provides clinicians with an organized framework for evaluation before or in conjunction with more traditional counseling methods.

Treatment: Female Sexual Dysfunction

GENERAL

An open discussion with the patient is important as couples may need to be educated about normal anatomy and physiologic responses, including the role of orgasm, in sexual encounters. Physiologic changes associated with aging and/or disease should be explained. Couples may need to be reminded that clitoral stimulation rather than coital intromission may be more beneficial.

Behavioral modification and nonpharmacologic therapies should be a first step. Patient and partner counseling may improve communication and relationship strains. Lifestyle changes involving known risk factors can be an important part of the treatment process. Emphasis on maximizing physical health and avoiding lifestyles (e.g., smoking, alcohol abuse) and medications likely to produce FSD is important (Table 48-3). The use of topical lubricants may address complaints of dyspareunia and dryness. Contributing medications such as antidepressants may need to be altered, including the use of medications with less impact on sexual function, dose reduction, medication switching, or drug holidays.

HORMONAL THERAPY

In postmenopausal women, estrogen replacement therapy may be helpful in treating vaginal atrophy, decreasing coital pain, and improving clitoral sensitivity (Chap. 348). Estrogen replacement in the form

of local cream is the preferred method, as it avoids systemic side effects. Androgen levels in women decline substantially before menopause. However, low levels of testosterone or DHEA are not effective predictors of a positive therapeutic outcome with androgen therapy. The widespread use of exogenous androgens is not supported by the literature except in select circumstances (premature ovarian failure or menopausal states) and in secondary arousal disorders.

ORAL AGENTS

The efficacy of PDE-5i in FDS has been a marked disappointment in light of the proposed role of nitric oxide–dependent physiology in the normal female sexual response. The use of PDE-5i for FSD should be discouraged pending proof that it is effective.

CLITORAL VACUUM DEVICE

In patients with arousal and orgasmic difficulties, the option of using a clitoral vacuum device may be explored. This handheld battery-operated device has a small soft plastic cup that applies a vacuum over the stimulated clitoris. This causes increased cavernosal blood flow, engorgement, and vaginal lubrication.

FURTHER READINGS

Araujo AB et al: Changes in sexual function in middle-aged and older men: Longitudinal data from the Massachusetts male aging study. J Am Geriatr Soc 52:1502, 2004[PMID: 15341552]

Bhasin S et al: Sexual dysfunction in men and women with endocrine disorders. Lancet 369:597, 2007[PMID: 17307107]

Burnett AL: Erectile dysfunction. J Urol 175:S25, 2006

Cappelleri JC et al: A 5-year review of research and clinical experience. Int J Impot Res 17:307, 2005[PMID: 15875061]

Davis SR et al: Endocrine aspects of female sexual dysfunction. J Sex Med 1:82, 2004[PMID: 16422987]

Doggrell SA: Comparison of clinical trials with sildenafil, vardenafil, and tadalafil in erectile dysfunction. Expert Opin Pharmacother 6:75, 2005[PMID: 15709885]

Esposito K et al: Effect of lifestyle changes on erectile dysfunction in obese men: A randomized controlled trial. JAMA 291:2978, 2004[PMID: 15213209]

Inman BA et al. A population-based, longitudinal study of erectile dysfunction and future coronary artery disease. Mayo Clin Proc 84:108, 2009[PMID: 19181643]

Pauls RN et al: Female sexual dysfunction: Principles of diagnosis and therapy. Obstet Gynecol Surv 60:196, 2005[PMID: 16570398]

Rees PM et al: Sexual function in men and women with neurological disorders. Lancet 369(9560):512, 2007[PMID: 17292771]

Thompson IM et al: Erectile dysfunction and subsequent cardiovascular disease. JAMA 294:2996, 2005[PMID: 16414947]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HIRSUTISM AND VIRILIZATION: INTRODUCTION

Hirsutism, which is defined as androgen-dependent excessive male-pattern hair growth, affects approximately 10% of women. Hirsutism is most often idiopathic or the consequence of androgen excess associated with the polycystic ovarian syndrome (PCOS). Less frequently, it may result from adrenal androgen overproduction as occurs in nonclassic congenital adrenal hyperplasia (CAH) **(Table 49-1)**. Rarely, it is a harbinger of a serious underlying condition. Cutaneous manifestations commonly associated with hirsutism include acne and male-pattern balding (androgenic alopecia). *Virilization* refers to a condition in which androgen levels are sufficiently high to cause additional signs and symptoms, such as deepening of the voice, breast atrophy, increased muscle bulk, clitoromegaly, and increased libido; virilization is an ominous sign that suggests the possibility of an ovarian or adrenal neoplasm.

Table 49-1 Causes of Hirsutism

Gonadal hyperandrogenism
Ovarian hyperandrogenism
Polycystic ovary syndrome/functional ovarian hyperandrogenism
Ovarian steroidogenic blocks
Syndromes of extreme insulin resistance
Ovarian neoplasms
Adrenal hyperandrogenism
Premature adrenarche
Functional adrenal hyperandrogenism
Congenital adrenal hyperplasia (nonclassic and classic)
Abnormal cortisol action/metabolism
Adrenal neoplasms
Other endocrine disorders
Cushing's syndrome
Hyperprolactinemia
Acromegaly
Peripheral androgen overproduction
Obesity
Idiopathic
Pregnancy-related hyperandrogenism
Hyperreactio luteinalis
Thecoma of pregnancy
Drugs
Androgens
Oral contraceptives containing androgenic progestins
Minoxidil
Phenytoin
Diazoxide
Cyclosporine
True hermaphroditism

Hair Follicle Growth and Differentiation

Hair can be categorized as either *vellus* (fine, soft, and not pigmented) or *terminal* (long, coarse, and pigmented). The number of hair follicles does not change over an individual's lifetime, but the follicle size and type of hair can change in response to numerous factors, particularly androgens. Androgens are necessary for terminal hair and sebaceous gland development and mediate differentiation of pilosebaceous units (PSUs) into either a terminal hair follicle or a sebaceous gland. In the former case, androgens transform the vellus hair into a terminal hair; in the latter case, the sebaceous component proliferates and the hair remains vellus.

There are three phases in the cycle of hair growth: (1) *anagen* (growth phase), (2) *catagen* (involution phase), and (3) *telogen* (rest phase). Depending on the body site, hormonal regulation may play an important role in the hair growth cycle. For example, the eyebrows, eyelashes, and vellus hairs are androgen-insensitive, whereas the axillary and pubic areas are sensitive to low levels of androgens. Hair growth on the face, chest, upper abdomen, and back requires higher levels of androgens and is therefore more characteristic of the pattern typically seen in men. Androgen excess in women leads to increased hair growth in most androgen-sensitive sites except in the scalp region, where hair loss occurs because androgens cause scalp hairs to spend less time in the anagen phase.

Although androgen excess underlies most cases of hirsutism, there is only a modest correlation between androgen levels and the quantity of hair growth. This is due to the fact that hair growth from the follicle also depends on local growth factors, and there is variability in end organ (PSU) sensitivity. Genetic factors and ethnic background also influence hair growth. In general, dark-haired individuals tend to be more hirsute than blond or fair individuals. Asians and Native Americans have relatively sparse hair in regions sensitive to high androgen levels, whereas people of Mediterranean descent are more hirsute.

Clinical Assessment

Historic elements relevant to the assessment of hirsutism include the age at onset and rate of progression of hair growth and associated symptoms or signs (e.g., acne). Depending on the cause, excess hair growth typically is first noted during the second and third decades of life. The growth is usually slow but progressive. Sudden development and rapid progression of hirsutism suggest the possibility of an androgen-secreting neoplasm, in which case virilization also may be present.

The age at onset of menstrual cycles (menarche) and the pattern of the menstrual cycle should be ascertained; irregular cycles from the time of menarche onward are more likely to result from ovarian rather than adrenal androgen excess. Associated symptoms such as galactorrhea should prompt evaluation for hyperprolactinemia (Chap. 339) and possibly hypothyroidism (Chap. 341). Hypertension, striae, easy bruising, centripetal weight gain, and weakness suggest hypercortisolism (Cushing's syndrome; Chap. 342). Rarely, patients with growth hormone excess (i.e., acromegaly) present with hirsutism. Use of medications such as phenytoin, minoxidil, and cyclosporine may be associated with androgen-independent excess hair growth (i.e., hypertrichosis). A family history of infertility and/or hirsutism may indicate disorders such as nonclassic CAH (Chap. 342).

Physical examination should include measurement of height and weight and calculation of body mass index (BMI). A BMI >25 kg/m^2 is indicative of excess weight for height, and values >30 kg/m^2 are often seen in association with hirsutism, probably the result of increased conversion of androgen precursors to testosterone. Notation should be made of blood pressure, as adrenal causes may be associated with hypertension. Cutaneous signs sometimes associated with androgen excess and insulin resistance include acanthosis nigricans and skin tags.

An objective clinical assessment of hair distribution and quantity is central to the evaluation in any woman presenting with hirsutism. This assessment permits the distinction between hirsutism and hypertrichosis and provides a baseline reference point to gauge the response to treatment. A simple and commonly used method to grade hair growth is the modified scale of Ferriman and Gallwey **(Fig. 49-1)**, in which each of nine androgen-sensitive sites is graded from 0 to 4. Approximately 95% of white women have a score below 8 on this scale; thus, it is normal for most women to have some hair growth in androgen-sensitive sites. Scores above 8 suggest excess androgen-mediated hair growth, a finding that should be assessed further by means of hormonal evaluation (see below). In racial/ethnic groups that are less likely to manifest hirsutism (e.g., Asian women), additional cutaneous evidence of androgen excess should be sought, including pustular acne and thinning scalp hair.

Figure 49-1

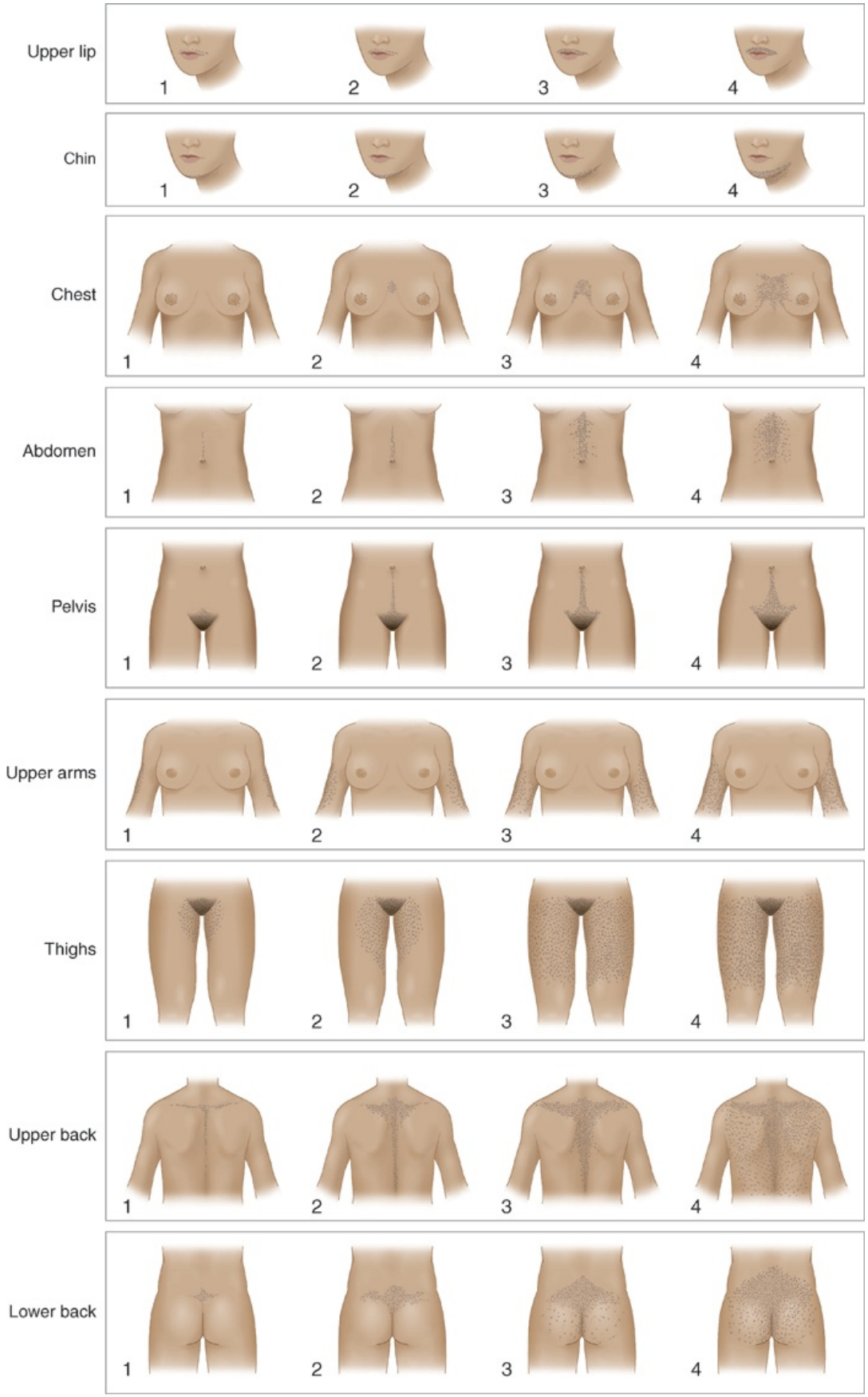

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hirsutism scoring scale of Ferriman and Gallwey. The nine body areas that have androgen-sensitive areas are graded from 0 (no terminal hair) to 4 (frankly virile) to obtain a total score. A normal hirsutism score is <8. [*Modified from DA Ehrmann et al: Hyperandrogenism, hirsutism, and polycystic ovary syndrome, in LJ DeGroot and JL Jameson (eds), Endocrinology, 5th ed. Philadelphia, Saunders, 2006; with permission.]*

Hormonal Evaluation

Androgens are secreted by the ovaries and adrenal glands in response to their respective tropic hormones: luteinizing hormone (LH) and adrenocorticotropic hormone (ACTH). The principal circulating steroids involved in the etiology of hirsutism are testosterone, androstenedione, and dehydroepiandrosterone (DHEA) and its sulfated form (DHEAS). The ovaries and adrenal glands normally contribute about equally to testosterone production. Approximately half of the total testosterone originates from direct glandular secretion, and the remainder is derived from the peripheral conversion of androstenedione and DHEA (Chap. 346).

Although it is the most important circulating androgen, testosterone is in effect the penultimate androgen in mediating hirsutism; it is converted to the more potent dihydrotestosterone (DHT) by the enzyme 5α-reductase, which is located in the PSU. DHT has a higher affinity for, and slower dissociation from, the androgen receptor. The local production of DHT allows it to serve as the primary mediator of androgen action at the level of the pilosebaceous unit. There are two isoenzymes of 5α-reductase: Type 2 is found in the prostate gland and in hair follicles, and type 1 is found primarily in sebaceous glands.

One approach to testing for hyperandrogenemia is depicted in **Fig. 49-2**. In addition to measuring blood levels of testosterone and DHEAS, it is important to measure the level of free (or unbound) testosterone. The fraction of testosterone that is not bound to its carrier protein, sex hormone–binding globulin (SHBG), is biologically available for conversion to DHT and binding to androgen receptors. Hyperinsulinemia and/or androgen excess decrease hepatic production of SHBG, resulting in levels of total testosterone within the high-normal range, whereas the unbound hormone is elevated more substantially. Although there is a decline in ovarian testosterone production after menopause, ovarian estrogen production decreases to an even greater extent, and the concentration of SHBG is reduced. Consequently, there is an increase in the relative proportion of unbound testosterone, and it may exacerbate hirsutism after menopause.

Figure 49-2

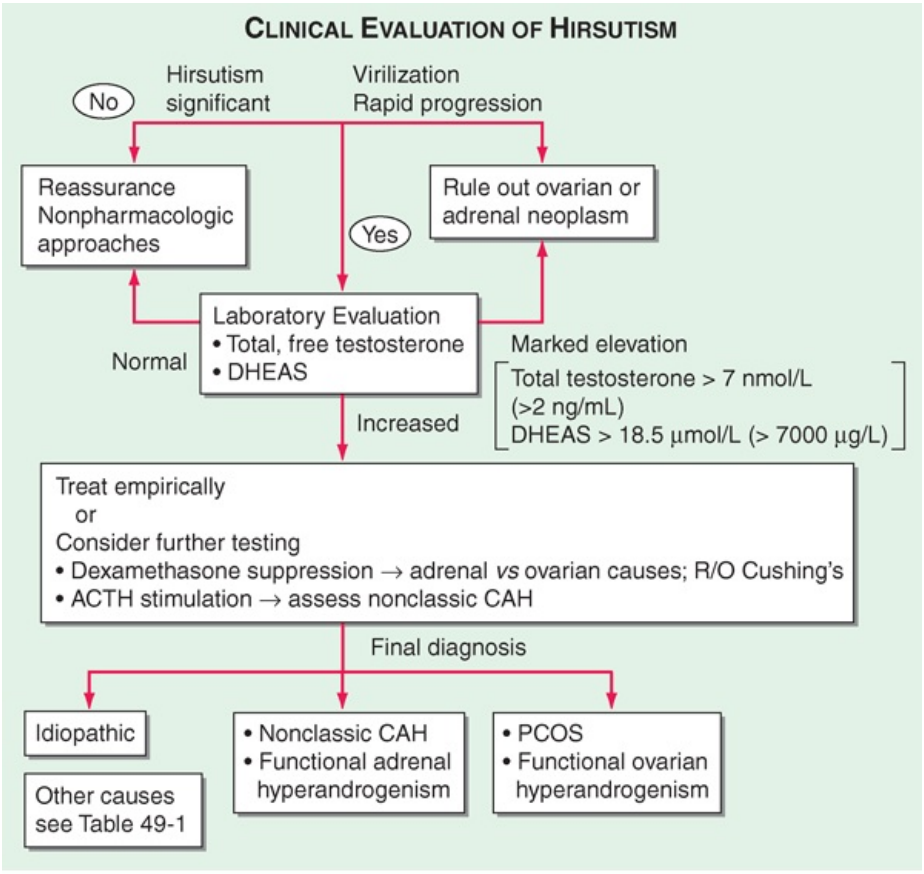

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the evaluation and differential diagnosis of hirsutism. ACTH, adrenocorticoctropic hormone; CAH, congential adrenal hyperplasia; DHEAS, sulfated form of dehydroepiandrosterone; PCOS, polycystic ovarian syndrome.

A baseline plasma total testosterone level >12 nmol/L (>3.5 ng/mL) usually indicates a virilizing tumor,

whereas a level >7 nmol/L (>2 ng/mL) is suggestive. A basal DHEAS level >18.5 μmol/L (>7000 μg/L) suggests an adrenal tumor. Although DHEAS has been proposed as a "marker" of predominant adrenal androgen excess, it is not unusual to find modest elevations in DHEAS among women with PCOS. Computed tomography (CT) or magnetic resonance imaging (MRI) should be used to localize an adrenal mass, and ultrasound usually suffices to identify an ovarian mass if clinical evaluation and hormonal levels suggest these possibilities.

PCOS is the most common cause of ovarian androgen excess (Chap. 347). An increased ratio of LH to follicle-stimulating hormone is characteristic in carefully studied patients with PCOS. However, because of the pulsatile nature of gonadotropin secretion, this finding may be absent in up to half of women with PCOS. Transvaginal ultrasound classically shows enlarged ovaries and increased stroma in women with PCOS. However, cystic ovaries also may be found in women without clinical or laboratory features of PCOS. Although usually limited to a research setting, a gonadotropin-releasing hormone agonist test can be used to make a specific diagnosis of ovarian hyperandrogenism. A peak 17-hydroxyprogesterone level ≥7.8 nmol/L (≥2.6 μg/L) after the administration of 100 μg nafarelin (or 10 μg/kg leuprolide) subcutaneously is virtually diagnostic of ovarian hyperandrogenism.

Because adrenal androgens are readily suppressed by low doses of glucocorticoids, the dexamethasone androgen-suppression test may broadly distinguish ovarian from adrenal androgen overproduction. A blood sample is obtained before and after the administration of dexa-methasone (0.5 mg orally every 6 h for 4 days). An adrenal source is suggested by suppression of unbound testosterone into the normal range; incomplete suppression suggests ovarian androgen excess. An overnight 1-mg dexamethasone suppression test, with measurement of 8:00 a.m. serum cortisol, is useful when there is clinical suspicion of Cushing's syndrome (Chap. 342).

Nonclassic CAH is most commonly due to 21-hydroxylase deficiency but also can be caused by autosomal recessive defects in other steroidogenic enzymes necessary for adrenal corticosteroid synthesis (Chap. 342). Because of the enzyme defect, the adrenal gland cannot secrete glucocorticoids (especially cortisol) efficiently. This results in diminished negative feedback inhibition of ACTH, leading to compensatory adrenal hyperplasia and the accumulation of steroid precursors that subsequently are converted to androgen. Deficiency of 21-hydroxylase can be reliably excluded by determining a morning 17-hydroxyprogesterone level <6 nmol/L (<2 μg/L) (drawn in the follicular phase). Alternatively, 21-hydroxylase deficiency can be diagnosed by measurement of 17-hydroxyprogesterone 1 h after the administration of 250 μg of synthetic ACTH (cosyntropin) intravenously.

Treatment: Hirsutism

Treatment of hirsutism may be accomplished pharmacologically or by mechanical means of hair removal. Nonpharmacologic treatments should be considered in all patients either as the only treatment or as an adjunct to drug therapy.

Nonpharmacologic treatments include (1) bleaching, (2) depilatory (removal from the skin surface) such as shaving and chemical treatments, and (3) epilatory (removal of the hair including the root) such as plucking, waxing, electrolysis, and laser therapy. Despite perceptions to the contrary, shaving does not increase the rate or density of hair growth. Chemical depilatory treatments may be useful for mild hirsutism that affects only limited skin areas, though they can cause skin irritation. Wax treatment removes hair temporarily but is uncomfortable. Electrolysis is effective for more permanent hair removal, particularly in the hands of a skilled electrologist. Laser phototherapy appears to be efficacious for hair removal. It delays hair regrowth and causes permanent hair removal in most patients. The long-term effects and complications associated with laser treatment are being evaluated.

Pharmacologic therapy is directed at interrupting one or more of the steps in the pathway of androgen synthesis and action: (1) suppression of adrenal and/or ovarian androgen production, (2) enhancement of androgen-binding to plasma-binding proteins, particularly SHBG, (3) impairment of the peripheral conversion of androgen precursors to active androgen, and (4) inhibition of androgen action at the target tissue level. Attenuation of hair growth is typically not evident until 4–6 months after initiation of medical treatment and in most cases leads to only a modest reduction in hair growth.

Combination estrogen-progestin therapy in the form of an oral contraceptive is usually the first-line endocrine treatment for hirsutism and acne, after cosmetic and dermatologic management. The estrogenic component of most oral contraceptives currently in use is either ethinyl estradiol or mestranol. The suppression of LH leads to reduced production of ovarian androgens. The reduced androgen levels also result in a dose-related increase in SHBG, thus lowering the fraction of unbound plasma testosterone. Combination therapy also has been demonstrated to decrease DHEAS, perhaps by reducing ACTH levels. Estrogens also have a direct, dose-dependent suppressive effect on sebaceous cell function.

The choice of a specific oral contraceptive should be predicated on the progestational component, as progestins vary in their suppressive effect on SHBG levels and in their androgenic potential. Ethynodiol diacetate has relatively low androgenic potential, whereas progestins such as norgestrel and levonor-gestrel are particularly androgenic, as judged from their attenuation of the estrogen-

induced increase in SHBG. Norgestimate exemplifies the newer generation of progestins that are virtually nonandrogenic. Drospirenone, an analogue of spironolactone that has both antimineralocorticoid and antiandrogenic activities, has been approved for use as a progestational agent in combination with ethinyl estradiol.

Oral contraceptives are contraindicated in women with a history of thromboembolic disease and women with increased risk of breast or other estrogen-dependent cancers (Chap. 348). There is a relative contraindication to the use of oral contraceptives in smokers and those with hypertension or a history of migraine headaches. In most trials, estrogen-progestin therapy alone improves the extent of acne by a maximum of 50–70%. The effect on hair growth may not be evident for 6 months, and the maximum effect may require 9–12 months owing to the length of the hair growth cycle. Improvements in hirsutism are typically in the range of 20%, but there may be an arrest of further progression of hair growth.

Adrenal androgens are more sensitive than cortisol to the suppressive effects of glucocorticoids. Therefore, glucocorticoids are the mainstay of treatment in patients with CAH. Although glucocorticoids have been reported to restore ovulatory function in some women with PCOS, this effect is highly variable. Because of side effects from excessive glucocorticoids, low doses should be used. Dexamethasone (0.2–0.5 mg) or prednisone (5–10 mg) should be taken at bedtime to achieve maximal suppression by inhibiting the nocturnal surge of ACTH.

Cyproterone acetate is the prototypic antiandrogen. It acts mainly by competitive inhibition of the binding of testosterone and DHT to the androgen receptor. In addition, it may enhance the metabolic clearance of testosterone by inducing hepatic enzymes. Although not available for use in the United States, cyproterone acetate is widely used in Canada, Mexico, and Europe. Cyproterone (50–100 mg) is given on days 1–15 and ethinyl estradiol (50 μg) is given on days 5–26 of the menstrual cycle. Side effects include irregular uterine bleeding, nausea, headache, fatigue, weight gain, and decreased libido.

Spironolactone, which usually is used as a mineralocorticoid antagonist, is also a weak antiandrogen. It is almost as effective as cyproterone acetate when used at high enough doses (100–200 mg daily). Patients should be monitored intermittently for hyperkalemia or hypotension, though these side effects are uncommon. Pregnancy should be avoided because of the risk of feminization of a male fetus. Spironolactone can also cause menstrual irregularity. It often is used in combination with an oral contraceptive, which suppresses ovarian androgen production and helps prevent pregnancy.

Flutamide is a potent nonsteroidal antiandrogen that is effective in treating hirsutism, but concerns about the induction of hepatocellular dysfunction have limited its use. Finasteride is a competitive inhibitor of 5α-reductase type 2. Beneficial effects on hirsutism have been reported, but the predominance of 5α-reductase type 1 in the PSU appears to account for its limited efficacy. Finasteride would also be expected to impair sexual differentiation in a male fetus, and it should not be used in women who may become pregnant.

Eflornithine cream (Vaniqa) has been approved as a novel treatment for unwanted facial hair in women, but long-term efficacy remains to be established. It can cause skin irritation under exaggerated conditions of use. Ultimately, the choice of any specific agent(s) must be tailored to the unique needs of the patient being treated. As noted previously, pharmacologic treatments for hirsutism should be used in conjunction with nonpharmacologic approaches. It is also helpful to review the pattern of female hair distribution in the normal population to dispel unrealistic expectations.

FURTHER READINGS

Ehrmann DA: Polycystic ovary syndrome. N Engl J Med 352:1223, 2005[PMID: 15788499]

Martin KA et al: Evaluation and treatment of hirsutism in premenopausal women: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 93:1105, 2008[PMID: 18252793]

Pall M et al: The phenotype of hirsute women: A comparison of polycystic ovary syndrome and 21-hydroxylase–deficient nonclassic adrenal hyperplasia. Fertil Steril 94:684, 2010[PMID: 19726039]

Rathnayake D, Sinclair: Innovative use of spironolactone as an antiandrogen in the treatment of female pattern hair loss. Dermatol Clin 28:611, 2010[PMID: 20510769]

Swiglo BA et al: Antiandrogens for the treatment of hirsutism: A systematic review and metaanalyses of randomized controlled trials. J Clin Endocrinol Metab 93:1153, 2008[PMID: 18252786]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MENSTRUAL DISORDERS AND PELVIC PAIN: INTRODUCTION

Menstrual dysfunction can signal an underlying abnormality that may have long-term health consequences. Although frequent or prolonged bleeding usually prompts a woman to seek medical attention, infrequent or absent bleeding may seem less troubling and the patient may not bring it to the attention of the physician. Thus, a focused menstrual history is a critical part of every encounter with a female patient. Pelvic pain is a common complaint that may relate to an abnormality of the reproductive organs but also may be of gastrointestinal, urinary tract, or musculoskeletal origin. Depending on its cause, pelvic pain may require urgent surgical attention.

MENSTRUAL DISORDERS

Definition and Prevalence

Amenorrhea refers to the absence of menstrual periods. Amenorrhea is classified as *primary* if menstrual bleeding has never occurred in the absence of hormonal treatment or *secondary* if menstrual periods are absent for 3–6 months. Primary amenorrhea is a rare disorder that occurs in <1% of the female population. However, between 3 and 5% of women experience at least 3 months of secondary amenorrhea in any specific year. There is no evidence that race or ethnicity influences the prevalence of amenorrhea. However, because of the importance of adequate nutrition for normal reproductive function, both the age at menarche and the prevalence of secondary amenorrhea vary significantly in different parts of the world.

Oligoamenorrhea is defined as a cycle length >35 days or <10 menses per year. Both the frequency and the amount of vaginal bleeding are irregular in oligoamenorrhea. It is often associated with anovulation, which also can occur with intermenstrual intervals <24 days or vaginal bleeding for >7 days. Frequent or heavy irregular bleeding is termed *dysfunctional uterine bleeding* if anatomic uterine lesions or a bleeding diathesis has been excluded.

PRIMARY AMENORRHEA

The absence of menses by age 16 has been used traditionally to define primary amenorrhea. However, other factors, such as growth, secondary sexual characteristics, the presence of cyclic pelvic pain, and the secular trend toward an earlier age of menarche, particularly in African-American girls, also influence the age at which primary amenorrhea should be investigated. Thus, an evaluation for amenorrhea should be initiated by age 15 or 16 in the presence of normal growth and secondary sexual characteristics; age 13 in the absence of secondary sexual characteristics or if height is less than the third percentile; age 12 or 13 in the presence of breast development and cyclic pelvic pain; or within 2 years of breast development if menarche, defined by the first menstrual period, has not occurred.

SECONDARY AMENORRHEA OR OLIGOAMENORRHEA

Anovulation and irregular cycles are relatively common for up to 2 years after menarche and for 1–2 years before the final menstrual period. In the intervening years, menstrual cycle length is ~28 days, with an intermenstrual interval normally ranging between 25 and 35 days. Cycle-to-cycle variability in an individual woman who is ovulating consistently is generally +/– 2 days. Pregnancy is the most common cause of amenorrhea and should be excluded early in any evaluation of menstrual irregularity. However, many women occasionally miss a single period. Three or more months of secondary amenorrhea should prompt an evaluation, as should a history of intermenstrual intervals >35 or <21 days or bleeding that persists for >7 days.

Diagnosis

Evaluation of menstrual dysfunction depends on understanding the interrelationships between the four critical components of the reproductive tract: (1) the hypothalamus, (2) the pituitary, (3) the ovaries, and (4) the uterus and outflow tract (**Fig. 50-1**; Chap. 347). This system is maintained by complex negative and positive feedback loops involving the ovarian steroids (estradiol and progesterone) and peptides (inhibin B and inhibin A) and the hypothalamic [gonadotropin-releasing hormone (GnRH)] and pituitary [follicle-stimulating hormone (FSH) and luteinizing hormone (LH)] components of this system (Fig. 50-1).

Figure 50-1

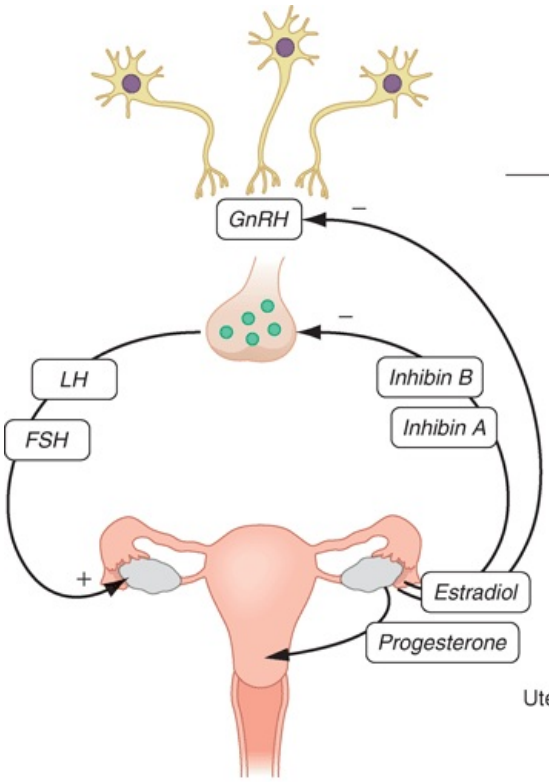

	Primary	Secondary
Hypothalamus	27%	36%
Pituitary	2%	15%
PCOS	7%	30%
Ovary	43%	12%
Uterus/outflow tract	19%	7%

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Role of the hypothalamic-pituitary-gonadal axis in the etiology of amenorrhea. Gonadotropin-releasing hormone (GnRH) secretion from the hypothalamus stimulates follicle-stimulating hormone (FSH) and luteinizing hormone (LH) secretion from the pituitary to induce ovarian folliculogenesis and steroidogenesis. Ovarian secretion of estradiol and progesterone controls the shedding of the endometrium, resulting in menses, and, in combination with the inhibins, provides feedback regulation of the hypothalamus and pituitary to control secretion of FSH and LH. The prevalence of amenorrhea resulting from abnormalities at each level of the reproductive system (hypothalamus, pituitary, ovary, uterus and outflow tract) varies depending on whether amenorrhea is primary or secondary. PCOS, polycystic ovarian syndrome.

Disorders of menstrual function can be thought of in two main categories: disorders of the uterus and outflow tract and disorders of ovulation. Many of the conditions that cause primary amenorrhea are congenital but go unrecognized until the time of normal puberty (e.g., genetic, chromosomal, and anatomic abnormalities). All causes of secondary amenorrhea also can cause primary amenorrhea.

DISORDERS OF THE UTERUS OR OUTFLOW TRACT

Abnormalities of the uterus and outflow tract typically present as primary amenorrhea. In patients with normal pubertal development and a blind vagina, the differential diagnosis includes *obstruction* by a transverse vaginal septum or imperforate hymen; *müllerian agenesis* (Mayer-Rokitansky-Kuster-Hauser syndrome), which has been associated with mutations in the *WNT4* gene; and *androgen insensitivity syndrome* (AIS), which is an X-linked recessive disorder that accounts for ~10% of all cases of primary amenorrhea (Chap. 346). Patients with AIS have a 46,XY karyotype, but because of the lack of androgen receptor responsiveness, they have severe underandrogenization and female external genitalia. The absence of pubic and axillary hair distinguishes them clinically from patients with müllerian agenesis. *Asherman syndrome* presents as secondary amenorrhea or hypomenorrhea and results from partial or complete obliteration of the uterine cavity by adhesions that prevent normal growth and shedding of the endometrium. Curettage performed for pregnancy complications accounts for >90% of cases; genital tuberculosis is an important cause in regions where it is endemic.

Treatment: Disorders of the Uterus or Outflow Tract

Obstruction of the outflow tract requires surgical correction. The risk of endometriosis is increased with this condition, perhaps because of retrograde menstrual flow. *Müllerian agenesis* also may require surgical intervention, although vaginal dilatation is adequate in some patients. Because ovarian function is normal, assisted reproductive techniques can be used with a surrogate carrier. *Androgen resistance syndrome* requires gonadectomy because there is risk of gonadoblastoma in the dysgenetic gonads. Whether this should be performed in early childhood or after completion of breast development is controversial. Estrogen replacement is indicated after gonadectomy, and vaginal dilatation may be required to allow sexual intercourse.

DISORDERS OF OVULATION

Once uterus and outflow tract abnormalities have been excluded, other causes of amenorrhea involve disorders of ovulation. The differential diagnosis is based on the results of initial tests, including a pregnancy test, gonadotropins (to determine whether the cause is likely to be ovarian or central), and assessment of hyperandrogenism (**Fig. 50-2**).

Figure 50-2

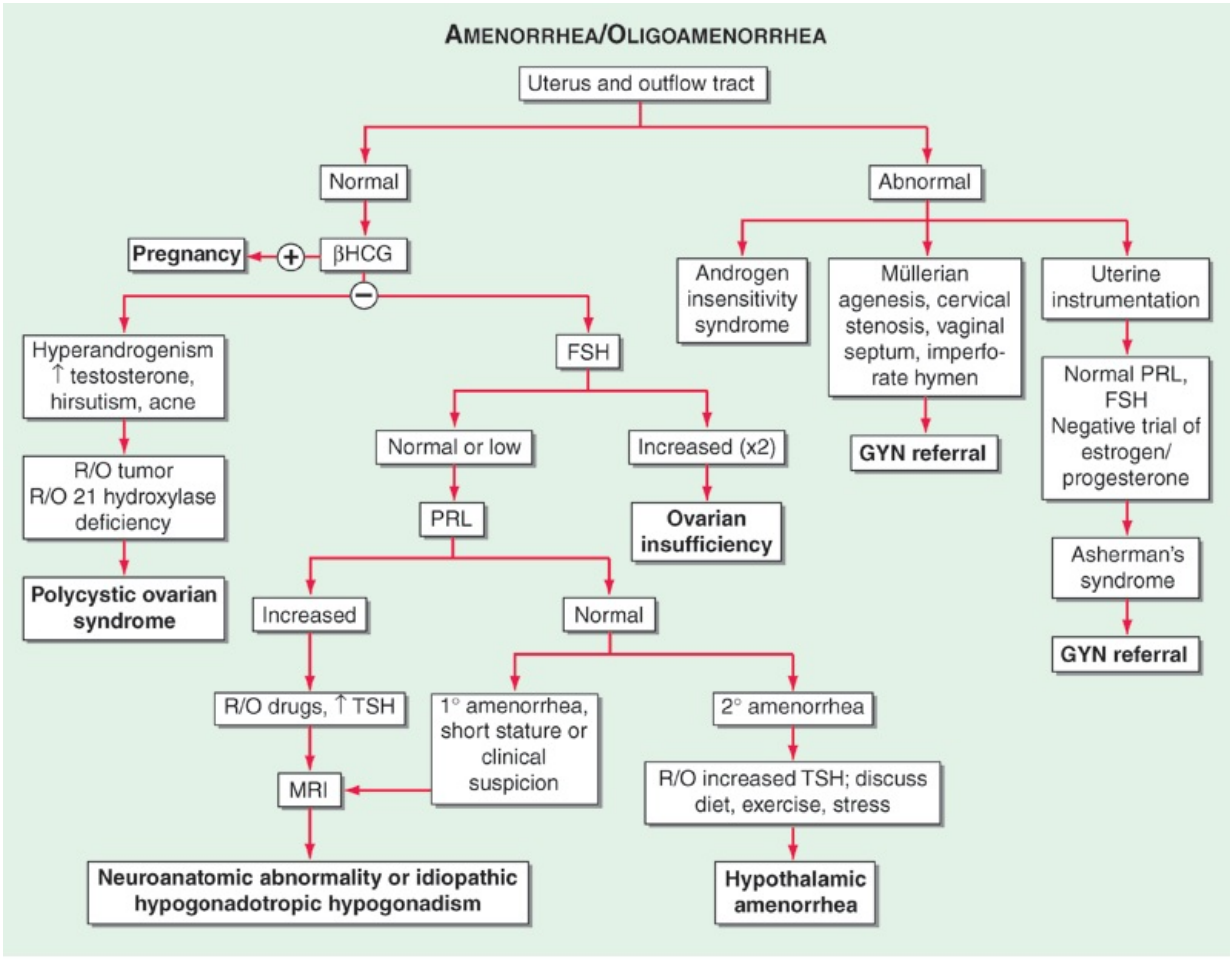

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for evaluation of amenorrhea. ß-hCG, human chorionic gonadotropin; FSH, follicle-stimulating hormone; PRL, prolactin; TSH, thyroid-stimulating hormone.

Hypogonadotropic Hypogonadism

Low estrogen levels in combination with normal or low levels of LH and FSH are seen with anatomic, genetic, or functional abnormalities that interfere with hypothalamic GnRH secretion or normal pituitary responsiveness to GnRH. Although relatively uncommon, tumors and infiltrative diseases should be considered in the differential diagnosis of hypogonadotropic hypogonadism (Chap. 339). These disorders may present with primary or secondary amenorrhea. They may occur in association with other features suggestive of hypothalamic or pituitary dysfunction, such as short stature, diabetes insipidus, galactorrhea, and headache. Hypogonadotropic hypogonadism also may be seen after cranial irradiation. In the postpartum period, it may be caused by pituitary necrosis (Sheehan's syndrome) or lymphocytic hypophysitis. Because reproductive dysfunction is commonly associated with hyperprolactinemia from neuroanatomic lesions or medications, prolactin should be measured in all patients with hypogonadotropic hypogonadism (Chap. 339).

Isolated hypogonadotropic hypogonadism (IHH) occurs in women, although it is more common in men. IHH generally presents with primary amenorrhea and is associated with anosmia in about 50% of women (termed Kallmann syndrome). Genetic causes of IHH have been identified in approximately 35% of patients (Chaps. 346 and 347).

Functional hypothalamic amenorrhea (HA) is caused by a mismatch between energy expenditure and energy intake. Recent studies suggest that variants in genes associated with IHH may increase susceptibility to these environmental inputs, accounting in part for the clinical variability in this disorder.

Leptin secretion may play a key role in transducing the signals from the periphery to the hypothalamus in HA. The hypothalamic-pituitary-adrenal axis also may play a role. The diagnosis of HA generally can be made on the basis of a careful history, a physical examination, and the demonstration of low levels of gonadotropins and normal prolactin levels. Eating disorders and chronic disease must be specifically excluded (Chap. 79). An atypical history, headache, signs of other hypothalamic dysfunction, or hyperprolactinemia, even if mild, necessitates cranial imaging with CT or MRI to exclude a neuroanatomic cause.

Hypergonadotropic Hypogonadism

Ovarian failure is considered premature when it occurs in women <40 years old and accounts for ~10% of secondary amenorrhea. *Primary ovarian insufficiency* (POI) has generally replaced the terms *premature menopause* and *premature ovarian failure* in recognition that this disorder represents a continuum of impaired ovarian function. Ovarian insufficiency is associated with the loss of negative-feedback restraint on the hypothalamus and pituitary, resulting in increased FSH and LH levels. FSH is a better marker of ovarian failure as its levels are less variable than those of LH. As with natural menopause, POI may wax and wane, and serial measurements may be necessary to establish the diagnosis.

Once the diagnosis of POI has been established, further evaluation is indicated because of other health problems that may be associated with POI. For example, POI occurs in association with a variety of chromosomal abnormalities, including Turner syndrome, autoimmune polyglandular failure syndromes, radio- and chemotherapy, and galactosemia. The recognition that early ovarian failure occurs in premutation carriers of the fragile X syndrome is important because of the increased risk of severe mental retardation in male children with *FMR1* mutations. In the majority of cases, however, a cause for POI is not determined.

Hypergonadotropic hypogonadism occurs rarely in other disorders, such as mutations in the FSH or LH receptors. Aromatase deficiency and 17α-hydroxylase deficiency are associated with elevated gonadotropins with hyperandrogenism and hypertension, respectively. Gonadotropin-secreting tumors in women of reproductive age generally present with high, rather than low, estrogen levels and cause ovarian hyperstimulation or dysfunctional bleeding.

Treatment: Hypo- and Hypergonadotropic Causes of Amenorrhea

Amenorrhea almost always is associated with chronically low levels of estrogen, whether it is caused by hypogonadotropic hypogonadism or ovarian insufficiency. Development of secondary sexual characteristics requires gradual titration of estradiol replacement with eventual addition of progestin. Symptoms of hypoestrogenism can be treated with hormone replacement therapy or oral contraceptive pills. Patients with hypogonadotropic hypogonadism who are interested in fertility require treatment with pulsatile GnRH or exogenous FSH and LH, whereas patients with ovarian failure can consider oocyte donation, which has a high chance of success in this population.

Polycystic Ovarian Syndrome (PCOS)

PCOS is diagnosed on the basis of a combination of clinical or biochemical evidence of hyperandrogenism, amenorrhea or oligomenorrhea, and the ultrasound appearance of polycystic ovaries. Approximately half of patients with PCOS are obese, and abnormalities in insulin dynamics are common, as is metabolic syndrome. Symptoms generally begin shortly after menarche and are slowly progressive. Lean patients with PCOS generally have high LH levels in the presence of normal to low levels of FSH and estradiol. The LH/FSH ratio is less pronounced in obese patients in whom insulin resistance is a more prominent feature.

Treatment: Polycystic Ovarian Syndrome

A major abnormality in patients with PCOS is the failure of regular, predictable ovulation. Thus, these patients are at risk for the development of dysfunctional bleeding and endometrial hyperplasia associated with unopposed estrogen exposure. Endometrial protection can be achieved with the use of oral contraceptives or progestins (medroxyprogesterone acetate, 5–10 mg, or prometrium, 200 mg daily for 10–14 days of each month). Oral contraceptives are also useful for management of hyperandrogenic symptoms, as is spironolactone, which functions as a weak androgen receptor antagonist. Management of the associated metabolic syndrome may be appropriate for some patients (Chap. 242). For patients interested in fertility, weight control is a critical first step. Clomiphene citrate is highly effective as a first-line treatment, with or without the addition of metformin. Exogenous gonadotropins can be used by experienced practitioners.

PELVIC PAIN

The mechanisms that cause pelvic pain are similar to those which cause abdominal pain (Chap. 13) and include inflammation of the parietal peritoneum, obstruction of hollow viscera, vascular disturbances, and pain originating in the abdominal wall. Pelvic pain may reflect pelvic disease per se but also may reflect extrapelvic disorders that refer pain to the pelvis. In up to 60% of cases, pelvic pain can be attributed to gastrointestinal problems, including appendicitis, cholecystitis, infections,

intestinal obstruction, diverticulitis, and inflammatory bowel disease. Urinary tract and musculoskeletal disorders are also common causes of pelvic pain.

Approach to the Patient: Pelvic Pain

A thorough history that includes the type, location, radiation, and status with respect to increasing or decreasing severity can help identify the cause of acute pelvic pain. Specific associations with vaginal bleeding, sexual activity, defecation, urination, movement, or eating should be specifically sought. A careful menstrual history is essential to assess the possibility of pregnancy. Determination of whether the pain is acute versus chronic and cyclic versus noncyclic will direct further investigation **(Table 50-1)**. However, disorders that cause cyclic pain occasionally may cause noncyclic pain, and the converse is also true.

Table 50-1 Causes of Pelvic Pain

	Acute	Chronic
Cyclic pelvic pain		Premenstrual symptoms Mittelschmerz
		Dysmenorrhea
Noncyclic pelvic pain	Pelvic inflammatory disease	Endometriosis Pelvic congestion syndrome
	Ruptured or hemorrhagic ovarian cyst or ovarian torsion	Adhesions and retroversion of the uterus
	Ectopic pregnancy	Pelvic malignancy
	Endometritis Acute growth or degeneration of uterine myoma	Vulvodynia History of sexual abuse

Acute Pelvic Pain

Pelvic inflammatory disease most commonly presents with bilateral lower abdominal pain. It is generally of recent onset and is exacerbated by intercourse or jarring movements. Fever is present in about half of these patients; abnormal uterine bleeding occurs in about one-third. New vaginal discharge, urethritis, and chills may be present but are less specific signs. *Adnexal pathology* can present acutely and may be due to rupture, bleeding or torsion of cysts, or, much less commonly, neoplasms of the ovary, fallopian tubes, or paraovarian areas. Fever may be present with ovarian torsion. *Ectopic pregnancy* is associated with right- or left-sided lower abdominal pain and vaginal bleeding, with clinical signs generally appearing 6–8 weeks after the last normal menstrual period. Orthostatic signs and fever may be present. Risk factors include the presence of known tubal disease, previous ectopic pregnancies, a history of infertility, diethylstilbestrol (DES) exposure of the mother in utero, or a history of pelvic infections. *Uterine pathology* includes endometritis and, less frequently, degenerating leiomyomas (fibroids). Endometritis often is associated with vaginal bleeding and systemic signs of infection. It occurs in the setting of sexually transmitted infections, uterine instrumentation, or postpartum infection.

A sensitive pregnancy test, complete blood count with differential, urinalysis, tests for chlamydial and gonococcal infections, and abdominal ultrasound aid in making the diagnosis and directing further management.

Treatment: Acute Pelvic Pain

Treatment of acute pelvic pain depends on the suspected etiology but may require surgical or gynecologic intervention. Conservative management is an important consideration for ovarian cysts, if torsion is not suspected, to avoid unnecessary pelvic surgery and the subsequent risk of infertility due to adhesions. The majority of unruptured ectopic pregnancies are now treated with methotrexate, which is effective in 84–96% of cases. However, surgical treatment may be required.

Chronic Pelvic Pain

Some women experience discomfort at the time of ovulation (*mittelschmerz*). The pain can be quite intense but is generally of short duration. The mechanism is thought to involve rapid expansion of the dominant follicle, although it also may be caused by peritoneal irritation by follicular fluid released at the time of ovulation. Many women experience premenstrual symptoms such as breast discomfort, food cravings, and abdominal bloating or discomfort. These moliminal symptoms are a good predictor

of ovulation, although their absence is less helpful.

DYSMENORRHEA

Dysmenorrhea refers to the crampy lower abdominal discomfort that begins with the onset of menstrual bleeding and gradually decreases over the next 12–72 h. It may be associated with nausea, diarrhea, fatigue, and headache and occurs in 60–93% of adolescents, beginning with the establishment of regular ovulatory cycles. Its prevalence decreases after pregnancy and with the use of oral contraceptives.

Primary dysmenorrhea results from increased stores of prostaglandin precursors, which are generated by sequential stimulation of the uterus by estrogen and progesterone. During menstruation these precursors are converted to prostaglandins, which cause intense uterine contractions, decreased blood flow, and increased peripheral nerve hypersensitivity, resulting in pain.

Secondary dysmenorrhea is caused by underlying pelvic pathology. *Endometriosis* results from the presence of endometrial glands and stroma outside the uterus. These deposits of ectopic endometrium respond to hormonal stimulation and cause dysmenorrhea, which generally precedes menstruation by several days. Endometriosis also may be associated with painful intercourse, painful bowel movements, and tender nodules in the uterosacral ligament. Fibrosis and adhesions can produce lateral displacement of the cervix. The CA125 level may be increased, but it has low negative predictive value. Definitive diagnosis requires laparoscopy. Symptomatology does not always predict the extent of endometriosis. Other secondary causes of dysmenorrhea include adenomyosis, a condition caused by the presence of ectopic endometrial glands and stroma within the myometrium. Cervical stenosis may result from trauma, infection, or surgery.

Treatment: Dysmenorrhea

Local application of heat; use of vitamins B_1, B_6, and E and magnesium; acupuncture; yoga; and exercise are of some benefit for the treatment of dysmenorrhea. However, nonsteroidal anti-inflammatory drugs (NSAIDs) are the most effective treatment and provide >80% sustained response rates. Ibuprofen, naproxen, ketoprofen, mefanamic acid, and nimesulide are all superior to placebo. Treatment should be started a day before expected menses and generally is continued for 2–3 days. Oral contraceptives also reduce symptoms of dysmenorrhea. Failure of response to NSAIDs and oral contraceptives is suggestive of a pelvic disorder such as endometriosis, and diagnostic laparoscopy should be considered to guide further treatment.

FURTHER READINGS

Gordon CM: Clinical practice: Functional hypothalamic amenorrhea. N Engl J Med 33:365, 2010

Hall JE: Neuroendocrine control of the menstrual cycle, in *Yen and Jaffe's Reproductive Endocrinology*, 6th ed, JF Strauss III, RL Barbieri (eds). Philadelphia, Elsevier, 2009, pp 139–154

Howard FM: Endometriosis and mechanisms of pelvic pain. J Minim Invasive Gynecol 16:540, 2009[PMID: 19835795]

Nelson LM: Clinical practice: Primary ovarian insufficiency. N Engl J Med 360:606, 2009[PMID: 19196677]

Pallais JC et al: Kallmann syndrome, in *GeneReviews,* RA Pagon et al (eds). Seattle, University of Washington, 1993–2007 [updated April 8, 2010]. http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene

Sultan C et al: Mayer-Rokitansky-Kuster-Hauser syndrome: Recent clinical and genetic findings. Gynecol Endocrinol 25:88, 2009

Wittenberger MD et al: The FMR1 premutation and reproduction. Fertil Steril 87:456, 2007[PMID: 17074338]

Zahradnik HP et al: Nonsteroidal anti-inflammatory drugs and hormonal contraceptives for pain relief from dysmenorrhea: A review. Contraception 81:185, 2010[PMID: 20159173]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH A SKIN DISORDER: INTRODUCTION

The challenge of examining the skin lies in distinguishing normal from abnormal, significant findings from trivial ones, and in integrating pertinent signs and symptoms into an appropriate differential diagnosis. The fact that the largest organ in the body is visible is both an advantage and a disadvantage to those who examine it. It is advantageous because no special instrumentation is necessary and because the skin can be biopsied with little morbidity. However, the casual observer can be misled by a variety of stimuli and overlook important, subtle signs of skin or systemic disease. For instance, the sometimes minor differences in color and shape that distinguish a melanoma **(Fig. 51-1)** from a benign nevomelanocytic nevus **(Fig. 51-2)** can be difficult to recognize. To aid in the interpretation of skin lesions, a variety of descriptive terms have been developed to characterize cutaneous lesions **(Tables 51-1**, **51-2**, and **51-3** as well as **Fig. 51-3)** and to formulate a differential diagnosis **(Table 51-4)**. For instance, the finding of scaling papules (present in patients with psoriasis or atopic dermatitis) places the patient in a different diagnostic category than would hemorrhagic papules, which may indicate vasculitis or sepsis (**Figs. 51-4** and **51-5**, respectively). It is also important to differentiate primary from secondary skin lesions. If the examiner focuses on linear erosions overlying an area of erythema and scaling, he or she may incorrectly assume that the erosion is the primary lesion and the redness and scale are secondary, while the correct interpretation would be that the patient has a pruritic eczematous dermatitis with erosions caused by scratching.

Figure 51-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Superficial spreading melanoma. This is the most common type of melanoma. Such lesions usually demonstrate asymmetry, border irregularity, color variegation (black, blue, brown, pink, and white), a diameter >6 mm, and a history of change (e.g., an increase in size or development of associated symptoms such as pruritus or pain).

Figure 51-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nevomelanocytic nevus. Nevi are benign proliferations of nevomelanocytes characterized by regularly shaped hyperpigmented macules or papules of a uniform color.

Figure 51-3

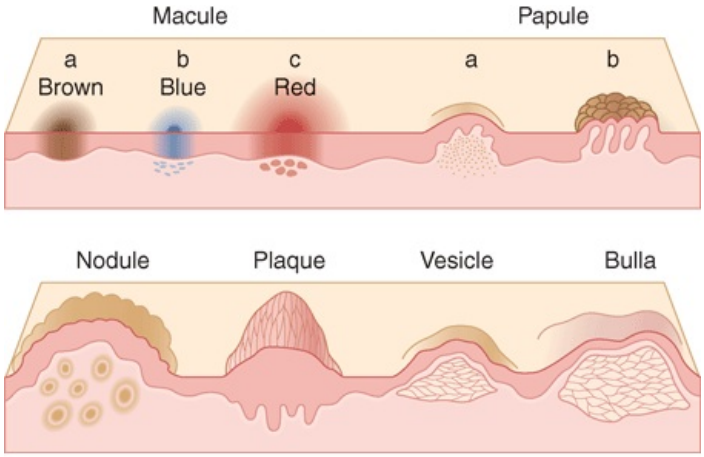

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A schematic representation of several common primary skin lesions (see Table 51-1).

Figure 51-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Necrotizing vasculitis. Palpable purpuric papules on the lower legs are seen in this patient with cutaneous small vessel vasculitis. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure 51-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Meningococcemia. An example of fulminant meningococcemia with extensive angular purpuric patches. *(Courtesy of Stephen E. Gellis, MD; with permission.)*

Table 51-1 Description of Primary Skin Lesions

Macule: A flat, colored lesion, <2 cm in diameter, not raised above the surface of the surrounding skin. A "freckle," or ephelid, is a prototype pigmented macule.
Patch: A large (>2 cm) flat lesion with a color different from the surrounding skin. This differs from a macule only in size.
Papule: A small, solid lesion, <0.5 cm in diameter, raised above the surface of the surrounding skin and, hence, palpable (e.g., a closed comedone, or whitehead, in acne).
Nodule: A larger (0.5–5.0 cm), firm lesion raised above the surface of the surrounding skin. This differs from a papule only in size (e.g., a dermal nevomelanocytic nevus).
Tumor: A solid, raised growth >5 cm in diameter.

Plaque: A large (>1 cm), flat-topped, raised lesion; edges may either be distinct (e.g., in psoriasis) or gradually blend with surrounding skin (e.g., in eczematous dermatitis).
Vesicle: A small, fluid-filled lesion, <0.5 cm in diameter, raised above the plane of surrounding skin. Fluid is often visible, and the lesions are translucent [e.g., vesicles in allergic contact dermatitis caused by *Toxicodendron* (poison ivy)].
Pustule: A vesicle filled with leukocytes. Note: The presence of pustules does not necessarily signify the existence of an infection.
Bulla: A fluid-filled, raised, often translucent lesion >0.5 cm in diameter.
Wheal: A raised, erythematous, edematous papule or plaque, usually representing short-lived vasodilatation and vasopermeability.
Telangiectasia: A dilated, superficial blood vessel.

Table 51-2 Description of Secondary Skin Lesions

Lichenification: A distinctive thickening of the skin that is characterized by accentuated skin-fold markings.
Scale: Excessive accumulation of stratum corneum.
Crust: Dried exudate of body fluids that may be either yellow (i.e., serous crust) or red (i.e., hemorrhagic crust).
Erosion: Loss of epidermis without an associated loss of dermis.
Ulcer: Loss of epidermis and at least a portion of the underlying dermis.
Excoriation: Linear, angular erosions that may be covered by crust and are caused by scratching.
Atrophy: An acquired loss of substance. In the skin, this may appear as a depression with intact epidermis (i.e., loss of dermal or subcutaneous tissue) or as sites of shiny, delicate, wrinkled lesions (i.e., epidermal atrophy).
Scar: A change in the skin secondary to trauma or inflammation. Sites may be erythematous, hypopigmented, or hyperpigmented depending on their age or character. Sites on hair-bearing areas may be characterized by destruction of hair follicles.

Table 51-3 Common Dermatologic Terms

Alopecia: Hair loss; it may be partial or complete.
Annular: Ring-shaped lesions.
Cyst: A soft, raised, encapsulated lesion filled with semisolid or liquid contents.
Herpetiform: Grouped lesions.
Lichenoid: Violaceous to purple, polygonal lesions that resemble those seen in lichen planus.
Milia: Small, firm, white papules filled with keratin.
Morbilliform: Generalized, small erythematous macules and/or papules that resemble lesions seen in measles.
Nummular: Coin-shaped lesions.
Poikiloderma: Skin that displays variegated pigmentation, atrophy, and telangiectases.
Polycyclic: A configuration of skin lesions formed from coalescing rings or incomplete rings.
Pruritus: A sensation that elicits the desire to scratch. Pruritus is often the predominant symptom of inflammatory skin diseases (e.g., atopic dermatitis, allergic contact dermatitis); it is also commonly associated with xerosis and aged skin. Systemic conditions that can be associated with pruritus include chronic renal disease, cholestasis, pregnancy, malignancy, thyroid disease, polycythemia vera, and delusions of parasitosis.

Table 51-4 Selected Common Dermatologic Conditions

Diagnosis	Common Distribution	Usual Morphology	Diagnosis	Common Distribution	Usual Morphology
Acnevulgaris	Face, upper back, chest	Open and closed comedones, erythematous papules, pustules, cysts	Seborrheic keratosis	Trunk, face	Brown plaques with adherent, greasy scale; "stuck on" appearance
Rosacea	Blush area of cheeks, nose	Erythema, telangiectases, papules, pustules	Folliculitis Impetigo	Any hair-bearing area	Follicular pustules

	nose, forehead, chin			Anywhere	Papules, vesicles, pustules, often with honey-colored crusts
Seborrheic dermatitis	Scalp, eyebrows, perinasal areas	Erythema with greasy yellow-brown scale	Herpes simplex	Lips, genitalia	Grouped vesicles progressing to crusted erosions
Atopic dermatitis	Antecubital and popliteal fossae; may be widespread	Patches and plaques of erythema, scaling, and lichenification; pruritus	Herpes zoster	Dermatomal, usually trunk but may be anywhere	Vesicles limited to a dermatome (often painful)
Stasis dermatitis	Ankles, lower legs over medial malleoli	Patches of erythema and scaling on background of hyperpigmentation associated with signs of venous insufficiency	Varicella	Face, trunk, relative sparing of extremities	Lesions arise in crops and quickly progress from erythematous macules, to papules, to vesicles, to pustules, to crusted sites
Dyshidrotic eczema	Palms, soles, sides of fingers and toes	Deep vesicles	Pityriasis rosea	Trunk (Christmas tree pattern); herald patch followed by multiple smaller lesions	Symmetric erythematous patches with a collarette of scale
Allergic contact dermatitis	Anywhere	Localized erythema, vesicles, scale, and pruritus (e.g., fingers, earlobes—nickel; dorsal aspect of foot—shoe; exposed surfaces—poison ivy)	Tinea versicolor	Chest, back, abdomen, proximal extremities	Scaly hyper- or hypopigmented macules
Psoriasis	Elbows, knees, scalp, lower back, fingernails (may be generalized)	Papules and plaques covered with silvery scale; nails have pits	Candidiasis	Groin, beneath breasts, vagina, oral cavity	Erythematous macerated areas with satellite pustules; white, friable patches on mucous membranes
Lichen planus	Wrists, ankles, mouth (may be widespread)	Violaceous flat-topped papules and plaques	Dermatophytosis	Feet, groin, beard, or scalp	Varies with site, (e.g., tinea corporis —scaly annular patch)
Keratosis pilaris	Extensor surfaces of arms and thighs, buttocks	Keratotic follicular papules with surrounding erythema	Scabies	Groin, axillae, between fingers and toes, beneath breasts	Excoriated papules, burrows, pruritus
Melasma	Forehead, cheeks, temples, upper lip	Tan to brown patches	Insect bites	Anywhere	Erythematous papules with central puncta
Vitiligo	Periorificial, trunk, extensor surfaces of extremities	Chalk-white macules	Cherry angioma Keloid Dermatofibroma	Trunk Anywhere (site of previous	Red, blood-filled papules Firm tumor, pink, purple, or

	extremities, flexor wrists, axillae			injury) Anywhere	brown Firm red to brown nodule that shows dimpling of overlying skin with lateral compression
Actinic keratosis	Sun-exposed areas	Skin-colored or red-brown macule or papule with dry, rough, adherent scale	Acrochordons (skin tags)	Groin, axilla, neck	Fleshy papules
Basal cell carcinoma	Face	Papule with pearly, telangiectatic border on sun-damaged skin	Urticaria	Anywhere	Wheals, sometimes with surrounding flare; pruritus
Squamous cell carcinoma	Face, especially lower lip, ears	Indurated and possibly hyperkeratotic lesions often showing ulceration and/or crusting	Transient acantholytic dermatosis	Trunk, especially anterior chest	Erythematous papules
			Xerosis	Extensor extremities, especially legs	Dry, erythematous, scaling patches; pruritus

Approach to the Patient: Skin Disorder

In examining the skin it is usually advisable to assess the patient before taking an extensive history. This way, the entire cutaneous surface is sure to be evaluated, and objective findings can be integrated with relevant historic data. Four basic features of a skin lesion must be noted and considered during a physical examination: the distribution of the eruption, the types of primary and secondary lesions, the shape of individual lesions, and the arrangement of the lesions. An ideal skin examination includes evaluation of the skin, hair, and nails as well as the mucous membranes of the mouth, eyes, nose, nasopharynx, and anogenital region. In the initial examination it is important that the patient be disrobed as completely as possible. This will minimize chances of missing important individual skin lesions and make it possible to assess the distribution of the eruption accurately. The patient should first be viewed from a distance of about 1.5–2 m (4–6 ft) so that the general character of the skin and the distribution of lesions can be evaluated. Indeed, the distribution of lesions often correlates highly with diagnosis **(Fig. 51-6)**. For example, a hospitalized patient with a generalized erythematous exanthem is more likely to have a drug eruption than is a patient with a similar rash limited to the sun-exposed portions of the face. Once the distribution of the lesions has been established, the nature of the primary lesion must be determined. Thus, when lesions are distributed on elbows, knees, and scalp, the most likely possibility based solely on distribution is psoriasis or dermatitis herpetiformis (**Figs. 51-7** and **51-8**, respectively). The primary lesion in psoriasis is a scaly papule that soon forms erythematous plaques covered with a white scale, whereas that of dermatitis herpetiformis is an urticarial papule that quickly becomes a small vesicle. In this manner, identification of the primary lesion directs the examiner toward the proper diagnosis. Secondary changes in skin can also be quite helpful. For example, scale represents excessive epidermis, while crust is the result of a discontinuous epithelial cell layer. Palpation of skin lesions can also yield insight into the character of an eruption. Thus, red papules on the lower extremities that blanch with pressure can be a manifestation of many different diseases, but hemorrhagic red papules that do not blanch with pressure indicate palpable purpura characteristic of necrotizing vasculitis (Fig. 51-4).

Figure 51-6

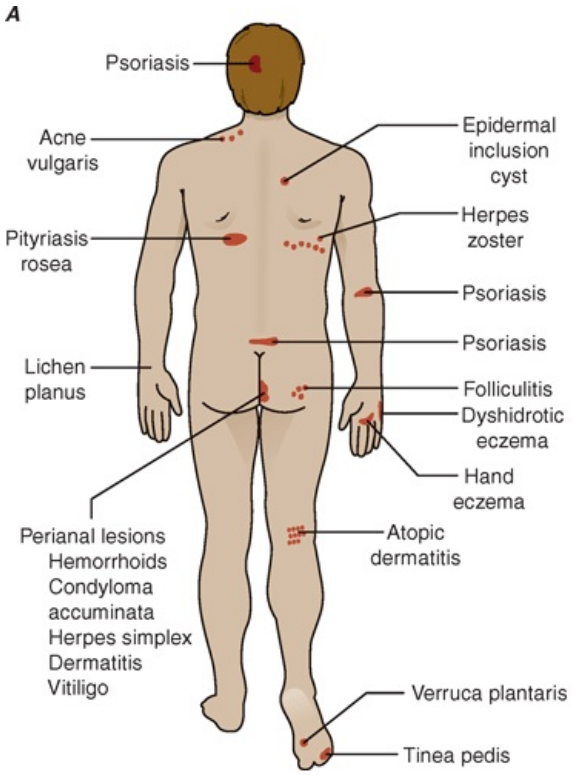

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

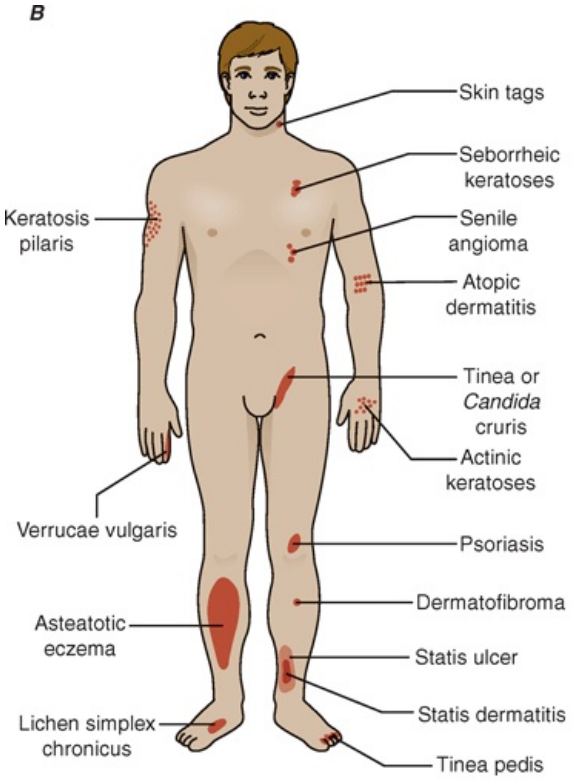

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

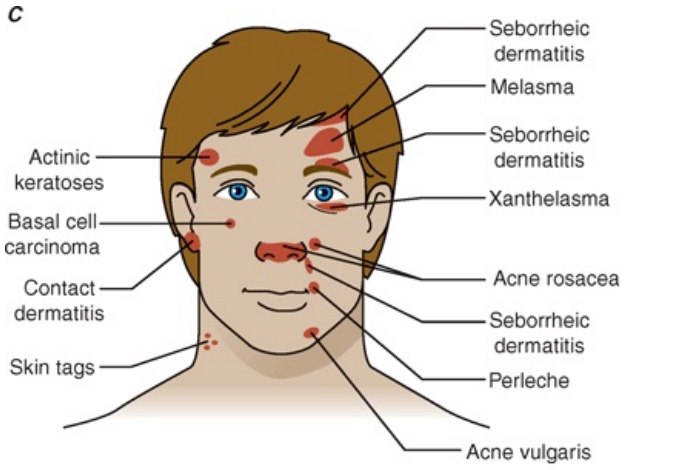

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

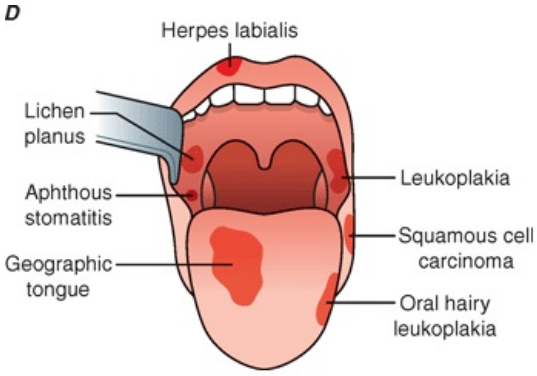

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A–D. **The distribution of some common dermatologic diseases and lesions.**

Figure 51-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Psoriasis. This papulosquamous skin disease is characterized by small and large erythematous papules and plaques with overlying adherent silvery scale.

Figure 51-8

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dermatitis herpetiformis. This disorder typically displays pruritic, grouped papulovesicles on elbows, knees, buttocks, and posterior scalp. Vesicles are often excoriated due to associated pruritus.

The shape of lesions is also an important feature. Flat, round, erythematous papules and plaques are common in many cutaneous diseases. However, target-shaped lesions that consist in part of erythematous plaques are specific for erythema multiforme **(Fig. 51-9)**. In the same way, the arrangement of individual lesions is important. Erythematous papules and vesicles can occur in many conditions, but their arrangement in a specific linear array suggests an external etiology such as allergic contact **(Fig. 51-10)** or primary irritant dermatitis. In contrast, lesions with a generalized arrangement are common and suggest a systemic etiology.

Figure 51-9

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema multiforme. This eruption is characterized by multiple erythematous plaques with a target or iris morphology. It usually represents a hypersensitivity reaction to drugs (e.g., sulfonylamides) or infections (e.g., HSV). *(Courtesy of the Yale Resident's Slide Collection; with permission.)*

Figure 51-10

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Allergic contact dermatitis (ACD). ***A.*** An example of ACD in its acute phase, with sharply demarcated, weeping, eczematous plaques in a perioral distribution. ***B.*** ACD in its chronic phase demonstrating an erythematous, lichenified, weeping plaque on skin chronically exposed to nickel in a metal snap. *(B, Courtesy of Robert Swerlick, MD; with permission.)*

As in other branches of medicine, a complete history should be obtained to emphasize the following features:

1. Evolution of lesions
    a. Site of onset
    b. Manner in which the eruption progressed or spread
    c. Duration
    d. Periods of resolution or improvement in chronic eruptions
2. Symptoms associated with the eruption
    a. Itching, burning, pain, numbness
    b. What, if anything, has relieved symptoms
    c. Time of day when symptoms are most severe
3. Current or recent medications (prescribed as well as over-the-counter)
4. Associated systemic symptoms (e.g., malaise, fever, arthralgias)
5. Ongoing or previous illnesses
6. History of allergies
7. Presence of photosensitivity
8. Review of systems

9. Family history (particularly relevant for patients with melanoma, atopy, psoriasis, or acne)
10. Social, sexual, or travel history as relevant to the patient

Diagnostic Techniques

Many skin diseases can be diagnosed on gross clinical appearance, but sometimes relatively simple diagnostic procedures can yield valuable information. In most instances, they can be performed at the bedside with a minimum of equipment.

SKIN BIOPSY

A skin biopsy is a straightforward minor surgical procedure; however, it is important to biopsy a lesion that is most likely to yield diagnostic findings. This decision may require expertise in skin diseases and knowledge of superficial anatomic structures in selected areas of the body. In this procedure, a small area of skin is anesthetized with 1% lidocaine with or without epinephrine. The skin lesion in question can be excised or saucerized with a scalpel or removed by punch biopsy. In the latter technique, a punch is pressed against the surface of the skin and rotated with downward pressure until it penetrates to the subcutaneous tissue. The circular biopsy is then lifted with forceps, and the bottom is cut with iris scissors. Biopsy sites may or may not need suture closure, depending on size and location.

KOH PREPARATION

A potassium hydroxide (KOH) preparation is performed on scaling skin lesions where a fungal infection is suspected. The edge of such a lesion is scraped gently with a no. 15 scalpel blade, and the removed scale is collected on a glass microscope slide then treated with 1 to 2 drops of a solution of 10–20% KOH. KOH dissolves keratin and allows easier visualization of fungal elements. Brief heating of the slide accelerates dissolution of keratin. When the preparation is viewed under the microscope, the refractile hyphae will be seen more easily when the light intensity is reduced and the condenser is lowered. This technique can be used to identify hyphae in dermatophyte infections, pseudohyphae and budding yeast in *Candida* infections (see Fig. 203-1), and "spaghetti and meatballs" yeast forms in tinea versicolor. The same sampling technique can be used to obtain scale for culture of selected pathogenic organisms.

TZANCK SMEAR

A Tzanck smear is a cytologic technique most often used in the diagnosis of herpesvirus infections [herpes simplex virus (HSV) or varicella zoster virus (VZV)] (see Figs. 180-1 and 180-3). An early vesicle, not a pustule or crusted lesion, is unroofed, and the base of the lesion is scraped gently with a scalpel blade. The material is placed on a glass slide, air-dried, and stained with Giemsa or Wright's stain. Multinucleated epithelial giant cells suggest the presence of HSV or VZV; culture or immunofluorescence, or genetic testing must be performed to identify the specific virus.

DIASCOPY

Diascopy is designed to assess whether a skin lesion will blanch with pressure as, for example, in determining whether a red lesion is hemorrhagic or simply blood-filled. Urticaria **(Fig. 51-11)** will blanch with pressure, whereas a purpuric lesion caused by necrotizing vasculitis (Fig. 51-4) will not. Diascopy is performed by pressing a microscope slide or magnifying lens against a lesion and noting the amount of blanching that occurs. Granulomas often have an opaque to transparent, brown-pink "apple jelly" appearance on diascopy.

Figure 51-11

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Urticaria. Discrete and confluent, edematous, erythematous papules and plaques are characteristic of this whealing eruption.

WOOD'S LIGHT

A Wood's lamp generates 360-nm ultraviolet (or "black") light that can be used to aid the evaluation of certain skin disorders. For example, a Wood's lamp will cause erythrasma (a superficial, intertriginous infection caused by *Corynebacterium minutissimum*) to show a characteristic coral pink color, and wounds colonized by *Pseudomonas* to appear pale blue. Tinea capitis caused by certain dermatophytes such as *Microsporum canis* or *M. audouini* exhibits a yellow fluorescence. Pigmented lesions of the epidermis such as freckles are accentuated, while dermal pigment such as postinflammatory hyperpigmentation fades under a Wood's light. Vitiligo **(Fig. 51-12)** appears totally white under a Wood's lamp, and previously unsuspected areas of involvement often become apparent. A Wood's lamp may also aid in the demonstration of tinea versicolor and in recognition of ash leaf spots in patients with tuberous sclerosis.

Figure 51-12

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vitiligo. Characteristic lesions display an acral distribution and striking depigmentation as a result of loss of melanocytes.

PATCH TESTS

Patch testing is designed to document sensitivity to a specific antigen. In this procedure, a battery of suspected allergens is applied to the patient's back under occlusive dressings and allowed to remain in contact with the skin for 48 h. The dressings are removed, and the area is examined for evidence of delayed hypersensitivity reactions (e.g., erythema, edema, or papulovesicles). This test is best performed by physicians with special expertise in patch testing and is often helpful in the evaluation of patients with chronic dermatitis.

FURTHER READINGS

Habif TP: *Clinical Dermatology: A Color Guide to Diagnosis and Therapy,* 4th ed. Philadelphia, Mosby, 2004

James WD et al: *Andrews Diseases of the Skin: Clinical Dermatology,* 10th ed. Philadelphia, Elsevier, 2006

Wolff K et al (eds): *Fitzpatricks Dermatology in General Medicine*, 7th ed. New York, McGraw-Hill, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ECZEMA AND DERMATITIS

Eczema is a type of dermatitis and these terms are often used synonymously (atopic eczema or atopic dermatitis). Eczema is a reaction pattern that presents with variable clinical findings and the common histologic finding of spongiosis (intercellular edema of the epidermis). Eczema is the final common expression for a number of disorders, including those discussed in the following sections. Primary lesions may include erythematous macules, papules, and vesicles, which can coalesce to form patches and plaques. In severe eczema, secondary lesions from infection or excoriation, marked by weeping and crusting, may predominate. In chronic eczematous conditions, lichenification (cutaneous hypertrophy and accentuation of normal skin markings) may alter the characteristic appearance of eczema.

Atopic Dermatitis

Atopic dermatitis (AD) is the cutaneous expression of the atopic state, characterized by a family history of asthma, allergic rhinitis, or eczema. The prevalence of AD is increasing worldwide. Some of its features are shown in **Table 52-1**.

Table 52-1 Clinical Features of Atopic Dermatitis

1. Pruritus and scratching
2. Course marked by exacerbations and remissions
3. Lesions typical of eczematous dermatitis
4. Personal or family history of atopy (asthma, allergic rhinitis, food allergies, or eczema)
5. Clinical course lasting longer than 6 weeks
6. Lichenification of skin

The etiology of AD is only partially defined, but there is a clear genetic predisposition. When both parents are affected by AD, >80% of their children manifest the disease. When only one parent is affected, the prevalence drops to slightly over 50%. Patients with AD may display a variety of immunoregulatory abnormalities including increased IgE synthesis; increased serum IgE; and impaired, delayed-type hypersensitivity reactions.

The clinical presentation often varies with age. Half of patients with AD present within the first year of life, and 80% present by 5 years of age. About 80% ultimately coexpress allergic rhinitis or asthma. The infantile pattern is characterized by weeping inflammatory patches and crusted plaques on the face, neck, and extensor surfaces. The childhood and adolescent pattern is marked by dermatitis of flexural skin, particularly in the antecubital and popliteal fossae **(Fig. 52-1)**. AD may resolve spontaneously, but approximately 40% of all individuals affected as children will have dermatitis in adult life. The distribution of lesions may be similar to those seen in childhood; however, adults frequently have localized disease, manifesting as lichen simplex chronicus or hand eczema (see below). In patients with localized disease, AD may be suspected because of a typical personal history, family history, or the presence of cutaneous stigmata of AD such as perioral pallor, an extra fold of skin beneath the lower eyelid (Dennie-Morgan folds), increased palmar skin markings, and an increased incidence of cutaneous infections, particularly with *Staphylococcus aureus*. Regardless of other manifestations, pruritus is a prominent characteristic of AD in all age groups and is exacerbated by dry skin. Many of the cutaneous findings in affected patients, such as lichenification, are secondary to rubbing and scratching.

Figure 52-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atopic dermatitis. Hyperpigmentation, lichenification, and scaling in the antecubital fossae are seen in this patient with atopic dermatitis. (*Courtesy of Robert Swerlick, MD; with permission* .)

TREATMENT: ATOPIC DERMATITIS

Therapy of AD should include avoidance of cutaneous irritants, adequate moisturizing through the application of emollients, judicious use of topical anti-inflammatory agents, and prompt treatment of secondary infection. Patients should be instructed to bathe no more often than daily, using warm or cool water, and to use only mild bath soap. Immediately after bathing, while the skin is still moist, a topical anti-inflammatory agent in a cream or ointment base should be applied to areas of dermatitis, and all other skin areas should be lubricated with a moisturizer. Approximately 30 g of a topical agent is required to cover the entire body surface of an average adult.

Low- to midpotency topical glucocorticoids are employed in most treatment regimens for AD. Skin atrophy and the potential for systemic absorption are constant concerns, especially with more potent agents. Low-potency topical glucocorticoids or nonglucocorticoid anti-inflammatory agents should be selected for use on the face and intertriginous areas to minimize the risk of skin atrophy. Two nonglucocorticoid anti-inflammatory agents are available: tacrolimus ointment and pimecrolimus cream. These agents are macrolide immunosuppressants that are approved by the U.S. Food and Drug Administration (FDA) for topical use in AD. Reports of broader effectiveness appear in the literature. These agents do not cause skin atrophy, nor do they suppress the hypothalamic-pituitary-adrenal axis. Recently, however, concerns have emerged regarding the potential for lymphomas in patients treated with these agents. Thus, caution should be exercised when considering these agents. Currently, they are also more costly than topical glucocorticoids. Barrier-repair products are also nonglucocorticoid agents and are gaining popularity in treating AD.

Secondary infection of eczematous skin may lead to exacerbation of AD. Crusted and weeping skin lesions may be infected with *S. aureus*. When secondary infection is suspected, eczematous lesions should be cultured and patients treated with systemic antibiotics active against *S. aureus*. The initial use of penicillinase-resistant penicillins or cephalosporins is preferable. Dicloxacillin or cephalexin (250 mg qid for 7–10 days) is generally adequate for adults; however, antibiotic selection must be directed by culture results and clinical response. More than 50% of *S. aureus* (SA) isolates are now methacillin resistant (MR) in some communities—community-acquired MRSA (CA-MRSA). Current recommendations for the treatment of CA-MRSA infection in adults include trimethoprim/sulfamethoxazole (1 double strength bid), minocycline (100 mg bid), doxycycline (100 mg bid), or clindamycin (300–450 mg qid). Duration of therapy should be 7–10 days. Inducible resistance may limit clindamycin's usefulness. The latter can be detected by the double-disk diffusion test, which should be ordered if the isolate is erythromycin-resistant and clindamycin-sensitive. As an adjunct, the use of antibacterial washes or dilute sodium hypochlorite baths (0.005% bleach) and intermittent nasal mupirocin may be useful.

Control of pruritus is essential for treatment, because AD often represents "an itch that rashes." Antihistamines are most often used to control pruritus, and mild sedation may be responsible for their antipruritic action. Sedation may also limit their usefulness; however, when used at bedtime, sedating antihistamines may improve the patient's sleep. Unlike their effects in urticaria, nonsedating antihistamines and selective H_2 blockers are of little use in controlling the pruritus of AD.

Treatment with systemic glucocorticoids should be limited to severe exacerbations unresponsive to topical therapy. In the patient with chronic AD, therapy with systemic glucocorticoids will generally clear the skin only briefly, and cessation of the systemic therapy will invariably be accompanied by return, if

not worsening, of the dermatitis. Patients who do not respond to conventional therapies should be considered for patch testing to rule out allergic contact dermatitis (ACD). The role of dietary allergens in atopic dermatitis is controversial, and there is little evidence they play any role outside of infancy where a small percentage of patients with AD may be affected by food allergens.

Lichen Simplex Chronicus

Lichen simplex chronicus may represent the end stage of a variety of pruritic and eczematous disorders, including atopic dermatitis. It consists of a circumscribed plaque or plaques of lichenified skin due to chronic scratching or rubbing. Common areas involved include the posterior nuchal region, dorsum of the feet, and ankles. Treatment of lichen simplex chronicus centers on breaking the cycle of chronic itching and scratching. High-potency topical glucocorticoids are helpful in most cases, but in recalcitrant cases, application of topical glucocorticoids under occlusion, or intralesional injection of glucocorticoids may be required. Oral antihistamines such as hydroxyzine (10–25 mg every 6 h) or tricyclic antidepressants with antihistaminic activity, such as doxepin (10–25 mg at bedtime), are useful primarily due to their sedating action. Higher doses of these agents may be required, but sedation can become bothersome. Patients need to be counseled regarding driving or operating heavy equipment after taking these medications.

Contact Dermatitis

Contact dermatitis is an inflammatory process in skin caused by an exogenous agent or agents that directly or indirectly injure the skin. This injury may be caused by an inherent characteristic of a compound—irritant contact dermatitis (ICD). An example of ICD would be dermatitis induced by a concentrated acid or base. Agents that cause ACD induce an antigen-specific immune response (poison ivy dermatitis). The clinical lesions of contact dermatitis may be acute (wet and edematous) or chronic (dry, thickened, and scaly), depending on the persistence of the insult (see Fig. 51-10).

IRRITANT CONTACT DERMATITIS

ICD is generally well demarcated and often localized to areas of thin skin (eyelids, intertriginous areas) or to areas where the irritant was occluded. Lesions may range from minimal skin erythema to areas of marked edema, vesicles, and ulcers. Chronic low-grade irritant dermatitis is the most common type of ICD, and the most common area of involvement is the hands (see below). The most common irritants encountered are chronic wet work, soaps, and detergents. Treatment should be directed to avoidance of irritants and use of protective gloves or clothing.

ALLERGIC CONTACT DERMATITIS

ACD is a manifestation of delayed-type hypersensitivity mediated by memory T lymphocytes in the skin. The most common cause of ACD is exposure to plants, especially to members of the family Anacardiaceae, including the genus *Toxicodendron*. Poison ivy, poison oak, and poison sumac are members of this genus and cause an allergic reaction marked by erythema, vesiculation, and severe pruritus. The eruption is often linear or angular, corresponding to areas where plants have touched the skin. The sensitizing antigen common to these plants is urushiol, an oleoresin containing the active ingredient pentadecylcatechol. The oleoresin may adhere to skin, clothing, tools, and pets, and contaminated articles may cause dermatitis even after prolonged storage. Blister fluid does not contain urushiol and is not capable of inducing skin eruption in exposed subjects.

TREATMENT: CONTACT DERMATITIS

If contact dermatitis is suspected and an offending agent is identified and removed, the eruption will resolve. Usually, treatment with high-potency topical glucocorticoids is enough to relieve symptoms while the dermatitis runs its course. For those patients who require systemic therapy, daily oral prednisone beginning at 1 mg/kg, but usually ≤60 mg/d, is sufficient. It should be tapered over 2 to 3 weeks, and each daily dose given in the morning with food.

Identification of a contact allergen can be a difficult and time-consuming task. Patients with dermatitis unresponsive to conventional therapy or with an unusual and patterned distribution should be suspected of having ACD. They should be questioned carefully regarding occupational exposures and topical medications. Common sensitizers include preservatives in topical preparations, nickel sulfate, potassium dichromate, thimerosal, neomycin sulfate, fragrances, formaldehyde, and rubber-curing agents. Patch testing is helpful in identifying these agents but should not be attempted on patients with widespread active dermatitis or on those taking systemic glucocorticoids.

Hand Eczema

Hand eczema is a very common, chronic skin disorder in which both exogenous and endogenous factors play important roles. It may be associated with other cutaneous disorders such as atopic dermatitis, and contact with various agents may be involved. It represents a large proportion of occupation-associated skin disease. Chronic, excessive exposure to water and detergents, harsh chemicals, or allergens may initiate or aggravate this disorder. It may present with dryness and cracking of the skin of the hands as well as with variable amounts of erythema and edema. Often, the dermatitis will begin under rings where water and irritants are trapped. Dyshidrotic eczema, a variant of hand eczema, presents with multiple, intensely pruritic, small papules and vesicles occurring on the

thenar and hypothenar eminences and the sides of the fingers **(Fig. 52-2)**. Lesions tend to occur in crops that slowly form crusts and then heal.

Figure 52-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dyshidrotic eczema. This example is characterized by deep-seated vesicles and scaling on palms and lateral fingers, and the disease is often associated with an atopic diathesis.

The evaluation of a patient with hand eczema should include an assessment of potential occupation-associated exposures. The history should be directed to identifying possible irritant or allergen exposures.

TREATMENT: HAND ECZEMA

Therapy of hand eczema is directed toward avoidance of irritants, identification of possible contact allergens, treatment of coexistent infection, and application of topical glucocorticoids. Whenever possible, the hands should be protected by gloves, preferably vinyl. The use of rubber gloves (latex) to protect dermatitic skin is sometimes associated with the development of hypersensitivity reactions to components of the gloves. Patients can be treated with cool moist compresses, followed by application of a mid- to high-potency topical glucocorticoid in a cream or ointment base. As with atopic dermatitis, treatment of secondary infection is essential for good control. In addition, patients with hand eczema should be examined for dermatophyte infection by KOH preparation and culture (see below).

Nummular Eczema

Nummular eczema is characterized by circular or oval "coinlike" lesions, beginning as small edematous papules that become crusted and scaly. The etiology of nummular eczema is unknown, but dry skin is a contributing factor. Common locations are the trunk or the extensor surfaces of the extremities, particularly on the pretibial areas or dorsum of the hands. It occurs more frequently in men and is most commonly seen in middle age. The treatment of nummular eczema is similar to that for atopic dermatitis.

Asteatotic Eczema

Asteatotic eczema, also known as *xerotic eczema* or *"winter itch,"* is a mildly inflammatory dermatitis that develops in areas of extremely dry skin, especially during the dry winter months. Clinically, there may be considerable overlap with nummular eczema. This form of eczema accounts for a large number of physician visits because of the associated pruritus. Fine cracks and scale, with or without erythema, characteristically develop in areas of dry skin, especially on the anterior surfaces of the lower extremities in elderly patients. Asteatotic eczema responds well to topical moisturizers and the avoidance of cutaneous irritants. Overbathing and the use of harsh soaps exacerbate asteatotic eczema.

Stasis Dermatitis and Stasis Ulceration

Stasis dermatitis develops on the lower extremities secondary to venous incompetence and chronic edema. Patients may give a history of deep venous thrombosis, have evidence of vein removal, or varicose veins. Early findings in stasis dermatitis consist of mild erythema and scaling associated with pruritus. The typical initial site of involvement is the medial aspect of the ankle, often over a distended vein **(Fig. 52-3)**.

Figure 52-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stasis dermatitis. An example of stasis dermatitis showing erythematous, scaly, and oozing patches over the lower leg. Several stasis ulcers are also seen in this patient.

Stasis dermatitis may become acutely inflamed, with crusting and exudate. In this state, it is easily confused with cellulitis. Chronic stasis dermatitis is often associated with dermal fibrosis that is recognized clinically as brawny edema of the skin. As the disorder progresses, the dermatitis becomes progressively pigmented, due to chronic erythrocyte extravasation leading to cutaneoushemosiderin deposition. Stasis dermatitis may be complicated by secondary infection and contact dermatitis. Severe stasis dermatitis may precede the development of stasis ulcers.

TREATMENT: STASIS DERMATITIS AND STASIS ULCERATION

Patients with stasis dermatitis and stasis ulceration benefit greatly from leg elevation and the routine use of compression stockings with a gradient of at least 30–40 mmHg. Stockings providing less compression, such as antiembolism hose, are poor substitutes. Use of emollients and/or midpotency topical glucocorticoids and avoidance of irritants are also helpful in treating stasis dermatitis. Protecting the legs from injury, including scratching, and control of chronic edema are essential to prevent ulcers. Diuretics may be required to adequately control chronic edema.

Stasis ulcers are difficult to treat, and resolution is slow. It is extremely important to elevate the affected limb as much as possible. The ulcer should be kept clear of necrotic material by gentle debridement and covered with a semipermeable dressing and a compression dressing or compression stocking. Glucocorticoids should not be applied to ulcers, because they may retard healing; however, they may be applied to the surrounding skin to control itching, scratching, and additional trauma. Secondarily infected lesions should be treated with appropriate oral antibiotics, but it should be noted that all ulcers will become colonized with bacteria, and the purpose of antibiotic therapy should not be to clear all bacterial growth. Care must be taken to exclude treatable causes of leg ulcers (hypercoagulation, vasculitis) before beginning the chronic management outlined above.

Seborrheic Dermatitis

Seborrheic dermatitis is a common, chronic disorder, characterized by greasy scales overlying erythematous patches or plaques. Induration and scale are generally less prominent than in psoriasis, but clinical overlap exists between these diseases—"sebopsoriasis." The most common location is in the scalp, where it may be recognized as severe dandruff. On the face, seborrheic dermatitis affects

the eyebrows, eyelids, glabella, and nasolabial folds **(Fig. 52-4)**. Scaling of the external auditory canal is common in seborrheic dermatitis. In addition, the postauricular areas often become macerated and tender. Seborrheic dermatitis may also develop in the central chest, axilla, groin, submammary folds, and gluteal cleft. Rarely, it may cause a widespread generalized dermatitis. Pruritus is variable.

Figure 52-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Seborrheic dermatitis. Central facial erythema with overlying greasy, yellowish scale is seen in this patient. (*Courtesy of Jean Bolognia, MD; with permission*.)

Seborrheic dermatitis may be evident within the first few weeks of life, and within this context it typically occurs in the scalp ("cradle cap"), face, or groin. It is rarely seen in children beyond infancy but becomes evident again during adult life. Although it is frequently seen in patients with Parkinson's disease, in those who have had cerebrovascular accidents, and in those with HIV infection, the overwhelming majority of individuals with seborrheic dermatitis have no underlying disorder.

TREATMENT: SEBORRHEIC DERMATITIS

Treatment with low-potency topical glucocorticoids in conjunction with a topical antifungal agent, such as ketoconazole cream or ciclopirox cream, is often effective. The scalp and beard areas may benefit from antidandruff shampoos, which should be left in place 3–5 min before rinsing. High-potency topical glucocorticoid solutions (betamethasone or clobetasol) are effective for control of severe scalp involvement. High-potency glucocorticoids should not be used on the face because this is often associated with steroid-induced rosacea or atrophy.

PAPULOSQUAMOUS DISORDERS

See Table 52-2.

Table 52-2 Papulosquamous Disorders

	Clinical Features	Other Notable Features	Histologic Features
Psoriasis	Sharply demarcated, erythematous plaques with mica-like scale; predominantly elbows, knees, and scalp; atypical forms may localize to intertriginous areas; eruptive forms may be associated with infection	May be aggravated by certain drugs, infection; severe forms seen associated with HIV	Acanthosis, vascular proliferation
Lichen planus	Purple polygonal papules marked by severe pruritus; lacy white markings, especially associated with mucous membrane lesions	Certain drugs may induce: thiazides, antimalarial drugs	Interface dermatitis
Pityriasis rosea	Rash often preceded by herald patch; oval to	Variable pruritus;	Pathologic

	round plaques with trailing scale; most often affects the trunk, and eruption lines up in skinfolds giving a "fir treelike" appearance; generally spares palms and soles	self-limited resolving in 2–8 weeks; may be imitated by secondary syphilis	features often nonspecific
Dermatophytosis	Polymorphous appearance depending on dermatophyte, body site, and host response; sharply defined to ill-demarcated scaly plaques with or without inflammation; may be associated with hair loss	KOH preparation may show branching hyphae; culture helpful	Hyphae and neutrophils in stratum corneum

Psoriasis

Psoriasis is one of the most common dermatologic diseases, affecting up to 1% of the world's population. It is a chronic inflammatory skin disorder clinically characterized by erythematous, sharply demarcated papules and rounded plaques, covered by silvery micaceous scale. The skin lesions of psoriasis are variably pruritic. Traumatized areas often develop lesions of psoriasis (Koebner's or isomorphic phenomenon). In addition, other external factors may exacerbate psoriasis including infections, stress, and medications (lithium, beta blockers, and antimalarials).

The most common variety of psoriasis is called *plaque-type*. Patients with plaque-type psoriasis will have stable, slowly enlarging plaques, which remain basically unchanged for long periods of time. The most commonly involved areas are the elbows, knees, gluteal cleft, and the scalp. Involvement tends to be symmetric. Plaque psoriasis generally develops slowly and runs an indolent course. It rarely remits spontaneously. *Inverse psoriasis* affects the intertriginous regions including the axilla, groin, submammary region, and navel; it also tends to affect the scalp, palms, and soles. The individual lesions are sharply demarcated plaques (see Fig. 51-7), but they may be moist and without scale due to their locations.

Guttate psoriasis (eruptive psoriasis) is most common in children and young adults. It develops acutely in individuals without psoriasis or in those with chronic plaque psoriasis. Patients present with many small erythematous, scaling papules, frequently after upper respiratory tract infection with β-hemolytic streptococci. The differential diagnosis should include pityriasis rosea and secondary syphilis.

Pustular psoriasis is another variant. Patients may have disease localized to the palms and soles, or the disease may be generalized. Regardless of the extent of disease, the skin is erythematous with pustules and variable scale. Localized to the palms and soles, it is easily confused with eczema. When generalized, episodes are characterized by fever [39°–40°C (102.2°–104.0°F)] lasting several days, an accompanying generalized eruption of sterile pustules, and a background of intense erythema; patients may become erythrodermic. Episodes of fever and pustules are recurrent. Local irritants, pregnancy, medications, infections, and systemic glucocorticoid withdrawal can precipitate this form of psoriasis. Oral retinoids are the treatment of choice in nonpregnant patients.

Fingernail involvement, appearing as punctate pitting, onycholysis, nail thickening, or subungual hyperkeratosis may be a clue to the diagnosis of psoriasis when the clinical presentation is not classic.

According to the National Psoriasis Foundation, up to 30% of patients with psoriasis have psoriatic arthritis (PsA). There are five subtypes of PsA: symmetric, asymmetric, distal interphalangeal predominant (DIP), spondylitis, and arthritis mutilans. Symmetric arthritis resembles rheumatoid arthritis, but is usually milder. Asymmetric arthritis can involve any joint and may present as "sausage digits." DIP is the classic form, but occurs in only about 5% of patients with PsA. It may involve fingers and toes. Spondylitis also occurs in about 5% of patients with PsA. Arthritis mutilans is severe and deforming. It affects primarily the small joints of the hands and feet. It accounts for less than 5% of PsA.

The etiology of psoriasis is still poorly understood, but there is clearly a genetic component to the disease. Over 50% of patients with psoriasis report a positive family history. Psoriatic lesions demonstrate infiltrates of activated T cells that are thought to elaborate cytokines responsible for keratinocyte hyperproliferation, which results in the characteristic clinical findings. Agents inhibiting T cell activation, clonal expansion, or release of proinflammatory cytokines are often effective for the treatment of severe psoriasis (see below).

TREATMENT: PSORIASIS

Treatment of psoriasis depends on the type, location, and extent of disease. All patients should be instructed to avoid excess drying or irritation of their skin and to maintain adequate cutaneous hydration. Most patients with localized, plaque-type psoriasis can be managed with midpotency topical glucocorticoids, although their long-term use is often accompanied by loss of effectiveness (tachyphylaxis) and atrophy of the skin. A topical vitamin D analogue (calcipotriene) and a retinoid (tazarotene) are also efficacious in the treatment of limited psoriasis and have largely replaced other topical agents such as coal tar, salicylic acid, and anthralin.

Ultraviolet light, natural or artificial, is an effective therapy for many patients with widespread psoriasis. Ultraviolet B (UV-B) light, narrowband UV-B, and ultraviolet A (UV-A) spectrum with either oral or topical

psoralens (PUVA) are also extremely effective. The long-term use of UV light may be associated with an increased incidence of nonmelanoma and melanoma skin cancer. UV-light therapy is contraindicated in patients receiving cyclosporine and should be used with great care in all immunocompromised patients due to an increased risk of developing skin cancers.

Various systemic agents can be used for severe, widespread psoriatic disease **(Table 52-3)**. Oral glucocorticoids should not be used for the treatment of psoriasis due to the potential for developing life-threatening pustular psoriasis when therapy is discontinued. Methotrexate is an effective agent, especially in patients with psoriatic arthritis. The synthetic retinoid acitretin is useful, especially when immunosuppression must be avoided; however, teratogenicity limits its use.

Table 52-3 FDA-Approved Systemic Therapy for Psoriasis

	Medication	Administration		
Agent	**Class**	**Route**	**Frequency**	**Adverse Events (Selected)**
Methotrexate	Antimetabolite	Oral	Weekly	Hepatotoxicity, pulmonary toxicity, pancytopenia, potential for increased malignancies, ulcerative stomatitis, nausea, diarrhea, teratogenicity
Acitretin	Retinoid	Oral	Daily	Teratogenicity, osteophyte formation, hyperlipidemia, flare of inflammatory bowel disease, hepatoxicity, depression
Cyclosporine	Calcineurin inhibitor	Oral	Twice daily	Renal dysfunction, hypertension, hyperkalemia, hyperuricemia, hypomagnesemia, hyperlipidemia, increased risk of malignancies

The evidence implicating psoriasis as a T cell–mediated disorder has directed therapeutic efforts to immunoregulation. Cyclosporine and other immunosuppressive agents can be very effective in the treatment of psoriasis, and much attention is currently directed toward the development of biologic agents with more selective immunosuppressive properties and better safety profiles **(Table 52-4)**. Experience with these agents is limited, and information regarding combination therapy and adverse events continues to emerge. Use of tumor necrosis factor (TNF-α) inhibitors may worsen congestive heart failure (CHF), and they should be used with caution in those at risk for or known to have CHF. Further, none of the immunosuppressive agents used in the treatment of psoriasis should be initiated if the patient has a severe infection; patients on such therapy should be routinely screened for tuberculosis. There have been reports of progressive multifocal leukoencephalopathy in association with treatment with the TNF-α inhibitors. Malignancies, including a risk or history of certain malignancies, may limit the use of these systemic agents.

Table 52-4 Biologics Approved for Psoriasis or Psoriatic Arthritis

	Mechanism		Administration		
Agent	**of Action**	**Indication**	**Route**	**Frequency**	**Warnings**
Alefacept	Anti-CD-2	Ps	IM	Once weekly x 12 weeks; may repeat	Lymphopenia, potential for increased malignancies, serious infections
Etanercept	Anti TNF-α	Ps, PsA	SC	Once or twice weekly	Serious infections, neurologic events, hematologic events, potential for increased malignancies
Adalimumab	Anti TNF-α	PsA	SC	Every other week	Serious infections, neurologic events, potential for increased malignancies, hypersensitivity reactions, hematologic events
Infliximab	Anti TNF-α	PsA	IV	Initial infusion followed by infusions at week 2, 6, then every 8 weeks	Serious infections, hepatotoxicity, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies

Abbreviations: IM, intramuscular; Ps, psoriasis; PsA, psoriatic arthritis; SC, subcutaneous; TNF, tumor necrosis factor.

Lichen Planus

Lichen planus (LP) is a papulosquamous disorder that may affect the skin, scalp, nails, and mucous membranes. The primary cutaneous lesions are pruritic, polygonal, flat-topped, violaceous papules. Close examination of the surface of these papules often reveals a network of gray lines (Wickham's striae). The skin lesions may occur anywhere but have a predilection for the wrists, shins, lower back,

and genitalia **(Fig. 52-5)**. Involvement of the scalp, lichen planopilaris, may lead to scarring alopecia, and nail involvement may lead to permanent deformity or loss of fingernails and toenails. LP commonly involves mucous membranes, particularly the buccal mucosa, where it can present a spectrum of disease from a mild, white, reticulate eruption of the mucosa to a severe, erosive stomatitis. Erosive stomatitis may persist for years and may be linked to an increased risk of oral squamous cell carcinoma. Cutaneous eruptions clinically resembling LP have been observed after administration of numerous drugs, including thiazide diuretics, gold, antimalarials, penicillamine, and phenothiazines, and in patients with skin lesions of chronic graft-versus-host disease. In addition, LP may be associated with hepatitis C infection. The course of LP is variable, but most patients have spontaneous remissions 6 months to 2 years after the onset of disease. Topical glucocorticoids are the mainstay of therapy.

Figure 52-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lichen planus. An example of lichen planus showing multiple flat-topped, violaceous papules and plaques. Nail dystrophy as seen in this patient's thumbnail may also be a feature. (*Courtesy of Robert Swerlick, MD; with permission.*)

Pityriasis Rosea

Pityriasis rosea (PR) is a papulosquamous eruption of unknown etiology occurring more commonly in the spring and fall. Its first manifestation is the development of a 2- to 6-cm annular lesion (the herald patch). This is followed in a few days to a few weeks by the appearance of many smaller annular or papular lesions with a predilection to occur on the trunk **(Fig. 52-6)**. The lesions are generally oval, with their long axis parallel to the skinfold lines. Individual lesions may range in color from red to brown and have a trailing scale. PR shares many clinical features with the eruption of secondary syphilis, but palm and sole lesions are extremely rare in PR and common in secondary syphilis. The eruption tends to be moderately pruritic and lasts 3 to 8 weeks. Treatment is directed at alleviating pruritus and consists of oral antihistamines; midpotency topical glucocorticoids; and, in some cases, the use of UV-B phototherapy.

Figure 52-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pityriasis rosea. In this patient with pityriasis rosea, multiple round to oval erythematous patches with fine central scale are distributed along the skin tension lines on the trunk.

CUTANEOUS INFECTIONS (TABLE 52-5)

See Table 52-5.

Table 52-5 Common Skin Infections

	Clinical Features	Etiologic Agent	Treatment
Impetigo	Honey-colored crusted papules, plaques, or bullae	Group A streptococcus and *Staphylococcus aureus*	Systemic or topical antistaphylococcal antibiotics
Dermatophytosis	Inflammatory or noninflammatory annular scaly plaques; may have hair loss; groin involvement spares scrotum; hyphae on KOH preparation	*Trichophyton, Epidermophyton*, or *Microsporum* sp.	Topical azoles, systemic griseofulvin, terbinafine, or azoles
Candidiasis	Inflammatory papules and plaques with satellite pustules, frequently in intertriginous areas; may involve scrotum; pseudohyphae on KOH preparation	*Candida albicans* and other *Candida* sp.	Topical nystatin or azoles; systemic azoles for resistant disease
Tinea versicolor	Hyper- or hypopigmented scaly patches on the trunk; characteristic mixture of hyphae and spores ("spaghetti and meatballs") on KOH preparation	*Malassezia furfur*	Topical selenium sulfide lotion or azoles

Impetigo, Ecthyma, and Furunculosis

Impetigo is a common superficial bacterial infection of skin caused most often by *S. aureus* (Chap. 135), and in some cases by group A β-hemolytic streptococci (Chap. 136). The primary lesion is a superficial pustule that ruptures and forms a characteristic yellow-brown honey-colored crust (Chap. 136). Lesions may occur on normal skin—primary infection—or in areas already affected by another skin disease—secondary infection. Lesions caused by staphylococci may be tense, clear bullae, and this less common form of the disease is called *bullous impetigo*. Blisters are caused by the production of exfoliative toxin by *S. aureus* phage type II. This is the same toxin responsible for staphylococcal scalded-skin syndrome (SSSS), often resulting in dramatic loss of the superficial epidermis due to blistering. SSSS is much more common in children than in adults; however, it should be considered along with toxic epidermal necrolysis and severe drug eruptions in patients with widespread blistering

of the skin. *Ecthyma* is a variant of impetigo that causes punched-out ulcerative lesions. It may result from neglected or inadequately treated impetigo. Treatment of both ecthyma and impetigo involves gentle debridement of adherent crusts, which is facilitated by the use of soaks and topical antibiotics, in conjunction with appropriate oral antibiotics. *Furunculosis* is also caused by *S. aureus*, and this disorder has gained prominence in the last decade because of CA-MRSA. A furuncle, or boil, is a painful, erythematous, nodule that can occur on any cutaneous surface. The lesions may be solitary but are most often multiple. Patients frequently believe they have been bitten by spiders or insects. Family members or close contacts may also be affected. Furuncles can rupture and drain spontaneously or may need incision and drainage, which may be adequate therapy for small solitary furuncles without cellulitis or systemic symptoms. Whenever possible, lesional material should be sent for culture. Current recommendations for methicillin-sensitive infections are β-lactam antibiotics. Therapy for CA-MRSA was discussed previously (see "Atopic Dermatitis"). Warm compresses and nasal mupirocin are helpful therapeutic additions. Severe infections may require IV antibiotics.

Erysipelas and Cellulitis

See Chap. 125.

Dermatophytosis

Dermatophytes are fungi that infect skin, hair, and nails and include members of the genera *Trichophyton*, *Microsporum*, and *Epidermophyton*. Tinea corporis, or infection of the relatively hairless skin of the body (glabrous skin), may have a variable appearance depending on the extent of the associated inflammatory reaction. Typical infections have an annular appearance that patients refer to as "ringworm." Deep inflammatory nodules or granulomas occur in some infections—especially in those infections inappropriately treated with mid- to high-potency topical glucocorticoids. Involvement of the groin (tinea cruris) is more common in males than females. It presents as a scaling, erythematous eruption sparing the scrotum. Infection of the foot (tinea pedis) is the most common dermatophyte infection and is often chronic; it is characterized by variable erythema, edema, scaling, pruritus, and, occasionally, vesiculation. Involvement may be widespread or localized but generally involves the web space between the fourth and fifth toes. Infection of the nails (tinea unguium or onychomycosis) occurs in many patients with tinea pedis and is characterized by opacified, thickened nails and subungual debris. The distal-lateral variant is most common. Proximal subungual onychomycosis may be a marker for HIV infection or other immunocompromised states. Dermatophyte infection of the scalp (tinea capitis) continues to be common, particularly affecting inner-city children, but it also affects adults. The predominant organism is *T. tonsurans*, which can produce a relatively noninflammatory infection with mild scale and hair loss that is diffuse or localized. *T. tonsurans* can also cause a markedly inflammatory dermatosis with edema and nodules. This latter presentation is a kerion.

The diagnosis of tinea can be made from skin scrapings, nail scrapings, or hair by culture or direct microscopic examination with potassium hydroxide (KOH). Nail clippings may be sent for histologic examination with periodic acid-Schiff (PAS) stain.

TREATMENT: DERMATOPHYTOSIS

Both topical and systemic therapies may be used to treat dermatophyte infections. Treatment depends on the site involved and the type of infection. Topical therapy is generally effective for uncomplicated tinea corporis, tinea cruris, and limited tinea pedis. It is not effective as a monotherapy for tinea capitis or onychomycosis. Topical imidazoles, triazoles, and allylamines may be effective therapies for dermatophyte infections, but nystatin is not active against dermatophytes. Topicals are generally applied twice daily, and treatment should continue 1 week beyond clinical resolution of the infection. Tinea pedis often requires longer treatment courses and it frequently relapses. Oral antifungal agents may be required for recalcitrant tinea pedis or tinea corporis.

Oral antifungal agents are required for dermatophyte infections involving the hair and nails and for other infections unresponsive to topical therapy. A fungal etiology should be confirmed by direct microscopic examination or by culture prior to prescribing oral antifungal agents. All of the oral agents may cause hepatotoxicity and should not be used in women who are pregnant or breast-feeding.

Griseofulvin is the only oral agent approved in the United States for dermatophyte infections involving the skin, hair, or nails. When griseofulvin is used, a daily dose of 500 mg microsized or 375 mg ultramicrosized griseofulvin administered with a fatty meal is an adequate dose for most dermatophyte infections. Higher doses are required for some cases of tinea pedis and tinea capitis. Markedly inflammatory tinea capitis may result in scarring and hair loss, and systemic or topical glucocorticoids may be helpful in preventing these sequelae. The duration of anti-fungal therapy may be 2 weeks for uncomplicated tinea corporis, 8–12 weeks for tinea capitis, or as long as 6–18 months for nail infections. Due to high relapse rates, griseofulvin is seldom used for nail infections. Common side effects of griseofulvin include gastrointestinal distress, headache, and urticaria.

Oral itraconazole and terbinafine are approved for onychomycosis. Itraconazole is given as either continuous daily therapy (200 mg/d) or pulses (200 mg bid for 1 week per month) administered with food. Fingernails require 2 months of continuous therapy or two pulses. Toenails require 3 months of

continuous therapy or three pulses. Itraconazole has the potential for serious interactions with other drugs requiring the P450 enzyme system for metabolism. Terbinafine (250 mg/d) is also effective for onychomycosis. Therapy with terbinafine is continued for 6 weeks for fingernail infections and 12 weeks for toenail infections. Terbinafine has fewer drug-drug interactions, but caution should be used with patients who are on multiple medications.

Tinea Versicolor

Tinea versicolor is caused by a nondermatophyte, dimorphic fungus, *Malassezia furfur*, a normal inhabitant of the skin. The expression of infection is promoted by heat and humidity. The typical lesions consist of oval scaly macules, papules, and patches concentrated on the chest, shoulders, and back, but only rarely on the face or distal extremities. On dark skin, they often appear as hypopigmented areas, while on light skin, they are slightly erythematous or hyperpigmented. A KOH preparation from scaling lesions will demonstrate a confluence of short hyphae and round spores ("spaghetti and meatballs"). Lotions or shampoos containing sulfur, salicylic acid, or selenium sulfide will clear the infection if used daily for 1 to 2 weeks and then weekly thereafter. These preparations are irritating if left on the skin for more than 10 minutes; thus, they should be washed off completely. Treatment with some oral antifungal agents is also effective, but they do not provide lasting results and they are not FDA-approved for this indication. Ketoconazole has been used as have itraconazole and fluconazole. Griseofulvin is not effective, and terbinafine is not reliably effective for tinea versicolor.

Candidiasis

Candidiasis is a fungal infection caused by a related group of yeasts, whose manifestations may be localized to the skin, or, rarely, may be systemic and life-threatening. The causative organism is usually *Candida albicans*, but may also be *C. tropicalis*, *C. parapsilosis*, or *C. krusei*. These organisms are normal saprophytic inhabitants of the gastrointestinal tract but may overgrow (usually due to broad-spectrum antibiotic therapy) and cause disease at a number of cutaneous sites. Other predisposing factors include diabetes mellitus, chronic intertrigo, oral contraceptive use, and cellular immune deficiency. Candidiasis is a very common infection in HIV-infected individuals (Chap. 189). The oral cavity is commonly involved. Lesions may occur on the tongue or buccal mucosa (thrush) and appear as white plaques. Microscopic examination of scrapings demonstrates both pseudohyphae and yeast forms. Fissured, macerated lesions at the corners of the mouth (perléche) are often seen in individuals with poorly fitting dentures and may also be associated with candidal infection. In addition, candidal infections have an affinity for sites that are chronically wet and macerated, including the skin around nails (onycholysis and paronychia) and in intertriginous areas. Intertriginous lesions are characteristically edematous, erythematous, and scaly, with scattered "satellite pustules." In males, there is often involvement of the penis and scrotum as well as the inner aspect of the thighs. In contrast to dermatophyte infections, candidal infections are frequently painful and accompanied by a marked inflammatory response. Diagnosis of candidal infection is based upon the clinical pattern and demonstration of yeast on KOH preparation or culture.

TREATMENT: CANDIDIASIS

Treatment involves removing any predisposing factors such as antibiotic therapy or chronic wetness and the use of appropriate topical or systemic antifungal agents. Effective topicals include nystatin or azoles (miconazole, clotrimazole, econazole, or ketoconazole). The associated inflammatory response accompanying candidal infection on glabrous skin can be treated with a mild glucocorticoid lotion or cream (2.5% hydrocortisone). Systemic therapy is usually reserved for immunosuppressed patients or individuals with chronic or recurrent disease who fail to respond to appropriate topical therapy. Oral agents approved for the treatment of candidiasis include itraconazole and fluconazole. Oral nystatin is only effective for candidiasis of the gastrointestinal tract. Griseofulvin and terbinafine are not effective.

Warts

Warts are cutaneousneoplasms caused by papilloma viruses. More than 100 different human papilloma viruses (HPV) have been described. A typical wart, verruca vulgaris, is sessile, dome-shaped, and usually about a centimeter in diameter. Its surface is hyperkeratotic consisting of many small filamentous projections. The HPV that cause typical verruca vulgaris also cause typical plantar warts, flat warts (or verruca plana), and filiform warts. Plantar warts are endophytic and are covered by thick keratin. Paring of the wart will generally demonstrate a central core of keratinized debris and punctate bleeding points. Filiform warts are most commonly seen on the face, neck, and skinfolds and present as papillomatous lesions on a narrow base. Flat warts are only slightly elevated and have a velvety, nonverrucous surface. They have a propensity for the face, arms, and legs and are often spread by shaving.

Genital warts begin as small papillomas that may grow to form large, fungating lesions. In women, they may involve either the labia, perineum, or perianal skin. In addition, the mucosa of the vagina, urethra, and anus can be involved as well as the cervical epithelium. In men, the lesions often occur initially in the coronal sulcus but may be seen on the shaft of the penis, the scrotum, perianal skin, or in the urethra.

Appreciable evidence has accumulated that suggests HPV plays a role in the development of

neoplasia of the uterine cervix and anogenital skin (Chap. 97). HPV types 16 and 18 have been most intensely studied and are the major risk factors for intraepithelial neoplasia and squamous cell carcinoma of the cervix, anus, vulva, and penis. The risk is higher in patients immunosuppressed after solid organ transplantation and in those infected with HIV. Recent evidence also implicates other types. Histologic examination of biopsies from affected sites may reveal changes associated with typical warts and/or features typical of intraepidermal carcinoma (Bowen's disease). Squamous cell carcinomas associated with HPV infections have also been observed in extragenital skin (Chap. 87). This is most commonly seen in patients immunosuppressed after organ transplantation. Patients on long-term immunosuppression should be monitored for the development of squamous cell carcinoma and other cutaneous malignancies.

TREATMENT: WARTS

Treatment of warts, other than anogenital warts, should be tempered by the observation that a majority of warts in normal individuals resolve spontaneously within 1 to 2 years. There are many modalities available to treat warts, but no single therapy is universally effective. Factors that influence the choice of therapy include the location of the wart, the extent of disease, the age and immunologic status of the patient, and the patient's desire for therapy. Perhaps the most useful and convenient method for treating warts in almost any location is cryotherapy with liquid nitrogen. Equally effective for nongenital warts, but requiring much more patient compliance, is the use of keratolytic agents such as salicylic acid plasters or solutions. For genital warts, in-office application of a podophyllin solution is moderately effective but may be associated with marked local reactions. Prescription preparations of dilute, purified podophyllin are available for home use. Topical imiquimod, a potent inducer of local cytokine release, has also been approved for use in genital warts. Conventional and laser surgical procedures may be required for recalcitrant warts. Recurrence of warts appears to be common to all these modalities. A highly effective vaccine for selected types of HPV has been recently approved by the FDA, and its use will likely reduce the incidence of anogenital and cervical carcinoma.

Herpes Simplex

See Chap. 179.

Herpes Zoster

See Chap. 180.

ACNE

Acne Vulgaris

Acnevulgaris is a self-limited disorder primarily of teenagers and young adults, although perhaps 10–20% of adults may continue to experience some form of the disorder. The permissive factor for the expression of the disease in adolescence is the increase in sebum production by sebaceous glands after puberty. Small cysts, called *comedones*, form in hair follicles due to blockage of the follicular orifice by retention of keratinous material and sebum. The activity of bacteria (*Propionibacterium acnes*) within the comedones releases free fatty acids from sebum, causes inflammation within the cyst, and results in rupture of the cyst wall. An inflammatory foreign-body reaction develops as result of extrusion of oily and keratinous debris from the cyst.

The clinical hallmark of acnevulgaris is the comedone, which may be closed (whitehead) or open (blackhead). Closed comedones appear as 1- to 2-mm pebbly white papules, which are accentuated when the skin is stretched. They are the precursors of inflammatory lesions of acne vulgaris. The contents of closed comedones are not easily expressed. Open comedones, which rarely result in inflammatory acne lesions, have a large dilated follicular orifice and are filled with easily expressible oxidized, darkened, oily debris. Comedones are usually accompanied by inflammatory lesions: papules, pustules, or nodules.

The earliest lesions seen in adolescence are generally mildly inflamed or noninflammatory comedones on the forehead. Subsequently, more typical inflammatory lesions develop on the cheeks, nose, and chin **(Fig. 52-7)**. The most common location for acne is the face, but involvement of the chest and back is common. Most disease remains mild and does not lead to scarring. A small number of patients develop large inflammatory cysts and nodules, which may drain and result in significant scarring. Regardless of the severity, acne may affect a patient's quality of life. If adequately treated, this may be a transient effect. In the case of severe, scarring acne, the effects can be permanent and profound. Early therapeutic intervention in severe acne is essential.

Figure 52-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acnevulgaris. An example of acne vulgaris with inflammatory papules, pustules, and comedones. (*Courtesy of Kalman Watsky, MD; with permission* .)

Exogenous and endogenous factors can alter the expression of acnevulgaris. Friction and trauma (from headbands or chin straps of athletic helmets), application of comedogenic topical agents (cosmetics or hair preparations), or chronic topical exposure to certain industrial compounds may elicit or aggravate acne. Glucocorticoids, topical or systemic, may also elicit acne. Other systemic medications such as oral contraceptive pills, lithium, isoniazid, androgenic steroids, halogens, phenytoin, and phenobarbital may produce acneiform eruptions or aggravate preexisting acne. Genetic factors and polycystic ovary disease may also play a role.

TREATMENT: ACNE VULGARIS

Treatment of acnevulgaris is directed toward elimination of comedones by normalization of follicular keratinization, decreasing sebaceous gland activity, decreasing the population of *P. acnes*, and decreasing inflammation. Minimal to moderate pauci-inflammatory disease may respond adequately to local therapy alone. Although areas affected with acne should be kept clean, overly vigorous scrubbing may aggravate acne due to mechanical rupture of comedones. Topical agents such as retinoic acid, benzoyl peroxide, or salicylic acid may alter the pattern of epidermal desquamation, preventing the formation of comedones and aiding in the resolution of preexisting cysts. Topical antibacterial agents such as azelaic acid, topical erythromycin, or clindamycin are also useful adjuncts to therapy.

Patients with moderate to severe acne with a prominent inflammatory component will benefit from the addition of systemic therapy, such as tetracycline in doses of 250–500 mg bid or doxycycline, 100 mg bid. Minocycline is also useful. Such antibiotics appear to have anti-inflammatory effects independent of their antibacterial effects. Female patients who do not respond to oral antibiotics may benefit from hormonal therapy. Several oral contraceptives are now approved by the FDA for use in the treatment of acne vulgaris.

Patients with severe nodulocystic acne unresponsive to the therapies discussed above may benefit from treatment with the synthetic retinoid isotretinoin. Its dose is based on the patient's weight, and it is given once daily for 5 months. Results are excellent in appropriately selected patients. Its use is highly regulated due to its potential for severe adverse events, primarily teratogenicity. In addition, patients receiving this medication develop extremely dry skin, cheilitis, and must be followed for development of hypertriglyceridemia.

Recently, there have also been concerns that isotretinoin is associated with severe depression in some patients. At present, prescribers must enroll in a program designed to prevent pregnancy and adverse events while patients are taking isotretinoin. These measures are imposed to ensure that all prescribers are familiar with the risks of isotretinoin; that all female patients have two negative pregnancy tests prior to initiating therapy and a negative pregnancy test prior to each refill; and that all patients have been warned about the risks associated with isotretinoin.

Acne Rosacea

Acne rosacea, commonly referred to as rosacea, is an inflammatory disorder predominantly affecting the central face. Those most often affected are Caucasians of northern European background, but it is seen in patients with dark skin also. It is seen almost exclusively in adults, only rarely affecting patients <30 years. Rosacea is more common in women, but those most severely affected are men. It is characterized by the presence of erythema, telangiectases, and superficial pustules **(Fig. 52-8)**, but is not associated with the presence of comedones. Rosacea rarely involves the chest or back.

Figure 52-8

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acne rosacea. Prominent facial erythema, telangiectasia, scattered papules, and small pustules are seen in this patient with acne rosacea. (*Courtesy of Robert Swerlick, MD; with permission* .)

There is a relationship between the tendency for facial flushing and the subsequent development of acne rosacea. Often, individuals with rosacea initially demonstrate a pronounced flushing reaction. This may be in response to heat, emotional stimuli, alcohol, hot drinks, or spicy foods. As the disease progresses, the flush persists longer and longer and may eventually become permanent. Papules, pustules, and telangiectases can become superimposed on the persistent flush. Rosacea of very long standing may lead to connective tissue overgrowth, particularly of the nose (rhinophyma). Rosacea may also be complicated by various inflammatory disorders of the eye, including keratitis, blepharitis, iritis, and recurrent chalazion. These ocular problems are potentially sight-threatening and warrant ophthalmologic evaluation.

TREATMENT: ACNE ROSACEA

Acne rosacea can be treated topically or systemically. Mild disease often responds to topical metronidazole or sodium sulfacetamide. More severe disease requires oral tetracyclines: tetracycline 250–500 mg bid, doxycycline 100 mg bid, or minocycline 50–100 mg bid. Residual telangiectasia may respond to laser therapy. Topical glucocorticoids, especially potent agents, should be avoided because chronic use of these preparations may elicit rosacea. Topical therapy of the skin is not effective treatment for ocular disease.

SKIN DISEASES AND SMALLPOX VACCINATION

Because of a higher incidence of adverse events associated with smallpox vaccination in patients with a history of certain skin diseases, including atopic dermatitis, eczema, and psoriasis, such vaccination is contraindicated in patients with these conditions in the absence of a bioterrorism attack and a real or potential exposure to smallpox. In the case of such exposure, the risk of smallpox infection outweighs the risk of adverse events from the vaccine (Chap. 221).

FURTHER READINGS

Bolognia JL, Jorizzo JL, Rapini RP (eds): *Dermatology*, 2nd ed. Philadelphia, Mosby, 2009

James WD et al: *Andrews Diseases of the Skin Clinical Dermatology,* 10th ed. Philadelphia, Saunders-Elsevier, 2006

Wolff K, Goldsmith LA, Katz SI, et al (eds): *Fitzpatricks Dermatology in General Medicine,* 7th ed. New York, McGraw Hill, 2008

Wolff K, Johnson RA: *Fitzpatricks Color Atlas and Synopsis of Clinical Dermatology,* 6th ed. New York, McGraw-Hill, 2009

Wolverton SE (ed): *Comprehensive Dermatologic Drug Therapy* 2nd ed. Philadelphia, Saunders, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SKIN MANIFESTATIONS OF INTERNAL DISEASE: INTRODUCTION

It is a generally accepted concept in medicine that the skin can develop signs of internal disease. Therefore, in textbooks of medicine, one finds a chapter describing in detail the major systemic disorders that can be identified by cutaneous signs. The underlying assumption of such a chapter is that the clinician has been able to identify the specific disorder in the patient and needs only to read about it in the textbook. In reality, concise differential diagnoses and the identification of these disorders are actually difficult for the nondermatologist because he or she is not well-versed in the recognition of cutaneous lesions or their spectrum of presentations. Therefore, this chapter covers this particular topic of cutaneous medicine not by discussing individual disorders, but by describing and discussing the various presenting clinical signs and symptoms that indicate the presence of these disorders. Concise differential diagnoses will be generated in which the significant diseases will be briefly discussed and distinguished from the more common disorders that have minimal or no significance for internal diseases. The latter disorders are reviewed in table form and always need to be excluded when considering the former. For a detailed description of individual diseases, the reader should consult a dermatologic text.

PAPULOSQUAMOUS SKIN LESIONS

(Table 53-1) When an eruption is characterized by elevated lesions, papules (<1 cm), or plaques (>1 cm), in association with scale, it is referred to as a *papulosquamous lesion*. The most common papulosquamous diseases—*psoriasis*, *tinea*, *pityriasis rosea*, and *lichen planus*—are primary cutaneous disorders (Chap. 52). When psoriatic lesions are accompanied by arthritis, the possibility of psoriatic arthritis or reactive arthritis (formerly known as Reiter's syndrome) should be considered. A history of oral ulcers, conjunctivitis, uveitis, and/or urethritis points to the latter diagnosis. Lithium, beta blockers, HIV or streptococcal infections, and a rapid taper of systemic glucocorticoids are known to exacerbate psoriasis. Emerging comorbidities in psoriasis include cardiovascular disease and metabolic syndrome.

Table 53-1 Selected Causes of Papulosquamous Skin Lesions

1. Primary cutaneous disorders
a. Psoriasis[a]
b. Tinea[a]
c. Pityriasis rosea[a]
d. Lichen planus[a]
e. Parapsoriasis
f. Bowen's disease (squamous cell carcinoma in situ)[b]
2. Drugs
3. Systemic diseases
a. Lupus erythematosus[c]
b. Cutaneous T cell lymphoma; in particular, mycosis fungoides[d]
c. Secondary syphilis
d. Reactive arthritis (formerly known as Reiter's syndrome)
e. Sarcoidosis[e]

[a] Discussed in detail in Chap. 52; cardiovascular disease and metabolic syndrome are emerging comorbidities in psoriasis; primarily in Europe, hepatitis C virus is associated with oral lichen planus.

[b] Associated with chronic sun exposure and exposure to arsenic.

[c] See also Red Lesions in "Papulonodular Skin Lesions."

[d] Also cutaneous lesions of HTLV-1-associated adult T cell leukemia/lymphoma.

[e] See also Red-Brown Lesions in "Papulonodular Skin Lesions."

Whenever the diagnosis of pityriasis rosea or lichen planus is made, it is important to review the patient's medications because the eruption can be treated by simply discontinuing the offending agent. Pityriasis rosea–like drug eruptions are seen most commonly with beta blockers, angiotensin-converting enzyme (ACE) inhibitors, and metronidazole, while the drugs that can produce a lichenoid eruption include thiazides, antimalarials, quinidine, beta blockers, and ACE inhibitors. In some populations, there is a higher prevalence of hepatitis C viral infection in patients with lichen planus. Lichen planus–like lesions are also observed in chronic graft-versus-host disease.

In its early stages, the mycosis fungoides (MF) form of *cutaneous T cell lymphoma* (CTCL) may be confused with eczema or psoriasis, but it often fails to respond to the appropriate therapy for those inflammatory diseases. MF can develop within lesions of large-plaque parapsoriasis and is suggested by an increase in the thickness of the lesions. The diagnosis of MF is established by skin biopsy in which collections of atypical T lymphocytes are found in the epidermis and dermis. As the disease progresses, cutaneous tumors and lymph node involvement may appear.

In *secondary syphilis*, there are scattered red-brown papules with thin scale. The eruption often involves the palms and soles and can resemble pityriasis rosea. Associated findings are helpful in making the diagnosis and include annular plaques on the face, nonscarring alopecia, condyloma lata (broad-based and moist), and mucous patches as well as lymphadenopathy, malaise, fever, headache, and myalgias. The interval between the primary chancre and the secondary stage is usually 4–8 weeks, and spontaneous resolution without appropriate therapy is seen.

ERYTHRODERMA

(Table 53-2)*Erythroderma* is the term used when the majority of the skin surface is erythematous (red in color). There may be associated scale, erosions, or pustules as well as shedding of the hair and nails. Potential systemic manifestations include fever, chills, hypothermia, reactive lymphadenopathy, peripheral edema, hypoalbuminemia, and high-output cardiac failure. The major etiologies of erythroderma are (1) cutaneous diseases such as psoriasis and dermatitis **(Table 53-3)**; (2) drugs; (3) systemic diseases, most commonly CTCL; and (4) idiopathic. In the first three groups, the location and description of the initial lesions, prior to the development of the erythroderma, aid in the diagnosis. For example, a history of red scaly plaques on the elbows and knees would point to psoriasis. It is also important to examine the skin carefully for a migration of the erythema and associated secondary changes such as pustules or erosions. Migratory waves of erythema studded with superficial pustules are seen in *pustular psoriasis*.

Table 53-2 Causes of Erythroderma

1. Primary cutaneous disorders
a. Psoriasis[a]
b. Dermatitis [atopic, contact >> stasis (with autosensitization) or seborrheic][a]
c. Pityriasis rubra pilaris
2. Drugs
3. Systemic diseases
a. Cutaneous T cell lymphoma
b. Lymphoma
4. Idiopathic
[a]Discussed in detail in Chap. 52.

Table 53-3 Erythroderma (Primary Cutaneous Disorders)

	Initial Lesions	Location of Initial Lesions	Other Findings	Diagnostic Aids	Treatment
Psoriasis[a]	Pink-red, silvery scale,	Elbows, knees, scalp,	Nail dystrophy, arthritis, pustules; SAPHO	Skin biopsy	Topical glucocorticoids,

	sharply demarcated	presacral area	syndrome, especially with palmoplantar pustulosis		vitamin D; UV-B (narrowband); oral retinoid and/or PUVA; MTX, cyclosporine, anti-TNF agents, anti-IL-12/23 Ab
Dermatitis[a]					
Atopic	Acute: Erythema, fine scale, crust, indistinct borders Chronic: Lichenification (increased skin markings)	Antecubital and popliteal fossae, neck, hands, eyelids	Pruritus Personal and/or family history of atopy, including asthma, allergic rhinitis or conjunctivitis, and atopic dermatitis Exclude secondary infection with *S. aureus* or HSV Exclude superimposed irritant or allergic contact dermatitis	Skin biopsy	Topical glucocorticoids, tacrolimus, pimecrolimus, tar, and antipruritics; oral antihistamines; open wet dressings; UV-B ± UV-A; PUVA; oral/IM glucocorticoids; MTX; mycophenolate mofetil; cyclosporine Topical or oral antibiotics
Contact	Local: Erythema, crusting, vesicles, and bullae	Depends on offending agent	Irritant—onset often within hours Allergic—delayed-type hypersensitivity; lag time of 48 h	Patch testing; open use test	Remove irritant or allergen; topical glucocorticoids; oral antihistamines; oral/IM glucocorticoids
	Systemic: Erythema, fine scale, crust	Generalized vs major intertriginous zones (especially groin)	Patient has history of allergic contact dermatitis to topical agent and then receives systemic medication that is structurally related, e.g., ethylenediamine (topical), aminophylline (IV)	Patch testing	Same as local
Seborrheic (rare) in adults	Pink-red, greasy scale	Scalp, nasolabial folds, eyebrows, intertriginous zones	Flares with stress, HIV infection Associated with Parkinson's disease	Skin biopsy	Topical glucocorticoids and imidazoles
Stasis (with autosensitization)	Erythema, crusting, excoriations	Lower extremities	Pruritus, lower extremity edema, varicosities, hemosiderin deposits History of venous ulcers, thrombophlebitis, and/or cellulitis Exclude cellulitis Exclude superimposed contact dermatitis, e.g., topical neomycin	Skin biopsy	Topical glucocorticoids; open wet dressings; leg elevation; pressure stockings
Pityriasis rubra pilaris	Orange-red, perifollicular papules	Generalized, but characteristic "skip" areas of normal skin	Wax-like palmoplantar keratoderma Exclude cutaneous T cell lymphoma	Skin biopsy	Isotretinoin or acitretin; MTX

[a]Discussed in detail in Chap. 52.

Note: Ab, antibody; HSV, herpes simplex virus; IL, interleukin; IM, intramuscular; MTX, methotrexate; PUVA, *p* soralens + *u*ltra*v*iolet *A* irradiation; SAPHO, *s*ynovitis, *a*cne, *p*ustulosis, *h*yperostosis and *o*steitis (also referred to as chronic recurrent multifocal osteomyelitis); TNF, tumor necrosis factor; UV-A, *u*ltra*v*iolet *A* irradiation; UV-B, *u*ltra*v*iolet *B* irradiation.

Drug-induced erythroderma (exfoliative dermatitis) may begin as an exanthematous (morbilliform) eruption (Chap. 55) or may arise as diffuse erythema. A number of drugs can produce an erythroderma, including penicillins, sulfonamides, carbamazepine, phenytoin, and allopurinol. Fever and peripheral eosinophilia often accompany the eruption, and there may also be facial swelling, hepatitis, myocarditis, and allergic interstitial nephritis; this constellation is frequently referred to as *drug reaction with eosinophilia and systemic symptoms* (DRESS). In addition, these reactions, especially to anticonvulsants, can lead to a pseudolymphoma syndrome (with adenopathy and circulating atypical lymphocytes), while reactions to allopurinol may be accompanied by gastrointestinal bleeding.

The most common malignancy that is associated with erythroderma is CTCL; in some series, up to 25% of the cases of erythroderma were due to CTCL. The patient may progress from isolated plaques and tumors, but, more commonly, the erythroderma is present throughout the course of the disease (Sézary syndrome). In the Sézary syndrome, there are circulating atypical T lymphocytes, pruritus, and lymphadenopathy. In cases of erythroderma where there is no apparent cause (idiopathic), longitudinal evaluation is mandatory to monitor for the possible development of CTCL. There have been isolated case reports of erythroderma secondary to some solid tumors—lung, liver, prostate, thyroid, and colon—but it is primarily during a late stage of the disease.

ALOPECIA

(Table 53-4) The two major forms of alopecia are scarring and nonscarring. In *scarring alopecia*, there are associated fibrosis, inflammation, and loss of hair follicles. A smooth scalp with a decreased number of follicular openings is usually observed clinically, but in some cases, the changes are seen only in biopsy specimens from affected areas. In *nonscarring alopecia*, the hair shafts are gone, but the hair follicles are preserved, explaining the reversible nature of nonscarring alopecia.

Table 53-4 Causes of Alopecia

I. Nonscarring alopecia
A. Primary cutaneous disorders
1. Telogen effluvium
2. Androgenetic alopecia
3. Alopecia areata
4. Tinea capitis
5. Traumatic alopecia[a]
B. Drugs
C. Systemic diseases
1. Systemic lupus erythematosus
2. Secondary syphilis
3. Hypothyroidism
4. Hyperthyroidism
5. Hypopituitarism
6. Deficiencies of protein, iron, biotin, and zinc
II. Scarring alopecia
A. Primary cutaneous disorders
1. Cutaneous lupus (chronic discoid lesions)[b]
2. Lichen planus
3. Central centrifugal cicatricial alopecia
4. Folliculitis decalvans
5. Linear scleroderma (morphea)
B. Systemic diseases
1. Discoid lesions in the setting of systemic lupus erythematosus[b]
2. Sarcoidosis
3. Cutaneous metastases

[a] Most patients with trichotillomania, pressure-induced alopecia, early stages of traction alopecia.

[b] While the majority of patients with discoid lesions have only cutaneous disease, these lesions do represent one of the 11 American College of Rheumatology criteria (1982) for systemic lupus erythematosus.

The most common causes of nonscarring alopecia include *telogen effluvium*, *androgenetic alopecia*, *alopecia areata*, *tinea capitis*, and the early phase of *traumatic alopecia***(Table 53-5)**. In women with androgenetic alopecia, an elevation in circulating levels of androgens may be seen as a result of ovarian or adrenal gland dysfunction. When there are signs of virilization, such as a deepened voice and enlarged clitoris, the possibility of an ovarian or adrenal gland tumor should be considered.

Table 53-5 Nonscarring Alopecia (Primary Cutaneous Disorders)

	Clinical Characteristics	Pathogenesis	Treatment
Telogen effluvium	Diffuse shedding of normal hairs Follows major stress (high fever, severe infection) or change in hormones (postpartum) Reversible without treatment	Stress causes more of the asynchronous growth cycles of individual hairs to become synchronous; therefore, larger numbers of growing (anagen) hairs simultaneously enter the dying (telogen) phase	Observation; discontinue any drugs that have alopecia as a side effect; must exclude underlying metabolic causes, e.g., hypothyroidism, hyperthyroidism
Androgenetic alopecia (male pattern; female pattern)	Miniaturization of hairs along the midline of the scalp Recession of the anterior scalp line in men and some women	Increased sensitivity of affected hairs to the effects of androgens Increased levels of circulating androgens (ovarian or adrenal source in women)	If no evidence of hyperandrogen state, then topical minoxidil; finasteride[a]; spironolactone (women); hair transplant
Alopecia areata	Well-circumscribed, circular areas of hair loss, 2–5 cm in diameter In extensive cases, coalescence of lesions and/or involvement of other hair-bearing surfaces of the body Pitting of the nails	The germinative zones of the hair follicles are surrounded by T lymphocytes Occasional associated diseases: hyperthyroidism, hypothyroidism, vitiligo, Down syndrome	Topical anthralin or tazarotene; intralesional glucocorticoids; topical contact sensitizers
Tinea capitis	Varies from scaling with minimal hair loss to discrete patches with "black dots" (broken hairs) to boggy plaque with pustules (kerion)[b]	Invasion of hairs by dermatophytes, most commonly *Trichophyton tonsurans*	Oral griseofulvin or terbinafine plus 2.5% selenium sulfide or ketoconazole shampoo; examine family members
Traumatic alopecia[c]	Broken hairs Irregular outline	Traction with curlers, rubber bands, braiding Exposure to heat or chemicals (e.g., hair straighteners) Mechanical pulling (trichotillomania)	Discontinuation of offending hair style or chemical treatments; trichotillomania may require hair clipping and observation of shaved hairs or biopsy for diagnosis, possibly followed by psychotherapy

[a] To date, FDA-approved for men.

[b] Scarring alopecia can occur at sites of kerions.

[c] May also be scarring, especially late-stage traction alopecia.

Exposure to various drugs can also cause diffuse hair loss, usually by inducing a telogen effluvium. An exception is the anagen effluvium observed with antimitotic agents such as daunorubicin. Alopecia is a side effect of the following drugs: warfarin, heparin, propylthiouracil, carbimazole, vitamin A, isotretinoin,

acitretin, lithium, beta blockers, colchicine, and amphetamines. Fortunately, spontaneous regrowth usually follows discontinuation of the offending agent.

Less commonly, nonscarring alopecia is associated with *lupus erythematosus* and *secondary syphilis.* In systemic lupus there are two forms of alopecia—one is scarring secondary to discoid lesions (see below), and the other is nonscarring. The latter form may be diffuse and involve the entire scalp or it may be localized to the frontal scalp, with the appearance of multiple short hairs ("lupus hairs") as a sign of initial regrowth. Scattered, poorly circumscribed patches of alopecia with a "moth-eaten" appearance are a manifestation of the secondary stage of syphilis. Diffuse thinning of the hair is also associated with hypothyroidism and hyperthyroidism (Table 53-4).

Scarring alopecia is more frequently the result of a primary cutaneous disorder such as *lichen planus*, *folliculitis decalvans*, *chroniccutaneous (discoid) lupus*, or *linear scleroderma* (*morphea*) than it is a sign of systemic disease. Although the scarring lesions of *discoid lupus* can be seen in patients with systemic lupus, in the majority of cases, the disease process is limited to the skin. Less common causes of scarring alopecia include *sarcoidosis* (see "Papulonodular Skin Lesions," below) and *cutaneous metastases*.

In the early phases of discoid lupus, lichen planus, and folliculitis decalvans, there are circumscribed areas of alopecia. Fibrosis and subsequent loss of hair follicles are observed primarily in the center of the individual lesions; the inflammatory process is most prominent at the periphery. The areas of active inflammation in discoid lupus are erythematous with scale, whereas the areas of previous inflammation are often hypopigmented with a rim of hyperpigmentation. In lichen planus, the peripheral perifollicular macules are usually violet-colored. A complete examination of the skin and oral mucosa combined with a biopsy and direct immunofluorescence microscopy will aid in distinguishing these two entities. The peripheral active lesions in folliculitis decalvans are follicular pustules; these patients can develop a reactive arthritis.

FIGURATE SKIN LESIONS

(Table 53-6) In *figurate eruptions*, the lesions form rings and arcs that are usually erythematous but can be skin-colored to brown. Most commonly, they are due to primary cutaneous diseases such as *tinea*, *urticaria*, *erythema annulare centrifugum*, and *granuloma annulare* (Chaps. 52 and 54). An underlying systemic illness is found in a second, less common group of migratory annular erythemas. It includes *erythema gyratum repens*, *erythema migrans*, *erythema marginatum*, and *necrolytic migratory erythema*.

Table 53-6 Causes of Figurate Skin Lesions

I. Primary cutaneous disorders
A. Tinea
B. Urticaria (primary in ≥90% of patients)
C. Erythema annulare centrifugum
D. Granuloma annulare
E. Psoriasis
II. Systemic diseases
A. Migratory
1. Erythema migrans
2. Urticaria (≤10% of patients)
3. Erythema gyratum repens
4. Erythema marginatum
5. Pustular psoriasis (generalized)
6. Necrolytic migratory erythema (glucagonoma syndrome)[a]
B. Nonmigratory
1. Sarcoidosis
2. Subacute cutaneous lupus erythematosus
3. Secondary syphilis
4. Cutaneous T cell lymphoma (especially mycosis fungoides)

[a] Migratory erythema with erosions; favors lower extremities and girdle area.

In erythema gyratum repens, one sees numerous mobile concentric arcs and wavefronts that resemble the grain in wood. A search for an underlying malignancy is mandatory in a patient with this eruption. Erythema migrans is the cutaneous manifestation of Lyme disease, which is caused by the

spirochete *Borrelia burgdorferi*. In the initial stage (3–30 days after tick bite), a single annular lesion is usually seen, which can expand to ≥10 cm in diameter. Within several days, approximately half the patients develop multiple smaller erythematous lesions at sites distant from the bite. Associated symptoms include fever, headache, photophobia, myalgias, arthralgias, and malar rash. Erythema marginatum is seen in patients with rheumatic fever, primarily on the trunk. Lesions are pink-red in color, flat to mildly elevated, and transient.

There are additional cutaneous diseases that present as annular eruptions but lack an obvious migratory component. Examples include *CTCL*, *subacute cutaneous lupus*, *secondary syphilis*, and *sarcoidosis* (see "Papulonodular Skin Lesions," below).

ACNE

(Table 53-7) In addition to *acne vulgaris* and *acne rosacea*, the two major forms of acne (Chap. 52), there are drugs and systemic diseases that can lead to acneiform eruptions.

Table 53-7 Causes of Acneiform Eruptions
I. Primary cutaneous disorders
A. Acne vulgaris
B. Acne rosacea
II. Drugs, e.g., anabolic steroids, glucocorticoids, lithium, EGFR[a] inhibitors, iodides
III. Systemic diseases
A. Increased androgen production
1. Adrenal origin, e.g., Cushing's disease, 21-hydroxylase deficiency
2. Ovarian origin, e.g., polycystic ovary syndrome
B. Cryptococcosis, disseminated
C. Dimorphic fungi
D. Behçet's disease

[a] EGFR, epidermal growth factor receptor.

Patients with the *carcinoid syndrome* have episodes of flushing of the head, neck, and sometimes the trunk. Resultant skin changes of the face, in particular telangiectasias, may mimic the clinical appearance of acne rosacea.

PUSTULAR LESIONS

Acneiform eruptions (see "Acne," above) and *folliculitis* represent the most common pustular dermatoses. An important consideration in the evaluation of follicular pustules is a determination of the associated pathogen, e.g., normal flora, *Staphylococcus aureus*, *Pseudomonas aeruginosa* ("hot tub" folliculitis), *Malassezia*, and dermatophytes (Majocchi's granuloma). Noninfectious forms of folliculitis include HIV-associated eosinophilic folliculitis and folliculitis secondary to drugs such as glucocorticoids, lithium, and epidermal growth factor receptor (EGFR) inhibitors. Administration of high-dose systemic glucocorticoids can result in a widespread eruption of follicular pustules on the trunk, characterized by lesions in the same stage of development. With regard to underlying systemic diseases, nonfollicular-based pustules are a characteristic component of pustular psoriasis (sterile) and can be seen in septic emboli of bacterial or fungal origin (see "Purpura," below). In patients with acute generalized exanthematous pustulosis (AGEP) due primarily to medications (e.g., cephalosporins), there are large areas of erythema studded with multiple sterile pustules in addition to neutrophilia.

TELANGIECTASIAS

(Table 53-8) To distinguish the various types of telangiectasias, it is important to examine the shape and configuration of the dilated blood vessels. *Linear telangiectasias* are seen on the face of patients with *actinically damaged skin* and *acne rosacea,* and they are found on the legs of patients with *venous hypertension* and *essential telangiectasia*. Patients with an unusual form of *mastocytosis* (telangiectasia macularis eruptiva perstans) and the *carcinoid syndrome* (see "Acne," above) also have linear telangiectasias. Lastly, linear telangiectasias are found in areas of cutaneous inflammation. For example, lesions of discoid lupus frequently have telangiectasias within them.

Table 53-8 Causes of Telangiectasias
I. Primary cutaneous disorders
A. Linear
1. Acne rosacea

2. Actinically damaged skin
3. Venous hypertension
4. Essential telangiectasia
5. Within basal cell carcinomas
B. Poikiloderma
1. Ionizing radiation[a] 2. Poikiloderma vasculare atrophicans
C. Spider angioma
1. Idiopathic
2. Pregnancy
II. Systemic diseases
A. Linear
1. Carcinoid
2. Ataxia-telangiectasia
3. Mastocytosis
B. Poikiloderma
1. Dermatomyositis
2. Cutaneous T cell lymphoma
3. Xeroderma pigmentosum
C. Mat
1. Scleroderma
D. Periungual
1. Lupus erythematosus
2. Scleroderma
3. Dermatomyositis
4. Hereditary hemorrhagic telangiectasia
E. Papular
1. Hereditary hemorrhagic telangiectasia
F. Spider angioma
1. Cirrhosis

[a]Becoming less common.

Poikiloderma is a term used to describe a patch of skin with: (1) reticulated hypo- and hyperpigmentation, (2) wrinkling secondary to epidermal atrophy, and (3) telangiectasias. Poikiloderma does not imply a single disease entity—although it is becoming less common, it is seen in skin damaged by *ionizing radiation* as well as in patients with autoimmune connective tissue diseases, primarily *dermatomyositis* (DM), and rare genodermatoses (e.g., Kindler syndrome).

In *systemic sclerosis (scleroderma)* the dilated blood vessels have a unique configuration and are known as *mat telangiectasias*. The lesions are broad macules that usually measure 2–7 mm in diameter but occasionally are larger. Mats have a polygonal or oval shape and their erythematous color may appear uniform, but, upon closer inspection, the erythema is the result of delicate telangiectasias. The most common locations for mat telangiectasias are the face, oral mucosa, and hands—peripheral sites that are prone to intermittent ischemia. The CREST (*c*alcinosis cutis, *R*aynaud's phenomenon, *e*sophageal dysmotility, *s*clerodactyly, and *t*elangiectasia) variant of scleroderma (Chap. 323) is associated with a chronic course and anticentromere antibodies. Mat telangiectasias are an important clue to the diagnosis of the CREST syndrome as well as systemic scleroderma because they may be the only cutaneous finding.

Periungual telangiectasias are pathognomonic signs of the three major autoimmune connective tissue diseases: *lupus erythematosus*, *scleroderma*, and *DM*. They are easily visualized by the naked eye and occur in at least two-thirds of these patients. In both DM and lupus, there is associated nailfold erythema, and in DM, the erythema is often accompanied by "ragged" cuticles and fingertip tenderness. Under 10x magnification, the blood vessels in the nailfolds of lupus patients are tortuous and resemble "glomeruli," whereas in scleroderma and DM, there is a loss of capillary loops and those that remain are markedly dilated.

In *hereditary hemorrhagic telangiectasia* (Osler-Rendu-Weber disease), the lesions usually appear during adulthood and are most commonly seen on the mucous membranes, face, and distal extremities, including under the nails. They represent arteriovenous (AV) malformations of the dermal

microvasculature, are dark red in color, and are usually slightly elevated. When the skin is stretched over an individual lesion, an eccentric punctum with radiating legs is seen. Although the degree of systemic involvement varies in this autosomal dominant disease (due primarily to mutations in either the endoglin or activin receptor–like kinase gene), the major symptoms are recurrent epistaxis and gastrointestinal bleeding. The fact that these mucosal telangiectasias are actually AV communications helps to explain their tendency to bleed.

HYPOPIGMENTATION

(Table 53-9) Disorders of hypopigmentation are often classified as either diffuse or localized. The classic example of *diffuse hypopigmentation* is *oculocutaneous albinism* (OCA). The most common forms are due to mutations in the tyrosinase gene (type I) or the *P* gene (type II); patients with type IA OCA have a total lack of enzyme activity. At birth, different forms of OCA can appear similar—white hair, gray-blue eyes, and pink-white skin. However, the patients with no tyrosinase activity maintain this phenotype, whereas those with decreased activity will acquire some pigmentation of the eyes, hair, and skin as they age. The degree of pigment formation is also a function of racial background, and the pigmentary dilution is more readily apparent when patients are compared to their first-degree relatives. The ocular findings in OCA correlate with the degree of hypopigmentation and include decreased visual acuity, nystagmus, photophobia, and a lack of normal binocular vision.

Table 53-9 Causes of Hypopigmentation

I. Primary cutaneous disorders
A. Diffuse
1. Generalized vitiligo[a]
B. Localized
1. Idiopathic guttate hypomelanosis
2. Postinflammatory
3. Tinea (pityriasis) versicolor
4. Vitiligo[a]
5. Chemical- or drug-enduced leukoderma
6. Nevus depigmentosus
7. Piebaldism
II. Systemic diseases
A. Diffuse
1. Oculocutaneous albinism[b]
2. Hermansky-Pudlak syndrome[b,c]
3. Chédiak-Higashi syndrome[b,d]
4. Phenylketonuria
5. Homocystinuria
B. Localized
1. Scleroderma
2. Melanoma-associated leukoderma
3. Sarcoidosis
4. Cutaneous T cell lymphoma (especially mycosis fungoides)
5. Tuberculoid and indeterminate leprosy
6. Onchocerciasis
7. Tuberous sclerosis
8. Hypomelanosis of Ito/mosaicism/linear nevoid hypopigmentation
9. Incontinentia pigmenti (stage IV)
10. Vogt-Koyanagi-Harada syndrome
11. Waardenburg syndrome

[a] Absence of melanocytes.

[b] Normal number of melanocytes.

[c] Platelet storage defect and restrictive lung disease secondary to deposits of ceroid-like material; due to mutations in ß subunit of adaptor protein 3 as well as subunits of *b*iogenesis of *l*ysosome-related *o*rganelles *c*omplex (BLOC)-1, -2, and -3.

[d] Giant lysosomal granules and recurrent infections.

The differential diagnosis of *localized hypomelanosis* includes the following primary cutaneous disorders: *idiopathic guttate hypomelanosis*, *postinflammatory hypopigmentation*, *tinea (pityriasis) versicolor*, *vitiligo*, *chemical- or drug-induced leukoderma*, *nevus depigmentosus* (see below), and *piebaldism* (Table 53-9). In this group of diseases, the areas of involvement are macules or patches with a decrease or absence of pigmentation. Patients with vitiligo also have an increased incidence of several autoimmune disorders, including Hashimoto's thyroiditis, Graves' disease, pernicious anemia, Addison's disease, uveitis, alopecia areata, chronic mucocutaneous candidiasis, and the polyglandular autoimmune syndromes (types I and II). Diseases of the thyroid gland are the most frequently associated disorders, occurring in up to 30% of patients with vitiligo. Circulating autoantibodies are often found, and the most common ones are antithyroglobulin, antimicrosomal, and antithyroid-stimulating hormone receptor antibodies.

There are four systemic diseases that should be considered in a patient with skin findings suggestive of vitiligo—*Vogt-Koyanagi- Harada syndrome*, *scleroderma*, *onchocerciasis*, and *melanoma-associated leukoderma*. A history of aseptic meningitis, nontraumatic uveitis, tinnitus, hearing loss, and/or dysacousia points to the diagnosis of the Vogt-Koyanagi-Harada syndrome. In these patients, the face and scalp are the most common locations of pigment loss. The vitiligo-like leukoderma seen in patients with scleroderma has a clinical resemblance to idiopathic vitiligo that has begun to repigment as a result of treatment; that is, perifollicular macules of normal pigmentation are seen within areas of depigmentation. The basis of this leukoderma is unknown; there is no evidence of inflammation in areas of involvement, but it can resolve if the underlying connective tissue disease becomes inactive. In contrast to idiopathic vitiligo, melanoma-associated leukoderma often begins on the trunk, and its appearance should prompt a search for metastatic disease. It is also seen in patients undergoing immunotherapy for melanoma, with cytotoxic T lymphocytes presumably recognizing cell surface antigens common to melanoma cells and melanocytes, and is associated with a greater likelihood of a clinical response.

There are two systemic disorders (neurocristopathies) that may have the cutaneous findings of piebaldism **(Table 53-10)**. They are *Shah-Waardenburg syndrome* and *Waardenburg syndrome*. A possible explanation for both disorders is an abnormal embryonic migration or survival of two neural crest–derived elements, one of them being melanocytes and the other myenteric ganglion cells (leading to Hirschsprung disease in Shah-Waardenburg syndrome) or auditory nerve cells (Waardenburg syndrome). The latter syndrome is characterized by congenital sensorineural hearing loss, dystopia canthorum (lateral displacement of the inner canthi but normal interpupillary distance), heterochromic irises, and a broad nasal root, in addition to the piebaldism. Patients with Waardenburg syndrome have been shown to have mutations in three genes, including two (*PAX-3* and *MITF*) that encode DNA-binding proteins, while patients with Hirschsprung disease plus white spotting have mutations in one of three genes—endothelin 3, endothelin B receptor, and *SOX-10*.

Table 53-10 Hypopigmentation (Primary Cutaneous Disorders, Localized)

	Clinical Characteristics	Wood's Lamp Examination (UV-A; Peak = 365 nm)	Skin Biopsy Specimen	Pathogenesis	Treatment
Idiopathic guttate hypomelanosis	Common; acquired; 1–4 mm in diameter Shins and extensor forearms	Less enhancement than vitiligo	Abrupt decrease in epidermal melanin content	Possible somatic mutations as a reflection of aging; UV exposure	None
Postinflammatory hypopigmentation	Can develop within active lesions, as in subacute cutaneous lupus, or after the lesion fades, as in dermatitis	Depends on particular disease Usually less enhancement than in vitiligo	Type of inflammatory infiltrate depends on specific disease	Block in transfer of melanin from melanocytes to keratinocytes could be secondary to edema or decrease in contact time Destruction of melanocytes if	Treat underlying inflammatory disease

				inflammatory cells attack basal layer of epidermis	
Pityriasis (tinea) versicolor	Common disorder Upper trunk and neck (shawl-like distribution), groin Young adults Macules have fine white scale when scratched	Golden fluorescence	Hyphal forms and budding yeast in stratum corneum	Invasion of stratum corneum by the yeast *Malassezia* Yeast is lipophilic and produces C_9 and C_{11} dicarboxylic acids, which in vitro inhibit tyrosinase	Selenium sulfide 2.5%; topical imidazoles; oral imidazoles or triazoles
Vitiligo	Acquired; progressive Symmetric areas of complete pigment loss Periorificial—around mouth, nose, eyes, nipples, umbilicus, anus Other areas—flexor wrists, extensor distal extremities Segmental form is less common—unilateral, dermatomal-like	More apparent Chalk-white	Absence of melanocytes Mild inflammation	Autoimmune phenomenon that results in destruction of melanocytes—primarily cellular (circulating skin-homing autoreactive T cells)	Topical glucocorticoids; topical calcineurin inhibitors; NBUV-B; PUVA; transplants, if stable; depigmentation (topical MBEH), if widespread
Chemical- or drug-induced leukoderma	Similar appearance to vitiligo Often begins on hands when associated with chemical exposure Satellite lesions in areas not exposed to chemicals	More apparent Chalk-white	Decreased number or absence of melanocytes	Exposure to chemicals that selectively destroy melanocytes, in particular phenols and catechols (germicides; adhesives) or ingestion of drugs such as imatinib Release of cellular antigens and activation of circulating lymphocytes may explain satellite phenomenon Possible inhibition of KIT receptor	Avoid exposure to offending agent, then treat as vitiligo Drug-induced variant may undergo repigmentation when medication is discontinued
Piebaldism	Autosomal dominant Congenital, stable White forelock Areas of hypomelanosis contain normally pigmented and hyperpigmented macules of various	Enhancement of leukoderma and hyperpigmented macules	Hypomelanotic areas—few to no melanocytes	Defect in migration of melanoblasts from neural crest to involved skin or failure of melanoblasts to survive or differentiate in	None; occasionally transplants

	sizes Symmetric involvement of central forehead, ventral trunk, and mid regions of upper and lower extremities			these areas Mutations within the c-*kit* protooncogene that encodes the tyrosine kinase receptor for stem cell growth factor (kit ligand)	

Abbreviations: MBEH, monobenzylether of hydroquinone; NBUV-B, *n*arrow*b* and *u*ltra*v*iolet *B*; PUVA, *p*soralens +*u*ltra*v*iolet *A* irradiation.

In *tuberous sclerosis*, the earliest cutaneous sign is an ash leaf spot. These lesions are often present at birth and are usually multiple; however, detection may require Wood's lamp examination, especially in fair-skinned individuals. The pigment within them is reduced, but not absent. The average size is 1–3 cm, and the common shapes are polygonal and lance-ovate. Examination of the patient for additional cutaneous signs such as multiple angiofibromas of the face (adenoma sebaceum), ungual and gingival fibromas, fibrous plaques of the forehead, and connective tissue nevi (shagreen patches) is recommended. It is important to remember that an ash leaf spot on the scalp will result in a circumscribed patch of lightly pigmented hair. Internal manifestations include seizures, mental retardation, central nervous system (CNS) and retinal hamartomas, pulmonary lymphangioleiomyomatosis (women), renal angiomyolipomas, and cardiac rhabdomyomas. The latter can be detected in up to 60% of children (<18 years) with tuberous sclerosis by echocardiography.

Nevus depigmentosus is a stable, well-circumscribed hypomelanosis that is present at birth. There is usually a single oval or rectangular lesion, but when there are multiple lesions, the possibility of tuberous sclerosis needs to be considered. In *linear nevoid hypopigmentation* or *pigmentary mosaicism*, terms that are replacing hypomelanosis of Ito and segmental or systematized nevus depigmentosus, streaks and swirls of hypopigmentation are observed. Up to a third of patients in a referral population had associated abnormalities involving the musculoskeletal system (asymmetry), the CNS (seizures and mental retardation), and the eyes (strabismus and hypertelorism). Chromosomal mosaicism has been detected in these patients, lending support to the hypothesis that the pattern is the result of the migration of two clones of primordial melanocytes, each with a different pigment potential.

Localized areas of decreased pigmentation are commonly seen as a result of cutaneous inflammation (Table 53-10) and have been observed in the skin overlying active lesions of sarcoidosis (see "Papulonodular Skin Lesions," below) as well as in CTCL. Cutaneous infections also present as disorders of hypopigmentation, and in *tuberculoid leprosy*, there are a few asymmetric patches of hypomelanosis that have associated anesthesia, anhidrosis, and alopecia. Biopsy specimens of the palpable border show dermal granulomas that contain rare, if any, *Mycobacterium leprae* organisms.

HYPERPIGMENTATION

(Table 53-11) Disorders of hyperpigmentation are also divided into two groups—localized and diffuse. The localized forms are due to an epidermal alteration, a proliferation of melanocytes, or an increase in pigment production. Both seborrheic keratoses and acanthosis nigricans belong to the first group. *Seborrheic keratoses* are common lesions, but in one rare clinical setting, they are a sign of systemic disease, and that setting is the sudden appearance of multiple lesions, often with an inflammatory base and in association with acrochordons (skin tags) and acanthosis nigricans. This is termed the *sign of Leser-Trélat* and alerts the clinician to search for an internal malignancy. *Acanthosis nigricans* can also be a reflection of an internal malignancy, most commonly of the gastrointestinal tract, and it appears as velvety hyperpigmentation, primarily in flexural areas. However, in the majority of patients, acanthosis nigricans is associated with obesity and insulin resistance, but it may be a reflection of an endocrinopathy such as acromegaly, Cushing's syndrome, polycystic ovary syndrome, or insulin-resistant diabetes mellitus (type A, type B, and lipoatrophic forms).

Table 53-11 Causes of Hyperpigmentation

I. Primary cutaneous disorders
A. Localized
1. Epidermal alteration
a. Seborrheic keratosis
b. Pigmented actinic keratosis
2. Proliferation of melanocytes
a. Lentigo
b. Melanocytic nevus (mole)

c. Melanoma
3. Increased pigment production
a. Ephelide (freckle)
b. Café au lait macule
c. Postinflammatory hyperpigmentation
B. Localized and diffuse
1. Drugs
II. Systemic diseases
A. Localized
1. Epidermal alteration
a. Seborrheic keratoses (sign of Leser-Trélat)
b. Acanthosis nigricans (insulin resistance, other endocrine disorders, paraneoplastic)
2. Proliferation of melanocytes
a. Lentigines (Peutz-Jeghers and LEOPARD syndromes; xeroderma pigmentosum)
b. Melanocytic nevi [Carney complex (LAMB and NAME syndromes)][a]
3. Increased pigment production
a. Café au lait macules (neurofibromatosis, McCune-Albright syndrome[b])
b. Urticaria pigmentosa[c]
4. Dermal pigmentation
a. Incontinentia pigmenti (stage III)
b. Dyskeratosis congenita
B. Diffuse
1. Endocrinopathies
a. Addison's disease
b. Nelson syndrome
c. Ectopic ACTH syndrome
2. Metabolic
a. Porphyria cutanea tarda
b. Hemochromatosis
c. Vitamin B12 , folate deficiency
d. Pellagra
e. Malabsorption, including Whipple's disease
3. Melanosis secondary to metastatic melanoma
4. Autoimmune
a. Biliary cirrhosis
b. Scleroderma
c. POEMS syndrome
d. Eosinophilia-myalgia syndrome[d]
5. Drugs and metals

[a] Also lentigines.

[b] Polyostotic fibrous dysplasia.

[c] See also "Papulonodular Skin Lesions."[d] Late 1980s.

Abbreviations: LAMB, *l*entigines, *a*trial myxomas, *m*ucocutaneous myxomas, and *b*lue nevi; LEOPARD, *l*entigines, *E*CG abnormalities, *o*cular hypertelorism, *p*ulmonary stenosis and subaortic valvular stenosis, *a*bnormal genitalia, *r* etardation of growth, and *d*eafness (sensorineural); NAME, *n*evi, *a*trial myxoma, *m*yxoid neurofibroma, and *e*phelides (freckles); POEMS, *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathies, *M*-protein, and *s*kin changes.

A proliferation of melanocytes results in the following pigmented lesions: *lentigo*, *melanocytic nevus*, and *melanoma* (Chap. 87). In an adult, the majority of lentigines are related to sun exposure, which explains their distribution. However, in the Peutz-Jeghers and LEOPARD [*l*entigines; *E*CG abnormalities, primarily

conduction defects; *o*cular hypertelorism; *p*ulmonary stenosis and subaortic valvular stenosis; *a*bnormal genitalia (cryptorchidism, hypospadias); *r*etardation of growth; and *d*eafness (sensorineural)] syndromes, lentigines do serve as a clue to systemic disease. In *LEOPARD syndrome*, hundreds of lentigines develop during childhood and are scattered over the entire surface of the body. The lentigines in patients with *Peutz-Jeghers syndrome* are located primarily around the nose and mouth, on the hands and feet, and within the oral cavity. While the pigmented macules on the face may fade with age, the oral lesions persist. However, similar intraoral lesions are also seen in Addison's disease, Laugier-Hunziker syndrome (no internal manifestations), and as a normal finding in darkly pigmented individuals. Patients with this autosomal dominant syndrome (due to mutations in a novel serine threonine kinase gene) have multiple benign polyps of the gastrointestinal tract, testicular or ovarian tumors, and an increased risk of developing gastrointestinal (primarily colon) and pancreatic cancers.

In the Carney complex, numerous lentigines are also seen but they are in association with cardiac myxomas. This autosomal dominant disorder is also known as the *LAMB* (*l*entigines, *a*trial myxomas, *m*ucocutaneous myxomas, and *b*lue nevi) *syndrome* or *NAME* [*n*evi, *a*trial myxoma, *m*yxoid neurofibroma, and *e*phelides (freckles)] *syndrome*. These patients can also have evidence of endocrine overactivity in the form of Cushing's syndrome (pigmented nodular adrenocortical disease) and acromegaly.

The third type of localized hyperpigmentation is due to a local increase in pigment production, and it includes *ephelides* and café au lait macules (CALM). While a single CALM can be seen in up to 10% of the normal population, the presence of multiple or large-sized CALM raises the possibility of an associated genodermatosis, e.g., neurofibromatosis (NF) or McCune-Albright syndrome. *CALM* are flat, uniformly brown in color (usually two shades darker than uninvolved skin), and can- vary in size from 0.5 –12 cm. Approximately 80–90% of adult patients with *type I NF* will have six or more CALM measuring ≥ 1.5 cm in diameter. Additional findings are discussed in the section on neurofibromas (see "Papulonodular Skin Lesions," below). In comparison with NF, the CALM in patients with *McCune-Albright syndrome* [polyostotic fibrous dysplasia with precocious puberty in females due to mosaicism for an activating mutation in a G protein ($G_{S}\alpha$) gene] are usually larger, more irregular in outline, and tend to respect the midline.

In incontinentia pigmenti, dyskeratosis congenita, and bleomycin pigmentation, the areas of localized hyperpigmentation form a pattern—swirled in the first, reticulated in the second, and flagellate in the third. In *dyskeratosis congenita*, atrophic reticulated hyperpigmentation is seen on the neck, trunk, and thighs and is accompanied by nail dystrophy, pancytopenia, and leukoplakia of the oral and anal mucosae. The latter often develops into squamous cell carcinoma. In addition to the flagellate pigmentation (linear streaks) on the trunk, patients receiving bleomycin often have hyperpigmentation overlying the elbows, knees, and small joints of the hand.

Localized hyperpigmentation is seen as a side effect of several other *systemic medications*, including those that produce fixed drug reactions [nonsteroidal anti-inflammatory drugs (NSAIDs), sulfonamides, barbiturates, and tetracyclines] and those that can complex with melanin (antimalarials) or iron (minocycline). Fixed drug eruptions recur in the exact same location as circular areas of erythema that can become bullous and then resolve as brown macules. The eruption usually appears within hours of administration of the offending agent, and common locations include the genitalia, extremities, and perioral region. Chloroquine and hydroxychloroquine produce gray-brown to blue-black discoloration of the shins, hard palate, and face, while blue macules (often misdiagnosed as bruises) can be seen on the lower extremities and in sites of inflammation with prolonged minocycline administration. Estrogen in oral contraceptives can induce melasma—symmetric brown patches on the face, especially the cheeks, upper lip, and forehead. Similar changes are seen in pregnancy and in patients receiving phenytoin.

In the diffuse forms of hyperpigmentation, the darkening of the skin may be of equal intensity over the entire body or may be accentuated in sun-exposed areas. The causes of diffuse hyperpigmentation can be divided into four major groups—endocrine, metabolic, autoimmune, and drugs. The endocrinopathies that frequently have associated hyperpigmentation include *Addisons disease*, *Nelson syndrome*, and *ectopic ACTH syndrome*. In these diseases, the increased pigmentation is diffuse but is accentuated in sun-exposed areas, the palmar creases, sites of friction, and scars. An overproduction of the pituitary hormones α-MSH (melanocyte-stimulating hormone) and ACTH can lead to an increase in melanocyte activity. These peptides are products of the proopiomelanocortin gene and exhibit homology; e.g., α-MSH and ACTH share 13 amino acids. A minority of the patients with Cushing's disease or hyperthyroidism have generalized hyperpigmentation.

The metabolic causes of hyperpigmentation include *porphyria cutanea tarda* (PCT), *hemochromatosis*, *vitamin B_{12} deficiency*, *folic acid deficiency*, *pellagra*, and *malabsorption*, including *Whipples disease*. In patients with PCT (see "Vesicles/Bullae," below), the skin darkening is seen in sun-exposed areas and is a reflection of the photoreactive properties of porphyrins. The increased level of iron in the skin of patients with hemochromatosis stimulates melanin pigment production and leads to the classic bronze color. Patients with pellagra have a brown discoloration of the skin, especially in sun-exposed areas, as a result of nicotinic acid (niacin) deficiency. In the areas of increased pigmentation, there is a thin, varnish-like scale. These changes are also seen in patients who are vitamin B6 deficient, have functioning carcinoid tumors (increased consumption of niacin), or take isoniazid. Approximately 50% of

the patients with Whipple's disease have an associated generalized hyperpigmentation in association with diarrhea, weight loss, arthritis, and lymphadenopathy. A diffuse, slate-blue color is seen in patients with *melanosis secondary to metastatic melanoma*. Although there is a debate as to whether the color is due to single-cell metastases in the dermis or to a widespread deposition of melanin resulting from the high concentration of circulating melanin precursors, there is more evidence to support the latter.

Of the autoimmune diseases associated with diffuse hyperpigmentation, *biliary cirrhosis* and *scleroderma* are the most common, and occasionally, both disorders are seen in the same patient. The skin is dark brown in color, especially in sun-exposed areas. In biliary cirrhosis, the hyperpigmentation is accompanied by pruritus, jaundice, and xanthomas, whereas in scleroderma, it is accompanied by sclerosis of the extremities, face, and, less commonly, the trunk. Additional clues to the diagnosis of scleroderma are mat and periungual telangiectasias, calcinosis cutis, Raynaud's phenomenon, and distal ulcerations (see "Telangiectasias," above). The differential diagnosis of cutaneous sclerosis with hyperpigmentation includes the POEMS [*p*olyneuropathy; *o*rganomegaly (liver, spleen, lymph nodes); *e*ndocrinopathies (impotence, gynecomastia); *M*-protein; and *s*kin changes] syndrome. The skin changes include hyperpigmentation, induration, hypertrichosis, and angiomas.

Diffuse hyperpigmentation that is due to drugs or metals can result from one of several mechanisms—induction of melanin pigment formation, complexing of the drug or its metabolites to melanin, and deposits of the drug in the dermis. Busulfan, cyclophosphamide, 5-fluorouracil, and inorganic arsenic induce pigment production. Complexes containing melanin or iron plus the drug or its metabolites are seen in patients receiving minocycline, and a diffuse, blue-gray, muddy appearance within sun-exposed areas may develop, in addition to pigmentation of the mucous membranes, teeth, nails, bones, and thyroid. Administration of amiodarone can result in both a phototoxic eruption (exaggerated sunburn) and/or a slate-gray to violaceous discoloration of sun-exposed skin. Biopsy specimens of the latter show yellow-brown granules in dermal macrophages, which represent intralysosomal accumulations of lipids, amiodarone, and its metabolites. Actual deposits of a particular drug or metal in the skin are seen with silver (argyria), where the skin appears blue-gray in color; gold (chrysiasis), where the skin has a brown to blue-gray color; and clofazimine, where the skin appears reddish brown. The associated hyperpigmentation is accentuated in sun-exposed areas, and discoloration of the eye is seen with gold (sclerae) and clofazimine (conjunctivae).

VESICLES/BULLAE

(Table 53-12) Depending on their size, cutaneous blisters are referred to as *vesicles* (<0.5 cm) or *bullae* (>0.5 cm). The primary autoimmune blistering disorders include *pemphigus vulgaris*, *pemphigus foliaceus*, *pemphigus erythematosus*, *paraneoplastic pemphigus*, *bullous pemphigoid*, *gestational pemphigoid*, *cicatricial pemphigoid*, *epidermolysis bullosa acquisita*, *linear IgA bullous dermatosis* (LABD), and *dermatitis herpetiformis* (Chap. 54).

Table 53-12 Causes of Vesicles/Bullae

I. Primary mucocutaneous diseases
A. Primary blistering diseases (autoimmune)
1. Pemphigus[a]
2. Bullous pemphigoid[b]
3. Gestational pemphigoid[b]
4. Cicatricial pemphigoid[b]
5. Dermatitis herpetiformis[b,c]
6. Linear IgA bullous dermatosis[b]
7. Epidermolysis bullosa acquisita[b,d]
B. Secondary blistering diseases
1. Contact dermatitis[a]
2. Erythema multiforme[e]
3. Stevens-Johnson syndrome[e]
4. Toxic epidermal necrolysis[e]
C. Infections
1. Varicella/zoster virus[a,f]
2. Herpes simplex virus[a,f]
3. Enteroviruses, e.g., hand-foot-and-mouth disease

4. Staphylococcal scalded-skin syndrome[a,g]
5. Bullous impetigo[a]
II. Systemic diseases
A. Autoimmune
1. Paraneoplastic pemphigus[a]
B. Infections
1. Cutaneous emboli[b]
C. Metabolic
1. Diabetic bullae[a,b]
2. Porphyria cutanea tarda[b]
3. Porphyria variegata[b]
4. Pseudoporphyria[b]
5. Bullous dermatosis of hemodialysis[b]
D. Ischemia
1. Coma bullae

[a] Intraepidermal.

[b] Subepidermal.

[c] Associated with gluten enteropathy.

[d] Associated with inflammatory bowel disease.

[e] Degeneration of cells within the basal layer of the epidermis can give impression split is subepidermal.

[f] Also systemic.

[g] In adults, associated with renal failure and immunocompromised state.

Vesicles and bullae are also seen in *contact dermatitis*, both allergic and irritant forms (Chap. 52). When there is a linear arrangement of vesicular lesions, an exogenous cause should be suspected. Bullous disease secondary to the ingestion of drugs can take one of several forms, including phototoxic eruptions, isolated bullae, Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN) (Chap. 55). Clinically, phototoxic eruptions resemble an exaggerated sunburn with diffuse erythema and bullae in sun-exposed areas. The most commonly associated drugs are doxycycline, quinolones, thiazides, NSAIDs, voriconazole, and psoralens. The development of a phototoxic eruption is dependent on the doses of both the drug and ultraviolet (UV)-A irradiation.

Toxic epidermal necrolysis is characterized by bullae that arise on widespread areas of erythema and then slough. This results in large areas of denuded skin. The associated morbidity, such as sepsis, and mortality rates are relatively high and are a function of the extent of epidermal necrosis. In addition, these patients may also have involvement of the mucous membranes and respiratory and intestinal tracts. Drugs are the primary cause of TEN, and the most common offenders are phenytoin, barbiturates, carbamazepine, sulfonamides, aminopenicillins, allopurinol, and NSAIDs. Severe acute graft-versus-host disease (grade 4), drug-induced LABD, and the acute syndrome of apoptotic pan-epidermolysis (ASAP) in patients with lupus can also resemble TEN.

In *erythema multiforme* (EM), the primary lesions are pink-red macules and edematous papules, the centers of which may become vesicular. In contrast to a morbilliform exanthem, the clue to the diagnosis of EM, and especially SJS, is the development of a "dusky" violet color in the center of the lesions. Target or iris lesions are also characteristic of EM and arise as a result of active centers and borders in combination with centrifugal spread. However, iris lesions need not be present to make the diagnosis of EM.

EM has been subdivided into two major groups: (1) EM minor due to herpes simplex virus (HSV) and (2) EM major due to HSV; *Mycoplasma pneumoniae*; or, occasionally, drugs. Involvement of the mucous membranes (oral, nasal, ocular, and genital) is seen more commonly in the latter form. Hemorrhagic crusts of the lips are characteristic of EM major and SJS as well as herpes simplex, pemphigus vulgaris, and paraneoplastic pemphigus. Fever, malaise, myalgias, sore throat, and cough may precede or accompany the eruption. The lesions of EM usually resolve over 2–4 weeks but may be recurrent, especially when due to HSV. In addition to HSV (in which lesions usually appear 7–12 days after the viral eruption), EM can also follow vaccinations, radiation therapy, and exposure to environmental toxins, including the oleoresin in poison ivy.

Induction of SJS is most often due to drugs, especially sulfonamides, phenytoin, barbiturates,

aminopenicillins, non-nucleoside reverse transcriptase inhibitors, and carbamazepine. Widespread dusky macules and significant mucosal involvement are characteristic of SJS, and the cutaneous lesions may or may not develop epidermal detachment. If the latter occurs, by definition, it is limited to <10% of the body surface area (BSA). Greater involvement leads to the diagnosis of SJS/TEN overlap (10–30% BSA) or TEN (>30% BSA).

In addition to primary blistering disorders and hypersensitivity reactions, bacterial and viral infections can lead to vesicles and bullae. The most common infectious agents are HSV (Chap. 179), varicella-zoster virus (Chap. 180), and *S. aureus* (Chap. 135).

Staphylococcal scalded-skin syndrome (SSSS) and *bullous impetigo* are two blistering disorders associated with staphylococcal (phage group II) infection. In SSSS, the initial findings are redness and tenderness of the central face, neck, trunk, and intertriginous zones. This is followed by short-lived flaccid bullae and a slough or exfoliation of the superficial epidermis. Crusted areas then develop, characteristically around the mouth. SSSS is distinguished from TEN by the following features: younger age group (primarily infants), more superficial site of blister formation, no oral lesions, shorter course, lower morbidity and mortality rates, and an association with staphylococcal exfoliative toxin ("exfoliatin"), not drugs. A rapid diagnosis of SSSS versus TEN can be made by a frozen section of the blister roof or exfoliative cytology of the blister contents. In SSSS the site of staphylococcal infection is usually extracutaneous (conjunctivitis, rhinorrhea, otitis media, pharyngitis, tonsillitis), and the cutaneous lesions are sterile, whereas in bullous impetigo, the skin lesions are the site of infection. Impetigo is more localized than SSSS and usually presents with honey-colored crusts. Occasionally, superficial purulent blisters also form. *Cutaneous emboli* from gram-negative infections may present as isolated bullae, but the base of the lesion is purpuric or necrotic, and it may develop into an ulcer (see "Purpura," below).

Several metabolic disorders are associated with blister formation, including diabetes mellitus, renal failure, and porphyria. Local hypoxemia secondary to decreased cutaneous blood flow can also produce blisters, which explains the presence of bullae over pressure points in comatose patients (coma bullae). In *diabetes mellitus*, tense bullae with clear viscous fluid arise on normal skin. The lesions can be as large as 6 cm in diameter and are located on the distal extremities. There are several types of porphyria, but the most common form with cutaneous findings is *porphyria cutanea tarda* (PCT). In sun-exposed areas (primarily the face and hands), the skin is very fragile, with trauma leading to erosions mixed with tense vesicles. These lesions then heal with scarring and formation of milia; the latter are firm, 1- to 2-mm white or yellow papules that represent epidermoid inclusion cysts. Associated findings can include hypertrichosis of the lateral malar region (men) or face (women) and, in sun-exposed areas, hyperpigmentation and firm sclerotic plaques. An elevated level of urinary uroporphyrins confirms the diagnosis and is due to a decrease in uroporphyrinogen decarboxylase activity. Precipitating agents include alcohol, iron, chlorinated hydrocarbons, hepatitis C infection, and hepatomas.

The differential diagnosis of PCT includes (1) *porphyria variegata*—the skin signs of PCT plus the systemic findings of acute intermittent porphyria; it has a diagnostic plasma porphyrin fluorescence emission at 626 nm; (2) *drug-induced* pseudoporphyria—the clinical and histologic findings are similar to PCT, but porphyrins are normal; etiologic agents include naproxen and other NSAIDs, furosemide, tetracycline, and retinoids; (3) *bullous dermatosis of hemodialysis*—the same appearance as PCT, but porphyrins are usually normal or occasionally borderline elevated; patients have chronic renal failure and are on hemodialysis; (4) PCT associated with hepatomas and hemodialysis; and (5) *epidermolysis bullosa acquisita* (Chap. 54).

EXANTHEMS

(Table 53-13) Exanthems are characterized by an acute generalized eruption. The most common presentation is erythematous macules and papules (morbilliform) and less often confluent blanching erythema (scarlatiniform). *Morbilliform* eruptions are usually due to either drugs or viral infections. For example, up to 5% of patients receiving penicillins, sulfonamides, phenytoin, or nevira- pine will develop a maculopapular eruption. Accompanying signs may include pruritus, fever, eosinophilia, and transient lymphadenopathy. Similar maculopapular eruptions are seen in the classic childhood viral exanthems, including (1) *rubeola* (measles)—a prodrome of coryza, cough, and conjunctivitis followed by Koplik's spots on the buccal mucosa; the eruption begins behind the ears, at the hairline, and on the forehead and then spreads down the body, often becoming confluent; (2) *rubella*—the eruption begins on the forehead and face and then spreads down the body; it resolves in the same order and is associated with retroauricular and suboccipital lymphadenopathy; and (3) *erythema infectiosum* (fifth disease)—erythema of the cheeks is followed by a reticulated pattern on the extremities; it is secondary to a parvovirus B19 infection, and an associated arthritis is seen in adults.

Table 53-13 Causes of Exanthems

I. Morbilliform
A. Drugs
B. Viral
1. Rubeola (measles)

2. Rubella
3. Erythema infectiosum (reticulated on extremities)
4. Epstein-Barr virus, echovirus, coxsackievirus, CMV, and adenovirus infections
5. Early HIV infection (plus mucosal ulcerations)
C. Bacterial
1. Typhoid fever
2. Early secondary syphilis
3. Early *Rickettsia* infections
4. Early meningococcemia
D. Acute graft-versus-host disease
E. Kawasaki's disease
II. Scarlatiniform
A. Scarlet fever
B. Toxic shock syndrome
C. Kawasaki's disease
D. Early staphylococcal scalded-skin syndrome

Abbreviations: CMV, cytomegalovirus; HIV, human immunodeficiency virus.

Both measles and rubella can occur in unvaccinated adults, and an atypical form of measles is seen in adults immunized with either killed measles vaccine or killed vaccine followed in time by live vaccine. In contrast to classic measles, the eruption of atypical measles begins on the palms, soles, wrists, and knuckles, and the lesions may become purpuric. The patient with atypical measles can have pulmonary involvement and be quite ill. Rubelliform and roseoliform eruptions are also associated with *Epstein-Barr virus* (5–15% of patients), *echovirus*, *coxsackievirus*, *cytomegalovirus*, *adenovirus*, *dengue virus*, and *West Nile virus* infections. Detection of specific IgM antibodies or fourfold elevations in IgG antibodies allows the proper diagnosis. Occasionally, a maculopapular drug eruption is a reflection of an underlying viral infection. For example, about 95% of the patients with infectious mononucleosis who are given ampicillin will develop a rash.

Of note, early in the course of infections with *Rickettsia* and meningococcus, prior to the development of purpura, the lesions may be erythematous macules and papules. This is also the case in chickenpox prior to the development of vesicles. Maculopapular eruptions are associated with early *HIV infection*, early secondary *syphilis*, *typhoid fever* , and *acute graft-versus-host disease*. In the last, lesions frequently begin on the dorsal hands and forearms; the macular rose spots of typhoid fever involve primarily the anterior trunk.

The prototypic *scarlatiniform* eruption is seen in *scarlet fever* and is due to an erythrotoxin produced by group A β-hemolytic streptococcal infections, most commonly pharyngitis. This eruption is characterized by diffuse erythema, which begins on the neck and upper trunk, and red follicular puncta. Additional findings include a white strawberry tongue (white coating with red papillae) followed by a red strawberry tongue (red tongue with red papillae); petechiae of the palate; a facial flush with circumoral pallor; linear petechiae in the antecubital fossae; and desquamation of the involved skin, palms, and soles 5–20 days after onset of the eruption. A similar desquamation of the palms and soles is seen in toxic shock syndrome (TSS), Kawasaki's disease, and after severe febrile illnesses. Certain strains of staphylococci also produce an erythrotoxin that leads to the same clinical findings as in streptococcal scarlet fever, except that the anti-streptolysin O or -DNase B titers are not elevated.

In *toxic shock syndrome*, staphylococcal (phage group I) infections produce an exotoxin (TSST-1) that causes the fever and rash as well as enterotoxins. Initially, the majority of cases were reported in menstruating women who were using tampons. However, other sites of infection, including wounds and nasal packing, can lead to TSS. The diagnosis of TSS is based on clinical criteria (Chap. 135), and three of these involve mucocutaneous sites (diffuse erythema of the skin, desquamation of the palms and soles 1–2 weeks after onset of illness, and involvement of the mucous membranes). The latter is characterized as hyperemia of the vagina, oropharynx, or conjunctivae. Similar systemic findings have been described in *streptococcal toxic shock syndrome* (Chap. 136), and although an exanthem is seen less often than in TSS due to a staphylococcal infection, the underlying infection is often in the soft tissue.

The cutaneous eruption in *Kawasakis disease* (mucocutaneous lymph node syndrome) (Chap. 326) is polymorphous, but the two most common forms are morbilliform and scarlatiniform. Additional mucocutaneous findings include bilateral conjunctival injection; erythema and edema of the hands and feet followed by desquamation; and diffuse erythema of the oropharynx, red strawberry tongue, and dry fissured lips. This clinical picture can resemble TSS and scarlet fever, but clues to the diagnosis of Kawasaki's disease are cervical lymphadenopathy, cheilitis, and thrombocytosis. The most serious associated systemic finding in this disease is coronary aneurysms secondary to arteritis. Aneurysms may lead to sudden death, primarily within the first 30 days of the illness. Scarlatiniform eruptions are also

seen in the early phase of SSSS (see "Vesicles/Bullae," above) and as reactions to drugs.

URTICARIA

(Table 53-14) *Urticaria* (hives) are transient lesions that are composed of a central wheal surrounded by an erythematous halo. Individual lesions are round, oval, or figurate and are often pruritic. Acute and chronic urticaria have a wide variety of allergic etiologies and reflect edema in the dermis. Urticarial lesions can also be seen in patients with mastocytosis (urticaria pigmentosa), hypo- or hyperthyroidism, and systemic-onset juvenile idiopathic arthritis (Still's disease). In both juvenile- and adult-onset Still's disease, the lesions coincide with the fever spike, are transient, and are due to dermal infiltrates of neutrophils.

Table 53-14 Causes of Urticaria and Angioedema

I. Primary cutaneous disorders
A. Acute and chronic urticaria[a]
B. Physical urticaria
1. Dermatographism
2. Solar urticaria[b]
3. Cold urticaria[b]
4. Cholinergic urticaria[b]
C. Angioedema (hereditary and acquired)[b]
II. Systemic diseases
A. Urticarial vasculitis
B. Hepatitis B or C infection
C. Serum sickness
D. Angioedema (hereditary and acquired)

[a] A small minority develop anaphylaxis.

[b] Also systemic.

The common *physical urticarias* include dermatographism, solar urticaria, cold urticaria, and cholinergic urticaria. Patients with *dermatographism* exhibit linear wheals following minor pressure or scratching of the skin. It is a common disorder, affecting ~5% of the population. *Solar urticaria* characteristically occurs within minutes of sun exposure and is a skin sign of one systemic disease— erythropoietic protoporphyria. In addition to the urticaria, these patients have subtle pitted scarring of the nose and hands. *Cold urticaria* is precipitated by exposure to the cold, and therefore exposed areas are usually affected. In occasional patients, the disease is associated with abnormal circulating proteins—more commonly cryoglobulins and less commonly cryofibrinogens. Additional systemic symptoms include wheezing and syncope, thus explaining the need for these patients to avoid swimming in cold water. *Cholinergic urticaria* is precipitated by heat, exercise, or emotion and is characterized by small wheals with relatively large flares. It is occasionally associated with wheezing.

Whereas urticarias are the result of dermal edema, subcutaneous edema leads to the clinical picture of *angioedema*. Sites of involvement include the eyelids, lips, tongue, larynx, and gastrointestinal tract as well as the subcutaneous tissue. Angioedema occurs alone or in combination with urticaria, including urticarial vasculitis and the physical urticarias. Both acquired and hereditary (autosomal dominant) forms of angioedema occur (Chap. 317), and in the latter, urticaria is rarely, if ever, seen.

Urticarial vasculitis is an immune complex disease that may be confused with simple urticaria. In contrast to simple urticaria, individual lesions tend to last longer than 24 h and usually develop central petechiae that can be observed even after the urticarial phase has resolved. The patient may also complain of burning rather than pruritus. On biopsy, there is a leukocytoclastic vasculitis of the small blood vessels. Although many cases of urticarial vasculitis are idiopathic in origin, it can be a reflection of an underlying systemic illness such as lupus erythematosus, Sjögren's syndrome, or hereditary complement deficiency. There is a spectrum of urticarial vasculitis that ranges from purely cutaneous to multisystem involvement. The most common systemic signs and symptoms are arthralgias and/or arthritis, nephritis, and crampy abdominal pain, with asthma and chronic obstructive lung disease seen less often. Hypocomplementemia occurs in one- to two-thirds of patients, even in the idiopathic cases. Urticarial vasculitis can also be seen in patients with *hepatitis B* and *hepatitis C* infections, *serum sickness*, and *serum sickness–like illnesses* (e.g., due to cefaclor, minocycline).

PAPULONODULAR SKIN LESIONS

(Table 53-15) In the *papulonodular diseases*, the lesions are elevated above the surface of the skin and

may coalesce to form plaques. The location, consistency, and color of the lesions are the keys to their diagnosis; this section is organized on the basis of color.

Table 53-15 Papulonodular Skin Lesions According to Color Groups
I. White
A. Calcinosis cutis
II. Skin-colored
A. Rheumatoid nodules
B. Neurofibromas (von Recklinghausen's disease)
C. Angiofibromas (tuberous sclerosis, MEN syndrome, type 1)
D. Neuromas (MEN syndrome, type 2b)
E. Adnexal tumors
1. Basal cell carcinomas (nevoid basal cell carcinoma syndrome)
2. Tricholemmomas (Cowden disease)
F. Osteomas (Gardner syndrome)
G. Primary cutaneous disorders
1. Epidermal inclusion cysts[a]
2. Lipomas
III. Pink/translucent[b]
A. Amyloidosis
B. Papular mucinosis
IV. Yellow
A. Xanthomas
B. Tophi
C. Necrobiosis lipoidica
D. Pseudoxanthoma elasticum
E. Sebaceous adenomas (Torre syndrome)
V. Red[b]
A. Papules
1. Angiokeratomas (Fabry disease)
2. Bacillary angiomatosis (primarily in AIDS)
B. Papules/plaques
1. Cutaneous lupus
2. Lymphoma cutis
3. Leukemia cutis
4. Sweet syndrome
C. Nodules
1. Panniculitis
2. Cutaneous polyarteritis nodosa
3. Systemic vasculitis
D. Primary cutaneous disorders
1. Arthropod bites
2. Cherry hemangiomas
3. Infections, e.g., erysipelas, sporotrichosis
4. Polymorphous light eruption
5. Lymphocytoma cutis (pseudolymphoma)
VI. Red-brown[b]
A. Sarcoidosis
B. Sweet's syndrome
C. Urticaria pigmentosa

D. Erythema elevatum diutinum (chronic leukocytoclastic vasculitis)
E. Lupus vulgaris
VII. Blue[b]
A. Venous malformations (e.g., blue rubber bleb syndrome)
B. Primary cutaneous disorders
1. Venous lake
2. Blue nevus
VIII. Violaceous
A. Lupus pernio (sarcoidosis)
B. Lymphoma cutis
C. Cutaneous lupus
IX. Purple
A. Kaposi's sarcoma
B. Angiosarcoma
C. Palpable purpura (see Table 53-16)
X. Brown-black[c]
XI. Any color
A. Metastases

[a] If multiple with childhood onset, consider Gardner syndrome.

[b] May have darker hue in more darkly pigmented individuals.

[c] See also "Hyperpigmentation."

Abbreviation: MEN, multiple endocrine neoplasia.

White Lesions

In *calcinosis cutis* there are firm white to white-yellow papules with an irregular surface. When the contents are expressed, a chalky white material is seen. *Dystrophic calcification* is seen at sites of previous inflammation or damage to the skin. It develops in acne scars as well as on the distal extremities of patients with scleroderma and in the subcutaneous tissue and intermuscular fascial planes in DM. The latter is more extensive and is more commonly seen in children. An elevated calcium phosphate product, most commonly due to secondary hyperparathyroidism in the setting of renal failure, can lead to nodules of *metastatic calcinosis cutis*, which tend to be subcutaneous and periarticular. These patients can also develop calcification of muscular arteries and subsequent ischemic necrosis (calciphylaxis).

Skin-Colored Lesions

There are several types of skin-colored lesions, including epidermoid inclusion cysts, lipomas, rheumatoid nodules, neurofibromas, angiofibromas, neuromas, and adnexal tumors such as tricholemmomas. Both *epidermoid inclusion cysts* and *lipomas* are very common mobile subcutaneous nodules—the former are rubbery and drain cheeselike material (sebum and keratin) if incised. Lipomas are firm and somewhat lobulated on palpation. When extensive facial epidermoid inclusion cysts develop during childhood or there is a family history of such lesions, the patient should be examined for other signs of Gardner syndrome, including osteomas and desmoid tumors. *Rheumatoid nodules* are firm 0.5- to 4-cm nodules that favor the extensor aspect of joints, especially the elbows. They are seen in ~20% of patients with rheumatoid arthritis and 6% of patients with Still's disease. Biopsies of the nodules show palisading granulomas. Similar lesions that are smaller and shorter-lived are seen in rheumatic fever.

Neurofibromas (benign Schwann cell tumors) are soft papules or nodules that exhibit the "button-hole" sign; that is, they invaginate into the skin with pressure in a manner similar to a hernia. Single lesions are seen in normal individuals, but multiple neurofibromas, usually in combination with six or more CALM measuring >1.5 cm (see "Hyperpigmentation," above), axillary freckling, and multiple Lisch nodules, are seen in von Recklinghausen's disease (NF type I; Chap. 379). In some patients, the neurofibromas are localized and unilateral due to somatic mosaicism.

Angiofibromas are firm pink to skin-colored papules that measure from 3 mm to a few centimeters in diameter. When multiple lesions are located on the central cheeks (adenoma sebaceum), the patient has tuberous sclerosis or multiple endocrine neoplasia (MEN) syndrome, type 1. The former is an autosomal disorder due to mutations in two different genes, and the associated findings are discussed in the section on ash leaf spots as well as in Chap. 379.

Neuromas (benign proliferations of nerve fibers) are also firm, skin-colored papules. They are more commonly found at sites of amputation and as rudimentary supernumerary digits. However, when there

are multiple neuromas on the eyelids, lips, distal tongue, and/or oral mucosa, the patient should be investigated for other signs of the MEN syndrome, type 2b. Associated findings include marfanoid habitus, protuberant lips, intestinal ganglioneuromas, and medullary thyroid carcinoma (>75% of patients; Chap. 351).

Adnexal tumors are derived from pluripotent cells of the epidermis that can differentiate toward hair, sebaceous, apocrine, or eccrine glands or remain undifferentiated. *Basal cell carcinomas* (BCCs) are examples of adnexal tumors that have little or no evidence of differentiation. Clinically, they are translucent papules with rolled borders, telangiectasias, and central erosion. BCCs commonly arise in sun-damaged skin of the head and neck as well as the upper trunk. When a patient has multiple BCCs, especially prior to age 30, the possibility of the nevoid basal cell carcinoma syndrome should be raised. It is inherited as an autosomal dominant trait and is associated with jaw cysts, palmar and plantar pits, frontal bossing, medulloblastomas, and calcification of the falx cerebri and diaphragma sellae. *Tricholemmomas* are also skin-colored adnexal tumors but differentiate toward hair follicles and can have a wartlike appearance. The presence of multiple tricholemmomas on the face and cobblestoning of the oral mucosa points to the diagnosis of Cowden disease (multiple hamartoma syndrome) due to mutations in the phosphatase and tensin homolog (*PTEN)* gene. Internal organ involvement (in decreasing order of frequency) includes fibrocystic disease and carcinoma of the breast, adenomas and carcinomas of the thyroid, and gastrointestinal polyposis. Keratoses of the palms, soles, and dorsal aspect of the hands are also seen.

Pink Lesions

The cutaneous lesions associated with primary systemic *amyloidosis* are often pink in color and translucent. Common locations are the face, especially the periorbital and perioral regions, and flexural areas. On biopsy, homogeneous deposits of amyloid are seen in the dermis and in the walls of blood vessels; the latter lead to an increase in vessel wall fragility. As a result, petechiae and purpura develop in clinically normal skin as well as in lesional skin following minor trauma, hence the term *pinch purpura*. Amyloid deposits are also seen in the striated muscle of the tongue and result in macroglossia.

Even though specific mucocutaneous lesions are rarely seen in secondary amyloidosis and are present in only ~30% of the patients with primary amyloidosis, a rapid diagnosis of systemic amyloidosis can be made by an examination of abdominal subcutaneous fat. By special staining, deposits are seen around blood vessels or individual fat cells in 40–50% of patients. There are also three forms of amyloidosis that are limited to the skin and that should not be construed as cutaneous lesions of systemic amyloidosis. They are macular amyloidosis (upper back), lichenoid amyloidosis (usually lower extremities), and nodular amyloidosis. In macular and lichenoid amyloidosis, the deposits are composed of altered epidermal keratin. Early-onset macular and lichenoid amyloidosis have been associated with MEN syndrome, type 2a.

Patients with *multicentric reticulohistiocytosis* also have pink-colored papules and nodules on the face and mucous membranes as well as on the extensor surface of the hands and forearms. They have a polyarthritis that can mimic rheumatoid arthritis clinically. On histologic examination, the papules have characteristic giant cells that are not seen in biopsies of rheumatoid nodules. Pink to skin-colored papules that are firm, 2–5 mm in diameter, and often in a linear arrangement are seen in patients with *papular mucinosis*. This disease is also referred to as *generalizedlichen myxedematosus* or *scleromyxedema*. The latter name comes from the induration of the face and extremities that may accompany the papular eruption. Biopsy specimens of the papules show localized mucin deposition, and serum protein electrophoresis and/or immunofixation electrophoresis demonstrates a monoclonal spike of IgG, usually with a λ light chain.

Yellow Lesions

Several systemic disorders are characterized by yellow-colored cutaneous papules or plaques —hyperlipidemia (xanthomas), gout (tophi), diabetes (necrobiosis lipoidica), pseudoxanthoma elasticum, and Torre syndrome (sebaceous tumors). Eruptive xanthomas are the most common form of *xanthomas* and are associated with hypertriglyceridemia (types I, III, IV, and V). Crops of yellow papules with erythematous halos occur primarily on the extensor surfaces of the extremities and the buttocks, and they spontaneously involute with a fall in serum triglycerides. Increased β-lipoproteins (primarily types II and III) result in one or more of the following types of xanthoma: xanthelasma, tendon xanthomas, and plane xanthomas. Xanthelasma are found on the eyelids, whereas tendon xanthomas are frequently associated with the Achilles and extensor finger tendons; plane xanthomas are flat and favor the palmar creases, face, upper trunk, and scars. Tuberous xanthomas are frequently associated with hypertriglyceridemia, but they are also seen in patients with hypercholesterolemia (type II) and are found most frequently over the large joints or hand. Biopsy specimens of xanthomas show collections of lipid-containing macrophages (foam cells).

Patients with several disorders, including biliary cirrhosis, can have a secondary form of hyperlipidemia with associated tuberous and planar xanthomas. However, patients with plasma cell dyscrasias have *normolipemic flat xanthomas*. This latter form of xanthoma may be ≥12 cm in diameter and is most frequently seen on the upper trunk or side of the neck. It is important to note that the most common setting for eruptive xanthomas is uncontrolled diabetes mellitus. The least specific sign for hyperlipidemia

is xanthelasma, because at least 50% of the patients with this finding have normal lipid profiles.

In *tophaceous gout*, there are deposits of monosodium urate in the skin around the joints, particularly those of the hands and feet. Additional sites of *tophi* formation include the helix of the ear and the olecranon and prepatellar bursae. The lesions are firm, yellow in color, and occasionally discharge a chalky material. Their size varies from 1 mm to 7 cm, and the diagnosis can be established by polarization of the aspirated contents of a lesion. Lesions of *necrobiosis lipoidica* are found primarily on the shins (90%), and patients can have diabetes mellitus or develop it subsequently. Characteristic findings include a central yellow color, atrophy (transparency), telangiectasias, and a red to red-brown border. Ulcerations can also develop within the plaques. Biopsy specimens show necrobiosis of collagen and granulomatous inflammation.

In *pseudoxanthoma elasticum* (PXE), due to mutations in the gene *ABCC6*, there is an abnormal deposition of calcium on the elastic fibers of the skin, eye, and blood vessels. In the skin, the flexural areas such as the neck, axillae, antecubital fossae, and inguinal area are the primary sites of involvement. Yellow papules coalesce to form reticulated plaques that have an appearance similar to that of plucked chicken skin. In severely affected skin, hanging, redundant folds develop. Biopsy specimens of involved skin show swollen and irregularly clumped elastic fibers with deposits of calcium. In the eye, the calcium deposits in Bruch's membrane lead to angioid streaks and choroiditis; in the arteries of the heart, kidney, gastrointestinal tract, and extremities, the deposits lead to angina, hypertension, gastrointestinal bleeding, and claudication, respectively. Long-term administration of d-penicillamine can lead to PXE-like skin changes as well as elastic fiber alterations in internal organs.

Adnexal tumors that have differentiated toward sebaceous glands include sebaceous adenoma, sebaceous carcinoma, and sebaceous hyperplasia. Except for sebaceous hyperplasia, which is commonly seen on the face, these tumors are fairly rare. Patients with Torre syndrome have one or more *sebaceous adenoma(s)*, and they can also have sebaceous carcinomas and sebaceous hyperplasia as well as keratoacanthomas. The internal manifestations of Torre syndrome include *multiple* carcinomas of the gastrointestinal tract (primarily colon) as well as cancers of the larynx, genitourinary tract, and breast.

Red Lesions

Cutaneous lesions that are red in color have a wide variety of etiologies; in an attempt to simplify their identification, they will be subdivided into papules, papules/plaques, and subcutaneous nodules. Common red papules include *arthropod bites* and *cherry hemangiomas*; the latter are small, bright-red, dome-shaped papules that represent benign proliferation of capillaries. In patients with AIDS (see Chapter 189), the development of multiple red hemangioma-like lesions points to bacillary angiomatosis, and biopsy specimens show clusters of bacilli that stain positive with the Warthin-Starry stain; the pathogens have been identified as *Bartonella henselae* and *B. quintana*. Disseminated visceral disease is seen primarily in immunocompromised hosts but can occur in immunocompetent individuals.

Multiple *angiokeratomas* are seen in Fabry disease, an X-linked recessive lysosomal storage disease that is due to a deficiency of α-galactosidase A. The lesions are red to red-blue in color and can be quite small in size (1–3 mm), with the most common location being the lower trunk. Associated findings include chronic renal failure, peripheral neuropathy, and corneal opacities (cornea verticillata). Electron photomicrographs of angiokeratomas and clinically normal skin demonstrate lamellar lipid deposits in fibroblasts, pericytes, and endothelial cells that are diagnostic of this disease. Widespread acute eruptions of erythematous papules are discussed in the section on exanthems.

There are several infectious diseases that present as erythematous papules or nodules in a lymphocutaneous or sporotrichoid pattern, i.e., in a linear arrangement along the lymphatic channels. The two most common etiologies are *Sporothrix schenckii* (sporotrichosis) and the atypical mycobacterium *M. marinum*. The organisms are introduced as a result of trauma, and a primary inoculation site is often seen in addition to the lymphatic nodules. Additional causes include *Nocardia*, *Leishmania*, and other dimorphic fungi; culture of lesional tissue will aid in the diagnosis.

The diseases that are characterized by erythematous plaques with scale are reviewed in the papulosquamous section, and the various forms of dermatitis are discussed in the section on erythroderma. Additional disorders in the differential diagnosis of red papules/plaques include *erysipelas*, *polymorphous light eruption* (PMLE), *cutaneous lymphoid hyperplasia* (lymphocytoma cutis), *cutaneous lupus*, *lymphoma cutis*, and *leukemia cutis*. The first three diseases represent primary cutaneous disorders. PMLE is characterized by erythematous papules and plaques in a primarily sun-exposed distribution—dorsum of the hand, extensor forearm, and upper trunk. Lesions follow exposure to UV-B and/or UV-A, and in higher latitudes, PMLE is most severe in the late spring and early summer. A process referred to as "hardening" occurs with continued UV exposure, and the eruption fades, but in temperate climates, it will recur in the spring. PMLE must be differentiated from cutaneous lupus, and this is accomplished by observation of the natural history, histologic examination, and direct immunofluorescence of the lesions. Cutaneous lymphoid hyperplasia (pseudolymphoma) is a *benign* polyclonal proliferation of lymphocytes in the skin that presents as infiltrated pink-red to red-purple papules and plaques; it must be distinguished from lymphoma cutis.

Several types of red plaques are seen in patients with systemic *lupus*, including (1) erythematous urticarial plaques across the cheeks and nose in the classic butterfly rash; (2) erythematous discoid

lesions with fine or "carpet-tack" scale, telangiectasias, central hypopigmentation, peripheral hyperpigmentation, follicular plugging, and atrophy located on the face, scalp, external ears, arms, and upper trunk; and (3) psoriasiform or annular lesions of subacute cutaneous lupus with hypopigmented centers located primarily on the extensor arms and upper trunk. Additional mucocutaneous findings include (1) a violaceous flush on the face and V of the neck; (2) photosensitivity; (3) urticarial vasculitis (see "Urticaria," above); (4) lupus panniculitis (see below); (5) diffuse alopecia; (6) alopecia secondary to discoid lesions; (7) periungual telangiectasias and erythema; (8) EM-like lesions that may become bullous; (9) oral ulcers; and (10) distal ulcerations secondary to Raynaud's phenomenon, vasculitis, or livedoid vasculopathy. Patients with only discoid lesions usually have the form of lupus that is limited to the skin. However, 2–10% of these patients eventually develop systemic lupus. Direct immunofluorescence of involved skin shows deposits of IgG or IgM and C3 in a granular distribution along the dermal-epidermal junction.

In *lymphoma cutis*, there is a proliferation of malignant lymphocytes in the skin, and the clinical appearance resembles that of cutaneous lymphoid hyperplasia—infiltrated pink-red to red-purple papules and plaques. Lymphoma cutis can occur anywhere on the surface of the skin, whereas the sites of predilection for lymphocytomas include the malar ridge, tip of the nose, and earlobes. Patients with non-Hodgkin's lymphomas have specific cutaneous lesions more often than those with Hodgkin's disease, and, occasionally, the skin nodules precede the development of extracutaneous non-Hodgkin's lymphoma or represent the only site of involvement (e.g., primary cutaneous B cell lymphoma). Arcuate lesions are sometimes seen in lymphoma and lymphocytoma cutis as well as in CTCL. *Adult T cell leukemia/lymphoma* that develops in association with HTLV-1 infection is characterized by cutaneous plaques, hypercalcemia, and circulating CD25+ lymphocytes. *Leukemia cutis* has the same appearance as lymphoma cutis, and specific lesions are seen more commonly in monocytic leukemias than in lymphocytic or granulocytic leukemias. Cutaneous chloromas (granulocytic sarcomas) may precede the appearance of circulating blasts in acute myelogenous leukemia and, as such, represent a form of aleukemic leukemia cutis.

Common causes of erythematous subcutaneous nodules include inflamed epidermoid inclusion cysts, acne cysts, and furuncles. *Panniculitis*, an inflammation of the fat, also presents as subcutaneous nodules and is frequently a sign of systemic disease. There are several forms of panniculitis, including erythema nodosum, erythema induratum/nodular vasculitis, lupus profundus, lipodermatosclerosis, α_1-antitrypsin deficiency, factitial, and fat necrosis secondary to pancreatic disease. Except for erythema nodosum, these lesions may break down and ulcerate or heal with a scar. The shin is the most common location for the nodules of erythema nodosum, whereas the calf is the most common location for lesions of erythema induratum. In erythema nodosum, the nodules are initially red but then develop a blue color as they resolve. Patients with erythema nodosum but no underlying systemic illness can still have fever, malaise, leukocytosis, arthralgias, and/or arthritis. However, the possibility of an underlying illness should be excluded, and the most common associations are streptococcal infections, upper respiratory viral infections, sarcoidosis, and inflammatory bowel disease in addition to drugs (oral contraceptives, sulfonamides, penicillins, bromides, iodides). Less common associations include bacterial gastroenteritis (*Yersinia*, *Salmonella*) and coccidioidomycosis followed by tuberculosis, histoplasmosis, brucellosis, and infections with *Chlamydophila pneumoniae* or *Chlamydia trachomatis, Mycoplasma pneumoniae,* or hepatitis B virus.

Erythema induratum and nodular vasculitis have overlapping features clinically and histologically, and whether they represent two separate entities or the ends of a single disease spectrum is a point of debate; in general, the latter is usually idiopathic and the former is associated with the presence of *M. tuberculosis* DNA by polymerase chain reaction (PCR) within skin lesions. The lesions of lupus panniculitis are found primarily on the cheeks, upper arms, and buttocks (sites of abundant fat) and are seen in both the cutaneous and systemic forms of lupus. The overlying skin may be normal, erythematous, or have the changes of discoid lupus. The subcutaneous fat necrosis that is associated with pancreatic disease is presumably secondary to circulating lipases and is seen in patients with pancreatic carcinoma as well as in patients with acute and chronic pancreatitis. In this disorder, there may be an associated arthritis, fever, and inflammation of visceral fat. Histologic examination of deep incisional biopsy specimens will aid in the diagnosis of the particular type of panniculitis.

Subcutaneous erythematous nodules are also seen in *cutaneous polyarteritis nodosa* (PAN) and as a manifestation of *systemic vasculitis*, e.g., systemic PAN, allergic granulomatosis, or granulomatosis with polyangiitis (Wegener's) (Chap. 326). Cutaneous PAN presents with painful subcutaneous nodules and ulcers within a red-purple, netlike pattern of livedo reticularis. The latter is due to slowed blood flow through the superficial horizontal venous plexus. The majority of lesions are found on the lower extremity, and while arthralgias and myalgias may accompany cutaneous PAN, there is no evidence of systemic involvement. In both the cutaneous and systemic forms of vasculitis, skin biopsy specimens of the associated nodules will show the changes characteristic of a vasculitis; the size of the vessel involved will depend on the particular disease.

Red-Brown Lesions

The cutaneous lesions in *sarcoidosis* (Chap. 329) are classically red to red-brown in color, and with diascopy (pressure with a glass slide), a yellow-brown residual color is observed that is secondary to the

granulomatous infiltrate. The waxy papules and plaques may be found anywhere on the skin, but the face is the most common location. Usually there are no surface changes, but occasionally the lesions will have scale. Biopsy specimens of the papules show "naked" granulomas in the dermis, i.e., granulomas surrounded by a minimal number of lymphocytes. Other cutaneous findings in sarcoidosis include annular lesions with an atrophic or scaly center, papules within scars, hypopigmented macules and papules, alopecia, acquired ichthyosis, erythema nodosum, and lupus pernio (see below).

The differential diagnosis of sarcoidosis includes foreign-body granulomas produced by chemicals such as beryllium and zirconium, late secondary syphilis, and *lupus vulgaris*. Lupus vulgaris is a form of cutaneous tuberculosis that is seen in previously infected and sensitized individuals. There is often underlying active tuberculosis elsewhere, usually in the lungs or lymph nodes. Lesions occur primarily in the head and neck region and are red-brown plaques with a yellow-brown color on diascopy. Secondary scarring and squamous cell carcinomas can develop within the plaques. Cultures or PCR analysis of the lesions should be performed because it is rare for the acid-fast stain to show bacilli within the dermal granulomas.

Sweets syndrome is characterized by red to red-brown plaques and nodules that are frequently painful and occur primarily on the head, neck, and upper (and, less often, lower) extremities. The patients also have fever, neutrophilia, and a dense dermal infiltrate of neutrophils in the lesions. In ~10% of the patients, there is an associated malignancy, most commonly acute myelogenous leukemia. Sweet's syndrome has also been reported with inflammatory bowel disease, systemic lupus, and solid tumors (primarily of the genitourinary tract) as well as drugs [e.g., *all-trans*-retinoic acid, granulocyte colony-stimulating factor (G-CSF)]. The differential diagnosis includes neutrophilic eccrine hidradenitis; atypical forms of pyoderma gangrenosum; and, occasionally, cellulitis. Extracutaneous sites of involvement include joints, muscles, eye, kidney (proteinuria, occasionally glomerulonephritis), and lung (neutrophilic infiltrates). The idiopathic form of Sweet's syndrome is seen more often in women, following a respiratory tract infection.

A generalized distribution of red-brown macules and papules is seen in the form of mastocytosis known as *urticaria pigmentosa* (Chap. 317). Each lesion represents a collection of mast cells in the dermis, with hyperpigmentation of the overlying epidermis. Stimuli such as rubbing cause these mast cells to degranulate, and this leads to the formation of localized urticaria (Darier's sign). Additional symptoms can result from mast cell degranulation and include headache, flushing, diarrhea, and pruritus. Mast cells also infiltrate various organs such as the liver, spleen, and gastrointestinal tract, and accumulations of mast cells in the bones may produce either osteosclerotic or osteolytic lesions on radiographs. In the majority of these patients, however, the internal involvement remains indolent. A subtype of chronic cutaneous small-vessel vasculitis, *erythema elevatum diutinum* (EED), also presents with papules that are red-brown in color. The papules coalesce into plaques on the extensor surfaces of knees, elbows, and the small joints of the hand. Flares of EED have been associated with streptococcal infections.

Blue Lesions

Lesions that are blue in color are the result of either vascular ectasias and tumors or melanin pigment in the dermis. *Venous lakes* (ectasias) are compressible dark-blue lesions that are found commonly in the head and neck region. *Venous malformations* are also compressible blue papulonodules and plaques that can occur anywhere on the body, including the oral mucosa. When there are multiple rather than single congenital lesions, the patient may have the blue rubber bleb syndrome or Maffucci's syndrome. Patients with the blue rubber bleb syndrome also have vascular anomalies of the gastrointestinal tract that may bleed, whereas patients with Maffucci's syndrome have associated dyschondroplasia and osteochondromas. *Blue nevi* (moles) are seen when there are collections of pigment-producing nevus cells in the dermis. These benign papular lesions are dome-shaped and occur most commonly on the dorsum of the hand or foot or in the head and neck region.

Violaceous Lesions

Violaceous papules and plaques are seen in *lupus pernio*, *lymphoma cutis*, and *cutaneous lupus*. Lupus pernio is a particular type of sarcoidosis that involves the tip and alar rim of the nose as well as the earlobes, with lesions that are violaceous in color rather than red-brown. This form of sarcoidosis is associated with involvement of the upper respiratory tract. The plaques of lymphoma cutis and cutaneous lupus may be red or violaceous in color and were discussed above.

Purple Lesions

Purple-colored papules and plaques are seen in vascular tumors, such as *Kaposis sarcoma* (Chap. 189) and *angiosarcoma*, and when there is extravasation of red blood cells into the skin in association with inflammation, as in *palpable purpura* (see "Purpura," below). Patients with congenital or acquired AV fistulas and venous hypertension can develop purple papules on the lower extremities that can resemble Kaposi's sarcoma clinically and histologically; this condition is referred to as pseudo-Kaposi sarcoma (acral angiodermatitis). Angiosarcoma is found most commonly on the scalp and face of elderly patients or within areas of chronic lymphedema and presents as purple papules and plaques. In the head and neck region, the tumor often extends beyond the clinically defined borders and may be accompanied by facial edema.

Brown and Black Lesions

Brown- and black-colored papules are reviewed in "Hyperpigmentation," above.

Cutaneous Metastases

These are discussed last because they can have a wide range of colors. Most commonly, they present as either firm, skin-colored subcutaneous nodules or firm, red to red-brown papulonodules. The lesions of lymphoma cutis range from pink-red to plum in color, whereas metastatic melanoma can be pink, blue, or black in color. Cutaneous metastases develop from hematogenous or lymphatic spread and are most often due to the following primary carcinomas: in men, melanoma, oropharynx, lung, and colon; and in women, breast, melanoma, and ovary. These metastatic lesions may be the initial presentation of the carcinoma, especially when the primary site is the lung.

PURPURA

(Table 53-16)*Purpura* are seen when there is an extravasation of red blood cells into the dermis and, as a result, the lesions do not blanch with pressure. This is in contrast to those erythematous or violet-colored lesions that are due to localized vasodilatation—they do blanch with pressure. Purpura (≤3 mm) and petechiae (≤2 mm) are divided into two major groups: palpable and nonpalpable. The most frequent causes of *nonpalpable* petechiae and purpura are primary cutaneous disorders such as *trauma*, *solar* (*actinic*) *purpura*, and *capillaritis*. Less common causes are *steroid purpura* and *livedoid vasculopathy* (see "Ulcers," below). Solar purpura are seen primarily on the extensor forearms, while steroid purpura secondary to potent topical glucocorticoids or endogenous or exogenous Cushing's syndrome can be more widespread. In both cases, there is alteration of the supporting connective tissue that surrounds the dermal blood vessels. In contrast, the petechiae that result from capillaritis are found primarily on the lower extremities. In capillaritis, there is an extravasation of erythrocytes as a result of perivascular lymphocytic inflammation. The petechiae are bright red, 1–2 mm in size, and scattered within annular or coin-shaped yellow-brown macules. The yellow-brown color is caused by hemosiderin deposits within the dermis.

Table 53-16 Causes of Purpura

I. Primary cutaneous disorders
A. Nonpalpable
1. Trauma
2. Solar (actinic, senile) purpura
3. Steroid purpura
4. Capillaritis
5. Livedoid vasculopathy in the setting of venous hypertension[a]
II. Systemic diseases
A. Nonpalpable
1. Clotting disturbances
a. Thrombocytopenia (including ITP)
b. Abnormal platelet function
c. Clotting factor defects
2. Vascular fragility
a. Amyloidosis
b. Ehlers-Danlos syndrome
c. Scurvy
3. Thrombi
a. Disseminated intravascular coagulation
b. Monoclonal cryoglobulinemia
c. Heparin-induced thrombocytopenia and thrombosis
d. Thrombocytosis
e. Thrombotic thrombocytopenic purpura
f. Antiphospholipid antibody syndrome
g. Warfarin reaction
h. Homozygousprotein C or protein S deficiency
4. Emboli
a. Cholesterol
b. Fat

5. Possible immune complex
a. Gardner-Diamond syndrome (autoerythrocyte sensitivity)
b. Waldenström's hypergammaglobulinemic purpura
B. Palpable
1. Vasculitis
a. Cutaneous small-vessel vasculitis
b. Polyarteritis nodosa
2. Emboli[b]
a. Acute meningococcemia
b. Disseminated gonococcal infection
c. Rocky Mountain spotted fever
d. Ecthyma gangrenosum

[a] Also associated with underlying disorders that lead to hypercoagulability, e.g. factor V Leiden, protein C dysfunction/deficiency.

[b] Bacterial, fungal, or parasitic.

Abbreviation: ITP, idiopathic thrombocytopenic purpura.

Systemic causes of nonpalpable purpura fall into several categories, and those secondary to clotting disturbances and vascular fragility will be discussed first. The former group includes *thrombocytopenia* (Chap. 115), *abnormal platelet function* as is seen in uremia, and *clotting factor defects* . The initial site of presentation for thrombocytopenia-induced petechiae is the distal lower extremity. Capillary fragility leads to nonpalpable purpura in patients with systemic *amyloidosis* (see "Papulonodular Skin Lesions," above), disorders of collagen production such as *Ehlers-Danlos syndrome*, and *scurvy*. In scurvy, there are flattened corkscrew hairs with surrounding hemorrhage on the lower extremities, in addition to gingivitis. Vitamin C is a cofactor for lysyl hydroxylase, an enzyme involved in the posttranslational modification of procollagen that is necessary for cross-link formation.

In contrast to the previous group of disorders, the purpura (noninflammatory with a retiform outline) seen in the following group of diseases are associated with thrombi formation within vessels. It is important to note that these thrombi are demonstrable in skin biopsy specimens. This group of disorders includes disseminated intravascular coagulation (DIC), monoclonal cryoglobulinemia, thrombocytosis, thrombotic thrombocytopenic purpura, antiphospholipid antibody syndrome, and reactions to warfarin and heparin (heparin-induced thrombocytopenia and thrombosis). DIC is triggered by several types of infection (gram-negative, gram-positive, viral, and rickettsial) as well as by tissue injury and neoplasms. Widespread purpura and hemorrhagic infarcts of the distal extremities are seen. Similar lesions are found in purpura fulminans, which is a form of DIC associated with fever and hypotension that occurs more commonly in children following an infectious illness such as varicella, scarlet fever, or an upper respiratory tract infection. In both disorders, hemorrhagic bullae can develop in involved skin.

Monoclonal cryoglobulinemia is associated with multiple myeloma, Waldenström's macroglobulinemia, chronic lymphocytic leukemia, and lymphoma. Purpura, primarily of the lower extremities, and hemorrhagic infarcts of the fingers and toes are seen in these patients. Exacerbations of disease activity can follow cold exposure or an increase in serum viscosity. Biopsy specimens show precipitates of the cryoglobulin within dermal vessels. Similar deposits have been found in the lung, brain, and renal glomeruli. Patients with *thrombotic thrombocytopenic purpura* can also have hemorrhagic infarcts as a result of intravascular thromboses. Additional signs include thrombocytopenic purpura, fever, and microangiopathic hemolytic anemia.

Administration of *warfarin* can result in painful areas of erythema that become purpuric and then necrotic with an adherent black eschar; the condition is referred to as warfarin-induced necrosis. This reaction is seen more often in women and in areas with abundant subcutaneous fat—breasts, abdomen, buttocks, thighs, and calves. The erythema and purpura develop between the third and tenth day of therapy, most likely as a result of a transient imbalance in the levels of anticoagulant and procoagulant vitamin K –dependent factors. Continued therapy does not exacerbate preexisting lesions, and patients with an inherited or acquired deficiency of protein C are at increased risk for this particular reaction as well as for purpura fulminans.

Purpura secondary to *cholesterol emboli* are usually seen on the lower extremities of patients with atherosclerotic vascular disease. They often follow anticoagulant therapy or an invasive vascular procedure such as an arteriogram but also occur spontaneously from disintegration of atheromatous plaques. Associated findings include livedo reticularis, gangrene, cyanosis, and ischemic ulcerations. Multiple step sections of the biopsy specimen may be necessary to demonstrate the cholesterol clefts within the vessels. Petechiae are also an important sign of *fat embolism* and occur primarily on the upper body 2–3 days after a major injury. By using special fixatives, the emboli can be demonstrated in

biopsy specimens of the petechiae. Emboli of tumor or thrombus are seen in patients with atrial myxomas and marantic endocarditis.

In the *Gardner-Diamond syndrome* (autoerythrocyte sensitivity), female patients develop large ecchymoses within areas of painful, warm erythema. Intradermal injections of autologous erythrocytes or phosphatidyl serine derived from the red cell membrane can reproduce the lesions in some patients; however, there are instances where a reaction is seen at an injection site of the forearm but not in the midback region. The latter has led some observers to view Gardner-Diamond syndrome as a cutaneous manifestation of severe emotional stress. More recently, the possibility of platelet dysfunction (as assessed via aggregation studies) has been raised. *Waldenströms hypergammaglobulinemic purpura* is a chronic disorder characterized by petechiae on the lower extremities. There are circulating complexes of IgG–anti-IgG molecules, and exacerbations are associated with prolonged standing or walking.

Palpable purpura are further subdivided into vasculitic and embolic. In the group of vasculitic disorders, cutaneous small-vessel vasculitis, also known as *leukocytoclastic vasculitis* (LCV), is the one most commonly associated with palpable purpura (Chap. 326). Underlying etiologies include drugs (e.g., antibiotics), infections (e.g., hepatitis C virus), and autoimmune connective tissue diseases (e.g. rheumatoid arthritis, Sjögren syndrome, lupus). *Henoch-Schönlein purpura* (HSP) is a subtype of acute LCV that is seen primarily in children and adolescents following an upper respiratory infection. The majority of lesions are found on the lower extremities and buttocks. Systemic manifestations include fever, arthralgias (primarily of the knees and ankles), abdominal pain, gastrointestinal bleeding, and nephritis. Direct immunofluorescence examination shows deposits of IgA within dermal blood vessel walls. Renal disease is of particular concern in adults with HSP. In *polyarteritis nodosa*, specific cutaneous lesions result from a vasculitis of arterial vessels (arteritis) or there may be an associated LCV. Arteritis leads to an infarct of the skin, and this explains the irregular outline of the purpura (see below).

Several types of infectious emboli can give rise to palpable purpura. These embolic lesions are usually *irregular* in outline as opposed to the lesions of LCV, which are *circular* in outline. The irregular outline is indicative of a cutaneous infarct, and the size corresponds to the area of skin that received its blood supply from that particular arteriole or artery. The palpable purpura in LCV are circular because the erythrocytes simply diffuse out evenly from the postcapillary venules as a result of inflammation. Infectious emboli are most commonly due to gram-negative cocci (meningococcus, gonococcus), gram-negative rods (Enterobacteriaceae), and gram-positive cocci (*Staphylococcus*). Additional causes include *Rickettsia* and, in immunocompromised patients, *Candida* and other opportunistic fungi.

The embolic lesions in *acute meningococcemia* are found primarily on the trunk, lower extremities, and sites of pressure, and a gunmetal-gray color often develops within them. Their size varies from a few millimeters to several centimeters, and the organisms can be cultured from the lesions. Associated findings include a preceding upper respiratory tract infection; fever; meningitis; DIC; and, in some patients, a deficiency of the terminal components of complement. In *disseminated gonococcal infection* (arthritis-dermatitis syndrome), a small number of papules and vesicopustules with central purpura or hemorrhagic necrosis are found on the distal extremities. Additional symptoms include arthralgias, tenosynovitis, and fever. To establish the diagnosis, a Gram stain of these lesions should be performed. *Rocky Mountain spotted fever* is a tick-borne disease that is caused by *R. rickettsii*. A several-day history of fever, chills, severe headache, and photophobia precedes the onset of the cutaneous eruption. The initial lesions are erythematous macules and papules on the wrists, ankles, palms, and soles. With time, the lesions spread centripetally and become purpuric.

Lesions of *ecthyma gangrenosum* begin as edematous, erythematous papules or plaques and then develop central purpura and necrosis. Bullae formation also occurs in these lesions, and they are frequently found in the girdle region. The organism that is classically associated with ecthyma gangrenosum is *Pseudomonas aeruginosa*, but other gram-negative rods such as *Klebsiella*, *Escherichia coli*, and *Serratia* can produce similar lesions. In immunocompromised hosts, the list of potential pathogens is expanded to include *Candida* and other opportunistic fungi (e.g., *Aspergillus*, *Fusarium)*.

ULCERS

The approach to the patient with a cutaneous ulcer is outlined in **Table 53-17**. Peripheral vascular diseases of the extremities are reviewed in Chap. 249, as is Raynaud's phenomenon.

Table 53-17 Causes of Mucocutaneous Ulcers

I. Primary cutaneous disorders
A. Peripheral vascular disease (Chap. 249)
1. Venous
2. Arterial[a]
B. Livedoid vasculopathy in the setting of venous hypertension[b]

C. Squamous cell carcinoma, e.g., within scars, basal cell carcinomas
D. Infections, e.g., ecthyma caused by *Streptococcus* (Chap. 136)
E. Physical, e.g., trauma, pressure
F. Drugs, e.g., hydroxyurea
II. Systemic diseases
A. Lower legs
1. Small-vessel and medium-vessel vasculitis[c]
2. Hemoglobinopathies (Chap. 104)
3. Cryoglobulinemia,[c] cryofibrinogenemia
4. Cholesterol emboli[c]
5. Necrobiosis lipoidica[d]
6. Antiphospholipid syndrome (Chap. 116)
7. Neuropathic[e] (Chap. 344)
8. Panniculitis
9. Kaposi's sarcoma, acral angiodermatitis
B. Hands and feet
1. Raynaud's phenomenon (Chap. 249)
2. Buerger disease
C. Generalized
1. Pyoderma gangrenosum, but most commonly legs
2. Calciphylaxis (Chap. 353)
3. Infections, e.g., dimorphic fungi, leishmaniasis
4. Lymphoma
D. Face, especially perioral, and anogenital
1. Chronic herpes simplex[f]
III. Mucosal
A. Behçet's syndrome (Chap. 327)
B. Erythema multiforme major, Stevens-Johnson syndrome, TEN
C. Primary blistering disorders (Chap. 54)
D. Lupus erythematosus, lichen planus
E. Inflammatory bowel disease
F. Acute HIV infection
G. Reactive arthritis (formerly known as Reiter's syndrome)

[a] Underlying atherosclerosis.

[b] Also associated with underlying disorders that lead to hypercoagulability, e.g., factor V Leiden, protein C dysfunction/deficiency, anti-phospholipid antibodies.

[c] Reviewed in section on Purpura.

[d] Reviewed in section on Papulonodular Skin Lesions.

[e] Favors plantar surface of the foot.

[f] Sign of immunosuppression.

Abbreviation: TEN, toxic epidermal necrolysis.

Livedoid vasculopathy (livedoid vasculitis; atrophie blanche) represents a combination of a vasculopathy plus intravascular thrombosis. Purpuric lesions and livedo reticularis are found in association with *painful* ulcerations of the lower extremities. These ulcers are often slow to heal, but when they do, irregularly shaped white scars form. The majority of cases are secondary to venous hypertension, but possible underlying illnesses include cryofibrinogenemia and disorders of hypercoagulability, e.g., the antiphospholipid antibody syndrome (Chaps. 117 and 320).

In *pyoderma gangrenosum*, the border of untreated ulcers has a characteristic appearance consisting of an undermined necrotic violaceous edge and a peripheral erythematous halo. The ulcers often begin as pustules that then expand rather rapidly to a size as large as 20 cm. Although these lesions are most commonly found on the lower extremities, they can arise anywhere on the surface of the body, including sites of trauma (pathergy). An estimated 30–50% of cases are idiopathic, and the most common

associated disorders are ulcerative colitis and Crohn's disease. Less commonly, pyoderma gangrenosum is associated with seropositive rheumatoid arthritis, acute and chronic myelogenous leukemia, hairy cell leukemia, and myelofibrosis. Additional findings in these patients, even those with idiopathic disease, are cutaneous anergy and a monoclonal gammopathy, usually IgA. Because the histology of pyoderma gangrenosum may be nonspecific (dermal infiltrate of neutrophils when in untreated state), the diagnosis requires clinicopathologic correlation, in particular, the exclusion of similar-appearing ulcers such as necrotizing vasculitis, Meleney's ulcer (synergistic infection at a site of trauma or surgery), dimorphic fungi, cutaneous amebiasis, spider bites, and factitial. In the myeloproliferative disorders, the ulcers may be more superficial with a pustulobullous border, and these lesions provide a connection between classic pyoderma gangrenosum and acute febrile neutrophilic dermatosis (Sweet's syndrome).

FEVER AND RASH

The major considerations in a patient with a fever and a rash are inflammatory diseases versus infectious diseases. In the hospital setting, the most common scenario is a patient who has a drug rash plus a fever secondary to an underlying infection. However, it should be emphasized that a drug reaction can lead to both a cutaneous eruption and a fever ("drug fever"), especially in the setting of DRESS or AGEP. Additional inflammatory diseases that are often associated with a fever include pustular psoriasis, erythroderma, and Sweet's syndrome. Lyme disease, secondary syphilis, and viral and bacterial exanthems (see "Exanthems," above) are examples of infectious diseases that produce a rash and a fever. Lastly, it is important to determine whether or not the cutaneous lesions represent septic emboli (see "Purpura," above). Such lesions usually have evidence of ischemia in the form of purpura, necrosis, or impending necrosis (gunmetal-gray color). In the patient with thrombocytopenia, however, purpura can be seen in inflammatory reactions such as morbilliform drug eruptions and infectious lesions.

FURTHER READINGS

Bolognia JL et al: *Dermatology*, 2nd ed. Philadelphia, Mosby, 2008

Braverman IM: *Skin Signs of Systemic Disease* , 3rd ed. Philadelphia, Saunders, 1998

Callen JP et al: *Dermatological Signs of Internal Disease* , 4th ed. Philadelphia, Saunders, 2009

Mckee PH et al: *Pathology of the Skin* , 3rd ed. London, Elsevier, 2005

Spitz JL: *Genodermatoses: A Clinical Guide to Genetic Skin Disorders,* 2nd ed. Philadelphia, Lippincott Williams & Wilkins, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

IMMUNOLOGICALLY MEDIATED SKIN DISEASES: INTRODUCTION

A number of immunologically mediated skin diseases and immunologically mediated systemic disorders with cutaneous manifestations are now recognized as distinct entities with consistent clinical, histologic, and immunopathologic findings. Clinically, these disorders are characterized by morbidity (pain, pruritus, disfigurement) and, in some instances, by mortality (largely due to loss of epidermal barrier function and/or secondary infection). The major features of the more common immunologically mediated skin diseases are summarized in this chapter **(Table 54-1)** as are the systemic disorders with cutaneous manifestations.

Table 54-1 Immunologically Mediated Blistering Diseases

Disease	Clinical	Histology	Immunopathology	Autoantigens[a]
Pemphigus foliaceus	Crusts and shallow erosions on scalp, central face, upper chest, and back	Acantholytic blister formed in superficial layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg1
Pemphigus vulgaris	Flaccid blisters, denuded skin, oromucosal lesions	Acantholytic blister formed in suprabasal layer of epidermis	Cell surface deposits of IgG on keratinocytes	Dsg3 (plus Dsg1 in patients with skin involvement)
Paraneoplastic pemphigus	Painful stomatitis with papulosquamous or lichenoid eruptions that progress to blisters	Acantholysis, keratinocyte necrosis and vacuolar interface dermatitis	Cell surface deposits of IgG and C3 on keratinocytes and (variably) similar immunoreactants in epidermal BMZ	Plakin protein family members and desmosomal cadherins (see text for details)
Bullous pemphigoid	Large tense blisters on flexor surfaces and trunk	Subepidermal blister with eosinophil-rich infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	BPAG1, BPAG2
Pemphigoid gestationis	Pruritic, urticarial plaques, rimmed by vesicles and bullae on the trunk and extremities	Teardrop-shaped, subepidermal blisters in dermal papillae; eosinophil-rich infiltrate	Linear band of C3 in epidermal BMZ	BPAG2 (plus BPAG1 in some patients)
Linear IgA disease	Pruritic small papules on extensor surfaces; occasionally larger, arciform blisters	Subepidermal blister with neutrophil-rich infiltrate	Linear band of IgA in epidermal BMZ	BPAG2 (see text for specific details)
Cicatricial pemphigoid	Erosive and/or blistering lesions of mucous membranes and possibly the skin; scarring of some sites	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG, IgA, and/or C3 in epidermal BMZ	BPAG2, laminin-332, or others
Epidermolysis bullosa acquisita	Blisters, erosions, scars, and milia on sites exposed to trauma; widespread, inflammatory, tense blisters may be seen initially	Subepidermal blister that may or may not include a leukocytic infiltrate	Linear band of IgG and/or C3 in epidermal BMZ	Type VII collagen
Dermatitis herpetiformis	Extremely pruritic small papules and vesicles on elbows, knees, buttocks, and posterior neck	Subepidermal blister with neutrophils in dermal papillae	Granular deposits of IgA in dermal papillae	Epidermal transglutaminase

[a] Autoantigens bound by these patients' autoantibodies are defined as follows: Dsg1, desmoglein 1; Dsg3, desmoglein 3; BPAG1, bullous pemphigoid antigen 1; BPAG2, bullous pemphigoid antigen 2.

Abbreviation: BMZ, basement membrane zone.

AUTOIMMUNE CUTANEOUS DISEASES

Pemphigus Vulgaris

Pemphigus refers to a group of autoantibody-mediated intraepidermal blistering diseases characterized by loss of cohesion between epidermal cells (a process termed *acantholysis*). Manual pressure to the skin of these patients may elicit the separation of the epidermis (Nikolsky's sign). This finding, while characteristic of pemphigus, is not specific to this group of disorders and is also seen in toxic epidermal necrolysis, Stevens-Johnson syndrome, and a few other skin diseases.

Pemphigus vulgaris (PV) is a mucocutaneous blistering disease that predominates in patients >40 years. PV typically begins on mucosal surfaces and often progresses to involve the skin. PV is characterized by fragile, flaccid blisters that rupture to produce extensive denudation of mucous membranes and skin **(Fig. 54-1)**. PV typically involves the mouth, scalp, face, neck, axilla, groin, and trunk. PV may be associated with severe skin pain; some patients experience pruritus as well. Lesions usually heal without scarring, except at sites complicated by secondary infection or mechanically induced dermal wounds. Postinflammatory hyperpigmentation is usually present at sites of healed lesions for some time.

Figure 54-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pemphigus vulgaris. ***A.*** Pemphigus vulgaris demonstrating flaccid bullae that are easily ruptured, resulting in multiple erosions and crusted plaques. ***B***. Pemphigus vulgaris almost invariably involves the oral mucosa and may present with erosions involving the gingiva, buccal mucosa, palate, posterior pharynx, or the tongue. (***B,*** *Courtesy of Robert Swerlick, MD;with permission.*)

Biopsies of early lesions demonstrate intraepidermal vesicle formation secondary to loss of cohesion between epidermal cells (i.e., acantholytic blisters). Blister cavities contain acantholytic epidermal cells, which appear as round homogeneous cells containing hyperchromatic nuclei. Basal keratinocytes remain attached to the epidermal basement membrane; hence, blister formation is within the suprabasal portion of the epidermis. Lesional skin may contain focal collections of intraepidermal eosinophils within blister cavities; dermal alterations are slight, often limited to an eosinophil-predominant leukocytic infiltrate. Direct immunofluorescence microscopy of lesional or intact patient skin shows deposits of IgG on the surface of keratinocytes; deposits of complement components are typically found in lesional but not uninvolved skin. Deposits of IgG on keratinocytes are derived from circulating autoantibodies directed against cell-surface autoantigens. Such circulating autoantibodies can be demonstrated in 80–90% of PV patients by indirect immunofluorescence microscopy; monkey esophagus is the optimal substrate for these studies. Patients with PV have IgG autoantibodies directed against *desmogleins* (Dsgs), transmembrane desmosomal glycoproteins that belong to the cadherin family of calcium-dependent adhesion molecules. Such autoantibodies can be precisely quantitated by enzyme-linked immunosorbent assay (ELISA). Patients with early PV (i.e., mucosal disease) have IgG autoantibodies directed against Dsg3; patients with advanced PV (i.e., mucocutaneous disease) have IgG autoantibodies directed against both Dsg3 and Dsg1. Experimental studies have shown that autoantibodies from patients with PV are pathogenic (i.e., responsible for blister formation) and that their titer correlates with disease activity. Recent studies have shown that the anti-Dsg autoantibody profile in these patients' sera as well as the tissue distribution of Dsg3 and Dsg1 determine the site of blister formation in patients with pemphigus. Coexpression of Dsg3 and Dsg1 by epidermal cells protects against pathogenic IgG directed against either cadherin but not pathogenic autoantibodies directed against both.

PV can be life threatening. Prior to the availability of glucocorticoids, the mortality ranged from 60–90%; the current mortality is ~5%. Common causes of morbidity and mortality are infection and complications of treatment with glucocorticoids. Bad prognostic factors include advanced age, widespread involvement, and the requirement for high doses of glucocorticoids (with or without other immunosuppressive agents) for control of disease. The course of PV in individual patients is variable and difficult to predict. Some patients achieve remission, while others may require long-term treatment or succumb to complications of their disease or its treatment. The mainstay of treatment is systemic glucocorticoids. Patients with moderate to severe PV are usually started on prednisone, 1 mg/kg per day. If new lesions continue to appear after 1–2 weeks of treatment, the dose may need to be increased and/or combined with other immunosuppressive agents such as azathioprine (2–2.5 mg/kg per d), mycophenolate mofetil (20–35 mg/kg per d), or cyclophosphamide (1–2 mg/kg per d). Patients with severe, treatment-resistant disease may derive benefit from plasmapheresis [six high-volume exchanges (i.e., 2–3 L per exchange) over ~2 weeks], IV immunoglobulin (IVIg) (2 g/kg over 3–5 days

every 6–8 weeks), or rituximab (375 mg/m^2 per week x 4, or 1,000 mg on days 1 and 15). It is important to bring severe or progressive disease under control quickly to lessen the severity and/or duration of this disorder. Accordingly, some have suggested that rituximab and daily glucocorticoids should be used early in PV patients to avert the development of treatment-resistant disease.

Pemphigus Foliaceus

Pemphigus foliaceus (PF) is distinguished from PV by several features. In PF, acantholytic blisters are located high within the epidermis, usually just beneath the stratum corneum. Hence, PF is a more superficial blistering disease than PV. The distribution of lesions in the two disorders is much the same, except that in PF, mucous membranes are almost always spared. Patients with PF rarely demonstrate intact blisters but rather exhibit shallow erosions associated with erythema, scale, and crust formation. Mild cases of PF resemble severe seborrheic dermatitis; severe PF may cause extensive exfoliation. Sun exposure (ultraviolet irradiation) may be an aggravating factor. *Fogo selvagem* (FS), an endemic form of PF thought to develop as a consequence of environmental stimuli (e.g., insect bites), is found in south central rural Brazil as well as selected sites in Latin America and Tunisia.

Patients with PF have immunopathologic features in common with PV. Specifically, direct immunofluorescence microscopy of perilesional skin demonstrates IgG on the surface of keratinocytes. Similarly, patients with PF have circulating IgG autoantibodies directed against the surface of keratinocytes. In PF, autoantibodies are directed against Dsg1, a 160-kDa desmosomal cadherin. These autoantibodies can be quantitated by ELISA. As noted for PV, the autoantibody profile in patients with PF (i.e., anti-Dsg1 IgG) and the tissue distribution of this autoantigen (i.e., expression in oral mucosa that is compensated by coexpression of Dsg3) are thought to account for the distribution of lesions in this disease.

Although pemphigus has been associated with several autoimmune diseases, its association with thymoma and/or myasthenia gravis is particularly notable. To date, >30 cases of thymoma and/or myasthenia gravis have been reported in association with pemphigus, usually with PF. Patients may also develop pemphigus as a consequence of drug exposure; drug-induced pemphigus usually resembles PF rather than PV. Drugs containing a thiol group in their chemical structure (e.g., penicillamine, captopril, enalapril) are most commonly associated with drug-induced pemphigus. Nonthiol drugs linked to pemphigus include penicillins, cephalosporins, and piroxicam. It has been suggested that thiol- and nonthiol-containing drugs induce pemphigus via biochemical and immunologic mechanisms, respectively. Hence, the better prognosis upon drug withdrawal in cases of pemphigus induced by thiol-containing medications. Some cases of drug-induced pemphigus are durable and require treatment with systemic glucocorticoids and/or immunosuppressive agents.

PF is generally a less severe disease than PV and carries a better prognosis. Localized disease can sometimes be treated with topical or intralesional glucocorticoids; more active cases can usually be controlled with systemic glucocorticoids. Patients with severe, treatment-resistant disease may require more aggressive interventions as described above for patients with PV.

Paraneoplastic Pemphigus

Paraneoplastic pemphigus (PNP) is an autoimmune acantholytic mucocutaneous disease associated with an occult or confirmed neoplasm. Patients with PNP typically show painful mucosal erosive lesions in association with papulosquamous and/or lichenoid eruptions that often progress to blisters. Palm and sole involvement are common in these patients and raise the possibility that prior reports of neoplasia-associated erythema multiforme actually may have represented unrecognized cases of PNP. Biopsies of lesional skin from these patients show varying combinations of acantholysis, keratinocyte necrosis, and vacuolar-interface dermatitis. Direct immunofluorescence microscopy of patient's skin shows deposits of IgG and complement on the surface of keratinocytes and (variably) similar immunoreactants in the epidermal basement membrane zone. Patients with PNP have IgG autoantibodies against cytoplasmic proteins that are members of the plakin family (e.g., desmoplakins I and II, bullous pemphigoid antigen 1, envoplakin, periplakin, and plectin) and cell-surface proteins that are members of the cadherin family (e.g., Dsg1 and Dsg3). Passive transfer studies have shown that autoantibodies from patients with PNP are pathogenic.

The predominant neoplasms associated with PNP are non-Hodgkin's lymphoma, chronic lymphocytic leukemia, thymoma, spindle cell tumors, Waldenström's macroglobulinemia, and Castleman's disease; the latter is particularly common among children with PNP. Rare cases of seronegative PNP have been reported in patients with B cell malignancies previously treated with rituximab. In addition to severe skin lesions, many patients with PNP develop life-threatening bronchiolitis obliterans. PNP is generally resistant to conventional therapies (i.e., those used to treat PV); rarely patients may improve (or even remit) following ablation or removal of underlying neoplasms.

Bullous Pemphigoid

Bullous pemphigoid (BP) is a polymorphic autoimmune subepidermal blistering disease usually seen in the elderly. Initial lesions may consist of urticarial plaques; most patients eventually display tense

blisters on either normal-appearing or erythematous skin **(Fig. 54-2)**. The lesions are usually distributed over the lower abdomen, groin, and flexor surface of the extremities; oral mucosal lesions are found in some patients. Pruritus may be nonexistent or severe. As lesions evolve, tense blisters tend to rupture and be replaced by erosions with or without surmounting crust. Nontraumatized blisters heal without scarring. The major histocompatibility complex class II allele HLA-DQβ1*0301 is prevalent in patients with BP. Despite isolated reports, several studies have shown that patients with BP do not have an increased incidence of malignancy in comparison with appropriately age- and gender-matched controls.

Figure 54-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bullous pemphigoid with tense vesicles and bullae on erythematous, urticarial bases. (*Courtesy of the Yale Resident's Slide Collection; with permission*).

Biopsies of early lesional skin demonstrate subepidermal blisters and histologic features that roughly correlate with the clinical character of the particular lesion under study. Lesions on normal-appearing skin generally show a sparse perivascular leukocytic infiltrate with some eosinophils; conversely, biopsies of inflammatory lesions typically show an eosinophil-rich infiltrate at sites of vesicle formation and in perivascular areas. In addition to eosinophils, cell-rich lesions also contain mononuclear cells and neutrophils. It is not possible to distinguish BP from other subepidermal blistering diseases by routine histologic studies alone.

Direct immunofluorescence microscopy of normal-appearing perilesional skin from patients with BP shows linear deposits of IgG and/or C3 in the epidermal basement membrane. The sera of ~70% of these patients contain circulating IgG autoantibodies that bind the epidermal basement membrane of normal human skin in indirect immunofluorescence microscopy. IgG from an even higher percentage of patients shows reactivity to the epidermal side of 1 *M* NaCl split skin [an alternative immunofluorescence microscopy test substrate used to distinguish circulating IgG anti-basement membrane autoantibodies in patients with BP from those in patients with similar, yet different, subepidermal blistering diseases (see below)]. In BP, circulating autoantibodies recognize 230- and 180-kDa hemidesmosome-associated proteins in basal keratinocytes [i.e., bullous pemphigoid antigen (BPAG)1 and BPAG2, respectively]. Autoantibodies against BPAG2 are thought to deposit in situ, activate complement, produce dermal mast cell degranulation, and generate granulocyte-rich infiltrates that cause tissue damage and blister formation.

BP may persist for months to years, with exacerbations or remissions. Extensive involvement may result in widespread erosions and compromise cutaneous integrity; deaths may occur in elderly and/or debilitated patients. The mainstay of treatment is systemic glucocorticoids. Patients with local or minimal disease can sometimes be controlled with topical glucocorticoids alone; patients with more extensive lesions generally respond to systemic glucocorticoids either alone or in combination with immunosuppressive agents. Patients will usually respond to prednisone, 0.75–1 mg/kg per d. In some

instances, azathioprine (2–2.5 mg/kg per d), mycophenolate mofetil (20–35 mg/kg per d), or cyclophosphamide (1–2 mg/kg per d) are necessary adjuncts.

Pemphigoid Gestationis

Pemphigoid gestationis (PG), also known as *herpes gestationis*, is a rare, nonviral, subepidermal blistering disease of pregnancy and the puerperium. PG may begin during any trimester of pregnancy or present shortly after delivery. Lesions are usually distributed over the abdomen, trunk, and extremities; mucous membrane lesions are rare. Skin lesions in these patients may be quite polymorphic and consist of erythematous urticarial papules and plaques, vesiculopapules, and/or frank bullae. Lesions are almost always very pruritic. Severe exacerbations of PG frequently occur after delivery, typically within 24–48 hours. PG tends to recur in subsequent pregnancies, often beginning earlier during such gestations. Brief flare-ups of disease may occur with resumption of menses and may develop in patients later exposed to oral contraceptives. Occasionally, infants of affected mothers demonstrate transient skin lesions.

Biopsies of early lesional skin show teardrop-shaped subepidermal vesicles forming in dermal papillae in association with an eosinophil-rich leukocytic infiltrate. Differentiation of PG from other subepidermal bullous diseases by light microscopy is difficult. However, direct immunofluorescence microscopy of perilesional skin from PG patients reveals the immunopathologic hallmark of this disorder—linear deposits of C3 in epidermal basement membrane. These deposits develop as a consequence of complement activation produced by low titer IgG anti-basement membrane autoantibodies directed against BPAG2, the same hemidesmosome-associated protein that is targeted by autoantibodies in patients with BP—a subepidermal bullous disease that resembles PG clinically, histologically, and immunopathologically.

The goals of therapy in patients with PG are to prevent the development of new lesions, relieve intense pruritus, and care for erosions at sites of blister formation. Many patients require treatment with moderate doses of daily glucocorticoids (i.e., 20–40 mg prednisone) at some point in their course. Mild cases (or brief flare-ups) may be controlled by vigorous use of potent topical glucocorticoids. Infants born of mothers with PG appear to be at increased risk of being slightly premature or "small for dates." Current evidence suggests that there is no difference in the incidence of uncomplicated live births in PG patients treated with systemic glucocorticoids and in those managed more conservatively. If systemic glucocorticoids are administered, newborns are at risk for development of reversible adrenal insufficiency.

Dermatitis Herpetiformis

Dermatitis herpetiformis (DH) is an intensely pruritic, papulovesicular skin disease characterized by lesions symmetrically distributed over extensor surfaces (i.e., elbows, knees, buttocks, back, scalp, and posterior neck) (see Fig. 51-8). Primary lesions in this disorder consist of papules, papulovesicles, or urticarial plaques. Because pruritus is prominent, patients may present with excoriations and crusted papules but no observable primary lesions. Patients sometimes report that their pruritus has a distinctive burning or stinging component; the onset of such local symptoms reliably heralds the development of distinct clinical lesions 12–24 hours later. Almost all DH patients have an associated, usually subclinical, gluten-sensitive enteropathy (Chap. 294), and >90% express the HLA-B8/DRw3 and HLA-DQw2 haplotypes. DH may present at any age, including childhood; onset in the second to fourth decades is most common. The disease is typically chronic.

Biopsy of early lesional skin reveals neutrophil-rich infiltrates within dermal papillae. Neutrophils, fibrin, edema, and microvesicle formation at these sites are characteristic of early disease. Older lesions may demonstrate nonspecific features of a subepidermal bulla or an excoriated papule. Because the clinical and histologic features of this disease can be variable and resemble other subepidermal blistering disorders, the diagnosis is confirmed by direct immunofluorescence microscopy of normal-appearing perilesional skin. Such studies demonstrate granular deposits of IgA (with or without complement components) in the papillarydermis and along the epidermal basement membrane zone. IgA deposits in the skin are unaffected by control of disease with medication; however, these immunoreactants diminish in intensity or disappear in patients maintained for long periods on a strict gluten-free diet (see below). Patients with DH have granular deposits of IgA in their epidermal basement membrane zone and should be distinguished from individuals with linear IgA deposits at this site (see below).

Although most DH patients do not report overt gastrointestinal symptoms or have laboratory evidence of malabsorption, biopsies of the small bowel usually reveal blunting of intestinal villi and a lymphocytic infiltrate in the lamina propria. As is true for patients with celiac disease, this gastrointestinal abnormality can be reversed by a gluten-free diet. Moreover, if maintained, this diet alone may control the skin disease and eventuate in clearance of IgA deposits from these patients' epidermal basement membrane zones. Subsequent gluten exposure in such patients alters the morphology of their small bowel, elicits a flare-up of their skin disease, and is associated with the reappearance of IgA in their epidermal basement membrane zones. As in patients with celiac disease, dietary gluten sensitivity in patients with DH is associated with IgA antiendomysial autoantibodies that target tissue transglutaminase. Recent studies indicate that patients with DH also have high-avidity IgA autoantibodies against epidermal transglutaminase 3 and that the latter is co-localized with granular

deposits of IgA in the papillarydermis of DH patients. Patients with DH also have an increased incidence of thyroid abnormalities, achlorhydria, atrophic gastritis, and antigastric parietal cell autoantibodies. These associations likely relate to the high frequency of the HLA-B8/DRw3 haplotype in these patients, because this marker is commonly linked to autoimmune disorders. The mainstay of treatment of DH is dapsone, a sulfone. Patients respond rapidly (24–48 hours) to dapsone (50–200 mg/d), but require careful pretreatment evaluation and close follow-up to ensure that complications are avoided or controlled. All patients on >100 mg/d dapsone will have some hemolysis and methemoglobinemia, which are expected pharmacologic side effects of this agent. Gluten restriction can control DH and lessen dapsone requirements; this diet must rigidly exclude gluten to be of maximal benefit. Many months of dietary restriction may be necessary before a beneficial result is achieved. Good dietary counseling by a trained dietitian is essential.

Linear IgA Disease

Linear IgA disease, once considered a variant form of dermatitis herpetiformis, is actually a separate and distinct entity. Clinically, these patients may resemble individuals with DH, BP, or other subepidermal blistering diseases. Lesions typically consist of papulovesicles, bullae, and/or urticarial plaques predominantly on central or flexural sites. Oral mucosal involvement occurs in some patients. Severe pruritus resembles that seen in patients with DH. Patients with linear IgA disease do not have an increased frequency of the HLA-B8/DRw3 haplotype or an associated enteropathy and, hence, are not candidates for treatment with a gluten-free diet.

Histologic alterations in early lesions may be virtually indistinguishable from those in DH. However, direct immunofluorescence microscopy of normal-appearing perilesional skin reveals linear deposits of IgA (and often C3) in the epidermal basement membrane zone. Most patients with linear IgA disease demonstrate circulating IgA anti-basement membrane autoantibodies directed against neoepitopes in the proteolytically processed extracellular domain of BPAG2. These patients generally respond to treatment with dapsone, 50–200 mg/d.

Epidermolysis Bullosa Acquisita

Epidermolysis bullosa acquisita (EBA) is a rare, noninherited, polymorphic, chronic, subepidermal blistering disease. (The inherited form is discussed in Chap. 363.) Patients with classic or noninflammatory EBA have blisters on noninflamed skin, atrophic scars, milia, nail dystrophy, and oral lesions. Because lesions generally occur at sites exposed to minor trauma, classic EBA is considered to be a mechanobullous disease. Other patients with EBA have widespread inflammatory scarring and bullous lesions that resemble severe BP. Inflammatory EBA may evolve into the classic, noninflammatory form of this disease. Rarely patients present with lesions that predominate on mucous membranes. The HLA-DR2 haplotype is found with increased frequency in EBA patients. Recent studies suggest that EBA is sometimes associated with inflammatory bowel disease (especially Crohn's disease).

The histology of lesional skin varies depending on the character of the lesion being studied. Noninflammatory bullae show subepidermal blisters with a sparse leukocytic infiltrate and resemble those in patients with porphyria cutanea tarda. Inflammatory lesions consist of neutrophil-rich subepidermal blisters. EBA patients have continuous deposits of IgG (and frequently C3) in a linear pattern within the epidermal basement membrane zone. Ultrastructurally, these immunoreactants are found in the sublamina densa region in association with anchoring fibrils. Approximately 50% of EBA patients have demonstrable circulating IgG anti-basement membrane autoantibodies directed against type VII collagen—the collagen species that comprises anchoring fibrils. Such IgG autoantibodies bind the dermal side of 1 *M* NaCl split skin (in contrast to IgG autoantibodies in patients with BP). Studies have shown that passive transfer of experimental or patient IgG directed against type VII collagen can produce lesions in mice that clinically, histologically, and immunopathologically resemble those seen in patients with inflammatory EBA.

Treatment of EBA is generally unsatisfactory. Some patients with inflammatory EBA may respond to systemic glucocorticoids, either alone or in combination with immunosuppressive agents. Other patients (especially those with neutrophil-rich inflammatory lesions) may respond to dapsone. The chronic, noninflammatory form of this disease is largely resistant to treatment, although some patients may respond to cyclosporine, azathioprine, or IVIg.

Cicatricial Pemphigoid

Cicatricial pemphigoid (CP) is a rare, acquired, subepithelial immunobullous disease characterized by erosive lesions of mucous membranes and skin that result in scarring of at least some sites of involvement. Common sites of involvement include the oral mucosa (especially the gingiva) and conjunctiva; other sites that may be affected include the nasopharyngeal, laryngeal, esophageal, and anogenital mucosa. Skin lesions (present in about one-third of patients) tend to predominate on the scalp, face, and upper trunk and generally consist of a few scattered erosions or tense blisters on an erythematous or urticarial base. CP is typically a chronic and progressive disorder. Serious complications may arise as a consequence of ocular, laryngeal, esophageal, or anogenital lesions. Erosive conjunctivitis may result in shortened fornices, symblephara, ankyloblepharon, entropion,

corneal opacities, and (in severe cases) blindness. Similarly, erosive lesions of the larynx may cause hoarseness, pain, and tissue loss that, if unrecognized and untreated, may eventuate in complete destruction of the airway. Esophageal lesions may result in stenosis and/or strictures that could place patients at risk for aspiration. Strictures might also complicate anogenital involvement.

Biopsies of lesional tissue generally demonstrate subepithelial vesiculobullae and a mononuclear leukocytic infiltrate. Neutrophils and eosinophils may be seen in biopsies of early lesions; older lesions may demonstrate a scant leukocytic infiltrate and fibrosis. Direct immunofluorescence microscopy of perilesional tissue typically demonstrates deposits of IgG, IgA, and/or C3 in the epidermal basement membrane of these patients. Because many of these patients show no evidence of circulating anti-basement membrane autoantibodies, testing of perilesional skin is important diagnostically. Although CP was once thought to be a single nosologic entity, it is now largely regarded as a disease phenotype that may develop as a consequence of an autoimmune reaction against a variety of different molecules in epidermal basement membrane (e.g., BPAG2, laminin-332, type VII collagen, and other antigens yet to be completely defined). Recent studies suggest that CP patients with anti-laminin-332 autoantibodies have an increased relative risk for cancer. Treatment of CP is largely dependent upon sites of involvement. Due to potentially severe complications, patients with ocular, laryngeal, esophageal, and/or anogenital involvement require aggressive systemic treatment with dapsone, prednisone, or the latter in combination with another immunosuppressive agent (e.g., azathioprine, mycophenolate mofetil, or cyclophosphamide) or IVIg. Less threatening forms of the disease may be managed with topical or intralesional glucocorticoids.

AUTOIMMUNE SYSTEMIC DISEASES WITH PROMINENT CUTANEOUS FEATURES

Dermatomyositis

The cutaneous manifestations of dermatomyositis (Chap. 388) are often distinctive, but, at times, they may resemble those of systemic lupus erythematosus (SLE) (Chap. 319), scleroderma (Chap. 323), or other overlapping connective tissue diseases (Chap. 323). The extent and severity of cutaneous disease may or may not correlate with the extent and severity of the myositis. The cutaneous manifestations of dermatomyositis are similar, whether the disease appears in children or the elderly, except that calcification of subcutaneous tissue is a common late sequela in childhood dermatomyositis.

The cutaneous signs of dermatomyositis may precede or follow the development of myositis by weeks to years. Cases lacking muscle involvement (i.e., dermatomyositis sine myositis) have also been reported. The most common manifestation is a purple-red discoloration of the upper eyelids, sometimes associated with scaling ("heliotrope" erythema; **Fig. 54-3**) and periorbital edema. Erythema on the cheeks and nose in a "butterfly" distribution may resemble the malar eruption of SLE. Erythematous or violaceous scaling patches are common on the upper anterior chest; posterior neck; scalp; and the extensor surfaces of the arms, legs, and hands. Erythema and scaling may be particularly prominent over the elbows, knees, and the dorsal interphalangeal joints. Approximately one-third of patients have violaceous, flat-topped papules over the dorsal interphalangeal joints that are pathognomonic of dermatomyositis (Gottron's sign or Gottron's papules; **Fig. 54-4**). These lesions can be contrasted with the erythema and scaling on the dorsum of the fingers that spares the skin over the interphalangeal joints of some SLE patients. Periungual telangiectasia may be prominent. Lacy or reticulated erythema may be associated with fine scaling on the extensor surfaces of the thighs and upper arms. Other patients, particularly those with long-standing disease, develop areas of hypopigmentation, hyperpigmentation, mild atrophy, and telangiectasia known as *poikiloderma*. Poikiloderma is rare in both SLE and scleroderma and, thus, can serve as a clinical sign that distinguishes dermatomyositis from these two diseases. Cutaneous changes may be similar in scleroderma and dermatomyositis and may include thickening and binding down of the skin of the hands (sclerodactyly) as well as Raynaud's phenomenon. However, the presence of severe muscle disease, Gottron's papules, heliotrope erythema, and poikiloderma serve to distinguish patients with dermatomyositis. Skin biopsy of erythematous, scaling lesions of dermatomyositis may reveal only mild nonspecific inflammation but sometimes may show changes indistinguishable from those found in SLE, including epidermal atrophy, hydropic degeneration of basal keratinocytes, edema of the upper dermis, and a mild mononuclear cell infiltrate. Direct immunofluorescence microscopy of lesional skin is usually negative, although granular deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone have been described in some patients. Treatment should be directed at the systemic disease. Topical glucocorticoids are sometimes useful; patients should avoid exposure to ultraviolet irradiation and aggressively use photoprotective measures including use of broad-spectrum sunscreens.

Figure 54-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dermatomyositis. Periorbital violaceous erythema characterizes the classic heliotrope rash. (*Courtesy of James Krell, MD; with permission.*)

Figure 54-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gottron's sign. Dermatomyositis often involves the hands as erythematous flat-topped papules over the knuckles (Gottron's sign). Periungual telangiectases are also evident.

Lupus Erythematosus

The cutaneous manifestations of lupus erythematosus (LE) (Chap. 319) can be divided into acute, subacute, and chronic types. *Acute cutaneous LE* is characterized by erythema of the nose and malar eminences in a "butterfly" distribution **(Fig. 54-5)**. The erythema is often sudden in onset, accompanied by edema and fine scale, and correlated with systemic involvement. Patients may have widespread involvement of the face as well as erythema and scaling of the extensor surfaces of the extremities and upper chest. These acute lesions, while sometimes evanescent, usually last for days and are often associated with exacerbations of systemic disease. Skin biopsy of acute lesions may show only a sparse dermal infiltrate of mononuclear cells and dermal edema. In some instances, cellular infiltrates around blood vessels and hair follicles are notable, as is hydropic degeneration of basal cells of the epidermis. Direct immunofluorescence microscopy of lesional skin frequently reveals deposits of immunoglobulin(s) and complement in the epidermal basement membrane zone. Treatment is aimed at

control of systemic disease; photoprotection in this as well as in other forms of LE is very important.

Figure 54-5

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. **Acute cutaneous lupus erythematosus** showing prominent, scaly, malar erythema. Involvement of other sun-exposed sites is also common. ***B.*** **Acute cutaneous LE** on the upper chest demonstrating brightly erythematous and slightly edematous papules and plaques. (***B,*** *Courtesy of Robert Swerlick, MD; with permission.*)

Subacutecutaneous lupus erythematosus (SCLE) is characterized by a widespread photosensitive, nonscarring eruption. Most of these patients have SLE in which renal and CNS involvement is mild or absent. SCLE may present as a papulosquamous eruption that resembles psoriasis or annular lesions that resemble those seen in erythema multiforme. In the papulosquamous form, discrete erythematous papules arise on the back, chest, shoulders, extensor surfaces of the arms, and the dorsum of the hands; lesions are uncommon on the face, flexor surfaces of the arms, and below the waist. These slightly scaling papules tend to merge into large plaques, some with a reticulate appearance. The annular form involves the same areas and presents with erythematous papules that evolve into oval, circular, or polycyclic lesions. The lesions of SCLE are more widespread but have less tendency for scarring than do lesions of discoid LE. Skin biopsy reveals a dense mononuclear cell infiltrate around

hair follicles and blood vessels in the superficial dermis, combined with hydropic degeneration of basal cells in the epidermis. Direct immunofluorescence microscopy of lesional skin reveals deposits of immunoglobulin(s) in the epidermal basement membrane zone in about one-half of these cases. A particulate pattern of IgG deposition throughout the epidermis has been associated with SCLE. Most SCLE patients have anti-Ro autoantibodies. Local therapy alone is usually unsuccessful. Most patients require treatment with aminoquinoline antimalarials. Low-dose therapy with oral glucocorticoids is sometimes necessary. Photoprotective measures against both ultraviolet B and A wavelengths are very important.

Discoid lupus erythematosus (DLE, also called *chronic cutaneous LE*) is characterized by discrete lesions, most often found on the face, scalp, and/or external ears. The lesions are erythematous papules or plaques with a thick, adherent scale that occludes hair follicles (follicular plugging). When the scale is removed, its underside shows small excrescences that correlate with the openings of hair follicles (so-called "carpet tacking"), a finding relatively specific for DLE. Long-standing lesions develop central atrophy, scarring, and hypopigmentation but frequently have erythematous, sometimes raised borders **(Fig. 54-6)**. These lesions persist for years and tend to expand slowly. Only 5–10% of patients with DLE meet the American Rheumatism Association criteria for SLE. However, typical discoid lesions are frequently seen in patients with SLE. Biopsy of DLE lesions shows hyperkeratosis, follicular plugging, atrophy of the epidermis, hydropic degeneration of basal keratinocytes, and a mononuclear cell infiltrate adjacent to epidermal, adnexal, and microvascular basement membranes. Direct immunofluorescence microscopy demonstrates immunoglobulin(s) and complement deposits at the basement membrane zone in ~90% of cases. Treatment is focused on control of local cutaneous disease and consists mainly of photoprotection and topical or intralesional glucocorticoids. If local therapy is ineffective, use of aminoquinoline antimalarials may be indicated.

Figure 54-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Discoid lupus erythematosus. Violaceous, hyperpigmented, atrophic plaques, often with evidence of follicular plugging, which may result in scarring, are characteristic of discoid lupus erythematosus (also called chronic cutaneous lupus erythematosus).

Scleroderma and Morphea

The skin changes of scleroderma (Chap. 323) usually begin on the hands, feet, and face, with episodes of recurrent nonpitting edema. Sclerosis of the skin begins distally on the fingers (sclerodactyly) and spreads proximally, usually accompanied by resorption of bone of the fingertips, which may have punched out ulcers, stellate scars, or areas of hemorrhage **(Fig. 54-7)**. The fingers may actually shrink in size and become sausage-shaped, and because the fingernails are usually unaffected, the nails may curve over the end of the fingertips. Periungual telangiectases are usually present, but periungual erythema is rare. In advanced cases, the extremities show contractures and calcinosis cutis. Facial involvement includes a smooth, unwrinkled brow, taut skin over the nose, shrinkage of tissue around the mouth, and perioral radial furrowing **(Fig. 54-8)**. Matlike telangiectases

are often present, particularly on the face and hands. Involved skin feels indurated, smooth, and bound to underlying structures; hyper- and hypopigmentation are also often present. Raynaud's phenomenon (i.e., cold-induced blanching, cyanosis, and reactive hyperemia) is present in almost all patients and can precede development of scleroderma by many years. Linear scleroderma is a limited form of disease that presents in a linear, bandlike distribution and tends to involve deep as well as superficial layers of skin. The combination of calcinosis cutis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia has been termed the *CREST syndrome*. Anticentromere antibodies have been reported in a very high percentage of patients with the CREST syndrome but in only a small minority of patients with scleroderma. Skin biopsy reveals thickening of the dermis and homogenization of collagen bundles. Direct immunofluorescence microscopy of lesional skin is usually negative.

Figure 54-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scleroderma showing acral sclerosis and focal digital ulcers.

Figure 54-8

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scleroderma often eventuates in development of an expressionless, masklike facies.

Morphea is characterized by localized thickening and sclerosis of skin; it dominates on the trunk. This disorder may affect children or adults. Morphea begins as erythematous or flesh-colored plaques that become sclerotic, develop central hypopigmentation, and demonstrate an erythematous border. In most cases, patients have one or a few lesions, and the disease is termed *localized morphea*. In some patients, widespread cutaneous lesions may occur without systemic involvement. This form is called *generalized morphea*. Many adults with generalized morphea have concomitant rheumatic or other autoimmune disorders. Skin biopsy of morphea is indistinguishable from that of scleroderma. Scleroderma and morphea are usually quite resistant to therapy. For this reason, physical therapy to prevent joint contractures and to maintain function is employed and is often helpful. Treatment options for early, rapidly progressive disease include phototherapy (UVA1 or PUVA) or methotrexate 15–20 mg per week alone or in combination with daily glucocorticoids.

Diffuse fasciitis with eosinophilia is a clinical entity that can sometimes be confused with scleroderma. There is usually the sudden onset of swelling, induration, and erythema of the extremities frequently following significant physical exertion. The proximal portions of extremities (arms, forearms, thighs, legs) are more often involved than are the hands and feet. While the skin is indurated, it is usually not bound down as in scleroderma; contractures may occur early secondary to fascial involvement. The latter may also cause muscle groups to be separated and veins to appear depressed (i.e., the "groove sign"). These skin findings are accompanied by peripheral blood eosinophilia, increased erythrocyte sedimentation rate, and sometimes hypergammaglobulinemia. Deep biopsy of affected areas of skin reveals inflammation and thickening of the deep fascia overlying muscle. An inflammatory infiltrate composed of eosinophils and mononuclear cells is usually found. Patients with eosinophilic fasciitis appear to be at increased risk to develop bone marrow failure or other hematologic abnormalities. While the ultimate course of eosinophilic fasciitis is uncertain, many patients respond favorably to treatment with prednisone in doses ranging from 40–60 mg/d.

The *eosinophilia-myalgia syndrome*, a disorder reported in epidemic numbers in 1989 and linked to ingestion of l-tryptophan manufactured by a single company in Japan, is a multisystem disorder characterized by debilitating myalgias and absolute eosinophilia in association with varying combinations of arthralgias, pulmonary symptoms, and peripheral edema. In a later phase (3–6 months after initial symptoms), these patients often develop localized sclerodermatous skin changes, weight loss, and/or neuropathy (Chap. 323). The precise cause of this syndrome, which may resemble other sclerotic skin conditions, is unknown. However, the implicated lots of l-tryptophan contained the contaminant 1,1-ethylidene bis[tryptophan]. This contaminant may be pathogenic or a marker for another substance that provokes the disorder.

FURTHER READINGS

Anhalt GJ: Paraneoplastic pemphigus. J Investig Dermatol Symp Proc 9:29, 2004 [PMID: 14870982]

Olasz EB, Yancey KB: Bullous pemphigoid and related subepidermal autoimmune blistering diseases, in *Current Directions in Autoimmunity*: *Dermatologic Immunity*, BJ Nickoloff, FO Nestle (eds). Basel, Karger Press, 2008, pp 141-166

Payne AS et al: Desmosomes and disease: Pemphigus and bullous impetigo. Curr Opin Cell Biol 16:536, 2004 [PMID: 15363804]

Sontheimer RD: Skin manifestations of systemic autoimmune connective tissue disease: Diagnostics and therapeutics. Best Pract Res Clin Rheumatol 18:429, 2004 [PMID: 15158749]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CUTANEOUS DRUG REACTIONS: INTRODUCTION

Cutaneous reactions are among the most frequent adverse reactions to drugs. Most are benign, but a few can be life threatening. Prompt recognition of severe reactions, drug withdrawal, and appropriate therapeutic interventions can minimize toxicity. This chapter focuses on adverse cutaneous reactions to systemic medications; it covers their incidence, patterns, and pathogenesis and provides some practical guidelines on treatment, assessment of causality, and future use of drugs.

USE OF PRESCRIPTION DRUGS IN THE UNITED STATES

In the United States, more than 3 billion prescriptions for over 60,000 drug products, which include more than 2000 different active agents, are dispensed annually. Hospital inpatients alone annually receive about 120 million courses of drug therapy, and half of adult Americans receive prescription drugs on a regular outpatient basis. Many patients use over-the-counter medicines that may cause adverse cutaneous reactions.

INCIDENCE OF CUTANEOUS REACTIONS

Several large cohort studies established that acute cutaneous reaction to drugs affected about 3% of hospital inpatients. Reactions usually occur a few days to 4 weeks after initiation of therapy.

Many drugs of common use are associated with a 1–2% rate of "rashes" during premarketing clinical trials. The risk is often higher when medications are used in general, unselected populations. The rate may reach 3–7% for amoxicillin, sulfamethoxazole, and many anticonvulsants. It may be even higher with anti-HIV agents.

In addition to acute eruptions, a variety of skin diseases can be induced or exacerbated by prolonged use of drugs (e.g., pruritus, pigmentation, nail or hair disorders, psoriasis, pemphigoid, and pemphigus). These drug reactions are not frequent, but neither their incidence nor their impact on public health has been evaluated.

In a series of 48,005 inpatients over a 20-year period, morbilliform rash (91%) and urticaria (6%) were the most frequent skin reactions. Severe reactions are actually too rare to be detected in such cohorts. Although rare, severe cutaneous reactions to drugs have an important impact on health and on the risk-versus-benefit evaluation of medicines because of significant sequelae, including mortality. In one prospective study in the hospital setting, adverse drug rash was responsible for hospitalization, increased the duration of hospital stay, or was life threatening. Some populations are at increased risk of drug reactions: patients with collagen vascular diseases, bone marrow graft recipients, and those with acute Epstein-Barr virus infection. The pathophysiology underlying this association is unknown. It has also been established that HIV infection increases the risk of drug allergy, including severe hypersensitivity reactions (Chap. 189). This was true for many drugs but has been evaluated mainly with sulfamethoxazole. Up to 40% of HIV-infected patients had skin reactions when treated with high doses, and about 15% reacted to the same dosage that induced eruption in 3–5% of non-HIV-infected populations. How HIV promotes sensitivity to certain medications or their metabolites remains unclear.

PATHOGENESIS OF DRUG REACTIONS

The skin is commonly affected by adverse drug reactions. The list of conditions that can be triggered by medications includes nearly all dermatologic diseases. Adverse cutaneous responses to drugs can arise as a result of immunologic or nonimmunologic mechanisms. Examples of responses that arise from nonimmunologic mechanisms are pigmentary changes related to accumulation in the dermis of amiodarone, antimalarials, minocycline, quinolones, alteration of hair follicles by antimetabolites, and lipodystrophy associated with metabolic effects of anti-HIV medications. These side effects are mostly toxic, predictable, and often can be avoided in part by simple preventive measures.

Immunologic Drug Reactions

Evidence suggests an immunologic basis for most acute drug eruptions, benign or severe. Drug-specific T cell clones can be derived from the blood or from skin lesions of patients with a variety of drug allergies, strongly suggesting that drugs can be recognized as antigens by human T cells and that these T cells play a role in drug allergy. Specific clones were obtained with penicillin G, amoxicillin, cephalosporins, sulfamethoxazole, phenobarbital, carbamazepine, lamotrigine (i.e., many of the medications that are frequently a cause of drug eruptions). Both CD4 and CD8 clones have been

obtained; however, their specific roles in the manifestations of allergy have not been elucidated. Drug presentation to T cells was MHC-restricted and may involve hapten-peptide complexes formed between drugs or reactive metabolites and endogenous cell-surface proteins.

Once a drug has induced an immune response, the final phenotype of the reaction probably depends on the nature of effectors: cytotoxic (CD8+) T cells in blistering and certain hypersensitivity reactions, chemokines for reactions mediated by neutrophils or eosinophils, and collaboration with B cells for production of specific antibodies for urticarial reactions.

IMMEDIATE REACTIONS

Immediate reactions depend on the release of mediators of inflammation by tissue mast cells or circulating basophilic leukocytes. These mediators include histamine, leukotrienes, prostaglandins, platelet-activating factor, enzymes, and proteoglycans. Drugs can trigger mediator release either directly ("anaphylactoid" reaction) or through IgE-specific antibodies. These reactions usually manifest in the skin and gastrointestinal, respiratory, and cardiovascular systems (Chap. 317). Primary symptoms and signs include pruritus, urticaria, nausea, vomiting, abdominal cramps, bronchospasm, laryngeal edema, and, occasionally, anaphylactic shock with hypotension and death. They occur within minutes of drug exposure. Nonsteroidal anti-inflammatory drugs (NSAIDs), including aspirin, and radiocontrast media are frequent causes of direct mast cell degranulation or anaphylactoid reactions, which can occur on first exposure. Penicillins and muscle relaxants used in general anesthesia are the most frequent causes of IgE-dependent reactions to drugs, which require prior sensitization. Release of mediators is triggered when polyvalent drug protein conjugates cross-link IgE molecules fixed to sensitized cells. Certain routes of administration favor different clinical patterns (e.g., gastrointestinal effects from oral route, circulatory effects from intravenous route).

IMMUNE COMPLEX–DEPENDENT REACTIONS

Serum sickness is produced by tissue deposition of circulating immune complexes with consumption of complement. It is characterized by fever, arthritis, nephritis, neuritis, edema, and an urticarial, papular, or purpuric rash (Chap. 326). First described following administration of nonhuman sera, it currently occurs in the setting of monoclonal antibodies and other similar medications. In classic serum sickness, symptoms develop 6 days or more after exposure to a drug, the latent period representing the time needed to synthesize antibody. Cephalosporin administration may be associated with a clinically similar "serum sickness-like" reaction in children. The mechanism of this reaction is unknown but is unrelated to complement activation and lacks immune complex formation, vasculitis, or renal disease.

Cutaneous or systemic vasculitis, a relatively rare cutaneous complication of drugs, may also be a result of immune complex deposition (Chap. 326).

DELAYED HYPERSENSITIVITY

Delayed hypersensitivity directed by drug-specific T cells is probably the most important mechanism in the etiology of the most common drug eruptions—morbilliform eruptions—and also of rare and severe forms such as hypersensitivity syndrome, acute generalized exanthematous pustulosis (AGEP), Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN). Drug-specific T cells have been detected in these types of drug eruptions. It remains unknown why T cell stimulation by medications leads to reactions that are clinically so diverse.

Contrary to what has been believed for years, the antigen is more often the native drug itself than its metabolites. Drug-specific cytotoxic T cells have been detected in the skin lesions of fixed drug eruptions and of TEN. In TEN, blisters that result from accumulation of interstitial fluid under the necrotic epidermis contain T lymphocytes reactive to autologous lymphocytes and keratinocytes in a drug-specific, HLA-restricted, and perforin/granzyme-mediated pathway.

Drug-specific clones producing CXCL8, a neutrophil-attracting chemokine, were obtained from skin biopsies of patients with AGEP, a neutrophil-mediated drug reaction.

GENETIC FACTORS AND CUTANEOUS DRUG REACTIONS

Genetic determinants may predispose individuals to severe drug reactions by affecting either drug metabolism or immune responses to drugs. Polymorphisms in cytochrome P450 enzymes may increase susceptibility to drug toxicity, highlighting a role for differential pharmacokinetic or pharmacodynamic effects. It has also been suspected that a slow acetylator phenotype increases the risk of rash from sulfonamides. However, in two large prospective cohorts of HIV-infected patients treated with sulfonamides, no association of drug eruption with acetylation genotype was found.

Associations between drug hypersensitivities and HLAhaplotypes also suggest a key role for immune mechanisms. Hypersensitivity to the anti-HIV medication abacavir is strongly associated with HLA B*5701. In Taiwan, within a homogeneous Han Chinese population, a 100% association was observed between SJS or TEN related to carbamazepine and HLA B*1502. In the same population, another 100% association was found between SJS, TEN, or hypersensitivity syndrome/drug reaction with eosinophilia and systemic symptoms (DRESS) related to allopurinol and HLA B*5801. Additional adverse cutaneous drug reactions have been linked to other HLA haplotypes. However, the strong

associations found in Taiwan have not been observed in other countries with more heterogeneous populations. Development of molecular screening tests for abacavir hypersensitivity is under way as a model for implementing individualized pharmacogenetic screening in clinical practice.

CLINICAL PRESENTATION OF CUTANEOUS DRUG REACTIONS

Nonimmune Cutaneous Reactions

EXACERBATION OR INDUCTION OF DERMATOLOGIC DISEASES

A variety of drugs can exacerbate preexisting diseases or sometimes induce a disease that may or may not disappear after withdrawal of the inducing medication. For example, NSAIDs, lithium, beta blockers, TNF-αcytokine antagonists, and angiotensin-converting enzyme (ACE) inhibitors can exacerbate plaque psoriasis, while antimalarials and withdrawal of systemic glucocorticoids can worsen pustular psoriasis. Acne may be induced by glucocorticoids, androgens, lithium, and antidepressants. Minocycline and thiazide diuretics may exacerbate subacute systemic lupus erythematosus; and pemphigus can be induced by D-penicillamine, captopril, and other ACE inhibitors. Furosemide is associated with drug-induced bullous pemphigoid. The hypothesis that a drug may be responsible should always be considered, especially in cases with atypical clinical presentation.

PHOTOSENSITIVITY ERUPTIONS

Photosensitivity eruptions are usually most marked in sun-exposed areas but may extend to sun-protected areas. The mechanism is almost always phototoxicity. Phototoxic reactions resemble sunburn and can occur with first exposure to a drug. Blistering may occur in drug-related pseudoporphyria. The severity of the reactions depends on the tissue level of the drug, its efficiency as a photosensitizer, and the extent of exposure to the activating wavelengths of ultraviolet light (Chap. 56).

Common orally administered photosensitizing drugs include many fluoroquinolones and tetracycline antibiotics. Other drugs less frequently encountered are chlorpromazine, thiazides, and several NSAIDs (ibuprofen, naproxen, piroxicam). Voriconazole may result in severe photosensitivity and accelerated photo-induced aging in certain organ transplant recipients.

Because UV-A and visible light, which trigger these reactions, are not easily absorbed by nonopaque sunscreens and are transmitted through window glass, photosensitivity reactions may be difficult to block. Photosensitivity reactions abate with removal of either the drug or ultraviolet radiation, use of high-potency sunscreens that block UV-A light, and treating the reaction as one would a sunburn. Rarely, individuals develop persistent reactivity to light, necessitating long-term avoidance of sun exposure.

PIGMENTATION CHANGES

Drugs, either systemic or topical, may cause a variety of pigmentary changes in the skin. Oral contraceptives may induce melasma. Long-term minocycline, pefloxacin, and amiodarone may cause blue-gray pigmentation. Long-term, high-dose phenothiazine results in gray-brown pigmentation of sun-exposed areas. Numerous cancer chemotherapeutic agents may be associated with characteristic patterns of pigmentation (e.g., bleomycin, busulfan, daunorubicin, cyclophosphamide, hydroxyurea, and methotrexate). Pigmentation changes may also occur in mucous membranes (busulfan), nails (zidovudine), hair, and teeth.

WARFARIN NECROSIS OF SKIN

This rare reaction usually occurs between the third and tenth days of therapy with warfarin, usually in women. Common sites are breasts, thighs, and buttocks (**Fig. 55-1**). Lesions are sharply demarcated, erythematous, indurated, and purpuric and may progress to form large, irregular, hemorrhagic bullae with eventual necrosis and slow-healing eschar formation. These lesions can be life threatening.

Figure 55-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Warfarin necrosis.

Development of the syndrome is unrelated to drug dose, and the course is not altered by discontinuation of the drug after onset of the eruption. Warfarin reactions are associated with protein C deficiency. Warfarin anticoagulation in heterozygotes for protein C deficiency causes a precipitous fall in circulating levels of protein C, permitting hypercoagulability and thrombosis in the cutaneous microvasculature, with consequent areas of necrosis. Similar reactions have been associated with heparin. Heparin-induced necrosis may have clinically similar features but is probably due to heparin-induced platelet aggregation with subsequent occlusion of blood vessels; it can affect areas adjacent to the injection site or more distant sites if infused.

Warfarin-induced cutaneous necrosis is treated with vitamin K, heparin, and intensive wound care. Treatment with protein C concentrates may also be helpful.

DRUG-INDUCED HAIR DISORDERS

Drug-Induced Hair Loss

Medications may affect hair follicles at two different phases of their growth cycle: anagen (growth) or telogen (resting). *Anagen effluvium* occurs within days of drug administration, especially with antimetabolite or other chemotherapeutic drugs. In contrast, in *telogen effluvium*, the delay is 2 to 4 months following initiation of a new medication, or after childbirth, acute illness, or severe stress. In drug-related cases, both present as diffuse nonscarring alopecia most often reversible after discontinuation of the responsible agent.

The prevalence and severity of alopecia depend on the drug as well as on an individual's predisposition. A considerable number of drugs have been reported to induce hair loss. These include antineoplastic agents (alkylating agents, bleomycin, vinca alkaloids, platinum compounds), anticonvulsants (carbamazepine, valproate), antihypertensive drugs (beta blockers), antidepressants, antithyroid drugs, interferons (IFNs), oral contraceptives, and cholesterol-lowering agents.

Hirsutism

Hirsutism is an excessive growth of terminal hair with masculine hair growth pattern in a female, most often on the face and trunk. Hirsutism results from androgenic stimulation of hormone-sensitive hair follicles.

Anabolic steroids, oral contraceptives of the nonsteroid progesterone type, testosterone, and corticotropin can induce hirsutism.

Hypertrichosis

Hypertrichosis differs from hirsutism by being located mainly on the forehead and temporal regions of the face. It is usually reversible. Drugs responsible for hypertrichosis include anti-inflammatory drugs, glucocorticoids, vasodilators (diazoxide, minoxidil), diuretics (acetazolamide), anticonvulsants (phenytoin), immunosuppressive agents, psoralens, and zidovudine.

Changes in hair color or structure are uncommon adverse effects from medications. Hair discoloration may occur with chloroquine, IFN-α, chemotherapeutic agents, and tyrosine kinase inhibitors. Changes in hair structure have been observed in patients given epidermal growth factor receptor (EGFR) inhibitors.

DRUG-INDUCED NAIL DISORDERS

Drug-related nail disorders usually involve several or all 20 nails and need months to resolve after

withdrawal of the offending agent. The pathogenesis is most often toxic. Drug-induced nail changes include Beau's line (transverse depression of the nail plate), onycholysis (detachment of the distal part of the nail plate), onychomadesis (detachment of the proximal part of the nail plate), pigmentation, and paronychia (inflammation of periungual skin).

Onycholysis

Onycholysis occurs with tetracyclines, fluoroquinolones, phenothiazines, and psoralens, as well as in persons taking NSAIDs, captopril, retinoids, sodium valproate, and many chemotherapeutic agents such as anthracyclines or taxanes including paclitaxel and docetaxel. The risk of onycholysis in patients receiving cytotoxic drugs can be increased by exposure to sunlight.

Onychomadesis

Onychomadesis is caused by temporary arrest of nail matrix mitotic activity. Common drugs reported to induce onychomadesis include carbamazepine, lithium, retinoids, and chemotherapeutic agents such as cyclophosphamide and vincristine.

Paronychia

Paronychia and multiple pyogenic granuloma with progressive and painful periungual abscess of fingers and toes are a side effect of systemic retinoids, lamivudine, indinavir, and anti-EGFR monoclonal antibodies (cetuximab, gefitinib).

Nail Discoloration

Some drugs—including anthracyclines, taxanes, fluorouracil, and zidovudine—may induce nail bed hyperpigmentation through melanocyte stimulation. It appears to be reversible and dose-dependent.

PRURITUS

Pruritus is a common symptom of most drug eruptions, but it may also occur without skin lesions as the only manifestation of drug intolerance. Severe pruritus may occur in up to 50% of African patients treated with antimalarials and leads to poor compliance. It is much rarer in Caucasians.

Immune Cutaneous Reactions: Benign

MACULOPAPULAR ERUPTIONS

Morbilliform or maculopapular eruptions **(Fig. 55-2)** are the most common of all drug-induced reactions, often start on the trunk or intertriginous areas, and consist of erythematous macules and papules that are frequently symmetric and may become confluent. Involvement of mucous membranes is unusual, with the exception of scaly lips; the eruption may be associated with moderate to severe pruritus and fever. Diagnosis is rarely assisted by laboratory testing. Skin biopsy is useless because it shows normal skin or very mild and nonspecific changes. A viral exanthem is the principal differential diagnostic consideration, especially in children, and graft-versus-host-disease in the proper clinical setting. Absence of enanthems; absence of symptoms in ears, nose, throat, and upper respiratory tract; and polymorphism of the skin lesions support a drug rather than a viral eruption.

Figure 55-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Morbilliform drug eruption.

Maculopapular reactions usually develop within 1 week of initiation of therapy and last less than 2 weeks. Occasionally, these eruptions decrease with continued use of the responsible drug. Because the eruption may also worsen, the suspect drug should be discontinued unless it is essential. Oral

antihistamines, emollients, and soothing baths may help relieve pruritus. Short courses of potent topical glucocorticoids can reduce inflammation and symptoms. Systemic glucocorticoid treatment is rarely indicated.

URTICARIA/ANGIOEDEMA

Urticaria is the second most frequent type of cutaneous reaction to drugs. However, "drug allergy" explains no more than 10–20% of acute urticaria cases. It is a skin reaction characterized by pruritic, red wheals of varying size. Individual lesions rarely last more than 24 hours. Deep edema within dermal and subcutaneous tissues is known as *angioedema*. Angioedema may involve respiratory and gastrointestinal mucous membranes. Urticaria and angioedema may be part of a life-threatening anaphylactic reaction.

Drug-induced urticaria may be caused by three mechanisms: an IgE-dependent mechanism, circulating immune complexes (serum sickness), and nonimmunologic activation of effector pathways. IgE-dependent urticarial reactions usually occur within 36 hours of drug exposure but can occur within minutes. Immune complex–induced urticaria associated with serum sickness usually occurs 6–12 days after first exposure. In this syndrome, the urticarial eruption may be accompanied by fever, hematuria, arthralgias, hepatic dysfunction, and neurologic symptoms. Certain drugs, such as NSAIDs, ACE inhibitors, angiotensin II antagonists, and radiographic dyes, may induce urticarial reactions, angioedema, and anaphylaxis in the absence of drug-specific antibody through direct mast-cell degranulation.

Although ACE inhibitors, aspirin, penicillin, and blood products are the most frequent causes of urticarial eruptions, urticaria has been observed in association with nearly all drugs. Drugs may also cause chronic urticaria, which lasts more than 6 weeks. Aspirin frequently exacerbates this problem.

The treatment of urticaria or angioedema depends on the severity of the reaction and the rate at which it is evolving. In severe cases with respiratory or cardiovascular compromise, epinephrine is the mainstay of therapy, but its effect is reduced in patients using beta blockers. Treatment with systemic glucocorticoids, sometimes administered IV, is helpful. In addition to drug withdrawal, for patients with only cutaneous symptoms and without symptoms of angioedema or anaphylaxis, oral antihistamines are usually sufficient.

FIXED DRUG ERUPTIONS

These reactions are characterized by one or more sharply demarcated, erythematous lesions, sometimes leading to a blister. Hyperpigmentation often results after resolution of the acute inflammation. With rechallenge, the lesion recurs in the same (i.e., fixed) location. Lesions often involve the lips, hands, legs, face, genitalia, and oral mucosa and cause a burning sensation. Most patients have multiple lesions. Fixed drug eruptions have been associated with pseudoephedrine (frequently a nonpigmented reaction), phenolphthalein (in laxatives), sulfonamides, tetracyclines, NSAIDs, and barbiturates. Patch testing has been used in Europe to help establish the etiology.

Immune Cutaneous Reactions: Severe

VASCULITIS

Cutaneous small-vessel vasculitis often presents as palpable purpuric lesions that may be generalized or limited to the lower extremities or other dependent areas (Chap. 326). Urticarial lesions, ulcers, and hemorrhagic blisters also occur. Vasculitis may involve other organs, including the liver, kidney, brain, and joints. Drugs are implicated as a cause of 10–15% of all cases of vasculitides. Infection, malignancy, and collagen vascular disease are responsible for the majority non-drug related cases.

Propylthiouracil induces a cutaneousvasculitis that is accompanied by leukopenia and splenomegaly. Direct immunofluorescent changes in these lesions suggest immune-complex deposition. Drugs implicated in vasculitis include allopurinol, thiazides, sulfonamides, other antimicrobials, and several NSAIDs. The presence of eosinophils in the perivascular infiltrate of skin biopsy suggests a drug etiology.

PUSTULAR ERUPTIONS

AGEP is a rare reaction pattern that is often associated with exposure to drugs. Usually beginning on the face or intertriginous areas, small nonfollicular pustules overlying erythematous and edematous skin may coalesce and lead to superficial ulceration. Differentiating this eruption from TEN in its initial stages may be difficult. A skin biopsy is important and shows scattered pustules in the upper part of the epidermis instead of the full-thickness necrosis that characterizes TEN. Fever is present with elevated neutrophil counts, and sepsis is often suspected. Acute pustular psoriasis is the principal differential diagnostic consideration. AGEP often begins within a few days of initiating drug treatment, most notably antibiotics. For other associated drugs (e.g., diltiazem, chloroquine, hydroxychloroquine, or terbinafine), AGEP begins later: 7–14 days after initiation of treatment.

HYPERSENSITIVITY SYNDROME

Initially described with phenytoin, hypersensitivity syndrome—a multiorgan drug-induced reaction—is also known as *DRESS (drug reaction with eosinophilia and systemic symptoms)* and *drug-induced*

hypersensitivity syndrome (*DIHS*). It presents as a widespread erythematous eruption that may become purpuric or lichenoid and is accompanied by many of the following features: fever, facial and periorbital edema, tender and generalized lymphadenopathy, leukocytosis (often with atypical lymphocytes and eosinophilia), hepatitis, and sometimes nephritis or pneumonitis. The cutaneous reaction usually begins 2 to 8 weeks after the drug is started and lasts longer than mild eruptions after drug cessation. Signs and symptoms may persist for several weeks, especially those associated with hepatitis. The eruption recurs with rechallenge, and cross-reactions among aromatic anticonvulsants, including phenytoin, carbamazepine, and barbiturates, are frequent. Other drugs causing this syndrome include lamotrigine, minocycline, dapsone, allopurinol, and sulfo- namides, as well as abacavir and zalcitabine in HIV-infected patients. Hypersensitivity to reactive drug metabolites, hydroxylamine for sulfamethoxazole, and arene oxide for aromatic anticonvulsants may be involved in the pathogenesis of DRESS. Reactivation of herpes viruses, especially of herpes virus 6, has been reported to be frequent in this syndrome. The role of virus infection is still unclear; it may contribute to long-lasting manifestations such as hepatitis or encephalitis. Mortality rates as high as 10% have been reported; mortality is highest in association with hepatitis. Systemic glucocorticoids (prednisone, 1.5–2 mg/kg per d) should be started with slow taper over 8–12 weeks. Topical, high-potency glucocorticoids may also be helpful. In all cases, rapid withdrawal of the suspected drug is required. Patients should be closely monitored for resolution of organ dysfunction and for development of late-onset autoimmune thyroiditis (up to 6 months).

STEVENS-JOHNSON SYNDROME AND TOXIC EPIDERMAL NECROLYSIS

SJS and *TEN* are characterized by blisters and epidermal detachment resulting from epidermal necrosis in the absence of substantial dermal inflammation. The term *Stevens-Johnson syndrome* is now used to describe cases with blisters developing on target lesions, dusky or purpuric macules in which mucosal involvement is significant, and total body surface area blistering and eventual detachment is <10%. The term *Stevens-Johnson syndrome/toxic epidermal necrolysis overlap* is used to describe cases with 10–30% detachment, and *TEN* is used to describe cases with >30% detachment. Erythema multiforme major is now considered by most to be different from SJS. Erythema multiforme (EM) is characterized by mucosal involvement and true "target" lesions or atypical raised target lesions often more acrally distributed and with limited extent of skin detachment. EM is most often triggered by infection, particularly herpes simplex virus or *mycoplasma pneumoniae.*

Patients with SJS, SJS/TEN, or TEN initially present with acute symptoms, painful skin lesions, fever >39°C (102.2°F), sore throat, and conjunctivitis resulting from mucosal membrane and ocular lesions. Intestinal and pulmonary involvements are associated with a poor prognosis, as are a greater extent of epidermal detachment and older age. About 10% and 30% of SJS- and TEN-affected persons die from their disease, respectively. Drugs that most commonly cause SJS or TEN are sulfonamides, nevirapine, allopurinol, lamotrigine, aromatic anticonvulsants, and the oxicam NSAIDs. At this time, SJS or TEN have no treatment of proven efficacy. The best results come from early diagnosis, immediate discontinuation of any suspected drug, supportive therapy, and paying close attention to ocular complications and infection. Frozen-section skin biopsy may aid in rapid diagnosis. Systemic glucocorticoid therapy (prednisone 1 to 2 mg/kg) may be useful early in the evolution of the disease, but long-term systemic glucocorticoid use has been associated with higher mortality. After initial enthusiasm for the use of intravenous immunoglobulin (IVIG) in the treatment of SJS/TEN, some recent data suggest IVIG is unlikely to benefit these patients. Randomized studies to more definitively assess the potential benefit of systemic glucocorticoids and IVIG are lacking and difficult to perform but are necessary.

DRUGS OF SPECIAL INTEREST

Allopurinol

Together with sulfonamides and antiepileptics, allopurinol is a frequent cause of mild maculopapular eruptions (in at least 3% of users) and may also cause more severe reactions including hypersensitivity/DRESS and SJS/TEN; patients with HLA B*5801 are highly susceptible to these severe reactions.

Anti-HIV Medications

In clinical trials, combinations of highly active antiretroviral treatments were frequently associated with ≥10% "drug eruptions." Two drugs, nevirapine and abacavir, have been associated with specific risks.

Nevirapine has both a high risk of maculopapular eruptions and a very high risk (about 1 in 1000) of SJS or TEN. Progressive escalation of daily doses has been shown to decrease the risk of mild eruption but does not abrogate the risk of severe reactions.

Abacavir is associated with a 4–8% risk of a hypersensitivity reaction, which is remarkable because of the association of symptoms suggesting a type I hypersensitivity reaction (dyspnea, diarrhea, low blood pressure, shock on rechallenge) as well as signs of delayed hypersensitivity (rash, late onset hepatitis). The risk is lower in patients of African ancestry and strongly correlated with HLA B*5701. Evidence supports patch testing for abacavir hypersensitivity, and additional molecular screening

methods are being developed to identify susceptible individuals.

Penicillin

Incidence of cutaneous reactions due to penicillin is about 1%. About 85% of cutaneous reactions are morbilliform, and about 10% are urticaria or angioedema. Anaphylaxis and serum sickness appear to be due to IgE antibodies in serum.

Delayed reactions, mainly maculopapular eruptions, are much more common with aminopenicillins, involving 4–7% of users. The question of cross-reactivity between β-lactam antibiotics and preventing the risk of anaphylaxis is discussed below ("Management of a Patient with a Drug Eruption"). Approximately 10% of patients with penicillin allergies will also develop allergic reactions to cephalosporin class antibiotics.

Nonsteroidal Anti-Inflammatory Drugs

Most NSAIDs, including aspirin, cause immediate, allergy-like symptoms in susceptible individuals. Approximately 1% of persons experience urticaria or angioedema, and rare individuals (0.5%) experience rhinosinusitis and asthma.

Urticaria/angioedema may be delayed up to 24 hours and may occur at any age. The rhinosinusitis-asthma syndrome generally develops within 1 hour of drug administration. Recurrences are frequent and can be complicated by nasal and sinus infection, polyposis, bloody discharge, and nasal eosinophilia. In many individuals with this syndrome, potentially life-threatening asthma may ensue whenever NSAIDs are subsequently ingested. Proof of the association of symptoms and NSAID use requires either clear-cut history of symptoms following drug ingestion or an oral challenge. That procedure must be conducted only in a hospital setting by experienced personnel. Cross-reactivity between NSAIDs that inhibit cyclooxygenase (COX) 1 is common, while reactivity to COX-2 inhibitors is less frequent. The reaction is pharmacologic, and patients who are sensitive to NSAIDs cannot be identified by assessment of IgE antibody to aspirin, lymphocyte sensitization, or in vitro immunologic testing.

Other reactions can also occur with NSAIDs, including phototoxicity with many agents, a pattern of pseudoporphyria being often related to naproxen, hypersensitivity/DRESS (oxicam derivatives, COX-2 inhibitors), and SJS or TEN (phenylbutazone, oxicam derivatives).

Radiocontrast Media

Many patients are exposed to radiocontrast agents, which can be a cause of adverse drug reactions. High-osmolality radiocontrast media were about five times more likely to induce urticaria (1%) or anaphylaxis than were newer low-osmolality media. Severe reactions are rare with either type of contrast media. About one-third of those with mild reactions to previous exposure react on re-exposure. Pretreatment with prednisone and diphenhydramine reduces reaction rates. Persons with a reaction to a high-osmolality contrast media should be given low-osmolality media if later contrast studies are required.

A new association between certain types of gadolinium contrast and nephrogenic systemic fibrosis, a condition of sclerosing skin with rare internal organ involvement, has been reported; advanced renal compromise may be an important risk factor.

Anticonvulsants

Phenobarbital, phenytoin, carbamazepine, and lamotrigine are associated with many types of severe reactions in adults and a high incidence of less-severe reactions in children. These drugs carry the highest risk of SJS, TEN, and hypersensitivity syndrome in immunologically normal patients. There are strong HLA associations with anticonvulsant-induced toxicity in Han Chinese (see above). Aromatic anticonvulsants can induce a pseudolymphoma syndrome and induce gingival hyperplasia.

The overall rate of lamotrigine associated rash is lower in patients whose doses are slowly increased, but whether or not the risk of more severe reactions is reduced by slow dose escalation is less clear. Patients who also use valproate, which increases lamotrigine levels and half life, have a higher risk of skin reactions.

Sulfonamides

Antibacterial sulfonamides cause cutaneous eruptions at a high rate and are among the drugs most frequently implicated in SJS and TEN. The combination of sulfamethoxazole and trimethoprim frequently induces adverse cutaneous reactions in patients with AIDS (Chap. 182). Desensitization is often successful in AIDS patients with morbilliform eruptions but is not recommended in AIDS patients who manifested erythroderma or a bullous reaction in response to their earlier sulfonamide exposure.

Reaction rates are much lower with nonantibiotic sulfonamides, including diuretics or antidiabetic agents. Cross-reactivity between antibiotic and nonantibiotic sulfonamides is at most infrequent.

Vancomycin

Vancomycin causes two unusual but recognizable cutaneous reactions: linear IgA bullous dermatosis

(a transient, blistering eruption) and *red man syndrome*. Red man syndrome occurs during rapid IV infusion of vancomycin. This is thought to be a histamine-related anaphylactoid reaction characterized by flushing, diffuse maculopapular eruption, hypotension, and, in rare cases, cardiac arrest.

Agents Used in Cancer Chemotherapy

Because many agents used in cancer chemotherapy inhibit cell division, rapidly proliferating elements of the skin, including hair, mucous membranes, and appendages, are sensitive to their effects. As a result, stomatitis and alopecia are among the most frequent dose-dependent side effects of chemotherapy. Various nail abnormalities have been described: onycholysis, dystrophy, Beau's lines, white lines, and pigmentation. Sterile cellulitis and phlebitis and ulceration of pressure areas occur with many of these agents. Also reported is acral erythema, which begins with dysesthesia followed by redness and a painful edematous eruption of the palms and soles, and is caused by cytarabine, doxorubicin, methotrexate, hydroxyurea, and 5-fluorouracil. Pyridoxine alleviates symptoms of acral erythema. Urticaria, angioedema, and exfoliative dermatitis also have been seen, as has local and diffuse hyperpigmentation.

Hypersensitivity to carboplatin or cisplatin is not rare (with an incidence of 10–20%) among patients receiving multiple treatments with these drugs. It is probably IgE mediated. Moderate to severe reactions, including respiratory distress and hypotension, are also observed in 10–20% of patients receiving paclitaxel regardless of premedication with glucocorticoids and histamine H(1) and H(2) antagonists.

Cetuximab and other EGF receptor antagonists induce follicular eruptions and nail toxicity after a mean interval of 10 days in a majority of patients. Xerosis, eczematous eruptions, and pruritus also frequently occur with EGF receptor antagonists. The severity of the eruption correlates with a better anticancer effect. Systemic antibiotics, topical antiacne treatments, and supportive care are helpful. Sorafenib, a tyrosine kinase inhibitor, may result in follicular eruptions and bullous palmoplantar eruptions with dysesthesia.

Glucocorticoids

Both systemic and topical glucocorticoids cause a variety of skin changes, including acneiform eruptions, atrophy, striae, and other stigmata of Cushing's syndrome. And, in sufficiently high doses, glucocorticoids can retard wound healing. Patients using glucocorticoids are at higher risk for bacterial, yeast, and fungal skin infections that may be misinterpreted as drug eruptions but are instead drug side effects. Allergy to glucocorticoids may also occur either as contact dermatitis to topical formulations or as systemic reactions, including anaphylaxis.

Biologic Therapies

These include cytokines, cytokine antagonists, and monoclonal antibodies.

Injection-site reactions are the most frequent adverse events. The severity varies from mild redness to deep inflammation and necrosis. In most cases, the treatment can be continued and the severity of reactions will decrease with time.

Like all foreign proteins, monoclonal antibodies may induce urticaria, angioedema, anaphylactic reactions, and serum sickness. Production of inactivating antibodies to the drug may occur, classically with infliximab, and may be prevented by concomitant administration of low-dose methotrexate.

Alopecia is a common complication of IFN-α. A nonspecific, highly pruritic "dermatitis" is frequent in patients receiving IFN and ribavirin for hepatitis C.

Induction or exacerbation of various immune-mediated disorders, especially lupus erythematosus, has been reported with interleukin 2, IFN-α, and antitumor necrosis factor α.

Granulocyte colony-stimulating factor may induce various neutrophilic dermatoses, including Sweet's syndrome and pyoderma gangrenosum, and can exacerbate psoriasis.

Antimalarial Agents

Antimalarial agents are used as therapy for several skin diseases, including the skin manifestations of lupus and polymorphous light eruption, but they can also induce cutaneous reactions. The most frequent is pruritus, which occurs in up to 50% of African patients receiving chloroquine and may be severe enough to lead to discontinuation of treatment.

Pigmentation disturbances, including black pigmentation of the face, mucous membranes, and pretibial and subungual areas, occur with antimalarials. Quinacrine (mepacrine) causes generalized, cutaneous yellow discoloration. Less frequent reactions include pustular eruptions (AGEP) and hypersensitivity/DRESS.

MANAGEMENT OF A PATIENT WITH A DRUG ERUPTION

There are four main questions to answer regarding an eruption:

1. Is it a drug reaction?

2. Is it a severe eruption or the onset of a form that may become severe?
3. Which drug(s) are suspected, and which drug(s) should be withdrawn?
4. What is recommended for future use of drugs?

Early Diagnosis of Severe Eruptions

Rapid recognition of adverse drug reactions that may become serious or life threatening is paramount. **Table 55-1** lists clinical and laboratory features that, if present, suggest that the reaction may be serious. **Table 55-2** provides key features of the most serious adverse cutaneous reactions. Intensity of symptoms and rapid progression of signs should raise the suspicion of a severe eruption. Any doubt should lead to prompt consultation with a dermatologist and/or referral of the patient to a specialized center.

Table 55-1 Clinical and Laboratory Findings Associated with More Serious Drug-Induced Cutaneous Clinical Findings

Cutaneous
Generalized erythema
Facial edema or central facial involvement
Skin pain
Palpable purpura
Target lesions
Skin necrosis
Blisters or epidermal detachment
Positive Nikolsky's sign
Mucous membrane erosions
Urticaria
Swelling of tongue
General
High fever [temperature >40°C (>104°F)]
Enlarged lymph nodes
Arthralgias or arthritis
Shortness of breath, wheezing, hypotension
Laboratory results
Eosinophil count >1000/μL
Lymphocytosis with atypical lymphocytes
Abnormal liver or kidney function tests

Source: Adapted from Roujeau and Stern.

Table 55-2 Clinical Features of Severe Cutaneous Drug Reactions

Diagnosis	Mucosal Lesions	Typical Skin Lesions	Frequent Signs and Symptoms	Alternative Causes Not Related to Drugs
Stevens-Johnson syndrome	Erosions usually at two or more sites	Small blisters on dusky purpuric macules or atypical targets; rare areas of confluence; detachment ≤10% of body surface area	Most cases involve fever	10-20% cause not determined
Toxic epidermal necrolysis[a]	Erosions usually at two or more sites	Individual lesions like those seen in Stevens-Johnson syndrome; confluent erythema; outer layer of epidermis separates readily from basal layer with lateral pressure; large sheet of necrotic epidermis; total detachment of >30% of body surface area	Nearly all cases involve fever, "acute skin failure," leukopenia	10–20% cause not determined
Hypersensitivity syndrome	Infrequent	Severe exanthematous rash (may become purpuric), exfoliative dermatitis, facial edema	30–50% of cases involve fever, lymphadenopathy, hepatitis, nephritis,	Cutaneous lymphoma

			carditis, eosinophilia, atypical lymphocytes	
Acute generalized exanthematous pustulosis	About 20% erosions mouth, tongue	Initially nonfollicular small pustules overlying edematous erythema, sometimes leading to superficial erosions	Fever, burning, pruritus, facial swelling, leukocytosis, hypocalcemia	Infection
Serum sickness or reactions resembling serum sickness	Absent	Morbilliform lesions, sometimes with urticaria	Fever, arthralgias	Infection
Anticoagulant-induced necrosis	Infrequent	Erythema then purpura and necrosis, especially of fatty areas	Pain in affected areas	Disseminated intravascular coagulopathy, septicemia
Angioedema	Often involved	Urticaria or swelling of central part of face	Respiratory distress, cardiovascular collapse	Insect stings, foods

[a] Overlap of Stevens-Johnson syndrome and toxic epidermal necrolysis with features of both and attachment of 10–30% of body surface area may occur.

Source: Adapted from Roujeau and Stern.

Confirmation of Drug Reaction

The probability of drug etiology varies with the pattern of the reaction. Only fixed drug eruptions are always drug-induced. Morbilliform eruptions are usually viral in children and drug-induced in adults. Among severe reactions, drugs account for 10–20% for anaphylaxis and vasculitis and between 70–90% for AGEP, DRESS, SJS, or TEN. Skin biopsy helps in characterizing the reaction but does not indicate drug causality. Blood counts and liver and renal function tests are important for evaluating organ involvement. The association of mild elevation of liver enzymes and high eosinophil count is frequent but not specific for a drug reaction. Blood tests that could identify an alternative cause, antihistone antibody tests, and serology or polymerase chain reaction for infections may be of great importance for final assessment of etiology.

What Drug(S) to Suspect and Withdraw

Most cases of drug eruptions occur during the first course of treatment with a new medication. A notable exception is IgE-mediated urticaria and anaphylaxis that need presensitization and develop a few minutes to a few hours after rechallenge. Characteristic time of onset to drug reaction: 4–14 days for morbilliform eruptions, 5–28 days for SJS/TEN, and 14–48 days for DRESS. Medications introduced for the first time in the relevant time frame are prime suspects. Two other important elements to suspect causality at this stage are (1) previous experience with the drug in the population and (2) alternative etiologic candidates.

The decision to continue or discontinue any medication will depend on the severity of the reaction, the severity of the primary disease, the degree of suspicion of causality, and the feasibility of an alternative safer treatment. In severe drug reactions, elimination of all possible suspect drugs or unnecessary medications should be attempted. Some rashes may disappear when "treating through" a benign drug-related eruption as seen in patients with AIDS treated for opportunistic infections with antibacterial sulfonamides. The decision to treat through an eruption should, however, remain the exception and withdrawal of every suspect drug the general rule. On the other hand, drugs that are not suspected and are important for the patient, (e.g., antihypertensive agents) generally should not be quickly withdrawn. This approach prevents reluctance to future use of these agents.

Recommendation for Future Use of Drugs

The aims are (1) to prevent the recurrence of the drug eruption and (2) not to compromise future treatments by contraindicating otherwise useful medications.

Begin with thorough assessment of drug causality. Drug causality is evaluated based on timing of the reaction, evaluation of other possible causes, effect of drug withdrawal or continuation, and knowledge of medications that have been associated with the observed reaction. Combination of these criteria leads to considering the causality as definite, probable, possible, or unlikely.

A drug with "unlikely" causality or that has been continued when the reaction improved or was reintroduced without a reaction can be administered safely.

A drug with a "definite" or "probable" causality should be contraindicated, and a warning card or

medical alert tag (e.g., wristband) given to the patient.

A drug with a "possible" causality may be submitted to further investigations depending on the expected need for future treatment.

The usefulness of drug tests is still debated. Many in vitro immunologic assays have been developed, but the predictive value of these tests has not been validated in any large series of affected patients; these tests exist primarily for research and not clinical purposes.

In patients with history suggesting immediate IgE-mediated reactions to penicillin, skin testing with major and minor determinants of penicillins or cephalosporins has proved useful for identifying patients at risk of anaphylactic reactions to these agents. However, skin tests themselves carry a small risk of anaphylaxis. Negative skin tests do not totally rule out IgE-mediated reactivity, but the risk of anaphylaxis in response to penicillin administration in patients with negative skin tests is about 1% while about two-thirds of patients with a positive skin test experience an allergic response on rechallenge.

For patients with late drug reactions, the clinical usefulness of skin tests is more questionable. At least one of a combination of several tests (prick, patch, and intradermal) is positive in 50–70% of patients with a reaction "definitely" attributed to a single medication. This low sensitivity probably contributes to the fact that readministration of drugs that had been tested negative resulted in up to 17% eruptions.

Cross-Sensitivity

Because of the possibility of cross-sensitivity among chemically related drugs, many physicians recommend avoidance of not only the medication that induced the reaction but also all drugs of the same pharmacologic class.

There are two types of cross-sensitivity. Reactions that depend on a pharmacologic interaction may recur with all drugs that target the same pathway, whether they are structurally similar or not. This is the case with angioedema caused by NSAIDs and ACE inhibitors. In this situation, the risk of recurrence varies from drug to drug in a particular class; however, avoidance of all drugs in the class is usually recommended. Immune recognition of structurally related drugs is the second mechanism by which cross-sensitivity occurs. A classic example is hypersensitivity to aromatic antiepileptics (barbiturates, phenytoin, carbamazepine) with up to 50% reaction to a second drug in patients who reacted to one. For other drugs, in vitro as well as in vivo data have suggested that cross-reactivity existed only between compounds with very similar chemical structures. Sulfamethoxazole-specific lymphocytes may occasionally recognize a few other antibacterial sulfonamides but not diuretics, antidiabetic drugs, or anti-COX2 NSAIDs with a sulfonamide group.

Recent data suggest that although the risk of a drug eruption to another drug was increased in persons with a prior reaction, "cross-sensitivity" was probably not the explanation. As an example, persons with a history of an allergic-like reaction to penicillin were at higher risk to develop a reaction to antibacterial sulfonamides than to cephalosporin.

These data suggest that the list of drugs to avoid after a drug reaction should be limited to the causative one(s) and to a few very similar medications.

Because of growing evidence that some severe cutaneous reactions to drugs are associated with HLAgenes, it is probably wise to recommend that first-degree family members of patients with severe cutaneous reactions also should avoid these causative medications.

Desensitization can be considered in those with a history of reaction to a medication that must be used again. Efficacy of such procedures has been demonstrated in cases of immediate reaction to penicillin and positive skin tests, anaphylactic reactions to platinum chemotherapy, and delayed reactions to sulfonamides in patients with AIDS. Various protocols are available, including oral and parenteral approaches. Oral desensitization appears to have a lower risk of serious anaphylactic reactions. However, desensitization carries the risk of anaphylaxis regardless of how it is performed and should be performed in monitored clinical settings such as an intensive care unit. After desensitization, many patients experience non-life-threatening reactions during therapy with the culprit drug.

Reporting

Any severe reaction to drugs should be reported to a regulatory agency or to pharmaceutical companies (e.g., MedWatch, http://www.fda.gov/Safety/MedWatch/default.htm). Because severe reactions are too rare to be detected in premarketing clinical trials, spontaneous reports are of critical importance for early detection of unexpected life-threatening events. To be useful, the report should contain enough details to permit ascertainment of severity and drug causality. This permits recognition of similar cases that may be reported from several different sources.

FURTHER READINGS

Auquier-Dunant A et al: Severe cutaneous adverse reactions. Correlations between clinical patterns and causes of erythema multiforme majus, Stevens-Johnson syndrome, and toxic epidermal necrolysis: Results of an international prospective study. Arch Dermatol 138:1019, 2002 [PMID: 12164739]

Bachot N, Roujeau JC: Differential diagnosis of severe cutaneous drug eruptions. *Am J Clin Dermatol* 4:561, 2003 [PMID: 12862499]

Bigby M: Rates of cutaneous reactions to drugs. Arch Dermatol 137:765, 2001 [PMID: 11405768]

Eliaszewicz M et al: Prospective evaluation of risk factors of cutaneous drug reactions to sulfonamides in patients with AIDS. J Am Acad Dermatol 47:40, 2002 [PMID: 12077579]

Fiszenson-Albala F et al: A 6 month prospective survey of cutaneous drug reactions in a hospital setting. *Br J Dermatol* 149:1018, 2003 [PMID: 14632808]

Fox LP: Nail toxicity associated with epidermal growth factor receptor inhibitor therapy. *J Am Acad Dermatol* 56:460, 2007 [PMID: 17147969]

Hirsch LJ et al: Predictors of lamotrigine-associated rash. *Epilepsia* 47:318, 2006 [PMID: 16499755]

Hung SI et al: HLA B*5801 allele as a genetic marker for severe cutaneous reactions caused by allopurinol. Proc Natl Acad Sci USA 102:4134, 2005 [PMID: 15743917]

Lange-Asschenfeldt C et al: Cutaneous adverse reactions to psychotropic drugs: data from a multicenter surveillance program. *J Clin Psychiatry* 70:1258, 2009 [PMID: 19538904]

Lin D, Tucker MJ, Rieder MJ: Increased adverse drug reactions to antimicrobials and anticonvulsants in patients with HIV infection. *Ann Pharm* 40:1594, 2006 [PMID: 16912251]

Mockenhaupt M et al: Stevens-Johnson syndrome and toxic epidermal necrolysis: Assessment of medication risks with emphasis on recently marketed drugs. The EuroSCAR-Study. *J Invest Dermatol* 128:35, 2008 [PMID: 17805350]

Perazella MA: Advanced kidney disease, gadolinium and nephrogenic systemic fibrosis: the perfect storm. *Curr Opin Nephrol Hypertens* Epub: 2009

Phillips EJ, Mallal SA: HLA and drug-induced toxicity. *Curr Opin Mol Therap* 11:231, 2009 [PMID: 19479656]

Racette AJ et al: Photoaging and phototoxicity from long-term voriconazole treatment in a 15-year-old girl. *J Am Acad Dermatol* 52:(S81), 2005

Roujeau JC, Stern RS: Severe adverse cutaneous reactions to drugs. N Engl J Med 331:1272, 1994 [PMID: 7794310]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SOLAR RADIATION

Sunlight is the most visible and obvious source of comfort in the environment. The sun provides the beneficial effects of warmth and vitamin D synthesis; however, acute and chronic sun exposure also have pathologic consequences. Few effects of sun exposure beyond those affecting the skin have been identified, but cutaneous exposure to sunlight is the major cause of human skin cancer and can have immunosuppressive effects as well.

The sun's energy reaching the earth's surface is limited to components of the ultraviolet (UV), the visible, and portions of the infrared spectra. The cutoff at the short end of the UV is at ~290 nm; this is due primarily to stratospheric ozone formed by highly energetic ionizing radiation, preventing penetration to the earth's surface of the shorter, more energetic, potentially more harmful wavelengths of solar radiation. Indeed, concern about destruction of the ozone layer by chlorofluorocarbons released into the atmosphere has led to international agreements to reduce production of those chemicals.

Measurements of solar flux indicate that there is a twentyfold regional variation in the amount of energy at 300 nm that reaches the earth's surface. This variability relates to seasonal effects, the path that sunlight traverses through ozone and air, the altitude (4% increase for each 300 m of elevation), the latitude (increasing intensity with decreasing latitude), and the amount of cloud cover, fog, and pollution.

The major components of the photobiologic action spectrum capable of affecting human skin include the UV and visible wavelengths between 290 and 700 nm. In addition, the wavelengths beyond 700 nm in the infrared spectrum primarily emit heat and in certain circumstances may exacerbate the pathologic effects of energy in the UV and visible spectra.

The UV spectrum reaching the earth represents <10% of total incident solar energy and is arbitrarily divided into two major segments, UV-B and UV-A, constituting the wavelengths from 290–400 nm. UV-B consists of wavelengths between 290 and 320 nm. This portion of the photobiologic action spectrum is the most efficient in producing redness or erythema in human skin and hence sometimes is known as the "sunburn spectrum." UV-A includes wavelengths between 320 and 400 nm and is ~1000-fold less efficient in producing skin redness than is UV-B.

The wavelengths between 400 and 700 nm are visible to the human eye. The photon energy in the visible spectrum is not capable of damaging human skin in the absence of a photosensitizing chemical. Without the absorption of energy by a molecule, there can be no photosensitivity. Thus, the *absorption spectrum* of a molecule is defined as the range of wavelengths absorbed by it, whereas the *action spectrum* for an effect of incident radiation is defined as the range of wavelengths that evoke the response.

Photosensitivity occurs when a photon-absorbing chemical (chromophore) present in the skin absorbs incident energy, becomes excited, and transfers the absorbed energy to various structures or to oxygen.

UV Radiation (UVR) and Skin Structure and Function

Skin consists of two major compartments: the outer epidermis, which is a stratified squamous epithelium, and the underlying dermis, which is rich in matrix proteins such as collagens and elastin. Both compartments are susceptible to damage from sun exposure. The epidermis and the dermis contain several chromophores capable of absorbing incident solar energy, including nucleic acids, proteins, and lipids. The outermost epidermal layer, the stratum corneum, is a major absorber of UV-B, and <10% of incident UV-B wavelengths penetrate through the epidermis to the dermis. Approximately 3% of radiation below 300 nm, 20% of radiation below 360 nm, and 33% of short visible radiation reaches the basal cell layer in untanned human skin. In contrast, UV-A readily penetrates to the dermis and is capable of altering structural and matrix proteins that contribute to photoaging of chronically sun-exposed skin, particularly in individuals of light complexion. Thus, longer wavelengths can penetrate more deeply into the skin.

MOLECULAR TARGETS FOR UVR-INDUCED SKIN EFFECTS

Epidermal DNA, predominantly in keratinocytes and in Langerhans cells (LCs), which are dendritic antigen-presenting cells, absorbs UV-B and undergoes structural changes between adjacent pyrimidine bases (thymine or cytosine), including the formation of cyclobutane dimers and 6,4-

photoproducts. These structural changes are potentially mutagenic and are found in most basal cell and squamous cell skin cancers. They can be repaired by cellular mechanisms that result in their recognition and excision and the restoration of normal base sequences. The efficient repair of these structural aberrations is crucial, since individuals with defective DNA repair are at high risk for the development of cutaneous cancer. For example, patients with xeroderma pigmentosum (XP), an autosomal recessive disorder, are characterized by variably deficient repair of UV-induced photoproducts, and their skin phenotype often manifests the dry, leathery appearance of prematurely photoaged skin as well as basal cell and squamous cell carcinomas and melanoma in the first two decades of life. Studies in mice using knockout gene technology have verified the importance of functional genes regulating these repair pathways in preventing the development of UV-induced cancer. Furthermore, incorporation of a bacterial DNA repair enzyme, T4 endonuclease V, into liposomes in a product applied to the skin of patients with XP selectively removes cyclobutane pyrimidine dimers and reduces the degree of solar damage and skin cancer. This approach also may benefit other patients who are susceptible to skin cancer, such as organ transplant recipients receiving chronic immunosuppressive drug therapy. DNA damage in LCs may contribute to the known immunosuppressive effects of UV-B (see "Immunologic Effects," below).

In addition to DNA, molecular oxygen is a target for incident solar UVR, leading to the generation of reactive oxygen species (ROS). These ROS can damage skin components, such as epidermal lipids, either free lipids in the stratum corneum or cell membrane lipids. UVR also can target proteins, leading to increased cross-linking and degradation of matrix proteins in the dermis and thus to photoaging changes known as solar elastosis.

CUTANEOUS OPTICS AND CHROMOPHORES

Chromophores are endogenous or exogenous chemical components that can absorb physical energy. Endogenous chromophores are of two types: (1) normal components of skin, including nucleic acids, proteins, lipids, and 7-dehydrocholesterol, the precursor of vitamin D, and (2) components that are synthesized elsewhere in the body that circulate in the bloodstream and diffuse into the skin, such as porphyrins. Normally, only trace amounts of porphyrins are present in the skin, but in selected diseases known as the porphyrias (Chap. 358), increased amounts are released into the circulation from the bone marrow and the liver and are transported to the skin, where they absorb incident energy both in the Soret band, around 400 nm (short visible), and to a lesser extent in the red portion of the visible spectrum (580–660 nm). This results in the generation of ROS that can mediate structural damage to the skin, manifested as erythema, edema, urticaria, or blister formation. It is of interest that photoexcited porphyrins are currently used therapeutically in treating nonmelanoma skin cancers and their precursor lesions, actinic keratoses. Known as photodynamic therapy (PDT), this modality generates ROS in the skin, leading to cell death. Topical photosensitizers used in PDT are the porphyrin precursors 5-aminolevulinic acid (ALA) and methyl aminolevulinate (MAL), which are converted to porphyrins in the skin. It is believed that PDT targets tumor cells for destruction more selectively than it targets adjacent nonneoplastic cells. The efficacy of PDT requires appropriate timing of the application of MAL and ALA to the affected skin followed by exposure to artificial sources of visible light. High-intensity blue light has been used successfully for the treatment of thin actinic keratoses. Red light has a longer wavelength and penetrates more deeply into the skin and is more beneficial in the treatment of superficial basal cell carcinomas.

ACUTE EFFECTS OF SUN EXPOSURE

The acute effects of skin exposure to sunlight include sunburn and vitamin D synthesis.

Sunburn

This painful skin condition is an acute inflammatory response of the skin predominantly to UV-B. Generally, an individual's ability to tolerate sunlight is inversely proportional to that individual's degree of melanin pigmentation. Melanin, a complex polymer of tyrosine derivatives, is synthesized in specialized epidermal dendritic cells known as melanocytes and is packaged into melanosomes that are transferred via dendritic process into keratinocytes, thereby providing photoprotection and simultaneously darkening the skin. Sun-induced melanogenesis is a consequence of increased tyrosinase activity in melanocytes. Central to the suntan response is the melanocortin-1 receptor (*MC1R*), and mutations in this gene account for the wide variation in human skin and hair color. The human *MC1R* gene encodes a 317-amino-acid G protein-coupled receptor (melanocortin receptor) that binds α-melanocyte-stimulating hormone (α-MSH), which is secreted in the skin mainly by keratinocytes in response to UVR. This UV-induced expression of α-MSH is controlled by the tumor suppressor p53, and absence of functional p53 is known to ablate the tanning response in a mouse model. Activation of the melanocortin receptor leads to increased intracellular cyclic adenosine 5'-monophosphate (cyclic AMP) and protein kinase A activation, followed by increased transcription of microphthalmia transcription factor (MITF), which stimulates melanogenesis. Since the precursor of α-MSH, proopiomelanocortin (POMC), is also the precursor of β-endorphin, UVR results in increased pigmentation and β-endorphin production, probably promoting sun-seeking behaviors.

The Fitzpatrick classification of human skin is a function of the efficiency of the epidermal-melanin unit and usually can be ascertained by asking an individual two questions: (1) Do you burn after sun

exposure? and (2) Do you tan after sun exposure? The answers to these questions permit division of the population into six skin types varying from type I (always burn, never tan) to type VI (never burn, always tan) **(Table 56-1)**.

Table 56-1 Skin Type and Sunburn Sensitivity (Fitzpatrick Classification)

Type	Description
I	Always burn, never tan
II	Always burn, sometimes tan
III	Sometimes burn, sometimes tan
IV	Sometimes burn, always tan
V	Never burn, sometimes tan
VI	Never burn, always tan

Sunburn erythema is due to vasodilation of dermal blood vessels. There is a lag in time between skin exposure to sunlight and the development of visible redness (usually 4–12 h), suggesting that an epidermal chromophore causes delayed production and/or release of vasoactive mediator(s), or cytokines, that diffuse to the dermal vasculature to evoke vasodilation.

The action spectrum for sunburn erythema includes UV-B and UV-A. Photons in the UV-B are at least 1000-fold more efficient than photons in the UV-A in evoking the response. However, UV-A may contribute to sunburn erythema at midday, when much more UV-A than UV-B is present in the solar spectrum. The erythema that accompanies the inflammatory response induced by UVR results from the orchestrated release of cytokines, along with growth factors or lipid mediators, and the generation of ROS. Furthermore, it is known that UV-induced activation of nuclear factor-κB (NF-κB)-dependent gene transcription can augment release of several proinflammatory cytokines and vasoactive mediators, including interleukin 1 (IL-1), IL-6, IL-8, IL-12, vascular endothelial growth factor, prostaglandin E2 , and tumor necrosis factor α. Local accumulation of these cytokines occurs in sunburned skin, providing chemotactic factors that attract neutrophils and macrophages, which can cause cell damage through generation of ROS. UVR also promotes infiltration of inflammatory cells through induced expression of adhesion molecules such as E-selectin and intercellular adhesion molecule-1 (ICAM-1) on endothelial cells and ICAM-1 on keratinocytes. It is of interest that nonsteroidal anti-inflammatory drugs (NSAIDs) can reduce sunburn erythema. UVR also has been shown to activate phospholipase A_2, resulting in increased eicosanoids, such as prostaglandin E_2, which is known to be a potent inducer of sunburn erythema, thereby explaining how cyclooxygenase inhibitors can reduce sunburn erythema.

Epidermal changes in sunburn include the induction of "sunburn cells," which are keratinocytes undergoing p53-dependent apoptosis as a defense for the elimination of cells that harbor UV- B-induced structural DNA damage.

Vitamin D Photochemistry

Cutaneous exposure to UV-B causes photolysis of epidermal 7-dehydrocholesterol, converting it to pre-vitamin D3 , which then undergoes a temperature-dependent isomerization to form the stable hormonevitamin D_3. This compound then diffuses to the dermal vasculature and circulates to the liver and kidney, where it is converted to the dihydroxylated functional hormone 1,25-dihydroxyvitamin D_3[1,25$(OH)_2D_3$]. Vitamin D metabolites from the circulation and those produced in the skin itself can augment epidermal differentiation signaling and inhibit keratinocyte proliferation. These effects on keratinocytes are used therapeutically in psoriasis with the topical application of synthetic vitamin D analogues. In addition, vitamin D is increasingly recognized to have beneficial effects in several other inflammatory conditions, and there is some evidence to suggest that it is associated with a reduced risk for various internal malignancies, aside from its classic physiologic effects on calcium metabolism and bone homeostasis. There is controversy regarding the risk/benefit of sun exposure in vitamin D homeostasis. At present, it is important to emphasize that there is no clear-cut evidence to suggest that the use of sunscreens substantially diminishes vitamin D levels. Since aging also substantially decreases the ability of human skin to photocatalytically produce vitamin D_3, the widespread use of sunscreens that filter out UV-B has led to the concern that vitamin D deficiency may become a significant clinical problem in the elderly. However, the amount of sunlight needed to produce sufficient vitamin D is small and does not justify increased sun exposure or tanning behavior. Increased nutritional supplementation of vitamin D is advocated in patients with vitamin D deficiency rather than increasing UV exposure that is associated with carcinogenic and other types of photodamage.

CHRONIC EFFECTS OF SUN EXPOSURE: NONMALIGNANT

The clinical features of photodamaged sun-exposed skin consist of wrinkling, blotchiness, and telangiectasia and a roughened, irregular, "weather-beaten" leathery appearance. Whether this photoaging represents accelerated chronologic aging or a separate and distinct process is not clear.

Within chronically sun-exposed epidermis, there is thickening (acanthosis) and morphologic heterogeneity within the basal cell layer. Higher but irregular melanosome content may be present in some keratinocytes, indicating prolonged residence of the cells in the basal cell layer. These structural changes may help explain the leathery texture and the blotchy discoloration of sun-damaged skin.

UV-A is important in the pathogenesis of photoaging in human skin, and ROS probably are involved. The dermis and its connective tissue matrix are the major site for sun-associated chronic damage, manifest as solar elastosis, a massive increase in thickened irregular masses of abnormal-appearing elastic fibers. Collagen fibers are also abnormally clumped in the deeper dermis of sun-damaged skin. The chromophore(s), the action spectra, and the specific biochemical events orchestrating these changes are only partially understood, although UV-A seems to be primarily involved. This could be due to the predominance of UV-A in the solar energy reaching the earth's surface as well as the fact that UV-A penetrates more deeply into the dermis. Chronologically aged sun-protected skin and photoaged skin share important molecular features, including connective tissue damage and elevated matrix metalloproteinases (MMPs). MMPs are enzymes involved in the degradation of the extracellular matrix, and UV-A induces expression of MMPs, including MMP-1 and MMP-3, leading to increased collagen breakdown. In addition, UV-A reduces type I procollagen mRNA expression. Thus, chronic UVR reduces functional collagen content in the dermis. Based on these observations, it is not surprising that high-dose UV-A phototherapy may have beneficial effects in some patients with localized fibrotic diseases of the skin, such as localized scleroderma.

CHRONIC EFFECTS OF SUN EXPOSURE: MALIGNANT

One of the major known consequences of chronic skin exposure to sunlight is nonmelanoma skin cancer. The two most common types of nonmelanoma skin cancer are *basal cell carcinoma* (BCC) and *squamous cell carcinoma* (SCC; Chap. 87). A model for skin cancer induction involves three major steps: initiation, promotion, and progression. Exposure of human skin to sunlight results in *initiation*, a step by which structural (mutagenic) changes in DNA evoke an irreversible alteration in the target cell (*keratinocyte*) that begins the tumorigenic process. Exposure to a tumor initiator such as UV-B is believed to be a necessary but not sufficient step in the malignant process, since initiated skin cells not exposed to tumor promoters generally do not develop tumors. The second stage in tumor development is *promotion*, a multistep process by which chronic exposure to sunlight evokes further changes that culminate in the clonal expansion of initiated cells and cause the development, over many years, of premalignant growths known as *actinic keratoses*, a minority of which may progress to form SCCs. Based on extensive studies, it seems clear that UV-B is a *complete carcinogen*, meaning that it can act as both a tumor initiator and a promoter.

The third and final step in the malignant process is *malignant conversion* of benign precursors into malignant lesions, a process thought to require additional genetic alterations.

On a molecular level skin carcinogenesis is thought to be caused by the accumulation of gene mutations that result in inactivation of tumor suppressors, activation of oncogenes, or reactivation of cellular signaling pathways that normally are expressed only during embryologic development of the epidermis.

Accumulation of mutations in the tumor-suppressor gene *p53* as a result of UV-induced DNA damage has been found in both SCCs and BCCs and is probably important in promoting skin carcinogenesis. Indeed, both human and murine UV-induced skin cancers have characteristic *p53* mutations (C → T and CC → TT transitions) that are present in the majority of these lesions. Studies in mice have shown that sunscreens can substantially reduce the frequency of these signature mutations in *p53* and dramatically inhibit the induction of tumors.

BCCs also manifest inactivating mutations in the tumor-suppressor gene known as *patched*, which results in activation of hedgehog signaling, and enhanced activity of *smoothened*, which in turn causes downstream activation of transcription factors that augment cell proliferation. Thus, these tumors can manifest mutations in both p53 and *patched* or smothered. New evidence links alterations in the Wnt/ß-catenin signaling pathway known to be critical for hair follicle development, to skin cancer as well. Thus interactions between this pathway and the hedgehog signaling pathway appear to be involved in both skin carcinogenesis and the embryologic development of the skin and hair follicles.

Studies in mouse models have strengthened the hypothesis that SCCs and BCCs are derived from cells in the epidermis with properties of stem cells, and the cells of origin for these skin cancers may be derived from the hair bulge region. The transcription factor Myc is important for stem cell maintenance in the skin, and oncogenic activation of Myc has been implicated in the development of BCCs and SCCs. Thus, nonmelanoma skin cancer involves mutations and alterations in multiple genes and pathways that occur as a result of the chronic accumulation of such changes promoted by exposure to environmental factors such as solar radiation.

Sun exposure causes nonmelanoma cancers and melanoma of the skin, although the evidence is far more direct for its role in nonmelanoma skin cancer (BCC and SCC) than its role in melanoma. Approximately 80% of nonmelanoma skin cancers develop on sun-exposed body areas, including the face, neck, and hands. Major risk factors include male sex, childhood sun exposures, older age, fair

skin, and residence at latitudes closer to the equator. Whites of darker complexions (e.g., Hispanics) have one-tenth the risk of developing such cancers compared with fair-skinned individuals. Blacks are at substantially reduced risk for skin cancer. More than 1.3 million individuals in the United States develop nonmelanoma skin cancer annually, and the lifetime risk for a fair-skinned individual to develop such a neoplasm is estimated at ~15%. A consensus exists that the incidence of nonmelanoma skin cancer in the population is increasing at a rate of 2–3% per year for unknown reasons. One potential explanation is the widespread use of indoor tanning. It is estimated that 30 million people tan indoors in the United States annually, including >2 million adolescents.

The relationship of sun exposure to melanoma development is less clear-cut, but suggestive evidence supports an association. The strongest risk factors for melanoma include positive family history for melanoma, multiple dysplastic nevi, and prior melanoma. Melanomas occasionally develop by the teenage years, indicating that the latent period for tumor growth is less than that for nonmelanoma skin cancer. Melanomas are among the most rapidly increasing among all human malignancies (Chap. 87). Epidemiologic studies of immigrant populations of similar ethnic background indicate that individuals who are born in one area or who migrate to the same locale before age 10 have higher age-specific melanoma rates than do individuals arriving later. It is thus reasonable to conclude that life in a sunny climate from birth or early childhood increases the risk of melanoma. In general, risk does not correlate with cumulative sun exposure but may be related to the duration and extent of exposure in childhood. Epidemiologic studies have shown that indoor tanning is a risk factor for melanoma.

Meta-analysis of 17 case-control studies in patients with melanoma concluded that the protective effect of sunscreens against this type of tumor could not be substantiated, but this probably is due to failure to control for confounding factors such as sunscreen stability and frequency of application. Since no prospective studies are available to address this issue, it seems reasonable to recommend that patients at risk for melanoma utilize photoprotection such as sun avoidance, high sun protective factor (SPF) sunscreens, and protective clothing.

IMMUNOLOGIC EFFECTS

Exposure to solar radiation causes local (inhibition of immune responses to antigens applied at the irradiated site) and systemic (inhibition of immune responses to antigens applied at remote unirradiated sites) immunosuppression. For example, administration of modest doses of UV-B to human skin reduces the degree of allergic sensitization to the potent contact allergen dinitrochlorobenzene. This is associated with depletion of epidermal LCs.

An example of the systemic immunosuppressive effects of higher doses of UVR is the diminished immunologic response to antigens introduced either epicutaneously or intracutaneously at sites distant from the irradiated site.

The major chromophores in the upper epidermis that initiate UV-mediated immunosuppression include DNA, *trans*-urocanic acid, and membrane components. The action spectrum for UV-induced immunosuppression closely mimics the absorption spectrum of DNA. Pyrimidine dimers in LCs may inhibit antigen presentation. The absorption spectrum of epidermal urocanic acid closely mimics the action spectrum for UV-B-induced immunosuppression as well. Urocanic acid is a metabolic product of the essential amino acidhistidine and accumulates in the upper epidermis through breakdown of the histidine-rich protein filaggrin due to the absence of its catabolizing enzyme in keratinocytes. It is synthesized as a trans-isomer, and UV-induced trans-cis isomerization of urocanic acid in the stratum corneum leads to its immunosuppressive effects. *Cis*-urocanic acid has been proposed to exert its immunosuppressive effects through a variety of mechanisms, including inhibition of antigen presentation by LCs.

One important consequence of chronic sun exposure and the concomitant immunosuppression is enhanced risk of skin cancer. The molecular mechanisms of photocarcinogenesis are complex and only partially understood. In part, UV-B activates regulatory T cells that suppress antitumor immune responses via IL-10 expression, whereas in the absence of high UV-B exposure, epidermal antigen-presenting cells present tumor-associated antigens and induce protective immunity, thereby inhibiting skin tumorigenesis. UV-induced DNA damage is a major molecular trigger of this immunosuppressive effect.

Perhaps the most graphic demonstration of the role of immunosuppression in enhancing the risk of nonmelanoma skin cancer has come from studies of patients who are organ transplant recipients who are chronically treated with immunosuppressive antirejection drug regimens. More than 50% of transplant patients develop BCCs and SCCs, and these cancers are the most common malignancies arising in immunosuppressed solid-organ transplant recipients. Rates of BCCs and SCCs increase with the duration and the degree of immunosuppression. These patients require close periodic monitoring and rigorous photoprotection through the use of sunscreens, protective clothing, and sun avoidance.

Photosensitivity Diseases

The diagnosis of photosensitivity requires a careful history to define the duration of the signs and symptoms, the length of time between exposure to sunlight and the development of subjective

complaints, and visible changes in the skin. The age of onset can also be a helpful clue; for example, the acute photosensitivity of erythropoietic protoporphyria almost always begins in childhood, whereas the chronic photosensitivity of porphyria cutanea tarda (PCT) typically begins in the fourth and fifth decades. A history of exposure to topical and systemic drugs and chemicals may provide important clues. Many classes of drugs can cause photosensitivity on the basis of either phototoxicity or photoallergy. Fragrances such as musk ambrette that were previously present in numerous cosmetic products are also potent photosensitizers.

Examination of the skin also may offer important clues. Anatomic areas that are naturally protected from direct sunlight, such as the hairy scalp, the upper eyelids, the retroauricular areas, and the infranasal and submental regions, may be spared, whereas exposed areas show characteristic features of the pathologic process. These anatomic localization patterns are often helpful but not infallible in making the diagnosis. For example, airborne contact sensitizers that are blown onto the skin may produce a dermatitis that can be difficult to distinguish from photosensitivity despite the fact that such material may trigger skin reactivity in areas shielded from direct sunlight.

Many dermatologic conditions may be caused or aggravated by sunlight **(Table 56-2)**. The role of light in evoking these responses may be dependent on genetic abnormalities ranging from well-described defects in DNA repair that occur in XP to the inherited abnormalities in heme synthesis that characterize the porphyrias. In certain photosensitivity diseases the chromophore has been identified, whereas in the majority, the energy-absorbing agent is unknown.

Table 56-2 Classification of Photosensitivity Diseases

Type	Disease
Genetic	Erythropoietic porphyria
	Erythropoietic protoporphyria
	Porphyria cutanea tarda—familial
	Variegate porphyria
	Hepatoerythropoietic porphyria
	Albinism
	Xeroderma pigmentosum
	Rothmund-Thomson syndrome
	Bloom syndrome
	Cockayne's disease
	Kindler syndrome
	Phenylketonuria
Metabolic	Porphyria cutanea tarda—sporadic
	Hartnup disease
	Kwashiorkor
	Pellagra
	Carcinoid syndrome
Phototoxic	
Internal	Drugs
External	Drugs, plants, food
Photoallergic	
Immediate	Solar urticaria
Delayed	Drug photoallergy
	Persistent light reaction/chronic actinic dermatitis
Neoplastic and degenerative	Photoaging
	Actinic keratosis
	Melanoma and nonmelanoma skin cancer
Idiopathic	Polymorphous light eruption
	Hydroa aestivale
	Actinic prurigo

Photoaggravated	Lupus erythematosus Systemic Subacutecutaneous Discoid Dermatomyositis Herpes simplex Lichen planus actinicus Acnevulgaris (aestivale)

POLYMORPHOUS LIGHT ERUPTION

After sunburn, the most common type of photosensitivity disease is *polymorphous light eruption* (PLE). Many affected individuals never seek medical attention because the condition is often transient, becoming manifest each spring with initial sun exposure but then subsiding spontaneously with continuing exposure, a phenomenon known as "hardening." The major manifestations of PLE include pruritic (often intensely so) erythematous papules that may coalesce into plaques in a patchy distribution on exposed areas of the trunk and forearms. The face is usually less seriously involved. Whereas the morphologic skin findings remain similar for each patient with subsequent recurrences, significant interindividual variations in skin findings are characteristic (hence the term "polymorphous").

The diagnosis can be confirmed by skin biopsy and by performing phototest procedures in which skin is exposed to multiple erythema doses of UV-A and UV-B. The action spectrum for PLE is usually within these portions of the solar spectrum.

Whereas the treatment of an acute flare of PLE may require topical or systemic glucocorticoids, approaches to prevent PLE are important and include the use of broad-spectrum sunscreens and the induction of hardening by the cautious administration of artificial UV-B (broad-band or narrow-band) and/or UV-A radiation or the use of psoralen plus UV-A (PUVA) photochemotherapy for 2–4 weeks before initial sun exposure. Such prophylactic phototherapy or photochemotherapy at the beginning of spring may prevent the occurrence of PLE throughout the summer.

PHOTOTOXICITY AND PHOTOALLERGY

These photosensitivity disorders are related to the topical or systemic administration of drugs and other chemicals. Both reactions require the absorption of energy by a drug or chemical resulting in the production of an excited-state photosensitizer that can transfer its absorbed energy to a bystander molecule or to molecular oxygen, thereby generating tissue-destructive chemical species, including ROS.

Phototoxicity is a nonimmunologic reaction caused by drugs and chemicals, a few of which are listed in **Table 56-3**. The usual clinical manifestations include erythema resembling a sunburn reaction that quickly desquamates, or "peels," within several days. In addition, edema, vesicles, and bullae may occur.

Table 56-3 Phototoxic Drugs

	Topical	Systemic
Amiodarone		+
Dacarbazine		+
Fluoroquinolones		+
5-Fluorouracil	+	+
Furosemide		+
Nalidixic acid		+
Phenothiazines		+
Psoralens	+	+
Retinoids	+/	+
Sulfonamides		+
Sulfonylureas		+
Tetracyclines		+
Thiazides		+
Vinblastine		+

Photoallergy is much less common and is distinct in that it is an immunopathologic process. The

excited-state photosensitizer may create highly unstable haptenic free radicals that bind covalently to macromolecules to form a functional antigen capable of evoking a delayed hypersensitivity response. Some of the drugs and chemicals that produce photoallergy are listed in **Table 56-4**. The clinical manifestations typically differ from those of phototoxicity in that an intensely pruritic eczematous dermatitis tends to predominate and evolves into lichenified, thickened, "leathery" changes in sun-exposed areas. A small subset (perhaps 5–10%) of patients with photoallergy may develop a persistent exquisite hypersensitivity to light even when the offending drug or chemical is identified and eliminated, a condition known as *persistent light reaction*.

Table 56-4 Photoallergic Drugs

	Topical	Systemic
6-Methylcoumarin	+	
Aminobenzoic acid and esters	+	
Bithionol	+	
Chlorpromazine		+
Diclofenac		+
Fluoroquinolones		+
Halogenated salicylanilides	+	
Hypericin (St John's wort)	+	+
Musk ambrette	+	
Piroxicam		+
Promethazine		+
Sulfonamides		+
Sulfonylureas		+

A very uncommon type of persistent photosensitivity is known as *chronic actinic dermatitis*. These patients are typically elderly men with a long history of preexisting allergic contact dermatitis or photosensitivity. They are usually exquisitely sensitive to UV-B, UV-A, and visible wavelengths.

Diagnostic confirmation of phototoxicity and photoallergy often can be obtained by using phototest procedures. In patients with suspected phototoxicity, determining the minimal erythema dose (MED) while the patient is exposed to a suspected agent and then repeating the MED after discontinuation of the agent may provide a clue to the causative drug or chemical. Photopatch testing can be performed to confirm the diagnosis of photoallergy. This is a simple variant of ordinary patch testing in which a series of known photoallergens is applied to the skin in duplicate and one set is irradiated with a suberythema dose of UV-A. Development of eczematous changes at sites exposed to sensitizer and light is a positive result. The characteristic abnormality in patients with persistent light reaction is a diminished threshold to erythema evoked by UV-B. Patients with chronic actinic dermatitis usually manifest a broad spectrum of UV hyperresponsiveness and require meticulous photoprotection, including avoiding sun exposure, high (>30) SPF sunscreens, and in severe cases systemic immunosuppression, preferably with azathioprine (1–2 mg/kg per day).

The management of drug photosensitivity involves first and foremost the elimination of exposure to the chemical agents responsible for the reaction and minimization of sun exposure. The acute symptoms of phototoxicity may be ameliorated by cool moist compresses, topical glucocorticoids, and systemically administered NSAIDs. In severely affected individuals, a rapidly tapered course of systemic glucocorticoids may be useful. Judicious use of analgesics may be necessary.

Photoallergic reactions require a similar management approach. Furthermore, patients with persistent light reaction and chronic actinic dermatitis must be meticulously protected against light exposure. In selected patients in whom chronic systemic high-dose glucocorticoids pose unacceptable risks, it may be necessary to employ an immunosuppressive drug such as azathioprine, cyclophosphamide, cyclosporine, or mycophenolate mofetil.

PORPHYRIA

The porphyrias (Chap. 358) are a group of diseases that have in common inherited or acquired derangements in the synthesis of heme. Heme is an iron-chelated tetrapyrrole or porphyrin, and the nonmetal chelated porphyrins are potent photosensitizers that absorb light intensely in both the short (400–410 nm) and the long (580–650 nm) portions of the visible spectrum.

Heme cannot be reutilized and must be synthesized continuously, and the two body compartments with the largest capacity for its production are the bone marrow and the liver. Accordingly, the porphyrias originate in one or the other of these organs, with the end result of excessive endogenous production of potent photosensitizing porphyrins. The porphyrins circulate in the bloodstream and diffuse into the skin, where they absorb solar energy, become photoexcited, generate ROS, and evoke cutaneous

photosensitivity. The mechanism of porphyrin photosensitization is known to be photodynamic, or oxygen-dependent, and is mediated by ROS such as singlet oxygen and superoxide anions.

Porphyria cutanea tarda (PCT) is the most common type of human porphyria and is associated with decreased activity of the enzyme uroporphyrinogen decarboxylase. There are two basic types of PCT: (1) the sporadic or acquired type, generally seen in individuals ingesting ethanol or receiving estrogens, and (2) the inherited type, in which there is autosomal dominant transmission of deficient enzyme activity. Both forms are associated with increased hepatic iron stores.

In both types of PCT, the predominant feature is a chronic photosensitivity characterized by increased fragility of sun-exposed skin, particularly areas subject to repeated trauma, such as the dorsa of the hands, the forearms, the face, and the ears. The predominant skin lesions are vesicles and bullae that rupture, producing moist erosions, often with a hemorrhagic base, that heal slowly with crusting and purplish discoloration of the affected skin. Hypertrichosis, mottled pigmentary change, and scleroderma-like induration are associated features. Biochemical confirmation of the diagnosis can be obtained by measurement of urinary porphyrin excretion, plasma porphyrin assay, and assay of erythrocyte and/or hepatic uroporphyrinogen decarboxylase. Multiple mutations of the uroporphyrinogen decarboxylase gene have been identified in human populations. Some patients with PCT have associated mutations in the *HFE* gene, which is linked to hemochromatosis. This could contribute to the iron overload seen in PCT, although iron status as measured by serum ferritin, iron levels, and transferrin saturation is no different from that in PCT patients without *HFE* mutations. Prior hepatitis C virus infection appears to be an independent risk factor for PCT.

Treatment of PCT consists of repeated phlebotomies to diminish the excessive hepatic iron stores and/or intermittent low doses of the antimalarial drugs chloroquine and hydroxychloroquine. Long-term remission of the disease can be achieved if the patient eliminates exposure to porphyrinogenic agents.

Erythropoietic protoporphyria originates in the bone marrow and is due to a decrease in the mitochondrial enzyme ferrochelatase secondary to numerous gene mutations. The major clinical features include an acute photosensitivity characterized by subjective burning and stinging of exposed skin that often develops during or just after sun exposure. There may be associated skin swelling and, after repeated episodes, a waxlike scarring.

The diagnosis is confirmed by demonstration of elevated levels of free erythrocyte protoporphyrin. Detection of increased plasma protoporphyrin helps differentiate lead poisoning and iron-deficiency anemia, in both of which elevated erythrocyte protoporphyrin levels occur in the absence of cutaneous photosensitivity and elevated plasma protoporphyrin levels.

Treatment consists of reducing sun exposure and the oral administration of the carotenoid β-carotene, which is an effective scavenger of free radicals. This drug increases tolerance to sun exposure in many affected individuals, although it has no effect on deficient ferrochelatase.

An algorithm for managing patients with photosensitivity is illustrated in **Fig. 56-1**.

Figure 56-1

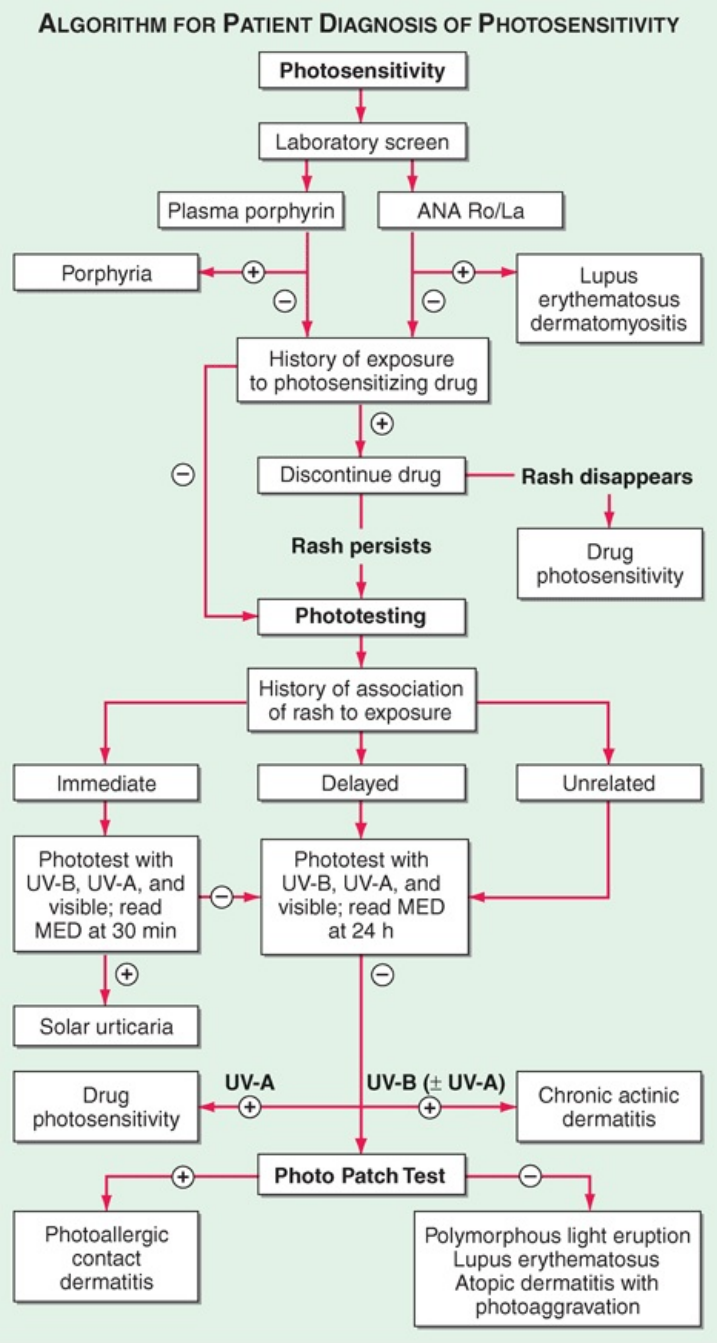

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An algorithm for the diagnosis of a patient with photosensitivity.

PHOTOPROTECTION

Since photosensitivity of the skin results from exposure to sunlight, it follows that absolute avoidance of the sun will eliminate these disorders. Unfortunately, contemporary lifestyles make this an impractical alternative for most individuals, and this has led to a search for better approaches to photoprotection.

Natural photoprotection is provided by structural proteins in the epidermis, particularly keratins and melanin. The amount of melanin and its distribution in cells are genetically regulated, and individuals of darker complexion (skin types IV–VI) are at decreased risk for the development of acute sunburn and cutaneousmalignancy.

Other forms of photoprotection include clothing and sunscreens. Clothing constructed of tightly woven sun-protective fabrics, irrespective of color, affords substantial protection. Wide-brimmed hats, long sleeves, and trousers all reduce direct exposure. Sunscreens are now considered over-the-counter drugs, and category I ingredients are recognized by the U.S. Food and Drug Administration (FDA) as monographed and safe and effective. Those ingredients are listed in **Table 56-5**. Sunscreens are

rated for their photoprotective effect by their sun protection factor. The SPF is simply a ratio of the time required to produce sunburn erythema with and without sunscreen application. Thus, the SPF of sunscreens reflects mainly protection from UV-B and does not reflect protection from UV-A. The monograph stipulates that sunscreens must be rated on a scale ranging from minimal (SPF ≥2 and <12) to moderate (SPF ≥12 and <30) to high (SPF ≥30, labeled as 30+).

Table 56-5 FDA Category 1 Monographed Sunscreen Ingredients[a]

Ingredients	Maximum Concentration, %
p-Aminobenzoic acid (PABA)	15
Avobenzone	3
Cinoxate	3
Dioxybenzone (benzophenone-8)	3
Ecamsule[b]	15
Homosalate	15
Menthyl anthranilate	5
Octocrylene	10
Octyl methoxycinnamate	7.5
Octyl salicylate	5
Oxybenzone (benzophenone-3)	6
Padimate O (octyl dimethyl PABA)	8
Phenylbenzimidazole sulfonic acid	4
Sulisobenzone (benzophenone-4)	10
Titanium dioxide	25
Trolamine salicylate	12
Zinc oxide	25

[a] FDA, U.S. Food and Drug Administration.

[b] Recently approved by the FDA.

Broad-spectrum sunscreens contain in addition to UV-B-absorbing chemicals UV-A-absorbing chemicals, such as avobenzone and ecamsule (terephthalylidene dicamphor sulfonic acid). These chemicals absorb UVR and transfer the absorbed energy to surrounding cells either as heat or by emission of fluorescence or phosphorescence. In contrast, physical UV blockers (zinc oxide and titanium dioxide) scatter or reflect UVR.

In addition to light absorption, a critical determinant of the sustained photoprotective effect of sunscreens is their water resistance. The FDA monograph has defined strict testing criteria for sunscreens making this claim.

Some degree of photoprotection can be achieved by limiting the time of exposure during the day. Since a large part of an individual's total lifetime sun exposure may occur by age 18, it is important to educate parents and young children about the hazards of sunlight. Simply eliminating exposure at midday will substantially reduce lifetime UV-B exposure.

PHOTOTHERAPY AND PHOTOCHEMOTHERAPY

UVR also can be used therapeutically. The administration of UV-B alone or in combination with topically applied agents can induce remissions of many dermatologic diseases, including psoriasis and atopic dermatitis. In particular, narrow-band UV-B treatments (with fluorescent bulbs emitting radiation at ~311 nm) have enhanced efficacy compared with broad-band UV-B in the treatment of psoriasis.

Photochemotherapy in which topically applied or systemically administered psoralens are combined with UV-A (PUVA) is also effective in treating psoriasis and the early stages of cutaneous T cell lymphoma and vitiligo. Psoralens are tricyclic furocoumarins that, when intercalated into DNA and exposed to UV-A, form adducts with pyrimidine bases and eventually form DNA cross-links. These structural changes are thought to decrease DNA synthesis and relate to the improvement that occurs in psoriasis. The reason why PUVA photochemotherapy is effective in cutaneous T cell lymphoma is not clear, but it has been shown to induce apoptosis of atypical T-lymphocyte populations in the skin. Consequently, direct treatment of circulating atypical lymphocytes by extracorporeal photochemotherapy (photopheresis) has been used in Sézary syndrome as well as in other severe systemic diseases with circulating atypical lymphocytes, such as graft-versus-host disease.

In addition to its effects on DNA, PUVA photochemotherapy stimulates epidermal thickening and melanin synthesis; the latter provides the rationale for its use in the depigmenting disease vitiligo, together with its anti-inflammatory effects. Oral 8-methoxypsoralen and UV-A appear to be most effective in this regard, but as many as 100 treatments extending over 12–18 months may be required to promote satisfactory repigmentation.

Not surprisingly the major side effects of long-term UV-B phototherapy and PUVA photochemotherapy mimic those seen in individuals with chronic sun exposure and include skin dryness, actinic keratoses, and an increased risk of skin cancer. Despite these risks, the therapeutic index of these modalities continues to be excellent. It is important to choose the most appropriate phototherapy treatment approach for a specific dermatologic disease. For example, narrow-band UV-B has been reported in several studies to be as effective as PUVA photochemotherapy in the treatment of psoriasis but to have a lower risk of skin cancer development than PUVA.

FURTHER READING

Benjamin CL, Ananthaswamy HN: p53 and the pathogenesis of skin cancer. Toxicol Appl Pharmacol 224:221, 2007

Cui R et al: Central role of p53 in the suntan response and pathologic hyperpigmentation. Cell 128:853, 2007 [PMID: 17350573]

Deeb KK et al: Vitamin D signaling pathways in cancer: Potential for anticancer therapeutics. Nat Rev Cancer 7:684, 2007 [PMID: 17721433]

Epstein EH: Basal cell carcinomas: Attack of the hedgehog. Nat Rev Cancer 8(10):1275, 2008

Laga AC, Murphy GF: The translational basis of human cutaneous photoaging. Am J Pathol 174:357, 2009 [PMID: 19147830]

Malanchi I et al: Cutaneous cancer stem cell maintenance is dependent on beta-catenin signalling. Nature 452:650, 2008 [PMID: 18385740]

Miller AJ, Mihm MC Jr: Melanoma. N Engl J Med 355:51, 2006 [PMID: 16822996]

Morison WL: Photosensitivity. N Engl J Med 350:111, 2004

Sage RJ, Lim HW: Therapeutic Hotline: Recommendations on photoprotection and vitamin D. Dermatol Ther 23:82, 2010?

Schade N et al: Ultraviolet B radiation–induced immunosuppression: Molecular mechanisms and cellular alterations. Photochem Photobiol Sci 3:699, 2005

Wong TH, Rees JL: The relation between melanocortin I receptor variation and generation of phenotypic diversity in the cutaneous response to ultraviolet radiation. Peptides 26:1965, 2005 [PMID: 15963603]

Yaar M, Gilchrest BA: Photoageing: Mechanism, prevention and therapy. Br J Dermatol 157: 874, 2007 [PMID: 17711532]

Yang SH et al: Pathological responses to oncogenic Hedgehog signaling in skin are dependent on canonical Wnt/β-catenin signaling. Nat Genet 40: 1130, 2008 [PMID: 19165927]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF SKIN MANIFESTATIONS OF INTERNAL DISEASE: INTRODUCTION

In the practice of medicine, virtually every clinician encounters patients with skin disease. Physicians of all specialties face the daily task of determining the nature and clinical implication of dermatologic disease. In patients with skin eruptions and rashes, the physician must confront the question of whether the cutaneous process is confined to the skin, representing a pure dermatologic event, or whether it is a manifestation of internal disease relating to the patient's overall medical condition. Evaluation and accurate diagnosis of skin lesions are also critical given the marked rise in both melanoma and nonmelanoma skin cancer. Dermatologic conditions can be classified and categorized in many different ways, and in this Atlas, a selected group of inflammatory skin eruptions and neoplastic conditions are grouped in the following manner: (A) common skin diseases and lesions, (B) nonmelanoma skin cancer, (C) melanoma and pigmented lesions, (D) infectious disease and the skin, (E) immunologically mediated skin disease, and (F) skin manifestations of internal disease.

COMMON SKIN DISEASES AND LESIONS

(Figs. e16-1 to e16-19) In this section, several common inflammatory skin diseases and benign neoplastic and reactive lesions are presented. While most of these dermatoses usually present as a predominantly dermatologic process, underlying systemic associations may be made in some settings. Atopic dermatitis is often present in patients with an atopic diathesis, including asthma and sinusitis. Psoriasis ranges from limited patches on the elbows and knees to severe erythrodermic involvement and associated psoriatic arthritis. Some patients with alopecia areata may have an underlying thyroid abnormality requiring screening. Finally, even acnevulgaris, one of the most common inflammatory dermatoses, can be associated with a systemic process such as polycystic ovarian syndrome.

Figure e16-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acnevulgaris with inflammatory papules, pustules, and comedones. *(Courtesy of Kalman Watsky, MD; with permission.)*

Figure e16-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acne rosacea with prominent facial erythema, telangiectasias, scattered papules, and small pustules. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Psoriasis is characterized by small and large erythematous plaques with adherent silvery scale.

Figure e16-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atopic dermatitis with hyperpigmentation, lichenification, and scaling in the antecubital fossae. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dyshidrotic eczema, characterized by deep-seated vesicles and scaling on palms and lateral fingers, is often associated with an atopic diathesis.

Figure e16-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Seborrheic dermatitis showing erythema and scale in the nasolabial fold *(Courtesy of Robert A. Swerlick, MD; with permission.)*

Figure e16-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stasis dermatitis showing erythematous, scaly, and oozing patches over the lower leg. Several stasis ulcers are also seen in this patient.

Figure e16-8

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***A.*Allergic contact dermatitis**, acute phase, with sharply demarcated, weeping, eczematous plaques in a perioral distribution. ***B.*Allergic contact dermatitis to nickel,** chronic phase demonstrating an erythematous, lichenified, weeping plaque on skin chronically exposed to a metal snap. *(**B,** Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-9

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lichen planus showing multiple flat-topped, violaceous papules and plaques. Nail dystrophy as seen

in this patient's thumbnail may also be a feature. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-10

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Seborrheic keratoses are seen as "stuck on," waxy, verrucous papules and plaques with a variety of colors ranging from light tan to black.

Figure e16-11

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vitiligo in a typical acral distribution demonstrating striking cutaneous depigmentation, as a result of loss of melanocytes.

Figure e16-12

Alopecia areata characterized by a sharply demarcated circular patch of scalp completely devoid of hairs. Follicular orifices are preserved, indicating a nonscarring alopecia. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-13

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pityriasis rosea. Multiple round to oval erythematous patches with fine central scale are distributed along the skin tension lines on the trunk.

Figure e16-14

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***A.*Urticaria** showing characteristic discrete and confluent, edematous, erythematous papules and plaques. ***B.*Dermatographism.** Erythema and whealing that developed after firm stroking of the skin. *(**B,** Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-15

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Epidermoid cysts. Several inflamed and noninflamed firm, cystic nodules are seen in this patient. Often a patulous follicular punctum is observed on the overlying epidermal surface.

Figure e16-16

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Keloids resulting from ear piercing, with firm exophytic flesh-colored to erythematous nodules of scar tissue.

Figure e16-17

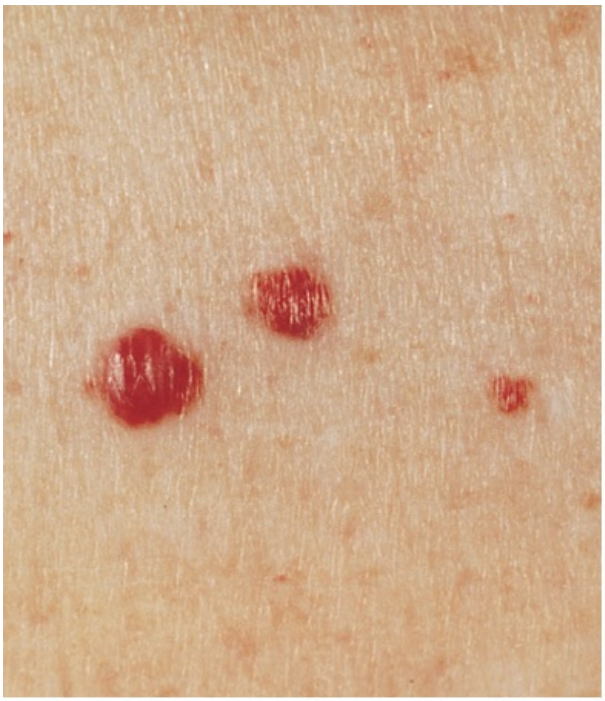

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cherry hemangiomas are very common and arise in middle-aged to older adults. They are characterized by multiple erythematous to dark-purple papules, usually located on the trunk.

Figure e16-18

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Frostbite with vesiculation, surrounded by edema and erythema. *(Courtesy of Daniel F. Danzl, MD; with permission.)*

Figure e16-19

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Frostbite with vesiculation, surrounded by edema and erythema. *(Courtesy of Daniel F. Danzl, MD; with permission.)*

NONMELANOMA SKIN CANCER

(Figs. e16-20 to e16-27) In more fair-skinned ethnic populations, nonmelanoma skin cancer is increasing at an alarming rate. Basal cell carcinoma is the most common cancer in humans and is strongly linked to ultraviolet radiation. Squamous cell carcinoma, including keratoacanthoma-type squamous cell carcinoma, is the second most common skin cancer in most ethnic populations and is also most commonly linked to ultraviolet radiation. Less common cutaneousmalignancies include cutaneous T cell lymphoma (mycosis fungoides) and carcinoma and lymphoma metastatic to skin.

Figure e16-20

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Kaposi's sarcoma in a patient with AIDS demonstrating patch, plaque, and tumor stages.

Figure e16-21

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Non-Hodgkin's lymphoma involving the skin with typical violaceous, "plum-colored" nodules. *(Courtesy of Jean Bolognia, MD; with permission.)*

Figure e16-22

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Basal cell carcinoma showing central ulceration and a pearly, rolled, telangiectatic tumor border.

Figure e16-23

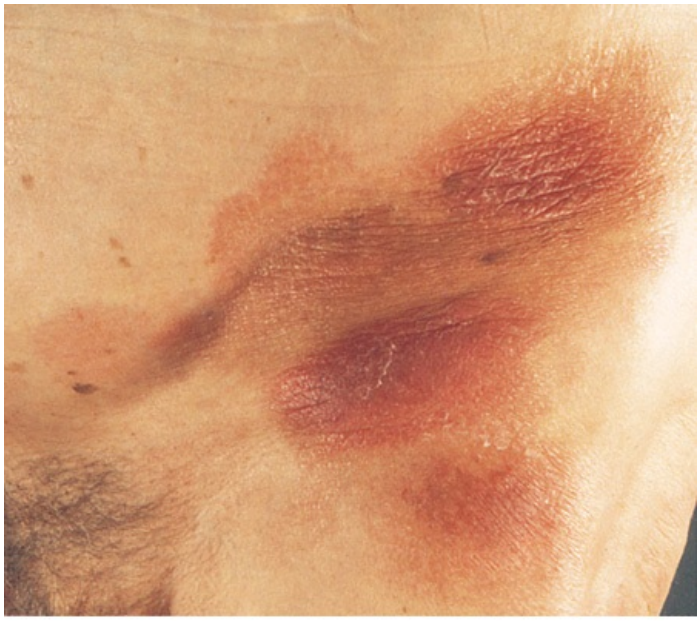

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mycosis fungoides is a cutaneous T cell lymphoma, and plaque stage lesions are seen in this patient.

Figure e16-24

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Metastatic carcinoma to the skin is characterized by inflammatory, often ulcerated dermal nodules.

Figure e16-25

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Keratoacanthoma is a low-grade squamous cell carcinoma that presents as an exophytic nodule with central keratinous debris.

Figure e16-26

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Squamous cell carcinoma seen here as a hyperkeratotic crusted and somewhat eroded plaque on the lower lip. Sun-exposed skin such as the head, neck, hands, and arms are other typical sites of involvement.

Figure e16-27

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Actinic keratoses consist of hyperkeratotic erythematous papules and patches on sun-exposed skin. They arise in middle-aged to older adults and have some potential for malignant transformation. *(Courtesy of Robert Swerlick, MD; with permission.)*

MELANOMA AND BENIGN PIGMENTED LESIONS

(Figs. e16-28 to e16-33) As the prognosis of melanoma is primarily related to microscopic depth of invasion, and early detection with surgical treatment can be curative in a high percentage of patients, it is essential that all clinicians acquire some facility in evaluating pigmented lesions. Three of the clinicopathologic subtypes of melanoma, superficial spreading, lentigo maligna, and acral lentiginous

melanoma, typically display features noted in the "ABCD rule." *A*symmetry—one half of the lesion varies from the other half. *B*order irregularity—the circumferential border exhibits an irregular, sometimes jagged appearance. *C*olor—there is uneven coloration and tone to the pigmented lesion with varying shades of brown, black, red, and white within different areas of the lesion. *D*iameter—is typically >6 mm. The more uncommon subtype, nodular melanoma, may not manifest all these features but present as a more symmetric, evenly pigmented, or amelanotic lesion. Dysplastic (atypical) melanocytic nevi may occur as solitary or multiple lesions as well as in the setting of familial melanoma. They display some degree of asymmetry, border irregularity, and color variation. Ordinary nevi may be acquired or congenital and are quite common.

Figure e16-28

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nevus. Nevi are benign proliferations of nevomelanocytes characterized by regularly shaped hyperpigmented macules or papules of a uniform color.

Figure e16-29

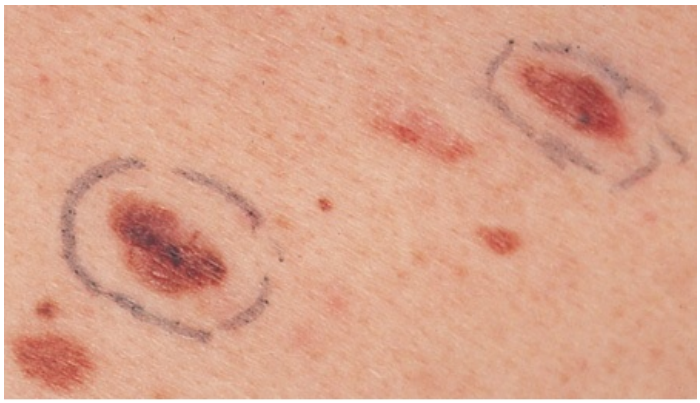

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dysplastic nevi are irregularly pigmented and shaped nevomelanocytic lesions that may be associated with familial melanoma.

Figure e16-30

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Superficial spreading melanoma is the most common type of malignant melanoma and demonstrates color variegation (black, blue, brown, pink, and white) and irregular borders.

Figure e16-31

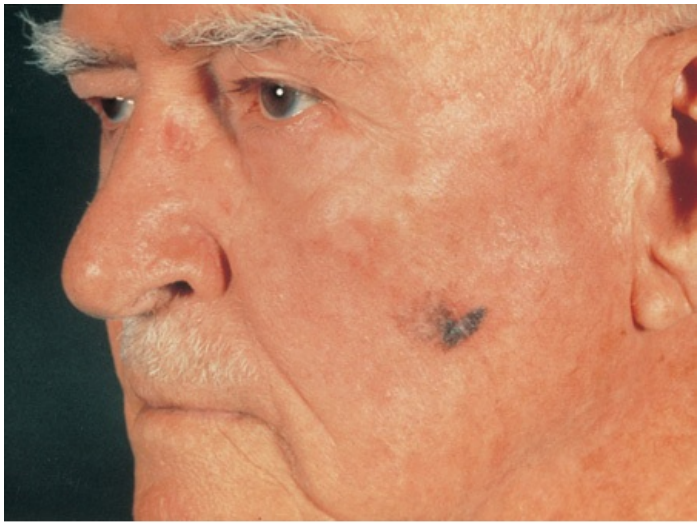

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lentigo maligna melanoma occurs on sun-exposed skin as a large, hyperpigmented macule or plaque with irregular borders and variable pigmentation. *(Courtesy of Alvin Solomon, MD; with permission.)*

Figure e16-32

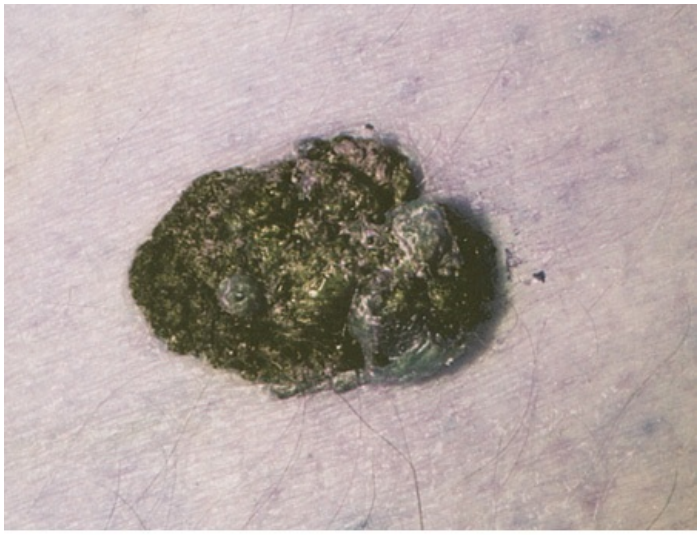

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nodular melanoma most commonly manifests itself as a rapidly growing, often ulcerated or crusted black nodule. *(Courtesy of S. Wright Caughman, MD; with permission.)*

Figure e16-33

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acral lentiginous melanoma is more common in blacks, Asians, and Hispanics and occurs as an enlarging hyperpigmented macule or plaque on the palms or soles. Lateral pigment diffusion is present.

INFECTIOUS DISEASE AND THE SKIN

(Figs. e16-34 to e16-58) One of the roles of the skin is to function as a barrier from the outside world. In this capacity, exposure to infectious agents occurs, and bacterial, viral, fungal, and parasitic infections may result. In addition, the skin may be secondarily involved and provides diagnostic clues to systemic infections such as meningococcemia, Rocky Mountain spotted fever, Lyme disease, and septic emboli. Most sexually transmitted bacterial and viral diseases exhibit cutaneous involvement, and examples include primary and secondary syphilis, chancroid, genital herpes simplex, and condyloma acuminatum.

Figure e16-34

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erysipelas is a streptococcal infection of the superficial dermis and consists of well-demarcated, erythematous, edematous, warm plaques.

Figure e16-35

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Varicella showing numerous lesions in various stages of evolution: vesicles on an erythematous base, umbilicated vesicles, and crusts. *(Courtesy of Robert Hartman, MD; with permission.)*

Figure e16-36

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Herpes zoster is seen in this HIV-infected patient as hemorrhagic vesicles and pustules on an erythematous base grouped in a dermatomal distribution. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-37

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Impetigo contagiosa is a superficial streptococcal or *Staphylococcus aureus* infection consisting of honey-colored crusts and erythematous weeping erosions. Occasionally, bullous lesions may be seen.

Figure e16-38

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tender vesicles and erosions in the mouth of a patient with hand-foot-and-mouth disease. *(Courtesy of Stephen D. Gellis, MD; with permission.)*

Figure e16-39

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lacy reticular rash of erythema infectiosum (fifth disease).

Figure e16-40

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Molluscum contagiosum is a cutaneous poxvirus infection characterized by multiple umbilicated flesh-colored or hypopigmented papules. *(Courtesy of Yale Resident's Slide Collection; with permission.)*

Figure e16-41

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oral hairy leukoplakia often presents as white plaques on the lateral tongue and is associated with Epstein-Barr virus infection. *(From K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005. www.accessmedicine.com.)*

Figure e16-42

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fulminant meningococcemia with extensive angular purpuric patches. *(Courtesy of Stephen D. Gellis, MD; with permission.)*

Figure e16-43

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rocky Mountain spotted fever demonstrating pinpoint petechial lesions on the palm and volar aspect of the wrist. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-44

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema chronicum migrans is the early cutaneous manifestation of Lyme disease and is characterized by erythematous annular patches, often with a central erythematous papule at the tick bite site. *(Courtesy of Yale Resident's Slide Collection; with permission.)*

Figure e16-45

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary syphilis with a firm, nontender chancre. *(Courtesy of Gregory Cox, MD; with permission.)*

Figure e16-46

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Secondary syphilis commonly affects the palms and soles with scaling, firm, red-brown papules. *(Courtesy of Alvin Solomon, MD; with permission.)*

Figure e16-47

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Condylomata lata are moist, somewhat verrucous intertriginous plaques seen in secondary syphilis. *(Courtesy of Yale Resident's Slide Collection; with permission.)*

Figure e16-48

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Secondary syphilis demonstrating the papulosquamous truncal eruption.

Figure e16-49

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tinea corporis is a superficial fungal infection, seen here as an erythematous annular scaly plaque with central clearing.

Figure e16-50

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scabies showing typical scaling erythematous papules and few linear burrows.

Figure e16-51

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Skin lesions caused by *Chironex fleckeri* sting. *(Courtesy of V. Pranava Murthy, MD; with permission.)*

Figure e16-52

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chancroid with characteristic penile ulcers and associated left inguinal adenitis (bubo).

Figure e16-53

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Condylomata acuminata are lesions induced by human papillomavirus and in this patient are seen as multiple verrucous papules coalescing into plaques. *(Courtesy of S. Wright Caughman, MD; with permission.)*

Figure e16-54

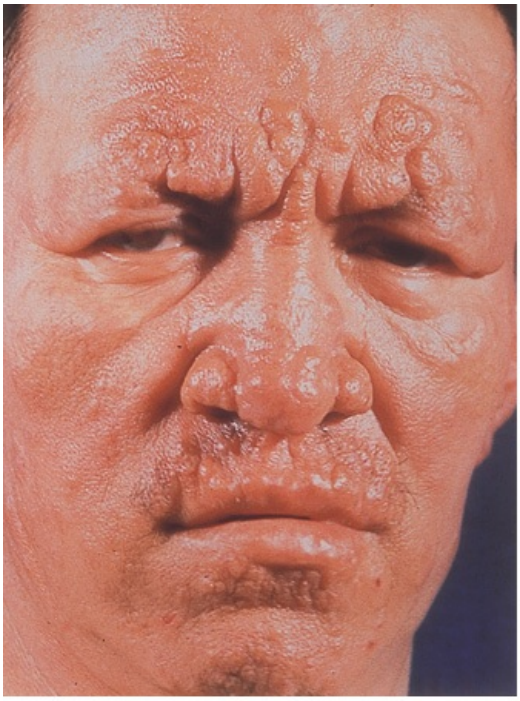

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A patient with features of polar lepromatous leprosy; multiple nodular skin lesions, particularly of the forehead, and loss of eyebrows. *(Courtesy of Robert Gelber, MD; with permission.)*

Figure e16-55

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

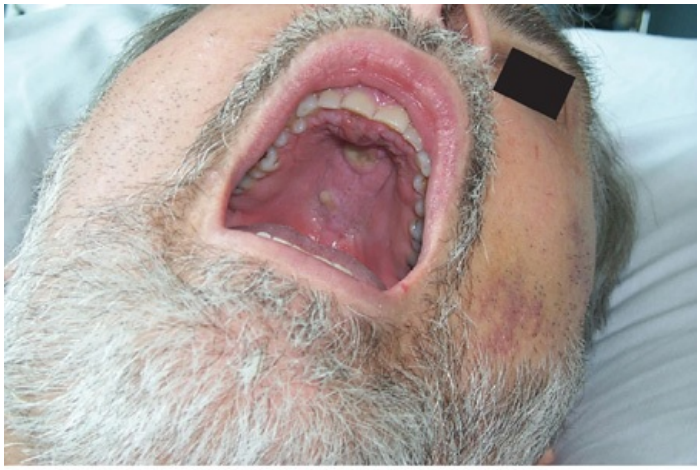

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Skin lesions of neutropenic patients. ***A.*** Hemorrhagic papules on the foot in a patient undergoing treatment for multiple myeloma. Biopsy and culture demonstrated *Aspergillosis* sp. ***B.*** Eroded nodule on the hard palate of a patient undergoing chemotherapy. Biopsy and culture demonstrated *Mucor* sp. ***C.*** Ecthyma gangrenosum in a neutropenic patient with *Pseudomonas aeruginosa* bacteremia.

Figure e16-56

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Septic emboli with hemorrhage and infarction due to acute *Staphylococcus aureus* endocarditis. *(Courtesy of L. Baden, MD; with permission.*)

Figure e16-57

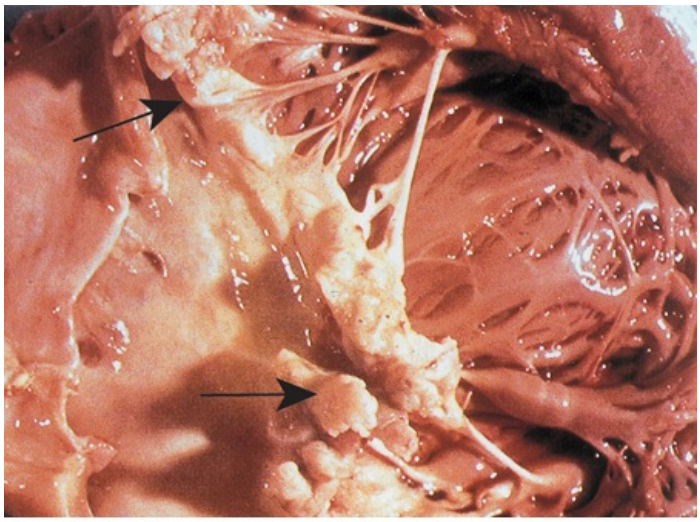

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vegetations*(arrows)* due to viridans streptococcal endocarditis involving the mitral valve *(Courtesy of AW Kerchner, MD; with permission)*

Figure e16-58

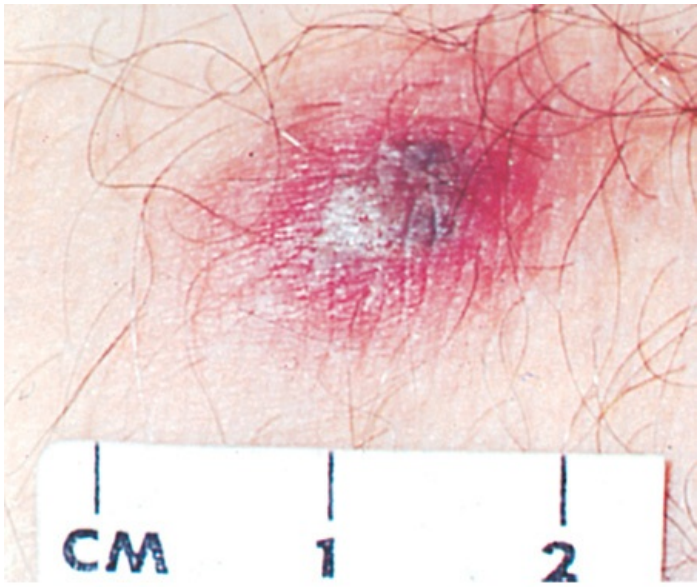

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Disseminated gonococcemia in the skin is seen as hemorrhagic papules and pustules with purpuric centers in an acral distribution. *(Courtesy of Daniel M. Musher, MD; with permission.)*

IMMUNOLOGICALLY MEDIATED SKIN DISEASE

(Figs. e16-59 to e16-70) Immunologically mediated skin disease may be largely localized to skin and mucous membranes and manifest with blisters and erosions such as pemphigus, pemphigoid, and dermatitis herpetiformis. In diseases such as systemic lupus erythematosus, dermatomyositis, and vasculitis, skin manifestations are often only one element of a widespread process.

Figure e16-59

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. **Systemic lupus erythematosus** showing prominent, scaly, malar erythema. Involvement of other sun-exposed sites is also common. ***B.*** Acute lupus erythematosus on the upper chest demonstrating brightly erythematous and slightly edematous coalescence of papules and plaques. *(**B,** Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-60

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Discoid lupus erythematosus. Atrophic, depigmented plaques and patches surrounded by hyperpigmentation and erythema associated with scarring and alopecia, are characteristic of this cutaneous form of lupus.

Figure e16-61

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dermatomyositis. Periorbital violaceous erythema characterizes the classic heliotrope rash. *(Courtesy of James Krell, MD; with permission.)*

Figure e16-62

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scleroderma characterized by typical expressionless, mask-like facies.

Figure e16-63

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dermatomyositis often involves the hands as erythematous flat-topped papules over the knuckles (Gottron's sign) and periungual telangiectasias.

Figure e16-64

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scleroderma showing acral sclerosis and focal digital ulcers.

Figure e16-65

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema multiforme is characterized by multiple erythematous plaques with a target or iris morphology and usually represents a hypersensitivity reaction to drugs or infections (especially herpes simplex virus). *(Courtesy of Yale Resident's Slide Collection; with permission.)*

Figure e16-66

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dermatitis herpetiformis manifested by pruritic, grouped vesicles in a typical location. The vesicles are often excoriated and may occur on knees, buttocks, elbows and posterior scalp.

Figure e16-67

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***A.*Pemphigus vulgaris** demonstrating eroded bullae on the back. ***B.*Pemphigus vulgaris** almost invariably involves the oral mucosa and may present with erosions involving the gingiva, buccal mucosa, palate, posterior pharynx, or the tongue. *(**B,** Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-68

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema nodosum is a panniculitis characterized by tender deep-seated nodules and plaques usually located on the lower extremities. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-69

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vasculitis. Palpable purpuric papules on the lower legs are seen in this patient with cutaneous small vessel vasculitis. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-70

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bullous pemphigoid with tense vesicles and bullae on an erythematous, urticarial base. *(Courtesy of Yale Resident's Slide Collection; with permission.)*

SKIN MANIFESTATIONS OF INTERNAL DISEASE

(Figs. e16-71 to e16-78) While many systemic diseases also have cutaneous manifestations, there are some well-recognized dermatologic markers of internal disease, and some are demonstrated in this section. Many of these dermatologic markers may precede, accompany, or follow diagnosis of systemic disease. Acanthosis nigricans is a prototypical dermatologic process often occurring in association with underlying systemic abnormalities, most commonly obesity and insulin resistance. It may also be associated with other endocrine disorders and several rare genetic syndromes. Malignant acanthosis nigricans may occur in association with several malignancies, especially adenocarcinoma of the gastrointestinal tract, lung, and breast. Other markers of internal disease in this section include pretibial myxedema, which is associated with thyroid disease, and Sweet's syndrome, which may be associated with hematologic malignancies, solid tumors, or inflammatory bowel disease. The skin is also involved in many systemic inflammatory diseases such as sarcoidosis, rheumatoid arthritis, and lupus erythematosus.

Figure e16-71

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acanthosis nigricans demonstrating typical hyperpigmented axillary plaques with a velvet-like, verrucous surface on the neck.

Figure e16-72

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pretibial myxedema manifesting as waxy, infiltrated plaques in a patient with Graves' disease.

Figure e16-73

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Plaque of Sweet's syndrome demonstrating an erythematous indurated plaque with a pseudo-vesicular border. *(Courtesy of Robert Swerlick, MD, with permission.)*

Figure e16-74

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bilateral rheumatoid nodules of the upper extremities. *(Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-75

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurofibromatosis demonstrating numerous flesh-colored cutaneous neurofibromas.

Figure e16-76

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coumarin necrosis showing cutaneous and subcutaneous necrosis of a breast. Other fatty areas such as buttocks and thighs are also common sites of involvement. *(Courtesy of Kim Yancey, MD; with permission.)*

Figure e16-77

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***A.*Sarcoid.** Infiltrated papules and plaques of variable color are seen in a typical paranasal and periorbital location. ***B.*Sarcoid.** Infiltrated, hyperpigmented, and slightly erythematous coalescent papules and plaques on the upper arm. *(**B,** Courtesy of Robert Swerlick, MD; with permission.)*

Figure e16-78

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pyoderma gangrenosum on the dorsal aspect of both hands demonstrating multiple necrotic ulcers surrounded by a violaceous and undermined border. *(Courtesy of Robert Swerlick, MD; with permission.)*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEMATOPOIESIS AND THE PHYSIOLOGIC BASIS OF RED CELL PRODUCTION

Hematopoiesis is the process by which the formed elements of blood are produced. The process is regulated through a series of steps beginning with the hematopoietic stem cell. Stem cells are capable of producing red cells, all classes of granulocytes, monocytes, platelets, and the cells of the immune system. The precise molecular mechanism—either intrinsic to the stem cell itself or through the action of extrinsic factors—by which the stem cell becomes committed to a given lineage is not fully defined. However, experiments in mice suggest that erythroid cells come from a common erythroid/megakaryocyte progenitor that does not develop in the absence of expression of the GATA-1 and FOG-1 (friend of GATA-1) transcription factors (Chap. 66). Following lineage commitment, hematopoietic progenitor and precursor cells come increasingly under the regulatory influence of growth factors and hormones. For red cell production, erythropoietin (EPO) is the regulatory hormone. EPO is required for the maintenance of committed erythroid progenitor cells that, in the absence of the hormone, undergo programmed cell death (*apoptosis*). The regulated process of red cell production is *erythropoiesis*, and its key elements are illustrated in Fig. 57-1.

Figure 57-1

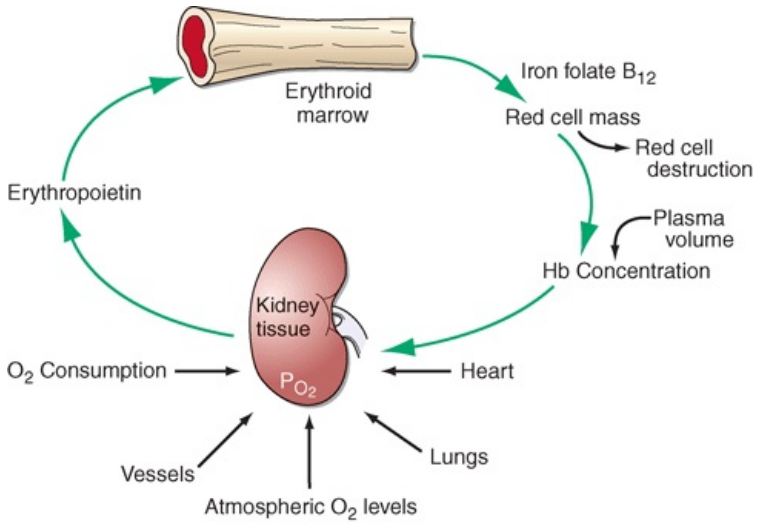

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The physiologic regulation of red cell production by tissue oxygen tension. Hb, hemoglobin.

In the bone marrow, the first morphologically recognizable erythroid precursor is the pronormoblast. This cell can undergo four to five cell divisions, which result in the production of 16–32 mature red cells. With increased EPO production, or the administration of EPO as a drug, early progenitor cell numbers are amplified and, in turn, give rise to increased numbers of erythrocytes. The regulation of EPO production itself is linked to tissue oxygenation.

In mammals, O_2 is transported to tissues bound to the hemoglobin contained within circulating red cells. The mature red cell is 8 μm in diameter, anucleate, discoid in shape, and extremely pliable in order to traverse the microcirculation successfully; its membrane integrity is maintained by the intracellular generation of ATP. Normal red cell production results in the daily replacement of 0.8–1% of all circulating red cells in the body, since the average red cell lives 100–120 days. The organ responsible for red cell production is called the *erythron*. The erythron is a dynamic organ made up of a rapidly proliferating pool of marrow erythroid precursor cells and a large mass of mature circulating red blood cells. The size of the red cell mass reflects the balance of red cell production and destruction. The physiologic basis of red cell production and destruction provides an understanding of the mechanisms that can lead to anemia.

The physiologic regulator of red cell production, the glycoprotein hormone EPO, is produced and released by peritubular capillary lining cells within the kidney. These cells are highly specialized epithelial-like cells. A small amount of EPO is produced by hepatocytes. The fundamental stimulus for EPO production is the availability of O_2 for tissue metabolic needs. Key to EPO gene regulation is hypoxia-inducible factor (HIF)-1α. In the presence of O_2, HIF-1α is hydroxylated at a key proline, allowing HIF-1α to be ubiquitinylated and degraded via the proteasome pathway. If O_2 becomes limiting, this critical hydroxylation step does not occur, allowing HIF-1α to partner with other proteins, translocate to the nucleus, and upregulate the EPO gene, among others.

Impaired O_2 delivery to the kidney can result from a decreased red cell mass (*anemia*), impaired O_2 loading of the hemoglobin molecule or a high O_2 affinity mutant hemoglobin (*hypoxemia*), or, rarely, impaired blood flow to the kidney (renal artery stenosis). EPO governs the day-to-day production of red cells, and ambient levels of the hormone can be measured in the plasma by sensitive immunoassays—the normal level being 10–25 U/L. When the hemoglobin concentration falls below 100–120 g/L (10–12 g/dL), plasma EPO levels increase in proportion to the severity of the anemia (Fig. 57-2). In circulation, EPO has a half-clearance time of 6–9 h. EPO acts by binding to specific receptors on the surface of marrow erythroid precursors, inducing them to proliferate and to mature. With EPO stimulation, red cell production can increase four- to fivefold within a 1- to 2-week period, but only in the presence of adequate nutrients, especially iron. The functional capacity of the erythron, therefore, requires normal renal production of EPO, a functioning erythroid marrow, and an adequate supply of substrates for hemoglobin synthesis. A defect in any of these key components can lead to anemia. Generally, anemia is recognized in the laboratory when a patient's hemoglobin level or hematocrit is reduced below an expected value (the normal range). The likelihood and severity of anemia are defined based on the deviation of the patient's hemoglobin/hematocrit from values expected for age- and sex-matched normal subjects. The hemoglobin concentration in adults has a Gaussian distribution. The mean hematocrit value for adult males is 47% (± SD 7) and that for adult females is 42% (± 5). Any single hematocrit or hemoglobin value carries with it a likelihood of associated anemia. Thus, a hematocrit of ≤39% in an adult male or <35% in an adult female has only about a 25% chance of being normal. Suspected low hemoglobin or hematocrit values are more easily interpreted if previous values for the same patient are known for comparison. The World Health Organization (WHO) defines anemia as a hemoglobin level <130 g/L (13 g/dL) in men and <120 g/L (12 g/dL) in women.

Figure 57-2

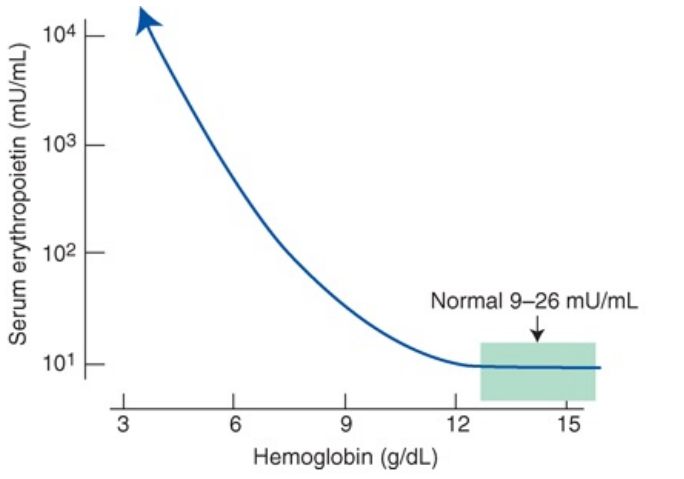

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythropoietin (EPO) levels in response to anemia. When the hemoglobin level falls to 120 g/L (12 g/dL), plasma EPO levels increase logarithmically. In the presence of chronic kidney disease or chronic inflammation, EPO levels are typically lower than expected for the degree of anemia. As individuals age, the level of EPO needed to sustain normal hemoglobin levels appears to increase. *(From Hillman et al.)*

The critical elements of erythropoiesis—EPO production, iron availability, the proliferative capacity of the bone marrow, and effective maturation of red cell precursors—are used for the initial classification of anemia (see below).

ANEMIA

Clinical Presentation of Anemia

SIGNS AND SYMPTOMS

Anemia is most often recognized by abnormal screening laboratory tests. Patients less commonly present with advanced anemia and its attendant signs and symptoms. Acute anemia is due to blood loss or hemolysis. If blood loss is mild, enhanced O_2 delivery is achieved through changes in the O_2–hemoglobin dissociation curve mediated by a decreased pH or increased CO_2 (*Bohr effect*). With acute blood loss, hypovolemia dominates the clinical picture and the hematocrit and hemoglobin levels do not reflect the volume of blood lost. Signs of vascular instability appear with acute losses of 10–15% of the total blood volume. In such patients, the issue is not anemia but hypotension and decreased organ perfusion. When >30% of the blood volume is lost suddenly, patients are unable to compensate with the usual mechanisms of vascular contraction and changes in regional blood flow. The patient prefers to remain supine and will show postural hypotension and tachycardia. If the volume of blood lost is >40% (i.e., >2 L in the average-sized adult), signs of hypovolemic shock including confusion, dyspnea, diaphoresis, hypotension, and tachycardia appear (Chap. 106). Such patients have significant deficits in vital organ perfusion and require immediate volume replacement.

With acute hemolysis, the signs and symptoms depend on the mechanism that leads to red cell destruction. Intravascular hemolysis with release of free hemoglobin may be associated with acute back pain, free hemoglobin in the plasma and urine, and renal failure. Symptoms associated with more chronic or progressive anemia depend on the age of the patient and the adequacy of blood supply to critical organs. Symptoms associated with moderate anemia include fatigue, loss of stamina, breathlessness, and tachycardia (particularly with physical exertion). However, because of the intrinsic compensatory mechanisms that govern the O_2–hemoglobin dissociation curve, the gradual onset of anemia—particularly in young patients—may not be associated with signs or symptoms until the anemia is severe [hemoglobin <70–80 g/L (7–8 g/dL)]. When anemia develops over a period of days or weeks, the total blood volume is normal to slightly increased, and changes in cardiac output and regional blood flow help compensate for the overall loss in O_2-carrying capacity. Changes in the position of the O_2–hemoglobin dissociation curve account for some of the compensatory response to anemia. With chronic anemia, intracellular levels of 2,3-bisphosphoglycerate rise, shifting the dissociation curve to the right and facilitating O_2 unloading. This compensatory mechanism can only maintain normal tissue O_2 delivery in the face of a 20–30 g/L (2–3 g/dL) deficit in hemoglobin concentration. Finally, further protection of O_2 delivery to vital organs is achieved by the shunting of blood away from organs that are relatively rich in blood supply, particularly the kidney, gut, and skin.

Certain disorders are commonly associated with anemia. Chronic inflammatory states (e.g., infection, rheumatoid arthritis, cancer) are associated with mild to moderate anemia, whereas lymphoproliferative disorders, such as chronic lymphocytic leukemia and certain other B cell neoplasms, may be associated with autoimmune hemolysis.

Approach to the Patient: Anemia

The evaluation of the patient with anemia requires a careful history and physical examination. Nutritional history related to drugs or alcohol intake and family history of anemia should always be assessed. Certain geographic backgrounds and ethnic origins are associated with an increased likelihood of an inherited disorder of the hemoglobin molecule or intermediary metabolism. Glucose-6-phosphate dehydrogenase (G6PD) deficiency and certain hemoglobinopathies are seen more commonly in those of Middle Eastern or African origin, including African Americans who have a high frequency of G6PD deficiency. Other information that may be useful includes exposure to certain toxic agents or drugs and symptoms related to other disorders commonly associated with anemia. These include symptoms and signs such as bleeding, fatigue, malaise, fever, weight loss, night sweats, and other systemic symptoms. Clues to the mechanisms of anemia may be provided on physical examination by findings of infection, blood in the stool, lymphadenopathy, splenomegaly, or petechiae. Splenomegaly and lymphadenopathy suggest an underlying lymphoproliferative disease, while petechiae suggest platelet dysfunction. Past laboratory measurements are helpful to determine a time of onset.

In the anemic patient, physical examination may demonstrate a forceful heartbeat, strong peripheral pulses, and a systolic "flow" murmur. The skin and mucous membranes may be pale if the hemoglobin is <80–100 g/L (8–10 g/dL). This part of the physical examination should focus on areas where vessels are close to the surface such as the mucous membranes, nail beds, and palmar creases. If the palmar creases are lighter in color than the surrounding skin when the hand is hyperextended, the hemoglobin level is usually <80 g/L (8 g/dL).

LABORATORY EVALUATION

Table 57–1 lists the tests used in the initial workup of anemia. A routine complete blood count (CBC) is required as part of the evaluation and includes the hemoglobin, hematocrit, and red cell indices: the mean cell volume (MCV) in femtoliters, mean cell hemoglobin (MCH) in picograms per cell, and mean concentration of hemoglobin per volume of red cells (MCHC) in grams per liter (non-SI: grams per deciliter). The red cell indices are calculated as shown in Table 57–2, and the normal variations in the

hemoglobin and hematocrit with age are shown in Table 57–3. A number of physiologic factors affect the CBC, including age, sex, pregnancy, smoking, and altitude. High-normal hemoglobin values may be seen in men and women who live at altitude or smoke heavily. Hemoglobin elevations due to smoking reflect normal compensation due to the displacement of O_2 by CO in hemoglobin binding. Other important information is provided by the reticulocyte count and measurements of iron supply including *serum iron*, *total iron-binding capacity* (TIBC; an indirect measure of the transferrin level), and *serum ferritin*. Marked alterations in the red cell indices usually reflect disorders of maturation or iron deficiency. A careful evaluation of the peripheral blood smear is important, and clinical laboratories often provide a description of both the red and white cells, a white cell differential count, and the platelet count. In patients with severe anemia and abnormalities in red blood cell morphology and/or low reticulocyte counts, a bone marrow aspirate or biopsy can assist in the diagnosis. Other tests of value in the diagnosis of specific anemias are discussed in chapters on specific disease states.

Table 57–1 Laboratory Tests in Anemia Diagnosis

I. Complete blood count (CBC)
A. Red blood cell count
1. Hemoglobin
2. Hematocrit
3. Reticulocyte count
B. Red blood cell indices
1. Mean cell volume (MCV)
2. Mean cell hemoglobin (MCH)
3. Mean cell hemoglobin concentration (MCHC)
4. Red cell distribution width (RDW)
C. White blood cell count
1. Cell differential
2. Nuclear segmentation of neutrophils
D. Platelet count
E. Cell morphology
1. Cell size
2. Hemoglobin content
3. Anisocytosis
4. Poikilocytosis
5. Polychromasia
II. Iron supply studies
A. Serum iron
B. Total iron-binding capacity
C. Serum ferritin
III. Marrow examination
A. Aspirate
1. M/E ratio[a]
2. Cell morphology
3. Iron stain
B. Biopsy
1. Cellularity
2. Morphology

[a] M/E ratio, ratio of myeloid to erythroid precursors.

Table 57–2 Red Blood Cell Indices

Index	Normal Value
Mean cell volume (MCV) = (hematocrit x 10)/(red cell count x 10^6)	90 ± 8 fL

Mean cell hemoglobin (MCH) = (hemoglobin x 10)/(red cell count x 10^6)	30 ± 3 pg
Mean cell hemoglobin concentration = (hemoglobin x 10)/hematocrit, or MCH/MCV	33 ± 2%

Table 57–3 Changes in Normal Hemoglobin/Hematocrit Values with Age and Pregnancy

Age/Sex	Hemoglobin g/dL	Hematocrit %
At birth	17	52
Childhood	12	36
Adolescence	13	40
Adult man	16 (±2)	47 (±6)
Adult woman (menstruating)	13 (±2)	40 (±6)
Adult woman (postmenopausal)	14 (±2)	42 (±6)
During pregnancy	12 (±2)	37 (±6)

Source: From Hillman et al.

The components of the CBC also help in the classification of anemia. *Microcytosis* is reflected by a lower than normal MCV (<80), whereas high values (>100) reflect *macrocytosis*. The MCH and MCHC reflect defects in hemoglobin synthesis (*hypochromia*). Automated cell counters describe the red cell volume distribution width (RDW). The MCV (representing the peak of the distribution curve) is insensitive to the appearance of small populations of macrocytes or microcytes. An experienced laboratory technician will be able to identify minor populations of large or small cells or hypochromic cells before the red cell indices change.

PERIPHERAL BLOOD SMEAR

The peripheral blood smear provides important information about defects in red cell production (Chap. e17). As a complement to the red cell indices, the blood smear also reveals variations in cell size (*anisocytosis*) and shape (*poikilocytosis*). The degree of anisocytosis usually correlates with increases in the RDW or the range of cell sizes. Poikilocytosis suggests a defect in the maturation of red cell precursors in the bone marrow or fragmentation of circulating red cells. The blood smear may also reveal *polychromasia*—red cells that are slightly larger than normal and grayish blue in color on the Wright-Giemsa stain. These cells are reticulocytes that have been prematurely released from the bone marrow, and their color represents residual amounts of ribosomal RNA. These cells appear in circulation in response to EPO stimulation or to architectural damage of the bone marrow (fibrosis, infiltration of the marrow by malignant cells, etc.) that results in their disordered release from the marrow. The appearance of nucleated red cells, Howell-Jolly bodies, target cells, sickle cells, and others may provide clues to specific disorders (Figs. 57-3 to 57-11).

Figure 57-3

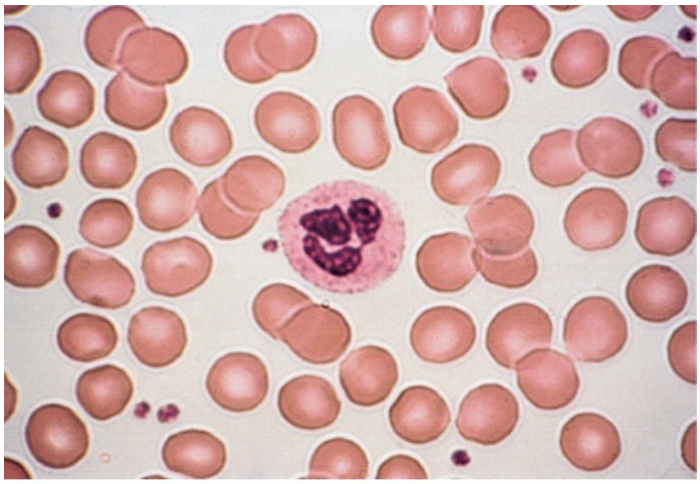

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal blood smear (Wright stain). High-power field showing normal red cells, a neutrophil, and a few platelets. *(From Hillman et al.)*

Figure 57-4

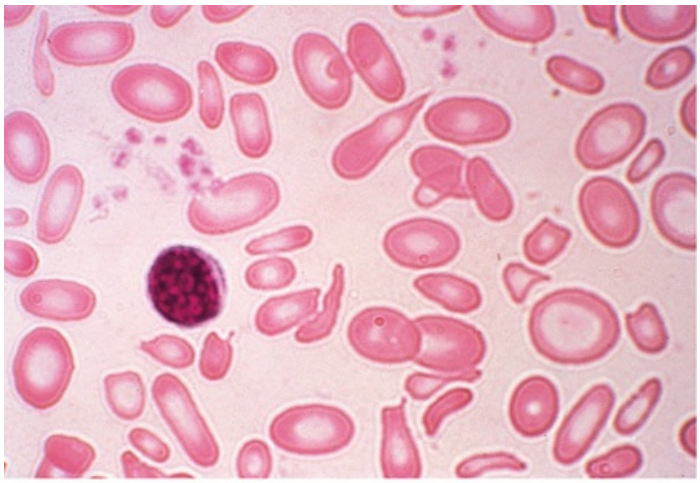

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe iron-deficiency anemia. Microcytic and hypochromic red cells smaller than the nucleus of a lymphocyte associated with marked variation in size (anisocytosis) and shape (poikilocytosis). *(From Hillman et al.)*

Figure 57-5

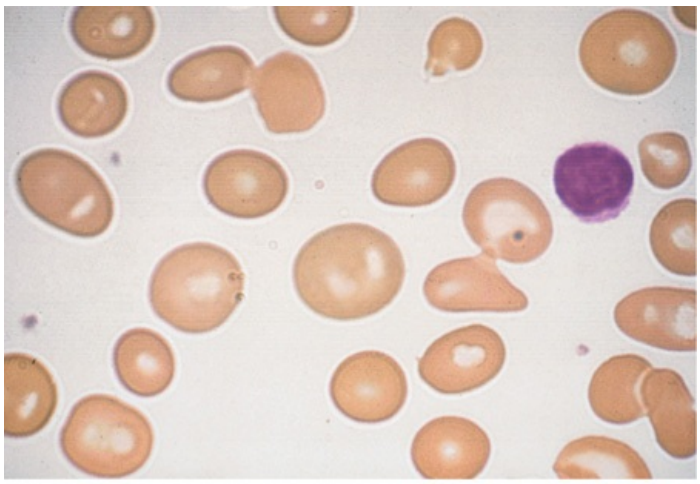

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Macrocytosis. Red cells are larger than a small lymphocyte and well hemoglobinized. Often macrocytes are oval shaped (macro-ovalocytes).

Figure 57-6

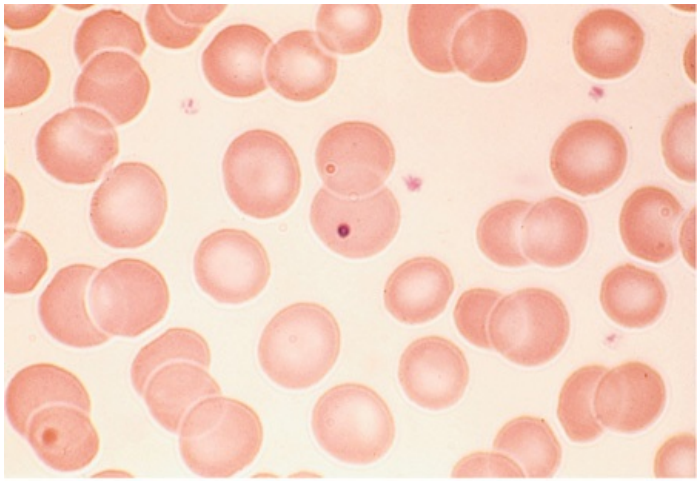

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Howell-Jolly bodies. In the absence of a functional spleen, nuclear remnants are not culled from the red cells and remain as small homogeneously staining blue inclusions on Wright stain. *(From Hillman et al.)*

Figure 57-7

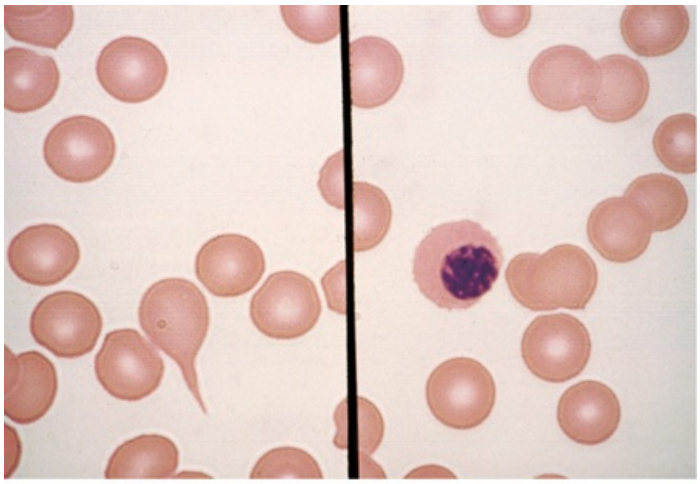

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Red cell changes in myelofibrosis. The left panel shows a teardrop-shaped cell. The right panel shows a nucleated red cell. These forms are seen in myelofibrosis.

Figure 57-8

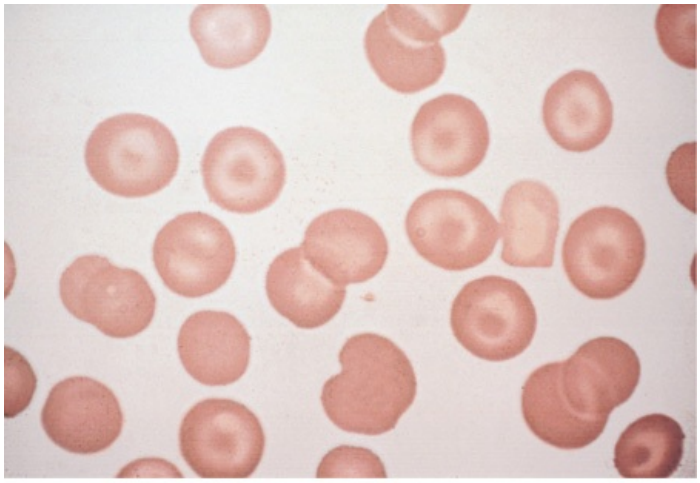

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Target cells. Target cells have a bull's-eye appearance and are seen in thalassemia and in liver disease. *(From Hillman et al.)*

Figure 57-9

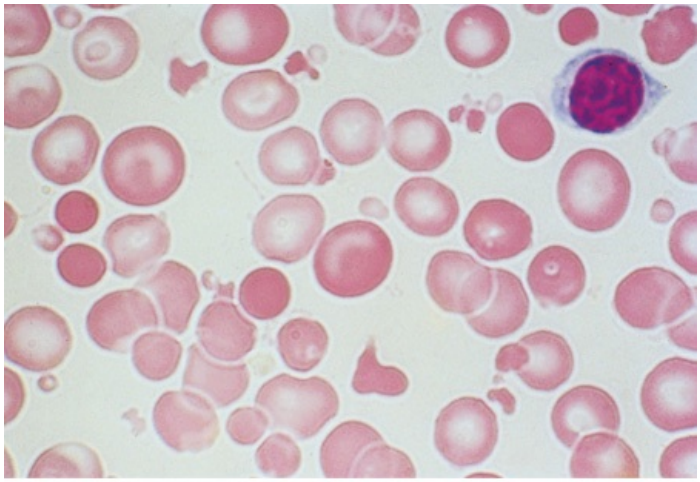

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Red cell fragmentation. Red cells may become fragmented in the presence of foreign bodies in the circulation, such as mechanical heart valves, or in the setting of thermal injury. *(From Hillman et al.)*

Figure 57-10

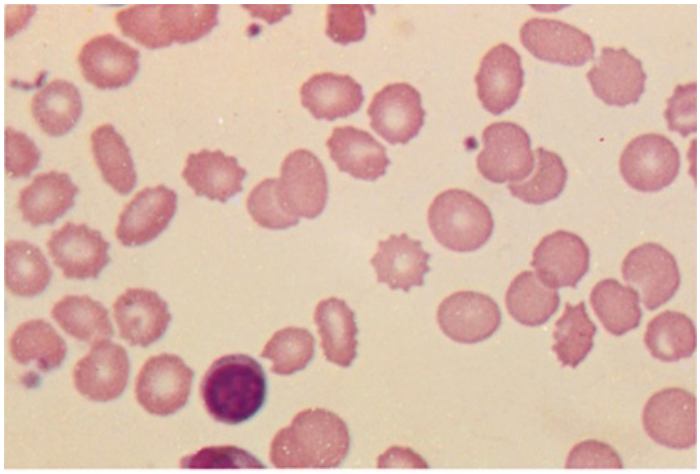

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Uremia. The red cells in uremia may acquire numerous regularly spaced, small, spiny projections. Such cells, called burr cells or echinocytes, are readily distinguishable from irregularly spiculated acanthocytes shown in **Fig. 57-11**.

Figure 57-11

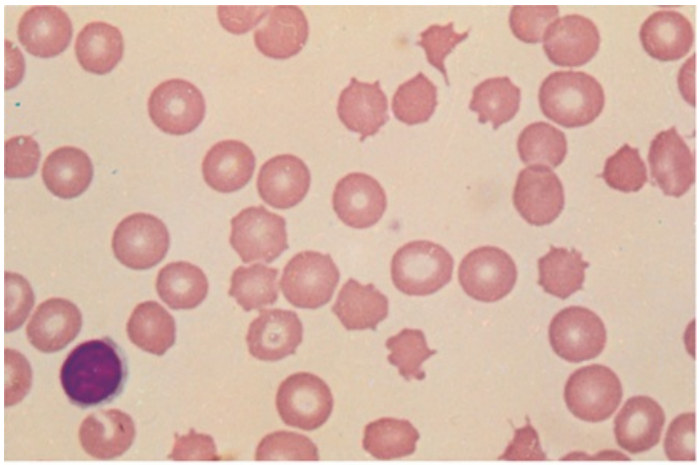

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spur cells. Spur cells are recognized as distorted red cells containing several irregularly distributed thornlike projections. Cells with this morphologic abnormality are also called acanthocytes. *(From Hillman et al.)*

RETICULOCYTE COUNT

An accurate reticulocyte count is key to the initial classification of anemia. Normally, reticulocytes are red cells that have been recently released from the bone marrow. They are identified by staining with a supravital dye that precipitates the ribosomal RNA (Fig. 57-12). These precipitates appear as blue or black punctate spots. This residual RNA is metabolized over the first 24–36 h of the reticulocyte's life span in circulation. Normally, the reticulocyte count ranges from 1 to 2% and reflects the daily replacement of 0.8–1.0% of the circulating red cell population. A corrected reticulocyte count provides a reliable measure of red cell production.

Figure 57-12

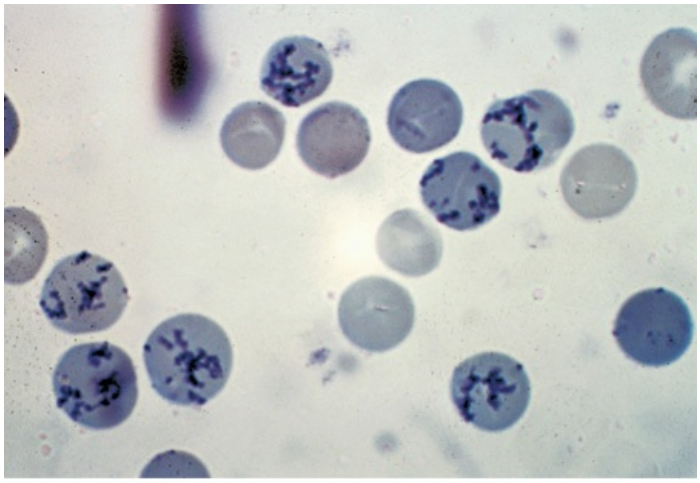

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Reticulocytes. Methylene blue stain demonstrates residual RNA in newly made red cells. *(From Hillman et al.)*

In the initial classification of anemia, the patient's reticulocyte count is compared with the expected reticulocyte response. In general, if the EPO and erythroid marrow responses to moderate anemia [hemoglobin < 100 g/L (10 g/dL)] are intact, the red cell production rate increases to two to three times normal within 10 days following the onset of anemia. In the face of established anemia, a reticulocyte response less than two to three times normal indicates an inadequate marrow response.

In order to use the reticulocyte count to estimate marrow response, two corrections are necessary. The first correction adjusts the reticulocyte count based on the reduced number of circulating red cells. With anemia, the percentage of reticulocytes may be increased while the absolute number is unchanged. To correct for this effect, the reticulocyte percentage is multiplied by the ratio of the patient's hemoglobin or hematocrit to the expected hemoglobin/hematocrit for the age and gender of the patient (Table 57–4). This provides an estimate of the reticulocyte count corrected for anemia. In order to convert the corrected reticulocyte count to an index of marrow production, a further correction is required, depending on whether some of the reticulocytes in circulation have been released from the marrow prematurely. For this second correction, the peripheral blood smear is examined to see if there are polychromatophilic macrocytes present.

Table 57–4 Calculation of Reticulocyte Production Index

Correction #1 for Anemia:
This correction produces the corrected reticulocyte count In a person whose reticulocyte count is 9%, hemoglobin 7.5 g/dL, hematocrit 23%, the absolute reticulocyte count = 9 x (7.5/15) [or x (23/45)]= 4.5%
Correction #2 for Longer Life of Prematurely Released Reticulocytes in the Blood:
This correction produces the reticulocyte production index In a person whose reticulocyte count is 9%, hemoglobin 7.5 gm/dL, hematocrit 23%, the reticulocyte production index $=9\times\frac{(7.5/15)(\text{hemoglobin correction})}{2(\text{maturation time correction})}=2.25$

These cells, representing prematurely released reticulocytes, are referred to as "shift" cells, and the relationship between the degree of shift and the necessary shift correction factor is shown in Fig. 57-13. The correction is necessary because these prematurely released cells survive as reticulocytes in circulation for >1 day, thereby providing a falsely high estimate of daily red cell production. If polychromasia is increased, the reticulocyte count, already corrected for anemia, should be divided again by 2 to account for the prolonged reticulocyte maturation time. The second correction factor varies from 1 to 3 depending on the severity of anemia. In general, a correction of 2 is commonly used. An appropriate correction is shown in Table 57–4. If polychromatophilic cells are not seen on the blood smear, the second correction is not required. The now doubly corrected reticulocyte count is the *reticulocyte production index*, and it provides an estimate of marrow production relative to normal.

Figure 57-13

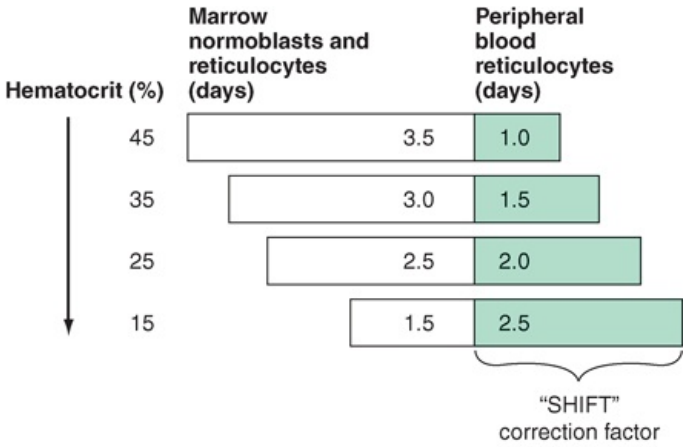

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Correction of the reticulocyte count. In order to use the reticulocyte count as an indicator of effective red cell production, the reticulocyte percentage must be corrected based on the level of anemia and the circulating life span of the reticulocytes. Erythroid cells take ~4.5 days to mature. At a normal hemoglobin, reticulocytes are released to the circulation with ~1 day left as reticulocytes. However, with different levels of anemia, reticulocytes (and even earlier erythroid cells) may be released from the marrow prematurely. Most patients come to clinical attention with hematocrits in the mid-20s, and thus a correction factor of 2 is commonly used because the observed reticulocytes will live for 2 days in the circulation before losing their RNA.

Premature release of reticulocytes is normally due to increased EPO stimulation. However, if the integrity of the bone marrow release process is lost through tumor infiltration, fibrosis, or other disorders, the appearance of nucleated red cells or polychromatophilic macrocytes should still invoke the second reticulocyte correction. The shift correction should always be applied to a patient with anemia and a very high reticulocyte count to provide a true index of effective red cell production. Patients with severe chronic hemolytic anemia may increase red cell production as much as six- to sevenfold. This measure alone, therefore, confirms the fact that the patient has an appropriate EPO response, a normally functioning bone marrow, and sufficient iron available to meet the demands for new red cell formation. Table 57–5 demonstrates the normal marrow response to anemia. If the reticulocyte production index is <2 in the face of established anemia, a defect in erythroid marrow proliferation or maturation must be present.

Table 57–5 Normal Marrow Response to Anemia

Hematocrit	Production Index	Reticulocytes (incl corrections)	Marrow M/E Ratio
45	1	1	3:1
35	2.0–3.0	4.8%/3.8/2.5	2:1–1:1
25	3.0–5.0	14%/8/4.0	1:1–1:2
15	3.0–5.0	30%/10/4.0	1:1–1:2

TESTS OF IRON SUPPLY AND STORAGE

The laboratory measurements that reflect the availability of iron for hemoglobin synthesis include the serum iron, the TIBC, and the percent transferrin saturation. The percent transferrin saturation is derived by dividing the serum iron level (x 100) by the TIBC. The normal serum iron ranges from 9 to 27 μmol/L (50–150 μg/dL), while the normal TIBC is 54–64 μmol/L (300–360 μg/dL); the normal transferrin saturation ranges from 25 to 50%. A diurnal variation in the serum iron leads to a variation in the percent transferrin saturation. The serum ferritin is used to evaluate total body iron stores. Adult males have serum ferritin levels that average ~100 μg/L, corresponding to iron stores of ~1 g. Adult females have lower serum ferritin levels averaging 30 μg/L, reflecting lower iron stores (~300 mg). A serum ferritin level of 10–15 μg/L represents depletion of body iron stores. However, ferritin is also an acute-phase reactant and, in the presence of acute or chronic inflammation, may rise several-fold above baseline levels. As a rule, a serum ferritin >200 μg/L means there is at least some iron in tissue stores.

BONE MARROW EXAMINATION

A bone marrow aspirate and smear or a needle biopsy can be useful in the evaluation of some patients with anemia. In patients with hypoproliferative anemia and normal iron status, a bone marrow is indicated. Marrow examination can diagnose primary marrow disorders such as myelofibrosis, a red cell maturation defect, or an infiltrative disease (Figs. 57-14 to 57-16). The increase or decrease of one cell lineage (myeloid vs. erythroid) compared to another is obtained by a differential count of

nucleated cells in a bone marrow smear [the myeloid/erythroid (M/E) ratio]. A patient with a hypoproliferative anemia (see below) and a reticulocyte production index <2 will demonstrate an M/E ratio of 2 or 3:1. In contrast, patients with hemolytic disease and a production index >3 will have an M/E ratio of at least 1:1. Maturation disorders are identified from the discrepancy between the M/E ratio and the reticulocyte production index (see below). Either the marrow smear or biopsy can be stained for the presence of iron stores or iron in developing red cells. The storage iron is in the form of ferritin or *hemosiderin*. On carefully prepared bone marrow smears, small ferritin granules can normally be seen under oil immersion in 20–40% of developing erythroblasts. Such cells are called *sideroblasts*.

Figure 57-14

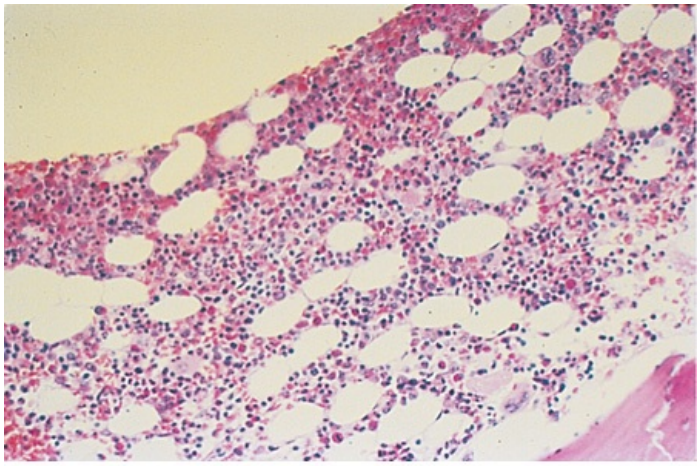

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal bone marrow. This is a low-power view of a section of a normal bone marrow biopsy stained with hematoxylin and eosin (H&E). Note that the nucleated cellular elements account for ~40–50% and the fat (clear areas) accounts for ~50–60% of the area. *(From Hillman et al.)*

Figure 57-15

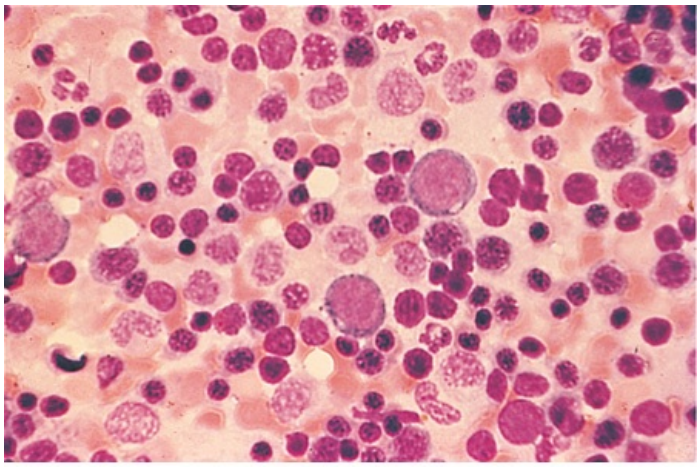

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythroid hyperplasia. This marrow shows an increase in the fraction of cells in the erythroid lineage as might be seen when a normal marrow compensates for acute blood loss or hemolysis. The M/E ratio is about 1:1. M/E, myeloid/erythroid. *(From Hillman et al.)*

Figure 57-16

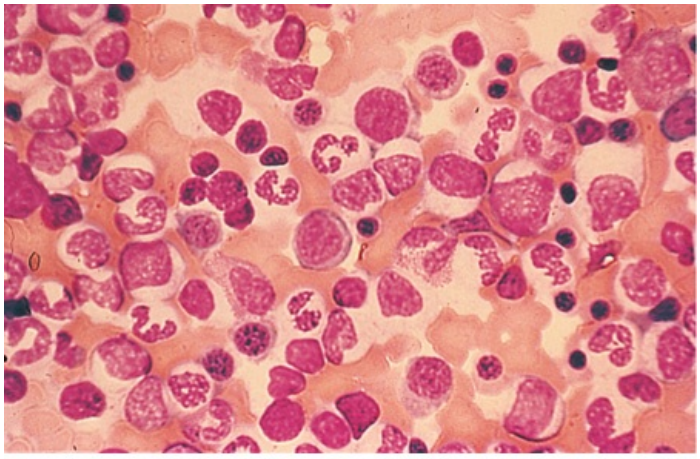

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Myeloid hyperplasia. This marrow shows an increase in the fraction of cells in the myeloid or granulocytic lineage as might be seen in a normal marrow responding to infection. The M/E ratio is >3:1. M/E, myeloid/erythroid. *(From Hillman et al.)*

OTHER LABORATORY MEASUREMENTS

Additional laboratory tests may be of value in confirming specific diagnoses. **For details of these tests and how they are applied in individual disorders, see Chaps. 103, 104, 105, 106, and 107.**

Definition and Classification of Anemia

INITIAL CLASSIFICATION OF ANEMIA

The functional classification of anemia has three major categories. These are (1) marrow production defects (*hypoproliferation*), (2) red cell maturation defects (*ineffective erythropoiesis*), and (3) decreased red cell survival (*blood loss/hemolysis*). The classification is shown in Fig. 57-17. A hypoproliferative anemia is typically seen with a low reticulocyte production index together with little or no change in red cell morphology (a normocytic, normochromic anemia) (Chap. 103). Maturation disorders typically have a slight to moderately elevated reticulocyte production index that is accompanied by either macrocytic (Chap. 105) or microcytic (Chaps. 103, 104) red cell indices. Increased red blood cell destruction secondary to hemolysis results in an increase in the reticulocyte production index to at least three times normal (Chap. 106), provided sufficient iron is available. Hemorrhagic anemia does not typically result in production indices of more than 2.0–2.5 times normal because of the limitations placed on expansion of the erythroid marrow by iron availability.

Figure 57-17

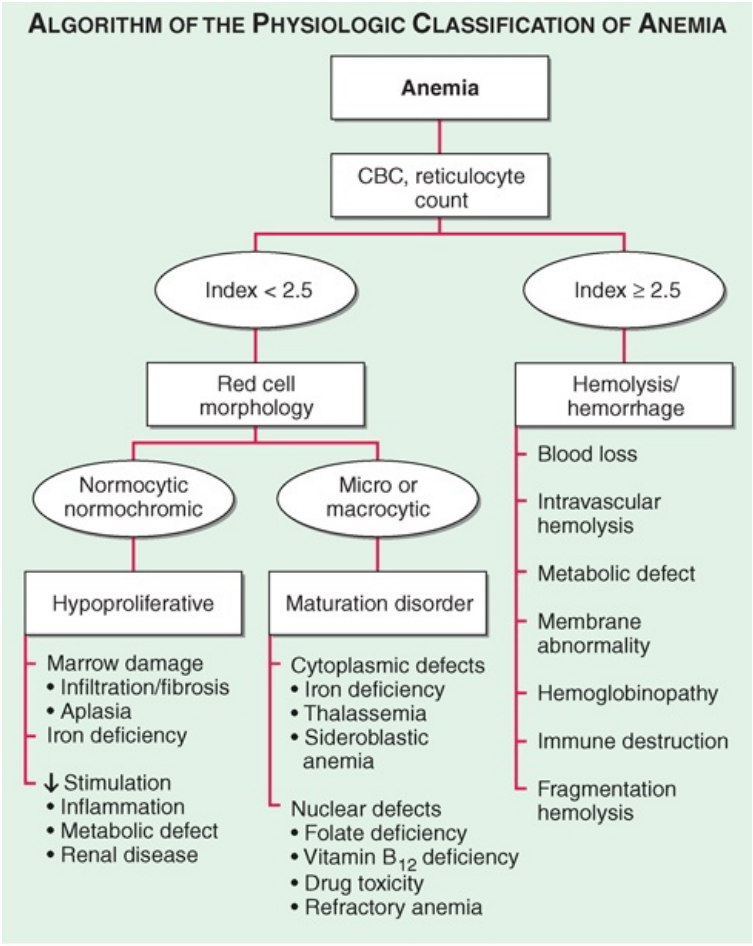

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The physiologic classification of anemia. CBC, complete blood count.

In the first branch point of the classification of anemia, a reticulocyte production index >2.5 indicates that hemolysis is most likely. A reticulocyte production index <2 indicates either a hypoproliferative anemia or maturation disorder. The latter two possibilities can often be distinguished by the red cell indices, by examination of the peripheral blood smear, or by a marrow examination. If the red cell indices are normal, the anemia is almost certainly hypoproliferative in nature. Maturation disorders are characterized by ineffective red cell production and a low reticulocyte production index. Bizarre red cell shapes—macrocytes or hypochromic microcytes—are seen on the peripheral blood smear. With a hypoproliferative anemia, no erythroid hyperplasia is noted in the marrow, whereas patients with ineffective red cell production have erythroid hyperplasia and an M/E ratio <1:1.

HYPOPROLIFERATIVE ANEMIAS

At least 75% of all cases of anemia are hypoproliferative in nature. A hypoproliferative anemia reflects absolute or relative marrow failure in which the erythroid marrow has not proliferated appropriately for the degree of anemia. The majority of hypoproliferative anemias are due to mild to moderate iron deficiency or inflammation. A hypoproliferative anemia can result from marrow damage, iron deficiency, or inadequate EPO stimulation. The last may reflect impaired renal function, suppression of EPO production by inflammatory cytokines such as interleukin 1, or reduced tissue needs for O_2 from metabolic disease such as hypothyroidism. Only occasionally is the marrow unable to produce red cells at a normal rate, and this is most prevalent in patients with renal failure. With diabetes mellitus or myeloma, the EPO deficiency may be more marked than would be predicted by the degree of renal insufficiency. In general, hypoproliferative anemias are characterized by normocytic, normochromic red cells, although microcytic, hypochromic cells may be observed with mild iron deficiency or long-standing chronic inflammatory disease. The key laboratory tests in distinguishing between the various forms of hypoproliferative anemia include the serum iron and iron-binding capacity, evaluation of renal and thyroid function, a marrow biopsy or aspirate to detect marrow damage or infiltrative disease, and serum ferritin to assess iron stores. An iron stain of the marrow will determine the pattern of iron distribution. Patients with the anemia of acute or chronic inflammation show a distinctive pattern of serum iron (low), TIBC (normal or low), percent transferrin saturation (low), and serum ferritin (normal or high). These changes in iron values are brought about by hepcidin, the iron regulatory hormone that is increased in inflammation (Chap. 103). A distinct pattern of results is noted in mild to moderate iron

deficiency (low serum iron, high TIBC, low percent transferrin saturation, low serum ferritin) (Chap. 103). Marrow damage by drugs, infiltrative disease such as leukemia or lymphoma, or marrow aplasia are diagnosed from the peripheral blood and bone marrow morphology. With infiltrative disease or fibrosis, a marrow biopsy is required.

MATURATION DISORDERS

The presence of anemia with an inappropriately low reticulocyte production index, macro- or microcytosis on smear, and abnormal red cell indices suggests a maturation disorder. Maturation disorders are divided into two categories: nuclear maturation defects, associated with macrocytosis, and cytoplasmic maturation defects, associated with microcytosis and hypochromia usually from defects in hemoglobin synthesis. The inappropriately low reticulocyte production index is a reflection of the ineffective erythropoiesis that results from the destruction within the marrow of developing erythroblasts. Bone marrow examination shows erythroid hyperplasia.

Nuclear maturation defects result from vitamin B12 or folic acid deficiency, drug damage, or myelodysplasia. Drugs that interfere with cellular DNA synthesis, such as methotrexate or alkylating agents, can produce a nuclear maturation defect. Alcohol, alone, is also capable of producing macrocytosis and a variable degree of anemia, but this is usually associated with folic acid deficiency. Measurements of folic acid and vitamin B_{12} are critical not only in identifying the specific vitamin deficiency but also because they reflect different pathogenetic mechanisms (Chap. 105).

Cytoplasmic maturation defects result from severe iron deficiency or abnormalities in globin or heme synthesis. Iron deficiency occupies an unusual position in the classification of anemia. If the iron-deficiency anemia is mild to moderate, erythroid marrow proliferation is blunted and the anemia is classified as hypoproliferative. However, if the anemia is severe and prolonged, the erythroid marrow will become hyperplastic despite the inadequate iron supply, and the anemia will be classified as ineffective erythropoiesis with a cytoplasmic maturation defect. In either case, an inappropriately low reticulocyte production index, microcytosis, and a classic pattern of iron values make the diagnosis clear and easily distinguish iron deficiency from other cytoplasmic maturation defects such as the thalassemias. Defects in heme synthesis, in contrast to globin synthesis, are less common and may be acquired or inherited (Chap. 358). Acquired abnormalities are usually associated with myelodysplasia, may lead to either a macro- or microcytic anemia, and are frequently associated with mitochondrial iron loading. In these cases, iron is taken up by the mitochondria of the developing erythroid cell but not incorporated into heme. The iron-encrusted mitochondria surround the nucleus of the erythroid cell, forming a ring. Based on the distinctive finding of so-called ringed sideroblasts on the marrow iron stain, patients are diagnosed as having a sideroblastic anemia—almost always reflecting myelodysplasia. Again, studies of iron parameters are helpful in the differential diagnosis of these patients.

BLOOD LOSS/HEMOLYTIC ANEMIA

In contrast to anemias associated with an inappropriately low reticulocyte production index, hemolysis is associated with red cell production indices ≥2.5 times normal. The stimulated erythropoiesis is reflected in the blood smear by the appearance of increased numbers of polychromatophilic macrocytes. A marrow examination is rarely indicated if the reticulocyte production index is increased appropriately. The red cell indices are typically normocytic or slightly macrocytic, reflecting the increased number of reticulocytes. Acute blood loss is not associated with an increased reticulocyte production index because of the time required to increase EPO production and, subsequently, marrow proliferation. Subacute blood loss may be associated with modest reticulocytosis. Anemia from chronic blood loss presents more often as iron deficiency than with the picture of increased red cell production.

The evaluation of blood loss anemia is usually not difficult. Most problems arise when a patient presents with an increased red cell production index from an episode of acute blood loss that went unrecognized. The cause of the anemia and increased red cell production may not be obvious. The confirmation of a recovering state may require observations over a period of 2–3 weeks, during which the hemoglobin concentration will be seen to rise and the reticulocyte production index fall (Chap. 106).

Hemolytic disease, while dramatic, is among the least common forms of anemia. The ability to sustain a high reticulocyte production index reflects the ability of the erythroid marrow to compensate for hemolysis and, in the case of extravascular hemolysis, the efficient recycling of iron from the destroyed red cells to support red cell production. With intravascular hemolysis, such as paroxysmal nocturnal hemoglobinuria, the loss of iron may limit the marrow response. The level of response depends on the severity of the anemia and the nature of the underlying disease process.

Hemoglobinopathies, such as sickle cell disease and the thalassemias, present a mixed picture. The reticulocyte index may be high but is inappropriately low for the degree of marrow erythroid hyperplasia (Chap. 104).

Hemolytic anemias present in different ways. Some appear suddenly as an acute, self-limited episode of intravascular or extravascular hemolysis, a presentation pattern often seen in patients with autoimmune hemolysis or with inherited defects of the Embden-Meyerhof pathway or the glutathione

reductase pathway. Patients with inherited disorders of the hemoglobin molecule or red cell membrane generally have a lifelong clinical history typical of the disease process. Those with chronic hemolytic disease, such as hereditary spherocytosis, may actually present not with anemia but with a complication stemming from the prolonged increase in red cell destruction such as symptomatic bilirubin gallstones or splenomegaly. Patients with chronic hemolysis are also susceptible to aplastic crises if an infectious process interrupts red cell production.

The differential diagnosis of an acute or chronic hemolytic event requires the careful integration of family history, the pattern of clinical presentation and—whether the disease is congenital or acquired—by a careful examination of the peripheral blood smear. Precise diagnosis may require more specialized laboratory tests, such as hemoglobin electrophoresis or a screen for red cell enzymes. Acquired defects in red cell survival are often immunologically mediated and require a direct or indirect antiglobulin test or a cold agglutinin titer to detect the presence of hemolytic antibodies or complement-mediated red cell destruction (Chap. 106).

Treatment: Anemia

An overriding principle is to initiate treatment of mild to moderate anemia only when a specific diagnosis is made. Rarely, in the acute setting, anemia may be so severe that red cell transfusions are required before a specific diagnosis is made. Whether the anemia is of acute or gradual onset, the selection of the appropriate treatment is determined by the documented cause(s) of the anemia. Often, the cause of the anemia is multifactorial. For example, a patient with severe rheumatoid arthritis who has been taking anti-inflammatory drugs may have a hypoproliferative anemia associated with chronic inflammation as well as chronic blood loss associated with intermittent gastrointestinal bleeding. In every circumstance, it is important to evaluate the patient's iron status fully before and during the treatment of any anemia. **Transfusion is discussed in Chap. 113; iron therapy is discussed in Chap. 103; treatment of megaloblastic anemia is discussed in Chap. 105; treatment of other entities is discussed in their respective chapters (sickle cell anemia, Chap. 104; hemolytic anemias, Chap. 106; aplastic anemia and myelodysplasia, Chap. 107).**

Therapeutic options for the treatment of anemias have expanded dramatically during the past 25 years. Blood component therapy is available and safe. Recombinant EPO as an adjunct to anemia management has transformed the lives of patients with chronic renal failure on dialysis and reduced transfusion needs of anemic cancer patients receiving chemotherapy. Eventually, patients with inherited disorders of globin synthesis or mutations in the globin gene, such as sickle cell disease, may benefit from the successful introduction of targeted genetic therapy (Chap. 68).

POLYCYTHEMIA

Polycythemia is defined as an increase in the hemoglobin above normal. This increase may be real or only apparent because of a decrease in plasma volume (spurious or relative polycythemia). The term *erythrocytosis* may be used interchangeably with polycythemia, but some draw a distinction between them: erythrocytosis implies documentation of increased red cell mass, whereas polycythemia refers to any increase in red cells. Often patients with polycythemia are detected through an incidental finding of elevated hemoglobin or hematocrit levels. Concern that the hemoglobin level may be abnormally high is usually triggered at 170 g/L (17 g/dL) for men and 150 g/L (15 g/dL) for women. Hematocrit levels >50% in men or >45% in women may be abnormal. Hematocrits >60% in men and >55% in women are almost invariably associated with an increased red cell mass. Given that the machine that quantitates red cell parameters actually measures hemoglobin concentrations and calculates hematocrits, hemoglobin levels may be a better index.

Features of the clinical history that are useful in the differential diagnosis include smoking history; current living at high altitude; or a history of congenital heart disease, sleep apnea, or chronic lung disease.

Patients with polycythemia may be asymptomatic or experience symptoms related to the increased red cell mass or the underlying disease process that leads to the increased red cell mass. The dominant symptoms from an increased red cell mass are related to hyperviscosity and thrombosis (both venous and arterial), because the blood viscosity increases logarithmically at hematocrits >55%. Manifestations range from digital ischemia to Budd-Chiari syndrome with hepatic vein thrombosis. Abdominal vessel thromboses are particularly common. Neurologic symptoms such as vertigo, tinnitus, headache, and visual disturbances may occur. Hypertension is often present. Patients with *polycythemia vera* may have aquagenic pruritus and symptoms related to hepatosplenomegaly. Patients may have easy bruising, epistaxis, or bleeding from the gastrointestinal tract. Peptic ulcer disease is common. Patients with hypoxemia may develop cyanosis on minimal exertion or have headache, impaired mental acuity, and fatigue.

The physical examination usually reveals a ruddy complexion. Splenomegaly favors polycythemia vera as the diagnosis (Chap. 108). The presence of cyanosis or evidence of a right-to-left shunt suggests congenital heart disease presenting in the adult, particularly tetralogy of Fallot's or Eisenmenger's syndrome (Chap. 236). Increased blood viscosity raises pulmonary artery pressure; hypoxemia can lead to increased pulmonary vascular resistance. Together, these factors can produce cor pulmonale.

Polycythemia can be spurious (related to a decrease in plasma volume; Gaisbock's syndrome), primary, or secondary in origin. The secondary causes are all associated with increases in EPO levels: either a physiologically adapted appropriate elevation based on tissue hypoxia (lung disease, high altitude, CO poisoning, high-affinity hemoglobinopathy) or an abnormal overproduction (renal cysts, renal artery stenosis, tumors with ectopic EPO production). A rare familial form of polycythemia is associated with normal EPO levels but hyperresponsive EPO receptors due to mutations.

Approach to the Patient: Polycythemia

As shown in Fig. 57-18, the first step is to document the presence of an increased red cell mass using the principle of isotope dilution by administering ^{51}Cr-labeled autologous red blood cells to the patient and sampling blood radioactivity over a 2-h period. If the red cell mass is normal (<36 mL/kg in men, <32 mL/kg in women), the patient has spurious or relative polycythemia. If the red cell mass is increased (>36 mL/kg in men, >32 mL/kg in women), serum EPO levels should be measured. If EPO levels are low or unmeasurable, the patient most likely has polycythemia vera. Tests that support this diagnosis include elevated white blood cell count, increased absolute basophil count, and thrombocytosis. A mutation in *JAK-2* (Val617Phe), a key member of the cytokine intracellular signaling pathway, can be found in 70–95% of patients with polycythemia vera.

Figure 57-18

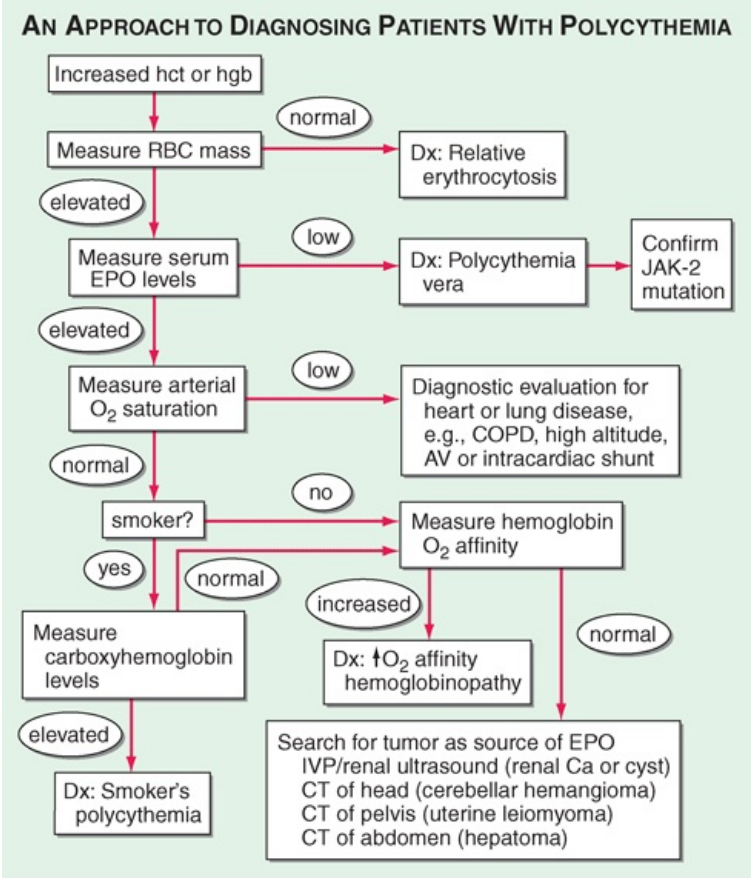

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An approach to the differential diagnosis of patients with an elevated hemoglobin (possible polycythemia). AV, atrioventricular; COPD, chronic obstructive pulmonary disease; EPO, erythropoietin; hct, hematocrit; IVP, intravenous pyelogram; RBC, red blood cell.

If serum EPO levels are elevated, one needs to distinguish whether the elevation is a physiologic response to hypoxia or is related to autonomous EPO production. Patients with low arterial O_2 saturation (<92%) should be further evaluated for the presence of heart or lung disease, if they are not living at high altitude. Patients with normal O_2 saturation who are smokers may have elevated EPO levels because of CO displacement of O_2. If carboxyhemoglobin (COHb) levels are high, the diagnosis is "smoker's polycythemia." Such patients should be urged to stop smoking. Those who cannot stop smoking require phlebotomy to control their polycythemia. Patients with normal O_2 saturation who do not smoke either have an abnormal hemoglobin that does not deliver O_2 to the tissues (evaluated by

finding elevated O_2–hemoglobin affinity) or have a source of EPO production that is not responding to the normal feedback inhibition. Further workup is dictated by the differential diagnosis of EPO-producing neoplasms. Hepatoma, uterine leiomyoma, and renal cancer or cysts are all detectable with abdominopelvic CT scans. Cerebellar hemangiomas may produce EPO, but they present with localizing neurologic signs and symptoms rather than polycythemia-related symptoms.

FURTHER READINGS

Hillman RS et al: *Hematology in Clinical Practice*, 5th ed. New York, McGraw-Hill, 2010

McMullin MF: The classification and diagnosis of erythrocytosis. Int J Lab Hematol 30:447, 2008[PMID: 18823397]

Rizzo JD et al: American Society of Hematology/American Society of Clinical Oncology clinical practice guideline update on the use of epoetin and darbepoetin in adult patients with cancer. Blood 116:4045, 2010[PMID: 20974674]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BLEEDING AND THROMBOSIS: INTRODUCTION

The human hemostatic system provides a natural balance between procoagulant and anticoagulant forces. The procoagulant forces include platelet adhesion and aggregation and fibrin clot formation; anticoagulant forces include the natural inhibitors of coagulation and fibrinolysis. Under normal circumstances, hemostasis is regulated to promote blood flow; however, it is also prepared to clot blood rapidly to arrest blood flow and prevent exsanguination. After bleeding is successfully halted, the system remodels the damaged vessel to restore normal blood flow. The major components of the hemostatic system, which function in concert, are (1) platelets and other formed elements of blood, such as monocytes and red cells; (2) plasma proteins (the coagulation and fibrinolytic factors and inhibitors); and (3) the vessel wall.

STEPS OF NORMAL HEMOSTASIS

Platelet Plug Formation

On vascular injury, platelets adhere to the site of injury, usually the denuded vascular intimal surface. Platelet adhesion is mediated primarily by von Willebrand factor (VWF), a large multimeric protein present in both plasma and the extracellular matrix of the subendothelial vessel wall, which serves as the primary "molecular glue," providing sufficient strength to withstand the high levels of shear stress that would tend to detach them with the flow of blood. Platelet adhesion is also facilitated by direct binding to subendothelial collagen through specific platelet membrane collagen receptors.

Platelet adhesion results in subsequent platelet activation and aggregation. This process is enhanced and amplified by humoral mediators in plasma (e.g., epinephrine, thrombin); mediators released from activated platelets (e.g., adenosine diphosphate, serotonin); and vessel wall extracellular matrix constituents that come in contact with adherent platelets (e.g., collagen, VWF). Activated platelets undergo the release reaction, during which they secrete contents that further promote aggregation and inhibit the naturally anticoagulant endothelial cell factors. During platelet aggregation (platelet-platelet interaction), additional platelets are recruited from the circulation to the site of vascular injury, leading to the formation of an occlusive platelet thrombus. The platelet plug is anchored and stabilized by the developing fibrin mesh.

The platelet glycoprotein (Gp) IIb/IIIa ($\alpha_{IIb}\beta_3$) complex is the most abundant receptor on the platelet surface. Platelet activation converts the normally inactive Gp IIb/IIIa receptor into an active receptor, enabling binding to fibrinogen and VWF. Because the surface of each platelet has about 50,000 Gp IIb/IIIa-binding sites, numerous activated platelets recruited to the site of vascular injury can rapidly form an occlusive aggregate by means of a dense network of intercellular fibrinogen bridges. Since this receptor is the key mediator of platelet aggregation, it has become an effective target for antiplatelet therapy.

Fibrin Clot Formation

Plasma coagulation proteins (*clotting factors*) normally circulate in plasma in their inactive forms. The sequence of coagulation protein reactions that culminate in the formation of fibrin was originally described as a *waterfall* or a *cascade*. Two pathways of blood coagulation have been described in the past: the so-called extrinsic, or tissue factor, pathway and the so-called intrinsic, or contact activation, pathway. We now know that coagulation is normally initiated through tissue factor (TF) exposure and activation through the classic *extrinsic pathway* but with critically important amplification through elements of the classic *intrinsic pathway*, as illustrated in **Fig. 58-1**. These reactions take place on phospholipid surfaces, usually the activated platelet surface. Coagulation testing in the laboratory can reflect other influences due to the artificial nature of the in vitro systems used (see below).

Figure 58-1

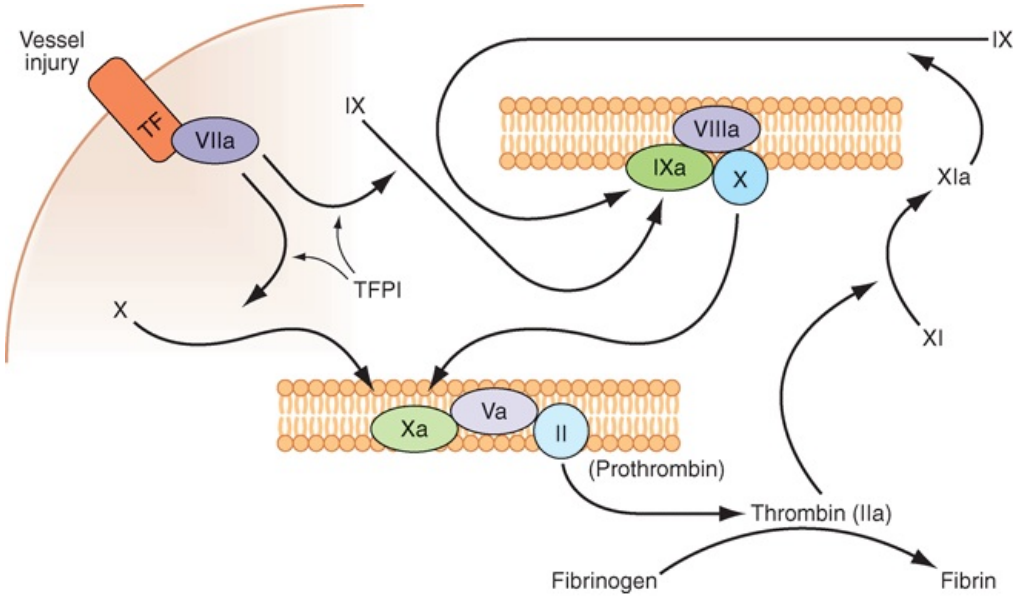

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coagulation is initiated by tissue factor (TF) exposure, which, with factor (F)VIIa, activates FIX and FX, which in turn, with FVIII and FV as cofactors, respectively, results in thrombin formation and subsequent conversion of fibrinogen to fibrin. Thrombin activates FXI, FVIII, and FV, amplifying the coagulation signal. Once the TF/FVIIa/FXa complex is formed, tissue factor pathway inhibitor (TFPI) inhibits the TF/FVIIa pathway, making coagulation dependent on the amplification loop through FIX/FVIII. Coagulation requires calcium (not shown) and takes place on phospholipid surfaces, usually the activated platelet membrane.

The immediate trigger for coagulation is vascular damage that exposes blood to TF that is constitutively expressed on the surfaces of subendothelial cellular components of the vessel wall, such as smooth muscle cells and fibroblasts. TF is also present in circulating microparticles, presumably shed from cells including monocytes and platelets. TF binds the serine protease factor VIIa; the complex activates factor X to factor Xa. Alternatively, the complex can indirectly activate factor X by initially converting factor IX to factor IXa, which then activates factor X. The participation of factor XI in hemostasis is not dependent on its activation by factor XIIa but rather on its positive feedback activation by thrombin. Thus, factor XIa functions in the propagation and amplification, rather than in the initiation, of the coagulation cascade.

Factor Xa can be formed through the actions of either the tissue factor/factor VIIa complex or factor IXa (with factor VIIIa as a cofactor) and converts prothrombin to thrombin, the pivotal protease of the coagulation system. The essential cofactor for this reaction is factor Va. Like the homologous factor VIIIa, factor Va is produced by thrombin-induced limited proteolysis of factor V. Thrombin is a multifunctional enzyme that converts soluble plasma fibrinogen to an insoluble fibrin matrix. Fibrin polymerization involves an orderly process of intermolecular associations (**Fig. 58-2**). Thrombin also activates factor XIII (fibrin-stabilizing factor) to factor XIIIa, which covalently cross-links and thereby stabilizes the fibrin clot.

Figure 58-2

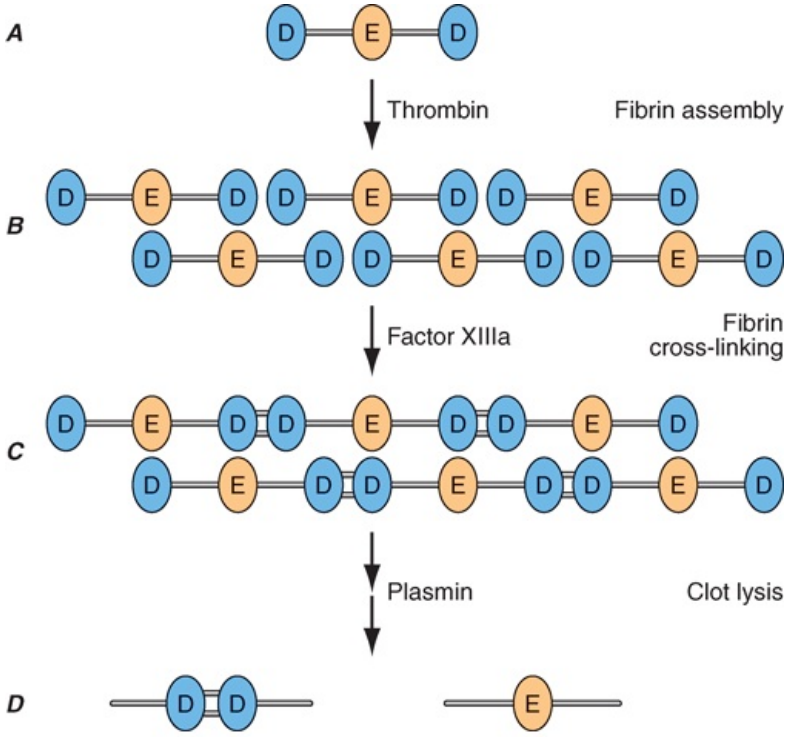

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fibrin formation and dissolution. (*A*). Fibrinogen is a trinodular structure consisting of 2 D domains and 1 E domain. Thrombin activation results in an ordered lateral assembly of protofibrils (*B*) with noncovalent associations. FXIIIa cross-links the D domains on adjacent molecules (*C*). Fibrin and fibrinogen (not shown) lysis by plasmin occurs at discrete sites and results in intermediary fibrin(ogen) degradation products (not shown). D-Dimers are the product of complete lysis of fibrin (*D*), maintaining the cross-linked D domains.

The assembly of the clotting factors on activated cell membrane surfaces greatly accelerates their reaction rates and also serves to localize blood clotting to sites of vascular injury. The critical cell membrane components, acidic phospholipids, are not normally exposed on resting cell membrane surfaces. However, when platelets, monocytes, and endothelial cells are activated by vascular injury or inflammatory stimuli, the procoagulant head groups of the membrane anionic phospholipids become translocated to the surfaces of these cells or released as part of microparticles, making them available to support and promote the plasma coagulation reactions.

ANTITHROMBOTIC MECHANISMS

Several physiologic antithrombotic mechanisms act in concert to prevent clotting under normal circumstances. These mechanisms operate to preserve blood fluidity and to limit blood clotting to specific focal sites of vascular injury. Endothelial cells have many antithrombotic effects. They produce prostacyclin, nitric oxide, and ectoADPase/CD39, which act to inhibit platelet binding, secretion, and aggregation. Endothelial cells produce anticoagulant factors including heparan proteoglycans, antithrombin, TF pathway inhibitor, and thrombomodulin. They also activate fibrinolytic mechanisms through the production of tissue plasminogen activator 1, urokinase, plasminogen activator inhibitor, and annexin-2. The sites of action of the major physiologic antithrombotic pathways are shown in **Fig. 58-3**.

Figure 58-3

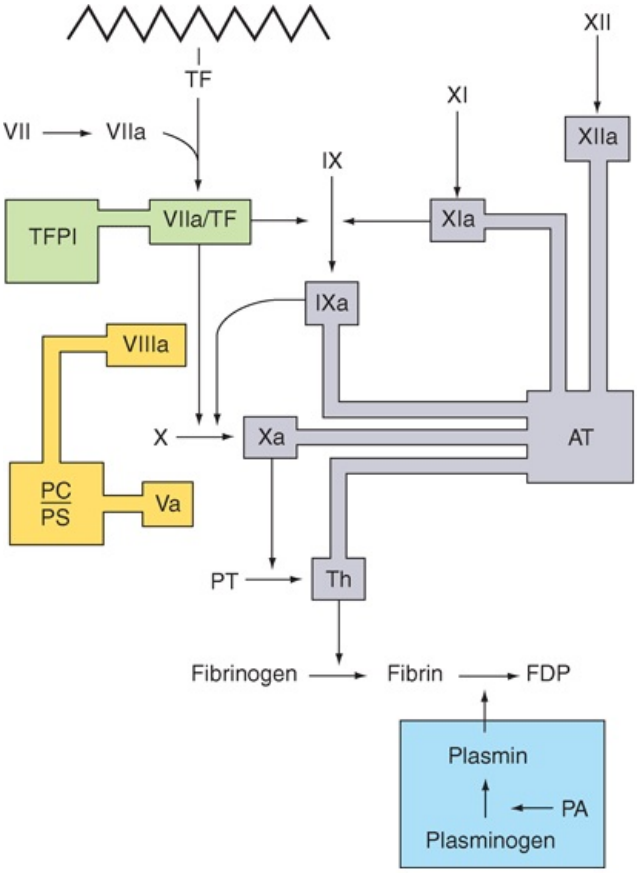

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sites of action of the four major physiologic antithrombotic pathways: antithrombin (AT); protein C/S (PC/PS); tissue factor pathway inhibitor (TFPI); and the fibrinolytic system, consisting of plasminogen, plasminogen activator (PA), and plasmin. PT, prothrombin; Th, thrombin; FDP, fibrin(ogen) degradation products. [*Modified from BA Konkle, AI Schafer, in DP Zipes et al (eds): Braunwald's Heart Disease, 7th ed. Philadelphia, Saunders, 2005.*]

Antithrombin (or antithrombin III) is the major plasma protease inhibitor of thrombin and the other clotting factors in coagulation. Antithrombin neutralizes thrombin and other activated coagulation factors by forming a complex between the active site of the enzyme and the reactive center of antithrombin. The rate of formation of these inactivating complexes increases by a factor of several thousand in the presence of heparin. Antithrombin inactivation of thrombin and other activated clotting factors occurs physiologically on vascular surfaces, where glycosoaminoglycans, including heparan sulfates, are present to catalyze these reactions. Inherited quantitative or qualitative deficiencies of antithrombin lead to a lifelong predisposition to venous thromboembolism.

Protein C is a plasma glycoprotein that becomes an anticoagulant when it is activated by thrombin. The thrombin-induced activation of protein C occurs physiologically on thrombomodulin, a transmembrane proteoglycan-binding site for thrombin on endothelial cell surfaces. The binding of protein C to its receptor on endothelial cells places it in proximity to the thrombin-thrombomodulin complex, thereby enhancing its activation efficiency. Activated protein C acts as an anticoagulant by cleaving and inactivating activated factors V and VIII. This reaction is accelerated by a cofactor, protein S, which, like protein C, is a glycoprotein that undergoes vitamin K–dependent posttranslational modification. Quantitative or qualitative deficiencies of protein C or protein S, or resistance to the action of activated protein C by a specific mutation at its target cleavage site in factor Va (factor V Leiden), lead to hypercoagulable states.

Tissue factor pathway inhibitor (TFPI) is a plasma protease inhibitor that regulates the TF–induced extrinsic pathway of coagulation. TFPI inhibits the TF/FVIIa/FXa complex, essentially turning off the TF/FVIIa initiation of coagulation, which then becomes dependent on the "amplification loop" via FXI and FVIII activation by thrombin. TFPI is bound to lipoprotein and can also be released by heparin from endothelial cells, where it is bound to glycosoaminoglycans, and from platelets. The heparin-mediated release of TFPI may play a role in the anticoagulant effects of unfractionated and low-molecular-weight heparins.

The Fibrinolytic System

Any thrombin that escapes the inhibitory effects of the physiologic anticoagulant systems is available to convert fibrinogen to fibrin. In response, the endogenous fibrinolytic system is then activated to

dispose of intravascular fibrin and thereby maintain or reestablish the patency of the circulation. Just as thrombin is the key proteaseenzyme of the coagulation system, plasmin is the major protease enzyme of the fibrinolytic system, acting to digest fibrin to fibrin degradation products. The general scheme of fibrinolysis and its control is shown in **Fig. 58-4**.

Figure 58-4

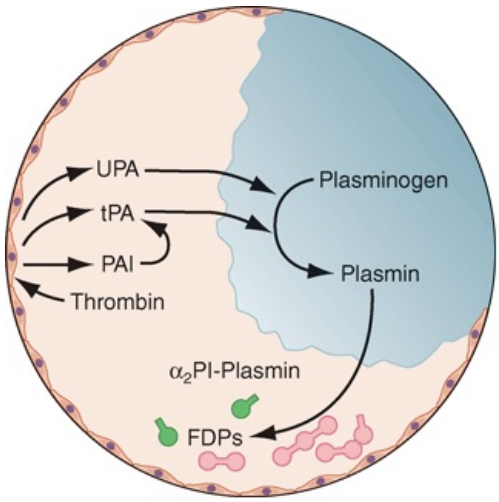

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A schematic diagram of the fibrinolytic system. Tissue plasminogen activator (tPA) is released from endothelial cells, binds the fibrin clot, and activates plasminogen to plasmin. Excess fibrin is degraded by plasmin to distinct degradation products (FDPs). Any free plasmin is complexed with α_2-antiplasmin (α_2PI).

The plasminogen activators, tissue type plasminogen activator (tPA) and the urokinase type plasminogen activator cleave (uPA), cleave the Arg560-Val561 bond of plasminogen to generate the active enzyme plasmin. The lysine-binding sites of plasmin (and plasminogen) permit it to bind to fibrin, so that physiologic fibrinolysis is "fibrin specific." Both plasminogen (through its lysine-binding sites) and tPA possess specific affinity for fibrin and thereby bind selectively to clots. The assembly of a ternary complex, consisting of fibrin, plasminogen, and tPA, promotes the localized interaction between plasminogen and tPA and greatly accelerates the rate of plasminogen activation to plasmin. Moreover, partial degradation of fibrin by plasmin exposes new plasminogen and tPA-binding sites in carboxy-terminus lysine residues of fibrin fragments to enhance these reactions further. This creates a highly efficient mechanism to generate plasmin focally on the fibrin clot, which then becomes plasmin's substrate for digestion to fibrin degradation products. Plasmin cleaves fibrin at distinct sites of the fibrin molecule leading to the generation of characteristic fibrin fragments during the process of fibrinolysis (Fig. 58-2). The sites of plasmin cleavage of fibrin are the same as those in fibrinogen. However, when plasmin acts on covalently cross-linked fibrin, D-dimers are released; hence, D-dimers can be measured in plasma as a relatively specific test of fibrin (rather than fibrinogen) degradation. D-dimer assays can be used as sensitive markers of blood clot formation, and some have been validated for clinical use to exclude the diagnosis of deep-venous thrombosis (DVT) and pulmonary embolism in selected populations.

Physiologic regulation of fibrinolysis occurs primarily at three levels: (1) plasminogen activator inhibitors (PAIs), specifically PAI-1 and PAI-2, inhibit the physiologic plasminogen activators; (2) the thrombin-activatable fibrinolysis inhibitor (TAFI) limits fibrinolysis; and (3) α_2-antiplasmin inhibits plasmin. PAI1 is the primary inhibitor of tPA and uPA in plasma. TAFI cleaves the N-terminal lysine residues of fibrin, which aid in localization of plasmin activity. α_2-Antiplasmin is the main inhibitor of plasmin in human plasma, inactivating any non-fibrin clot-associated plasmin.

CLINICAL PRESENTATION

Disorders of hemostasis may be either inherited or acquired. A detailed personal and family history is key in determining the chronicity of symptoms and the likelihood of the disorder being inherited, and it provides clues to underlying conditions that have contributed to the bleeding or thrombotic state. In addition, the history can give clues as to the etiology by determining (1) the bleeding (mucosal and/or joint) or thrombosis (arterial and/or venous) site and (2) whether an underlying bleeding or clotting tendency was enhanced by another medical condition or the introduction of medications or dietary supplements.

HISTORY OF BLEEDING

A history of bleeding is the most important predictor of bleeding risk. In evaluating a patient for a bleeding disorder, a history of at-risk situations, including the response to past surgeries, should be assessed. Does the patient have a history of spontaneous or trauma/surgery-induced bleeding? Spontaneous hemarthroses are a hallmark of moderate and severe factor VIII and IX deficiency and, in rare circumstances, of other clotting factor deficiencies. Mucosal bleeding symptoms are more suggestive of underlying platelet disorders or von Willebrand disease (VWD), termed *disorders of primary hemostasis or platelet plug formation*. Disorders affecting primary hemostasis are shown in **Table 58–1**.

Table 58–1 Primary Hemostatic (Platelet Plug) Disorders

Defects of Platelet Adhesion
Von Willebrand disease
Bernard-Soulier syndrome (absence of dysfunction of GpIb-IX-V)
Defects of Platelet Aggregation
Glanzmann's thrombasthenia (absence or dysfunction of GpIIbIIIa)
Afibrinogenemia
Defects of Platelet Secretion
Decreased cyclooxygenase activity
Drug-induced (aspirin, nonsteroidal anti-inflammatory agents, thienopyridines)
Inherited
Granule storage pool defects
Inherited Acquired
Non-specific inherited secretory defects
Non-specific drug effects
Uremia
Platelet coating (e.g., paraprotein, penicillin)
Defect of Platelet Coagulant Activity
Scott's syndrome

A bleeding score has been validated as a tool to predict patients more likely to have Type 1 VWD. Studies are under way to validate additional formats, including ones that are easier to administer and improving performance in pediatric populations. Bleeding symptoms that appear to be more common in patients with bleeding disorders include prolonged bleeding with surgery, dental procedures and extractions, and/or trauma, menorrhagia or postpartum hemorrhage, and large bruises (often described with lumps).

Easy bruising and menorrhagia are common complaints in patients with and without bleeding disorders. Easy bruising can also be a sign of medical conditions in which there is no identifiable coagulopathy; instead, the conditions are caused by an abnormality of blood vessels or their supporting tissues. In Ehlers-Danlos syndrome there may be posttraumatic bleeding and a history of joint hyperextensibility. Cushing's syndrome, chronic steroid use, and aging result in changes in skin and subcutaneous tissue, and subcutaneous bleeding occurs in response to minor trauma. The latter has been termed *senile purpura*.

Epistaxis is a common symptom, particularly in children and in dry climates, and may not reflect an underlying bleeding disorder. However it is the most common symptom in hereditary hemorrhagic telangiectasia and in boys with VWD. Clues that epistaxis is a symptom of an underlying bleeding disorder include lack of seasonal variation and bleeding that requires medical evaluation or treatment, including cauterization. Bleeding with eruption of primary teeth is seen in children with more severe bleeding disorders, such as moderate and severe hemophilia. It is uncommon in children with mild bleeding disorders. Patients with disorders of primary hemostasis (platelet adhesion) may have increased bleeding after dental cleanings and other procedures that involve gum manipulation.

Menorrhagia is defined quantitatively as a loss of >80 mL of blood per cycle, based on blood loss required to produce iron-deficiency anemia. A complaint of heavy menses is subjective and has a poor correlation with excessive blood loss. Predictors of menorrhagia include bleeding resulting in iron-deficiency anemia or a need for blood transfusion, passage of clots >1 inch in diameter and changing a pad or tampon more than hourly. Menorrhagia is a common symptom in women with underlying bleeding disorders and is reported in the majority of women with VWD and factor XI deficiency and in symptomatic carriers of hemophilia A. Women with underlying bleeding disorders are more likely to have other bleeding symptoms, including bleeding after dental extractions, postoperative bleeding, and

postpartum bleeding, and are much more likely to have menorrhagia beginning at menarche than women with menorrhagia due to other causes.

Postpartum hemorrhage (PPH) is a common symptom in women with underlying bleeding disorders. In women with Type 1 VWD and symptomatic carriers of hemophilia in whom levels of VWF and FVIII usually normalize during pregnancy, PPH may be delayed. Women with a history of postpartum hemorrhage have a high risk of recurrence with subsequent pregnancies. Rupture of ovarian cysts with intraabdominal hemorrhage has also been reported in women with underlying bleeding disorders.

Tonsillectomy is a major hemostatic challenge, as intact hemostatic mechanisms are essential to prevent excessive bleeding from the tonsillar bed. Bleeding may occur early after surgery or after approximately 7 days postoperatively, with loss of the eschar at the operative site. Similar delayed bleeding is seen after colonic polyp resection. Gastrointestinal (GI) bleeding and hematuria are usually due to underlying pathology, and procedures to identify and treat the bleeding site should be undertaken, even in patients with known bleeding disorders. VWD, particularly types 2 and 3, has been associated with angiodysplasia of the bowel and GI bleeding.

Hemarthroses and spontaneous muscle hematomas are characteristic of moderate or severe congenital factor VIII or IX deficiency. They can also be seen in moderate and severe deficiencies of fibrinogen, prothrombin, and of factors V, VII, and X. Spontaneous hemarthroses occur rarely in other bleeding disorders except for severe VWD, with associated FVIII levels <5%. Muscle and soft tissue bleeds are also common in acquired FVIII deficiency. Bleeding into a joint results in severe pain and swelling, as well as loss of function, but is rarely associated with discoloration from bruising around the joint. Life-threatening sites of bleeding include bleeding into the oropharynx, where bleeding can obstruct the airway, into the central nervous system, and into the retroperitoneum. Central nervous system bleeding is the major cause of bleeding-related deaths in patients with severe congenital factor deficiencies.

PROHEMORRHAGIC EFFECTS OF MEDICATIONS AND DIETARY SUPPLEMENTS

Aspirin and other nonsteroidal inflammatory drugs (NSAIDs) that inhibit cyclooxygenase 1 impair primary hemostasis and may exacerbate bleeding from another cause or even unmask a previously occult mild bleeding disorder such as VWD. All NSAIDs, however, can precipitate gastrointestinal bleeding, which may be more severe in patients with underlying bleeding disorders. The aspirin effect on platelet function as assessed by aggregometry can persist for up to 7 days, although it has frequently returned to normal by 3 days after the last dose. The effect of other NSAIDs is shorter, as the inhibitor effect is reversed when the drug is removed. Thienopyridines (clopidogrel and prasugrel) inhibit ADP-mediated platelet aggregation and like NSAIDs can precipitate or exacerbate bleeding symptoms.

Many herbal supplements can impair hemostatic function (**Table 58–2**). Some are more convincingly associated with a bleeding risk than others. Fish oil or concentrated omega 3 fatty acid supplements impair platelet function. They alter platelet biochemistry to produce more PGI3, a more potent platelet inhibitor than prostacyclin (PGI2), and more thromboxane A3, a less potent platelet activator than thromboxane A2. In fact, diets naturally rich in omega 3 fatty acids can result in a prolonged bleeding time and abnormal platelet aggregation studies, but the actual associated bleeding risk is unclear. Vitamin E appears to inhibit protein kinase C–mediated platelet aggregation and nitric oxide production. In patients with unexplained bruising or bleeding, it is prudent to review any new medications or supplements and discontinue those that may be associated with bleeding.

Table 58–2 Herbal Supplements Associated with Increased Bleeding

Herbs With Potential Anti-Platelet Activity
Ginkgo (*Ginkgo biloba L.*)
Garlic (*Allium sativum*)
Bilberry (*Vaccinium myrtillus*)
Ginger (*Gingiber officinale*)
Dong quai (*Angelica sinensis*)
Feverfew (*Tanacetum parthenium*)
Asian ginseng (*Panax ginseng*)
American ginseng (*Panax quinquefolius*)
Siberian ginseng/eleuthero (*Eleutherococcus senticosus*)
Turmeric (*Circuma longa*)
Meadowsweet (*Filipendula ulmaria*)
Willow (*Salix* spp.)
Coumarin-Containing Herbs

Motherworth (*Leonurus cardiaca*)
Chamomile (*Matricaria recutita, Chamaemelum mobile*)
Horse chestnut (*Aesculus hippocastanum*)
Red clover (*Trifolium pratense*)
Fenugreek (*Trigonella foenum-graecum*)

UNDERLYING SYSTEMIC DISEASES THAT CAUSE OR EXACERBATE A BLEEDING TENDENCY

Acquired bleeding disorders are commonly secondary to, or associated with, systemic disease. The clinical evaluation of a patient with a bleeding tendency must therefore include a thorough assessment for evidence of underlying disease. Bruising or mucosal bleeding may be the presenting complaint in liver disease, severe renal impairment, hypothyroidism, paraproteinemias or amyloidosis, and conditions causing bone marrow failure. All coagulation factors are synthesized in the liver, and hepatic failure results in combined factor deficiencies. This is often compounded by thrombocytopenia from splenomegaly due to portal hypertension. Coagulation factors II, VII, IX, X and proteins C, S, and Z are dependent on vitamin K for posttranslational modification. Although vitamin K is required in both procoagulant and anticoagulant processes, the phenotype of vitamin K deficiency or the warfarin effect on coagulation is bleeding.

The normal blood platelet count is 150,000–450,000/μL. Thrombocytopenia results from decreased production, increased destruction, and/or sequestration. Although the bleeding risk varies somewhat by the reason for the thrombocytopenia, bleeding rarely occurs in isolated thrombocytopenia at counts <50,000/μL and usually not until <10,000–20,000/μL. Coexisting coagulopathies, as is seen in liver failure or disseminated coagulation; infection, platelet-inhibitory drugs; and underlying medical conditions can all increase the risk of bleeding in the thrombocytopenic patient. Most procedures can be performed in patients with a platelet count of 50,000/μL. The level needed for major surgery will depend on the type of surgery and the patients' underlying medical state, although a count of approximately 80,000/μL is likely sufficient.

HISTORY OF THROMBOSIS

The risk of thrombosis, like that of bleeding, is influenced by both genetic and environmental influences. The major risk factor for arterial thrombosis is atherosclerosis, while for venous thrombosis the risk factors are immobility, surgery, underlying medical conditions such as malignancy, medications such as hormonal therapy, obesity, and genetic predispositions. Factors that increase risks for venous and for both venous and arterial thromboses are shown in **Table 58–3**.

Table 58–3 Risk Factors for Thrombosis

Venous	Venous and Arterial
Inherited	**Inherited**
Factor V Leiden	Homocystinuria
Prothrombin G20210A	Dysfibrinogenemia
Antithrombin deficiency	**Mixed (inherited and acquired)**
Protein C deficiency	Hyperhomocysteinemia
Protein S deficiency	**Acquired**
Elevated FVIII	Malignancy
Acquired	Antiphospholipid antibody syndrome
Age	Hormonal therapy
Previous thrombosis	Polycythemia vera
Immobilization	Essential thrombocythemia
Major surgery	Paroxysmal nocturnal hemoglobinuria
Pregnancy and puerperium	Thrombotic thrombocytopenic purpura
Hospitalization	Heparin-induced thrombocytopenia
Obesity	Disseminated intravascular coagulation
Infection	
APC resistance, nongenetic	
Smoking	
Unknown*	

Elevated factor II, IX, XI Elevated TAFI levels Low levels of TFPI	

* Unknown whether risk is inherited or acquired.

Abbreviations: APC, activated protein C; TAFI, thrombin-activatable fibrinolysis inhibitor; TFPI, tissue factor pathway inhibitor.

The most important point in a history related to venous thrombosis is determining whether the thrombotic event was idiopathic (meaning there was no clear precipitating factor) or was a precipitated event. In patients without underlying malignancy, having an idiopathic event is the strongest predictor of recurrence of venous thromboembolism. In patients who have a vague history of thrombosis, a history of being treated with warfarin suggests a past DVT. Age is an important risk factor for venous thrombosis—the risk of DVT increasing per decade, with an approximate incidence of 1/100,000 per year in early childhood to 200 per year among octogenarians. Family history is helpful in determining if there is a genetic predisposition and how strong that predisposition appears to be. A genetic thrombophilia that confers a relatively small increased risk, such as being a heterozygote for the prothrombin G20210A or factor V Leiden mutation, may be a minor determinant of risk in an elderly individual undergoing a high-risk surgical procedure. As illustrated in **Fig. 58-5**, a thrombotic event usually has more than one contributing factor. Predisposing factors must be carefully assessed to determine the risk of recurrent thrombosis and, with consideration of the patient's bleeding risk, determine the length of anticoagulation. Similar consideration should be given in determining the need to test the patient and family members for thrombophilias.

Figure 58-5

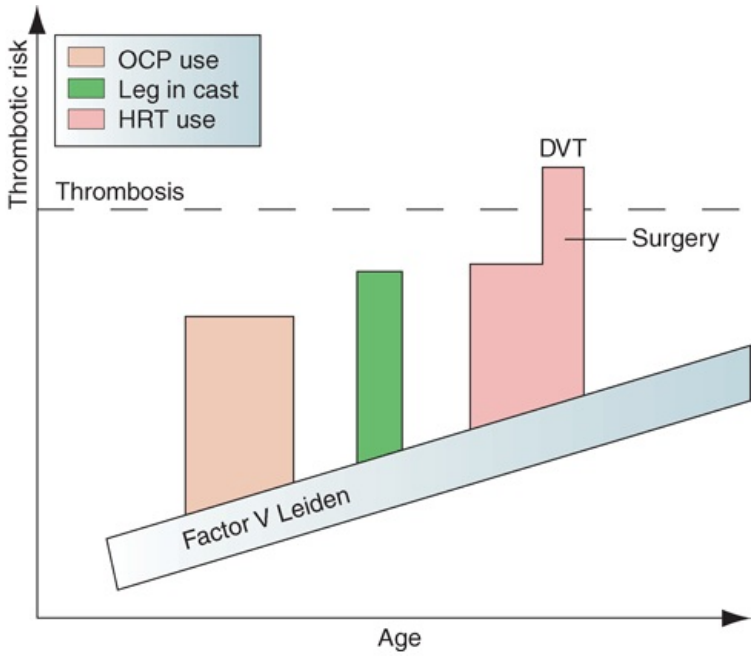

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thrombotic risk over time. Shown schematically is an individual's thrombotic risk over time. An underlying factor V Leiden mutation provides a "theoretically" constant increased risk. The thrombotic risk increases with age and, intermittently, with oral contraceptive (OCP) or hormone replacement (HRT) use; other events may increase the risk further. At some point the cumulative risk may increase to the threshold for thrombosis and result in deep-venous thrombosis (DVT). Note: The magnitude and duration of risk portrayed in the figure is meant for example only and may not precisely reflect the relative risk determined by clinical study. *[From BA Konkle, A Schafer, in DP Zipes et al (eds): Braunwald's Heart Disease, 7th ed. Philadelphia, Saunders, 2005; modified with permission from FR Rosendaal: Venous thrombosis: A multicausal disease. Lancet 353:1167, 1999.]*

LABORATORY EVALUATION

Careful history taking and clinical examination are essential components in the assessment of bleeding and thrombotic risk. The use of laboratory tests of coagulation complement, but cannot substitute for, clinical assessment. No test exists that provides a global assessment of hemostasis. The bleeding time has been used to assess bleeding risk; however, it does not predict bleeding risk with surgery and it is

not recommended for this indication. The PFA-100, an instrument that measures platelet-dependent coagulation under flow conditions, is more sensitive and specific for platelet disorders and VWD than the bleeding time; however it is not sensitive enough to rule out underlying mild bleeding disorders. Also, its utility in predicting bleeding risk has not been determined.

For routine preoperative and preprocedure testing, an abnormal prothrombin time (PT) may detect liver disease or vitamin K deficiency that had not been previously appreciated. Studies have not confirmed the usefulness of an activated partial thromboplastin time (aPTT) in preoperative evaluations in patients with a negative bleeding history. The primary use of coagulation testing should be to confirm the presence and type of bleeding disorder in a patient with a suspicious clinical history.

Because of the nature of coagulation assays, proper sample acquisition and handling is critical to obtaining valid results. In patients with abnormal coagulation assays who have no bleeding history, repeat studies with attention to these factors frequently results in normal values. Most coagulation assays are performed in sodium citrate anticoagulated plasma that is recalcified for the assay. Because the anticoagulant is in liquid solution and needs to be added to blood in proportion to the plasma volume, incorrectly filled or inadequately mixed blood collection tubes will give erroneous results. Vacutainer tubes should be filled to >90% of the recommended fill, which is usually denoted by a line on the tube. An elevated hematocrit (>55%) can result in a false value due to a decreased plasma to anticoagulant ratio.

SCREENING ASSAYS

The most commonly used screening tests are the PT, aPTT, and platelet count. The PT assesses the factors I (fibrinogen), II (prothrombin), V, VII, and X **(Fig. 58-6)**. The PT measures the time for clot formation of the citrated plasma after recalcification and addition of thromboplastin, a mixture of TF and phospholipids. The sensitivity of the assay varies by the source of thromboplastin. The relationship between defects in secondary hemostasis (fibrin formation) and coagulation test abnormalities is shown in **Table 58–4**. To adjust for this variability, the overall sensitivity of different thromboplastins to reduction of the vitamin K–dependent clotting factors II, VII, IX, and X in anticoagulation patients is now expressed as the International Sensitivity Index (ISI). An inverse relationship exists between ISI and thromboplastin sensitivity. The international normalized ratio (INR) is then determined based on the formula: $INR = (PT_{patient}/PT_{normal\ mean})^{ISI}$.

Figure 58-6

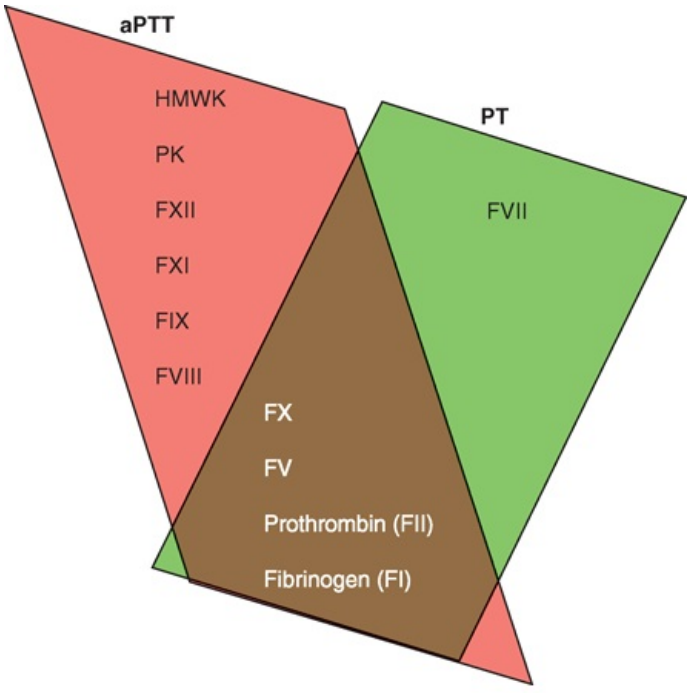

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coagulation factor activity tested in the activated partial thromboplastin time (aPTT) in red and prothrombin time (PT) in green, or both. F, factor; HMWK, high-molecular-weight kininogen; PK, prekallikrein.

Table 58–4 Hemostatic Disorders and Coagulation Test Abnormalities

Prolonged Activated Partial Thromboplastin Time (aPTT)
No clinical bleeding—↓ factors XII, high-molecular-weight kininogen, prekallikrein
Variable, but usually mild, bleeding—↓ factor XI, mild↑ FVIII and FIX
Frequent, severe bleeding—severe deficiencies of FVIII and FIX
Heparin
Prolonged Prothrombin Time (PT)
Factor VII deficiency
Vitamin K deficiency—early
Warfarin anticoagulation
Prolonged aPTT and PT
Factor II, V, X, or fibrinogen deficiency
Vitamin K deficiency—late
Direct thrombin inhibitors
Prolonged Thrombin Time
Heparin or heparin-like inhibitors
Mild or no bleeding—dysfibrinogenemia
Frequent, severe bleeding—afibrinogenemia
Prolonged PT and/or aPTT Not Correct with Mixing with
Normal Plasma
Bleeding—specific factor inhibitor
No symptoms, or clotting and/or pregnancy loss—lupus anticoagulant
Disseminated intravascular coagulation
Heparin or direct thrombin inhibitor
Abnormal Clot Solubility
Factor XIII deficiency
Inhibitors or defective cross-linking
Rapid Clot Lysis
Deficiency of α_2-antiplasmin or plasminogen activator inhibitor 1
Treatment with fibrinolytic therapy

The INR was developed to assess anticoagulation due to reduction of vitamin K–dependent coagulation factors; it is commonly used in the evaluation of patients with liver disease. While it does allow comparison between laboratories, reagent sensitivity as used to determine the ISI is not the same in liver disease as with warfarin anticoagulation. In addition, progressive liver failure is associated with variable changes in coagulation factors; the degree of prolongation of either the PT or the INR only roughly predicts the bleeding risk. Thrombin generation has been shown to be normal in many patients with mild to moderate liver dysfunction. As the PT only measures one aspect of hemostasis affected by liver dysfunction, we likely overestimate the bleeding risk of a mildly elevated INR in this setting.

The aPTT assesses the intrinsic and common coagulation pathways, factors XI, IX, VIII, X, V, II, fibrinogen, and also prekallikrein, high-molecular-weight kininogen and factor XII (Fig. 58-6). The aPTT reagent contains phospholipids derived from either animal or vegetable sources that function as a platelet substitute in the coagulation pathways and includes an activator of the intrinsic coagulation system, such as nonparticulate ellagic acid or the particulate activators kaolin, celite, or micronized silica.

The phospholipid composition of aPTT reagents varies, which influences the sensitivity of individual reagents to clotting factor deficiencies and to inhibitors such as heparin and lupus anticoagulants. Thus, aPTT results will vary from one laboratory to another and the normal range in the laboratory where the testing occurs should be used in the interpretation. Local laboratories cam relate their aPTT values to the therapeutic heparin anticoagulation by correlating aPTT values with direct measurements of heparin activity (anti-Xa or protamine titration assays) in samples from heparinized patients, although correlation between these assays is often poor. The aPTT reagent will vary in sensitivity to individual factor deficiencies and usually becomes prolonged with individual factor deficiencies of 30 –50%.

MIXING STUDIES

Mixing studies are used to evaluate a prolonged aPTT or, less commonly PT, to distinguish between a factor deficiency and an inhibitor. In this assay, normal plasma and patient plasma are mixed in a 1:1 ratio, and the aPTT or PT determined immediately and after incubation at 37°C for varying times, typically 30, 60, and/or 120 min. With isolated factor deficiencies, the aPTT will correct with mixing and stay corrected with incubation. With aPTT prolongation due to a lupus anticoagulant, the mixing and incubation will show no correction. In acquired neutralizing factor antibodies, such as an acquired factor VIII inhibitor, the initial assay may or may not correct immediately after mixing but will prolong or remain prolonged with incubation at 37°C. Failure to correct with mixing can also be due to the presence of other inhibitors or interfering substances such as heparin, fibrin split products, and paraproteins.

SPECIFIC FACTOR ASSAYS

Decisions to proceed with specific clotting factor assays will be influenced by the clinical situation and the results of coagulation screening tests. Precise diagnosis and effective management of inherited and acquired coagulation deficiencies necessitate quantitation of the relevant factors. When bleeding is severe, specific assays are often urgently required to guide appropriate therapy. Individual factor assays are usually performed as modifications of the mixing study, where the patient's plasma is mixed with plasma deficient in the factor being studied. This will correct all factor deficiencies to >50%, thus making prolongation of clot formation due to a factor deficiency dependent on the factor missing from the added plasma.

TESTING FOR ANTIPHOSPHOLIPID ANTIBODIES

Antibodies to phospholipids (cardiolipin) or phospholipid-binding proteins (β_2-microglobulin and others) are detected by ELISA. When these antibodies interfere with phospholipid-dependent coagulation tests, they are termed *lupus anticoagulants*. The aPTT has variability sensitivity to lupus anticoagulants, depending in part on the aPTT reagents used. An assay utilizing a sensitive reagent has been termed an *LA-PTT*. The dilute Russell viper venom test (dRVVT) and the tissue thromboplastin inhibition (TTI) test are modifications of standard tests with the phospholipid reagent decreased, thus increasing the sensitivity to antibodies that interfere with the phospholipid component. The tests, however, are not specific for lupus anticoagulants, as factor deficiencies or other inhibitors will also result in prolongation. Documentation of a lupus anticoagulant requires not only prolongation of a phospholipid-dependent coagulation test but also lack of correction when mixed with normal plasma and correction with the addition of activated platelet membranes or certain phospholipids, e.g. hexagonal phase.

OTHER COAGULATION TESTS

The thrombin time and the reptilase time measure fibrinogen conversion to fibrin and are prolonged when the fibrinogen level is low (usually <80–100 mg/dL), qualitatively abnormal, as seen in inherited or acquired dysfibrinogenemias, or when fibrin/fibrinogen degradation products interfere. The thrombin time, but not the reptilase time, is prolonged in the presence of heparin. Measurement of anti–factor Xa plasma inhibitory activity is a test frequently used to assess low-molecular-weight heparin (LMWH) levels or as a direct measurement of unfractionated heparin (UFH) activity. Heparin in the patient sample inhibits the enzymatic conversion of a Xa-specific chromogenic substrate to colored product by factor Xa. Standard curves are created using multiple concentrations of UFH and LMWH and are used to calculate the concentration of anti-Xa activity in the patient plasma.

LABORATORY TESTING FOR THROMBOPHILIA.

Laboratory assays to detect thrombophilic states include molecular diagnostics and immunologic and functional assays. These assays vary in their sensitivity and specificity for the condition being tested. Furthermore, acute thrombosis, acute illnesses, inflammatory conditions, pregnancy, and medications affect levels of many coagulation factors and their inhibitors. Antithrombin is decreased by heparin and in the setting of acute thrombosis. Protein C and S levels may be increased in the setting of acute thrombosis and are decreased by warfarin. Antiphospholipid antibodies are frequently transiently positive in acute illness. Testing for genetic thrombophilias should, in general, only be performed when there is a strong family history of thrombosis and results would affect clinical decision making.

As thrombophilia evaluations are usually performed to assess the need to extend anticoagulation, testing should be performed in a steady state, remote from the acute event. In most instances, warfarin anticoagulation can be stopped after the initial 3–6 months of treatment, and testing performed at least 3 weeks later. Sensitive markers of coagulation activation, notably the D-dimer assay and the thrombin generation test, hold promise as predictors, when elevated, of recurrent thrombosis when measured at least 1 month from discontinuation of warfarin.

MEASURES OF PLATELET FUNCTION.

The bleeding time has been used to assess bleeding risk; however, it has not been found to predict bleeding risk with surgery, and it is not recommended for use for this indication. The PFA-100 and similar instruments that measure platelet-dependent coagulation under flow conditions are generally more sensitive and specific for platelet disorders and VWD than the bleeding time; however, data are

insufficient to support their use to predict bleeding risk or monitor response to therapy. When they are used in the evaluation of a patient with bleeding symptoms, abnormal results, as with the bleeding time, require specific testing, such as VWF assays and/or platelet aggregation studies. Since all of these "screening" assays may miss patients with mild bleeding disorders, further studies are needed to define their role in hemostasis testing.

For classic platelet aggregometry, various agonists are added to the patient's platelet-rich plasma and platelet aggregation is observed. Tests of platelet secretion in response to agonists can also be measured. These tests are affected by many factors, including numerous medications, and the association between minor defects in aggregation or secretion in these assays and bleeding risk is not clearly established.

ACKNOWLEDGMENT

Robert I. Handin, MD, contributed this chapter in the 16th edition and some material from that chapter has been retained here.

FURTHER READINGS

Baglin T et al: Clinical guidelines for testing for heritable thrombophilia. Br J Haematol 149:209, 2010[PMID: 20128794]

Cosmi B et al: Usefulness of repeated D-dimer testing after stopping anticoagulation for a first episode of unprovoked venous thromboembolism: The PROLONG II prospective study. Blood 115:481, 2010[PMID: 19965693]

Galli M: The antiphospholipid triangle. J Thromb Haemost 8:234, 2010[PMID: 19912514]

Lijfering WM et al: Risk factors for venous thrombosis—current understanding from an epidemiological point of view. Br J Haematol 149:824, 2010[PMID: 20456358]

Pengo V et al: Antiphospholipid syndrome: Critical analysis of the diagnostic path. Lupus 19:428, 2010[PMID: 20353982]

Wagenman BL et al: The laboratory approach to inherited and acquired coagulation factor deficiencies. Clin Lab Med 29:229, 2009[PMID: 19665676]

Wheeler AP, Rice TW: Coagulopathy in critically ill patients: Soluble clotting factors and hemostatic testing. Chest 137:185, 2010[PMID: 20051403]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ENLARGEMENT OF LYMPH NODES AND SPLEEN: INTRODUCTION

This chapter is intended to serve as a guide to the evaluation of patients who present with enlargement of the lymph nodes (*lymphadenopathy*) or the spleen (*splenomegaly*). Lymphadenopathy is a rather common clinical finding in primary care settings, whereas palpable splenomegaly is less so.

LYMPHADENOPATHY

Lymphadenopathy may be an incidental finding in patients being examined for various reasons, or it may be a presenting sign or symptom of the patient's illness. The physician must eventually decide whether the lymphadenopathy is a normal finding or one that requires further study, up to and including biopsy. Soft, flat, submandibular nodes (<1 cm) are often palpable in healthy children and young adults; healthy adults may have palpable inguinal nodes of up to 2 cm, which are considered normal. Further evaluation of these normal nodes is not warranted. In contrast, if the physician believes the node(s) to be abnormal, then pursuit of a more precise diagnosis is needed.

Approach to the Patient: Lymphadenopathy

Lymphadenopathy may be a primary or secondary manifestation of numerous disorders, as shown in Table 59–1. Many of these disorders are infrequent causes of lymphadenopathy. In primary care practice, more than two-thirds of patients with lymphadenopathy have nonspecific causes or upper respiratory illnesses (viral or bacterial) and <1% have a malignancy. In one study, 84% of patients referred for evaluation of lymphadenopathy had a "benign" diagnosis. The remaining 16% had a malignancy (lymphoma or metastatic adenocarcinoma). Of the patients with benign lymphadenopathy, 63% had a nonspecific or reactive etiology (no causative agent found), and the remainder had a specific cause demonstrated, most commonly infectious mononucleosis, toxoplasmosis, or tuberculosis. Thus, the vast majority of patients with lymphadenopathy will have a nonspecific etiology requiring few diagnostic tests.

Table 59–1 Diseases Associated with Lymphadenopathy

1. Infectious diseases	
a. Viral—infectious mononucleosis syndromes (EBV, CMV), infectious hepatitis, herpes simplex, herpesvirus-6, varicella-zoster virus, rubella, measles, adenovirus, HIV, epidemic keratoconjunctivitis, vaccinia, herpesvirus-8	
b. Bacterial—streptococci, staphylococci, cat-scratch disease, brucellosis, tularemia, plague, chancroid, melioidosis, glanders, tuberculosis, atypical mycobacterial infection, primary and secondary syphilis, diphtheria, leprosy	
c. Fungal—histoplasmosis, coccidioidomycosis, paracoccidioidomycosis	
d. Chlamydial—lymphogranuloma venereum, trachoma	
e. Parasitic—toxoplasmosis, leishmaniasis, trypanosomiasis, filariasis	
f. Rickettsial—scrub typhus, rickettsialpox, Q fever	2. Immunologic diseases
a. Rheumatoid arthritis	
b. Juvenile rheumatoid arthritis	
c. Mixed connective tissue disease	
d. Systemic lupus erythematosus	
e. Dermatomyositis	
f. Sjögren's syndrome	
g. Serum sickness	
h. Drug hypersensitivity—diphenylhydantoin, hydralazine, allopurinol, primidone, gold, carbamazepine, etc.	
i. Angioimmunoblastic lymphadenopathy	
j. Primary biliary cirrhosis	
k. Graft-vs.-host disease	

l. Silicone-associated
m. Autoimmune lymphoproliferative syndrome
3.Malignant diseases
a. Hematologic—Hodgkin's disease, non-Hodgkin's lymphomas, acute or chronic lymphocytic leukemia, hairy cell leukemia, malignant histiocytosis, amyloidosis
b. Metastatic—from numerous primary sites
4.Lipid storage diseases—Gaucher's, Niemann-Pick, Fabry, Tangier
5. Endocrine diseases—hyperthyroidism
6. Other disorders
a. Castleman's disease (giant lymph node hyperplasia)
b. Sarcoidosis
c. Dermatopathic lymphadenitis
d. Lymphomatoid granulomatosis
e. Histiocytic necrotizing lymphadenitis (Kikuchi's disease)
f. Sinus histiocytosis with massive lymphadenopathy (Rosai-Dorfman disease)
g. Mucocutaneous lymph node syndrome (Kawasaki's disease)
h. Histiocytosis X
i. Familial Mediterranean fever
j. Severe hypertriglyceridemia
k. Vascular transformation of sinuses
l. Inflammatory pseudotumor of lymph node
m. Congestive heart failureAbbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus.

CLINICAL ASSESSMENT

The physician will be aided in the pursuit of an explanation for the lymphadenopathy by a careful medical history, physical examination, selected laboratory tests, and perhaps an excisional lymph node biopsy.

The *medical history* should reveal the setting in which lymphadenopathy is occurring. Symptoms such as sore throat, cough, fever, night sweats, fatigue, weight loss, or pain in the nodes should be sought. The patient's age, sex, occupation, exposure to pets, sexual behavior, and use of drugs such as diphenylhydantoin are other important historic points. For example, children and young adults usually have benign (i.e., nonmalignant) disorders that account for the observed lymphadenopathy such as viral or bacterial upper respiratory infections; infectious mononucleosis; toxoplasmosis; and, in some countries, tuberculosis. In contrast, after age 50, the incidence of malignant disorders increases and that of benign disorders decreases.

The *physical examination* can provide useful clues such as the extent of lymphadenopathy (localized or generalized), size of nodes, texture, presence or absence of nodal tenderness, signs of inflammation over the node, skin lesions, and splenomegaly. A thorough ear, nose, and throat (ENT) examination is indicated in adult patients with cervical adenopathy and a history of tobacco use. Localized or regional adenopathy implies involvement of a single anatomic area. Generalized adenopathy has been defined as involvement of three or more noncontiguous lymph node areas. Many of the causes of lymphadenopathy (Table 59–1) can produce localized *or* generalized adenopathy, so this distinction is of limited utility in the differential diagnosis. Nevertheless, generalized lymphadenopathy is frequently associated with nonmalignant disorders such as infectious mononucleosis [Epstein-Barr virus (EBV) or cytomegalovirus (CMV)], toxoplasmosis, AIDS, other viral infections, systemic lupus erythematosus (SLE), and mixed connective tissue disease. Acute and chronic lymphocytic leukemias and malignant lymphomas also produce generalized adenopathy in adults.

The site of localized or regional adenopathy may provide a useful clue about the cause. Occipital adenopathy often reflects an infection of the scalp, and preauricular adenopathy accompanies conjunctival infections and cat-scratch disease. The most frequent site of regional adenopathy is the neck, and most of the causes are benign—upper respiratory infections, oral and dental lesions, infectious mononucleosis, or other viral illnesses. The chief malignant causes include metastatic cancer from head and neck, breast, lung, and thyroid primaries. Enlargement of supraclavicular and scalene nodes is always abnormal. Because these nodes drain regions of the lung and retroperitoneal space, they can reflect lymphomas, other cancers, or infectious processes arising in these areas. Virchow's node is an enlarged left supraclavicular node infiltrated with metastatic cancer from a gastrointestinal primary. Metastases to supraclavicular nodes also occur from lung, breast, testis, or ovarian cancers. Tuberculosis, sarcoidosis, and toxoplasmosis are nonneoplastic causes of

supraclavicular adenopathy. Axillary adenopathy is usually due to injuries or localized infections of the ipsilateral upper extremity. Malignant causes include melanoma or lymphoma and, in women, breast cancer. Inguinal lymphadenopathy is usually secondary to infections or trauma of the lower extremities and may accompany sexually transmitted diseases such as lymphogranuloma venereum, primary syphilis, genital herpes, or chancroid. These nodes may also be involved by lymphomas and metastatic cancer from primary lesions of the rectum, genitalia, or lower extremities (melanoma).

The size and texture of the lymph node(s) and the presence of pain are useful parameters in evaluating a patient with lymphadenopathy. Nodes <1.0 cm^2 in area (1.0 cm x 1.0 cm or less) are almost always secondary to benign, nonspecific reactive causes. In one retrospective analysis of younger patients (9–25 years) who had a lymph node biopsy, a maximum diameter of >2 cm served as one discriminant for predicting that the biopsy would reveal malignant or granulomatous disease. Another study showed that a lymph node size of 2.25 cm^2 (1.5 cm x 1.5 cm) was the best size limit for distinguishing malignant or granulomatous lymphadenopathy from other causes of lymphadenopathy. Patients with node(s) ≤1.0 cm^2 should be observed after excluding infectious mononucleosis and/or toxoplasmosis unless there are symptoms and signs of an underlying systemic illness.

The texture of lymph nodes may be described as soft, firm, rubbery, hard, discrete, matted, tender, movable, or fixed. Tenderness is found when the capsule is stretched during rapid enlargement, usually secondary to an inflammatory process. Some malignant diseases such as acute leukemia may produce rapid enlargement and pain in the nodes. Nodes involved by lymphoma tend to be large, discrete, symmetric, rubbery, firm, mobile, and nontender. Nodes containing metastatic cancer are often hard, nontender, and nonmovable because of fixation to surrounding tissues. The coexistence of splenomegaly in the patient with lymphadenopathy implies a systemic illness such as infectious mononucleosis, lymphoma, acute or chronic leukemia, SLE, sarcoidosis, toxoplasmosis, cat-scratch disease, or other less common hematologic disorders. The patient's story should provide helpful clues about the underlying systemic illness.

Nonsuperficial presentations (thoracic or abdominal) of adenopathy are usually detected as the result of a symptom-directed diagnostic workup. Thoracic adenopathy may be detected by routine chest radiography or during the workup for superficial adenopathy. It may also be found because the patient complains of a cough or wheezing from airway compression; hoarseness from recurrent laryngeal nerve involvement; dysphagia from esophageal compression; or swelling of the neck, face, or arms secondary to compression of the superior vena cava or subclavian vein. The differential diagnosis of mediastinal and hilar adenopathy includes primary lung disorders and systemic illnesses that characteristically involve mediastinal or hilar nodes. In the young, mediastinal adenopathy is associated with infectious mononucleosis and sarcoidosis. In endemic regions, histoplasmosis can cause unilateral paratracheal lymph node involvement that mimics lymphoma. Tuberculosis can also cause unilateral adenopathy. In older patients, the differential diagnosis includes primary lung cancer (especially among smokers), lymphomas, metastatic carcinoma (usually lung), tuberculosis, fungal infection, and sarcoidosis.

Enlarged intraabdominal or retroperitoneal nodes are usually malignant. Although tuberculosis may present as mesenteric lymphadenitis, these masses usually contain lymphomas or, in young men, germ cell tumors.

LABORATORY INVESTIGATION

The laboratory investigation of patients with lymphadenopathy must be tailored to elucidate the etiology suspected from the patient's history and physical findings. One study from a family practice clinic evaluated 249 younger patients with "enlarged lymph nodes, not infected" or "lymphadenitis." No laboratory studies were obtained in 51%. When studies were performed, the most common were a complete blood count (CBC) (33%), throat culture (16%), chest x-ray (12%), or monospot test (10%). Only eight patients (3%) had a node biopsy, and half of those were normal or reactive. The CBC can provide useful data for the diagnosis of acute or chronic leukemias, EBV or CMV mononucleosis, lymphoma with a leukemic component, pyogenic infections, or immune cytopenias in illnesses such as SLE. Serologic studies may demonstrate antibodies specific to components of EBV, CMV, HIV, and other viruses; *Toxoplasma gondii*; *Brucella*; etc. If SLE is suspected, antinuclear and anti-DNA antibody studies are warranted.

The chest x-ray is usually negative, but the presence of a pulmonary infiltrate or mediastinal lymphadenopathy would suggest tuberculosis, histoplasmosis, sarcoidosis, lymphoma, primary lung cancer, or metastatic cancer and demands further investigation.

A variety of imaging techniques (CT, MRI, ultrasound, color Doppler ultrasonography) have been employed to differentiate benign from malignant lymph nodes, especially in patients with head and neck cancer. CT and MRI are comparably accurate (65–90%) in the diagnosis of metastases to cervical lymph nodes. Ultrasonography has been used to determine the long (L) axis, short (S) axis, and a ratio of long to short axis in cervical nodes. An L/S ratio of <2.0 has a sensitivity and a specificity of 95% for distinguishing benign and malignant nodes in patients with head and neck cancer. This ratio has greater specificity and sensitivity than palpation or measurement of either the long or the

short axis alone.

The indications for lymph node biopsy are imprecise, yet it is a valuable diagnostic tool. The decision to biopsy may be made early in a patient's evaluation or delayed for up to two weeks. Prompt biopsy should occur if the patient's history and physical findings suggest a malignancy; examples include a solitary, hard, nontender cervical node in an older patient who is a chronic user of tobacco; supraclavicular adenopathy; and solitary or generalized adenopathy that is firm, movable, and suggestive of lymphoma. If a primary head and neck cancer is suspected as the basis of a solitary, hard cervical node, then a careful ENT examination should be performed. Any mucosal lesion that is suspicious for a primary neoplastic process should be biopsied first. If no mucosal lesion is detected, an excisional biopsy of the largest node should be performed. Fine-needle aspiration should not be performed as the first diagnostic procedure. Most diagnoses require more tissue than such aspiration can provide, and it often delays a definitive diagnosis. Fine-needle aspiration should be reserved for thyroid nodules and for confirmation of relapse in patients whose primary diagnosis is known. If the primary physician is uncertain about whether to proceed to biopsy, consultation with a hematologist or medical oncologist should be helpful. In primary care practices, <5% of lymphadenopathy patients will require a biopsy. That percentage will be considerably larger in referral practices, i.e., hematology, oncology, or ENT.

Two groups have reported algorithms that they claim will identify more precisely those lymphadenopathy patients who should have a biopsy. Both reports were retrospective analyses in referral practices. The first study involved patients 9–25 years of age who had a node biopsy performed. Three variables were identified that predicted those young patients with peripheral lymphadenopathy who should undergo biopsy; lymph node size >2 cm in diameter and abnormal chest x-ray had positive predictive values, whereas recent ENT symptoms had negative predictive values. The second study evaluated 220 lymphadenopathy patients in a hematology unit and identified five variables [lymph node size, location (supraclavicular or nonsupraclavicular), age (>40 years or <40 years), texture (nonhard or hard), and tenderness] that were used in a mathematical model to identify those patients requiring a biopsy. Positive predictive value was found for age >40 years, supraclavicular location, node size >2.25 cm^2, hard texture, and lack of pain or tenderness. Negative predictive value was evident for age <40 years, node size <1.0 cm^2, nonhard texture, and tender or painful nodes. Ninety-one percent of those who required biopsy were correctly classified by this model. Because both of these studies were retrospective analyses and one was limited to young patients, it is not known how useful these models would be if applied prospectively in a primary care setting.

Most lymphadenopathy patients do not require a biopsy, and at least half require no laboratory studies. If the patient's history and physical findings point to a benign cause for lymphadenopathy, careful follow-up at a 2- to 4-week interval can be employed. The patient should be instructed to return for reevaluation if there is an increase in the size of the nodels). Antibiotics are not indicated for lymphadenopathy unless strong evidence of a bacterial infection is present. Glucocorticoids should not be used to treat lymphadenopathy because their lympholytic effect obscures some diagnoses (lymphoma, leukemia, Castleman's disease) and they contribute to delayed healing or activation of underlying infections. An exception to this statement is the life-threatening pharyngeal obstruction by enlarged lymphoid tissue in Waldeyer's ring that is occasionally seen in infectious mononucleosis.

SPLENOMEGALY

Structure and Function of the Spleen

The spleen is a reticuloendothelial organ that has its embryologic origin in the dorsal mesogastrium at about five weeks' gestation. It arises in a series of hillocks, migrates to its normal adult location in the left upper quadrant (LUQ), and is attached to the stomach via the gastrolienal ligament and to the kidney via the lienorenal ligament. When the hillocks fail to unify into a single tissue mass, accessory spleens may develop in around 20% of persons. The function of the spleen has been elusive. Galen believed it was the source of "black bile" or melancholia, and the word *hypochondria* (literally, beneath the ribs) and the idiom "to vent one's spleen" attest to the beliefs that the spleen had an important influence on the psyche and emotions. In humans, its normal physiologic roles seem to be the following:

1. Maintenance of quality control over erythrocytes in the red pulp by removal of senescent and defective red blood cells. The spleen accomplishes this function through a unique organization of its parenchyma and vasculature **(Fig. 59-1)**.
2. Synthesis of antibodies in the white pulp.
3. The removal of antibody-coated bacteria and antibody-coated blood cells from the circulation.

Figure 59-1

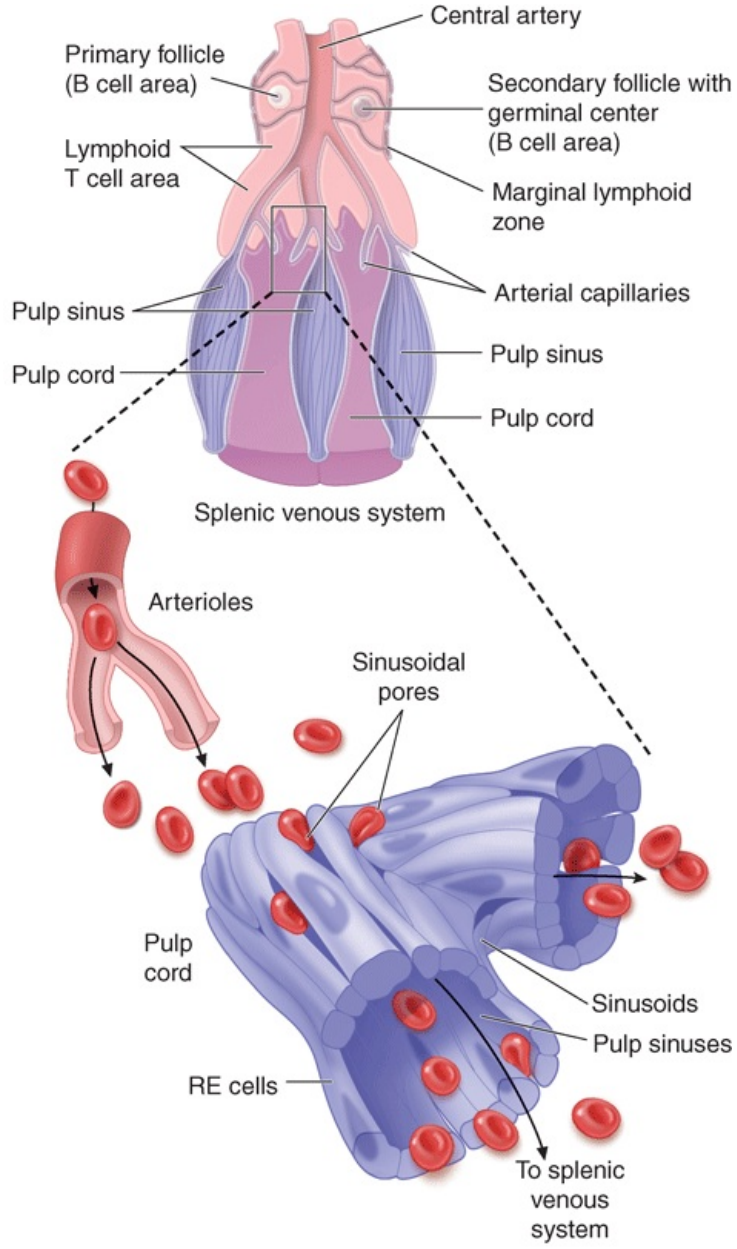

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic spleen structure. The spleen comprises many units of red and white pulp centered around small branches of the splenic artery, called central arteries. White pulp is lymphoid in nature and contains B cell follicles, a marginal zone around the follicles, and T cell–rich areas sheathing arterioles. The red pulp areas include pulp sinuses and pulp cords. The cords are dead ends. In order to regain access to the circulation, red blood cells must traverse tiny openings in the sinusoidal lining. Stiff, damaged, or old red cells cannot enter the sinuses. RE, reticuloendthelial. (*Bottom portion of figure from RS Hillman, KA Ault: Hematology in Clinical Practice, 4th ed. New York, McGraw-Hill, 2005.*)

An increase in these normal functions may result in splenomegaly.

The spleen is composed of *red pulp* and *white pulp*, which are Malpighi's terms for the red blood–filled sinuses and reticuloendothelial cell–lined cords and the white lymphoid follicles arrayed within the red pulp matrix. The spleen is in the portal circulation. The reason for this is unknown but may relate to the fact that lower blood pressure allows less rapid flow and minimizes damage to normal erythrocytes. Blood flows into the spleen at a rate of about 150 mL/min through the splenic artery, which ultimately ramifies into central arterioles. Some blood goes from the arterioles to capillaries and then to splenic veins and out of the spleen, but the majority of blood from central arterioles flows into the macrophage-lined sinuses and cords. The blood entering the sinuses reenters the circulation through the splenic venules, but the blood entering the cords is subjected to an inspection of sorts. To return to the circulation, the blood cells in the cords must squeeze through slits in the cord lining to enter the sinuses that lead to the venules. Old and damaged erythrocytes are less deformable and are retained in the cords, where they are destroyed and their components recycled. Red cell–inclusion bodies such

as parasites (Chaps. 210 and e27), nuclear residua (Howell-Jolly bodies, see Fig. 57-6), or denatured hemoglobin (Heinz bodies) are pinched off in the process of passing through the slits, a process called *pitting*. The culling of dead and damaged cells and the pitting of cells with inclusions appear to occur without significant delay because the blood transit time through the spleen is only slightly slower than in other organs.

The spleen is also capable of assisting the host in adapting to its hostile environment. It has at least three adaptive functions: (1) clearance of bacteria and particulates from the blood, (2) the generation of immune responses to certain pathogens, and (3) the generation of cellular components of the blood under circumstances in which the marrow is unable to meet the needs (i.e., extramedullary hematopoiesis). The latter adaptation is a recapitulation of the blood-forming function the spleen plays during gestation. In some animals, the spleen also serves a role in the vascular adaptation to stress because it stores red blood cells (often hemoconcentrated to higher hematocrits than normal) under normal circumstances and contracts under the influence of β-adrenergic stimulation to provide the animal with an autotransfusion and improved oxygen-carrying capacity. However, the normal human spleen does not sequester or store red blood cells and does not contract in response to sympathetic stimuli. The normal human spleen contains approximately one-third of the total body platelets and a significant number of marginated neutrophils. These sequestered cells are available when needed to respond to bleeding or infection.

Approach to the Patient: Splenomegaly

CLINICAL ASSESSMENT

The most common *symptoms* produced by diseases involving the spleen are pain and a heavy sensation in the LUQ. Massive splenomegaly may cause early satiety. Pain may result from acute swelling of the spleen with stretching of the capsule, infarction, or inflammation of the capsule. For many years, it was believed that splenic infarction was clinically silent, which, at times, is true. However, Soma Weiss, in his classic 1942 report of the self-observations by a Harvard medical student on the clinical course of subacute bacterial endocarditis, documented that severe LUQ and pleuritic chest pain may accompany thromboembolic occlusion of splenic blood flow. Vascular occlusion, with infarction and pain, is commonly seen in children with sickle cell crises. Rupture of the spleen, from either trauma or infiltrative disease that breaks the capsule, may result in intraperitoneal bleeding, shock, and death. The rupture itself may be painless.

A palpable spleen is the major *physical sign* produced by diseases affecting the spleen and suggests enlargement of the organ. The normal spleen weighs <250 g, decreases in size with age, normally lies entirely within the rib cage, has a maximum cephalocaudad diameter of 13 cm by ultrasonography or maximum length of 12 cm and/or width of 7 cm by radionuclide scan, and is usually not palpable. However, a palpable spleen was found in 3% of 2200 asymptomatic, male, freshman college students. Follow-up at 3 years revealed that 30% of those students still had a palpable spleen without any increase in disease prevalence. Ten-year follow-up found no evidence for lymphoid malignancies. Furthermore, in some tropical countries (e.g., New Guinea), the incidence of splenomegaly may reach 60%. Thus, the presence of a palpable spleen does not always equate with presence of disease. Even when disease is present, splenomegaly may not reflect the primary disease but rather a reaction to it. For example, in patients with Hodgkin's disease, only two-thirds of the palpable spleens show involvement by the cancer.

Physical examination of the spleen uses primarily the techniques of palpation and percussion. Inspection may reveal fullness in the LUQ that descends on inspiration, a finding associated with a massively enlarged spleen. Auscultation may reveal a venous hum or friction rub.

Palpation can be accomplished by bimanual palpation, ballotment, and palpation from above (Middleton maneuver). For bimanual palpation, which is at least as reliable as the other techniques, the patient is supine with flexed knees. The examiner's left hand is placed on the lower rib cage and pulls the skin toward the costal margin, allowing the fingertips of the right hand to feel the tip of the spleen as it descends while the patient inspires slowly, smoothly, and deeply. Palpation is begun with the right hand in the left lower quadrant with gradual movement toward the left costal margin, thereby identifying the lower edge of a massively enlarged spleen. When the spleen tip is felt, the finding is recorded as centimeters below the left costal margin at some arbitrary point, i.e., 10–15 cm, from the midpoint of the umbilicus or the xiphisternal junction. This allows other examiners to compare findings or the initial examiner to determine changes in size over time. Bimanual palpation in the right lateral decubitus position adds nothing to the supine examination.

Percussion for splenic dullness is accomplished with any of three techniques described by Nixon, Castell, or Barkun:

1. *1. Nixon's method*: The patient is placed on the right side so that the spleen lies above the colon and stomach. Percussion begins at the lower level of pulmonary resonance in the posterior axillary line and proceeds diagonally along a perpendicular line toward the lower midanterior costal margin. The upper border of dullness is normally 6–8 cm above the costal margin. Dullness >8 cm in an adult is presumed to indicate splenic enlargement.

2. *2. Castell's method*: With the patient supine, percussion in the lowest intercostal space in the anterior axillary line (8th or 9th) produces a resonant note if the spleen is normal in size. This is true during expiration or full inspiration. A dull percussion note on full inspiration suggests splenomegaly.
3. *3. Percussion of Traube's semilunar space* : The borders of Traube's space are the sixth rib superiorly, the left midaxillary line laterally, and the left costal margin inferiorly. The patient is supine with the left arm slightly abducted. During normal breathing, this space is percussed from medial to lateral margins, yielding a normal resonant sound. A dull percussion note suggests splenomegaly.

Studies comparing methods of percussion and palpation with a standard of ultrasonography or scintigraphy have revealed sensitivity of 56–71% for palpation and 59–82% for percussion. Reproducibility among examiners is better for palpation than percussion. Both techniques are less reliable in obese patients or patients who have just eaten. Thus, the physical examination techniques of palpation and percussion are imprecise at best. It has been suggested that the examiner perform percussion first and, if positive, proceed to palpation; if the spleen is palpable, then one can be reasonably confident that splenomegaly exists. However, not all LUQ masses are enlarged spleens; gastric or colon tumors and pancreatic or renal cysts or tumors can mimic splenomegaly.

The presence of an enlarged spleen can be more precisely determined, if necessary, by liver-spleen radionuclide scan, CT, MRI, or ultrasonography. The latter technique is the current procedure of choice for routine assessment of spleen size (normal = a maximum cephalocaudad diameter of 13 cm) because it has high sensitivity and specificity and is safe, noninvasive, quick, mobile, and less costly. Nuclear medicine scans are accurate, sensitive, and reliable but are costly, require greater time to generate data, and use immobile equipment. They have the advantage of demonstrating accessory splenic tissue. CT and MRI provide accurate determination of spleen size, but the equipment is immobile and the procedures are expensive. MRI appears to offer no advantage over CT. Changes in spleen structure such as mass lesions, infarcts, inhomogeneous infiltrates, and cysts are more readily assessed by CT, MRI, or ultrasonography. None of these techniques is very reliable in the detection of patchy infiltration (e.g., Hodgkin's disease).

DIFFERENTIAL DIAGNOSIS

Many of the diseases associated with splenomegaly are listed in Table 59–2. They are grouped according to the presumed basic mechanisms responsible for organ enlargement:

Table 59–2 Diseases Associated with Splenomegaly Grouped by Pathogenic Mechanism

Enlargement Due to Increased Demand for Splenic Function
Reticuloendothelial system hyperplasia (for removal of defective erythrocytes)
Spherocytosis
Early sickle cell anemia
Ovalocytosis
Thalassemia major
Hemoglobinopathies
Paroxysmal nocturnal hemoglobinuria
Pernicious anemia
Immune hyperplasia
Response to infection (viral, bacterial, fungal, parasitic)
Infectious mononucleosis
AIDS
Viral hepatitis
Cytomegalovirus
Subacute bacterial endocarditis
Bacterial septicemia
Congenital syphilis
Splenic abscess
Tuberculosis
Histoplasmosis
Malaria
Leishmaniasis
Trypanosomiasis
Ehrlichiosis

Disordered immunoregulation
Rheumatoid arthritis (Felty's syndrome)
Systemic lupus erythematosus
Collagen vascular diseases
Serum sickness
Immune hemolytic anemias
Immune thrombocytopenias
Immune neutropenias
Drug reactions
Angioimmunoblastic lymphadenopathy
Sarcoidosis
Thyrotoxicosis (benign lymphoid hypertrophy)
Interleukin 2 therapy
Extramedullary hematopoiesis
Myelofibrosis
Marrow damage by toxins, radiation, strontium
Marrow infiltration by tumors, leukemias, Gaucher's disease
Enlargement Due to Abnormal Splenic or Portal Blood Flow
Cirrhosis
Hepatic vein obstruction
Portal vein obstruction, intrahepatic or extrahepatic
Cavernous transformation of the portal vein
Splenic vein obstructionSplenic artery aneurysm
Hepatic schistosomiasis
Congestive heart failure
Hepatic echinococcosis
Portal hypertension (any cause including the above): "Banti's disease"
Infiltration of the Spleen
Intracellular or extracellular depositions
Amyloidosis
Gaucher's disease
Niemann-Pick disease
Tangier disease
Hurler's syndrome and other mucopolysaccharidoses
Hyperlipidemias
Benign and malignant cellular infiltrations
Leukemias (acute, chronic, lymphoid, myeloid, monocytic)
Lymphomas
Hodgkin's disease
Myeloproliferative syndromes (e.g., polycythemia vera, essential thrombocytosis)
Angiosarcomas
Metastatic tumors (melanoma is most common)
Eosinophilic granuloma
Histiocytosis X
Hamartomas
Hemangiomas, fibromas, lymphangiomas
Splenic cysts
Unknown Etiology
Idiopathic splenomegaly
Berylliosis
Iron-deficiency anemia

1. Hyperplasia or hypertrophy related to a particular splenic function such as reticuloendothelial

hyperplasia (work hypertrophy) in diseases such as hereditary spherocytosis or thalassemia syndromes that require removal of large numbers of defective red blood cells; immune hyperplasia in response to systemic infection (infectious mononucleosis, subacute bacterial endocarditis) or to immunologic diseases (immune thrombocytopenia, SLE, Felty's syndrome).

2. Passive congestion due to decreased blood flow from the spleen in conditions that produce portal hypertension (cirrhosis, Budd-Chiari syndrome, congestive heart failure).
3. Infiltrative diseases of the spleen (lymphomas, metastatic cancer, amyloidosis, Gaucher's disease, myeloproliferative disorders with extramedullary hematopoiesis).

The differential diagnostic possibilities are much fewer when the spleen is "massively enlarged," palpable more than 8 cm below the left costal margin or its drained weight is ≥1000 g (Table 59–3). The vast majority of such patients will have non-Hodgkin's lymphoma, chronic lymphocytic leukemia, hairy cell leukemia, chronic myeloid leukemia, myelofibrosis with myeloid metaplasia, or polycythemia vera.

Table 59–3 Diseases Associated with Massive Splenomegaly*

Chronic myeloid leukemia
Lymphomas
Hairy cell leukemia
Myelofibrosis with myeloid
metaplasia
Polycythemia vera
Gaucher's disease
Chronic lymphocytic leukemia
Sarcoidosis
Autoimmune hemolytic anemia
Diffuse splenic hemangiomatosis

* The spleen extends greater than 8 cm below left costal margin and/or weighs more than 1000 g.

LABORATORY ASSESSMENT

The major laboratory abnormalities accompanying splenomegaly are determined by the underlying systemic illness. Erythrocyte counts may be normal, decreased (thalassemia major syndromes, SLE, cirrhosis with portal hypertension), or increased (polycythemia vera). Granulocyte counts may be normal, decreased (Felty's syndrome, congestive splenomegaly, leukemias), or increased (infections or inflammatory disease, myeloproliferative disorders). Similarly, the platelet count may be normal, decreased when there is enhanced sequestration or destruction of platelets in an enlarged spleen (congestive splenomegaly, Gaucher's disease, immune thrombocytopenia), or increased in the myeloproliferative disorders such as polycythemia vera.

The CBC may reveal cytopenia of one or more blood cell types, which should suggest *hypersplenism*. This condition is characterized by splenomegaly, cytopenia(s), normal or hyperplastic bone marrow, and a response to splenectomy. The latter characteristic is less precise because reversal of cytopenia, particularly granulocytopenia, is sometimes not sustained after splenectomy. The cytopenias result from increased destruction of the cellular elements secondary to reduced flow of blood through enlarged and congested cords (congestive splenomegaly) or to immune-mediated mechanisms. In hypersplenism, various cell types usually have normal morphology on the peripheral blood smear, although the red cells may be spherocytic due to loss of surface area during their longer transit through the enlarged spleen. The increased marrow production of red cells should be reflected as an increased reticulocyte production index, although the value may be less than expected due to increased sequestration of reticulocytes in the spleen.

The need for additional laboratory studies is dictated by the differential diagnosis of the underlying illness of which splenomegaly is a manifestation.

Splenectomy

Splenectomy is infrequently performed for diagnostic purposes, especially in the absence of clinical illness or other diagnostic tests that suggest underlying disease. More often, splenectomy is performed for symptom control in patients with massive splenomegaly, for disease control in patients with traumatic splenic rupture, or for correction of cytopenias in patients with hypersplenism or immune-mediated destruction of one or more cellular blood elements. Splenectomy is necessary for staging of patients with Hodgkin's disease only in those with clinical stage I or II disease in whom radiation

therapy alone is contemplated as the treatment. Noninvasive staging of the spleen in Hodgkin's disease is not a sufficiently reliable basis for treatment decisions because one-third of normal-sized spleens will be involved with Hodgkin's disease and one-third of enlarged spleens will be tumor-free. The widespread use of systemic therapy to test all stages of Hodgkin's disease has made staging laparotomy with splenectomy unnecessary. Although splenectomy in chronic myeloid leukemia (CMC) does not affect the natural history of disease, removal of the massive spleen usually makes patients significantly more comfortable and simplifies their management by significantly reducing transfusion requirements. The improvements in therapy of CML have reduced the need for splenectomy for symptom control. Splenectomy is an effective secondary or tertiary treatment for two chronic B cell leukemias, hairy cell leukemia and prolymphocytic leukemia, and for the very rare splenic mantle cell or marginal zone lymphoma. Splenectomy in these diseases may be associated with significant tumor regression in bone marrow and other sites of disease. Similar regressions of systemic disease have been noted after splenic irradiation in some types of lymphoid tumors, especially chronic lymphocytic leukemia and prolymphocytic leukemia. This has been termed the *abscopal effect*. Such systemic tumor responses to local therapy directed at the spleen suggest that some hormone or growth factor produced by the spleen may affect tumor cell proliferation, but this conjecture is not yet substantiated. A common therapeutic indication for splenectomy is traumatic or iatrogenic splenic rupture. In a fraction of patients with splenic rupture, peritoneal seeding of splenic fragments can lead to *splenosis*—the presence of multiple rests of spleen tissue not connected to the portal circulation. This ectopic spleen tissue may cause pain or gastrointestinal obstruction, as in endometriosis. A large number of hematologic, immunologic, and congestive causes of splenomegaly can lead to destruction of one or more cellular blood elements. In the majority of such cases, splenectomy can correct the cytopenias, particularly anemia and thrombocytopenia. In a large series of patients seen in two tertiary care centers, the indication for splenectomy was diagnostic in 10% of patients, therapeutic in 44%, staging for Hodgkin's disease in 20%, and incidental to another procedure in 26%. Perhaps the only contraindication to splenectomy is the presence of marrow failure, in which the enlarged spleen is the only source of hematopoietic tissue.

The absence of the spleen has minimal long-term effects on the hematologic profile. In the immediate postsplenectomy period, leukocytosis (up to 25,000/μL) and thrombocytosis (up to 1 x 10^6/μL) may develop, but within 2–3 weeks, blood cell counts and survival of each cell lineage are usually normal. The chronic manifestations of splenectomy are marked variation in size and shape of erythrocytes (anisocytosis, poikilocytosis) and the presence of Howell-Jolly bodies (nuclear remnants), Heinz bodies (denatured hemoglobin), basophilic stippling, and an occasional nucleated erythrocyte in the peripheral blood. When such erythrocyte abnormalities appear in a patient whose spleen has not been removed, one should suspect splenic infiltration by tumor that has interfered with its normal culling and pitting function.

The most serious consequence of splenectomy is increased susceptibility to bacterial infections, particularly those with capsules such as *Streptococcus pneumoniae*, *Haemophilus influenzae*, and some gram-negative enteric organisms. Patients under age 20 years are particularly susceptible to overwhelming sepsis with *S. pneumoniae*, and the overall actuarial risk of sepsis in patients who have had their spleens removed is about 7% in 10 years. The case-fatality rate for pneumococcal sepsis in splenectomized patients is 50–80%. About 25% of patients without spleens will develop a serious infection at some time in their life. The frequency is highest within the first three years after splenectomy. About 15% of the infections are polymicrobial, and lung, skin, and blood are the most common sites. No increased risk of viral infection has been noted in patients who have no spleen. The susceptibility to bacterial infections relates to the inability to remove opsonized bacteria from the bloodstream and a defect in making antibodies to T cell–independent antigens such as the polysaccharide components of bacterial capsules. Pneumococcal vaccine should be administered to all patients two weeks before elective splenectomy. The Advisory Committee on Immunization Practices recommends that these patients receive repeat vaccination five years post-splenectomy. Efficacy has not been proven for this group, and the recommendation discounts the possibility that administration of the vaccine may actually lower the titer of specific pneumococcal antibodies. A more effective pneumococcal conjugate vaccine that involves T cells in the response is now available (Prevenar, 7-valent). The vaccine to *Neisseria meningitidis* should also be given to patients in whom elective splenectomy is planned. Although efficacy data for *Haemophilus influenzae* type b vaccine are not available for older children or adults, it may be given to patients who have had a splenectomy.

Splenectomized patients should be educated to consider any unexplained fever as a medical emergency. Prompt medical attention with evaluation and treatment of suspected bacteremia may be life-saving. Routine chemoprophylaxis with oral penicillin can result in the emergence of drug-resistant strains and is not recommended.

In addition to an increased susceptibility to bacterial infections, splenectomized patients are also more susceptible to the parasitic disease babesiosis. The splenectomized patient should avoid areas where the parasite *Babesia* is endemic (e.g., Cape Cod, MA).

Surgical removal of the spleen is an obvious cause of hyposplenism. Patients with sickle cell disease

often suffer from autosplenectomy as a result of splenic destruction by the numerous infarcts associated with sickle cell crises during childhood. Indeed, the presence of a palpable spleen in a patient with sickle cell disease after age five suggests a coexisting hemoglobinopathy, e.g., thalassemia or hemoglobin C. In addition, patients who receive splenic irradiation for a neoplastic or autoimmune disease are also functionally hyposplenic. The term *hyposplenism* is preferred to *asplenism* in referring to the physiologic consequences of splenectomy because asplenia is a rare, specific, and fatal congenital abnormality in which there is a failure of the left side of the coelomic cavity (which includes the splenic anlagen) to develop normally. Infants with asplenia have no spleens, but that is the least of their problems. The right side of the developing embryo is duplicated on the left so there is liver where the spleen should be, there are two right lungs, and the heart comprises two right atria and two right ventricles.

FURTHER READINGS

Barkun AN et al: The bedside assessment of splenic enlargement. Am J Med 91:512, 1991[PMID: 1951414]

Center for Disease Control and Prevention: Recommended adult immunization schedule—United States, 2009. MMWR 57:Q-1, 2009

Facchetti F: Tumors of the spleen. Int J Surg Pathol 18:136S, 2010

Graves SA et al: Does this patient have splenomegaly? JAMA 270:2218, 1993

Kraus MD et al: The spleen as a diagnostic specimen: A review of ten years' experience at two tertiary care institutions. Cancer 91:2001, 2001[PMID: 11391578]

McIntyre OR, Ebaugh FG Jr: Palpable spleens: Ten year follow-up. Ann Intern Med 90:130, 1979[PMID: 420452]

Mikocka-Walus A et al: Management of spleen injuries: The current profile. ANZ J Surg 80:157, 2010[PMID: 20575917]

Pangalis GA et al: Clinical approach to lymphadenopathy. Semin Oncol 20:570, 1993[PMID: 8296196]

Williamson HA Jr: Lymphadenopathy in a family practice: A descriptive study of 240 cases. J Fam Pract 20:449, 1985[PMID: 3989485]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF GRANULOCYTES AND MONOCYTES: INTRODUCTION

Leukocytes, the major cells comprising inflammatory and immune responses, include neutrophils, T and B lymphocytes, natural killer (NK) cells, monocytes, eosinophils, and basophils. These cells have specific functions, such as antibody production by B lymphocytes or destruction of bacteria by neutrophils, but in no single infectious disease is the exact role of the cell types completely established. Thus, whereas neutrophils are classically thought to be critical to host defense against bacteria, they may also play important roles in defense against viral infections.

The blood delivers leukocytes to the various tissues from the bone marrow, where they are produced. Normal blood leukocyte counts are 4.3–10.8 x 10^9/L, with neutrophils representing 45–74% of the cells, bands 0–4%, lymphocytes 16–45%, monocytes 4–10%, eosinophils 0–7%, and basophils 0–2%. Variation among individuals and among different ethnic groups can be substantial, with lower leukocyte numbers for certain African-American ethnic groups. The various leukocytes are derived from a common stem cell in the bone marrow. Three-fourths of the nucleated cells of bone marrow are committed to the production of leukocytes. Leukocyte maturation in the marrow is under the regulatory control of a number of different factors, known as colony-stimulating factors (CSFs) and interleukins (ILs). Because an alteration in the number and type of leukocytes is often associated with disease processes, total white blood cell (WBC) count (cells per μL) and differential counts are informative. This chapter focuses on neutrophils, monocytes, and eosinophils. **Lymphocytes and basophils are discussed in Chaps. 314 and 317, respectively.**

NEUTROPHILS

Maturation

Important events in neutrophil life are summarized in Fig. 60-1. In normal humans, neutrophils are produced only in the bone marrow. The minimum number of stem cells necessary to support hematopoiesis is estimated to be 400–500 at any one time. Human blood monocytes, tissue macrophages, and stromal cells produce CSFs, hormones required for the growth of monocytes and neutrophils in the bone marrow. The hematopoietic system not only produces enough neutrophils (~1.3 x 10^{11} cells per 80-kg person per day) to carry out physiologic functions but also has a large reserve stored in the marrow, which can be mobilized in response to inflammation or infection. An increase in the number of blood neutrophils is called *neutrophilia*, and the presence of immature cells is termed a *shift to the left*. A decrease in the number of blood neutrophils is called *neutropenia*.

Figure 60-1

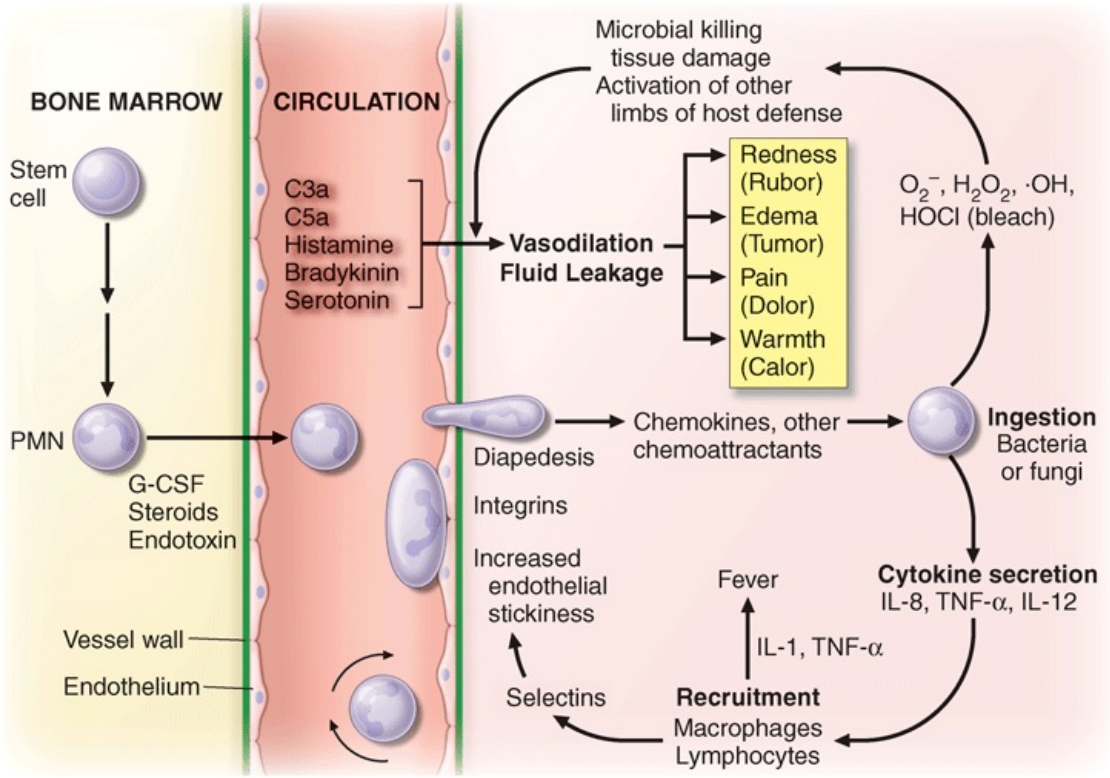

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic events in neutrophil production, recruitment, and inflammation. The four cardinal signs of inflammation (rubor, tumor, calor, dolor) are indicated, as are the interactions of neutrophils with other cells and cytokines. PMN, polymorphonuclear leukocyte; G-CSF, granulocyte colony-stimulating factor; IL, interleukin; TNF-α, tumor necrosis factor α.

Neutrophils and monocytes evolve from pluripotent stem cells under the influence of cytokines and CSFs (Fig. 60-2). The proliferation phase through the metamyelocyte takes about 1 week, while the maturation phase from metamyelocyte to mature neutrophil takes another week. The myeloblast is the first recognizable precursor cell and is followed by the *promyelocyte*. The promyelocyte evolves when the classic lysosomal granules, called the *primary*, or *azurophil, granules* are produced. The primary granules contain hydrolases, elastase, myeloperoxidase, cathepsin G, cationic proteins, and bactericidal/permeability-increasing protein, which is important for killing gram-negative bacteria. Azurophil granules also contain *defensins*, a family of cysteine-rich polypeptides with broad antimicrobial activity against bacteria, fungi, and certain enveloped viruses. The promyelocyte divides to produce the *myelocyte*, a cell responsible for the synthesis of the *specific*, or *secondary, granules*, which contain unique (specific) constituents such as lactoferrin, vitamin B_{12}–binding protein, membrane components of the reduced nicotinamide-adenine dinucleotide phosphate (NADPH) oxidase required for hydrogen peroxide production, histaminase, and receptors for certain chemoattractants and adherence-promoting factors (CR3) as well as receptors for the basement membrane component, laminin. The secondary granules do not contain acid hydrolases and therefore are not classic lysosomes. Packaging of secondary granule contents during myelopoiesis is controlled by CCAAT/enhancer binding protein-ε. Secondary granule contents are readily released extracellularly, and their mobilization is important in modulating inflammation. During the final stages of maturation, no cell division occurs, and the cell passes through the metamyelocyte stage and then to the band neutrophil with a sausage-shaped nucleus (Fig. 60-3). As the band cell matures, the nucleus assumes a lobulated configuration. The nucleus of neutrophils normally contains up to four segments (Fig. 60-4). Excessive segmentation (more than five nuclear lobes) may be a manifestation of folate or vitamin B12 deficiency or the congenital neutropenia syndrome of warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) described below. The Pelger-Hüet anomaly (Fig. 60-5), an infrequent dominant benign inherited trait, results in neutrophils with distinctive bilobed nuclei that must be distinguished from band forms. Acquired bilobed nuclei, pseudo Pelger-Hüet anomaly, can occur with acute infections or in myelodysplastic syndromes. The physiologic role of the normal multilobed nucleus of neutrophils is unknown, but it may allow great deformation of neutrophils during migration into tissues at sites of inflammation.

Figure 60-2

Cell	Stage	Surface Markers[a]	Characteristics
	MYELOBLAST	CD33, CD13, CD15	Prominent nucleoli
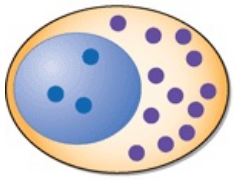	PROMYELOCYTE	CD33, CD13, CD15	Large cell Primary granules appear
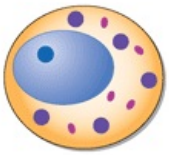	MYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Secondary granules appear
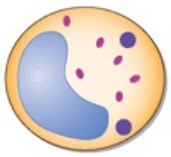	METAMYELOCYTE	CD33, CD13, CD15, CD14, CD11b	Kidney bean–shaped nucleus
	BAND FORM	CD33, CD13, CD15, CD14, CD11b CD10, CD16	Condensed, band–shaped nucleus
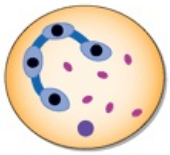	NEUTROPHIL	CD33, CD13, CD15, CD14, CD11b CD10, CD16	Condensed, multilobed nucleus

[a]CD= Cluster Determinant; ● Nucleolus; ● Primary granule; ● Secondary granule.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stages of neutrophil development shown schematically. G-CSF (granulocyte colony-stimulating factor) and GM-CSF (granulocyte-macrophage colony-stimulating factor) are critical to this process. Identifying cellular characteristics and specific cell-surface markers are listed for each maturational stage.

Figure 60-4

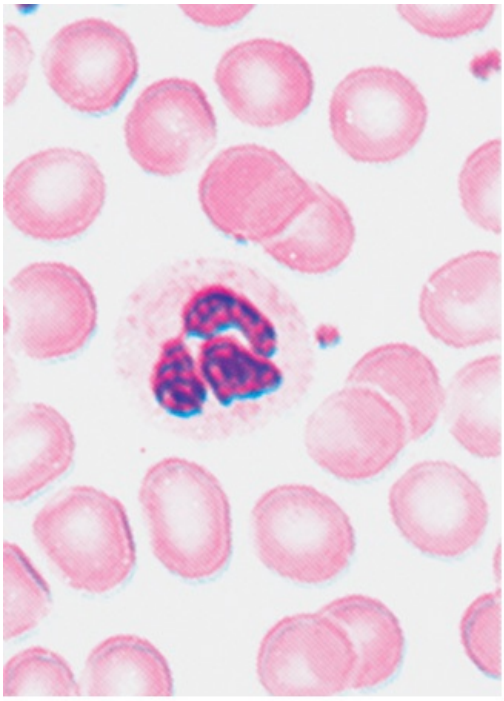

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal granulocyte. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm.

In severe acute bacterial infection, prominent neutrophil cytoplasmic granules, called *toxic granulations*, are occasionally seen. Toxic granulations are immature or abnormally staining azurophil granules. Cytoplasmic inclusions, also called *Döhle bodies* (Fig. 60-3), can be seen during infection and are fragments of ribosome-rich endoplasmic reticulum. Large neutrophil vacuoles are often present in acute bacterial infection and probably represent pinocytosed (internalized) membrane.

Figure 60-3

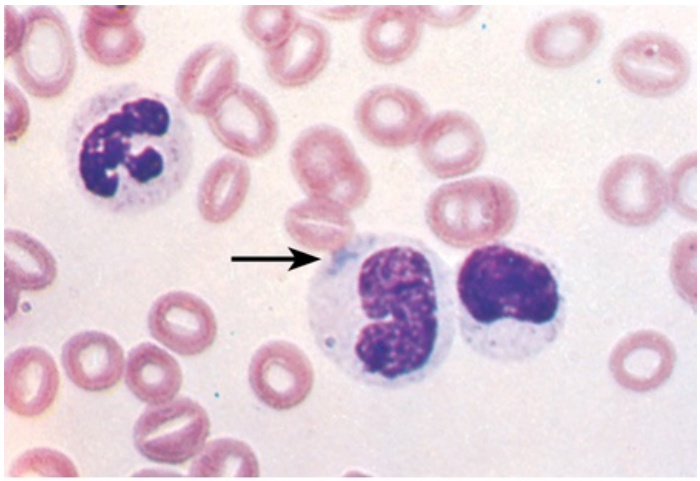

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies are discrete, blue-staining nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum.

Neutrophils are heterogeneous in function. Monoclonal antibodies have been developed that recognize only a subset of mature neutrophils. The meaning of neutrophil heterogeneity is not known.

The morphology of eosinophils and basophils is shown in Fig. 60-6.

Figure 60-5

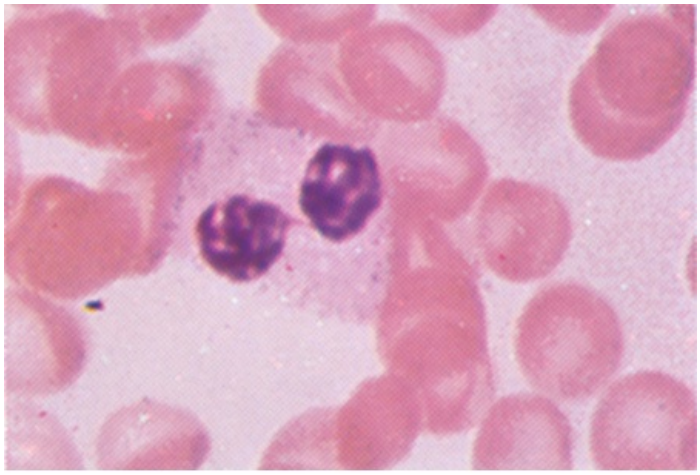

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pelger-Hüet anomaly. In this benign disorder, the majority of granulocytes are bilobed. The nucleus frequently has a spectacle-like, or "pince-nez," configuration.

Marrow Release and Circulating Compartments

Specific signals, including IL-1, tumor necrosis factor α (TNF-α), the CSFs, complement fragments, and chemokines, mobilize leukocytes from the bone marrow and deliver them to the blood in an unstimulated state. Under normal conditions, ~90% of the neutrophil pool is in the bone marrow, 2–3% in the circulation, and the remainder in the tissues (Fig. 60-7).

Figure 60-6

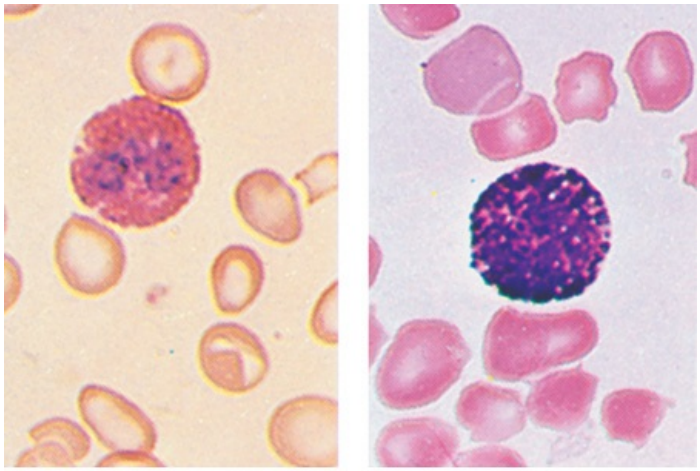

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal eosinophil and basophil. The eosinophil contains large, bright orange granules and usually a bilobed nucleus. The basophil contains large purple-black granules that fill the cell and obscure the nucleus.

Figure 60-7

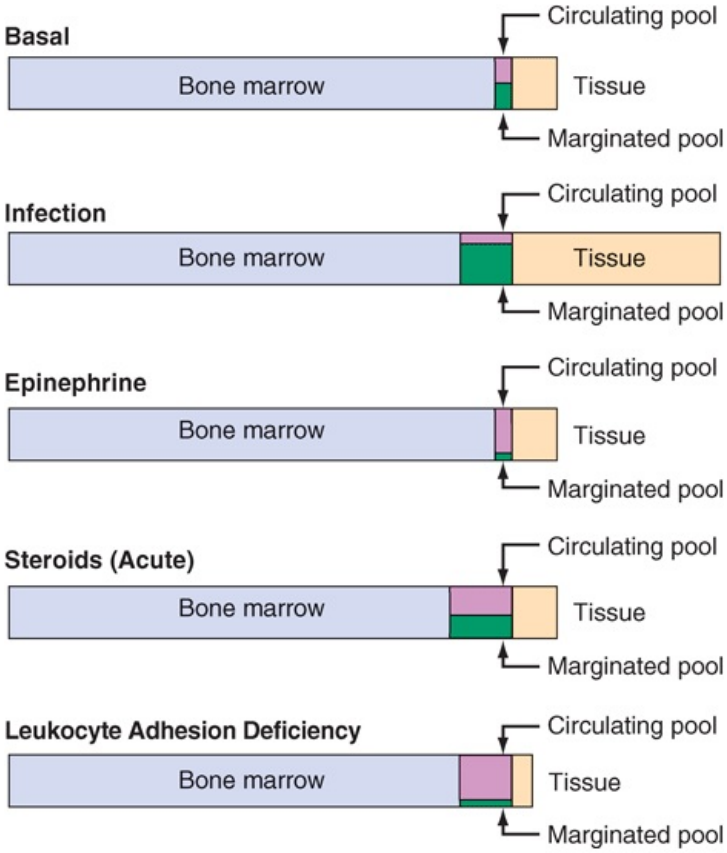

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic neutrophil distribution and kinetics between the different anatomic and functional pools.

The circulating pool exists in two dynamic compartments: one freely flowing and one marginated. The freely flowing pool is about one-half the neutrophils in the basal state and is composed of those cells that are in the blood and not in contact with the endothelium. Marginated leukocytes are those that are in close physical contact with the endothelium (Fig. 60-8). In the pulmonary circulation, where an extensive capillary bed (~1000 capillaries per alveolus) exists, margination occurs because the capillaries are about the same size as a mature neutrophil. Therefore, neutrophil fluidity and deformability are necessary to make the transit through the pulmonary bed. Increased neutrophil rigidity and decreased deformability lead to augmented neutrophil trapping and margination in the lung. In contrast, in the systemic postcapillary venules, margination is mediated by the interaction of specific cell-surface molecules called *selectins*. Selectins are glycoproteins expressed on neutrophils and endothelial cells, among others, that cause a low-affinity interaction, resulting in "rolling" of the neutrophil along the endothelial surface. On neutrophils, the molecule L-selectin [cluster determinant (CD) 62L] binds to glycosylated proteins on endothelial cells [e.g., glycosylation-dependent cell adhesion molecule (GlyCAM1) and CD34]. Glycoproteins on neutrophils, most importantly sialyl-Lewis^x (SLe^x, CD15s), are targets for binding of selectins expressed on endothelial cells [E-selectin (CD62E) and P-selectin (CD62P)] and other leukocytes. In response to chemotactic stimuli from injured tissues (e.g., complement product C5a, leukotriene B_4, IL-8) or bacterial products [e.g., *N*-formylmethionylleucylphenylalanine (f-metleuphe)], neutrophil adhesiveness increases, and the cells "stick" to the endothelium through *integrins*. The integrins are leukocyte glycoproteins that exist as complexes of a common CD18 β chain with CD11a (LFA-1), CD11b (called Mac-1, CR3, or the C3bi receptor), and CD11c (called p150,95 or CR4). CD11a/CD18 and CD11b/CD18 bind to specific endothelial receptors [intercellular adhesion molecules (ICAM) 1 and 2].

Figure 60-8

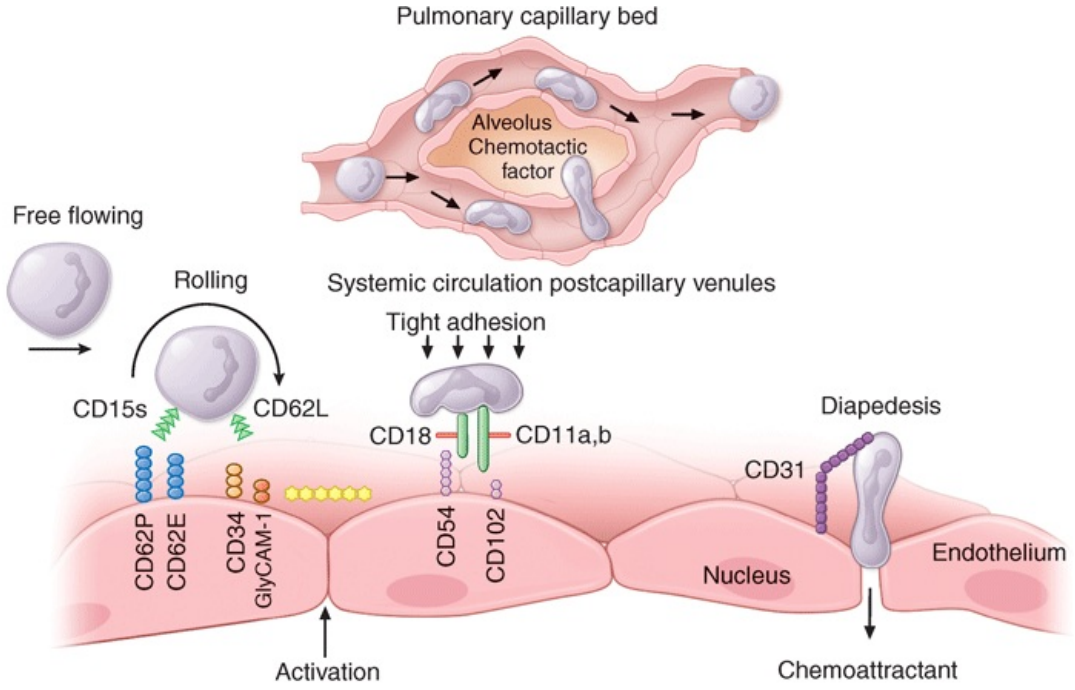

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neutrophil travel through the pulmonary capillaries is dependent on neutrophil deformability. Neutrophil rigidity (e.g., caused by C5a) enhances pulmonary trapping and response to pulmonary pathogens in a way that is not so dependent on cell-surface receptors. Intraalveolar chemotactic factors, such as those caused by certain bacteria (e.g., *Streptococcus pneumoniae*), lead to diapedesis of neutrophils from the pulmonary capillaries into the alveolar space. Neutrophil interaction with the endothelium of the systemic postcapillary venules is dependent on molecules of attachment. The neutrophil "rolls" along the endothelium using selectins: neutrophil CD15s (sialyl-Lewis^x) binds to CD62E (E-selectin) and CD62P (P-selectin) on endothelial cells; CD62L (l-selectin) on neutrophils binds to CD34 and other molecules (e.g., GlyCAM-1) expressed on endothelium. Chemokines or other activation factors stimulate integrin-mediated "tight adhesion": CD11a/CD18 (LFA-1) and CD11b/CD18 (Mac-1, CR3) bind to CD54 (ICAM-1) and CD102 (ICAM-2) on the endothelium. Diapedesis occurs between endothelial cells: CD31 (PECAM-1) expressed by the emigrating neutrophil interacts with CD31 expressed at the endothelial cell-cell junction. CD, cluster determinant; GlyCAM, glycosylation-dependent cell adhesion molecule; ICAM, intercellular adhesion molecule; PECAM, platelet/endothelial cell adhesion molecule.

On cell stimulation, L-selectin is shed from neutrophils, and E-selectin increases in the blood, presumably because it is shed from endothelial cells; receptors for chemoattractants and opsonins are mobilized; and the phagocytes orient toward the chemoattractant source in the extravascular space, increase their motile activity (chemokinesis), and migrate directionally (chemotaxis) into tissues. The process of migration into tissues is called *diapedesis* and involves the crawling of neutrophils between postcapillary endothelial cells that open junctions between adjacent cells to permit leukocyte passage. Diapedesis involves platelet/endothelial cell adhesion molecule (PECAM) 1 (CD31), which is expressed on both the emigrating leukocyte and the endothelial cells. The endothelial responses (increased blood flow from increased vasodilation and permeability) are mediated by anaphylatoxins (e.g., C3a and C5a) as well as vasodilators such as histamine, bradykinin, serotonin, nitric oxide, vascular endothelial growth factor (VEGF), and prostaglandins E and I. Cytokines regulate some of these processes [e.g., TNF-α induction of VEGF, interferon (IFN) γ inhibition of prostaglandin E].

In the healthy adult, most neutrophils leave the body by migration through the mucous membrane of the gastrointestinal tract. Normally, neutrophils spend a short time in the circulation (half-life, 6–7 h). Senescent neutrophils are cleared from the circulation by macrophages in the lung and spleen. Once in the tissues, neutrophils release enzymes, such as collagenase and elastase, which may help establish abscess cavities. Neutrophils ingest pathogenic materials that have been opsonized by IgG and C3b. Fibronectin and the tetrapeptide tuftsin also facilitate phagocytosis.

With phagocytosis comes a burst of oxygen consumption and activation of the hexose-monophosphate shunt. A membrane-associated NADPH oxidase, consisting of membrane and cytosolic components, is assembled and catalyzes the reduction of oxygen to superoxide anion, which is then converted to hydrogen peroxide and other toxic oxygen products (e.g., hydroxyl radical). Hydrogen peroxide + chloride + neutrophil myeloperoxidase generate hypochlorous acid (bleach), hypochlorite, and chlorine. These products oxidize and halogenate microorganisms and tumor cells and, when uncontrolled, can damage host tissue. Strongly cationic proteins, defensins, elastase, cathepsins, and probably nitric oxide also participate in microbial killing. Lactoferrin chelates iron, an important growth

factor for microorganisms, especially fungi. Other enzymes, such as lysozyme and acid proteases, help digest microbial debris. After 1–4 days in tissues, neutrophils die. The apoptosis of neutrophils is also cytokine-regulated; granulocyte colony-stimulating factor (G-CSF) and IFN-γ prolong their life span. Under certain conditions, such as in delayed-type hypersensitivity, monocyte accumulation occurs within 6–12 h of initiation of inflammation. Neutrophils, monocytes, microorganisms in various states of digestion, and altered local tissue cells make up the inflammatory exudate, pus. Myeloperoxidase confers the characteristic green color to pus and may participate in turning off the inflammatory process by inactivating chemoattractants and immobilizing phagocytic cells.

Neutrophils respond to certain cytokines [IFN-γ, granulocyte-macrophage colony-stimulating factor (GM-CSF), IL-8] and produce cytokines and chemotactic signals [TNF-α, IL-8, macrophage inflammatory protein (MIP) 1] that modulate the inflammatory response. In the presence of fibrinogen, f-metleuphe or leukotriene B_4 induces IL-8 production by neutrophils, providing autocrine amplification of inflammation. *Chemokines* (*chemo*attractant cyto*kines*) are small proteins produced by many different cell types, including endothelial cells, fibroblasts, epithelial cells, neutrophils, and monocytes, that regulate neutrophil, monocyte, eosinophil, and lymphocyte recruitment and activation. Chemokines transduce their signals through heterotrimeric G protein–linked receptors that have seven cell membrane–spanning domains, the same type of cell-surface receptor that mediates the response to the classic chemoattractants f-metleuphe and C5a. Four major groups of chemokines are recognized based on the cysteine structure near the N terminus: C, CC, CXC, and CXXXC. The CXC cytokines such as IL-8 mainly attract neutrophils; CC chemokines such as MIP-1 attract lymphocytes, monocytes, eosinophils, and basophils; the C chemokine lymphotactin is T cell tropic; the CXXXC chemokine fractalkine attracts neutrophils, monocytes, and T cells. These molecules and their receptors not only regulate the trafficking and activation of inflammatory cells, but specific chemokine receptors serve as co-receptors for HIV infection (Chap. 189) and have a role in other viral infections such as West Nile infection and atherogenesis.

Neutrophil Abnormalities

Defects in the neutrophil life cycle can lead to dysfunction and compromised host defenses. Inflammation is often depressed, and the clinical result is often recurrent, with severe bacterial and fungal infections. Aphthous ulcers of mucous membranes (gray ulcers without pus) and gingivitis and periodontal disease suggest a phagocytic cell disorder. Patients with congenital phagocyte defects can have infections within the first few days of life. Skin, ear, upper and lower respiratory tract, and bone infections are common. Sepsis and meningitis are rare. In some disorders, the frequency of infection is variable, and patients can go for months or even years without major infection. Aggressive management of these congenital diseases has extended the life span of patients well beyond 30 years.

NEUTROPENIA

The consequences of absent neutrophils are dramatic. Susceptibility to infectious diseases increases sharply when neutrophil counts fall below 1000 cells/μL. When the absolute neutrophil count (ANC; band forms and mature neutrophils combined) falls to <500 cells/μL, control of endogenous microbial flora (e.g., mouth, gut) is impaired; when the ANC is <200/μL, the local inflammatory process is absent. Neutropenia can be due to depressed production, increased peripheral destruction, or excessive peripheral pooling. A falling neutrophil count or a significant decrease in the number of neutrophils below steady-state levels, together with a failure to increase neutrophil counts in the setting of infection or other challenge, requires investigation. Acute neutropenia, such as that caused by cancer chemotherapy, is more likely to be associated with increased risk of infection than neutropenia of long duration (months to years) that reverses in response to infection or carefully controlled administration of endotoxin (see "Laboratory Diagnosis and Management," below).

Some causes of inherited and acquired neutropenia are listed in Table 60–1. The most common neutropenias are iatrogenic, resulting from the use of cytotoxic or immunosuppressive therapies for malignancy or control of autoimmune disorders. These drugs cause neutropenia because they result in decreased production of rapidly growing progenitor (stem) cells of the marrow. Certain antibiotics such as chloramphenicol, trimethoprim-sulfamethoxazole, flucytosine, vidarabine, and the antiretroviral drug zidovudine may cause neutropenia by inhibiting proliferation of myeloid precursors. Azathioprine and 6-mercaptopurine are metabolized by the enzyme thiopurine methyltransferase (TMPT), hypofunctional polymorphisms in which can lead to accumulation of 6-thioguanine and profound marrow toxicity. The marrow suppression is generally dose-related and dependent on continued administration of the drug. Cessation of the offending agent and recombinant human G-CSF usually reverse these forms of neutropenia.

Table 60–1 Causes of Neutropenia

Decreased Production
Drug-induced—alkylating agents (nitrogen mustard, busulfan, chlorambucil, cyclophosphamide); antimetabolites (methotrexate, 6-mercaptopurine, 5-flucytosine); noncytotoxic agents [antibiotics

(chloramphenicol, penicillins, sulfonamides), phenothiazines, tranquilizers (meprobamate), anticonvulsants (carbamazepine), antipsychotics (clozapine), certain diuretics, anti-inflammatory agents, antithyroid drugs, many others]
Hematologic diseases—idiopathic, cyclic neutropenia, Chédiak-Higashi syndrome, aplastic anemia, infantile genetic disorders (see text)
Tumor invasion, myelofibrosis
Nutritional deficiency—vitamin B_{12}, folate (especially alcoholics)
Infection—tuberculosis, typhoid fever, brucellosis, tularemia, measles, infectious mononucleosis, malaria, viral hepatitis, leishmaniasis, AIDS
Peripheral Destruction
Antineutrophil antibodies and/or splenic or lung trapping
Autoimmune disorders—Felty's syndrome, rheumatoid arthritis, lupus erythematosus
Drugs as haptens—aminopyrine, α-methyldopa, phenylbutazone, mercurial diuretics, some phenothiazines
Granulomatosis with polyangiitis (Wegener's)
Peripheral Pooling (Transient Neutropenia)
Overwhelming bacterial infection (acute endotoxemia)
Hemodialysis
Cardiopulmonary bypass

Another important mechanism for iatrogenic neutropenia is the effect of drugs that serve as immune haptens and sensitize neutrophils or neutrophil precursors to immune-mediated peripheral destruction. This form of drug-induced neutropenia can be seen within 7 days of exposure to the drug; with previous drug exposure, resulting in preexisting antibodies, neutropenia may occur a few hours after administration of the drug. Although any drug can cause this form of neutropenia, the most frequent causes are commonly used antibiotics, such as sulfa-containing compounds, penicillins, and cephalosporins. Fever and eosinophilia may also be associated with drug reactions, but often these signs are not present. Drug-induced neutropenia can be severe, but discontinuation of the sensitizing drug is sufficient for recovery, which is usually seen within 5–7 days and is complete by 10 days. Readministration of the sensitizing drug should be avoided, since abrupt neutropenia will often result. For this reason, diagnostic challenge should be avoided.

Autoimmune neutropenias caused by circulating antineutrophil antibodies are another form of acquired neutropenia that results in increased destruction of neutrophils. Acquired neutropenia may also be seen with viral infections, including infection with HIV. Acquired neutropenia may be cyclic in nature, occurring at intervals of several weeks. Acquired cyclic or stable neutropenia may be associated with an expansion of large granular lymphocytes (LGLs), which may be T cells, NK cells, or NK-like cells. Patients with large granular lymphocytosis may have moderate blood and bone marrow lymphocytosis, neutropenia, polyclonal hypergammaglobulinemia, splenomegaly, rheumatoid arthritis, and absence of lymphadenopathy. Such patients may have a chronic and relatively stable course. Recurrent bacterial infections are frequent. Benign and malignant forms of this syndrome occur. In some patients, a spontaneous regression has occurred even after 11 years, suggesting an immunoregulatory defect as the basis for at least one form of the disorder. Glucocorticoids, cyclosporine, and methotrexate are commonly used to manage these cytopenias.

HEREDITARY NEUTROPENIAS

Hereditary neutropenias are rare and may manifest in early childhood as a profound constant neutropenia or agranulocytosis. Congenital forms of neutropenia include Kostmann's syndrome (neutrophil count <100/μL), which is often fatal and due to mutations in the anti-apoptosis gene HAX-1; severe chronic neutropenia (neutrophil count of 300–1500/μL) due to mutations in neutrophil elastase (ELA-2); hereditary cyclic neutropenia, or, more appropriately, cyclic hematopoiesis, also due to mutations in neutrophil elastase (ELA-2); the cartilage-hair hypoplasia syndrome due to mutations in the mitochondrial RNA-processing endoribonuclease RMRP; Shwachman-Diamond syndrome associated with pancreatic insufficiency due to mutations in the Shwachman-Bodian-Diamond syndrome gene *SBDS*; the WHIM [*w*arts, *h*ypogammaglobulinemia, *i*nfections, *m*yelokathexis (retention of WBCs in the marrow)] syndrome, characterized by neutrophil hypersegmentation and bone marrow myeloid arrest due to mutations in the chemokine receptor CXCR4; and neutropenias associated with other immune defects, such as X-linked agammaglobulinemia, Wiskott-Aldrich syndrome, and CD40 ligand deficiency. Mutations in the G-CSF receptor can develop in severe congenital neutropenia and are linked to leukemia. Absence of both myeloid and lymphoid cells is seen in reticular dysgenesis, due to mutations in the nuclear genome-encoded mitochondrial enzyme adenylate kinase-2 (AK2).

Maternal factors can be associated with neutropenia in the newborn. Transplacental transfer of IgG

directed against antigens on fetal neutrophils can result in peripheral destruction. Drugs (e.g., thiazides) ingested during pregnancy can cause neutropenia in the newborn by either depressed production or peripheral destruction.

In Felty's syndrome—the triad of rheumatoid arthritis, splenomegaly, and neutropenia (Chap. 321)—spleen-produced antibodies can shorten neutrophil life span, while LGLs can attack marrow neutrophil precursors. Splenectomy may increase the neutrophil count in Felty's syndrome and lower serum neutrophil-binding IgG. Some Felty's syndrome patients also have neutropenia associated with an increased number of LGLs. Splenomegaly with peripheral trapping and destruction of neutrophils is also seen in lysosomal storage diseases and in portal hypertension.

NEUTROPHILIA

Neutrophilia results from increased neutrophil production, increased marrow release, or defective margination (Table 60–2). The most important acute cause of neutrophilia is infection. Neutrophilia from acute infection represents both increased production and increased marrow release. Increased production is also associated with chronic inflammation and certain myeloproliferative diseases. Increased marrow release and mobilization of the marginated leukocyte pool are induced by glucocorticoids. Release of epinephrine, as with vigorous exercise, excitement, or stress, will demarginate neutrophils in the spleen and lungs and double the neutrophil count in minutes. Cigarette smoking can elevate neutrophil counts above the normal range. Leukocytosis with cell counts of 10,000–25,000/μL occurs in response to infection and other forms of acute inflammation and results from both release of the marginated pool and mobilization of marrow reserves. Persistent neutrophilia with cell counts of ≥30,000–50,000/μL is called a *leukemoid reaction*, a term often used to distinguish this degree of neutrophilia from leukemia. In a leukemoid reaction, the circulating neutrophils are usually mature and not clonally derived.

Table 60–2 Causes of Neutrophilia

Increased Production
Idiopathic
Drug-induced—glucocorticoids, G-CSF
Infection—bacterial, fungal, sometimes viral
Inflammation—thermal injury, tissue necrosis, myocardial and pulmonary infarction, hypersensitivity states, collagen vascular diseases
Myeloproliferative diseases—myelocytic leukemia, myeloid metaplasia, polycythemia vera
Increased Marrow Release
Glucocorticoids Acute infection (endotoxin) nflammation—thermal injury
Decreased or Defective Margination
Drugs—epinephrine, glucocorticoids, nonsteroidal anti-inflammatory agents Stress, excitement, vigorous exercise Leukocyte adhesion deficiency type 1 (CD18); leukocyte adhesion deficiency type 2 (selectin ligand, CD15s); leukocyte adhesion deficiency type 3 (Kindlin-3)
Miscellaneous
Metabolic disorders—ketoacidosis, acute renal failure, eclampsia, acute poisoning
Drugs—lithium
Other—metastatic carcinoma, acute hemorrhage or hemolysis
Abbreviation: G-CSF, granulocyte colony-stimulating factor.

ABNORMAL NEUTROPHIL FUNCTION

Inherited and acquired abnormalities of phagocyte function are listed in Table 60–3. The resulting diseases are best considered in terms of the functional defects of adherence, chemotaxis, and microbicidal activity. The distinguishing features of the important inherited disorders of phagocyte function are shown in Table 60–4.

Table 60–3 Types of Granulocyte and Monocyte Disorders

Cause of Indicated Dysfunction

Function	Drug-Induced	Acquired	Inherited
Adherence-aggregation	Aspirin, colchicine, alcohol, glucocorticoids, ibuprofen, piroxicam	Neonatal state, hemodialysis	Leukocyte adhesion deficiency types 1, 2, and 3
Deformability		Leukemia, neonatal state, diabetes mellitus, immature neutrophils	
Chemokinesis-chemotaxis	Glucocorticoids (high dose), auranofin, colchicine (weak effect), phenylbutazone, naproxen, indomethacin, interleukin 2	Thermal injury, malignancy, malnutrition, periodontal disease, neonatal state, systemic lupus erythematosus, rheumatoid arthritis, diabetes mellitus, sepsis, influenza virus infection, herpes simplex virus infection, acrodermatitis enteropathica, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, hyper IgE–recurrent infection (Job's) syndrome (in some patients), Down syndrome, α-mannosidase deficiency, leukocyte adhesion deficiencies, Wiskott-Aldrich syndrome
Microbicidal activity	Colchicine, cyclophosphamide, glucocorticoids (high dose), TNF-α blocking antibodies	Leukemia, aplastic anemia, certain neutropenias, tuftsin deficiency, thermal injury, sepsis, neonatal state, diabetes mellitus, malnutrition, AIDS	Chédiak-Higashi syndrome, neutrophil-specific granule deficiency, chronic granulomatous disease, defects in IFN- /IL-12 axis

Abbreviations: IFN, interferon; IL, interleukin; TNF-α, tumor necrosis factor alpha.

Table 60–4 Inherited Disorders of Phagocyte Function: Differential Features

Clinical Manifestations	Cellular or Molecular Defects	Diagnosis
Chronic Granulomatous Diseases (70% X-linked, 30% Autosomal Recessive)		
Severe infections of skin, ears, lungs, liver, and bone with catalase-positive microorganisms such as *Staphylococcus aureus, Burkholderia cepacia, Aspergillus* spp., *Chromobacterium violaceum*; often hard to culture organism; excessive inflammation with granulomas, frequent lymph node suppuration; granulomas can obstruct GI or GU tracts; gingivitis, aphthous ulcers, seborrheic dermatitis	No respiratory burst due to the lack of one of five NADPH oxidase subunits in neutrophils, monocytes, and eosinophils	DHR or NBT test; no superoxide and H_2O_2 production by neutrophils; immunoblot for NADPH oxidase components; genetic detection
Chédiak-Higashi Syndrome (Autosomal Recessive)		
Recurrent pyogenic infections, especially with *S. aureus;* many patients get lymphoma-like illness during adolescence; periodontal disease; partial oculocutaneous albinism, nystagmus, progressive peripheral neuropathy, mental retardation in some patients	Reduced chemotaxis and phagolysosome fusion, increased respiratory burst activity, defective egress from marrow, abnormal skin window; defect in *CHS1*	Giant primary granules in neutrophils and other granule-bearing cells (Wright's stain); genetic detection
Specific Granule Deficiency (Autosomal Recessive)		
Recurrent infections of skin, ears, and sinopulmonary tract; delayed wound healing; decreased inflammation; bleeding diathesis	Abnormal chemotaxis, impaired respiratory burst and bacterial killing, failure to upregulate chemotactic and adhesion receptors with stimulation, defect in transcription of granule proteins; defect in C/EBPε	Lack of secondary (specific) granules in neutrophils (Wright's stain), no neutrophil-specific granule contents (i.e., lactoferrin), no defensins, platelet α granule abnormality; genetic detection

Myeloperoxidase Deficiency (Autosomal Recessive)		
Clinically normal except in patients with underlying disease such as diabetes mellitus; then candidiasis or other fungal infections	No myeloperoxidase due to pre- and posttranslational defects in myeloperoxidase deficiency	No peroxidase in neutrophils; genetic detection
Leukocyte Adhesion Deficiency		
Type 1: Delayed separation of umbilical cord, sustained neutrophilia, recurrent infections of skin and mucosa, gingivitis, periodontal disease Type 2: Mental retardation, short stature, Bombay (hh) blood phenotype, recurrent infections, neutrophilia Type 3: Petechial hemorrhage, recurrent infections	Impaired phagocyte adherence, aggregation, spreading, chemotaxis, phagocytosis of C3bi-coated particles; defective production of CD18 subunit common to leukocyte integrins Impaired phagocyte rolling along endothelium due to defects in fucose transporter Impaired signaling for integrin activation resulting in impaired adhesion due to mutation in *FERMT3*	Reduced phagocyte surface expression of the CD18-containing integrins with monoclonal antibodies against LFA-1 (CD18/CD11a), Mac-1 or CR3 (CD18/CD11b), p150,95 (CD18/CD11c); genetic detection Reduced phagocyte surface expression of Sialyl-Lewis^X, with monoclonal antibodies against CD15s; genetic detection Reduced signaling for adhesion through integrins; genetic detection
Phagocyte Activation Defects (X-linked and Autosomal Recessive)		
NEMO deficiency: mild hypohidrotic ectodermal dysplasia; broad-based immune defect: pyogenic and encapsulated bacteria, viruses, *Pneumocystis*, mycobacteria; X-linked IRAK4 and MyD88 deficiency: susceptibility to pyogenic bacteria such as staphylococci, streptococci, clostridia; resistant to candida; autosomal recessive	Impaired phagocyte activation by IL-1, IL-18, TLR, CD40L, TNF-α leading to problems with inflammation and antibody production Impaired phagocyte activation by endotoxin through TLR and other pathways; TNF-α signaling preserved	Poor in vitro response to endotoxin; lack of NF-κB activation; genetic detection Poor in vitro response to endotoxin; lack of NF-κB activation by endotoxin; genetic detection
Hyper IgE–Recurrent Infection Syndrome (Autosomal Dominant) (Job's Syndrome)		
Eczematoid or pruritic dermatitis, "cold" skin abscesses, recurrent pneumonias with *S . aureus* with bronchopleural fistulas and cyst formation, mild eosinophilia, mucocutaneous candidiasis, characteristic facies, restrictive lung disease, scoliosis, delayed primary dental deciduation DOCK8 deficiency (autosomal recessive) Severe eczema, atopic dermatitis, cutaneous abscesses, HSV, HPV, and molluscum infections, severe allergies, cancer	Reduced chemotaxis in some patients, reduced suppressor T cell activity. Mutation in *STAT3* Impaired T cell proliferation to mitogens	Somatic and immune features involving lungs, skeleton, and immune system; serum IgE > 2000 IU/mL; genetic testing Severe allergies, viral infections, high IgE, eosinophilia, low IgM, progressive lymphopenia, genetic detection
Mycobacteria Susceptibility (Autosomal Dominant and Recessive Forms)		
Severe extrapulmonary or disseminated infections with bacille Calmette-Guérin (BCG), nontuberculous mycobacteria, salmonella, histoplasmosis, coccidioidomycosis, poor granuloma formation	Inability to kill intracellular organisms due to low IFN-γ production or response; mutations in IFN-γ receptors, IL-12 receptor, IL-12 p40, STAT1, NEMO	Low or very high levels of IFN-γ receptor 1; functional assays of cytokine production and response; genetic detection

Abbreviations: C/EBPε, CCAAT/enhancer binding protein-ε; DHR, dihydrorhodamine (oxidation test); DOCK8, dedicator of cytokinesis 8; GI, gastrointestinal; GU, genitourinary; HPV, human papilloma virus; HSV, herpes simplex virus; IFN, interferon; IL, interleukin; IRAK4, IL-1 receptor–associated kinase 4; LFA-1, leukocyte function–associated antigen 1; MyD88, myeloid differentiation primary response gene 88; NADPH, nicotinamide–adenine dinueleotide phosphate; NBT, nitroblue tetrazolium (dye test); NEMO, NF-κB essential modulator; NF-κB, nuclear factor κB; STAT1, –3, signal transducer and activator of transcription 1, –3; TLR, Toll-like receptor; TNF, tumor necrosis factor.

Disorders of Adhesion

Three main types of leukocyte adhesion deficiency (LAD) have been described. All are autosomal

recessive and result in the inability of neutrophils to exit the circulation to sites of infection, leading to leukocytosis and increased susceptibility to infection (Fig. 60-8). Patients with LAD 1 have mutations in CD18, the common component of the integrins LFA-1, Mac-1, and p150,95, leading to a defect in tight adhesion between neutrophils and the endothelium. The heterodimer formed by CD18/CD11b (Mac-1) is also the receptor for the complement-derived opsonin C3bi (CR3). The *CD18* gene is located on distal chromosome 21q. The severity of the defect determines the severity of clinical disease. Complete lack of expression of the leukocyte integrins results in a severe phenotype in which inflammatory stimuli do not increase the expression of leukocyte integrins on neutrophils or activated T and B cells. Neutrophils (and monocytes) from patients with LAD 1 adhere poorly to endothelial cells and protein-coated surfaces and exhibit defective spreading, aggregation, and chemotaxis. Patients with LAD 1 have recurrent bacterial infections involving the skin, oral and genital mucosa, and respiratory and intestinal tracts; persistent leukocytosis (resting neutrophil counts of 15,000–20,000/μL) because cells do not marginate; and, in severe cases, a history of delayed separation of the umbilical stump. Infections, especially of the skin, may become necrotic with progressively enlarging borders, slow healing, and development of dysplastic scars. The most common bacteria are *Staphylococcus aureus* and enteric gram-negative bacteria. LAD 2 is caused by an abnormality of fucosylation of SLe^X (CD15s), the ligand on neutrophils that interacts with selectins on endothelial cells and is responsible for neutrophil rolling along the endothelium. Infection susceptibility in LAD 2 appears to be less severe than in LAD 1. LAD 2 is also known as *congenital disorder of glycosylation IIc* (CDGIIc) due to mutation in a GDP-fucose transporter (SLC35C1). LAD 3 is characterized by infection susceptibility, leukocytosis, and petechial hemorrhage due to impaired integrin activation caused by mutations in the gene FERMT3.

Disorders of Neutrophil Granules

The most common neutrophil defect is myeloperoxidase deficiency, a primary granule defect inherited as an autosomal recessive trait; the incidence is ~1 in 2000 persons. Isolated myeloperoxidase deficiency is not associated with clinically compromised defenses, presumably because other defense systems such as hydrogen peroxide generation are amplified. Microbicidal activity of neutrophils is delayed but not absent. Myeloperoxidase deficiency may make other acquired host defense defects more serious. An acquired form of myeloperoxidase deficiency occurs in myelomonocytic leukemia and acute myeloid leukemia.

Chédiak-Higashi syndrome (CHS) is a rare disease with autosomal recessive inheritance due to defects in the lysosomal transport protein LYST, encoded by the gene *CHS1* at 1q42. This protein is required for normal packaging and disbursement of granules. Neutrophils (and all cells containing lysosomes) from patients with CHS characteristically have large granules (Fig. 60-9), making it a systemic disease. Patients with CHS have nystagmus, partial oculocutaneous albinism, and an increased number of infections resulting from many bacterial agents. Some CHS patients develop an "accelerated phase" in childhood with a hemophagocytic syndrome and an aggressive lymphoma requiring bone marrow transplantation. CHS neutrophils and monocytes have impaired chemotaxis and abnormal rates of microbial killing due to slow rates of fusion of the lysosomal granules with phagosomes. NK cell function is also impaired. CHS patients may develop a severe disabling peripheral neuropathy in adulthood that can lead to bed confinement.

Figure 60-9

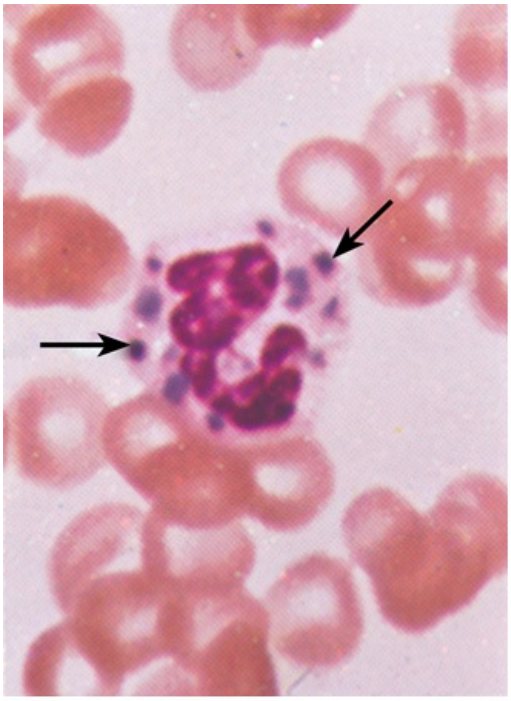

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chédiak-Higashi syndrome. The granulocytes contain huge cytoplasmic granules formed from aggregation and fusion of azurophilic and specific granules. Large abnormal granules are found in other granule-containing cells throughout the body.

Specific granule deficiency is a rare autosomal recessive disease in which the production of secondary granules and their contents, as well as the primary granule component defensins, is defective. The defect in bacterial killing leads to severe bacterial infections. One type of specific granule deficiency is due to a mutation in the CCAAT/enhancer binding protein-ε, a regulator of expression of granule components.

Chronic Granulomatous Disease

Chronic granulomatous disease (CGD) is a group of disorders of granulocyte and monocyte oxidative metabolism. Although CGD is rare, with an incidence of 1 in 200,000 individuals, it is an important model of defective neutrophil oxidative metabolism. Most often CGD is inherited as an X-linked recessive trait; 30% of patients inherit the disease in an autosomal recessive pattern. Mutations in the genes for the five proteins that assemble at the plasma membrane account for all patients with CGD. Two proteins (a 91-kDa protein, abnormal in X-linked CGD, and a 22-kDa protein, absent in one form of autosomal recessive CGD) form the heterodimer cytochrome b-558 in the plasma membrane. Three other proteins (40-, 47-, and 67-kDa, abnormal in the other autosomal recessive forms of CGD) are cytoplasmic in origin and interact with the cytochrome after cell activation to form NADPH oxidase, required for hydrogen peroxide production. Leukocytes from patients with CGD have severely diminished hydrogen peroxide production. The genes involved in each of the defects have been cloned and sequenced and the chromosome locations identified. Patients with CGD characteristically have increased numbers of infections due to catalase-positive microorganisms (organisms that destroy their own hydrogen peroxide). When patients with CGD become infected, they often have extensive inflammatory reactions, and lymph node suppuration is common despite the administration of appropriate antibiotics. Aphthous ulcers and chronic inflammation of the nares are often present. Granulomas are frequent and can obstruct the gastrointestinal or genitourinary tracts. The excessive inflammation reflects failure to downregulate inflammation, reflecting failure to inhibit the synthesis, degradation of or response to chemoattractants or residual antigens, leading to persistent neutrophil accumulation. Impaired killing of intracellular microorganisms by macrophages may lead to persistent cell-mediated immune activation and granuloma formation. Autoimmune complications such as immune thrombocytopenic purpura and juvenile rheumatoid arthritis are also increased in CGD. In addition, discoid lupus is more common in X-linked carriers. Late complications including nodular regenerative hyperplasia and portal hypertension are increasingly recognized in long-term survivors of severe CGD.

Disorders of Phagocyte Activation

Phagocytes depend on cell-surface stimulation to induce signals that evoke multiple levels of the inflammatory response, including cytokine synthesis, chemotaxis, and antigen presentation. Mutations affecting the major pathway that signals through NF-κB have been noted in patients with a variety of infection susceptibility syndromes. If the defects are at a very late stage of signal transduction, in the

protein critical for NF-κB activation known as the NF-κB essential modulator (NEMO), then affected males develop ectodermal dysplasia and severe immune deficiency with susceptibility to bacteria, fungi, mycobacteria, and viruses. If the defects in NF-κB activation are closer to the cell-surface receptors, in the proteins transducing Toll-like receptor signals, IL-1 receptor–associated kinase 4 (IRAK4), and myeloid differentiation primary response gene 88 (MyD88), then children have a marked susceptibility to pyogenic infections early in life but develop resistance to infection later.

MONONUCLEAR PHAGOCYTES

The mononuclear phagocyte system is composed of monoblasts, promonocytes, and monocytes, in addition to the structurally diverse tissue macrophages that make up what was previously referred to as the reticuloendothelial system. Macrophages are long-lived phagocytic cells capable of many of the functions of neutrophils. They are also secretory cells that participate in many immunologic and inflammatory processes distinct from neutrophils. Monocytes leave the circulation by diapedesis more slowly than neutrophils and have a half-life in the blood of 12–24 h.

After blood monocytes arrive in the tissues, they differentiate into macrophages ("big eaters") with specialized functions suited for specific anatomic locations. Macrophages are particularly abundant in capillary walls of the lung, spleen, liver, and bone marrow, where they function to remove microorganisms and other noxious elements from the blood. Alveolar macrophages, liver Kupffer cells, splenic macrophages, peritoneal macrophages, bone marrow macrophages, lymphatic macrophages, brain microglial cells, and dendritic macrophages all have specialized functions. Macrophage-secreted products include lysozyme, neutral proteases, acid hydrolases, arginase, complement components, enzyme inhibitors (plasmin, α_2-macroglobulin), binding proteins (transferrin, fibronectin, transcobalamin II), nucleosides, and cytokines (TNF-α; IL-1, -8, -12, -18). IL-1 (Chaps. 16 and 314) has many functions, including initiating fever in the hypothalamus, mobilizing leukocytes from the bone marrow, and activating lymphocytes and neutrophils. TNF-α is a pyrogen that duplicates many of the actions of IL-1 and plays an important role in the pathogenesis of gram-negative shock (Chap. 271). TNF-α stimulates production of hydrogen peroxide and related toxic oxygen species by macrophages and neutrophils. In addition, TNF-α induces catabolic changes that contribute to the profound wasting (cachexia) associated with many chronic diseases.

Other macrophage-secreted products include reactive oxygen and nitrogen metabolites, bioactive lipids (arachidonic acid metabolites and platelet-activating factors), chemokines, CSFs, and factors stimulating fibroblast and vessel proliferation. Macrophages help regulate the replication of lymphocytes and participate in the killing of tumors, viruses, and certain bacteria (*Mycobacterium tuberculosis* and *Listeria monocytogenes*). Macrophages are key effector cells in the elimination of intracellular microorganisms. Their ability to fuse to form giant cells that coalesce into granulomas in response to some inflammatory stimuli is important in the elimination of intracellular microbes and is under the control of IFN-γ. Nitric oxide induced by IFN-γ is an important effector against intracellular parasites, including tuberculosis and *Leishmania*.

Macrophages play an important role in the immune response (Chap. 314). They process and present antigen to lymphocytes and secrete cytokines that modulate and direct lymphocyte development and function. Macrophages participate in autoimmune phenomena by removing immune complexes and other substances from the circulation. Polymorphisms in macrophage receptors for immunoglobulin (Fc γRII) determine susceptibility to some infections and autoimmune diseases. In wound healing, they dispose of senescent cells, and they contribute to atheroma development. Macrophage elastase mediates development of emphysema from cigarette smoking.

Disorders of the Mononuclear Phagocyte System

Many disorders of neutrophils extend to mononuclear phagocytes. Thus, drugs that suppress neutrophil production in the bone marrow can cause monocytopenia. Transient monocytopenia occurs after stress or glucocorticoid administration. Monocytosis is associated with tuberculosis, brucellosis, subacute bacterial endocarditis, Rocky Mountain spotted fever, malaria, and visceral leishmaniasis (kala azar). Monocytosis also occurs with malignancies, leukemias, myeloproliferative syndromes, hemolytic anemias, chronic idiopathic neutropenias, and granulomatous diseases such as sarcoidosis, regional enteritis, and some collagen vascular diseases. Patients with LAD, hyperimmunoglobulin E –recurrent infection (Job's) syndrome, CHS, and CGD all have defects in the mononuclear phagocyte system.

Monocyte cytokine production or response is impaired in some patients with disseminated nontuberculous mycobacterial infection who are not infected with HIV. Genetic defects in the pathways regulated by IFN-γ and IL-12 lead to impaired killing of intracellular bacteria, mycobacteria, salmonellae, and certain viruses (Fig. 60-10).

Figure 60-10

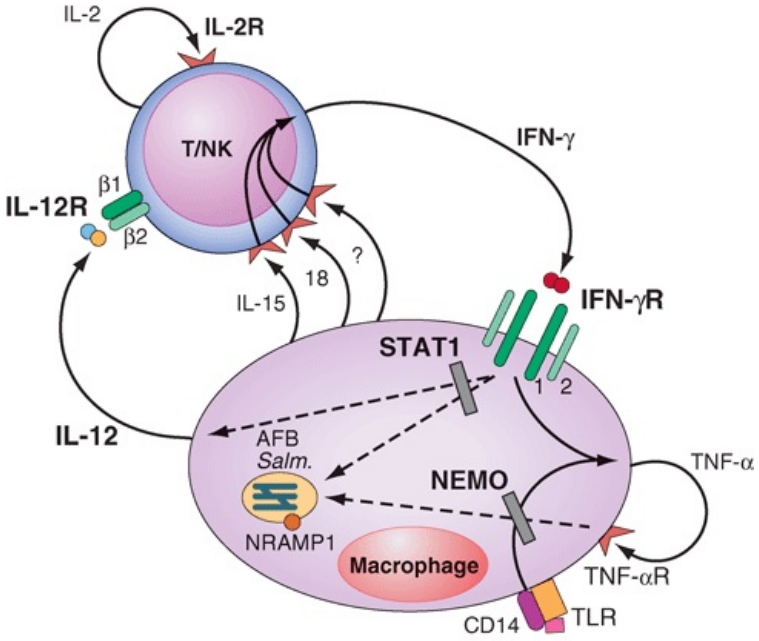

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lymphocyte-macrophage interactions underlying resistance to mycobacteria and other intracellular parasites such as *Salmonella*. Mycobacteria infect macrophages, leading to the production of IL-12, which activates T or NK cells through its receptor, leading to production of IL-2 and IFN-γ. IFN-γ acts through its receptor on macrophages to upregulate TNF-γ and IL-12 and kill intracellular parasites. Mutant forms of the cytokines and receptors shown in large type have been found in severe cases of nontuberculous mycobacterial infection and salmonellosis. AFB, acid fast bacilli; IFN, interferon; IL, interleukin; NEMO, NF-κB essential modulator; NK, natural killer; STAT1, signal transducer and activator of transcription 1; TLR, Toll-like receptor; TNF, tumor necrosis factor.

Certain viral infections impair mononuclear phagocyte function. For example, influenza virus infection causes abnormal monocyte chemotaxis. Mononuclear phagocytes can be infected by HIV using CCR5, the chemokine receptor that acts as a co-receptor with CD4 for HIV. T lymphocytes produce IFN-γ, which induces FcR expression and phagocytosis and stimulates hydrogen peroxide production by mononuclear phagocytes and neutrophils. In certain diseases, such as AIDS, IFN-γ production may be deficient, whereas in other diseases, such as T cell lymphomas, excessive release of IFN-γ may be associated with erythrophagocytosis by splenic macrophages.

Autoinflammatory diseases are characterized by abnormal cytokine regulation, leading to excess inflammation in the absence of infection. These diseases can mimic infectious or immunodeficient syndromes. Gain-of-function mutations in the TNF-α receptor cause TNF-α receptor–associated periodic syndrome (TRAPS), which is characterized by recurrent fever in the absence of infection, due to persistent stimulation of the TNF-α receptor (Chap. 330). Diseases with abnormal IL-1 regulation leading to fever include familial Mediterranean fever due to mutations in *PYRIN*. Mutations in *cold-induced autoinflammatory syndrome 1* (*CIAS1*) lead to neonatal-onset multisystem autoinflammatory disease, familial cold urticaria, and Muckle-Wells syndrome. The syndrome of pyoderma gangrenosum, acne, and sterile pyogenic arthritis (PAPA syndrome) is caused by mutations in CD2BP1. In contrast to these syndromes of overexpression of proinflammatory cytokines, blockade of TNF-α by the antagonists infliximab, adalimumab, certolizumab, or etanercept has been associated with severe infections due to tuberculosis, nontuberculous mycobacteria, and fungi (Chap. 330).

Monocytopenia occurs with acute infections, with stress, and after treatment with glucocorticoids. Monocytopenia also occurs in aplastic anemia, hairy cell leukemia, acute myeloid leukemia, and as a direct result of myelotoxic drugs.

EOSINOPHILS

Eosinophils and neutrophils share similar morphology, many lysosomal constituents, phagocytic capacity, and oxidative metabolism. Eosinophils express a specific chemoattractant receptor and respond to a specific chemokine, eotaxin, but little is known about their required role. Eosinophils are much longer lived than neutrophils, and unlike neutrophils, tissue eosinophils can recirculate. During most infections, eosinophils appear unimportant. However, in invasive helminthic infections, such as hookworm, schistosomiasis, strongyloidiasis, toxocariasis, trichinosis, filariasis, echinococcosis, and cysticercosis, the eosinophil plays a central role in host defense. Eosinophils are associated with bronchial asthma, cutaneous allergic reactions, and other hypersensitivity states.

The distinctive feature of the red-staining (Wright's stain) eosinophil granule is its crystalline core

consisting of an arginine-rich protein (major basic protein) with histaminase activity, important in host defense against parasites. Eosinophil granules also contain a unique eosinophil peroxidase that catalyzes the oxidation of many substances by hydrogen peroxide and may facilitate killing of microorganisms.

Eosinophil peroxidase, in the presence of hydrogen peroxide and halide, initiates mast cell secretion in vitro and thereby promotes inflammation. Eosinophils contain cationic proteins, some of which bind to heparin and reduce its anticoagulant activity. Eosinophil-derived neurotoxin and eosinophil cationic protein are ribonucleases that can kill respiratory syncytial virus. Eosinophil cytoplasm contains Charcot-Leyden crystal protein, a hexagonal bipyramidal crystal first observed in a patient with leukemia and then in sputum of patients with asthma; this protein is lysophospholipase and may function to detoxify certain lysophospholipids.

Several factors enhance the eosinophil's function in host defense. T cell–derived factors enhance the ability of eosinophils to kill parasites. Mast cell–derived eosinophil chemotactic factor of anaphylaxis (ECFa) increases the number of eosinophil complement receptors and enhances eosinophil killing of parasites. Eosinophil CSFs (e.g., IL-5) produced by macrophages increase eosinophil production in the bone marrow and activate eosinophils to kill parasites.

Eosinophilia

Eosinophilia is the presence of >500 eosinophils per μL of blood and is common in many settings besides parasite infection. Significant tissue eosinophilia can occur without an elevated blood count. A common cause of eosinophilia is allergic reaction to drugs (iodides, aspirin, sulfonamides, nitrofurantoin, penicillins, and cephalosporins). Allergies such as hay fever, asthma, eczema, serum sickness, allergic vasculitis, and pemphigus are associated with eosinophilia. Eosinophilia also occurs in collagen vascular diseases (e.g., rheumatoid arthritis, eosinophilic fasciitis, allergic angiitis, and periarteritis nodosa) and malignancies (e.g., Hodgkin disease; mycosis fungoides; chronic myeloid leukemia; and cancer of the lung, stomach, pancreas, ovary, or uterus), as well as in Job's syndrome, DOCK8 deficiency (see below), and CGD. Eosinophilia is commonly present in helminthic infections. IL-5 is the dominant eosinophil growth factor. Therapeutic administration of the cytokinesIL-2 and GM-CSF frequently leads to transient eosinophilia. The most dramatic hypereosinophilic syndromes are Loeffler's syndrome, tropical pulmonary eosinophilia, Loeffler's endocarditis, eosinophilic leukemia, and idiopathic hypereosinophilic syndrome (50,000–100,000/μL). IL-5 is the dominant eosinophil growth factor, and can be specifically inhibited with the monoclonal antibody mepolizumab.

The idiopathic hypereosinophilic syndrome represents a heterogeneous group of disorders with the common feature of prolonged eosinophilia of unknown cause and organ system dysfunction, including the heart, central nervous system, kidneys, lungs, gastrointestinal tract, and skin. The bone marrow is involved in all affected individuals, but the most severe complications involve the heart and central nervous system. Clinical manifestations and organ dysfunction are highly variable. Eosinophils are found in the involved tissues and likely cause tissue damage by local deposition of toxic eosinophil proteins such as eosinophil cationic protein and major basic protein. In the heart, the pathologic changes lead to thrombosis, endocardial fibrosis, and restrictive endomyocardiopathy. The damage to tissues in other organ systems is similar. Some cases are due to mutations involving the platelet-derived growth factor receptor, and these are extremely sensitive to the tyrosine kinase inhibitor imatinib. Glucocorticoids, hydroxyurea, and IFN-α each have been used successfully, as have therapeutic antibodies against IL-5. Cardiovascular complications are managed aggressively.

The *eosinophilia-myalgia syndrome* is a multisystem disease, with prominent cutaneous, hematologic, and visceral manifestations, that frequently evolves into a chronic course and can occasionally be fatal. The syndrome is characterized by eosinophilia (eosinophil count >1000/μL) and generalized disabling myalgias without other recognized causes. Eosinophilic fasciitis, pneumonitis, and myocarditis; neuropathy culminating in respiratory failure; and encephalopathy may occur. The disease is caused by ingesting contaminants in l-tryptophan–containing products. Eosinophils, lymphocytes, macrophages, and fibroblasts accumulate in the affected tissues, but their role in pathogenesis is unclear. Activation of eosinophils and fibroblasts and the deposition of eosinophil-derived toxic proteins in affected tissues may contribute. IL-5 and transforming growth factor β have been implicated as potential mediators. Treatment is withdrawal of products containing l-tryptophan and the administration of glucocorticoids. Most patients recover fully, remain stable, or show slow recovery, but the disease can be fatal in up to 5% of patients.

Eosinopenia

Eosinopenia occurs with stress, such as acute bacterial infection, and after treatment with glucocorticoids. The mechanism of eosinopenia of acute bacterial infection is unknown but is independent of endogenous glucocorticoids, since it occurs in animals after total adrenalectomy. There is no known adverse effect of eosinopenia.

HYPERIMMUNOGLOBULIN E–RECURRENT INFECTION SYNDROME

The hyperimmunoglobulin E–recurrent infection syndrome, or Job's syndrome, is a rare multisystem

disease in which the immune and somatic systems are affected, including neutrophils, monocytes, T cells, B cells, and osteoclasts. Autosomal dominant mutations in signal transducer and activator of transcription 3 (*STAT3*) lead to inhibition of normal STAT signaling with broad and profound effects. Patients have characteristic facies with broad nose, kyphoscoliosis and osteoporosis, and eczema. The primary teeth erupt normally but do not deciduate, often requiring extraction. Patients develop recurrent sinopulmonary and cutaneous infections that tend to be much less inflamed than appropriate for the degree of infection and have been referred to as "cold abscesses." Characteristically, pneumonias cavitate, leading to pneumatoceles. Coronary artery aneurysms are common, as are cerebral demyelinated plaques that accumulate with age. Importantly, IL-17–producing cells, which are thought responsible for protection against extracellular and mucosal infections, are profoundly reduced in Job's syndrome. Despite very high IgE levels, these patients do not have elevated levels of allergy. An important syndrome with clinical overlap with STAT3 deficiency is due to autosomal recessive defects in dedicator of cytokinesis 8 (DOCK8). In DOCK8 deficiency, IgE elevation is joined to severe allergy, viral susceptibility, and increased rates of cancer.

LABORATORY DIAGNOSIS AND MANAGEMENT

Initial studies of WBC and differential and often a bone marrow examination may be followed by assessment of bone marrow reserves (steroid challenge test), marginated circulating pool of cells (epinephrine challenge test), and marginating ability (endotoxin challenge test) (Fig. 60-7). In vivo assessment of inflammation is possible with a Rebuck skin window test or an in vivo skin blister assay, which measures the ability of leukocytes and inflammatory mediators to accumulate locally in the skin. In vitro tests of phagocyte aggregation, adherence, chemotaxis, phagocytosis, degranulation, and microbicidal activity (for *S. aureus*) may help pinpoint cellular or humoral lesions. Deficiencies of oxidative metabolism are detected with either the nitroblue tetrazolium (NBT) dye test or the dihydrorhodamine (DHR) oxidation test. These tests are based on the ability of products of oxidative metabolism to alter the oxidation states of reporter molecules so that they can be detected microscopically (NBT) or by flow cytometry (DHR). Qualitative studies of superoxide and hydrogen peroxide production may further define neutrophil oxidative function.

Patients with leukopenias or leukocyte dysfunction often have delayed inflammatory responses. Therefore, clinical manifestations may be minimal despite overwhelming infection, and unusual infections must always be suspected. Early signs of infection demand prompt, aggressive culturing for microorganisms, use of antibiotics, and surgical drainage of abscesses. Prolonged courses of antibiotics are often required. In patients with CGD, prophylactic antibiotics (trimethoprim-sulfamethoxazole) and antifungals (itraconazole) markedly diminish the frequency of life-threatening infections. Glucocorticoids may relieve gastrointestinal or genitourinary tract obstruction by granulomas in patients with CGD. Although TNF-α blocking agents may markedly relieve inflammatory bowel symptoms, extreme caution must be exercised in their use in CGD inflammatory bowel disease, as it profoundly increases these patients' already heightened susceptibility to infection. Recombinant human IFN-γ, which nonspecifically stimulates phagocytic cell function, reduces the frequency of infections in patients with CGD by 70% and reduces the severity of infection. This effect of IFN-γ in CGD is additive to the effect of prophylactic antibiotics. The recommended dose is 50 $\mu g/m^2$ subcutaneously three times weekly. IFN-γ has also been used successfully in the treatment of leprosy, nontuberculous mycobacteria, and visceral leishmaniasis.

Rigorous oral hygiene reduces but does not eliminate the discomfort of gingivitis, periodontal disease, and aphthous ulcers; chlorhexidine mouthwash and tooth brushing with a hydrogen peroxide–sodium bicarbonate paste helps many patients. Oral antifungal agents (fluconazole, itraconazole, voriconazole, posaconazole) have reduced mucocutaneous candidiasis in patients with Job's syndrome. Androgens, glucocorticoids, lithium, and immunosuppressive therapy have been used to restore myelopoiesis in patients with neutropenia due to impaired production. Recombinant G-CSF is useful in the management of certain forms of neutropenia due to depressed neutrophil production, especially those related to cancer chemotherapy. Patients with chronic neutropenia with evidence of a good bone marrow reserve need not receive prophylactic antibiotics. Patients with chronic or cyclic neutrophil counts <500/μL may benefit from prophylactic antibiotics and G-CSF during periods of neutropenia. Oral trimethoprim-sulfamethoxazole (160/800 mg) twice daily can prevent infection. Increased numbers of fungal infections are not seen in patients with CGD on this regimen. Oral quinolones such as levofloxacin and ciprofloxacin are alternatives.

In the setting of cytotoxic chemotherapy with severe, persistent neutropenia, trimethoprim-sulfamethoxazole prevents *Pneumocystis jirovecii* pneumonia. These patients, and patients with phagocytic cell dysfunction, should avoid heavy exposure to airborne soil, dust, or decaying matter (mulch, manure), which are often rich in *Nocardia* and the spores of *Aspergillus* and other fungi. Restriction of activities or social contact has no proven role in reducing risk of infection.

Although aggressive medical care for many patients with phagocytic disorders can allow them to go for years without a life-threatening infection, there may still be delayed effects of prolonged antimicrobials and other inflammatory complications. Cure of most congenital phagocyte defects is possible by bone

marrow transplantation, and rates of success are improving (Chap. 114). The identification of specific gene defects in patients with LAD 1, CGD, and other immunodeficiencies has led to gene therapy trials in a number of genetic white cell disorders.

FURTHER READINGS

Heimall J et al: Pathogenesis of hyper IgE syndrome. Clin Rev Allergy Immunol 38:32, 2010 Holland SM: Chronic granulomatous disease. Clin Rev Allergy Immunol 38:3, 2010

International Union of Immunological Societies Expert Committee on Primary Immunodeficiencies: Primary immunodeficiencies: 2009 update. J Allergy Clin Immunol 124:1161, 2009

Klein C, Welte K: Genetic insights into congenital neutropenia. Clin Rev Allergy Immunol 38:68, 2010[PMID: 19440858]

Kuhns DB, et al: Residual NADPH opidase and survival in chronic granulomatous disease. N Engl J Med 363:2600, 2010[PMID: 21190454]

Notarangelo LD: Primary immunodeficiencies. J Allergy Clin Immunol 125:S182, 2010.

Ogbogu PU et al: Hypereosinophilic syndrome: A multicenter, retrospective analysis of clinical characteristics and response to therapy. J Allergy Clin Immunol 124:1319, 2009[PMID: 19910029]

van de Vosse E et al: Genetic deficiencies of innate immune signalling in human infectious disease. Lancet Infect Dis 9:688, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF HEMATOLOGY AND ANALYSIS OF PERIPHERAL BLOOD SMEARS: INTRODUCTION

Some of the relevant findings in peripheral blood, enlarged lymph nodes, and bone marrow are illustrated in this chapter. Systematic histologic examination of the bone marrow and lymph nodes is beyond the scope of a general medicine textbook. However, every internist should know how to examine a peripheral blood smear.

The examination of a peripheral blood smear is one of the most informative exercises a physician can perform. Although advances in automated technology have made the examination of a peripheral blood smear by a physician seem less important, the technology is not a completely satisfactory replacement for a blood smear interpretation by a trained medical professional who also knows the patient's clinical history, family history, social history, and physical findings. It is useful to ask the laboratory to generate a Wright's-stained peripheral blood smear and examine it.

The best place to examine blood cell morphology is the feathered edge of the blood smear where red cells lie in a single layer, side by side, just barely touching one another but not overlapping. The author's approach is to look at the smallest cellular elements, the platelets, first and work his way up in size to red cells and then white cells.

Using an oil immersion lens that magnifies the cells 100-fold, one counts the platelets in five to six fields, averages the number per field, and multiplies by 20,000 to get a rough estimate of the platelet count. The platelets are usually 1–2 μm in diameter and have a blue granulated appearance. There is usually 1 platelet for every 20 or so red cells. Of course, the automated counter is much more accurate, but gross disparities between the automated and manual counts should be assessed. Large platelets may be a sign of rapid platelet turnover, as young platelets are often larger than old ones; alternatively, certain rare inherited syndromes can produce large platelets. Platelet clumping visible on the smear can be associated with falsely low automated platelet counts. Similarly, neutrophil fragmentation can be a source of falsely elevated automated platelet counts.

Next one examines the red blood cells. One can gauge their size by comparing the red cell to the nucleus of a small lymphocyte. Both are normally about 8 μm wide. Red cells that are smaller than the small lymphocyte nucleus may be microcytic; those larger than the small lymphocyte nucleus may be macrocytic. Macrocytic cells also tend to be more oval than spherical in shape and are sometimes called macroovalocytes. The automated mean corpuscular volume (MCV) can assist in making a classification. However, some patients may have both iron and vitamin B_{12} deficiency, which will produce an MCV in the normal range but wide variation in red cell size. When the red cells vary greatly in size, *anisocytosis* is said to be present. When the red cells vary greatly in shape, *poikilocytosis* is said to be present. The electronic cell counter provides an independent assessment of variability in red cell size. It measures the range of red cell volumes and reports the results as "red cell distribution width" (RDW). This value is calculated from the MCV; thus, cell width is not being measured but cell volume is. The term is derived from the curve displaying the frequency of cells at each volume, also called the distribution. The width of red cell volume distribution curve is what determines the RDW. The RDW is calculated as follows: RDW = (standard deviation of MCV ÷ mean MCV) x 100. In the presence of morphologic anisocytosis, RDW (normally 11–14%) increases to 15–18%. The RDW is useful in at least two clinical settings. In patients with microcytic anemia, the differential diagnosis is generally between iron deficiency and thalassemia. In thalassemia, the small red cells are generally of uniform size with a normal small RDW. In iron deficiency, the size variability and the RDW are large. In addition, a large RDW can suggest a dimorphic anemia when a chronic atrophic gastritis can produce both vitamin B_{12} malabsorption to produce macrocytic anemia and blood loss to produce iron deficiency. In such settings, RDW is also large. An elevated RDW also has been reported as a risk factor for all-cause mortality in population-based studies (Patel KV et al: Arch Intern Med 169:515, 2009), a finding that is unexplained currently.

After red cell size is assessed, one examines the hemoglobin content of the cells. They are either normal in color (*normochromic*) or pale in color (*hypochromic*). They are never "hyperchromic." If more than the normal amount of hemoglobin is made, the cells get larger—they do not become darker. In addition to hemoglobin content, the red cells are examined for inclusions. Red cell inclusions are the following:

1. *Basophilic stippling*—diffuse fine or coarse blue dots in the red cell usually representing RNA residue—especially common in lead poisoning

2. *Howell-Jolly bodies*—dense blue circular inclusions that represent nuclear remnants—their presence implies defective splenic function
3. *Nuclei*—red cells may be released or pushed out of the marrow prematurely before nuclear extrusion—often implies a myelophthisic process or a vigorous narrow response to anemia, usually hemolytic anemia
4. *Parasites*—red cell parasites include malaria and babesia (Chap. e27)
5. *Polychromatophilia*—the red cell cytoplasm has a bluish hue, reflecting the persistence of ribosomes still actively making hemoglobin in a young red cell

Vital stains are necessary to see precipitated hemoglobin called *Heinz bodies*.

Red cells can take on a variety of different shapes. All abnormally shaped red cells are *poikilocytes*. Small red cells without the central pallor are *spherocytes*; they can be seen in hereditary spherocytosis, hemolytic anemias of other causes, and clostridial sepsis. *Dacrocytes* are teardrop-shaped cells that can be seen in hemolytic anemias, severe iron deficiency, thalassemias, myelofibrosis, and myelodysplastic syndromes. *Schistocytes* are helmet-shaped cells that reflect microangiopathic hemolytic anemia or fragmentation on an artificial heart valve. *Echinocytes* are spiculated red cells with the spikes evenly spaced; they can represent an artifact of abnormal drying of the blood smear or reflect changes in stored blood. They also can be seen in renal failure and malnutrition and are often reversible. *Acanthocytes* are spiculated red cells with the spikes irregularly distributed. This process tends to be irreversible and reflects underlying renal disease, abetalipoproteinemia, or splenectomy. *Elliptocytes* are elliptical-shaped red cells that can reflect an inherited defect in the red cell membrane, but they also are seen in iron deficiency, myelodysplastic syndromes, megaloblastic anemia, and thalassemias. *Stomatocytes* are red cells in which the area of central pallor takes on the morphology of a slit instead of the usual round shape. Stomatocytes can indicate an inherited red cell membrane defect and also can be seen in alcoholism. *Target cells* have an area of central pallor that contains a dense center, or bull's-eye. These cells are seen classically in thalassemia, but they are also present in iron deficiency, cholestatic liver disease, and some hemoglobinopathies. They also can be generated artifactually by improper slide making.

One last feature of the red cells to assess before moving to the white blood cells is the distribution of the red cells on the smear. In most individuals, the cells lie side by side in a single layer. Some patients have red cell clumping (called *agglutination*) in which the red cells pile upon one another; it is seen in certain paraproteinemias and autoimmune hemolytic anemias. Another abnormal distribution involves red cells lying in single cell rows on top of one another like stacks of coins. This is called *rouleaux formation* and reflects abnormal serum protein levels.

Finally, one examines the white blood cells. Three types of granulocytes are usually present: neutrophils, eosinophils, and basophils, in decreasing frequency. Neutrophils are generally the most abundant white cell. They are round, are 10–14 μm wide, and contain a lobulated nucleus with two to five lobes connected by a thin chromatin thread. Bands are immature neutrophils that have not completed nuclear condensation and have a U-shaped nucleus. Bands reflect a left shift in neutrophil maturation in an effort to make more cells more rapidly. Neutrophils can provide clues to a variety of conditions. Vacuolated neutrophils may be a sign of bacterial sepsis. The presence of 1- to 2-μm blue cytoplasmic inclusions, called *Döhle bodies*, can reflect infections, burns, or other inflammatory states. If the neutrophil granules are larger than normal and stain a darker blue, "toxic granulations" are said to be present, and they also suggest a systemic inflammation. The presence of neutrophils with more than five nuclear lobes suggests megaloblastic anemia. Large misshapen granules may reflect the inherited Chédiak-Higashi syndrome.

Eosinophils are slightly larger than neutrophils, have bilobed nuclei, and contain large red granules. Diseases of eosinophils are associated with too many of them rather than any morphologic or qualitative change. They normally total less than one-thirtieth the number of neutrophils. Basophils are even more rare than eosinophils in the blood. They have large dark blue granules and may be increased as part of chronic myeloid leukemia.

Lymphocytes can be present in several morphologic forms. Most common in healthy individuals are small lymphocytes with a small dark nucleus and scarce cytoplasm. In the presence of viral infections, more of the lymphocytes are larger, about the size of neutrophils, with abundant cytoplasm and a less condensed nuclear chromatin. These cells are called *reactive lymphocytes*. About 1% of lymphocytes are larger and contain blue granules in a light blue cytoplasm; they are called *large granular lymphocytes*. In chronic lymphoid leukemia, the small lymphocytes are increased in number, and many of them are ruptured in making the blood smear, leaving a smudge of nuclear material without a surrounding cytoplasm or cell membrane; they are called *smudge cells* and are rare in the absence of chronic lymphoid leukemia.

Monocytes are the largest white blood cells, ranging from 15 to 22 μm in diameter. The nucleus can take on a variety of shapes but usually appears to be folded; the cytoplasm is gray.

Abnormal cells may appear in the blood. Most often the abnormal cells originate from neoplasms of bone marrow–derived cells, including lymphoid cells, myeloid cells, and occasionally red cells. More rarely, other types of tumors can get access to the bloodstream, and rare epithelial malignant cells may be identified. The chances of seeing such abnormal cells is increased by examining blood smears made from buffy coats, the layer of cells that is visible on top of sedimenting red cells when blood is left in the test tube for an hour. Smears made from finger sticks may include rare endothelial cells.

Figure e17-1

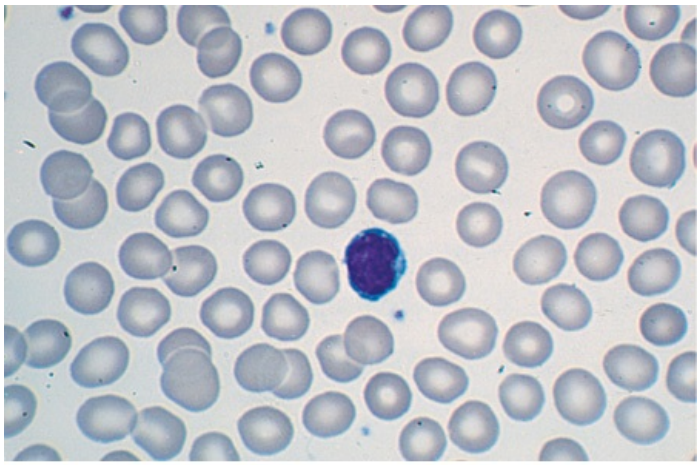

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal peripheral blood smear. Small lymphocyte in center of field. Note that the diameter of the red blood cell is similar to the diameter of the small lymphocyte nucleus.

Figure e17-2

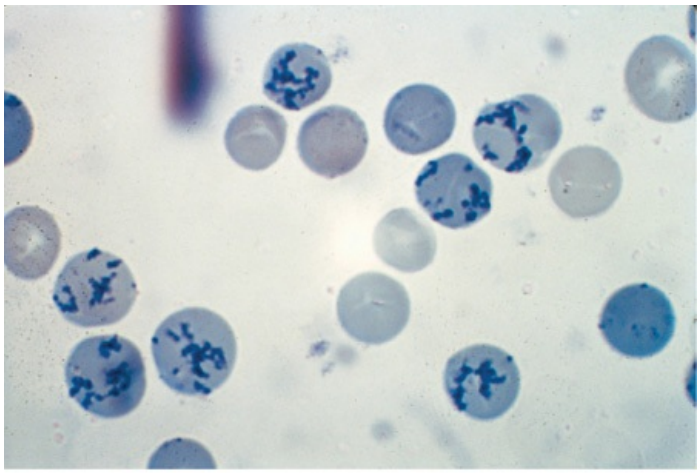

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Reticulocyte count preparation. This new methylene blue–stained blood smear shows large numbers of heavily stained reticulocytes (the cells containing the dark blue–staining RNA precipitates).

Figure e17-3

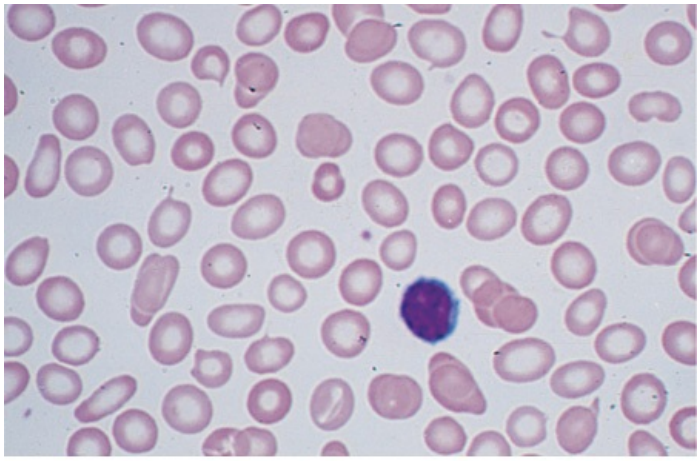

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypochromic microcytic anemia of iron deficiency. Small lymphocyte in field helps assess the red blood cell size.

Figure e17-4

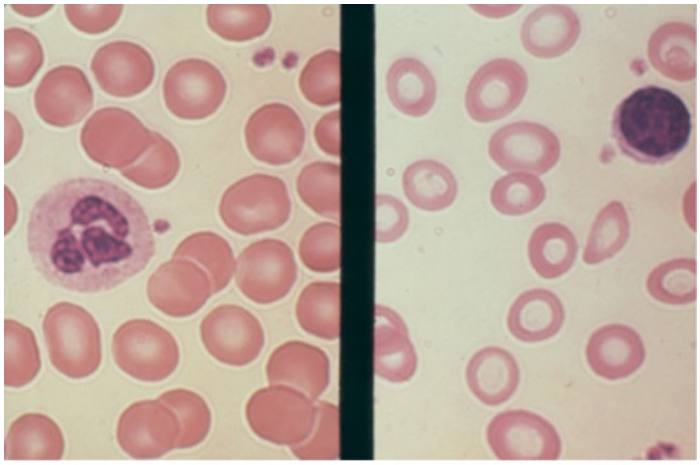

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Iron deficiency anemia next to normal red blood cells. Microcytes (*right panel*) are smaller than normal red blood cells (cell diameter <7 μm) and may or may not be poorly hemoglobinized (hypochromic).

Figure e17-5

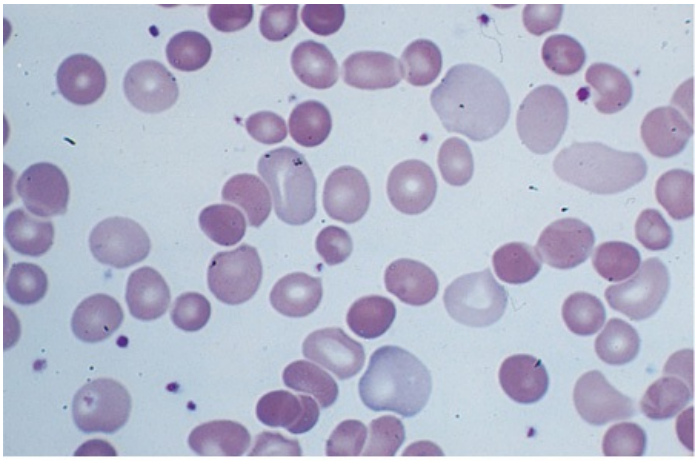

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Polychromatophilia. Note large red cells with light purple coloring.

Figure e17-6

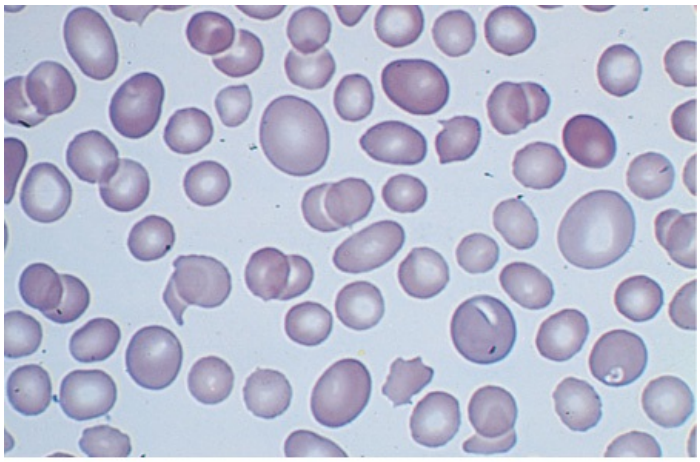

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Macrocytosis. These cells are both larger than normal (mean corpuscular volume >100) and somewhat oval in shape. Some morphologists call these cells macroovalocytes.

Figure e17-7

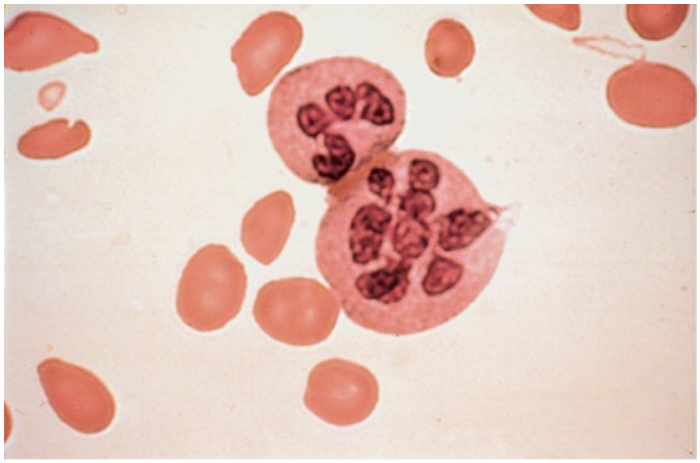

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypersegmented neutrophils. Hypersegmented neutrophils (multilobed polymorphonuclear leukocytes) are larger than normal neutrophils with five or more segmented nuclear lobes. They are commonly seen with folic acid or vitamin B_{12} deficiency.

Figure e17-8

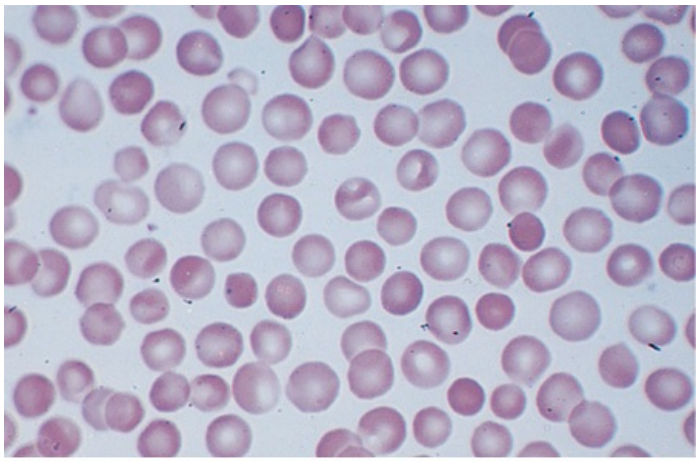

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spherocytosis. Note small hyperchromatic cells without the usual clear area in the center.

Figure e17-9

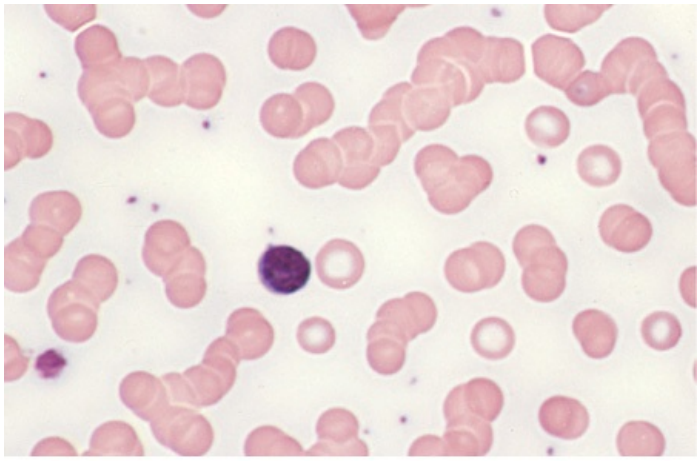

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rouleaux formation. Small lymphocyte in center of field. These red cells align themselves in stacks and are related to increased serum protein levels.

Figure e17-10

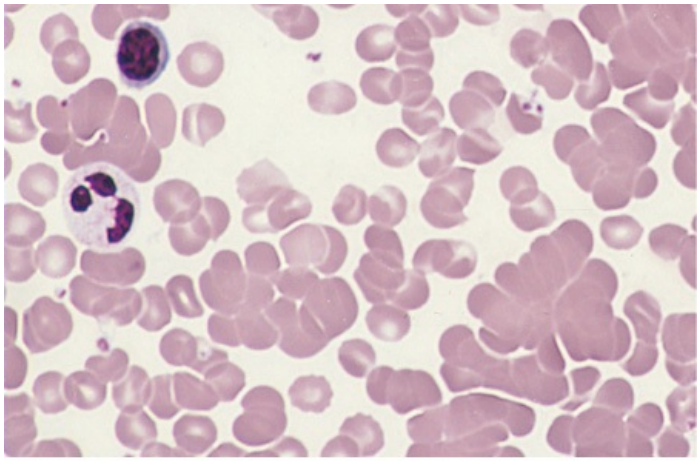

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Red cell agglutination. Small lymphocyte and segmented neutrophil in upper left center. Note irregular collections of aggregated red cells.

Figure e17-11

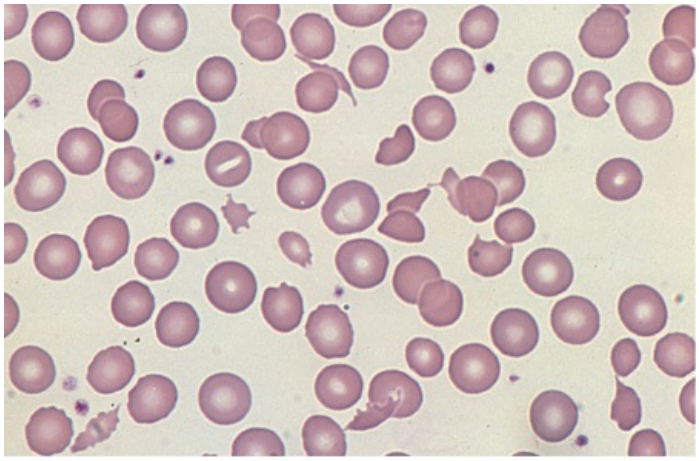

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fragmented red cells. Heart valve hemolysis.

Figure e17-12

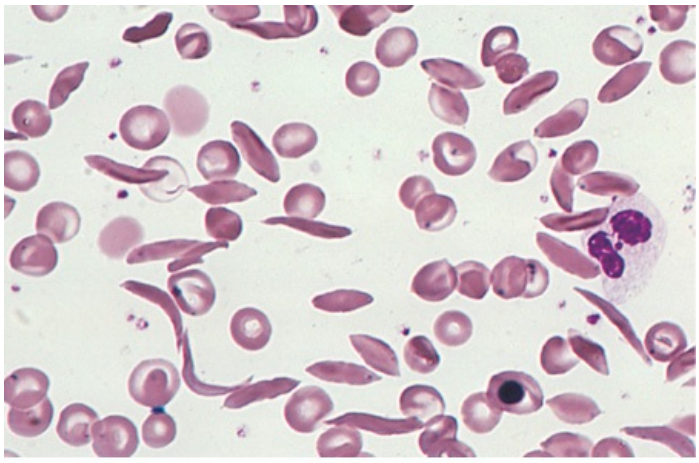

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sickle cells. Homozygoussickle cell disease. A nucleated red cell and neutrophil are also in the field.

Figure e17-13

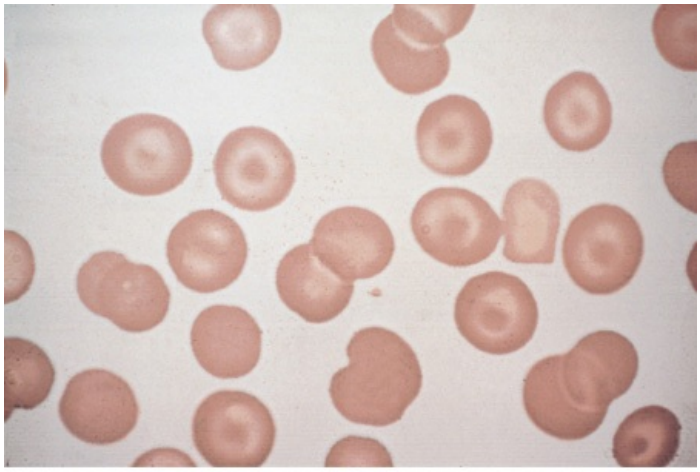

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Target cells. Target cells are recognized by the bull's-eye appearance of the cell. Small numbers of target cells are seen with liver disease and thalassemia. Larger numbers are typical of hemoglobin C disease.

Figure e17-14

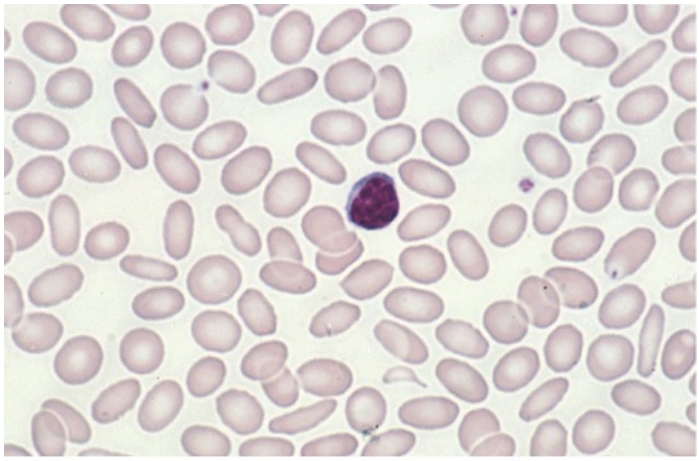

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Elliptocytosis. Small lymphocyte in center of field. Elliptical shape of red cells related to weakened membrane structure, usually due to mutations in spectrin.

Figure e17-15

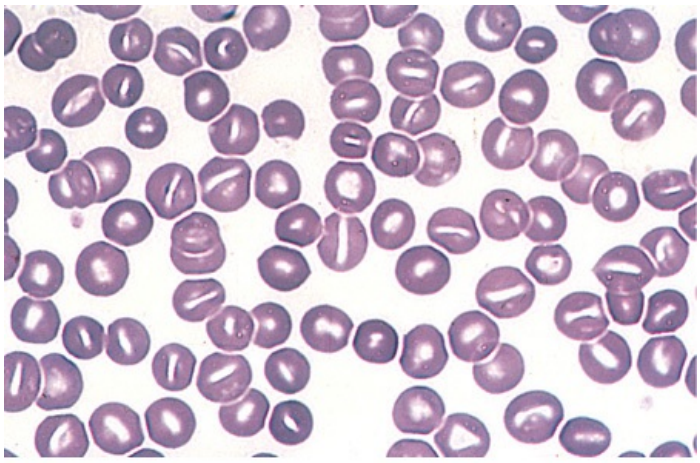

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stomatocytosis. Red cells characterized by a wide transverse slit or stoma. This often is seen as an artifact in a dehydrated blood smear. These cells can be seen in hemolytic anemias and in conditions in which the red cell is overhydrated or dehydrated.

Figure e17-16

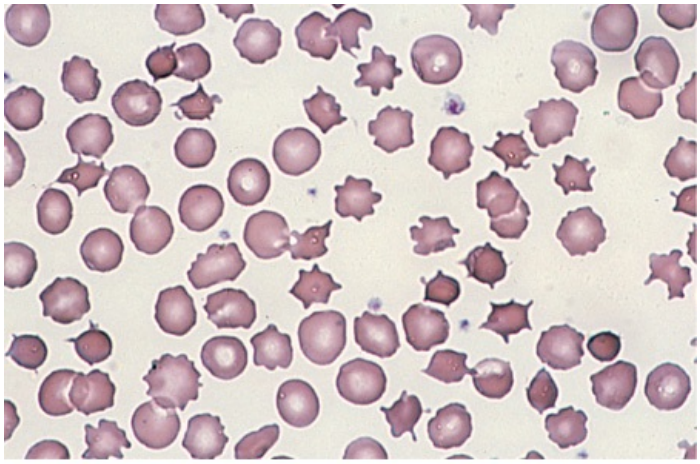

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acanthocytosis. Spiculated red cells are of two types: *acanthocytes* are contracted dense cells with irregular membrane projections that vary in length and width; *echinocytes* have small, uniform, and evenly spaced membrane projections. Acanthocytes are present in severe liver disease, in patients with abetalipoproteinemia, and in rare patients with McLeod blood group. Echinocytes are found in patients with severe uremia, in glycolytic red cell enzyme defects, and in microangiopathic hemolytic anemia.

Figure e17-17

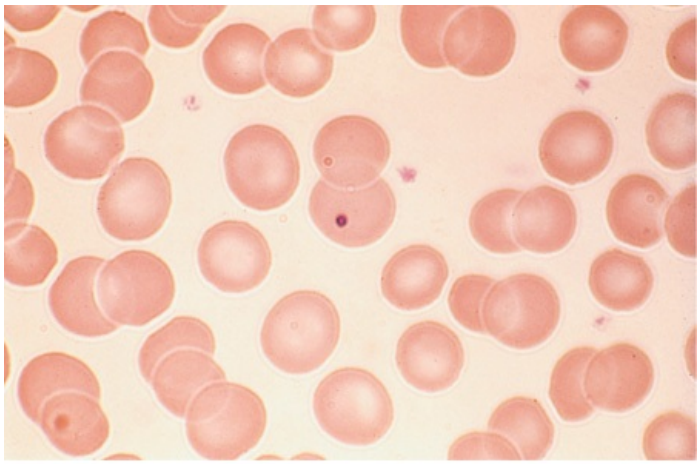

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Howell-Jolly bodies. Howell-Jolly bodies are tiny nuclear remnants that normally are removed by the spleen. They appear in the blood after splenectomy (defect in removal) and with maturation/dysplastic disorders (excess production).

Figure e17-18

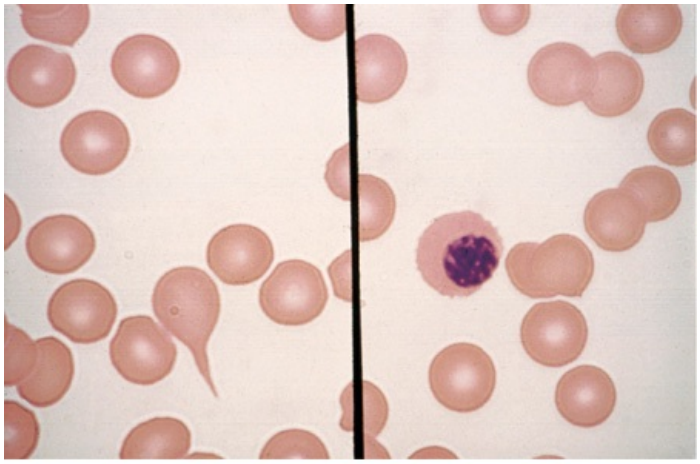

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Teardrop cells and nucleated red blood cells characteristic of myelofibrosis. A teardrop-shaped red blood cell (*left panel*) and a nucleated red blood cell (*right panel*) as typically seen with myelofibrosis and extramedullary hematopoiesis.

Figure e17-19

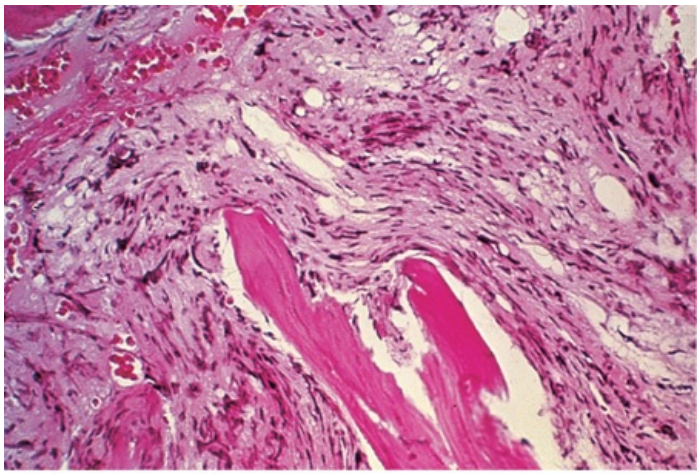

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Myelofibrosis of the bone marrow. Total replacement of marrow precursors and fat cells by a dense infiltrate of reticulin fibers and collagen (H&E stain).

Figure e17-20

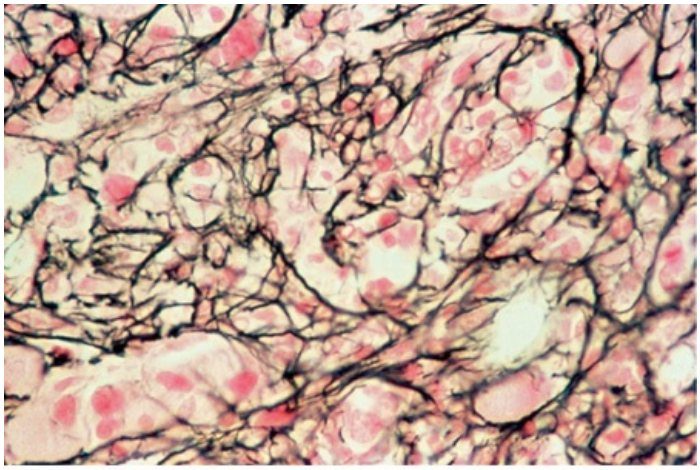

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Reticulin stain of marrow myelofibrosis. Silver stain of a myelofibrotic marrow showing an increase in reticulin fibers (black-staining threads).

Figure e17-21

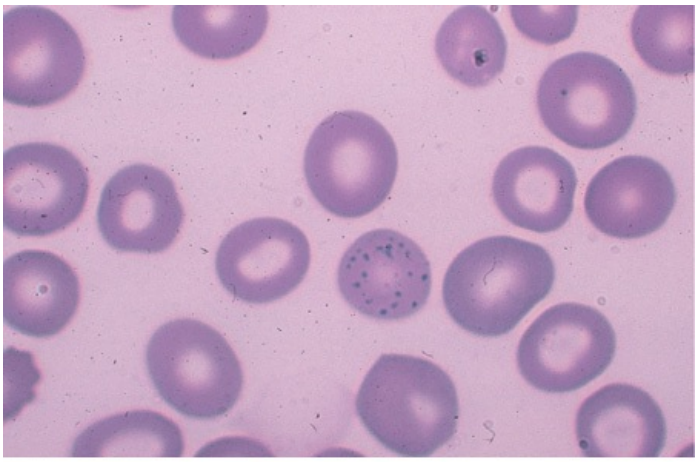

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stippled red cell in lead poisoning. Mild hypochromia. Coarsely stippled red cell.

Figure e17-22

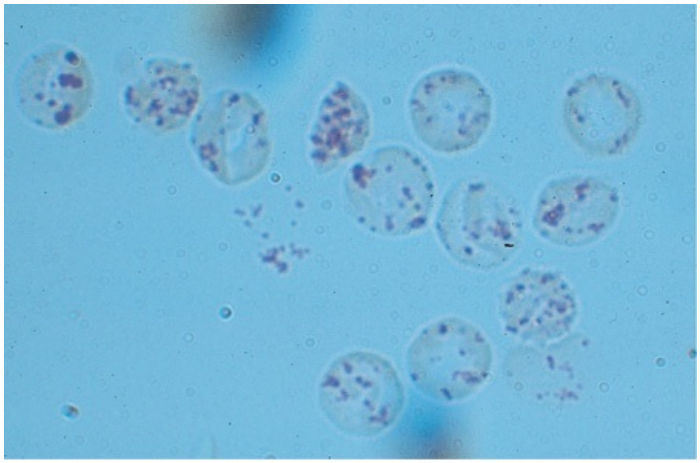

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Heinz bodies. Blood mixed with hypotonic solution of crystal violet. The stained material is precipitates of denatured hemoglobin within cells.

Figure e17-23

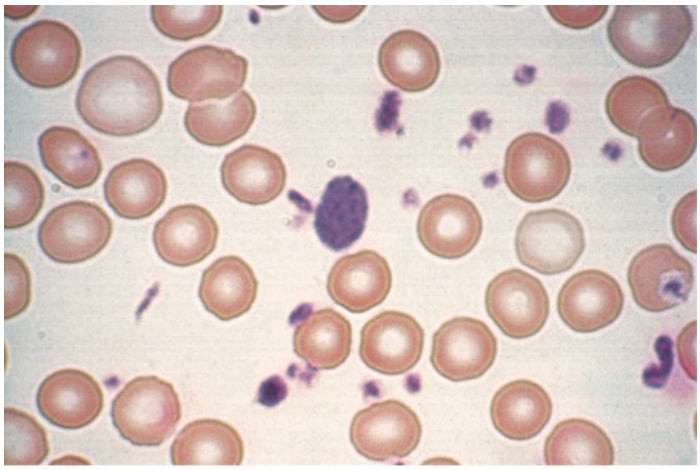

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Giant platelets. Giant platelets, together with a marked increase in the platelet count, are seen in myeloproliferative disorders, especially primary thrombocythemia.

Figure e17-24

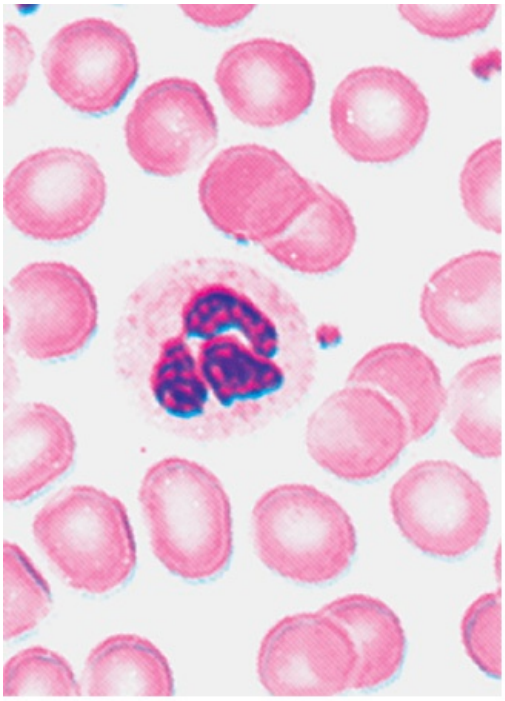

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal granulocytes. The normal granulocyte has a segmented nucleus with heavy, clumped chromatin; fine neutrophilic granules are dispersed throughout the cytoplasm.

Figure e17-25

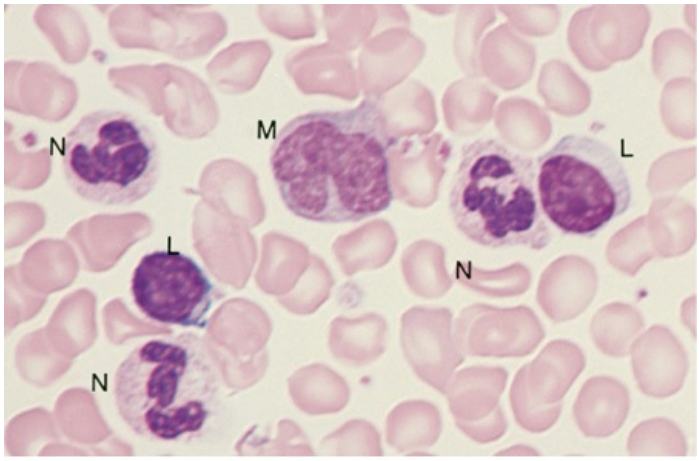

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal monocytes. The film was prepared from the buffy coat of the blood from a normal donor. L, lymphocyte; M monocyte; N, neutrophil.

Figure e17-26

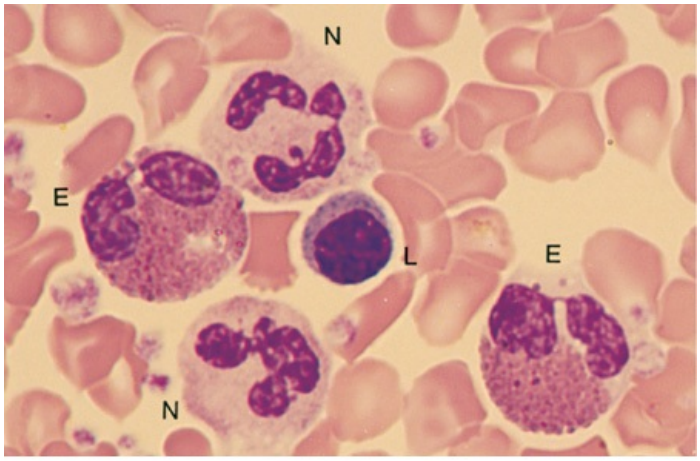

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal eosinophils. The film was prepared from the buffy coat of the blood from a normal donor. N, neutrophil; E, eosinophil; L, lymphocyte.

Figure e17-27

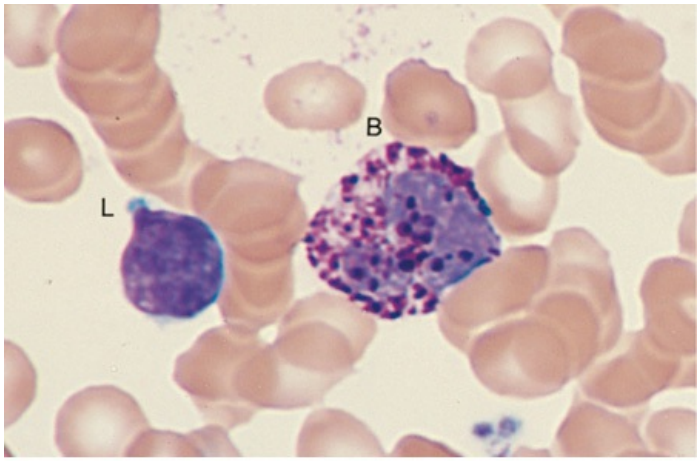

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal basophil. The film was prepared from the buffy coat of the blood from a normal donor. L, lymphocyte; B, basophil.

Figure e17-28

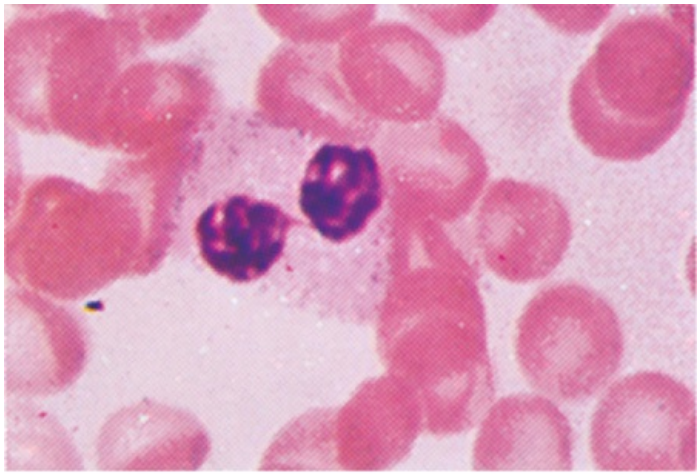

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pelger-Hüet anomaly. In this benign disorder, the majority of granulocytes are bilobed. The nucleus frequently has a spectacle-like, or "pince-nez," configuration.

Figure e17-29

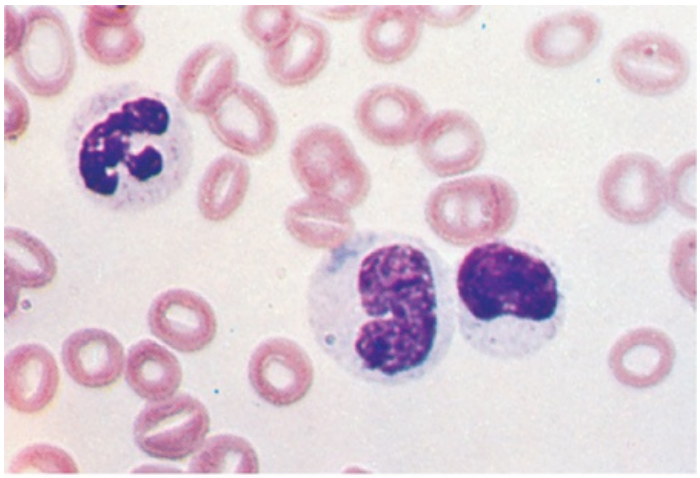

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Döhle body. Neutrophil band with Döhle body. The neutrophil with a sausage-shaped nucleus in the center of the field is a band form. Döhle bodies are discrete, blue-staining nongranular areas found in the periphery of the cytoplasm of the neutrophil in infections and other toxic states. They represent aggregates of rough endoplasmic reticulum.

Figure e17-30

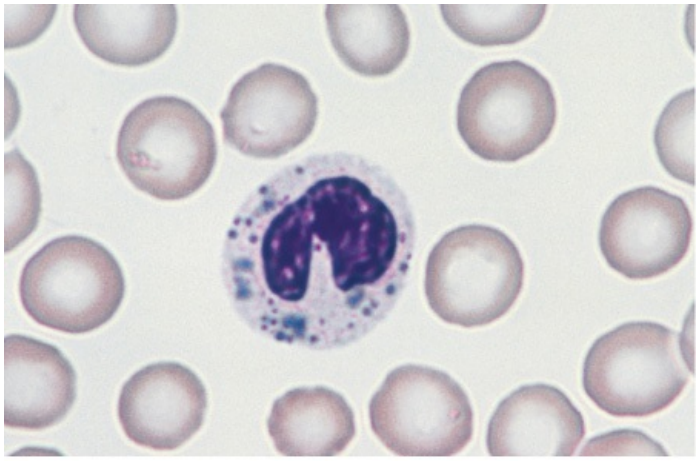

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chédiak-Higashi disease. Note giant granules in neutrophil.

Figure e17-31

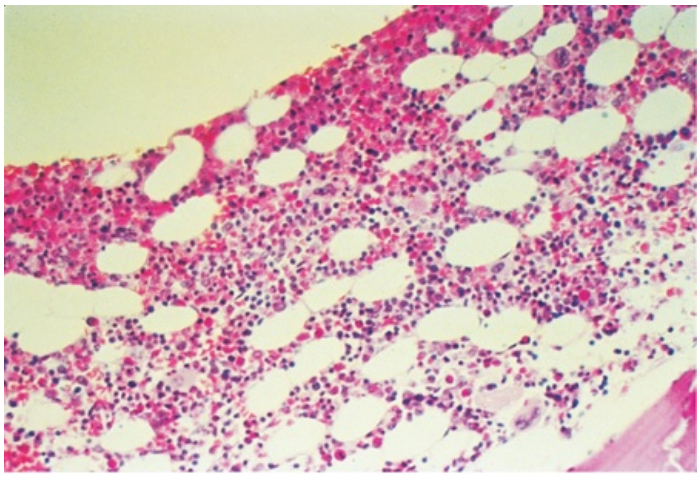

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal bone marrow. Low-power view of normal adult marrow (H&E stain), showing a mix of fat cells (clear areas) and hematopoietic cells. The percentage of the space that consists of hematopoietic cells is referred to as *marrow cellularity*. In adults, normal marrow cellularity is 35–40%. If demands for increased marrow production occur, cellularity may increase to meet the demand. As people age, the marrow cellularity decreases and the marrow fat increases. Patients >70 years old may have a 20–30% marrow cellularity.

Figure e17-32

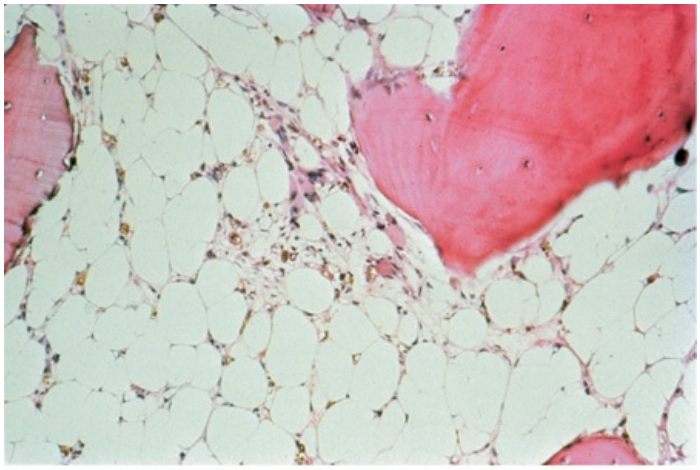

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Aplastic anemia bone marrow. Normal hematopoietic precursor cells are virtually absent, leaving behind fat cells, reticuloendothelial cells, and the underlying sinusoidal structure.

Figure e17-33

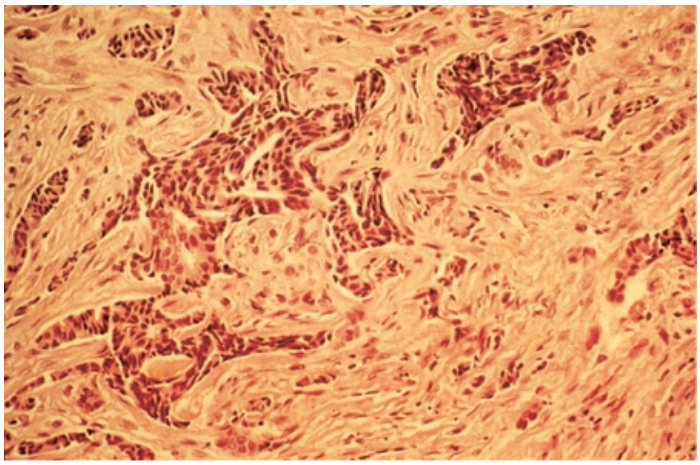

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Metastatic cancer in the bone marrow. Marrow biopsy specimen infiltrated with metastatic breast cancer and reactive fibrosis (H&E stain).

Figure e17-34

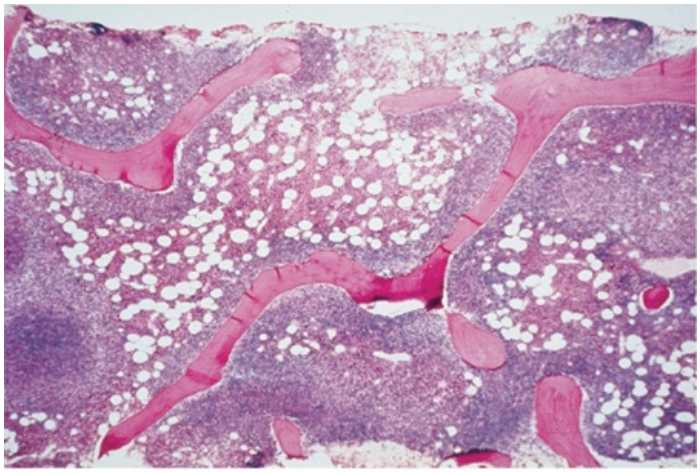

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lymphoma in the bone marrow. Nodular (follicular) lymphoma infiltrate in a marrow biopsy specimen. Note the characteristic paratrabecular location of the lymphoma cells.

Figure e17-35

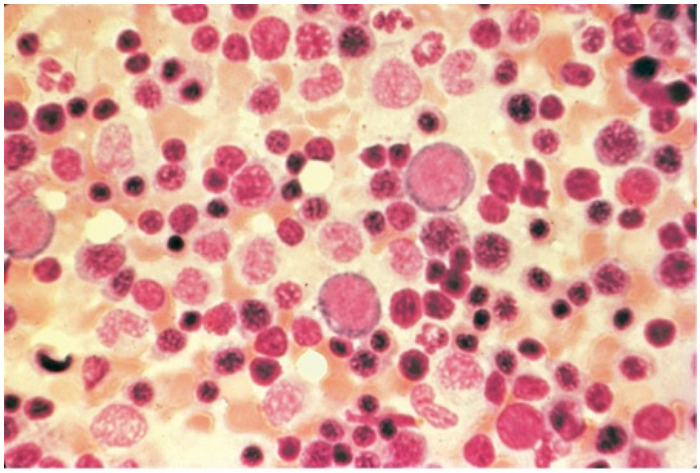

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythroid hyperplasia of the marrow. Marrow aspirate specimen with a myeloid/erythroid ratio (M/E ratio) of 1:1–2, typical for a patient with a hemolytic anemia or one recovering from blood loss.

Figure e17-36

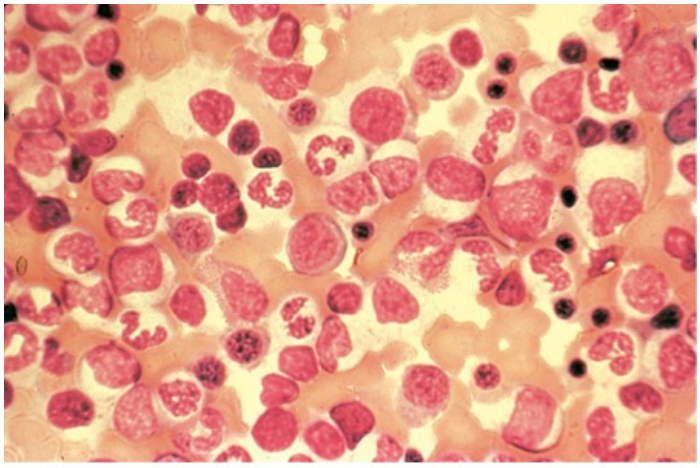

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Myeloid hyperplasia of the marrow. Marrow aspirate specimen showing a myeloid/erythroid ratio of ≥3:1, suggesting either a loss of red blood cell precursors or an expansion of myeloid elements.

Figure e17-37

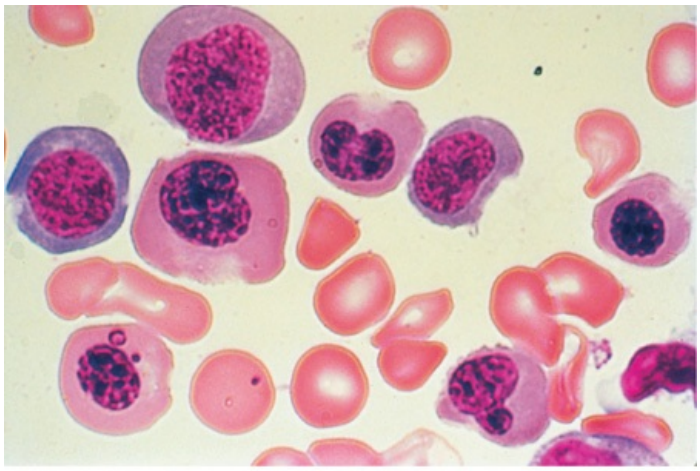

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Megaloblastic erythropoiesis. High-power view of megaloblastic red blood cell precursors from a patient with a macrocytic anemia. Maturation is delayed, with late normoblasts showing a more immature-appearing nucleus with a lattice-like pattern with normal cytoplasmic maturation.

Figure e17-38

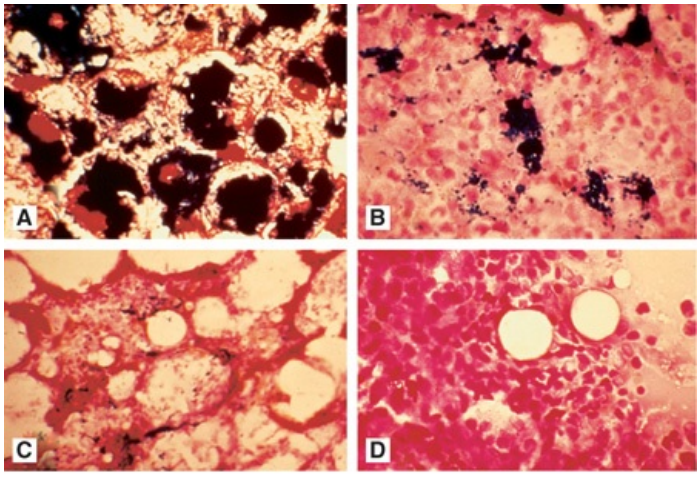

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prussian blue staining of marrow iron stores. Iron stores can be graded on a scale of 0 to 4+. ***A:*** a marrow with excess iron stores (>4+); ***B:*** normal stores (2–3+); ***C:*** minimal stores (1+); and ***D:*** absent iron stores (0).

Figure e17-39

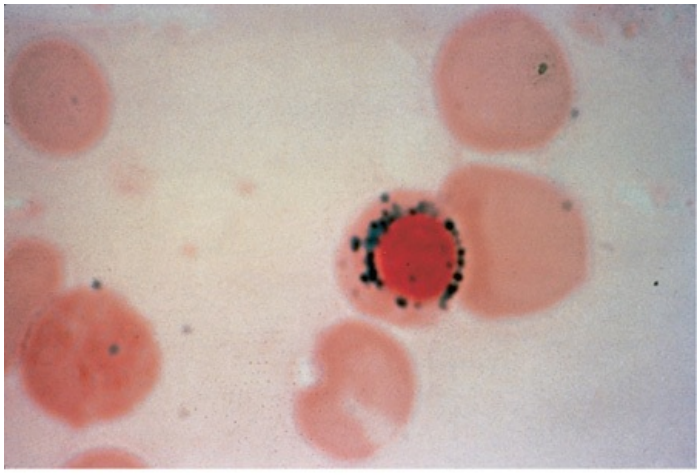

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ringed sideroblast. An orthochromatic normoblast with a collar of blue granules (mitochondria encrusted with iron) surrounding the nucleus.

Figure e17-40

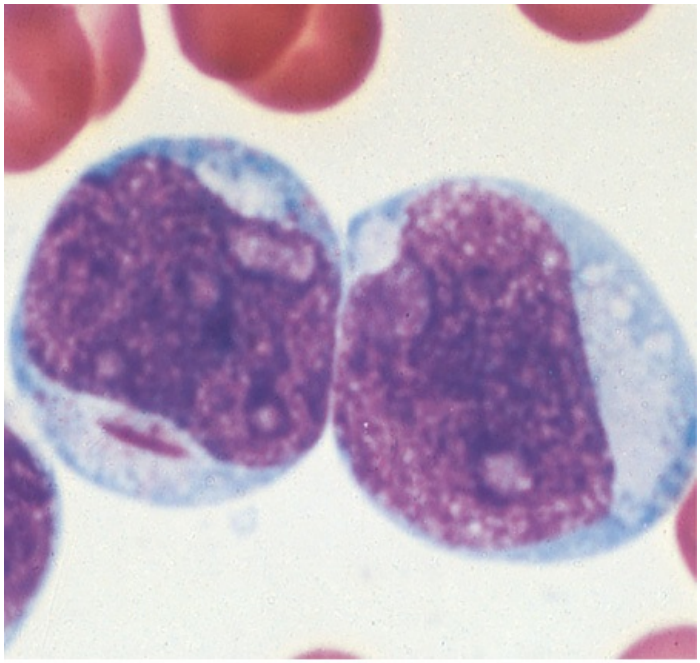

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute myeloid leukemia. Leukemic myeloblast with an Auer rod. Note two to four large, prominent nucleoli in each cell.

Figure e17-41

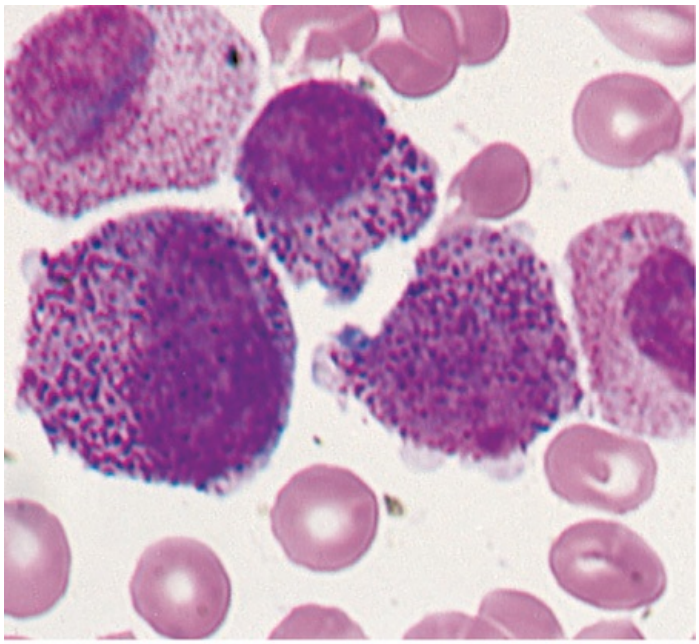

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute promyelocytic leukemia. Note prominent cytoplasmic granules in the leukemia cells.

Figure e17-42

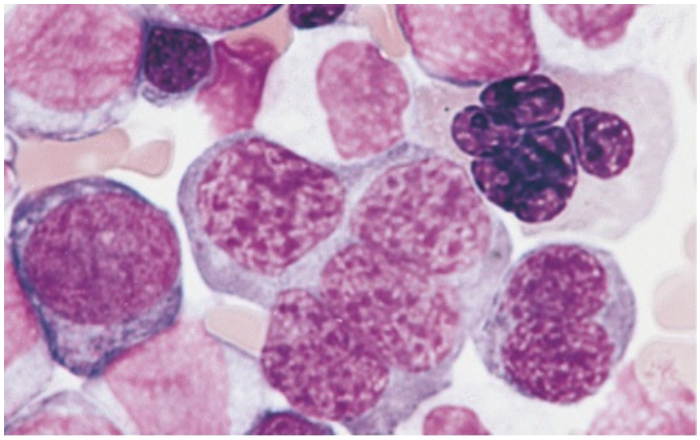

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute erythroleukemia. Note giant dysmorphic erythroblasts; two are binucleate, and one is multinucleate.

Figure e17-43

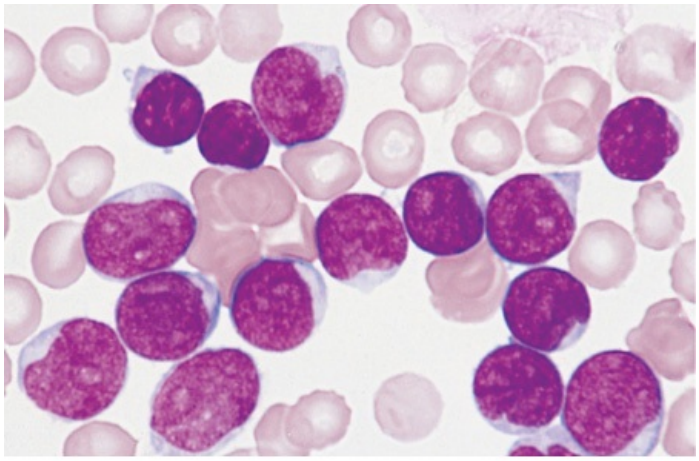

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute lymphoblastic leukemia.

Figure e17-44

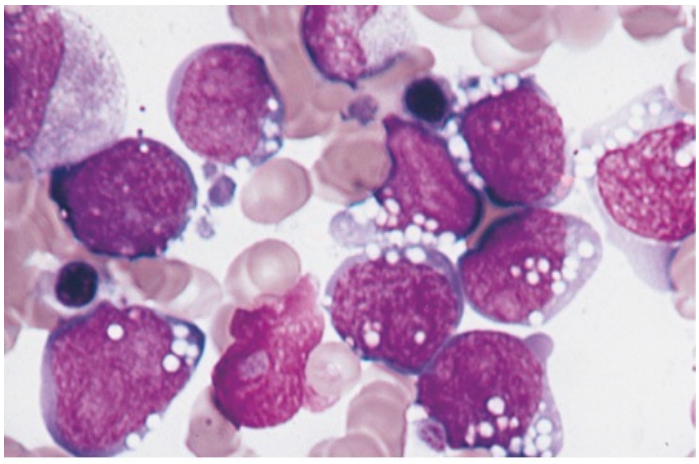

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Burkitt's leukemia, acute lymphoblastic leukemia.

Figure e17-45

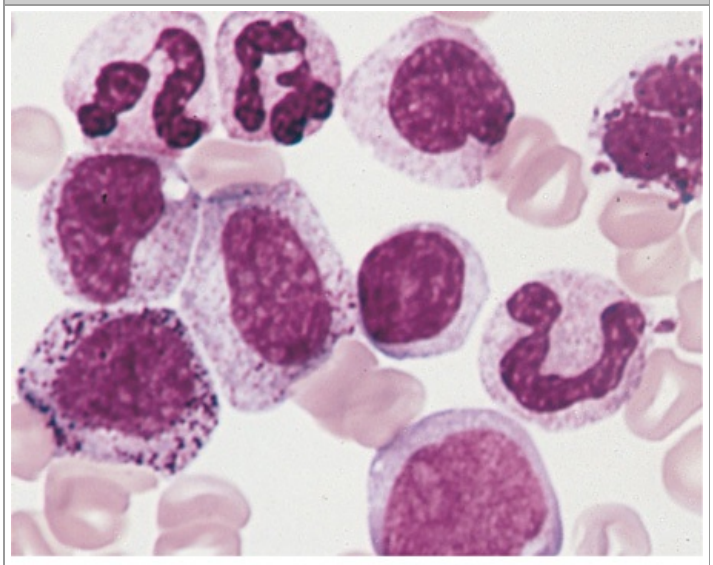

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic myeloid leukemia in the peripheral blood.

Figure e17-46

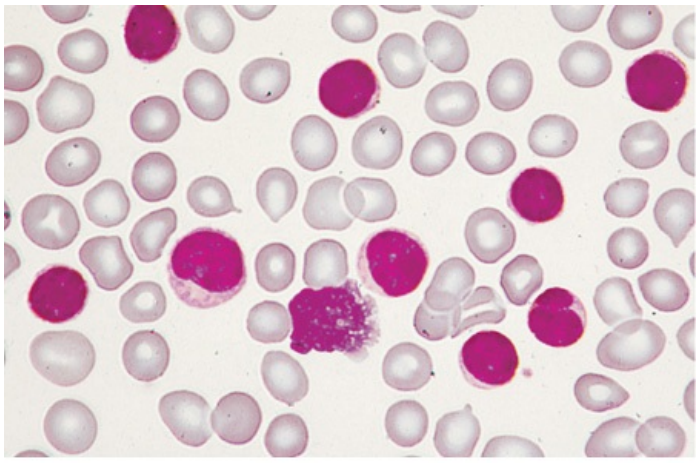

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic lymphoid leukemia in the peripheral blood.

Figure e17-47

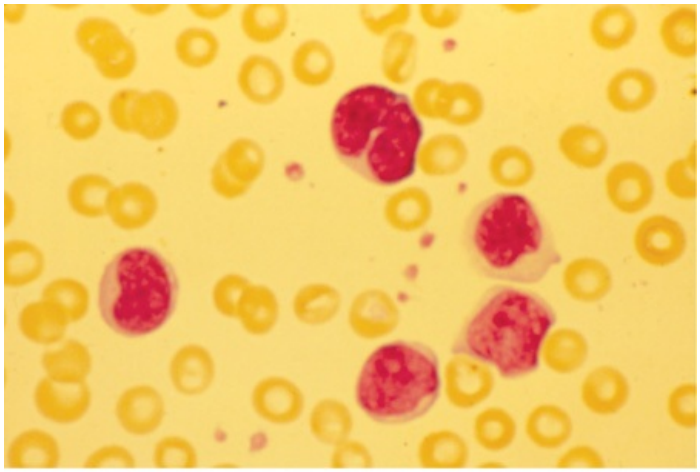

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sézary's syndrome. Lymphocytes with frequently convoluted nuclei (Sézary cells) in a patient with advanced mycosis fungoides.

Figure e17-48

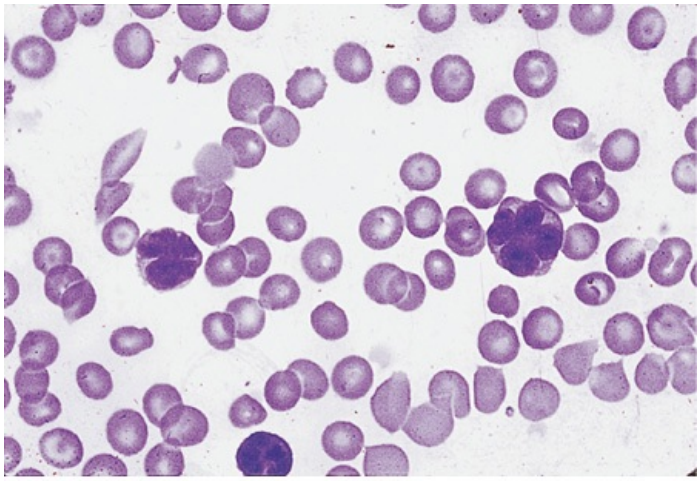

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adult T cell leukemia. Peripheral blood smear showing leukemia cells with typical "flower-shaped" nucleus.

Figure e17-49

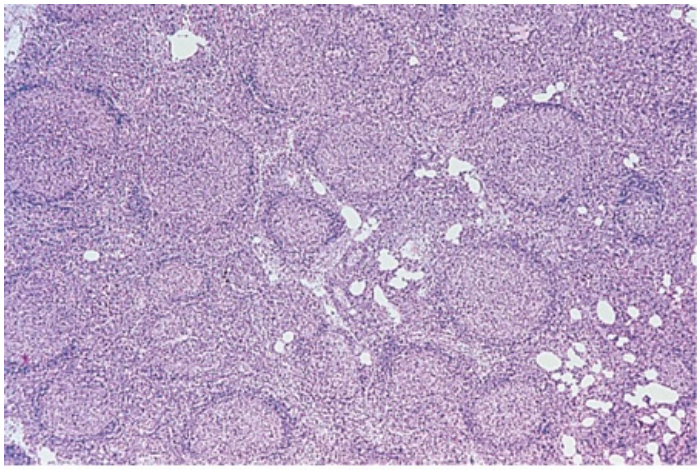

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Follicular lymphoma in a lymph node. The normal nodal architecture is effaced by nodular expansions of tumor cells. Nodules vary in size and contain predominantly small lymphocytes with cleaved nuclei along with variable numbers of larger cells with vesicular chromatin and prominent nucleoli.

Figure e17-50

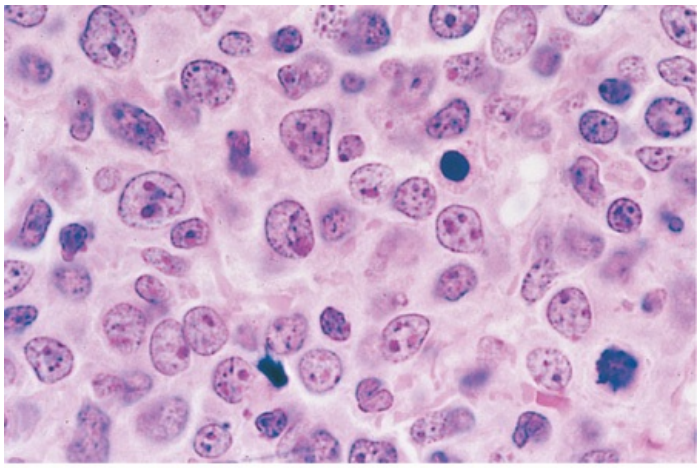

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diffuse large B cell lymphoma in a lymph node. The neoplastic cells are heterogeneous but predominantly large cells with vesicular chromatin and prominent nucleoli.

Figure e17-51

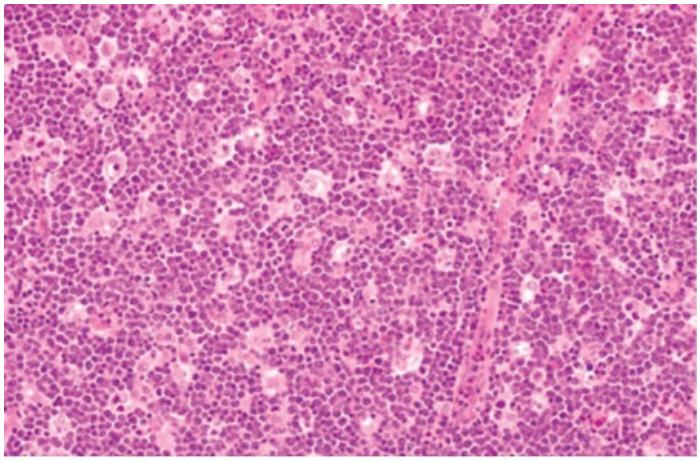

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Burkitt's lymphoma in a lymph node. Burkitt's lymphoma with starry-sky appearance. The lighter areas are macrophages attempting to clear dead cells.

Figure e17-52

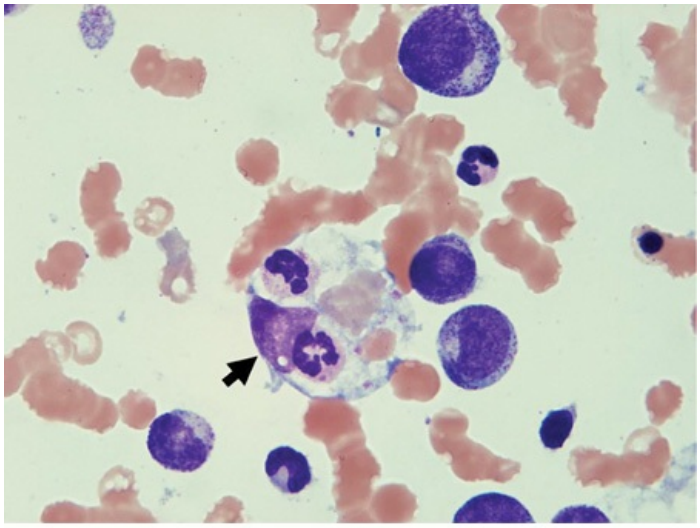

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythrophagocytosis accompanying aggressive lymphoma. The central macrophage is ingesting red cells, neutrophils, and platelets. *(Courtesy of Dr. Kiyomi Tsukimori, Kyushu University, Fukuoka, Japan.)*

Figure e17-53

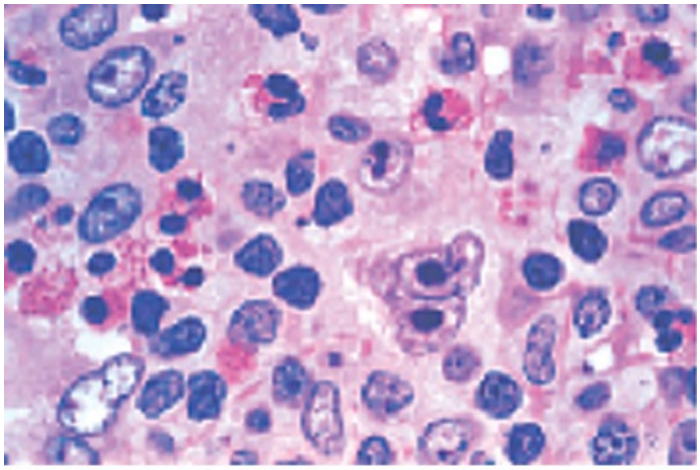

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hodgkin's disease. A Reed-Sternberg cell is present near the center of the field; a large cell with a bilobed nucleus and prominent nucleoli giving an "owl's eyes" appearance. The majority of the cells are normal lymphocytes, neutrophils, and eosinophils that form a pleiomorphic cellular infiltrate.

Figure e17-54

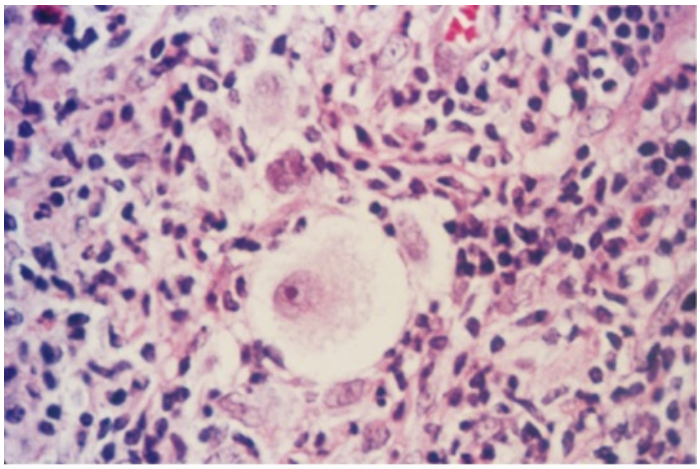

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lacunar cell; Reed-Sternberg cell variant in nodular sclerosing Hodgkin's disease. High-power view of single mononuclear lacunar cell with retracted cytoplasm in a patient with nodular sclerosing Hodgkin's disease.

Figure e17-55

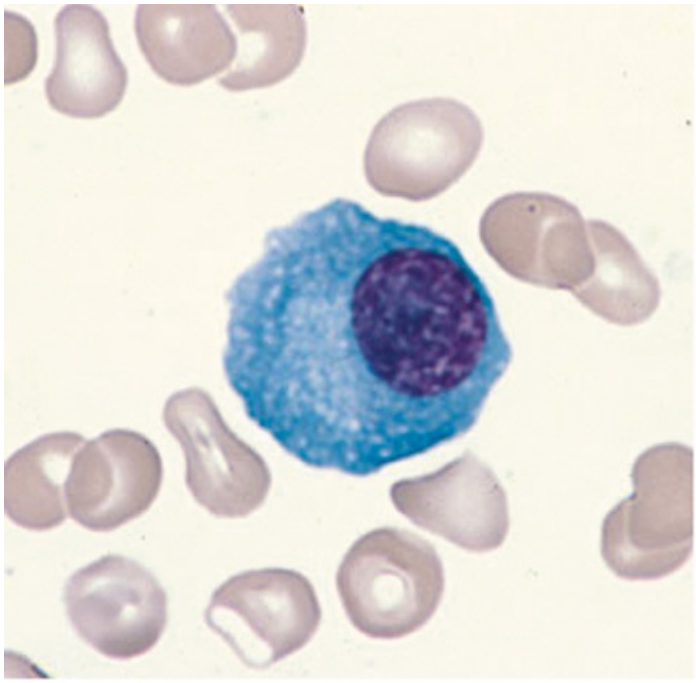

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal plasma cell.

Figure e17-56

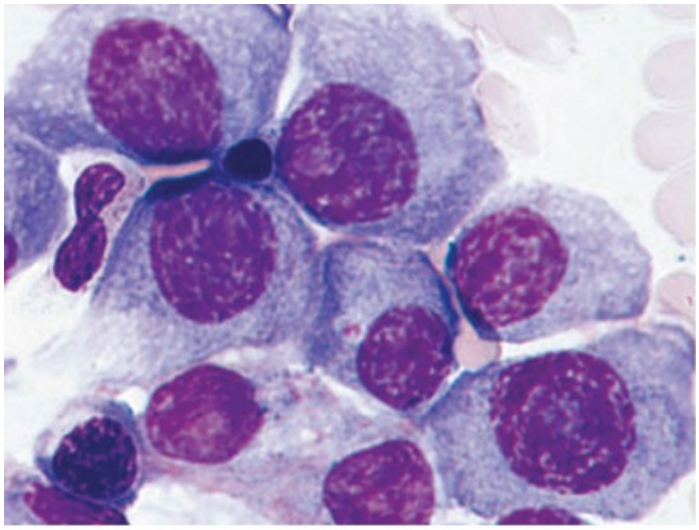

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple myeloma.

Figure e17-57

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Color serum in hemoglobinemia. The distinctive red coloration of plasma (hemoglobinemia) in a spun blood sample in a patient with intravascular hemolysis.

ACKNOWLEDGMENT

Figures in this e-chapter were borrowed from *Williams Hematology*, 7th edition, M Lichtman et al (eds). New York, McGraw-Hill, 2005; *Hematology in General Practice*, 4th edition, RS Hillman, KA Ault, New York, McGraw-Hill, 2005.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

IMPACT OF GENETICS ON MEDICAL PRACTICE

The beginning of the new millennium was marked by the announcement that the vast majority of the human genome had been sequenced. This milestone in the exploration of the human genome was preceded by numerous conceptual and technologic advances. They include, among others, the elucidation of the DNA double-helix structure, the discovery of restriction enzymes and the polymerase chain reaction (PCR), the development and automatization of DNA sequencing, and the generation of genetic and physical maps by the Human Genome Project (HGP). The consequences of this wealth of knowledge for the practice of medicine are profound. First, the most significant impact of genetics has been to enhance our understanding of disease etiology and pathogenesis. However, genetics is playing an increasingly prominent role in the diagnosis, prevention, and treatment of disease (Chap. 63). Genetic approaches have proven invaluable for the detection of infectious pathogens and are used clinically to identify agents that are difficult to culture such as mycobacteria, viruses, and parasites. In many cases, molecular genetics has improved the feasibility and accuracy of diagnostic testing and is beginning to open new avenues for therapy, including gene and cellular therapy (Chaps. 68 and 67). Molecular genetics has significantly changed the treatment of human disease. Peptide hormones, growth factors, cytokines, and vaccines can now be produced in large amounts using recombinant DNA technology. Targeted modifications of these peptides provide the practitioner with improved therapeutic tools, as illustrated by genetically modified insulin analogues with more favorable kinetics. There is hope that a better understanding of the genetic basis of human disease will also have an increasing impact on disease prevention.

Genetics has traditionally been viewed through the window of relatively rare single-gene diseases. Taken together, these disorders account for ~10% of pediatric admissions and childhood mortality. It is, however, increasingly apparent that virtually every medical condition has a genetic component. As is often evident from a patient's family history, many common disorders such as hypertension, heart disease, asthma, diabetes mellitus, and mental illnesses are significantly influenced by the genetic background. These polygenic or multifactorial (complex) disorders involve the contributions of many different genes, as well as environmental factors that can modify disease risk (Chap. 63). Genome-wide association studies (GWAS) have elucidated numerous disease-associated loci and are providing novel insights into the allelic architecture of complex traits. These studies have been facilitated by the availability of comprehensive catalogues of human single-nucleotide polymorphism (SNP) haplotypes generated through the HapMap Project.

Cancer has a genetic basis since it results from acquired somatic mutations in genes controlling growth, apoptosis, and cellular differentiation (Chap. 83). In addition, the development of many cancers is associated with a hereditary predisposition. The prevalence of genetic diseases, combined with their severity and chronic nature, imposes great financial, social, and emotional burdens on society.

Genetics has historically focused predominantly on chromosomal and metabolic disorders, reflecting the long-standing availability of techniques to diagnose these conditions. For example, conditions such as trisomy 21 (Down syndrome) or monosomy X (Turner's syndrome) can be diagnosed using cytogenetics (Chap. 62). Likewise, many metabolic disorders (e.g., phenylketonuria, familial hyper–cholesterolemia) are diagnosed using biochemical analyses. Recent advances in DNA diagnostics have extended the field of genetics to include virtually all medical specialties. In cardiology, for example, the molecular basis of inherited cardiomyopathies and ion channel defects that predispose to arrhythmias is being defined (Chaps. 233 and 238). In neurology, genetics has unmasked the pathophysiology of a startling number of neurodegenerative disorders (Chap. 366). Hematology has evolved dramatically, from its incipient genetic descriptions of hemoglobinopathies to the current understanding of the molecular basis of red cell membrane defects, clotting disorders, and thrombotic disorders (Chaps. 104 and 116).

New concepts derived from genetic studies can sometimes clarify the pathogenesis of disorders that were previously opaque. For example, although many different genetic defects can cause peripheral neuropathies, disruption of the normal folding of the myelin sheaths is frequently a common final pathway (Chap. 384). Several genetic causes of obesity appear to converge on a physiologic pathway that involves products of the proopiomelanocortin polypeptide and the MC4R receptor, thus identifying a key mechanism for appetite control (Chap. 77). A similar phenomenon is emerging for genetically distinct forms of Alzheimer's disease, several of which lead to the formation of neurofibrillary tangles (Chap. 371). The identification of defective genes often leads to the detection of cellular pathways

involved in key physiologic processes. Examples include identification of the cystic fibrosis conductance regulator (*CFTR*) gene; the Duchenne's muscular dystrophy (*DMD*) gene, which encodes dystrophin; and the fibroblast growth factor receptor-3 (*FGFR3*) gene, which is responsible for achondroplastic dwarfism. Similarly, transgenic (over)expression, and targeted gene "knock-out" and "knock-in" models help to unravel the physiologic function of genes.

The astounding rate at which new genetic information is being generated creates a major challenge for physicians, health care providers, and basic investigators. The terminology and techniques used for discovery evolve continuously. Much genetic information resides in databases or is being published in basic science journals. Databases provide easy access to the expanding information about the human genome, genetic disease, and genetic testing **(Table 61-1)**. For example, several thousand monogenic disorders are summarized in a large, continuously evolving compendium, referred to as the *Online Mendelian Inheritance in Man* (OMIM) catalogue (Table 61-1). The ongoing refinement of bioinformatics is simplifying the access to this daunting onslaught of new information.

Table 61-1 Selected Databases Relevant for Genomics and Genetic Disorders

Site	URL	Comment
National Center for Biotechnology Information (NCBI)	*http://www.ncbi.nlm.nih.gov.proxy.cc.uic.edu/*	Broad access to biomedical and genomic information, literature (PubMed), sequence databases, software for analyses of nucleotides and proteins Extensive links to other databases, genome resources, and tutorials
National Human Genome Research Institute	*http://www.genome.gov/*	Web links providing information about the human genome sequence, genomes of other organisms, and genomic research
Ensembl Genome browser	*http://www.ensembl.org*	Maps and sequence information of eukaryotic genomes
Online Mendelian Inheritance in Man	*http://www.ncbi.nlm.nih.gov.proxy.cc.uic.edu/omim*	Online compendium of Mendelian disorders and human genes causing genetic disorders
Office of Biotechnology Activities National Institutes of Health	*http://oba.od.nih.gov.proxy.cc.uic.edu/oba*	Information about recombinant DNA and gene transfer Medical, ethical, legal, and social issues raised by genetic testing Medical, ethical, legal, and social issues raised by xenotransplantation
American College of Medical Genetics	*http://www.acmg.net/*	Extensive links to other databases relevant for the diagnosis

		treatment, and prevention of genetic disease
Cancer Genome Anatomy Project (CGAP)	*http://cgap.nci.nih.gov.proxy.cc.uic.edu/*	Information about gene expression profiles of normal, precancer, and cancer cells
GeneTests	*http://www.genetests.org/*	International directory of genetic testing laboratories and prenatal diagnosis clinics Reviews and educational materials
Genomes Online Database (GOLD)	*http://www.genomesonline.org/*	Information on published and unpublished genomes
HUGO Gene Nomenclature	*http://www.genenames.org/*	Gene names and symbols
MITOMAP, a human mitochondrial genome database	*http://www.mitomap.org/*	A compendium of polymorphisms and mutations of the human mitochondrial DNA
Mitochondrial disorders	*http://www.neuro.wustl.edu/neuromuscular/mitosyn.html*	Overview on clinical syndromes associated with mtDNA mutations
DNA repeat sequences & disease	*http://www.neuro.wustl.edu/neuromuscular/mother/dnarep.htm*	Overview on clinical syndromes associated with DNA repeats
Online Mendelian Inheritance in Animals (OMIA)	*http://omia.angis.org.au/*	Online compendium of Mendelian disorders in animals
The Jackson Laboratory	*http://www.jax.org/*	Information about murine models and the mouse genome
International HapMap Project	*http://www.hapmap.org/*	Catalogue of haplotypes in different ethnic groups relevant for association studies and pharmacogenomics
Nuclear Receptor Signaling Atlas	*http://nursa.org*	Atlas of nuclear receptors, coregulators, and ligands
Dolan DNA Learning Center, Cold Spring Harbor Laboratories	*http://www.dnalc.org/*	Educational material about selected genetic disorders, DNA, eugenics, and genetic origin
The Online Metabolic and Molecular Bases of Inherited Disease (OMMBID)	*http://www.ommbid.com/*	Online version of the comprehensive text on The Metabolic and Molecular Bases of Inherited Disease

Note: Databases are evolving constantly. Pertinent information may be found by using links listed in

the few selected databases. Instructions for the use of genome-related databases have been published [Nat Genet 32(Suppl):1–79, 2002].

Chromosomes and DNA Replication

ORGANIZATION OF DNA INTO CHROMOSOMES

Size of the Human Genome

The human genome is divided into 23 different chromosomes, including 22 autosomes (numbered 1–22) and the X and Y sex chromosomes. Adult cells are *diploid*, meaning they contain two homologous sets of 22 autosomes and a pair of sex chromosomes. Females have two X chromosomes (XX), whereas males have one X and one Y chromosome (XY). As a consequence of meiosis, germ cells (sperm or oocytes) are haploid and contain one set of 22 autosomes and one of the sex chromosomes. At the time of fertilization, the diploid genome is reconstituted by pairing of the homologous chromosomes from the mother and father. With each cell division (mitosis), chromosomes are replicated, paired, segregated, and divided into two daughter cells (Chap. 62).

Although the exact number of genes encoded by the human genome is still unknown, current estimates predict about 23,000 to 25,000 protein-coding genes, a number that is substantially smaller than initially predicted. A *gene* is a functional unit that is regulated by transcription (see below) and encodes an RNA product, which is most commonly, but not always, translated into a protein that exerts activity within or outside the cell. Historically, genes were identified because they conferred specific traits that are transmitted from one generation to the next. Increasingly, they are characterized based on expression in various tissues (transcriptome). The number of genes greatly underestimates the complexity of genetic expression, as single genes can generate multiple spliced messenger RNA (mRNA) products, which are translated into proteins that are subject to complex posttranslational modification such as phosphorylation. *Proteomics,* the study of the proteome using technologies of large-scale protein separation and identification, is focused on protein variation and function. Similarly, the field of *metabolomics* aims at determining the composition and modifications of the *metabolome*, the complement of low-molecular-weight molecules, many of which participate in various metabolic functions. The human *microbiome* refers to the constellation of viruses, bacteria, and fungi that colonize various human tissues (Chap. 64). Comprehensive characterization of the microbiome has been made feasible by the availability of high-throughput DNA-sequencing methods. Analyses of genomics, proteomics, metabolomics, and the microbiome are heavily dependent on bioinformatics, and they are beginning to reveal how physiologic or pathologic alterations affect *modular networks* rather than *linear pathways* (Chap. e19).

Human DNA consists of ~3 billion base pairs (bp) of DNA per haploid genome. DNA length is normally measured in units of 1000 bp (kilobases, kb) or 1,000,000 bp (megabases, Mb). Not all DNA encodes genes. In fact, genes account for only ~10–15% of DNA. Much of the remaining DNA consists of highly repetitive sequences, the function of which is poorly understood. These repetitive DNA regions, along with nonrepetitive sequences that do not encode genes, may serve a structural role in the packaging of DNA into chromatin [i.e., DNA bound to histone proteins, and chromosomes **(Fig. 61-1)**]. If only 10% of DNA is expressed and there are 25,000 genes, the average gene would be ~12 kb in length. Although many genes are about this size, the range is quite broad. For example, some genes are only a few hundred bp, whereas others such as the *DMD* gene, are extraordinarily large (2 Mb).

Figure 61-1

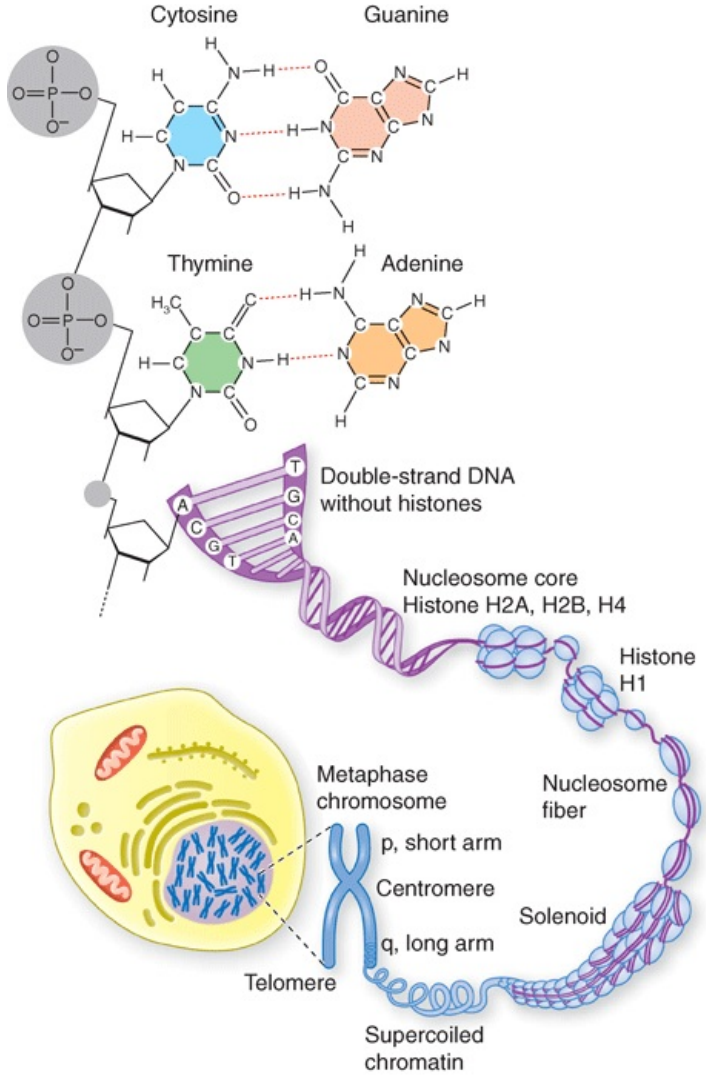

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Structure of chromatin and chromosomes. Chromatin is composed of double-strand DNA that is wrapped around histone and non–histone proteins forming nucleosomes. The nucleosomes are further organized into solenoid structures. Chromosomes assume their characteristic structure, with short (p) and long (q) arms at the metaphase stage of the cell cycle.

Structure of DNA

Each gene is composed of a linear polymer of DNA. DNA is a double-stranded helix composed of four different bases: adenine (A), thymidine (T), guanine (G), and cytosine (C). Adenine is paired to thymidine, and guanine is paired to cytosine, by hydrogen bond interactions that span the double helix. DNA has several remarkable features that make it ideal for the transmission of genetic information. It is relatively stable, at least in comparison to RNA or proteins. The double-stranded nature of DNA and its feature of strict base-pair complementarity permit faithful replication during cell division. As described below, complementarity also allows the transmission of genetic information from DNA → RNA → protein **(Fig. 61-2)**. mRNA is encoded by the so-called sense or coding strand of the DNA double helix and is translated into proteins by ribosomes.

Figure 61-2

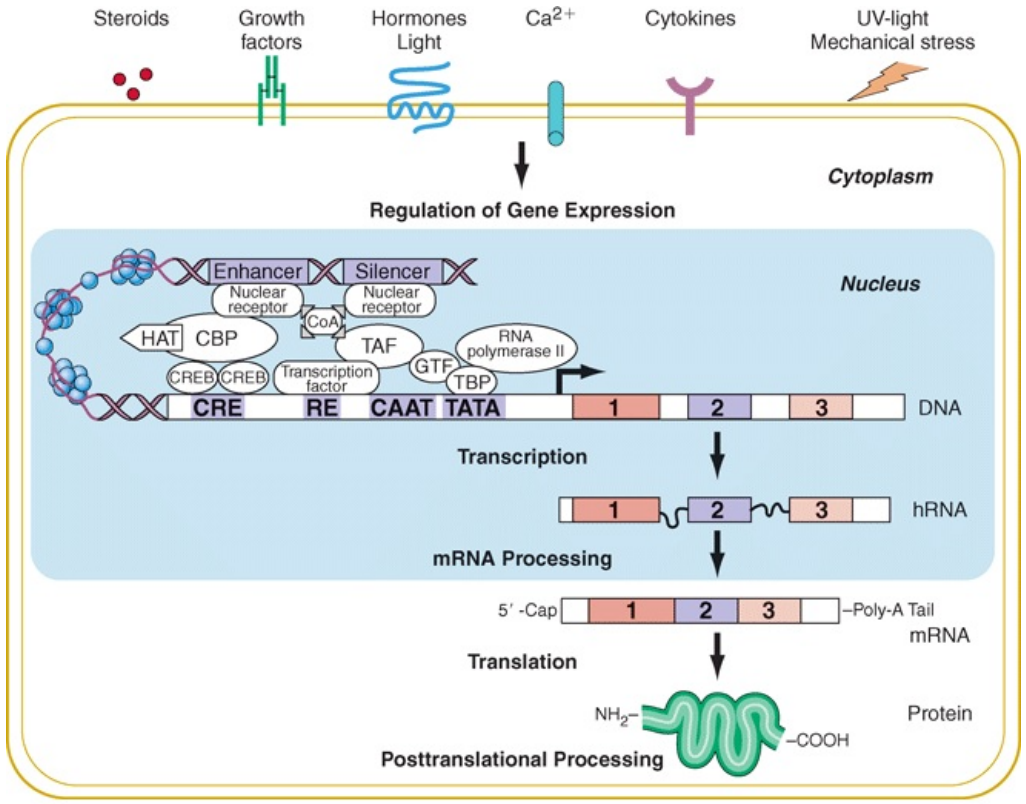

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flow of genetic information. Multiple extracellular signals activate intracellular signal cascades that result in altered regulation of gene expression through the interaction of transcription factors with regulatory regions of genes. RNA polymerase transcribes DNA into RNA that is processed to mRNA by excision of intronic sequences. The mRNA is translated into a polypeptide chain to form the mature protein after undergoing posttranslational processing. HAT, histone acetyl transferase; CBP, CREB-binding protein; CREB, cyclic AMP response element–binding protein; CRE, cyclic AMP responsive element; CoA, Co activator; TAF, TBP-associated factors; GTF, general transcription factors; TBP, TATA-binding protein; TATA, TATA box; RE, response element; NH_2, aminoterminus; COOH, carboxyterminus.

The presence of four different bases provides surprising genetic diversity. In the protein-coding regions of genes, the DNA bases are arranged into codons, a triplet of bases that specifies a particular amino acid. It is possible to arrange the four bases into 64 different triplet codons (4^3). Each codon specifies 1 of the 20 different amino acids, or a regulatory signal such as initiation and stop of translation. Because there are more codons than amino acids, the genetic code is degenerate; that is, most amino acids can be specified by several different codons. By arranging the codons in different combinations and in various lengths, it is possible to generate the tremendous diversity of primary protein structure.

REPLICATION OF DNA AND MITOSIS

Genetic information in DNA is transmitted to daughter cells under two different circumstances: (1) somatic cells divide by mitosis, allowing the diploid ($2n$) genome to replicate itself completely in conjunction with cell division; and (2) germ cells (sperm and ova) undergo meiosis, a process that enables the reduction of the diploid ($2n$) set of chromosomes to the haploid state ($1n$) (Chap. 62).

Prior to mitosis, cells exit the resting, or G_0 state, and enter the cell cycle (Chap. 84). After traversing a critical checkpoint in G_1, cells undergo DNA synthesis (S phase), during which the DNA in each chromosome is replicated, yielding two pairs of sister chromatids ($2n \rightarrow 4n$). The process of DNA synthesis requires stringent fidelity in order to avoid transmitting errors to subsequent generations of cells. Genetic abnormalities of DNA mismatch/repair include xeroderma pigmentosum, Bloom's syndrome, ataxia telangiectasia, and hereditary nonpolyposis colon cancer (HNPCC), among others. Many of these disorders strongly predispose to neoplasia because of the rapid acquisition of additional mutations (Chap. 83). After completion of DNA synthesis, cells enter G_2 and progress through a second checkpoint before entering mitosis. At this stage, the chromosomes condense and are aligned along the equatorial plate at metaphase. The two identical sister chromatids, held together at the centromere, divide and migrate to opposite poles of the cell (Fig. 61-3). After formation of a nuclear membrane around the two separated sets of chromatids, the cell divides and two daughter cells are formed, thus restoring the diploid ($2n$) state.

Figure 61-3

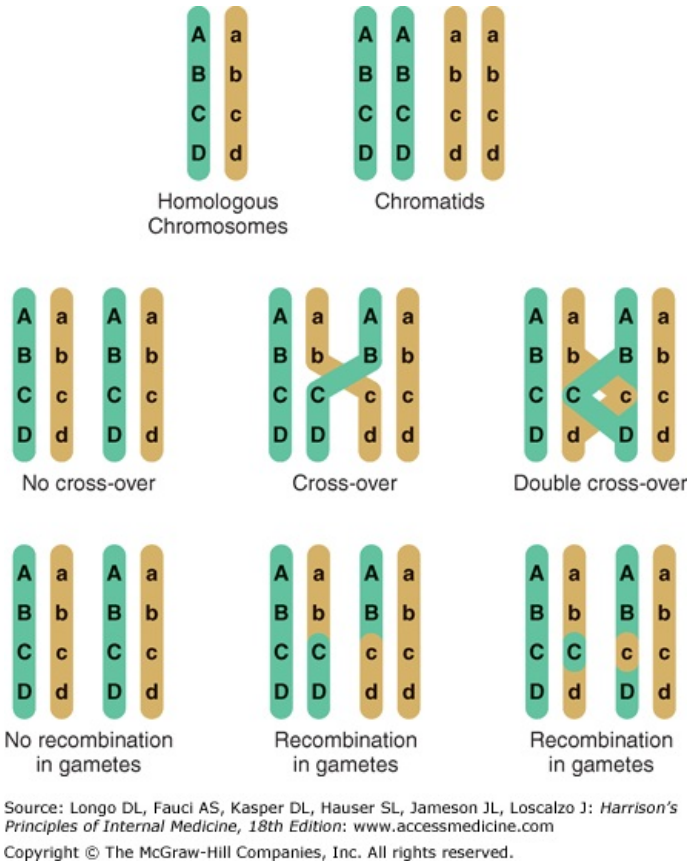

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Crossing-over and genetic recombination. During chiasma formation, either of the two sister chromatids on one chromosome pairs with one of the chromatids of the homologous chromosome. Genetic recombination occurs through crossing-over and results in recombinant and nonrecombinant chromosome segments in the gametes. Together with the random segregation of the maternal and paternal chromosomes, recombination contributes to genetic diversity and forms the basis of the concept of linkage.

ASSORTMENT AND SEGREGATION OF GENES DURING MEIOSIS

Meiosis occurs only in germ cells of the gonads. It shares certain features with mitosis but involves two distinct steps of cell division that reduce the chromosome number to the haploid state. In addition, there is active recombination that generates genetic diversity. During the first cell division, two sister chromatids ($2n \rightarrow 4n$) are formed for each chromosome pair and there is an exchange of DNA between homologous paternal and maternal chromosomes. This process involves the formation of *chiasmata*, structures that correspond to the DNA segments that cross over between the maternal and paternal homologues **(Fig. 61-3)**. Usually there is at least one crossover on each chromosomal arm; recombination occurs more frequently in female meiosis than in male meiosis. Subsequently, the chromosomes segregate randomly. Because there are 23 chromosomes, there exist 2^{23} (>8 million) possible combinations of chromosomes. Together with the genetic exchanges that occur during recombination, chromosomal segregation generates tremendous diversity, and each gamete is genetically unique. The process of recombination, and the independent segregation of chromosomes, provide the foundation for performing linkage analyses, whereby one attempts to correlate the inheritance of certain chromosomal regions (or linked genes) with the presence of a disease or genetic trait (see below).

After the first meiotic division, which results in two daughter cells ($2n$), the two chromatids of each chromosome separate during a second meiotic division to yield four gametes with a haploid state ($1n$). When the egg is fertilized by sperm, the two haploid sets are combined, thereby restoring the diploid state ($2n$) in the zygote.

Regulation of Gene Expression

Mechanisms that regulate gene expression play a critical role in the function of genes. The transcription of genes is controlled primarily by *transcription factors* that bind to DNA sequences in the regulatory regions of genes. As described below, mutations in transcription factors cause a significant number of genetic disorders. Gene expression is also influenced by *epigenetic events (epigenetics)*. DNA and histone modifications can result in the activation or silencing of gene expression **(Fig. 61-4)**.

They include heritable changes such as X-inactivation and imprinting, and dynamic alterations in response to environmental influences such as diet, age, or drugs. Epigenomics addresses epigenetic changes across the whole genome in a cell or organism. Several genetic disorders such as Prader-Willi syndrome (neonatal hypotonia, developmental delay, obesity, short stature, and hypogonadism) and Albright's hereditary osteodystrophy (resistance to parathyroid hormone, short stature, brachydactyly, resistance to other hormones in certain subtypes), exhibit the consequences of genomic imprinting. Acquired modifications in DNA methylation occur during development and are found in cancer cells, likely contributing to cancer development and progression. Most studies of gene expression have focused on the regulatory DNA elements of genes that control transcription. However, it should be emphasized that gene expression requires a series of steps, including mRNA processing, protein translation, and posttranslational modifications, all of which are actively regulated (Fig. 61-2).

Figure 61-4

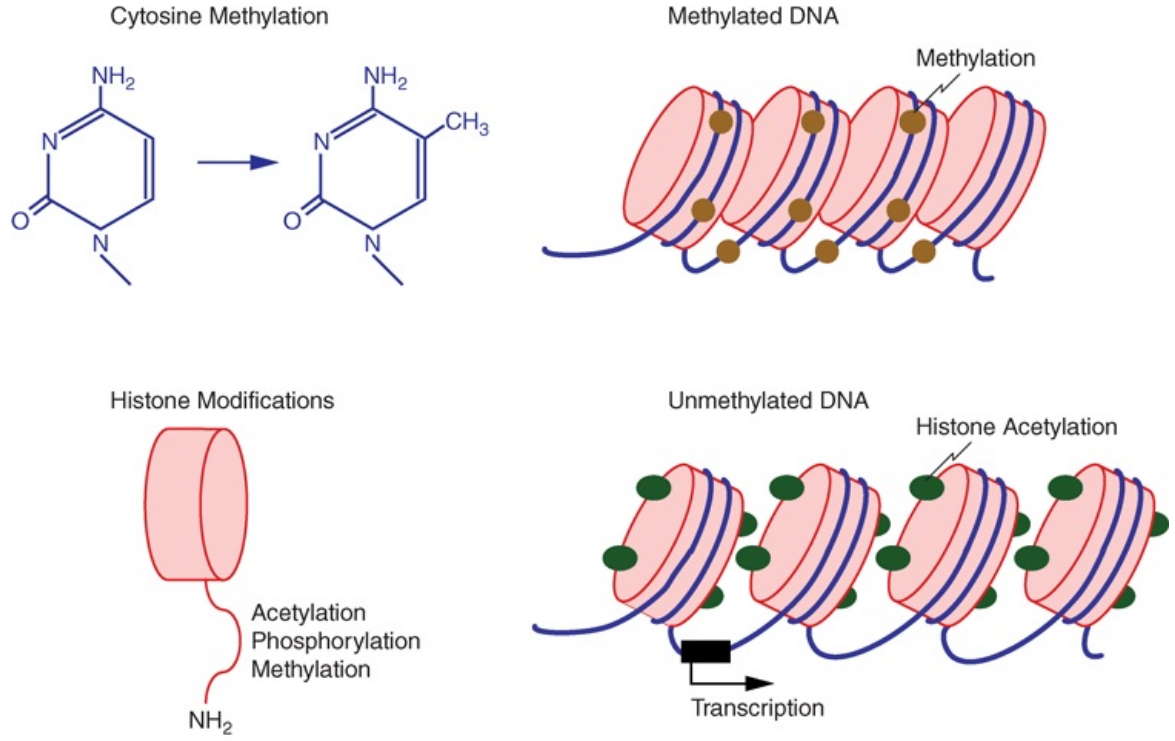

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Epigenetic modifications of DNA and histones. Methylation of cytosine residues is associated with gene silencing. Methylation of certain genomic regions is inherited (imprinting) and is involved in the silencing of one of the two X chromosomes in females (X-inactivation). Alterations in methylation can also be acquired (e.g. in cancer cells). Covalent posttranslational modifications of histones and other proteins play an important role in altering chromatin structure and, hence, transcription. Histones can be reversibly modified in their amino-terminal tails, which protrude from the nucleosome core particle, by acetylation of lysine, phosphorylation of serine, or methylation of lysine and arginine residues. Acetylation of histones by histone acetylases (HATs), for example, leads to unwinding of chromatin and accessibility to transcription factors. Conversely, deacetylation by histone deacetylases (HDACs) results in a compact chromatin structure and silencing of transcription.

The field of *functional genomics* is based on the concept that understanding alterations of gene expression under various physiologic and pathologic conditions provides insight into the underlying processes, and by revealing certain gene expression profiles, this knowledge may be of diagnostic and therapeutic relevance. The large-scale study of expression profiles, which takes advantage of microarray and bead array technologies, is also referred to as *transcriptomics* because the complement of mRNAs transcribed by the cellular genome is called the *transcriptome*.

STRUCTURE OF GENES

A gene product is usually a protein but can occasionally consist of RNA that is not translated (e.g., microRNAs). *Exons* refer to the portion of genes that are eventually spliced together to form mRNA. *Introns* refer to the spacing regions between the exons that are spliced out of precursor RNAs during RNA processing. The gene locus also includes regions that are necessary to control its expression (Fig. 61-2).

The number of DNA sequences and transcription factors that regulate transcription is much greater than originally anticipated. Most genes contain at least 15–20 discrete regulatory elements within 300

bp of the transcription start site. This densely packed promoter region often contains binding sites for ubiquitous transcription factors such as CAAT box/enhancer binding protein (C/EBP), cyclic AMP response element–binding (CREB) protein, selective promoter factor 1 (Sp-1), or activator protein 1 (AP-1). However, factors involved in cell-specific expression may also bind to these sequences. For example, basic helix-loop-helix (bHLH) proteins bind to E-boxes in the promoters of myogenic genes, and steroidogenic factor 1 (SF-1) binds to a specific recognition site in the regulatory region of multiple steroidogenic enzyme genes. Key regulatory elements may also reside at a large distance from the proximal promoter. The globin and the immunoglobulin genes, (e.g., contain *locus control regions* that are several kilobases away from the structural sequences of the gene). Specific groups of transcription factors that bind to these promoter and enhancer sequences provide a combinatorial code for regulating transcription. In this manner, relatively ubiquitous factors interact with more restricted factors to allow each gene to be expressed and regulated in a unique manner that is dependent on developmental state, cell type, and numerous extracellular stimuli. The transcription factors that bind to DNA actually represent only the first level of regulatory control. Other proteins—*co-activators* and *co-repressors*—interact with the DNA-binding transcription factors to generate large regulatory complexes. These complexes are subject to control by numerous cell-signaling pathways, including phosphorylation, acetylation, sumoylation, and ubiquitination. Ultimately, the recruited transcription factors interact with, and stabilize, components of the basal transcription complex that assembles at the site of the TATA box and initiator region. This basal transcription factor complex consists of >30 different proteins. Gene transcription occurs when RNA polymerase begins to synthesize RNA from the DNA template. A growing number of identified genetic diseases involve transcription factors **(Table 61-2)**.

Table 61-2 Selected Examples of Diseases Caused by Mutations and Rearrangements in Transcription Factor Classes

Transcription Factor Class	Example	Associated Disorder
Nuclear receptors	Androgen receptor	Complete or partial androgen insensitivity (recessive missense mutations) Spinobulbar muscular atrophy (CAG repeat expansion)
Zinc finger proteins	WT1	WAGR syndrome: Wilms' tumor, aniridia, genitourinary malformations, mental retardation
Basic helix-loop-helix	MITF	Waardenburg's syndrome type 2A
Homeobox	IPF1	Maturity onset of diabetes mellitus type 4 (heterozygous mutation/haploinsufficiency) Pancreatic agenesis (homozygous mutation)
Leucine zipper	Retina leucine zipper (NRL)	Autosomal dominant retinitis pigmentosa
High mobility group (HMG) proteins	SRY	Sex-reversal
Fork-head	HNF4α, HNF1α, HNF1β	Maturity-onset of diabetes mellitus types 1, 3, 5
Paired box	PAX3	Waardenburg's syndrome types 1 and 3
T-box	TBX5	Holt-Oram syndrome (thumb anomalies, atrial or ventricular septum defects, phocomelia)
Cell cycle control proteins	P53	Li-Fraumeni syndrome, other cancers
Coactivators	CREB binding protein (CBP)	Rubinstein-Taybi syndrome
General transcription factors	TATA-binding protein (TBP)	Spinocerebellar ataxia 17 (CAG expansion)
Transcription elongation factor	VHL	von Hippel–Lindau syndrome (renal cell carcinoma, pheochromocytoma, pancreatic tumors, hemangioblastomas) Autosomal dominant inheritance, somatic inactivation of second allele (Knudson two-hit model)
Runt	CBFA2	Familial thrombocytopenia with propensity to acute myelogenous leukemia
Chimeric pro-teins due to	PML–RAR	Acute promyelocytic leukemia t(15;17)(q22;q11.2-q12) translocation

translocations		

Abbreviations: CREB, cAMP responsive element–binding protein; HNF, hepatocyte nuclear factor; PML, -promyelocytic leukemia; RAR, retinoic acid receptor; SRY, sex-determining region Y; VHL, von Hip-pel–Lindau.

Mutations can occur in all domains of a gene **(Fig. 61-5)**. A point mutation occurring within the coding region leads to an amino acid substitution if the codon is altered. Point mutations that introduce a premature stop codon result in a truncated protein. Large deletions may affect a portion of a gene or an entire gene, whereas small deletions and insertions alter the reading frame if they do not represent a multiple of three bases. These "frameshift" mutations lead to an entirely altered carboxy terminus. Mutations occurring in regulatory or intronic regions may result in altered expression or splicing of genes. Examples are shown in **Fig. 61-6**.

Figure 61-5

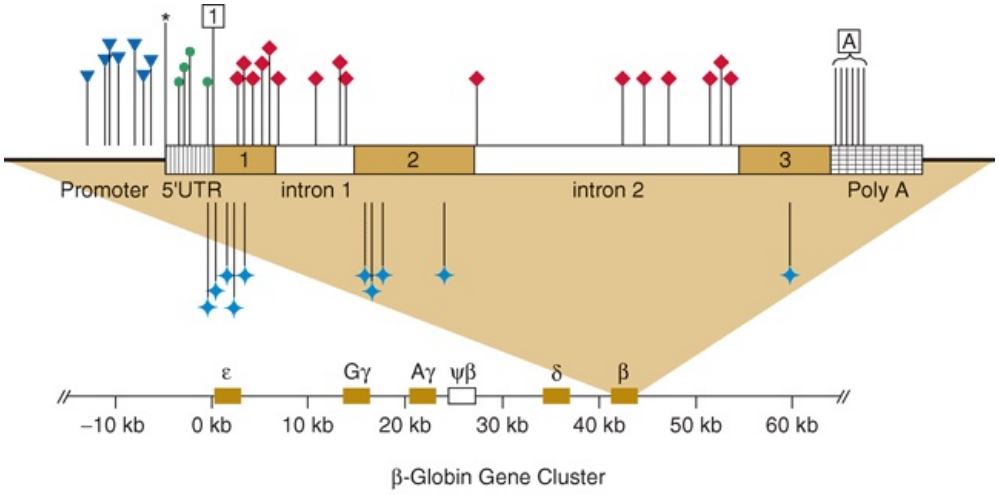

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Point mutations causing β-thalassemia as example of allelic heterogeneity. The β-globin gene is located in the globin gene cluster. Point mutations can be located in the promoter, the CAP site, the 5'-untranslated region, the initiation codon, each of the three exons, the introns, or the polyadenylation signal. Many mutations introduce missense or nonsense mutations, whereas others cause defective RNA splicing. Not shown here are deletion mutations of the β-globin gene or larger deletions of the globin locus that can also result in thalassemia. ▾, Promoter mutations; *, CAP site; •, 5'UTR; 1, Initiation codon; ◆, Defective RNA processing; ✦, Missense and nonsense mutations; A, Poly A signal.

Figure 61-6

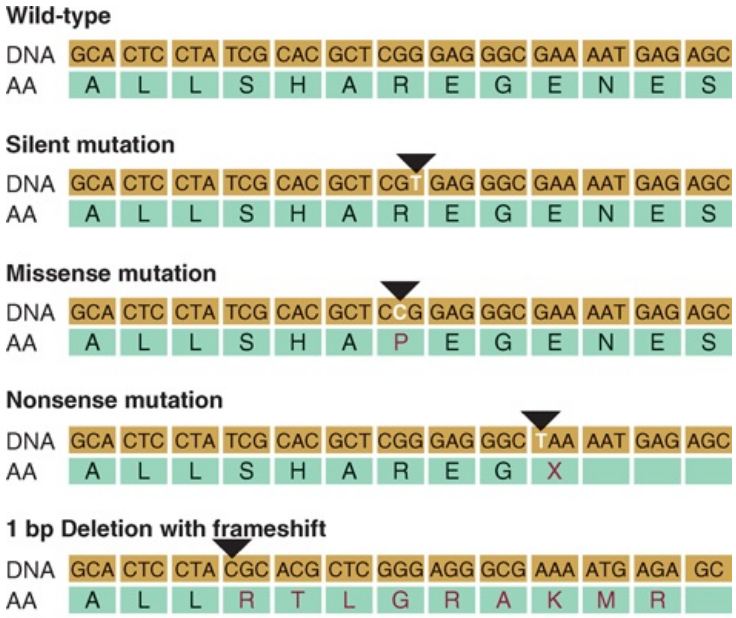

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

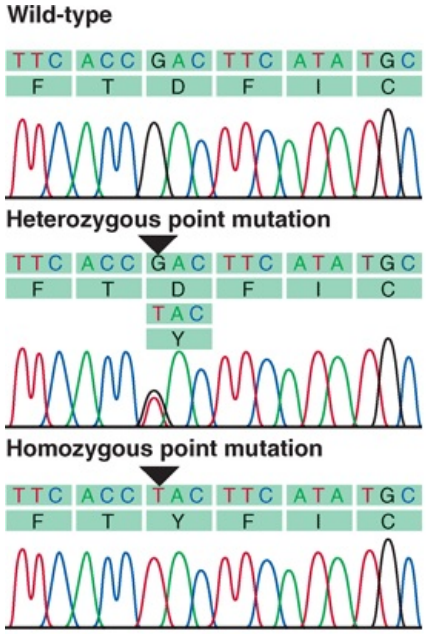

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A*.Examples of mutations.** The coding strand is shown with the encoded amino acid sequence. ***B. Chromatograms of sequence analyses after amplification of genomic DNA by polymerase chain reaction.

Cloning and Sequencing DNA

A description of recombinant DNA techniques, the methodology used for the manipulation, analysis, and characterization of DNA segments, is beyond the scope of this chapter. These methods are now widely used in genetics and molecular diagnostics.

Transgenic Mice as Models of Genetic Disease

Several organisms have been studied extensively as genetic models, including *Mus musculus* (mouse), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (nematode), *Saccharomyces cerevisiae* (baker's yeast), and *Escherichia coli* (colonic bacterium). The ability to use these evolutionarily distant organisms as genetic models that are relevant to human physiology reflects a

surprising conservation of genetic pathways and gene function. Transgenic mouse models have been particularly valuable, because many human and mouse genes exhibit similar structure and function, and because manipulation of the mouse genome is relatively straightforward compared to those of other mammalian species.

Transgenic strategies in mice can be divided into two main approaches: (1) expression of a gene by random insertion into the genome, and (2) deletion or targeted mutagenesis of a gene by homologous recombination with the native endogenous gene (knock-out, knock-in) **(Table 61-3)**.

Table 61-3 Genetically Modified Animals

Commonly Used Descrip-tion	Technical Principle	Remarks
Transgenic	Pronuclear injection of transgene	Commonly used Genomic DNA or cDNA constructs Random integration of transgene Variable copy numbers of transgene Variable expression in each individual founder Gain-of-function models due to overexpression using tissue-specific promoters Loss-of-function models using antisense and dominant negative transgenes Inducible expression possible (tetracycline, ecdysone) Applicable to several species
(Targeted) Knock-out	Substitution of functional gene with inactive gene by homologous recombination in embryonic stem cells	Predominantly used in mice Tissue-specific knock-out possible (Cre/lox) Absence of phenotype possible due to redundancy
(Targeted) Knock-in	Introduction of subtle mutation(s) into gene by substitution of endogenous gene with gene carrying a specific mutation. Homologous recombination in embryonic stem cells	Predominantly used in mice Can accurately model human disease
Forward genetics	Mutations created randomly by ENU (*N*-ethyl-*N*-nitrourea)	Selection of phenotype followed by genetic characterization Useful for identifying novel genes
Congenic strains	Mating of an inbred *donor* strain with a disease phenotype with an inbred *recipient* strain in order to define the genomic region responsible for the disorder	Useful for mapping disease-causing genes
Cloning	Introduction of nucleus into enucleated eggs (nuclear transfer)	Successful in several mammalian species including sheep (Dolly), mice, cows, monkeys Cloning of genetically identical individuals May affect lifespan Ethical concerns

Transgenic expression of genes can be useful for studying disorders that are sensitive to gene

dosage. Overexpression of *PMP22*, for example, mimics a common duplication of this gene in type IA Charcot-Marie-Tooth disease (Chap. 384). Duplication of the *PMP22* gene results in high levels of expression of peripheral myelin protein 22, and this dosage effect is responsible for the demyelinating neuropathy. Expression of the Y chromosome–specific gene, *SRY*, in XX females demonstrates that *SRY* is sufficient to induce the formation of testes. This finding confirms the pathogenic role of *SRY* translocations to the X chromosome in sex-reversed XX females. Huntington's disease is an autosomal dominant disorder caused by expansion of a CAG trinucleotide repeat that encodes a polyglutamine tract. Targeted deletion of the Huntington disease's(*HD*) gene does not induce the neurologic disorder. On the other hand, transgenic expression of the entire gene or of the first exon containing the sequence encoding the expanded polyglutamine repeat is sufficient to cause many features of the neurologic disorder, indicating a gain-of-function property for the expanded polyglutamine-containing protein. Transgenic strategies can also be used as precursors to gene therapy. Expression of dystrophin, the protein that is deleted in Duchenne's muscular dystrophy, partially corrects the disorder in a mouse model of Duchenne's. Targeted expression of oncogenes has been valuable to study mechanisms of neoplasia and to generate immortalized cell lines. For example, expression of the simian virus 40 (SV40) large T antigen under the direction of the insulin promoter induces the formation of islet cell tumors.

The creation of gene knock-out and knock-in models takes advantage of the fact that a segment of DNA can be substituted by another that is identical (homologous), or nearly identical, by recombination. This permits integration of deletions that disrupt the gene (knock-out) or selected mutations (knock-in) into the target gene of choice. Many of these gene knock-outs do not have an apparent phenotype, either because of redundant functions of the other genes or because the phenotype is subtle. For example, deletion of the hypoxanthine phosphoribosyltransferase (HPRT) gene (*Hprt*) does not cause characteristic features of Lesch-Nyhan syndrome in mice because of their reliance on adenine phosphoribosyltransferase (APRT) in the purine salvage pathway. Deletion of the retinoblastoma (*Rb*) gene encoding p105 does not lead to retinoblastoma or other tumors that characterize the human syndrome. However, mice with combinatorial deletion of several Rb-related proteins exhibit features similar to the human disorder. These examples underscore the fact that the functions of genes, and their interactions with genetic background and the environment, are not necessarily identical in mice and humans. On the other hand, the deletion of many genes provides a remarkably faithful model of human disorders. In addition to clarifying pathophysiology, these models facilitate the development of therapies, both genetic and pharmaceutical.

Many variations of these basic approaches now exist that allow genes to be expressed or deleted in specific cell types, at different times during development, or at varying levels. Consequently, transgenic technology has emerged as a powerful strategy for defining the physiologic effects of deleting or overexpressing a gene, as well as providing unique genetic models for dissecting pathophysiology or testing therapies. In addition to transgenic animal models, naturally occurring mutations in mice and other species continue to provide fundamental insights into human disease. A compendium of natural and transgenic animal models is provided in continuously evolving databases (Table 61-1).

Implications of the Human Genome Project

The HGP was initiated in the mid-1980s as an ambitious effort to characterize the human genome, culminating in a complete DNA sequence. The initial main goals were (1) creation of genetic maps, (2) development of physical maps, and (3) determination of the complete human DNA sequence. Some analogies help in appreciating the scope of the HGP. The 23 pairs of human chromosomes encode ~23,000–30,000 genes. The total length of DNA is ~3 billion bp, which is nearly 1000-fold greater than that of the *E. coli* genome. If the human DNA sequence were printed out, it would correspond to about 120 volumes of *Harrison's Principles of Internal Medicine* .

The identification of the ~10 million SNPs estimated to occur in the human genome has generated a catalogue of common genetic variants that occur in human beings from distinct ethnic backgrounds **(Fig. 61-7)**. SNPs that are in close proximity are inherited together (i.e., they are linked) and are referred to as *haplotypes*, hence the name HapMap **(Fig. 61-8)**. The HapMap describes the nature and location of these SNP haplotypes and how they are distributed among individuals within and among populations. The HapMap information is greatly facilitating GWAS designed to elucidate the complex interactions among multiple genes and lifestyle factors in multifactorial disorders (see below). Moreover, haplotype analyses will be useful to assess variations in responses to medications (*pharmacogenomics*) and environmental factors, as well as the prediction of disease predisposition.

Figure 61-7

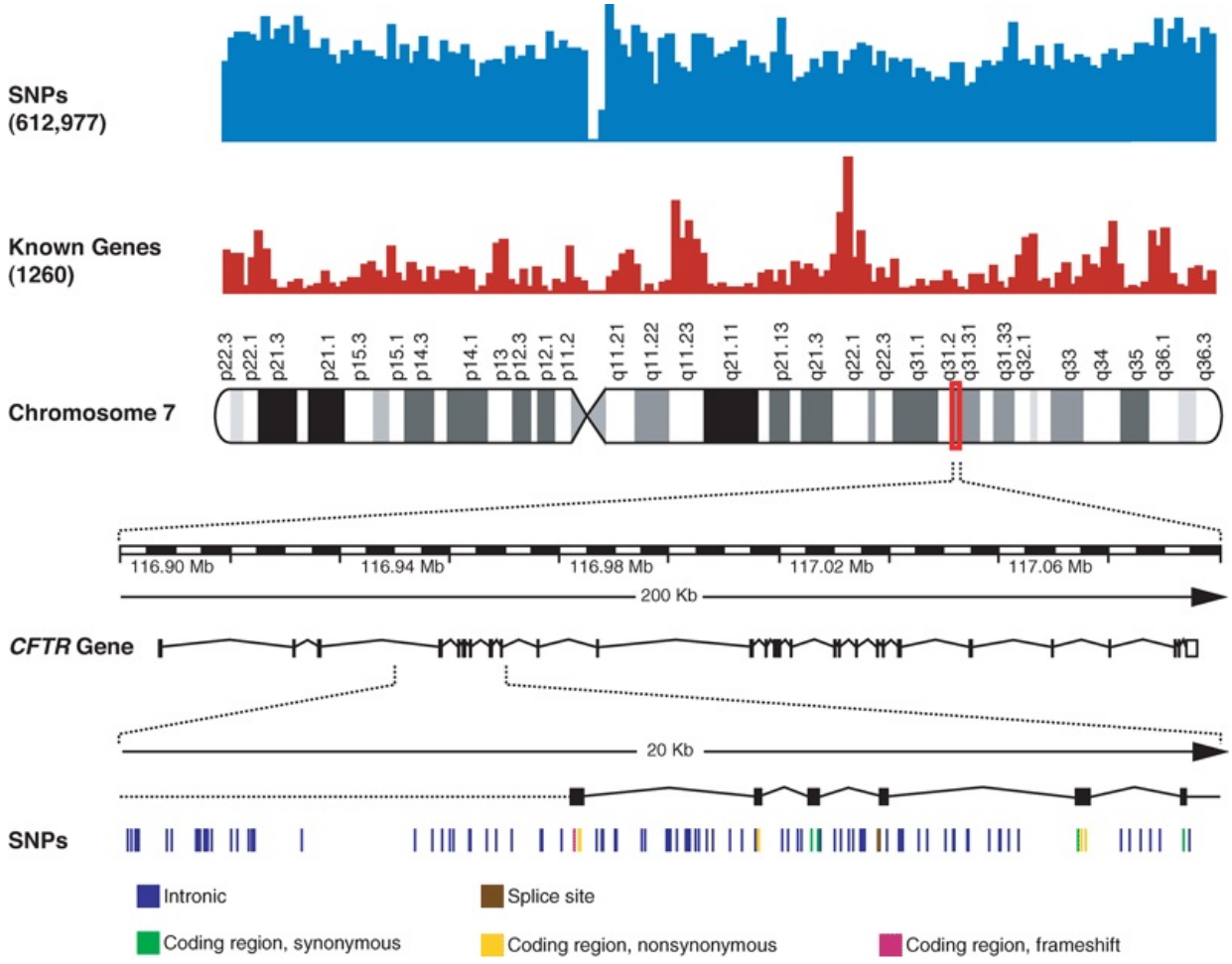

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chromosome 7 is shown with the density of single nucleotide polymorphisms (SNPs) and genes above. A 200-kb region in 7q31.2 containing the *CFTR* gene is shown below. The *CFTR* gene contains 27 exons. More than 1790 mutations in this gene have been found in patients with cystic fibrosis. A 20-kb region encompassing exons 4–9 is shown further amplified in order to illustrate the SNPs in this region.

Figure 61-8

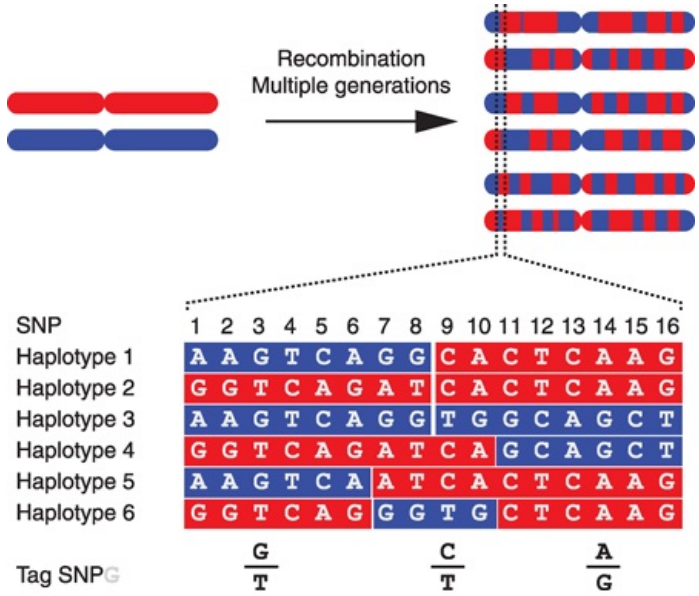

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The origin of haplotypes is due to repeated recombination events occurring in multiple generations. Over time, this leads to distinct haplotypes. These haplotype blocks can often be characterized by genotyping selected Tag single-nucleotide polymorphisms, an approach that now facilitates performing genome-wide association studies (GWAS).

THE HUMAN DNA SEQUENCE

The complete DNA sequence of each chromosome provides the highest-resolution physical map. The primary focus of the HGP was to obtain DNA sequence for the entire human genome as well as model organisms. Although the prospect of determining the complete sequence of the human genome seemed daunting several years ago, technical advances in DNA sequencing and bioinformatics led to the completion of a draft human sequence in June 2000, well in advance of the original goal year of 2003. High-quality reference sequences, completed in 2003, further closed gaps and reduced remaining ambiguities, and the HGP announced the completion of the DNA sequence for the last of the human chromosomes in May 2006. The Personal Genome Project (PGP) launched in 2006 aims to completely sequence the genomes from multiple individuals and to associate them with health and physical information to enhance our understanding of human (patho)physiology. A variety of technical advances in DNA sequencing and computational analyses may soon reduce the cost of obtaining a complete human genome sequence to close to US $1,000. In addition to the human genome, the genomes of numerous organisms have been sequenced completely (~ 1000) or partially (~ 5500) [Genomes Online Database (GOLD); Table 61-1]. They include, among others, eukaryotes such as man and mouse; *S. cerevisiae*, *C. elegans*, and *D. melanogaster*; bacteria (e.g., *E. coli*); and archeae, viruses, organelles (mitochondriae, chloroplasts), and plants (e.g., *Arabidopsis thaliana*). This information, together with technological advances and refinement of computational bioinformatics, has led to a fast-paced transition from the study of single genes to whole genomes. The current directions arising from the HGP include, among others, (1) the comparison of entire genomes (*comparative genomics*), (2) the study of large-scale expression of RNAs (*functional genomics*) and proteins (*proteomics*) in order to detect differences between various tissues in health and disease, (3) the characterization of the variation among individuals by establishing catalogues of sequence variations and SNPs (HapMap project) and the Personal Genome Project, and (4) the identification of genes that play critical roles in the development of polygenic and multifactorial disorders.

ETHICAL ISSUES

Implicit in the HGP is the concept that identifying disease-causing genes can lead to improvements in diagnosis, treatment, and prevention. Most individuals can be expected to harbor several serious recessive gene mutations. Completion of the human genome sequence, determination of the association of genetic defects with disease, and studies of genetic variation raise many new issues with implications for the individual and mankind. The controversies concerning the cloning of mammals, the establishment of human ES cells, and the creation of synthetic organisms underscore the relevance of these questions. Moreover, the information gleaned from genotypic results can have quite different impacts, depending on the availability of strategies to modify the course of disease. For example, the identification of mutations that cause multiple endocrine neoplasia (MEN) type 2 or hemochromatosis allows specific interventions for affected family members. On the other hand, at present, the identification of an Alzheimer's or Huntington's disease gene does not alter therapy and outcomes. However, the progress in this area is unpredictable, as underscored by the finding that

angiotensin II receptor blockers may slow disease progression in Marfan's syndrome.

Genetic test results can generate anxiety in affected individuals and family members, and there is the possibility of discrimination on the basis of the test results. Most genetic disorders are likely to fall into an intermediate category where the opportunity for prevention or treatment is significant but limited (Chap. 63). For these reasons, the scientific components of the HGP have been paralleled by efforts to examine ethical, social, and legal implications as new issues arise.

The Genetic Information Nondiscrimination Act (GINA), signed into law in 2008, aims to protect individuals against the misuse of genetic information for health insurance and employment. The impact of genetic testing on health care costs is currently unclear. It is likely to vary among disorders and depend on the availability of effective therapeutic modalities. A significant problem arises from the marketing of genetic testing directly to consumers by commercial companies. The validity of these tests has not been defined and there are numerous concerns about the lack of appropriate regulatory oversight, the accuracy and confidentiality of genetic information, the availability of counseling, and the handling of these results.

Many issues raised by the genome project are familiar, in principle, to medical practitioners. For example, an asymptomatic patient with increased low-density lipoprotein (LDL) cholesterol, high blood pressure, or a strong family history of early myocardial infarction is known to be at increased risk of coronary heart disease. In such cases, it is clear that the identification of risk factors and an appropriate intervention are beneficial. Likewise, patients with phenylketonuria, cystic fibrosis, or sickle cell anemia are often identified as having a genetic disease early in life. These precedents can be helpful for adapting policies that relate to genetic information. We can anticipate similar efforts, whether based on genotypes or other markers of genetic predisposition, to be applied to many disorders. One confounding aspect of the rapid expansion of information is that our ability to make clinical decisions often lags behind initial insights into genetic mechanisms of disease. For example, when genes that predispose to breast cancer such as *BRCA1* are described, they generate tremendous public interest in the potential to predict disease, but many years of clinical research are still required to rigorously establish genotype and phenotype correlations.

Whether related to informed consent, participation in research, or the management of a genetic disorder that affects an individual or their families, there is a great need for more information about fundamental principles of genetics. The pervasive nature of the role of genetics in medicine makes it imperative for physicians and other health care professionals to become more informed about genetics and to provide advice and counseling in conjunction with trained genetic counselors (Chap. 63). The application of screening and prevention strategies will therefore require intensive patient and physician education, changes in health care financing, and legislation to protect patient's rights.

GENOMICS AND GLOBAL HEALTH

Genomics may contribute to improvements in global health by providing a better understanding of pathogens and diagnostics, and through contributions to drug development. There is, however, concern about the development of a "genomics divide" because of the costs associated with these developments and uncertainty as to whether these advances will be accessible to the populations of developing countries. The World Health Organization has summarized the current issues and inequities surrounding genomic medicine in a detailed report on *Genomics and World Health*.

Transmission of Genetic Disease

ORIGINS AND TYPES OF MUTATIONS

A *mutation* can be defined as any change in the primary nucleotide sequence of DNA regardless of its functional consequences. Some mutations may be lethal, others are less deleterious, and some may confer an evolutionary advantage. Mutations can occur in the germline (sperm or oocytes); these can be transmitted to progeny. Alternatively, mutations can occur during embryogenesis or in somatic tissues. Mutations that occur during development lead to *mosaicism*, a situation in which tissues are composed of cells with different genetic constitutions. If the germline is mosaic, a mutation can be transmitted to some progeny but not others, which sometimes leads to confusion in assessing the pattern of inheritance. Somatic mutations that do not affect cell survival can sometimes be detected because of variable phenotypic effects in tissues (e.g., pigmented lesions in McCune-Albright syndrome). Other somatic mutations are associated with neoplasia because they confer a growth advantage to cells. Epigenetic events, heritable changes that do not involve changes in gene sequence (e.g., altered DNA methylation), may influence gene expression or facilitate genetic damage. With the exception of triplet nucleotide repeats, which can expand (see below), mutations are usually stable.

Mutations are structurally diverse'they can involve the entire genome, as in triploidy (one extra set of chromosomes), or gross numerical or structural alterations in chromosomes or individual genes (Chap. 62). Large deletions may affect a portion of a gene or an entire gene, or, if several genes are involved, they may lead to a *contiguous gene syndrome*. Unequal crossing-over between homologous genes can result in fusion gene mutations, as illustrated by color blindness (Chap. 28). Mutations involving single nucleotides are referred to as *point mutations*. Substitutions are called *transitions* if a purine is

replaced by another purine base (A ↔ G) or if a pyrimidine is replaced by another pyrimidine (C ↔ T). Changes from a purine to a pyrimidine, or vice versa, are referred to as *transversions*. If the DNA sequence change occurs in a coding region and alters an amino acid, it is called a *missense mutation*. Depending on the functional consequences of such a missense mutation, amino acid substitutions in different regions of the protein can lead to distinct phenotypes. *Polymorphisms* are sequence variations that have a frequency of at least 1%. Usually, they do not result in a perceptible phenotype. Often they consist of single base-pair substitutions that do not alter the protein coding sequence because of the degenerate nature of the genetic code (synonymous polymorphism), although it is possible that some might alter mRNA stability, translation, or the amino acid sequence (nonsynonymous polymorphism) (Fig. 61-7). These types of base substitutions are encountered frequently during genetic testing and must be distinguished from true mutations that alter protein expression or function. Small nucleotide deletions or insertions cause a shift of the codon reading frame (*frameshift*). Most commonly, reading frame alterations result in an abnormal protein segment of variable length before termination of translation occurs at a stop codon (*nonsense mutation*) (Fig. 61-6). Mutations in intronic sequences or in exon junctions may destroy or create splice donor or splice acceptor sites. Mutations may also be found in the regulatory sequences of genes, resulting in reduced gene transcription.

Mutation Rates

As noted before, mutations represent an important cause of genetic diversity as well as disease. Mutation rates are difficult to determine in humans because many mutations are silent and because testing is often not adequate to detect the phenotypic consequences. Mutation rates vary in different genes but are estimated to occur at a rate of $\sim 10^{-10}$/bp per cell division. Germline mutation rates (as opposed to somatic mutations) are relevant in the transmission of genetic disease. Because the population of oocytes is established very early in development, only ~20 cell divisions are required for completed oogenesis, whereas spermatogenesis involves ~30 divisions by the time of puberty and 20 cell divisions each year thereafter. Consequently, the probability of acquiring new point mutations is much greater in the male germline than the female germline, in which rates of aneuploidy are increased (Chap. 62). Thus, the incidence of new point mutations in spermatogonia increases with paternal age (e.g., achondrodysplasia, Marfan's syndrome, neurofibromatosis). It is estimated that about 1 in 10 sperm carries a new deleterious mutation. The rates for new mutations are calculated most readily for autosomal dominant and X-linked disorders and are $\sim 10^{-5}$–10^{-6}/locus per generation. Because most monogenic diseases are relatively rare, new mutations account for a significant fraction of cases. This is important in the context of genetic counseling, as a new mutation can be transmitted to the affected individual but does not necessarily imply that the parents are at risk to transmit the disease to other children. An exception to this is when the new mutation occurs early in germline development, leading to *gonadal mosaicism*.

Unequal Crossing-over

Normally, DNA recombination in germ cells occurs with remarkable fidelity to maintain the precise junction sites for the exchanged DNA sequences (Fig. 61-3). However, mispairing of homologous sequences leads to unequal crossover, with gene duplication on one of the chromosomes and gene deletion on the other chromosome. A significant fraction of growth hormone (*GH*) gene deletions, for example, involve unequal crossing-over (Chap. 339). The *GH* gene is a member of a large gene cluster that includes a gh variant gene as well as several structurally related chorionic somatomammotropin genes and pseudogenes (highly homologous but functionally inactive relatives of a normal gene). Because such gene clusters contain multiple homologous DNA sequences arranged in tandem, they are particularly prone to undergo recombination and, consequently, gene duplication or deletion. On the other hand, duplication of the *PMP22* gene because of unequal crossing-over results in increased gene dosage and type IA Charcot-Marie-Tooth disease. Unequal crossing-over resulting in deletion of *PMP22* causes a distinct neuropathy called *hereditary liability to pressure palsy* (Chap. 384).

Glucocorticoid-remediable aldosteronism (GRA) is caused by a rearrangement involving the genes that encode aldosterone synthase (*CYP11B2*) and steroid 11β-hydroxylase (*CYP11B1*), normally arranged in tandem on chromosome 8q. These two genes are 95% identical, predisposing to gene duplication and deletion by unequal crossing-over. The rearranged gene product contains the regulatory regions of 11β-hydroxylase fused to the coding sequence of aldosterone synthetase. Consequently, the latter enzyme is expressed in the adrenocorticotropic hormone (ACTH)–dependent zona fasciculata of the adrenal gland, resulting in overproduction of mineralocorticoids and hypertension (Chap. 342).

Gene conversion refers to a nonreciprocal exchange of homologous genetic information; it is probably more common than generally recognized. In human genetics, gene conversion has been used to explain how an internal portion of a gene is replaced by a homologous segment copied from another allele or locus; these genetic alterations may range from a few nucleotides to a few thousand nucleotides. As a result of gene conversion, it is possible for short DNA segments of two chromosomes to be identical, even though these sequences are distinct in the parents. A practical consequence of

this phenomenon is that nucleotide substitutions can occur during gene conversion between related genes, often altering the function of the gene. In disease states, gene conversion often involves intergenic exchange of DNA between a gene and a related pseudogene. For example, the 21-hydroxylase gene (*CYP21A2*) is adjacent to a nonfunctional pseudogene (*CYP21A1P*). Many of the nucleotide substitutions that are found in the *CYP21A2* gene in patients with congenital adrenal hyperplasia correspond to sequences that are present in the *CYP21A1P* pseudogene, suggesting gene conversion as a mechanism of mutagenesis. In addition, mitotic gene conversion has been suggested as a mechanism to explain revertant mosaicism in which an inherited mutation is "corrected" in certain cells. For example, patients with autosomal recessive generalized atrophic benign epidermolysis bullosa have acquired reverse mutations in one of the two mutated *COL17A1* alleles, leading to clinically unaffected patches of skin.

Insertions and Deletions

Although many instances of insertions and deletions occur as a consequence of unequal crossing-over, there is also evidence for internal duplication, inversion, or deletion of DNA sequences. The fact that certain deletions or insertions appear to occur repeatedly as independent events suggests that specific regions within the DNA sequence predispose to these errors. For example, certain regions of the *DMD* gene appear to be hot spots for deletions. Some regions within the human genome are rearrangement hot spots and lead to copy number variations (CNVs).

Errors in DNA Repair

Because mutations caused by defects in DNA repair accumulate as somatic cells divide, these types of mutations are particularly important in the context of neoplastic disorders (Chap. 84). Several genetic disorders involving DNA repair enzymes underscore their importance. Patients with xeroderma pigmentosum have defects in DNA damage recognition or in the nucleotide excision and repair pathway (Chap. 87). Exposed skin is dry and pigmented and is extraordinarily sensitive to the mutagenic effects of ultraviolet irradiation. More than 10 different genes have been shown to cause the different forms of xeroderma pigmentosum. This finding is consistent with the earlier classification of this disease into different complementation groups in which normal function is rescued by the fusion of cells derived from two different forms of xeroderma pigmentosum.

Ataxia telangiectasia causes large telangiectatic lesions of the face, cerebellar ataxia, immunologic defects, and hypersensitivity to ionizing radiation (Chap. 373). The discovery of the ataxia telangiectasia mutated (*ATM*) gene reveals that it is homologous to genes involved in DNA repair and control of cell cycle checkpoints. Mutations in the *ATM* gene give rise to defects in meiosis as well as increasing susceptibility to damage from ionizing radiation. Fanconi's anemia is also associated with an increased risk of multiple acquired genetic abnormalities. It is characterized by diverse congenital anomalies and a strong predisposition to develop aplastic anemia and acute myelogenous leukemia (Chap. 109). Cells from these patients are susceptible to chromosomal breaks caused by a defect in genetic recombination. At least 13 different complementation groups have been identified, and several loci and genes associated with Fanconi's anemia have been mapped or cloned. HNPCC (Lynch's syndrome) is characterized by autosomal dominant transmission of colon cancer, young age (<50 years) of presentation, predisposition to lesions in the proximal large bowel, and associated malignancies such as uterine cancer and ovarian cancer. HNPCC is predominantly caused by mutations in one of several different mismatch repair (MMR) genes including MutS homologue 2 (*MSH2*), MutL homologue 1 and 6 (*MLH1, MLH6*), *MSH6*, *PMS1*, and *PMS2* (Chap. 91). These proteins are involved in the detection of nucleotide mismatches and in the recognition of slipped-strand trinucleotide repeats. Germline mutations in these genes lead to microsatellite instability and a high mutation rate in colon cancer. Genetic screening tests for this disorder are now being used for families considered to be at risk (Chap. 63). Recognition of HNPCC allows early screening with colonoscopy and the implementation of prevention strategies using nonsteroidal anti-inflammatory drugs.

Dipyrimidine and CpG Sequences

Certain DNA sequences are particularly susceptible to mutagenesis. Successive pyrimidine residues (e.g., T-T or C-C) are subject to the formation of ultraviolet light–induced photoadducts. If these pyrimidine dimers are not repaired by the nucleotide excision repair pathway, mutations will be introduced after DNA synthesis. The dinucleotide C-G, or CpG, is also a hot spot for a specific type of mutation. In this case, methylation of the cytosine is associated with an enhanced rate of deamination to uracil, which is then replaced with thymine. This C → T transition (or G → A on the opposite strand) accounts for at least one-third of point mutations associated with polymorphisms and mutations. Many of the *MSH2* mutations in HNPCC, for example, involve CpG sequences. In addition to the fact that certain types of mutations (C → T or G → A) are relatively common, the nature of the genetic code also results in overrepresentation of certain amino acid substitutions.

Unstable DNA Sequences

Trinucleotide repeats may be unstable and expand beyond a critical number. Mechanistically, the expansion is thought to be caused by unequal recombination and slipped mispairing. A premutation represents a small increase in trinucleotide copy number. In subsequent generations, the expanded

repeat may increase further in length and result in an increasingly severe phenotype, a process called *dynamic mutation* (see below for discussion of anticipation). Trinucleotide expansion was first recognized as a cause of the fragile X syndrome, one of the most common causes of mental retardation. Other disorders arising from a similar mechanism include Huntington's disease (Chap. 371), X-linked spinobulbar muscular atrophy (Chap. 374), and myotonic dystrophy (Chap. 387). Malignant cells are also characterized by genetic instability, indicating a breakdown in mechanisms that regulate DNA repair and the cell cycle.

FUNCTIONAL CONSEQUENCES OF MUTATIONS

Functionally, mutations can be broadly classified as gain-of-function and loss-of-function mutations. Gain-of-function mutations are typically dominant (i.e., they result in phenotypic alterations when a single allele is affected). Inactivating mutations are usually recessive, and an affected individual is homozygous or compound heterozygous (e.g., carrying two different mutant alleles of the same gene) for the disease-causing mutations. Alternatively, mutation in a single allele can result in *haploinsufficiency*, a situation in which one normal allele is not sufficient to maintain a normal phenotype. Haploinsufficiency is a commonly observed mechanism in diseases associated with mutations in transcription factors (Table 61-2). Remarkably, the clinical features among patients with an identical mutation in a transcription factor often vary significantly. One mechanism underlying this variability consists in the influence of modifying genes. Haploinsufficiency can also affect the expression of rate-limiting enzymes. For example, haploinsufficiency in enzymes involved in heme synthesis can cause porphyrias (Chap. 358).

An increase in dosage of a gene product may also result in disease, as illustrated by the duplication of the *DAX1* gene in dosage-sensitive sex-reversal (Chap. 349). Mutation in a single allele can also result in loss of function due to a dominant-negative effect. In this case, the mutated allele interferes with the function of the normal gene product by one of several different mechanisms: (1) a mutant protein may interfere with the function of a multimeric protein complex, as illustrated by mutations in type 1 collagen (*COL1A1*, *COL1A2*) genes in osteogenesis imperfecta (Chap. 363); (2) a mutant protein may occupy binding sites on proteins or promoter response elements, as illustrated by thyroid hormone resistance, a disorder in which inactivated thyroid hormone receptor binds to target genes and functions as an antagonist of normal receptors (Chap. 341); or (3) a mutant protein can be cytotoxic as in α_1 antitrypsin deficiency (Chap. 260) or autosomal dominant neurohypophyseal diabetes insipidus (Chap. 340), in which the abnormally folded proteins are trapped within the endoplasmic reticulum and ultimately cause cellular damage.

GENOTYPE AND PHENOTYPE

Alleles, Genotypes, and Haplotypes

An observed trait is referred to as a *phenotype*; the genetic information defining the phenotype is called the *genotype*. Alternative forms of a gene or a genetic marker are referred to as *alleles*. Alleles may be polymorphic variants of nucleic acids that have no apparent effect on gene expression or function. In other instances, these variants may have subtle effects on gene expression, thereby conferring the adaptive advantages associated with genetic diversity. On the other hand, allelic variants may reflect mutations in a gene that clearly alter its function. The common Glu6Val (E6V) sickle cell mutation in the β-*globin* gene and the ΔF508 deletion of phenylalanine (F) in the *CFTR* gene are examples of allelic variants of these genes that result in disease. Because each individual has two copies of each chromosome (one inherited from the mother and one inherited from the father), he or she can have only two alleles at a given locus. However, there can be many different alleles in the population. The normal or common allele is usually referred to as *wild type*. When alleles at a given locus are identical, the individual is *homozygous*. Inheriting identical copies of a mutant allele occurs in many autosomal recessive disorders, particularly in circumstances of consanguinity. If the alleles are different on the maternal and the paternal copy of the gene, the individual is *heterozygous* at this locus (Fig. 61-6). If two different mutant alleles are inherited at a given locus, the individual is said to be a *compound heterozygote*. *Hemizygous* is used to describe males with a mutation in an X chromosomal gene or a female with a loss of one X chromosomal locus.

Genotypes describe the specific alleles at a particular locus. For example, there are three common alleles (E2, E3, E4) of the apolipoprotein E (*APOE*) gene. The genotype of an individual can therefore be described as *APOE3/4* or *APOE4/4* or any other variant. These designations indicate which alleles are present on the two chromosomes in the *APOE* gene at locus 19q13.2. In other cases, the genotype might be assigned arbitrary numbers (e.g., 1/2) or letters (e.g., B/b) to distinguish different alleles.

A *haplotype* refers to a group of alleles that are closely linked together at a genomic locus (Fig. 61-8). Haplotypes are useful for tracking the transmission of genomic segments within families and for detecting evidence of genetic recombination, if the crossover event occurs between the alleles (Fig. 61-3). As an example, various alleles at the histocompatibility locus antigen (HLA) on chromosome 6p are used to establish haplotypes associated with certain disease states. For example, 21-hydroxylase deficiency, complement deficiency, and hemochromatosis are each associated with specific HLA

haplotypes. It is now recognized that these genes lie in close vicinity to the HLA locus, which explains why HLA associations were identified even before the disease genes were cloned and localized. In other cases, specific HLA associations with diseases such as ankylosing spondylitis (HLA-B27) or type 1 diabetes mellitus (HLA-DR4) reflect the role of specific HLA allelic variants in susceptibility to these autoimmune diseases. The characterization of common SNP haplotypes in numerous populations from different parts of the world through the HapMap project is providing a novel tool for association studies designed to detect genes involved in the pathogenesis of complex disorders (Table 61-1). The presence or absence of certain haplotypes may also become relevant for the customized choice of medical therapies (pharmacogenomics) or for preventive strategies.

Allelic Heterogeneity

Allelic heterogeneity refers to the fact that different mutations in the same genetic locus can cause an identical or similar phenotype. For example, many different mutations of the β-globin locus can cause β-thalassemia **(Table 61-4)** (Fig. 61-5). In essence, allelic heterogeneity reflects the fact that many different mutations are capable of altering protein structure and function. For this reason, maps of inactivating mutations in genes usually show a near-random distribution. Exceptions include (1) a founder effect, in which a particular mutation that does not affect reproductive capacity can be traced to a single individual; (2) "hot spots" for mutations, in which the nature of the DNA sequence predisposes to a recurring mutation; and (3) localization of mutations to certain domains that are particularly critical for protein function. Allelic heterogeneity creates a practical problem for genetic testing because one must often examine the entire genetic locus for mutations, as these can differ in each patient. For example, there are currently 1795 reported mutations in the *CFTR* gene (Fig. 61-7). The mutational analysis initially focuses on a panel of mutations that are particularly frequent (often taking the ethnic background of the patient into account), but a negative result does not exclude the presence of a mutation elsewhere in the gene. One should also be aware that mutational analyses generally focus on the coding region of a gene without considering regulatory and intronic regions. Because disease-causing mutations may be located outside the coding regions, negative results should be interpreted with caution. It is expected that the advent of more comprehensive sequencing technologies will greatly facilitate mutational analyses. It will, however, also introduce challenges because the detection of a sequence alteration alone is not always sufficient to establish that it has a causal role.

Table 61-4 Selected Examples of Locus Heterogeneity and Phe-Notypic Heterogeneity

Phenotypic Heterogeneity

Gene, Protein	Phenotype	Inheritance	OMIM
LMNA, Lamin A/C	Emery-Dreifuss muscular dystrophy (AD)	AD	181350
	Familial partial lipodystrophy Dunnigan	AD	151660
	Hutchinson-Gilford progeria	AD	176670
	Atypical Werner's syndrome	AD	150330
	Dilated cardiomyopathy	AD	115200
	Early-onset atrial fibrillation	AD	607554
	Emery-Dreifuss muscular dystrophy (AR)	AR	604929
	Limb-girdle muscular dystrophy type 1B	AR	159001
	Charcot-Marie-Tooth type 2B1	AR	605588
KRAS	Noonan syndrome	AD	163950
	Cardio-facio-cutaneous syndrome	AD	115150

Locus Heterogeneity

Phenotype	Gene	Chromosomal Location	Protein
Familial hypertrophic cardiomyopathy	*MYH7*	14q12	Myosin heavy chain beta
	TNNT2	1q2	Troponin-T2
Genes encoding sarcomeric proteins	*TPM1*	15q22.1	Tropomyosin alpha
	MYBPC3	11p11q	Myosin-binding protein C

	TNNI3	19q13.4	Troponin 1
	MYL2	12q23-24.3	Myosin light chain 2
	MYL3	3p	Myosin light chain 3
	TTN	2q24.3	Cardiac titin
	ACTC	15q11	Cardiac alpha actin
	MYH6	14q1	Myosin heavy chain alpha
	MYLK2	20q13.3	Myosin light-peptide kinase
	CAV3	3p25	Caveolin 3
Genes encoding nonsarcomeric proteins	*MTT1*	Mitochondrial	tRNA isoleucine
	MTTG	Mitochondrial	tRNA glycine
	PRKAG2	7q35-q36	AMP-activated protein kinase γ2 subunit
	DMPK	19q13.2-13.3	Myotonin protein kinase (myotonic dystrophy)
	FRDA	9q13	Frataxin (Friedreich ataxia)
Polycystic kidney disease	*PKD1*	16p13.3-13.12	Polycystin 1 (AD)
	PKD2	4q21.-23	Polycystin 2 (AD)
	PKHD1	6p21.1-p12	Fibrocystin (AR)
Noonan syndrome	PTPN11	12q24.1	Protein-tyrosine phosphatase 2c
	KRAS	12p12.1	KRAS

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

Phenotypic Heterogeneity

Phenotypic heterogeneity occurs when more than one phenotype is caused by allelic mutations (i.e., different mutations in the same gene) (Table 61-4). For example, laminopathies are monogenic multisystem disorders that result from mutations in the *LMNA* gene, which encodes the nuclear lamins A and C. Twelve autosomal dominant and four autosomal recessive disorders are caused by mutations in the *LMNA* gene. They include several forms of lipodystrophies, Emery-Dreifuss muscular dystrophy, progeria syndromes, a form of neuronal Charcot-Marie-Tooth disease (type 2B1), and a group of overlapping syndromes. Remarkably, hierarchical cluster analysis has revealed that the phenotypes vary depending on the position of the mutation. Similarly, identical mutations in the *FGFR2* gene can result in very distinct phenotypes: Crouzon's syndrome (craniofacial synostosis) or Pfeiffer's syndrome (acrocephalopolysyndactyly).

Locus or Nonallelic Heterogeneity and Phenocopies

Nonallelic or locus heterogeneity refers to the situation in which a similar disease phenotype results from mutations at different genetic loci. This often occurs when more than one gene product produces different subunits of an interacting complex or when different genes are involved in the same genetic cascade or physiologic pathway. For example, osteogenesis imperfecta can arise from mutations in two different procollagen genes (*COL1A1* or *COL1A2*) that are located on different chromosomes (Chap. 363). The effects of inactivating mutations in these two genes are similar because the protein products comprise different subunits of the helical collagen fiber. Similarly, muscular dystrophy syndromes can be caused by mutations in various genes, consistent with the fact that it can be transmitted in an X-linked (Duchenne or Becker), autosomal dominant (limb-girdle muscular dystrophy type 1), or autosomal recessive (limb-girdle muscular dystrophy type 2) manner (Chap. 387). Mutations in the X-linked *DMD* gene, which encodes dystrophin, are the most common cause of muscular dystrophy. This feature reflects the large size of the gene as well as the fact that the phenotype is expressed in hemizygous males because they have only a single copy of the X chromosome. Dystrophin is associated with a large protein complex linked to the membrane-associated cytoskeleton in muscle. Mutations in several different components of this protein complex can also cause muscular dystrophy syndromes. Although the phenotypic features of some of these disorders are distinct, the phenotypic spectrum caused by mutations in different genes overlaps, thereby leading to nonallelic heterogeneity. It should be noted that mutations in dystrophin also cause allelic heterogeneity. For example, mutations in the *DMD* gene can cause either Duchenne's or the less severe Becker's muscular dystrophy, depending on the severity of the protein defect.

Recognition of nonallelic heterogeneity is important for several reasons: (1) the ability to identify disease loci in linkage studies is reduced by including patients with similar phenotypes but different genetic disorders; (2) genetic testing is more complex because several different genes need to be

considered along with the possibility of different mutations in each of the candidate genes; and (3) novel information is gained about how genes or proteins interact, providing unique insights into molecular physiology.

Phenocopies refer to circumstances in which nongenetic conditions mimic a genetic disorder. For example, features of toxin- or drug-induced neurologic syndromes can resemble those seen in Huntington's disease, and vascular causes of dementia share phenotypic features with familial forms of Alzheimer's dementia (Chap. 371). Children born with activating mutations of the thyroid-stimulating hormone receptor (TSH-R) exhibit goiter and thyrotoxicosis similar to that seen in neonatal Graves' disease, which is caused by the transfer of maternal autoantibodies to the fetus (Chap. 341). As in nonallelic heterogeneity, the presence of phenocopies has the potential to confound linkage studies and genetic testing. Patient history and subtle differences in phenotype can often provide clues that distinguish these disorders from related genetic conditions.

Variable Expressivity and Incomplete Penetrance

The same genetic mutation may be associated with a phenotypic spectrum in different affected individuals, thereby illustrating the phenomenon of *variable expressivity*. This may include different manifestations of a disorder variably involving different organs (e.g., MEN), the severity of the disorder (e.g., cystic fibrosis), or the age of disease onset (e.g., Alzheimer's dementia). MEN-1 illustrates several of these features. Families with this autosomal dominant disorder develop tumors of the parathyroid gland, endocrine pancreas, and the pituitary gland (Chap. 351). However, the pattern of tumors in the different glands, the age at which tumors develop, and the types of hormones produced vary among affected individuals, even within a given family. In this example, the phenotypic variability arises, in part, because of the requirement for a second mutation in the normal copy of the *MEN1* gene, as well as the large array of different cell types that are susceptible to the effects of *MEN1* gene mutations. In part, variable expression reflects the influence of modifier genes, or genetic background, on the effects of a particular mutation. Even in identical twins, in whom the genetic constitution is essentially the same, one can occasionally see variable expression of a genetic disease.

Interactions with the environment can also influence the course of a disease. For example, the manifestations and severity of hemochromatosis can be influenced by iron intake (Chap. 357), and the course of phenylketonuria is affected by exposure to phenylalanine in the diet (Chap. 364). Other metabolic disorders such as hyperlipidemias and porphyria, also fall into this category. Many mechanisms, including genetic effects and environmental influences, can therefore lead to variable expressivity. In genetic counseling, it is particularly important to recognize this variability, as one cannot always predict the course of disease, even when the mutation is known.

Penetrance refers to the proportion of individuals with a mutant genotype that express the phenotype. If all carriers of a mutant express the phenotype, penetrance is complete, whereas it is said to be *incomplete* or *reduced* if some individuals do not have any features of the phenotype. Dominant conditions with incomplete penetrance are characterized by skipping of generations with unaffected carriers transmitting the mutant gene. For example, hypertrophic obstructive cardiomyopathy (HCM) caused by mutations in the *myosin-binding protein C* gene is a dominant disorder with clinical features in only a subset of patients who carry the mutation (Chap. 237). Patients who have the mutation but no evidence of the disease can still transmit the disorder to subsequent generations. In many conditions with postnatal onset, the proportion of gene carriers who are affected varies with age. Thus, when describing penetrance, one has to specify age. For example, for disorders such as Huntington's disease or familial amyotrophic lateral sclerosis, which present late in life, the rate of penetrance is influenced by the age at which the clinical assessment is performed. *Imprinting* can also modify the penetrance of a disease (see below). For example, in patients with Albright's hereditary osteodystrophy, mutations in the Gsα subunit (*GNAS1* gene) are expressed clinically only in individuals who inherit the mutation from their mother (Chap. 353).

Sex-Influenced Phenotypes

Certain mutations affect males and females quite differently. In some instances, this is because the gene resides on the X or Y sex chromosomes (X-linked disorders and Y-linked disorders). As a result, the phenotype of mutated X-linked genes will be expressed fully in males but variably in heterozygous females, depending on the degree of X-inactivation and the function of the gene. For example, most heterozygous female carriers of factor VIII deficiency (hemophilia A) are asymptomatic because sufficient factor VIII is produced to prevent a defect in coagulation (Chap. 116). On the other hand, some females heterozygous for the X-linked lipid storage defect caused by α-galactosidase A deficiency (Fabry's disease) experience mild manifestations of painful neuropathy, as well as other features of the disease (Chap. 361). Because only males have a Y chromosome, mutations in genes such as *SRY*, which causes male-to-female sex-reversal, or *DAZ* (deleted in azoospermia), which causes abnormalities of spermatogenesis, are unique to males (Chap. 349).

Other diseases are expressed in a sex-limited manner because of the differential function of the gene product in males and females. Activating mutations in the luteinizing hormone receptor cause dominant

male-limited precocious puberty in boys (Chap. 346). The phenotype is unique to males because activation of the receptor induces testosterone production in the testis, whereas it is functionally silent in the immature ovary. Biallelic inactivating mutations of the follicle-stimulating hormone (FSH) receptor cause primary ovarian failure in females because the follicles do not develop in the absence of FSH action. In contrast, affected males have a more subtle phenotype, because testosterone production is preserved (allowing sexual maturation) and spermatogenesis is only partially impaired (Chap. 346). In congenital adrenal hyperplasia, most commonly caused by 21-hydroxylase deficiency, cortisol production is impaired and ACTH stimulation of the adrenal gland leads to increased production of androgenic precursors (Chap. 342). In females, the increased androgen level causes ambiguous genitalia, which can be recognized at the time of birth. In males, the diagnosis may be made on the basis of adrenal insufficiency at birth, because the increased adrenal androgen level does not alter sexual differentiation, or later in childhood, because of the development of precocious puberty. Hemochromatosis is more common in males than in females, presumably because of differences in dietary iron intake and losses associated with menstruation and pregnancy in females (Chap. 357).

CHROMOSOMAL DISORDERS

Chromosomal or cytogenetic disorders are caused by numerical or structural aberrations in chromosomes. Deviations in chromosome number are common causes of abortions, developmental disorders, and malformations. *Contiguous gene syndromes* (i.e., large deletions affecting several genes), have been useful for identifying the location of new disease-causing genes. Because of the variable size of gene deletions in different patients, a systematic comparison of phenotypes and locations of deletion breakpoints allows positions of particular genes to be mapped within the critical genomic region.

For discussion of disorders of chromosome number and structure, see Chap. 62.

MONOGENIC MENDELIAN DISORDERS

Monogenic human diseases are frequently referred to as *Mendelian disorders* because they obey the principles of genetic transmission originally set forth in Gregor Mendel's classic work. The continuously updated OMIM catalogue lists several thousand of these disorders and provides information about the clinical phenotype, molecular basis, allelic variants, and pertinent animal models (Table 61-1). The mode of inheritance for a given phenotypic trait or disease is determined by pedigree analysis. All affected and unaffected individuals in the family are recorded in a pedigree using standard symbols **(Fig. 61-9)**. The principles of allelic segregation, and the transmission of alleles from parents to children, are illustrated in **Fig. 61-10**. One dominant (A) allele and one recessive (a) allele can display three Mendelian modes of inheritance: autosomal dominant, autosomal recessive, and X-chromosomal. About 65% of human monogenic disorders are autosomal dominant, 25% are autosomal recessive, and 5% are X-linked. Genetic testing is now available for many of these disorders and plays an increasingly important role in clinical medicine (Chap. 63).

Figure 61-9

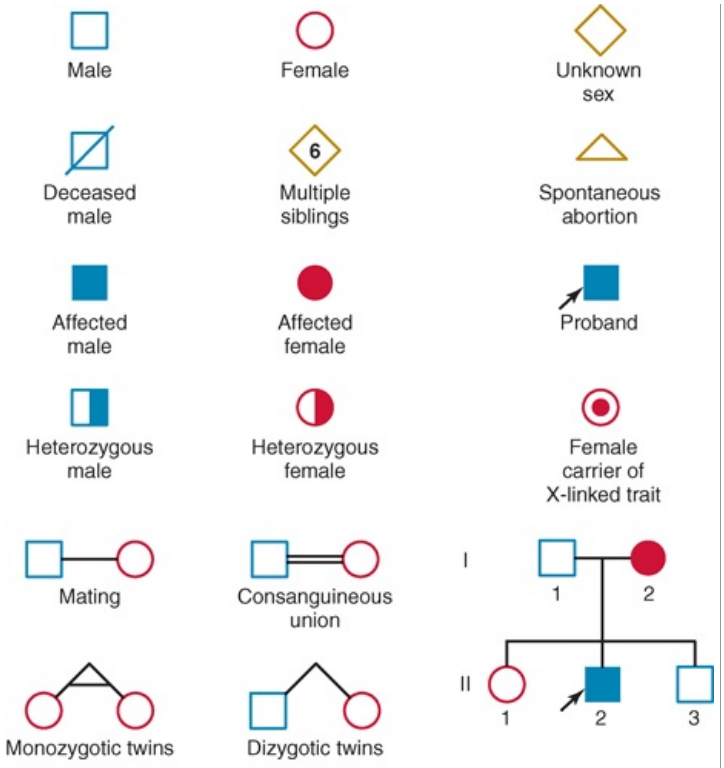

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Standard pedigree symbols.

Figure 61-10

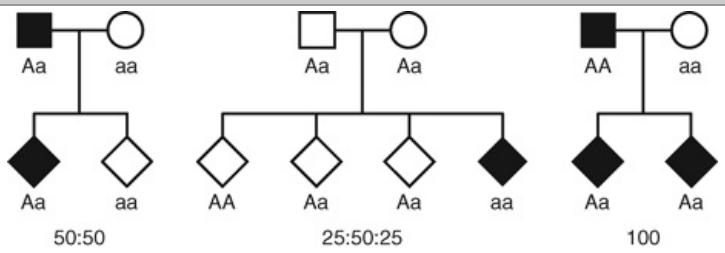

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Segregation of alleles. Segregation of genotypes in the offspring of parents with one dominant (A) and one recessive (a) allele. The distribution of the parental alleles to their offspring depends on the combination present in the parents. Filled symbols represent affected individuals.

Autosomal Dominant Disorders

Autosomal dominant disorders assume particular relevance because mutations in a single allele are sufficient to cause the disease. In contrast to recessive disorders, in which disease pathogenesis is relatively straightforward because there is loss of gene function, dominant disorders can be caused by various disease mechanisms, many of which are unique to the function of the genetic pathway involved.

In autosomal dominant disorders, individuals are affected in successive generations; the disease does not occur in the offspring of unaffected individuals. Males and females are affected with equal frequency because the defective gene resides on one of the 22 autosomes **(Fig. 61-11*A*)**. Autosomal dominant mutations alter one of the two alleles at a given locus. Because the alleles segregate randomly at meiosis, the probability that an offspring will be affected is 50%. Unless there is a new germline mutation, an affected individual has an affected parent. Children with a normal genotype do not transmit the disorder. Due to differences in penetrance or expressivity (see above), the clinical manifestations of autosomal dominant disorders may be variable. Because of these variations, it is sometimes challenging to determine the pattern of inheritance.

Figure 61-11

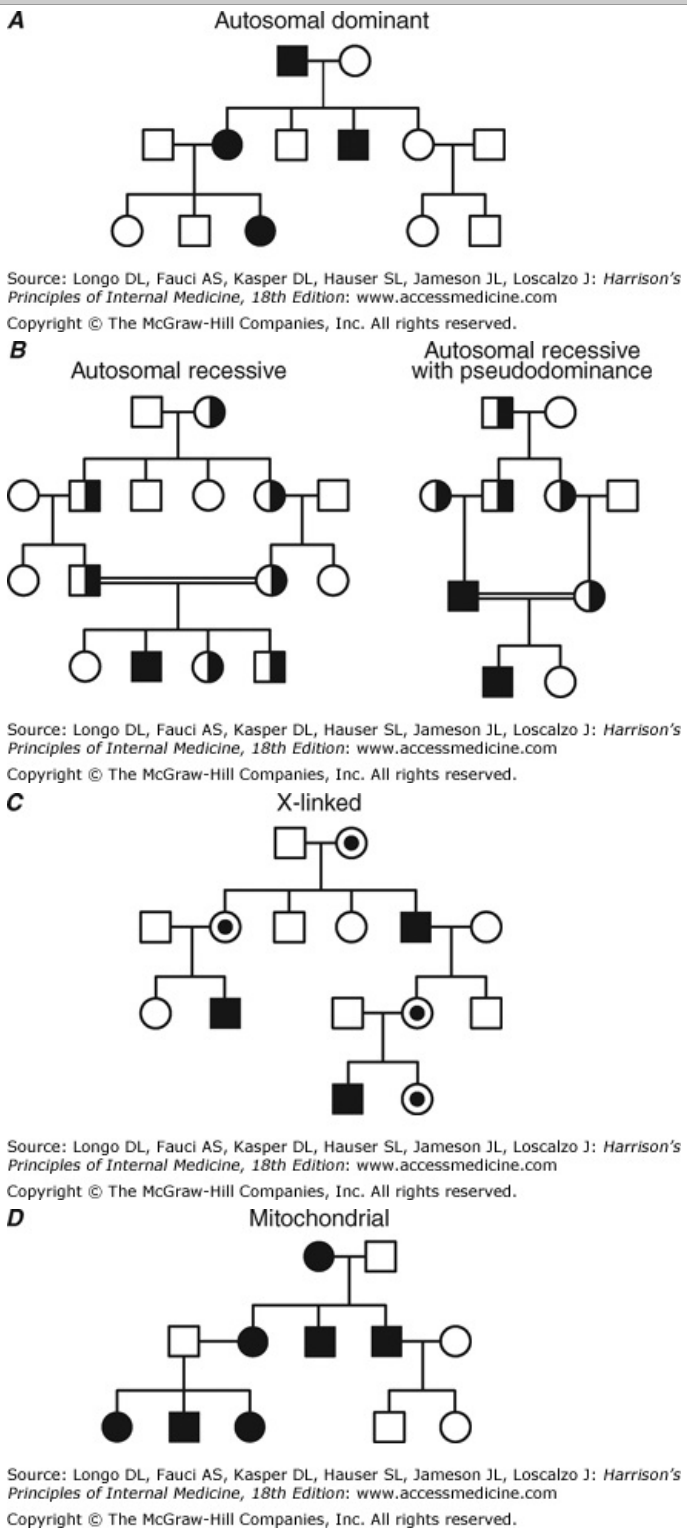

Dominant, recessive, X-linked, and mitochondrial (matrilinear) inheritance.

It should be recognized, however, that some individuals acquire a mutated gene from an unaffected parent. De novo germline mutations occur more frequently during later cell divisions in gametogenesis, which explains why siblings are rarely affected. As noted before, new germline mutations occur more frequently in fathers of advanced age. For example, the average age of fathers with new germline mutations that cause Marfan's syndrome is ~37 years, whereas fathers who transmit the disease by inheritance have an average age of ~30 years.

Autosomal Recessive Disorders

In recessive disorders, the mutated alleles result in a complete or partial loss of function. They frequently involve enzymes in metabolic pathways, receptors, or proteins in signaling cascades. In an autosomal recessive disease, the affected individual, who can be of either sex, is a homozygote or compound heterozygote for a single-gene defect. With a few important exceptions, autosomal

recessive diseases are rare and often occur in the context of parental consanguinity. The relatively high frequency of certain recessive disorders such as sickle cell anemia, cystic fibrosis, and thalassemia, is partially explained by a selective biologic advantage for the heterozygous state (see below). Though heterozygous carriers of a defective allele are usually clinically normal, they may display subtle differences in phenotype that only become apparent with more precise testing or in the context of certain environmental influences. In sickle cell anemia, for example, heterozygotes are normally asymptomatic. However, in situations of dehydration or diminished oxygen pressure, sickle cell crises can also occur in heterozygotes (Chap. 104).

In most instances, an affected individual is the offspring of heterozygous parents. In this situation, there is a 25% chance that the offspring will have a normal genotype, a 50% probability of a heterozygous state, and a 25% risk of homozygosity for the recessive alleles (Figs. 61-10, 61-11*B*). In the case of one unaffected heterozygous and one affected homozygous parent, the probability of disease increases to 50% for each child. In this instance, the pedigree analysis mimics an autosomal dominant mode of inheritance (*pseudodominance*). In contrast to autosomal dominant disorders, new mutations in recessive alleles are rarely manifest because they usually result in an asymptomatic carrier state.

X-Linked Disorders

Males have only one X chromosome; consequently, a daughter always inherits her father's X chromosome in addition to one of her mother's two X chromosomes. A son inherits the Y chromosome from his father and one maternal X chromosome. Thus, the characteristic features of X-linked inheritance are (1) the absence of father-to-son transmission, and (2) the fact that all daughters of an affected male are obligate carriers of the mutant allele (Fig. 61-11*C*). The risk of developing disease due to a mutant X-chromosomal gene differs in the two sexes. Because males have only one X chromosome, they are hemizygous for the mutant allele; thus, they are more likely to develop the mutant phenotype, regardless of whether the mutation is dominant or recessive. A female may be either heterozygous or homozygous for the mutant allele, which may be dominant or recessive. The terms *X-linked dominant* or *X-linked recessive* are therefore only applicable to expression of the mutant phenotype in women. In addition, the expression of X-chromosomal genes is influenced by X chromosome inactivation (see below).

Y-Linked Disorders

The Y chromosome has a relatively small number of genes. One such gene, the sex-region determining Y factor (*SRY*), which encodes the testis-determining factor (*TDF*), is crucial for normal male development. Normally there is infrequent exchange of sequences on the Y chromosome with the X chromosome. The *SRY* region is adjacent to the pseudoautosomal region, a chromosomal segment on the X and Y chromosomes with a high degree of homology. A crossing-over occasionally involves the *SRY* region with the distal tip of the X chromosome during meiosis in the male. Translocations can result in XY females with the Y chromosome lacking the *SRY* gene or XX males harboring the *SRY* gene on one of the X chromosomes (Chap. 349). Point mutations in the *SRY* gene may also result in individuals with an XY genotype and an incomplete female phenotype. Most of these mutations occur de novo. Men with oligospermia/azoospermia frequently have microdeletions on the long arm of the Y chromosome that involve one or more of the azoospermia factor (*AZF*) genes.

EXCEPTIONS TO SIMPLE MENDELIAN INHERITANCE PATTERNS

Mitochondrial Disorders

Mendelian inheritance refers to the transmission of genes encoded by DNA contained in the nuclear chromosomes. In addition, each mitochondrion contains several copies of a small circular chromosome. The mitochondrial DNA (mtDNA) is ~16.5 kb and encodes transfer and ribosomal RNAs and 13 proteins that are components of the respiratory chain involved in oxidative phosphorylation and ATP generation. The mitochondrial genome does not recombine and is inherited through the maternal line because sperm does not contribute significant cytoplasmic components to the zygote. A noncoding region of the mitochondrial chromosome, referred to as D-loop, is highly polymorphic. This property, together with the absence of mtDNA recombination, makes it a valuable tool for studies tracing human migration and evolution, and it is also used for specific forensic applications.

Inherited mitochondrial disorders are transmitted in a matrilineal fashion; all children from an affected mother will inherit the disease, but it will not be transmitted from an affected father to his children (Fig. 61-11*D*). Alterations in the mtDNA affecting enzymes required for oxidative phosphorylation lead to reduction of ATP supply, generation of free radicals, and induction of apoptosis. Several syndromic disorders arising from mutations in the mitochondrial genome are known in humans and they affect both protein-coding and tRNA genes (Tables 61-1 and **61-5**). The broad clinical spectrum often involves (cardio)myopathies and encephalopathies because of the high dependence of these tissues on oxidative phosphorylation. The age of onset and the clinical course are highly variable because of the unusual mechanisms of mtDNA transmission, which replicates independently from nuclear DNA. During cell replication, the proportion of wild-type and mutant mitochondria can drift among different cells and tissues. The resulting heterogeneity in the proportion of mitochondria with and without a

mutation is referred to as *heteroplasmia* and underlies the phenotypic variability that is characteristic of mitochondrial diseases.

Table 61-5 Selected Mitochondrial Diseases

Disease/Syndrome	OMIM #
MELAS syndrome: mitochondrial myopathy with encephalopathy, lactacidosis, and stroke	540000
Leber's optic atrophy: hereditary optical neuropathy	535000
Kearns-Sayre syndrome (KSS): ophthalmoplegia, pigmental degeneration of the retina, cardiomyopathy	530000
MERRF syndrome: myoclonic epilepsy and ragged-red fibers	545000
Neurogenic muscular weakness with ataxia and retinitis pigmentosa (NARP)	551500
Chronic progressive external ophthalmoplegia (CEOP)	258470
Pearson's syndrome (PEAR): bone marrow and pancreatic failure	557000
Autosomal dominant inherited mitochondrial myopathy with mitochondrial deletion (ADMIMY)	157640
Somatic mutations in cytochrome *b* gene: exercise intolerance, lactic acidosis, complex III deficiency, muscle pain, ragged-red fibers	516020

Acquired somatic mutations in mitochondria are thought to be involved in several age-dependent degenerative disorders affecting predominantly muscle and the peripheral and central nervous system (e.g., Alzheimer's and Parkinson's diseases). Establishing that an mtDNA alteration is causal for a clinical phenotype is challenging because of the high degree of polymorphism in mtDNA and the phenotypic variability characteristic of these disorders. Certain pharmacologic treatments may have an impact on mitochondria and/or their function. For example, treatment with the antiretroviral compound azidothymidine (AZT) causes an acquired mitochondrial myopathy through depletion of muscular mtDNA.

Mosaicism

Mosaicism refers to the presence of two or more genetically distinct cell lines in the tissues of an individual. It results from a mutation that occurs during embryonic, fetal, or extrauterine development. The developmental stage at which the mutation arises will determine whether germ cells and/or somatic cells are involved. Chromosomal mosaicism results from nondisjunction at an early embryonic mitotic division, leading to the persistence of more than one cell line, as exemplified by some patients with Turner's syndrome (Chap. 349). Somatic mosaicism is characterized by a patchy distribution of genetically altered somatic cells. The McCune-Albright syndrome, for example, is caused by activating mutations in the stimulatory G protein α ($G_{s}\alpha$) that occur early in development (Chap. 353). The clinical phenotype varies depending on the tissue distribution of the mutation; manifestations include ovarian cysts that secrete sex steroids and cause precocious puberty, polyostotic fibrous dysplasia, café-au-lait skin pigmentation, growth hormone–secreting pituitary adenomas, and hypersecreting autonomous thyroid nodules (Chap. 347).

X-Inactivation, Imprinting, and Uniparental Disomy

According to traditional Mendelian principles, the parental origin of a mutant gene is irrelevant for the expression of the phenotype. There are, however, important exceptions to this rule. *X-inactivation* prevents the expression of most genes on one of the two X-chromosomes in every cell of a female. Gene inactivation also occurs on selected chromosomal regions of autosomes. This phenomenon, referred to as *genomic imprinting*, leads to inheritable preferential expression of one of the parental alleles. It is of pathophysiologic importance in disorders where the transmission of disease is dependent on the sex of the transmitting parent and, thus, plays an important role in the expression of certain genetic disorders. Two classic examples are the Prader-Willi syndrome and Angelman's syndrome (Chap. 62). Prader-Willi syndrome is characterized by diminished fetal activity, obesity, hypotonia, mental retardation, short stature, and hypogonadotropic hypogonadism. Deletions of the paternal copy of the Prader-Willi locus located on the short arm of chromosome 15 result in a contiguous gene syndrome involving missing paternal copies of the *necdin* and *SNRPN* genes, among others. In contrast, patients with Angelman's syndrome, characterized by mental retardation, seizures, ataxia, and hypotonia, have deletions involving the maternal copy of this region on chromosome 15. These two syndromes may also result from *uniparental disomy*. In this case, the syndromes are not caused by deletions on chromosome 15 but by the inheritance of either two maternal chromosomes (Prader-Willi syndrome) or two paternal chromosomes (Angelman's syndrome).

Imprinting and the related phenomenon of allelic exclusion may be more common than currently documented, as it is difficult to examine levels of mRNA expression from the maternal and paternal alleles in specific tissues or in individual cells. Genomic imprinting, or uniparental disomy, is involved in the pathogenesis of several other disorders and malignancies (Chap. 62). For example, hydatidiform

moles contain a normal number of diploid chromosomes, but they are all of paternal origin. The opposite situation occurs in ovarian teratomata, with 46 chromosomes of maternal origin. Expression of the imprinted gene for insulin-like growth factor II (IGF-II) is involved in the pathogenesis of the cancer-predisposing Beckwith-Wiedemann syndrome (BWS) (Chap. 83). These children show somatic overgrowth with organomegalies and hemihypertrophy, and they have an increased risk of embryonal malignancies such as Wilms' tumor. Normally, only the paternally derived copy of the *IGF-II* gene is active and the maternal copy is inactive. Imprinting of the *IGF-II* gene is regulated by *H19*, which encodes an RNA transcript that is not translated into protein. Disruption or lack of *H19* methylation leads to a relaxation of *IGF-II* imprinting and expression of both alleles.

Meiotically and mitotically heritable changes in gene expression not associated with DNA sequence alterations are referred to as *epigenetic effects*. These changes involve DNA methylation, histone modifications, and RNA-mediated silencing, resulting in gene repression without a change in the coding sequence. Epigenetic alterations are increasingly recognized to play a role in human diseases such as cancer, mental retardation, hematologic disorders, and possibly in aging. For example, de novo methylation of CpG islands, regions of >500 bp in size with a GC content >55% in promoter regions that are normally unmethylated, is a hallmark of human cancers. Inhibitors of enzymes controlling epigenetic modifications such as histone deacetylases and DNA methyltransferases reverse gene silencing and represent a promising new group of antineoplastic agents.

Somatic Mutations

Cancer can be defined as a genetic disease at the cellular level (Chap. 83). Cancers are monoclonal in origin, indicating that they have arisen from a single precursor cell with one or several mutations in genes controlling growth (proliferation or apoptosis) and/or differentiation. These acquired somatic mutations are restricted to the tumor and its metastases and are not found in the surrounding normal tissue. The molecular alterations include dominant gain-of-function mutations in oncogenes, recessive loss-of-function mutations in tumor-suppressor genes and DNA repair genes, gene amplification, and chromosome rearrangements. Rarely, a single mutation in certain genes may be sufficient to transform a normal cell into a malignant cell. In most cancers, however, the development of a malignant phenotype requires several genetic alterations for the gradual progression from a normal cell to a cancerous cell, a phenomenon termed *multistep carcinogenesis* (Chaps. 83 and 84). Genomewide analyses of cancers using deep sequencing often reveal somatic rearrangements and mutations in multiple genes. Most human tumors express telomerase, an enzyme formed of a protein and an RNA component, which adds telomere repeats at the ends of chromosomes during replication. This mechanism impedes shortening of the telomeres, which is associated with senescence in normal cells, and is associated with enhanced replicative capacity in cancer cells. Telomerase inhibitors may provide a novel strategy for treating advanced human cancers.

In many cancer syndromes, there is an inherited *predisposition* to tumor formation. In these instances, a germline mutation is inherited in an autosomal dominant fashion inactivating one allele of an autosomal tumor-suppressor gene. If the second allele is inactivated by a somatic mutation or by epigenetic silencing in a given cell, this will lead to neoplastic growth (Knudson two-hit model). Thus, the defective allele in the germline is transmitted in a dominant mode, though tumorigenesis results from a biallelic loss of the tumor-suppressor gene in an affected tissue. The classic example to illustrate this phenomenon is retinoblastoma, which can occur as a sporadic or hereditary tumor. In sporadic retinoblastoma, both copies of the retinoblastoma (*RB*) gene are inactivated through two somatic events. In hereditary retinoblastoma, one mutated or deleted *RB* allele is inherited in an autosomal dominant manner and the second allele is inactivated by a subsequent somatic mutation. This two-hit model applies to other inherited cancer syndromes such as MEN-1 (Chap. 351) and neurofibromatosis type 2 (Chap. 379).

Nucleotide Repeat Expansion Disorders

Several diseases are associated with an increase in the number of nucleotide repeats above a certain threshold **(Table 61-6)**. The repeats are sometimes located within the coding region of the genes, as in Huntington's disease or the X-linked form of spinal and bulbar muscular atrophy (SBMA, Kennedy's syndrome). In other instances, the repeats probably alter gene regulatory sequences. If an expansion is present, the DNA fragment is unstable and tends to expand further during cell division. The length of the nucleotide repeat often correlates with the severity of the disease. When repeat length increases from one generation to the next, disease manifestations may worsen or be observed at an earlier age; this phenomenon is referred to as *anticipation*. In Huntington's disease, for example, there is a correlation between age of onset and length of the triplet codon expansion (Chap. 366). Anticipation has also been documented in other diseases caused by dynamic mutations in trinucleotide repeats (Table 61-6). The repeat number may also vary in a tissue-specific manner. In myotonic dystrophy, the CTG repeat may be tenfold greater in muscle tissue than in lymphocytes (Chap. 387).

Table 61-6 Selected Trinucleotide Repeat Disorders

Disease	Locus	Repeat	Triplet Length	Inheritance	Gene Product

			(Normal/Disease)		
X-chromosomal spinobulbar muscular atrophy (SBMA)	Xq11-q12	CAG	11–34/40–62	XR	Androgen receptor
Fragile X-syndrome (FRAXA)	Xq27.3	CGG	6–50/200–300	XR	FMR-1 protein
Fragile X-syndrome (FRAXE)	Xq28	GCC	6–25/>200	XR	FMR-2 protein
Dystrophia myotonica (DM)	19q13.2-q13.3	CTG	5–30/200–1000	AD, variable penetrance	Myotonin protein kinase
Huntington's disease (HD)	4p16.3	CAG	6–34/37–180	AD	Huntingtin
Spinocerebellar ataxia type 1 (SCA1)	6p21.3-21.2	CAG	6–39/40–88	AD	Ataxin 1
Spinocerebellar ataxia type 2 (SCA2)	12q24.1	CAG	15–31/34–400	AD	Ataxin 2
Spinocerebellar ataxia type 3 (SCA3); Machado-Joseph disease (MD)	14q21	CAG	13–36/55–86	AD	Ataxin 3
Spinocerebellar ataxia type 6 (SCA6, CACNAIA)	19p13.1-13.2	CAG	4–16/20–33	AD	Alpha 1A voltage-dependent L-type calcium chan-nel
Spinocerebellar ataxia type 7 (SCA7)	3p21.1-p12	CAG	4–19/37 to >300	AD	Ataxin 7
Spinocerebellar ataxia type 12 (SCA12)	5q31	CAG	6–26/66–78	AD	Protein phosphatase 2A
Dentorubral pallidoluysian atrophy (DRPLA)	12p	CAG	7–23/49–75	AD	Atrophin 1
Friedreich ataxia (FRDA1)	9q13-21	GAA	7–22/200–900	AR	Frataxin

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; XR, X-linked recessive.

COMPLEX GENETIC DISORDERS

The expression of many common diseases such as cardiovascular disease, hypertension, diabetes, asthma, psychiatric disorders, and certain cancers is determined by a combination of genetic background, environmental factors, and lifestyle. A trait is called *polygenic* if multiple genes contribute to the phenotype or *multifactorial* if multiple genes are assumed to interact with environmental factors. Genetic models for these complex traits need to account for genetic heterogeneity and interactions with other genes and the environment. Complex genetic traits may be influenced by modifying genes that are not linked to the main gene involved in the pathogenesis of the trait. This type of gene-gene interaction, or *epistasis*, plays an important role in polygenic traits that require the simultaneous presence of variations in multiple genes to result in a pathologic phenotype.

Type 2 diabetes mellitus provides a paradigm for considering a multifactorial disorder, as genetic, nutritional, and lifestyle factors are intimately interrelated in disease pathogenesis **(Table 61-7)** (Chap. 344). The identification of genetic variations and environmental factors that either predispose to or protect against disease is essential for predicting disease risk, designing preventive strategies, and developing novel therapeutic approaches. The study of rare monogenic diseases may provide insight into some of genetic and molecular mechanisms important in the pathogenesis of complex diseases. For example, the identification of the hepatocyte nuclear factor 1α (HNF1α) in maturity-onset of diabetes type 4 defined it as a *candidate gene* in the pathogenesis of diabetes mellitus type 2 (Tables 61-2 and **61-7**). Genome scans have identified various loci that may be associated with susceptibility to development of diabetes mellitus in certain populations. Efforts to identify susceptibility genes require very large sample sizes, and positive results may depend on ethnicity, ascertainment criteria, and statistical analysis. Association studies analyzing the potential influence of (biologically functional) SNPs and SNP haplotypes on a particular phenotype are providing new insights into the genes involved in the pathogenesis of these common disorders. Large variants [(micro)deletions, duplications, and inversions] present in the human population also contribute to the pathogenesis of complex disorders, but their contributions remain poorly understood.

Table 61-7 Genes and Loci Involved in Mono- and Polygenic Forms of Diabetes

Disorder	Genes or Susceptibility Locus	Chromosomal Location	Other Factors
Monogenic forms of diabetes			
MODY 1	HNF4α (hepatocyte nuclear factor	20q12-q13.1	AD

	4α)		inheritance
MODY 2	GCK (glucokinase)	7p15-p13	
MODY 3	HNF1α (hepatocyte nuclear factor 1α)	12q24.2	
MODY 4	IPF1 (insulin receptor substrate)	13q12.1	
MODY 5 (renal cysts, diabetes)	HNF1β (hepatocyte nuclear factor 1β)	17cen-q21.3	
MODY 6	NeuroD1 (neurogenic differention factor 1)	2q32	
Diabetes mellitus type 2; loci and genes linked and/or associated with susceptibility for diabetes mellitus type 2	Genes and loci identified by linkage/association studies		
	CPN10 (Calpain-10)	2q37.3	Diet
	HNF4α (hepatocyte nuclear factor 4α)	20q12-q13.1	Energy expenditure
	PTPN1 (protein-tyrosine phosphatase)	20q13.1-q13.2	Obesity
	PKLR (liver pyruvate kinase)	1q21	
	CASQ1 (calsequestrin 1) APM1 (adiponectin)	1q21 3q27	
	TCF7L2 (transcription factor 7-like 2)	10q25.3	
	1q21-23	1q21-23	
	2q	2q	
	3q22-27	3q22-27	
	8p21-23	8p21-23	
	11q	11q	
	12q24	12q24	
	15 18p11	15 18p11	
	20q	20q	
	20p	20p	
	Selected candidate genes with possible contribution		
	PPARγ (Peroxisome proliferator receptor γ)	3p25	
	KCNJ11(ATP-sensitive K channel Kir6.2)	11p15.1	
	ABCC8 (ATP-binding cassette, subfamily c, member 8)	11p15.1	
	Insulin VNTR	11p15	
	IRS-1 (insulin receptor substrate)	2q36	
	PGC1α (PPAR γ coactivator α)	4p15.1	
	ENPP1 (ectonucleotide pyrophosphatase/phosphodiesterase 1)	6q22-23	

Abbreviations: AD, autosomal dominant; MODY, maturity onset diabetes of the young; VNTR, variable number of tandem repeats.

LINKAGE AND ASSOCIATION STUDIES

There are two primary strategies for mapping genes that cause or increase susceptibility to human disease: (1) classic linkage can be performed based on a known genetic model or, when the model is unknown, by studying pairs of affected relatives; or (2) disease genes can be mapped using allelic association studies (Table 61-8).

Table 61-8 Genetic Approaches for Identifying Disease Genes

Method	Indications and Advantages	Limitations

Linkage Studies		
Classical linkage analysis (parametric methods)	Analysis of monogenic traits Suitable for genome scan Control population not required Useful for multifactorial disorders in isolated populations	Difficult to collect large informative pedigrees Difficult to obtain sufficient statistical power for complex traits
Allele-sharing methods (nonparametric methods) Affected sib and relative pair analyses Sib pair analysis	Suitable for identification of susceptibility genes in polygenic and multifactorial disorders Suitable for genome scan Control population not required if allele frequencies are known Statistical power can be increased by including parents and relatives	Difficult to collect sufficient number of subjects Difficult to obtain sufficient statistical power for complex traits Reduced power compared to classical linkage, but not sensitive to specification of genetic mode
Association Studies		
Case-control studies Linkage disequilibrium Transmission disequilibrium test (TDT) Whole-genome association studies	Suitable for identification of susceptibility genes in polygenic and multifactorial disorders Suitable for testing specific allelic variants of known candidate loci Facilitated by HapMap data, making GWAS more feasible Does not necessarily need relatives	Requires large sample size and matched control population False-positive results in the absence of suitable control population Candidate gene approach does not permit to detect novel genes and pathways Whole-genome association studies very expensive

Abbreviation: GWAS, genome-wide association study

Genetic Linkage

Genetic linkage refers to the fact that genes are physically connected, or linked, to one another along the chromosomes. Two fundamental principles are essential for understanding the concept of linkage: (1) when two genes are close together on a chromosome, they are usually transmitted together, unless a recombination event separates them (Figs. 61-3 and 61-7); and (2) the odds of a crossover, or recombination event, between two linked genes is proportional to the distance that separates them. Thus, genes that are farther apart are more likely to undergo a recombination event than genes that are very close together. The detection of chromosomal loci that segregate with a disease by linkage can be used to identify the gene responsible for the disease (*positional cloning*) and to predict the odds of disease gene transmission in genetic counseling.

Polymorphisms are essential for linkage studies because they provide a means to distinguish the maternal and paternal chromosomes in an individual. On average, 1 out of every 1000 bp varies from one person to the next. Although this degree of variation seems low (99.9% identical), it means that >3 million sequence differences exist between any two unrelated individuals and the probability that the sequence at such loci will differ on the two homologous chromosomes is high (often >70–90%). These sequence variations include variable number of tandem repeats (VNTRs), short tandem repeats (STRs), and SNPs. Most STRs, also called *polymorphic microsatellite markers*, consist of di-, tri-, or tetranucleotide repeats that can be measured readily using PCR **(Fig. 61-12)**. Characterization of SNPs, using DNA chips or beads, permit comprehensive analyses of genetic variation, linkage, and association studies. Although these sequence variations usually have no apparent functional consequences, they provide much of the basis for variation in genetic traits.

Figure 61-12

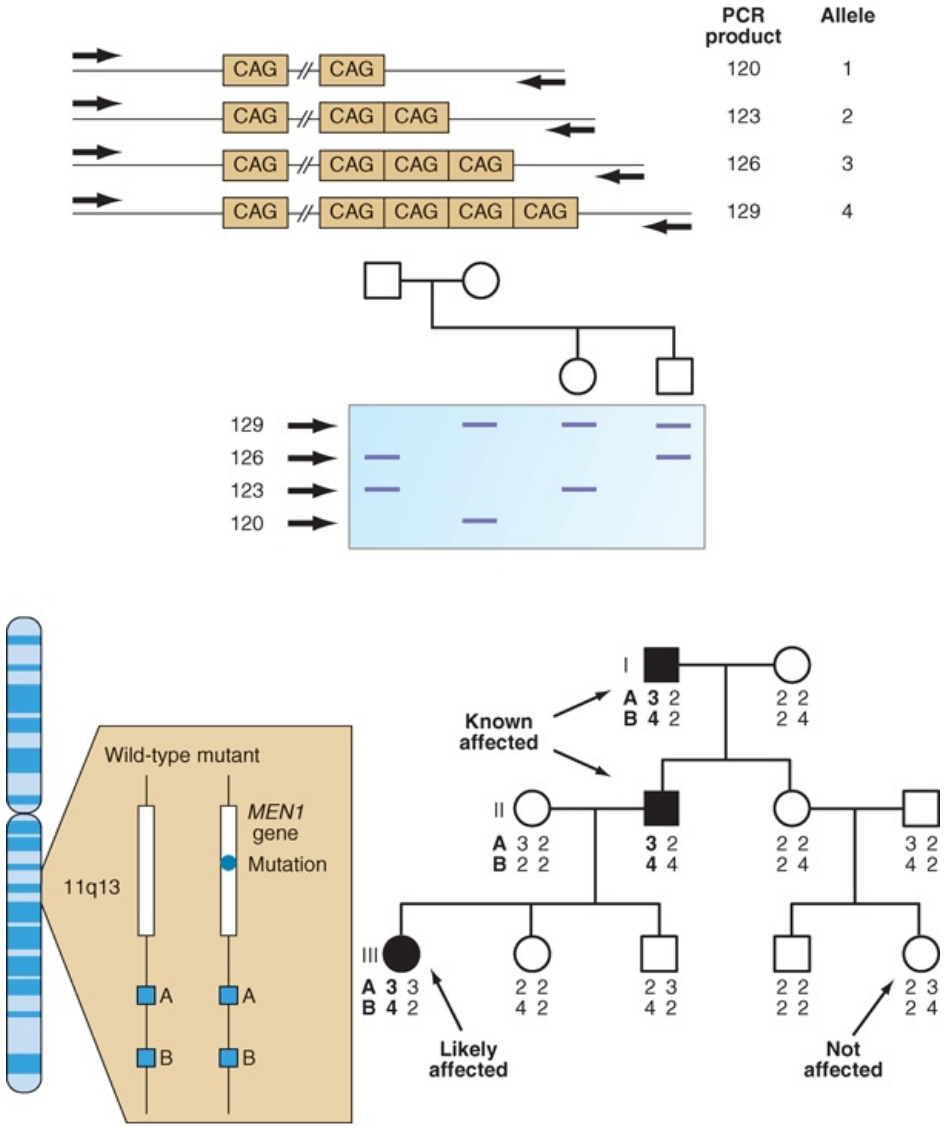

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CAG repeat length and linkage analysis in multiple endocrine neoplasia (MEN) type 1. *Upper panel.* Detection of different alleles using polymorphic microsatellite markers. The example depicts a CAG trinucleotide repeat. PCR with primers flanking the polymorphic region results in products of variable length, depending on the number of CAG repeats. After characterization of the alleles in the parents, transmission of the paternal and maternal alleles can be determined. ***Lower panel.*** Genotype analysis using microsatellite markers in a family with MEN-1. Two microsatellite markers, A and B, are located in close proximity to the *MEN1* gene on chromosome 11q13. For each individual, the A and B alleles have been determined. Based on this analysis, the genotype A3,B4 is linked to the disease because it occurs in the two affected individuals I-1 and II-2 but not in unaffected siblings. Because the disease allele is linked to A3,B4 within the affected family, it is likely that the individual III-1 is a carrier of the mutated *MEN1* gene. Although III-5 also has the A3,B4 genotype, she has inherited the allele from her unaffected father (II-4), who is not related to the original family. The A3,B4 genotype is only associated with MEN-1 in the original family, but not in the general population. Therefore, individual III-5 is not at risk for developing the disease.

In order to identify a chromosomal locus that segregates with a disease, it is necessary to characterize polymorphic DNA markers from affected and unaffected individuals of one or several pedigrees. One can then assess whether certain marker alleles cosegregate with the disease. Markers that are closest to the disease gene are less likely to undergo recombination events and therefore receive a higher linkage score. Linkage is expressed as a lod (logarithm of odds) score'the ratio of the probability that the disease and marker loci are linked rather than unlinked. Lod scores of +3 (1000:1) are generally accepted as supporting linkage, whereas a score of –2 is consistent with the absence of linkage.

An example of the use of linkage analysis is shown in Fig. 61-12. In this case, the gene for the autosomal dominant disorder MEN-1 is known to be located on chromosome 11q13. Using positional cloning, the *MEN1* gene was identified and shown to encode menin, a tumor suppressor. Affected individuals inherit a mutant form of the *MEN1* gene, predisposing them to certain types of tumors (parathyroid, pituitary, pancreatic islet) (Chap. 351). In the tissues that develop a tumor, a "second hit"

occurs in the normal copy of the *MEN1* gene. This somatic mutation may be a point mutation, a microdeletion, or loss of a chromosomal fragment (detected as loss of heterozygosity, LOH). Within a given family, linkage to the *MEN1* gene locus can be assessed without necessarily knowing the specific mutation in the *MEN1* gene. Using polymorphic STRs that are close to the *MEN1* gene, one can assess transmission of the different *MEN1* alleles and compare this pattern to development of the disorder to determine which allele is associated with risk of MEN-1. In the pedigree shown, the affected grandfather in generation I carries alleles 3 and 4 on the chromosome with the mutated *MEN1* gene and alleles 2 and 2 on his other chromosome 11. Consistent with linkage of the 3/4 genotype to the *MEN1* locus, his son in generation II is affected, whereas his daughter (who inherits the 2/2 genotype from her father) is unaffected. In the third generation, transmission of the 3/4 genotype indicates risk of developing MEN-1, assuming that no genetic recombination between the 3/4 alleles and the *MEN1* gene has occurred. After a specific mutation in the *MEN1* gene is identified within a family, it is possible to track transmission of the mutation itself, thereby eliminating uncertainty caused by recombination.

Allelic Association, Linkage Disequilibrium, and Haplotypes

Allelic association refers to a situation in which the frequency of an allele is significantly increased or decreased in individuals affected by a particular disease in comparison to controls. Linkage and association differ in several aspects. Genetic linkage is demonstrable in families or sibships. Association studies, on the other hand, compare a population of affected individuals with a control population. Association studies can be performed as case-control studies that include unrelated affected individuals and matched controls, or as family-based studies that compare the frequencies of alleles transmitted or not transmitted to affected children.

Allelic association studies are particularly useful for identifying susceptibility genes in complex diseases. When alleles at two loci occur more frequently *in combination* than would be predicted (based on known allele frequencies and recombination fractions), they are said to be in *linkage disequilibrium*. Evidence for linkage disequilibrium can be helpful in mapping disease genes because it suggests that the two loci are tightly linked.

Detecting the genetic factors contributing to the pathogenesis of common complex disorders remains a great challenge. In many instances, these are low-penetrancealleles (i.e., variations that individually only have a subtle effect on disease development, and they can only be identified by unbiased GWAS). Most variants are in noncoding or regulatory sequences but do not alter protein structure. The analysis of complex disorders is further complicated by ethnic differences in disease prevalence, differences in allele frequencies in known susceptibility genes among different populations, locus and allelic heterogeneity, gene-gene and gene-environment interactions, and the possibility of phenocopies. The data generated by the HapMap Project are greatly facilitating GWAS for the characterization of complex disorders. Adjacent SNPs are inherited together as blocks, and these blocks can be identified by genotyping selected marker SNPs, so-called *Tag SNPs*, thereby reducing cost and workload (Fig. 61-8). The availability of this information permits the characterization of a limited number of SNPs to identify the set of haplotypes present in an individual (e.g., in cases and controls). This, in turn, permits GWAS by searching for associations of certain haplotypes with a disease phenotype of interest, an essential step for unraveling the genetic factors contributing to complex disorders.

Population Genetics

In population genetics, the focus changes from alterations in an individual's genome to the distribution pattern of different genotypes in the population. In a case where there are only two alleles, A and a, the frequency of the genotypes will be $p^2 + 2pq + q^2 = 1$, with p^2 corresponding to the frequency of AA, $2pq$ to the frequency of Aa, and q^2 to aa. When the frequency of an allele is known, the frequency of the genotype can be calculated. Alternatively, one can determine an allele frequency, if the genotype frequency has been determined.

Allele frequencies vary among ethnic groups and geographic regions. For example, heterozygous mutations in the *CFTR* gene are relatively common in populations of European origin but are rare in the African population. Allele frequencies may vary because certain allelic variants confer a selective advantage. For example, heterozygotes for the sickle cell mutation, which is particularly common in West Africa, are more resistant to malarial infection because the erythrocytes of heterozygotes provide a less favorable environment for *Plasmodium* parasites. Though homozygosity for the sickle cell gene is associated with severe anemia and sickle crises (Chap. 104), heterozygotes have a higher probability of survival because of the reduced morbidity and mortality from malaria; this phenomenon has led to an increased frequency of the mutant allele. Recessive conditions are more prevalent in geographically isolated populations because of the more restricted gene pool.

Approach to the Patient: Inherited Disorders

For the practicing clinician, the family history remains an essential step in recognizing the possibility of a hereditary component. When taking the history, it is useful to draw a detailed pedigree of the first-

degree relatives (e.g., parents, siblings, and children), since they share 50% of genes with the patient. Standard symbols for pedigrees are depicted in Fig. 61-9. The family history should include information about ethnic background, age, health status, and (infant) deaths. Next, the physician should explore whether there is a family history of the same or related illnesses to the current problem. An inquiry focused on commonly occurring disorders such as cancers, heart disease, and diabetes mellitus should follow. Because of the possibility of age-dependent expressivity and penetrance, the family history will need intermittent updating. If the findings suggest a genetic disorder, the clinician will have to assess whether some of the patient's relatives may be at risk of carrying or transmitting the disease. In this circumstance, it is useful to confirm and extend the pedigree based on input from several family members. This information may form the basis for carrier detection, genetic counseling, early intervention, and prevention of a disease in relatives of the index patient (Chap. 63).

In instances where a diagnosis at the molecular level may be relevant, the physician will have to identify an appropriate laboratory that can perform the test. Genetic testing is available for a rapidly growing number of monogenic disorders through commercial laboratories. For uncommon disorders, the test may only be performed in a specialized research laboratory. Approved laboratories offering testing for inherited disorders can be identified in continuously updated on-line resources (GeneTests; Table 61-1). If genetic testing is considered, the patient and the family should be informed about the potential implications of positive results, including psychological distress and the possibility of discrimination. The patient or caretakers should be informed about the meaning of a negative result, technical limitations, and the possibility of false-negative and inconclusive results. For these reasons, genetic testing should only be performed after obtaining *informed consent*. Published ethical guidelines address the specific aspects that should be considered when testing children and adolescents. Genetic testing should usually be limited to situations in which the results may have an impact on the medical management.

IDENTIFYING THE DISEASE-CAUSING GENE

Genomic medicine aims to enhance the quality of medical care through the use of genotypic analysis (DNA testing) to identify genetic predisposition to disease, to select more specific pharmacotherapy, and to design individualized medical care based on genotype. Genotype can be deduced by analysis of protein (e.g., hemoglobin, apoprotein E), mRNA, or DNA. However, technologic advances have made DNA analysis particularly useful because it can be readily applied.

DNA testing is performed by mutational analysis or linkage studies in individuals at risk for a genetic disorder known to be present in a family. Mass screening programs require tests of high sensitivity and specificity to be cost effective. Prerequisites for the success of genetic screening programs include the following: that the disorder is potentially serious; that it can be influenced at a presymptomatic stage by changes in behavior, diet, and/or pharmaceutical manipulations; and that the screening does not result in any harm or discrimination. Screening in Jewish populations for the autosomal recessive neurodegenerative storage disease Tay-Sachs has reduced the number of affected individuals. In contrast, screening for sickle cell trait/disease in African Americans has led to unanticipated problems of discrimination by health insurers and employers. Mass screening programs harbor additional potential problems. For example, screening for the most common genetic alteration in cystic fibrosis, the ΔF508 mutation with a frequency of ~70% in northern Europe is feasible and seems to be effective. One has to keep in mind, however, that there is pronounced allelic heterogeneity and that the disease can be caused by >1700 other mutations. The search for these less common mutations would substantially increase costs but not the effectiveness of the screening program as a whole. Next-generation genome sequencing will permit comprehensive and cost-effective mutational analyses after selective enrichment of candidate genes. For example, tests that sequence all the common genes causing hereditary deafness are already commercially available. Occupational screening programs aim to detect individuals with increased risk for certain professional activities (e.g., α_1 antitrypsin deficiency and smoke or dust exposure).

MUTATIONAL ANALYSES

DNA sequence analysis is now widely used as a diagnostic tool and has significantly enhanced diagnostic accuracy. It is used for determining carrier status and for prenatal testing in monogenic disorders (Chap. 63). Numerous techniques are available for the detection of mutations **(Table 61-9)**. In a very broad sense, one can distinguish between techniques that allow for screening for the absence or presence of known mutations (screening mode) or techniques that definitively characterize mutations. Analyses of large alterations in the genome are possible using classic methods such as cytogenetics, fluorescent in situ hybridization (FISH), and Southern blotting (Chap. 62), as well as more sensitive novel techniques that search for multiple single exon deletions or duplications.

Table 61-9 Techniques Commonly Used for Mutation Detection

Method	Principle	Type of Mutation Detected
Cytogenetic analysis	Unique visual appearance of various chromosomes	Numerical or structural abnormalities in chromosomes

Fluorescent in situ hybridization (FISH)	Hybridization to chromosomes with fluorescently labeled probes	Numerical or structural abnormalities in chromosomes
Southern blot	Hybridization with genomic probe or cDNA probe after digestion of high-molecular-weight DNA	Large deletion, insertion, rearrangement, expansions of triplet repeat, amplification
Polymerase chain reaction (PCR)	Amplification of DNA segment	Expansion of triplet repeats, variable number of tandem repeats (VNTR), gene rearrangements, translocations; prepare DNA for other mutation methods
Reverse transcriptase PCR (RT-PCR)	Reverse transcription, amplification of DNA segment → absence or reduction of mRNA transcription	Analyze expressed mRNA (cDNA) sequence; detect loss of expression
Traditional DNA sequencing	Direct sequencing of PCR products Sequencing of DNA segments cloned into plasmid vectors	Point mutations, small deletions and insertions
Next-generation DNA sequencing	Sequencing of large contiguous genomic regions, exome of single or all chromosomes.	Deep sequencing for detection of mutations and SNPs. Sequencing of whole genomes of microorganisms
Microarrays	Hybridization of PCR products to wild-type or mutated oligonucleotides	Point mutations, small deletions and insertions Genotyping of SNPs

More discrete sequence alterations rely heavily on the use of the PCR, which allows rapid gene amplification and analysis. Moreover, PCR makes it possible to perform genetic testing and mutational analysis with small amounts of DNA extracted from leukocytes or even from single cells, buccal cells, or hair roots. DNA sequencing can be performed directly on PCR products or on fragments cloned into plasmid vectors amplified in bacterial host cells. Sequencing of all exons of the genome or selected chromosomes, or sequencing of numerous candidate genes in a single run, is now possible with next-generation sequencing platforms.

The majority of traditional diagnostic methods were gel-based. Novel technologies for the analysis of mutations, genotyping, large-scale sequencing, and mRNA expression profiles are currently undergoing rapid evolution. DNA chip technologies allow hybridization of DNA or RNA to hundreds of thousands of probes simultaneously. Microarrays are being used clinically for mutational analysis of several human disease genes, as well as for the identification of viral or bacterial sequence variations. With advances in high-throughput DNA-sequencing technology, complete sequencing of the genome of an individual is expected to cost about $1000 within this decade. Although comprehensive sequencing of large genomic regions or multiple genes is already a reality, the subsequent bioinformatics analysis, assembly of sequence fragments, and comparative alignments remains a significant and commonly underestimated challenge for data handling.

The huge amount of data generated by massive parallel sequencing underscores the need for ongoing attention to ethical and legal concerns. They include, among others, issues related to consent, sequence interpretation, the discovery of incidental findings that are predictors of serious disorders, the ability to link the information to an individual despite the absence of any conventional personal identifier, and data storage.

A general algorithm for the approach to mutational analysis is outlined in Fig. 61-13. The importance of a detailed clinical phenotype cannot be overemphasized. This is the step where one should also consider the possibility of genetic heterogeneity and phenocopies. If obvious candidate genes are suggested by the phenotype, they can be analyzed directly. After identification of a mutation, it is essential to demonstrate that it segregates with the phenotype. The functional characterization of novel mutations is labor intensive and may require analyses in vitro or in transgenic models in order to document the relevance of the genetic alteration.

Figure 61-13

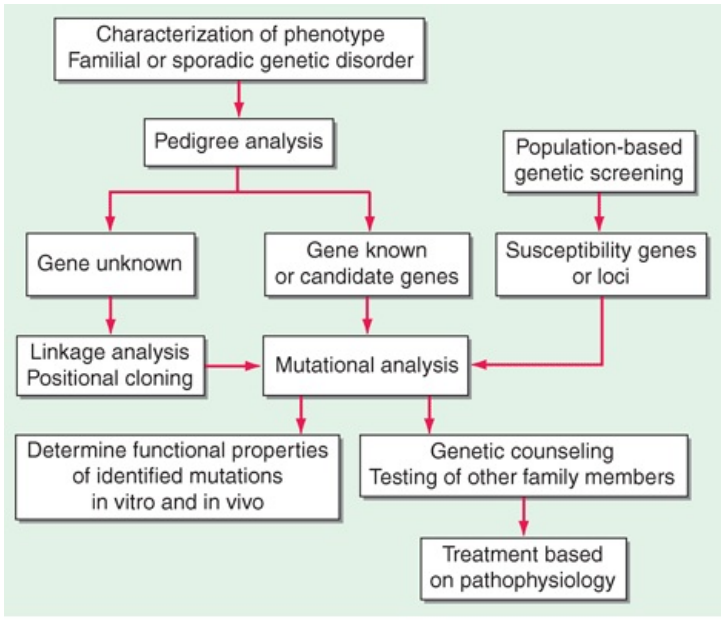

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to genetic disease.

Prenatal diagnosis of numerous genetic diseases in instances with a high risk for certain disorders is now possible by direct DNA analysis. *Amniocentesis* involves the removal of a small amount of amniotic fluid, usually at 16 weeks of gestation. Cells can be collected and submitted for karyotype analyses, FISH, and mutational analysis of selected genes. The main indications for amniocentesis include advanced maternal age (>35 years), an abnormal serum triple marker test (α-fetoprotein, β human chorionic gonadotropin, pregnancy-associated plasma protein A, or unconjugated estriol), a family history of chromosomal abnormalities, or a Mendelian disorder amenable to genetic testing. Prenatal diagnosis can also be performed by *chorionic villus sampling* (CVS), in which a small amount of the chorion is removed by a transcervical or transabdominal biopsy. Chromosomes and DNA obtained from these cells can be submitted for cytogenetic and mutational analyses. CVS can be performed earlier in gestation (weeks 9–12) than amniocentesis, an aspect that may be of relevance when termination of pregnancy is a consideration. Later in pregnancy, beginning at about 18 weeks of gestation, percutaneous umbilical blood sampling (PUBS) permits collection of fetal blood for lymphocyte culture and analysis. In combination with in vitro fertilization (IVF) techniques, it is even possible to perform genetic diagnoses in a single cell removed from the four- to eight-cell embryo or to analyze the first polar body from an oocyte. Preconceptual diagnosis thereby avoids therapeutic abortions but is extremely costly and labor intensive. Lastly, it has to be emphasized that excluding a specific disorder by any of these approaches is never equivalent to the assurance of having a normal child.

Mutations in certain cancer susceptibility genes such as *BRCA1* and *BRCA2* may identify individuals with an increased risk for the development of malignancies and result in risk-reducing interventions. The detection of mutations is an important diagnostic and prognostic tool in leukemias and lymphomas. The demonstration of the presence or absence of mutations and polymorphisms is also relevant for the rapidly evolving field of pharmacogenomics, including the identification of differences in drug treatment response or metabolism as a function of genetic background. For example, the thiopurine drugs 6-mercaptopurine and azathioprine are commonly used cytotoxic and immunosuppressive agents. They are metabolized by thiopurine methyltransferase (TPMT), an enzyme with variable activity associated with genetic polymorphisms in 10% of whites and complete deficiency in about 1/300 individuals. Patients with intermediate or deficient TPMT activity are at risk for excessive toxicity, including fatal myelosuppression. Characterization of these polymorphisms allows mercaptopurine doses to be modified based on TPMT genotype. Pharmacogenomics may increasingly permit individualized drug therapy, improve drug effectiveness, reduce adverse side effects, and provide cost-effective pharmaceutical care.

FURTHER READINGS

Antonarakis SE et al: Mendelian disorders and multifactorial traits: The big divide or one for all? Nat Rev Genet 11:380, 2010[PMID: 20395971] [Full Text]

Eichler EE et al: Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11:446, 2010[PMID: 20479774] [Full Text]

Guttmacher AE et al: Personalized genomic information: Preparing for the future of genetic medicine. Nat Rev Genet 11:161, 2010[PMID: 20065954] [Full Text]

Hamburg MA, Collins FS: The path to personalized medicine. N Engl J Med 363:301, 2010[PMID: 20551152] [Full Text]

Hudson KL et al: Keeping pace with the times—The Genetic Information Nondiscrimination Act of 2008. N Engl J Med 358:2661, 2008[PMID: 18565857] [Full Text]

La Spada AR, Taylor JP: Repeat expansion disease: Progress and puzzles in disease pathogenesis. Nat Rev Genet 11:247, 2010

Metzker ML: Sequencing technologies—the next generation. Nat Rev Genet 11:31, 2010[PMID: 19997069] [Full Text]

Pang T: The impact of genomics on global health. Am J Public Health 92:1077, 2002[PMID: 12084683] [Full Text]

Smith RD et al: Genomics knowledge and equity: A global public goods perspective of the patent system. Bull World Health Organ 82:385, 2004[PMID: 15298230] [Full Text]

Waalen J, Beutler E: Genetic screening for low-penetrance variants in protein-coding genes. Annu Rev Genomic Hum Genet 10:431, 2009[PMID: 19715441] [Full Text]

Zhang F et al: Copy number variation in human health, disease, and evolution. Annu Rev Genomic Hum Genet 10:451, 2009[PMID: 19775440] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHROMOSOME DISORDERS: INTRODUCTION

In humans, the normal diploid number of chromosomes is 46, consisting of 22 pairs of autosomal chromosomes (numbered 1–22 in decreasing size) and one pair of sex chromosomes (XX in females and XY in males). The genome is estimated to contain approximately 25,000 genes. Even the smallest autosome contains between 200 and 300 genes. Not surprisingly, duplications or deletions of chromosomes, or even small chromosome segments, have profound consequences on normal gene expression, leading to severe developmental and physiologic abnormalities.

Deviations in number or structure of the 46 human chromosomes are astonishingly common, despite severe deleterious consequences. Chromosomal disorders occur in an estimated 10–25% of all pregnancies. They are the leading cause of fetal loss and, among pregnancies surviving to term, the leading known cause of birth defects and mental retardation.

In recent years, the practice of cytogenetics has shifted from conventional cytogenetic methodology to a union of cytogenetic and molecular techniques. Formerly the province of research laboratories, fluorescence in situ hybridization (FISH), genomic array analysis, and related molecular cytogenetic technologies have been incorporated into everyday practice in clinical laboratories. As a result, there is an increased appreciation of the importance of "subtle" constitutional cytogenetic abnormalities such as micro-deletions and imprinting disorders, as well as previously recognized translocations and disorders of chromosome number.

VISUALIZING CHROMOSOMES

Conventional Cytogenetic Analysis

In theory, chromosome preparations can be obtained from any actively dividing tissue by causing the cells to arrest in metaphase, the stage of the cell cycle when chromosomes are maximally condensed. In practice, only a small number of tissues are used for routine chromosome analysis: amniocytes or chorionic villi for prenatal testing and blood, bone marrow, or skin fibroblasts for postnatal studies. Samples of blood, bone marrow, and chorionic villi can be processed using short-term culture techniques that yield results in 1–3 days. Analysis of other tissue types typically involves long-term cell culture, requiring 1–3 weeks of processing before cytogenetic analysis is possible.

Cells are isolated at metaphase or prometaphase and treated chemically or enzymatically to reveal chromosome "bands" **(Fig. 62-1)**. Analysis of the number of chromosomes in the cell and the distribution of bands on individual chromosomes allow the identification of numerical or structural abnormalities. This strategy is useful for characterizing the normal chromosome complement and determining the incidence and types of major chromosome abnormalities.

Figure 62-1

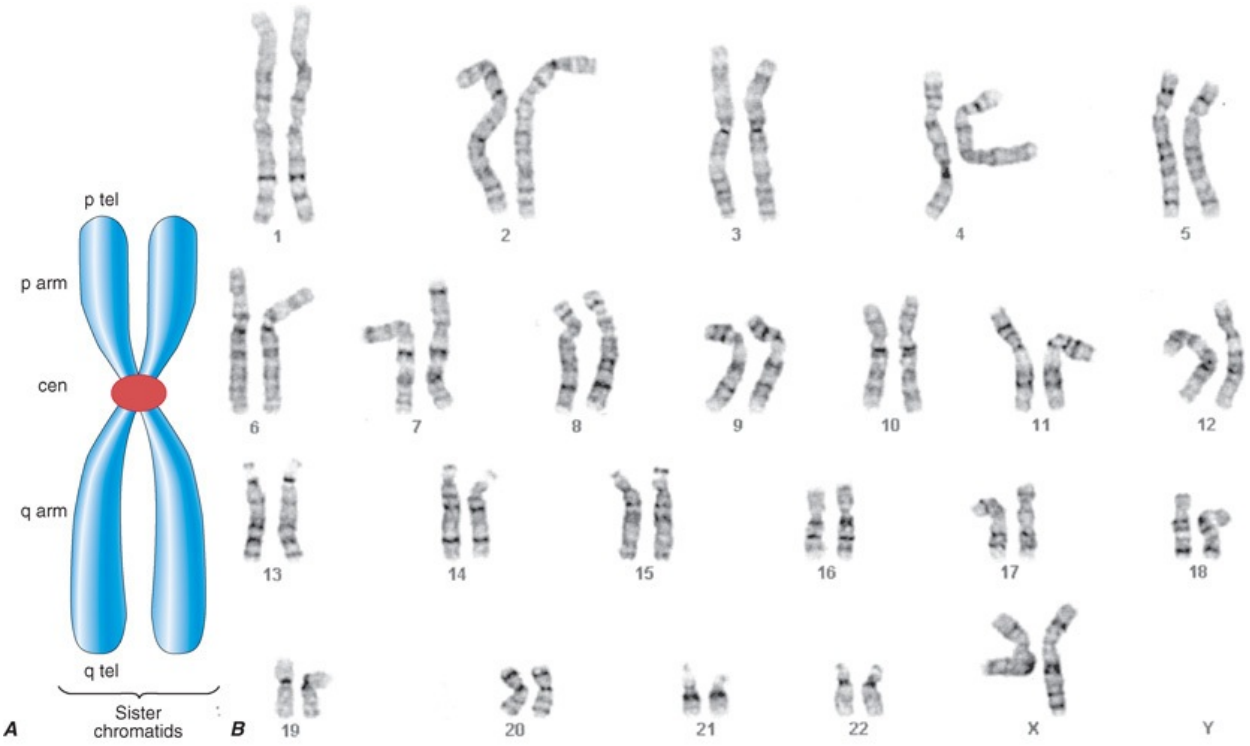

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. An idealized human chromosome, showing the centromere (cen), long (q) and short (p) arms, and telomeres (tel). ***B.*** A G-banded human karyotype from a normal (46,XX) female.

Each human chromosome contains two specialized structures: a centromere and two telomeres. The centromere, or primary constriction, divides the chromosome into short (p) and long (q) arms and is responsible for the segregation of chromosomes during cell division. The telomeres, or chromosome ends, "cap" the p and q arms and are important for allowing DNA replication at the ends of the chromosomes. Prior to DNA replication, each chromosome consists of a single chromatid copy of the DNA double helix. After DNA replication and continuing until the time of cell division (including metaphase, when chromosomes are typically visualized), each chromosome consists of two identical sister chromatids (Fig. 62-1).

Molecular Cytogenetics

The introduction of FISH methodologies in the late 1980s revolutionized the field of cytogenetics. In principle, FISH is similar to other DNA–DNA hybridization methodologies. In most instances, the probe is labeled directly with a fluorochrome to allow detection. After the hybridization step, the specimen is counterstained and the preparations are visualized with a fluorescence microscope.

TYPES OF FISH PROBES

A variety of probes are available for use with FISH, including chromosome-specific paints (chromosome libraries), repetitive probes, and single-copy probes **(Fig. 62-2)**. Chromosome libraries hybridize to sequences that span the entirety of the chromosome from which they are derived and, as a result, they can be used to "paint" individual chromosomes.

Figure 62-2

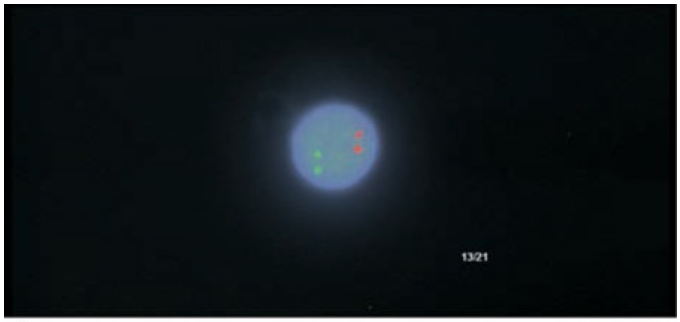

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

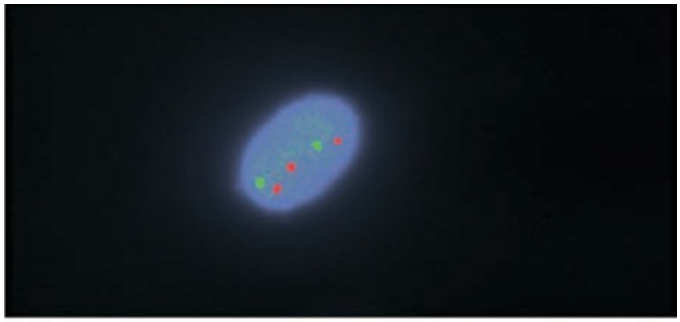

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

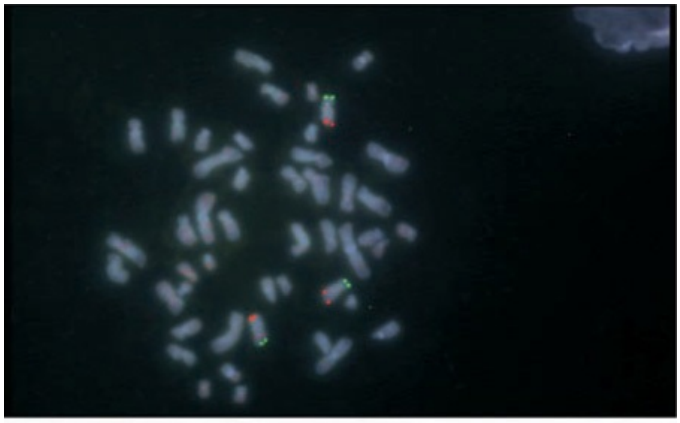

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

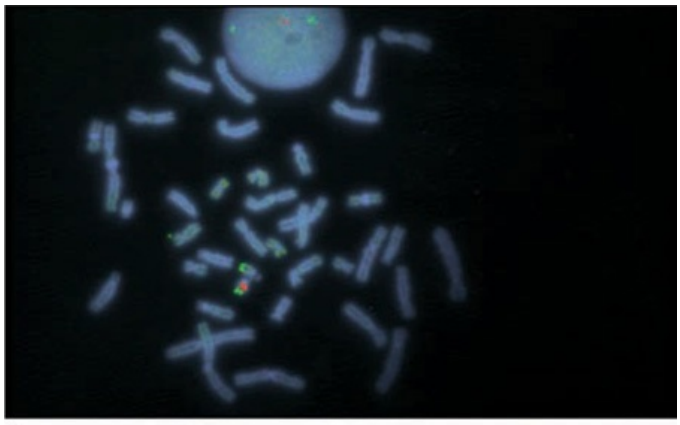

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

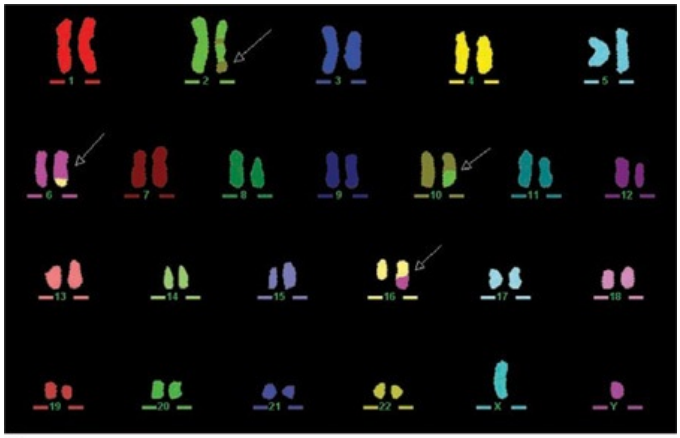

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examples of different applications of fluorescence in situ hybridization (FISH) to human metaphase and interphase preparations. ***A, B.*** Aneuploidy detection: Interphase FISH using chromosome 13 (green) and chromosome 21 (red) unique sequence probes on interphase cells from direct amniotic fluid preparations. In "A" (a normal cell), two signals for both chromosomes 13 and 21 are seen; in "B," three signals for chromosome 21 are seen, indicating trisomy 21 in the fetus. ***C.*** Aneuploidy detection: Two-color FISH with telomere probes from the short arm (green) and the long arm (red) of chromosome 8. Hybridization with these probes shows fluorescence of both probes to three separate chromosomes, indicating the presence of trisomy 8 in this individual. ***D.*** Microdeletion detection: Two-color FISH is used to detect a microdeletion of chromosome 22 associated with velocardiofacial (VCF) syndrome. A probe for ARSA (a locus on the distal portion of chromosome 22, visualized as a green signal) is observed on both chromosomes 22. However, a probe for TUPLE1 (a locus within the VCF region of chromosome 22, visualized in red) hybridizes to only the normal chromosome. ***E.*** Characterization of structural rearrangements: M-FISH (multicolor FISH) is used to detect a complex chromosome rearrangement involving a translocation between chromosomes 6 and 16, as well as a translocation and inversion involving chromosomes 2 and 10.

Repetitive probes recognize amplified DNA sequences present in chromosomes. The most common are α-satellite DNA probes that are complementary to DNA sequences found at the centromeric regions of all human chromosomes. A vast number of *single-copy probes* are now available as a result of the human genome project. These probes can be as small as 1 kb, though normally they are much larger and are packaged into fosmids (40 kb), or bacterial artificial chromosomes (BACs) (100–200 kb). The most widely used probes are available commercially, including probes for specific rearrangements identified in a variety of cancers.

APPLICATIONS OF FISH AND ARRAY TECHNOLOGY

The majority of FISH applications involve hybridization of one or two probes of interest as an adjunctive procedure to conventional chromosomal banding techniques. In this regard, FISH can be utilized to identify specific chromosomes, characterize de novo duplications or deletions, and clarify subtle chromosomal rearrangements. Its greatest utilization in constitutional analysis, however, is in the detection of microdeletions (see below). In cancer cytogenetics, it is used extensively in the analysis of structural rearrangements. Though conventional cytogenetic studies can detect some microdeletions, initial detection and/or confirmation with FISH is essential. In fact, since appropriate FISH probes have become available, detection of microdeletion syndromes has increased significantly.

In addition to metaphase FISH, cells can be analyzed at a variety of stages. Interphase analysis, for example, can be used to make a rapid diagnosis in instances when metaphase chromosome preparations are not yet available (e.g., amniotic fluid interphase analysis). Interphase analysis also increases the number of cells available for examination, allows for investigation of nuclear organization, and provides results when cells do not progress to metaphase. One specialized type of interphase analysis involves the application of FISH to paraffin-embedded sections, thereby preserving the architecture of the tissue.

The use of interphase FISH has increased recently, especially for analyses of amniocentesis samples. These studies are performed on uncultured amniotic fluid, typically using DNA probes specific for the chromosomes most commonly identified in trisomies (chromosomes 13, 18, 21, and the X and Y). These studies can be performed rapidly (24–48 hours) and will ascertain about 60% of the abnormalities detected prenatally. Another area in which interphase analysis is routinely utilized is cancer cytogenetics (Chap. 83). Many site-specific translocations are associated with specific types of malignancies. For example, there are probes available for both the Abelson (Abl) oncogene and breakpoint cluster region (bcr) involved in chronic myelogenous leukemia (CML); these probes are labeled in red and green, respectively; the fusion of these genes in CML combines the fluorescent colors and appears as a yellow hybridization signal.

In addition to standard metaphase and interphase FISH analyses, a number of enhanced techniques have been developed for specific types of analysis, including multicolor FISH techniques, reverse

painting, fiber FISH, and comparative genomic hybridization. *Spectral karyotyping* (SKY) and multicolor FISH (m-FISH) techniques use combinatorially labeled probes that create a unique color for individual chromosomes. This technology is useful in the identification of unknown chromosome material (such as accessory marker chromosomes containing additional material) but has been most commonly used with the complex rearrangements seen in cancer specimens.

Fiber FISH is a technique in which chromosomes are mechanically stretched, using a variety of different methods. It provides a higher resolution of analysis than conventional FISH.

FISH comparative genomic hybridization (CGH) is a method that can be used only when DNA is available from a specimen of interest. The entire DNA specimen from the sample of interest is labeled in one color (e.g., green), and the normal control DNA specimen is indicated by another color (e.g., red). These are mixed in equal amounts and hybridized to normal metaphase chromosomes. The red-to-green ratio is analyzed by a computer program that determines where the DNA of interest may have gains or losses of material. Many of the FISH technologies are now being supplanted by the utilization of array analysis as described below.

The major advance for examining human chromosomes is an extension of the FISH CGH technologies. Specifically, the initial development of CGH arrays used protocols that are similar to standard FISH CGH, except that test DNA is hybridized to DNAs that are spread on arrays, rather than hybridized to normal chromosomes. These arrays are a natural outgrowth of both cytogenetic and FISH analysis, as they can provide a whole-genome analysis (as cytogenetic analysis provides), but at a higher resolution. There are several different types of arrays. CGH arrays utilize either BACs or oligonucleotides, whereas genotyping arrays (SNP arrays) utilize single nucleotide polymorphisms **(Fig. 62-3)**. The resolution of these arrays can be up to 50 kb, far greater than for standard chromosome analysis. This technology has been used to study cryptic chromosomal imbalances in patients with mental retardation and multiple congenital anomalies, as well as in prenatal diagnosis. It has also been used to detect microdeletions and microduplications in cancer and in previously unidentified genomic disorders. This technology is still in development, but it will ultimately be the initial method of study for constitutional abnormalities.

Figure 62-3

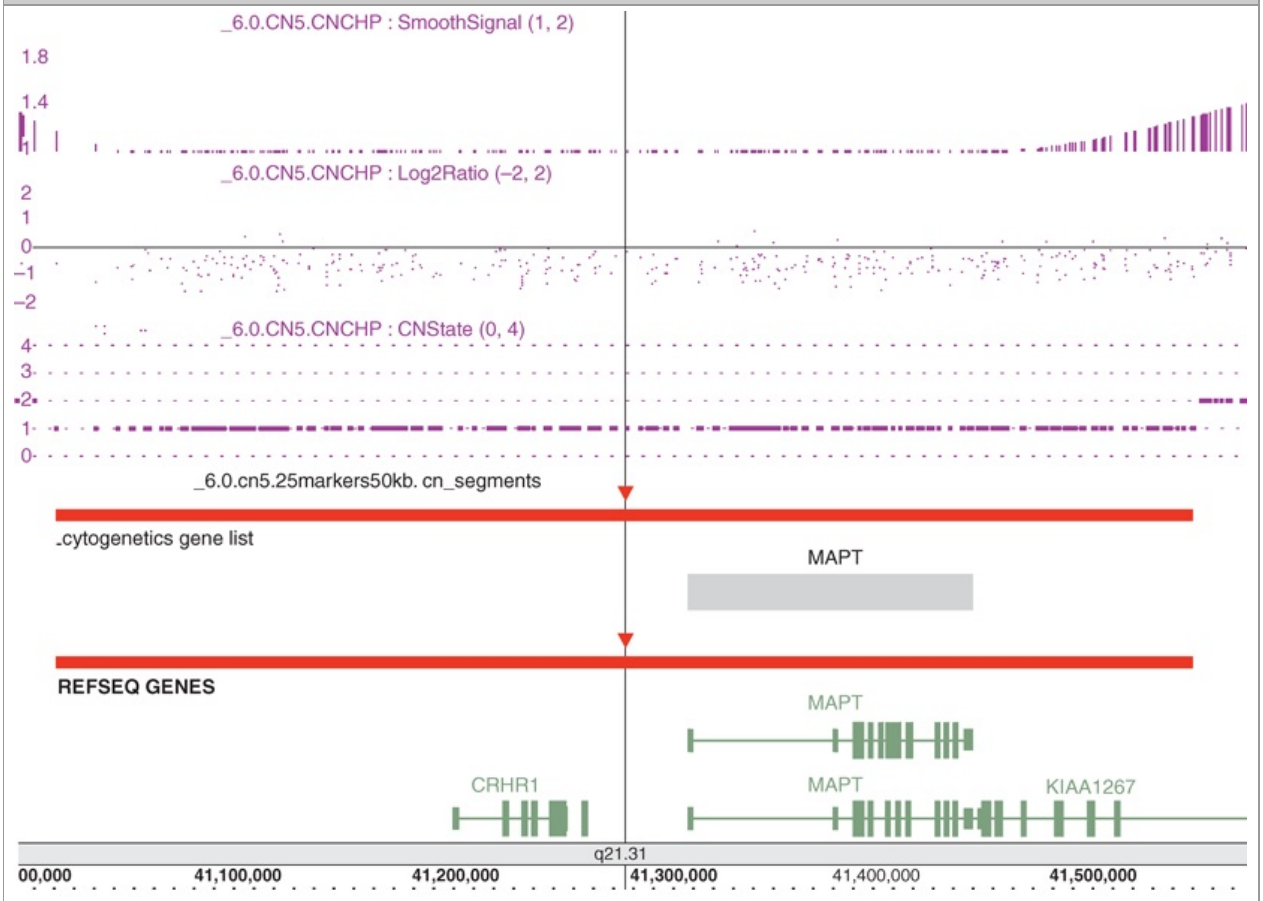

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Array analysis to diagnose human chromosome abnormalities. Detection of a 533-kb deletion in 17q21.31 that includes the *MAPT* gene that is responsible for a newly described microdeletion syndrome. The SNP array illustrates the deletion by changes in the log2 ratio, allele difference, smooth signal, and copy number change.

INDICATIONS FOR CYTOGENETIC ANALYSIS

Primary indications for karyotypic analysis vary according to the developmental stage/age of the conceptus/individual under investigation. One especially important application is in prenatal diagnosis (particularly for pregnancies involving older women or women identified at increased risk by screening modalities), assaying for chromosomal abnormalities in either chorionic villi of first-trimester fetuses or amniotic fluid of second-trimester fetuses. Tissue specimens from spontaneously aborted fetuses or stillbirths can also be examined for chromosome abnormalities. Interphase cytogenetics (using FISH) is increasingly being used to study individual blastomeres of preimplantation embryos (with in vitro fertilization-derived pregnancies). This makes it possible to detect aneuploid or structurally unbalanced embryos or, in the case of sex-linked disorders, to identify male conceptuses; such embryos would not be used to initiate pregnancies.

Among infants and children, peripheral blood is examined, most often in individuals with specific phenotypic abnormalities. For example, karyotypic analysis can be used for the confirmation or exclusion of a specific chromosomal syndrome (e.g., trisomy 21); in patients with unexplained psychomotor retardation with or without dysmorphic features; in cases of monogenic disorders associated with mental retardation and/or dysmorphic features; and with abnormalities of sexual differentiation and development. Additionally, the array analysis is being increasingly used in patients with autism.

In adults, peripheral blood can be examined in patients with infertility or recurrent miscarriages, since chromosome abnormalities can lead to meiotic arrest or to genetically unbalanced gametes. An important branch of cytogenetics is concerned with analyses of bone marrow, unstimulated peripheral blood, and lymph nodes of tumors, as chromosomal abnormalities are a common correlate of leukemia, lymphoma, and solid tumors (Chap. 83).

Cytogenetic Testing in Prenatal Diagnosis

The vast majority of prenatal diagnostic studies are performed to rule out a chromosomal abnormality, but cells may also be propagated for biochemical studies or molecular analyses of DNA. Three procedures are used to obtain samples for prenatal diagnosis: amniocentesis, chorionic villus sampling (CVS), and fetal blood sampling. Amniocentesis is the most common procedure and is routinely performed at 15–17 weeks of gestation. On some occasions, early amniocentesis at 12–14 weeks is performed to expedite results, although less fluid is obtained at this time. Early amniocentesis carries a greater risk of spontaneous abortion or fetal injury but provides results at an earlier stage of pregnancy.

The vast majority of amniocenteses are performed in the context of advanced maternal age, the best-known correlate of trisomy (see below). Additional reasons for amniocentesis referral include an abnormal "triple- or quad-marker assay" and/or detection of ultrasound abnormalities. In the second trimester, levels of human chorionic gonadotropin (HCG), α-fetoprotein, and unconjugated estriol (and, in the quad assay, inhibin) in the maternal serum are quantified and used to adjust the maternal age–predicted risk of a trisomy 21 or trisomy 18 fetus. More recently, first-trimester screening, involving the measurement of nuchal translucency, and the levels of PAPP-A and HCG are being used to identify women at increased risk. When integrated together, first- and second-trimester testing will identify approximately 93% of trisomy 21 pregnancies. Specific ultrasound abnormalities, when detected at midtrimester, can also be associated with chromosomal defects. When a nonspecific ultrasound abnormality is present, the estimated risk of a chromosomal defect is ~16%. Associations of chromosomal abnormalities and specific types of abnormal ultrasound findings are listed in **Table 62-1**.

Table 62-1 Frequency of Chromosome Abnormalities, Identified on the Basis of Abnormal Ultrasound Findings

	Chromosomal Abnormalities (Frequency)	
Ultrasound Finding	**Average, %**	**Range in Different Studies, %**
Abnormal ultrasound (nonspecific)	16	13–35
Omphalocele	39	26–54
Cystic hygroma	68	46–78
Congenital heart disease	30	8–40
Choroid plexus cyst	5	4–10

CVS is the second most common procedure for genetic prenatal diagnosis. Because this procedure is routinely performed at about 10–12 weeks of gestation, it allows for an earlier detection of abnormalities and a safer pregnancy termination, if desired. CVS is a relatively safe procedure (spontaneous abortions, <0.5–1%). Because there is an increased association of limb defects when the procedure is performed earlier (<10 weeks of gestation), CVS is applicable during a narrow time frame of gestation. CVS involves the use of a catheter inserted transvaginally; ~25 mg of villi are aspirated from the chorion frondosum (the fetal portion of the placenta). By adding colchicine directly to the rapidly dividing cytotrophoblasts, results can be obtained within 24–48 hours. Findings from these procedures should be confirmed by analyses of cultured mesenchymal cells, as they are more reliably derived from the fetus.

Percutaneous umbilical blood sampling (PUBS) is a method for obtaining fetal blood during the second

and third trimesters of pregnancy. PUBS is usually performed when ultrasound abnormalities are detected late in the second trimester. PUBS is also used when cytogenetic results from amniocentesis need clarification, such as in the detection of mosaicism.

CHROMOSOME ABNORMALITIES

Chromosomes in Cell Division

To understand the etiology of chromosome abnormalities, it is important to review the movement of chromosomes during cell division. In somatic tissues, chromosomes are replicated during the S-phase of the cell cycle, so that each replicated chromosome consists of two identical sister chromatids. When the cell enters mitosis, each of the 46 chromosomes align on the metaphase plate, with the centromeres co-oriented toward opposite spindle poles **(Fig. 62-4)**. At anaphase, the sister chromatids separate, with each of the daughter cells receiving one sister chromatid from each of the 46 chromosomes.

Figure 62-4

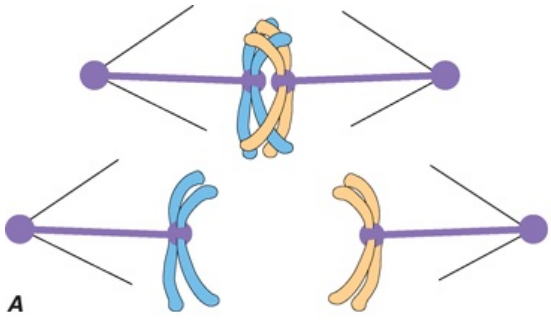

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

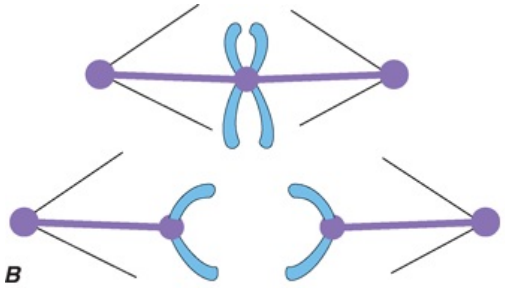

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chromosome segregation in meiosis. *A.* In meiosis I, each of the 23 pairs of chromosomes finds its "partner," or homologue, and exchanges genetic material (recombines) with it. At metaphase, each homologous pair aligns on the equatorial plate; at anaphase, each member of the homologous pair segregates from its partner. Thus, at the end of meiosis I, each daughter cell contains 23 chromosomes, with each chromosome consisting of two sister chromatids. ***B.*** In meiosis II, each chromosome aligns on the metaphase plate, and at anaphase, each of the two sister chromatids divides from the other. Thus, at the end of meiosis II, each daughter cell (e.g., the oocyte or spermatocyte) contains 23 chromosomes, with each chromosome consisting of one sister chromatid. In mitosis, the chromosomes behave exactly as they do in meiosis II, except that somatically dividing cells contain 46 chromosomes, not the 23 that are present in the meiosis II cell.

Chromosome segregation is more complicated in germ cell division, since the number of chromosomes must be reduced from 46 to 23 in the mature sperm and eggs. This is accomplished by two rounds of division'meiosis I and meiosis II (Fig. 62-4). In meiosis I, homologous chromosomes pair and exchange genetic material, then align on the metaphase plate, and finally separate from one another. Thus, by the end of meiosis I, only 23 of the original 46 chromosomes are represented in each of the two daughter cells. Meiosis II quickly follows meiosis I and is essentially a "haploid mitosis," involving separation of the sister chromatids in each of the 23 chromosomes.

Although the fundamentals of meiosis are the same in males and females, there are important distinctions, particularly in the timing of meiotic divisions. In males, meiosis begins with puberty and continues throughout the individual's lifetime. In females, meiosis begins prenatally, with oocytes proceeding through the first stages of meiosis I but arresting at mid-prophase. At the time of birth, the first meiotic division is suspended in oocytes. Only after ovulation many years later do oocytes complete meiosis I and proceed to the metaphase stage of meiosis II; if fertilized, the oocyte then completes the second meiotic division. Thus, in females, the first meiotic division takes at least 10–15 years and as many as 40–45 years to complete. Maternal age-related increases in the incidence of trisomy are likely the consequence of this protracted process of cell division.

Incidence and Types of Chromosome Abnormalities

Errors in meiosis, or in early cleavage divisions, occur with extraordinary frequency. At least 10–25% of all pregnancies, for example, involve chromosomally abnormal conceptions. A large proportion of these terminate in the earliest stages of pregnancy, many of which go unrecognized. Nevertheless, even among clinically recognized pregnancies, nearly 10% of fetuses are chromosomally unbalanced. For the three types of clinically recognized pregnancies—spontaneous abortions, stillbirths, and livebirths—the frequencies of different chromosomal abnormalities are summarized in **Table 62-2**. The most common abnormalities are numerical, involving fetuses with additional (trisomy) or missing (monosomy) chromosomes, or those with one (triploidy) or two (tetraploidy) additional sets of chromosomes. Structural chromosome abnormalities are much less common, although several of the most important clinical chromosomal disorders involve structural rearrangements (see below).

Table 62-2 Frequency and Distribution of Chromosome Abnormalities in Different Types of Clinically Recognizable Pregnancies

	Frequency of Abnormality			
Chromosome Abnormality	**Spontaneous Abortion**	**Stillbirth**	**Livebirth**	**Probability of Surviving to Term, %**
Trisomy, all	25.1	4.0	0.3	5
+13, 18, 21	4.5	2.7	0.14	15
+16	7.5	—	—	0
Sex chromosome monosomy (45,X)	8.7	0.1	0.01	1
Triploidy	6.4	0.2	—	0
Tetraploidy	2.4	—	—	0
Structural abnormality	2.0	0.8	0.3	45
Total abnormalities	50.0	5.1	0.6	5

By far the most common abnormality is trisomy, which is identified in ~25% of spontaneous abortions and 0.3% of newborns. Trisomies for all chromosomes have now been identified in embryos or fetuses, but there is considerable variation in frequency for various chromosomes. For example, trisomy 16 is extraordinarily common, accounting for about one-third of all trisomies in spontaneous abortions, whereas trisomies 1, 5, 11, and 19 have been identified less often. Available evidence suggests two reasons for this variation: (1) some chromosomes (e.g., chromosome 16) are more likely to segregate abnormally or undergo nondisjunction during meiosis than are others; and (2) the potential for development varies widely among different trisomic conditions, with some being eliminated very early in gestation, others surviving to the time of clinical pregnancy recognition, and some (e.g., trisomies 13, 18, and 21 and sex chromosome trisomies) being compatible with survival to term.

CHROMOSOMAL SYNDROMES

While most chromosomally abnormal conceptions perish in utero, several conditions are compatible with survival to term. The best-characterized of these are numerical abnormalities, involving loss or gain of individual chromosomes, and abnormalities resulting from unbalanced translocations. FISH, array analysis, and other molecular studies have led to the identification of two "new" types of chromosome abnormalities, commonly referred to as *microdeletion syndromes* and *imprinting syndromes*.

Numerical Abnormalities

Virtually all types of numerical abnormalities are eliminated prenatally, so that only those involving small, gene-poor autosomes or the sex chromosomes are identified with any frequency among liveborns. Clinically, the most important of these is trisomy 21, the most frequent cause of Down syndrome. Depending on the maternal age structure of the population and the utilization of prenatal testing, the incidence of trisomy 21 ranges from 1/600 to 1/1000 livebirths, making it the most common chromosome abnormality in live-born individuals. Like most trisomies, the incidence of trisomy 21 is highly correlated with maternal age, increasing from about 1/1500 livebirths for women 20 years of age to 1/30 for women ≥ 45 years.

In addition to trisomy 21, only two other autosomal trisomies, 13 and 18, occur with any frequency in livebirths. Incidence rates for trisomies 13 and 18 in livebirths are 1/20,000 and 1/10,000, respectively. Unlike trisomy 21 that is associated with near-normal life expectancy, both trisomies 13 and 18 are associated with death in infancy, typically occurring during the first year of life.

Three sex chromosome trisomies—the 47,XXX, 47,XXY (Klinefelter's syndrome), and 47,XYY conditions—are quite common, with each occurring in about 1/2000 newborns. Of all the trisomic conditions, these three have the fewest phenotypic complications. In fact, with the exception of infertility in Klinefelter's syndrome (Chap. 349), it is likely that most individuals with such trisomic conditions would go undetected. The additional Y chromosome in the 47,XYY condition is small and contains only a few genes. Most Y-linked genes are involved in testicular development or spermatogenesis. Thus, dosage imbalance of Y-

linked genes has relatively little effect on other developmental processes. The 47,XYY genotype is associated with increased height. Its role in antisocial behavior, postulated initially because of an increased prevalence among some penalized populations, is unclear.

For the 47,XXX and 47,XXY conditions, the situation is different—the X chromosome contains >1000 genes, many of them essential for normal development. How, then, are 47,XXX and 47,XXY individuals spared from the catastrophic consequences of dosage imbalance? The answer lies in the biology of X chromosome gene expression. In normal females, one of the chromosomes undergoes *X inactivation* in somatic cells. The inactivation of the paternal or maternal X chromosome occurs randomly in each somatic cell and thereby serves as a mechanism of dosage compensation, ensuring that males and females have equal expression of most X-linked genes. The inactivation process occurs at the blastocyst stage of development; prior to this, both X chromosomes are active. In addition, not all X-linked genes are inactivated. Some genes on the X chromosome "escape" the inactivating mechanism and are expressed from both X chromosomes. In disorders such as Klinefelter's syndrome, some genes may be expressed from both X chromosomes, resulting in its phenotypic features.

As a rule, monosomic conditions are incompatible with fetal development and, consequently, autosomal monosomies are only rarely identified in spontaneous abortions and are not found among live-born individuals. In fact, the only monosomy compatible with livebirth is the 45,X condition that causes Turner's syndrome. The 45,X chromosome constitution occurs with surprisingly high frequency, present in at least 1–2% of all pregnancies. More than 99% of all 45,X conceptions are spontaneously aborted. Thus, live-born individuals with a 45,X chromosome constitution represent a rare group of survivors. The 45,X phenotype is mild, presumably because the second copy of many X chromosomal genes is normally inactivated. Nonetheless, Turner's syndrome causes gonadal dysgenesis, resulting in infertility and failure to undergo secondary sexual development, along with a number of other phenotypic features (Chap. 349). Several other structural abnormalities of the X chromosome such as deletions, isochromosome X, or ring chromosomes can cause Turner's syndrome. Mosaicism, including 45,X/46,XX, 45,X/47,XXX, 45,X/46,XY, and others, also occurs (see below) and contributes to the phenotypic spectrum in Turner's syndrome.

Because numerical abnormalities originate in meiosis **(Table 62-3)**, affected individuals have missing or extra chromosomes in all cells. In a small proportion of cases, a mitotic nondisjunctional event occurs at an early stage in an individual with an initially normal chromosome constitution. Alternatively, a "normalizing" mitotic nondisjunctional event may result in a normal chromosome complement in some cells of an embryo. In either case, the embryo is a mosaic, with some cells bearing a normal chromosome constitution and others an aneuploid number of chromosomes. The phenotypic consequences are difficult to predict because they depend on the timing of nondisjunction and the distribution of normal and abnormal cells in different tissues. Nevertheless, mosaicism may lead to clinical abnormalities indistinguishable from those of nonmosaic individuals (e.g., nearly 5% of all cases of Down syndrome involve individuals with mosaic trisomy 21, and about 15% of individuals with Turner's syndrome are mosaic for various sex chromosomal constitutions as described above).

Table 62-3 Studies of the Parent and Meiotic/Mitotic Stage of Origin of Human Trisomies and Sex Chromosome Monosomy

	Origin, %				
	Paternal		Maternal		
	I	II	I	II	Mitotic
Trisomy					
2	28	—	54	13	6
7	—	—	17	26	57
15	—	15	76	9	—
16	—	1	96	3	—
18	—	—	33	56	11
21	3	5	67	22	2
22	3	—	94	3	—
XXY	46	—	38	14	3
XXX	—	6	60	16	18
Monosomy					
X*	80		20		

*Results pertain to nonmosaic 45,X individuals.

THE ORIGIN AND ETIOLOGY OF NUMERICAL ABNORMALITIES

Over the past decade, a number of studies have used DNA polymorphisms to investigate the origin of

different types of chromosome abnormalities **(Fig. 62-5)**. The most thoroughly investigated types have been numerical abnormalities **(Table 62-4)**. Sex chromosome monosomy usually results from loss of the paternal sex chromosome, regardless of whether the conception is liveborn or spontaneously aborted.

Figure 62-5

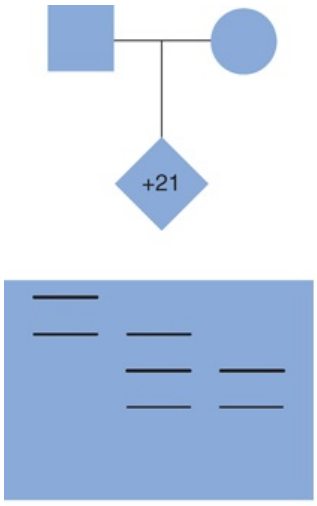

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

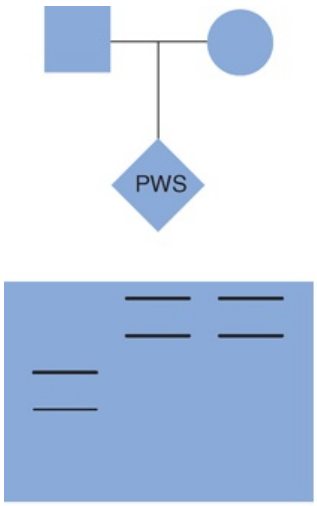

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Use of DNA technology to determine the origin of chromosome abnormalities. ***A.*** Analysis of a chromosome 21–specific DNA polymorphism demonstrates that the trisomic individual received two chromosomes 21 from his mother and one from his father; thus, the extra chromosome 21 resulted from an error in oogenesis. ***B.*** Inheritance of a chromosome 15–specific DNA polymorphism in an individual with Prader-Willi syndrome (PWS). The affected individual has received two maternal, but no paternal, chromosomes 15; thus, the individual is said to have maternal uniparental disomy 15, a common cause of PWS.

Table 62-4 Some Commonly Identified MI-Crodeletion and Microduplication Syndromes—FISH Analysis

Syndrome	Cytogenetic Location	Principal Features	Imprinting Effects
Langer-Giedion syndrome	8q24.1 (del)	Sparse hair, bulbous nose, variable mental retardation	No
WAGR complex	11p13 (del)	Wilms' tumor, aniridia, genitourinary disorders, mental retardation	No
Beckwith-Wiedemann syndrome	11p15 (dup)	Macrosomia, macroglossia, omphalocele	Yes, occasionally associated with "paternal

			uniparental disomy" (see text)
Retinoblastoma	13q14.11 (del)	Retinoblastoma due to homozygous loss of functional RB allele	No obvious effect, although abnormal RB allele more likely to be paternal
Prader-Willi syndrome	15q11-13 (del)	Obesity, hypogonadism, mental retardation	Yes, prototypic imprinting disorder (see text)
Angelman syndrome	15q11-13 (del)	Ataxic gait	With Prader-Willi syndrome, prototypic imprinting disorder (see text)
α-Thalassemia and mental retardation	16p13.3 (del)	α-Thalassemia and mental retardation, due to deletion of distal 16p, including α-globin locus	No
Smith-Magenis syndrome	17p11.2 (del)	Brachycephaly, midface hypoplasia, mental retarda-tion	No
Miller-Dieker syndrome	17p13 (del)	Dysmorphic facies, lissencephaly	No
Charcot-Marie-Tooth syndrome type 1A	17p11.2 (dup)	Progressive neuropathy due to microduplication	No
DiGeorge syndrome/velocardiofacial syndrome	22q11 (del)	Abnormalities of third and fourth branchial arches	No

Trisomies show remarkable variation in parental origin. For example, paternal nondisjunction is responsible for nearly 50% of 47,XXY but only 5–10% of cases of trisomies 13, 14, 15, 21, and 22; it is rarely, if ever, the source of the additional chromosome in trisomy 16. Similarly, there is considerable variability in the meiotic stage of origin. For example, all cases of trisomy 16 may be due to meiosis I errors, whereas for trisomy 21, one-third of cases are associated with meiosis II errors, and for trisomy 18, the majority of cases are apparently due to meiosis II nondisjunction. In spite of this variation in parental and meiotic origin, nondisjunction at maternal meiosis I appears to be the most common source of trisomy.

MATERNAL AGE AND TRISOMY

The association between increasing maternal age and trisomy is the most important etiologic factor in congenital chromosomal disorders. Among women under the age of 25, ~2% of all clinically recognized pregnancies are trisomic; by the age of 36, however, this figure increases to 10% and by the age of 42, to >33% **(Fig. 62-6)**. This association between maternal age and trisomy is exerted without respect to race, geography, or socioeconomic factors and likely affects segregation of all chromosomes.

Figure 62-6

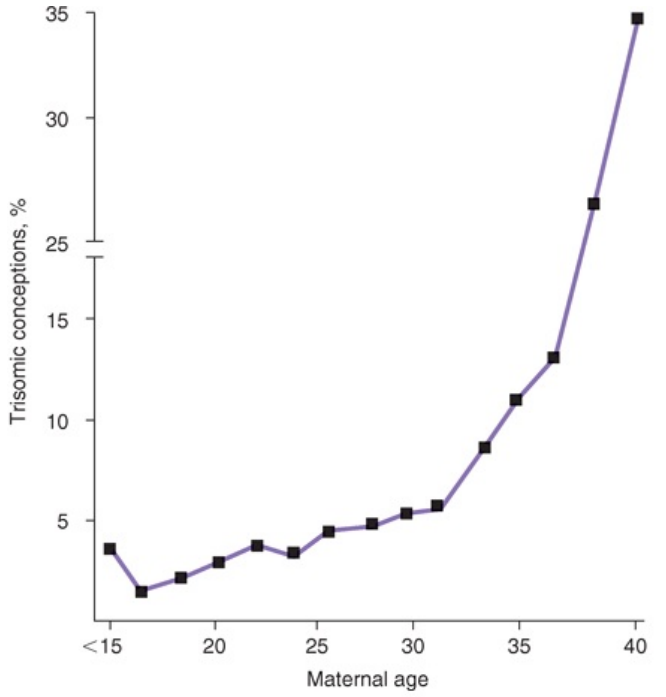

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated maternal age–adjusted rates of trisomy among all clinically recognized pregnancies (e.g., spontaneous abortions, stillbirths, and livebirths). Among women in their forties, more than 25% of all

spontaneous abortion, stillbirths, and livebirths). Among women in their forties, more than 25% of all pregnancies are estimated to involve a trisomic conception; the vast majority of these spontaneously abort, with only trisomies 13, 18, and 21 and sex chromosome trisomies surviving to term with any appreciable frequency.

Despite the importance of increasing age, little is known about the mechanism by which aging leads to abnormal chromosomal segregation. As noted above, it is thought to originate in maternal meiosis I owing to the protracted time to completion (often ≥40 years) in females, and recent studies suggest that it may be associated with alterations in meiotic crossing-over. In trisomy 21, for example, crossover patterns appear to be similarly abnormal in younger and older mothers of trisomic conceptions. Thus, it has been suggested that two distinct steps, or "hits," may be involved in maternal age-related nondisjunction. The first hit, which is age independent, involves the establishment of a "vulnerable" crossover configuration in the fetal oocyte; the second hit, which is age dependent, involves abnormal processing of the vulnerable bivalent structure at metaphase I. If this model is correct, it suggests that the nondisjunctional process is the same in younger and older women, but it occurs more frequently with aging, possibly because of age-dependent degradation of meiotic proteins.

Structural Chromosome Abnormalities

Structural rearrangements involve breakage and reunion of chromosomes. Although less common than numerical abnormalities, they present additional challenges from a genetic counseling standpoint. This is because structural abnormalities, unlike numerical abnormalities, can be present in "balanced" form in clinically normal individuals but transmitted in "unbalanced" form to progeny, thereby resulting in a hereditary form of chromosome abnormality.

Rearrangements may involve exchanges of material between different chromosomes (translocations) or loss, gain, or rearrangements of individual chromosomes (e.g., deletions, duplications, inversions, rings, or isochromosomes). Of particular clinical importance are *translocations* that involve two basic types: Robertsonian and reciprocal. Robertsonian rearrangements are a special class of translocation, in which the long arms of two acrocentric chromosomes (chromosomes 13, 14, 15, 21, and 22) join together, generating a fusion chromosome that contains virtually all of the genetic material of the original two chromosomes. If the Robertsonian translocation is present in unbalanced form, a monosomic or trisomic conception ensues. For example, ~3% of Down syndrome cases are attributable to unbalanced Robertsonian translocations, most often involving chromosomes 14 and 21. In this instance, the affected individual has 46 chromosomes, including one structurally normal chromosome 14, two structurally normal chromosomes 21, and one fusion 14/21 chromosome. This effect leads to a normal diploid dosage for chromosome 14 and to a triplication of chromosome 21, thus resulting in Down syndrome. Similarly, a small proportion of individuals with trisomy 13 syndrome are clinically affected because of an unbalanced Robertsonian translocation involving chromosome 13.

Reciprocal translocations involve mutual exchanges between any two chromosomes. In this circumstance, the phenotypic consequences associated with unbalanced translocations depend on the location of the breakpoints that dictate the amount of material that has been "exchanged" between the two chromosomes. Because most reciprocal translocations involve unique sets of breakpoints, it is difficult to predict the phenotypic consequences in any one situation. In general, severity is determined by the amount of excess or missing chromosome material in individuals with unbalanced translocations.

In addition to rearrangements between chromosomes, there are several examples of intrachromosome structural abnormalities. The most common and deleterious of these involve loss of chromosome material due to deletions. The two best-characterized deletion syndromes, Wolf-Hirschhorn syndrome and cri-du-chat syndrome, result from loss of relatively small chromosomal segments on chromosomes 4p and 5p, respectively. Nonetheless, each is associated with multiple congenital anomalies, developmental delays, profound retardation, and reduced lifespan.

MICRODELETION SYNDROMES—FISH

The term *contiguous gene syndrome* refers to genetic disorders that mimic a combination of single-gene disorders. They result from the deletion of a small number of tightly clustered genes. Because some are too small to be detected cytogenetically, they are termed *microdeletions*. The application of molecular techniques has led to the identification of at least 18 of these microdeletion syndromes (Table 62-4) that have been diagnosed using a directed FISH analysis. Some of the more common ones include the Wilms' tumor–aniridia complex (WAGR), Miller-Dieker syndrome (MDS), and velocardiofacial (VCF) syndrome. WAGR is characterized by mental retardation and involvement of multiple organs, including kidney (Wilms' tumor), eye (aniridia), and the genitourinary system. The cytogenetic abnormality involves a deletion of a part of the short arm of chromosome 11 (11p13), which typically is detectable on well-banded chromosome preparations. In MDS, a disorder characterized by mental retardation, dysmorphic facies, and lissencephaly, the deletion involves chromosome 17 (17p13.3). Using FISH, 17p deletions have been detected in >90% of patients with MDS as well as in 20% of cases of isolated lissencephaly.

Deletions involving the long arm of chromosome 22 (22q11.21) are the most common microdeletions identified to date, present in ~1/3000 newborns. VCF syndrome, the most commonly associated syndrome, consists of learning disabilities or mild mental retardation, palatal defects, a hypoplastic aloe nasi and long nose, and congenital heart defects (conotruncal defect). Some individuals with 22q11.21 deletion are more severely affected and present with DiGeorge syndrome that involves abnormalities in the development of the third and fourth branchial arches leading to thymic hypoplasia, parathyroid

hypoplasia, and conotruncal heart defects. In ~30% of these cases, a deletion at 22q11.21 can be detected with high-resolution banding; by combing conventional cytogenetics, FISH, and molecular detection techniques (i.e., Southern blotting or polymerase chain reaction analyses), these rates improve to >90%. Additional studies have demonstrated a surprisingly high frequency of 22q11.21 deletions in individuals with nonsyndromic conotruncal defects. Approximately 10% of individuals with a 22q11.21 deletion inherited it from a parent with a similar deletion.

Smith-Magenis syndrome involves a microdeletion localized to the proximal region of the short arm of chromosome 17 (17p11.2). Affected individuals have mental retardation, dysmorphic facial features, delayed speech, peripheral neuropathy, and behavior abnormalities. Most of these deletions can be detected with cytogenetic analysis, although FISH is available to confirm these findings. In contrast, William syndrome, a chromosome 7 (7q11.23) microdeletion, cannot be diagnosed with standard or high-resolution analysis; it is only detectable utilizing FISH or other molecular methods. Williams' syndrome involves a deletion of the elastin gene and is characterized by mental retardation, dysmorphic features, a gregarious personality, premature aging, and congenital heart disease (usually supravalvular aortic stenosis).

In addition to microdeletion syndromes, there are several well-described microduplication syndromes, one of which is Charcot-Marie-Tooth type 1A (CMT1A). This is a nerve conduction disease previously thought to be transmitted as a simple autosomal dominant disorder. Recent molecular studies have demonstrated that affected individuals are heterozygous for duplication of a small region of chromosome 17 (17p12). Although it is not yet clear why increased gene dosage would result in CMT1A, the inheritance pattern is explained by the fact that one-half of the offspring of affected individuals inherit the duplication-carrying chromosome.

MICRODELETION SYNDROMES—ARRAY ANALYSIS

All of the above-mentioned microdeletions and microduplications were initially identified clinically because of specific phenotypic features. After these were mapped to specific cytogenetic regions, FISH probes were developed to confirm the clinical diagnosis. With the advent of array analysis, new microdeletion and microduplication syndromes have been identified, often having less specific diagnostic criteria. These microdeletions and microduplications include changes in 16p11.2, 16p13.1, 15q13.3, 1q21.1, and are often ascertained because of autism. Often, the detected genetic changes are familial and parents do not always demonstrate any phenotypic abnormalities. Many of these are believed to be susceptibility genes that might increase an individual's risk for developing a disorder. Other changes such as a microdeletion involving 17q11.21, was not diagnosed before the advent of array analysis, and involves loss of the *MAPT* gene and is seen in patients with mental retardation, dysmorphic features, and hypopigmentation (Fig. 62-3).

Imprinting Disorders

Two microdeletion syndromes, Prader-Willi syndrome (PWS) and Angelman syndrome (AS), exhibit parent-of-origin, or "imprinting," effects. For many years, it has been known that cytogenetically detectable deletions of chromosome 15 occur in a proportion of patients with PWS, as well as in those with AS. This seemed curious, as the clinical manifestations of the two syndromes are very dissimilar, but the deletions appeared identical. PWS is characterized by obesity, hypogonadism, and mild to moderate mental retardation, whereas AS is associated with microcephaly, ataxic gait, seizures, inappropriate laughter, and severe mental retardation. New insight into the pathogenesis of these disorders has been provided by the recognition that parental origin of the deletion determines which phenotype ensues: If the deletion is paternal, the result is PWS, whereas if the deletion is maternal, the result is AS (Fig. 62-5B).

This scenario is complicated further by the recognition that not all individuals with PWS or AS carry the chromosome 15 deletion. For such individuals, the parental origin of the chromosome 15 region is again the important determinant. In PWS, for example, nondeletion patients invariably have two maternal and no paternal chromosomes 15 [maternal uniparental disomy (UPD)], whereas for some nondeletion AS patients the reverse is true (paternal UPD). This indicates that at least some genes on chromosome 15 are differently expressed, depending on which parent contributed the chromosome. Additionally, this means that normal fetal development requires the presence of one maternal and one paternal copy of chromosome 15.

Approximately 70% of PWS cases are due to paternal deletions of 15q11-q13, whereas 25% are due to maternal UPD, and about 5% are caused by mutations in a chromosome 15 imprinting center. In AS, 75% of cases are due to maternal deletions, and only 2% are due to paternal UPD. The remaining cases are presumably caused by imprinting mutations (5%), or mutations in the *UBE3A* gene, which is associated with AS. The UPD cases are mostly caused by meiotic nondisjunction resulting in trisomy 15, subsequently followed by a normalizing mitotic nondisjunction event ("trisomy rescue") resulting in two normal chromosomes 15, both from the same parent. *UBE3A* is the only maternally imprinted gene known in the critical region of chromosome 15. However, several paternally imprinted genes, or expressed-sequence tags (ESTs), have been identified, including *ZNF127*, *IPW*, *SNRPN*, *SNURF*, *PAR1*, and *PAR5*. Recently it has been postulated that a nontranscribed snoRNA (HBII-85), localized to the paternally imprinted region, may be responsible for the PWS phenotype.

Chromosomal regions that behave in the manner observed in PWS and AS are said to be *imprinted*. This phenomenon is involved in differential expression of certain genes on different chromosomes.

Chromosome 11 is one of these with an imprinted region, since it is known that a small proportion of individuals with the Beckwith-Wiedemann overgrowth syndrome have two paternal but no maternal copies of this chromosome.

ACQUIRED CHROMOSOME ABNORMALITIES IN CANCER

In addition to the constitutional cytogenetic chromosomal abnormalities that are present at birth, somatic chromosomal changes can be acquired later in life and are often associated with malignant conditions. As with constitutional abnormalities, somatic changes can include the net loss of chromosomal material (due to a deletion or loss of a chromosome), net gain of material (duplication or gain of a chromosome), and relocation of DNA sequences (translocation). Cytogenetic changes have been particularly well studied in (1) leukemias, e.g., Philadelphia chromosome translocation in CML [t(9;22)(q34.1;q11.2)]; and (2) lymphomas, e.g., translocations of *MYC* in Burkitt's [t(8;14)(q24;q32)]. These and other translocations are useful for diagnosis, classification, and prognosis. Analyses of cytogenetic changes are also useful in certain solid tumors. For example, a complex karyotype with Wilms' tumor, diploidy in medulloblastoma, and Her-2/neu amplification in breast cancer are poor prognostic signs. **For detailed discussion of cancer genetics, see Chap. 83.**

FURTHER READINGS

Bejjani BA, Shaffer LG: Clinical utility of contemporary molecular cytogenetics. Annu Rev Genomics Hum Genet 9:71, 2008[PMID: 18949852] [Full Text]

Ferguson-Smith MA: Cytogenetics and the evolution of cytogenetics. Genet Med 10:553, 2008[PMID: 18641515] [Full Text]

Hassold T, Hunt P: Maternal age and chromosomally abnormal pregnancies: What we know and what we wish we knew. Curr Opin Pediatr 21:703, 2009[PMID: 19881348] [Full Text]

Lee C et al: Multicolor fluorescence in situ hybridization in clinical cytogenetic diagnostics. Curr Opin Pediatr 13:550, 2002

Maya I et al: Diagnostic utility of array-based comparative genomic hybridization (aCGH) in a prenatal setting. Prenat Diagn 30:1131, 2010[PMID: 20925131] [Full Text]

Mefford HC, Eichler EE: Duplication hotspots, rare genomic disorders and common disease. Curr Opin Genet Dev 19:196, 2009[PMID: 19477115] [Full Text]

Miller DT et al: Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 86:749, 2010[PMID: 20466091] [Full Text]

Nasmyth K: Segregating sister genomes: The molecular biology of chromosome separation. Science 297:559, 2002[PMID: 12142526] [Full Text]

Rimoin DL et al (eds): *Emery and Rimoin's Principles and Practice of Medical Genetics* , 5th ed. Philadelphia, Churchill Livingstone, 2007

Sharp AJ et al: Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 38:1038, 2006[PMID: 16906162] [Full Text]

Vorsanova SG et al: Human interphase chromosomes: A review of available molecular cytogenetic technologies. Mol Cytogenet 3:1, 2010[PMID: 20180947] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPLICATIONS OF MOLECULAR GENETICS IN CLINICAL MEDICINE

The field of medical genetics has traditionally focused on chromosomal abnormalities (Chap. 62) and Mendelian disorders (Chap. 61). However, there is genetic susceptibility to many common adult-onset diseases, including atherosclerosis, cardiac disorders, asthma, hypertension, autoimmune diseases, diabetes mellitus, macular degeneration, Alzheimer's disease, psychiatric disorders, and many forms of cancer. Genetic contributions to these common disorders involve more than the ultimate expression of the condition; these genes can also influence the severity of illness, progression of disease, and effect of treatment.

The primary care clinician is now faced with the role of recognizing and counseling patients at risk for a number of genetically influenced diseases. Among the greater than 20,000 genes in the human genome, it is estimated that each of us harbors several potentially deleterious mutations. Fortunately, many of these genetic alterations are recessive or clinically silent. An even greater number, however, represent genetic variants that alter disease susceptibility, course, or response to therapy.

Genetic medicine is changing the way diseases are classified, enhancing our understanding of pathophysiology, providing practical information concerning drug metabolism and therapeutic response, and allowing for individualized screening and health care management programs. In view of these changes, the physician must integrate personal medical history, family history, and diagnostic molecular testing into the overall care of individual patients and their families. Patients turn to their primary care providers for guidance about genetic disorders, even though they may also be seeing other specialists. The primary care provider has an important role in educating patients about the indications, benefits, risks, and limitations of genetic testing in the management of a number of diverse diseases. This is a difficult task, because scientific advances in genetic medicine are outpacing the translation of these discoveries into standards of clinical care.

COMMON ADULT-ONSET GENETIC DISORDERS

Multifactorial Inheritance

The risk for many adult-onset disorders reflects the combined effects of genetic factors at multiple loci that may function independently or in combination with other genes or environmental factors. Our understanding of the genetic basis of these disorders is incomplete, despite the clear recognition of genetic susceptibility. In Type 2 diabetes mellitus, for example, the concordance rate in monozygotic twins ranges between 50 and 90%. Diabetes or impaired glucose tolerance occurs in 40% of siblings and in 30% of the offspring of an affected individual. Despite the fact that diabetes affects 5% of the population and exhibits a high degree of heritability, only a few genetic mutations (most of which are rare) that might account for the familial nature of the disease have been identified. They include certain mitochondrial DNA disorders (Chap. 61), mutations in a cascade of genes that control pancreatic islet cell development and function (*HNF4α*, *HNF1α*, *IPF1*, *TCF7L2, glucokinase*), insulin receptor mutations, and others (Chap. 344). In addition to these known genes, a large number of additional genetic loci that confer disease susceptibility have been identified. Superimposed on this genetic background are environmental or medical influences such as diet, exercise, pregnancy, and medications.

Identifying susceptibility genes associated with multifactorial adult-onset disorders is a formidable task. Nonetheless, a reasonable goal for these types of diseases is to identify genes that increase (or decrease) disease risk by a factor of two or more. For common diseases such as diabetes or heart disease, this level of risk has important implications for health. In much the same way that cholesterol is currently used as a biochemical marker of cardiovascular risk, we can anticipate the development of genetic panels with similar predictive power. The availability of DNA-microarray systems represents an important technology that makes large-scale testing feasible (Chap. 61). Whether to perform a genetic test for a particular inherited adult-onset disorder, such as hemochromatosis, multiple endocrine neoplasia (MEN) type 1, prolonged QT syndrome, or Huntington's disease, is a complex decision; it depends on the clinical features of the disorder, the desires of the patient and family, and whether the results of genetic testing will alter medical decision-making or treatment (see below).

POPULATION SCREENING

Mass genetic screening programs require tests with high enough sensitivity and specificity to be cost-effective. An effective screening program should fulfill the following criteria: the tested disorder is

prevalent and serious; it can be influenced presymptomatically through lifestyle changes, screening, medications, or other risk-reducing interventions; and identification of risk does not result in undue discrimination or harm. Screening individuals of Jewish descent for the autosomal recessive neurodegenerative disorder Tay-Sachs disease has resulted in a dramatic decline in the incidence of this syndrome in the United States. On the other hand, screening for sickle cell disease or trait in the African-American population has sometimes resulted in insurance and employment discrimination.

Mass screening for complex genetic disorders can result in potential problems. For example, cystic fibrosis is most commonly associated with the ΔF508 mutation. This variant accounts for 30–80% of mutant alleles, depending on the ethnic group. Nevertheless, cystic fibrosis is associated with pronounced genetic heterogeneity with more than 1000 disease-related mutations. The American College of Medical Genetics recommends a panel of 23 alleles, including the ΔF508 allele, for routine carrier testing. Analysis for the less common cystic fibrosis–associated mutations would greatly impact the cost of testing without significantly influencing the effectiveness of mass screening. Nevertheless, the individual who carries one of the less common cystic fibrosis–associated alterations will not benefit if testing is limited to a routine panel.

Occupational health screening programs hold promise but also raise concerns about employment discrimination. These concerns were brought to light when it was discovered that a railroad company was testing its employees, without consent, for a rare genetic condition that results in susceptibility to carpal tunnel syndrome. The Equal Employment Opportunity Commission argued that the tests were unlawful under the Americans with Disabilities Act.

The Family History

When two or more first-degree relatives are affected with asthma, cardiovascular disease, Type 2 diabetes, breast cancer, colon cancer, or melanoma, the relative risk ranges from two- to fivefold, underscoring the importance of family history for these prevalent disorders. Pending further advances in genetic testing, in most circumstances, the key to assessing the inherited risk for common adult-onset diseases rests in the collection and interpretation of a detailed personal and family medical history in conjunction with a directed physical examination. For example, a history of multiple family members with early-onset coronary artery disease, glucose intolerance, and hypertension should suggest increased risk for genetic, and perhaps environmental, predisposition to metabolic syndrome (Chap. 242). Individual patients with this family history should be monitored for the possible development of high blood pressure, diabetes, and hyperlipidemia. They should be counseled about the importance of avoiding additional risk factors such as obesity, physical inactivity, and cigarette smoking.

Family history should be recorded in the form of a pedigree. At a minimum, pedigrees should convey health-related data on all first-degree relatives and selected second-degree relatives, including grandparents. When pedigrees appear to suggest an inherited disease, they should be extended to include additional family members. The determination of risk for an asymptomatic individual will vary depending on the size of the pedigree, the number of unaffected relatives, and the types of diagnoses, as well as the age of disease onset. For example, a woman with two first-degree relatives with breast cancer is at greater risk for a Mendelian disorder if she has a total of three female first-degree relatives than if she has a total of ten female first-degree relatives. Additional variables that should be documented in the pedigree include the presence or absence of nonhereditary risk factors among those affected with diseases and the finding of multiple diseases in an individual patient. For instance, a woman with a history of both colon cancer and endometrial cancer is at risk for Lynch syndrome regardless of her family history.

When assessing the personal and family history, the physician should be alert to a younger age of disease onset than is usually seen in the general population. A 30-year-old with acute myocardial infarction should be considered at risk for a hereditary trait, even if there is no family history of premature coronary artery disease (Chap. 241). The absence of the nonhereditary risk factors typically associated with a disease also raises the prospect of genetic causation. A personal or family history of deep vein thrombosis, in the absence of known environmental or medical risk factors, suggests a hereditary thrombotic disorder (Chap. 117). The physical examination also may provide important clues about the risk for a specific inherited disorder. A patient presenting with xanthomas at a young age should prompt consideration of familial hypercholesterolemia. Some adult-onset disease-causing mutations are more prevalent in certain ethnic groups. For instance, >2% of the Ashkenazi population carry one of three specific mutations in the *BRCA1* or *BRCA2* genes. The prevalence of the factor V Leiden allele ranges from 3 to 7% in Caucasians but is much lower in Africans or Asians.

Recall of family history is often inaccurate. This is especially so when the history is remote and families become more dispersed geographically. It can be helpful to ask patients to fill out family history forms before or after their visits, as this provides them with an opportunity to contact relatives. Ideally, this information should be embedded in electronic health records and updated intermittently. Attempts should be made to confirm the illnesses reported in the family history before making important and, in certain circumstances, irreversible management decisions. This process is often labor intensive and ideally involves interviews of additional family members or reviewing medical records, autopsy reports,

and death certificates.

Although many inherited disorders will be suggested by the clustering of relatives with the same or related conditions, it is important to note that *disease penetrance* is incomplete for most multifactorial genetic disorders. As a result, the pedigree obtained in such families may not exhibit a clear Mendelian inheritance pattern, as not all family members carrying the disease-associated alleles will manifest a clinical disorder. Furthermore, genes associated with some of these disorders often exhibit *variable expression* of disease. For example, the breast cancer–associated gene *BRCA1* can predispose to several different malignancies in the same family, including cancers of the breast, ovary, and prostate (Chap. 83). For common diseases such as breast cancer, some family members without the disease-causing mutation may also develop breast cancer, representing another confounding variable in the pedigree analysis.

Some of the aforementioned features of the family history are illustrated in **Fig. 63-1**. In this example, the proband, a 36-year-old woman (IV-1), has a strong history of breast and ovarian cancer on the paternal side of her family. The early age of onset, as well as the co-occurrence of breast and ovarian cancer in this family, suggests the possibility of an inherited mutation in *BRCA1* or *BRCA2*. It is unclear though—without genetic testing—whether her father harbors such a mutation and transmitted it to her. After appropriate genetic counseling of the proband and her family, the most informative approach to DNA analysis in this family is to test the cancer-affected 42-year-old living cousin for the presence of a *BRCA1* or *BRCA2* mutation. If a mutation is found, then it is possible to test for this particular alteration in other family members, if they so desire. In the example shown, if the proband's father has the *BRCA1* mutation, there is a 50:50 probability that the mutation was transmitted to her, and genetic testing can be used to establish the absence or presence of this alteration. In this same example, if a mutation is not detected in the cancer-affected cousin, testing would not be indicated for cancer-unaffected relatives.

Figure 63-1

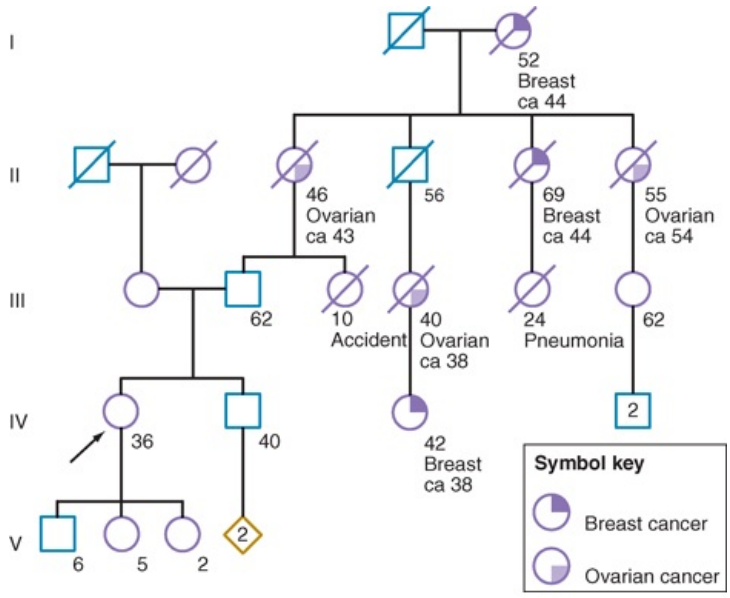

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A 36-year-old woman (*arrow*) seeks consultation because of her family history of cancer. The patient expresses concern that the multiple cancers in her relatives imply an inherited predisposition to develop cancer. The family history is recorded and records of the patient's relatives confirm the reported diagnoses.

GENETIC TESTING FOR ADULT-ONSET DISORDERS

A critical first step before initiating genetic testing is to ensure that the correct clinical diagnosis has been made, whether it is based on family history, characteristic physical findings, or biochemical testing. Careful clinical assessment can define the *phenotype*, thereby preventing unnecessary testing and directing testing toward the most probable candidate genes **(Fig. 63-2)**. For patients identified by population-based screening (e.g., diabetes, hypercholesterolemia), testing might involve known candidate genes, or genome-wide linkage studies of the population could be used as part of a research study to identify susceptibility alleles. For patients with a strong family history (e.g., breast cancer, hemochromatosis), testing often includes known candidate genes, or traditional linkage analyses within pedigrees can identify candidate disease-causing genes. Once disease-related genes

are known, mutational analyses can be performed after pretest genetic counseling (see below).

Figure 63-2

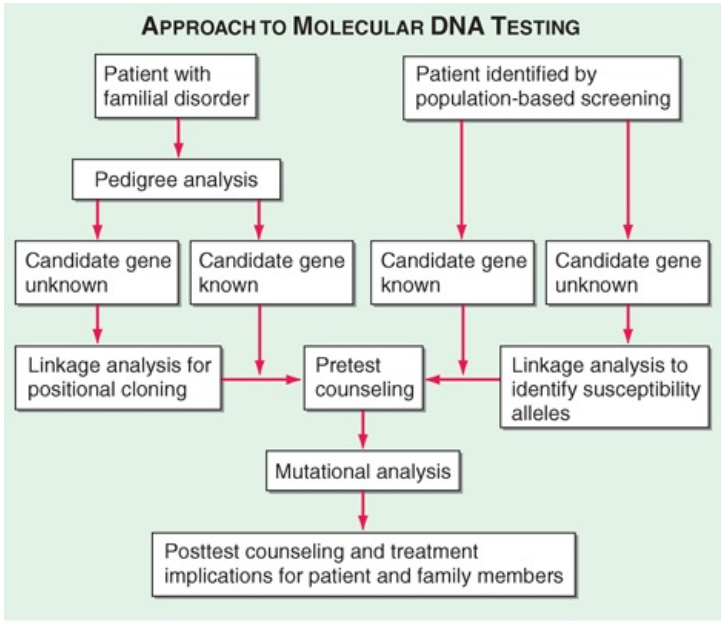

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to identifying a disease-causing gene.

Many disorders exhibit the feature of *locus heterogeneity*, which refers to the fact that mutations in different genes can cause phenotypically similar disorders. For example, osteogenesis imperfecta (Chap. 363), long QT syndrome (Chap. 233), muscular dystrophy (Chap. 387), homocystinuria (Chap. 364), retinitis pigmentosa (Chap. 28), and hereditary predisposition to colon cancer (Chap. 91) or breast cancer (Chap. 90) can each be caused by mutations in distinct genes. The pattern of disease transmission, disease risk, clinical course, and treatment may differ significantly, depending on the specific gene affected. In these cases, the choice of which genes to test is often determined by unique clinical and family history features, the relative prevalence of mutations in various genes, or test availability.

Methodologic Approaches to Genetic Testing

Genetic testing is performed in much the same way as other specialized laboratory tests. In the United States, genetic testing laboratories are Clinical Laboratory Improvement Act (CLIA) approved to ensure that they meet quality and proficiency standards. A useful information source for various genetic tests is *www.genetests.org*.

DNA testing is most commonly performed by DNA sequence analysis for mutations, although genotype can also be deduced through the study of RNA or protein (e.g., apoprotein E, hemoglobin, immunohistochemistry). For example, immunohistochemical analysis of colorectal cancers for absence of expression of mismatch repair proteins has been proposed as a strategy for universal Lynch syndrome screening. The determination of DNA sequence alterations relies heavily on the use of polymerase chain reaction (PCR), which allows rapid amplification and analysis of the gene of interest. In addition, PCR enables genetic testing on minimal amounts of DNA extracted from a wide range of tissue sources including leukocytes, mucosal epithelial cells, and archival tissues. Amplified DNA can be analyzed directly by DNA sequencing or it can be hybridized to DNA chips or blots to detect the presence of normal and altered DNA sequences. Direct DNA sequencing is increasingly used for prenatal diagnosis as well as for determination of hereditary disease susceptibility. Analyses of large alterations in the genome are possible using cytogenetics, fluorescent in situ hybridization (FISH), or Southern blotting (Chap. 62).

Protein truncation tests (PTTs) are used to detect mutations that result in the premature termination of a polypeptide occurring during protein synthesis. In this assay, the isolated complementary DNA (cDNA) is transcribed and translated in vitro, and the protein is analyzed by gel electrophoresis. The truncated (mutant) gene product is readily identified as its electrophoretic mobility differs from that of the normal protein. This test has been used most commonly for analyses of large genes with significant genetic heterogeneity such as the *APC* gene.

Like all laboratory analyses, there are limitations to the accuracy and interpretation of genetic tests. In

addition to technical errors, genetic tests are sometimes designed to detect only the most common mutations. In this case, a negative result must be qualified by the possibility that the individual may have a mutation that is not included in the test. In addition, a negative result does not mean that there is not a mutation in some other gene that causes a similar inherited disorder. A negative result, in those at risk for the disorder, is typically classified as *uninformative*.

In addition to molecular testing for established disease, genetic testing for susceptibility to chronic disease is being increasingly integrated into the practice of medicine. In most cases, however, the discovery of disease-associated genes has greatly outpaced studies that assess the clinical utility of genetic testing. Until such evidence-based studies are available, predictive molecular testing must be approached with caution and should be offered only to patients who have been adequately counseled and have provided informed consent. In the majority of cases, genetic testing should be offered only to individuals with a suggestive personal or family medical history or in the context of a clinical trial.

Predictive genetic testing falls into two distinct categories. *Presymptomatic testing* applies to diseases where a specific genetic alteration is associated with a near 100% likelihood of developing disease. In contrast, *predisposition testing* predicts a risk for disease that is less than 100%. For example, presymptomatic testing is available for those at risk for Huntington's disease, whereas predisposition testing is considered for those at risk for hereditary breast cancer. It is important to note that, for the majority of adult-onset, multifactorial genetic disorders, testing is only predictive. Test results cannot reveal with confidence whether, when, or how the disease will manifest itself. For example, not everyone with the apolipoprotein E allele (ε4) will develop Alzheimer's disease, and individuals without this genetic marker can still develop the disorder (Chap. 371).

Molecular analysis is generally more informative if testing is initiated in a symptomatic family member because the identification of a mutation can direct the testing of other at-risk family members (whether they are symptomatic or not). In the absence of additional familial or environmental risk factors, individuals who test negative for the mutation found in the affected family member can be informed that they are at general population risk for that particular disease. Furthermore, they can be reassured that they are not at risk for passing the mutation on to their children. On the other hand, asymptomatic family members who test positive for the known mutation must be informed that they are at increased risk for disease development and for transmitting the alteration to their children.

Clinicians providing pretest counseling and education should assess the patient's ability to understand and cope with test results. Individuals who demonstrate signs and symptoms of emotional distress should have their psychosocial needs addressed before proceeding with molecular testing. Generally, genetic testing should not be offered at a time of personal crisis or acute illness within the family. Patients will derive more benefit from test results if they are emotionally able to comprehend and absorb the information. It is important to assess a patient's preconceived notions of their personal likelihood of disease in preparing pretest educational strategies. Often, patients harbor unwarranted fear or denial of their likelihood of genetic risk.

Genetic testing has the potential of affecting the way individual family members relate to one another, both negatively and positively. As a result, patients addressing the option of molecular testing must consider how test results might impact their relationships with relatives, partners, spouses, and friends. In families with a known genetic mutation, those who test positive must consider the impact of their carrier status on their present and future lifestyles; those who test negative may manifest *survivor guilt*. Family members are likely to differ in their emotional and social responses to the same information. Counseling should also address the potential consequences of test results on relationships with a spouse or child. Parents who are found to have a disease-associated mutation often express considerable anxiety and despair as they address the issue of risk to their children.

When a condition does not manifest until adulthood, clinicians will be faced with the question of whether at-risk children should be offered molecular testing and, if so, at what age. Although the matter is debated, several professional organizations have cautioned that genetic testing for adult-onset disorders should not be offered to children. Many of these conditions are not preventable; consequently, such information can pose significant psychosocial risk to the child. In addition, there is concern that testing during childhood violates a child's right to make an informed decision regarding testing upon reaching adulthood. On the other hand, testing should be offered in childhood for disorders that may manifest early in life, especially when management options are available. For example, children at risk for familial adenomatous polyposis (FAP), associated with alterations in the *APC* gene, may develop polyps as early as their teens, and progression to an invasive cancer can occur by their twenties. Likewise, children at risk for MEN type 2, which is caused by mutations in the *RET* protooncogene, may develop medullary thyroid cancer early in childhood, and the issue of prophylactic thyroidectomy should be addressed with the parents of children with documented mutations (Chap. 351).

Informed Consent

When the issue of testing is addressed, patients should be strongly encouraged to involve other relatives in the decision-making process, because molecular diagnostics will likely have an impact on

the entire family. Informed consent for molecular testing begins with detailed education and counseling **(Fig. 63-3)**. The patient must fully understand the risks, benefits, and limitations of undergoing the analysis. Informed consent should include a written document, drafted clearly and concisely in a language and format that is comprehensible to the patient, who should be made aware of the disposition of test results. Informed consent should also include a discussion of the mechanics of testing. Most molecular testing for hereditary disease involves DNA-based analysis of peripheral blood. In the majority of circumstances, test results should be given only to the individual, in person, and preferably with a support person in the room.

Figure 63-3

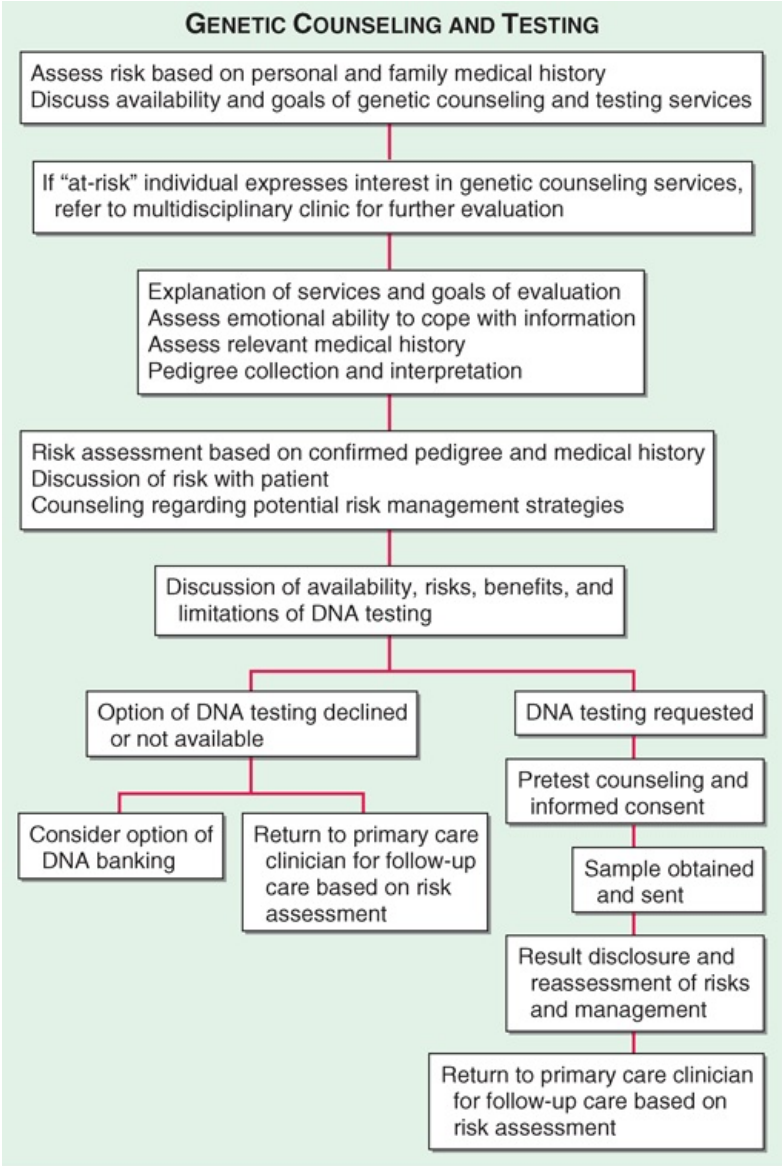

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for genetic counseling in association with genetic testing.

Because molecular testing of an asymptomatic individual often allows prediction of future risk, the patient should understand any potential long-term medical, psychological, and social implications of this decision. In the United States, legislation affecting health insurance genetic discrimination has evolved through the Genetics Information Nondiscrimination Act. It is important to explore with the patient the potential impact that test results may have on employment and future health as well as disability and life insurance coverage.

Patients should understand that alternatives to molecular analysis remain available if they decide not to proceed with this option. They should also be notified that testing is available in the future if they are not currently prepared to undergo analysis. The option of DNA banking should be presented so that samples are readily available for future use by family members, if needed.

Follow-Up Care after Testing

Depending on the nature of the genetic disorder, posttest interventions may include (1) cautious surveillance and appropriate health care screening, (2) specific medical interventions, (3) chemoprevention, (4) risk avoidance, and (5) referral to support services. For example, patients with known pathologic mutations in *BRCA1* or *BRCA2* are strongly encouraged to pursue risk-reducing bilateral salpingo-oophorectomy and are offered intensive screening as well as the option of risk-reducing mastectomy. In addition, such women may be eligible for preventive treatment with tamoxifen or enrollment in a chemoprevention clinical trial. In contrast, those at known risk for Huntington's disease are offered continued follow-up and supportive services, including physical and occupational therapy, and social services or support groups, as indicated. Specific interventions will change as translational research continues to enhance our understanding of these genetic diseases and as more is learned about the functions of the gene products involved.

Individuals who test negative for a mutation in a disease-associated gene identified in an affected family member must be reminded that they may still be at risk for the disease. This is of particular importance for common diseases such as diabetes mellitus, cancer, and coronary artery disease. For example, a woman who finds that she does not carry the disease-associated mutation in *BRCA2* previously discovered in her family must be reminded that she still requires the same breast cancer screening recommended for the general population.

GENETIC COUNSELING AND EDUCATION

Genetic counseling should be distinguished from genetic testing and screening, even though genetic counselors are often involved in issues related to testing. Genetic counseling refers to *a communication process that deals with human problems associated with the occurrence or risk of a genetic disorder in a family*. Genetic risk assessment is complex and often involves elements of uncertainty. Counseling, therefore, includes genetic education as well as psychosocial counseling. Genetic counselors may be called upon by other health care professionals (or by individual patients and families) to address a broad range of issues directly and indirectly related to genetic disease **(Table 63-1)**. The role of the genetic counselor includes the following:

- Gather and document a detailed family history
- Educate patients about general genetic principles related to disease risk, both for themselves and for others in their family
- Assess and enhance the patient's ability to cope with the genetic information offered
- Discuss how nongenetic factors may relate to the ultimate expression of disease
- Address medical management issues
- Assist in determining the role of genetic testing for the individual and family
- Ensure that the patient is aware of the indications, process, risks, benefits, and limitations of the various genetic testing options
- Assist the patient, family, and referring physician in the interpretation of the test results
- Refer the patient and other at-risk family members for additional medical and support services, if necessary

Table 63-1 Indications for Genetic Counseling

Advanced maternal (>35) or paternal (>50) age
Consanguinity
Previous history of a child with birth defects or a genetic disorder
Personal or family history suggestive of a genetic disorder
High-risk ethnic groups; known carriers of genetic alterations
Documented genetic alteration in a family member
Ultrasound or prenatal testing suggesting a genetic disorder

The complexity of genetic counseling and the broad scope of genetic diseases have led to the development of specialized, multidisciplinary clinics designed to provide broad-based support and medical care for those at risk and their family members. Such specialty clinics are well established in the areas of cancer and neurodegenerative disorders and are developing in other areas such as cardiology. The multidisciplinary teams are often composed of medical geneticists, specialist physicians, genetic counselors, nurses, psychologists, social workers, and biomedical ethicists who work together to consider difficult diagnostic, treatment, and testing decisions. Such a format also provides primary care physicians with invaluable support and assistance as they follow and treat at-risk patients.

The approach to genetic counseling has important ethical, social, and financial implications. Philosophies related to genetic counseling vary widely by country and center. Counseling is generally offered in a nondirective manner, wherein patients learn to understand how their values factor into a particular medical decision. Nondirective counseling is particularly appropriate when there are no data demonstrating a clear benefit associated with a particular intervention or when an intervention is considered experimental. For example, nondirective genetic counseling is employed when a person is deciding whether to undergo genetic testing for Huntington's disease (Chap. 371). At this time, there is no clear benefit (in terms of medical outcome) to an at-risk individual undergoing genetic testing for this disease, because its course cannot be altered by therapeutic interventions. However, testing can have an important impact on this individual's perception of the future and his or her interpersonal relationships and plans for reproduction. Therefore, the decision to pursue testing rests on the individual's belief system and values. On the other hand, a more directive approach is appropriate when a condition can be treated. In a family with FAP, colon cancer screening and prophylactic colectomy should be recommended for known *APC* mutation carriers. The counselor and clinician following this family must ensure that the at-risk family members have access to the resources necessary to adhere to these recommendations.

Genetic education is central to an individual's ability to make an informed decision regarding testing options and treatment. Although genetic counselors represent one source of genetic education, other health care providers also need to contribute to patient education. Patients at risk for genetic disease should understand fundamental medical genetic principles and terminology relevant to their situation. This includes the concept of genes, how they are transmitted, and how they confer hereditary disease risk. An adequate knowledge of patterns of inheritance will allow patients to understand the probability of disease risk for themselves and other family members. It is also important to impart the concepts of disease penetrance and expression. For most complex adult-onset genetic disorders, asymptomatic patients should be advised that a positive test result does not always translate into future disease development. In addition, the role of nongenetic factors, such as environmental exposures, must be discussed in the context of multifactorial disease risk and disease prevention. Finally, patients should understand the natural history of the disease as well as the potential options for intervention, including screening, prevention, and—in certain circumstances—pharmacologic treatment or prophylactic surgery.

THERAPEUTIC INTERVENTIONS BASED ON GENETIC RISK FOR DISEASE

Specific treatments are now available for an increasing number of genetic disorders, whether identified through population-based screening or directed testing **(Table 63-2)**. Although the strategies for therapeutic interventions are best developed for childhood hereditary metabolic diseases, these principles have made their way into the diagnosis and management of adult-onset disorders. Hereditary hemochromatosis illustrates many of the issues raised by the availability of genetic screening in the adult population. For instance, hemochromatosis is relatively common (approximately 1 in 200 individuals of northern European descent are homozygous), and its complications are potentially preventable through phlebotomy (Chap. 357). The identification of the *HFE* gene, mutations of which are associated with this syndrome, has sparked interest in the use of DNA-based testing for presymptomatic diagnosis of the disorder. However, up to one-third of individuals who are homozygous for the *HFE* mutation do not have evidence of iron overload. Consequently, in the absence of a positive family history, current recommendations include phenotypic screening for evidence of iron overload followed by genetic testing. Whether genetic screening for hemochromatosis will someday be coupled to assessment of phenotypic expression awaits further studies. In contrast to the issue of population screening, it is important to test and counsel other family members when the diagnosis of hemochromatosis has been made in a proband. Testing allows the physician to exclude family members who are not at risk. It also permits presymptomatic detection of iron overload and the institution of treatment (phlebotomy) before the development of organ damage.

Table 63-2 Examples of Genetic Testing and Possible Interventions

Genetic Disorder	Inheritance	Genes	Interventions
Oncologic			
Lynch syndrome	AD	*MSH2, MLH1, MSH6, PMS2,*	Early endoscopic screening
Familial adenomatous polyposis	AD	*APC*	Early endoscopic screening
			Nonsteroidal anti-inflammatory drugs
Hereditary breast and ovarian cancer	AD	*BRCA1, BRCA2*	Colectomy Estrogen receptor antagonists Early screening by exams, mammography, and breast MRI

			Risk-reducing surgery
Familial malignant melanoma	AD	*CDKN2A, CDK4*	Avoidance of UV light Screening and biopsies
Basal cell nevus syndrome	AD	*PTCH*	Avoidance of UV light Screening and biopsies
Hematologic			
Factor V Leiden	AD	*F5*	Avoidance of thrombogenic risk factors and oral contraceptives
Hemophilia A	XL	*F8*	Factor VIII replacement
Hemophilia B	XL	*F9*	Factor IX replacement
Glucose 6-PO_4dehydrogenase deficiency	XL	*G6PD*	Avoidance of oxidant drugs
Cardiovascular			
Hypertrophic cardiomyopathy	AD	*MYH7, MYBPC3, TNNT2, TPM1*	Echocardiographic screening Early pharmacologic intervention Myomectomy
Long QT syndrome	AD, AR	*KCNQ1, SCN5A, KCNE1, KCNE2*	Electrocardiographic screening Early pharmacologic intervention Implantable cardioverter defibrillator devices
Marfan syndrome	AD	*FBN1*	Echocardiographic screening Prophylactic beta blockers or ACE inhibitors
Gastrointestinal			
Familial Mediterranean fever	AR	*MEFV*	Colchicine treatment
Hemochromatosis	AR	*HFE*	Phlebotomy
Pulmonary			
α_1 Antitrypsin deficiency	AR	*SERPINA1*	Avoidance of smoking Avoidance of occupational and environmental toxins
Primary pulmonary hypertension	AD	*BMPR2*	Pharmacologic intervention Lung transplantation
Renal			
Polycystic kidney disease	AD	*PKD1, PKD2*	Prevention of hypertension Prevention of urinary tract infections Kidney transplantation
Nephrogenic diabetes insipidus	XL, AR	*AVPR2, AQP2*	Fluid replacement Thiazides with or without amiloride
Endocrine			
Neurohypophyseal diabetes insipidus	AD	*AVP*	Replace vasopressin
Maturity-onset diabetes of the young	AD	Multiple genes	Screen and treat for diabetes
Familial hypocalciuric hypercalcemia	AD	*CASR*	Avoidance of parathyroidectomy
Kallmann syndrome	XL	*KAL*	Induction of puberty with hormone replacement
Multiple endocrine neoplasia type 2	AD	*RET*	Prophylactic thyroidectomy Screening for pheochromocytoma and

			hyperparathyroidism
21-hydroxylase deficiency	AR	*CYP21*	Glucocorticoid and mineralocorticoid treatment
Neurologic			
Malignant hyperthermia	AD	*RYR1, CACNA1S*	Avoidance of precipitating anesthetics
Hyperkalemic periodic paralysis	AD	*SCN4A*	Diet rich in carbohydrates and low in potassium Avoidance of fasting, strenuous work, or cold temperatures Thiazides or acetazolamide
Adrenoleukodystrophy	XL	*ABCD1*	Possible bone marrow transplantation for childhood cerebral form
Duchenne and Becker muscular dystrophy	XL	*DMD*	Corticosteroids
Familial Parkinson disease	AD, AR	*SNCA, PARK2, PINK1, PARK7, LRRK2*	Amantadine, anticholinergics, levodopa, monoamine oxidase B inhibitors
Wilson disease	AR	*ATP7B*	Zinc, trientene

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; CNS, central nervous system; MRI, magnetic resonance imaging; XL, X-linked.

Preventive measures and therapeutic interventions are not restricted to metabolic disorders. Identification of familial forms of long QT syndrome, associated with ventricular arrhythmias, allows early electrocardiographic testing and the use of prophylactic antiarrhythmic therapy, overdrive pacemakers, or defibrillators (Chap. 233). Individuals with familial hypertrophic cardiomyopathy can be screened by ultrasound, treated with beta blockers or other drugs, and counseled about the importance of avoiding strenuous exercise and dehydration (Chap. 238). Likewise, individuals with Marfan syndrome can be treated with beta blockers or ACE inhibitors and monitored for the development of aortic aneurysms (Chap. 248). Individuals with α_1 antitrypsin deficiency can be strongly counseled to avoid cigarette smoking and exposure to environmental pulmonary and hepatotoxins. Various host genes influence the pathogenesis of certain infectious diseases in humans, including HIV (Chap. 189). The factor V Leiden allele increases risk of thrombosis (Chap. 58). Approximately 3% of the worldwide population is heterozygous for this mutation. Moreover, it is found in up to 25% of patients with recurrent deep vein thrombosis or pulmonary embolism. Women who are heterozygous or homozygous for this allele should, therefore, avoid the use of oral contraceptives.

The field of pharmacogenomics seeks to identify genes that alter drug metabolism or confer susceptibility to toxic drug reactions. Pharmacogenomics permits individualized drug therapy, resulting in improved treatment outcomes, reduced toxicities, and more cost-effective pharmaceutical care. Examples include succinylcholine sensitivity, thiopurine methyltransferase (TPMT) deficiency, malignant hyperthermia, dihydropyrimidine dehydrogenase deficiency, the porphyrias, and glucose-6-phosphate dehydrogenase (G6PD) deficiency.

As noted above, the identification of genes that increase the risk of specific types of neoplasia is rapidly changing the management of many cancers. Identifying family members with mutations that predispose to FAP or Lynch syndrome can lead to recommendations of early cancer screening or prophylactic surgery (Chap. 91). Similar principles apply to familial forms of melanoma; basal cell carcinoma; and cancers of the breast, ovary, and thyroid gland. It should be recognized, however, that most cancers harbor several distinct genetic abnormalities by the time they acquire invasive or metastatic potential (Chaps. 83 and 84). Consequently, the major impact of genetic testing in these cases is to allow more intensive management, including disease prevention and screening, as it remains very challenging to predict disease penetrance, expression, or clinical course.

Although genetic diagnosis of these and other disorders is still evolving in the clinical setting, predictive testing holds the promise of allowing earlier and more targeted interventions that can reduce morbidity and mortality rates. We can expect the availability of genetic tests to expand, including direct-to-consumer options. A critical challenge for physicians and other health care providers is to keep pace with these advances in genetic medicine and to implement testing judiciously.

FURTHER READINGS

Brand A et al: The impact of genetics and genomics on public health. Eur J Hum Genet 16:5, 2008[PMID: 17957226] [Full Text]

Clayton EW: Ethical, legal, and social implications of genomic medicine. N Engl J Med 349:562,

2003[PMID: 12904522] [Full Text]

Feero WG et al: Genomic medicine–an updated primer. N Engl J Med 362:2001, 2010[PMID: 20505179] [Full Text]

Green NS et al: Newborn screening: Complexities in universal genetic testing. Am J Public Health 96 1955, 2006

Harper PS: *Practical Genetic Counseling*, 5th ed. Oxford, Butterworth Heinmann, 1998

Lumbreras B, et al: Assessing the social meaning, value and implications of research in genomics. J Epidemiol Community Health 61:755, 2007[PMID: 17699526] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MITOCHONDRIAL DNA AND HERITABLE TRAITS AND DISEASES: INTRODUCTION

Mitochondria are cytoplasmic organelles whose major function is to generate ATP by the process of oxidative phosphorylation in aerobic conditions. This process is mediated by the respiratory electron transport chain (ETC) multiprotein enzyme complexes I–V and the two electron carriers, coenzyme Q (CoQ) and cytochrome c. Other cellular processes to which mitochondria make a major contribution include apoptosis (programmed cell death), along with additional cell-type specific functions **(Table e18-1)**. The efficiency of the mitochondrial ETC in ATP production is a major determinant of overall body energy balance and thermogenesis. In addition, mitochondria are the predominant source for generating reactive oxygen species (ROS), whose rate of production also relates to the coupling of ATP production to oxygen consumption. In light of the centrality of oxidative phosphorylation to the normal activities of almost all cells, it is not surprising that mitochondrial dysfunction can affect almost any organ system **(Fig. e18-1)**. Thus, physicians in many specialties may encounter patients with mitochondrial diseases and should be aware of the existence and characteristics of those diseases.

Table e18-1 Functions of Mitochondria

All Cells and Tissues
Oxidative phosphorylation
Apoptosis (programmed cell death)
Tissue- or Cell-Specific
Cholesterol metabolism
Amino and organic acid metabolism
Fatty acid beta oxidation
Sex steroid synthesis
Heme synthesis
Hepatic ammonia detoxification
Neurotransmitter metabolism

Figure e18-1

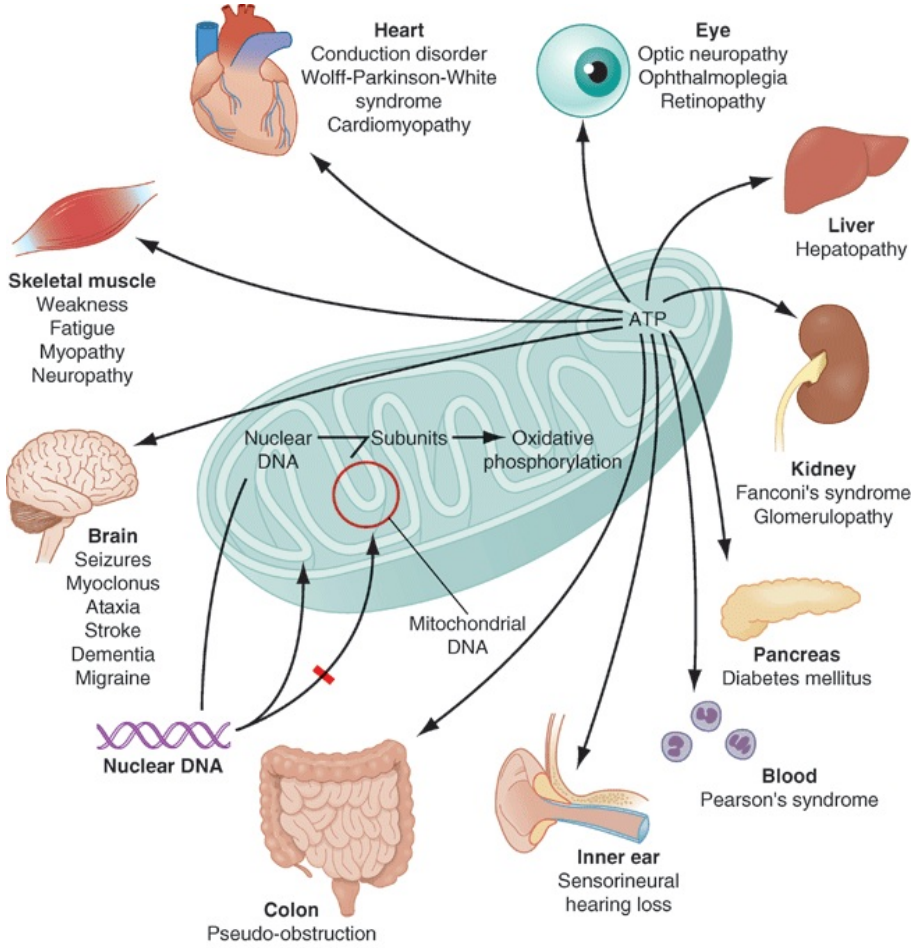

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dual genetic control and multiple organ system manifestations of mitochondrial disease. *(Reproduced with permission from DR Johns: N Engl J Med 333:638, 1995.)*

The integrated activity of an estimated 1500 gene products is required for normal mitochondrial biogenesis, function, and integrity. Most of these products are encoded by nuclear genes and thus follow the rules and patterns of nuclear genomic inheritance (Chap. 63). These nuclear-encoded proteins are synthesized in the cell cytoplasm and imported to their location of activity in mitochondria through a complex biochemical process. In addition, the mitochondria have their own genome, which consists of numerous copies (polyploidy) per mitochondrion of a circular, double-strand mitochondrial DNA (mtDNA) molecule consisting of a 16,569-nucleotide sequence. This mtDNA sequence contains a total of 37 genes, of which 13 encode mitochondrial protein components of the ETC. The remaining 22 tRNA- and 2 rRNA-encoding genes are dedicated to the process of translating the 13 mtDNA-encoded proteins. This dual genetic control of mitochondrial function results in unique and diagnostically challenging patterns of inheritance. This chapter focuses on heritable traits and diseases related to the mtDNA component of the dual genetic control of mitochondrial function. The reader is referred to Chaps. 63 and 387 for consideration of mitochondrial disease originating from mutations in the nuclear genome. These mutations include (1) nuclear genomic mutations that disrupt the integrity of the mitochondrial genome itself (mtDNA deletion and depletion states), (2) disorders due to mutations in nuclear genes that encode structural components or assembly factors of the oxidative phosphorylation complexes, and (3) mitochondrial disorders due to mutations in nuclear genes that encode proteins indirectly related to oxidative phosphorylation.

MITOCHONDRIAL DNA STRUCTURE AND FUNCTION

As a result of its circular structure and extranuclear location, the replication and transcription mechanisms of mtDNA differ from the corresponding mechanisms in the nuclear genome, whose nucleosomal packaging and structure are more complex. Since each mitochondrion contains many copies of mtDNA and because the number of mitochondria can vary during the lifetime of each cell through the processes of fission, fusion, and mitochondrial biogenesis, mtDNA copy number is not directly coordinated with the cell cycle. Thus, vast differences in mtDNA copy number are observed between different cell types and tissues and during the lifetime of a cell. Another important feature of the mtDNA replication process is a greatly reduced stringency of proofreading and replication error correction, leading to a greater degree of sequence variation compared with the nuclear genome. This fidelity limitation is due to the presence of one replicase, polymerase γ, which is solely responsible for both DNA replication and repair in mitochondria. Some of these sequence variants are silent polymorphisms that do not have the potential for a phenotypic or pathogenic effect, whereas others may be considered pathogenic mutations.

With respect to transcription, initiation can occur on both strands and proceeds through the production of an intronless polycistronic precursor RNA that then is processed to produce the 13 individual mRNA and 24 individual

tRNA and rRNA products. The 37 mtDNA genes account for fully 93% of the 16,569 nucleotides of the mtDNA in what is known as the *coding region*. The *control region* consists of ~1.1 kilobases (kb) of noncoding DNA that is thought to play a major role in replication and transcription initiation. The mutation rate is considerably higher in the control region, which contains a displacement, or D loop, which in turn contains two adjacent hypervariable regions (HVR-I and HVR-II) that give rise to large interindividual variability within the human population. Indeed, mtDNA sequence variants at both the coding and control regions are more highly partitioned across geographically defined populations than are sequence variants in other parts of the genome, and combinations of these sequence variants define phylogeographic mtDNA haplogroups and haplotypes. Accumulating evidence supports the notion that differences in these haplotypes are of medical significance in regard to predisposition to common diseases. The foregoing structural and functional features of mtDNA lead to the expectation that phenotypic inheritance and disease patterns for disorders related to mtDNA sequence variations and mutations should be quite different from the more familiar inheritance and disease patterns attributed to variation and mutation in nuclear DNA. Intensive research during the last two decades has confirmed that this is the case.

Maternal Inheritance and Lack of Recombination

In contrast to the homologous pair recombination that takes place in the nucleus, mtDNA molecules do not undergo recombination, and so mutational events represent the only source of mtDNA genetic diversification. Moreover, with very rare exceptions, it is only the maternal DNA that is transmitted to the offspring. The fertilized oocyte degrades mtDNA carried from the sperm in a complex process that involves the ubiquitin proteasome system. Thus, whereas mothers transmit their mtDNA to both their sons and their daughters, only the daughters are able to transmit the inherited mtDNA to future generations. Accordingly, mtDNA sequence variation and associated phenotypic traits and diseases are inherited exclusively along maternal lines.

As noted below, because of the complex relationship between mtDNA mutations and disease expression, sometimes this maternal inheritance is difficult to recognize at the clinical or pedigree level. However, evidence of paternal transmission almost certainly rules out an mtDNA genetic origin of phenotypic variation or disease; conversely, a disease affecting both sexes without evidence of paternal transmission strongly suggests a heritable mtDNA disorder **(Fig. e18-2)**.

Figure e18-2

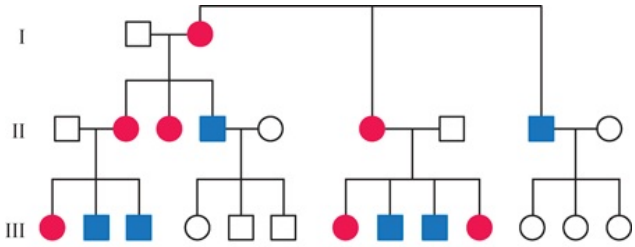

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Maternal inheritance of mtDNA disorders and heritable traits. Affected women (filled circles) transmit the trait to their children. Affected men (filled squares) do not transmit the trait to any of their offspring.

Multiple Copy Number (Polyploidy), Mitotic Segregation, and High Mutation Rate

Each aerobic cell in the body has multiple mitochondria, often numbering many hundreds or more in cells with extensive energy production requirements. Furthermore, the number of copies of mtDNA within each mitochondrion varies from several to hundreds; this is true of both somatic and germ cells, including oocytes in females. In the case of somatic cells, this means that the impact of most newly acquired somatic mutations is likely to be very small in terms of total cellular or organ system function; however, because of the manyfold higher mutation rate during mtDNA replication, numerous different mutations may accumulate with the aging of the organism. It has been proposed that the total cumulative burden of acquired somatic mtDNA mutations with age may result in an overall perturbation of mitochondrial function, contributing to an age-related reduction in the efficiency of oxidative phosphorylation and increased production of damaging ROS. According to this formulation, the high somatic mtDNA mutation rate and the global effect of numerous cumulative mutations that disrupt mitochondrial function counterbalance the reduced impact of the multiple copy number of each individual mtDNA mutation. The accumulation of such acquired somatic mtDNA mutations with aging may contribute to age-related diseases, such as metabolic syndrome and diabetes, cancer, neurodegenerative disease, and cardiovascular disease, in any specific individual. However, such somatic mtDNA mutations are not carried forward to the next generation, and the hereditary impact of mtDNA mutagenesis requires separate consideration of events in the female germ line.

The multiple mtDNA copy number within the maternal germ cells results in the phenomenon of heteroplasmy of inherited mtDNA mutations. Heteroplasmy for a particular mtDNA sequence variant or mutation results from the coexistence within the oocyte of mtDNA molecules bearing both versions of the sequence variant **(Fig. e18-3)**. In the case of pathogenic mutations, this means coexistence within the oocyte of both the wild-type and mutant versions. For each oocyte, the percentage of mtDNA molecules bearing each version of the polymorphic sequence variant or mutation depends on stochastic events related to partitioning of mtDNA molecules during the process of oogenesis itself. Thus, oocytes differ from each other in the degree of heteroplasmy for that sequence variant or mutation. In turn, the heteroplasmic state is carried forward to the zygote and to the organism as a whole to varying degrees, depending on mitotic segregation of mtDNA molecules during organ system development and maintenance. For this reason, in vitro fertilization followed by preimplantation genetic diagnosis (PGD) is not as predictive of the genetic health of the offspring in the case of mtDNA mutations as it is in the case

of the nuclear genome.

Figure e18-3

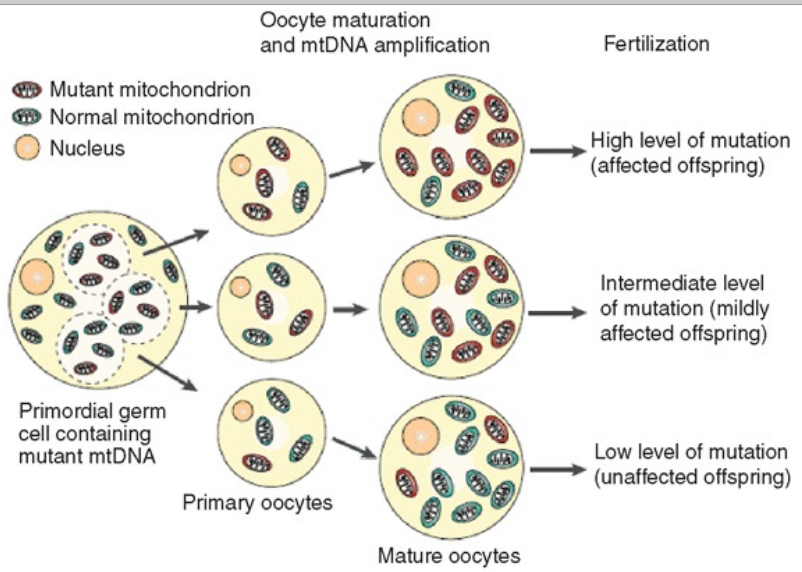

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Heteroplasmy and the mitochondrial genetic bottleneck. During the production of primary oocytes, a selected number of mitochondrial DNA (mtDNA) molecules are transferred into each oocyte. Oocyte maturation is associated with the rapid replication of this mtDNA population. This restriction-amplification event can lead to a random shift of mtDNA mutational load between generations and is responsible for the variable levels of mutated mtDNA observed in affected offspring from mothers with pathogenic mtDNA mutations. Mitochondria that contain mutated mtDNA are shown in red, and those with normal mtDNA are shown in green. *(Reproduced with permission from R Taylor, D Turnbull: Nature Reviews: Genetics 6:389, 2005.)*

Mitotic segregation refers to the unequal distribution of wild-type and mutant versions of the mtDNA molecules during all cell divisions that occur during prenatal development and subsequently throughout the lifetime of an individual. Such mtDNA sequence variants may be entirely silent in terms of disease phenotype or predisposition or, alternatively, potentially affect one or more aspects of mitochondrial function in a manner that gives rise to a disease phenotype or predisposition if present in high copy number. The phenotypic effect or disease impact thus will be a function not only of the inherent disruptive effect (pathogenicity) on the mtDNA-encoded gene (coding region mutations) or integrity of the mtDNA molecule (control region mutations) but also of its distribution among the multiple copies of mtDNA in the various mitochondria, cells, and tissues of the affected individual. This leads to a "threshold" effect in which the actual expression of disease depends on the relative percentage of mitochondria whose function is disrupted by mtDNA mutations. Consequently, there is tremendous heterogeneity in disease penetrance and severity as well as complexity of organ system involvement among the offspring of women with pathogenic heteroplasmic mutations. This heterogeneity arises from differences in the degree of heteroplasmy among oocytes of the affected woman and with subsequent mitotic segregation of the pathogenic mutation during tissue and organ development and throughout the lifetime of the individual offspring. This may confound the diagnosis of an mtDNA genetic cause of disease.

In contrast to classic mtDNA diseases, most of which begin in childhood and are the result of heteroplasmic mutations as noted above, during the course of human evolution, certain mtDNA sequence variants have drifted to a state of homoplasmy in which all the mtDNA molecules in the organism contain the new sequence variant. This arises due to a "bottleneck" effect followed by genetic drift during the very process of oogenesis itself (Fig. e18-3). In other words, during certain stages of oogenesis, the mtDNA copy number is reduced so substantially that the particular mtDNA species bearing the novel or derived sequence variant may become the increasingly predominant, and eventually exclusive, version of the mtDNA for that particular nucleotide site. All the offspring of a woman bearing an mtDNA sequence variant or mutation that has become homoplasmic will also be homoplasmic for that variant and will transmit the sequence variant forward in subsequent generations. This is the main process that establishes a new mtDNA haplotype in the human population.

Considerations of reproductive fitness limit the evolutionary or population emergence of homoplasmic mutations that are lethal or cause severe disease in infancy or childhood. Thus, with a number of notable exceptions (e.g., mtDNA mutations that cause Leber hereditary optic neuropathy; see below), most homoplasmic mutations were considered neutral markers of human evolution—useful and interesting in the population genetics analysis of shared maternal ancestry but with little significance in human phenotypic variation or disease predisposition. However, recent research and clinical attention have focused on the potential for certain of these homoplasmic mtDNA sequence variants to contribute to the evolutionary adaptation of populations to their climatic environment or to predispose to heritable late postreproductive and age-associated diseases.

MITOCHONDRIAL DNA DISEASE

The true prevalence of mtDNA disease is difficult to estimate because of the phenotypic heterogeneity that occurs

as a function of heteroplasmy, the challenge of detecting and assessing heteroplasmy in different affected tissues, and the other unique features of mtDNA function and inheritance described above. It is estimated that at least 1 in 200 healthy humans harbors a pathogenic mtDNA mutation that potentially causes disease but that heteroplasmic germ-line pathogenic mtDNA mutations affect up to approximately 1 in 5000 individuals. The true overall impact of mtDNA mutation in human health and disease may be much greater if the potential contribution of homoplasmic mtDNA sequence variation to common complex diseases that appear in the postreproductive age also is considered.

The true disease burden will be known only with the ability to distinguish a completely neutral sequence variant from a true phenotype-modifying or pathogenic mutation, when an accurate assessment of heteroplasmy can be determined with fidelity, and when the epistatic interactions of mtDNA sequence variations with mutations in the nuclear genome can be expressed using a systems biology approach (Chap. e19).

Overview of Clinical and Pathologic Features of Human $_{MT}$DNA Disease

In light of the vital roles of mitochondria in all nucleated cells, it is not surprising that mtDNA mutations can affect numerous tissues with pleiotropic effects. More than 200 different disease-causing, mostly heteroplasmic mtDNA mutations have been described that affect ETC function. Figure e18-4 provides a partial mtDNA map of some of the better characterized of these disorders. A number of clinical clues can increase the index of suspicion for a heteroplasmic mtDNA mutation as an etiology of a heritable trait or disease, including (1) familial clustering with absence of paternal transmission, (2) adherence to one of the classic syndromes (see below) or paradigmatic combinations of disease phenotypes involving several organ systems that normally do not fit together within a single nuclear genomic mutation category, (3) a complex of laboratory and pathologic abnormalities that reflect disruption in cellular energetics (e.g., lactic acidosis and neurodegenerative and myodegenerative symptoms with the finding of ragged red fibers, reflecting the accumulation of abnormal mitochondria under the muscle sarcolemmal membrane), and (4) a mosaic pattern reflecting a heteroplasmic state.

Figure e18-4

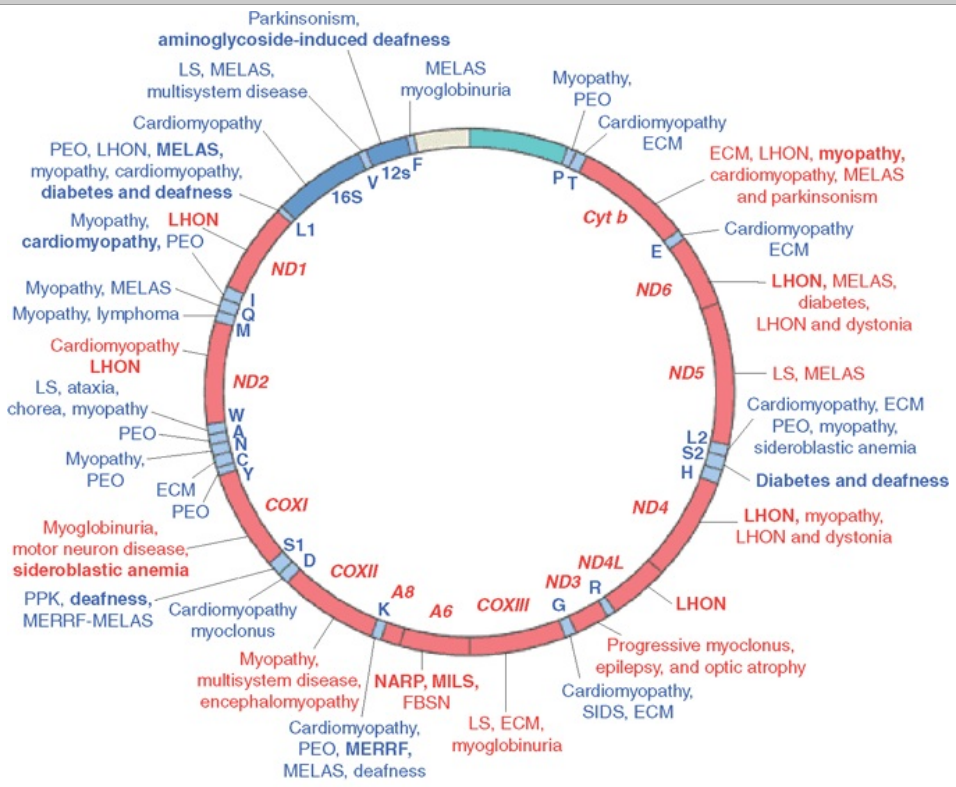

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mutations in the human mitochondrial genome known to cause disease. Disorders that are frequently or prominently associated with mutations in a particular gene are shown in boldface. Diseases due to mutations that impair mitochondrial protein synthesis are shown in blue. Diseases due to mutations in protein-coding genes are shown in red. ECM, encephalomyopathy; FBSN, familial bilateral striatal necrosis; LHON, Leber hereditary optic neuropathy; LS, Leigh syndrome; MELAS, mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes; MERRF, myoclonic epilepsy with ragged red fibers; MILS, maternally inherited Leigh syndrome; NARP, neuropathy, ataxia, and retinitis pigmentosa; PEO, progressive external ophthalmoplegia; PPK, palmoplantar keratoderma; SIDS, sudden infant death syndrome. *(Reproduced with permission from S DiMauro, E Schon: N Engl J Med 348: 2656, 2003.)*

Heteroplasmy sometimes can be demonstrated elegantly at the tissue level by using histochemical staining for enzymes in the oxidative phosphorylation pathway, with a mosaic pattern indicating heterogeneity of the genotype for the coding region for the mtDNA-encoded enzyme. Complex II, CoQ, and cytochrome c are encoded exclusively by nuclear DNA. In contrast, complexes I, III, IV, and V contain at least some subunits encoded by

mtDNA. Just 3 of the 13 subunits of the ETC complex IV enzyme, cytochrome c oxidase, are encoded by mtDNA; therefore, this enzyme has the lowest threshold for dysfunction when a threshold of mutated mtDNA is reached. Histochemical staining for cytochrome c oxidase activity in tissues of patients affected with heteroplasmic inherited mtDNA mutations (or with the somatic accumulation of mtDNA mutations; see below) can show a mosaic pattern of reduced histochemical staining in comparison with histochemical staining for the complex II enzyme, succinate dehydrogenase **(Fig. e18-5)**. Heteroplasmy also can be detected at the genetic level through direct mtDNA genotyping in special conditions, though clinically significant low levels of heteroplasmy can escape detection in genomic samples extracted from whole blood using conventional genotyping and sequencing techniques.

Figure e18-5

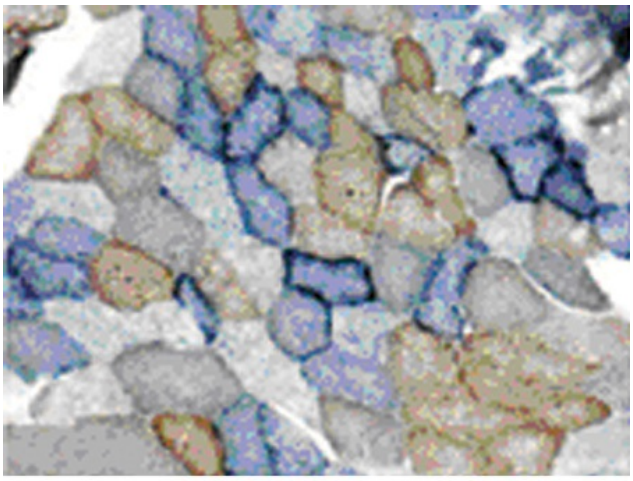

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

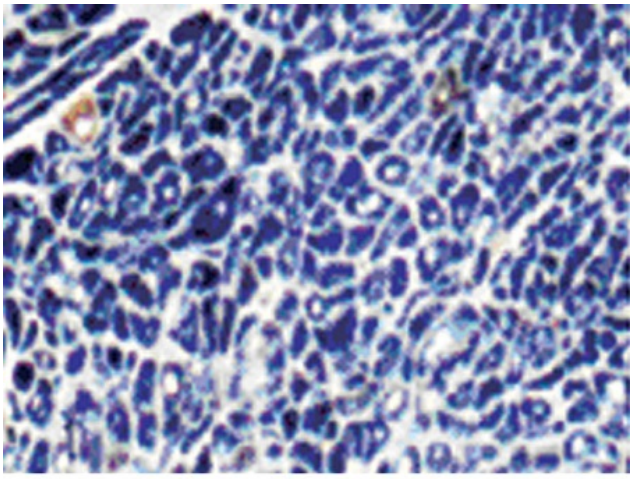

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

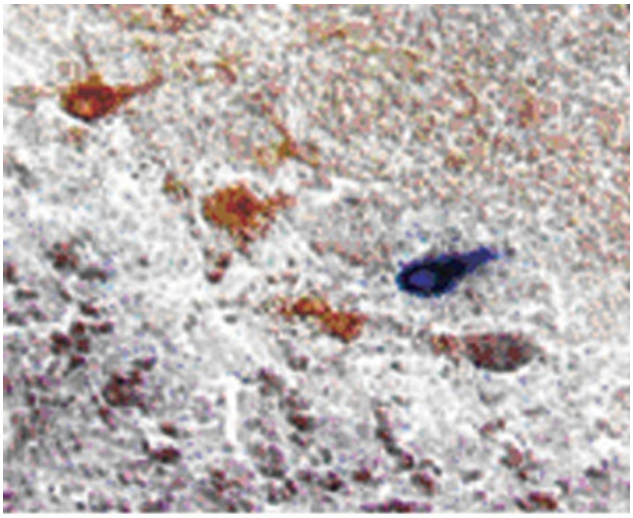

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

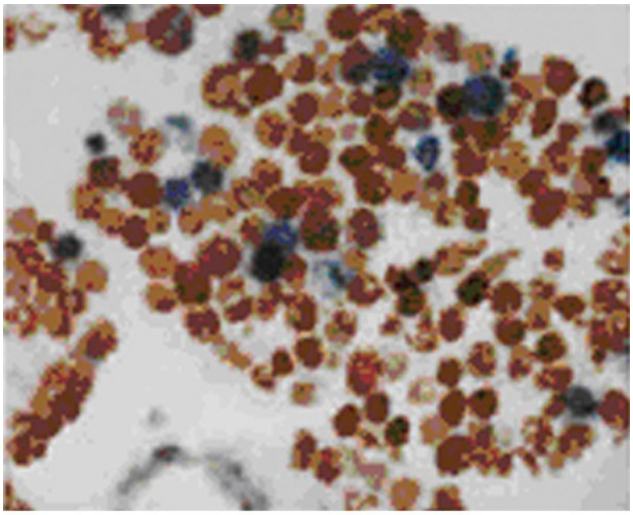

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

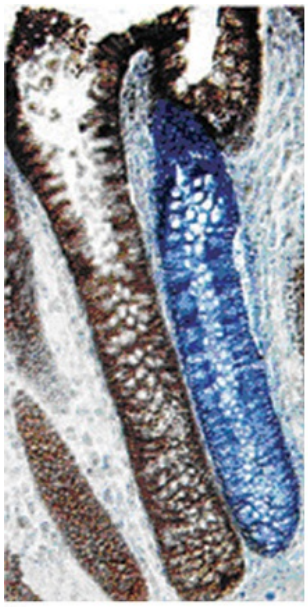

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cytochrome c oxidase deficiency in mitochondrial DNA–associated disease. Transverse tissue sections that have been stained for cytochrome c oxidase (COX) and succinate dehydrogenase (SDH) activities sequentially, with COX-positive cells shown in brown and COX-deficient cells shown in blue. ***A.*** Skeletal muscle from a patient with a heteroplasmic mitochondrial tRNA point mutation. The section shows a typical "mosaic" pattern of COX activity, with many muscle fibers harboring levels of mutated mtDNA that are above the crucial threshold to produce a functional enzyme complex. ***B.*** Cardiac tissue (left ventricle) from a patient with a homoplasmic tRNA mutation that causes hypertrophic cardiomyopathy, which demonstrates an absence of COX in most cells. ***C.*** A section of cerebellum from a patient with mtDNA rearrangement that highlights the presence of COX-deficient neurons. ***D, E.*** Tissues that show COX deficiency due to clonal expansion of somatic mtDNA mutations within single cells—a phenomenon that is seen in both postmitotic cells (***D***; extraocular muscles) and rapidly dividing cells (***E***; colonic crypt) in aging humans. *(Reproduced with permission from R Taylor, D Turnbull: Nature Reviews: Genetics 6:389, 2005.)*

Clinically, the most striking overall characteristic of mitochondrial genetic disease is the phenotypic heterogeneity associated with mtDNA mutations. This extends to intrafamilial phenotypic heterogeneity for the same mtDNA pathogenic mutation and, conversely, to the overlap of phenotypic disease manifestations with distinct mutations. Thus, although fairly consistent and well-defined "classic" syndromes have been attributed to specific mutations, frequently "nonclassic" combinations of disease phenotypes ranging from isolated myopathy to extensive multisystem disease are encountered, rendering genotype-phenotype correlation challenging. In both classic and nonclassic mtDNA disorders, there is often a clustering of some combination of abnormalities affecting the neurologic system (including optic nerve atrophy, pigment retinopathy, sensorineural hearing loss), cardiac and skeletal muscle (including extraocular muscles), and endocrine and metabolic systems (including diabetes mellitus). Additional organ systems that may be affected include the hematopoietic, renal, hepatic, and gastrointestinal systems, though these systems are involved more frequently in infants and children. Disease-causing mtDNA coding region mutations can affect either one of the 13 protein encoding genes or one of the 24 protein synthetic genes. Clinical manifestations do not readily distinguish these two categories, though lactic acidosis and muscle pathologic findings tend to be more prominent in the latter. In all cases, either defective ATP production due to disturbances in the ETC or enhanced generation of reactive oxygen species has been invoked as the mediating biochemical mechanism between mtDNA mutation and disease manifestation.

MTDNA Disease Presentations

The clinical presentation of adult patients with mtDNA disease can be divided into three categories: (1) clinical features suggestive of mitochondrial disease **(Table e18-2)** but not a well-defined classic syndrome, (2) classic mtDNA syndromes, and (3) clinical presentation confined to one organ system (e.g., isolated sensorineural deafness, cardiomyopathy, or diabetes mellitus).

Table e18-2 Common Features of mtDNA-Associated Diseases in Adults
Neurologic: stroke, epilepsy, migraine headache, peripheral neuropathy, cranial neuropathy (optic atrophy, sensorineural deafness, dysphagia, dysphasia)
Skeletal myopathy: ophthalmoplegia, exercise intolerance, myalgia
Cardiac: conduction block, cardiomyopathy
Respiratory: hypoventilation, aspiration pneumonitis
Endocrine: diabetes mellitus, premature ovarian failure, hypothyroidism, hypoparathyroidism
Ophthalmologic: cataracts, pigment retinopathy, neurologic and myopathic (optic atrophy, ophthalmoplegia)

Table e18-3 provides a summary of eight illustrative classic mtDNA syndromes or disorders that affect adult patients and highlights some of the most interesting features of mtDNA disease in terms of molecular

pathogenesis, inheritance, and clinical presentation. The first five of these syndromes result from heritable point mutations in either protein-encoding or protein synthetic mtDNA genes; the other three result from rearrangements or deletions that usually do not involve the germ line.

Table e18-3 Mitochondrial Diseases Due to mtDNA Point Mutations and Large-Scale Rearrangements

Disease	Phenotype	Most Common mtDNA Mutations	Homoplasmic (usually)	Maternal
Leber Hereditary Optic Neuropathy (LHON)	Bilateral subacute or acute painless optic atrophy	G11778A, T14484C, G3460A	Homoplasmic	Maternal
NARP, Leigh disease	Loss of central vision leading to blindness in young adult life	G1778A, T14484C, G3460A	Heteroplasmic	Maternal
MELAS	*M*itochondrial *e*ncephalomyopathy, *l*actic *a*cidosis, and *s*trokelike episodes; may manifest only as diabetes	Point mutation in $tRNA^{leu}$	Heteroplasmic	Maternal
MERRF	*M*yoclonic *e*pilepsy, *r*agged *r*ed *f*ibers in muscle, ataxia, increased CSF protein, sensorineural deafness, dementia	Point mutation in $tRNA^{lys}$	Heteroplasmic	Maternal
Deafness	Progressive sensorineural deafness, often induced by aminoglycoside antibiotics.	A1555G mutation in 12S rRNA	Homoplasmic	Maternal
	Nonsyndromic sensorineural deafness	A7445G mutation in 12S rRNA	Homoplasmic	Maternal
Chronic *p*rogressive *e*xternal *o*phthalmoplegia (PEO)	Late-onset bilateral ptosis and ophthalmoplegia, proximal muscle weakness, and exercise intolerance	Single deletions or duplications	Heteroplasmic	Mostly sporadic, somatic mutations
Pearson syndrome	Pancreatic insufficiency, pancytopenia, lactic acidosis	Large deletion	Heteroplasmic	Sporadic, somatic mutations
Kearn-Sayre syndrome (KSS)	External ophthalmoplegia, heart block, retinal pigmentation, ataxia	The 5-kb "common deletion"	Heteroplasmic	Sporadic, somatic mutations

Abbreviations: CSF, cerebrospinal fluid; NARP, neuropathy, ataxia, retinitis pigmentosa.

Leber hereditary optic neuropathy (LHON) is a common cause of maternally inherited visual failure. LHON typically presents during young adulthood with subacute painless loss of vision in one eye, with symptoms developing in the other eye 6–12 weeks after the initial onset. In some instances, cerebellar ataxia, peripheral neuropathy, and cardiac conduction defects are observed. In >95% of cases, LHON is due to one of three homoplasmic point mutations of mtDNA that affect genes that encode different subunits of complex I of the mitochondrial ETC; however, not all individuals who inherit a primary LHON mtDNA mutation develop optic neuropathy, indicating that additional environmental (e.g., tobacco exposure) or genetic factors are important in the etiology of the disorder. Both the nuclear and the mitochondrial genomic background modify disease penetrance. Indeed, a region of the X chromosome containing a high-risk haplotype for LHON has been identified, supporting the formulation that nuclear genes act as modifiers and affording an explanation for the male prevalence of LHON. This haplotype can be used in predictive genomic testing and prenatal screening for this disease. In contrast to the other classic mtDNA disorders, it is of interest that patients with this syndrome are often homoplasmic for the disease-causing mutation. The somewhat later onset in young adulthood and the modifying effect of protective background nuclear genomic haplotypes may have enabled homoplasmic pathogenic mutations to escape evolutionary censoring.

Mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes (MELAS) is probably the most common mtDNA disease, consisting of a progressive encephalomyopathy characterized by repeated strokelike events involving mainly posterior cerebral areas. Of note, brain lesions do not respect the distribution of vascular territories. Recurrent migraine-like headache and vomiting, exercise intolerance, seizures, short stature, and lactic acidosis are other common clinical features. The most commonly described pathogenic point mutations are A3243G and T3271C in the gene encoding the leucine tRNA.

Myoclonic epilepsy with ragged red fibers (MERRF) is a multisystem disorder characterized by myoclonus, seizures, ataxia, and myopathy with ragged red fibers. Hearing loss, exercise intolerance, neuropathy, and short stature are often present. Almost all MERRF patients have mutation in the mtDNA $tRNA^{lys}$ gene, and the A8344G mutation in the mtDNA gene encoding the lysine amino acid tRNA is responsible for 80–90% of MERRF cases.

Neurogenic weakness, ataxia, and retinitis pigmentosa (NARP) is characterized by moderate diffuse cerebral and cerebellar atrophy and symmetric lesions of the basal ganglia on MRI. A heteroplasmic T8993G mutation in the gene ATPase 6 subunit gene has been identified as causative. Ragged red fibers are not observed in muscle biopsy. When >95% of mtDNA molecules are mutant, a more severe clinical neuroradiologic and neuropathologic picture (Leigh syndrome) emerges. Point mutations in the mtDNA gene encoding the 12S rRNA result in heritable nonsyndromic hearing loss. One such mutation causes heritable ototoxic susceptibility to aminoglycoside

antibiotics, which opens a pathway for a simple pharmacogenetic test in the appropriate clinical settings.

Kearns-Sayre syndrome (KSS), sporadic progressive external ophthalmoplegia (PEO), and Pearson syndrome are three disease phenotypes caused by large-scale mtDNA rearrangements, including partial deletions or partial duplication. The majority of single large-scale rearrangements of mtDNA are thought to result from clonal amplification of a single sporadic mutational event occurring in the maternal oocyte or during early embryonic development. Since germ-line involvement is rare, most cases are sporadic rather than inherited.

KSS is characterized by the triad of onset before age 20, chronic progressive external ophthalmoplegia, and pigmentary retinopathy. Cerebellar syndrome, heart block, increased cerebrospinal fluid protein content, diabetes, and short stature are also part of the syndrome. Single deletions/duplication also can result in milder phenotypes such as PEO, characterized by late-onset progressive external ophthalmoplegia, proximal myopathy, and exercise intolerance. In both KSS and PEO, diabetes mellitus and hearing loss are common accompaniments. Pearson syndrome also is characterized by diabetes mellitus from pancreatic insufficiency, together with pancytopenia and lactic acidosis, caused by the large-scale sporadic deletion of several mtDNA genes.

Two important dilemmas in classic mtDNA disease have benefited from recent important research insights. The first relates to the greater involvement of neuronal, muscular, renal, hepatic, and pancreatic manifestations in mtDNA disease in these syndromes. This observation appropriately has been attributed mostly to the high energy utilization of the involved tissues and organ systems and, hence, greater dependency on mitochondrial ETC integrity and health. However, since mutations are stochastic events, mitochondrial mutations should occur in any organ during embryogenesis and development. Recently, additional explanations have been suggested based on studies of the common A3243G transition. The proportion of this mutation in peripheral blood cells was shown to decrease exponentially with age. A selective process acting at the stem cell level with a strong bias against the mutated form would have its greatest effect in reducing the mutant mtDNA only in highly proliferating cells, such as those derived from the hematopoietic system. Tissues and organs with lower cell turnover, such as those involved with mtDNA mutations, would not benefit from this effect and thus would be affected the most.

Another important question of interest arises from the observation that only a subset of mtDNA mutations account for the majority of the familial mtDNA diseases. The random occurrence of mutations in the mtDNA sequence should yield a more uniform distribution of disease-causing mutations. However, recent studies utilizing the introduction of one severe and one mild point mutation into the female germ line of experimental animals demonstrated selective elimination during oogenesis of the severe and selective retention of the milder mutations, with the emergence of mitochondrial disease in offspring after multiple generations. Thus, oogenesis itself can act as an "evolutionary" filter for mtDNA disease.

The Investigation of Suspected MTDNA Disease

The clinical presentations of classic syndromes, groupings of disease manifestations in multiple organ systems, or unexplained isolated presentations of one of the disease features of a classic mtDNA syndrome should prompt a systematic clinical investigation, as outlined in **Fig. e18-6**. Despite the centrality of disruptive oxidative phosphorylation, an elevated blood lactate level is neither specific nor sensitive because there are many causes of blood lactic acidosis, and many patients with mtDNA defects that present in adulthood have normal blood lactate. An elevated cerebrospinal fluid lactate is a more specific test for mitochondrial disease if there is central nervous system involvement. The serum creatine kinase may be elevated but is often normal even in the presence of a proximal myopathy. Urinary organic and amino acids also may be abnormal, reflecting metabolic and kidney proximal tubule dysfunction. Every patient with seizures or cognitive decline should have an electroencephalogram. A brain CT scan may show calcified basal ganglia or bilateral hypodense regions with cortical atrophy. MRI is indicated in patients with brainstem signs or strokelike episodes.

Figure e18-6

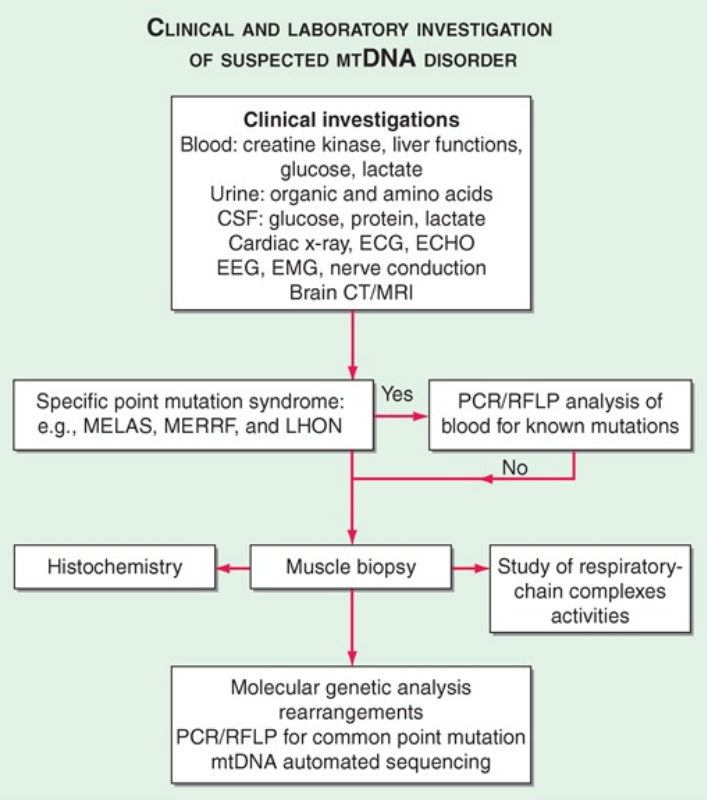

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical and laboratory investigation of suspected mtDNA disorder. CSF, cerebrospinal fluid; ECG, electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; MELAS, mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes; MERFF, myoclonic epilepsy with ragged red fibers; LHON, Leber hereditary optic neuropathy; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism.

For some mitochondrial diseases, it is possible to obtain an accurate diagnosis with a simple molecular genetic screen. For examples, 95% of patients with LHON harbor one of three mtDNA point mutations (A11778G, A3460G, and T14484C). These patients have very high levels of mutated mtDNA in peripheral blood cells, and it is therefore appropriate to send a blood sample for molecular genetic analysis by polymerase chain reaction (PCR) or restriction fragment length polymorphism. The same is true for most MERRF patients who harbor a point mutation in the lysine tRNA gene at position 8344. In contrast, patients with the A3243G MELAS mutation often have low levels of mutated mtDNA in blood. If clinical suspicion is strong enough to warrant peripheral blood testing, patients with a negative result should be investigated further with a skeletal muscle biopsy.

Muscle biopsy histochemical analysis is the cornerstone for investigation of patients with suspected mitochondrial disease. Histochemical analysis may show subsarcolemmal accumulation of mitochondria with the appearance of ragged red fibers. Electron microscopy may show abnormal mitochondria with paracrystalline inclusions. Muscle histochemistry may show cytochrome c oxidase (COX)–deficient fibers, which indicate mitochondrial dysfunction (Fig. e18-5). Respiratory chain complex assays also may show reduced enzyme function. Either of these two abnormalities confirms the presence of a mitochondrial disease, to be followed by an in-depth molecular genetic analysis.

Relevant evidence has provided important insights into the importance of nuclear-mtDNA genomic cross-talk and has provided a descriptive framework for classifying and understanding disorders that emanate from perturbations in this cross-talk. Although these are not strictly considered as mtDNA genetic disorders, their manifestations overlap with those highlighted above.

IMPACT OF HOMOPLASMIC SEQUENCE VARIATION ON HERITABLE TRAITS AND DISEASE

The relationship among the degree of heteroplasmy, tissue distribution of the mutant mtDNA, and disease phenotype simplifies the inference of a clear causative relationship between heteroplasmic mutation and disease. With the exception of certain mutations (e.g., those causing most cases of LHON), the drift to homoplasmy of such mutations would be precluded normally by the severity of impaired oxidative phosphorylation and the consequent reduction in reproductive fitness. Therefore, it previously was thought that sequence variants that have reached homoplasmy should be neutral in terms of human evolution and useful only for tracing human evolution, demography, and migration; however, recent studies have suggested that some homoplasmic mtDNA sequence variants may affect heritable traits or health through one or more mechanisms.

The first such mechanism relates to locally adaptive evolutionary forces. As noted above, homoplasmic mtDNA sequence variants that partition population groups are designated as defining maternal "haplogroups" or "haplotypes." Striking discontinuities have been observed in mtDNA haplogroup distribution among climatic zones around the globe. This phenomenon has been attributed to adaptive differences in the efficiency of oxidative phosphorylation and consequent thermogenesis in accordance with differences in prevailing climates in different

global geographic regions during much of human evolution. A potential health implication of this finding is the possibility that these mutations might result in deleterious effects on energy metabolism and caloric balance in the current era of human transglobal migration and ability to modulate residential climate.

A much broader extrapolation of the foregoing mechanism states that many homoplasmic mtDNA mutations affect human health in the postreproductive age only and therefore escaped evolutionary censoring altogether. In the modern era of increased median life span, such mutations are postulated to account for a considerable burden of age-associated common complex disease. During the previous century alone, mean life expectancy rose from ~47 years to ~77 years in many parts of the developing world; therefore, late-onset effects of a subset of homoplasmic mtDNA mutations may contribute significantly to the burden of human illness only in the current era, when a relatively higher percentage of the population is surviving beyond reproductive age. The challenge is to identify those homoplasmic mutations that modify mtDNA function and contribute to late-onset, common complex disease. Indeed, in light of the finding that global populations are more differentiated at the level of mtDNA than they are at the level of the nuclear genome, it is also attractive to postulate that population differences with the predisposition to certain late-onset common complex metabolic diseases may be attributed in part to population-based mtDNA sequence variation. The diseases that have been of particular interest are those involving the very organ systems familiar from the known classic heteroplasmic mtDNA syndromes described above.

Metabolic Syndrome and Type 2 Diabetes Mellitus (T2DM)

Insulin release by pancreatic beta cells is modulated in response to ATP metabolism, and insulin action is perturbed by metabolites of mitochondrial fatty acid oxidation. This has led investigators to consider mtDNA itself as a potential genomic locus for susceptibility to T2DM. A rather clear-cut case is that of a mutation in mtDNA nucleotide 3243 encoding the mitochondrial tRNA for the amino acid leucine. Even a low level of heteroplasmy for a particular point mutation in the mtDNA tRNA gene encoding the leucine tRNA is thought to contribute to the pathogenesis of up to 1% of all cases of T2DM. This and other findings at the biochemical and population genetics levels have motivated the search for more definitive evidence of the role of homoplasmic variants in the predisposition to metabolic syndrome and T2DM. Such evidence has been obtained with the finding of significant segregation of a homoplasmic mtDNA tRNA mutation (T to C transition in the nucleotide immediately 5' to the isoleucine tRNA anticodon) with metabolic syndrome phenotypes in a large Caucasian kindred.

Since the metabolic syndrome is so common and can result from numerous different genetic susceptibility loci and environmental causes and since many nuclear genetic susceptibility loci and environmental risk factors have been identified, special features in this particular reported kindred enabled the accurate distinction of affected from unaffected individuals for purposes of determining the causality of the mtDNA variant. The affected individuals had signs of hypomagnesemia, hypertension, and hypercholesterolemia. This particular mutation in a tRNA-encoding mtDNA gene also highlights the expected difference in the phenotypic impact of mutations in genomic regions encoding tRNAs in the mitochondrial versus the nuclear genome, since a mutation in the latter would affect too many gene products to be compatible with life.

Beyond the immediate medical and biologic significance, the importance of the findings in this kindred is to highlight the fact that the contribution of common homoplasmic mutations to common complex late-onset human disease syndromes is probably underestimated. A common variant mtDNA sequence variant (T16189C) has been related to low birth weight, impaired glucose tolerance, and metabolic syndrome in specific populations. However, rigorous population-based association studies using case-control designs have not provided definitive evidence for a relationship between mtDNA haplogroups and susceptibility to T2DM or its complications.

Neurodegenerative Disease

The prominence of neurologic injury in classic mtDNA diseases, together with the presumed role of reactive oxygen species in neuronal injury and the late age of onset of neurodegenerative diseases, has led investigators to consider the possibility that homoplasmic variants in mtDNA sequence that define population haplogroups also may modify the susceptibility to neurodegenerative diseases such as Parkinson's disease and Alzheimer's disease. Particular configurations of mtDNA sequence polymorphisms that define population haplogroups designated in phylogenetics by the labels J, T, U, and K have been reported to be potentially protective against Parkinson's disease in different populations. In the case of Alzheimer's disease, some studies have shown haplogroup J to increase risk, with haplogroup D decreasing risk. Mutations in the mtDNA control region do not produce defective polypeptide products but affect both the light and heavy strand promoters, as well as the heavy strand origin of replication, and thus may modulate mtDNA replication and transcription. Mitochondrial DNA control region sequence variants (e.g., T414G) have been identified in Alzheimer's disease brains in association with a significant reduction in mtDNA copy number and a reduction in specific transcripts. A number of studies have focused on the interaction of mtDNA haplogroup–designating mutations with the well-established Alzheimer's disease risk alleles at the nuclear apoE4 locus. From those studies it was postulated that the ETC-uncoupling mutations that minimize ROS production are those which confer protection against neuronal injury, but definitive proof of this postulate awaits further studies.

Other Diseases and Nondisease Heritable Traits

Consideration of the potential contribution of mtDNA mutations to numerous heritable traits and common complex diseases requires consideration of the common variant–common phenotype model (including disease phenotype) versus the rare variant–common phenotype model, which are also applicable to the nuclear genome. According to the common variant–common phenotype model, DNA sequence variants inherited identically by descent and present in large numbers of individuals within one or more populations may predispose to common phenotypes. In the rare variant–common phenotype model, different mutations within one or more genetic loci involved in a particular molecular pathway may predispose to a common phenotype or disease. In this regard, the entire mtDNA can be considered a single genomic locus. Genomewide association studies have been utilized to try to map common variants responsible for common diseases, using case-control or multiplex family approaches. These approaches have been applied to common variants in mtDNA sequence as well, as noted above for metabolic syndrome and neurodegenerative disease. Additional examples include the variable length of an mtDNA control

region polycytosine stretch (16189 variant) as a contributing genomic influence in the onset of age-related cardiomyopathy with T2DM. An association of mtDNA haplogroup T and a polymorphism at position 13368 with hypertrophic cardiomyopathy has been reported in a European population, and a number of studies have suggested an association between mtDNA mutations and mitochondrial dysfunction in predisposition to heart failure. In the case of age-related cancers as well, the association of a number of heritable homoplasmic mtDNA mutations with certain cancers has been reported, including prostate, kidney, and breast cancer.

The association of mtDNA haplogroups with at least two nondisease heritable traits also has been studied: life expectancy and exercise endurance. Several mtDNA control region mutations, including the C150T mutation that shifts the heavy chain origin of replication, have been reported to accumulate with age in specific tissues, including lymphocytes of centenarians and their twins. The relationship between the C150T mutation and longevity has been replicated in Italian, Finnish, and Japanese populations, suggesting a common ancient origin. The alternative of evolutionary convergence of this mutation for longevity seems less likely, as the trait does not confer reproductive advantage. The association of haplogroup J and its subhaplogroups with longevity has been demonstrated in north Italian, north Irish, and Finnish sample sets. At least in the Italian study, this association was shown to be population-specific, since it was not reproduced in sample sets of southern Italian communities. Furthermore, in the case of the north Italian communities, an additional interaction of the mtDNA haplogroup designated as J2 with several different mutations adjacent to replication origins, including the aforementioned C150T, has been noted.

The functional importance of one or more of the mutations designated as haplogroup J is strengthened further by the finding of the interesting interaction with the mtDNA mutations that cause LHON, as noted previously. Reduced disease predilection suggests that one or more of the ancient sequence variants designated as haplogroup J appear to attenuate predisposition to degenerative disease in the face of other risk factors. It has been proposed that the mtDNA haplogroups associated with exceptional longevity favor a relatively uncoupled state of the ETC, with reduced efficiency of production of ATP and ROS and increased thermogenesis. Although this has not been demonstrated biochemically, the notion is strengthened by the finding of a relative paucity of these mtDNA haplogroups among successful endurance athletes, for whom maximum efficiency of oxidative phosphorylation confers an athletic competitive advantage.

It should be noted that not all studies have replicated associations of mtDNA haplogroups with longevity, athletic performance, or other heritable phenotypes. Most of these studies are limited by small sample sizes, possible genotyping inaccuracies, and the possibility of population stratification or ethnic ancestry bias. Since mtDNA haplogroups are so prominently partitioned along phylogeographic lines, it is difficult to rule out the possibility that a haplogroup for which an association has been found is simply a marker for differences in populations with a societal or environmental difference or with different allele frequencies at other genomic loci that are actually causally related to the heritable trait or disease of interest. The difficulty in generating cellular or animal models to test the functional influence of homoplasmic sequence variants (as a result of mtDNA polyploidy) further compounds the challenge. The most likely formulation is that the risk conferred by different mtDNA haplogroup –defining homoplasmic mutations for common diseases depends on the concomitant nuclear genomic background, together with environmental influences. Progress in minimizing potentially misleading associations in mtDNA heritable trait and disease studies should include ensuring adequate sample size taken from a large sample recruitment base, together with the use of carefully matched controls and population structure determination, along with analysis that takes into account epistatic interactions with other genomic loci and environmental factors.

IMPACT OF ACQUIRED SOMATIC MTDNA MUTATION ON HUMAN HEALTH AND DISEASE

Studies on aging humans and animals have shown a potentially important correlation of age with the accumulation of heterogeneous mtDNA mutations, especially in the organ systems that undergo the most prominent age-related degenerative tissue phenotype. Sequencing of PCR-amplified single mtDNA molecules has demonstrated an average of two to three point mutations per molecule in elderly subjects compared with younger ones. Point mutations observed include those responsible for known heritable heteroplasmic mtDNA disorders such as the A3344G and A3243G mutations responsible for the MERRF and MELAS syndromes, respectively. However, the cumulative burden of these acquired somatic point mutations with age was observed to remain well below the threshold expected for phenotypic expression (<2%). Point mutations at other sites not normally involved in inherited mtDNA disorders also have been shown to accumulate to much higher levels in some tissues of elderly individuals, with the description of tissue-specific "hot spots" for mtDNA point mutations. Along the same lines, an age-associated and tissue-specific accumulation of mtDNA deletions has been observed, including deletions involved in known heritable mtDNA disorders, as well as others. The accumulation of functional mtDNA deletions in a given tissue is expected to be associated with mitochondrial dysfunction, as reflected in an age-associated patchy and reduced cytochrome c oxidase activity on histochemical staining, especially in skeletal and cardiac muscle and brain. A particularly well-studied and potentially important example is the accumulation of mtDNA deletions and cytochrome c oxidase deficiency observed in neurons of the substantia nigra in Parkinson's disease patients.

The progressive accumulation of ROS has been proposed as the key factor connecting mtDNA mutations with aging and age-related disease pathogenesis **(Fig. e18-7)**. As noted above, ROS are a by-product of oxidative phosphorylation and are removed by detoxifying antioxidants into less harmful moieties; however, exaggerated production of ROS or impaired removal results in their accumulation. One of the main targets for ROS-mediated injury is DNA, and mtDNA is particularly vulnerable because of its lack of protective histones and less efficient injury repair systems compared with nuclear DNA. In turn, accumulation of mtDNA mutations results in inefficient oxidative phosphorylation, with the potential for excessive production of ROS, generating a "vicious cycle" of cumulative mtDNA damage. Indeed, measurement of the oxidative stress biomarker 8-hydroxy-2-deoxyguanosine has been used to measure age-dependent increases in mtDNA oxidative damage at a rate exceeding that of nuclear DNA. It should be noted that mtDNA mutation potentially can occur in postmitotic cells as well, since mtDNA replication is not synchronized with the cell cycle. Two other proposed links between mtDNA mutation and

aging, besides ROS-mediated tissue injury, are the perturbations in efficiency of oxidative phosphorylation with disturbed cellular aerobic function and perturbations in apoptotic pathways whose execution steps involve mitochondrial activity.

Figure e18-7

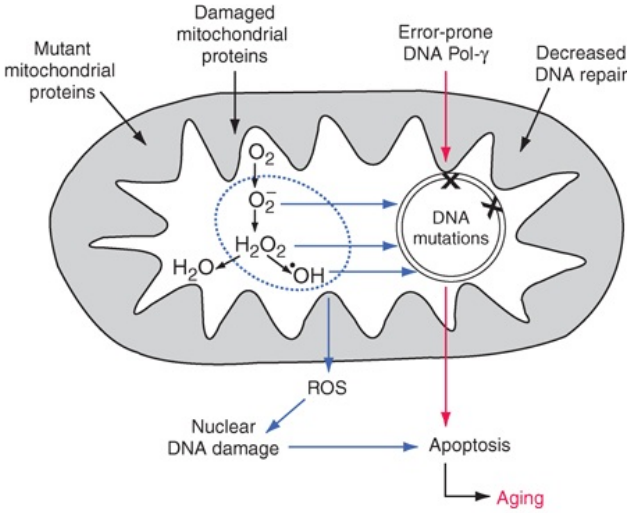

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple pathways of DNA damage and aging. Mitochondrial DNA damage and aging. Multiple factors may impinge on the integrity of mitochondria that lead to loss of cell function, apoptosis, and aging. The classic pathway is indicated with blue arrows; the generation of ROS (superoxide anion, hydrogen peroxide, and hydroxyl radicals), as a by-product of mitochondrial oxidative phosphorylation, results in damage to mitochondrial macromolecules, including the mtDNA, with the latter leading to deleterious mutations. When these factors damage the mitochondrial energy-generating apparatus beyond a functional threshold, proteins are released from the mitochondria that activate the caspase pathway, leading to apoptosis, cell death, and aging. ROS, reactive oxygen species. *(Reproduced with permission from L Loeb et al: Proc Natl Acad Sci USA 102:18769, 2005.)*

Genetic intervention studies in animal models have sought to clarify the potential causative relationship between acquired somatic mtDNA mutation and the aging phenotype, the role of ROS in particular. Replication of the mitochondrial genome is mediated by the activity of the nuclear-encoded polymerase γ gene. A transgenic homozygous mouse knock-in mutation of this gene renders the polymerase enzyme deficient in proofreading and results in a three- to fivefold increase in the mtDNA mutation rate. Such mice develop a premature aging phenotype, which includes subcutaneous lipoatrophy, alopecia, kyphonia, and weight loss with premature death. Although the finding of increases in mtDNA mutation and mitochondrial dysfunction with age has been solidly established, the causative role and specific contribution of mitochondrial ROS to aging and age-related disease in humans have yet to be proved. Similarly, although many tumors display higher levels of heterogeneous mtDNA mutations, a causal relationship to tumorigenesis has not been proved.

Besides the age-dependent acquired accumulation in somatic cells of heterogeneous point mutations and deletions, a quite different effect of nonheritable and acquired mtDNA mutation has been described affecting tissue stem cells. In particular, disease phenotypes attributed to acquired mtDNA mutation have been observed in sporadic and apparently nonfamilial cases involving a single individual or even tissue, usually skeletal muscle. The presentation consists of decreased exercise tolerance and myalgias, sometimes progressing to rhabdomyolysis. As in the case of the sporadic heteroplasmic large-scale deletion classic syndromes of chronic PEO, Pearson syndrome, and KSS, the absence of a maternal inheritance pattern, together with the finding of limited tissue distribution, suggests a molecular pathogenic mechanism emanating from mutations arising de novo in muscle stem cells after germ-line differentiation (somatic mutations that are not sporadic and occur in tissue-specific stem cells during fetal development or in the postnatal maintenance or postinjury repair stage). Such mutations would be expected to be propagated only within the progeny of that stem cell and affect a particular tissue within a specific individual, without evidence of heritability.

PROSPECTS FOR PREVENTION AND TREATMENT OF MTDNA DISEASE

Genetic Counseling in MTDNA Disorders

The provision of accurate genetic counseling and reproductive options to families with mtDNA mutations is complicated by the unique genetic features of mtDNA that distinguish it from Mendelian genetics. Although there is no risk of disease transmission from an affected male, the risk of maternal transmission of disease phenotypes associated with heteroplasmic mutations is a function of differential segregation and copy number of mutant mtDNA during oogenesis and, subsequently, after tissue and organ development in the offspring. This is rarely predictable with any degree of accuracy. In addition, interactions with the mtDNA haplotype background or nuclear human genome (as in the case of LHON) serve as an additional important determinant of disease penetrance. Environmental interactions are also of importance, such as the strong and consistent association between visual loss in LHON and smoking. A clinical penetrance of 93% was found in men who smoked. Asymptomatic carriers of a LHON mtDNA mutation therefore should be strongly advised not to smoke and to moderate their alcohol intake. Although not a cure, these interventions might stave off the devastating clinical manifestations of the LHON

mutation. Another example is the familial syndrome of ototoxic susceptibility to aminoglycosides in the case of the mtDNA A1555G mutation of the 12S rRNA encoding gene.

The value of prenatal genetic testing is also questionable, partly owing to the absence of data on the rules that govern the segregation of wild-type and mutant mtDNA species (heteroplasmy) among tissue in the developing embryo. Three factors are required to ensure the reliability of prenatal testing: (1) a close correlation between the mutant load and the disease severity, (2) a uniform distribution of mutant load among tissues, and (3) no major change in mutant load with time. These criteria are suggested to be fulfilled for the NARP T8993G mutation but do not seem to apply to other mtDNA disorders. In fact, the level of mutant mtDNA in a chorionic villus or amniotic fluid sample may be very different from the level in the fetus, and it would be difficult to deduce whether the mutational load in the prenatal samples provide clinically useful information regarding the postnatal and adult state.

Treatment of MTDNA Disorders

The polyploid nature of the mitochondrial genome, the inability to deliver therapeutic nucleic acids to the organelle through mitochondrial transfection, and the phenomenon of heteroplasmy coupled with the relative unavailability of useful preclinical experimental models have hampered progress in the development of curative treatments for mtDNA disease. One possible approach to "diluting" or even entirely eliminating the mutant mtDNA is applicable only in the earliest embryonic state and in effect represents a form of germ-line preventive therapy (**Fig. e18-8**). This would involve cytoplasmic transfer of normal mitochondria into the oocyte of a woman affected with the heteroplasmic mtDNA mutation or, more radically, the use of a normal donor oocyte that has been enucleated and into which the nucleus of the affected recipient mother has been introduced. These approaches have not met with widely reported clinical success. In a newer study, the nucleus (containing both maternal and paternal DNA with a few adherent mitochondria) from a severely abnormal human embryo was transplanted into a donated oocyte after removal of its nuclear DNA. The resulting embryo developed normally for the first 6 days, at which time it was set aside for nonmedical reasons. The long-term future feasibility of this approach is strengthened by similar mtDNA genetic replacement studies in ova of nonhuman primates that have been taken to successful birth of healthy offspring. The threshold effect, in which even a small increase in nonmutant mtDNA copies or limited amelioration of metabolic dysfunction might greatly improve the disease phenotype and benefit patient health, further bolsters hope that this approach may reduce the burden of human mtDNA disease in the future.

Figure e18-8

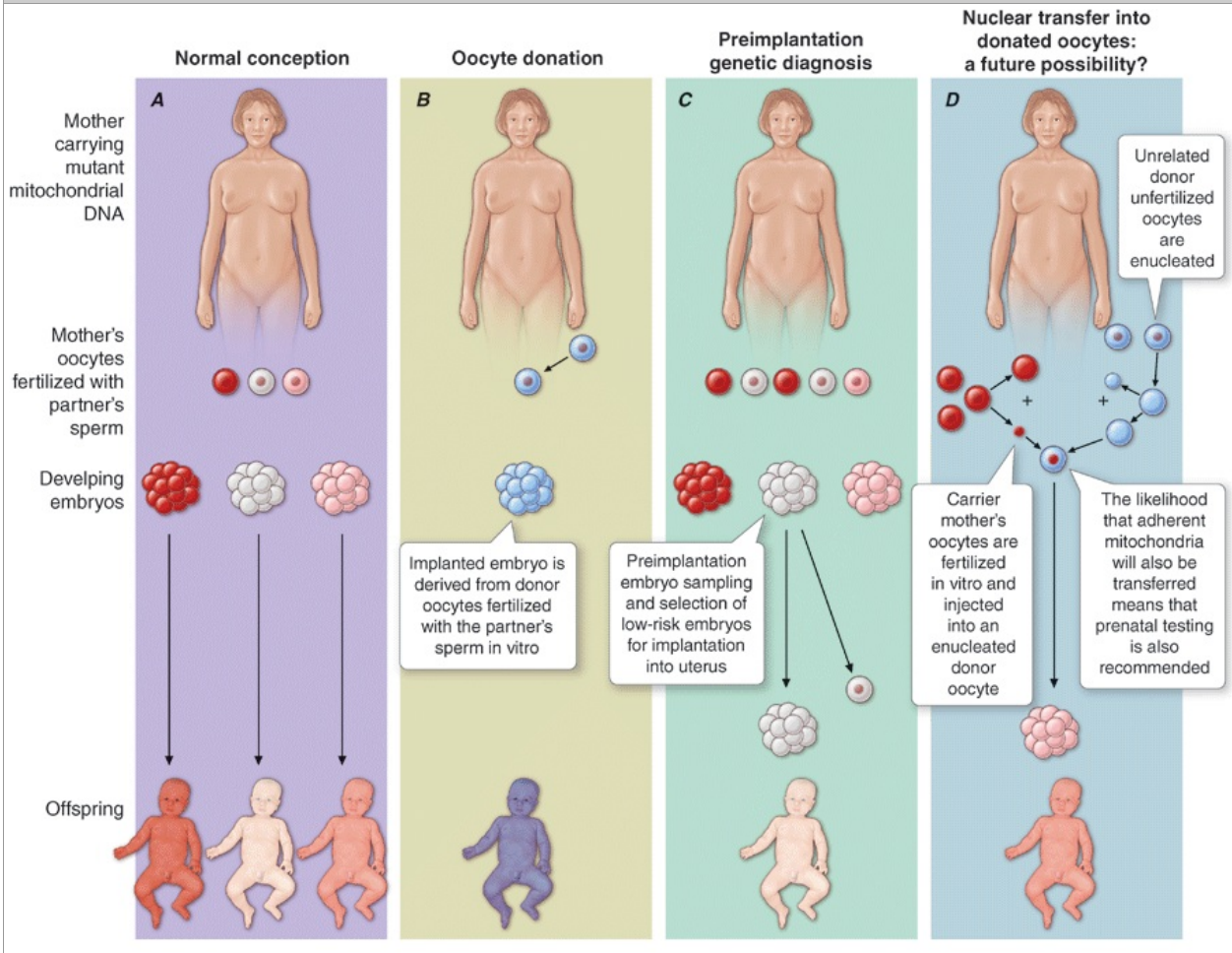

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Possible approaches for prevention of mtDNA disease. ***A***. No intervention: Offspring's mutant mitochondrial DNA load will vary greatly. ***B***. Oocyte donation: currently permitted in some constituencies but limited by the

DNA load will vary greatly. ***C***. Oocyte donation: currently permitted in some constituencies but limited by the availability of oocyte donors. ***C***. Preimplantation genetic diagnosis: available for some mtDNA diseases (reliable in determining background nuclear genomic haplotype risk). ***D***. Nuclear transfer: research stage, including initial studies in nonhuman primates. Red represents mutant mitochondrial DNA, pink and white represent successively higher proportions of normal mitochondrial DNA, and blue represents genetic material from an unrelated donor. *(Adapted with permission from Poulton et al.)*

Nevertheless, until such human genetic intervention studies reach the routine clinical arena, counteracting the most damaging metabolic changes currently represents the mainstay of treatment. The avoidance of well-characterized environmental risk factors as noted above, such as tobacco use in the case of LHON, and the avoidance of aminoglycosides in the case of mtDNA A1555G mutation have very major benefits. Less specific interventions in the case of other disorders involve combined treatment strategies and include dietary intervention and removal of toxic metabolites. Cofactors and vitamin supplements are used widely in the treatment of diseases of mitochondrial oxidative phosphorylation, although there is little evidence, apart from anecdotal reports, to support their use. This includes administration of artificial electron acceptors, including vitamin K_3, vitamin C, and ubiquinone (coenzyme Q_{10}); administration of cofactors (coenzymes), including riboflavin, carnitine, and creatine; and use of oxygen radical scavengers, such as vitamin E, copper, selenium, ubiquinone, and idebenone. Drugs that could interfere with mitochondrial defects, such as the anesthetic agent propofol, barbiturates, and high doses of valproate, should be avoided. Supplementation with the nitric oxide synthase substrate l-arginine has been advocated as a vasodilator treatment during strokelike episodes.

In the case of homoplasmic mtDNA variants that predispose to late-onset common complex disease, it is more realistic to think of using their identification in a specific patient as a nonmodifiable risk factor, which guides the aggressiveness of medical intervention for the associated modifiable risk factors for the same disorder. For example, the identification of a haplogroup-defining homoplasmic mtDNA mutation that confers added risk for metabolic syndrome should trigger intensive dietary, lifestyle, and medical intervention to reduce other factors that promote the metabolic syndrome and its complications. In the case of acquired somatic mutations—to the extent that a vicious cycle of ROS production with mtDNA mutation plays a role—effective antioxidant and ROS scavenging therapeutic strategies may prove to be of benefit.

FURTHER READINGS

Bandelt HJ et al: Exaggerated status of "novel" and "pathogenic" mtDNA sequence variants due to inadequate database searches. Hum Mutat 30:191, 2009[PMID: 18800376] [Full Text]

Bredenoord AL et al: PGD to reduce reproductive risk: The case of mitochondrial DNA disorders. Hum Reprod 23:2392, 2008[PMID: 18664474] [Full Text]

Calvo S et al: Systematic identification of human mitochondrial disease genes through integrative genomics. Nat Genet 38: 576, 2006[PMID: 16582907] [Full Text]

Craven L et al: Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease. Nature 465:82, 2010[PMID: 20393463] [Full Text]

Cree LM et al: A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. Nat Genet 40:249, 2008[PMID: 18223651] [Full Text]

Di Donato S: Multisystem manifestations of mitochondrial disorders. J Neurol 256:693, 2009

Dimauro S: Mitochondrial DNA medicine. Biosci Rep 27:5, 2007[PMID: 17484047] [Full Text]

Filosto M, Manusco M: Mitochondrial diseases: A nosological update. Acta Neurol Scand 115:211, 2007[PMID: 17376118] [Full Text]

Hudson G et al: Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. Am J Hum Genet 81:228, 2007[PMID: 17668373] [Full Text]

———: Mutation of OPA1 causes dominant optic atrophy with external ophthalmoplegia, ataxia, deafness and multiple mitochondrial DNA deletions: A novel disorder of mtDNA maintenance. Brain 131:329, 2008

Mancuso M et al: Mitochondrial DNA-related disorders: Natural selection shaped regional mtDNA variation in humans. Biosci Rep 27:31, 2007[PMID: 17484046] [Full Text]

McKenzie M et al: Mitochondrial disease: Mutations and mechanisms. Neurochem Res 29:589, 2004[PMID: 15038606] [Full Text]

Neimi AK, Majamaa K: Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. Eur J Hum Genet 13:965, 2005

Poulton J et al: Preventing transmission of maternally inherited mitochondrial DNA diseases. BMJ 338:b94, 2009

Salas A et al: A critical reassessment of the role of mitochondria in tumorigenesis. PLoS Med 2:1158, 2005

Spinazzola A, Zeviani M: Disorders from perturbations of nuclear-mitochondrial intergenomic cross-talk. J Intern Med 265:174, 2009[PMID: 19192035] [Full Text]

Suissa S et al: Ancient mtDNA genetic variants modulate mtDNA transcription and replication. PLoS Genet

5:e1000474, 2009

Wallace D: The mitochondrial genome in human adaptive radiation and disease: On the road to therapeutics and performance enhancement. Gene 354:169, 2005[PMID: 16024186] [Full Text]

Wilson FH et al: A cluster of metabolic defects caused by mutation in a mitochondrial tRNA. Science 306:1190, 2004[PMID: 15498972] [Full Text]

Zeviani M, Carelli V: Mitochondrial disorders. Curr Opin Neurol 20:564, 2007[PMID: 17885446] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE HUMAN MICROBIOME: INTRODUCTION

The words *well-being* and *microbes* typically are not spoken in the same breath. Microbes have a strong negative connotation in contemporary societies and are viewed in a warlike context. Disease-producing or pathogenic microbes are indeed serious threats to human health and have received justifiable attention from the inception of the field of microbiology. The list of known and notorious pathogens is long. However, most human encounters with microbes are not hostile but benign or even beneficial. Advances in DNA sequencing and computational biology now permit comprehensive description of the composition of and the roles played by the microbial communities (*microbiota*) associated with the human body.

Like all animals on the planet, humans have had to adapt to a microbe-dominated biosphere. The number of microbes on Earth is staggering. It has been estimated that 10^{30} microbes live in the ocean, in surface and subsurface terrestrial ecosystems, and on and inside animals and plants. Microbes, which are defined here as microscopic living organisms that belong to any of the three known domains of life on Earth (Bacteria, Archaea, and Eukarya), inhabit the exposed exterior and interior surfaces of the human body (e.g., the skin, mouth, airways, gastrointestinal tract, and vagina). The uterus traditionally has been thought to be microbe-free, although new evidence may prompt a reevaluation of the idea that this ecosystem is completely sterile throughout gestation. Colonization with microbes begins no later than during parturition; in the ensuing years, microbes come to outnumber human cells by an estimated tenfold in the human body. Therefore, a comprehensive view of humans as a life form entails consideration of the body's microbial and *Homo sapiens* cells together as a connected network'a coevolved symbiosis ("living together")'in which various body habitats serve as homes to microbial communities. These habitats harbor microbiota composed of members that function as mutualists (both host and microbe benefit from the other's presence), commensals (one partner benefits, and the other is seemingly unaffected), and potential or overt pathogens (one partner benefits, and the other is harmed).

Human Microbiome Projects

Human microbiome projects (HMPs) reflect this view of the human body as an amalgamation of human and microbial cells as well as human and microbial genes. These projects represent a confluence of ongoing technical and computational advances in the genome sciences. The newest generation of massively parallel DNA sequencers can be used to document—with unprecedented speed and economy—which microbes compose a microbiota and to characterize a microbiota's gene content (its *microbiome*). Key terms relevant to HMPs are defined in **Table 64-1**.

Table 64-1 Glossary of Terms Used in Discussion of the Human Microbiome

Term	Definition
Culture-independent analysis	A type of analysis that does not require culture of microbes. Information is extracted directly from environmental samples.
Diversity	The distribution of different kinds of organisms in a specific habitat or habitats. *Alpha* diversity is, broadly speaking, the number of kinds of organisms in a single sample; *beta* diversity describes how the types of organisms are partitioned among samples.
Domains of life	The three major branches of life on Earth: the Eukarya (including humans), the Bacteria, and the Archaea
Gnotobiotics	Rearing of animals under sterile (germ-free) conditions. Animals subsequently can be colonized at various stages of their life cycle with defined collections of microbes.
Human microbiome	In ecology, *biome* refers to a habitat and the organisms in it. In this sense, the human *microbiome* would be defined as the collection of microorganisms associated with the human body. However, the term *microbiome* is also used to refer to the collective genomes and genes present in members of a particular microbiota, and the human *metagenome* is the sum of the human genome and microbial genes (microbiome). A *core* human microbiome is defined as everything shared in a par-ticular body habitat among all or the vast majority of human microbiomes. A core microbiome may include a common set of genomes and genes encoding various protein families and/or metabolic capabilities. Microbial genes that are variably represented in different

	humans may contribute to distinctive physiologic/metabolic phenotypes.
Metagenomics	An emerging field encompassing culture-independent studies of the structures and functions of microbial communities and their interactions with the habitats they occupy. Metagenomics includes (1) shotgun sequencing of microbial DNA isolated directly from a particular environment and (2) high-throughput screening of expression libraries constructed from cloned community DNA to identify specific functions such as antibiotic resistance (*functional* metagenomics). DNA-level analyses provide the foundation for profiling of mRNAs and proteins produced by a microbiome (*meta-transcriptomics* and *meta-proteomics*) and for identification of a community's metabolic network (*meta-metabolomics*).
Microbiota	A microbial community, including Bacteria, Archaea, Eukarya, and viruses, that occupies a specific habitat
Pan-genome	The group of genes found in genomes that make up a particular microbial phylotype, including *core* genes found in all genomes and *dispensable* genes found in a subset of genomes within the phylotype
Phylogenetic analysis	Characterization of the evolutionary relationships between organisms and their gene products
Phylogenetic tree	A "tree" in which organisms are shown according to their relationships to hypothetical common ancestors. When built from molecular sequences, the branch lengths are proportional to the amount of evolutionary change separating each ancestor-descendant pair.
Phylotype	A phylogenetic group of microbes, currently defined by a threshold percentage identity shared among their small subunit rRNA genes (e.g., ≥97% for a species-level phylotype)
Rarefaction	A procedure in which subsampling is used to assess whether all the diversity present in a specific sample or set of samples has been observed at a specific sampling depth and to extrapolate how much additional sampling would be needed to observe all the diversity
Resilience	A community's ability to return to its initial state after a perturbation

Computing power and software tools are evolving rapidly to mine the vast amount of data generated by these sequencers. The coevolution of software development and ever-increasing data-set generation is not surprising. In systems ranging from artificial organisms to bacteria to metazoa, coevolution of predators and prey greatly accelerates the evolutionary rate. Dawkins and Krebs (1979) introduced the concept of an "arms race" in which the predator has a clear advantage in developing better means for consuming its prey and the prey has a clear advantage in developing better means to avoid being eaten. Adaptation to evade a population that is itself adapting proceeds at a far more rapid rate than does adaptation to an environment that can change but cannot evolve in an adaptive way. The same dynamics are apparent in software development: the "consumers" of software produce ever-larger data sets that break the software, in part because the availability of these resources prompts new strategies for exploitation (i.e., the design of experiments that could not have been conceived without the availability of the improved tools). This intimate association between tool users and tool developers is essential for rapid development in terms of what the software can accomplish and what experiments can be conceived and accomplished in HMPs. HMPs also reflect a more ecologic focus of microbiology on the properties and functions of microbial communities; that is, these projects go beyond the properties and functions that individual component species of these communities exhibit when studied in isolation—i.e., outside their native environments.

The ability to characterize the structures and functions of whole microbial communities without culturing their component members has spawned a new field of science known as *metagenomics* (Table 64-1). Metagenomics involves the sequencing of DNA isolated directly from a microbial community residing in a particular environment; the resulting information permits the application of other systems-level techniques, such as profiling of mRNA and protein products expressed by a microbiome and characterization of a community's metabolic activities. The goal of metagenomics (these allied systemwide characterizations) and of HMPs in general is an understanding of the ecologic principles and the various factors that determine how microbial communities are assembled, are maintained, and operate. The results promise to provide a deeper understanding of how habitats and microbial communities coevolve and, in the case of humans, how these communities vary both compositionally and functionally over time at different anatomic sites within an individual and in different groups of people living in different cultural contexts as well as how these communities contribute to human physiologic status, physiologic variation, disease predisposition, and disease pathogenesis.

Furthermore, HMPs address one of the most fundamental questions in genetics: How does environment influence the structure and function of "human genes"? Over a lifetime, each human encounters a unique environment. Part of this personally experienced environment is incorporated into the body habitat–associated microbiota. In this sense, HMPs will expand the conceptualization of "human" genetic potential from a relatively fixed deterministic view in which individuals are seen as

inheriting only a defined set of ~20,000 genes from their parents to a view in which each human acquires a microbiome containing a varied assemblage of genes several orders of magnitude larger than the collected *H. sapiens* genes through a process influenced by family, lifestyle, and life experiences as well as by *H. sapiens* genes. International HMPs probably will help determine whether there is a dimension of human evolution that is occurring at the level of the human microbiota and microbiome and—if so—whether, how, and how fast this microbial evolution may be affecting human biology. Finally, these projects probably will raise a number of important questions about personal identity, how to define the origins of health disparities, and issues related to privacy and confidentiality.

A Toolbox for Culture-Independent Metagenomic Analyses of Microbial Communities

Most members of complex microbial communities cannot be cultured by conventional laboratory techniques. The vast microbial diversity that exists inside and on the human body cannot be characterized with culture-based approaches, in large part because the metabolic milieu fashioned by these communities in their native habitats cannot be duplicated in vitro at this time. Therefore, investigators have turned to culture-independent methods to identify which organisms are present in a microbiota and in what abundance. The gene widely used to identify microorganisms and classify their evolutionary relationships encodes the major RNA component of the small subunit (SSU) of ribosomes. The SSU rRNA gene has been highly conserved among all known life forms on Earth. This conservation allows SSU rRNA genes from different organisms to be aligned accurately so that regions of nucleotide sequence variation can be identified readily. Pairwise comparisons of SSU rRNA gene sequences from different microbes allow the construction of a phylogenetic tree that represents an evolutionary map; previously unknown organisms can then be assigned a location (coordinate) on that map. This approach, known as *molecular phylogenetics*, allows each organism to be characterized on the basis of its evolutionary distance from other organisms. Different phylogenetic types (phylotypes) can be viewed as branches on an evolutionary tree.

The most straightforward way to define which microbes are present in microbial communities associated with the human body is to amplify SSU rRNA genes by polymerase chain reaction (PCR), using primers directed at regions with nucleotide sequences that are conserved among all bacteria (or archaea or eukaryota) and that flank more variable regions. These variable regions can be used to discriminate among different kinds of organisms belonging to each of the three domains of life. Because bacteria dominate human microbial communities, most efforts have been devoted to defining bacterial diversity in the microbiota.

The coevolution of high-throughput next-generation DNA sequencing with new software tools has led to an SSU rRNA renaissance that allows simultaneous characterization of the diversity present in hundreds to thousands of microbial communities. Sequencing of a small fragment of the 1500-bp (base pair) bacterial 16S rRNA gene has been found to be sufficient for many types of analysis. For example, 250-base reads encompassing a variable region of the gene are suitable for taxonomic assignments and microbial community comparisons provided that the region is chosen carefully. Primer design for PCR of bacterial 16S rRNA genes is a critical factor: differential annealing with different primer pairs designed to amplify different variable regions can lead to over- or underrepresentation of specific taxonomic groups (*taxa*), and different regions within the gene can have somewhat different patterns of evolution. Therefore, caution must be exercised in comparisons of the relative abundance of taxa in samples characterized in different studies using different methods.

A key innovation has been multiplex sequencing with highly parallel DNA sequencers that can generate large numbers of sequences with read lengths of ≥200 nucleotides. Amplicons generated from each microbial-community DNA sample are tagged by incorporation of a unique oligonucleotide barcode into the primer. Amplicons that harbor these sample-specific barcodes can be pooled together so that multiple samples representing multiple communities can be sequenced simultaneously (**Fig. 64-1**).

Figure 64-1

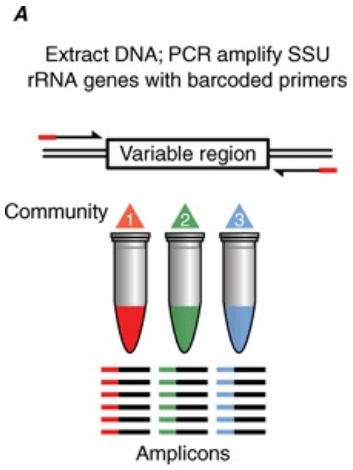

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

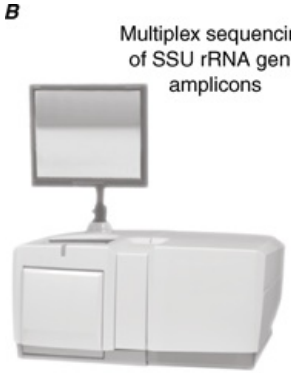

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

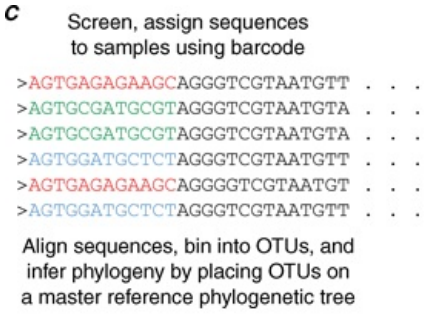

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

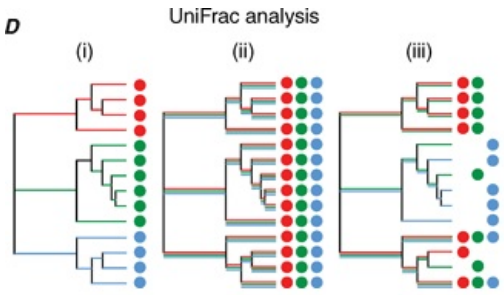

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

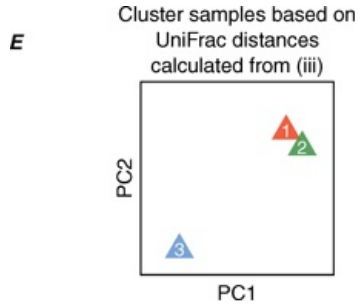

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pipeline for culture-independent studies of a microbiota. ***(A)*** DNA is extracted directly from a sampled human body habitat–associated microbial community. The precise location of the community and relevant patient meta-data are collected. Polymerase chain reaction (PCR) is used to amplify portions of the bacterial 16S rRNA gene containing one or more variable regions. Primers with sample-specific, error-correcting barcodes are designed to recognize the more conserved regions of the 16S rRNA gene that flank the targeted variable region(s). ***(B)*** Barcoded amplicons from multiple samples (communities 1–3) are pooled and sequenced in batches in a highly parallel next-generation DNA sequencer. ***(C)*** The resulting reads are processed. Barcodes denote which sample the sequence came from. After barcode sequences are removed in silico, reads are aligned and grouped according to a specified level of shared identity; e.g., sequences that share ≥97% nucleotide sequence identity are regarded as representing a species. Once reads are binned in this fashion, they are placed on a phylogenetic tree of all known bacteria to infer their phylogeny. ***(D)*** Communities can be compared to one another by either taxon-based methods, in which phylogeny is not considered and the number of shared taxa is simply scored, or phylogenetic methods, in which community similarity is considered in light of the evolutionary relationships of community members. The UniFrac metric is commonly used for phylogenetic-based comparisons. In the three stylized examples here, communities with varying degrees of similarity are shown. Each circle represents an operational taxonomic unit (OTU) which is colored according to its community of origin and placed on a master phylogenetic tree that includes all lineages from all communities. Branches (*horizontal lines*) are colored with each community that contains members from that branch. Examples (i), (ii), and (iii) vary in the amount of branch length shared between the OTUs from each community. In (i), there is no shared branch length, and the three communities have a similarity score of 0. In (ii), the communities are identical and are assigned a similarity score of 1. In (iii), there is an intermediate level of similarity. Communities represented in red and green share more branch length and thus have a higher similarity score than red versus blue or green versus blue. The amount of shared branch length in each pairwise community comparison provides a distance matrix. ***(E)*** The results of taxon- or phylogenetic-based distance matrices can be displayed by principal coordinates analysis (PCoA), in which each community is plotted spatially such that the largest component of variance is captured on the *x* axis (PC1) and the second largest component of variance is displayed on the *y* axis (PC2). In the example shown, the three communities in (iii) from panel D are compared. Note that for shotgun sequencing of whole-community DNA (microbiome analysis), reads are compared with genes present in the genomes of sequenced cultured microbes and/or with genes that have been annotated by hierarchical classification schemes in various databases, such as KEGG. Communities can then be compared according to the distribution of functional groups in their microbiomes (in a manner analogous to taxon-based methods for 16S rRNA-based comparisons) and the results plotted by PCoA.

One important choice in multiplex barcoded sequencing involves a trade-off between the number of samples that can be processed simultaneously and the number of sequences per sample, which in turn depends on the expected size of the differences between microbial communities. Differences in the microbiota between individuals or between communities occupying different body habitats in the same individual are large; therefore, relatively few (<1000) SSU rRNA reads are required for discrimination among communities. However, the identification of systematic differences in community ecology that correlate with physiologic or pathophysiologic status is confounded by this immense interpersonal variation. For example, 1000 sequences per sample means that species present at 1% abundance can be identified with reasonable confidence, although this level of coverage will fail to identify many of the rarer species-level phylotypes that may provide critical functions for using specific nutrients or for triggering immune responses that greatly affect other components of the microbiota.

The capacity to perform multiplex sequencing of bacterial 16S rRNA genes with highly parallel sequencers creates a problem: traditional tools for aligning sequences and defining taxonomic groups by their sequence similarity [a process known as *picking operational taxonomic units* (OTUs)] and traditional methods for phylogenetic analyses cannot handle the vast data sets involved. Emerging tools for performing large-scale alignments and large-scale taxon-based and phylogenetic analyses are starting to resolve this issue.

Both taxon-based methods (analyses based only on the OTUs present regardless of their evolutionary relationships) and phylogenetic methods (compositional analyses considered in light of the evolutionary relationships of community members) can be useful. The advantages of taxon-based methods are that they reveal directly which taxa contribute to the similarities and differences among samples (communities) and do not rely on possibly inaccurate tree reconstructions. The advantages of

phylogenetic methods are that, with the same input data, they use a more accurate picture of evolution to provide clearer results than do taxon-based methods. Unlike taxon-based methods, phylogenetic methods do not assume that all taxa are equally related to one another.

UniFrac, a commonly used phylogenetic method that compares the evolutionary history encompassed within different microbial communities, can be used to compare any two communities by noting the degree to which they share branch length on a master tree of microbial life: the more similar communities are to one another, the more branch length they will share (Fig. 64-1). A matrix of UniFrac-based measurements of distances between each pair of communities can be generated and the results graphed by principal coordinates analysis (PCoA), nonmetric multidimensional scaling (NMDS), or other geometric techniques that project a high-dimensional data set down onto a small number of dimensions that can be visualized and analyzed conveniently. The resulting dimensions show, in descending order, orthogonal contributions to variation in the full data set (Fig. 64-1).

Another level of analysis entails estimation of the richness, or diversity, of a microbial community by plotting the number of different types of SSU rRNA sequences at a specific phylogenetic level (e.g., species, genus) that are identified in a sample as a function of the number of sequences collected. Diversity estimates typically are based on rarefaction procedures, which assess how many species (or genera, etc.) would have been observed in a given sample if only 100, 200, 300, or more sequences had been collected.

Assembling a Bacterial 16S rRNA-Based Atlas of the Human Body

At several levels, humans are very much alike: their *H. sapiens* genomes are >99% identical, and they have similar collections of human cells. However, microbial communities differ drastically both between people and between habitats within a single human body. The variation is greatest between body sites; for example, the difference between the microbial communities residing in a person's mouth and those residing in that person's gut is comparable to the difference between the communities found in soil and in seawater. Even within a body site, the differences between people are not subtle: both gut and hand communities can differ by 80–90% at the bacterial species level, although the degree of variation in the mouth appears to be somewhat less (see below). The poet John Donne said that "no man is an island"; from a microbial perspective, however, each person consists of not just one isolated island but a whole archipelago of distinct habitats that exchange microbes with one another and with the "outside" at some undetermined level.

As with other ecosystems, human body habitat–associated microbial communities vary over time, and an understanding of this variation is probably essential to a functional understanding of the human microbiota. For example, studies of forests would prove puzzling without an understanding of the succession of events during which plant communities change systematically over time from weedy species colonizing fields to large mature trees. One exciting area opened up by high-throughput sequencing is the potential for tracking multiple body locations in multiple individuals over time, with direct visualization of the flow of microbes among different body habitats and different individuals in the presence or absence of various perturbations, such as antibiotic administration.

International HMPs must address a number of issues during the cataloging of microbial communities that inhabit humans. How many people have to be sampled (breadth) and how extensively (depth) to get a true measure of the extent of microbial diversity in humans? Where and how can investigators consistently sample a defined region of body habitat? In sampling relatively inaccessible microbial communities deep in the interior of the body, how can the risk to the donor be minimized? Defining the spatial features of microbial community structure in the mouth and the skin poses a particularly daunting challenge; for example, there is evidence that each tooth in an individual has a distinctive microbiota and that brushing produces a dramatic, immediate reduction in diversity. How often and over what period should a human being be sampled? What are the effects of gender? What is the impact of a person's relationship to family members who may or may not be sharing living space? What demographic factors should be evaluated (e.g., rural versus urban)? What is the impact of culture, lifestyle, health status, medications, and *H. sapiens* genotype?

Ecologists, who study the macroscopic world of plants and animals, have shown that the composition of a community depends on the order of initial entry of the component species. The same thing is probably true in the microbial world. Studies of bacterial diversity in fecal samples obtained from young adult female mono- and dizygotic twin pairs and their mothers over time have revealed that (1) the communities with the greatest degree of similarity are those derived from the same individual, (2) the similarity in the gut bacterial communities of adult monozygotic twin pairs is not significantly different in degree from the similarity of those of dizygotic twin pairs, and (3) fecal communities are more similar within family members than between members of different families. These results emphasize that early environmental exposures are a critical determinant of adult-gut microbial ecology. In humans, initial exposures depend on the mode of delivery. Babies sampled within 20 min of birth have relatively undifferentiated microbial communities in the mouth, the skin, and the gut. For vaginally delivered babies, these communities resemble the specific microbial communities found in the mother's vagina; for babies delivered by cesarean section, the communities resemble the mother's skin communities. The infant-gut microbiota changes to resemble the adult-gut community over the first 3 years of life

and may continue to change throughout life. The stages at which communities in other body habitats reach their highly differentiated adult forms have not been determined.

An initial integrated view of the spatial and temporal distribution of bacteria in the human body has been obtained from a survey of communities of microbiota occupying 27 sites in a few healthy unrelated men and women. The body habitats sampled (on a given day for 2 consecutive days on two occasions separated by 3 months) included the gut (feces), the oral cavity, the external auditory canals, the inside of the nares, and 18 distinct skin locations (**Fig. 64-2**). Across all body habitats, members of 22 bacterial phyla were detected, but the vast majority of sequences belonged to only four bacterial phyla: Actinobacteria (37%), Firmicutes (34%), Proteobacteria (12%), and Bacteroidetes (10%). UniFrac-based PCoA revealed strong primary clustering by body habitat; community composition was significantly less varied within than between habitats. Within habitats, the degree of variation was significantly less within individuals sampled over time than between individuals sampled on a specific day. Finally, after habitat and individual had been taken into account, the degree of variation was significantly less over 1 day than over 3 months. People's daily composite "whole-body" communities revealed perfect grouping by host and month, further emphasizing that the personal microbiota signature remains relatively stable over time.

Figure 64-2

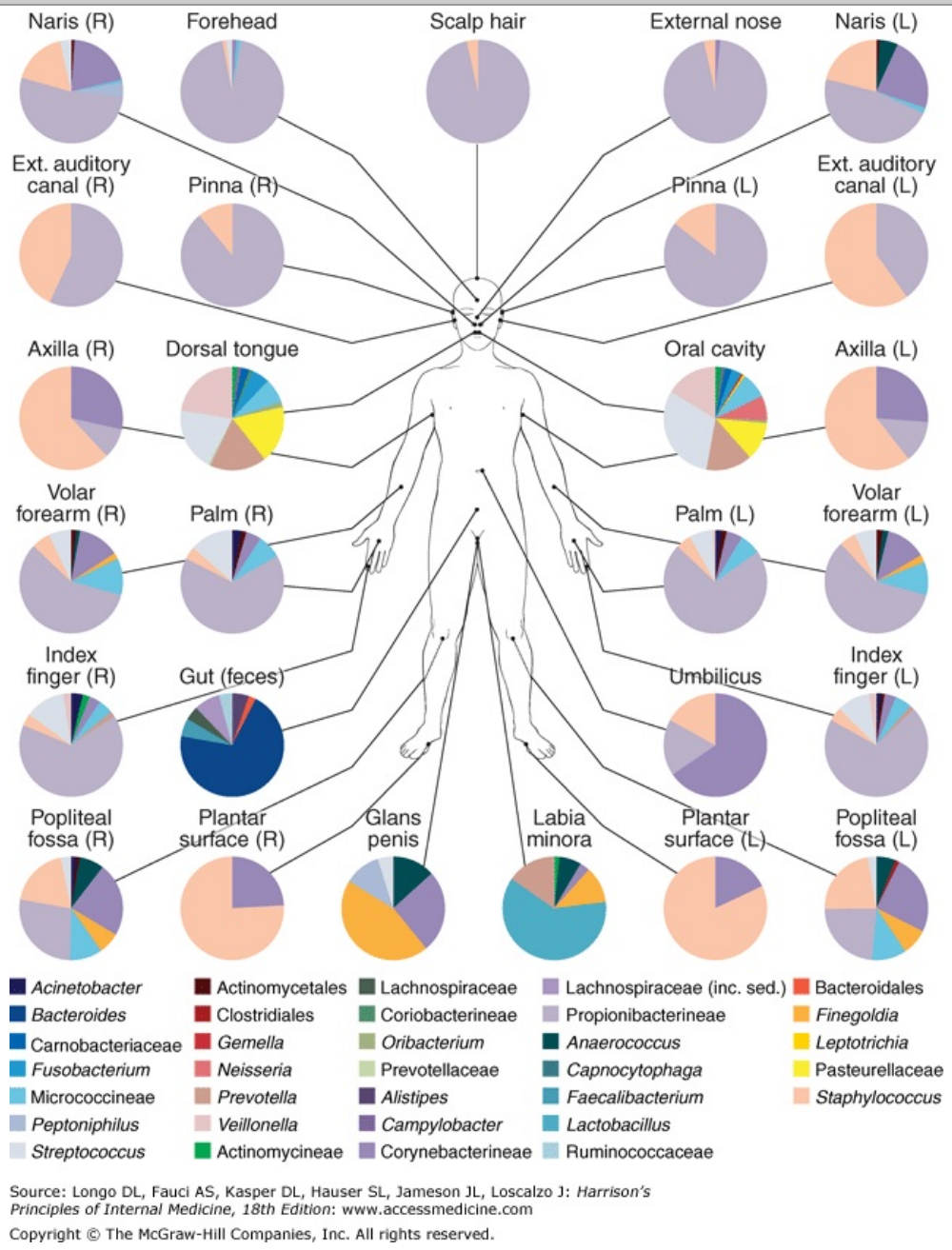

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Results of a 16S rRNA gene sequence–based study of bacterial diversity in 27 body habitats of nine unrelated healthy human adults. The relative proportions of major phylogenetic groups present in the various microbial communities are shown. Note that skin communities occupying

present in the various microbial communities are shown. Plots that skin communities occupying comparable positions on the right and left sides of the body are similar. *(Adapted from Costello et al.)*

Several surveys have emphasized that the skin harbors communities with predictable, albeit complex, biogeographic features (Fig. 64-2). To determine whether these differences reflect differences in local environmental factors, the history of the exposure of a specific site to microbes, or both, reciprocal microbiota transplantation has been performed in which microbial communities from one region of the skin are depleted by treatment with germicidal agents and the region (plot) is inoculated with a "foreign" microbiota harvested from different regions of the skin or from different body habitats from the same person or another individual. Community assembly at the site of transplantation subsequently is tracked over time. Remarkably, assembly proceeds differently at different sites: forearm plots receiving a tongue microbiota remained more similar to tongue communities than to native forearm communities in terms of their composition and diversity. However, forehead plots inoculated with tongue bacteria evolved to become more similar to native forehead communities. Thus, it appears that environmental factors operating at the forehead plot, in addition to the history of exposure to tongue bacteria, shaped community assembly.

These results underscore the need to specify body habitat when conducting microbial surveillance studies designed to examine the flow of normal and pathogenic organisms into and out of different body sites in inpatients and their health care providers. They also tie together several emerging themes from studies of human-associated microbial communities. Notably, there is a high level of *inter*personal variability in every body habitat studied to date. *Intra*personal variation in a specific body habitat is less marked, making longitudinal studies of microbial community ecology in a person before, during, or after a perturbation (e.g., dietary, pharmacologic) an attractive study option.

The resilience of human-associated microbial communities after perturbation has been addressed in several other contexts. One study showed that hand washing led to profound changes in the microbial community, greatly increasing diversity (presumably because of the preferential removal of high-abundance dominant phylotypes such as *Propionibacterium*). Within 6 h, the microbiota had rebounded to resemble the original hand communities. In adults, studies of a few individuals who took a 5-day course of ciprofloxacin showed that restoration of the fecal community after antibiotic administration took several months. Both the nature and the time course of restoration (reconfiguration) varied with the individual. Most previously present taxa returned, although in different proportions. The effects of antibiotics given during the first year of life on the assembly of the microbiota are ill defined. Are these effects transient or persistent? Is the diversity of the adult microbiota affected? Intriguingly, the "hygiene hypothesis" links the development of the human immune system to microbial exposures early in life. Increasing evidence indicates that fewer exposures and less conditioning are associated with an increased risk of allergic disorders such as asthma and various food sensitivities.

Bacterial 16S rRNA–based studies of the midvaginal microbiota in sexually active asymptomatic women have documented significant differences in community configurations among four self-reported ethnic groups (women of white, black, Hispanic, and Asian ancestry). Unlike most other body habitats surveyed, this ecosystem is dominated by a single genus, *Lactobacillus*. Four species of this genus together make up more than one-half the bacteria in most of these communities. Five community groups, designated I–V, have been defined. Four groups are dominated by *L. iners*, *L. crispatus*, *L. gasseri*, or *L. jensii,* whereas the fifth group includes proportionally fewer members of this genus and more anaerobes. The representation of these community categories was very distinctive within each of the four ethnic groups and correlated with vaginal pH and Nugent scores (a biomarker for bacterial vaginosis). Identification of the factors that determine the assembly of these distinct communities within and between ethnic groups as well as their resistance to or resilience after various physiologic or pathologic perturbations will require extensive longitudinal studies within individuals, including assessment of the impact of menstrual cycle, age, pregnancy, vaginal infections, and antibiotic use.

Who Else Is There? Other Branches of the Tree of Life Represented in the Human Body

Surveys based on SSU rRNA sequencing have focused largely on bacteria, yet the census of "who's there" in human body habitat–associated communities ultimately must take into account the other two domains of life (Archaea and Eukarya) as well as viruses. The representation of the four major Archaea phyla (Euryarchaeota, Crenarchaeota, Nanoarchaeota, and Korarchaeota) in human microbial communities and their contributions to community functions are largely ill defined, in part because of the difficulty in optimizing Archaea-directed primer design. Some archaeons are known to play key roles in community metabolism. For example, methane-producing archaeons (methanogens) make up as many as 10% of all anaerobes in the feces of some humans, yet archaeal diversity in the gut microbiota appears to be low, with *Methanobrevibacter smithii* being the dominant species. Gut methanogens improve the efficiency of dietary polysaccharide degradation and fermentation by preventing the buildup of hydrogen gas, an end product of fermentation. It is not a particularly rewarding job: the task of coupling H_2 oxidation with CO_2 reduction to produce methane (CH_4) is one of the lowest-energy-yielding reactions known in biology.

Culture-independent surveys of eukaryotic diversity in body habitat–associated communities of

microbiota have been very limited to date. Information about the representation of members of this domain of life has important implications for defining "normal"; for example, eukaryotic parasites are represented in millions of hosts living in various parts of the world and may be construed as a component of "normal" in these regions. Culture-independent analyses have relied on targeting eukaryotic SSU rRNA genes that encode 18S rRNA as well as the internal transcribed spacer (ITS) regions of fungal rRNA genes. Metagenomic studies of the fecal microbiota of a few healthy human adults indicate that the degree of eukaryotic diversity is lower than that of bacterial diversity, with prominent representation of members of the genus *Blastocystis*—obligate anaerobes with a wide host range whose role as parasites or pathogens in the human gut ecosystem is not clear. The fecal microbiota of these healthy humans also harbored other fungal genera (*Galactomyces, Paecilomyces*, and *Gloeotinia*). As with bacteria, culture-independent and culture-based surveys provide very different pictures of community composition, with *Candida* species appearing dominant among cultivable fecal eukaryotes.

Viruses are *the* major predators on this microbe-dominated planet, determining which microbial strains survive. Most genetic diversity on Earth is viral: viruses move DNA from microbial host to microbial host, harboring host-derived genes as they evolve. The current view is that there are ~10 virus-like particles (VLPs) per microbial cell in virtually all microbial communities. Sequencing of VLP DNA from purified feces from a group of monozygotic twins and their mothers disclosed that prophages and phages constitute the majority of the virome, with a majority of the phages belonging to the Podoviridae. One survey found that most viral diversity in the distal gut was novel (<20% of VLP sequencing reads exhibited homology to known viruses) and that VLP sequences represented <5% of all fecal microbiome sequences. Time-course studies of the fecal virome revealed that viral populations—like bacterial populations—are most similar within an individual; i.e., interpersonal variation is greater than intrapersonal variation. Unlike bacterial populations, in which community composition is more similar among family members than among unrelated individuals, interpersonal variations in VLP-associated viral populations are not appreciably less pronounced within than between families.

The Microbiome: Convergent Functions from Different Species Assemblages

Characterization of the microbiome by shotgun sequencing is important because, unlike SSU rRNA analysis, this method provides a direct readout of the genes that are available to perform particular functions in a specific community. The central issues are (1) to what degree variation in species-level assemblages occupying particular body habitats correlates with variation in community gene content and (2) whether groups of genes are represented in a particular body habitat–associated community in most or all individuals. The neutral theory of community assembly developed by macroecologists posits that most species in a community will have the same general niche (profession) or will adopt the broadest niche possible, endowing the community with functional redundancy. If applicable to the microbial world, neutral community assembly would predict a high level of variation in the types of microbial lineages that occupy a specific body habitat in different individuals (as defined by SSU rRNA sequencing), although the broad functions encoded in the microbiomes of these communities could be quite similar. In addition, chemical food webs are generated when the metabolic product(s) of one type of microbe become the substrate(s) for other microbes. These webs can be incredibly elaborate and can change as microbes adjust their patterns of gene expression and metabolism in response to alterations in nutrient availability. Thus, the sum of all activities of members of a microbial community can be viewed as an emergent rather than a fixed property (Chap. e19).

There are several key challenges in dealing with data obtained from shotgun sequencing of microbiomes. The first challenge is to attain a biomass sufficient for DNA recovery. Most human microbiome characterization has used fecal samples because they can be obtained readily in bulk quantities, 50% of the biomass of stool is microbial, and feces are an excellent proxy for depicting interpersonal differences in gut microbial ecology. At present, microbiome-level analyses typically are based on counts of reads assigned to specific taxa or functions. It is challenging to reconstruct metabolic pathways realistically for prediction of symbiotic, syntrophic, or antagonistic relationships among organisms. A number of databases are available for functional assignments; they employ various schemes for hierarchical classification. Unfortunately, the vast majority of these functional assignments are based on the very limited number of cultured organisms that have been subjected to direct experimental analyses.

Shotgun sequencing of the fecal microbiome has revealed that different microbial communities (species assemblages) converge on the same functional state. In other words, there is a group of microbial genes represented in the guts of unrelated as well as related individuals. This "core" microbiome is enriched in functions related to microbial survival in the gut (e.g., translation; nucleotide, carbohydrate, and amino acid metabolism) and in functions that benefit the host (nutrient and energy partitioning from the diet to microbes *and* host). Microbial genes whose proportional representation in gut communities varies among individuals and make up a "variable" microbiome. Pairwise comparisons have shown that family members have functionally more similar gut microbiomes than do unrelated individuals. Thus, intrafamilial transmission of a gut microbiome—which probably contains >100-fold more genes than the human genome—within a specific generation and across multiple generations could shape the biologic features of humans belonging to a kinship. This second dimension of human

gene flow (the other occurring at the level of *H. sapiens* genes) could modulate/mediate risks for a variety of pathogenic states.

The generalizability of the observation that vastly different species assemblages converge on similar gene functional repertoires in human body–associated habitats will have profound implications for an understanding of microbial functions in these different environments and the way these functions shape "human" physiologic and metabolic phenotypes. Issues such as the level of intra- and interpersonal variation in the gene content of the microbiome in various body habitats, the sampling depth required to characterize these differences, and the role of the *H. sapiens* genotype in shaping the microbiome genotype will be critical for informing this understanding.

Early Examples of Associations between the Human Microbiota and Disease States

HMPs probably will expand the scientific view of what constitutes a pathogenic microbe. Pathogens currently are viewed as discrete phylotypes that are able to elicit disease in susceptible hosts; this attribution is, in the best circumstances, based on satisfying Koch's postulates. A more ecologic view is that pathogens do not function in isolation; rather, their invasions and/or emergence as well as their effects on the host reflect interactions with other members of a microbiota. An even more expansive view is that a number of co-occurring organisms can "conspire" to produce pathogenic effects in certain host and environmental contexts, forming a "pathologic community."

The relationships of microbiota and microbiome composition and their function to human diseases are being investigated in many HMPs. The rationale for hypotheses invoking a disease–microbiota/microbiome relationship in many cases emanates from gnotobiotic mouse models (i.e., mice raised in germ-free environments—with no exposure to microbes—and then colonized at specific stages of life with different microbial communities). Germ-free animals are compared with conventionally raised animals of similar genotype or with germ-free mice that have received a microbiota from a conventionally raised mouse donor with a defined phenotype. These comparisons have shown that the gut microbiota plays a key role in the maturation of the innate as well as the adaptive components of the immune system, that the microbiota is a key trigger in the development of inflammatory bowel disease (IBD) in animals that harbor mutations in genes associated with disease risk in humans, and that surface components of certain members of the gut microbiota can specifically modify the activity of the immune system to mitigate or prevent IBD. The risk of development of type 1 diabetes in genetically susceptible mice is modified by the gut microbiota; this fact provides additional evidence for the role of this microbial community in the pathogenesis of certain types of autoimmune disorders.

Gnotobiotic mice provide an excellent system for controlling host genotype, microbial community composition, diet, and housing conditions. Microbial communities harvested from donor mice of defined genotypes and physiologic phenotypes can be used to determine the impact of these communities on formerly germ-free recipients as well as the effect of the recipient on the transplanted microbiota and its microbiome. In this respect, gnotobiotic mice provide an opportunity to combine comparative metagenomic studies of donor communities with functional assays of community properties. Studies of gnotobiotic mice also have revealed that the gut microbiota plays a role in regulating the efficiency of energy and nutrient harvest from the diet; it does so not only by processing otherwise indigestible components of the diet (e.g., polysaccharides) but also by regulating host genes that affect energy storage in adipocytes. The microbiota influences the rate of epithelial turnover in the gut (the rate is slower in germ-free animals), modulates the development of the elaborate microvasculature that underlies the epithelium (capillary network density is markedly reduced in adult germ-free animals but can be restored to normal levels within 2 weeks after gut microbiota transplantation), is a key determinant of whether radiation enteritis follows abdominal or whole-body irradiation (germ-free mice are resistant), and influences gut motility. The impact of the gut microbiota extends beyond the gastrointestinal tract. Heart weight, whether measured echocardiographically or as wet mass and normalized to tibial length or lean body weight, is reduced significantly in germ-free mice; this difference is eliminated within 2 weeks after colonization with a gut microbiota. The presence or absence of a gut microbiota influences certain aspects of behavior, including locomotor activity. This observation raises the question of whether in this coevolved relationship microbes have developed strategies for manipulating certain features of host behavior that are mutually beneficial. For example, it is intriguing that plasma levels of serotonin are several-fold higher in conventionally raised mice than in germ-free mice.

Initial metagenomic studies of relatively small numbers of humans have provided insights into bacterial communities associated with several diseases, offering, for example, new ways to classify bacterial vaginosis, identifying alterations in the cutaneous bacterial communities of patients with psoriatic lesions, demonstrating shifts in the representation of various bacterial taxa in IBD, and revealing differences in the composition of the microbiota and microbiome in obese versus lean individuals as well as before and after bariatric surgery. The challenge provided by such observations is not only to expand the level of sampling to different populations of individuals but also to determine whether these associations are causal or only side effects of other processes.

Ongoing studies are exploring the impact of the microbiota on nutritional status, the development of

Clostridium difficile colitis and its risk of relapse, the occurrence of necrotizing enterocolitis in premature newborns, the pathogenesis of diabetes, and various other metabolic phenotypes (*metabotypes*). These studies include assessment of the impact of the microbiota on drug metabolism. Comparisons of metabolites in the blood of germ-free and conventionally raised mice by nontargeted mass spectrometry have disclosed hundreds of compounds that are detectable in one but not the other type of animal and whose concentrations are affected markedly by the microbiota. These include compounds that are conjugated with sulfate, glycine, glucuronide, or other charged adducts that also modify xenobiotics, potentially altering the capacity of the host to metabolize these foreign molecules. Analogous events are being observed in humans. For example, metabolism of the analgesic acetaminophen to either sulfonated or glucuronidated forms is associated with predose levels of *p*-cresol, a microbial tyrosine metabolite that competes with acetaminophen for its sulfonate donor and the relevant sulfotransferase enzyme. Higher levels of *p*-cresol correlate with decreased sulfonation and increased glucuronidation of acetaminophen.

An additional series of metagenomic studies is exploring antibiotic resistance reservoirs in human microbial communities. To this end, a recent study of fecal samples used expression cloning: taking random fragments of the microbiome, placing them in expression vectors in a bacterial host, and screening for antibiotic resistance phenotypes. A related study demonstrated that bacteria subsisting on antibiotics are widely distributed in the environment. Capable of degrading a variety of antibiotics, these organisms are extensively drug resistant, phylogenetically diverse, and in many cases related to pathogens; they also harbor many resistance genes identical to those in clinical pathogens.

The role of microbes in therapeutics extends to clinical trials of the impact of probiotics—most recovered from fermented dairy products—on various forms of IBD [pouchitis that occurs in both ileoanal (pull-through) pouches and continent ileostomies or in recurrent *C. difficile* colitis]. A number of questions raised about probiotics typify the issues that surround studies of the role of the microbiota and microbiome in disease pathogenesis and therapeutics. Is there a consistent configuration of the microbiota definable in the study population that is associated with a particular disease state? The answer has diagnostic and mechanistic implications. How is the configuration affected by the intervention? Is there an identifiable reconfiguration? If so, how does it proceed? Are there many routes? Are the pathways related to the initial pretreatment state? If a reconfiguration does occur, is it sustained after cessation of treatment? How is host biology related to the configuration or reconfiguration? As with all studies involving human microbial ecology, the issue of what constitutes a suitable reference control is extremely important: the person himself or herself? family members? age- and sex-matched individuals living in the same locale with similar cultural traditions?

Impact of HMPs on Genetic Determinism, Personal Identity, Cultural Traditions, and Personalized Medicine

Sets of mono- and dizygotic twins and their family members will be an extremely valuable resource for initially teasing out relationships between environmental exposures, *H. sapiens* genotype, and human microbial ecology. Similarly, monozygotic twins discordant for certain disease states represent a powerful paradigm that enhances the ability to determine whether various diseases (e.g., asthma, IBD, metabolic disorders) can be linked to a person's microbiota and microbiome. It will be important to explore microbial ecology in groups of individuals living in developing countries that are undergoing rapid transformations in lifestyle and experiencing the emergence of a variety of Western diseases. Birth cohort studies (including studies of twins) that are initiated every 10 years in these countries may capture the impact of changes in lifestyle, including diet, on human microbial ecology.

Defining the human metagenome (the genes embedded in the *H. sapiens* genome plus the microbiome) will provide an entirely new level of refinement to the human description of self as well as a potential microbial legacy of personal lifestyle choices. Although this information may promote an appreciation of the origins of certain health disparities, care must be taken to avoid stigmatization of individuals or groups of individuals who have different cultural norms or express different behaviors. Metagenomics is a new field through which to view the influence of cultural traditions on lifestyles and choices that in turn affect human microbial ecology. Cultural anthropologists must examine the impact of this field on the ways in which study volunteers who live in various cultural settings view the natural world that envelops them and the ways this field and their cultural traditions interact to influence their perceptions of forces that affect their lives or their connections to one another within the context of a family or community. The union of metagenomics and cultural anthropology can help reveal how cultural traditions (e.g., the way infants are handled and cared for in early life) influence the flow of microbes between generations of humans, thus shaping the physiologic and genetic features of a kinship.

Finally, although microbiome-directed/related diagnostics and therapeutics would represent a new and different dimension of personalized medicine, sensitivity to the societal impact of this work is essential. This field is changing the human sense of personal identity, relationships with the world, and genetic determinism. Studies of human microbiomes promise to uncover new gene and protein families and probably will reveal novel biotransformations by microbial communities that could have commercial and therapeutic implications. Assessing the microbiome's capacity to metabolize orally administered drugs

may be highly instructive for the pharmaceutical industry as it investigates new and more accurate ways to predict drug bioavailability and toxicity. The chemical entities that human microbial communities synthesize to support mutually beneficial relationships with the host may become new classes of drugs; the human genes that these chemical entities target (manipulate) may represent new targets for drug discovery. Microbial strains harvested from the microbiota of people living in various parts of the world with varied diets and lifestyles could expand the repertoire of probiotic strains, including strains that could be added to food to enhance its nutrient value. Microbes, microbial genes, and microbial products may serve as new and valuable biomarkers of physiologic status and distinctive biological properties. Thus, a microbial ecology scan may become a standard component of regular health examinations or a tool for forensic scientists. All these speculations raise questions about how microbial strains and whole microbial communities obtained from volunteers should be archived and distributed as well as about who owns these reagents and discoveries emanating from them.

SUMMARY

HMPs are an important manifestation of progress in the genome sciences, a timely step in the quest to achieve a better understanding of the place of humans in the natural world, and a reflection of the evolving focus of twenty-first-century medicine on disease prevention, new definitions of health, new ways to determine the origins of individual biological differences, and new approaches to elucidate how changes in lifestyle and biosphere affect human biology.

FURTHER READINGS

Clayton TA et al: Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. Proc Natl Acad Sci USA 106:14728, 2009[PMID: 19667173] [Full Text]

Costello EK et al: Bacterial community variation in human body habitats across space and time. Science 326:1694, 2009[PMID: 19892944] [Full Text]

Dantas G et al: Bacteria subsisting on antibiotics. Science 320:100, 2008[PMID: 18388292] [Full Text]

Dawkins R, Krebs JR: Arms races between and within species. Proc R Soc Lond B Biol Sci 205:489, 1979[PMID: 42057] [Full Text]

Dethlefsen L et al: The pervasive effects of an antibiotic on the human gut microbiota. PLoS Biol 6:e280, 2008

Eckburg PB et al: Diversity of the human intestinal microbial flora. Science 308:1635, 2005[PMID: 15831718] [Full Text]

Fierer N et al: The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci USA 105:17994, 2008[PMID: 19004758] [Full Text]

Grice EA et al: Topographical and temporal diversity of the human skin microbiome. Science 324:1190, 2009[PMID: 19478181] [Full Text]

Human Microbiome Jumpstart Reference Strains Consortium: A catalog of reference genomes from the human microbiome. Science 328:994, 2010

Lee YK et al: Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyeitis. Proc Natl Acad Sci USA 108(Suppl 1):4615, 2011

Mazmanian SI et al: A microbial symbiosis factor prevents inflammatory bowel disease. Nature 453:620, 2008[PMID: 18509436] [Full Text]

Qin J et al: A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59, 2010[PMID: 20203603] [Full Text]

Ravel J et al: Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci USA 108(Suppl 1):4680, 2011

Reyes A et al: Viruses in the fecal microbiota of monozygotic twins and their mothers. Nature 466:334, 2010[PMID: 20631792] [Full Text]

Scanlan PD, Marchesi jr: Micro-eukaryotic diversity of the human distal gut microbiota: Qualitative assessment using culture-dependent and -independent analysis of faeces. ISME J 2:1183, 2008[PMID:

18670396] [Full Text]

Sommer MOA et al: Functional characterization of the antibiotic reservoir in the human microflora. Science 325:1128, 2009[PMID: 19713526] [Full Text]

Turnbaugh PJ et al: A core gut microbiome in obese and lean twins. Nature 457:480, 2009[PMID: 19043404] [Full Text]

Volkov I et al: Neutral theory and relative species abundance in ecology. Nature 424:1035, 2003[PMID: 12944964] [Full Text]

Wen L et al: Innate immunity and intestinal microbiota in the development of type 1 diabetes. Nature 255:1109, 2008

Wikoff WR et al: Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proc Natl Acad Sci USA 106:3698, 2009[PMID: 19234110] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SYSTEMS BIOLOGY IN HEALTH AND DISEASE: INTRODUCTION

The field of human biology has progressed over the last three centuries, largely as a result of the reductionist approach to the scientific problems that challenge the discipline. Biologists study the experimental response of a variable of interest in a cell or organism while holding all other variables constant. In this way, it is possible to dissect the individual components of a biologic system and assume that a thorough understanding of a specific component (e.g., an enzyme or a transcription factor) will provide sufficient insight to explain the global behavior of that system (e.g., a metabolic pathway or a gene network, respectively). Biologic systems are, however, much more complex and manifest behaviors that frequently (if not invariably) cannot be predicted from knowledge of their component parts characterized in isolation. Growing recognition of this shortcoming of conventional biologic research has led to the development of a new discipline, *systems biology,* that is defined as the holistic study of living organisms or their cellular or molecular network components to predict precisely their response to perturbations. Concepts of systems biology can be applied readily to human disease and therapy and define the field of *systems pathobiology,* in which genetic or environmental perturbations produce disease and drug perturbations restore normal system behavior.

Systems biology evolved from the field of systems engineering, in which a linked collection of component parts constitute a network whose output the engineer wishes to predict. The simple example of an electronic circuit can be used to illustrate some basic systems engineering concepts. All the individual elements of the circuit—resistors, capacitors, transistors—have well-defined properties that can be characterized precisely. However, they can be linked (wired or configured) in a variety of ways, each of which yields a circuit whose response to voltage applied across it is different from the response of every other configuration. To predict the circuit's (i.e., system's) behavior, the engineer must study its response to perturbation (e.g., voltage applied across it) holistically rather than its individual components' responses to that perturbation. Viewed another way, the resulting behavior of the system is greater than (or different from) the simple sum of its parts, and systems engineering utilizes rigorous mathematical approaches to predict these complex, often nonlinear, responses. By analogy to biologic systems, one can reason that detailed knowledge of a single enzyme in a metabolic pathway or of a single transcription factor in a gene network will not provide sufficient detail to predict the output of that metabolic pathway or transcriptional network, respectively. Only a systems-based approach will suffice.

It has taken biologists a long time to appreciate the importance of systems approaches to biomedical problems. Reductionism has reigned supreme for many decades, largely because it is experimentally and analytically simpler than holism, and because it has provided insights into biologic mechanisms and disease pathogenesis, and has led to successful therapies. However, reductionism cannot solve all biomedical problems. For example, the so-called off-target effects of new drugs that frequently limit their approval likely reflect the failure of a drug to be studied in holistic context, i.e., the failure to explore all possible actions aside from the principal target action for which it was developed. Other approaches to understanding biology therefore are clearly needed. With the growing body of genomic, proteomic, and metabolomic data sets in which dynamic changes in the expression of many genes and many metabolites are recorded after a perturbation and with the growth of rigorous mathematical approaches to analyzing those changes, the stage has been set for applying systems engineering principles to modern biology.

Physiologists historically have had more of a (bio)engineering perspective on the conduct of their studies and have been among the first systems biologists. Yet, with few exceptions, they, too, have focused on comparatively simple physiologic systems that are tractable using conventional reductionist approaches. Efforts at integrative modeling of human physiologic systems, as first attempted by Guyton for blood pressure regulation, represent one application of systems engineering to human biology. These dynamic physiologic models often focus on the acute response of a measurable physiologic parameter to a system perturbation, and do so from a classic analytic perspective in which all the conventional physiologic determinants of the output parameter are known and can be modeled quantitatively.

Until recently, molecular systems analysis has been limited owing to inadequate knowledge of the molecular determinants of a biologic system of interest. Although biochemists have approached metabolic pathways from a systems perspective for over 50 years, their efforts have been limited by the inadequacy of key information for each enzyme (K_M, k_{cat}, and concentration) and substrate (concentration) in the pathway. With increasingly rich molecular data sets available for systems-based analyses, including genomic, transcriptomic, proteomic, and metabolomic data, biochemists are now poised to use systems biology approaches to explore biologic and pathobiologic phenomena.

Properties of Complex Biologic Systems

To understand how best to apply the principles of systems biology to human biomedicine, it is necessary to review briefly the building blocks of any biologic system and the determinants of system complexity. All systems can be analyzed by defining their static topology (architecture) and their dynamic (i.e., time-dependent) response to perturbation. In the discussion that follows, system properties are described that derive from the consequences of topology (form) for dynamic response (function). Any system of interacting elements can be represented schematically as a network in which the individual elements are depicted as nodes and their connections are depicted as links. The nature of the links among nodes reflects the degree of complexity of the system. *Simple systems* are those in which the nodes are linearly linked with occasional feedback or feedforward loops modulating system throughput in highly predictable ways. By contrast, *complex systems* are nodes that are linked in more complicated, nonlinear networks; the behavior of these systems by definition is inherently more difficult to predict owing to the nature of the interacting links, the dependence of the system's behavior on its initial conditions, and the inability to measure the overall state of the system at any specific time with great precision. Complex systems can be depicted as a network of lower-complexity interacting components or modules, each of which can be reduced further to simpler analyzable canonical motifs (such as feedback and feedforward loops, or negative and positive autoregulation); however, a central property of complex systems is that simplifying their structures by identifying and characterizing the simpler substructures does not yield a predictable understanding of a system's behavior. Thus, the functioning system is greater than (or different from) the sum of its individual, tractable parts.

Defined in this way, most biologic systems are complex systems whose behaviors are not readily predictable from simple reductionist principles. The nodes, for example, can be metabolites that are linked by the enzymes that cause their transformations, transcription factors that are linked by the genes whose expression they influence, or proteins in an interaction network that are linked by cofactors that facilitate interactions or by thermodynamic forces that facilitate their biochemical

association. Biologic systems typically are organized as *scale-free*, rather than stochastic, networks of nodes. Scale-free systems are those in which a few nodes have many links to other nodes (highly linked nodes, or hubs) but most nodes have only a few links (weakly linked nodes). The term *scale-free* refers to the fact that the distribution of nodes in the network is not influenced by the magnitude or scale of the links considered. This is quite different from two other common network architectures: random (Poisson) and exponential distributions. Scale-free networks can be mathematically described by a power law that defines the probability of the number of links per node [$P(k) = k^{-(t)}$, where k is the number of links per node and is the slope of the log P(k) versus log(k) plot]; this unique property of most biologic networks is a reflection of their self-similarity or fractal nature (**Fig. e19-1**).

Figure e19-1

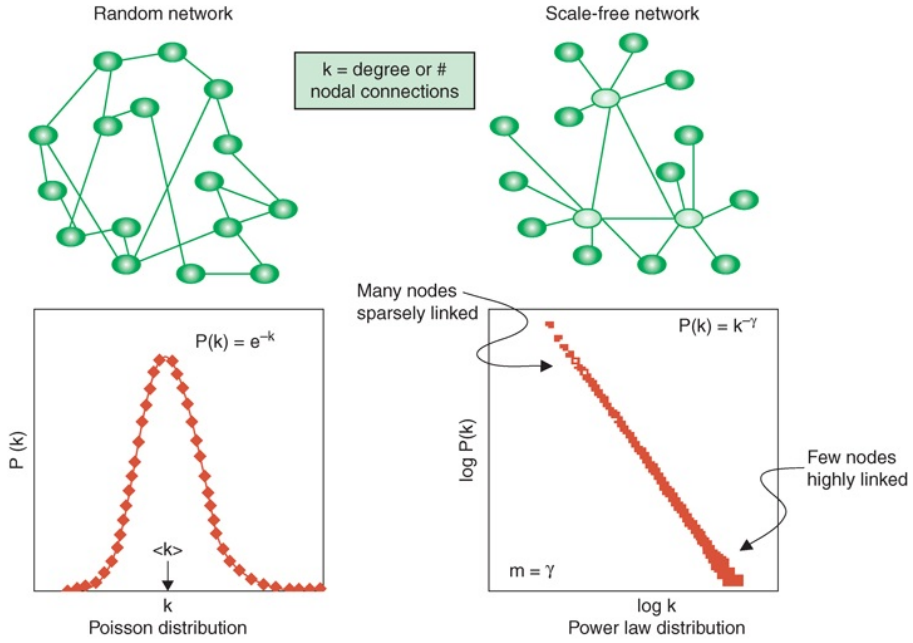

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Network representations and their distributions. A random network is depicted on the left, and its Poisson distribution of the number of nodal connections (k) is shown in the graph below it. A scale-free network is depicted on the right, and its power law distribution of the number of nodal connections (k) is shown in the graph below it. Highly connected nodes (hubs) are lightly shaded.

There are unique properties of scale-free biologic systems that reflect their evolution and promote their adaptability and survival. Biologic networks likely evolved one node at a time in a process in which new nodes are more likely to link to a highly connected node than to a sparsely connected node. Furthermore, scale-free networks can become sparsely linked to one another, yielding more complex, *modular scale-free topologies*. This evolutionary growth of biologic networks has three important properties that affect system function and survival. First, this scale-free addition of new nodes promotes *system redundancy*, which minimizes the consequences of errors and accommodates adverse perturbations to the system robustly with minimal effects on critical functions (unless the highly connected nodes are the focus of the perturbation). Second, this resulting network redundancy provides a survival advantage to the system. In complex gene networks, for example, mutations or polymorphisms in weakly linked genes account for biodiversity and biologic variability without disrupting the critical functions of the system; only mutations in highly linked (*essential*) genes (hubs) can shut down the system and cause embryonic lethality. Third, scale-free biologic systems facilitate the flow of information (e.g., metabolite flux) across the system compared with randomly organized biologic systems; this so-called "small-world" property of the system (in which the clustered nature of the highly linked hubs defines a local neighborhood within the network that communicates through weaker, less frequent links to other clusters) minimizes the energy cost for the dynamic action of the system (e.g., minimizes the transition time between states in a metabolic network).

These basic organizing principles of complex biologic systems lead to three unique properties that require emphasis. First, biologic systems are *robust*, which means that they are quite stable in response to most changes in external conditions or internal modification. Second, a corollary to the property of robustness is that complex biologic systems are *sloppy*, which means that they are insensitive to changes in external conditions or internal modification except under certain uncommon conditions (i.e., when a hub is involved in the change). Third, complex biologic systems exhibit *emergent properties*, which means that they manifest behaviors that cannot be predicted from the reductionist principles used to characterize their component parts. Examples of emergent behavior in biologic systems include spontaneous, self-sustained oscillations in glycolysis; spiral and scroll waves of depolarization in cardiac tissue that cause reentrant arrhythmias; and self-organizing patterns in biochemical systems governed by diffusion and chemical reaction.

Applications of Systems Biology to Pathobiology

The principles of systems biology have been applied to complex pathologic processes with some early successes. The key to these applications is the identification of emergent properties of the system under study in order to define novel, otherwise unpredictable (i.e., from the reductionist perspective) methods for regulating the system's response. Systems biology approaches have been used to characterize epidemics and ways to control them, taking advantage of the scale-free properties of the network of infected individuals that constitute the epidemic. Through the use of a systems analysis of a neural protein-protein interaction network, unique disease-modifying proteins have been identified that are common to a wide range of cerebellar neurodegenerative disorders that cause inherited ataxias. Systems biology models have been used to dissect the dynamics of the inflammatory response using oscillatory changes in the transcription factor NF-kB as the system output. Systems biology principles also have been used to predict the development of an idiotypy–anti-idiotypy antibody network, describe the dynamics of species growth in microbial biofilms, and analyze the innate immune response. In each of these

examples, a systems (patho)biology approach provided insights into the behavior of these complex systems that could not have been recognized with conventional scientific reductionism.

A unique application of systems biology to biomedicine is in the area of drug development. Conventional drug development involves identifying a potential target protein and then designing or screening compounds to identify those that inhibit the function of that target. This reductionist analysis has identified many potential drug targets and drugs, yet only when a drug is tested in animal models or humans are the systems consequences of the drug's action apparent; not uncommonly, so-called off-target effects may become apparent and be sufficiently adverse for researchers to cease development of the agent. A good example of this problem is the unexpected outcomes of the vitamin B–based regimens for lowering homocysteine levels. In these trials, plasma homocysteine levels were reduced effectively; however, there was no effect of this reduction on clinical vascular endpoints. One explanation for this outcome is that one of the B vitamins in the regimen, folate, has a panoply of effects on cell proliferation and metabolism that probably offset its homocysteine-lowering benefits, promoting progressive atherosclerotic plaque growth and its consequences for clinical events. In addition to these types of unexpected outcomes exerted through pathways that were not considered ab initio, conventional approaches to drug development typically do not take into consideration the possibility of emergent behaviors of the organism or the metabolic pathway or the transcriptional network of interest. Thus, a systems-based analysis of potential drugs (drug-target network analysis) can benefit the development paradigm both by enhancing the likelihood that a compound of interest will not manifest unforeseen adverse effects and by promoting novel analytic methods for identifying unique control points in metabolic or genetic networks that would benefit from drug-based modulation.

Systems Pathobiology and Human Disease Classification

Perhaps most important, systems pathobiology can be used to revise and refine the definition of human disease. The classification of human disease used in this and all medical textbooks derives from the correlation between pathologic analysis and clinical syndromes that began in the nineteenth century. Although this approach has been very successful, serving as the basis for the development of many effective therapies, it has major shortcomings. Those shortcomings include a lack of sensitivity in defining preclinical disease, a primary focus on overtly manifest disease, failure to recognize different and potentially differentiable causes of common late-stage pathophenotypes, and a limited ability to incorporate the growing body of molecular and genetic determinants of pathophenotype into the conventional classification scheme.

Two examples will illustrate the weakness of simple correlation analyses grounded in the reductionist principle of simplification (Occam's razor) in defining human disease. Sickle cell anemia, the "classic" Mendelian disorder, is caused by a Val6Gln substitution in the β chain of hemoglobin. If conventional genetic teaching holds, this single mutation should lead to a single phenotype in patients who harbor it (genotype-phenotype correlation). This assumption is, however, false, as patients with sickle cell disease manifest a variety of pathophenotypes, including hemolytic anemia, stroke, acute chest syndrome, boney infarction, and painful crisis, as well as an overtly normal phenotype. The reasons for these different phenotypic presentations include the presence of disease-modifying genes or gene products (e.g., hemoglobin F, hemoglobin C, glucose-6-phosphate dehydrogenase), exposure to adverse environmental factors (e.g., hypoxia, dehydration), and the genetic and environmental determinants of common intermediate pathophenotypes (i.e., variations in those generic pathologic mechanisms underlying all human disease—inflammation, thrombosis/hemorrhage, fibrosis, cell proliferation, apoptosis/necrosis, immune response).

A second example of note is familial pulmonary arterial hypertension. This disorder is associated with 50 different mutations in three members of the transforming growth factor β (TGF-β) superfamily: bone morphogenetic protein receptor-2 (BMPR-2), activin receptor-like kinase-1 (Alk-1), and endoglin. All these different genotypes are associated with a common pathophenotype, and each leads to that pathophenotype by molecular mechanisms that range from haploinsufficiency to dominant negative effects. As only approximately one-fourth of individuals in families that harbor these mutations manifest the pathophenotype, other disease-modifying genes (e.g., the serotonin receptor 5-HT2B, the serotonin transporter 5-HTT), genomic and environmental determinants of common intermediate pathophenotypes, and environmental exposures [e.g., hypoxia, infective agents (HIV), anorexigens] probably account for the incomplete penetrance of the disorder.

On the basis of these and many other related examples, one can approach human disease from a systems pathobiology perspective in which each "disease" can be depicted as a network that includes the following modules: the primary disease-determining elements of the genome (or proteome, if posttranslationally modified), the disease-modifying elements of the genome or proteome, environmental determinants, and genomic and environmental determinants of the generic intermediate pathophenotypes. **Figure e19-2** graphically depicts these genotype-phenotype relationships for the six common disease types with specific examples for each type. **Figure e19-3** shows a network-based depiction of sickle cell disease using this kind of modular approach.

Figure e19-2

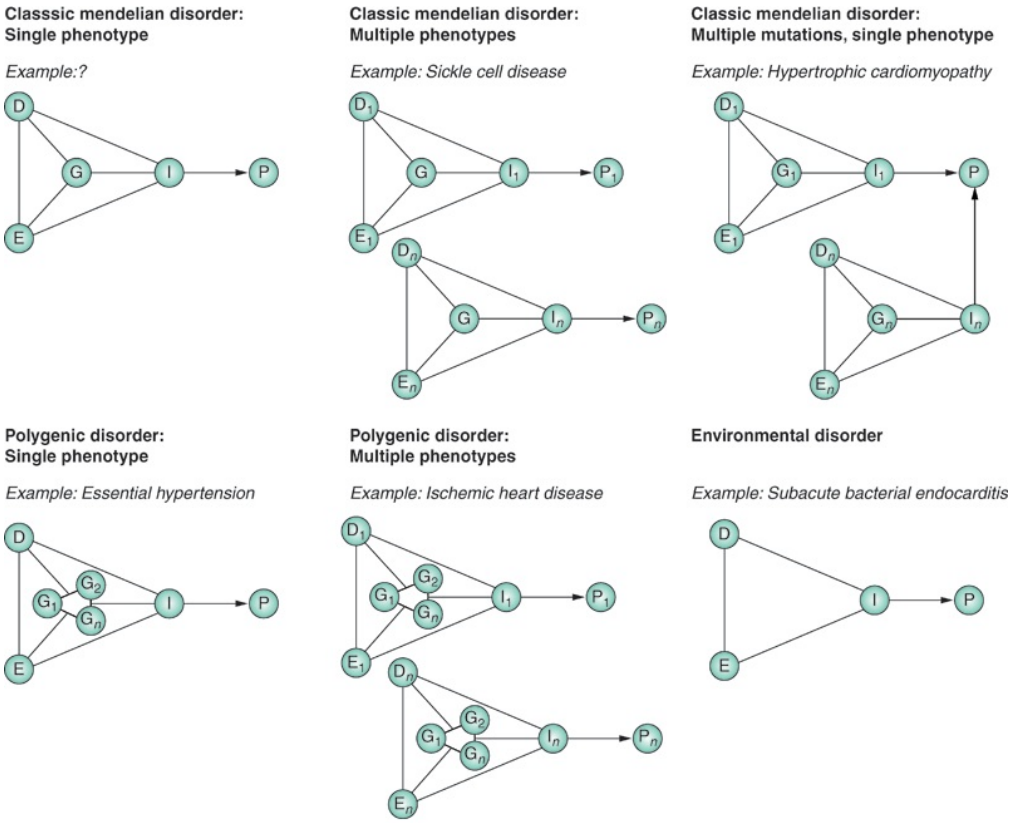

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examples of modular representations of human disease. G, primary human disease genome or proteome; D, secondary human disease genome or proteome; E, environmental determinants; I, intermediate phenotype; P, pathophenotype. *(Reproduced with permission from Loscalzo et al.)*

Figure e19-3

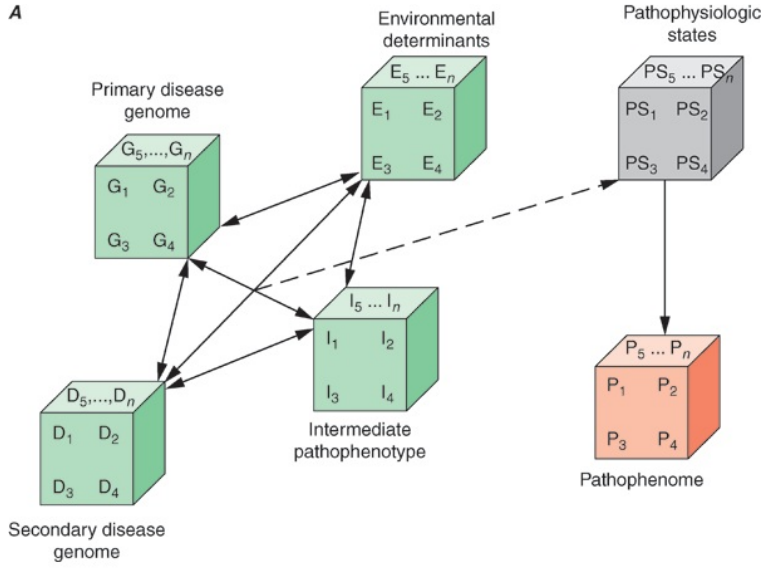

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

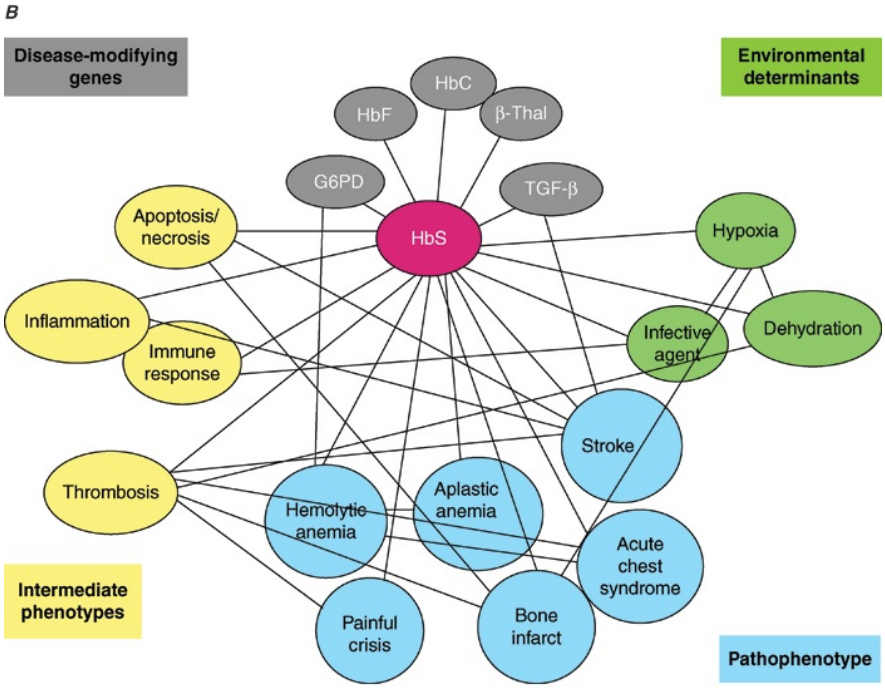

A. **Theoretical human disease network illustrating the relationships among genetic and environmental determinants of the pathophenotypes**. Key: G, primary disease genome or proteome; D, secondary disease genome or proteome; I, intermediate phenotype; E, environmental determinants; PS, pathophysiologic states leading to P, pathophenotype. ***B***. Example of this theoretical construct applied to sickle cell disease. Key: Red, primary molecular abnormality; gray, disease-modifying genes; yellow, intermediate phenotypes; green, environmental determinants; blue, pathophenotypes. *(Reproduced with permission from Loscalzo et al.)*

Goh and colleagues developed the concept of a human disease network (**Fig. e19-4**) in which they used a systems approach to characterize the disease-gene associations listed in the Online Mendelian Inheritance in Man database. Their analysis showed that genes linked to similar disorders are more likely to have products that associate and greater similarity between their transcription profiles than do genes not associated with similar disorders. In addition, proteins associated with the same pathophenotype are significantly more likely to interact with one another than with other proteins not associated with the pathophenotype. Finally, these authors showed that the great majority of disease-associated genes are not highly connected genes (i.e., not hubs) and are typically weakly linked nodes within the functional periphery of the network in which they operate.

Figure e19-4

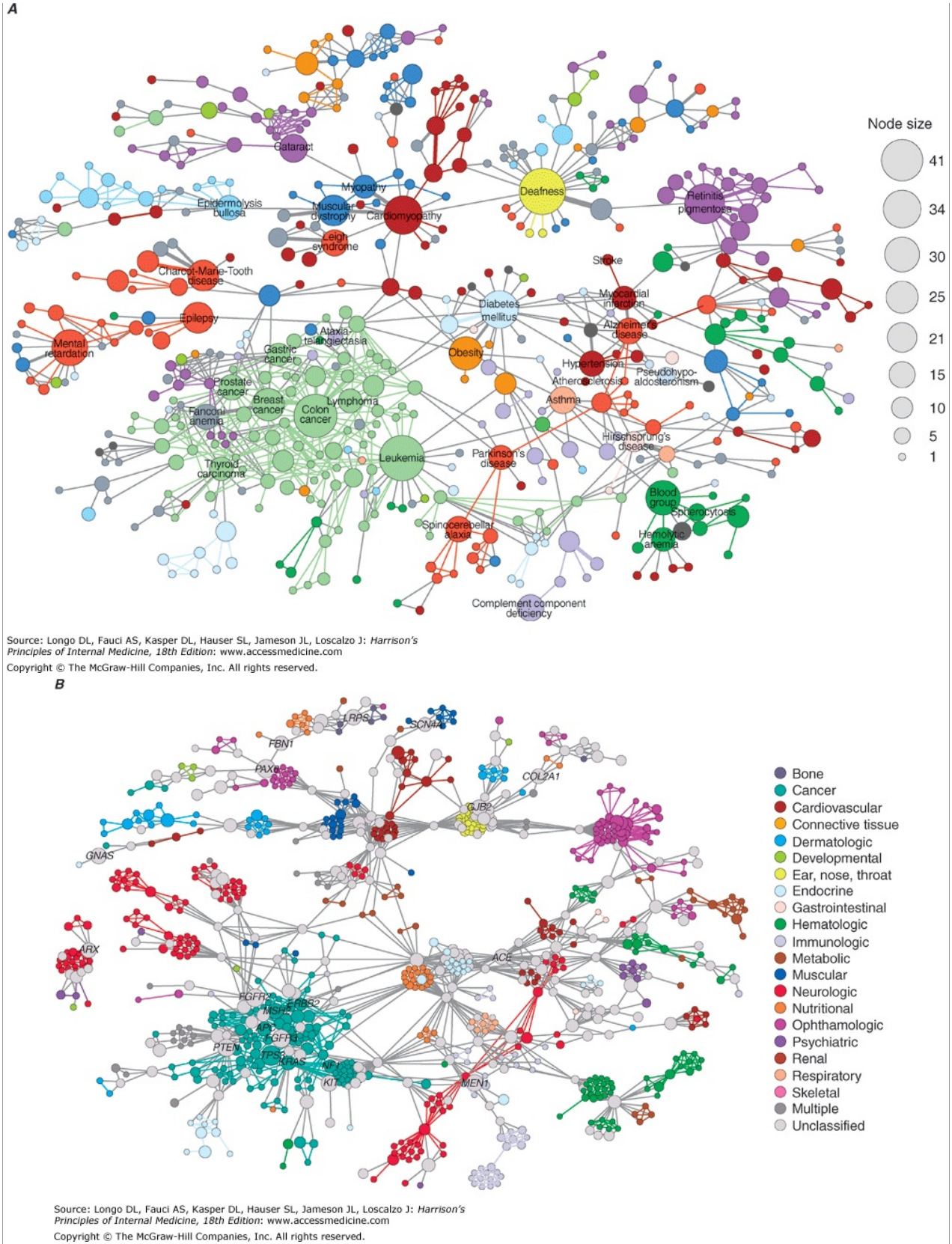

A. **Human disease network**. Each node corresponds to a specific disorder colored by class (22 classes, shown in the key to ***B***). The size of each node is proportional to the number of genes contributing to the disorder. Edges between disorders in the same disorder class are colored with the same (lighter) color, and edges connecting different disorder classes are colored gray, with the thickness of the edge proportional to the number of genes shared by the disorders connected by it. ***B***. Disease gene network. Each node is a single gene, and any two genes are connected if implicated in the same disorder. In this network map, the size of each node is proportional to the number of specific disorders in which the gene is implicated. *(From Goh et al. Reproduced with permission from the National Academies Press.)*

This type of analysis validates the potential importance of defining disease on the basis of its systems pathobiologic determinants. Clearly, doing this will require a more careful dissection of the molecular elements in the relevant pathways (i.e., more precise molecular pathophenotyping), less reliance on overt manifestations of disease for their classification, and an understanding of the dynamics (not just the static architecture) of the pathobiologic networks that underlie pathophenotypes defined in this way.

As yet another potential consideration, one can argue that disease reflects the later-stage consequences of the predilection of an organ system to manifest a particular intermediate pathophenotype in response to injury. This paradigm reflects a reverse causality view in which a disease is defined as a tendency to heightened inflammation, thrombosis, or fibrosis after an injurious perturbation. Where the process is manifest (i.e., the organ in which it occurs) is less important than that it occurs (with the exception of the organ-specific pathophysiologic consequences that may require acute attention). For example, from this perspective, acute myocardial infarction (AMI) and its consequences are a reflection of thrombosis (in the coronary artery), inflammation (in the acutely injured myocardium), and fibrosis (at the site or sites of cardiomyocyte death). In effect, the major therapies for AMI address these intermediate pathophenotypes (e.g., antithrombotics, statins) rather than any organ-specific disease-determining process. This paradigm would argue for a systems-based analysis that would first identify the intermediate pathophenotypes to which a person is predisposed, then determine how and when to intervene to attenuate that adverse predisposition, and finally limit the likelihood that a major organ-specific event will occur. Evidence for the validity of this approach is found in the work of Rzhetsky and colleagues, who reviewed 1.5 million patient records and 161 diseases and found that these disease phenotypes form a network of strong pairwise correlations. This result is consistent with the notion that underlying genetic predispositions to intermediate pathophenotypes form the predicate basis for conventionally defined end organ diseases.

Regardless of the specific nature of the systems pathobiologic approach used, these analyses will lead to a drastic revision of the way human disease is defined and treated. This will be a lengthy and complicated process but ultimately will lead to better disease prevention and therapy and probably do so from an increasingly personalized perspective. The analysis of pathobiology from a systems-based perspective is likely to help define specific subsets of patients more likely to respond to particular interventions based on shared disease mechanisms. This approach is being applied to certain conditions, for example, the responsiveness of lung cancer patients with mutations in the epidermal growth factor receptor (EGFR) to erlotinib, an agent that targets EGFR. Although it is unlikely that the extreme of "individualized medicine" will ever be practical (or even desirable), complex diseases can be mechanistically subclassified and interventions may be tailored to those settings in which they are more likely to work.

FURTHER READINGS

Barabási AL et al: Network medicine: A Network-based approach to human disease. Net Rev Genet 12:56, 2011[PMID: 18184624] [Full Text]

Brynildsen MP, Collins JJ: Systems biology makes it personal. Mol Cell 34:137, 2009[PMID: 19394290] [Full Text]

Burggren WW, Monticino MG: Assessing physiological complexity. J Exp Biol 208:3221, 2005[PMID: 16109885] [Full Text]

Daniels BC et al: Sloppiness, robustness, and evolvability in systems biology. Curr Opin Biotechnol 19:389, 2008[PMID: 18620054] [Full Text]

de Boer RJ, Perelson AS: Size and connectivity as emergent properties of a developing immune network. J Theor Biol 149:381, 1991

Dezso Z, Barabasi AL: Halting viruses in scale-free networks. Phys Rev E Stat Nonlin Soft Matter Phys 65:055103 (e-published), 2002

Goh KI et al: The human disease network. Proc Natl Acad Sci USA 104:8685, 2007[PMID: 17502601] [Full Text]

Guyton AC et al: Systems analysis of arterial pressure regulation and hypertension. Ann Biomed Eng 1: 254, 1972[PMID: 4358506] [Full Text]

Hart E et al: How affinity influences tolerance in an idiotypic network. J Theor Biol 249:422, 2007[PMID: 17904580] [Full Text]

He X, Zhang J: Why do hubs tend to be essential in protein networks? PLoS Genet 2:e88, 2006

Lim J et al: A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 126:801, 2006

Loscalzo J: Homocysteine trials—clear outcomes for complex reasons. N Engl J Med 354:1629, 2006[PMID: 16531615] [Full Text]

——— et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 3:1, 2007

Pastor-Satorras R, Vespignani A: Epidemic spreading in scale-free networks. Phys Rev Lett 86:3200, 2001[PMID: 11290142] [Full Text]

Rzhetsky A et al: Probing molecular overlap among complex human phenotypes. Proc Natl Acad Sci USA 104:11694, 2007[PMID: 17609372] [Full Text]

Schadt EE et al: A network view of disease and compound screening. Nature 8:286, 2009[PMID: 19337271] [Full Text]

Weiss JN et al: Network perspectives of cardiovascular metabolism. J Lipid Res 47:2355, 2006[PMID: 16946414] [Full Text]

Xavier JB, Foster KR: Cooperation and conflict in microbial biofilms. Proc Natl Acad Sci USA 104:876, 2007[PMID: 17210916] [Full Text]

Yildirim MA et al: Drug-target network. Nat Biotechnol 25:1119, 2007[PMID: 17921997] [Full Text]

Yue H et al: Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: A case study of an NK-kB signaling pathway. Mol Biosyst 2:640, 2006[PMID: 17216045] [Full Text]

Zak DE, Aderem A: Systems biology of innate immunity. Immunol Rev 227:264, 2009[PMID: 19120490] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

STEM CELL BIOLOGY: INTRODUCTION

Stem cell biology is a rapidly expanding field that explores the characteristics and possible clinical applications of a variety of stem cells that serve as the progenitors of more differentiated cell types. In addition to potential therapeutic applications (Chap. 67), patient-derived stem cells can also be used as disease models and a means to test drug effectiveness. Stem cells and their niche are becoming a major focus of medical research because they play central roles in tissue and organ homeostasis and repair, which are important aspects of aging and disease.

Identification, Isolation, and Derivation of Stem Cells

RESIDENT STEM CELLS

The definition of stem cells remains elusive. Stem cells were originally postulated as *unspecified* or *undifferentiated* cells that provide a source of renewal of skin, intestine, and blood cells throughout life. These *resident stem cells* have been identified in a variety of organs (e.g., epithelia of the skin and digestive system, bone marrow, blood vessels, brain, skeletal muscle, liver, testis, and pancreas) based on their specific locations, morphology, and biochemical markers.

ISOLATED STEM CELLS

Unequivocal identification of stem cells requires their separation and purification, usually based on a combination of specific cell-surface markers. These *isolated stem cells* [e.g., hematopoietic stem (HS) cells] can be studied in detail and used in clinical applications, such as bone marrow transplantation (Chap. 66). However, the lack of specific cell-surface markers for other types of stem cells has made it difficult to isolate them in large quantities. This challenge has been partially addressed in animal models by genetically marking different cell types with green-fluorescence protein driven by cell-specific promoters. Alternatively, putative stem cells have been isolated from a variety of tissues as side population (SP) cells using fluorescence-activated cell sorting after staining with Hoechst 33342 dye.

CULTURED STEM CELLS

It is desirable to culture and expand stem cells in vitro to obtain a sufficient quantity for analysis and potential therapeutic use. Although the derivation of stem cells in vitro has been a major obstacle in stem cell biology, the number and types of *cultured stem cells* have increased progressively **(Table 65–1)**. Cultured stem cells derived from resident stem cells are often called *adult stem cells* to distinguish them from *embryonic stem* (ES) and *embryonic germ* (EG) *cells*. However, considering the existence of embryo-derived, tissue-specific stem cells [e.g., trophoblast stem (TS) cells] and the possible derivation of similar cells from an embryo/fetus [e.g., neural stem (NS) cells], it is more appropriate to use the term, *tissue stem cells*.

Table 65–1 Classification of Cultured Stem Cells

Name	Source	Properties
Embryonic stem cells (ES, ESC)	Blastocysts or immuno-surgically isolated inner cell mass (ICM) from blastocysts.	ES cells grow as tightly adherent multicellular colonies with a population doubling time of ~12 h, maintain a stable euploid karyotype even after extensive culture and manipulation, can differentiate into a variety of cell types *in vitro*, and can contribute to all cell types, including functional sperm and oocytes, when injected into a blastocyst (m). ES cells form relatively flat, compact colonies with a population doubling time of 35–40 h (h).
Embryonic germ cells (EG, EGC)	Primordial germ cells (PGCs) from embryos at E8.5 –E12.5 (m). Gonadal tissues from 5–11 week post-fertilization embryo/fetus (h).	EG cells show essentially the same pluripotency as ES cells when injected into mouse blastocysts (m). The only known difference is the imprinting status of some genes (e.g., Igf2r): Imprinting is normally erased during germline development, and thus, the imprinting status of EG cells is different from that of ES cells.
Trophoblast	Trophectoderm of	TS cells can differentiate into trophoblast giant cells *in vitro* (m).

stem cells (TS, TSC)	E3.5 blastocysts, extraembryonic ectoderm of E6.5 embryos, and chorionic ectoderm of E7.5 embryos.	TS can contribute exclusively to all trophoblast subtypes when injected into blastocysts (m).
Extraembryonic endoderm cells (XEN)	ICM from blastocysts	XEN cells can contribute only to the parietal endoderm lineage when injected into a blastocyst (m).
Embryonal carcinoma cells (EC)	Teratocarcinoma—a type of cancer that develops in the testes and ovaries	EC cells rarely show pluripotency *in vitro*, but they can contribute to nearly all cell types when injected into blastocysts (m). EC cells often have an aneuploid karyotype and other genome alterations (m, h).
Mesenchymal stem cells (MS, MSC)	Bone marrow, muscle, adipose tissue, peripheral blood, and umbilical cord blood (m, h)	MS cells can differentiate into mesenchymal cell types, including adipocytes, osteocytes, chondrocytes, and myocytes (m, h).
Multipotent adult stem cells (MAPC)	Bone marrow mononuclear cells (m, h); postnatal muscle and brain (m)	MAPCs are very rare cells that are present within MSC cultures from postnatal bone marrow (m, h). MAPCs can be cultured for >120 population doublings, can differentiate into all tissues *in vivo* when injected into a mouse blastocyst, and can differentiate into various cell lineages of mesodermal, ectodermal, and endodermal origin *in vitro* (m).
Spermatogonial stem cells (SS, SSC)	Newborn testis (m)	SS cells can reconstitute long-term spermatogenesis after transplantation into recipient testes and restore fertility (m).
Germline stem cells (GS, GSC)	Neonatal testis (m)	GS cells can differentiate into three germlayers *in vitro* and contribute to a variety of tissues, including germline, when injected into blastocysts (m).
Multipotent adult germline stem cells (maGSC)	Adult testis (m)	maGSC can differentiate into three germlayers *in vitro* and can contribute to a variety of tissues, including germline, when injected into blastocysts.
Neural stem cells (NS, NSC)	Fetal and adult brain (subventricular zone, ventricular zone, and hippocampus)	NS cells can be cultured as a heterogeneous cell population of monolayer or floating cell clusters called *neurospheres*. NS cells can differentiate into neuron and glia *in vivo* and *in vitro*. Recently, the culture of pure population of symmetrically dividing adherent NS cells became possible.
Unrestricted somatic stem cells (USSC)	Mononuclear fraction of cord blood (h)	USSCs can differentiate into a variety of cell types *in vitro* and can contribute a variety of cells types in *in vivo* transplantation experiments in rat, mouse, and sheep (h). USSCs are $CD45^{-}$ adherent cells and can be expanded to 10^{15} cells without losing pluripotency (h).
Epistem cells (EpiSC)	Early postimplantation epiblast (m)	EpiSCs can differentiate into three germlayers *in vitro* and form teratomas but cannot contribute to normal tissues when injected into blastocysts (m). EpiSCs (m) are shown to be more similar to ESC (h) than ESC (m).
Induced pluripotent stem cells (iPS, iPSC)	Variety of terminally differentiated cells and tissue stem cells (m, h)	ESC-like cells originally derived by introducing four transcription factors (Klf4, Pou5f1/Oct4, Sox2, and Myc) into mouse embryo fibroblasts by retroviral vectors (m). iPS cells are essentially indistinguishable from ES cells. A number of somatic cell types can be converted into iPS cells using different combinations of transcription factors and treatment with small molecules.

Abbreviations: m, mouse; h, human.

Successful derivation of cultured stem cells (both embryonic and tissue stem cells) often requires the identification of necessary growth factors and culture conditions, mimicking the microenvironment or *niche* of the resident stem cells. For example, the derivation of mouse TS cells, once considered impossible, became feasible by using FGF4, a ligand expressed by adjacent cells in the developing trophoblast in vivo. Therefore, it may be possible to culture other resident stem cells (e.g., intestinal stem cells) or isolated stem cells (e.g., HS cells) by studying the factors that constitute their normal niche.

Self-Renewal and Proliferation of Stem Cells

SYMMETRIC AND ASYMMETRIC CELL DIVISION

The most widely accepted stem cell definition is a cell with a unique capacity to produce unaltered daughter cells (*self-renewal*) and to generate specialized cell types (*potency*). Self-renewal can be achieved in two ways. *Asymmetric cell division* produces one daughter cell that is identical to the parental cell and one daughter cell that is different from the parental cell and is a progenitor or differentiated cell. Asymmetric cell division does not increase the number of stem cells. *Symmetric cell division* produces two identical daughter cells. For stem cells to proliferate in vitro, they must divide symmetrically.

UNLIMITED EXPANSION IN VITRO

Resident stem cells are often quiescent and divide infrequently. However, once the stem cells are successfully cultured in vitro, they often acquire the capacity to divide continuously and the ability to proliferate beyond the normal passage limit typical of primary cultured cells (sometimes called *immortality*). These features are primarily seen in ES cells but have also been demonstrated for NS cells, MS cells, MAPCs, maGSCs (adult-derived tissue stem cells), and USSCs (newborn-derived tissue stem cells), thereby enhancing the potential of these cells for therapeutic use (Table 65–1).

STABILITY OF GENOTYPE AND PHENOTYPE

The capacity to actively proliferate is often associated with the accumulation of chromosomal abnormalities and mutations. Mouse ES cells appear to be an exception to this rule and tend to maintain their euploid karyotype and genome integrity. By contrast, human ES cells appear to be more susceptible to mutations after long-term culture. However, it is also important to note that even euploid mouse ES cells can form teratomas when injected into immuno-suppressed animals, raising concerns about the possible formation of tumors after transplanting actively dividing stem cells.

Potency and Differentiation of Stem Cells

DEVELOPMENTAL POTENCY

The term *potency* is used to indicate a cell's ability to differentiate into specialized cell types. The current lack of knowledge about the molecular nature of potency requires the experimental manipulation of stem cells to demonstrate their potency. For example, in vivo testing can be done by injecting stem cells into mouse blastocysts or immunosuppressed adult mice and determining how many different cell types are formed from the injected cells. In vitro testing can be done by differentiating cells in various culture conditions to determine how many different cell types are formed from the cells. However, these in vivo assays are not applicable to human stem cells. The formal test of self-renewal and potency is performed by demonstrating that a single cell possesses such abilities in vitro (*clonality*). Cultured stem cells are tentatively grouped according to their potency **(Fig. 65-1)**.

Figure 65-1

Stage / Potency	Preimplantation	Embryonic, fetal	Postnatal	Adult
Totipotent	Zygote^{m,h}			
Pluripotent	ES^{m,h}	EG^{m,h} EpiSC^{m}	GS^{m} iPS^{m,h} USSC^{h} MAPC^{m,h}	EC^{m,h} iPS^{m,h} maGSC^{m} MAPC^{m,h}
Multipotent				MS^{m,h}
Oligopotent	TS^{m}			NS^{m,h}
Unipotent	XEN^{m}			SSC^{m}
Terminally differentiated cells				

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Potency and source developmental stage of cultured stem cells. For abbreviations of stem cells, see Table 65–1. Note that stem cells are often abbreviated with or without "cells," e.g., ES cells or ESCs for embryonic stem cells. m, mouse; h, human.

FROM TOTIPOTENCY TO UNIPOTENCY

Totipotent cells can form an entire organism autonomously. Only a fertilized egg (zygote) possesses this feature. *Pluripotent cells* (e.g., ES cells) can form almost all of the body's cell lineages (endoderm,

mesoderm, and ectoderm), including germ cells. *Multipotent cells* (e.g., HS cells) can form multiple cell lineages but cannot form all of the body's cell lineages. *Oligopotent cells* (e.g., NS cells) can form more than one cell lineage but are more restricted than multipotent cells. Oligopotent cells are sometimes called *progenitor cells* or *precursor cells*; however, these terms are often more strictly used to define partially differentiated or lineage-committed cells (e.g., myeloid progenitor cells) that can divide into different cell types but lack self-renewing capacity. *Unipotent cells* or *monopotent cells* [e.g., spermatogonial stem (SS) cells] can form a single differentiated cell lineage. Terminally differentiated cells, such as fibroblasts, also have a capacity to proliferate (which may be called self-renewal) but maintain the same cell type (i.e., they have no ability to form another cell type) and are not, therefore, considered unipotent cells.

NUCLEAR REPROGRAMMING

Development naturally progresses from totipotent fertilized eggs to pluripotent epiblast cells to multipotent cells and, finally, to terminally differentiated cells. According to Waddington's epigenetic landscape, this is analogous to a ball moving down a slope. The reversal of the terminally differentiated cells to totipotent or pluripotent cells (called *nuclear reprogramming*) can thus be seen as an uphill gradient that never occurs in normal conditions. However, nuclear reprogramming has been achieved using *nuclear transplantation*, or *nuclear transfer* (NT), procedures (often called "cloning"), where the nucleus of a differentiated cell is transferred into an enucleated oocyte. Although this is an error-prone procedure with a very low success rate, live animals have been produced using adult somatic cells as donors in sheep, mice, and other mammals. In mice, it has been demonstrated that ES cells derived from blastocysts made by somatic cell NT are indistinguishable from normal ES cells. NT can potentially be used to produce *patient-specific ES cells* carrying a genome identical to that of the patient. However, the successful implementation of this procedure has not been reported in humans. Setting aside technical and ethical issues, the limited supply of human oocytes will be a major problem for clinical applications of NT. Alternatively, successful nuclear reprogramming of somatic cells by fusing them with ES cells has been demonstrated in mice and humans. However, it is not yet clear how ES-derived DNA can be removed from hybrid cells.

An approach that has become increasingly successful is the direct conversion of terminally differentiated cells into ES-like cells [called induced pluripotent (iPS) cells] by transiently overexpressing a combination of key transcription factors (TFs). The original method was to infect mouse embryonic fibroblast cells with retrovirus vectors carrying four TFs [*Pou5f1*(*Oct4*), *Sox2*, *Klf4*, and *Myc*] and to identify rare ES-like cells in culture. It was soon adapted to human cells, followed by a more refined procedure (e.g., the use of fewer TFs, different cell types, and different gene-delivery methods). Use of protein cocktails and a variety of small molecules has also been actively pursued, as the goal is to produce patient-specific iPS cells without altering their genetic makeup.

STEM CELL PLASTICITY OR TRANSDIFFERENTIATION

The prevailing paradigm in developmental biology is that once cells are differentiated, their phenotypes are stable. However, a number of reports have shown that tissue stem cells, which have traditionally been thought to be lineage-committed multipotent cells, possessing the capacity to differentiate into cell types outside their lineage restrictions (called *transdifferentiation*). For example, HS cells may be converted into neurons as well as germ cells. This feature may provide a means to use tissue stem cells derived directly from a patient for therapeutic purposes, thereby eliminating the need to use embryonic stem cells or elaborate procedures such as nuclear reprogramming of a patient's somatic cells. However, more strict criteria and rigorous validation are required to establish tissue stem cell plasticity. For example, observations of transdifferentiation may reflect cell fusion, contamination with progenitor cells from other cell lineages, or persistence of pluripotent embryonic cells in adult organs. Therefore, the assignment of potency to each cultured stem cell in Fig. 65-1 should be taken with caution. Whether transdifferentiation exists and can be used for therapeutic purposes remains to be determined conclusively.

DIRECTED DIFFERENTIATION OF STEM CELLS

Pluripotent stem cells (e.g., ES and iPS cells) can differentiate into multiple cell types, but in culture, they normally differentiate into heterogeneous cell populations in a stochastic manner. However, for therapeutic uses, it is desirable to direct stem cells into specific cell types (e.g., insulin-secreting beta cells). This is an active area of stem cell research, and protocols are being developed to achieve this goal. In any of these directed cell differentiation systems, the cell phenotype must be evaluated critically.

Molecular Characterization of Stem Cells

GENOMICS AND PROTEOMICS

In addition to standard molecular biological approaches, high-throughput genomics and proteomics have been extensively applied to the analysis of stem cells. For example, DNA microarray analyses have revealed the expression levels of essentially all genes and identified specific markers for some stem cells. Chromatin-immunoprecipitation coupled with next-generation sequencing technologies,

capable of producing tens of millions of sequence reads in a single run, have revealed chromatin modifications ("epigenetic marks") relevant to stem cell properties. Similarly, the protein profiles of stem cells have been assessed by using mass spectrometry. These methods are beginning to provide a novel means to characterize and classify various stem cells and the molecular mechanisms that give them their unique characteristics.

ES CELL REGULATION

It is important to identify genes involved in the regulation of stem cell function and to examine the effects of altered gene expression on ES and other stem cells. For example, core networks of TFs such as *Pou5f1* (*Oct4*), *Nanog,* and *Sox*2, govern key gene regulatory pathways/networks for the maintenance of self-renewal and pluripotency of mouse and human ES cells. These TF networks are modulated by specific external factors through signal transduction pathways, such as leukemia inhibitory factor (*Lif*)/*Stat3*, mitogen-activated protein kinase 1/3 (*Mapk1/3*), TGFβ superfamily, and *Wnt*/glycogen synthase kinase 3 beta (*Gsk3b*). Inhibitors of Mapk1/3 and Gsk3b signaling enhance the derivation of ES cells and help to maintain ES cells in a full pluripotency ("ground" or "naive state"). Recent data also indicate that 20–25 nucleotide RNAs, called microRNAs (miRNAs), play an important role in regulating stem cell function by repressing the translation of their target genes. For example, it has been shown that miR-21 regulates cell cycle progression in ES cells and miR-128 prevents the differentiation of hematopoietic progenitor cells. These types of analyses should provide molecular clues about the function of stem cells and lead to a more effective means to manipulate stem cells for future therapeutic use.

FURTHER READINGS

Department of Health and Human Services: NIH Stem Cell Information Home Page, 2010. *http://stemcells.nih.gov.proxy.cc.uic.edu/index*

Gangaraju VK, Lin H: MicroRNAs: Key regulators of stem cells. Nat Rev Mol Cell Biol 10:116, 2009[PMID: 19165214] [Full Text]

Lanza R et al (eds): Essentials of Cell Biology, 2nd ed. San Diego, Academic Press, 2009 Nichols J, Smith A: Naive and primed pluripotent state. Cell Stem Cell 4:487, 2009

Nishiyama A et al: Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. Cell Stem Cell 5:420, 2009[PMID: 19796622] [Full Text]

Saha K, Jaenisch R: Technical challenges in using human induced pluripotent stem cells to Model Disease. Cell Stem Cell 5:584, 2009[PMID: 19951687] [Full Text]

Yamanaka S, Blau HM: Nuclear reprogramming to a pluripotent state by three approaches. Nature 465:704, 2010[PMID: 20535199] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEMATOPOIETIC STEM CELLS: INTRODUCTION

All of the cell types in the peripheral blood and some cells in every tissue of the body are derived from hematopoietic (*hemo*: blood; *poiesis*: creation) stem cells. If the hematopoietic stem cell is damaged and can no longer function (e.g., due to a nuclear accident), a person would survive 2–4 weeks in the absence of extraordinary support measures. With the clinical use of hematopoietic stem cells, tens of thousands of lives are saved each year (Chap. 114). Stem cells produce tens of billions of blood cells daily from a stem cell pool that is estimated to be only in the hundreds of thousands. How stem cells do this, how they persist for many decades despite the production demands, and how they may be better used in clinical care are important issues in medicine.

The study of blood cell production has become a paradigm for how other tissues may be organized and regulated. Basic research in hematopoiesis that includes defining stepwise molecular changes accompanying functional changes in maturing cells, aggregating cells into functional subgroups, and demonstrating hematopoietic stem cell regulation by a specialized microenvironment are concepts worked out in hematology, but they offer models for other tissues. Moreover, these concepts may not be restricted to normal tissue function but extend to malignancy. Stem cells are rare cells among a heterogeneous population of cell types, and their behavior is assessed mainly in experimental animal models involving reconstitution of hematopoiesis. Thus, much of what we know about stem cells is imprecise and based on inferences from genetically manipulated animals.

CARDINAL FUNCTIONS OF HEMATOPOIETIC STEM CELLS

All stem cell types have two cardinal functions: self-renewal and differentiation (Fig. 66-1). Stem cells exist to generate, maintain, and repair tissues. They function successfully if they can replace a wide variety of shorter-lived mature cells over prolonged periods. The process of self-renewal (see below) assures that a stem cell population can be sustained over time. Without self-renewal, the stem cell pool would become exhausted and tissue maintenance would not be possible. The process of differentiation leads to production of the effectors of tissue function: mature cells. Without proper differentiation, the integrity of tissue function would be compromised and organ failure would ensue.

Figure 66-1

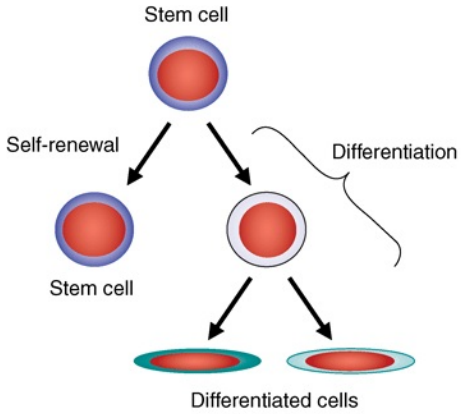

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Signature characteristics of the stem cell. Stem cells have two essential features: the capacity to differentiate into a variety of mature cell types and the capacity for self-renewal. Intrinsic factors associated with self-renewal include expression of Bmi-1, Gfi-1, PTEN, STAT5, Tel/Atv6, p21, p18, MCL-1, Mel-18, RAE28, and HoxB4. Extrinsic signals for self-renewal include Notch, Wnt, SHH, and Tie2/Ang-1. Based mainly on murine studies, hematopoietic stem cells express the following cell surface molecules: CD34, Thy-1 (CD90), c-Kit receptor (CD117), CD133, CD164, and c-Mpl (CD110, also known as the thrombopoietin receptor).

In the blood, mature cells have variable average life spans, ranging from 7 h for mature neutrophils to a few months for red blood cells to many years for memory lymphocytes. However, the stem cell pool is the central, durable source of all blood and immune cells, maintaining a capacity to produce a broad range of cells from a single cell source, yet keeping itself vigorous over decades of life. As an individual stem cell divides, it has the capacity to accomplish one of three division outcomes: two stem cells, two cells destined for differentiation, or one stem cell and one differentiating cell. The former two outcomes are the result of symmetric cell division, whereas the latter indicates a different outcome for the two daughter cells —an event termed *asymmetric cell division*. The relative balance for these types of outcomes may change during development and under particular kinds of demands on the stem cell pool.

Developmental Biology of Hematopoietic Stem Cells

During development, blood cells are produced at different sites. Initially, the yolk sac provides oxygen-carrying red blood cells, and then the placenta and several sites of intraembryonic blood cell production become involved. These intraembryonic sites engage in sequential order, moving from the genital ridge at a site where the aorta, gonadal tissue, and mesonephros are emerging to the fetal liver and then, in the second trimester, to the bone marrow and spleen. As the location of stem cells changes, the cells they produce also change. The yolk sac provides red cells expressing embryonic hemoglobins while intraembryonic sites of hematopoiesis generate red cells, platelets, and the cells of innate immunity. The production of the cells of adaptive immunity occurs when the bone marrow is colonized and the thymus forms. Stem cell

proliferation remains high, even in the bone marrow, until shortly after birth, when it appears to dramatically decline. The cells in the bone marrow are thought to arrive by the bloodborne transit of cells from the fetal liver after calcification of the long bones has begun. The presence of stem cells in the circulation is not unique to a time window in development. Rather, hematopoietic stem cells appear to circulate throughout life. The time that cells spend freely circulating appears to be brief (measured in minutes in the mouse), but the cells that do circulate are functional and can be used for transplantation. The number of stem cells that circulate can be increased in a number of ways to facilitate harvest and transfer to the same or a different host.

Mobility of Hematopoietic Stem Cells

Cells entering and exiting the bone marrow do so through a series of molecular interactions. Circulating stem cells (through CD162 and CD44) engage the lectins P- and E-selectin on the endothelial surface to slow the movement of the cells to a rolling phenotype. Stem cell integrins are then activated and accomplish firm adhesion between the stem cell and vessel wall, with a particularly important role for stem cell VCAM-1 engaging endothelial VLA-4. The chemokine CXCL12 (SDF1) interacting with stem cell CXCR4 receptors also appears to be important in the process of stem cells getting from the circulation to where they engraft in the bone marrow. This is particularly true in the developmental move from fetal liver to bone marrow; however, the role for this molecule in adults appears to be more related to retention of stem cells in the bone marrow rather the process of getting them there. Interrupting that retention process through either specific molecular blockers of the CXCR4/CXCL12 interaction, cleavage of CXCL12, or downregulation of the receptor can all result in the release of stem cells into the circulation. This process is an increasingly important aspect of recovering stem cells for therapeutic use as it has permitted the harvesting process to be done by leukapheresis rather than bone marrow punctures in the operating room. Refining our knowledge of how stem cells get into and out of the bone marrow may improve our ability to obtain stem cells and make them more efficient at finding their way to the specific sites for blood cell production, the so-called stem cell niche.

Hematopoietic Stem Cell Microenvironment

The concept of a specialized microenvironment, or stem cell niche, was first proposed to explain why cells derived from the bone marrow of one animal could be used in transplantation and again be found in the bone marrow of the recipient. This niche is more than just a housing site for stem cells, however. It is an anatomic location where regulatory signals are provided that allow the stem cells to thrive, to expand if needed, and to provide varying amounts of descendant daughter cells. In addition, unregulated growth of stem cells may be problematic based on their undifferentiated state and self-renewal capacity. Thus, the niche must also regulate the number of stem cells produced. In this manner, the niche has the dual function of serving as a site of nurture but imposing limits for stem cells: in effect, acting as both a nutritive and constraining home.

The niche for blood stem cells changes with each of the sites of blood production during development, but for most of human life it is located in the bone marrow. Within the bone marrow, at least two niche sites have been proposed: on trabecular bone surfaces and in the perivascular space. Stem cells may be found in both places by histologic analysis, and functional regulation has been shown at the highly vascular bone surface. Specifically, bone-forming mesenchymal cells, osteoblastic cells, participate in hematopoietic stem cell function, affecting their location, proliferation, and number. The basis for this interaction is through a number of molecules mediating location, such as the chemokine CXCL12 (SDF1), through proliferation signals mediated by angiopoietin 1, and signaling to modulate self-renewal or survival by factors such as Notch ligands, kit ligand, and Wnts. Other bone components, such as the extracellular matrix glycoprotein, osteopontin, and the high ionic calcium found at trabecular surfaces, contribute to the unique microenvironment, or stem cell niche, on trabecular bone. This physiology has practical applications. First, medications altering niche components may have an effect on stem cell function. This has now been shown for a number of compounds, and some are being clinically tested. Second, it is now possible to assess whether the niche participates in disease states and to examine whether targeting the niche with medications may alter the outcome of certain diseases.

Excess Capacity of Hematopoietic Stem Cells

In the absence of disease, one never runs out of hematopoietic stem cells. Indeed, serial transplantation studies in mice suggest that sufficient stem cells are present to reconstitute several animals in succession, with each animal having normal blood cell production. The fact that allogeneic stem cell transplant recipients also never run out of blood cells in their life span, which can extend for decades, argues that even the limiting numbers of stem cells provided to them are sufficient. How stem cells respond to different conditions to increase or decrease their mature cell production remains poorly understood. Clearly, negative feedback mechanisms affect the level of production of most of the cells, leading to the normal tightly regulated blood cell counts. However, many of the regulatory mechanisms that govern production of more mature progenitor cells do not apply or apply differently to stem cells. Similarly, most of the molecules shown to be able to change the size of the stem cell pool have little effect on more mature blood cells. For example, the growth factor erythropoietin, which stimulates red blood cell production from more mature precursor cells, has no effect on stem cells. Similarly, granulocyte colony-stimulating factor drives the rapid proliferation of granulocyte precursors but has little or no effect on the cell cycling of stem cells. Rather, it changes the location of stem cells by indirect means, altering molecules such as CXCL12 that tether stem cells to their niche. Molecules shown to be important for altering the proliferation, self renewal or survival of stem cells, such as cyclin-dependent kinase inhibitors, transcription factors like Bmi-1, or microRNAs like miR125a, have little or different effects on progenitor cells. Hematopoietic stem cells have governing mechanisms that are distinct from the cells they generate.

Hematopoietic Stem Cell Differentiation

Hematopoietic stem cells sit at the base of a branching hierarchy of cells culminating in the many mature cell types that compose the blood and immune system **(Fig. 66-2)**. The maturation steps leading to terminally differentiated and functional blood cells take place both as a consequence of intrinsic changes in gene expression and niche-directed and cytokine-directed changes in the cells. Our knowledge of the details remains incomplete. As stem cells mature to progenitors, precursors, and, finally, mature effector cells, they undergo a series of functional changes. These include the obvious acquisition of functions defining mature blood cells, such as phagocytic capacity or hemoglobin synthesis. They also include the progressive loss of plasticity (i.e., the ability to become other cell types). For example, the myeloid progenitor can make all cells in the myeloid series but none in the lymphoid series. As common myeloid progenitors mature, they become precursors for either monocytes and granulocytes or erythrocytes and megakaryocytes, but not both. Some amount of reversibility of this process may exist early in the differentiation cascade, but that is lost beyond a distinct stage. As cells differentiate, they may also lose proliferative capacity **(Fig. 66-3)**. Mature granulocytes are incapable of proliferation and only increase in number by increased production from precursors. Lymphoid cells retain the capacity to proliferate but have linked their proliferation to the recognition of particular proteins or peptides by specific antigen receptors on their surface. In most tissues the proliferative cell population is a more immature progenitor population. In general, cells within the highly proliferative progenitor cell compartment are also relatively short-lived, making their way through the differentiation process

in a defined molecular program involving the sequential activation of particular sets of genes. For any particular cell type, the differentiation program is difficult to speed up. The time it takes for hematopoietic progenitors to become mature cells is ~10–14 days in humans, evident clinically by the interval between cytotoxic chemotherapy and blood count recovery in patients.

Figure 66-3

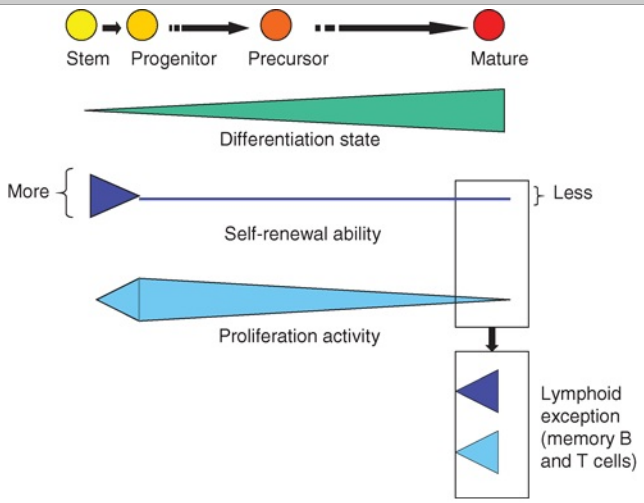

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relative function of cells in the hematopoietic hierarchy. The boxes represent distinct functional features of cells in the myeloid (*upper box*) versus lymphoid (*lower box*) lineages.

Figure 66-2

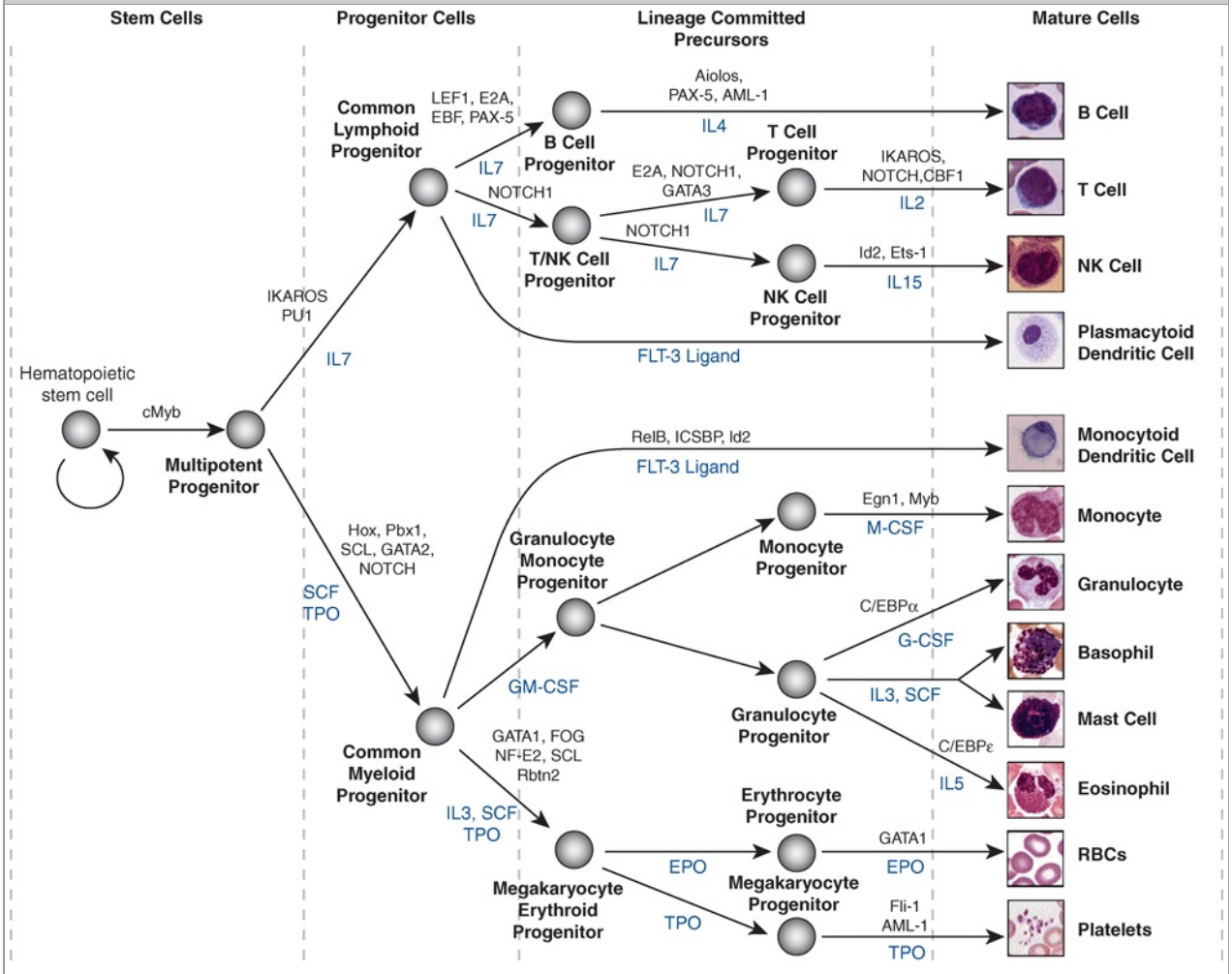

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hierarchy of hematopoietic differentiation. *Stem cells* are multipotent cells that are the source of all descendant cells

and have the capacity to provide either long-term (measured in years) or short-term (measured in months) cell production. *Progenitor cells* have a more limited spectrum of cells they can produce and are generally a short-lived, highly proliferative population also known as transient amplifying cells. *Precursor cells* are cells committed to a single blood cell lineage but with a continued ability to proliferate; they do not have all the features of a fully mature cell. *Mature cells* are the terminally differentiated product of the differentiation process and are the effector cells of specific activities of the blood and immune system. Progress through the pathways is mediated by alterations in gene expression. The regulation of the differentiation by soluble factors and cell-cell communications within the bone marrow niche are still being defined. The transcription factors that characterize particular cell transitions are illustrated on the arrows; the soluble factors that contribute to the differentiation process are in blue. EPO, erythropoietin; SCF, stem cell factor; TPO, thrombopoietin.

Self-Renewal

The hematopoietic stem cell must balance its three potential fates: apoptosis, self-renewal, and differentiation. The proliferation of cells is generally not associated with the ability to undergo a self-renewing division except among memory T and B cells and among stem cells. Self-renewal capacity gives way to differentiation as the only option after cell division when cells leave the stem cell compartment, until they have the opportunity to become memory lymphocytes. In addition to this self-renewing capacity, stem cells have an additional feature characterizing their proliferation machinery. Stem cells in many mature adult tissues may be heterogeneous with some being deeply quiescent, serving as a deep reserve, while others are more proliferative and replenish the short-lived progenitor population. In the hematopoietic system, stem cells are generally cytokine-resistant, remaining dormant even when cytokines drive bone marrow progenitors to proliferation rates measured in hours. Stem cells, in contrast, are thought to divide at far longer intervals measured in months to years, for the most quiescent cells. This quiescence is difficult to overcome in vitro, limiting the ability to effectively expand human hematopoietic stem cells. The process may be controlled by particularly high levels of cyclin-dependent kinase inhibitors that restrict entry of stem cells into cell cycle, blocking the G1-S transition. Exogenous signals from the niche also appear to enforce quiescence, including the activation of the tyrosine kinase receptor Tie2 on stem cells by angiopoietin 1 on osteoblasts.

The regulation of stem cell proliferation also appears to change with age. In mice, the cyclin-dependent kinase inhibitor p16INK4a accumulates in stem cells in older animals and is associated with a change in five different stem cell functions, including cell cycling. Lowering expression of p16INK4a in older animals improves stem cell cycling and capacity to reconstitute hematopoiesis in adoptive hosts, making them similar to younger animals. Mature cell numbers are unaffected. Therefore, molecular events governing the specific functions of stem cells are being gradually made clear and offer the potential of new approaches to changing stem cell function for therapy. One critical stem cell function that remains poorly defined is the molecular regulation of self-renewal.

For medicine, self-renewal is perhaps the most important function of stem cells because it is critical in regulating the number of stem cells. Stem cell number is a key limiting parameter for both autologous and allogeneic stem cell transplantation. Were we to have the ability to use fewer stem cells or expand limited numbers of stem cells ex vivo, it might be possible to reduce the morbidity and expense of stem cell harvests and enable use of other stem cell sources. Specifically, umbilical cord blood is a rich source of stem cells. However, the volume of cord blood units is extremely small and, therefore, the total number of hematopoietic stem cells that can be obtained is generally only sufficient to transplant an individual of <40 kg. This limitation restricts what would otherwise be an extremely promising source of stem cells. Two features of cord blood stem cells are particularly important. (1) They are derived from a diversity of individuals that far exceeds the adult donor pool and therefore can overcome the majority of immunologic cross-matching obstacles. (2) Cord blood stem cells have a large number of T cells associated with them, but (paradoxically) they appear to be associated with a lower incidence of graft-versus-host disease when compared with similarly mismatched stem cells from other sources. If stem cell expansion by self-renewal could be achieved, the number of cells available might be sufficient for use in larger adults. An alternative approach to this problem is to improve the efficiency of engraftment of donor stem cells. Graft engineering is exploring methods of adding cell components that may enhance engraftment. Furthermore, at least some data suggest that depletion of host NK (natural killer) cells may lower the number of stem cells necessary to reconstitute hematopoiesis.

Some limited understanding of self-renewal exists and, intriguingly, implicates gene products that are associated with the chromatin state, a high-order organization of chromosomal DNA that influences transcription. These include members of the polycomb family, a group of zinc finger–containing transcriptional regulators that interact with the chromatin structure, contributing to the accessibility of groups of genes for transcription. One member, Bmi-1, is important in enabling hematopoietic stem cell self-renewal through modification of cell cycle regulators such as the cyclin-dependent kinase inhibitors. In the absence of *Bmi-1* or of the transcriptional regulator, Gfi-1, hematopoietic stem cells decline in number and function. In contrast, dysregulation of *Bmi-1* has been associated with leukemia; it may promote leukemic stem cell self-renewal when it is overexpressed. Other transcription regulators have also been associated with self-renewal, particularly homeobox, or "hox," genes. These transcription factors are named for their ability to govern large numbers of genes, including those determining body patterning in invertebrates. HoxB4 is capable of inducing extensive self-renewal of stem cells through its DNA-binding motif. Other members of the hox family of genes have been noted to affect normal stem cells, but they are also associated with leukemia. External signals that may influence the relative self-renewal versus differentiation outcomes of stem cell cycling include the Notch ligands and specific Wnt ligands. Intracellular signal transducing intermediates are also implicated in regulating self-renewal but, interestingly, are not usually associated with the pathways activated by Notch or Wnt receptors. They include PTEN, an inhibitor of the AKT pathway, and STAT5, both of which are usually downstream of activated growth factor receptors and necessary for normal stem cell functions including self-renewal, at least in mouse models. The connections between these molecules remain to be defined, and their role in physiologic regulation of stem cell self-renewal is still poorly understood.

CANCER IS SIMILAR TO AN ORGAN WITH SELF-RENEWING CAPACITY

The relationship of stem cells to cancer is an important evolving dimension of adult stem cell biology. Cancer may share principles of organization with normal tissues. Cancer might have the same hierarchical organization of cells with a base of stem-like cells capable of the signature stem-cell features, self-renewal, and differentiation. These stem-like cells might be the basis for perpetuation of the tumor and represent a slowly dividing, rare population with distinct regulatory mechanisms, including a relationship with a specialized microenvironment. A subpopulation of self-renewing cells has been defined for some, but not all, cancers. A more sophisticated understanding of the stem-cell organization of cancers may lead to improved strategies for developing new therapies for the many common and difficult-to-treat types of malignancies that have been relatively refractory to interventions aimed at dividing cells.

Does the concept of cancer stem cells provide insight into the cellular origin of cancer? The fact that some cells within a cancer have stem cell–like properties does not necessarily mean that the cancer arose in the stem cell itself. Rather, more mature cells could have acquired the self-renewal characteristics of stem cells. Any single genetic event is unlikely to be sufficient to enable full transformation of a normal cell to a frankly malignant one. Rather, cancer is a multistep process, and for the multiple steps to accumulate, the cell of origin must be able to persist for prolonged periods. It must also be able to

generate large numbers of daughter cells. The normal stem cell has these properties and, by virtue of its having intrinsic self-renewal capability, may be more readily converted to a malignant phenotype. This hypothesis has been tested experimentally in the hematopoietic system. Taking advantage of the cell-surface markers that distinguish hematopoietic cells of varying maturity, stem cells, progenitors, precursors, and mature cells can be isolated. Powerful transforming gene constructs were placed in these cells, and it was found that the cell with the greatest potential to produce a malignancy was dependent on the transforming gene. In some cases it was the stem cell, but in others, the progenitor cell functioned to initiate and per-petuate the cancer. This shows that cells can acquire stem cell-like properties in malignancy.

WHAT ELSE CAN HEMATOPOIETIC STEM CELLS DO?

Some experimental data have suggested that hematopoietic stem cells or other cells mobilized into the circulation by the same factors that mobilize hematopoietic stem cells are capable of playing a role in healing the vascular and tissue damage associated with stroke and myocardial infarction. These data are controversial, and the applicability of a stem-cell approach to nonhematopoietic conditions remains experimental. However, the application of the evolving knowledge of hematopoietic stem cell biology may lead to wide-ranging clinical uses.

The stem cell, therefore, represents a true dual-edged sword. It has tremendous healing capacity and is essential for life. Uncontrolled, it can threaten the life it maintains. Understanding how stem cells function, the signals that modify their behavior, and the tissue niches that modulate stem cell responses to injury and disease are critical for more effectively developing stem cell–based medicine. That aspect of medicine will include the use of the stem cells and the use of drugs to target stem cells to enhance repair of damaged tissues. It will also include the careful balance of interventions to control stem cells where they may be dysfunctional or malignant.

FURTHER READINGS

Dick JE: Stem cell concepts renew cancer research. Blood 112:4793, 2008[PMID: 19064739] [Full Text]

Carlesso N, Cardoso AA: Stem cell regulatory niches and their role in normal and malignant hematopoiesis. Curr Opin Hematol 17:281, 2010[PMID: 20473160] [Full Text]

Laslo P et al: Gene regulatory networks directing myeloid and lymphoid cell fates with the immune system. Semin Immunol 20:228, 2008[PMID: 18771937] [Full Text]

Ottersbach K et al: Ontogeny of haematopoiesis: Recent advances and open questions. Br J Haematol 148:343, 2010[PMID: 19863543] [Full Text]

Schulz C et al: Hematopoietic stem and progenitor cells: Their mobilization and homing to bone marrow and peripheral tissue. Immunol Res 44:160, 2009[PMID: 19340403] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPLICATIONS OF STEM CELL BIOLOGY IN CLINICAL MEDICINE: INTRODUCTION

Damage to an organ initiates a series of events that lead to the reconstruction of the damaged tissue, including proliferation, differentiation and migration of various cell types, release of cytokines and chemokines, and remodeling of the extracellular matrix. Endogenous stem and progenitor cells are among the cell populations that are involved in the injury responses. In normal steady-state conditions, an equilibrium is maintained in which endogenous stem cells intrinsic to the tissue replenish dying cells. After tissue injury, stem cells in organs, such as the liver and skin, have a remarkable ability to regenerate the organ, whereas other stem cell populations, such as those in the heart and brain, have a much more limited capability for self-repair. In rare circumstances, circulating stem cells may contribute to regenerative responses by migrating into a tissue and differentiating into organ-specific cell types. The goal of stem cell therapies is to promote cell replacement in organs that are damaged beyond their ability for self-repair.

GENERAL STRATEGIES FOR STEM CELL REPLACEMENT

At least three different therapeutic concepts for cell replacement can be envisaged (**Fig. 67-1**). One therapeutic approach involves direct administration of stem cells. This may involve injection of the cells directly into the damaged organ, where they can differentiate into the desired cell type. Alternatively, stem cells may be injected systemically since they have the capacity to home in on damaged tissues by following gradients of cytokines and chemokines released by the diseased organ. A second approach involves transplantation of differentiated cells derived from stem cells. For example, pancreatic islet cells would be generated from stem cells before transplantation into diabetic patients and cardiomyocytes would be generated to treat ischemic heart disease. A third approach involves stimulation of endogenous stem cells to facilitate repair. This might be accomplished by administration of appropriate growth factors and drugs that amplify the number of endogenous stem/progenitor cells and/or direct them to differentiate into the desired cell types. Therapeutic stimulation of precursor cells is already a clinical reality in the hematopoietic system, where factors such as erythropoietin, granulocyte colony-stimulating factor (G-CSF), and granulocyte-macrophage colony-stimulating factor (GM-CSF) are used to increase production of specific blood elements. In addition to these strategies for cell replacement, a number of other approaches could involve stem cells for ex vivo or in situ generation of tissues, a process termed tissue engineering (Chap. 69). Stem cells are also excellent candidates as vehicles for cellular gene therapy (Chap. 68). Finally, transplanted stem cells may exert paracrine effects on damaged tissues without differentiating and replacing lost cells.

Figure 67-1

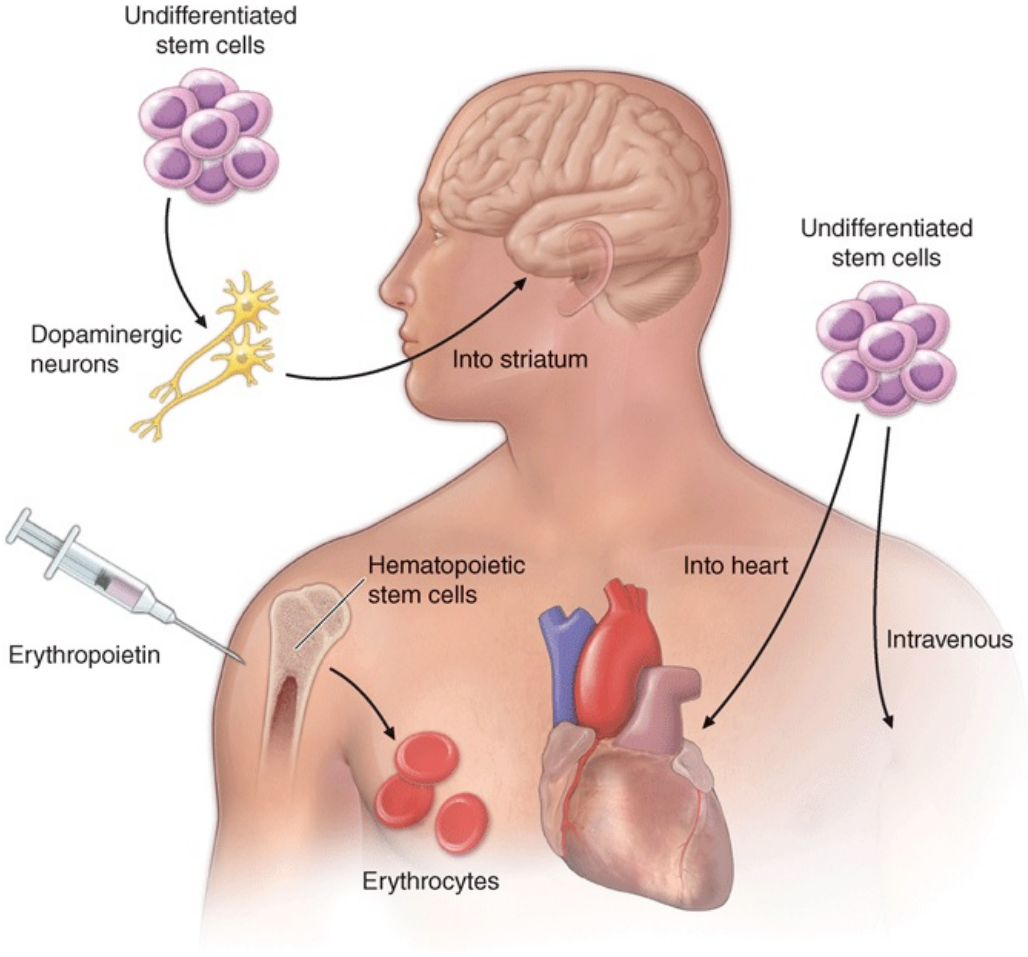

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Strategies for transplantation of stem cells. 1. Undifferentiated or partially differentiated stem cells may be injected directly into the target organ or intravenously. 2. Stem cells may be -differentiated ex vivo before injection into the target organ. 3. Growth factors or other drugs may be injected to stimulate endogenous stem cell -populations.

Stem cell transplantation is not a new concept and is already part of established medical practice. Hematopoietic stem cells (Chap. 66) are responsible for the long-term repopulation of all blood elements in recipients of bone marrow transplants, and hematopoietic stem cell transplantation is the gold standard against which other stem cell transplantation therapies will be measured. Transplantation of differentiated cells is also a clinical reality, and donated organs and tissues often are used to replace damaged tissues. However, the need for transplantable tissues and organs far outweighs the available supply, and organ transplantation has limited potential for some tissues, such as the brain. Stem cells offer the possibility of a renewable source of replacement cells for virtually all organs.

SOURCES OF STEM CELLS FOR TISSUE REPAIR

A variety of different types of stem cells (Chap. 65) could be used in regenerative strategies, including embryonic stem (ES) cells, induced pluripotent stem cells (iPS cells), umbilical cord blood stem cells, organ-specific somatic stem cells (e.g., neural stem cells for treatment of the brain), and somatic stem cells that generate cell types specific for the target organ rather than the donor organ (e.g., bone marrow mesenchymal stem cells for cardiac repair). Each cell type has potential advantages and disadvantages, and there are a number of generic problems in developing any of these cell types into a useful and reliable clinical tool.

Embryonic Stem Cells

Embryonic stem cells have the potential to generate all the cell types in the body, and so in theory there are no restrictions on the organs that could be regenerated. They can self-renew endlessly so that a single cell line with carefully characterized traits potentially could generate almost limitless numbers of cells. In the absence of moral or ethical constraints, unused human blastocysts from fertility clinics could be used to derive new ES cell lines that are matched immunologically with potential transplant recipients. Alternatively, somatic cell nuclear transfer ("therapeutic cloning") could

be used to create ES cell lines that are genetically identical to those of the patient. However, human ES cells are difficult to culture and grow slowly. Techniques for differentiating them into specific cell types are just beginning to be developed. Cells tend to develop abnormal karyotypes and other abnormalities with increased time in culture, and ES cells have the potential to form teratomas if all cells are not committed to the desired cell types before transplantation. Further, human ES cells are ethically controversial and, on these grounds, would be unacceptable to some patients and physicians despite their therapeutic potential.

Induced Pluripotent Stem Cells

The field of stem cell biology was transformed by the discovery that adult somatic cells can be converted ("reprogrammed") into pluripotent cells by overexpressing four transcription factors normally expressed in pluripotent cells (Chap. 65). These iPS cells share most properties with ES cells, although there are distinct differences in gene expression between ES and iPS cells. Viruses were used initially to insert the transcription factors into somatic cells, making them unsuitable for clinical use. However, a number of strategies have been developed to circumvent this problem, including the insertion of proteins rather than cDNAs; insertion of transposons with the programming factors, followed by their subsequent removal; and use of floxed viral constructs followed by Cre recombinase to excise the viral constructs. None of these approaches is sufficiently developed for cells to be used clinically, but appropriate techniques are likely to evolve quickly. The potential advantages of such cells are that use of somatic cells from patients would generate iPS cells genetically identical to those of the patient and that the cells do not have the same ethical constraints as ES cells. It is not clear whether the differences in gene expression between ES and iPS cells will have any impact on their potential clinical utility, and studies of both cell types will be essential to resolve this issue.

Umbilical Cord Stem Cells

Umbilical cord blood stem/progenitor cells are widely and easily available, and they appear to be associated with less graft-versus-host disease than are some other cell types, such as marrow stem cells. They have less human leukocyte antigen (HLA) restriction than adult marrow stem cells and are less likely to be contaminated with herpesvirus. However, it is unclear how many different cell types can be generated from umbilical cord blood stem cells, and methods for differentiating these cells into nonhematopoietic phenotypes are largely lacking. Further, the quantity of obtainable cells may be limiting.

Organ-Specific Multipotent Stem Cells

Organ-specific multipotent stem cells have the advantage that they are already somewhat specialized and inducement of desired cell types thus may be easier. Cells potentially could be obtained from the patient and amplified in cell culture, circumventing the problems associated with immune rejection. Stem cells are relatively easy to harvest from some tissues, such as bone marrow and blood, but difficult in the case of other tissues, such as heart and brain. However, these populations of cells are more limited in potentiality than are pluripotent ES or iPS cells, and they may be difficult to obtain in large quantities from many organs. Therefore, substantial efforts have been devoted to developing techniques for using more easily obtainable stem cell populations, such as bone marrow mesenchymal stem cells (MSCs) and adipose stem cells, for use in regenerative strategies. Tissue culture evidence suggests that these stem cell populations may be able to generate differentiated cell types unrelated to their organ source, including myocytes, chondrocytes, tendon cells, osteoblasts, cardiomyocytes, adipocytes, hepatocytes, and neurons, a process known as transdifferentiation. However, it is still unclear whether these stem cells are capable of generating differentiated cell types that integrate into organs, survive, and function after transplantation in vivo. A number of early studies of bone marrow–derived stem cells transplanted into heart, liver, and other organs suggested that the cells had differentiated into organ-specific cell types. Subsequent studies, however, revealed that the stem cells had simply fused with cells resident in the organs. Further studies will be necessary to determine whether transdifferentiation of MSCs, adipose stem cells, or other stem cell populations occurs at a high enough frequency to be useful for stem cell replacement therapy.

Regardless of the source of the stem cell used in regenerative strategies, there are a number of generic problems that must be overcome for the development of successful clinical applications. These problems include development of methods for reliably generating large numbers of specific cell types, minimizing the risk of tumor formation or proliferation of inappropriate cell types, ensuring the viability and function of the engrafted cells, overcoming the problems of immune rejection when autografts are not used, and facilitating revascularization of the regenerated tissue. Each organ system will also pose tissue-specific problems for stem cell therapies.

DISEASE-SPECIFIC APPLICATIONS OF STEM CELLS

Ischemic Heart Disease and Cardiomyocyte Regeneration

Because of the prevalence of ischemic heart disease, extensive efforts have been devoted to the development of strategies for cell replacement of cardiomyocytes. Historically, the adult heart has been viewed as a terminally differentiated organ without the capacity for regeneration. However, recent

studies have demonstrated that the heart has the ability to achieve low levels of cardiomyocyte regeneration (Chap. 224). This regeneration appears to be accomplished by cardiac stem cells resident in the heart and possibly by cells originating in the bone marrow. Cardiac stem cells might be an ideal source for therapeutic use, but techniques for isolating, characterizing, and amplifying large numbers of these cells have not been perfected. To effect myocardial repair, stem cell therapy must deliver cells either systemically or locally, and the cells must survive, engraft, and differentiate into functional cardiomyocytes that couple mechanically and electrically with the recipient myocardium. The optimal method for cell delivery is not clear, and various experimental and clinical studies have successfully employed intramyocardial, transendocardial, intravenous, and intracoronary injections. In experimental myocardial infarction, functional improvements have been achieved after transplantation of a variety of different cell types, including ES cells, bone marrow stem cells, endothelial stem cells, and adipose stem cells. Early studies suggested that each of these cell types might have the potential to engraft and generate cardiomyocytes. However, most investigators have found that the generation of new cardiomyocytes by these cells is at best a rare event and that graft survival over long periods is poor. The preponderance of evidence suggests that the beneficial effects of most experimental therapies were not derived from direct stem cell generation of cardiomyocytes but rather from indirect effects of the stem cells on resident cells. It is not clear whether this reflects the release of soluble growth factors, induction of angiogenesis, or another mechanism. The ready availability of bone marrow stem cells facilitated a series of limited clinical trials in human ischemic heart disease. A wide variety of cell delivery methods, cell types, and doses of cells have been used in the various studies, and the fate of the cells and the mechanisms by which they altered cardiac function are open questions. In aggregate, however, these studies have shown a small but measurable improvement in cardiac function and, in some cases, reduction in infarct size. The limited available evidence suggests that the beneficial effects probably reflect an indirect effect of the transplanted cells rather than genuine cell replacement.

Diabetes

Successes with islet cell and pancreas transplantation have provided proof of the concept for cell-based therapies for type 1 diabetes. However, the demand for donor pancreases far exceeds the number available, and maintenance of long-term graft survival is a problem. The search for a renewable source of stem cells capable of regenerating pancreatic islets has therefore been intensive. Pancreatic beta cell turnover occurs even in the normal pancreas, although the source of the new beta cells remains controversial. This suggests that in principle, it should be possible to develop strategies for reconstituting the beta cell population in diabetics. Attempts to devise techniques for promoting endogenous regenerative processes by using combinations of growth factors, drugs, and gene therapy have failed thus far, but this remains a potentially viable approach. A number of different cell types are candidates for use in stem cell replacement strategies, including iPS cells, ES cells, hepatic progenitor cells, pancreatic ductal progenitor cells, and bone marrow stem cells. Successful therapy will depend on the development of a source of cells that can be amplified to produce large numbers of progeny that have the ability to synthesize, store, and release insulin when it is required, primarily in response to changes in the ambient level of glucose. The proliferative capacity of the replacement cells must be tightly regulated to avoid excessive expansion of beta cell numbers with the consequent development of hyperinsulinemia/hypoglycemia, and the cells must withstand immune rejection. Although it has been reported that ES and iPS cells can be differentiated into cells that produce insulin, these cells have a low content of insulin and a high rate of apoptosis and generally lack the capacity to normalize blood glucose in diabetic animals. Thus, ES and iPS cells have not been useful for the large-scale production of differentiated islet cells. During embryogenesis, the pancreas, liver, and gastrointestinal tract are all derived from the anterior endoderm, and transdifferentiation of the pancreas to liver and vice versa has been observed in a number of pathologic conditions. There is also substantial evidence that multipotential stem cells reside within gastric glands and intestinal crypts. These observations suggest that hepatic, pancreatic and/or gastrointestinal precursor cells may be reasonable candidates for cell-based therapy for diabetes, although it is unclear whether insulin-producing cells derived from pancreatic stem cells or liver progenitors can be expanded in vitro to clinically useful numbers. Bone marrow stem cells and neural stem cells both reportedly have the capacity to generate insulin-producing cells, but there is no convincing evidence that either cell type will be clinically useful.

Nervous System

Substantial progress has been made in developing methodologies for generating neural cells from different stem cell populations. Human ES or iPS cells can be induced to generate cells with the properties of neural stem cells, and these cells in turn give rise to neurons, oligodendroglia, and astrocytes. Reasonably large numbers of these cells can be transplanted into the rodent brain with formation of appropriate cell types and no tumor formation. Multipotent stem cells present in the adult brain also can be easily amplified in number and used to generate all the major neural cell types, but the invasive procedures that would be necessary to obtain autologous cells are a major limitation. Fetal neural stem cells derived from miscarriages or abortions are an alternative but may raise ethical concerns. Nevertheless, a clinical trial of fetal neural stem cells in Batten's disease is in progress.

Transdifferentiation of bone marrow and adipose stem cells into neural stem cells, and vice versa, has been reported by numerous investigators, and clinical trials of such cells have begun for a number of neurologic disorders. Clinical trials of a conditionally immortalized human cell line and of human umbilical cord blood cells in stroke are planned. Some evidence suggests that epidermal stem cells also have the potential for neural regeneration. Neurologic disorders that have been targeted for stem cell therapies include spinal cord injury, amyotrophic lateral sclerosis (ALS), stroke, traumatic brain injury, Batten's disease, and Parkinson's disease. In Parkinson's disease the major motor features of the disorder result from the loss of a single cell population—dopaminergic neurons within the substantia nigra—suggesting that cell replacement should be relatively straightforward. However, two clinical trials of fetal nigral transplantation failed to meet their primary endpoint and were complicated by the development of dyskinesia. Transplantation of stem cell–derived dopamine-producing cells offers a number of potential advantages over the fetal transplants, including the ability of stem cells to migrate and disperse within tissue, the potential for engineering regulatable release of dopamine, and the ability to engineer cells to produce factors that will enhance cell survival. Nevertheless, the experience with fetal transplants points out the difficulties that may be encountered.

At least some of the neurologic dysfunction after spinal cord injury reflects demyelination, and both ES cells and marrow-derived stem cells are able to facilitate remyelination after experimental spinal cord injury. Clinical trials of marrow-derived stem cells in this disorder have commenced in a number of countries, and this may be the first disease targeted for the clinical use of ES cells. Marrow-derived stem cells also are being used in the treatment of stroke, traumatic brain injury, and ALS, in which possible benefits are more likely to be due to indirect trophic effects or remyelination than to neuron replacement. At present no population of transplanted stem cells has been shown to have the capacity to generate neurons that extend axons over long distances to form synaptic connections (as would be necessary for replacement of upper motor neurons in ALS, stroke, or other disorders).

Liver

Liver transplantation is currently the only successful treatment for end-stage liver diseases, but the shortage of liver grafts is a serious problem. Clinical trials of hepatocyte transplantation demonstrate that it potentially can substitute for organ transplantation, but this approach is limited by the paucity of available cells. Potential sources of stem cells for regenerative strategies include endogenous liver stem cells (such as oval cells), ES cells, bone marrow cells, and umbilical cord blood cells. Although a series of studies in humans as well as animals suggested that transplanted bone marrow stem cells can generate hepatocytes, fusion of the transplanted cells with endogenous liver cells, giving the erroneous appearance of new hepatocytes, appears to be the underlying event in most circumstances. The available evidence suggests that transplanted hematopoietic cells can generate hepatocyte-like cells in the liver only at a very low frequency. ES cells can be differentiated into hepatocytes and transplanted in animal models of liver failure without the formation of teratomas.

Other Organ Systems and the Future

The use of stem cells in regenerative strategies has been studied for many other organ systems and cell types, including skin, eye, cartilage, bone, kidney, lung, endometrium, vascular endothelium, smooth muscle, and striated muscle, among others. In fact, the potential for stem cell regeneration of damaged organs and tissues is virtually limitless. However, there are numerous obstacles to be overcome before stem cell therapies can become a widespread clinical reality. Only hematopoietic stem cells have been adequately characterized by surface markers so that they can be unambiguously identified, a prerequisite for reliable clinical applications. The pathways for differentiating stem cells into specific cellular phenotypes are largely unknown, and there is little ability at present to control the migration of transplanted cells or predict the response of the cells to the environment of diseased organs. Some strategies may employ the coadministration of scaffolding, artificial extracellular matrix, and/or growth factors to orchestrate differentiation of stem cells and their organization into appropriate constituents of the organ. There is currently no way to image stem cells in vivo after transplantation into humans, and it will be necessary to develop techniques to accomplish this. Fortunately, stem cells can be engineered before transplantation to contain a contrast agent that may make this feasible. The potential for tumor formation and the problems associated with immune rejection are impediments, and it will also be necessary to develop techniques for ensuring vascularization of regenerated tissues. There already are many strategies for cell replacement, including vasoactive endothelial growth factor (VEGF) coadministration to foster vascularization of the transplant. Some strategies also include genetically engineering stem cells to have an inducible suicide gene so that the cells can be easily eradicated in the event of tumor formation or another complication. The potential for stem cell therapies to revolutionize medical care is extraordinary, and disorders such as myocardial infarction, diabetes, and Parkinson's disease, among many others, will become potentially curable. However, such stem cell–based therapies are still at a very early stage of development, and perfection of techniques for clinical transplantation of predictable, well-characterized cells is going to be a difficult and lengthy undertaking.

Ethical Issues

Stem cell therapies raise ethical and socially contentious issues that must be addressed in parallel with

the scientific and medical opportunities. This society has great diversity with respect to religious beliefs, concepts of individual rights, tolerance for uncertainty and risk, and boundaries for how scientific interventions should be used to alter the outcome of disease. In the United States, the federal government has authorized research using existing human ES cell lines but still restricts the use of federal funds for developing new human ES cell lines. Ongoing studies of existing lines have indicated that they develop abnormalities with time in culture and that they may be contaminated with mouse proteins. These findings point out the need to develop new human ES cell lines. The development of iPS cell technology may lessen the need for deriving new ES cell lines, but it is still not clear whether the differences in gene expression by ES and iPS cells are important for potential clinical use.

In considering ethical issues associated with the use of stem cells, it is helpful to draw from experience with other scientific advances, such as organ transplantation, recombinant DNA technology, implantation of mechanical devices, neuroscience and cognitive research, in vitro fertilization, and prenatal genetic testing. These and other precedents have pointed to the importance of understanding and testing fundamental biology in the laboratory setting and in animal models before applying new techniques in carefully controlled clinical trials. When these trials occur, they must include full informed consent and provide for careful oversight by external review groups.

Ultimately, there will be medical interventions that are scientifically feasible but ethically or socially unacceptable to some members of a society. Stem cell research raises fundamentally difficult questions about the definition of human life, and it has raised deep fears about the ability to balance issues of justice and safety with the needs of critically ill patients. Health care providers and experts with backgrounds in ethics, law, and sociology, must help guard against the premature or inappropriate application of stem cell therapies and the inappropriate use of vulnerable population groups. However, these therapies offer important new strategies for the treatment of otherwise irreversible disorders. An open dialogue among the scientific community, physicians, patients and their advocates, lawmakers, and the lay population is critically important to raise and address important ethical issues and balance the benefits and risks associated with stem cell transfer.

FURTHER READINGS

Higgs DR: A new dawn for stem-cell therapy. N Engl J Med 358:964, 2008[PMID: 18305272] [Full Text]

Kiskinis E, Eggan K: Progress toward the clinical application of patient-specific pluripotent stem cells. J Clin Invest 120:51, 2010[PMID: 20051636] [Full Text]

Park IH et al: Disease-specific induced pluripotent stem cells. Cell 134:877, 2008[PMID: 18691744] [Full Text]

Parmacek MS, Epstein JA: Cardiomyocyte renewal. N Engl J Med 361:86, 2009[PMID: 19571289] [Full Text]

Riazi AM et al: Stem cell sources for regenerative medicine. Methods Mol Biol 482:55, 2009[PMID: 19089350] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GENE THERAPY IN CLINICAL MEDICINE: INTRODUCTION

Gene transfer is a novel area of therapeutics in which the active agent is a nucleic acid rather than a protein or small molecule. Because delivery of naked DNA or RNA to a cell is an inefficient process, most gene transfer is carried out using a vector, or gene delivery vehicle. These vehicles have generally been engineered from viruses by deleting some or all of the viral genome and replacing it with the therapeutic gene of interest under the control of a suitable promoter **(Table 68–1)**. Gene transfer strategies can be described in terms of three essential elements: (1) a vector; (2) a gene to be delivered, sometimes called the transgene; and (3) a relevant target cell to which the DNA or RNA is delivered. The series of steps in which the donated DNA enters the target cell and expresses the transgene is referred to as *transduction*. Gene delivery can take place in vivo, in which the vector is directly injected into the patient or, in the case of hematopoietic and some other target cells, ex vivo, with removal of the target cells from the patient, followed by return of the modified autologous cells after gene transfer in the laboratory. The latter approach offers opportunities to integrate gene transfer techniques with cellular therapies (Chap. 67).

Table 68–1 Characteristics of Gene Delivery Vehicles

		Viral Vectors						
Features	**Retroviral**	**Lentiviral**	**Adenoviral**	**AAV**	**Human Foamy Virus**	**HSV-1**	**SV-40**	**Alphaviruses**
Viral genome	RNA	RNA	DNA	DNA	RNA	DNA	DNA	RNA
Cell division requirement	Yes	G1 phase	No	No	No	No	No	No
Packaging limitation	8 kb	8 kb	8–30 kb	5 kb	8.5 kb	40–150 kb	5 kb	5 kb
Immune responses to vector	Few	Few	Extensive	Few	Few	Few in recombinant virus	Few	Few
Genome integration	Yes	Yes	Poor	Poor	Yes	No	Poor	No
Long-term expression	Yes	Yes	No	Yes	Yes	No	No	No
Main advantages	Persistent gene transfer in dividing cells	Persistent gene transfer in transduced tissues	Highly effective in transducing various tissues	Elicits few inflammatory responses, nonpathogenic	Persistent gene expression in both dividing and nondividing cells	Large packaging capacity with persistent gene transfer	Wide host cell range; lack of immunogenicity	Limited immune responses against the vector
Main disadvantages	Theoretical risk of insertional mutagenesis (occurred in three cases)	Might induce oncogenesis in some cases	Viral capsid elicits strong immune responses	Limited packaging capacity	In need of a stable packaging system	Residual cytotoxicity with neuron specificity	Limited packaging capacity	Transduced gene expression is transient

Abbreviations: AAV, adeno-associated virus; HSV, herpes simplex virus; SV, sarcoma virus.

Gene transfer technology is still under development, although licensing applications for gene therapy products have been filed. Gene therapy is one of the most complex therapeutic modalities yet attempted, and each new disease represents a therapeutic problem for which dosing, safety, and efficacy must be defined. Nonetheless, gene transfer remains one of the most powerful concepts in modern molecular medicine and has the potential to address a host of diseases for which there are currently no cures or, in some cases, no available treatment. More than 5000 subjects have been enrolled in gene transfer studies, and serious adverse events have been rare. Gene therapies are being developed for a wide variety of disease entities **(Fig. 68-1)**.

Figure 68-1

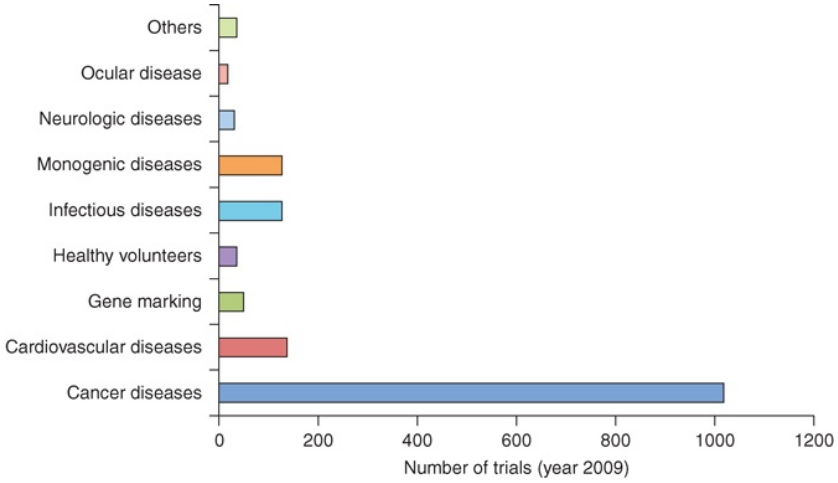

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Indications in gene therapy clinical trials. The bar graph classifies clinical gene transfer studies by disease. A majority of trials have addressed cancer, with monogenic disorders and cardiovascular diseases the next largest categories. *(Adapted from J Gene Med. New Jersey, Wiley, 2009.)*

Gene Transfer for Genetic Disease

Gene transfer strategies for genetic disease generally involve gene addition therapy. This approach most commonly involves transfer of the missing gene to a physiologically relevant target cell. However, other strategies are possible, including supplying a gene that achieves a similar biologic effect through an alternative pathway (e.g., factor VIIa for hemophilia A), supplying an antisense oligonucleotide to splice out a mutant exon if the sequence is not critical to the function of the protein (as has been done with the dystrophin gene in Duchenne's muscular dystrophy), or downregulating a harmful response through an siRNA. Two distinct strategies are used to achieve long-term gene expression: one is to transduce stem cells with an integrating vector, so that all progeny cells will carry the donated gene; the other is to transduce long-lived cells such as skeletal muscle or neural cells. In the case of long-lived cells, integration into the target cell genome is unnecessary. Instead, because the cells are nondividing, the donated DNA can be stabilized in an episomal form, avoiding problems related to integration and insertional mutagenesis.

IMMUNODEFICIENCY DISORDERS: PROOF OF PRINCIPLE

Early attempts to effect gene replacement into hematopoietic stem cells (HSCs) were stymied by the relatively low transduction efficiency of retroviral vectors, which require dividing target cells for integration. Because HSCs are normally quiescent, they are a formidable transduction target. However, identification of cytokines that induced cell division without promoting differentiation of stem cells, along with technical improvements in the isolation and transduction of HSCs, led to modest but real gains in transduction efficiency.

The first convincing therapeutic effect from gene transfer occurred with X-linked severe combined immunodeficiency disease (SCID), which results from mutations in the gene (*IL2RG*) encoding the γc subunit of cytokine receptors required for normal development of T and NK cells (Chap. 316). Affected infants present in the first few months of life with overwhelming infections and/or failure to thrive. In this disorder, it was recognized that the transduced cells, even if few in number, would have a proliferative advantage compared to the nontransduced cells, which lack receptors for the cytokines required for lymphocyte development and maturation. Complete reconstitution of the immune system, including documented responses to standard childhood vaccinations, clearing of infections, and remarkable gains in growth occurred in most of the treated children. However, among 20 children treated in two separate trials, 5 developed a syndrome similar to T cell acute lymphocytic leukemia, with splenomegaly, rising white blood cell counts, and the emergence of a single clone of T cells. In most of these children, the retroviral vector had integrated within a gene, *LMO-2* (LIM only-2), which encodes a component of a transcription factor complex involved in hematopoietic development. The retroviral long terminal repeat is thought to increase the expression of *LMO-2*.

The X-linked SCID studies were a watershed event in the evolution of gene therapy. They demonstrated conclusively that gene therapy could cure disease; of the 20 children eventually treated in these trials, 18 achieved correction of the immunodeficiency disorder. Unfortunately, 5 later developed a leukemia-like disorder, and 1 died of this complication, but the rest are alive and free of complications for up to 10 years after initial treatment. These studies also demonstrated that insertional mutagenesis leading to cancer was more than a hypothetical possibility **(Table 68–2)**. As a result of the experience in these trials, all protocols using integrating vectors in hematopoietic cells must include a plan for monitoring sites of insertion and clonal proliferation. Strategies to overcome this complication have included employing a "suicide" gene cassette in the vector, so that errant clones can be quickly ablated; or using "insulator" elements in the cassette, which can limit the activation of genes surrounding the insertion site. Lentiviral vectors, which can efficiently transduce nondividing target cells, may prove to be safer than retroviral vectors, based on patterns of integration; the field is thus gradually using these as an alternative to retroviral vectors.

Table 68–2 Potential Complications of Gene Therapy

Gene silencing: repression of promoter
Genotoxicity: complications arising from insertional mutagenesis
Phenotoxicity: complications arising from overexpression or ectopic expression of the transgene
Immunotoxicity: harmful immune response to either the vector or transgene
Risks of horizontal transmission: shedding of infectious vector into environment
Risks of vertical transmission: germline transmission of donated DNA

More clear-cut success has been achieved in a gene therapy trial for another form of SCID, adenosine deaminase (ADA) deficiency (Chap. 316). ADA-SCID is clinically similar to X-linked SCID, although it can be treated by enzyme replacement therapy with a pegylated form of the enzyme (PEG-ADA), which leads to immune reconstitution but not always to normal T cell counts. Enzyme replacement therapy is expensive (annual costs: $200,000–$300,000 in U.S. dollars). The initial trials of gene therapy for ADA-SCID were unsuccessful, but modifications of this protocol to include the use of HSCs rather than T cells as the

target for transduction; discontinuation of PEG-ADA at the time of vector infusion, so that the transduced cells have a proliferative advantage over the nontransduced; and the use of a mild conditioning regimen to facilitate engraftment of the transduced cells. There have been no complications in the 10 children treated on the Milan protocol, with a median follow-up of >4 years. ADA-SCID, then, is an example where gene therapy has changed therapeutic options for patients. For those with an HLA-identical sibling, bone marrow transplantation is still the best treatment option, but this applies to only a minority of those affected. For those without an HLA-identical match, gene therapy has comparable efficacy to PEG-ADA and does not run the risk of neutralizing antibodies to the bovine enzyme.

OTHER APPLICATIONS OF INTEGRATING VECTORS

More recently, therapeutic success has also been reported in the setting of X-linked adrenoleukodystrophy, a fatal demyelinating disease of the central nervous system (CNS) caused by mutations in the gene encoding an adenosine triphosphate–binding cassette transporter. Deficiency of this protein leads to adrenal insufficiency and accumulation of very-long-chain fatty acids in oligodendrocytes and microglia, disrupting myelin maintenance by these cells. Affected boys present with clinical and neuroradiographic evidence of disease at age 6–8 years and usually die before adolescence. Aubourg and colleagues have shown dramatic stabilization of disease both clinically and radiographically following lentiviral transduction of hematopoietic stem cells in young boys with disease. Other diseases likely to be amenable to transduction of HSCs include Wiskott-Aldrich syndrome, chronic granulomatous disease, thalassemia, and sickle cell disease, with trials under way currently for the first three of these disorders.

LONG-TERM EXPRESSION IN GENETIC DISEASE: IN VIVO GENE TRANSFER WITH RECOMBINANT ADENO-ASSOCIATED VIRAL (AAV) VECTORS

Recombinant AAV vectors have emerged as attractive gene delivery vehicles for genetic disease. Engineered from a small replication-defective DNA virus, they are devoid of viral coding sequences and trigger very little immune response in experimental animals. They are capable of transducing nondividing target cells, and the donated DNA is stabilized primarily in an episomal form, thus minimizing risks associated with insertional mutagenesis. Because the vector has a tropism for certain long-lived cell types, such as skeletal muscle, the central nervous system, and hepatocytes, long-term expression can be achieved even in the absence of integration.

Clinical trials using recombinant AAV vectors are now ongoing for muscular dystrophies, α_1-antitrypsin deficiency, Parkinson's disease, lipoprotein lipase deficiency, hemophilia B, and a form of congenital blindness called Leber's congenital amaurosis. Hemophilia is often considered a promising disease model for gene transfer, as the gene product does not require precise regulation of expression and biologically active clotting factors can be synthesized in a variety of tissue types, permitting latitude in choice of target tissue. Moreover, raising circulating factor levels from <1% (levels seen in those severely affected) into the range of 5% greatly improves the phenotype of the disease. Preclinical studies with recombinant AAV vectors infused into skeletal muscle or liver have resulted in long-term (>5 years) expression of factor VIII or factor IX in the hemophilic dog model. Administration to skeletal muscle of an AAV vector expressing factor IX in patients with hemophilia was safe and resulted in long-term expression as measured by muscle biopsy, but circulating levels never rose >1% for sustained periods, and a large number of IM injections (>80–100) was required to access a large muscle mass. Intravascular vector delivery has been employed to access large areas of skeletal muscle in animal models of hemophilia and will likely be tested in upcoming trials. Administration of an AAV vector expressing factor IX to the liver in humans with hemophilia resulted in therapeutic circulating levels at the highest dose tested, but expression at these levels (>5%) lasted for only 6–10 weeks before declining to baseline (<1%). A memory T cell response to viral capsid, present in humans but not in other animal species (which are not natural hosts for the virus), may be a contributing factor in the loss of expression (Table 68–2). Trials currently under way are testing two potential strategies to overcome this obstacle—in the first, subjects receive a short course of immunosuppression coadministered with the vector to block the immune response to capsid until the capsid proteins have been degraded and metabolized. In the second, the use of a more efficient expression cassette with a codon-optimized, self-complementary design should allow expression of factor IX at lower doses of vector, which may also elude the immune response.

Early experience with AAV in liver in the hemophilia trial suggested that introduction of small doses of vector into an immunoprivileged site might allow long-term expression, even in humans previously exposed to AAV. This has been most elegantly demonstrated in the setting of the retinal degenerative disease Leber's congenital amaurosis (LCA). LCA, characterized by early-onset blindness, is not currently treatable, and is caused by mutations in several different genes; ~15% of cases of LCA are due to a mutation in a gene, *RPE65*, encoding a retinal pigment epithelial-associated 65 kilodalton protein. In dogs with a null mutation in *RPE65*, sight has been restored after subretinal injection of an AAV vector expressing *RPE65*. Transgene expression appears to be stable, with the first animals treated >10 years ago continuing to manifest electrophysiologic and behavioral evidence of visual function. As is the case for X-linked SCID, gene transfer must occur relatively early in life to achieve optimal correction of the genetic disease, although the exact limitations imposed by age have not yet been defined. AAV-RPE65 trials carried out in both the United States and the United Kingdom have shown restoration of visual and retinal function in more than 20 subjects, with the most marked improvement occurring in the youngest subjects treated (age 8 years). Other inherited retinal degenerative disorders may also be amenable to correction by gene transfer, as are certain complex acquired disorders such as age-related macular degeneration, which affects several million people worldwide. The neovascularization that occurs in age-related macular degeneration can be inhibited by expression of vascular endothelial growth factor (VEGF) inhibitors such as angiostatin, or through the use of RNAi-mediated knockdown of VEGF. Early-phase trials of siRNAs that target VEGF RNA are under way, but these require repeated intravitreal injection of the siRNAs; an AAV vector–mediated approach, which would allow long-term inhibition of the biologic effects of VEGF through a soluble VEGF receptor is now in early-phase clinical testing.

Gene Therapy for Cancer

The majority of clinical gene transfer experience has been in subjects with cancer (Fig. 68-1). As a general rule, a feature that distinguishes gene therapies from conventional cancer therapeutics is that the former are less toxic, in some cases because they are delivered locally (e.g., intratumoral injections), and in other cases because they are targeted specifically to features of the tumor (immunotherapies, antiangiogenic approaches).

Cancer gene therapies can be divided into local and systemic approaches **(Table 68–3)**. Some of the earliest cancer gene therapy trials focused on local delivery of a prodrug or a suicide gene that would increase sensitivity of tumor cells to cytotoxic drugs. A frequently used strategy has been intratumoral injection of an adenoviral vector expressing the thymidine kinase (*TK*) gene. Cells that take up and express the *TK* gene can be killed after the administration of gancyclovir, which is phosphorylated to a toxic nucleoside by *TK*. Because cell division is required for the toxic nucleoside to affect cell viability, this strategy was initially used in aggressive brain tumors (glioblastoma multiforme) where the cycling tumor cells were affected but the nondividing normal neurons were not. More recently, this approach has been explored for locally recurrent prostate, breast, and colon tumors, among others.

Table 68–3 Gene Therapy Strategies in Cancer
Local/regional approaches

Suicide gene/prodrug Suppressor oncogene
Oncolytic virus
Systemic approaches Chemoprotections
Immunomodulation
Antiangiogenesis

Another local approach uses adenoviral-mediated expression of the tumor suppressor p53, which is mutated in a wide variety of cancers. This strategy has shown complete and partial responses in squamous cell carcinoma of the head and neck, esophageal cancer, and non-small-cell lung cancer after direct intratumoral injection of the vector. Response rates (~15%) are comparable to those of other single agents. The use of oncolytic viruses that selectively replicate in tumor cells but not in normal cells has also shown promise in squamous cell carcinoma of the head and neck and in other solid tumors. This approach is based on the observation that deletion of certain viral genes abolishes their ability to replicate in normal cells but not in tumor cells. An advantage of this strategy is that the replicating vector can proliferate and spread within the tumor, facilitating eventual tumor clearance. However, physical limitations to viral spread, including fibrosis, intermixed normal cells, basement membranes, and necrotic areas within the tumor, may reduce clinical efficacy. Oncolytic viruses are licensed and available in some countries, but not in the United States.

Because metastatic disease rather than uncontrolled growth of the primary tumor is the source of mortality for most cancers, there has been considerable interest in developing systemic gene therapy approaches. One strategy has been to promote more efficient recognition of tumor cells by the immune system. Approaches have included transduction of tumor cells with immune-enhancing genes encoding cytokines, chemokines, or co-stimulatory molecules. Sustained clinical responses provide evidence that the transduced cells can act as a vaccine. In a related approach, patient lymphocytes have been transduced with genes encoding a T cell receptor–like molecule, with a tumor antigen–binding domain fused to an intracellular signaling domain to allow T cell activation, thereby converting normal lymphocytes into cells capable of recognizing and destroying tumor cells. A third immunotherapy approach relies on ex vivo manipulation of dendritic cells to enhance the presentation of tumor antigens. These immunologic approaches may be of particular value in treating minimal residual disease after other anticancer modalities.

Gene transfer strategies have also been developed for inhibiting tumor angiogenesis. These have included constitutive expression of angiogenesis inhibitors such as angiostatin and endostatin; use of siRNA to reduce levels of VEGF or VEGF receptor; and combined approaches in which autologous T cells are genetically modified to recognize antigens specific to tumor vasculature. These studies are still in early-phase testing.

Another novel systemic approach is the use of gene transfer to protect normal cells from the toxicities of chemotherapy. The most extensively studied of these approaches has been transduction of hematopoietic cells with genes encoding resistance to chemotherapeutic agents, including the multidrug resistance gene *MDRI* or the gene encoding O^6-methylguanine DNA methyltransferase (*MGMT*). Ex vivo transduction of hematopoietic cells, followed by autologous transplantation, is being investigated as a strategy for allowing administration of higher doses of chemotherapy than would otherwise be tolerated.

Gene Therapy for Vascular Disease

The third major category addressed by gene transfer studies is cardiovascular disease. The most extensive experience has been in trials designed to increase blood flow to either skeletal (critical limb ischemia) or cardiac muscle (angina/myocardial ischemia). Initial treatment options for both of these groups include mechanical revascularization or medical management, but a subset of patients are not candidates for, or fail, these approaches. These patients have formed the first cohorts for evaluation of gene transfer to achieve therapeutic angiogenesis. The major transgene used has been VEGF, attractive because of its specificity for endothelial cells; other transgenes have included fibroblast growth factor (FGF) and hypoxia-inducible factor 1, α subunit (HIF-1α). The design of most of the trials has included direct IM (or myocardial) injection of either a plasmid or an adenoviral vector expressing the transgene. Both of these vectors are likely to result in only short-term expression of VEGF. This strategy may be adequate, however, as there is no need for continued transgene expression once the new vessels have formed. Direct injection favors local expression, which should help to avoid systemic effects such as retinal neovascularization or new vessel formation in a nascent tumor. Initial trials of adeno-VEGF or plasmid-VEGF injection have resulted in improvement over baseline in terms of frequency of claudication/angina or amounts of nitroglycerin consumption. Study designs including placebo control groups and more objective endpoints (exercise duration at 3 or 6 months, rest and stress cardiac perfusion scans, and regional wall motion assessed by nonfluoroscopic electroanatomic mapping) continue to suggest a beneficial effect of gene transfer, although definitive conclusions will require larger studies. Continuing areas of investigation include choice of the optimal vector (adenoviral vs. plasmid), the optimal transgene (VEGF, HIF-1α, FGF, etc.), the optimal method of delivery in cardiac indications (intracoronary vs. direct myocardial), ideal objective endpoints, and whether concurrent administration of cytokines to mobilize endothelial progenitor cells will augment the therapeutic effect.

Other Diseases

The power and versatility of gene transfer approaches are such that there are few serious disease entities for which gene transfer therapies are *not* under development. Besides those already discussed, other areas of interest include gene therapies for HIV and for neurodegenerative disorders. The latter include studies in patients with Parkinson's disease, where AAV vectors expressing enzymes required for enhanced production of dopamine, or of the inhibitory neurotransmitter γ-aminobutyric acid, have been introduced into affected areas of the brain (striatum, subthalamic nucleus) by stereotactic neurosurgery. In Alzheimer's disease, an ex vivo approach in which autologous fibroblasts are transduced with a retroviral vector expressing nerve growth factor, then reimplanted into the basal forebrain, has slowed the rate of cognitive decline in a small phase I study.

Summary

The development of new classes of therapeutics typically takes two to three decades; monoclonal antibodies and recombinant proteins are recent examples. Gene therapeutics, which entered clinical testing in the early 1990s, are well along in the course of development, and are likely to become increasingly important as a therapeutic modality in the twenty-first century. A central question to be addressed is the long-term safety of gene transfer, and regulatory agencies have mandated a 15-year follow-up for subjects enrolled in gene therapy trials **(Table 68–4)**. Realization of the therapeutic benefits of the Human Genome Project, and of new discoveries such as RNAi, will depend on continued progress in gene transfer technology.

Table 68–4 Taking History from Subjects Enrolled in Gene Transfer Studies

Elements of History for Subjects Enrolled in Gene Transfer Trials
1. What vector was administered? Is it predominantly integrating [retroviral, lentiviral, herpesvirus (latency and reactivation)], or nonintegrating (plasmid, adenoviral, AAV)?

2. What was the route of administration of the vector?
3. What was the target tissue?
4. What gene was transferred in? A disease-related gene? A marker?
5. Were there any adverse events noted after gene transfer?
Screening Questions for Long-Term Follow-Up in Gene Transfer Subjects*
1. Has a new malignancy been diagnosed?
2. Has a new neurologic/ophthalmologic disorder, or exacerbation of a preexisting disorder, been diagnosed?
3. Has a new autoimmune or rheumatologic disorder been diagnosed?
4. Has a new hematologic disorder been diagnosed?

*Factors influencing long-term risk include: integration of the vector into the genome; vector persistence without integration; and transgene-specific effects.

FURTHER READINGS

Aiuti A et al: Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med 360:447, 2009[PMID: 19179314] [Full Text]

Breitbach CJ et al: Navigating the clinical development landscape for oncolytic viruses and other cancer therapeutics: No shortcuts on the road to approval. Cytokine Growth Factor Rev 21:85, 2010[PMID: 20472490] [Full Text]

Cartier N et al: Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science 326:818, 2009[PMID: 19892975] [Full Text]

Gene Therapy Clinical Trials Worldwide. J Gene Med. John Wiley & Sons, 2009. Accessed at *http://www.wiley.co.uk/genetherapy/clinical/*

Hacein-Bey-Abina S et al: LMO2-associated clonal T cell prolifer-ation in two patients after gene therapy for SCID-XI. Science 302:415, 2003[PMID: 14564000] [Full Text]

Maguire AM et al: Age-dependent effects of REP65 gene therapy for Leber's congenital amaurosis: A phase 1 dose escalation trial. Lancet 374:1597, 2009[PMID: 19854499] [Full Text]

Manno CS et al: Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med 12:342, 2006[PMID: 16474400] [Full Text]

Sadelain M: T-cell engineering for cancer immunotherapy. Cancer 15:451, 2009[PMID: 20010162] [Full Text]

Tongers GR, Losordo DW: Human studies of angiogenic gene therapy. Circ Res 105:724, 2009[PMID: 19815827] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

As medicine therapy advances, we can now realistically set goals for fully functional living replacement of nearly every diseased organ or tissue. In the past several decades, the limitations of nonliving mechanical solutions to organ and tissue dysfunction are now recognized and include dialysis, mechanical heart valves, metallic orthopedic implants, and nonresorbable hernia mesh. Tissue engineering has broad goals including organ development, the elimination of the waiting time for transplants such as the liver and kidney, the creation of living tissue replacements for soft tissue, bone, cartilage, fascia, and virtually every structure in the body.

MATERIALS AND SCAFFOLDS FOR TISSUE ENGINEERING

The principal mechanical supporting structure of any engineered tissue is the scaffold. These three-dimensional constructs are often composed of several materials and support the living cells required to generate a functional tissue **(Fig. 69-1)**. The mechanical properties of the scaffolds, such as strength and elasticity, must correspond to the mechanical properties of the target tissue. Moreover, cells respond to environmental cues such that the scaffold should mimic the target tissue to achieve the desired cell alignment and the three-dimensional arrangement of the cells.

Figure 69-1

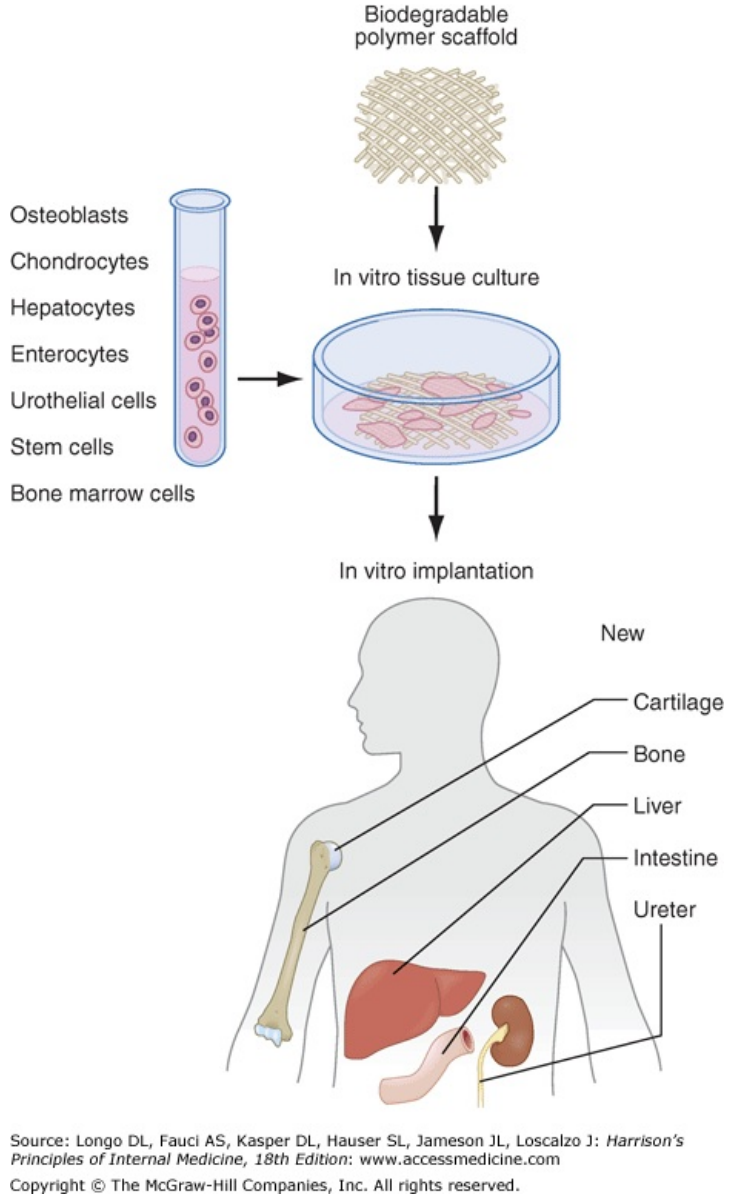

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic of basic principles of tissue engineering. (*From R Langer, J Vacanti: Science 260:1993; with permission.*)

The ideal scaffold materials for engineered tissues are resorbable materials that break down over time. During resorption, the engineered tissue is remodeled by normal healing processes, leaving only living cellular tissue with natural supporting connective tissue. Many implants or organ assist devices are under development utilizing the principles of tissue engineering with nonresorbable materials. These technologies such as lung assist devices, liver assist devices, and even composite implants with resorbable and nonresorbable components for orthopedic and hernia repair represent an important step toward developing fully resorbable scaffolds for all tissues. Current research is focused on resorbable synthetic polymers (e.g., polyglycolic acid, resorbable polyurethanes, polyglycerol sebacic acid); naturally occurring polymers (e.g., collagens, fibrin); and minerals (e.g., calcium triphosphate). Scaffold materials can be supplemented with growth factors or other cytokines to improve cellular incorporation and differentiation. Other surface modifications including surface texturing and protein or antibody coatings may improve cellular adhesion and migration.

Decellularized connective tissue, or organs, represents a rapidly developing area of materials for engineered tissues. Decellularized hearts reseeded with myocardial cells have been demonstrated to function in vitro, and research is focused on expanding this approach to other organs such as the lung and liver as well as soft tissues using material such as decellularized dermis.

CELLS FOR TISSUE ENGINEERING

The cellular components of engineered tissues must achieve organ functionality (e.g., hepatocytes), maintain organ structure (e.g., fibroblasts and stromal cells), and deliver blood to the tissue (e.g., endothelial cells and pericytes). The three-dimensional arrangement of the cells is critical for proper function; therefore, coordination of scaffold design and cellular components is essential. Maintenance of appropriate cell-to-cell interactions and cell-to-extracellular matrix interactions establishes the

microenvironment that drives the differentiation, proliferation potential, and function of many cell types.

Using autologous cells is a principal goal when developing any engineered tissue. This avoids the tremendous burden of immunosuppression. Induced pluripotent stem cells and adult mesenchymal stem cells (e.g., bone marrow or adipose derived) are two promising autologous progenitor cells sources. Differentiation cues including soluble signals and incorporating genes can drive the progenitor cells into cell lineages and end-organ cells (Chaps. 67 and 68). Finally, achieving an adequate number of cells is important. For example, there are approximately 5×10^{10} hepatocytes in a normal adult liver. Therefore, any cellular differentiation strategy must take into account the number of cells needed to seed the scaffold for appropriate tissue function.

Allogenic cells remain an option for many tissues and have been the principal source of clinically successful engineered tissues to date. Embryonic stem cells continue to hold great promise as an allogenic cell option. Xenogenic cells are important to consider as they are available in large numbers, but their use has been mostly limited to organ support devices (i.e., liver devices) and they are not directly implanted into a patient.

IN VITRO MATURATION OF ENGINEERED TISSUES

Nearly all tissues have baseline mechanical requirements and many tissues such as heart valves, blood vessels, bone, and tendons must have adequate mechanical properties to achieve function. Mechanical forces are important to induce cell alignment and in the production of an extracellular matrix. Bioreactors have been developed to impart the needed mechanical forces to engineered tissues, including shear stress, pusatile flow, and pressure for valves and blood vessels, and axial tension and compression for bone, cartilage, and tendons.

CLINICAL APPLICATIONS

Engineered skin substitutes were the first true clinical success of the principles of tissue engineering **(Table 69–1)**. These clinically available products use autologous fibroblasts grown on a resorbable polymer scaffold for a single-layer product or fibroblasts covered with keratinocytes for a two-layer product. Recent products for promoting the healing of skin and dermal wounds also incorporate autologous fibroblasts delivered into the wound. Autologous chondrocytes are being used to heal damaged joints, and they have shown excellent results. Acellular collagen-based materials derived from human or animal dermis are being implanted for soft tissue reconstruction or hernia repair and become cellularized in vivo with autologous cells. Many engineered tissues are in clinical trials **(Table 69–2)** or under development, including those for bone, cartilage, nerve, skeletal muscle, small-diameter blood vessels, heart valves, and vital organs, including heart, liver, lung, and kidney.

Table 69–1 Examples of FDA-Approved Tissue Engineering Products

Alloderm (LifeCell)	Acellular dermal matrix for tissue repair
Apligraf (Organogenesis)	Living skin equivalent approved for the treatment of venous leg ulcers and diabetic foot ulcers
Carticel (Genzyme Biosurgery)	Autologous chondrocytes approved for cartilage repair
Dermagraft (Smith and Nephew)	Living skin equivalent approved for full-thickness diabetic foot ulcers
Durasis (Cook Surgical Products)	Porcine small-intestine submucosa for replacement of dura mater
Epicel (Genzyme Biosurgery)	Living skin equivalent approved for burn patients
OrCel (Ortec)	Living skin equivalent approved for burn patients
Surgisis (Cook Surgical Products)	Porcine small-intestine submucosa for dermal wounds and reinforcement of weakened tissue
FlexHD (Ethicon)	Acellular dermal matrix for tissue repair
Strattice (LifeCell)	Acellular dermal matrix for tissue repair
DermaMatrix (Synthes)	Acellular dermal matrix for tissue repair

Table 69–2 Tissue Engineering Products in Clinical Trials

Dermagen (Laboratoires Genévrier)	Living skin equivalent
Lifeline (Cytograft)	Living small-diameter blood vessel
Hepatocyte Matrix Implant (Baermed)	Resorbable matrix and autologous hepatocytes grafted into omentum to support liver function

TRC (Aastrom)	Autologous adult bone marrow cells for bone grafting
Myocell (Bioheart)	Encapsulated autologous muscle cells for myocardial infarction
MarkII (Excorp)	Extracorporeal liver assist device
HuCNS-SC (Stem Cell Inc)	Human central nervous system stem cells
NT-501 (Neurotech SA)	Encapsulated cell technology for long-term delivery of therapeutic factors to retina
Procord (Proneuron)	Autologous activated macrophage therapy for patients with acute complete spinal cord injury
ChondroCelect (Tigenix)	Autologous chondrocyte implantation
ELAD (Vigagen)	Extracorporeal liver assist device
Neo-Bladder (Tengion)	Living engineered bladder

CHALLENGES AHEAD

Thin tissues such as skin, cartilage, heart valves, and blood vessels are advantageous in that they can survive on diffusion of oxygen and nutrients, while new blood vessels sprout for permanent tissue formation. Thick tissues and solid organs depend on an elaborate network of blood vessels that course within 150–200 μm of each cell in that tissue. The heart requires so much oxygen that often there are blood vessels on both sides of an individual cardiomyocyte. Development of blood vessel networks within tissues is the principal challenge for engineered tissues. Angiogenesis can be used to create capillary networks for musculoskeletal and soft tissues while engineered vascular networks utilizing microfluidic principles are being developed for solid organs such as the liver and lung.

In vitro expansion and differentiation of progenitor cells to tissue-specific cells also remains an important milestone. This includes optimization of the different cell types and ratios for each tissue for long-term tissue function. Conditions necessary for in vitro culture and preimplant logistics such as delivery and storage are just beginning to be addressed for complex tissues.

FURTHER READINGS

Badylak SF et al: Extracellular matrix as a biological scaffold material: Structure and function. Acta Biomater 5:1, 2009[PMID: 18938117] [Full Text]

Grayson WL et al: Biomimetic approach to tissue engineering. Semin Cell Dev Biol 20:665, 2009[PMID: 19146967] [Full Text]

Hoganson DM et al: Tissue engineering and organ structure: A vascularized approach to liver and lung. Pediatr Res 63:520, 2008[PMID: 18427297] [Full Text]

Khademhosseini A et al: Progress in tissue engineering. Sci Am 300:64, 2009[PMID: 19438051] [Full Text]

Ott HC et al: Perfusion-decellularized matrix: Using nature's platform to engineer a bioartificial heart. Nat Med 14:213, 2008[PMID: 18193059] [Full Text]

Sacks MS et al: Bioengineering challenges for heart valve tissue engineering. Annu Rev Biomed Eng 11:289, 2009[PMID: 19413511] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

WORLD DEMOGRAPHY OF AGING: INTRODUCTION

Population aging is transforming the world in dramatic and fundamental ways. The age distributions of populations have changed and will continue to change radically, due to long-term declines in fertility rates and improvements in mortality rates **(Table 70-1)**. This transformation, known as the Demographic Transition, is also accompanied by an epidemiological transition, in which noncommunicable chronic diseases are becoming the major causes of death and contributors to the burden of disease and disability. A concomitant of population aging is the change in key ratios expressing "dependency" of one form or another—the ratio of adults in the workforce to those typically out of the workforce, such as infants, children, retired "young old" (those still active in many ways other than paid work) and the oldest old. Global aging will affect economic growth, migration, patterns of work and retirement, family structures, pension and health systems, and even trade and the relative standing of nations. Both numbers (the size of an age group) and ratios (the ratio of those in the labor force age group to dependants such as the young or retired, or the ratio of children to older people) are important. The size of population age groups might affect the number of hospital beds needed, while the ratio of children to older people, the relative demand for pediatricians and geriatricians.

Table 70-1 Selected Indicators of Population Aging, Estimates for 2009 and Projections to 2050; Selected Regions and Countries

	Population Age 60+ (in millions)		Percentage of Population 60+		Life Expectancy at Birth		Life Expectancy at 60		Old Age Support Ratio[a]	
	2009	2050	2009	2050	Male	Female	Male	Female	2009	2050
World	737	2008	11	22	65.4	69.8	18.1	21.2	9	4
More Developed Regions	264	416	21	33	73.6	80.5	19.6	23.7	4	2
Less Developed Regions	473	1592	8	20	63.9	67.4	17.3	19.6	11	4
Least Developed Countries	43	185	5	11	54.7	57.2	15.3	16.8	17	9
Africa	53.8	212.8	5	11	52.9	55.3	15.2	17.1	16	9
Egypt	6.0	24.8	7	19	68.3	71.8	16.2	18.3	14	5
Tanzania	2.1	8.4	5	8	54.6	56.2	16.1	17.6	17	13
Kenya	1.6	7.9	4	9	53.7	54.5	16	17.8	21	11
Nigeria	7.6	27.7	5	10	47.3	48.3	14.5	15.7	17	11
South Africa	3.6	8.1	7	14	49.9	53.2	13.6	17.8	14	7
Asia	399.9	1236.1	10	24	67.1	75.7	17.6	20.3	10	4
China	160.2	440.4	12	31	71.3	74.8	18.2	20.7	9	3
Japan	37.8	44.9	30	44	79	86.2	22.1	27.8	3	1
Rep. of Korea	7.3	18.0	15	41	75.9	82.5	19.7	24.6	7	2
India	88.6	315.6	7	20	62.1	65	16	17.9	13	5
Pakistan	11.1	49.8	6	15	66	66.7	17.6	18.4	15	7
Indonesia	20.2	71.6	9	25	68.7	72.6	16.5	18.5	11	3
Australia	4.1	8.5	19	30	79.1	83.8	22.3	25.8	5	3
New Zealand	0.8	1.6	18	29	78.2	82.2	21.6	24.7	5	3
Europe	158.5	236.4	22	34	71.1[b]	79.1[b]	18.3	22.6	4	2
Russian	25.0	36.8	18	32	60.3	73.1	14.3	19.2	6	3

Federation										
United Kingdom	13.8	20.9	22	29	77.2	81.6	20.4	24	4	3
Italy	15.8	22.3	26	39	78.1	84.1	21.5	25.9	3	2
Germany	21.1	27.9	26	40	77.1	82.4	20.4	24.6	3	2
France	14.1	22.0	23	33	77.6	84.7	21.6	26.6	4	2
Americas										
United States	56.2	110.5	18	27	76.9	81.4	21.2	24.6	5	3
Canada	6.5	14.1	20	32	78.3	82.9	21.7	25.2	5	2
Mexico	10.0	36.4	9	28	73.8	78.6	20.6	22.9	10	3
Argentina	5.9	12.7	15	25	71.6	79.1	18.1	23	6	3
Brazil	19.1	64.0	10	29	68.7	76	19.5	22.4	10	3

[a]UN Population Division defines Old Age Support Ratio as the number of persons aged 65 or older per one hundred persons aged 15 to 64 years.

[b]The UN includes all European regions in its overall statistics; life expectancy at birth for males ranges from 63.8 years in Eastern Europe to 77.4 years in Western Europe. For women it ranges from 74.8 to 83.1 years in Western Europe.

Source: United Nations Population Division, *World Population Ageing* 2009.

While the increase in life expectancy, resulting from a series of social, economic, public health, and medical victories over disease, might very well be considered the crowning achievement of the past century and a half, the increased length of life coupled with the shifts in the size of dependent groups relative to the size of the labor force, present formidable long-term challenges.

The pace of the change is accelerating. In countries where the Demographic Transition began earlier, the process was slower: it took France 115 years for the proportion of the age group 65 and older to increase from 7 to 14% of the total population, and the United States will soon have completed this same increase in 69 years. But in countries that started the transition later, the process is occurring much more rapidly: Japan took 26 years to go from 7 to 14% age 65 and older, while China and Brazil are projected to require just 24 years.

Sometime around the year 2020, for the first time ever, the number of people aged 65 and older in the world is expected to exceed that of children under the age of 5. Around the middle of the twentieth century, the under-5 age group constituted almost 15% of the total population and the over-65 age group 5%. It took about 70 years for these two to reach equal proportions. But population forecasts predict it will take only another 30 years for the 65 and older age group to equal about 15% and the younger age group, 5%. By the middle of their careers, medical students in most countries should expect to be practicing in far older populations. Preparations for these changes need to begin decades in advance, and the costs and penalties of waiting can be very high. While some governments have started planning for the long term, many, if not most, have yet to begin.

HISTORICAL

Population aging around the world in recent decades has followed a broadly similar pattern, starting with a decline in infant and childhood mortality that precedes a decline in fertility; at later stages, mortality at older ages declines as well. Declining fertility began as early as the beginning of the nineteenth century in the United States and France and extended to the rest of Europe and North America and parts of East Asia by the middle of the twentieth century. Since World War II, fertility declines have started in all other world regions. In fact, more than half the world's population now lives in countries or provinces with fertility rates below the replacement level of just over two live births per woman. Mortality rates also began to change, relatively slowly at first, in Western Europe and North America during the nineteenth century. At first, changes were most evident at the youngest ages. Improvements in water supply and sewage handling, as well as in nutrition and housing, accounted for most of the improvement before the 1940s, when antibiotics and vaccines, and increasing education of mothers began to make a major impact. Since the middle of the twentieth century, the "Child Survival Revolution" has spread to all parts of the world. Children almost everywhere in the world are much more likely to reach late middle age now than in previous generations.

Especially since around 1960, mortality at older ages has improved rapidly in most of the developed countries. This improvement has been primarily due to advances in care of heart disease and stroke and in control of conditions like hypertension and hypercholesterolemia that lead to circulatory diseases. In some parts of the world, smoking rates have declined, and these declines have led to lower incidence of many cancers, heart disease, and stroke.

The initial decline in fertility resulted in older age groups becoming a larger fraction of the total population. Declines in adult and old age mortality contributed to the later stages of the process. Life

expectancy at birth—the average age to which someone is expected to live, under prevailing mortality conditions—has been calculated at around 28 years in ancient Greece, perhaps 30 years in medieval Britain, and less than 25 years in the colony of Virginia in North America. In the United States, life expectancy climbed slowly during the nineteenth century, reaching 49 years for white women by 1900. White men had a life expectancy 2 years lower than that for white women, and black Americans had life expectancy 14 years lower than did white Americans in 1900. By the early twenty-first century, life expectancy in the United States had improved dramatically for all, with the sex gap wider and the racial gaps narrower than at the beginning of the century: 76 years for white men in 2006; 81 years for white women; and 70 and 76 years for black men and women, respectively.

At later stages of the demographic transition, mortality declines at the oldest ages, leading to increases in the 65 and older population, and the oldest old, those older than age 85 years. Migration can also affect population aging. An influx of young migrants with high birth rates can slow (though not stop) the process, as it has in the United States and Canada; or the out-migration of the young leaving older people behind can accelerate aging at the population level, as it has in many rural areas of the world. Wars and pandemic diseases such as AIDS can also change age composition by decimating particular adult age groups.

REGIONAL AGING–NUMBERS AND PERCENTAGES OLDER THAN AGE 60 YEARS

Regions of the world are at very different stages of the demographic transition **(Fig. 70-1)**. Of a world population of 6.8 billion in 2009, approximately 11% were older than age 60 years, with Japan (30%) and Europe (22%) being the oldest regions (Germany, Italy, and Sweden have the highest percentages, 25–26%), and the United States and Canada having 21% and 20%, respectively. The percentage of the population older than age 60 years in the United States has remained lower than in Europe, due both to modestly higher fertility rates and to higher rates of immigration. The Caribbean and some Latin American countries average 10% older than age 60 years, with countries such as Uruguay, Cuba, and Argentina in the 15–18% range. Asia has about 10% older than age 60 years, with the population giants close to the average—China (12%), Indonesia (9%), and India (7%). Middle Eastern and African countries have the lowest proportions of older people (5% or lower).

Figure 70-1

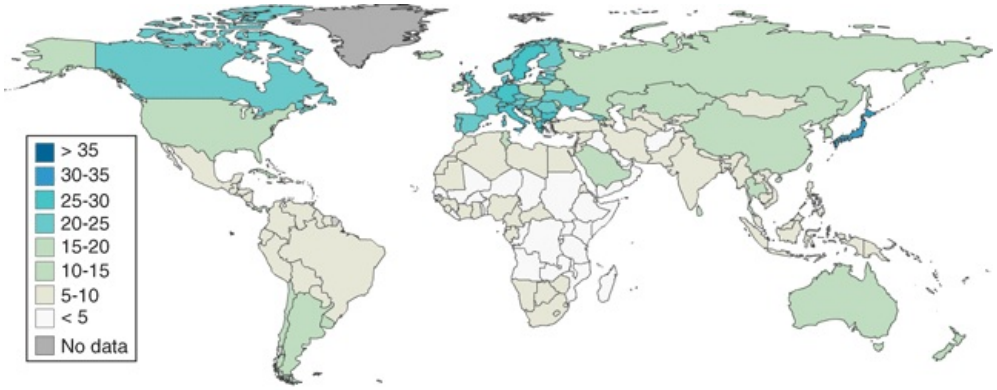

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Percentages of national populations aged 60+, in 2010. *(From the US Census Bureau, International Database. StatPlanet Mapping Software.)*

Based on estimates from the United Nations Population Division, 737 million people were older than age 60 years in 2009, of whom 264 million lived in more developed countries, 473 million in less developed countries (43 million of the latter living in the least developed countries, as classified by the United Nations). The countries with the largest populations of those aged 60 and older were China (160 million), India (89 million), and the United States (56 million).

NUMBERS–POPULATION SIZE PROJECTIONS

Population projections make use of expected fertility, mortality, and migration rates and should be regarded as uncertain when applied 40 or more years in the future. However, the population that will be aged 60 and older in 2050 have all been born and survived childhood in 2010, so uncertainty about their numbers (as distinct from their proportion of the total population) is not great. Comparing the maps of the world in 2010 (Fig. 70-1) and 2050 **(Fig. 70-2)**, it is apparent that the middle- and low-income countries in Latin America, Asia, and much of Africa will be joining the "oldest" category. In the four decades between 2010 and 2050, the United Nations Population Division projects that the world population aged 60 and older will almost triple to 2.01 billion, with the least developed regions more than quadrupling. China's 60+ population is projected to reach 440 million, India's 316 million, and the

United States's 111 million. Over the next 40 years, the median age of the world's population is expected to increase by 10 years.

Figure 70-2

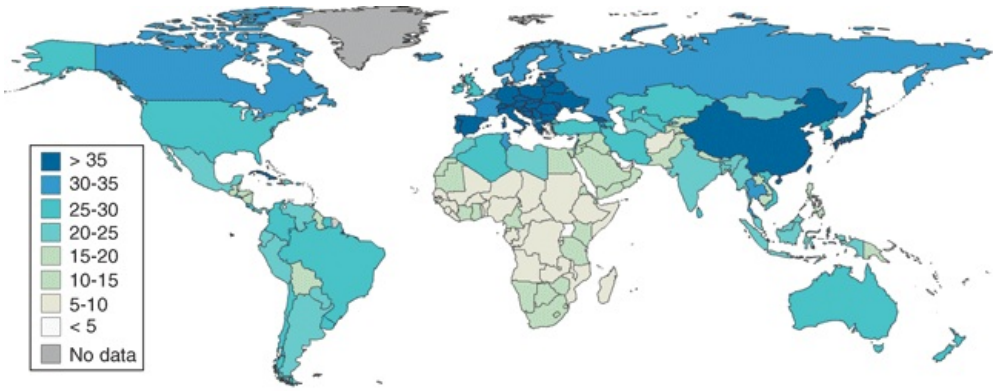

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Percentages of national populations aged 60 +, in 2050 (projections). *(From the US Census Bureau, International Database. StatPlanet Mapping Software.)*

Current global life expectancy at birth is estimated to be 65.4 for men and 69.8 for women, with the comparable figures for the more developed region being 73.6 and 80.5 years. Life expectancy in the least developed countries averaged only 57.2 for women and 54.7 for men. Life expectancy at birth is heavily influenced by infant and child mortality, which is considerably higher in poor countries. At older ages, the gap between rich and poor nations is narrower; so while women who have reached age 60 in wealthy countries can expect 23.7 more years of life on average, women at age 60 in poor countries live 16.8 years on average—a significant difference but not so stark as the difference in life expectancy at birth. At the lowest levels of per capita GNP, life expectancy shows a powerful positive association with this measure of economic development but then the slope of the relationship flattens out; for countries with average incomes above about $20,000 per years life expectancy is not closely related to income. At each level of economic development, there is significant variation in life expectancy, indicating that many other factors influence life expectancy.

Japan, France, Italy, and Australia currently have life expectancy among the highest in the world, while the United States has lagged behind other high-income countries since about 1980, especially in the case of white women. The causes of this lag are being explored, but the cumulative number of years that people have smoked tobacco by the time they reach older ages appears to play an important role.

GROWTH OF THE OLDEST OLD POPULATION–THOSE OVER AGE 85

A modern feature of population aging has been the almost explosive growth of the age group known as the oldest old, variously defined as those over age 80 or age 85. This is the age group with the highest burden of noncommunicable degenerative disease and related disability. Thirty years ago, this group attracted little attention because they were hidden within the overall older population in most statistical reports; for example, the U.S. Census Bureau merged them into a 65+ category. The reduction of mortality at older ages coupled with larger birth cohorts surviving into old age led to the rapid growth of the oldest old. This age group is predicted to grow at a significantly higher rate than the 60+ population and one estimate has the current 102 million aged 80+ increasing to almost 400 million by 2050 **(Table 70-2)**. Projected increases are astounding: China's 80+ population might increase from 18 to 101 million, India from 8 million to 43 million, the United States from 12 to 32 million, and Japan from 8 to 16 million. The numbers of centenarians are increasing at an even faster rate.

Table 70-2 Estimates (2009) and Projections (2050) for the Population Aged 80 Years and Older: Selected Regions and Countries

	Population Age 80 and Over (in thousands)	
	2009	**2050**
World	101,873	394,663
More Developed Regions	51,827	120,982
Less Developed Regions	50,046	273,771
Least Developed Countries	3,282	18,221

Africa	4,136	21,042
Kenya	151	686
Congo	20	65
Tanzania	164	810
South Africa	277	1,226
Asia	45,235	227,916
China	18,463	101,354
Japan	7,693	15,864
Rep. of Korea	895	5,618
India	7,754	42,583
Pakistan	1,041	6,156
Indonesia	1,670	11,515
Australia	814	2587
New Zealand	147	472
Europe	29,994	66,147
Russian Federation	3,986	6,984
United Kingdom	2,869	6,235
Italy	3,486	7,670
Germany	4,096	9,926
France	3,324	7,634
Americas		
United States	11,894	31,501
Canada	1,282	4,302
Mexico	1,477	7,944
Argentina	1,030	2,692
Brazil	2,754	14,133

Source: United Nations Population Division, *World Population Ageing* 2009.

THE FUTURE OF LIFE EXPECTANCY

The members of the population who could potentially become aged 80 and older in 2050 are those aged 40+ who are alive today. The actual numbers who will be age 80 and older in 2050 will therefore depend almost solely on adult and old age mortality rates over the next 40 years. The history of the decline of mortality and increase in life expectancy suggests that improvements in the standard of living including increased and improved education and improved nutrition coupled with improvements in public health stemming from an understanding of the germ theory of disease initially led to the improvements, with medical achievements such as antibiotics and improved understanding of risk factors for cardiovascular and circulatory diseases becoming factors only in the post–World War II period with the largest strides for cardiovascular disease coming only in more recent decades. The improvements in educational attainment of succeeding generations have been credited in large part for improvements in child mortality during the past century, since educated mothers are especially likely to understand and take advantage of measures to reduce infection. The effects of continuing progress will likely be seen in coming decades as well, since educational attainment is associated with improved health and survival at older ages. Countries vary in the extent to which the "future elderly" cohorts will be more educated. China in particular will have a much more educated elderly population in 2050 (with more than two-thirds of the 65+ population having completed secondary school) than it did in 2000 (when only 10% of older people had a secondary education). In the United States and other rich nations, this change has largely taken place already; future changes in educational attainment of the elderly population will be less dramatic.

Holding aside the possibility of new infectious diseases ravaging populations as AIDS did in some African countries, debates about future life expectancy revolve around the balance and influence of risk factors such as obesity, the possibility of reducing the deaths from current killers such as cancer, heart disease, and diabetes, whether there is some natural limit to life expectancy, and the distant though nonzero possibility that science will find a way to slow the basic processes of aging.

While some have posited natural limits to human life expectancy, the limits have been surpassed with some regularity and at the very oldest ages in the leading countries with the highest life expectancy, there appears to be little evidence of any approaching asymptote. Indeed a surprising discovery was that life expectancy in the leading country over the last century and a half, with different countries taking the lead in different epochs, could be represented almost perfectly by a straight line, with the

increase for females showing a steady and astonishing increase of three months per year or 2.5 years per decade **(Fig. 70-3)**. No single country kept that pace of improvement the entire time, but this trend does call into question the notion that improvement must slow down, at least in the near future.

Figure 70-3

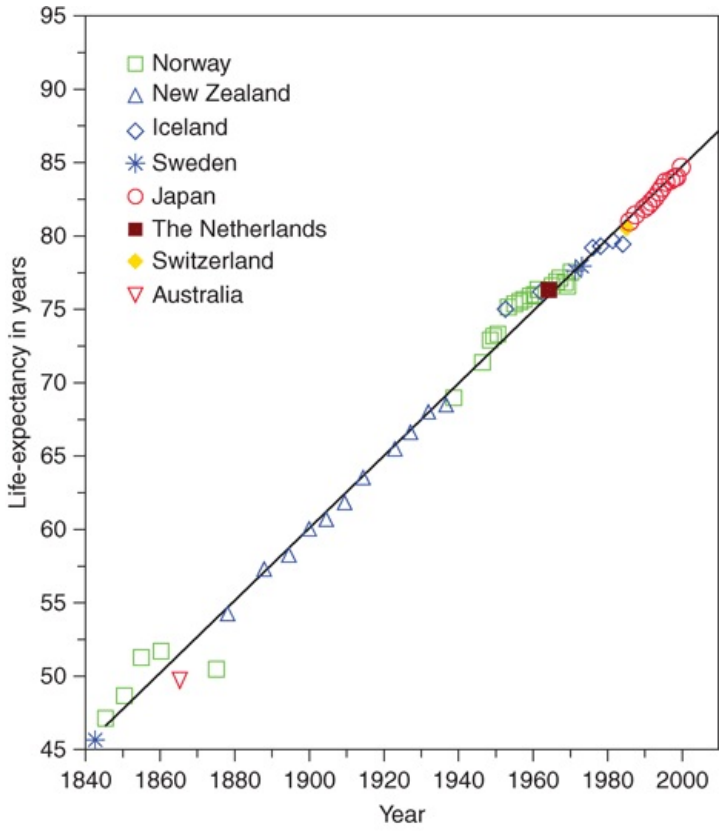

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Life expectancy in most advanced nations, 1800-2000, females. *(From J Oeppen, JW Vaupel: Science 296:1029, 2002.)*

There remains a great deal of diversity in health conditions both among and within national populations. There is nothing inevitable about the mortality transition—in several African countries, the prevalence of AIDS has been high enough to cause life expectancy to fall below the levels of 1980. Though none has so far reached a scale to rival the AIDS epidemic, periodic outbreaks of new influenza viruses or "emerging infectious" agents remind us that infectious diseases could again come to the fore. Progress against chronic disease is also reversible: In Russia and some other countries that formed part of the Soviet Union before 1992, life expectancy for men has been declining, now reaching levels below those of men in South Asia. Much of the gap between Russian and Western European men is explainable by much greater heart disease and injuries among the former.

DEPENDENCY AND CAREGIVING RATIOS

Ratios of different age groups provide useful though crude indicators of potential demands on resources and resource availability. One set of ratios, known variously as dependency or support ratios, compare the age groups who are most likely to be in the labor force with the age groups typically dependent on the productive capacity of those working—the young and the old, or just the old. A commonly used ratio is the number of persons aged 15–64 per persons aged 65 and older. Even though many in some countries do not enter the labor force until significantly older than age 15, retire before age 65 or work past age 65, the ratios do summarize important facts, especially in countries where financial support for the retired comes partially or mainly from those currently in the labor force through either a formal pension system or through the family. While many countries, including China and most African countries do not have formal pension systems except in specialized sectors such as the government sector, in Europe public pensions are quite generous, and face dramatic changes in their dependency ratios. Over the next 40 years, Western Europe faces a drop in the ratio from 4 to 2. In other words, while in crude terms there are today 4 workers supporting the pensions and other costs of each older person, by 2050 there will only be 2. China faces an even steeper drop from 9 of working age to only 3, while Japan declines from 3 to just 1. Even in India, projected to become the most populous country, the decline is quite steep from 13 to 5.

The dramatically declining number of workers per older person (however determined) is at the crux of the economic challenge of population aging. The extra years of life that can be considered the crowning achievement in medicine and public health of the last 150 years have to be financed. The economic model of the life cycle assumes that people are economically productive for a limited number of years and that the proceeds of their work during those years have to be smoothed over to finance consumption during less economically productive ages, either within families or by institutions such as the state in order to provide for the young, the old, and the infirm. There are only so many ways to meet the challenge of an extended period of dependency, including increasing the productivity of those in the labor force, saving more, reducing consumption, increasing the number of years worked most especially by increasing the age of retirement, increasing the voluntary nonmonetary productive contributions of the retired, and immigration of very large numbers of young workers into the "old" countries. Pressures to increase retirement ages in industrialized countries and to reduce benefits are increasing. But no single one of these measures can bear the full load of adaptation to population aging, since the changes would have to be so severe and disruptive as to be politically impossible. More likely will be some combination of these measures.

Population health and the ability to function at work and in everyday life interact with these population ratios in significant ways. The physical and cognitive capacity to continue to work at older ages is crucial if the age of retirement is raised. Similarly, caregiving often requires significant physical and emotional stamina. Further, healthier older populations require less caregiving and medical services. Just two decades ago, the prevalent view of aging was highly pessimistic. Epidemiologists held that while modern medicine could keep older people alive, nothing much could be done to prevent, delay, or significantly treat the degenerative chronic diseases of aging. The result would be that more and more older people with chronic diseases would be kept from dying, with the consequent piling up of the older people disabled by chronic disease. Surprisingly, between 1984 and about 2000, the prevalence of disability in the 65+ population in the United States declined by about 25%, suggesting that in this respect, aging was more plastic than had been previously believed **(Fig. 70-4)**. All the causes of this significant shift in disability are not yet understood, but rising levels of education, improved treatment of cardiovascular diseases and cataracts, greater availability of assistive devices, and less physically demanding occupations have been found to contribute. One calculation showed that if the rate of improvement could be maintained until 2050, that the numbers of disabled in the older population could be kept constant despite the aging of the baby boomers and the older population itself growing older. Unfortunately, concern is increasing that the rapid increase in obesity rates will negate and perhaps even reverse this most positive trend. Because of the absence of comparable data in other countries, it is uncertain whether the same improvement in disability rates (or recent deceleration) is occurring outside of the United States, but a global network of longitudinal studies on aging, health, and retirement is now providing data to answer the question.

Figure 70-4

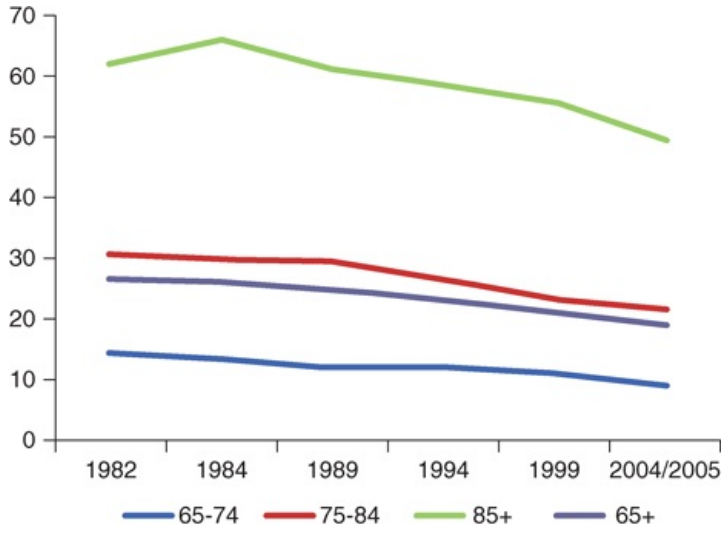

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Disability prevalence, various years 1982-2005, by age group over 65, United States. (*Adapted from KG Manton et al: Proc Natl Acad Sci U S A. 103:18374, 2006.*)

Population aging, and related demographic changes including changes in family structure, could also have powerful though indirect effects through the "supply side" and health care financing. In every country, long-term care of the disabled and the chronically ill relies heavily on informal, typically unpaid

caregivers—usually spouses, children, and, increasingly in more developed countries, caregivers for the oldest old are in their 60s and early 70s. Though there are many men who provide care, on the population level, informal caregiving is still mainly done by women. Because both men and women are living longer than in the past, the "young old" in most countries are less likely than in the past to be widows. But because women live longer than men, lack of a spousal caregiver is especially likely to be a problem for older women. Both men and women have fewer children on whom they can call for informal caregiving, because of the worldwide decline in fertility rates. An increasing proportion of older men in Europe and North America have spent much or all of their adult lives apart from their biological children. Lower fertility rates, delayed marriage, and increasing divorce rates among their children mean that people approaching old age may be less likely to have or to have close ties with daughters and daughters-in-law—the adults who have in the past been the most common caregivers apart from spouses. Adult women who in the past have provided uncompensated care (and much other essential volunteer work) are now more likely than in the past to be working for pay, and thus have fewer hours to devote to the unpaid roles.

These broad demographic and economic trends do not dictate particular social adaptations or policy responses, of course. One can imagine many different responses to the challenges of caring for the disabled—increased reliance on home health agencies and assisted living communities, "naturally occurring retirement communities" in which neighbors fulfill many of the roles once reserved for close kin; private or even publicly financed direct payments to compensate formerly unpaid family caregivers (a reform that has proved very popular in Germany)—these and other responses to the challenge of long-term care are being tested in aging countries, and others will no doubt be needed.

THE EPIDEMIOLOGIC TRANSITION–CHANGES IN THE BURDEN OF DISEASE AND RISK FACTORS

The secular improvements in ages at death have been accompanied by changes in causes of death. In the broadest terms, the proportion of deaths due to infectious disease and conditions associated with pregnancy and delivery has fallen, and the proportion due to chronic, noncommunicable diseases, such as heart and cerebrovascular diseases, diabetes, cancers, and age-related neurodegenerative diseases such as Alzheimer's and Parkinson's diseases, has increased steadily and is expected to continue to increase. **Figure 70-5** shows results from an international comparative project that drew on a wide variety of data sources to provide estimates of the global burden of disease at the beginning of this century, with projections to future years based on recent trends in disease prevalence and demographic rates. Burden of disease in these pie charts is a composite measure, one that takes into account both the number of deaths due to a particular disease or condition and the timing of such deaths—an infant death represents a loss of more potential life-years lived than does the death of a very old person. Nor is death the only outcome that matters; most diseases or conditions cause significant disability and suffering even when nonfatal, so this measure of burden captures nonfatal outcomes using statistical weighting. As Table 70–3 shows, the "modern plagues" of chronic noncommunicable diseases are already among the leading causes of premature death and disability even in low-income countries. This is due to a mix of factors—lower fertility rates mean fewer infants and children at prime ages of susceptibility to infections; more people reaching older ages when chronic disease incidence is high; and often changing incidence rates due to increased exposure to tobacco, Western diets, and inactivity. Noncommunicable diseases, once thought of as "diseases of affluence"—are projected to account for more than half of the disease burden even in low- and middle-income countries by the year 2030 (Fig. 70-5).

Figure 70-5

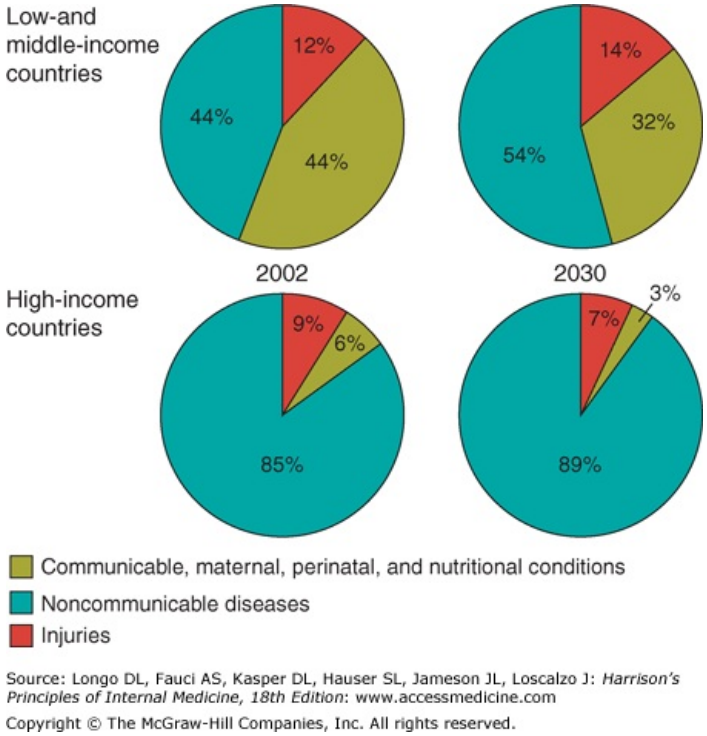

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Leading causes of burden of illness in world regions 2002 and projected for 2030. (*Adapted from Mathers and Loncar.*)

Table 70-3 Diseases and Conditions Causing the Greatest Burden of Disease for Low-, Middle-, and High-Income Countries

	Low-Income Countries	Medium-Income Countries	High-Income Countries
1	Lower respiratory infections	Unipolar depressive disorders	Unipolar depressive disorders
2	Diarrheal diseases	Ischemic heart disease	Ischemic heart disease
3	HIV/AIDS	Cerebrovascular disease	Cerebrovascular disease
4	Malaria	Road traffic accidents	Alzheimer's and other dementias
5	Prematurity and low birth weight	Lower respiratory infections	Alcohol use disorders
6	Neonatal infections and other	COPD	Hearing-loss, adult onset
7	Birth asphyxia and birth trauma	HIV/AIDS	COPD
8	Unipolar depressive disorders	Alcohol use disorders	Diabetes mellitus
9	Ischemic heart disease	Refractive errors	Trachea, bronchus, lung cancers
10	Tuberculosis	Diarrheal disease	Road traffic accidents

Note: "Burden of Disease" takes into account years of life lost due to premature death, and also a weighted estimate for years spent with disability, pain, or impairments due to conditions. These estimates are aggregated from many different national reporting systems and special surveys or surveillance systems, with adjustments for incomplete coverage and different reporting schemes. Especially for older people with multiple chronic conditions, it is difficult even for physicians familiar with their case to designate underlying and precipitating causes. For example, Alzheimer's disease and other dementias are treated very differently and cause-of-death data are of little use in studying trends and differences in dementiaprevalence.

Abbreviation: COPD, chronic obstructive pulmonary disease.

Source: World Health Organization: *The Global Burden of Disease: 2004 Update.* © World Health Organization 2008. Accessed February 28, 2011 at http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html.

SUMMARY

Population aging is a global phenomenon with profound short- and long-term implications for health and long-term care needs, and indeed for the economic and social wellbeing of nations. The timing and context of aging vary across and within world regions and countries; the industrialized nations became wealthy before they aged significantly, while many of the low resource regions will age before they become significantly industrialized. The variation at both the population level and individual levels indicates that there is a significant degree of flexibility in successful aging, but meeting the challenges

will require significant advance planning and preparation. The extent to which research can find solutions that reduce physical and cognitive disability at older ages will be a significant factor in how countries cope with this fundamental transformation of populations and societies.

FURTHER READINGS

Cutler D et al: The determinants of mortality. J Econ Perspectives 20:97, 2006

Kinsella K, He W: US Census Bureau, International Population reports, P95/09-1, in *An Aging World: 2008*. Washington DC, Govt Printing Office 2009

Lopez AD et al: The global and regional burden of disease and risk factors in 2001: Health priorities at the beginning of the millennium. Lancet 367:1747, 2006[PMID: 16731270] [Full Text]

Mathers CD, Loncar D: Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3:e442, 2006

Samir KC et al: Projection of populations by level of educational attainment, age and sex for 120 countries for 2005-2050. Demographic Research 22:Article 15, 2010 (online at www.demographic-research.org)

Schoeni RF et al: Why is late-life disability declining? Milbank Q 86:47, 2008[PMID: 18307477] [Full Text]

United Nations Dept of International Economic and Social Affairs, Population Division (2010). *World Population Ageing 2009*. United Nations, NY, 2010

Vaupel JW: Biodemography of human ageing. Nature 464:536, 2010[PMID: 20336136] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE BIOLOGY OF AGING: INTRODUCTION

Thanks largely to the power of genetic analysis in model organisms such as *Caenorhabditis elegans* (a nematode), *Drosophila melanogaster* (a fruit fly), and the laboratory mouse, major advances have been made in the elucidation of what can be termed "public" modulations of intrinsic biological aging—that is to say, commonalities of gene actions across widely diverse phyla that explain, in part, the plasticity of processes of aging. There are hints that at least one such conserved pathway may be operative in our own species. These observations, together with related research on other biochemical pathways, a long history of research on the beneficial effects of dietary restriction (most recently including an initial report of its beneficial effects on healthspan and lifespan in a primate) (**Fig. 71-1**), and spectacular advances in genomics raise the possibility that we may one day be able to delay the times of onset and decrease the rates of progression of aging processes. Such interventions have the potential to extend the healthspans and, therefore, the functional lifespans of a large proportion of our population. This new knowledge, however, is still very distant from clinical translation. Many remain skeptical of the relevance of these experimental findings. Moreover, we need much more information on the pathophysiology of aging, especially in the invertebrate models that have provided most of our new knowledge concerning genetic modulations of lifespan. We will also require more detailed information on the impact of longevity enhancements upon what can be described as the "terminal decline" of the life course, the stage of life in humans responsible for protracted morbidity, frailty, and the consequent loss of the ability to live independently. These terminal declines account for a very substantial proportion of all health care costs. Finally, the promising new knowledge needs fuller discussions by ethicists, economists, sociologists, and political scientists, among others, as to the impacts upon society of any large-scale clinical translations.

Figure 71-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Photographs of an old (age 27.6 years) calorically restricted (CR) Rhesus monkey (A, B) compared to an age-matched control (C, D). The mean lifespans of control animals in captivity is 27 years and the maximum lifespan is about 40 years. CR (30% of ad libitum fed controls) was initiated as adults (ages 7–14 years). Restricted animals exhibited superior insulin sensitivity, less fat, more muscle mass, and fewer diseases (diabetes, neoplasia, cardiovascular, brain atrophy). Preliminary analyses of survival curves were consistent with enhanced longevity of the CR cohort. (*Reproduced from RJ Colman et al: Science 235:201, 2009; with permission.*)

Definitions of Aging: Senescent Phenotypes

Mammalian gerontologists usually define aging in terms of the gradual, insidious, and progressive declines in structure and function (involving molecules, cells, tissues, organs, and organisms) that begin to unfold after the achievement of sexual maturity. These declines affect the germ line as well as the soma. For large populations of individuals, exponential declines in the probability of survival are observed. The aging organism is less successful in its reaction to injury and has increasing difficulties in maintaining physiological homeostasis. The organism, therefore, becomes increasingly vulnerable to a wide range of environmental perturbations.

Biological aging is the major risk factor for essentially all of the major geriatric disorders, including dementias of the Alzheimer type (DAT), Parkinson's disease, age-related macular degeneration, ocular cataracts, presbycusis, all forms of arteriosclerosis, type 2 diabetes mellitus, congestive heart failure, sarcopenia, osteoporosis, osteoarthritis, degenerative intervertebral disk disease, immunosenescence, benign prostatic hyperplasia, and most forms of cancer. These and many other disorders (**Table 71-1**) can be referred to as "senescent phenotypes."

Table 71-1 Alterations in Proliferative Homeostasis in the Tissues of Aging Humans

Integument	epidermal atrophy, "liver spots", seborrheic keratoses, basal cell ca, squamous cell carcinoma, graying and loss of hair, eccrine sweat gland atrophy, apocrine sweat gland hyperplasia, stasis dermatitis, regional subcutaneous atrophy, hyperplasia
Sensory	lacrimal gland atrophy, corneal degenerations, ocular cataracts, age-related macular degeneration, presbycusis, olfactory loss
Musculoskeletal	sarcopenia, "fatty infiltration" of muscle, osteoarthritis, osteoporosis
Hematopoietic/Immune	anemias, myelodysplastic syndromes, leukemia, lymphoma, monoclonal gammopathy and multiple myeloma, autoimmune disorders (e.g., atrophic gastritis and polycythemia vera), immunosenescence (accelerated in AIDS)
CNS	reactive gliosis, dural and meningeal fibrosis
Cardiovascular	atherosclerosis, arteriolosclerosis, myocardial interstitial fibrosis
Pulmonary	interstitial fibrosis, emphysema
Renal	glomerulosclerosis, interstitial fibrosis
Male Reproductive	benign prostatic hyperplasia (smooth muscle and glands), adenocarcinoma of prostate, testicular atrophy
Female Reproductive	ovarian atrophy and theca cell hyperplasia, endometrial atrophy and hyperplasia, endometrial carcinoma, smooth muscle atrophy, leiomyomas of uterus

Endocrine	parenchymal atrophy with interstitial fibrosis, cell type-specific hyperplasias, adenomas
GI	mucosal and smooth muscle atrophy, hyperplastic polyps, adenomas, adenocarcinomas of colon and rectum

Source: After GM Martin et al: PLoS Genet 3:e125, 2007.

The Classical Evolutionary Biological Theory of Why We Age

A compelling theory as to why aging occurs has been developed by a series of contributions by evolutionary biologists, beginning with JBS Haldane. Haldane wondered why certain late-onset disorders, such as Huntington's disease, seemed to be so prevalent in England—perhaps of the order of one per thousand instead of what might have been expected from germ-line mutation rates —perhaps one per million. He concluded that this was because the disease had largely escaped the force of natural selection, because the commonest forms of the disease did not manifest until after reproduction had ceased. In age-structured populations (i.e., populations consisting of individuals with a wide range of ages), most of the reproduction is carried out by the younger cohorts. This is because, historically, few individuals living in the wild escaped the effects of infections, predation, nutritional deprivations, and accidents to achieve old age. As such, even late-acting *good* alleles will have only minor contributions to the gene pools of subsequent generations. Peter Medawar extended this idea, arguing that there are numerous such constitutional mutations, an idea that has come to be known as the "mutation accumulation" theory of aging. (These are mutations that one is born with, not somatic mutations.) By this argument, most of us are likely to have been born with some special vulnerability to a late-life disorder or disorders. A second major theory, known as "antagonistic pleiotropy," was developed by George C. Williams. He argued that there are likely to be many genetic alleles that were selected because of their enhancement of reproductive fitness *early* in the life course, but that have negative effects *late* in life, when the force of natural selection will have greatly diminished. A more general conceptualization of tradeoffs between reproduction and lifespan (the "disposable soma" theory) was developed by TBL Kirkwood. The evolutionary theory was quantified by William D. Hamilton and elaborated on by Brian Charlesworth and Michael Rose.

Perhaps the best indication that the field of biogerontology has finally matured as a science is the fact that its most cherished theory, what can now be termed as the "classical" evolutionary theory, has undergone several challenges. First, demographers have noted that at the extremes of old age for organisms as diverse as roundworms, fruit flies, med flies, and humans, rates of *declines* in the force of natural selection diminish. One response to this important challenge (the "cocoon" hypothesis) is that these declines may simply be related to the virtual cessation of locomotor activities at extreme ages. When flies cease flying and worms cease moving, fewer opportunities for serious injuries may result. These "plateaus" in the rates of mortality at very advanced ages are much less striking for people; they might also be related, in part, to diminished motility, as well as to secular trends in the development of central heating, air conditioning, and immunizations. A second challenge comes from geneticists who have discovered that, to our great surprise, many single-gene mutations can substantially increase the lifespans of Baker's yeast, nematodes, and fruit flies This issue has been most systematically explored in *C. elegans*; a meta-analysis of an initial set of unbiased, genome-wide RNAi screens ("knockdowns," but not "knockouts" of gene expression) has tabulated hundreds of single-gene loci capable of enhancing lifespan when they are dialed down. These genes fall into a finite number of pathways and, moreover, many of them fit within the context of antagonistic pleiotropic mechanisms of aging. For example, hypomorphic mutations in genes within the most famous of these pathways, the insulin-like growth factor (IGF)/insulin signaling pathway, may be reporting on an evolutionarily conserved diapause—nature's way of taking time out from the business of development and reproduction during hard times, such as severe shortages of food. The gene actions associated with such diapauses (many of which are still unexplored at the biochemical genetic level) understandably result in enhanced resistance to various stresses. A large number of these *C. elegans* "longevity genes" converge upon a single transcription factor, *daf2*, or influence mitochondrial function. Moreover, it is possible that many such genes may be specific for this highly inbred laboratory model. In general, these investigations support, rather than refute, the evolutionary theory of aging.

A third challenge to the theory comes mainly from anthropologists and economists and emphasizes intergenerational transfers of resources. As an oversimplification, this idea is often referred to as "the grandmother hypothesis." Older members of population groups have survived a number of threats to their existence and can pass along to their younger family members information about successful avoidance or adaptation. Field experiments with prides of lions and olive baboons have failed to support this hypothesis. Moreover, while many grandparents are now contributing to the reproductive fitness of their children and grandchildren via transfer of resources, current evidence, while still incomplete, indicates that such elders were exceedingly rare within populations of our remote ancestors, when the evolution of species-specific gene actions would have evolved. Any such favorable alleles not expressed until those late stages of life would, therefore, have been vastly diluted by the alleles of their progeny.

A final challenge to evolutionary theory comes from a reexamination of the assumptions made by Hamilton in his influential 1966 paper. Baudisch, using different assumptions, demonstrates that, under some conditions, the force of natural selection can *increase* during aging. For example, species of rockfish that continue to grow beyond sexual maturity are much more likely to become predators rather than prey; as such, the force of natural selection would, indeed, become stronger, not weaker, as they age.

The message is that aging is *non-adaptive*. It did not evolve via a program of determinative gene actions that are designed to lead to the death of aging organisms because it is good for the species. Moreover, evolution has taught us that lifespans and their associated healthspans are plastic, thus providing a rationale for the potential effectiveness of future interventions.

Classes of Gene Action that Modulate Rates of Aging

If we are ever to intelligently intervene in one or more aging processes, we must be thoroughly familiar with the nature of the underlying gene actions. We must also be aware of the influences of idiosyncratic constitutional mutations, genetic polymorphisms, gene-gene interactions, environmental agents, gene-environmental interactions, and stochastic events. All of these complexities support what most experienced clinicians have learned in the course of their practices—namely that no two patients (even identical twins) age in precisely the same way; they share some commonalities but also have unique subsets of structural and functional impairments. Enhancements in the prevention and treatment of geriatric disorders will, therefore, require a more sophisticated and comprehensive understanding of the cellular, molecular, and integrative physiological underpinnings of these intra-specific variations in the patterns of aging.

The evolutionary biological theory of why we age provides clues as to how we age (i.e., the nature of the underlying gene actions). We can, in fact, outline twelve distinct classes of gene actions suggested by the evolutionary theory.

CLASS ONE: GOOD ALLELES WITH GOOD EFFECTS EARLY AND LATE: *LONGEVITY ASSURANCE GENES*

There are many examples of such genes. For instance, about 150 distinct human genetic loci have already been identified for the repair of DNA. **Table 71-2** gives an example of the many different types of genetic loci involved in the oxidative-damage theory of aging. According to that venerable theory, various sources of oxidative damage to macromolecules, notably those produced as byproducts of the oxidative metabolism of mitochondria, are the primary causes of intrinsic biological aging. There is now mounting evidence, however, that this theory is an insufficient explanation for variations in longevity. Nevertheless, increasing evidence also suggests that the modulation of oxidative damage is important for major aspects of healthspan. A cogent example is the evidence that the genetic engineering of mice to provide high concentrations of a human cDNA for catalase directed to the mitochondria greatly ameliorates a form of congestive heart failure responsible for many geriatric hospital admissions. By the criteria of anatomic pathology and echocardiography, the features in aging mice are quite comparable to those observed in aging human subjects. This finding gives support to the proposition that even potentially good novel gene actions can also escape the force of natural selection. Nature has targeted catalase to peroxisomes, but it does not supplement mitochondrial protection by also directing some optimal amount of catalase to the mitochondria to decrease the steady state levels of a dangerous by-product of mitochondrial metabolism, peroxide (H_2O_2). In the presence of iron, H_2O_2 results in the synthesis of the highly reactive hydroxyl radical, with resulting damage to all classes of macromolecules in its immediate vicinity.

Table 71-2 Categories of Gene Action Relevant to the Oxidative Damage Theory of Aging

Category I	Structural and regulatory genes modulating genesis of free radicals *Examples:* Cytochrome C Oxidase; P450 family
Category II	Structural and regulatory genes for scavenger enzymes *Examples:* SOD-1,2,3; Catalase; γ Glutamyl cysteine synthetase
Category III	Genes regulating flux of nonenzymatic free radical scavengers *Examples:* uric acid synthetic enzymes
Category IV	Genes regulating target copy number *Examples:* Genetic regulation of mitochondrial DNA replication, fusion, fission
Category V	Genes specifying target structure *Examples:* Structural genes for chromatin proteins and membrane lipoproteins
Category VI	Structural and regulatory genes for repair of target macromolecules

VI	*Examples:* Specification of machinery for reversal, repair, tolerance of DNA damage
Category VII	Genes specifying the orderly replacement of effete cells *Examples:* Genes modulating DNA replication and cell cycle progression, apoptosis, growth factors, growth factor receptors, and stem cell biology

Source: After GM Martin et al: Nature Genetics 13:25, 1996.

The last category (VII) of gene actions listed in Table 71-2 is relevant to a very important aspect of mammalian aging—the maintenance of proliferative homeostasis. As mammals age, there is a puzzling juxtaposition of both atrophy and inappropriate hyperplasia, often side by side. This can involve multiple tissues and is associated with numerous geriatric pathologies (Table 71-1). The hyperplasias may act as tumor promoters, leading to benign and malignantneoplasms.

Tissue atrophy may be attributable, in part, to the accumulating effects of cell apoptosis and necrosis from various causes and the failure of stem cells to compensate for the gradual attenuation of the replicative potentials of somatic cells, a process known as replicative senescence or the "Hayflick limit." It remains controversial as to whether or not an in vivo phenomenon exists that reflects the in vitro phenomenon of replicative senescence. The dominant (but not the only) mechanism responsible for replicative senescence is the loss of telomere repeat units from the ends of chromosomes. The germ line, many stem cells, and most cancers are protected from the erosion of telomeres by an enzyme known as telomerase, but this enzyme is absent from most somatic cells. Scientists shared a 2009 Nobel Prize for research in this field, the first such prize to recognize modern research in biogerontology. Cells that have exited the cell cycle typically do not undergo necrosis or apoptosis. They have a variety of phenotypes that characterize them as senescent; most notably a senescence-associated secretory phenotype (SASP) that has important consequences for regional pathology. In brief, such cells secrete a range of pro-inflammatory cytokines, metallothioneins, and mitogens. While such cells may be few in number within a given tissue, they can have a "field effect," thus driving the proliferation of neighboring epithelial cells, altering the connective tissue matrix, and contributing to sustained chronic inflammation. The latter is receiving increasing attention as a major concomitant of intrinsic biological aging.

As noted above, Category VII is also highly relevant to the emerging field of regenerative medicine. A possible generalization emerging from research on the interface of stem cell biology with the biology of aging is that a dominant factor in the decline of the success in mobilizing stem cells for the response to injury resides in the microenvironment of stem cells. This has been most effectively demonstrated in the repair of injury to skeletal muscle, in which satellite stem cells are mobilized quite effectively for repair in young mice but not in old mice. A clever parabiotic experiment provided compelling evidence for a non-cellular factor circulating from young to old mice that was capable of markedly ameliorating the deficient repair in the old mice. The satellite stem cell model has also demonstrated a switch, during adult myogenesis, from the Notch pathway—required for stem cell proliferation—to the Wnt pathway—required for effective differentiation. As is the case with so many signal transduction pathways, there is cross-talk between these two pathways. The situation is more complex, however, with evidence of an important role for a member of the transforming growth factor beta family of cytokines. There may also be a role for Klotho, a transmembrane protein with a structure suggestive of certain glycosidases. In any case, the clinical implications of this field of research are clear: as a much less complex and much safer alternative to the transplantation of stem cells, including those derived from the patient's own cells, clinicians might one day be able to inject small molecular weight compounds to "wake up" the patient's endogenous stem cells.

CLASS TWO: BAD ALLELES WITH LATE-LIFE PENETRANCE: *IDIOSYNCRATIC CONSTITUTIONAL MUTATIONS*

There is a national debate concerning health care reform, including the need for legislation for the protection of children with pre-existing genetic disorders. It seems likely, however, that we *all* have pre-existing conditions—namely gene actions that will lead to variable times of onset and variable severities of late-life disorders. Among these gene actions are individually rare but numerically numerous mutations that do not reach some phenotypic level of expression until middle age or beyond, when the effects will have escaped the force of natural selection. A prototypic example is Huntington's disease, one of a number of triplet repeat diseases; those unfortunate enough to have been born with the requisite number of CAG repeats (coding for a run of polyglutamines) in the affected locus will develop the disease after the peak of reproduction. More cogent examples, perhaps, come from rare, but pathogenetically informative, autosomal dominant mutations at three distinct loci resulting in "early" onset DAT. ("Early" here means younger than age 60; most patients are in late middle age when they're diagnosed with these forms of Alzheimer's disease.) To get some estimate of the genetic load of mutations that were known to lead to late-onset dementias of various types, the following is a systematic analysis of several editions of *McKusick's Mendelian Inheritance in Man*. In the 1975 edition, some 55 loci were identified from among 2336 listed at that time. Thirty years ago, the conventional wisdom was that there were ~100,000 protein-coding genes in the human genome. When the results of full genome sequencing were about to appear, geneticists participated in a contest

(GeneSweep) to see who got closest to the sequencing results; the low bidder (~26,000 genes) won the event. Assuming that number is, in fact, correct (some new estimates suggest somewhat lower or higher figures), we can conclude that about 2.4% of these protein-coding genes, or a total of 624, have the potential to modulate one's susceptibility to dementing disorders of late life. That has to be an absolutely lower limit of how we can get in trouble with cognition as we age; however, as there are thousands of functionally distinct splice variants and thousands of DNA sequences that code for various families of RNA molecules (the "dark matter" of DNA), most of which seem likely to have a role in the regulation of gene expression. We are also learning about so-called "moonlighting proteins" —single proteins with two distinct functions.

CLASS THREE: BAD ALLELES EARLY WITH GOOD EFFECTS LATE: *PARADOXICAL ANTAGONISTIC PLEIOTROPY*

In theory, such genes can persist in a population either because of a comparatively recent founder effect, or because the allele participates in a balanced polymorphism. Elevated frequencies of the 4G allele and of the homozygous 4G4G genotype were found in centenarians and were associated with high levels of plasminogen activator inhibitor-1. Such high levels are predictive of recurring myocardial infarction in young men. It is, therefore, paradoxical that these high levels are associated with extreme longevity. Like all such studies with centenarians, however, one requires independent confirmations and more sophisticated controls, such as the use of centenarian progeny and their spouses.

CLASS FOUR: BAD ALLELES EARLY AND LATE: *SEGMENTAL PROGEROID SYNDROMES*

These conditions can be defined as genetic disorders that mimic, to various extents, many, but not all, senescent phenotypes found in the general population. Prototypic examples are the Werner syndrome (WS) ("Progeria of the Adult") and the Hutchinson-Gilford syndrome (HGPS) ("Progeria of Childhood"). WS results from homozygosity for null mutations at the *WRN* locus, which codes for a member of the RecQ family of helicases. To "do business" with DNA, one must first unwind the double helix. The WRN protein appears to have several functions in DNA replication, transcription, recombination, and repair. Telomeres are particularly favored substrates. Almost all HGPS patients suffer from the same C-terminal mutation in the Lamin A gene (*LMNA*), which codes for an intermediate filament that lines the nuclear membrane. The mutation results in the preferential use of a cryptic splice site, resulting in the deletion of a sequence of 50 amino acids. The abnormal gene product (progerin) no longer acts as a substrate for the enzymatic removal of a post-translational modification (farnesylation). This contributes to structurally distorted nuclei and abnormalities in gene expression and has, therefore, led to clinical trials employing farnesylation inhibitors. Other factors are likely to contribute to the pathology, however. Many segmental progeroid syndromes are characterized by genomic instability. For the case of WS, there is a 10–100 fold increase in mutation rates; large deletions are particularly common. There is a striking limitation of the replicative lifespans of somatic cells from such patients, probably related to unrepaired damage of a major product of oxidative damage, 8-oxo-2'-deoxyguanosine, at telomeres. Deficiency of the WRN helicase also results in pro-inflammatory gene actions characteristic of normative aging.

For the case of HGPS, in addition to the genomic instability related to the abnormal structures of nuclei, there are likely to be important aberrations in the regulation of gene expression. Evidence that the study of HGPS can provide insight into normative aging comes from the discovery that small, potentially pathogenetically relevant amounts of progerin can be found in cells from normal individuals and that the effects of this abnormal protein may be particularly relevant for stem cells.

CLASS FIVE: GOOD ALLELES EARLY WITH BAD EFFECTS LATE: *ANTAGONISTIC PLEIOTROPY*

This category of gene action may contribute to three of our most devastating geriatric disorders —cancer, atherosclerosis, and DAT. The evidence that it plays a role in the pathogenesis of cancer comes from research referred to above, largely from the laboratory of Judith Campisi. The underlying hypothesis, as noted above, is that the repression of telomerase in somatic cells and alternative modalities of expediting exit from the mitotic cell cycle (such as DNA damage and oncogenic stimuli), evolved because they were adaptive for young, actively reproducing organisms, where they act as tumor suppressors. Later in life, however, the accumulation of replicative senescent cells may act in tumor promotion, both via mitogenic effects upon neighboring epithelial cells and via degradative actions upon the associated matrix, which may enhance local invasion by neoplastic cells. The arguments for roles in atherosclerosis and in Alzheimer's disease are much more speculative but worthy of more research. For the case of atherosclerosis, an argument can be made that macrophages play a primary pathogenetic role, particularly in the presence of functionally impaired, aged endothelial cells. The primary functions of macrophages are to engulf and destroy pathogens, for which purpose they employ a range of receptors including promiscuous "flypaper" receptors that can also recognize oxidized lipoproteins, which are of relevance to the pathogenesis of atherosclerosis. Given the exposure of modern humans to high-fat diets, late-life deleterious effects of the phagocytosis of oxidized lipoproteins may be an unfortunate tradeoff. Evidence for such a tradeoff has come from experiments with a mouse model of atherosclerosis (APOE knockout mice on a high-fat Western diet). When these mice were crossed with mice deficient in a class-A scavenger receptor

(*MSR-A*), there was the expected amelioration of atherosclerosis. These hybrid mice, however, were shown to be highly susceptible to infection with *Listeria monocytogenes* or herpes simplex virus.

The evidence for a role of antagonistic pleiotropic gene action in the pathogenesis of DAT remains modest but is of considerable interest. Most such discussions revolve around polymorphic alleles of the *APOE* gene. The ancestral allele found within primates is the notorious epsilon 4 allele, the single-most important genetic susceptibility factor for common "sporadic" forms of DAT. In populations within the developed world, this has become a minor allele, however. Arguments have been developed that this allele had been (and still is, in some parts of the world) a major allele of *Homo sapiens* because it is under selection for its putative protection against a variety of infectious agents, either because of its enhancement of the immune/inflammatory response or because of its less efficient delivery of lipids to the membranes of infectious agents such as *Trypanosoma brucei*, which must obtain its lipids from its infected host. Yet another antagonistic pleiotropic hypothesis for the existence of DAT in our species comes from studies of polymorphic forms of a locus coding for an adaptor protein of importance in the metabolism of the beta amyloid precursor protein (APP), widely considered to be of central importance in the pathogenesis of DAT. The gene coding for this protein is formally designated as the *APBB1* (Amyloid Beta A4 Precursor Protein-Binding, Family B, Member 1), but is widely referred to as *FE65.* Two laboratories have provided evidence that polymorphic alleles of this locus modulate susceptibility to what has been termed "very-late-onset Alzheimer's disease," that is to say, dementias with onsets after the peak ages of onset of the sporadic late-onset forms of Alzheimer's disease that are associated with the epsilon 4 allele of *APOE* (an association that peaks between 65 and 75 years of age). The *FE65* polymorphism was shown to be independent of the impact of the *APOE* polymorphism. Its minor allele, the dominant allele found in other mammals and primates (species resistant to DAT) was shown to bind with much less avidity to APP. This is likely to significantly alter the modulation of functions of that protein, potentially including the role of the APP/FE65/TIP60 complex in transcription. The authors suggested that the new allele emerged as part of a suite of gene actions to enhance cognitive functions, but that this came with deleterious effects of APP metabolism in late life.

CLASS SIX: BAD ALLELES EARLY AND LATE: *NUCLEAR AND MITOCHONDRIAL SOMATIC MUTATIONS*

The accumulation of somatic mutations in the tissues of aging mammals has been well-documented and can be surprisingly high. For example, *non-leaky* mutations (i.e., severe loss of function mutations) have been shown to rise exponentially in the renal tubular epithelium of human kidneys, reaching levels—by about age 80—between 10^{-3} and 10^{-4}. If one assumes that, for each such severe mutation, there are of the order of ten "leaky" mutations (i.e., mutations with diminished function), one can conclude that the levels of mutations at the single locus that was investigated could approach one in a hundred cells at advanced ages. In aging mice, the rates of increase and the types of nuclear somatic mutations were found to vary substantially from tissue to tissue. The frequencies of chromosomal mutations from the kidneys of aged F1 hybrid mice were found to be as high as one in three cells, but these remarkable results were likely related, in part, to transient exposures of the cells to ambient oxygen, which is now known to be particularly cytotoxic for murine cells. The late Howard J. Curtis, in an early example of the use of comparative gerontology, demonstrated that the levels of carbon tetrachloride-induced chromosomal mutations in mammals were inversely related to their lifespan potentials.

About 1500 genes (encoded by nuclei and mitochondrial DNA) contribute to the functioning of mitochondria. These targets clearly also contribute to the load of somatic mutations during aging and age-related diseases. The frequency of somatic mutation in mitochondrial DNA is ~500–1000-fold higher than it is in nuclear DNA. The tissues of individuals may, in fact, provide unique and dynamic mosaics of patterns of these mutations, serving as a sort of forensic fingerprint. While we have noted above that there is a waning of support for the idea that lifespan is limited by oxidative damage related to mitochondrial metabolism, arguments are still garnered for a primary role in aging processes. We will require more research on the characterization of the specific types of mitochondrial mutations that might be important actors in normative aging, because certain mutations may be more likely to out-compete wild type molecules within cells. Perhaps mutations with rearrangements of the mitochondrial genome produce more than one origin of replication and may lead to dominance ("homoplasty") of deleterious mutant molecules and thus the death of cells bearing that type of mutation.

In contrast to the present debate on the role of mitochondrial mutations and mitochondrial dysfunctions in the genesis of normative cellular aging, support for the roles of mitochondrial dysfunction in common geriatric diseases is becoming stronger. It is noted above that the evidence for an important role of mitochondrial dysfunction in a common form of geriatric congestive heart failure. The evidence of important roles of dysfunctional mitochondria in the genesis of Parkinson's disease is now quite compelling, given the roles of mutations at the *DJ-1* and *PINK1* loci. There is also great interest in observations of marked increases in the prevalence of mutations in the mitochondrial control region in the brains of patients with DAT. Given the key role of mitochondria in the control of apoptosis, aberrations in this pathway can skew the balance of cell proliferation and cell death and, therefore,

modulate carcinogenesis; there is, in fact, growing interest in targeting mitochondria as treatments for cancer. Mitochondrial deletions appear to be key events in the genesis of sarcopenia. Mitochondrial mutations or dysfunctions may also be involved in the pathogenesis of atherosclerosis. Finally, a novel hypothesis has been proposed implicating changes that occur in utero as influences on diseases of aging. Epidemiologic data show an association of small birth weight with the increased risks of developing type 2 diabetes, metabolic syndrome, and cardiovascular disease in adulthood. Leduc and Levy propose that placental mitochondrial dysfunction is present in cases of placental insufficiency and may be a critical influence on the fetus leading to atherosclerosis in later life.

Epigenetic Shifts in Gene Expression

The next six classes of gene action (half of the total) can all be grouped under this heading. The term "epigenetic" refers to covalent chemical alterations in the expressions of DNA that are "on top of" the DNA—that is to say, in contrast to mutations or polymorphisms, they do not change the primary sequence or arrangement of nucleotides. These alterations produce the striking specificities of gene expression that define the numerous functionally distinct mammalian cell types and are essential components in the reactions to various injuries. These chemical changes are of two broad types. One type involves methylations of cytosines, typically at "islands" of CpG runs within domains, such as promoters and enhancers, that regulate gene expression. Such methylations are associated with gene silencing. A second general type involves alterations of specific amino acids within the histone proteins that coat the DNA, such as acetylations (associated with gene activation), deacetylations (associated with gene silencing), phosphorylations, methylations, ubiquitinylations, and ADP ribosylations. Methods for molecular epigenetic analysis are being applied at the single cell level, a key technical development for gerontology, because there are many shifts in the population heterogeneity of tissues during aging. Such studies are of interest in the assessment of yet another antagonistic pleiotropic mechanism of aging (Class Twelve).

CLASS SEVEN: GOOD ALLELES DOWNREGULATED EARLY FOR GOOD REASONS: *ADAPTIVE SILENCING*

Given our evolutionary biological premise that nature does not really "care" much about the impact of gene actions that are going on late in the life course (i.e., when the force of natural selection is very weak), any switches in the degrees of gene expression that are initiated at some earlier stage of life because of their adaptive nature could, in principle, have a life of their own and could, therefore, continue to be downregulated or upregulated to a degree that could eventually become deleterious. The following are two examples of such downregulations, one observed in lab mice and the other in people. The first example occurs at the period of life when the somatic growth of mice is dramatically slowed and when resources are switched toward the business of reproduction. Beginning at around that time, a subset of genes that code for the synthesis of ribosomal proteins is silenced. Consequently, the rates of protein synthesis and protein turnover decline in most tissues. This leads to the accumulation of post-translationally modified proteins, including alterations that lead to diminished functions of those proteins, a process likely to contribute to senescent phenotypes. The second example involves the silencing, probably beginning with human puberty, of the estrogen receptor in human colonic mucosa, including regions of the colon that are particularly susceptible to the development of adenocarcinoma. This silencing continues throughout the adult lifespan and, given other lines of evidence for a role of this locus in the regulation of gene expression, likely plays a role in the development of colon cancer the elderly. Both examples involve gene silencing via methylations of CpG dinucleotides.

CLASS EIGHT: GOOD ALLELES UPREGULATED EARLY FOR GOOD REASONS: *ADAPTIVE EXPRESSION*

Here is the counterpoint for the mechanism mentioned above. An example is the enhanced expression of androgenic loci after sexual maturation, a likely contributor to the eventual emergence of benign prostatic hyperplasia later in life.

CLASS NINE: GOOD ALLELES INAPPROPRIATELY UPREGULATED IN LATE LIFE: *NON-ADAPTIVE LOSS OF SILENCING*

The best example of this mechanism comes from research on aging laboratory mice. To cite only one of several examples, a locus (*Atp7a*) on the X-inactivated chromosome (a normal dosage compensation event) increases expression in the spleen with increasing age. While the links to specific pathophysiological effects of this and other aberrant upregulations of genes remains to be elucidated, it seems likely that that there will be contributions to senescent phenotypes, given large number of such alterations.

CLASS TEN: GOOD ALLELES INAPPROPRIATELY DOWNREGULATED IN LATE LIFE: *NON-ADAPTIVE EPIGENETIC LOSS OF EXPRESSION*

An example comes from studies of monozygotic human twins. Both gains and losses of expression were seen in a large number of loci. While it is possible that a number of these alterations were adaptive responses to age-related changes in physiology ("sageing"), given the very large number of these epigenetic alterations in gene expression, many are likely to have resulted in pathophysiological

effects.

CLASS ELEVEN: GOOD ALLELES FOR FEMALES, BAD EFFECTS FOR MALES (AND VICE-VERSA): *SEX-BASED ANTAGONISTIC PLEIOTROPY*

It would seem likely that evolution will have optimized gene actions somewhat differently in males versus females to enhance the fitness for these behaviorally, morphologically, and physiologically distinct organisms. Thus, alleles that evolve because of optimization for females may not be optimal for males, and vice versa—a sort of evolutionary battle of the sexes. In fruit flies, a large proportion of the alleles that have been optimized for male fertilization are detrimental for the fecundity of females, and vice versa. Arguments have also been made that, given the exclusive inheritance of mitochondria via the female germ line, mitochondrial structure, and function, including numerous interactions with the nuclear genome, may not work as well in males as they do in females. These effects can translate into differential longevities and patterns of late life dysfunction.

CLASS TWELVE: GOOD OR BAD ALLELES EARLY WITH BAD EFFECTS LATE: *EPIGENETIC GAMBLING AND EPIGENETIC DRIFT*

The histograms of **Fig. 71-2** demonstrate a phenomenon that has puzzled gerontologists for generations. Why is it that, despite every effort to control genetics and environment, there are still marked variations in the lifespans of cohorts of experimental animals ranging from worms to mice? By far the best job of controlling for both genetics and environment has been achieved in experiments with *C. elegans* (Fig. 71-2). These organisms are hermaphrodites and, therefore, every diploid locus is driven to homozygosity—in other words, populations of these worms essentially consist of identical twins. They also can be grown in axenic media (devoid of bacteria) in temperature-controlled suspension cultures, including those with magnetic stirring rods, so that each individual worm "sees" the same food and waste and has equal opportunities for contact with other worms. Nevertheless, lifespans among such genetically defined organisms show considerable variability. This intrinsic variation is sufficient to produce a remarkable degree of overlap in the distributions of lifespans among wild type controls and a long-lived mutant population. These variations in lifespan are not heritable. Therefore, when considering the relative contributions of nature, nurture, and chance to intraspecific variations in lifespan, chance seems to be the dominating factor. Phenotypic evidence for stochastic variations is seen in the differences in the rates of development of aberrations in the ultrastructure of the skeletal muscles of aging worms. Some of these stochastic factors could involve somatic mutations—possibly in mitochondria—but their probable frequencies seem insufficient to explain these striking variations. Therefore, a theory suggests that, based upon the proposition that stochastic variations in gene expression develop in cohorts of all organisms, they enhance survival of the population. In a given environment, some worms may have developed a "lucky" set of gene expressions, while the patterns of gene expression may be quite nonadaptive in some of their identical twins. These individual outcomes could be changed in different environments. It is speculated that such "epigenetic gambling" may have evolved before the development of meiosis as a mechanism to ensure the survival of the species under unpredictable environments. Once initiated, "epigenetic drift" would ensue, eventually leading to departures from physiological homeostasis and senescent phenotypes. Published evidence indicates that epigenetic drift occurs within families of isogenic single cells.

Figure 71-2

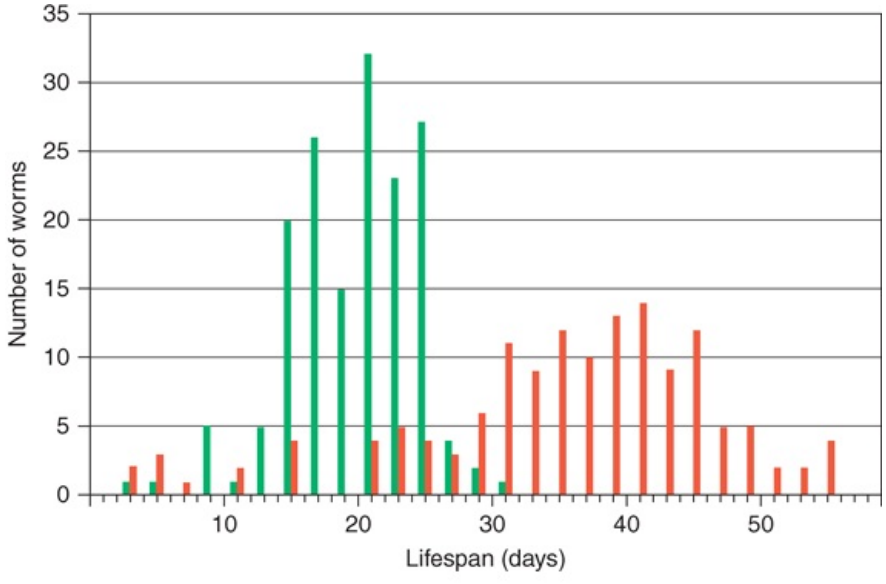

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Life-span distributions for individual *Caenorhabditis elegans* nematodes in isogenic populations of wild-type (green bars) and age-1 (red bars) strains. (*Reproduced from TB Kirkwood et al: Mech Ageing Dev 126:439, 2005; with permission.*)

The Identification of Signal Transduction Pathways Capable of Modulating Lifespan and Their Potential as Guides to Drug Targets

Earlier in the chapter evidence was cited for the existence of biochemical genetic pathways (or "signal transduction pathways") that, when appropriately modified, could extend the lifespans of several distantly related species—in other words, there is now evidence for the existence of "public" or shared mechanisms of aging. Because of claims that drugs that act in such pathways may enhance lifespans in model organisms (and, by extrapolation, perhaps in people), we need to review these pathways and the evidence for pharmacological interventions in more detail. We must keep in mind, however, that there is likely to be extensive cross-talk among these various pathways. Therefore, the "tweaking" of one pathway may have unanticipated effects on distant types of gene actions, especially given the high degree of genetic polymorphism in our species.

The best-documented such longevity-related signaling pathway is the IGF-insulin signaling pathway. The details of this pathway vary somewhat between nematodes, fruit flies, mice, and men, but a key common element is a downstream transcription factor (*daf16* in worms and members of the *FOXO* family in mammals). In response to the regulation of its phosphorylation state, the transcription factors can enter the nucleus and expedite the transcription of a very large suite of genes with diverse activities that serve to enhance the protection of the soma from macromolecular damage. Downregulation of IGF-insulin signaling, as in the case of upstream hypomorphic mutations, releases inhibition of this transcription factor. Such decreased signaling is associated with enhanced longevity. It has also been shown to protect *C. elegans* from the types of proteotoxicity associated with a variety of human neurodegenerative disorders. It is apparent, however, that an optimum level of functioning of this pathway must have evolved to enhance reproductive fitness and survival in both good times and bad times, including the prevention of diabetes mellitus. That such optimization is subject to modulations by genetic polymorphisms in human subjects is suggested by initial findings that exceptionally healthy and exceptionally long-lived subjects (centenarians) exhibit an enrichment of variant hypomorphic alleles for the IGF1 receptor. In that study, it was interesting that the progeny of these centenarians carrying these alleles were more likely to have shorter stature. Such short stature has been associated with a specific haplotype for IGF1 and with longer lifespans of breeds of dogs.

A second pathway (actually two related pathways) involves mTOR signaling. The nomenclature derives from the discovery that this protein, a serine/threonine kinase, is a target of rapamycin, a drug in clinical use as an immunosuppressive agent, for the prevention of coronary artery restenosis and for the treatment of malignantneoplasms. The pathways are involved in a plethora of vital functions, including cell proliferation and survival, the recycling of intracellular materials, aspects of metabolism including nutrient sensing, the generation of ribosomes, and the regulation of protein translation. It has been referred to as the conductor of the cell-signaling symphony. Rapamycin enhances the lifespans and healthspans of mice even when administered at ages roughly comparable to those of 60-year-old humans. While there appeared to be no impact of this treatment upon the numbers of neoplasms, the dominant cause of death of these strains of lab mice, it remains to be seen if the increased lifespan was attributable, at least in part, to a slowing of the rate of growth of these neoplasms.

A third pathway of growing interest to biogerontologists involves a family of histone deacetylases known as sirtuins. Early experiments in yeast demonstrated that such enzymes are associated with the silencing of genes. Their functions are intimately related to the metabolism of nicotinamide adenine dinucleotide (NAD), a coenzyme used as an oxidizing or reducing agent in a variety of essential metabolic processes. Although still controversial, an important member of this family is SIRT1, homologues of which enhance the lifespans of model organisms. SIRT1 is a putative target of resveratrol, which is thought to activate the enzyme and, therefore, might enhance lifespan and, presumably healthspan as well. Health benefits of large doses of resveratrol have been noted in obese, diabetic mice with fatty livers. Resveratrol, a component of red wine, has caught the imagination of the general public and the investment dollars of some pharmaceutical companies. The latter are searching for more potent variants of the molecule; phase IIa clinical trials in patients with type 2 diabetes and other disorders are in progress.

FURTHER READINGS

Austad SN: *Why we age: what science Is Discovering About the Body's Journey Through Life* . New York, John Wiley & Sons, 1997

Colman RJ et al: Caloric restriction delays disease onset and mortality in rhesus monkeys. Science 325:201, 2009[PMID: 19590001] [Full Text]

Conboy IM et al: Rejuvenation of aged progenitor cells by exposure to a young systemic environment. Nature 433:760, 2005[PMID: 15716955] [Full Text]

Coppe JP et al: The senescence-associated secretory phenotype: the dark side of tumor suppression. Annu Rev Pathol 5:99, 2010[PMID: 20078217] [Full Text]

Harrison DE et al: Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. Nature 460:392, 2009[PMID: 19587680] [Full Text]

Imai SI, Guarente L: Ten years of NAD-dependent SIR2 family deacetylases: implications for metabolic diseases. Trends Pharmacol Sci 31:212, 2010[PMID: 20226541] [Full Text]

Kennedy BK: The genetics of ageing: insight from genome-wide approaches in invertebrate model organisms. J Intern Med 263:142, 2008[PMID: 18226092] [Full Text]

Kenyon CJ: The genetics of ageing. Nature 464:504, 2010[PMID: 20336132] [Full Text]

Leduc L et al: Fetal programming of atherosclerosis: Possible role of the mitochondria. Eur J Obstet Gynecol Reprod Biol 149:127, 2010[PMID: 20053495] [Full Text]

Martin GM: *Modalities of Gene Action Predicted by the Classical Evolutionary Theory of Aging* , in VL Bengston et al (Eds). *Handbook of Theories of Aging* , 2nd ed, New York, Springer, 2009, pp 179-191

Partridge L, Gems D: Beyond the evolutionary theory of ageing, from functional genomics to evo-gero. Trends Ecol Evol 21:334, 2006[PMID: 16769434] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CLINICAL PROBLEMS OF AGING: INTRODUCTION

While an in-depth understanding of internal medicine serves as a foundation, proper care of older adults should be complemented by insight into the multidimensional effects of aging on disease manifestations, consequences, and response to treatment. In younger adults, individual diseases tend to have a more distinct pathophysiology with well-defined risk factors; the same diseases in older persons may have a less distinct pathophysiology and are often the result of failed homeostatic mechanisms. Causes and clinical manifestations are less specific and can vary widely between individuals. Therefore, the care of older patients demands an understanding of the effects of aging on human physiology and a broader perspective that incorporates geriatric syndromes, disability, social contexts, and goals of care. For example, care planning for the older patient cannot ignore the influence of life expectancy. In fact, the expected remaining years of life can guide recommendations about appropriate preventive and other long-term interventions, and shape discussions about treatment alternatives.

Demography (Chap. 70) Population aging emerged on a worldwide scale for the first time in history within the last century. Since aging influences many facets of life, governments and societies now face new social and economic challenges that impact health care, as well as family and community responsibilities. **Figure 72-1** highlights recent and predicted changes in U.S. population structure. The overall number of children has remained relatively stable, but explosive growth has occurred among older populations. The percentage growth is particularly dramatic among the oldest old. For example, the 80–89-year-old group increased more than threefold between 1960 and 2010 and will increase almost tenfold between 1960 and 2050. Women already outlive men by many years and the sex discrepancy in longevity is projected to increase further in the future.

Figure 72-1

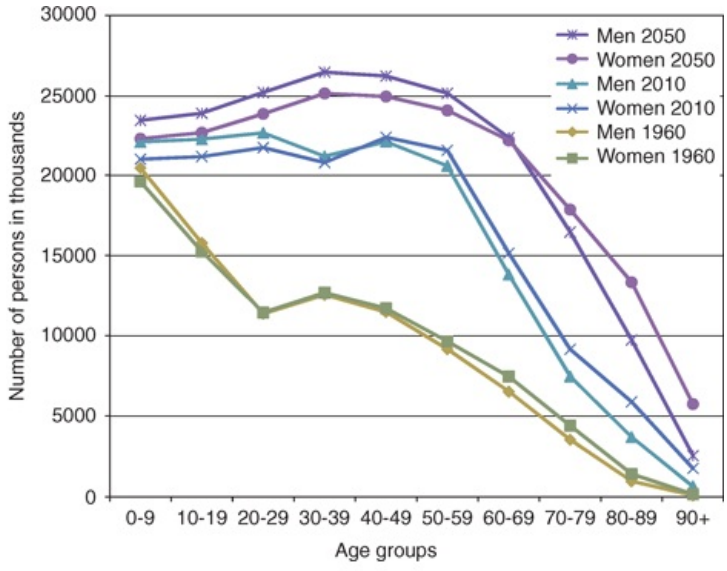

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Change in the structure of the U.S. population between 1960 and 2050. *(From United Nations World Population Prospects: The 2008 Revision, http://esa.un.org/unpp.)*

Population aging occurs at different rates in varying geographic regions of the world. Over the last century, Europe, Australia, and North America have had the populations with the greatest proportions of older persons, but Asia and South America are aging rapidly, with a population structure that will resemble the "older" countries by around 2050 **(Fig. 72-2)**. Among older persons, the oldest old (those older than age 80 years) are the fastest growing segment of the population **(Fig. 72-3)**, and the pace of aging is projected to accelerate in most countries in the next 50 years. There is no evidence that the

rate of population aging is decreasing.

Figure 72-2

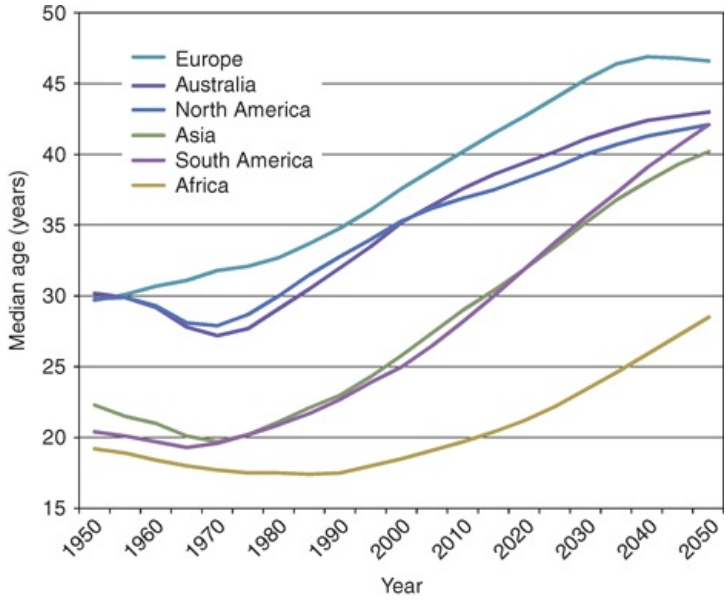

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Population aging in different geographic regions. *(From United Nations World Population Prospects: The 2008 Revision, http://esa.un.org/unpp.)*

Figure 72-3

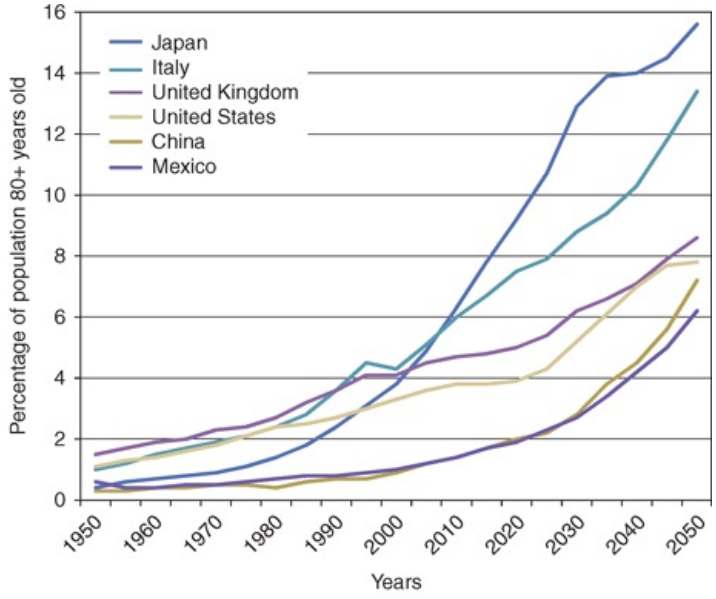

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Percentage of the population age >80 years from 1950 to 2050 in different representative nations. The pace of aging will accelerate. *(From United Nations World Population Prospects: The 2008 Revision, http://esa.un.org/unpp.)*

POPULATION AGING AND HEALTH

Many chronic diseases increase in prevalence with age. It is not unusual for older persons to have multiple chronic diseases **(Fig. 72-4)** although some seem more susceptible to co-occurring problems compared to others. Functional problems with difficulty or need for help in performing basic activities of daily living (ADLs) **(Table 72-1)** increase with age and are more common in women than men. In recent decades, the age-specific prevalence of disability has declined, especially in the oldest old. Estimated

rates are shown in **Fig. 72-5** as the percentage of persons who reported severe difficulty or needed help in bathing, but data on other basic activities of daily living show similar trends. The rate of decline in disability is decreasing, but the magnitude of this decline is small compared to the overwhelming effect of population aging. Thus, the number of people with disability in the United States and other countries is rapidly expanding. Rates of cognitive impairments, such as memory problems, also increase with aging **(Fig. 72-6)**. Chronic disease and disability lead to increased use of health care resources. Health care expenditures increase with age, increase more with disability, and are highest in the last year of life. However, new medical technologies and expensive medications are greater influences on health care costs than population aging alone. General practitioners and internists with little specific training in geriatric medicine provide the bulk of care for older persons.

Figure 72-4

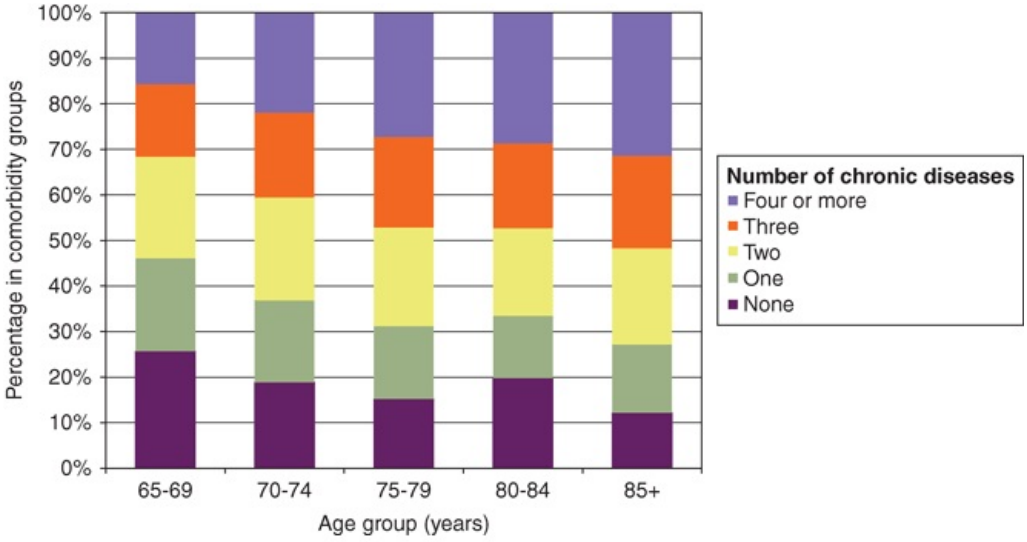

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prevalence of comorbidity by age group in persons 65 years and older living in the U.S. and enrolled in Medicare parts A and B in 1999. *(From JL Wolff et al: Arch Intern Med 162:2269, 2002.)*

Figure 72-5

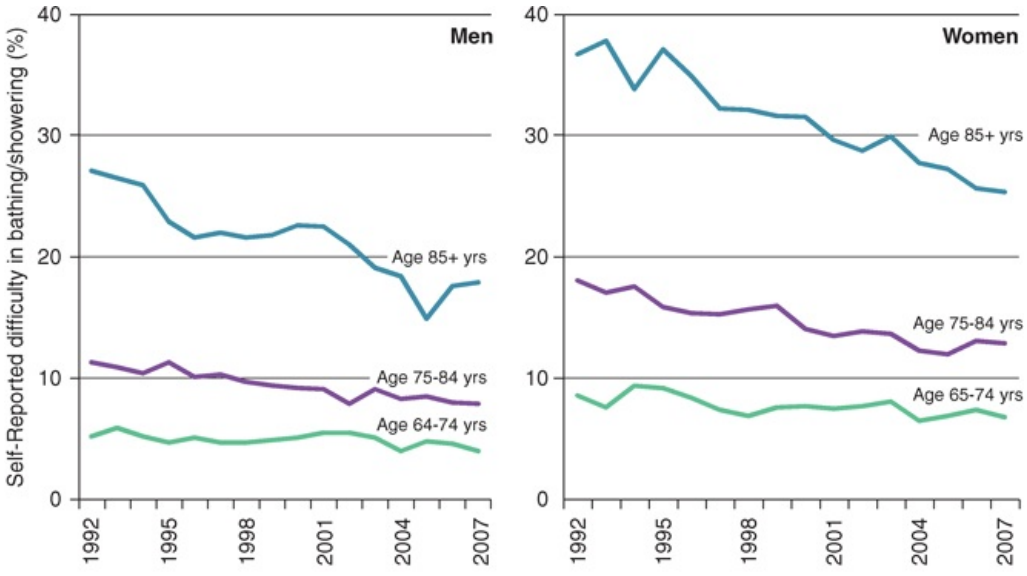

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Self-reported prevalence of disability (severe difficulty) in bathing/showering between 1992 and 2007, according to age and sex. *(From Medicare Current Beneficiary Survey 1992–2007. Accessed May 26, 2010 at http://205.207.175.93/HDI/TableViewer/.)*

Figure 72-6

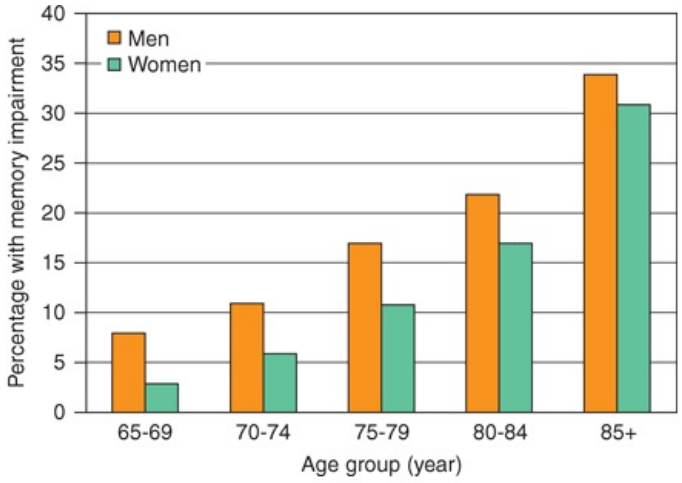

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rates of memory impairments in different age groups. The definition of "moderate or severe memory impairment" is 4 or fewer words recalled out of 20. *(Source: Health and Retirement Survey. Accessed February 7, 2011 at http://aoa.gov/agingstatsdotnet/Main_Site/Data/2000_Documents/healthstatus.aspx.)*

Table 72-1 Basic and Instrumental Activities of Daily Living

The Basic Activities of daily living (ADLs) consist of these self-care tasks

- Personal hygiene
- Dressing and undressing
- Eating
- Transferring from bed to chair, and back
- Voluntarily controlling urinary and fecal discharge
- Using the toilet
- Moving around (as opposed to being bedridden)

Instrumental activities of daily living (IADLs) are not necessary for fundamental functioning, but they let an individual live independently in a community

- Doing light housework
- Preparing meals
- Taking medications
- Shopping for groceries or clothes
- Using the telephone
- Managing money
- Using technology (older generations may not be that technologically savvy since they were not as exposed to it during their lifetime.)

SYSTEMIC EFFECTS OF AGING

Systemic consequences of aging are widespread but can be clustered into four main domains or processes **(Fig. 72-7)**: (1) body composition; (2) balance between energy availability and energy demand; (3) signaling networks that maintain homeostasis; and (4) neurodegeneration. Each domain can be assessed using routine clinical tests, although more detailed research techniques are also available **(Table 72-2)**.

Figure 72-7

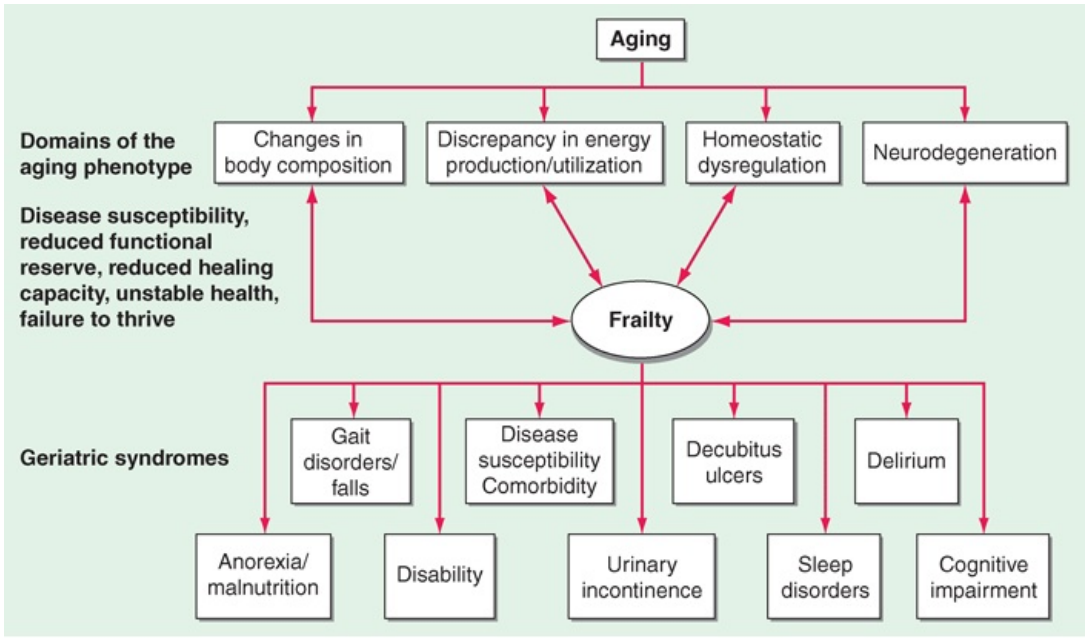

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A unifying model of aging, frailty, and the geriatric syndromes.

Table 72-2 Example of Assessment of the Different Domains of the Aging Phenotype

Body Composition	Energetics	Homeostatic Regulation	Neurodegeneration
Anthropometrics (weight, height, BMI, waist circumference, arm and leg circumference, skin folds)	Self-reported questionnaires investigating physical activity, sense of fatigue/exhaustion, exercise tolerance	Baseline levels of biomarkers and hormone levels	Standard neurologic exam, including assessment of global cognition [Mini Mental State, Montreal Cognitive Assessment (MoCa)]
Muscle strength testing (isometric and isokinetic)	Resting metabolic rate	Inflammatory markers (ESR, CRP, IL-6, TNF-alpha, etc)	Objective assessment of gait, balance, reaction time, coordination
Biomarkers (24-h creatinuria or 3-methyl-histidine)	Performance-based tests of physical function	Nutritional biomarkers (vitamins, antioxidants, etc)	Electroneurography and electromyography
CT and MRI, dual-energy x-ray absorptiometry (DEXA)	Treadmill testing	Response to provocative tests, such as oral glucose tolerance test, dexamethasone test, and others	MRI, fMRI, PET, and other dynamic imaging techniques
Hydrostatic weighing	Objective measures of physical activity (accelerometers, double labeled water)	Stress response	Evoked potentials

Body Composition

Profound changes in body composition may be the most evident and inescapable effect of aging **(Fig. 72-8)**. Over the life span, body weight tends to increase through childhood, puberty, and adulthood until late middle age. Weight tends to decline in men between ages 65 and 70 years, and somewhat later in women. Lean body mass, composed predominantly of muscle and visceral organs, decreases steadily after the third decade. In muscle, this atrophy is greater in fast-twitch compared to slow-twitch fibers. Fat mass tends to increase in middle age and then declines in late life, reflecting the trajectory of weight change. Interestingly, waist circumference continues to increase across the life span, suggesting that visceral fat, which is responsible for most of the pathologic consequences of obesity, continues to accumulate. In some individuals, fat also accumulates inside muscle where it affects muscle quality and function. With age, fibro-connective tissue tends to increase in many organ systems. In muscle, fibro-connective tissue buildup also affects muscle quality and function. In combination, the loss of muscle mass and quality result in reduced muscle strength, with ultimate

impact on functional capacity and mobility. Muscle strength declines with aging and not only affects functional status but is also a strong independent predictor of mortality **(Fig. 72-9)**. Progressive demineralization and architectural modification occurs in bone, resulting in a decline of bone strength. Loss of bone strength increases the risk of fracture. Sex differences in the effects of aging on bone mass are due to sex differences in peak bone mass and the effects of gonadal hormones on bone. Overall, compared to men, women tend to lose bone at an earlier age and more quickly reach the threshold of low bone strength that increases fracture risk. All of these changes in body composition can be attributed to disruptions in the links between synthesis, degradation, and repair that normally serve to remodel tissues. Body composition can be approximated in clinical practice using weight, height, body mass index (weight in kilograms divided by height in meters squared), and waist circumference, or more precisely using dual-energy x-ray absorptiometry (DEXA), computed tomography or MRI.

Figure 72-8

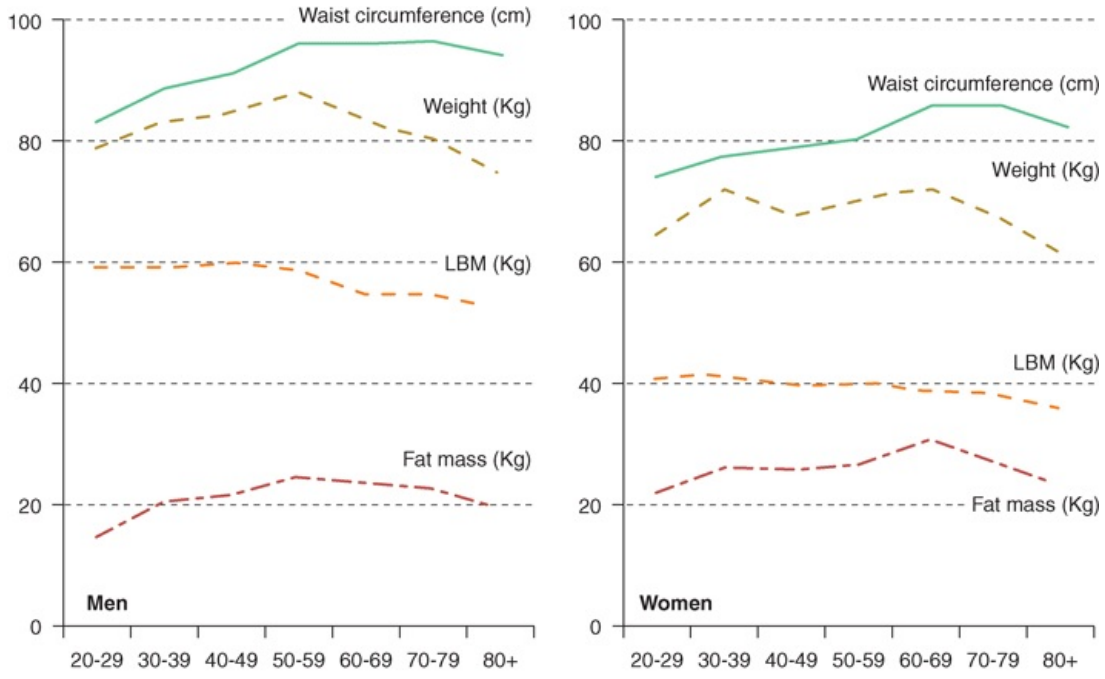

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Longitudinal changes of weight, body composition, and waist circumference over the life span estimated in 1167 participants of the Baltimore Longitudinal Study of Aging. *Lean Body Mass (LBM) and Fat Mass estimated with DEXA. *[Source: The Baltimore Longitudinal study of Aging 2010 (unpublished data).]*

Figure 72-9

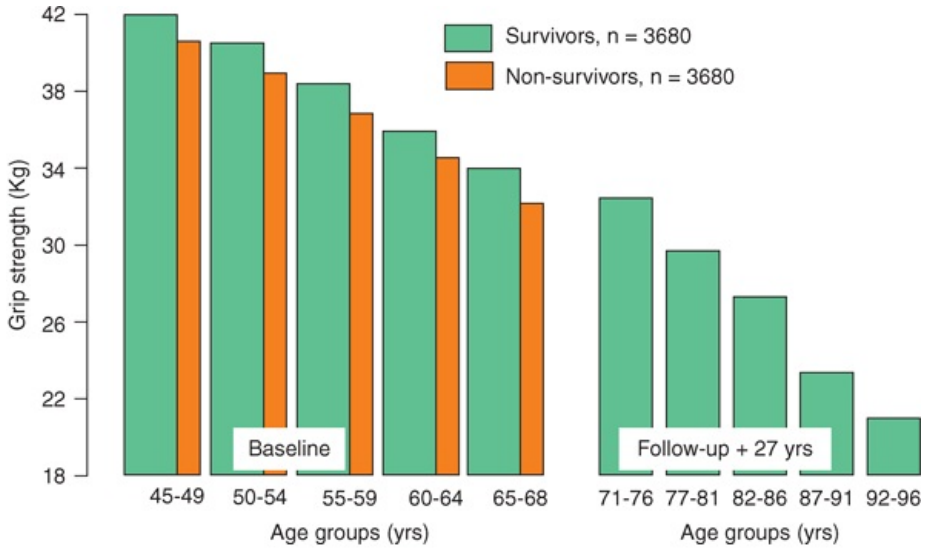

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cross-sectional differences and longitudinal changes in muscle strength over a 27-year follow-up. Note that subjects who died during the follow-up had lower baseline muscle strength. *(From T Rantanen et al: J Appl Physiol 85:2047, 1998.)*

Balance between Energy Availability and Energy Demand

Release of phosphate from ATP provides every living cell with the energy required for life. However, the storage of ATP is only enough for 6 seconds; therefore, ATP is constantly resynthesized. Although ATP resynthesis can be done during anaerobic glycolysis, most of the energy used in the body is generated through aerobic metabolism. Therefore, energy comsumption is usually estimated indirectly by oxygen consumption (indirect calorimetry). There is currently no method to measure true "fitness," which is the maximum energy that can be produced by an organism over extended time periods. Thus, fitness is estimated indirectly from peak oxygen consumption (MVo_{2peak}), often during a maximal treadmill test. Longitudinal studies have demonstrated that MVo_{2peak} declines progressively with aging **(Fig. 72-10)**, and the rate of decline is accelerated in those who are sedentary and in those affected by chronic diseases. A large portion of energy is consumed as the "resting metabolic rate" (RMR), the amount of energy expended at rest in a neutral temperature environment and in a postabsorptive state. In healthy individuals, RMR declines with aging, mostly due to a decline in the highly metabolically active tissues of lean body mass **(Fig. 72-11)**. However, persons with unstable homeostasis due to illness require additional energy for compensatory mechanisms. Indeed, an inadequately high RMR is a marker of illness, is an independent risk factor for mortality, and may contribute to the weight loss that often accompanies severe illness. Finally, for reasons that are not yet completely clear, older age, pathology, and physical impairment increase the energy cost of motor activities such as walking. Overall, older individuals with multiple chronic conditions have low available energy levels and require more energy both at rest and during physical activity. Thus, sick older people may consume all their available energy performing the most basic activities of daily living, leading to symptoms of fatigue and restriction to a sedentary existence. Energetic status can be assessed clinically by simply asking the patient about their perceived level of fatigue during daily activities such as walking or dressing. Energy capacity can be assessed more precisely by exercise tolerance during a walking test or a treadmill test coupled with spirometry.

Figure 72-10

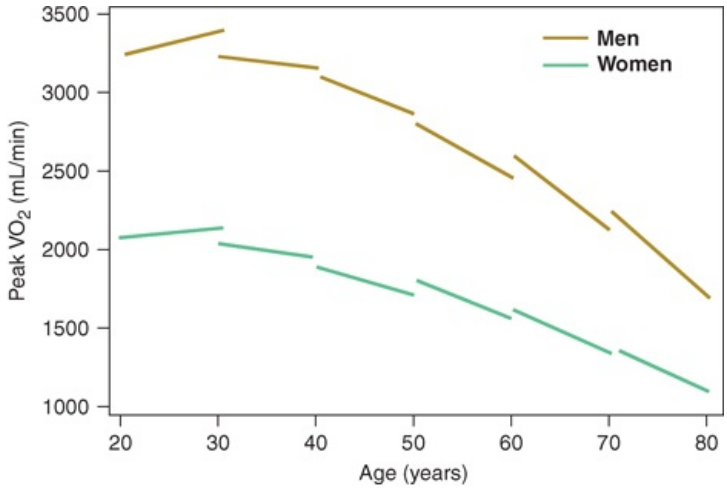

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Longitudinal changes in aerobic capacity in participants of the Baltimore Longitudinal Study of Aging. *(From JL Fleg: Circulation 112:674, 2005.)*

Figure 72-11

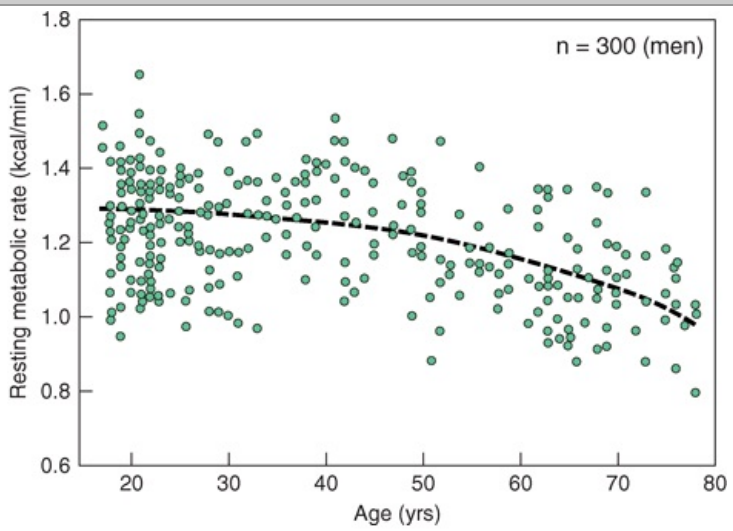

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Changes in resting metabolic rate with aging. *[Data from the Baltimore Longitudinal Study of Aging (unpublished).]*

Signaling Networks that Maintain Homeostasis

The main signaling pathways that control homeostasis involve hormones, inflammatory mediators, and antioxidants; all are profoundly affected by aging. Sex hormone levels, such as testosterone, decrease with age in both men **(Fig. 72-12)** and women, while other hormone systems may change more subtly **(Table 72-3)**. Most aging individuals, even those who remain healthy and fully functional, tend to develop a mild proinflammatory state characterized by high levels of proinflammatory markers, including IL-6 and CRP **(Fig. 72-13)**. Aging is also thought to be associated with increased oxidative stress damage, either because the production of reactive oxygen species increases or because antioxidant buffers are less effective. Since hormones, inflammatory markers, and antioxidants are integrated into complex signaling networks, levels of individual biomarkers may well reflect adaptation within homeostatic feedback loops rather than true causative factors. Thus, the therapeutic strategy of single-molecule replacement may be ineffective or even counterproductive. The presence of such signaling networks and feedback loops may help explain why single-hormone "replacement therapy" for problems of aging has demonstrated little benefit. The focus of research in this area is now on multiple-hormonal dysregulation. For example, taken one at a time, levels of testosterone, dehydroepiandrosterone (DHEAs), and IGF-1 do not predict mortality, but in combination they are highly predictive of longevity. This combination effect is especially strong in the setting of congestive

heart failure. Similarly, several micronutrients, such as vitamins (especially vitamin D), minerals (selenium and magnesium), and antioxidants (vitamins D and E), also regulate aspects of metabolism. Low levels of these micronutrients have been associated with accelerated aging and high risk of adverse outcomes. However, except for vitamin D, no clear evidence suggests that supplementation has positive effects on health. Unfortunately, no standard criteria exist that allow the detection and quantification of homeostatic dysregulation.

Table 72-3 Hormones that Decrease, Remain Stable, and Increase with Aging

Decrease	No Change	Increase
Growth hormone	Prolactin	Cholecystokinin
Luteinizing hormone (men)	Thyrotropin	Luteinizing hormone (women)
Insulin growth factor I	Thyroid hormones	Follicle-stimulating hormone
Testosterone	Epinephrine	Cortisol
Estradiol	Glucagon-like peptide 1	Prolactin
DHEA and DHEAs	Gastric inhibitory polypeptide	Norepinephrine
Pregnenolone		Insulin
25(OH) vitamin D		Parathormone
Aldosterone		
Vasoactive intestinal peptide Melatonin		

Figure 72-12

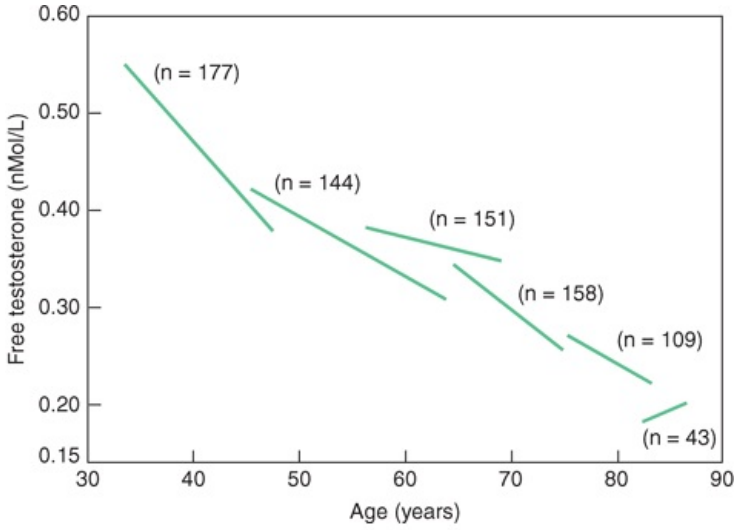

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Longitudinal changes of free testosterone levels in healthy men. *(From SM Harman et al: J Clin Endocrinol Metab 86:724, 2001.)*

Figure 72-13

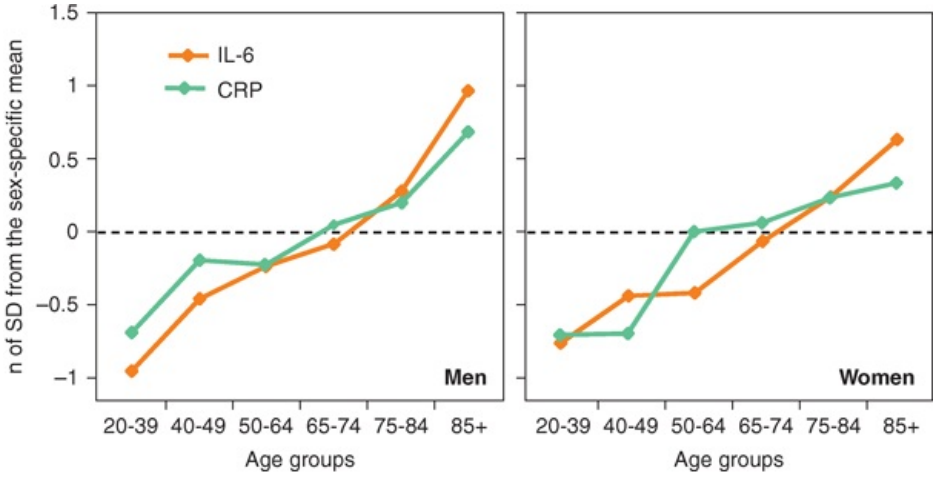

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Change in IL-6 and c-reactive protein with aging. *Values are expressed as Z-Scores to make them comparable. *(From L Ferrucci et al: Blood 105:2294, 2005.)*

Neurodegeneration

Neurons stop reproducing shortly after birth and their number declines throughout life. Brain atrophy occurs with aging after the age of 60 years. Atrophy proceeds at varying rates in different parts of the brain **(Fig. 72-14)** and is often accompanied by an inflammatory response and microglial activation. Age-associated brain atrophy may contribute to age-related declines in cognitive and motor function. Atrophy may also be a factor in some brain diseases that may occur with aging, such as mild cognitive impairment (MCI), in which persons have mild but detectable impairments on tests of cognition but no severe disability in daily activities. In MCI, atrophy has been found mostly in the prefrontal cortex and hippocampus, but these findings are not specific and their diagnostic utility is unclear **(Fig. 72-15)**. Other neurophysiologic changes in the brain also frequently occur with aging and may contribute to cognitive decline. Functional imaging studies have shown that some older people have diminished coordination between the brain regions responsible for higher-order cognitive functions and that such diminished coordination is correlated with poor cognitive performance. In young healthy individuals, the brain activity associated with executive cognitive functions (e.g., problem solving, decision making) is very well localized while in healthy older individuals, the pattern of cortical activation is more diffuse. Brain pathology has typically been associated with specific diseases; amyloid plaques and neurofibrillary tangles are considered the pathologic hallmarks of Alzheimer's disease. However, these pathologic markers have been found at autopsy in many older individuals who had cognitive testing in the year before death and were found to be normal. Taken together, trends in brain changes with aging suggest that some neurophysiologic manifestations are compensatory adaptations, rather than primary contributors to age-related declines. Because the brain is capable of reorganization and compensation, extensive neurodegeneration may not be clinically evident. Therefore, early detection requires careful testing. Clinically, cortical and subcortical changes are reflected in the high prevalence of "soft," nonspecific neurologic signs, often with slow and unstable gait, poor balance, and slow reaction times. These movement changes can be elicited more overtly using "dual tasks," in which a cognitive and a motor task are performed simultaneously. In a simple version of a dual task, an older adult who has to stop walking in order to talk has been shown to predict increased risk of falls. Poor dual task performance has been interpreted as a marker of reduced overall capacity for central processing, so that simultaneous processing is more constrained. Beyond the brain, the spinal cord also experiences changes after age 60 years, including reduced numbers of motor neurons and damage to myelin. The motor neurons that survive compensate by increased branching complexity, and by serving larger motor units. As motor units become larger, they decline in number at a rate of about 1% per year, starting after the third decade. These larger motor units contribute to reductions in fine-motor control and manual dexterity. Age-related changes also occur in the autonomic nervous system, affecting cardiovascular and splanchnic function.

Figure 72-14

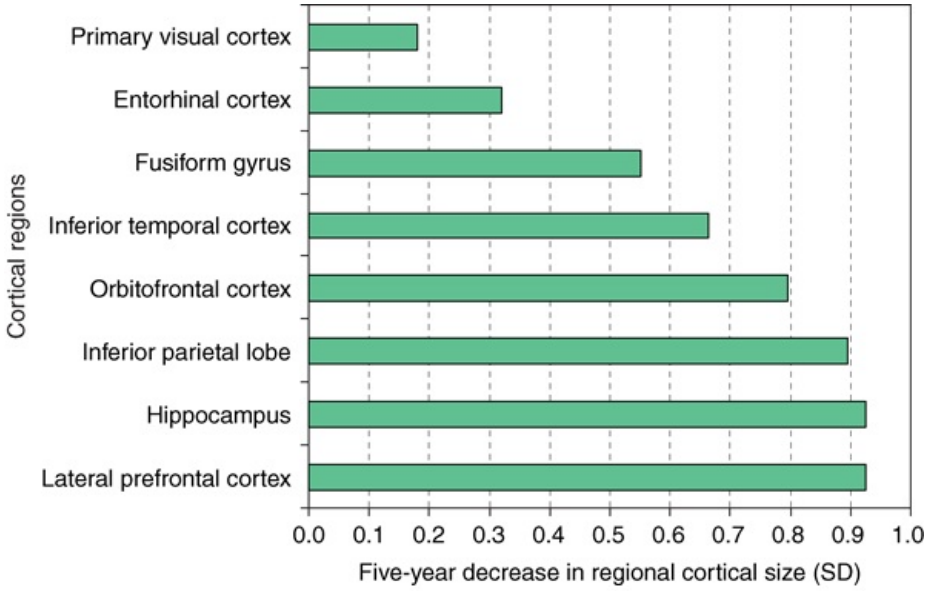

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Five-year decline in mean volumes measured in standard deviation units (Cohen's *d*) of different brain regions. The primary visual cortex shows the smallest average shrinkage and the prefrontal and inferior parietal cortex and hippocampus showed the highest average shrinkage. *(From N Raz et al: Ann N Y Acad Sci 1097:84, 2007.)*

Figure 72-15

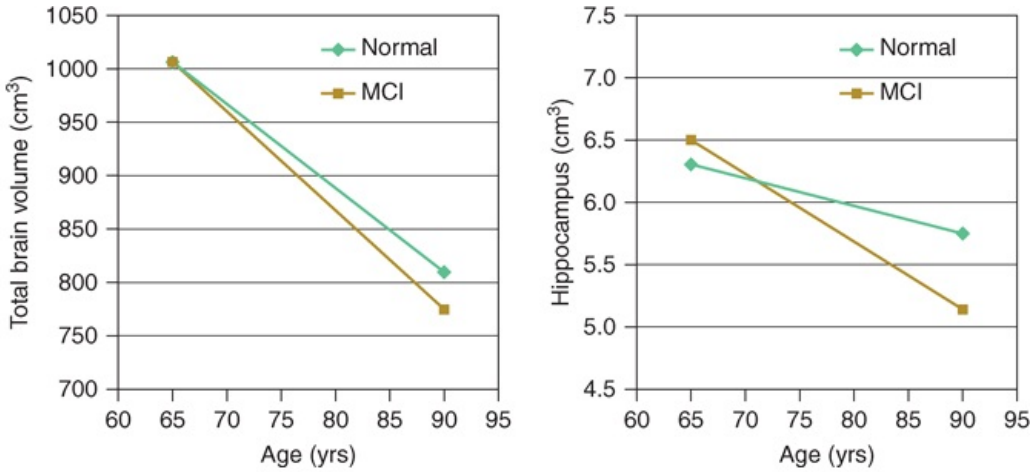

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Longitudinal changes of regional brain volumes in normal aging and MCI. *(From I Driscoll et al: Neurology 72:1906, 2009.)*

System Changes Coexist and Affect Each Other: The Phenotype of Aging Is the Final Common Pathway of This Interaction

While age-related system changes were described individually, in reality, these changes develop in parallel and affect each other through many feed-forward and feedback loops. Some system interactions are well understood, while others are under investigation. For example, body composition interacts with energy balance and signaling. Higher lean body mass increases energy consumption and improves insulin sensitivity and carbohydrate metabolism. Higher fat mass, especially visceral fat mass, is the culprit in the metabolic syndrome and is associated with low testosterone, high SHBG, and increased levels of proinflammatory markers such as C-reactive protein and IL-6. Altered signaling can affect neurodegeneration; insulin resistance and adipokines such as leptin and adiponectin are associated with declines in cognitive function. A state of inflammation, reduced levels of testosterone and IGF-1, combined with loss of motor neurons and dysfunction of the motor unit, have been linked to accelerated decline of muscle mass and strength. Normal intersystem coordination is also affected by aging. The hypothalamus normally functions as a central regulator of metabolism and energy use and

coordinates physiologic responses of the entire organism through hormonal signaling; aging-related changes in the hypothalamus alter this control. The central nervous system also controls adaptive sympathetic/parasympathetic activity, so that age-related CNS degeneration may have implications for autonomic function.

The phenotype that results from the aging process is characterized by increased susceptibility to diseases, high risk of multiple coexisting diseases, impaired response to stress (including limited ability to heal or recover after an acute disease), emergence of "geriatric syndromes" (characterized by stereotyped clinical manifestations but multifactorial causes), altered response to treatment, high risk of disability, and loss of personal autonomy with all its psychological and social consequences. In addition, these key aging processes may interfere with the typical pathophysiology of specific diseases, thereby altering expected clinical manifestations and confounding diagnosis. Clinically, patients may present with obvious problems within only one of these domains, but since systems interact, all four main domains should be evaluated and considered potential therapeutic targets. When patients present with obvious problems in multiple main systems affected by aging, they tend to extreme degrees of susceptibility and loss of resilience, a condition that is globally referred to as "frailty."

Frailty

Frailty has been described as a physiologic syndrome characterized by decreased reserve and diminished resistance to stressors, resulting from cumulative decline across multiple physiologic systems, causing vulnerability to adverse outcomes and high risk of death. A proposed definition characterized by weight loss, fatigue, impaired grip strength, diminished physical activity, and slow gait has shown good internal consistency and strong predictive validity, and has been used in many clinical and epidemiologic studies. However, alternative schools of thought have different diagnostic criteria. For example, frailty has been suggested to be a random accumulation of multiple impairments with aging, and, therefore, no standard criteria for diagnosis can be developed. Regardless of the definition, an extensive literature shows that older persons who are considered frail by any definition have overt changes in the same four main processes—body composition, homeostatic dysregulation, energetic failure, and neurodegeneration, the characteristics of the aging "phenotype." A classic clinical case would be an older woman with sarcopenic obesity characterized by increased body fat and decreased muscle (body composition changes); extremely low exercise tolerance and extreme fatigue (energetic failure); high insulin; low IGF-1; inadequate intake of calories; low vitamin D, E, and carotenoids (signal dysregulation); and memory problems, slow gait, and unstable balance (neurodegeneration). This woman is likely to show all the manifestations of frailty, including high risk of multiple diseases, disability, urinary incontinence, falls, delirium, depression, and other geriatric syndromes.

Conceptualizing frailty through the four main underlying processes stems from accumulated evidence and recognizes the heterogeneity and dynamic nature of the aging phenotype. Aging is universal but proceeds at highly variable rates, with wide heterogeneity in the emergence of the aging phenotype. Thus, the question is not whether an older patient is frail, but rather whether the severity of frailty is beyond the threshold of clinical and behavioral relevance. Understanding frailty through the lens of four interacting underlying processes also provides an interface with diseases that, like aging itself, affect the "aging phenotype." For example, congestive heart failure is associated with low energy availability, multiple hormonal derangements, and a proinflammatory state, thereby contributing to frailty severity. Parkinson's disease is an example of neurodegeneration that, in an advanced state, affects body composition, energy metabolism, and homeostatic signaling, resulting in a syndrome that closely resembles frailty. Diabetes is especially important to aging and frailty because it harms body composition, energy metabolism, homeostatic dysregulation, and neuronal integrity. Accordingly, a number of studies have found that type 2 diabetes is a strong risk factor for frailty and for many of its consequences. Since disease and aging interact, careful and appropriate treatment of disease is critical to prevent or reduce frailty.

Consequences of Aging Processes, the Aging Phenotype, and Frailty

While the pathophysiology of frailty is still being elucidated, its consequences have been well characterized in prospective studies. Four main consequences are important for clinical practice: (1) ineffective or incomplete homeostatic response to stress, (2) multiple coexisting diseases (multi- or comorbidity) and polypharmacy, (3) physical disability, and (4) the so-called geriatric syndromes. We will briefly address each one of them.

Low Resistance to Stress

Frailty can be considered a progressive loss of reserve in multiple physiologic functions. At an early stage and in the absence of stress, mildly frail older individuals may appear to be normal. However, they have reduced ability to cope with challenges, such as acute diseases, traumas, surgical procedures, or chemotherapy. Acute illness involving a hospital stay is associated with undernutrition and inactivity, which sometimes may be of such magnitude that the residual muscle mass fails to meet the minimal requirement for walking. Even when nutrition is reinstated, energy reserves may be insufficient to adequately rebuild muscle mass. Older persons have a reduced ability to tolerate

infections, in part because they are less able to build a dynamic inflammatory response to vaccination or infectious exposure, so that infections are more likely to become severe and systemic and resolve more slowly. In the context of tolerance to stress, assessing aspects of frailty can help estimate ability to withstand the rigors of aggressive treatments, to respond to interventions aimed at infection, to anticipate and prevent complications of hospitalization, and generally to estimate prognosis. Accordingly, treatment plans may be adjusted to improve tolerance and safety; bed rest and hospitalization should be used sparingly; and infections should be prevented, anticipated, and managed assertively.

Comorbidity and Polypharmacy

Older age is associated with high rates of many chronic diseases (Fig. 72-4). Thus, not unexpectedly, the percentage of individuals affected by multiple medical conditions (co- or multimorbidity) also increases with age. In frail older individuals, comorbidity occurs at higher rates than would be expected from the combined probability of the component conditions. It is likely that frailty and comorbidity affect each other, so that multiple diseases contribute to frailty and frailty increases susceptibility to diseases. Clinically, patients with multiple conditions present unique diagnostic and treatment challenges. Standard diagnostic criteria may not be informative because there are additional confusing signs and symptoms. A classic example is the coexistence of iron and vitamin B12 deficiency, creating an apparently normocytic anemia. The risk/benefit ratio for many medical and surgical treatment options may be reduced in the face of other diseases. Drug treatment planning is made more complex because comorbid diseases may affect the absorption, volume of distribution, protein binding, and, especially, elimination of many drugs, leading to fluctuation in therapeutic levels and increased risk of under- or overdosing. Drug excretion is affected by renal and liver changes with aging that may not be detectable with usual clinical tests. Formulas for estimating glomerular filtration rate in older patients are available while estimating changes in hepatic excretion is still a challenge. Patients with many diseases are usually prescribed multiple drugs, especially when they are cared for by multiple specialists who do not communicate. The risk of adverse drug reactions, drug-drug interactions, and poor compliance increases geometrically with the number of drugs prescribed and with the severity of frailty. Some general rules to minimize the chances of adverse drug events are as follows: (a) always ask patients to bring in all medications, including prescription, over-the-counter, vitamins supplements, and herbal preparations (the "brown bag test"); (b) screen for unnecessary drugs—those without a clear indication should be stopped; (c) simplify the regimen in terms of number of agents and schedules, try to avoid frequent changes, and use single daily dose regimens whenever possible; (d) avoid drugs that are expensive or not covered by insurance whenever possible; (e) minimize the number of drugs to those that are absolutely essential and always check for possible interactions; (f) make sure that the patient or an available caregiver understands the administration regimen and provide legible written instructions; and (g) schedule periodic medication reviews.

Disability and Impaired Recovery from Acute-Onset Disability

The prevalence of disability in self-care and home management increases steeply with aging and tends to be higher in women than in men (Fig. 72-5). Physical and cognitive function in older persons reflects overall health status, and predicts health care utilization, institutionalization, and mortality more accurately than any other known biomedical measure. Thus, assessing function and disability and predicting the risk of disability are cornerstones of geriatric medicine. Frailty, regardless of the criteria used for its definition, is a robust and powerful risk factor for disability. Because of this strong relationship, measures of physical function and mobility have been proposed as standard criteria for frailty. However, disability occurs late in the frailty process, after reserve and compensation are exhausted. Early in the development of frailty, body composition change, reduced fitness, homeostatic deregulation, and neurodegeneration can begin without impact on daily function. As opposed to disability in younger persons, where the rule is to look for a clear dominant cause, disability in frail, older persons is almost always multifactorial. Multiple disrupted aging processes are usually involved, even when the precipitating cause seems unique. Excess fat mass, poor muscle strength, reduced lean body mass, poor fitness, reduced energy efficiency, poor nutritional intake, low circulating levels of antioxidant micronutrients, high levels of proinflammatory markers, objective signs of neurologic dysfunction, and cognitive impairment all contribute to disability. The multifactorial nature of disability in frail older persons reduces the capacity for compensation and interferes with functional recovery. For example, a small lacunar stroke that causes problems with balance in a young hypertensive individual can be overcome by standing and walking with the feet further apart, a strategy that requires the brain to adapt, strong muscles, and lots of energy capacity. The same small, lacunar stroke may cause catastrophic disability in an older person already affected by neurodegeneration and weakness who is less able to compensate. As a consequence, interventions aimed at preventing and reducing disability in older persons should have a dual focus on both the precipitating cause and the systems needed for compensation. In the case of the lacunar stroke, interventions to promote mobility function might include stroke prevention, balance rehabilitation, and strength training. As a rule of thumb, the assessment of contributing causes and the design of intervention strategies for disability in older persons should always consider the four main aging processes that contribute to frailty. One of the most popular approaches to disability measurement is a modification of the International Classification

of Impairments, Disabilities and Handicaps (World Health Organization, 1980) proposed by the Institute of Medicine (IOM, 1992). This classification infers a causal pathway in four steps: pathology (diseases), impairment (the physical manifestation of diseases), functional limitation (global functions such as walking, grasping, climbing stairs), and disability (ability to fulfill social roles in the environment). In practice, the assessment of functional limitation and disability is performed either by: 1) self reported questionnaire concerning the degree of ability to perform basic self-care or more complex activities of daily living; 2) performance-based measures of physical function that assess specific domains, such as balance, gait, manual dexterity, coordination, flexibility, and endurance. A concise list of standard tools that can be used to assess physical function in older persons is reported in **Table 72-4**. In 2001, the WHO officially endorsed a new classification system, the International Classification of Functioning, Disability and Health, known more commonly as ICF. In the ICF, health measures are classified from body, individual, and societal perspectives by means of two lists: a list of body functions and structure and a list of domains of activity and participation. Since an individual's functioning and disability occurs in a context, the ICF also includes a list of environmental factors. A detailed list of codes that allow the classification of body functions, activities, and participation is being developed. The ICF system is widely implemented in Europe and is gaining popularity in the United States.

Table 72-4 Tools for Functional Assessment in Older Patients

Measurement Instrument	Evaluation	Activities/ Reference	Notes
Index of independence in Activities of Daily Living (ADLs)	Self-reported	Difficulty/need for help in bathing, dressing, using toilet, transferring, continence, feeding. *Katz S et al: The index of ADL: A standardized measure of biological and psychosocial function. JAMA 1963*	Short and simple but subjective.
Instrumental Activities of Daily Living (IADLs)	Self-reported	Difficulty using the telephone, using car/public transportation, shopping, preparing meals, housework, managing medications, financial management. *Lawton MP et al: Instrumental activities of daily living (IADL) scale: original observer-rated version. Psychopharmacol Bull 1988*	Short and simple. Gender-biased and cultural-biased items.
Functional Independence Measure (FIM)	Consensus multidisciplinary team	Motor (eating, grooming, bathing, dressing, toileting, bladder/bowel management, transfers, walking, climbing stairs); Cognitive (auditory comprehension, verbal expression, social interaction, problem solving, memory). *Keith RA et al: The functional independence measure: A new tool for rehabilitation. Adv Clin Rehabil 1987*	Administered by trained health professionals.
Barthel Index	Professionally evaluated	Independence and need for help in feeding, transferring from bed to chair and back, grooming, transferring to and from toilet, bathing, walking, stairs, dressing, continence. *Mahoney FI et al: Functional evaluation: The Barthel Index. Md State Med J 1965*	Administered by trained health professionals.
Mobility Questionnaire	Self-reported	Severe difficulty walking 1/4 mile and/or climbing stairs.	Short and simple
Short Physical Performance Battery	Objective performance based	Time to walk 4 m, rise from a chair 5 times, maintain balance for 10 sec in the side-by-side, semi-tandem and tandem positions. *Guralnik JM et al: A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994*	Some training required.
Berg Balance Scale	Objective and professionally evaluated	Performance in 14 different tasks related to balance. *Berg KO et al: Clinical and laboratory measures of postural balance in an elderly population. Arch Phys Med Rehabil 1992*	Typically used by physical therapists.
Walking Speed	Objective performance	Measure walking speed over a 4 m course. *Studenski S: Bradypedia: Is gait speed ready for clinical use? J Nutr Health Aging 2009*	Simple and powerful but limited to patients who can walk.
6-Minute Walk	Objective performance based	Distance covered in 6 min. *Guyatt GH: The 6-minute walk: A new measure of exercise capacity in patients with chronic heart failure. Can Med Assoc J 1985*	Good measure of fitness, walking capacity/

			endurance.
Long Distance Corridor Walk (400 m)	Objective performance based	Time to fast walk 400 m. *Newman AB et al: Association of long-distance corridor walk performance with mortality, cardiovascular disease, mobility limitation, and disability. JAMA 2006*	More challenging than the 6-min walk.

Impaired cognition is a very important cause of disability but is treated in detail in Chap. 371.

Geriatric Syndromes

The term *geriatric syndrome* is used to capture clinical conditions that are frequently encountered in older persons, have a deleterious effect on function and quality of life, have multifactorial pathophysiology, often involving systems unrelated to the apparent chief complaint, and are manifested by stereotypical clinical presentations. The list of geriatric syndromes includes incontinence, delirium, falls, pressure ulcers, sleep disorders, problems with eating or feeding, pain, and depressed mood. Dementia and physical disability are also sometimes considered to be geriatric syndromes. Using the term *syndrome* is somewhat misleading since this term is normally used to describe a pattern of symptoms and signs that have a single underlying cause. Geriatric syndromes, by contrast, refer to "multifactorial health conditions that occur when the accumulated effects of impairments in multiple systems render an older person vulnerable to situational challenges." According to this definition, geriatric syndromes reflect the complex interactions between an individual's vulnerabilities and exposure to stressors or challenges. This definition aligns well with the concept that geriatric syndromes should be considered as phenotypic consequences of frailty and that a limited number of shared risk factors contribute to their etiology. Indeed, in various combinations and frequencies, virtually all geriatric syndromes are characterized by body composition changes, energy gaps, signaling disequilibria, and neurodegeneration. For example, detrusor (bladder) underactivity is a multifactorial geriatric condition that contributes to urinary retention in the frail elderly. It is characterized by detrusor muscle loss, fibrosis, and axonal degeneration. A proinflammatory state, and a lack of estrogen signaling, cause bladder muscle loss and detrusor underactivity while a chronic urinary infection may cause detrusor hyperactivity, all factors that may contribute to urinary incontinence.

Due to limited space, only delirium, falls, incontinence, chronic pain, and anorexia are addressed here. Interested readers can consult the references at the end of the chapter for details on other geriatric syndromes.

Delirium

(Chap. 25) Delirium is an acute disorder of disturbed attention that fluctuates with time. It affects between 15 and 55% of hospitalized older patients and is associated with high in-hospital mortality and sometimes with permanent brain damage. **Figure 72-16** shows brief guidelines for assessment and management of delirium in hospitalized older patients. The clinical presentation of delirium is heterogeneous, but frequent features are (a) rapid decline in level of consciousness with difficulty focusing, shifting, or sustaining attention; (b) cognitive change (rumbling incoherent speech, memory gaps, disorientation, hallucinations) not explained by dementia; and (c) medical history suggestive of preexisting cognitive impairment, frailty, and comorbidity. The strongest predisposing factors for delirium are dementia, any other condition associated with chronic or transient neurologic dysfunction (neurologic diseases, dehydration, alcohol consumption, psychoactive drugs), and sensory (visual and hearing) deprivation, suggesting that delirium is a condition of brain function susceptibility (neurodegeneration or transient neuronal impairment) that cannot avoid decompensation when hit by a stressful event. Many stressful conditions have been implicated as precipitating factors, including surgery, anesthesia, persistent pain, opiates, narcotics, anticholinergics, sleep deprivation, immobilization, hypoxia, malnutrition, and metabolic and electrolyte derangements. Delirium, both onset and severity, can be reduced by anticipatory screening and preventive strategies aimed at reducing precipitating causes. The Confusion Assessment Method (CAM) is a simple, validated tool for screening in the hospital setting. Immediate identification and treatment of precipitating factors, withdrawal of drugs that may have facilitated the insurgence of delirium, and supportive care (management of hypoxia, hydration and nutrition, mobilization, and environmental modifications) are the three pillars of treatment. Whether patients in special Delirium Units have better outcomes is still in question. Physical restraints should be avoided because they tend to increase agitation and injury. Whenever possible, drug treatment should be avoided because it may prolong or aggravate delirium in some cases. The treatment of choice is low-dose haloperidol.

Falls and Balance Disorders

Unstable gait and falls are serious concerns in the older adult because they lead not only to injury but also to restricted activity, increased health care utilization, and even death. Like all geriatric syndromes, problems with balance and falls tend to be multifactorial and are strongly connected with the disrupted aging systems that contribute to frailty. Poor muscle strength, neural damage in the basal ganglia and cerebellum, diabetes, and peripheral neuropathy are all recognized risk factors for falls. Therefore, evaluation and management require a structured multisystem approach that spans the

entire frailty spectrum and beyond. Accordingly, interventions to prevent or reduce instability and falls usually require a mix of medical, rehabilitative, and environmental modification approaches. Guidelines for the evaluation and management of falls, released by the American Geriatrics Society, recommend asking all older adults about falls and perceived gait instability **(Fig. 72-17)**. Patients with a positive history of multiple falls, in addition to persons who have sustained one or more injurious falls, should undergo an evaluation of gait and balance as well as a targeted history and physical examination to detect sensory, nervous system, brain, cardiovascular, and musculoskeletal contributors. Interventions depend on the factors identified but often include medication adjustment, physical therapy, and home modifications. Meta-analyses of falls prevention strategies have found that multifactorial risk assessment and management, as well as individually targeted therapeutic exercise, are effective in reducing falls. Supplementation with vitamin D at 800 IU daily may help reduce falls, especially in older persons with reduced levels.

Figure 72-16

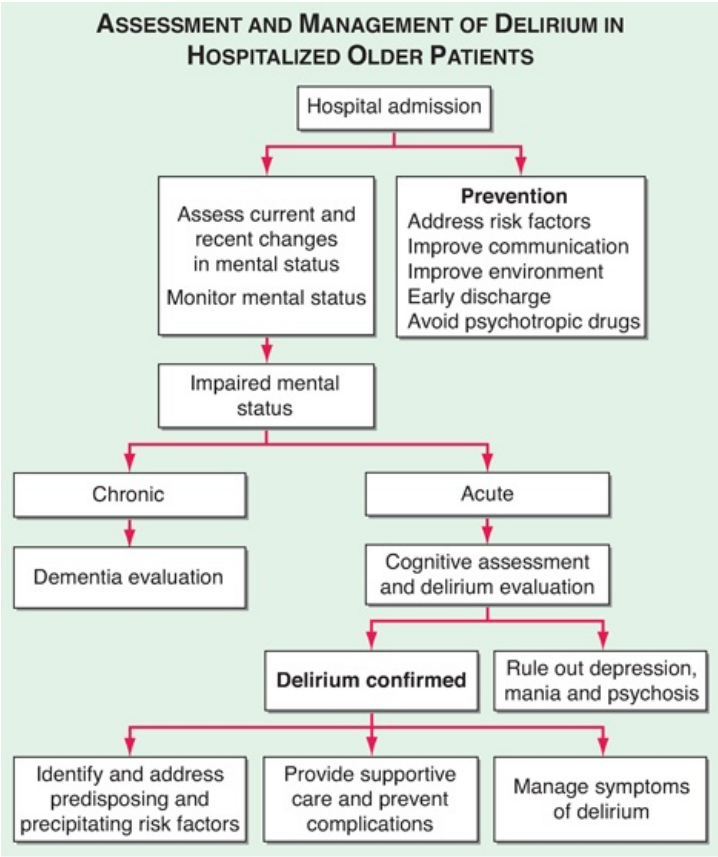

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm depicting assessment and management of delirium in hospitalized older patients. *(Modified from SK Inouye: N Engl J Med 354:1157, 2006.)*

Figure 72-17

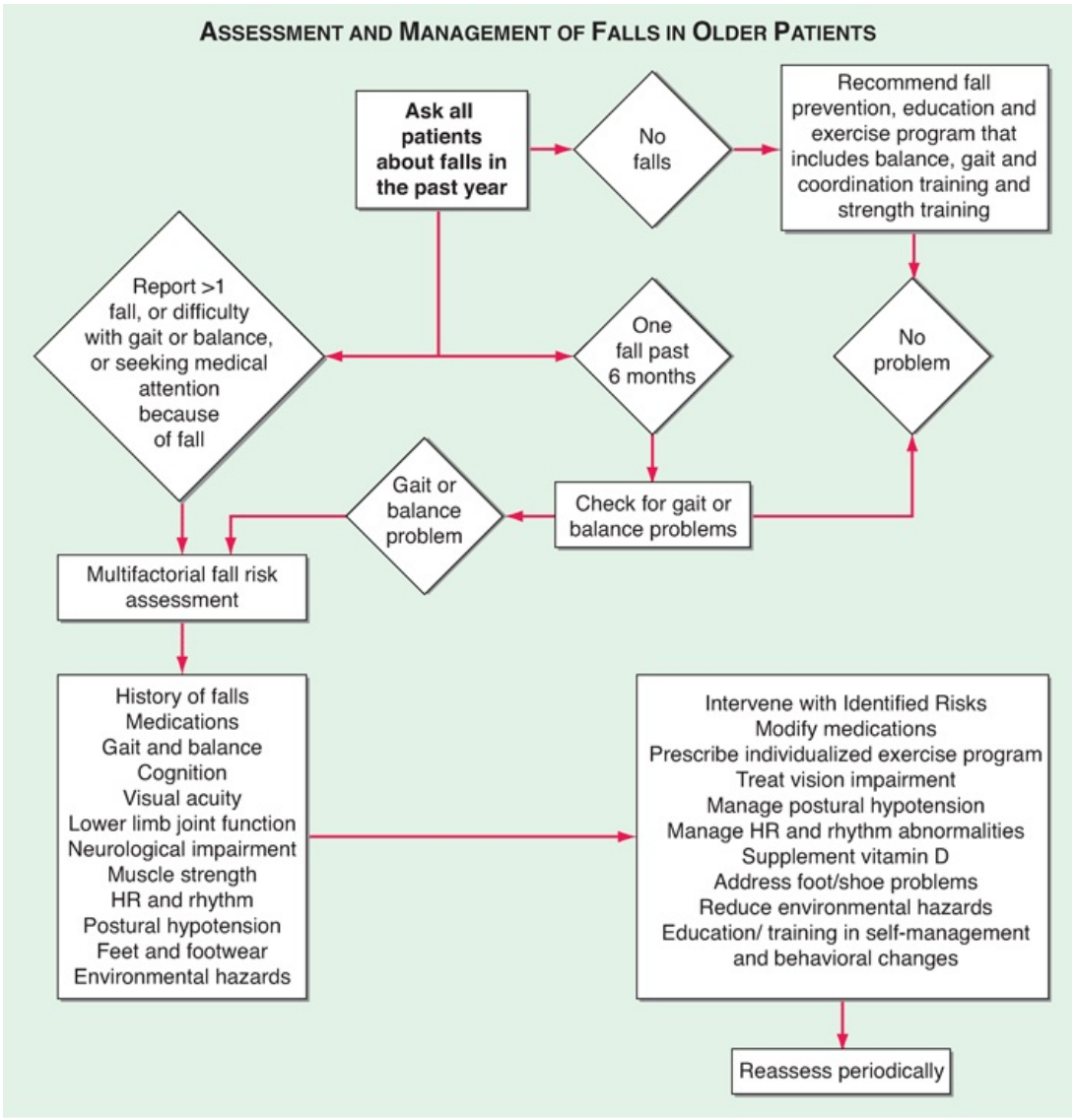

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm depicting assessment and management of falls in older patients. *(From American Geriatrics Society and British Geriatrics Society: Clinical Practice Guideline for the Prevention of Falls in Older Persons. New York, American Geriatric Society, 2010.)*

Persistent Pain

Pain from multiple sources is the most common symptom complaint of older adults in primary care settings and is also common in acute care, long-term care, and palliative care settings. Acute pain and cancer pain problems are beyond the scope of this chapter. Persistent pain results in restricted activity, depression, sleep disorders, social isolation, and increases the risk of medication adverse events. The most common causes of persistent pain are musculoskeletal problems, but neuropathic and ischemic pain occur frequently, and multiple concurrent causes are often found. Alterations in mechanical and structural elements of the skeleton commonly lead to secondary problems in other parts of the body, especially to soft tissue or myofascial components. A structured history should obtain information about the quality, severity, and temporal patterns of pain. Physical examination should focus on back and joints as well as trigger points and periarticular areas, as well as evidence of radicular neurologic patterns and peripheral vascular disease. Pharmacologic management should follow standard progressions as recommended by the World Health Organization (Chap. 11), but side effects to the central nervous system are more likely and must be monitored. For persistent pain, regular analgesic schedules are appropriate and should be combined with nonpharmacologic approaches such as splints, physical exercise, heat, and other modalities. A variety of adjuvant analgesics such as antidepressants and anticonvulsants may be used, but again, effects on reaction time and alertness may be dose limiting, especially in older persons with cognitive impairment. Joint or soft tissue injections may be helpful. Patient education and mutual goal setting are important since pain is usually not fully eliminated, but rather controlled to a tolerable level that maximizes function while minimizing adverse effects.

Urinary Incontinence

Urinary incontinence (UI), the involuntary leakage of urine, is highly prevalent in older persons, especially in older women, and has a profound negative impact on quality of life. Approximately 50% of American women will suffer from some form of UI over a lifetime. Increasing age, white race, childbirth, obesity, and medical comorbidity are all risk factors for UI. The three main clinical forms of UI are as follows: (1) *Stress* incontinence is the failure of the sphincteric mechanism to remain closed when there is a sudden increase in intraabdominal pressure, such as a cough or sneeze. In women, this condition is due to insufficient strength of the pelvic floor muscles, while in men it is almost exclusively secondary to prostate surgery. (2) *Urge* incontinence is the loss of urine accompanied by a sudden sensation of need to urinate and is due to detrusor muscle overactivity (lack of inhibition) due to loss of neurologic control or local irritation. (3) *Overflow* incontinence is characterized by urinary dribbling, either constantly or for some period after urination. This condition is due to impaired detrusor contractility (due usually to denervation, for example, in diabetes) or bladder outlet obstruction (prostate hypertrophy in men and cystocele in women). Thus, not surprisingly, the pathogenesis of urinary incontinence is connected to the disrupted aging systems that contribute to frailty, body composition changes (atrophy of the bladder and pelvic floor muscle), and neurodegeneration (both central and peripheral nervous systems). Frailty is a strong risk factor for urinary incontinence. Indeed, older women are more likely to have mixed (urge+stress) incontinence than any pure form **(Fig. 72-18)**. In analogy with the other geriatric syndromes, UI derives from a predisposing condition superimposed on a stressful, precipitating factor. Accordingly, treatment of UI should address both. The first line of treatment is bladder training associated with pelvic muscle exercise (Kegel exercises) that sometimes should be associated with electrical stimulation. Those with possible vaginal or uterine prolapse should be referred to a specialist. Urinary infections should be investigated and eventually treated. A long list of medications can precipitate urinary incontinence, including diuretics, antidepressants, sedative hypnotics, adrenergic agonists or blockers, anticholinergic, and calcium channel blockers. Whenever possible, these medications should be discontinued. Until recently, it was believed that estrogen oral or local treatments improved the UI symptoms in postmenopausal women, but this notion is now controversial. Antimuscarinic drugs such as tolterodine, darifenacin, and fesoterodine are modestly effective for mixed incontinence, but they all can affect cognition and so must be used with caution and careful follow-up monitoring of cognitive status. In some cases, surgical treatment should be considered. Chronic catheterization has many adverse effects and should be limited to chronic urinary retention that cannot be managed in any other way. Bacteriuria always occurs and should be treated only if symptomatic.

Figure 72-18

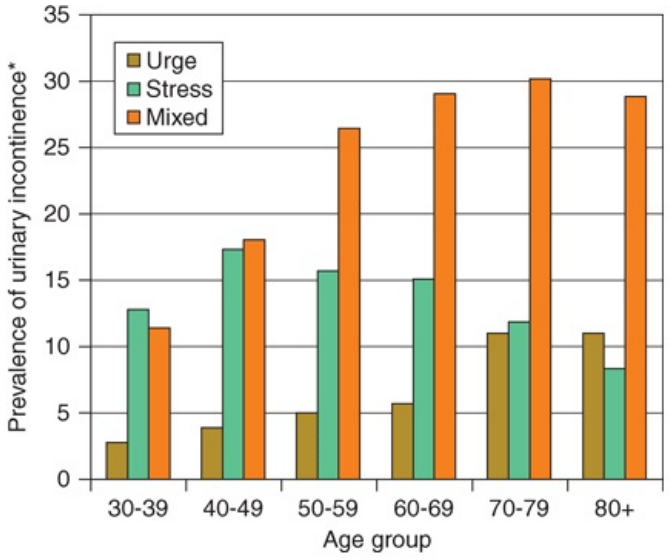

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rates of urge, stress and mixed incontinence by age group in a sample of 3552 women. *Based on a sample of 3553 participants. *(From JL Melville et al: Arch Intern Med 165:537, 2005.)*

Undernutrition and Anorexia

Normal aging is associated with a decline in food intake that is more marked in men than in women. To some extent, food intake is reduced because energy demand declines as a result of a combination of lower physical activity, decline in lean body mass, and slowed rates of protein turnover. Other contributors to decreased food intake include losses of taste sensation, reduced stomach compliance,

higher circulating levels of cholecystokinin, and, in men, low testosterone associated with increased leptin. When food intake decreases to a level below the reduced energy demand, the result is energy malnutrition. Malnutrition in older persons should be considered a geriatric syndrome because it is the result of intrinsic susceptibility due to aging, complicated by multiple superimposed precipitating causes. In addition, many older individuals tend to consume a monotonous diet that lacks sufficient fresh food, fruits, and vegetables, so that intake of important micronutrients is inadequate. Undernutrition in older people is associated with multiple adverse health consequences, including impaired muscle function, decreased bone mass, immune dysfunction, anemia, reduced cognitive function, poor wound healing, delayed recovery from surgery, and increased risk of falls, disability, and mortality. Despite these serious consequences, undernutrition often remains unrecognized until it is very advanced because weight loss tends to be ignored by both patients and physicians. Muscle wasting is a frequent feature of weight loss and malnutrition, often associated with loss of subcutaneous fat. The main causes of weight loss are anorexia, cachexia, sarcopenia, malabsorption, hypermetabolism, and dehydration, almost always in various combinations. Many of these causes can be detected and corrected. Cancer accounts for only 10–15% of cases of weight loss and anorexia in older people. Other important causes include a recent move to a long-term care setting, acute illness (often with inflammation), hospitalization with bed rest for as little as 1–2 days, depression, drugs that cause anorexia and nausea (e.g., digoxin and antibiotics), swallowing problems, oral infections, dental problems, GI pathology, thyroid and other hormonal problems, poverty, and isolation, with reduced access to food. Weight loss may also result from dehydration, possibly related to excess sweating, diarrhea, vomiting, or reduced intake. Early identification is paramount and requires careful weight monitoring. Patients or caregivers should be taught to record weight regularly at home, the patient should be weighed at each clinical encounter, and a record of serial weights should be maintained in the medical record. If malnutrition is suspected, formal assessment should begin with a standardized screening instrument such as the Mini Nutritional Assessment (MNA), the Malnutrition Universal Screening Tool, or the Simplified Nutritional Appetite Questionnaire. The MNA includes questions on appetite, timing of eating, and frequency of meals and taste, and has sensitivity and specificity >75% for future weight loss of ≥5% in older people. Many nutritional supplementations are available and should be initiated early to prevent more severe weight loss and its consequences. When an older patient has malnutrition, the diet should be liberalized and dietary restrictions should be lifted as much as possible. Nutritional supplements should be given between meals to avoid interference with food intake at mealtime. Limited evidence supports the use of any pharmacologic intervention to treat weight loss. The two antianorexic drugs most often prescribed in older persons are megesterol and dronabinol. Both can increase weight, although the gain is mostly fat not muscle, and both have serious side effects. Dronabinol is an excellent drug in the palliative care setting. There is little evidence that intentional weight loss in overweight, older people prolongs life. Weight loss after the age of 70 should probably be limited to those with extreme obesity and should always be medically supervised.

How the Phenotypes of Aging Affect Disease Presentation

Common diseases in older adults may have unexpected and atypical clinical features. Most age-related changes in clinical presentation, evolution, and response to treatment are due to the interaction between the pathophysiology of disease and the development of age-related system dysregulation. Some diseases directly impact aging systems and, therefore, have a devastating impact on frailty and its consequences. Parkinson's disease (PD) and diabetes are described as examples.

Parkinson's disease (Chap. 372) Most cases of PD begin after age 60 years, and incidence increases up to about age 80 years. Brain aging and PD have long been thought to be related. The nigrostriatal system deteriorates with aging, and many older persons tend to develop a mild form of movement disorder characterized by bradykinesia and stooped posture, that mimics mild PD. Interestingly, in PD, older age at presentation is associated with a more severe and rapid decline in gait, balance, postural problems, and cognition. These age-related motor and cognitive manifestations of PD tend to be poorly responsive to levodopa or dopamine agonist treatments, especially in older persons. Interestingly, age at presentation does not correlate with the severity and progression of other classic PD symptoms such as tremor, rigidity, and bradykinesia, and it does not affect response of these symptoms to levodopa. The pattern of PD features in older persons suggests that late-life PD may reflect a failure of the normal cellular compensatory mechanisms in vulnerable brain regions, and this vulnerability is increased by age-related neurodegeneration, making PD symptoms particularly resistant to levodopa treatment. In addition to the motor symptoms, older PD patients tend to have reduced muscle mass (sarcopenia), eating disorders, and poor levels of fitness. Accordingly, PD is a powerful risk factor for frailty and its consequences, including disability, comorbidity, falls, incontinence, chronic pain, and delirium. Use of levodopa and dopaminergic agonists in older PD patients requires complex dosing schedules and therefore slow-release preparations should be preferred. Both dopaminergic and anticholinergic agents increase the risk of confusion and hallucinations. Use of anti-cholinergic agents should generally be avoided. For dopaminergic agents, cognitive side effects can be dose-limiting.

Diabetes (Chap. 344) Both the incidence and prevalence of diabetes mellitus increase with aging.

Among persons age 65 years and older, the prevalence is about 12%, and is higher in African Americans and Hispanics, reflecting the effects of population aging and the obesity epidemic. Diabetes affects all four main aging systems that contribute to frailty. Obesity, especially visceral obesity, is a strong risk factor for diabetes. Diabetes is associated with both reduced muscle mass and accelerated rates of muscle wasting. Diabetic patients have a higher RMR and poor fitness. Diabetes is associated with multiple hormone dysregulation, a pro-inflammatory state, and excess oxidative stress. Finally, diabetes-induced neurodegeneration involves both the central and peripheral nervous systems. Given these characteristics, not surprisingly, patients with DM are more likely to be frail and to have a very high risk of developing physical disability, depression, delirium, cognitive impairment, urinary incontinence, injurious falls, and persistent pain. Thus, the assessment of older diabetic patients should always include screening and risk factor evaluation for these conditions.

In young and adult patients, the main treatment goal has been strict glycemic control aimed at bringing the hemoglobin A_{1c} level to within normal values. However, the risk/benefit ratio is optimized by using less-aggressive glycemic targets. In fact, in the context of a randomized clinical trial, strict glycemic control was associated with higher mortality. Thus, a more reasonable goal for A_{1c} is 7% or slightly below. Treatment goals are altered further in frail older adults who have a high risk of complications of hypoglycemia and a life expectancy of less than 5 years. In these cases, an even less stringent target such as 7–8% should be considered, with A_{1c} monitored every 6 or 12 months. Hypoglycemia is particularly difficult to identify in older diabetic patients because autonomic and nervous system symptoms occur at a lower blood sugar level compared to younger diabetics, although the metabolic reactions and neurologic injury effects are similar in young and older diabetics. The autonomic symptoms of hypoglycemia are often masked by beta blockers. Frail older adults are at even higher risk for serious hypoglycemia than are healthier, higher functioning older adults. In older patients with type 2 diabetes, a history of severe hypoglycemic episodes is associated with higher mortality, more severe microvascular complications, and greater risk of dementia. Thus, patients with suspected or documented episodes of hypoglycemia, especially those who are frail or disabled, need more liberal glucose control goals, careful education about hypoglycemia, and close follow-up with the health provider, possibly accompanied by a caregiver. Chlorpropamide has a prolonged half-life, particularly in older adults, and should be avoided because it is associated with high risk for hypoglycemia. Metformin should be use with caution and only in patients free of severe renal insufficiency. Renal insufficiency should be assessed by calculated glomerular filtration rate, or in very old patients who have reduced muscle mass, by a direct measure of creatinine clearance from a 24-h urine collection. Lifestyle changes in diet and exercise and losing a little weight can prevent or delay diabetes in high-risk individuals and is substantially more effective than metformin. Risk of type 2 diabetes decreased by 58% in a study of diet and exercise and this effect was similar in all ages and all ethnic groups. For comparison, the risk reduction with standard care plus metformin was 31 percent.

Approach to the Care of Older Persons

Organization of Health Care for Older Adults

The complex underlying physiology of aging leads to multiple coexisting medical problems and functional consequences that are often chronic with recurrent exacerbations and remissions. Combined with social consequences of aging such as widowhood and lack of an available caregiver, older adults must sometimes use nonmedical services to meet functional needs. The end result of these medical, functional, and social factors is that older adults use many health care and social support services and settings. Thus it is incumbent on the internist, whether a generalist or specialist, to be familiar with the scope of settings and services that are used by their patients. For many settings, Medicare reimbursement requires a medical order based on specific indications, so the hospitalist or referring physician must be familiar with eligibility requirements. **Table 72-5** summarizes the types of services and payment sources for common settings of care. Older adults who have experienced new disability during a hospitalization are eligible for rehabilitation services. Inpatient rehabilitation requires at least 3 h per day of active rehabilitation and is limited to specific diagnoses. More and more rehabilitative services are provided in postacute settings, where the required intensity of service is less stringent. Postacute settings are also used for complex nursing services such as provision and supervision of long-term parenteral medications or wound care. Under current policy, Medicare only covers postacute care if there is an eligible medical, nursing, or rehabilitation service. Otherwise, nursing home care is not covered by Medicare and must be paid for by expending personal assets until resources are consumed, at which time Medicaid coverage becomes available. Medicaid is a state–federal partnership whose greatest single cost is nursing home care. Thus the need for chronic daily assistance with personal care in a nursing home consumes a large part of most state Medicaid budgets as well as personal assets. Therefore, alternatives to chronic nursing home care are of great interest to states, patients, and families. Some states have developed Medicaid-funded day care programs, sometimes based on the Program for All-inclusive Care of the Elderly (PACE) model. In this situation, older adults who are eligible for both Medicare and Medicaid, and otherwise eligible for chronic nursing home care, can receive coordinated medical and functional services along with a day care setting. For most older adults, a caregiver must be available to provide assistance on weeknights and weekends. Home health services under current policy do not provide chronic functional assistance

in the home, but rather are targeted at episodes of care for medical or rehabilitative services for older adults who are considered home bound. Within the past decade, there has been tremendous growth in a broad spectrum of assisted living settings. Such settings do not provide the degree of 24-h nursing supervision or personal aide care that is provided in traditional nursing homes, although distinctions are becoming blurred. Most assisted living settings provide meals, medication supervision, and homemaking services but often require that residents be capable of transporting themselves to a congregate meal site. Most of these settings accept only private pay from residents and their families and thus are hard to access for older adults with limited resources. Some states are exploring coverage for lower-cost residential care services such as family care homes. Some community agencies, whether private or public, can provide homemaker and home aide services to assist the home-bound older adult with functional needs, but there may be income requirements or need for expensive private payment.

Table 72-5 Care Setting for Older Patients

Setting	Services	Payment Source
Hospital acute care	Medical, surgical, and psychiatric services that cannot be provided in less complex settings	Medicare, Medicaid, and private insurance
Emergency room	Resuscitation, stabilization, triage, disposition	Medicare, Medicaid, and private insurance
Inpatient rehabilitation	Hospital-based residential program providing team-based, physician-supervised, intensive therapeutic rehabilitation for specific diagnoses	Medicare, Medicaid, and private insurance
Outpatient clinic	Chronic, urgent, and preventive services	Medicare, Medicaid, and private insurance
Postacute care	Medical, nursing, and rehabilitative services after hospitalization, often based in hospitals or nursing homes	Medicare up to 100 days with eligibility requirements
Long-term care	Residential program with daily nursing and aide care for persons who are dependent in self-care	Medicaid, private pay, long-term care insurance
Assisted living	Residential program with daily aide care and housing for persons who are dependent in household management	Private pay
Home health care	Nursing and rehabilitative services for episodes of care provided to persons in the community	Medicare, Medicaid
Day programs	Supervised settings providing nursing and aide care for scheduled hours	Private pay, Medicaid

Models of Care Coordination

The complexity and fragmentation of care for complex older adults results in both increased costs and increased risk of iatrogenic complications such as missed diagnoses, adverse medication events, further worsening of function, and even death. These serious consequences have led to a strong interest in care coordination through teams of providers, with the goal to reduce unnecessary costs and prevent adverse events. **Table 72-6** lists examples of evidence-based models of care coordination that were recommended in a recent Institute of Medicine report. While not mentioned as a specific type of team care, modern information technology has substantial promise to provide consistent readily available information across settings and providers. All such team programs are targeted at prevention and management of chronic and complex problems. Each model has demonstrated evidence for benefit in clinical trials or quasi-experimental studies, and some have sufficient findings to support meta-analyses. The evidence for benefit is not always consistent between studies or types of care, but includes some support for improved quality of care, quality of life, function, survival, and health care costs and use. Some models of care are disease-specific and focus on common chronic conditions such as diabetes mellitus, congestive heart failure, chronic obstructive pulmonary disease, or stroke. One challenge to these models is that a majority of older adults will have multiple simultaneous conditions, and thus need services from multiple programs that may not communicate among themselves.

Table 72-6 Evidence-Based Models of Care Coordination for Older Patients (Institute of Medicine 2009)

Model	Team Members	Services
Interdisciplinary primary care "Medical Home"	Primary care physician plus social worker, nurse, nurse practitioner, or other	Coordinate medical and social needs across settings

	care coordinators	
Case management	Nurse or social worker	Provide education and information to patients and families, may communicate with providers and settings
Disease management	Nurse	Health education and follow-up support for specific chronic diseases
Preventive home visits	Physician, nurse, social worker, and others	Structured assessment of physical, mental, functional, and social status in the home setting, with recommendations for care and prevention
Comprehensive outpatient geriatric assessment and management	Physician, nurse, social worker, plus sometimes others such as pharmacist, rehabilitation therapists, psychologist	Structured assessment of physical, mental, functional, and social status in the outpatient setting, with recommendations for care and prevention. Some programs also take responsibility for implementing the recommendations
Pharmaceutical care management	Pharmacist	Review and recommendations regarding the total medication regimen, provided in any setting
Chronic disease self-management	Nurse, health educator, or other health professional	Health education and coaching for specific chronic conditions
Prevention rehabilitation	Rehabilitation therapist	Anticipatory evaluation, therapeutic exercise, and assistive technology in the home or outpatient setting for older adults with physical disability
Caregiver services	Social worker, psychologist, or other health professional	Education, counseling, and resource referral for caregivers of older adults with chronic functional and mental health problems
Hospital discharge/transition coordination	Nurse, nurse practitioner	Care planning and education for patient and family before and after hospital discharge
Hospital at home	Physician, nurse, pharmacist	Diagnostic testing and medical treatments that can replace hospitalization or reduce length of stay for target conditions
Nursing home care coordination	Nurse practitioner or physician assistant	Scheduled assessment and care planning, as well as education for health workers in chronic care settings
Hospital delirium comprehensive care	Physician, nurse	Prevention, screening, and management of delirium in the hospital setting
Comprehensive inpatient geriatric assessment and management	Physician, nurse, social worker plus sometimes others such as pharmacist, rehabilitation therapists, psychologist	Specialized inpatient settings such as acute care of the elderly (ACE) unit or roving multidisciplinary teams that provide evaluation and recommendations for medical, mental health, functional, and social needs. ACE units and some teams take responsibility for implementation of recommendations

Source: Reproduced with permission from Boult et al.

Most models of care are difficult to implement in today's health care system because non physician services are not reimbursed, nor is physician effort that is not incorporated into "face to face" time. Thus several models have largely been developed in the Department of Veterans Affairs Health Care system, Medicare Managed Care providers, or other sponsoring agencies. Medicare has developed a series of demonstration projects that can help further build the evidence base and serve policy makers.

Screening and Prevention in Older Persons

In older adults, prevention tests and interventions are less consistently recommended for all asymptomatic patients. The guidelines fail to address the influence of health status and life expectancy on recommendations, although the benefits of prevention are clearly affected by life expectancy. For example, in most types of cancer, screening provides no benefit in patients with a life expectancy of 5 years or less. More research is needed to build an appropriate evidence base for age- and life expectancy–adapted preventive services. Health behavior modification, especially increasing physical activity and improving nutrition, are probably the two interventions with the highest potential to promote healthy aging.

SCREENING TESTS

- Osteoporosis: Bone mineral density (BMD) at least once after the age of 65 years. There is little evidence that regular monitoring of BMD improves the prediction of fractures. However, because of limitations in the precision of DEXA the minimal interval between evaluations

should be 2–3 years.

- Hypertension: Blood pressure at least once a year, more often in patients with hypertension.
- Diabetes: Serum glucose and hemoglobin A_{1c} every 3 years, more often in patients who are obese or hypertensive.
- Lipid disorders: Lipid panel every 5 years, more often in patients with diabetes or any cardiovascular disease.
- Colorectal cancer: Fecal occult blood test, sigmoidoscopy or colonoscopy, regular schedule up to age 75 years. No consensus guidelines after age 75 years.
- Breast cancer: Mammography every 2 years between ages 50 and 74 years. No consensus guidelines after age 75 years.
- Cervical cancer: Pap smear every 3 years up to age 65 years.

PREVENTIVE INTERVENTIONS

- Influenza: Immunization annually.
- Shingles: Herpes zoster immunization once after age 50 years.
- Pneumonia: Pneumococcal immunization once at age 65 years.
- Myocardial infarction: Daily aspirin in patients with prevalent cardiovascular disease or with poor cardiovascular risk profile.
- Osteoporosis: Calcium 1200 mg daily and vitamin D at least 800 UI daily.

Exercise

Rates of regular physical activity decrease with age and are lowest in older persons. This is unfortunate because physical activity has clear benefits in aging. In older adults, increased physical activity improves physical function, muscle strength, mood, sleep, and metabolic risk profile. Some studies suggest that exercise can improve cognition and prevent dementia, but this is still controversial. Exercise programs, both aerobic and strength training, are feasible and beneficial even in very old and frail individuals. Regular, moderate-intensity exercise can reduce the rate of age-associated decline in physical function. The U.S. Centers for Disease Control and Prevention recommend that older persons should have at least 150 min per week of moderate-intensity aerobic activity (such as brisk walking) and muscle-strengthening activities that work all major muscle groups (legs, hips, back, abdomen, chest, shoulders, and arms) on 2 or more days a week. In the absence of contraindications, more intense and prolonged physical activity provides higher benefits. Frail and sedentary persons may need supervision, at least at the start of the exercise program, to avoid falls and exercise-related injuries.

Nutrition

Older persons are particularly vulnerable to malnutrition and many problems that affect older patients can be addressed by dietary modification. In spite of this, evidence-based guidelines for individualized dietary modification in the elderly are lacking and older people tend to be poorly compliant with dietary recommendations. Basic principles of a healthy diet that are also valid for older persons are as follows:

- Encourage consumption of fruits and vegetables; they are rich in micronutrients, mineral, and fibers. Whole grains are also a good source of fiber. Keep in mind that some of these foods are costly and thus less accessible to low-income persons.
- Good hydration is essential. Fluid intake should be at least 1000 mL daily.
- Encourage the use of fat-free and low-fat dairy products, legumes, poultry, and lean meats. Encourage consumption of fish at least once a week.
- Match intake of energy (calories) to overall energy needs in order to maintain a healthy weight and body mass index (BMI 20–27). If BMI >27, implement a 5–10% calorie restriction.
- Limit consumption of foods with high caloric density, high sugar, and high salt content (less than 6 g per day).
- Limit the intake of foods with a high content of saturated fatty acids and cholesterol.
- Limit alcohol consumption (1 drink per day or less).
- Older persons who have little exposure to UVB radiation are at risk of vitamin D insufficiency. Thus, vitamin D–fortified foods and/or vitamin D supplements should be introduced in the diet.
- Make sure that the diet includes adequate food-related intake of magnesium, vitamin A, and vitamin B_{12}.
- For constipation: increase dietary fiber to 10–25 g and fluid intake to 1500 mL daily. A bulk laxative (methylcellulose or psyllium) can be added.

Novel Interventions to Modify Aging Processes

Aging is a complex process with multiple manifestations at the molecular, cellular, organ, and whole

organism level. The nature of the aging process is still not fully understood, but aging and its effects may be modulated by appropriate interventions. Dietary and genetic alterations can increase healthy life span and prevent the development of dysregulated systems and the aging phenotype in laboratory model organisms. The mechanisms responsible for life span expansion are "food" sensors typically activated in situations of food shortage, such as IGF (insulin-like growth factor)/insulin and the TOR (target of rapamycin) pathways. Accordingly, a reduction in food intake without malnutrition extends the life span by 10–50% in diverse organisms, from yeast to rhesus monkeys. Mechanisms that mediate the effects of caloric restriction are under intensive study because they are potential targets for interventions aimed at counteracting the emergence of the aging phenotype and its deleterious effects in humans. For example, resveratrol, a natural compound found in grape skin that mimics some of the effects of dietary restriction, increases longevity and improves health when fed to mice on a high-fat diet but has little effect in mice on a standard diet. Other compounds that potentially mimic caloric restriction are currently in development and testing. A high prevalence of IGF-1 receptor gene mutation has been found in Ashkenazi Jewish centenarians and in long-lived individuals, suggesting that the downregulation of IGF-1 signaling may promote human longevity. A 20-year 30% dietary restriction applied to adult rhesus monkeys was shown to be associated with reduced cardiovascular and cancer morbidity, reduced signs of aging, and higher longevity, though a second such study did not find increased longevity. In humans, dietary restriction is effective against obesity and reduces insulin resistance, inflammation, blood pressure, c-reactive protein, and intima-media thickness of the carotid arteries. However, the beneficial effects of dietary restriction in humans are still controversial, and some potential negative effects have not been sufficiently studied. An interesting effect of caloric restriction in humans is mitochondrial biogenesis. Mitochondrial dysfunction has emerged as a potentially important underlying contributor to aging. Reduced expression of mitochondrial genes is a strongly conserved feature of aging across different species. Interestingly, mitochondria are the machinery for chemical energy production, and brain and muscle are particularly susceptible to defective mitochondrial function. Thus, declining mitochondrial function may be a direct cause of at least three of the main dysregulated systems contributing to the phenotype of aging.

Other Aspects of Aging

This chapter touched on some of the fundamental aspects of human aging, focusing mostly on those that are relevant to the care of older patients. Many aspects of geriatric medicine could not be addressed due to space limitations. Valuable topics had to be left out, including details of comprehensive geriatric assessment, depression and anxiety, hypertension, orthostatic hypotension, dementia, vision and hearing impairment, osteoporosis, palliative care, prostate disorders, foot problems, and women's health. Some of these topics are treated extensively elsewhere in this text, sometimes with comments on age-specific issues. The interested reader can find resources for further information at the end of this chapter.

Conclusions

The universal process of aging is becoming better understood. There appear to be shared underlying cellular and molecular processes that induce widespread dysregulation in key systems. This dysregulation contributes to clinical manifestations of a frailty phenotype and can be used to understand how to evaluate and manage the older patient.

Acknowledgment

We would like to thank our colleagues who provided criticisms and suggestions for improvement of this chapter. We are particularly indebted to Dr. John Morley for his valuable suggestions on the "Anorexia-Malnutrition" section.

FURTHER READINGS

American Geriatrics Society Panel on Pharmacological Management of Persistent Pain in Older Persons. Pharmacological management of persistent pain in older persons. J Am Geriatr Soc 57:1331, 2009

Boult C et al: Successful models of comprehensive care for older adults with chronic conditions: Evidence for the Institute of Medicine's "retooling for an aging America" report. J Am Geriatr Soc 57:2328, 2009[PMID: 20121991] [Full Text]

Halter JB et al: *Hazzard's Geriatric Medicine & Gerontology*, 6th ed. New York, McGraw-Hill, 2009

Reuben DB et al: *Geriatrics at Your Fingertips: 2010* , 12th ed. New York, American Geriatric Society, 2010

WEBSITES

The American Geriatric Society http://www.americangeriatrics.org/

The Gerontological Society of America http://www.geron.org/

World Health Organization Website on Aging http://www.who.int/topics/ageing/en/

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NUTRIENT REQUIREMENTS AND DIETARY ASSESSMENT: INTRODUCTION

Nutrients are substances that are not synthesized in sufficient amounts in the body and therefore must be supplied by the diet. Nutrient requirements for groups of healthy persons have been determined experimentally. For good health, we require energy-providing nutrients (protein, fat, and carbohydrate), vitamins, minerals, and water. Human requirements for organic nutrients include 9 essential amino acids, several fatty acids, glucose, 4 fat-soluble vitamins, 10 water-soluble vitamins, dietary fiber, and choline. Several inorganic substances, including 4 minerals, 7 trace minerals, 3 electrolytes, and the ultra trace elements, must also be supplied by diet.

The required amounts of the essential nutrients differ by age and physiologic state. Conditionally essential nutrients are not required in the diet but must be supplied to individuals who do not synthesize them in adequate amounts, such as those with genetic defects, those having pathologic states with nutritional implications, and developmentally immature infants. Many other organic and inorganic compounds present in foods have health effects. For example, lead and pesticide residues may have toxic effects.

ESSENTIAL NUTRIENT REQUIREMENTS

Energy

For weight to remain stable, energy intake must match energy output. The major components of energy output are resting energy expenditure (REE) and physical activity; minor sources include the energy cost of metabolizing food (thermic effect of food or specific dynamic action) and shivering thermogenesis (e.g., cold-induced thermogenesis). The average energy intake is about 2600 kcal/d for American men and about 1900 kcal/d for American women, though these estimates vary with body size and activity level. Formulas for estimating REE are useful for assessing the energy needs of an individual whose weight is stable. Thus, for males, REE = 900 + 10*m*, and for females, REE = 700 + 7*m*, where *m* is mass in kilograms. The calculated REE is then adjusted for physical activity level by multiplying by 1.2 for sedentary, 1.4 for moderately active, or 1.8 for very active individuals. The final figure provides an estimate of total caloric needs in a state of energy balance. For further discussion of energy balance in health and disease, see Chap.75.

Protein

Dietary protein consists of both essential and nonessential amino acids that are required for protein synthesis. The nine essential amino acids are histidine, isoleucine, leucine, lysine, methionine/ cystine, phenylalanine/tyrosine, threonine, tryptophan, and valine. Certain amino acids, such as alanine, can also be used for energy and gluconeogenesis. When energy intake is inadequate, protein intake must be increased, because ingested amino acids are diverted into pathways of glucose synthesis and oxidation. In extreme energy deprivation, protein-calorie malnutrition may ensue (Chap. 75).

For adults, the recommended dietary allowance (RDA) for protein is about 0.6 g/kg desirable body mass per day, assuming that energy needs are met and that the protein is of relatively high biologic value. Current recommendations for a healthy diet call for at least 10 to 14% of calories from protein. Most American diets provide at least those amounts. Biologic value tends to be highest for animal proteins, followed by proteins from legumes (beans), cereals (rice, wheat, corn), and roots. Combinations of plant proteins that complement one another in biologic value, or combinations of animal and plant proteins, can increase biologic value and lower total protein requirements.

Protein needs increase during growth, pregnancy, lactation, and rehabilitation after injury or malnutrition. Tolerance to dietary protein is decreased in renal insufficiency (causing uremia) and in liver failure. Normal protein intake can precipitate encephalopathy in patients with cirrhosis of the liver.

Fat and Carbohydrate

Fats are a concentrated source of energy and constitute, on average, 34% of calories in U.S. diets. However, for optimal health, fat intake should total no more than 30% of calories. Saturated fat and trans-fat should be limited to <10% of calories, and polyunsaturated fats to <10% of calories, with monounsaturated fats comprising the remainder of fat intake. At least 45–55% of total calories should be derived from carbohydrates. The brain requires about 100 g/d of glucose for fuel; other tissues use about 50 g/d. Some tissues (e.g., brain and red blood cells) rely on glucose supplied either exogenously or from muscle proteolysis. Over time, adaptations in carbohydrate needs are possible during hypocaloric states.

Water

For adults, 1 to 1.5 mL water per kcal of energy expenditure is sufficient under usual conditions to allow for normal variations in physical activity, sweating, and solute load of the diet. Water losses include 50 to 100 mL/d in the feces; 500 to 1000 mL/d by evaporation or exhalation; and, depending on the renal solute load, ≥1000 mL/d in the urine. If external losses increase, intakes must increase accordingly to avoid underhydration. Fever increases water losses by approximately 200 mL/d per °C; diarrheal losses vary, but may be as great as 5 L/d in severe diarrhea. Heavy sweating and vomiting also increase water losses. When renal function is normal and solute intakes are adequate, the kidneys can adjust to increased water intake by excreting up to 18 L/d of excess water (Chap. 340). However, obligatory urine outputs can compromise hydration status when there is inadequate intake or when losses increase in disease or kidney damage.

Infants have high requirements for water because of their large ratio of surface area to volume, the limited capacity of the immature kidney to handle high renal solute loads, and their inability to communicate their thirst. Increased water needs during pregnancy are about 30 mL/d. During lactation, milk production increases water requirements so that approximately 1000 mL/d of additional water is needed, or 1 mL for each mL of milk produced. Special attention must be paid to the water needs of the elderly, who have reduced total body water and blunted thirst sensation, and are more likely to be taking medications such as diuretics.

Other Nutrients

See Chap. 74 for detailed descriptions of vitamins and trace minerals.

DIETARY REFERENCE INTAKES AND RECOMMENDED DIETARY ALLOWANCES

Fortunately, human life and well-being can be maintained within a fairly wide range for most nutrients. However, the capacity for adaptation is not infinite—too much, as well as too little, intake of a nutrient could have adverse effects or alter the health benefits conferred by another nutrient. Therefore, benchmark recommendations regarding nutrient intakes have been developed to guide clinical practice. These quantitative estimates of nutrient intakes are collectively referred to as the *dietary reference intakes* (DRIs). The DRIs supplant the *recommended daily allowances* (RDAs), the single reference values used in the United States until the early 1990s. DRIs include the *estimated average requirement* (EAR) for nutrients as well as other reference values used for dietary planning for individuals: the RDA, the *adequate intake* (AI), and the tolerable *upper level* (UL). The DRI also include acceptable macronutrient distribution ranges (AMDR) for protein, fat, and carbohydrate. The current DRIs for vitamins and elements are provided in Tables 73-1 and 73-2, respectively.

Table 73-1 Dietary Reference Intakes: Recommended Intakes for Individuals—Vitamins

Life-Stage Group	Vitamin, μg/d					Thiamine, mg/d	Riboflavin, mg/d	Niacin, mg/d[e]	Vitamin B_6, mg/d	Folate, μg/d[f]	Vitamin B_{12}, μg/d	Pantothenic Acid, mg/d	Biotin, μg/d	Choline, mg/d[g]
	A[a]	C	D[b,c]	E[d]	K									
Infants														
0–6 mo	400	40	5	4	2.0	0.2	0.3	2	0.1	65	0.4	1.7	5	125
712 mo	500	50	5	5	2.5	0.3	0.4	4	0.3	80	0.5	1.8	6	150

Children															
1–3 y	**300**	**15**	5	**6**	30	**0.5**	**0.5**	**6**	**0.5**	**150**	**0.9**	2		8	200
4–8 y	**400**	**25**	5	**7**	55	**0.6**	**0.6**	**8**	**0.6**	**200**	**1.2**	3		12	250
Males															
9–13 y	**600**	**45**	5	**11**	60	**0.9**	**0.9**	**12**	**1.0**	**300**	**1.8**	4		20	375
14–18 y	**900**	**75**	5	**15**	75	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5		25	550
19–30 y	**900**	**90**	5	**15**	120	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5		30	550
31–50 y	**900**	**90**	5	**15**	**120**	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5		30	550
51–70 y	**900**	**90**	10	**15**	120	**1.2**	**1.3**	**16**	**1.7**	**400**	**2.4**[h]	5		30	550
>70 y	**900**	90	**15**	**15**	120	**1.2**	**1.3**	**16**	**1.7**	**400**	**2.4**[h]	5		30	550
Females															
9–13 y	**600**	**45**	5	**11**	60	**0.9**	**0.9**	**12**	**1.0**	**300**	**1.8**	4		20	375
14–18 y	**700**	**65**	5	**15**	75	**1.0**	**1.0**	**14**	**1.2**	**400**[i]	**2.4**	5		25	400
19–30 y	**700**	**75**	5	**15**	90	**1.1**	**1.1**	**14**	**1.3**	**400**[i]	**2.4**	5		30	425
31–50 y	**700**	**75**	5	**15**	90	**1.1**	**1.1**	**14**	**1.3**	**400**[i]	**2.4**	5		30	425
51–70 y	**700**	**75**	10	**15**	90	**1.1**	**1.1**	**14**	**1.5**	**400**	**2.4**[h]	5		30	425
>70 y	**700**	**75**	15	**15**	90	**1.1**	**1.1**	**14**	**1.5**	**400**	**2.4**[h]	5		30	425
Pregnancy															
≤18 y	**750**	**80**	5	**15**	75	**1.4**	**1.4**	**18**	**1.6**	**600**[j]	**2.6**	6		30	450
19–30 y	**770**	**85**	5	**15**	90	**1.4**	**1.4**	**18**	**1.9**	**600**[j]	**2.6**	6		30	450
31–50 y	**770**	**85**	5	**15**	90	**1.4**	**1.4**	**18**	**1.9**	**600**[j]	**2.6**	6		30	450
Lactation															
≤18 y	**1200**	**115**	5	**19**	75	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7		35	550
19–30 y	**1300**	**120**	5	**19**	90	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7		35	550
31–50 y	**1300**	**120**	5	**19**	90	**1.4**	**1.6**	**17**	**2.0**	**500**	**2.8**	7		35	550

Note: This table presents recommended dietary allowances (RDAs) in **bold type** and adequate intakes (AIs) in ordinary type. RDAs and AIs may both be used as goals for individual intake. RDAs are set to meet the needs of almost all individuals (97 to 98%) in a group. For healthy breastfed infants, the AI is the mean intake. The AI for other life stage and gender groups is believed to cover needs of all individuals in the group, but lack of data or uncertainty in the data prevent being able to specify with confidence the percentage of individuals covered by this intake.

[a]As retinol activity equivalents (RAEs). 1 RAE = 1 μg retinol, 12 μg β-carotene, 24 μg α-carotene, or 24 μg β-cryptoxanthin. To calculate RAEs from retinol equivalents (REs) of provitamin A carotenoids in foods, divide the REs by 2. For preformed vitamin A in foods or supplements and for provitamin A carotenoids in supplements, 1 RE = 1 RAE.

[b]As calciferol. 1 μg calciferol = 40 IU vitamin D.

[c]In the absence of adequate exposure to sunlight.

[d]As α-tocopherol. α-Tocopherol includes RRR-α-tocopherol, the only form of α-tocopherol that occurs naturally in foods, and the 2R-stereoisomeric forms of α-tocopherol (RRR-, RSR-, RRS-, and RSS-α-tocopherol) that occur in fortified foods and supplements. It does not include the 2S-stereoisomeric forms of -tocopherol (SRR-, SSR-, SRS-, and SSS-α-tocopherol), also found in fortified foods and supplements.

[e]As niacin equivalents (NE). 1 mg of niacin = 60 mg of tryptophan; 0–6 months = preformed niacin (not NE).

[f]As dietary folate equivalents (DFEs). 1 DFE = 1 μg food folate = 0.6 μg of folic acid from fortified food or as a supplement consumed with food = 0.5 μg of a supplement taken on an empty stomach.

[g]Although AIs have been set for choline, there are few data to assess whether a dietary supply of choline is needed at all stages of the life cycle, and it may be that the choline requirement can be met by endogenous synthesis at some of these stages.

[h]Because 10 to 30% of older people may malabsorb food-bound B_{12}, it is advisable for those >50 years to meet their RDA mainly by consuming foods fortified with B_{12} or a supplement containing B_{12}.

[i]In view of evidence linking inadequate folate intake with neural tube defects in the fetus, it is recommended that all women capable of becoming pregnant consume 400 μg from supplements or fortified foods in addition to intake of food folate from a varied diet.

[j]It is assumed that women will continue consuming 400 μg from supplements or fortified food until their pregnancy is confirmed and they enter prenatal care, which ordinarily occurs after the end of the periconceptional period—the critical time for formation of the neural tube.

Source: Food and Nutrition Board, Institute of Medicine—National Academy of Sciences Dietary Reference Intakes, 2000, 2002, reprinted with permission. Courtesy of the National Academy Press, Washington, DC. http://www.nap.edu.

Table 73-2 Dietary Reference Intakes: Recommended Intakes for Individuals—Elements

Life-Stage Group	Calcium, mg/d	Chromium, μg/d	Copper, μg/d	Fluoride, mg/d	Iodine, μg/d	Iron, mg/d	Magnesium, mg/d	Manganese, mg/d	Molybdenum, μg/d	Phosphorus mg/d	Selenium, μg/d	Zinc, mg/d
Infants												
0–6 mo	210	0.2	200	0.01	110	0.27	30	0.003	2	100	15	2
7–12 mo	270	5.5	220	0.5	130	11	75	0.6	3	275	20	3
Children												
1–3 y	500	11	**340**	0.7	**90**	**7**	**80**	1.2	**17**	**460**	**20**	**3**
4–8 y	800	15	**440**	1	**90**	**10**	**130**	1.5	**22**	**500**	**30**	**5**
Males												
9–13 y	1300	25	**700**	2	**120**	**8**	**240**	1.9	**34**	**1250**	**40**	**8**
14–18 y	1300	35	**890**	3	**150**	**11**	**410**	2.2	**43**	**1250**	**55**	**11**
19–30 y	1000	35	**900**	4	**150**	**8**	**400**	2.3	**45**	**700**	**55**	**11**
31–50 y	1000	35	**900**	4	**150**	**8**	**420**	2.3	**45**	**700**	**55**	**11**
51–70 y	1200	30	**900**	4	**150**	**8**	**420**	2.3	**45**	**700**	**55**	**11**
>70 y	1200	30	**900**	4	**150**	**8**	**420**	2.3	**45**	**700**	**55**	**11**
Females												
9–13 y	1300	21	**700**	2	**120**	**8**	**240**	1.6	**34**	**1250**	**40**	**8**
14–18 y	1300	24	**890**	3	**150**	**15**	**360**	1.6	**43**	**1250**	**55**	**9**

19–30 y	1000	25	**900**	3	**150**	**18**	**310**	1.8	**45**	**700**	**55**	**8**
31–50 y	1000	25	**900**	3	**150**	**18**	**320**	1.8	**45**	**700**	**55**	**8**
51–70 y	1200	20	**900**	3	**150**	**8**	**320**	1.8	**45**	**700**	**55**	**8**
>70 y	1200	20	**900**	3	**150**	**8**	**320**	1.8	**45**	**700**	**55**	**8**
Pregnancy												
≤18 y	1300	29	**1000**	3	**220**	**27**	**400**	2.0	**50**	**1250**	**60**	**12**
19–30 y	1000	30	**1000**	3	**220**	**27**	**350**	2.0	**50**	**700**	**60**	**11**
31–50 y	1000	30	**1000**	3	**220**	**27**	**360**	2.0	**50**	**700**	**60**	**11**
Lactation												
≤18 y	1300	44	**1300**	3	**290**	**10**	**360**	2.6	**50**	**1250**	**70**	**13**
19–30 y	1000	45	**1300**	3	**290**	**9**	**310**	2.6	**50**	**700**	**70**	**12**
31–50 y	1000	45	**1300**	3	**290**	**9**	**320**	2.6	**50**	**700**	**70**	**12**

Note: This table presents recommended dietary allowances (RDAs) in **bold type** and adequate intakes (AIs) in ordinary type. RDAs and AIs may both be used as goals for individual intake. RDAs are set to meet the needs of almost all individuals (97 to 98%) in a group. For healthy breastfed infants, the AI is the mean intake. The AI for other life stage and gender groups is believed to cover needs of all individuals in the group, but lack of data or uncertainty in the data prevent being able to specify with confidence the percentage of individuals covered by this intake.

Source: Food and Nutrition Board, Institute of Medicine—National Academy of Sciences Dietary Reference Intakes, 2000, 2002, reprinted with permission. Courtesy of the National Academy Press, Washington, DC. http://www.nap.edu.

Estimated Average Requirement

When florid manifestations of the classic dietary deficiency diseases such as rickets (deficiency of vitamin D and calcium), scurvy (deficiency of vitamin C), xerophthalmia (deficiency of vitamin A), and protein-calorie malnutrition were common, nutrient adequacy was inferred from the absence of their clinical signs. Later, biochemical and other changes were found to be evident long before the clinical deficiency became apparent. Consequently, criteria of adequacy are now based on biologic markers when they are available. Current efforts focus on the amount of a nutrient that reduces the risk of chronic degenerative diseases. Priority is given to sensitive biochemical, physiologic, or behavioral tests that reflect early changes in regulatory processes; maintenance of body stores of nutrients; or, if available, the amount of a nutrient that minimizes risk of chronic degenerative disease.

The EAR is the amount of a nutrient estimated to be adequate for half of the healthy individuals of a specific age and sex. The types of evidence and criteria used to establish nutrient requirements vary by nutrient, age, and physiologic group. The EAR is not an effective estimate of nutrient adequacy in individuals because it is a median requirement for a group; 50% of individuals in a group fall below the requirement and 50% fall above it. Thus, a person with a usual intake at the EAR has a 50% risk of an inadequate intake. For these reasons, other standards, described below, are more useful for clinical purposes.

Recommended Dietary Allowances

The RDA is the average daily dietary intake level that meets the nutrient requirements of nearly all healthy persons of a specific sex, age, life stage, or physiologic condition (such as pregnancy or lactation). The RDA is the nutrient-intake goal for planning diets of individuals.

The RDA is defined statistically as two standard deviations (SD) above the EAR to ensure that the needs of any given individual are met. Recommendations for individuals of a given age, sex, and weight are easily obtained from a Web-based calculator at *http://fnic.nal.usda.gov/interactiveDRI/*. This online tool allows health professionals to calculate daily nutrient recommendations for dietary planning based on the DRIs for individuals.

The RDAs are used to formulate food guides such as the U.S. Department of Agriculture (USDA) Food Guide Pyramid for individuals, to create food-exchange lists for therapeutic diet planning, and as a standard for describing the nutritional content of processed foods and nutrient-containing dietary supplements. The nutrient content in a food is stated by weight or as a percent of the daily value (DV), a variant of the RDA used in food labeling on the nutrition facts panel that, for an adult, represents the highest RDA for an adult consuming 2000 kcal/d.

The risk of dietary inadequacy increases as intake falls below the RDA. However, the RDA is an overly generous criterion for evaluating nutrient adequacy. For example, by definition the RDA exceeds the actual requirements of all but about 2 to 3% of the population. Therefore, many people whose intake falls below the RDA may still be getting enough of the nutrient.

Adequate Intake

It is not possible to set an RDA for some nutrients that do not have an established EAR. In this circumstance, the AI is based on observed, or experimentally determined, approximations of nutrient intakes in healthy people. In the DRIs, AIs rather than RDAs are proposed for infants up to age 1 year, as well as for calcium, chromium, vitamin D, fluoride, manganese, pantothenic acid, biotin, and choline for persons of all ages. Vitamin D and calcium are currently being reevaluated, and more precise values may be available in the near future.

Tolerable Upper Levels of Nutrient Intake

Excessive nutrient intake can disturb body functions and cause acute, progressive, or permanent disabilities. The tolerable UL is the highest level of chronic nutrient intake (usually daily) that is unlikely to pose a risk of adverse health effects for most of the population. Data on the adverse effects of large amounts of many nutrients are unavailable or too limited to establish a UL. Therefore, the lack of a UL does *not* mean that the risk of adverse effects from high intake is nonexistent. Healthy individuals derive no established benefit from consuming nutrient levels above the RDA or AI. Nutrients in commonly eaten foods rarely exceed the UL. However, highly fortified foods and dietary supplements provide more concentrated amounts of nutrients per serving and, thus, pose a potential risk of toxicity. Nutrient supplements are labeled with supplement facts that express the amount of nutrient in absolute units or as the percent of the DV provided per recommended serving size. Total nutrient consumption, including foods; supplements; and over-the-counter medications, such as antacids, should not exceed RDA levels.

Acceptable Macronutrient Distribution Ranges

The AMDR is a range of energy providing intakes that the Food and Nutrition Board considers to be healthful for macronutrients. These ranges are 10–35% calories for protein, 20–35% calories for fat, and 45–65% of calories for carbohydrate. Alcohol, which also provides energy, is not a nutrient and recommendations, therefore, are not provided for it.

FACTORS ALTERING NUTRIENT NEEDS

The DRIs are affected by age, sex, rate of growth, pregnancy, lactation, physical activity, concomitant diseases, drugs, and dietary composition. If requirements for nutrient sufficiency are close to levels indicating excess, dietary planning is difficult.

Physiologic Factors

Growth, strenuous physical activity, pregnancy, and lactation increase needs for energy and several essential nutrients. Energy needs rise during pregnancy due to the demands of fetal growth and during lactation because of the increased energy required for milk production. Energy needs decrease with loss of lean body mass, the major determinant of REE. Because both health and physical activity tend to decline with age, energy needs of older persons, especially those over 70, tend to be less than those of younger persons.

Dietary Composition

Dietary composition affects the biologic availability and use of nutrients. For example, the absorption of iron may be impaired by high amounts of calcium or lead; also, non-heme iron uptake may be impaired by the lack of ascorbic acid and amino acids in the meal. Protein use by the body may be decreased when essential amino acids are not present in sufficient amounts. Animal foods, such as milk, eggs, and meat, have high biologic values with most of the needed amino acids present in adequate amounts. Plant proteins in corn (maize), soy, and wheat have lower biologic values and must be combined with other plant or animal proteins to achieve optimal use by the body.

Route of Administration

The RDAs apply only to oral intakes. When nutrients are administered parenterally, similar values can sometimes be used for amino acids, carbohydrates, fats, sodium, chloride, potassium, and most of the vitamins, because their intestinal absorption is nearly 100%. However, the oral

bioavailability of most mineral elements may be only half that obtained by parenteral administration. For some nutrients that are not readily stored in the body or cannot be stored in large amounts, timing of administration may also be important. For example, amino acids cannot be used for protein synthesis if they are not supplied together; instead, they will be used for energy production.

Disease

Specific dietary deficiency diseases include: protein-calorie malnutrition; iron, iodine, and vitamin A deficiency; megaloblastic anemia due to vitamin B12 or folic acid deficiency; vitamin D–deficiency rickets and osteomalacia; and scurvy, beriberi, and pellagra (Chaps. 74 and 75). Each deficiency disease is characterized by imbalances at the cellular level between the supply of nutrients or energy and the body's nutritional needs for growth, maintenance, and other functions. Imbalances and excess in nutrient intakes are recognized as risk factors for certain chronic degenerative diseases, such as saturated fat and cholesterol in coronary artery disease; sodium in hypertension; obesity in hormone-dependent endometrial and breast cancers; and ethanol in alcoholism. Because the etiology and pathogenesis of these disorders are multifactorial, diet is only one of many risk factors. Osteoporosis, for example, is associated with calcium deficiency, as well as risk factors related to environment (e.g., smoking, sedentary lifestyle), physiology (e.g., estrogen deficiency), genetic determinants (e.g., defects in collagen metabolism), and drug use (chronic steroids) (Chap. 354).

DIETARY ASSESSMENT

In clinical situations, nutritional assessment is an iterative process that involves: (1) screening for malnutrition, (2) assessing the diet and other data to establish either the absence or presence of malnutrition and its possible causes, (3) planning and implementing the most appropriate nutritional therapy, and (4) reassessing intakes to make sure that they were consumed. Some disease states affect the bioavailability, requirements, use, or excretion of specific nutrients. In these circumstances, specific measurements of various nutrients or their biomarkers may be required to ensure adequate replacement (Chap. 73).

Most health care facilities have nutrition-screening processes in place for identifying possible malnutrition after hospital admission. Nutritional screening is required by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), but there are no universally recognized or validated standards. The factors that are usually assessed include: abnormal weight for height or body mass index (e.g., BMI <19 or >25); reported weight change (involuntary loss or gain of >5 kg in the past 6 months) (Chap. 80); diagnoses with known nutritional implications (metabolic disease, any disease affecting the gastrointestinal tract, alcoholism, and others); present therapeutic dietary prescription; chronic poor appetite; presence of chewing and swallowing problems or major food intolerances; need for assistance with preparing or shopping for food, eating, or other aspects of self care; and social isolation. Reassessment of nutrition status should occur periodically in hospitalized patients—at least once every week.

A more complete dietary assessment is indicated for patients who exhibit a high risk of or frank malnutrition on nutrition screening. The type of assessment varies based on the clinical setting, severity of the patient's illness, and stability of his or her condition.

Acute Care Settings

In acute-care settings, anorexia, various diseases, test procedures, and medications can compromise dietary intake. Under such circumstances, the goal is to identify and avoid inadequate intake and assure appropriate alimentation. Dietary assessment focuses on what patients are currently eating, whether or not they are able and willing to eat, and whether or not they experience any problems with eating. Dietary intake assessment is based on information from observed intakes; medical record; history; clinical examination; and anthropometric, biochemical, and functional status. The objective is to gather enough information to establish the likelihood of malnutrition due to poor dietary intake or other causes to assess whether nutritional therapy is indicated (Chap. 76).

Simple observations may suffice to suggest inadequate oral intake. These include dietitians' and nurses' notes, the amount of food eaten on trays, frequent tests and procedures that are likely to cause meals to be skipped, nutritionally inadequate diet orders such as clear liquids or full liquids for more than a few days, fever, gastrointestinal distress, vomiting, diarrhea, a comatose state, and diseases or treatments that involve any part of the alimentary tract. Acutely ill patients with diet-related diseases such as diabetes need assessment because an inappropriate diet may exacerbate these conditions and adversely affect other therapies. Abnormal biochemical values [serum albumin levels <35 g/L (<3.5 mg/dL); serum cholesterol levels <3.9 mmol/L (<150 mg/dL)] are nonspecific but may also indicate a need for further nutritional assessment.

Most therapeutic diets offered in hospitals are calculated to meet individual nutrient requirements and the RDA *if they are eaten*. Exceptions include clear liquids, some full-liquid diets, and test diets (such as preparation for gastrointestinal procedures), which are inadequate for several nutrients and should not be used, if possible, for more than 24 h. As much as half of the food served to hospitalized patients is not eaten, and so it cannot be assumed that the intakes of hospitalized patients are adequate. Dietary assessment should compare how much and what food the patient has consumed with the diet that has been provided. Major deviations in intakes of energy, protein, fluids, or other nutrients of special concern for the patient's illness should be noted and corrected.

Nutritional monitoring is especially important for patients who are very ill and who have extended lengths of stay. Patients who are fed by special enteral and parenteral routes also require special nutritional assessment and monitoring by physicians and/or dietitians with certification in nutrition support (Chap. 76).

Ambulatory Settings

The aim of dietary assessment in the outpatient setting is to determine whether or not the patient's usual diet is a health risk in itself or if it contributes to existing chronic disease-related problems. Dietary assessment also provides the basis for planning a diet that fulfills therapeutic goals while ensuring patient adherence. The outpatient dietary assessment should review the adequacy of present and usual food intakes, including vitamin and mineral supplements, medications, and alcohol, because all of these may affect the patient's nutritional status. The assessment should focus on the dietary constituents that are most likely to be involved or compromised by a specific diagnosis, as well as any co-morbidities that are present. More than one day's intake should be reviewed to provide a better representation of the usual diet.

There are many ways to assess the adequacy of the patient's habitual diet. These include a food guide, a food-exchange list, a diet history, or a food-frequency questionnaire. A commonly used food guide for healthy persons is the USDA's food pyramid, which is useful as a basis for identifying inadequate intakes of essential nutrients, as well as likely excesses in fat, saturated fat, sodium, sugar, and alcohol (Table 73-3). The Web version of the guide provides a calculator that tailors the number of servings suggested for healthy patients of different weights, sexes, ages, and life-cycle stages: http://www.mypyramidtracker.gov/planner/launchpage.aspx. Patients who follow ethnic or unusual dietary patterns may need extra instruction on how foods should be categorized, as well as the appropriate portion sizes that constitute a serving. The process of reviewing the guide with patients helps them transition to healthier dietary patterns and identifies food groups eaten in excess of recommendations or in insufficient quantities. For those on therapeutic diets, assessment against food-exchange lists may be useful. These include, for example, the American Diabetes Association food-exchange lists for diabetes or the American Dietetic Association food-exchange lists for renal disease.

Table 73-3 My Pyramid: The USDA Food Guide Pyramid for Healthy Persons

Servings and Examples of Standard Portion Sizes	Lower: 1600 kcal	Moderate: 2200 kcal	Higher: 2800 kcal
Fruits, cups	1.5	2	2.5
Vegetables, cups	2	3	3.5
Grains, oz eq	5	7	10
(1 slice bread, 1 cup ready to eat cereal, 0.5 cup cooked rice, pasta, cooked cereal)			
Meat and beans, oz eq	5	6	7
(1 oz lean meat, poultry, or fish; 1 egg, 1 Tbsp. peanut butter, 0.25 cup cooked dry beans, or 0.5 oz nuts or seeds)			
Milk, cups	3	3	3
(1 cup milk or yogurt, 1.5 oz natural or 2 oz processed cheese)			
Oils, tsp	5	6	8
Discretionary calorie allowance, kcal (remaining calories after accounting for all of the above)	132	290	426

Abbreviation: oz eq, ounce equivalent.

Source: Data from United States Department of Agriculture. http://www.MyPyramid.com.

NUTRITIONAL STATUS ASSESSMENT

Full nutritional status assessment is reserved for seriously ill patients and those at very high nutritional risk when the cause of malnutrition is still uncertain after initial clinical evaluation and dietary assessment. It involves multiple dimensions, including documentation of dietary intake, anthropometric measurements, biochemical measurements of blood and urine, clinical examination, health history, and functional status. Therapeutic dietary prescriptions and menu plans for most diseases are available from most hospitals and from the American Dietetic Association. For further discussion of nutritional assessment, see Chap. 75.

GLOBAL CONSIDERATIONS

The DRI such as the EAR, UL, and energy needs are estimates of physiological requirements based on experimental evidence. Assuming appropriate adjustments are made for age, sex, body size, and physical-activity level, they should be applicable to individuals in most parts of the world. However, the AI are based on customary and adequate intakes in U.S. and Canadian populations, which appear to be compatible with good health, rather than being based on a large body of direct experimental evidence. Similarly, the AMDR represent expert opinion of approximate intakes of energy providing nutrients that are healthful in these North American populations. As such, they should be used with caution in other settings. Nutrient-based standards like the DRI have also been developed by the World Health Organization/Food and Agricultural Organization of the United Nations (WHO/FAO) and are available at their Web site: http://wwww.who.int/nutrition/topics/nutrecomm/en/index.html. The different standards have many similarities in their basic concepts, definitions, and nutrient-recommendation levels, but there are some differences from the DRI, due to the functional criteria chosen, environmental differences, the timeliness of the evidence reviewed, and expert judgment.

FURTHER READINGS

Kaiser MJ et al: Validation of the mini-nutritional assessment short form (MNA-SF): A practical tool for identification of nutritional status. J Nutr Health Aging 13: 782, 2009[PMID: 19812868] [Full Text]

King JC, Garza C: Harmonization of nutrient intake values. Food Nutr Bull 28:S1, 2007

Gibson RS: *Principles of Nutritional Assessment* , 2nd ed, London, Oxford University Press, 2005

Murphy SP et al: Multivitamin-multimineral supplements' effect on total nutrient intake Am J Clin Nutr 85: 280S, 2007

Otten JJ et al (eds): *The Dietary Reference Intakes: The Essential Guide to Nutrient Requirements* Washington Washington, DC, National Academy Press, 2006

Shils ME et al (eds): *Modern Nutrition in Health and Disease,* 10th ed, Philadelphia, Lippincott, Williams and Wilkins, 2005

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VITAMIN AND TRACE MINERAL DEFICIENCY AND EXCESS: INTRODUCTION

Vitamins and trace minerals are required constituents of the human diet since they are inadequately synthesized or not synthesized in the human body. Only small amounts of these substances are needed to carry out essential biochemical reactions (e.g., by acting as coenzymes or prosthetic groups). Overt vitamin or trace mineral deficiencies are rare in Western countries due to a plentiful, varied, and inexpensive food supply; however, multiple nutrient deficiencies may appear together in persons who are chronically ill or alcoholic. After gastric bypass surgery, patients are at high risk for multiple nutrient deficiencies. Moreover, subclinical vitamin and trace mineral deficiencies, as diagnosed by laboratory testing, are quite common in the normal population, especially in the geriatric age group.

Victims of famine, emergency-affected and displaced populations, and refugees are at increased risk for protein-energy malnutrition and classic micronutrient deficiencies (vitamin A, iron, iodine) as well as for thiamine (beriberi), riboflavin, vitamin C (scurvy), and niacin (pellagra) overt deficiencies.

Body stores of vitamins and minerals vary tremendously. For example, vitamin B12 and vitamin A stores are large, and an adult may not become deficient for 1 or more years after being on a deficient diet. However, folate and thiamine may become depleted within weeks among those eating a deficient diet. Therapeutic modalities can deplete essential nutrients from the body; for example, hemodialysis removes water-soluble vitamins, which must be replaced by supplementation.

There are several roles for vitamins and trace minerals in diseases: (1) deficiencies of vitamins and minerals may be caused by disease states such as malabsorption; (2) both deficiency and excess of vitamins and minerals can cause disease in and of themselves (e.g., vitamin Aintoxication and liver disease); and (3) vitamins and minerals in high doses may be used as drugs (e.g., niacin for hypercholesterolemia). The hematologic-related vitamins and minerals (Chaps. 103, 105) either are not considered or are considered only briefly in this chapter, as are the bone-related vitamins and minerals (vitamin D, calcium, phosphorus; Chap. 352), since they are covered elsewhere **(Tables 74-1** and **74-2** and **Fig. 74-1)**.

Table 74-1 Principal Clinical Findings of Vitamin Malnutrition

Nutrient	Clinical Finding	Dietary Level per Day Associated with Overt Deficiency in Adults	Contributing Factors to Deficiency
Thiamine	Beriberi: neuropathy, muscle weakness and wasting, cardiomegaly, edema, ophthalmoplegia, confabulation	<0.3 mg/1000 kcal	Alcoholism, chronic diuretic use, hyperemesis
Riboflavin	Magenta tongue, angular stomatitis, seborrhea, cheilosis	<0.6 mg	—
Niacin	Pellagra: pigmented rash of sun-exposed areas, bright red tongue, diarrhea, apathy, memory loss, disorientation	<9.0 niacin equivalents	Alcoholism, vitamin B6 deficiency, riboflavin deficiency, tryptophan deficiency
Vitamin B6	Seborrhea, glossitis, convulsions, neuropathy, depression, confusion, microcytic anemia	<0.2 mg	Alcoholism, isoniazid
Folate	Megaloblastic anemia, atrophic glossitis, depression, ↑ homocysteine	<100 μg/d	Alcoholism, sulfasalazine, pyrimethamine, triamterene
Vitamin B12	Megaloblastic anemia, loss of vibratory and position sense, abnormal gait, dementia, impotence, loss of bladder and bowel control, ↑ homocysteine, ↑	<1.0 μg/d	Gastric atrophy (pernicious anemia), terminal ileal disease, strict vegetarianism, acid-reducing drugs (e.g., H2

	methylmalonic acid		blockers)
Vitamin C	Scurvy: petechiae, ecchymosis, coiled hairs, inflamed and bleeding gums, joint effusion, poor wound healing, fatigue	<10 mg/d	Smoking, alcoholism
Vitamin A	Xerophthalmia, night blindness, Bitot's spots, follicular hyperkeratosis, impaired embryonic development, immune dysfunction	<300 μg/d	Fat malabsorption, infection, measles, alcoholism, protein-energy malnutrition
Vitamin D	Rickets: skeletal deformation, rachitic rosary, bowed legs; osteomalacia	<2.0 μg/d	Aging, lack of sunlight exposure, fat malabsorption, deeply pigmented skin
Vitamin E	Peripheral neuropathy, spinocerebellar ataxia, skeletal muscle atrophy, retinopathy	Not described unless underlying contributing factor is present	Occurs only with fat malabsorption or genetic abnormalities of vitamin E metabolism/transport
Vitamin K	Elevated prothrombin time, bleeding	<10 μg/d	Fat malabsorption, liver disease, antibiotic use

Table 74-2 Deficiencies and Toxicities of Metals

Element	Deficiency	Toxicity	Tolerable Upper (Dietary) Intake Level
Boron	No biologic function determined	Developmental defects, male sterility, testicular atrophy	20 mg/d (extrapolated from animal data)
Calcium	Reduced bone mass, osteoporosis	Renal insufficiency (milk-alkali syndrome), nephrolithiasis, impaired iron absorption	2500 mg/d (milk-alkali)
Copper	Anemia, growth retardation, defective keratinization and pigmentation of hair, hypothermia, degenerative changes in aortic elastin, osteopenia, mental deterioration	Nausea, vomiting, diarrhea, hepatic failure, tremor, mental deterioration, hemolytic anemia, renal dysfunction	10 mg/d (liver toxicity)
Chromium	Impaired glucose tolerance	Occupational: renal failure, dermatitis, pulmonary cancer	ND
Fluoride	↑ Dental caries	Dental and skeletal fluorosis, osteosclerosis	10 mg/d (fluorosis)
Iodine	Thyroid enlargement, ↓ T4, cretinism	Thyroid dysfunction, acne-like eruptions	1100 μg/d (thyroid dysfunction)
Iron	Muscle abnormalities, koilonychia, pica, anemia, ↓ work performance, impaired cognitive development, premature labor, ↑ perinatal maternal mortality	Gastrointestinal effects (nausea, vomiting, diarrhea, constipation), iron overload with organ damage, acute systemic toxicity	45 mg/d of elemental iron (GI side effects)
Manganese	Impaired growth and skeletal development, reproduction, lipid and carbohydrate metabolism; upper body rash	General: Neurotoxicity, Parkinson-like symptoms Occupational: Encephalitis-like syndrome, Parkinson-like syndrome, psychosis, pneumoconiosis	11 mg/d (neurotoxicity)
Molybdenum	Severe neurologic abnormalities	Reproductive and fetal abnormalities	2 mg/d extrapolated from animal

			data
Selenium	Cardiomyopathy, heart failure, striated muscle degeneration	General: Alopecia, nausea, vomiting, abnormal nails, emotional lability, peripheral neuropathy, lassitude, garlic odor to breath, dermatitis Occupational: Lung and nasal carcinomas, liver necrosis, pulmonary inflammation	400 μg/d (hair, nail changes)
Phosphorus	Rickets (osteomalacia), proximal muscle weakness, rhabdomyolysis, paresthesia, ataxia, seizure, confusion, heart failure, hemolysis, acidosis	Hyperphosphatemia	4000 mg/d
Zinc	Growth retardation, ↓ taste and smell, alopecia, dermatitis, diarrhea, immune dysfunction, failure to thrive, gonadal atrophy, congenital malformations	General: Reduced copper absorption, gastritis, sweating, fever, nausea, vomiting Occupational: Respiratory distress, pulmonary fibrosis	40 mg/d (impaired copper metabolism)

Abbreviations: GI, gastrointestinal; ND, not determined.

Figure 74-1

Vitamin	Active derivative or cofactor form	Principal function
Thiamine (B_1) 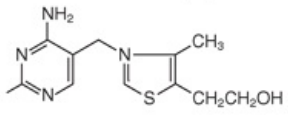	Thiamine pyrophosphate	Coenzyme for cleavage of carbon-carbon bonds; amino acid and carbohydrate metabolism
Riboflavin (B_2) 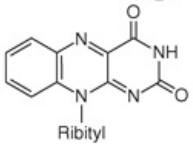	Flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD)	Cofactor for oxidation, reduction reactions, and covalently attached prosthetic groups for some enzymes
Niacin 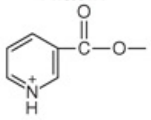	Nicotinamide adenine dinucleotide phosphate (NADP) and nicotinamide adenine dinucleotide (NAD)	Coenzymes for oxidation and reduction reactions
Vitamin B_6 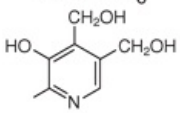	Pyridoxal phosphate	Cofactor for enzymes of amino acid metabolism
Folate 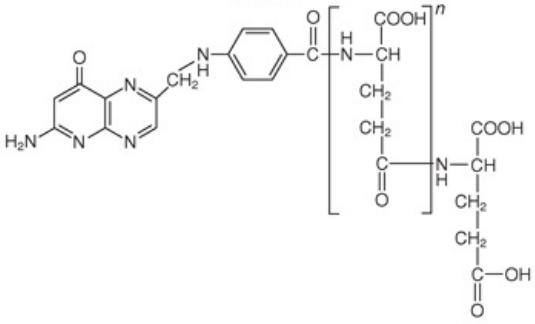	Polyglutamate forms of (5, 6, 7, 8) tetrahydrofolate with carbon unit attachments	Coenzyme for one carbon transfer in nucleic acid and amino acid metabolism
Vitamin B_{12} 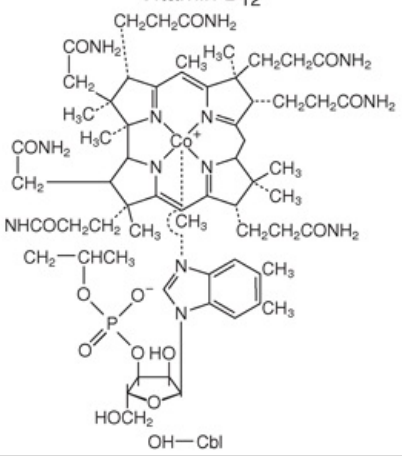	Methylcobalamine Adenosylcobalamin	Coenzyme for methionine synthase and L-methylmalonyl-CoA mutase

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vitamin	Active derivative or cofactor form	Principal function
Vitamin C 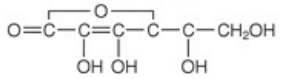	Ascorbic acid and dehydroascorbic acid	Participation as a redox ion in many biologic oxidation and hydrogen transfer reactions
Vitamin A 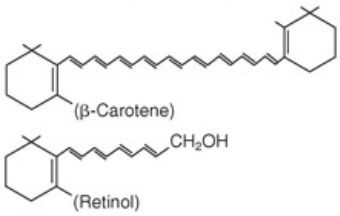	Retinol, retinaldehyde, and retinoic acid	Formation of rhodopsin (vision) and glycoproteins (epithelial cell function); also regulates gene transcription
Vitamin D 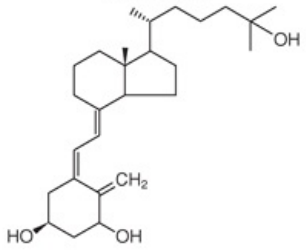	1,25-Dihydroxyvitamin D	Maintenance of blood calcium and phosphorus levels; antiproliferative hormone
Vitamin E 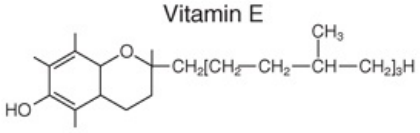	Tocopherols and tocotrienols	Antioxidants
Vitamin K 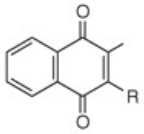	Vitamin K hydroquinone	Cofactor for posttranslation carboxylation of many proteins including essential clotting factors

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The structures and principal functionsof vitamins associated with human disorders.

VITAMINS

Thiamine (Vitamin B_1)

Thiamine was the first B vitamin to be identified and therefore is referred to as vitamin B1. Thiamine functions in the decarboxylation of α-ketoacids, such as pyruvate α-ketoglutarate, and branched-chain amino acids and thus is essential for energy generation. In addition, thiamine pyrophosphate acts as a coenzyme for a transketolase reaction that mediates the conversion of hexose and pentose phosphates. It has been postulated that thiamine plays a role in peripheral nerve conduction, although the exact chemical reactions underlying this function are not known.

FOOD SOURCES

The median intake of thiamine in the United States from food alone is 2 mg/d. Primary food sources for thiamine include yeast, organ meat, pork, legumes, beef, whole grains, and nuts. Milled rice and grains contain little thiamine, if any. Thiamine deficiency is therefore more common in cultures that rely heavily on a rice-based diet. Tea, coffee (regular and decaffeinated), raw fish, and shellfish contain thiaminases, which can destroy the vitamin. Thus, drinking large amounts of tea or coffee can theoretically lower thiamine body stores.

DEFICIENCY

Most dietary deficiency of thiamine worldwide is the result of poor dietary intake. In Western countries, the primary causes of thiamine deficiency are alcoholism and chronic illnesses such as cancer. Alcohol

interferes directly with the absorption of thiamine and with the synthesis of thiamine pyrophosphate. Thiamine should always be replenished when a patient with alcoholism is being refed, as carbohydrate repletion without adequate thiamine can precipitate acute thiamine deficiency with lactic acidosis. Other at-risk populations are women with prolonged hyperemesis gravidarum and anorexia, patients with overall poor nutritional status on parenteral glucose, patients after bariatric bypass surgery, and patients on chronic diuretic therapy due to increased urinary thiamine losses. Maternal thiamine deficiency can lead to infantile beriberi in breast-fed children. Thiamine deficiency should be considered in the setting of motor vehicle accidents associated with head injury.

Thiamine deficiency in its early stage induces anorexia and nonspecific symptoms (e.g., irritability, decrease in short-term memory). Prolonged thiamine deficiency causes beriberi, which is classically categorized as wet or dry, although there is considerable overlap. In either form of beriberi, patients may complain of pain and paresthesia. *Wet beriberi* presents primarily with cardiovascular symptoms, due to impaired myocardial energy metabolism and dysautonomia, and can occur after 3 months of a thiamine-deficient diet. Patients present with an enlarged heart, tachycardia, high-output congestive heart failure, peripheral edema, and peripheral neuritis. Patients with *dry beriberi* present with a symmetric peripheral neuropathy of the motor and sensory systems with diminished reflexes. The neuropathy affects the legs most markedly, and these patients have difficulty rising from a squatting position.

Alcoholic patients with chronic thiamine deficiency also may have central nervous system (CNS) manifestations known as *Wernicke's encephalopathy*, consisting of horizontal nystagmus, ophthalmoplegia (due to weakness of one or more extraocular muscles), cerebellar ataxia, and mental impairment (Chap. 392). When there is an additional loss of memory and a confabulatory psychosis, the syndrome is known as *Wernicke-Korsakoff syndrome*. Despite the typical clinical picture and history, Wernicke-Korsakoff syndrome is underdiagnosed.

The laboratory diagnosis of thiamine deficiency usually is made by a functional enzymatic assay of transketolase activity measured before and after the addition of thiamine pyrophosphate. A >25% stimulation by the addition of thiamine pyrophosphate (an activity coefficient of 1.25) is interpreted as abnormal. Thiamine or the phosphorylated esters of thiamine in serum or blood also can be measured by high-performance liquid chromatography (HPLC) to detect deficiency.

TREATMENT: THIAMINE DEFICIENCY

In acute thiamine deficiency with either cardiovascular or neurologic signs, 100 mg/d of thiamine should be given parenterally for 7 days, followed by 10 mg/d orally until there is complete recovery. Cardiovascular and ophthalmoplegic improvement occurs within 24 h. Other manifestations gradually clear, although psychosis in Wernicke-Korsakoff syndrome may be permanent or persist for several months.

TOXICITY

Although anaphylaxis has been reported after high doses of thiamine, no adverse effects have been recorded from either food or supplements at high doses. Thiamine supplements may be bought over the counter in doses of up to 50 mg/d.

Riboflavin (Vitamin B_2)

Riboflavin is important for the metabolism of fat, carbohydrate, and protein, reflecting its role as a respiratory coenzyme and an electron donor. Enzymes that contain flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) as prosthetic groups are known as *flavoenzymes* (e.g., succinic acid dehydrogenase, monoamine oxidase, glutathione reductase). FAD is a cofactor for methyltetrahydrofolate reductase and therefore modulates homocysteine metabolism. The vitamin also plays a role in drug and steroid metabolism, including detoxification reactions.

Although much is known about the chemical and enzymatic reactions of riboflavin, the clinical manifestations of riboflavin deficiency are nonspecific and are similar to those of other deficiences of B vitamins. Riboflavin deficiency is manifested principally by lesions of the mucocutaneous surfaces of the mouth and skin (Table 74-1). In addition to the mucocutaneous lesions, corneal vascularization, anemia, and personality changes have been described with riboflavin deficiency.

DEFICIENCY AND EXCESS

Riboflavin deficiency almost always is due to dietary deficiency. Milk, other dairy products, and enriched breads and cereals are the most important dietary sources of riboflavin in the United States, although lean meat, fish, eggs, broccoli, and legumes are also good sources. Riboflavin is extremely sensitive to light, and milk should be stored in containers that protect against photodegradation. Laboratory diagnosis of riboflavin deficiency can be made by measurement of red blood cell or urinary riboflavin concentrations or by measurement of erythrocyte glutathione reductase activity, with and without added FAD. Because the capacity of the gastrointestinal tract to absorb riboflavin is limited (~ 20 mg if given in one oral dose), riboflavin toxicity has not been described.

Niacin (Vitamin B_3)

The term *niacin* refers to nicotinic acid and nicotinamide and their biologically active derivatives. Nicotinic acid and nicotinamide serve as precursors of two coenzymes, nicotinamide adenine dinucleotide (NAD) and NAD phosphate (NADP), which are important in numerous oxidation and reduction reactions in the body. In addition, NAD and NADP are active in adenine diphosphate–ribose transfer reactions involved in DNA repair and calcium mobilization.

METABOLISM AND REQUIREMENTS

Nicotinic acid and nicotinamide are absorbed well from the stomach and small intestine. Niacin bioavailability is high from beans, milk, meat, and eggs; bioavailability from cereal grains is lower. Since flour is enriched with the "free" niacin (i.e., non-coenzyme form), bioavailability is excellent. Median intakes of niacin in the United States considerably exceed the recommended dietary allowance (RDA).

The amino acid tryptophan can be converted to niacin with an efficiency of 60:1 by weight. Thus, the RDA for niacin is expressed in niacin equivalents. A lower conversion of tryptophan to niacin occurs in vitamin B6 and/or riboflavin deficiencies and in the presence of isoniazid. The urinary excretion products of niacin include 2-pyridone and 2-methyl nicotinamide, measurements of which are used in the diagnosis of niacin deficiency.

DEFICIENCY

Niacin deficiency causes *pellagra*, which is found mostly among people eating corn-based diets in parts of China, Africa, and India. Pellagra in North America is found mainly among alcoholics; in patients with congenital defects of intestinal and kidney absorption of tryptophan (Hartnup disease; Chap. 364); and in patients with carcinoid syndrome (Chap. 350), in which there is increased conversion of tryptophan to serotonin. In the setting of famine or population displacement, the occurrence of pellagra results from the absolute lack of niacin but also from the deficiency of micronutrients required for the conversion of tryptophan to niacin (e.g., iron, riboflavin, and pyridoxine). The early symptoms of pellagra include loss of appetite, generalized weakness and irritability, abdominal pain, and vomiting. Bright red glossitis then ensues, followed by a characteristic skin rash that is pigmented and scaling, particularly in skin areas exposed to sunlight. This rash is known as *Casal's necklace* because it forms a ring around the neck; it is seen in advanced cases. Vaginitis and esophagitis also may occur. Diarrhea (in part due to proctitis and in part due to malabsorption), depression, seizures, and dementia are also part of the pellagra syndrome—the four *D*s: *d*ermatitis, *d*iarrhea, and *d*ementia leading to *d*eath.

TREATMENT PELLAGRA

Treatment of pellagra consists of oral supplementation of 100–200 mg of nicotinamide or nicotinic acid three times daily for 5 days. High doses of nicotinic acid (2 g/d in a time-release form) are used for the treatment of elevated cholesterol and triglyceride levels and/or a low high-density lipoprotein (HDL) cholesterol level (Chap. 356).

TOXICITY

Prostaglandin-mediated flushing due to binding of the vitamin to a G protein–coupled receptor has been observed at daily doses as low as 50 mg of niacin when taken as a supplement or as therapy for dyslipidemia. There is no evidence of toxicity from niacin derived from food sources. Flushing always starts in the face and may be accompanied by skin dryness, itching, paresthesia, and headache. Pharmaceutical preparations of nicotinic acid combined with laropiprant, a selective prostaglandin D_2 receptor 1 antagonist, or premedication with aspirin may alleviate these symptoms. Flushing is subject to tachyphylaxis and often improves with time. Nausea, vomiting, and abdominal pain also occur at similar doses of niacin. Hepatic toxicity is the most serious toxic reaction caused by niacin and may present as jaundice with elevated aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels. A few cases of fulminant hepatitis requiring liver transplantation have been reported at doses of 3–9 g/d. Other toxic reactions include glucose intolerance, hyperuricemia, macular edema, and macular cysts. The combination of nicotinic acid preparations for dyslipidemia with 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors may increase the risk of rhabdomyolysis. The upper limit for daily niacin intake has been set at 35 mg. However, this upper limit does not pertain to the therapeutic use of niacin.

Pyridoxine (Vitamin B_6)

Vitamin B6 refers to a family of compounds that include pyridoxine, pyridoxal, pyridoxamine, and their 5'-phosphate derivatives. 5'-Pyridoxal phosphate (PLP) is a cofactor for more than 100 enzymes involved in amino acid metabolism. Vitamin B_6 also is involved in heme and neurotransmitter synthesis and in the metabolism of glycogen, lipids, steroids, sphingoid bases, and several vitamins, including the conversion of tryptophan to niacin.

DIETARY SOURCES

Plants contain vitamin B6 in the form of pyridoxine, whereas animal tissues contain PLP and pyridoxamine phosphate. The vitamin B_6 contained in plants is less bioavailable than that in animal

tissues. Rich food sources of vitamin B_6 include legumes, nuts, wheat bran, and meat, although it is present in all food groups.

DEFICIENCY

Symptoms of vitamin B6 deficiency include epithelial changes, as seen frequently with other B vitamin deficiencies. In addition, severe vitamin B_6 deficiency can lead to peripheral neuropathy, abnormal electroencephalograms, and personality changes that include depression and confusion. In infants, diarrhea, seizures, and anemia have been reported. Microcytic hypochromic anemia is due to diminished hemoglobin synthesis, since the first enzyme involved in heme biosynthesis (aminolevulinate synthase) requires PLP as a cofactor (Chap. 103). In some case reports, platelet dysfunction has been reported. Since vitamin B_6 is necessary for the conversion of homocysteine to cystathionine, it is possible that chronic low-grade vitamin B_6 deficiency may result in hyperhomocysteinemia and increased risk of cardiovascular disease (Chaps. 241 and 364). Independent of homocysteine, low levels of circulating vitamin B_6 have been associated with inflammation and elevated levels of C-reactive protein.

Certain medications, such as isoniazid, l-dopa, penicillamine, and cycloserine, interact with PLP due to a reaction with carbonyl groups. Pyridoxine should be given concurrently with isoniazid to avoid neuropathy. The increased ratio of AST to ALT seen in alcoholic liver disease reflects the relative vitamin B6 dependence of ALT. Vitamin B_6 dependency syndromes that require pharmacologic doses of vitamin B_6 are rare; they include cystathionine β-synthase deficiency, pyridoxine-responsive (primarily sideroblastic) anemias, and gyrate atrophy with chorioretinal degeneration due to decreased activity of the mitochondrial enzymeornithine aminotransferase. In these situations, 100–200 mg/d of oral vitamin B_6 is required for treatment.

High doses of vitamin B6 have been used to treat carpal tunnel syndrome, premenstrual syndrome, schizophrenia, autism, and diabetic neuropathy but have not been found to be effective.

The laboratory diagnosis of vitamin B6 deficiency is generally made on the basis of low plasma PLP values (<20 nmol/L). Treatment of vitamin B_6 deficiency is done with 50 mg/d; higher doses of 100–200 mg/d are given if the deficiency is related to medication use. Vitamin B_6 should not be given with l-dopa, since the vitamin interferes with the action of this drug.

TOXICITY

The safe upper limit for vitamin B6 has been set at 100 mg/d, although no adverse effects have been associated with high intakes of vitamin B_6 from food sources only. When toxicity occurs, it causes a severe sensory neuropathy, leaving patients unable to walk. Some cases of photosensitivity and dermatitis have been reported.

Folate, Vitamin B_{12}

See Chap. 105.

Vitamin C

Both ascorbic acid and its oxidized product dehydroascorbic acid are biologically active. Actions of vitamin C include antioxidant activity, promotion of nonheme iron absorption, carnitine biosynthesis, the conversion of dopamine to norepinephrine, and the synthesis of many peptide hormones. Vitamin C is also important for connective tissue metabolism and cross-linking (proline hydroxylation), and it is a component of many drug-metabolizing enzyme systems, particularly the mixed-function oxidase systems.

ABSORPTION AND DIETARY SOURCES

Almost complete absorption of vitamin C occurs if <100 mg is administered in a single dose; however, only 50% or less is absorbed at doses >1 g. Enhanced degradation and fecal and urinary excretion of vitamin C occur at higher intake levels.

Good dietary sources of vitamin C include citrus fruits, green vegetables (especially broccoli), tomatoes, and potatoes. Consumption of five servings of fruits and vegetables a day provides vitamin C in excess of the RDA of 90 mg/d for males and 75 mg/d for females. In addition, approximately 40% of the U.S. population consumes vitamin C as a dietary supplement in which "natural forms" of the vitamin are no more bioavailable than synthetic forms. Smoking, hemodialysis, pregnancy, and stress (e.g., infection, trauma) appear to increase vitamin C requirements.

DEFICIENCY

Vitamin C deficiency causes scurvy. In the United States, this is seen primarily among the poor and elderly, in alcoholics who consume <10 mg/d of vitamin C, and in individuals consuming macrobiotic diets. Vitamin C deficiency also can occur in young adults who eat severely unbalanced diets. In addition to generalized fatigue, symptoms of scurvy primarily reflect impaired formation of mature connective tissue and include bleeding into skin (petechiae, ecchymoses, perifollicular hemorrhages);

inflamed and bleeding gums; and manifestations of bleeding into joints, the peritoneal cavity, the pericardium, and the adrenal glands. In children, vitamin C deficiency may cause impaired bone growth. Laboratory diagnosis of vitamin C deficiency is made on the basis of low plasma or leukocyte levels.

Administration of vitamin C (200 mg/d) improves the symptoms of scurvy within a matter of several days. High-dose vitamin C supplementation (e.g., 1–2 g/d) may slightly decrease the symptoms and duration of upper respiratory tract infections. Vitamin C supplementation has also been reported to be useful in Chédiak-Higashi syndrome (Chap. 60) and osteogenesis imperfecta (Chap. 363). Diets high in vitamin C have been claimed to lower the incidence of certain cancers, particularly esophageal and gastric cancers. If proved, this effect may be due to the fact that vitamin C can prevent the conversion of nitrites and secondary amines to carcinogenic nitrosamines. However, an intervention study from China did not show vitamin C to be protective. Parenteral ascorbic acid has been suggested to have a potential therapeutic role in the treatment of advanced cancers.

TOXICITY

Taking >2 g of vitamin C in a single dose may result in abdominal pain, diarrhea, and nausea. Since vitamin C may be metabolized to oxalate, it is feared that chronic high-dose vitamin C supplementation could result in an increased prevalence of kidney stones. However, this has not been borne out in several trials, except in patients with preexisting renal disease. Thus, it is reasonable to advise patients with a past history of kidney stones not to take large doses of vitamin C. There is also an unproven but possible risk that chronic high doses of vitamin C could promote iron overload in patients taking supplemental iron. High doses of vitamin C can induce hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency, and doses >1 g/d can cause false-negative guaiac reactions as well as interfere with tests for urinary glucose. High doses may interfere with certain drugs (e.g., bortezomib in myeloma patients).

Biotin

Biotin is a water-soluble vitamin that plays a role in gene expression, gluconeogenesis, and fatty acid synthesis and serves as a CO_2 carrier on the surface of both cytosolic and mitochondrial carboxylase enzymes. The vitamin also functions in the catabolism of specific amino acids (e.g., leucine). Excellent food sources of biotin include organ meat such as liver or kidney, soy, beans, yeast, and egg yolks; however, egg white contains the protein avidin, which strongly binds the vitamin and reduces its bioavailability.

Biotin deficiency due to low dietary intake is rare; rather, deficiency is due to inborn errors of metabolism. Biotin deficiency has been induced by experimental feeding of egg white diets and in patients with short bowels who received biotin-free parenteral nutrition. In adults, biotin deficiency results in mental changes (depression, hallucinations), paresthesia, anorexia, and nausea. A scaling, seborrheic, and erythematous rash may occur around the eyes, nose, and mouth as well as on the extremities. In infants, biotin deficiency presents as hypotonia, lethargy, and apathy. In addition, infants may develop alopecia and a characteristic rash that includes the ears. The laboratory diagnosis of biotin deficiency can be established on the basis of a decreased urinary concentration or an increased urinary excretion of 3-hydroxyisovaleric acid after a leucine challenge. Treatment requires pharmacologic doses of biotin, using up to 10 mg/d. No toxicity is known.

Pantothenic Acid (Vitamin B_5)

Pantothenic acid is a component of coenzyme A and phosphopantetheine, which are involved in fatty acid metabolism and the synthesis of cholesterol, steroid hormones, and all compounds formed from isoprenoid units. In addition, pantothenic acid is involved in the acetylation of proteins. The vitamin is excreted in the urine, and the laboratory diagnosis of deficiency is made on the basis of low urinary vitamin levels.

The vitamin is ubiquitous in the food supply. Liver, yeast, egg yolks, whole grains, and vegetables are particularly good sources. Human pantothenic acid deficiency has been demonstrated only in experimental feeding of diets low in pantothenic acid or by giving a specific pantothenic acid antagonist. The symptoms of pantothenic acid deficiency are nonspecific and include gastrointestinal disturbance, depression, muscle cramps, paresthesia, ataxia, and hypoglycemia. Pantothenic acid deficiency is believed to have caused the burning feet syndrome seen in prisoners of war during World War II. No toxicity of this vitamin has been reported.

Choline

Choline is a precursor for acetylcholine, phospholipids, and betaine. Choline is necessary for the structural integrity of cell membranes, cholinergic neurotransmission, lipid and cholesterol metabolism, methyl-group metabolism, and transmembrane signaling. Recently, a recommended adequate intake was set at 550 mg/d for adult males and 425 mg/d for adult females, although certain genetic polymorphisms can increase an individual's requirement. Choline is thought to be a "conditionally essential" nutrient in that de novo synthesis occurs in the liver and is less than the vitamin's utilization only under certain stress conditions (e.g., alcoholic liver disease). The dietary requirement of choline

depends on the status of other methyl-group donors (folate, vitamin B12, and methionine) and thus varies widely. Choline is widely distributed in food (e.g., egg yolk, wheat germ, organ meat, milk) in the form of lecithin (phosphatidylcholine). Choline deficiency has occurred in patients receiving parenteral nutrition devoid of choline. Deficiency results in fatty liver, elevated transaminase levels, and skeletal muscle damage with high creatine phosphokinase values. The diagnosis of choline deficiency is currently made on the basis of low plasma levels, although nonspecific conditions (e.g., heavy exercise) may suppress plasma levels.

Toxicity from choline results in hypotension, cholinergic sweating, diarrhea, salivation, and a fishy body odor. The upper limit for choline has been set at 3.5 g/d. Therapeutically, choline has been suggested for patients with dementia and patients at high risk of cardiovascular disease, due to its ability to lower cholesterol and homocysteine levels. However, such benefits have not been firmly documented. Choline- and betaine-restricted diets are of therapeutic value in trimethylaminuria (fish odor syndrome).

Flavonoids

Flavonoids constitute a large family of polyphenols that contribute to the aroma, taste, and color of fruits and vegetables. Major groups of dietary flavonoids include anthocyanidins in berries; catechins in green tea and chocolate; flavonols (e.g., quercitin) in broccoli, kale, leeks, onion, and the skins of grapes and apples; and isoflavones (e.g., genistein) in legumes. Isoflavones have a low bioavailability and are partially metabolized by the intestinal flora. The dietary intake of flavonoids is estimated to be between 10 and 100 mg/d, although this is almost certainly an underestimate due to the lack of knowledge of their concentrations in many foods. Several flavonoids have been shown to have antioxidant activity and to affect cell signaling. From observational epidemiologic studies and limited clinical human and animal studies, flavonoids have been postulated to play a role in the prevention of several chronic diseases, including neurodegenerative disease, diabetes, and osteoporosis. The ultimate importance and usefulness of their compounds against human disease have not been demonstrated.

Vitamin A

Vitamin A, in the strictest sense, refers to retinol. However, the oxidized metabolites, retinaldehyde and retinoic acid, are also biologically active compounds. The term *retinoids* includes all molecules (including synthetic molecules) that are chemically related to retinol. Retinaldehyde (11-*cis*) is the essential form of vitamin A that is required for normal vision, whereas retinoic acid is necessary for normal morphogenesis, growth, and cell differentiation. Retinoic acid does not function in vision and, in contrast to retinol, is not involved in reproduction. Vitamin A also plays a role in iron utilization, humoral immunity, T cell–mediated immunity, natural killer cell activity, and phagocytosis. Vitamin A is commercially available in esterified forms (e.g., acetate, palmitate) since it is more stable as an ester.

There are more than 600 carotenoids in nature, and approximately 50 of them can be metabolized to vitamin A. β-Carotene is the most prevalent carotenoid in the food supply that has provitamin A activity. In humans, significant fractions of carotenoids are absorbed intact and are stored in liver and fat. It is now estimated that 12 μg or more (range, 4–27 μg) of dietary all-trans β-carotene is equivalent to 1 μg of retinol activity, whereas 24 μg or more of other dietary provitamin A carotenoids (e.g., cryptoxanthin, α-carotene) is equivalent to 1 μg of retinol activity. The vitamin A equivalency for a β-carotene supplement in an oily solution is 2:1.

METABOLISM

The liver contains approximately 90% of the vitamin A reserves and secretes vitamin A in the form of retinol, which is bound to retinol-binding protein. Once this has occurred, the retinol-binding protein complex interacts with a second protein, transthyretin. This trimolecular complex functions to prevent vitamin A from being filtered by the kidney glomerulus to protect the body against the toxicity of retinol and to allow retinol to be taken up by specific cell-surface receptors that recognize retinol-binding protein. A certain amount of vitamin A enters peripheral cells even if it is not bound to retinol-binding protein. After retinol is internalized by the cell, it becomes bound to a series of cellular retinol-binding proteins, which function as sequestering and transporting agents as well as co-ligands for enzymatic reactions. Certain cells also contain retinoic acid–binding proteins, which have sequestering functions but also shuttle retinoic acid to the nucleus and enable its metabolism.

Retinoic acid is a ligand for certain nuclear receptors that act as transcription factors. Two families of receptors (RAR and RXR receptors) are active in retinoid-mediated gene transcription. Retinoid receptors regulate transcription by binding as dimeric complexes to specific DNA sites, the retinoic acid response elements, in target genes (Chap. 338). The receptors can either stimulate or repress gene expression in response to their ligands. RAR binds all-*trans* retinoic acid and 9-*cis*-retinoic acid, whereas RXR binds only 9-*cis*-retinoic acid.

The retinoid receptors play an important role in controlling cell proliferation and differentiation. Retinoic acid is useful in the treatment of promyelocytic leukemia (Chap. 109) and also is used in the treatment of cystic acne because it inhibits keratinization, decreases sebum secretion, and possibly alters the inflammatory reaction (Chap. 52). RXRs dimerize with other nuclear receptors to function as coregulators of genes responsive to retinoids, thyroid hormone, and calcitriol. RXR agonists induce

insulin sensitivity experimentally, perhaps because RXR is a cofactor for the peroxisome-proliferator-activated receptors (PPARs), which are targets for thiazolidinedione drugs such as rosiglitazone and troglitazone (Chap. 344).

DIETARY SOURCES

The retinol activity equivalent (RAE) is used to express the vitamin A value of food. One RAE is defined as 1 μg of retinol (0.003491 mmol), 12 μg of β-carotene, and 24 μg of other provitamin A carotenoids. In older literature, vitamin A often was expressed in international units (IU), with 1 μg of retinol being equal to 3.33 IU of retinol and 20 IU of β-carotene, but these units are no longer in scientific use.

Liver, fish, and eggs are excellent food sources for preformed vitamin A; vegetable sources of provitamin A carotenoids include dark green and deeply colored fruits and vegetables. Moderate cooking of vegetables enhances carotenoid release for uptake in the gut. Carotenoid absorption is also aided by some fat in a meal. Infants are particularly susceptible to vitamin A deficiency because neither breast nor cow's milk supplies enough vitamin A to prevent deficiency. In developing countries, chronic dietary deficiency is the main cause of vitamin A deficiency and is exacerbated by infection. In early childhood, low vitamin A status results from inadequate intakes of animal food sources and edible oils, both of which are expensive, coupled with seasonal unavailability of vegetables and fruits and lack of marketed fortified food products. Concurrent zinc deficiency can interfere with the mobilization of vitamin A from liver stores. Alcohol interferes with the conversion of retinol to retinaldehyde in the eye by competing for alcohol (retinol) dehydrogenase. Drugs that interfere with the absorption of vitamin A include mineral oil, neomycin, and cholestyramine.

DEFICIENCY

Vitamin A deficiency is endemic in areas where diets are chronically poor, especially in southern Asia, sub-Saharan Africa, some parts of Latin America, and the western Pacific, including parts of China. Vitamin A status is usually assessed by measuring serum retinol [normal range, 1.05–3.50 μmol/L (30–100 μg/dL)] or blood spot retinol or by tests of dark adaptation. There are stable isotopic or invasive liver biopsy methods to estimate total body stores of vitamin A. Based on deficient serum retinol [<0.70 μmol/L (20 μg/dL)], there are >90 million preschool-age children with vitamin A deficiency, among whom >4 million have an ocular manifestation of deficiency termed *xerophthalmia*. This condition includes milder stages of night blindness and conjunctival xerosis (dryness) with Bitot's spots (white patches of keratinizedepithelium appearing on the sclera) as well as rare, potentially blinding corneal ulceration and necrosis. Keratomalacia (softening of the cornea) leads to corneal scarring that blinds at least a quarter of a million children each year and is associated with a fatality rate of 4–25%. However, vitamin A deficiency at any stage poses an increased risk of mortality from diarrhea, dysentery, measles, malaria, and respiratory disease. Vitamin A deficiency can compromise barrier and innate and acquired immune defenses to infection. Vitamin A supplementation can markedly reduce risk of child mortality (23–34%, on average) in areas where deficiency is widely prevalent. About 10% of pregnant women in undernourished settings also develop night blindness, assessed by history, during the latter half of pregnancy, and this moderate vitamin A deficiency is associated with an increased risk of maternal infection and mortality rate.

TREATMENT VITAMIN A DEFICIENCY

Any stage of xerophthalmia should be treated with 60 mg of vitamin A in oily solution, usually contained in a soft-gel capsule. The same dose is repeated 1 and 14 days later. Doses should be reduced by half for patients 6–11 months of age. Mothers with night blindness or Bitot's spots should be given vitamin A orally, either 3 mg daily or 7.5 mg twice a week for 3 months. These regimens are efficacious, and they are less expensive and more widely available than injectable water-miscible vitamin A. A common approach to prevention is to supplement young children in high-risk areas with 60 mg every 4–6 months, with a half dose given to infants 6–11 months of age.

Uncomplicated vitamin A deficiency rarely occurs in industrialized countries. One high-risk group, extremely low-birth-weight infants (<1000 g), is likely to be vitamin A–deficient and should be supplemented with 1500 μg (or RAE) of vitamin A three times a week for 4 weeks. Severe measles in any society can lead to secondary vitamin A deficiency. Children hospitalized with measles should receive two 60-mg doses of vitamin A on two consecutive days. Vitamin A deficiency most often occurs in patients with malabsorptive diseases (e.g., celiac sprue, short-bowel syndrome) who have abnormal dark adaptation or symptoms of night blindness without other ocular changes. Typically, such patients are treated for 1 month with 15 mg/d of a water-miscible preparation of vitamin A. This is followed by a lower maintenance dose, with the exact amount determined by monitoring serum retinol.

There are no specific deficiency signs or symptoms that result from carotenoid deficiency. It was postulated that β-carotene would be an effective chemopreventive agent for cancer because numerous epidemiologic studies had shown that diets high in β-carotene were associated with lower incidences of cancers of the respiratory and digestive systems. However, intervention studies in smokers found that treatment with high doses of β-carotene actually resulted in more lung cancers than did treatment with placebo. Non–provitamin A carotenoids such as lutein and zeaxanthin have been suggested to

protect against macular degeneration, and large-scale intervention studies have been undertaken to test this hypothesis. The non–provitamin A carotenoid lycopene has been proposed to protect against prostate cancer. However, the effectiveness of these agents has not been proved by intervention studies, and the mechanisms underlying these purported biologic actions are unknown.

Selective plant breeding techniques that lead to a higher provitamin A content of staple foods may improve vitamin A malnutrition in low-income countries. Moreover, a recently developed genetically modified food (Golden Rice) showed an improved β-carotene to vitamin A conversion ratio of ~3:1.

TOXICITY

Acute toxicity of vitamin A was first noted in Arctic explorers who ate polar bear liver and has also been seen after administration of 150 mg in adults or 100 mg in children. Acute toxicity is manifested by increased intracranial pressure, vertigo, diplopia, bulging fontanels in children, seizures, and exfoliative dermatitis; it may result in death. In children being treated for vitamin A deficiency according to the protocols outlined above, transient bulging of fontanels occurs in 2% of infants, and transient nausea, vomiting, and headache occur in 5% of preschoolers. Chronic vitamin A intoxication is largely a concern in industrialized countries and has been seen in normal adults who ingest 15 mg/d and children who ingest 6 mg/d of vitamin A over a period of several months. Manifestations include dry skin, cheilosis, glossitis, vomiting, alopecia, bone demineralization and pain, hypercalcemia, lymph node enlargement, hyperlipidemia, amenorrhea, and features of pseudotumor cerebri with increased intracranial pressure and papilledema. Liver fibrosis with portal hypertension and bone demineralization may result from chronic vitamin A intoxication. When vitamin A is provided in excess to pregnant women, congenital malformations have included spontaneous abortions, craniofacial abnormalities, and valvular heart disease. In pregnancy, the daily dose of vitamin A should not exceed 3 mg. Commercially available retinoid derivatives are also toxic, including 13-*cis*-retinoic acid, which has been associated with birth defects. As a result, contraception should be continued for a least 1 year and possibly longer in women who have taken 13-*cis*-retinoic acid.

In malnourished children, vitamin A supplements (100,000–200,000 IU) as a function of age in several rounds over 2 years are considered to amplify nonspecific effects of vaccines. However, for unclear reasons, there may be a negative effect on mortality rates in incompletely vaccinated girls.

High doses of carotenoids do not result in toxic symptoms but should be avoided in smokers due to an increased risk of lung cancer. Very high doses of β-carotene (~200 mg/d) have been used to treat or prevent the skin rashes of erythropoietic protoporphyria. Carotenemia, which is characterized by a yellowing of the skin (creases of the palms and soles) but not the sclerae, may be present after ingestion of >30 mg of β-carotene daily. Hypothyroid patients are particularly susceptible to the development of carotenemia due to impaired breakdown of carotene to vitamin A. Reduction of carotenes from the diet results in the disappearance of skin yellowing and carotenemia over a period of 30–60 days.

Vitamin D

The metabolism of the fat-soluble vitamin D is described in detail in Chap. 352. The biologic effects of this vitamin are mediated by vitamin D receptors, which are found in most tissues, thus potentially expanding vitamin D actions on nearly all cell systems and organs (e.g., immune cells, brain, breast, colon, and prostate) as well as exerting classic endocrine effects on calcium metabolism and bone health. Vitamin D is thought to be important for maintaining normal function of many nonskeletal tissues such as muscle (including heart muscle), immune function, and inflammation as well as cell proliferation and differentiation. Studies have shown that it may be useful as adjunctive treatment for tuberculosis, psoriasis, and multiple sclerosis or for the prevention of certain cancers. Vitamin D insufficiency may increase the risk of Type 1 diabetes mellitus, cardiovascular disease (insulin resistance, hypertension, or low-grade inflammation), or brain dysfunction (e.g., depression). However, the importance of the exact physiologic role of vitamin D in these non skeletal diseases has not been clarified.

A major source of vitamin D is its synthesis in the skin upon ultraviolet B (UV-B) (wavelength, 290–315 nm) exposure. Except for fish, food (unless fortified) contains only limited amounts of vitamin D. Vitamin D2 (ergocalciferol) is obtained from plant sources and is the chemical form found in some supplements.

DEFICIENCY

Vitamin D status has been assessed by measuring serum 1,25-dihydroxyvitamin D [1,25$(OH)_2$ vitamin D] levels; however, there is no consensus on a uniform assay methodology or on optimal serum levels. The optimal level might, in fact, differ according to the targeted disease entity. Based on epidemiologic and experimental data, a 1,25$(OH)_2$ vitamin D level >20 ng/mL (≥50 nmol/L; to convert ng/mL to nmol/L, multiply by 2.496) is sufficient for good bone health. Some experts advocate higher serum levels (e.g., >30ng/mL) for other desirable endpoints of vitamin D action.

Risk factors for vitamin D deficiency are old age, lack of sun exposure, dark skin (especially among those living in northern latitudes), fat malabsorption, and obesity. Rickets represents the classic

disease of vitamin D deficiency. Signs of deficiency are muscle soreness, weakness, and bone pain. Some of these effects are independent of calcium intake.

The U.S. National Academy of Science recently concluded that the majority of North Americans are receiving adequate amounts of vitamin D (RDA = 15 μg/d or 600 IU/d; Chap. 73). However, for people older than 70 years, the RDA is set at 20 μg/d (800 IU/d). The consumption of fortified or enriched foods as well as suberythemal sun exposure should be encouraged for people at risk for vitamin D deficiency. If an adequate intake cannot be achieved, vitamin D supplements should be taken, especially during the winter months. Vitamin D deficiency can be treated by the oral administration of 50,000 IU/week for 6–8 weeks followed by a maintenance dose of 800 IU/d (100 μg/d) from food and supplements after achievement of normal plasma levels. The physiologic effects of vitamin D2 and D3 are identical when ingested over long periods.

TOXICITY

The upper limit of intake has been set at 4000 IU/d. Contrary to earlier beliefs, acute vitamin D intoxication is rare and usually is caused by the uncontrolled and excessive ingestion of supplements or faulty food fortification practices. High plasma $1,25(OH)_2$ vitamin D and high plasma calcium levels are central features of toxicity. Stopping vitamin D and calcium supplements is mandatory, and treatment of hypercalcemia may be required.

Vitamin E

Vitamin E is a collective name for all stereoisomers of tocopherols and tocotrienols, although only the RR tocopherols meet human requirements. Vitamin E acts as a chain-breaking antioxidant and is an efficient pyroxyl radical scavenger that protects low-density lipoproteins (LDLs) and polyunsaturated fats in membranes from oxidation. A network of other antioxidants (e.g., vitamin C, glutathione) and enzymes maintains vitamin E in a reduced state. Vitamin E also inhibits prostaglandin synthesis and the activities of protein kinase C and phospholipase A_2.

ABSORPTION AND METABOLISM

After absorption, vitamin E is taken up from chylomicrons by the liver, and a hepatic α-tocopherol transport protein mediates intracellular vitamin E transport and incorporation into very low density lipoprotein (VLDL). The transport protein has particular affinity for the RRR isomeric form of α-tocopherol; thus, this natural isomer has the most biologic activity.

REQUIREMENT

Vitamin E is widely distributed in the food supply and is particularly high in sunflower oil, safflower oil, and wheat germ oil; γ-tocotrienols are notably present in soybean and corn oils. Vitamin E is also found in meats, nuts, and cereal grains, and small amounts are present in fruits and vegetables. Vitamin E pills containing doses of 50–1000 mg are ingested by about 10% of the U.S. population. The RDA for vitamin E is 15 mg/d (34.9 μmol or 22.5 IU) for all adults. Diets high in polyunsaturated fats may necessitate a slightly higher intake of vitamin E.

Dietary deficiency of vitamin E does not exist. Vitamin E deficiency is seen in only severe and prolonged malabsorptive diseases, such as celiac disease, or after small-intestinal resection. Children with cystic fibrosis or prolonged cholestasis may develop vitamin E deficiency characterized by areflexia and hemolytic anemia. Children with abetalipoproteinemia cannot absorb or transport vitamin E and become deficient quite rapidly. A familial form of isolated vitamin E deficiency also exists; it is due to a defect in the α-tocopherol transport protein. Vitamin E deficiency causes axonal degeneration of the large myelinated axons and results in posterior column and spinocerebellar symptoms. Peripheral neuropathy is initially characterized by areflexia, with progression to an ataxic gait, and by decreased vibration and position sensations. Ophthalmoplegia, skeletal myopathy, and pigmented retinopathy may also be features of vitamin E deficiency. Either vitamin E or selenium deficiency in the host has been shown to increase certain viral mutations and, therefore, virulence. The laboratory diagnosis of vitamin E deficiency is made on the basis of low blood levels of α-tocopherol (<5 μg/mL, or <0.8 mg of α-tocopherol per gram of total lipids).

TREATMENT: VITAMIN E DEFICIENCY

Symptomatic vitamin E deficiency should be treated with 800–1200 mg of α-tocopherol per day. Patients with abetalipoproteinemia may need as much as 5000–7000 mg/d. Children with symptomatic vitamin E deficiency should be treated with 400 mg/d orally of water-miscible esters; alternatively, 2 mg/kg per d may be administered intramuscularly. Vitamin E in high doses may protect against oxygen-induced retrolental fibroplasia and bronchopulmonary dysplasia as well as intraventricular hemorrhage of prematurity. Vitamin E has been suggested to increase sexual performance, treat intermittent claudication, and slow the aging process, but evidence for these properties is lacking. When given in combination with other antioxidants, vitamin E may help prevent macular degeneration. High doses (60–800 mg/d) of vitamin E have been shown in controlled trials to improve parameters of immune function and reduce colds in nursing home residents, but intervention studies using vitamin E to prevent cardiovascular disease or cancer have not shown efficacy, and at doses >400 mg/d, vitamin E may even increase all-cause mortality rates.

TOXICITY

All forms of vitamin E are absorbed and could contribute to toxicity. *High* doses of vitamin E (>800 mg/d) may reduce platelet aggregation and interfere with vitamin K metabolism and are therefore contraindicated in patients taking warfarin and antiplatelet agents (such as aspirin or clopidogrel). Nausea, flatulence, and diarrhea have been reported at doses >1 g/d.

Vitamin K

There are two natural forms of vitamin K: vitamin K1, also known as *phylloquinone*, from vegetable and animal sources, and vitamin K_2, or *menaquinone*, which is synthesized by bacterial flora and found in hepatic tissue. Phylloquinone can be converted to menaquinone in some organs.

Vitamin K is required for the posttranslational carboxylation of glutamic acid, which is necessary for calcium binding to γ-carboxylated proteins such as prothrombin (factor II); factors VII, IX, and X; protein C; protein S; and proteins found in bone (osteocalcin) and vascular smooth muscle (e.g., matrix Gla protein). However, the importance of vitamin K for bone mineralization and prevention of vascular calcification is not known. Warfarin-type drugs inhibit γ-carboxylation by preventing the conversion of vitamin K to its active hydroquinone form.

DIETARY SOURCES

Vitamin K is found in green leafy vegetables such as kale and spinach, and appreciable amounts are also present in margarine and liver. Vitamin K is present in vegetable oils; olive, canola, and soybean oils are particularly rich sources. The average daily intake by Americans is estimated to be approximately 100 μg/d.

DEFICIENCY

The symptoms of vitamin K deficiency are due to hemorrhage, and newborns are particularly susceptible because of low fat stores, low breast milk levels of vitamin K, sterility of the infantile intestinal tract, liver immaturity, and poor placental transport. Intracranial bleeding, as well as gastrointestinal and skin bleeding, can occur in vitamin K–deficient infants 1–7 days after birth. Thus, vitamin K (1 mg IM) is given prophylactically at the time of delivery.

Vitamin K deficiency in adults may be seen in patients with chronic small-intestinal disease (e.g., celiac disease, Crohn's disease), in those with obstructed biliary tracts, or after small-bowel resection. Broad-spectrum antibiotic treatment can precipitate vitamin K deficiency by reducing gut bacteria, which synthesize menaquinones, and by inhibiting the metabolism of vitamin K. In patients with warfarin therapy, the antiobesity drug orlistat can lead to international normalized ratio (INR) changes due to vitamin K malabsorption. The diagnosis of vitamin K deficiency usually is made on the basis of an elevated prothrombin time or reduced clotting factors, although vitamin K may also be measured directly by HPLC. Vitamin K deficiency is treated by using a parenteral dose of 10 mg. For patients with chronic malabsorption, 1–2 mg/d of vitamin K should be given orally, or 1–2 mg/week can be taken parenterally. Patients with liver disease may have an elevated prothrombin time because of liver cell destruction as well as vitamin K deficiency. If an elevated prothrombin time does not improve on vitamin K therapy, it can be deduced that it is not the result of vitamin K deficiency.

TOXICITY

Toxicity from dietary phylloquinones and menaquinones has not been described. High doses of vitamin K can impair the actions of oral anticoagulants.

MINERALS

Table 74-2.

Calcium

See Chap. 352.

Zinc

Zinc is an integral component of many metalloenzymes in the body; it is involved in the synthesis and stabilization of proteins, DNA, and RNA and plays a structural role in ribosomes and membranes. Zinc is necessary for the binding of steroid hormone receptors and several other transcription factors to DNA. Zinc is absolutely required for normal spermatogenesis, fetal growth, and embryonic development.

ABSORPTION

The absorption of zinc from the diet is inhibited by dietary phytate, fiber, oxalate, iron, and copper, as well as by certain drugs, including penicillamine, sodium valproate, and ethambutol. Meat, shellfish, nuts, and legumes are good sources of bioavailable zinc, whereas zinc in grains and legumes is less available for absorption.

DEFICIENCY

Mild zinc deficiency has been described in many diseases, including diabetes mellitus, HIV/AIDS, cirrhosis, alcoholism, inflammatory bowel disease, malabsorption syndromes, and sickle cell disease. In these diseases, mild chronic zinc deficiency can cause stunted growth in children, decreased taste sensation (hypogeusia), and impaired immune function. Severe chronic zinc deficiency has been described as a cause of hypogonadism and dwarfism in several Middle Eastern countries. In these children, hypopigmented hair is also part of the syndrome. Acrodermatitis enteropathica is a rare autosomal recessive disorder characterized by abnormalities in zinc absorption. Clinical manifestations include diarrhea, alopecia, muscle wasting, depression, irritability, and a rash involving the extremities, face, and perineum. The rash is characterized by vesicular and pustular crusting with scaling and erythema. Occasional patients with Wilson's disease have developed zinc deficiency as a consequence of penicillamine therapy (Chap. 360).

The diagnosis of zinc deficiency is usually made by a serum zinc level <12 μmol/L (<70 μg/dL). Pregnancy and birth control pills may cause a slight depression in serum zinc levels, and hypoalbuminemia from any cause can result in hypozincemia. In acute stress situations, zinc may be redistributed from serum into tissues. Zinc deficiency may be treated with 60 mg elemental zinc, orally twice a day. Zinc gluconate lozenges (13 mg elemental zinc every 2 h while awake) have been reported to reduce the duration and symptoms of the common cold in adults, but studies are conflicting.

Zinc deficiency is prevalent in many developing countries and usually coexists with other micronutrient deficiencies (especially iron). Zinc (20 mg/d) may be an effective adjunctive therapeutic strategy for diarrheal disease and pneumonia in children.

TOXICITY

Acute zinc toxicity after oral ingestion causes nausea, vomiting, and fever. Zinc fumes from welding may also be toxic and cause fever, respiratory distress, excessive salivation, sweating, and headache. Chronic large doses of zinc may depress immune function and cause hypochromic anemia as a result of copper deficiency. Intranasal zinc preparations should be avoided because they may lead to irreversible damage of nasal mucosa and anosmia.

Copper

Copper is an integral part of numerous enzyme systemsz, including amine oxidases, ferroxidase (ceruloplasmin), cytochrome-*c* oxidase, superoxide dismutase, and dopamine hydroxylase. Copper is also a component of ferroprotein, a transport protein involved in the basolateral transfer of iron during absorption from the enterocyte. As such, copper plays a role in iron metabolism, melanin synthesis, energy production, neurotransmitter synthesis, and CNS function; the synthesis and cross-linking of elastin and collagen; and the scavenging of superoxide radicals. Dietary sources of copper include shellfish, liver, nuts, legumes, bran, and organ meats.

DEFICIENCY

Dietary copper deficiency is relatively rare, although it has been described in premature infants who are fed milk diets and in infants with malabsorption (Table 74-2). Copper-deficiency anemia has been reported in patients with malabsorptive diseases and nephrotic syndrome and in patients treated for Wilson's disease with chronic high doses of oral zinc, which can interfere with copper absorption. Menkes' kinky hair syndrome is an X-linked metabolic disturbance of copper metabolism characterized by mental retardation, hypocupremia, and decreased circulating ceruloplasmin (Chap. 363). It is caused by mutations in the copper-transporting *ATP7A* gene. Children with this disease often die within 5 years because of dissecting aneurysms or cardiac rupture. Aceruloplasminemia is a rare autosomal recessive disease characterized by tissue iron overload, mental deterioration, microcytic anemia, and low serum iron and copper concentrations.

The diagnosis of copper deficiency is usually made on the basis of low serum levels of copper (<65 μg/dL) and low ceruloplasmin levels (<20 mg/dL). Serum levels of copper may be elevated in pregnancy or stress conditions since ceruloplasmin is an acute-phase reactant and 90% of circulating copper is bound to ceruloplasmin.

TOXICITY

Copper toxicity is usually accidental (Table 74-2). In severe cases, kidney failure, liver failure, and coma may ensue. In Wilson's disease, mutations in the copper-transporting *ATP7B* gene lead to accumulation of copper in the liver and brain, with low blood levels due to decreased ceruloplasmin (Chap. 360).

Selenium

Selenium, in the form of selenocysteine, is a component of the enzymeglutathione peroxidase, which serves to protect proteins, cell membranes, lipids, and nucleic acids from oxidant molecules. As such, selenium is being actively studied as a chemopreventive agent against certain cancers, such as prostate cancer. Selenocysteine is also found in the deiodinase enzymes, which mediate the deiodination of thyroxine to triiodothyronine (Chap. 341). Rich dietary sources of selenium include

seafood, muscle meat, and cereals, although the selenium content of cereal is determined by the soil concentration. Countries with low soil concentrations include parts of Scandinavia, China, and New Zealand. *Keshan disease* is an endemic cardiomyopathy found in children and young women residing in regions of China where dietary intake of selenium is low (<20 μg/d). Concomitant deficiencies of iodine and selenium may worsen the clinical manifestations of cretinism. Chronic ingestion of high amounts of selenium leads to selenosis, characterized by hair and nail brittleness and loss, garlic breath odor, skin rash, myopathy, irritability, and other abnormalities of the nervous system.

Chromium

Chromium potentiates the action of insulin in patients with impaired glucose tolerance, presumably by increasing insulin receptor–mediated signaling, although its usefulness in treating Type 2 diabetes is uncertain. In addition, improvement in blood lipid profiles has been reported in some patients. The usefulness of chromium supplements in muscle building has not been substantiated. Rich food sources of chromium include yeast, meat, and grain products. Chromium in the trivalent state is found in supplements and is largely nontoxic; however, chromium-6 is a product of stainless steel welding and is a known pulmonary carcinogen as well as a cause of liver, kidney, and CNS damage.

Magnesium

See Chap. 352.

Fluoride, Manganese, and Ultratrace Elements

An essential function for fluoride in humans has not been described, although it is useful for the maintenance of structure in teeth and bone. Adult fluorosis results in mottled and pitted defects in tooth enamel as well as brittle bone (skeletal fluorosis).

Manganese and molybdenum deficiencies have been reported in patients with rare genetic abnormalities and in a few patients receiving prolonged total parenteral nutrition. Several manganese-specific enzymes have been identified (e.g., manganese superoxide dismutase). Deficiencies of manganese have been reported to result in bone demineralization, poor growth, ataxia, disturbances in carbohydrate and lipid metabolism, and convulsions.

Ultratrace elements are defined as those needed in amounts <1 mg/d. Essentiality has not been established for most ultratrace elements, although selenium, chromium, and iodine are clearly essential (Chap. 341). *Molybdenum* is necessary for the activity of sulfite and xanthine oxidase, and molybdenum deficiency may result in skeletal and brain lesions.

FURTHER READINGS

Aasheim ET et al: Vitamin status after bariatric surgery: A randomized study of gastric bypass and duodenal switch. Am J Clin Nutr 90:15, 2009[PMID: 19439456] [Full Text]

Booth SL: Roles for vitamin K beyond coagulation. Annu Rev Nutr 29:89, 2009[PMID: 19400704] [Full Text]

Holick MF: Deficiency of sunlight and vitamin D. BMJ 336:1318, 2008[PMID: 18556276] [Full Text]

Lichtenstein AH, Russell RM: Essential nutrients: Food or supplements? Where should the emphasis be? JAMA 294:351, 2005[PMID: 16030280] [Full Text]

Neuhouser ML et al. Multivitamin use and risk of cancer and cardiovascular disease in the Women's Health Initiative Cohorts. Arch Intern Med 169:294, 2009[PMID: 19204221] [Full Text]

Penniston KL, Tanumihardjo SA: The acute and chronic toxic effects of vitamin A. Am J Clin Nutr 83:191, 2006[PMID: 16469975] [Full Text]

Ross AC et al (eds): *Dietary Reference Intakes for Calcium and Vitamin D* . Washington, DC, The National Academies Press, 2011

Tang G et al: Golden Rice is an effective source of vitamin A. Am J Clin Nutr 89:1776, 2009[PMID: 19369372] [Full Text]

Zeisel SH, Da Costa KA: Choline: An essential nutrient for public health. Nutr Rev 67:615, 2009[PMID: 19906248] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MALNUTRITION AND NUTRITIONAL ASSESSMENT: INTRODUCTION

Malnutrition can arise from primary or secondary causes, with the former resulting from inadequate or poor-quality food intake and the latter from diseases that alter food intake or nutrient requirements, metabolism, or absorption. Primary malnutrition occurs mainly in developing countries and under conditions of political unrest, war, or famine. Secondary malnutrition, the main form encountered in industrialized countries, was largely unrecognized until the early 1970s, when it was appreciated that persons with adequate food supplies can become malnourished as a result of acute or chronic diseases that alter nutrient intake or metabolism, particularly diseases that cause acute or chronic inflammation. Various studies have shown that protein-energy malnutrition (PEM) affects one-third to one-half of patients on general medical and surgical wards in teaching hospitals. The consistent finding that nutritional status influences patient prognosis underscores the importance of preventing, detecting, and treating malnutrition.

PROTEIN-ENERGY MALNUTRITION

Definitions for forms of PEM are in flux. Traditionally, the two major types of PEM have been *marasmus* and *kwashiorkor*. These conditions are compared in **Table 75-1**. Marasmus has been considered the end result of a long-term deficit of dietary energy, whereas kwashiorkor has been understood to result from a protein-poor diet. Although the former concept remains essentially correct, evidence is accumulating that PEM syndromes are distinguished by two main features: dietary intake and underlying inflammatory processes. Energy-poor diets with minimal inflammation cause gradual erosion of body mass, resulting in classic marasmus. By contrast, inflammation from acute illnesses such as injury or sepsis and chronic illnesses such as cancer, lung or heart disease, and HIV can erode lean body mass even in the presence of relatively sufficient dietary intake, leading to a kwashiorkor-like state. Quite often, inflammatory illnesses impair appetite and dietary intake, producing combinations of the two.

Table 75-1 Comparison of Marasmus/Cachexia and Kwashiorkor/Protein-Calorie Malnutrition

	Marasmus or Cachexia	Kwashiorkor or Protein-Calorie Malnutri-tion[a]
Clinical setting	↓ Energy intake	↓ Protein intake during stress state
Time course to develop	Months or years	Weeks
Clinical features	Starved appearance	Well-nourished appear-ance
	Weight <80% standard for height	Easy hair pluckabil-ity[b]
	Triceps skinfold <3 mm	Edema
	Midarm muscle circumference <15 cm	
Laboratory findings	Creatinine-height index <60% standard	Se-rum albumin <2.8 g/dL
		Total iron-binding capacity <200 μg/dL
		Lymphocytes <1500/μL
		Anergy
Clinical course	Reasonably preserved responsiveness to short-term stress	Infections
		Poor wound healing, decubitus ulcers, skin break-down
Mortality	Low unless related to underlying disease	High
Diagnostic criteria	Triceps skinfold <3 mm	Serum albu-min <2.8 g/dL
	Midarm muscle circumference <15 cm	At least one of the following:
		Poor wound healing, decubitus ulcers, or skin breakdown

		Easy hair pluckabil-ity[b]
		Edema

[a]The findings used to diagnose kwashiorkor must be unexplained by other causes.

[b]Tested by *firmly* pulling a lock of hair from the top (not the sides or back), grasping with the thumb and forefinger. An average of three or more hairs removed easily and painlessly is considered abnormal hair pluckability.

An international consensus committee has proposed the following revised definitions. *Starvation-related malnutrition* is suggested for instances of chronic starvation without inflammation, *chronic disease–related malnutrition* when inflammation is chronic and of mild to moderate degree, and *acute disease– or injury-related malnutrition* when inflammation is acute and of a severe degree. However, because distinguishing diagnostic criteria for these conditions have not been elaborated, this chapter outlines criteria that have served well and are embedded in the medical literature.

Marasmus or Cachexia

Marasmus is a state in which virtually all available body fat stores have been exhausted due to starvation. Cachexia is a state that involves substantial loss of lean body mass due to chronic systemic inflammation. Conditions that produce cachexia in high-income countries tend to be chronic and indolent, such as cancer and chronic pulmonary disease, whereas marasmus occurs in patients with anorexia nervosa. These conditions are relatively easy to detect because of the patient's starved appearance. The diagnosis is based on fat and muscle wastage resulting from prolonged calorie deficiency and/or inflammation. Diminished skinfold thickness reflects the loss of fat reserves; reduced arm muscle circumference with temporal and interosseous muscle wasting reflects the catabolism of protein throughout the body, including vital organs such as the heart, liver, and kidneys.

Routine laboratory findings in cachexia/marasmus are relatively unremarkable. The creatinine-height index (24-h urinary creatinine excretion compared with normal values based on height) is low, reflecting the loss of muscle mass. Occasionally, the serum albumin level is reduced, but it stays above 2.8 g/dL in uncomplicated cases. Despite a morbid appearance, immunocompetence, wound healing, and the ability to handle short-term stress are reasonably well preserved in most patients.

Pure starvation-related malnutrition is a chronic, fairly well adapted form of starvation rather than an acute illness; it should be treated cautiously in an attempt to reverse the downward trend gradually. Although nutritional support is necessary, overly aggressive repletion can result in severe, even life-threatening metabolic imbalances such as hypophosphatemia and cardiorespiratory failure (refeeding syndrome). When possible, oral or enteral nutritional support is preferred; treatment started slowly allows readaptation of metabolic and intestinal functions (Chap. 76).

Kwashiorkor or Protein-Calorie Malnutrition (PCM)

By contrast, kwashiorkor or PCM in developed countries occurs mainly in connection with acute, life-threatening illnesses such as trauma and sepsis. The physiologic stress produced by these illnesses increases protein and energy requirements at a time when intake is often limited. A classic scenario for PCM is an acutely stressed patient who receives only 5% dextrose solutions for periods as brief as 2 weeks; this gives rise to the proposed term *acute disease– or injury-related malnutrition* . Although the etiologic mechanisms are not fully known, the protein-sparing response normally seen in starvation is blocked by the stressed state and by carbohydrate infusion.

In its early stages, the physical findings of kwashiorkor/PCM are few and subtle. Fat reserves and muscle mass are initially unaffected, giving the deceptive appearance of adequate nutrition. Signs that support the diagnosis of kwashiorkor/PCM include easy hair pluckability, edema, skin breakdown, and poor wound healing. The major sine qua non is severe reduction of levels of serum proteins such as albumin (<2.8 g/dL) and transferrin (<150 mg/dL) or iron-binding capacity (<200 μg/dL). Cellular immune function is depressed, reflected by lymphopenia (<1500 lymphocytes/μL in adults and older children) and lack of response to skin test antigens (anergy).

The prognosis of adult patients with full-blown kwashiorkor/ PCM is not good even with aggressive nutritional support. Surgical wounds often dehisce (fail to heal), pressure sores develop, gastroparesis and diarrhea can occur with enteral feeding, the risk of gastrointestinal bleeding from stress ulcers is increased, host defenses are compromised, and death from overwhelming infection may occur despite antibiotic therapy. Unlike treatment in marasmus, aggressive nutritional support is indicated to restore better metabolic balance rapidly (Chap. 76). Although kwashiorkor in children is less foreboding, perhaps because a lesser degree of stress is required to precipitate the disorder, it is still a serious condition.

Physiologic Characteristics of Hypometabolic and Hypermetabolic States

The metabolic characteristics and nutritional needs of hypermetabolic patients who are stressed from injury, infection, or chronic inflammatory illness differ from those of hypometabolic patients who are

unstressed but chronically starved. In both cases, nutritional support is important, but misjudgments in selecting the appropriate approach may have serious adverse consequences.

The hypometabolic patient is typified by the relatively less stressed but mildly catabolic and chronically starved individual who, with time, will develop cachexia/marasmus. The hypermetabolic patient stressed from injury or infection is catabolic (experiencing rapid breakdown of body mass) and is at high risk for developing PCM/kwashiorkor if nutritional needs are not met and/or the illness does not resolve quickly. As summarized in **Table 75-2**, the two states are distinguished by differing perturbations of metabolic rate, rates of protein breakdown (proteolysis), and rates of gluconeogenesis. These differences are mediated by proinflammatory cytokines and counterregulatory hormones—tumor necrosis factor, interleukins 1 and 6, C-reactive protein, catecholamines (epinephrine and norepinephrine), glucagon, and cortisol—that are relatively reduced in hypometabolic patients and increased in hypermetabolic patients. Although insulin levels are also elevated in stressed patients, insulin resistance in the target tissues blocks insulin-mediated anabolic effects.

Table 75-2 Physiologic Characteristics of Hypometabolic and Hypermetabolic States

Physiologic Characteristics	Hypometabolic, Nonstressed Patient (Marasmic)	Hypermetabolic, Stressed Patient (Kwashiorkor Risk*)
Cytokines, catecholamines, glucagon, cortisol, insulin	↓	↑
Metabolic rate, O_2 consumption	↓	↑
Proteolysis, gluconeogenesis Ureagenesis, urea excretion	↓ ↓	↑ ↑
Fat catabolism, fatty acid utilization	Relative ↑	Absolute ↑
Adaptation to starvation	Normal	Abnormal

*These changes characterize the stressed, kwashiorkor-risk patient seen in developed countries; they differ in some respects from the characteristics of primary kwashiorkor seen in developing countries.

Metabolic Rate

In starvation and semistarvation, the resting metabolic rate falls between 10 and 30% as an adaptive response to energy restriction, slowing the rate of weight loss. By contrast, resting metabolic rate rises in the presence of physiologic stress in proportion to the degree of the insult. It may increase by about 10% after elective surgery, 20–30% after bone fractures, 30–60% with severe infections such as peritonitis or gram-negative septicemia, and as much as 110% after major burns.

If the metabolic rate (energy requirement) is not matched by energy intake, weight loss results, slowly in hypometabolism and quickly in hypermetabolism. Losses of up to 10% of body mass are unlikely to be detrimental; however, losses greater than this in acutely ill hypermetabolic patients may be associated with rapid deterioration in body function.

Protein Catabolism

The rate of endogenous protein breakdown (catabolism) to supply energy needs normally falls during uncomplicated energy deprivation. After about 10 days of total starvation, an unstressed individual loses about 12–18 g/d protein (equivalent to approximately 2 oz of muscle tissue or 2–3 g of nitrogen). By contrast, in injury and sepsis, protein breakdown accelerates in proportion to the degree of stress, reaching 30–60 g/d after elective surgery, 60–90 g/d with infection, 100–130 g/d with severe sepsis or skeletal trauma, and >175 g/d with major burns or head injuries. These losses are reflected by proportional increases in the excretion of urea nitrogen, the major by-product of protein breakdown.

Gluconeogenesis

The major aim of protein catabolism during a state of starvation is to provide the glucogenic amino acids (especially alanine and glutamine) that serve as substrates for endogenous glucose production (gluconeogenesis) in the liver. In the hypometabolic/starved state, protein breakdown for gluconeogenesis is minimized, especially as ketones derived from fatty acids become the substrate preferred by certain tissues. In the hypermetabolic/stress state, gluconeogenesis increases dramatically and in proportion to the degree of the insult to increase the supply of glucose (the major fuel of reparation). Glucose is the only fuel that can be utilized by hypoxemic tissues (anaerobic glycolysis), white blood cells, and newly generated fibroblasts. Infusions of glucose partially offset a negative energy balance but do not significantly suppress the high rates of gluconeogenesis in catabolic patients. Hence, adequate supplies of protein are needed to replace the amino acids utilized

for this metabolic response.

In summary, a hypometabolic patient is adapted to starvation and conserves body mass by reducing the metabolic rate and using fat as the primary fuel (rather than glucose and its precursor amino acids). A hypermetabolic patient also uses fat as a fuel but rapidly breaks down body protein to produce glucose, causing loss of muscle and organ tissue and endangering vital body functions.

MICRONUTRIENT MALNUTRITION

The same illnesses and reductions in nutrient intake that lead to PEM often produce deficiencies of vitamins and minerals as well (Chap. 74). Deficiencies of nutrients that are stored in small amounts (such as the water-soluble vitamins) are lost through external secretions, such as zinc in diarrhea fluid or burn exudate, and are probably more common than generally recognized.

Deficiencies of vitamin C, folic acid, and zinc are reasonably common in sick patients. Signs of scurvy such as corkscrew hairs on the lower extremities are found frequently in chronically ill and/or alcoholic patients. The diagnosis can be confirmed with plasma vitamin C levels. Folic acid intakes and blood levels are often less than optimal, even among healthy persons; when illness, alcoholism, poverty, or poor dentition is present, these deficiencies are common. Low blood zinc levels are prevalent in patients with malabsorption syndromes such as inflammatory bowel disease. Patients with zinc deficiency often exhibit poor wound healing, pressure ulcer formation, and impaired immunity. Thiamine deficiency is a common complication of alcoholism but may be prevented by therapeutic doses of thiamine in patients treated for alcohol abuse.

Patients with low plasma vitamin C levels usually respond to the doses in multivitamin preparations, but patients with deficiencies should be supplemented with 250–500 mg/d. Folic acid is absent from some oral multivitamin preparations; patients with deficiencies should be supplemented with about 1 mg/d. Patients with zinc deficiencies resulting from large external losses sometimes require oral supplementation with 220 mg of zinc sulfate one to three times daily. For these reasons, laboratory assessments of the micronutrient status of patients at high risk are desirable.

Hypophosphatemia develops in hospitalized patients with remarkable frequency and generally results from rapid intracellular shifts of phosphate in cachectic or alcoholic patients receiving intravenous glucose (Chap. 45). The adverse clinical sequelae are numerous; some, such as acute cardiopulmonary failure, are collectively called refeeding syndrome and can be life-threatening.

GLOBAL CONSIDERATIONS

Many developing countries are still faced with high prevalences of the classic forms of PEM: marasmus and kwashiorkor. *Food insecurity*, which characterizes many poor countries, prevents consistent dietary sufficiency and/or quality and leads to endemic or cyclic malnutrition. Factors threatening food security include marked seasonal variations in agricultural productivity (rainy season –dry season cycles), periodic droughts, political unrest or injustice, and disease epidemics, especially HIV/AIDS. The coexistence of malnutrition and disease epidemics exacerbates the latter and increases complications and mortality rates, creating vicious cycles of malnutrition and disease.

As economic prosperity improves, developing countries have been observed to undergo an epidemiologic transition, a component of which has been termed the *nutrition transition*. As improved economic resources make greater dietary diversity possible, middle-income populations (e.g., southern Asia, China, and Latin America) typically begin to adopt lifestyle habits of industrialized nations, with increased consumption of energy and fat and decreased levels of physical activity. This leads to rising levels of obesity, metabolic syndrome, diabetes, cardiovascular disease, and cancer, sometimes coexisting in populations with persistent undernutrition.

Micronutrient deficiencies also remain prevalent in many countries of the world, impairing functional status and productivity and increasing mortality rates. Vitamin A deficiency affects perhaps 20% of the world's population, impairing vision and increasing morbidity and mortality rates from infections, for example, measles. Community vitamin A supplementation programs have significantly reduced measles mortality rates in vulnerable populations. Mild to moderate iron deficiency may be prevalent in up to 50% of the world, resulting from poor dietary diversity coupled with periodic blood loss and pregnancies. Iodine deficiency remains prevalent in about 35% of the world's population, causing goiter, hypothyroidism, and cretinism. Zinc deficiency is endemic in many populations, producing growth retardation, hypogonadism, and dermatoses, and impairing wound healing.

NUTRITIONAL ASSESSMENT

Because interactions between illness and nutrition are complex, many physical and laboratory findings reflect both underlying disease and nutritional status. Therefore, the nutritional evaluation of a patient requires an integration of the history, physical examination, anthropometrics, and laboratory studies. This approach helps both to detect nutritional problems and to prevent concluding that isolated findings indicate nutritional problems when they do not. For example, hypoalbuminemia caused by an underlying illness does not necessarily indicate malnutrition.

Nutritional History

A nutritional history is directed toward identifying underlying mechanisms that put patients at risk for nutritional depletion or excess. These mechanisms include inadequate intake, impaired absorption, decreased utilization, increased losses, and increased requirements of nutrients.

Individuals with the characteristics listed in **Table 75-3** are at particular risk for nutritional deficiencies.

Table 75-3 The High-Risk Patient

Underweight (body mass index <18.5) and/or recent loss of ≥10% of usual body mass
Poor intake: anorexia, food avoidance (e.g., psychiatric condition), or NPO status for more than about 5 days
Protracted nutrient losses: malabsorption, enteric fistulas, draining abscesses or wounds, renal dialysis
Hypermetabolic states: sepsis, protracted fever, extensive trauma or burns
Alcohol abuse or use of drugs with antinutrient or catabolic properties: steroids, antimetabolites (e.g., methotrexate), immunosuppressants, antitumor agents
Impoverishment, isolation, advanced age

Physical Examination

Physical findings that suggest vitamin, mineral, and protein-energy deficiencies and excesses are outlined in **Table 75-4**. Most of the physical findings are not specific for individual nutrient deficiencies and must be integrated with the historic, anthropometric, and laboratory findings. For example, the finding of follicular hyperkeratosis on the back of the arms is a fairly common, normal finding. However, if it is widespread in a person who consumes little fruit and vegetables and smokes regularly (increasing ascorbic acid requirements), vitamin C deficiency is likely. Similarly, easily pluckable hair may be a consequence of chemotherapy, but in a hospitalized patient who has poorly healing surgical wounds and hypoalbuminemia, it suggests PCM/kwashiorkor.

Table 75-4 Physical Findings of Nutritional Deficiencies

Clinical Findings	Possible Deficiency*	Possible Excess
Hair, Nails		
Corkscrew hairs and unemerged coiled hairs	Vitamin C	
Easily pluckable hair	Protein	
Flag sign (transverse depigmentation of hair)	Protein	
Sparse hair	Protein, biotin, zinc	Vitamin A
Transverse ridging of nails	Protein	
Skin		
Cellophane appearance	Protein	
Cracking (flaky paint or crazy pavementdermatosis)	Protein	
Follicular hyperkeratosis	Vitamins A, C	
Petechiae (especially perifollicular)	Vitamin C	
Purpura	Vitamins C, K	
Pigmentation, scaling of sun-exposed areas	Niacin	
Poor wound healing, decubitus ulcers	Protein, vitamin C, zinc	
Scaling	Vitamin A, essential fatty acids, biotin Zinc (hyperpigmented)	Vitamin A Carotene

Yellow pigmentation sparing sclerae (benign)		
Eyes		
Night blindness	Vitamin A	
Papilledema		Vitamin A
Perioral		
Angular stomatitis	Riboflavin, pyridoxine, niacin	
Cheilosis (dry, cracking, ulcerated lips)	Riboflavin, pyridoxine, niacin	
Oral		
Atrophic lingual papillae (slick tongue)	Riboflavin, niacin, folate, vitamin B12, protein, iron	
Glossitis (scarlet, raw tongue)	Riboflavin, niacin, pyridoxine, folate, vitamin B12	
Hypogeusesthesia, hyposmia	Zinc	
Swollen, retracted, bleeding gums (if teeth present)	Vitamin C	
Bones, Joints		
Beading of ribs, epiphyseal swelling, bowlegs	Vitamin D	
Tenderness, subperiosteal hemorrhage in children	Vitamin C	
Neurologic		
Confabulation, disorientation	Thiamine (Korsakoff's psychosis)	
Drowsiness, lethargy, vomiting		Vitamin A
Dementia	Niacin, vitamin B12, folate	
Headache		Vitamin A
Ophthalmoplegia	Thiamine, phosphorus	
Peripheral neuropathy (e.g., weakness, paresthesias, ataxia, footdrop, and decreased tendon reflexes, fine tactile sense, vibratory sense, and position sense)	Thiamine, pyridoxine, vitamin B12	Pyridoxine
Tetany	Calcium, magnesium	
Other		
Edema	Protein, thiamine	
Heart failure	Thiamine ("wet" beriberi), phosphorus	
Hepatomegaly	Protein	Vitamin A
Parotid enlargement	Protein (consider also bulimia)	
Sudden heart failure, death	Vitamin C	

*In this table, "protein deficiency" is used to signify kwashiokor/PCM.

Anthropometrics

Anthropometric measurements provide information on body muscle mass and fat reserves. The most

practical and commonly used measurements are body weight, height, triceps skinfold (TSF), and midarm muscle circumference (MAMC). Body weight is one of the most useful nutritional parameters to follow in patients who are acutely or chronically ill. Unintentional weight loss during illness often reflects loss of lean body mass (muscle and organ tissue), especially if it is rapid and is not caused by diuresis. This can be an ominous sign since it indicates use of vital body protein stores for metabolic fuel. The reference standard for normal body weight, body mass index (BMI: weight in kilograms divided by height, in meters, squared), is discussed in Chap 78. BMIs <18.5 are considered underweight, 18.5–24.9 are normal, 25–29.9 are overweight, and ≥30 are obese.

Measurement of skinfold thickness is useful for estimating body fat stores, because about 50% of body fat is normally in the subcutaneous region. Skinfold thickness can also permit discrimination of fat mass from muscle mass. The TSF is a convenient site that is generally representative of the body's overall fat level. A thickness <3 mm suggests virtually complete exhaustion of fat stores. The MAMC can be used to estimate skeletal muscle mass, calculated as follows:

MAMC (cm) = upper arm circumference (cm) – [0.314 x TSF (mm)]

Laboratory Studies

A number of laboratory tests used routinely in clinical medicine can yield valuable information about a patient's nutritional status if a slightly different approach to their interpretation is used. For example, abnormally low serum albumin levels, total iron-binding capacity, and anergy may have a distinct explanation, but collectively they may represent kwashiorkor. In the clinical setting of a hypermetabolic, acutely ill patient who is edematous and has easily pluckable hair and inadequate protein intake, the diagnosis of PCM/kwashiorkor is clear-cut. Commonly used laboratory tests for assessing nutritional status are outlined in **Table 75-5**. The table also provides tips to avoid assigning nutritional significance to tests that may be abnormal for nonnutritional reasons.

Table 75-5 Laboratory Tests for Nutritional Assessment

Test (Normal Values)	Nutritional Use	Causes of Normal Value Despite Mal-nutrition	Other Causes of Abnormal Value
Serum albumin (3.5–5.5 g/dL)	2.8–3.5: Compromised protein status	Dehydration	**Low**
	<2.8: Possible kwashiorkor	Infusion of albumin, fresh-frozen plasma, or whole blood	Common:
	Increasing value reflects positive protein bal-ance		
			Infection and other stress, especially with poor protein intake
			Burns, trauma
			Congestive heart failure Fluid overload
			Severe liver disease
			Uncommon:
			Nephrotic syndrome
			Zinc deficiency
			Bacterial stasis/overgrowth of small intestine
Serum prealbumin, also called transthyretin (20–40 mg/dL; lower in prepubertal chil-dren)	10–15 mg/dL: Mild protein depletion	Chronic renal failure	Similar to serum albumin
	5–10 mg/dL: Moderate protein deple-tion		
	<5 mg/dL: Severe protein depletion		

	Increasing value reflects positive protein bal-ance		
Serum total iron-binding capacity (TIBC) 240 –450 μg/dL	<200: Compromised protein status, possi-ble kwashiorkor	Iron deficiency	**Low**
			Similar to serum albumin
	Increasing value reflects positive protein bal-ance		**High**
	More labile than albumin		Iron deficiency
Prothrombin time 12.0–15.5 s	Prolongation: vitamin K deficiency		**Prolonged**
			Anticoagulant therapy (war-farin)
			Severe liver disease
Serum creatinine 0.6–1.6 mg/dL	<0.6: Muscle wasting due to prolonged energy deficit Reflects muscle mass		**High** Despite muscle wasting:
			Renal failure
			Severe dehydration
24-h urinary creatinine 500–1200 mg/d (standardized for height and sex)	Low value: muscle wasting due to prolonged energy deficit	>24-h collection	**Low**
		Decreasing serum creatinine	Incomplete urine collection
			Increasing serum creatinine
			Neuromuscular wasting
24-h urinary urea nitrogen (UUN) <5 g/d (depends on level of protein intake)	Determine level of catabolism (as long as protein intake is ≥10 g below calculated protein loss or <20 g total, but at least 100 g carbohydrate is provided)		
	5–10 g/d = mild catabolism or normal fed state		
	10–15 g/d = moderate catabolism		
	>15 g/d = severe catabolism		
	Estimate protein balance		
	Protein balance = protein intake – protein loss where protein loss (protein catabolic rate) = [24-h UUN (g) + 4] x 6.25		
	Adjustments required in burn patients and others with large nonurinary nitrogen losses and in patients with fluctuating BUN levels (e.g., renal failure)		
Blood urea nitrogen (BUN) 8–23 mg/dL	<8: Possibly inadequate protein intake		**Low**
	12–23: Possibly adequate protein intake		Severe liver disease
	>23: Possibly excessive protein intake		Anabolic state
	If serum creatinine is normal, use BUN		Syndrome of inappropriate antidiuretic hormone
	If serum creatinine is elevated, use BUN/creatinine ratio (normal range is		**High**

	essen-tially the same as for BUN)		
			Despite poor protein intake:
			Renal failure (use BUN/creatinine ratio)
			Congestive heart failure
			Gastrointestinal hemorrhage

ASSESSMENT OF CIRCULATING (VISCERAL) PROTEINS

The serum proteins most commonly used to assess nutritional status include albumin, total iron-binding capacity (or transferrin), thyroxine-binding prealbumin (or transthyretin), and retinol-binding protein. Because they have differing synthesis rates and half-lives—the half-life of serum albumin is about 21 days, whereas those of prealbumin and retinol-binding protein are about 2 days and 12 h, respectively—some of these proteins reflect changes in nutritional status more quickly than do others. However, rapid fluctuations can also make shorter-half-life proteins less reliable.

Levels of circulating proteins are influenced by their rates of synthesis and catabolism, "third spacing" (loss into interstitial spaces), and, in some cases, external loss. Although an adequate intake of calories and protein is necessary to achieve optimal circulating protein levels, serum protein levels generally do not reflect protein intake. For example, a drop in the serum level of albumin or transferrin often accompanies significant physiologic stress (e.g., from infection or injury) and is not necessarily an indication of malnutrition or poor intake. A low serum albumin level in a burned patient with both hypermetabolism and increased dermal losses of protein may not indicate malnutrition. However, adequate nutritional support of the patient's calorie and protein needs is critical for returning circulating proteins to normal levels as stress resolves. Thus low values by themselves do not define malnutrition, but they often point to increased risk of malnutrition because of the hypermetabolic stress state. As long as significant physiologic stress persists, serum protein levels remain low, even with aggressive nutritional support. However, if the levels do not rise after the underlying illness improves, the patient's protein and calorie needs should be reassessed to ensure that intake is sufficient.

ASSESSMENT OF VITAMIN AND MINERAL STATUS

The use of laboratory tests to confirm suspected micronutrient deficiencies is desirable because the physical findings for those deficiencies are often equivocal or nonspecific. Low blood micronutrient levels can predate more serious clinical manifestations and also may indicate drug-nutrient interactions.

ESTIMATING ENERGY AND PROTEIN REQUIREMENTS

A patient's basal energy expenditures (BEE, measured in kilocalories per day) can be estimated from height, weight, age, and gender by using the Harris-Benedict equations:

Men: BEE = 66.47 + 13.75W + 5.00H – 6.76A

Women: BEE = 655.10 + 9.56W + 1.85H – 4.68A

where W is weight in kilograms; H is height in centimeters, and A is age in years. After these equations are solved, total energy requirements are estimated by multiplying BEE by a factor that accounts for the stress of illness. Multiplying by 1.1–1.4 yields a range 10–40% above basal that estimates the 24-h energy expenditure of the majority of patients. The lower value (1.1) is used for patients without evidence of significant physiologic stress; the higher value (1.4) is appropriate for patients with marked stress such as sepsis or trauma. The result is used as a 24-h energy goal for feeding.

When it is important to have a more accurate assessment of energy expenditure, it can be measured at the bedside by using indirect calorimetry. This technique is useful in patients who are believed to be hypermetabolic from sepsis or trauma and whose body weights cannot be obtained accurately. Indirect calorimetry can also be useful in patients who have difficulty weaning from a ventilator, as their energy needs should not be exceeded to avoid excessive CO_2 production. Patients at the extremes of weight (e.g., obese persons) and/or age are good candidates as well, because the Harris-Benedict equations were developed from measurements in adults with roughly normal body weights.

Because urea is a major by-product of protein catabolism, the amount of urea nitrogen excreted each day can be used to estimate the rate of protein catabolism and determine whether protein intake is adequate to offset it. Total protein loss and protein balance can be calculated from urinary urea nitrogen (UUN) as follows:

Protein catabolic rate (g/d) = [24-h UUN (g) + 4] x 6.25 (g protein/g nitrogen)

The value of 4 g added to the UUN represents a liberal estimate of the unmeasured nitrogen lost in the urine (e.g., creatinine and uric acid), sweat, hair, skin, and feces. When protein intake is low (e.g., less than about 20 g/d), the equation indicates both the patient's protein requirement and the severity of the catabolic state (Table 75-5). More substantial protein intakes can raise the UUN because some of the ingested (or infused) protein is catabolized and converted to UUN. Thus, at lower protein intakes the equation is useful for estimating *requirements*, and at higher protein intakes it is useful for assessing protein *balance*.

Protein balance (g/d) = protein intake – protein catabolic rate

FURTHER READINGS

Bales CW, Ritchie CS (eds): *Handbook of Clinical Nutrition and Aging* , 2nd ed. New York, Humana, 2009

Gottschlich MM (ed): *The A.S.P.E.N. Nutrition Support Core Curriculum: A Case-Based Approach—The Adult Patient*. Silver Spring, MD, A.S.P.E.N., 2007. Available online at www.nutritioncare.org

Heimburger DC, ARD JD (eds): *Handbook of Clinical Nutrition,* 4th ed. Philadelphia, Mosby Elsevier, 2006

Jensen GL et al: Adult starvation and disease-related malnutrition: A proposal for etiology-based diagnosis in the clinical practice setting from the International Consensus Guideline Committee. J Parenter Enteral Nutr 34:156, 2010[PMID: 20375423] [Full Text]

Shils ME et al (eds): *Modern Nutrition in Health and Disease*, 10th ed. Baltimore, Lippincott Williams & Wilkins, 2005

West KP et al: Nutrition, in *International Public Health: Diseases, Programs, Systems, and Policies*, 2nd ed, MH Merson et al (eds). Sudbury, MA, Jones and Bartlett, 2006, pp 207–269

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ENTERAL AND PARENTERAL NUTRITION THERAPY: INTRODUCTION

The ability to provide specialized nutritional support (SNS) represents a major advance in medical therapy. Nutritional support, via either enteral or parenteral routes, is used in two main settings: (1) to provide adequate nutritional intake during the recuperative phase of illness or injury, when the patient's ability to ingest or absorb nutrients is impaired, and (2) to support the patient during the systemic response to inflammation, injury, or infection during an extended critical illness. SNS is also used in patients with permanent loss of intestinal length or function. In addition, an increasing number of elderly patients living in nursing homes and chronic care facilities receive enteral feeding, usually as a consequence of inadequate nutritional intake.

Enteral refers to feeding via a tube placed into the gut to deliver liquid formulas containing all essential nutrients. *Parenteral* refers to the infusion of complete nutrient solutions into the bloodstream via a peripheral vein or, more commonly, by central venous access to meet nutritional needs. Enteral feeding is generally the preferred route because of benefits derived from maintaining the digestive, absorptive, and immunologic barrier functions of the gastrointestinal tract. Small-bore pliable tubes have largely replaced large-bore rubber tubes, making placement easier and more acceptable to patients. Infusion pumps have also improved the delivery of nutrient solutions.

For short-term use, enteral tubes can be placed via the nose into the stomach, duodenum, or jejunum. For long-term use, these sites can be accessed through the abdominal wall using endoscopic, radiologic, or surgical procedures. Intestinal tolerance of tube feeding may be limited during acute illness by gastric retention or diarrhea. Parenteral feeding has greater risk of infection, reflecting the need for venous access, and a greater propensity for inducing hyperglycemia. However, these risks can generally be managed successfully by SNS teams. For the postoperative patient with preexisting malnutrition, or in trauma patients who were previously well nourished, SNS is cost-effective. In the most critically ill patient in the intensive care unit, SNS can enhance survival. Although enteral nutrition (EN) can be provided by most health care teams caring for hospitalized patients, safe and effective parenteral nutrition (PN) usually requires specialized teams.

Approach to the Patient: Requirements for Specialized Nutritional Support

INDICATIONS FOR SPECIALIZED NUTRITIONAL SUPPORT

Although at least 15–20% of patients in acute care hospitals have evidence of significant malnutrition, only a small fraction will benefit from SNS. For others, wasting is an inevitable component of a terminal disease and the course of the disease will not be altered by SNS. The decision to use SNS should be based on the likelihood that preventing protein-calorie malnutrition (PCM) will increase the likelihood of recovery, reduce infection rates, improve healing, or otherwise shorten the hospital stay. In the case of the elderly or chronically ill patient for whom full recovery is not anticipated, the decision to feed is usually based on whether SNS will extend the duration or quality of life. The decision-making process used to assess whether to use SNS is depicted in **Fig. 76-1**.

Figure 76-1

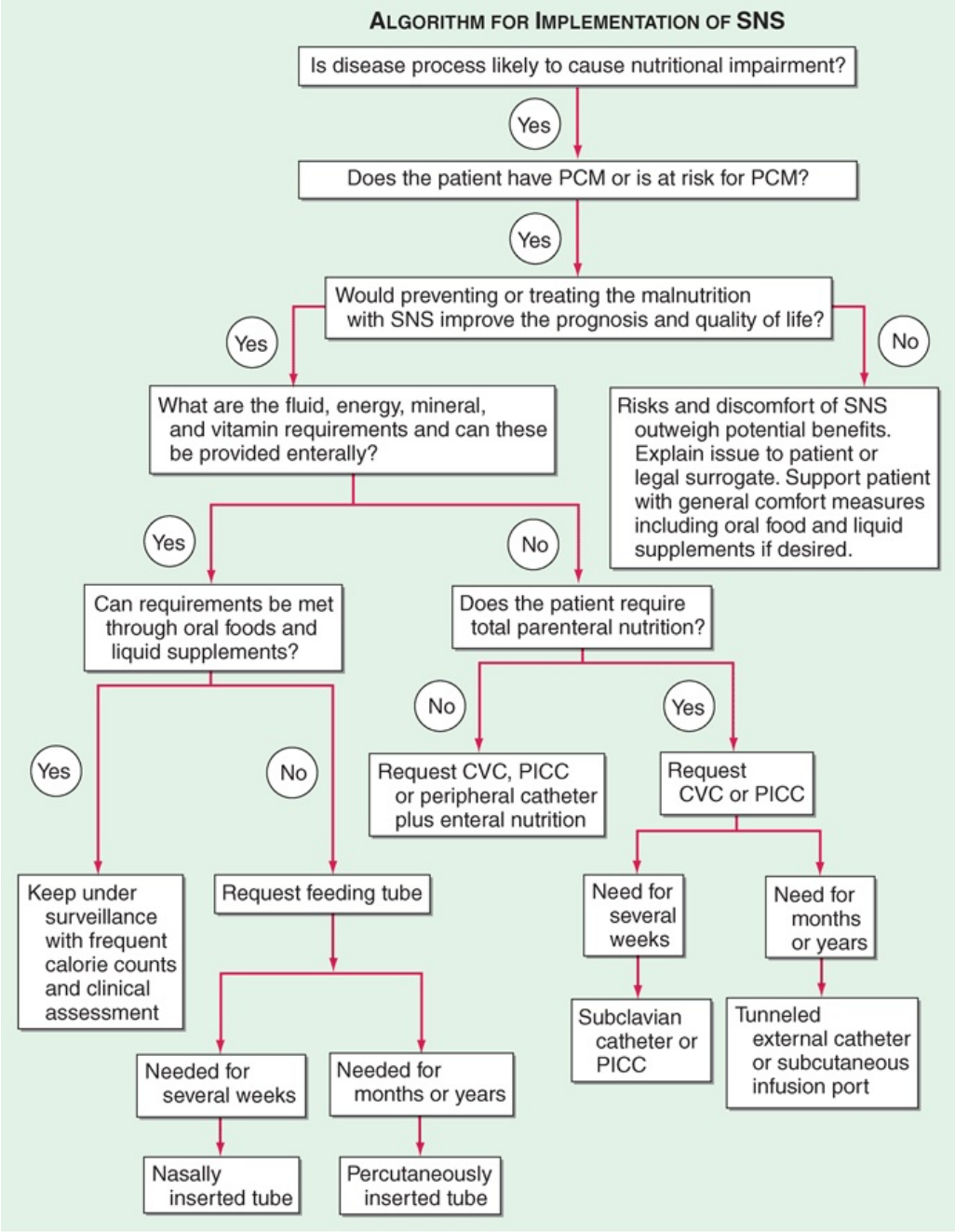

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Decision-making for the implementation of specialized nutrition support (SNS). CVC, central venous catheter; PICC, peripherally inserted central catheter. *(Adapted from chapter in Harrison's Principles of Internal Medicine, 16e, by Lyn Howard, MD.)*

The first step in deciding to administer SNS is to consider the nutritional implications of the disease process. Is the condition or its treatment likely to impair food intake and absorption for a prolonged period of time? For example, a well-nourished individual can tolerate approximately 7 days of starvation while experiencing a systemic response to inflammation (SRI). The second step is to determine if the patient is already significantly malnourished to the degree that critical functions such as wound healing, immune responses, or ventilatory function are impaired (Chap. 75). An unintentional weight loss of >10% during the previous 6 months or a weight/height <90% of standard, when associated with physiologic impairment, represents significant PCM. Weight loss >20% of usual or <80% of standard reflects severe PCM. The presence or absence of SRI should be noted, since inflammation, injury, and infection increase the rate of lean tissue loss. SRI also has pathophysiologic effects that influence nutritional responses such as fluid retention and hyperglycemia, as well as impairment of anabolic responses to nutritional support.

Once it is determined that a patient is already or at risk of becoming malnourished, the next step is to decide whether SNS will impact positively on the patient's response to disease. In the end stages of many chronic illnesses with accompanying PCM, particularly those due to cancer or terminal neurologic disorders, nutrition may not reverse the PCM or improve quality of life. While the provision of food and water is part of basic medical care, nutrition delivered by tube or catheter, either enterally or parenterally,

is associated with risk and discomfort. Thus, SNS should be recommended only when potential benefits exceed risks, and it should be undertaken with the consent of the patient. Like other life support measures, enteral or parenteral therapy is difficult to withdraw once started. Initiating nutrition support may be appropriate before a final prognosis can be determined, but this should not preclude its subsequent withdrawal. If preventing or treating PCM with SNS is appropriate, nutritional requirements and the method of delivery should be determined. The optimal route depends on the degree of gut function and somewhat on the available technical resources.

The timing of nutritional support is based on evaluation of the preexisting nutritional status, the presence and extent of SRI, and the anticipated clinical course. SRI is identified by the standard clinical signs of leukocytosis, tachycardia, tachypnea, and/or temperature elevation or depression. Although the degree of hypoalbuminemia provides an estimate of SRI severity, normal serum albumin levels will not be restored by adequate nutritional support until the SRI remits, even though nutritional benefits can be achieved by adequate feeding.

The SRI can be graded as severe, moderate, or mild. Examples of severe SRI include sepsis or other inflammatory conditions like pancreatitis requiring ICU care, multiple trauma with an Injury Severity Score > 20–25 or Acute Physiology and Chronic Health Evaluation II (APACHE II) > 25, closed head injury with a Glasgow Coma Scale < 8, or major third-degree burns of >40% of body surface area. Moderate SRI includes less severe infections, injuries, or inflammatory conditions like pneumonia, major surgery, acute hepatic or renal insufficiency, and exacerbations of ulcerative colitis or regional enteritis requiring hospitalization. PCM should also be defined as severe, moderate, or minimal as assessed by weight/height, percent recent weight loss, and body mass index. The body mass index in relation to nutritional status is listed in **Table 76-1**. A patient with a severe SRI requires early feeding within the first several days of care because the condition is likely to produce inadequate spontaneous intake over the next 7 days. A moderate SRI, as commonly seen during a postoperative period without oral intake that exceeds 5 days, benefits from adequate feeding by day 5–7 if the patient was initially well nourished. If severely malnourished, candidates for elective major surgery benefit from preoperative nutritional repletion for 5–7 days. However, this is not often possible. Thus, early postoperative feeding is indicated. Patients with a moderate SRI and moderate PCM also benefit from earlier feeding within the first several days.

Table 76-1 Body Mass Index (BMI) and Nutritional Status

BMI	Nutritional Status
>30 kg/m^2	Obese
>25–30 kg/m^2	Overweight
20–25 kg/m^2	Normal
<18.5 kg/m^2	Moderate malnutrition
<16 kg/m^2	Severe malnutrition
<13 kg/m^2	Lethal in males
<11 kg/m^2	Lethal in females

Source: From D Driscoll, B Bistrian: Parenteral and enteral nutrition in the intensive care unit, in *Intensive Care Medicine*, R Irwin, J Rippe (eds). Lippincott Williams & Wilkins, Philadelphia, 2003.

EFFICACY OF SNS IN DIFFERENT DISEASE STATES

Efficacy studies have shown that malnourished patients undergoing major thoracoabdominal surgery benefit from SNS. Critical illnesses requiring ICU care, including major burns, major trauma, severe sepsis, closed head injury, and severe pancreatitis (positive CT scan and APACHE II > 10), all benefit from early SNS, as indicated by reduced mortality and morbidity. In critical illness, initiation of SNS within 24 h of injury or ICU admission is associated with a ~50% reduction in mortality. Patients with nitrogen accumulation disorders of renal and hepatic failure have a likelihood of PCM of >50% and at least a moderate SRI. SNS is associated with improvements in morbidity, including infection rates, encephalopathy, liver or renal function, and length of hospital stay. Inflammatory bowel disease—including Crohn's disease particularly and, to a lesser degree, ulcerative colitis—often produce PCM. In the outpatient setting, SNS in Crohn's disease can improve nutritional status, quality of life, and the likelihood of remission. With pulmonary disease in the critically ill, SNS improves ventilatory status, and in acute lung injury the use of omega 3 fats as a component of SNS improves gas exchange and respiratory dynamics and reduces the need for mechanical ventilation. Low body mass in chronic obstructive pulmonary disease is associated with diminished pulmonary status and exercise capacity and higher mortality rates. However, there is little convincing evidence that SNS as caloric supplementation improves nutrition or pulmonary function. PCM is also common in the course of cancer and HIV disease, although less so in the latter with

the advent of highly active antiretroviral therapy. When PCM develops as a consequence of SRI in these conditions, there is limited likelihood of substantial efficacy or benefit from SNS. However, when PCM develops as a consequence of gastrointestinal dysfunction, SNS can be effective. Although no randomized trials have been performed for SNS provided for hyperemesis gravidarum, there is considerable clinical evidence that it improves pregnancy outcomes.

RISKS AND BENEFITS OF SPECIALIZED NUTRITION SUPPORT

The risks are determined primarily by patient factors such as state of alertness, swallowing competence, the route of delivery, underlying conditions, and the experience of the supervising clinical team. The safest and least costly approach is to avoid SNS by close attention to oral food intake, by adding an oral liquid supplement, or in certain chronic conditions by using medications to stimulate appetite. Nutrient intake monitoring by frequent calorie counts or oral formula selection is best performed by a nutritionist.

Enteral tube feeding is often required in patients with anorexia, impaired swallowing, or bowel disease. The bowel and its associated digestive organs derive 70% of their required nutrients directly from food in the lumen. Arginine, glutamine, short-chain fatty acids, long-chain omega 3 fatty acids, and nucleotides available in some specialty enteral formulas are particularly important for maintaining immunity. Enteral feeding also supports gut function by stimulating splanchnic blood flow, neuronal activity, IgA antibody release, and secretion of gastrointestinal hormones that stimulate gut trophic activity. These factors support the gut as an immunologic barrier against enteric pathogens. For these reasons, some luminal nutrition should be provided, even when PN is required to provide most of the nutritional support. The combination of some enteral feeding either by mouth or by enteral tube with parenteral feeding often shortens the transition to full enteral feeding, which can generally be used when >50% of requirements can be met enterally. Substantial nutritional benefit can be achieved by providing ~50% of energy needs for periods of up to 10 days, if protein and other essential nutrient requirements are met. For longer periods of time, it may be preferable to provide 75–80% of energy needs, rather than full feeding, if this improves gastrointestinal tolerance, glycemic control, and avoidance of excess fluid administration.

In the past, bowel rest through PN was the cornerstone of treatment for many severe gastrointestinal disorders. However, the value of providing even minimal amounts of EN is now widely accepted. Protocols to facilitate more widespread use of EN include initiation within 24 h of ICU admission; aggressive use of the head-upright position; postpyloric and nasojejunal feeding tubes; prokinetic agents; more rapid increases in feeding rates; tolerance of higher gastric residuals; and nurse-administered algorithms for feeding progression. PN alone is generally necessary only for severe gut dysfunction due to prolonged ileus, obstruction, or severe hemorrhagic pancreatitis. In the critically ill, feeding adequately by PN beginning within the first 24 h of care improves mortality and is more effective than delayed EN. Early feeding of the critically ill in the ICU is associated with a 50% reduction in mortality, but there is also a 50% increase in infection risk. Much of the increase in morbidity related to PN and EN is due to hyperglycemia, which can be significantly reduced by intensive insulin therapy. The level of glycemia necessary to accomplish this goal, whether <110 mg/dL or only <150 mg/dL, is not yet defined. Surgical patients being adequately fed may benefit from the lower glucose range, but studies of intensive insulin therapy alone without full feeding have shown improved morbidity and mortality outcome with looser control of glucose < 180 mg/dL.

Although PN was initially relatively expensive, its components are now often less expensive than specialty enteral formulas. Percutaneous placement of a central venous catheter into the subclavian or internal jugular vein with advancement into the superior vena cava can be accomplished at the bedside by trained personnel using sterile techniques. Peripherally inserted central catheters (PICCs) can also be placed within the lumen in the central vein, but this technique is usually more appropriate for non-ICU patients. Subclavian or internal jugular catheters carry a greater risk, including pneumothorax or serious vascular damage, but they are well tolerated and can be exchanged over a wire rather than requiring reinsertion when ruling out catheter infection. The peripherally inserted catheters are subject to position-related flow, and the catheter cannot be changed over a wire. Inserting a nasogastric tube is a bedside procedure, but many critically ill patients have impaired gastric emptying that increases the risk of aspiration pneumonia. This risk can be reduced by feeding directly into the jejunum beyond the ligament of Treitz. This usually requires fluoroscopic guidance or endoscopic placement. In patients who have planned laparotomies or other conditions likely to require a prolonged need for SNS, it is advantageous to place a jejunal feeding tube at the time of surgery.

Although most SNS is delivered in hospitals, some patients require it on a long-term basis. If they have a safe environment and a willingness to learn the self-care techniques, SNS can be administered at home. The clinical outcomes of patients with severe intestinal disorders treated with home PN or EN are summarized in **Table 76-2**. PN infused at home is usually cycled overnight to give greater daytime freedom. Other important considerations in determining the appropriateness of home PN or EN are that the patient's prognosis is longer than several months and that the therapy benefits quality of life. Recent advances in surgical techniques and immunosuppressive therapies have made intestinal transplantation a viable alternative for some patients who require life-long home parenteral nutrition. Although the quality of life can be improved with intestinal transplantation relative to home PN, long-term survival even in the most accomplished centers is still somewhat less with transplantation.

Table 76-2 Summary of Outcomes for Patients on Home Parenteral and Enteral Nutrition (HPEN)

				Therapy Status, % at 1 Year[b]			Rehabilitation Status, % in 1st Year[c]			Complications per Patient-Year[d]	
Diagnosis	**Number in Group**	**Age in Years**	**Survival on Therapy, %[a]**	**Full Oral Nutrition**	**Continued on HPEN Rx**	**Died**	**C**	**P**	**M**	**HPEN**	**NonHPEN**
Home Parenteral Nutrition											
Crohn's disease	562	36	96	70	25	2	60	38	2	0.9	1.1
Ischemic bowel disease	331	49	87	27	48	19	53	41	6	1.4	1.1
Motility disorder	299	45	87	31	44	21	49	39	12	1.3	1.1
Congenital bowel defect	172	5	94	42	47	9	63	27	11	2.1	1.0
Hyperemesis gravidarum	112	28	100	100	0	0	83	16	1	1.5	3.5
Chronic pancreatitis	156	42	90	82	10	5	60	38	2	1.2	2.5
Radiation enteritis	145	58	87	28	49	22	42	49	9	0.8	1.1
Chronic adhesive obstructions	120	53	83	47	34	13	23	68	10	1.7	1.4
Cystic fibrosis	51	17	50	38	13	36	24	66	16	0.8	3.7
Cancer	2122	44	20	26	8	63	29	57	14	1.1	3.3
AIDS	280	33	10	13	6	73	8	63	29	1.6	3.3
Home Enteral Nutrition											
Neurologic disorders of swallowing	1134	65	55	19	25	48	5	24	71	0.3	0.9
Cancer	1644	61	30	30	6	59	21	59	21	0.4	2.7

[a] Survival rates on therapy are values at 1 year, calculated by the life table method. This will differ from the percentage listed as died under Therapy Status, since all patients with known endpoints are considered in this latter measure. The ratio of observed versus expected deaths is equivalent to a Standard Mortality Ratio.

[b] Not shown are those patients who were back in hospital or who had changed therapy type by 12 months.

[c] Rehabilitation is designated complete (C), partial (P), or minimal (M), relative to the patient's ability to sustain normal age-related activity.

[d] Complications refer only to those complications that resulted in rehospitalization.

Source: Derived from North American HPEN Registry. Adapted from chapter in *Harrison's Principles of Internal Medicine*, 16e, by Lyn Howard, MD.

DISEASE-SPECIFIC NUTRITIONAL SUPPORT

SNS is basically a support therapy and is primary therapy only for the treatment or prevention of malnutrition. Certain conditions require modification of nutritional support because of organ or system impairment. For instance, in nitrogen accumulation disorders, protein intake may need to be reduced. However, in renal disease, except for brief periods of several days, protein intakes should approach requirement levels of at least 0.8 g/kg or higher up to 1.2 g/kg as long as the blood urea nitrogen does not exceed 100 mg/dL. If this is not possible, then dialysis or other renal replacement therapy should be considered to allow better feeding. In hepatic failure, intakes of 1.2–1.4 g/kg up to the optimal 1.5 g/kg should be attempted, as long as encephalopathy due to protein intolerance is not encountered. In the presence of protein intolerance, formulas containing 33–50% branched-chain amino acids are available and should be provided at the 1.2–1.4-g/kg level. Cardiac patients, and many severely stressed patients, often benefit from fluid and sodium restriction to levels of 1000 mL of total parenteral nutrition (TPN) formula and 5–20 meq of sodium per day. In patients with severe chronic PCM characterized by severe weight loss and tissue wasting, TPN must be instituted gradually because of the profound antinatriuresis, antidiuresis, and intracellular accumulation of potassium, magnesium, and phosphorus that develop as a

consequence of high insulin levels. This modification of TPN is usually accomplished by limiting fluid intakes initially to about 1000 mL containing modest carbohydrate content of 10–20% dextrose, low sodium, and ample potassium, magnesium, and phosphorus, with careful daily assessment of fluid and electrolyte status. Protein need not be restricted.

THE DESIGN OF INDIVIDUAL REGIMENS

Fluid Requirements

The normal daily requirement for fluid is 30 mL/kg of body mass from all sources (IV infusions, per tube, or oral intake), plus any replacement of abnormal losses such as from an osmotic diuresis, nasogastric drainage, wound output, or diarrheal/ostomy losses. Electrolyte and mineral losses can be estimated or measured and also need to be replaced **(Table 76-3)**. Fluid restriction may be necessary in patients with fluid overload, and fluid inputs can be limited to 1200 mL/d if urine is the only significant fluid output. When severe fluid overload is present, the optimal TPN solution for central venous administration is a concentrated 1-L solution of 7% crystalline amino acids (70 g) and 21% dextrose (210 g), which provides an amount of nitrogen and glucose that is optimally effective at protein-sparing.

Table 76-3 Enteric Fluid Volumes and Their Electrolyte Content[a]

	L/d	Na	K	Cl	HCO_3	H
Oral intake	2–3					
Enteric secretions						
Saliva	1–2	15	30	15	50	—
Gastric juice	1.5–2	50–70	5–15	90–120	0	70–100
Bile	0.5–1.5	120–150	5–15	80–120	30–50	—
Pancreatic	0.5–1	100–140	10	70–100	60–110	—
Small intestine	1–2	80–140	10–20	80–120	20–40	—

[a]All in meq/L.

Source: Adapted from chapter in *Harrison's Principles of Internal Medicine* , 16e, by Lyn Howard, MD.

Patients requiring PN or EN in the acute care setting generally have some element of associated hormonal adaptations to their underlying critical illness (e.g., increased secretion of antidiuretic hormone, aldosterone, insulin, glucagon, or cortisol) that cause fluid retention and hyperglycemia. Weight gain in the critically ill, whether receiving SNS or not, is invariably the consequence of fluid retention, since lean tissue accretion, even with feeding, is minimal in the acute phase of illness. Because excess fluid removal can be difficult, limiting fluid intake to allow for balanced intake and output is more effective.

Energy Requirements

Total energy expenditure comprises resting energy expenditure, activity energy expenditure, and the thermal effect of feeding (Chap. 75). Resting energy expenditure (two-thirds) includes the calories necessary for basal metabolism at bed rest. Activity energy expenditure represents one-fourth to one-third of the total, and the thermal effect of feeding is about 10% of the total energy expenditure. For normally nourished healthy individuals, the total energy expenditure is about 30–35 kcal/kg. Although critical illness increases resting energy expenditure, only in initially well- nourished individuals with the highest systemic inflammatory response, such as that from severe multiple trauma, burns, closed head injury, or sepsis, do total energy expenditures reach 40–45 kcal/kg. The chronically ill patient with lean tissue loss has reduced basal energy expenditure, as well as inactivity, which results in a total energy expenditure of about 20–25 kcal/kg. About 95% of such patients need <30 kcal/kg to achieve energy balance. Because providing about 50% of measured energy expenditure as SNS is at least equally efficacious for the first 10 days of critical illness, actual measurement of energy expenditure is not generally necessary in the early period of SNS. However, in patients who remain critically ill beyond several weeks, in the severely malnourished for whom estimates of energy expenditure are unreliable, or in those who are difficult to wean from ventilators, it is reasonable to actually measure energy expenditure and to aim for energy balance to 1.2 times measured expenditure with SNS.

Insulin resistance due to SRI is associated with increased gluconeogenesis and reduced peripheral glucose utilization, predisposing a patient to hyperglycemia. This is aggravated in patients receiving exogenous carbohydrate from SNS. Normalization of blood glucose levels by insulin infusion in critically ill patients receiving SNS reduces morbidity and mortality. In mild or moderately malnourished patients, a reasonable goal is to provide metabolic support to improve protein synthesis and maintain metabolic homeostasis. Hypocaloric nutrition providing only about 1000 kcal/d and 70 g protein for up to 10 days requires less fluid and reduces the likelihood of poor glycemic control. Energy content can be advanced to

20–25 kcal/kg with 1.5 g protein/kg as metabolic conditions permit and definitely during the second week of SNS. Patients with multiple trauma, closed head injury, and severe burns often have much higher energy expenditures, but there is little evidence that providing more than 30 kcal/kg has additional benefit, and it substantially increases the risks of hyperglycemia.

Generally, because glucose is an essential tissue fuel, glucose and amino acids are provided parenterally until the level of resting energy expenditure is reached. At this point, adding fat becomes beneficial, since more parenteral glucose stimulates de novo lipogenesis by the liver—an energy-inefficient process. Polyunsaturated long-chain triglycerides as soybean oil are the chief ingredient in most parenteral fat emulsions and the majority of the fat in enteral feeding formulas. These vegetable oil–based emulsions provide essential fatty acids. Enteral feeding formulas have fat content that ranges from 3% of calories up to as much as 50% of calories, while parenteral fat comes in separate containers as 10%, 20%, and 30% emulsions that can be infused separately or mixed by the pharmacy under controlled conditions as all-in-one or total nutrient admixture with glucose, amino acids, lipid, electrolytes, vitamins, and minerals. Although parenteral fat is required at only about 3% of energy requirements to meet essential fatty acid requirements, when provided as an all-in-one mixture of carbohydrate, fat, and protein, 2–3% fat in the TPN mixtures, representing about 20–30% of calories as fat, is provided to ensure emulsion stability. If given separately, parenteral fat should not be provided at rates exceeding 0.11 g/kg body mass per h or about 100 g over 12 h—equivalent to 1 L of 10% parenteral fat and 500 mL of 20% parenteral fat.

Medium-chain triglycerides, which contain saturated fatty acids with chain lengths of 6, 8, 10, or 12 carbons, are provided in a number of enteral feeding formulas because they are absorbed preferentially. Fish oil contains polyunsaturated fatty acids of the omega 3 family, which have been shown to improve immune function and reduce the inflammatory response.

Carbohydrates are provided as hydrous glucose providing 3.4 kcal/g in PN formulas. In enteral formulas, glucose is the carbohydrate source in so-called monomeric diets. These diets provide protein as amino acids and fat in minimal amounts (3%) to meet essential fatty acid requirements. Monomeric formulas are designed to optimize absorption in the seriously compromised gut. These formulas, like the immune-enhancing diets, are quite expensive. In polymeric diets, the carbohydrate source is usually an osmotically less active polysaccharide, protein is usually soy or casein protein, and fat is present in amounts from 25 to 50%. Such formulas are usually well tolerated by patients with normal intestinal length, and some are acceptable for oral consumption.

Protein or Amino Acid Requirements

Although the recommended dietary allowance for protein is 0.8 g/kg per d, maximal rates of repletion occur with 1.5 g/kg in the malnourished. In the severely catabolic patient, this higher level minimizes protein loss. In patients requiring SNS in the acute care setting, at least 1 g/kg is recommended, with greater amounts up to 1.5 g/kg as volume, renal, and hepatic tolerances allow. The standard parenteral and enteral formulas contain protein of high biologic value and meet the requirements for the eight essential amino acids when nitrogen needs are met. In protein-intolerant conditions such as renal and hepatic failure, modified amino acid formulas should be considered. In hepatic failure, higher branched-chain amino acid –enriched formulas appear to improve outcomes. Conditionally essential amino acids like arginine and glutamine may also have some benefit in supplemental amounts.

Protein (nitrogen) balance provides a measure of feeding efficacy of PN or EN. It is calculated as protein intake/6.25 because proteins are on average 16% nitrogen (N), minus the 24-h urine urea N (UUN) plus 4 g N, which reflects other N losses. In the critically ill, a mild negative balance of 2–4 g N/d is usually achievable with a similarly mild positive balance in the recuperating patient. Each g N represents approximately 30 g lean tissue.

Mineral and Vitamin Requirements

Parenteral electrolyte, vitamin, and trace mineral requirements are summarized in **Tables 76-4**, **76-5**, and **76-6**. Electrolyte modifications are necessary with substantial gastrointestinal losses from nasogastric drainage or intestinal losses from fistulas, diarrhea, or ostomy outputs. Such losses also imply extra calcium, magnesium, and zinc losses. Excessive urine or potassium losses with amphotericin, or magnesium losses with cisplatin or in renal failure, necessitate adjustments in sodium, potassium, magnesium, phosphorus, and acid-base balance. Vitamin and trace element requirements are met by the daily provision of a complete parenteral vitamin supplement and trace elements for PN, and with the provision of adequate amounts of enteral feeding formulas that contain these micronutrients.

Table 76-4 Usual Daily Electrolyte Additions to Parenteral Nutrition

Electrolyte	Parenteral Equivalent of RDA	Usual Intake
Sodium		1–2 meq/kg + replacement, but can be as low as 5–40 meq/d
Potassium		40–100 meq/d + replacement of unusual losses
Chloride		As needed for acid-base balance, but usually 2:1 to 1:1 with acetate

Acetate		As needed for acid-base balance
Calcium	10 meq	10–20 meq/d
Magnesium	10 meq	8–16 meq/d
Phosphorus	30 mmol	20–40 mmol

Table 76-5 Parenteral Multivitamin Requirements for Adults

Vitamin	Recently Revised Value
Vitamin A	3300 IU
Thiamin (B_1)	6 mg
Riboflavin (B_2)	3.6 mg
Niacin (B_3)	40 mg
Folic acid	600 g
Pantothenic acid	15 mg
Pyridoxine (B_6)	6 mg
Cyanocobalamin (B_{12})	5 μg
Biotin	60 μg
Ascorbic acid (C)	200 mg
Vitamin D	200 IU
Vitamin E	10 IU
Vitamin K[a]	150 μg

[a] A product is available without vitamin K. Vitamin K supplementation is recommended at 2–4 mg/week in patients not receiving oral anticoagulation therapy if using this product.

Table 76-6 Parenteral Trace Metal Supplementation for Adults[a]

Trace Mineral	Intake
Zinc	2.5–4 mg/d, an additional 10–15 mg/d per L of stool or ileostomy output
Copper	0.5–1.5 mg/d, possibility of retention in biliary tract obstruction
Manganese	0.1–0.3 mg/d, possibility of retention in biliary tract obstruction
Chromium	10–15 μg/d
Selenium	20–100 μg/d, necessary for long-term PN, optional for short-term TPN
Molybdenum	20–120 μg/d, necessary for long-term PN, optional for short-term PN
Iodine	75–150 μg/d, necessary for long-term PN, optional for short-term PN

[a] Commercial products are available with the first four, first five, and all seven of these metals in recommended amounts.

Note: PN, parenteral nutrition; TPN, total parenteral nutrition.

PARENTERAL NUTRITION

Infusion Technique and Patient Monitoring

Parenteral feeding through a peripheral vein is limited by osmolality and volume constraints. Solutions that contain more than 3% amino acids and 5% glucose (290 kcal/L) are poorly tolerated peripherally. Parenteral fat (20%) can be given to increase the calories delivered. The total volume required to provide

a marginal protein intake of 60 g and 1680 total kcal is 2.5 L. Moreover, the risk of significant morbidity and mortality from incompatibilities of calcium and phosphate salts is greatest in these low-osmolality, low-glucose regimens. Parenteral feeding via a peripheral vein is generally intended as a supplement to oral feeding and is not optimal for the critically ill. Peripheral parenteral nutrition may benefit from small amounts of heparin at 1000 U/L and co-infusion with parenteral fat to reduce osmolality, but volume constraints still limit the value of this therapy. PICCs can be used for the short term to provide concentrated glucose parenteral solutions of 20–25% dextrose and 4–7% amino acids, while avoiding some of the complications of catheter placement via a large central vein. With PICC lines, however, flow can be position-related, and the lines cannot be exchanged over a wire for infection monitoring. For these reasons, in the critically ill, centrally placed catheters are preferred. The subclavian approach is best tolerated by the patient and is the easiest to dress. The jugular approach is less likely to lead to a pneumothorax. The femoral approach is discouraged because of the greater risk of catheter infection. For long-term feeding in the home, tunneled catheters and implanted ports reduce infection risk and are more acceptable to patients. However, tunneled catheters require placement in the operating room.

Catheters are made of silastic, polyurethane, or polyvinyl chloride. Silastic catheters are less thrombogenic and are best for tunneled catheters. Polyurethane is best for temporary catheters. Dressing changes with dry gauze at regular intervals should be performed by nurses skilled in catheter care to avoid infection. Chlorhexidine solution is more effective than alcohol or iodine compounds. Appropriate monitoring for patients receiving PN is summarized in **Table 76-7**.

Table 76-7 Monitoring the Patient on Parenteral Nutrition

Clinical Data Monitored Daily	
General sense of well-being	
Strength as evidenced in getting out of bed, walking, resistance exercise as appropriate	
Vital signs including temperature, blood pressure, pulse, and respiratory rate	
Fluid balance: weight at least several times weekly, fluid intake (parenteral and enteral) vs. fluid output (urine, stool, gastric drainage, wound, ostomy)	
Parenteral nutrition delivery equipment: tubing, pump, filter, catheter, dressing	
Nutrient solution composition	
Laboratory Daily	
Finger-stick glucose	Three times daily until stable
Blood glucose, Na, K, Cl, HCO_3, BUN	Daily until stable and fully advanced, then twice weekly
Serum creatinine, albumin, PO_4, Ca, Mg, Hb/Hct, WBC	Baseline, then twice weekly
INR	Baseline, then weekly
Micronutrient tests	As indicated

Note: BUN, blood urea nitrogen; Hb, hemoglobin; Hct, hematocrit; INR, international normalized ratio; WBC, white blood cell count.

Source: Adapted from chapter in *Harrison's Principles of Internal Medicine* , 16e, by Lyn Howard, MD.

Complications

MECHANICAL

The insertion of a central venous catheter should be performed by trained and experienced personnel using aseptic techniques to limit the major common complications of pneumothorax and inadvertent arterial puncture or injury. Catheter position should be radiographically confirmed to be in the superior vena cava distal to the junction with the jugular or subclavian vein and not directly against the vessel wall. Thrombosis related to the catheter may occur at the site of entry into the vein and extend to encase the catheter. Catheter infection predisposes to thrombosis, as does the SRI. The addition of 6000 U of heparin in the daily parenteral formula in hospitalized patients with temporary catheters reduces the risk of fibrin sheath formation and catheter infection. Temporary catheters that develop a thrombus should be removed and, based on clinical findings, treated with anticoagulants. Thrombolytic therapy can be considered for patients with permanent catheters, depending on the ease of replacement and presence of alternate, reasonably acceptable venous access sites. Low-dose warfarin therapy of 1 mg/d reduces the risk of thrombosis in permanent catheters used for home PN, but full anticoagulation may be required in patients who have recurrent thrombosis related to permanent catheters. A recent U.S. Food and Drug Administration mandate to reformulate parenteral multivitamins to include vitamin K at a dose of 150 μg

daily may affect the efficacy of low-dose warfarin therapy. There is a "no vitamin K" version available for patients receiving this therapy. Catheters can become mechanically occluded and may also become occluded by fibrin at the tip, or by fat, minerals, or drugs intraluminally. These occlusions can be managed with low-dose alteplase for fibrin, with indwelling 70% alcohol for fat, with 0.1 *N* hydrochloric acid for mineral precipitates, and with either 0.1 *N* hydrochloric acid or 0.1 *N* sodium hydroxide for drugs, depending on their pH.

METABOLIC

The most common problems related to PN are fluid overload and hyperglycemia **(Table 76-8)**. Hypertonic dextrose stimulates a much higher insulin level than meal feeding. Because insulin is a potent antinatriuretic and antidiuretic hormone, hyperinsulinemia leads to sodium and fluid retention. In the absence of gastrointestinal losses or renal dysfunction, net fluid retention is likely when total fluid intake exceeds 2000 mL/d. Close monitoring of body mass, as well as fluid intake and output, is necessary to prevent this complication. In the absence of significant renal impairment, the sodium content of the urine is likely to be <10 meq/L. Providing sodium in limited amounts of 40 meq/d and the use of both glucose and fat in the PN mixture to lower total glucose and sodium will help reduce fluid retention. The elevated insulin also increases the intracellular transport of potassium, magnesium, and phosphorus, which can precipitate a dangerous refeeding syndrome if the total glucose content of the PN solution is advanced too quickly in severely malnourished patients. It is generally best to start PN with <200 g glucose/d to assess glucose tolerance. Regular insulin can be added to the PN formula to establish glycemic control, and the insulin doses can be increased proportionately as the glucose is advanced. As a general rule, patients with insulin-dependent diabetes require about twice their usual home insulin doses when they are receiving TPN at 20–25 kcal/kg, largely as a consequence of parenteral glucose administration and some loss of insulin to the TPN container. As a rough estimate, the amount of insulin can be provided in a similar proportion to the amount of calories provided as TPN relative to full feeding, and the insulin can be placed in the TPN formula. Subcutaneous (SC) regular insulin can be provided to improve glucose control as assessed by measurements of blood glucose every 6 h. About two-thirds of the total 24-h amount can be added to the next day's order, with SC insulin supplements as needed. Advances in TPN concentration should be made when reasonable glucose control is established, and the insulin dose adjusted proportionately to the calories added as glucose and amino acids. These are general rules, and they are conservative. Given the adverse clinical impact of hyperglycemia, it may be necessary to use intensive insulin therapy as a separate infusion with a standard protocol to initially establish control. Once established, this insulin dose can be added to the PN formula. Acid-base imbalance is also common during PN therapy. Amino acid formulas are buffered, but critically ill patients are prone to metabolic acidosis, often due to renal tubular impairment. The use of sodium and potassium acetate salts in the PN formula may address this problem. Bicarbonate salts should not be used because they are incompatible with TPN formulations. Nasogastric drainage produces a hypochloremic alkalosis that can be managed by attention to chloride balance. Occasionally, hydrochloric acid may be required for a more rapid response or when diuretic therapy limits the ability to provide substantial sodium chloride. Up to 100 meq/L and up to 150 meq of hydrochloric acid per day may be placed in a fat-free TPN formula.

Table 76-8 Selected Metabolic Disturbances and Their Correction

Disturbance	Cause	Corrective Action with PN
Hyponatremia	Increased total body water or decreased total body sodium	Decrease free water or increase sodium
Hypernatremia	Occurs commonly with excessive isotonic or hypertonic fluid followed by diuretic administration with free water clearance; can also occur with dehydration and normal total body sodium	Increase free water to produce net positive fluid balance maintaining sodium and chloride balance
Hypokalemia	Inadequate intake relative to need	Use supplements
	Excessive diuresis, tubular dysfunction	Use supplements
	Magnesium deficiency	Increase PNmagnesium
	Metabolic alkalosis	Correct alkalosis
	Hyperinsulinemia	Maintain constant PN, increase potassium
Hyperkalemia	Excessive provision	Reduce supplements
	Metabolic acidosis	Evaluate acidosis, treat with PN acetate salt and decrease potassium
	Renal deterioration	Evaluate patient and adjust PN as indicated

Hypocalcemia	Reciprocal response to phosphorus repletion	Increase calcium
	Critical illness effect	Increase calcium
	Severe malabsorption	Supplement calcium
Hypercalcemia	Excessive administration or pathologic (cancer, hyperparathyroidism)	Reduce or eliminate calcium
Hypomagnesemia	Increased requirements due to diuretic use, alcoholism, malabsorption, malnutrition	Supplement magnesium
	Critical illness	Supplement magnesium
Hypophosphatemia	Inadequate intake relative to needs related to malnutrition, alcohol use	Supplement phosphorus
	Increased calcium intake	Use supplements
Hyperphosphatemia	Excessive administration or worsening renal function	Reduce phosphorus
Azotemia	Excessive amino acid infusion or worsening renal function	Reduce amino acid level but consider renal replacement therapy if cannot provide 1 g protein per kg for prolonged periods

Abbreviation: PN, parenteral nutrition.

INFECTIOUS

Infections of the central access catheter rarely occur in the first 72 h. Fever during this period is usually from infection elsewhere or another cause. Fever that develops during PN can be addressed by checking the catheter site and, if the site looks clean, exchanging the catheter over a wire with cultures taken through the catheter and at the catheter tip. If these cultures are negative, as they are most of the time, the new catheter can continue to be used. If a culture is positive for a relatively nonpathogenic bacteria like *Staphylococcus epidermidis*, consider a second exchange over a wire with repeat cultures or replace the catheter depending on the clinical circumstances. If cultures are positive for more pathogenic bacteria, or for fungi like *Candida albicans*, it is generally best to replace the catheter at a new site. Whether antibiotic treatment is required is a clinical decision, but *C. albicans* grown from the blood culture in a patient receiving PN should always be treated because the consequences of failure to treat can be dire.

Catheter infections can be minimized by dedicating the feeding catheter to TPN, without blood sampling or medication administration. Central catheter infections are a serious complication with an attributed mortality of 12–25%. Infections in central venous catheters dedicated to feeding should occur less frequently than 3 per 1000 catheter-days. Home TPN catheters that become infected may be treated through the catheter without removal of the catheter, particularly if the offending organism is *S. epidermidis*. Clearing of the biofilm and fibrin sheath by local treatment of the catheter with indwelling alteplase may increase the likelihood of eradication. Antibiotic lock therapy with high concentrations of antibiotic, with or without heparin in addition to systemic therapy, may improve efficacy. Sepsis with hypotension should precipitate catheter removal in either the temporary or permanent TPN setting.

ENTERAL NUTRITION

Tube Placement and Patient Monitoring

The types of enteral feeding tubes, methods of insertion, their clinical uses, and potential complications are outlined in **Table 76-9**. The different types of enteral formulas are listed in **Table 76-10**. Patients receiving EN are at risk for many of the same metabolic complications as those who receive PN and should be monitored in the same manner. EN can be a source of similar problems, but not to the same degree, because the insulin response to EN is about half of that seen with PN. Enteral feeding formulas have fixed electrolyte compositions that are generally modest in sodium and somewhat higher in potassium content. Acid-base disturbances can be addressed to a more limited extent with EN. Acetate salts can be added to the formula to treat chronic metabolic acidosis. Calcium chloride can be added to treat mild chronic metabolic alkalosis. Medications and other additives to enteral feeding formulas can clog the tubes (e.g., calcium chloride may interact with casein-based formulas to produce insoluble calcium caseinate products) and may reduce the efficacy of some drugs (e.g., phenytoin). Since small-bore tubes are easily displaced, tube position should be checked at intervals by aspirating and measuring the pH of the gut fluid (<4 in the stomach, >6 in the jejunum).

Table 76-9 Enteral Feeding Tubes

Type/Insertion Technique	Clinical Uses	Potential Complications

Nasogastric Tube		
External measurement: nostril, ear, xiphisternum; tube stiffened by ice water or stylet; position verified by injecting air and auscultating, or by x-ray	Short-term clinical situation (weeks) or longer periods with intermittent insertion; bolus feeding simpler, but continuous drip with pump better tolerated	Aspiration; ulceration of nasal and esophageal tissues, leading to stricture
Nasoduodenal Tube		
External measurement: nostril, ear, anterior superior iliac spine; tube stiffened by stylet and passed through pylorus under fluoroscopy or with endoscopic loop	Short-term clinical situations where gastric emptying impaired or proximal leak suspected; requires continuous drip with pump	Spontaneous pulling back into stomach (position verified by aspirating content, pH >6); diarrhea common, fiber-containing formulas may help
Gastrostomy Tube		
Percutaneous placement endoscopically, radiologically, or surgically; after tract established, can be converted to a gastric "button"	Long-term clinical situations, swallowing disorders, or impaired small-bowel absorption requiring continuous drip	Aspiration; irritation around tube exit site; peritoneal leak; balloon migration and obstruction of pylorus
Jejunostomy Tube		
Percutaneous placement endoscopically or radiologically via pylorus or endoscopically or surgically directly into the jejunum	Long-term clinical situations where gastric emptying impaired; requires continuous drip with pump; direct endoscopic placement (PEJ) is the most comfortable for patient	Clogging or displacement of tube; jejunal fistula if large-bore tube used; diarrhea from dumping; irritation of surgical anchoring suture
Combined Gastrojejunostomy Tube		
Percutaneous placement endoscopically, radiologically, or surgically; intragastric arm for continuous or intermittent gastric suction; jejunal arm for enteral feeding	Used for patients with impaired gastricemptying and at high risk for aspiration or patients with acute pancreatitis or proximal leaks	Clogging: especially of small-bore jejunal tube

Note: All small tubes are at risk for clogging, especially if used for crushed medications. In long-term enteral patients, gastrostomy and jejunostomy tubes can be exchanged for a low-profile "button" once the tract is established. PEJ, percutaneous endoscopic jejunostomy.

Source: Adapted from chapter in *Harrison's Principles of Internal Medicine* , 16e, by Lyn Howard, MD.

Table 76-10 Enteral Formulas

Composition Characteristics	Clinical Indications
Standard Enteral Formula	
1. Complete dietary products (+)[a] a. Caloric density 1 kcal/mL	Suitable for most patients requiring tube feeding; some can be used orally
b. Protein ~14% cals, caseinates, soy, lactalbumin	
c. CHO ~60% cals, hydrolyzed corn starch, maltodextrin, sucrose d. Fat ~30% cals, corn, soy, safflower oils	
e. Recommended daily intake of all minerals and vitamins in >1500 kcal/d	
f. Osmolality (mosmol/kg): ~300	
Modified Enteral Formulas	
1. Caloric density 1.5–2 kcal/mL (+) 2. a. High protein ~20–25% protein (+)	Fluid-restricted patients Critically ill patients
b. Hydrolyzed protein to small peptides (+)	Impaired absorption
c. ↑ Arginine, glutamine, nucleotides, ω3 fat (+++)	Immune-enhancing diets
d. ↑ Branched-chain amino acids, ↓ aromatic amino acids (+++)	Liver failure patients intolerant of 0.8 g/kg protein

e. Low protein of high biologic value	Renal failure patient for brief periods if critically ill
3. a. Low-fat partial MCT substitution (+) b. ↑ Fat >40% cals (++)	Fat malabsorption Pulmonary failure with CO_2 retention on standard formula, limited utility
c. ↑ Fat from MUFA (++)	Improvement in glycemic index control in diabetes
d. ↑ Fat from ω3 and ↓ ω6 linoleic acid (+++) 4. Fiber provided as soy polysaccharide (+)	Improved ventilation in ARDS Improved laxation

[a]Cost: + inexpensive; ++ moderately expensive; +++ very expensive.

Note: ARDS, acute respiratory distress syndrome; CHO, carbohydrate; MCT, medium-chain triglyceride; MUFA, monounsaturated fatty acids; ω3 or ω6, polyunsaturated fat with first double bond at carbon 3 (fish oils) or carbon 6 (vegetable oils).

Source: Adapted from chapter in *Harrison's Principles of Internal Medicine* , 16e, by Lyn Howard, MD.

Complications

ASPIRATION

The debilitated patient with poor gastric emptying and impairment of swallowing and cough is at risk for aspiration; this is particularly true for those who are mechanically ventilated. Tracheal suctioning induces coughing and gastric regurgitation, and cuffs on endotracheal or tracheostomy tubes seldom protect against aspiration. Preventive measures include elevating the head of the bed to 30 degrees, using nurse-directed algorithms for formula advancement, combining enteral with parenteral feeding, and using post–ligament of Treitz feeding. Tube feeding should not be discontinued for gastric residuals of <300 mL unless there are other signs of gastrointestinal intolerance such as nausea, vomiting, or abdominal distention. Continuous feeding using pumps is better tolerated intragastrically and is essential for feeding into the jejunum. For small-bowel feeding, residuals are not assessed but abdominal pain and distention should be monitored.

DIARRHEA

Enteral feeding often leads to diarrhea, especially if bowel function is compromised by disease or drugs, particularly broad-spectrum antibiotics. Diarrhea may be controlled by the use of a continuous drip, with a fiber-containing formula, or by adding an antidiarrheal agent to the formula. However, *Clostridium difficile*, which is a common cause of diarrhea in patients being tube fed, should be ruled out before using antidiarrheal agents. H2 blockers may also assist in reducing the net fluid presented to the colon. Diarrhea associated with enteral feeding does not necessarily imply inadequate absorption of nutrients other than water and electrolytes. Amino acids and glucose are particularly well absorbed in the upper small bowel except in the most diseased or shortest bowel. Since luminal nutrients exert trophic effects on the gut mucosa, it is often appropriate to persist with tube feeding, despite the diarrhea, even when this necessitates supplemental parenteral fluid support.

GLOBAL CONSIDERATIONS

In the United States the only parenteral lipid emulsion available is made with soybean oil, and it has been suggested that its constituent fatty acids may be immunosuppressive under certain circumstances. In Europe and Japan there are a number of other lipid emulsions available, including those containing fish oil only; mixtures of fish oil, medium-chain triglycerides, and long-chain triglycerides as olive oil and/or soybean oil; mixtures of medium-chain triglycerides and long-chain triglycerides as soybean oil; and long-chain triglyceride mixtures of olive oil and soybean oil, which may be more beneficial in terms of metabolism and hepatic and immune function. Furthermore, a glutamine-containing dipeptide for inclusion in TPN formulas is available in Europe and may be helpful in terms of immune function and resistance to infection.

ACKNOWLEDGMENT

The authors acknowledge the contributions of Lyn Howard, MD, the author in earlier editions of HPIM, to material in this chapter.

FURTHER READINGS

Bistrian B, McCowen K: Nutritional support in the adult intensive care unit: Key controversies. Crit Care Med 34:1525, 2006[PMID: 16557154] [Full Text]

Centers for Disease Control and Prevention: Reduction in central line–associated bloodstream infections among patients in intensive care units—Pennsylvania, April 2001–March 2005. MMWR Morb Mortal Wkly Rep 54:1013, 2005

Fishbein T: Intestinal transplantation. N Engl J Med 361:998, 2009[PMID: 19726774] [Full Text]

Koretz RL et al: Does enteral nutrition affect clinical outcome? A systematic review of the randomized

trials. Am J Gastroenterol 102:412, 2007[PMID: 17311654] [Full Text]

McClave SA et al: Guidelines for the provision and assessment of nutrition support therapy in the adult critically ill patient: Society of Critical Care Medicine (SCCM) and American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.) JPEN J Parenter Enteral Nutr 33:277, 2009[PMID: 19398613] [Full Text]

Milne A et al: Meta-analysis: Protein and energy supplementation in older people. Ann Intern Med 144:37, 2006[PMID: 16389253] [Full Text]

The Nice-Sugar Study Investigators: Intensive versus conventional glucose control in critically ill patients. N Engl J Med 360:1283, 2009

Simpson F, Doig GS: Parenteral vs. enteral nutrition in the critically ill patient: A meta-analysis of trials using the intention to treat principle. Intensive Care Med 31:12, 2005[PMID: 15592814] [Full Text]

Singer P et al: ESPEN guidelines on parenteral nutrition: Intensive care. Clin Nutr 28:387, 2009[PMID: 19505748] [Full Text]

van den Berghe G et al: Intensive insulin therapy in the critically ill patients. N Engl J Med 345:1359, 2001

Zeigler TR: Parenteral nutrition in the critically ill patient. N Engl J Med 361:1088, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BIOLOGY OF OBESITY: INTRODUCTION

In a world where food supplies are intermittent, the ability to store energy in excess of what is required for immediate use is essential for survival. Fat cells, residing within widely distributed adipose tissue depots, are adapted to store excess energy efficiently as triglyceride and, when needed, to release stored energy as free fatty acids for use at other sites. This physiologic system, orchestrated through endocrine and neural pathways, permits humans to survive starvation for as long as several months. However, in the presence of nutritional abundance and a sedentary lifestyle, and influenced importantly by genetic endowment, this system increases adipose energy stores and produces adverse health consequences.

Definition and Measurement

Obesity is a state of excess adipose tissue mass. Although often viewed as equivalent to increased body weight, this need not be the case—lean but very muscular individuals may be overweight by numerical standards without having increased adiposity. Body weights are distributed continuously in populations, so that choice of a medically meaningful distinction between lean and obese is somewhat arbitrary. Obesity is therefore more effectively defined by assessing its linkage to morbidity or mortality.

Although not a direct measure of adiposity, the most widely used method to gauge obesity is the *body mass index* (BMI), which is equal to weight/height^2 (in kg/m^2) **(Fig. 77-1)**. Other approaches to quantifying obesity include anthropometry (skinfold thickness), densitometry (underwater weighing), CT or MRI, and electrical impedance. Using data from the Metropolitan Life Tables, BMIs for the midpoint of all heights and frames among both men and women range from 19 to 26 kg/m^2; at a similar BMI, women have more body fat than men. Based on data of substantial morbidity, a BMI of 30 is most commonly used as a threshold for obesity in both men and women. Large-scale epidemiologic studies suggest that all-cause, metabolic, cancer, and cardiovascular morbidity begin to rise (albeit at a slow rate) when BMIs are ≥25, suggesting that the cutoff for obesity should be lowered. Most authorities use the term *overweight* (rather than obese) to describe individuals with BMIs between 25 and 30. A BMI between 25 and 30 should be viewed as medically significant and worthy of therapeutic intervention, especially in the presence of risk factors that are influenced by adiposity such as hypertension and glucose intolerance.

Figure 77-1

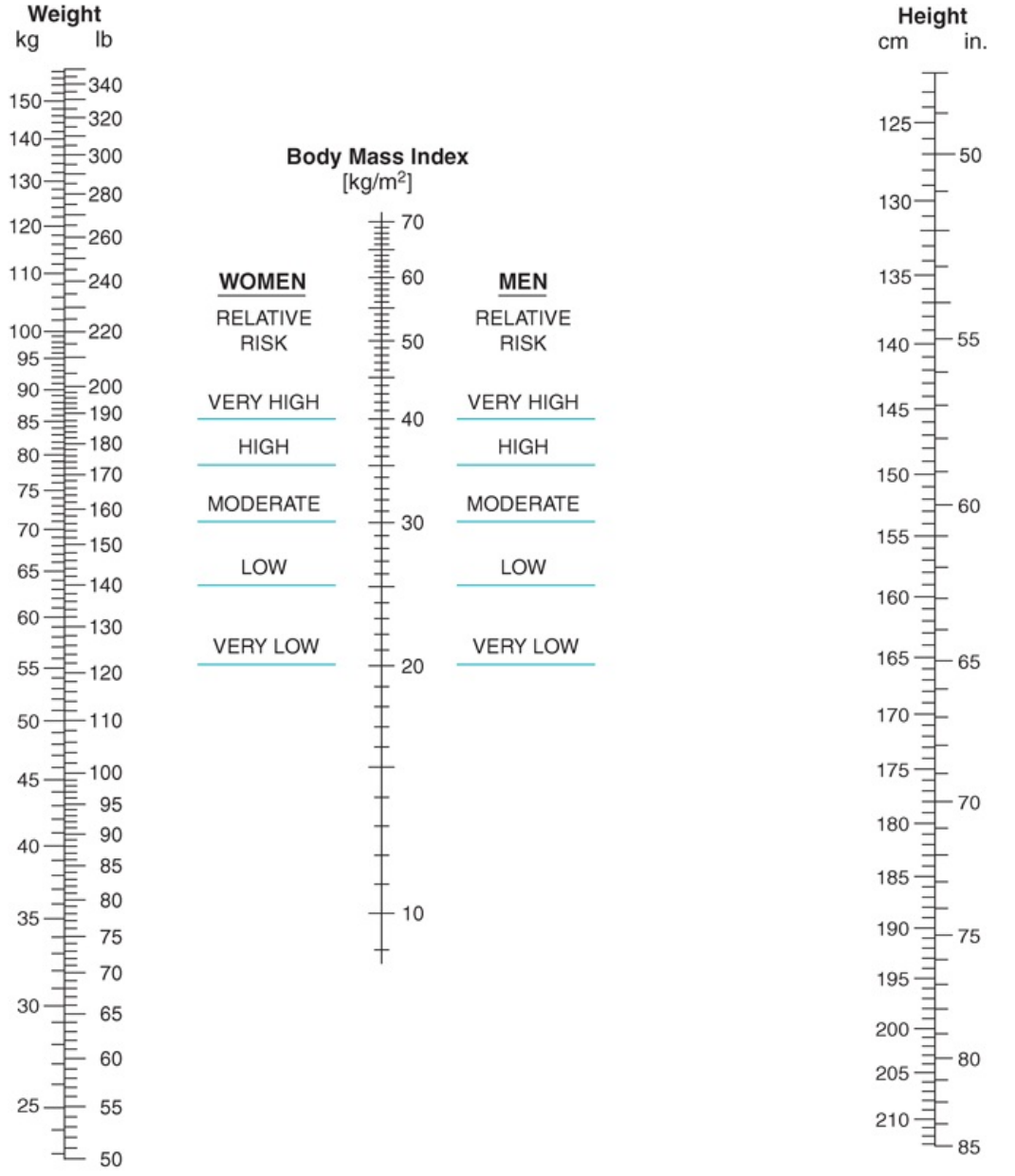

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nomogram for determining body mass index. To use this nomogram, place a ruler or other straight edge between the body weight (without clothes) in kilograms or pounds located on the left-hand line and the height (without shoes) in centimeters or inches located on the right-hand line. The body mass index is read from the middle of the scale and is in metric units. *(Copyright 1979, George A. Bray, MD; used with permission.)*

The distribution of adipose tissue in different anatomic depots also has substantial implications for morbidity. Specifically, intraabdominal and abdominal subcutaneous fat have more significance than subcutaneous fat present in the buttocks and lower extremities. This distinction is most easily made clinically by determining the waist-to-hip ratio, with a ratio >0.9 in women and >1.0 in men being abnormal. Many of the most important complications of obesity such as insulin resistance, diabetes, hypertension, hyperlipidemia, and hyperandrogenism in women, are linked more strongly to intraabdominal and/or upper body fat than to overall adiposity (Chap. 242). The mechanism underlying this association is unknown but may relate to the fact that intraabdominal adipocytes are more lipolytically active than those from other depots. Release of free fatty acids into the portal circulation has adverse metabolic actions, especially on the liver. Whether adipokines and cytokines secreted by visceral adipocytes play an additional role in systemic complications of obesity is an area of active investigation.

Prevalence

Data from the National Health and Nutrition Examination Surveys (NHANES) show that the percentage of the American adult population with obesity (BMI >30) has increased from 14.5% (between 1976 and 1980) to 33.9% (between 2007 and 2008). As many as 68% of U.S. adults aged ≥20 years were

overweight (defined as BMI >25) between the years of 2007 and 2008. Extreme obesity (BMI ≥40) has also increased and affects 5.7% of the population. The increasing prevalence of medically significant obesity raises great concern. Obesity is more common among women and in the poor, and among blacks and Hispanics; the prevalence in children is also rising at a worrisome rate.

Physiologic Regulation of Energy Balance

Substantial evidence suggests that body weight is regulated by both endocrine and neural components that ultimately influence the effector arms of energy intake and expenditure. This complex regulatory system is necessary because even small imbalances between energy intake and expenditure will ultimately have large effects on body weight. For example, a 0.3% positive imbalance over 30 years would result in a 9-kg (20-lb) weight gain. This exquisite regulation of energy balance cannot be monitored easily by calorie-counting in relation to physical activity. Rather, body weight regulation or dysregulation depends on a complex interplay of hormonal and neural signals. Alterations in stable weight by forced overfeeding or food deprivation induce physiologic changes that resist these perturbations: with weight loss, appetite increases and energy expenditure falls; with overfeeding, appetite falls and energy expenditure increases. This latter compensatory mechanism frequently fails, however, permitting obesity to develop when food is abundant and physical activity is limited. A major regulator of these adaptive responses is the adipocyte-derived hormone leptin, which acts through brain circuits (predominantly in the hypothalamus) to influence appetite, energy expenditure, and neuroendocrine function (see below).

Appetite is influenced by many factors that are integrated by the brain, most importantly within the hypothalamus **(Fig. 77-2)**. Signals that impinge on the hypothalamic center include neural afferents, hormones, and metabolites. Vagal inputs are particularly important, bringing information from viscera, such as gut distention. Hormonal signals include leptin, insulin, cortisol, and gut peptides. Among the latter is ghrelin, which is made in the stomach and stimulates feeding, and peptide YY (PYY) and cholecystokinin, which is made in the small intestine and signal to the brain through direct action on hypothalamic control centers and/or via the vagus nerve. Metabolites, including glucose, can influence appetite, as seen by the effect of hypoglycemia to induce hunger; however, glucose is not normally a major regulator of appetite. These diverse hormonal, metabolic, and neural signals act by influencing the expression and release of various hypothalamic peptides [e.g., neuropeptide Y (NPY), Agouti-related peptide (AgRP), α-melanocyte-stimulating hormone (α-MSH), and melanin-concentrating hormone (MCH)] that are integrated with serotonergic, catecholaminergic, endocannabinoid, and opioid signaling pathways (see below). Psychological and cultural factors also play a role in the final expression of appetite. Apart from rare genetic syndromes involving leptin, its receptor, and the melanocortin system, specific defects in this complex appetite control network that influence common cases of obesity are not well defined.

Figure 77-2

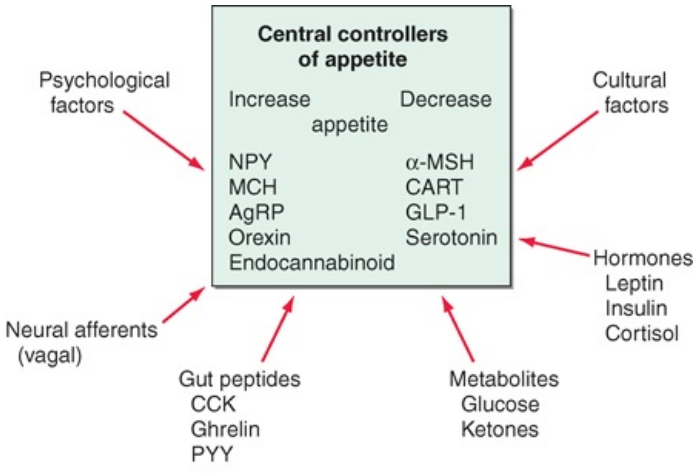

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The factors that regulate appetite through effects on central neural circuits. Some factors that increase or decrease appetite are listed. NPY, neuropeptide Y; MCH, melanin-concentrating hormone; AgRP, Agouti-related peptide; α-MSH, α-melanocyte-stimulating hormone; CART, cocaine- and amphetamine-related transcript; GLP-1, glucagon-related peptide-1; CCK, cholecystokinin.

Energy expenditure includes the following components: (1) resting or basal metabolic rate; (2) the energy cost of metabolizing and storing food; (3) the thermic effect of exercise; and (4) adaptive thermogenesis, which varies in response to long-term caloric intake (rising with increased intake). Basal metabolic rate accounts for ~70% of daily energy expenditure, whereas active physical activity

contributes 5–10%. Thus, a significant component of daily energy consumption is fixed.

Genetic models in mice indicate that mutations in certain genes (e.g., targeted deletion of the insulin receptor in adipose tissue) protect against obesity, apparently by increasing energy expenditure. Adaptive thermogenesis occurs in *brown adipose tissue* (BAT), which plays an important role in energy metabolism in many mammals. In contrast to white adipose tissue, which is used to store energy in the form of lipids, BAT expends stored energy as heat. A mitochondrial *uncoupling protein* (UCP-1) in BAT dissipates the hydrogen ion gradient in the oxidative respiration chain and releases energy as heat. The metabolic activity of BAT is increased by a central action of leptin, acting through the sympathetic nervous system that heavily innervates this tissue. In rodents, BAT deficiency causes obesity and diabetes; stimulation of BAT with a specific adrenergic agonist (β_3 agonist) protects against diabetes and obesity. BAT exists in humans (especially neonates), and although its physiologic role is not yet established, identification of functional BAT in many adults using PET imaging has increased interest in the implications of the tissue for pathogenesis and therapy of obesity.

The Adipocyte and Adipose Tissue

Adipose tissue is composed of the lipid-storing adipose cell and a stromal/vascular compartment in which cells including preadipocytes and macrophages reside. Adipose mass increases by enlargement of adipose cells through lipid deposition, as well as by an increase in the number of adipocytes. Obese adipose tissue is also characterized by increased numbers of infiltrating macrophages. The process by which adipose cells are derived from a mesenchymal preadipocyte involves an orchestrated series of differentiation steps mediated by a cascade of specific transcription factors. One of the key transcription factors is *peroxisome proliferator-activated receptor* γ (PPARγ), a nuclear receptor that binds the thiazolidinedione class of insulin-sensitizing drugs used in the treatment of type 2 diabetes (Chap. 344).

Although the adipocyte has generally been regarded as a storage depot for fat, it is also an endocrine cell that releases numerous molecules in a regulated fashion **(Fig. 77-3)**. These include the energy balance–regulating hormone leptin, cytokines such as tumor necrosis factor (TNF)-α and interleukin (IL)-6, complement factors such as factor D (also known as *adipsin*), prothrombotic agents such as plasminogen activator inhibitor I, and a component of the blood pressure–regulating system, angiotensinogen. Adiponectin, an abundant adipose-derived protein whose levels are reduced in obesity, enhances insulin sensitivity and lipid oxidation and it has vascular-protective effects, whereas resistin and RBP4, whose levels are increased in obesity, may induce insulin resistance. These factors, and others not yet identified, play a role in the physiology of lipid homeostasis, insulin sensitivity, blood pressure control, coagulation, and vascular health, and are likely to contribute to obesity-related pathologies.

Figure 77-3

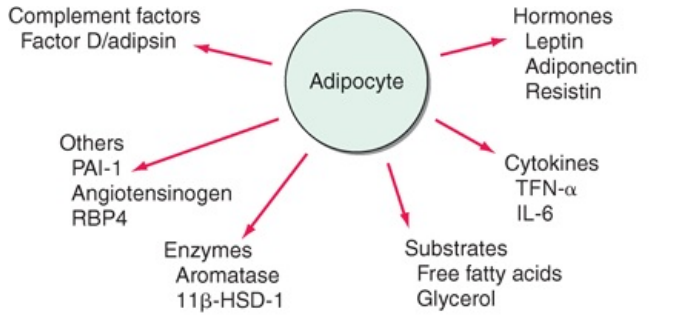

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Factors released by the adipocyte that can affect peripheral tissues. PAI, plasminogen activator inhibitor; TNF, tumor necrosis factor; RBP4, retinal binding protein 4.

Etiology of Obesity

Although the molecular pathways regulating energy balance are beginning to be illuminated, the causes of obesity remain elusive. In part, this reflects the fact that obesity is a heterogeneous group of disorders. At one level, the pathophysiology of obesity seems simple: a chronic excess of nutrient intake relative to the level of energy expenditure. However, due to the complexity of the neuroendocrine and metabolic systems that regulate energy intake, storage, and expenditure, it has been difficult to quantitate all the relevant parameters (e.g., food intake and energy expenditure) over time in human subjects.

ROLE OF GENES VERSUS ENVIRONMENT

Obesity is commonly seen in families, and the heritability of body weight is similar to that for height.

Inheritance is usually not Mendelian, however, and it is difficult to distinguish the role of genes and environmental factors. Adoptees more closely resemble their biologic than adoptive parents with respect to obesity, providing strong support for genetic influences. Likewise, identical twins have very similar BMIs whether reared together or apart, and their BMIs are much more strongly correlated than those of dizygotic twins. These genetic effects appear to relate to both energy intake and expenditure.

Whatever the role of genes, it is clear that the environment plays a key role in obesity, as evidenced by the fact that famine prevents obesity in even the most obesity-prone individual. In addition, the recent increase in the prevalence of obesity in the United States is far too rapid to be due to changes in the gene pool. Undoubtedly, genes influence the susceptibility to obesity in response to specific diets and availability of nutrition. Cultural factors are also important—these relate to both availability and composition of the diet and to changes in the level of physical activity. In industrial societies, obesity is more common among poor women, whereas in underdeveloped countries, wealthier women are more often obese. In children, obesity correlates to some degree with time spent watching television. Although the role of diet composition in obesity continues to generate controversy, it appears that high-fat diets may promote obesity when combined with diets rich in simple, rapidly absorbed carbohydrates.

Additional environmental factors may contribute to the increasing obesity prevalence. Both epidemiologic correlations and experimental data suggest that sleep deprivation leads to increased obesity. Changes in gut microbiome with capacity to alter energy balance are receiving experimental support from animal studies, and a possible role for obesigenic viral infections continues to receive sporadic attention.

SPECIFIC GENETIC SYNDROMES

For many years, obesity in rodents has been known to be caused by a number of distinct mutations distributed through the genome. Most of these single-gene mutations cause both hyperphagia and diminished energy expenditure, suggesting a physiologic link between these two parameters of energy homeostasis. Identification of the *ob* gene mutation in genetically obese (ob/ob) mice represented a major breakthrough in the field. The ob/ob mouse develops severe obesity, insulin resistance, and hyperphagia, as well as efficient metabolism (e.g., it gets fat even when ingesting the same number of calories as lean litter mates). The product of the *ob* gene is the peptide leptin, a name derived from the Greek root *leptos*, meaning thin. Leptin is secreted by adipose cells and acts primarily through the hypothalamus. Its level of production provides an index of adipose energy stores **(Fig. 77-4)**. High leptin levels decrease food intake and increase energy expenditure. Another mouse mutant, db/db, which is resistant to leptin, has a mutation in the leptin receptor and develops a similar syndrome. The *ob* gene is present in humans where it is also expressed in fat. Several families with morbid, early-onset obesity caused by inactivating mutations in either leptin or the leptin receptor have been described, thus demonstrating the biologic relevance of the leptin pathway in humans. Obesity in these individuals begins shortly after birth, is severe, and is accompanied by neuroendocrine abnormalities. The most prominent of these is hypogonadotropic hypogonadism, which is reversed by leptin replacement in the leptin-deficient subset. Central hypothyroidism and growth retardation are seen in the mouse model, but their occurrence in leptin-deficient humans is less clear. To date, there is no evidence that mutations in the leptin or leptin receptor genes play a prominent role in common forms of obesity.

Figure 77-4

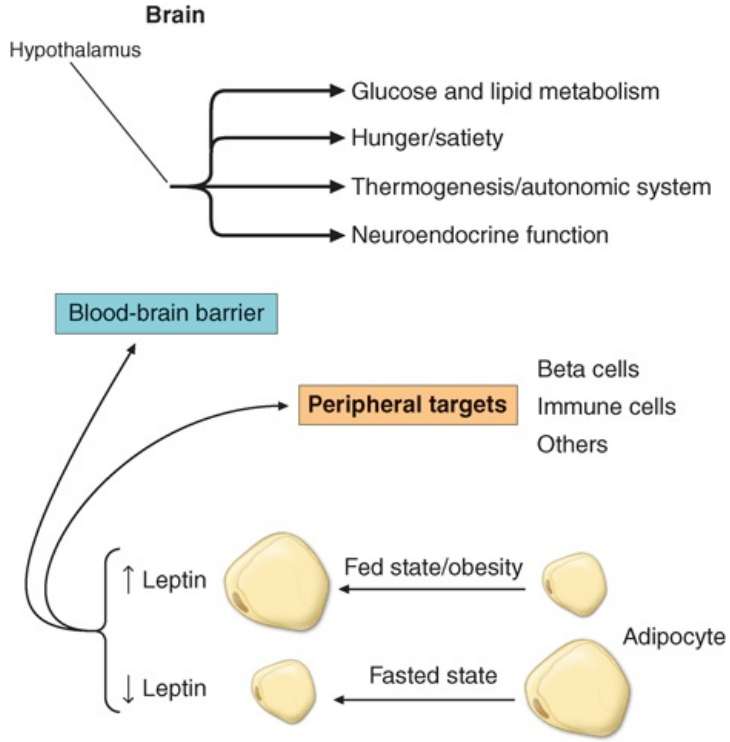

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The physiologic system regulated by leptin. Rising or falling leptin levels act through the hypothalamus to influence appetite, energy expenditure, and neuroendocrine function and through peripheral sites to influence systems such as the immune system.

Mutations in several other genes cause severe obesity in humans **(Table 77-1)**; each of these syndromes is rare. Mutations in the gene encoding proopiomelanocortin (POMC) cause severe obesity through failure to synthesize α-MSH, a key neuropeptide that inhibits appetite in the hypothalamus. The absence of POMC also causes secondary adrenal insufficiency due to absence of adrenocorticotropic hormone (ACTH), as well as pale skin and red hair due to absence of α-MSH. Proenzyme convertase 1 (PC-1) mutations are thought to cause obesity by preventing synthesis of α-MSH from its precursor peptide, POMC. α-MSH binds to the type 4 melanocortin receptor (MC4R), a key hypothalamic receptor that inhibits eating. Heterozygous loss-of-function mutations of this receptor account for as much as 5% of severe obesity. These five genetic defects define a pathway through which leptin (by stimulating POMC and increasing α-MSH) restricts food intake and limits weight **(Fig. 77-5)**. The results of genomewide association studies to identify genetic loci responsible for obesity in the general population have so far been disappointing. More than 10 replicated loci linked to obesity have been identified, but together they account for less than 3% of interindividual variation in BMI. The most replicated of these is a gene named FTO, which is of unknown function, but like many of the other recently described candidates, is expressed in the brain. Since the heritability of obesity is estimated to be 40–70%, it is likely that many more loci remain to be identified.

Table 77-1 Some Obesity Genes in Humans and Mice

Gene	Gene Product	Mechanism of Obesity	In Human	In Rodent
Lep (*ob*)	Leptin, a fat-derived hormone	Mutation prevents leptin from delivering satiety signal; brain perceives starvation	Yes	Yes
LepR (*db*)	Leptin receptor	Same as above	Yes	Yes
POMC	Proopiomelanocortin, a precursor of several hormones and neuropeptides	Mutation prevents synthesis of melanocyte-stimulating hormone (MSH), a satiety signal	Yes	Yes
MC4R	Type 4 receptor for MSH	Mutation prevents reception of satiety signal from MSH	Yes	Yes
AgRP	Agouti-related peptide, a neuropeptide expressed in the hypothalamus	Overexpression inhibits signal through *MC4R*	No	Yes

PC-1	Prohormone convertase 1, a processing enzyme	Mutation prevents synthesis of neuropeptide, probably MSH	Yes	No
Fat	Carboxypeptidase E, a processing enzyme	Same as above	No	Yes
Tub	Tub, a hypothalamic protein of unknown function	Hypothalamic dysfunction	No	Yes
TrkB	TrkB, a neurotrophin receptor	Hyperphagia due to uncharacterized hypothalamic defect	Yes	Yes

Figure 77-5

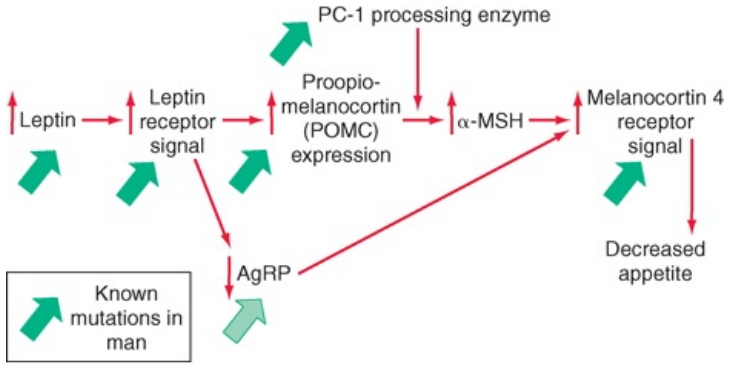

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A central pathway through which leptin acts to regulate appetite and body weight. Leptin signals through proopiomelanocortin (POMC) neurons in the hypothalamus to induce increased production of α-melanocyte-stimulating hormone (α-MSH), requiring the processing enzyme PC-1 (proenzyme convertase 1). α-MSH acts as an agonist on melanocortin-4 receptors to inhibit appetite, and the neuropeptide AgRp (Agouti-related peptide) acts as an antagonist of this receptor. Mutations that cause obesity in humans are indicated by the solid green arrows.

In addition to these human obesity genes, studies in rodents reveal several other molecular candidates for hypothalamic mediators of human obesity or leanness. The *tub* gene encodes a hypothalamic peptide of unknown function; mutation of this gene causes late-onset obesity. The *fat* gene encodes carboxypeptidase E, a peptide-processing enzyme; mutation of this gene is thought to cause obesity by disrupting production of one or more neuropeptides. AgRP is coexpressed with NPY in arcuate nucleus neurons. AgRP antagonizes α-MSH action at MC4 receptors, and its overexpression induces obesity. In contrast, a mouse deficient in the peptide MCH, whose administration causes feeding, is lean.

A number of complex human syndromes with defined inheritance are associated with obesity **(Table 77-2)**. Although specific genes have limited definition at present, their identification will likely enhance our understanding of more common forms of human obesity. In the Prader-Willi syndrome, a multigenic neurodevelopmental disorder, obesity coexists with short stature, mental retardation, hypogonadotropic hypogonadism, hypotonia, small hands and feet, fish-shaped mouth, and hyperphagia. Most patients have a deletion in the 15q11-13 chromosomal region, and reduced expression of the signaling protein necdin may be an important cause of defective hypothalamic neural development in this disorder (Chap. 62). Bardet-Biedl syndrome (BBS) is a genetically heterogeneous disorder characterized by obesity, mental retardation, retinitis pigmentosa, diabetes, renal and cardiac malformations, polydactyly, and hypogonadotropic hypogonadism. At least 12 genetic loci have been identified, and most of the encoded proteins form two multiprotein complexes that are involved in ciliary function and microtubule-based intracellular transport. Recent evidence suggests that mutations might disrupt leptin receptor trafficking in key hypothalamic neurons, causing leptin resistance.

Table 77-2 A Comparison of Syndromes of Obesity—Hypogonadism and Mental Retardation

	Syndrome				
Feature	**Prader-Willi**	**Laurence-Moon-Biedl**	**Ahlstrom's**	**Cohen's**	**Carpenter's**
Inheritance	Sporadic; two-thirds have defect	Autosomal recessive	Autosomal recessive	Probably autosomal recessive	Autosomal recessive
Stature	Short	Normal; infrequently	Normal; infrequently	Short or tall	Normal

		short	short		
Obesity	Generalized	Generalized	Truncal	Truncal	Truncal, gluteal
	Moderate to severe	Early onset, 1–2 years	Early onset, 2–5 years	Mid-childhood, age 5	
	Onset 1–3 years				
Craniofacies	Narrow bifrontal diameter	Not distinctive	Not distinctive	High nasal bridge	Acrocephaly
	Almond-shaped eyes			Arched palate	Flat nasal bridge
	Strabismus			Open mouth	High-arched palate
	V-shaped mouth			Short philtrum	
	High-arched palate				
Limbs	Small hands and feet	Polydactyly	No abnormalities	Hypotonia	Polydactyly
	Hypotonia			Narrow hands and feet	Syndactyly
					Genu valgum
Reproductive status	1° Hypogonadism	1° Hypogonadism	Hypogonadism in males but not in females	Normal gonadal function or hypogonadotrophic hypogonadism	2° Hypogonadism
Other features	Enamel hypoplasia			Dysplastic ears	
	Hyperphagia			Delayed puberty	
	Temper tantrums				
	Nasal speech				
Mental retardation	Mild to moderate		Normal intelligence	Mild	Slight

OTHER SPECIFIC SYNDROMES ASSOCIATED WITH OBESITY

Cushing's syndrome

Although obese patients commonly have central obesity, hypertension, and glucose intolerance, they lack other specific stigmata of Cushing's syndrome (Chap. 342). Nonetheless, a potential diagnosis of Cushing's syndrome is often entertained. Cortisol production and urinary metabolites (17OH steroids) may be increased in simple obesity. Unlike in Cushing's syndrome, however, cortisol levels in blood and urine in the basal state and in response to corticotropin-releasing hormone (CRH) or ACTH are normal; the overnight 1-mg dexamethasone suppression test is normal in 90%, with the remainder being normal on a standard 2-day low-dose dexamethasone suppression test. Obesity may be associated with excessive local reactivation of cortisol in fat by 11β-hydroxysteroid dehydrogenase 1, an enzyme that converts inactive cortisone to cortisol.

Hypothyroidism

The possibility of hypothyroidism should be considered, but it is an uncommon cause of obesity; hypothyroidism is easily ruled out by measuring thyroid-stimulating hormone (TSH). Much of the weight gain that occurs in hypothyroidism is due to myxedema (Chap. 341).

Insulinoma

Patients with insulinoma often gain weight as a result of overeating to avoid hypoglycemic symptoms (Chap. 345). The increased substrate plus high insulin levels promote energy storage in fat. This can be marked in some individuals but is modest in most.

Craniopharyngioma and other disorders involving the hypothalamus

Whether through tumors, trauma, or inflammation, hypothalamic dysfunction of systems controlling satiety, hunger, and energy expenditure can cause varying degrees of obesity (Chap. 339). It is uncommon to identify a discrete anatomic basis for these disorders. Subtle hypothalamic dysfunction is probably a more common cause of obesity than can be documented using currently available imaging techniques. Growth hormone (GH), which exerts lipolytic activity, is diminished in obesity and is

increased with weight loss. Despite low GH levels, insulin-like growth factor (IGF)-I (somatomedin) production is normal, suggesting that GH suppression is a compensatory response to increased nutritional supply.

PATHOGENESIS OF COMMON OBESITY

Obesity can result from increased energy intake, decreased energy expenditure, or a combination of the two. Thus, identifying the etiology of obesity should involve measurements of both parameters. However, it is difficult to perform direct and accurate measurements of energy intake in free-living individuals; and the obese, in particular, often underreport intake. Measurements of chronic energy expenditure are possible using doubly labeled water or metabolic chamber/rooms. In subjects at stable weight and body composition, energy intake equals expenditure. Consequently, these techniques allow assessment of energy intake in free-living individuals. The level of energy expenditure differs in established obesity, during periods of weight gain or loss, and in the pre- or postobese state. Studies that fail to take note of this phenomenon are not easily interpreted.

There is continued interest in the concept of a body weight "set point." This idea is supported by physiologic mechanisms centered around a sensing system in adipose tissue that reflects fat stores and a receptor, or "adipostat," that is in the hypothalamic centers. When fat stores are depleted, the adipostat signal is low, and the hypothalamus responds by stimulating hunger and decreasing energy expenditure to conserve energy. Conversely, when fat stores are abundant, the signal is increased, and the hypothalamus responds by decreasing hunger and increasing energy expenditure. The recent discovery of the *ob* gene, and its product leptin, and the *db* gene, whose product is the leptin receptor, provides important elements of a molecular basis for this physiologic concept (see above).

WHAT IS THE STATUS OF FOOD INTAKE IN OBESITY? (DO THE OBESE EAT MORE THAN THE LEAN?)

This question has stimulated much debate, due in part to the methodologic difficulties inherent in determining food intake. Many obese individuals believe that they eat small quantities of food, and this claim has often been supported by the results of food intake questionnaires. However, it is now established that average energy expenditure increases as individuals get more obese, due primarily to the fact that metabolically active lean tissue mass increases with obesity. Given the laws of thermodynamics, the obese person must therefore eat more than the average lean person to maintain their increased weight. It may be the case, however, that a subset of individuals who are predisposed to obesity have the capacity to become obese initially without an absolute increase in caloric consumption.

WHAT IS THE STATE OF ENERGY EXPENDITURE IN OBESITY?

The average total daily energy expenditure is higher in obese than lean individuals when measured at stable weight. However, energy expenditure falls as weight is lost, due in part to loss of lean body mass and to decreased sympathetic nerve activity. When reduced to near-normal weight and maintained there for awhile, (some) obese individuals have lower energy expenditure than (some) lean individuals. There is also a tendency for those who will develop obesity as infants or children to have lower resting energy expenditure rates than those who remain lean.

The physiologic basis for variable rates of energy expenditure (at a given body weight and level of energy intake) is essentially unknown. A mutation in the human β_3-adrenergic receptor may be associated with increased risk of obesity and/or insulin resistance in certain (but not all) populations.

One recently described component of thermogenesis, called *nonexercise activity thermogenesis* (NEAT), has been linked to obesity. It is the thermogenesis that accompanies physical activities other than volitional exercise such as the activities of daily living, fidgeting, spontaneous muscle contraction, and maintaining posture. NEAT accounts for about two-thirds of the increased daily energy expenditure induced by overfeeding. The wide variation in fat storage seen in overfed individuals is predicted by the degree to which NEAT is induced. The molecular basis for NEAT and its regulation is unknown.

LEPTIN IN TYPICAL OBESITY

The vast majority of obese persons have increased leptin levels but do not have mutations of either leptin or its receptor. They appear, therefore, to have a form of functional "leptin resistance." Data suggesting that some individuals produce less leptin per unit fat mass than others or have a form of relative leptin deficiency that predisposes to obesity are at present contradictory and unsettled. The mechanism for leptin resistance, and whether it can be overcome by raising leptin levels or combining leptin with other treatments in a subset of obese individuals, is not yet established. Some data suggest that leptin may not effectively cross the blood-brain barrier as levels rise. It is also apparent from animal studies that leptin signaling inhibitors, such as SOCS3 and PTP1b, are involved in the leptin-resistant state.

Pathologic Consequences of Obesity

(See also Chap. 78) Obesity has major adverse effects on health. Obesity is associated with an

increase in mortality, with a 50–100% increased risk of death from all causes compared to normal-weight individuals, mostly due to cardiovascular causes. Obesity and overweight together are the second leading cause of preventable death in the United States, accounting for 300,000 deaths per year. Mortality rates rise as obesity increases, particularly when obesity is associated with increased intraabdominal fat (see above). Life expectancy of a moderately obese individual could be shortened by 2–5 years, and a 20- to 30-year-old male with a BMI >45 may lose 13 years of life. It is also apparent that the degree to which obesity affects particular organ systems is influenced by susceptibility genes that vary in the population.

INSULIN RESISTANCE AND TYPE 2 DIABETES MELLITUS

Hyperinsulinemia and insulin resistance are pervasive features of obesity, increasing with weight gain and diminishing with weight loss (Chap. 242). Insulin resistance is more strongly linked to intraabdominal fat than to fat in other depots. Molecular links between obesity and insulin resistance in fat, muscle, and liver have been sought for many years. Major factors include: (1) insulin itself, by inducing receptor downregulation; (2) free fatty acids that are increased and capable of impairing insulin action; (3) intracellular lipid accumulation; and (4) several circulating peptides produced by adipocytes, including the cytokinesTNF-α and IL-6, RBP4, and the "adipokines" adiponectin and resistin that have altered expression in obese adipocytes, and can modify insulin action. Additional mechanisms are obesity-linked inflammation, including infiltration of macrophages into tissues including fat, and induction of the endoplasmic reticulum stress response, that can bring about resistance to insulin action in cells. Despite the prevalence of insulin resistance, most obese individuals do not develop diabetes, suggesting that diabetes requires an interaction between obesity-induced insulin resistance and other factors such as impaired insulin secretion (Chap. 344). Obesity, however, is a major risk factor for diabetes, and as many as 80% of patients with type 2 diabetes mellitus are obese. Weight loss and exercise, even of modest degree, increase insulin sensitivity and often improve glucose control in diabetes.

REPRODUCTIVE DISORDERS

Disorders that affect the reproductive axis are associated with obesity in both men and women. Male hypogonadism is associated with increased adipose tissue, often distributed in a pattern more typical of females. In men whose weight is >160% ideal body weight (IBW), plasma testosterone and sex hormone–binding globulin (SHBG) are often reduced, and estrogen levels (derived from conversion of adrenal androgens in adipose tissue) are increased (Chap. 346). Gynecomastia may be seen. However, masculinization, libido, potency, and spermatogenesis are preserved in most of these individuals. Free testosterone may be decreased in morbidly obese men whose weight is >200% IBW.

Obesity has long been associated with menstrual abnormalities in women, particularly in women with upper body obesity (Chap. 347). Common findings are increased androgen production, decreased SHBG, and increased peripheral conversion of androgen to estrogen. Most obese women with oligomenorrhea have the polycystic ovarian syndrome (PCOS), with its associated anovulation and ovarian hyperandrogenism; 40% of women with PCOS are obese. Most nonobese women with PCOS are also insulin-resistant, suggesting that insulin resistance, hyperinsulinemia, or the combination of the two are causative or contribute to the ovarian pathophysiology in PCOS in both obese and lean individuals. In obese women with PCOS, weight loss or treatment with insulin-sensitizing drugs often restores normal menses. The increased conversion of androstenedione to estrogen, which occurs to a greater degree in women with lower body obesity, may contribute to the increased incidence of uterine cancer in postmenopausal women with obesity.

CARDIOVASCULAR DISEASE

The Framingham Study revealed that obesity was an independent risk factor for the 26-year incidence of cardiovascular disease in men and women [including coronary disease, stroke, and congestive heart failure (CHF)]. The waist-to-hip ratio may be the best predictor of these risks. When the additional effects of hypertension and glucose intolerance associated with obesity are included, the adverse impact of obesity is even more evident. The effect of obesity on cardiovascular mortality in women may be seen at BMIs as low as 25. Obesity, especially abdominal obesity, is associated with an atherogenic lipid profile; with increased low-density lipoprotein cholesterol, very low density lipoprotein, and triglyceride; and with decreased high density lipoprotein cholesterol and decreased levels of the vascular protective adipokine adiponectin (Chap. 356). Obesity is also associated with hypertension. Measurement of blood pressure in the obese requires use of a larger cuff size to avoid artifactual increases. Obesity-induced hypertension is associated with increased peripheral resistance and cardiac output, increased sympathetic nervous system tone, increased salt sensitivity, and insulin-mediated salt retention; it is often responsive to modest weight loss.

PULMONARY DISEASE

Obesity may be associated with a number of pulmonary abnormalities. These include reduced chest wall compliance, increased work of breathing, increased minute ventilation due to increased metabolic rate, and decreased functional residual capacity and expiratory reserve volume (Chap. 252). Severe obesity may be associated with obstructive sleep apnea and the "obesity hypoventilation syndrome"

with attenuated hypoxic and hypercapnic ventilatory responses (Chap. 264). Sleep apnea can be obstructive (most common), central, or mixed and is associated with hypertension. Weight loss (10–20 kg) can bring substantial improvement, as can major weight loss following gastric bypass or restrictive surgery. Continuous positive airway pressure has been used with some success.

HEPATOBILIARY DISEASE

Obesity is frequently associated with the common disorder nonalcoholic fatty liver disease (NAFLD). This hepatic fatty infiltration of NAFLD can progress in a subset to inflammatory nonalcoholic steatohepatitis (NASH) and more rarely to cirrhosis and hepatocellular carcinoma. Steatosis has been noted to improve following weight loss, secondary to diet or bariatric surgery. The mechanism for the association remains unclear. Obesity is associated with enhanced biliary secretion of cholesterol, supersaturation of bile, and a higher incidence of gallstones, particularly cholesterol gallstones (Chap. 311). A person 50% above IBW has about a sixfold increased incidence of symptomatic gallstones. Paradoxically, fasting increases supersaturation of bile by decreasing the phospholipid component. Fasting-induced cholecystitis is a complication of extreme diets.

CANCER

Obesity in males is associated with higher mortality from cancer, including cancer of the esophagus, colon, rectum, pancreas, liver, and prostate; obesity in females is associated with higher mortality from cancer of the gallbladder, bile ducts, breasts, endometrium, cervix, and ovaries. Some of the latter may be due to increased rates of conversion of androstenedione to estrone in adipose tissue of obese individuals. Other possible mechanistic links are other hormones whose levels are linked to nutritional state, including insulin, leptin, adiponectin, and IGF-1. It has been estimated that obesity accounts for 14% of cancer deaths in men and 20% in women in the United States.

BONE, JOINT, AND CUTANEOUS DISEASE

Obesity is associated with an increased risk of osteoarthritis, no doubt partly due to the trauma of added weight bearing, but potentially linked as well to activation of inflammatory pathways that could promote synovial pathology. The prevalence of gout may also be increased (Chap. 333). Among the skin problems associated with obesity is acanthosis nigricans, manifested by darkening and thickening of the skinfolds on the neck, elbows, and dorsal interphalangeal spaces. Acanthosis reflects the severity of underlying insulin resistance and diminishes with weight loss. Friability of skin may be increased, especially in skinfolds, enhancing the risk of fungal and yeast infections. Finally, venous stasis is increased in the obese.

FURTHER READINGS

Bochukova EG et al: Large, rare chromosomal deletions associated with severe early-onset obesity. Nature 463:666, 2010[PMID: 19966786] [Full Text]

Cecil JE et al: An obesity-associated FTO gene variant and increased energy intake in children. N Engl J Med 359:2558, 2008[PMID: 19073975] [Full Text]

Fisher FM et al: Obesity is an FGF21 resistant state. Diabetes 59:2781, 2010[PMID: 20682689] [Full Text]

Froy O: Metabolism and circadian rhythms—Implications for obesity. Endocr Rev 31:1, 2010[PMID: 19854863] [Full Text]

Misra A, Khurana L: Obesity and the metabolic syndrome in developing countries. J Clin Endocrinol Metab 93:S9, 2008

Morton GJ et al: Central nervous system control of food intake and body weight. Nature 443:289, 2006[PMID: 16988703] [Full Text]

Virtanen KA et al: Functional brown adipose tissue in healthy adults. N Engl J Med 360:1518, 2009[PMID: 19357407] [Full Text]

Walters RG et al: A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature 463:671, 2010[PMID: 20130649] [Full Text]

Zobel DP et al: Variants near MC4R are associated with obesity and influence obesity-related quantitative traits in a population of middle-aged people: Studies of 14,940 Danes. Diabetes 58:757, 2009[PMID: 19073769] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EVALUATION AND MANAGEMENT OF OBESITY: INTRODUCTION

Over 66% of U.S. adults are categorized as overweight or obese, and the prevalence of obesity is increasing rapidly in most of the industrialized world. Children and adolescents also are becoming more obese, indicating that the current trends will accelerate over time. Obesity is associated with an increased risk of multiple health problems, including hypertension, Type 2 diabetes, dyslipidemia, degenerative joint disease, and some malignancies. Thus, it is important for physicians to identify, evaluate, and treat patients for obesity and associated comorbid conditions.

Evaluation

Physicians should screen all adult patients for obesity and offer intensive counseling and behavioral interventions to promote sustained weight loss. The five main steps in the evaluation of obesity, as described below, are (1) focused obesity-related history, (2) physical examination to determine the degree and type of obesity, (3) comorbid conditions, (4) fitness level, and (5) the patient's readiness to adopt lifestyle changes.

THE OBESITY-FOCUSED HISTORY

Information from the history should address the following six questions:

- What factors contribute to the patient's obesity?
- How is the obesity affecting the patient's health?
- What is the patient's level of risk from obesity?
- What are the patient's goals and expectations?
- Is the patient motivated to begin a weight management program?
- What kind of help does the patient need?

Although the vast majority of cases of obesity can be attributed to behavioral features that affect diet and physical activity patterns, the history may suggest secondary causes that merit further evaluation. Disorders to consider include polycystic ovarian syndrome, hypothyroidism, Cushing's syndrome, and hypothalamic disease. Drug-induced weight gain also should be considered. Common causes include medications for diabetes (insulin, sulfonylureas, thiazolidinediones); steroid hormones; psychotropic agents; mood stabilizers (lithium); antidepressants (tricyclics, monoamine oxidase inhibitors, paroxetine, mirtazapine); and antiepileptic drugs (valproate, gabapentin, carbamazepine). Other medications, such as nonsteroidal anti-inflammatory drugs and calcium channel blockers, may cause peripheral edema but do not increase body fat.

The patient's current diet and physical activity patterns may reveal factors that contribute to the development of obesity in addition to identifying behaviors to target for treatment. This type of historic information is best obtained by using a questionnaire in combination with an interview.

BMI AND WAIST CIRCUMFERENCE

Three key anthropometric measurements are important to evaluate the degree of obesity: weight, height, and waist circumference. The body mass index (BMI), calculated as weight (kg)/height $(m)^2$, or weight (lbs)/height $(inches)^2$ x 703, is used to classify weight status and risk of disease **(Tables 78-1** and **78-2)**. BMI is used since it provides an estimate of body fat and is related to risk of disease. Lower BMI thresholds for overweight and obesity have been proposed for the Asia-Pacific region since this population appears to be at risk for glucose and lipid abnormalities at lower body weights.

Table 78-1 Body Mass Index (BMI) Table

BMI	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**
Height, inches	**Body Weight, pounds**																
58	91	96	100	105	110	115	119	124	129	134	138	143	148	153	158	162	167
59	94	99	104	109	114	119	124	128	133	138	143	148	153	158	163	168	173
60	97	102	107	112	118	123	128	133	138	143	148	153	158	163	168	174	179

61	100	106	111	116	122	127	132	137	143	148	153	158	164	169	174	180	185	
62	104	109	115	120	126	131	136	142	147	153	158	164	169	175	180	186	191	
63	107	113	118	124	130	135	141	146	152	158	163	169	175	180	186	191	197	
64	110	116	122	128	134	140	145	151	157	163	169	174	180	186	192	197	204	
65	114	120	126	132	138	144	150	156	162	168	174	180	186	192	198	204	210	
66	118	124	130	136	142	148	155	161	167	173	179	186	192	198	204	210	216	
67	121	127	134	140	146	153	159	166	172	178	185	191	198	204	211	217	223	
68	125	131	138	144	151	158	164	171	177	184	190	197	203	210	216	223	230	
69	128	135	142	149	155	162	169	176	182	189	196	203	209	216	223	230	236	
70	132	139	146	153	160	167	174	181	188	195	202	209	216	222	229	236	243	
71	136	143	150	157	165	172	179	186	193	200	208	215	222	229	236	243	250	
72	140	147	154	162	169	177	184	191	199	206	213	221	228	235	242	250	258	
73	144	151	159	166	174	182	189	197	204	212	219	227	235	242	250	257	265	
74	148	155	163	171	179	186	194	202	210	218	225	233	241	249	256	264	272	
75	152	160	168	176	184	192	200	208	216	224	232	240	248	256	264	272	279	
76	156	164	172	180	189	197	205	213	221	230	238	246	254	263	271	279	287	
BMI	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**
58	172	177	181	186	191	196	201	205	210	215	220	224	229	234	239	244	248	253	258
59	178	183	188	193	198	203	208	212	217	222	227	232	237	242	247	252	257	262	267
60	184	189	194	199	204	209	215	220	225	230	235	240	245	250	255	261	266	271	276
61	190	195	201	206	211	217	222	227	232	238	243	248	254	259	264	269	275	280	285
62	196	202	207	213	218	224	229	235	240	246	251	256	262	267	273	278	284	289	295
63	203	208	214	220	225	231	237	242	248	254	259	265	270	278	282	287	293	299	304
64	209	215	221	227	232	238	244	250	256	262	267	273	279	285	291	296	302	308	314
65	216	222	228	234	240	246	252	258	264	270	276	282	288	294	300	306	312	318	324
66	223	229	235	241	247	253	260	266	272	278	284	291	297	303	309	315	322	328	334
67	230	236	242	249	255	261	268	274	280	287	293	299	306	312	319	325	331	338	344
68	236	243	249	256	262	269	276	282	289	295	302	308	315	322	328	335	341	348	354
69	243	250	257	263	270	277	284	291	297	304	311	318	324	331	338	345	351	358	365
70	250	257	264	271	278	285	292	299	306	313	320	327	334	341	348	355	362	369	376
71	257	265	272	279	286	293	301	308	315	322	329	338	343	351	358	365	372	379	386
72	265	272	279	287	294	302	309	316	324	331	338	346	353	361	368	375	383	390	397
73	272	280	288	295	302	310	318	325	333	340	348	355	363	371	378	386	393	401	408
74	280	287	295	303	311	319	326	334	342	350	358	365	373	381	389	396	404	412	420
75	287	295	303	311	319	327	335	343	351	359	367	375	383	391	399	407	415	423	431
76	295	304	312	320	328	336	344	353	361	369	377	385	394	402	410	418	426	435	443

Table 78-2 Classification of Weight Status and Risk of Disease

	BMI (kg/m^2)	Obesity Class	Risk of Disease
Underweight	<18.5		
Healthy weight	18.5–24.9		
Overweight	25.0–29.9		Increased
Obesity	30.0–34.9	I	High
Obesity	35.0–39.9	II	Very high
Extreme Obesity	≥40	III	Extremely high

Source: Adapted from National Institutes of Health, National Heart, Lung, and Blood Institute: *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults*. U.S. Department of Health and Human Services, Public Health Service, 1998.

Excess abdominal fat, assessed by measurement of waist circumference or waist-to-hip ratio, is independently associated with higher risk for diabetes mellitus and cardiovascular disease. Measurement of the waist circumference is a surrogate for visceral adipose tissue and should be

performed in the horizontal plane above the iliac crest **(Table 78-3)**.

Table 78-3 Ethnic-Specific Values for Waist Circumference

Ethnic Group	Waist Circumference
Europeans	
Men	>94 cm (37 in)
Women	>80 cm (31.5 in)
South Asians and Chinese	
Men	>90 cm (35 in)
Women	>80 cm (31.5 in)
Japanese	
Men	>85 cm (33.5 in)
Women	>90 cm (35 in)
Ethnic South and Central Americans	Use south Asian recommendations until more specific data are available.
Sub-Saharan Africans	Use European data until more specific data are available.
Eastern Mediterranean and Middle East (Arab) populations	Use European data until more specific data are available.

Source: From KGMM Alberti et al for the IDF Epidemiology Task Force Consensus Group: Lancet 366:1059, 2005.

PHYSICAL FITNESS

Several prospective studies have demonstrated that physical fitness, reported by questionnaire or measured by a maximal treadmill exercise test, is an important predictor of all-cause mortality rate independent of BMI and body composition. These observations highlight the importance of taking an exercise history during examination as well as emphasizing physical activity as a treatment approach.

OBESITY-ASSOCIATED COMORBID CONDITIONS

The evaluation of comorbid conditions should be based on presentation of symptoms, risk factors, and index of suspicion. All patients should have a fasting lipid panel [total, low-density lipoprotein (LDL), and high-density lipoprotein (HDL) cholesterol and triglyceride levels] and fasting blood glucose along with blood pressure determination. Symptoms and diseases that are directly or indirectly related to obesity are listed in **Table 78-4**. Although individuals vary, the number and severity of organ-specific comorbid conditions usually rise with increasing levels of obesity. Patients at very high absolute risk include those with the following: established coronary heart disease; presence of other atherosclerotic diseases, such as peripheral arterial disease, abdominal aortic aneurysm, and symptomatic carotid artery disease; Type 2 diabetes; and sleep apnea.

Table 78-4 Obesity-Related Organ Systems Review

Cardiovascular	Respiratory
Hypertension	Dyspnea
Congestive heart failure	Obstructive sleep apnea
Cor pulmonale	Hypoventilation syndrome
Varicose veins	Pickwickian syndrome
Pulmonary embolism	Asthma
Coronary artery disease	**Gastrointestinal**
Endocrine	Gastroesophageal reflux disease
Metabolic syndrome	Nonalcoholic fatty liver disease
Type 2 diabetes	Cholelithiasis
Dyslipidemia	Hernias
Polycystic ovarian syndrome	Colon cancer
Musculoskeletal	**Genitourinary**
Hyperuricemia and gout	Urinary stress incontinence
Immobility	Obesity-related glomerulopathy
Osteoarthritis (knees and hips)	Hypogonadism (male)
Low back pain	Breast and uterine cancer

Carpal tunnel syndrome	Pregnancy complications
Psychological	**Neurologic**
Depression/low self-esteem	Stroke
Body image disturbance	Idiopathicintracranial hypertension
Social stigmatization	Meralgia paresthetica
Integument	Dementia
Striae distensae	
Stasis pigmentation of legs	
Lymphedema	
Cellulitis	
Intertrigo, carbuncles	
Acanthosis nigricans Acrochordon (skin tags)	
Hidradenitis suppurativa	

ASSESSING THE PATIENT'S READINESS TO CHANGE

An attempt to initiate lifestyle changes when the patient is not ready usually leads to frustration and may hamper future weight-loss efforts. Assessment includes patient motivation and support, stressful life events, psychiatric status, time availability and constraints, and appropriateness of goals and expectations. Readiness can be viewed as the balance of two opposing forces: (1) motivation, or the patient's desire to change, and (2) resistance, or the patient's resistance to change.

A helpful method to begin a readiness assessment is to "anchor" the patient's interest and confidence to change on a numerical scale. With this technique, the patient is asked to rate his or her level of interest and confidence on a scale from 0 to 10, with 0 being not so important (or confident) and 10 being very important (or confident) to lose weight at this time. This exercise helps establish readiness to change and also serves as a basis for further dialogue.

Treatment: Obesity

THE GOAL OF THERAPY

The primary goal of treatment is to improve obesity-related comorbid conditions and reduce the risk of developing future comorbidities. Information obtained from the history, physical examination, and diagnostic tests is used to determine risk and develop a treatment plan **(Fig. 78-1)**. The decision of how aggressively to treat the patient and which modalities to use is determined by the patient's risk status, expectations, and available resources. Therapy for obesity always begins with lifestyle management and may include pharmacotherapy or surgery, depending on BMI risk category **(Table 78-5)**. Setting an initial weight-loss goal of 10% over 6 months is a realistic target.

Table 78-5 A Guide to Selecting Treatment

	BMI Category				
Treatment	**25–26.9**	**27–29.9**	**30–35**	**35–39.9**	**≥40**
Diet, exercise, behavior therapy	With comorbidities	With comorbidities	+	+	+
Pharmacotherapy		With comorbidities	+	+	+
Surgery				With comorbidities	+

Source: From National Heart, Lung, and Blood Institute, North American Association for the Study of Obesity (2000).

Figure 78-1

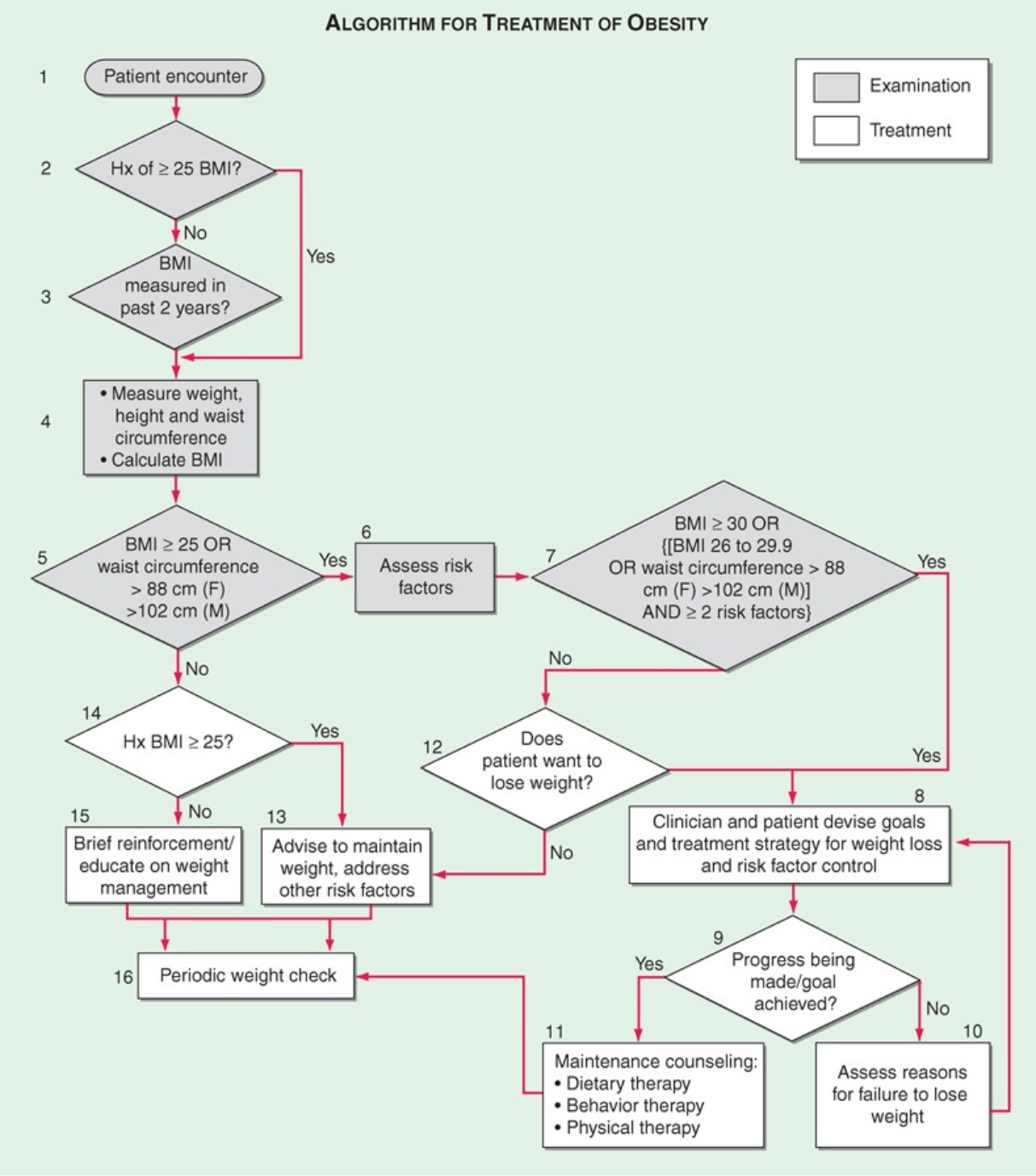

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment algorithm. This algorithm applies only to the assessment for overweight and obesity and subsequent decisions on that assessment. It does not reflect any initial overall assessment for other conditions that the physician may wish to perform. BMI, body mass index; Ht, height; Hx, history; Wt, weight. *(From National, Heart, Lung, and Blood Institute: Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: The evidence report. Washington, DC, US Department of Health and Human Services, 1998.)*

LIFESTYLE MANAGEMENT

Obesity care involves attention to three essential elements of lifestyle: dietary habits, physical activity, and behavior modification. Because obesity is fundamentally a disease of energy imbalance, all patients must learn how and when energy is consumed (diet), how and when energy is expended (physical activity), and how to incorporate this information into their daily lives (behavior therapy). Lifestyle management has been shown to result in a modest (typically 3–5 kg) weight loss compared with no treatment or usual care.

Diet Therapy

The primary focus of diet therapy is to reduce overall calorie consumption. The National Heart, Lung, and Blood Institute (NHLBI) guidelines recommend initiating treatment with a calorie deficit of 500

–1000 kcal/d compared with the patient's habitual diet. This reduction is consistent with a goal of losing approximately 1–2 lb per week. This calorie deficit can be accomplished by suggesting substitutions or alternatives to the diet. Examples include choosing smaller portion sizes, eating more fruits and vegetables, consuming more whole-grain cereals, selecting leaner cuts of meat and skimmed dairy products, reducing fried foods and other added fats and oils, and drinking water instead of caloric beverages. It is important that the dietary counseling remain patient-centered and that the goals be practical, realistic, and achievable.

The macronutrient composition of the diet will vary with the patient's preference and medical condition. The 2005 U.S. Department of Agriculture Dietary Guidelines for Americans (Chap. 73), which focus on health promotion and risk reduction, can be applied to treatment of overweight or obese patients. The recommendations include maintaining a diet rich in whole grains, fruits, vegetables, and dietary fiber; consuming two servings (8 oz) of fish high in omega 3 fatty acids per week; decreasing sodium to <2300 mg/d; consuming 3 cups of milk (or equivalent low-fat or fat-free dairy products) per day; limiting cholesterol to <300 mg/d; and keeping total fat between 20 and 35% of daily calories and saturated fats to <10% of daily calories. Application of these guidelines to specific calorie goals can be found on the websitewww.mypyramid.gov. The revised Dietary Reference Intakes for Macronutrients released by the Institute of Medicine recommends 45–65% of calories from carbohydrates, 20–35% from fat, and 10–35% from protein. The guidelines also recommend daily fiber intake of 38 g (men) and 25 g (women) for persons over 50 years of age and 30 g (men) and 21 g (women) for those under age 50.

Since portion control is one of the most difficult strategies for patients to manage, the use of pre-prepared products such as meal replacements is a simple and convenient suggestion. Examples include frozen entrees, canned beverages, and bars. Use of meal replacements in the diet has been shown to result in a 7–8% weight loss.

An ongoing area of investigation is the use of low-carbohydrate, high-protein diets for weight loss. These diets are based on the concept that carbohydrates are the primary cause of obesity and lead to insulin resistance. Most low-carbohydrate diets (e.g., South Beach, Zone, and Sugar Busters!) recommend a carbohydrate level of approximately 40–46% of energy. The Atkins diet contains 5–15% carbohydrate, depending on the phase of the diet. Low-carbohydrate, high-protein diets appear to be more effective in lowering BMI; improving coronary heart disease risk factors, including an increase in HDL cholesterol and a decrease in triglyceride levels; and controlling satiety in the short term compared with low-fat diets. However, after 12 months, there is no significance difference among diets. Multiple studies have shown that sustained adherence to the diet rather than diet type is likely to be the best predictor of weight-loss outcome.

Another dietary approach to consider is the concept of energy density, which refers to the number of calories (energy) a food contains per unit of weight. People tend to ingest a constant volume of food regardless of caloric or macronutrient content. Adding water or fiber to a food decreases its energy density by increasing weight without affecting caloric content. Examples of foods with low-energy density include soups, fruits, vegetables, oatmeal, and lean meats. Dry foods and high-fat foods such as pretzels, cheese, egg yolks, potato chips, and red meat have a high-energy density. Diets containing low-energy dense foods have been shown to control hunger and result in decreased caloric intake and weight loss.

Occasionally, very low calorie diets (VLCDs) are prescribed as a form of aggressive dietary therapy. The primary purpose of a VLCD is to promote a rapid and significant (13–23 kg) short-term weight loss over a 3- to 6-month period. These propriety formulas typically supply ≤800 kcal, 50–80 g protein, and 100% of the recommended daily intake for vitamins and minerals. According to a review by the National Task Force on the Prevention and Treatment of Obesity, indications for initiating a VLCD include well-motivated individuals who are moderately to severely obese (BMI >30), have failed at more conservative approaches to weight loss, and have a medical condition that would be immediately improved with rapid weight loss. These conditions include poorly controlled Type 2 diabetes, hypertriglyceridemia, obstructive sleep apnea, and symptomatic peripheral edema. The risk for gallstone formation increases exponentially at rates of weight loss >1.5 kg/week (3.3 lb/week). Prophylaxis against gallstone formation with ursodeoxycholic acid, 600 mg/d, is effective in reducing this risk. Because of the need for close metabolic monitoring, these diets usually are prescribed by physicians specializing in obesity care.

Physical Activity Therapy

Although exercise alone is only moderately effective for weight loss, the combination of dietary modification and exercise is the most effective behavioral approach for the treatment of obesity. The most important role of exercise appears to be in the maintenance of the weight loss. The 2008 Physical Activity Guidelines for Americans recommends that adults should engage in 150 min a week of moderate-intensity or 75 minutes a week of vigorous-intensity aerobic physical activity performed in episodes of at least 10 min, preferably spread throughout the week. The guidelines can be found at www.health.gov/paguidelines. Focusing on simple ways to add physical activity into the normal daily routine through leisure activities, travel, and domestic work should be suggested. Examples include walking, using the stairs, doing home and yard work, and engaging in sport activities. Asking the

patient to wear a pedometer to monitor total accumulation of steps as part of the activities of daily living is a useful strategy. Step counts are highly correlated with activity level. Studies have demonstrated that lifestyle activities are as effective as structured exercise programs for improving cardiorespiratory fitness and weight loss. A high amount of physical activity (more than 300 min of moderate-intensity activity a week) is often needed to lose weight and sustain weight loss. These exercise recommendations are daunting to most patients and need to be implemented gradually. Consultation with an exercise physiologist or personal trainer may be helpful.

Behavioral Therapy

Cognitive behavioral therapy is used to help change and reinforce new dietary and physical activity behaviors. Strategies include self-monitoring techniques (e.g., journaling, weighing, and measuring food and activity); stress management; stimulus control (e.g., using smaller plates, not eating in front of the television or in the car); social support; problem solving; and cognitive restructuring to help patients develop more positive and realistic thoughts about themselves. When recommending any behavioral lifestyle change, have the patient identify what, when, where, and how the behavioral change will be performed. The patient should keep a record of the anticipated behavioral change so that progress can be reviewed at the next office visit. Because these techniques are time-consuming to implement, they are often provided by ancillary office staff such as a nurse clinician or registered dietitian.

PHARMACOTHERAPY

Adjuvant pharmacologic treatments should be considered for patients with a BMI >30 kg/m^2 or a BMI >27 kg/m^2 for those who also have concomitant obesity-related diseases and for whom dietary and physical activity therapy has not been successful. When an antiobesity medication is prescribed, patients should be actively engaged in a lifestyle program that provides the strategies and skills needed to use the drug effectively since this support increases total weight loss.

There are several potential targets of pharmacologic therapy for obesity. The most thoroughly explored treatment is suppression of appetite via centrally active medications that alter monoamine neurotransmitters. A second strategy is to reduce the absorption of selective macronutrients from the gastrointestinal (GI) tract, such as fat.

Centrally Acting Anorexiant Medications

Appetite-suppressing drugs, or anorexiants, affect satiety—the absence of hunger after eating—and hunger—a biologic sensation that initiates eating. By increasing satiety and decreasing hunger, these agents help patients reduce caloric intake without a sense of deprivation. The target site for the actions of anorexiants is the ventromedial and lateral hypothalamic regions in the central nervous system (Chap. 77). Their biologic effect on appetite regulation is produced by augmenting the neurotransmission of three monoamines: norepinephrine; serotonin [5-hydroxytryptamine (5-HT)]; and, to a lesser degree, dopamine. The classic sympathomimetic adrenergic agents (benzphetamine, phendimetrazine, diethylpropion, mazindol, and phentermine) function by stimulating norepinephrine release or by blocking its reuptake. In contrast, sibutramine (Meridia) functions as a serotonin and norepinephrine reuptake inhibitor. Unlike other previously used anorexiants, sibutramine is not pharmacologically related to amphetamine and has no addictive potential.

Sibutramine was the only available anorexiant approved by the U.S. Food and Drug Administration (FDA) for long-term use until it was voluntarily withdrawn from the U.S. market by the manufacturer in October 2010, due to an increased risk of nonfatal myocardial infarction and nonfatal stroke among individuals with preexisting cardiovascular disease.

Peripherally Acting Medications

Orlistat (Xenical) is a synthetic hydrogenated derivative of a naturally occurring lipase inhibitor, lipostatin, produced by the mold *Streptomyces toxytricini*. Orlistat is a potent, slowly reversible inhibitor of pancreatic, gastric, and carboxylester lipases and phospholipase A_2, which are required for the hydrolysis of dietary fat into fatty acids and monoacylglycerols. The drug acts in the lumen of the stomach and small intestine by forming a covalent bond with the active site of these lipases. Taken at a therapeutic dose of 120 mg tid, orlistat blocks the digestion and absorption of about 30% of dietary fat. After discontinuation of the drug, fecal fat usually returns to normal concentrations within 48–72 h.

Multiple randomized, double-blind, placebo-controlled studies have shown that after 1 year, orlistat produces a weight loss of about 9–10%, compared with a 4–6% weight loss in the placebo-treated groups. Because orlistat is minimally (<1%) absorbed from the GI tract, it has no systemic side effects. Tolerability to the drug is related to the malabsorption of dietary fat and subsequent passage of fat in the feces. GI tract adverse effects are reported in at least 10% of orlistat-treated patients. These effects include flatus with discharge, fecal urgency, fatty/oily stool, and increased defecation. These side effects generally are experienced early, diminish as patients control their dietary fat intake, and infrequently cause patients to withdraw from clinical trials. Psyllium mucilloid is helpful in controlling the orlistat-induced GI side effects when taken concomitantly with the medication. Serum concentrations of the fat-soluble vitamins D and E and β-carotene may be reduced, and vitamin supplements are recommended to prevent potential deficiencies. Orlistat was approved for over-the-counter use in

2007.

The Endocannabinoid System

Cannabinoid receptors and their endogenous ligands have been implicated in a variety of physiologic functions, including feeding, modulation of pain, emotional behavior, and peripheral lipid metabolism. Cannabis and its main ingredient, Δ^9-tetrahydrocannabinol (THC), is an exogenous cannabinoid compound. Two endocannabinoids have been identified: anandamide and 2-arachidonyl glyceride. Two cannabinoid receptors have been identified: CB_1 (abundant in the brain) and CB_2 (present in immune cells). The brain endocannabinoid system is thought to control food intake by reinforcing motivation to find and consume foods with high incentive value and to regulate actions of other mediators of appetite. The first selective cannabinoid CB_1 receptor antagonist, rimonabant, was discovered in 1994. The medication antagonizes the orexigenic effect of THC and suppresses appetite. Several large prospective, randomized controlled trials have demonstrated the effectiveness of rimonabant as a weight-loss agent with concomitant improvements in waist circumference and cardiovascular risk factors. However, increased risk of neurologic and psychiatric side effects —seizures, depression, anxiety, insomnia, aggressiveness, and suicidal thoughts among patients randomized to rimonabant—resulted in a ruling against approval of the drug by the FDA in June 2007. Although the drug was available in 56 countries around the world in 2008, approval was officially withdrawn by the European Medicines Agency (EMEA) in January 2009, stating that the benefits of rimonabant no longer outweighed its risks. Development of CB_1 antagonists that do not enter the brain and selectively target the peripheral endocannabinoid system is needed.

Antiobesity Drugs in Development

An emerging theme in pharmacotherapy for obesity is to target several points in the regulatory pathways that control body weight. Several combination drug therapies have completed phase III trials and have been submitted to the FDA for approval. Bupropion and naltrexone (Contrave), a dopamine and norepinephrine reuptake inhibitor and an opioid receptor antagonist, respectively, are combined to dampen the motivation/reinforcement that food brings (dopamine effect) and the pleasure/palatability of eating (opioid effect). Another formulation of bupropion with zonisamide (Empatic) combines bupropion with an anticonvulsant that has serotonergic and dopaminergic activity. Lastly, a formulation of phentermine and topiramate (Qnexa) combines a catecholamine releaser and an anticonvulsant, respectively, that have independently been shown to result in weight loss. The mechanism responsible for topiramate's weight loss is uncertain but is thought to be mediated through its modulation of γ-aminobutyric acid (GABA) receptors, inhibition of carbonic anhydrase, and antagonism of glutamate to reduce food intake. In October 2010, the FDA rejected Qnexa's initial application as a new drug, citing clinical concerns regarding the potential teratogenic risks of topiramate in women of childbearing age. An additional investigational drug, lorcaserin, a $5\text{-}HT_2C$ receptor agonist, has completed phase III trials as a single agent. The FDA rejected Lorcaserin's initial application as a new drug, citing clinical concerns that the weight loss efficacy in overweight and obese individuals without type 2 diabetes is marginal, and non-clinical concerns related to mammary adenocarcinomas in female rats.

SURGERY

Bariatric surgery can be considered for patients with severe obesity (BMI ≥40 kg/m^2) or those with moderate obesity (BMI ≥35 kg/m^2) associated with a serious medical condition. Surgical weight loss functions by reducing caloric intake and, depending on the procedure, macronutrient absorption.

Weight-loss surgeries fall into one of two categories: restrictive and restrictive-malabsorptive **(Fig. 78-2)**. Restrictive surgeries limit the amount of food the stomach can hold and slow the rate of gastric emptying. The vertical banded gastroplasty (VBG) is the prototype of this category but is currently performed on a very limited basis due to lack of effectiveness in long-term trials. Laparoscopic adjustable silicone gastric banding (LASGB) has replaced the VBG as the most commonly performed restrictive operation. The first banding device, the LAP-BAND, was approved for use in the United States in 2001, and the second, the REALIZE band, in 2007. In contrast to previous devices, the diameters of these bands are adjustable by way of their connection to a reservoir that is implanted under the skin. Injection or removal of saline into the reservoir tightens or loosens the band's internal diameter, thus changing the size of the gastric opening.

Figure 78-2

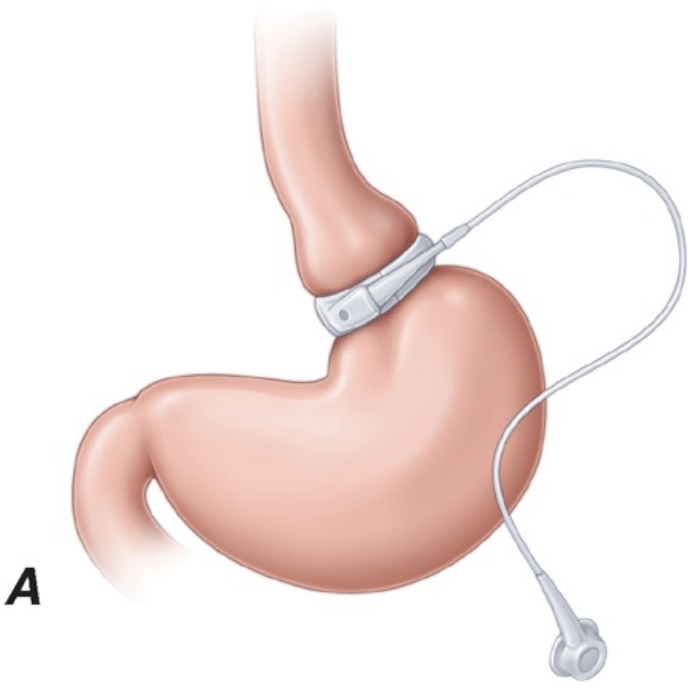

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

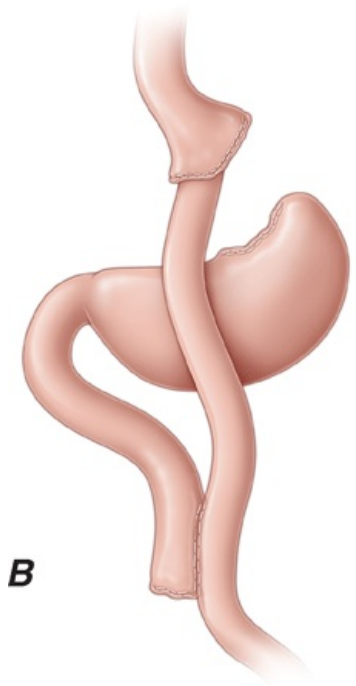

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

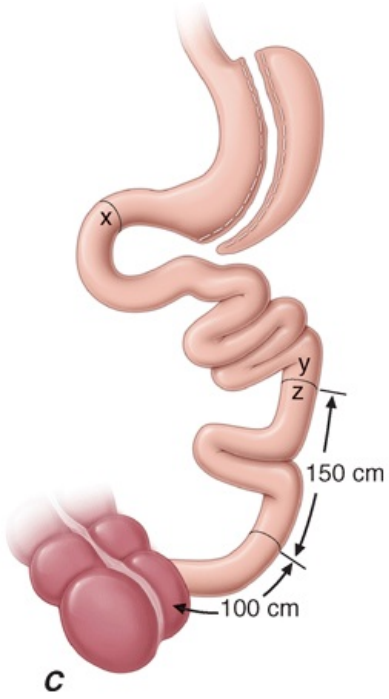

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

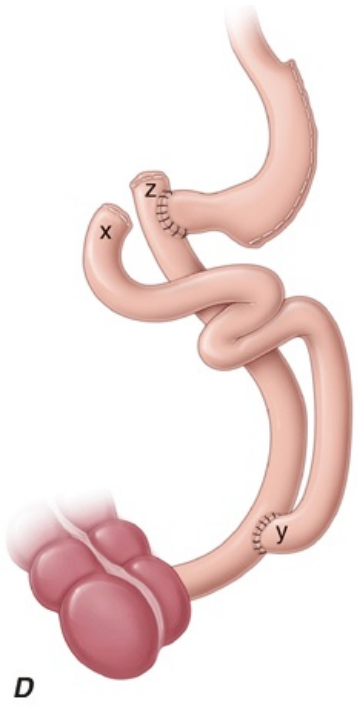

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bariatric surgical procedures. Examples of operative interventions used for surgical manipulation of the gastrointestinal tract. ***A.*** Laparoscopic gastric band (LAGB). ***B.*** The Roux-en-Y gastric bypass. ***C.*** Biliopancreatic diversion with duodenal switch. ***D.*** Biliopancreatic diversion. *(From ML Kendrick, GF Dakin: Mayo Clin Proc 815:518, 2006; with permission.)*

The three restrictive-malabsorptive bypass procedures combine the elements of gastric restriction and selective malabsorption. These procedures include Roux-en-Y gastric bypass (RYGB), biliopancreatic diversion (BPD), and biliopancreatic diversion with duodenal switch (BPDDS) (Fig. 78-2). RYGB is the most commonly performed and accepted bypass procedure. It may be performed with an open incision or laparoscopically.

Although no recent randomized controlled trials compare weight loss after surgical and nonsurgical interventions, data from meta-analyses and large databases, primarily obtained from observational studies, suggest that bariatric surgery is the most effective weight-loss therapy for those with clinically severe obesity. These procedures generally produce a 30–35% average total body weight loss that is

maintained in nearly 60% of patients at 5 years. In general, mean weight loss is greater after the combined restrictive-malabsorptive procedures than after the restrictive procedures. An abundance of data supports the positive impact of bariatric surgery on obesity-related morbid conditions, including diabetes mellitus, hypertension, obstructive sleep apnea, dyslipidemia, and nonalcoholic fatty liver disease. The rapid improvement seen in diabetes after restrictive-malabsorptive procedures is thought to be due to surgery-specific, weight-independent effects on glucose homeostasis brought about by alteration of gut hormones.

Surgical mortality rate from bariatric surgery is generally <1% but varies with the procedure, patient's age and comorbid conditions, and experience of the surgical team. The most common surgical complications include stomal stenosis or marginal ulcers (occurring in 5–15% of patients) that present as prolonged nausea and vomiting after eating or inability to advance the diet to solid foods. These complications typically are treated by endoscopic balloon dilatation and acid suppression therapy, respectively. For patients who undergo LASGB, there are no intestinal absorptive abnormalities other than mechanical reduction in gastric size and outflow. Therefore, selective deficiencies occur uncommonly unless eating habits become unbalanced. In contrast, the restrictive-malabsorptive procedures increase risk for micronutrient deficiencies of vitamin B12, iron, folate, calcium, and vitamin D. Patients with restrictive-malabsorptive procedures require lifelong supplementation with these micronutrients.

FURTHER READINGS

Bray GA, Greenway FL: Pharmacologic treatment of the overweight patient. Pharmacol Rev 59:151, 2007[PMID: 17540905] [Full Text]

———, Wilson JF: In the clinic: Obesity. Ann Intern Med 149:ITC4-1, 2008

DeMaria EJ: Bariatric surgery for morbid obesity. N Engl J Med 356:2176, 2007[PMID: 17522401] [Full Text]

Eckel RH: Nonsurgical management of obesity in adults. N Engl J Med 358:1941, 2008[PMID: 18450605] [Full Text]

Kushner RF: Roadmaps for clinical practice: Case studies in disease prevention and health promotion —assessment and management of adult obesity: A primer for physicians. Chicago, American Medical Association, 2003. Available online at www.ama-assn.org/ama/pub/category/10931.html

———: Anti-obesity drugs. Expert Opin Pharmacother 9:1339, 2008

McTigue KM et al: Screening and interventions for obesity in adults: Summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 139:933, 2003. (Appendix tables available at www.annals.org)

National Heart, Lung, and Blood Institute, North American Association for the Study of Obesity: Practical guide: Identification, evaluation, and treatment of overweight and obesity in adults. Bethesda, MD, National Institutes of Health pub number 00-4084, Oct. 2000. Available online at www.nhlbi.nih.gov/guidelines/obesity/practgde.htm

Vetter ML et al: Narrative review: Effect of bariatric surgery on type 2 diabetes mellitus. Ann Intern Med 150:94, 2009[PMID: 19153412] [Full Text]

Wadden TA et al: Lifestyle modification for the management of obesity. Gastroenterology 132:2226, 2007[PMID: 17498514] [Full Text]

Wee CC: A 52-year-old woman with obesity: Review of bariatric surgery. JAMA 302:1097, 2009[PMID: 19671886] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EATING DISORDERS: INTRODUCTION

Anorexia nervosa and bulimia nervosa are characterized by severe disturbances of eating behavior. The salient feature of *anorexia nervosa* (AN) is a voluntary restriction of food intake relative to caloric requirements leading to an inappropriately low body weight. *Bulimia nervosa* (BN) is characterized by recurrent episodes of binge eating followed by abnormal compensatory behaviors, such as self-induced vomiting. AN and BN are distinct clinical syndromes but share common features. Both disorders occur primarily among previously healthy young women who become overly concerned with body shape and weight. Many patients with BN have past histories of AN, and many patients with AN engage in binge eating and purging behavior. In the current diagnostic system, the critical distinction between AN and BN depends on body weight: patients with AN are, by definition, significantly underweight, whereas patients with BN have body weights in the normal range or above. *Binge eating disorder* (BED) is a more recently described syndrome characterized by repeated episodes of binge eating, similar to those of BN, in the absence of inappropriate compensatory behavior.

ANOREXIA NERVOSA

Epidemiology

Among women, the lifetime prevalence of the full syndrome of AN is approximately 1%. AN is much less common in males. AN is more prevalent in cultures where food is plentiful and being thin is associated with attractiveness. Individuals who pursue interests that place a premium on thinness, such as ballet and modeling, are at greater risk. The incidence of AN has increased in recent decades.

Etiology

The etiology of AN is unknown but appears to involve a combination of psychological, biologic, and cultural risk factors. Some factors, such as sexual or physical abuse and a family history of mood disturbance, are best viewed as nonspecific risk factors that increase vulnerability to a range of psychiatric disorders, including AN.

Patients who develop AN are inclined to be more obsessional and perfectionist than their peers. The disorder often begins as a diet not distinguishable at the outset from those undertaken by many adolescents and young women. As weight loss progresses, the fear of gaining weight grows; dieting becomes stricter; and psychological, behavioral, and medical aberrations increase. Eating disorders, including AN, may develop among individuals with type 1 diabetes mellitus and are associated with poorer glycemic control and an increased frequency of complications (Chap. 344).

Numerous physiologic disturbances, including abnormalities in a variety of neurotransmitter systems, have been described in AN (see below). It is difficult to distinguish neurochemical, metabolic, and hormonal changes that may have a role in the initiation or perpetuation of the syndrome from those that are secondary to the disorder. The resolution of most of these abnormalities with weight restoration argues against an etiologic role.

Genetic factors contribute to the risk of development of AN, as its incidence is greater in families with one affected member and the concordance in monozygotic twins is greater than in dizygotic twins. However, specific genes or risk factor loci have not been identified.

Clinical Features

AN typically begins in mid to late adolescence, sometimes in association with a stressful life event such as leaving home for school **(Table 79-1)**. The disorder occasionally develops in early puberty, before menarche, but seldom begins after age 40. Despite being underweight, patients with AN are irrationally afraid of gaining weight. They also exhibit a distortion of body image; despite being emaciated, patients with AN may believe that their body as a whole, or some part of their body, is too fat. Further weight loss is viewed by the patient as a fulfilling accomplishment, whereas weight gain is seen as a personal failure. Patients with AN rarely complain of hunger or fatigue and often exercise extensively. Despite the denial of hunger, one-quarter to one-half of patients with AN engage in eating binges. Patients tend to become socially withdrawn and increasingly committed to work or study, dieting, and exercise. As weight loss progresses, thoughts of food dominate mental life and idiosyncratic rules develop around eating. Patients with AN may obsessively collect cookbooks and recipes and be drawn to food-related occupations.

Table 79-1 Common Characteristics of Anorexia Nervosa, Bulimia Nervosa, and Binge Eating

Disorder

	Anorexia Nervosa[a]	Bulimia Nervosa	Binge Eating Disorder[b]
Clinical characteristics			
Onset	Mid-adolescence	Late adolescence/early adulthood	Late adolescence/early adulthood
Female:male	10:1	10:1	2:1
Lifetime prevalence	1% of women	1–3% of women	4% of men and women
Weight	Markedly decreased	Usually normal	Usually obese
Menstruation	Absent	Usually normal	Usually normal
Binge eating	25–50%	Required for diagnosis	Required for diagnosis
Mortality	~5% per decade	Low	Low
Physical and laboratory findings[a]			
Skin/extremities	Lanugo	Callus/abrasion on dorsum of hand	
	Acrocyanosis Edema		
Cardiovascular	Bradycardia		
	Hypotension		
Gastrointestinal	Salivary gland enlargement	Salivary gland enlargement	
	Slow gastric emptying	Dental erosion	
	Constipation		
	Elevated liver enzymes		
Hematopoietic	Normochromic, normocytic anemia		
	Leukopenia		
Fluid/Electrolyte	Increased BUN, creatinine Hypokalemia	Hypokalemia Hypochloremia	
	Hypophosphatemia,	Alkalosis	
	Hypomagnesemia		
Endocrine	Hypoglycemia		
	Low estrogen or testosterone		
	Low LH and FSH		
	Low-normal thyroxine		
	Normal TSH		
Bone	Increased cortisol Osteopenia		

[a]Patients with anorexia nervosa who frequently induce purging may also exhibit the physical and laboratory findings associated with bulimia nervosa.

[b]Obese patients with binge eating disorder are at risk for complications of obesity.

Abbreviations: BUN, blood urea nitrogen; FSH, follicle stimulating hormone; LH, luteinizing hormone; TSH, thyroid stimulating hormone.

PHYSICAL FEATURES

Patients with AN typically have few physical complaints but may note cold intolerance. Gastrointestinal motility is diminished, leading to reduced gastric emptying and constipation. Some women who develop AN after menarche report that their menses ceased before significant weight loss occurred. Weight and height should be measured to allow calculation of body mass index (BMI; kg/m^2). Vital signs may reveal bradycardia, hypotension, and mild hypothermia. Soft, downy hair growth (lanugo) sometimes occurs, as does alopecia. Salivary gland enlargement, which is associated with starvation as well as

with binge eating and vomiting, may make the face appear surprisingly full in contrast to the marked general wasting. Acrocyanosis of the digits is common, and peripheral edema can be seen in the absence of hypoalbuminemia, particularly when the patient begins to regain weight. Consumption of large amounts of vegetables containing vitamin A can result in a yellow tint to the skin (*hypercarotenemia*), which is especially notable on the palms.

LABORATORY ABNORMALITIES

Mild normochromic, normocytic anemia is frequent, as is mild to moderate leukopenia, with a disproportionate reduction of polymorphonuclear leukocytes. Dehydration may result in slightly increased levels of blood urea nitrogen and creatinine. Serum transaminase levels may increase, especially during the early phases of refeeding. The level of serum proteins is usually normal. Blood sugar is often low and serum cholesterol may be moderately elevated. Hypokalemia, often accompanied by alkalosis, suggests self-induced vomiting or use of diuretics. Hyponatremia is common and may result from excess fluid intake and disturbances in the secretion of antidiuretic hormone. Hypophosphatemia and hypomagnesemia may be present in severe AN, especially as part of a refeeding syndrome.

ENDOCRINE ABNORMALITIES

The regulation of virtually every endocrine system is altered in AN, but the most striking changes occur in the reproductive system. Amenorrhea is hypothalamic in origin and reflects diminished production of gonadotropin-releasing hormone (GnRH). The resulting gonadotropin deficiency causes low plasma estrogen in women and reduced testosterone in men. The hypothalamic GnRH pulse generator is exquisitely sensitive, particularly in women, to body weight, stress, and exercise, each of which may contribute to *hypothalamic amenorrhea* in AN (Chap. 347).

Serum leptin levels are markedly reduced in AN as a result of undernutrition and decreased body-fat mass. The reduction in leptin is the primary factor responsible for the disturbances of the hypothalamic-pituitary-gonadal axis, and an important mediator of the other neuroendocrine abnormalities characteristic of AN (Chap. 77).

Serum cortisol and 24-h urine-free cortisol levels are generally elevated but without characteristic clinical signs of cortisol excess. Thyroid function tests resemble the pattern seen in euthyroid sick syndrome (Chap. 341). Thyroxine (T4) and free T_4 levels are usually in the low-normal range, triiodothyronine (T_3) levels are reduced, and reverse T_3 (rT_3) is elevated. The level of thyroid-stimulating hormone (TSH) is normal or partially suppressed. Growth hormone is increased, but insulin-like growth factor 1 (IGF-1), which is produced mainly by the liver, is reduced, as in other conditions of starvation. Diminished bone density is routinely observed in AN and reflects the effects of multiple nutritional deficiencies, reduced gonadal steroids, increased cortisol, and reduced IGF-1. The degree of bone-density reduction is proportional to the length of the illness, and patients are at risk for the development of symptomatic fractures. The occurrence of AN during adolescence may lead to the premature cessation of linear bone growth and a failure to achieve expected adult height.

CARDIAC ABNORMALITIES

Cardiac output is reduced, and congestive heart failure occurs rarely during rapid refeeding. The electrocardiogram usually shows sinus bradycardia, reduced QRS voltage, and nonspecific ST-T-wave abnormalities. Some patients develop a prolonged QT_c interval, which may predispose to serious arrhythmias, particularly when electrolyte abnormalities are present.

Diagnosis

The diagnosis of AN is based on the presence of characteristic behavioral, psychological, and physical attributes **(Table 79-2)**. Widely accepted diagnostic criteria are provided by the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). These criteria include maintenance of a less than minimally normal body weight for age and height. Weights less than 85% of that expected, roughly equivalent to a BMI of 18.5 kg/m^2, are commonly considered to meet this criterion, but a patient weighing somewhat more who meets all other diagnostic criteria would still merit the diagnosis of AN. The current diagnostic criteria require that women with AN not have spontaneous menses, patients who have other characteristics of AN but report menstrual activity probably merit the diagnosis.

Table 79-2 Diagnostic Features of Anorexia Nervosa

Refusal to maintain body weight at or above a minimally normal weight for age and height. (This includes a failure to achieve weight gain expected during a period of growth leading to an abnormally low body weight.)
Intense fear of weight gain or becoming fat.
Distortion of body image (e.g., feeling fat despite an objectively low weight or minimizing the seriousness of low weight).

Amenorrhea. (This criterion is met if menstrual periods occur only following hormone—e.g., estrogen—administration.)

The diagnosis of AN can usually be made confidently in a patient with a history of weight loss accomplished by restrictive dieting and excessive exercise accompanied by a marked reluctance to gain weight. Patients with AN often deny that they have a serious problem and may be brought to medical attention by concerned family or friends. Especially in atypical presentations, other causes of significant weight loss in previously healthy young people should be considered, including inflammatory bowel disease, gastric outlet obstruction, diabetes mellitus, CNS tumors, or neoplasm (Chap. 80).

Prognosis

The course and outcome of AN are highly variable. One-quarter to one-half of patients eventually recover fully, with few psychological or physical sequelae. However, many patients have persistent difficulties with weight maintenance, depression, and eating disturbances, including BN. The development of obesity following AN is rare. The long-term mortality of AN is among the highest associated with any psychiatric disorder. Approximately 5% of patients die per decade of follow-up, primarily due to the physical effects of chronic starvation or by suicide.

Virtually all of the physiologic abnormalities associated with AN are observed in other forms of starvation and markedly improve or disappear with weight gain. A worrisome exception is the reduction in bone mass, which may not recover fully, particularly if AN occurs during adolescence when peak bone mass is normally achieved.

TREATMENT: ANOREXIA NERVOSA

Because of the profound physiologic and psychological effects of starvation, there is a broad consensus that weight restoration to at least 90% of predicted weight is the primary goal in the treatment of AN. Unfortunately, because most patients resist this goal, the management of AN is often accompanied by frustration for the patient, the family, and the physician. Patients typically exaggerate their food intake and minimize their symptoms. Some patients resort to subterfuge to make their weights appear higher, for example, by water-loading before they are weighed. In attempting to engage the patient in treatment, it may be useful to elicit the patient's physical concerns (e.g., about osteoporosis, weakness, or fertility), and provide education about the importance of normalizing nutritional status in order to address those concerns. The physician should reassure the patient that weight gain will not be permitted to get out of control but simultaneously emphasize that weight restoration is medically and psychologically imperative.

The intensity of the initial treatment, including the need for hospitalization, is determined by the patient's current weight, the rapidity of recent weight loss, and the severity of medical and psychological complications **(Fig. 79-1)**. Hospitalization should be strongly considered for patients weighing <75% of that expected age and height, even if the results of routine blood studies are within normal limits. Acute medical problems, such as severe electrolyte imbalances, should be identified and addressed. Nutritional restoration can almost always be successfully accomplished by oral feeding, and parenteral methods are rarely required. For severely underweight patients, sufficient calories (approximately 1200–1800 kcal/d) should be provided initially in divided meals as food or liquid supplements to maintain weight and permit physiological stabilization. Calories can then be gradually increased to achieve a weight gain of 1–2 kg (2–4 lb) per week, typically requiring an intake of 3000–4000 kcal/d. Meals must be supervised, ideally by personnel who are firm regarding the necessity of food consumption, empathic regarding the challenges entailed, and reassuring about the patient's eventual recovery. Patients have great psychological difficulty complying with the need for increased caloric consumption, and the assistance of psychiatrists or psychologists experienced in the treatment of AN is usually necessary.

Figure 79-1

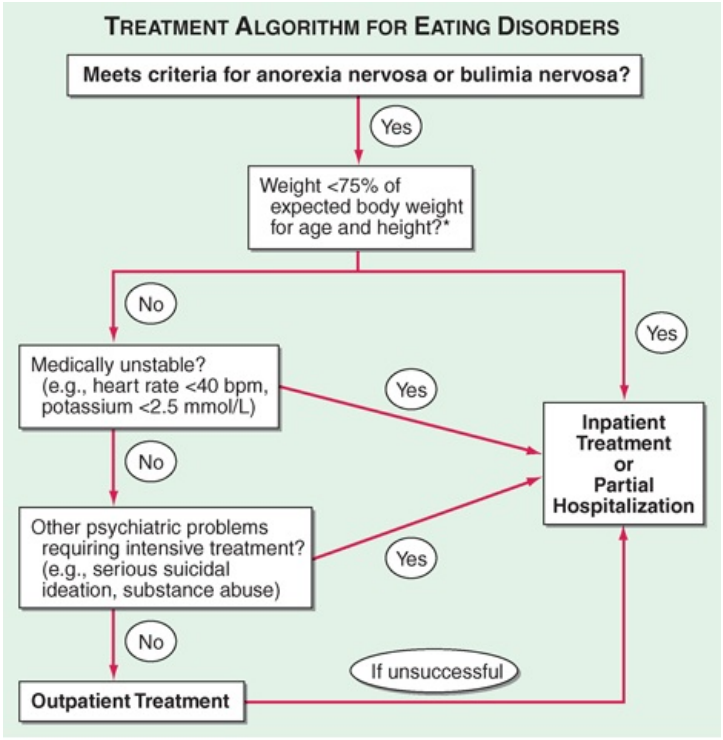

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An algorithm for basic treatment decisions regarding patients with anorexia nervosa or bulimia nervosa. Based on the American Psychiatric Association practice guidelines for the treatment of patients with eating disorders. *Although outpatient management may be considered for patients with anorexia nervosa weighing more than 75% of expected, there should be a low threshold for using more intensive interventions if the weight loss has been rapid or if current weight is <80% of expected.

Less severely affected patients may be treated in a partial hospitalization program where medical and psychiatric supervision is available and several meals can be monitored each day. Outpatient treatment may suffice for mildly ill patients. Weight must be monitored at frequent intervals, and explicit goals agreed on for weight gain, with the understanding that more intensive treatment will be required if the level of care initially employed is not successful. For younger patients, the active involvement of the family in treatment is crucial regardless of treatment setting. Outpatient interventions that help parents refeed their child have been be quite successful at achieving weight restoration.

Psychiatric treatment focuses primarily on two issues. First, patients require much emotional support during the period of weight gain. They often intellectually agree with the need to gain weight, but strenuously resist increases in caloric intake, and often surreptitiously discard food that is provided. Second, patients must learn to base their self-esteem not on the achievement of an inappropriately low weight, but on the development of satisfying personal relationships and the attainment of reasonable academic and occupational goals. While this is often possible, some patients with AN develop other serious emotional and behavioral symptoms such as depression, self-mutilation, obsessive-compulsive behavior, and suicidal ideation. These symptoms may require additional therapeutic interventions, in the form of psychotherapy, medication, or hospitalization.

Medical complications occasionally occur during refeeding. Especially in the early stages of treatment, severely malnourished patients may develop a "refeeding syndrome" characterized by hypophosphatemia, hypomagnesemia, and cardiovascular instability. Acute gastric dilatation has been described when refeeding is rapid. As in other forms of malnutrition, fluid retention and peripheral edema may occur, but they generally do not require specific treatment in the absence of cardiac, renal, or hepatic dysfunction. Transient modest elevations in serum liver enzyme levels occasionally occur. Multivitamins should be given, and an adequate intake of vitamin D (400 IU/d) and calcium (1500 mg/d) should be provided.

No psychotropic medications are of established value in the treatment of AN, although there is recent preliminary evidence that the atypical antipsychotic medication olanzapine may assist some patients by increasing the rate of weight gain and decreasing obsessive thinking. Medications that may prolong the QT_C interval should be avoided. The alterations of cortisol and thyroid hormone metabolism do not require specific treatment and correct with weight gain. Estrogen treatment appears to offer no benefit to bone density in underweight patients, and the small benefit of bisphosphonate treatment appears to be outweighed by the potential risks of such agents in young women.

BULIMIA NERVOSA

Epidemiology

In women, the full syndrome of BN occurs with a lifetime prevalence of 1–3%. Variants of the disorder, such as occasional binge eating or purging, are much more common and occur in 5–10% of young women. The frequency of BN among men is less than one-tenth of that among women. The prevalence of BN increased dramatically in the early 1970s and 1980s but may have leveled off or declined somewhat in recent years.

Etiology

As with AN, the etiology of BN is likely to be multifactorial. Patients who develop BN describe a higher-than-expected prevalence of childhood and parental obesity, suggesting that a predisposition toward obesity may increase vulnerability to this eating disorder. The marked increase in the number of cases of BN during the past 30 years and the rarity of BN in underdeveloped countries imply that cultural factors are important.

Clinical Features

The typical patient presenting for treatment of BN is a woman of normal weight in her mid-twenties who reports binge eating and purging 5–10 times a week for 5–10 years **(Table 79-3)**. The disorder usually begins in late adolescence or early adulthood during or following a diet, often in association with depressed mood. The self-imposed caloric restriction leads to increased hunger and to overeating. In an attempt to avoid weight gain, the patient induces vomiting, takes laxatives or diuretics, or engages in some other form of compensatory behavior. During binges, patients with this disorder tend to consume large amounts of sweet foods with a high fat content, such as dessert items. The most frequent compensatory behaviors are self-induced vomiting and laxative abuse, but a wide variety of techniques have been described, including the omission of insulin injections by individuals with type 1 diabetes mellitus. Initially, patients may experience a sense of satisfaction that appealing food can be eaten without weight gain. However, as the disorder progresses, patients perceive diminished control overeating. Binges increase in size and frequency and are provoked by a variety of stimuli, such as transient depression, anxiety, or a sense that too much food has been consumed in a normal meal. Between binges, patients restrict caloric intake, which increases hunger and sets the stage for the next binge. Typically, patients with BN are ashamed of their behavior and endeavor to keep their disorder hidden from family and friends. Like patients with AN, those with BN place an unusual emphasis on weight and shape as a basis for their self-esteem. Many patients with BN have mild symptoms of depression. Some patients exhibit serious mood and behavioral disturbances, such as suicide attempts, sexual promiscuity, and drug and alcohol abuse. Although vomiting may be triggered initially by manual stimulation of the gag reflex, most patients with BN develop the ability to induce vomiting at will. Rarely, patients resort to the regular use of syrup of ipecac. Laxatives and diuretics are frequently taken in impressive quantities, such as 30 or 60 laxative pills on a single occasion. The resulting fluid loss produces dehydration and a feeling of emptiness but has little impact on caloric balance.

Table 79-3 Diagnostic Features of Bulimia Nervosa

Recurrent episodes of binge eating, which is characterized by the consumption of a large amount of food in a short period of time and a feeling that the eating is out of control.
Recurrent inappropriate behavior to compensate for the binge eating, such as self-induced vomiting.
The occurrence of both the binge eating and the inappropriate compensatory behavior at least twice weekly, on average, for 3 months.
Overconcern with body shape and weight.

Note: If the diagnostic criteria for anorexia nervosa are simultaneously met, only the diagnosis of anorexia nervosa is given.

The physical abnormalities associated with BN primarily result from the purging behavior. Painless bilateral salivary gland hypertrophy (sialadenosis) may be noted. A scar or callus on the dorsum of the hand may develop due to repeated trauma from the teeth among patients who manually stimulate the gag reflex. Recurrent vomiting and the exposure of the lingual surfaces of the teeth to stomach acid lead to loss of dental enamel and eventually to chipping and erosion of the front teeth. Laboratory abnormalities are surprisingly infrequent, but hypokalemia, hypochloremia, and hyponatremia are observed occasionally. Repeated vomiting may lead to alkalosis, whereas repeated laxative abuse may produce a mild metabolic acidosis. Serum amylase may be slightly elevated due to an increase in the salivary isoenzyme.

Serious physical complications resulting from BN are rare. Oligomenorrhea and amenorrhea are more frequent than among women without eating disorders. Arrhythmias occasionally occur secondary to electrolyte disturbances. Tearing of the esophagus and rupture of the stomach have been reported

and constitute life-threatening events. Some patients who chronically abuse laxatives or diuretics develop transient peripheral edema when this behavior ceases, presumably due to high levels of aldosterone secondary to persistent fluid and electrolyte depletion.

Diagnosis

The critical diagnostic features of BN are repeated episodes of binge eating followed by inappropriate behaviors aimed at avoiding weight gain **(Table 79-3).** The diagnosis of BN requires a candid history from the patient detailing frequent, large eating binges followed by the purposeful use of inappropriate mechanisms to avoid weight gain. Most patients with BN who present for treatment are distressed by their inability to control their eating behavior and are able to provide such details if queried in a supportive and nonjudgmental fashion.

Prognosis

The prognosis of BN is much more favorable than that of AN. Mortality is low, and full recovery occurs in approximately 50% of patients within 10 years. Approximately 25% of patients have persistent symptoms of BN over many years. Few patients progress from BN to AN.

Treatment: Bulimia Nervosa

BN can usually be treated on an outpatient basis (Fig. 79-1). Cognitive behavioral therapy (CBT) is a short-term (4–6 months) psychological treatment that focuses on the intense concern with shape and weight, the persistent dieting, and the binge eating and purging that characterize this disorder. Patients are directed to monitor the circumstances, thoughts, and emotions associated with binge/purge episodes, to eat regularly, and to challenge their assumptions linking weight to self-esteem. CBT produces symptomatic remission in 25–50% of patients.

Numerous double-blind, placebo-controlled trials have documented that antidepressant medications are useful in the treatment of BN but are probably somewhat less effective than CBT. Although efficacy has been established for virtually all chemical classes of antidepressants, only the selective serotonin reuptake inhibitor fluoxetine (Prozac) has been approved for use in BN by the U.S. Food and Drug Administration. Antidepressant medications are helpful even for patients with BN who are not depressed, and the dose of fluoxetine recommended for BN (60 mg/d) is higher than that typically used to treat depression. These observations suggest that different mechanisms may underlie the utility of these medications in BN and in depression.

A subset of patients does not respond to CBT, antidepressant medication, or their combination. More intensive forms of treatment, including hospitalization, may be required.

BINGE EATING DISORDER

Binge eating disorder (BED) is characterized by frequent episodes of eating unusually large amounts of food accompanied by feeling loss of control. In contrast to those with BN, patients with BED do not frequently engage in appropriate behavior to compensate for binge eating. In addition, BED is commonly associated with obesity **(Table 79-1)**. BED occurs more frequently than AN and BN in both clinical and community population samples, and males constitute a greater fraction of affected individuals. Compared with obese individuals without BED, those with BED have higher rates of anxiety, depression, and health care use. A range of psychological treatments, such as CBT and interpersonal therapy (IPT) and medications, including antidepressants and weight-loss agents, appear helpful in reducing binge eating. Surprisingly, cessation of binge eating is not routinely followed by loss of weight.

GLOBAL CONSIDERATIONS

Eating disorders are more common in cultures where food is available and thinness is idealized. Nevertheless, eating disorders have been reported across the world, including in many parts of Asia and Africa. Cultural variation may contribute to differing presentations of eating disorder symptoms. For example, in some cultural groups, the rationale for food refusal of low weight individuals who exhibit many symptoms of AN may not include the DSM IV criterion, "intense fear of gaining weight or becoming fat."

FURTHER READINGS

Attia E, Walsh BT: Behavioral management for anorexia nervosa. N Engl J Med 30:500, 2009

Katzman DK: Medical complications in adolescents with anorexia: A review of the literature. Int J Eat Disord 37(Suppl):S52, 2005

Keski-Rahkonen A et al: Epidemiology and course of anorexia nervosa in the community. Am J Psychiatry 164:1259, 2007[PMID: 17671290] [Full Text]

Mehler PS: Clinical practice. Bulimia nervosa. N Engl J Med 349:875, 2003[PMID: 12944574] [Full Text]

Mehler PS, Mackenzie TD: Treatment of osteopenia and osteoporosis in anorexia nervosa: A systematic review of the literature. Int J Eat Disord 42:195, 2009[PMID: 18951456] [Full Text]

Treasure J et al: Eating Disorders. Lancet 375:583, 2010[PMID: 19931176] [Full Text]

Yager J et al: Practice guideline for the treatment of patients with eating disorders, 3rd ed, in *American Psychiatric Association Practice Guidelines for the Treatment of Psychiatric Disorders Compendium 2006*, American Psychiatric Association (ed). Arlington, Virginia, American Psychiatric Association, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INVOLUNTARY WEIGHT LOSS: INTRODUCTION

Involuntary weight loss (IWL) is frequently insidious and can have important implications, often serving as a harbinger of serious underlying disease. Clinically important weight loss is defined as the loss of 10 pounds (4.5 kg) or >5% of one's body weight over a period of 6–12 months. IWL is encountered in up to 8% of all adult outpatients and 27% of frail persons age 65 years and older. There is no identifiable cause in up to one-quarter of patients despite extensive investigation. Conversely, up to half of people who claim to have lost weight have no documented evidence of weight loss. People with no known cause of weight loss generally have a better prognosis than do those with known causes, particularly when the source is neoplastic. Weight loss in older persons is associated with a variety of deleterious effects, including hip fracture, pressure ulcers, impaired immune function, decreased functional status, and death. Not surprisingly, significant weight loss is associated with increased mortality, which can range from 9% to as high as 38% within 1 to 2.5 years in the absence of clinical awareness and attention.

Physiology of Weight Regulation with Aging

(See also Chaps. 71 and 77) Among healthy aging people, total body weight peaks in the sixth decade of life and generally remains stable until the ninth decade, after which it gradually falls. In contrast, lean body mass (fat-free mass) begins to decline at a rate of 0.3 kg per year in the third decade, and the rate of decline increases further beginning at age 60 in men and age 65 in women. These changes in lean body mass largely reflect the age-dependent decline in growth hormone secretion and, consequently, circulating levels of insulin-like growth factor type I (IGF-I) that occur with normal aging. In the healthy elderly, an increase in fat tissue balances the loss in lean body mass until very old age, when loss of both fat and skeletal muscle occurs. Age-dependent changes also occur at the cellular level. Telomeres shorten, and body cell mass—the fat-free portion of cells—declines steadily with aging.

Between ages 20 and 80, mean energy intake is reduced by up to 1200 kcal/d in men and 800 kcal/d in women. Decreased hunger is a reflection of reduced physical activity and loss of lean body mass, producing lower demand for calories and food intake. Several important age-associated physiologic changes also predispose elderly persons to weight loss, such as declining chemosensory function (smell and taste), reduced efficiency of chewing, slowed gastric emptying, and alterations in the neuroendocrine axis, including changes in levels of leptin, cholecystokinin, neuropeptide Y, and other hormones and peptides. These changes are associated with early satiety and a decline in both appetite and the hedonistic appreciation of food. Collectively, they contribute to the "anorexia of aging."

Causes of Involuntary Weight Loss

Most causes of IWL belong to one of four categories: (1) malignantneoplasms, (2) chronic inflammatory or infectious diseases, (3) metabolic disorders (e.g., hyperthyroidism and diabetes), or (4) psychiatric disorders (**Table 80-1**). Not infrequently, more than one of these causes can be responsible for IWL. In most series, IWL is caused by malignant disease in a quarter of patients and by organic disease in one-third, with the remainder due to psychiatric disease, medications, or uncertain causes.

Table 80-1 Causes of Involuntary Weight Loss

Cancer
Colon
Hepatobiliary
Hematologic Lung
Breast
Genitourinary
Ovarian
Prostate
Gastrointestinal disorders

Malabsorption
Peptic ulcer
Inflammatory bowel disease
Pancreatitis
Obstruction/constipation
Pernicious anemia
Endocrine and metabolic
Hyperthyroidism
Diabetes mellitus
Pheochromocytoma
Adrenal insufficiency
Cardiac disorders
Chronic ischemia
Chronic congestive heart failure
Respiratory disorders
Emphysema
Chronic obstructive pulmonary disease
Renal insufficiency Rheumatologic disease Infections
HIV
Tuberculosis
Parasitic infection
Subacute bacterial endocarditis
Medications
Sedatives
Antibiotics
Nonsteroidal anti-inflammatory drugs
Serotonin reuptake inhibitors
Metformin
Levodopa
Angiotensin-converting enzyme inhibitors
Other drugs
Disorders of the mouth and teeth
Caries
Dygeusia
Age-related factors
Physiologic changes
Visual impairment
Decreased taste and smell
Functional disabilities
Neurologic
Stroke Parkinson's disease
Neuromuscular disorders
Dementia
Social
Isolation
Economic hardship
Psychiatric and behavioral
Depression
Anxiety
Paranoia
Bereavement
Alcoholism
Eating disorders

Increased activity or exercise
Idiopathic

The most common malignant causes of IWL are gastrointestinal, hepatobiliary, hematologic, lung, breast, genitourinary, ovarian, and prostate. Half of all patients with cancer lose some body weight; one-third lose more than 5% of their original body weight, and up to 20% of all cancer deaths are caused directly by cachexia (through immobility and/or cardiac/respiratory failure). The greatest incidence of weight loss is seen among patients with solid tumors. Malignancy that reveals itself through significant weight loss usually has a very poor prognosis.

In addition to malignancies, gastrointestinal causes are among the most prominent causes of IWL. Peptic ulcer disease, inflammatory bowel disease, dysmotility syndromes, chronic pancreatitis, celiac disease, constipation, and atrophic gastritis are some of the more common entities. Oral and dental problems are easily overlooked and may manifest with halitosis, poor oral hygiene, xerostomia, inability to chew, reduced masticatory force, nonocclusion, temporomandibular joint syndrome, edentulousness, and pain due to caries or abscesses.

Tuberculosis, fungal diseases, parasites, subacute bacterial endocarditis, and HIV are well-documented causes of IWL. Cardiovascular and pulmonary diseases cause unintentional weight loss through increased metabolic demand and decreased appetite and caloric intake. Uremia produces nausea, anorexia, and vomiting. Connective tissue diseases may increase metabolic demand and disrupt nutritional balance. As the incidence of diabetes mellitus increases with aging, the associated glucosuria can contribute to weight loss. Hyperthyroidism in the elderly may have less prominent sympathomimetic features and may present as "apathetic hyperthyroidism" or T_3 toxicosis (Chap. 341).

Neurologic injuries such as stroke, quadriplegia, and multiple sclerosis may lead to visceral and autonomic dysfunction that can impair caloric intake. Dysphagia from these neurologic insults is a common mechanism. Functional disability that compromises activities of daily living (ADLs) is a common cause of undernutrition in the elderly. Visual impairment from ophthalmic or central nervous system disorders such as a tremor can limit the ability of people to prepare and eat meals. IWL may be one of the earliest manifestations of Alzheimer's dementia.

Isolation and depression are significant causes of IWL that may manifest as an inability to care for oneself, including nutritional needs. A cytokine-mediated inflammatory metabolic cascade can be both a cause of and a manifestation of depression. Bereavement can be a cause of IWL and, when present, is more pronounced in men. More intense forms of mental illness such as paranoid disorders may lead to delusions about food and cause weight loss. Alcoholism can be a significant source of weight loss and malnutrition.

Elderly persons living in poverty may have to choose between purchasing food and purchasing medications. Institutionalization is an independent risk factor, as up to 30–50% of nursing home patients have inadequate food intake.

Medications can cause anorexia, nausea, vomiting, gastrointestinal distress, diarrhea, dry mouth, and changes in taste. This is particularly an issue in the elderly, many of whom take five or more medications.

Assessment

The four major manifestations of IWL are (1) anorexia (loss of appetite), (2) sarcopenia (loss of muscle mass), (3) cachexia (a syndrome that combines weight loss, loss of muscle and adipose tissue, anorexia, and weakness), and (4) dehydration. The current obesity epidemic adds complexity, as excess adipose tissue can mask the development of sarcopenia and delay awareness of the development of cachexia. If it is not possible to measure weight directly, a change in clothing size, corroboration of weight loss by a relative or friend, and a numeric estimate of weight loss provided by the patient are suggestive of true weight loss.

Initial assessment includes a comprehensive history and physical, a complete blood count, tests of liver enzyme levels, a C-reactive protein, erythrocyte sedimentation rate, renal function studies, thyroid function tests, chest radiography, and an abdominal ultrasound **(Table 80-2**). Age, sex, and risk factor –specific cancer screening tests, such as mammography and colonoscopy, should be performed (Chap. 82). Patients at risk should have HIV testing. All elderly patients with weight loss should undergo screening for dementia and depression by using instruments such as the Mini-Mental Status Examination and the Geriatric Depression Scale, respectively (Chap. 72). The Mini Nutritional Assessment (www.mna-elderly.com) and the Nutrition Screening Initiative (www.aafp.org/afp/980301ap/edits.html) are also available for the nutritional assessment of elderly patients. Almost all patients with a malignancy and >90% of those with other organic diseases have at least one laboratory abnormality. In patients presenting with substantial IWL, major organic and malignant diseases are unlikely when a baseline evaluation is completely normal. Careful follow-up rather than undirected testing is advised since the prognosis of weight loss of undetermined cause is generally favorable.

Table 80-2 Assessment and Testing for Involuntary Weight Loss

Indications	**Laboratory**
5% weight loss in 30 d	Complete blood count
10% weight loss in 180 d Body mass index <21 25% of food left uneaten after 7 d	Comprehensive electrolyte and metabolic panel, including liver and renal function tests Thyroid function tests
Change in fit of clothing Change in appetite, smell, or taste	Erythrocyte sedimentation rate C-reactive protein
Abdominal pain, nausea, vomiting, diarrhea, constipation, dysphagia	Ferritin HIV testing, if indicated
Assessment	**Radiology**
Complete physical exam, including dental evaluation	Chest x-ray Abdominal ultrasound
Medication review	
Recommended cancer screening	
Mini-Mental State Examination*	
Mini-Nutritional Assessment* Nutrition Screening Initiative*	
Simplified Nutritional Assessment Questionnaire*	
Observation of eating*	
Activities of daily living?* Instrumental activities of daily living*	

*May be more specific to assess weight loss in the elderly.

Treatment: Unintentional Weight Loss

The first priority in managing weight loss is to identify and treat the underlying causes systematically. Treatment of underlying metabolic, psychiatric, infectious, or other systemic disorders may be sufficient to restore weight and functional status gradually. Medications that cause nausea or anorexia should be withdrawn or changed, if possible. For those with unexplained IWL, oral nutritional supplements such as high-energy drinks sometimes reverse weight loss. Advising patients to consume supplements between meals rather than with a meal may help minimize appetite suppression and facilitate increased overall intake. Orexigenic, anabolic, and anticytokine agents are under investigation. In selected patients, the antidepressant mirtazapine results in a significant increase in body weight, body fat mass, and leptin concentration. Patients with wasting conditions who can comply with an appropriate exercise program gain muscle protein mass, strength, and endurance and may be more capable of performing ADL.

FURTHER READINGS

Alibhai SM et al: An approach to the management of unintentional weight loss in elderly people. CMAJ 172:773, 2005[PMID: 15767612] [Full Text]

Chen SP et al: Evaluating probability of cancer among older people with unexplained, unintentional weight loss. Arch Gerontol Geriatr 50 Suppl 1:S27, 2010

Miller SL, Wolfe RR: The danger of weight loss in the elderly. J Nutr Health Aging 12:487, 2008[PMID: 18615231] [Full Text]

Rolland Y et al: Office management of weight loss in older persons. Am J Med 119:1019, 2006[PMID: 17145241] [Full Text]

Vanderschueren S et al: The diagnostic spectrum of unintentional weight loss. Eur J Intern Med 16:160, 2005[PMID: 15967329] [Full Text]

Visvanathan R, Chapman IM: Undernutrition and anorexia in the older person. Gastroenterol Clin North Am 38:393, 2009[PMID: 19699404] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH CANCER: INTRODUCTION

The application of current treatment techniques (surgery, radiation therapy, chemotherapy, and biologic therapy) results in the cure of nearly two of three patients diagnosed with cancer. Nevertheless, patients experience the diagnosis of cancer as one of the most traumatic and revolutionary events that has ever happened to them. Independent of prognosis, the diagnosis brings with it a change in a person's self-image and in his or her role in the home and workplace. The prognosis of a person who has just been found to have pancreatic cancer is the same as the prognosis of the person with aortic stenosis who develops the first symptoms of congestive heart failure (median survival, ~8 months). However, the patient with heart disease may remain functional and maintain a self-image as a fully intact person with just a malfunctioning part, a diseased organ ("a bum ticker"). By contrast, the patient with pancreatic cancer has a completely altered self-image and is viewed differently by family and anyone who knows the diagnosis. He or she is being attacked and invaded by a disease that could be anywhere in the body. Every ache or pain takes on desperate significance. Cancer is an exception to the coordinated interaction among cells and organs. In general, the cells of a multicellular organism are programmed for collaboration. Many diseases occur because the specialized cells fail to perform their assigned task. Cancer takes this malfunction one step further. Not only is there a failure of the cancer cell to maintain its specialized function but it also strikes out on its own; the cancer cell competes to survive using natural mutability and natural selection to seek advantage over normal cells in a recapitulation of evolution. One consequence of the traitorous behavior of cancer cells is that the patient feels betrayed by his or her body. The cancer patient feels that he or she, and not just a body part, is diseased.

THE MAGNITUDE OF THE PROBLEM

No nationwide cancer registry exists; therefore, the incidence of cancer is estimated on the basis of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database, which tabulates cancer incidence and death figures from nine sites, accounting for about 10% of the U.S. population, and from population data from the U.S. Census Bureau. In 2010, 1.530 million new cases of invasive cancer (789,620 men, 739,940 women) were diagnosed and 569,490 persons (299,200 men, 270,290 women) died from cancer. The percent distribution of new cancer cases and cancer deaths by site for men and women are shown in **Table 81–1**. Cancer incidence has been declining by about 2% each year since 1992.

Table 81–1 Distribution of Cancer Incidence and Deaths for 2010

Male			Female		
Sites	**%**	**Number**	**Sites**	**%**	**Number**
Cancer Incidence					
Prostate	28	217,730	Breast	28	207,090
Lung	15	116,750	Lung	14	105,770
Colorectal	9	72,090	Colorectal	10	70,480
Bladder	7	52,760	Endometrial	6	43,470
Melanoma	5	38,870	Thyroid	5	33,930
Lymphoma	4	35,380	Lymphoma	4	30,160
Kidney	4	35,370	Melanoma	4	29,260
Oral cavity	3	25,420	Kidney	3	22,870
Leukemia	3	24,690	Ovary	3	21,880
Pancreas	3	21,370	Pancreas	3	21,770
All others	19	149,190	All others	20	153,260
All sites	100	789,620	All sites	100	739,940
Cancer Deaths					
Lung	29	86,220	Lung	26	71,080
Prostate	11	32,050	Breast	15	39,840

Colorectal	9	26,580	Colorectal	9	24,790
Pancreas	6	18,770	Pancreas	7	18,030
Liver	4	12,720	Ovary	5	13,850
Leukemia	4	12,660	Lymphoma	4	9500
Esophagus	4	11,650	Leukemia	3	9180
Lymphoma	4	10,710	Endometrial	3	7950
Bladder	3	10,410	Liver	2	6190
Kidney	3	8210	CNS	2	5720
All others	23	69,220	All others	24	64,160
All sites	100	299,200	All sites	100	270,290

The most significant risk factor for cancer overall is age; two-thirds of all cases were in those older than age 65 years. Cancer incidence increases as the third, fourth, or fifth power of age in different sites. For the interval between birth and age 39 years, 1 in 70 men and 1 in 48 women will develop cancer; for the interval between ages 40 and 59 years, 1 in 12 men and 1 in 11 women will develop cancer; and for the interval between ages 60 and 79 years, 1 in 3 men and 1 in 5 women will develop cancer. Overall, men have a 44% risk of developing cancer at some time during their lives; women have a 38% lifetime risk.

Cancer is the second leading cause of death behind heart disease. Deaths from heart disease have declined 45% in the United States since 1950 and continue to decline. Cancer has overtaken heart disease as the number one cause of death in persons younger than age 85 years **(Fig. 81-1)**. After a 70-year period of increase, cancer deaths began to decline in 1990-1991 **(Fig. 81-2)**. Between 1990 and 2006, cancer deaths decreased by 21% among men and 12.3% among women. The five leading causes of cancer deaths are shown for various populations in **Table 81–2**. The 5-year survival for white patients was 39% in 1960–1963 and 69% in 1999–2005. Cancers are more often deadly in blacks; the 5-year survival was 59% for the 1999–2005 interval. Incidence and mortality vary among racial and ethnic groups **(Table 81–3)**. The basis for these differences is unclear.

Figure 81-1

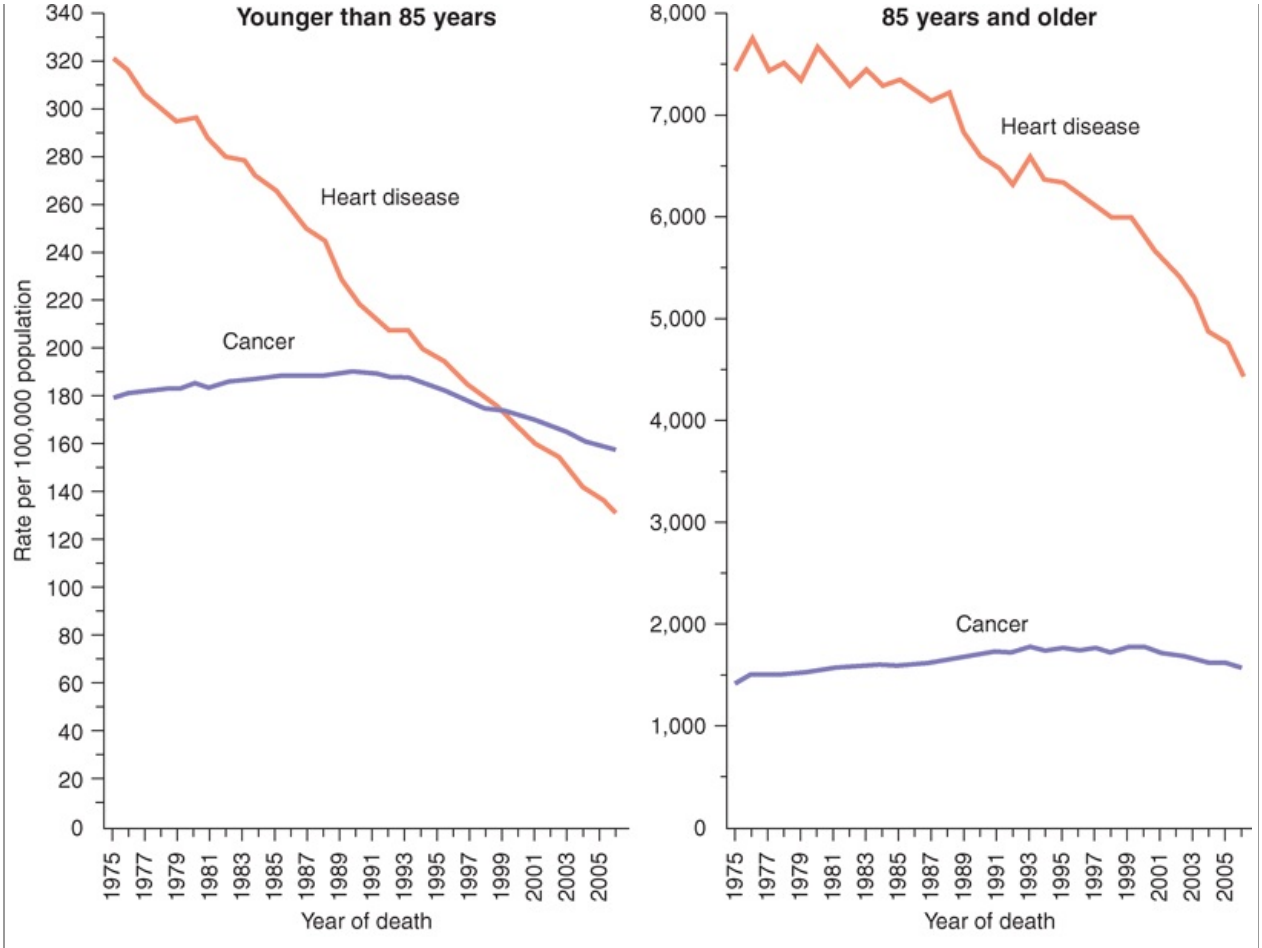

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Death rates for heart disease and cancer among people younger and older than age 85 years. ***A.*** In people younger than age 85 years, cancer has overtaken heart disease as the largest cause of death. ***B.*** In people older than age 85 years, heart disease is by far the major cause of death. *(From Jemal et al.)*

Figure 81-2

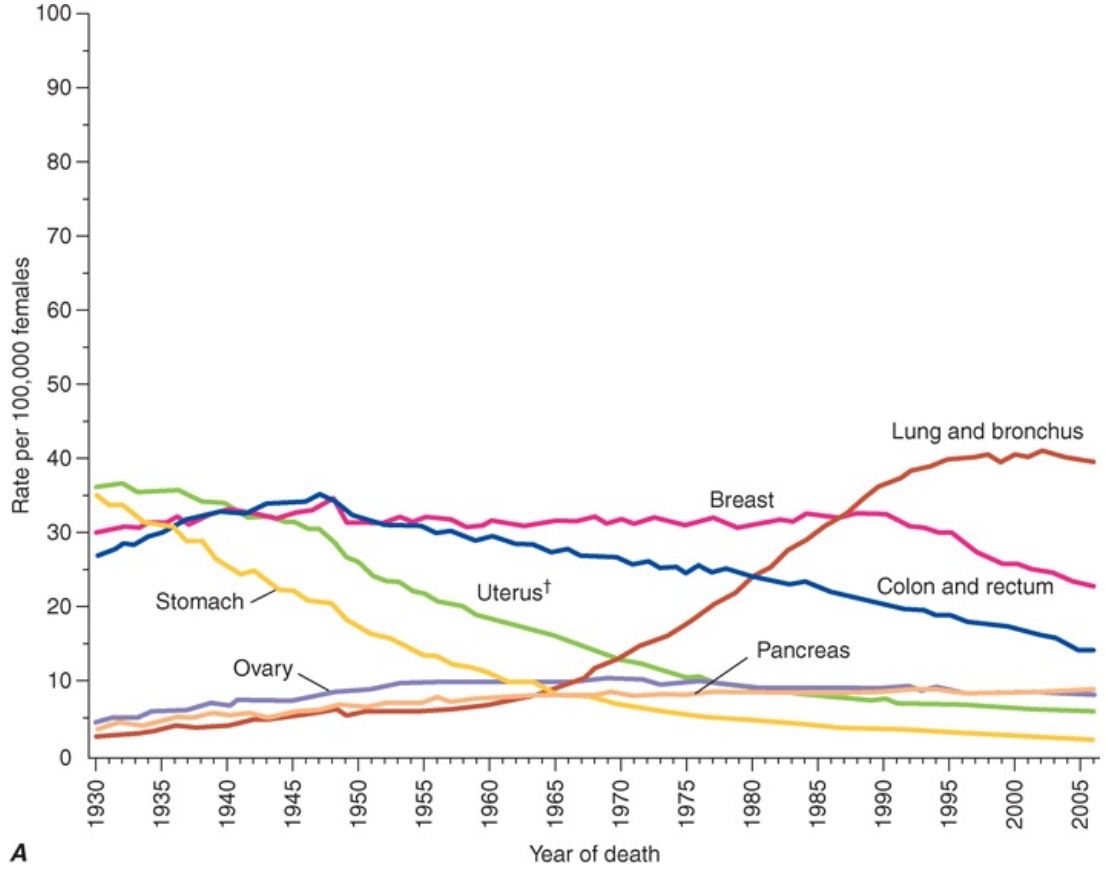

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

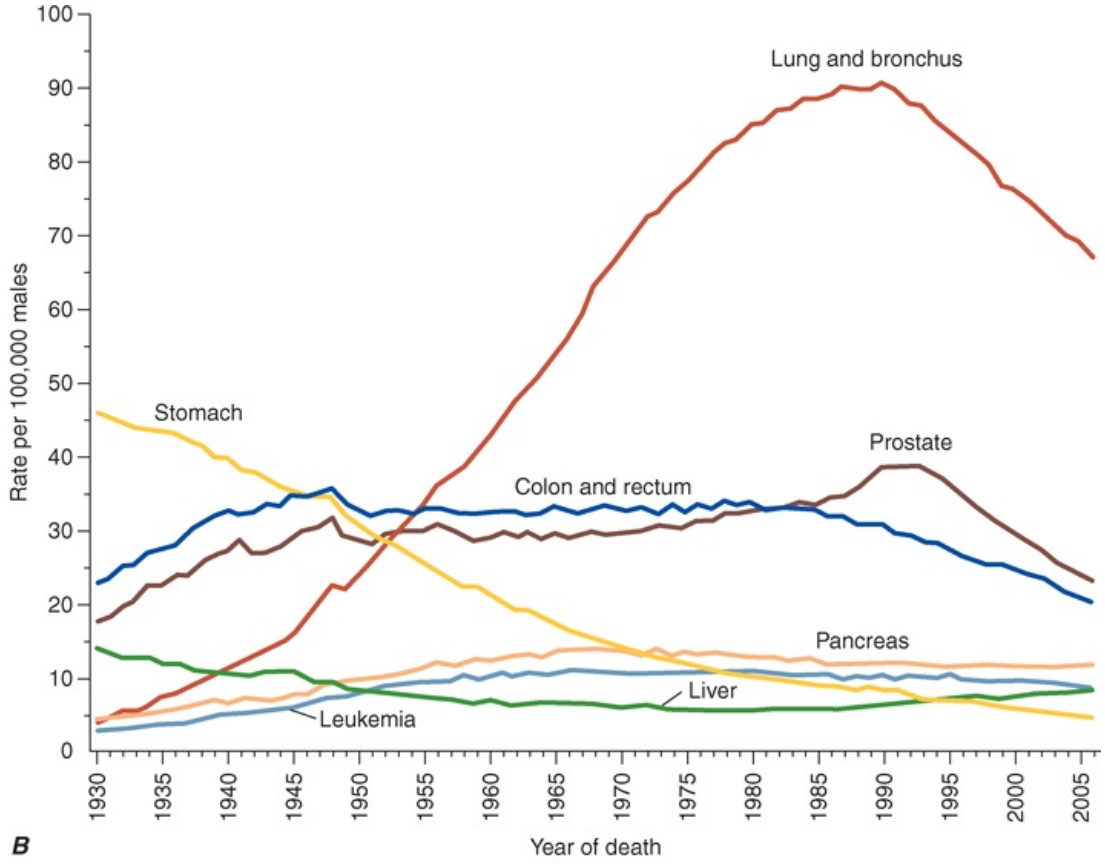

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sixty-five-year trend in cancer death rates for (***A***) women and (***B***) men by site in the United States, 1930–2006. Rates are per 100,000 age-adjusted to the 2000 U.S. standard population. *(From*

Jemal et al.)

Table 81–2 The Five Leading Primary Tumor Sites for Patients Dying of Cancer Based on Age and Sex in 2007

		Age, years					
Rank		**All Ages**	**Under 20**	**20–39**	**40–59**	**60–79**	**>80**
1	M	Lung	Leukemia	Leukemia	Lung	Lung	Lung
	F	Lung	Leukemia	Breast	Breast	Lung	Lung
2	M	Prostate	CNS	CNS	Colorectal	Colorectal	Prostate
	F	Breast	CNS	Cervix	Lung	Breast	Colorectal
3	M	Colorectal	Bone sarcoma	Colorectal	Liver	Prostate	Colorectal
	F	Colorectal	Endocrine	Leukemia	Colorectal	Colorectal	Breast
4	M	Pancreas	Endocrine	Lymphoma	Pancreas	Pancreas	Bladder
	F	Pancreas	Bone sarcoma	Colorectal	Ovary	Pancreas	Pancreas
5	M	Leukemia	Soft tissue sarcoma	Lung	Esophagus	Esophagus	Pancreas
	F	Ovary	Soft tissue sarcoma	CNS	Pancreas	Ovary	Lymphoma

Abbreviations: M, male; F, female.

Table 81–3 Cancer Incidence and Mortality in Racial and Ethnic Groups, U.S., 2002–2006

Site		**White**	**Black**	**Asian/Pacific Islander**	**American Indian**	**Hispanic**
Incidence per 100,000 Population						
All	M	550.1	626.8	334.5	318.4	430.3
	F	420.0	389.5	276.3	265.1	326.8
Breast		123.5	113.0	81.6	67.2	90.2
Colorectal	M	58.2	68.4	44.1	38.1	50.0
	F	42.6	51.7	33.1	30.7	35.1
Kidney	M	19.7	20.6	9.0	16.6	18.2
	F	10.3	10.6	4.5	10.6	10.3
Liver	M	8.0	12.5	21.4	8.9	15.9
	F	2.8	3.8	8.1	4.6	6.2
Lung	M	85.9	104.8	50.6	57.9	49.2
	F	57.1	50.7	27.6	41.3	26.5
Prostate		146.3	231.9	82.3	82.7	131.1
Deaths per 100,000 Population						
All	M	226.7	304.2	135.4	183.3	154.8
	F	157.3	183.7	95.1	140.1	103.9
Breast		23.9	33.0	12.5	17.6	15.5
Colorectal	M	21.4	31.4	13.8	20.0	16.2
	F	14.9	21.6	10.0	13.7	10.7
Kidney	M	6.1	6.0	2.4	9.0	5.2
	F	2.8	2.7	1.2	4.2	2.4
Liver	M	6.8	10.8	15.0	10.3	11.2
	F	2.9	3.9	6.6	6.5	5.1
Lung	M	69.9	90.1	36.9	48.0	33.9
	F	41.9	40.0	18.2	33.5	14.4
Prostate		23.6	56.3	10.6	20.0	19.6

Abbreviations: M, male; F, female.

Cancer Around the World

In 2002, 11 million new cancer cases and 7 million cancer deaths were estimated worldwide. When broken down by region of the world, ~45% of cases were in Asia, 26% in Europe, 14.5% in North America, 7.1% in Central/South America, 6% in Africa, and 1% in Australia/New Zealand **(Fig. 81-3)**.

Lung cancer is the most common cancer and the most common cause of cancer death in the world. Its incidence is highly variable, affecting only 2 per 100,000 African women but as many as 61 per 100,000 North American men. Breast cancer is the second most common cancer worldwide; however, it ranks fifth as a cause of death behind lung, stomach, liver, and colorectal cancer. Among the eight most common forms of cancer, lung (2-fold), breast (3-fold), prostate (2.5-fold), and colorectal (3-fold) cancers are more common in more developed countries than in less developed countries. By contrast, liver (2-fold), cervical (2-fold), and esophageal (2- to 3-fold) cancers are more common in less developed countries. Stomach cancer incidence is similar in more and less developed countries but is much more common in Asia than North America or Africa. The most common cancers in Africa are cervical, breast, and liver cancers. It has been estimated that nine modifiable risk factors are responsible for more than one-third of cancers worldwide. These include smoking, alcohol consumption, obesity, physical inactivity, low fruit and vegetable consumption, unsafe sex, air pollution, indoor smoke from household fuels, and contaminated injections.

Figure 81-3

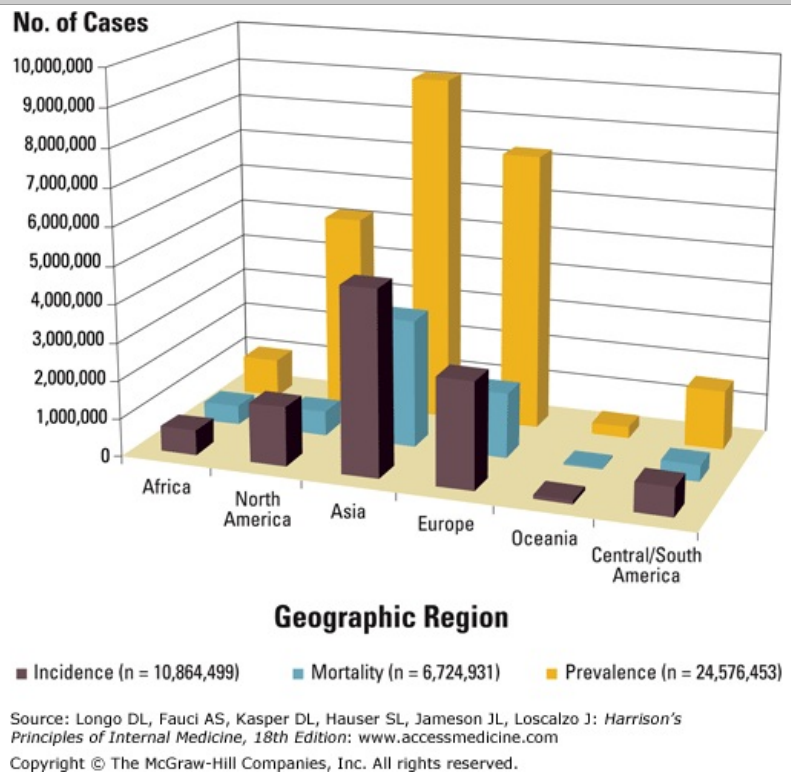

Worldwide overall annual cancer incidence, mortality and 5-year prevalence for the period of 1993–2001. *(From Kamangar et al.)*

PATIENT MANAGEMENT

Important information is obtained from every portion of the routine history and physical examination. The duration of symptoms may reveal the chronicity of disease. The past medical history may alert the physician to the presence of underlying diseases that may affect the choice of therapy or the side effects of treatment. The social history may reveal occupational exposure to carcinogens or habits, such as smoking or alcohol consumption, that may influence the course of disease and its treatment. The family history may suggest an underlying familial cancer predisposition and point out the need to begin surveillance or other preventive therapy for unaffected siblings of the patient. The review of systems may suggest early symptoms of metastatic disease or a paraneoplastic syndrome.

Diagnosis

The diagnosis of cancer relies most heavily on invasive tissue biopsy and should never be made without obtaining tissue; no noninvasive diagnostic test is sufficient to define a disease process as cancer. Although in rare clinical settings (e.g., thyroid nodules) fine-needle aspiration is an acceptable diagnostic procedure, the diagnosis generally depends on obtaining adequate tissue to permit careful evaluation of the histology of the tumor, its grade, and its invasiveness and to yield further molecular diagnostic information, such as the expression of cell-surface markers or intracellular proteins that typify a particular cancer, or the presence of a molecular marker, such as the t(8;14) translocation of Burkitt's lymphoma. Increasing evidence links the expression of certain genes with the prognosis and response to therapy (Chaps. 83 and 84).

Occasionally a patient will present with a metastatic disease process that is defined as cancer on

biopsy but has no apparent primary site of disease. Efforts should be made to define the primary site based on age, sex, sites of involvement, histology and tumor markers, and personal and family history. Particular attention should be focused on ruling out the most treatable causes (Chap. 99).

Once the diagnosis of cancer is made, the management of the patient is best undertaken as a multidisciplinary collaboration among the primary care physician, medical oncologists, surgical oncologists, radiation oncologists, oncology nurse specialists, pharmacists, social workers, rehabilitation medicine specialists, and a number of other consulting professionals working closely with each other and with the patient and family.

Defining the Extent of Disease and the Prognosis

The first priority in patient management after the diagnosis of cancer is established and shared with the patient is to determine the extent of disease. The curability of a tumor usually is inversely proportional to the tumor burden. Ideally, the tumor will be diagnosed before symptoms develop or as a consequence of screening efforts (Chap. 82). A very high proportion of such patients can be cured. However, most patients with cancer present with symptoms related to the cancer, caused either by mass effects of the tumor or by alterations associated with the production of cytokines or hormones by the tumor.

For most cancers, the extent of disease is evaluated by a variety of noninvasive and invasive diagnostic tests and procedures. This process is called *staging*. There are two types. *Clinical staging* is based on physical examination, radiographs, isotopic scans, CT scans, and other imaging procedures; *pathologic staging* takes into account information obtained during a surgical procedure, which might include intraoperative palpation, resection of regional lymph nodes and/or tissue adjacent to the tumor, and inspection and biopsy of organs commonly involved in disease spread. Pathologic staging includes histologic examination of all tissues removed during the surgical procedure. Surgical procedures performed may include a simple lymph node biopsy or more extensive procedures such as thoracotomy, mediastinoscopy, or laparotomy. Surgical staging may occur in a separate procedure or may be done at the time of definitive surgical resection of the primary tumor.

Knowledge of the predilection of particular tumors for spreading to adjacent or distant organs helps direct the staging evaluation.

Information obtained from staging is used to define the extent of disease either as localized, as exhibiting spread outside of the organ of origin to regional but not distant sites, or as metastatic to distant sites. The most widely used system of staging is the TNM (tumor, node, metastasis) system codified by the International Union Against Cancer and the American Joint Committee on Cancer. The TNM classification is an anatomically based system that categorizes the tumor on the basis of the size of the primary tumor lesion (T1–4, where a higher number indicates a tumor of larger size), the presence of nodal involvement (usually N0 and N1 for the absence and presence, respectively, of involved nodes, although some tumors have more elaborate systems of nodal grading), and the presence of metastatic disease (M0 and M1 for the absence and presence, respectively, of metastases). The various permutations of T, N, and M scores (sometimes including tumor histologic grade G) are then broken into stages, usually designated by the roman numerals I through IV. Tumor burden increases and curability decreases with increasing stage. Other anatomic staging systems are used for some tumors, e.g., the Dukes classification for colorectal cancers, the International Federation of Gynecologists and Obstetricians classification for gynecologic cancers, and the Ann Arbor classification for Hodgkin's disease.

Certain tumors cannot be grouped on the basis of anatomic considerations. For example, hematopoietic tumors such as leukemia, myeloma, and lymphoma are often disseminated at presentation and do not spread like solid tumors. For these tumors, other prognostic factors have been identified (Chaps. 109, 110, and 111).

In addition to tumor burden, a second major determinant of treatment outcome is the physiologic reserve of the patient. Patients who are bedridden before developing cancer are likely to fare worse, stage for stage, than fully active patients. Physiologic reserve is a determinant of how a patient is likely to cope with the physiologic stresses imposed by the cancer and its treatment. This factor is difficult to assess directly. Instead, surrogate markers for physiologic reserve are used, such as the patient's age or Karnofsky performance status **(Table 81–4)** or Eastern Cooperative Oncology Group (ECOG) performance status **(Table 81–5)**. Older patients and those with a Karnofsky performance status <70 or ECOG performance status ≥3 have a poor prognosis unless the poor performance is a reversible consequence of the tumor.

Table 81–4 Karnofsky Performance Index

Performance Status	Functional Capability of the Patient
100	Normal; no complaints; no evidence of disease
90	Able to carry on normal activity; minor signs or symptoms of disease
80	Normal activity with effort; some signs or symptoms of disease

70	Cares for self; unable to carry on normal activity or do active work
60	Requires occasional assistance but is able to care for most needs
50	Requires considerable assistance and frequent medical care
40	Disabled; requires special care and assistance
30	Severely disabled; hospitalization is indicated although death is not imminent
20	Very sick; hospitalization necessary; active supportive treatment is necessary
10	Moribund, fatal processes progressing rapidly
0	Dead

Table 81–5 The Eastern Cooperative Oncology Group (ECOG) Performance Scale

ECOG Grade 0: Fully active, able to carry on all predisease performance without restriction
ECOG Grade 1: Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light housework, office work
ECOG Grade 2: Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours
ECOG Grade 3: Capable of only limited self-care, confined to bed or chair more than 50% of waking hours
ECOG Grade 4: Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair
ECOG Grade 5: Dead

Source: From MM Oken et al: Am J Clin Oncol 5:649, 1982.

Increasingly, biologic features of the tumor are being related to prognosis. The expression of particular oncogenes, drug-resistance genes, apoptosis-related genes, and genes involved in metastasis are being found to influence response to therapy and prognosis. The presence of selected cytogenetic abnormalities may influence survival. Tumors with higher growth fractions, as assessed by expression of proliferation-related markers such as proliferating cell nuclear antigen, behave more aggressively than tumors with lower growth fractions. Information obtained from studying the tumor itself will increasingly be used to influence treatment decisions. Host genes involved in drug metabolism can influence the safety and efficacy of particular treatments.

Making a Treatment Plan

From information on the extent of disease and the prognosis and in conjunction with the patient's wishes, it is determined whether the treatment approach should be curative or palliative in intent. Cooperation among the various professionals involved in cancer treatment is of the utmost importance in treatment planning. For some cancers, chemotherapy or chemotherapy plus radiation therapy delivered before the use of definitive surgical treatment (so-called neoadjuvant therapy) may improve the outcome, as seems to be the case for locally advanced breast cancer and head and neck cancers. In certain settings in which combined modality therapy is intended, coordination among the medical oncologist, radiation oncologist, and surgeon is crucial to achieving optimal results. Sometimes the chemotherapy and radiation therapy need to be delivered sequentially, and other times concurrently. Surgical procedures may precede or follow other treatment approaches. It is best for the treatment plan either to follow a standard protocol precisely or else to be part of an ongoing clinical research protocol evaluating new treatments. Ad hoc modifications of standard protocols are likely to compromise treatment results.

The choice of treatment approaches was formerly dominated by the local culture in both the university and the practice settings. However, it is now possible to gain access electronically to standard treatment protocols and to every approved clinical research study in North America through a personal computer interface with the Internet.[1]

The skilled physician also has much to offer the patient for whom curative therapy is no longer an option. Often a combination of guilt and frustration over the inability to cure the patient and the pressure of a busy schedule greatly limit the time a physician spends with a patient who is receiving only palliative care. Resist these forces. In addition to the medicines administered to alleviate symptoms (see below), it is important to remember the comfort that is provided by holding the patient's hand, continuing regular examinations, and taking time to talk.

Management of Disease and Treatment Complications

Because cancer therapies are toxic (Chap. 85), patient management involves addressing complications of both the disease and its treatment as well as the complex psychosocial problems associated with cancer. In the short term during a course of curative therapy, the patient's functional status may decline. Treatment-induced toxicity is less acceptable if the goal of therapy is palliation. The most

common side effects of treatment are nausea and vomiting (see below), febrile neutropenia (Chap. 86), and myelosuppression (Chap. 85). Tools are now available to minimize the acute toxicity of cancer treatment.

New symptoms developing in the course of cancer treatment should always be assumed to be reversible until proven otherwise. The fatalistic attribution of anorexia, weight loss, and jaundice to recurrent or progressive tumor could result in a patient dying from a reversible intercurrent cholecystitis. Intestinal obstruction may be due to reversible adhesions rather than progressive tumor. Systemic infections, sometimes with unusual pathogens, may be a consequence of the immunosuppression associated with cancer therapy. Some drugs used to treat cancer or its complications (e.g., nausea) may produce central nervous system symptoms that look like metastatic disease or may mimic paraneoplastic syndromes such as the syndrome of inappropriate antidiuretic hormone. A definitive diagnosis should be pursued and may even require a repeat biopsy.

A critical component of cancer management is assessing the response to treatment. In addition to a careful physical examination in which all sites of disease are physically measured and recorded in a flow chart by date, response assessment usually requires periodic repeating of imaging tests that were abnormal at the time of staging. If imaging tests have become normal, repeat biopsy of previously involved tissue is performed to document complete response by pathologic criteria. Biopsies are not usually required if there is macroscopic residual disease. A *complete response* is defined as disappearance of all evidence of disease, and a *partial response* as >50% reduction in the sum of the products of the perpendicular diameters of all measurable lesions. The determination of partial response may also be based on a 30% decrease in the sums of the longest diameters of lesions (Response Evaluation Criteria in Solid Tumors, or RECIST, criteria). *Progressive disease* is defined as the appearance of any new lesion or an increase of >25% in the sum of the products of the perpendicular diameters of all measurable lesions (or an increase of 20% in the sums of the longest diameters by RECIST). Tumor shrinkage or growth that does not meet any of these criteria is considered *stable disease*. Some sites of involvement (e.g., bone) or patterns of involvement (e.g., lymphangitic lung or diffuse pulmonary infiltrates) are considered unmeasurable. No response is complete without biopsy documentation of their resolution, but partial responses may exclude their assessment unless clear objective progression has occurred.

[1] The National Cancer Institute maintains a database called PDQ (Physician Data Query) that is accessible on the Internet under the name CancerNet at *www.icic.nci.nih.gov/health.htm*. Information can be obtained through a facsimile machine using CancerFax by dialing 301-402-5874. Patient information is also provided by the National Cancer Institute in at least three formats: on the Internet via CancerNet at *www.icic.nci.nih.gov/patient.htm*, through the CancerFax number listed above, or by calling 1-800-4-CANCER. The quality control for the information provided through these services is rigorous.

Tumor markers may be useful in patient management in certain tumors. Response to therapy may be difficult to gauge with certainty. However, some tumors produce or elicit the production of markers that can be measured in the serum or urine, and in a particular patient, rising and falling levels of the marker are usually associated with increasing or decreasing tumor burden, respectively. Some clinically useful tumor markers are shown in **Table 81–6**. Tumor markers are not in themselves specific enough to permit a diagnosis of malignancy to be made, but once a malignancy has been diagnosed and shown to be associated with elevated levels of a tumor marker, the marker can be used to assess response to treatment.

Table 81–6 Tumor Markers

Tumor Markers	Cancer	Nonneoplastic Conditions
Hormones		
Human chorionic gonadotropin	Gestational trophoblastic disease, gonadal germ cell tumor	Pregnancy
Calcitonin	Medullary cancer of the thyroid	
Catecholamines	Pheochromocytoma	
Oncofetal Antigens		
α Fetoprotein	Hepatocellular carcinoma, gonadal germ cell tumor	Cirrhosis, hepatitis
Carcinoembryonic antigen	Adenocarcinomas of the colon, pancreas, lung, breast, ovary	Pancreatitis, hepatitis, inflammatory bowel disease, smoking
Enzymes		
Prostatic acid phosphatase	Prostate cancer	Prostatitis, prostatic hypertrophy
Neuron-specific enolase	Small cell cancer of the lung, neuroblastoma	

Lactate dehydrogenase	Lymphoma, Ewing's sarcoma	Hepatitis, hemolytic anemia, many others
Tumor-Associated Proteins		
Prostate-specific antigen	Prostate cancer	Prostatitis, prostatic hypertrophy
Monoclonal immunoglobulin	Myeloma	Infection, MGUS
CA-125	Ovarian cancer, some lymphomas	Menstruation, peritonitis, pregnancy
CA 19-9	Colon, pancreatic, breast cancer	Pancreatitis, ulcerative colitis
CD30	Hodgkin's disease, anaplastic large cell lymphoma	—
CD25	Hairy cell leukemia, adult T cell leukemia/lymphoma	—

Abbreviation: MGUS, monoclonal gammopathy of uncertain significance.

The recognition and treatment of depression are important components of management. The incidence of depression in cancer patients is ~25% overall and may be greater in patients with greater debility. This diagnosis is likely in a patient with a depressed mood (dysphoria) and/or a loss of interest in pleasure (anhedonia) for at least 2 weeks. In addition, three or more of the following symptoms are usually present: appetite change, sleep problems, psychomotor retardation or agitation, fatigue, feelings of guilt or worthlessness, inability to concentrate, and suicidal ideation. Patients with these symptoms should receive therapy. Medical therapy with a serotonin reuptake inhibitor such as fluoxetine (10–20 mg/d), sertraline (50–150 mg/d), or paroxetine (10–20 mg/d) or a tricyclic antidepressant such as amitriptyline (50–100 mg/d) or desipramine (75–150 mg/d) should be tried, allowing 4–6 weeks for response. Effective therapy should be continued at least 6 months after resolution of symptoms. If therapy is unsuccessful, other classes of antidepressants may be used. In addition to medication, psychosocial interventions such as support groups, psychotherapy, and guided imagery may be of benefit.

Many patients opt for unproven or unsound approaches to treatment when it appears that conventional medicine is unlikely to be curative. Those seeking such alternatives are often well educated and may be early in the course of their disease. Unsound approaches are usually hawked on the basis of unsubstantiated anecdotes and not only cannot help the patient but may be harmful. Physicians should strive to keep communications open and nonjudgmental, so that patients are more likely to discuss with the physician what they are actually doing. The appearance of unexpected toxicity may be an indication that a supplemental therapy is being taken.[2]

[2] Information about unsound methods may be obtained from the National Council Against Health Fraud, Box 1276, Loma Linda, CA 92354, or from the Center for Medical Consumers and Health Care Information, 237 Thompson Street, New York, NY 10012.

Long-Term Follow-Up/Late Complications

At the completion of treatment, sites originally involved with tumor are reassessed, usually by radiography or imaging techniques, and any persistent abnormality is biopsied. If disease persists, the multidisciplinary team discusses a new salvage treatment plan. If the patient has been rendered disease-free by the original treatment, the patient is followed regularly for disease recurrence. The optimal guidelines for follow-up care are not known. For many years, a routine practice has been to follow the patient monthly for 6–12 months, then every other month for a year, every 3 months for a year, every 4 months for a year, every 6 months for a year, and then annually. At each visit, a battery of laboratory and radiographic and imaging tests were obtained on the assumption that it is best to detect recurrent disease before it becomes symptomatic. However, where follow-up procedures have been examined, this assumption has been found to be untrue. Studies of breast cancer, melanoma, lung cancer, colon cancer, and lymphoma have all failed to support the notion that asymptomatic relapses are more readily cured by salvage therapy than symptomatic relapses. In view of the enormous cost of a full battery of diagnostic tests and their manifest lack of impact on survival, new guidelines are emerging for less frequent follow-up visits, during which the history and physical examination are the major investigations performed.

As time passes, the likelihood of recurrence of the primary cancer diminishes. For many types of cancer, survival for 5 years without recurrence is tantamount to cure. However, important medical problems can occur in patients treated for cancer and must be examined (Chap. 102). Some problems emerge as a consequence of the disease and some as a consequence of the treatment. An understanding of these disease- and treatment-related problems may help in their detection and management.

Despite these concerns, most patients who are cured of cancer return to normal lives.

Supportive Care

In many ways, the success of cancer therapy depends on the success of the supportive care. Failure to control the symptoms of cancer and its treatment may lead patients to abandon curative therapy. Of equal importance, supportive care is a major determinant of quality of life. Even when life cannot be prolonged, the physician must strive to preserve its quality. Quality-of-life measurements have become common endpoints of clinical research studies. Furthermore, palliative care has been shown to be cost-effective when approached in an organized fashion. A credo for oncology could be to cure sometimes, to extend life often, and to comfort always.

PAIN

Pain occurs with variable frequency in the cancer patient: 25–50% of patients present with pain at diagnosis, 33% have pain associated with treatment, and 75% have pain with progressive disease. The pain may have several causes. In ~70% of cases, pain is caused by the tumor itself—by invasion of bone, nerves, blood vessels, or mucous membranes or obstruction of a hollow viscus or duct. In ~20% of cases, pain is related to a surgical or invasive medical procedure, to radiation injury (mucositis, enteritis, or plexus or spinal cord injury), or to chemotherapy injury (mucositis, peripheral neuropathy, phlebitis, steroid-induced aseptic necrosis of the femoral head). In 10% of cases, pain is unrelated to cancer or its treatment.

Assessment of pain requires the methodical investigation of the history of the pain, its location, character, temporal features, provocative and palliative factors, and intensity (Chap. 11); a review of the oncologic history and past medical history as well as personal and social history; and a thorough physical examination. The patient should be given a 10-division visual analogue scale on which to indicate the severity of the pain. The clinical condition is often dynamic, making it necessary to reassess the patient frequently. Pain therapy should not be withheld while the cause of pain is being sought.

A variety of tools are available with which to address cancer pain. About 85% of patients will have pain relief from pharmacologic intervention. However, other modalities, including antitumor therapy (such as surgical relief of obstruction, radiation therapy, and strontium-89 or samarium-153 treatment for bone pain), neurostimulatory techniques, regional analgesia, or neuroablative procedures are effective in an additional 12% or so. Thus, very few patients will have inadequate pain relief if appropriate measures are taken. **A specific approach to pain relief is detailed in Chap. 9.**

NAUSEA

Emesis in the cancer patient is usually caused by chemotherapy (Chap. 85). Its severity can be predicted from the drugs used to treat the cancer. Three forms of emesis are recognized on the basis of their timing with regard to the noxious insult. *Acute emesis*, the most common variety, occurs within 24 h of treatment. *Delayed emesis* occurs 1–7 days after treatment; it is rare, but, when present, usually follows cisplatin administration. *Anticipatory emesis* occurs before administration of chemotherapy and represents a conditioned response to visual and olfactory stimuli previously associated with chemotherapy delivery.

Acute emesis is the best understood form. Stimuli that activate signals in the chemoreceptor trigger zone in the medulla, the cerebral cortex, and peripherally in the intestinal tract lead to stimulation of the vomiting center in the medulla, the motor center responsible for coordinating the secretory and muscle contraction activity that leads to emesis. Diverse receptor types participate in the process, including dopamine, serotonin, histamine, opioid, and acetylcholine receptors. The serotonin receptor antagonists ondansetron and granisetron are the most effective drugs against highly emetogenic agents, but they are expensive.

As with the analgesia ladder, emesis therapy should be tailored to the situation. For mildly and moderately emetogenic agents, prochlorperazine, 5–10 mg PO or 25 mg PR, is effective. Its efficacy may be enhanced by administering the drug before the chemotherapy is delivered. Dexamethasone, 10–20 mg IV, is also effective and may enhance the efficacy of prochlorperazine. For highly emetogenic agents such as cisplatin, mechlorethamine, dacarbazine, and streptozocin, combinations of agents work best and administration should begin 6–24 h before treatment. Ondansetron, 8 mg PO every 6 h the day before therapy and IV on the day of therapy, plus dexamethasone, 20 mg IV before treatment, is an effective regimen. Addition of oral aprepitant (a substance P/neurokinin 1 receptor antagonist) to this regimen (125 mg on day 1, 80 mg on days 2 and 3) further decreases the risk of both acute and delayed vomiting. Like pain, emesis is easier to prevent than to alleviate.

Delayed emesis may be related to bowel inflammation from the therapy and can be controlled with oral dexamethasone and oral metoclopramide, a dopamine receptor antagonist that also blocks serotonin receptors at high dosages. The best strategy for preventing anticipatory emesis is to control emesis in the early cycles of therapy to prevent the conditioning from taking place. If this is unsuccessful, prophylactic antiemetics the day before treatment may help. Experimental studies are evaluating behavior modification.

EFFUSIONS

Fluid may accumulate abnormally in the pleural cavity, pericardium, or peritoneum. Asymptomatic malignant effusions may not require treatment. Symptomatic effusions occurring in tumors responsive to systemic therapy usually do not require local treatment but respond to the treatment for the underlying tumor. Symptomatic effusions occurring in tumors unresponsive to systemic therapy may require local treatment in patients with a life expectancy of at least 6 months.

Pleural effusions due to tumors may or may not contain malignant cells. Lung cancer, breast cancer, and lymphomas account for ~75% of malignant pleural effusions. Their exudative nature is usually gauged by an effusion/serum protein ratio of ≥0.5 or an effusion/serum lactate dehydrogenase ratio of ≥0.6. When the condition is symptomatic, thoracentesis is usually performed first. In most cases, symptomatic improvement occurs for <1 month. Chest tube drainage is required if symptoms recur within 2 weeks. Fluid is aspirated until the flow rate is <100 mL in 24 h. Then either 60 units of bleomycin or 1 g of doxycycline is infused into the chest tube in 50 mL of 5% dextrose in water; the tube is clamped; the patient is rotated on four sides, spending 15 min in each position; and, after 1–2 h, the tube is again attached to suction for another 24 h. The tube is then disconnected from suction and allowed to drain by gravity. If <100 mL drains over the next 24 h, the chest tube is pulled, and a radiograph taken 24 h later. If the chest tube continues to drain fluid at an unacceptably high rate, sclerosis can be repeated. Bleomycin may be somewhat more effective than doxycycline but is very expensive. Doxycycline is usually the drug of first choice. If neither doxycycline nor bleomycin is effective, talc can be used.

Symptomatic pericardial effusions are usually treated by creating a pericardial window or by stripping the pericardium. If the patient's condition does not permit a surgical procedure, sclerosis can be attempted with doxycycline and/or bleomycin.

Malignantascites is usually treated with repeated paracentesis of small volumes of fluid. If the underlying malignancy is unresponsive to systemic therapy, peritoneovenous shunts may be inserted. Despite the fear of disseminating tumor cells into the circulation, widespread metastases are an unusual complication. The major complications are occlusion, leakage, and fluid overload. Patients with severe liver disease may develop disseminated intravascular coagulation.

NUTRITION

Cancer and its treatment may lead to a decrease in nutrient intake of sufficient magnitude to cause weight loss and alteration of intermediary metabolism. The prevalence of this problem is difficult to estimate because of variations in the definition of cancer cachexia, but most patients with advanced cancer experience weight loss and decreased appetite. A variety of both tumor-derived factors (e.g., bombesin, adrenocorticotropic hormone) and host-derived factors (e.g., tumor necrosis factor, interleukins 1 and 6, growth hormone) contribute to the altered metabolism, and a vicious cycle is established in which protein catabolism, glucose intolerance, and lipolysis cannot be reversed by the provision of calories.

It remains controversial how to assess nutritional status and when and how to intervene. Efforts to make the assessment objective have included the use of a prognostic nutritional index based on albumin levels, triceps skinfold thickness, transferrin levels, and delayed-type hypersensitivity skin testing. However, a simpler approach has been to define the threshold for nutritional intervention as >10% unexplained body weight loss, serum transferrin level <1500 mg/L (150 mg/dL), and serum albumin <34 g/L (3.4 g/dL).

The decision is important, because it appears that cancer therapy is substantially more toxic and less effective in the face of malnutrition. Nevertheless, it remains unclear whether nutritional intervention can alter the natural history. Unless some pathology is affecting the absorptive function of the gastrointestinal tract, enteral nutrition provided orally or by tube feeding is preferred over parenteral supplementation. However, the risks associated with the tube may outweigh the benefits. Megestrol acetate, a progestational agent, has been advocated as a pharmacologic intervention to improve nutritional status. Research in this area may provide more tools in the future as cytokine-mediated mechanisms are further elucidated.

PSYCHOSOCIAL SUPPORT

The psychosocial needs of patients vary with their situation. Patients undergoing treatment experience fear, anxiety, and depression. Self-image is often seriously compromised by deforming surgery and loss of hair. Women who receive cosmetic advice that enables them to look better also feel better. Loss of control over how one spends time can contribute to the sense of vulnerability. Juggling the demands of work and family with the demands of treatment may create enormous stresses. Sexual dysfunction is highly prevalent and needs to be discussed openly with the patient. An empathetic health care team is sensitive to the individual patient's needs and permits negotiation where such flexibility will not adversely affect the course of treatment.

Cancer survivors have other sets of difficulties. Patients may have fears associated with the termination of a treatment they associate with their continued survival. Adjustments are required to physical losses and handicaps, real and perceived. Patients may be preoccupied with minor physical problems. They perceive a decline in their job mobility and view themselves as less desirable workers.

They may be victims of job and/or insurance discrimination. Patients may experience difficulty reentering their normal past life. They may feel guilty for having survived and may carry a sense of vulnerability to colds and other illnesses. Perhaps the most pervasive and threatening concern is the ever-present fear of relapse (the Damocles syndrome).

Patients in whom therapy has been unsuccessful have other problems related to the end of life.

DEATH AND DYING

The most common causes of death in patients with cancer are infection (leading to circulatory failure), respiratory failure, hepatic failure, and renal failure. Intestinal blockage may lead to inanition and starvation. Central nervous system disease may lead to seizures, coma, and central hypoventilation. About 70% of patients develop dyspnea preterminally. However, many months usually pass between the diagnosis of cancer and the occurrence of these complications, and during this period the patient is severely affected by the possibility of death. The path of unsuccessful cancer treatment usually occurs in three phases. First, there is optimism at the hope of cure; when the tumor recurs, there is the acknowledgment of an incurable disease, and the goal of palliative therapy is embraced in the hope of being able to live with disease; finally, at the disclosure of imminent death, another adjustment in outlook takes place. The patient imagines the worst in preparation for the end of life and may go through stages of adjustment to the diagnosis. These stages include denial, isolation, anger, bargaining, depression, acceptance, and hope. Of course, patients do not all progress through all the stages or proceed through them in the same order or at the same rate. Nevertheless, developing an understanding of how the patient has been affected by the diagnosis and is coping with it is an important goal of patient management.

It is best to speak frankly with the patient and the family regarding the likely course of disease. These discussions can be difficult for the physician as well as for the patient and family. The critical features of the interaction are to reassure the patient and family that everything that can be done to provide comfort will be done. They will not be abandoned. Many patients prefer to be cared for in their homes or in a hospice setting rather than a hospital. The American College of Physicians has published a book called *Home Care Guide for Cancer: How to Care for Family and Friends at Home* that teaches an approach to successful problem-solving in home care. With appropriate planning, it should be possible to provide the patient with the necessary medical care as well as the psychological and spiritual support that will prevent the isolation and depersonalization that can attend in-hospital death.

The care of dying patients may take a toll on the physician. A "burnout" syndrome has been described that is characterized by fatigue, disengagement from patients and colleagues, and a loss of self-fulfillment. Efforts at stress reduction, maintenance of a balanced life, and setting realistic goals may combat this disorder.

END-OF-LIFE DECISIONS

Unfortunately, a smooth transition in treatment goals from curative to palliative may not be possible in all cases because of the occurrence of serious treatment-related complications or rapid disease progression. Vigorous and invasive medical support for a reversible disease or treatment complication is assumed to be justified. However, if the reversibility of the condition is in doubt, the patient's wishes determine the level of medical care. These wishes should be elicited before the terminal phase of illness and reviewed periodically. Information about advance directives can be obtained from the American Association of Retired Persons, 601 E Street, NW, Washington, DC 20049, 202-434-2277 or Choice in Dying, 250 West 57th Street, New York, NY 10107, 212-366-5540. **A full discussion of end-of-life management is in Chap. 9.**

FURTHER READINGS

Edge SB et al (eds): *American Joint Commission on Cancer Staging Manual*, 7th ed. New York, Springer, 2010

Eisenhauer EA et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 45:228, 2009[PMID: 19097774]

Green E et al: Cancer-related pain management: A report of evidence-based recommendations to guide practice. Clin J Pain 26:449, 2010[PMID: 20551720]

Jemal A et al: Cancer statistics, 2010. CA Cancer J Clin 60:277, 2010[PMID: 20610543]

Kamangar F et al: Patterns of cancer incidence, mortality, and prevalence across five continents: Defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol 24:2137, 2006[PMID: 16682732]

U.S. Department of Health and Human Services: *Clinical Practice Guideline Number 9, Management of Cancer Pain*. U.S. Department of Health and Human Services, Agency for Health Care Policy and Research publication no. 94-0592, 1994

Walsh D et al: The symptoms of advanced cancer: Relationship to age, gender, and performance status in 1000 patients. Support Care Cancer 8:175, 2000[PMID: 10789956]

Wickham R: Best practice management of CINV in oncology patients: II. Antiemetic guidelines and rationale for use. J Support Oncol 8:10, 2010[PMID: 20629453]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PREVENTION AND EARLY DETECTION OF CANCER: INTRODUCTION

Improved understanding of carcinogenesis has allowed cancer prevention and early detection (also known as cancer control) to expand beyond the identification and avoidance of carcinogens. Specific interventions to prevent cancer in those at risk, and effective screening for early detection of cancer, are the goals.

Carcinogenesis is not simply an event but a process, a continuum of discrete tissue and cellular changes over time resulting in more autonomous cellular processes. Prevention concerns the identification and manipulation of the biologic, environmental, and genetic factors in the causal pathway of cancer.

EDUCATION AND HEALTHFUL HABITS

Public education on the avoidance of identified risk factors for cancer and encouraging healthy habits contributes to cancer prevention and control. The clinician is a powerful messenger in this process. The patient-provider encounter provides an opportunity to teach patients about the hazards of smoking, the features of a healthy lifestyle, use of proven cancer screening methods, and sun avoidance.

Smoking Cessation

Tobacco smoking is a strong, modifiable risk factor for cardiovascular disease, pulmonary disease, and cancer. Smokers have an approximately 1 in 3 lifetime risk of dying prematurely from a tobacco-related cancer or cardiovascular or pulmonary disease. Tobacco use causes more deaths from cardiovascular disease than from cancer. Lung cancer and cancers of the larynx, oropharynx, esophagus, kidney, bladder, pancreas, and stomach are all tobacco-related.

The number of cigarettes smoked per day and the level of inhalation of cigarette smoke are correlated with risk of lung cancer mortality. Light- and low-tar cigarettes are not safer because smokers tend to inhale them more frequently and deeply.

Those who stop smoking have a 30–50% lower 10-year lung cancer mortality rate compared to those who continue smoking, despite the fact that some carcinogen-induced gene mutations persist for years after smoking cessation. Smoking cessation and avoidance have the potential to save more lives than any other public health activity.

The risk of tobacco smoke is not limited to the smoker. Environmental tobacco smoke, known as secondhand or passive smoke, causes lung cancer and other cardiopulmonary diseases in nonsmokers.

Tobacco prevention is a pediatric issue. More than 80% of adult American smokers began smoking before the age of 18 years. Approximately 20% of Americans in Grades 9 through 12 have smoked a cigarette in the past month. Counseling of adolescents and young adults is critical to prevent smoking. A clinician's simple advice to not start smoking or to quit smoking can be of benefit. Providers should query patients on tobacco use and offer smokers assistance in quitting.

Current approaches to smoking cessation recognize that smoking is an addiction (Chap. 395). The smoker who is quitting goes through a process with identifiable stages that include contemplation of quitting, an action phase in which the smoker quits, and a maintenance phase. Smokers who quit completely are more likely to be successful than those who gradually reduce the number of cigarettes smoked or change to lower-tar or lower-nicotine cigarettes. More than 90% of the Americans who have successfully quit smoking did so on their own, without participation in an organized cessation program, but cessation programs are helpful for some smokers. The Community Intervention Trial for Smoking Cessation (COMMIT) was a 4-year program showing that light smokers (<25 cigarettes per day) were more likely to benefit from simple cessation messages and cessation programs than those who did not receive an intervention. Quit rates were 30.6% in the intervention group and 27.5% in the control group. The COMMIT interventions were not successful in heavy smokers (>25 cigarettes per day). Heavy smokers may need an intensive broad-based cessation program that includes counseling, behavioral strategies, and pharmacologic adjuncts, such as nicotine replacement (gum, patches, sprays, lozenges, and inhalers), bupropion, and/or varenicline.

The health risks of cigars are similar to those of cigarettes. Smoking one or two cigars daily doubles the risk for oral and esophageal cancers; three or four cigars daily increases the risk of oral cancers

more than eightfold and esophageal cancer fourfold. The risks of occasional use are unknown.

Smokeless tobacco also represents a substantial health risk. Chewing tobacco is a carcinogen linked to dental caries, gingivitis, oral leukoplakia, and oral cancer. The systemic effects of smokeless tobacco (including snuff) may increase risks for other cancers. Esophageal cancer is linked to carcinogens in tobacco dissolved in saliva and swallowed.

Physical Activity

Physical activity is associated with a decreased risk of colon and breast cancer. A variety of mechanisms have been proposed. However, such studies are prone to confounding factors such as recall bias, association of exercise with other health-related practices, and effects of preclinical cancers on exercise habits (reverse causality).

Diet Modification

International epidemiologic studies suggest that diets high in fat are associated with increased risk for cancers of the breast, colon, prostate, and endometrium. These cancers have their highest incidence and mortalities in western cultures, where fat comprises an average of one-third of the total calories consumed.

Despite correlations, dietary fat has not been proven to cause cancer. Case-control and cohort epidemiologic studies give conflicting results. In addition, diet is a highly complex exposure to many nutrients and chemicals. Low-fat diets are associated with many dietary changes beyond simple subtraction of fat. Other lifestyle changes are also associated with adherence to a low-fat diet.

In observational studies, dietary fiber is associated with a reduced risk of colonic polyps and invasive cancer of the colon. However, cancer-protective effects of increasing fiber and lowering dietary fat have not been proven in the context of a prospective clinical trial. The putative protective mechanisms are complex and speculative. Fiber binds oxidized bile acids and generates soluble fiber products, such as butyrate, that may have differentiating properties. Fiber does not increase bowel transit times. High-fiber diets could lower the risk of breast and prostate cancer by absorbing and inactivating dietary estrogenic and androgenic cancer promoters. However, two large prospective cohort studies of >100,000 health professionals showed no association between fruit and vegetable intake and risk of cancer.

The Polyp Prevention Trial randomly assigned 2000 elderly persons, who had polyps removed, to a low-fat, high-fiber diet versus routine diet for 4 years. No differences were noted in polyp formation.

The U.S. National Institutes of Health Women's Health Initiative, launched in 1994, was a long-term clinical trial enrolling >100,000 women aged 45–69 years. It placed women in 22 intervention groups. Participants received calcium/vitamin D supplementation; hormone-replacement therapy; and counseling to increase exercise, eat a low-fat diet with increased consumption of fruits, vegetables, and fiber, and cease smoking. The study showed that while dietary fat intake was lower in the diet intervention group, invasive breast cancers were not reduced over an 8-year follow-up period compared to the control group. No reduction was seen in the incidence of colorectal cancer in the dietary intervention arm. The difference in dietary fat averaged ~10% between the two groups. Evidence does not currently establish the anticarcinogenic value of vitamin, mineral, or nutritional supplements in amounts greater than those provided by a balanced diet.

Energy Balance

Risk of cancer appears to increase as body mass index increases to more than 25 kg/m^2. Obesity is associated with increased risk for cancers of the colon, breast (female postmenopausal), endometrium, kidney (renal cell), and esophagus, although causality has not been established.

In observational studies, relative risks of colon cancer are increased in obesity by 1.5–2 for men and 1.2–1.5 for women. Obese postmenopausal women have a 30–50% increased risk of breast cancer. A hypothesis for the association is that adipose tissue serves as a depot for aromatase that facilitates estrogen production.

Sun Avoidance

Nonmelanoma skin cancers (basal cell and squamous cell) are induced by cumulative exposure to ultraviolet (UV) radiation. Intermittent acute sun exposure and sun damage have been linked to melanoma, but the evidence is inconsistent. Sunburns, especially in childhood and adolescence, may be associated with an increased risk of melanoma in adulthood. Reduction of sun exposure through use of protective clothing and changing patterns of outdoor activities can reduce skin cancer risk. Sunscreens decrease the risk of actinic keratoses, the precursor to squamous cell skin cancer, but melanoma risk may not be reduced. Sunscreens prevent burning, but they may encourage more prolonged exposure to the sun and may not filter out wavelengths of energy that cause melanoma.

Educational interventions to help individuals accurately assess their risk of developing skin cancer have some impact. Self-examination for skin pigment characteristics associated with skin cancer, such as freckling, may be useful in identifying people at high risk. Those who recognize themselves as

being at risk tend to be more compliant with sun-avoidance recommendations. Risk factors for melanoma include a propensity to sunburn, a large number of benign melanocytic nevi, and atypical nevi.

CANCER CHEMOPREVENTION

Chemoprevention involves the use of specific natural or synthetic chemical agents to reverse, suppress, or prevent carcinogenesis before the development of invasive malignancy.

Cancer develops through an accumulation of tissue abnormalities associated with genetic and epigenetic changes that are potential points of intervention to prevent cancer. The initial changes are termed *initiation*. The alteration can be inherited or acquired through the action of physical, infectious, or chemical carcinogens. Like most human diseases, cancer arises from an interaction between genetics and environmental exposures (Table 82–1). Influences that cause the initiated cell to progress through the carcinogenic process and change phenotypically are termed *promoters*. Promoters include hormones such as androgens, linked to prostate cancer, and estrogen, linked to breast and endometrial cancer. The distinction between an initiator and promoter is sometimes arbitrary; some components of cigarette smoke are "complete carcinogens," acting as both initiators and promoters. Cancer can be prevented or controlled through interference with the factors that cause cancer initiation, promotion, or progression. Compounds of interest in chemoprevention often have antimutagenic, hormone modulation, anti-inflammatory, antiproliferative, or pro-apoptotic activity (or a combination).

Table 82–1 Suspected Carcinogens

Carcinogens[a]	Associated Cancer or Neoplasm
Alkylating agents	Acute myeloid leukemia, bladder cancer
Androgens	Prostate cancer
Aromatic amines (dyes)	Bladder cancer
Arsenic	Cancer of the lung, skin
Asbestos	Cancer of the lung, pleura, peritoneum
Benzene	Acute myelocytic leukemia
Chromium	Lung cancer
Diethylstilbestrol (prenatal)	Vaginal cancer (clear cell)
Epstein-Barr virus	Burkitt's lymphoma, nasal T cell lymphoma
Estrogens	Cancer of the endometrium, liver, breast
Ethyl alcohol	Cancer of the liver, esophagus, head and neck
Helicobacter pylori	Gastric cancer, gastric MALT lymphoma
Hepatitis B or C virus	Liver cancer
Human immunodeficiency virus	Non-Hodgkin's lymphoma, Kaposi's sarcoma, squamous cell carcinomas (especially of the urogenital tract)
Human papilloma virus	Cervix cancer, head and neck cancer
Human T cell lymphotropic virus type I (HTLV-I)	Adult T cell leukemia/lymphoma
Immunosuppressive agents (azathioprine, cyclosporine, glucocorticoids)	Non-Hodgkin's lymphoma
Ionizing radiation (therapeutic or diagnostic)	Breast, bladder, thyroid, soft tissue, bone, hematopoietic, and many more
Nitrogen mustard gas	Cancer of the lung, head and neck, nasal sinuses
Nickel dust	Cancer of the lung, nasal sinuses
Phenacetin	Cancer of the renal pelvis and bladder
Polycyclic hydrocarbons	Cancer of the lung, skin (especially squamous cell carcinoma of scrotal skin)
Schistosomiasis	Bladder cancer (squamous cell)
Sunlight (ultraviolet)	Skin cancer (squamous cell and melanoma)
Tobacco (including smokeless)	Cancer of the upper aerodigestive tract, bladder
Vinyl chloride	Liver cancer (angiosarcoma)

[a]Agents that are thought to act as cancer initiators and/or promoters.

Chemoprevention of Cancers of the Upper Aerodigestive Tract

Smoking causes diffuse epithelial injury in the oral cavity, neck, esophagus, and lung. Patients cured of squamous cell cancers of the lung, esophagus, oral cavity, and neck are at risk (as high as 5% per year) of developing second cancers of the upper aerodigestive tract. Cessation of cigarette smoking does not markedly decrease the cured cancer patient's risk of second malignancy, even though it does lower the cancer risk in those who have never developed a malignancy. Smoking cessation may halt the early stages of the carcinogenic process (such as metaplasia), but it may have no effect on late stages of carcinogenesis. This "field carcinogenesis" hypothesis for upper aerodigestive tract cancer has made "cured" patients an important population for chemoprevention of second malignancies.

Oral human papilloma virus (HPV) infection, particularly HPV-16, increases the risk for cancers of the oropharynx. This association exists even in the absence of other risk factors such as smoking or alcohol use (although the magnitude of increased risk appears greater than additive when HPV infection and smoking are both present). Oral HPV infection is believed to be largely sexually acquired. The introduction of the HPV vaccine might eventually reduce oropharyngeal cancer rates.

Oral leukoplakia, a premalignant lesion commonly found in smokers, has been used as an intermediate marker allowing demonstration of chemopreventive activity in smaller shorter-duration, randomized, placebo-controlled trials. Response was associated with upregulation of retinoic acid receptor-β (RAR-β). Therapy with high, relatively toxic doses of isotretinoin (13-*cis*-retinoic acid) causes regression of oral leukoplakia. However, the lesions recur when the therapy is withdrawn, suggesting the need for long-term administration. More tolerable doses of isotretinoin have not proven beneficial in the prevention of head and neck cancer. Isotretinoin also failed to prevent second malignancies in patients cured of early-stage non-small cell lung cancer; mortality rates were actually increased in current smokers.

Several large-scale trials have assessed agents in the chemoprevention of lung cancer in patients at high risk. In the α-tocopherol/β-carotene (ATBC) Lung Cancer Prevention Trial, participants were male smokers, ages 50–69 years at entry. Participants had smoked an average of one pack of cigarettes per day for 35.9 years. Participants received α-tocopherol, β-carotene, and/or placebo in a randomized, two-by-two factorial design. After median follow-up of 6.1 years, lung cancer incidence and mortality were statistically significantly increased in those receiving β-carotene. α-Tocopherol had no effect on lung cancer mortality, and no evidence suggested interaction between the two drugs. Patients receiving α-tocopherol had a higher incidence of hemorrhagic stroke.

The β-Carotene and Retinol Efficacy Trial (CARET) involved 17,000 American smokers and workers with asbestos exposure. Entrants were randomly assigned to one of four arms and received β-carotene, retinol, and/or placebo in a two-by-two factorial design. This trial also demonstrated harm from β-carotene: a lung cancer rate of 5 per 1000 subjects per year for those taking placebo and of 6 per 1000 subjects per year for those taking β-carotene.

The ATBC and CARET results demonstrate the importance of testing chemoprevention hypotheses thoroughly before their widespread implementation as the results contradict a number of observational studies. The Physicians' Health Trial showed no change in the risk of lung cancer for those taking β-carotene; however, fewer of its participants were smokers than those in the ATBC and CARET studies.

Chemoprevention of Colon Cancer

Many colon cancer prevention trials are based on the premise that most colorectal cancers develop from adenomatous polyps. These trials use adenoma recurrence or disappearance as a surrogate endpoint (not yet validated) for colon cancer prevention. Early clinical trial results suggest that nonsteroidal anti-inflammatory drugs (NSAIDs), such as piroxicam, sulindac, and aspirin, may prevent adenoma formation or cause regression of adenomatous polyps. The mechanism of action of NSAIDs is unknown, but they are presumed to work through the cyclooxygenase pathway. Pooled findings from observational cohort studies demonstrate a relative reduction in colorectal cancer incidence of approximately 22%, and a relative reduction in colorectal adenoma incidence of about 28%, with regular aspirin use; however, in two randomized controlled trials (the Physicians' Health Study and the Women's Health Study), aspirin had no effect on colon cancer or adenoma incidence in persons with no previous history of colonic lesions, at up to 10 years of therapy. The randomized controlled trials did show an approximately 18% relative risk reduction for colonic adenoma incidence in persons with a previous history of adenomas after 1 year's therapy.

Cyclooxygenase-2 (COX-2) inhibitors have also been considered for colorectal cancer and polyp prevention. Trials with COX-2 inhibitors were initiated but an increased risk of cardiovascular events in those taking the COX-2 inhibitors was noted, suggesting that these agents are not suitable for chemoprevention in the general population.

Epidemiologic studies suggest that diets high in calcium lower colon cancer risk. Calcium binds bile and fatty acids, which cause proliferation of colonic epithelium. It is hypothesized that calcium reduces intraluminal exposure to these compounds. The randomized controlled Calcium Polyp Prevention Study found that calcium supplementation decreased the absolute risk of adenomatous polyp recurrence by 7% at 4 years; extended observational follow-up demonstrated a 12% absolute risk

reduction 5 years after cessation of treatment. However, in the Women's Health Initiative, combined use of calcium carbonate and vitamin D twice daily did not reduce the incidence of invasive colorectal cancer compared with placebo after 7 years.

The Women's Health Initiative demonstrated that postmenopausal women taking estrogen plus progestin have a 44% lower risk of colorectal cancer compared to women taking placebo. Of >16,600 women randomized and followed for a median of 5.6 years, 43 invasive colorectal cancers occurred in the hormone group and 72 in the placebo group. The positive effect on colon cancer is mitigated by the modest increase in cardiovascular and breast cancer risks associated with combined estrogen plus progestin therapy.

A case-control study suggested that statins decrease the incidence of colorectal cancer; however, several subsequent case-control and cohort studies have not demonstrated an association between regular statin use and a reduced risk of colorectal cancer. No randomized controlled trials have addressed this hypothesis. A meta-analysis of statin use showed no protective effect of statins on overall cancer incidence or death.

Chemoprevention of Breast Cancer

Tamoxifen is an antiestrogen with partial estrogen agonistic activity in some tissues, such as endometrium and bone. One of its actions is to upregulate transforming growth factor β, which decreases breast cell proliferation. In randomized placebo-controlled trials to assess tamoxifen as adjuvant therapy for breast cancer, tamoxifen reduced the number of new breast cancers in the opposite breast by more than a third. In a randomized placebo-controlled prevention trial involving >13,000 women at high risk, tamoxifen decreased the risk of developing breast cancer by 49% (from 43.4 to 22 per 1000 women) after a median follow-up of nearly 6 years. Tamoxifen also reduced bone fractures; a small increase in risk of endometrial cancer, stroke, pulmonary emboli, and deep vein thrombosis was noted. The International Breast Cancer Intervention Study (IBIS-I) and the Italian Randomized Tamoxifen Prevention Trial also demonstrated a reduction in breast cancer incidence with tamoxifen use. Tamoxifen has been approved by the U.S. Food and Drug Administration for reduction of breast cancer in women at high risk for the disease (1.66% risk at 5 years based on the Gail risk model: *http://www.nci.nih.gov/cancertopics/pdq/genetics/breast-and-ovarian/healthprofessional#Section_66*).

A trial comparing tamoxifen with another selective estrogen receptor modulator, raloxifene, showed that raloxifene is comparable to tamoxifen in cancer prevention. This trial only included postmenopausal women. Raloxifene was associated with more noninvasive breast cancer than tamoxifen; the drugs are similar in risks of other cancers, fractures, ischemic heart disease, and stroke. Because the aromatase inhibitors are even more effective than tamoxifen in adjuvant breast cancer therapy, it is hoped that they would be more effective in breast cancer prevention. However, no data are yet available on this point.

Chemoprevention of Prostate Cancer

Finasteride is a 5-α-reductase inhibitor. It inhibits conversion of testosterone to dihydrotestosterone (DHT), a potent stimulator of prostate cell proliferation. The Prostate Cancer Prevention Trial (PCPT) randomly assigned men aged 55 years or older at average risk of prostate cancer to finasteride or placebo. All men in the trial were being regularly screened with PSA and digital rectal examination. After 7 years of therapy, the incidence of prostate cancer was 18.4% in the finasteride arm and 24.8% in the placebo arm, a statistically significant difference. However, the finasteride group had more patients with tumors of Gleason score 7 and higher compared to the placebo arm (6.4 vs 5.1%). The clinical significance of this finding, if any, is unknown. The observed increase in high-grade tumors was spurious and likely due to an increased sensitivity of PSA and digital rectal exam for high-grade tumors in men receiving finasteride.

Another 5-α-reductase inhibitor, dutasteride, has also been evaluated as a preventive agent for prostate cancer. The Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial was a randomized double-blind trial in which approximately 8200 men with an elevated PSA (2.5–10 ng/mL for men aged 50–60 years and 3–10 ng/mL for men aged 60 years or older) and negative prostate biopsy on enrollment received daily 0.5 mg dutasteride or placebo. A preliminary report from this trial noted a statistically significant 23% relative risk reduction in the incidence of biopsy-detected prostate cancer in the dutasteride arm at 4 years of treatment (659 cases vs 857 cases, respectively). Unlike the PCPT, no difference was observed in the rates of high-grade prostate cancer. Since all men in both the PCPT and REDUCE trials were being screened and since screening approximately doubles the rate of prostate cancer, it is not known if finasteride or dutasteride decrease the risk of prostate cancer in men who are not being screened.

Several favorable laboratory and observational studies led to the formal evaluation of selenium and α-tocopherol (vitamin E) as potential prostate cancer preventives. The Selenium and Vitamin E Cancer Prevention Trial (SELECT) assigned 35,533 men to receive 200 μg/d selenium, 400 IU/d α-tocopherol, selenium plus vitamin E, or placebo. After a median follow-up of 5.5 years, no significant difference in the prostate cancer incidence rate was observed for any group. In fact, compared to placebo, a trend

toward an increased risk of developing prostate cancer was observed for those men taking vitamin E alone (hazard ratio 1.13, 95% confidence interval, 0.99–1.29)

Vaccines and Cancer Prevention

A number of infectious agents cause cancer. Hepatitis B and C are linked to liver cancer; some human papilloma virus (HPV) strains are linked to cervical and head and neck cancer; and *Helicobacter pylori* is associated with gastric adenocarcinoma and gastric lymphoma. Vaccines to protect against these agents may reduce the risk of their associated cancers.

The hepatitis B vaccine is effective in preventing hepatitis and hepatomas due to chronic hepatitis B infection. Public health officials are encouraging widespread administration of the hepatitis B vaccine, especially in Asia, where the disease is epidemic.

A quadrivalent HPV vaccine (covering HPV strains 6, 11, 16, and 18) and a bivalent vaccine (covering HPV strains 16 and 18) are available for use in the United States. HPV types 16 and 18 cause cervical cancer, and types 6 and 11 cause genital papillomas. For females not previously infected with these HPV strains, the vaccines demonstrate high efficacy in preventing persistent strain-specific HPV infections. Trials that evaluated the vaccines' ability to prevent cervical cancer relied on surrogate outcome measures [cervical intraepithelial neoplasia (CIN) I, II, and III], and no cases of cervical cancer were observed in either the vaccine or control arms. The vaccines do not appear to impact preexisting infections; efficacy was markedly lower for populations that had previously been exposed to vaccine-specific HPV strains. The vaccine is recommended for girls and women ages 9–26 years. Reduction in these HPV types could prevent >70% of cervical cancers worldwide.

SURGICAL PREVENTION OF CANCER

Some organs in some individuals are at such high risk of developing cancer that surgical removal of the organ at risk may be considered. Women with severe cervical dysplasia are treated with conization and occasionally even hysterectomy. Colectomy is used to prevent colon cancer in patients with familial polyposis or ulcerative colitis.

Prophylactic bilateral mastectomy may be chosen for breast cancer prevention among women with genetic predisposition to breast cancer. In a prospective series of 139 women with *BRCA1* and *BRCA2* mutations, 76 chose to undergo prophylactic mastectomy and 63 chose close surveillance. At 3 years, no cases of breast cancer had been diagnosed in those opting for surgery, but 8 in the surveillance group had developed breast cancer. A larger (n = 639) retrospective cohort study reported that 3 patients developed breast cancer after prophylactic mastectomy compared with an expected incidence of 30–53 cases: a 90–94% reduction in breast cancer risk. The effect of the procedure on mortality is unknown.

Prophylactic oophorectomy may also be employed for the prevention of ovarian and breast cancers among high-risk women. A case-control study of women with *BRCA1* or *BRCA2* mutations found that 6 (2.8%) of 259 women who underwent bilateral prophylactic oophorectomy had stage I ovarian cancer at the time of surgery and 2 (0.8%) developed papillary serous peritoneal carcinoma over 9 years. By comparison, 58 (19.9%) of 292 women in the matched control group developed ovarian cancer: this corresponds to a 96% relative risk reduction for ovarian cancer with the use of prophylactic surgery. Studies of prophylactic oophorectomy for prevention of breast cancer in women with genetic mutations have shown relative risk reductions of approximately 50%.

At present, all of the evidence concerning the use of prophylactic mastectomy and oophorectomy for prevention of breast and ovarian cancer in high-risk women has been observational in nature; such studies are prone to a variety of biases, including case selection bias, family relationships between patients and controls, and inadequate information about hormone use. Thus, they may give an overestimate of the magnitude of benefit.

Orchiectomy is an effective method of androgen deprivation in prostate cancer.

Cancer Screening

Screening is a means of detecting disease early in asymptomatic individuals, with the goal of decreasing morbidity and mortality. While screening can potentially reduce disease-specific deaths and has been shown to do so in cervical, colon, and breast cancer, it is also subject to a number of biases that can suggest a benefit when actually there is none. Biases can even mask net harm. Early detection does not in itself confer benefit. To be of value, screening must detect disease earlier, and treatment of earlier disease must yield a better outcome than treatment at the onset of symptoms. Cause-specific mortality, rather than survival after diagnosis, is the preferred endpoint (see below).

Because screening is done on asymptomatic, healthy persons, it should offer substantial likelihood of benefit that outweighs harm. Screening tests and their appropriate use should be carefully evaluated before their use is widely encouraged in screening programs, as a matter of public policy.

A large and increasing number of genetic mutations and nucleotide polymorphisms have been associated with an increased risk of cancer. Testing for these genetic mutations could in theory define

a high-risk population. However, most of the identified mutations have very low penetrance and individually provide minimal predictive accuracy. The ability to predict the development of a particular cancer may some day present therapeutic options as well as ethical dilemmas. It may eventually allow for early intervention to prevent a cancer or limit its severity. People at high risk may be ideal candidates for chemoprevention and screening; however, efficacy of these interventions in the high-risk population should be investigated. Currently, persons at high risk for a particular cancer can engage in intensive screening. While this course is clinically reasonable, it is not known if it saves lives in these populations.

THE ACCURACY OF SCREENING

A screening test's accuracy or ability to discriminate disease is described by four indices: sensitivity, specificity, positive predictive value, and negative predictive value (Table 82–2). *Sensitivity*, also called the true positive rate, is the proportion of persons with the disease who test positive in the screen (i.e., the ability of the test to detect disease when it is present). *Specificity*, or 1 minus the false positive rate, is the proportion of persons who do not have the disease and test negative in the screening test (i.e., the ability of a test to correctly identify that the disease is not present). The *positive predictive value* is the proportion of persons who test positive and actually have the disease. Similarly, *negative predictive value* is the proportion testing negative who do not have the disease. The sensitivity and specificity of a test are independent of the underlying prevalence (or risk) of the disease in the population screened, but the predictive values depend strongly on the prevalence of the disease.

Table 82–2 Assessment of the Value of a Diagnostic Test [a]

	Condition Present	Condition Absent
Positive test	*a*	*b*
Negative test	*c*	*d*
a = true positive *b* = false positive *c* = false negative *d* = true negative		
Sensitivity		The proportion of persons with the condition who test positive: $a/(a + c)$
Specificity		The proportion of persons without the condition who test negative: $d/(b + d)$
Positive predictive value (PPV)		The proportion of persons with a positive test who have the condition: $a/(a + b)$
Negative predictive value		The proportion of persons with a negative test who do not have the condition: $d/(c + d)$
Prevalence, sensitivity, and specificity determine PPV $PPV = \frac{\text{prevalence} \times \text{sensitivity}}{(\text{prevalence} \times \text{sensitivity}) + (1 - \text{prevalence})(1 - \text{specificity})}$		

[a] For diseases of low prevalence, such as cancer, poor specificity has a dramatic adverse effect on PPV such that only a small fraction of positive tests are true positives.

Screening is most beneficial, efficient, and economical when the target disease is common in the population being screened. To be valuable, the screening test should have a high specificity; sensitivity need not be very high.

POTENTIAL BIASES OF SCREENING TESTS

Common biases of screening are lead time, length-biased sampling, and selection. These biases can make a screening test seem beneficial when actually it is not (or even causes net harm). Whether beneficial or not, screening can create the false impression of an epidemic by increasing the number of cancers diagnosed. It can also produce a shift in proportion of patients diagnosed at an early stage and inflated survival statistics without reducing mortality (i.e., the number of deaths from a given cancer relative to the number of those at risk for the cancer). In such a case, the *apparent* duration of survival (measured from date of diagnosis) increases without lives being saved or life expectancy changed.

Lead-time bias occurs when a test does not influence the natural history of the disease; the patient is merely diagnosed at an earlier date. When lead-time bias occurs, survival *appears* increased, but life

is not really prolonged. The screening test only prolongs the time the subject is aware of the disease and spends as a patient.

Length-biased sampling occurs because screening tests generally can more easily detect slow-growing, less aggressive cancers than fast-growing cancers. Cancers diagnosed due to the onset of symptoms between scheduled screenings are on average more aggressive, and treatment outcomes are not as favorable. An extreme form of length bias sampling is termed *overdiagnosis*, the detection of "pseudo disease." The reservoir of some undetected slow-growing tumors is large. Many of these tumors fulfill the histologic criteria of cancer but will never become clinically significant or cause death. This problem is compounded by the fact that the most common cancers appear most frequently at ages when competing causes of death are more frequent.

Selection bias must be considered in assessing the results of any screening effort. The population most likely to seek screening may differ from the general population to which the screening test might be applied. In general, volunteers for studies are more health conscious and likely to have a better prognosis or lower mortality rate, irrespective of the screening result. This is termed the *healthy volunteer effect*.

POTENTIAL DRAWBACKS OF SCREENING

Risks associated with screening include harm caused by the screening intervention itself, harm due to the further investigation of persons with positive tests (both true and false positives), and harm from the treatment of persons with a true-positive result, even if life is extended by treatment. The diagnosis and treatment of cancers that would never have caused medical problems can lead to the harm of unnecessary treatment and give patients the anxiety of a cancer diagnosis. The psychosocial impact of cancer screening can also be substantial when applied to the entire population.

ASSESSMENT OF SCREENING TESTS

Good clinical trial design can offset some biases of screening and demonstrate the relative risks and benefits of a screening test. A randomized controlled screening trial with cause-specific mortality as the endpoint provides the strongest support for a screening intervention. Overall mortality should also be reported to detect an adverse effect of screening and treatment on other disease outcomes (e.g., cardiovascular disease). In a randomized trial, two like populations are randomly established. One is given the usual standard of care (which may be no screening at all) and the other receives the screening intervention being assessed. The two populations are compared over time. Efficacy for the population studied is established when the group receiving the screening test has a better cause-specific mortality rate than the control group. Studies showing a reduction in the incidence of advanced-stage disease, an improved survival, or a stage shift are weaker (and possibly misleading) evidence of benefit. These latter criteria are necessary but not sufficient to establish the value of a screening test.

Although a randomized, controlled screening trial provides the strongest evidence to support a screening test, it is not perfect. Unless the trial is population-based, it does not remove the question of generalizability to the target population. Screening trials generally involve thousands of persons and last for years. Less definitive study designs are therefore often used to estimate the effectiveness of screening practices. However, every non-randomized study design is subject to strong confounders. In descending order of strength, evidence may also be derived from the findings of internally controlled trials using intervention allocation methods other than randomization (e.g., allocation by birth date, date of clinic visit); the findings of cohort or case-control analytic observational studies; or the results of multiple time series studies with or without the intervention.

SCREENING FOR SPECIFIC CANCERS

Widespread screening for cervical, colon, and breast cancer is beneficial for certain age groups. A number of organizations have considered whether or not to endorse routine use of certain screening tests. Because these groups have not used the same criteria to judge whether a screening test should be endorsed, they have arrived at different recommendations. The American Cancer Society (ACS) and the U.S. Preventive Services Task Force (USPSTF) publish screening guidelines (Table 82–3); the American College of Physicians (ACP) and the American Academy of Family Practitioners (AAFP) generally follow/endorse the USPSTF recommendations. Special surveillance of those at high risk for a specific cancer because of a family history or a genetic risk factor may be prudent, but few studies have assessed the influence on mortality.

Table 82–3 Screening Recommendations for Asymptomatic Normal-Risk Subjects[a]

Test or Procedure	USPSTF	ACS
Sigmoidoscopy	Adults 50–75 years: every 5 years ("A")[b] Adults 76–85 years: "C"	Adults ≥50 years: Screen every 5 years

	Adults ≥85 years: "D"	
Fecal occult blood testing (FOBT)	Adults 50–75 years: Annually ("A") Adults 76–85 years: "C" Adults ≥85 years: "D"	Adults ≥50 years: Screen every year
Colonoscopy	Adults 50–75 years: every 10 years ("A") Adults 76–85 years: "C" Adults ≥85 years: "D"	Adults ≥50 years: Screen every 10 years
Fecal DNA testing	"I"	Adults ≥50 years: Screen, but interval uncertain
Fecal immunochemical testing (FIT)	"I"	Adults ≥50 years: Screen every year
CT colonography	"I"	Adults ≥50 years: Screen every 5 years
Digital rectal examination (DRE)	No recommendation	Men ≥50 years, with a 10-year life expectancy; men ≥45 years, if African-American, or men with a first-degree relative diagnosed with prostate cancer <65 years; ≥40, if has several relatives with prostate cancer <65 years: Discuss and offer (with PSA testing) annually
Prostate-specific antigen (PSA)	Men <75 years: "I" Men ≥75 years: "D"	As for DRE
Pap test	Women <65 years: Beginning 3 years after first intercourse or by age 21, screen at least every 3 years ("A") Women ≥65 years, with adequate, normal recent Pap screenings: "D" Women after total hysterectomy for noncancerous causes: "D"	Women <30 years: Beginning 3 years after first intercourse or by age 21. Yearly for standard Pap; every 2 years with liquid test. Women 30–70 years: Every 2–3 years if last 3 tests normal Women ≥70 years: May stop screening if no abnormal Pap in past 10 years Women after total hysterectomy for noncancerous causes: Do not screen
Breast self-examination	"D"	Women ≥20 years: Breast self-exam is an option
Breast clinical examination	Women ≥40 years: "I" (as a stand-alone without mammography)	Women 20–40 years: Perform every 3 years Women ≥40 years: Perform annually
Mammography	Women 40–49 years: The decision should be an individual one, and take patient context into account ("C") Women 50–74 years: every 2 years ("B") Women ≥75 years: ("I")	Women ≥40 years: Screen annually
Magnetic resonance imaging (MRI)	"I"	Women >20% lifetime risk of breast cancer: Screen with MRI plus mammography annually Women 15–20% lifetime risk of breast cancer: Discuss option of MRI plus mammography annually Women <15% lifetime risk of breast cancer: Do not screen annually with MRI
Complete skin examination	"I"	Self-examination monthly; clinical exam as part of routine cancer-related checkup

[a]Summary of the screening procedures recommended for the general population by the U.S. Preventive Services Task Force and the American Cancer Society. These recommendations refer to

asymptomatic persons who have no risk factors, other than age or gender, for the targeted condition. [b]USPSTF lettered recommendations are defined as follows: "A": The USPSTF strongly recommends that clinicians provide (the service) to eligible patients; "B": The USPSTF recommends that clinicians provide (this service) to eligible patients; "C": The USPSTF makes no recommendation for or against routine provision of (the service); "D": The USPSTF recommends against routinely providing [the service] to asymptomatic patients; "I": The USPSTF concludes that the evidence is insufficient to recommend for or against routinely providing (the service). ***Abbreviations:*** ACS, American Cancer Society; USPSTF, U.S. Preventive Services Task Force.

Breast Cancer

Breast self-examination, clinical breast examination by a caregiver, mammography, and MRI have all been variably advocated as useful screening tools.

A number of trials have suggested that annual or biennial screening with mammography or mammography plus clinical breast examination in normal-risk women older than age 50 years decreases breast cancer mortality. Each trial has been criticized for design flaws. In most trials, breast cancer mortality rate is decreased by 15–30%. Experts disagree on whether average-risk women aged 40–49 years should receive regular screening (Table 82–3). The U.K. Age Trial, the only randomized trial of breast cancer screening to specifically evaluate the impact of mammography in women aged 40–49 years, found no statistically significant difference in breast cancer mortality for screened women versus controls after about 11 years of follow-up (RR, 0.83; 95% CI, 0.66–1.04); however, less than 70% of women received screening in the intervention arm, potentially diluting the observed effect. A meta-analysis of eight large randomized trials showed a 15% relative reduction in mortality (RR, 0.85; 95% CI, 0.75–0.96) from mammography screening for women aged 39–49 years after 11–20 years of follow-up. This is equivalent to a number needed to invite to screening of 1904 over 10 years to prevent one breast cancer death. At the same time, nearly half of women aged 40–49 years screened annually will have false-positive mammograms necessitating further evaluation, often including biopsy. Estimates of overdiagnosis range from 10 to 40% of diagnosed invasive cancers.

No study of breast self-examination has shown it to decrease mortality. A randomized controlled trial of approximately 266,000 women in China demonstrated no difference in mortality between a group that received intensive breast self-exam instruction and reinforcement/reminders and controls at 10 years of follow-up. However, more benign breast lesions were discovered and more breast biopsies were performed in the self-examination arm.

Genetic screening for *BRCA1* and *BRCA2* mutations and other markers of breast cancer risk has identified a group of women at high risk for breast cancer. Unfortunately, when to begin and the optimal frequency of screening have not been defined. Mammography is less sensitive at detecting breast cancers in women carrying *BRCA1* and *-2* mutations, possibly because such cancers occur in younger women, in whom mammography is known to be less sensitive. MRI screening may be more sensitive than mammography in women at high risk due to genetic predisposition or in women with very dense breast tissue, but specificity may be lower. An increase in overdiagnosis may accompany the higher sensitivity. The impact of MRI on breast cancer mortality with or without concomitant use of mammography has not been evaluated in a randomized controlled trial.

Cervical Cancer

Screening with Papanicolaou smears decreases cervical cancer mortality. The cervical cancer mortality rate has fallen substantially since the widespread use of the Pap smear. Screening guidelines recommend regular Pap testing for all women who have reached the age of 21; some organizations advocate beginning earlier depending on sexual history. With the onset of sexual activity comes the risk of sexual transmission of HPV, the most common etiologic factor for cervical cancer. The recommended interval for Pap screening varies from 1 to 3 years. At age 30, women who have had three normal test results in a row may get screened every 2–3 years. An upper age limit at which screening ceases to be effective is not known, but women aged 65–70 years with no abnormal results in the previous 10 years may choose to stop screening. Screening should be discontinued in women who have undergone a hysterectomy for non-cancerous reasons.

Although the efficacy of the Papanicolaou smear in reducing cervical cancer mortality has never been directly confirmed in a randomized, controlled setting, a clustered randomized trial in India evaluated the impact of one-time cervical visual inspection and immediate colposcopy, biopsy, and/or cryotherapy (where indicated) versus counseling on cervical cancer deaths in women aged 30–59 years. After 7 years of follow-up, the age-standardized rate of death due to cervical cancer was 39.6 per 100,000 person-years in the intervention group versus 56.7 per 100,000 person-years in controls.

Colorectal Cancer

Fecal occult blood testing (FOBT), digital rectal examination (DRE), rigid and flexible sigmoidoscopy, colonoscopy, and CT colonography have been considered for colorectal cancer screening. Annual FOBT could reduce colorectal cancer mortality by a third. The sensitivity for fecal occult blood is increased if specimens are re-hydrated before testing, but at the cost of lower specificity. The false-positive rate for rehydrated FOBT is high; 1–5% of persons tested have a positive test. Only 2–10% of

those with occult blood in the stool have cancer and 20–30% have adenomas. The high false-positive rate of FOBT dramatically increases the number of colonoscopies performed.

Fecal immunochemical tests appear to have higher sensitivity for colorectal cancer than nonrehydrated FOBT tests. Fecal DNA testing is an emerging testing modality; it appears to have increased sensitivity and comparable specificity to FOBT and could potentially reduce harms associated with follow-up of false-positive tests. The body of evidence on the operating characteristics and effectiveness of fecal DNA tests in reducing colorectal cancer mortality is limited.

Two case-control studies suggest that regular screening of those older than age 50 years with sigmoidoscopy decreases mortality. This type of study is prone to selection biases. A quarter to a third of polyps can be discovered with the rigid sigmoidoscope; half are found with a 35-cm flexible scope and two-thirds to three-quarters are found with a 60-cm scope. Diagnosis of adenomatous polyps by sigmoidoscopy should lead to evaluation of the entire colon with colonoscopy. The most efficient interval for screening sigmoidoscopy is unknown, but 5 years is often recommended. Case-control studies suggest that intervals of up to 15 years may confer benefit.

One-time colonoscopy detects ~25% more advanced lesions (polyps >10 mm, villous adenomas, adenomatous polyps with high-grade dysplasia, invasive cancer) than one-time FOBT with sigmoidoscopy. Perforation rates are about 3/1000 for colonoscopy and 1/1000 for sigmoidoscopy. Debate continues on whether colonoscopy is too expensive and invasive for widespread use as a screening tool in standard-risk populations. Two observational studies suggest that efficacy of colonoscopy to decrease colorectal cancer mortality is restricted to the left side of the colon. CT colonography, if done at expert centers, appears to have a sensitivity for polyps ≥6 mm comparable to colonoscopy. However, the rate of extracolonic findings of abnormalities of uncertain significance that must nevertheless be worked up is high (~15–30%); the long-term cumulative radiation risk of repeated colonography screenings is also a concern.

Lung Cancer

Chest x-ray and sputum cytology have been evaluated in randomized lung cancer screening trials. No reduction in lung cancer mortality has been seen, although all controlled trials have had low statistical power. Preliminary (unpublished) findings from the National Lung Screening Trial, a randomized controlled trial of screening for lung cancer in approximately 53,000 persons aged 55-74 years with a 30+ pack-year smoking history, have shown a statistically significant 20% reduction in lung cancer mortality in the spiral CT arm (354 deaths) compared to the chest x-ray arm (442 deaths). However, the mortality benefits must be weighed against the disadvantages of spiral CT for a given population. These include the potential radiation risks associated with multiple scans, the discovery of incidental findings of unclear significance, and a high rate of false-positive test results. Both incidental findings and false-positive tests can lead to invasive diagnostic procedures associated with anxiety, expense, and complications (e.g., pneumo- or hemothorax after lung biopsy).

Ovarian Cancer

Adnexal palpation, transvaginal ultrasound, and serum CA-125 assay have been considered for ovarian cancer screening. These tests alone and in combination do not have sufficiently high sensitivity or specificity to be recommended for routine screening of ovarian cancer. The risks and costs associated with the high number of false-positive results is an impediment to routine use of these modalities for screening. A large randomized controlled trial has shown that of female participants receiving at least one false-positive serum CA-125 test, 14% underwent a major surgical procedure (e.g., laparotomy with oophorectomy) for benign disease. For transvaginal ultrasound, the rate was close to 40%.

Prostate Cancer

The most common prostate cancer screening modalities are DRE and serum prostate-specific antigen (PSA) assay. Newer serum tests, such as measurement of bound to free serum PSA, have yet to be fully evaluated. An emphasis on PSA screening has caused prostate cancer to become the most common non-skin cancer diagnosed in American males. This disease is prone to lead-time bias, length bias, and overdiagnosis, and substantial debate rages among experts as to whether it is effective. Prostate cancer screening clearly detects many asymptomatic cancers, but the ability to distinguish tumors that are lethal but still curable from those that pose little or no threat to health is limited. Men older than age 50 years have a high prevalence of indolent, clinically insignificant prostate cancers.

Two randomized controlled trials of the impact of PSA screening on prostate cancer mortality have been published. The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial was a multicenter U.S. trial that randomized almost 77,000 men ages 55–74 years to receive annual PSA testing for 6 years or usual care. At 7 years of follow-up, no statistically significant difference in the number of prostate cancer deaths were noted between the arms (rate ratio, 1.13; 95% CI, 0.75–1.90). The data at 10 years (67% complete) showed similar results. Approximately 44% of men in the control arm received at least one PSA test during the trial, which may have potentially diluted an observed effect.

The European Randomized Study of Screening for Prostate Cancer (ERSPC) was a multinational

study that randomized approximately 162,000 men between ages 50 and 74 years (with a predefined "core" screening group of men ages 55–69 years) to receive PSA testing every 4 years or no screening. Recruitment and randomizationprocedures and actual frequency of PSA testing varied by country. After a median follow-up of 9 years, a 20% relative reduction in the risk of prostate cancer death in the screened arm was noted in the "core" screening group (no difference in mortality was observed in the overall study population). The trial also found that 1140 men would need to be screened, and 48 additional cases treated to avert 1 death from prostate cancer.

The effectiveness of treatments for low-stage prostate cancer is under study. However, both surgery and radiation therapy may cause significant morbidity, such as impotence and urinary incontinence. Comparison of radical prostatectomy to "watchful waiting" in clinically diagnosed (not screen-detected) prostate cancers showed a small decrease in prostate cancer death rate in the surgery arm, but no statistically significant decrease in overall mortality was seen after 11 years of follow-up. Benefits were restricted to men younger than age 65 years. Urinary incontinence and sexual impotence were more common in the surgery arm. A man should have a life expectancy of at least 10 years to be eligible for screening. The USPSTF has found insufficient evidence to recommend prostate cancer screening for men younger than age 75 years; it recommends against screening for prostate cancer in men age 75 years or older ("D" recommendation) (Table 82–3).

Endometrial Cancer

Transvaginal ultrasound and endometrial sampling have been advocated as screening tests for endometrial cancer. Benefit from routine screening has not been shown. Transvaginal ultrasound and endometrial sampling are indicated for workup of vaginal bleeding in postmenopausal women but are not considered as screening tests in symptomatic women.

Skin Cancer

Visual examination of all skin surfaces by the patient or by a health care provider is used in screening for basal and squamous cell cancers and melanoma. No prospective randomized study has been performed to look for a mortality decrease. Unfortunately, screening is associated with a substantial rate of overdiagnosis.

FURTHER READINGS

Andriole GL et al: Mortality results from a randomized prostate-cancer screening trial. N Engl J Med 360:1310, 2009[PMID: 19297565]

Atkin WS et al: Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: A multicentre randomised controlled trial. Lancet 375:1624, 2010[PMID: 20430429]

Bach PB et al: Computed tomography screening and lung cancer outcomes. JAMA 297:953, 2007[PMID: 17341709]

Barrett-Connor E et al: Effects of raloxifene on cardiovascular events and breast cancer in postmenopausal women. N Engl J Med 355:125, 2006[PMID: 16837676]

Kramer BS et al: Cancer screening: The clash of science and intuition. Annu Rev Med 60:125, 2009.[PMID: 18803476]

Nelson HD et al: Screening for breast cancer: An update for the U.S. Preventive Services Task Force. Ann Intern Med 151:727, 2009[PMID: 19920273]

Prentice RL et al: Low-fat dietary pattern and risk of invasive breast cancer. The Women's Health Initiative randomized controlled dietary modification trial. JAMA 295:629, 2006[PMID: 16467232]

Schroeder FH et al: Screening and prostate-cancer mortality in a randomized European study. N Engl J Med 360: 1320, 2009

Web Sites

The U.S. Preventive Services Task Force: *http://www.ahrq.gov/clinic/uspstfix.htm*

The National Cancer Institute Cancernet: *http://www.cancer.gov/cancertopics*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CANCER IS A GENETIC DISEASE

Cancer arises through a series of somatic alterations in DNA that result in unrestrained cellular proliferation. Most of these alterations involve actual sequence changes in DNA (i.e., mutations). They may originate as a consequence of random replication errors, exposure to carcinogens (e.g., radiation), or faulty DNA repair processes. While most cancers arise sporadically, familial clustering of cancers occurs in certain families that carry a germline mutation in a cancer gene.

HISTORICAL PERSPECTIVE

The idea that cancer progression is driven by sequential somatic mutations in specific genes has only gained general acceptance in the past 25 years. Before the advent of the microscope, cancer was believed to be composed of aggregates of mucus or other noncellular matter. By the middle of the nineteenth century, it became clear that tumors were masses of cells and that these cells arose from the normal cells of the tissue from which the cancer originated. However, the molecular basis for the uncontrolled proliferation of cancer cells was to remain a mystery for another century. During that time, a number of theories for the origin of cancer were postulated. The great biochemist Otto Warburg proposed the combustion theory of cancer, which stipulated that cancer was due to abnormal oxygen metabolism. In addition, some believed that all cancers were caused by viruses, and that cancer was in fact a contagious disease.

In the end, observations of cancer occurring in chimney sweeps, studies of x-rays, and the overwhelming data demonstrating cigarette smoke as a causative agent in lung cancer, together with Ames's work on chemical mutagenesis, provided convincing evidence that cancer originated through changes in DNA. Although the viral theory of cancer did not prove to be generally accurate (with the exception of human papillomaviruses, which can cause cervical cancer in human), the study of retroviruses led to the discovery of the first human *oncogenes* in the late 1970s. Soon after, the study of families with genetic predisposition to cancer was instrumental in the discovery of *tumor-suppressor genes*. The field that studies the type of mutations, as well as the consequence of these mutations in tumor cells, is now known as *cancer genetics*.

THE CLONAL ORIGIN AND MULTISTEP NATURE OF CANCER

Nearly all cancers originate from a single cell; this clonal origin is a critical discriminating feature between neoplasia and hyperplasia. Multiple cumulative mutational events are invariably required for the progression of a tumor from normal to fully malignant phenotype. The process can be seen as Darwinian microevolution in which, at each successive step, the mutated cells gain a growth advantage resulting in an increased representation relative to their neighbors (Fig. 83-1). Based on observations of cancer frequency increases during aging, as well as recent molecular genetics work, it is believed that 5 to 10 accumulated mutations are necessary for a cell to progress from the normal to the fully malignant phenotype.

Figure 83-1

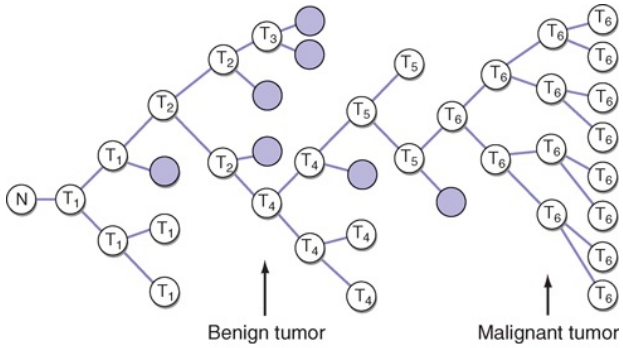

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multistep clonal development of malignancy. In this diagram a series of five cumulative mutations (T1, T2, T4, T5, T6), each with a modest growth advantage acting alone, eventually results in a malignant tumor. Note that not all such alterations result in progression; for example, the T3 clone is a dead end. The actual number of cumulative mutations necessary to transform from the normal to the malignant state is unknown in most tumors. *(After P Nowell: Science 194:23, 1976, with permission.)*

We are beginning to understand the precise nature of the genetic alterations responsible for some malignancies and to get a sense of the order in which they occur. The best studied example is colon cancer, in which analyses of DNA from tissues extending from normal colon epithelium through adenoma to carcinoma have identified some of the genes mutated in the process (Fig. 83-2). Other malignancies are believed to progress in a similar step-wise fashion, although the order and identity of genes affected may be different.

Figure 83-2

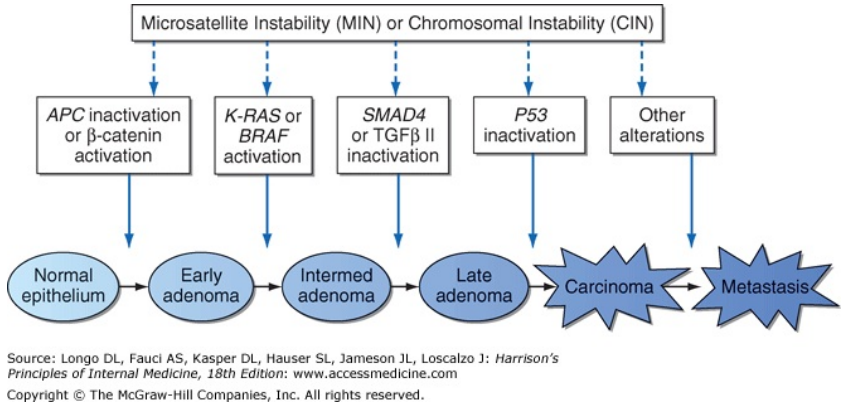

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Progressive somatic mutational steps in the development of colon carcinoma. The accumulation of alterations in a number of different genes results in the progression from normal epithelium through adenoma to full-blown carcinoma. Genetic instability (microsatellite or chromosomal) accelerates the progression by increasing the likelihood of mutation at each step. Patients with familial polyposis are already one step into this pathway, since they inherit a germline alteration of the *APC* gene. TGF, transforming growth factor.

Figure 83-3

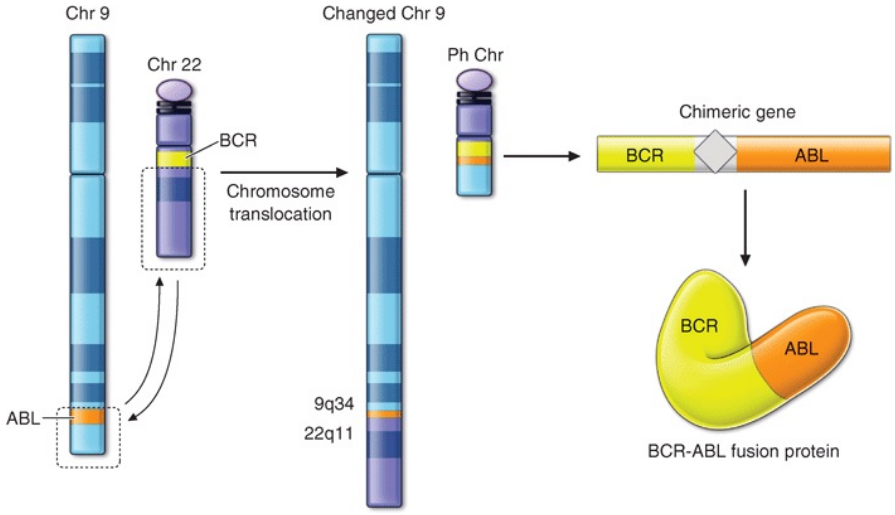

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Specific translocation seen in chronic myelogenous leukemia (CML). The Philadelphia chromosome (Ph) is derived from a reciprocal translocation between chromosomes 9 and 22 with the breakpoint joining the sequences of the *ABL* oncogene with the *BCR* gene. The fusion of these DNA sequences allows the generation of an entirely novel fusion protein with modified function.

TWO TYPES OF CANCER GENES: ONCOGENES AND TUMOR-SUPPRESSOR GENES

There are two major types of cancer genes. The first type comprises genes that positively influence tumor formation and are known as *oncogenes*. The second type of cancer genes negatively impact tumor growth and have been named *tumor-suppressor genes*. Both oncogenes and tumor-suppressor genes exert their effects on tumor growth through their ability to control cell division (cell birth) or cell death (apoptosis), although the mechanisms can be extremely complex. While tightly regulated in normal cells, oncogenes acquire mutations in cancer cells, and the mutations typically relieve this control and lead to increased activity of the gene products. This mutational event typically occurs in a single allele of the oncogene and acts in a dominant fashion. In contrast, the normal function of tumor-suppressor genes is usually to restrain cell growth, and this function is lost in cancer. Because of the diploid nature of mammalian cells, both alleles must be inactivated for a cell to completely lose the function of a tumor-suppressor gene, leading to a recessive mechanism at the cellular level. From these ideas and studies on the inherited form of retinoblastoma, Knudson and others formulated the *two-hit hypothesis*, which in its modern version states that both copies of a tumor-suppressor gene must be inactivated in cancer.

There is a subset of tumor-suppressor genes, the *caretaker genes*, which do not affect cell growth directly, but rather control the ability of the cell to maintain the integrity of its genome. Cells with a deficiency in these genes have an increased rate of mutations throughout their genomes, including in oncogenes and tumor-suppressor genes. This "mutator" phenotype was first hypothesized by Loeb to explain how the multiple mutational events required for tumorigenesis can occur in the lifetime of an individual. A mutator phenotype has now been observed in some forms of cancer, such as those associated with deficiencies in DNA mismatch repair. The great majority of cancers, however, do not harbor repair deficiencies, and their rate of mutation is similar to that observed in normal cells. Many of these cancers, however, appear to harbor a different kind of genetic instability, affecting the loss or gains of whole chromosomes or large parts thereof (as explained in more detail below).

ONCOGENES IN HUMAN CANCER

Work by Peyton Rous in the early 1900s revealed that a chicken sarcoma could be transmitted from animal to animal in cell-free extracts, suggesting that cancer could be induced by an agent acting positively to promote tumor formation. The agent responsible for the transmission of the cancer was a retrovirus (Rous sarcoma virus, RSV) and the oncogene responsible was

identified 75 years later as *v-src*. Other oncogenes were also discovered through their presence in the genomes of retroviruses that are capable of causing cancers in chickens, mice, and rats. The cellular homologues of these viral genes are called protooncogenes and are often targets of mutation or aberrant regulation in human cancer. Whereas many oncogenes were discovered because of their presence in retroviruses, other oncogenes, particularly those involved in translocations characteristic of particular leukemias and lymphomas, were isolated through genomic approaches. Investigators cloned the sequences surrounding the chromosomal translocations observed cytogenetically and then deduced the nature of the genes that were the targets of these translocations (see below). Some of these were oncogenes known from retroviruses (like *ABL*, involved in chronic myeloid leukemia [CML]), while others were new (like *BCL2*, involved in B cell lymphoma). In the normal cellular environment, protooncogenes have crucial roles in cell proliferation and differentiation. Table 83–1 is a partial list of oncogenes known to be involved in human cancer.

Table 83–1 Common Oncogenes Altered in Human Cancers

Oncogene	Function	Alteration in Cancer	Neoplasm
AKT1	Serine/threonine kinase	Amplification	Stomach
AKT2	Serine/threonine kinase	Amplification	Ovarian, breast, pancreatic
BRAF	Serine/threonine kinase	Point mutation	Melanoma, lung, colorectal
CTNNB1	Signal transduction	Point mutation	Colon, prostate, melanoma, skin, others
FOS	Transcription factor	Overexpression	Osteosarcomas
ERBB2	Receptor tyrosine kinase	Point mutation, amplification	Breast, ovary, stomach, neuroblastoma
JUN	Transcription factor	Overexpression	Lung
MET	Receptor tyrosine kinase	Point mutation, rearrangement	Osteocarcinoma, kidney, glioma
MYB	Transcription factor	Amplification	AML, CML, colorectal, melanoma
C-MYC	Transcription factor	Amplification	Breast, colon, gastric, lung
L-MYC	Transcription factor	Amplification	Lung, bladder
N-MYC	Transcription factor	Amplification	Neuroblastoma, lung
HRAS	GTPase	Point mutation	Colon, lung, pancreas
KRAS	GTPase	Point mutation	Melanoma, colorectal, AML
NRAS	GTPase	Point mutation	Various carcinomas, melanoma
REL	Transcription factor	Rearrangement, amplification	Lymphomas
WNT1	Growth factor	Amplification	Retinoblastoma

Abbreviations: AML, acute myeloid leukemia; CML, chronic myeloid leukemia.

The normal growth and differentiation of cells is controlled by growth factors that bind to receptors on the surface of the cell. The signals generated by the membrane receptors are transmitted inside the cells through signaling cascades involving kinases, G proteins, and other regulatory proteins. Ultimately, these signals affect the activity of transcription factors in the nucleus, which regulate the expression of genes crucial in cell proliferation, cell differentiation, and cell death. Oncogene products have been found to function at critical steps in these pathways (Chap. 84), and inappropriate activation of these pathways can lead to tumorigenesis.

MECHANISMS OF ONCOGENE ACTIVATION

Point Mutation

Point mutation is a common mechanism of oncogene activation. For example, mutations in one of the *RAS* genes (*HRAS*, *KRAS*, or *NRAS*) are present in up to 85% of pancreatic cancers and 45% of colon cancers but are less common in other cancer types, although they can occur at significant frequencies in leukemia, lung, and thyroid cancers. Remarkably—and in contrast to the diversity of mutations found in tumor-suppressor genes (see below)—most of the activated *RAS* genes contain point mutations in codons 12, 13, or 61 (these mutations reduce RAS GTPase activity, leading to constitutive activation of the mutant RAS protein). The restricted pattern of mutations observed in oncogenes compared to that of tumor-suppressor genes reflects the fact that gain-of-function mutations are less likely to occur than mutations that simply lead to loss of activity. Indeed, inactivation of a gene can in theory be accomplished through the introduction of a stop codon anywhere in the coding sequence, whereas activations require precise substitutions at residues that can somehow lead to an increase in the activity of the encoded protein. Importantly, the specificity of oncogene mutations provides diagnostic opportunities, as tests that identify mutations at defined positions are easier to design than tests aimed at detecting random changes in a gene.

DNA Amplification

The second mechanism for activation of oncogenes is DNA sequence amplification, leading to overexpression of the gene product. This increase in DNA copy number may cause cytologically recognizable chromosome alterations referred to as *homogeneous staining regions* (HSRs) if integrated within chromosomes, or *double minutes* (dmins) if extrachromosomal. The recognition of DNA amplification is accomplished through various cytogenetic techniques such as comparative genomic hybridization (CGH) or fluorescence in situ hybridization (FISH), which allow the visualization of chromosomal aberrations using fluorescent dyes. In addition, noncytogenetic, microarray-based approaches are now available for identifying changes in copy number at high resolution. Newer short-tag–based sequencing approaches have been used to evaluate amplifications. When paired with next-generation sequencing instruments, this approach offers the highest degree of resolution and quantification available. With both microarray and sequencing technologies, the entire genome can be surveyed for gains and losses of DNA sequences, thus pinpointing chromosomal regions likely to contain genes important in the development or progression of cancer.

Numerous genes have been reported to be amplified in cancer. Several of these genes, including *NMYC* and *LMYC*, were identified through their presence within the amplified DNA sequences of a tumor and had homology to known oncogenes. Because the region amplified often includes hundreds of thousands of base pairs, multiple oncogenes may be amplified in a single amplicon in some cancers (particularly in sarcomas). Indeed, *MDM2*, *GLI*, *CDK4*, and *SAS* at chromosomal location 12q13-15 have been shown to be simultaneously amplified in several types of sarcomas and other tumors. Amplification of a cellular gene is often a predictor of poor prognosis; for example, *ERBB2*/*HER2* and *NMYC* are often amplified in aggressive breast cancers and neuroblastoma, respectively.

Chromosomal Rearrangement

Chromosomal alterations provide important clues to the genetic changes in cancer. The chromosomal alterations in human solid tumors such as carcinomas are heterogeneous and complex and occur as a result of the frequent chromosomal instability (CIN) observed in these tumors (see below). In contrast, the chromosome alterations in myeloid and lymphoid tumors are often simple translocations, i.e., reciprocal transfers of chromosome arms from one chromosome to another. Consequently, many detailed and informative chromosome analyses have been performed on hematopoietic cancers. The breakpoints of recurring chromosome abnormalities usually occur at the site of cellular oncogenes. Table 83–2 lists representative examples of recurring chromosome alterations in malignancy and the associated gene(s) rearranged or deregulated by the chromosomal rearrangement. Translocations are particularly common in lymphoid tumors, probably because these cell types have the capability to rearrange their DNA to generate antigen receptors. Indeed, antigen receptor genes are commonly involved in the translocations, implying that an imperfect regulation of receptor gene rearrangement may be involved in the pathogenesis. An interesting example is Burkitt's lymphoma, a B cell tumor characterized by a reciprocal translocation between chromosomes 8 and 14. Molecular analysis of Burkitt's lymphomas demonstrated that the breakpoints occurred within or near the *MYC* locus on chromosome 8 and within the immunoglobulin heavy chain locus on chromosome 14, resulting in the transcriptional activation of *MYC*. Enhancer activation by translocation, although not universal, appears to play an important role in malignant progression. In addition to transcription factors and signal transduction molecules, translocation may result in the overexpression of cell cycle regulatory proteins or proteins such as cyclins and of proteins that regulate cell death.

Table 83–2 Representative Oncogenes at Chromosomal Translocations

Gene (Chromosome)	Translocation	Malignancy
ABL (9q34.1)–*BCR* (22q11)	(9;22)(q34;q11)	Chronic myeloid leukemia
ATF1 (12q13)–*EWS* (22q12)	(12;22)(q13;q12)	Malignant melanoma of soft parts
BCL1 (11q13.3)–*IgH* (14q32)	(11;14)(q13;q32)	Mantle cell lymphoma
BCL2 (18q21.3)–*IgH* (14q32)	(14;18)(q32;q21)	Follicular lymphoma
FLI1 (11q24)–*EWS* (22q12)	(11;22)(q24;q12)	Ewing's sarcoma
LCK (1p34)–*TCRB* (7q35)	(1;7)(p34;q35)	T cell acute lymphocytic leukemia
MYC (8q24)–*IgH* (14q32)	(8;14)(q24;q32)	Burkitt's lymphoma, B cell acute lymphocytic leukemia
PAX3 (2q35)–*FKHR/ALV* (13q14)	(2;13)(q35;q14)	Alveolar rhabdomyosarcoma
PAX7 (1p36)–*KHR/*ALV(13q14)	(1;13)(p36;q14)	Alveolar rhabdomyosarcoma
REL (2p13)–*NRG* (2p11.2-14)	Inv(2(p13;p11.2-14)	Non-Hodgkin's lymphoma
RET (10q11.2)–*PKAR1A* (17q23)	(10;17)(q11.2;q23)	Thyroid carcinoma
TAL1(1p32)–*TCTA* (3p21)	(1;3)(p34;p21)	Acute T cell leukemia
TRK (1q23-1q24)–*TPM3* (1q31)	Inv1(q23;q31)	Colon carcinoma
WT1 (11p13)–*EWS* (22q12)	(11;22)(p13;q12)	Desmoplastic small round cell tumor

Source: From Hesketh R: *The Oncogene and Tumour Suppressor Gene Facts Book* , 2nd ed. San Diego, Academic Press, 1997; with permission.

The first reproducible chromosome abnormality detected in human malignancy was the Philadelphia chromosome detected in CML. This cytogenetic abnormality is generated by reciprocal translocation involving the *ABL* oncogene on chromosome 9, encoding a tyrosine kinase, being placed in proximity to the *BCR* (breakpoint cluster region) gene on chromosome 22. Figure 83-3 illustrates the generation of the translocation and its protein product. The consequence of expression of the *BCR-ABL* gene product is the activation of signal transduction pathways leading to cell growth independent of normal external signals. Imatinib (marketed as Gleevec), a drug that specifically blocks the activity of *BCR-ABL*, has shown remarkable efficacy with little toxicity in patients with CML. It is hoped that knowledge of genetic alterations in other cancers will likewise lead to mechanism-based design and development of a new generation of chemotherapeutic agents.

CHROMOSOMAL INSTABILITY IN SOLID TUMORS

Solid tumors are generally highly aneuploid, containing an abnormal number of chromosomes; these chromosomes also exhibit structural alterations such as translocations, deletions, and amplifications. These abnormalities are collectively referred to as chromosomal instability (CIN). Normal cells possess several cell cycle checkpoints, essentially quality-control requirements that have to be met before subsequent events are allowed to take place. The mitotic checkpoint, which ensures proper chromosome attachment to the mitotic spindle before allowing the sister chromatids to separate, is altered in certain cancers. The molecular basis of CIN remains unclear, although a number of mitotic checkpoint genes are found mutated or abnormally expressed in various tumors. The exact effects of these changes on the mitotic checkpoint are unknown, and both weakening and overactivation of the checkpoint have been proposed. The identification of the cause of CIN in tumors will likely be a formidable task, considering that several hundred genes are thought to control the mitotic checkpoint and other cellular processes ensuring proper chromosome segregation. Regardless of the mechanisms underlying CIN, the measurement of the number of chromosomal alterations present in tumors is now possible with both cytogenetic and molecular techniques, and several studies have shown that this information can be useful for prognostic purposes. In addition, since the mitotic checkpoint is essential for cellular viability, it may become a target for novel therapeutic approaches.

TUMOR-SUPPRESSOR GENE INACTIVATION IN CANCER

The first indication of the existence of tumor-suppressor genes came from experiments showing that fusion of mouse cancer cells with normal mouse fibroblasts led to a nonmalignant phenotype in the fused cells. The normal role of tumor-suppressor genes is to restrain cell growth, and the function of these genes is inactivated in cancer. The two major types of somatic lesions observed in tumor-suppressor genes during tumor development are *point mutations* and *large deletions*. Point mutations in the coding region of tumor-suppressor genes will frequently lead to truncated protein products or otherwise nonfunctional proteins. Similarly, deletions lead to the loss of a functional product and sometimes encompass the entire gene or even the entire chromosome arm, leading to loss of heterozygosity (LOH) in the tumor DNA compared to the corresponding normal tissue DNA (Fig. 83-4). LOH in tumor DNA is considered a hallmark for the presence of a tumor-suppressor gene at a particular chromosomal location, and LOH studies have been useful in the positional cloning of many tumor-suppressor genes.

Figure 83-4

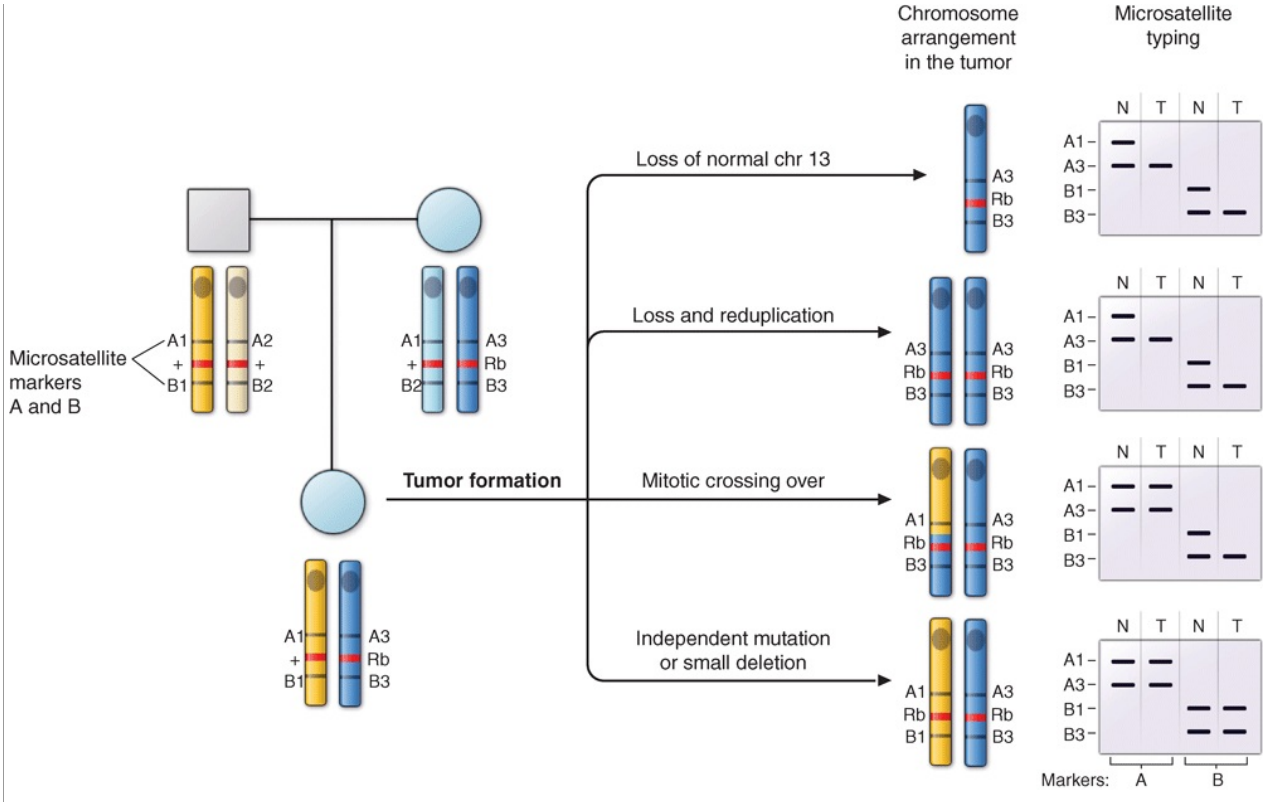

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of possible mechanisms for tumor formation in an individual with hereditary (familial) retinoblastoma. On the left is shown the pedigree of an affected individual who has inherited the abnormal (Rb) allele from her affected mother. The normal allele is shown as a (+). The four chromosomes of her two parents are drawn to indicate their origin. Flanking the retinoblastoma locus are microsatellite markers (A and B) also analyzed in this family. Markers A3 and B3 are on the chromosome carrying the -retinoblastoma disease gene. Tumor formation results when the normal allele, which this patient inherited from her father, is inactivated. On the right are shown four possible ways in which this could occur. In each case, the resulting chromosome 13 arrangement is shown, as well as the results of PCR typing using the microsatellite markers comparing normal tissue (N) with tumor tissue (T). Note that in the first three situations the normal allele (B1) has been lost in the tumor tissue, which is referred to as loss of heterozygosity (LOH) at this locus.

Gene silencing, an epigenetic change that leads to the loss of gene expression and occurs in conjunction with hypermethylation of the promoter and histone deacetylation, is another mechanism of tumor-suppressor gene inactivation. (An *epigenetic modification* refers to a change in the genome, heritable by cell progeny, that does not involve a change in the DNA sequence. The inactivation of the second X chromosome in female cells is an example of an epigenetic silencing that prevents gene expression from the inactivated chromosome). During embryologic development, regions of chromosomes from one parent are silenced and gene expression proceeds from the chromosome of the other parent. For most genes, expression occurs from both alleles or randomly from one allele or the other. The preferential expression of a particular gene exclusively from the allele contributed by one parent is called *parental imprinting* and is thought to be regulated by covalent modifications of chromatin protein and DNA (often methylation) of the silenced allele.

The role of epigenetic control mechanisms in the development of human cancer is unclear. However, a general decrease in the level of DNA methylation has been noted as a common change in cancer. In addition, numerous genes, including some tumor-suppressor genes, appear to become hypermethylated and silenced during tumorigenesis. *VHL* and *p16INK4* are well-studied examples of such tumor-suppressor genes. Overall, epigenetic mechanisms may be responsible for reprogramming the expression of a large number of genes in cancer and, together with the mutation of specific genes, are likely to be crucial in the development of human malignancies.

FAMILIAL CANCER SYNDROMES

A small fraction of cancers occur in patients with a genetic predisposition. In these families, the affected individuals have a predisposing loss-of-function mutation in one allele of a tumor-suppressor gene. The tumors in these patients show a loss of the remaining normal allele as a result of somatic events (point mutations or deletions), in agreement with the two-hit hypothesis (Fig. 83-4). Thus, most cells of an individual with an inherited loss-of-function mutation in a tumor-suppressor gene are functionally normal, and only the rare cells that develop a mutation in the remaining normal allele will exhibit uncontrolled regulation.

Roughly 100 syndromes of familial cancer have been reported, although many are rare. The majority are inherited as autosomal dominant traits, although some of those associated with DNA repair abnormalities (xeroderma pigmentosum, Fanconi's anemia, ataxia telangiectasia) are autosomal recessive. Table 83–3 shows a number of cancer predisposition syndromes and the responsible genes. The current paradigm states that the genes mutated in familial syndromes can also be targets for somatic mutations in sporadic (noninherited) tumors. The study of cancer syndromes has thus provided invaluable insights into the mechanisms of progression for many tumor types. This section examines the case of inherited colon cancer in detail, but similar general lessons can be applied to many of the cancer syndromes listed in Table 83–3. In particular, the study of inherited colon cancer will clearly illustrate the difference between two types of tumor-suppressor genes: the *gatekeepers*, which directly regulate the growth of tumors, and the *caretakers*, which, when mutated, lead to genetic instability and therefore act indirectly on tumor growth.

Table 83–3 Cancer Predisposition Syndromes and Associated Genes

Syndrome	Gene	Chromosome	Inheritance	Tumors
Ataxia telangiectasia	*ATM*	11q22-q23	AR	Breast
Autoimmune lymphoproliferative syndrome	*FAS FASL*	10q24 1q23	AD	Lymphomas
Bloom syndrome	*BLM*	15q26.1	AR	Several types
Cowden syndrome	*PTEN*	10q23	AD	Breast, thyroid
Familial adenomatous polyposis	*APC*	5q21	AD	Intestinal adenoma, colorectal
Familial melanoma	*p16INK4*	9p21	AD	Melanoma, pancreatic
Familial Wilms' tumor	*WT1*	11p13	AD	Kidney (pediatric)
Hereditary breast/ovarian cancer	*BRCA1* *BRCA2*	17q21 13q12.3	AD	Breast, ovarian, colon, prostate
Hereditary diffuse gastric cancer	*CDH1*	16q22	AD	Stomach
Hereditary multiple exostoses	*EXT1* *EXT2*	8q24 11p11-12	AD	Exostoses, chondrosarcoma
Hereditary prostate cancer	*HPC1*	1q24-25	AD	Prostate
Hereditary retinoblastoma	*RB1*	13q14.2	AD	Retinoblastoma, osteosarcoma
Hereditary nonpolyposis colon cancer (HNPCC)	*MSH2* *MLH1* *MSH6* *PMS2*	2p16 3p21.3 2p16 7p22	AD	Colon, endometrial, ovarian, stomach, small bowel, ureter carcinoma
Hereditary papillary renal carcinoma	*MET*	7q31	AD	Papillary kidney
Juvenile polyposis	*SMAD4*	18q21	AD	Gastrointestinal, pancreatic
Li-Fraumeni	*TP53*	17p13.1	AD	Sarcoma, breast
Multiple endocrine neoplasia type 1	*MEN1*	11q13	AD	Parathyroid, endocrine, pancreas, and pituitary
Multiple endocrine neoplasia type 2a	*RET*	10q11.2	AD	Medullary thyroid carcinoma, pheochromocytoma
Neurofibromatosis type 1	*NF1*	17q11.2	AD	Neurofibroma, neurofibrosarcoma, brain
Neurofibromatosis type 2	*NF2*	22q12.2	AD	Vestibular schwannoma, meningioma, spine
Nevoid basal cell carcinoma syndrome (Gorlin's syndrome)	*PTCH*	9q22.3	AD	Basal cell carcinoma, medulloblastoma, jaw cysts
Tuberous sclerosis	*TSC1* *TSC2*	9q34 16p13.3	AD	Angiofibroma, renal angiomyolipoma
von Hippel–Lindau	*VHL*	3p25-26	AD	Kidney, cerebellum, pheochromocytoma

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

Familial adenomatous polyposis (FAP) is a dominantly inherited colon cancer syndrome due to germline mutations in the adenomatous polyposis coli (*APC*) tumor-suppressor gene on chromosome 5. Patients with this syndrome develop hundreds to thousands of adenomas in the colon. Each of these adenomas has lost the normal remaining allele of *APC* but has not yet accumulated the required additional mutations to generate fully malignant cells (Fig. 83-2). The loss of the second functional *APC* allele in tumors from FAP families often occurs through loss of heterozygosity. However, out of these thousands of benign adenomas, several will invariably acquire further abnormalities and a subset will even develop into fully malignant cancers. *APC* is thus considered to be a gatekeeper for colon tumorigenesis: in the absence of mutation of this gatekeeper (or a gene acting within the same pathway), a colorectal tumor simply cannot form. Figure 83-5 shows germline and somatic mutations found in the *APC* gene. The function of the APC protein is still not completely understood, but it likely provides differentiation and apoptotic cues to colonic cells as they migrate up the crypts. Defects in this process may lead to abnormal accumulation of cells that should normally undergo apoptosis.

Figure 83-5

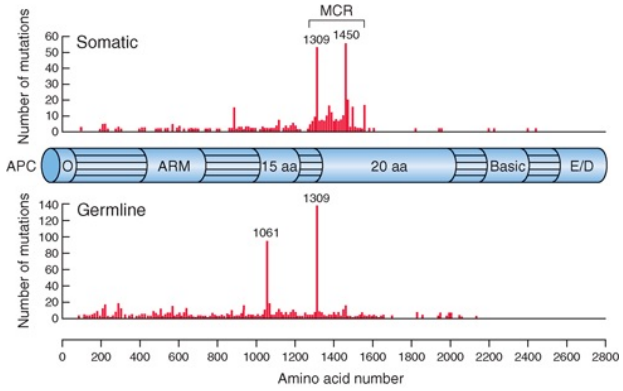

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Germline and somatic mutations in the tumor-suppressor gene *APC*. *APC* encodes a 2843-amino-acid protein with 6 major domains: an oligomerization region (O), armadillo repeats (ARM), 15-amino-acid repeats (15 aa), 20-amino-acid

major domains: an oligomerization region (O), armadillo repeats (~10 aa), 15-amino-acid repeats (20 aa), a basic region, and a domain involved in binding EB1 and the *Drosophila* discs large homologue (E/D). Shown are the positions within the *APC* gene of a total of 650 somatic and 826 germline mutations (from the *APC* database at *http://www.umd.be/APC*). The vast majority of these mutations result in the truncation of the APC protein. Germline mutations are found to be relatively evenly distributed up to codon 1600 except for two mutation hotspots at amino acids 1061 and 1309, which together account for one-third of the mutations found in familial adenomatous polyposis (FAP) families. Somatic *APC* mutations in colon tumors cluster in an area of the gene known as the *mutation cluster region* (MCR). The location of the MCR suggests that the 20-amino-acid domain plays a crucial role in tumor suppression.

In contrast to patients with FAP, patients with hereditary nonpolyposis colon cancer (HNPCC, or Lynch syndrome) do not develop multiple polyposis but instead develop only one or a small number of adenomas that rapidly progress to cancer. Most HNPCC cases are due to mutations in one of four DNA mismatch repair genes (Table 83–3), which are components of a repair system that is normally responsible for correcting errors in freshly replicated DNA. Germline mutations in *MSH2* and *MLH1* account for more than 90% of HNPCC cases, while mutations in *MSH6* and *PMS2* are much less frequent. When a somatic mutation inactivates the remaining wild-type allele of a mismatch repair gene, the cell develops a hypermutable phenotype characterized by profound genomic instability, especially for the short repeated sequences called *microsatellites*. This microsatellite instability (MSI) favors the development of cancer by increasing the rate of mutations in many genes, including oncogenes and tumor-suppressor genes (Fig. 83-2). These genes can thus be considered caretakers. Interestingly, chromosomal instability (CIN) can also be found in colon cancer, but MSI and CIN appear to be mutually exclusive, suggesting that they represent alternative mechanisms for the generation of a mutator phenotype in this cancer (Fig. 83-2). Other cancer types rarely exhibit MSI but most exhibit CIN.

While most autosomal dominant inherited cancer syndromes are due to mutations in tumor-suppressor genes (Table 83–3), there are a few interesting exceptions. Multiple endocrine neoplasia type II, a dominant disorder characterized by pituitary adenomas, medullary carcinoma of the thyroid, and (in some pedigrees) pheochromocytoma, is due to gain-of-function mutations in the protooncogene *RET* on chromosome 10. Similarly, gain-of-function mutations in the tyrosine kinase domain of the *MET* oncogene lead to hereditary papillary renal carcinoma. Interestingly, loss-of-function mutations in the *RET* gene cause a completely different disease, Hirschsprung's disease (aganglionic megacolon [Chaps. 297 and 351]).

Although the Mendelian forms of cancer have taught us much about the mechanisms of growth control, most forms of cancer do not follow simple patterns of inheritance. In many instances (e.g., lung cancer), a strong environmental contribution is at work. Even in such circumstances, however, some individuals may be more genetically susceptible to developing cancer, given the appropriate exposure, due to the presence of modifier alleles.

GENETIC TESTING FOR FAMILIAL CANCER

The discovery of cancer susceptibility genes raises the possibility of DNA testing to predict the risk of cancer in individuals of affected families. An algorithm for cancer risk assessment and decision making in high-risk families using genetic testing is shown in Fig. 83-6. Once a mutation is discovered in a family, subsequent testing of asymptomatic family members can be crucial in patient management. A negative gene test in these individuals can prevent years of anxiety in the knowledge that their cancer risk is no higher than that of the general population. On the other hand, a positive test may lead to alteration of clinical management, such as increased frequency of cancer screening and, when feasible and appropriate, prophylactic surgery. Potential negative consequences of a positive test result include psychological distress (anxiety, depression) and discrimination, although the Genetic Information Nondiscrimination Act (GINA) makes it illegal for predictive genetic information to be used to discriminate in health insurance or employment. Testing should therefore not be conducted without counseling before and after disclosure of the test result. In addition, the decision to test should depend on whether effective interventions exist for the particular type of cancer to be tested. Despite these caveats, genetic cancer testing for some cancer syndromes already appears to have greater benefits than risks, and many companies now offer testing for various genes associated with the predisposition to breast cancer (*BRCA1* and *BRCA2*), melanoma (*p16INK4*), and colon cancer (*APC* and the HNPCC genes).

Figure 83-6

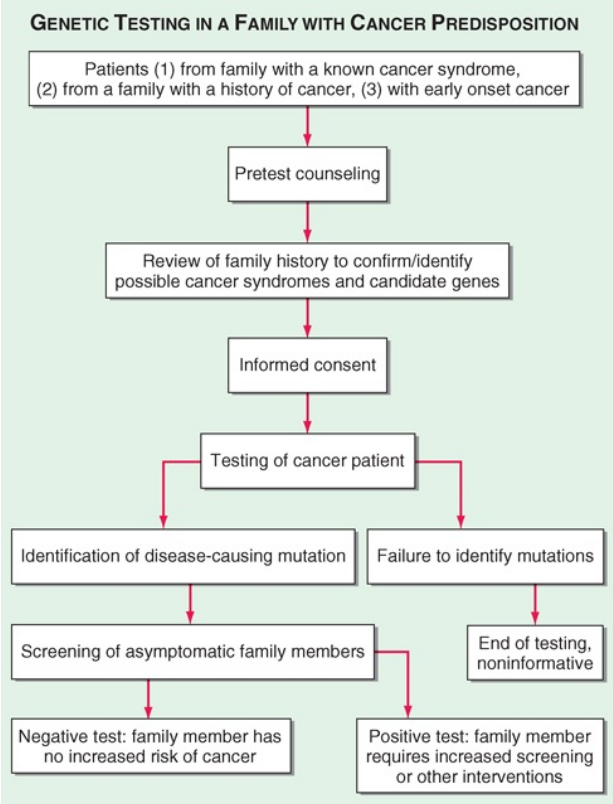

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for genetic testing in a family with cancer predisposition. The key step is the identification of a mutation in a cancer patient, which allows testing of asymptomatic family members. Asymptomatic family members who test positive may require increased screening or surgery, whereas others are at no greater risk for cancer than the general population.

Because of the inherent problems of genetic testing such as cost, specificity, and sensitivity, it is not yet appropriate to offer these tests to the general population. However, testing may be appropriate in some subpopulations with a known increased risk, even without a defined family history. For example, two mutations in the breast cancer susceptibility gene *BRCA1*, 185delAG and 5382insC, exhibit a sufficiently high frequency in the Ashkenazi Jewish population that genetic testing of an individual of this ethnic group may be warranted.

As noted above, it is important that genetic test results be communicated to families by trained genetic counselors, especially for high-risk high-penetrance conditions such as the hereditary breast/ovarian cancer syndrome (*BRCA1/BRCA2*). To ensure that the families clearly understand its advantages and disadvantages and the impact it may have on disease management and psyche, genetic testing should never be done before counseling. Significant expertise is needed to communicate the results of genetic testing to families. For example, one common mistake is to misinterpret the result of negative genetic tests. For many cancer predisposition genes, the sensitivity of genetic testing is less than 70% (i.e., of 100 kindreds tested, disease-causing mutations can be identified in 70 at most). Therefore, such testing should in general begin with an affected member of the kindred (the youngest family member still alive who has had the cancer of interest). If a mutation is not identified in this individual, then the test should be reported as noninformative (Fig. 83-6) rather than negative (because it is possible that, for technical reasons, the mutation in this individual is not detectable by standard genetic assays). On the other hand, if a mutation can be identified in this individual, then testing of other family members can be performed, and the sensitivity of such subsequent tests will be 100% (because the mutation in the family is in this case known to be detectable by the method used).

MICRORNAS AND CANCER

MicroRNAs (miRNAs) are small noncoding RNAs 20–22 nucleotides in length that are involved in posttranscriptional gene regulation. Studies in chronic lymphocytic leukemia first suggested a link between miRNAs and cancer when *miR-15* and *miR-16* were found to be deleted or downregulated in the vast majority of tumors. Various miRNAs have since been found abnormally expressed in several human malignancies. Aberrant expression of miRNAs in cancer has been attributed to several mechanisms, such as chromosomal rearrangements, genomic copy number change, epigenetic modifications, defects in miRNA biogenesis pathway, and regulation by transcriptional factors.

Functionally, miRNAs have been suggested to contribute to tumorigenesis through their ability to regulate oncogenic signaling pathways. For example, *miR-15* and *miR-16* have been shown to target the *BCL2* oncogene, leading to its downregulation in leukemic cells and apoptosis. As another example of miRNAs' involvement in oncogenic pathways, the p53 tumor suppressor can transcriptionally induce *miR-34* following genotoxic stress, and this induction is important in mediating p53 function. The expression of miRNAs is extremely specific and there is evidence that miRNA expression patterns may be useful in distinguishing lineage and differentiation state, as well as cancer diagnosis and outcome prediction. However, no miRNA gene has ever been observed to be mutated in a cancer, either in the germline or somatically. At present, the only absolutely reliable way to causally implicate a gene in the human neoplastic process is through evidence of its mutant status. Hundreds of other genes, besides miRNAs, are expressed at higher or lower levels in cancers compared to corresponding normal tissues, and the roles of any of these genes in human cancer remain conjectural.

VIRUSES IN HUMAN CANCER

Certain human malignancies are associated with viruses. Examples include Burkitt's lymphoma (Epstein-Barr virus), hepatocellular carcinoma (hepatitis viruses), cervical cancer [human papillomavirus (HPV)], and T cell leukemia (retroviruses). The mechanisms of action of these viruses are varied but always involve activation of growth-promoting pathways or inhibition

of tumor-suppressor products in the infected cells. For example, HPV proteins E6 and E7 bind and inactivate cellular tumor suppressors p53 and pRB, respectively. Viruses are not sufficient for cancer development, but constitute one alteration in the multistep process of cancer progression.

GENE EXPRESSION IN CANCER

The tumorigenesis process, driven by alterations in tumor suppressors, oncogenes, and epigenetic regulation, is accompanied by changes in gene expression. The advent of powerful techniques for high-throughput gene expression profiling, based on sequencing or microarrays, has allowed the comprehensive study of gene expression in neoplastic cells. It is indeed possible to identify the expression levels of thousands of genes expressed in normal and cancer tissues. Figure 83-7 shows a typical microarray experiment examining gene expression in cancer. This global knowledge of gene expression allows the identification of differentially expressed genes and, in principle, the understanding of the complex molecular circuitry regulating normal and neoplastic behaviors. Such studies have led to molecular profiling of tumors, which has suggested general methods for distinguishing tumors of various biologic behaviors (molecular classification), elucidating pathways relevant to the development of tumors, and identifying molecular targets for the detection and therapy of cancer. The first practical applications of this technology have suggested that global gene expression profiling can provide prognostic information not evident from other clinical or laboratory tests. The Sanger Cancer Genome Project (http://www.sanger.ac.uk/genetics/CGP/) maintains a database dedicated to collect data on gene expression in normal and malignant tissues and make it available on the Internet. The Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) is another online data repository for expression profiling experiments.

Figure 83-7

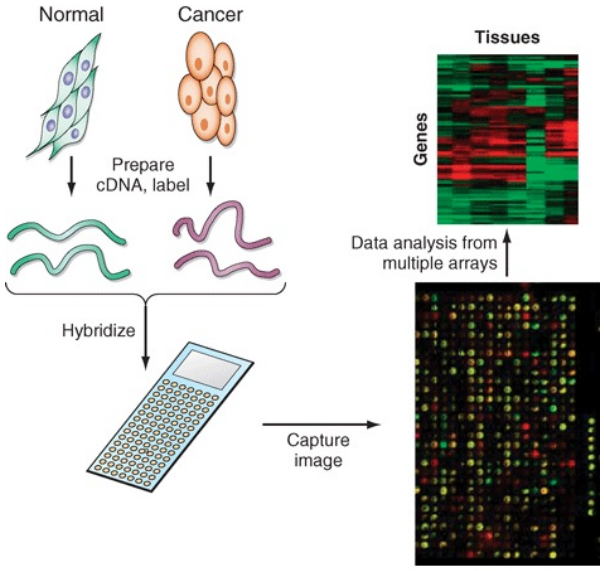

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A microarray experiment. RNA is prepared from cells, reverse transcribed to cDNA, and labeled with fluorescent dyes (typically green for normal cells and red for cancer cells). The fluorescent probes are mixed and hybridized to a cDNA array. Each spot on the array is an oligonucleotide (or cDNA fragment) that represents a different gene. The image is then captured with a fluorescence camera; red spots indicate higher expression in tumor cells compared with reference, while green spots represent the lower expression in tumor cells. Yellow signals indicate equal expression levels in normal and tumor specimens. After clustering analysis of multiple arrays, the results are typically represented graphically using a visualization software, which shows, for each sample, a color-coded representation of gene expression for every gene on the array.

GENOMEWIDE MUTATIONAL PROFILING IN CANCER

With the completion of the Human Genome Project and advances in sequencing technologies, systematic mutational analysis of the cancer genome has become possible. All protein-encoding genes known to be present in the human genome have been sequenced in breast, pancreatic, brain, and colorectal tumors. Interestingly it was found that there are generally 40 to 100 genetic alterations that affect protein sequence in a typical cancer, although statistical analyses suggested that only 8–15 are functionally involved in tumorigenesis. The picture that emerges from these studies is that most genes found mutated in tumors are actually mutated at relatively low frequencies (<5%), while a small number of genes (such as *p53*, *KRAS*) are mutated in a large proportion of tumors (Fig. 83-8). In the past, the focus of research has been on the frequently mutated genes, but it appears that the large number of genes that are infrequently mutated in cancer are major contributors to the cancer phenotype. Understanding the signaling pathways altered by mutations in these genes, as well as the functional relevance of these different mutations, represents the next challenge in the field. The Cancer Genome Atlas (http://cancergenome.nih.gov) is a coordinated effort from the National Cancer Institute and the National Human Genome Research Institute to systematically characterize the entire spectrum of genomic changes involved in human cancers.

Figure 83-8

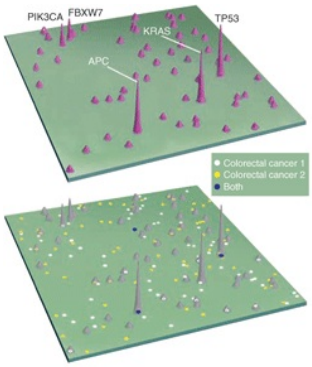

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A two-dimensional maps of genes mutated in colo-rectal cancer. The two-dimensional landscape represent the positions of the RefSeq genes along the chromosomes and the height of the peaks represent the mutation frequency. On the top map, the taller peaks represent the genes that are commonly mutated in colon cancer while the large number of smaller hills indicates the genes that are mutated at lower frequency. On the lower map, the mutations of two individual tumors are indicated. Note that there is little overlap between the mutated genes of the two colorectal tumors shown. These differences may represent the basis for the heterogeneity in terms of behavior and responsiveness to therapy observed in human cancer. *(From Wood et al: Science 318:1108, 2007, with permission.)*

PERSONALIZED CANCER TREATMENT BASED ON MOLECULAR PROFILES

Gene expression profiling and genomewide sequencing approaches have allowed for an unprecedented understanding of cancer at the molecular level. It has been suggested that individualized knowledge of pathways or genes deregulated in a given tumor (personalized genomics) may provide a guide for therapeutic options on this tumor, thus leading to personalized therapy. As tumor behavior is highly heterogeneous, even within a tumor type, personalized information-based medicine may provide a viable alternative to current one-size-fits-all therapy, especially in the case of tumors resistant to conventional therapeutic approaches. The success of this approach will be dependent on accumulated information on cancer behavior and phenotypes. For example, the identification of a particular mutation, such as that in *BRAF*, can indicate whether a certain tumor (such as melanoma) is likely to be susceptible to a specific drug that targets the mutant *BRAF* gene. Similarly, the identification of another mutation, in *KRAS*, can indicate that tumor is unlikely to be sensitive to an antibody targeting EGFR. Gene expression also offers the potential to predict drug sensitivities as well as provide prognostic information. Commercial diagnostic tests, such as Mammaprint and Oncotype DX for breast cancer, are available to help the patients and their physicians make treatment decisions. Personalized medicine is an exciting new avenue for cancer treatment based on molecular profiling, and this approach is in the process of changing our approaches to cancer therapy in fundamental ways.

THE FUTURE

A revolution in cancer genetics has occurred in the past 25 years. Identification of cancer genes has led to a deep understanding of the tumorigenesis process and has had important repercussions on all fields of cancer biology. In particular, the advancement of powerful techniques for genomewide expression profiling and mutation analyses has provided a detailed picture of the molecular defects present in individual tumors. In addition, individualized treatment based on the specific genetic alterations within some tumor types has already become possible. While these advances have not yet translated into overall changes in cancer prevention, prognosis, or treatment, it is expected that breakthroughs in these areas will continue to emerge and be applicable to an ever-increasing number of cancers.

FURTHER READINGS

Foulkes WD: Inherited susceptibility to common cancers. N Engl J Med 359:2148, 2008

Ley TJ et al: DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456:66, 2008[PMID: 18987736]

Markowitz SD, Bertagnolli M: Molecular basis of colorectal cancer. N Engl J Med 361:2449, 2009[PMID: 20018966]

Mousses S et al: Using biointelligence to search the cancer genome: an epistemological perspective on knowledge recovery strategies to enable precision medical genomics. Oncogene 27:S58, 2009

Schvartzman JM et al: Mitotic chromosomal instability and cancer: mouse modelling of the human disease. Nat Rev Cancer 10:102, 2010[PMID: 20094045]

Timp W et al: A new link between epigenetic progenitor lesions in cancer and the dynamics of signal transduction. Cell Cycle 8:383, 2009[PMID: 19177016]

Vogelstein B, Kinzler KW: The multistep nature of cancer. Trends Genet 9:138, 1993[PMID: 8516849]

Wood LD et al: The genomic landscapes of human breast and colorectal cancers. Science 318:1108, 2007[PMID: 17932254]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CANCER CELL BIOLOGY

Cancers are characterized by unregulated cell growth, tissue invasion, and metastasis. A neoplasm is *benign* when it grows in an unregulated fashion without tissue invasion. The presence of both features is characteristic of *malignant* neoplasms. Cancers are named based on their origin: those derived from epithelial tissue are called *carcinomas*, those derived from mesenchymal tissues are *sarcomas*, and those derived from hematopoietic tissue are *leukemias* or *lymphomas*.

Cancers nearly always arise as a consequence of genetic alterations. Choriocarcinoma may be an exception to this rule in that experimental insertion of a choriocarcinoma cell into an animal blastocyst can result in the neoplastic cell giving rise to normal body structures under the inductive influence of the developing embryo. Such an occurrence would be unlikely in the setting of irreversible genetic damage.

Occasional cancers appear to be caused by an alteration in a dominant gene that drives uncontrolled cell proliferation. Examples include chronic myeloid leukemia (*abl*) and Burkitt's lymphoma (*c-myc*). The genes that can promote cell growth when altered are often called *oncogenes*. They were first identified as critical elements of viruses that cause animal tumors; later it was found that the viral genes had normal counterparts with important functions in the cell and had been captured and mutated by viruses as they passed from host to host.

However, the vast majority of human cancers are characterized by multiple genetic abnormalities, each of which contributes to the loss of control of cell proliferation and differentiation and the acquisition of capabilities, such as tissue invasion and angiogenesis. Many cancers go through recognizable steps of progressively more abnormal phenotypes: hyperplasia, to adenoma, to dysplasia, to carcinoma in situ, to invasive cancer **(Table 84–1)**. These properties are not found in the normal adult cell from which the tumor is derived. Indeed, normal cells have a large number of safeguards against uncontrolled proliferation and invasion.

Table 84–1 Phenotypic Characteristics of Malignant Cells

Deregulated cell proliferation: Loss of function of negative growth regulators (suppressor oncogenes, i.e., Rb, p53), and increased action of positive growth regulators (oncogenes, i.e., *Ras, Myc*). Leads to aberrant cell cycle control and includes loss of normal checkpoint responses.
Failure to differentiate: Arrest at a stage before terminal differentiation. May retain stem cell properties. (Frequently observed in leukemias due to transcriptional repression of developmental programs by the gene products of chromosomal translocations.)
Loss of normal apoptosis pathways: Inactivation of p53, increases in Bcl-2 family members. This defect enhances the survival of cells with oncogenic mutations and genetic instability and allows clonal expansion and diversification within the tumor without activation of physiologic cell death pathways.
Genetic instability: Defects in DNA repair pathways leading to either single or oligo-nucleotide mutations (as in microsatellite instability, MIN) or more commonly chromosomal instability (CIN) leading to aneuploidy. Caused by loss of function of p53, BRCA1/2, mismatch repair genes, DNA repair enzymes, and the spindle checkpoint.
Loss of replicative senescence: Normal cells stop dividing in vitro after 25–50 population doublings. Arrest is mediated by the Rb, $p16^{INK4a}$, and p53 pathways. Further replication leads to telomere loss, with crisis. Surviving cells often harbor gross chromosomal abnormalities. Relevance to human in vivo cancer remains uncertain. Many human cancers express telomerase.
Increased angiogenesis: Due to increased gene expression of proangiogenic factors (VEGF, FGF, IL-8) by tumor or stromal cells, or loss of negative regulators (endostatin, tumstatin, thrombospondin).
Invasion: Loss of cell-cell contacts (gap junctions, cadherins) and increased production of matrix metalloproteinases (MMPs). Often takes the form of epithelial-to-mesenchymal transition (EMT), with anchored epithelial cells becoming more like motile fibroblasts.
Metastasis: Spread of tumor cells to lymph nodes or distant tissue sites. Limited by the ability of tumor cells to survive in a foreign environment.
Evasion of the immune system: Downregulation of MHC class I and II molecules; induction of T cell tolerance; inhibition of normal dendritic cell and/or T cell function; antigenic loss variants and clonal heterogeneity; increase in regulatory T cells.

Abbreviations: FGF, fibroblast growth factor; IL, interleukin; MHC, major histocompatibility complex; VEGF, vascular endothelial growth factor.

In most organs, only primitive nonfunctional cells are capable of proliferating and the cells lose the capacity to proliferate as they differentiate and acquire functional capability. The expansion of the primitive cells is linked to some functional need in the host through receptors that receive signals from the local environment or through hormonal influences delivered by the vascular supply. In the absence of such signals, the cells are at rest. We have a poor understanding of the signals that keep the primitive cells at rest. These signals, too, must be environmental, based on the observations that a regenerating liver stops growing when it has replaced the portion that has been surgically removed and regenerating bone marrow stops growing when the peripheral

blood counts return to normal. Cancer cells clearly have lost responsiveness to such controls and do not recognize when they have overgrown the niche normally occupied by the organ from which they are derived. We know very little about this mechanism of growth regulation.

Cell Cycle Checkpoints

Normal cells have a number of control mechanisms that are targeted by specific genetic alterations in cancer. The progression of a cell through the cell division cycle is regulated at a number of checkpoints by a wide array of genes. In the first phase, G_1, preparations are made to replicate the genetic material. The cell stops before entering the DNA synthesis phase or S phase to take inventory. Are we ready to replicate our DNA? Is the DNA repair machinery in place to fix any mutations that are detected? Are the DNA replicating enzymes available? Is there an adequate supply of nucleotides? Is there sufficient energy? The main brake on the process is the retinoblastoma protein, Rb. When the cell determines that it is prepared to move ahead, sequential activation of cyclin-dependent kinases (CDKs) results in the inactivation of the brake, Rb, by phosphorylation. Phosphorylated Rb releases the S-phase-regulating transcription factor, E2F/DP1, and genes required for S phase progression are expressed. If the cell determines that it is unready to move ahead with DNA replication, a number of inhibitors are capable of blocking the action of the CDKs, including $p21^{Cip2/Waf1}$, $p16^{Ink4a}$, and $p27^{Kip1}$. *Nearly every cancer has one or more genetic lesions in the G_1 checkpoint that permits progression to S phase*.

At the end of S phase, when the cell has exactly duplicated its DNA content, a second inventory is taken at the S checkpoint. Have all of the chromosomes been fully duplicated? Were any segments of DNA copied more than once? Do we have the right number of chromosomes and the right amount of DNA? If so, the cell proceeds to G_2, in which the cell prepares for division by synthesizing mitotic spindle and other proteins needed to produce two daughter cells. When DNA damage is detected, the p53 pathway is normally activated. Called the guardian of the genome, p53 is a transcription factor that is normally present in the cell in very low levels. Its level is generally regulated through its rapid turnover. Normally p53 is bound to mdm2, which transports p53 out of the nucleus for degradation in the proteosome. When damage is sensed, the ATM (ataxia-telangiectasia mutated) pathway is activated; ATM phosphorylates mdm2, which no longer binds to p53, and p53 then stops cell cycle progression, directs the synthesis of repair enzymes, or if the damage is too great, initiates apoptosis of the cell to prevent the propagation of a damaged cell **(Fig. 84-1)**.

Figure 84-1

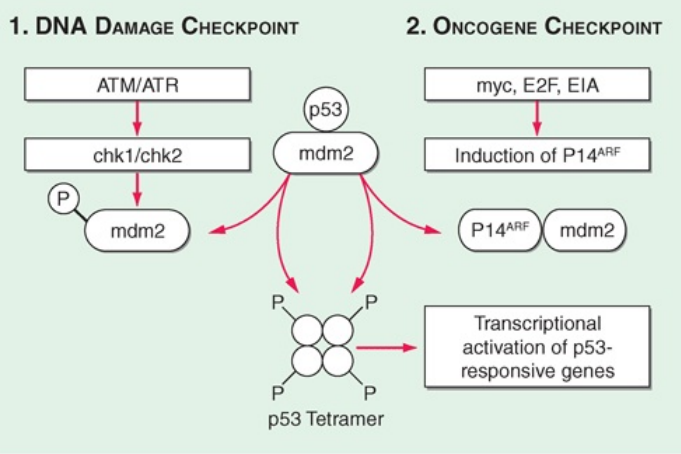

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Induction of p53 by the DNA damage and oncogene checkpoints. In response to noxious stimuli, p53 and mdm2 are phosphor-ylated by the ataxia-telangiectasia mutated (ATM) and related ATR serine/threonine kinases, as well as the immediate downstream checkpoint kinases, Chk1 and Chk2. This causes dissociation of p53 from mdm2, leading to increased p53 protein levels and transcription of genes leading to cell cycle arrest ($p21^{Cip1/Waf1}$) or apoptosis (e.g., the proapoptotic Bcl-2 family members Noxa and Puma). Inducers of p53 include hypoxemia, DNA damage (caused by ultraviolet radiation, gamma irradiation, or chemotherapy), ribonucleotide depletion, and telomere shortening. A second mechanism of p53 induction is activated by oncogenes such as *Myc*, which promote aberrant G_1/S transition. This pathway is regulated by a second product of the Ink4a locus, $p14^{ARF}$ (p19 in mice), which is encoded by an *a*lternative *r*eading *f*rame of the same stretch of DNA that codes for $p16^{Ink4a}$. Levels of ARF are upregulated by *Myc* and E2F, and ARF binds to mdm2 and rescues p53 from its inhibitory effect. This *oncogene checkpoint* leads to the death or senescence (an irreversible arrest in G_1 of the cell cycle) of renegade cells that attempt to enter S phase without appropriate physiologic signals. Senescent cells have been identified in patients whose premalignant lesions harbor activated oncogenes, for instance, dysplastic nevi that encode an activated form of BRAF (see below), demonstrating that induction of senescence is a protective mechanism that operates in humans to prevent the outgrowth of neoplastic cells.

A second method of activating p53 involves the induction by oncogenes of $p14^{ARF}$ (p19 in the mouse). ARF competes with p53 for binding to mdm2, allowing p53 to escape the effects of mdm2 and accumulate in the cell. Then p53 stops cell cycle progression by activating CDK inhibitors such as p21 and/or initiating the apoptosis pathway. Mutations in the gene for p53 on chromosome 17p are found in more than 50% of human cancers. Most

commonly these mutations are acquired in the malignant tissue in one allele and the second allele is deleted, leaving the cell unprotected from DNA-damaging agents. Some environmental exposures produce signature mutations in p53; for example, aflatoxin exposure leads to mutation of arginine to serine at codon 249 and leads to hepatocellular carcinoma. In rare instances, p53 mutations are in the germ line (Li-Fraumeni syndrome) and produce a familial cancer syndrome. The absence of p53 leads to chromosome instability and the accumulation of DNA damage including the acquisition of properties that give the abnormal cell a proliferative and survival advantage. *Like Rb dysfunction, most cancers have mutations that disable the p53 pathway*. Indeed, the importance of p53 and Rb in the development of cancer is underscored by the neoplastic transformation mechanism of human papillomavirus. This virus has two main oncogenes, E6 and E7. E6 acts to increase the rapid turnover of p53, and E7 acts to inhibit Rb function; inhibition of these two targets is sufficient to lead to neoplasia.

Another cell cycle checkpoint exists when the cell is undergoing division, the spindle checkpoint. The details of this checkpoint are still being discovered; however, it appears that if the spindle apparatus does not properly align the chromosomes for division, if the chromosome number is abnormal (i.e., greater or less than 4n), if the centromeres are not properly paired with their duplicated partners, the cell initiates a cell death pathway to prevent the production of aneuploid progeny. Abnormalities in the spindle checkpoint facilitate the development of aneuploidy. In some tumors, aneuploidy is a predominant genetic feature. In others, microsatellite instability is the primary genetic lesion. Microsatellite instability arises from defects in DNA mismatch repair genes. In general, tumors either have defects in chromosome number or microsatellite instability, but not both. Defects that lead to cancer include abnormal cell cycle checkpoints, inadequate DNA repair, and failure to preserve genome integrity.

Efforts are underway to therapeutically restore the defects in cell cycle regulation that characterize cancer.

Cancer as an Organ that Ignores Its Niche

The fundamental cellular defects that create a malignantneoplasm act at the cellular level. However, that is not the entire story. Cancers behave as organs that have lost their specialized function and stopped responding to signals that normally limit their growth. Human cancers usually become clinically detectable when a primary mass is at least 1 cm in diameter—such a mass consists of about 10^9 cells. More commonly patients present with tumors that are 10^{10} cells or greater. A lethal tumor burden is about 10^{12} cells. If all tumor cells were dividing at the time of diagnosis, patients would reach a lethal tumor burden in a very short time. However, human tumors grow by Gompertzian kinetics—this means that not every daughter cell produced by a cell division is itself capable of dividing. The growth fraction of a tumor declines exponentially with time. The growth fraction of the first malignant cell is 100%, and by the time a patient presents for medical care, the growth fraction is 2–3% or less. This fraction is similar to the growth fraction of normal bone marrow and normal intestinal epithelium, the most highly proliferative normal tissues in the human body, a fact that may explain the dose-limiting toxicities of agents that target dividing cells.

The implication of these data is that the tumor is slowing its own growth over time. How does it do this? The tumor cells have multiple genetic lesions that tend to promote proliferation, yet by the time the tumor is clinically detectable, its capacity for proliferation has declined. We need to better understand how a tumor stops its own growth. A number of factors can contribute to the failure of tumor cells to proliferate in vivo. Some cells are hypoxemic and have inadequate supply of nutrients and energy. Some have sustained too much genetic damage to complete the cell cycle and have lost the capacity to undergo apoptosis. However, an important subset is not actively dividing but retains the capacity to divide and starts dividing again when the tumor mass is reduced by treatments. Just as the bone marrow increases its rate of proliferation in response to bone marrow–damaging agents, so too does the tumor seem to sense when the tumor cell numbers have been reduced and responds by increasing growth rate. However, the marrow stops growing when it has reached its production goals. Tumors do not.

It is not in the long-term interest of a cancer to kill its host. It errs when it overshoots the limits imposed by the organ niche it occupies. Additional tumor cell vulnerabilities are likely to be detected when we learn more about how normal cells respond to "stop" signals from their environment and how tumor cells fail to heed such signals.

Is In Vitro Senescence Relevant to Carcinogenesis?

When normal cells are placed in culture in vitro, most are not capable of sustained growth. Fibroblasts are an exception to this rule. When they are cultured, fibroblasts may divide 30–50 times and then they undergo what has been termed a "crisis" during which the majority of cells stop dividing (usually due to an increase in p21 expression, a CDK inhibitor), many die, and a small fraction emerge that have acquired genetic changes that permit their uncontrolled growth. The cessation of growth of normal cells in culture has been termed "senescence" and whether this phenomenon is relevant to any physiologic event in vivo is debated.

Among the cellular changes during in vitro propagation is telomere shortening. DNA polymerase is unable to replicate the tips of chromosomes, resulting in the loss of DNA at the specialized ends of chromosomes (called *telomeres*) with each replication cycle. At birth, human telomeres are 15- to 20-kb pairs long and are composed of tandem repeats of a six-nucleotide sequence (TTAGGG) that associates with specialized telomere-binding proteins to form a T-loop structure that protects the ends of chromosomes from being mistakenly recognized as damaged. The loss of telomeric repeats with each cell division cycle causes gradual telomere shortening, leading to growth arrest (called *senescence*) when one or more critically short telomeres trigger a p53-regulated DNA-damage checkpoint response. Cells can bypass this growth arrest if pRb and p53 are nonfunctional, but cell death ensues when the unprotected ends of chromosomes lead to chromosome fusions or other catastrophic DNA rearrangements. *The ability to bypass telomere-based growth limitations is thought to be a critical step in the evolution of most malignancies*. This occurs by the reactivation of telomerase expression in cancer cells. Telomerase is an enzyme that adds TTAGGG repeats onto the 3' ends of chromosomes. It contains a catalytic subunit with reverse transcriptase activity (hTERT) and an RNA component that provides the template for telomere extension. Most normal somatic cells do not express sufficient telomerase to prevent telomere attrition

with each cell division. Exceptions include stem cells (such as those found in hematopoietic tissues, gut and skin epithelium, and germ cells) that require extensive cell division to maintain tissue homeostasis. More than 90% of human cancers express high levels of telomerase that prevent telomere shortening to critical levels and allow indefinite cell proliferation. In vitro experiments indicate that inhibition of telomerase activity leads to tumor cell apoptosis. Major efforts are underway to develop methods to inhibit telomerase activity in cancer cells. The reverse transcriptase activity of telomerase is a prime target for small-molecule pharmaceuticals. In addition, the protein component of telomerase (hTERT) may act as a tumor-associated antigen and be targeted by vaccine approaches.

All of the known functions of telomerase relate to cell division. Thus, it is unclear how short telomeres interfere with the differentiated functions of normal cells. Nevertheless, a major growth industry in medical research has been discovering an association between short telomeres and human diseases ranging from diabetes and coronary artery disease to Alzheimer's disease. The picture is further complicated by the fact that rare genetic defects in the telomerase enzyme seem to cause pulmonary fibrosis, but not hematopoietic failure or defects in nutrient absorption in the gut, the sites that might be presumed to be most sensitive to defective cell proliferation. Much remains to be learned about how telomere shortening and telomere maintenance is related to human illness in general and cancer in particular.

Signal Transduction Pathways in Cancer Cells

Signals that affect cell behavior come from adjacent cells, the stroma in which the cells are located, hormonal signals that originate remotely, and from the cells themselves (autocrine signaling). These signals generally exert their influence on the receiving cell through activation of signal transduction pathways that have as their end result the induction of activated transcription factors that mediate a change in cell behavior or function or the acquisition of effector machinery to accomplish a new task. Although signal transduction pathways can lead to a wide variety of outcomes, many such pathways rely on cascades of signals that sequentially activate different proteins or glycoproteins and lipids or glycolipids, and the activation steps often involve the addition or removal of one or more phosphate groups on a downstream target. Other chemical changes can result from signal transduction pathways, but phosphorylation and dephosphorylation play a major role. The protein kinases are generally of two distinct classes; one class acts on tyrosine residues and the other acts on serine/threonine residues. The tyrosine kinases often play critical roles in signal transduction pathways; they may be receptor tyrosine kinases or they may be linked to other cell-surface receptors through associated docking proteins **(Fig. 84-2)**.

Figure 84-2

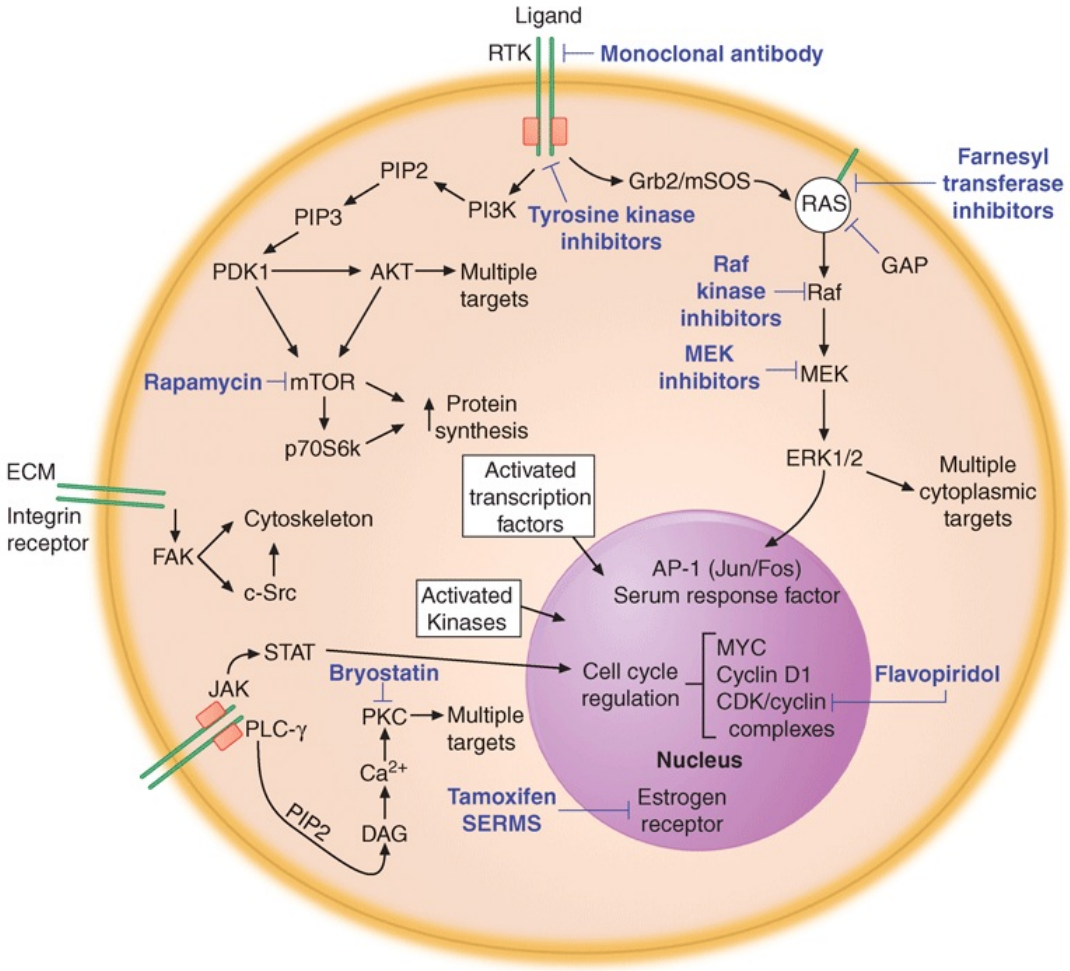

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Therapeutic targeting of signal transduction pathways in cancer cells. Three major signal transduction pathways are activated by receptor tyrosine kinases (RTK). **1.** The protooncogene Ras is activated by the Grb2/mSOS guanine nucleotide exchange factor, which induces an association with Raf and activation of downstream kinases (MEK and ERK1/2). **2.** Activated PI3K phosphorylates the membrane lipid PIP_2 to generate PIP_3, which acts as a membrane-docking site for a number of cellular proteins including the serine/threonine kinases PDK1 and Akt. PDK1 has numerous cellular targets, including Akt and mTOR. Akt phosphorylates target proteins that promote resistance to apoptosis and enhance cell cycle progression, while mTOR and its target p70S6K upregulate protein synthesis to potentiate cell growth. **3.** Activation of PLC-γ leads the formation of diacylglycerol (DAG) and increased intracellular calcium, with activation of multiple isoforms of PKC and other enzymes regulated by the calcium/calmodulin system. Other important signaling pathways involve non-RTKs that are activated by cytokine or integrin receptors. Janus kinases (JAK) phosphorylate STAT (signal transducer and activator of transcription) transcription factors, which translocate to the nucleus and activate target genes. Integrin receptors mediate cellular interactions with the extracellular matrix (ECM), inducing activation of FAK (focal adhesion kinase) and c-Src, which activate multiple downstream pathways, including modulation of the cell cytoskeleton. Many activated kinases and transcription factors migrate into the nucleus, where they regulate gene transcription, thus completing the path from extracellular signals, such as growth factors, to a change in cell phenotype, such as induction of differentiation or cell proliferation. The nuclear targets of these processes include transcription factors (e.g., Myc, AP-1, and serum response factor) and the cell cycle machinery (CDKs and cyclins). Inhibitors of many of these pathways have been developed for the treatment of human cancers. Examples of inhibitors that are currently being evaluated in clinical trials are shown in purple type.

Normally, tyrosine kinase activity is short-lived and reversed by protein tyrosine phosphatases (PTPs). However, in many human cancers, tyrosine kinases or components of their downstream pathways are activated by mutation, gene amplification, or chromosomal translocations. Because these pathways regulate proliferation, survival, migration, and angiogenesis, they have been identified as important targets for cancer therapeutics.

Inhibition of kinase activity is effective in the treatment of a number of neoplasms. Lung cancers with mutations in the epidermal growth factor receptor are highly responsive to erlotinib and gefitinib (Table 84–2). Lung cancers with activation of the anaplastic lymphoma kinase (ALK) respond to crizotinib, an ALK inhibitor. A BRAF inhibitor is highly effective in melanomas and thyroid cancers in which BRAF is overexpressed. Janus kinase inhibitors are active in myeloproliferative syndromes in which JAK2 activation is a pathogenetic event. Imatinib is an effective agent in tumors that overexpress c-Abl (such as chronic myeloid leukemia), c-Kit (gastrointestinal stromal cell tumors), or platelet-derived growth factor receptor (PDGFR; chronic myelomonocytic leukemia); second-generation congeners, dasatinib, and nilotinib are even more effective. Sorafenib and sunitinib, agents that inhibit a large number of kinases, are being widely tested and have shown promising antitumor activity in renal cell cancer and hepatocellular carcinoma. Inhibitors of the mammalian target of rapamycin (mTOR) such as

temsirolimus are also active in renal cell cancer. The list of active agents and treatment indications is growing rapidly. These new agents have ushered in a new era of personalized therapy. It is becoming more routine for resected tumors to be assessed for specific molecular changes that predict response and to have clinical decision-making guided by those results.

Table 84–2 Some FDA-Approved Molecularly Targeted Agents for the Treatment of Cancer

Drug	Molecular Target	Disease	Mechanism of Action
All-*trans* retinoic acid (ATRA)	PML-RARα oncogene	Acute promyelocytic leukemia M3 AML; t(15;17)	Inhibits transcriptional repression by PML-RARα
Imatinib (Gleevec) Dasatinib (Sprycel) Nilotinib (Tasigna)	Bcr-Abl, c-Abl, c-Kit, PDGFR-α/β	Chronic myeloid leukemia; GIST	Blocks ATP binding to tyrosine kinase active site
Sunitinib (Sutent)	c-Kit, VEGFR-2, PDGFR-β, Flt-3	GIST; renal cell cancer	Inhibits activated c-Kit and PDGFR in GIST; inhibits VEGFR in RCC
Sorafenib (Nexavar)	RAF, VEGFR-2, PDGFR-α/β, Flt-3, c-Kit	RCC; hepatocellular carcinoma	Targets VEGFR pathways in RCC. Possible activity against BRAF in melanoma, colon cancer, and others
Erlotinib (Tarceva)	EGFR	Non-small cell lung cancer; pancreatic cancer	Competitive inhibitor of the ATP-binding site of the EGFR
Gefitinib (Iressa)	EGFR	Non-small cell lung cancer	Inhibitor of EGFR tyrosine kinase
Bortezomib (Velcade)	Proteasome	Multiple myeloma	Inhibits proteolytic degradation of multiple cellular proteins
Monoclonal Antibodies			
Trastuzumab (Herceptin)	HER2/neu (ERBB2)	Breast cancer	Binds HER2 on tumor cell surface and induces receptor internalization
Cetuximab (Erbitux)	EGFR	Colon cancer, squamous cell carcinoma of the head and neck	Binds extracellular domain of EGFR and blocks binding of EGF and TGF-α; induces receptor internalization. Potentiates the efficacy of chemotherapy and radiotherapy
Panitumumab (Vectibix)	EGFR	Colon cancer	Like cetuximab; likely to be very similar in clinical activity
Rituximab (Rituxan)	CD20	B cell lymphomas and leukemias that express CD20	Multiple potential mechanisms, including direct induction of tumor cell apoptosis and immune mechanisms
Alemtuzumab (Campath)	CD52	Chronic lymphocytic leukemia and CD52-expressing lymphoid tumors	Immune mechanisms
Bevacizumab (Avastin)	VEGF	Colon, lung, breast cancers; data pending in other tumors	Inhibits angiogenesis by high-affinity binding to VEGF

Abbreviations: AML, acute myeloid leukemia; EGFR, epidermal growth factor receptor; Flt-3, fms-like tyrosine kinase-3; GIST, gastrointestinal stromal tumor; PDGFR, platelet-derived growth factor receptor; PML-RARα, promyelocytic leukemia-retinoic acid receptor-alpha; RCC, renal cell cancer; t(15;17), translocation between chromosomes 15 and 17; TGF-α, transforming growth factor-alpha; VEGFR, vascular endothelial growth factor receptor.

However, it must be acknowledged that none of these therapies is curative in any malignancy. The reasons for the failure to cure are not all defined. However, at least some causes of resistance are known. In some tumors, resistance to kinase inhibitors is related to an acquired mutation in the target kinase that inhibits drug binding. Many of these kinase inhibitors act as competitive inhibitors of the ATP-binding pocket. ATP is the phosphate donor in these phosphorylation reactions. Mutation in the BCR-ABL kinase in the ATP-binding pocket (such as the tyrosine to isoleucine change at codon 315) can prevent imatinib binding. Other resistance mechanisms include altering other signal transduction pathways to bypass the inhibited pathway. Some kinase inhibitors are less specific for an oncogenic target than was hoped, and toxicities related to off-target kinase inhibition limit the use of the agent at a dose that would inhibit the cancer-relevant kinase. As resistance mechanisms become better defined, rational strategies to overcome resistance will emerge.

Another strategy to enhance the antitumor effects of targeted agents is to use them in rational combinations with each other and in empiric combinations with chemotherapy agents that kill cells in ways distinct from targeted

agents. For example, in the c-Kit overexpressing gastrointestinal stromal tumor (GIST), resistance to imatinib develops due to secondary mutations in c-Kit, and many of these tumors are susceptible to treatment with the multitargeted tyrosine kinase (TK) inhibitor sunitinib that has activity against c-Kit as well as the PDGF and vascular endothelial growth factor (VEGF) receptors. Sunitinib is approved by the U.S. Food and Drug Administration for treatment of patients with imatinib-resistant GIST or who are intolerant of imatinib (Table 84–2). Interestingly, tumors with mutations in exon 11 of c-Kit's juxtamembrane region are particularly sensitive to imatinib, whereas those with exon 9 mutations (extracellular domain) respond better to sunitinib than imatinib. In the future, primary therapy for GIST may be determined by the specific molecular defect in c-Kit.

While targeted therapies have not yet resulted in cures when used alone, their use in the adjuvant setting and when combined with other effective treatments has substantially increased the fraction of patients cured. For example, the addition of rituximab, an anti-CD20 antibody, to combination chemotherapy in patients with diffuse large B-cell lymphoma improves cure rates by 15–20%. The addition of trastuzumab, antibody to HER2, to combination chemotherapy in the adjuvant treatment of HER2-positive breast cancer reduces relapse rates by 50%.

Targeted therapies are being developed for the ras/mitogen-activated protein (MAP) kinase pathways, the hedgehog pathway, various angiogenesis pathways, and phospholipid signaling pathways such as the phosphatidylinositol-3-kinase (PI3K) and phospholipase C-gamma pathways, which are involved in a large number of cellular processes that are important in cancer development and progression.

One of the strategies for new drug development is to take advantage of so-called oncogene addiction. This situation **(Fig. 84-3)** is created when a tumor cell develops an activating mutation in an oncogene that becomes a dominant pathway with reduced contributions from auxiliary pathways. This dependency on a single pathway creates a cell that is vulnerable to inhibitors of the oncogene pathway. For example, cells harboring mutations in *BRAF* are very sensitive to MEK inhibitors.

Figure 84-3

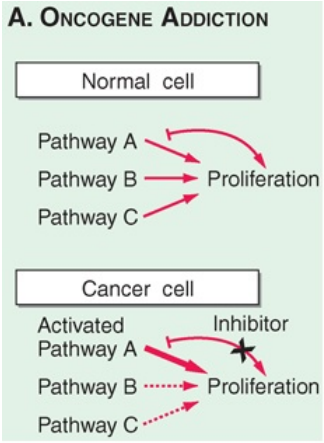

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

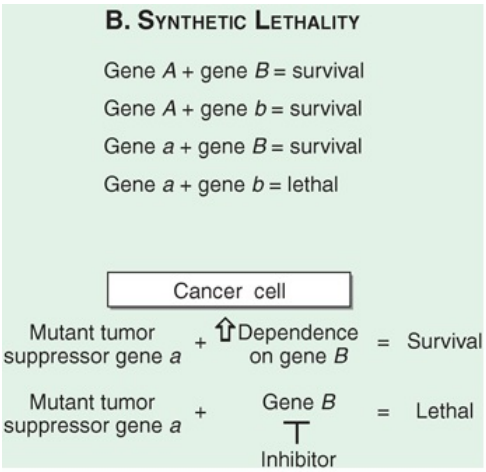

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oncogene addiction and synthetic lethality: keys to discovery of new anticancer drugs. Panel ***A.*** Normal cells receive environmental signals that activate signaling pathways (pathways A, B, and C) that together promote G_1 to S phase transition and passage through the cell cycle. Inhibition of one pathway (such as pathway A by a targeted inhibitor) has no significant effect due to redundancy provided by pathways B and C. In cancer cells, oncogenic mutations lead over time to dependency on the activated pathway, with loss of significant input from pathways B and C. The dependency or addiction of the cancer cell to pathway A makes it

highly vulnerable to inhibitors that target components of this pathway. Clinically relevant examples include Bcr-Abl (CML), amplified HER2/*neu* (breast cancer), overexpressed or mutated EGF receptors (lung cancer), and mutated *BRAF* (melanoma). Panel ***B.*** Genes are said to have a synthetic lethal relationship when mutation of either gene alone is tolerated by the cell, but mutation of both genes leads to lethality. Thus, in the example, mutant *gene a* and *gene b* have a synthetic lethal relationship, implying that the loss of one gene makes the cell dependent on the function of the other gene. In cancer cells, loss of function of a tumor-suppressor gene (wild-type designated *gene A*; mutant designated *gene a*) may render the cancer cells dependent on an alternative pathway of which *gene B* is a component. As shown in the figure, if an inhibitor of *gene B* can be identified, this can cause death of the cancer cell, without harming normal cells (which maintain wild-type function for *gene A*). High-throughput screens can now be performed using isogenic cell line pairs in which one cell line has a defined defect in a tumor-suppressor pathway. Compounds can be identified that selectively kill the mutant cell line; targets of these compounds have a synthetic lethal relationship to the tumor-suppressor pathway, and are potentially important targets for future therapeutics. Note that this approach allows discovery of drugs that indirectly target deleted tumor-suppressor genes and hence greatly expands the list of physiologically relevant cancer targets.

Many transcription factors are activated by phosphorylation, which can be prevented by tyrosine- or serine/threonine kinase inhibitors. The transcription factor NF-κB is a heterodimer composed of p65 and p50 subunits that associate with an inhibitor, IκB, in the cell cytoplasm. In response to growth factor or cytokine signaling, a multi-subunit kinase called IKK (IκB-kinase) phosphorylates IκB and directs its degradation by the ubiquitin/proteasome system. NF-κB, free of its inhibitor, translocates to the nucleus and activates target genes, many of which promote the survival of tumor cells. Novel drugs called *proteasome inhibitors* block the proteolysis of IκB, thereby preventing NF-κB activation. For unexplained reasons, this is selectively toxic to tumor cells. The antitumor effects of proteasome inhibitors are more complicated and involve the inhibition of the degradation of multiple cellular proteins. Proteasome inhibitors [bortezomib (Velcade)] have activity in patients with multiple myeloma, including partial and complete remissions. Inhibitors of IKK are also in development, with the hope of more selectively blocking the degradation of IκB, thus "locking" NF-κB in an inhibitory complex and rendering the cancer cell more susceptible to apoptosis-inducing agents.

Estrogen receptors (ERs) and androgen receptors, members of the steroid hormone family of nuclear receptors, are targets of inhibition by drugs used to treat breast and prostate cancers, respectively. Tamoxifen, a partial agonist and antagonist of ER function, can mediate tumor regression in metastatic breast cancer and can prevent disease recurrence in the adjuvant setting. Tamoxifen binds to the ER and modulates its transcriptional activity, inhibiting activity in the breast but promoting activity in bone and uterine epithelium. Selective estrogen receptor modulators (SERMs) have been developed with the hope of a more beneficial modulation of ER activity, i.e., antiestrogenic activity in the breast, uterus, and ovary, but estrogenic for bone, brain, and cardiovascular tissues. Aromatase inhibitors, which block the conversion of androgens to estrogens in breast and subcutaneous fat tissues, have demonstrated improved clinical efficacy compared with tamoxifen and are often used as first-line therapy in patients with ER-positive disease (Chap. 90).

EPIGENETIC INFLUENCES ON CANCER GENE TRANSCRIPTION

Chromatin structure regulates the hierarchical order of sequential gene transcription that governs differentiation and tissue homeostasis. Disruption of chromatin remodeling leads to aberrant gene expression and can induce proliferation of undifferentiated cells. *Epigenetics* is defined as changes that alter the pattern of gene expression that persist across at least one cell division but are not caused by changes in the DNA code. Epigenetic changes include alterations of chromatin structure mediated by methylation of cytosine residues in CpG dinucleotides, modification of histones by acetylation or methylation, or changes in higher-order chromosome structure **(Fig. 84-4)**. The transcriptional regulatory regions of active genes often contain a high frequency of CpG dinucleotides (referred to as *CpG islands*), which are normally unmethylated. Expression of these genes is controlled by transient association with repressor or activator proteins that regulate transcriptional activation. However, hypermethylation of promoter regions is a common mechanism by which tumor-suppressor loci are epigenetically silenced in cancer cells. Thus one allele may be inactivated by mutation or deletion (as occurs in loss of heterozygosity), while expression of the other allele is epigenetically silenced. The mechanisms that target tumor-suppressor genes for this form of gene silencing are unknown.

Figure 84-4

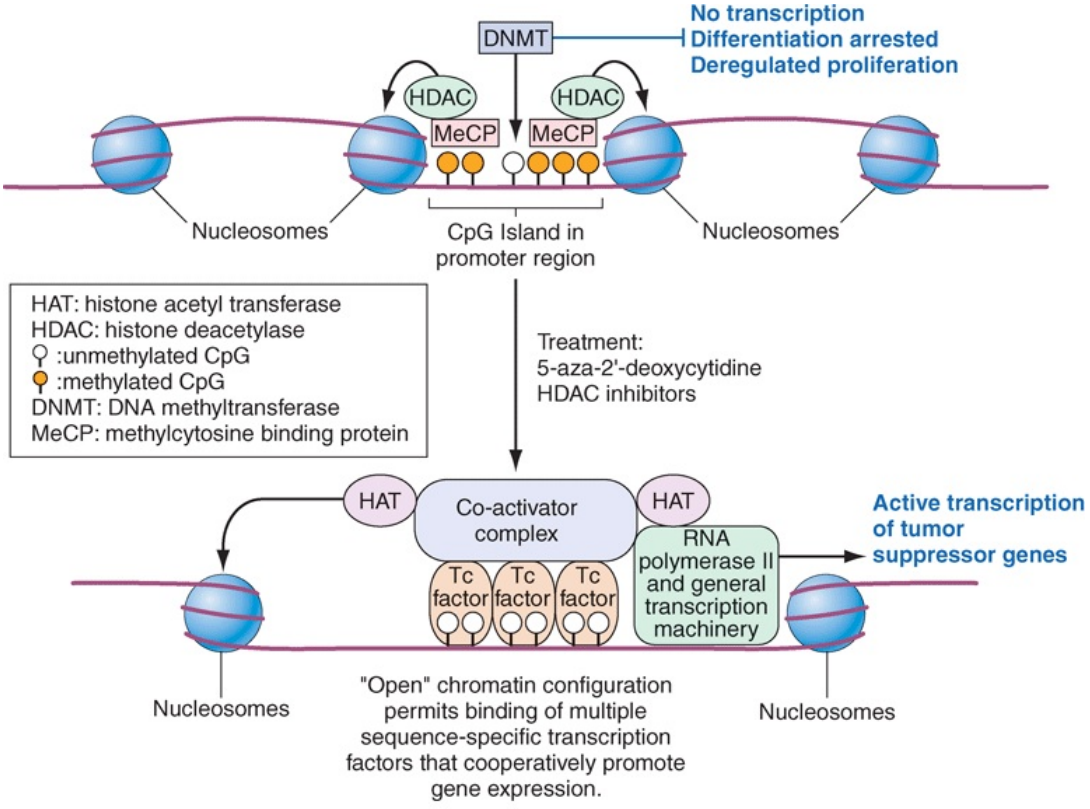

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Epigenetic regulation of gene expression in cancer cells. Tumor-suppressor genes are often epigenetically silenced in cancer cells. In the upper portion, a CpG island within the promoter and enhancer regions of the gene has been methylated, resulting in the recruitment of methyl-cytosine binding proteins (MeCP) and complexes with histone deacetylase (HDAC) activity. Chromatin is in a condensed, nonpermissive conformation that inhibits transcription. Clinical trials are under way utilizing the combination of demethylating agents such as 5-aza-2'-deoxycytidine plus HDAC inhibitors, which together confer an open, permissive chromatin structure (*lower portion*). Transcription factors bind to specific DNA sequences in promoter regions and, through protein-protein interactions, recruit coactivator complexes containing histone acetyl transferase (HAT) activity. This enhances transcription initiation by RNA polymerase II and associated general transcription factors. The expression of the tumor-suppressor gene commences, with phenotypic changes that may include growth arrest, -differentiation, or apoptosis.

Acetylation of the amino terminus of the core histones H3 and H4 induces an open chromatin conformation that promotes transcription initiation. Histone acetylases are components of coactivator complexes recruited to promoter/enhancer regions by sequence-specific transcription factors during the activation of genes (Fig. 84-4). Histone deacetylases (HDACs; at least 17 are encoded in the human genome) are recruited to genes by transcriptional repressors and prevent the initiation of gene transcription. Methylated cytosine residues in promoter regions become associated with methyl cytosine–binding proteins that recruit protein complexes with HDAC activity. The balance between permissive and inhibitory chromatin structure is therefore largely determined by the activity of transcription factors in modulating the "histone code" and the methylation status of the genetic regulatory elements of genes.

The pattern of gene transcription is aberrant in all human cancers, and in many cases, epigenetic events are responsible. Unlike genetic events that alter DNA primary structure (e.g., deletions), epigenetic changes are potentially reversible and appear amenable to therapeutic intervention. In certain human cancers, including pancreatic cancer and multiple myeloma, the $p16^{Ink4a}$ promoter is inactivated by methylation, thus permitting the unchecked activity of CDK4/cyclin D and rendering pRb nonfunctional. In sporadic forms of renal, breast, and colon cancer, the von Hippel–Lindau (*VHL*), breast cancer 1 (*BRCA1*), and serine/threonine kinase 11 (*STK11*) genes, respectively, are epigenetically silenced. Other targeted genes include the $p15^{Ink4b}$ CDK inhibitor, glutathione-S-transferase (which detoxifies reactive oxygen species), and the E-cadherin molecule (important for junction formation between epithelial cells). Epigenetic silencing can occur in premalignant lesions and can affect genes involved in DNA repair, thus predisposing to further genetic damage. Examples include MLH1 (mut L homologue) in hereditary nonpolyposis colon cancer (HNPCC, also called Lynch's syndrome), which is critical for repair of mismatched bases that occur during DNA synthesis, and O^6-methylguanine-DNA methyltransferase, which removes alkylated guanine adducts from DNA and is often silenced in colon, lung, and lymphoid tumors.

Human leukemias often have chromosomal translocations that code for novel fusion proteins with enzymatic activities that alter chromatin structure. The promyelocytic leukemia–retinoic acid receptor (PML-RAR) fusion protein, generated by the t(15;17) observed in most cases of acute promyelocytic leukemia (APL), binds to promoters containing retinoic acid response elements and recruits HDAC to these promoters, effectively inhibiting gene expression. This arrests differentiation at the promyelocyte stage and promotes tumor cell proliferation and survival. Treatment with pharmacologic doses of all-*trans* retinoic acid (ATRA), the ligand for RARα, results in the

release of HDAC activity and the recruitment of coactivators, which overcome the differentiation block. This induced differentiation of APL cells has improved treatment of these patients but also has led to a novel treatment toxicity when newly differentiated tumor cells infiltrate the lungs. However, ATRA represents a treatment paradigm for the reversal of epigenetic changes in cancer. For other leukemia-associated fusion proteins, such as acute myeloid leukemia (AML)-eight-twenty-one (ETO) and the MLL fusion proteins seen in AML and ALL, no ligand is known. Therefore, efforts are ongoing to determine the structural basis for interactions between translocation fusion proteins and chromatin-remodeling proteins and to use this information to rationally design small molecules that will disrupt specific protein-protein associations. Drugs that block the enzymatic activity of HDAC are being tested. HDAC inhibitors have demonstrated antitumor activity in clinical studies against cutaneous T cell lymphoma (e.g., vorinostat) and some solid tumors. HDAC inhibitors may target cancer cells via a number of mechanisms, including upregulation of death receptors (DR4/5, FAS, and their ligands) and $p21^{Cip1/Waf1}$, as well as inhibition of cell cycle checkpoints.

Efforts are also under way to reverse the hypermethylation of CpG islands that characterizes many solid tumors. Drugs that induce DNA demethylation, such as 5-aza-2'-deoxycytidine, can lead to reexpression of silenced genes in cancer cells with restoration of function. However, 5-aza-2'-deoxycytidine has limited aqueous solubility and is myelosuppressive. Other inhibitors of DNA methyltransferases are in development. In ongoing clinical trials, inhibitors of DNA methylation are being combined with HDAC inhibitors. The hope is that by reversing coexisting epigenetic changes, the deregulated patterns of gene transcription in cancer cells will be at least partially reversed.

Another epigenetic form of gene regulation is microRNAs. These are short (average 22 nucleotides in length) RNA molecules that silence gene expression after transcription by binding and inhibiting the translation or promoting the degradation of mRNA transcripts. It is estimated that more than 1000 microRNAs are encoded in the human genome. Each tissue has a distinctive repertoire of microRNA expression and this pattern is altered in specific ways in cancers. However, specific correlations between microRNA expression and tumor biology and clinical behavior are just now emerging. Therapies targeting microRNAs are not currently at hand but represent a novel area of treatment development.

APOPTOSIS

Tissue homeostasis requires a balance between the death of aged, terminally differentiated cells and their renewal by proliferation of committed progenitors. Genetic damage to growth-regulating genes of stem cells could lead to catastrophic results for the host as a whole. However, genetic events causing activation of oncogenes or loss of tumor suppressors, which would be predicted to lead to unregulated cell proliferation, may instead activate signal transduction pathways that block aberrant cell proliferation. These pathways can lead to programmed cell death (*apoptosis*) or irreversible growth arrest (*senescence*). Much as a panoply of intra- and extracellular signals impinge upon the core cell cycle machinery to regulate cell division, so too these signals are transmitted to a core enzymatic machinery that regulates cell death and survival.

Apoptosis is induced by two main pathways **(Fig. 84-5)**. The extrinsic pathway of apoptosis is activated by cross-linking members of the tumor necrosis factor (TNF) receptor superfamily, such as CD95 (Fas) and death receptors DR4 and DR5, by their ligands, Fas ligand or TRAIL (TNF-related apoptosis-inducing ligand), respectively. This induces the association of FADD (Fas-associated death domain) and procaspase-8 to death domain motifs of the receptors. Caspase-8 is activated and then cleaves and activates effector caspases-3 and -7, which then target cellular constituents (including caspase-activated DNAse, cytoskeletal proteins, and a number of regulatory proteins), inducing the morphologic appearance characteristic of apoptosis, which pathologists term "karyorrhexis." The intrinsic pathway of apoptosis is initiated by the release of cytochrome c and SMAC (second mitochondrial activator of caspases) from the mitochondrial intermembrane space in response to a variety of noxious stimuli, including DNA damage, loss of adherence to the extracellular matrix (ECM), oncogene-induced proliferation, and growth factor deprivation. Upon release into the cytoplasm, cytochrome c associates with dATP, procaspase-9, and the adaptor protein APAF-1, leading to the sequential activation of caspase-9 and effector caspases. SMAC binds to and blocks the function of inhibitor of apoptosis proteins (IAP), negative regulators of caspase activation.

Figure 84-5

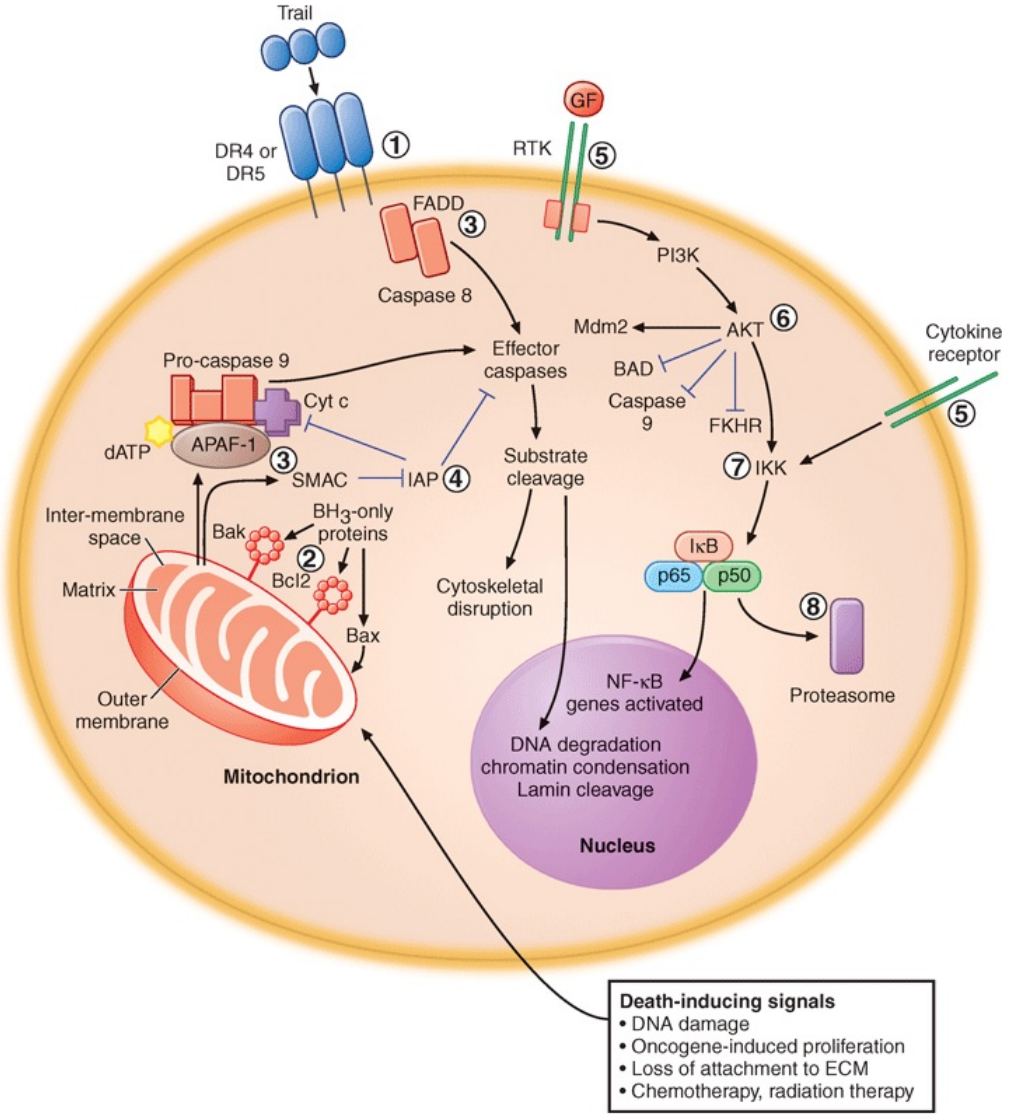

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Therapeutic strategies to overcome aberrant survival pathways in cancer cells.1. The extrinsic pathway of apoptosis can be selectively induced in cancer cells by TRAIL (the ligand for death receptors 4 and 5) or by agonistic monoclonal antibodies. **2.** Inhibition of antiapoptotic Bcl-2 family members with antisense oligonucleotides or inhibitors of the BH_3-binding pocket will promote formation of Bak- or Bax-induced pores in the mitochondrial outer membrane. **3.** Epigenetic silencing of APAF-1, caspase-8, and other proteins can be overcome using demethylating agents and inhibitors of histone deacetylases. **4.** Inhibitor of apoptosis proteins (IAP) blocks activation of caspases; small-molecule inhibitors of IAP function (mimicking SMAC action) should lower the threshold for apoptosis. **5.** Signal transduction pathways originating with activation of receptor tyrosine kinase receptors (RTKs) or cytokine receptors promote survival of cancer cells by a number of mechanisms. Inhibiting receptor function with monoclonal antibodies, such as trastuzumab or cetuximab, or inhibiting kinase activity with small-molecule inhibitors can block the pathway. **6.** The Akt kinase phosphorylates many regulators of apoptosis to promote cell survival; inhibitors of Akt may render tumor cells more sensitive to apoptosis-inducing signals; however, the possibility of toxicity to normal cells may limit the therapeutic value of these agents. **7** and **8**. Activation of the transcription factor NF-κB (composed of p65 and p50 subunits) occurs when its inhibitor, IκB, is phosphorylated by IκB-kinase (IKK), with subsequent degradation of IκB by the proteasome. Inhibition of IKK activity should selectively block the activation of NF-κB target genes, many of which promote cell survival. Inhibitors of proteasome function are FDA approved and may work in part by preventing destruction of IκB, thus blocking NF-κB nuclear localization. NF-κB is unlikely to be the only target for proteasome inhibitors.

The release of apoptosis-inducing proteins from the mitochondria is regulated by pro- and antiapoptotic members of the Bcl-2 family. Antiapoptotic members (e.g., Bcl-2, Bcl-XL, and Mcl-1) associate with the mitochondrial outer membrane via their carboxyl termini, exposing to the cytoplasm a hydrophobic binding pocket composed of Bcl-2 homology (BH) domains 1, 2, and 3 that is crucial for their activity. Perturbations of normal physiologic processes in specific cellular compartments lead to the activation of BH3-only proapoptotic family members (such as Bad, Bim, Bid, Puma, Noxa, and others) that can alter the conformation of the outer-membrane proteins Bax and Bak, which then oligomerize to form pores in the mitochondrial outer membrane resulting in cytochrome c release. If proteins comprised only by BH3 domains are sequestered by Bcl-2, Bcl-XL, or Mcl-1, pores do not form and apoptosis-inducing proteins are not released from the mitochondria. The ratio of levels of antiapoptotic Bcl-2 family members and the levels of proapoptotic BH3-only proteins at the mitochondrial membrane determines the activation state of the intrinsic pathway. The mitochondrion must therefore be recognized not only as an organelle with vital roles in intermediary metabolism and oxidative phosphorylation but also as a central regulatory structure of the apoptotic process.

The evolution of tumor cells to a more malignant phenotype requires the acquisition of genetic changes that subvert apoptosis pathways and promote cancer cell survival and resistance to anticancer therapies. However, cancer cells may be more vulnerable than normal cells to therapeutic interventions that target the apoptosis pathways that cancer cells depend upon. For instance, overexpression of Bcl-2 as a result of the t(14;18) translocation contributes to follicular lymphoma. Upregulation of Bcl-2 expression is also observed in prostate, breast, and lung cancers and melanoma. Targeting of antiapoptotic Bcl-2 family members has been accomplished by the identification of several low-molecular-weight compounds that bind to the hydrophobic pockets of either Bcl-2 or Bcl-XL and block their ability to associate with death-inducing BH3-only proteins. These compounds inhibit the antiapoptotic activities of Bcl-2 and Bcl-XL at nanomolar concentrations in the laboratory and are entering clinical trials.

Preclinical studies targeting death receptors DR4 and -5 have demonstrated that recombinant, soluble, human TRAIL or humanized monoclonal antibodies with agonist activity against DR4 or -5 can induce apoptosis of tumor cells while sparing normal cells. The mechanisms for this selectivity may include expression of decoy receptors or elevated levels of intracellular inhibitors (such as FLIP, which competes with caspase-8 for FADD) by normal cells but not tumor cells. Synergy has been shown between TRAIL-induced apoptosis and chemotherapeutic agents. For instance, some colon cancers encode mutated Bax protein as the result of mismatch repair (MMR) defects and are resistant to TRAIL. However, upregulation of Bak by chemotherapy restores the ability of TRAIL to activate the mitochondrial pathway of apoptosis. However, clinical studies have not yet shown that clinical activity correlates with activation of the extrinsic pathway of apoptosis.

Many of the signal transduction pathways perturbed in cancer promote tumor cell survival (Fig. 84-5). These include activation of the PI3K/Akt pathway, increased levels of the NF-κB transcription factor, and epigenetic silencing of genes such as APAF-1 and caspase-8. Each of these pathways is a target for therapeutic agents that, in addition to affecting cancer cell proliferation or gene expression, may render cancer cells more susceptible to apoptosis, thus promoting synergy when combined with other chemotherapeutic agents.

Some tumor cells resist drug-induced apoptosis by expression of one or more members of the ABC family of ATP-dependent efflux pumps that mediate the multidrug-resistance (MDR) phenotype. The prototype, P-glycoprotein (PGP), spans the plasma membrane 12 times and has two ATP-binding sites. Hydrophobic drugs (e.g., anthracyclines and vinca alkaloids) are recognized by PGP as they enter the cell and are pumped out. Numerous clinical studies have failed to demonstrate that drug resistance can be overcome using inhibitors of PGP. However, ABC transporters have different substrate specificities, and inhibition of a single family member may not be sufficient to overcome the MDR phenotype. Efforts to reverse PGP-mediated drug resistance continue.

Metastasis

The three major features of tissue invasion are cell adhesion to the basement membrane, local proteolysis of the membrane, and movement of the cell through the rent in the membrane and the extracellular matrix (ECM). Malignant cells that gain access to the circulation must then repeat those steps at a remote site, find a hospitable niche in a foreign tissue, avoid detection by host defenses, and induce the growth of new blood vessels. Few drugs directly target the process of metastasis. Metalloproteinase inhibitors (see "Tumor Angiogenesis," below) represent an initial attempt to inhibit the migration of tumor cells into blood and lymphatic vessels. The rate-limiting step for metastasis is the ability for tumor cells to survive and expand in the novel microenvironment of the metastatic site, and multiple host-tumor interactions determine the ultimate outcome **(Fig. 84-6)**.

Figure 84-6

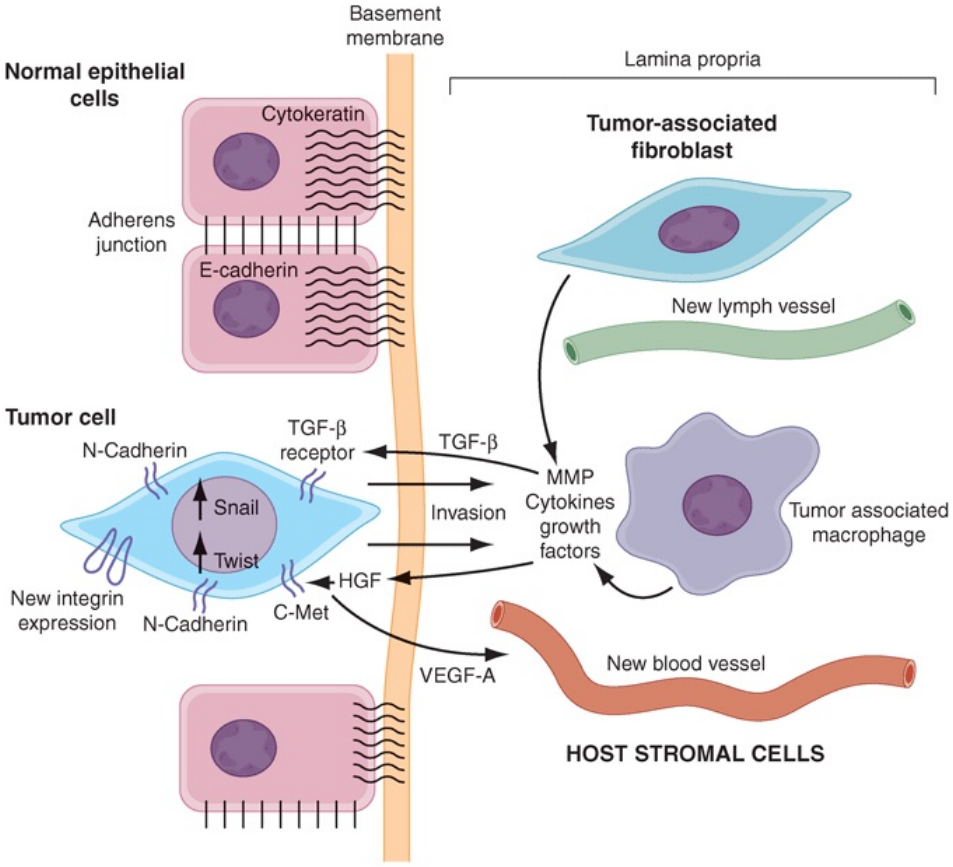

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oncogene signaling pathways are activated during tumor progression and promote metastatic potential. This figure shows a cancer cell that has undergone epithelial to mesenchymal transition (EMT) under the influence of several environmental signals. Critical components include activated transforming growth factor beta (TGF-β) and the hepatocyte growth factor (HGF)/c-Met pathways, as well as changes in the expression of adhesion molecules that mediate cell-cell and cell–extracellular matrix interactions. Important changes in gene expression are mediated by the Snail and Twist family of transcriptional repressors (whose expression is induced by the oncogenic pathways), leading to reduced expression of E-cadherin, a key component of adherens junctions between epithelial cells. This, in conjunction with upregulation of N-cadherin, a change in the pattern of expression of integrins (which mediate cell–extracellular matrix associations that are important for cell motility), and a switch in intermediate filament expression from cytokeratin to vimentin, results in the phenotypic change from adherent highly organized epithelial cells to motile and invasive cells with a fibroblast or mesenchymal morphology. EMT is thought to be an important step leading to metastasis in some human cancers. Host stromal cells, including tumor-associated fibroblasts and macrophages, play an important role in modulating tumor cell behavior through secretion of growth factors and proangiogenic cytokines, and matrix metalloproteinases that degrade the basement membrane. VEGF-A, -C, and -D are produced by tumor cells and stromal cells in response to hypoxemia or oncogenic signals, and induce production of new blood vessels and lymphatic channels through which tumor cells metastasize to lymph nodes or tissues.

The metastatic phenotype is likely restricted to a small fraction of tumor cells (Fig. 84-6). Some data suggest that cells with the appropriate capability express chemokine receptors. A number of candidate metastasis-suppressor genes have been identified. The loss of function of these genes enhances metastasis, and although the molecular mechanisms are in many cases uncertain, one common theme is enhancing the ability of the metastatic tumor cells to overcome apoptosis signals. Gene expression profiling is being used to study the metastatic process and other properties of tumor cells that may predict susceptibilities.

Bone metastases are extremely painful, cause fractures of weight-bearing bones, can lead to hypercalcemia, and are a major cause of morbidity for cancer patients. Osteoclasts and their monocyte-derived precursors express the surface receptor RANK (receptor activator of NF-κB), which is required for terminal differentiation and activation of osteoclasts. Osteoblasts and other stromal cells express RANK ligand, as both a membrane-bound and soluble cytokine. Osteoprotegerin (OPG), a soluble receptor for RANK ligand produced by stromal cells, acts as a decoy receptor to inhibit RANK activation. The relative balance of RANK ligand and OPG determines the activation state of RANK on osteoclasts. Many tumors increase osteoclast activity by secretion of substances such as parathyroid hormone (PTH), PTH-related peptide, interleukin (IL)-1, or Mip1 that perturb the homeostatic balance of bone remodeling by increasing RANK signaling. One example is multiple myeloma, where tumor cell –stromal cell interactions activate osteoclasts and inhibit osteoblasts, leading to the development of multiple lytic bone lesions. Inhibition of RANK ligand by an antibody (denosumab) can prevent further bone destruction. Bisphosphonates are also effective inhibitors of osteoclast function that are used in the treatment of cancer patients with bone metastases.

Cancer Stem Cells

Only a small proportion of the cells within a tumor are capable of initiating colonies in vitro or forming tumors at high efficiency when injected into immunocompromised NOD/SCID mice. Acute and chronic myeloid leukemias

(AML and CML) have a small population of cells (<1%) that have properties of stem cells, such as unlimited self-renewal and the capacity to cause leukemia when serially transplanted in mice. These cells have an undifferentiated phenotype ($Thy1^{-}CD34^{+}CD38^{-}$ and do not express other differentiation markers) and resemble normal stem cells in many ways, but are no longer under homeostatic control **(Fig. 84-7)**. Solid tumors may also contain a population of stem cells. Cancer stem cells, like their normal counterparts, have unlimited proliferative capacity and paradoxically traverse the cell cycle at a very slow rate; cancer growth occurs largely due to expansion of the stem cell pool, the unregulated proliferation of an amplifying population, and failure of apoptosis pathways (Fig. 84-7). Slow cell cycle progression and high levels of expression of antiapoptotic Bcl-2 family members and drug efflux pumps of the MDR family render cancer stem cells less vulnerable to cancer chemotherapy or radiation therapy. Implicit in the cancer stem cell hypothesis is the idea that failure to cure most human cancers is due to the fact that current therapeutic agents do not kill the stem cells. If cancer stem cells can be identified and isolated, then aberrant signaling pathways that distinguish these cells from normal tissue stem cells can be identified and targeted.

Figure 84-7

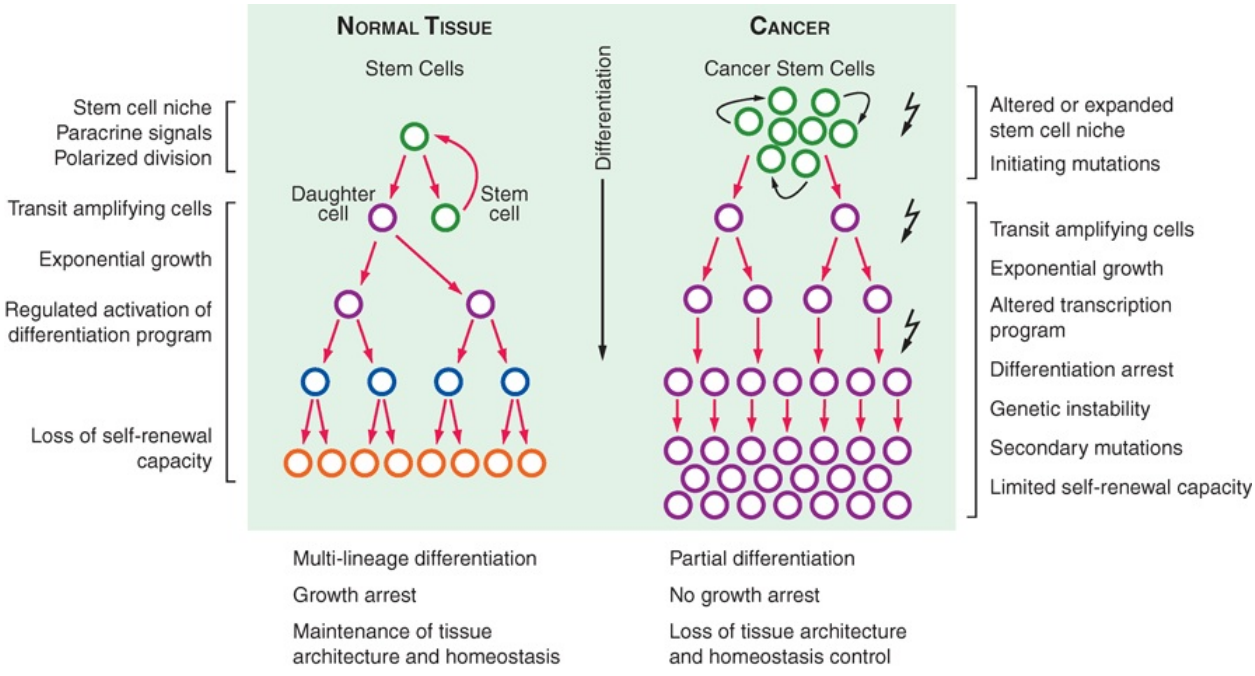

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cancer stem cells play a critical role in the initiation, progression, and resistance to therapy of malignantneoplasms. In normal tissues (left), homeostasis is maintained by asymmetric division of stem cells leading to one progeny cell that will differentiate and one cell that will maintain the stem cell pool. This occurs within highly specific niches unique to each tissue, such as in close apposition to osteoblasts in bone marrow, or at the base of crypts in the colon. Here, paracrine signals from stromal cells, such as sonic hedgehog or Notch-ligands, as well as upregulation of β-catenin and telomerase, help to maintain stem cell features of unlimited self-renewal while preventing differentiation or cell death. This occurs in part through upregulation of the transcriptional repressor Bmi-1 and inhibition of the $p16^{Ink4a}$/Arf and p53 pathways. Daughter cells leave the stem cells niche and enter a proliferative phase (referred to as *transit-amplifying*) for a specified number of cell divisions, during which time a developmental program is activated, eventually giving rise to fully differentiated cells that have lost proliferative potential. Cell renewal equals cell death, and homeostasis is maintained. In this hierarchical system, only stem cells are long-lived. The hypothesis is that cancers harbor stem cells that make up a small fraction (i.e., 0.001–1%) of all cancer cells. These cells share several features with normal stem cells, including an undifferentiated phenotype, unlimited self-renewal potential, a capacity for some degree of differentiation; however, due to initiating mutations (mutations are indicated by lightning bolts), they are no longer regulated by environmental cues. The cancer stem cell pool is expanded, and rapidly proliferating progeny, through additional mutations, may attain stem cell properties, although most of this population is thought to have a limited proliferative capacity. Differentiation programs are dysfunctional due to reprogramming of the pattern of gene transcription by oncogenic signaling pathways. Within the cancer transit-amplifying population, genomic instability generates aneuploidy and clonal heterogeneity as cells attain a fully malignant phenotype with metastatic potential. The cancer stem cell hypothesis has led to the idea that current cancer therapies may be effective at killing the bulk of tumor cells but do not kill tumor stem cells, leading to a regrowth of tumors that is manifested as tumor recurrence or disease progression. Research is in progress to identify unique molecular features of cancer stem cells that can lead to their direct targeting by novel therapeutic agents.

Oncogene Addiction and Synthetic Lethality

The concepts of oncogene addiction and synthetic lethality have spurred new drug development targeting oncogene- and tumor-suppressor pathways. As discussed earlier in this chapter and outlined in Fig. 84-3, cancer cells become dependent upon signaling pathways containing activated oncogenes; this can effect proliferation (i.e., mutated Ras, BRAF, overexpressed Myc, or activated tyrosine kinases), survival (overexpression of Bcl-2 or NF-κB), cell metabolism (as occurs when hypoxemia-inducible factor (HIF)-1α and Akt increase dependence on

glycolysis), and perhaps angiogenesis (production of VEGF, e.g., renal cell cancer). In such cases, targeted inhibition of the pathway can lead to specific killing of the cancer cells. However, targeting defects in tumor-suppressor genes has been much more difficult, since the target of the mutation is often deleted. However, identifying genes that have a synthetic lethal relationship to tumor-suppressor pathways may allow targeting of proteins required uniquely by the tumor cells (Fig. 84-3, panel *B*). Several examples of this have been identified. For instance, the von Hippel–Lindau tumor-suppressor protein is inactivated in 60% of renal cell cancers, leading to overexpression of HIF-1α and the subsequent activation of downstream genes that promote angiogenesis, proliferation, survival, and altered glucose metabolism. HIF-1α mRNA has a complex 5'-terminus that indirectly requires the activity of mTOR (via activation of p70S6K and inhibition of 4E-BP) for efficient protein translation. Inhibitors of mTOR block HIF-1α translation and have significant clinical activity in renal cell cancer. In this case, mTOR is synthetic lethal to VHL loss (Fig. 84-3), and its inhibition results in selective killing of cancer cells. Conceptually, this provides a framework for genetic screens to identify other synthetic lethal combinations involving known tumor-suppressor genes, and development of novel therapeutic agents to target dependent pathways.

TUMOR ANGIOGENESIS

The growth of primary and metastatic tumors to larger than a few millimeters requires the recruitment of blood vessels and vascular endothelial cells to support their metabolic requirements. The diffusion limit for oxygen in tissues is ~100 mm. A critical element in the growth of primary tumors and formation of metastatic sites is the *angiogenic switch*: the ability of the tumor to promote the formation of new capillaries from preexisting host vessels. The angiogenic switch is a phase in tumor development when the dynamic balance of pro- and antiangiogenic factors is tipped in favor of vessel formation by the effects of the tumor on its immediate environment. Stimuli for tumor angiogenesis include hypoxemia, inflammation, and genetic lesions in oncogenes or tumor suppressors that alter tumor cell gene expression. Angiogenesis consists of several steps, including the stimulation of endothelial cells (ECs) by growth factors, the degradation of the ECM by proteases, proliferation of ECs and migration into the tumor, and the eventual formation of new capillary tubes.

Tumor blood vessels are not normal; they have chaotic architecture and blood flow. Due to an imbalance of angiogenic regulators such as VEGF and angiopoietins (see below), tumor vessels are tortuous and dilated with an uneven diameter, excessive branching, and shunting. Tumor blood flow is variable, with areas of hypoxemia and acidosis leading to the selection of variants that are resistant to hypoxemia-induced apoptosis (often due to the loss of p53 expression). Tumor vessel walls have numerous openings, widened interendothelial junctions, and discontinuous or absent basement membrane; this contributes to the high vascular permeability of these vessels and, together with lack of functional intratumoral lymphatics, causes increased interstitial pressure within the tumor (which also interferes with the delivery of therapeutics to the tumor; Figs. 84-8, 84-9, and 84-10). Tumor blood vessels lack perivascular cells such as pericytes and smooth-muscle cells that normally regulate flow in response to tissue metabolic needs.

Unlike normal blood vessels, the vascular lining of tumor vessels is not a homogeneous layer of ECs but often consists of a mosaic of ECs and tumor cells; the concept of cancer cell–derived vascular channels, which may be lined by ECM secreted by the tumor cells, is referred to as *vascular mimicry* . It is unclear whether tumor cells actually form structural elements of vascular channels or represent tumor cells in transit into or out of the vessel. However, the former is supported by evidence that in some human colon cancers, tumor cells can comprise up to 15% of vessel walls. The ECs of angiogenic blood vessels are unlike quiescent ECs found in adult vessels, where only 0.01% of ECs are dividing. During tumor angiogenesis, ECs are highly proliferative and express a number of plasma membrane proteins that are characteristic of activated endothelium, including growth factor receptors and adhesion molecules such as integrins.

Mechanisms of Tumor Vessel Formation

Tumors use a number of mechanisms to promote vascularization, subverting normal angiogenic processes for this purpose (Fig. 84-8). Primary or metastatic tumor cells sometimes arise in proximity to host blood vessels and grow around these vessels, parasitizing nutrients by co-opting the local blood supply. However, most tumor blood vessels arise by the process of *sprouting*, in which tumors secrete trophic angiogenic molecules, the most potent being VEGF, that induce the proliferation and migration of host ECs into the tumor. Sprouting in normal and pathogenic angiogenesis is regulated by three families of transmembrane receptor tyrosine kinases (RTKs) expressed on ECs and their ligands (VEGFs, angiopoietins, ephrins; Fig. 84-9), which are produced by tumor cells, inflammatory cells, or stromal cells in the tumor microenvironment.

Figure 84-8

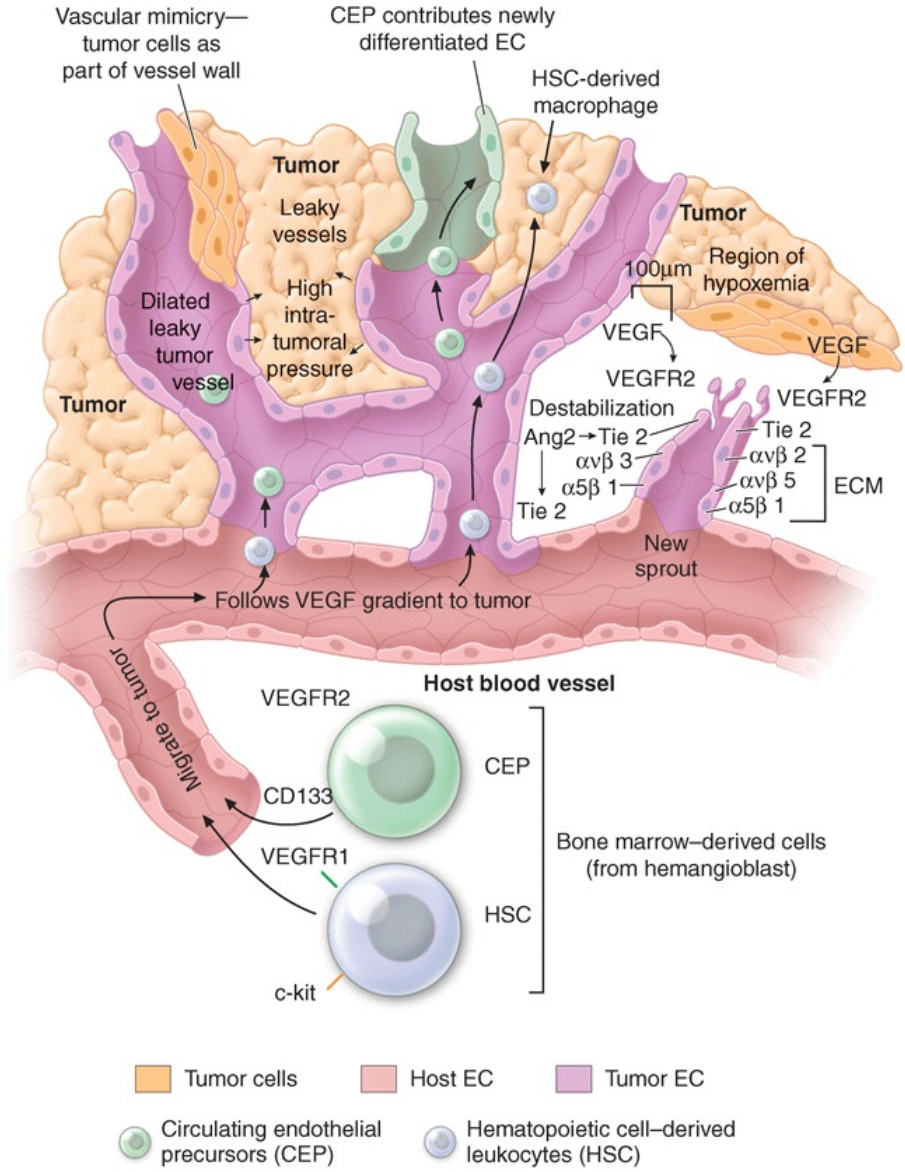

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tumor angiogenesis is a complex process involving many different cell types that must proliferate, migrate, invade, and differentiate in response to signals from the tumor microenvironment. Endothelial cells (ECs) sprout from host vessels in response to VEGF, bFGF, Ang2, and other proangiogenic stimuli. Sprouting is stimulated by VEGF/VEGFR2, Ang2/Tie-2, and integrin/extracellular matrix (ECM) interactions. Bone marrow–derived circulating endothelial precursors (CEPs) migrate to the tumor in response to VEGF and differentiate into ECs, while hematopoietic stem cells differentiate into leukocytes, including tumor-associated macrophages that secrete angiogenic growth factors and produce MMPs that remodel the ECM and release bound growth factors. Tumor cells themselves may directly form parts of vascular channels within tumors. The pattern of vessel formation is haphazard: vessels are tortuous, dilated, leaky, and branch in random ways. This leads to uneven blood flow within the tumor, with areas of acidosis and hypoxemia (which stimulate release of angiogenic factors) and high intratumoral pressures that inhibit delivery of therapeutic agents.

Figure 84-9

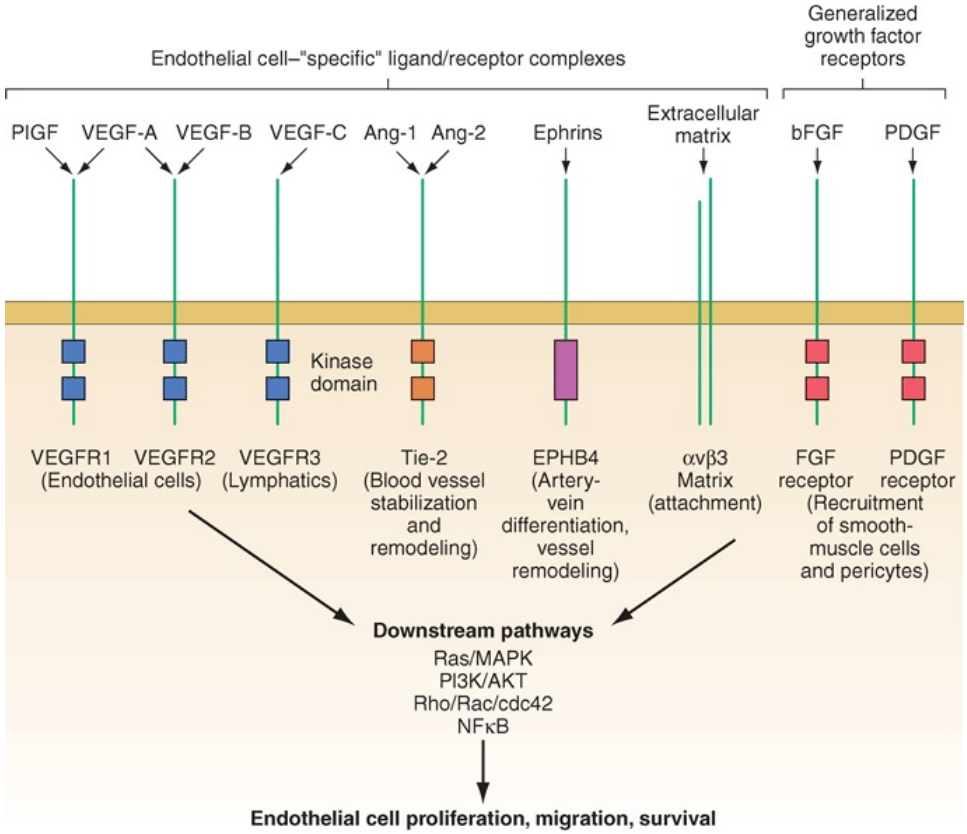

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Critical molecular determinants of endothelial cell biology. Angiogenic endothelium expresses a number of receptors not found on resting endothelium. These include receptor tyrosine kinases (RTKs) and integrins that bind to the extracellular matrix and mediate endothelial cell (EC) adhesion, migration, and invasion. ECs also express RTK (i.e., the FGF and PDGF receptors) that are found on many other cell types. Critical functions mediated by activated RTK include proliferation, migration, and enhanced survival of endothelial cells, as well as regulation of the recruitment of perivascular cells and bloodborne circulating endothelial precursors and hematopoietic stem cells to the tumor. Intracellular signaling via EC-specific RTK utilizes molecular pathways that may be targets for future antiangiogenic therapies.

Figure 84-10

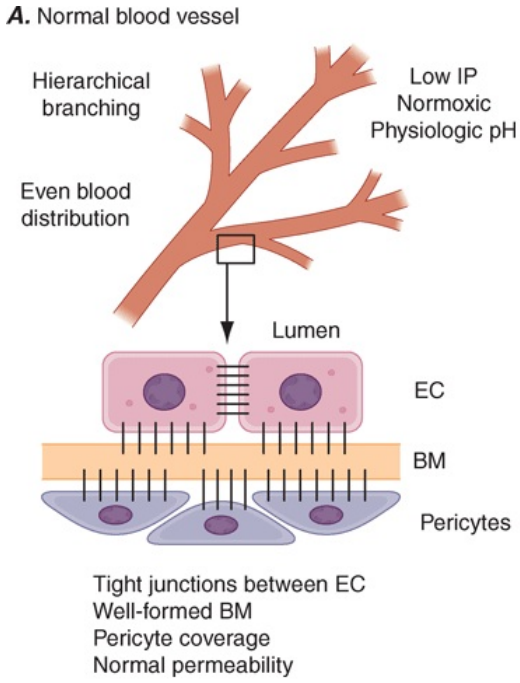

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

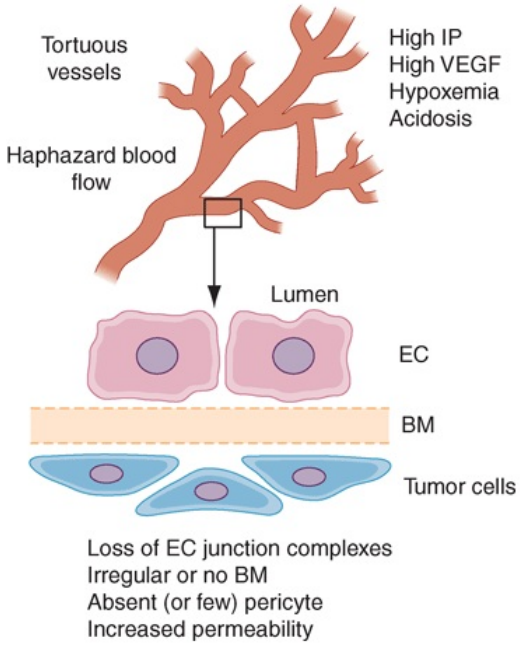

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

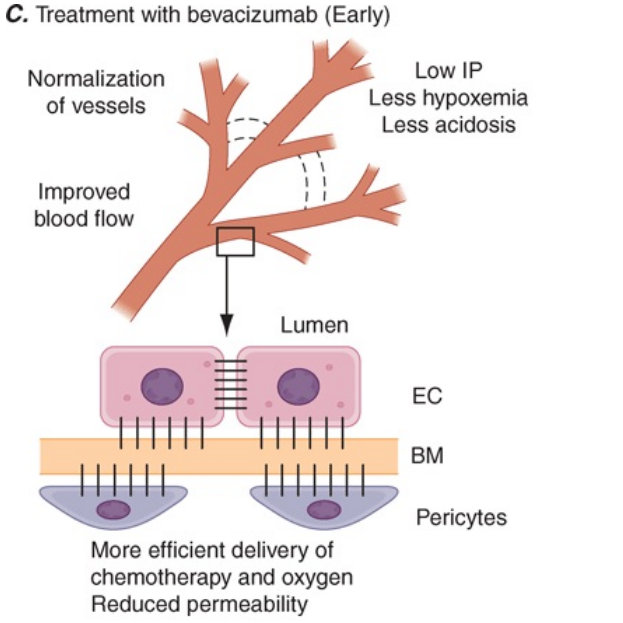

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

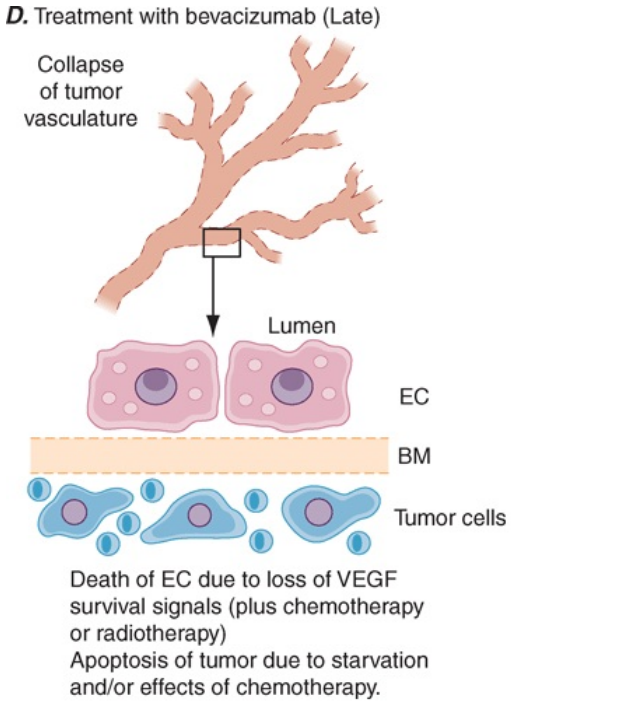

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normalization of tumor blood vessels due to inhibition of VEGF signaling. ***A.*** Blood vessels in normal tissues exhibit a regular hierarchical branching pattern that delivers blood to tissues in a spatially and temporally efficient manner to meet the metabolic needs of the tissue (top). At the microscopic level, tight junctions are maintained between endothelial cells (ECs), which are adherent to a thick and evenly distributed basement membrane (BM). Pericytes form a surrounding layer that provides trophic signals to the EC and helps maintain proper vessel tone. Vascular permeability is regulated, interstitial fluid pressure is low, and oxygen tension and pH are physiologic. ***B.*** Tumors have abnormal vessels with tortuous branching and dilated, irregular interconnecting branches, causing uneven blood flow with areas of hypoxemia and acidosis. This harsh environment selects genetic events that result in resistant tumor variants, such as the loss of p53. High levels of VEGF (secreted by tumor cells) disrupt gap junction communication, tight junctions, and adherens junctions between EC via src-mediated phosphor-ylation of proteins such as connexin 43, zonula occludens-1, VE-cadherin, and α/β-catenins. Tumor vessels have thin, irregular BM, and pericytes are sparse or absent. Together, these molecular abnormalities result in a vasculature that is permeable to serum macromolecules, leading to high tumor interstitial pressure, which can prevent the delivery of drugs to the tumor cells. This is made worse by the binding and activation of platelets at sites of exposed BM, with release of stored VEGF and microvessel clot formation, creating more abnormal blood flow and regions of hypoxemia. ***C.*** In experimental systems, treatment with bevacizumab or blocking antibodies to VEGFR2 leads to changes in the tumor vasculature that has been termed *vessel normalization*. During the first week of treatment, abnormal vessels are eliminated or pruned (dotted lines), leaving a more normal branching pattern. ECs partially regain features such as cell-cell junctions, adherence to a more normal BM, and pericyte coverage. These changes lead to a decrease in vascular permeability, reduced interstitial pressure, and a transient increase in blood flow within the tumor. Note that in murine models, this normalization period lasts only for ~5–6 days. ***D.*** After continued anti-

VEGF/VEGFR therapy (which is often combined with chemo- or radiotherapy), ECs die, leading to tumor cell death (either due to direct effects of the chemotherapy or lack of blood flow).

When tumor cells arise in or metastasize to an avascular area, they grow to a size limited by hypoxemia and nutrient deprivation. Hypoxemia, a key regulator of tumor angiogenesis, causes the transcriptional induction of the gene encoding VEGF by a process that involves stabilization of HIF-1α. Under normoxemic conditions, HIF-1α levels are maintained at a low level by proteasome-mediated destruction regulated by a ubiquitin E3-ligase encoded by the VHL tumor-suppressor locus. However, under hypoxemic conditions, HIF-1α is not hydroxylated and association with VHL does not occur; therefore HIF-1 levels increase, and target genes including VEGF, nitric oxide synthetase (NOS), and Ang2 are induced. Loss of the *VHL* genes, as occurs in familial and sporadic renal cell carcinomas, results in HIF-1α stabilization and induction of VEGF. Most tumors have hypoxemic regions due to poor blood flow, and tumor cells in these areas stain positive for HIF-1α expression; in renal cancers with *VHL* deletion, all of the tumor cells express high levels of HIF-1α, and VEGF-induced angiogenesis leads to high microvascular density.

VEGF and its receptors are required for embryonic *vasculogenesis*, and normal (wound healing, corpus luteum formation) and pathologic angiogenesis (tumor angiogenesis, inflammatory conditions such as rheumatoid arthritis). VEGF-A is a heparin-binding glycoprotein with at least four isoforms (splice variants) that regulates blood vessel formation by binding to the RTKs VEGFR1 and VEGFR2, which are expressed on all ECs in addition to a subset of hematopoietic cells (Fig. 84-8). VEGFR2 regulates EC proliferation, migration, and survival, while VEGFR1 may act as an antagonist of R2 in ECs but is probably also important for angioblast differentiation during embryogenesis. Tumor vessels may be more dependent on VEGFR signaling for growth and survival than normal ECs. While VEGF signaling is a critical initiator of angiogenesis, this is a complex process regulated by additional signaling pathways (Fig. 84-9). The angiopoietin, Ang1, produced by stromal cells, binds to the EC RTK Tie-2 and promotes the interaction of ECs with the ECM and perivascular cells, such as pericytes and smooth-muscle cells, to form tight, nonleaky vessels. PDGF and basic fibroblast growth factor (bFGF) help to recruit these perivascular cells. Ang1 is required for maintaining the quiescence and stability of mature blood vessels and prevents the vascular permeability normally induced by VEGF and inflammatory cytokines.

For tumor cell–derived VEGF to initiate sprouting from host vessels, the stability conferred by the Ang1/Tie2 pathway must be perturbed; this occurs by the secretion of Ang2 by ECs that are undergoing active remodeling. Ang2 binds to Tie2 and is a competitive inhibitor of Ang1 action: under the influence of Ang2, preexisting blood vessels become more responsive to remodeling signals, with less adherence of ECs to stroma and associated perivascular cells and more responsiveness to VEGF. Therefore, Ang2 is required at early stages of tumor angiogenesis for destabilizing the vasculature by making host ECs more sensitive to angiogenic signals. Since tumor ECs are blocked by Ang2, there is no stabilization by the Ang1/Tie2 interaction, and tumor blood vessels are leaky, hemorrhagic, and have poor association of ECs with underlying stroma. Sprouting tumor ECs express high levels of the transmembrane protein ephrin-B2 and its receptor, the RTK EPH, whose signaling appears to work with the angiopoietins during vessel remodeling. During embryogenesis, EPH receptors are expressed on the endothelium of primordial venous vessels while the transmembrane ligand ephrin-B2 is expressed by cells of primordial arteries; the reciprocal expression may regulate differentiation and patterning of the vasculature.

A number of ubiquitously expressed host molecules play critical roles in normal and pathologic angiogenesis. Proangiogenic cytokines, chemokines, and growth factors secreted by stromal cells or inflammatory cells make important contributions to neovascularization, including bFGF, transforming growth factor-α (TGF-α), TNF-α, and IL-8. In contrast to normal endothelium, angiogenic endothelium overexpresses specific members of the integrin family of ECM-binding proteins that mediate EC adhesion, migration, and survival. Specifically, expression of integrins $\alpha_v\beta_3$, $\alpha_v\beta_5$, and $\alpha_5\beta_1$ mediates spreading and migration of ECs and is required for angiogenesis induced by VEGF and bFGF, which in turn can upregulate EC integrin expression. The $\alpha_v\beta_3$ integrin physically associates with VEGFR2 in the plasma membrane and promotes signal transduction from each receptor to promote EC proliferation (via focal adhesion kinase, src, PI3K, and other pathways) and survival (by inhibition of p53 and increasing the Bcl-2/Bax expression ratio). In addition, $\alpha_v\beta_3$ forms cell-surface complexes with matrix metalloproteinases (MMPs), zinc-requiring proteases that cleave ECM proteins, leading to enhanced EC migration and the release of heparin-binding growth factors, including VEGF and bFGF. EC adhesion molecules can be upregulated (i.e., by VEGF, TNF-α) or downregulated (by TGF-β); this, together with chaotic blood flow, explains poor leukocyte-endothelial interactions in tumor blood vessels and may help tumor cells avoid immune surveillance.

Cells derived from hematopoietic progenitors in the host bone marrow contribute to tumor angiogenesis in a process linked to the secretion of VEGF and PlGF (placenta-derived growth factor) by tumor cells and their surrounding stroma. VEGF promotes the mobilization and recruitment of circulating endothelial cell precursors (CEPs) and hematopoietic stem cells (HSCs) to tumors where they co-localize and appear to cooperate in neovessel formation. CEPs express VEGFR2, while HSCs express VEGFR1, a receptor for VEGF and PlGF. Both CEPs and HSCs are derived from a common precursor, the hemangioblast. CEPs are thought to differentiate into ECs, whereas the role of HSC-derived cells (such as tumor-associated macrophages) may be to secrete angiogenic factors required for sprouting and stabilization of ECs (VEGF, bFGF, angiopoietins) and to activate MMPs, resulting in ECM remodeling and growth factor release. In mouse tumor models and in human cancers, increased numbers of CEPs and subsets of VEGFR-expressing HSCs can be detected in the circulation, which may correlate with increased levels of serum VEGF. It is not yet known whether levels of these cells have prognostic value or if changes during treatment correlate with inhibition of tumor angiogenesis. Whether CEPs and VEGFR1-expressing HSCs are required to maintain the long-term integrity of established tumor vessels is also unknown.

Lymphatic vessels also exist within tumors. Development of tumor lymphatics is associated with expression of VEGFR3 and its ligands VEGF-C and VEGF-D. The role of these vessels in tumor cell metastasis to regional lymph nodes remains to be determined, since, as discussed above, interstitial pressures within tumors are high

and most lymphatic vessels may exit in a collapsed and nonfunctional state. However, VEGF-C levels correlate significantly with metastasis to regional lymph nodes in lung, prostate, and colorectal cancers.

Antiangiogenic Therapy

ECs comprising the tumor vasculature are genetically stable and do not share genetic changes with tumor cells; the EC apoptosis pathways are therefore intact. Each EC of a tumor vessel helps provide nourishment to many tumor cells, and although tumor angiogenesis can be driven by a number of exogenous proangiogenic stimuli, experimental data indicate that at least in some tumor types, blockade of a single growth factor (e.g., VEGF) may inhibit tumor-induced vascular growth. Angiogenesis inhibitors function by targeting the critical molecular pathways involved in EC proliferation, migration, and/or survival, many of which are unique to the activated endothelium in tumors. Inhibition of growth factor and adhesion-dependent signaling pathways can induce EC apoptosis with concomitant inhibition of tumor growth. Different types of tumors use distinct molecular mechanisms to activate the angiogenic switch. Therefore, it is doubtful that a single antiangiogenic strategy will suffice for all human cancers; rather, a number of agents will be needed, each responding to distinct programs of angiogenesis used by different human cancers.

Bevacizumab, an antibody to VEGF, appears to potentiate the effects of many different types of active chemotherapeutic regimens used to treat a variety of different tumor types. It lacks single-agent antitumor activity and its strategy, the sopping up of locally produced VEGF after systemic administration, does not seem as likely to be effective as a therapy that interferes with the VEGF receptor on target cells. Bevacizumab appears to augment the antitumor effects of chemotherapy in colon cancer and additional testing in other tumor types is underway.

Bevacizumab is administered IV every 2–3 weeks (its half-life is nearly 20 days) and is generally well tolerated. Hypertension has been noted in most trials that utilize inhibitors of VEGF receptors, but only 10% of patients require treatment with antihypertensive agents and this rarely requires discontinuation of therapy. A mechanism for the hypertension may be a bevacizumab-induced decrease in vessel production of nitric oxide, resulting in vasoconstriction and increased blood pressure. Rare but serious side effects of bevacizumab include an increased risk of arterial thromboembolic events including stroke and myocardial infarction, usually in patients older than age 65 with a history of cardiovascular disease. An increased risk of hemorrhage was noted in lung cancer patients with a squamous histology and large central tumors near the major mediastinal blood vessels. Cavitation of the tumor with vessel rupture and massive hemoptysis lead to the exclusion of squamous cell cancers from treatment with bevacizumab. This potentially fatal side effect may actually reflect an increased activity of bevacizumab plus chemotherapy in squamous cell cancers. Other serious complications include bowel perforations that have been observed in 1–3% of patients (mainly those with colon and ovarian cancers).

The bevacizumab experience suggests that inhibition of the VEGF pathway will be most efficacious when combined with agents that directly target tumor cells. This also appears to be the case in the development of small-molecule inhibitors (SMI) that target VEGF receptor tyrosine kinase activity but are also inhibitory to other kinases that are expressed by tumor cells and important for their proliferation and survival. Sunitinib, FDA approved for the treatment of GIST (see above and Table 84–2), has activity directed against mutant c-Kit receptors, but also targets VEGFR and PDGFR, and has shown significant antitumor activity against metastatic renal cell carcinoma (RCC), presumably on the basis of its antiangiogenic activity. Similarly, sorafenib, originally developed as a Raf kinase inhibitor but with potent activity against VEGF and PDGF receptors, increases progression-free survival in RCC. Thus, agents that target both angiogenesis and tumor-specific signaling pathways may have greater efficacy against a broad range of cancers. A caveat is that RCC and GIST are highly dependent upon single signaling pathways (VEGF and c-Kit, respectively), whereas most solid tumors use a panoply of interconnected proliferation and survival pathways that are redundant and likely to be less amenable to single-agent targeting.

The success in targeting tumor angiogenesis has led to enhanced enthusiasm for the development of drugs that target other aspects of the angiogenic process; some of these therapeutic approaches are outlined in **Fig. 84-11**.

Figure 84-11

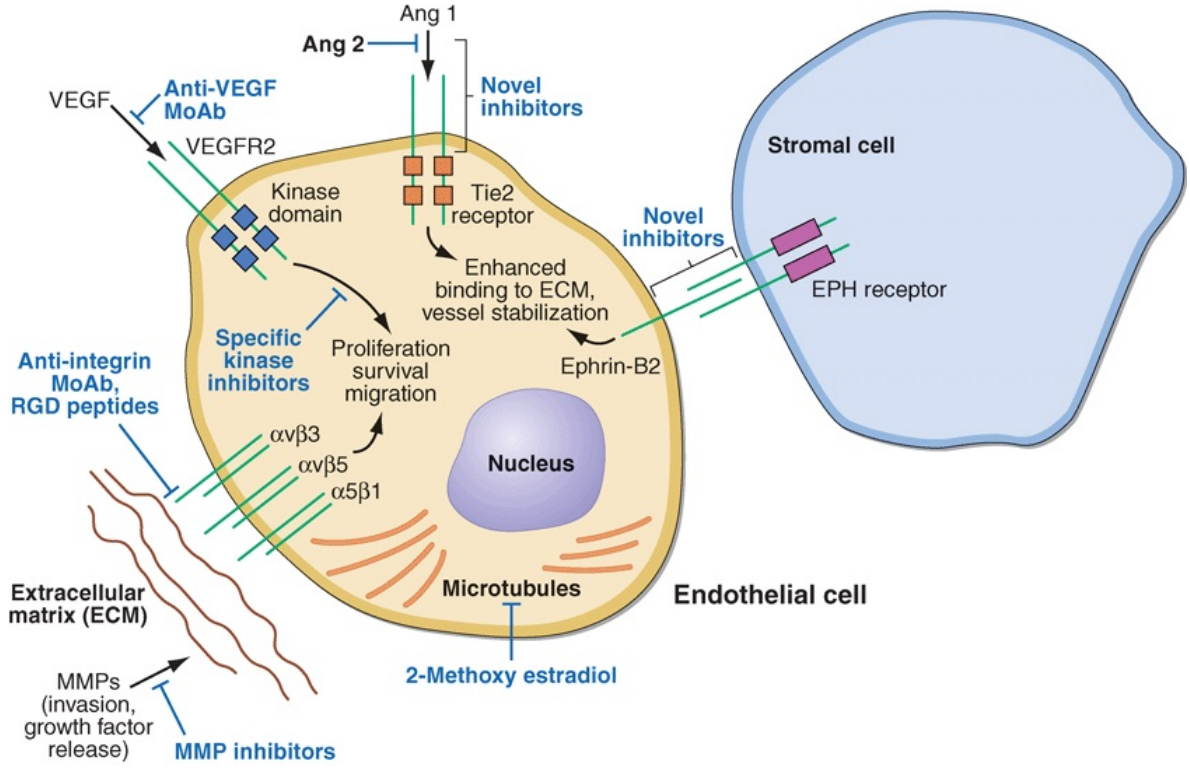

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Knowledge of the molecular events governing tumor angiogenesis has led to a number of therapeutic strategies to block tumor blood vessel formation. The successful therapeutic targeting of VEGF is described in the text. Other endothelial cell–specific receptor tyrosine kinase pathways (e.g., angiopoietin/Tie2 and ephrin/EPH) are likely targets for the future. Ligation of the $\alpha_v\beta_3$ integrin is required for EC survival. Integrins are also required for EC migration and are important regulators of matrix metalloproteinase (MMP) activity, which modulates EC movement through the ECM as well as release of bound growth factors. Targeting of integrins includes development of blocking antibodies, small peptide inhibitors of integrin signaling, and arg-gly-asp–containing peptides that prevent integrin:ECM binding. Peptides derived from normal proteins by proteolytic cleavage, including endostatin and tumstatin, inhibit angiogenesis by mechanisms that include interfering with integrin function. Signal transduction pathways that are dysregulated in tumor cells indirectly regulate EC function. Inhibition of EGF-family receptors, whose signaling activity is upregulated in a number of human cancers (e.g., breast, colon, and lung cancers), results in downregulation of VEGF and IL-8, while increasing expression of the antiangiogenic protein thrombospondin-1. The Ras/MAPK, PI3K/Akt, and Src kinase pathways constitute important antitumor targets that also regulate the proliferation and survival of tumor-derived EC. The discovery that EC from normal tissues express tissue-specific "vascular addressins" on their cell surface suggests that targeting specific EC subsets may be possible.

SUMMARY

The explosion of information on tumor cell biology, metastasis, and angiogenesis has ushered in a new era of rational targeted therapy for cancer. Furthermore, it has become clear that specific molecular factors detected in individual tumors (specific gene mutations, gene-expression profiles, microRNA expression) can be used to tailor therapy and maximize antitumor effects.

ACKNOWLEDGEMENT

Robert G. Fenton contributed to this chapter in prior editions and important material from those prior chapters has been included here.

FURTHER READINGS

Bild AH et al: Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 439:353, 2006[PMID: 16273092]

Brough R et al: Searching for synthetic lethality in cancer. Curr Opin Genet Dev 21:34, 2011[PMID: 21255997]

Dai y, Grant S: Targeting multiple arms of the apoptotic regulatory machinery. Cancer Res 67:2908, 2007[PMID: 17409392]

Finkel T et al: The common biology of cancer and ageing. Nature 448:767, 2007[PMID: 17700693]

Flaherty KT et al: Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med 363:809, 2010[PMID: 20818844]

Huber MA et al: Molecular requirements for epithelial-mesenchymal transition during tumor progression. Curr Opin Cell Biol 17:548, 2005[PMID: 16098727]

Nana-Sinkam SP, Croce CM: MicroRNA in chronic lymphocytic leukemia: Transitioning from laboratory-based investigation to clinical application. Cancer Genet Cytogenet 203:127, 2010[PMID: 21156224]

Panares RL, Garcia AA: Bevacizumab in the management of solid tumors. Expert Rev Anticancer Ther 7:434, 2007

Sharma SV et al: Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 7:169, 2007[PMID: 17318210]

Sherbenou DW, Drucker BJ: Applying the discovery of the Philadelphia chromosome. J Clin Invest 117:2068, 2007

Vousden KH, Lane DP: p53 in health and disease. Nat Rev Mol Cell Biol 8:275, 2007[PMID: 17380161]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRINCIPLES OF CANCER TREATMENT: INTRODUCTION

The goal of cancer treatment is first to eradicate the cancer. If this primary goal cannot be accomplished, the goal of cancer treatment shifts to palliation, the amelioration of symptoms, and preservation of quality of life while striving to extend life. The dictum *primum non nocere* may not always be the guiding principle of cancer therapy. When cure of cancer is possible, cancer treatments may be undertaken despite the certainty of severe and perhaps life-threatening toxicities. Every cancer treatment has the potential to cause harm, and treatment may be given that produces toxicity with no benefit. The therapeutic index of many interventions is quite narrow, and most treatments are given to the point of toxicity. Conversely, when the clinical goal is palliation, careful attention to minimizing the toxicity of potentially toxic treatments becomes a significant goal. Irrespective of the clinical scenario, the guiding principle of cancer treatment should be *primum succerrere*, "first hasten to help." Radical surgical procedures, large-field hyperfractionated radiation therapy, high-dose chemotherapy, and maximum tolerable doses of cytokines such as interleukin (IL) 2 are all used in certain settings where 100% of the patients will experience toxicity and side effects from the intervention and only a fraction of the patients will experience benefit. One of the challenges of cancer treatment is to use the various treatment modalities alone and together in a fashion that maximizes the chances for patient benefit.

Cancer treatments are divided into four main types: surgery, radiation therapy (including photodynamic therapy), chemotherapy (including hormonal therapy and molecularly targeted therapy), and biologic therapy (including immunotherapy and gene therapy). The modalities are often used in combination, and agents in one category can act by several mechanisms. For example, cancer chemotherapy agents can induce differentiation, and antibodies (a form of immunotherapy) can be used to deliver radiation therapy. Surgery and radiation therapy are considered local treatments, though their effects can influence the behavior of tumor at remote sites. Chemotherapy and biologic therapy are usually systemic treatments. *Oncology*, the study of tumors including treatment approaches, is a multidisciplinary effort with surgical-, radiotherapy-, and internal medicine–related areas of expertise. Treatments for patients with hematologicmalignancies are often shared by hematologists and medical oncologists.

In many ways, cancer mimics an organ attempting to regulate its own growth. However, cancers have not set an appropriate limit on how much growth should be permitted. Normal organs and cancers share the property of having (1) a population of cells in cycle and actively renewing and (2) a population of cells not in cycle. In cancers, cells that are not dividing are heterogeneous; some have sustained too much genetic damage to replicate but have defects in their death pathways that permit their survival, some are starving for nutrients and oxygen, and some are out of cycle but poised to be recruited back into cycle and expand if needed (i.e., reversibly growth-arrested). Severely damaged and starving cells are unlikely to kill the patient. The problem is that the cells that are reversibly not in cycle are capable of replenishing tumor cells physically removed or damaged by radiation and chemotherapy. These include *cancer stem cells*, whose properties are being elucidated. The stem cell fraction may define new targets for therapies that will retard their ability to reenter the cell cycle.

Tumors follow a Gompertzian growth curve (Fig. 85-1); the growth fraction of a neoplasm starts at 100% with the first transformed cell and declines exponentially over time until at the time of diagnosis, with a tumor burden of 1–5 x 10^9 tumor cells, the growth fraction is usually 1–4%. Thus, peak growth rate occurs before the tumor is detectable. A key feature of a successful tumor is the ability to stimulate the development of a new supporting stroma through angiogenesis and production of proteases to allow invasion through basement membranes and normal tissue barriers (Chap. 84). Specific cellular mechanisms promote entry or withdrawal of tumor cells from the cell cycle. For example, when a tumor recurs after surgery or chemotherapy, frequently its growth is accelerated and the growth fraction of the tumor is increased. This pattern is similar to that seen in regenerating organs. Partial resection of the liver results in the recruitment of cells into the cell cycle, and the resected liver volume is replaced. Similarly, chemotherapy-damaged bone marrow increases its growth to replace cells killed by chemotherapy. However, cancers do not recognize a limit on their expansion. Monoclonal gammopathy of uncertain significance may be an example of a clonal neoplasm with intrinsic features that stop its growth before a lethal tumor burden is reached. A fraction of patients with this disorder go on to develop fatal multiple myeloma, but probably this occurs because of the accumulation of additional genetic lesions. Elucidation of the mechanisms that regulate this "organ-like" behavior of tumors may provide additional clues to cancer control and treatment.

Figure 85-1

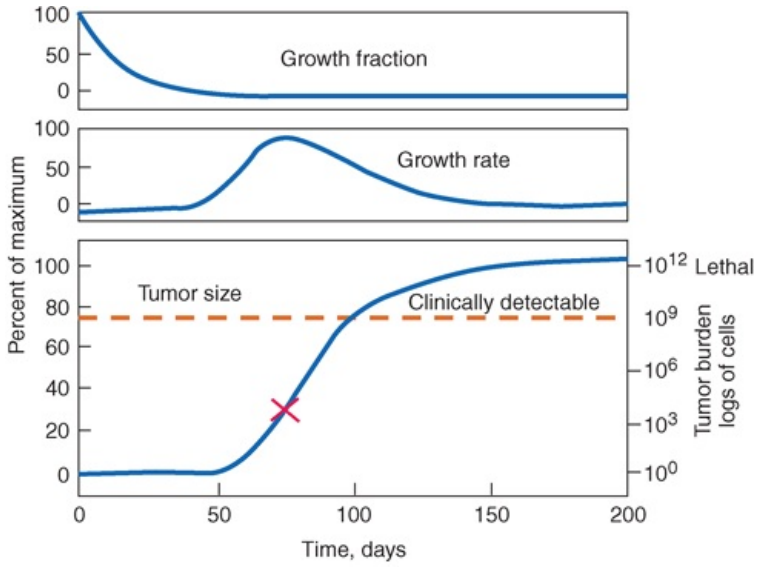

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gompertzian tumor growth. The growth fraction of a tumor declines exponentially over time *(top)*. The growth rate of a tumor peaks before it is clinically detectable *(middle)*. Tumor size increases slowly, goes through an exponential phase, and slows again as the tumor reaches the size at which limitation of nutrients or auto- or host regulatory influences can occur. The maximum growth rate occurs at 1/e, the point at which the tumor is about 37% of its maximum size *(marked with an X)*. Tumor becomes detectable at a burden of about 10^9 (1 cm^3) cells and kills the patient at a tumor cell burden of about 10^{12} (1 kg). Efforts to treat the tumor and reduce its size can result in an increase in the growth fraction and an increase in growth rate.

Figure 85-2

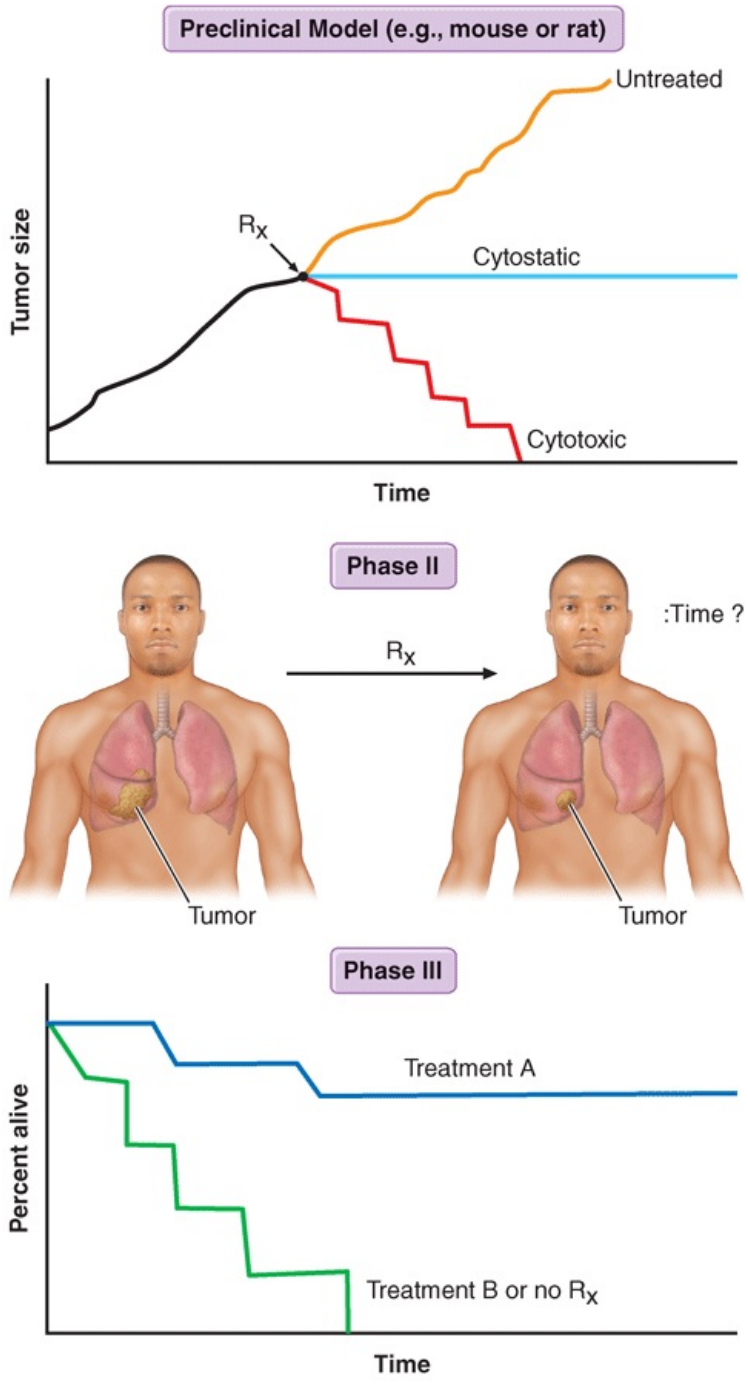

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Steps in cancer drug discovery and development. Preclinical activity *(top)* in animal models of cancers may be used as evidence to support the entry of the drug candidate into phase I trials in humans to define a correct dose and observe any clinical antitumor effect that may occur. The drug may then be advanced to phase II trials directed against specific cancer types, with rigorous quantitation of antitumor effects *(middle)*. Phase III trials then may reveal activity superior to standard or no treatment *(lowest panel)*.

PRINCIPLES OF CANCER SURGERY

Surgery is used in cancer prevention, diagnosis, staging, treatment (for both localized and metastatic disease), palliation, and rehabilitation.

Prophylaxis

Cancer can be prevented by surgery in people who have premalignant lesions resected (e.g., premalignant lesions of skin, colon, cervix) and in those who are at increased risk of cancer from either an underlying disease (colectomy in those with pancolonic involvement with ulcerative colitis), the presence of genetic lesions (colectomy for familial polyposis, thyroidectomy for multiple endocrine neoplasia type 2, bilateral mastectomy or oophorectomy for familial breast or ovarian cancer

syndromes), or a developmental anomaly (orchiectomy in those with an undescended testis). In some cases, prophylactic surgery is more radical than the surgical procedures used to treat the cancer after it develops. The assessment of risk involves many factors and should be undertaken with care before advising a patient to undergo such a major procedure. For breast cancer prevention, many experts use a 20% risk of developing breast cancer over the next 5 years as a threshold. However, patient fears play a major role in defining candidates for cancer prevention surgery. Counseling and education may not be enough to allay the fears of someone who has lost close family members to a malignancy.

Diagnosis

The underlying principle in cancer diagnosis is to obtain as much tissue as safely possible. Owing to tumor heterogeneity, pathologists are better able to make the diagnosis when they have more tissue to examine. In addition to light-microscopic inspection of a tumor for pattern of growth, degree of cellular atypia, invasiveness, and morphologic features that aid in the differential diagnosis, sufficient tissue is of value in searching for genetic abnormalities and protein expression patterns, such as hormone receptor expression in breast cancers, that may aid in differential diagnosis or provide information about prognosis or likely response to treatment. Efforts to define "personalized" information from the biology of each patient's tumor and pertinent to each patient's treatment plan are becoming increasingly important in selecting treatment options. Histologically similar tumors may have very different gene expression patterns when assessed by such techniques as microarray analysis using gene chips, with important differences in response to treatment. Such testing requires that the tissue be handled properly (e.g., immunologic detection of proteins is more effective in fresh-frozen tissue rather than in formalin-fixed tissue). Coordination among the surgeon, pathologist, and primary care physician is essential to ensure that the amount of information learned from the biopsy material is maximized.

These goals are best met by an *excisional biopsy* in which the entire tumor mass is removed with a small margin of normal tissue surrounding it. If an excisional biopsy cannot be performed, *incisional biopsy* is the procedure of second choice. A wedge of tissue is removed, and an effort is made to include the majority of the cross-sectional diameter of the tumor in the biopsy to minimize sampling error. The biopsy techniques that involve cutting into tumor carry with them a risk of facilitating the spread of the tumor. *Core-needle biopsy* usually obtains considerably less tissue, but this procedure often provides enough information to plan a definitive surgical procedure. *Fine-needle aspiration* generally obtains only a suspension of cells from within a mass. This procedure is minimally invasive, and if positive for cancer it may allow inception of systemic treatment when metastatic disease is evident, or it can provide a basis for planning a more meticulous and extensive surgical procedure.

Staging

As noted in Chap. 81, an important component of patient management is defining the extent of disease. Radiographic and other imaging tests can be helpful in defining the clinical stage; however, pathologic staging requires defining the extent of involvement by documenting the histologic presence of tumor in tissue biopsies obtained through a surgical procedure. Axillary lymph node sampling in breast cancer and lymph node sampling at laparotomy for testicular, colon, and other intraabdominal cancers may provide crucial information for treatment planning and may determine the extent and nature of primary cancer treatment.

Treatment

Surgery is the most effective means of treating cancer. Today about 40% of cancer patients are cured by surgery. Unfortunately, a large fraction of patients with solid tumors (perhaps 60%) have metastatic disease that is not accessible for removal. However, even when the disease is not curable by surgery alone, the removal of tumor can obtain important benefits, including local control of tumor, preservation of organ function, debulking that permits subsequent therapy to work better, and staging information on extent of involvement. Cancer surgery aiming for cure is usually planned to excise the tumor completely with an adequate margin of normal tissue (the margin varies with the tumor and the anatomy), touching the tumor as little as possible to prevent vascular and lymphatic spread, and minimizing operative risk. Extending the procedure to resect draining lymph nodes obtains prognostic information, but such resections alone generally do not improve survival.

Increasingly, laparoscopic approaches are being used to address primary abdominal and pelvic tumors. Lymph node spread may be assessed using the sentinel node approach, in which the first draining lymph node a spreading tumor would encounter is defined by injecting a dye into the tumor site at operation and then resecting the first node to turn blue. The sentinel node assessment is continuing to undergo clinical evaluation but appears to provide reliable information without the risks (lymphedema, lymphangiosarcoma) associated with resection of all the regional nodes. Advances in adjuvant chemotherapy and radiation therapy following surgery have permitted a substantial decrease in the extent of primary surgery necessary to obtain the best outcomes. Thus, lumpectomy with radiation therapy is as effective as modified radical mastectomy for breast cancer, and limb-sparing surgery followed by adjuvant radiation therapy and chemotherapy has replaced radical primary surgical procedures involving amputation and disarticulation for childhood rhabdomyosarcomas. More

limited surgery is also being employed to spare organ function, as in larynx and bladder cancer. The magnitude of operations necessary to optimally control and cure cancer has also been diminished by technical advances; for example, the circular anastomotic stapler has allowed narrower (<2-cm) margins in colon cancer without compromise of local control rates, and many patients who would have had colostomies are able to maintain normal anatomy.

In some settings—e.g., bulky testicular cancer or stage III breast cancer—surgery is not the first treatment modality employed. After an initial diagnostic biopsy, chemotherapy and/or radiation therapy is delivered to reduce the size of the tumor and clinically control undetected metastatic disease. Such therapy is followed by a surgical procedure to remove residual masses; this is called *neoadjuvant therapy*. Because the sequence of treatment is critical to success and is different from the standard surgery-first approach, coordination among the surgical oncologist, radiation oncologist, and medical oncologist is crucial.

Surgery may be curative in a subset of patients with metastatic disease. Patients with lung metastases from osteosarcoma may be cured by resection of the lung lesions. In patients with colon cancer who have fewer than five liver metastases restricted to one lobe and no extrahepatic metastases, hepatic lobectomy may produce long-term disease-free survival in 25% of selected patients. Surgery can also be associated with systemic antitumor effects. In the setting of hormonally responsive tumors, oophorectomy and/or adrenalectomy may control estrogen production, and orchiectomy may reduce androgen production; both have effects on metastatic tumor growth. If resection of the primary lesion takes place in the presence of metastases, acceleration of metastatic growth may occur, perhaps based on the removal of a source of angiogenesis inhibitors and mass-related growth regulators in the tumor.

In selecting a surgeon or center for primary cancer treatment, consideration must be given to the volume of cancer surgeries undertaken by the site. Studies in a variety of cancers have shown that increased annual procedure volume appears to correlate with outcome. In addition, facilities with extensive support systems—e.g., for joint thoracic and abdominal surgical teams with cardiopulmonary bypass, if needed—may allow resection of certain tumors that would otherwise not be possible.

Palliation

Surgery is employed in a number of ways for supportive care: insertion of central venous catheters, control of pleural and pericardial effusions and ascites, caval interruption for recurrent pulmonary emboli, stabilization of cancer-weakened weight-bearing bones, and control of hemorrhage, among others. Surgical bypass of gastrointestinal, urinary tract, or biliary tree obstruction can alleviate symptoms and prolong survival. Surgical procedures may provide relief of otherwise intractable pain or reverse neurologic dysfunction (cord decompression). Splenectomy may relieve symptoms and reverse hypersplenism. Intrathecal or intrahepatic therapy relies on surgical placement of appropriate infusion portals. Surgery may correct other treatment-related toxicities such as adhesions or strictures.

Rehabilitation

Surgical procedures are also valuable in restoring a cancer patient to full health. Orthopedic procedures may be necessary to ensure proper ambulation. Breast reconstruction can make an enormous impact on the patient's perception of successful therapy. Plastic and reconstructive surgery can correct the effects of disfiguring primary treatment.

PRINCIPLES OF RADIATION THERAPY

Physical Properties and Biologic Effects

Radiation is a physical form of treatment that damages any tissue in its path; its selectivity for cancer cells may be due to defects in a cancer cell's ability to repair sublethal DNA and other damage. Radiation causes breaks in DNA and generates free radicals from cell water that may damage cell membranes, proteins, and organelles. Radiation damage is augmented dependent on only oxygen; hypoxemic cells are more resistant. Augmentation of oxygen is the basis for radiation sensitization. Sulfhydryl compounds interfere with free radical generation and may act as radiation protectors.

Most radiation-induced cell damage is due to the formation of hydroxyl radicals:

The dose-response curve for cells has both linear and exponential components. The linear component is from double-strand DNA breaks produced by single hits. The exponential component represents breaks produced by multiple hits. Plotting the fraction of surviving cells against doses of x-rays or gamma radiation, the curve has a shoulder that reflects the cell's repair of sublethal damage, followed by a linear portion reflecting greater cell kill with larger doses. The features that make a particular cell more sensitive or more resistant to the biologic effects of radiation are not completely defined.

Therapeutic radiation is delivered in three ways: (1) *teletherapy*, with beams of radiation generated at a distance and aimed at the tumor within the patient; (2) *brachytherapy*, with encapsulated sources of radiation implanted directly into or adjacent to tumor tissues; and (3) *systemic therapy*, with radionuclides targeted in some fashion to a site of tumor. Teletherapy is the most commonly used form of radiation therapy.

X-rays and gamma rays are the forms of radiation most commonly used to treat cancer. They are both electromagnetic, nonparticulate waves that cause the ejection of an orbital electron when absorbed. This orbital electron ejection is called *ionization*. X-rays are generated by linear accelerators; gamma rays are generated from decay of atomic nuclei in radioisotopes such as cobalt and radium. These waves behave biologically as packets of energy, called *photons*. Particulate forms of radiation are also used in certain circumstances. Electron beams have a very low tissue penetrance and are used to treat skin conditions such as mycosis fungoides. Proton beams are becoming more widely available, and may be more able to delimit dose to tumor in certain anatomic locations. However, aside from these specialized uses, particulate forms of radiation such as neutrons, protons, and negative mesons, which should do more tissue damage because of their higher linear energy transfer and lesser dependence on oxygen, are in most applications not superior to X- or gamma rays in clinical studies reported thus far.

A number of parameters influence the damage done to tissue by radiation. Hypoxemic cells are relatively resistant. Nondividing cells are more resistant than dividing cells. In addition to these biologic parameters, physical parameters of the radiation are also crucial. The energy of the radiation determines its ability to penetrate tissue. Low-energy orthovoltage beams (150–400 kV) scatter when they strike the body, much like light diffuses when it strikes particles in the air. Such beams result in more damage to adjacent normal tissues and less radiation delivered to the tumor. Megavoltage radiation (>1 MeV) has very low lateral scatter; this produces a skin-sparing effect, more homogeneous distribution of the radiation energy, and greater deposit of the energy in the tumor, or *target volume*. The tissues that the beam passes through to get to the tumor are called the *transit volume*. The maximum dose in the target volume is often the cause of complications to tissues in the transit volume, and the minimum dose in the target volume influences the likelihood of tumor recurrence. Dose homogeneity in the target volume is the goal. Computational approaches and delivery of many beams to converge on a target lesion are the basis for "gamma knife" and related approaches to deliver high dose to small volumes of tumor, sparing normal tissue.

Radiation is quantitated on the basis of the amount of radiation absorbed in the patient; it is not based on the amount of radiation generated by the machine. The *rad* (*r*adiation *a*bsorbed *d*ose) is defined as 100 ergs of energy per gram of tissue. The International System (SI) unit for rad is the Gray (Gy); 1 Gy = 100 rad. Radiation dose is measured by placing detectors at the body surface or calculating the dose based on radiating phantoms that resemble human form and substance. Radiation dose has three determinants: total absorbed dose, number of fractions, and time. A frequent error is to omit the number of fractions and the duration of treatment. This is analogous to saying that a runner completed a race in 20 s; without knowing how far he or she ran, the result is difficult to interpret. The time could be very good for a 200-m race or very poor for a 100-m race. Thus, a typical course of radiation therapy should be described as 4500 cGy delivered to a particular target (e.g., mediastinum) over 5 weeks in 180-cGy fractions. Most curative radiation treatment programs are delivered once a day, 5 days a week in 150- to 200-cGy fractions.

Certain drugs used in cancer treatment may also act as radiation sensitizers. For example, compounds that incorporate into DNA and alter its stereochemistry (e.g., halogenated pyrimidines, cisplatin) augment radiation effects, as does hydroxyurea, another DNA synthesis inhibitor.

Application to Patients Teletherapy

Radiation therapy can be used alone or together with chemotherapy to produce cure of localized tumors and control of the primary site of disease in tumors that have disseminated. Therapy is planned based on the use of a simulator with the treatment field or fields designed to accommodate an individual patient's anatomic features. Individualized treatment planning employs lead shielding tailored to shape the field and limit the radiation exposure of normal tissue. Often the radiation is delivered from two or three different positions. Conformal three-dimensional treatment planning permits the delivery of higher doses of radiation to the target volume without increasing complications in the transit volume.

Radiation therapy is a component of curative therapy for a number of diseases, including breast cancer, Hodgkin's disease, head and neck cancer, prostate cancer, and gynecologic cancers. Radiation therapy can also palliate disease symptoms in a variety of settings: relief of bone pain from metastatic disease, control of brain metastases, reversal of spinal cord compression and superior vena caval obstruction, shrinkage of painful masses, and opening of threatened airways. In high-risk settings, radiation therapy can prevent the development of leptomeningeal disease and brain metastases in acute leukemia and lung cancer.

BRACHYTHERAPY

Brachytherapy involves placing a sealed source of radiation into or adjacent to the tumor and withdrawing the radiation source after a period of time precisely calculated to deliver a chosen dose of radiation to the tumor. This approach is often used to treat prostate tumors and cervical cancer. The difficulty with brachytherapy is the short range of radiation effects (the inverse square law) and the inability to shape the radiation to fit the target volume. Normal tissue may receive toxic exposure to the radiation, with attendant radiation enteritis or cystitis in cervix cancer or brain injury in brain tumors.

RADIONUCLIDES AND RADIOIMMUNOTHERAPY

Nuclear medicine physicians or radiation oncologists may administer radionuclides with therapeutic effects. Iodine 131 is used to treat thyroid cancer since iodine is naturally taken up preferentially by the thyroid; it emits gamma rays that destroy the normal thyroid as well as the tumor. Strontium 89 and samarium 153 are two radionuclides that are preferentially taken up in bone, particularly sites of new bone formation. Both are capable of controlling bone metastases and the pain associated with them, but the dose-limiting toxicity is myelosuppression.

Monoclonal antibodies and other ligands can be attached to radioisotopes by conjugation (for nonmetal isotopes) or by chelation (for metal isotopes), and the targeting moiety can result in the accumulation of the radionuclide preferentially in tumor. Iodine 131–labeled anti-CD20 and yttrium 90 –labeled anti-CD20 are active in B cell lymphoma, and other labeled antibodies are being evaluated. Thyroid uptake of labeled iodine is blocked by cold iodine. Dose-limiting toxicity is myelosuppression.

PHOTODYNAMIC THERAPY

Some chemical structures (porphyrins, phthalocyanines) are selectively taken up by cancer cells by mechanisms not fully defined. When light, usually delivered by a laser, is shone on cells containing these compounds, free radicals are generated and the cells die. Hematoporphyrins and light are being used with increasing frequency to treat skin cancer; ovarian cancer; and cancers of the lung, colon, rectum, and esophagus. Palliation of recurrent locally advanced disease can sometimes be dramatic and last many months.

Toxicity

Though radiation therapy is most often administered to a local region, systemic effects, including fatigue, anorexia, nausea, and vomiting, may develop that are related in part to the volume of tissue irradiated, dose fractionation, radiation fields, and individual susceptibility. Bone is among the most radioresistant organs, radiation effects being manifested mainly in children through premature fusion of the epiphyseal growth plate. By contrast, the male testis, female ovary, and bone marrow are the most sensitive organs. Any bone marrow in a radiation field will be eradicated by therapeutic irradiation. Organs with less need for cell renewal, such as heart, skeletal muscle, and nerves, are more resistant to radiation effects. In radiation-resistant organs, the vascular endothelium is the most sensitive component. Organs with more self-renewal as a part of normal homeostasis, such as the hematopoietic system and mucosal lining of the intestinal tract, are more sensitive. Acute toxicities include mucositis, skin erythema (ulceration in severe cases), and bone marrow toxicity. Often these can be alleviated by interruption of treatment.

Chronic toxicities are more serious. Radiation of the head and neck region often produces thyroid failure. Cataracts and retinal damage can lead to blindness. Salivary glands stop making saliva, which leads to dental caries and poor dentition. Taste and smell can be affected. Mediastinal irradiation leads to a threefold increased risk of fatal myocardial infarction. Other late vascular effects include chronic constrictive pericarditis, lung fibrosis, viscus stricture, spinal cord transection, and radiation enteritis. A serious late toxicity is the development of second solid tumors in or adjacent to the radiation fields. Such tumors can develop in any organ or tissue and occur at a rate of about 1% per year beginning in the second decade after treatment. Some organs vary in susceptibility to radiation carcinogenesis. A woman who receives mantle field radiation therapy for Hodgkin's disease at age 25 years has a 30% risk of developing breast cancer by age 55 years. This is comparable in magnitude to genetic breast cancer syndromes. Women treated after age 30 years have little or no increased risk of breast cancer. No data suggest that a threshold dose of therapeutic radiation exists below which the incidence of second cancers is decreased. High rates of second tumors occur in people who receive as little as 1000 cGy.

PRINCIPLES OF CHEMOTHERAPY

Medical oncology is the subspecialty of internal medicine that cares for and designs treatment approaches to patients with cancer, in conjunction with surgical and radiation oncologists. The core skills of the medical oncologist include the use of drugs that may have a beneficial effect on the natural history of the patient's illness or favorably influence the patient's quality of life.

Endpoints of Drug Action

The concept that systemically administered drugs may have a useful effect on cancers was historically derived from three sets of observations. Paul Ehrlich in the nineteenth century observed that different dyes reacted with different cell and tissue components. He hypothesized the existence of compounds that would be "magic bullets" that might bind to tumors, owing to the affinity of the agent for the tumor. A second observation was the toxic effects of certain mustard gas derivatives on the bone marrow during World War I, leading to the idea that smaller doses of these agents might be used to treat tumors of marrow-derived cells. Finally, the observation that certain tumors from hormone-responsive tissues, e.g. breast tumors, could shrink after oophorectomy led to the idea that endogenous substances promoting the growth of a tumor might be antagonized. Chemicals achieving each of the goals are actually or intellectually the forbearers of the currently used cancer chemotherapy agents.

Chemotherapy agents may be used for the treatment of active, clinically apparent cancer. Table 85–1, *A* lists those tumors considered curable by conventionally available chemotherapeutic agents when used to address disseminated or metastatic cancers. If a tumor is localized to a single site, serious consideration of surgery or primary radiation therapy should be given, as these treatment modalities may be curative as local treatments. Chemotherapy may be employed after the failure of these modalities to eradicate a local tumor or as part of multimodality approaches to offer primary treatment to a clinically localized tumor. In this event, it can allow organ preservation when given with radiation, as in the larynx or other upper airway sites; or sensitize tumors to radiation when given, e.g., to patients concurrently receiving radiation for lung or cervix cancer (Table 85–1, *B*). Chemotherapy can be administered as an adjuvant, i.e., in addition to surgery (Table 85–1, *C*) or radiation, after all clinically apparent disease has been removed. This use of chemotherapy may have curative potential in breast and colorectal neoplasms, as it attempts to eliminate clinically unapparent tumor that may have already disseminated. As noted above, small tumors frequently have high growth fractions and therefore may be intrinsically more susceptible to the action of antiproliferative agents. Chemotherapy is routinely used in "conventional" dose regimens. In general, these doses produce reversible acute side effects, primarily consisting of transient myelosuppression with or without gastrointestinal toxicity (usually nausea), which are readily managed. High-dose chemotherapy regimens are predicated on the observation that the dose-response curve for many anticancer agents is rather steep, and increased dose can produce markedly increased therapeutic effect, although at the cost of potentially life-threatening complications that require intensive support, usually in the form of hematopoietic stem cell support from the patient (*autologous*) or from donors matched for histocompatibility loci (*allogeneic*). High-dose regimens have definite curative potential in defined clinical settings (Table 85–1, *D*).

Table 85–1 Curability of Cancers with Chemotherapy

A. Advanced Cancers With Possible Cure
Acute lymphoid and acute myeloid leukemia (pediatric/adult)
Hodgkin's disease (pediatric/adult)
Lymphomas—certain types (pediatric/adult)
Germ cell neoplasms
Embryonal carcinoma
Teratocarcinoma
Seminoma or dysgerminoma
Choriocarcinoma
Gestational trophoblastic neoplasia
Pediatric neoplasms
Wilms' tumor
Embryonal rhabdomyosarcoma
Ewing's sarcoma
Peripheral neuroepithelioma
Neuroblastoma
Small cell lung carcinoma
Ovarian carcinoma
B. Advanced Cancers Possibly Cured by Chemotherapy and Radiation
Squamous carcinoma (head and neck)
Squamous carcinoma (anus)
Breast carcinoma
Carcinoma of the uterine cervix
Non-small cell lung carcinoma (stage lll)
Small cell lung carcinoma
C. Cancers Possibly Cured With Chemotherapy as Adjuvant to Surgery
Breast carcinoma
Colorectal carcinomaa
Osteogenic sarcoma
Soft tissue sarcoma

D. Cancers Possibly Cured with "High-Dose" Chemotherapy With Stem Cell Support
Relapsed leukemias, lymphoid and myeloid
Relapsed lymphomas, Hodgkin's and non-Hodgkin's
Chronic myeloid leukemia
Multiple myeloma
E. Cancers Responsive With Useful Palliation, But Not Cure, by Chemotherapy
Bladder carcinoma
Chronic myeloid leukemia
Hairy cell leukemia
Chronic lymphocytic leukemia
Lymphoma—certain types
Multiple myeloma
Gastric carcinoma
Cervix carcinoma
Endometrial carcinoma
Soft tissue sarcoma
Head and neck cancer
Adrenocortical carcinoma
Islet-cell neoplasms
Breast carcinoma
Colorectal carcinoma
Renal carcinoma
F. Tumor Poorly Responsive in Advanced Stages to Chemotherapy
Pancreatic carcinoma
Biliary-tract neoplasms
Thyroid carcinoma
Carcinoma of the vulva
Non-small cell lung carcinoma
Prostate carcinoma
Melanoma
Hepatocellular carcinoma
Salivary gland cancer

Karnofsky was among the first to champion the evaluation of a chemotherapeutic agent's benefit by carefully quantitating its effect on tumor size and using these measurements to objectively decide the basis for further treatment of a particular patient or further clinical evaluation of a drug's potential. A partial response (PR) is defined conventionally as a decrease by at least 50% in a tumor's bidimensional area; a complete response (CR) connotes disappearance of all tumor; progression of disease signifies an increase in size of existing lesions by >25% from baseline or best response or development of new lesions; and "stable" disease fits into none of the above categories. Newer evaluation systems such as RECIST (Response Evaluation Criteria In Solid Tumors) utilize unidimensional measurement, but the intent is similar in rigorously defining evidence for the activity of the agent in assessing its value to the patient.

If cure is not possible, chemotherapy may be undertaken with the goal of palliating some aspect of the tumor's effect on the host. Common tumors that may be meaningfully addressed with palliative intent are listed in Table 85–1, *E*. Usually, tumor-related symptoms may manifest as pain, weight loss, or some local symptom related to the tumor's effect on normal structures. Patients treated with palliative intent should be aware of their diagnosis and the limitations of the proposed treatments, have access to supportive care, and have suitable "performance status," according to assessment algorithms such as the one developed by Karnofsky or by the Eastern Cooperative Oncology Group (ECOG). ECOG performance status 0 (PS0) patients are without symptoms; PS1 patients are ambulatory but restricted in strenuous physical activity; PS2 patients are ambulatory but unable to work and are up and about 50% or more of the time; PS3 patients are capable of limited self-care and are up <50% of the time; PS4 patients are totally confined to bed or chair and incapable of self-care. Only PS0, PS1, and PS2

patients are generally considered suitable for palliative (noncurative) treatment. If there is curative potential, even poor–performance status patients may be treated, but their prognosis is usually inferior to that of good–performance status patients treated with similar regimens.

An important perspective the primary care provider may bring to patients and their families facing incurable cancer is that, given the limited value of chemotherapeutic approaches at some point in the natural history, *palliative care* or *hospice-based* approaches, with meticulous and ongoing attention to symptom relief and with family, psychological, and spiritual support, should receive prominent attention as a valuable therapeutic plan (Chap. 9). Optimizing the quality of life rather than attempting to extend it becomes a valued intervention. Patients facing the impending progression of disease in a life-threatening way frequently choose to undertake toxic treatments of little to no potential value, and support provided by the primary caregiver in accessing palliative and hospice-based options in contrast to receiving toxic and ineffective regimen can be critical in providing a basis for patients to make sensible choices.

Cancer Drugs: Overview and Principles for Use

Cancer drug treatments are of four broad types. *Conventional chemotherapy agents* were historically derived by the empirical observation that these "small molecules" (generally with molecular mass <1500 Da) could cause major regression of experimental tumors growing in animals. These agents mainly target DNA structure or segregation of DNA as chromosomes in mitosis. *Targeted agents* refer to small molecules or "biologicals" (generally macromolecules such as antibodies or cytokines) designed and developed to interact with a defined molecular target important in either maintaining the malignant state or selectively expressed by the tumor cells. As described in Chap. 84, successful tumors have activated biochemical pathways that lead to uncontrolled proliferation through the action of, e.g., oncogene products, loss of cell cycle inhibitors, or loss of cell death regulation, and have acquired the capacity to replicate chromosomes indefinitely, invade, metastasize, and evade the immune system. Targeted therapies seek to capitalize on the biology behind the aberrant cellular behavior as a basis for therapeutic effects. *Hormonal therapies* (the first form of targeted therapy) capitalize on the biochemical pathways underlying estrogen and androgen function and action as a therapeutic basis for approaching patients with tumors of breast, prostate, uterus, and ovarian origin. *Biologic therapies* are often macromolecules that have a particular target (e.g., antigrowth factor or cytokine antibodies) or may have the capacity to regulate growth of tumor cells or induce a host immune response to kill tumor cells. Thus, biologic therapies include not only antibodies but cytokines and gene therapies.

The usefulness of any drug is governed by the extent to which a given dose causes a useful result (therapeutic effect; in the case of anticancer agents, toxicity to tumor cells) as opposed to a toxic effect to the host. The *therapeutic index* is the degree of separation between toxic and therapeutic doses. Really useful drugs have large therapeutic indices, and this usually occurs when the drug target is expressed in the disease-causing compartment as opposed to the normal compartment. Classically, selective toxicity of an agent for an organ is governed by the expression of an agent's target or by differential accumulation into or elimination from compartments where toxicity is experienced or ameliorated, respectively. Currently used chemotherapeutic agents have the unfortunate property that their targets are present in both normal and tumor tissues. Therefore, they have relatively narrow therapeutic indices.

Figure 85-2 illustrates steps in cancer drug discovery and development. Following demonstration of antitumor activity in animal models, potentially useful anticancer agents are further evaluated to define an optimal schedule of administration and arrive at a drug formulation designed for a given route and schedule. Safety testing in two species on an analogous schedule of administration defines the starting dose for a phase I trial in humans. This is established as a fraction, usually one-sixth to one-tenth, of the dose just causing easily reversible toxicity in the more sensitive animal species. Escalating doses of the drug are then given during the human phase I trial until reversible toxicity is observed. Dose-limiting toxicity (DLT) defines a dose that conveys greater toxicity than would be acceptable in routine practice, allowing definition of a lower maximal tolerated dose (MTD). The occurrence of toxicity is, if possible, correlated with plasma drug concentrations. The MTD or a dose just lower than the MTD is usually the dose suitable for phase II trials, where a fixed dose is administered to a relatively homogeneous set of patients with a particular tumor type in an effort to define whether the drug causes regression of tumors. An "active" agent conventionally has PR rates of at least 20–25% with reversible non-life-threatening side effects, and it may then be suitable for study in phase III trials to assess efficacy in comparison to standard or no therapy.

Response, defined as tumor shrinkage, is but the most immediate indicator of drug effect. To be clinically valuable, responses must translate into clinical benefit. This is conventionally established by a beneficial effect on overall survival, or at least an increased time to further progression of disease. Active efforts are being made to quantitate effects of anticancer agents on quality of life. Cancer drug clinical trials conventionally use a toxicity grading scale where grade I toxicities do not require treatment, grade II often require symptomatic treatment but are not life-threatening, grade III toxicities are potentially life-threatening if untreated, grade IV toxicities are actually life-threatening, and grade V

toxicities are those that result in the patient's death.

Development of "targeted agents" may proceed quite differently. While phase I–III trials are still conducted, molecular analysis of human tumors may allow the precise definition of target expression in a patient's tumor that is necessary for or relevant to the drug's action. This information might then allow selection of patients expressing the drug target for participation in all trial phases. These patients may then have a greater chance of developing a useful response to the drug by virtue of expressing the target in the tumor. Clinical trials may be designed to incorporate an assessment of the behavior of the target in relation to the drug (pharmacodynamic studies). Ideally, the plasma concentration that affects the drug target is known, so escalation to MTD may not be necessary. Rather, the correlation of host toxicity while achieving an "optimal biologic dose" becomes a more relevant endpoint for phase I and early phase II trials with targeted agents.

Useful cancer drug treatment strategies using conventional chemotherapy agents, targeted agents, hormonal treatments, or biologicals have one of two valuable outcomes. They can induce cancer cell death, resulting in tumor shrinkage with corresponding improvement in patient survival, or increase the time until the disease progresses. Another potential outcome is to induce cancer cell *differentiation* or dormancy with loss of tumor cell replicative potential and reacquisition of phenotypic properties resembling normal cells. Blocking tumor cell differentiation may be a key feature in the pathogenesis of certain leukemias.

Cell death is a closely regulated process. *Necrosis* refers to cell death induced, for example, by physical damage with the hallmarks of cell swelling and membrane disruption. *Apoptosis*, or programmed cell death, refers to a highly ordered process whereby cells respond to defined stimuli by dying, and it recapitulates the necessary cell death observed during the ontogeny of the organism. *Anoikis* refers to the death of epithelial cells after removal from the normal milieu of substrate, particularly from cell-to-cell contact. Cancer chemotherapeutic agents can cause both necrosis and apoptosis. Apoptosis is characterized by chromatin condensation (giving rise to "apoptotic bodies"); cell shrinkage; and, in living animals, phagocytosis by surrounding stromal cells without evidence of inflammation. This process is regulated either by signal transduction systems that promote a cell's demise after a certain level of insult is achieved, or in response to specific cell-surface receptors that mediate cell death signals. Modulation of apoptosis by manipulation of signal transduction pathways has emerged as a basis for understanding the actions of drugs and designing new strategies to improve their use. *Autophagy* is a cellular response to injury where the cell does not initially die but catabolizes itself in a way that can lead to loss of replicative potential.

A general view of how cancer treatments work is that the interaction of a chemotherapeutic drug with its target induces a "cascade" of further signaling steps. These signals ultimately lead to cell death by triggering an "execution phase" where proteases, nucleases, and endogenous regulators of the cell death pathway are activated (Fig. 85-3).

Figure 85-3

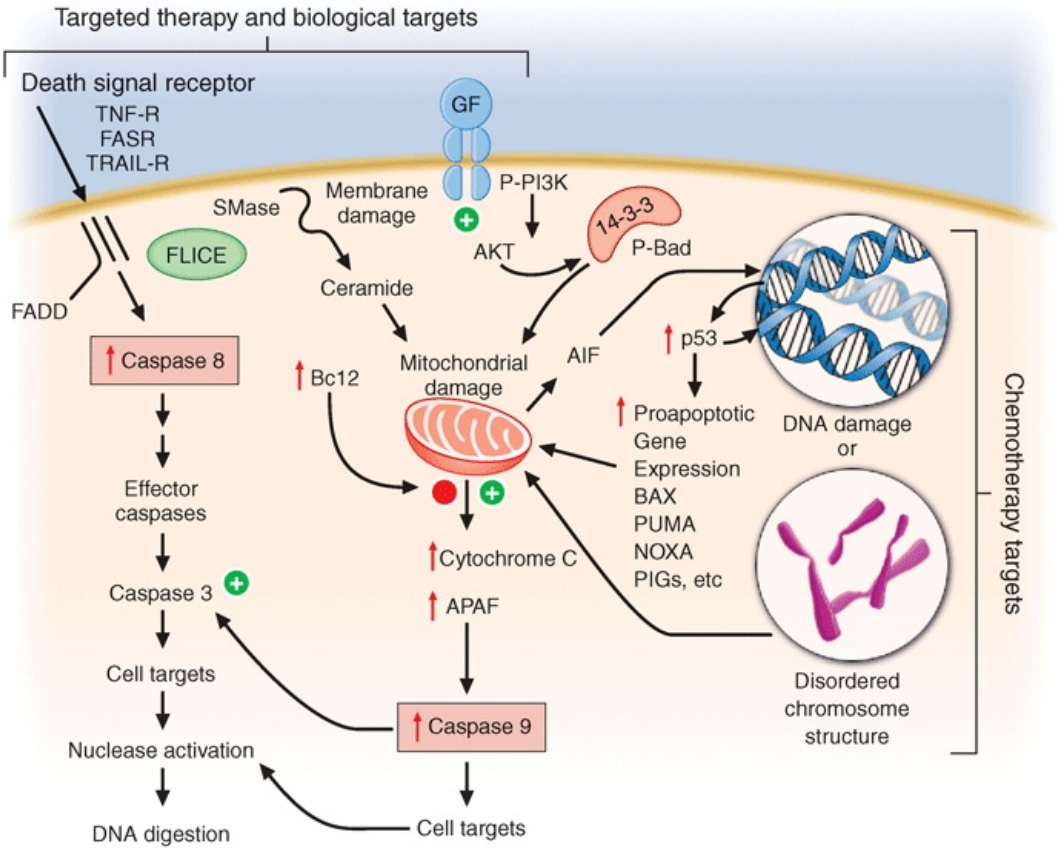

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Integration of cell death responses. Cell death through an apoptotic mechanism requires active participation of the cell. In response to interruption of growth factor (GF) or propagation of certain cytokine death signals (e.g., tumor necrosis factor receptor, TNF-R), there is activation of "upstream" cysteine aspartyl proteases (caspases), which then directly digest cytoplasmic and nuclear proteins, resulting in activation of "downstream" caspases; these cause activation of nucleases, resulting in the characteristic DNA fragmentation that is a hallmark of apoptosis. Chemotherapy agents that create lesions in DNA or alter mitotic spindle function seem to activate aspects of this process by damage ultimately conveyed to the mitochondria, perhaps by activating the transcription of genes whose products can produce or modulate the toxicity of free radicals. In addition, membrane damage with activation of sphingomyelinases results in the production of ceramides that can have a direct action at mitochondria. The antiapoptotic protein bcl2 attenuates mitochondrial toxicity, while proapoptotic gene products such as bax antagonize the action of bcl2. Damaged mitochondria release cytochrome C and apoptosis-activating factor (APAF), which can directly activate caspase 9, resulting in propagation of a direct signal to other downstream caspases through protease activation. Apoptosis-inducing factor (AIF) is also released from the mitochondrion and then can translocate to the nucleus, bind to DNA, and generate free radicals to further damage DNA. An additional proapoptotic stimulus is the bad protein, which can heterodimerize with *bcl2* gene family members to antagonize apoptosis. Importantly, though, bad protein function can be retarded by its sequestration as phospho-bad through the 14-3-3 adapter proteins. The phosphorylation of bad is mediated by the action of the AKT kinase in a way that defines how growth factors that activate this kinase can retard apoptosis and promote cell survival.

Figure 85-4

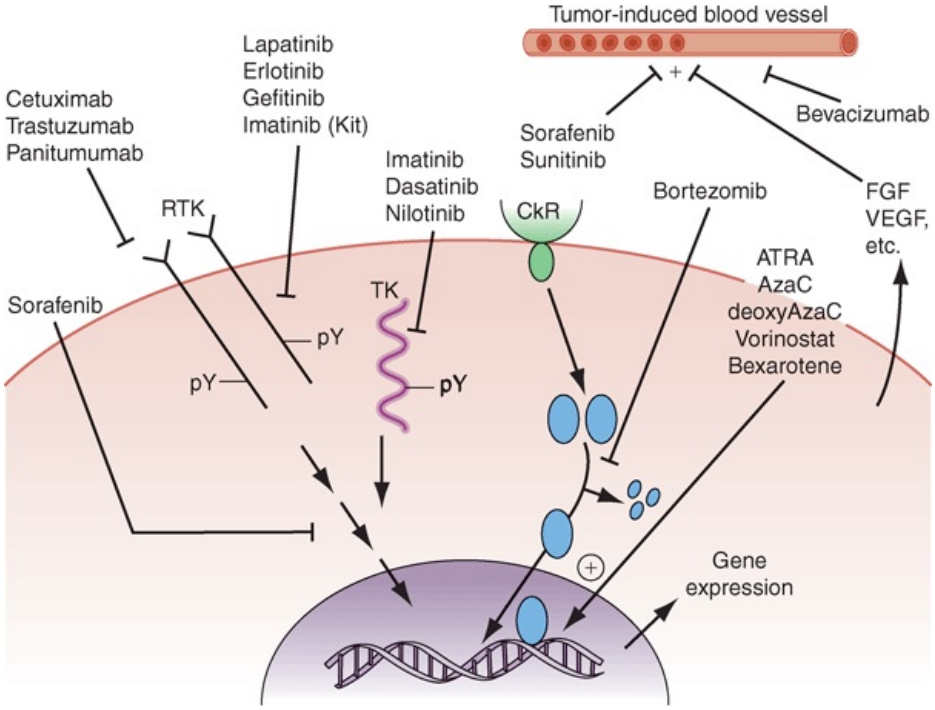

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Site of action of targeted agents. Signals proceeding from growth factor–related receptor tyrosine kinases (RTKs) such as EGF-R, erbB2, or c-kit can be interrupted by lapatinib, erlotinib, gefitinib, and imatinib, acting at the ATP binding site; or by cetuximab, trastuzumab, or panitumumab acting at the receptor. Tyrosine kinases (TKs) that are not directly stimulated by growth factors such as p210 bcr-abl or src can be inhibited by imatinib, dasatinib, or nilotinib. Signals projected downstream from growth factor receptors can be affected by the multitargeted kinase inhibitor sorafenib, acting on c-raf, and, upon arrival at the nucleus, affect gene expression, which can be affected by the targeted transcriptional modulators vorinostat (targeting histone deacetylase), azacytidine derivatives (targeting DNA methyltransferase), or retinoid receptor modulators all-*trans*-retinoic acid (ATRA) or bexarotene. Cytokine receptors (CkRs) are one stimulus for degradation of the inhibitory subunit of the $NF_{\kappa}B$ transcription factor by the proteosome. Bortezomib inhibits this process and can prevent activation of $NF_{\kappa}B$-dependent genes, among other growth-related effects. Sorafenib and sunitinib, acting as inhibitors of vascular endothelial growth factor (VEGF) receptors, can modulate tumor blood vessel function through their action on endothelial cells, while bevacizumab targets the same process by combining with VEGF itself.

Targeted agents differ from chemotherapy agents in that they do not indiscriminately cause macromolecular lesions but regulate the action of particular pathways. For example, the $p210^{bcr\text{-}abl}$ fusion protein tyrosine kinase drives chronic myeloid leukemia (CML), and HER-2/neu stimulates the proliferation of certain breast cancers. The tumor has been described as "addicted" to the function of these molecules in the sense that without the pathway's continued action, the tumor cell cannot survive. In this way, targeted agents may alter the "threshold" tumors have for undergoing apoptosis without actually creating any molecular lesions such as direct DNA strand breakage or altered membrane function.

While apoptotic mechanisms are important in regulating cellular proliferation and the behavior of tumor cells in vitro, in vivo it is unclear whether all of the actions of chemotherapeutic agents to cause cell death can be attributed to apoptotic mechanisms. However, changes in molecules that regulate apoptosis are correlated with clinical outcomes (e.g., *bcl2* overexpression in certain lymphomas conveys poor prognosis; proapoptotic bax expression is associated with a better outcome after chemotherapy for ovarian carcinoma). A better understanding of the relationship of cell death and cell survival mechanisms is needed.

Resistance to chemotherapy drugs has been postulated to arise either from cells not being in the appropriate phase of the cell cycle to allow drug lethality, or from decreased uptake, increased efflux, metabolism of the drug, or alteration of the target, e.g., by mutation or overexpression. Indeed, p170PGP (p170 P-glycoprotein; *mdr* gene product) was recognized from experiments with cells growing in tissue culture as mediating the efflux of chemotherapeutic agents in resistant cells. Certain neoplasms, particularly hematopoietic tumors, have an adverse prognosis if they express high levels of p170PGP, and modulation of this protein's function has been attempted by a variety of strategies.

"Combination chemotherapy" refers to the use of regimens where different drugs are combined with the goal of achieving at least an additive and hopefully supra-additive effect. The component drugs in such regimens ideally have distinct, nonoverlapping toxicities to the host, are each individually active

to some degree, and have been shown in a clinical trial to be tolerable and convey clinical value in contrast to the use of single agents.

Chemotherapeutic Agents Used for Cancer Treatment

Table 85–2 lists commonly used cancer chemotherapy agents and pertinent clinical aspects of their use. The drugs and schedules listed are examples that have proved tolerable and useful; the specific doses that may be used in a particular patient may vary somewhat with the particular protocol, or plan, of treatment. Significant variation from these dose ranges should be carefully verified to avoid or anticipate toxicity. Not included in Table 85–2 are hormone receptor–directed agents, as the side effects are generally those expected from the interruption or augmentation of hormonal effect, and doses used in most cases are those that adequately saturate the intended hormone receptor. The drugs listed may be usefully grouped into three general categories: those affecting DNA, those affecting microtubules, and molecularly targeted agents.

Table 85–2 Commonly Used Cancer Chemotherapy Agents

Drug	Examples of Usual Doses	Toxicity	Interactions, Issues
Direct DNA-Interacting Agents			
Alkylators			
Cyclophosphamide	400–2000 mg/m^2 IV 100 mg/m^2 PO qd	Marrow (relative platelet sparing) Cystitis Common alkylator[a] Cardiac (high dose)	Liver metabolism required to activate to phosphoramide mustard + acrolein Mesna protects against "high-dose" bladder damage
Mechlorethamine	6 mg/m^2 IV day 1 and day 8	Marrow Vesicant Nausea	Topical use in cutaneous lymphoma
Chlorambucil	1–3 mg/m^2 qd PO	Marrow Common alkylator[a]	
Melphalan	8 mg/m^2 qd x 5, PO	Marrow (delayed nadir) GI (high dose)	Decreased renal function delays clearance
Carmustine (BCNU)	200 mg/m^2 IV 150 mg/m^2 PO	Marrow (delayed nadir) GI, liver (high dose) Renal	
Lomustine (CCNU)	100–300 mg/m^2 PO	Marrow (delayed nadir)	
Ifosfamide	1.2 g/m^2 per day qd x 5 + mesna	Myelosuppressive Bladder Neurologic Metabolic acidosis Neuropathy	Isomeric analogue of cyclophosphamide More lipid soluble Greater activity vs testicular neoplasms and sarcomas Must use mesna
Procarbazine	100 mg/m^2 per day qd x 14	Marrow Nausea Neurologic Common alkylator[a]	Liver and tissue metabolism required Disulfiram-like effect with ethanol Acts as MAOI HBP after tyrosinase-rich foods
Dacarbazine (DTIC)	375 mg/m^2 IV day 1 and day 15	Marrow Nausea	Metabolic activation

		Flulike	
Temozolomide	150–200 mg/m^2 qd x 5 q28d *or* 75 mg/m^2 qd x 6–7 weeks	Nausea/vomiting Headache/fatigue Constipation	Infrequent myelosuppression
Altretamine (formerly hexamethylmelamine)	260 mg/m^2 per day qd x 14–21 as 4 divided oral doses	Nausea Neurologic (mood swing) Neuropathy Marrow (less)	Liver activation Barbiturates enhance/cimetidine diminishes
Cisplatin	20 mg/m^2 qd x 5 IV 1 q3–4 weeks or 100–200 mg/m^2 per dose IV q3–4 weeks	Nausea Neuropathy Auditory Marrow platelets > WBCs Renal Mg^{2+}, Ca^{2+}	Maintain high urine flow; osmotic diuresis, monitor intake/output K^+, Mg^{2+} Emetogenic—prophylaxis needed Full dose if CrCl > 60 mL/min and tolerate fluid push
Carboplatin	365 mg/m^2 IV q3–4 weeks as adjusted for CrCl	Marrow platelets > WBCs Nausea Renal (high dose)	Reduce dose according to CrCl: to AUC of 5–7 mg/mL per min [AUC = dose/(CrCl + 25)]
Oxaliplatin	130 mg/m^2 q3 weeks over 2 h *or* 85 mg/m^2 q2 weeks	Nausea Anemia	Acute reversible neurotoxicity; chronic sensory neurotoxicity cumulative with dose; reversible laryngopharyngeal spasm
Antitumor Antibiotics and Topoisomerase Poisons			
Bleomycin	15–25 mg/d qd x 5 IV bolus *or* continuous IV	Pulmonary Skin effects Raynaud's Hypersensitivity	Inactivate by bleomycin hydrolase (decreased in lung/skin) O_2 enhances pulmonary toxicity Cisplatin-induced decrease in CrCl may increase skin/lung toxicity Reduce dose if CrCl < 60 mL/min
Actinomycin D	10–15 μg/kg per day qd x 5 IV bolus	Marrow Nausea Mucositis Vesicant Alopecia	Radiation recall
Etoposide (VP16-213)	100–150 mg/m^2 IV qd x 3–5d *or* 50 mg/m^2 PO qd x 21d *or* up to 1500 mg/m^2 per dose (high dose with stem cell support)	Marrow (WBCs > platelet) Alopecia Hypotension Hypersensitivity (rapid IV) Nausea Mucositis (high dose)	Hepatic metabolism—renal 30% Reduce doses with renal failure Schedule-dependent (5 day better than 1 day) Late leukemogenic Accentuate antimetabolite action
Topotecan	20 mg/m^2 IV q3–4 weeks over 30 min *or* 1.5–3	Marrow Mucositis	Reduce dose with renal failure

	mg/m^2 q3–4 weeks over 24 h *or* 0.5 mg/m^2 per day over 21 days	Nausea Mild alopecia	No liver toxicity
Irinotecan (CPT II)	100–150 mg/m^2 IV over 90 min q3–4 weeks *or* 30 mg/m^2 per day over 120 h	Diarrhea: "early onset" with cramping, flushing, vomiting; "late onset" after several doses Marrow Alopecia Nausea Vomiting Pulmonary	Prodrug requires enzymatic clearance to active drug "SN 38" Early diarrhea likely due to biliary excretion Late diarrhea, use "high-dose" loperamide (2 mg q2–4 h)
Doxorubicin and daunorubicin	45–60 mg/m^2 dose q3–4 weeks *or* 10–30 mg/m^2 dose q week *or* continuous-infusion regimen	Marrow Mucositis Alopecia Cardiovascular acute/chronic Vesicant	Heparin aggregate; coadministration increases clearance Acetaminophen, BCNU increase liver toxicity Radiation recall
Idarubicin	10–15 mg/m^2 IV q 3 weeks *or* 10 mg/m^2 IV qd x 3	Marrow Cardiac (less than doxorubicin)	None established
Epirubicin	150 mg/m^2 IV q3 weeks	Marrow Cardiac	None established
Mitoxantrone	12 mg/m^2 qd x 3 *or* 12–14 mg/m^2 q3 weeks	Marrow Cardiac (less than doxorubicin) Vesicant (mild) Blue urine, sclerae, nails	Interacts with heparin Less alopecia, nausea than doxorubicin Radiation recall
Indirect DNA-Interacting Agents			
Antimetabolites			
Deoxycoformycin	4 mg/m^2 IV every other week	Nausea Immunosuppression Neurologic Renal	Excretes in urine Reduce dose for renal failure Inhibits adenosine deaminase
6-Mercaptopurine	75 mg/m^2 PO *or* up to 500 mg/m^2 PO (high dose)	Marrow Liver Nausea	Variable bioavailability Metabolize by xanthine oxidase Decrease dose with allopurinol Increased toxicity with thiopurine methyltransferase deficiency
6-Thioguanine	2–3 mg/kg per day for up to 3–4 weeks	Marrow Liver Nausea	Variable bioavailability Increased toxicity with thiopurine methyltransferase deficiency
Azathioprine	1–5 mg/kg per day	Marrow Nausea	Metabolizes to 6MP, therefore reduce dose

		Nausea Liver	with allopurinol Increased toxicity with thiopurine methyltransferase deficiency
2-Chlorodeoxyadenosine	0.09 mg/kg per day qd x 7 as continuous infusion	Marrow Renal Fever	Notable use in hairy cell leukemia
Hydroxyurea	20–50 mg/kg (lean body weight) PO qd *or* 1–3 g/d	Marrow Nausea Mucositis Skin changes Rare renal, liver, lung, CNS	Decrease dose with renal failure Augments antimetabolite effect
Methotrexate	15–30 mg PO or IM qd x 3–5 *or* 30 mg IV days 1 and 8 *or* 1.5–12g/m^2 per day (with leucovorin)	Marrow Liver/lung Renal tubular Mucositis	Rescue with leucovorin Excreted in urine Decrease dose in renal failure NSAIDs increase renal toxicity
5-Fluorouracil (5FU)	375 mg/m^2 IV qd x 5 *or* 600 mg/m^2 IV days 1 and 8	Marrow Mucositis Neurologic Skin changes	Toxicity enhanced by leucovorin Dihydropyrimidine dehydrogenase deficiency increases toxicity Metabolizes in tissues
Capecitabine	665 mg/m^2 bid continuous; 1250 mg/m^2 bid 2 weeks on / 1 off; 829 mg/m^2 bid 2 weeks on / 1 off + 60 mg/d leucovorin	Diarrhea Hand-foot syndrome	Prodrug of 5FU due to intratumoral metabolism
Cytosine arabinoside	100 mg/m^2 per day qd x 7 by continuous infusion *or* 1–3 g/m^2 dose IV bolus	Marrow Mucositis Neurologic (high dose) Conjunctivitis (high dose) Noncardiogenic pulmonary edema	Enhances activity of alkylating agents Metabolizes in tissues by deamination
Azacytidine	750 mg/m^2 per week or 75–200 mg/m^2 per day x 5–10 (bolus) or (continuous IV or subcutaneous)	Marrow Nausea Liver Neurologic Myalgia	Use limited to leukemia Altered methylation of DNA alters gene expression
Gemcitabine	1000 mg/m^2 IV weekly x 7	Marrow Nausea Hepatic Fever/"flu syndrome"	
Fludarabine phosphate	25 mg/m^2 IV qd x 5	Marrow Neurologic Lung	Dose reduction with renal failure Metabolized to F-ara converted to F-ara ATP in cells by deoxycytidine kinase

Asparaginase	25,000 IU/m^2 q3–4 weeks *or* 6000 IU/m^2 per day qod for 3–4 weeks *or* 1000–2000 IU/m^2 for 10–20 days	Protein synthesis Clotting factors Glucose Albumin Hypersensitivity CNS Pancreatitis Hepatic	Blocks methotrexate action
Pemetrexed	200 mg/m^2 q3 weeks	Anemia Neutropenia Thrombocytopenia	Supplement folate/B_{12} Caution in renal failure
Antimitotic Agents			
Vincristine	1–1.4 mg/m^2 per week (frequently cap at 2 mg total dose)	Vesicant Marrow Neurologic GI: ileus/constipation; bladder hypotoxicity; SIADH Cardiovascular	Hepatic clearance Dose reduction for bilirubin >1.5 mg/dL Prophylactic bowel regimen
Vinblastine	6–8 mg/m^2 per week	Vesicant Marrow Neurologic (less common but similar spectrum to other vincas) Hypertension Raynaud's	Hepatic clearance Dose reduction as with vincristine
Vinorelbine	15–30 mg/m^2 per week	Vesicant Marrow Allergic/bronchospasm (immediate) Dyspnea/cough (subacute) Neurologic (less prominent but similar spectrum to other vincas)	Hepatic clearance
Paclitaxel	135–175 mg/m^2 per 24-h infusion *or* 175 mg/m^2 per 3-h infusion *or* 140 mg/m^2 per 96-h infusion *or* 250 mg/m^2 per 24-h infusion plus G-CSF	Hypersensitivity Marrow Mucositis Alopecia Sensory neuropathy CV conduction disturbance Nausea—infrequent	Premedicate with steroids, H_1 and H_2 blockers Hepatic clearance Dose reduction as with vincas
Docetaxel	100 mg/m^2 per 1-h infusion q3 weeks	Hypersensitivity Fluid retention syndrome Marrow Dermatologic Sensory neuropathy Nausea infrequent Some stomatitis	Premedicate with steroids, H_1 and H_2 blockers

Estramustine phosphate	14 mg/kg per day in 3–4 divided doses with water >2 h after meals Avoid Ca^{2+}-rich foods	Nausea Vomiting Diarrhea CHF Thrombosis Gynecomastia	
Nab-paclitaxel (protein bound)	260 mg/m^2 q3 weeks	Neuropathy Anemia Neutropenia Thrombocytopenia	Caution in hepatic insufficiency
Ixabepilone	40 mg/m^2 q3 weeks	Myelosuppression Neuropathy	
Molecularly Targeted Agents			
Retinoids			
Tretinoin	45 mg/m^2 per day until complete response + anthracycline-based regimen in APL	Teratogenic Cutaneous	APL differentiation syndrome: pulmonary dysfunction/infiltrate, pleural/pericardial effusion, fever
Bexarotene	300–400 mg/m^2 per day, continuous	Hypercholesterolemia Hypertriglyceridemia Cutaneous Teratogenic	Central hypothyroidism
Targeted Toxins			
Denileukin diftitox	9–18 μg/kg per day x 5 d q3 weeks	Nausea/vomiting Chills/fever Asthenia Hepatic	Acute hypersensitivity: hypotension, vasodilation, rash, chest tightness Vascular leak: hypotension, edema, hypoalbuminemia, thrombotic events (MI, DVT, CVA)
Tyrosine Kinase Inhibitors			
Imatinib	400 mg/d, continuous	Nausea Periorbital edema	Myelosuppression not frequent in solid tumor indications
Gefitinib	250 mg PO per day	Rash Diarrhea	In U.S., only with prior documented benefit
Erlotinib	150 mg PO per day	Rash Diarrhea	1 h before, 2 h after meals
Dasatinib	70 mg PO bid; 100 mg PO per day	Liver changes Rash Neutropenia Thrombocytopenia	
Sorafenib	400 mg PO bid	Diarrhea Hand-foot syndrome Other rash	
Sunitinib	50 mg PO qd for 4 of 6 weeks	Fatigue Diarrhea Neutropenia	

Proteosome Inhibitors			
Bortezomib	1.3 mg/m^2 day 1,4	Neuropathy Thrombocytopenia	
Histone Deacetylase Inhibitors			
Vorinostat	400 mg/day	Fatigue Diarrhea Thrombocytopenia Embolism	
Romidepsin	14 mg/m^2 day 1, 8, 15	Nausea Vomiting Cytopenias Cardiac conduction	
mTOR Inhibitors			
Temsirolimus	25 mg weekly	Stomatitis Thrombocytopenia Nausea Anorexia, fatigue Metabolic (glucose, lipid)	
Everolimus	10 mg daily	Stomatitis Fatigue	
Miscellaneous			
Arsenic trioxide	0.16 mg/kg per day up to 50 days in APL	↑ QT_C Peripheral neuropathy Musculoskeletal pain Hyperglycemia	APL differentiation syndrome (see under tretinoin)

[a]Common alkylator: alopecia, pulmonary, infertility, plus teratogenesis.

Abbreviations: APL, acute promyelocytic leukemia; AUC, area under the curve; CHF, congestive heart failure; CNS, central nervous system; CrCl, creatinine clearance; CV, cardiovascular; CVA, cerebrovascular accident; DVT, deep venous thrombosis; G-CSF, granulocyte colony-stimulating factor; GI, gastrointestinal; HBP, high blood pressure; MAOI, monoamine oxidase inhibitors; MI, myocardial infarction; 6MP, 6-mercaptopurine; mTOR, mammalian target of rapamycin; NSAIDs, nonsteroidal anti-inflammatory drugs; SIADH, syndrome of inappropriate antidiuretic hormone; WBCs, white blood cells.

DIRECT DNA-INTERACTIVE AGENTS

DNA replication occurs during the synthesis or S-phase of the cell cycle, with chromosome segregation of the replicated DNA occurring in the M, or mitosis, phase. The G_1 and G_2 "gap phases" precede S and M, respectively. Historically, chemotherapeutic agents have been divided into "phase-nonspecific" agents, which can act in any phase of the cell cycle, and "phase-specific" agents, which require the cell to be at a particular cell cycle phase to cause greatest effect. Once the agent has acted, cells may progress to "checkpoints" in the cell cycle where the drug-related damage may be assessed and either repaired or allowed to initiate apoptosis. An important function of certain tumor-suppressor genes such as p53 may be to modulate checkpoint function.

Formation of Covalent DNA Adducts

Alkylating agents as a class are cell cycle phase–nonspecific agents. They break down, either spontaneously or after normal organ or tumor cell metabolism, to reactive intermediates that covalently modify bases in DNA. This leads to cross-linkage of DNA strands or the appearance of breaks in DNA as a result of repair efforts. "Broken" or cross-linked DNA is intrinsically unable to complete normal replication or cell division; in addition, it is a potent activator of cell cycle checkpoints and further activates cell-signaling pathways that can precipitate apoptosis. As a class, alkylating agents share similar toxicities: myelosuppression, alopecia, gonadal dysfunction, mucositis, and pulmonary fibrosis. They differ greatly in a spectrum of normal organ toxicities. As a class they share the capacity to cause "second" neoplasms, particularly leukemia, many years after use, particularly when used in low doses for protracted periods.

Cyclophosphamide is inactive unless metabolized by the liver to 4-hydroxy-cyclophosphamide, which decomposes into an alkylating species, as well as to chloroacetaldehyde and acrolein. The latter causes chemical cystitis; therefore, excellent hydration must be maintained while using cyclophosphamide. If severe, the cystitis may be effectively treated by mesna (2-mercaptoethanesulfonate). Liver disease impairs drug activation. Sporadic interstitial pneumonitis leading to pulmonary fibrosis can accompany the use of cyclophosphamide, and high doses used in conditioning regimens for bone marrow transplant can cause cardiac dysfunction. Ifosfamide is a cyclophosphamide analogue also activated in the liver, but more slowly, and it requires coadministration of mesna to prevent bladder injury. Central nervous system (CNS) effects, including somnolence, confusion, and psychosis, can follow ifosfamide use; the incidence appears related to low body surface area or decreased creatinine clearance.

Several alkylating agents are less commonly used. Nitrogen mustard (mechlorethamine) is the prototypic agent of this class, decomposing rapidly in aqueous solution to potentially yield a bifunctional carbonium ion. It must be administered shortly after preparation into a rapidly flowing intravenous line. It is a powerful vesicant, and infiltration may be symptomatically ameliorated by infiltration of the affected site with 1/6 *M* thiosulfate. Even without infiltration, aseptic thrombophlebitis is frequent. It can be used topically as a dilute solution in cutaneous lymphomas, with a notable incidence of hypersensitivity reactions. It causes moderate nausea after intravenous administration. Bendamustine is a nitrogen mustard derivative with evidence of activity in chronic lymphocytic leukemia and certain lymphomas.

Chlorambucil causes predictable myelosuppression, azoospermia, nausea, and pulmonary side effects. Busulfan can cause profound myelosuppression, alopecia, and pulmonary toxicity but is relatively "lymphocyte sparing." Its routine use in treatment of CML has been curtailed in favor of imatinib (Gleevec) or dasatinib, but it is still employed in transplant preparation regimens. Melphalan shows variable oral bioavailability and undergoes extensive binding to albumin and α_1-acidic glycoprotein. Mucositis appears more prominently; however, it has prominent activity in multiple myeloma.

Nitrosoureas break down to carbamylating species that not only cause a distinct pattern of DNA base pair–directed toxicity but also can covalently modify proteins. They share the feature of causing relatively delayed bone marrow toxicity, which can be cumulative and long-lasting. Methyl CCNU (lomustine) causes direct glomerular as well as tubular damage, cumulatively related to dose and time of exposure.

Procarbazine is metabolized in the liver and possibly in tumor cells to yield a variety of free radical and alkylating species. In addition to myelosuppression, it causes hypnotic and other CNS effects, including vivid nightmares. It can cause a disulfiram-like syndrome on ingestion of ethanol. Altretamine (formerly hexamethylmelamine) and thiotepa can chemically give rise to alkylating species, although the nature of the DNA damage has not been well characterized in either case. Dacarbazine (DTIC) is activated in the liver to yield the highly reactive methyl diazonium cation. It causes only modest myelosuppression 21–25 days after a dose but causes prominent nausea on day 1. Temozolomide is structurally related to dacarbazine but was designed to be activated by nonenzymatic hydrolysis in tumors and is bioavailable orally.

Cisplatin was discovered fortuitously by observing that bacteria present in electrolysis solutions could not divide. Only the *cis* diamine configuration is active as an antitumor agent. It is hypothesized that in the intracellular environment, a chloride is lost from each position, being replaced by a water molecule. The resulting positively charged species is an efficient bifunctional interactor with DNA, forming Pt-based cross-links. Cisplatin requires administration with adequate hydration, including forced diuresis with mannitol to prevent kidney damage; even with the use of hydration, gradual decrease in kidney function is common, along with noteworthy anemia. Hypomagnesemia frequently attends cisplatin use and can lead to hypocalcemia and tetany. Other common toxicities include neurotoxocity with stocking-and-glove sensorimotor neuropathy. Hearing loss occurs in 50% of patients treated with conventional doses. Cisplatin is intensely emetogenic, requiring prophylactic antiemetics. Myelosuppression is less evident than with other alkylating agents. Chronic vascular toxicity (Raynaud's phenomenon, coronary artery disease) is a more unusual toxicity. Carboplatin displays less nephro-, oto-, and neurotoxicity. However, myelosuppression is more frequent, and as the drug is exclusively cleared through the kidney, adjustment of dose for creatinine clearance must be accomplished through use of various dosing nomograms. Oxaliplatin is a platinum analogue with noteworthy activity in colon cancers refractory to other treatments. It is prominently neurotoxic.

Antitumor Antibiotics and Topoisomerase Poisons

Antitumor antibiotics are substances produced by bacteria that in nature appear to provide a chemical defense against other hostile microorganisms. As a class they bind to DNA directly and can frequently undergo electron transfer reactions to generate free radicals in close proximity to DNA, leading to DNA damage in the form of single-strand breaks or cross-links. Topoisomerase poisons include natural products or semisynthetic species derived ultimately from plants, and they modify enzymes that

regulate the capacity of DNA to unwind to allow normal replication or transcription. These include topoisomerase I, which creates single-strand breaks that then rejoin following the passage of the other DNA strand through the break. Topoisomerase II creates double-strand breaks through which another segment of DNA duplex passes before rejoining. DNA damage from these agents can occur in any cell cycle phase, but cells tend to arrest in S-phase or G_2 of the cell cycle in cells with p53 and Rb pathway lesions as the result of defective checkpoint mechanisms in cancer cells. Owing to the role of topoisomerase I in the procession of the replication fork, topoisomerase I poisons cause lethality if the topoisomerase I–induced lesions are made in S-phase.

Doxorubicin can intercalate into DNA, thereby altering DNA structure, replication, and topoisomerase II function. It can also undergo reduction reactions by accepting electrons into its quinone ring system, with the capacity to undergo reoxidation to form reactive oxygen radicals after reoxidation. It causes predictable myelosuppression, alopecia, nausea, and mucositis. In addition, it causes acute cardiotoxicity in the form of atrial and ventricular dysrhythmias, but these are rarely of clinical significance. In contrast, cumulative doses >550 mg/m^2 are associated with a 10% incidence of chronic cardiomyopathy. The incidence of cardiomyopathy appears to be related to schedule (peak serum concentration), with low-dose, frequent treatment or continuous infusions better tolerated than intermittent higher-dose exposures. Cardiotoxicity has been related to iron-catalyzed oxidation and reduction of doxorubicin, and not to topoisomerase action. Cardiotoxicity is related to peak plasma dose; thus, lower doses and continuous infusions are less likely to cause heart damage. Doxorubicin's cardiotoxicity is increased when given together with trastuzumab (Herceptin), the anti-HER2/neu antibody. Radiation recall or interaction with concomitantly administered radiation to cause local site complications is frequent. The drug is a powerful vesicant, with necrosis of tissue apparent 4–7 days after an extravasation; therefore, it should be administered into a rapidly flowing intravenous line. Dexrazoxane is an antidote to doxorubicin-induced extravasation. Doxorubicin is metabolized by the liver, so doses must be reduced by 50–75% in the presence of liver dysfunction. Daunorubicin is closely related to doxorubicin and was actually introduced first into leukemia treatment, where it remains part of curative regimens and has been shown preferable to doxorubicin owing to less mucositis and colonic damage. Idarubicin is also used in acute myeloid leukemia treatment and may be preferable to daunorubicin in activity. Encapsulation of daunorubicin into a liposomal formulation has attenuated cardiac toxicity and antitumor activity in Kaposi's sarcoma and ovarian cancer.

Bleomycin refers to a mixture of glycopeptides that have the unique feature of forming complexes with Fe^{2+} while also bound to DNA. It remains an important component of curative regimens for Hodgkin's disease and germ cell neoplasms. Oxidation of Fe^{2+} gives rise to superoxide and hydroxyl radicals. The drug causes little, if any, myelosuppression. The drug is cleared rapidly, but augmented skin and pulmonary toxicity in the presence of renal failure has led to the recommendation that doses be reduced by 50–75% in the face of a creatinine clearance <25 mL/min. Bleomycin is not a vesicant and can be administered intravenously, intramuscularly, or subcutaneously. Common side effects include fever and chills, facial flush, and Raynaud's phenomenon. Hypertension can follow rapid intravenous administration, and the incidence of anaphylaxis with early preparations of the drug has led to the practice of administering a test dose of 0.5–1 unit before the rest of the dose. The most feared complication of bleomycin treatment is pulmonary fibrosis, which increases in incidence at >300 cumulative units administered and is minimally responsive to treatment (e.g., glucocorticoids). The earliest indicator of an adverse effect is a decline in the DL_{CO}, although cessation of drug immediately upon documentation of a decrease in DL_{CO} may not prevent further decline in pulmonary function. Bleomycin is inactivated by a bleomycin hydrolase, whose concentration is diminished in skin and lung. Because bleomycin-dependent electron transport is dependent on O_2, bleomycin toxicity may become apparent after exposure to transient very high Pi_{O2}. Thus, during surgical procedures, patients with prior exposure to bleomycin should be maintained on the lowest Pi_{O2} consistent with maintaining adequate tissue oxygenation.

Mitoxantrone is a synthetic compound that was designed to recapitulate features of doxorubicin but with less cardiotoxicity. It is quantitatively less cardiotoxic (comparing the ratio of cardiotoxic to therapeutically effective doses) but is still associated with a 10% incidence of cardiotoxicity at cumulative doses of >150 mg/m^2. It also causes alopecia. Cases of acute promyelocytic leukemia (APL) have arisen shortly after exposure of patients to mitoxantrone, particularly in the adjuvant treatment of breast cancer. While chemotherapy-associated leukemia is generally of the acute myeloid type, APL arising in the setting of prior mitoxantrone treatment had the typical t(15;17) chromosome translocation associated with APL, but the breakpoints of the translocation appeared to be at topoisomerase II sites that would be preferred sites of mitoxantrone action, clearly linking the action of the drug to the generation of the leukemia.

Etoposide was synthetically derived from the plant product podophyllotoxin; it binds directly to topoisomerase II and DNA in a reversible ternary complex. It stabilizes the covalent intermediate in the enzyme's action where the enzyme is covalently linked to DNA. This "alkali-labile" DNA bond was historically a first hint that an enzyme such as a topoisomerase might exist. The drug therefore causes

a prominent G_2 arrest, reflecting the action of a DNA damage checkpoint. Prominent clinical effects include myelosuppression, nausea, and transient hypotension related to the speed of administration of the agent. Etoposide is a mild vesicant but is relatively free from other large-organ toxicities. When given at high doses or very frequently, topoisomerase II inhibitors may cause acute leukemia associated with chromosome 11q23 abnormalities in up to 1% of exposed patients.

Camptothecin was isolated from extracts of a Chinese tree and had notable antileukemia activity in preclinical mouse models. Early human clinical studies with the sodium salt of the hydrolyzed camptothecin lactone showed evidence of toxicity with little antitumor activity. Identification of topoisomerase I as the target of camptothecins and the need to preserve lactone structure allowed additional efforts to identify active members of this series. Topoisomerase I is responsible for unwinding the DNA strand by introducing single-strand breaks and allowing rotation of one strand about the other. In S-phase, topoisomerase I–induced breaks that are not promptly resealed lead to progress of the replication fork off the end of a DNA strand. The DNA damage is a potent signal for induction of apoptosis. Camptothecins promote the stabilization of the DNA linked to the enzyme in a so-called cleavable complex, analogous to the action of etoposide with topoisomerase II. Topotecan is a camptothecin derivative approved for use in gynecologic tumors and small cell lung cancer. Toxicity is limited to myelosuppression and mucositis. CPT-11, or irinotecan, is a camptothecin with evidence of activity in colon carcinoma. In addition to myelosuppression, it causes a secretory diarrhea related to the toxicity of a metabolite called SN-38. The diarrhea can be treated effectively with loperamide or octreotide.

INDIRECT EFFECTORS OF DNA FUNCTION: ANTIMETABOLITES

A broad definition of antimetabolites would include compounds with structural similarity to precursors of purines or pyrimidines, or compounds that interfere with purine or pyrimidine synthesis. Antimetabolites can cause DNA damage indirectly, through misincorporation into DNA, abnormal timing or progression through DNA synthesis, or altered function of pyrimidine and purine biosynthetic enzymes. They tend to convey greatest toxicity to cells in S-phase, and the degree of toxicity increases with duration of exposure. Common toxic manifestations include stomatitis, diarrhea, and myelosuppression. Second malignancies are not associated with their use.

Methotrexate inhibits dihydrofolate reductase, which regenerates reduced folates from the oxidized folates produced when thymidine monophosphate is formed from deoxyuridine monophosphate. Without reduced folates, cells die a "thymine-less" death. N5-tetrahydrofolate or N5-formyltetrahydrofolate (leucovorin) can bypass this block and rescue cells from methotrexate, which is maintained in cells by polyglutamylation. The drug and other reduced folates are transported into cells by the folate carrier, and high concentrations of drug can bypass this carrier and allow diffusion of drug directly into cells. These properties have suggested the design of "high-dose" methotrexate regimens with leucovorin rescue of normal marrow and mucosa as part of curative approaches to osteosarcoma in the adjuvant setting and hematopoietic neoplasms of children and adults. Methotrexate is cleared by the kidney via both glomerular filtration and tubular secretion, and toxicity is augmented by renal dysfunction and drugs such as salicylates, probenecid, and nonsteroidal anti-inflammatory agents that undergo tubular secretion. With normal renal function, 15 mg/m^2 leucovorin will rescue 10^{-8} to 10^{-6} *M* methotrexate in three to four doses. However, with decreased creatinine clearance, doses of 50–100 mg/m^2 are continued until methotrexate levels are <5 x 10^{-8} *M*. In addition to bone marrow suppression and mucosal irritation, methotrexate can cause renal failure itself at high doses owing to crystallization in renal tubules; therefore, high-dose regimens require alkalinization of urine with increased flow by hydration. Methotrexate can be sequestered in third-space collections and leach back into the general circulation, causing prolonged myelosuppression. Less-frequent adverse effects include reversible increases in transaminases and hypersensitivity-like pulmonary syndrome. Chronic low-dose methotrexate can cause hepatic fibrosis. When administered to the intrathecal space, methotrexate can cause chemical arachnoiditis and CNS dysfunction.

Pemetrexed is a novel folate-directed antimetabolite. It is "multitargeted" in that it inhibits the activity of several enzymes, including thymidylate synthetase, dihydrofolate reductase, and glycinamide ribonucleotide formyltransferase, thereby affecting the synthesis of both purine and pyrimidine nucleic acid precursors. To avoid significant toxicity to the normal tissues, patients receiving pemetrexed should also receive low-dose folate and vitamin B12 supplementation. Pemetrexed has notable activity against certain lung cancers and, in combination with cisplatin, also against mesotheliomas. Palatrexate is an antifolate approved for use in T cell lymphoma that is very efficiently transported into cancer cells.

5-Fluorouracil (5FU) represents an early example of "rational" drug design in that it originated from the observation that tumor cells incorporate radiolabeled uracil more efficiently into DNA than normal cells, especially gut. 5FU is metabolized in cells to 5'FdUMP, which inhibits thymidylate synthetase (TS). In addition, misincorporation can lead to single-strand breaks, and RNA can aberrantly incorporate FUMP. 5FU is metabolized by dihydropyrimidine dehydrogenase, and deficiency of this enzyme can lead to excessive toxicity from 5FU. Oral bioavailability varies unreliably, but orally administered

analogues of 5FU such as capecitabine have been developed that allow at least equivalent activity to many parenteral 5FU-based approaches. Intravenous administration of 5FU leads to bone marrow suppression after short infusions but to stomatitis after prolonged infusions. Leucovorin augments the activity of 5FU by promoting formation of the ternary covalent complex of 5FU, the reduced folate, and TS. Less-frequent toxicities include CNS dysfunction, with prominent cerebellar signs, and endothelial toxicity manifested by thrombosis, including pulmonary embolus and myocardial infarction.

Cytosine arabinoside (ara-C) is incorporated into DNA after formation of ara-CTP, resulting in S-phase–related toxicity. Continuous infusion schedules allow maximal efficiency, with uptake maximal at 5–7 μM. Ara-C can be administered intrathecally. Adverse effects include nausea, diarrhea, stomatitis, chemical conjunctivitis, and cerebellar ataxia. Gemcitabine is a cytosine derivative that is similar to ara-C in that it is incorporated into DNA after anabolism to the triphosphate, rendering DNA susceptible to breakage and repair synthesis, which differs from that in ara-C in that gemcitabine-induced lesions are very inefficiently removed. In contrast to ara-C, gemcitabine appears to have useful activity in a variety of solid tumors, with limited nonmyelosuppressive toxicities. 6-Thioguanine and 6-mercaptopurine (6MP) are used in the treatment of acute lymphoid leukemia. Although administered orally, they display variable bioavailability. 6MP is metabolized by xanthine oxidase and therefore requires dose reduction when used with allopurinol.

Fludarabine phosphate is a prodrug of F-adenine arabinoside (F-ara-A), which in turn was designed to diminish the susceptibility of ara-A to adenosine deaminase. F-ara-A is incorporated into DNA and can cause delayed cytotoxicity even in cells with low growth fraction, including chronic lymphocytic leukemia and follicular B cell lymphoma. CNS and peripheral nerve dysfunction and T cell depletion leading to opportunistic infections can occur in addition to myelosuppression. 2-Chlorodeoxyadenosine is a similar compound with activity in hairy cell leukemia. 2-Deoxycoformycin inhibits adenosine deaminase, with resulting increase in dATP levels. This causes inhibition of ribonucleotide reductase as well as augmented susceptibility to apoptosis, particularly in T cells. Renal failure and CNS dysfunction are notable toxicities in addition to immunosuppression. Hydroxyurea inhibits ribonucleotide reductase, resulting in S-phase block. It is orally bioavailable and useful for the acute management of myeloproliferative states.

Asparaginase is a bacterial enzyme that causes breakdown of extracellular asparagine required for protein synthesis in certain leukemic cells. This effectively stops tumor cell DNA synthesis, as DNA synthesis requires concurrent protein synthesis. The outcome of asparaginase action is therefore very similar to the result of the small-molecule antimetabolites. As asparaginase is a foreign protein, hypersensitivity reactions are common, as are effects on organs such as pancreas and liver that normally require continuing protein synthesis. This may result in decreased insulin secretion with hyperglycemia, with or without hyperamylasemia and clotting function abnormalities. Close monitoring of clotting functions should accompany use of asparaginase. Paradoxically, owing to depletion of rapidly turning over anticoagulant factors, thromboses particularly affecting the CNS may also be seen with asparaginase.

MITOTIC SPINDLE INHIBITORS

Microtubules are cellular structures that form the mitotic spindle, and in interphase cells they are responsible for the cellular "scaffolding" along which various motile and secretory processes occur. Microtubules are composed of repeating noncovalent multimers of a heterodimer of α and β isoform of the protein tubulin. Vincristine binds to the tubulin dimer with the result that microtubules are disaggregated. This results in the block of growing cells in M-phase; however, toxic effects in G_1 and S-phase are also evident, reflecting effects on normal cellular activities of microtubules. Vincristine is metabolized by the liver, and dose adjustment in the presence of hepatic dysfunction is required. It is a powerful vesicant, and infiltration can be treated by local heat and infiltration of hyaluronidase. At clinically used intravenous doses, neurotoxicity in the form of glove-and-stocking neuropathy is frequent. Acute neuropathic effects include jaw pain, paralytic ileus, urinary retention, and the syndrome of inappropriate antidiuretic hormone secretion. Myelosuppression is not seen. Vinblastine is similar to vincristine, except that it tends to be more myelotoxic, with more frequent thrombocytopenia and also mucositis and stomatitis. Vinorelbine is a vinca alkaloid that appears to have differences in resistance patterns in comparison to vincristine and vinblastine; it may be administered orally.

The taxanes include paclitaxel and docetaxel. These agents differ from the vinca alkaloids in that the taxanes stabilize microtubules against depolymerization. The "stabilized" microtubules function abnormally and are not able to undergo the normal dynamic changes of microtubule structure and function necessary for cell cycle completion. Taxanes are among the most broadly active antineoplastic agents for use in solid tumors, with evidence of activity in ovarian cancer, breast cancer, Kaposi's sarcoma, and lung tumors. They are administered intravenously, and paclitaxel requires use of a Cremophor-containing vehicle that can cause hypersensitivity reactions. Premedication with dexamethasone (8–16 mg orally or intravenously 12 and 6 h before treatment) and diphenhydramine (50 mg) and cimetidine (300 mg), both 30 min before treatment, decreases but does not eliminate the risk of hypersensitivity reactions to the paclitaxel vehicle. Docetaxel uses a polysorbate 80 formulation, which can cause fluid retention in addition to hypersensitivity reactions, and dexamethasone

premedication with or without antihistamines is frequently used. A protein-bound formulation of paclitaxel (called *nab-paclitaxel*) has at least equivalent antineoplastic activity and decreased risk of hypersensitivity reactions. Paclitaxel may also cause hypersensitivity reactions, myelosuppression, neurotoxicity in the form of glove-and-stocking numbness, and paresthesia. Cardiac rhythm disturbances were observed in phase I and II trials, most commonly asymptomatic bradycardia but also, much more rarely, varying degrees of heart block. These have not emerged as clinically significant in the majority of patients. Docetaxel causes comparable degrees of myelosuppression and neuropathy. Hypersensitivity reactions, including bronchospasm, dyspnea, and hypotension, are less frequent but occur to some degree in up to 25% of patients. Fluid retention appears to result from a vascular leak syndrome that can aggravate preexisting effusions. Rash can complicate docetaxel administration, appearing prominently as a pruritic maculopapular rash affecting the forearms, but it has also been associated with fingernail ridging, breakdown, and skin discoloration. Stomatitis appears to be somewhat more frequent than with paclitaxel.

Resistance to taxanes has been related to the emergence of efficient efflux of taxanes from tumor cells through the p170 P-glycoprotein (*mdr* gene product) or the presence of variant or mutant forms of tubulin. Epothilones represent a class of novel microtubule-stabilizing agents that have been conscientiously optimized for activity in taxane-resistant tumors. Ixabepilone has clear evidence of activity in breast cancers resistant to taxanes and anthracyclines such as doxorubicin. It retains acceptable expected side effects, including myelosuppression, and can also cause peripheral sensory neuropathy.

Estramustine was originally synthesized as a mustard derivative that might be useful in neoplasms that possessed estrogen receptors. However, no evidence of interaction with DNA was observed. Surprisingly, the drug caused metaphase arrest, and subsequent study revealed that it binds to microtubule-associated proteins, resulting in abnormal microtubule function. Estramustine binds to estramustine-binding proteins (EMBPs), which are notably present in prostate tumor tissue. The drug is used in patients with prostate cancer. Gastrointestinal and cardiovascular adverse effects related to the estrogen moiety occur in up to 10% of patients, including worsened heart failure and thromboembolic phenomena. Gynecomastia and nipple tenderness can also occur.

HORMONAL AGENTS

Steroid hormone receptor–related molecules have emerged as prominent targets for small molecules useful in cancer treatment. When bound to their cognate ligands, these receptors can alter gene transcription and, in certain tissues, induce apoptosis. The pharmacologic effect is a mirror or parody of the normal effects of the agents acting on nontransformed normal tissues, although the effects on tumors are mediated by indirect effects in some cases.

Glucocorticoids are generally given in "pulsed" high doses in leukemias and lymphomas, where they induce apoptosis in tumor cells. Cushing's syndrome or inadvertent adrenal suppression on withdrawal from high-dose glucocorticoids can be significant complications, along with infections common in immunosuppressed patients, in particular *Pneumocystis* pneumonia, which classically appears a few days after completing a course of high-dose glucocorticoids.

Tamoxifen is a partial estrogen receptor antagonist; it has a tenfold greater antitumor activity in breast cancer patients whose tumors express estrogen receptors than in those who have low or no levels of expression. It might be considered the prototypic "molecularly targeted" agent. Owing to its agonistic activities in vascular and uterine tissue, side effects include a somewhat increased risk of cardiovascular complications, such as thromboembolic phenomena, and a small increased incidence of endometrial carcinoma, which appears after chronic use (usually >5 years). Progestational agents —including medroxyprogesterone acetate, androgens including fluoxymesterone (Halotestin), and, paradoxically, estrogens—have approximately the same degree of activity in primary hormonal treatment of breast cancers that have elevated expression of estrogen receptor protein. Estrogen itself is not used often owing to prominent cardiovascular and uterotropic activity.

Aromatase refers to a family of enzymes that catalyze the formation of estrogen in various tissues, including the ovary and peripheral adipose tissue and some tumor cells. Aromatase inhibitors are of two types, the irreversible steroid analogues such as exemestane and the reversible inhibitors such as anastrozole or letrozole. Anastrozole is superior to tamoxifen in the adjuvant treatment of breast cancer in postmenopausal patients with estrogen receptor–positive tumors. Letrozole treatment affords benefit following tamoxifen treatment. Adverse effects of aromatase inhibitors may include an increased risk of osteoporosis.

Prostate cancer is classically treated by androgen deprivation. Diethylstilbestrol (DES) acting as an estrogen at the level of the hypothalamus to downregulate hypothalamic luteinizing hormone (LH) production results in decreased elaboration of testosterone by the testicle. For this reason, orchiectomy is equally as effective as moderate-dose DES, inducing responses in 80% of previously untreated patients with prostate cancer but without the prominent cardiovascular side effects of DES, including thrombosis and exacerbation of coronary artery disease. In the event that orchiectomy is not accepted by the patient, testicular androgen suppression can also be effected by luteinizing hormone –releasing hormone (LHRH) agonists such as leuprolide and goserelin. These agents cause tonic

stimulation of the LHRH receptor, with the loss of its normal pulsatile activation resulting in decreased output of LH by the anterior pituitary. Therefore, as primary hormonal manipulation in prostate cancer, one can choose orchiectomy or leuprolide, but not both. The addition of androgen receptor blockers, including flutamide or bicalutamide, is of uncertain additional benefit in extending overall response duration; the combined use of orchiectomy or leuprolide plus flutamide is referred to as *total androgen blockade*.

Tumors that respond to a primary hormonal manipulation may frequently respond to second and third hormonal manipulations. Thus, breast tumors that had previously responded to tamoxifen have, on relapse, notable response rates to withdrawal of tamoxifen itself or to subsequent addition of an aromatase inhibitor or progestin. Likewise, initial treatment of prostate cancers with leuprolide plus flutamide may be followed after disease progression by response to withdrawal of flutamide. These responses may result from the removal of antagonists from mutant steroid hormone receptors that have come to depend on the presence of the antagonist as a growth-promoting influence.

Additional strategies to treat refractory breast and prostate cancers that possess steroid hormone receptors may also address adrenal capacity to produce androgens and estrogens, even after orchiectomy or oophorectomy, respectively. Thus, aminoglutethimide or ketoconazole can be used to block adrenal synthesis by interfering with the enzymes of steroid hormone metabolism. Administration of these agents requires concomitant hydrocortisone replacement and additional glucocorticoid doses administered in the event of physiologic stress.

Humoral mechanisms can also result in complications from an underlying malignancy producing the hormone. Adrenocortical carcinomas can cause Cushing's syndrome as well as syndromes of androgen or estrogen excess. Mitotane can counteract these by decreasing synthesis of steroid hormones. Islet cell neoplasms can cause debilitating diarrhea, treated with the somatostatin analogue octreotide. Prolactin-secreting tumors can be effectively managed by the dopaminergic agonist bromocriptine.

Targeted Therapies

A better understanding of cancer cell biology has suggested many new targets for cancer drug discovery and development. These include the products of oncogenes and tumor-suppressor genes, regulators of cell death pathways, mediators of cellular immortality such as telomerase, and molecules responsible for microenvironmental molding such as proteases or angiogenic factors. The essential difference in the development of agents that would target these processes is that the basis for discovery of the candidate drug is the a priori importance of the target in the biology of the tumor, rather than the initial detection of drug candidates based on the phenomenon of tumor cell regression in tissue culture or in animals. The following examples reflect the rapidly evolving clinical research activity in this area. Figure 85-4 summarizes how FDA-approved targeted agents act.

HEMATOPOIETIC NEOPLASMS

Imatinib targets the ATP binding site of the $p210^{bcr\text{-}abl}$ protein tyrosine kinase that is formed as the result of the chromosome 9,22 translocation producing the Philadelphia chromosome in CML. Imatinib is superior to interferon plus chemotherapy in the initial treatment of the chronic phase of this disorder. It has lesser activity in the blast phase of CML, where the cells may have acquired additional mutations in $p210^{bcr\text{-}abl}$ itself or other genetic lesions. Its side effects are relatively tolerable in most patients and include hepatic dysfunction, diarrhea, and fluid retention. Rarely, patients receiving imatinib have decreased cardiac function, which may persist after discontinuation of the drug. The quality of response to imatinib enters into the decision about when to refer patients with CML for consideration of transplant approaches. Nilotinib is a tyrosine protein kinase inhibitor with a similar spectrum of activity to imatinib, but with increased potency and perhaps better tolerance by certain patients.

Dasatinib, another inhibitor of the $p210^{bcr\text{-}abl}$ oncoproteins, is active in certain mutant variants of $p210^{bcr\text{-}abl}$ that are refractory to imatinib and arise during therapy with imatinib or are present de novo. Dasatinib also has inhibitory action against kinases belonging to the src tyrosine protein kinase family; this activity may contribute to its effects in hematopoietic tumors and suggest a role in solid tumors where src kinases are active. Only the T315I mutant is resistant to dasatinib; a new class of inhibitors called aurora kinase inhibitors is in development to address this problem.

All-*trans*-retinoic acid (ATRA) targets the PML-retinoic acid receptor (RAR) α fusion protein, which is the result of the chromosome 15,17 translocation pathogenic for most forms of APL. Administered orally, it causes differentiation of the neoplastic promyelocytes to mature granulocytes and attenuates the rate of hemorrhagic complications. Adverse effects include headache with or without pseudotumor cerebri and gastrointestinal and cutaneous toxicities. Another active retinoid is the synthetic retinoid X receptor ligand bexarotene, which has activity in cutaneous T cell lymphoma.

Bortezomib is an inhibitor of the proteasome, the multisubunit assembly of protease activities responsible for the selective degradation of proteins important in regulating activation of transcription factors, including NF-κB and proteins regulating cell cycle progression. It has activity in multiple myeloma and certain lymphomas. Adverse effects include neuropathy, orthostatic hypotension with or

without hyponatremia, and reversible thrombocytopenia.

Vorinostat is an inhibitor of histone deacetylases, responsible for maintaining the proper orientation of histones on DNA, with resulting capacity for transcriptional readiness. Acetylated histones allow entry of transcription factors and therefore increased expression of genes that are selectively repressed in tumors. The result can be differentiation with the emergence of a more normal cellular phenotype, or cell cycle arrest with expression of endogenous regulators of cell cycle progression. Vorinostat is approved for clinical use in cutaneous T cell lymphoma, with dramatic skin clearing and very few side effects. Romidepsin is a distinct molecular class of histone deacetylase inhibitor also active in cutaneous T cell lymphoma.

DNA methyltransferase inhibitors including 5-aza-cytidine and 2'-deoxy-5-azacytidine (decitabine) can also increase transcription of genes "silenced" during the pathogenesis of a tumor by causing demethylation of the methylated cytosines that are acquired as an "epigenetic" (i.e., after the DNA is replicated) modification of DNA. These drugs were originally considered antimetabolites but have clinical value in myelodysplastic syndromes and certain leukemias when administered at low doses. Combinations of DNA methyltransferase inhibitors and histone deacetylase inhibitors may offer new approaches to regulate chromatin function.

Targeted toxins utilize macromolecules such as antibodies or cytokines with high affinity for defined tumor cell-surface molecules, such as a leukemia differentiation antigen, to which a therapeutic antibody can deliver a covalently linked potent cytotoxin, or a growth factor such as IL-2 to deliver a toxin (in the form of diphtheria toxin in denileukin diftitox) to cells bearing the IL-2 receptor. The value of such targeted approaches is that in addition to maximizing the therapeutic index by differential expression of the target in tumor (as opposed to nonrenewable normal cells), selection of patients for clinical use can capitalize on assessing the target in the tumor.

SOLID TUMORS

Small-molecule epidermal growth factor (EGF) antagonists act at the ATP binding site of the EGF receptor tyrosine kinase. In early clinical trials, gefitinib showed evidence of responses in a small fraction of patients with non-small cell lung cancer (NSCLC). Side effects were generally acceptable, consisting mostly of rash and diarrhea. Subsequent analysis of responding patients revealed a high frequency of activating mutations in the EGF receptor. Often patients who developed resistance to gefitinib have acquired additional mutations in the enzyme, similar to what was seen in imatinib-resistant CML. Erlotinib is another EGF receptor tyrosine kinase antagonist with somewhat superior outcome in clinical trials in NSCLC. Even patients with wild-type EGF receptors may benefit from erlotinib treatment. Lapatinib is a combined EGF receptor and erbB2 tyrosine kinase antagonist with activity in breast cancers refractory to anti-erbB2 antibodies.

In addition to the $p210^{bcr\text{-}abl}$ kinase, imatinib also has activity against the c-kit tyrosine kinase, activated in gastrointestinal stromal sarcoma, and the platelet-derived growth factor receptor (PDGF-R), activated by translocation in certain sarcomas. Imatinib has found clinical utility in these neoplasms previously refractory to chemotherapeutic approaches.

"Multitargeted" kinase antagonists are small-molecule ATP site-directed antagonists that inhibit more than one protein kinase. Drugs of this type with prominent activity against the vascular endothelial growth factor receptor (VEGF-R) tyrosine kinase have activity in renal cell carcinoma. Sorafenib is a VEGF-R antagonist with activity against the *raf* serine-threonine protein kinase as well. Sunitinib has anti-VEGF-R as well as anti-PDGF-R and anti-c-kit activity. It causes prominent responses as well as stabilization of disease in renal cell cancers and gastrointestinal stromal tumors. Side effects for both agents are mostly acceptable, with fatigue and diarrhea encountered with both agents. The "hand-foot syndrome" with erythema and desquamation of the distal extremities, in some cases requiring dose modification, may be seen with sorafenib. Temsirolimus and everolimus are mammalian target of rapamycin (mTOR) inhibitors with activity in renal cancers. They produce stomatitis, fatigue, and some hyperlipidemia (10%), myelosuppression (10%), and rare lung toxicity.

PERSONALIZED CANCER TREATMENT

The recognition that targeted therapies may benefit subsets of patients with an identical histologic diagnosis, but whose tumor is dependent for viability on the target's function, has spurred research to define molecular diagnostic approaches to define potentially responding patients. In addition, a patient's germ-line DNA may contain indicators of differential capacity to metabolize cancer chemotherapy agents and thus be susceptible to drug-induced toxicity. While efforts in this area are still a focus of both clinical and basic research, the following conclusions can be drawn, and are applicable to patients being initially managed in the primary care setting.

All patients undergoing initial diagnostic evaluation for breast cancer should have their tumor tested for the expression of the estrogen receptor (ER), progesterone receptor (PR), and the c-erbB2 (HER2; HER2/neu) oncoprotein by immunohistochemistry or fluorescence in situ hybridization (FISH). Patients expressing the ER and/or PR are candidates for adjuvant hormone receptor–directed therapies. Patients with evidence of abundant HER2 expression or HER2 gene amplification will likely derive benefit from trastuzumab. In addition, Oncotype Dx is a 21-gene expression test that has been

approved by the FDA for defining patients without lymph node involvement but with ER+ tumors who may have the greatest chance of benefiting from adjuvant chemotherapy added to adjuvant estrogen therapy. The MammaPrint test is similar in intent for node-negative patients but without reference to ER expression status.

The value of characterizing the mutational status of the epidermal growth factor (EGF) receptor pathway in patients with lung cancer is also a matter of current clinical investigations. While the tyrosine kinase inhibitor erlotinib is approved for use in all patients with NSCLC who have had progression of disease despite treatment with platinum-based chemotherapy, subsets of patients, such as female Asian nonsmokers, have a high evidence of EGF-R mutations resulting in marked sensitivity to erlotinib. And it is possible that in the larger population of patients with NSCLC, such testing may allow selection of patients in whom initial use of erlotinib may also be considered. Conversely, a mutated K-ras oncogene in patients with lung adenocarcinoma is associated with no benefit from erlotinib treatment.

In patients with colon cancer, a mutated K-ras oncogene is clearly associated with no benefit to the use of the EGF-R–directed antibody cetuximab, and characterization of K-ras mutational status should be undertaken as part of the routine diagnostic evaluation of patients with newly diagnosed metastatic or newly recurrent colon cancer. Patients undergoing diagnostic evaluation for initial treatment of metastatic colon cancer might usefully undergo evaluation of their germ-line uridine diphosphate glucuronosyl transferase (UGT) 1A1 allele status, as the expression of variant alleles at that locus influences susceptibility to irinotecan-induced hematologic toxicity. Patients with known Gilbert's disease should receive irinotecan very cautiously or perhaps not at all.

Acute Complications of Cancer Chemotherapy

MYELOSUPPRESSION

The common cytotoxic chemotherapeutic agents almost invariably affect bone marrow function. Titration of this effect determines the MTD of the agent on a given schedule. The normal kinetics of blood cell turnover influences the sequence and sensitivity of each of the formed elements. Polymorphonuclear leukocytes (PMNs; $t_{1/2}$ = 6–8 h), platelets ($t_{1/2}$ = 5–7 days), and red blood cells (RBCs; $t_{1/2}$ = 120 days) respectively have most, less, and least susceptibility to usually administered cytotoxic agents. The nadir count of each cell type in response to classes of agents is characteristic. Maximal neutropenia occurs 6–14 days after conventional doses of anthracyclines, antifolates, and antimetabolites. Alkylating agents differ from each other in the timing of cytopenias. Nitrosoureas, DTIC, and procarbazine can display delayed marrow toxicity, first appearing 6 weeks after dosing.

Complications of myelosuppression result from the predictable sequelae of the missing cells' function. *Febrile neutropenia* refers to the clinical presentation of fever (one temperature ≥38.5°C or three readings ≥38°C but ≤38.5°C per 24 h) in a neutropenic patient with an uncontrolled neoplasm involving the bone marrow or, more usually, in a patient undergoing treatment with cytotoxic agents. Mortality from uncontrolled infection varies inversely with the neutrophil count. If the nadir neutrophil count is >1000/μL, there is little risk; if <500/μL, risk of death is markedly increased. Management of febrile neutropenia has conventionally included empirical coverage with antibiotics for the duration of neutropenia (Chap. 86). Selection of antibiotics is governed by the expected association of infections with certain underlying neoplasms; careful physical examination (with scrutiny of catheter sites, dentition, mucosal surfaces, and perirectal and genital orifices by gentle palpation); chest x-ray; and Gram stain and culture of blood, urine, and sputum (if any) to define a putative site of infection. In the absence of any originating site, a broadly acting β-lactam with anti-*Pseudomonas* activity, such as ceftazidime, is begun empirically. The addition of vancomycin to cover potential cutaneous sites of origin (until these are ruled out or shown to originate from methicillin-sensitive organisms) or metronidazole or imipenem for abdominal or other sites favoring anaerobes reflects modifications tailored to individual patient presentations. The coexistence of pulmonary compromise raises a distinct set of potential pathogens, including *Legionella*, *Pneumocystis*, and fungal agents that may require further diagnostic evaluations, such as bronchoscopy with bronchoalveolar lavage. Febrile neutropenic patients can be stratified broadly into two prognostic groups. The first, with expected short duration of neutropenia and no evidence of hypotension or abdominal or other localizing symptoms, may be expected to do well even with oral regimens, e.g., ciprofloxacin or moxifloxacin, or amoxicillin plus clavulanic acid. A less favorable prognostic group is patients with expected prolonged neutropenia, evidence of sepsis, and end organ compromise, particularly pneumonia. These patients require tailoring of their antibiotic regimen to their underlying presentation, with frequent empirical addition of antifungal agents if fever persists for 7 days without identification of an adequately treated organism or site.

Transfusion of granulocytes has no role in the management of febrile neutropenia, owing to their exceedingly short half-life, mechanical fragility, and clinical syndromes of pulmonary compromise with leukostasis after their use. Instead, colony-stimulating factors (CSFs) are used to augment bone marrow production of PMNs. Early-acting factors such as IL-1, IL-3, and stem cell factor have not been as useful clinically as late-acting, lineage-specific factors such as G-CSF (granulocyte colony-stimulating factor) or GM-CSF (granulocyte-macrophage colony-stimulating factor), erythropoietin

(EPO), thrombopoietin, IL-6, and IL-11. CSFs may easily become overused in oncology practice. The settings in which their use has been proved effective are limited. G-CSF, GM-CSF, EPO, and IL-11 are currently approved for use. The American Society of Clinical Oncology has developed practice guidelines for the use of G-CSF and GM-CSF (Table 85–3).

Table 85–3 Indications for the Clinical Use of G-CSF or GM-CSF

Preventive Uses
With the first cycle of chemotherapy (so-called primary CSF administration)
Not needed on a routine basis
Use if the probability of febrile neutropenia is ≥20%
Use if patient has preexisting neutropenia or active infection
Age >65 years treated for lymphoma with curative intent or other tumor treated by similar regimens
Poor performance status
Extensive prior chemotherapy
Dose-dense regimens in a clinical trial or with strong evidence of benefit
With subsequent cycles if febrile neutropenia has previously occurred (so-called secondary CSF administration)
Not needed after short-duration neutropenia without fever
Use if patient had febrile neutropenia in previous cycle
Use if prolonged neutropenia (even without fever) delays therapy
Therapeutic Uses
Afebrile neutropenic patients
No evidence of benefit
Febrile neutropenic patients
No evidence of benefit
May feel compelled to use in the face of clinical deterioration from sepsis, pneumonia, or fungal infection, but benefit unclear
In bone marrow or peripheral blood stem cell transplantation
Use to mobilize stem cells from marrow
Use to hasten myeloid recovery
In acute myeloid leukemia
G-CSF of minor or no benefit
GM-CSF of no benefit and may be harmful
In myelodysplastic syndromes
Not routinely beneficial
Use intermittently in subset with neutropenia and recurrent infection
What Dose and Schedule Should Be Used?
G-CSF: 5 mg/kg per day subcutaneously GM-CSF: 250 mg/m^2 per day subcutaneously
Peg-filgrastim: one dose of 6 mg 24 h after chemotherapy
When Should Therapy Begin and End?
When indicated, start 24–72 h after chemotherapy
Continue until absolute neutrophil count is 10,000/μL
Do not use concurrently with chemotherapy or radiation therapy

Abbreviations: G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor.

Source: From the American Society of Clinical Oncology.

Primary prophylaxis (i.e., shortly after completing chemotherapy to reduce the nadir) administers G-CSF to patients receiving cytotoxic regimens associated with a 20% incidence of febrile neutropenia. "Dose-dense" regimens, where cycling of chemotherapy is intended to be completed without delay of administered doses, may also benefit, but such patients should be on a clinical trial. Administration of G-CSF in these circumstances has reduced the incidence of febrile neutropenia in several studies by about 50%. Most patients, however, receive regimens that do not have such a high risk of expected febrile neutropenia, and therefore most patients initially should not receive G-CSF or GM-CSF. Special

circumstances—such as a documented history of febrile neutropenia with the regimen in a particular patient or categories of patients at increased risk, such as patients older than age 65 years with aggressive lymphoma treated with curative chemotherapy regimens; extensive compromise of marrow by prior radiation or chemotherapy; or active, open wounds or deep-seated infection—may support primary treatment with G-CSF or GM-CSF. Administration of G-CSF or GM-CSF to afebrile neutropenic patients or to patients with low-risk febrile neutropenia is not recommended, and patients receiving concomitant chemoradiation treatment, particularly those with thoracic neoplasms, likewise are not generally recommended for treatment. In contrast, administration of G-CSF to high-risk patients with febrile neutropenia and evidence of organ compromise including sepsis syndrome, invasive fungal infection, concurrent hospitalization at the time fever develops, pneumonia, profound neutropenia (<0.1 x 10^9/L), or age >65 years is reasonable.

Secondary prophylaxis refers to the administration of CSFs in patients who have experienced a neutropenic complication from a prior cycle of chemotherapy; dose reduction or delay may be a reasonably considered alternative. G-CSF or GM-CSF is conventionally started 24–72 h after completion of chemotherapy and continued until a PMN count of 10,000/μL is achieved, unless a "depot" preparation of G-CSF such as pegfilgrastim is used, where one dose is administered at least 14 days before the next scheduled administration of chemotherapy. Also, patients with myeloid leukemias undergoing induction therapy may have a slight reduction in the duration of neutropenia if G-CSF is commenced after completion of therapy and may be of particular value in elderly patients, but the influence on long-term outcome has not been defined. GM-CSF probably has a more restricted utility than G-CSF, with its use currently limited to patients after autologous bone marrow transplants, although proper head-to-head comparisons with G-CSF have not been conducted in most instances. GM-CSF may be associated with more systemic side effects.

Dangerous degrees of thrombocytopenia do not frequently complicate the management of patients with solid tumors receiving cytotoxic chemotherapy (with the possible exception of certain carboplatin-containing regimens), but they are frequent in patients with certain hematologicneoplasms where marrow is infiltrated with tumor. Severe bleeding related to thrombocytopenia occurs with increased frequency at platelet counts <20,000/μL and is very prevalent at counts <5000/μL.

The precise "trigger" point at which to transfuse patients is being evaluated in a randomized study. This issue is important not only because of the costs of frequent transfusion, but unnecessary platelet transfusions expose the patient to the risks of allosensitization and loss of value from subsequent transfusion owing to rapid platelet clearance, as well as the infectious and hypersensitivity risks inherent in any transfusion. Prophylactic transfusions to keep platelets >20,000/μL are reasonable in patients with leukemia who are stressed by fever or concomitant medical conditions (the threshold for transfusion is 10,000/μL in patients with solid tumors and no other bleeding diathesis or physiologic stressors such as fever or hypotension, a level that might also be reasonably considered for leukemia patients who are thrombocytopenic but not stressed or bleeding). In contrast, patients with myeloproliferative states may have functionally altered platelets despite normal platelet counts, and transfusion with normal donor platelets should be considered for evidence of bleeding in these patients. Careful review of medication lists to prevent exposure to nonsteroidal anti-inflammatory agents and maintenance of clotting factor levels adequate to support near-normal prothrombin and partial thromboplastin time tests are important in minimizing the risk of bleeding in the thrombocytopenic patient.

Certain cytokines in clinical investigation have shown an ability to increase platelets (e.g., IL-6, IL-1, thrombopoietin), but clinical benefit and safety are not yet proven. IL-11 (oprelvekin) is approved for use in the setting of expected thrombocytopenia, but its effects on platelet counts are small, and it is associated with side effects such as headache, fever, malaise, syncope, cardiac arrhythmias, and fluid retention.

Anemia associated with chemotherapy can be managed by transfusion of packed RBCs. Transfusion is not undertaken until the hemoglobin falls to <80 g/L (8 g/dL) or if compromise of end organ function occurs, or an underlying condition (e.g., coronary artery disease) calls for maintenance of hemoglobin >90 g/L (9 g/dL). Patients who are to receive therapy for >2 months on a "stable" regimen and who are likely to require continuing transfusions are also candidates for erythropoietin (EPO). Randomized trials in certain tumors have raised the possibility that EPO use may promote tumor-related adverse events. This information should be considered in the care of individual patients. In the event EPO treatment is undertaken, maintenance of hemoglobin of 90–100 g/L (9–10 g/dL) should be the target. In the setting of adequate iron stores and serum EPO levels <100 ng/mL, EPO, 150 U three times a week, can produce a slow increase in hemoglobin over about 2 months of administration. Depot formulations can be administered less frequently. It is unclear whether higher hemoglobin levels, up to 110–120 g/L (11–12 g/dL), are associated with improved quality of life to a degree that justifies the more intensive EPO use. Efforts to achieve levels at or above 120 g/L (12 g/dL) have been associated with increased thromboses and mortality rates. EPO may rescue hypoxemic cells from death and contribute to tumor radioresistance.

NAUSEA AND VOMITING

The most common side effect of chemotherapy administration is nausea, with or without vomiting. Nausea may be acute (within 24 h of chemotherapy), delayed (>24 h), or anticipatory of the receipt of chemotherapy. Patients may be likewise stratified for their risk of susceptibility to nausea and vomiting, with increased risk in young, female, heavily pretreated patients without a history of alcohol or drug use but with a history of motion or morning sickness. Antineoplastic agents vary in their capacity to cause nausea and vomiting. Highly emetogenic drugs (>90%) include mechlorethamine, streptozotocin, DTIC, cyclophosphamide at >1500 mg/m^2, and cisplatin; moderately emetogenic drugs (30–90% risk) include carboplatin, cytosine arabinoside (>1 mg/m^2), ifosfamide, conventional-dose cyclophosphamide, and anthracyclines; low-risk (10–30%) agents include fluorouracil, taxanes, etoposide, and bortezomib, with minimal risk (<10%) afforded by treatment with antibodies, bleomycin, busulfan, fludarabine, and vinca alkaloids. *Emesis* is a reflex caused by stimulation of the vomiting center in the medulla. Input to the vomiting center comes from the chemoreceptor trigger zone (CTZ) and afferents from the peripheral gastrointestinal tract, cerebral cortex, and heart. The different emesis "syndromes" require distinct management approaches. In addition, a conditioned reflex may contribute to anticipatory nausea arising after repeated cycles of chemotherapy. Accordingly, antiemetic agents differ in their locus and timing of action. Combining agents from different classes or the sequential use of different classes of agent is the cornerstone of successful management of chemotherapy-induced nausea and vomiting. Of great importance are the prophylactic administration of agents, and such psychological techniques as the maintenance of a supportive milieu, counseling, and relaxation to augment the action of antiemetic agents.

Serotonin antagonists (5-HT3) and neurokine (NK1) receptor antagonists are useful in "high-risk" chemotherapy regimens. The combination acts at both peripheral gastrointestinal as well as CNS sites that control nausea and vomiting. For example, the 5-HT3 blocker dolasetron (Anzamet), 100 mg intravenously or orally; dexamethasone, 12 mg; and the NK1 antagonist aprepitant, 125 mg orally, are combined on the day of administration of severely emetogenic regimens, with repetition of dexamethasone (8 mg) and aprepitant (80 mg) on days 2 and 3 for delayed nausea. Alternate 5-HT3 antagonists include ondansetron (Zofran), given as 0.15 mg/kg intravenously for three doses just before and at 4 and 8 h after chemotherapy; palonosetron (Aloxi) at 0.25 mg over 30 s, 30 min prechemotherapy; and granisetron (Kytril), given as a single dose of 0.01 mg/kg just before chemotherapy. Emesis from moderately emetic chemotherapy regimens may be prevented with a 5-HT3 antagonist and dexamethasone alone for patients not receiving doxorubicin and cyclophosphamide combinations; the latter combination requires the 5-HT3/dexamethasone/aprepitant on day 1 but aprepitant alone on days 2 and 3. Emesis from low-emetic-risk regimens may be prevented with 8 mg of dexamethasone alone, or with non-5-HT3, non-NK1 antagonist approaches including the following.

Antidopaminergic phenothiazines act directly at the CTZ and include prochlorperazine (Compazine), 10 mg intramuscularly or intravenously, 10–25 mg orally or 25 mg per rectum every 4–6 h for up to four doses; and thiethylperazine (Torecan), 10 mg by potentially all the above routes every 6 h. Haloperidol (Haldol) is a butyrophenone dopamine antagonist given at 1.0 to 1 mg intramuscularly or orally every 8 h. Antihistamines such as diphenhydramine (Benadryl) have little intrinsic antiemetic capacity but are frequently given to prevent or treat dystonic reactions that can complicate use of the antidopaminergic agents. Lorazepam (Ativan) is a short-acting benzodiazepine that provides an anxiolytic effect to augment the effectiveness of a variety of agents when used at 1–2 mg intramuscularly, intravenously, or orally every 4–6 h. Metoclopramide (Reglan) acts on peripheral dopamine receptors to augment gastric emptying and is used in high doses for highly emetogenic regimens (1–2 mg/kg intravenously 30 min before chemotherapy and every 2 h for up to three additional doses as needed); intravenous doses of 10–20 mg every 4–6 h as needed or 50 mg orally 4 h before and 8 and 12 h after chemotherapy are used for moderately emetogenic regimens. 5-9-Tetrahydrocannabinol (Marinol) is a rather weak antiemetic compared to other available agents, but it may be useful for persisting nausea and is used orally at 10 mg every 3–4 h as needed.

DIARRHEA

Regimens that include fluorouracil infusions and/or irinotecan may produce severe diarrhea. Similar to the vomiting syndromes, chemotherapy-induced diarrhea may be immediate or can occur in a delayed fashion up to 48–72 h after the drugs. Careful attention to maintained hydration and electrolyte repletion, intravenously if necessary, along with antimotility treatments such as "high-dose" loperamide, commenced with 4 mg at the first occurrence of diarrhea, with 2 mg repeated every 2 h until 12 h without loose stools, not to exceed a total daily dose of 16 mg. Octreotide (100–150 μg), a somatostatin analogue, or opiate-based preparations may be considered for patients not responding to loperamide.

MUCOSITIS

Irritation and inflammation of the mucous membranes particularly afflicting the oral and anal mucosa, but potentially involving the gastrointestinal tract, may accompany cytotoxic chemotherapy. Mucositis is due to damage to the proliferating cells at the base of the mucosal squamous epithelia or in the intestinal crypts. Topical therapies, including anesthetics and barrier-creating preparations, may provide symptomatic relief in mild cases. Palifermin or keratinocyte growth factor, a member of the

fibroblast growth factor family, is effective in preventing severe mucositis in the setting of high-dose chemotherapy with stem cell transplantation for hematologicmalignancies. It may also prevent or ameliorate mucositis from radiation.

ALOPECIA

Chemotherapeutic agents vary widely in causing alopecia, with anthracyclines, alkylating agents, and topoisomerase inhibitors reliably causing near-total alopecia when given at therapeutic doses. Antimetabolites are more variably associated with alopecia. Psychological support and the use of cosmetic resources are to be encouraged, and "chemo caps" that reduce scalp temperature to decrease the degree of alopecia should be discouraged, particularly during treatment with curative intent of neoplasms, such as leukemia or lymphoma, or in adjuvant breast cancer therapy. The richly vascularized scalp can certainly harbor micrometastatic or disseminated disease.

GONADAL DYSFUNCTION AND PREGNANCY

Cessation of ovulation and azoospermia reliably result from alkylating agent– and topoisomerase poison–containing regimens. The duration of these effects varies with age and sex. Males treated for Hodgkin's disease with mechlorethamine- and procarbazine-containing regimens are effectively sterile, whereas fertility usually returns after regimens that include cisplatin, vinblastine, or etoposide and after bleomycin for testicular cancer. Sperm banking before treatment may be considered to support patients likely to be sterilized by treatment. Females experience amenorrhea with anovulation after alkylating agent therapy; they are likely to recover normal menses if treatment is completed before age 30 but unlikely to recover menses after age 35. Even those who regain menses usually experience premature menopause. As the magnitude and extent of decreased fertility can be difficult to predict, patients should be counseled to maintain effective contraception, preferably by barrier means, during and after therapy. Resumption of efforts to conceive should be considered in the context of the patient's likely prognosis. Hormone replacement therapy should be undertaken in women who do not have a hormonally responsive tumor. For those patients who have had a hormone-sensitive tumor primarily treated by a local modality, conventional practice would counsel against hormone replacement, but this issue is under investigation.

Chemotherapy agents have variable effects on the success of pregnancy. All agents tend to have increased risk of adverse outcomes when administered during the first trimester, and strategies to delay chemotherapy, if possible, until after this milestone should be considered if the pregnancy is to continue to term. Patients in their second or third trimester can be treated with most regimens for the common neoplasms afflicting women in their childbearing years, with the exception of antimetabolites, particularly antifolates, which have notable teratogenic or fetotoxic effects throughout pregnancy. The need for anticancer chemotherapy per se is infrequently a clear basis to recommend termination of a concurrent pregnancy, although each treatment strategy in this circumstance must be tailored to the individual needs of the patient. **Chronic effects of cancer treatment are reviewed in Chap. 102.**

BIOLOGIC THERAPY

The goal of biologic therapy is to manipulate the host–tumor interaction in favor of the host, potentially at an optimum biologic dose that might be different than an MTD. As a class, biologic therapies may be distinguished from molecularly targeted agents in that many biologic therapies require an active response (e.g., reexpression of silenced genes, or antigen expression) on the part of the tumor cell or on the part of the host (e.g., immunologic effects) to allow therapeutic effect. This may be contrasted with the more narrowly defined antiproliferative or apoptotic response that is the ultimate goal of molecularly targeted agents discussed above. However, there is much commonality in the strategies to evaluate and use molecularly targeted and biologic therapies.

Immune Mediators of Antitumor Effects

Tumors have a variety of means of avoiding the immune system: (1) they are often only subtly different from their normal counterparts; (2) they are capable of downregulating their major histocompatibility complex antigens, effectively masking them from recognition by T cells; (3) they are inefficient at presenting antigens to the immune system; (4) they can cloak themselves in a protective shell of fibrin to minimize contact with surveillance mechanisms; and (5) they can produce a range of soluble molecules, including potential immune targets, that can distract the immune system from recognizing the tumor cell or can kill the immune effector cells. Some of the cell products initially polarize the immune response away from cellular immunity (shifting from T_H1 to T_H2 responses; Chap. 314) and ultimately lead to defects in T cells that prevent their activation and cytotoxic activity. Cancer treatment further suppresses host immunity. A variety of strategies are being tested to overcome these barriers.

CELL-MEDIATED IMMUNITY

The strongest evidence that the immune system can exert clinically meaningful antitumor effects comes from allogeneic bone marrow transplantation. Adoptively transferred T cells from the donor expand in the tumor-bearing host, recognize the tumor as being foreign, and can mediate impressive antitumor effects (graft-versus-tumor effects). Three types of experimental interventions are being developed to take advantage of the ability of T cells to kill tumor cells.

1. Allogeneic T cells are transferred to cancer-bearing hosts in three major settings: in the form of allogeneic bone marrow transplantation, as pure lymphocyte transfusions following bone marrow recovery after allogeneic bone marrow transplantation, and as pure lymphocyte transfusions following immunosuppressive (but not myeloablative) therapy (so-called minitransplants). In each of these settings, the effector cells are donor T cells that recognize the tumor as being foreign, probably through minor histocompatibility differences. The main risk of such therapy is the development of graft-versus-host disease because of the minimal difference between the cancer and the normal host cells. This approach has been highly effective in certain hematologic cancers.
2. Autologous T cells are removed from the tumor-bearing host, manipulated in several ways in vitro, and given back to the patient. The two major classes of autologous T cell manipulation are (a) to develop tumor antigen–specific T cells and expand them to large numbers over many weeks ex vivo before administration, and (b) to activate the cells with polyclonal stimulators such as anti-CD3 and anti-CD28 after a short period ex vivo and try to expand them in the host after adoptive transfer with stimulation by IL-2, for example. Short periods removed from the patient permit the cells to overcome the tumor-induced T cell defects, and such cells traffic and home to sites of disease better than cells that have been in culture for many weeks.
3. Tumor vaccines are aimed at boosting T cell immunity. The finding that mutant oncogenes that are expressed only intracellularly can be recognized as targets of T cell killing greatly expanded the possibilities for tumor vaccine development. No longer is it difficult to find something different about tumor cells. However, major difficulties remain in getting the tumor-specific peptides presented in a fashion to prime the T cells. Tumors themselves are very poor at presenting their own antigens to T cells at the first antigen exposure (*priming*). Priming is best accomplished by professional antigen-presenting cells (dendritic cells). Thus, a number of experimental strategies are aimed at priming host T cells against tumor-associated peptides. Vaccine adjuvants such as GM-CSF appear capable of attracting antigen-presenting cells to a skin site containing a tumor antigen. Such an approach has been documented to eradicate microscopic residual disease in follicular lymphoma and give rise to tumor-specific T cells. Purified antigen-presenting cells can be pulsed with tumor, its membranes, or particular tumor antigens and delivered as a vaccine. One such vaccine, Sipuleucel-T, is approved for use in patients with hormone-independent prostate cancer. In this approach, the patient undergoes leukapheresis, wherein mononuclear cells (that include antigen-presenting cells) are removed from the patient's blood. The cells are pulsed in a laboratory with an antigenic fusion protein comprising a protein frequently expressed by prostate cancer cells, prostate acid phosphatase, fused to GM-CSF, and matured to increase their capacity to present the antigen to immune effector cells. The cells are then returned to the patient, in a well-tolerated treatment. While no objective tumor response was documented, median survival was increased about 4 months. Tumor cells can also be transfected with genes that attract antigen-presenting cells. Vaccines against viruses that cause cancers are safe and effective. Hepatitis B vaccine prevents hepatocellular carcinoma, and a tetravalent human papilloma virus vaccine prevents infection by virus types currently accounting for 70% of cervical cancer. These vaccines are ineffective at treating patients who have developed a virus-induced cancer.

ANTIBODIES

In general, antibodies are not very effective at killing cancer cells. Because the tumor seems to influence the host toward making antibodies rather than generating cellular immunity, it is inferred that antibodies are easier for the tumor to fend off. Many patients can be shown to have serum antibodies directed at their tumors, but these do not appear to influence disease progression. However, the ability to grow very large quantities of high-affinity antibody directed at a tumor by the hybridoma technique has led to the application of antibodies to the treatment of cancer.

Clinical antitumor efficacy has been obtained using antibodies where the antigen-combining regions are grafted onto human immunoglobulin gene products (chimerized or humanized) or derive de novo from mice bearing human immunoglobulin gene loci. Such humanized antibodies against the CD20 molecule expressed on B cell lymphomas (rituximab) and against the HER-2/neu receptor overexpressed on epithelial cancers, especially breast cancer (trastuzumab), have become reliable tools in the oncologist's armamentarium. Each used alone can cause tumor regression (rituximab more than trastuzumab), and both appear to potentiate the effects of combination chemotherapy given just after antibody administration. Antibodies to CD52 are active in chronic lymphoid leukemia and T cell malignancies. EGF-R–directed antibodies (such as cetuximab and panitumumab) have activity in colorectal cancer refractory to chemotherapy, particularly when utilized to augment the activity of an additional chemotherapy program, and in the primary treatment of head and neck cancers treated with radiation therapy. The mechanism of action is unclear. Direct effects on the tumor may mediate an antiproliferative effect as well as stimulate the participation of host mechanisms involving immune cell or complement-mediated response to tumor cell–bound antibody. Alternatively, the antibody may alter the release of paracrine factors promoting tumor cell survival.

The anti-VEGF antibody bevacizumab shows little evidence of antitumor effect when used alone, but when combined with chemotherapeutic agents it improves the magnitude of tumor shrinkage and time to disease progression in colorectal, lung, and breast cancer. The mechanism for the effect is unclear and may relate to the capacity of the antibody to alter delivery and tumor uptake of the active

chemotherapeutic agent.

Side effects include infusion-related hypersensitivity reactions, usually limited to the first infusion, which can be managed with glucocorticoid and/or antihistamine prophylaxis. In addition, distinct syndromes have emerged with different antibodies. Anti-EGF-R antibodies produce an acneiform rash that poorly responds to steroid cream treatment. Trastuzumab (anti-HER2) can inhibit cardiac function, particularly in those patients with prior exposure to anthracyclines. Bevacizumab has a number of side effects of medical significance, including hypertension, thrombosis, proteinuria, hemorrhage, and gastrointestinal perforations with or without prior surgeries.

Conjugation of antibodies to drugs and toxins is discussed above; conjugates of antibodies with isotopes, photodynamic agents, and other killing moieties may also be effective. Radioconjugates targeting CD20 on lymphomas have been approved for use [ibritumomab tiuxetan (Zevalin), using yttrium-90 or ^{131}I-tositumomab]. Other conjugates are associated with problems that have not yet been solved (e.g., antigenicity, instability, poor tumor penetration).

CYTOKINES

There are >70 separate proteins and glycoproteins with biologic effects in humans: interferon (IFN) α, β, γ; IL-1 through -29 (so far); the tumor necrosis factor (TNF) family [including lymphotoxin, TNF-related apoptosis-inducing ligand (TRAIL), CD40 ligand, and others]; and the chemokine family. Only a fraction of these has been tested against cancer; only IFN-α and IL-2 are in routine clinical use.

About 20 different genes encode IFN-α, and their biologic effects are indistinguishable. Interferon induces the expression of many genes, inhibits protein synthesis, and exerts a number of different effects on diverse cellular processes. The two recombinant forms that are commercially available are IFN-α2a and -α2b. Interferon is not curative for any tumor but can induce partial responses in follicular lymphoma, hairy cell leukemia, CML, melanoma, and Kaposi's sarcoma. It has been used in the adjuvant setting in stage II melanoma, multiple myeloma, and follicular lymphoma, with uncertain effects on survival. It produces fever, fatigue, a flulike syndrome, malaise, myelosuppression, and depression and can induce clinically significant autoimmune disease.

IL-2 must exert its antitumor effects indirectly through augmentation of immune function. Its biologic activity is to promote the growth and activity of T cells and natural killer (NK) cells. High doses of IL-2 can produce tumor regression in certain patients with metastatic melanoma and renal cell cancer. About 2–5% of patients may experience complete remissions that are durable, unlike any other treatment for these tumors. IL-2 is associated with myriad clinical side effects: intravascular volume depletion, capillary leak syndrome, adult respiratory distress syndrome, hypotension, fever, chills, skin rash, and impaired renal and liver function. Patients may require blood pressure support and intensive care to manage the toxicity. However, once the agent is stopped, most of the toxicities reverse completely within 3–6 days.

GENE THERAPIES

No gene therapy has been approved for routine clinical use. Several strategies are under evaluation, including the use of viruses that cannot replicate to express genes that can allow the action of drugs or directly inhibit cancer cell growth, viruses that can actually replicate but only in the context of the tumor cell, or viruses that can express antigens in the context of the tumor and therefore provoke a host-mediated immune response. Key issues in the success of these approaches will be in defining safe viral vector systems that escape host immune function and effectively target the tumor or tumor cell milieu. Other gene therapy strategies would utilize therapeutic oligonucleotides to target the expression of genes important in the maintenance of tumor cell viability.

ACKNOWLEDGMENTS

Stephen M. Hahn, MD, and Eli Glatstein, MD, contributed a chapter on radiation therapy in a prior edition, and some of their material has been incorporated into this chapter.

FURTHER READINGS

American Society of Clinical Oncology: 2006 Update of recommendations for the use of white blood cell growth factors: An evidence-based clinical practice guideline. J Clin Oncol 24:3187, 2006

Chabner ba, Longo DL (eds): *Cancer Chemotherapy and Biotherapy: Principles and Practice* , 5th ed. Philadelphia, Lippincott Williams & Wilkins, 2011

Ettinger DS et al: Antiemesis: Clinical practice guidelines in oncology. J Natl Compr Cancer Network 7:572, 2009[PMID: 19460282]

Mcdermott U, Settleman J: Personalized cancer therapy with selective kinase inhibitors: An emerging paradigm in medical oncology. J Clin Oncol 27: 5650, 2009[PMID: 19858389]

Olopade OI et al. Advances in breast cancer: Pathways to personalized medicine. Clin Cancer Res 14: 7988, 2008[PMID: 19088015]

Rizzo JD et al: ASCO special article: Use of epoetin and darbepoetin in patients with cancer: 2007 American Society of Clinical Oncology / American Society of Hematology Clinical Practice Guideline update. J Clin Oncol 26, 132, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS IN PATIENTS WITH CANCER: INTRODUCTION

Infections are a common cause of death and an even more common cause of morbidity in patients with a wide variety of neoplasms. Autopsy studies show that most deaths from acute leukemia and half of deaths from lymphoma are caused directly by infection. With more intensive chemotherapy, patients with solid tumors have also become more likely to die of infection. Fortunately, an evolving approach to prevention and treatment of infectious complications of cancer has decreased infection-associated mortality rates and will probably continue to do so. This accomplishment has resulted from three major steps:

1. The concept of "early empirical" antibiotics reduced mortality rates among patients with leukemia and bacteremia from 84% in 1965 to 44% in 1972. This dramatic improvement is attributed to early intervention with appropriate antimicrobial therapy.
2. "Empirical" antifungal therapy has lowered the incidence of disseminated fungal infection; in trial settings, mortality rates now range from 7% to 21%. An antifungal agent is administered—on the basis of likely fungal infection—to neutropenic patients who, after 4–7 days of antibiotic therapy, remain febrile but have no positive cultures. In one study, the 7-day survival rate was ~85% among patients who had fever and neutropenia as a result of cancer chemotherapy and who required antifungal therapy.
3. Use of antibiotics for afebrile neutropenic patients as broad-spectrum prophylaxis against infections has decreased both mortality and morbidity even further. The current approach to treatment of severely neutropenic patients (e.g., those receiving high-dose chemotherapy for leukemia or high-grade lymphomas) is based on initial prophylactic therapy at the onset of neutropenia, with subsequent "empirical" antibacterial therapy targeting the organisms whose involvement is likely in light of physical findings (most often fever alone), and finally "empirical" antifungal therapy based on the known likelihood that fungal infection will become a serious issue after 4–7 days of broad-spectrum antibacterial therapy.

A physical predisposition to infection in patients with cancer (Table 86–1) can be a result of the neoplasm's production of a break in the skin. For example, a squamous cell carcinoma may cause local invasion of the epidermis, which allows bacteria to gain access to the subcutaneous tissue and permits the development of cellulitis. The artificial closing of a normally patent orifice can also predispose to infection; for example, obstruction of a ureter by a tumor can cause urinary tract infection, and obstruction of the bile duct can cause cholangitis. Part of the host's normal defense against infection depends on the continuous emptying of a viscus; without emptying, a few bacteria that are present as a result of bacteremia or local transit can multiply and cause disease.

Table 86–1 Disruption of Normal Barriers that May Predispose to Infections in Patients with Cancer

Type of Defense	Specific Lesion	Cells Involved	Organism	Cancer Association	Disease
Physical barrier	Breaks in skin	Skin epithelial cells	Staphylococci, streptococci	Head and neck, squamous cell carcinoma	Cellulitis, extensive skin infection
Emptying of fluid collections	Occlusion of orifices: ureters, bile duct, colon	Luminal epithelial cells	Gram-negative bacilli	Renal, ovarian, biliary tree, metastatic diseases of many cancers	Rapid, overwhelming bacteremia; urinary tract infection
Lymphatic function	Node dissection	Lymph nodes	Staphylococci, streptococci	Breast cancer surgery	Cellulitis
Splenic clearance of microorganisms	Splenectomy	Splenic reticuloendothelial cells	*Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis, Babesia.*	Hodgkin's disease, leukemia, idiopathic thrombocytopenic purpura	Rapid, overwhelming sepsis

			Capnocytophaga canimorsus		
Phagocytosis	Lack of granulocytes	Granulocytes (neutrophils)	Staphylococci, streptococci, enteric organisms, fungi	Hairy cell, acute myelocytic, and acute lymphocytic leukemias	Bacteremia
Humoral immunity	Lack of antibody	B cells	*S. pneumoniae, H. influenzae, N. meningitidis*	Chronic lymphocytic leukemia, multiple myeloma	Infections with encapsulated organisms, sinusitis, pneumonia
Cellular immunity	Lack of T cells	T cells and macrophages	*Mycobacterium tuberculosis, Listeria,* herpesviruses, fungi, intracellular parasites	Hodgkin's disease, leukemia, T cell lymphoma	Infections with intracellular bacteria, fungi, parasites

A similar problem can affect patients whose lymph node integrity has been disrupted by radical surgery, particularly patients who have had radical node dissections. A common clinical problem following radical mastectomy is the development of cellulitis (usually caused by streptococci or staphylococci) because of lymphedema and/or inadequate lymph drainage. In most cases, this problem can be addressed by local measures designed to prevent fluid accumulation and breaks in the skin, but antibiotic prophylaxis has been necessary in refractory cases.

A life-threatening problem common to many cancer patients is the loss of the reticuloendothelial capacity to clear microorganisms after splenectomy, which may be performed as part of the management of hairy cell leukemia, chronic lymphocytic leukemia (CLL), and chronic myelocytic leukemia (CML) and in Hodgkin's disease. Even after curative therapy for the underlying disease, the lack of a spleen predisposes such patients to rapidly fatal infections. The loss of the spleen through trauma similarly predisposes the normal host to overwhelming infection throughout life. The splenectomized patient should be counseled about the risks of infection with certain organisms, such as the protozoan *Babesia* (Chap. 211) and *Capnocytophaga canimorsus*, a bacterium carried in the mouths of animals (Chaps. 146 and **e24**). Since encapsulated bacteria (*Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*) are the organisms most commonly associated with postsplenectomy sepsis, splenectomized persons should be vaccinated (and revaccinated; Table 86–2 and Chap. 122) against the capsular polysaccharides of these organisms. Many clinicians recommend giving splenectomized patients a small supply of antibiotics effective against *S. pneumoniae*, *N. meningitidis*, and *H. influenzae* to avert rapid, overwhelming sepsis in the event that they cannot present for medical attention immediately after the onset of fever or other signs or symptoms of bacterial infection. A few amoxicillin/clavulanic acid tablets are a reasonable choice for this purpose.

Table 86–2 Vaccination of Cancer Patients Receiving Chemotherapy[a]

	Use in Indicated Patients		
Vaccine	**Intensive Chemotherapy**	**Hodgkin's Disease**	**Hematopoietic Stem Cell Transplantation**
Diphtheria-tetanus[b]	Primary series and boosters as necessary	No special recommendation	3 doses given 6–12 months after transplantation
Poliomyelitis[c]	Complete primary series and boosters	No special recommendation	3 doses given 6–12 months after transplantation
Haemophilus influenzae type b conjugate	Primary series and booster for children	Immunization before treatment and booster 3 months afterward	3 doses given 6–12 months after transplantation
Human papillomavirus	3 doses for girls and women through 26 years of age	3 doses for girls and women through 26 years of age	3 doses for girls and women through 26 years of age
Hepatitis A	As indicated for normal hosts based on occupation and lifestyle	As indicated for normal hosts based on occupation and lifestyle	As indicated for normal hosts based on occupation and lifestyle
Hepatitis B	Same as for normal	As indicated for normal	3 doses given 6–12

	hosts	hosts based on occupation and lifestyle	months after transplantation
23-Valent pneumococcal polysaccharide[d]	Every 5 years	Immunization before treatment and booster 3 months afterward	1 or 2 doses given 6–12 months after transplantation
4-Valent meningococcal vaccine[e]	Should be administered to splenectomized patients and patients living in endemic areas, including college students in dormitories	Should be administered to splenectomized patients and patients living in endemic areas, including college students in dormitories	Should be administered to splenectomized patients and patients living in endemic areas, including college students in dormitories
Influenza	Seasonal immunization	Seasonal immunization	Seasonal immunization
Measles/mumps/rubella	Contraindicated	Contraindicated during chemotherapy	After 24 months in patients without graft-versus-host disease
Varicella-zoster virus[f]	Contraindicated[g]	Contraindicated	Contraindicated

[a]The latest recommendations by the Advisory Committee on Immunization Practices and the CDC guidelines can be found at http://www.cdc.gov/vaccines.

[b] The Td (tetanus-diphtheria) combination was recommended for adults. Pertussis vaccine was not recommended for people >6 years of age in the past. However, recent data indicate that the Tdap (tetanus–diphtheria–acellular pertussis) product is both safe and efficacious in adults. A single Tdap booster is now recommended for adults.

[c]Live-virus vaccine is contraindicated; inactivated vaccine should be used.

[d]The 7- and 13-valent pneumococcal conjugate vaccines are currently recommended for children.

[e]Meningococcal conjugate vaccine (MCV4) is recommended for adults ≤55 years old and meningococcal polysaccharide vaccine (MPSV4) for those ≥56 years old.

[f]Includes both varicella vaccine for children and zoster vaccine for adults.

[g]Contact the manufacturer for more information on use in children with acute lymphocytic leukemia.

The level of suspicion of infections with certain organisms should depend on the type of cancer diagnosed (Table 86–3). Diagnosis of multiple myeloma or CLL should alert the clinician to the possibility of hypogammaglobulinemia. While immunoglobulin replacement therapy can be effective, in most cases prophylactic antibiotics are a cheaper, more convenient method of eliminating bacterial infections in CLL patients with hypogammaglobulinemia. Patients with acute lymphocytic leukemia (ALL), patients with non-Hodgkin's lymphoma, and all cancer patients treated with high-dose glucocorticoids (or glucocorticoid-containing chemotherapy regimens) should receive antibiotic prophylaxis for *Pneumocystis* infection (Table 86–3) for the duration of their chemotherapy. In addition to exhibiting susceptibility to certain infectious organisms, patients with cancer are likely to manifest their infections in characteristic ways. For example, fever—generally a sign of infection in normal hosts—continues to be a reliable indicator in neutropenic patients. In contrast, patients receiving glucocorticoids and agents that impair T cell function and cytokine secretion may have serious infections in the absence of fever. Similarly, neutropenic patients commonly present with cellulitis without purulence and with pneumonia without sputum or even x-ray findings (see below).

Table 86–3 Infections Associated with Specific Types of Cancer

Cancer	Underlying Immune Abnormality	Organisms Causing Infection
Multiple myeloma	Hypogammaglobulinemia	*Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis*
Chronic lymphocytic leukemia	Hypogammaglobulinemia	*S. pneumoniae, H. influenzae, N. meningitidis*
Acute myelocytic or lymphocytic leukemia	Granulocytopenia, skin and mucous-membrane lesions	Extracellular gram-positive and gram-negative bacteria, fungi
Hodgkin's disease	Abnormal T cell function	Intracellular pathogens *(Mycobacterium*

		tuberculosis, Listeria, Salmonella, Cryptococcus, Mycobacterium avium)
Non-Hodgkin's lymphoma and acute lymphocytic leukemia	Glucocorticoid chemotherapy, T and B cell dysfunction	*Pneumocystis*
Colon and rectal tumors	Local abnormalities[a]	*Streptococcus bovis* (bacteremia)
Hairy cell leukemia	Abnormal T cell function	Intracellular pathogens (*M. tuberculosis, Listeria, Cryptococcus, M. avium*)

[a]The reason for this association is not well defined.

The use of monoclonal antibodies that target B and T cells as well as drugs that interfere with lymphocyte signal transduction events is associated with reactivation of latent infections. The use of rituximab, the antibody to CD20 (a B–cell surface protein), is associated with the development of reactivation tuberculosis as well as hepatitis B, cytomegalovirus (CMV) infection, and other latent infections. Like organ transplant recipients (Chap. 132), patients with positive purified protein derivative tests and underlying viral infection should be carefully monitored for reactivation disease.

SYSTEM-SPECIFIC SYNDROMES

Skin-Specific Syndromes

Skin lesions are common in cancer patients, and the appearance of these lesions may permit the diagnosis of systemic bacterial or fungal infection. While cellulitis caused by skin organisms such as *Streptococcus* or *Staphylococcus* is common, neutropenic patients—i.e., those with <500 functional polymorphonuclear leukocytes (PMNs)/μL—and patients with impaired blood or lymphatic drainage may develop infections with unusual organisms. Innocent-looking macules or papules may be the first sign of bacterial or fungal sepsis in immunocompromised patients (Fig. 86-1). In the neutropenic host, a macule progresses rapidly to ecthyma gangrenosum (Fig. e7-35), a usually painless, round, necrotic lesion consisting of a central black or gray-black eschar with surrounding erythema. Ecthyma gangrenosum, which is located in nonpressure areas (as distinguished from necrotic lesions associated with lack of circulation), is often associated with *Pseudomonas aeruginosa* bacteremia (Chap. 152) but may be caused by other bacteria.

Figure 86-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. **Papules related to *Escherichia coli*** bacteremia in a neutropenic patient with acute lymphocytic leukemia. ***B.*** The same lesion the following day.

Candidemia (Chap. 203) is also associated with a variety of skin conditions (Fig. e7-38) and commonly presents as a maculopapular rash. Punch biopsy of the skin may be the best method for diagnosis.

Cellulitis, an acute spreading inflammation of the skin, is most often caused by infection with group A *Streptococcus* or *Staphylococcus aureus*, virulent organisms normally found on the skin (Chap. 125). Although cellulitis tends to be circumscribed in normal hosts, it may spread rapidly in neutropenic patients. A tiny break in the skin may lead to spreading cellulitis, which is characterized by pain and erythema; in the affected patients, signs of infection (e.g., purulence) are often lacking. What might be a furuncle in a normal host may require amputation because of uncontrolled infection in a patient presenting with leukemia. A dramatic response to an infection that might be trivial in a normal host can mark the first sign of leukemia. Fortunately, granulocytopenic patients are likely to be infected with certain types of organisms (Table 86–4); thus the selection of an antibiotic regimen is somewhat easier than it might otherwise be (see "Antibacterial Therapy," below). It is essential to recognize cellulitis early and to treat it aggressively. Patients who are neutropenic or have previously received antibiotics for other reasons may develop cellulitis with unusual organisms (e.g., *Escherichia coli*, *Pseudomonas*, or fungi). Early treatment, even of innocent-looking lesions, is essential to prevent necrosis and loss of tissue. Debridement to prevent spread may sometimes be necessary early in the course of disease, but it can often be performed after chemotherapy, when the PMN count increases.

Table 86–4 Organisms Likely to Cause Infections in Granulocytopenic Patients

Gram-positive cocci	*Enterobacter* spp.
Staphylococcus epidermidis *Staphylococcus aureus*	*Serratia* spp.
Viridans *Streptococcus*	*Acinetobacter* spp.[a]
Enterococcus faecalis	*Citrobacter* spp.
Streptococcus pneumoniae	Gram-positive bacilli
Gram-negative bacilli	Diphtheroids
Escherichia coli	JK bacillus[a]
Klebsiella spp.	Fungi
Pseudomonas aeruginosa	*Candida* spp.
Non-*aeruginosa Pseudomonas* spp.[a]	*Aspergillus* spp.

[a]Often associated with intravenous catheters.

Sweet's syndrome, or *febrile neutrophilic dermatosis*, was originally described in women with elevated white blood cell (WBC) counts. The disease is characterized by the presence of leukocytes in the lower dermis, with edema of the papillary body. Ironically, this disease now is usually seen in neutropenic patients with cancer, most often in association with acute leukemia but also in association

with a variety of other malignancies. Sweet's syndrome usually presents as red or bluish-red papules or nodules that may coalesce and form sharply bordered plaques (Fig. e7-41). The edema may suggest vesicles, but on palpation the lesions are solid, and vesicles probably never arise in this disease. The lesions are most common on the face, neck, and arms. On the legs, they may be confused with erythema nodosum (Fig. e7-40). The development of lesions is often accompanied by high fevers and an elevated erythrocyte sedimentation rate. Both the lesions and the temperature elevation respond dramatically to glucocorticoid administration. Treatment begins with high doses of glucocorticoids (60 mg/d of prednisone) followed by tapered doses over the next 2–3 weeks.

Data indicate that *erythema multiforme* (Fig. e7-25) with mucous membrane involvement is often associated with herpes simplex virus (HSV) infection and is distinct from Stevens-Johnson syndrome, which is associated with drugs and tends to have a more widespread distribution. Since cancer patients are both immunosuppressed [and therefore susceptible to herpes infections) and heavily treated with drugs and therefore subject to Stevens-Johnson syndrome (Fig. e12-4)], both of these conditions are common in this population.

Cytokines, which are used as adjuvants or primary treatments for cancer, can themselves cause characteristic rashes, further complicating the differential diagnosis. This phenomenon is a particular problem in bone marrow transplant recipients (Chap. 132), who, in addition to having the usual chemotherapy-, antibiotic-, and cytokine-induced rashes, are plagued by graft-versus-host disease.

Catheter-Related Infections

Because IV catheters are commonly used in cancer chemotherapy and are prone to infection (Chap. 131), they pose a major problem in the care of patients with cancer. Some catheter-associated infections can be treated with antibiotics, while in others the catheter must be removed (Table 86–5). If the patient has a "tunneled" catheter (which consists of an entrance site, a subcutaneous tunnel, and an exit site), a red streak over the subcutaneous part of the line (the tunnel) is grounds for immediate device removal. Failure to remove catheters under these circumstances may result in extensive cellulitis and tissue necrosis.

Table 86–5 Approach to Catheter Infections in Immunocompromised Patients

Clinical Presentation	Catheter Removal	Antibiotics	Comments
Evidence of Infection, Negative Blood Cultures			
Exit-site erythema	Not necessary if infection responds to treatment	Usually begin treatment for gram-positive cocci.	Coagulase-negative staphylococci are most common.
Tunnel-site erythema	Required	Treat for gram-positive cocci pending culture results.	Failure to remove the catheter may lead to complications.
Blood Culture–Positive Infections			
Coagulase-negative staphylococci	Line removal optimal but may be unnecessary if patient is clinically stable and responds to antibiotics	Usually start with vancomycin. (Linezolid, quinupristin/dalfopristin, and daptomycin are all appropriate.)	If there are no contraindications to line removal, this course of action is optimal. If the line is removed, antibiotics may not be necessary.
Other gram-positive cocci (e.g., *Staphylococcus aureus, Enterococcus*); gram-positive rods (*Bacillus, Corynebacterium* spp.)	Recommended	Treat with antibiotics to which the organism is sensitive, with duration based on the clinical setting.	The incidence of metastatic infections following *S.aureus* infection and the difficulty of treating enterococcal infection make line removal the recommended course of action. In addition, gram-positive rods do not respond readily to antibiotics alone.
Gram-negative bacteria	Recommended	Use an agent to which the organism is shown to be sensitive.	Organisms like *Stenotrophomonas, Pseudomonas,* and *Burkholderia* are notoriously hard to treat.
Fungi	Recommended	—	Fungal infections of catheters are extremely difficult to treat.

More common than tunnel infections are exit-site infections, often with erythema around the area where the line penetrates the skin. Most authorities (Chap. 135) recommend treatment (usually with vancomycin) for an exit-site infection caused by coagulase-negative *Staphylococcus*. Treatment of

coagulase-positive staphylococcal infection is associated with a poorer outcome, and it is advisable to remove the catheter if possible. Similarly, many clinicians remove catheters associated with infections due to *P. aeruginosa* and *Candida* species, since such infections are difficult to treat and bloodstream infections with these organisms are likely to be deadly. Catheter infections caused by *Burkholderia cepacia*, *Stenotrophomonas* spp., *Agrobacterium* spp., and *Acinetobacter baumannii* as well as *Pseudomonas* spp. other than *aeruginosa* are likely to be very difficult to eradicate with antibiotics alone. Similarly, isolation of *Bacillus*, *Corynebacterium*, and *Mycobacterium* spp. should prompt removal of the catheter.

Gastrointestinal Tract–Specific Syndromes

UPPER GASTROINTESTINAL TRACT DISEASE

Infections of the Mouth

The oral cavity is rich in aerobic and anaerobic bacteria (Chap. 164) that normally live in a commensal relationship with the host. The antimetabolic effects of chemotherapy cause a breakdown of host defenses, leading to ulceration of the mouth and the potential for invasion by resident bacteria. Mouth ulcerations afflict most patients receiving cytotoxic chemotherapy and have been associated with viridans streptococcal bacteremia. *Candida* infections of the mouth are very common. Fluconazole is clearly effective in the treatment of both local infections (thrush) and systemic infections (esophagitis) due to *Candida albicans*. Other azoles (e.g., voriconazole) as well as echinocandins offer similar efficacy as well as activity against the fluconazole-resistant organisms that are associated with extensive fluconazole treatment (Chap. 203).

Noma (*cancrum oris*), commonly seen in malnourished children, is a penetrating disease of the soft and hard tissues of the mouth and adjacent sites, with resulting necrosis and gangrene. It has a counterpart in immunocompromised patients and is thought to be due to invasion of the tissues by *Bacteroides*, *Fusobacterium*, and other normal inhabitants of the mouth. Noma is associated with debility, poor oral hygiene, and immunosuppression.

Viruses, particularly HSV, are a prominent cause of morbidity in immunocompromised patients, in whom they are associated with severe mucositis. The use of acyclovir, either prophylactically or therapeutically, is of value.

Esophageal Infections

The differential diagnosis of esophagitis (usually presenting as substernal chest pain upon swallowing) includes herpes simplex and candidiasis, both of which are readily treatable.

LOWER GASTROINTESTINAL TRACT DISEASE

Hepatic candidiasis (Chap. 203) results from seeding of the liver (usually from a gastrointestinal source) in neutropenic patients. It is most common among patients being treated for acute leukemia and usually presents symptomatically around the time the neutropenia resolves. The characteristic picture is that of persistent fever unresponsive to antibiotics, abdominal pain and tenderness or nausea, and elevated serum levels of alkaline phosphatase in a patient with hematologicmalignancy who has recently recovered from neutropenia. The diagnosis of this disease (which may present in an indolent manner and persist for several months) is based on the finding of yeasts or pseudohyphae in granulomatous lesions. Hepatic ultrasound or CT may reveal bull's-eye lesions. In some cases, MRI reveals small lesions not visible by other imaging modalities. The pathology (a granulomatous response) and the timing (with resolution of neutropenia and an elevation in granulocyte count) suggest that the host response to *Candida* is an important component of the manifestations of disease. In many cases, although organisms are visible, cultures of biopsied material may be negative. The designation *hepatosplenic candidiasis* or *hepatic candidiasis* is a misnomer because the disease often involves the kidneys and other tissues; the term *chronic disseminated candidiasis* may be more appropriate. Because of the risk of bleeding with liver biopsy, diagnosis is often based on imaging studies (MRI, CT). Treatment should be directed to the causative agent (usually *C.albicans* but sometimes *C. tropicalis* or other less common *Candida* spp).

TYPHLITIS

Typhlitis (also referred to as necrotizing colitis, neutropenic colitis, necrotizing enteropathy, ileocecal syndrome, and cecitis) is a clinical syndrome of fever and right-lower-quadrant tenderness in an immunosuppressed host. This syndrome is classically seen in neutropenic patients after chemotherapy with cytotoxic drugs. It may be more common among children than among adults and appears to be much more common among patients with acute myelocytic leukemia (AML) or ALL than among those with other types of cancer; a similar syndrome has been reported in patients infected with HIV type 1. Physical examination reveals right-lower-quadrant tenderness, with or without rebound tenderness. Associated diarrhea (often bloody) is common, and the diagnosis can be confirmed by the finding of a thickened cecal wall on CT, MRI, or ultrasonography. Plain films may reveal a right-lower-quadrant mass, but CT with contrast or MRI is a much more sensitive means of diagnosis. Although surgery is sometimes attempted to avoid perforation from ischemia, most cases resolve with medical therapy

alone. The disease is sometimes associated with positive blood cultures (which usually yield aerobic gram-negative bacilli), and therapy is recommended for a broad spectrum of bacteria (particularly gram-negative bacilli, which are likely to be found in the bowel flora). Surgery is indicated in the case of perforation.

CLOSTRIDIUM DIFFICILE–INDUCED DIARRHEA

Patients with cancer are predisposed to the development of *C. difficile* diarrhea (Chap. 129) as a consequence of chemotherapy alone. Thus, they may have positive toxin tests before receiving antibiotics. Obviously, such patients are also subject to *C. difficile*–induced diarrhea as a result of antibiotic pressure. *C. difficile* should always be considered as a possible cause of diarrhea in cancer patients who have received antibiotics.

Central Nervous System–Specific Syndromes

MENINGITIS

The presentation of meningitis in patients with lymphoma or CLL, patients receiving chemotherapy (particularly with glucocorticoids) for solid tumors, and patients who have received bone marrow transplants suggests a diagnosis of cryptococcal or listerial infection. As noted previously, splenectomized patients are susceptible to rapid, overwhelming infection with encapsulated bacteria (including *S. pneumoniae*, *H. influenzae*, and *N. meningitidis*). Similarly, patients who are antibody-deficient (e.g., those with CLL, those who have received intensive chemotherapy, or those who have undergone bone marrow transplantation) are likely to have infections caused by these bacteria. Other cancer patients, however, because of their defective cellular immunity, are likely to be infected with other pathogens (Table 86–3).

ENCEPHALITIS

The spectrum of disease resulting from viral encephalitis is expanded in immunocompromised patients. A predisposition to infections with intracellular organisms similar to those encountered in patients with AIDS (Chap. 189) is seen in cancer patients receiving (1) high-dose cytotoxic chemotherapy, (2) chemotherapy affecting T cell function (e.g., fludarabine), or (3) antibodies that eliminate T cells (e.g., anti-CD3, alemtuzumab, anti-CD52) or cytokine activity (anti-tumor necrosis factor agents or interleukin 1 receptor antagonists). Infection with varicella-zoster virus (VZV) has been associated with encephalitis that may be caused by VZV-related vasculitis. Chronic viral infections may also be associated with dementia and encephalitic presentations, and a diagnosis of progressive multifocal leukoencephalopathy (Chap. 381) should be considered when a patient who has received chemotherapy presents with dementia (Table 86–6). Other abnormalities of the central nervous system (CNS) that may be confused with infection include normal-pressure hydrocephalus and vasculitis resulting from CNS irradiation. It may be possible to differentiate these conditions by MRI.

Table 86–6 Differential Diagnosis of Central Nervous System Infections in Patients with Cancer

Findings on CT or MRI	Underlying Predisposition	
	Prolonged Neutropenia	Defects in Cellular Immunity[a]
Mass lesions	*Aspergillus, Nocardia,* or *Cryptococcus* brain abscess	Toxoplasmosis EBV-LPD
Diffuse encephalitis	PML (JC virus)	Infection with VZV, CMV, HSV, HHV-6, JC virus (PML), *Listeria*

[a]High-dose glucocorticoid therapy, cytotoxic chemotherapy. *Abbreviations:* CMV, cytomegalovirus; EBV-LPD, Epstein-Barr virus lymphoproliferative disease; HHV-6, human herpesvirus type 6; HSV, herpes simplex virus; PML, progressive multifocal leukoencephalopathy; VZV, varicella-zoster virus.

BRAIN MASSES

Mass lesions of the brain most often present as headache with or without fever or neurologic abnormalities. Infections associated with mass lesions may be caused by bacteria (particularly *Nocardia*), fungi (particularly *Cryptococcus* or *Aspergillus*), or parasites (*Toxoplasma*). Epstein-Barr virus (EBV)–associated lymphoproliferative disease may also present as single or multiple mass lesions of the brain. A biopsy may be required for a definitive diagnosis.

Pulmonary Infections

Pneumonia (Chap. 257) in immunocompromised patients may be difficult to diagnose because conventional methods of diagnosis depend on the presence of neutrophils. Bacterial pneumonia in

neutropenic patients may present without purulent sputum—or, in fact, without any sputum at all—and may not produce physical findings suggestive of chest consolidation (rales or egophony).

In granulocytopenic patients with persistent or recurrent fever, the chest x-ray pattern may help to localize an infection and thus to determine which investigative tests and procedures should be undertaken and which therapeutic options should be considered (Table 86–7). In this setting, a simple chest x-ray is a screening tool; because the impaired host response results in less evidence of consolidation or infiltration, high-resolution CT is recommended for the diagnosis of pulmonary infections. The difficulties encountered in the management of pulmonary infiltrates relate in part to the difficulties of performing diagnostic procedures on the patients involved. When platelet counts can be increased to adequate levels by transfusion, microscopic and microbiologic evaluation of the fluid obtained by endoscopic bronchial lavage is often diagnostic. Lavage fluid should be cultured for *Mycoplasma*, *Chlamydia*, *Legionella*, *Nocardia*, more common bacterial pathogens, and fungi. In addition, the possibility of *Pneumocystis* pneumonia should be considered, especially in patients with ALL or lymphoma who have not received prophylactic trimethoprim-sulfamethoxazole (TMP-SMX). The characteristics of the infiltrate may be helpful in decisions about further diagnostic and therapeutic maneuvers. Nodular infiltrates suggest fungal pneumonia (e.g., that caused by *Aspergillus* or *Mucor*). Such lesions may best be approached by visualized biopsy procedures.

Table 86–7 Differential Diagnosis of Chest Infiltrates in Immunocompromised Patients

	Cause of Pneumonia	
Infiltrate	**Infectious**	**Noninfectious**
Localized	Bacteria (including *Legionella,* mycobacteria)	Local hemorrhage or embolism, tumor
Nodular	Fungi (e.g., *Aspergillus* or *Mucor), Nocardia*	Recurrent tumor
Diffuse	Viruses (especially CMV), *Chlamydia, Pneumocystis, Toxoplasma gondii,* mycobacteria	Congestive heart failure, radiation pneumonitis, drug-induced lung injury, diffuse alveolar hemorrhage (described after BMT)

Abbreviations: BMT, bone marrow transplantation; CMV, cytomegalovirus.

Aspergillus species (Chap. 204) can colonize the skin and respiratory tract or cause fatal systemic illness. Although this fungus may cause aspergillomas in a previously existing cavity or may produce allergic bronchopulmonary disease, the major problem posed by this genus in neutropenic patients is invasive disease due to *A. fumigatus* or *A. flavus*. The organisms enter the host following colonization of the respiratory tract, with subsequent invasion of blood vessels. The disease is likely to present as a thrombotic or embolic event because of this ability of the fungi to invade blood vessels. The risk of infection with *Aspergillus* correlates directly with the duration of neutropenia. In prolonged neutropenia, positive surveillance cultures for nasopharyngeal colonization with *Aspergillus* may predict the development of disease.

Patients with *Aspergillus* infection often present with pleuritic chest pain and fever, which are sometimes accompanied by cough. Hemoptysis may be an ominous sign. Chest x-rays may reveal new focal infiltrates or nodules. Chest CT may reveal a characteristic halo consisting of a mass-like infiltrate surrounded by an area of low attenuation. The presence of a "crescent sign" on chest x-ray or chest CT, in which the mass progresses to central cavitation, is characteristic of invasive *Aspergillus* infection but may develop as the lesions are resolving.

In addition to causing pulmonary disease, *Aspergillus* may invade through the nose or palate, with deep sinus penetration. The appearance of a discolored area in the nasal passages or on the hard palate should prompt a search for invasive *Aspergillus*. This situation is likely to require surgical debridement. Catheter infections with *Aspergillus* usually require both removal of the catheter and antifungal therapy.

Diffuse interstitial infiltrates suggest viral, parasitic, or *Pneumocystis* pneumonia. If the patient has a diffuse interstitial pattern on chest x-ray, it may be reasonable, while considering invasive diagnostic procedures, to institute empirical treatment for *Pneumocystis* with TMP-SMX and for *Chlamydia*, *Mycoplasma*, and *Legionella* with a quinolone or an erythromycin derivative (e.g., azithromycin). Noninvasive procedures, such as staining of sputum smears for *Pneumocystis*, serum cryptococcal antigen tests, and urine testing for *Legionella* antigen, may be helpful. Serum galactomannan and ß-d-glucan tests may be helpful in diagnosing *Aspergillus* infection, but their utility is limited by their lack of sensitivity. In transplant recipients who are seropositive for CMV, a determination of CMV load in the serum should be considered. Viral load studies (which allow physicians to quantitate viruses) have superseded simple measurement of serum IgG, which merely documents prior exposure to virus. Infections with viruses that cause only upper respiratory symptoms in immunocompetent hosts, such as respiratory syncytial virus (RSV), influenza viruses, and parainfluenza viruses, may be associated with

fatal pneumonitis in immunocompromised hosts. Polymerase chain reaction testing now allows rapid diagnosis of viral pneumonia, which can lead to treatment in some cases (e.g., influenza).

Bleomycin is the most common cause of chemotherapy-induced lung disease. Other causes include alkylating agents (such as cyclophosphamide, chlorambucil, and melphalan), nitrosoureas [carmustine (BCNU), lomustine (CCNU), and methyl-CCNU], busulfan, procarbazine, methotrexate, and hydroxyurea. Both infectious and noninfectious (drug- and/or radiation-induced) pneumonitis can cause fever and abnormalities on chest x-ray; thus, the differential diagnosis of an infiltrate in a patient receiving chemotherapy encompasses a broad range of conditions (Table 86–7). The treatment of radiation pneumonitis (which may respond dramatically to glucocorticoids) or drug-induced pneumonitis is different from that of infectious pneumonia, and a biopsy may be important in the diagnosis. Unfortunately, no definitive diagnosis can be made in ~30% of cases, even after bronchoscopy.

Open-lung biopsy is the gold standard of diagnostic techniques. Biopsy via a visualized thoracostomy can replace an open procedure in many cases. When a biopsy cannot be performed, empirical treatment can be undertaken; a quinolone or an erythromycin derivative (azithromycin) and TMP-SMX are used in the case of diffuse infiltrates, and an antifungal agent is administered in the case of nodular infiltrates. The risks should be weighed carefully in these cases. If inappropriate drugs are administered, empirical treatment may prove toxic or ineffective; either of these outcomes may be riskier than biopsy.

Cardiovascular Infections

Patients with Hodgkin's disease are prone to persistent infections by *Salmonella*, sometimes (and particularly often in elderly patients) affecting a vascular site. The use of IV catheters deliberately lodged in the right atrium is associated with a high incidence of bacterial endocarditis, presumably related to valve damage followed by bacteremia. Nonbacterial thrombotic endocarditis has been described in association with a variety of malignancies (most often solid tumors) and may follow bone marrow transplantation as well. The presentation of an embolic event with a new cardiac murmur suggests this diagnosis. Blood cultures are negative in this disease of unknown pathogenesis.

Endocrine Syndromes

Infections of the endocrine system have been described in immunocompromised patients. *Candida* infection of the thyroid may be difficult to diagnose during the neutropenic period. It can be defined by indium-labeled WBC scans or gallium scans after neutrophil counts increase. CMV infection can cause adrenalitis with or without resulting adrenal insufficiency. The presentation of a sudden endocrine anomaly in an immunocompromised patient may be a sign of infection in the involved end organ.

Musculoskeletal Infections

Infection that is a consequence of vascular compromise, resulting in gangrene, can occur when a tumor restricts the blood supply to muscles, bones, or joints. The process of diagnosis and treatment of such infection is similar to that in normal hosts, with the following caveats:

1. *In terms of diagnosis* , a lack of physical findings resulting from a lack of granulocytes in the granulocytopenic patient should make the clinician more aggressive in obtaining tissue rather than relying on physical signs.
2. *In terms of therapy* , aggressive debridement of infected tissues may be required, but it is usually difficult to operate on patients who have recently received chemotherapy, both because of a lack of platelets (which results in bleeding complications) and because of a lack of WBCs (which may lead to secondary infection). A blood culture positive for *Clostridium perfringens*—an organism commonly associated with gas gangrene—can have a number of meanings (Chap. 142). Bloodstream infections with intestinal organisms such as *Streptococcus bovis* and *C. perfringens* may arise spontaneously from lower gastrointestinal lesions (tumor or polyps); alternatively, these lesions may be harbingers of invasive disease. The clinical setting must be considered in order to define the appropriate treatment for each case.

Renal and Ureteral Infections

Infections of the urinary tract are common among patients whose ureteral excretion is compromised (Table 86–1). *Candida*, which has a predilection for the kidney, can invade either from the bloodstream or in a retrograde manner (via the ureters or bladder) in immunocompromised patients. The presence of "fungus balls" or persistent candiduria suggests invasive disease. Persistent funguria (with *Aspergillus* as well as *Candida*) should prompt a search for a nidus of infection in the kidney.

Certain viruses are typically seen only in immunosuppressed patients. BK virus (polyomavirus hominis 1) has been documented in the urine of bone marrow transplant recipients and, like adenovirus, may be associated with hemorrhagic cystitis. BK-induced cystitis usually remits with decreasing immunosuppression. Anecdotal reports have described the treatment of infections due to adenovirus and BK virus with cidofovir.

ABNORMALITIES THAT PREDISPOSE TO INFECTION

(Table 86–1)

The Lymphoid System

It is beyond the scope of this chapter to detail how all the immunologic abnormalities that result from cancer or from chemotherapy for cancer lead to infections. Disorders of the immune system are discussed in other sections of this book. As has been noted, patients with antibody deficiency are predisposed to overwhelming infection with encapsulated bacteria (including *S. pneumoniae*, *H. influenzae*, and *N. meningitidis*). Infections that result from the lack of a functional cellular immune system are described in Chap. 189. It is worth mentioning, however, that patients undergoing intensive chemotherapy for any form of cancer will have not only defects due to granulocytopenia but also lymphocyte dysfunction, which may be profound. Thus, these patients—especially those receiving glucocorticoid-containing regimens or drugs that inhibit either T cell activation (calcineurin inhibitors or drugs like fludarabine, which affect lymphocyte function) or cytokine induction—should be given prophylaxis for *Pneumocystis* pneumonia.

The Hematopoietic System

Initial studies in the 1960s revealed a dramatic increase in the incidence of infections (fatal and nonfatal) among cancer patients with a granulocyte count of <500/μL. The use of prophylactic antibacterial agents has reduced the number of bacterial infections, but 35–78% of febrile neutropenic patients being treated for hematologicmalignancies develop infections at some time during chemotherapy. Aerobic pathogens (both gram-positive and gram-negative) predominate in all series, but the exact organisms isolated vary from center to center. Infections with anaerobic organisms are uncommon. Geographic patterns affect the types of fungi isolated. Tuberculosis and malaria are common causes of fever in the developing world and may present in this setting as well.

Neutropenic patients are unusually susceptible to infection with a wide variety of bacteria; thus, antibiotic therapy should be initiated promptly to cover likely pathogens if infection is suspected. Indeed, early initiation of antibacterial agents is mandatory to prevent deaths. Like most immunocompromised patients, neutropenic patients are threatened by their own microbial flora, including gram-positive and gram-negative organisms found commonly on the skin and in the bowel (Table 86–4). Because treatment with narrow-spectrum agents leads to infection with organisms not covered by the antibiotics used, the initial regimen should target all pathogens likely to be initial causes of bacterial infection in neutropenic hosts. As noted in the algorithm shown in Fig. 86-2, administration of antimicrobial agents is routinely continued until neutropenia resolves—i.e., the granulocyte count is sustained above 500/μL for at least 2 days. In some cases, patients remain febrile after resolution of neutropenia. In these instances, the risk of sudden death from overwhelming bacteremia is greatly reduced, and the following diagnoses should be seriously considered: (1) fungal infection, (2) bacterial abscesses or undrained foci of infection, and (3) drug fever (including reactions to antimicrobial agents as well as to chemotherapy or cytokines). In the proper setting, viral infection or graft-versus-host disease should be considered. In clinical practice, antibacterial therapy is usually discontinued when the patient is no longer neutropenic and all evidence of bacterial disease has been eliminated. Antifungal agents are then discontinued if there is no evidence of fungal disease. If the patient remains febrile, a search for viral diseases or unusual pathogens is conducted while unnecessary cytokines and other drugs are systematically eliminated from the regimen.

Figure 86-2

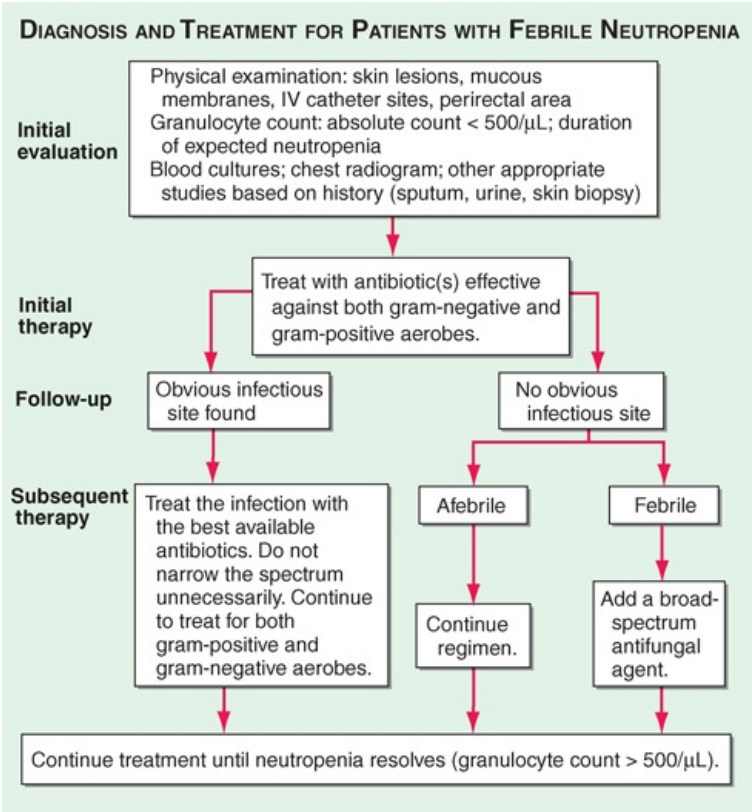

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the diagnosis and treatment of febrile -neutropenic patients.

Treatment: Infections in Cancer Patients

ANTIBACTERIAL THERAPY

Hundreds of antibacterial regimens have been tested for use in patients with cancer. The major risk of infection is related to the degree of neutropenia seen as a consequence of either the disease or the therapy. Many of the relevant studies have involved small populations in which the outcomes have generally been good, and most have lacked the statistical power to detect differences among the regimens studied. Each febrile neutropenic patient should be approached as a unique problem, with particular attention given to previous infections and recent antibiotic exposures. Several general guidelines are useful in the initial treatment of neutropenic patients with fever (Fig. 86-2):

1. In the initial regimen, it is necessary to use antibiotics active against both gram-negative and gram-positive bacteria (Table 86–4).
2. Monotherapy with an aminoglycoside or an antibiotic without good activity against gram-positive organisms (e.g., ciprofloxacin or aztreonam) is not adequate in this setting.
3. The agents used should reflect both the epidemiology and the antibiotic resistance pattern of the hospital.
4. If the pattern of resistance justifies its use, a single third-generation cephalosporin constitutes an appropriate initial regimen in many hospitals.
5. Most standard regimens are designed for patients who have not previously received prophylactic antibiotics. The development of fever in a patient who has received antibiotics affects the choice of subsequent therapy, which should target resistant organisms and organisms known to cause infections in patients being treated with the antibiotics already administered.
6. Randomized trials have indicated the safety of oral antibiotic regimens in the treatment of "low-risk" patients with fever and neutropenia. Outpatients who are expected to remain neutropenic for <10 days and who have no concurrent medical problems (such as hypotension, pulmonary compromise, or abdominal pain) can be classified as low risk and treated with a broad-spectrum oral regimen.
7. Several large-scale studies indicate that prophylaxis with a fluoroquinolone (ciprofloxacin or levofloxacin) decreases morbidity and mortality rates among afebrile patients who are anticipated to have neutropenia of long duration.

The initial antibacterial regimen should be refined on the basis of culture results (Fig. 86-2). Blood cultures are the most relevant on which to base therapy; surface cultures of skin and mucous

membranes may be misleading. In the case of gram-positive bacteremia or another gram-positive infection, it is important that the antibiotic be optimal for the organism isolated. Although it is not desirable to leave the patient unprotected, the addition of more and more antibacterial agents to the regimen is not appropriate unless there is a clinical or microbiologic reason to do so. Planned progressive therapy (the serial, empirical addition of one drug after another without culture data) is not efficacious in most settings and may have unfortunate consequences. Simply adding another antibiotic for fear that a gram-negative infection is present is a dubious practice. The synergy exhibited by β-lactams and aminoglycosides against certain gram-negative organisms (especially *P. aeruginosa*) provides the rationale for using two antibiotics in this setting, but recent analyses suggest that efficacy is not enhanced by the addition of aminoglycosides, while toxicity may be increased. Mere "double coverage," with the addition of a quinolone or another antibiotic that is not likely to exhibit synergy, has not been shown to be of benefit and may cause additional toxicities and side effects. Cephalosporins can cause bone marrow suppression, and vancomycin is associated with neutropenia in some healthy individuals. Furthermore, the addition of multiple cephalosporins may induce β-lactamase production by some organisms; cephalosporins and double β-lactam combinations should probably be avoided altogether in *Enterobacter* infections.

ANTIFUNGAL THERAPY

Fungal infections in cancer patients are most often associated with neutropenia. Neutropenic patients are predisposed to the development of invasive fungal infections, most commonly those due to *Candida* and *Aspergillus* species and occasionally those caused by *Fusarium*, *Trichosporon*, and *Bipolaris*. Cryptococcal infection, which is common among patients taking immunosuppressive agents, is uncommon among neutropenic patients receiving chemotherapy for AML. Invasive candidal disease is usually caused by *C. albicans* or *C. tropicalis* but can be caused by *C. krusei*, *C. parapsilosis*, and *C. glabrata*.

For decades it has been common clinical practice to add amphotericin B to antibacterial regimens if a neutropenic patient remains febrile despite 4–7 days of treatment with antibacterial agents. The rationale for this empirical addition is that it is difficult to culture fungi before they cause disseminated disease and that mortality rates from disseminated fungal infections in granulocytopenic patients are high. Before the introduction of newer azoles into clinical practice, amphotericin B was the mainstay of antifungal therapy. The insolubility of amphotericin B has resulted in the marketing of several lipid formulations that are less toxic than the amphotericin B deoxycholate complex. Echinocandins (e.g., caspofungin) are useful in the treatment of infections caused by azole-resistant *Candida* as well as in therapy for aspergillosis and have been shown to be equivalent to liposomal amphotericin B for the empirical treatment of patients with prolonged fever and neutropenia. Newer azoles have also been demonstrated to be effective in this setting. Although fluconazole is efficacious in the treatment of infections due to many *Candida* spp., its use against serious fungal infections in immunocompromised patients is limited by its narrow spectrum: it has no activity against *Aspergillus* or against several non-*albicansCandida* spp. The broad-spectrum azoles (e.g., voriconazole and posaconazole) provide another option for the treatment of *Aspergillus* infection (Chap. 204), including CNS infection, in which amphotericin B has usually failed. Clinicians should be aware that the spectrum of each azole is somewhat different and that no drug can be assumed to be efficacious against all fungi. For example, while voriconazole is active against *Pseudallescheria boydii*, amphotericin B is not; however, voriconazole has no activity against *Mucor*. Posaconazole, which is administered orally, is useful as a prophylactic agent in patients with prolonged neutropenia. Studies in progress are assessing the use of these agents in combinations. **For a full discussion of antifungal therapy, see Chap. 198.**

ANTIVIRAL THERAPY

The availability of a variety of agents active against herpes-group viruses, including some new agents with a broader spectrum of activity, has heightened focus on the treatment of viral infections, which pose a major problem in cancer patients. Viral diseases caused by the herpes group are prominent. Serious (and sometimes fatal) infections due to HSV and CMV are well documented, and VZV infections may be fatal to patients receiving chemotherapy. The roles of human herpesvirus (HHV)-6, HHV-7, and HHV-8 (Kaposi's sarcoma–associated herpesvirus) in cancer patients are still being defined (Chap. 182). While clinical experience is most extensive with acyclovir, which can be used therapeutically or prophylactically, a number of derivative drugs offer advantages over this agent (Table 86–8).

Table 86–8 Antiviral Agents Active Against Herpesviruses

Agent	Description	Spectrum	Toxicity	Other Issues
Acyclovir	Inhibits HSV polymerase	HSV, VZV (± CMV, EBV)	Rarely has side effects; crystalluria can occur at high doses	Long history of safety; original antiviral agent
Famciclovir	Prodrug of	HSV, VZV (±	Associated with	Longer effective half-life

	penciclovir (a guanosine analogue)	CMV)	cancerin rats	than acyclovir
Valacyclovir	Prodrug of acyclovir; better absorption	HSV, VZV (± CMV)	Associated with thrombotic microangiopathy in one study of immunocompromised patients	Better oral absorption and longer effective half-life than acyclovir; can be given as a single daily dose for prophylaxis
Ganciclovir	More potent polymerase inhibitor; more toxic than acyclovir	HSV, VZV, CMV, HHV-6	Bone marrow suppression	Neutropenia may respond to G-CSF or GM-CSF
Valganciclovir	Prodrug of ganciclovir; better absorption	HSV, VZV, CMV, HHV-6	Bone marrow suppression	—
Cidofovir	Nucleotide analogue of cytosine	HSV, VZV, CMV; good in vitro activity against adenovirus and others	Nephrotoxic marrow suppression	Given IV once a week
Foscarnet	Phosphonoformic acid; inhibits viral DNA polymerase	HSV, VZV, CMV, HHV-6	Nephrotoxic; electrolyte abnormalities common	IV only

Abbreviations: ±, agent has some activity but not enough for the treatment of infections; CMV, cytomegalovirus; EBV, Epstein-Barr virus; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; HHV, human herpesvirus; HSV, herpes simplex virus; VZV, varicella-zoster virus.

In addition to the herpes group, several respiratory viruses (especially RSV) may cause serious disease in cancer patients. While influenza vaccination is recommended (see below), it may be ineffective in this patient population. The availability of antiviral drugs with activity against influenza viruses gives the clinician additional options for the treatment of these patients (Table 86–9).

Table 86–9 Other Antiviral Agents Useful in the Treatment of Infections in Cancer Patients

Agent	Description	Spectrum	Toxicity	Other Issues
Amantadine, rimantadine	Interfere with uncoating	Influenza A only	5–10% fewer CNS effects with rimantadine	May be given prophylactically
Zanamivir	Neuraminidase inhibitor	Influenza A and B	Usually well tolerated	Inhalation only
Oseltamivir	Neuraminidase inhibitor	Influenza A and B	Usually well tolerated	PO dosing
Pleconaril	Blocks enterovirus binding and uncoating	90% of enteroviruses, 80% of rhinoviruses	Generally well tolerated	Decreases duration of meningitis; available for compassionate use only
Interferons	Cytokines with broad spectrum of activity	Used locally for warts, systemically for hepatitis	Fever, myalgias, bone marrow suppression	Not shown to be helpful in CMV infection; use limited by toxicity
Ribavirin	Purine analogue (precise mechanism of action unknown)	Broad theoretical spectrum; documented use against RSV, Lassa fever virus, and hepatitis viruses (with interferon)	IV form causes anemia	Given by aerosol for RSV infection (efficacy in doubt); approved for use in children with heart/lung disease; given with interferon for hepatitis C

Abbreviations: CMV, cytomegalovirus; CNS, central nervous system; RSV, respiratory syncytial virus.

OTHER THERAPEUTIC MODALITIES

Another way to address the problems of the febrile neutropenic patient is to replenish the neutrophil

population. Although granulocyte transfusions are effective in the treatment of refractory gram-negative bacteremia, they do not have a documented role in prophylaxis. Because of the expense, the risk of leukoagglutinin reactions (which has probably been decreased by improved cell-separation procedures), and the risk of transmission of CMV from unscreened donors (which has been reduced by the use of filters), granulocyte transfusion is reserved for patients unresponsive to antibiotics. This modality is efficacious for documented gram-negative bacteremia refractory to antibiotics, particularly in situations where granulocyte numbers will be depressed for only a short period. The demonstrated usefulness of granulocyte colony-stimulating factor (G-CSF) in mobilizing neutrophils and advances in preservation techniques may make this option more useful than in the past.

A variety of cytokines, including G-CSF and granulocyte-macrophage colony-stimulating factor (GM-CSF), enhance granulocyte recovery after chemotherapy and consequently shorten the period of maximal vulnerability to fatal infections. Interferon γ has been demonstrated to be effective in some infections caused by intracellular organisms, presumably because of its ability to activate macrophages. The role of these cytokines in routine practice is still a matter of some debate. Most authorities recommend their use only when neutropenia is both severe and prolonged. The cytokines themselves may have adverse effects, including fever, hypoxemia, and pleural effusions or serositis in other areas (Chap. 314).

Once neutropenia has resolved, the risk of infection decreases dramatically. However, depending on what drugs they receive, patients who continue on chemotherapeutic protocols remain at high risk for certain diseases. Any patient receiving more than a maintenance dose of glucocorticoids (including many treatment regimens for diffuse lymphoma) should also receive prophylactic TMP-SMX because of the risk of *Pneumocystis* infection; those with ALL should receive such prophylaxis for the duration of chemotherapy.

PREVENTION OF INFECTION IN CANCER PATIENTS

Effect of the Environment

Outbreaks of fatal *Aspergillus* infection have been associated with construction projects and materials in several hospitals. The association between spore counts and risk of infection suggests the need for a high-efficiency air-handling system in hospitals that care for large numbers of neutropenic patients. The use of laminar-flow rooms and prophylactic antibiotics has decreased the number of infectious episodes in severely neutropenic patients. However, because of the expense of such a program and the failure to show that it dramatically affects mortality rates, most centers do not routinely use laminar flow to care for neutropenic patients. Some centers use "reverse isolation," in which health care providers and visitors to a patient who is neutropenic wear gowns and gloves. Since most of the infections these patients develop are due to organisms that colonize the patients' own skin and bowel, the validity of such schemes is dubious, and limited clinical data do not support their use. Hand washing by all staff caring for neutropenic patients should be required to prevent the spread of resistant organisms.

The presence of large numbers of bacteria (particularly *P. aeruginosa*) in certain foods, especially fresh vegetables, has led some authorities to recommend a special "low-bacteria" diet. A diet consisting of cooked and canned food is satisfactory to most neutropenic patients and does not involve elaborate disinfection or sterilization protocols. However, there are no studies to support even this type of dietary restriction. Counseling of patients to avoid leftovers, deli foods, and unpasteurized dairy products is recommended.

Physical Measures

Although few studies address this issue, patients with cancer are predisposed to infections resulting from anatomic compromise (e.g., lymphedema resulting from node dissections after radical mastectomy). Surgeons who specialize in cancer surgery can provide specific guidelines for the care of such patients, and patients benefit from commonsense advice about how to prevent infections in vulnerable areas.

Immunoglobulin Replacement

Many patients with multiple myeloma or CLL have immunoglobulin deficiencies as a result of their disease, and all allogeneic bone marrow transplant recipients are hypogammaglobulinemic for a period after transplantation. However, current recommendations reserve intravenous immunoglobulin replacement therapy for those patients with severe (<400 mg/dL), prolonged hypogammaglobulinemia. Antibiotic prophylaxis has been shown to be cheaper and is efficacious in preventing infections in most CLL patients with hypogammaglobulinemia. Routine use of immunoglobulin replacement is not recommended.

Sexual Practices

The use of condoms is recommended for severely immunocompromised patients. Any sexual practice that results in oral exposure to feces is not recommended. Neutropenic patients should be advised to avoid any practice that results in trauma, as even microscopic cuts may result in bacterial invasion and

fatal sepsis.

Antibiotic Prophylaxis

Several studies indicate that the use of oral fluoroquinolones prevents infection and decreases mortality rates among severely neutropenic patients. Fluconazole prevents *Candida* infections when given prophylactically to patients receiving bone marrow transplants. The use of broader-spectrum antifungal agents (e.g., posaconazole) appears to be more efficacious. Prophylaxis for *Pneumocystis* is mandatory for patients with ALL and for all cancer patients receiving glucocorticoid-containing chemotherapy regimens.

Vaccination of Cancer Patients

In general, patients undergoing chemotherapy respond less well to vaccines than do normal hosts. Their greater need for vaccines thus leads to a dilemma in their management. Purified proteins and inactivated vaccines are almost never contraindicated and should be given to patients even during chemotherapy. For example, all adults should receive diphtheria–tetanus toxoid boosters at the indicated times as well as seasonal influenza vaccine. However, if possible, vaccination should not be undertaken concurrent with cytotoxic chemotherapy. If patients are expected to be receiving chemotherapy for several months and vaccination is indicated (e.g., influenza vaccination in the fall), the vaccine should be given midcycle—as far apart in time as possible from the antimetabolic agents that will prevent an immune response. The meningococcal and pneumococcal polysaccharide vaccines should be given to patients before splenectomy, if possible. The *H. influenzae* type b conjugate vaccine should be administered to all splenectomized patients.

In general, live virus (or live bacterial) vaccines should not be given to patients during intensive chemotherapy because of the risk of disseminated infection. Recommendations on vaccination are summarized in Table 86–2.

FURTHER READINGS

Gafter-Gvili A et al: Antibiotic prophylaxis for bacterial infections in afebrile neutropenic patients following chemotherapy. Cochrane Database Syst Rev 4:CD004386, 2009

Gupta A et al: Infections in acute myeloid leukemia: An analysis of 382 febrile episodes. Med Oncol, published online October 15, 2009

Mermel LA et al: Clinical practice guidelines for the diagnosis and management of intravenous catheter –related infection: 2009 update by the Infectious Diseases Society of America. Clin Infect Dis 49:1, 2009[PMID: 19489710]

Paul M et al: Beta-lactam versus beta-lactam–aminoglycoside combination therapy in cancer patients with neutropenia. Cochrane Database Syst Rev 3:CD003038, 2010

Pickering LK et al: Immunization programs for infants, children, adolescents, and adults: Clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 49:817, 2009[PMID: 19659433]

Tomblyn M et al: Guidelines for preventing infectious complications among hematopoietic cell transplantation recipients: A global perspective. Biol Blood Marrow Transplant 15:1143, 2009[PMID: 19747629]

Ullmann AJ et al: Posaconazole or fluconazole for prophylaxis in severe graft-versus-host disease. N Engl J Med 356:335, 2007[PMID: 17251530]

Vento S et al: Lung infections after cancer chemotherapy. Lancet Oncol 9:982, 2008[PMID: 19071255]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MELANOMA

Pigmented lesions are among the most common findings on skin examination. The challenge is to distinguish cutaneous melanomas, which account for the overwhelming majority of deaths resulting from skin cancer, from the remainder, which with rare exceptions are benign. Cutaneous melanoma can occur in adults of all ages, even young individuals, and people of all colors; it is located on the skin, where it is visible; and it has distinct clinical features that make it detectable at a time when complete surgical excision is possible. Examples of malignant and benign pigmented lesions are shown in Fig. 87-1.

Figure 87-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atypical and malignant pigmented lesions. The most common melanoma is superficial spreading melanoma (not pictured). ***A.*** Acral lentiginous melanoma is the most common melanoma in blacks, Asians, and Hispanics and occurs as an enlarging hyperpigmented macule or plaque on the palms and soles. Lateral pigment diffusion is present. ***B.*** Nodular melanoma most commonly manifests as a rapidly growing, often ulcerated or crusted black nodule. ***C.*** Lentigo maligna melanoma occurs on sun-exposed skin as a large, hyperpigmented macule or plaque with -irregular borders and variable pigmentation. ***D.*** Dysplastic nevi are irregularly pigmented and shaped nevomelanocytic lesions that may be associated with familial melanoma.

Epidemiology

Melanoma is an aggressive malignancy of melanocytes: pigment-producing cells that originate from the neural crest and migrate to the skin, meninges, mucous membranes, upper esophagus, and eyes. Melanocytes in each of these locations have the potential for malignant transformation. In the United States, nearly 69,000 individuals were expected to develop melanoma and approximately 9,000 were expected to die in 2010. Although the overall incidence and mortality have increased over the last decades, the mortality rates for younger patients have flattened whereas those rates for individuals over age 65 have continued to increase. It is predominantly a malignancy of white-skinned people (98% of cases), and the incidence correlates with latitude of residence, providing strong evidence for the role of sun exposure. Men are affected slightly more than women (1.3:1), and the median age at diagnosis is the late fifties. Dark-skinned populations (such as those of India and Puerto Rico), blacks, and East Asians also develop melanoma, albeit at rates 10–20 times lower than those in whites. Cutaneous melanomas in these populations are diagnosed more often at a higher stage, and patients tend to have worse outcomes. Furthermore, in nonwhite populations, there is a much higher frequency of acral (subungual, plantar, palmar) and mucosal melanomas.

Risk Factors

The strongest risk factors for melanoma are the presence of multiple benign or atypical nevi and a family or personal history of melanoma (Table 87–1). The presence of melanocytic nevi, common or dysplastic, is a marker for increased risk of melanoma. Nevi have been referred to as precursor lesions because they can transform into melanomas; however, the actual risk for any specific nevus is exceedingly low. About one-quarter of melanomas are histologically associated with nevi, but the majority arise de novo. Table 87–2 lists the characteristic features of clinically atypical moles and the

features that differentiate them from benign acquired nevi. The number of clinically atypical moles may vary from one to several hundred, and they usually differ from one another in appearance. The borders are often hazy and indistinct, and the pigment pattern is more highly varied than that in benign acquired nevi. Individuals with clinically atypical moles and a strong family history of melanoma have been reported to have a >50% lifetime risk for developing melanoma and warrant close follow-up with a dermatologist. Of the 90% of melanoma patients whose disease is regarded as sporadic (i.e., who lack a family history of melanoma), ~40% have clinically atypical moles, compared with an estimated 5–10% of the population at large.

Table 87–1 Factors Associated with Increased Risk of Melanoma

Total body nevi (higher number = higher risk)
Family or personal history Dysplastic nevi
Light skin/hair/eye color
Poor tanning ability
Freckling
UV exposure/sunburns/tanning booths
CDKN2A mutation
MC1R variants

Table 87–2 Pigmented Lesions that Must Be Distinguished from Cutaneous Melanoma and Its Precursors

Blue nevus	Gunmetal or cerulean blue, blue-gray. Stable over time. One-half occur on dorsa of hands and feet. Lesions are usually single, small, 3 mm– < 1 cm. Must be distinguished from nodular melanoma.
Compound nevus	Round or oval shape, well demarcated, smooth-bordered. May be dome-shaped or papillomatous; colors range from flesh-colored to very dark brown, with individual nevi being relatively homogeneous in color.
Hemangioma	Dome-shaped reddish, purple, blue nodule. Compression with a glass microscope slide may result in blanching. Must be distinguished from nodular melanoma.
Junctional nevus Lentigo	Flat to barely raised brown lesion. Sharp border. Fine pigmentary stippling visible, especially upon magnification.
Juvenile	Flat, uniformly medium or dark brown lesion with sharp border.
Solar	Solar lentigines are acquired lesions on sites of chronic solar exposure (face and backs of hands). Lesions are 2 mm–≥1 cm. Solar lentigines have reticulate pigmentation upon magnification.
Pigmented basal cell carcinoma	Papular border. May have central ulceration. Usually on a sun-exposed surface in an older patient. Patient usually has dark brown eyes and dark brown or black hair.
Pigmented dermatofibroma	Lesion is not well demarcated visually, is firm, and dimples downward when compressed laterally. Usually on extremities. Usually <6 mm.
Seborrheic keratosis	Rough, sharp-bordered lesions that feel waxy and "stuck on"; range in color from flesh to tan to dark brown. Presence of keratin plugs in surface is helpful for discriminating especially dark lesions from melanoma.
Subungual hematoma	Maroon (red-brown) coloration. As lesion grows out from nail fold, a curving clear area is seen.
Tattoo (medical or traumatic)	In medical tattoos, lesions are small pigmentary dots, often blue or green, which make a regular pattern (rectangle). Traumatic tattoos are irregular, and pigmentation may appear black.

Congenital melanocytic nevi, which are classified as small (≤1.5 cm), medium (1.5–20 cm), and giant (>20 cm), can be precursors for melanoma. The risk is highest for the giant melanocytic nevus, also called the bathing trunk nevus, which is a rare malformation that affects 1 in 30,000–100,000 individuals, with a lifetime risk of melanoma development estimated to be as high as 6%. At present there are no uniform management guidelines for giant congenital nevi, but because of the potential for malignancy, prophylactic excision early in life is prudent. This usually requires staged removal with

coverage by split-thickness skin grafts. Surgery cannot remove all at-risk nevus cells, as some may penetrate into the muscles or central nervous system (CNS) below the nevus. Small- to medium-size congenital melanocytic nevi affect approximately 1% of persons; the risk of melanoma developing in these lesions is not known but appears to be relatively low. The management of small- to medium-size congenital melanocytic nevi remains controversial.

PERSONAL AND FAMILY HISTORY

Perhaps the single greatest risk factor for melanoma is a personal history of melanoma. Once diagnosed, patients with melanoma require a lifetime of surveillance because their risk is 10 times that of the general population. First-degree relatives have a higher risk of developing melanoma than do individuals without a family history, but only 5–10% of all melanomas are truly familial. In familial melanoma, patients tend to be younger at first diagnosis, lesions are thinner, survival is improved, and multiple primary melanomas are common.

GENETIC SUSCEPTIBILITY

Approximately 20–40% of cases of hereditary melanoma (0.2–2% of all melanomas) are due to germline mutations in the cell cycle regulatory gene cyclin-dependent kinase inhibitor 2A (*CDKN2A*). In fact, 70% of all cutaneous melanomas have somatic mutations or deletions affecting the CDKN2A locus on chromosome 9p21. This locus encodes two distinct tumor suppressor proteins from alternate reading frames: p16 and ARF ($p14^{ARF}$). The p16 protein inhibits CDK4/6-mediated phosphorylation and inactivation of the retinoblastoma (RB) protein, whereas ARF inhibits MDM2 ubiquitin-mediated degradation of p53. The end result of the loss of *CDKN2A* is inactivation of two critical tumor suppressor pathways, RB and p53, which control entry of cells into the cell cycle. Several studies have shown an increased risk of pancreatic cancer among melanoma-prone families with *CDKN2A* mutations.

The melanocortin-1 receptor (*MC1R*) gene is also an inherited melanoma susceptibility factor. Solar radiation stimulates the production of melanocortin [α-melanocyte-stimulating hormone (α-MSH)], the ligand for MC1R, which is a G-protein-coupled receptor that signals via cyclic AMP and regulates the amount and type of pigment produced. *MC1R* is highly polymorphic, and among its 80 variants are those that result in partial loss of signaling and lead to the production of pheomelanin, which is not sun-protective and produces red hair. This red hair color (RHC) phenotype is associated with fair skin, red hair, freckles, increased sun sensitivity, and increased risk of melanoma.

Clinical Classification

Traditionally, four major types of cutaneous melanoma have been recognized (Table 87–3). In three of these types—*superficial spreading melanoma*, *lentigo maligna melanoma*, and *acral lentiginous melanoma*—the lesion has a period of superficial (so-called radial) growth during which it increases in size but does not penetrate deeply. It is during this period that the melanoma is most capable of being cured by surgical excision. The fourth type—*nodular melanoma*—does not have a recognizable radial growth phase and usually presents as a deeply invasive lesion that is capable of early metastasis. When tumors begin to penetrate deeply into the skin, they are in the so-called vertical growth phase. Melanomas with a radial growth phase are characterized by irregular and sometimes notched borders, variation in pigment pattern, and variation in color. An increase in size or change in color is noted by the patient in 70% of early lesions. Bleeding, ulceration, and pain are late signs and are of little help in early recognition. Superficial spreading melanoma is the most common variant observed in the white population. The back is the most common site for melanoma in men. In women, the back and the lower leg (from knee to ankle) are common sites. Nodular melanomas are dark brown-black to blue-black nodules. Lentigo maligna melanoma usually is confined to chronically sun-damaged, sun-exposed sites (face, neck, back of hands) in older individuals. Acral lentiginous melanoma occurs on the palms, soles, nail beds, and mucous membranes. Although this type occurs in whites, it occurs most frequently (along with nodular melanoma) in blacks and East Asians. A fifth type of melanoma, *desmoplastic melanoma*, is associated with a fibrotic response, neural invasion, and a greater tendency for local recurrence. Occasionally, melanomas appear clinically to be amelanotic, in which case the diagnosis is established histologically after biopsy of a new or a changing skin nodule or because of suspicion of a basal cell carcinoma.

Table 87–3 Classification of Malignant Melanoma

Type	Site	Average Age at Diagnosis, Years	Duration of Known Existence, Years	Color
Lentigo maligna melanoma	Sun-exposed surfaces, particularly malar region of cheek and temple	70	5–20 or longer*	In flat portions, shades of brown and tan predominate, but whitish gray occasionally present; in nodules, shades of reddish brown, bluish gray, bluish black

Superficial spreading melanoma	Any site (more common on upper back and, in women, lower legs)	40–50	1–7	Shades of brown mixed with bluish red (violaceous), bluish black, reddish brown, and often whitish pink, and the border of lesion is at least in part visibly and/or palpably elevated
Nodular melanoma	Any	40–50	Months–<5 years	Reddish blue (purple) or bluish black; either uniform in color or mixed with brown or black
Acral lentiginous melanoma	Palm, sole, nail bed, mucous membrane	60	1–10	In flat portions, dark brown predominantly; in raised lesions (plaques), brown-black or blue-black predominantly

*During much of this time, the precursor stage, lentigo maligna, is confined to the epidermis. *Source:* Adapted from AJ Sober, in NA Soter, HP Baden (eds): *Pathophysiology of Dermatologic Diseases* . New York, McGraw-Hill, 1984.

Although melanoma subtypes are clinically and histopathologically distinct, this classification does not have independent prognostic value and has fallen out of favor. Histologic subtype is not part of American Joint Committee on Cancer (AJCC) staging and often is not identified in current pathology reports. Future classification schemes will be based on molecular features of each melanoma (see below). The molecular analysis of individual melanomas will provide a basis for distinguishing benign nevi from melanomas, identify distinct subclasses of melanoma on the basis of the anatomic site, indicate the extent of ultraviolet (UV) exposure, and determine the mutational status of the tumor, which will help elucidate the molecular mechanisms of tumorigenesis and identify targets that will serve as a basis for selection of therapy.

Pathogenesis and Molecular Classification

Considerable evidence from epidemiologic and molecular studies suggests that cutaneous melanomas arise via multiple pathways. There are both environmental and genetic components. Uv solar radiation causes genetic changes in the skin, impairs cutaneous immune function, increases the production of growth factors, and induces the formation of DNA-damaging reactive oxygen species that affect keratinocytes and melanocytes. A comprehensive catalog of somatic mutations from a human melanoma revealed more than 33,000 base mutations with damage to almost 300 protein-coding segments compared with normal cells from the same patient. The dominant mutational signature reflected DNA damage due to UV light exposure. The melanoma also contained previously described driver mutations (i.e., mutations that confer selective clonal growth advantage and are implicated in oncogenesis). These driver mutations affect pathways that promote cell proliferation and inhibit normal pathways of apoptosis in response to DNA repair (see below). The altered melanocytes accumulate DNA damage, and selection occurs for all the attributes that constitute the malignant phenotype: invasion, metastasis, and angiogenesis.

An understanding of the molecular changes that occur during the transformation of normal melanocytes into malignant melanoma not only would help classify patients in similar prognostic groups but also would contribute to the understanding of etiology and help identify new therapeutic options. A genomewide assessment of melanomas classified into four groups based on their location and degree of exposure to the sun has confirmed that there are distinct genetic pathways in the development of melanoma. The four groups were melanomas on skin without chronic sun-induced damage, melanomas on skin with chronic sun-induced damage, mucosal melanomas, and acral melanomas. Remarkably, distinct patterns of DNA alterations were noted that varied with the site of origin and were independent of the histologic subtype of the tumor. What that work and research done by others have shown is that the overall pattern of mutation, amplification, and loss of cancer genes indicate that although the genetic changes are diverse, they have convergent effects on key biochemical pathways involved in proliferation, senescence, and apoptosis. The p16 mutation that leads to cell cycle arrest and the ARF mutation that results in defective apoptotic responses to genotoxic damage were described earlier. The proliferative pathways affected were the mitogen-activated protein (MAP) kinase and phosphatidylinositol 3' kinase/AKT pathways (Fig. 87-2).

Figure 87-2

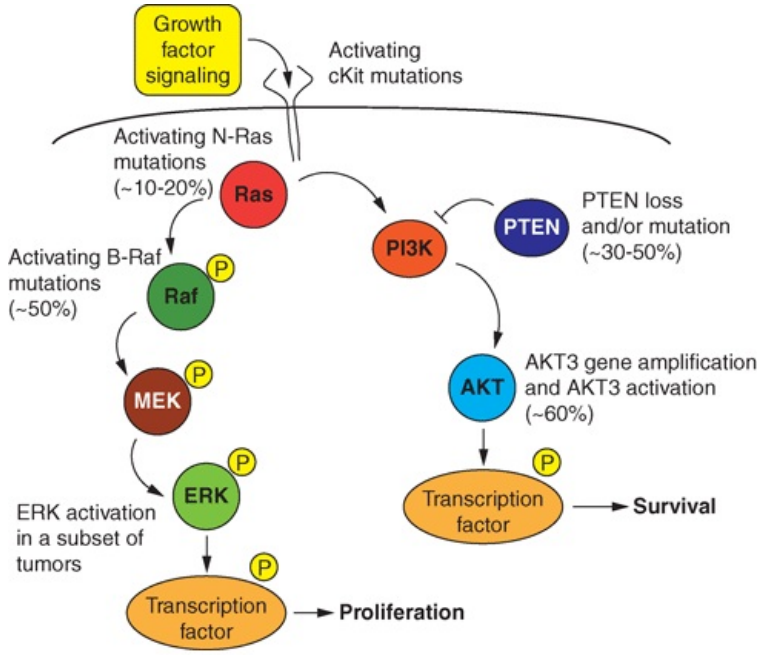

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Major pathways involved in melanoma. The MAP kinase and AKT pathways, which promote proliferation and inhibit apoptosis, respectively, are subject to mutations in melanoma. ERK, extracellular signalregulated kinase; MEK, methylethylketone; PTEN, pentaerythritol tetranitrate.

The RAS family and BRAF, members of the MAP kinase pathway, which classically mediates the transcription of genes involved in cell proliferation and survival, undergo somatic mutation in melanoma. N-RAS is mutated in approximately 20% of melanomas, and somatic activating BRAF mutations are found in most benign nevi and 40–60% of melanomas. Neither mutation by itself appears to be sufficient to cause melanoma; they often are accompanied by other mutations, (e.g., CDKN2A) or phosphatidylinositol 3' kinase pathway (e.g., loss of PTEN). The BRAF mutation is almost always a point mutation (T→A nucleotide change) that results in a valine-to-glutamate amino acid substitution (V600E). V600E BRAF mutations do not have the standard UV signature mutation (pyrimidine dimer) but are present in most melanomas that arise on sites with intermittent sun exposure and are absent in melanomas from chronically sun-damaged skin.

Melanomas also contain mutations in AKT (primarily in AKT3) and PTEN (phosphatase and tensin homolog). AKT can be amplified, and PTEN may be deleted or undergo epigenetic silencing that leads to constitutive activation of the PI3K/AKT pathway and enhanced cell survival by antagonizing the intrinsic pathway of apoptosis. Loss of PTEN, which dysregulates AKT activity, and mutation of AKT3 prolong survival through inactivation of BAD, Bc12-antagonist of cell death, and activation of the forkhead transcription factor FOXO1, which leads to synthesis of prosurvival genes. In melanoma, these two signaling pathways enhance tumorigenesis, chemoresistance, migration, and cell cycle dysregulation. Targeted agents are being employed that inhibit each pathway, but it is likely that effective antimelanoma therapy will require simultaneous inhibition of both MAPK and PI3K.

Diagnosis

The main goal is to diagnose melanoma early in its natural history, before tumor invasion and life-threatening metastases have occurred. Early detection of melanoma may be facilitated by applying the ABCDEs: *a*symmetry (benign lesions are usually symmetric); *b*order irregularity (most nevi have clear-cut borders); *c*olor variegation (benign lesions usually have uniform light or dark pigment); *d*iameter >6 mm (the size of a pencil eraser); *e*volving (any change in size, shape, color, or elevation or new symptoms such as bleeding, itching, and crusting). The aim of differential diagnosis is to distinguish benign pigmented lesions from melanoma and its precursor. If melanoma is a consideration, biopsy is appropriate. Some benign look-alikes may be removed in the process of trying to detect melanoma. Several factors may help distinguish benign nevi from atypical moles:

1. Size: Benign nevi usually are <6 mm in diameter; atypical moles usually are >6 mm in diameter.
2. Shape: Benign nevi usually are round with distinct borders and may be flat or elevated; atypical moles usually have irregular borders with pigment fading off at the edge.
3. Color: Benign nevi usually are uniformly brown or tan; atypical moles usually have variable mixtures of brown, tan, black, and reddish pigment and differ from one another.
4. Location: Benign nevi usually appear on sun-exposed skin above the waist, rarely involving scalp,

breasts, or buttocks; atypical moles usually appear on sun-exposed skin, most often on the back, but can involve the scalp, breasts, or buttocks.

5. Number: Benign nevi are present in 85% of adults, with 10–40 moles scattered over the body; atypical nevi can be present in the hundreds.

The entire cutaneous surface, including the scalp and mucous membranes, as well as the nails should be examined in each patient. Bright room illumination is important, and a hand lens is helpful for evaluating variation in pigment pattern. Any suspicious lesions should be biopsied, evaluated by a specialist, or recorded by chart and/or photography for follow-up. A focused method for examining individual lesions, dermoscopy, employs low-level magnification of the epidermis and may allow a more precise visualization of patterns of pigmentation than is possible with the naked eye. Complete physical examination with attention to the regional lymph nodes is part of the initial evaluation in a patient with suspected melanoma. The patient should be advised to have other family members screened if either melanoma or clinically atypical moles (dysplastic nevi) are present. Patients who fit into high-risk groups should be instructed to perform monthly self-examinations.

BIOPSY

Any pigmented cutaneous lesion that has changed in size or shape or has other features suggestive of malignant melanoma is a candidate for biopsy. The recommended technique is an excisional biopsy, which facilitates pathologic assessment of the lesion, permits accurate measurement of thickness if the lesion is melanoma, and constitutes treatment if the lesion is benign. For large lesions or lesions on anatomic sites where excisional biopsy may not be feasible (such as the face, hands, and feet), an incisional biopsy through the most nodular or darkest area of the lesion is acceptable; this should include the vertical growth phase of the primary tumor, if present. Incisional biopsy does not appear to facilitate the spread of melanoma. For suspicious lesions, every attempt should be made to preserve the ability to assess the deep and peripheral margins and to perform immunohistochemistry. Shave biopsies and cauterization should be avoided. The biopsy should be read by a pathologist experienced in pigmented lesions, and the minimal elements of the report should include Breslow thickness, mitoses per square millimeter for lesions ≤1mm, presence or absence of ulceration, and peripheral and deep margin status. Breslow thickness is the greatest thickness of a primary cutaneous melanoma measured on the slide from the top of the epidermal granular layer, or from the ulcer base, to the bottom of the tumor. To distinguish melanomas from benign nevi in cases with challenging histology, fluorescence in situ hybridization (FISH) with multiple probes can be helpful.

Prognostic Factors

The prognostic factors of greatest importance to a newly diagnosed patient are included in the staging classification (Table 87–4). The best predictor of metastatic risk is the lesion's Breslow thickness. The Clark level, which defines melanomas on the basis of the layer of skin to which a melanoma has invaded, does not add significant prognostic information and no longer is used. Other important factors recognized via the staging classification include presence of ulceration, evidence of nodal involvement, serum lactate dehydrogenase (LDH), and presence and site of distant metastases. The effects of these important prognostic factors on survival can be seen in Fig. 8

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEAD AND NECK CANCER: INTRODUCTION

Epithelial carcinomas of the head and neck arise from the mucosal surfaces in the head and neck area and typically are squamous cell in origin. This category includes tumors of the paranasal sinuses, the oral cavity, and the nasopharynx, oropharynx, hypopharynx, and larynx. Tumors of the salivary glands differ from the more common carcinomas of the head and neck in etiology, histopathology, clinical presentation, and therapy. **Thyroid malignancies are described in Chap. 341.**

INCIDENCE AND EPIDEMIOLOGY

The number of new cases of head and neck cancers in the United States was 36,540 in 2010, accounting for about 3% of adult malignancies; 7880 people died from the disease. The worldwide incidence exceeds half a million cases annually. In North America and Europe, the tumors usually arise from the oral cavity, oropharynx, or larynx, whereas nasopharyngeal cancer is more commonly seen in the Mediterranean countries and in the Far East.

ETIOLOGY AND GENETICS

Alcohol and tobacco use are the most significant risk factors for head and neck cancer in the United States. Smokeless tobacco is an etiologic agent for oral cancers. Other potential carcinogens include marijuana and occupational exposures such as nickel refining, exposure to textile fibers, and woodworking.

Dietary factors may contribute. The incidence of head and neck cancer is higher in people with the lowest consumption of fruits and vegetables. Certain vitamins, including carotenoids, may be protective if included in a balanced diet. Supplements of retinoids such as *cis*-retinoic acid have not been shown to prevent head and neck cancers (or lung cancer) and may increase the risk in active smokers.

Some head and neck cancers have a viral etiology. Epstein-Barr virus (EBV) infection is frequently associated with nasopharyngeal cancer. Nasopharyngeal cancer occurs endemically in some countries of the Mediterranean and Far East, where EBV antibody titers can be measured to screen high-risk populations. Nasopharyngeal cancer has also been associated with consumption of salted fish.

In Western countries, the human papilloma virus (HPV) is associated with approximately 50% of tumors arising from the oropharynx, i.e., the tonsillar bed and base of tongue. Similar to cervical cancer, HPV 16 and 18 are the commonly associated viral subtypes. The incidence of oropharyngeal cancers is increasing in Western counties. Epidemiologically HPV-related oropharyngeal cancer occurs in a younger patient population and is associated with increased numbers of sexual partners and oral sexual practices.

No specific risk factors or environmental carcinogens have been identified for salivary gland tumors.

HISTOPATHOLOGY, CARCINOGENESIS, AND MOLECULAR BIOLOGY

Squamous cell head and neck cancers can be divided into well-differentiated, moderately well-differentiated, and poorly differentiated categories. Poorly differentiated tumors have a worse prognosis than well-differentiated tumors. For nasopharyngeal cancers, the less common differentiated squamous cell carcinoma is distinguished from nonkeratinizing and undifferentiated carcinoma (lymphoepithelioma) that contains infiltrating lymphocytes and is commonly associated with EBV.

Salivary gland tumors can arise from the major (parotid, submandibular, sublingual) or minor salivary glands (located in the submucosa of the upper aerodigestive tract). Most parotid tumors are benign, but half of submandibular and sublingual gland tumors and most minor salivary gland tumors are malignant. Malignant tumors include mucoepidermoid and adenoid cystic carcinomas and adenocarcinomas.

The mucosal surface of the entire pharynx is exposed to alcohol- and tobacco-related carcinogens and is at risk for the development of a premalignant or malignant lesion. Erythroplakia (a red patch) or leukoplakia (a white patch) can be histopathologically hyperplasia, dysplasia, carcinoma in situ, or carcinoma. However, most head and neck cancers do not present with a history of premalignant lesions. Multiple synchronous or metachronous cancers can also be observed. In fact, over time patients with early-stage head and neck cancer are at greater risk of dying from a second malignancy than from a recurrence of the primary disease.

Second head and neck malignancies are usually not therapy-induced; they reflect the exposure of the upper aerodigestive mucosa to the same carcinogens that caused the first cancer. These second primaries develop in the head and neck area, the lung, or the esophagus. Rarely, patients can develop a radiation therapy–induced sarcoma after having undergone prior radiotherapy for a head and neck cancer.

The molecular carcinogenesis of head and neck cancer is a developing story. Activation of oncogenes and inactivation of tumor suppressor genes (frequently of p53) have been described. Overexpression of the epidermal growth factor receptor (EGFR) is common and of prognostic importance.

Resected tumor specimens with histopathologically negative margins ("complete resection") can have residual tumor cells with persistent p53 mutations at the margins. Thus, a tumor-specific p53 mutation can be detected in some phenotypically "normal" surgical margins, indicating residual disease. Patients with such submicroscopic marginal involvement may have a worse prognosis than patients with truly negative margins.

CLINICAL PRESENTATION AND DIFFERENTIAL DIAGNOSIS

Most head and neck cancers occur in patients older than age 50 years. HPV-related malignancies are frequently diagnosed in patients in their 40s while EBV-related nasopharyngeal cancer can occur in all ages, including teenagers. The manifestations vary according to the stage and primary site of the tumor. Patients with nonspecific signs and symptoms in the head and neck area should be evaluated with a thorough otolaryngologic exam, particularly if symptoms persist longer than 2–4 weeks.

Cancer of the nasopharynx typically does not cause early symptoms. However, on occasion it may cause unilateral serous otitis media due to obstruction of the eustachian tube, unilateral or bilateral nasal obstruction, or epistaxis. Advanced nasopharyngeal carcinoma causes neuropathies of the cranial nerves due to skull base involvement.

Carcinomas of the oral cavity present as nonhealing ulcers, changes in the fit of dentures, or painful lesions. Tumors of the tongue base or oropharynx can cause decreased tongue mobility and alterations in speech. Cancers of the oropharynx or hypopharynx rarely cause early symptoms, but they may cause sore throat and/or otalgia.

Hoarseness may be an early symptom of laryngeal cancer, and persistent hoarseness requires referral to a specialist for indirect laryngoscopy and/or radiographic studies. If a head and neck lesion treated initially with antibiotics does not resolve in a short period, further workup is indicated; to simply continue the antibiotic treatment may be to lose the chance of early diagnosis of a malignancy.

Advanced head and neck cancers in any location can cause severe pain, otalgia, airway obstruction, cranial neuropathies, trismus, odynophagia, dysphagia, decreased tongue mobility, fistulas, skin involvement, and massive cervical lymphadenopathy, which may be unilateral or bilateral. Some patients have enlarged lymph nodes even though no primary lesion can be detected by endoscopy or biopsy; these patients are considered to have carcinoma of unknown primary (Fig. 88-1). If the enlarged nodes are located in the upper neck and the tumor cells are of squamous cell histology, the malignancy probably arose from a mucosal surface in the head or neck. Tumor cells in supraclavicular lymph nodes may also arise from a primary site in the chest or abdomen.

Figure 88-1

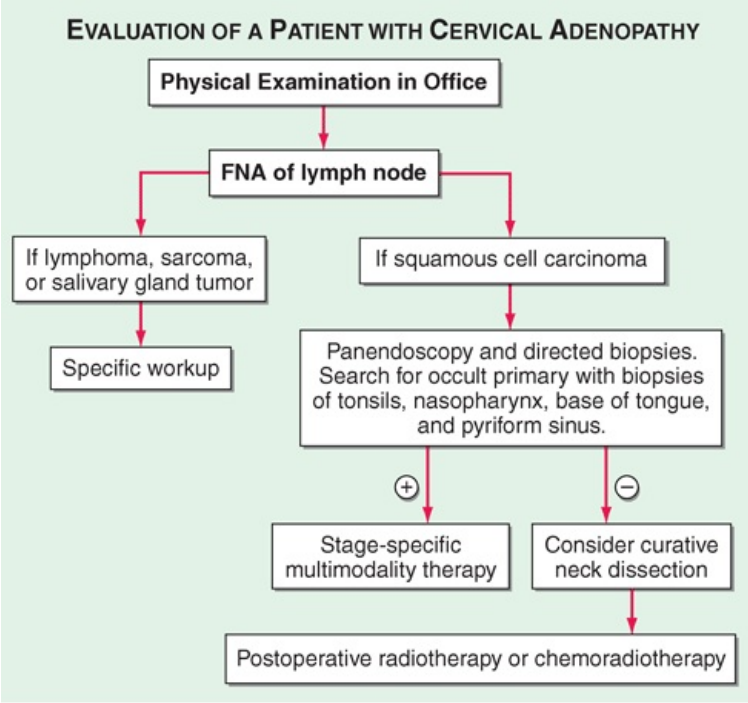

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evaluation of a patient with cervical adenopathy without a primary mucosal lesion; a diagnostic workup. FNA, fine-needle aspiration.

The physical examination should include inspection of all visible mucosal surfaces and palpation of the floor of mouth and tongue and of the neck. In addition to tumors themselves, leukoplakia (a white mucosal patch) or erythroplakia (a red mucosal patch) may be observed; these "premalignant" lesions can represent hyperplasia, dysplasia, or carcinoma in situ and require biopsy. Further examination should be performed by a specialist. Additional staging procedures include CT of the head and neck to identify the extent of the disease. Patients with lymph node involvement should have chest radiography and a bone scan to screen for distant metastases. A positron emission tomographic scan may also be administered and can help to identify or exclude distant metastases. The definitive staging procedure is an endoscopic examination under anesthesia, which may include laryngoscopy, esophagoscopy, and bronchoscopy; during this procedure, multiple biopsy samples are obtained to establish a primary diagnosis, define the extent of primary disease, and identify any additional premalignant lesions or second primaries.

Head and neck tumors are classified according to the TNM system of the American Joint Committee on Cancer. This classification varies according to the specific anatomic subsite (Tables 88–1 and 88–2). Distant metastases are found in <10% of patients at initial diagnosis and are more common in patients with advanced lymph nodal stage; microscopic involvement of the lungs, bones, or liver is more common, particularly in patients with advanced neck lymph node disease. Modern imaging techniques may increase the number of patients with clinically detectable distant metastases in the future.

Table 88–1 TNM Classification for Head and Neck Cancer

Primary Tumor Site (Example)		
T Grade	**Oropharynx**	**Hypopharynx**
T1	0–2 cm	0–2 cm
T2	2.1–4 cm	>1 site, 2.1–4 cm
T3	>4 cm	>4 cm or fixation of hemilarynx
T4a	Invasion of larynx, muscle of tongue, medial pterygoid, hard palate, mandible	Invasion of thyroid/cricoid cartilage, hyoid bone, thyroid gland, esophagus, or central compartment soft tissue invasion
T4b	Invasion of lateral pterygoid muscle, pterygoid plates, lateral nasopharynx, or skull base or encases carotid artery	Invasion of prevertebral fascia, encases carotid artery, or involvement of mediastinal structures

Regional Lymph Nodes (N)	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Unilateral metastasis in lymph node(s), ≤3 cm in greatest dimension
N2	Single ipsilateral lymph node >3.1, ≤6 cm, or multiple ipsilateral, or contralateral lymph nodes ≤6 cm. Bilateral metastasis in lymph node(s), ≤6 cm in greatest dimension, above the supraclavicular fossa
N3	Lymph node >6 cm in greatest dimension

Stage Grouping			
Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage II	T2	N0	M0
Stage III	T3	N0	M0
	T1-T3	N1	M0
Stage IVA	T4a	N0	M0
	T4a	N1	M0
	T1-T4a	N2	M0
Stage IVB	T4b	Any N	M0
	Any T	N3	M0
Stage IVC	Any T	Any N	M1

Table 88–2 Definition of TNM-Nasopharynx

Primary Tumor (T)	Stage Grouping
Tis	Carcinoma in situ
T1	Tumor confined to the nasopharynx
T2	Tumor extends to parapharyngeal soft tissues
T3	Tumor involves bony structures of skull base and/or paranasal sinuses
T4	Tumor with intracranial extension and/or involvement of cranial nerves, infratemporal fossa, hypopharynx, orbit, or masticator space

Regional Lymph Nodes (N)	
	The distribution and the prognostic impact of regional lymph node spread from nasopharynx cancer, particularly of the undifferentiated type, are different from those of other head and neck mucosal cancers and justify the use of a different N classification scheme.
N0	No regional lymph node metastasis
N1	Unilateral metastasis in lymph node(s), ≤6 cm in greatest dimension, above the supraclavicular fossa
N2	Bilateral metastasis in lymph node(s), ≤6 cm in greatest dimension, above the supraclavicular fossa
N3	Metastasis in lymph node(s), >6 cm and/or to supraclavicular fossa
N3a	Greater than 6 cm in dimension
N3b	Extension to the supraclavicular fossa

Anatomic Stage/Prognostic Groups			
Nasopharynx			
Stage 0	Tis	NO	MO
Stage I	T1	N0	M0
Stage II	T1	N1	M0
	T2	N0-N1	M0
Stage III	T1	N2	M0
	T2	N2	M0
	T3	N0-N2	M0
Stage IVA	T4	N0-N2	M0

Stage IVB	Any T	N3	M0
Stage IVC	Any T	Any N	M1

In patients with lymph node involvement and no visible primary, the diagnosis should be made by lymph node excision. If the results indicate squamous cell carcinoma, a panendoscopy should be performed, with biopsy of all suspicious-appearing areas and directed biopsies of common primary sites, such as the nasopharynx, tonsil, tongue base, and pyriform sinus.

Treatment: Head and Neck Cancer

Patients with head and neck cancer can be grossly categorized into three clinical groups: those with localized disease, those with locally or regionally advanced disease, and those with recurrent and/or metastatic disease. Comorbidities associated with tobacco and alcohol abuse can affect treatment outcome and define long-term risks for patients who are cured of their disease.

LOCALIZED DISEASE

Nearly one-third of patients have localized disease, that is, T1 or T2 (stage I or stage II) lesions without detectable lymph node involvement or distant metastases. These lesions are treated with curative intent by either surgery or radiation therapy. The choice of modality differs according to anatomic location and institutional expertise. Radiation therapy is often preferred for laryngeal cancer to preserve voice function, and surgery is preferred for small lesions in the oral cavity to avoid the long-term complications of radiation, such as xerostomia and dental decay. Overall 5-year survival is 60–90%. Most recurrences occur within the first 2 years following diagnosis and are usually local.

LOCALLY OR REGIONALLY ADVANCED DISEASE

Locally or regionally advanced disease—disease with a large primary tumor and/or lymph node metastases—is the stage of presentation for >50% of patients. Such patients can also be treated with curative intent, but not with surgery or radiation therapy alone. Combined modality therapy including surgery, radiation therapy, and chemotherapy is most successful. It can be administered as induction chemotherapy (chemotherapy before surgery and/or radiotherapy) or as concomitant (simultaneous) chemotherapy and radiation therapy. The latter is currently most commonly used and best evidence–supported. In patients with intermediate stage (stage III and early stage IV) concomitant chemoradiotherapy is given postoperatively. It can be administered either as a primary treatment for patients with unresectable disease, to pursue an organ-preserving approach, or in the postoperative setting for intermediate-stage resectable tumors.

Induction Chemotherapy

In this strategy, patients receive chemotherapy [current standard is a three-drug regimen of docetaxel, cisplatin, and fluorouracil (5-FU)] before surgery and radiation therapy. Most patients who receive three cycles show tumor reduction, and the response is clinically "complete" in up to half. This "sequential" multimodality therapy allows for organ preservation (omission of surgery) in patients with laryngeal and hypopharyngeal cancer, and it has been shown to result in higher cure rates compared with radiotherapy alone.

Concomitant Chemoradiotherapy

With the concomitant strategy, chemotherapy and radiation therapy are given simultaneously rather than in sequence. Tumor recurrences from head and neck cancer develop most commonly locoregionally (in the head and neck area of the primary and draining lymph nodes). The concomitant approach is aimed at enhancing tumor cell killing by radiation therapy in the presence of chemotherapy (radiation enhancement). Toxicity (especially mucositis, grade 3 or 4 in 70–80%) is increased with concomitant chemoradiotherapy. However, meta-analyses of randomized trials document an improvement in 5-year survival of 8% with concomitant chemotherapy and radiation therapy. Results seem more favorable in recent trials as more active drugs or more intensive radiotherapy schedules are used. Five-year survival is 34–50%. In addition, concomitant chemoradiotherapy produces better laryngectomy-free survival (organ preservation) than radiation therapy alone in patients with advanced larynx cancer. The use of radiation therapy together with cisplatin has also produced improved survival in patients with advanced nasopharyngeal cancer. The outcome of HPV-related cancers seems to be especially favorable following cisplatin-based chemoradiotherapy.

The success of concomitant chemoradiotherapy in patients with unresectable disease has led to the testing of a similar approach in patients with resected intermediate-stage disease as a postoperative therapy. Concomitant chemoradiotherapy produces a significant improvement over postoperative radiation therapy alone for patients whose tumors demonstrate higher risk features, such as extracapsular spread beyond involved lymph nodes, involvement of multiple lymph nodes, or positive margins at the primary site following surgery.

A monoclonal antibody to the EGFR (cetuximab) increases survival rates when administered during radiotherapy. EGFR blockade results in radiation sensitization and has milder systemic side effects than traditional chemotherapy agents, although an acneiform skin rash is commonly observed. The

integration of cetuximab into current standard chemoradiotherapy regimens is under investigation.

RECURRENT AND/OR METASTATIC DISEASE

Ten percent of patients present with metastatic disease, and more than half of patients with locoregionally advanced disease have recurrence, frequently outside the head and neck region. Patients with recurrent and/or metastatic disease are, with few exceptions, treated with palliative intent. Some patients may require local or regional radiation therapy for pain control, but most are given chemotherapy. Response rates to chemotherapy average only 30–50%; the duration of response averages only 3 months, and the median survival time is 6–8 months. Therefore, chemotherapy provides transient symptomatic benefit. Drugs with single-agent activity in this setting include methotrexate, 5-FU, cisplatin, paclitaxel, and docetaxel. Combinations of cisplatin with 5-FU, carboplatin with 5-FU, and cisplatin or carboplatin with paclitaxel or docetaxel are frequently used.

EGFR-directed therapies, including monoclonal antibodies (e.g., cetuximab) and tyrosine kinase inhibitors (TKIs) of the EGFR signaling pathway (e.g., erlotinib or gefitinib) have single-agent activity of approximately 10%. Side effects are usually limited to an acneiform rash and diarrhea (for the TKIs). The addition of cetuximab to standard combination chemotherapy with cis- or carboplatin and 5-FU was shown to result in a significant increase in median survival.

COMPLICATIONS

Complications from treatment of head and neck cancer are usually correlated to the extent of surgery and exposure of normal tissue structures to radiation. Currently, the extent of surgery has been limited or completely replaced by chemotherapy and radiation therapy as the primary approach. Acute complications of radiation include mucositis and dysphagia. Long-term complications include xerostomia, loss of taste, decreased tongue mobility, second malignancies, dysphagia, and neck fibrosis. The complications of chemotherapy vary with the regimen used but usually include myelosuppression, mucositis, nausea and vomiting, and nephrotoxicity (with cisplatin).

The mucosal side effects of therapy can lead to malnutrition and dehydration. Many centers address issues of dentition before starting treatment, and some place feeding tubes to ensure control of hydration and nutrition intake. About 50% of patients develop hypothyroidism from the treatment; thus, thyroid function should be monitored.

SALIVARY GLAND TUMORS

Most benign salivary gland tumors are treated with surgical excision, and patients with invasive salivary gland tumors are treated with surgery and radiation therapy. These tumors may recur regionally; adenoid cystic carcinoma has a tendency to recur along the nerve tracks. Distant metastases may occur as late as 10–20 years after the initial diagnosis. For metastatic disease, therapy is given with palliative intent, usually chemotherapy with doxorubicin and/or cisplatin. Identification of novel agents with activity in these tumors is a high priority.

FURTHER READINGS

Ang KK et al: Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 363:24, 2010[PMID: 20530316]

Bonner JA et al: Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. Lancet 11:21, 2010[PMID: 19897418]

Pharynx. In *Cancer Staging Handbook: From the AJCC Cancer Staging Manual* , 7th ed, SB Edge et al (eds). New York, Springer, 2009

Haddad RI et al: Recent advances in head and neck cancer. N Engl J Med 359:1143, 2008[PMID: 18784104]

Lonneux M et al: Positron emission tomography with [^{18}F] fluorodeoxyglucose improves staging and patient management in patients with head and neck squamous cell carcinoma: a multicenter prospective study. J Clin Oncol 27:1190, 2010

Pfister DG et al: American Society of Clinical Oncology clinical practice guideline for the use of larynx-preservation strategies in the treatment of laryngeal cancer. J Clin Oncol 24:3693, 2006[PMID: 16832122]

Pointreau Y et al: Randomized trial of induction chemotherapy with cisplatin and 5-flourouracil with or without docetaxel for larynx preservation. J Natl Cancer Inst 101: 498, 2009[PMID: 19318632]

Tobias JS et al: Chemoradiotherapy for locally advanced head and neck cancer: 10-year follow-up of the UK Head and Neck (UKHAN1) trial. Lancet 11:66, 2010[PMID: 19875337]

Vermorken JB et al: Platinum-based chemotherapy plus cetuximab in head and neck cancer. N Engl J Med 359:1116, 2008[PMID: 18784101]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NEOPLASMS OF THE LUNG: INTRODUCTION

Lung cancer is largely a disease of modern man and was considered quite rare before 1900, with fewer than 400 cases described in the medical literature. However, by the mid-twentieth century lung cancer had become epidemic and firmly established as the leading cause of cancer-related death in North America and Europe, killing more than three times as many men as prostate cancer and nearly twice as many women as breast cancer. This fact is particularly distressing since lung cancer is one of the most preventable of all of the common malignancies. Tobacco consumption is the primary cause of lung cancer, a fact firmly established in the mid-twentieth century and codified with the release of the U.S. Surgeon General's 1964 report on the health effects of tobacco smoking. Following the report, cigarette use started to decline in North America and parts of Europe, and with it so did the incidence of lung cancer. To date, the decline in lung cancer is seen most clearly in men; only recently has the decline become apparent among women in the United States. Unfortunately, in many parts of the world, especially in countries with developing economies, cigarette use continues to increase, and along with it, the incidence of lung cancers is also rising. While tobacco smoking remains the primary cause of lung cancer worldwide, more than 60% of new lung cancers occur in never smokers (smoked <100 cigarettes per lifetime) or former smokers (smoked ≥100 cigarettes per lifetime, quit ≥1 year), many of whom quit decades ago. Moreover, 1 in 5 women and 1 in 12 men diagnosed with lung cancer have never smoked. Given the magnitude of the problem, it is incumbent that every internist has a broad knowledge of lung cancer and its management.

EPIDEMIOLOGY

Lung cancer is the most common cause of cancer death among American men and women. More than 220,000 individuals will be diagnosed with lung cancer in the United States in 2010. The incidence of lung cancer peaked among men in the late 1980s and has plateaued in women. Lung cancer is rare below age 40, with rates increasing until age 80, after which the rate tapers off. The projected lifetime probability of developing lung cancer is estimated to be approximately 8% among males and approximately 6% among females. The incidence of lung cancer varies by racial and ethnic group, with the highest age-adjusted incidence rates among African Americans. The excess in age-adjusted rates among African Americans occurs only among men, but age-specific rates show that below age 50 mortality from lung cancer is more than 25% higher among African American than Caucasian women. Incidence and mortality rates among Hispanic and Native and Asian Americans are approximately 40–50% those of whites.

Risk Factors

While the large majority (80–90%) of lung cancers is caused by cigarette smoking, several other factors have been implicated, although none to the extent of tobacco. Cigarette smokers have a tenfold or greater increase in risk of this cancer compared to those who have never smoked. A deep sequencing study suggested that one genetic mutation is induced for every 15 cigarettes smoked. The risk of lung cancer is lower among persons who quit smoking than among those who continue smoking; former smokers have a ninefold increased risk of developing lung cancer compared to men who have never smoked versus the twentyfold excess in those who continue to smoke. The size of the risk reduction increases with the length of time the person has quit smoking, although generally even long-term former smokers have higher risks of lung cancer than those who never smoked. Cigarette smoking increases the risk of all the major lung cancer cell types. Environmental tobacco smoke (ETS) or secondhand smoke is also an established cause of lung cancer. The risk from ETS is less than from active smoking, with a 20–30% increase in lung cancer observed among never smokers married for many years to smokers, in comparison to the 2000% increase among continuing active smokers.

While cigarette smoking is the dominant cause of lung cancer, several other risk factors have been identified, including occupational exposures to asbestos, arsenic, bischloromethyl ether, hexavalent chromium, mustard gas, nickel (as in certain nickel-refining processes), and polycyclic aromatic hydrocarbons. Occupational studies also have provided insight into possible mechanisms of lung cancer induction. For example, the risk of lung cancer among asbestos-exposed workers is increased primarily among those with underlying asbestosis, raising the possibility that the scarring and inflammation produced by this fibrotic nonmalignant lung disease may in many cases (though likely not in all) be the trigger for asbestos-induced lung cancer. Several other occupational exposures have been associated with increased rates of lung cancer, but the causal nature of the association is not as clear.

The risk of lung cancer appears higher among individuals with low fruit and vegetable intake during adulthood. This observation led to hypotheses that specific nutrients, in particular retinoids and carotenoids, might have chemopreventive effects for lung cancer. However, randomized trials failed to validate this hypothesis. In fact, studies found the incidence of lung cancer was increased among smokers with supplementation. Ionizing radiation is also an established lung carcinogen, most convincingly demonstrated from studies showing increased rates of lung cancer among survivors of the atom bombs dropped on Hiroshima and Nagasaki and large excesses among workers exposed to alpha irradiation from radon in underground uranium mining. Prolonged exposure to low-level radon in homes might impart a risk of lung cancer equal or greater than that of ETS. Prior lung diseases such as chronic bronchitis, emphysema, and tuberculosis have been linked to increased risks of lung cancer as well.

SMOKING CESSATION

Given the undeniable link between cigarette smoking and lung cancer (not even addressing other tobacco-related illnesses), physicians must promote tobacco abstinence. Physicians also must help their patients who smoke to stop smoking. Smoking cessation, even well into middle age, can minimize an individual's subsequent risk of lung cancer. Stopping tobacco use before middle age avoids more than 90% of the lung cancer risk attributable to tobacco. However, little health benefit is derived from just "cutting back." Importantly, smoking cessation can even be beneficial in individuals with an established diagnosis of lung cancer, as it is associated with improved survival, fewer side effects from therapy, and an overall improvement in quality of life. Moreover, smoking can alter the metabolism of many chemotherapy drugs, potentially adversely altering the toxicities and therapeutic benefits of the agents. Consequently, it is important to promote smoking cessation even *after* the diagnosis of lung cancer is established.

Physicians need to understand the essential elements of smoking cessation therapy. The individual must want to stop smoking and must be willing to work hard to achieve the goal of smoking abstinence. Self-help strategies alone only marginally affect quit rates, whereas individual and combined pharmacotherapies in combination with counseling can significantly increase rates

of cessation. Therapy with an antidepressant (e.g., bupropion) or nicotine replacement therapy (varenicline, an $\alpha_4\beta_2$ nicotinic acetylcholine receptor partial agonist), are approved by the U.S. Food and Drug Administration (FDA) as first-line treatments for nicotine dependence. However, both drugs have been reported to increase suicidal ideation and must be used with caution. In a randomized trial, varenicline was more efficacious than bupropion or placebo. Prolonged use of varenicline beyond the initial induction phase proved useful in maintaining smoking abstinence. Clonidine and nortriptyline are recommended as second-line treatments (Chap. 395).

INHERITED PREDISPOSITION TO LUNG CANCER

Exposure to environmental carcinogens, such as those found in tobacco smoke, induce or facilitate the transformation from bronchoepithelial cells to the malignant phenotype. The contribution of carcinogens on transformation is modulated by polymorphic variations in genes that affect aspects of carcinogen metabolism. Certain genetic polymorphisms of the P450 enzyme system, specifically CYP1A1, or chromosome fragility are associated with the development of lung cancer. These genetic variations occur at relatively high frequency in the population but their contribution to an individual's lung cancer risk is generally low. However, because of their population frequency, the overall impact on lung cancer risk could be high. In addition, environmental factors, as modified by inherited modulators, likely affect specific genes by deregulating important pathways to enable the cancer phenotype.

First-degree relatives of lung cancer probands have a two- to threefold excess risk of lung cancer and other cancers, many of which are not smoking-related. These data suggest that specific genes and/or genetic variants may contribute to susceptibility to lung cancer. However, very few such genes have yet been identified. Individuals with inherited mutations in *RB* (patients with retinoblastoma living to adulthood) and *p53* (Li-Fraumeni syndrome) genes may develop lung cancer. Three genetic loci for lung cancer risk have been identified by genomewide association studies, including 5p15 (TERT-CLPTM1L), 15q25(CHRNA5-CHRNA-3 nicotinic acetylcholine receptor subunits), and 6p21 (BAT3-MSH5). A rare germline mutation (T790M) involving the epidermal growth factor receptor (EGFR) maybe be linked to lung cancer susceptibility in never smokers. Currently, however, no molecular criteria are used to select patients for more intense screening regimens or for specific chemopreventive strategies.

Pathology

The term *lung cancer* is used for tumors arising from the respiratory epithelium (bronchi, bronchioles, and alveoli). Mesotheliomas, lymphomas, and stromal tumors (sarcomas) are distinct from epithelial lung cancers. According to the World Health Organization classification, epithelial lung cancers consist of four major cell types: small cell lung cancer (SCLC) and the so-called non-small cell lung cancer (NSCLC) histologies including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Fig. 89-1). These four histologies account for approximately 90% of all epithelial lung cancers. The remainder include undifferentiated carcinomas, carcinoids, bronchial gland tumors (including adenoid cystic carcinomas and mucoepidermoid tumors), and rarer tumor types. Tumors may occur as single or mixed-type histology.

Figure 89-1

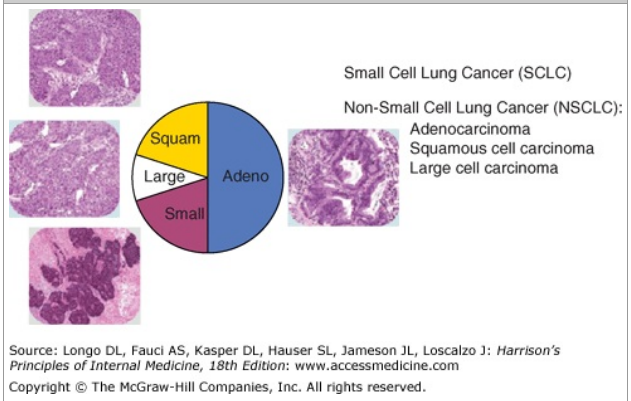

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Traditional view of lung cancer.

All histologic types of lung cancer can be found in current and former smokers. Historically, the histologies associated with heavy tobacco use are squamous and small cell carcinomas. Squamous carcinoma was the most commonly diagnosed form of NSCLC; however, with the steady decline in cigarette consumption over the past four decades and changes in cigarette manufacturing (including use of different types of filters), adenocarcinoma has replaced squamous cell carcinoma as the most frequent histologic subtype in North America. The incidence of small cell carcinoma is also on the decline. In lifetime never smokers, all histologic forms of lung cancer can be found, although adenocarcinoma tends to predominate. Among women and young adults (<60 years), adenocarcinoma tends also to be the most common form of lung cancer.

Small cell carcinoma is a poorly differentiated neuroendocrine tumor that tends to occur as a central mass with endobronchial growth and is strongly associated with smoking. Small cell carcinoma cells have scant cytoplasm, small hyperchromatic nuclei with a fine ("salt and pepper") chromatin pattern and prominent nucleoli. Tumors might be arranged in diffuse sheets of cells or may show neuroendocrine patterns such as rosettes, trabeculae, or peripheral palisading of cells at the periphery of nests. There is often widespread cellular necrosis. Small cell carcinomas, more often than non-small cell carcinomas, may produce specific peptide hormones such as adrenocorticotrophic hormone (ACTH), arginine vasopressin (AVP), atrial natriuretic factor (ANF), and gastrin-releasing peptide (GRP). These hormones may be associated with distinctive paraneoplastic syndromes that prompt workup and eventual diagnosis (Chap. 100).

Squamous cell carcinomas of the lung are morphologically identical to extrapulmonary (i.e., head and neck) squamous cell carcinomas and require clinical correlation to differentiate. These tumors tend to occur centrally and are classically associated with a history of smoking. Histologically, the most common pattern is that of an infiltrating nest of tumor cells that lack intercellular bridges. Keratin can usually be seen when present.

Adenocarcinomas often occur in more peripheral lung locations and may be associated with a history of smoking. However, adenocarcinomas are the most common type of lung cancer occurring in never smokers. Histologically, the tissue may contain the presence of glands, papillary structure, bronchioloalveolar pattern, cellular mucin, or solid pattern if poorly differentiated. Variants of adenocarcinomas include signet-ring, clear cell, and mucinous and fetal adenocarcinomas. Bronchioloalveolar carcinoma (BAC) is a subtype of adenocarcinoma that grows along the alveoli without invasion and can present

radiographically as a single mass, as a diffuse multinodular lesion, as a fluffy infiltrate, and on screening CT scans as a "ground-glass" opacity (GGO). Pure BAC is relatively rare. More common is adenocarcinoma with BAC features. BAC may present in a mucinous form, which tends to be multicentric, and a nonmucinous form, which tends to be solitary.

Large cell carcinomas tend to occur peripherally and are defined as poorly differentiated carcinomas of the lung composed of larger malignant cells without evidence of squamous, glandular differentiation, or features of small cell carcinoma by light microscopy. These tumors usually consist of sheets of large malignant cells, often with associated necrosis. Cytologically, the tumor is also arranged in syncytial groups and single cells. Variants of large cell, carcinoma include basaloid carcinoma, which may present as an endobronchial lesion and may resemble a high-grade neuroendocrine tumor, and lymphoepithelioma-like carcinoma, which is similar to the same-named tumor of other sites and is Epstein-Barr virus–related.

Historically, for treatment and prognostication purposes, the major distinction has been between SCLC and NSCLC, as these tumors have quite different natural histories and therapeutic approaches. SCLC is typically widely disseminated at diagnosis. Even if localized, it is rarely curable by surgery. By contrast, NSCLC can be potentially cured by resection in up to 30% of cases. Small cell cancers tend to respond more favorably to traditional cytotoxic chemotherapy agents. Intrinsic drug resistance is the norm for both SCLC and NSCLC. As knowledge of tumor biology improves, more sophisticated classification schemas are under development, including ones based in part on the presence of specific mutations and molecular alterations (Fig. 89-2). Recognition of these molecular distinctions may help guide therapy in the future.

Figure 89-2

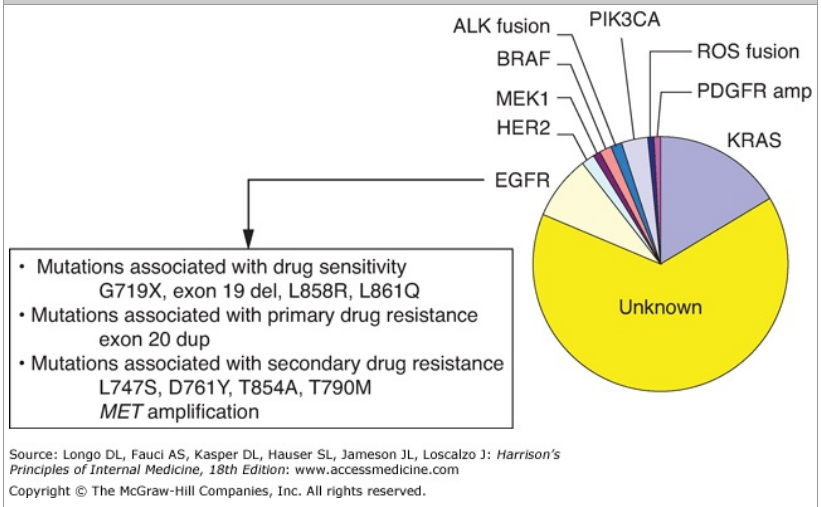

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

2010: Lung adenocarcinoma—multiple molecular subsets.

Immunohistochemistry

The diagnosis of lung cancer most often rests on the morphologic or cytologic features correlated with clinical and radiographic findings. Immunohistochemistry may be used to verify neuroendocrine differentiation within a tumor, with markers such as neuron-specific enolase (NSE), CD56 or neural cell adhesion molecule (NCAM), synaptophysin, chromogranin, and Leu7 (Table 89–1). Immunohistochemistry is also helpful in differentiating primary from metastatic adenocarcinomas. Thyroid transcription factor 1 (TTF-1), identified in tumors of thyroid and pulmonary origin, is positive in more than 70% of pulmonary adenocarcinomas and is a reliable indicator of primary lung cancer, provided a thyroid primary has been excluded. A negative TTF-1, however, does not exclude the possibility of a lung primary. TTF-1 is also positive in neuroendocrine tumors of pulmonary and extrapulmonary origin. Cytokeratins 7 and 20 used in combination can help narrow the differential diagnosis; nonsquamous NSCLC, SCLC, and mesothelioma may stain positive for CK7 and negative for CK20, while squamous cell lung cancer will be both CK7 and CK20 negative. Mesothelioma can be easily identified ultrastructurally, but it has historically been difficult to differentiate from adenocarcinoma through morphology and immunohistochemical staining. Several markers in the past few years have proven to be more helpful, including CK5/6, calretinin, and Wilms' tumor gene 1 (WT-1), all of which show positivity in mesothelioma.

Table 89–1 Common Immunohistochemical Markers Used in the Diagnosis of Lung Tumors

Histology	Positive Immunohistochemical Markers
Squamous cell carcinoma	Cytokeratin (CK) cocktail, e.g., AE1/AE3 CK5/6 CK7 rare
Adenocarcinoma	Cytokeratin cocktail, e.g., AE1/AE3 CK7 TTF-1 Neuroendocrine markers rare, e.g., CD56, NSE
Large cell carcinoma	Cytokeratin TTF-1 rare Neuroendocrine markers rare (e.g., CD56, NSE)
Large cell neuroendocrine carcinoma	Cytokeratin cocktail, e.g., AE1/AE3 TTF-1 CD56 Chromogranin

	Synaptophysin
Small cell carcinoma	Cytokeratin cocktail (tends to be patchy) TTF-1 CD56 Chromogranin Synaptophysin

Abbreviation: WT-1, Wilms' tumor gene 1; NSE, neuron-specific enolase; NCAM, neural cell adhesion molecule; TTF-1, thyroid transcription factor 1.

Molecular Pathogenesis

Cancer is a disease involving dynamic changes in the genome. As proposed by Hanahan and Weinberg, virtually all cancer cells acquire six hallmark capabilities: self-sufficiency in growth signals, insensitivity to antigrowth signals, evading apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. The order in which these hallmark capabilities are acquired appears quite variable and can differ from tumor to tumor. Events leading to acquisition of these hallmarks can vary widely; although broadly, cancers arise as a result of accumulations of gain-of-function mutations in oncogenes and loss-of-function mutations in tumor suppressor genes. Further complicating the study of lung cancer, the sequence of events that lead to disease is clearly different for the various histopathologic entities.

The exact cell of origin for lung cancers is not known. Whether one cell of origin leads to all histologic forms of lung cancer is unclear. However, at least for lung adenocarcinoma, type II epithelial cells (or alveolar epithelial cells) have the capacity to give rise to tumors. For SCLC, cells of neuroendocrine origin have been implicated as precursors.

For cancers in general, one theory holds that a small subset of the cells within a tumor (i.e., "stem cells") are responsible for the full malignant behavior of the tumor. As part of this concept, the large bulk of the cells in a cancer are "offspring" of these cancer stem cells. While clonally related to the cancer stem cell subpopulation, most cells by themselves cannot regenerate the full malignant phenotype. The stem cell concept may explain the failure of standard medical therapies to eradicate lung cancers, even when there is a clinical complete response. Disease recurs because therapies do not eliminate the stem cell component, which may be more resistant to chemotherapy. Precise human lung cancer stem cells have yet to be identified.

Lung cancer cells harbor multiple chromosomal abnormalities, including mutations, amplifications, insertions, deletions, and translocations. One of the earliest set of oncogenes found to be aberrant was the MYC family of transcription factors (*MYC, MYCN*, and *MYCL*). *MYC* is most frequently activated via gene amplification or transcriptional dysregulation in both SCLC and NSCLC, whereas abnormalities of *MYCN* and *MYCL* generally occur in SCLC. Currently, there are no MYC-specific drugs.

To date, among lung cancer histologies, adenocarcinomas have been the most extensively catalogued for recurrent genomic gains and losses as well as for somatic mutations. While multiple different kinds of aberrations have been found, a major class involves "driver mutations"—mutations that occur in genes encoding signaling proteins that when aberrant, drive initiation and maintenance of tumor cells (Table 89–2). Importantly, driver mutations can serve as Achilles' heels for tumors if their gene products can be targeted appropriately. For example, one set of mutations involves the epidermal growth factor receptor (EGFR), which belongs to the ERBB (HER) family of protooncogenes, including *EGFR* (ERBB1), *Her2/neu* (ERBB2), *HER3* (ERBB3), and *HER4* (ERBB4). These genes encode cell-surface receptors consisting of an extracellular ligand-binding domain, a transmembrane structure, and an intracellular tyrosine kinase (TK) domain. The binding of ligand to receptor activates receptor dimerization and TK autophosphorylation, initiating a cascade of intracellular events, leading to increased cell proliferation, angiogenesis, metastasis, and a decrease in apoptosis. Lung adenocarcinomas can arise when tumors express mutant *EGFR*. These same tumors display high sensitivity to small molecule EGFR tyrosine kinase inhibitors. Additional examples of driver mutations in lung adenocarcinoma include those involving the signaling molecules downstream of EGFR, e.g., the tyrosine kinase *HER2*; the GTPase, *KRAS*; the serine-threonine kinase, *BRAF*; and the lipid kinase, *PIK3CA*. In 2007, other subsets of lung adenocarcinoma were found to be defined by the presence of specific translocations fusing tyrosine kinases such as ALK and ROS to aberrant upstream partners. Notably, at least *EGFR*, *KRAS*, and *EML4-ALK* mutations are mutually exclusive, suggesting that acquisition of one of these driver mutations is sufficient to promote tumorigenesis. Thus far, potentially targetable driver mutations have mostly been identified in lung adenocarcinomas as opposed to lung cancers displaying other types of histologies.

Table 89–2 List of Some Genes Somatically Altered in Different Histologic Subtypes of Lung Cancer

Histology	Oncogene	Tumor-Suppressor Genes
Adenocarcinoma	EGFR KRAS ALK	TP53 CDKN2A/B (p16, p14) LKB1 (STK11)
Squamous cell carcinoma	EGFR PIK3CA IGF-1R	TP53 TP63
Small cell carcinoma	MYC BCL-2	TP53 RB1 FHIT
Large cell carcinoma (not well studied)		

Abbreviation: EGFR, epidermal growth factor receptor; ALK, anaplastic lymphoma kinase; IGF-1R, insulin-like growth factor 1 receptor; RB1, retinoblastoma protein 1.

A large number of tumor-suppressor genes (recessive oncogenes) have also been identified that are inactivated during the pathogenesis of lung cancer (Table 89–2). Such genes include *TP53*, *RB1*, *RASSF1A*, *CDKN2A/B*, *LKB1* (*STK11*), and *FHIT*. Nearly 90% of SCLCs harbor mutations in *TP53* and *RB1*. Several tumor-suppressor genes on chromosome 3p appear to be involved in nearly all lung cancers. Allelic loss for this region occurs very early in lung cancer pathogenesis, including in histologically normal smoking-damaged lung epithelium.

EARLY DETECTION AND SCREENING

The clinical outcome for lung cancer is related to the stage at diagnosis. Accordingly, it is presumed that early detection of occult tumors will lead to improved survival. Early detection is a process that involves screening tests, surveillance, diagnosis, and early treatment. By contrast, screening is defined as a systematic testing of asymptomatic individuals for preclinical disease. The majority of patients with lung cancer present with advanced disease, raising the question as to whether screening could detect lung tumors at earlier stages when they are theoretically more curable. In order for a screening program to be successful, the burden of disease within the population must be high, effective treatment must be available that can reduce mortality rate, and the test must be accessible, cost-effective, and both sensitive and specific. With any screening procedure, one must keep in mind the possible influence of lead-time bias (i.e., detecting the cancer earlier without an effect on survival), length time bias (i.e., indolent cancers are detected on screening and may actually not affect survival, while aggressive cancers are likely to cause symptoms earlier in patients and are less likely to be detected), and over diagnosis (i.e., diagnosing cancers so slow growing that they are unlikely to cause the death of the patient)(Chap. 82).

Randomized controlled trials from the 1960s to the 1980s reported no impact on lung cancer–specific mortality rate using screening chest radiographs with or without sputum cytology in high-risk patients (age >50 years or history of smoking). Although these studies have been criticized for their design, statistical analyses, and outdated imaging modalities, they resulted in the current recommendations not to use these tools to screen for lung cancer. The more recent Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial has completed accrual. This study involved more than 150,000 patients randomized to standard care or a baseline single-view posterior-anterior chest radiograph. A total of 5991 (8.9%) baseline chest x-rays were reported as suspicious for lung cancer, highest in current and former smokers. Two hundred six patients had a biopsy, of which 126 (61%) were positive for lung cancer. Among those with positive biopsies, 52% were stage I, 12% were stage II, and 22% were stage III. Long-term follow-up is required to determine the effect on mortality rate, if any.

Low-dose, noncontrast, thin-slice helical or spiral chest CT has emerged as a possible new tool for lung cancer screening. In a spiral chest CT scan, only the pulmonary parenchyma is examined, thus negating the use of intravenous contrast and the necessity of a physician being present at the exam. The scan can usually be done quickly (within a breath) and involves low doses of radiation. However, the benefits of screening with such technology remain to be determined. The International Early Lung Cancer Action Project (I-ECLAP) screened 31,567 asymptomatic patients at high risk for lung cancer (age ≥60 years with history of at least a 10–pack-year smoking history) using low-dose baseline CT and annual screening in 27,456 study participants. Suspicious lesions requiring biopsies were indicated in 535 participants. Lung cancer was diagnosed in 484 participants; 405 at baseline, 74 on subsequent screening, and 5 participants due to symptoms between annual visits. Of the 484 participants who received a diagnosis of lung cancer, 412 (85%) were clinical stage I with an estimated 10-year survival rate of 88% regardless of treatment and 92% among the 302 participants who underwent resection within 1 month after diagnosis. A second trial randomized 1276 patients to low-dose CT screening and 1196 to baseline chest x-ray followed by yearly medical examination. At 3-year follow-up, this study found a trend toward more patients being diagnosed with stage I lung cancer in the low-dose CT arm, with no difference in the number of patients diagnosed with advanced lung cancer or lung cancer deaths. More mature data from all trials are required to determine whether screening reduces lung cancer mortality rates.

A major challenge confronting advocates of CT screening is the high false-positive rate; on initial screening of at-risk populations, false-positive rates range between 10 and 20% but can be as high as 50%, depending on the geographical region. Positive predictive values range from 2.8 to 11.6%. False positives can have a substantial impact on patients through the expense and risk of unneeded further evaluation and emotional stress. False-positive rates and positive predictive values are somewhat improved in annual follow-up CT scans, but there is still significant room for improvement. Based on extant data, it appears that nodules <0.5 mm are unlikely to be cancerous and those 5–10 mm in diameter (25–40% of noncalcified nodules detected) are of uncertain significance. The management of these patients usually consists of serial CT scans over time to see if the nodules grow, attempted fine-needle aspirates, or surgical resection (Fig. 89-3).

Figure 89-3

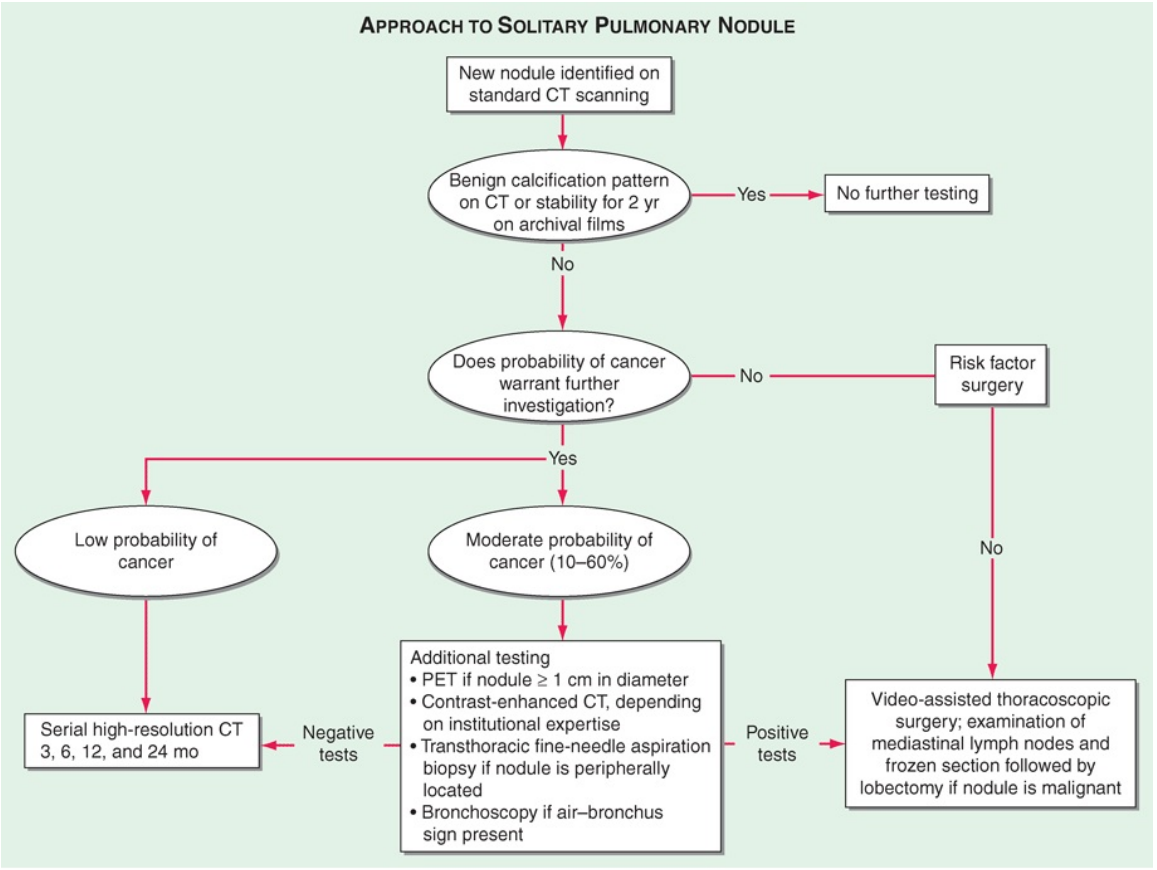

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to solitary pulmonary nodule.

Two additional screening studies are ongoing, the National Lung Cancer Screening Trial (NLST), a prospective comparison of spiral CT and standard chest x-ray in 50,000 current or ex-smokers and a similar study in Europe comparing CT scanning with standard of care in subjects with a history of heavy smoking. Until these data and more mature data from the above-mentioned trials become available, routine CT screening for lung cancer cannot be recommended for any risk group. For those patients who want to be screened, physicians need to discuss the possible benefits and risks of screening. While lung cancers may be found, patients are at risk for more radiation exposure and false-positive results. The latter can result in multiple follow-up CTs and possible invasive procedures, with potential added costs, anxiety, and morbidity and mortality rates. There are no data on screening never smokers for lung cancer.

CLINICAL MANIFESTATIONS

More than half of all patients diagnosed with lung cancer present with advanced disease at the time of diagnosis. The majority of patients present with signs, symptoms, or laboratory abnormalities that can be attributed to the primary lesion, local tumor growth, invasion or obstruction of adjacent structures, growth at distant metastatic sites, or a paraneoplastic syndrome (Tables 89–3 and 89–4). The prototypical lung cancer patient is a current or former smoker of either sex, usually in the seventh decade of life. A history of chronic cough with or without hemoptysis in a current or former smoker with chronic obstructive pulmonary disease (COPD) aged 40 years or older should prompt a thorough investigation for lung cancer even in the face of a normal chest x-ray. A persistent pneumonia without constitutional symptoms and unresponsive to repeated courses of antibiotics also should prompt an evaluation for the underlying cause.

Table 89–3 Presenting Signs and Symptoms of Lung Cancer

Symptom and Signs	Range of frequency
Cough	8–75%
Weight loss	0–68%
Dyspnea	3–60%
Chest pain	20–49%
Hemoptysis	6–35%
Bone pain	6–25%
Clubbing	0–20%
Fever	0–20%
Weakness	0–10%
SVCO	0–4%
Dysphagia	0–2%
Wheezing and stridor	0–2%

Abbreviation: SVCO, superior vena cava obstruction.

Source: Reproduced with permission from MA Beckles et al: Chest 123:97, 2003.

Table 89–4 Clinical Findings Suggestive of Metastatic Disease	
Symptoms elicited in history	• Constitutional: weight loss >10 lb • Musculoskeletal: focal skeletal pain • Neurologic: headaches, syncope, seizures, extremity weakness, recent change in mental status
Signs found on physical examination	• Lymphadenopathy (>1 cm) • Hoarseness, superior vena cava syndrome • Bone tenderness • Hepatomegaly (>13 cm span) • Focal neurologic signs, papilledema • Soft-tissue mass
Routine laboratory tests	• Hematocrit: <40% in men, <35% in women • Elevated alkaline phosphatase, GGT, SGOT, and calcium levels

Abbreviation: GGT, gamma-glutamyltransferase; SGOT, serum glutamic-oxaloacetic transaminase.

Source: Reproduced with permission from GA Silvestri et al: Chest 123(1 Suppl):147S, 2003.

Lung cancer arising in a lifetime never smoker is more common in women and East Asians. Such patients also tend to be younger than their smoking counterparts at the time of diagnosis. The clinical presentation of lung cancer in never smokers tends to mirror that of current and former smokers.

Patients with central or endobronchial growth of the primary tumor may present with cough, hemoptysis, wheeze, stridor, dyspnea, or postobstructive pneumonitis. Peripheral growth of the primary tumor may cause pain from pleural or chest wall involvement, dyspnea on a restrictive basis, and symptoms of a lung abscess resulting from tumor cavitation. Regional spread of tumor in the thorax (by contiguous growth or by metastasis to regional lymph nodes) may cause tracheal obstruction, esophageal compression with dysphagia, recurrent laryngeal paralysis with hoarseness, phrenic nerve palsy with elevation of the hemidiaphragm and dyspnea, and sympathetic nerve paralysis with Horner's syndrome (enophthalmos, ptosis, miosis, and anhidrosis). Malignant pleural effusions can cause pain or dyspnea. Pancoast (or superior sulcus tumor) syndromes result from local extension of a tumor growing in the apex of the lung with involvement of the eighth cervical and first and second thoracic nerves, with shoulder pain that characteristically radiates in the ulnar distribution of the arm, often with radiologic destruction of the first and second ribs. Often Horner's syndrome and Pancoast syndrome coexist. Other problems of regional spread include superior vena cava syndrome from vascular obstruction; pericardial and cardiac extension with resultant tamponade, arrhythmia, or cardiac failure; lymphatic obstruction with resultant pleural effusion; and lymphangitic spread through the lungs with hypoxemia and dyspnea. In addition, lung cancer can spread transbronchially, producing tumor growth along multiple alveolar surfaces with impairment of gas exchange, respiratory insufficiency, dyspnea, hypoxemia, and sputum production. Constitutional symptoms may include anorexia, weight loss, weakness, fever, and night sweats. Apart from the brevity of symptom duration, these parameters fail to clearly distinguish SCLC from NSCLC or even from neoplasms metastatic to lungs.

Extrathoracic metastatic disease is found at autopsy in >50% of patients with squamous carcinoma, 80% of patients with adenocarcinoma and large cell carcinoma, and >95% of patients with SCLC. Approximately one-third of patients present with symptoms as a result of distant metastases. Lung cancer metastases may occur in virtually every organ system, and the site of metastatic involvement largely determines other symptoms. Patients with brain metastases may present with headache, nausea and vomiting, or neurologic deficits. Patients with bone metastases may present with pain, pathologic fractures, or cord compression. The latter may also occur with epidural metastases. Individuals with bone marrow invasion may present with cytopenias or leukoerythroblastosis. Those with liver metastases may present with hepatomegaly, right upper quadrant pain, anorexia, and weight loss. Liver dysfunction or biliary obstructions are rare. Adrenal metastases are common but rarely cause pain or adrenal insufficiency unless they are large.

Paraneoplastic syndromes are common in patients with lung cancer, especially those with SCLC, and may be the presenting finding or the first sign of recurrence. In addition, paraneoplastic syndromes may mimic metastatic disease and, unless detected, lead to inappropriate palliative rather than curative treatment. Often the paraneoplastic syndrome may be relieved with successful treatment of the tumor. In some cases, the pathophysiology of the paraneoplastic syndrome is known, particularly when a hormone with biologic activity is secreted by a tumor. However, in many cases the pathophysiology is not known. Systemic symptoms of anorexia, cachexia, weight loss (seen in 30% of patients), fever, and suppressed immunity are paraneoplastic syndromes of unknown etiology or at least not well defined. Weight loss greater than 10% of total body weight is considered a bad prognostic sign. Endocrine syndromes are seen in 12% of patients; hypercalcemia resulting from ectopic production of parathyroid hormone (PTH), or more commonly, PTH-related peptide, is the most common life-threatening metabolic complication of malignancy, primarily occurring with squamous cell carcinomas of the lung. Clinical symptoms include nausea, vomiting, abdominal pain, constipation, polyuria, thirst, and altered mental status.

Hyponatremia may be caused by the syndrome of inappropriate secretion of antidiuretic hormone (SIADH) or possibly atrial natriuretic peptide (ANP). SIADH resolves within 1–4 weeks of initiating chemotherapy in the vast majority of cases. During this period, serum sodium can usually be managed and maintained above 128 meq/L via fluid restriction. Demeclocycline can be a useful adjunctive measure when fluid restriction alone is insufficient. Of note, patients with ectopic ANP may have worsening hyponatremia if sodium intake is not concomitantly increased. Accordingly, if hyponatremia fails to improve or worsens after 3–4 days of adequate fluid restriction, plasma levels of ANP should be measured to determine the causative syndrome.

Ectopic secretion of ACTH by SCLC and pulmonary carcinoids usually results in additional electrolyte disturbances, especially hypokalemia, rather than the changes in body habitus that occur in Cushing's syndrome from a pituitary adenoma. Treatment with standard medications, such as metyrapone and ketoconazole, is largely ineffective due to extremely high cortisol levels. The most effective strategy for management of Cushing's syndrome is effective treatment of the underlying SCLC. Bilateral adrenalectomy may be considered in extreme cases.

Skeletal–connective tissue syndromes include clubbing in 30% of cases (usually NSCLCs) and hypertrophic primary osteoarthropathy in 1–10% of cases (usually adenocarcinomas). Patients may develop periostitis, causing pain, tenderness, and swelling over the affected bones and a positive bone scan. Neurologic–myopathic syndromes are seen in only 1% of patients but are dramatic and include the myasthenic Eaton-Lambert syndrome and retinal blindness with SCLC, while peripheral neuropathies, subacute cerebellar degeneration, cortical degeneration, and polymyositis are seen with all lung

cancer types. Many of these are caused by autoimmune responses such as the development of anti–voltage-gated calcium channel antibodies in Eaton-Lambert syndrome. Patients with this disorder present with proximal muscle weakness, usually in the lower extremities, occasional autonomic dysfunction, and rarely with cranial nerve symptoms or involvement of the bulbar or respiratory muscles. Depressed deep tendon reflexes are frequently present. In contrast to patients with myasthenia gravis, strength improves with serial effort. Some patients who respond to chemotherapy will have resolution of the neurologic abnormalities. Thus, chemotherapy is the initial treatment of choice. Paraneoplastic encephalomyelitis and sensory neuropathies, cerebellar degeneration, limbic encephalitis, and brainstem encephalitis occur in SCLC in association with a variety of antineuronal antibodies such as anti-Hu, anti-CRMP5, and ANNA-3. Paraneoplastic cerebellar degeneration may be associated with anti-Hu, anti-Yo, or P/Q calcium channel autoantibodies. Coagulation, thrombotic, or other hematologic manifestations occur in 1–8% of patients and include migratory venous thrombophlebitis (Trousseau's syndrome), nonbacterial thrombotic (marantic) endocarditis with arterial emboli, and disseminated intravascular coagulation with hemorrhage, anemia, granulocytosis, and leukoerythroblastosis. Thrombotic disease complicating cancer is usually a poor prognostic sign. Cutaneous manifestations such as dermatomyositis and acanthosis nigricans are uncommon (1%), as are the renal manifestations of nephrotic syndrome and glomerulonephritis (≤1%).

DIAGNOSING LUNG CANCER

Tissue sampling is required to confirm a diagnosis in all patients with suspected lung cancer. Tumor tissue may be obtained via minimally invasive techniques such as bronchial or transbronchial biopsy during fiberoptic bronchoscopy, by fine-needle aspiration (FNA) or percutaneous biopsy using image guidance, or via endobronchial ultrasound (EBUS)-guided biopsy. Depending on the location, lymph node sampling may occur via transesophageal endoscopic ultrasound guide biopsy (EUS), EBUS, or blind biopsy. In patients with clinically palpable lymph nodes, an FNA may be obtained. In patients with suspected metastatic disease, a diagnosis may be confirmed by percutaneous biopsy of a soft tissue mass, lytic bone lesion, bone marrow, pleural or liver lesion, or an adequate cell block obtained from a malignant pleural effusion. In patients with a suspected malignant pleural effusion, if the initial thoracentesis is negative, a repeat thoracentesis is recommended. While the majority of pleural effusions are due to malignant disease, particularly if they are exudative or bloody, some may be parapneumonic. In this case, patients should be considered for possible curative treatment.

The diagnostic yield of any biopsy depends on several factors, including location (accessibility) of the tumor, tumor size, tumor type, and technical aspects of the diagnostic procedure including the experience level of the bronchoscopist and pathologist. In general, central lesions, such as squamous cell carcinomas, small cell carcinomas, or endobronchial lesions, such as carcinoid tumors, are more readily diagnosed by bronchoscopic examination, while peripheral lesions such as adenocarcinomas and large cell carcinomas are more amenable to transthoracic FNA. Diagnostic accuracy for SCLC versus NSCLC for most specimens is excellent, with lesser accuracy for subtypes of NSCLC.

Bronchoscopic specimens include bronchial brush, bronchial wash, bronchioloalveolar lavage, and transbronchial FNA. Of these, transbronchial FNA consistently demonstrates the highest sensitivity, surpassed only by the use of a combination of bronchoscopic specimens. Overall sensitivity for combined use of bronchoscopic methods is approximately 80%, and together with tissue biopsy, the yield increases to 85–90%. Like transbronchial FNA specimens, transthoracic FNA specimens are also very good, yielding diagnostic material in 70–95% of cases. Sensitivity is highest for larger lesions and peripheral tumors. In general, FNA specimens, whether transbronchial, transthoracic, or endoscopic ultrasound-guided, are superior to other specimen types. This is primarily due to the higher percentage of tumor cells with fewer confounding factors, such as obscuring inflammation and reactive nonneoplastic cells. For more accurate histologic classification, mutation analysis, or for investigational purposes, reasonable efforts (e.g., a core-needle biopsy) should be made to obtain more tissue than what is contained in a routine cytology specimen obtained by FNA.

Sputum cytology is inexpensive and noninvasive but has a lower yield than other specimen types due to poor preservation of the cells and more variability in acquiring a good-quality specimen. The yield for sputum cytology is highest for larger and centrally located tumors such as squamous cell carcinoma and small cell carcinoma histology. The specificity for sputum cytology averages close to 100%, although sensitivity is generally less than 70%. The accuracy of sputum cytology improves with increased numbers of specimens analyzed. Consequently, analysis of at least three sputum specimens is recommended.

STAGING LUNG CANCER

Lung cancer staging consists of two parts: first, a determination of the location of the tumor and possible metastatic sites (anatomic staging), and second, an assessment of a patient's ability to withstand various antitumor treatments (physiologic staging). All patients with lung cancer should have a complete history and physical examination, with evaluation of all other medical problems, determination of performance status, and history of weight loss. The most significant dividing line is between those patients who are candidates for surgical resection and those who are inoperable but will benefit from chemotherapy, radiation therapy, or both. Staging with regard to a patient's potential for surgical resection is principally applicable to NSCLC.

Anatomic Staging of Patients with Lung Cancer

The accurate staging of patients with NSCLC is essential for determining the appropriate treatment in patients with resectable disease and avoiding unnecessary surgical procedures in patients with advanced disease (Fig. 89-4). All patients with NSCLC should undergo initial radiographic imaging with CT scan, positron emission tomography (PET), or preferably CT-PET. PET scanning attempts to identify sites of malignancy based on glucose metabolism by measuring the uptake of fluorodeoxyglucose F18. Rapidly dividing cells, presumably in the lung tumors, will preferentially take up ^{18}F-FDG and appear as a "hot spot." To date, PET has been mostly used for staging and detection of metastases in lung cancer and in the detection of nodules >15 mm in diameter. Combined ^{18}F-FDG PET-CT imaging has been shown to improve the accuracy of staging in NSCLC compared to visual correlation of PET and CT or either study alone. CT-PET has been found to be superior in identifying pathologically enlarged mediastinal lymph nodes and extrathoracic metastases. A standardized uptake value (SUV) of >2.5 on PET is highly suspicious for malignancy. False negatives can be seen in diabetes, in lesions <8 mm, in slow-growing tumors, and in concurrent infections such as tuberculosis. False positives can be seen in infections and granulomatous disease. Thus, PET should never be used alone to diagnose lung cancer, mediastinal involvement, or metastases. Confirmation with tissue biopsy is required. For brain metastases, MRI is the most effective method. MRI can also be useful in selected circumstances, such as for superior sulcus tumors, to rule out brachial plexus involvement, but in general does not play a major role in NSCLC staging.

Figure 89-4

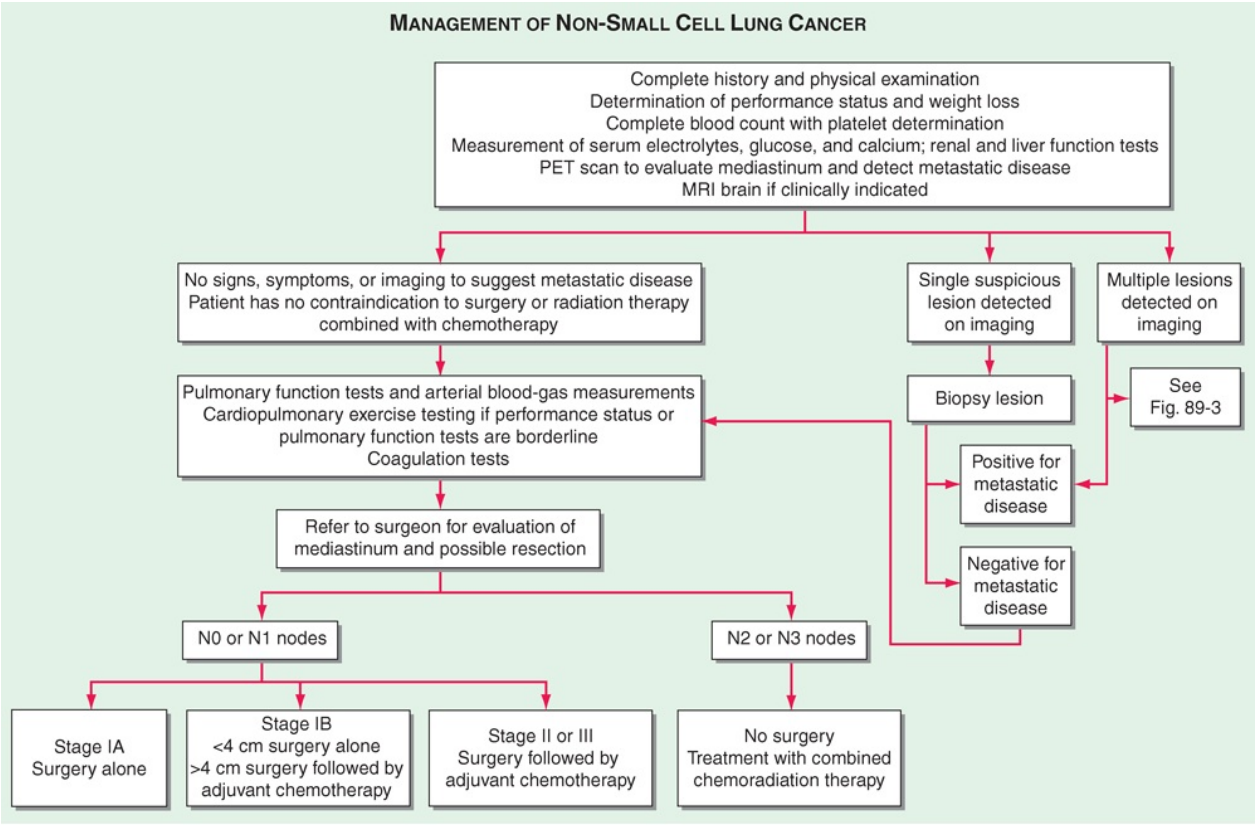

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for management of non-small cell lung cancer.

In patients with NSCLC, the following are major contraindications to potential curative resection: extrathoracic metastases; superior vena cava syndrome; vocal cord and, in most cases, phrenic nerve paralysis; malignant pleural effusion; cardiac tamponade; tumor within 2 cm of the carina (potentially curable with combined chemoradiotherapy); metastasis to the contralateral lung; metastases to supraclavicular lymph nodes; contralateral mediastinal node metastases (potentially curable with combined chemoradiotherapy); and involvement of the main pulmonary artery. In situations where it will make a difference in treatment, abnormal scan findings require tissue confirmation of malignancy so that patients are not precluded from having potentially curative surgery.

The best predictor of metastatic disease remains a careful history and physical examination. If signs, symptoms, or findings from physical examination suggest the presence of malignancy, then sequential imaging starting with the most appropriate study should be performed. If the findings from the clinical evaluation are negative, then imaging studies beyond CT-PET are unnecessary and the search for metastatic disease is complete. More controversial is how one should assess patients with known stage III disease. Because these patients are more likely to have asymptomatic occult metastatic disease, current guidelines recommend a more extensive imaging evaluation, including imaging of the brain with either CT scan or MRI. In patients in whom distant metastatic disease has been ruled out, lymph node status needs to be assessed via a combination of radiographic imaging and/or minimally invasive techniques such as those mentioned above and/or invasive techniques such as mediastinoscopy, mediastinotomy, thoracoscopy, and thoracotomy. About a quarter to a half of patients diagnosed with NSCLC will have mediastinal lymph node metastases at the time of diagnosis. Lymph node sampling is recommended in all patients with enlarged nodes detected by CT or PET scan and in patients with large tumors or tumors occupying the inner third of the lung. The extent of mediastinal lymph node involvement is important in determining the appropriate treatment strategy: surgical resection followed by adjuvant chemotherapy versus combined chemoradiotherapy alone (see below). A standard nomenclature for referring to the location of lymph nodes involved with lung cancer has evolved (Fig. 89-5).

Figure 89-5

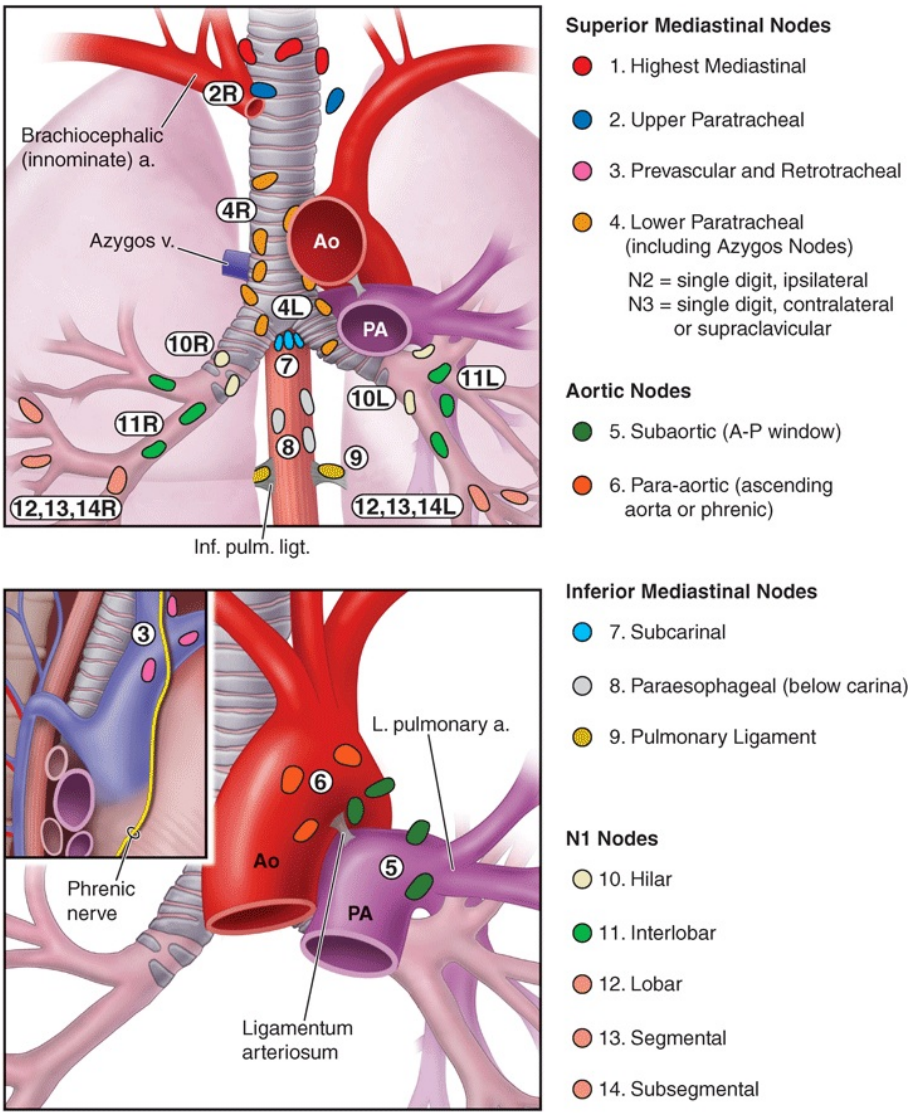

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lymph node stations in staging non-small cell lung cancer.

There are limited data on the use of CT-PET in the staging of patients with SCLC (Fig. 89-6). Current staging recommendations include a CT scan of the chest and abdomen (because of the high frequency of hepatic and adrenal involvement), MRI of the brain (positive in 10% of asymptomatic patients), and radionuclide (bone) scan if symptoms or signs suggest disease involvement in these areas. Bone marrow biopsies and aspirations are rarely performed given the low incidence of isolated bone marrow metastases. Confirmation of metastatic disease, ipsilateral or contralateral lung nodules, or metastases beyond the mediastinum may be achieved by the same modalities recommended above for patients with NSCLC.

Figure 89-6

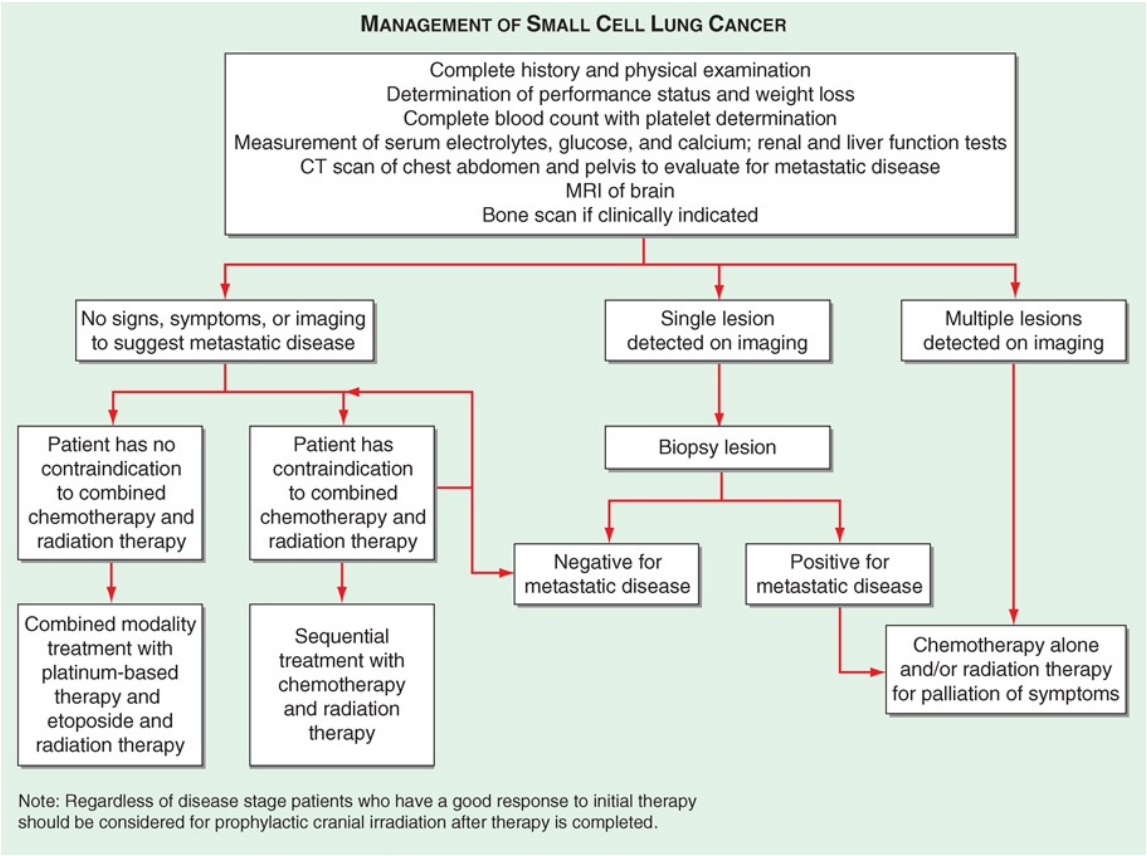

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for management of small cell lung cancer.

If a patient has signs or symptoms of spinal cord compression (pain, weakness, paralysis, urinary retention), a spinal CT or MRI scan and examination of the cerebrospinal fluid cytology should be performed. If metastases are evident on imaging, a neurosurgeon should be consulted for possible palliative surgical resection and/or a radiation oncologist should be consulted for palliative radiotherapy to the site of compression. If signs of symptoms of leptomeningitis develop at any time in a patient with lung cancer, an MRI of the brain and spinal cord should be performed as well as a spinal tap for detection of malignant cells. If the spinal tap is negative, a repeat spinal tap should be considered. There is currently no approved therapy for the treatment of leptomeningeal disease.

The Staging System for Non-Small Cell Lung Cancer

The TNM International Staging System provides useful prognostic information and is used to stage all patients with NSCLC. The various T (tumor size), N (regional node involvement) and M (presence or absence of distant metastasis) are combined to form different stage groups (Table 89–5). The former tumor node metastasis (TNM) staging system for lung cancer (sixth edition) was developed based on a relatively small database of patients from a single institution. In 1999, the International Association for the Study of Lung Cancer established the lung cancer staging project and collected data on more than 68,000 cases from 46 sources in more than 19 countries to develop the new TNM (seventh edition) staging system, which has come into use as of 2010. As seen in Tables 89–5 and 89–6, the major distinction between the sixth and seventh edition staging system is within the T classification; T1 tumors are divided into tumors ≤2 cm in size, as these patients were found to have a better prognosis compared to tumors >2 cm but ≤3 cm. T2 tumors are divided into those that are >3 cm but ≤5 cm and those that are >5 cm but ≤7 cm. T3 tumors are >7 cm. T4 tumors include those that have additional nodules in the same lobe or tumors that have a malignant pleural effusion. No changes have been made to the current classification of lymph node involvement (N). Patients with metastasis may be classified as M1a, malignant pleural or pericardial effusion, pleural nodules or nodules in the contralateral lung, or M1b distant metastasis (e.g., bone, liver, adrenal, or brain metastasis). Based on these data, approximately one-third of patients have localized disease that can be treated with curative attempt (surgery or radiotherapy), one-third have local or regional disease that may or may not be amenable to a curative attempt, and one-third have metastatic disease at the time of diagnosis.

Table 89–5 Comparison of the Sixth and Seventh Edition TNM Staging Systems for Non-Small Cell Lung Cancer

	Sixth Edition	Seventh Edition
Tumor (T)		
T1	Tumor ≤3 cm diameter without invasion more proximal than lobar bronchus	Tumor ≤2 cm diameter, surrounded by lung or visceral pleura, without invasion more proximal than lobar bronchus
T1a		Tumor ≤2 cm in diameter
T1b		Tumor >2 cm but ≤3 cm in diameter
T2	Tumor >3 cm diameter OR tumor of any size with any of the following: Visceral pleural invasion Atelectasis of less than entire lung Proximal extent at least 2 cm from carina	Tumor >3 cm but ≤7 cm with any of the following: Involves main bronchus, ≥2 cm distal to carina Invades visceral pleura Associated with atelectasis or obstructive pneumonitis extending to hilar region but not involving the entire lung

T2a		Tumor >3 cm but ≤5 cm in diameter
T2b		Tumor >5 cm but ≤7 cm in diameter
T3	Tumor of any size that invades any of the following: chest wall, diaphragm, mediastinal pleura, parietal pericardium Tumor <2 cm distal to carina	Tumor >7 cm or directly invades any of the following: chest wall (including superior sulcus tumors), phrenic nerve, mediastinal pleura, parietal pericardium Tumor <2 cm distal to carina but without involvement of carina Tumor with associated atelectasis or obstructive pneumonitis of entire lung Separate tumor nodule(s) in same lobe
T4	Tumor of any size that invades any of the following: mediastinum, heart or great vessels, trachea, esophagus, vertebral body, carina Tumor with malignant pleural or pericardial effusion Separate tumor nodules in same lobe	Tumor of any size that invades any of the following: mediastinum, heart or great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina Separate tumor nodule(s) in a different ipsilateral lobe
Nodes (N)		
N0	No regional lymph node metastasis	No regional lymph node metastasis
N1	Metastasis in ipsilateral peribronchial and/or hilar lymph node(s)	Metastasis in ipsilateral peribronchial and/or hilar lymph node(s) and intrapulmonary node(s), including involvement by direct extensions
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene or supraclavicular lymph node(s)	Metastasis in contralateral mediastinal, hilar, ipsilateral or contralateral scalene or supraclavicular lymph node(s)
Metastasis (M)		
M0	No distant metastasis	No distant metastasis
M1	Distant metastasis (includes tumor nodules in different lobe from primary)	Distant metastasis
M1a		Separate tumor nodules in a contralateral lobe Tumor with pleural nodules or malignant pleural or pericardial effusion
M1b		Distant metastasis

Source: Reproduced with permission from P Goldstraw et al: J Thorac Oncol 2:706, 2007.

Table 89–6 Comparison of Survival by Stage in TNM Sixth and Seventh Editions

Stage	TNM Sixth Edition	TNM Seventh Edition	5-Year Survival (%)*
IA	T1N0M0	T1a-T1bN0M0	73
IB	T2N0M0	T2aN0M0	58
IIA	T1N1M0	T1a-T2aN1M0 or	46
IIB	T2N1M0 or T3N0M0	T2bN0M0 T2bN1M0 or T3N0M0	36
IIIA	T3N1M0 or	T1a-T3N2M0 or	24
IIIB	T1-3N2M0 Any T N3M0 T4 Any N M0	T3N1M0 or T4N0-1M0 T4N2M0 or T1a-T4N3M0	9
IV	Any T Any N M1	Any T Any N M1a or M1b	13

*Survival according to the seventh edition.

The Staging System for Small Cell Lung Cancer

Small cell lung cancer has a distinct two-stage system. Patients with limited-stage disease (LD) have cancer that is confined to the ipsilateral hemithorax and can be encompassed within a tolerable radiation port. Thus, contralateral supraclavicular nodes, recurrent laryngeal nerve involvement, and superior vena caval obstruction can all be part of limited-stage disease. Patients with extensive-stage disease (ED) have overt metastatic disease by imaging or physical examination. Cardiac tamponade, malignant pleural effusion, and bilateral pulmonary parenchymal involvement generally qualify disease as extensive-stage, because the involved organs cannot be encompassed safely or effectively within a single radiation therapy port. Sixty to 70% of patients are diagnosed with extensive disease at presentation.

Physiologic Staging

Patients with lung cancer often have other comorbid conditions related to smoking, including cardiovascular disease and COPD. To improve their preoperative condition, correctable problems (e.g., anemia, electrolyte and fluid disorders, infections, cardiac disease, and arrhythmias) should be addressed, appropriate chest physical therapy instituted, and patients should be encouraged to stop smoking. Since it is not always possible to predict whether a lobectomy or pneumonectomy will be required until the time of operation, a conservative approach is to restrict surgical resection to patients who could potentially tolerate a pneumonectomy. Patients with an FEV_1 (forced expiratory volume in 1 s) of greater than 2 L or greater than 80% of predicted

can tolerate a pneumonectomy, and those with an FEV_1 greater than 1.5 L have adequate reserve for a lobectomy. In patients with borderline lung function but a resectable tumor, cardiopulmonary exercise testing could be performed as part of the physiologic evaluation. This test allows an estimate of the maximal oxygen consumption (Vo_2max). A Vo_2max <15 mL/(kg·min) predicts for a higher risk of postoperative complications. Patients deemed unable to tolerate lobectomy or pneumonectomy from a pulmonary functional standpoint may be candidates for more limited resections, such as wedge or anatomic segmental resection, although such procedures are associated with significantly higher rates of local recurrence and a trend toward decreased overall survival. All patients should be assessed for cardiovascular risk using American College of Cardiology and American Heart Association guidelines. A myocardial infarction within the past 3 months is a contraindication to thoracic surgery because 20% of patients will die of reinfarction. An infarction in the past 6 months is a relative contraindication. Other major contraindications include uncontrolled arrhythmias, an FEV_1 of less than 1 L, CO_2 retention (resting PCO_2 >45 mmHg), DL_{CO} <40%, and severe pulmonary hypertension.

Treatment: Non-Small Cell Lung Cancer

The overall treatment approach to patients with NSCLC is shown in Fig. 89-4.

MANAGEMENT OF OCCULT AND STAGE 0 CARCINOMAS

Patients with severe atypia on sputum cytology have an increased risk of developing lung cancer compared to those without atypia. In the uncommon circumstance where malignant cells are identified in a sputum or bronchial washing specimen but the chest imaging appears normal (TX tumor stage), the lesion must be localized. More than 90% of tumors can be localized by meticulous examination of the bronchial tree with a fiberoptic bronchoscope under general anesthesia and collection of a series of differential brushings and biopsies. Surgical resection following bronchoscopic localization improves survival compared to no treatment. Close follow-up of these patients is indicated because of the high incidence of second primary lung cancers (5% per patient per year).

SOLITARY PULMONARY NODULE AND "GROUND-GLASS" OPACITIES

A solitary pulmonary nodule is defined as an x-ray density completely surrounded by normal aerated lung with circumscribed margins, of any shape, usually 1–6 cm at greatest diameter. The approach to a patient with a solitary pulmonary nodule is based on an estimate of the probability of cancer, determined according to the patient's smoking history, age, and characteristics on imaging (Table 89–7). Prior chest x-rays and CT scans should be obtained if available for comparison. A PET scan may be useful if the lesion is greater than 7–8 mm in diameter. If no diagnosis is apparent, Mayo Clinic investigators reported that clinical characteristics (age, cigarette smoking status, and prior cancer diagnosis) and three radiologic characteristics (nodule diameter, spiculation, and upper lobe location) were independent predictors of malignancy. At present, only two radiographic criteria are thought to predict the benign nature of a solitary pulmonary nodule: lack of growth over a period >2 years and certain characteristic patterns of calcification. Calcification alone, however, does not exclude malignancy; a dense central nidus, multiple punctate foci, and "bull's-eye" (granuloma) and "popcorn ball" (hamartoma) calcifications are highly suggestive of a benign lesion. In contrast, a relatively large lesion, lack of or asymmetric calcification, chest symptoms, associated atelectasis, pneumonitis, or growth of the lesion revealed by comparison with an old x-ray or CT scan or a positive PET scan are suggestive of a malignant process and warrant further attempts to establish a histologic diagnosis. An algorithm for assessing these lesions is shown in Fig. 89-3.

Table 89–7 Assessment of Risk of Cancer in Patients with Solitary Pulmonary Nodules

	Risk		
Variable	**Low**	**Intermediate**	**High**
Diameter (cm)	<1.5	1.5–2.2	≥2.3
Age (years)	<45	45–60	>60
Smoking status	Never smoker	Current smoker (<20 cigarettes/d)	Current smoker (>20 cigarettes/d)
Smoking cessation status	Quit ≥7 years ago or quit	Quit <7 years ago	Never quit
Characteristics of nodule margins	Smooth	Scalloped	Corona radiata or spiculated

Source: Reproduced with permission from D Ost et al: N Engl J Med 348:2535, 2003.

Since the advent of screening CTs, small GGOs have often been observed, particularly as the increased sensitivity of CTs enables detection of smaller lesions. Many of these GGOs, when biopsied, are found to be BAC. Some of the GGOs are semiopaque and referred to as "partial" GGOs. These are often more slow-growing and harbor atypical adenomatous hyperplasia histology, and are thought to be precursors to adenocarcinoma. By contrast, "solid" GGOs have faster growth rates and are usually typical adenocarcinoma histologically.

MANAGEMENT OF STAGES I AND II NSCLC

Surgical Resection for Stages I and II NSCLC

Surgical resection by an experienced surgeon is the treatment of choice for patients with clinical stage I or II NSCLC who are able to tolerate the procedure. A retrospective review indicated that operative mortality rates for patients whose tumors were resected by noncardiothoracic or cardiothoracic surgeons were lower compared to general surgeons (5.8% vs 5.6% vs 7.6%, $p = .001$). The extent of resection is a matter of surgical judgment based on findings at exploration. A clinical trial in patients with stage IA NSCLC found that lobectomy was superior to wedge resection in reducing the rate of local recurrence, with a trend toward improvement in overall survival. A retrospective review of the Surveillance, Epidemiology, and End Results (SEER) database also reported a survival benefit for lobectomy compared to wedge resection. A limited resection, wedge resection, and segmentectomy [potentially by video-assisted thoracic surgery (VATS)] may be more appropriate in patients with comorbidities, including compromised pulmonary reserve and small peripheral lesions. Pneumonectomy is reserved for patients with very central tumors and should only be performed in patients with excellent pulmonary reserve. The 5-year survival rates are 60–80% for patients with stage I NSCLC and 40–50% for patients with stage II NSCLC.

Accurate pathologic staging requires adequate segmental, hilar, and mediastinal lymph node sampling. Mediastinal lymph node dissection provides for a significantly larger amount of material, which can refine pathologic (nodal) stage. On the right side, mediastinal stations 2R, 4R, 7, 8R, and 9R should be dissected; on the left side, stations 5, 6, 7, 8L, and 9L should be dissected (Fig. 89-5). Hilar lymph nodes are typically resected and sent with the specimen, although it is helpful to specifically dissect and label level 10 lymph nodes when possible. On the left side, level 2 and sometimes level 4 lymph nodes are generally obscured by the aorta. Although the therapeutic benefit of nodal dissection versus nodal sampling remains

controversial, in a recent pooled analysis of three trials, 4-year survival was superior in patients undergoing resection with stages I–IIIA NSCLC who had complete mediastinal lymph node dissection compared with lymph node sampling. Moreover, a complete mediastinal lymphadenectomy adds little morbidity to a pulmonary resection for lung cancer. Thus, the recommendation at this time is that patients should have a complete mediastinal node dissection.

Radiation Therapy in Stages I and II NSCLC

There is currently no role for adjuvant radiation therapy in patients following resection of stage I or II NSCLC. Patients with stage I or II disease who refuse or are not candidates for pulmonary resection should be considered for radiation therapy with curative intent. The decision to administer high-dose radiotherapy is based on the extent of disease and the volume of the chest that requires radiation. A systematic review reported 5-year survival rates of 13–39% in patients with stage I or II NSCLC treated with radical radiotherapy. Stereotactic radiation therapy and cryoablation are relatively new techniques that are being used in the treatment of patients with isolated pulmonary nodules who are not candidates for or refuse surgical resection, but their use may be limited by tumor size: ≤5 cm for stereotactic radiotherapy and ≤3 cm for cryoablation therapy.

Chemotherapy in Stages I and II NSCLC

A multitude of trials have evaluated the role of adjuvant chemotherapy in patients with resected stage IA–IIIA NSCLC with conflicting results (Table 89–8). A meta-analysis, the Lung Adjuvant Cisplatin Evaluation Study (LACE), reported a 5.4% improvement in 5-year survival for adjuvant chemotherapy compared to surgery alone in patients with stage I–IIIA NSCLC. The effect of cisplatin plus vinorelbine appeared marginally better than other cisplatin-based doublet regimens. The analysis, however, reported a harmful effect for chemotherapy in patients with stage IA disease, with questionable benefit in patients with stage IB disease. Chemotherapy also appeared to be detrimental in patients with poor performance status (ECOG PS2). The results of these studies have led to the recommendation for adjuvant chemotherapy only in patients with stage II or III NSCLC. Chemotherapy should start 6 to 8 weeks after surgery, if the patient has recovered, and should be administered for four cycles. All patients should be treated with a cisplatin-based regimen. Carboplatin is a reasonable consideration in patients who are unlikely to tolerate cisplatin for reasons such as reduced renal function, presence of neuropathy, or hearing impairment.

Table 89–8 Adjuvant Chemotherapy Trials in Non-Small Cell Lung Cancer

Trial	Stage	Treatment	N	5-Year Survival	*p*
IALT	I–III	Cisplatin-based	932	44.5	<.03
		Control	835	40.4	
BR10	IB–II	Cisplatin + vinorelbine	242	69	.03
		Control	240	54	
ANITA	IB–IIIA	Cisplatin + vinorelbine	407	60	.017
		Control	433	58	
ALPI	I–III	MVP	548	50	.49
		Control	540	45	
BLT	I–III	Cisplatin-based	192	60	.90
		Control	189	58	
CALGB	IB	Carboplatin + paclitaxel	173	59	.10
			171	57	

Abbreviation: IALT, International Adjuvant Lung Cancer Trial; ANITA, Adjuvant Navelbine International Trialist Association; ALPI, Adjuvant Lung Cancer Project Italy; BLT, Big Lung Trial; CALGB, Cancer and Lung Cancer Group B; MVP, mitomycin, vindesine, and cisplatin.

The treatment of patients with stage IB NSCLC remains controversial. Retrospective subset analyses of randomized phase III trials have reported no benefit for adjuvant chemotherapy in patients with stage IB disease. The only trial to evaluate adjuvant chemotherapy in patients with stage IB NSCLC reported no improvement in overall survival; however, a retrospective analysis of the trial reported a benefit in patients with tumors that were ≥4 cm. At this time, the risks and benefits of chemotherapy should be considered on an individual patient basis.

Four trials have evaluated neoadjuvant chemotherapy (chemotherapy before surgery) in patients with stage I–III NSCLC, of which three reported a trend toward improvement in progression-free and overall survival. However, at this time, no data support the use of neoadjuvant chemotherapy in NSCLC patients.

All patients with resected NSCLC are at high risk of recurrence or developing a second primary lung cancer. Thus, it is reasonable to follow these patients with regular imaging. The most appropriate modality and frequency has not been defined. Given that the majority of patients recur within the first 2 years after therapy, one guideline suggests CT scans of the chest with contrast every 6 months for the first 2 years after surgery, followed by yearly CT scans of the chest without contrast thereafter.

MANAGEMENT OF STAGE III NSCLC

The interpretation of the results of clinical trials involving patients with stage III NSCLC has been clouded by a number of issues, including changing diagnostic techniques, different staging systems, and heterogeneous patient populations. In prior studies, patients may have had tumors ranging from nonbulky stage IIIA (clinical N1 nodes with N2 nodes discovered only at the time of surgery, despite a negative mediastinoscopy) to bulky N2 nodes (lymph nodes >2 cm clearly visible on imaging, or multilevel ipsilateral mediastinal nodes) to clearly inoperable nodes.

Surgery followed by adjuvant chemotherapy is the treatment of choice for patients with stage IIIA disease due to hilar nodal involvement (T3N1). Surgery for N2 disease is more controversial. A randomized phase III trial demonstrated an improvement in progression-free survival but no improvement in overall survival when patients with pathologically staged N2 NSCLC were treated with concurrent chemoradiotherapy (cisplatin and etoposide) and 45 Gy of radiation followed by surgery compared to chemotherapy and 61 Gy of radiotherapy without surgery. Treatment-related mortality rate greater in the surgery arm (8% vs 2%), with the majority of deaths occurring in patients undergoing pneumonectomy. In subset analysis, the investigators found survival was improved if a lobectomy was performed but not pneumonectomy compared to chemoradiotherapy alone.

In spite of a careful preoperative staging evaluation, as many as a quarter of patients will be found to have metastases to N2

nodes on frozen-section examination at the time of thoracotomy or on final pathologic examination of the surgical specimen. For patients with an occult, single-station mediastinal node metastasis recognized at thoracotomy in which a complete resection of the nodes and primary tumor is technically possible, most thoracic surgeons proceed with the planned lung resection and a mediastinal lymphadenectomy. If a complete resection is not possible or there is multistation or bulky nodal disease or extracapsular nodal disease, then the planned lung resection should be aborted. These patients can then be considered for combined chemoradiotherapy as described below. Although incomplete resection rarely results in long-term survival, collected results indicate that surgery alone in stage IIIA disease (N2 disease) is associated with a 14–30% 5-year survival. The best survival rate is seen in cases with minimal N2 disease and complete resection.

Chemotherapy plus radiation therapy is the treatment of choice for patients with N3 nodal involvement or bulky stage IIIA disease. In general, patients with histologically involved lymph nodes >2 cm in short-axis diameter measured by CT, who have extranodal involvement or multistation disease along with groups of multiple smaller lymph nodes involved, are considered to have bulky, unresectable disease. Randomized phase III trials initially demonstrated an improvement in median and long-term survival for chemotherapy followed by radiation therapy, compared with radiation therapy alone. Subsequent trials demonstrated administering concurrent chemotherapy and radiation therapy results in improved survival compared to sequential therapy, albeit with more side effects, such as fatigue, esophagitis, and neutropenia. Therefore, combined modality treatment with chemotherapy and radiation therapy is recommended in patients who are able to tolerate the treatment.

Superior Sulcus or Pancoast Tumors

Superior sulcus tumors arise in the apex of the lung and invade adjacent structures producing Pancoast's syndrome: Horner's syndrome, shoulder and/or arm pain, and weakness and atrophy of the muscles of the hand. Patients with these tumors should undergo the same staging procedures as all patients with stage II or III NSCLC. Neoadjuvant chemotherapy or combined chemotherapy and radiation therapy is typically reserved for those patients with N0 or N1 involvement. This approach results in a 33-month median survival and 44% 5-year survival for all patients, and a 94-month median survival and 54% 5-year survival in patients with an R0 resection. For patients with Pancoast tumors that have metastatic disease at the time of presentation, radiation therapy with or without chemotherapy may be offered for palliation of symptoms.

TREATMENT OF METASTATIC NON-SMALL CELL LUNG CANCER

Approximately two-thirds of NSCLC patients present with advanced disease (stage IIIB with a pleural effusion or stage IV) at the time of diagnosis. These patients have a median survival of 4–5 months and a 1-year survival of 10% when managed with best supportive care alone. In addition, a significant number of patients who present with early-stage NSCLC eventually relapse with distant disease. Patients who have recurrent disease have a better prognosis than those presenting with metastatic disease at the time of diagnosis. Standard medical management, the judicious use of pain medications, and the appropriate use of radiotherapy and chemotherapy form the cornerstone of management.

Chemotherapy palliates symptoms, improves the quality of life, and improves survival in patients with stage IV NSCLC, particularly in patients with good performance status. In addition, economic analysis has found chemotherapy to be cost-effective palliation for stage IV NSCLC. However, the use of chemotherapy for NSCLC requires clinical experience and careful judgment to balance potential benefits and toxicities.

First-Line Chemotherapy for Metastatic or Recurrent Non-Small Cell Lung Cancer

The first indication of the benefit of chemotherapy in patients with advanced NSCLC came from a meta-analysis published in 1995 that reported a survival advantage in patients treated with cisplatin-based chemotherapy compared to those receiving supportive care alone (HR = 0.73, $p < .0001$). This led to a multitude of clinical trials comparing different cisplatin-based regimens in patients with advanced NSCLC all reporting a similar magnitude of benefit; 20–30% response rate and an 8- to 10-month median survival (Table 89–9). Chemotherapy was well tolerated in all studies in patients with a good performance status, ECOG PS 0–1.

Table 89–9 First-Line Chemotherapy Trials for Metastatic Non-Small Cell Lung Cancer

Trial	Regimen	*N*	RR (%)	Median Survival (months)
ECOG1594	Cisplatin + paclitaxel	288	21	7.8
	Cisplatin + gemcitabine	288	22	8.1
	Cisplatin + docetaxel	289	17	7.4
	Carboplatin + paclitaxel	290	17	8.1
TAX-326	Cisplatin + docetaxel	406	32	11.3
	Cisplatin + vinorelbine	394	25	10.1
	Carboplatin + docetaxel	404	24	9.4
EORTC	Cisplatin + paclitaxel	159	32	8.1
	Cisplatin + gemcitabine	160	37	8.9
	Paclitaxel + gemcitabine	161	28	6.7
ILCP	Cisplatin + gemcitabine	205	30	9.8
	Carboplatin + paclitaxel	204	32	9.9
	Cisplatin + vinorelbine	203	30	9.5
SWOG	Cisplatin + vinorelbine	202	28	8.0
	Carboplatin + paclitaxel	206	25	8.0
FACS	Cisplatin + irinotecan	145	31	13.9
	Carboplatin + paclitaxel	145	32	12.3
	Cisplatin + gemcitabine	146	30	14.0
	Cisplatin + vinorelbine	145	33	11.4
Scagliotti	Cisplatin + gemcitabine	863	28	10.3
	Cisplatin + pemetrexed	862	31	10.3
iPASS*	Carboplatin + paclitaxel	608	32	17.3

	Gefitinib	609	43	18.6

*Enrolled selected patients: 18 years of age or older, had histologic or cytologically confirmed stage IIIB or IV non-small cell lung cancer with histologic features of adenocarcinoma (including bronchioloalveolar carcinoma), were nonsmokers (defined as patients who had smoked <100 cigarettes in their lifetime) or former light smokers (those who had stopped smoking at least 15 years previously and had a total of ≤10 pack-years of smoking), and had had no previous chemotherapy or biologic or immunologic therapy.

Abbreviation: ECOG, Eastern Cooperative Oncology Group; EORTC, European Organization for Research and Treatment of Cancer; ILCP, Italian Lung Cancer Project; SWOG, South-Western Oncology Group; FACS, Follow-up After Colorectal Surgery; iPASS, Iressa Pan-Asian Study.

An ongoing debate in the treatment of patients with NSCLC is the appropriate duration of platinum-based chemotherapy. Several large phase III randomized trials have failed to show a benefit for increasing the duration of platinum-based doublet chemotherapy beyond four to six cycles. In fact, longer duration of chemotherapy has been associated with increased toxicities and impaired quality of life. Therefore, prolonged therapy (beyond four to six cycles) with platinum-based regimens is not recommended in patients with advanced NSCLC.

Tumor histology has emerged as an important consideration in the treatment of patients with NSCLC. A randomized phase III trial found that patients with nonsquamous NSCLC had an improved survival when treated with cisplatin and pemetrexed compared to cisplatin and gemcitabine, while patients with squamous carcinoma had an improved survival when treated with cisplatin and gemcitabine. This difference in survival is thought to be related to the differential expression of thymidylate synthase, one of the targets of pemetrexed, between tumor types. Bevacizumab, a monoclonal antibody against VEGF, when combined with chemotherapy improves response rate, progression-free survival, and overall survival in patients with advanced disease (see below). However, bevacizumab cannot be given to patients with squamous cell histology NSCLC because of the risk of serious hemorrhagic effects.

Second-Line Chemotherapy and beyond

As first-line chemotherapy regimens improve, a substantial number of patients will maintain a good performance status and a desire for further therapy when they develop recurrent disease. At present only three drugs are FDA-approved for second-line therapy of NSCLC in the United States, i.e., docetaxel, pemetrexed, and erlotinib. In general, these agents have similar overall response rates of 5–10% (depending on the patient's prior exposure to taxanes and platinum) and yield median survivals of 6–8 months. However, the available drugs have distinct toxicity profiles that can influence their use in the second-line setting. Hematologic toxicity including febrile neutropenia is greater for docetaxel compared with pemetrexed and erlotinib, whereas nonhematologic toxicity, namely rash and diarrhea, is greater with erlotinib. Most of the survival benefit for any of these agents is realized in those patients who maintain a good performance status.

AGENTS THAT INHIBIT ANGIOGENESIS

Bevacizumab was the first antiangiogenic agent approved for the treatment of patients with advanced NSCLC in the United States. This drug primarily acts by sponging up VEGF and blocking the growth of new blood vessels, which are required for tumor viability. Two randomized phase III trials of chemotherapy with or without bevacizumab had conflicting results. The first trial, conducted in North America, compared carboplatin/paclitaxel with or without bevacizumab in patients with recurrent or advanced nonsquamous NSCLC and reported a significant improvement in response rate, progression-free survival, and overall survival for chemotherapy-plus-bevacizumab–treated patients compared to chemotherapy alone. Toxicities were more frequent in bevacizumab-treated patients. The second trial, conducted in Europe, compared cisplatin/gemcitabine with or without bevacizumab in patients with recurrent or advanced nonsquamous NSCLC and reported a significant improvement in progression-free survival but no improvement in overall survival for bevacizumab-treated patients. Therefore, at this time carboplatin/paclitaxel and bevacizumab is an approved regimen for first-line treatment of nonsquamous NSCLC in the United States but not in Europe.

AGENTS THAT INHIBIT THE EPIDERMAL GROWTH FACTOR RECEPTOR

Erlotinib and gefitinib are oral small-molecule kinase inhibitors that inhibit signaling via EGFR. These were the first EGFR inhibitors to be approved for the treatment of patients with NSCLC. A randomized phase III trial compared erlotinib to placebo in previously treated patients with advanced NSCLC and reported an improvement in overall survival for erlotinib compared to placebo. Gefitinib received premarketing approval by the FDA after impressive results seen in phase II trials in patients with previously treated NSCLC; however, a randomized phase III trial found no difference in overall survival between patients treated with gefitinib compared to placebo. These results led to a U.S. FDA-mandated change in the gefitinib indication to include only patients who have previously benefited from this drug. However, gefitinib is still available for the treatment of NSCLC patients in Europe and Asia. Clinical features that have been shown to correlate with responsiveness to EGFR TKI treatment include female sex, never smoking status, adenocarcinoma histology, and Asian ethnicity. Somatic mutations in the kinase domain of EGFR and high EGFR copy number have also been shown to correlate with response and improved survival with oral EGFR inhibitors.

Two randomized phase III trials conducted in Asia have compared gefitinib to platinum-based chemotherapy in patients with NSCLC. The first trial compared first-line gefitinib to carboplatin/paclitaxel in never or light ex-smokers with newly diagnosed advanced NSCLC. Treatment with gefitinib was associated with a significant improvement in response rate and 12-month progression-free survival. In patients with tumors available for mutation analysis, treatment with gefitinib was favored over chemotherapy in patients with tumors that harbored an EGFR mutation and chemotherapy was favored in patients with tumors that were EGFR mutation negative. Quality of life favored treatment with gefitinib. The second trial enrolled only patients with tumors that were EGFR mutation positive and reported a significant improvement in progression-free survival and disease control for patients treated with gefitinib compared to cisplatin/docetaxel. These and related results suggest standard chemotherapy regimens or gefitinib and erlotinib could be considered for first-line therapy in a subset of advanced NSCLC patients with tumors that harbor the EGFR mutation.

Cetuximab is an intravenously administered chimeric antibody directed against EGFR. A randomized phase III trial evaluated treatment with cisplatin/vinorelbine with or without cetuximab in patients with advanced NSCLC and at least one EGFR-positive cell as determined by immunohistochemistry. The results showed no difference in progression-free survival but a significant improvement in response rate and overall survival in patients treated with cetuximab compared to placebo. A prespecified subgroup analysis showed no improvement in overall survival among patients of Asian ethnicity receiving cetuximab compared with placebo. However, a significant improvement in overall survival was noted among Caucasian patients receiving cetuximab; this appeared true regardless of histology. Contrary to patients with colon cancer, KRAS mutation status did not predict response to therapy with cetuximab, although the number of cases examined at the molecular level was suboptimal. Development of acneiform rash was associated with improved overall survival compared to patients with no rash. A second phase III trial in patients with advanced NSCLC with no required EGFR testing reported no difference in overall survival between patients randomized to carboplatin/paclitaxel or docetaxel with or without cetuximab.

MAINTENANCE THERAPY

Maintenance chemotherapy in nonprogressing patients (patients with a complete response, partial response, or stable disease) is a controversial topic in the treatment of NSCLC patients. Two studies have investigated maintenance single-agent chemotherapy with docetaxel or pemetrexed in nonprogressing patients following treatment with first-line platinum-based chemotherapy. Both trials randomized patients to immediate single-agent therapy versus observation and reported improvements in progression-free and overall survival. In both trials, a significant portion of patients in the observation arm did not receive therapy with the agent under investigation upon disease progression; 37% of study patients never received docetaxel in the docetaxel study and 81% of patients never received pemetrexed in the pemetrexed study. In the trial of maintenance docetaxel versus observation, survival was identical to the treatment group in the subset of patients who received docetaxel on progression, indicating this is an active agent in NSCLC. These data are not available for the pemetrexed study. Currently maintenance pemetrexed is the only therapy approved by the U.S. FDA following platinum-based chemotherapy in patients with advanced NSCLC. However, maintenance chemotherapy is not without toxicity and at this time should be considered on an individual patient basis.

Two randomized controlled trials have reported improvements in progression-free survival from maintenance treatment with erlotinib compared to placebo in patients with advanced NSCLC following platinum-based chemotherapy.

Treatment: Small Cell Lung Cancer

TREATMENT OF LIMITED DISEASE SMALL CELL LUNG CANCER

Surgery

SCLC is a highly aggressive disease characterized by its rapid doubling time, high growth fraction, early development of disseminated disease, and dramatic response to first-line chemotherapy and radiation. Surgical resection is not routinely recommended for patients because even those patients with LD-SCLC still have occult micrometastases. If the histologic diagnosis of SCLC is made in patients on review of a resected surgical specimen, such patients should receive standard SCLC chemotherapy as described below. If one employs classic TNM staging categories, two retrospective series have reported high cure rates for adjuvant chemotherapy following resection in patients with stage I or II SCLC.

Chemotherapy

Chemotherapy significantly prolongs survival in patients with SCLC. Combination chemotherapy with a platinum agent (cisplatin or carboplatin) and etoposide for four to six cycles is the mainstay of treatment and has not changed in almost three decades. Cyclophosphamide, doxorubicin (Adriamycin), and vincristine (CAV) may be an alternative for patients who are unable to tolerate a platinum-based regimen. Despite response rates to first-line therapy as high as 80%, the median survival ranges from 12 to 20 months for patients with LD and from 7 to 11 months for patients with ED. Regardless of disease extent, the majority of patients relapse and develop chemotherapy-resistant disease. Only 6–12% of patients with LD- and 2% of patients with ED-SCLC live beyond 5 years. The prognosis is especially poor for patients who relapse within the first 3 months of therapy; these patients are said to have platinum-*resistant disease*. Patients are said to have *sensitive disease* if they relapse more than 3 months after their initial therapy and are thought to have a somewhat better overall survival. Those patients with sensitive disease are thought to have the greatest potential benefit from second-line chemotherapy. Topotecan is the only FDA-approved agent with modest activity as second-line therapy in patients with SCLC.

Radiation Therapy

Patients with LD-SCLC are treated with combined modality therapy with cisplatin and etoposide chemotherapy and radiation therapy. A retrospective analysis of patients with SCLC treated with once-daily fractionation found improved local control rates as the total dose delivered was increased from 30 to 50 Gy. Chemotherapy when given concurrently with radiation is more effective than sequential chemoradiation but is associated with significantly more esophagitis and hematologic toxicity. The addition of radiation therapy early on is preferred. Twice-daily (hyperfractionated) radiation has been shown to improve survival in patients with LD-SCLC but is associated with higher rates of grade 3 esophagitis and pulmonary toxicity. It is feasible to deliver once-daily radiation therapy doses up to at least 70 Gy when administered concurrently with cisplatin-based chemotherapy. This higher dose of once-daily radiotherapy may be equivalent or superior to the 45-Gy twice-daily radiotherapy dose. Patients should be carefully selected for concurrent chemoradiation therapy based on good performance status and pulmonary reserve.

Prophylactic Cranial Irradiation

Prophylactic cranial irradiation (PCI) should be considered in all patients with LD- and ED-SCLC who have responded to initial therapy. A meta-analysis including 7 trials and 987 patients with LD-SCLC who had achieved a complete remission following primary chemotherapy reported a 5.4% improvement in overall survival for patients treated with PCI. In patients with ED-SCLC who had responded to first-line chemotherapy, PCI reduced the occurrence of symptomatic brain metastases and prolonged disease-free and overall survival compared to no radiation therapy. Long-term toxicities including deficits in cognition have been reported following PCI and are difficult to sort out from the effects of chemotherapy or normal aging.

Molecularly Tailored Lung Cancer Therapy

In the past 40 years, clinical research in lung cancer has demonstrated that surgery, systemic chemotherapy, and radiation therapy can all be used to prolong patient survival and/or improve quality of life. However, conventional approaches, especially those that classify patients according to disease histology alone, appear to have reached a therapeutic plateau of effectiveness. One promising future approach to improve the outcome for patients with lung cancer is tailored therapy based on individualized phenotypic or genotypic tumor characteristics. Such a strategy is based upon an understanding of the molecular underpinnings of the disease, recognizing that although tumors may appear similar at the histologic level, they do differ from individual to individual. It is hoped that better outcomes can be achieved by matching the most appropriate therapy to a patient at the right time.

For example, one subset of lung cancer can be defined by somatic mutations in *EGFR*. *EGFR* mutations are almost exclusively found in lung adenocarcinoma and are more common in females, never smokers compared to former or current smokers, and in East Asians compared to Western populations (30–70% vs 8%). These mutations, primarily in-frame deletions in exon 19 and point mutations in exon 21 (L858R), result in constitutive activation of the receptor and are associated with very high response rates (60–90%) to the specific tyrosine kinase inhibitors gefitinib and erlotinib. Almost all patients with these dramatic responses, however, develop acquired resistance. In about half of patients, resistance can be attributed to the emergence of clones harboring a second-site mutation in exon 20 (T790M), which alters binding of drug to the receptor. About 20% of *EGFR* mutant tumors from patients with acquired resistance display amplification of a gene encoding a different tyrosine kinase, MET. As a result of these findings, many trials are being conducted in these patients using second-generation EGFR inhibitors that can overcome T790M-mediated resistance or MET inhibitors to target MET-dependent cells.

Another subset of lung adenocarcinoma can be defined by EML4-ALK fusion proteins. These translocations arise from a small inversion within chromosome 2p that leads to the formation of a fusion-gene comprising the N terminal of the echinoderm microtubule–associated protein-like 4 (EML4) gene and the intracellular tyrosine kinase domain of the anaplastic lymphoma

kinase (ALK) gene. Patients with lung cancers harboring an ALK fusion protein have demonstrated dramatic responses to small-molecule ALK inhibitors in early clinical trials. The ALK fusion protein is relatively rare, occurring in 3–7% of NSCLCs. Clinical characteristics associated with EML4-ALK–positive lung cancer appears to be a younger age at diagnosis, minimal smoking history, male sex, and adenocarcinoma histology with signet-ring features.

Other biomarkers being explored include molecules that may predict outcomes with conventional chemotherapy. For example, low expression of the DNA repair gene excision repair cross-complementation group 1 (ERCC1) correlates with improved survival after treatment with platinum drugs, whereas tumors that have high expression of ERCC1 are less sensitive to therapy with a platinum agent. In the absence of treatment, lung cancer with low ERCC1 expression has a poorer prognosis.

Ribonucleotide reductase M1 (RRM1) encodes the regulatory subunit of ribonucleotide reductase, the rate-limiting enzyme in DNA synthesis. Ribonucleotide reductase converts ribonucleotide 5-diphosphate to deoxyribonucleotide 5-diphosphate. Notably, gemcitabine, an agent commonly used in the treatment of NSCLC, competes with ribonucleotide 5-diphosphate for incorporation into DNA. Levels of RRM1 expression are significantly and inversely correlated with disease response after two cycles of gemcitabine and carboplatin in patients with locally advanced NSCLC. In addition, low RRM1 mRNA expression levels are associated with a significantly longer median survival compared to high levels.

Thymidylate synthase (TS) catalyzes the methylation of dUMP to dTMP and is the rate-limiting irreversible step in de novo DNA synthesis. TS is one of the targets of the novel folate-based drug pemetrexed, an agent that is FDA-approved as second-line treatment in patients with NSCLC. TS expression is an independent prognostic and predictive factor in several cancers, including lung cancers, and overexpression of TS has been linked to resistance to pemetrexed, an agent commonly employed as second-line treatment in patients with nonsquamous NSCLC. TS mRNA and protein levels are significantly higher in squamous cell carcinomas and small cell carcinomas of the lung as compared with adenocarcinomas. A randomized phase III trial reported that cisplatin plus gemcitabine was more effective in squamous cell carcinomas while, cisplatin plus pemetrexed was found to be more effective in adenocarcinomas and large cell carcinomas. Molecular markers are likely to play an increasing role in helping guide treatment decisions.

BENIGN LUNG NEOPLASMS

Benign tumors account for about 5% of all lung cancers. About half are hamartomas; the lungs are the site of about 90% of all hamartomas. The other half are bronchial adenomas.

Hamartomas

Lung hamartomas are usually peripheral lung masses composed of normal pulmonary tissue components such as smooth muscle and collagen. They are more common in men than in women and have a peak incidence in the 60s. They are often incidental radiographic findings as solitary nodules. They have a pathognomonic "popcorn" pattern of calcification in some cases; however, without such a finding, resection is necessary to rule out malignancy, especially in smokers.

Bronchial Adenomas

These are centrally located slow-growing endobronchial lesions that are generally carcinoid tumors (≥80%; Chap. 350), adenocystic tumors (so called cylindromas, 10–15%), or mucoepidermoid tumors (2–3%). Mean age at presentation is 45 years (range 15–60). Patients often give a history of chronic cough, intermittent hemoptysis, or repeated episodes of airway obstruction with atelectasis, or pneumonias with abscess formation due to endobronchial lesions obstructing the airway. They are usually visible at bronchoscopy but are highly vascular and may bleed profusely after a bronchoscopic biopsy. They are largely curable by surgical resection (local excision), but they may recur locally or become invasive and metastasize. Five-year survival after resection is 95% if the disease is localized. For bronchial adenomas that spread, the course of disease can become highly aggressive, such as SCLC, or somewhat slower, such as carcinoid tumors. Therapy is generally dictated by the pace of the disease.

FURTHER READINGS

Azzoli CG et al: American Society of Clinical Oncology clinical practice guideline update on chemotherapy for stage IV non-small cell lung cancer. J Clin Oncol 27:6251, 2009[PMID: 19917871]

Hanahan D, Weinberg RA: The hallmarks of cancer. Cell 100:57, 2000[PMID: 10647931]

Herbst RS et al: Lung cancer. N Engl J Med 359: 1367, 2008[PMID: 18815398]

Jackman DM, Johnson BE: Small-cell lung cancer. Lancet 366:1385, 2005[PMID: 16226617]

Kwak EL et al: Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 363:1693, 2010[PMID: 20979469]

Milroy R: New American College of Chest Physicians Lung Cancer Guidelines: An important addition to the lung cancer guidelines armamentarium. Chest 132:744, 2007[PMID: 17873186]

Ost D et al: The solitary pulmonary nodule. N Engl J Med 348:2535, 2003[PMID: 12815140]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BREAST CANCER: INTRODUCTION

Breast cancer is a malignant proliferation of epithelial cells lining the ducts or lobules of the breast. In the year 2010, about 180,000 cases of invasive breast cancer and 40,000 deaths will occur in the United States. In addition, about 2000 men will be diagnosed with breast cancer. Epithelial malignancies of the breast are the most common cause of cancer in women (excluding skin cancer), accounting for about one-third of all cancer in women. As a result of improved treatment and earlier detection, mortality rate from breast cancer has begun to decrease very

substantially in the United States. This chapter will not consider rare malignancies presenting in the breast, such as sarcomas and lymphomas, but will focus on the epithelial cancers. Human breast cancer is a clonal disease; a single transformed cell—the product of a series of somatic (acquired) or germ-line mutations—is eventually able to express full malignant potential. Thus, breast cancer may exist for a long period as either a noninvasive disease or an invasive but nonmetastatic disease. These facts have significant clinical ramifications.

GENETIC CONSIDERATIONS

Not more than 10% of human breast cancers can be linked directly to germ-line mutations. Several genes have been implicated in familial cases. The Li-Fraumeni syndrome is characterized by inherited mutations in the p53 tumor-suppressor gene, which lead to an increased incidence of breast cancer, osteogenic sarcomas, and other malignancies. Inherited mutations in *PTEN* have also been reported in breast cancer.

Another tumor-suppressor gene, *BRCA-1*, has been identified at the chromosomal locus 17q21; this gene encodes a zinc finger protein, and the product therefore may function as a transcription factor. The gene appears to be involved in gene repair. Women who inherit a mutated allele of this gene from either parent have at least a 60–80% lifetime chance of developing breast cancer and about a 33% chance of developing ovarian cancer. The risk is higher among women born after 1940, presumably due to promotional effects of hormonal factors. Men who carry a mutant allele of the gene have an increased incidence of prostate cancer and breast cancer. A fourth gene, termed *BRCA-2*, which has been localized to chromosome 13q12, is also associated with an increased incidence of breast cancer in men and women.

Germ-line mutations in *BRCA-1* and *BRCA-2* can be readily detected; patients with these mutations can be counseled appropriately. All women with strong family histories for breast cancer should be referred to genetic screening programs, particularly women of Ashkenazi Jewish descent who have a high likelihood of a specific *BRCA-1* mutation (substitution of adenine for guanine at position 185).

Even more important than the role these genes play in inherited forms of breast cancer may be their role in sporadic breast cancer. A p53 mutation is present in nearly 40% of human breast cancers as an acquired defect. Acquired mutations in *PTEN* occur in about 10% of the cases. *BRCA-1* mutation in sporadic primary breast cancer has not been reported. However, decreased expression of *BRCA-1* mRNA (possibly via gene methylation) and abnormal cellular location of the *BRCA-1* protein have been found in some breast cancers. Loss of heterozygosity of *BRCA-1* and *BRCA-2* suggests that tumor-suppressor activity may be inactivated in sporadic cases of human breast cancer. Finally, increased expression of a dominant oncogene plays a role in about a quarter of human breast cancer cases. The product of this gene, a member of the epidermal growth factor receptor superfamily, is called *erbB2* (HER/2 neu) and is overexpressed in these breast cancers due to gene amplification; this overexpression can contribute to transformation of human breast epithelium and is the target of effective systemic therapy in adjuvant and metastatic disease settings.

EPIDEMIOLOGY

Breast cancer is a hormone-dependent disease. Women without functioning ovaries who never receive estrogen-replacement therapy do not develop breast cancer. The female-male ratio is about 150:1. For most epithelial malignancies, a log-log plot of incidence versus age shows a single-component straight-line increase with every year of life. A similar plot for breast cancer shows two components: a straight-line increase with age but with a decrease in slope beginning at the age of menopause. The three dates in a woman's life that have a major impact on breast cancer incidence are age at menarche, age at first full-term pregnancy, and age at menopause. Women who experience menarche

at age 16 years have only 50–60% of the breast cancer risk of a woman having menarche at age 12 years; the lower risk persists throughout life. Similarly, menopause occurring 10 years before the median age of menopause (52 years), whether natural or surgically induced, reduces lifetime breast cancer risk by about 35%. Women who have a first full-term pregnancy by age 18 years have a 30–40% lower risk of breast cancer compared with nulliparous women. Thus, length of menstrual life—particularly the fraction occurring before first full-term pregnancy—is a substantial component of the total risk of breast cancer. These three factors (menarche, age of first full-term pregnancy, and menopause) can account for 70–80% of the variation in breast cancer frequency in different countries. A meta-analysis has shown that duration of maternal nursing correlates with substantial risk reduction independent of either parity or age at first full-term pregnancy.

International variation in incidence has provided some of the most important clues on hormonal carcinogenesis. A woman living to age 80 years in North America has one chance in nine of developing invasive breast cancer. Asian women have one-fifth to one-tenth the risk of breast cancer of women in North America or Western Europe. Asian women have substantially lower concentrations of estrogens and progesterone. These differences cannot be explained on a genetic basis because Asian women living in a Western environment have sex steroid hormone concentrations and risks identical to those of their Western counterparts. These migrant women, and more notably their daughters, also differ markedly in height and weight from Asian women in Asia; height and weight are critical regulators of age of menarche and have substantial effects on plasma concentrations of estrogens.

The role of diet in breast cancer etiology is controversial. While there are associative links between total caloric and fat intake and breast cancer risk, the exact role of fat in the diet is unproven. Increased caloric intake contributes to breast cancer risk in multiple ways: earlier menarche, later age at menopause, and increased postmenopausal estrogen concentrations reflecting enhanced aromatase activities in fatty tissues. Moderate alcohol intake also increases the risk by an unknown mechanism. Folic acid supplementation appears to modify risk in women who use alcohol but is not additionally protective in abstainers. Recommendations favoring abstinence from alcohol must be weighed against other social pressures and the possible cardioprotective effect of moderate alcohol intake. Chronic low-dose aspirin use also appears associated with a decreased incidence of breast cancer.

Understanding the potential role of exogenous hormones in breast cancer is of extraordinary importance because millions of American women regularly use oral contraceptives and postmenopausal hormone replacement therapy. The most credible meta-analyses of oral contraceptive use suggest that these agents cause a small increased risk of breast cancer. By contrast, oral contraceptives offer a substantial protective effect against ovarian epithelial tumors and endometrial cancers. Hormone replacement therapy (HRT) has a powerful effect on breast cancer risk. Data from the Women's Health Initiative (WHI) trial showed that conjugated equine estrogens plus progestins increased the risk of breast cancer and adverse cardiovascular events but with decreases in bone fractures and colorectal cancer. On balance, there were more negative events with HRT; 6 to 7 years of HRT nearly doubled the risk of breast cancer. A parallel WHI trial with >12,000 women enrolled testing conjugated estrogens alone (ERT in women who have had hysterectomies) showed no significant increase in breast cancer incidence. A meta-analysis of nonrandomized HRT studies suggests that most of the previously attributed benefit of HRT can be accounted for by higher socioeconomic status among users, which is presumably associated with better access to health care and healthier behaviors. Certain potential benefits of HRT were not assessed in WHI. HRT is an area of rapid reevaluation, but it would appear (at least from breast cancer and cardiovascular disease vantage points) that there are serious concerns about long-term HRT use. HRT in women previously diagnosed with breast cancer increases recurrence rates. Rapid decrease in the number of women on HRT has already led to a coincident decrease in breast cancer incidence.

In addition to the other factors, radiation is a risk factor in younger women. Women who have been exposed before age 30 years to radiation in the form of multiple fluoroscopies (200–300 cGy) or treatment for Hodgkin's disease (>3600 cGy) have a substantial increase in risk of breast cancer, whereas radiation exposure after age 30 years appears to have a minimal carcinogenic effect on the breast.

EVALUATION OF BREAST MASSES IN MEN AND WOMEN

Because the breasts are a common site of potentially fatal malignancy in women, examination of the breast is an essential part of the physical examination. Unfortunately, internists frequently do not examine breasts in men, and, in women, they are apt to defer this evaluation to gynecologists. Because of the plausible association between early detection and improved outcome, it is the duty of every physician to identify breast abnormalities at the earliest possible stage and to institute a diagnostic workup. Women should be trained in breast self-examination (BSE). Although breast cancer in men is unusual, unilateral lesions should be evaluated in the same manner as in women, with the recognition that gynecomastia in men can sometimes begin unilaterally and is often asymmetric.

Virtually all breast cancer is diagnosed by biopsy of a nodule detected either on a mammogram or by

palpation. Algorithms have been developed to enhance the likelihood of diagnosing breast cancer and reduce the frequency of unnecessary biopsy (Fig. 90-1).

Figure 90-1

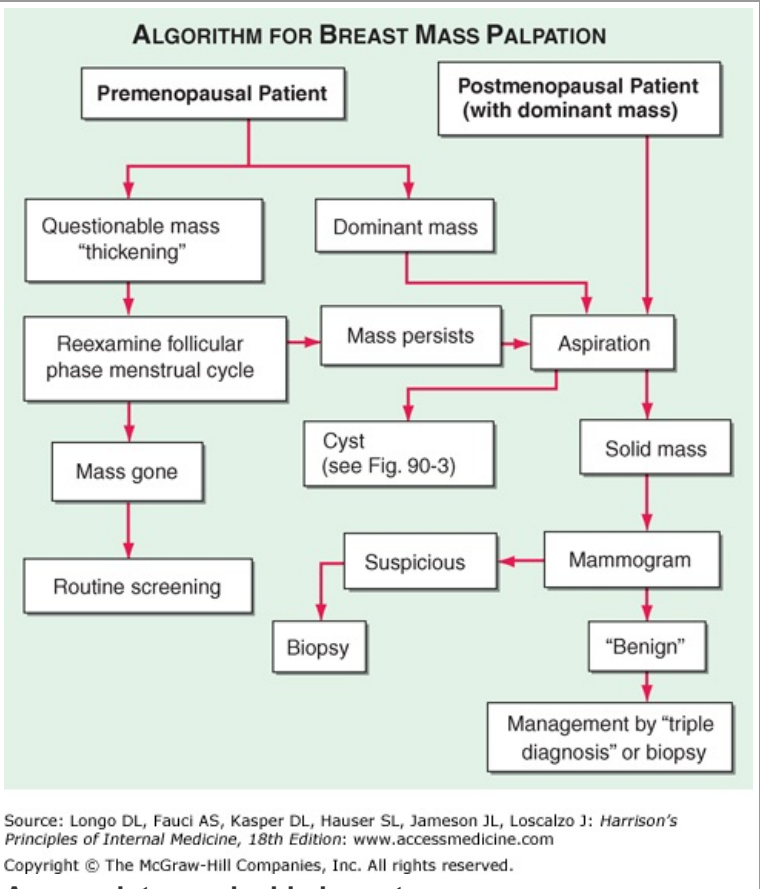

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to a palpable breast mass.

The Palpable Breast Mass

Women should be strongly encouraged to examine their breasts monthly. A potentially flawed study from China has suggested that BSE does not alter survival, but given its safety, the procedure should still be encouraged. At worst, this practice increases the likelihood of detecting a mass at a smaller size when it can be treated with more limited surgery. Breast examination by the physician should be performed in good light so as to see retractions and other skin changes. The nipple and areolae should be inspected, and an attempt should be made to elicit nipple discharge. All regional lymph node groups should be examined, and any lesions should be measured. Physical examination alone cannot exclude malignancy. Lesions with certain features are more likely to be cancerous (hard, irregular, tethered or fixed, or painless lesions). A negative mammogram in the presence of a persistent lump in the breast does not exclude malignancy. Palpable lesions require additional diagnostic procedures, including biopsy.

In premenopausal women, lesions that are either equivocal or nonsuspicious on physical examination should be reexamined in 2–4 weeks, during the follicular phase of the menstrual cycle. Days 5–7 of the cycle are the best time for breast examination. A dominant mass in a postmenopausal woman or a dominant mass that persists through a menstrual cycle in a premenopausal woman should be aspirated by fine-needle biopsy or referred to a surgeon. If nonbloody fluid is aspirated, the diagnosis (cyst) and therapy have been accomplished together. Solid lesions that are persistent, recurrent, complex, or bloody cysts require mammography and biopsy, although in selected patients the so-called triple diagnostic techniques (palpation, mammography, aspiration) can be used to avoid biopsy (Figs. 90-1, 90-2, and 90-3). Ultrasound can be used in place of fine-needle aspiration to distinguish cysts from solid lesions. Not all solid masses are detected by ultrasound; thus, a palpable mass that is not visualized on ultrasound must be presumed to be solid.

Figure 90-2

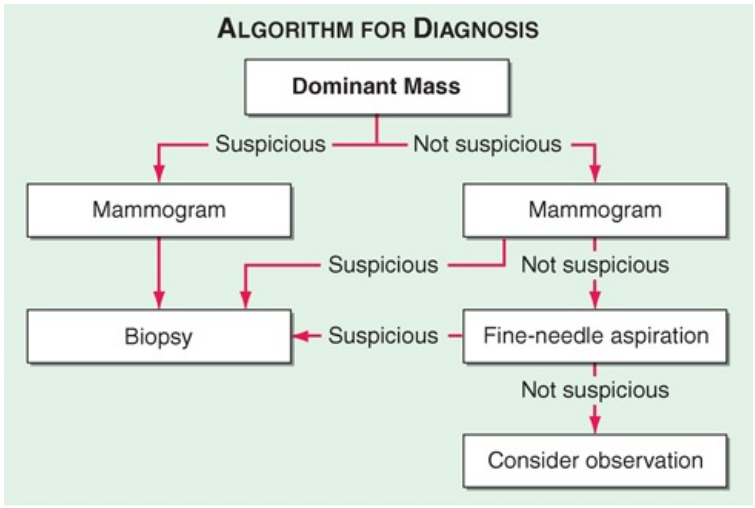

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The "triple diagnosis" technique.

Figure 90-3

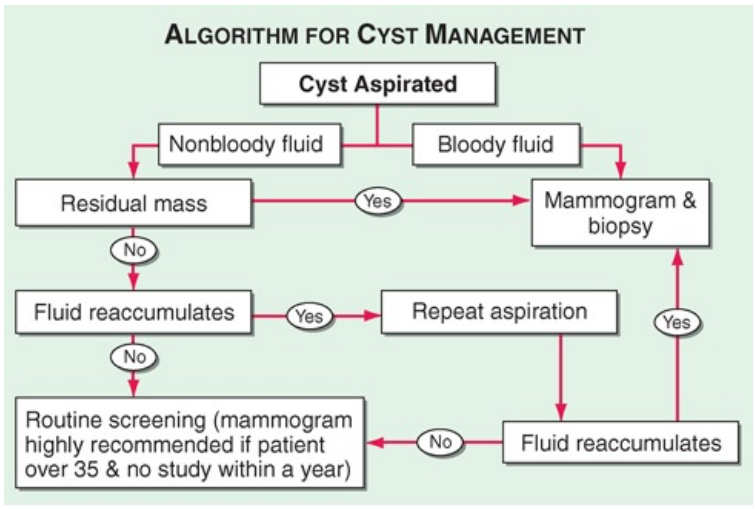

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of a breast cyst.

Several points are essential in pursuing these management decision trees. First, risk-factor analysis is not part of the decision structure. No constellation of risk factors, by their presence or absence, can be used to exclude biopsy. Second, fine-needle aspiration should be used only in centers that have proven skill in obtaining such specimens and analyzing them. The likelihood of cancer is low in the setting of a "triple negative" (benign-feeling lump, negative mammogram, and negative fine-needle aspiration), but it is not zero. The patient and physician must be aware of a 1% risk of false negatives. Third, additional technologies such as MRI, ultrasound, and sestamibi imaging cannot be used to exclude the need for biopsy, although in unusual circumstances they may provoke a biopsy.

The Abnormal Mammogram

Diagnostic mammography should not be confused with *screening mammography*, which is performed after a palpable abnormality has been detected. Diagnostic mammography is aimed at evaluating the rest of the breast before biopsy is performed or occasionally is part of the triple-test strategy to exclude immediate biopsy.

Subtle abnormalities that are first detected by screening mammography should be evaluated carefully by compression or magnified views. These abnormalities include clustered microcalcifications, densities (especially if spiculated), and new or enlarging architectural distortion. For some nonpalpable lesions, ultrasound may be helpful either to identify cysts or to guide biopsy. If there is no palpable lesion and detailed mammographic studies are unequivocally benign, the patient should have routine

follow-up appropriate to the patient's age. It cannot be stressed too strongly that in the presence of a breast lump a negative mammogram does not rule out cancer.

If a nonpalpable mammographic lesion has a low index of suspicion, mammographic follow-up in 3–6 months is reasonable. Workup of indeterminate and suspicious lesions has been rendered more complex by the advent of stereotactic biopsies. Morrow and colleagues have suggested that these procedures are indicated for lesions that require biopsy but are likely to be benign—that is, for cases in which the procedure probably will eliminate additional surgery. When a lesion is more probably malignant, open biopsy should be performed with a needle localization technique. Others have proposed more widespread use of stereotactic core biopsies for nonpalpable lesions on economic grounds and because diagnosis leads to earlier treatment planning. However, stereotactic diagnosis of a malignant lesion does not eliminate the need for definitive surgical procedures, particularly if breast conservation is attempted. For example, after a breast biopsy with needle localization (i.e., local excision) of a stereotactically diagnosed malignancy, reexcision may still be necessary to achieve negative margins. To some extent, these issues are decided on the basis of referral pattern and the availability of the resources for stereotactic core biopsies. A reasonable approach is shown in Fig. 90-4.

Figure 90-4

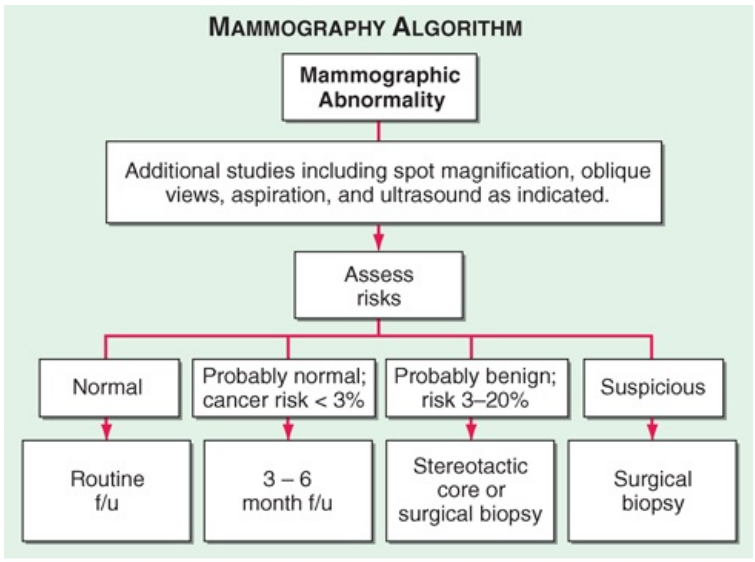

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approaches to abnormalities detected by mammogram.

Breast Masses in the Pregnant or Lactating Woman

During pregnancy, the breast grows under the influence of estrogen, progesterone, prolactin, and human placental lactogen. Lactation is suppressed by progesterone, which blocks the effects of prolactin. After delivery, lactation is promoted by the fall in progesterone levels, which leaves the effects of prolactin unopposed. The development of a dominant mass during pregnancy or lactation should never be attributed to hormonal changes. A dominant mass must be treated with the same concern in a pregnant woman as any other. Breast cancer develops in 1 in every 3000–4000 pregnancies. Stage for stage, breast cancer in pregnant patients is no different from premenopausal breast cancer in nonpregnant patients. However, pregnant women often have more advanced disease because the significance of a breast mass was not fully considered and/or because of endogenous hormone stimulation. Persistent lumps in the breast of pregnant or lactating women *cannot* be attributed to benign changes based on physical findings; such patients should be promptly referred for diagnostic evaluation.

Benign Breast Masses

Only about 1 in every 5–10 breast biopsies leads to a diagnosis of cancer, although the rate of positive biopsies varies in different countries and clinical settings. (These differences may be related to interpretation, medicolegal considerations, and availability of mammograms.) The vast majority of benign breast masses are due to "fibrocystic" disease, a descriptive term for small fluid-filled cysts and modest epithelial cell and fibrous tissue hyperplasia. However, fibrocystic disease is a histologic, not a clinical, diagnosis, and women who have had a biopsy with benign findings are at greater risk of developing breast cancer than those who have not had a biopsy. The subset of women with ductal or lobular cell proliferation (about 30% of patients), particularly the small fraction (3%) with atypical

hyperplasia, have a fourfold greater risk of developing breast cancer than those women who have not had a biopsy, and the increase in the risk is about ninefold for women in this category who also have an affected first-degree relative. Thus, careful follow-up of these patients is required. By contrast, patients with a benign biopsy without atypical hyperplasia are at little risk and may be followed routinely.

SCREENING

Breast cancer is virtually unique among the epithelial tumors in adults in that screening (in the form of annual mammography) improves survival. Meta-analysis examining outcomes from every randomized trial of mammography conclusively shows a 25–30% reduction in the chance of dying from breast cancer with annual screening after age 50 years; the data for women between ages 40 and 50 years are almost as positive; however, since the incidence is much lower in younger women, there are more false positives. While controversy continues to surround the assessment of screening mammography, the preponderance of data strongly supports the benefits of screening mammography. New analyses of older randomized studies have occasionally suggested that screening may not work. While the design defects in some older studies cannot be retrospectively corrected, most experts, including panels of the American Society of Clinical Oncology and the American Cancer Society (ACS), continue to believe that screening conveys substantial benefit. Furthermore, the profound drop in breast cancer mortality rate seen over the past decade is unlikely to be solely attributable to improvements in therapy. It seems prudent to recommend annual or biannual mammography for women past the age of 40 years. Although no randomized study of BSE has ever shown any improvement in survival, its major benefit is identification of tumors appropriate for conservative local therapy. Better mammographic technology, including digitized mammography, routine use of magnified views, and greater skill in mammographic interpretation, combined with newer diagnostic techniques (MRI, magnetic resonance spectroscopy, positron emission tomography, etc.) may make it possible to identify breast cancers even more reliably and earlier. Screening by any technique other than mammography is not indicated; however, the ACS recommends younger women who are *BRCA-1* or *BRCA-2* carriers or their untested first-degree relative; history of radiation therapy to the chest between ages 10 and 30 years; a lifetime risk of breast cancer of at least 20%; or a history of Li-Fraumeni, Cowden, or Bannayan-Riley-Ruvalcaba syndromes benefit from MRI screening, where the higher sensitivity may outweigh the loss of specificity.

STAGING

Correct staging of breast cancer patients is of extraordinary importance. Not only does it permit an accurate prognosis, but in many cases therapeutic decision-making is based largely on the TNM (primary tumor, regional nodes, metastasis) classification (Table 90–1). Comparison with historic series should be undertaken with caution, as the staging has changed several times in the past 20 years. The current staging is complex and results in significant changes in outcome by stage as compared with prior staging systems.

Table 90–1 Staging of Breast Cancer

Primary Tumor (T)	
T0	No evidence of primary tumor
TIS	Carcinoma in situ
T1	Tumor ≤2 cm
T1a	Tumor >0.1 cm but ≤0.5 cm
T1b	Tumor >0.5 but ≤1 cm
T1c	Tumor >1 cm but ≤2 cm
T2	Tumor >2 cm but ≤5 cm
T3	Tumor >5 cm
T4	Extension to chest wall, inflammation, satellite lesions, ulcerations
Regional Lymph Nodes (N)	
PN0(i–)	No regional lymph node metastasis histologically, negative IHC
PN0(i+)	No regional lymph node metastasis histologically, positive IHC, no IHC cluster greater than 0.2 mm
PN0(mol–)	No regional lymph node metastasis histologically, negative molecular findings (RT-PCR)
PN0(mol+)	No regional lymph node metastasis histologically, positive molecular findings (RT-PCR)
PN1	Metastasis in one to three axillary lymph nodes, or in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent
PN1mi	Micrometastasis (>0.2 mm, none >2 mm)

PN1a	Metastasis in one to three axillary lymph nodes
PN1b	Metastasis in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not *clinically apparent?*[a]
PN1c	Metastasis in one to three axillary lymph nodes and in internal mammary lymph nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent.[a] (If associated with greater than three positive axillary lymph nodes, the internal mammary nodes are classified as pN3b to reflect increased tumor burden.)
pN2	Metastasis in four to nine axillary lymph nodes, or in clinically apparent internal mammary lymph nodes in the *absence* of axillary lymph node metastasis
pN3	Metastasis in 10 or more axillary lymph nodes, or in infraclavicular lymph nodes, or in clinically apparent[a] ipsilateral internal mammary lymph nodes in the *presence* of 1 or more positive axillary lymph nodes; or in more than 3 axillary lymph nodes with clinically negative microscopic metastasis in internal mammary lymph nodes; or in ipsilateral subcarinal lymph nodes
Distant Metastasis (M)	
M0	No distant metastasis
M1	Distant metastasis (includes spread to ipsilateral supraclavicular nodes)

Stage Grouping			
Stage 0	TIS	N0	M0
Stage I	T1	N0	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1, N2	M0
Stage IIIB	T4	Any N	M0
	Any T	N3	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

[a]Clinically apparent is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination. ***Abbreviation:*** IHC, immunohistochemistry; RT-PCR, reverse transcriptase/polymerase chain reaction. ***Source:*** Used with permission of the American Joint Committee on Cancer (AJCC), Chicago, Illinois. The original source for this material is the *AJCC Cancer Staging Manual*, 7th ed. New York, Springer, 2010; *www.springeronline.com.*

Treatment: Breast Cancer

One of the most exciting aspects of breast cancer biology has been its recent subdivision into at least five subtypes based upon gene expression profiling.

1. 1.**Luminal A:** The luminal tumors express cytokeratins 8 and 18, have the highest levels of estrogen receptor expression, tend to be low-grade, are most likely to respond to endocrine therapy, and have a favorable prognosis. They tend to be less responsive to chemotherapy.
2. 2. **Luminal B:** Tumor cells are also of luminal epithelial origin, but with a gene expression pattern distinct from luminal A. Prognosis is somewhat worse that luminal A.
3. 3. **Normal breast—like**: These tumors have a gene expression profile reminiscent of nonmalignant "normal" breast epithelium. Prognosis is similar to the luminal B group.
4. 4. **HER2 amplified**: These tumors have amplification of the HER2 gene on chromosome 17q and frequently exhibit coamplification and overexpression of other genes adjacent to HER2. Historically the clinical prognosis of such tumors was poor. However, with the advent of trastuzumab, the clinical outcome of HER2-positive patients is markedly improving.
5. 5. **Basal:** These estrogen-receptor/progesterone receptor-negative and HER2-negative tumors (so-called triple negative) are characterized by markers of basal/myoepithelial cells. They tend to

be high-grade, and express cytokeratins 5/6 and 17 as well as vimentin, p63, CD10, α-smooth muscle actin, and epidermal growth factor receptor (EGFR). Patients with *BRCA* mutations also fall within this molecular subtype. They also have stem cell characteristics.

PRIMARY BREAST CANCER

Breast-conserving treatments, consisting of the removal of the primary tumor by some form of lumpectomy with or without irradiating the breast, result in a survival that is as good as (or slightly superior to) that after extensive surgical procedures, such as mastectomy or modified radical mastectomy, with or without further irradiation. Postlumpectomy breast irradiation greatly reduces the risk of recurrence in the breast. While breast conservation is associated with a possibility of recurrence in the breast, 10-year survival is at least as good as that after more extensive surgery. Postoperative radiation to regional nodes following mastectomy is also associated with an improvement in survival. Since radiation therapy can also reduce the rate of local or regional recurrence, it should be strongly considered following mastectomy for women with high-risk primary tumors (i.e., T2 in size, positive margins, positive nodes). At present, nearly one-third of women in the United States are managed by lumpectomy. Breast-conserving surgery is not suitable for all patients: it is not generally suitable for tumors >5 cm (or for smaller tumors if the breast is small), for tumors involving the nipple areola complex, for tumors with extensive intraductal disease involving multiple quadrants of the breast, for women with a history of collagen-vascular disease, and for women who either do not have the motivation for breast conservation or do not have convenient access to radiation therapy. However, these groups probably do not account for more than one-third of patients who are treated with mastectomy. Thus, a great many women still undergo mastectomy who could safely avoid this procedure and probably would if appropriately counseled.

An extensive intraductal component is a predictor of recurrence in the breast, and so are several clinical variables. Both axillary lymph node involvement and involvement of vascular or lymphatic channels by metastatic tumor in the breast are associated with a higher risk of relapse in the breast but are not contraindications to breast-conserving treatment. When these patients are excluded, and when lumpectomy with negative tumor margins is achieved, breast conservation is associated with a recurrence rate in the breast of substantially <10%. The survival of patients who have recurrence in the breast is somewhat worse than that of women who do not. Thus, recurrence in the breast is a negative prognostic variable for long-term survival. However, recurrence in the breast is not the *cause* of distant metastasis. If recurrence in the breast caused metastatic disease, then women treated with lumpectomy, who have a higher rate of recurrence in the breast, should have poorer survival than women treated with mastectomy, and they do not. Most patients should consult with a radiation oncologist before making a final decision concerning local therapy. However, a multimodality clinic in which the surgeon, radiation oncologist, medical oncologist, and other caregivers cooperate to evaluate the patient and develop a treatment is usually considered a major advantage by patients.

Adjuvant Therapy

The use of systemic therapy after local management of breast cancer substantially improves survival. More than half of the women who would otherwise die of metastatic breast cancer remain disease-free when treated with the appropriate systemic regimen. These data have grown more and more impressive with longer follow-up and more effective regimens.

Prognostic Variables

The most important prognostic variables are provided by *tumor staging* . The size of the tumor and the status of the axillary lymph nodes provide reasonably accurate information on the likelihood of tumor relapse. The relation of pathologic stage to 5-year survival is shown in Table 90–2. For most women, the need for adjuvant therapy can be readily defined on this basis alone. In the absence of lymph node involvement, involvement of microvessels (either capillaries or lymphatic channels) in tumors is nearly equivalent to lymph node involvement. The greatest controversy concerns women with intermediate prognoses. *There is rarely justification for adjuvant chemotherapy in most women with tumors <1 cm in size whose axillary lymph nodes are negative. HER2-positive tumors are a potential exception.* Detection of breast cancer cells either in the circulation or bone marrow is associated with an increased relapse rate. The most exciting development in this area is the use of gene expression arrays to analyze patterns of tumor gene expression. Several groups have independently defined gene sets that reliably predict disease-free and overall survival far more accurately than any single prognostic variable including the Oncotype DX® analysis of 21 genes. Also the use of such standardized risk assessment tools such as Adjuvant! Online (*www.adjuvantonline.com*) are very helpful. These tools are highly recommended in otherwise ambiguous circumstances. *Estrogen and progesterone receptor status* are of prognostic significance. Tumors that lack either or both of these receptors are more likely to recur than tumors that have them.

Table 90–2 5-Year Survival Rate for Breast Cancer by Stage

Stage	5-Year Survival, %

0	99
I	92
IIA	82
IIB	65
IIIA	47
IIIB	44
IV	14

Source: Modified from data of the National Cancer Institute: Surveillance, Epidemiology, and End Results (SEER).

Several *measures of tumor growth rate* correlate with early relapse. S-phase analysis using flow cytometry is the most accurate measure. Indirect S-phase assessments using antigens associated with the cell cycle, such as PCNA (Ki67), are also valuable. Tumors with a high proportion (more than the median) of cells in S-phase pose a greater risk of relapse; chemotherapy offers the greatest survival benefit for these tumors. Assessment of DNA content in the form of ploidy is of modest value, with nondiploid tumors having a somewhat worse prognosis.

Histologic classification of the tumor has also been used as a prognostic factor. Tumors with a poor nuclear grade have a higher risk of recurrence than tumors with a good nuclear grade. Semiquantitative measures such as the Elston score improve the reproducibility of this measurement.

Molecular changes in the tumor are also useful. Tumors that overexpress *erbB2* (HER2/neu) or have a mutated p53 gene have a worse prognosis. Particular interest has centered on *erbB2* overexpression as measured by histochemistry or by fluorescence in situ hybridization. Tumors that overexpress *erbB2* are more likely to respond to higher doses of doxorubicin-containing regimens and predict those tumors that will respond to HER2/neu antibodies (trastuzumab) (herceptin) and HER2/neu kinase inhibitors.

To grow, tumors must generate a neovasculature (Chap. 84). The presence of more microvessels in a tumor, particularly when localized in so-called hot spots, is associated with a worse prognosis. This may assume even greater significance in light of blood vessel—targeting therapies such as bevacizumab (avastin). While the benefits of bevacizumab in metastatic disease have been modest, close attention should be paid to the soon-to-be reported studies evaluating its role in adjuvant therapy.

Other variables that have also been used to evaluate prognosis include proteins associated with invasiveness, such as type IV collagenase, cathepsin D, plasminogen activator, plasminogen activator receptor, and the metastasis-suppressor gene *nm23*. None of these has been widely accepted as a prognostic variable for therapeutic decision-making. One problem in interpreting these prognostic variables is that most of them have not been examined in a study using a large cohort of patients.

Adjuvant Regimens

Adjuvant therapy is the use of systemic therapies in patients whose known disease has received local therapy but who are at risk of relapse. Selection of appropriate adjuvant chemotherapy or hormone therapy is highly controversial in some situations. Meta-analyses have helped to define broad limits for therapy but do not help in choosing optimal regimens or in choosing a regimen for certain subgroups of patients. A summary of recommendations is shown in Table 90–3. In general, premenopausal women for whom any form of adjuvant systemic therapy is indicated should receive multidrug chemotherapy. Antihormone therapy improves survival in premenopausal patients with positive estrogen receptors and should be added following completion of chemotherapy. Prophylactic castration may also be associated with a substantial survival benefit (primarily in estrogen receptor—positive patients) but is not widely used in this country.

Table 90–3 Suggested Approaches to Adjuvant Therapy

Age Group	Lymph Node Status[a]	Estrogen Receptor (ER) Status	Tumor	Recommendation
Premenopausal	Positive	Any	Any	Multidrug chemotherapy + tamoxifen if ER-positive + trastuzumab in HER2/neu –positive tumors
Premenopausal	Negative	Any	>2 cm, or 1–2 cm with other poor prognostic variables	Multidrug chemotherapy + tamoxifen if ER-positive + trastuzumab in HER2/neu –positive tumors
Postmenopausal	Positive	Negative	Any	Multidrug chemotherapy + trastuzumab

[illegible]	[illegible]	[illegible]	[illegible]	in HER2/neu–positive tumors
Postmenopausal	Positive	Positive	Any	Aromatase inhibitors and tamoxifen with or without chemotherapy + trastuzumab in HER2/neu–positive tumors
Postmenopausal	Negative	Positive	>2 cm, or 1–2 cm with other poor prognostic variables	Aromatase inhibitors and tamoxifen + trastuzumab in HER2/neu–positive tumors
Postmenopausal	Negative	Negative	>2 cm, or 1–2 cm with other poor prognostic variables	Consider multidrug chemotherapy + trastuzumab in HER2/neu–positive tumors

[a]As determined by pathologic examination.

Data on postmenopausal women are also controversial. The impact of adjuvant chemotherapy is quantitatively less clear-cut than in premenopausal patients, particularly in estrogen receptor—positive cases, although survival advantages have been shown. The first decision is whether chemotherapy or endocrine therapy should be used. While adjuvant endocrine therapy (aromatase inhibitors and tamoxifen) improves survival regardless of axillary lymph node status, the improvement in survival is modest for patients in whom multiple lymph nodes are involved. For this reason, it has been usual to give chemotherapy to postmenopausal patients who have no medical contraindications and who have more than one positive lymph node; hormone therapy is commonly given subsequently. For postmenopausal women for whom systemic therapy is warranted but who have a more favorable prognosis (based more commonly on analysis such as the Oncotype DX methodology), hormone therapy may be used alone. Large clinical trials have shown superiority for aromatase inhibitors over tamoxifen alone in the adjuvant setting. Unfortunately the optimal plan is unclear. Tamoxifen for 5 years followed by an aromatase inhibitor, the reverse strategy, or even switching to an aromatase inhibitor after 2–3 years of tamoxifen has been shown to be better than tamoxifen alone. No valid information currently permits selection among the three clinically approved aromatase inhibitors. Large clinical trials currently underway will help address these questions. Concomitant use of bisphosphonates is almost always warranted; however, it is not finally settled as to whether their prophylactic use increases survival in addition to just decreasing recurrences in bone.

Most comparisons of adjuvant chemotherapy regimens show little difference among them, although small advantages for doxorubicin-containing regimens and "dose sense" regimens are usually seen.

One approach—so-called neoadjuvant chemotherapy—involves the administration of adjuvant therapy before definitive surgery and radiation therapy. Because the objective response rates of patients with breast cancer to systemic therapy in this setting exceed 75%, many patients will be "down-staged" and may become candidates for breast-conserving therapy. However, overall survival has not been improved using this approach. Patients who achieve a pathologic complete remission after neoadjuvant chemotherapy not unexpectedly have a substantially improved survival. The neoadjuvant setting also provides a wonderful opportunity for the evaluation of new agents.

Other adjuvant treatments under investigation include the use of taxanes, such as paclitaxel and docetaxel, and therapy based on alternative kinetic and biologic models. In such approaches, high doses of single agents are used separately in relatively dose-intensive cycling regimens. Node-positive patients treated with doxorubicin-cyclophosphamide for four cycles followed by four cycles of a taxane have a substantial improvement in survival as compared with women receiving doxorubicin-cyclophosphamide alone, particularly in women with estrogen receptor—negative tumors. In addition, administration of the same drug combinations at the same dose but at more frequent intervals (q2 weeks with cytokine support as compared with the standard q3 weeks) is even more effective. Among the 25% of women whose tumors overexpress HER2/neu, addition of trastuzumab given concurrently with a taxane and then for a year after chemotherapy produces significant improvement in survival. Though longer follow-up will be important, this is now the standard care for most women with HER2/neu—positive breast cancers. Cardiotoxicity, immediate and long-term, remains a concern, and further efforts to exploit non—anthracycline-containing regimens are being pursued. Very high dose therapy with stem cell transplantation in the adjuvant setting has not proved superior to standard-dose therapy and should not be routinely used. A variety of exciting approaches are close to adoption and the literature needs to be followed attentively. These include the use of antiangiogenics such as bevacizumab. In addition, tyrosine kinase inhibitors such as lapatinib that target the HER2 kinase are very promising. Finally, as described in the next section, a novel class of agents targeting DNA repair —the so-called poly—ADP ribose polymerase [PARP] inhibitors—is likely to have a major impact on breast cancers either caused by *BRCA-1* or *-2* mutations or sharing similar defects in DNA repair in their etiology.

SYSTEMIC THERAPY OF METASTATIC DISEASE

About one-third of patients treated for apparently localized breast cancer develop metastatic disease.

Although a small number of these patients enjoy long remissions when treated with combinations of systemic and local therapy, most eventually succumb to metastatic disease. The median survival for all patients diagnosed with mestastatic breast cancer is less than 3 years. Soft tissue, bony, and visceral (lung and liver) metastases each account for approximately one-third of sites of initial relapses. However, by the time of death, most patients will have bony involvement. Recurrences can appear at any time after primary therapy. A very cruel fact about breast cancer recurrences is that at least half of all breast cancer recurrences occur >5 years after initial therapy.

Because the diagnosis of metastatic disease alters the outlook for the patient so drastically, it should rarely be made without a confirmatory biopsy. Every oncologist has seen patients with tuberculosis, gallstones, sarcoidosis, or other nonmalignant diseases misdiagnosed and treated as though they had metastatic breast cancer or even second malignancies such as multiple myeloma thought to be recurrent breast cancer. This is a catastrophic mistake and justifies biopsy for virtually every patient at the time of initial suspicion of metastatic disease.

The choice of therapy requires consideration of local therapy needs, the overall medical condition of the patient, and the hormone receptor status of the tumor, as well as clinical judgment. Because therapy of systemic disease is palliative, the potential toxicities of therapies should be balanced against the response rates. Several variables influence the response to systemic therapy. For example, the presence of estrogen and progesterone receptors is a strong indication for endocrine therapy. On the other hand, patients with short disease-free intervals, rapidly progressive visceral disease, lymphangitic pulmonary disease, or intracranial disease are unlikely to respond to endocrine therapy.

In many cases, systemic therapy can be withheld while the patient is managed with appropriate local therapy. Radiation therapy and occasionally surgery are effective at relieving the symptoms of metastatic disease, particularly when bony sites are involved. Many patients with bone-only or bone-dominant disease have a relatively indolent course. Under such circumstances, systemic chemotherapy has a modest effect, whereas radiation therapy may be effective for long periods. Other systemic treatments, such as strontium 89 and/or bisphosphonates, may provide a palliative benefit without inducing objective responses. Most patients with metastatic disease, and certainly all who have bone involvement, should receive concurrent bisphosphonates. Since the goal of therapy is to maintain well-being for as long as possible, emphasis should be placed on avoiding the most hazardous complications of metastatic disease, including pathologic fracture of the axial skeleton and spinal cord compression. New back pain in patients with cancer should be explored aggressively on an emergent basis; to wait for neurologic symptoms is a potentially catastrophic error. Metastatic involvement of endocrine organs can cause profound dysfunction, including adrenal insufficiency and hypopituitarism. Similarly, obstruction of the biliary tree or other impaired organ function may be better managed with a local therapy than with a systemic approach.

Endocrine Therapy

Normal breast tissue is estrogen dependent. Both primary and metastatic breast cancer may retain this phenotype. The best means of ascertaining whether a breast cancer is hormone dependent is through analysis of estrogen and progesterone receptor levels on the tumor. Tumors that are positive for the estrogen receptor and negative for the progesterone receptor have a response rate of ~30%. Tumors that have both receptors have a response rate approaching 70%. If neither receptor is present, the objective response rates are <5%. Receptor analyses provide information as to the correct ordering of endocrine therapies as opposed to chemotherapy. Because of their lack of toxicity and because some patients whose receptor analyses are reported as negative respond to endocrine therapy, an endocrine treatment should be attempted in virtually every patient with metastatic breast cancer. Potential endocrine therapies are summarized in Table 90–4. The choice of endocrine therapy is usually determined by toxicity profile and availability. In most patients, the initial endocrine therapy should be an aromatase inhibitor rather than tamoxifen. For the subset of postmenopausal women who are estrogen receptor—positive but also HER2/neu positive, response rates to aromatase inhibitors are substantially higher than to tamoxifen. Newer "pure" antiestrogens that are free of agonistic effects are also effective. Cases in which tumors shrink in response to tamoxifen withdrawal (as well as withdrawal of pharmacologic doses of estrogens) have been reported. Endogenous estrogen formation may be blocked by analogues of luteinizing hormone—releasing hormone in premenopausal women. Additive endocrine therapies, including treatment with progestogens, estrogens, and androgens, may also be tried in patients who respond to initial endocrine therapy; the mechanism of action of these latter therapies is unknown. Patients who respond to one endocrine therapy have at least a 50% chance of responding to a second endocrine therapy. It is not uncommon for patients to respond to two or three sequential endocrine therapies; however, combination endocrine therapies do not appear to be superior to individual agents, and combinations of chemotherapy with endocrine therapy are not useful. The median survival of patients with metastatic disease is approximately 2 years, and many patients, particularly older persons and those with hormone-dependent disease, may respond to endocrine therapy for 3–5 years or longer.

Table 90–4 Endocrine Therapies for Breast Cancer

Therapy	Comments
Castration Surgical LHRH agonists	For premenopausal women
Antiestrogens	
Tamoxifen	Useful in pre- and postmenopausal women
"Pure" antiestrogens	Responses in tamoxifen-resistant and aromatase inhibitor–resistant patients
Surgical adrenalectomy	Rarely employed second-line choice
Aromatase inhibitors	Low toxicity; now first choice for metastatic disease
High-dose progestogens	Common fourth-line choice after AIs, tamoxifen and fulvestrant
Hypophysectomy	Rarely used
Additive androgens or estrogens	Plausible fourth-line therapies; potentially toxic

Abbreviation: AI, aromatase inhibitor; LHRH, luteinizing hormone–releasing hormone.

Chemotherapy

Unlike many other epithelial malignancies, breast cancer responds to multiple chemotherapeutic agents, including anthracyclines, alkylating agents, taxanes, and antimetabolites. Multiple combinations of these agents have been found to improve response rates somewhat, but they have had little effect on duration of response or survival. The choice among multidrug combinations frequently depends on whether adjuvant chemotherapy was administered and, if so, what type. While patients treated with adjuvant regimens such as cyclophosphamide, methotrexate, and fluorouracil (CMF regimens) may subsequently respond to the same combination in the metastatic disease setting, most oncologists use drugs to which the patients have not been previously exposed. Once patients have progressed after combination drug therapy, it is most common to treat them with single agents. Given the significant toxicity of most drugs, the use of a single effective agent will minimize toxicity by sparing the patient exposure to drugs that would be of little value. No method to select the drugs most efficacious for a given patient has been demonstrated to be useful.

Most oncologists use either an anthracycline or paclitaxel following failure with the initial regimen. However, the choice has to be balanced with individual needs. One randomized study has suggested docetaxel may be superior to paclitaxel. A nanoparticle formulation of paclitaxel (Abraxane) is also effective.

The use of a humanized antibody to *erbB2* [trastuzumab (Herceptin)] combined with paclitaxel can improve response rate and survival for women whose metastatic tumors overexpress *erbB2*. The magnitude of the survival extension is modest in patients with metastatic disease. Similarly, the use of bevacizumab (avastin) has improved the response rate and response duration to paclitaxel. Objective responses in previously treated patients may also be seen with gemcitabine, vinca alkaloids, capecitabine, Navelbine, and oral etoposide and a new class of agents, epothilones.

High-Dose Chemotherapy Including Autologous Bone Marrow Transplantation

Autologous bone marrow transplantation combined with high doses of single agents can produce objective responses even in heavily pretreated patients. However, such responses are rarely durable and do not alter the clinical course for most patients with advanced metastatic disease.

STAGE III BREAST CANCER

Between 10 and 25% of patients present with so-called locally advanced, or stage III, breast cancer at diagnosis. Many of these cancers are technically operable, whereas others, particularly cancers with chest wall involvement, inflammatory breast cancers, or cancers with large matted axillary lymph nodes, cannot be managed with surgery initially. Although no randomized trials have proved the efficacy of neoadjuvant chemotherapy, this approach has gained widespread use. More than 90% of patients with locally advanced breast cancer show a partial or better response to multidrug chemotherapy regimens that include an anthracycline. Early administration of this treatment reduces the bulk of the disease and frequently makes the patient a suitable candidate for salvage surgery and/or radiation therapy. These patients should be managed in multimodality clinics to coordinate surgery, radiation therapy, and systemic chemotherapy. Such approaches produce long-term disease-free survival in about 30–50% of patients.

BREAST CANCER PREVENTION

Women who have one breast cancer are at risk of developing a contralateral breast cancer at a rate of approximately 0.5% per year. When adjuvant tamoxifen is administered to these patients, the rate of

development of contralateral breast cancers is reduced. In other tissues of the body, tamoxifen has estrogen-like effects that are beneficial: preservation of bone mineral density and long-term lowering of cholesterol. However, tamoxifen has estrogen-like effects on the uterus, leading to an increased risk of uterine cancer (0.75% incidence after 5 years on tamoxifen). Tamoxifen also increases the risk of cataract formation. The Breast Cancer Prevention Trial (BCPT) revealed a >49% reduction in breast cancer among women with a risk of at least 1.66% taking the drug for 5 years. Raloxifene has shown similar breast cancer prevention potency but may have different effects on bone and heart. The two agents have been compared in a prospective randomized prevention trial [the Study of Tamoxifen and Raloxifene (STAR) trial]. The agents are approximately equivalent in preventing breast cancer with fewer thromboembolic events and endometrial cancers with raloxifene; however, raloxifene did not reduce noninvasive cancers as effectively as tamoxifen, so no clear winner has emerged. A newer selective estrogen receptor modulator (SERM), lasofoxifene has recently been shown to reduce cardiovascular events in addition to breast cancer and fractures, and further studies of this agent should be watched with interest. It should be recalled that prevention of contralateral breast cancers in women diagnosed with *onc* cancer is a reasonable surrogate for breast cancer prevention as these are second primaries not recurrences. In this regard, the aromatase inhibitors are all considerably more effective than tamoxifen; however, they are not approved for primary breast cancer prevention. It remains puzzling that agents with the safety profile of raloxifene, which can reduce breast cancer risk by 50% with additional benefits in preventing osteoporotic fracture, are still so infrequently prescribed.

NONINVASIVE BREAST CANCER

Breast cancer develops as a series of molecular changes in the epithelial cells that lead to ever more malignant behavior. Increased use of mammography has led to more frequent diagnoses of noninvasive breast cancer. These lesions fall into two groups: ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (lobular neoplasia). The management of both entities is controversial.

Ductal Carcinoma In Situ

Proliferation of cytologically malignant breast epithelial cells within the ducts is termed *DCIS*. Atypical hyperplasia may be difficult to differentiate from DCIS. At least one-third of patients with untreated DCIS develop invasive breast cancer within 5 years. For many years, the standard treatment for this disease was mastectomy. However, treatment of this condition by lumpectomy and radiation therapy gives survival that is as good as the survival for invasive breast cancer treated by mastectomy. In one randomized trial, the combination of wide excision plus irradiation for DCIS caused a substantial reduction in the local recurrence rate as compared with wide excision alone with negative margins, though survival was identical in the two arms. No studies have compared either of these regimens to mastectomy. Addition of tamoxifen to any DCIS surgical/radiation therapy regimen further improves local control. Data for aromatase inhibitors in this setting are not available.

Several prognostic features may help to identify patients at high risk for local recurrence after either lumpectomy alone or lumpectomy with radiation therapy. These include extensive disease; age <40; and cytologic features such as necrosis, poor nuclear grade, and comedo subtype with overexpression of *erbB2*. Some data suggest that adequate excision with careful determination of pathologically clear margins is associated with a low recurrence rate. When surgery is combined with radiation therapy, recurrence (which is usually in the same quadrant) occurs with a frequency of ≤10%. Given the fact that half of these recurrences will be invasive, about 5% of the initial cohort will eventually develop invasive breast cancer. A reasonable expectation of mortality for these patients is about 1%, a figure that approximates the mortality rate for DCIS managed by mastectomy. Although this train of reasoning has not formally been proved valid, it is reasonable to recommend that patients who desire breast preservation, and in whom DCIS appears to be reasonably localized, be managed by adequate surgery with meticulous pathologic evaluation, followed by breast irradiation and tamoxifen. For patients with localized DCIS, axillary lymph node dissection is unnecessary. More controversial is the question of what management is optimal when there is any degree of invasion. Because of a significant likelihood (10–15%) of axillary lymph node involvement even when the primary lesion shows only microscopic invasion, it is prudent to do at least a level 1 and 2 axillary lymph node dissection for all patients with any degree of invasion, or sentinel node biopsy may be substituted. Further management is dictated by the presence of nodal spread.

Lobular Neoplasia

Proliferation of cytologically malignant cells within the lobules is termed *lobular neoplasia*. Nearly 30% of patients who have had adequate local excision of the lesion develop breast cancer (usually infiltrating ductal carcinoma) over the next 15–20 years. Ipsilateral and contralateral cancers are equally common. Therefore, lobular neoplasia may be a premalignant lesion that suggests an elevated risk of subsequent breast cancer, rather than a form of malignancy itself, and aggressive local management seems unreasonable. Most patients should be treated with an SERM for 5 years and followed with careful annual mammography and semiannual physical examinations. Additional molecular analysis of these lesions may make it possible to discriminate between patients who are at risk of further progression and require additional therapy and those in whom simple follow-up is adequate.

MALE BREAST CANCER

Breast cancer is about 1/150th as frequent in men as in women; 1720 men developed breast cancer in 2006. It usually presents as a unilateral lump in the breast and is frequently not diagnosed promptly. Given the small amount of soft tissue and the unexpected nature of the problem, locally advanced presentations are somewhat more common. When male breast cancer is matched to female breast cancer by age and stage, its overall prognosis is identical. Although gynecomastia may initially be unilateral or asymmetric, any unilateral mass in a man older than age 40 years should receive a careful workup including biopsy. On the other hand, bilateral symmetric breast development rarely represents breast cancer and is almost invariably due to endocrine disease or a drug effect. It should be kept in mind, nevertheless, that the risk of cancer is much greater in men with gynecomastia; in such men, gross asymmetry of the breasts should arouse suspicion of cancer. Male breast cancer is best managed by mastectomy and axillary lymph node dissection (modified radical mastectomy). Patients with locally advanced disease or positive nodes should also be treated with irradiation. Approximately 90% of male breast cancers contain estrogen receptors, and approximately 60% of cases with metastatic disease respond to endocrine therapy. No randomized studies have evaluated adjuvant therapy for male breast cancer. Two historic experiences suggest that the disease responds well to adjuvant systemic therapy, and, if not medically contraindicated, the same criteria for the use of adjuvant therapy in women should be applied to men.

The sites of relapse and spectrum of response to chemotherapeutic drugs are virtually identical for breast cancers in either sex.

FOLLOW-UP OF BREAST CANCER PATIENTS

Despite the availability of sophisticated and expensive imaging techniques and a wide range of serum tumor marker tests, survival is not influenced by early diagnosis of relapse. Surveillance guidelines are given in Table 90–5. Despite pressure from patients and their families, routine CAT scans etc. are not recommended.

Table 90–5 Breast Cancer Surveillance Guidelines

Test	Frequency
Recommended	
History; eliciting symptoms; physical examination	q3–6 months x 3 years; q6–12 months x 2 years; then annually
Breast self-examination	Monthly
Mammography	Annually
Pelvic examination	Annually
Patient education about symptoms of recurrence	Ongoing
Coordination of care	Ongoing
Not Recommended	
Complete blood count	
Serum chemistry studies Chest radiographs	
Bone scans	
Ultrasound examination of the liver	
Computed tomography of chest, abdomen, or pelvis	
Tumor marker CA 15-3, CA 27-29	
Tumor marker CEA	

Source: Recommended Breast Cancer Surveillance Guidelines, ASCO Education Book, Fall, 1997.

FURTHER READINGS

Chlebowski RT et al: Breast cancer after use of estrogen plus progestin in postme nopausal women. N Engl J Med 360:6, 2009

Clarke M et al: Adjuvant chemotherapy in oestrogen-receptor-poor breast cancer: Patient-level meta-analysis of randomised trials. Lancet 371:29, 2008[PMID: 18177773]

Foulkes WD et al: Triple-negative breast cancer. N Engl J Med 363:1938, 2010[PMID: 21067385]

Geyer CE et al: Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med

355:2733, 2006[PMID: 17192538]

Mansel RE et al: Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: The ALMANAC Trial. J Natl Cancer Inst98:599, 2006[PMID: 16670385]

Miller K et al: Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. N Engl J Med 357:26, 2007

Olivotto IA et al: Population-based validation of the prognostic model ADJUVANT! for early breast cancer. J Clin Oncol 23:2716, 2005[PMID: 15837986]

Paik S et al: Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor—positive breast cancer. J Clin Oncol 24:3726, 2006[PMID: 16720680]

Shipitsin M et al: Molecular definition of breast tumor heterogeneity. Cancer Cell 11:259, 2007[PMID: 17349583]

Sorlie T et al: Gene expression patterns of breast carcinomas distinguished tumor subclasses with clinical implications. Proc Natl Acad Sci USA 98:10869, 2001[PMID: 11553815]

Sotiriou C, Pusztai L: Gene-expression signatures in breast cancer. N Engl J Med360:8, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GASTROINTESTINAL TRACT CANCER: INTRODUCTION

The gastrointestinal tract is the second most common noncutaneous site for cancer and the second major cause of cancer-related mortality in the United States.

ESOPHAGEAL CANCER

Incidence and Etiology

Cancer of the esophagus is a relatively uncommon but extremely lethal malignancy. The diagnosis was made in 16,640 Americans in 2010 and led to 14,500 deaths. Worldwide, the incidence, of esophageal cancer varies strikingly. It occurs frequently within a geographic region extending from the southern shore of the Caspian Sea on the west to northern China on the east and encompassing parts of Iran, Central Asia, Afghanistan, Siberia, and Mongolia. Familial increased risk has been seen in regions with high incidence, though gene associations are not yet defined. High-incidence "pockets" of the disease are also present in such disparate locations as Finland, Iceland, Curaçao, southeastern Africa, and northwestern France. In North America and western Europe, the disease is more common in blacks than whites and in males than females; it appears most often after age 50 and seems to be associated with a lower socioeconomic status.

A variety of causative factors have been implicated in the development of the disease **(Table 91–1)**. In the United States, esophageal cancer cases are either squamous cell carcinomas or adenocarcinomas. The etiology of squamous cell esophageal cancer is related to excess alcohol consumption and/or cigarette smoking. The relative risk increases with the amount of tobacco smoked or alcohol consumed, with these factors acting synergistically. The consumption of whiskey is linked to a higher incidence than the consumption of wine or beer. Squamous cell esophageal carcinoma has also been associated with the ingestion of nitrites, smoked opiates, and fungal toxins in pickled vegetables, as well as mucosal damage caused by such physical insults as long-term exposure to extremely hot tea, the ingestion of lye, radiation-induced strictures, and chronic achalasia. The presence of an esophageal web in association with glossitis and iron deficiency (i.e., Plummer-Vinson or Paterson-Kelly syndrome) and congenital hyperkeratosis and pitting of the palms and soles (i.e., tylosis palmaris et plantaris) have each been linked with squamous cell esophageal cancer, as have dietary deficiencies of molybdenum, zinc, selenium, and vitamin A. Bisphosphonates may increase the risk in patients with Barrett's esophagus. Patients with head and neck cancer are at increased risk of squamous cell cancer of the esophagus.

Table 91–1 Some Etiologic Factors Believed to Be Associated with Esophageal Cancer

Excess alcohol consumption
Cigarette smoking
Other ingested carcinogens
Nitrates (converted to nitrites)
Smoked opiates
Fungal toxins in pickled vegetables
Mucosal damage from physical agents
Hot tea
Lye ingestion
Radiation-induced strictures
Chronic achalasia
Host susceptibility
Esophageal web with glossitis and iron deficiency (i.e., Plummer-Vinson or Paterson-Kelly syndrome)
Congenital hyperkeratosis and pitting of the palms and soles (i.e., tylosis palmaris et plantaris)
Dietary deficiencies of selenium, molybdenum, zinc, and vitamin A
Celiac sprue
Chronic gastric reflux (i.e., Barrett's esophagus) for adenocarcinoma

For unclear reasons, the incidence of squamous cell esophageal cancer has decreased somewhat in both the black and white population in the United States over the past 30 years, while the rate of adenocarcinoma has risen dramatically, particularly in white males (M:F 6:1). Adenocarcinomas arise in the distal esophagus in the presence of chronic gastric reflux and gastric metaplasia of the epithelium (Barrett's esophagus), which is more common in obese persons. Adenocarcinomas arise within dysplastic columnar epithelium in the distal esophagus. Even before frank neoplasia is detectable, aneuploidy and p53 mutations are found in the dysplastic epithelium. These adenocarcinomas behave clinically like gastric adenocarcinoma and now account for >70% of esophageal cancers.

Clinical Features

About 10% of esophageal cancers occur in the upper third of the esophagus (cervical esophagus), 35% in the middle third, and 55% in the lower third. Squamous cell carcinomas and adenocarcinomas cannot be distinguished radiographically or endoscopically.

Progressive dysphagia and weight loss of short duration are the initial symptoms in the vast majority of patients. Dysphagia initially occurs with solid foods and gradually progresses to include semisolids and liquids. By the time these symptoms develop, the disease is usually incurable, since difficulty in swallowing does not occur until >60% of the esophageal circumference is infiltrated with cancer. Dysphagia may be associated with pain on swallowing (odynophagia), pain radiating to the chest and/or back, regurgitation or vomiting, and aspiration pneumonia. The disease most commonly spreads to adjacent and supraclavicular lymph nodes, liver, lungs, pleura, and bone. Tracheoesophageal fistulas may develop as the disease advances, leading to severe suffering. As with other squamous cell carcinomas, hypercalcemia may occur in the absence of osseous metastases, probably from parathormone-related peptide secreted by tumor cells (Chap. 100).

Diagnosis

Attempts at endoscopic and cytologic screening for carcinoma in patients with Barrett's esophagus, while effective as a means of detecting high-grade dysplasia, have not yet been shown to improve the prognosis in individuals found to have a carcinoma. Routine contrast radiographs effectively identify esophageal lesions large enough to cause symptoms. In contrast to benign esophageal leiomyomas, which result in esophageal narrowing with preservation of a normal mucosal pattern, esophageal carcinomas show ragged, ulcerating changes in the mucosa in association with deeper infiltration, producing a picture resembling achalasia. Smaller, potentially resectable tumors are often poorly visualized despite technically adequate esophagograms. Because of this, esophagoscopy should be performed in all patients suspected of having an esophageal abnormality, to visualize the tumor and to obtain histopathologic confirmation of the diagnosis. Because the population of persons at risk for squamous cell carcinoma of the esophagus (i.e., smokers and drinkers) also has a high rate of cancers of the lung and the head and neck region, endoscopic inspection of the larynx, trachea, and bronchi should also be done. A thorough examination of the fundus of the stomach (by retroflexing the endoscope) is imperative as well. Endoscopic biopsies of esophageal tumors fail to recover malignant tissue in one-third of cases because the biopsy forceps cannot penetrate deeply enough through normal mucosa pushed in front of the carcinoma. Taking multiple biopsies increases the yield. Cytologic examination of tumor brushings complements standard biopsies and should be performed routinely. The extent of tumor spread to the mediastinum and para-aortic lymph nodes should be assessed by CT scans of the chest and abdomen and by endoscopic ultrasound. Positron emission tomography scanning provides a useful assessment of resectability, offering accurate information regarding spread to mediastinal lymph nodes. Most patients have advanced disease at presentation.

Treatment: Esophageal Cancer

The prognosis for patients with esophageal carcinoma is poor. Fewer than 5% of patients survive 5 years after the diagnosis; thus, management focuses on symptom control. Surgical resection of all gross tumor (i.e., total resection) is feasible in only 45% of cases, with residual tumor cells frequently present at the resection margins. Such esophagectomies have been associated with a postoperative mortality rate of approximately 5% due to anastomotic fistulas, subphrenic abscesses, and respiratory complications. About 20% of patients who survive a total resection live 5 years. The efficacy of primary radiation therapy (5500–6000 cGy) for squamous cell carcinomas is similar to that of radical surgery, sparing patients perioperative morbidity but often resulting in less satisfactory palliation of obstructive symptoms. The evaluation of chemotherapeutic agents in patients with esophageal carcinoma has been hampered by ambiguity in the definition of "response" and the debilitated physical condition of many treated individuals. Nonetheless, significant reductions in the size of measurable tumor masses have been reported in 15–25% of patients given single-agent treatment and in 30–60% of patients treated with drug combinations that include cisplatin. Combination chemotherapy and radiation therapy as the initial therapeutic approach, either alone or followed by an attempt at operative resection, seems to be beneficial. When administered along with radiation therapy, chemotherapy produces a better survival outcome than radiation therapy alone. The use of preoperative chemotherapy and radiation therapy followed by esophageal resection appears to prolong survival as compared with

controls in small, randomized trials, and some reports suggest that no additional benefit accrues when surgery is added if significant shrinkage of tumor has been achieved by the chemoradiation combination.

For the incurable, surgically unresectable patient with esophageal cancer, dysphagia, malnutrition, and the management of tracheoesophageal fistulas are major issues. Approaches to palliation include repeated endoscopic dilatation, the surgical placement of a gastrostomy or jejunostomy for hydration and feeding, and endoscopic placement of an expansive metal stent to bypass the tumor. Endoscopic fulguration of the obstructing tumor with lasers is the most promising of these techniques.

TUMORS OF THE STOMACH

Gastric Adenocarcinoma

INCIDENCE AND EPIDEMIOLOGY

For unclear reasons, the incidence and mortality rates for gastric cancer have decreased markedly worldwide during the past 75 years. The mortality rate from gastric cancer in the United States has dropped in men from 28 to 5.8 per 100,000 persons, while in women the rate has decreased from 27 to 2.8 per 100,000. Nonetheless, 21,000 new cases of stomach cancer were diagnosed in the United States, and 10,570 Americans died of the disease in 2010. Gastric cancer incidence has decreased worldwide but remains high in Japan, China, Chile, and Ireland.

The risk of gastric cancer is greater among lower socioeconomic classes. Migrants from high- to low-incidence nations maintain their susceptibility to gastric cancer, while the risk for their offspring approximates that of the new homeland. These findings suggest that an environmental exposure, probably beginning early in life, is related to the development of gastric cancer, with dietary carcinogens considered the most likely factor(s).

PATHOLOGY

About 85% of stomach cancers are adenocarcinomas, with 15% due to lymphomas and gastrointestinal stromal tumors (GIST) and leiomyosarcomas. Gastric adenocarcinomas may be subdivided into two categories: a *diffuse type*, in which cell cohesion is absent, so that individual cells infiltrate and thicken the stomach wall without forming a discrete mass; and an *intestinal type*, characterized by cohesive neoplastic cells that form glandlike tubular structures. The diffuse carcinomas occur more often in younger patients, develop throughout the stomach (including the cardia), result in a loss of distensibility of the gastric wall (so-called linitis plastica, or "leather bottle" appearance), and carry a poorer prognosis. Diffuse cancers have defective intercellular adhesion, mainly as a consequence of loss of expression of E-cadherin. Intestinal-type lesions are frequently ulcerative, more commonly appear in the antrum and lesser curvature of the stomach, and are often preceded by a prolonged precancerous process, often initiated by *Helicobacter pylori* infection. While the incidence of diffuse carcinomas is similar in most populations, the intestinal type tends to predominate in the high-risk geographic regions and is less likely to be found in areas where the frequency of gastric cancer is declining. Thus, different etiologic factor(s) are likely involved in these two subtypes. In the United States, ~30% of gastric cancers originate in the distal stomach, ~20% arise in the midportion of the stomach, and ~37% originate in the proximal third of the stomach. The remaining 13% involve the entire stomach.

ETIOLOGY

The long-term ingestion of high concentrations of nitrates in dried, smoked, and salted foods appears to be associated with a higher risk. The nitrates are thought to be converted to carcinogenic nitrites by bacteria **(Table 91–2)**. Such bacteria may be introduced exogenously through the ingestion of partially decayed foods, which are consumed in abundance worldwide by the lower socioeconomic classes. Bacteria such as *H. pylori* may also contribute to this effect by causing chronic gastritis, loss of gastric acidity, and bacterial growth in the stomach. The effect of *H. pylori* eradication on the subsequent risk for gastric cancer in high-incidence areas is under investigation. Loss of acidity may occur when acid-producing cells of the gastric antrum have been removed surgically to control benign peptic ulcer disease or when achlorhydria, atrophic gastritis, and even pernicious anemia develop in the elderly. Serial endoscopic examinations of the stomach in patients with atrophic gastritis have documented replacement of the usual gastric mucosa by intestinal-type cells. This process of intestinal metaplasia may lead to cellular atypia and eventual neoplasia. Since the declining incidence of gastric cancer in the United States primarily reflects a decline in distal, ulcerating, intestinal-type lesions, it is conceivable that better food preservation and the availability of refrigeration to all socioeconomic classes have decreased the dietary ingestion of exogenous bacteria. *H. pylori* has not been associated with the diffuse, more proximal form of gastric carcinoma.

Table 91–2 Nitrate-Converting Bacteria as a Factor in the Causation of Gastric Carcinoma[a]

Exogenous sources of nitrate-converting bacteria:

Bacterially contaminated food (common in lower socioeconomic classes, who have a higher incidence of the disease; diminished by improved food preservation and refrigeration)
?*Helicobacter pylori* infection
Endogenous factors favoring growth of nitrate-converting bacteria in the stomach:
Decreased gastric acidity
Prior gastric surgery (antrectomy) (15- to 20-year latency period)
Atrophic gastritis and/or pernicious anemia
? Prolonged exposure to histamine H_2-receptor antagonists

[a]Hypothesis: Dietary nitrates are converted to carcinogenic nitrites by bacteria.

Several additional etiologic factors have been associated with gastric carcinoma. Gastric ulcers and adenomatous polyps have occasionally been linked, but data on a cause-and-effect relationship are unconvincing. The inadequate clinical distinction between benign gastric ulcers and small ulcerating carcinomas may, in part, account for this presumed association. The presence of extreme hypertrophy of gastric rugal folds (i.e., Ménétrier's disease), giving the impression of polypoid lesions, has been associated with a striking frequency of malignant transformation; such hypertrophy, however, does not represent the presence of true adenomatous polyps. Individuals with blood group A have a higher incidence of gastric cancer than persons with blood group O; this observation may be related to differences in the mucous secretion, leading to altered mucosal protection from carcinogens. A germ-line mutation in the E-cadherin gene (*CDH1*), inherited in an autosomal dominant pattern and coding for a cell adhesion protein, has been linked to a high incidence of occult diffuse-type gastric cancers in young asymptomatic carriers. Duodenal ulcers are not associated with gastric cancer.

In keeping with the stepwise model of carcinogenesis, K-ras mutations appear to be early events in intestinal-type gastric cancer. C-met expression is amplified in about 1 in 5 cases and correlates with advanced stage. About half of intestinal-type tumors have mutations in tumor suppressor genes such as *TP53*, *TP73*, *APC* (adenomatous polyposis coli), *TFF* (trefoid factor family), *DCC* (deleted in colon cancer), and *FHIT* (fragile histidine triad). Cyclin E overexpression is associated with progression from dysplasia. Epigenetic changes (especially increased methylation) has been correlated with higher risk of invasive disease. Beta-catenin has been found in the nucleus of tumor cells at the leading edge of invasion.

CLINICAL FEATURES

Gastric cancers, when superficial and surgically curable, usually produce no symptoms. As the tumor becomes more extensive, patients may complain of an insidious upper abdominal discomfort varying in intensity from a vague, postprandial fullness to a severe, steady pain. Anorexia, often with slight nausea, is very common but is not the usual presenting complaint. Weight loss may eventually be observed, and nausea and vomiting are particularly prominent with tumors of the pylorus; dysphagia and early satiety may be the major symptoms caused by diffuse lesions originating in the cardia. There are no early physical signs. A palpable abdominal mass indicates long-standing growth and predicts regional extension.

Gastric carcinomas spread by direct extension through the gastric wall to the perigastric tissues, occasionally adhering to adjacent organs such as the pancreas, colon, or liver. The disease also spreads via lymphatics or by seeding of peritoneal surfaces. Metastases to intraabdominal and supraclavicular lymph nodes occur frequently, as do metastatic nodules to the ovary (Krukenberg's tumor), periumbilical region ("Sister Mary Joseph node"), or peritoneal cul-de-sac (Blumer's shelf palpable on rectal or vaginal examination); malignantascites may also develop. The liver is the most common site for hematogenous spread of tumor.

The presence of iron-deficiency anemia in men and of occult blood in the stool in both sexes mandates a search for an occult gastrointestinal tract lesion. A careful assessment is of particular importance in patients with atrophic gastritis or pernicious anemia. Unusual clinical features associated with gastric adenocarcinomas include migratory thrombophlebitis, microangiopathic hemolytic anemia, diffuse seborrheic keratoses (so-called Leser-Trélat sign), and acanthosis nigricans.

DIAGNOSIS

A double-contrast radiographic examination is the simplest diagnostic procedure for the evaluation of a patient with epigastric complaints. The use of double-contrast techniques helps to detect small lesions by improving mucosal detail. The stomach should be distended at some time during every radiographic examination, since decreased distensibility may be the only indication of a diffuse infiltrative carcinoma. Although gastric ulcers can be detected fairly early, distinguishing benign from malignant lesions radiographically is difficult. The anatomic location of an ulcer is not in itself an indication of the presence or absence of a cancer.

Gastric ulcers that appear benign by radiography present special problems. Some physicians believe

that gastroscopy is not mandatory if the radiographic features are typically benign, if complete healing can be visualized by x-ray within 6 weeks, and if a follow-up contrast radiograph obtained several months later shows a normal appearance. However, we recommend gastroscopic biopsy and brush cytology for all patients with a gastric ulcer in order to exclude a malignancy. Malignant gastric ulcers must be recognized before they penetrate into surrounding tissues, because the rate of cure of early lesions limited to the mucosa or submucosa is >80%. Since gastric carcinomas are difficult to distinguish clinically or radiographically from gastric lymphomas, endoscopic biopsies should be made as deeply as possible, due to the submucosal location of lymphoid tumors.

The staging system for gastric carcinoma is shown in **Table 91–3**.

Table 91–3 Staging System for Gastric Carcinoma

			Data from ACS	
Stage	**TNM**	**Features**	**No. of Cases, %**	**5-Year Survival, %**
0	$T_{is}N0M0$	Node negative; limited to mucosa	1	90
IA	T1N0M0	Node negative; invasion of lamina propria or submucosa	7	59
IB	T2N0M0 T1N1M0	Node negative; invasion of muscularis propria	10	44
II	T1N2M0 T2N1M0	Node positive; invasion beyond mucosa but within wall	17	29
		or		
	T3N0M0	Node negative; extension through wall		
IIIA	T2N2M0 T3N1-2M0	Node positive; invasion of muscularis propria or through wall	21	15
IIIB	T4N0-1M0	Node negative; adherence to surrounding tissue	14	9
IIIC	T4N2-3M0 T3N3M0	>3 nodes positive; invasion of serosa or adjacent structures 7 or more positive nodes; penetrates wall without invading serosa or adjacent structures		
IV	T4N2M0	Node positive; adherence to surrounding tissue	30	3
		or		
	T1-4N0-2 M1	Distant metastases		

Abbreviation: ACS, American Cancer Society; TNM, tumor, node, metastasis.

TREATMENT: GASTRIC ADENOCARCINOMA

Complete surgical removal of the tumor with resection of adjacent lymph nodes offers the only chance for cure. However, this is possible in less than a third of patients. A subtotal gastrectomy is the treatment of choice for patients with distal carcinomas, while total or near-total gastrectomies are required for more proximal tumors. The inclusion of extended lymph node dissection in these procedures appears to confer an added risk for complications without enhancing survival. The prognosis following complete surgical resection depends on the degree of tumor penetration into the stomach wall and is adversely influenced by regional lymph node involvement, vascular invasion, and abnormal DNA content (i.e., aneuploidy), characteristics found in the vast majority of American patients. As a result, the probability of survival after 5 years for the 25–30% of patients able to undergo complete resection is ~20% for distal tumors and <10% for proximal tumors, with recurrences continuing for at least 8 years after surgery. In the absence of ascites or extensive hepatic or peritoneal metastases, even patients whose disease is believed to be incurable by surgery should be offered resection of the primary lesion. Reduction of tumor bulk is the best form of palliation and may enhance the probability of benefit from subsequent therapy.

Gastric adenocarcinoma is a relatively radioresistant tumor, and adequate control of the primary tumor requires doses of external beam irradiation that exceed the tolerance of surrounding structures, such as bowel mucosa and spinal cord. As a result, the major role of radiation therapy in patients has been

palliation of pain. Radiation therapy alone after a complete resection does not prolong survival. In the setting of surgically unresectable disease limited to the epigastrium, patients treated with 3500–4000 cGy did not live longer than similar patients not receiving radiotherapy; however, survival was prolonged slightly when 5-fluorouracil (5-FU) plus leucovorin was given in combination with radiation therapy (3-year survival 50% vs 41% for radiation therapy alone). In this clinical setting, the 5-FU may be functioning as a radiosensitizer.

The administration of combinations of cytotoxic drugs to patients with advanced gastric carcinoma has been associated with partial responses in 30–50% of cases; responders appear to benefit from treatment. Such drug combinations have generally included cisplatin combined with epirubicin or docetaxel and infusional 5-FU, or with irinotecan. Despite this encouraging response rate, complete remissions are uncommon, the partial responses are transient, and the overall influence of multidrug therapy on survival has been unclear. The use of adjuvant chemotherapy alone following the complete resection of a gastric cancer has only minimally improved survival. However, combination chemotherapy administered before and after surgery (*perioperative treatment*) as well as postoperative chemotherapy combined with radiation therapy reduces the recurrence rate and prolongs survival.

Primary Gastric Lymphoma

Primary lymphoma of the stomach is relatively uncommon, accounting for <15% of gastric malignancies and ~2% of all lymphomas. The stomach is, however, the most frequent extranodal site for lymphoma, and gastric lymphoma has increased in frequency during the past 30 years. The disease is difficult to distinguish clinically from gastric adenocarcinoma; both tumors are most often detected during the sixth decade of life; present with epigastric pain, early satiety, and generalized fatigue; and are usually characterized by ulcerations with a ragged, thickened mucosal pattern demonstrated by contrast radiographs. The diagnosis of lymphoma of the stomach may occasionally be made through cytologic brushings of the gastric mucosa but usually requires a biopsy at gastroscopy or laparotomy. Failure of gastroscopic biopsies to detect lymphoma in a given case should not be interpreted as being conclusive, since superficial biopsies may miss the deeper lymphoid infiltrate. The macroscopic pathology of gastric lymphoma may also mimic adenocarcinoma, consisting of either a bulky ulcerated lesion localized in the corpus or antrum or a diffuse process spreading throughout the entire gastric submucosa and even extending into the duodenum. Microscopically, the vast majority of gastric lymphoid tumors are non-Hodgkin's lymphomas of B cell origin; Hodgkin's disease involving the stomach is extremely uncommon. Histologically, these tumors may range from well-differentiated, superficial processes [mucosa-associated lymphoid tissue (MALT)] to high-grade, large-cell lymphomas. Like gastric adenocarcinoma, infection with *H. pylori* increases the risk for gastric lymphoma in general and MALT lymphomas in particular. Gastric lymphomas spread initially to regional lymph nodes (often to Waldeyer's ring) and may then disseminate. Gastric lymphomas are staged like other lymphomas (Chap. 110).

TREATMENT: PRIMARY GASTRIC LYMPHOMA

Primary gastric lymphoma is a far more treatable disease than adenocarcinoma of the stomach, a fact that underscores the need for making the correct diagnosis. Antibiotic treatment to eradicate *H. pylori* infection has led to regression of about 75% of gastric MALT lymphomas and should be considered before surgery, radiation therapy, or chemotherapy are undertaken in patients having such tumors. A lack of response to such antimicrobial treatment has been linked to a specific chromosomal abnormality, i.e., t(11;18). Responding patients should undergo periodic endoscopic surveillance because it remains unclear whether the neoplastic clone is eliminated or merely suppressed, although the response to antimicrobial treatment is quite durable. Subtotal gastrectomy, usually followed by combination chemotherapy, has led to 5-year survival rates of 40–60% in patients with localized high-grade lymphomas. The need for a major surgical procedure has been questioned, particularly in patients with preoperative radiographic evidence of nodal involvement, for whom chemotherapy [CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone)] plus rituximab is effective therapy. A role for radiation therapy is not defined because most recurrences develop at distant sites.

Gastric (Nonlymphoid) Sarcoma

Leiomyosarcomas and GISTs make up 1–3% of gastric neoplasms. They most frequently involve the anterior and posterior walls of the gastric fundus and often ulcerate and bleed. Even those lesions that appear benign on histologic examination may behave in a malignant fashion. These tumors rarely invade adjacent viscera and characteristically do not metastasize to lymph nodes, but they may spread to the liver and lungs. The treatment of choice is surgical resection. Combination chemotherapy should be reserved for patients with metastatic disease. All such tumors should be analyzed for a mutation in the c-*kit* receptor. GISTs are unresponsive to conventional chemotherapy; yet ~50% of patients experience objective response and prolonged survival when treated with imatinib mesylate (Gleevec) (400–800 mg PO daily), a selective inhibitor of the c-*kit* tyrosine kinase. Many patients with GIST whose tumors have become refractory to imatinib subsequently benefit from sunitinib (Sutent), another inhibitor of the c-*kit* tyrosine kinase.

COLORECTAL CANCER

Incidence

Cancer of the large bowel is second only to lung cancer as a cause of cancer death in the United States: 142,570 new cases occurred in 2010, and 51,370 deaths were due to colorectal cancer. The incidence rate has decreased significantly during the past 20 years, likely due to enhanced and more compliantly followed screening practices. Similarly, mortality rates in the United States have decreased by approximately 25%, resulting largely from improved treatment and earlier detection.

Polyps and Molecular Pathogenesis

Most colorectal cancers, regardless of etiology, arise from adenomatous polyps. A polyp is a grossly visible protrusion from the mucosal surface and may be classified pathologically as a nonneoplastic hamartoma (*juvenile polyp*), a hyperplastic mucosal proliferation (*hyperplastic polyp*), or an adenomatous polyp. Only adenomas are clearly premalignant, and only a minority of such lesions becomes cancer. Adenomatous polyps may be found in the colons of ~30% of middle-aged and ~50% of elderly people; however, <1% of polyps ever become malignant. Most polyps produce no symptoms and remain clinically undetected. Occult blood in the stool is found in <5% of patients with polyps.

A number of molecular changes are noted in adenomatous polyps, dysplastic lesions, and polyps containing microscopic foci of tumor cells (carcinoma in situ), which are thought to reflect a multistep process in the evolution of normal colonic mucosa to life-threatening invasive carcinoma. These developmental steps toward carcinogenesis include, but are not restricted to, point mutations in the K-*ras* protooncogene; hypomethylation of DNA, leading to gene activation; loss of DNA (*allelic loss*) at the site of a tumor-suppressor gene [the adenomatous polyposis coli (*APC*) gene] on the long arm of chromosome 5 (5q21); allelic loss at the site of a tumor-suppressor gene located on chromosome 18q [the deleted in colorectal cancer (*DCC*) gene]; and allelic loss at chromosome 17p, associated with mutations in the p53 tumor-suppressor gene (see Fig. 83-2). Thus, the altered proliferative pattern of the colonic mucosa, which results in progression to a polyp and then to carcinoma, may involve the mutational activation of an oncogene followed by and coupled with the loss of genes that normally suppress tumorigenesis. It remains uncertain whether the genetic aberrations always occur in a defined order. Based on this model, however, cancer is believed to develop only in those polyps in which most (if not all) of these mutational events take place.

Clinically, the probability of an adenomatous polyp becoming a cancer depends on the gross appearance of the lesion, its histologic features, and its size. Adenomatous polyps may be pedunculated (stalked) or sessile (flat-based). Cancers develop more frequently in sessile polyps. Histologically, adenomatous polyps may be tubular, villous (i.e., papillary), or tubulovillous. Villous adenomas, most of which are sessile, become malignant more than three times as often as tubular adenomas. The likelihood that any polypoid lesion in the large bowel contains invasive cancer is related to the size of the polyp, being negligible (<2%) in lesions <1.5 cm, intermediate (2–10%) in lesions 1.5–2.5 cm, and substantial (10%) in lesions >2.5 cm in size.

Following the detection of an adenomatous polyp, the entire large bowel should be visualized endoscopically or radiographically, since synchronous lesions are noted in about one-third of cases. Colonoscopy should then be repeated periodically, even in the absence of a previously documented malignancy, since such patients have a 30–50% probability of developing another adenoma and are at a higher-than-average risk for developing a colorectal carcinoma. Adenomatous polyps are thought to require >5 years of growth before becoming clinically significant; colonoscopy need not be carried out more frequently than every 3 years.

Etiology and Risk Factors

Risk factors for the development of colorectal cancer are listed in **Table 91–4**.

Table 91–4 Risk Factors for the Development of Colorectal Cancer

Diet: Animal fat
Hereditary syndromes (autosomal dominant inheritance)
Polyposis coli
Nonpolyposis syndrome (Lynch syndrome)
Inflammatory bowel disease
Streptococcus bovis bacteremia
Ureterosigmoidostomy
? Tobacco use

DIET

The etiology for most cases of large-bowel cancer appears to be related to environmental factors. The

disease occurs more often in upper socioeconomic populations who live in urban areas. Mortality from colorectal cancer is directly correlated with per capita consumption of calories, meat protein, and dietary fat and oil as well as elevations in the serum cholesterol concentration and mortality from coronary artery disease. Geographic variations in incidence are unrelated to genetic differences, since migrant groups tend to assume the large-bowel cancer incidence rates of their adopted countries. Furthermore, population groups such as Mormons and Seventh Day Adventists, whose lifestyle and dietary habits differ somewhat from those of their neighbors, have significantly lower-than-expected incidence and mortality rates for colorectal cancer. Colorectal cancer has increased in Japan since that nation has adopted a more "Western" diet. At least three hypotheses have been proposed to explain the relationship to diet, none of which is fully satisfactory.

Animal Fats

One hypothesis is that the ingestion of animal fats found in red meats and processed meat leads to an increased proportion of anaerobes in the gut microflora, resulting in the conversion of normal bile acids into carcinogens. This provocative hypothesis is supported by several reports of increased amounts of fecal anaerobes in the stools of patients with colorectal cancer. Diets high in animal (but not vegetable) fats are also associated with high serum cholesterol, which is also associated with enhanced risk for the development of colorectal adenomas and carcinomas.

Insulin Resistance

The large number of calories in Western diets coupled with physical inactivity has been associated with a higher prevalence of obesity. Obese persons develop insulin resistance with increased circulating levels of insulin, leading to higher circulating concentrations of insulin-like growth factor type I (IGF-I). This growth factor appears to stimulate proliferation of the intestinal mucosa.

Fiber

Contrary to prior beliefs, the results of randomized trials and case-controlled studies have failed to show any value for dietary fiber or diets high in fruits and vegetables in preventing the recurrence of colorectal adenomas or the development of colorectal cancer. The weight of epidemiologic evidence, however, implicates diet as being the major etiologic factor for colorectal cancer, particularly diets high in animal fat and in calories.

Hereditary Factors and Syndromes

Up to 25% of patients with colorectal cancer have a family history of the disease, suggesting a hereditary predisposition. Inherited large-bowel cancers can be divided into two main groups: the well-studied but uncommon polyposis syndromes and the more common nonpolyposis syndromes **(Table 91–5)**.

Table 91–5 Hereditable (Autosomal Dominant) Gastrointestinal Polyposis Syndromes

Syndrome	Distribution of Polyps	Histologic Type	Malignant Potential	Associated Lesions
Familial adenomatous polyposis	Large intestine	Adenoma	Common	None
Gardner's syndrome	Large and small intestines	Adenoma	Common	Osteomas, fibromas, lipomas, epidermoid cysts, ampullary cancers, congenital hypertrophy of retinal pigment epithelium
Turcot's syndrome	Large intestine	Adenoma	Common	Brain tumors
Nonpolyposis syndrome (Lynch syndrome)	Large intestine (often proximal)	Adenoma	Common	Endometrial and ovarian tumors
Peutz-Jeghers syndrome	Small and large intestines, stomach	Hamartoma	Rare	Mucocutaneous pigmentation; tumors of the ovary, breast, pancreas, endometrium
Juvenile polyposis	Large and small intestines, stomach	Hamartoma, rarely progressing to adenoma	Rare	Various congenital abnormalities

POLYPOSIS COLI

Polyposis coli (familial polyposis of the colon) is a rare condition characterized by the appearance of thousands of adenomatous polyps throughout the large bowel. It is transmitted as an autosomal dominant trait; the occasional patient with no family history probably developed the condition due to a

spontaneous mutation. Polyposis coli is associated with a deletion in the long arm of chromosome 5 [including the *APC* (adenomatous polyposis coli) gene] in both neoplastic (somatic mutation) and normal (germ-line mutation) cells. The loss of this genetic material (i.e., allelic loss) results in the absence of tumor-suppressor genes whose protein products would normally inhibit neoplastic growth. The presence of soft tissue and bony tumors, congenital hypertrophy of the retinal pigment epithelium, mesenteric desmoid tumors, and ampullary cancers in addition to the colonic polyps characterizes a subset of polyposis coli known as *Gardner's syndrome*. The appearance of malignant tumors of the central nervous system accompanying polyposis coli defines *Turcot's syndrome*. The colonic polyps in all these conditions are rarely present before puberty but are generally evident in affected individuals by age 25. If the polyposis is not treated surgically, colorectal cancer will develop in almost all patients before age 40. Polyposis coli results from a defect in the colonic mucosa, leading to an abnormal proliferative pattern and impaired DNA repair mechanisms. Once the multiple polyps are detected, patients should undergo a total colectomy. Medical therapy with nonsteroidal anti-inflammatory drugs (NSAIDs) such as sulindac and cyclooxygenase-2 inhibitors such as celecoxib can decrease the number and size of polyps in patients with polyposis coli; however, this effect on polyps is only temporary, and NSAIDs are not proven to reduce the risk of cancer. Colectomy remains the primary therapy/prevention. The offspring of patients with polyposis coli, who often are prepubertal when the diagnosis is made in the parent, have a 50% risk for developing this premalignant disorder and should be carefully screened by annual flexible sigmoidoscopy until age 35. Proctosigmoidoscopy is a sufficient screening procedure because polyps tend to be evenly distributed from cecum to anus, making more-invasive and expensive techniques such as colonoscopy or barium enema unnecessary. Testing for occult blood in the stool is an inadequate screening maneuver. An alternative method for identifying carriers is testing DNA from peripheral blood mononuclear cells for the presence of a mutated *APC* gene. The detection of such a germ-line mutation can lead to a definitive diagnosis before the development of polyps.

HEREDITARY NONPOLYPOSIS COLON CANCER

Hereditary nonpolyposis colon cancer (HNPCC), also known as *Lynch syndrome*, is another autosomal dominant trait. It is characterized by the presence of three or more relatives with histologically documented colorectal cancer, one of whom is a first-degree relative of the other two; one or more cases of colorectal cancer diagnosed before age 50 in the family; and colorectal cancer involving at least two generations. In contrast to polyposis coli, HNPCC is associated with an unusually high frequency of cancer arising in the proximal large bowel. The median age for the appearance of an adenocarcinoma is <50 years, 10–15 years younger than the median age for the general population. Despite having a poorly differentiated histologic appearance, the proximal colon tumors in HNPCC have a better prognosis than sporadic tumors from patients of similar age. Families with HNPCC often include individuals with multiple primary cancers; the association of colorectal cancer with either ovarian or endometrial carcinomas is especially strong in women. It has been recommended that members of such families undergo biennial colonoscopy beginning at age 25 years, with intermittent pelvic ultrasonography and endometrial biopsy for afflicted women; such a screening strategy has not yet been validated. HNPCC is associated with germ-line mutations of several genes, particularly *hMSH2* on chromosome 2 and *hMLH1* on chromosome 3. These mutations lead to errors in DNA replication and are thought to result in DNA instability because of defective repair of DNA mismatches resulting in abnormal cell growth and tumor development. Testing tumor cells through molecular analysis of DNA or immunohistochemical staining of paraffin-fixed tissue for "microsatellite instability" (sequence changes reflecting defective mismatch repair) in patients younger than age 50 with colorectal cancer and a positive family history for colorectal or endometrial cancer may identify probands with HNPCC.

Inflammatory Bowel Disease

(Chap. 295) Large-bowel cancer is increased in incidence in patients with long-standing inflammatory bowel disease (IBD). Cancers develop more commonly in patients with ulcerative colitis than in those with granulomatous colitis, but this impression may result in part from the occasional difficulty of differentiating these two conditions. The risk of colorectal cancer in a patient with IBD is relatively small during the initial 10 years of the disease, but then it appears to increase at a rate of ~0.5–1% per year. Cancer may develop in 8–30% of patients after 25 years. The risk is higher in younger patients with pancolitis.

Cancer surveillance in patients with IBD is unsatisfactory. Symptoms such as bloody diarrhea, abdominal cramping, and obstruction, which may signal the appearance of a tumor, are similar to the complaints caused by a flare-up of the underlying disease. In patients with a history of IBD lasting ≥15 years who continue to experience exacerbations, the surgical removal of the colon can significantly reduce the risk for cancer and also eliminate the target organ for the underlying chronic gastrointestinal disorder. The value of such surveillance techniques as colonoscopy with mucosal biopsies and brushings for less-symptomatic individuals with chronic IBD is uncertain. The lack of uniformity regarding the pathologic criteria that characterize dysplasia and the absence of data that such surveillance reduces the development of lethal cancers have made this costly practice an area of controversy.

Other High-Risk Conditions

STREPTOCOCCUS BOVIS BACTEREMIA

For unknown reasons, individuals who develop endocarditis or septicemia from this fecal bacterium have a high incidence of occult colorectal tumors and, possibly, upper gastrointestinal cancers as well. Endoscopic or radiographic screening appears advisable.

TOBACCO USE

Cigarette smoking is linked to the development of colorectal adenomas, particularly after >35 years of tobacco use. No biologic explanation for this association has yet been proposed.

Primary Prevention

Several orally administered compounds have been assessed as possible inhibitors of colon cancer. The most effective class of chemopreventive agents is aspirin and other NSAIDs, which are thought to suppress cell proliferation by inhibiting prostaglandin synthesis. Regular aspirin use reduces the risk of colon adenomas and carcinomas as well as death from large-bowel cancer; such use also appears to diminish the likelihood for developing additional premalignant adenomas following treatment for a prior colon carcinoma. This effect of aspirin on colon carcinogenesis increases with the duration and dosage of drug use. Oral folic acid supplements and oral calcium supplements reduce the risk of adenomatous polyps and colorectal cancers in case-controlled studies. The value of vitamin D as a form of chemo-prevention is under study. Antioxidant vitamins such as ascorbic acid, tocopherols, and ß-carotene are ineffective at reducing the incidence of subsequent adenomas in patients who have undergone the removal of a colon adenoma. Estrogen-replacement therapy has been associated with a reduction in the risk of colorectal cancer in women, conceivably by an effect on bile acid synthesis and composition or by decreasing synthesis of IGF-I. The otherwise unexplained reduction in colorectal cancer mortality rate in women may be a result of the widespread use of estrogen replacement in postmenopausal individuals.

Screening

The rationale for colorectal cancer screening programs is that earlier detection of localized, superficial cancers in asymptomatic individuals will increase the surgical cure rate. Such screening programs are important for individuals having a family history of the disease in first-degree relatives. The relative risk for developing colorectal cancer increases to 1.75 in such individuals and may be even higher if the relative was afflicted before age 60. The prior use of proctosigmoidoscopy as a screening tool was based on the observation that 60% of early lesions are located in the rectosigmoid. For unexplained reasons, however, the proportion of large-bowel cancers arising in the rectum has been decreasing during the past several decades, with a corresponding increase in the proportion of cancers in the more proximal descending colon. As such, the potential for rigid proctosigmoidoscopy to detect a sufficient number of occult neoplasms to make the procedure cost-effective has been questioned. Flexible, fiberoptic sigmoidoscopes permit trained operators to visualize the colon for up to 60 cm, which enhances the capability for cancer detection. However, this technique still leaves the proximal half of the large bowel unscreened.

Most programs directed at the early detection of colorectal cancers have focused on digital rectal examinations and fecal occult blood testing. The digital examination should be part of any routine physical evaluation in adults older than age 40 years, serving as a screening test for prostate cancer in men, a component of the pelvic examination in women, and an inexpensive maneuver for the detection of masses in the rectum. The development of the Hemoccult test has greatly facilitated the detection of occult fecal blood. Unfortunately, even when performed optimally, the Hemoccult test has major limitations as a screening technique. About 50% of patients with documented colorectal cancers have a negative fecal Hemoccult test, consistent with the intermittent bleeding pattern of these tumors. When random cohorts of asymptomatic persons have been tested, 2–4% have Hemoccult-positive stools. Colorectal cancers have been found in <10% of these "test-positive" cases, with benign polyps being detected in an additional 20–30%. Thus, a colorectal neoplasm will not be found in most asymptomatic individuals with occult blood in their stool. Nonetheless, persons found to have Hemoccult-positive stool routinely undergo further medical evaluation, including sigmoidoscopy, barium enema, and/or colonoscopy—procedures that are not only uncomfortable and expensive but also associated with a small risk for significant complications. The added cost of these studies would appear justifiable if the small number of patients found to have occult neoplasms because of Hemoccult screening could be shown to have an improved prognosis and prolonged survival. Prospectively controlled trials showed a statistically significant reduction in mortality rate from colorectal cancer for individuals undergoing annual screening. However, this benefit only emerged after >13 years of follow-up and was extremely expensive to achieve, since all positive tests (most of which were false-positive) were followed by colonoscopy. Moreover, these colonoscopic examinations quite likely provided the opportunity for cancer prevention through the removal of potentially premalignant adenomatous polyps since the eventual development of cancer was reduced by 20% in the cohort undergoing annual screening.

Screening techniques for large-bowel cancer in asymptomatic persons remain unsatisfactory. Compliance with any screening strategy within the general population is poor. At present, the American Cancer Society suggests fecal Hemoccult screening annually and flexible sigmoidoscopy every 5 years beginning at age 50 for asymptomatic individuals having no colorectal cancer risk factors. The American Cancer Society has also endorsed a "total colon examination" (i.e., colonoscopy or double-contrast barium enema) every 10 years as an alternative to Hemoccult testing with periodic flexible sigmoidoscopy. Colonoscopy has been shown to be superior to double-contrast barium enema and also to have a higher sensitivity for detecting villous or dysplastic adenomas or cancers than the strategy employing occult fecal blood testing and flexible sigmoidoscopy. Whether colonoscopy performed every 10 years beginning after age 50 will prove to be cost-effective and whether it may be supplanted as a screening maneuver by sophisticated radiographic techniques ("virtual colonoscopy") remains unclear. More effective techniques for screening are needed, perhaps taking advantage of the molecular changes that have been described in these tumors. Analysis of fecal DNA for multiple mutations associated with colorectal cancer is being tested.

Clinical Features

PRESENTING SYMPTOMS

Symptoms vary with the anatomic location of the tumor. Since stool is relatively liquid as it passes through the ileocecal valve into the right colon, cancers arising in the cecum and ascending colon may become quite large without resulting in any obstructive symptoms or noticeable alterations in bowel habits. Lesions of the right colon commonly ulcerate, leading to chronic, insidious blood loss without a change in the appearance of the stool. Consequently, patients with tumors of the ascending colon often present with symptoms such as fatigue, palpitations, and even angina pectoris and are found to have a hypochromic, microcytic anemia indicative of iron deficiency. Since the cancer may bleed intermittently, a random fecal occult blood test may be negative. As a result, the unexplained presence of iron-deficiency anemia in any adult (with the possible exception of a premenopausal, multiparous woman) mandates a thorough endoscopic and/or radiographic visualization of the entire large bowel **(Fig. 91-1)**.

Figure 91-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Double-contrast air-barium enema revealing a sessile tumor of the cecum in a patient with iron-deficiency anemia and guaiac-positive stool. The lesion at surgery was a stage II adenocarcinoma.

Since stool becomes more formed as it passes into the transverse and descending colon, tumors arising there tend to impede the passage of stool, resulting in the development of abdominal cramping, occasional obstruction, and even perforation. Radiographs of the abdomen often reveal characteristic annular, constricting lesions ("apple-core" or "napkin-ring") **(Fig. 91-2)**.

Figure 91-2

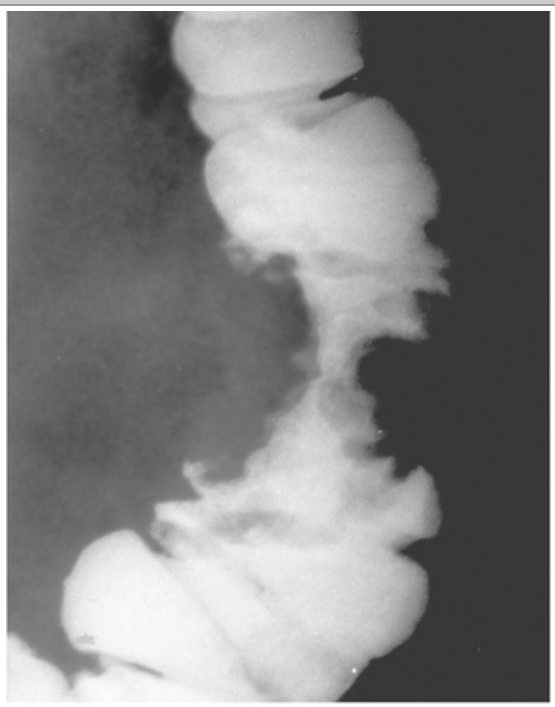

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Annular, constricting adenocarcinoma of the descending colon. This radiographic appearance is referred to as an "apple-core" lesion and is always highly suggestive of malignancy.

Cancers arising in the rectosigmoid are often associated with hematochezia, tenesmus, and narrowing of the caliber of stool; anemia is an infrequent finding. While these symptoms may lead patients and their physicians to suspect the presence of hemorrhoids, the development of rectal bleeding and/or altered bowel habits demands a prompt digital rectal examination and proctosigmoidoscopy.

STAGING, PROGNOSTIC FACTORS, AND PATTERNS OF SPREAD

The prognosis for individuals having colorectal cancer is related to the depth of tumor penetration into the bowel wall and the presence of both regional lymph node involvement and distant metastases. These variables are incorporated into the staging system introduced by Dukes and applied to a TNM classification method, in which T represents the depth of tumor penetration, N the presence of lymph node involvement, and M the presence or absence of distant metastases **(Fig. 91-3)**. Superficial lesions that do not involve regional lymph nodes and do not penetrate through the submucosa (T_1) or the muscularis (T_2) are designated as *stage I* ($T_{1–2}N_0M_0$) disease; tumors that penetrate through the muscularis but have not spread to lymph nodes are *stage II* disease ($T_3N_0M_0$); regional lymph node involvement defines *stage III* ($T_XN_1M_0$) disease; and metastatic spread to sites such as liver, lung, or bone indicates *stage IV* ($T_XN_XM_1$) disease. Unless gross evidence of metastatic disease is present, disease stage cannot be determined accurately before surgical resection and pathologic analysis of the operative specimens. It is not clear whether the detection of nodal metastases by special immunohistochemical molecular techniques has the same prognostic implications as disease detected by routine light microscopy.

Figure 91-3

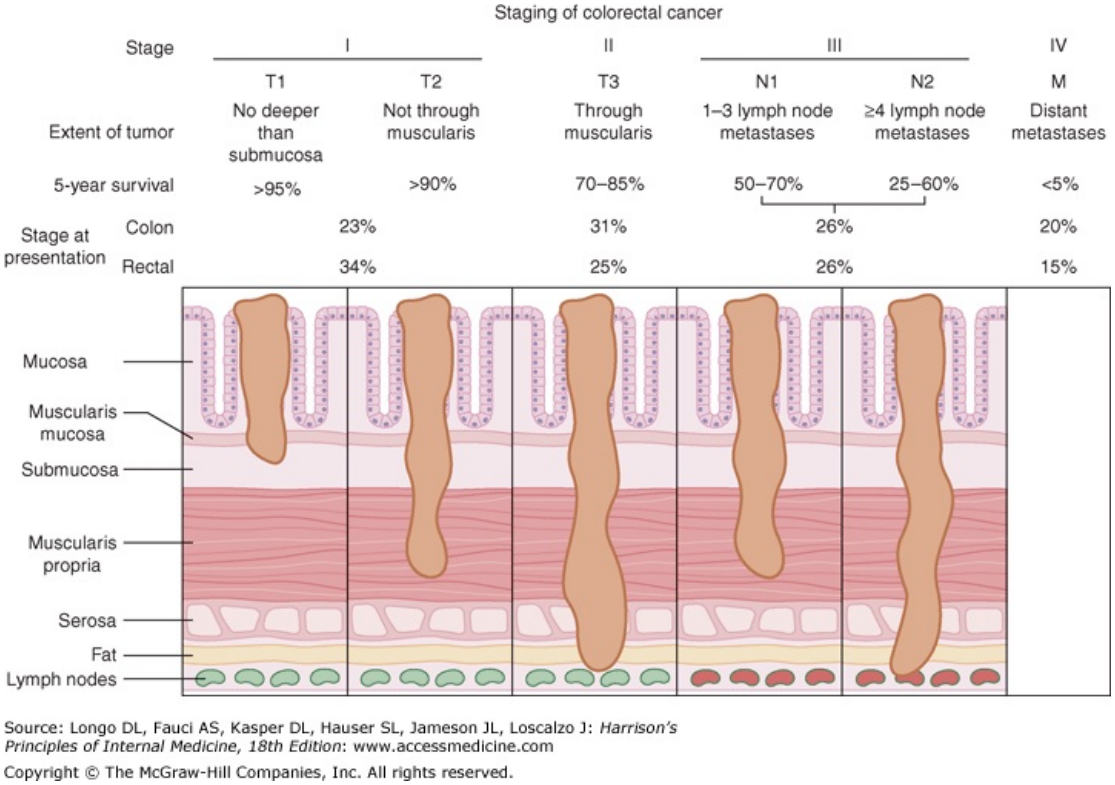

Staging and prognosis for patients with colorectal cancer.

Most recurrences after a surgical resection of a large-bowel cancer occur within the first 4 years, making 5-year survival a fairly reliable indicator of cure. The likelihood for 5-year survival in patients with colorectal cancer is stage-related (Fig. 91-3). That likelihood has improved during the past several decades when similar surgical stages have been compared. The most plausible explanation for this improvement is more thorough intraoperative and pathologic staging. In particular, more exacting attention to pathologic detail has revealed that the prognosis following the resection of a colorectal cancer is not related merely to the presence or absence of regional lymph node involvement. Prognosis may be more precisely gauged by the number of involved lymph nodes (one to three lymph nodes versus four or more lymph nodes) and the number of nodes examined. A minimum of 12 sampled lymph nodes is thought necessary to accurately define tumor stage, and the more nodes examined the better. Other predictors of a poor prognosis after a total surgical resection include tumor penetration through the bowel wall into pericolic fat, poorly differentiated histology, perforation and/or tumor adherence to adjacent organs (increasing the risk for an anatomically adjacent recurrence), and venous invasion by tumor **(Table 91–6)**. Regardless of the clinicopathologic stage, a preoperative elevation of the plasma carcinoembryonic antigen (CEA) level predicts eventual tumor recurrence. The presence of aneuploidy and specific chromosomal deletions, such as allelic loss in chromosome 18q (involving the *DCC* gene) in tumor cells, appears to predict a higher risk for metastatic spread, particularly in patients with stage II (T3N0M0) disease. Conversely, the detection of microsatellite instability in tumor tissue indicates a more favorable outcome. In contrast to most other cancers, the prognosis in colorectal cancer is not influenced by the size of the primary lesion when adjusted for nodal involvement and histologic differentiation.

Table 91–6 Predictors of Poor Outcome Following Total Surgical Resection of Colorectal Cancer

Tumor spread to regional lymph nodes
Number of regional lymph nodes involved
Tumor penetration through the bowel wall
Poorly differentiated histology
Perforation
Tumor adherence to adjacent organs Venous invasion
Preoperative elevation of CEA titer (<5 ng/mL)
Aneuploidy
Specific chromosomal deletion (e.g., allelic loss on chromosome 18q)

Abbreviation: CEA, carcinoembryonic antigen.

Cancers of the large bowel generally spread to regional lymph nodes or to the liver via the portal venous circulation. The liver represents the most frequent visceral site of metastasis; it is the initial site of distant spread in one-third of recurring colorectal cancers and is involved in more than two-thirds of such patients at the time of death. In general, colorectal cancer rarely spreads to the lungs, supraclavicular lymph nodes, bone, or brain without prior spread to the liver. A major exception to this rule occurs in patients having primary tumors in the distal rectum, from which tumor cells may spread through the paravertebral venous plexus, escaping the portal venous system and thereby reaching the lungs or supraclavicular lymph nodes without hepatic involvement. The median survival after the detection of distant metastases has ranged in the past from 6–9 months (hepatomegaly, abnormal liver chemistries) to 24–30 months (small liver nodule initially identified by elevated CEA level and subsequent CT scan), but effective systemic therapy is improving the prognosis.

Efforts to use gene expression profiles to identify patients at risk of recurrence or those particularly likely to benefit from adjuvant therapy have not yet yielded practice-changing results. Despite a burgeoning literature examining a host of prognostic factors, pathologic stage at diagnosis is the best predictor of long-term prognosis. Patients with lymphovascular invasion and high preoperative CEA levels are likely to have a more aggressive clinical course.

Treatment: Colorectal Cancer

Total resection of tumor is the optimal treatment when a malignant lesion is detected in the large bowel. An evaluation for the presence of metastatic disease, including a thorough physical examination, chest radiograph, biochemical assessment of liver function, and measurement of the plasma CEA level, should be performed before surgery. When possible, a colonoscopy of the entire large bowel should be performed to identify synchronous neoplasms and/or polyps. The detection of metastases should not preclude surgery in patients with tumor-related symptoms such as gastrointestinal bleeding or obstruction, but it often prompts the use of a less radical operative procedure. At the time of laparotomy, the entire peritoneal cavity should be examined, with thorough inspection of the liver, pelvis, and hemidiaphragm and careful palpation of the full length of the large bowel. Following recovery from a complete resection, patients should be observed carefully for 5 years by semiannual physical examinations and yearly blood chemistry measurements. If a complete colonoscopy was not performed preoperatively, it should be carried out within the first several postoperative months. Some authorities favor measuring plasma CEA levels at 3-month intervals because of the sensitivity of this test as a marker for otherwise undetectable tumor recurrence. Subsequent endoscopic or radiographic surveillance of the large bowel, probably at triennial intervals, is indicated, since patients who have been cured of one colorectal cancer have a 3–5% probability of developing an additional bowel cancer during their lifetime and a >15% risk for the development of adenomatous polyps. Anastomotic ("suture-line") recurrences are infrequent in colorectal cancer patients, provided the surgical resection margins are adequate and free of tumor. The value of periodic CT scans of the abdomen, assessing for an early, asymptomatic indication of tumor recurrence, is an area of uncertainty, with some experts recommending the test be performed annually for the first 3 postoperative years.

Radiation therapy to the pelvis is recommended for patients with rectal cancer because it reduces the 20–25% probability of regional recurrences following complete surgical resection of stage II or III tumors, especially if they have penetrated through the serosa. This alarmingly high rate of local disease recurrence is believed to be due to the fact that the contained anatomic space within the pelvis limits the extent of the resection and because the rich lymphatic network of the pelvic side wall immediately adjacent to the rectum facilitates the early spread of malignant cells into surgically inaccessible tissue. The use of sharp rather than blunt dissection of rectal cancers (*total mesorectal excision*) appears to reduce the likelihood of local disease recurrence to ~10%. Radiation therapy, either pre- or postoperatively, reduces the likelihood of pelvic recurrences but does not appear to prolong survival. Combining postoperative radiation therapy with 5-fluorouracil-based chemotherapy lowers local recurrence rates and improves overall survival. Preoperative radiotherapy is indicated for patients with large, potentially unresectable rectal cancers; such lesions may shrink enough to permit subsequent surgical removal. Radiation therapy is not effective in the primary treatment of colon cancer.

Systemic therapy for patients with colorectal cancer has become more effective. 5-FU remains the backbone of treatment for this disease. Partial responses are obtained in 15–20% of patients. The probability of tumor response appears to be somewhat greater for patients with liver metastases when chemotherapy is infused directly into the hepatic artery, but intraarterial treatment is costly and toxic and does not appear to appreciably prolong survival. The concomitant administration of folinic acid (leucovorin) improves the efficacy of 5-FU in patients with advanced colorectal cancer, presumably by enhancing the binding of 5-FU to its target enzyme, thymidylate synthase. A threefold improvement in the partial response rate is noted when folinic acid is combined with 5-FU; however, the effect on survival is marginal, and the optimal dose schedule remains to be defined. 5-FU is generally administered intravenously but may also be given orally in the form of capecitabine (Xeloda) with

seemingly similar efficacy.

Irinotecan (CPT-11), a topoisomerase 1 inhibitor, prolongs survival when compared to supportive care in patients whose disease has progressed on 5-FU. Furthermore, the addition of irinotecan to 5-FU and leucovorin (LV) improves response rates and survival of patients with metastatic disease. The *FOLFIRI regimen* is as follows: irinotecan, 180 mg/m^2 as a 90-min infusion on day 1; LV, 400 mg/m^2 as a 2-h infusion during irinotecan administration; immediately followed by 5-FU bolus, 400 mg/m^2, and 46-h continuous infusion of 2.4–3 g/m^2 every 2 weeks. Diarrhea is the major side effect from irinotecan. Oxaliplatin, a platinum analogue, also improves the response rate when added to 5-FU and LV as initial treatment of patients with metastatic disease. The *FOLFOX regimen* is the following: 2-h infusion of LV (400 mg/m^2 per day) followed by a 5-FU bolus (400 mg/m^2 per day) and 22-h infusion (1200 mg/m^2) every 2 weeks, together with oxaliplatin, 85 mg/m^2 as a 2-h infusion on day 1. Oxaliplatin frequently causes a dose-dependent sensory neuropathy that often resolves following the cessation of therapy. FOLFIRI and FOLFOX are equal in efficacy. In metastatic disease, these regimens may produce median survivals of 2 years.

Monoclonal antibodies are also effective in patients with advanced colorectal cancer. Cetuximab (Erbitux) and panitumumab (Vectibix) are directed against the epidermal growth factor receptor (EGFR), a transmembrane glycoprotein involved in signaling pathways affecting growth and proliferation of tumor cells. Both cetuximab and panitumumab, when given alone, have been shown to benefit a small proportion of previously treated patients, and cetuximab appears to have therapeutic synergy with such chemotherapeutic agents as irinotecan, even in patients previously resistant to this drug; this suggests that cetuximab can reverse cellular resistance to cytotoxic chemotherapy. The antibodies are not effective in the subset of colon tumors that contain mutated *K-ras*. The use of both cetuximab and panitumumab can lead to an acne-like rash, with the development and severity of the rash being correlated with the likelihood of antitumor efficacy. Inhibitors of the EGFR tyrosine kinase such as erlotinib (Tarceva) do not appear to be effective in colorectal cancer.

Bevacizumab (Avastin) is a monoclonal antibody directed against the vascular endothelial growth factor (VEGF) and is thought to act as an anti-angiogenesis agent. The addition of bevacizumab to irinotecan-containing combinations and to FOLFOX initially appeared to improve the outcome observed with chemotherapy alone, but subsequent studies have been less convincing. The use of bevacizumab can lead to hypertension, proteinuria, and an increased likelihood of thromboembolic events.

Patients with solitary hepatic metastases without clinical or radiographic evidence of additional tumor involvement should be considered for partial liver resection, because such procedures are associated with 5-year survival rates of 25–30% when performed on selected individuals by experienced surgeons.

The administration of 5-FU and LV for 6 months after resection of tumor in patients with stage III disease leads to a 40% decrease in recurrence rates and 30% improvement in survival. The likelihood of recurrence has been further reduced when oxaliplatin has been combined with 5-FU and LV (e.g., FOLFOX); unexpectedly, the addition of irinotecan to 5-FU and LV as well as the addition of either bevacizumab or cetuximab to FOLFOX did not enhance outcome. Patients with stage II tumors do not appear to benefit applicably from adjuvant therapy with the use of such treatment generally restricted to those patients having biologic characteristics (e.g., perforated tumors, Ty lesions, lymphovascular invasion) that place them at higher than usual risk for recurrence. In rectal cancer, the delivery of preoperative or postoperative combined modality therapy (5-FU plus radiation therapy) reduces the risk of recurrence and increases the chance of cure for patients with stages II and III tumors, with the preoperative approach being better tolerated. The 5-FU acts as a radiosensitizer when delivered together with radiation therapy. Life-extending adjuvant therapy is used in only about half of patients older than age 65 years. This age bias is completely inappropriate as the benefits and likely the tolerance of adjuvant therapy in patients aged 65+ years appear similar to those seen in younger individuals.

TUMORS OF THE SMALL INTESTINE

Small-bowel tumors comprise <3% of gastrointestinal neoplasms. Because of their rarity, a correct diagnosis is often delayed. Abdominal symptoms are usually vague and poorly defined, and conventional radiographic studies of the upper and lower intestinal tract often appear normal. Small-bowel tumors should be considered in the differential diagnosis in the following situations: (1) recurrent, unexplained episodes of crampy abdominal pain; (2) intermittent bouts of intestinal obstruction, especially in the absence of IBD or prior abdominal surgery; (3) intussusception in the adult; and (4) evidence of chronic intestinal bleeding in the presence of negative conventional contrast radiographs. A careful small-bowel barium study is the diagnostic procedure of choice; the diagnostic accuracy may be improved by infusing barium through a nasogastric tube placed into the duodenum (enteroclysis).

Benign Tumors

The histology of benign small-bowel tumors is difficult to predict on clinical and radiologic grounds alone. The symptomatology of benign tumors is not distinctive, with pain, obstruction, and hemorrhage being the most frequent symptoms. These tumors are usually discovered during the fifth and sixth decades of life, more often in the distal rather than the proximal small intestine. The most common benign tumors are adenomas, leiomyomas, lipomas, and angiomas.

ADENOMAS

These tumors include those of the islet cells and Brunner's glands as well as polypoid adenomas. *Islet cell adenomas* are occasionally located outside the pancreas; the associated syndromes are discussed in Chap. 350. *Brunner's gland adenomas* are not truly neoplastic but represent a hypertrophy or hyperplasia of submucosal duodenal glands. These appear as small nodules in the duodenal mucosa that secrete a highly viscous alkaline mucus. Most often, this is an incidental radiographic finding not associated with any specific clinical disorder.

POLYPOID ADENOMAS

About 25% of benign small-bowel tumors are polypoid adenomas (Table 91–5). They may present as single polypoid lesions or, less commonly, as papillary villous adenomas. As in the colon, the sessile or papillary form of the tumor is sometimes associated with a coexisting carcinoma. Occasionally, patients with Gardner's syndrome develop premalignant adenomas in the small bowel; such lesions are generally in the duodenum. Multiple polypoid tumors may occur throughout the small bowel (and occasionally the stomach and colorectum) in the Peutz-Jeghers syndrome. The polyps are usually hamartomas (juvenile polyps) having a low potential for malignant degeneration. Mucocutaneous melanin deposits as well as tumors of the ovary, breast, pancreas, and endometrium are also associated with this autosomal dominant condition.

LEIOMYOMAS

These neoplasms arise from smooth-muscle components of the intestine and are usually intramural, affecting the overlying mucosa. Ulceration of the mucosa may cause gastrointestinal hemorrhage of varying severity. Cramping or intermittent abdominal pain is frequently encountered.

LIPOMAS

These tumors occur with greatest frequency in the distal ileum and at the ileocecal valve. They have a characteristic radiolucent appearance and are usually intramural and asymptomatic, but on occasion cause bleeding.

ANGIOMAS

While not true neoplasms, these lesions are important because they frequently cause intestinal bleeding. They may take the form of telangiectasia or hemangiomas. Multiple intestinal telangiectasias occur in a nonhereditary form confined to the gastrointestinal tract or as part of the hereditary Osler-Rendu-Weber syndrome. Vascular tumors may also take the form of isolated hemangiomas, most commonly in the jejunum. Angiography, especially during bleeding, is the best procedure for evaluating these lesions.

Malignant Tumors

While rare, small-bowel malignancies occur in patients with long-standing regional enteritis and celiac sprue as well as in individuals with AIDS. Malignant tumors of the small bowel are frequently associated with fever, weight loss, anorexia, bleeding, and a palpable abdominal mass. After ampullary carcinomas (many of which arise from biliary or pancreatic ducts), the most frequently occurring small-bowel malignancies are adenocarcinomas, lymphomas, carcinoid tumors, and leiomyosarcomas.

ADENOCARCINOMAS

The most common primary cancers of the small bowel are adenocarcinomas, accounting for ~50% of malignant tumors. These cancers occur most often in the distal duodenum and proximal jejunum, where they tend to ulcerate and cause hemorrhage or obstruction. Radiologically, they may be confused with chronic duodenal ulcer disease or with Crohn's disease if the patient has long-standing regional enteritis. The diagnosis is best made by endoscopy and biopsy under direct vision. Surgical resection is the treatment of choice.

LYMPHOMAS

Lymphoma in the small bowel may be primary or secondary. A diagnosis of a primary intestinal lymphoma requires histologic confirmation in a clinical setting in which palpable adenopathy and hepatosplenomegaly are absent and no evidence of lymphoma is seen on chest radiograph, CT scan, or peripheral blood smear or on bone marrow aspiration and biopsy. Symptoms referable to the small bowel are present, usually accompanied by an anatomically discernible lesion. Secondary lymphoma of the small bowel consists of involvement of the intestine by a lymphoid malignancy extending from involved retroperitoneal or mesenteric lymph nodes (Chap. 110).

Primary intestinal lymphoma accounts for ~20% of malignancies of the small bowel. These neoplasms

are non-Hodgkin's lymphomas; they usually have a diffuse, large-cell histology and are of T cell origin. Intestinal lymphoma involves the ileum, jejunum, and duodenum, in decreasing frequency—a pattern that mirrors the relative amount of normal lymphoid cells in these anatomic areas. The risk of small-bowel lymphoma is increased in patients with a prior history of malabsorptive conditions (e.g., celiac sprue), regional enteritis, and depressed immune function due to congenital immunodeficiency syndromes, prior organ transplantation, autoimmune disorders, or AIDS.

The development of localized or nodular masses that narrow the lumen results in periumbilical pain (made worse by eating) as well as weight loss, vomiting, and occasional intestinal obstruction. The diagnosis of small-bowel lymphoma may be suspected from the appearance on contrast radiographs of patterns such as infiltration and thickening of mucosal folds, mucosal nodules, areas of irregular ulceration, or stasis of contrast material. The diagnosis can be confirmed by surgical exploration and resection of involved segments. Intestinal lymphoma can occasionally be diagnosed by peroral intestinal mucosal biopsy, but since the disease mainly involves the lamina propria, full-thickness surgical biopsies are usually required.

Resection of the tumor constitutes the initial treatment modality. While postoperative radiation therapy has been given to some patients following a total resection, most authorities favor short-term (three cycles) systemic treatment with combination chemotherapy. The frequent presence of widespread intraabdominal disease at the time of diagnosis and the occasional multicentricity of the tumor often make a total resection impossible. The probability of sustained remission or cure is ~75% in patients with localized disease but is ~25% in individuals with unresectable lymphoma. In patients whose tumors are not resected, chemotherapy may lead to bowel perforation.

A unique form of small-bowel lymphoma, diffusely involving the entire intestine, was first described in oriental Jews and Arabs and is referred to as *immunoproliferative small intestinal disease* (IPSID), *Mediterranean lymphoma*, or *α heavy chain disease*. This is a B cell tumor. The typical presentation includes chronic diarrhea and steatorrhea associated with vomiting and abdominal cramps; clubbing of the digits may be observed. A curious feature in many patients with IPSID is the presence in the blood and intestinal secretions of an abnormal IgA that contains a shortened α heavy chain and is devoid of light chains. It is suspected that the abnormal α chains are produced by plasma cells infiltrating the small bowel. The clinical course of patients with IPSID is generally one of exacerbations and remissions, with death frequently resulting from either progressive malnutrition and wasting or the development of an aggressive lymphoma. The use of oral antibiotics such as tetracycline appears to be beneficial in the early phases of the disorder, suggesting a possible infectious etiology. Combination chemotherapy has been administered during later stages of the disease, with variable results. Results are better when antibiotics and chemotherapy are combined.

CARCINOID TUMORS

Carcinoid tumors arise from argentaffin cells of the crypts of Lieberkühn and are found from the distal duodenum to the ascending colon, areas embryologically derived from the midgut. More than 50% of intestinal carcinoids are found in the distal ileum, with most congregating close to the ileocecal valve. Most intestinal carcinoids are asymptomatic and of low malignant potential, but invasion and metastases may occur, leading to the carcinoid syndrome (Chap. 350).

LEIOMYOSARCOMAS

Leiomyosarcomas often are >5 cm in diameter and may be palpable on abdominal examination. Bleeding, obstruction, and perforation are common. Such tumors should be analyzed for the expression of mutant c-*kit* receptor (defining GIST), and in the presence of metastatic disease, justifying treatment with imatinib mesylate (Gleevec) or, in imatinib-refractory patients, sunitinib (Sutent).

CANCERS OF THE ANUS

Cancers of the anus account for 1–2% of the malignant tumors of the large bowel. Most such lesions arise in the anal canal, the anatomic area extending from the anorectal ring to a zone approximately halfway between the pectinate (or dentate) line and the anal verge. Carcinomas arising proximal to the pectinate line (i.e., in the transitional zone between the glandular mucosa of the rectum and the squamous epithelium of the distal anus) are known as *basaloid*, *cuboidal*, or *cloacogenic* tumors; about one-third of anal cancers have this histologic pattern. Malignancies arising distal to the pectinate line have squamous histology, ulcerate more frequently, and constitute ~55% of anal cancers. The prognosis for patients with basaloid and squamous cell cancers of the anus is identical when corrected for tumor size and the presence or absence of nodal spread.

The development of anal cancer is associated with infection by human papillomavirus, the same organism etiologically linked to cervical cancer. The virus is sexually transmitted. The infection may lead to anal warts (condyloma acuminata), which may progress to anal intraepithelial neoplasia and on to squamous cell carcinoma. The risk for anal cancer is increased among homosexual males, presumably related to anal intercourse. Anal cancer risk is increased in both men and women with AIDS, possibly because their immunosuppressed state permits more severe papillomavirus infection.

Anal cancers occur most commonly in middle-aged persons and are more frequent in women than men. At diagnosis, patients may experience bleeding, pain, sensation of a perianal mass, and pruritus.

Radical surgery (abdominal-perineal resection with lymph node sampling and a permanent colostomy) was once the treatment of choice for this tumor type. The 5-year survival rate after such a procedure was 55–70% in the absence of spread to regional lymph nodes and <20% if nodal involvement was present. An alternative therapeutic approach combining external beam radiation therapy with concomitant chemotherapy has resulted in biopsy-proven disappearance of all tumor in >80% of patients whose initial lesion was <3 cm in size. Tumor recurrences develop in <10% of these patients, meaning that ~70% of patients with anal cancers can be cured with nonoperative treatment. Surgery should be reserved for the minority of individuals who are found to have residual tumor after being managed initially with radiation therapy combined with chemotherapy.

FURTHER READINGS

Abbas A et al: Management of anal cancer in 2010: Parts 1 and 2. Oncology 24:364 and 417, 2010
Cunningham D et al: Colorectal cancer. Lancet 375:1030, 2010

Enzinger PC, Mayer RJ: Esophageal cancer. N Engl J Med 349:2241, 2003[PMID: 14657432]

Hartgrink HH et al: Gastric cancer. Lancet 374:477, 2009[PMID: 19625077]

Lieberman DA: Clinical practice. Screening for colorectal cancer. N Engl J Med 2009 361:1179, 2009[PMID: 19759380]

Lynch HT, De la Chapelle A: Hereditary colorectal cancer. N Engl J Med 348:919, 2003[PMID: 12621137]

Raderer M, Paul De Boer J: Role of chemotherapy in gastric MALT lymphoma, diffuse large B-cell lymphoma and other lymphomas. Best Pract Clin Gastroenterol 24:19, 2010[PMID: 20206105]

Rostrum A et al: Nonsteroidal anti-inflammatory drugs and cyclooxygenase-2 inhibitors for primary prevention of colorectal cancer: A systematic review prepared for the US Preventive Services Task Force. Ann Intern Med 146:376, 2007

Rothwell PM et al: Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet 376:1741, 2010[PMID: 20970847]

Shaheen NJ et al: Radiofrequency ablation in Barrett's esophagus with dysplasia. N Engl J Med 360:2277, 2009[PMID: 19474425]

Sharma P: Clinical practice. Barrett's esophagus. N Engl J Med 361:2548, 2009[PMID: 20032324]

Spechler SJ: Barrett's esophagus. N Engl J Med 346:836, 2002[PMID: 11893796]

Uemura N et al: *Helicobacter pylori* infection and the development of gastric cancer. N Engl J Med 345:784, 2001[PMID: 11556297]

Wolpin BM et al: Adjuvant treatment of colorectal cancer. CA Cancer Clin J 57:168, 2007[PMID: 17507442]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEPATOCELLULAR CARCINOMA

Incidence

Hepatocellular carcinoma (HCC) is one of the most common malignancies worldwide. The annual global incidence is approximately 1 million cases, with a male to female ratio of approximately 4:1 (1:1 without cirrhosis to 9:1 in many high-incidence countries). The incidence rate equals the death rate. In the United States, approximately 22,000 new cases are diagnosed annually, with 18,000 deaths. The death rates in males in low-incidence countries such as the United States are 1.9 per 100,000 per year; in intermediate areas such as Austria and South Africa, they range from 5.1–20; and in high-incidence areas such as in the Orient (China and Korea) as high as 23.1–150 per 100,000 per year (Table 92–1). The incidence of HCC in the United States is approximately 3 per 100,000 persons, with significant gender, ethnic, and geographic variations. These numbers are rapidly increasing and may be an underestimate. Approximately 4 million chronic hepatitis C virus (HCV) carriers are in the United States alone. Approximately 10% of them or 400,000 are likely to develop cirrhosis. Approximately 5% or 20,000 of these may develop HCC annually. Add to this the two other common predisposing factors—hepatitis B virus (HBV) and chronic alcohol consumption—and 60,000 new HCC cases annually seem possible. Future advances in HCC survival will likely depend in part on immunization strategies for HBV (and HCV) and earlier diagnosis by screening of patients at risk of HCC development.

Table 92–1 Age-Adjusted Incidence Rates for Hepatocellular Carcinoma

	Persons per 100,000 per Year	
Country	**Male**	**Female**
Argentina	6.0	2.5
Brazil, Recife	9.2	8.3
Brazil, Sao Paulo	3.8	2.6
Mozambique	112.9	30.8
South Africa, Cape: Black	26.3	8.4
South Africa, Cape: White	1.2	0.6
Senegal	25.6	9.0
Nigeria	15.4	3.2
Gambia	33.1	12.6
Burma	25.5	8.8
Japan	7.2	2.2
Korea	13.8	3.2
China, Shanghai	34.4	11.6
India, Bombay	4.9	2.5
India, Madras	2.1	0.7
Great Britain	1.6	0.8
France	6.9	1.2
Italy, Varese	7.1	2.7
Norway	1.8	1.1
Spain, Navarra	7.9	4.7

CURRENT DIRECTIONS

With the U.S. HCV epidemic, HCC is increasing in most states, and obesity-associated liver disease (nonalcoholic steatohepatitis [NASH]) is increasingly recognized as a cause.

Epidemiology

There are two general types of epidemiologic studies of HCC—those of country-based incidence rates (Table 92–1) and those of migrants. Endemic hot spots occur in areas of China and sub-Saharan Africa, which are associated both with high endemic hepatitis B carrier rates as well as mycotoxin contamination of foodstuffs (aflatoxin B_1), stored grains, drinking water, and soil. Environmental factors are important, for example, Japanese in Japan have a higher incidence than those living in Hawaii, who in turn have a higher incidence than those living in California.

Etiologic Factors

CHEMICAL CARCINOGENS

Causative agents for HCC have been studied along two general lines. First are agents identified as carcinogenic in experimental animals (particularly rodents) that are thought to be present in the human environment (Table 92–2). Second, is the association of HCC with various other clinical conditions. Probably the best-studied and most potent ubiquitous natural chemical carcinogen is a product of the *Aspergillus* fungus, called aflatoxin B_1. This mold and aflatoxin product can be found in a variety of stored grains in hot, humid places, where peanuts and rice are stored in unrefrigerated conditions. Aflatoxin contamination of foodstuffs correlates well with incidence rates in Africa and to some extent in China. In endemic areas of China, even farm animals such as ducks have HCC. The most potent carcinogens appear to be natural products of plants, fungi, and bacteria, such as bush trees containing pyrrolizidine alkaloids as well as tannic acid and safrole. Pollutants such as pesticides and insecticides are known rodent carcinogens.

Table 92–2 Factors Associated with an Increased Risk of Developing HCC

Common	Unusual
Cirrhosis from any cause	Primary biliary cirrhosis
Hepatitis B or C chronic infection	Hemochromatosis
Ethanol chronic consumption	α_1 Antitrypsin deficiency
NASH/NAFL	Glycogen storage diseases
Aflatoxin B_1 or other mycotoxins	Citrullinemia
	Porphyria cutanea tarda
	Hereditary tyrosinemia
	Wilson's disease

Abbreviation: NAFL, nonalcoholic fatty liver; NASH, nonalcoholic steatohepatitis.

HEPATITIS

Both case-control and cohort studies have shown a strong association between chronic hepatitis B carrier rates and increased incidence of HCC. In Taiwanese male postal carriers who were hepatitis B surface antigen (HBsAg)-positive, a 98-fold greater risk for HCC was found compared to HBsAg-negative individuals. The incidence of HCC in Alaskan natives is markedly increased related to a high prevalence of HBV infection. HBV-based HCC may involve rounds of hepatic destruction with subsequent proliferation and not necessarily frank cirrhosis. The increase in Japanese HCC incidence rates in the last three decades is thought to be from hepatitis C. A large-scale World Health Organization (WHO)-sponsored intervention study is currently underway in Asia involving HBV vaccination of the newborn. HCC in African blacks is not associated with severe cirrhosis but is poorly differentiated and very aggressive. Despite uniform HBV carrier rates among the South African Bantu, there is a ninefold difference in HCC incidence between Mozambicans living along the coast and inland. These differences are attributed to the additional exposure to dietary aflatoxin B_1 and other carcinogenic mycotoxins. A typical interval between HCV-associated transfusion and subsequent HCC is approximately 30 years. HCV-associated HCC patients tend to have more frequent and advanced cirrhosis, but in HBV-associated HCC, only half the patients have cirrhosis; the remainder having chronic active hepatitis (Chap. 306).

OTHER ETIOLOGIC CONDITIONS

The 75–85% association of HCC with underlying cirrhosis has long been recognized, more typically with macronodular cirrhosis in Southeast Asia, but also with micronodular cirrhosis (alcohol) in Europe and the United States (Chap. 308). It is still not clear whether cirrhosis itself is a predisposing factor to the development of HCC or whether the underlying causes of the cirrhosis are actually the carcinogenic factors. However, ~20% of U.S. patients with HCC do not have underlying cirrhosis. Several underlying conditions are associated with an increased risk for cirrhosis-associated HCC (Table 92–2), including hepatitis, alcohol, autoimmune chronic active hepatitis, cryptogenic cirrhosis, and NASH. A less common association is with primary biliary cirrhosis and several metabolic diseases including hemochromatosis, Wilson disease, α_1-antitrypsin deficiency, tyrosinemia, porphyria cutanea tarda, glycogenesis types 1 and 3, citrullinemia. and orotic aciduria. The etiology of HCC in those 20% of patients who have no cirrhosis is currently unclear, and their HCC natural history is not well-defined.

CURRENT DIRECTIONS

Many patients have multiple etiologies, and the interactions of either hepatitis or alcohol and smoking, or with aflatoxins, are just beginning to be explored.

CLINICAL FEATURES

SYMPTOMS

These include abdominal pain, weight loss, weakness, abdominal fullness and swelling, jaundice, and nausea (Table 92–3). Presenting signs and symptoms differ somewhat between high- and low-incidence areas. In high-risk areas, especially in South African blacks, the most common symptom is abdominal pain; by contrast, only 40–50% of Chinese and Japanese patients present with abdominal pain. Abdominal swelling may occur as a consequence of ascites due to the underlying chronic liver disease or may be due to a rapidly expanding tumor. Occasionally, central necrosis or acute hemorrhage into the peritoneal cavity leads to death. In countries with

an active surveillance program, HCC tends to be identified at an earlier stage, when symptoms may be due only to the underlying disease. Jaundice is usually due to obstruction of the intrahepatic ducts from underlying liver disease. Hematemesis may occur due to esophageal varices from the underlying portal hypertension. Bone pain is seen in 3–12% of patients, but necropsies show pathologic bone metastases in ~20% of patients. However, 25% of patients may be asymptomatic.

Table 92–3 Hepatocellular Carcinoma Clinical Presentation (*N* = 547)

Symptom	Patient # (%)
No symptom	129(24)
Abdominal pain	219(40)
Other (workup of anemia and various diseases)	64(12)
Routine physical exam finding, elevated LFTs	129(24)
Weight loss	112(20)
Appetite loss	59 (11)
Weakness/malaise	83(15)
Jaundice	30(5)
Routine CT scan screening of known cirrhosis	92(17)
Cirrhosis symptoms (ankle swelling, abdominal bloating, increased girth, pruritus, GI bleed)	98(18)
Diarrhea	7 (1)
Tumor rupture	1
Patient Characteristics	
Mean age (yr)	56 ± 13
Male:Female	3:1
Ethnicity	
White	72 %
Middle Eastern	10 %
Asian	13 %
African American	5 %
Cirrhosis	81 %
No cirrhosis	19 %
Tumor Characteristics	
Hepatic tumor numbers	
1	20 %
2	25 %
3 or more	65 %
Portal vein invasion	75 %
Unilobar	25 %
Bilobar	75 %

Abbreviation: GI, gastrointestinal; LFT, liver function test.

PHYSICAL SIGNS

Hepatomegaly is the most common physical sign, occurring in 50–90% of the patients. Abdominal bruits are noted in 6–25%, and ascites occurs in 30–60% of patients. Ascites should be examined by cytology. Splenomegaly is mainly due to portal hypertension. Weight loss and muscle wasting are common, particularly with rapidly growing or large tumors. Fever is found in 10–50% of patients, from unclear cause. The signs of chronic liver disease may often be present, including jaundice, dilated abdominal veins, palmar erythema, gynecomastia, testicular atrophy, and peripheral edema. Budd-Chiari syndrome can occur due to HCC invasion of the hepatic veins, with tense ascites and a large tender liver (Chap. 308).

PARANEOPLASTIC SYNDROMES

Most paraneoplastic syndromes in HCC are biochemical abnormalities without associated clinical consequences. They include hypoglycemia (also caused by end-stage liver failure), erythrocytosis, hypercalcemia, hypercholesterolemia, dysfibrinogenemia, carcinoid syndrome, increased thyroxin-binding globulin, changes in secondary sex characteristics (gynecomastia, testicular atrophy, and precocious puberty), and porphyria cutanea tarda. Mild hypoglycemia occurs in rapidly growing HCC as part of terminal illness, and profound hypoglycemia may occur, although the cause is unclear. Erythrocytosis occurs in 3–12% of patients and hypercholesterolemia in 10–40%. A high percent of patients have thrombocytopenia or leukopenia, resulting from portal hypertension, and not from cancer infiltration of bone marrow, as in other tumor types.

Staging

Multiple clinical staging systems for HCC have been described. A widely used one has been the American Joint Commission for Cancer (AJCC)/tumor-node-metastasis (TNM) classification. However, the Cancer of the Liver Italian Program (CLIP) system is now popular as it takes the cirrhosis into account, based on the Okuda system

(Table 92–4). Other staging systems have been proposed, and a consensus is needed. They are all based on combining the prognostic features of liver damage with those of tumor aggressiveness and include systems from Spain (Barcelona Clinic Liver Cancer [BCLC]), Japan, Hong Kong, and others (Chinese University Prognostic Index [CUPI], Japan Integrated Staging [JIS], and SLiDe which stands for S, stage; Li, liver damage; De, des-γ-carboxy prothrombin). The best prognosis is for stage I, solitary tumors of less than 2-cm diameter without vascular invasion. Adverse prognostic features include ascites, jaundice, vascular invasion, and elevated α-fetoproteins (AFPs). Vascular invasion in particular has profound effects on prognosis and may be microscopic or macroscopic (visible on CT scans). Most large tumors have microscopic vascular invasion, so full staging can usually be made only after surgical resection. Stage III disease contains a mixture of lymph node—positive and -negative tumors. Stage III patients with positive lymph node disease have a poor prognosis, and few patients survive 1 year. The prognosis of stage IV is poor after either resection or transplantation and 1-year survival is rare. A working staging system based entirely on clinical grounds that incorporates the contribution of the underlying liver disease was originally developed by Okuda et al. (Table 92–4). Patients with Okuda stage III have a dire prognosis because they usually cannot be curatively resected, and the condition of their liver typically precludes chemotherapy.

Table 92–4 Clip and Okuda Staging Systems for Hepatocellular Carcinoma

CLIP Classification			
	Points		
Variables	**0**	**1**	**2**
i. Tumor number	Single	Multiple	–
Hepatic replacement by tumor (%)	<50	<50	>50
ii. Child-Pugh score	A	B	C
iii. α Fetoprotein level (ng/mL)	<400	≥400	–
iv. Portal vein thrombosis (CT)	No	Yes	–
CLIP stages (score = sum of points): CLIP 0, 0 points; CLIP 1, 1 point; CLIP 2, 2 points; CLIP 3, 3 points.			

Okuda Classification							
Tumor extent[a]		**Ascites**		**Albumin (g/L)**		**Bilirubin (mg/dL)**	
≥50%	<50	+	–	≤3	>3	≥3	<3
(+)	(–)	(+)	(–)	(+)	(–)	(+)	(–)
Okuda stages: stage 1, all (–); stage 2, 1 or 2 (+); stage 3, 3 or 4 (+)							

Abbreviation: CLIP, Cancer of the Liver Italian Program.

[a]Extent of liver occupied by tumor

NEW DIRECTIONS

Consensus is needed on staging. These systems will soon be upended by proteomics.

Approach to the Patient: Dysphagia

HISTORY AND PHYSICAL The history is important in evaluating putative predisposing factors, including a history of hepatitis or jaundice, blood transfusion, or use of intravenous drugs. A family history of HCC or hepatitis should be sought and a detailed social history taken to include job descriptions for industrial exposure to possible carcinogenic drugs as well as contraceptive hormones. Physical examination should include assessing stigmata of underlying liver disease such as jaundice, ascites, peripheral edema, spider nevi, palmar erythema, and weight loss. Evaluation of the abdomen for hepatic size, masses or ascites, hepatic nodularity and tenderness, and splenomegaly is needed, as is assessment of overall performance status and psychosocial evaluation.

SEROLOGIC ASSAYS AFP is a serum tumor marker for HCC; however, it is only increased in approximately one-half of U.S. patients. The lens culinaris agglutinin-reactive fraction of AFP (AFP-L3) assay is thought to be more specific. The other widely used assay is that for des-γ-carboxy prothrombin (DCP), a protein induced by vitamin K absence (PIVKA-2). This protein is increased in as many as 80% of HCC patients but may also be elevated in patients with vitamin K deficiency; it is always elevated after Coumadin use. It may predict for portal vein invasion. Both AFP-L3 and DCP are FDA-approved. Many other assays have been developed, such as glypican-3, but none have greater aggregate sensitivity and specificity. In a patient presenting with either a new hepatic mass or other indications of recent hepatic decompensation, carcinoembryonic antigen (CEA), vitamin B12 , AFP, ferritin, PIVKA-2, and anti-mitochondrial Ab should be measured, and standard liver function tests should be performed, including prothrombin time (PT), partial thromboplastin time (PTT), albumin, transaminases, γ-glutamyl transpeptidase, and alkaline phosphatase. Decreases in platelet count and white blood cell count may reflect portal hypertension and associated hypersplenism. Hepatitis A, B, and C serology should be measured. If HBV or HCV serology is positive, quantitative measurements of HBV DNA or HCV RNA are needed.

New Directions Newer biomarkers are being evaluated, especially tissue- and serum-based genomics profiling.

RADIOLOGY An ultrasound examination of the liver is an excellent screening tool. The two characteristic

vascular abnormalities are hypervascularity of the tumor mass (neovascularization or abnormal tumor-feeding arterial vessels) and thrombosis by tumor invasion of otherwise normal portal veins. To determine tumor size and extent and the presence of portal vein invasion accurately, a helical/triphasic CT scan of the abdomen and pelvis, with fast-contrast bolus technique should be performed to detect the vascular lesions typical of HCC. Portal vein invasion is normally detected as an obstruction and expansion of the vessel. A chest CT is used to exclude metastases. MRI can also provide detailed information, especially with the newer contrast agents. Ethiodol (Lipiodol) is an ethiodized oil emulsion retained by liver tumors that can be delivered by hepatic artery injection (5–15 ml) for CT imaging 1 week later. For small tumors, Ethiodol injection is very helpful before biopsy because the histological presence of the dye constitutes proof that the needle biopsied the mass under suspicion. A prospective comparison of triphasic CT, gadolinium-enhanced MRI, ultrasound, and fluorodeoxyglucose positron emission tomography (FDG-PET) showed similar results for CT, MRI, and ultrasound; PET imaging was unsuccessful.

New Directions The altered tumor vascularity that is a consequence of molecularly targeted therapies is the basis for newer imaging techniques including contrast-enhanced ultrasound (CEUS) and dynamic MRI.

PATHOLOGIC DIAGNOSIS Histologic proof of the presence of HCC is obtained through a core liver biopsy of the liver mass under ultrasound guidance, as well as random biopsy of the underlying liver. Bleeding risk is increased compared to other cancers because (1) the tumors are hypervascular, and (2) patients often have thrombocytopenia and decreased liver-dependent clotting factors. Bleeding risk is further increased in the presence of ascites. Tracking of tumor has an uncommon problem. Fine-needle aspirates can provide sufficient material for diagnosis of cancer, but core biopsies are preferred. Tissue architecture allows the distinction between HCC and adenocarcinoma. Laparoscopic approaches can also be used. For patients suspected of having portal vein involvement, a core biopsy of the portal vein may be performed safely. If positive, this is regarded as an exclusion criterion for transplantation for HCC.

New Directions Immunohistochemistry has become mainstream. Prognostic subgroupings are being defined based on growth signaling pathway proteins and genotyping strategies. Furthermore, molecular profiling of the underlying liver has provided evidence for a "field-effect" of cirrhosis in generating recurrent or new HCCs after primary resection.

Screening High-Risk Populations

Screening has not been shown to save lives. Prospective studies in high-risk populations showed that ultrasound was more sensitive than AFP elevations. An Italian study in patients with cirrhosis identified a yearly HCC incidence of 3% but showed no increase in the rate of detection of potentially curable tumors with aggressive screening. Prevention strategies including universal vaccination against hepatitis are more likely to be effective than screening efforts. Despite absence of formal guidelines, most practitioners obtain 6-monthly AFP and CT (or ultrasound) when following high-risk patients (HBV carriers, HCV cirrhosis, family history of HCC).

CURRENT DIRECTIONS

Cost-benefit analysis is not yet convincing, even though screening is intuitively sound. However, studies from areas of high HBV carrier rates have shown a survival benefit for screening as a result of earlier stage at diagnosis. Gamma-glutamyl transpeptidase appears useful for detecting small tumors.

Treatment: Hepatocellular Carcinoma

Most HCC patients have two liver diseases, cirrhosis and HCC, each of which is an independent cause of death. The presence of cirrhosis usually places constraints on resection surgery, ablative therapies, and chemotherapy. Thus patient assessment and treatment planning has to take the severity of the nonmalignant liver disease into account. The clinical management choices for HCC can be complex (Fig. 92-1, Table 92–5, and Table 92–6). The natural history of HCC is highly variable. Patients presenting with advanced tumors (vascular invasion, symptoms, extrahepatic spread) have a median survival of ~4 months, with or without treatment. Treatment results from the literature are difficult to interpret. Survival is not always a measure of the efficacy of therapy because of the adverse effects on survival of the underlying liver disease. A multidisciplinary team, including a hepatologist, interventional radiologist, surgical oncologist, transplant surgeon, and medical oncologist, is important for the comprehensive management of HCC patients.

Figure 92-1

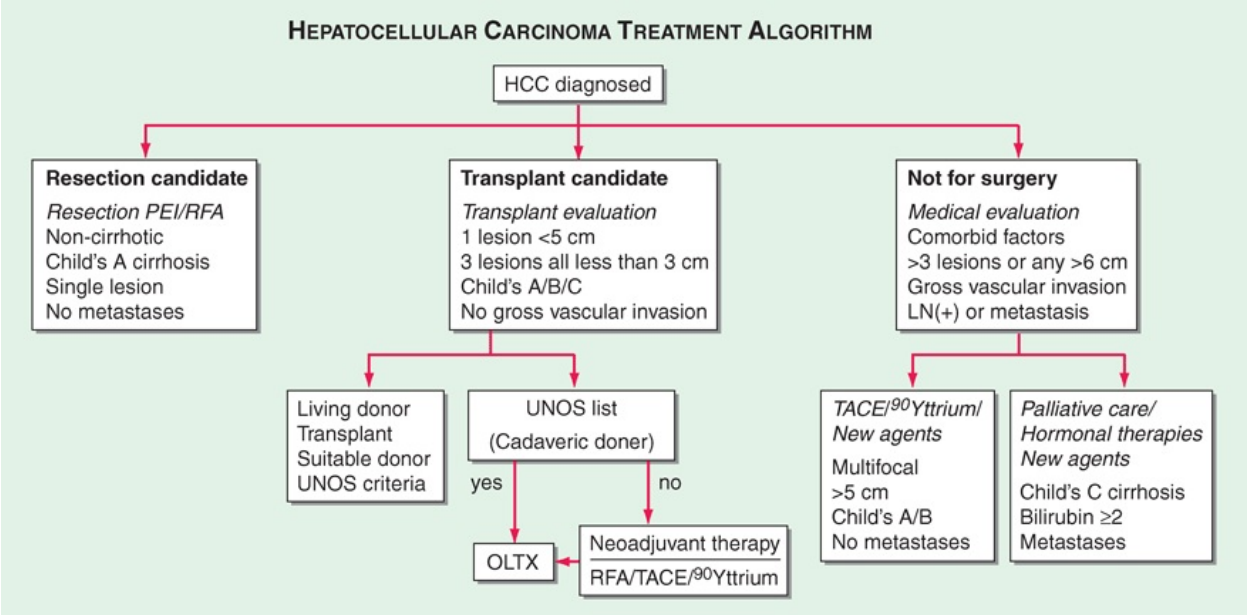

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hepatocellular carcinoma treatment algorithm. Treatment approach to patients with hepatocellular carcinoma. The initial clinical evaluation is aimed at assessing the extent of the tumor and the underlying functional compromise of the liver by cirrhosis. Patients are classified as having resectable disease, unresectable disease, or as transplantation candidates. LN, lymph node; OLTX, orthotopic liver transplantation; PEI, percutaneous ethanol injection; RFA, radiofrequency ablation; TACE, transarterial chemoembolization; UNOS, United Network for Organ Sharing. Child's A/B/C refers to the Child-Pugh classification of liver failure.

Table 92–5 Treatment Options for Hepatocellular Carcinoma

Surgery
Resection
Liver transplantation
Local Ablative Therapies
Cryosurgery
Radiofrequency ablation (RFA)
Percutaneous ethanol injection (PEI)
Regional Therapies: Hepatic Artery Transcatheter Treatments
Transarterial chemotherapy
Transarterial embolization
Transarterial chemoembolization
Transarterial drug-eluting beads
Transarterial radiotherapies:
^{90}Yttrium microspheres
^{131}Iodine - Ethiodol
Conformal External-Beam Radiation
Therapy systemic therapies
Molecularly targeted therapies (sorafenib, etc.)
Chemotherapy
Immunotherapy
Hormonal therapy + growth control
Supportive therapies

Table 92–6 Some Randomized Clinical Trials Involving Transhepatic Artery Chemoembolization (TACE) for Hepatocellular Carcinoma

Author	Year	Agents 1	Agents 2	Survival Effect
Kawaii	1992	Doxorubicin + Embo	Embo	No

Chang	1994	Cisplatin + Embo	Embo	No
Hatanaka	1995	Cisplatin, Doxorubicin + Embo	Same + Lipiodol	No
Uchino	1993	Cisplatin, Doxorubicin + oral FU	Same + Tamoxifen	No
Lin	1988	Embo	Embo + IV FU	No
Yoshikawa	1994	Epirubicin + Ethiodol	Epirubicin	No
Pelletier	1990	Doxorubicin + Gelfoam	None	No
Trinchet	1995	Cisplatin + Gelfoam	None	No
Bruix	1998	Coils and Gelfoam	None	No
Pelletier	1998	Cisplatin + Ethiodol	None	No
Trinchet	1995	Cisplatin + Gelfoam	None	No
Lo	2002	Cisplatin + Ethiodol	None	Yes
Llovet	2002	Doxorubicin + Ethiodol	None	Yes

Abbreviation: Embo, embolization; FU, 5-fluorouracil.

STAGES I AND II HCC

Early-stage tumors are successfully treated using various techniques, including surgical resection, local ablation (thermal or radiofrequency [RFA]), and local injection therapies (Table 92–6). Because the majority of patients with HCC suffer from a field defect in the cirrhotic liver, they are at risk for subsequent multiple primary liver tumors. Many will also have significant underlying liver disease and may not tolerate major surgical loss of hepatic parenchyma, and they may be eligible for orthotopic liver transplant (OLTX). Living related donor transplants have increased in popularity resulting in absence of waiting for a transplant. An important principle in treating early-stage HCC is to use liver-sparing treatments and to focus on treatment of both the tumor and the cirrhosis.

Surgical Excision

The risk of major hepatectomy is high (5–10% mortality rate) due to the underlying liver disease and the potential for liver failure, but acceptable in selected cases. Preoperative portal vein occlusion can sometimes be performed to cause atrophy of the HCC-involved lobe and compensatory hypertrophy of the noninvolved liver, permitting safer resection. Intraoperative ultrasound (US) is useful for planning the surgical approach. The US can image the proximity of major vascular structures that may be encountered during the dissection. In cirrhotic patients, any major liver surgery can result in liver failure. The Child-Pugh classification of liver failure is still a reliable prognosticator for tolerance of hepatic surgery and only Child A patients should be considered for surgical resection. Child B and C patients with stages I and II HCC should be referred for OLTX if appropriate, as well as patients with ascites or a recent history of variceal bleeding. Although open surgical excision is the most reliable, the patient may be better served with a laparoscopic approach to resection, using RFA or percutaneous ethanol injection (PEI). No adequate comparisons of these different techniques have been undertaken and the choice of treatment is usually based on physician skill.

Local Ablation Strategies

Radiofrequency ablation (RFA) uses heat to ablate tumors. The maximum size of the probe arrays allows for a 7-cm zone of necrosis, which would be adequate for a 3- to 4-cm tumor. The heat reliably kills cells within the zone of necrosis. Treatment of tumors close to the main portal pedicles can lead to bile duct injury and obstruction. This limits the location of tumors that are anatomically suited for this technique. RFA can be performed percutaneously with CT or ultrasound guidance, or at the time of laparoscopy with ultrasound guidance.

Local Injection Therapy

Numerous agents have been used for local injection into tumors, most commonly, ethanol (PEI). The relatively soft HCC within the hard background cirrhotic liver allows for injection of large volumes of ethanol into the tumor without diffusion into the hepatic parenchyma or leakage out of the liver. PEI causes direct destruction of cancer cells, but it is not selective for cancer and will destroy normal cells in the vicinity. However, it usually requires multiple injections (average three), in contrast to one for RFA. The maximum size of tumor reliably treated is 3 cm, even with multiple injections.

Current Directions

Resection and RFA each obtain similar results.

Liver Transplantation (Oltx)

A viable option for stages I and II tumors in the setting of cirrhosis is OLTX, with survival approaching that for noncancer cases. OLTX for patients with a single lesion ≤5 cm or three or fewer nodules, each ≤3 cm (Milan criteria), resulted in excellent tumor-free survival (≥70% at 5 years). For advanced HCC, OLTX has been abandoned due to high tumor recurrence rates. Priority scoring for OLTX previously led to HCC patients waiting too long for their OLTX, resulting in some tumors becoming too advanced during the patient's wait for a donated liver. A variety of therapies were used as a "bridge" to OLTX, including RFA, polyethylenimine, and transcatheter arterial chemoembolization (TACE). It seems clear that these pretransplant treatments allow patients to remain on the waiting list longer, giving them greater opportunities to be transplanted. What remains unclear, however, is whether this translates into prolonged survival after transplant. Further, it is not known whether patients who have had their tumor(s) treated preoperatively follow the recurrence pattern predicted by their tumor status at the time of transplant (i.e., post—local ablative therapy), or if they follow the course set by their tumor parameters present before such treatment. The United Network for Organ Sharing (UNOS) point

system for priority scoring of OLTX recipients now includes additional points for patients with HCC. The success of living related donor liver transplantation programs has also led to patients receiving transplantation earlier for HCC and often with greater than minimal tumors.

Current Directions

Expanded criteria for larger HCCs beyond the Milan criteria (one lesion <5 cm or three lesions, each less than 3 cm) are being increasingly accepted by various UNOS areas for OLTX with satisfactory longer-term survival. Furthermore, downstaging HCCs by medical therapy (TACE) is increasingly recognized as acceptable treatment before OLTX.

Adjuvant Therapy

The role of adjuvant chemotherapy for patients after resection or OLTX remains unclear. Both adjuvant and neoadjuvant approaches have been studied, but no clear advantage in disease-free or overall survival has been found. However, a meta-analysis of several trials revealed a significant improvement in disease-free and overall survival. Although analysis of postoperative adjuvant systemic chemotherapy trials demonstrated no disease-free or overall survival advantage, single studies of TACE and neoadjuvant ^{131}I-Ethiodol showed enhanced survival post-resection.

Current Directions

A large adjuvant trial examining resection with or without sorafenib (see below) is in progress.

STAGES III AND IV HCC

Fewer surgical options exist for stage III tumors involving major vascular structures. In patients without cirrhosis, a major hepatectomy is feasible, although prognosis is poor. Patients with Child's A cirrhosis may be resected, but a lobectomy is associated with significant morbidity and mortality rates, and long-term prognosis is poor. Nevertheless, a small percentage of patients will achieve long-term survival, justifying an attempt at resection when feasible. Because of the advanced nature of these tumors, even successful resection can be followed by rapid recurrence. These patients are not considered candidates for transplantation because of the high tumor recurrence rates, unless their tumors can first be down-staged with neoadjuvant therapy. Decreasing the size of the primary tumor allows for less surgery, and the delay in surgery allows for extrahepatic disease to manifest on imaging studies and avoid unhelpful OLTX. The prognosis is poor for stage IV tumors, and no surgical treatment is recommended.

Systemic Chemotherapy

A large number of controlled and uncontrolled clinical studies have been performed with most of the major classes of cancer chemotherapy. No single agent or combination of agents given systemically reproducibly leads to even a 25% response rate or has any effect on survival.

Regional Chemotherapy

In contrast to the dismal results of systemic chemotherapy, a variety of agents given via the hepatic artery have activity for HCC confined to the liver (Table 92–6). Two randomized controlled trials have shown a survival advantage for TACE in a selected subset of patients. One used doxorubicin and the other used cisplatin. Despite the fact that increased hepatic extraction of chemotherapy has been shown for very few drugs, some drugs such as cisplatin, doxorubicin, mitomycin C, and possibly neocarzinostatin, produce substantial objective responses when administered regionally. Few data are available on continuous hepatic arterial infusion for HCC, although pilot studies with cisplatin have shown encouraging responses. Because the reports have not usually stratified responses or survival based on TNM staging, it is difficult to know long-term prognosis in relation to tumor extent. Most of the studies on regional hepatic arterial chemotherapy also use an embolizing agent such as Ethiodol, gelatin sponge particles (Gelfoam), starch (Spherex), or microspheres. Two products are composed of microspheres of defined size ranges—Embospheres (Biospheres) and Contour SE—using particles of 40–120, 100–300, 300–500, and 500–1000 μm in size. The optimal diameter of the particles for TACE has yet to be defined. Consistently higher objective response rates are reported for arterial administration of drugs together with some form of hepatic artery occlusion compared with any form of systemic chemotherapy to date. The widespread use of some form of embolization in addition to chemotherapy has added to its toxicities. These include a frequent, but transient fever, abdominal pain, and anorexia (all in >60% of patients). In addition, >20% of patients have increased ascites or transient elevation of transaminases. Cystic artery spasm and cholecystitis are also not uncommon. However, higher responses have also been obtained. The hepatic toxicities associated with embolization may be ameliorated by the use of degradable starch microspheres, with 50–60% response rates. Two randomized studies of TACE vs. placebo showed a survival advantage for treatment (Table 92–6). In addition, it is not clear that formal oncologic CT response criteria are adequate for HCC. A loss of vascularity on CT without size change may be an index of loss of viability and thus of response to TACE. A major problem that TACE trials have had in showing a survival advantage is that many HCC patients die of their underlying cirrhosis, not the tumor. However, improving quality of life is a legitimate goal of regional therapy.

New Therapies

The major finding has been a survival advantage for oral sorafenib (Nexavar) vs. placebo controls in two randomized trials, leading to its approval by the FDA. However, tumor responses were negligible, and the survival in the treatment arm in Asians was below the placebo arm in the Western trial (Table 92–7). Furthermore, prolonged survival has been reported in phase II trials using newer agents, such as bevacizumab plus erlotinib. Several forms of *radiation therapy* have been used in the treatment of HCC, including external beam radiation and conformal radiation therapy. Radiation hepatitis remains a dose-limiting problem. The pure beta emitter ^{90}Yttrium attached to either glass or resin microspheres has been assessed in phase II trials of

HCC and has encouraging survival effects with minimal toxicities. Randomized trials have yet to be performed. Vitamin K has been assessed in clinical trials at high dosage for its HCC-inhibitory actions. This idea is based on the characteristic biochemical defect in HCC of elevated plasma levels of immature prothrombin (DCP or PIVKA-2), due to a defect in the activity of prothrombin carboxylase, a vitamin K-dependent enzyme. Two vitamin K randomized controlled trials from Japan show decreased tumor occurrence.

Table 92–7 Targeted Therapies in HCC: Trials

Phase III	**Target**	**Survival (mo)**
Sorafenib vs placebo	Raf, VEGFR, PDGFR	10.7 vs. 7.9
Sorafenib vs placebo (Asians)	Raf, WGFR, PDGFR	6.5 vs. 4.2
Phase II		
Sorafenib		9
Sorafenib (Asians)		5
Sunitinib		9.8, 8 (2 trials)
Bevacizumab	VEGF	12.4
Bevacizumab plus erlotinib	VEGF plus EGFR	15.6
Bevacizumab plus capecitabine		8
Erlotinib	EGFR	13, 10.7 (2 trials)
Linifanib	VEGFR, PDGF	9.7
Brivanib	VEGFR, FGFR	10

Abbreviation: EGFR, epidermal growth factor receptor; FGFR, fibroblast growth factor receptor; PDGF, platelet-derived growth factor; PDGFR, platelet-derived growth factor receptor; Raf, rapidly accelerated fibrosarcoma; VEGF, vascular endothelial growth factor; VEGFR vascular endothelial growth factor receptor.

Current Directions A number of new treatments are being evaluated for HCC (Table 92–8). These include the biologicals, such as Raf kinase and vascular endothelial growth factor (VEGF) inhibitors, ^{90}Yttrium looks promising without chemotherapy toxicities, and vitamin K_2 appears to prevent recurrences post-resection. The bottleneck of liver donors for OLTX is at last widening with increasing use of living donors, and criteria for OLTX for larger HCCs are slowly expanding. Patient participation in clinical trials assessing new therapies is encouraged *(www.clinicaltrials.gov)*.

Table 92–8 Some Novel Medical Treatments for Hepatocellular Carcinoma

EGF receptor antagonists: Erlotinib, gefitinib, lapatinib, cetuximab, brivanib
Multi-kinase antagonists: sorafenib, sunitinib
VEGF antagonist: bevacizumab
VEGFR antagonist: ABT-869 (linifanib)
mTOR antagonists: sirolimus, temsirolimus, everolimus
Proteasome inhibitors: bortezomib
Vitamin K
^{131}I – Ethiodol (lipiodol)
^{131}I – Ferritin
^{90}Yttrium microspheres (TheraSphere, SIR-spheres)
^{166}Holmium, ^{188}Rhenium
Three-dimensional conformal radiation
Proton beam high-dose radiotherapy
Gamma knife, CyberKnife
New targets: inhibitors of cyclin dependent kinases (Cdk) and caspases

Abbreviation: EGF, epidermal growth factor; VEGF, vascular endothelial growth factor; VEGFR vascular endothelial growth factor receptor.

Summary (Table 92–5)

THE MOST COMMON MODES OF PATIENT PRESENTATION

1. 1. A patient with known history of hepatitis, jaundice, or cirrhosis, with an abnormality on US or CT scan, or rising AFP or DCP (PIVKA-2)
2. 2. A patient with an abnormal liver function test as part of a routine examination

3. 3. Radiologic workup for liver transplant for cirrhosis 4. Symptoms of HCC including cachexia, abdominal pain, or fever

HISTORY AND PHYSICAL EXAMINATION:

1. Clinical jaundice, asthenia, itching (scratches), tremors, or disorientation
2. Hepatomegaly, splenomegaly, ascites, peripheral edema, skin signs of liver failure

CLINICAL EVALUATION:

1. Blood tests: full blood count (splenomegaly), liver function tests, ammonia levels, electrolytes, AFP and DCP (PIVKA-2), Ca^{2+} and Mg^{2+}; hepatitis B, C, and D serology (and quantitative HBV DNA or HCV RNA, if either is positive); neurotensin (specific for fibrolamellar HCC)
2. Triphasic dynamic helical (spiral) CT scan of liver (if inadequate, then follow with an MRI); chest CT scan; upper and lower gastrointestinal endoscopy (for varices, bleeding, ulcers); and brain scan (only if symptoms suggest)
3. Core biopsy: of the tumor and separate biopsy of the underlying liver

THERAPY (TABLES 92–5 AND 92–6):

1. HCC <2 cm: RFA, PEI, or resection
2. HCC >2 cm, no vascular invasion: liver resection, RFA, or OLTX
3. Multiple unilobar tumors or tumor with vascular invasion: TACE or sorafenib
4. Bilobar tumors, no vascular invasion: TACE with OLTX for patients with tumor response
5. Extrahepatic HCC or elevated bilirubin: sorafenib or bevacizumab plus erlotinib (combination agent trials are in progress)

OTHER PRIMARY LIVER TUMORS

Fibrolamellar HCC (Fl-Hcc)

This rarer variant of HCC has a quite different biology than adult-type HCC. None of the known HCC causative factors seem important here. It is typically a disease of younger adults, often teenagers and predominantly females. It is AFP-negative, but patients typically have elevated blood neurotensin levels, normal liver function tests, and no cirrhosis. Radiology is similar for HCC, except that characteristic adult-type portal vein invasion is less common. Although it is often multifocal in the liver, and therefore not resectable, metastases are common, especially to lungs and locoregional lymph nodes, but survival is often much better than with adult-type HCC. Resectable tumors are associated with 5-year survival >50%. Patients often present with a huge liver or unexplained weight loss, fever, or elevated liver function tests on routine evaluations. These huge masses suggest quite slow growth for many tumors. Surgical resection is the best management option, even for metastases, as these tumors respond much less well to chemotherapy than adult-type HCC. Although several series of OLTX for FL-HCC have been reported, the patients seem to die from tumor recurrences, with a 2- to 5-year lag compared with OLTX for adult-type HCC. Anecdotal responses to gemcitabine plus cisplatin-TACE are reported.

Epithelioid Hemangioendothelioma (Ehe)

This rare vascular tumor of adults is also usually multifocal and can also be associated with prolonged survival, even in the presence of metastases, which are commonly in the lung. There is usually no underlying cirrhosis. Histologically, these tumors are usually of borderline malignancy and express factor VIII, confirming their endothelial origin. OLTX may produce prolonged survival.

Cholangiocarcinoma (Ccc)

CCC typically refers to mucin-producing adenocarcinomas (different from HCC) that arise from the bile ducts. They are grouped by their anatomic site of origin, as intrahepatic, hilar (central, ~65% of CCCs), and peripheral (or distal, ~30% of CCCs). They arise on the basis of cirrhosis less frequently than HCC, excepting primary biliary cirrhosis. Nodular tumors arising at the bifurcation of the common bile duct are called *Klatskin* tumors and are often associated with a collapsed gallbladder, a finding that mandates visualization of the entire biliary tree. The approach to management of central and peripheral CCC is quite different. Incidence is increasing. Although most CCCs have no obvious cause, a number of predisposing factors have been identified. Predisposing diseases include primary sclerosing cholangitis (10–20% of primary sclerosing cholangitis [PSC] patients), an autoimmune disease, and liver fluke in Asians, especially *Opisthorchis viverrini* and *Clonorchis sinensis.* CCC seems also to be associated with any cause of chronic biliary inflammation and injury, with alcoholic liver disease, choledocholithiasis, choledochal cysts (10%), and Caroli's disease (a rare inherited form of bile duct ectasia). CCC most typically presents as painless jaundice, often with pruritus or weight loss. Diagnosis is made by biopsy, percutaneously for peripheral liver lesions, or more commonly via endoscopic retrograde cholangiopancreatography (ERCP) under direct vision for central lesions. The tumors often stain positively for cytokeratins 7, 8, and 19 and negatively for cytokeratin 20. However, histology alone cannot usually distinguish CCC from metastases from colon or pancreas primary tumors. Serologic tumor markers appear to be nonspecific, but CEA, CA 19–9, and CA-125 are often elevated in CCC patients and are useful for following response to therapy. Radiologic evaluation typically starts with ultrasound, which is very useful in visualizing dilated bile ducts, and then proceeds with either MRI or magnetic resonance

cholangiopancreatography (MRCP) or helical CT scans. Invasive cholangiopancreatography (ERCP) is then needed to define the biliary tree and obtain a biopsy or is needed therapeutically to decompress an obstructed biliary tree with internal stent placement. If that fails, then percutaneous biliary drainage will be needed, with the biliary drainage flowing into an external bag. Central tumors often invade the porta hepatis, and loco-regional lymph node involvement by tumor is frequent.

Treatment: Cholangiocarcinoma

Hilar CCC is resectable in ~30% of patients and usually involves bile duct resection and lymphadenectomy. Typical survival is approximately 24 months, with recurrences being mainly in the operative bed but with ~30% in the lungs and liver. Distal CCC, which involves the main ducts, is normally treated by resection of the extrahepatic bile ducts, often with pancreaticoduodenectomy. Survival is similar. Due to the high rates of locoregional recurrences or positive surgical margins, many patients receive postoperative adjuvant radiotherapy. Its effect on survival has not been assessed. Intraluminal brachyradiotherapy has also shown some promise. However, photodynamic therapy enhanced survival in one study. In this technique, sodium porfimer is injected intravenously and then subjected to intraluminal red light laser photoactivation. OLTX has been assessed for treatment of unresectable CCC. Five-year survival was ~20%, so enthusiasm waned. However, neoadjuvant radiotherapy with sensitizing chemotherapy has shown better survival rates for CCC treated by OLTX and is currently used by UNOS for perihilar CCC, size <3 cm with neither intrahepatic or extrahepatic metastases. Multiple chemotherapeutic agents have been assessed for activity and survival in unresectable CCC. Most have been inactive. However, both systemic and hepatic arterial gemcitabine have shown promising results. The combination of cisplatin plus gemcitabine has produced a survival advantage compared with gemcitabine alone and is considered standard therapy for unresectable CCC.

Gallbladder Cancer (Gb Ca)

GB Ca has an even worse prognosis than CCC, and with typical survival ~6 months or less. Women are affected much more commonly than men (4:1), unlike HCC or CCC, and GB Ca occurs more frequently than CCC. Most patients have a history of antecedent gallstones, but very few patients with gallstones develop GB Ca (~0.2%). It presents similarly to CCC and is often diagnosed unexpectedly during gallstone or cholecystitis surgery. Presentation is typically that of chronic cholecystitis, chronic right upper quadrant pain and weight loss. Useful but nonspecific serum markers include CEA and CA 19-9. CT scans or MRCP typically reveal a gallbladder mass. The mainstay of treatment is surgical, either simple or radical cholecystectomy for stages I or II disease, respectively. Survival rates are near 100% at 5 years for stage I, and range from 60–90% at 5 years for stage II. More advanced GB Ca has worse survival, and many patients are unresectable. Adjuvant radiotherapy, used in the presence of local lymph node disease, has not been shown to enhance survival. Chemotherapy is not useful in advanced or metastatic GB Ca.

Carcinoma of the Ampulla of Vater

This tumor arises within 2 cm of the distal end of the common bile duct and is mainly (90%) an adenocarcinoma. Locoregional lymph nodes are commonly involved (50%), and the liver is the most frequent site for metastases. The most common clinical presentation is jaundice, and many patients also have pruritus, weight loss, and epigastric pain. Initial evaluation is performed with an abdominal ultrasound to assess vascular involvement, biliary dilation, and liver lesions. This is followed by a CT scan, or MRI and especially MRCP. The most effective therapy is resection by pylorus-sparing pancreaticoduodenectomy, an aggressive procedure resulting in better survival rates than with local resection. Survival rates are ~25% at 5 years in operable patients with involved lymph nodes and ~50% in patients without involved nodes. Unlike CCC, approximately 80% of patients are thought to be resectable at diagnosis. Adjuvant chemotherapy or radiotherapy has not been shown to enhance survival. For metastatic tumors, chemotherapy is currently experimental.

Tumors Metastatic to the Liver

These are predominantly from colon, pancreas, and breast primary tumors but can originate from any organ primary. Ocular melanomas are prone to liver metastasis. Tumor spread to the liver normally carries a poor prognosis for that tumor type. Colorectal and breast hepatic metastases were previously treated with continuous hepatic arterial infusion chemotherapy. However, more effective systemic drugs for each of these two cancers, especially the addition of oxaliplatin to colorectal cancer regimens, have reduced the use of hepatic artery infusion therapy. In a large randomized study of systemic versus infusional plus systemic chemotherapy for resected colorectal metastases to the liver, the patients receiving infusional therapy had no survival advantage, mainly due to extrahepatic tumor spread. ^{90}Yttrium resin beads are approved in the United States for treatment of colorectal hepatic metastases. The role of this modality, either alone or in combination with chemotherapy, is being evaluated in many centers. Palliation my be obtained from chemoembolization, PEI, or RFA.

Benign Liver Tumors

Three common benign tumors occur and all are found predominantly in women. They are *hemangiomas, adenomas,* and *focal nodular hyperplasia* (FNH). FNH is typically benign, and usually no treatment is needed. Hemangiomas are the most common and are entirely benign. Treatment is unnecessary unless their expansion causes symptoms. Adenomas are associated with contraceptive hormone use. They can cause pain and can bleed or rupture, causing acute problems. Their main interest for the physician is a low potential for malignant change and a 30% risk of bleeding. For this reason, considerable effort has gone into differentiating these three entities radiologically. On discovery of a liver mass, patients are usually advised to stop taking sex steroids, as adenoma regression may then occasionally occur. Adenomas can often be large masses ranging from 8–15 cm. Due to their size and definite, but low, malignant potential and potential for bleeding, adenomas are typically resected. The most useful diagnostic differentiating tool is a triphasic CT scan performed with HCC fast bolus protocol for arterial-phase imaging, together with subsequent delayed venous-phase imaging.

Adenomas usually do not appear on the basis of cirrhosis, although both adenomas and HCCs are intensely vascular on the CT arterial phase and both can exhibit hemorrhage (40% of adenomas). However, adenomas have smooth, well-defined edges, and enhance homogeneously, especially in the portal venous phase on delayed images, when HCCs no longer enhance. FNHs exhibit a characteristic central scar that is hypovascular on the arterial-phase and hypervascular on the delayed-phase CT images. MRI is even more sensitive in depicting the characteristic central scar of FNH.

FURTHER READINGS

Carr BI et al: Therapeutic equivalence in survival for hepatic arterial chemoembolization and ^{90}Yttrium microspheres (Y90) treatments in unresectable HCC. Cancer 116:1305, 2010 [PMID: 20066715]

Hoshida Y et al: Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med 359:1995, 2008 [PMID: 18923165]

Llovet JM et al: Sorafenib in advanced hepatocellular carcinoma. N Engl J Med 359:378, 2008 [PMID: 18650514]

Mazzaferro V et al: Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: A retrospective, exploratory analysis. Lancet Oncol 10:35, 2009 [PMID: 19058754]

Noda I et al: Regular surveillance by imaging for early detection and better prognosis of hepatocellular carcinoma in patients infected with hepatitis C virus. J Gastroenterol 45:105, 2010 [PMID: 19866332]

Pancoska P et al: Network-based analysis of survival for unresectable hepatocellular carcinoma. Semin Oncol 37:170, 2010 [PMID: 20494709]

Sherman M: Hepatocellular carcinoma: Epidemiology, surveillance and diagnosis. Semin Liver Dis 30:3, 2010 [PMID: 20175029]

Valle J et al: Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. N Engl J Med 362:1273, 2010 [PMID: 20375404]

Villaneuva A et al: Hepatocellular carcinoma: Novel molecular approaches for diagnosis, prognosis and therapy. Annu Rev Med 61:317, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PANCREATIC CANCER: INTRODUCTION

Pancreatic cancer is the fourth leading cause of cancer death in the United States and is associated with a poor prognosis. Endocrine tumors affecting the pancreas are discussed in Chap. 350. Infiltrating ductal adenocarcinomas, the subject of this chapter, account for the vast majority of cases and arise most frequently in the head of pancreas. At the time of diagnosis 85–90% of patients have inoperable or metastatic disease, which is reflected in the 5-year survival rate of only 5% for all stages combined. An improved 5-year survival of up to 20% may be achieved when the tumor is detected at an early stage and when complete surgical resection is accomplished.

EPIDEMIOLOGY

Pancreatic cancer represents 3% of all newly diagnosed malignancies in the United States. The most common age group at diagnosis is 60–79 years for both sexes. Pancreatic cancer will be diagnosed in approximately 43,140 patients and account for 36,800 deaths in 2010. Over the past 30 years, 5-year survival rates have not improved substantially.

RISK FACTORS

Cigarette smoking may be the cause of up to 20–25% of all pancreatic cancers and is the most common environmental risk factor for this disease. Other risk factors are not well established due to inconsistent results from epidemiological studies, but include chronic pancreatitis and diabetes. It is difficult to evaluate whether these conditions are causally related, or develop as a consequence of cancer. Alcohol does not appear to be a risk factor unless excess consumption gives rise to chronic pancreatitis.

GENETIC CONSIDERATIONS

Pancreatic cancer is associated with a number of well-defined molecular hallmarks. The most frequent genetic aberrations comprise *KRAS* mutations, mostly affecting codon 12, which are observed in 60–75% of pancreatic cancers. The tumor-suppressor genes *p16, p53,* and *SMAD4* are frequently inactivated; the *p16* gene locus on chromosome 9p21 is deleted in up to 95% of tumors, the *p53* gene is inactivated by mutation or deleted in 50–70% of tumors, and the *SMAD4* gene is deleted in 55% of pancreatic tumors. Furthermore, *SMAD4* gene inactivation is associated with poorer survival in patients with surgically resected pancreatic adenocarcinoma. *IGF-1R* and focal adhesion kinase (*FAK*) interact to promote cell proliferation and survival, and their simultaneous inhibition synergistically inhibits pancreatic cell growth. Overexpression and/or aberrant activation of *c-Src* is frequently observed, which results in cell adhesion, enhanced migration, invasion, and cell proliferation. Survivin is overexpressed in more than 80% of pancreatic tumors, which results in resistance to apoptosis, and genomic sequencing has identified *PALB2* as a susceptibility gene for pancreatic cancer.

Up to 16% of pancreatic cancers are thought to be inherited. This occurs in three separate clinical settings: (1) familial multi-organ cancer syndromes, (2) genetically driven chronic diseases, and (3) familial pancreatic cancer with as yet unidentified genetic abnormalities, which comprise the largest proportion of inherited pancreatic cancer. The familial multi-organ cancer syndromes consist of Peutz-Jeghers syndrome, familial atypical multiple mole melanoma (FAMMM), familial breast-ovarian cancer associated with germline mutations in *BRCA1* and *BRCA2*, hereditary nonpolyposis colorectal cancer (HNPCC), familial adenomatous polyposis (FAP), and Li-Fraumeni syndrome. Peutz-Jeghers, associated with mutations in the *STK11* gene, carries the highest lifetime risk of pancreatic cancer with a relative risk of approximately 132-fold above that of the general population. Genetically driven chronic causes of pancreatic cancer include hereditary pancreatitis, cystic fibrosis, and ataxia telangiectasia. The absolute number of affected first-degree relatives is also correlated with increased cancer risk, and patients with at least two first-degree relatives with pancreatic cancer should be considered to have familial pancreatic cancer until proven otherwise.

SCREENING AND EARLY DETECTION

Screening is not routinely recommended as putative tumor markers such as Ca 19-9 and CEA have insufficient sensitivity, and computed tomography (CT) has inadequate resolution to detect pancreatic dysplasia. Endoscopic ultrasound (EUS) is a more promising screening tool, and preclinical efforts are focused on identifying biomarkers that may detect pancreatic cancer at an early stage. Consensus

practice recommendations based largely on expert opinion have chosen a threshold of >tenfold increased risk for developing pancreatic cancer to select individuals who may benefit from screening. This includes family members with ≥3 first-degree relatives with pancreatic cancer, and patients with FAMMM, Peutz-Jeghers syndrome, or hereditary pancreatitis.

CLINICAL FEATURES

Clinical Presentation

Obstructive jaundice occurs frequently when the cancer is located in the head of pancreas. This may be accompanied by symptoms of abdominal discomfort, pruritus, lethargy, and weight loss. Less common presenting features include epigastric pain, backache, new onset diabetes mellitus, and acute pancreatitis caused by pressure effects on the pancreatic duct. Nausea and vomiting, resulting from gastroduodenal obstruction, may also be a symptom of this disease.

Physical Signs

Patients can present with jaundice and cachexia, and scratch marks may be present. Of patients with operable tumors 25% have a palpable gall bladder (Courvoisier's sign). Physical signs related to the development of distant metastases include hepatomegaly, ascites, left supraclavicular lymphadenopathy (Virchow's node), and periumbilical lymphadenopathy (Sister Mary Joseph's nodes).

DIAGNOSIS

Diagnostic Imaging

Patients who present with clinical features suggestive of pancreatic cancer undergo imaging to confirm the presence of a tumor, and to establish whether the mass is likely to be inflammatory or malignant in nature. Other imaging objectives include the local and distant staging of the tumor, which will determine resectability and provide prognostic information. Dual phase, contrast-enhanced spiral CT is the imaging modality of choice (Fig. 93-1). It provides accurate visualization of surrounding viscera, vessels, and lymph nodes, thus determining tumor resectability. Intestinal infiltration, and liver and lung metastases are also reliably depicted on CT. There is no advantage of magnetic resonance imaging (MRI) over CT in predicting tumor resectability, but selected cases may benefit from MRI to characterize the nature of small indeterminate liver lesions and to evaluate the cause of biliary dilatation when no obvious mass is seen on CT. Endoscopic retrograde cholangiopancreatography (ERCP) is useful for revealing small pancreatic lesions, identifying stricture or obstruction in pancreatic or common bile ducts, and facilitates stent placement (Fig. 93-2). Magnetic resonance cholangiopancreatography (MRCP) is a noninvasive method for accurately depicting the level and degree of bile and pancreatic duct dilatation. EUS is highly sensitive in detecting lesions less than 3 cm in size, and is useful as a local staging tool for assessing vascular invasion and lymph node involvement. Positron-emission tomography with fluorodeoxyglucose positron emission tomography (FDG-PET) should be considered before surgery or radical chemoradiotherapy (CRT), as it is superior to conventional imaging in detecting distant metastases.

Figure 93-1

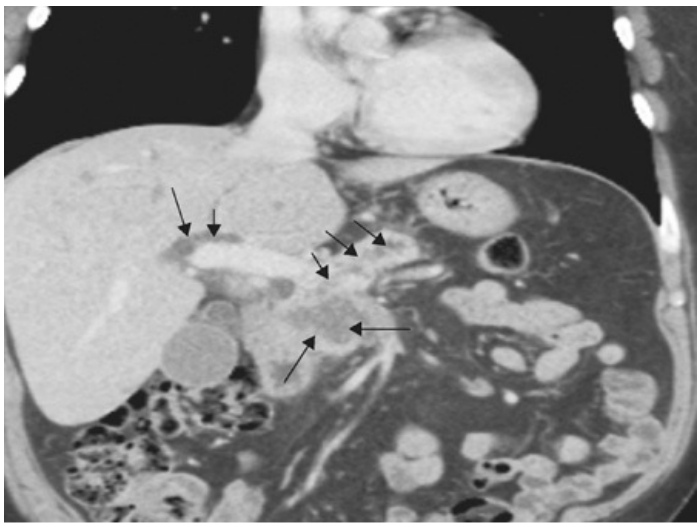

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronal CT showing pancreatic cancer and dilated intrahepatic and pancreatic ducts (arrows).

Figure 93-2

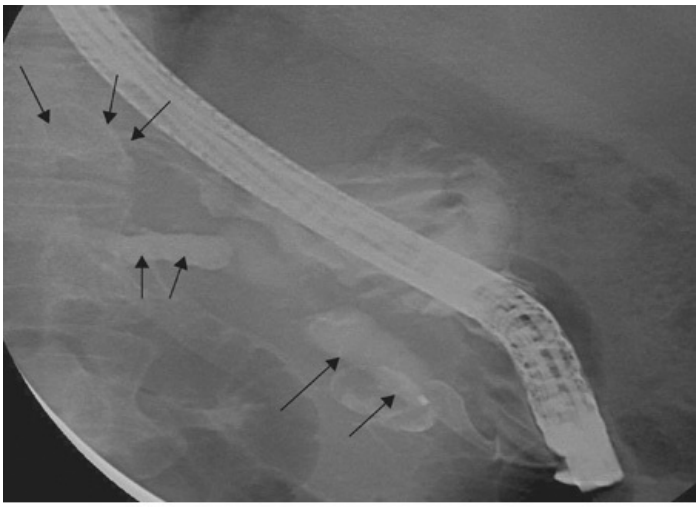

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

ERCP showing contrast in dilated pancreatic duct (arrows).

Tissue Diagnosis and Cytology

Preoperative confirmation of malignancy is not always necessary in patients with radiological appearances consistent with operable pancreatic cancer. However, EUS-guided fine-needle aspiration is the technique of choice when there is any doubt, and also for use in patients who require neoadjuvant treatment. It has an accuracy of approximately 90% and has a smaller risk of intraperitoneal dissemination compared with the percutaneous route. Percutaneous biopsy of the pancreatic primary or liver metastases is only acceptable in patients with inoperable or metastatic disease. ERCP is a useful method for obtaining ductal brushings, but the diagnostic value of pancreatic juice sampling is only in the order of 25–30%.

Serum Markers

Tumor-associated carbohydrate antigen 19-9 (CA 19-9) is elevated in approximately 70–80% of patients with pancreatic carcinoma, but is not recommended as a routine diagnostic or screening test as its sensitivity and specificity are inadequate for accurate diagnosis. Preoperative CA 19-9 levels correlate with tumor stage, and postresection CA 19-9 level has prognostic value. It is an indicator of asymptomatic recurrence in patients with completely resected tumors and is used as a biomarker of response in patients with advanced disease undergoing chemotherapy. A number of studies have established a high pretreatment CA 19-9 level as an independent prognostic factor.

STAGING

The American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging of pancreatic cancer takes into account the location and size of the tumor, the involvement of lymph nodes, and distant metastasis. This information is then combined to assign a stage (Fig. 93-3). From a practical standpoint, patients are grouped according to whether the cancer is resectable, locally advanced (unresectable, but without distant spread), or metastatic.

Figure 93-3

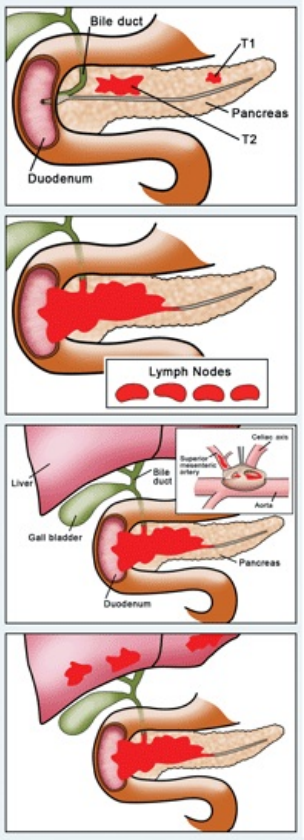

AJCC Stage	TNM Stage	Extent of Tumor	5 year Survival	Stage at Presentation (14% Unknown)
I	T1/N0	Limited to pancreas ≤ 2cm	20%	7%
	T2/N0	Limited to pancreas > 2cm		
II	T3 or N1	Beyond pancreas or regional lymph node metastases	8%	26%
III	T4 any N	Involves celiac axis or superior mesenteric artery		
IV	M1	Distant metastases	2%	53%

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Staging of pancreatic cancer, and survival according to stage. (*Illustration by Stephen Millward.*)

Treatment: Pancreatic Cancer

RESECTABLE DISEASE

Approximately 10% of patients present with localized nonmetastatic disease that is potentially suitable for surgical resection. Approximately 30% of patients have R1 resection (microscopic residual disease) following surgery. Those who undergo R0 resection (no microscopic or macroscopic residual tumor), and who receive adjuvant treatment have the best chance of cure, with an estimated median survival of 20–23 months and a 5-year survival of approximately 20%. Outcomes are more favorable in patients with small (< 3cm), well-differentiated tumors, and lymph node-negative disease.

Patients should have surgery in dedicated pancreatic centers that have lower postoperative morbidity and mortality rates. The standard surgical procedure for patients with tumors of the pancreatic head or uncinate process is a pylorus-preserving pancreaticoduodenectomy (modified Whipple's procedure). The procedure of choice for tumors of the pancreatic body and tail is a distal pancreatectomy, which routinely includes splenectomy.

Postoperative treatment, either chemotherapy or CRT, improves long-term outcomes in this group of patients. Adjuvant chemotherapy, comprising six cycles of fluorouracil (5FU) and folinic acid (FA) or gemcitabine, is common practice in Europe based on data from three randomized controlled trials (Table 93–1): Results from the European Study Group for Pancreatic Cancer 1 trial (ESPAC-1) revealed a median survival improvement from 14.7 months with surgery alone to 20.1 months with surgery plus adjuvant 5FU/FA, and patients did not benefit from CRT in this study. The Charité Onkologie trial (CONKO 001) found that the use of gemcitabine after complete resection significantly delayed the development of recurrent disease compared with surgery alone. The ESPAC-3 trial, which investigated the benefit of adjuvant 5FU/FA versus gemcitabine, revealed no survival difference between the two drugs. However, the safety profile of adjuvant gemcitabine, with respect to the incidence of stomatitis and diarrhea, was superior to 5FU/FA.

Table 93–1 Phase III Studies of Adjuvant Chemotherapy in Resected Pancreatic Cancer

			Survival	
Study	**Comparator Arm**	**Patient Number**	**PFS/DFS (months)**	**Median survival (months)**
ESPAC 1 Neoptolemos et al. (2004)	Chemotherapy (Folinic acid + bolus 5FU) vs No chemotherapy	550	PFS 15.3 vs 9.4. (p = 0.02)	20.1 vs 14.7 (HR 0.71, 95% CI 0.55 to 0.92, p = 0.009)
CONKO 001 Oettle et al. (2007)	Gemcitabine vs Observation	368	Median DFS 13.4 vs 6.9 (p < 0.001)	22.1 vs 20.2 (p = 0.06)
ESPAC 3 Neoptolemos et al. (2010)	5FU/LV vs Gemcitabine	1088		23 vs 23.6 (HR 0.94, 95% CI 0.81 to 1.08, p = 0.39)

Abbreviation: CI, confidence interval; CONKO, charite ONKOlogie; DFS, disease free survival; ESPAC, European Study Group for Pancreatic Cancer; 5FU, fluorouracil; HR, hazard ratio; LV, leucovorin; PFS, progression-free survival.

A different treatment strategy using adjuvant 5FU based CRT following gemcitabine as advocated by the Radiation Therapy Oncology Group (RTOG) 97-04 trial is preferred in the United States. This approach may be most beneficial in patients with bulky tumors involving the pancreatic head, and in patients with R1 resection.

INOPERABLE LOCALLY ADVANCED DISEASE

Approximately 30% of patients present with locally advanced unresectable but nonmetastatic pancreatic carcinoma. The median survival with gemcitabine is 9 months, and patients who respond to or achieve stable disease after 3–6 months of gemcitabine may derive benefit from consolidation radiotherapy.

METASTATIC DISEASE

Approximately 60% of patients with pancreatic cancer present with metastatic disease. Patients with poor performance status do not benefit from chemotherapy. Gemcitabine is the standard treatment with a median survival of 6 months and a 1-year survival rate of only 20%. The toxicities associated with gemcitabine need to be weighed against the potential benefits of treatment.

Adding other drugs to gemcitabine to improve outcome has been generally unsuccessful with the exception of erlotinib, an oral HER1/EGFR tyrosine kinase inhibitor. The combination of erlotinib with gemcitabine resulted in an improved 1-year survival compared with gemcitabine alone (23% versus 17%,p = 0.023) (Table 93–2). Capecitabine, an oral fluoropyrimidine, has been combined with gemcitabine (GEM-CAP) in a phase III trial that showed an improvement in response rate and progression-free survival over single-agent gemcitabine, but no survival benefit. However, pooling of two other randomized controlled trials with this trial in a meta-analysis resulted in a survival advantage with GEM-CAP.

Table 93–2 Selected Phase III Studies Evaluating Chemotherapy Treatment in Advanced Pancreatic Cancer

			Survival	
Study	**Comparator Arm**	**Patient Number**	**PFS (months)**	**Median survival (months)**
Moore M et al. (2007)	Gemcitabine vs Gemcitabine + erlotinib	569	3.55 vs 3.75 (HR 0.77, 95% CI 0.64 to 0.92, p = 0.004)	5.91 vs 6.24 (HR 0.82, 95% CI 0.69 to 0.99, p = 0.038)
GEM-CAP Cunningham et al. (2009)	Gemcitabine vs Gemcitabine + capecitabine (GEM-CAP)	533	3.8 vs 5.3 (HR 0.78, 95% CI 0.66 to 0.93, p = 0.004)	6.2 vs 7.1 (HR 0.86, 95% CI 0.72 to 1.02, p = 0.08)
GEM-CAP meta-analysis Cunningham et al. (2009)	Gemcitabine vs GEM-CAP	935		Overall survival in favor of GEM-CAP (HR 0.86, 95% CI 0.75 to 0.98, p = 0.02)

A trial in good performance status patients with metastatic pancreatic cancer showed improved survival with the combination of 5FU/FA, irinotecan and oxaliplatin (FOLFIRINOX) compared with gemcitabine, but with increased toxicity. Nab-paclitaxel (Abraxane), an albumin bound nano-particle formulation of paclitaxel, given with gemcitabine also shows promising activity.

FUTURE DIRECTIONS

The early detection and future treatment of pancreatic cancer relies on an improved understanding of molecular pathways involved in the development of this disease. This will ultimately lead to the discovery of novel agents, and the identification of patient groups who are likely to benefit most from targeted therapy.

FURTHER READINGS

Cascinu S, Jelic S: Pancreatic cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol 20:37, 2009[PMID: 19454458]

Cunningham D et al: Phase III randomised comparison of gemcitabine versus gemcitabine plus capecitabine in patients with advanced pancreatic cancer. J Clin Oncol 27:5513, 2009[PMID: 19858379]

Edge SB, Byrd DR (eds): *Exocrine and Endocrine Pancreas, AJCC Cancer Staging Manual* , 7th ed, New York, Springer, 2010, pp 241–249

Moore MJ et al: Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: A phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 26:1960, 2007

Neoptolemos JP et al: Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: A randomised controlled trial. JAMA 304:1073, 2010[PMID: 20823433]

Oettle H et al: Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: A randomized controlled trial. JAMA 297:267, 2007[PMID: 17227978]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BLADDER CANCER

A transitional cell epithelium lines the urinary tract from the renal pelvis to the ureter, urinary bladder, and the proximal two-thirds of the urethra. Cancers can occur at any point: 90% of malignancies develop in the bladder, 8% in the renal pelvis, and the remaining 2% in the ureter or urethra. Bladder cancer is the fourth most common cancer in men and the thirteenth in women, with an estimated 70,530 new cases and 14,680 deaths in the United States predicted for the year 2010. The almost 5:1 ratio of incidence to mortality reflects the higher frequency of the less lethal superficial variants compared to the more lethal invasive and metastatic variants. The incidence is three times higher in men than in women and twofold higher in whites than blacks, with a median age at diagnosis of 65 years.

Once diagnosed, urothelial tumors exhibit polychronotropism—the tendency to recur over time and in new locations in the urothelial tract. As long as urothelium is present, continuous monitoring of the tract is required.

Epidemiology

Cigarette smoking is believed to contribute to up to 50% of the diagnosed urothelial cancers in men and up to 40% in women. The risk of developing a urothelial malignancy in male smokers is increased two- to fourfold relative to nonsmokers and continues for 10 years or longer after cessation. Other implicated agents include the aniline dyes, the drugs phenacetin and chlornaphazine, and external beam radiation. Chronic cyclophosphamide exposure may also increase risk, whereas vitamin A supplements appear to be protective. Exposure to *Schistosoma haematobium*, a parasite found in many developing countries, is associated with an increase in both squamous and transitional cell carcinomas of the bladder.

Pathology

Clinical subtypes are grouped into three categories: 75% are superficial, 20% invade muscle, and 5% are metastatic at presentation. Staging of the tumor within the bladder is based on the pattern of growth and depth of invasion: Ta lesions grow as exophytic lesions; carcinoma in situ (CIS) lesions start on the surface and tend to invade. The revised tumor, node, metastasis (TNM) staging system is illustrated in **Fig. 94-1**. About half of invasive tumors presented originally as superficial lesions that later progressed. Tumors are also rated by grade. Grade I lesions (highly differentiated tumors) rarely progress to a higher stage, whereas grade III tumors do.

Figure 94-1

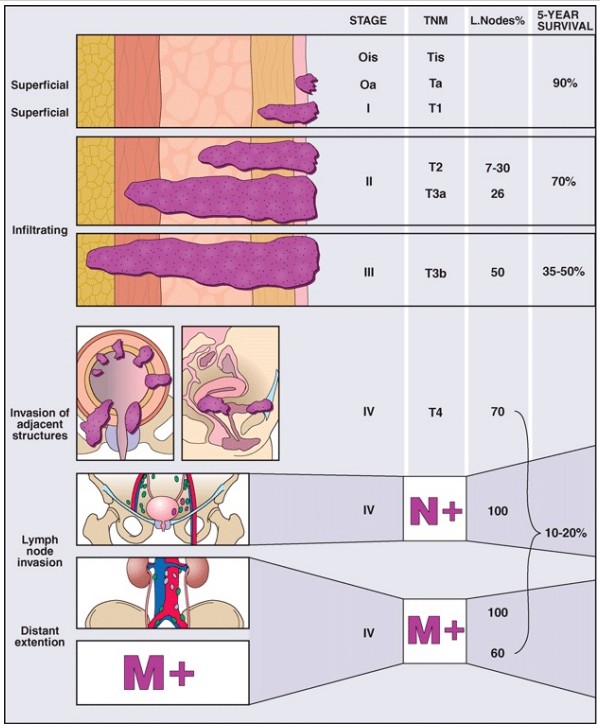

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bladder staging. TNM, tumor, node, metastasis.

More than 95% of urothelial tumors in the United States are transitional cell in origin. Pure squamous cancers with keratinization constitute 3%, adenocarcinomas 2%, and small cell tumors (with paraneoplastic syndromes) <1%. Adenocarcinomas develop primarily in the urachal remnant in the dome of the bladder or in the periurethral tissues; some assume a signet cell histology. Lymphomas and melanomas are rare. Of the transitional cell tumors, low-grade papillary lesions that grow on a central stalk are most common. These tumors are very friable, have a tendency to bleed, are at high risk for recurrence, and yet rarely progress to the more lethal invasive variety. In contrast, CIS is a high-grade tumor that is considered a precursor of the more lethal muscle-invasive disease.

Pathogenesis

The multicentric nature of the disease and high rate of recurrence has led to the hypothesis of a field defect in the urothelium that results in a predisposition to cancer. Molecular genetic analyses suggest that the superficial and invasive lesions develop along distinct molecular pathways in which primary tumorigenic aberrations precede secondary changes associated with progression to a more advanced stage. Low-grade papillary tumors that do not tend to invade or metastasize harbor constitutive activation of the receptor-tyrosine kinase-Ras signal transduction pathway and a high frequency of fibroblast growth factor receptor 3 (FGFR3) mutations. In contrast, CIS and invasive tumors have a higher frequency of *TP53* and *RB* gene alternations. Within all clinical stages, including Tis, T1, and T2 or greater lesions, tumors with alterations in *p53*, *p21*, and/or *RB* have a higher probability of recurrence, metastasis, and death from disease.

Clinical Presentation, Diagnosis, and Staging

Hematuria occurs in 80–90% of patients and often reflects exophytic tumors. The bladder is the most common source of gross hematuria (40%), but benign cystitis (22%) is a more common cause than bladder cancer (15%) (Chap. 44). Microscopic hematuria is more commonly of prostate origin (25%); only 2% of bladder cancers produce microscopic hematuria. Once hematuria is documented, a urinary cytology, visualization of the urothelial tract by CT or intravenous pyelogram, and cystoscopy are recommended if no other etiology is found. Screening asymptomatic individuals for hematuria increases the diagnosis of tumors at an early stage but has not been shown to prolong life. After hematuria, irritative symptoms are the next most common presentation, which may reflect in situ disease. Obstruction of the ureters may cause flank pain. Symptoms of metastatic disease are rarely the first presenting sign.

The endoscopic evaluation includes an examination under anesthesia to determine whether a palpable mass is present. A flexible endoscope is inserted into the bladder, and bladder barbotage is performed. The visual inspection includes mapping the location, size, and number of lesions, as well as a description of the growth pattern (solid vs. papillary). An intraoperative video is often recorded. All visible tumors should be resected, and a sample of the muscle underlying the tumor should be obtained to assess the depth of invasion. Normal-appearing areas are biopsied at random to ensure no field defect. A notation is made as to whether a tumor was completely or incompletely resected. Selective catheterization and visualization of the upper tracts should be performed if the cytology is positive and no disease is visible in the bladder. Ultrasonography, CT, and/or MRI may help to determine whether a tumor extends to perivesical fat (T3) and to document nodal spread. Distant metastases are assessed by CT of the chest and abdomen, MRI, or radionuclide imaging of the skeleton.

Treatment: Bladder Cancer

Management depends on whether the tumor invades muscle and whether it has spread to the regional lymph nodes and beyond. The probability of spread increases with increasing T stage.

SUPERFICIAL DISEASE

At a minimum, the management of a superficial tumor is complete endoscopic resection with or without intravesical therapy. The decision to recommend intravesical therapy depends on the histologic subtype, number of lesions, depth of invasion, presence or absence of CIS, and antecedent history. Recurrences develop in upward of 50% of cases, of which 5–20% progress to a more advanced stage. In general, solitary papillary lesions are managed by transurethral surgery alone. CIS and recurrent disease are treated by transurethral surgery followed by intravesical therapy.

Intravesical therapies are used in two general contexts: as an adjuvant to a complete endoscopic resection to prevent recurrence or, less commonly, to eliminate disease that cannot be controlled by endoscopic resection alone. Intravesical treatments are advised for patients with recurrent disease, >40% involvement of the bladder surface by tumor, diffuse CIS, or T1 disease. The standard intravesical therapy, based on randomized comparisons, is bacillus Calmette-Guerin (BCG) in six weekly instillations, followed by monthly maintenance administrations for ≥1 year. Other agents with activity include mitomycin-C, interferon (IFN), and gemcitabine. The side effects of intravesical therapies include dysuria, urinary frequency, and, depending on the drug, myelosuppression or contact dermatitis. Rarely, intravesical BCG may produce a systemic illness associated with granulomatous infections in multiple sites that requires antituberculin therapy.

Following the endoscopic resection, patients are monitored for recurrence at 3-month intervals during the first year. Recurrence may develop anywhere along the urothelial tract, including the renal pelvis, ureter, or urethra. A consequence of the "successful" treatment of tumors in the bladder is an increase in the frequency of extravesical recurrences (e.g., urethra or ureter). Those with persistent disease in the bladder or new tumors are generally considered for a second course of BCG or for intravesical chemotherapy with valrubicin or gemcitabine. In some cases, cystectomy is recommended, although the specific indications vary. Tumors in the ureter or renal pelvis are typically managed by resection during retrograde examination or, in some cases, by instillation through the renal pelvis. Tumors of the prostatic urethra may require cystectomy if the tumor cannot be resected completely.

INVASIVE DISEASE

The treatment of a tumor that has invaded muscle can be separated into control of the primary tumor and, depending on the pathologic findings at surgery, systemic chemotherapy to treat micrometastatic disease. Radical cystectomy is the standard, although in selected cases a bladder-sparing approach is used; this approach includes complete endoscopic resection; partial cystectomy; or a combination of resection, systemic chemotherapy, and external beam radiation therapy. In some countries, external beam radiation therapy is considered standard. In the United States, its role is limited to those patients deemed unfit for cystectomy, those with unresectable local disease, or as part of an experimental bladder-sparing approach.

Indications for cystectomy include muscle-invading tumors not suitable for segmental resection; low-stage tumors unsuitable for conservative management (e.g., due to multicentric and frequent recurrences resistant to intravesical instillations); high-grade tumors (T1G3) associated with CIS; and bladder symptoms, such as frequency or hemorrhage, that impair quality of life.

Radical cystectomy is major surgery that requires appropriate preoperative evaluation and management. The procedure involves removal of the bladder and pelvic lymph nodes and creation of a conduit or reservoir for urinary flow. Grossly abnormal lymph nodes are evaluated by frozen section. If metastases are confirmed, the procedure is often aborted. In males, radical cystectomy includes the removal of the prostate, seminal vesicles, and proximal urethra. Impotence is universal unless the nerves responsible for erectile function are preserved. In females, the procedure includes removal of the bladder, urethra, uterus, fallopian tubes, ovaries, anterior vaginal wall, and surrounding fascia.

Previously, urine flow was managed by directing the ureters to the abdominal wall, where it was collected in an external appliance. Currently, most patients receive either a continent cutaneous reservoir constructed from detubularized bowel or an orthotopic neobladder. Some 70% of men receive a neobladder. With a continent reservoir, 65–85% of men will be continent at night and 85–90% during the day. Cutaneous reservoirs are drained by intermittent catheterization; orthotopic neobladders are drained more naturally. Contraindications to a neobladder include renal insufficiency, an inability to self-catheterize, or an exophytic tumor or CIS in the urethra. Diffuse CIS in the bladder is a relative contraindication based on the risk of a urethral recurrence. Concurrent ulcerative colitis or Crohn's disease may hinder the use of resected bowel.

A partial cystectomy may be considered when the disease is limited to the dome of the bladder, a margin of at least 2 cm can be achieved, there is no CIS in other sites, and the bladder capacity is adequate after the tumor has been removed. This occurs in 5–10% of cases. Carcinomas in the ureter or in the renal pelvis are treated with nephroureterectomy with a bladder cuff to remove the tumor.

The probability of recurrence following surgery is predicted on the basis of pathologic stage, presence, or absence of lymphatic or vascular invasion, and nodal spread. Among those whose cancers recur, the recurrence develops in a median of 1 year (range 0.04–11.1 years). Long-term outcomes vary by pathologic stage and histology **(Table 94–1)**. The number of lymph nodes removed is also prognostic, whether or not the nodes contained tumor.

Table 94–1 Survival Following Surgery for Bladder Cancer

Pathologic Stage	5-Year Survival, %	10-Year Survival, %
T2,N0	89	87
T3a,N0	78	76
T3b,N0	62	61
T4,N0	50	45
Any T,N1	35	34

Chemotherapy (described below) has been shown to prolong the survival of patients with invasive disease, but only when combined with definitive treatment of the bladder by radical cystectomy or radiation therapy. Thus, for the majority of patients, chemotherapy alone is inadequate to clear the bladder of disease. Experimental studies are evaluating bladder preservation strategies by combining chemotherapy and radiation therapy in patients whose tumors were endoscopically removed.

METASTATIC DISEASE

The primary goal of treatment for metastatic disease is to achieve complete remission with chemotherapy alone or with a combined-modality approach of chemotherapy followed by surgical resection of residual disease, as is done routinely for the treatment of germ cell tumors. One can define a goal in terms of cure or palliation on the basis of the probability of achieving a complete response to chemotherapy using prognostic factors, such as Karnofsky Performance Status (KPS) (<80%), and whether the pattern of spread is nodal or visceral (liver, lung, or bone). For those with zero, one, or two risk factors, the probability of complete remission is 38, 25, and 5%, respectively, and median survival is 33, 13.4, and 9.3 months, respectively. Patients who are functionally compromised or who have visceral disease or bone metastases rarely achieve long-term survival. The toxicities also vary as a function of risk, and treatment-related mortality rates are as high as 3–4% using some combinations in these poor-risk patient groups.

CHEMOTHERAPY

A number of chemotherapeutic drugs have shown activity as single agents; cisplatin, paclitaxel, and gemcitabine are considered most active. Standard therapy consists of two-, three-, or four-drug combinations. Overall response rates of >50% have been reported using combinations such as methotrexate, vinblastine, doxorubicin, and cisplatin (M-VAC); cisplatin and paclitaxel (PT); gemcitabine and cisplatin (GC); or gemcitabine, paclitaxel, and cisplatin (GTC). M-VAC was considered standard, but the toxicities of neutropenia and fever, mucositis, diminished renal and auditory function, and peripheral neuropathy led to the development of alternative regimens. At present, GC is used more commonly than M-VAC, based on the results of a comparative trial of M-VAC versus GC that showed less neutropenia and fever, and less mucositis for the GC regimen. Anemia and thrombocytopenia were more common with GC. GTC is not more effective than GC.

Chemotherapy has also been evaluated in the neoadjuvant and adjuvant settings. In a randomized trial, patients receiving three cycles of neoadjuvant M-VAC followed by cystectomy had a significantly better median (6.2 years) and 5-year survival (57%) compared to cystectomy alone (median survival 3.8 years; 5-year survival 42%). Similar results were obtained in an international study of three cycles of cisplatin, methotrexate, and vinblastine (CMV) followed by either radical cystectomy or radiation therapy. The decision to administer adjuvant therapy is based on the risk of recurrence after cystectomy. Indications for adjuvant chemotherapy include the presence of nodal disease, extravesical tumor extension, or vascular invasion in the resected specimen. Another study of adjuvant therapy found that four cycles of CMV delayed recurrence, although an effect on survival was less clear. Additional trials are studying taxane- and gemcitabine-based combinations.

The management of bladder cancer is summarized in **Table 94–2**.

Table 94–2 Management of Bladder Cancer

Nature of Lesion	Management Approach
Superficial	Endoscopic removal, usually with intravesical therapy
Invasive disease	Cystectomy ± systemic chemotherapy (before or after surgery)
Metastatic disease	Curative or palliative chemotherapy (based on prognostic factors) ± surgery

Table 94–3 Classification of Epithelial Neoplasms Arising from the Kidney

Carcinoma Type	Growth Pattern	Cell of Origin	Cytogenetics
Clear cell	Acinar or sarcomatoid	Proximal tubule	3p–
Papillary	Papillary or sarcomatoid	Proximal tubule	+7, +17, –Y
Chromophobic	Solid, tubular, or sarcomatoid	Cortical collecting duct	Hypodiploid
Oncocytic	Tumor nests	Cortical collecting duct	Undetermined
Collecting duct	Papillary or sarcomatoid	Medullary collecting duct	Undetermined

CARCINOMA OF THE RENAL PELVIS AND URETER

About 2500 cases of renal pelvis and ureter cancer occur each year; nearly all are transitional cell carcinomas similar to bladder cancer in biology and appearance. This tumor is also associated with chronic phenacetin abuse and with Balkan nephropathy, a chronic interstitial nephritis endemic in Bulgaria, Greece, Bosnia-Herzegovina, and Romania.

The most common symptom is painless gross hematuria, and the disease is usually detected on intravenous pyelogram during the workup for hematuria. Patterns of spread are like those in bladder cancer. For low-grade disease localized to the renal pelvis and ureter, nephroureterectomy (including excision of the distal ureter with a portion of the bladder) is associated with 5-year survival of 80–90%. More invasive or histologically poorly differentiated tumors are more likely to recur locally and to metastasize. Metastatic disease is treated with the chemotherapy used in bladder cancer, and the outcome is similar to that of metastatic transitional-cell cancer of bladder origin.

RENAL CELL CARCINOMA

Renal cell carcinomas account for 90–95% of malignantneoplasms arising from the kidney. Notable features include resistance to cytotoxic agents, infrequent responses to biologic response modifiers such as interleukin (IL) 2, robust activity to antiangiogenesis targeted agents, and a variable clinical course for patients with metastatic disease, including anecdotal reports of spontaneous regression.

Epidemiology

The incidence of renal cell carcinoma continues to rise and is now nearly 58,000 cases annually in the United States, resulting in 13,000 deaths. The male to female ratio is 2:1. Incidence peaks between the ages of 50 and 70 years, although this malignancy may be diagnosed at any age. Many environmental factors have been investigated as possible contributing causes; the strongest association is with cigarette smoking. Risk is also increased for patients who have acquired cystic disease of the kidney associated with end-stage renal disease, and for those with tuberous sclerosis. Most cases are sporadic, although familial forms have been reported. One is associated with von Hippel-Lindau (VHL) syndrome. VHL syndrome is an autosomal dominant disorder. Genetic studies identified the *VHL* gene on the short arm of chromosome 3. Approximately 35% of individuals with VHL disease develop clear cell renal cell carcinoma. Other associated neoplasms include retinal hemangioma, hemangioblastoma of the spinal cord and cerebellum, pheochromocytoma, neuroendocrine tumors and cysts, and cysts in the epididymis of the testis in men and the broad ligament in women. Subtypes vary according to low risk (type 1) or high risk (type 2) of developing pheochromocytoma.

Pathology and Genetics

Renal cell neoplasia represents a heterogeneous group of tumors with distinct histopathologic, genetic, and clinical features ranging from benign to high-grade malignant **(Table 94–3)**. They are classified on the basis of morphology and histology. Categories include clear cell carcinoma (60% of cases), papillary tumors (5–15%), chromophobic tumors (5–10%), oncocytomas (5–10%), and collecting or Bellini duct tumors (<1%). Papillary tumors tend to be bilateral and multifocal. Chromophobic tumors have a more indolent clinical course, and oncocytomas are considered benign neoplasms. In contrast, Bellini duct carcinomas, which are thought to arise from the collecting ducts within the renal medulla, are very rare but very aggressive. Clear cell tumors, the predominant histology, are found in >80% of patients who develop metastases. Clear cell tumors arise from the epithelial cells of the proximal tubules and usually show chromosome 3p deletions. Deletions of 3p21–26 (where the *VHL* gene maps) are identified in patients with familial as well as sporadic tumors. *VHL* encodes a tumor-suppressor protein that is involved in regulating the transcription of vascular endothelial growth factor (VEGF), platelet-derived growth factor (PDGF), and a number of other hypoxia-inducible proteins. Inactivation of *VHL* leads to overexpression of these agonists of the VEGF and PDGF receptors, which promote tumor angiogenesis and tumor growth. Agents that inhibit proangiogenic growth factor activity show antitumor effects.

Clinical Presentation

The presenting signs and symptoms include hematuria, abdominal pain, and a flank or abdominal mass. This classic triad occurs in 10–20% of patients. Other symptoms are fever, weight loss, anemia, and a varicocele. The tumor is most commonly detected as an incidental finding on a radiograph. Widespread use of radiologic cross-sectional imaging procedures (CT, ultrasound, MRI) contributes to earlier detection, including incidental renal masses detected during evaluation for other medical conditions. The increasing number of incidentally discovered low-stage tumors has contributed to an improved 5-year survival for patients with renal cell carcinoma and increased use of nephron-sparing surgery (partial nephrectomy). A spectrum of paraneoplastic syndromes has been associated with these malignancies, including erythrocytosis, hypercalcemia, nonmetastatic hepatic dysfunction (Stauffer syndrome), and acquired dysfibrinogenemia. Erythrocytosis is noted at presentation in only about 3% of patients. Anemia, a sign of advanced disease, is more common.

The standard evaluation of patients with suspected renal cell tumors includes a CT scan of the abdomen and pelvis, chest radiograph, urine analysis, and urine cytology. If metastatic disease is suspected from the chest radiograph, a CT of the chest is warranted. MRI is useful in evaluating the inferior vena cava in cases of suspected tumor involvement or invasion by thrombus. In clinical practice, any solid renal masses should be considered malignant until proven otherwise; a definitive diagnosis is required. If no metastases are demonstrated, surgery is indicated, even if the renal vein is invaded. The differential diagnosis of a renal mass includes cysts, benign neoplasms (adenoma, angiomyolipoma, oncocytoma), inflammatory lesions (pyelonephritis or abscesses), and other primary or metastatic cancers. Other malignancies that may involve the kidney include transitional cell carcinoma of the renal pelvis, sarcoma, lymphoma, and Wilms' tumor. All of these are less common causes of renal masses than is renal cell cancer.

Staging and Prognosis

Staging is based on the American Joint Committee on Cancer (AJCC) staging system **(Fig. 94-2)**. Stage I tumors are <7 cm in greatest diameter and confined to the kidney, stage II tumors are ≥7 cm and confined to the kidney, stage III tumors extend through the renal capsule but are confined to Gerota's fascia (IIIa) or involve a single hilar lymph node (N1), and stage IV disease includes tumors that have invaded adjacent organs (excluding the adrenal gland) or involve multiple lymph nodes or distant metastases. The rate of 5-year survival varies by stage: >90% for stage I, 85% for stage II, 60% for stage III, and 10% for stage IV.

Figure 94-2

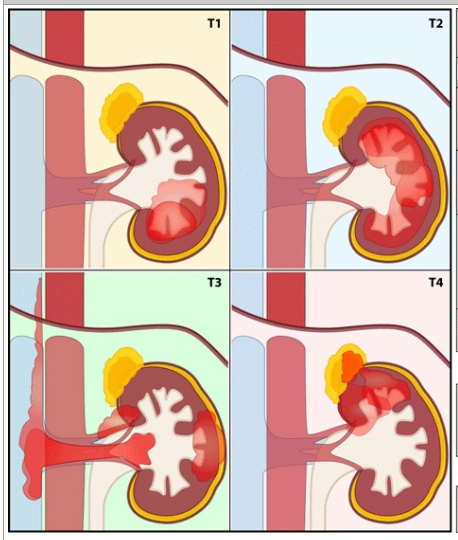

TNM	Involvement	Extent of Disease
TX	Primary not involved	
T1	≤ 7 cm	limited to kidney
T1a	≤ 4 cm	
T1b	≥ 4 cm	
T2	> 7 cm	limited to kidney
T2a	> 7 cm to ≤ 10 cm	
T2b	>10 cm	
T3	into major veins or perinephric tissues	not beyond Gerota's fascia
T3a	in renal vein or renal sinus fat	not beyond Gerota's fascia
T3b	into vena cava	below diaphragm
T3c	into vena cava	above diaphragm
T4	invasion beyond Gerota's Fascia	including contiguous extentions & into ipsilateral adrenal gland

Regional

NX	Regional lymph not assessed
N0	No lymph involvment
N1	Regional lymph involvment

Distant Metastases

M0	No distant metastases
M1	Distant metastases

Anatomic Stage / Prognostic Groups

I	T1	N0	M0
II	T2	N0	M0
III	T1 or T2,	N1	M0
	T3	N0 or N1	M0
IV	T4	Any N	M0
	Any T	Any N	M1

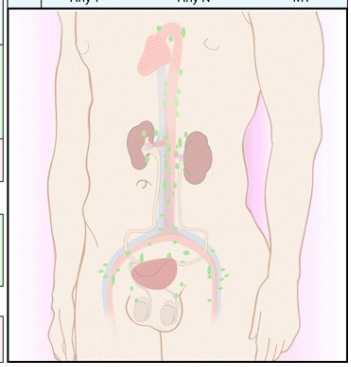

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Renal cell carcinoma staging. TNM, tumor, node, metastasis

Treatment: Renal Cell Carcinoma

LOCALIZED TUMORS

The standard management for stage I or II tumors and selected cases of stage III disease is radical nephrectomy. This procedure involves en bloc removal of Gerota's fascia and its contents, including the kidney, the ipsilateral adrenal gland, and adjacent hilar lymph nodes. The role of a regional lymphadenectomy is controversial. Extension into the renal vein or inferior vena cava (stage III disease) does not preclude resection even if cardiopulmonary bypass is required. If the tumor is resected, half of these patients have prolonged survival.

Nephron-sparing approaches via open or laparoscopic surgery may be appropriate for patients who have only one kidney, depending on the size and location of the lesion. A nephron-sparing approach can also be used for patients with bilateral tumors, accompanied by a radical nephrectomy on the opposite side. Partial nephrectomy techniques are applied electively to resect small masses for patients with a normal contralateral kidney. Adjuvant therapy following this surgery does not improve outcome, even in cases with a poor prognosis.

ADVANCED DISEASE

Surgery has a limited role for patients with metastatic disease. However, long-term survival may occur in patients who relapse after nephrectomy in a solitary site that can be removed. One indication for nephrectomy with metastases at initial presentation is to alleviate pain or hemorrhage of a primary tumor. Also, a cytoreductive nephrectomy before systemic treatment improves survival for carefully selected patients with stage IV tumors.

Metastatic renal cell carcinoma is highly refractory to chemotherapy. Cytokine therapy with IL-2 or IFN-α produces regressions in 10–20% of patients. IL-2 produces durable complete remission in a small proportion of cases. In general, cytokine therapy is considered unsatisfactory for most patients.

The situation changed dramatically when two large-scale randomized trials established a role for antiangiogenic therapy in this disease, as predicted by the genetic studies. These trials separately evaluated two orally administered antiangiogenic agents, sorafenib and sunitinib, that inhibited receptor tyrosine kinase signaling through the VEGF and PDGF receptors. Both showed efficacy as second-line treatment following progression during cytokine treatment, resulting in approval by regulatory authorities for the treatment of advanced renal cell carcinoma. A randomized phase 3 trial comparing sunitinib to IFN-α showed superior efficacy for sunitinib with an acceptable safety profile. The trial resulted in a change in the standard first-line treatment from IFN to sunitinib. Sunitinib is usually given orally at a dose of 50 mg/d for 4 weeks out of 6. Diarrhea is the main toxicity. Sorafenib is usually given orally at a dose of 400 mg bid. In addition to diarrhea, toxicities include rash, fatigue, and hand-foot syndrome. Temsirolimus and everolimus, inhibitors of the mammalian target of rapamycin (mTOR), show activity in patients with untreated poor-prognosis tumors and in sunitinib/sorafenib refractory tumors.

The prognosis of metastatic renal cell carcinoma is variable. In one analysis, no prior nephrectomy, a KPS <80, low hemoglobin, high corrected calcium, and abnormal lactate dehydrogenase were poor prognostic factors. Patients with zero, one or two, and three or more factors had a median survival of 24, 12, and 5 months, respectively. These tumors may follow an unpredictable and protracted clinical course. It may be best to document progression before considering systemic treatment.

FURTHER READINGS

Bladder Cancer: NCCN Clinical Practice Guidelines in Oncology, version 1.2011 www.nccn.org

Cheng L et al: Bladder cancer: Translating molecular genetic insights into clinical practice. Hum Pathol 42:455, 2011[PMID: 21106220]

Hutson TE et al: Targeted therapies for metastatic renal cell carcinoma: An overview of toxicity and dosing strategies. Oncologist 13:1084, 2008[PMID: 18838439]

Iyer G et al: Novel strategies for treating relapsed/refractory urothelial carcinoma. Expert Rev Anticancer Ther 10:1917, 2010[PMID: 21110758]

Linehan WM et al: Molecular diagnosis and therapy of kidney cancer. Annu Rev Med 61:329, 2010[PMID: 20059341]

Motzer RJ et al: Efficacy of everolimus in advanced renal cell carcinoma: A double-blind, randomised, placebo-controlled phase III trial. Lancet 372:449, 2008[PMID: 18653228]

Mulders P: Vascular endothelial growth factor and mTOR pathways in renal cell carcinoma: Differences and synergies of two targeted mechanisms. BJU Int 104:158, 2009

Pal SK et al: Breaking through a plateau in renal cell carcinoma therapeutics: Development and incorporation of biomarkers. Mol Cancer Ther 9:3115, 2010[PMID: 21078774]

RouprÊt M et al: European guidelines for the diagnosis and management of upper urinary tract urothelial cell carcinomas: 2011 update. Eur Urol 59:584, 2011

Solsona E et al: Feasibility of radical transurethral resection as monotherapy for selected patients with muscle invasive bladder cancer. J Urol 184:475, 2010[PMID: 20620402]

Zlotta AR et al: BCAN Think Tank session 1: Overview of risks for and causes of bladder cancer. Urol Oncol 28:329, 2010[PMID: 20439032]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ANATOMY AND PATHOLOGY

The prostate is located in the pelvis and is surrounded by the rectum, the bladder, the periprostatic and dorsal vein complexes and neurovascular bundles that are responsible for erectile function, and the urinary sphincter that is responsible for passive urinary control. The prostate is composed of branching tubuloalveolar glands arranged in lobules surrounded by fibromuscular stroma. The acinar unit includes an epithelial compartment made up of epithelial, basal, and neuroendocrine cells and separated by a basement membrane, a stromal compartment that includes fibroblasts and smooth-muscle cells. Prostate-specific antigen (PSA) and prostatic acid phosphatase (PAP) are produced in the epithelial cells. Both prostate epithelial cells and stromal cells express androgen receptors (ARs) and depend on androgens for growth. Testosterone, the major circulating androgen, is converted by the enzyme5α-reductase to dihydrotestosterone in the gland.

The periurethral portion of the gland increases in size during puberty and after the age of 55 years due to the growth of nonmalignant cells in the transition zone of the prostate that surrounds the urethra. Most cancers develop in the peripheral zone, and cancers in this location can often be palpated during a digital rectal examination (DRE).

PROSTATE CANCER

In 2010 approximately 217,730 prostate cancer cases were diagnosed, and 32,050 men died from prostate cancer in the United States. The absolute number of prostate cancer deaths has decreased in the past 5 years, which has been attributed by some to the widespread use of PSA-based detection strategies. However, the benefit of screening on survival is unclear. The paradox of management is that although 1 in 6 men will eventually be diagnosed with the disease, and the disease remains the second leading cause of cancer deaths in men, only 1 man in 30 with prostate cancer will die of his disease.

Epidemiology

Epidemiologic studies show that the risk of being diagnosed with prostate cancer increases by a factor of two if one first-degree relative is affected and by four if two or more are affected. Current estimates are that 40% of early-onset and 5–10% of all prostate cancers are hereditary. Prostate cancer affects ethnic groups differently. Matched for age, African-American males compared to white males have both a greater number of typically multifocal and highly unstable prostatic intraepithelial neoplasia (PIN) lesions, which are precursors to cancer, and larger tumors, possibly related to the higher levels of testosterone seen in African American males. Polymorphic variants of the AR, the cytochrome P450 C17, and the steroid 5α-reductase type II (SRD5A2) genes have also been implicated in the variations in incidence.

The prevalence of autopsy-detected cancers is similar around the world, while the incidence of clinical disease varies. Thus, environmental factors may play a role. High consumption of dietary fats, such as α-linoleic acid, or the polycyclic aromatic hydrocarbons that form when red meats are cooked is believed to increase risk. Similar to breast cancer in Asian women, the risk of prostate cancer in Asian men increases when they move to Western environments. Protective factors include consumption of the isoflavonoid genistein (which inhibits 5α-reductase) found in many legumes, cruciferous vegetables that contain the isothiocyanate sulforaphane, retinoids such as lycopene found in tomatoes, and inhibitors of cholesterol biosynthesis (e.g., statin drugs). The development of prostate cancer is a multistep process. One early change is hypermethylation of the GSTP1 gene promoter, which leads to loss of function of a gene that detoxifies carcinogens. The finding that many prostate cancers develop adjacent to a lesion termed PIA (proliferative inflammatory atrophy) suggests a role for inflammation.

Diagnosis and Treatment by Clinical State

The prostate cancer continuum—from the appearance of a preneoplastic and invasive lesion localized to the prostate, to a metastatic lesion that results in symptoms and, ultimately, mortality—can span decades. To facilitate disease management, competing risks are considered in the context of a series of clinical states (Fig. 95-1). The states are defined operationally on the basis of whether or not a cancer diagnosis has been established and, for those with a diagnosis, whether or not metastases are detectable on imaging studies and the measured level of testosterone in the blood. With this approach, an individual resides in only one state and remains in that state until he has progressed. At each assessment, the decision to offer treatment and the specific form of treatment is based on the risk posed by the cancer relative to competing causes of mortality that may be present in that individual. It

follows that the more advanced the disease, the greater the need for treatment.

For those without a cancer diagnosis, the decision to undergo testing to detect a cancer is based on the individual's estimated life expectancy and, separately, the probability that a clinically significant cancer may be present. For those with a prostate cancer diagnosis, the states model considers the probability of developing symptoms or dying from disease. Thus, a patient with localized prostate cancer who has had all cancer removed surgically remains in the state of localized disease as long as the PSA remains undetectable. The time within a state becomes a measure of the efficacy of an intervention, though the effect may not be assessable for years. As many men with active cancer are not at risk for developing metastases, symptoms, or death, the states model allows a distinction between *cure*—the elimination of all cancer cells, the primary therapeutic objective when treating most cancers—and *cancer control* , in which the tempo of the illness is altered and symptoms controlled until the patient dies of other causes. These can be equivalent therapeutically from a patient standpoint if the patient has not experienced symptoms of the disease or the treatment needed to control it. Even when a recurrence is documented, immediate therapy is not always necessary. Rather, as at the time of diagnosis, the need for intervention is based on the tempo of the illness as it unfolds in the individual, relative to the risk-to-benefit ratio of the therapy being considered.

No Cancer Diagnosis

PREVENTION

The results from several large double-blind, randomized chemoprevention trials have established 5α-reductase inhibitors (5ARI) as the predominant therapy to reduce the future risk of a prostate cancer diagnosis. The Prostate Cancer Prevention Trial (PCPT), in which men older than age 55 years received the 5α-reductase inhibitor finasteride, which inhibits the type 1 isoform, or a placebo, showed a 25% (95% confidence interval 19–31%) reduction in the period prevalence of prostate cancer across all age groups in favor of finasteride (18.4%) over placebo (24.4%). In REDUCE (Reduction by Dutasteride of Prostate Cancer Events Trial), a similar 23% reduction in the 4-year period prevalence was observed in favor of dutasteride ($p = 0.001$). Dutasteride inhibits both the type 1 and type 2 5ARI isoforms. These results contrast with those of the Selenium and Vitamin E Cancer Prevention Trial (SELECT) in which African American men aged ≥50 years and others aged ≥55 years were enrolled, which showed no difference in cancer incidence in patients receiving vitamin E (4.6%) or selenium (4.9%) alone or in combination (4.6%) relative to placebo (4.4%). A similar lack of benefit for vitamin E, vitamin C, and selenium was seen in the Physicians Health Study II.

PHYSICAL EXAMINATION

The need to pursue a diagnosis of prostate cancer is based on symptoms, an abnormal DRE, or, more typically, a change in or an elevated serum PSA. The urologic history should focus on symptoms of outlet obstruction, continence, potency, or change in ejaculatory pattern.

The DRE focuses on prostate size and consistency and abnormalities within or beyond the gland. Many cancers occur in the peripheral zone and can be palpated on DRE. Carcinomas are characteristically hard, nodular, and irregular, while induration may also be due to benign prostatic hypertrophy (BPH) or calculi. Overall, 20–25% of men with an abnormal DRE have cancer.

PROSTATE-SPECIFIC ANTIGEN

PSA (kallikrein-related peptidase 3; KLK3) is a kallikrein-related serine protease that causes liquefaction of seminal coagulum. It is produced by both nonmalignant and malignantepithelial cells and, as such, is prostate-specific, not prostate cancer–specific, and serum levels may also increase from prostatitis and BPH. Serum levels are not affected by DRE but the performance of a prostate biopsy can increase PSA levels up to tenfold for 8–10 weeks. PSA circulating in the blood is inactive and mainly occurs as a complex with the protease inhibitor α1-antichymotrypsin SERPIN A3 and as free unbound PSA forms. The formation of complexes between PSA, α2-macroglobulin, or other protease inhibitors is less significant. Free PSA is rapidly eliminated from the blood by glomerular filtration with an estimated half-life of 12-18 hours. Elimination of PSA bound to α1-antichymotrypsin is slow (estimated half-life of 1-2 weeks) as it is too large to be cleared by the kidneys. Levels should be undetectable after about six weeks if the prostate has been removed. Immunohistochemical staining for PSA can be used to establish a prostate cancer diagnosis.

PSA testing was approved by the U.S. FDA in 1994 for early detection of prostate cancer, and the widespread use of the test has played a significant role in the proportion of men diagnosed with early-stage cancers: more than 70–80% of newly diagnosed cancers are organ confined. The level of PSA in blood is strongly associated with the risk and outcome of prostate cancer. A single PSA measured at age 60 is associated (AUC of 0.90) with lifetime risk of death from prostate cancer. Most (90%) prostate cancer deaths occur among men with PSA levels in top quartile (>2 ng/mL), although only a minority of men with PSA >2 ng/mL will develop lethal prostate cancer. Despite this and mortality rate reductions reported from large randomized prostate cancer screening trials, routine use of the test remains controversial. The American Cancer Society (ACS) recommends that physicians offer PSA testing and a DRE on an annual basis for men older than age 50 years with an anticipated survival of >10 years; this includes men up to age 76 years. For African Americans and men with a family history of prostate

cancer, testing is advised to begin at age 45 years. The American Urologic Association recommendations are similar, with the proviso that the risks and benefits of the performance of these tests are not defined. The American College of Physicians recommends that physicians "describe the potential benefits and known harms of screening" and to "individualize the decision to screen." The National Comprehensive Cancer Network (NCCN) guidelines mirror those of the ACS, with the proviso that "physicians and potential participants must thoroughly discuss the pros and cons of screening." The NCCN also advises that men who opt to participate obtain a baseline PSA and DRE in their values and use the value to stratify future risk. As PSA values may fluctuate for no apparent reason, it is advised that isolated abnormal values should be confirmed before proceeding with further testing.

The PSA criteria used to recommend a diagnostic prostate biopsy have evolved over time. However, based on the commonly used outpoint for prostate biopsy CPSA≥4 mg/mL, most men with a PSA elevation do not have histologic widence of prostate cancer at biopsy, and commenly, many men with PSA levels below this cut point harbor cancer cells in their prostate. The goal is to increase the sensitivity of the test for younger men more likely to die of the disease and to reduce the frequency of detecting cancers of low malignant potential in elderly men more likely to die of other causes. Previously, the threshold for performance of a biopsy was 4.0 ng/mL, which has been reduced to 3 mg/mL or 2.6 ng/mL for men aged <60 years by many groups based on the finding that nearly half of the men with PSAs who reached this level increased to 4 ng/mL within a relatively short (4-year) time frame and that, once diagnosed, in nearly one-third it had spread beyond the confines of the gland.

Most PSA is complexed to α_1-antichymotrypsin (ACT); only a small percentage is "free," and lower in men with cancer. Free and complexed PSA measurements are used when levels are between 4 and 10 ng/mL to decide whether a biopsy is needed. The risk of cancer is under 10% if the free PSA is >25% but as high as 56% for those with a free PSA <10%. PSA density (PSAD) measurements were developed to correct for the contribution of BPH to the total PSA level. PSAD is calculated by dividing the serum PSA by the prostate weight estimated from transrectal ultrasound (TRUS). Values <0.10 ng/mL per cm^3 are consistent with BPH, while those >0.15 ng/mL per cm^3 suggest cancer. *PSA dynamics* is the rate of change in PSA levels over time and is expressed most commonly as the PSA velocity or PSA doubling time. It is particularly useful for men with seemingly normal values that are rising. For men with a PSA level higher than 4 ng/mL, rates of rise >0.75 ng/mL per year suggest cancer, while for those with lower PSA levels, rates >0.5 ng/mL per year should be used to advise a biopsy. As an example, an increase from 2.5 to 3.2 ng/mL in a 1-year period would warrant further testing.

PSA-based detection strategies have changed the clinical spectrum of the disease. Now, 95–99% of newly diagnosed cancers are clinically localized, 40% are not palpable, and of these, 70% are pathologically organ-confined. However, the benefits of PSA screening remain controversial due to the overdetection of cancers with low malignant potential that may lead to overtreatment and unnecessary morbidity. To this end, the U.S. Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening trial found no mortality benefit from combined PSA screening and DRE in 76,693 randomized men (annual exam vs. standard care) with a median follow-up of 11 years. However, important caveats about the PLCO study include (1) many screening participants had already undergone PSA screening before the trial; (2) contamination from PSA testing among controls increased from 40% in year one to 52% in year six and; (3) and the biopsycompliance was low. These factors make interpretation difficult. A subgroup analysis of this trial showed a reduction in cancer mortality among screened men with little or no comorbidity. The European Randomized Study of Screening for Prostate Cancer (ERSPC) trial followed 182,000 men a median of 9 years randomized either to PSA screening every 4 years or to a group not receiving regular PSA screening. In this study, PSA screening without DRE corresponded to a 20% relative reduction of the rate of death from prostate cancer. A report from the Swedish subgroup of this study based on 14 years follow-up suggested that PSA screening may reduce cancer-specific mortality by nearly half with less overdiagnosis and treatment than was noted in the European Study as a whole. Men remain advised to make an informed decision on an individual basis about whether to undergo testing.

A diagnostic algorithm based on the DRE and PSA findings is illustrated in **Fig. 95-2**. In general, a biopsy is recommended if the DRE or PSA is abnormal. Twenty-five percent of men with a PSA >4 ng/mL and an abnormal DRE have cancer, as do 17% of men with a PSA of 2.5–4 ng/mL and normal DRE.

Figure 95-2

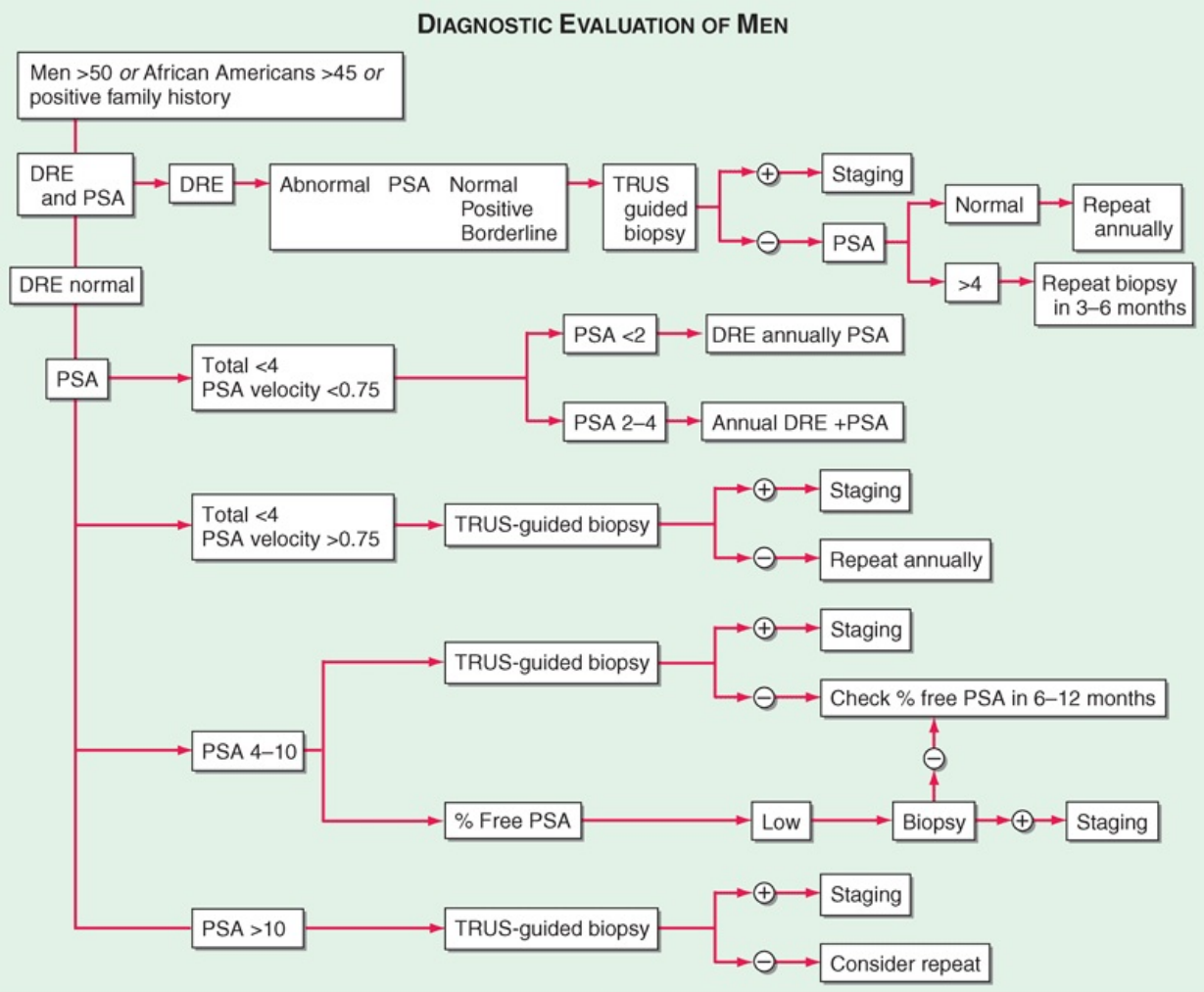

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for diagnostic evaluation of men based on digital rectal examination (DRE) and prostate-specific antigen (PSA) levels; TRUS, transrectal ultrasound.

PROSTATE BIOPSY

A diagnosis of cancer is established by a TRUS-guided needle biopsy. Direct visualization by ultrasound or MRI assures that all areas of the gland are sampled. A minimum of six separate cores, three from the right and three from the left, is advised, as is a separate biopsy of the transition zone if clinically indicated. Contemporary schemas advise an extended-pattern 12- to 14-core biopsy that includes the sextant sampling above plus 6 cores from the lateral peripheral zones as well as a lesion-directed palpable nodule or suspicious image-guided sampling. Patients with prostatitis should have a course of antibiotics before biopsy. Men with an abnormal PSA and negative biopsy are advised to undergo a repeat biopsy.

Each core of the biopsy is examined for the presence of cancer, and the amount of cancer is quantified based on the length of the tumor within the core and the percentage of the core involved.

PATHOLOGY

The noninvasive proliferation of epithelial cells within ducts is termed *prostatic intraepithelial neoplasia*. PIN is a precursor of cancer, but not all PIN lesions develop into invasive cancers. Of the cancers identified, >95% are adenocarcinomas; the rest are squamous or transitional cell tumors or, rarely, carcinosarcomas. Metastases to the prostate are rare, but in some cases colon cancers or transitional cell tumors of the bladder invade the gland by direct extension.

When prostate cancer is diagnosed, a measure of histologic aggressiveness is assigned using the *Gleason grading system*, in which the dominant and secondary glandular histologic patterns are scored from 1 (well-differentiated) to 5 (undifferentiated) and summed to give a total score of 2–10 for each tumor. The most poorly differentiated area of tumor (i.e., the area with the highest histologic grade) often determines biologic behavior. The presence or absence of perineural invasion and extracapsular spread are also recorded.

PROSTATE CANCER STAGING

The TNM staging system includes categories for cancers that are palpable on DRE, those identified solely on the basis of an abnormal PSA (T1c), those that are palpable but clinically confined to the gland (T2), and those that have extended outside the gland (T3 and T4) **(Table 95–1, Fig. 95-3)**. DRE alone is inaccurate in determining the extent of disease within the gland, the presence or absence of capsular invasion, involvement of seminal vesicles, and extension of disease to lymph nodes. Because of the inadequacy of DRE for staging, the TNM staging system was modified to include the results of imaging. Unfortunately, no single test has proven to accurately indicate the stage or the presence of organ-confined disease, seminal vesicle involvement, or lymph node spread.

Table 95-1 TNM Classification

TNM Staging System for Prostate Cancer[a]	
Tx	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Localized Disease	
T1	Clinically inapparent tumor, neither palpable nor visible by imaging
T1a	Tumor incidental histologic finding in ≤5% of resected tissue; not palpable
T1b	Tumor incidental histologic finding in >5% of resected tissue
T1c	Tumor identified by needle biopsy (e.g., because of elevated PSA)
T2	Tumor confined within prostate[b]
T2a	Tumor involves half of one lobe or less
T2b	Tumor involves more than one half of one lobe, not both lobes
T2c	Tumor involves both lobes
Local Extension	
T3	Tumor extends through the prostate capsule[c]
T3a	Extracapsular extension (unilateral or bilateral)
T3b	Tumor invades seminal vesicles(s)
T4	Tumor is fixed or invades adjacent structures other than seminal vesicles such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall.
Metastatic Disease	
N1	Positive regional lymph nodes
M1	Distant metastases

[a]Revised from SB Edge et al (eds): *AJCC Cancer Staging Manual* , 7th ed. New York, Springer, 2010.

[b]Tumor found in one or both lobes by needle biopsy, but not palpable or reliably visible by imaging, is classified as T1c.

[c]Invasion into the prostatic apex or into (but not beyond) the prostatic capsule is classified not as T3 but as T2.

Abbreviation: PSA, prostate-specific antigen.

Figure 95-3

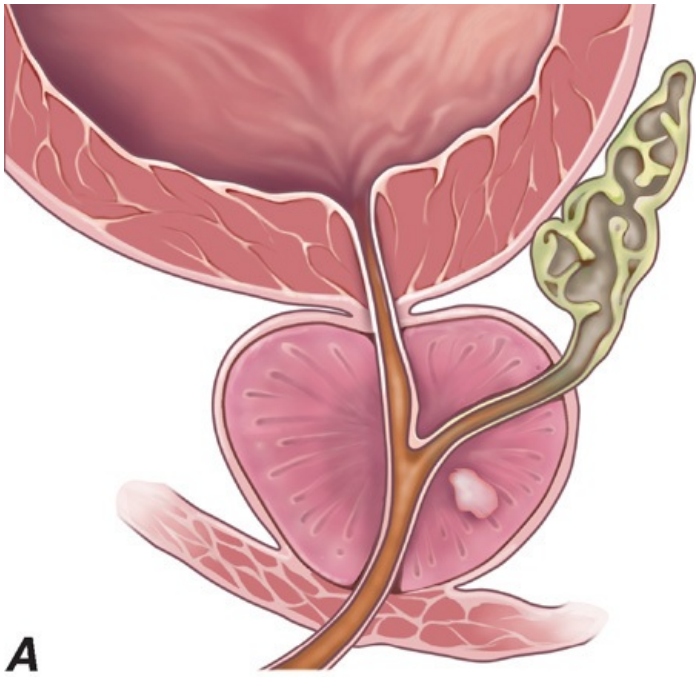

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

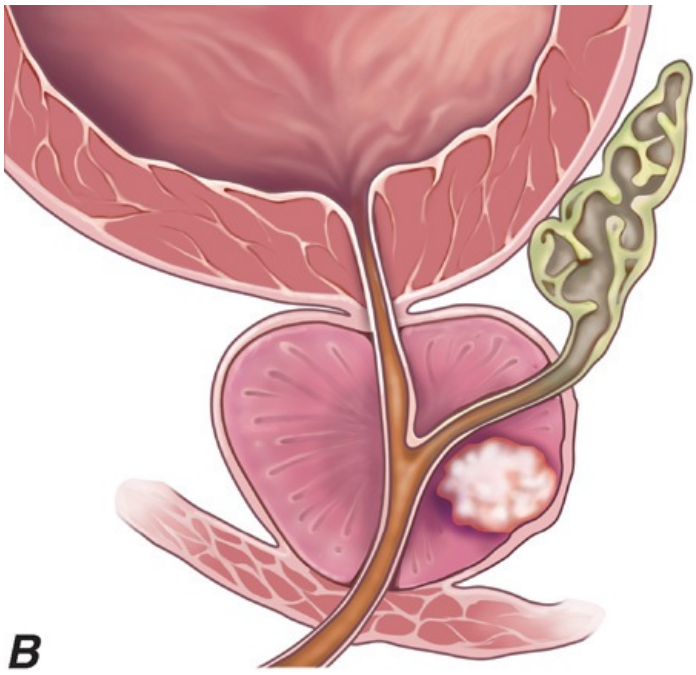

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

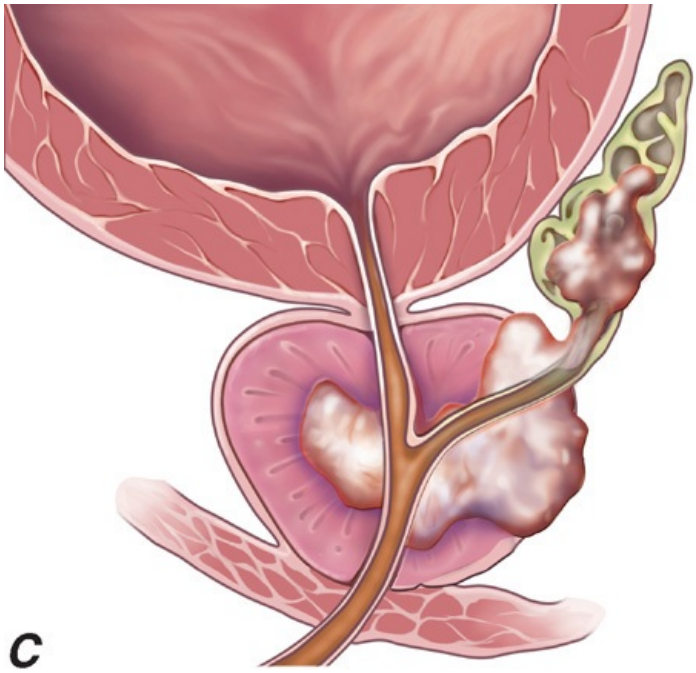

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

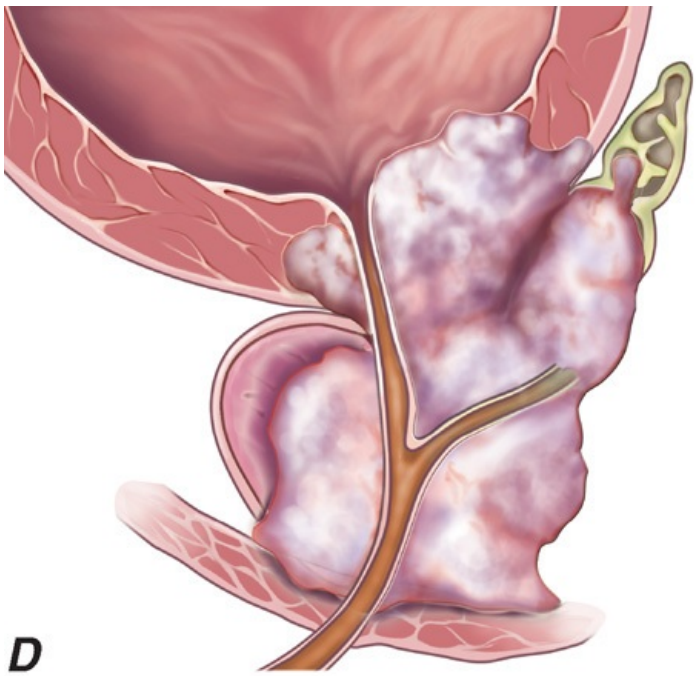

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

T stages of prostate cancer. ***(A)*** T1—Clinically inapparent tumor, neither palpable nor visible by imaging; ***(B)*** T2—Tumor confined within prostate; ***(C)*** T3—Tumor extends through prostate capsule and may invade the seminal vesicles; ***(D)*** T4—Tumor is fixed or invades adjacent structures. Eighty percent of patients present with local disease (T1 and T2), which is associated with a 5-year survival rate of 100%. An additional 12% of patients present with regional disease (T3 and T4 without metastases), which is also associated with a 100% survival rate after 5 years. Four percent of patients present with distant disease (T4 with metastases), which is associated with a 30% 5-year survival rate. (Three percent of patients are ungraded.) *(Data from AJCC, http://seer.cancer.gov/statfacts/html/prost.html. Figure © Memorial Sloan-Kettering Cancer Center Medical Graphics; used with permission.)*

TRUS is the imaging technique most frequently used to assess the primary tumor, but its chief use is directing prostate biopsies, not staging. No TRUS finding consistently indicates cancer with certainty. CT lacks sensitivity and specificity to detect extraprostatic extension and is inferior to MRI in visualization of lymph nodes. In general, MRI performed with an endorectal coil is superior to CT to detect cancer in the prostate and to assess local disease extent. T1-weighted images pr

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TESTICULAR CANCER: INTRODUCTION

Primary germ cell tumors (GCTs) of the testis arising by the malignant transformation of primordial germ cells constitute 95% of all testicular neoplasms. Infrequently, GCTs arise from an extragonadal site, including the mediastinum, retroperitoneum, and, very rarely, the pineal gland. This disease is notable for the young age of the afflicted patients, the totipotent capacity for differentiation of the tumor cells, and its curability; approximately 95% of newly diagnosed patients are cured. Experience in the management of GCTs leads to improved outcome.

INCIDENCE AND EPIDEMIOLOGY

In 2010, 8480 new cases of testicular GCT were diagnosed in the United States and 350 men died. The tumor occurs most frequently in men between the ages of 20 and 40 years. A testicular mass in a male ≥50 years should be regarded as a lymphoma until proved otherwise. GCT is at least four to five times more common in white than in African-American males, and a higher incidence has been observed in Scandinavia and New Zealand than in the United States.

ETIOLOGY AND GENETICS

Cryptorchidism is associated with a several-fold higher risk of GCT. Abdominal cryptorchid testes are at a higher risk than inguinal cryptorchid testes. Orchiopexy should be performed before puberty, if possible. Early orchiopexy reduces the risk of GCT and improves the ability to save the testis. An abdominal cryptorchid testis that cannot be brought into the scrotum should be removed. Approximately 2% of men with GCTs of one testis will develop a primary tumor in the other testis. Testicular feminization syndromes increase the risk of testicular GCT, and Klinefelter's syndrome is associated with mediastinal GCT.

An isochromosome of the short arm of chromosome 12 [i(12p)] is pathognomonic for GCT of all histologic types. Excess 12p copy number, either in the form of i(12p) or as increased 12p on aberrantly banded marker chromosomes, occurs in nearly all GCTs, but the gene(s) on 12p involved in the pathogenesis are not yet defined.

CLINICAL PRESENTATION

A painless testicular mass is pathognomonic for a testicular malignancy. More commonly, patients present with testicular discomfort or swelling suggestive of epididymitis and/or orchitis. In this circumstance, a trial of antibiotics is reasonable. However, if symptoms persist or a residual abnormality remains, then testicular ultrasound examination is indicated.

Ultrasound of the testis is indicated whenever a testicular malignancy is considered and for persistent or painful testicular swelling. If a testicular mass is detected, a radical inguinal orchiectomy should be performed. Because the testis develops from the gonadal ridge, its blood supply and lymphatic drainage originate in the abdomen and descend with the testis into the scrotum. An inguinal approach is taken to avoid breaching anatomic barriers and permitting additional pathways of spread.

Back pain from retroperitoneal metastases is common and must be distinguished from musculoskeletal pain. Dyspnea from pulmonary metastases occurs infrequently. Patients with increased serum levels of human chorionic gonadotropin (hCG) may present with gynecomastia. A delay in diagnosis is associated with a more advanced stage and possibly worse survival.

The staging evaluation for GCT includes a determination of serum levels of α fetoprotein (AFP), hCG, and lactate dehydrogenase (LDH). After orchiectomy, a chest radiograph and a CT scan of the abdomen and pelvis should be performed. A chest CT scan is required if pulmonary nodules or mediastinal or hilar disease is suspected. Stage I disease is limited to the testis, epididymis, or spermatic cord. Stage II disease is limited to retroperitoneal (regional) lymph nodes. Stage III disease is disease outside the retroperitoneum, involving supradiaphragmatic nodal sites or viscera. The staging may be "clinical"—defined solely by physical examination, blood marker evaluation, and radiographs—or "pathologic"—defined by an operative procedure.

The regional draining lymph nodes for the testis are in the retroperitoneum, and the vascular supply originates from the great vessels (for the right testis) or the renal vessels (for the left testis). As a result, the lymph nodes that are involved first by a right testicular tumor are the interaortocaval lymph nodes just below the renal vessels. For a left testicular tumor, the first involved lymph nodes are lateral to the aorta (para-aortic) and below the left renal vessels. In both cases, further nodal spread is inferior, contralateral, and, less commonly, above the renal hilum. Lymphatic involvement can extend cephalad to the retrocrural, posterior mediastinal, and supraclavicular lymph nodes. Treatment is determined by tumor histology (seminoma versus nonseminoma) and clinical stage (Fig. 96–1).

Figure 96-1

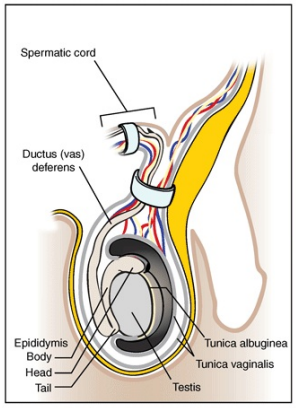

Stage	Extent of Disease
pT1	Tumor limited to the testis and epididymis without vascular/lymphatic invasion; tumor may invade into the tunica albuginea but not the tunica vaginalis
pT2	Tumor limited to the testis and epididymis with vascular/lymphatic invasion; tumor extending through the tunica albuginea with involvement of the tunica vaginalis
pT3	Tumor invades the spermatic cord with or without vascular/lymphatic invasion
pT4	Tumor invades the scrotum with or without vascular/lymphatic invasion

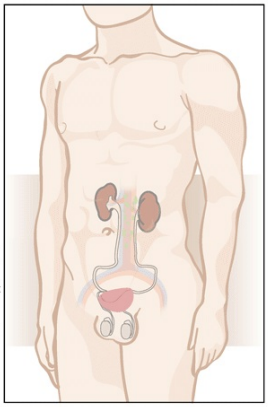

Stage	Extent of Disease	Treatment Option: Seminoma	Treatment Option: Nonseminoma
IA	Testis only, no vascular/lymphatic invasion (T1)	Observation Chemotheapy or RT	RPLND or Observation
IB	Testis only, with vascular/lymphatic invasion (T2), or extension through tunica albuginea (T2), or involvement of spermatic cord (T3) or scrotum (T4)	Observation Chemotheapy or RT	RPLND or Chemotherapy
IIA	Retroperitoneal Nodes < 2 cm	RT	RPLND +/– adjuvant Chemotherapy or Chemotherapy followed by RPLND
IIB	Retroperitoneal Nodes 2-5 cm	RT or Chemotherapy	Chemotherapy, often followed by RPLND
IIC	Retroperitoneal Nodes > 5 cm	Chemotherapy	Chemotherapy, often followed by RPLND
Distant Metastases			
III	Common sites include distant (or extra-abdominal) lymph nodes, lung, liver, bone, and brain	Chemotherapy	Chemotherapy, often followed by surgery (biopsy or resection)

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Germ cell tumor staging and treatment.

PATHOLOGY

GCTs are divided into nonseminoma and seminoma subtypes. Nonseminomatous GCTs are most frequent in the third decade of life and can display the full spectrum of embryonic and adult cellular differentiation. This entity comprises four histologies: embryonal carcinoma, teratoma, choriocarcinoma, and endodermal sinus (yolk sac) tumor. Choriocarcinoma, consisting of both cytotrophoblasts and syncytiotrophoblasts, represents malignant trophoblastic differentiation and is invariably associated with secretion of hCG. Endodermal sinus tumor is the malignant counterpart of the fetal yolk sac and is associated with secretion of AFP. Pure embryonal carcinoma may secrete AFP or hCG, or both; this pattern is biochemical evidence of differentiation. Teratoma is composed of somatic cell types derived from two or more germ layers (ectoderm, mesoderm, or endoderm). Each of these histologies may be present alone or in combination with others. Nonseminomatous GCTs tend to metastasize early to sites such as the retroperitoneal lymph nodes and lung parenchyma. One-third of patients present with disease limited to the testis (stage I), one-third with retroperitoneal metastases (stage II), and one-third with more extensive supradiaphragmatic nodal or visceral metastases (stage III).

Seminoma represents approximately 50% of all GCTs, has a median age in the fourth decade, and generally follows a more indolent clinical course. Most patients (70%) present with stage I disease, approximately 20% with stage II disease, and 10% with stage III disease; lung or other visceral metastases are rare. When a tumor contains both seminoma and nonseminoma components, patient management is directed by the more aggressive nonseminoma component.

TUMOR MARKERS

Careful monitoring of the serum tumor markers AFP and hCG is essential in the management of patients with GCT, as these markers are important for diagnosis, as prognostic indicators, in monitoring treatment response, and in the detection of early relapse. Approximately 70% of patients presenting with disseminated nonseminomatous GCT have increased serum concentrations of AFP and/or hCG. Although hCG concentrations may be increased in patients with either nonseminoma or seminoma histology, the AFP concentration is increased only in patients with nonseminoma. The presence of an increased AFP level in a patient whose tumor shows only seminoma indicates that an occult nonseminomatous component exists, and the patient should be treated for nonseminomatous GCT. LDH levels are not as specific as AFP or hCG but are increased in 50–60% patients with metastatic nonseminoma and in up to 80% of patients with advanced seminoma.

AFP, hCG, and LDH levels should be determined before and after orchiectomy. Increased serum AFP and hCG concentrations decay according to first-order kinetics; the half-life is 24–36 h for hCG and 5–7 days for AFP. AFP and hCG should be assayed serially during and after treatment. The reappearance of hCG and/or AFP or the failure of these markers to decline according to the predicted half-life is an indicator of persistent or recurrent tumor.

Treatment: Testicular Cancer

STAGE I NONSEMINOMA

If, after an orchiectomy (for clinical stage I disease), radiographs and physical examination show no evidence of disease and serum AFP and hCG concentrations are either normal

or declining to normal according to the known half-life, patients may be managed by either a nerve-sparing retroperitoneal lymph node dissection (RPLND) or surveillance. The retroperitoneal lymph nodes are involved by GCT (pathologic stage II) in 20–50% of these patients. The choice of surveillance or RPLND is based on the pathology of the primary tumor. If the primary tumor shows no evidence for lymphatic or vascular invasion and is limited to the testis (T1), then either option is reasonable. If lymphatic or vascular invasion is present or the tumor extends into the tunica, spermatic cord, or scrotum (T2 through T4), then surveillance should not be offered. Either approach should cure >95% of patients.

RPLND is the standard operation for removal of the regional lymph nodes of the testis (retroperitoneal nodes). The operation removes the lymph nodes draining the primary site and the nodal groups adjacent to the primary landing zone. The standard (modified bilateral) RPLND removes all node-bearing tissue down to the bifurcation of the great vessels, including the ipsilateral iliac nodes. The major long-term effect of this operation is retrograde ejaculation and infertility. Nerve-sparing RPLND, usually accomplished by identification and dissection of individual nerve fibers, may avoid injury to the sympathetic nerves responsible for ejaculation. Normal ejaculation is preserved in ~90% of patients. Patients with pathologic stage I disease are observed, and only the <10% who relapse require additional therapy. If retroperitoneal nodes are found to be involved at RPLND, then a decision regarding adjuvant chemotherapy is made on the basis of the extent of retroperitoneal disease (see below).

Surveillance is an option in the management of clinical stage I disease when no vascular/lymphatic invasion is found (T1). Only 20–30% of patients have pathologic stage II disease, implying that most RPLNDs in this situation are not therapeutic. Surveillance and RPLND lead to equivalent long-term survival rates. Patient compliance is essential if surveillance is to be successful. Patients must be carefully followed with periodic chest radiography, physical examination, CT scan of the abdomen, and serum tumor marker determinations. The median time to relapse is approximately 7 months, and late relapses (>2 years) are rare. The 70–80% of patients who do not relapse require no intervention after orchiectomy; treatment is reserved for those who do relapse. When the primary tumor is classified as T2 through T4 (extension beyond testis and epididymis or lymphatic/vascular invasion is identified), nerve-sparing RPLND is preferred. Approximately 50% of these patients have pathologic stage II disease and are destined to relapse without the RPLND.

STAGE II NONSEMINOMA

Patients with limited, ipsilateral retroperitoneal adenopathy (nodes usually ≤3 cm in largest diameter) and normal levels of AFP and hCG generally undergo a modified bilateral RPLND as primary management. Increased levels of either AFP or hCG or both imply metastatic disease outside the retroperitoneum; chemotherapy is used in this setting. The local recurrence rate after a properly performed RPLND is very low. Depending on the extent of disease, the postoperative management options include either surveillance or two cycles of adjuvant chemotherapy. Surveillance is the preferred approach for patients with resected "low-volume" metastases (tumor nodes ≤2 cm in diameter *and* <6 nodes involved) because the probability of relapse is one-third or less. For those who relapse, risk-directed chemotherapy is indicated (see below). Because relapse occurs in ≥50% of patients with "high-volume" metastases (>6 nodes involved, *or* any involved node >2 cm in largest diameter, *or* extranodal tumor extension), two cycles of adjuvant chemotherapy should be considered, as it results in a cure in ≥98% of patients. Regimens consisting of etoposide (100 mg/m^2 daily on days 1–5) plus cisplatin (20 mg/m^2 daily on days 1–5) with or without bleomycin (30 units per day on days 2, 9, and 16) given at 3-week intervals are effective and well tolerated.

STAGES I AND II SEMINOMA

Inguinal orchiectomy followed by retroperitoneal radiation therapy or surveillance cures nearly 100% of patients with stage I seminoma. Historically, radiation was the mainstay of treatment, But the reported association between radiation and secondary malignancies and the absence of a survival advantage of radiation over surveillance has led many to favor surveillance for patients committed to long-term follow-up. Studies have shown that approximately 15% of patients relapse, and rete testis involvement and size >4 cm have been associated with a higher relapse rate. The relapse is usually treated with chemotherapy. Long-term followup is essential, because approximately 30% of relapses occur after 2 years and 5% after 5 years. A single dose of carboplatin has also been investigated as an alternative to radiation therapy; the outcome was similar, but long-term safety data are lacking, and the retroperitoneum remained the most frequent site of relapse.

Nonbulky retroperitoneal disease (stage IIA and most IIB) is treated with retroperitoneal radiation therapy. Approximately 90% of patients achieve relapse-free survival with retroperitoneal masses <5 cm in diameter. Because at least one-third of patients with bulkier disease relapse, initial chemotherapy is preferred for all stage IIC and some stage IIB with bulkier or multifocal disease.

CHEMOTHERAPY FOR ADVANCED GCT

Regardless of histology, patients with stage IIC and stage III GCT are treated with chemotherapy. Combination chemotherapy programs based on cisplatin at doses of 100 mg/m^2 plus etoposide at doses of 500 mg/m^2 per cycle cure 70–80% of such patients, with or without bleomycin, depending on risk stratification (see below). A complete response (the complete disappearance of all clinical evidence of tumor on physical examination and radiography plus normal serum levels of AFP and hCG for ≥1 month) occurs after chemotherapy alone in ~60% of patients, and another 10–20% become disease-free with surgical resection of residual masses containing viable GCT. Lower doses of cisplatin result in inferior survival rates.

The toxicity of four cycles of the bleomycin, etoposide, and cisplatin (BEP) regimen is substantial. Nausea, vomiting, and hair loss occur in most patients, although nausea and vomiting have been markedly ameliorated by modern antiemetic regimens. Myelosuppression is frequent, and symptomatic bleomycin pulmonary toxicity occurs in ~5% of patients. Treatment-induced mortality due to neutropenia with septicemia or bleomycin-induced pulmonary failure occurs in 1–3% of patients. Dose reductions for myelosuppression are rarely indicated. Long-term permanent toxicities include nephrotoxicity (reduced glomerular filtration and persistent magnesium wasting), ototoxicity, and peripheral neuropathy. When bleomycin is administered by weekly bolus injection, Raynaud's phenomenon appears in 5–10% of patients. Other evidence of small blood vessel damage is seen less often, including transient ischemic attacks and myocardial infarction.

RISK-DIRECTED CHEMOTHERAPY

Because not all patients are cured and treatment may cause significant toxicities, patients are stratified into "good-risk" and "poor-risk" groups according to pretreatment clinical features. For good-risk patients, the goal is to achieve maximum efficacy with minimal toxicity. For poor-risk patients, the goal is to identify more effective therapy with tolerable toxicity.

The International Germ Cell Cancer Consensus Group developed criteria to assign patients to three risk groups (good, intermediate, poor) **(Table 96–1)**. The marker cut offs have been incorporated into the revised TNM (primary tumor, regional nodes, metastasis) staging of GCT. Hence, TNM stage groupings are now based on both anatomy (site and extent of disease) and biology (marker status and histology). Seminoma is either good or intermediate risk, based on the absence or presence of nonpulmonary visceral metastases. No poor-risk category exists for seminoma. Marker levels play no role in defining risk for seminoma. Nonseminomas have good-, intermediate-, and poor-risk categories based on the site of the primary tumor, the presence or absence of nonpulmonary visceral metastases, and marker levels.

Table 96–1 IGCCCG Risk Classification for Advanced Germ Cell Tumors

Risk	Nonseminoma	Seminoma
Good	Gonadal or retroperitoneal primary site	Any primary site
	Absent nonpulmonary visceral metastases	Absent nonpulmonary visceral metastases
	AFP <1000 ng/mL	Any LDH, hCG
	Beta-hCG <5000 mIU/mL	
	LDH <1.5 x upper limit or normal (ULN)	
Intermediate	Gonadal or retroperitoneal primary site	Any primary site
	Absent nonpulmonary visceral metastases	Presence of nonpulmonary visceral metastases
	AFP 1000–10,000 ng/mL	Any LDH, hCG
	Beta-hCG 5000–50,000 mIU/mL	
	LDH 1.5–10 x ULN	
Poor	Mediastinal primary site	No patients classified as poor prognosis
	Presence of nonpulmonary visceral metastases	
	AFP ≥10,000 ng/mL	
	Beta-hCG	
	>50,000 mIU/mL	
	LDH > 10 x ULN	

Abbreviations: AFP, α fetoprotein; hCG, human chorionic gonadotropin; IGCCCG, International Germ Cell Consensus Classification Group; LDH, lactate dehydrogenase. ***Source:*** From International Germ Cell Cancer Consensus Group.

For ~90% of patients with good-risk GCTs, four cycles of etoposide plus cisplatin (EP) or three cycles of BEP produce durable complete responses, with minimal acute and chronic toxicity. Pulmonary toxicity is absent when bleomycin is not used and is rare when therapy is limited to 9 weeks; myelosuppression with neutropenic fever is less frequent; and the treatment mortality rate is negligible. Approximately 75% of intermediate-risk patients and 45% of poor-risk patients achieve durable complete remission with four cycles of BEP, and no regimen has proved superior. More effective therapy is needed.

POSTCHEMOTHERAPY SURGERY

Resection of residual metastases after the completion of chemotherapy is an integral part of therapy. If the initial histology is nonseminoma and the marker values have normalized, all sites of residual disease should be resected. In general, residual retroperitoneal disease requires a modified bilateral RPLND. Thoracotomy (unilateral or bilateral) and neck dissection are less frequently required to remove residual mediastinal, pulmonary parenchymal, or cervical nodal disease. Viable tumor (seminoma, embryonal carcinoma, yolk sac tumor, or choriocarcinoma) will be present in 15%, mature teratoma in 40%, and necrotic debris and fibrosis in 45% of resected specimens. The frequency of teratoma or viable disease is highest in residual mediastinal tumors. If necrotic debris or mature teratoma is present, no further chemotherapy is necessary. If viable tumor is present but is completely excised, two additional cycles of chemotherapy are given.

If the initial histology is pure seminoma, mature teratoma is rarely present, and the most frequent finding is necrotic debris. For residual retroperitoneal disease, a complete RPLND is technically difficult owing to extensive postchemotherapy fibrosis. Observation is recommended when no radiographic abnormality exists on CT scan. Positive findings on a positron emission tomography (PET) scan correlate with viable seminoma in residua, and mandate surgical excision or biopsy.

SALVAGE CHEMOTHERAPY

Of patients with advanced GCT, 20–30% fail to achieve a durable complete response to first-line chemotherapy. A combination of vinblastine, ifosfamide, and cisplatin (VeIP) will cure approximately 25% of patients as a second-line therapy. Substitution of paclitaxel for vinblastine may be more effective in this setting. Patients are more likely to achieve a durable complete response if they had a testicular primary tumor and relapsed from a prior complete remission to first-line cisplatin-containing chemotherapy. In contrast, if the patient failed to achieve a complete response or has a primary mediastinal nonseminoma, then standard-dose salvage therapy is rarely beneficial. Treatment options for such

patients include dose-intensive treatment, experimental therapies, and surgical resection.

Chemotherapy consisting of dose-intensive, high-dose carboplatin (≥1500 mg/m^2) plus etoposide (≥1200 mg/m^2), with or without cyclophosphamide, with peripheral blood stem cell support, induces a complete response in 25–40% of patients who have progressed after ifosfamide-containing salvage chemotherapy. Approximately one-half of the complete responses will be durable. High-dose therapy is standard of care for this patient population and has been suggested as treatment of choice for all patients with relapsed or refractory disease. Paclitaxel is also active in previously treated patients and shows promise in high-dose combination programs. Cure is still possible in some relapsed patients.

EXTRAGONADAL GCT AND MIDLINE CARCINOMA OF UNCERTAIN HISTOGENESIS

The prognosis and management of patients with extragonadal GCT depends on the tumor histology and site of origin. All patients with a diagnosis of extragonadal GCT should have a testicular ultrasound examination. Nearly all patients with retroperitoneal or mediastinal seminoma achieve a durable complete response to BEP or EP. The clinical features of patients with primary retroperitoneal nonseminoma GCT are similar to those of patients with a primary of testis origin, and careful evaluation will find evidence of a primary testicular GCT in about two-thirds of cases. In contrast, a primary mediastinal nonseminomatous GCT is associated with a poor prognosis; one-third of patients are cured with standard therapy (four cycles of BEP). Patients with newly diagnosed mediastinal nonseminoma are considered to have poor-risk disease and should be considered for clinical trials testing regimens of possibly greater efficacy. In addition, mediastinal nonseminoma is associated with hematologic disorders, including acute myelogenous leukemia, myelodysplastic syndrome, and essential thrombocytosis unrelated to previous chemotherapy. These hematologic disorders are very refractory to treatment. Nonseminoma of any primary site may change into other malignant histologies such as embryonal rhabdomyosarcoma or adenocarcinoma. This is called *malignant transformation*. i(12p) has been identified in the transformed cell type, indicating GCT clonal origin.

A group of patients with poorly differentiated tumors of unknown histogenesis, midline in distribution, and not associated with secretion of AFP or hCG has been described; a few (10–20%) are cured by standard cisplatin-containing chemotherapy. i(12p) is present in ~25% of such tumors (the fraction that are cisplatin-responsive), confirming their origin from primitive germ cells. This finding is also predictive of the response to cisplatin-based chemotherapy and resulting long-term survival. These tumors are heterogeneous; neuroepithelial tumors and lymphoma may also present in this fashion.

FERTILITY

Infertility is an important consequence of the treatment of GCTs. Preexisting infertility or impaired fertility is often present. Azoospermia and/or oligospermia are present at diagnosis in at least 50% of patients with testicular GCTs. Ejaculatory dysfunction is associated with RPLND, and germ cell damage may result from cisplatin-containing chemotherapy. Nerve-sparing techniques to preserve the retroperitoneal sympathetic nerves have made retrograde ejaculation less likely in the subgroups of patients who are candidates for this operation. Spermatogenesis does recur in some patients after chemotherapy. However, because of the significant risk of impaired reproductive capacity, semen analysis and cryopreservation of sperm in a sperm bank should be recommended to all patients before treatment.

FURTHER READINGS

Bosl Gj et al: Cancer of the testis, in Cancer: Principles and Practice of Oncology, 7th ed, VT DeVita et al (eds). Philadelphia, Lippincott Williams & Wilkins, 2008, pp 1463–1485

Ehrlich Y et al: Serum tumor markers in testicular cancer. Urol Oncol 3 Sep 2010 [Epub ahead of print]

Einhorn et al: High-dose chemotherapy and stem-cell rescue for metastatic germ-cell tumors. N Engl J Med 357:340, 2007[PMID: 17652649]

Feldman Dr et al: Medical treatment of advanced testicular cancer. JAMA 299:672, 2008[PMID: 18270356]

International Germ Cell Cancer Consensus Group: International Germ Cell Consensus Classification: A prognostic factor-based staging system for metastatic germ cell cancers. J Clin Oncol 15:594, 1997

Sonpaude G et al: Management of recurrent testicular germ cell tumors. Oncologist 12:51, 2007

Travis LB et al: Testicular cancer survivorship: Research strategies and recommendations. J Natl Cancer Inst 102:1114, 2010[PMID: 20585105]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OVARIAN CANCER

Incidence and Pathology

Ovarian cancer is the most lethal malignancy of gynecologic origin in the United States and other countries that have organized and effective cervical cancer screening programs. In 2010, 21,880 cases of ovarian cancer with 13,850 deaths are expected in the United States. The ovary is a complex and dynamic organ and, between the ages of approximately 11 and 50 years, is responsible for follicle maturation associated with egg maturation, ovulation, and cyclical sex steroid hormone production. These complex and linked biologic functions are coordinated through a variety of cells within the ovary, each of which possesses neoplastic potential. By far the most common and most lethal of the ovarian neoplasms arise from the ovarian epithelium found both on the surface of the ovary and in subsurface locations, known as cortical inclusion cysts, believed to be entrapped epithelium from the healing associated with prior follicle rupture during ovulation. The ovarian epithelium in good health appears as a simple epithelium, but with neoplastic transformation, it undergoes metaplastic changes into what is termed *müllerian epithelium*. The müllerian epithelium has a variety of subtypes each of which provide a specific phenotype of the tumor and in some cases different clinical presentations. Epithelial tumors are the most common ovarian neoplasm; they may be benign (50%), malignant (33%), or of borderline malignancy (16%). Age influences risk of malignancy; tumors in younger women are more likely benign. The most common of the ovarian epithelial malignancies are serous tumors (50%); tumors of mucinous (25%), endometrioid (15%), clear cell (5%), and transitional cell histology or Brenner tumor (1%) represent smaller proportions of epithelial ovarian tumors. In contrast, stromal tumors arise from the steroid hormone-producing cells and likewise have different phenotypes and clinical presentations largely dependent on the type and quantity of hormone production. Tumors arising in the germ cell are most similar in biology and behavior to testicular tumors in males (Chap. 96).

Tumors may also metastasize to the ovary from breast, colon, gastric, and pancreatic primaries. Bilateral ovarian masses from metastatic mucin-secreting gastrointestinal cancers are termed *Krukenberg tumors*.

Ovarian Cancer of Epithelial Origin

EPIDEMIOLOGY

A female has approximately a 1 in 72 lifetime risk (1.6%) of developing ovarian cancer, with the majority of affected women developing epithelial tumors. Epithelial tumors of the ovary have a peak incidence in women in their sixties, although age at presentation can range across the extremes of adult life, with cases being reported in women in their twenties to nineties. Known risk factors that increase the chance of subsequent ovarian cancer include epidemiologic, environmental, and genetic factors such as nulliparity, use of talc agents applied to the perineum, obesity, and probably hormone replacement therapy. Protective factors include the use of oral contraceptives, multiparity, and breast-feeding. These protective factors are thought to work through suppression of ovulation and perhaps reduction of ovarian inflammation and damage associated with the repair of the ovarian cortex associated with ovulation, and perhaps suppression of gonadotropins. Other protective factors, such as fallopian tube ligation are thought to protect the ovarian epithelium (or perhaps the distal fallopian tube fimbriae) from carcinogens that migrate from the vagina to the tubes and ovarian surface epithelium (see below).

GENETIC RISK FACTORS

A variety of genetic syndromes substantially increases a woman's risk of developing ovarian cancer. Approximately 10% of women with ovarian cancer have a somatic mutation in one of two DNA repair genes: BRCA1 (chromosome 17q12-21) or BRCA2 (chromosome 13q12-13). Individuals inheriting a single copy of a mutant allele have a very high incidence of breast and ovarian cancer. Most of these women have a family history that is notable for multiple cases of breast and/or ovarian cancer, although inheritance through male members of the family can camouflage this genotype through several generations. The most common malignancy in these women is breast carcinoma, although women harboring germ-line BRCA1 mutations have a marked increased risk of developing ovarian malignancies in their forties and fifties with a 30–50% lifetime risk of developing ovarian cancer. Women harboring a mutation in BRCA2 have a lower penetrance of ovarian cancer with perhaps a 20–40% chance of developing this malignancy, with onset typically in their fifties or sixties. Women with a

BRCA2 mutation also are at slightly increased risk of pancreatic cancer. Screening studies in this select population suggest that current screening techniques, including serial evaluation of the CA-125 tumor marker and ultrasound, are insufficient at detecting early-stage and curable disease, so women with these germ-line mutations are advised to undergo prophylactic removal of ovaries and fallopian tubes typically after completing childbearing and ideally before ages 35–40. Early prophylactic oophorectomy also protects these women from subsequent breast cancer with a reduction of breast cancer risk of approximately 50%.

Ovarian cancer is also one form of cancer (along with colorectal and endometrial cancer) that may develop in women with Lynch syndrome, type II, caused by mutations in DNA mismatch repair genes (*MSH2, MLH1, MLH6, PMS1, PMS2*). Ovarian cancer may appear in women younger than 50 years of age in this syndrome.

PRESENTATION

Neoplasms of the ovary tend to be painless unless they undergo torsion. Symptoms are therefore typically related to compression of local organs or due to symptoms from metastatic disease. Women with tumors localized to the ovary do have an increased incidence of symptoms including pelvic discomfort, bloating, and perhaps changes in a woman's typical urinary or bowel pattern. Unfortunately, these symptoms are frequently dismissed by either the woman or her health care team. It is believed that high-grade tumors metastasize early in the neoplastic process. Unlike other epithelial malignancies, these tumors tend to exfoliate throughout the peritoneal cavity and thus present with symptoms associated with disseminated intraperitoneal tumors. The most common symptoms at presentation include a multimonth period of progressive complaints that typically include some combination of heartburn, nausea, early satiety, indigestion, constipation, and abdominal pain. Signs include the rapid increase in abdominal girth due to the accumulation of ascites that typically alerts the patient and her physician that the concurrent gastrointestinal symptoms are likely associated with serious pathology. Radiologic evaluation typically demonstrates a complex adnexal mass and ascites. Laboratory evaluation demonstrates a markedly elevated CA-125, a shed mucin (Muc 16) associated with, but not specific for, ovarian cancer. Hematogenous and lymphatic spread are seen but are not the typical presentation. Ovarian cancers are divided into four stages, with stage I tumors confined to the ovary, stage II malignancies confined to the pelvis, and stage III confined to the peritoneal cavity (Table 97–1). These three stages are subdivided, with the most common presentation, stage IIIc, defined as tumors with bulky intraperitoneal disease. About 70% of women present with stage IIIc disease. Stage IV disease includes women with parenchymal metastases (liver, lung, spleen) or, alternatively, abdominal wall or pleural disease. The 30% not presenting with stage IIIc disease are roughly evenly distributed among the other stages.

Table 97–1 Staging and Survival in Gynecologic Malignancies

Stage	Ovarian	5-Year Survival, %	Endometrial	5-Year Survival, %	Cervix	5-Year Survival, %
0	—		—		Carcinoma in situ	100
I	Confined to ovary	90–95	Confined to corpus	89	Confined to uterus	85
II	Confined to pelvis	70–80	Involves corpus and cervix	73	Invades beyond uterus but not to pelvic wall	65
III	Intraabdominal spread	20–50	Extends outside the uterus but not outside the true pelvis	52	Extends to pelvic wall and/or lower third of vagina, or hydronephrosis	35
IV	Spread outside abdomen	1–5	Extends outside the true pelvis or involves the bladder or rectum	17	Invades mucosa of bladder or rectum or extends beyond the true pelvis	7

SCREENING

Ovarian cancer is the fifth most lethal malignancy in women in the United States, curable in early stages, and seldom curable in advanced stages; hence, screening is of considerable interest. Furthermore the ovary is well visualized with a variety of imaging techniques, most notably transvaginal ultrasound. Early-stage tumors often produce proteins that can be measured in the blood such as CA-125 and HE-4. Nevertheless, the incidence of ovarian cancer in the middle-aged female population is low, with only approximately 1 in 2000 women between the ages of 50 and 60 carrying an asymptomatic and undetected tumor. Thus effective screening techniques must be sensitive but, more importantly, highly specific so to minimize the number of false positives. Even a screening test with 98% specificity and 50% sensitivity would have a positive predictive value of only about 1%. Despite these formidable barriers, ongoing studies are evaluating the utility of various screening strategies.

However, screening for ovarian cancer is currently not recommended outside of a clinical trial.

Treatment: Ovarian Cancer

In women presenting with a localized ovarian mass, the principal diagnostic and therapeutic maneuver is to determine if the tumor is benign or malignant and, in the event that the tumor is malignant, whether the tumor arises in the ovary or is a site of metastatic disease. Metastatic disease to the ovary can be seen from primary tumors of the colon, appendix, stomach (Krukenberg tumors), and breast. Typically women undergo a unilateral salpingo-oophorectomy, and if pathology reveals a primary ovarian malignancy, then the procedure is followed by a hysterectomy, removal of the remaining tube and ovary, omentectomy, and pelvic node sampling along with some random biopsies of the peritoneal cavity. This extensive surgical procedure is performed because approximately 30% of tumors that by visual inspection appear to be confined to the ovary have already disseminated to the peritoneal cavity and/or surrounding lymph nodes.

If there is evidence of bulky intraabdominal disease, a comprehensive attempt at maximal tumor cytoreduction is attempted even if it involves partial bowel resection, splenectomy, and in certain cases more extensive upper abdominal surgery. The ability to debulk metastatic ovarian cancer to minimal visible disease is associated with an improved prognosis as compared to women left with visible disease. Patients without gross residual disease after resection have a median survival of 39 months, compared to 17 months for those left with macroscopic tumor. Once tumors have been surgically debulked, women receive therapy with a platinum agent typically with a taxane. Debate continues as to whether this therapy should be delivered intravenously or alternatively whether some of the therapy should be delivered directly into the peritoneal cavity via a catheter. Three randomized studies have demonstrated improved survival with intraperitoneal therapy, but this approach is still not widely accepted due to technical challenges associated with this delivery route and increased toxicity. In women who present with bulky disease, an alternative approach is to treat with platinum plus a taxane for several cycles (neoadjuvant therapy). Subsequent surgical procedures are more effective at leaving the patient without gross residual tumor, and survival is comparable to surgery followed by chemotherapy.

With optimal debulking surgery and platinum-based chemotherapy [usually carboplatin dosed to an area under the curve (AUC) of 7.5 plus paclitaxel 175 mg/m^2 by 3-h infusion in monthly cycles], 70% of women who present with advanced-stage tumors respond, and 40–50% experience a complete remission with normalization of their CA-125, CT scans, and physical examination. Unfortunately, only half the complete responders remain in remission. Disease recurs within 1 to 4 years from the completion of their primary therapy in half the complete responders. CA-125 levels often increase as a first sign of relapse; however, data are not clear that early intervention influences survival. Recurrent disease is effectively managed, but not cured, with a variety of chemotherapeutic agents. Eventually all of these women develop chemotherapy-refractory disease at which point refractory ascites, poor bowel motility, and obstruction or pseudoobstruction due to a tumor-infiltrated aperistaltic bowel are common. Limited surgery to relieve intestinal obstruction, localized radiation therapy to relieve pressure or pain from masses, or palliative chemotherapy may be helpful. Agents with >15% response rates include gemcitabine, topotecan, liposomal doxorubicin, and bevacizumab. Approximately 20% of ovarian cancers are HER2/neu positive, and trastuzumab may induce responses in this subset.

Five-year survival correlates with the stage of disease: stage I, 90–95%; stage II, 70–80%; stage III, 20–50%; stage IV, 1–5% (Table 97–1). Prognosis is also influenced by histologic grade: 5-year survival is 88% for well-differentiated tumors, 58% for moderately differentiated tumors, and 27% for poorly differentiated tumors. Histologic type has less influence on outcome. Patients with tumors of low malignant potential are managed by surgery; chemotherapy and radiation therapy do not improve survival.

Ovarian Sex Cord and Stromal Tumors

EPIDEMIOLOGY, PRESENTATION, AND PREDISPOSING SYNDROMES

Approximately 7% of ovarian neoplasms are stromal or sex cord tumors, with approximately 1800 cases expected each year in the United States. Ovarian stromal tumors or sex cord tumors are most common in women in their fifties or sixties, but tumors can present in the extremes of age, including the pediatric population. These tumors arise from the mesenchymal components of the ovary, including steroid-producing cells as well as fibroblasts. Essentially all of these tumors are of low malignant potential and present as unilateral solid masses. Three clinical presentations are common: the detection of an abdominal mass abdominal pain due to ovarian torsion, intratumoral hemorrhage, or rupture or signs and symptoms due to hormonal production by these tumors.

The most common hormone-producing tumors include thecomas, granulosa cell tumor, or juvenile granulosa tumors in children. These estrogen-producing tumors often present with breast tenderness as well as isosexual precocious pseudopuberty in children, menometrorrhagia, oligomenorrhea, or amenorrhea in premenopausal women, or alternatively as postmenopausal bleeding in older women. In some women, estrogen-associated secondary malignancies, such as endometrial or breast cancer may present as synchronous malignancies. Alternatively, endometrial cancer may serve as the

presenting malignancy with evaluation subsequently identifying a unilateral solid ovarian neoplasm that proves to be an occult granulosa cell tumor. Sertoli-Leydig tumors often present with hirsutism, virilization, and occasionally Cushing's syndrome due to increased production of testosterone, androstenedione, or other 17-ketosteroids. Hormonally inert tumors include fibroma that presents as a solitary mass often in association with ascites and occasionally hydrothorax also known as Meigs' syndrome. A subset of these tumors present in individuals with a variety of inherited disorders that predispose them to mesenchymal neoplasia. Associations include juvenile granulosa cell tumors and perhaps Sertoli-Leydig tumors with Ollier's disease (multiple enchondromatosis) or Maffucci's syndrome, ovarian sex cord tumors with annular tubules with Peutz-Jeghers syndrome, and fibromas with Gorlin disease.

Treatment: Sex Cord Tumors

The mainstay of treatment for sex cord tumors is surgical resection. Most women present with tumors confined to the ovary. For the small subset of women who present with metastatic disease or develop evidence of tumor recurrence after primary resection, survival is still typically long, often in excess of a decade. Because these tumors are slow growing and relatively refractory to chemotherapy, women with metastatic disease are often debulked as disease is usually peritoneal-based (as with epithelial ovarian cancer). Definitive data that surgical debulking of metastatic or recurrent disease prolongs survival are lacking, but ample data document women who have survived years or in some cases decades after resection of recurrent disease. In addition, large peritoneal-based metastases also have a proclivity for hemorrhage, sometimes with catastrophic complications. Chemotherapy is occasionally effective, and women tend to receive regimens designed to treat epithelial or germ cell tumors. These tumors often produce high levels of müllerian inhibiting substance (MIS), inhibin, and in the case of Sertoli-Leydig tumors α fetoprotein (AFP). These proteins are detectable in serum and can be used as tumor markers to monitor women for recurrent disease as the increase and decrease of these proteins in the serum tend to reflect the changing bulk of systemic tumor.

GERM CELL TUMORS OF THE OVARY

Germ cell tumors, like their counterparts in the testis, are cancers of germ cells. These totipotent cells contain the programming for differentiation to essentially all tissue types, and hence the germ cell tumors include a histologic menagerie of bizarre tumors, including benign teratomas and a variety of malignant tumors, such as immature teratomas, dysgerminomas, yolk sac malignancies, and choriocarcinomas. Benign teratoma (or dermoid cyst) is the most common germ cell neoplasm of the ovary and often presents in young woman. These tumors include a complex mixture of differentiated tissue including tissues from all three germ layers. In older women these differentiated tumors can develop malignant transformation, most commonly squamous cell carcinomas. Malignant germ cell tumors include dysgerminomas, yolk sac tumors, immature teratomas, as well as embryonal and choriocarcinomas. There are no known genetic abnormalities that unify these tumors. A subset of dysgerminomas harbor mutations in c-Kit oncogenes [as seen in gastrointestinal stromal tumors (GIST)], whereas a subset of germ cell tumors have isochromosome 12 abnormalities as seen in testicular malignancies. In addition, a subset of dysgerminomas is associated with dysgenetic ovaries. Identification of a dysgerminoma arising in genotypic XY gonads is important in that it highlights the need to identify and remove the contralateral gonad due to risk of gonadoblastoma.

PRESENTATION

Germ cell tumors can present at all ages, but the peak age of presentation tends to be in females in their late teens or early twenties. Typically these tumors will become large ovarian masses, which eventually present as palpable low abdominal or pelvic masses. Like sex cord tumors, torsion or hemorrhage may present urgently or emergently as acute abdominal pain. Some of these tumors produce elevated levels of human chorionic gonadotropin (hCG) that can lead to isosexual precocious puberty when tumors present in younger girls. Unlike epithelial ovarian cancer, these tumors have a higher proclivity for nodal or hematogenous metastases. As with testicular tumors some of these tumors tend to produce AFP (yolk sac tumors) or hCG (embryonal and choriocarcinomas as well as some dysgerminomas) that are reliable tumor markers.

Treatment: Germ Cell Tumors

Germ cell tumors typically present in women who are still of childbearing age, and because bilateral tumors are uncommon (except in dysgerminoma, 10–15%), the typical treatment is unilateral oophorectomy or salpingo-oophorectomy. Because nodal metastases to pelvic and para-aortic nodes are common and may affect treatment choices, these nodes should be carefully inspected, and if enlarged, should be resected if possible. Women with malignant germ cell tumors typically receive bleomycin, etoposide, and cisplatin (BEP) chemotherapy. In the majority of women, even those with advanced-stage disease, cure is expected. Close follow-up without adjuvant therapy of women with stage I tumors is reasonable if there is high confidence that the patient and health care team are committed to compulsive and careful follow-up, as chemotherapy at the time of tumor recurrence is likely to be curative.

Dysgerminoma is the ovarian counterpart of testicular seminoma. The 5-year disease-free survival is

100% in early-stage patients and 61% in stage III disease. Although the tumor is highly radiation-sensitive, radiation produces infertility in many patients. BEP chemotherapy is as effective or more so without causing infertility. The use of BEP following incomplete resection is associated with 95%, 2-year disease-free survival. This chemotherapy is now the treatment of choice for dysgerminoma.

FALLOPIAN TUBE CANCER

Transport of the egg to the uterus occurs via transit through the fallopian tube, with the distal ends of these tubes composed of fimbriae that drape about the ovarian surface and capture the egg as it erupts from the ovarian cortex. Fallopian tube malignancies typically have the same histologic pattern as ovarian malignancies, with the most common epithelial malignancy being of serous histology. Previous teaching was that these malignancies were rare, but more careful histologic examination suggests that many "ovarian malignancies" might actually arise in the distal fimbria of the fallopian tube. Data supporting this theory are strongest in the population of women who carry BRCA1 or BRCA2 somatic mutations. These women often present with adnexal masses, and like ovarian cancer, these tumors spread relatively early throughout the peritoneal cavity and respond to platinum and taxane therapy and have a natural history that is essentially identical to ovarian cancer (Table 97–1).

CERVICAL CANCER

Global Considerations

Cervical cancer is the second most common and most lethal malignancy in women worldwide likely due to the widespread infection with high-risk strains of human papillomavirus (HPV) and limited utilization or access to Pap smear screening in many nations throughout the world. Nearly 500,000 cases of cervical cancer are expected worldwide with approximately 240,000 deaths annually. Cancer incidence is particularly high in women residing in central and South America, the Caribbean, and southern and eastern Africa. Mortality rate is disproportionately high in Africa. In the United States, 12,200 women were diagnosed with cervical cancer and 4210 women died. Whereas efforts in developed countries have looked at high-technology screening techniques for HPV involving polymerase chain reaction and other molecular technologies, there is an urgent need for high-throughput low-technology strategies to identify and treat women bearing high-risk but treatable cervical dysplasia. The development of effective vaccines for high-risk HPV types makes it imperative to determine economical, socially acceptable, and logistically feasible strategies to deliver and distribute this vaccine to girls and perhaps boys before their engagement in sexual activity.

HPV Infection and Preventive Vaccination

HPV is the primary neoplastic-initiating event in the vast majority of women with invasive cervical cancer. This double-strand DNA virus infects epithelium near the transformation zone of the cervix. More than 60 types of HPV are known, with approximately 20 types having the ability to generate high-grade dysplasia and malignancy. HPV16 and 18 are the types most frequently associated with high-grade dysplasia and targeted by both FDA-approved vaccines. The large majority of sexually active adults are exposed to HPV, and most women clear the infection without specific intervention. The 8-kilobase HPV genome encodes seven early genes, most notably E6 and E7, which can bind to RB and p53, respectively. High-risk types of HPV encode E6 and E7 molecules that are particularly effective at inhibiting the normal cell cycle checkpoint functions of these regulatory proteins, leading to immortalization but not full transformation of cervical epithelium. A minority of woman will fail to clear the infection with subsequent HPV integration into the host genome. Over the course of as short as months but more typically years, some of these women develop high-grade dysplasia. The time from dysplasia to carcinoma is likely years to more than a decade and almost certainly requires the acquisition of other poorly defined genetic mutations within the infected and immortalized epithelium.

Risk factors include a high number of sexual partners, age of first intercourse, and history of venereal disease. Smoking is a cofactor; heavy smokers have a higher risk of dysplasia with HPV infection. HIV infection, especially when associated with low CD4+ T cell counts, is associated with a higher rate of high-grade dysplasia and likely a shorter latency period between infection and invasive disease.

Currently approved vaccines include the recombinant proteins to the late proteins, L1 and L2 of HPV-16 and -18. Vaccination of women before the initiation of sexual activity dramatically reduces the rate of HPV-16 and -18 infection and subsequent dysplasia. There is also partial protection against other HPV types, although vaccinated women are still at risk for HPV infection and still require standard Pap smear screening. Although no randomized trial data demonstrate the utility of Pap smears, the dramatic drop in cervical cancer incidence and death in developed countries employing wide-scale screening provides strong evidence for its effectiveness. The incorporation of HPV testing by polymerase chain reaction (PCR) or other molecular techniques increases the sensitivity of detecting cervical pathology but at the cost of lower sensitivity in that it identifies many women with transient infections who require no specific medical intervention.

Clinical Presentations

The majority of cervical malignancies are squamous cell carcinomas associated with HPV.

Adenocarcinomas are also HPV-related and arise deep in the endocervical canal; they are typically not seen by visual inspection of the cervix and thus often missed by Pap smear screening. A variety of rarer malignancies including atypical epithelial tumors, carcinoids, small cell carcinomas, sarcomas, and lymphomas have also been reported.

The principal role of Pap smear testing is the detection of asymptomatic preinvasive cervical dysplasia of squamous epithelial lining. Invasive carcinomas often have symptoms or signs including postcoital spotting or intermenstrual cycle bleeding or menometrorrhagia. Foul-smelling or persistent yellow discharge may also be seen. Presentations that include pelvic or sacral pain suggest lateral extension of the tumor into pelvic nerve plexus by either the primary tumor or a pelvic node and are signs of advanced-stage disease. Likewise, flank pain from hydronephrosis from ureteral compression or deep venous thrombosis from iliac vessel compression suggests either extensive nodal disease or direct extension of the primary tumor to the pelvic sidewall. The most common finding of physical exam is a visible tumor on the cervix.

Treatment: Cervical Cancer

Scans are not part of the formal clinical staging of cervical cancer yet are very useful in planning appropriate therapy. CT can detect hydronephrosis indicative of pelvic sidewall disease but is not accurate at evaluating other pelvic structures. MRI is more accurate at estimating uterine extension and paracervical extension of disease into soft tissues typically bordered by broad and cardinal ligaments that support the uterus in the central pelvis. Positron emission tomography (PET) scan may be the most accurate technique for evaluating the pelvis and more importantly nodal (pelvic, para-aortic, and scalene) sites for disease. This technique seems more prognostic (and probably accurate) than CT, MRI, or lymphangiogram, especially in the para-aortic region.

Stage I cervical tumors are confined to the cervix, whereas stage II tumors extend into the upper vagina or paracervical soft tissue (**Fig. 97-1**). Stage III tumors extend to the lower vagina or the pelvic sidewalls, whereas stage IV tumors invade the bladder or rectum or have spread to distant sites. Very small stage I cervical tumors can be treated with a variety of surgical procedures. In young women desiring to maintain fertility, radical trachelectomy removes the cervix with subsequent anastomosis of the upper vagina to the uterine corpus. Larger cervical tumors confined to the cervix can be treated with either surgical resection or radiation therapy in combination with cisplatin-based chemotherapy with a high chance of cure. Larger tumors that extend down the vagina or into the paracervical soft tissues or the pelvic sidewalls are treated with combination chemotherapy and radiation therapy. The treatment of recurrent or metastatic disease is unsatisfactory due to the relative resistance of these tumors to chemotherapy and currently available biological agents.

Figure 97-1

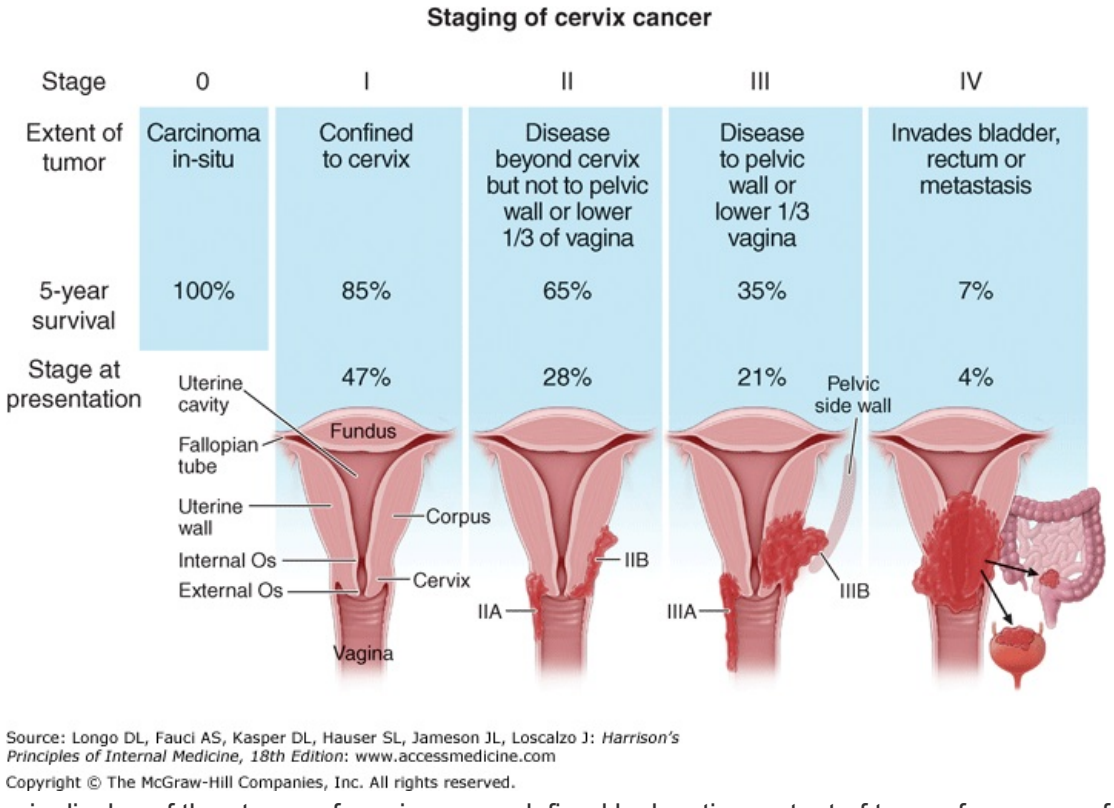

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anatomic display of the stages of cervix cancer defined by location, extent of tumor, frequency of presentation, and 5-year survival.

UTERINE CANCER

Epidemiology

Several different tumor types arise in uterine corpus. Most tumors arise in the glandular lining and are endometrial adenocarcinomas. Tumors can also arise in the smooth muscle; most are benign (uterine leiomyoma), with a small minority of tumors being sarcomas. The endometrioid histologic subtype of endometrial cancer is the most common gynecologic malignancy in the United States. In 2010, it was diagnosed in 43,470 women and 7950 women died from the disease. Development of these tumors is a multistep process with estrogen playing an important early role in driving endometrial gland proliferation. Relative overexposure to this class of hormones is a risk factor for the subsequent development of endometrioid tumors. In contrast progestins drive glandular maturation and are protective. Hence, women with high endogenous or pharmacologic exposure to estrogens, especially if unopposed by progesterone are at high risk for endometrial cancer. Obese women, women treated with unopposed estrogens, or women with estrogen-producing tumors (such as granulosa cell tumors of the ovary) are at higher risk for endometrial cancer. In addition, treatment with tamoxifen, which has antiestrogenic effects in breast tissue but estrogenic effects in uterine epithelium, is associated with an increased risk of endometrial cancer. Secondary events such as the loss of the PTEN or Cables tumor suppressor genes likely serve as secondary events in carcinogenesis. The molecular events that underlie less common endometrial cancers such as clear cell and papillary serous tumors of the uterine corpus are unknown.

Women with mutation in one of a series of DNA mismatch repair genes associated with the Lynch syndrome, also known as hereditary nonpolyposis colon cancer (HNPCC) syndrome, are at increased risk for endometrioid endometrial carcinoma. These individuals have germ-line mutations in MSH2, MLH1, and in rare cases PMS1, and PMS2. Individuals who carry these mutations typically have a family history of cancer and are at markedly increased risk for colon cancer and modestly increased risk for ovarian cancer and a variety of other tumors. Middle-aged women with HNPCC carry a 4% annual risk of endometrial cancer and a relative overall risk of approximately 200-fold as compared to age-matched women without HNPCC.

Pathology

Approximately 75–80% of endometrial cancers are adenocarcinomas. Prognosis depends on stage, histologic grade, and depth of myometrial invasion. Approximately 10% of patients have tumors with areas of squamous cell differentiation. When the tumor is well differentiated, it is called *adenoacanthoma*; when poorly differentiated, *adenosquamous* carcinoma. Less common histologies include mucinous carcinoma (5%) and a papillary serous tumor (<10%) that behaves like ovarian cancer.

Presentations

The majority of women with tumors of the uterine corpus present with postmenopausal vaginal bleeding due to shedding of the malignant endometrial lining. Premenopausal women often will present with atypical bleeding between typical menstrual cycles. These signs typically bring a woman to the attention of a health care professional, and hence the majority of women present with early-stage disease in which the tumor is confined to the uterine corpus. Diagnosis is typically established by endometrial biopsy. Epithelial tumors may spread to pelvic or para-aortic lymph nodes. Pulmonary metastases can appear later in the natural history of this disease but are very uncommon at initial presentation. Serous tumors tend to have patterns of spread much more reminiscent of ovarian cancer with many patients presenting with omental disease and sometimes ascites. Some women presenting with uterine sarcomas will present with pelvic pain. Nodal metastases are uncommon with sarcomas, which are more likely to present with either intraabdominal disease or pulmonary metastases.

Treatment: Uterine Cancer

Most women with endometrial cancer have disease that is localized to the uterus (75% are stage I, Table 97–1), and definitive treatment typically involves a hysterectomy with removal of the ovaries and fallopian tubes. The resection of lymph nodes does not improve outcome but does provide prognostic information. Node involvement defines stage III disease, present in 13% of patients. Tumor grade and depth of invasion are the two key prognostic variables in early-stage tumors, and women with low-grade and/or minimally invasive tumors are typically observed after definitive surgical therapy. Patients with high-grade tumors or tumors that are deeply invasive (stage IB, 13%) are at higher risk for pelvic recurrence or recurrence at the vaginal cuff, which is typically prevented by vaginal vault brachytherapy.

Women with regional metastases or metastatic disease (3% of patients) with low-grade tumors can be treated with progesterone. Poorly differentiated tumors are typically resistant to hormonal manipulation and thus are treated with chemotherapy. The role of chemotherapy in the adjuvant setting is currently under investigation. Chemotherapy for metastatic disease is delivered with palliative intent.

Five-year survival is 89% for stage I, 73% for stage II, 52% for stage III, and 17% for stage IV disease

(Table 97–1).

GESTATIONAL TROPHOBLASTIC TUMORS

Global Considerations

Gestational trophoblastic diseases represent a spectrum of neoplasia from benign hydatidiform mole to choriocarcinoma due to persistent trophoblastic disease associated most commonly with molar pregnancy but occasionally seen after normal gestation. The most common presentations of trophoblastic tumors are partial and complete molar pregnancies. These represent approximately 1 in 1500 conceptions in developed Western countries. The incidence widely varies globally, with areas in Southeast Asia having a much higher incidence of molar pregnancy. Regions with high molar pregnancy rates are often associated with diets low in carotene and animal fats.

Risk Factors

Trophoblastic tumors result from the outgrowth or persistence of placental tissue. They arise most commonly in the uterus but can also arise in other sites such as the fallopian tubes due to ectopic pregnancy. Risk factors include poorly defined dietary and environmental factors as well as conceptions at the extremes of reproductive age, with the incidence particularly high in females conceiving younger than age 16 or older than age 50. In older women, the incidence of molar pregnancy might be as high as one in three, likely due to increased risk of abnormal fertilization of the aged ova. Most trophoblastic neoplasms are associated with complete moles, diploid tumors with all genetic material from the paternal donor (known as parental disomy). This is thought to occur when a single sperm fertilizes an enucleate egg that subsequently duplicates the paternal DNA. Trophoblastic proliferation occurs with exuberant villous stroma. If pseudopregnancy extends out past the 12th week, fluid progressively accumulates within the stroma leading to "hydropic changes." There is no fetal development in complete moles.

Partial moles arise from the fertilization of an egg with two sperm, hence two-thirds of genetic material is paternal in these triploid tumors. Hydropic changes are less dramatic, and fetal development can often occur through late first trimester or early second trimester at which point spontaneous abortion is common. Laboratory findings will include excessively high hCG and high AFP. The risk of persistent gestational trophoblastic disease after partial mole is approximately 5%.

Complete and partial moles can be noninvasive or invasive. Myometrial invasion occurs in no more than one in six complete moles and a lower portion of partial moles.

Presentation of Invasive Trophoblastic Disease

The clinical presentation of molar pregnancy is changing in developed countries due to the early detection of pregnancy with home pregnancy kits and the very early use of Doppler and ultrasound to evaluate the early fetus and uterine cavity for evidence of a viable fetus. Thus, in these countries, the majority of women presenting with trophoblastic disease have their moles detected early and have typical symptoms of early pregnancy including nausea, amenorrhea, and breast tenderness. With uterine evacuation of early complete and partial moles, most women experience spontaneous remission of their disease as monitored by serial hCG levels. These women require no chemotherapy. Patients with persistent elevation of hCG or rising hCG postevacuation have persistent or actively growing gestational trophoblastic disease and require therapy. Most series suggest that between 15 and 25% of women will have evidence of persistent gestational trophoblastic disease after molar evacuation.

In women who lack access to prenatal care, presenting symptoms can be life threatening including the development of preeclampsia or even eclampsia. Hyperthyroidism can also be seen. Evacuation of large moles can be associated with life-threatening complications including uterine perforation, volume loss, high-output cardiac failure, and adult respiratory distress syndrome (ARDS).

For women with evidence of rising hCG or radiologic confirmation of metastatic or persistent regional disease, prognosis can be estimated through a variety of scoring algorithms that identify those women at low, intermediate, and high risk for requiring multi-agent chemotherapy. In general, women with widely metastatic nonpulmonary disease, very elevated hCG, and prior normal antecedent term pregnancy are considered at high risk and typically require multi-agent chemotherapy for cure.

Treatment: Invasive Trophoblastic Disease

The management for a persistent and rising hCG postevacuation of a molar conception is typically chemotherapy, although surgery can play an important role for disease that is persistently isolated in the uterus (especially if childbearing is complete) or to control hemorrhage. For women wishing to maintain fertility or with metastatic disease, the preferred treatment is chemotherapy. Chemotherapy is guided by the hCG level, which typically drops to undetectable levels with effective therapy. Single-agent treatment with methotrexate, or actinomycin D 90% cures of women with low-risk disease. Patients with high-risk disease (high hCG levels, presentation 4 or more months after pregnancy, brain or liver metastases, failure of methotrexate therapy) are typically treated with multi-agent chemotherapy [e.g., etoposide, methotrexate, and actinomycin D alternating with cyclophosphamide

and vincristine (EMA-CO)], which is typically curative even in those women with extensive metastatic disease. Cisplatin, bleomycin, and either etoposide or vinblastine are also active combinations. Survival in high-risk disease exceeds 80%. Cured women may get pregnant again without evidence of increased fetal or maternal complications.

FURTHER READINGS

Berkowitz Rs, Goldstein Dp: Clinical practice. Molar pregnancy. N Engl J Med 360:1639, 2009[PMID: 19369669]

Clarke-Pearson DL: Clinical practice. Screening for ovarian cancer. N Engl J Med 361:170, 2009[PMID: 19587342]

Jatoi I, Anderson WF: Management of women who have a genetic predisposition for breast cancer. Surg Clin North Am 88:845, 2008[PMID: 18672143]

Kahn JA: HPV vaccination for the prevention of cervical intraepithelial neoplasia. N Engl J Med 361:271, 2009[PMID: 19605832]

Ledermann JA, Kristeleit RS: Optimal treatment for relapsing ovarian cancer. Ann Oncol 21 Suppl 7:vii218, 2010

Moxley KM, McMeekin DS: Endometrial carcinoma: A review of chemotherapy, drug resistance, and the search for new agents. Oncologist 15:1026, 2010[PMID: 20930101]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SOFT TISSUE AND BONE SARCOMAS AND BONE METASTASES: INTRODUCTION

Sarcomas are rare (<1% of all malignancies) mesenchymal neoplasms that arise in bone and soft tissues. These tumors are usually of mesodermal origin, although a few are derived from neuroectoderm, and they are biologically distinct from the more common epithelial malignancies. Sarcomas affect all age groups; 15% are found in children <15 years of age, and 40% occur after age 55 years. Sarcomas are one of the most common solid tumors of childhood and are the fifth most common cause of cancer deaths in children. Sarcomas may be divided into two groups, those derived from bone and those derived from soft tissues.

SOFT TISSUE SARCOMAS

Soft tissues include muscles, tendons, fat, fibrous tissue, synovial tissue, vessels, and nerves. Approximately 60% of soft tissue sarcomas arise in the extremities, with the lower extremities involved three times as often as the upper extremities. Thirty percent arise in the trunk, the retroperitoneum accounting for 40% of all trunk lesions. The remaining 10% arise in the head and neck.

Incidence

Approximately 11,000 new cases of soft tissue sarcomas occurred in the United States in 2010. The annual age-adjusted incidence is 3 per 100,000 population, but the incidence varies with age. Soft tissue sarcomas constitute 0.7% of all cancers in the general population and 6.5% of all cancers in children.

Epidemiology

Malignant transformation of a benign soft tissue tumor is extremely rare, with the exception that malignant peripheral nerve sheath tumors (neurofibrosarcoma, malignant schwannoma) can arise from neurofibromas in patients with neurofibromatosis. Several etiologic factors have been implicated in soft tissue sarcomas.

ENVIRONMENTAL FACTORS

Trauma or previous injury is rarely involved, but sarcomas can arise in scar tissue resulting from a prior operation, burn, fracture, or foreign body implantation. Chemical carcinogens such as polycyclic hydrocarbons, asbestos, and dioxin may be involved in the pathogenesis.

IATROGENIC FACTORS

Sarcomas in bone or soft tissues occur in patients who are treated with radiation therapy. The tumor nearly always arises in the irradiated field. The risk increases with time.

VIRUSES

Kaposi's sarcoma (KS) in patients with HIV type 1, classic KS, and KS in HIV-negative homosexual men is caused by human herpesvirus (HHV) 8 (Chap. 182). No other sarcomas are associated with viruses.

IMMUNOLOGIC FACTORS

Congenital or acquired immunodeficiency, including therapeutic immunosuppression, increases the risk of sarcoma.

Genetic Considerations

Li-Fraumeni syndrome is a familial cancer syndrome in which affected individuals have germ-line abnormalities of the tumor-suppressor gene p53 and an increased incidence of soft tissue sarcomas and other malignancies, including breast cancer, osteosarcoma, brain tumor, leukemia, and adrenal carcinoma (Chap. 83). Neurofibromatosis 1 (NF-1, peripheral form, von Recklinghausen's disease) is characterized by multiple neurofibromas and café au lait spots. Neurofibromas occasionally undergo malignant degeneration to become malignant peripheral nerve sheath tumors. The gene for NF-1 is located in the pericentromeric region of chromosome 17 and encodes neurofibromin, a tumor-suppressor protein with guanosine 5'-triphosphate (GTP)ase-activating activity that inhibits Ras function (Chap. 379). Germ-line mutation of the *Rb-1* locus (chromosome 13q14) in patients with inherited retinoblastoma is associated with the development of osteosarcoma in those who survive the retinoblastoma and of soft tissue sarcomas unrelated to radiation therapy. Other soft tissue tumors, including desmoid tumors, lipomas, leiomyomas, neuroblastomas, and paragangliomas, occasionally

show a familial predisposition.

Ninety percent of synovial sarcomas contain a characteristic chromosomal translocation t(X;18)(p11;q11) involving a nuclear transcription factor on chromosome 18 called *SYT* and two breakpoints on X. Patients with translocations to the second X breakpoint (*SSX2*) may have longer survival than those with translocations involving *SSX1*.

Insulin-like growth factor (IGF) type II is produced by some sarcomas and may act as an autocrine growth factor and as a motility factor that promotes metastatic spread. IGF-II stimulates growth through IGF-I receptors, but its effects on motility are through different receptors. If secreted in large amounts, IGF-2 may produce hypoglycemia (Chaps. 100, 345).

Classification

Approximately 20 different groups of sarcomas are recognized on the basis of the pattern of differentiation toward normal tissue. For example, rhabdomyosarcoma shows evidence of skeletal muscle fibers with cross-striations; leiomyosarcomas contain interlacing fascicles of spindle cells resembling smooth muscle; and liposarcomas contain adipocytes. When precise characterization of the group is not possible, the tumors are called *unclassified sarcomas*. All of the primary bone sarcomas can also arise from soft tissues (e.g., extraskeletal osteosarcoma). The entity *malignant fibrous histiocytoma (MFH)* includes many tumors previously classified as fibrosarcomas or as pleomorphic variants of other sarcomas and is characterized by a mixture of spindle (fibrous) cells and round (histiocytic) cells arranged in a storiform pattern with frequent giant cells and areas of pleomorphism. As immunohistochemical suggestion of differentiation, particularly myogenic differentiation, may be found in a significant fraction of these patients, many are now characterized as poorly differentiated leiomyosarcomas, and the terms *undifferentiated pleomorphic sarcoma* and *myxofibrosarcoma* are replacing MFH and myxoid MFH.

For purposes of treatment, most soft tissue sarcomas can be considered together. However, some specific tumors have distinct features. For example, *liposarcoma* can have a spectrum of behaviors. Pleomorphic liposarcomas and dedifferentiated liposarcomas behave like other high-grade sarcomas; in contrast, well-differentiated liposarcomas (better termed *atypical lipomatous tumors*) lack metastatic potential, and myxoid liposarcomas metastasize infrequently, but, when they do, they have a predilection for unusual metastatic sites containing fat, such as the retroperitoneum, mediastinum, and subcutaneous tissue. Rhabdomyosarcomas, Ewing's sarcoma, and other small-cell sarcomas tend to be more aggressive and are more responsive to chemotherapy than other soft tissue sarcomas.

Gastrointestinal stromal cell tumors (GISTs), previously classified as gastrointestinal leiomyosarcomas, are now recognized as a distinct entity within soft tissue sarcomas. Its cell of origin resembles the interstitial cell of Cajal, which controls peristalsis. The majority of malignant GISTs have activating mutations of the c-*kit* gene that result in ligand-independent phosphorylation and activation of the KIT receptor tyrosine kinase, leading to tumorigenesis.

Diagnosis

The most common presentation is an asymptomatic mass. Mechanical symptoms referable to pressure, traction, or entrapment of nerves or muscles may be present. All new and persistent or growing masses should be biopsied, either by a cutting needle (core-needle biopsy) or by a small incision, placed so that it can be encompassed in the subsequent excision without compromising a definitive resection. Lymph node metastases occur in 5%, except in synovial and epithelioid sarcomas, clear-cell sarcoma (melanoma of the soft parts), angiosarcoma, and rhabdomyosarcoma, where nodal spread may be seen in 17%. The pulmonary parenchyma is the most common site of metastases. Exceptions are GISTs, which metastasize to the liver; myxoid liposarcomas, which seek fatty tissue; and clear-cell sarcomas, which may metastasize to bones. Central nervous system metastases are rare, except in alveolar soft part sarcoma.

RADIOGRAPHIC EVALUATION

Imaging of the primary tumor is best with plain radiographs and MRI for tumors of the extremities or head and neck and by CT for tumors of the chest, abdomen, or retroperitoneal cavity. A radiograph and CT scan of the chest are important for the detection of lung metastases. Other imaging studies may be indicated, depending on the symptoms, signs, or histology.

Staging and Prognosis

The histologic grade, relationship to fascial planes, and size of the primary tumor are the most important prognostic factors. The current American Joint Commission on Cancer (AJCC) staging system is shown in Table 98–1. Prognosis is related to the stage. Cure is common in the absence of metastatic disease, but a small number of patients with metastases can also be cured. Most patients with stage IV disease die within 12 months, but some patients may live with slowly progressive disease for many years.

Table 98–1 AJCC Staging System for Sarcomas

Histologic Grade (G)	Tumor Size (T)	Node Status (N)	Metastases (M)
Well differentiated (G1)	≤5 cm (T1)	Not involved (N0)	Absent (M0)
Moderately differentiated (G2)	>5 cm (T2)	Involved (N1)	Present (M1)
Poorly differentiated (G3)	Superficial fascial involvement (Ta)		
Undifferentiated (G4)	Deep fascial involvement (Tb)		
Disease Stage		**5-Year Survival, %**	
Stage I		98.8	
A: G1,2; T1a,b; N0; M0			
B: G1,2; T2a; N0; M0			
Stage II A: G1,2; T2b; N0; M0		81.8	
B: G3,4; T1; N0; M0			
C: G3,4; T2a; N0; M0			
Stage III G3,4; T2b; N0; M0		51.7	
Stage IV		<20	
A: any G; any T; N1; M0			
B: any G; any T; any N; M1			

Treatment: Soft Tissue Sarcomas

AJCC stage I patients are adequately treated with surgery alone. Stage II patients are considered for adjuvant radiation therapy. Stage III patients may benefit from adjuvant chemotherapy. Stage IV patients are managed primarily with chemotherapy, with or without other modalities.

SURGERY

Soft tissue sarcomas tend to grow along fascial planes, with the surrounding soft tissues compressed to form a pseudocapsule that gives the sarcoma the appearance of a well-encapsulated lesion. This is invariably deceptive because "shelling out," or marginal excision, of such lesions results in a 50–90% probability of local recurrence. Wide excision with a negative margin, incorporating the biopsy site, is the standard surgical procedure for local disease. The adjuvant use of radiation therapy and/or chemotherapy improves the local control rate and permits the use of limb-sparing surgery with a local control rate (85–90%) comparable to that achieved by radical excisions and amputations. Limb-sparing approaches are indicated except when negative margins are not obtainable, when the risks of radiation are prohibitive, or when neurovascular structures are involved so that resection will result in serious functional consequences to the limb.

RADIATION THERAPY

External beam radiation therapy is an adjuvant to limb-sparing surgery for improved local control. Preoperative radiation therapy allows the use of smaller fields and smaller doses but results in a higher rate of wound complications. Postoperative radiation therapy must be given to larger fields, as the entire surgical bed must be encompassed, and in higher doses to compensate for hypoxia in the operated field. This results in a higher rate of late complications. Brachytherapy or interstitial therapy, in which the radiation source is inserted into the tumor bed, is comparable in efficacy (except in low-grade lesions), less time-consuming, and less expensive.

ADJUVANT CHEMOTHERAPY

Chemotherapy is the mainstay of treatment for Ewing's primitive neuroectodermal tumors (PNET) and rhabdomyosarcomas. Meta-analysis of 14 randomized trials revealed a significant improvement in local control and disease-free survival in favor of doxorubicin-based chemotherapy. Overall survival improvement was 4% for all sites and 7% for the extremity site. An updated meta-analysis including four additional trials with doxorubicin and ifosfamide combination has reported a statistically significant 6% survival advantage in favor of chemotherapy. A chemotherapy regimen including an anthracycline and ifosfamide with growth factor support improved overall survival by 19% for high-risk (high-grade, ≥ 5 cm primary, or locally recurrent) extremity soft tissue sarcomas.

ADVANCED DISEASE

Metastatic soft tissue sarcomas are largely incurable, but up to 20% of patients who achieve a complete response become long-term survivors. The therapeutic intent, therefore, is to produce a complete remission with chemotherapy (<10%) and/or surgery (30–40%). Surgical resection of metastases, whenever possible, is an integral part of the management. Some patients benefit from

repeated surgical excision of metastases. The two most active chemotherapeutic agents are doxorubicin and ifosfamide. These drugs show a steep dose-response relationship in sarcomas. Gemcitabine with or without docetaxel has become an established second-line regimen and is particularly active in patients with leiomyosarcomas. Dacarbazine also has some modest activity. Taxanes have selective activity in angiosarcomas, and vincristine, etoposide, and irinotecan are effective in rhabdomyosarcomas and Ewing's sarcomas. Imatinib targets the KIT and platelet-derived growth factor (PDGF) tyrosine kinase activity and is standard therapy for advanced/metastatic GISTs and dermatofibrosarcoma protuberans. Imatinib is now also indicated as adjuvant therapy for completely resected primary GISTs.

BONE SARCOMAS

Incidence and Epidemiology

Bone sarcomas are rarer than soft tissue sarcomas; they accounted for only 0.2% of all new malignancies and 2600 new cases in the United States in 2010. Several benign bone lesions have the potential for malignant transformation. Enchondromas and osteochondromas can transform into chondrosarcoma; fibrous dysplasia, bone infarcts, and Paget's disease of bone can transform into either malignant fibrous histiocytoma or osteosarcoma.

Classification

BENIGN TUMORS

The common benign bone tumors include enchondroma, osteochondroma, chondroblastoma, and chondromyxoid fibroma, of cartilage origin; osteoid osteoma and osteoblastoma, of bone origin; fibroma and desmoplastic fibroma, of fibrous tissue origin; hemangioma, of vascular origin; and giant cell tumor, of unknown origin.

MALIGNANT TUMORS

The most common malignant tumors of bone are plasma cell tumors (Chap. 111). The four most common malignant nonhematopoietic bone tumors are osteosarcoma, chondrosarcoma, Ewing's sarcoma, and malignant fibrous histiocytoma. Rare malignant tumors include chordoma (of notochordal origin), malignant giant cell tumor and adamantinoma (of unknown origin), and hemangioendothelioma (of vascular origin).

MUSCULOSKELETAL TUMOR SOCIETY STAGING SYSTEM

Sarcomas of bone are staged according to the Musculoskeletal Tumor Society staging system based on grade and compartmental localization. A Roman numeral reflects the tumor grade: stage I is low grade, stage II is high grade, and stage III includes tumors of any grade that have lymph node or distant metastases. In addition, the tumor is given a letter reflecting its compartmental localization. Tumors designated A are intracompartmental (i.e., confined to the same soft tissue compartment as the initial tumor), and tumors designated B are extracompartmental (i.e., extending into the adjacent soft tissue compartment or into bone). The tumor node metastasis (TNM) staging system is shown in Table 98–2.

Table 98–2 Staging System for Bone Sarcomas

Primary tumor (T)	TX	Primary tumor cannot be assessed
	T0	No evidence of primary tumor
	T1	Tumor ≤8 cm in greatest dimension
	T2	Tumor >8 cm in greatest dimension
	T3	Discontinuous tumors in the primary bone site
Regional lymph nodes (N)	NX	Regional lymph nodes cannot be assessed
	N0	No regional lymph node metastasis
	N1	Regional lymph node metastasis
Distant metastasis (M)	MX	Distant metastasis cannot be assessed
	M0	No distant metastasis
	M1	Distant metastasis
	M1a	Lung
	M1b	Other distant sites
Histologic grade (G)	GX	Grade cannot be assessed
	G1	Well differentiated—low grade
	G2	Moderately differentiated—low grade
	G3	Poorly differentiated—high grade
	G4	Undifferentiated—high grade (Ewing's is always classed G4)

Stage Grouping				
Stage IA	T1	N0	M0	G1,2 low grade
Stage IB	T2	N0	M0	G1,2 low grade
Stage IIA	T1	N0	M0	G3,4 high grade
Stage IIB	T2	N0	M0	G3,4 high grade
Stage III	T3	N0	M0	Any G
Stage IVA	Any T	N0	M1a	Any G
Stage IVB	Any T	N1	Any M	Any G
	Any T	Any N	M1b	Any G

Osteosarcoma

Osteosarcoma, accounting for almost 45% of all bone sarcomas, is a spindle cell neoplasm that produces osteoid (unmineralized bone) or bone. Approximately 60% of all osteosarcomas occur in children and adolescents in the second decade of life, and approximately 10% occur in the third decade of life. Osteosarcomas in the fifth and sixth decades of life are frequently secondary to either radiation therapy or transformation in a preexisting benign condition, such as Paget's disease. Males are affected 1.5–2 times as often as females. Osteosarcoma has a predilection for metaphyses of long bones; the most common sites of involvement are the distal femur, proximal tibia, and proximal humerus. The classification of osteosarcoma is complex, but 75% of osteosarcomas fall into the "classic" category, which include osteoblastic, chondroblastic, and fibroblastic osteosarcomas. The remaining 25% are classified as "variants" on the basis of (1) clinical characteristics, as in the case of osteosarcoma of the jaw, postradiation osteosarcoma, or Paget's osteosarcoma; (2) morphologic characteristics, as in the case of telangiectatic osteosarcoma, small cell osteosarcoma, or epithelioid osteosarcoma; or (3) location, as in parosteal or periosteal osteosarcoma. Diagnosis usually requires a synthesis of clinical, radiologic, and pathologic features. Patients typically present with pain and swelling of the affected area. A plain radiograph reveals a destructive lesion with a moth-eaten appearance, a spiculated periosteal reaction (sunburst appearance), and a cuff of periosteal new bone formation at the margin of the soft tissue mass (Codman's triangle). A CT scan of the primary tumor is best for defining bone destruction and the pattern of calcification, whereas MRI is better for defining intramedullary and soft tissue extension. A chest radiograph and CT scan are used to detect lung metastases. Metastases to the bony skeleton should be imaged by a bone scan, or by fluorodeoxyglucose positron emission tomography (FDG-PET). Almost all osteosarcomas are hypervascular. Angiography is not helpful for diagnosis, but it is the most sensitive test for assessing the response to preoperative chemotherapy. Pathologic diagnosis is established either with a core-needle biopsy, where feasible, or with an open biopsy with an appropriately placed incision that does not compromise future limb-sparing resection. Most osteosarcomas are high-grade. The most important prognostic factor for long-term survival is response to chemotherapy. Preoperative chemotherapy followed by limb-sparing surgery (which can be accomplished in >80% of patients) followed by postoperative chemotherapy is standard management. The effective drugs are doxorubicin, ifosfamide, cisplatin, and high-dose methotrexate with leucovorin rescue. The various combinations of these agents that have been used have all been about equally successful. Long-term survival rates in extremity osteosarcoma range from 60 to 80%. Osteosarcoma is radioresistant; radiation therapy has no role in the routine management. Malignant fibrous histiocytoma is considered a part of the spectrum of osteosarcoma and is managed similarly.

Chondrosarcoma

Chondrosarcoma, which constitutes ~20–25% of all bone sarcomas, is a tumor of adulthood and old age with a peak incidence in the fourth to sixth decades of life. It has a predilection for the flat bones, especially the shoulder and pelvic girdles, but can also affect the diaphyseal portions of long bones. Chondrosarcomas can arise de novo or as a malignant transformation of an enchondroma or, rarely, of the cartilaginous cap of an osteochondroma. Chondrosarcomas have an indolent natural history and typically present as pain and swelling. Radiographically, the lesion may have a lobular appearance with mottled or punctate or annular calcification of the cartilaginous matrix. It is difficult to distinguish low-grade chondrosarcoma from benign lesions by x-ray or histologic examination. The diagnosis is therefore influenced by clinical history and physical examination. A new onset of pain, signs of inflammation, and progressive increase in the size of the mass suggest malignancy. The histologic classification is complex, but most tumors fall within the classic category. Like other bone sarcomas, high-grade chondrosarcomas spread to the lungs. Most chondrosarcomas are resistant to chemotherapy, and surgical resection of primary or recurrent tumors, including pulmonary metastases, is the mainstay of therapy. This rule does not hold for two histologic variants. Dedifferentiated chondrosarcoma has a high-grade osteosarcoma or a malignant fibrous histiocytoma component that responds to chemotherapy. Mesenchymal chondrosarcoma, a rare variant composed of a small cell element, also is responsive to systemic chemotherapy and is treated like Ewing's sarcoma.

Ewing's Sarcoma

Ewing's sarcoma, which constitutes ~10–15% of all bone sarcomas, is common in adolescence and has a peak incidence in the second decade of life. It typically involves the diaphyseal region of long bones and also has an affinity for flat bones. The plain radiograph may show a characteristic "onion peel" periosteal reaction with a generous soft tissue mass, which is better demonstrated by CT or MRI. This mass is composed of sheets of monotonous, small, round, blue cells and can be confused with lymphoma, embryonal rhabdomyosarcoma, and small-cell carcinoma. The presence of p30/32, the product of the *mic-2* gene (which maps to the pseudoautosomal region of the X and Y chromosomes) is a cell-surface marker for Ewing's sarcoma (and other members of the Ewing's family of tumors, sometimes called PNETs). Most PNETs arise in soft tissues; they include peripheral neuroepithelioma, Askin's tumor (chest wall), and esthesioneuroblastoma. Glycogen-filled cytoplasm detected by staining with periodic acid–Schiff is also characteristic of Ewing's sarcoma cells. The classic cytogenetic abnormality associated with this disease (and other PNETs) is a reciprocal translocation of the long arms of chromosomes 11 and 22, t(11;22), which creates a chimeric gene product of unknown function with components from the *fli-1* gene on chromosome 11 and *ews* on 22. This disease is very aggressive, and it is therefore considered a systemic disease. Common sites of metastases are lung, bones, and bone marrow. Systemic chemotherapy is the mainstay of therapy, often being used before surgery. Doxorubicin, cyclophosphamide or ifosfamide, etoposide, vincristine, and dactinomycin are active drugs. Topotecan or irinotecan in combination with an alkylating agent are often used in relapsed patients. Targeted therapy with an anti IGF1r antibody appears to have promising activity in refractory cases. Local treatment for the primary tumor includes surgical resection, usually with limb salvage or radiation therapy. Patients with lesions below the elbow and below the mid-calf have a 5-year survival rate of 80% with effective treatment. Ewing's sarcoma at first presentation is a curable tumor, even in the presence of obvious metastatic disease, especially in children <11 years old.

TUMORS METASTATIC TO BONE

Bone is a common site of metastasis for carcinomas of the prostate, breast, lung, kidney, bladder, and thyroid and for lymphomas and sarcomas. Prostate, breast, and lung primaries account for 80% of all bone metastases. Metastatic tumors of bone are more common than primary bone tumors. Tumors usually spread to bone hematogenously, but local invasion from soft tissue masses also occurs. In descending order of frequency, the sites most often involved are the vertebrae, proximal femur, pelvis, ribs, sternum, proximal humerus, and skull. Bone metastases may be asymptomatic or may produce pain, swelling, nerve root or spinal cord compression, pathologic fracture, or myelophthisis (replacement of the marrow). Symptoms of hypercalcemia may be noted in cases of bony destruction.

Pain is the most frequent symptom. It usually develops gradually over weeks, is usually localized, and often is more severe at night. When patients with back pain develop neurologic signs or symptoms, emergency evaluation for spinal cord compression is indicated (Chap. 276). Bone metastases exert a major adverse effect on quality of life in cancer patients.

Cancer in the bone may produce osteolysis, osteogenesis, or both. Osteolytic lesions result when the tumor produces substances that can directly elicit bone resorption (vitamin D–like steroids, prostaglandins, or parathyroid hormone–related peptide) or cytokines that can induce the formation of osteoclasts (interleukin 1 and tumor necrosis factor). Osteoblastic lesions result when the tumor produces cytokines that activate osteoblasts. In general, purely osteolytic lesions are best detected by plain radiography, but they may not be apparent until they are >1 cm. These lesions are more commonly associated with hypercalcemia and with the excretion of hydroxyproline-containing peptides indicative of matrix destruction. When osteoblastic activity is prominent, the lesions may be readily detected using radionuclide bone scanning (which is sensitive to new bone formation), and the radiographic appearance may show increased bone density or sclerosis. Osteoblastic lesions are associated with higher serum levels of alkaline phosphatase and, if extensive, may produce hypocalcemia. Although some tumors may produce mainly osteolytic lesions (e.g., kidney cancer) and others mainly osteoblastic lesions (e.g., prostate cancer), most metastatic lesions produce both types of lesion and may go through stages where one or the other predominates.

In older patients, particularly women, it may be necessary to distinguish metastatic disease of the spine from osteoporosis. In osteoporosis, the cortical bone may be preserved, whereas cortical bone destruction is usually noted with metastatic cancer.

Treatment: Metastatic Bone Disease

Treatment of metastatic bone disease depends on the underlying malignancy and the symptoms. Some metastatic bone tumors are curable (lymphoma, Hodgkin's disease), and others are treated with palliative intent. Pain may be relieved by local radiation therapy. Hormonally responsive tumors are responsive to hormone inhibition (antiandrogens for prostate cancer, antiestrogens for breast cancer). Strontium 89 and samarium 153 are bone-seeking radionuclides that can exert antitumor effects and relieve symptoms. Bisphosphonates such as pamidronate may relieve pain and inhibit bone resorption, thereby maintaining bone mineral density and reducing risk of fractures in patients with osteolytic metastases from breast cancer and multiple myeloma. Careful monitoring of serum electrolytes and

creatinine is recommended. Monthly administration prevents bone-related clinical events and may reduce the incidence of bone metastases in women with breast cancer. When the integrity of a weight-bearing bone is threatened by an expanding metastatic lesion that is refractory to radiation therapy, prophylactic internal fixation is indicated. Overall survival is related to the prognosis of the underlying tumor. Bone pain at the end of life is particularly common; an adequate pain relief regimen including sufficient amounts of narcotic analgesics is required. The management of hypercalcemia is discussed in Chap. 353.

FURTHER READINGS

Borden EC et al: Soft tissue sarcomas of adults: State of the translational science. Clin Canc Res 9:1941, 2003[PMID: 12796356]

Dematteo R et al: Adjuvant imatinib mesylate after resection of localized primary GIST: A

randomized double blind placebo controlled trial. Lancet 373:1097, 2009

Helman Lj, Meltzer P: Mechanisms of sarcoma development. Nat Rev Canc 3:685, 2003[PMID: 12951587]

Mocellin S et al: Adult soft tissue sarcomas: Conventional therapies and molecularly targeted approaches. Canc Treat Rev 32:9, 2006[PMID: 16338075]

Patel SR, Zalcberg J. Optimization of dose of Imatinib for the treatment of GIST—Lessons learned from the phase 3 trials. Eur J Cancer 44:501, 2008[PMID: 18234488]

Pisters PW et al: Evidence-based recommendations for local therapy for soft tissue sarcomas. J Clin Oncol 25:1003, 2007[PMID: 17350950]

Scurr M, Judson I: How to treat the Ewing's family of sarcomas in adult patients. Oncologist 11:65, 2006[PMID: 16401715]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CARCINOMA OF UNKNOWN PRIMARY

Carcinoma of unknown primary (CUP) is a biopsy-proven (mainly epithelial) malignancy for which the anatomic site of origin remains unidentified after an intensive search. CUP is one of the 10 most frequently diagnosed cancers worldwide, accounting for approximately 3–5% of all cancers. Most investigators do not include lymphomas, metastatic melanomas, and metastatic sarcomas that present without a known primary tumor as CUP because these cancers have specific stage- and histology-based treatments that guide management.

With the increasing availability of additional sophisticated imaging, invasive diagnostic techniques, and the emergence of effective targeted therapies in several cancers, an individualized management algorithm with an impact on quality of life and survival is critical. The reasons cancers present as CUP remain unclear. One hypothesis is that the primary tumor either regresses after seeding the metastasis or remains so small that it is not detected. It is possible that CUP falls on the continuum of cancer presentation where the primary has been contained or eliminated by the natural body defenses. Alternatively, CUP may represent a specific malignant event that results in an increase in metastatic spread or survival relative to the primary. Whether the CUP metastases truly define a clone that is genetically and phenotypically unique to this diagnosis remains to be determined.

CUP BIOLOGY

No characteristics that are unique to CUP relative to metastases from known primaries have been identified. Abnormalities in chromosomes 1 and 12 and other complex cytogenetic abnormalities have been reported. Aneuploidy has been described in 70% of CUP patients with metastatic adenocarcinoma or undifferentiated carcinoma. The overexpression of various genes, including *Ras, bcl-2* (40%), *her-2* (11%), and *p53* (26–53%), has been studied in CUP samples, but they have no effect on response to therapy or survival. The extent of angiogenesis in CUP relative to that in metastases from known primaries has also been evaluated, but no consistent findings have emerged.

Clinical Evaluation

Obtaining a thorough medical history from CUP patients is essential, paying particular attention to previous surgeries, removed lesions, and family medical history to assess potential hereditary cancers. Physical examination, including a digital rectal examination in men and breast and pelvic examinations in women, should be performed. Determining the patient's performance status, nutritional status, comorbid illnesses, and cancer-induced complications is essential since these may affect treatment planning.

ROLE OF SERUM TUMOR MARKERS AND CYTOGENETICS

Most tumor markers, including CEA, CA-125, CA 19-9, and CA 15-3, when elevated, are nonspecific and not helpful in determining the primary tumor site. Men who present with adenocarcinoma and osteoblastic metastasis should undergo a prostate-specific antigen (PSA) test. In patients with undifferentiated or poorly differentiated carcinoma (especially with a midline tumor), elevated β-human chorionic gonadotropin (βhCG) and α fetoprotein (AFP) levels suggest the possibility of an extragonadal germ cell (testicular) tumor. Cytogenetic studies had a larger role in the past, although interpretation of these older studies can be challenging. In our opinion, with the availability of immunohistochemical stains, cytogenetic analyses are indicated only occasionally. We reserve them for undifferentiated neoplasms with inconclusive immunohistochemical stains and those for which a high suspicion of lymphoma exists.

ROLE OF IMAGING STUDIES

Chest x-rays are always obtained in CUP workups but are often negative, especially with low-volume disease. A CT scan of the chest, abdomen, and pelvis is indicated in the search for the primary, evaluate the extent of disease, and select the most favorable biopsy site. Older studies suggested that the primary tumor site is detected in 20–35% of patients who undergo a CT scan of the abdomen and pelvis, although by current definition these patients would not be considered as having CUP. Older studies also suggest a latent primary tumor prevalence of 20%; with more sophisticated imaging, this prevalence is <5% today.

Mammography should be performed in all women who present with metastatic adenocarcinoma, especially in those with adenocarcinoma and isolated axillary lymphadenopathy. MRI of the breast is a recognized follow-up modality in patients with suspected occult primary breast carcinoma following a negative mammography and sonography. The results of these imaging modalities can influence surgical management; a negative breast MRI result predicts a low tumor yield at mastectomy.

A conventional workup for a squamous cell carcinoma and cervical CUP (neck lymphadenopathy with no known primary tumor) includes a CT scan or MRI and invasive studies, including indirect and direct laryngoscopy, bronchoscopy, and upper endoscopy. Ipsilateral (or bilateral) tonsillectomy (with histopathology) has been recommended for these patients. Fluorodeoxyglucose positron emission tomography (FDG-PET) scans are useful in this patient population and may help guide the biopsy; determine the extent of disease; facilitate the appropriate treatment, including planning radiation fields; and help with disease surveillance. A smaller radiation field encompassing the primary (when found) and metastatic adenopathy decreases the risk of chronic xerostomia. Several studies have evaluated the utility of PET in patients with cervical CUP. These trials have included a small number of patients; primary tumors were identified in ~21–30%.

The diagnostic contribution of PET to the evaluation of other CUP (outside of the neck adenopathy indication) is controversial. PET-CT can be helpful for patients who are candidates for surgical intervention for solitary metastatic disease because the presence of disease outside the primary site may affect surgical planning.

Invasive studies, including upper endoscopy, colonoscopy, and bronchoscopy, should be limited to symptomatic patients or those with laboratory, imaging or pathologic abnormalities that suggest that these techniques will result in a high yield in search for a primary cancer.

Pathologic Diagnosis of Cup

A detailed pathologic examination of the most accessible biopsied tissue specimen is mandatory in CUP patients. Pathologic evaluation typically consists of hematoxylin-and-eosin stains and immunohistochemical tests. Electron microscopy and cytogenetics are rarely useful.

LIGHT MICROSCOPY EVALUATION

Adequate tissue obtained by fine-needle aspiration or core-needle biopsy should first be stained with hematoxylin and eosin and subjected to light microscopic examination. On light microscopy, 60–65% of CUP are adenocarcinoma, and 5% are squamous cell carcinoma. The remaining 30–35% are poorly differentiated adenocarcinoma, poorly differentiated carcinoma, poorly differentiated neoplasm. A small percentage of lesions are diagnosed as neuroendocrine cancers (2%), mixed tumors (adenosquamous, or sarcomatoid carcinomas), or undifferentiated neoplasms (Table 99–1).

Table 99–1 Major Histologies in Cup

Histology	Proportion, %
Well to moderately differentiated adenocarcinoma	60
Squamous cell cancer	5
Poorly differentiated adenocarcinoma, poorly differentiated carcinoma	30
Neuroendocrine	2
Undifferentiated malignancy	3

ROLE OF IMMUNOHISTOCHEMICAL ANALYSIS

Immunohistochemical stains are peroxidase-labeled antibodies against specific tumor antigens that are used to define tumor lineage. The number of available immunohistochemical stains is ever-increasing. However, in CUP cases, more is not necessarily better, and immunohistochemical stains should be used in conjunction with the patient's clinical presentation and imaging studies to select the best therapy. Communication between the clinician and pathologist is essential. No stain is 100% specific, and overinterpretation should be avoided. PSA and thyroglobulin tissue markers, which are positive in prostate and thyroid cancer, respectively, are the most specific of the current marker panel. However, these cancers rarely present as CUP, so the yield of these tests may be low. Fig. 99-1 delineates a simple algorithm for immunohistochemical staining in CUP cases. Table 99–2 lists additional tests that may be useful to further define the tumor lineage. A more comprehensive algorithm may improve the diagnostic accuracy but can make the process complex. With the use of immunohistochemical markers, electron microscopic analysis, which is time-consuming and expensive, is rarely needed.

Figure 99-1

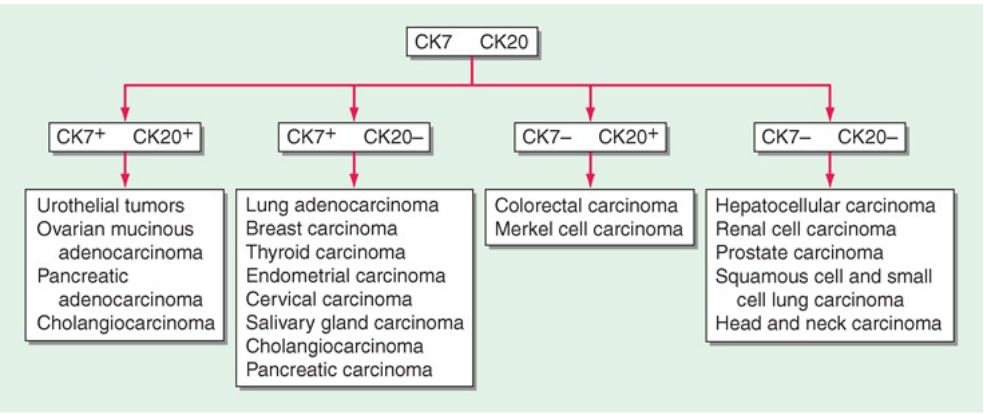

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to cytokeratin (CK7 and CK20) markers used in CUP.

Figure 99-2

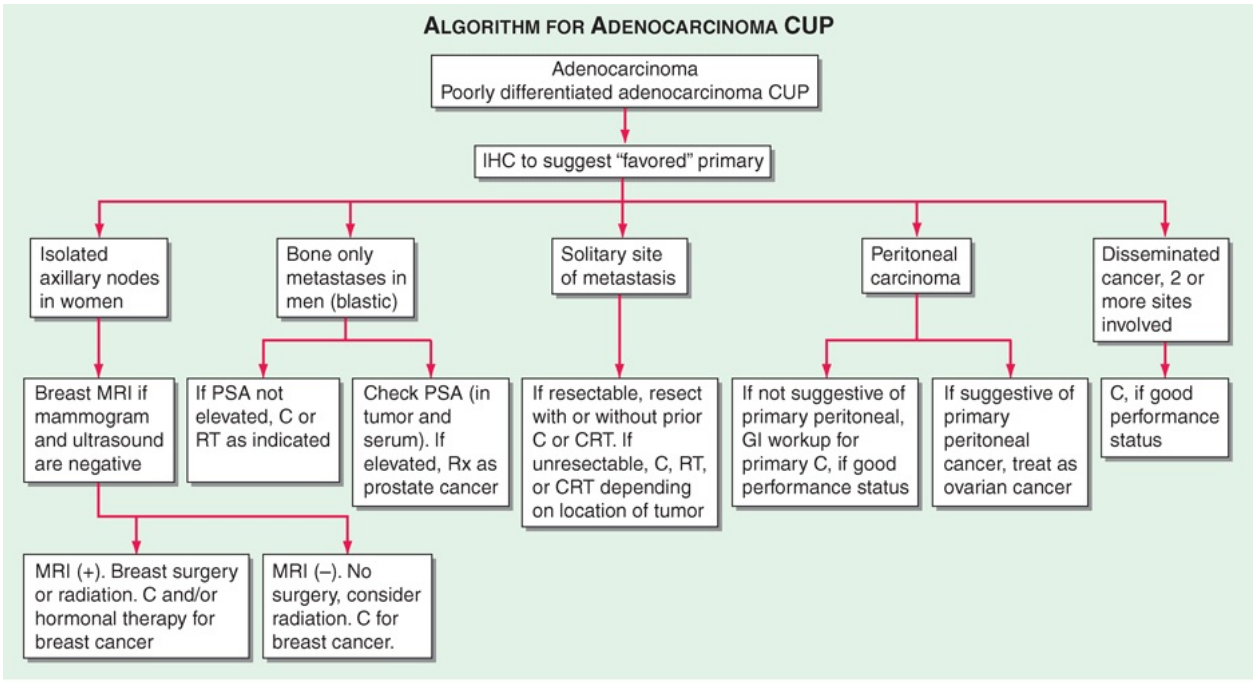

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment algorithm for adenocarcinoma and poorly differentiated adenocarcinoma CUP. C, chemotherapy; CRT, chemoradiation; GI, gastrointestinal; IHC, immunohistochemistry; MRI, magnetic resonance imaging; PSA, prostate-specific antigen; RT, radiation.

Figure 99-3

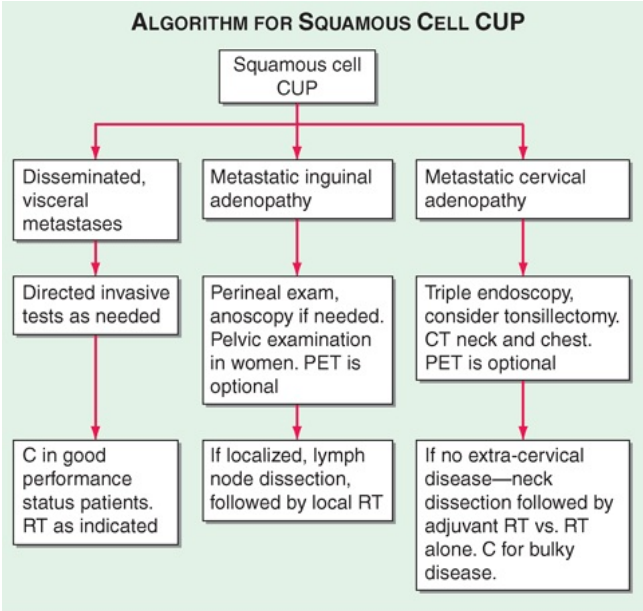

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment algorithm for squamous cell CUP. C, chemotherapy; CT, computed tomography; PET, positron emission tomography; RT, radiation.

Table 99–2 Additional Immunohistochemical Stains Useful in the Diagnosis of Cup

Tissue Marker	Diagnosis
Estrogen and progesterone receptors	Breast cancer
BRST-1	Breast cancer
Gross cystic disease fibrous protein-15	Breast cancer
Thyroid transcription factor 1	Lung and thyroid cancer
Thyroglobulin	Thyroid cancer

Chromogranin, synaptophysin, CD56	Neuroendocrine cancer
CDX-2	Gastrointestinal cancer
Calretinin, mesothelin	Mesothelioma
Leukocyte common antigen	Lymphoma
HMB-45, tyrosinase, Melan-A	Melanoma
URO-III, thrombomodulin	Bladder cancer
α Fetoprotein	Hepatocellular cancer, germ cell cancer
β-Human chronic gonadotropin	Germ cell cancer
Prostate specific antigen	Prostate cancer
WT-1, estrogen receptor (ER)	Müllerian/ovarian cancer
RCC, CD 10	Renal cell carcinoma

There are >20 subtypes of cytokeratin (CK) intermediate filaments with different molecular weights and differential expression in various cell types and cancers. Monoclonal antibodies to specific CK subtypes have been used to help classify tumors according to their site of origin; commonly used CK stains in CUP are CK7 and CK20. CK7 is found in tumors of the lung, ovary, endometrium, and breast and not in those of the lower gastrointestinal tract, whereas CK20 is normally expressed in the gastrointestinal epithelium, urothelium, and Merkel cells. CK20+/CK7– strongly suggests a primary tumor of the colon; 75–95% of colon tumors show this pattern of staining. CK20–/CK7+ suggests cancer of the lung, breast, ovary, endometrium, and pancreaticobiliary tract; some of these can also be CK20+. The nuclear CDX-2 transcription factor, which is the product of a homeobox gene necessary for intestinal organogenesis, is often used to aid in the diagnosis of gastrointestinal adenocarcinomas.

Thyroid transcription factor 1 (TTF-1) is a 38-kDa homeodomain-containing nuclear protein that plays a role in transcriptional activation during embryogenesis in the thyroid, diencephalon, and respiratory epithelium. TTF-1 nuclear staining is typically positive in lung and thyroid cancers. Approximately 68% of adenocarcinomas and 25% of squamous cell lung cancers stain positive for TTF-1, which helps differentiate a lung primary tumor from metastatic adenocarcinoma in a pleural effusion, the mediastinum, or the lung parenchyma.

Distinguishing pleural mesothelioma from lung adenocarcinoma can be challenging. Calretinin, Wilms' tumor gene-1 (WT-1), and mesothelin have been suggested as useful markers for mesothelioma. Gross cystic disease fibrous protein-15, a 15-kDa monomer protein, is a marker of apocrine differentiation that is detected in 62–72% of breast carcinomas. UROIII, high-molecular-weight cytokeratin, thrombomodulin, and CK20 are the markers used to diagnose lesions of urothelial origin.

Role of DNA Microarray and Reverse Transcriptase Polymerase Chain Reaction (Rt-PCR) in Cup

In the absence of a known primary, developing therapeutic strategies for CUP is challenging. The current diagnostic yield with imaging and immunochemistry is ~20–30% for CUP patients. The use of gene expression studies holds the promise of substantially increasing this yield. Gene expression profiles are most commonly generated using quantitative RT-PCR or DNA microarray.

Neural network programs have been used to develop predictive algorithms from the gene expression profiles. Typically, a training set of gene profiles from known cancers (preferably from metastatic sites) are used to train the software. The program can then be used to predict the putative origin of a test tumor, and presumably of true CUP. Comprehensive gene expression databases that have become available for common malignancies may also be useful in CUP. Investigators have used expression data from normal differentiated tissues to identify conserved expression profiles found in malignant tissue as a basis for predicting the tissue of origin (ToO). These approaches have been effective in blind testing against known primary cancers and their metastasis. However, because, by definition, the primary tumor site is not identifiable in CUP, validation of site prediction in this setting can be challenging, and any predictions currently must be supported by clinical and pathologic correlation. Prospective validation trials are currently evaluating the role of molecular studies identifying ToO in CUP and its impact on management. Early trials suggest that the profiling approach is feasible from archived formalin-fixed paraffin-embedded (FFPE) core-needle biopsies. Quantitative RT-PCR on fine-needle aspiration samples is very useful in clinical practice. Data from some studies suggest that a putative primary profile can be applied to 80–85% of cases. At present, the best confirmation of molecular profiling studies in CUP is an indirect validation based on the patient's presentation and clinical course. Current studies are geared to understanding profiling tools' accuracy, clinical effectiveness, and how these assays complement immunohistochemistry and help guide therapy.

Treatment: Carcinoma of Unknown Primary

General Considerations The treatment of CUP continues to evolve, albeit slowly. The median survival duration of most patients with disseminated CUP is ~6–10 months. Systemic chemotherapy is the primary treatment modality in most cases, but the careful integration of surgery, radiation therapy, and even periods of observation are important in the overall management of this condition (Figs. 99-2 and 99-3). Prognostic factors include performance status, site and number of metastases, response to chemotherapy, and serum lactate dehydrogenase (LDH) levels. Culine and colleagues developed a prognostic model using performance status and serum LDH levels, which allowed the assignment of patients into two subgroups with divergent outcomes. Future prospective trials using this prognostic model are warranted. Clinically, several CUP diagnoses fall into a favorable prognostic subset. Others, including those with disseminated CUP that do not fit a subset, have a more unfavorable prognosis.

TREATMENT OF FAVORABLE SUBSETS OF CUP

Women with isolated axillary adenopathy Women with isolated axillary adenopathy with adenocarcinoma or carcinoma are usually treated for stage II or III breast cancer based on pathologic findings. These patients should undergo a breast MRI if mammogram and ultrasound are negative. Radiation therapy to the ipsilateral breast is indicated if the breast MRI is positive. Chemotherapy and/or hormonal therapy is indicated based on patient's age (premenopausal or postmenopausal), nodal disease bulk, and hormone receptor status (Chap. 90). It is important to verify that the pathology does represent a breast cancer profile (morphology, immunohistochemical markers including HER-2, gene expression) before embarking on a breast cancer therapeutic program.

Women with peritoneal carcinomatosis The term *primary peritoneal papillary serous carcinoma* (PPSC) has been used to describe CUP with carcinomatosis with the pathologic and laboratory (elevated CA-125 antigen) characteristics of ovarian cancer but no ovarian primary tumor identified on transvaginal sonography or laparotomy. Studies suggest that ovarian cancer and PPSC, which are both of müllerian origin, have similar gene expression profiles. Similar to patients with ovarian cancer, patients with PPSC are candidates for cytoreductive surgery, followed by adjuvant taxane and platinum-based chemotherapy. In one retrospective study of 258 women with peritoneal carcinomatosis who had undergone cytoreductive surgery and chemotherapy, 22% of patients had a complete response to chemotherapy; the median survival duration was 18 months (range 11–24 months). However, not all peritoneal carcinomatosis in women is PPSC. Careful pathologic evaluation can help diagnose a colon cancer profile (CDX-2+, CK-20+, CK7–) or a pancreaticobiliary cancer.

Poorly differentiated carcinoma with midline adenopathy Men with poorly differentiated or undifferentiated carcinoma that presents as a midline adenopathy should be evaluated for extragonadal germ cell malignancy. If diagnosed and treated as such, they often experience a good response to treatment with platinum-based combination chemotherapy. Response rates of >50% have been noted, and 10–15% long-term survivors have been reported. Older patients (especially smokers) who present with mediastinal adenopathy are more likely to have a lung or head—and-neck cancer profile.

Neuroendocrine carcinoma Low-grade neuroendocrine carcinoma often has an indolent course, and treatment decisions are based on symptoms and tumor bulk. Urine 5-HIAA and serum chromogranin may be elevated and can be followed as markers. Often the patient is treated with somatostatin analogues alone for hormone-related symptoms (diarrhea, flushing, nausea). Specific local therapies or systemic therapy would only be indicated if the patient is symptomatic with local pain secondary to significant growth of the metastasis or the hormone-related symptoms are not controlled with endocrine therapy. Patients with high-grade neuroendocrine carcinoma are treated as having small cell lung cancer and are responsive to chemotherapy; 20–25% show a complete response, and up to 10% patients survive more than 5 years.

Squamous cell carcinoma presenting as neck adenopathy Patients with early-stage squamous cell carcinoma involving the cervical lymph nodes are candidates for node dissection and radiation therapy, which can result in long-term survival. The role of chemotherapy in these patients is undefined, although chemoradiation therapy or induction chemotherapy is often used and is beneficial in bulky N2/N3 lymph node disease.

Solitary metastatic site Patients with solitary metastases can also experience good treatment outcomes. Some patients who present with locoregional disease are candidates for aggressive trimodality management; both prolonged disease-free interval and occasionally cure are possible.

Men with blastic skeletal metastases and elevated PSA Blastic bone-only metastasis is a rare presentation, and elevated serum PSA or tumor staining with PSA may provide confirmatory evidence of prostate cancer in these patients. Those with elevated levels are candidates for hormonal therapy for prostate cancer, although it is important to rule out other primary tumors (lung most common).

Management of Disseminated CUP Patients who present with liver, brain, and adrenal metastatic disease usually have a poor prognosis. Beside primary peritoneal carcinoma, carcinomatosis presenting as CUP in other settings is not uncommon. Gastric, appendicular, colon, pancreas, and cholangiocarcinoma are all possible primaries, and imaging, endoscopy, and pathologic data help in the evaluation.

Traditionally, platinum-based combination chemotherapy regimens have been used to treat patients with CUP. In a phase II study by Hainsworth and colleagues, 55 mostly chemotherapy-naive patients were treated with paclitaxel, carboplatin, and oral etoposide every 3 weeks. The overall response rate was 47%, with median overall survival duration of 13.4 months. Briasoulis and colleagues reported similar response rates and survival durations in 77 patients with CUP, who had been treated with paclitaxel and carboplatin. In this study, patients with nodal or pleural disease and women with peritoneal carcinomatosis had higher response rates and overall survival durations of 13 and 15 months, respectively. Studies incorporating newer agents, including gemcitabine, irinotecan, and targeted agents, are showing higher response rates. In a phase II randomized trial by Culine and colleagues, 80 patients were randomly assigned to receive gemcitabine with cisplatin or irinotecan with cisplatin; 78 patients were assessable for efficacy and toxicity. Objective responses were observed in 21 patients (55%) in the gemcitabine and cisplatin arm and in 15 patients (38%) in the irinotecan and cisplatin arm. The median survival was 8 months for gemcitabine and cisplatin and 6 months for irinotecan and cisplatin.

The role of second-line chemotherapy in CUP is poorly defined. Gemcitabine as a single agent has shown a partial response rate of 8%, and 25% of patients had minor responses or stable disease, with improved symptoms. Combination chemotherapy as a second- and third-line treatment may result in a slightly improved response and therapy options should be guided by pathology and the patient's performance status.

Hainsworth and colleagues studied the combination of bevacizumab and erlotinib in 51 patients; 25% were chemotherapy-naive and had advanced bone or liver metastases, while the rest had been treated with 1 or 2 chemotherapy regimens. Responses were noted in 4 patients (8%), and 30 patients (59%) experienced stable disease or a minor response. The median overall survival was 8.9 months, with 42% of patients alive at 1 year.

Historically, patients with CUP have been treated with broad spectrum regimens that work for a variety of primary cancers; a "one treatment fits all" approach. With incremental improved responses over the past decade in known cancer types, we anticipate overall better response rates with newer regimens for selected CUP patients. With a more robust immunohistochemical panel (directed approach) and new molecular profiling tools, one may hope to create a more tailored treatment algorithm for CUP patients.

SUMMARY

Patients with CUP should undergo a directed diagnostic search for the primary tumor on the basis of clinical and pathologic data. Subsets of patients have prognostically favorable disease, as defined by clinical or histologic criteria, and may substantially benefit from aggressive treatment and expect prolonged survival. However, for most patients who present with advanced CUP, the prognosis remains poor, with early resistance to available cytotoxic therapy. The current focus has shifted away from empirical chemotherapeutic trials to understanding the metastatic phenotype, ToO profiling, and evaluating molecular targets in CUP patients. Our understanding of this challenging disease is growing steadily in the era of sophisticated diagnostics and therapeutics. A strategy that integrates information from a patient's

clinical presentation including imaging and pathology with directed immunohistochemistry and molecular profiling in selected cases, will help us select the best treatment for our patients.

FURTHER READINGS

Abbruzzese JL et al: The biology of unknown primary tumors. Semin Oncol 20:238, 1993[PMID: 8503019]

Bugat R et al: Summary of the standards, options and recommendations for the management of patients with carcinoma of unknown primary site. Br J Cancer 89:S59, 2003

Culine S et al: Cisplatin in combination with either gemcitabine or irinotecan in carcinomas of unknown primary site: Results of a randomized phase II study—trial for the French Study Group on Carcinomas of Unknown Primary (GEFCAPI 01). J Clin Oncol 21:3479, 2003[PMID: 12972523]

Dennis JL, Oien KA: Hunting the primary: Novel strategies for defining the origin of tumours. J Pathol 205:236, 2005[PMID: 15641019]

Greco FA. Therapy of adenocarcinoma of unknown primary: Are we making progress? J Natl Compr Cancer Netw 2008 10:1061, 2008.

Hainsworth JD et al: Carcinoma of unknown primary site: Treatment with 1-hour paclitaxel, carboplatin, and extended-schedule etoposide. J Clin Oncol 15:2385, 1997[PMID: 9196154]

Kende AI et al: Expression of cytokeratins 7 and 20 in carcinomas of the gastrointestinal tract. Histopathology 42:137, 2003[PMID: 12558745]

Monzon FA et al: Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin. J Clin Oncol 27:2503, 2009[PMID: 19332734]

Olson JA Jr et al: Magnetic resonance imaging facilitates breast conservation for occult breast cancer. Ann Surg Oncol 7:411, 2000[PMID: 10894136]

Roh MS, Hong SH: Utility of thyroid transcription factor-1 and cytokeratin 20 in identifying the origin of metastatic carcinomas of cervical lymph nodes. J Korean Med Sci 17:512, 2002[PMID: 12172048]

Rusthoven KE et al: The role of fluorodeoxyglucose positron emission tomography in cervical lymph node metastases from an unknown primary tumor. Cancer 101:2641, 2004[PMID: 15517576]

Tothill RW et al: An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. Cancer Res 65:4031, 2005[PMID: 15899792]

Varadhachary GR et al: Molecular profiling of carcinoma of unknown primary and correlation with clinical evaluation. J Clin Oncol 26:4442, 2008[PMID: 18802157]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PARANEOPLASTIC SYNDROMES: ENDOCRINOLOGIC/HEMATOLOGIC: INTRODUCTION

In addition to local tissue invasion and metastasis, neoplastic cells can produce a variety of products that can stimulate hormonal, hematologic, dermatologic, and neurologic responses. *Paraneoplastic syndromes* is the term used to refer to the disorders that accompany benign or malignant tumors but are not directly related to mass effects or invasion. Tumors of neuroendocrine origin, such as small cell lung carcinoma (SCLC) and carcinoids, produce a wide array of peptide hormones and are common causes of paraneoplastic syndromes. However, almost every type of tumor has the potential to produce hormones or cytokines or to induce immunologic responses. Careful studies of the prevalence of paraneoplastic syndromes indicate that they are more common than is generally appreciated. The signs, symptoms, and metabolic alterations associated with paraneoplastic disorders may be overlooked in the context of a malignancy and its treatment. Consequently, atypical clinical manifestations in a patient with cancer should prompt consideration of a paraneoplastic syndrome. The most common endocrinologic and hematologic syndromes associated with underlying neoplasia will be discussed here.

ENDOCRINE PARANEOPLASTIC SYNDROMES

ETIOLOGY

Hormones can be produced from eutopic or ectopic sources. *Eutopic* refers to the expression of a hormone from its normal tissue of origin, whereas *ectopic* refers to hormone production from an atypical tissue source. For example, adrenocorticotropic hormone (ACTH) is expressed eutopically by the corticotrope cells of the anterior pituitary, but it can be expressed ectopically in SCLC. Many hormones are produced at low levels from a wide array of tissues in addition to the classic endocrine source. Thus, ectopic expression is often a quantitative change rather than an absolute change in tissue expression. Nevertheless, the term *ectopic expression* is firmly entrenched and conveys the abnormal physiology associated with hormone production by neoplastic cells. In addition to high levels of hormones, ectopic expression typically is characterized by abnormal regulation of hormone production (e.g., defective feedback control) and peptide processing (resulting in large, unprocessed precursors).

A diverse array of molecular mechanisms has been suggested to cause ectopic hormone production. In rare instances, genetic rearrangements explain aberrant hormone expression. For example, translocation of the *parathyroid hormone (PTH)* gene can result in high levels of PTH expression in tissues other than the parathyroid gland, apparently because the genetic rearrangement brings the *PTH* gene under the control of atypical regulatory elements. A related phenomenon is well documented in many forms of leukemia and lymphoma, in which somatic genetic rearrangements confer a growth advantage and alter cellular differentiation and function (Chap. 110). Although genetic rearrangements may cause selected cases of ectopic hormone production, this mechanism is probably rare, as many tumors are associated with excessive production of numerous peptides. Cellular dedifferentiation probably underlies most cases of ectopic hormone production. Many cancers are poorly differentiated, and certain tumor products, such as human chorionic gonadotropin (hCG), parathyroid hormone–related protein (PTHrP), and α fetoprotein, are characteristic of gene expression at earlier developmental stages. In contrast, the propensity of certain cancers to produce particular hormones (e.g., squamous cell carcinomas produce PTHrP) suggests that dedifferentiation is partial or that selective pathways are derepressed. These expression profiles probably reflect alterations in transcriptional repression, changes in DNA methylation, or other factors that govern cell differentiation.

In SCLC, the pathway of differentiation has been relatively well defined. The neuroendocrine phenotype is dictated in part by the basic-helix-loop-helix (bHLH) transcription factor human achaete-scute homologue 1 (hASH-1), which is expressed at abnormally high levels in SCLC associated with ectopic ACTH. The activity of hASH-1 is inhibited by hairy enhancer of split 1 (HES-1) and by Notch proteins, which also are capable of inducing growth arrest. Thus, abnormal expression of these developmental transcription factors appears to provide a link between cell proliferation and differentiation.

Ectopic hormone production would only be an epiphenomenon associated with cancer if it did not result in clinical manifestations. Excessive and unregulated production of hormones such as ACTH, PTHrP, and vasopressin can lead to substantial morbidity and complicate the cancer treatment plan. Moreover, the paraneoplastic endocrinopathies are sometimes the presenting feature of underlying

malignancy and may prompt the search for an unrecognized tumor.

A large number of paraneoplastic endocrine syndromes have been described, linking overproduction of particular hormones with specific types of tumors. However, certain recurring syndromes emerge from this group (Table 100–1). The most common paraneoplastic endocrine syndromes include hypercalcemia from overproduction of PTHrP and other factors, hyponatremia from excess vasopressin, and Cushing's syndrome from ectopic ACTH.

Table 100–1 Paraneoplastic Syndromes Caused by Ectopic Hormone Production

Paraneoplastic Syndrome	Ectopic Hormone	Typical Tumor Types[a]
Common		
Hypercalcemia of malignancy	Parathyroid hormone-related protein (PTHrP)	Squamous cell (head and neck, lung, skin), breast, genitourinary, gastrointestinal
	1,25 dihydroxyvitamin D	Lymphomas
	Parathyroid hormone (PTH) (rare)	Lung, ovary
	Prostaglandin E2 (PGE2) (rare)	Renal, lung
Syndrome of inappropriate antidiuretic hormone secretion (SIADH)	Vasopressin	Lung (squamous, small cell), gastrointestinal, genitourinary, ovary
Cushing's syndrome	Adrenocorticotropic hormone (ACTH)	Lung (small cell, bronchial carcinoid, adenocarcinoma, squamous), thymus, pancreatic islet, medullary thyroid carcinoma
	Corticotropin-releasing hormone (CRH) (rare)	Pancreatic islet, carcinoid, lung, prostate
	Ectopic expression of gastric inhibitory peptide (GIP), luteinizing hormone (LH)/human chorionic gonadotropin (hCG), other G protein–coupled receptors (rare)	Macronodular adrenal hyperplasia
Less Common		
Non-islet cell hypoglycemia	Insulin-like growth factor (IGF-II)	Mesenchymal tumors, sarcomas, adrenal, hepatic, gastrointestinal, kidney, prostate
	Insulin (rare)	Cervix (small cell carcinoma)
Male feminization	hCGb	Testis (embryonal, seminomas), germinomas, choriocarcinoma, lung, hepatic, pancreatic islet
Diarrhea or intestinal hypermotility	Calcitoninc	Lung, colon, breast, medullary thyroid carcinoma
	Vasoactive intestinal peptide (VIP)	Pancreas, pheochromocytoma, esophagus
Rare		
Oncogenic osteomalacia	Phosphatonin [fibroblast growth factor 23 (FGF23)]	Hemangiopericytomas, osteoblastomas, fibromas, sarcomas, giant cell tumors, prostate, lung
Acromegaly	Growth hormone–releasing hormone (GHRH)	Pancreatic islet, bronchial and other carcinoids
	Growth hormone (GH)	Lung, pancreatic islet
Hyperthyroidism	Thyroid-stimulating hormone (TSH)	Hydatidiform mole, embryonal tumors, struma ovarii
Hypertension	Renin	Juxtaglomerular tumors, kidney, lung, pancreas, ovary

[a]Only the most common tumor types are listed. For most ectopic hormone syndromes, an extensive list of tumors has been reported to produce one or more hormones.

[b]hCG is produced eutopically by trophoblastic tumors. Certain tumors produce disproportionate

amounts of the hCG α or hCG β subunit. High levels of hCG rarely cause hyperthyroidism because of weak binding to the TSH receptor.cCalcitonin is produced eutopically by medullary thyroid carcinoma and is used as a tumor marker.

Hypercalcemia Caused by Ectopic Production of PthrP

(See also Chap. 353)

ETIOLOGY

Humoral hypercalcemia of malignancy (HHM) occurs in up to 20% of patients with cancer. HHM is most common in cancers of the lung, head and neck, skin, esophagus, breast, and genitourinary tract and in multiple myeloma and lymphomas. Although several distinct humoral causes of HHM occur, it is caused most commonly by overproduction of PTHrP. In addition to acting as a circulating humoral factor, bone metastases (e.g., breast, multiple myeloma) may produce PTHrP, leading to local osteolysis and hypercalcemia.

PTHrP is structurally related to PTH and binds to the PTH receptor, explaining the similar biochemical features of HHM and hyperparathyroidism. PTHrP plays a key role in skeletal development and regulates cellular proliferation and differentiation in other tissues, including skin, bone marrow, breast, and hair follicles. The mechanism of PTHrP induction in malignancy is incompletely understood; however, tumor-bearing tissues commonly associated with HHM normally produce PTHrP during development or cell renewal. PTHrP expression is stimulated by hedgehog pathways and Gli transcription factors that are active in many malignancies. Transforming growth factor β (TGF-β), which is produced by many tumors, also stimulates PTHrP, in part by activating the Gli pathway. Mutations in certain oncogenes, such as *Ras*, also can activate PTHrP expression. In adult T cell lymphoma, the transactivating Tax protein produced by human T cell lymphotropic virus I (HTLV-I) stimulates PTHrP promoter activity. Metastatic lesions to bone are more likely to produce PTHrP than are metastases in other tissues, suggesting that bone produces factors (e.g., TGF-β) that enhance PTHrP production or that PTHrP-producing metastases have a selective growth advantage in bone. Thus, PTHrP production can be stimulated by mutations in oncogenes, altered expression of viral or cellular transcription factors, and local growth factors.

Another relatively common cause of HHM is excess production of 1,25-dihydroxyvitamin D. Like granulomatous disorders associated with hypercalcemia, lymphomas can produce an enzyme that converts 25-hydroxyvitamin D to the more active 1,25-dihydroxyvitamin D, leading to enhanced gastrointestinal calcium absorption. Other causes of HHM include tumor-mediated production of osteolytic cytokines and inflammatory mediators.

CLINICAL MANIFESTATIONS

The typical presentation of HHM is a patient with a known malignancy who is found to be hypercalcemic on routine laboratory tests. Less often, hypercalcemia is the initial presenting feature of malignancy. Particularly when calcium levels are markedly increased [>3.5 mmol/L (>14 mg/dL)], patients may experience fatigue, mental status changes, dehydration, or symptoms of nephrolithiasis.

DIAGNOSIS

Features that favor HHM, as opposed to primary hyperparathyroidism, include known malignancy, recent onset of hypercalcemia, and very high serum calcium levels. Like hyperparathyroidism, hypercalcemia caused by PTHrP is accompanied by hypercalciuria and hypophosphatemia. Patients with HHM typically have metabolic alkalosis rather than hyperchloremic acidosis, as is seen in hyperparathyroidism. Measurement of PTH is useful to exclude primary hyperparathyroidism; the PTH level should be suppressed in HHM. An elevated PTHrP level confirms the diagnosis, and it is increased in ~80% of hypercalcemic patients with cancer. 1,25-Dihydroxyvitamin D levels may be increased in patients with lymphoma.

TREATMENT: HUMORAL HYPERCALCEMIA OF MALIGNANCY

The management of HHM begins with removal of excess calcium in the diet, medications, or IV solutions. Oral phosphorus (e.g., 250 mg Neutra-Phos 3–4 times daily) should be given until serum phosphorus is >1 mmol/L (>3 mg/dL). Saline rehydration is used to dilute serum calcium and promote calciuresis. Forced diuresis with furosemide or other loop diuretics can enhance calcium excretion but provides relatively little value except in life-threatening hypercalcemia. When used, loop diuretics should be administered only after complete rehydration and with careful monitoring of fluid balance. Bisphosphonates such as pamidronate (60–90 mg IV), zoledronate (4–8 mg IV), and etidronate (7.5 mg/kg per day PO for 3–7 consecutive days) can reduce serum calcium within 1–2 days and suppress calcium release for several weeks. Bisphosphonate infusions can be repeated, or oral bisphosphonates can be used for chronic treatment. Dialysis should be considered in severe hypercalcemia when saline hydration and bisphosphonate treatments are not possible or are too slow in onset. Previously used agents such as calcitonin and mithramycin have little utility now that bisphosphonates are available. Calcitonin (2–8 U/kg SC every 6–12 h) should be considered when rapid correction of severe hypercalcemia is needed. Hypercalcemia associated with lymphomas, multiple myeloma, or leukemia may respond to glucocorticoid treatment (e.g., prednisone 40–100 mg

PO in four divided doses).

Ectopic Vasopressin: Tumor-Associated SIADH

(See also Chap. 45)

ETIOLOGY

Vasopressin is an antidiuretic hormone normally produced by the posterior pituitary gland. Ectopic vasopressin production by tumors is a common cause of the syndrome of inappropriate antidiuretic hormone (SIADH), occurring in at least half of patients with SCLC. SIADH also can be caused by a number of nonneoplastic conditions, including central nervous system (CNS) trauma, infections, and medications (Chap. 340). Compensatory responses to SIADH, such as decreased thirst, may mitigate the development of hyponatremia. However, with prolonged production of excessive vasopressin, the osmostat controlling thirst and hypothalamic vasopressin secretion may become reset. In addition, intake of free water, orally or intravenously, can quickly worsen hyponatremia because of reduced renal diuresis.

Tumors with neuroendocrine features, such as SCLC and carcinoids, are the most common sources of ectopic vasopressin production, but it also occurs in other forms of lung cancer and with CNS lesions, head and neck cancer, and genitourinary, gastrointestinal, and ovarian cancers. The mechanism of activation of the vasopressin gene in these tumors is unknown but often involves concomitant expression of the adjacent oxytocin gene, suggesting derepression of this locus.

CLINICAL MANIFESTATIONS

Most patients with ectopic vasopressin secretion are asymptomatic and are identified because of the presence of hyponatremia on routine chemistry testing. Symptoms may include weakness, lethargy, nausea, confusion, depressed mental status, and seizures. The severity of symptoms reflects the rapidity of onset as well as the extent of hyponatremia. Hyponatremia usually develops slowly but may be exacerbated by the administration of IV fluids or the institution of new medications.

DIAGNOSIS

The diagnostic features of ectopic vasopressin production are the same as those of other causes of SIADH (Chaps. 45 and 340). Hyponatremia and reduced serum osmolality occur in the setting of an inappropriately normal or increased urine osmolality. Urine sodium excretion is normal or increased unless volume depletion is present. Other causes of hyponatremia should be excluded, including renal, adrenal, or thyroid insufficiency. Physiologic sources of vasopressin stimulation (CNS lesions, pulmonary disease, nausea), adaptive circulatory mechanisms (hypotension, heart failure, hepatic cirrhosis), and medications, including many chemotherapeutic agents, also should be considered as possible causes of hyponatremia. Vasopressin measurements are not usually necessary to make the diagnosis.

TREATMENT: ECTOPIC VASOPRESSIN: TUMOR-ASSOCIATED SIADH

Most patients with ectopic vasopressin production develop hyponatremia over several weeks or months. The disorder should be corrected gradually unless mental status is altered or there is risk of seizures. Treatment of the underlying malignancy may reduce ectopic vasopressin production, but this response is slow if it occurs at all. Fluid restriction to less than urine output, plus insensible losses, is often sufficient to correct hyponatremia partially. However, strict monitoring of the amount and types of liquids consumed or administered intravenously is required for fluid restriction to be effective. Salt tablets and saline are not helpful unless volume depletion is also present. Demeclocycline (150–300 mg orally three to four times daily) can be used to inhibit vasopressin action on the renal distal tubule, but its onset of action is relatively slow (1–2 weeks). Conivaptan, a nonpeptide V_2-receptor antagonist, can be administered either PO (20–120 mg bid) or IV (10–40 mg) and is particularly effective when used in combination with fluid restriction in euvolemic hyponatremia. Severe hyponatremia (Na <115 meq/L) or mental status changes may require treatment with hypertonic (3%) or normal saline infusion together with furosemide to enhance free water clearance. The rate of sodium correction should be slow (0.5–1 meq/L per h) to prevent rapid fluid shifts and the possible development of central pontine myelinolysis.

Cushing's Syndrome Caused by Ectopic ACTH Production

(See also Chap. 342)

ETIOLOGY

Ectopic ACTH production accounts for 10–20% of cases of Cushing's syndrome. The syndrome is particularly common in neuroendocrine tumors. SCLC (>50%) is by far the most common cause of ectopic ACTH, followed by thymic carcinoid (15%), islet cell tumors (10%), bronchial carcinoid (10%), other carcinoids (5%), and pheochromocytomas (2%). Ectopic ACTH production is caused by increased expression of the proopiomelanocortin (POMC) gene, which encodes ACTH, along with melanocyte-stimulating hormone (MSH), ß lipotropin, and several other peptides. In many tumors, there is abundant but aberrant expression of the POMC gene from an internal promoter, proximal to the third exon, which encodes ACTH. However, because this product lacks the signal sequence necessary for

protein processing, it is not secreted. Increased production of ACTH arises instead from less abundant, but unregulated, POMC expression from the same promoter site used in the pituitary. However, because the tumors lack many of the enzymes needed to process the POMC polypeptide, it is typically released as multiple large, biologically inactive fragments along with relatively small amounts of fully processed, active ACTH.

Rarely, corticotropin-releasing hormone (CRH) is produced by pancreatic islet cell tumors, SCLC, medullary thyroid cancer, carcinoids, or prostate cancer. When levels are high enough, CRH can cause pituitary corticotrope hyperplasia and Cushing's syndrome. Tumors that produce CRH sometimes also produce ACTH, raising the possibility of a paracrine mechanism for ACTH production.

A distinct mechanism for ACTH-independent Cushing's syndrome involves ectopic expression of various G protein–coupled receptors in the adrenal nodules. Ectopic expression of the gastric inhibitory peptide (GIP) receptor is the best-characterized example of this mechanism. In this case, meals induce GIP secretion, which inappropriately stimulates adrenal growth and glucocorticoid production.

CLINICAL MANIFESTATIONS

The clinical features of hypercortisolemia are detected in only a small fraction of patients with documented ectopic ACTH production. Patients with ectopic ACTH syndrome generally exhibit less marked weight gain and centripetal fat redistribution, probably because the exposure to excess glucocorticoids is relatively short and because cachexia reduces the propensity for weight gain and fat deposition. The ectopic ACTH syndrome is associated with several clinical features that distinguish it from other causes of Cushing's syndrome (e.g., pituitary adenomas, adrenal adenomas, iatrogenic glucocorticoid excess). The metabolic manifestations of ectopic ACTH syndrome are dominated by fluid retention and hypertension, hypokalemia, metabolic alkalosis, glucose intolerance, and occasionally steroid psychosis. The very high ACTH levels often cause increased pigmentation, and melanotrope-stimulating hormone (MSH) activity derived from the POMC precursor peptide is also increased. The extraordinarily high glucocorticoid levels in patients with ectopic sources of ACTH can lead to marked skin fragility and easy bruising. In addition, the high cortisol levels often overwhelm the renal 11β-hydroxysteroid dehydrogenase type II enzyme, which normally inactivates cortisol and prevents it from binding to renal mineralocorticoid receptors. Consequently, in addition to the excess mineralocorticoids produced by ACTH stimulation of the adrenal gland, high levels of cortisol exert activity through the mineralocorticoid receptor, leading to severe hypokalemia.

DIAGNOSIS

The diagnosis of ectopic ACTH syndrome is usually not difficult in the setting of a known malignancy. Urine free cortisol levels fluctuate but are typically greater than two to four times normal, and the plasma ACTH level is usually >22 pmol/L (>100 pg/mL). A suppressed ACTH level excludes this diagnosis and indicates an ACTH-independent cause of Cushing's syndrome (e.g., adrenal or exogenous glucocorticoid). In contrast to pituitary sources of ACTH, most ectopic sources of ACTH do not respond to glucocorticoid suppression. Therefore, high-dose dexamethasone (8 mg PO) suppresses 8:00 A.M. serum cortisol (50% decrease from baseline) in ~80% of pituitary ACTH-producing adenomas but fails to suppress ectopic ACTH in ~90% of cases. Bronchial and other carcinoids are well-documented exceptions to these general guidelines, as these ectopic sources of ACTH may exhibit feedback regulation indistinguishable from pituitary adenomas, including suppression by high-dose dexamethasone, and ACTH responsiveness to adrenal blockade with metyrapone. If necessary, petrosal sinus catheterization can be used to evaluate a patient with ACTH-dependent Cushing's syndrome when the source of ACTH is unclear. After CRH stimulation, a 3:1 petrosal sinus:peripheral ACTH ratio strongly suggests a pituitary ACTH source. Imaging studies are also useful in the evaluation of suspected carcinoid lesions, allowing biopsy and characterization of hormone production using special stains.

TREATMENT: DYSPNEA CUSHING'S SYNDROME CAUSED BY ECTOPIC ACTH PRODUCTION

The morbidity associated with the ectopic ACTH syndrome can be substantial. Patients may experience depression or personality changes because of extreme cortisol excess. Metabolic derangements, including diabetes mellitus and hypokalemia, can worsen fatigue. Poor wound healing and predisposition to infections can complicate the surgical management of tumors, and opportunistic infections caused by organisms such as *Pneumocystis carinii* and mycoses are often the cause of death in patients with ectopic ACTH production. Depending on prognosis and treatment plans for the underlying malignancy, measures to reduce cortisol levels are often indicated. Treatment of the underlying malignancy may reduce ACTH levels but is rarely sufficient to reduce cortisol levels to normal. Adrenalectomy is not practical for most of these patients but should be considered if the underlying tumor is not resectable and the prognosis is otherwise favorable (e.g., carcinoid). Medical therapy with ketoconazole (300–600 mg PO bid), metyrapone (250–500 mg PO every 6 h), mitotane (3–6 g PO in four divided doses, tapered to maintain low cortisol production), or other agents that block steroid synthesis or action is often the most practical strategy for managing the hypercortisolism associated with ectopic ACTH production (Chap. 339). Glucocorticoid replacement should be provided to prevent adrenal insufficiency. Unfortunately, many patients eventually progress despite medical

blockade.

Tumor-Induced Hypoglycemia Caused by Excess Production of IGF-II

(See also Chap. 345)

Mesenchymal tumors, hemangiopericytomas, hepatocellular tumors, adrenal carcinomas, and a variety of other large tumors have been reported to produce excessive amounts of insulin-like growth factor type II (IGF-II) precursor, which binds weakly to insulin receptors and strongly to IGF-I receptors, leading to insulin-like actions. The gene encoding IGF-II resides on a chromosome 11p15 locus that is normally imprinted (that is, expression is exclusively from a single parental allele). Biallelic expression of the IGF-II gene occurs in a subset of tumors, suggesting loss of methylation and loss of imprinting as a mechanism for gene induction. In addition to increased IGF-II production, IGF-II bioavailability is increased due to complex alterations in circulating binding proteins. Increased IGF-II suppresses growth hormone (GH) and insulin, resulting in reduced IGF binding protein 3 (IGFBP-3), IGF-I, and acid-labile subunit (ALS). The reduction in ALS and IGFBP-3, which normally sequester IGF-II, causes it to be displaced to a small circulating complex that has greater access to insulin target tissues. For this reason, circulating IGF-II levels may not be markedly increased despite causing hypoglycemia. In addition to IGF-II–mediated hypoglycemia, tumors may occupy enough of the liver to impair gluconeogenesis.

In most cases, the tumor causing hypoglycemia is clinically apparent (usually >10 cm in size) and hypoglycemia develops in association with fasting. The diagnosis is made by documenting low serum glucose and suppressed insulin levels in association with symptoms of hypoglycemia. Serum IGF-II levels may not be increased (IGF-II assays may not detect IGF-II precursors). Increased IGF-II mRNA expression is found in most of these tumors. Any medications associated with hypoglycemia should be eliminated. Treatment of the underlying malignancy, if possible, may reduce the predisposition to hypoglycemia. Frequent meals and IV glucose, especially during sleep or fasting, are often necessary to prevent hypoglycemia. Glucagon and glucocorticoids have also been used to enhance glucose production.

Human Chorionic Gonadotropin

hCG is composed of α and β subunits and can be produced as intact hormone, which is biologically active, or as uncombined biologically inert subunits. Ectopic production of intact hCG occurs most often in association with testicular embryonal tumors, germ cell tumors, extragonadal germinomas, lung cancer, hepatoma, and pancreatic islet tumors. Eutopic production of hCG occurs with trophoblastic malignancies. hCG α subunit production is particularly common in lung cancer and pancreatic islet cancer. In men, high hCG levels stimulate steroidogenesis and aromatase activity in testicular Leydig cells, resulting in increased estrogen production and the development of gynecomastia. Precocious puberty in boys or gynecomastia in men should prompt measurement of hCG and consideration of a testicular tumor or another source of ectopic hCG production. Most women are asymptomatic. hCG is easily measured. Treatment should be directed at the underlying malignancy.

Oncogenic Osteomalacia

Hypophosphatemic oncogenic osteomalacia, also called tumor-induced osteomalacia (TIO), is characterized by markedly reduced serum phosphorus and renal phosphate wasting, leading to muscle weakness, bone pain, and osteomalacia. Serum calcium and PTH levels are normal, and 1,25-dihydroxyvitamin D is low. Oncogenic osteomalacia is usually caused by benign mesenchymal tumors, such as hemangiopericytomas, fibromas, and giant cell tumors, often of the skeletal extremities or head. It has also been described in sarcomas and in patients with prostate and lung cancer. Resection of the tumor reverses the disorder, confirming its humoral basis. The circulating phosphaturic factor is called *phosphatonin*—a factor that inhibits renal tubular reabsorption of phosphate and renal conversion of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D. Phosphatonin has been identified as fibroblast growth factor 23 (FGF23). FGF23 levels are increased in some, but not all, patients with osteogenic osteomalacia. The disorder exhibits biochemical features similar to those seen with inactivating mutations in the *PHEX* gene, the cause of hereditary X-linked hypophosphatemia. The *PHEX* gene encodes a protease that inactivates FGF23. Treatment involves removal of the tumor, if possible, and supplementation with phosphate and vitamin D. Octreotide treatment reduces phosphate wasting in some patients with tumors that express somatostatin receptor subtype 2. Octreotide scans may also be useful in detecting these tumors.

HEMATOLOGIC SYNDROMES

The elevation of granulocyte, platelet, and eosinophil counts in most patients with myeloproliferative disorders is caused by the proliferation of the myeloid elements due to the underlying disease rather than to a paraneoplastic syndrome. The paraneoplastic hematologic syndromes in patients with solid tumors are less well characterized than are the endocrine syndromes because the ectopic hormone(s) or cytokines responsible have not been identified in most of these tumors (Table 100–2). The extent of the paraneoplastic syndromes parallels the course of the cancer.

Table 100–2 Paraneoplastic Hematologic Syndromes

Syndrome	Proteins	Cancers Typically Associated with Syndrome
Erythrocytosis	Erythropoietin	Renal cancers Hepatocarcinoma Cerebellar hemangioblastomas
Granulocytosis	G-CSFGM-CSFIL-6	Lung cancer Gastrointestinal cancer Ovarian cancer Genitourinary cancer Hodgkin's disease
Thrombocytosis	IL-6	Lung cancer Gastrointestinal cancer Breast cancer Ovarian cancer Lymphoma
Eosinophilia	IL-5	Lymphoma Leukemia Lung cancer
Thrombophlebitis	Unknown	Lung cancer Pancreatic cancer Gastrointestinal cancer Breast cancer Genitourinary cancer Ovarian cancer Prostate cancer Lymphoma

Abbreviations: G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage CSF; IL, interleukin.

Erythrocytosis

Ectopic production of erythropoietin by cancer cells causes most paraneoplastic erythrocytosis. The ectopically produced erythropoietin stimulates the production of red blood cells (RBCs) in the bone marrow and raises the hematocrit. Other lymphokines and hormones produced by cancer cells may stimulate erythropoietin release but have not been proved to cause erythrocytosis.

Most patients with erythrocytosis have an elevated hematocrit (>52% in men, >48% in women) that is detected on a routine blood count. Approximately 3% of patients with renal cell cancer, 10% of patients with hepatoma, and 15% of patients with cerebellar hemangioblastomas have erythrocytosis. In most cases the erythrocytosis is asymptomatic.

Patients with erythrocytosis due to a renal cell cancer, hepatoma, or CNS cancer should have measurement of red cell mass. If the red cell mass is elevated, the serum erythropoietin level should be measured. Patients with an appropriate cancer, elevated erythropoietin levels, and no other explanation for erythrocytosis (e.g., hemoglobinopathy that causes increased O_2 affinity; Chap. 57) have the paraneoplastic syndrome.

TREATMENT: ERYTHROCYTOSIS

Successful resection of the cancer usually resolves the erythrocytosis. If the tumor cannot be resected or treated effectively with radiation therapy or chemotherapy, phlebotomy may control any symptoms related to erythrocytosis.

Granulocytosis

Approximately 30% of patients with solid tumors have granulocytosis (granulocyte count >8000/μL). In about half of patients with granulocytosis and cancer, the granulocytosis has an identifiable nonparaneoplastic etiology (infection, tumor necrosis, glucocorticoid administration, etc.). The other patients have proteins in urine and serum that stimulate the growth of bone marrow cells. Tumors and tumor cell lines from patients with lung, ovarian, and bladder cancers have been documented to produce granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), and/or interleukin 6 (IL-6). However, the etiology of granulocytosis has not been characterized in most patients.

Patients with granulocytosis are nearly all asymptomatic, and the differential white blood cell count does not have a shift to immature forms of neutrophils. Granulocytosis occurs in 40% of patients with lung and gastrointestinal cancers, 20% of patients with breast cancer, 30% of patients with brain tumors and ovarian cancers, 20% of patients with Hodgkin's disease, and 10% of patients with renal cell carcinoma. Patients with advanced-stage disease are more likely to have granulocytosis than are those with early-stage disease.

Paraneoplastic granulocytosis does not require treatment. The granulocytosis resolves when the underlying cancer is treated.

Thrombocytosis

Some 35% of patients with thrombocytosis (platelet count >400,000/μL) have an underlying diagnosis of cancer. IL-6, a candidate molecule for the etiology of paraneoplastic thrombocytosis, stimulates the production of platelets in vitro and in vivo. Some patients with cancer and thrombocytosis have

elevated levels of IL-6 in plasma. Another candidate molecule is thrombopoietin, a peptide hormone that stimulates megakaryocyte proliferation and platelet production. The etiology of thrombocytosis has not been established in most cases.

Patients with thrombocytosis are nearly all asymptomatic. Thrombocytosis is not clearly linked to thrombosis in patients with cancer. Thrombocytosis is present in 40% of patients with lung and gastrointestinal cancers; 20% of patients with breast, endometrial, and ovarian cancers; and 10% of patients with lymphoma. Patients with thrombocytosis are more likely to have advanced-stage disease and have a poorer prognosis than do patients without thrombocytosis. Paraneoplastic thrombocytosis does not require treatment.

Eosinophilia

Eosinophilia is present in ~1% of patients with cancer. Tumors and tumor cell lines from patients with lymphomas or leukemia may produce IL-5, which stimulates eosinophil growth. Activation of IL-5 transcription in lymphomas and leukemias may involve translocation of the long arm of chromosome 5, to which the genes for IL-5 and other cytokines map.

Patients with eosinophilia are typically asymptomatic. Eosinophilia is present in 10% of patients with lymphoma, 3% of patients with lung cancer, and occasional patients with cervical, gastrointestinal, renal, and breast cancer. Patients with markedly elevated eosinophil counts (>5000/μL) can develop shortness of breath and wheezing. A chest radiograph may reveal diffuse pulmonary infiltrates from eosinophil infiltration and activation in the lungs.

TREATMENT: EOSINOPHILIA

Definitive treatment is directed at the underlying malignancy: Tumors should be resected or treated with radiation or chemotherapy. In most patients who develop shortness of breath related to eosinophilia, symptoms resolve with the use of oral or inhaled glucocorticoids.

Thrombophlebitis

Deep venous thrombosis and pulmonary embolism are the most common thrombotic conditions in patients with cancer. Migratory or recurrent thrombophlebitis may be the initial manifestation of cancer. Nearly 15% of patients who develop deep venous thrombosis or pulmonary embolism have a diagnosis of cancer (Chap. 117). The coexistence of peripheral venous thrombosis with visceral carcinoma, particularly pancreatic cancer, is called *Trousseau's syndrome*.

PATHOGENESIS

Patients with cancer are predisposed to thromboembolism because they are often at bed rest or immobilized, and tumors may obstruct or slow blood flow. Chronic IV catheters also predispose to clotting. In addition, clotting may be promoted by release of procoagulants or cytokines from tumor cells or associated inflammatory cells or by platelet adhesion or aggregation. The specific molecules that promote thromboembolism have not been identified.

In addition to cancer causing secondary thrombosis, primary thrombophilic diseases may be associated with cancer. For example, the antiphospholipid antibody syndrome is associated with a wide range of pathologic manifestations (Chap. 320). About 20% of patients with this syndrome have cancers. Among patients with cancer and antiphospholipid antibodies, 35–45% develop thrombosis.

CLINICAL MANIFESTATIONS

Patients with cancer who develop deep venous thrombosis usually develop swelling or pain in the leg, and physical examination reveals tenderness, warmth, and redness. Patients who present with pulmonary embolism develop dyspnea, chest pain, and syncope, and physical examination shows tachycardia, cyanosis, and hypotension. Some 5% of patients with no history of cancer who have a diagnosis of deep venous thrombosis or pulmonary embolism will have a diagnosis of cancer within 1 year. The most common cancers associated with thromboembolic episodes include lung, pancreatic, gastrointestinal, breast, ovarian, and genitourinary cancers; lymphomas; and brain tumors. Patients with cancer who undergo surgical procedures requiring general anesthesia have a 20–30% risk of deep venous thrombosis.

DIAGNOSIS

The diagnosis of deep venous thrombosis in patients with cancer is made by impedance plethysmography or bilateral compression ultrasonography of the leg veins. Patients with a noncompressible venous segment have deep venous thrombosis. If compression ultrasonography is normal and there is a high clinical suspicion for deep venous thrombosis, venography should be done to look for a luminal filling defect. Elevation of D-dimer is not as predictive of deep venous thrombosis in patients with cancer as it is in patients without cancer; elevations are seen in people over age 65 years without concomitant evidence of thrombosis, probably as a consequence of increased thrombin deposition and turnover in aging.

Patients with symptoms and signs suggesting a pulmonary embolism should be evaluated with a chest radiograph, electrocardiogram, arterial blood gas analysis, and ventilation-perfusion scan. Patients

with mismatched segmental perfusion defects have a pulmonary embolus. Patients with equivocal ventilation-perfusion findings should be evaluated as described above for deep venous thrombosis in their legs. If deep venous thrombosis is detected, they should be anticoagulated. If deep venous thrombosis is not detected, they should be considered for a pulmonary angiogram.

Patients without a diagnosis of cancer who present with an initial episode of thrombophlebitis or pulmonary embolus need no additional tests for cancer other than a careful history and physical examination. In light of the many possible primary sites, diagnostic testing in asymptomatic patients is wasteful. However, if the clot is refractory to standard treatment or is in an unusual site or if the thrombophlebitis is migratory or recurrent, efforts to find an underlying cancer are indicated.

TREATMENT: THROMBOPHLEBITIS

Patients with cancer and a diagnosis of deep venous thrombosis or pulmonary embolism should be treated initially with IV unfractionated heparin or low-molecular-weight heparin for at least 5 days, and warfarin should be started within 1 or 2 days. The warfarin dose should be adjusted so that the international normalized ratio (INR) is 2–3. Patients with proximal deep venous thrombosis and a relative contraindication to heparin anticoagulation (hemorrhagic brain metastases or pericardial effusion) should be considered for placement of a filter in the inferior vena cava (Greenfield filter) to prevent pulmonary embolism. Warfarin should be administered for 3–6 months. An alternative approach is to use low-molecular-weight heparin for 6 months. Patients with cancer who undergo a major surgical procedure should be considered for heparin prophylaxis or pneumatic boots. Breast cancer patients undergoing chemotherapy and patients with implanted catheters should be considered for prophylaxis (1 mg/d warfarin).

Cutaneous paraneoplastic syndromes are discussed in Chap. 53. Neurologic paraneoplastic syndromes are discussed in Chap. 101.

ACKNOWLEDGMENT

The authors acknowledge the contributions of Bruce E. Johnson to prior versions of this chapter.

FURTHER READINGS

AL-Tourah AJ et al: Paraneoplastic erythropoietin-induced polycythemia associated with small lymphocytic lymphoma. J Clin Oncol 24:2388, 2006[PMID: 16710039]

Gosden RG, Feinberg AP: Genetics and epigenetics—nature's pen-and-pencil set. N Engl J Med 356:731, 2007[PMID: 17301306]

Mundy GR, Edwards JR: PTH-related peptide (PTHrP) in hypercalcemia. J Am Soc Nephrol 19:672, 2008[PMID: 18256357]

Pelosof LC, Gerber DE: Paraneoplastic syndromes: An approach to diagnosis and treatment. Mayo Clin Proc 85:838, 2010[PMID: 20810794]

Rikhof B ET AL: The insulin-like growth factor system and sarcomas. J Pathol 217:469, 2009[PMID: 19148905]

Shaikh A et al: Regulation of phosphate homeostasis by the phosphatonins and other novel mediators. Pediatr Nephrol 23:1203, 2008[PMID: 18288501]

Stewart AF: Clinical practice: Hypercalcemia associated with cancer. N Engl J Med 352:373, 2005[PMID: 15673803]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PARANEOPLASTIC NEUROLOGIC SYNDROMES: INTRODUCTION

Paraneoplastic neurologic disorders (PNDs) are cancer-related syndromes that can affect any part of the nervous system (Table 101–1). They are caused by mechanisms other than metastasis or by any of the complications of cancer such as coagulopathy, stroke, metabolic and nutritional conditions, infections, and side effects of cancer therapy. In 60% of patients the neurologic symptoms precede the cancer diagnosis. Clinically disabling PNDs occur in 0.5–1% of all cancer patients, but they affect 2–3% of patients with neuroblastoma or small cell lung cancer (SCLC) and 30–50% of patients with thymoma or sclerotic myeloma.

Table 101–1 Paraneoplastic Syndromes of the Nervous System

Classic Syndromes: Usually Occur With Cancer Association	Nonclassic Syndromes: May Occur With and Without Cancer Association
Encephalomyelitis	Brainstem encephalitis
Limbic encephalitis	Stiff-person syndrome
Cerebellar degeneration (adults)	Necrotizing myelopathy
Opsoclonus-myoclonus	Motor neuron disease
Subacute sensory neuronopathy	Guillain-Barré syndrome
Gastrointestinal paresis or pseudo-obstruction	Subacute and chronic mixed sensory-motor neuropathies
Dermatomyositis (adults)	Neuropathy associated with plasma cell dyscrasias and lymphoma
Lambert-Eaton myasthenic syndrome	Vasculitis of nerve
Cancer or melanoma associated retinopathy	Pure autonomic neuropathy
	Acute necrotizing myopathy
	Polymyositis
	Vasculitis of muscle
	Optic neuropathy
	BDUMP

Abbreviation: BDUMP, bilateral diffuse uveal melanocytic proliferation.

PATHOGENESIS

Most PNDs are mediated by immune responses triggered by neuronal proteins (onconeuronal antigens) expressed by tumors. In PNDs of the central nervous system (CNS), many antibody-associated immune responses have been identified (Table 101–2). These antibodies react with the patient's tumor, and their detection in serum or cerebrospinal fluid (CSF) usually predicts the presence of cancer. When the antigens are intracellular, most syndromes are associated with extensive infiltrates of CD4+ and CD8+ T cells, microglial activation, gliosis, and variable neuronal loss. The infiltrating T cells are often in close contact with neurons undergoing degeneration, suggesting a primary pathogenic role. T cell—mediated cytotoxicity may contribute directly to cell death in these PNDs. Thus both humoral and cellular immune mechanisms participate in the pathogenesis of many PNDs. This complex immunopathogenesis may underlie the resistance of many of these conditions to therapy.

Table 101–2 Antibodies to Intracellular Antigens, Syndromes, and Associated Cancers

Antibody	Associated Neurologic Syndrome(s)	Tumors
Anti-Hu	Encephalomyelitis, subacute sensory neuronopathy	SCLC
Anti-Yo	Cerebellar degeneration	Ovary, breast

Anti-Ri	Cerebellar degeneration, opsoclonus	Breast, gynecologic, SCLC
Anti-Tr	Cerebellar degeneration	Hodgkin lymphoma
Anti-CV_2/CRMP5	Encephalomyelitis, chorea, optic neuritis, uveitis, peripheral neuropathy	SCLC, thymoma, other
Anti-Ma proteins	Limbic, hypothalamic, brainstem encephalitis	Testicular (Ma2), other (Ma)
Anti-amphiphysin	Stiff-person syndrome, encephalomyelitis	Breast, SCLC
Recoverin, bipolar cell antibodies, others[a]	Cancer-associated retinopathy (CAR) Melanoma-associated retinopathy (MAR)	SCLC (CAR), melanoma (MAR)
Anti-GAD	Stiff-person, cerebellar syndromes	Infrequent tumor association (thymoma)

[a]A variety of target antigens have been identified.

Abbreviations: CRMP, collapsing response-mediator protein; SCLC, small cell lung cancer.

In contrast to the disorders associated with immune responses against intracellular antigens, those associated with antibodies to antigens expressed on the neuronal cell surface of the CNS or at neuromuscular synapses are more responsive to immunotherapy (Table 101–3, Fig. 101-1). These disorders occur with and without a cancer association, and there is increasing evidence that they are mediated by the antibodies.

Table 101–3 Antibodies to Cell Surface or Synaptic Antigens, Syndromes, and Associated Tumors

Antibody	Neurologic Syndrome	Tumor Type when Associated
Anti-AChR (muscle)[a]	Myasthenia gravis	Thymoma
Anti-AChR (neuronal)[a]	Autonomic neuropathy	SCLC
Anti-VGKC- related proteins[b] (LGI1, Caspr2)	Neuromyotonia, limbic encephalitis	Thymoma, SCLC
Anti-VGCC[c]	LEMS, cerebellar degeneration	SCLC
Anti-NMDAR[d]	Anti-NMDAR encephalitis	Teratoma
Anti-AMPAR[d]	Limbic encephalitis with relapses	SCLC, thymoma, breast
Anti-$GABA_BR$[d]	Limbic encephalitis, seizures	SCLC, neuroendocrine
Glycine receptor[d]	Encephalomyelitis with rigidity, stiff-person syndrome	Lung cancer

[a]A direct pathogenic role of these antibodies has been demonstrated.

[b]Anti-VGKC-related proteins are pathogenic for some types of neuromyotonia.

[c]Anti-VGCC antibodies are pathogenic for LEMS.

[d] These antibodies are strongly suspected to be pathogenic.

Abbreviations: AChR, acetylcholine receptor; AMPAR, α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor; $GABA_BR$, Gamma-amino-butyric acid B receptor; GAD, glutamic acid decarboxylase; LEMS, Lambert-Eaton myasthenic syndrome; NMDAR, *N*-methyl-D-aspartate receptor; SCLC, small cell lung cancer; VGCC, voltage-gated calcium channel; VGKC, voltage-gated potassium channel.

Other PNDs are likely immune-mediated, although their antigens are unknown. These include several syndromes of inflammatory neuropathies and myopathies. In addition, many patients with typical PND syndromes are antibody-negative. For still other PNDs, the cause remains quite obscure. These include, among others, several neuropathies that occur in the terminal stages of cancer and a number of neuropathies associated with plasma cell dyscrasias or lymphoma without evidence of inflammatory infiltrates or deposits of immunoglobulin, cryoglobulin, or amyloid.

Figure 101-1

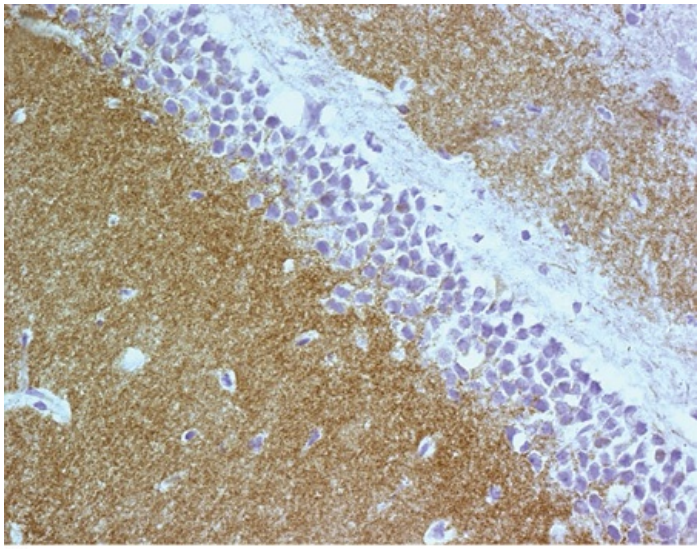

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

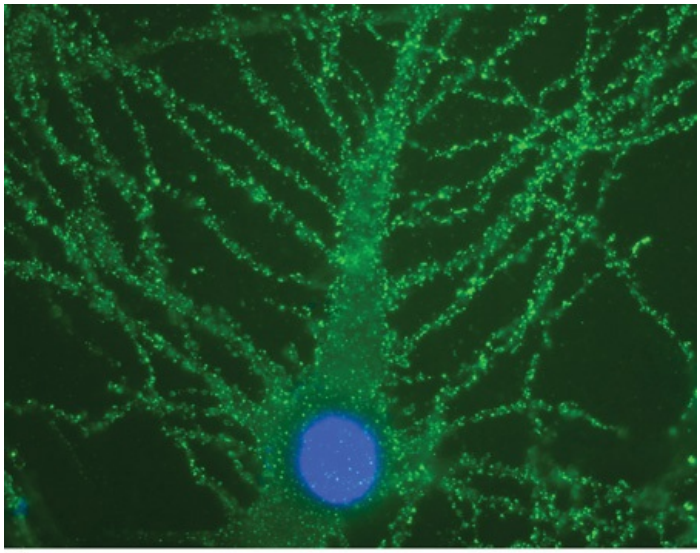

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Antibodies to NR1/NR2 subunits of the NMDA receptor in a patient with paraneoplastic encephalitis and ovarian teratoma. *Panel A* is a section of dentate gyrus of rat hippocampus immunolabeled (brown staining) with the patient's antibodies. The reactivity predominates in the molecular layer, which is highly enriched in dendritic processes. *Panel B* shows the antibody reactivity with cultures of rat hippocampal neurons; the intense green -immunolabeling is due to the antibodies against the NR1 subunits of NMDA receptors.

Approach to the Patient: Paraneoplastic Neurologic Disorders

Three key concepts are important for the diagnosis and management of PNDs. First, it is common for symptoms to appear before the presence of a tumor is known; second, the neurologic syndrome usually develops rapidly, producing severe deficits in a short period of time; and third, there is evidence that prompt tumor control improves the neurologic outcome. Therefore, the major concern of the physician is to recognize a disorder promptly as paraneoplastic to identify and treat the tumor.

PND OF THE CENTRAL NERVOUS SYSTEM AND DORSAL ROOT GANGLIA

When symptoms involve brain, spinal cord, or dorsal root ganglia, the suspicion of PND is usually based on a combination of clinical, radiologic, and CSF findings. In these cases, a biopsy of the affected tissue is often difficult to obtain, and although useful to rule out other disorders (e.g., metastasis, infection), neuropathologic findings are not specific for PND. Furthermore, there are no specific radiologic or electrophysiologic tests that are diagnostic of PND. The presence of antineuronal antibodies (Tables 101–2 and 101–3) may help in the diagnosis, but only 60–70% of PNDs of the CNS and less than 20% of those involving the peripheral nervous system have neuronal or neuromuscular antibodies that can be used as diagnostic tests. MRI and CSF studies are important to rule out

neurologic complications due to the direct spread of cancer, particularly metastatic and leptomeningeal disease. In most PNDs the MRI findings are nonspecific. Paraneoplastic limbic encephalitis is usually associated with characteristic MRI abnormalities in the mesial temporal lobes (see below), but similar findings can occur with other disorders [e.g., nonparaneoplastic autoimmune limbic encephalitis, and human herpesvirus type 6 (HHV-6) encephalitis] (Fig. 101-2). The CSF profile of patients with PND of the CNS or dorsal root ganglia typically consists of mild to moderate pleocytosis (<200 mononuclear cells, predominantly lymphocytes), an increase in the protein concentration, intrathecal synthesis of IgG, and a variable presence of oligoclonal bands.

Figure 101-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fluid-attenuated inversion recovery sequence MRI of a patient with limbic encephalitis and LGI1 antibodies. Note the abnormal hyperintensity involving the medial aspect of the temporal lobes.

PND OF NERVE AND MUSCLE

If symptoms involve peripheral nerve, neuromuscular junction, or muscle, the diagnosis of a specific PND is usually established on clinical, electrophysiologic, and pathologic grounds. The clinical history, accompanying symptoms (e.g., anorexia, weight loss), and type of syndrome dictate the studies and degree of effort needed to demonstrate a neoplasm. For example, the frequent association of Lambert-Eaton myasthenic syndrome (LEMS) with SCLC should lead to a chest and abdomen CT or body positron emission tomography (PET) scan and, if negative, periodic tumor screening for at least 3 years after the neurologic diagnosis. In contrast, the weak association of polymyositis with cancer calls into question the need for repeated cancer screenings in this situation. Serum and urine immunofixation studies should be considered in patients with peripheral neuropathy of unknown cause; detection of a monoclonal gammopathy suggests the need for additional studies to uncover a B cell or plasma cell malignancy. In paraneoplastic neuropathies, diagnostically useful antineuronal antibodies are limited to anti-CV_2/CRMP5 and anti-Hu. For any type of PND, if antineuronal antibodies are negative, the diagnosis relies on the demonstration of cancer and the exclusion of other cancer-related or independent neurologic disorders. Combined CT and PET scans often uncover tumors undetected by other tests.

For germ-cell tumors of the testis and teratomas of the ovary ultrasound and MRI may reveal tumors undetectable by PET.

SPECIFIC PARANEOPLASTIC NEUROLOGIC SYNDROMES

Paraneoplastic Encephalomyelitis and Focal Encephalitis

The term *encephalomyelitis* describes an inflammatory process with multifocal involvement of the nervous system, including brain, brainstem, cerebellum, and spinal cord. It is often associated with dorsal root ganglia and autonomic dysfunction. For any given patient, the clinical manifestations are

determined by the areas predominantly involved, but pathologic studies almost always reveal abnormalities beyond the symptomatic regions. Several clinicopathologic syndromes may occur alone or in combination: (1) *cortical encephalitis*, which may present as "epilepsia partialis continua"; (2) *limbic encephalitis*, characterized by confusion, depression, agitation, anxiety, severe short-term memory deficits, partial complex seizures, and sometimes dementia (the MRI usually shows unilateral or bilateral medial temporal lobe abnormalities, best seen with T2 and fluid-attenuated inversion recovery sequences, and occasionally enhancing with gadolinium); (3) *brainstem encephalitis,* resulting in eye movement disorders (nystagmus, opsoclonus, supranuclear or nuclear paresis), cranial nerve paresis, dysarthria, dysphagia, and central autonomic dysfunction; (4) *cerebellar gait and limb ataxia*; (5) *myelitis*, which may cause lower or upper motor neuron symptoms, myoclonus, muscle rigidity, and spasms; and (6) *autonomic dysfunction* as a result of involvement of the neuraxis at multiple levels, including hypothalamus, brainstem, and autonomic nerves (see autonomic neuropathy). Cardiac arrhythmias, postural hypotension, or central hypoventilation are frequent causes of death in patients with encephalomyelitis.

Paraneoplastic encephalomyelitis and focal encephalitis are usually associated with SCLC, but many other cancers have also been reported. Patients with SCLC and these syndromes usually have anti-Hu antibodies in serum and CSF. Anti-CV_2/CRMP5 antibodies occur less frequently; some of these patients may develop chorea, uveitis, or optic neuritis. Antibodies to Ma proteins are associated with limbic, hypothalamic, and brainstem encephalitis and occasionally with cerebellar symptoms (Fig. 101-3); some patients develop hypersomnia, cataplexy, and severe hypokinesia. MRI abnormalities are frequent, including those described with limbic encephalitis and variable involvement of the hypothalamus, basal ganglia, or upper brainstem. The oncologic associations of these antibodies are shown in Table 101–2.

Figure 101-3

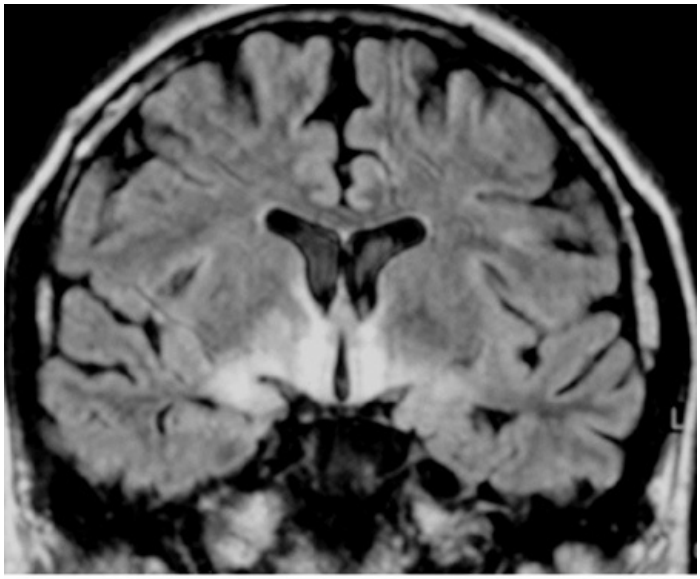

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

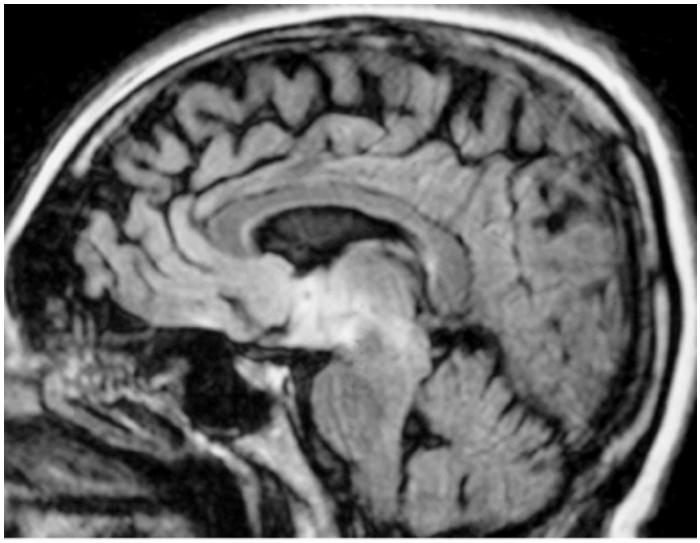

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

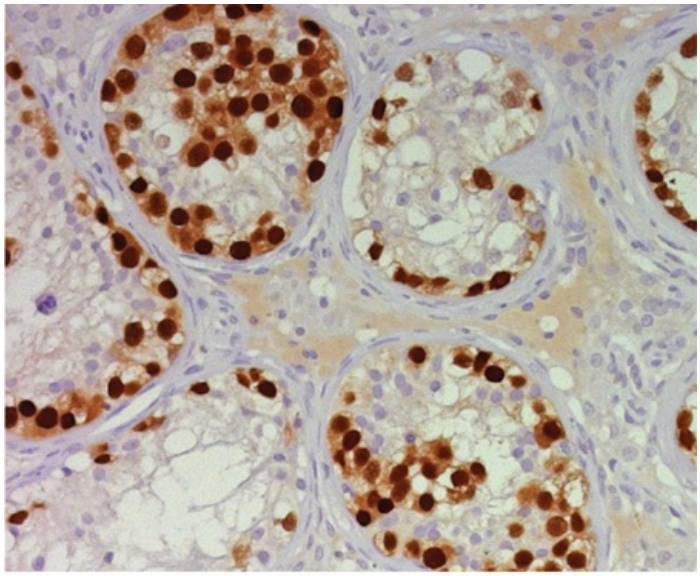

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI and tumor of a patient with anti-Ma2-associated encephalitis. Panels *A* and *B* are fluid-attenuated inversion recovery MRI sequences showing abnormal hyperintensities in the medial temporal lobes, hypothalamus, and upper brainstem. ***Panel C*** corresponds to a section of the patient's orchiectomy incubated with a specific marker (Oct4) of germ-cell tumors. The positive (brown) cells correspond to an intratubular germ-cell neoplasm.

TREATMENT: ENCEPHALOMYELITIS AND FOCAL ENCEPHALITIS

Most types of paraneoplastic encephalitis and encephalomyelitis respond poorly to treatment. Stabilization of symptoms or partial neurologic improvement may occasionally occur, particularly if there is a satisfactory response of the tumor to treatment. The roles of plasma exchange, intravenous immunoglobulin (IVIg), and immunosuppression have not been established. Approximately 30% of patients with anti-Ma2-associated encephalitis respond to treatment of the tumor (usually a germ-cell neoplasm of the testis) and immunotherapy.

Encephalitides with Antibodies to Cell-Surface or Synaptic Proteins (Table 101–3)

These disorders are important for three reasons: (1) they can occur with and without tumor association, (2) some syndromes predominate in young individuals and children, and (3) despite the severity of the symptoms patients usually respond to treatment of the tumor, if found, and immunotherapy (glucocorticoids, plasma exchange, IVIg, rituximab, or cyclophosphamide).

Encephalitis with antibodies to voltage-gated potassium channels (VGKC)-related proteins (LGI1,

Caspr2) predominates in men and frequently presents with memory loss and seizures (limbic encephalopathy), along with hyponatremia and sleep and autonomic dysfunction. Less commonly, patients develop neuromyotonia or a mixed clinical picture (Morvan's syndrome). Approximately 20% of patients with antibodies to VGKC-related proteins have an underlying tumor, usually SCLC or thymoma.

Encephalitis with N-*methyl-d-aspartate (NMDA) receptor antibodies* (Fig. 101-1) usually occurs in young women and children, but men and older patients of both sexes can be affected. The disorder has a characteristic pattern of symptom progression that includes a prodrome resembling a viral process, followed in a few days by the onset of severe psychiatric symptoms, memory loss, seizures, decreased level of consciousness, abnormal movements (orofacial, limb, and trunk dyskinesias, dystonic postures), autonomic instability, and frequent hypoventilation. The syndrome is often misdiagnosed as a viral or idiopathic encephalitis, neuroleptic malignant syndrome, or encephalitis lethargica, and many patients are initially evaluated by psychiatrists with the suspicion of drug abuse or an acute psychosis. The detection of an associated ovarian teratoma is age-dependant; 50% of female patients older than age 18 have uni- or bilateral ovarian teratomas, while less than 9% of girls younger than 14 years have a teratoma. In male patients the detection of a tumor is rare.

Encephalitis with α*-amino-3-hydroxy-5-methylisoxazole-4-propionate (AMPA) receptor antibodies* affects middle-aged women, who develop acute limbic dysfunction or less frequently prominent psychiatric symptoms; 70% of the patients have an underlying tumor in the lung, breast, or thymus. The neurologic disorder responds to treatment of the tumor and immunotherapy. Neurologic relapses may occur; these also respond to immunotherapy and are not necessarily associated with tumor recurrence.

Encephalitis with γ*-aminobutyric acid type B (GABA$_B$) receptor antibodies* usually presents with limbic encephalitis and seizures; 50% of the patients have SCLC or a neuroendocrine tumor of the lung. Neurologic symptoms often respond to immunotherapy and treatment of the tumor if found. Patients may have additional antibodies to glutamic acid decarboxylase (GAD), of unclear significance. Other antibodies to nonneuronal proteins are often found in these patients as well as in patients with AMPA receptor antibodies, indicating a general tendency to autoimmunity.

Paraneoplastic Cerebellar Degeneration

This disorder is often preceded by a prodrome that may include dizziness, oscillopsia, blurry or double vision, nausea, and vomiting. A few days or weeks later, patients develop dysarthria, gait and limb ataxia, and variable dysphagia. The examination usually shows downbeating nystagmus and, rarely, opsoclonus. Brainstem dysfunction, upgoing toes, or a mild neuropathy may occur, but more often the clinical features are restricted to the cerebellum. Early in the course, MRI studies are usually normal; later, the MRI typically reveals cerebellar atrophy. The disorder results from extensive degeneration of Purkinje cells, with variable involvement of other cerebellar cortical neurons, deep cerebellar nuclei, and spinocerebellar tracts. The tumors more frequently involved are SCLC, cancer of the breast and ovary, and Hodgkin lymphoma.

Anti-Yo antibodies in patients with breast and gynecologic cancers and anti-Tr antibodies in patients with Hodgkin lymphoma are the two immune responses typically associated with prominent or pure cerebellar degeneration. Antibodies to P/Q-type voltage-gated calcium channels (VGCC) occur in some patients with SCLC and cerebellar dysfunction; only some of these patients develop LEMS. A variable degree of cerebellar dysfunction can be associated with virtually any of the antibodies and PND of the CNS shown in Table 101–2. A number of single case reports have described neurologic improvement after tumor removal, plasma exchange, IVIg, cyclophosphamide, rituximab, or glucocorticoids. However, large series of patients with antibody-positive paraneoplastic cerebellar degeneration show that this disorder rarely improves with any treatment.

PARANEOPLASTIC OPSOCLONUS-MYOCLONUS SYNDROME

Opsoclonus is a disorder of eye movement characterized by involuntary, chaotic saccades that occur in all directions of gaze; it is frequently associated with myoclonus and ataxia. Opsoclonus-myoclonus may be cancer-related or idiopathic. When the cause is paraneoplastic, the tumors involved are usually cancer of the lung and breast in adults and neuroblastoma in children. The pathologic substrate of opsoclonus-myoclonus is unclear, but studies suggest that disinhibition of the fastigial nucleus of the cerebellum is involved. Most patients do not have detectable antineuronal antibodies. A small subset of patients with ataxia, opsoclonus, and other eye-movement disorders develop anti-Ri antibodies; in rare instances muscle rigidity, autonomic dysfunction, and dementia also occur. The tumors most frequently involved in anti-Ri-associated syndromes are breast and ovarian cancer. If the tumor is not successfully treated, the neurologic syndrome in adults often progresses to encephalopathy, coma, and death. In addition to treating the tumor, symptoms may respond to immunotherapy (glucocorticoids, plasma exchange, and/or IVIg).

At least 50% of children with opsoclonus-myoclonus have an underlying neuroblastoma. Hypotonia, ataxia, behavioral changes, and irritability are frequent accompanying symptoms. Neurologic

symptoms often improve with treatment of the tumor and glucocorticoids, adrenocorticotropic hormone (ACTH), plasma exchange, IVIg, and rituximab. Many patients are left with psychomotor retardation and behavioral and sleep problems.

Paraneoplastic Syndromes of the Spinal Cord

The number of reports of paraneoplastic spinal cord syndromes, such as *subacute motor neuronopathy* and *acute necrotizing myelopathy*, has decreased in recent years. This may represent a true decrease in incidence, due to improved and prompt oncologic interventions, or the identification of nonparaneoplastic etiologies.

Some patients with cancer develop *upper* or *lower motor neuron dysfunction* or both, resembling amyotrophic lateral sclerosis. It is unclear whether these disorders have a paraneoplastic etiology or simply coincide with the presence of cancer. There are isolated case reports of cancer patients with motor neuron dysfunction who had neurologic improvement after tumor treatment. A search for lymphoma should be undertaken in patients with a rapidly progressive motor neuron syndrome and a monoclonal protein in serum or CSF.

Paraneoplastic myelitis may present with upper or lower motor neuron symptoms, segmental myoclonus, and rigidity, and can be the first manifestation of encephalomyelitis.

Paraneoplastic myelopathy can also produce several syndromes characterized by prominent muscle stiffness and rigidity. The spectrum ranges from focal symptoms in one or several extremities (*stiff-limb syndrome* or *stiff-person syndrome*) to a disorder that also affects the brainstem (known as *encephalomyelitis with rigidity*) and likely has a different pathogenesis. Some patients with encephalomyelitis and rigidity have glycine receptor antibodies.

Paraneoplastic Stiff-Person Syndrome

This disorder is characterized by progressive muscle rigidity, stiffness, and painful spasms triggered by auditory, sensory, or emotional stimuli. Rigidity mainly involves the lower trunk and legs, but it can affect the upper extremities and neck. Symptoms improve with sleep and general anesthetics. Electrophysiologic studies demonstrate continuous motor unit activity. Antibodies associated with the stiff-person syndrome target proteins (GAD, amphiphysin) involved in the function of inhibitory synapses utilizing γ-aminobutyric acid (GABA) or glycine as neurotransmitters. Paraneoplastic stiff-person syndrome and amphiphysin antibodies are often related to SCLC and breast cancer. By contrast, antibodies to GAD may occur in some cancer patients but are much more frequently present in the nonparaneoplastic disorder.

TREATMENT: STIFF-PERSON SYNDROME

Optimal treatment of stiff-person syndrome requires therapy of the underlying tumor, glucocorticoids, and symptomatic use of drugs that enhance GABA-ergic transmission (diazepam, baclofen, sodium valproate, tiagabine, vigabatrin). A benefit of IVIg has been demonstrated for the nonparaneoplastic disorder but remains to be established for the paraneoplastic syndrome.

Paraneoplastic Sensory Neuronopathy or Dorsal Root Ganglionopathy

This syndrome is characterized by sensory deficits that may be symmetric or asymmetric, painful dysesthesias, radicular pain, and decreased or absent reflexes. All modalities of sensation and any part of the body including face and trunk can be involved. Specialized sensations such as taste and hearing can also be affected. Electrophysiologic studies show decreased or absent sensory nerve potentials with normal or near-normal motor conduction velocities. Symptoms result from an inflammatory, likely immune-mediated, process that targets the dorsal root ganglia, causing neuronal loss, proliferation of satellite cells, and secondary degeneration of the posterior columns of the spinal cord. The dorsal and less frequently the anterior nerve roots and peripheral nerves may also be involved. This disorder often precedes or is associated with encephalomyelitis and autonomic dysfunction and has the same immunologic and oncologic associations, e.g., anti-Hu antibodies and SCLC.

TREATMENT: SENSORY NEURONOPATHY

As with anti-Hu-associated encephalomyelitis, the therapeutic approach focuses on prompt treatment of the tumor. Glucocorticoids occasionally produce clinical stabilization or improvement. The benefit of IVIg and plasma exchange is not proved.

PARANEOPLASTIC PERIPHERAL NEUROPATHIES

These disorders may develop any time during the course of the neoplastic disease. Neuropathies occurring at late stages of cancer or lymphoma usually cause mild to moderate sensorimotor deficits due to axonal degeneration of unclear etiology. These neuropathies are often masked by concurrent neurotoxicity from chemotherapy and other cancer therapies. In contrast, the neuropathies that develop in the early stages of cancer frequently show a rapid progression, sometimes with a relapsing and remitting course, and evidence of inflammatory infiltrates and axonal loss or demyelination in biopsy studies. If demyelinating features predominate (Chap. 384), IVIg, plasma exchange, or

glucocorticoids may improve symptoms. Occasionally anti-CV_2/CRMP5 antibodies are present; detection of anti-Hu suggests concurrent dorsal root ganglionitis.

Guillain-Barré syndrome and *brachial plexitis* have occasionally been reported in patients with lymphoma, but there is no clear evidence of a paraneoplastic association.

Malignant monoclonal gammopathies include: (1) multiple myeloma and sclerotic myeloma associated with IgG or IgA monoclonal proteins; and (2) Waldenström's macroglobulinemia, B cell lymphoma, and chronic B cell lymphocytic leukemia associated with IgM monoclonal proteins. These disorders may cause neuropathy by a variety of mechanisms, including compression of roots and plexuses by metastasis to vertebral bodies and pelvis, deposits of amyloid in peripheral nerves, and paraneoplastic mechanisms. The paraneoplastic variety has several distinctive features. Approximately half of patients with sclerotic myeloma develop a sensorimotor neuropathy with predominantly motor deficits, resembling a chronic inflammatory demyelinating neuropathy (Chap. 385); some patients develop elements of the POEMS syndrome (*p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M* protein, *s*kin changes). Treatment of the plasmacytoma or sclerotic lesions usually improves the neuropathy. In contrast, the sensorimotor or sensory neuropathy associated with multiple myeloma rarely responds to treatment. Between 5 and 10% of patients with Waldenström's macroglobulinemia develop a distal symmetric sensorimotor neuropathy with predominant involvement of large sensory fibers. These patients may have IgM antibodies in their serum against myelin-associated glycoprotein and various gangliosides (Chap. 385). In addition to treating the Waldenström's macroglobulinemia, other therapies may improve the neuropathy, including plasma exchange, IVIg, chlorambucil, cyclophosphamide, fludarabine, or rituximab.

Vasculitis of the nerve and muscle causes a painful symmetric or asymmetric distal axonal sensorimotor neuropathy with variable proximal weakness. It predominantly affects elderly men and is associated with an elevated erythrocyte sedimentation rate and increased CSF protein concentration. SCLC and lymphoma are the primary tumors involved. Glucocorticoids and cyclophosphamide often result in neurologic improvement.

Peripheral nerve hyperexcitability (*neuromyotonia,* or *Isaacs' syndrome*) is characterized by spontaneous and continuous muscle fiber activity of peripheral nerve origin. Clinical features include cramps, muscle twitching (fasciculations or myokymia), stiffness, delayed muscle relaxation (pseudomyotonia), and spontaneous or evoked carpal or pedal spasms. The involved muscles may be hypertrophic, and some patients develop paresthesias and hyperhidrosis. CNS dysfunction, including mood changes, sleep disorder, or hallucinations, may occur. The electromyogram (EMG) shows fibrillations; fasciculations; and doublet, triplet, or multiplet single-unit (myokymic) discharges that have a high intraburst frequency. Approximately 20% of patients have serum antibodies to Caspr2-related proteins. The disorder often occurs without cancer; if paraneoplastic, benign, and malignant thymomas and SCLC are the usual tumors. Phenytoin, carbamazepine, and plasma exchange improve symptoms.

Paraneoplastic autonomic neuropathy usually develops as a component of other disorders, such as LEMS and encephalomyelitis. It may rarely occur as a pure or predominantly autonomic neuropathy with adrenergic or cholinergic dysfunction at the pre- or postganglionic levels. Patients can develop several life-threatening complications, such as gastrointestinal paresis with pseudoobstruction, cardiac dysrhythmias, and postural hypotension. Other clinical features include abnormal pupillary responses, dry mouth, anhidrosis, erectile dysfunction, and problems in sphincter control. The disorder occurs in association with several tumors, including SCLC, cancer of the pancreas or testis, carcinoid tumors, and lymphoma. Because autonomic symptoms can be the presenting feature of encephalomyelitis, serum anti-Hu and anti-CV_2/CRMP5 antibodies should be sought. Antibodies to ganglionic (alpha3-type) neuronal acetylcholine receptors are the cause of autoimmune autonomic ganglionopathy, a disorder that frequently occurs without cancer association (Chap. 375).

Lambert-Eaton Myasthenic Syndrome

LEMS is discussed in Chap. 386.

Myasthenia Gravis

Myasthenia gravis is discussed in Chap. 386.

Polymyositis-Dermatomyositis

Polymyositis and dermatomyositis are discussed in detail in Chap. 388.

Acute Necrotizing Myopathy

Patients with this syndrome develop myalgias and rapid progression of weakness involving the extremities and the pharyngeal and respiratory muscles, often resulting in death. Serum muscle enzymes are elevated, and muscle biopsy shows extensive necrosis with minimal or absent inflammation and sometimes deposits of complement. The disorder occurs as a paraneoplastic manifestation of a variety of cancers including SCLC and cancer of the gastrointestinal tract, breast, kidney, and prostate, among others. Glucocorticoids and treatment of the underlying tumor rarely control the disorder.

Paraneoplastic Visual Syndromes

This group of disorders involves the retina and, less frequently, the uvea and optic nerves. The term *cancer-associated retinopathy* is used to describe paraneoplastic cone and rod dysfunction characterized by photosensitivity, progressive loss of vision and color perception, central or ring scotomas, night blindness, and attenuation of photopic and scotopic responses in the electroretinogram (ERG). The most commonly associated tumor is SCLC. Melanoma-associated retinopathy affects patients with metastatic cutaneous melanoma. Patients develop acute onset of night blindness and shimmering, flickering, or pulsating photopsias that often progress to visual loss. The ERG shows reduced b waves with normal dark adapted a waves. Paraneoplastic optic neuritis and uveitis are very uncommon and can develop in association with encephalomyelitis. Some patients with paraneoplastic uveitis harbor anti-CV_2/CRMP5 antibodies.

Some paraneoplastic retinopathies are associated with serum antibodies that specifically react with the subset of retinal cells undergoing degeneration, supporting an immune-mediated pathogenesis (Table 101–2). Paraneoplastic retinopathies usually fail to improve with treatment, although rare responses to glucocorticoids, plasma exchange, and IVIg have been reported.

FURTHER READINGS

Antoine JC, Camdessanché JP: Peripheral nervous system involvement in patients with cancer. Lancet Neurol 6:75, 2007[PMID: 17166804]

Graus F et al: Antibodies and neuronal autoimmune disorders of the CNS. J Neurol 257:509, 2010[PMID: 20035430]

Lai M et al: Investigation of LGI1 as the antigen in limbic encephalitis previously attributed to potassium channels: A case series. Lancet Neurol 9:776, 2010[PMID: 20580615]

Rosenfeld MR, Dalmau J: Update on paraneoplastic and autoimmune disorders of the central nervous system. Semin Neurol 30:320, 2010[PMID: 20577938]

Titulaer MJ et al: Screening for tumours in paraneoplastic syndromes: Report of an EFNS task force. Eur J Neurol 18:19, 2011[PMID: 20880069]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THYMOMA: INTRODUCTION

The thymus is derived from the third and fourth pharyngeal pouches and is located in the anterior mediastinum. It is composed of epithelial and stromal cells derived from the pharyngeal pouch and lymphoid precursors derived from mesodermal cells. It is the site to which bone marrow precursors that are committed to differentiate into T cells migrate to complete their differentiation. Like many organs, it is organized into functional regions, in this case the cortex and the medulla. The cortex of the thymus contains ~85% of the lymphoid cells, and the medulla contains ~15%. It appears that the primitive bone marrow progenitors enter the thymus at the corticomedullary junction and migrate first through the cortex toward the periphery of the gland and then toward the medulla as they mature. Medullary thymocytes have a phenotype that cannot be distinguished readily from that of mature peripheral blood and lymph node T cells.

Several things can go wrong with the thymus, but thymic abnormalities are very rare. If the thymus does not develop properly, serious deficiencies in T cell development ensue and severe immunodeficiency is seen (e.g., DiGeorge syndrome, Chap. 316). If a lymphoid cell within the thymus becomes neoplastic, the disease that develops is a lymphoma. The majority of lymphoid tumors that develop in the thymus are derived from the precursor T cells, and the tumor is a precursor T cell lymphoblastic lymphoma (Chap. 110). Rare B cells exist in the thymus, and when they become neoplastic, the tumor is a mediastinal (thymic) B cell lymphoma (Chap. 110). Hodgkin's disease, particularly the nodular sclerosing subtype, often involves the anterior mediastinum. Extranodal marginal zone (MALT) lymphomas have been reported to involve the thymus in the setting of Sjögren's syndrome or other autoimmune disorders, and the lymphoma cells often express IgA instead of IgM on their surface. Castleman's disease can involve the thymus. Germ cell tumors and carcinoid tumors occasionally may arise in the thymus. If the epithelial cells of the thymus become neoplastic, the tumor that develops is a *thymoma*.

Clinical Presentation and Differential Diagnosis

Thymoma is the most common cause of an anterior mediastinal mass in adults, accounting for ~40% of all mediastinal masses. The other major causes of anterior mediastinal masses are lymphomas, germ cell tumors, and substernal thyroid tumors. Carcinoid tumors, lipomas, and thymic cysts also may produce radiographic masses. After combination chemotherapy for another malignancy, teenagers and young adults may develop a rebound thymic hyperplasia in the first few months after treatment. Granulomatous inflammatory diseases (tuberculosis, sarcoidosis) can produce thymic enlargement. Thymomas are most common in the fifth and sixth decades, are uncommon in children, and are distributed evenly between men and women.

About 40–50% of patients are asymptomatic; masses are detected incidentally on routine chest radiographs. When symptomatic, patients may have cough, chest pain, dyspnea, fever, wheezing, fatigue, weight loss, night sweats, or anorexia. Occasionally, thymomas may obstruct the superior vena cava. About 40% of patients with thymoma have another systemic autoimmune illness related to the thymoma. About 30% of patients with thymoma have myasthenia gravis, 5–8% have pure red cell aplasia, and ~5% have hypogammaglobulinemia. Thymoma with hypogammaglobulinemia also is called Good's syndrome. Among patients with myasthenia gravis, ~10–15% have a thymoma. Thymoma more rarely may be associated with polymyositis, systemic lupus erythematosus, thyroiditis, Sjögren's syndrome, ulcerative colitis, pernicious anemia, Addison's disease, scleroderma, and panhypopituitarism. In one series, 70% of patients with thymoma were found to have another systemic illness.

Diagnosis and Staging

Once a mediastinal mass is detected, a surgical procedure is required for definitive diagnosis. An initial mediastinoscopy or limited thoracotomy can be undertaken to get sufficient tissue to make an accurate diagnosis. Fine-needle aspiration is poor at distinguishing between lymphomas and thymomas but is more reliable in diagnosing germ cell tumors and metastatic carcinoma. Thymomas and lymphomas require sufficient tissue to examine the tumor architecture to assure an accurate diagnosis and obtain prognostic information.

Once a diagnosis of thymoma is defined, subsequent staging generally occurs at surgery. However, chest CT scans can assess local invasiveness in some instances. MRI has a defined role in the staging of posterior mediastinal tumors, but it is not clear that it adds important information to the CT

scan in anterior mediastinal tumors. Somatostatin receptor imaging with indium-labeled somatostatin analogues may be of value. If invasion is not distinguished by noninvasive testing, an effort to resect the entire tumor should be undertaken. If invasion is present, neoadjuvant chemotherapy may be warranted before surgery (see "Approach to Treatment," below).

Some 90% of thymomas are in the anterior mediastinum, but some may be in other mediastinal sites or even the neck, based on aberrant migration of the developing thymic enlage.

The staging system for thymoma was developed by Masaoka and colleagues **(Table e20-1)**. It is an anatomic system in which the stage is increased on the basis of the degree of invasiveness. The 5-year survival of patients in the various stages is as follows: stage I, 96%; stage II, 86%; stage III, 69%; stage IV, 50%. The French Study Group on Thymic Tumors (GETT) has proposed modifications to the Masaoka scheme based on the degree of surgical removal because the extent of surgery has been noted to be a prognostic indicator. In their system, stage I tumors are divided into A and B on the basis of whether the surgeon suspects adhesions to adjacent structures; stage III tumors are divided into A and B based on whether disease was subtotally resected or only biopsied. The concurrence between the two systems is high.

Table e20-1 Masaoka Staging System for Thymo-Mas

Stage	Diagnostic Criteria		
I	Macroscopically and mi-croscopically completely encapsulated; no invasion through cap-sule		
II			
IIA	Microscopic invasion out-side the capsule		
IIB	Macroscopic invasion into surrounding fat or grossly adherent to pleura or pericardium		
III			
IIIA	Macroscopic invasion into neighboring organs, pericardium, or pleura but not great ves-sels		
IIIB	Macroscopic invasion into neighboring organs that includes great ves-sels		
IV			
IVA	Pleural or pericardial dis-semination		
IVB	Lymphatic or hematoge-nous metasta-ses		
	Stage Distribution, %	**5-Year Survival, %**	**10-Year Survival, %**
I	36	95–100	86–100
II	26	70–100	50–100
III	22	68–89	47–60
IV	10	47–69	0–11

Source: From A Masaoka et al: Cancer 48:2485, 1981. Updated from S Tomaszek et al: Ann Thorac Surg 87:1973, 2009, and CB Falkson et al: J Thorac Oncol 4:911, 2009

Pathology and Etiology

Thymomas are epithelial tumors, and all of them have malignant potential. It is not worthwhile to try to divide them into benign and malignant forms; the key prognostic feature is whether they are noninvasive or invasive. About 65% of thymomas are encapsulated and noninvasive, and about 35% are invasive. They may have a variable percentage of lymphocytes within the tumor, but genetic studies suggest that the lymphocytes are benign polyclonal cells. The epithelial component of the tumor may consist primarily of round or oval cells derived mainly from the cortex or spindle-shaped cells derived mainly from the medulla or combinations of the two types **(**Table e20-2**)**. Cytologic features are not reliable predictors of biologic behavior. About 90% of A, AB, and B1 tumors are localized. A very small number of patients have aggressive histology features characteristic of carcinomas. Thymic carcinomas are invasive and have a poor prognosis.

Table e20-2 World Health Organization (WHO) Histologic Classification of Thymus Tumors

Type	Histologic Description
A	Medullary thymoma
AB	Mixed thymoma
B1	Predominantly cortical thymoma

B2	Cortical thymoma	
B3	Well-differentiated thymic carcinoma	
C	Thymic carcinoma	
Type	**Distribution, %**	**Prognosis (10-year disease-free survival), %**
A	8	100
AB	26	90—100
B1	15	78—94
B2	28	83
B3	15	36
C	8	0–35

Source: From S Tomaszek et al: Ann Thorac Surg 87:1973, 2009.

The genetic lesions in thymomas are not well characterized. Some data suggest that Epstein-Barr virus may be associated with thymomas. Some tumors overexpress the p21 *ras* gene product. However, molecular pathogenesis remains undefined. A thymoma susceptibility locus has been defined on *rat* chromosome 7, but the relationship between this gene locus, termed *Tsr1*, and human thymoma has not been examined.

Treatment: Thymoma

Treatment is determined by the stage of disease. For patients with encapsulated tumors and stage I disease, complete resection is sufficient to cure 96% of patients. For patients with stage II disease, complete resection may be followed by 30–60 Gy of postoperative radiation therapy to the site of the primary tumor. However, the value of radiation therapy in this setting has not been established. The main predictors of long-term survival are Masaoka stage and completeness of resection. For patients with stage III and IV disease, the use of neoadjuvant chemotherapy followed by radical surgery, with or without additional radiation therapy, and additional consolidation chemotherapy has been associated with excellent survival. Chemotherapy regimens that are most effective generally include a platinum compound (either cisplatin or carboplatin) and an anthracycline. Addition of cyclophosphamide, vincristine, and prednisone seems to improve response rates. Response rates of 50–93% have been reported in series of patients each of which involved fewer than 40 patients. A single most effective regimen has not been defined. If surgery after neoadjuvant chemotherapy fails to produce a complete resection of residual disease, radiation therapy (50–60 Gy) may help reduce recurrence rates.

This multimodality approach appears to be superior to the use of surgery followed by radiation therapy alone, which produces a 5-year survival of ≤50% in patients with advanced-stage disease.

Some thymic carcinomas express *c-kit*, and one patient whose *c-kit* locus was mutated responded dramatically to imatinib. Many thymomas express epidermal growth factor receptors, but the antibodies to the receptor and the kinase inhibitors that block its action have not been evaluated systematically. Octreotide plus prednisone produces responses in about one-third of patients.

Influence of Thymectomy on the Course of Accompanying Diseases

Patients with myasthenia gravis have a high incidence of thymic abnormalities (~80%), but overt thymoma is present in only ~10–15% of patients with myasthenia gravis. It is thought that the thymus plays a role in breaking self-tolerance and generating T cells that recognize the acetylcholine receptor as a foreign antigen. Although patients with thymoma and myasthenia gravis are less likely to have a remission in the myasthenia as a consequence of thymectomy than are patients with thymic abnormalities other than thymoma, the course of myasthenia gravis is not significantly different in patients with or without thymoma. Thymectomy produces at least some symptomatic improvement in ~65% of patients with myasthenia gravis. In one large series, thymoma patients with myasthenia gravis had a better long-term survival from thymoma resection than did those without myasthenia gravis.

About 30–50% of patients with pure red cell aplasia have a thymoma. Thymectomy results in the resolution of pure red cell aplasia in ~30% of patients. About 10% of patients with hypogammaglobulinemia have a thymoma, but hypogammaglobulinemia rarely responds to thymectomy.

FURTHER READINGS

Bril V et al: The long-term clinical outcome of myasthenia gravis in patients with thymoma. Neurology 51:1198, 1998[PMID: 9781560]

Chen YD et al: Role of adjuvant radiotherapy for stage II thymoma after complete tumor resection. Int J Radiat Oncol Biol Phys 2010 [Epub ahead of print]

Falkson CB et al: The management of thymoma: A systematic review and practice guideline. J Thorac

Oncol 4:911, 2009[PMID: 19557895]

Kelesidis T, Yang O: Good's syndrome remains a mystery after 55 years: A systematic review of the scientific evidence. Clin Immunol 135:347, 2010[PMID: 20149753]

Margaritora S et al: Thirty-five-year follow-up analysis of clinical and pathologic outcomes of thymoma surgery. Ann Thorac Surg 89:245, 2010[PMID: 20103246]

Okereke IC et al: Prognostic indicators after surgery for thymoma. Ann Thorac Surg 89:1071, 2010[PMID: 20338309]

Rajan A, Giaccone G: Treatment of advanced thymoma and thymic carcinoma. Curr Treat Options Oncol 9:277, 2008 Shelly S et al: Thymoma and autoimmunity. Cell Mol Immunol Feb 14, 2011 [Epub ahead of print]

Shin DM et al: A multidisciplinary approach to therapy for unresectable malignant thymoma. Ann Intern Med 129:100, 1998[PMID: 9669967]

Souadjian JV et al: The spectrum of diseases associated with thymoma. Arch Intern Med 134:374, 1974[PMID: 4602050]

Wright CD: Management of thymomas. Crit Rev Oncol Hematol 65:109, 2008[PMID: 17570676]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LATE CONSEQUENCES OF CANCER AND ITS TREATMENT: INTRODUCTION

Over 10 million Americans are cancer survivors. The vast majority of these people will bear some mark of their cancer and/or its treatment, and a large proportion will experience long-term consequences that include medical problems, psychosocial dysfunction, economic hardship, sexual dysfunction, and discrimination in employment and insurance. Many of these problems are directly related to cancer treatment. As patients with more types of malignancies survive longer, the biologic toll that very imperfect therapies take in terms of morbidity and mortality rates is being recognized increasingly. These consequences of therapy confront the patients and the cancer specialists and general internists who manage them every day. Although long-term survivors of childhood leukemias, Hodgkin's lymphoma, and testicular cancer have increased knowledge about the consequences of cancer treatment, researchers and physicians keep learning more as patients survive longer with newer therapies. The pace of the development of therapies that mitigate treatment-related consequences has been slow, partly due to an understandable aversion to altering regimens that work and partly due to a lack of new, effective, less toxic therapeutic agents with less "collateral damage" to replace known agents with known toxicities. The types of damage from cancer treatment vary. Often, a final common pathway is irreparable damage to DNA. Surgery can create dysfunction, including blind gut loops that lead to absorption problems and loss of function of removed body parts. Radiation may damage end-organ function, for example, loss of potency in prostate cancer patients, pulmonary fibrosis, neurocognitive impairment, acceleration of atherosclerosis, and second cancers. Cancer chemotherapy may act as a carcinogen and has a kaleidoscope of other toxicities, as discussed in this chapter. **Table 102–1** lists the long-term effects of treatment.

Table 102–1 Late Effects of Cancer Therapy

Surgical Procedure	**Effect**	
Amputation	Functional loss	
Lymph node dissection		Risk of lymphedema
Ostomy	Psychosocial impact	
Splenectomy	Risk of sepsis	
Adhesions	Risk of obstruction	
Bowel anastomoses	Malabsorption syndromes	
Radiation Therapy	Effect	
Organ		
Bone	Premature termination of growth, osteonecrosis	
Soft tissues	Atrophy, fibrosis	
Brain	Neuropsychiatric deficits, cognitive dysfunction	
Thyroid	Hypothyroidism, Graves' disease, cancer	
Salivary glands	Dry mouth, caries, dysgeusia	
Eyes	Cataracts	
Heart	Pericarditis, myocarditis, coronary artery disease	
Lung	Pulmonary fibrosis	
Kidney	Decreased function, hypertension	
Liver	Decreased function	
Intestine	Malabsorption, stricture	
Gonads	Infertility, premature menopause	
Any	Secondary neoplasia	

Chemotherapy		**Effect**
Organ	**Drug**	
Bone	Glucocorticoids	Osteoporosis, avascular necrosis

Brain	Methotrexate, cytarabine (Ara-C), others	Neuropsychiatric deficits, cognitive decline?
Peripheral nerves	Vincristine, platinum, taxanes	Neuropathy, hearing loss
Eyes	Glucocorticoids	Cataracts
Heart	Anthracyclines, trastuzumab	Cardiomyopathy
Lung	Bleomycin	Pulmonary fibrosis
	Methotrexate	Pulmonary hypersensitivity
Kidney	Platinum, others	Decreased function, hypomagnesemia
Liver	Various	Altered function
Gonads	Alkylating agents, others	Infertility, premature menopause
Bone marrow	Various	Aplasia, myelodysplasia, secondary leukemia

The first goal of therapy is to eradicate or control the malignancy. Late treatment consequences are, indeed, testimony to the increasing success of such treatment. Their occurrence sharply underlines the necessity to develop more effective therapies with less long-term morbidity and mortality. At the same time, a sense of perspective and relative risk is necessary; fear of long-term complications should not prevent the application of effective (particularly curative) cancer treatment.

CARDIOVASCULAR DYSFUNCTION

Chemotherapeutic Agents

The cardiovascular toxicity of cancer chemotherapeutic agents includes dysrhythmias, cardiomyopathic congestive heart failure (CHF), pericardial disease, and peripheral vascular disease. Because these cardiac toxicities are difficult to distinguish from disease that is not associated with cancer treatment, determining the clear etiologic implication of cancer chemotherapeutic agents may be difficult. Cardiovascular complications occurring in an unexpected clinical setting in patients who have undergone cancer therapy is often important in raising suspicion. Dose-dependent myocardial toxicity of anthracyclines with characteristic myofibrillar dropout is pathologically pathagnomonic on endomyocardial biopsy. Anthracycline cardiotoxity occurs through a root mechanism of chemical free-radical damage. Fe(III)-doxorubicin complexes damage DNA, nuclear and cytoplasmic membranes, and mitochondria. About 5% of patients receiving >450–550 mg/m^2 doxorubicin will develop CHF. Cardiotoxicity in relation to the dose of anthracycline is clearly not a step function but rather a continuous function, and occasional patients are seen with CHF at substantially lower doses. Advanced age, other concomitant cardiac disease, hypertension, diabetes, and thoracic radiation therapy are all important cofactors in promoting anthracycline-associated CHF. Anthracycline-related CHF is difficult to reverse; the mortality rate is as high as 50%, making prevention crucial. Some anthracyclines, such as mitoxantrone, are associated with less cardiotoxicity, and continuous infusion regimens or doxorubicin encapsulated in liposomes are associated with less cardiotoxicity. Dexrazoxane, an intracellular iron chelator, may limit anthracycline toxicity, but concern about limiting chemotherapeutic efficacy has limited its use. Monitoring patients for cardiac toxicity typically involves periodic gated nuclear cardiac blood pool ejection fraction testing [multi-gated acquisition scan (MUGA)] or cardiac ultrasonography. Cardiac MRI has been used but is not standard or widespread. Testing is performed more frequently at higher cumulative doses, with additional risk factors, certainly for any newly developing CHF or other symptoms of cardiac dysfunction.

Trastuzumab after anthracyclines is currently the next most commonly used cardiotoxic drug. Trastuzumab is commonly used in adjuvant therapy or for advanced HER2-positive breast cancer, sometimes in conjunction with anthracyclines, and is believed to result in additive or possibly synergistic toxicity. In contrast to anthracyclines, cardiotoxicity from trastuzumab is not dose-related, is usually reversible, is not associated with pathologic changes of anthracyclines on cardiac myofibrils, and has a different biochemical mechanism inhibiting intrinsic cardiac repair mechanisms. Toxicity is monitored routinely every three to four doses with functional cardiac testing as described above for anthracyclines.

Other cardiotoxic drugs include lapatinib, phosphoramide mustards (cyclophosphamide), ifosfamide, interleukin 2, imatinib, and sunitinib.

Radiation Therapy

Radiation therapy that includes the heart can cause interstitial myocardial fibrosis, acute and chronic pericarditis, valvular disease, and accelerated premature atherosclerotic coronary artery disease. Repeated or high (>6000 cGy) radiation doses are associated with greater risk, as is concomitant or distant cardiotoxic cancer chemotherapy exposure. Symptoms of acute pericarditis, which peaks about 9 months after treatment, include dyspnea, chest pain, and fever. Chronic constrictive pericarditis may develop 5–10 years after radiation therapy. Cardiac valvular disease includes aortic insufficiency from

fibrosis and papillary muscle dysfunction that results in mitral regurgitation. A threefold increased risk of fatal myocardial infarction is associated with mantle field radiation, with accelerated coronary artery disease. Carotid radiation similarly increases the risk of embolic stroke.

Treatment: Chemotherapeutic/Radiation-Induced Cardiovascular Disease

Therapy for chemotherapeutic/radiation-induced cardiovascular disease is essentially the same as therapy for disease that is not associated with cancer treatment. Discontinuation of the offending agent is the first step. Diuretics, fluid and sodium restriction, and antiarylthmic agents are often useful for treating acute symptoms. Afterload reduction with angiotensin-converting enzyme (ACE) inhibitors or, in some cases, β-adrenergic blockers (carvedilol) often is of significant benefit, and digitalis may be helpful as well.

Pulmonary Dysfunction

CHEMOTHERAPEUTIC AGENTS

Bleomycin generates activated free-radical oxygen species and causes pneumonitis associated with a radiographic or interstitial ground-glass appearance diffusely throughout both lungs, often worse in the lower lobes. This toxicity is dose-related and dose-limiting. The diffusion capacity of the lungs for carbon dioxide (DLCO) is a sensitive measure of toxicity and recovery, and a baseline value generally is obtained for future comparison before bleomycin therapy. Additive or synergistic risk factors include age, prior lung disease, and concomitant use of other chemotherapy, along with lung irradiation and high concentrations of inspired oxygen. Other chemotherapeutic agents notable for pulmonary toxicity include mitomycin, nitrosoureas, doxorubicin with radiation, gemcitabine combined with weekly docetaxel, methotrexate, and fludarabine. High-dose alkylating agents, cyclophosphamide, ifosfamide, and melphalan are used frequently in the hematopoietic stem cell transplant setting, often with whole-body radiation. This therapy may result in severe pulmonary fibrosis and/or pulmonary venoocclusive disease.

RADIATION THERAPY

Risk factors for radiation pneumonitis include advanced age, poor performance status, preexisting compromised pulmonary function, radiation volume, and dose. The dose "threshold" for lung damage is thought to be in the range of 5–20 Gy. Hypoxemia and dyspnea on exertion are characteristic. Fine, high-pitched "Velcro rales" may be an accompanying physical finding, and fever, cough, and pleuritic chest pain are common symptoms. The DLCO is the most sensitive measure of pulmonary functional impairment, and ground-glass infiltrates often correspond with relatively sharp edges to the irradiated volume, although the pneumonitis may progress beyond the field and occasionally involve the contralateral unirradiated lung.

Treatment: Pulmonary Dysfunction

Chemotherapy and radiation-induced pneumonitis are generally glucocorticoid-responsive, except in the case of nitrosoureas. Prednisone, 1 mg/kg, often is used to control acute symptoms and pulmonary dysfunction, with a generally slow taper. Prolonged glucocorticoid therapy requires gastrointestinal protection with proton pump inhibitors, management of hyperglycemia, heightened infection management, and prevention or treatment of steroid-induced osteoporosis. Antibiotics, bronchodilators, oxygen in only necessary doses, and diuretics may all play an important role in management of pneumonitis, and consultation with a pulmonologist should be undertaken routinely. Amifostine has been studied as a pulmonary radioprotectant, with inconclusive results, and is associated with skin rash, fatigue, and nausea; hence, it is not considered standard therapy at this time. Transforming growth factor β (TGF-β) is believed to be a major inducer of radiation fibrosis and represents a therapeutic target for development of anti-TGF-β therapies. As a general rule, patients who have received chest radiation therapy or bleomycin should receive supplemental O_2 when only absolutely necessary and at the lowest FiO_2 possible.

Neurologic Dysfunction

CHEMOTHERAPEUTIC AGENTS

Chemotherapy and radiation-induced neurologic dysfunction are increasing in both incidence and severity as a result of improved supportive care leading to more aggressive regimens and longer cancer survival allowing the development of late toxicity. Direct effects on myelin, glial cells, and neurons have been implicated, with alterations in cellular cytoskeleton, axonal transport, and cellular metabolism as mechanisms.

Vinca alkaloids produce a characteristic "stocking-glove" neuropathy, with numbness and tingling advancing to loss of motor function, which is highly dose-related. Distal sensorimotor polyneuropathy prominently involves loss of deep tendon reflexes with initially loss of pain and temperature sensation, followed by proprioceptive and vibratory loss. This requires a careful patient history and physical examination by experienced oncologists to decide when the drug must be stopped due to toxicity. Milder toxicity often slowly resolves completely. Vinca alkaloids may be associated with jaw claudication, autonomic neuropathy, ileus, cranial nerve palsies, and, in severe cases,

encephalopathy, seizures, and coma.

Cisplatin is associated with sensorimotor neuropathy as well as hearing loss, especially at doses >400 mg/m^2, requiring audiometry in patients with preexisting hearing compromise. Carboplatin often is substituted in such cases because of its lesser effect on hearing.

Neurocognitive dysfunction has been well described in childhood survivors of acute lymphocytic leukemia (ALL) treatment, including intrathecal methotrexate or cytarabine in conjunction with prophylactic cranial irradiation. Methotrexate alone may cause acute leucoencephalopathy characterized by somnolence and confusion that is often reversible. Acute toxicity is dose-related, especially at doses > 3 g/m^2, with younger patients being at greater risk. Subacute methotrexate toxicity occurs weeks after therapy and often is ameliorated with glucocorticoid therapy. Chronic methotrexate toxicity (leucoencephalopathy) develops months or years after treatment and is characterized clinically as progressive loss of cognitive function and focal neurologic signs that are irreversible, are promoted by synchronous or metachronous radiation therapy, and are more pronounced at a younger age.

Neurocognitive decline after chemotherapy alone occurs notably in breast cancer patients receiving adjuvant chemotherapy; this has been referred to as "chemo brain." It is clinically associated with impaired memory, learning, attention, and speed of information processing. The magnitude of the problem has been difficult to assess in light of the fact that cognitive decline is a normal feature of aging. It is not entirely clear that women receiving adjuvant therapy for breast cancer have a more rapid cognitive decline than do age-matched controls. Furthermore, its causes are unexplained given the poor penetrance of chemotherapy agents into the central nervous system (CNS). No prevention or therapy has been developed. This entity is attracting the attention of investigators.

Many cancer patients experience intrusive or debilitating concerns about cancer recurrence after successful therapy. In addition, these patients may experience job, insurance, stress, relationship, financial, and sexual difficulties. Physicians need to ask about and address these issues explicitly with cancer survivors, with referral to the appropriate counseling or support systems. Suicidal ideation and suicide have an increased incidence in cancer patients and survivors.

RADIATION THERAPY

Acute radiation CNS toxicity occurs within weeks and is characterized by nausea, drowsiness, hypersomnia, and ataxia, symptoms that most often abate with time. Early-delayed toxicity occurring weeks to 3 months after therapy is associated with symptoms similar to those of acute toxicity and is pathologically associated with reversible demyelination. Chronic, late radiation injury occurs 9 months to up to 10 years after therapy. Focal necrosis is a common pathologic finding, and glucocorticoid therapy may be helpful. Diffuse radiation injury is associated with global CNS neurologic dysfunction and diffuse white matter changes on CT or MRI. Pathologically, small vessel changes are prominent. Glucocorticoids may be symptomatically useful but do not alter the course. Necrotizing encephalopathy is the most severe form of radiation injury and almost always is associated with chemotherapy, notably methotrexate.

Cranial radiation also may be associated with an array of endocrine abnormalities with disruption of normal pituitary-hypothalamic axis function. A high index of suspicion needs to be maintained to identify and treat this toxicity.

Radiation-associated spinal cord injury (myelopathy) is highly dose-dependent and rarely occurs with modern radiation therapy. An early, self-limited form involving electric sensations down the spine on neck flexion (Lhermitte's sign) is seen 6–12 weeks after treatment, and generally resolves over weeks. Peripheral nerve toxicity is quite rare owing to relative radiation resistance.

Hepatic Dysfunction

CHEMOTHERAPEUTIC AGENTS

Long-term hepatic damage from standard chemotherapy regimens is rare. Long-term methotrexate or high-dose chemotherapy alone or with radiation therapy—e.g., in preparative regimens for bone marrow transplantation—may result in venoocclusive disease of the liver. This potentially lethal complication classically presents with anicteric ascites, elevated alkaline phosphatase, and hepatosplenomegaly. Pathologically, venous congestion, epithelial cell proliferation, and hepatocyte atrophy progressing to frank fibrosis are noted. Frequent monitoring of liver function tests during any type of chemotherapy is necessary to avoid both idiosyncratic and expected toxicities. Ursodiol may prevent cholestasis in the setting of high-dose therapy before bone marrow transplant.

RADIATION THERAPY

Hepatic radiation damage depends on dose, volume, fractionation, preexisting liver disease, and synchronous or metachronous chemotherapy. In general, radiation doses to the liver >1500 cGy can produce hepatic dysfunction with a steep dose-injury curve. Radiation-induced liver disease (RILD) closely mimics hepatic venoocclusive disease.

Renal/Bladder Dysfunction

Cisplatin produces reversible decrements in renal function but also may produce severe irreversible toxicity in the presence of renal disease and may predispose to accentuated damage with subsequent renal insults. Magnesium wasting with hypomagnesemia may be seen. Cyclophosphamide and ifosphamide, as prodrugs primarily activated in the liver, have cleavage products (acrolein) that can produce hemorrhagic cystitis. This can be prevented with the free-radical scavenger 2-mercaptoethane sulfonate (MESNA), which is required for ifosfamide administration. Hemorrhagic cystitis caused by these agents may predispose these patients to bladder cancer.

Reproductive and Endocrine Dysfunction

CHEMOTHERAPEUTIC AGENTS

Alkylating agents are associated with the highest rates of male and female infertility, which is directly dependent on age, dose, and duration of treatment. The age at treatment is an important determinant of fertility outcome, with prepubertal patients having the highest tolerance. Ovarian failure is age-related, and females who resume menses after treatment are still at increased risk for premature menopause. Males generally have reversible azospermia during lower-intensity alkylator chemotherapy, and long-term infertility is associated with total doses of cyclophosphamide >9 g/m^2 and with high-intensity therapy such as that used in hematopoietic stem cell transplantation. Males undergoing potentially sterilizing chemotherapy should be offered sperm banking. However, some cancers are associated with defective spermatogenesis. Gonadotropin-releasing hormone (GnRH) analogues to preserve ovarian function remain experimental. Assisted reproductive technologies may be helpful to couples with chemotherapy-induced infertility.

RADIATION THERAPY

The testicles and ovaries of prepubertal patients are less sensitive to radiation damage; spermatogenesis is affected by low doses of radiation, and complete azospermia occurs at 600–700 cGy. Leydig cell dysfunction, in contrast, occurs at <2000 cGy; hence, endocrine function is lost at much higher radiation doses than spermatogenesis. Erectile dysfunction occurs in up to 80% of men treated with external-beam radiation therapy for prostate cancer. Sildenafil (and congeners) may be useful in reversing erectile dysfunction. Ovarian function damage with radiation is age-related and occurs at doses of 150–500 cGY. Premature induction of menopause can have serious medical and psychological sequelae. Hormone replacement therapy often is contraindicated, as in estrogen receptor–positive breast cancer. Attention must be paid to maintenance of bone mass with calcium and vitamin D supplements and oral bisphosphonates; this is monitored by bone density determinations. Paroxetine, clonidine, pregabalin, and other drugs may be useful in symptomatically controlling hot flashes. In those without contraindications, estrogen may be used for periods <5 years at the lowest dose that relieves symptoms.

Long–term survivors of childhood cancer (e.g., ALL) who have received cranial radiation may have altered leptin biology and growth hormone deficiency, leading to obesity and reduced strength, exercise tolerance, and bone density.

Radiation therapy to the neck, for example, in Hodgkin's lymphoma, may lead to hypothyroidism, Graves' disease, thyroiditis, and thyroid malignancies. Thyroid-stimulating hormone (TSH) is followed routinely in these patients and suppressed with synthroid when elevated to prevent hypothyroidism and suppress the TSH drive, which may cause thyroid cancer.

Ocular Complications

Cataracts may be caused by glucocorticoids, depending on the duration and dose; radiation therapy; and, uncommonly, tamoxifen. Orbital radiation therapy may cause blindness.

Oral Complications

Radiation therapy can produce xerostomia (dry mouth) with an attendant increase in caries and poor dentition. Taste and appetite may be suppressed. Bisphosphonate use may result in osteonecrosis of the jaw.

Raynaud's Phenomenon

Up to 40% of patients treated with bleomycin may develop Raynaud's phenomenon. The mechanism is unknown.

SECOND MALIGNANCIES

Second malignancies in patients cured of cancer are a major cause of death, and treated cancer patients must be monitored for their occurrence. The induction of second malignancies is governed by the complex interplay of a number of factors, including age, sex, environmental exposures, genetic susceptibility, and cancer treatment itself. In a number of settings, the events leading to the primary cancer themselves increase the risk of second malignancies. Patients with lung cancer are at increased risk of esophageal and head and neck cancers, and vice versa, due to shared risk factors, including alcohol and tobacco abuse. Indeed, the risk of developing a second primary head and neck,

esophageal, or lung cancer is also increased in these patients. Patients with breast cancer are at increased risk of cancer in the opposite breast. Patients with Hodgkin's lymphoma are at risk for non-Hodgkin's lymphomas. Genetic cancer syndromes, for example, multiple endocrine neoplasia and Li-Fraumeni, Lynch, Cowden, and Gardner's syndromes, are examples of genetically based second malignancies of specific types. Cancer treatment itself does not appear to be responsible for the risk of these secondary malignancies. Deficient DNA repair can greatly increase the risk of cancers from DNA-damaging agents, as in ataxia-telangectasia. Importantly, the risk of treatment-related second malignancies is at least additive and often is synergistic with combined chemotherapy and radiation therapy; hence, for such combined therapy treatment approaches, it is important to establish the necessity of each modality in the treatment program. All these patients require special surveillance or in some cases prophylactic surgery as part of appropriate treatment and follow-up.

Chemotherapeutic Agents

Chemotherapy is significantly associated with two fatal second malignancies: acute leukemia and myelodysplastic syndromes. Two types of leukemia have been described. In patients treated with alkylating agents, acute myeloid leukemia is associated with deletions in chromosomes 5 or 7. The lifetime risk is about 1–5%, is increased by radiation therapy, and increases with age. The incidence of these leukemias peaks at 4–6 years, with risk returning close to baseline at 10 years. The other type of acute myeloid leukemia is related to therapy with topoisomerase inhibitors, is associated with chromosome 10q23 translocations, has an incidence <1%, and generally occurs 1.5–3 years after treatment. Both of these acute leukemias are refractory to treatment and have a high mortality rate. The development of myelodysplastic syndromes is increased after chemotherapy, particularly chronic alkylating agent therapy, and these syndromes often are associated with leukemic progression with a dismal prognosis.

Radiation Therapy

Patients receiving radiation have an increasing and lifelong risk of second malignancies that is 1–2% per year in the second decade after treatment but increases to >25% after 25 years. These malignancies include cancers of the thyroid and breast, sarcomas, and CNS cancers, which often tend to be aggressive and have a poor prognosis. An example of organ-, age-, and sex-dependent radiation-induced secondary malignancy is breast cancer, in which the risk is small with radiation 30 years. A 25-year-old woman treated with mantle radiation for Hodgkin's lymphoma has a 29% actuarial risk of developing breast cancer by age 55.

HORMONAL THERAPY

Treatment of breast cancer with tamoxifen for 5 years or longer is associated with a 1–2% risk of endometrial cancer. Surveillance is generally effective at finding these cancers at an early stage. The risk of mortality from tamoxifen-induced endometrial cancer is low compared to the benefit of tamoxifen as adjuvant therapy for breast cancer.

IMMUNOSUPRESSIVE THERAPY

Immunosupressive therapy, as used in allogeneic bone marrow transplantation, particularly with T cell depletion and using antithymocyte globulin or other means, increases the risk of Epstein Barr virus –associated B cell lymphoproliferative disorder. The incidence at 10 years after T cell depletion is 9 –12%. Discontinuation of immunosuppressive therapy, if possible, often is associated with complete disease regression.

RECOMMENDATIONS FOR FOLLOW-UP

All former cancer patients should be followed indefinitely. This is done most often by oncologists, but demographic changes suggest that more primary care physicians will need to be trained in the follow-up of treated cancer patients in remission. Cancer patients need to be educated about signs and symptoms of recurrence and potentially adverse effects related to therapy. Localized pain or palpable abnormality in a previously radiated field should prompt radiographic evaluation. Screening tests, when available and validated, should be used on a routine and regular basis, for example, mammography and Pap smear, particularly in patients receiving radiation to specific organs. Annual mammography should start no later than 10 years after breast radiation. Patients receiving radiation fields that encompass thyroid tissue should have regular thyroid exams and TSH testing. Patients treated with alkylating agents or topoisomerase inhibitors should have a complete blood count every 6–12 months, and cytopenias, abnormal cells on peripheral smear, or macrocytosis should be evaluated with bone marrow biopsy and aspirate, along with cytogenetics, flow cytometry, or fluorescence in situ hybridization (FISH) studies as appropriate.

Cancer survival, as the population lives longer and expands, has become an increasingly recognized subject. The Institute of Medicine and the National Research Council of the National Academy of Sciences have published a monograph titled *From Cancer Patient to Cancer Survivor: Lost in Transition.* It proposes a plan to inform clinicians caring for cancer survivors in complete detail about their previous treatments, the complications of those treatments, signs and symptoms of late effects, and recommended screening and follow-up procedures. **Table 102–2** describes long-term treatment

effects by cancer type.

Table 102–2 Long-Term Treatment Effects by Cancer Type

Cancer Type	Late Effects
Pediatric cancers	Majority have at least one late effect; 30% with moderate/severe problems Cardiovascular: radiation, anthracyclines Lungs: radiation Skeletal abnormalities: radiation Psychological, cognitive, and sexual problems Second neoplasms significant cause of death
Hodgkin's lymphoma	Thyroid dysfunction: radiation Premature coronary artery disease: radiation Gonadal dysfunction: chemotherapy Postsplenectomy sepsis Myelodysplasia Acute myeloid leukemia Non-Hodgkin's lymphomas Breast cancer, lung cancer, and melanoma Fatigue, psychological, and sexual problems Peripheral neuropathy
Non-Hodgkin's lymphoma	Myelodysplasia Acute leukemia Bladder cancer Peripheral neuropathy
Acute leukemia	Second malignancies: hematologic, solid tumors Neurophychiatric dysfunction Subnormal growth Thyroid abnormalities Infertility
Bone marrow stem cell transplantation	Infertility Graft-versus-host disease (allogeneic transplant) Psychosexual dysfunction
Head and neck cancer	Poor dentition, dry mouth, poor nutrition: radiation
Breast cancer	Tamoxifen: endometrial cancer, blood clots Aromatase inhibitors: osteoporosis, arthritis. Cardiomyopathy: anthracycline ± radiation, trastuzaumab Acute leukemia Hormone-deficiency symptoms: hot flashes, vaginal dryness, dyspareunia Psychosocial dysfunction "Chemo brain"
Testicular cancer	Raynaud's phenomenon Renal dysfunction Pulmonary dysfunction Retrograde ejaculation: surgery

	15% sexual dysfunction
Colon cancer	Major risk is second colon cancer Quality of life high in survivors
Prostate cancer	Impotence Urinary incontinence (0–15%) Chronic prostatitis/cystitis: radiation

OUTLOOK

Clearly, the challenge for the future is to combine chemotherapy, targeted agents, biologic therapies, radiation, and surgery to produce better outcomes with less toxicity, including late effects of therapy. This is easily said and less easily accomplished. As treatment becomes more effective in new patient populations (ovarian, bladder, anal, and laryngeal cancers, for example), one can expect to discover new populations at risk for late effects. These populations will have to be followed carefully so that such effects are recognized and treated. Cancer survivors represent an underutilized resource for prevention studies. Childhood cancer survivors especially have multiple chronic health impairments. The incidence of these late treatment consequences appears to have no plateau with age, throwing in stark relief the necessity of close monitoring and therapies with fewer late consequences of treatment.

FURTHER READINGS

Ahles T et al: Neuropsychiatric impact of standard-dose chemotherapy in long-term survivors of breast cancer and lymphoma. J Clin Oncol 20:485, 2002[PMID: 11786578]

Bookman MA et al: Late complications of curative treatment in Hodgkin's disease. JAMA 260:680, 1988[PMID: 3392794]

Brown LM et al: Risk of second non-hematologic malignancies among 376,825 breast cancer survivors. Breast Cancer Res Treat 106:439, 2007[PMID: 17277968]

Hewitt M et al (eds): *From Cancer Patient to Cancer Survivor: Lost in Transition* . Committee on Cancer Survivorship: Improving Care and Quality of Life. Washington, DC, National Academies Press, 2006

Liles A et al: Monitoring pulmonary complications in long-term childhood cancer survivors for the primary care physician. Cleve Clin J Med 75:531, 2008[PMID: 18646589]

———: Factors affecting late mortality from heart disease after treatment of Hodgkin's disease. JAMA 270:1949, 1993

Murphy BA: Advances in quality of life and symptom management for head and neck cancer patients. Curr Opin Oncol 21:242, 2009[PMID: 19322081]

Ng AK et al: Secondary malignancies across the age spectrum. Semin Radiat Oncol 20:67, 2010[PMID: 19959033]

Stubblefield MD, O'Dell MW (eds): *Cancer Rehabilitation: Principles and Practice* New York, Demos Medical, 2009

Travis LB et al: Cumulative absolute breast cancer risk for young women treated for Hodgkin lymphoma. J Natl Cancer Inst 97:1428, 2005[PMID: 16204692]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

IRON DEFICIENCY AND OTHER HYPOPROLIFERATIVE ANEMIAS: INTRODUCTION

Anemias associated with normocytic and normochromic red cells and an inappropriately low reticulocyte response (reticulocyte index <2–2.5) are *hypoproliferative anemias*. This category includes early iron deficiency (before hypochromic microcytic red cells develop), acute and chronic inflammation (including many malignancies), renal disease, hypometabolic states such as protein malnutrition and endocrine deficiencies, and anemias from marrow damage.

Marrow damage states are discussed in Chap. 107.

Hypoproliferative anemias are the most common anemias, and anemia associated with chronic inflammation is the most common of these. The anemia of inflammation, similar to iron deficiency, is related in part to abnormal iron metabolism. The anemias associated with renal disease, inflammation, cancer, and hypometabolic states are characterized by an abnormal erythropoietin response to the anemia.

IRON METABOLISM

Iron is a critical element in the function of all cells, although the amount of iron required by individual tissues varies during development. At the same time, the body must protect itself from free iron, which is highly toxic in that it participates in chemical reactions that generate free radicals such as singlet O_2 or OH^-. Consequently, elaborate mechanisms have evolved that allow iron to be made available for physiologic functions while at the same time conserving this element and handling it in such a way that toxicity is avoided.

The major role of iron in mammals is to carry O_2 as part of hemoglobin. O_2 is also bound by myoglobin in muscle. Iron is a critical element in iron-containing enzymes, including the cytochrome system in mitochondria. Iron distribution in the body is shown in **Table 103-1**. Without iron, cells lose their capacity for electron transport and energy metabolism. In erythroid cells, hemoglobin synthesis is impaired, resulting in anemia and reduced O_2 delivery to tissue.

Table 103-1 Body Iron Distribution

	Iron Content, mg	
	Adult Male, 80 kg	**Adult Female, 60 kg**
Hemoglobin	2500	1700
Myoglobin/enzymes	500	300
Transferrin iron	3	3
Iron stores	600–1000	0–300

The Iron Cycle in Humans

Figure 103-1 outlines the major pathways of internal iron exchange in humans. Iron absorbed from the diet or released from stores circulates in the plasma bound to *transferrin*, the iron transport protein. Transferrin is a bilobed glycoprotein with two iron binding sites. Transferrin that carries iron exists in two forms—*monoferric* (one iron atom) or *diferric* (two iron atoms). The turnover (half-clearance time) of transferrin-bound iron is very rapid—typically 60–90 min. Because almost all of the iron transported by transferrin is delivered to the erythroid marrow, the clearance time of transferrin-bound iron from the circulation is affected most by the plasma iron level and the erythroid marrow activity. When erythropoiesis is markedly stimulated, the pool of erythroid cells requiring iron increases and the clearance time of iron from the circulation decreases. The half-clearance time of iron in the presence of iron deficiency is as short as 10–15 min. With suppression of erythropoiesis, the plasma iron level typically increases and the half-clearance time may be prolonged to several hours. Normally, the iron bound to transferrin turns over 6–8 times per day. Assuming a normal plasma iron level of 80–100 μg/dL, the amount of iron passing through the transferrin pool is 20–24 mg/d.

Figure 103-1

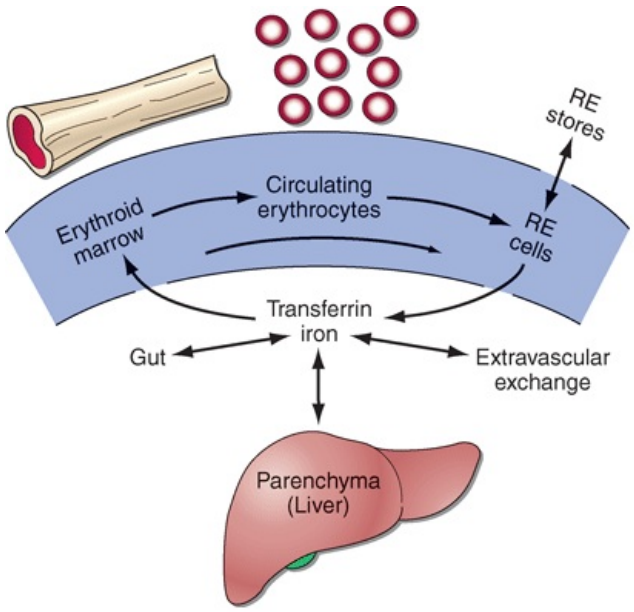

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Internal iron exchange. Normally approximately 80% of iron passing through the plasma transferrin pool is recycled from broken-down red cells. Absorption of approximately 1 mg/d is required from the diet in men, and 1.4 mg/d in women to maintain homeostasis. As long as transferrin saturation is maintained between 20–60% and erythropoiesis is not increased, use of iron stores is not required. However, in the event of blood loss, dietary iron deficiency, or inadequate iron absorption, up to 40 mg/d of iron can be mobilized from stores. RE, reticuloendothelial.

The iron-transferrin complex circulates in the plasma until it interacts with specific *transferrin receptors* on the surface of marrow erythroid cells. Diferric transferrin has the highest affinity for transferrin receptors; apotransferrin (not carrying iron) has very little affinity. Although transferrin receptors are found on cells in many tissues within the body—and all cells at some time during development will display transferrin receptors—the cell having the greatest number of receptors (300,000 to 400,000/cell) is the developing erythroblast.

Once the iron-bearing transferrin interacts with its receptor, the complex is internalized via clathrin-coated pits and transported to an acidic endosome, where the iron is released at the low pH. The iron is then made available for heme synthesis while the transferrin-receptor complex is recycled to the surface of the cell, where the bulk of the transferrin is released back into circulation and the transferrin receptor reanchors into the cell membrane. At this point a certain amount of the transferrin receptor protein may be released into circulation and can be measured as soluble transferrin receptor protein. Within the erythroid cell, iron in excess of the amount needed for hemoglobin synthesis binds to a storage protein, *apoferritin*, forming *ferritin*. This mechanism of iron exchange also takes place in other cells of the body expressing transferrin receptors, especially liver parenchymal cells where the iron can be incorporated into heme-containing enzymes or stored. The iron incorporated into hemoglobin subsequently enters the circulation as new red cells are released from the bone marrow. The iron is then part of the red cell mass and will not become available for reutilization until the red cell dies.

In a normal individual, the average red cell life span is 120 days. Thus, 0.8–1% of red cells turn over each day. At the end of its life span, the red cell is recognized as senescent by the cells of the *reticuloendothelial (RE) system*, and the cell undergoes phagocytosis. Once within the RE cell, the hemoglobin from the ingested red cell is broken down, the globin and other proteins are returned to the amino acid pool, and the iron is shuttled back to the surface of the RE cell, where it is presented to circulating transferrin. It is the efficient and highly conserved recycling of iron from senescent red cells that supports steady state (and even mildly accelerated) erythropoiesis.

Because each milliliter of red cells contains 1 mg of elemental iron, the amount of iron needed to replace those red cells lost through senescence amounts to 20 mg/d (assuming an adult with a red cell mass of 2 L). Any additional iron required for daily red cell production comes from the diet. Normally, an adult male will need to absorb at least 1 mg of elemental iron daily to meet needs, while females in the childbearing years will need to absorb an average of 1.4 mg/d. However, to achieve a maximum proliferative erythroid marrow response to anemia, additional iron must be available. With markedly stimulated erythropoiesis, demands for iron are increased by as much as six- to eightfold. With extravascular hemolytic anemia, the rate of red cell destruction is increased, but the iron recovered from the red cells is efficiently reutilized for hemoglobin synthesis. In contrast, with intravascular hemolysis or blood loss anemia, the rate of red cell production is limited by the amount of iron that can

be mobilized from stores. Typically, the rate of mobilization under these circumstances will not support red cell production more than 2.5 times normal. If the delivery of iron to the stimulated marrow is suboptimal, the marrow's proliferative response is blunted, and hemoglobin synthesis is impaired. The result is a hypoproliferative marrow accompanied by microcytic, hypochromic anemia.

Whereas blood loss or hemolysis places a demand on the iron supply, inflammatory conditions interfere with iron release from stores and can result in a rapid decrease in the serum iron (see below).

Nutritional Iron Balance

The balance of iron in humans is tightly controlled and designed to conserve iron for reutilization. There is no regulated excretory pathway for iron, and the only mechanisms by which iron is lost are blood loss (via gastrointestinal bleeding, menses, or other forms of bleeding) and the loss of epithelial cells from the skin, gut, and genitourinary tract. Normally, the only route by which iron comes into the body is via absorption from food or from medicinal iron taken orally. Iron may also enter the body through red-cell transfusions or injection of iron complexes. The margin between the amount of iron available for absorption and the requirement for iron in growing infants and the adult female is narrow; this accounts for the great prevalence of iron deficiency worldwide—currently estimated at one-half billion people.

The amount of iron required from the diet to replace losses averages approximately 10% of body iron content a year in men and 15% in women of childbearing age. Dietary iron content is closely related to total caloric intake (approximately 6 mg of elemental iron per 1000 calories). Iron bioavailability is affected by the nature of the foodstuff, with heme iron (e.g., red meat) being most readily absorbed. In the United States, the average iron intake in an adult male is 15 mg/d with 6% absorption; for the average female, the daily intake is 11 mg/d with 12% absorption. An individual with iron deficiency can increase iron absorption to approximately 20% of the iron present in a meat-containing diet but only 5–10% of the iron in a vegetarian diet. As a result, one-third of the female population in the United States has virtually no iron stores. Vegetarians are at an additional disadvantage because certain foodstuffs that include phytates and phosphates reduce iron absorption by approximately 50%. When ionizable iron salts are given together with food, the amount of iron absorbed is reduced. When the percentage of iron absorbed from individual food items is compared with the percentage for an equivalent amount of ferrous salt, iron in vegetables is only about one-twentieth as available, egg iron one-eighth, liver iron one-half, and heme iron one-half to two-thirds.

Infants, children, and adolescents may be unable to maintain normal iron balance because of the demands of body growth and lower dietary intake of iron. During the last two trimesters of pregnancy, daily iron requirements increase to 5–6 mg. That is the reason why iron supplements are strongly recommended for pregnant women in developed countries.

Iron absorption takes place largely in the proximal small intestine and is a carefully regulated process. For absorption, iron must be taken up by the luminal cell. That process is facilitated by the acidic contents of the stomach, which maintains the iron in solution. At the brush border of the absorptive cell, the ferric iron is converted to the ferrous form by a ferrireductase. Transport across the membrane is accomplished by divalent metal transporter type 1 [DMT-1, also known as natural resistance macrophage-associated protein type 2 (Nramp 2) or DCT-1]. DMT-1 is a general cation transporter. Once inside the gut cell, iron may be stored as ferritin or transported through the cell to be released at the basolateral surface to plasma transferrin through the membrane-embedded iron exporter, ferroportin. The function of ferroportin is negatively regulated by hepcidin, the principal iron regulatory hormone. In the process of release, iron interacts with another ferroxidase, hephaestin, which oxidizes the iron to the ferric form for transferrin binding. Hephaestin is similar to ceruloplasmin, the copper-carrying protein.

Iron absorption is influenced by a number of physiologic states. Erythroid hyperplasia stimulates iron absorption even in the face of normal or increased iron stores, and hepcidin levels are inappropriately low. The molecular mechanism underlying this relationship is not known. Thus, patients with anemias associated with high levels of ineffective erythropoiesis absorb excess amounts of dietary iron. Over time, this may lead to iron overload and tissue damage. In iron deficiency, hepcidin levels are low and iron is much more efficiently absorbed; the contrary is true in states of secondary iron overload. The normal individual can reduce iron absorption in situations of excessive intake or medicinal iron intake; however, while the percentage of iron absorbed goes down, the absolute amount goes up. This accounts for the acute iron toxicity occasionally seen when children ingest large numbers of iron tablets. Under these circumstances, the amount of iron absorbed exceeds the transferrin binding capacity of the plasma, resulting in free iron that affects critical organs such as cardiac muscle cells.

IRON-DEFICIENCY ANEMIA

Iron deficiency is one of the most prevalent forms of malnutrition. Globally, 50% of anemia is attributable to iron deficiency and accounts for approximately 841,000 deaths annually worldwide. Africa and parts of Asia bear 71% of the global mortality burden; North America represents only 1.4% of the total morbidity and mortality associated with iron deficiency.

Stages of Iron Deficiency

The progression to iron deficiency can be divided into three stages (**Fig. 103-2**). The first stage is *negative iron balance*, in which the demands for (or losses of) iron exceed the body's ability to absorb iron from the diet. This stage results from a number of physiologic mechanisms, including blood loss, pregnancy (in which the demands for red cell production by the fetus outstrip the mother's ability to provide iron), rapid growth spurts in the adolescent, or inadequate dietary iron intake. Blood loss in excess of 10–20 mL of red cells per day is greater than the amount of iron that the gut can absorb from a normal diet. Under these circumstances the iron deficit must be made up by mobilization of iron from RE storage sites. During this period, iron stores—reflected by the serum ferritin level or the appearance of stainable iron on bone marrow aspirations—decrease. As long as iron stores are present and can be mobilized, the serum iron, total iron-binding capacity (TIBC), and red cell protoporphyrin levels remain within normal limits. At this stage, red cell morphology and indices are normal.

Figure 103-2

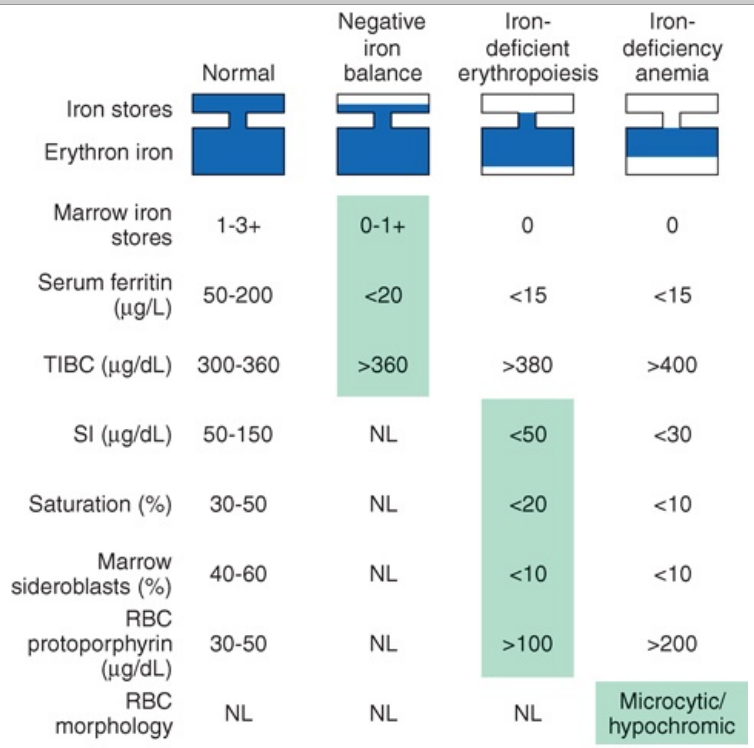

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Laboratory studies in the evolution of iron deficiency. Measurements of marrow iron stores, serum ferritin, and total iron-binding capacity (TIBC) are sensitive to early iron-store depletion. Iron-deficient erythropoiesis is recognized from additional abnormalities in the serum iron (SI), percent transferrin saturation, the pattern of marrow sideroblasts, and the red cell protoporphyrin level. Patients with iron-deficiency anemia demonstrate all the same abnormalities plus hypochromic microcytic anemia. *(From Hillman and Finch, with permission.)*

When iron stores become depleted, the serum iron begins to fall. Gradually, the TIBC increases, as do red cell protoporphyrin levels. By definition, marrow iron stores are absent when the serum ferritin level is <15 µg/L. As long as the serum iron remains within the normal range, hemoglobin synthesis is unaffected despite the dwindling iron stores. Once the transferrin saturation falls to 15–20%, hemoglobin synthesis becomes impaired. This is a period of *iron-deficient erythropoiesis*. Careful evaluation of the peripheral blood smear reveals the first appearance of microcytic cells, and if the laboratory technology is available, one finds hypochromic reticulocytes in circulation. Gradually, the hemoglobin and hematocrit begin to fall, reflecting *iron-deficiency anemia*. The transferrin saturation at this point is 10–15%.

When moderate anemia is present (hemoglobin 10–13 g/dL), the bone marrow remains hypoproliferative. With more severe anemia (hemoglobin 7–8 g/dL), hypochromia and microcytosis become more prominent, target cells and misshapen red cells (poikilocytes) appear on the blood smear as cigar- or pencil-shaped forms, and the erythroid marrow becomes increasingly ineffective. Consequently, with severe prolonged iron-deficiency anemia, erythroid hyperplasia of the marrow develops, rather than hypoproliferation.

Causes of Iron Deficiency

Conditions that increase demand for iron, increase iron loss, or decrease iron intake or absorption can produce iron deficiency (**Table 103-2**).

Table 103-2 Causes of Iron Deficiency

Increased Demand for Iron
Rapid growth in infancy or adolescence
Pregnancy
Erythropoietin therapy
Increased Iron Loss
Chronic blood loss
Menses
Acute blood loss
Blood donation
Phlebotomy as treatment for polycythemia vera
Decreased Iron Intake or Absorption
Inadequate diet
Malabsorption from disease (sprue, Crohn's disease)
Malabsorption from surgery (postgastrectomy)
Acute or chronic inflammation

Clinical Presentation of Iron Deficiency

Certain clinical conditions carry an increased likelihood of iron deficiency. Pregnancy, adolescence, periods of rapid growth, and an intermittent history of blood loss of any kind should alert the clinician to possible iron deficiency. A cardinal rule is that the appearance of iron deficiency in an adult male means gastrointestinal blood loss until proven otherwise. Signs related to iron deficiency depend on the severity and chronicity of the anemia in addition to the usual signs of anemia—fatigue, pallor, and reduced exercise capacity. *Cheilosis* (fissures at the corners of the mouth) and *koilonychia* (spooning of the fingernails) are signs of advanced tissue iron deficiency. The diagnosis of iron deficiency is typically based on laboratory results.

Laboratory Iron Studies

SERUM IRON AND TOTAL IRON-BINDING CAPACITY

The serum iron level represents the amount of circulating iron bound to transferrin. The TIBC is an indirect measure of the circulating transferrin. The normal range for the serum iron is 50–150 μg/dL; the normal range for TIBC is 300–360 μg/dL. Transferrin saturation, which is normally 25–50%, is obtained by the following formula: serum iron x 100 ÷ TIBC. Iron-deficiency states are associated with saturation levels below 20%. There is a diurnal variation in the serum iron. A transferrin saturation >50% indicates that a disproportionate amount of the iron bound to transferrin is being delivered to nonerythroid tissues. If this persists for an extended time, tissue iron overload may occur.

SERUM FERRITIN

Free iron is toxic to cells, and the body has established an elaborate set of protective mechanisms to bind iron in various tissue compartments. Within cells, iron is stored complexed to protein as ferritin or hemosiderin. Apoferritin binds to free ferrous iron and stores it in the ferric state. As ferritin accumulates within cells of the RE system, protein aggregates are formed as hemosiderin. Iron in ferritin or hemosiderin can be extracted for release by the RE cells, although hemosiderin is less readily available. Under steady-state conditions, the serum ferritin level correlates with total body iron stores; thus, the serum ferritin level is the most convenient laboratory test to estimate iron stores. The normal value for ferritin varies according to the age and gender of the individual (**Fig. 103-3**). Adult males have serum ferritin values averaging 100 μg/L, while adult females have levels averaging 30 μg/L. As iron stores are depleted, the serum ferritin falls to <15 μg/L. Such levels are diagnostic of absent body iron stores.

Figure 103-3

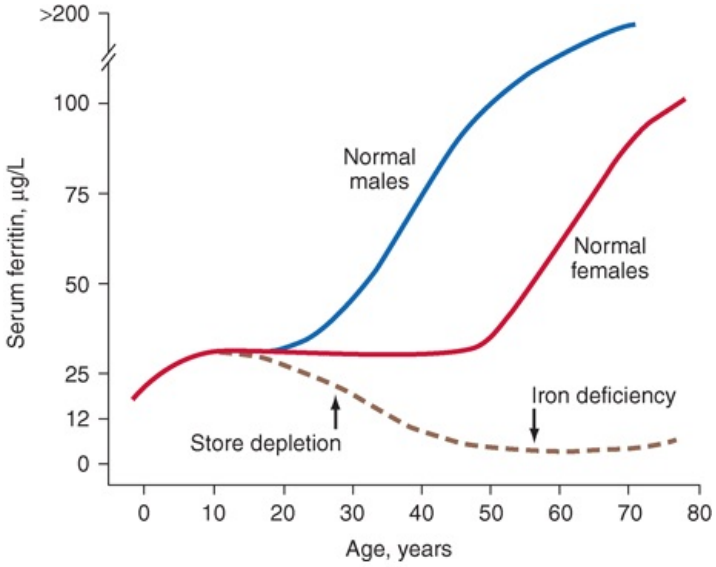

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Serum ferritin levels as a function of sex and age. Iron store depletion and iron deficiency are accompanied by a decrease in serum ferritin level below 20 µg/L. *(From Hillman et al, with permission.)*

EVALUATION OF BONE MARROW IRON STORES

Although RE cell iron stores can be estimated from the iron stain of a bone marrow aspirate or biopsy, the measurement of serum ferritin has largely supplanted bone marrow aspirates for determination of storage iron (**Table 103-3**). The serum ferritin level is a better indicator of iron overload than the marrow iron stain. However, in addition to storage iron, the marrow iron stain provides information about the effective delivery of iron to developing erythroblasts. Normally, when the marrow smear is stained for iron, 20–40% of developing erythroblasts—called *sideroblasts*—will have visible ferritin granules in their cytoplasm. This represents iron in excess of that needed for hemoglobin synthesis. In states in which release of iron from storage sites is blocked, RE iron will be detectable, and there will be few or no sideroblasts. In the myelodysplastic syndromes, mitochondrial dysfunction can occur, and accumulation of iron in mitochondria appears in a necklace fashion around the nucleus of the erythroblast. Such cells are referred to as *ringed sideroblasts*.

Table 103-3 Iron Store Measurements

Iron Stores	Marrow Iron Stain, 0-4+	Serum Ferritin, µg/L
0	0	<15
1–300 mg	Trace to 1+	15–30
300–800 mg	2+	30–60
800–1000 mg	3+	60–150
1–2 g	4+	>150
Iron overload	—	>500–1000

RED CELL PROTOPORPHYRIN LEVELS

Protoporphyrin is an intermediate in the pathway to heme synthesis. Under conditions in which heme synthesis is impaired, protoporphyrin accumulates within the red cell. This reflects an inadequate iron supply to erythroid precursors to support hemoglobin synthesis. Normal values are <30 µg/dL of red cells. In iron deficiency, values in excess of 100 µg/dL are seen. The most common causes of increased red cell protoporphyrin levels are absolute or relative iron deficiency and lead poisoning.

SERUM LEVELS OF TRANSFERRIN RECEPTOR PROTEIN

Because erythroid cells have the highest numbers of transferrin receptors of any cell in the body, and because transferrin receptor protein (TRP) is released by cells into the circulation, serum levels of TRP reflect the total erythroid marrow mass. Another condition in which TRP levels are elevated is absolute iron deficiency. Normal values are 4–9 µg/L determined by immunoassay. This laboratory test is becoming increasingly available and, along with the serum ferritin, has been proposed to distinguish between iron deficiency and the anemia of chronic inflammation (see below).

Differential Diagnosis

Other than iron deficiency, only three conditions need to be considered in the differential diagnosis of a hypochromic microcytic anemia (**Table 103-4**). The first is an inherited defect in globin chain synthesis: the thalassemias. These are differentiated from iron deficiency most readily by serum iron values; normal or increased serum iron levels and transferrin saturation are characteristic of the thalassemias. In addition, red blood cell distribution width (RDW) index is generally small in thalassemia and elevated in iron deficiency.

Table 103-4 Diagnosis of Microcytic Anemia

Tests	Iron Deficiency	Inflammation	Thalassemia	Sideroblastic Anemia
Smear	Micro/hypo	Normal micro/hypo	Micro/hypo with targeting	Variable
SI	<30	<50	Normal to high	Normal to high
TIBC	>360	<300	Normal	Normal
Percent saturation	<10	10–20	30–80	30–80
Ferritin (μg/L)	<15	30–200	50–300	50–300
Hemoglobin pattern on electrophoresis	Normal	Normal	Abnormal with β thalassemia; can be normal with α thalassemia	Normal

Abbreviations: SI, serum iron; TIBC, total iron-binding capacity.

The second condition is the anemia of chronic inflammation with inadequate iron supply to the erythroid marrow. The distinction between true iron-deficiency anemia and the anemia associated with chronic inflammation is among the most common diagnostic problem encountered by clinicians (see below). Usually the anemia of chronic inflammation is normocytic and normochromic. The iron values usually make the differential diagnosis clear, as the ferritin level is normal or increased and the percent transferrin saturation and TIBC are typically below normal.

Finally, the myelodysplastic syndromes represent the third and least common condition. Occasionally, patients with myelodysplasia have impaired hemoglobin synthesis with mitochondrial dysfunction, resulting in impaired iron incorporation into heme. The iron values again reveal normal stores and more than an adequate supply to the marrow, despite the microcytosis and hypochromia.

Treatment: Iron-Deficiency Anemia

The severity and cause of iron-deficiency anemia will determine the appropriate approach to treatment. As an example, symptomatic elderly patients with severe iron-deficiency anemia and cardiovascular instability may require red cell transfusions. Younger individuals who have compensated for their anemia can be treated more conservatively with iron replacement. The foremost issue for the latter patient is the precise identification of the cause of the iron deficiency.

For the majority of cases of iron deficiency (pregnant women, growing children and adolescents, patients with infrequent episodes of bleeding, and those with inadequate dietary intake of iron), oral iron therapy will suffice. For patients with unusual blood loss or malabsorption, specific diagnostic tests and appropriate therapy take priority. Once the diagnosis of iron-deficiency anemia and its cause is made, there are three major therapeutic approaches.

RED CELL TRANSFUSION

Transfusion therapy is reserved for individuals who have symptoms of anemia, cardiovascular instability, continued and excessive blood loss from whatever source, and require immediate intervention. The management of these patients is less related to the iron deficiency than it is to the consequences of the severe anemia. Not only do transfusions correct the anemia acutely, but the transfused red cells provide a source of iron for reutilization, assuming they are not lost through continued bleeding. Transfusion therapy will stabilize the patient while other options are reviewed.

ORAL IRON THERAPY

In the asymptomatic patient with established iron-deficiency anemia, treatment with oral iron is usually adequate. Multiple preparations are available, ranging from simple iron salts to complex iron compounds designed for sustained release throughout the small intestine (**Table 103-5**). Although the various preparations contain different amounts of iron, they are generally all absorbed well and are effective in treatment. Some come with other compounds designed to enhance iron absorption, such as ascorbic acid. It is not clear whether the benefits of such compounds justify their costs. Typically, for iron replacement therapy, up to 300 mg of elemental iron per day is given, usually as three or four iron tablets (each containing 50–65 mg elemental iron) given over the course of the day. Ideally, oral iron preparations should be taken on an empty stomach, since food may inhibit iron absorption. Some

patients with gastric disease or prior gastric surgery require special treatment with iron solutions, as the retention capacity of the stomach may be reduced. The retention capacity is necessary for dissolving the shell of the iron tablet before the release of iron. A dose of 200–300 mg of elemental iron per day should result in the absorption of iron up to 50 mg/d. This supports a red cell production level of two to three times normal in an individual with a normally functioning marrow and appropriate erythropoietin stimulus. However, as the hemoglobin level rises, erythropoietin stimulation decreases, and the amount of iron absorbed is reduced. The goal of therapy in individuals with iron-deficiency anemia is not only to repair the anemia, but also to provide stores of at least 0.5–1 g of iron. Sustained treatment for a period of 6–12 months after correction of the anemia will be necessary to achieve this.

Table 103-5 Oral Iron Preparations

Generic Name	Tablet (Iron Content), mg	Elixir (Iron Content), mg in 5 mL
Ferrous sulfate	325 (65) 195 (39)	300 (60) 90 (18)
Extended release	525 (105)	
Ferrous fumarate	325 (107) 195 (64)	 100 (33)
Ferrous gluconate	325 (39)	300 (35)
Polysaccharide iron	150 (150) 50 (50)	100 (100)

Of the complications of oral iron therapy, gastrointestinal distress is the most prominent and is seen in 15–20% of patients. Abdominal pain, nausea, vomiting, or constipation may lead to noncompliance. Although small doses of iron or iron preparations with delayed release may help somewhat, the gastrointestinal side effects are a major impediment to the effective treatment of a number of patients.

The response to iron therapy varies, depending on the erythropoietin stimulus and the rate of absorption. Typically, the reticulocyte count should begin to increase within 4–7 days after initiation of therapy and peak at 1–11/2 weeks. The absence of a response may be due to poor absorption, noncompliance (which is common), or a confounding diagnosis. A useful test in the clinic to determine the patient's ability to absorb iron is the *iron tolerance test.* Two iron tablets are given to the patient on an empty stomach, and the serum iron is measured serially over the subsequent 2 hours. Normal absorption will result in an increase in the serum iron of at least 100 μg/dL. If iron deficiency persists despite adequate treatment, it may be necessary to switch to parenteral iron therapy.

PARENTERAL IRON THERAPY

Intravenous iron can be given to patients who are unable to tolerate oral iron; whose needs are relatively acute; or who need iron on an ongoing basis, usually due to persistent gastrointestinal blood loss. Parenteral iron use has been increasing rapidly in the last several years with the recognition that recombinant erythropoietin (EPO) therapy induces a large demand for iron—a demand that frequently cannot be met through the physiologic release of iron from RE sources or oral iron absorption. The safety of parenteral iron—particularly iron dextran—has been a concern. The serious adverse reaction rate to intravenous high-molecular weight iron dextran is 0.7%. Fortunately, newer iron complexes are available in the United States, such as sodium ferric gluconate (Ferrlecit) and iron sucrose (Venofer) that have much lower rates of adverse effects.

Parenteral iron is used in two ways: one is to administer the total dose of iron required to correct the hemoglobin deficit and provide the patient with at least 500 mg of iron stores; the second is to give repeated small doses of parenteral iron over a protracted period. The latter approach is common in dialysis centers, where it is not unusual for 100 mg of elemental iron to be given weekly for 10 weeks to augment the response to recombinant EPO therapy. The amount of iron needed by an individual patient is calculated by the following formula:

Body weight (kg) x 2.3 x (15–patient's hemoglobin, g/dL) + 500 or 1000 mg (for stores).

In administering intravenous irondextran, anaphylaxis is a concern. Anaphylaxis is much rarer with the newer preparations. The factors that have correlated with an anaphylactic-like reaction include a history of multiple allergies or a prior allergic reaction to dextran (in the case of iron dextran). Generalized symptoms appearing several days after the infusion of a large dose of iron can include arthralgias, skin rash, and low-grade fever. These may be dose-related, but they do not preclude the further use of parenteral iron in the patient. To date, patients with sensitivity to iron dextran have been safely treated with iron gluconate. If a large dose of iron dextran is to be given (>100 mg), the iron preparation should be diluted in 5% dextrose in water or 0.9% NaCl solution. The iron solution can then be infused over a 60- to 90-minute period (for larger doses) or at a rate convenient for the

attending nurse or physician. Although a test dose (25 mg) of parenteral iron dextran is recommended, in reality a slow infusion of a larger dose of parenteral iron solution will afford the same kind of early warning as a separately injected test dose. Early in the infusion of iron, if chest pain, wheezing, a fall in blood pressure, or other systemic symptoms occur, the infusion of iron should be stopped immediately.

OTHER HYPOPROLIFERATIVE ANEMIAS

In addition to mild to moderate iron-deficiency anemia, the hypoproliferative anemias can be divided into four categories: (1) chronic inflammation, (2) renal disease, (3) endocrine and nutritional deficiencies (hypometabolic states), and (4) marrow damage (Chap. 107). With chronic inflammation, renal disease, or hypometabolism, endogenous EPO production is inadequate for the degree of anemia observed. For the anemia of chronic inflammation, the erythroid marrow also responds inadequately to stimulation, due in part to defective *iron reutilization*. As a result of the lack of adequate EPO stimulation, an examination of the peripheral blood smear will disclose only an occasional polychromatophilic ("shift") reticulocyte. In cases of iron deficiency or marrow damage, appropriate elevations in endogenous EPO levels are typically found, and shift reticulocytes will be present on the blood smear.

Anemia of Acute and Chronic Inflammation/Infection (the Anemia of Inflammation)

The anemia of inflammation—which encompasses inflammation, infection, tissue injury, and conditions (such as cancer) associated with the release of proinflammatory cytokines—is one of the most common forms of anemia seen clinically. It is the most important anemia in the differential diagnosis of iron deficiency, because many of the features of the anemia are brought about by inadequate iron delivery to the marrow, despite the presence of normal or increased iron stores. This is reflected by a low serum iron, increased red cell protoporphyrin, a hypoproliferative marrow, transferrin saturation in the range of 15–20%, and a normal or increased serum ferritin. The serum ferritin values are often the most distinguishing features between true iron-deficiency anemia and the iron-restricted erythropoiesis associated with inflammation. Typically, serum ferritin values increase threefold over basal levels in the face of inflammation. These changes are due to the effects of inflammatory cytokines and hepcidin, the key iron regulatory hormone, acting at several levels of erythropoiesis (**Fig. 103-4**).

Figure 103-4

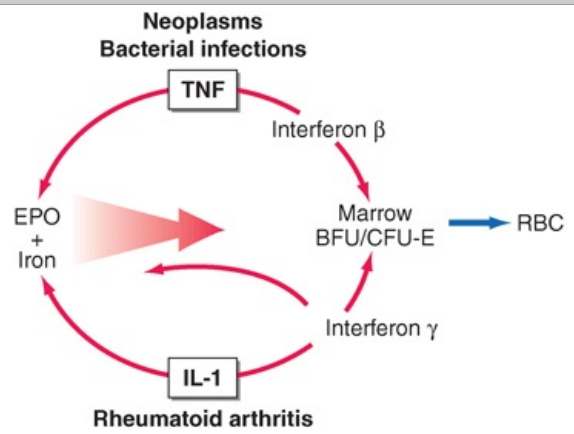

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Suppression of erythropoiesis by inflammatory cytokines. Through the release of tumor necrosis factor (TNF) and interferon γ (IFN-γ), neoplasms and bacterial infections suppress erythropoietin (EPO) production and the proliferation of erythroid progenitors [erythroid burst-forming units and erythroid colony-forming units (BFU/CFU-E)]. The mediators in patients with vasculitis and rheumatoid arthritis include interleukin 1 (IL-1) and IFN-γ The red arrows indicate sites of inflammatory cytokine inhibitory effects.

Interleukin 1 (IL-1) directly decreases EPO production in response to anemia. IL-1, acting through accessory cell release of interferon γ (IFN-γ), suppresses the response of the erythroid marrow to EPO—an effect that can be overcome by EPO administration in vitro and in vivo. In addition, tumor necrosis factor (TNF), acting through the release of IFN-γ by marrow stromal cells, also suppresses the response to EPO. Hepcidin, made by the liver, is increased in inflammation and acts to suppress iron absorption and iron release from storage sites. The overall result is a chronic hypoproliferative anemia with classic changes in iron metabolism. The anemia is further compounded by a mild to moderate shortening in red cell survival.

With chronic inflammation, the primary disease will determine the severity and characteristics of the anemia. For example, many patients with cancer also have anemia that is typically normocytic and

normochromic. In contrast, patients with long-standing active rheumatoid arthritis or chronic infections such as tuberculosis will have a microcytic, hypochromic anemia. In both cases, the bone marrow is hypoproliferative, but the differences in red cell indices reflect differences in the availability of iron for hemoglobin synthesis. Occasionally, conditions associated with chronic inflammation are also associated with chronic blood loss. Under these circumstances, a bone marrow aspirate stained for iron may be necessary to rule out absolute iron deficiency. However, the administration of iron in this case will correct the iron deficiency component of the anemia and leave the inflammatory component unaffected.

The anemia associated with acute infection or inflammation is typically mild but becomes more pronounced over time. Acute infection can produce a decrease in hemoglobin levels of 2–3 g/dL within 1 or 2 days; this is largely related to the hemolysis of red cells near the end of their natural life span. The fever and cytokines released exert a selective pressure against cells with more limited capacity to maintain the red cell membrane. In most individuals the mild anemia is reasonably well tolerated, and symptoms, if present, are associated with the underlying disease. Occasionally, in patients with preexisting cardiac disease, moderate anemia (hemoglobin 10–11 g/dL) may be associated with angina, exercise intolerance, and shortness of breath. The erythropoietic profile that distinguishes the anemia of inflammation from the other causes of hypoproliferative anemias is shown in **Table 103-6**.

Table 103-6 Diagnosis of Hypoproliferative Anemias

Tests	Iron Deficiency	Inflammation	Renal Disease	Hypometabolic States
Anemia	Mild to severe	Mild	Mild to severe	Mild
MCV (fL)	60–90	80–90	90	90
Morphology	Normo-microcytic	Normocytic	Normocytic	Normocytic
SI	<30	<50	Normal	Normal
TIBC	>360	<300	Normal	Normal
Saturation (%)	<10	10–20	Normal	Normal
Serum ferritin (?g/L)	<15	30–200	115–150	Normal
Iron stores	0	2–4+	1–4+	Normal

Abbreviations: MCV, mean corpuscular volume; SI, serum iron; TIBC, total iron-binding capacity.

Anemia of Chronic Kidney Disease (CKD)

Progressive CKD is usually associated with a moderate to severe hypoproliferative anemia; the level of the anemia correlates with the stage of CKD. Red cells are typically normocytic and normochromic, and reticulocytes are decreased. The anemia is primarily due to a failure of EPO production by the diseased kidney and a reduction in red cell survival. In certain forms of acute renal failure, the correlation between the anemia and renal function is weaker. Patients with the hemolytic-uremic syndrome increase erythropoiesis in response to the hemolysis, despite renal failure requiring dialysis. Polycystic kidney disease also shows a smaller degree of EPO deficiency for a given level of renal failure. By contrast, patients with diabetes or myeloma have more severe EPO deficiency for a given level of renal failure.

Assessment of iron status provides information to distinguish the anemia of CKD from the other forms of hypoproliferative anemia (Table 103-6) and to guide management. Patients with the anemia of CKD usually present with normal serum iron, TIBC, and ferritin levels. However, those maintained on chronic hemodialysis may develop iron deficiency from blood loss through the dialysis procedure. Iron must be replenished in these patients to ensure an adequate response to EPO therapy (see below).

Anemia in Hypometabolic States

Patients who are starving, particularly for protein, and those with a variety of endocrine disorders that produce lower metabolic rates, may develop a mild to moderate hypoproliferative anemia. The release of EPO from the kidney is sensitive to the need for O_2, not just O_2 levels. Thus, EPO production is triggered at lower levels of blood O_2 content in disease states (such as hypothyroidism and starvation) where metabolic activity, and thus O_2 demand, is decreased.

ENDOCRINE DEFICIENCY STATES

The difference in the levels of hemoglobin between men and women is related to the effects of androgen and estrogen on erythropoiesis. Testosterone and anabolic steroids augment erythropoiesis; castration and estrogen administration to males decrease erythropoiesis. Patients who are hypothyroid or have deficits in pituitary hormones also may develop a mild anemia. Pathogenesis may be complicated by other nutritional deficiencies because iron and folic acid absorption can be affected by these disorders. Usually, correction of the hormone deficiency reverses the anemia.

Anemia may be more severe in Addison's disease, depending on the level of thyroid and androgen hormone dysfunction; however, anemia may be masked by decreases in plasma volume. Once such patients are given cortisol and volume replacement, the hemoglobin level may fall rapidly. Mild anemia complicating hyperparathyroidism may be due to decreased EPO production as a consequence of the renal effects of hypercalcemia or to impaired proliferation of erythroid progenitors.

PROTEIN STARVATION

Decreased dietary intake of protein may lead to mild to moderate hypoproliferative anemia; this form of anemia may be prevalent in the elderly. The anemia can be more severe in patients with a greater degree of starvation. In marasmus, where patients are both protein and calorie deficient, the release of EPO is impaired in proportion to the reduction in metabolic rate; however, the degree of anemia may be masked by volume depletion and becomes apparent after refeeding. Deficiencies in other nutrients (iron, folate) may also complicate the clinical picture but may not be apparent at diagnosis. Changes in the erythrocyte indices on refeeding should prompt evaluation of iron, folate, and B_{12} status.

ANEMIA IN LIVER DISEASE

A mild hypoproliferative anemia may develop in patients with chronic liver disease from nearly any cause. The peripheral blood smear may show spur cells and stomatocytes from the accumulation of excess cholesterol in the membrane from a deficiency of lecithin-cholesterol acyltransferase. Red cell survival is shortened, and the production of EPO is inadequate to compensate. In alcoholic liver disease, nutritional deficiencies are common and complicate the management. Folate deficiency from inadequate intake, as well as iron deficiency from blood loss and inadequate intake, can alter the red cell indices.

Treatment: Hypoproliferative Anemias

Many patients with hypoproliferative anemias experience recovery of normal hemoglobin levels when the underlying disease is appropriately treated. For those in whom such reversals are not possible —such as patients with end-stage kidney disease, cancer, and chronic inflammatory diseases —symptomatic anemia requires treatment. The two major forms of treatment are transfusions and EPO.

TRANSFUSIONS

Thresholds for transfusion should be altered based on the patient's symptoms. In general, patients without serious underlying cardiovascular or pulmonary disease can tolerate hemoglobin levels above 8 g/dL and do not require intervention until the hemoglobin falls below that level. Patients with more physiologic compromise may need to have their hemoglobin levels kept above 11 g/dL. A typical unit of packed red cells increases the hemoglobin level by 1 g/dL. Transfusions are associated with certain infectious risks (Chap. 113), and chronic transfusions can produce iron overload. Importantly, the liberal use of blood has been associated with increased morbidity and mortality, particularly in the intensive care setting. Therefore, in the absence of documented tissue hypoxia, a conservative approach to the use of red cell transfusions is preferable.

ERYTHROPOIETIN (EPO)

EPO is particularly useful in anemias in which endogenous EPO levels are inappropriately low, such as CKD or the anemia of chronic inflammation. Iron status must be evaluated and iron repleted to obtain optimal effects from EPO. In patients with CKD, the usual dose of EPO is 50–150 U/kg three times a week intravenously. Hemoglobin levels of 10–12 g/dL are usually reached within 4–6 weeks if iron levels are adequate; 90% of these patients respond. Once a target hemoglobin level is achieved, the EPO dose can be decreased. A decrease in hemoglobin level occurring in the face of EPO therapy usually signifies the development of an infection or iron depletion. Aluminum toxicity and hyperparathyroidism can also compromise the EPO response. When an infection intervenes, it is best to interrupt the EPO therapy and rely on transfusion to correct the anemia until the infection is adequately treated. The dose needed to correct the anemia in patients with cancer is higher, up to 300 U/kg three times a week, and only approximately 60% of patients respond. Because of evidence that tumor progression may result from EPO administration, the risks and benefits of using EPO in patients with chemotherapy-induced anemia must be weighed carefully, and the target hemoglobin should be that necessary to avoid transfusions.

Longer-acting preparations of EPO can reduce the frequency of injections. Darbepoetin alfa, a molecularly modified EPO with additional carbohydrate, has a half-life in the circulation that is three to four times longer than recombinant human EPO, permitting weekly or every other week dosing.

FURTHER READINGS

Andrews, NC. Forging a field: The golden age of iron biology. Blood 112:219, 2008[PMID: 18606887]

Bailie GR et al: Parenteral iron use in the management of anemia in end-stage renal disease patients. Am J Kidney Dis 35:1, 2000[PMID: 10620537]

Brugnara C: Iron deficiency and erythropoiesis: New diagnostic approaches. Clin Chem 49:1573, 2003[PMID: 14500582]

Ferrucci L et al: Proinflammatory state, hepcidin, and anemia in older persons. Blood 115:3810, 2010[PMID: 20081092]

Ganz T: Hepcidin, a key regulator of iron metabolism and mediator of inflammation. Blood 102:783, 2003[PMID: 12663437]

Hillman RS et al: *Hematology in Clinical Practice,* 5th ed, New York, McGraw-Hill, 2010

Stoltzfus RF: Iron deficiency: Global prevalence and consequences. Food Nutr Bull 24:S99, 2003

Thomas C et al: The diagnostic plot: A concept for identifying different states of iron deficiency and monitoring the response to epoetin therapy. Med Oncol 23:23, 2006[PMID: 16645227]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF HEMOGLOBIN: INTRODUCTION

Hemoglobin is critical for normal oxygen delivery to tissues; it is also present in erythrocytes in such high concentrations that it can alter red cell shape, deformability, and viscosity. Hemoglobinopathies are disorders affecting the structure, function, or production of hemoglobin. These conditions are usually inherited and range in severity from asymptomatic laboratory abnormalities to death in utero. Different forms may present as hemolytic anemia, erythrocytosis, cyanosis, or vasoocclusive stigmata.

PROPERTIES OF THE HUMAN HEMOGLOBINS

Hemoglobin Structure

Different hemoglobins are produced during embryonic, fetal, and adult life **(Fig. 104-1)**. Each consists of a tetramer of globin polypeptide chains: a pair of α-like chains 141 amino acids long and a pair of β-like chains 146 amino acids long. The major adult hemoglobin, HbA, has the structure $\alpha_2\beta_2$. HbF ($\alpha_2\gamma_2$) predominates during most of gestation, and HbA_2 ($\alpha_2\delta_2$) is minor adult hemoglobin. Embryonic hemoglobins need not be considered here.

Figure 104-1

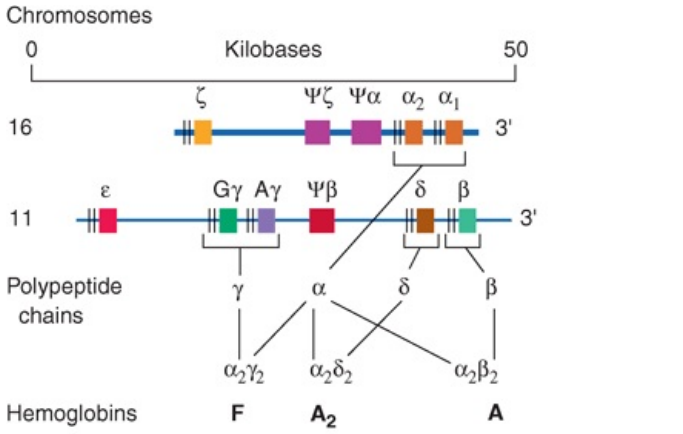

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The globin genes. The α-like genes (α,ζ) are encoded on chromosome 16; the β-like genes (β,γ,δ,ε) are encoded on chromosome 11. The ζ and ε genes encode embryonic globins.

Each globin chain enfolds a single heme moiety, consisting of a protoporphyrin IX ring complexed with a single iron atom in the ferrous state (Fe^{2+}). Each heme moiety can bind a single oxygen molecule; a molecule of hemoglobin can transport up to four oxygen molecules.

The amino acid sequences of the various globins are highly homologous to one another. Each has a highly helical *secondary structure*. Their globular *tertiary structures* cause the exterior surfaces to be rich in polar (hydrophilic) amino acids that enhance solubility, and the interior to be lined with nonpolar groups, forming a hydrophobic pocket into which heme is inserted. The tetrameric *quaternary structure* of HbA contains two αβ dimers. Numerous tight interactions (i.e., $\alpha_1\beta_1$ contacts) hold the α and β chains together. The complete tetramer is held together by interfaces (i.e., $\alpha_1\beta_2$ contacts) between the α-like chain of one dimer and the non-α chain of the other dimer.

The hemoglobin tetramer is highly soluble but individual globin chains are insoluble. Unpaired globin precipitates, forming inclusions that damage the cell. Normal globin chain synthesis is balanced so that each newly synthesized α or non-α globin chain will have an available partner with which to pair.

Solubility and reversible oxygen binding are the key properties deranged in hemoglobinopathies. Both depend most on the hydrophilic surface amino acids, the hydrophobic amino acids lining the heme pocket, a key histidine in the F helix, and the amino acids forming the $\alpha_1\beta_1$ and $\alpha_1\beta_2$ contact points.

Mutations in these strategic regions tend to be the ones that alter oxygen affinity or solubility.

Function of Hemoglobin

To support oxygen transport, hemoglobin must bind O_2 efficiently at the partial pressure of oxygen (PO_2) of the alveolus, retain it, and release it to tissues at the PO_2 of tissue capillary beds. Oxygen acquisition and delivery over a relatively narrow range of oxygen tensions depend on a property inherent in the tetrameric arrangement of heme and globin subunits within the hemoglobin molecule called *cooperativity* or *heme-heme interaction*.

At low oxygen tensions, the hemoglobin tetramer is fully deoxygenated **(Fig. 104-2)**. Oxygen binding begins slowly as O_2 tension rises. However, as soon as some oxygen has been bound by the tetramer, an abrupt increase occurs in the slope of the curve. Thus, hemoglobin molecules that have bound some oxygen develop a higher oxygen affinity, greatly accelerating their ability to combine with more oxygen. This S-shaped oxygen equilibrium curve (Fig. 104-2), along which substantial amounts of oxygen loading *and unloading* can occur over a narrow range of oxygen tensions, is physiologically more useful than the high-affinity hyperbolic curve of individual monomers.

Figure 104-2

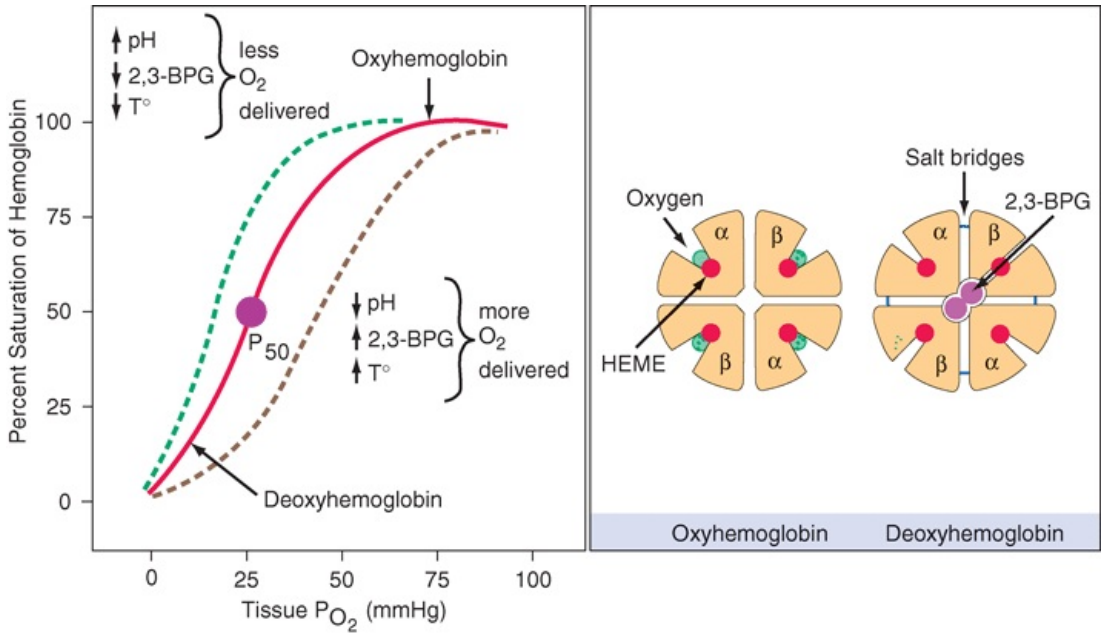

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hemoglobin-oxygen dissociation curve. The hemoglobin tetramer can bind up to four molecules of oxygen in the iron-containing sites of the heme molecules. As oxygen is bound, 2,3-BPG and CO_2 are expelled. Salt bridges are broken, and each of the globin molecules changes its conformation to facilitate oxygen binding. Oxygen release to the tissues is the reverse process, salt bridges being formed and 2,3-BPG and CO_2 bound. Deoxyhemoglobin does not bind oxygen efficiently until the cell returns to conditions of higher pH, the most important modulator of O_2 affinity (Bohr effect). When acid is produced in the tissues, the dissociation curve shifts to the right, facilitating oxygen release and CO_2 binding. Alkalosis has the opposite effect, reducing oxygen delivery.

Oxygen affinity is modulated by several factors. The Bohr effect is the ability of hemoglobin to deliver more oxygen to tissues at low pH. It arises from the stabilizing action of protons on deoxyhemoglobin, which binds protons more readily than oxyhemoglobin because the latter is a weaker acid (Fig. 104-2). Thus, hemoglobin has a lower oxygen affinity at low pH. The major small molecule that alters oxygen affinity in humans is 2,3-bisphosphoglycerate (2,3-BPG, formerly 2,3-DPG), which lowers oxygen affinity when bound to hemoglobin. HbA has a reasonably high affinity for 2,3-BPG. HbF does not bind 2,3-BPG, so it tends to have a higher oxygen affinity in vivo. Hemoglobin also binds nitric oxide reversibly; this interaction influences vascular tone, but its clinical relevance remains controversial.

Proper oxygen transport depends on the tetrameric structure of the proteins, the proper arrangement of the charged amino acids, and interaction with protons or 2,3-BPG.

Developmental Biology of Human Hemoglobins

Red cells first appearing at about 6 weeks after conception contain the embryonic hemoglobins Hb Portland ($\zeta_2\gamma_2$), Hb Gower I ($\zeta_2\varepsilon_2$), and Hb Gower II ($\alpha_2\varepsilon_2$). At 10–11 weeks, fetal hemoglobin (HbF; α_2

γ_2) becomes predominant. The switch to nearly exclusive synthesis of adult hemoglobin (HbA; $\alpha_2\beta_2$) occurs at about 38 weeks (Fig. 104-1). Fetuses and newborns therefore require α-globin but not β-globin for normal gestation. A major advance in understanding the HbF to HbA transition has been the demonstration that the transcription factor Bcl11a plays a pivotal role in its regulation. Small amounts of HbF are produced during postnatal life. A few red cell clones called *F cells* are progeny of a small pool of immature committed erythroid precursors (BFU-e) that retain the ability to produce HbF. Profound erythroid stresses, such as severe hemolytic anemias, bone marrow transplantation, or cancer chemotherapy, cause more of the F-potent BFU-e to be recruited. HbF levels thus tend to rise in some patients with sickle cell anemia or thalassemia. This phenomenon probably explains the ability of hydroxyurea to increase levels of HbF in adults. Agents such as butyrate and histone deacetylase inhibitors can also activate fetal globin genes partially after birth.

Genetics and Biosynthesis of Human Hemoglobin

The human hemoglobins are encoded in two tightly linked gene clusters; the α-like globin genes are clustered on chromosome 16 and the β-like genes on chromosome 11 (Fig. 104-1). The α-like cluster consists of two α-globin genes and a single copy of the ζ gene. The non-α gene cluster consists of a single ε gene, the Gγ and Aγ fetal globin genes, and the adult δ and β genes.

Important regulatory sequences flank each gene. Immediately upstream are typical promoter elements needed for the assembly of the transcription initiation complex. Sequences in the 5' flanking region of the γ and the β genes appear to be crucial for the correct developmental regulation of these genes, while elements that function like classic enhancers and silencers are in the 3' flanking regions. The locus control region (LCR) elements located far upstream appear to control the overall level of expression of each cluster. These elements achieve their regulatory effects by interacting with *trans*-acting transcription factors. Some of these factors are ubiquitous (e.g., Sp1 and YY1), while others are more or less limited to erythroid cells or hematopoietic cells (e.g., GATA-1, NFE-2, and EKLF). The LCR controlling the α-globin gene cluster is modulated by a SWI/SNF-like protein called *ATRX*; this protein appears to influence chromatin remodeling and DNA methylation. The association of α thalassemia with mental retardation and myelodysplasia in some families appears to be related to mutations in the ATRX pathway. This pathway also modulates genes specifically expressed during erythropoiesis, such as those that encode the enzymes for heme biosynthesis. Normal red blood cell (RBC) differentiation requires the coordinated expression of the globin genes with the genes responsible for heme and iron metabolism. RBC precursors contain a protein, α-hemoglobin-stabilizing protein (AHSP), that enhances the folding and solubility of α globin, which is otherwise easily denatured, leading to insoluble precipitates. These precipitates play an important role in the thalassemia syndromes and certain unstable hemoglobin disorders. Polymorphic variation in the amounts and/or functional capacity of AHSP might explain some of the clinical variability seen in patients inheriting identical thalassemia mutations.

CLASSIFICATION OF HEMOGLOBINOPATHIES

There are five major classes of hemoglobinopathies **(Table 104-1)**. *Structural hemoglobinopathies* occur when mutations alter the amino acid sequence of a globin chain, altering the physiologic properties of the variant hemoglobins and producing the characteristic clinical abnormalities. The most clinically relevant variant hemoglobins polymerize abnormally, as in sickle cell anemia, or exhibit altered solubility or oxygen-binding affinity. *Thalassemia syndromes* arise from mutations that impair production or translation of globin mRNA, leading to deficient globin chain biosynthesis. Clinical abnormalities are attributable to the inadequate supply of hemoglobin and the imbalances in the production of individual globin chains, leading to premature destruction of erythroblasts and RBC. *Thalassemic hemoglobin variants* combine features of thalassemia (e.g., abnormal globin biosynthesis) and of structural hemoglobinopathies (e.g., an abnormal amino acid sequence). *Hereditary persistence of fetal hemoglobin* (HPFH) is characterized by synthesis of high levels of fetal hemoglobin in adult life. *Acquired hemoglobinopathies* include modifications of the hemoglobin molecule by toxins (e.g., acquired methemoglobinemia) and clonal abnormalities of hemoglobin synthesis (e.g., high levels of HbF production in preleukemia and α thalassemia in myeloproliferative disorders).

Table 104-1 Classification of Hemoglobinopathies

I. Structural hemoglobinopathies—hemoglobins with altered amino acid sequences that result in deranged function or altered physical or chemical properties

- A. Abnormal hemoglobin polymerization—HbS, hemoglobin sickling
- B. Altered O2 affinity
  - 1. High affinity—polycythemia
  - 2. Low affinity—cyanosis, pseudoanemia
- C. Hemoglobins that oxidize readily

1. Unstable hemoglobins—hemolytic anemia, jaundice
2. M hemoglobins—methemoglobinemia, cyanosis

II. Thalassemias—defective biosynthesis of globin chains
- A. α Thalassemias
- B. β Thalassemias
- C. δβ, γδβ, αβ Thalassemias

III. Thalassemic hemoglobin variants—structurally abnormal Hb associated with co-inherited thalassemic phenotype
- A. HbE
- B. Hb Constant Spring
- C. Hb Lepore

IV. Hereditary persistence of fetal hemoglobin—persistence of high levels of HbF into adult life

V. Acquired hemoglobinopathies
- A. Methemoglobin due to toxic exposures
- B. Sulfhemoglobin due to toxic exposures
- C. Carboxyhemoglobin
- D. HbH in erythroleukemia
- E. Elevated HbF in states of erythroid stress and bone marrow dysplasia

Epidemiology

Hemoglobinopathies are especially common in areas in which malaria is endemic. This clustering of hemoglobinopathies is assumed to reflect a selective survival advantage for the abnormal RBC, which presumably provide a less hospitable environment during the obligate RBC stages of the parasitic life cycle. Very young children with α thalassemia are *more* susceptible to infection with the nonlethal *Plasmodium vivax*. Thalassemia might then favor a natural protection against infection with the more lethal *P. falciparum*.

Thalassemias are the most common genetic disorders in the world, affecting nearly 200 million people worldwide. About 15% of American blacks are silent carriers for α thalassemia; α-thalassemia trait (minor) occurs in 3% of American blacks and in 1–15% of persons of Mediterranean origin. β-Thalassemia has a 10–15% incidence in individuals from the Mediterranean and Southeast Asia and 0.8% in American blacks. The number of severe cases of thalassemia in the United States is about 1000. Sickle cell disease is the most common structural hemoglobinopathy, occurring in heterozygous form in ~8% of American blacks and in homozygous form in 1 in 400. Between 2 and 3% of American blacks carry a hemoglobin C allele.

Inheritance and Ontogeny

Hemoglobinopathies are autosomal codominant traits—compound heterozygotes who inherit a different abnormal mutant allele from each parent exhibit composite features of each. For example, patients inheriting sickle β thalassemia exhibit features of β thalassemia and sickle cell anemia. The α chain is present in HbA, HbA_2, and HbF; α-chain mutations thus cause abnormalities in all three. The α-globin hemoglobinopathies are symptomatic in utero and after birth because normal function of the α-globin gene is required throughout gestation and adult life. In contrast, infants with β-globin hemoglobinopathies tend to be asymptomatic until 3–9 months of age, when HbA has largely replaced HbF. Prevention or partial reversion of the switch should thus be an effective therapeutic strategy for β-chain hemoglobinopathies.

DETECTION AND CHARACTERIZATION OF HEMOGLOBINOPATHIES—GENERAL METHODS

Electrophoretic techniques are widely used for hemoglobin analysis. Electrophoresis at pH 8.6 on cellulose acetate membranes is especially simple, inexpensive, and reliable for initial screening. Agar gel electrophoresis at pH 6.1 in citrate buffer is often used as a complementary method because each method detects different variants. Some important variants are electrophoretically silent. These mutant hemoglobins can usually be characterized by more specialized techniques such as isoelectric focusing and/or high-pressure liquid chromatography (HPLC), which is rapidly replacing electrophoresis for initial analysis.

Quantitation of the hemoglobin profile is often desirable. HbA_2 is frequently elevated in β-thalassemia trait and depressed in iron deficiency. HbF is elevated in HPFH, some β-thalassemia syndromes, and occasional periods of erythroid stress or marrow dysplasia. For characterization of sickle cell trait,

sickle thalassemia syndromes, or HbSC disease, and for monitoring the progress of exchange transfusion therapy to lower the percentage of circulating HbS, quantitation of individual hemoglobins is also required. In most laboratories, quantitation is performed only if the test is specifically ordered. Complete characterization, including amino acid sequencing or gene cloning and sequencing, is readily available from several reference laboratories.

Because some variants can comigrate with HbA or HbS (sickle hemoglobin), electrophoretic assessment should always be regarded as incomplete unless functional assays for hemoglobin sickling, solubility, or oxygen affinity are also performed, as dictated by the clinical presentation. The best sickling assays involve measurement of the degree to which the hemoglobin sample becomes insoluble, or gelated, as it is deoxygenated (i.e., sickle solubility test). Unstable hemoglobins are detected by their precipitation in isopropanol or after heating to 50°C. High-O_2 affinity and low-O_2 affinity variants are detected by quantitating the P_{50}, the partial pressure of oxygen at which the hemoglobin sample becomes 50% saturated with oxygen. Direct tests for the percentage carboxyhemoglobin and methemoglobin, employing spectrophotometric techniques, can readily be obtained from most clinical laboratories on an urgent basis.

Laboratory evaluation remains an adjunct, rather than the primary diagnostic aid. Diagnosis is best established by recognition of a characteristic history, physical findings, peripheral blood smear morphology, and abnormalities of the complete blood cell count (e.g., profound microcytosis with minimal anemia in thalassemia trait).

STRUCTURALLY ABNORMAL HEMOGLOBINS

Sickle Cell Syndromes

The sickle cell syndromes are caused by a mutation in the β-globin gene that changes the sixth amino acid from glutamic acid to valine. HbS ($\alpha_2\beta_2^{6\ Glu\rightarrow Val}$) polymerizes reversibly when deoxygenated to form a gelatinous network of fibrous polymers that stiffen the RBC membrane, increase viscosity, and cause dehydration due to potassium leakage and calcium influx **(Fig. 104-3)**. These changes also produce the sickle shape. Sickled cells lose the pliability needed to traverse small capillaries. They possess altered "sticky" membranes that are abnormally adherent to the endothelium of small venules. These abnormalities provoke unpredictable episodes of microvascular vasoocclusion and premature RBC destruction (hemolytic anemia). Hemolysis occurs because the spleen destroys the abnormal RBC. The rigid adherent cells clog small capillaries and venules, causing tissue ischemia, acute pain, and gradual end-organ damage. This venoocclusive component usually dominates the clinical course. Prominent manifestations include episodes of ischemic pain (i.e., painful crises) and ischemic malfunction or frank infarction in the spleen, central nervous system, bones, liver, kidneys, and lungs (Fig. 104-3).

Figure 104-3

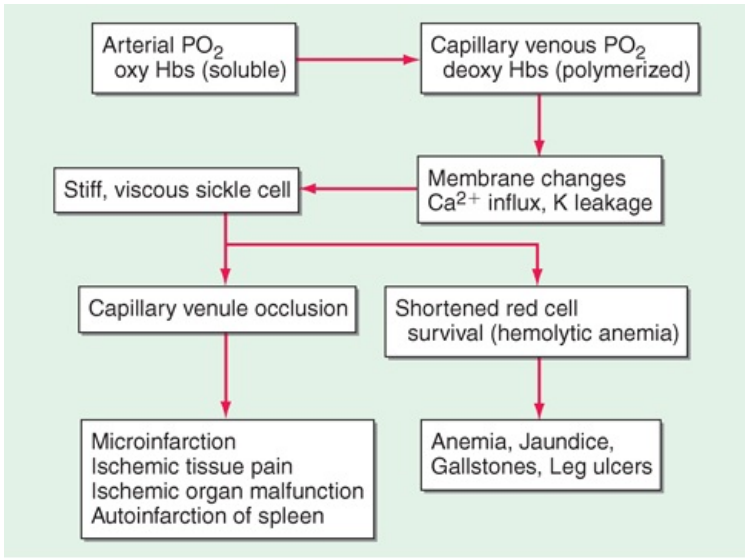

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathophysiology of sickle cell crisis.

Several sickle syndromes occur as the result of inheritance of HbS from one parent and another hemoglobinopathy, such as β thalassemia or HbC ($\alpha_2\beta_2^{6\ Glu\rightarrow Lys}$), from the other parent. The

prototype disease, sickle cell anemia, is the homozygous state for HbS **(Table 104-2)**.

Table 104-2 Clinical Features of Sickle Hemoglobinopathies

Condition	Clinical Abnormalities	Hemoglobin Level g/L (g/dL)	MCV, fL	Hemoglobin Electrophoresis
Sickle cell trait	None; rare painless hematuria	Normal	Normal	Hb S/A:40/60
Sickle cell anemia	Vasoocclusive crises with infarction of spleen, brain, marrow, kidney, lung; aseptic necrosis of bone; gallstones; priapism; ankle ulcers	70–100 (7–10)	80–100	Hb S/A:100/0 Hb F:2–25%
S/β° thalassemia	Vasoocclusive crises; aseptic necrosis of bone	70–100 (7–10)	60–80	Hb S/A:100/0 Hb F:1–10%
S/β+ thalassemia	Rare crises and aseptic necrosis	100–140 (10–14)	70–80	Hb S/A:60/40
Hemoglobin SC	Rare crises and aseptic necrosis; painless hematuria	100–140 (10–14)	80–100	Hb S/A:50/0 Hb C:50%

CLINICAL MANIFESTATIONS OF SICKLE CELL ANEMIA

Most patients with sickling syndromes suffer from hemolytic anemia, with hematocrits from 15 to 30%, and significant reticulocytosis. Anemia was once thought to exert protective effects against vasoocclusion by reducing blood viscosity. However, natural history and drug therapy trials suggest that an *increase* in the hematocrit and feedback inhibition of reticulocytosis might be beneficial, even at the expense of increased blood viscosity. The role of adhesive reticulocytes in vasoocclusion might account for these paradoxical effects.

Granulocytosis is common. The white count can fluctuate substantially and unpredictably during and between painful crises, infectious episodes, and other intercurrent illnesses.

Vasoocclusion causes protean manifestations. Intermittent episodes of vasoocclusion in connective and musculoskeletal structures produce ischemia manifested by acute pain and tenderness, fever, tachycardia, and anxiety. These recurrent episodes, called *painful crises*, are the most common clinical manifestation. Their frequency and severity vary greatly. Pain can develop almost anywhere in the body and may last from a few hours to 2 weeks. Repeated crises requiring hospitalization (>3 per year) correlate with reduced survival in adult life, suggesting that these episodes are associated with accumulation of chronic end organ damage. Provocative factors include infection, fever, excessive exercise, anxiety, abrupt changes in temperature, hypoxia, or hypertonic dyes.

Repeated micro-infarction can destroy tissues having microvascular beds prone to sickling. Thus, the spleen is frequently lost within the first 18–36 months of life, causing susceptibility to infection, particularly by pneumococci. Acute venous obstruction of the spleen (*splenic sequestration crisis*), a rare occurrence in early childhood, may require emergency transfusion and/or splenectomy to prevent trapping of the entire arterial output in the obstructed spleen. Occlusion of retinal vessels can produce hemorrhage, neovascularization, and eventual detachments. Renal papillary necrosis invariably produces isosthenuria. More widespread renal necrosis leads to renal failure in adults, a common late cause of death. Bone and joint ischemia can lead to aseptic necrosis, especially of the femoral or humeral heads; chronic arthropathy; and unusual susceptibility to osteomyelitis, which may be caused by organisms, such as *Salmonella*, rarely encountered in other settings. The *hand-foot syndrome* is caused by painful infarcts of the digits and dactylitis. Stroke is especially common in children; a small subset tend to suffer repeated episodes. Stroke is less common in adults and is often hemorrhagic. A particularly painful complication in males is priapism, due to infarction of the penile venous outflow tracts; permanent impotence is a frequent consequence. Chronic lower leg ulcers probably arise from ischemia and superinfection in the distal circulation.

Acute chest syndrome is a distinctive manifestation characterized by chest pain, tachypnea, fever, cough, and arterial oxygen desaturation. It can mimic pneumonia, pulmonary emboli, bone marrow infarction and embolism, myocardial ischemia, or in situ lung infarction. Acute chest syndrome is thought to reflect in situ sickling within the lung, producing pain and temporary pulmonary dysfunction. Often it is difficult or impossible to distinguish among other possibilities. Pulmonary infarction and pneumonia are the most frequent underlying or concomitant conditions in patients with this syndrome. Repeated episodes of acute chest pain correlate with reduced survival. Acutely, reduction in arterial oxygen saturation is especially ominous because it promotes sickling on a massive scale. Chronic acute or subacute pulmonary crises lead to pulmonary hypertension and cor pulmonale, an increasingly common cause of death as patients survive longer. Considerable controversy exists about the possible role played by free plasma HbS in scavenging NO_2, thus raising pulmonary vascular tone.

Trials of sildenafil to restore NO_2 levels were terminated because of adverse effects.

Sickle cell syndromes are remarkable for their clinical heterogeneity. Some patients remain virtually asymptomatic into or even through adult life, while others suffer repeated crises requiring hospitalization from early childhood. Patients with sickle thalassemia and sickle-HbE tend to have similar, slightly milder, symptoms, perhaps because of the ameliorating effects of production of other hemoglobins within the RBC. Hemoglobin SC disease, one of the more common variants of sickle cell anemia, is frequently marked by lesser degrees of hemolytic anemia and a greater propensity for the development of retinopathy and aseptic necrosis of bones. In most respects, however, the clinical manifestations resemble sickle cell anemia. Some rare hemoglobin variants actually aggravate the sickling phenomenon.

The clinical variability in different patients inheriting the same disease-causing mutation (sickle hemoglobin) has made sickle cell disease the focus of efforts to identify modifying genetic polymorphisms in other genes that might account for the heterogeneity. The complexity of the data obtained thus far has dampened the expectation that genomewide analysis will yield individualized profiles that predict a patient's clinical course. Nevertheless, a number of interesting patterns have emerged from these modifying gene analyses. For example, genes affecting the inflammatory response or cytokine expression appear to be modifying candidates. Genes that affect transcriptional regulation of lymphocytes may be involved. Thus, it appears likely that key polymorphic changes in the patient's inflammatory response to the damages provoked by sickle red cells or in the response to chronic or recurrent infections may prove to be important for prognosticating the clinical severity of disease.

CLINICAL MANIFESTATIONS OF SICKLE CELL TRAIT

Sickle cell trait is usually asymptomatic. Anemia and painful crises are rare. An uncommon but highly distinctive symptom is painless hematuria often occurring in adolescent males, probably due to papillary necrosis. Isosthenuria is a more common manifestation of the same process. Sloughing of papillae with urethral obstruction has been reported, as have isolated cases of massive sickling or sudden death due to exposure to high altitudes or extremes of exercise and dehydration. Avoidance of dehydration or extreme physical stress should be advised.

DIAGNOSIS

Sickle cell syndromes are suspected on the basis of hemolytic anemia, RBC morphology **(Fig. 104-4)**, and intermittent episodes of ischemic pain. Diagnosis is confirmed by hemoglobin electrophoresis and the sickling tests already discussed. Thorough characterization of the exact hemoglobin profile of the patient is important, because sickle thalassemia and hemoglobin SC disease have distinct prognoses or clinical features. Diagnosis is usually established in childhood, but occasional patients, often with compound heterozygous states, do not develop symptoms until the onset of puberty, pregnancy, or early adult life. Genotyping of family members and potential parental partners is critical for genetic counseling. Details of the childhood history establish prognosis and need for aggressive or experimental therapies. Factors associated with increased morbidity and reduced survival are more than three crises requiring hospitalization per year, chronic neutrophilia, a history of splenic sequestration or hand-foot syndrome, and second episodes of acute chest syndrome. Patients with a history of cerebrovascular accidents are at higher risk for repeated episodes and require partial exchange transfusion and especially close monitoring using Doppler carotid flow measurements. Patients with severe or repeated episodes of acute chest syndrome may need lifelong transfusion support, utilizing partial exchange transfusion, if possible.

Figure 104-4

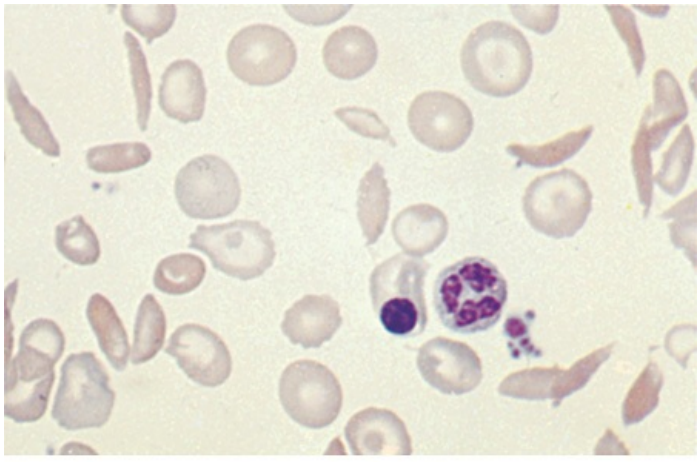

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sickle cell anemia. The elongated and crescent-shaped red blood cells seen on this smear represent circulating irreversibly sickled cells. Target cells and a nucleated red blood cell are also seen.

TREATMENT: SICKLE CELL SYNDROMES

Patients with sickle cell syndromes require ongoing continuity of care. Familiarity with the pattern of symptoms provides the best safeguard against excessive use of the emergency room, hospitalization, and habituation to addictive narcotics. Additional preventive measures include regular slit-lamp examinations to monitor development of retinopathy; antibiotic prophylaxis appropriate for splenectomized patients during dental or other invasive procedures; and vigorous oral hydration during or in anticipation of periods of extreme exercise, exposure to heat or cold, emotional stress, or infection. Pneumococcal and *Haemophilus influenzae* vaccines are less effective in splenectomized individuals. Thus, patients with sickle cell anemia should be vaccinated early in life.

The management of acute painful crisis includes vigorous hydration, thorough evaluation for underlying causes (such as infection), and aggressive analgesia administered by a standing order and/or patient-controlled analgesia (PCA) pump. Morphine (0.1–0.15 mg/kg every 3–4 h) should be used to control severe pain. Bone pain may respond as well to ketorolac (30–60 mg initial dose, then 15–30 mg every 6–8 h). Inhalation of nitrous oxide can provide short-term pain relief, but great care must be exercised to avoid hypoxia and respiratory depression. Nitrous oxide also elevates O_2 affinity, reducing O_2 delivery to tissues. Its use should be restricted to experts. Many crises can be managed at home with oral hydration and oral analgesia. Use of the emergency room should be reserved for especially severe symptoms or circumstances in which other processes, e.g., infection, are strongly suspected. Nasal oxygen should be employed as appropriate to protect arterial saturation. Most crises resolve in 1–7 days. Use of blood transfusion should be reserved for extreme cases: transfusions do not shorten the duration of the crisis.

No tests are definitive to diagnose acute painful crisis. Critical to good management is an approach that recognizes that most patients reporting crisis symptoms do indeed have crisis or another significant medical problem. Diligent diagnostic evaluation for underlying causes is imperative, even though these are found infrequently. In adults, the possibility of aseptic necrosis or sickle arthropathy must be considered, especially if pain and immobility become repeated or chronic at a single site. Nonsteroidal anti-inflammatory agents are often effective for sickle cell arthropathy.

Acute chest syndrome is a medical emergency that may require management in an intensive care unit. Hydration should be monitored carefully to avoid the development of pulmonary edema, and oxygen therapy should be especially vigorous for protection of arterial saturation. Diagnostic evaluation for pneumonia and pulmonary embolism should be especially thorough, since these may occur with atypical symptoms. Critical interventions are transfusion to maintain a hematocrit >30, and emergency exchange transfusion if arterial saturation drops to <90%. As patients with sickle cell syndrome increasingly survive into their fifth and sixth decades, end-stage renal failure and pulmonary hypertension are becoming increasingly prominent causes of end-stage morbidity. A sickle cell cardiomyopathy and/or premature coronary artery disease may compromise cardiac function in later years. Sickle cell patients have received kidney transplants, but they often experience an increase in the frequency and severity of crises, possibly due to increased infection as a consequence of immunosuppression.

The most significant advance in the therapy of sickle cell anemia has been the introduction of hydroxyurea as a mainstay of therapy for patients with severe symptoms. Hydroxyurea (10–30 mg/kg per day) increases fetal hemoglobin and may also exert beneficial affects on RBC hydration, vascular wall adherence, and suppression of the granulocyte and reticulocyte counts; dosage is titrated to maintain a white cell count between 5000 and 8000 per μL. White cells and reticulocytes may play a major role in the pathogenesis of sickle cell crisis, and their suppression may be an important benefit of hydroxyurea therapy.

Hydroxyurea should be considered in patients experiencing repeated episodes of acute chest syndrome or with more than three crises per year requiring hospitalization. The utility of this agent for reducing the incidence of other complications (priapism, retinopathy) is under evaluation, as are the long-term side effects. Hydroxyurea offers broad benefits to most patients whose disease is severe enough to impair their functional status, and it may improve survival. HbF levels increase in most patients within a few months.

The antitumor drug 5-azacytidine was the first agent found to elevate HbF. It never achieved widespread use because of concerns about acute toxicity and carcinogenesis. However, low doses of the related agent 5-deoxyazacytidine (decitabine) can elevate HbF with more acceptable toxicity.

Bone marrow transplantation can provide definitive cures but is known to be effective and safe only in children. Partially myeloablative conditioning regimens ("mini" transplants) may allow more widespread use of this modality. Prognostic features justifying bone marrow transplant are the presence of repeated crises early in life, a high neutrophil count, or the development of hand-foot syndrome. Children at risk for stroke can now be identified through the use of Doppler ultrasound techniques.

Prophylactic exchange transfusion appears to substantially reduce the risk of stroke in this population. Children who do suffer a cerebrovascular accident should be maintained for at least 3–5 years on a program of vigorous exchange transfusion, as the risk of second strokes is extremely high.

Gene therapy for sickle cell anemia is being intensively pursued, but no safe measures are currently available. Agents blocking RBC dehydration or vascular adhesion, such as clotrimazole or magnesium, may have value as an adjunct to hydroxyurea therapy, pending the completion of ongoing trials. Combinations of clotrimazole and magnesium are being evaluated.

Unstable Hemoglobins

Amino acid substitutions that reduce solubility or increase susceptibility to oxidation produce unstable hemoglobins that precipitate, forming inclusion bodies injurious to the RBC membrane. Representative mutations are those that interfere with contact points between the α and β subunits [e.g., Hb Philly ($\beta^{35Tyr\rightarrow Phe}$)], alter the helical segments [e.g., Hb Genova ($\beta^{28Leu\rightarrow Pro}$)], or disrupt interactions of the hydrophobic pockets of the globin subunits with heme [e.g., Hb Koln ($\beta^{98Val\rightarrow Met}$)] **(Table 104-3)**. The inclusions, called *Heinz bodies*, are clinically detectable by staining with supravital dyes such as crystal violet. Removal of these inclusions by the spleen generates pitted, rigid cells that have shortened life spans, producing hemolytic anemia of variable severity, sometimes requiring chronic transfusion support. Splenectomy may be needed to correct the anemia. Leg ulcers and premature gallbladder disease due to bilirubin loading are frequent stigmata.

Table 104-3 Representative Abnormal Hemoglobins with Altered Synthesis or Function

Designation	Mutation	Population	Main Clinical Effects[a]
Sickle or S	$\beta^{6Glu\rightarrow Val}$	African	Anemia, ischemic infarcts
C	$\beta^{6Glu\rightarrow Lys}$	African	Mild anemia; interacts with HbS
E	$\beta^{26Glu\rightarrow Lys}$	Southeast Asian	Microcytic anemia, splenomegaly, thalassemic phenotype
Köln	$\beta^{98Val\rightarrow Met}$	Sporadic	Hemolytic anemia, Heinz bodies when splenectomized
Yakima	$\beta^{99Asp\rightarrow His}$	Sporadic	Polycythemia
Kansas	$\beta^{102Asn\rightarrow Lys}$	Sporadic	Mild anemia
M. Iwata	$\alpha^{87His\rightarrow Tyr}$	Sporadic	Methemoglobinemia

[a]See text for details.

Unstable hemoglobins occur sporadically, often by spontaneous new mutations. Heterozygotes are often symptomatic because a significant Heinz body burden can develop even when the unstable variant accounts for only a portion of the total hemoglobin. Symptomatic unstable hemoglobins tend to be β-globin variants, because sporadic mutations affecting only one of the four α globins would generate only 20–30% abnormal hemoglobin.

Hemoglobins with Altered Oxygen Affinity

High-affinity hemoglobins [e.g., Hb Yakima ($\beta^{99Asp\rightarrow His}$)] bind oxygen more readily but deliver less O_2 to tissues at normal capillary PO_2 levels (Fig. 104-2). Mild tissue hypoxia ensues, stimulating RBC production and erythrocytosis (Table 104-3). In extreme cases, the hematocrits can rise to 60–65%, increasing blood viscosity and producing typical symptoms (headache, somnolence, or dizziness). Phlebotomy may be required. Typical mutations alter interactions within the heme pocket or disrupt the Bohr effect or salt-bond site. Mutations that impair the interaction of HbA with 2,3-BPG can increase O_2 affinity because 2,3-BPG binding lowers O_2 affinity.

Low-affinity hemoglobins [e.g., Hb Kansas ($\beta^{102Asn\rightarrow Lys}$)] bind sufficient oxygen in the lungs, despite their lower oxygen affinity, to achieve nearly full saturation. At capillary oxygen tensions, they lose sufficient amounts of oxygen to maintain homeostasis at a low hematocrit (Fig. 104-2) (*pseudoanemia*). Capillary hemoglobin desaturation can also be sufficient to produce clinically apparent cyanosis. Despite these findings, patients usually require no specific treatment.

Methemoglobinemias

Methemoglobin is generated by oxidation of the heme iron moieties to the ferric state, causing a characteristic bluish-brown muddy color resembling cyanosis. Methemoglobin has such high oxygen affinity that virtually no oxygen is delivered. Levels >50–60% are often fatal.

Congenital methemoglobinemia arises from globin mutations that stabilize iron in the ferric state [e.g., HbM Iwata ($\alpha^{87His \rightarrow Tyr}$), Table 104-3] or from mutations that impair the enzymes that reduce methemoglobin to hemoglobin (e.g., methemoglobin reductase, NADP diaphorase). Acquired methemoglobinemia is caused by toxins that oxidize heme iron, notably nitrate and nitrite-containing compounds, including drugs commonly used in cardiology and anesthesiology.

Diagnosis and Management of Patients with Unstable Hemoglobins, High-Affinity Hemoglobins, and Methemoglobinemia

Unstable hemoglobin variants should be suspected in patients with nonimmune hemolytic anemia, jaundice, splenomegaly, or premature biliary tract disease. Severe hemolysis usually presents during infancy as neonatal jaundice or anemia. Milder cases may present in adult life with anemia or only as unexplained reticulocytosis, hepatosplenomegaly, premature biliary tract disease, or leg ulcers. Because spontaneous mutation is common, family history of anemia may be absent. The peripheral blood smear often shows anisocytosis, abundant cells with punctate inclusions, and irregular shapes (i.e., poikilocytosis).

The two best tests for diagnosing unstable hemoglobins are the Heinz body preparation and the isopropanol or heat stability test. Many unstable Hb variants are electrophoretically silent. A normal electrophoresis does not rule out the diagnosis.

Severely affected patients may require transfusion support for the first 3 years of life, because splenectomy before age 3 is associated with a significantly higher immune deficit. Splenectomy is usually effective thereafter, but occasional patients may require lifelong transfusion support. After splenectomy, patients can develop cholelithiasis and leg ulcers, hypercoagulable states, and susceptibility to overwhelming sepsis. Splenectomy should be avoided or delayed unless it is the only alternative. Precipitation of unstable hemoglobins is aggravated by oxidative stress, e.g., infection and antimalarial drugs, which should be avoided where possible.

High-O_2 affinity hemoglobin variants should be suspected in patients with erythrocytosis. The best test for confirmation is measurement of the P_{50}. A high-O_2 affinity Hb causes a significant left shift (i.e., lower numeric value of the P_{50}); confounding conditions, e.g., tobacco smoking or carbon monoxide exposure, can also lower the P_{50}.

High-affinity hemoglobins are often asymptomatic; rubor or plethora may be telltale signs. When the hematocrit reaches to 55–60%, symptoms of high blood viscosity and sluggish flow (headache, lethargy, dizziness, etc.) may be present. These persons may benefit from judicious phlebotomy. Erythrocytosis represents an appropriate attempt to compensate for the impaired oxygen delivery by the abnormal variant. Overzealous phlebotomy may stimulate increased erythropoiesis or aggravate symptoms by thwarting this compensatory mechanism. The guiding principle of phlebotomy should be to improve oxygen delivery by reducing blood viscosity and increasing blood flow rather than restoration of a normal hematocrit. Modest iron deficiency may aid in control.

Low-affinity hemoglobins should be considered in patients with cyanosis or a low hematocrit with no other reason apparent after thorough evaluation. The P_{50} test confirms the diagnosis. Counseling and reassurance are the interventions of choice.

Methemoglobin should be suspected in patients with hypoxic symptoms who appear cyanotic but have a PaO_2 sufficiently high that hemoglobin should be fully saturated with oxygen. A history of nitrite or other oxidant ingestions may not always be available; some exposures may be inapparent to the patient, and others may result from suicide attempts. The characteristic muddy appearance of freshly drawn blood can be a critical clue. The best diagnostic test is methemoglobin assay, which is usually available on an emergency basis.

Methemoglobinemia often causes symptoms of cerebral ischemia at levels >15%; levels >60% are usually lethal. Intravenous injection of 1 mg/kg of methylene blue is effective emergency therapy. Milder cases and follow-up of severe cases can be treated orally with methylene blue (60 mg three to four times each day) or ascorbic acid (300–600 mg/d).

THALASSEMIA SYNDROMES

The thalassemia syndromes are inherited disorders of α- or β-globin biosynthesis. The reduced supply of globin diminishes production of hemoglobin tetramers, causing hypochromia and microcytosis. Unbalanced accumulation of α and β subunits occurs because the synthesis of the unaffected globins proceeds at a normal rate. Unbalanced chain accumulation dominates the clinical phenotype. Clinical severity varies widely, depending on the degree to which the synthesis of the affected globin is impaired, altered synthesis of other globin chains, and coinheritance of other abnormal globin alleles.

Clinical Manifestations of β-Thalassemia Syndromes

Mutations causing thalassemia can affect any step in the pathway of globin gene expression: transcription, processing of the mRNA precursor, translation, and posttranslational metabolism of the β-globin polypeptide chain. The most common forms arise from mutations that derange splicing of the mRNA precursor or prematurely terminate translation of the mRNA.

Hypochromia and microcytosis characterize all forms of β thalassemia because of the reduced amounts of hemoglobin tetramers **(Fig. 104-5)**. In heterozygotes (β-thalassemia trait), this is the only abnormality seen. Anemia is minimal. In more severe homozygous states, unbalanced α- and β-globin accumulation causes accumulation of highly insoluble unpaired α chains. They form toxic inclusion bodies that kill developing erythroblasts in the marrow. Few of the proerythroblasts beginning erythroid maturation survive. The surviving RBCs bear a burden of inclusion bodies that are detected in the spleen, shortening the RBC life span and producing severe hemolytic anemia. The resulting profound anemia stimulates erythropoietin release and compensatory erythroid hyperplasia, but the marrow response is sabotaged by the ineffective erythropoiesis. Anemia persists. Erythroid hyperplasia can become exuberant and produce masses of extramedullary erythropoietic tissue in the liver and spleen.

Figure 104-5

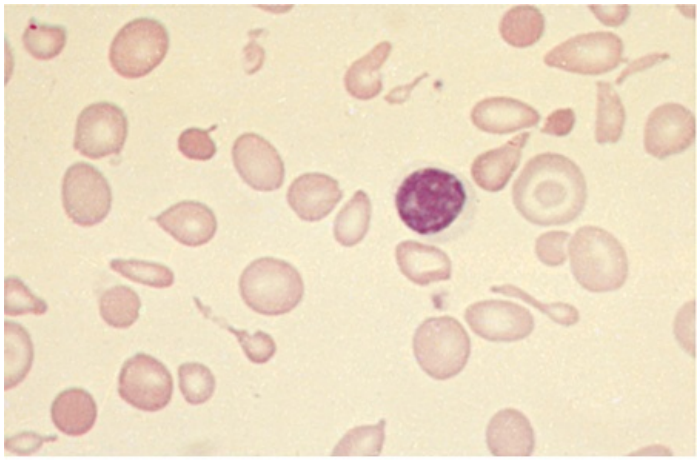

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

β-Thalassemia intermedia. Microcytic and hypochromic red blood cells are seen that resemble the red blood cells of severe iron deficiency anemia. Many elliptical and teardrop-shaped red blood cells are noted.

Massive bone marrow expansion deranges growth and development. Children develop characteristic "chipmunk" facies due to maxillary marrow hyperplasia and frontal bossing. Thinning and pathologic fracture of long bones and vertebrae may occur due to cortical invasion by erythroid elements and profound growth retardation. Hemolytic anemia causes hepatosplenomegaly, leg ulcers, gallstones, and high-output congestive heart failure. The conscription of caloric resources to support erythropoiesis leads to inanition, susceptibility to infection, endocrine dysfunction, and in the most severe cases, death during the first decade of life. Chronic transfusions with RBCs improve oxygen delivery, suppress the excessive ineffective erythropoiesis, and prolong life, but the inevitable side effects, notably iron overload, usually prove fatal by age 30 years.

Severity is highly variable. Known modulating factors are those that ameliorate the burden of unpaired α-globin inclusions. Alleles associated with milder synthetic defects and coinheritance of α-thalassemia trait reduce clinical severity by reducing accumulation of excess α globin. HbF persists to various degrees in β thalassemias. γ-Globin gene chains can substitute for β chains, generating more hemoglobin and reducing the burden of α-globin inclusions. The terms β-*thalassemia major* and β-*thalassemia intermedia* are used to reflect the clinical heterogeneity. Patients with β-thalassemia major require intensive transfusion support to survive. Patients with β-thalassemia intermedia have a somewhat milder phenotype and can survive without transfusion. The terms β-*thalassemia minor* and β-*thalassemia trait* describe asymptomatic heterozygotes for β thalassemia.

Thalassemia Syndromes

The four classic α thalassemias, most common in Asians, are α-thalassemia-2 trait, in which one of the four α-globin loci is deleted; α-thalassemia-1 trait, with two deleted loci; HbH disease, with three loci deleted; and hydrops fetalis with Hb Barts, with all four loci deleted **(Table 104-4)**. Nondeletion forms of α thalassemia also exist.

Table 104-4 The α Thalassemias

Condition	Hemoglobin A, %	Hemoglobin H (β4), %	Hemoglobin Level, g/L (g/dL)	MCV, fL
Normal	97	0	150 (15)	90
Silent thalassemia: –α/αα	98–100	0	150 (15)	90
Thalassemia trait: –α/–α homozygous α-thal-2[a] or —/αα heterozygous α-thal-1[a]	85–95	Rare red blood cell inclusions	120–130 (12–13)	70–80
Hemoglobin H disease: —/–α heterozygous α-thal-1/α-thal-2	70–95	5–30	60–100 (6–10)	60–70
Hydrops fetalis: —/— homozygous α-thal-1	0	5–10b	Fatal in utero or at birth	

[a]When both α alleles on one chromosome are deleted, the locus is called α-thal-1; when only a single α allele on one chromosome is deleted, the locus is called α-thal-2.

[b]90–95% of the hemoglobin is hemoglobin Barts (tetramers of γ chains).

α-*Thalassemia-2 trait* is an asymptomatic, silent carrier state. α-*Thalassemia-1 trait* resembles β-thalassemia minor. Offspring doubly heterozygous for α-thalassemia-2 and α-thalassemia-1 exhibit a more severe phenotype called *HbH disease*. Heterozygosity for a deletion that removes both genes from the same chromosome (*cis* deletion) is common in Asians and in those from the Mediterranean region, as is homozygosity for α-thalassemia-2 (*trans* deletion). Both produce asymptomatic hypochromia and microcytosis.

In *HbH disease*, HbA production is only 25–30% normal. Fetuses accumulate some unpaired γ chains (Hb Barts; γ-chain tetramers). In adults, unpaired β chains accumulate and are soluble enough to form β_4 tetramers called HbH. HbH forms few inclusions in erythroblasts and precipitates in circulating RBC. Patients with HbH disease have thalassemia intermedia characterized by moderately severe hemolytic anemia but milder ineffective erythropoiesis. Survival into midadult life without transfusions is common.

The homozygous state for the α-thalassemia-1 *cis* deletion (hydrops fetalis) causes total absence of α-globin synthesis. No physiologically useful hemoglobin is produced beyond the embryonic stage. Excess γ globin forms tetramers called *Hb Barts* (γ_4), which has a very high oxygen affinity. It delivers almost no O_2 to fetal tissues, causing tissue asphyxia, edema (hydrops fetalis), congestive heart failure, and death in utero. α-Thalassemia-2 trait is common (15–20%) among people of African descent. The *cis* α-thalassemia-1 deletion is almost never seen, however. Thus, α-thalassemia-2 and the *trans* form of α-thalassemia-1 are very common, but HbH disease and hydrops fetalis are rare.

It has been known for some time that some patients with myelodysplasia or erythroleukemia produce RBC clones containing HbH. This phenomenon is due to mutations in the ATRX pathway that affect the LCR of the α-globin gene cluster.

Diagnosis and Management of Thalassemias

The diagnosis of β-thalassemia major is readily made during childhood on the basis of severe anemia accompanied by the characteristic signs of massive ineffective erythropoiesis: hepatosplenomegaly, profound microcytosis, a characteristic blood smear (Fig. 104-5), and elevated levels of HbF, HbA_2, or both. Many patients require chronic hypertransfusion therapy designed to maintain a hematocrit of at least 27–30% so that erythropoiesis is suppressed. Splenectomy is required if the annual transfusion requirement (volume of RBCs per kilogram of body weight per year) increases by >50%. Folic acid supplements may be useful. Vaccination with Pneumovax in anticipation of eventual splenectomy is advised, as is close monitoring for infection, leg ulcers, and biliary tract disease. Many patients develop endocrine deficiencies as a result of iron overload. Early endocrine evaluation is required for glucose intolerance, thyroid dysfunction, and delayed onset of puberty or secondary sexual characteristics.

Patients with β-thalassemia intermedia exhibit similar stigmata but can survive without chronic hypertransfusion. Management is particularly challenging because a number of factors can aggravate the anemia, including infection, onset of puberty, and development of splenomegaly and hypersplenism. Some patients may eventually benefit from splenectomy. The expanded erythron can

cause absorption of excessive dietary iron and hemosiderosis, even without transfusion.

β-Thalassemia minor (i.e., thalassemia trait) usually presents as profound microcytosis and hypochromia with target cells, but only minimal or mild anemia. The mean corpuscular volume is rarely >75 fL; the hematocrit is rarely <30–33%. Hemoglobin analysis classically reveals an elevated HbA_2 (3.5–7.5%), but some forms are associated with normal HbA_2 and/or elevated HbF. Genetic counseling and patient education are essential. Patients with β-thalassemia trait should be warned that their blood picture resembles iron deficiency and can be misdiagnosed. They should eschew empirical use of iron, yet iron deficiency can develop during pregnancy or from chronic bleeding.

Persons with α-thalassemia trait may exhibit mild hypochromia and microcytosis usually without anemia. HbA_2 and HbF levels are normal. Affected individuals usually require only genetic counseling. HbH disease resembles β-thalassemia intermedia, with the added complication that the HbH molecule behaves like moderately unstable hemoglobin. Patients with HbH disease should undergo splenectomy if excessive anemia or a transfusion requirement develops. Oxidative drugs should be avoided. Iron overload leading to death can occur in more severely affected patients.

Prevention

Antenatal diagnosis of thalassemia syndromes is now widely available. DNA diagnosis is based on PCR amplification of fetal DNA, obtained by amniocentesis or chorionic villus biopsy followed by hybridization to allele-specific oligonucleotides probes or direct DNA sequencing.

THALASSEMIC STRUCTURAL VARIANTS

Thalassemic structural variants are characterized by both defective synthesis and abnormal structure.

Hemoglobin Lepore

Hb Lepore [$\alpha_2(\delta\beta)_2$] arises by an unequal crossover and recombination event that fuses the proximal end of the δ-gene with the distal end of the closely linked β-gene. It is common in the Mediterranean basin. The resulting chromosome contains only the fused δβ gene. The Lepore (δβ) globin is synthesized poorly because the fused gene is under the control of the weak δ-globin promoter. Hb Lepore alleles have a phenotype like β thalassemia, except for the added presence of 2–20% Hb Lepore. Compound heterozygotes for Hb Lepore and a classic β-thalassemia allele may also have severe thalassemia.

Hemoglobin E

HbE (i.e., $\alpha_2\beta_2^{26Glu\rightarrow Lys}$) is extremely common in Cambodia, Thailand, and Vietnam. The gene has become far more prevalent in the United States as a result of immigration of Asian persons, especially in California, where HbE is the most common variant detected. HbE is mildly unstable but not enough to affect RBC life span significantly. Heterozygotes resemble individuals with a mild β-thalassemia trait. Homozygotes have somewhat more marked abnormalities but are asymptomatic. Compound heterozygotes for HbE and a β-thalassemia gene can have β-thalassemia intermedia or β-thalassemia major, depending on the severity of the coinherited thalassemic gene.

The β^E allele contains a single base change in codon 26 that causes the amino acid substitution. However, this mutation activates a cryptic RNA splice site, generating a structurally abnormal globin mRNA that cannot be translated from about 50% of the initial pre-mRNA molecules. The remaining 40–50% are normally spliced and generate functional mRNA that is translated into β^E-globin because the mature mRNA carries the base change that alters codon 26.

Genetic counseling of the persons at risk for HbE should focus on the interaction of HbE with β thalassemia rather than HbE homozygosity, a condition associated with asymptomatic microcytosis, hypochromia, and hemoglobin levels rarely <100 g/L (<10 g/dL).

Hereditary Persistence of Fetal Hemoglobin

HPFH is characterized by continued synthesis of high levels of HbF in adult life. No deleterious effects are apparent, even when all of the hemoglobin produced is HbF. These rare patients demonstrate convincingly that prevention or reversal of the fetal to adult hemoglobin switch would provide effective therapy for sickle cell anemia and β thalassemia.

Acquired Hemoglobinopathies

The two most important acquired hemoglobinopathies are carbon monoxide poisoning and methemoglobinemia (see above). Carbon monoxide has a higher affinity for hemoglobin than does oxygen; it can replace oxygen and diminish O_2 delivery. Chronic elevation of carboxyhemoglobin levels to 10 or 15%, as occurs in smokers, can lead to secondary polycythemia. Carboxyhemoglobin is cherry red in color and masks the development of cyanosis usually associated with poor O_2 delivery to tissues.

Abnormalities of hemoglobin biosynthesis have also been described in blood dyscrasias. In some

patients with myelodysplasia, erythroleukemia, or myeloproliferative disorders, elevated HbF or a mild form of HbH disease may also be seen. The abnormalities are not severe enough to alter the course of the underlying disease.

TREATMENT: TRANSFUSIONAL HEMOSIDEROSIS

Chronic blood transfusion can lead to bloodborne infection, alloimmunization, febrile reactions, and lethal iron overload (Chap. 113). A unit of packed RBCs contains 250–300 mg iron (1 mg/mL). The iron assimilated by a single transfusion of two units of packed RBCs is thus equal to a 1- to 2-year intake of iron. Iron accumulates in chronically transfused patients because no mechanisms exist for increasing iron excretion: an expanded erythron causes especially rapid development of iron overload because accelerated erythropoiesis promotes excessive absorption of dietary iron. Vitamin C should not be supplemented because it generates free radicals in iron excess states.

Patients who receive >100 units of packed RBCs usually develop hemosiderosis. The ferritin level rises, followed by early endocrine dysfunction (glucose intolerance and delayed puberty), cirrhosis, and cardiomyopathy. Liver biopsy shows both parenchymal and reticuloendothelial iron. The superconducting quantum-interference device (SQUID) is accurate at measuring hepatic iron but not widely available. Cardiac toxicity is often insidious. Early development of pericarditis is followed by dysrhythmia and pump failure. The onset of heart failure is ominous, often presaging death within a year (Chap. 357).

The decision to start long-term transfusion support should also prompt one to institute therapy with iron-chelating agents. Desferoxamine (Desferal) is for parenteral use. Its iron-binding kinetics require chronic slow infusion via a metering pump. The constant presence of the drug improves the efficiency of chelation and protects tissues from occasional releases of the most toxic fraction of iron—low-molecular-weight iron—which may not be sequestered by protective proteins.

Desferoxamine is relatively nontoxic. Occasional cataracts, deafness, and local skin reactions, including urticaria, occur. Skin reactions can usually be managed with antihistamines. Negative iron balance can be achieved, even in the face of a high transfusion requirement, but this alone does not prevent long-term morbidity and mortality in chronically transfused patients. Irreversible end-organ deterioration develops at relatively modest levels of iron overload, even if symptoms do not appear for many years thereafter. To enjoy a significant survival advantage, chelation must begin before 5–8 years of age in β-thalassemia major.

Deferasirox is an oral iron-chelating agent. Single daily doses of 20 to 30 mg/kg deferasirox produced reductions in liver iron concentration comparable to desferoxamine in long-term transfused adult and pediatric patients. Deferasirox produces some elevations in liver enzymes and slight but persistent increases in serum creatinine, without apparent clinical consequence. Other toxicities are similar to those of desferoxamine. Its toxicity profile is acceptable, although long-term effects are still being evaluated.

EXPERIMENTAL THERAPIES

Bone Marrow Transplantation, Gene Therapy, and Manipulation of HBF

Bone marrow transplantation provides stem cells able to express normal hemoglobin; it has been used in a large number of patients with β thalassemia and a smaller number of patients with sickle cell anemia. Early in the course of disease, before end-organ damage occurs, transplantation is curative in 80–90% of patients. In highly experienced centers, the treatment-related mortality is <10%. Since survival into adult life is possible with conventional therapy, the decision to transplant is best made in consultation with specialized centers.

Gene therapy of thalassemia and sickle cell disease has proved to be an elusive goal. Uptake of gene vectors into the nondividing hematopoietic stem cells has been inefficient. Lentiviral-type vectors that can transduce nondividing cells may solve this problem.

Reestablishing high levels of fetal hemoglobin synthesis should ameliorate the symptoms of β chain hemoglobinopathies. Cytotoxic agents such as hydroxyurea and cytarabine promote high levels of HbF synthesis, probably by stimulating proliferation of the primitive HbF-producing progenitor cell population (i.e., F cell progenitors). Unfortunately, this regimen has not yet been effective in β thalassemia. Butyrates stimulate HbF production, but only transiently. Pulsed or intermittent administration has been found to sustain HbF induction in the majority of patients with sickle cell disease. It is unclear whether butyrates will have similar activity in patients with β thalassemia.

APLASTIC AND HYPOPLASTIC CRISIS IN PATIENTS WITH HEMOGLOBINOPATHIES

Patients with hemolytic anemias sometimes exhibit an alarming decline in hematocrit during and immediately after acute illnesses. Bone marrow suppression occurs in almost everyone during acute inflammatory illnesses. In patients with short RBC life spans, suppression can affect RBC counts more dramatically. These hypoplastic crises are usually transient and self-correcting before intervention is required.

Aplastic crisis refers to a profound cessation of erythroid activity in patients with chronic hemolytic anemias. It is associated with a rapidly falling hematocrit. Episodes are usually self-limited. Aplastic crises are caused by infection with a particular strain of parvovirus, B19A. Children infected with this virus usually develop permanent immunity. Aplastic crises do not often recur and are rarely seen in adults. Management requires close monitoring of the hematocrit and reticulocyte count. If anemia becomes symptomatic, transfusion support is indicated. Most crises resolve spontaneously within 1–2 weeks.

FURTHER READINGS

Ataga KI, Orringa EP: Hypercoagulability in sickle cell disease: A curious paradox. Am J Med 115:721, 2003[PMID: 14693325]

Desimone J et al: Maintenance of elevated fetal hemoglobin levels by decitabine during dose interval treatment of sickle cell anemia. Blood 99:3905, 2002[PMID: 12010787]

Neufeld EJ: Oral chelators deferasirox and deferiprone for transfusional iron overload in thalassemia major: New data, new questions. Blood 107:3436, 2006[PMID: 16627763]

Prabhakar H et al: Sickle cell disease in the United States: Looking back and forward at 100 years of progress in management and survival. Am J Hematol 85:346, 2010[PMID: 20425797]

Quek L, Thein SL: Molecular therapies in beta-thalassaemia. Br J Haematol 136:353, 2007[PMID: 17129232]

Smiers FJ et al: Hematopoietic stem cell transplantation for hemoglobinopathies: Current practice and emerging trends. Pediatr Clin North Am 57:181, 2010[PMID: 20307718]

Steinberg MH: Pathophysiologically based drug treatment of sickle cell disease. Trends Pharmacol Sci 27:204, 2006[PMID: 16530854]

Switzer JA et al: Pathophysiology and treatment of stroke in sickle-cell disease: Present and future. Lancet Neurol 5:501, 2006[PMID: 16713922]

Ware RE et al: Predictors of fetal hemoglobin response in children with sickle cell anemia receiving hydroxyurea therapy. Blood 99:10, 2002[PMID: 11756146]

Weatherall DJ: The inherited diseases of hemoglobin are an emerging global health burden. Blood 115:4331, 2010[PMID: 20233970]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MEGALOBLASTIC ANEMIAS: INTRODUCTION

The megaloblastic anemias are a group of disorders characterized by the presence of distinctive morphologic appearances of the developing red cells in the bone marrow. The marrow is usually cellular and the anemia is based on ineffective erythropoiesis. The cause is usually a deficiency of either cobalamin (vitamin B12) or folate, but megaloblastic anemia may occur because of genetic or acquired abnormalities that affect the metabolism of these vitamins or because of defects in DNA synthesis not related to cobalamin or folate **(Table 105–1)**. Cobalamin and folate absorption and metabolism are described next, followed by the biochemical basis, clinical and laboratory features, causes, and treatment of megaloblastic anemia.

Table 105–1 Causes of Megaloblastic Anemia

Cobalamin deficiency or abnormalities of cobalamin metabolism (see Tables 105–3 and 105–4)
Folate deficiency or abnormalities of folate metabolism (see Table 105–5)
Therapy with antifolate drugs (e.g., methotrexate)
Independent of either cobalamin or folate deficiency and refractory to cobalamin and folate therapy:
Some cases of acute myeloid leukemia, myelodysplasia
Therapy with drugs interfering with synthesis of DNA [e.g., cytosine arabinoside, hydroxyurea, 6-mercaptopurine, azidothymidine (AZT)]
Orotic aciduria (responds to uridine)
Thiamine-responsive

COBALAMIN

Cobalamin (vitamin B12) exists in a number of different chemical forms. All have a cobalt atom at the center of a corrin ring. In nature, the vitamin is mainly in the 2-deoxyadenosyl (ado) form, which is located in mitochondria. It is the cofactor for the enzyme methylmalonyl coenzyme A (CoA) mutase. The other major natural cobalamin is methylcobalamin, the form in human plasma and in cell cytoplasm. It is the cofactor for methionine synthase. There are also minor amounts of hydroxocobalamin to which methyl- and adocobalamin are converted rapidly by exposure to light.

DIETARY SOURCES AND REQUIREMENTS

Cobalamin is synthesized solely by microorganisms. Ruminants obtain cobalamin from the foregut, but the only source for humans is food of animal origin, e.g., meat, fish, and dairy products. Vegetables, fruits, and other foods of nonanimal origin are free from cobalamin unless they are contaminated by bacteria. A normal Western diet contains 5–30 μg of cobalamin daily. Adult daily losses (mainly in the urine and feces) are 1–3 μg (~0.1% of body stores), and as the body does not have the ability to degrade cobalamin, daily requirements are also about 1–3 μg. Body stores are of the order of 2–3 mg, sufficient for 3–4 years if supplies are completely cut off.

ABSORPTION

Two mechanisms exist for cobalamin absorption. One is passive, occurring equally through buccal, duodenal, and ileal mucosa; it is rapid but extremely inefficient, with <1% of an oral dose being absorbed by this process. The normal physiologic mechanism is active; it occurs through the ileum and is efficient for small (a few micrograms) oral doses of cobalamin, and it is mediated by gastric intrinsic factor (IF). Dietary cobalamin is released from protein complexes by enzymes in the stomach, duodenum, and jejunum; it combines rapidly with a salivary glycoprotein that belongs to the family of cobalamin-binding proteins known as haptocorrins (HCs). In the intestine, the haptocorrin is digested by pancreatic trypsin and the cobalamin is transferred to IF.

IF (gene at chromosome 11q13 coding for 9 exons) is produced in the gastric parietal cells of the fundus and body of the stomach, and its secretion parallels that of hydrochloric acid. Normally, there is a vast excess of IF. The IF-cobalamin complex passes to the ileum, where IF attaches to a specific receptor (cubilin) on the microvillus membrane of the enterocytes. Cubilin also is present in yolk sac and renal proximal tubular epithelium. Cubulin appears to traffic by means of amnionless (AMN), an endocytic receptor protein that directs sublocalization and endocytosis of cubulin with its ligand IF-cobalamin complex. The cobalamin-IF complex enters the ileal cell, where IF is destroyed. After a delay of about 6 h, the cobalamin appears in portal blood attached to transcobalamin (TC) II.

Between 0.5 and 5 μg of cobalamin enter the bile each day. This binds to IF, and a major portion of biliary cobalamin normally is reabsorbed together with cobalamin derived from sloughed intestinal cells. Because of the appreciable amount of cobalamin undergoing enterohepatic circulation, cobalamin deficiency develops more rapidly in individuals who malabsorb cobalamin than it does in vegans, in whom reabsorption of biliary cobalamin is intact.

TRANSPORT

Two main cobalamin transport proteins exist in human plasma; they both bind cobalamin—one molecule for one molecule. One HC, known as TC I, is closely related to other cobalamin-binding HCs in milk, gastric juice, bile, saliva, and other fluids. The gene *TCNL* is at chromosome 11q11-q12.3 and has 9 exons. These HCs differ from each other only in the carbohydrate moiety of the molecule. TC I is derived primarily from the specific granules in neutrophils. Normally, it is about two-thirds saturated with cobalamin, which it binds tightly. TC I does not enhance cobalamin entry into tissues. Glycoprotein receptors on liver cells are involved in the removal of TC I from plasma, and TC I may play a role in the transport of cobalamin analogues (which it binds more effectively than IF) to the liver for excretion in bile.

The other major cobalamin transport protein in plasma is TC II. The gene is on chromosome 22q11-q13.1. As for IF and HC, there are 9 exons. The three proteins are likely to have a common ancestral origin. TC II is synthesized by liver and by other tissues, including macrophages, ileum, and vascular endothelium. It normally carries only 20–60 ng of cobalamin per liter of plasma and readily gives up cobalamin to marrow, placenta, and other tissues, which it enters by receptor-mediated endocytosis involving the TC II receptor and megalin (encoded by the *LRP-2* gene). The TC II cobalamin is internalized by endocytosis via clathrin-coated pits; the complex is degraded, but the receptor probably is recycled to the cell membrane as is the case for transferrin. Export of "free" cobalamin is via the ATP-binding cassette drug transporter alias multidrug resistance protein 1.

FOLATE

DIETARY FOLATE

Folic (pteroylglutamic) acid is a yellow, crystalline, water-soluble substance. It is the parent compound of a large family of natural folate compounds, which differ from it in three respects: (1) they are partly or completely reduced to di- or tetrahydrofolate (THF) derivatives, (2) they usually contain a single carbon unit **(Table 105–2)**, and (3) 70–90% of natural folates are folate-polyglutamates.

Table 105–2 Biochemical Reactions of Folate Coenzymes

Reaction	Coenzyme Form of Folate Involved	Single Carbon Unit Transferred	Importance
Formate activation	THF	–CHO	Generation of 10-formyl-THF
Purine synthesis			
Formation of glycinamide ribonucleotide	5,10-MethyleneTHF	–CHO	Formation of purines needed for DNA, RNA synthesis, but reactions probably not rate-limiting
Formylation of aminoimidazole carboxamide ribonucleotide (AICAR)	10-Formyl (CHO)THF		
Pyrimidine synthesis			
Methylation of deoxyuridine monophosphate (dUMP) to thymidine monophosphate (dTMP)	5,10-MethyleneTHF	$-CH_3$	Rate limiting in DNA synthesis Oxidizes THF to DHF Some breakdown of folate at the C-9–N-10 bond
Amino acid interconversion			
Serine–glycine interconversion	THF	$=CH_2$	Entry of single carbon units into active pool
Homocysteine to methionine	5-Methyl(M)THF	$-CH_3$	Demethylation of 5-MTHF to THF; also requires cobalamin, flavine adenine dinucleotide, ATP, and adenosylmethionine
Forminoglutamic acid to glutamic acid in histidine catabolism	THF	–HN–CH=	

Abbreviations: DHF, dihydrofolate; THF, tetrahydrofolate.

Most foods contain some folate. The highest concentrations are found in liver, yeast, spinach, other greens, and nuts (>100 μg/100 g). The total folate content of an average Western diet is ~250 μg daily, but the amount varies widely according to the type of food eaten and the method of cooking. Folate is easily destroyed by heating, particularly in large volumes of water. Total-body folate in the adult is ~10 mg, with the liver containing the largest store. Daily adult requirements are ~100 μg, and so stores are sufficient for only 3–4 months in normal adults and severe folate deficiency may develop rapidly.

ABSORPTION

Folates are absorbed rapidly from the upper small intestine. The absorption of folate polyglutamates is less efficient than that of monoglutamates; on average, ~50% of food folate is absorbed. Polyglutamate forms are hydrolyzed to the monoglutamate derivatives either in the lumen of the intestine or within the mucosa. All dietary folates are converted to 5-methylTHF (5-MTHF) within the small-intestinal mucosa before entering portal plasma. The monoglutamates are actively transported across the enterocyte by a carrier-mediated mechanism. Pteroylglutamic acid at doses >400 μg is absorbed largely unchanged and converted to natural folates in the liver. Lower doses are converted to 5-MTHF during absorption through the intestine.

About 60–90 μg of folate enters the bile each day and is excreted into the small intestine. Loss of this folate, together with the folate of sloughed intestinal cells, accelerates the speed with which folate deficiency develops in malabsorption conditions.

TRANSPORT

Folate is transported in plasma; about one-third is loosely bound to albumin, and two-thirds is unbound. In all body fluids (plasma, cerebrospinal fluid, milk, bile) folate is largely, if not entirely, 5-MTHF in the monoglutamate form. Two types of folate-binding protein are involved in the entry of MTHF into cells. A high-affinity proton-coupled folate receptor (PCFT/HCPI) takes folate into cells by endocytosis and is internalized by clathrin-coated pits or in a vesicle (caveola), which is then acidified, releasing folate. It accounts for the bulk of folate absorption. Folate then is carried by the membrane folate transporter into the cytoplasm. The high-affinity receptor is attached to the outer surface of the cell membrane by glycosyl phosphatidylinositol linkages. It may be involved in transport of oxidized folates and folate breakdown products to the liver for excretion in bile. An independent low-affinity reduced-folate carrier also mediates uptake of physiologic folates into cells but also regulates the uptake of methotrexate.

BIOCHEMICAL FUNCTIONS

Folates (as the intracellular polyglutamate derivatives) act as coenzymes in the transfer of single-carbon units (Fig. 105-1

and Table 105–2). Two of these reactions are involved in purine synthesis and one in pyrimidine synthesis necessary for DNA and RNA replication. Folate is also a coenzyme for methionine synthesis, in which methylcobalamin is also involved and in which THF is regenerated. THF is the acceptor of single carbon units newly entering the active pool via conversion of serine to glycine. Methionine, the other product of the methionine synthase reaction, is the precursor for *S*-adenosylmethionine (SAM), the universal methyl donor involved in >100 methyltransferase reactions (Fig. 105-1).

Figure 105-1

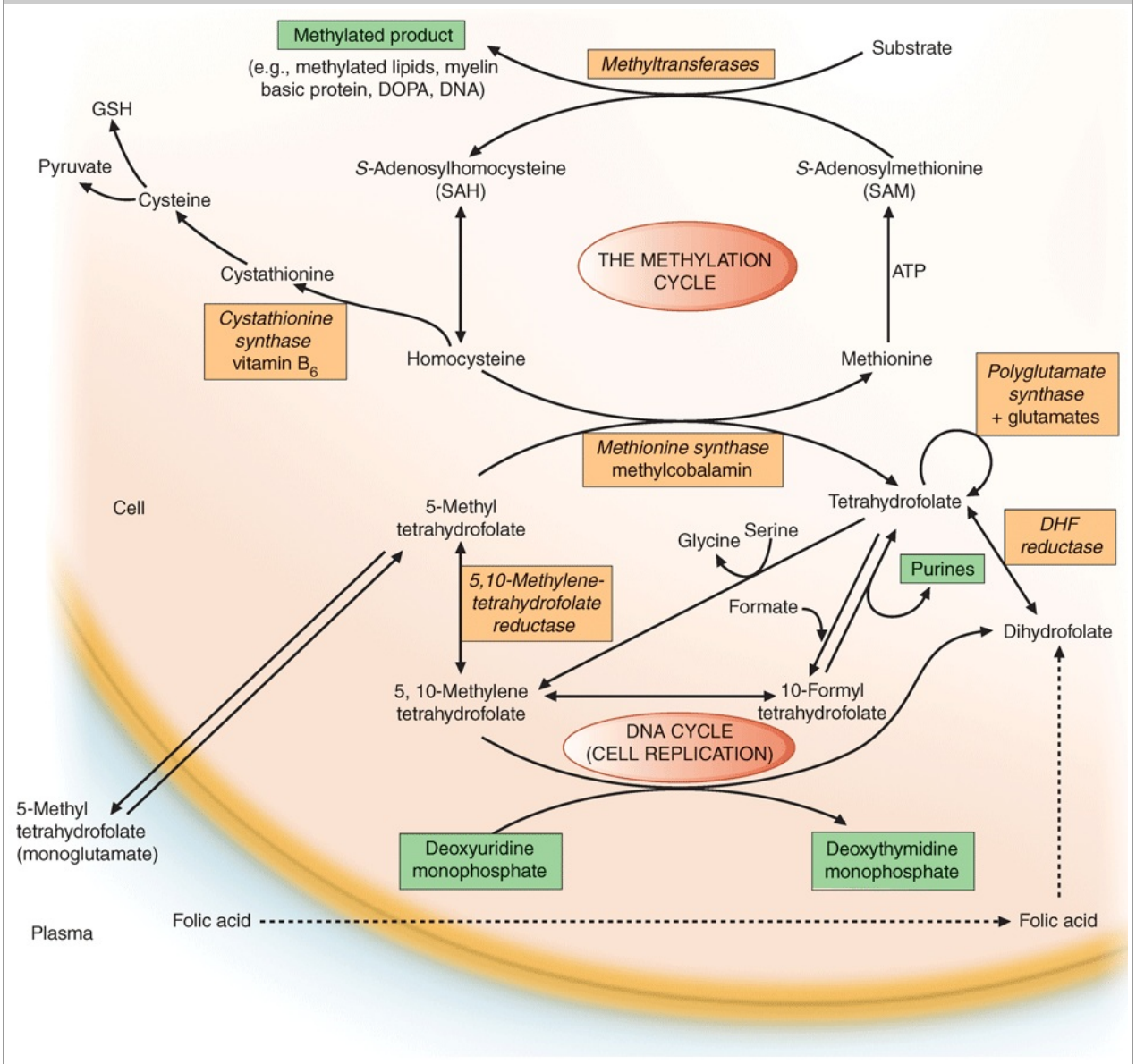

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The role of folates in DNA synthesis and in formation of *S*-adenosylmethionine (SAM), which is involved in numerous methylation reactions. *[Reprinted from AV Hoffbrand et al (eds), Postgraduate Haematology, 5th ed, Oxford, UK, Blackwell Publishing, 2005; with permission.]*

During thymidylate synthesis, 5,10-methylene-THF is oxidized to DHF (dihydrofolate). The enzyme DHF reductase converts this to THF. The drugs methotrexate, pyrimethamine, and (mainly in bacteria) trimethoprim inhibit DHF reductase and so prevent formation of active THF coenzymes from DHF. A small fraction of the folate coenzyme is not recycled during thymidylate synthesis but is degraded.

BIOCHEMICAL BASIS OF MEGALOBLASTIC ANEMIA

The common feature of all megaloblastic anemias is a defect in DNA synthesis that affects rapidly dividing cells in the bone marrow. All conditions that give rise to megaloblastic changes have in common a disparity in the rate of synthesis or availability of the four immediate precursors of DNA: the deoxyribonucleoside triphosphates (dNTPs)—dA(adenine)TP and dG(guanine)TP (purines), dT(thymine)TP and dC(cytosine)TP (pyrimidines). In deficiencies of either folate or cobalamin, there is failure to convert deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP), the precursor of dTTP (Fig. 105-1). This is the case because folate is needed as the coenzyme 5,10-methylene-THF polyglutamate for conversion of dUMP to dTMP; the availability of 5,10-methylene-THF is reduced in either cobalamin or folate deficiency. An alternative theory for megaloblastic anemia in cobal-amin or folate deficiency is misincorporation of uracil into DNA because of a buildup of deoxyuridine triphosphate (dUTP) at the DNA replication fork as a consequence of the block in conversion of dUMP to dTMP.

COBALAMIN-FOLATE RELATIONS

Folate is required for many reactions in mammalian tissues. Only two reactions in the body are known to require cobalamin. Methylmalonyl CoA isomerization requires adocobalamin, and the methylation of homocysteine to methionine requires both methylcobalamin and 5-MTHF (Fig. 105-1). This reaction is the first step in the pathway by which 5-MTHF, which enters bone marrow and other cells from plasma, is converted into all the intracellular folate coenzymes. The coenzymes are all polyglutamated (the larger size aiding retention in the cell), but the enzyme folate polyglutamate synthase can use only THF, not MTHF, as substrate. In cobalamin deficiency, MTHF accumulates in plasma, and intracellular folate concentrations fall due to failure of formation of THF, the substrate on which folate polyglutamates are built. This has been termed *THF starvation*, or the *methylfolate trap*.

This theory explains the abnormalities of folate metabolism that occur in cobalamin deficiency [high serum folate, low cell folate, positive purine precursor aminoimidazole carboxamide ribonucleotide (AICAR) excretion; Table 105–2] and also why the anemia of cobalamin deficiency responds to folic acid in large doses.

CLINICAL FEATURES

Many symptomless patients are detected through the finding of a raised mean corpuscular volume (MCV) on a routine blood count. The main clinical features in more severe cases are those of anemia. Anorexia is usually marked, and there may be weight loss, diarrhea, or constipation. Glossitis, angular cheilosis, a mild fever in more severely anemic patients, jaundice (unconjugated), and reversible melanin skin hyperpigmentation also may occur with a deficiency of either folate or cobalamin. Thrombocytopenia sometimes leads to bruising, and this may be aggravated by vitamin C deficiency or alcohol in malnourished patients. The anemia and low leukocyte count may predispose to infections, particularly of the respiratory and urinary tracts. Cobalamin deficiency has also been associated with impaired bactericidal function of phagocytes.

GENERAL TISSUE EFFECTS OF COBALAMIN AND FOLATE DEFICIENCIES

Epithelial Surfaces

After the marrow, the next most frequently affected tissues are the epithelial cell surfaces of the mouth, stomach, and small intestine and the respiratory, urinary, and female genital tracts. The cells show macrocytosis, with increased numbers of multinucleate and dying cells. The deficiencies may cause cervical smear abnormalities.

Complications of Pregnancy

The gonads are also affected, and infertility is common in both men and women with either deficiency. Maternal folate deficiency has been implicated as a cause of prematurity, and both folate deficiency and cobalamin deficiency have been implicated in recurrent fetal loss and neural tube defects, as discussed below.

Neural Tube Defects

Folic acid supplements at the time of conception and in the first 12 weeks of pregnancy reduce by ~70% the incidence of neural tube defects (NTDs) (anencephaly, meningomyelocele, encephalocele, and spina bifida) in the fetus. Most of this protective effect can be achieved by taking folic acid, 0.4 mg daily at the time of conception.

The incidence of cleft palate and harelip also can be reduced by prophylactic folic acid. There is no clear simple relationship between maternal folate status and these fetal abnormalities, although overall the lower the maternal folate, the greater the risk to the fetus. NTDs also can be caused by antifolate and antiepileptic drugs.

An underlying maternal folate metabolic abnormality has also been postulated. One abnormality has been identified: reduced activity of the enzyme 5,10-methylene-THF reductase (MTHFR) (Fig. 105-1) caused by a common C677T polymorphism in the *MTHFR* gene. In one study, the prevalence of this polymorphism was found to be higher in the parents of NTD fetuses and in the fetuses themselves: Homozygosity for the TT mutation was found in 13% compared with 5% in control subjects. The polymorphism codes for a thermolabile form of MTHFR. The homozygous state results in a lower mean serum and red cell folate level compared with control subjects, as well as significantly higher serum homocysteine levels. Tests for mutations in other enzymes possibly associated with NTDs, e.g., methionine synthase and serine–glycine hydroxymethylase, have been negative. Autoantibodies to folate receptors were suggested to be more common in mothers of NTD babies, but this has been disproved.

Cardiovascular Disease

Children with severe homocystinuria (blood levels ≥100 μmol/L) due to deficiency of one of three enzymes, methionine synthase, MTHFR, or cystathionine synthase (Fig. 105-1), have vascular disease, e.g., ischemic heart disease, cerebrovascular disease, or pulmonary embolus as teenagers or in young adulthood. Lesser degrees of raised serum homocysteine and low levels of serum folate and homozygous inherited mutations of MTHFR have been found to be associated with cerebrovascular, peripheral vascular, and coronary heart disease and with deep vein thrombosis. Prospective randomized trials of lowering homocysteine levels with supplements of folic acid, vitamin B12, and vitamin B6 against placebo over a 5-year period in patients with vascular disease or diabetes have not, however, shown a reduction of major cardiovascular events, nor have these supplements reduced the risk of recurrent cardiovascular disease after an acute myocardial infarct. It is possible that these trials were not sufficiently powered to detect a small (e.g., 10%) benefit or that another underlying factor is responsible for both the vascular damage and the raised homocysteine. Alternatively, the beneficial effects of lowering homocysteine were offset in these trials by the vitamins stimulating endothelial cell proliferation. Meta-analysis has suggested that folic acid supplementation reduces the risk of stroke by 18%. The results of longer and larger trials are needed to resolve these uncertainties.

Malignancy

Prophylactic folic acid in pregnancy has been found in some but not all studies to reduce the subsequent incidence of acute lymphoblastic leukemia (ALL) in childhood. A significant negative association has also been found with the *MTHFR* C677T polymorphism and leukemias with mixed lineage leukemia (MLL) translocations, but a positive association with hyperdiploidy in infants with ALL or acute myeloid leukemia or with childhood ALL. A second polymorphism in the *MTHFR* gene, A1298C, is also strongly associated with hyperdiploid leukemia. There are various positive and negative associations between polymorphisms in folate-dependent enzymes and the incidence of adult ALL. The C677T polymorphism is thought to lead to increased thymidine pools and "better quality" of DNA synthesis by shunting 1-carbon groups toward thymidine and purine synthesis. This may explain its reported association with a lower risk for colorectal cancer. Most but not all studies suggest that prophylactic folic acid also protects against colon adenomas. Other tumors that have been associated with folate polymorphisms or status include follicular lymphoma, breast cancer, and gastric

cancer. Because folic acid may "feed" tumors, it probably should be avoided in those with established tumors unless there is severe megaloblastic anemia due to folate deficiency.

Neurologic Manifestations

Cobalamin deficiency may cause a bilateral peripheral neuropathy or degeneration (demyelination) of the posterior and pyramidal tracts of the spinal cord and, less frequently, optic atrophy or cerebral symptoms.

The patient, more frequently male, presents with paresthesias, muscle weakness, or difficulty in walking and sometimes dementia, psychotic disturbances, or visual impairment. Long-term nutritional cobalamin deficiency in infancy leads to poor brain development and impaired intellectual development. Folate deficiency has been suggested to cause organic nervous disease, but this is uncertain, although methotrexate injected into the cerebrospinal fluid may cause brain or spinal cord damage.

An important clinical problem is the nonanemic patient with neurologic or psychiatric abnormalities and a low or borderline serum cobalamin level. In such patients, it is necessary to try to establish whether there is significant cobalamin deficiency, e.g., by careful examination of the blood film, tests for serum gastrin level and for antibodies to IF or parietal cells, along with serum methylmalonic acid (MMA) measurement if available. A trial of cobalamin therapy for at least 3 months will usually also be needed to determine whether the symptoms improve.

The biochemical basis for cobalamin neuropathy remains obscure. Its occurrence in the absence of methylmalonic aciduria in TC II deficiency suggests that the neuropathy is related to the defect in homocysteine-methionine conversion. Accumulation of *S*-adenosylhomocysteine in the brain, resulting in inhibition of transmethylation reactions, has been suggested.

Psychiatric disturbance is common in both folate and cobalamin deficiencies. This, like the neuropathy, has been attributed to a failure of the synthesis of SAM, which is needed in methylation of biogenic amines (e.g., dopamine) as well as that of proteins, phospholipids, and neurotransmitters in the brain (Fig. 105-1). Associations between lower serum folate or cobalamin levels and higher homocysteine levels and the development of decreased cognitive function and dementia in Alzheimer's disease have been reported. A 2-year double-blind placebo-controlled randomized clinical trial involving healthy subjects >65 years old given folate, cobalamin, and vitamin B6 supplements showed no benefit for cognitive performance, whereas a 3-year (FACIT) study did show benefit.

HEMATOLOGIC FINDINGS

PERIPHERAL BLOOD

Oval macrocytes, usually with considerable anisocytosis and poikilocytosis, are the main feature **(Fig. 105-2*A*)**. The MCV is usually >100 fL unless a cause of microcytosis (e.g., iron deficiency or thalassemia trait) is present. Some of the neutrophils are hypersegmented (more than five nuclear lobes). There may be leukopenia due to a reduction in granulocytes and lymphocytes, but this is usually >1.5 x 10^9/L; the platelet count may be moderately reduced, rarely to <40 x 10^9/L. The severity of all these changes parallels the degree of anemia. In a nonanemic patient, the presence of a few macrocytes and hypersegmented neutrophils in the peripheral blood may be the only indication of the underlying disorder.

Figure 105-2

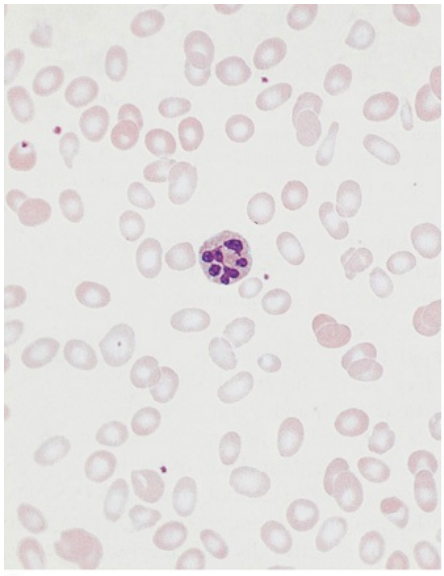

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

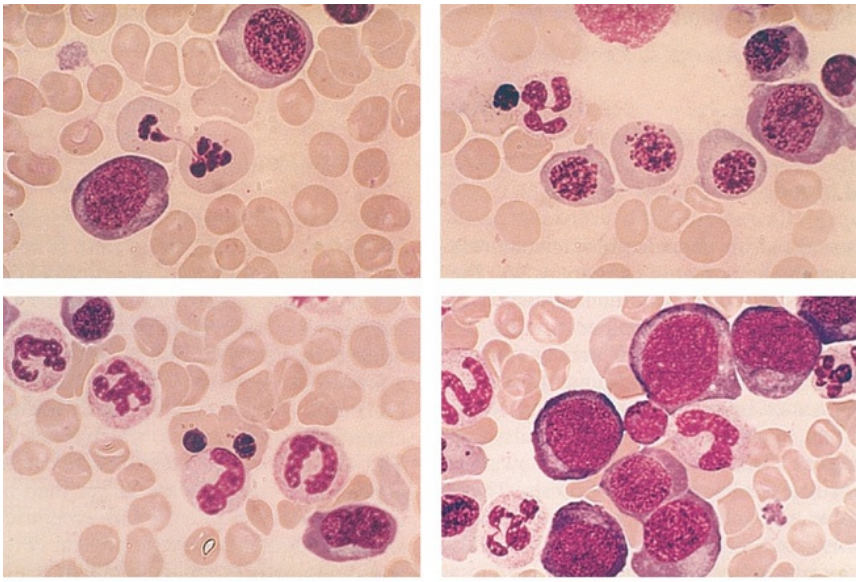

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. The peripheral blood in severe megaloblastic anemia. ***B.*** The bone marrow in severe megaloblastic anemia. *[Reprinted from AV Hoffbrand et al (eds), Postgraduate Haematology, 5th ed, Oxford, UK, Blackwell Publishing, 2005; with permission.]*

BONE MARROW

In a severely anemic patient, the marrow is hypercellular with an accumulation of primitive cells due to selective death by apoptosis of more mature forms. The erythroblast nucleus maintains a primitive appearance despite maturation and hemoglobinization of the cytoplasm. The cells are larger than normoblasts, and an increased number of cells with eccentric lobulated nuclei or nuclear fragments may be present (Fig. 105-2*B*). Giant and abnormally shaped metamyelocytes and enlarged hyperpolyploid megakaryocytes are characteristic. In less anemic patients, the changes in the marrow may be difficult to recognize. The terms *intermediate*, *mild*, and *early* have been used. The term *megaloblastoid* does not mean mildly megaloblastic. It is used to describe cells with both immature-appearing nuclei and defective hemoglobinization and is usually seen in myelodysplasia.

CHROMOSOMES

Bone marrow cells, transformed lymphocytes, and other proliferating cells in the body show a variety of changes, including random breaks, reduced contraction, spreading of the centromere, and exaggeration of secondary chromosomal constrictions and overprominent satellites. Similar abnormalities may be produced by antimetabolite drugs (e.g., cytosine arabinoside, hydroxyurea, and methotrexate) that interfere with either DNA replication or folate metabolism and that also cause megaloblastic appearances.

INEFFECTIVE HEMATOPOIESIS

There is an accumulation of unconjugated bilirubin in plasma due to the death of nucleated red cells in the marrow (ineffective erythropoiesis). Other evidence for this includes raised urine urobilinogen, reduced haptoglobins and positive urine hemosiderin, and a raised serum lactate dehydrogenase. A weakly positive direct antiglobulin test due to complement can lead to a false diagnosis of autoimmune hemolytic anemia.

CAUSES OF COBALAMIN DEFICIENCY

Cobalamin deficiency is usually due to malabsorption. The only other cause is inadequate dietary intake.

INADEQUATE DIETARY INTAKE

Adults

Dietary cobalamin deficiency arises in vegans who omit meat, fish, eggs, cheese, and other animal products from their diet. The largest group in the world consists of Hindus, and it is likely that many millions of Indians are at risk of deficiency of cobalamin on a nutritional basis. Subnormal serum cobalamin levels are found in up to 50% of randomly selected, young, adult Indian vegans, but the deficiency usually does not progress to megaloblastic anemia since the diet of most vegans is not totally lacking in cobalamin and the enterohepatic circulation of cobalamin is intact. Dietary cobalamin deficiency may also arise rarely in nonvegetarian individuals who exist on grossly inadequate diets because of poverty or psychiatric disturbance.

Infants

Cobalamin deficiency has been described in infants born to severely cobalamin-deficient mothers. These infants develop megaloblastic anemia at about 3–6 months of age, presumably because they are born with low stores of cobalamin and because they are fed breast milk with low cobalamin content. The babies have also shown growth retardation, impaired psychomotor development, and other neurologic sequelae.

GASTRIC CAUSES OF COBALAMIN MALABSORPTION

See Tables 105–3 and 105–4.

Table 105–3 Causes of Cobalamin Deficiency Sufficiently Severe to Cause Megaloblastic Anemia

Nutritional	Vegans

Malabsorption	Pernicious anemia
Gastric causes	Congenital absence of intrinsic factor or functional abnormality Total or partial gastrectomy
Intestinal causes	Intestinal stagnant loop syndrome: jejunal diverticulosis, ileocolic fistula, anatomic blind loop, intestinal stricture, etc. Ileal resection and Crohn's disease Selective malabsorption with proteinuria Tropical sprue Transcobalamin II deficiency Fish tapeworm

Table 105–4 Malabsorption of Cobalamin May Occur in the Following Conditions But Is Not Usually Sufficiently Severe and Prolonged to Cause Megaloblastic Anemia

Gastric causes
Simple atrophic gastritis (food cobalamin malabsorption)
Zollinger–Ellison syndrome
Gastric bypass surgery
Use of proton pump inhibitors
Intestinal causes
Gluten-induced enteropathy
Severe pancreatitis
HIV infection
Radiotherapy
Graft-versus-host disease
Deficiencies of cobalamin, folate, protein, ?riboflavin, ?nicotinic acid
Therapy with colchicine, para-aminosalicylate, neomycin, slow-release potassium chloride, anticonvulsant drugs, metformin, phenformin, cytotoxic drugs
Alcohol

Pernicious Anemia

Pernicious anemia (PA) may be defined as a severe lack of IF due to gastric atrophy. It is a common disease in north Europeans but occurs in all countries and ethnic groups. The overall incidence is about 120 per 100,000 population in the United Kingdom (UK). The ratio of incidence in men and women among whites is ~1:1.6, and the peak age of onset is 60 years, with only 10% of patients being <40 years of age. However, in some ethnic groups, notably black individuals and Latin Americans, the age at onset of PA is generally lower. The disease occurs more commonly than by chance in close relatives and in persons with other organ-specific autoimmune diseases, e.g., thyroid diseases, vitiligo, hypoparathyroidism, and Addison's disease. It is also associated with hypogammaglobulinemia, with premature graying or blue eyes, and persons of blood group A. An association with human leukocyte antigen (HLA) 3 has been reported in some but not all series and, in those with endocrine disease, with HLA-B8, -B12, and -BW15. Life expectancy is normal in women once regular treatment has begun. Men have a slightly subnormal life expectancy as a result of a higher incidence of carcinoma of the stomach than in control subjects. Gastric output of hydrochloric acid, pepsin, and IF is severely reduced. The serum gastrin level is raised, and serum pepsinogen I levels are low.

Gastric Biopsy

This usually shows atrophy of all layers of the body and fundus, with loss of glandular elements, an absence of parietal and chief cells and replacement by mucous cells, a mixed inflammatory cell infiltrate, and perhaps intestinal metaplasia. The infiltrate of plasma cells and lymphocytes contains an excess of CD4 cells. The antral mucosa is usually well preserved. *Helicobacter pylori* infection occurs infrequently in PA, but it has been suggested that *H. pylori* gastritis occurs at an early phase of atrophic gastritis and presents in younger patients as iron-deficiency anemia but in older patients as PA. *H. pylori* is suggested to stimulate an autoimmune process directed against parietal cells, with the *H. pylori* infection then being gradually replaced, in some individuals, by an autoimmune process.

Serum Antibodies

Two types of IF immunoglobulin G antibody may be found in the sera of patients with PA. One, the "blocking," or type I, antibody, prevents the combination of IF and cobalamin, whereas the "binding," or type II, antibody prevents attachment of IF to ileal mucosa. Type I occurs in the sera of ~55% of patients, and type II in 35%. IF antibodies cross the placenta and may cause temporary IF deficiency in a newborn infant. Patients with PA also show cell-mediated immunity to IF. Type I antibody has been detected rarely in the sera of patients without PA but with thyrotoxicosis, myxedema, Hashimoto's disease, or diabetes mellitus and in relatives of PA patients. IF antibodies also have been detected in gastric juice in ~80% of PA patients. These gastric antibodies may reduce absorption of dietary cobalamin by combining with small amounts of remaining IF.

Parietal cell antibody is present in the sera of almost 90% of adult patients with PA but is frequently present in other subjects. Thus, it occurs in as many as 16% of randomly selected female subjects age >60 years. The parietal cell antibody is directed against the α and β subunits of the gastric proton pump (H^+,K^+-ATPase).

JUVENILE PERNICIOUS ANEMIA

This usually occurs in older children and resembles PA of adults. Gastric atrophy, achlorhydria, and serum IF antibodies

are all present, although parietal cell antibodies are usually absent. About one-half of these patients show an associated endocrinopathy such as autoimmune thyroiditis, Addison's disease, or hypoparathyroidism; in some, mucocutaneous candidiasis occurs.

CONGENITAL INTRINSIC FACTOR DEFICIENCY OR FUNCTIONAL ABNORMALITY

An affected child usually presents with megaloblastic anemia in the first to third year of life; a few have presented as late as the second decade. The child usually has no demonstrable IF but has a normal gastric mucosa and normal secretion of acid. The inheritance is autosomal recessive. Parietal cell and IF antibodies are absent. Variants have been described in which the child is born with IF that can be detected immunologically but is unstable or functionally inactive, unable to bind cobalmin or to facilitate its uptake by ileal receptors.

GASTRECTOMY

After total gastrectomy, cobalamin deficiency is inevitable, and prophylactic cobalamin therapy should be commenced immediately after the operation. After partial gastrectomy, 10–15% of patients also develop this deficiency. The exact incidence and time of onset are most influenced by the size of the resection and the preexisting size of cobalamin body stores.

FOOD COBALAMIN MALABSORPTION

Failure of release of cobalamin from binding proteins in food is believed to be responsible for this condition, which is more common in the elderly. It is associated with low serum cobalamin levels, with or without raised serum levels of MMA and homocysteine. Typically, these patients have normal cobalamin absorption, as measured with crystalline cobalamin, but show malabsorption when a modified test using food-bound cobalamin is used. The frequency of progression to severe cobalamin deficiency and the reasons for this progression are not clear.

INTESTINAL CAUSES OF COBALAMIN MALABSORPTION

Intestinal Stagnant Loop Syndrome

Malabsorption of cobalamin occurs in a variety of intestinal lesions in which there is colonization of the upper small intestine by fecal organisms. This may occur in patients with jejunal diverticulosis, enteroanastomosis, or an intestinal stricture or fistula or with an anatomic blind loop due to Crohn's disease, tuberculosis, or an operative procedure.

Ileal Resection

Removal of ≥1.2 m of terminal ileum causes malabsorption of cobalamin. In some patients after ileal resection, particularly if the ileocecal valve is incompetent, colonic bacteria may contribute further to the onset of cobalamin deficiency.

Selective Malabsorption of Cobalamin with Proteinuria (Imerslund Syndrome: Imerslund-Gräsbeck Syndrome; Congenital Cobalamin Malabsorption; Autosomal Recessive Megaloblastic Anemia, MGA1)

This autosomally recessive disease is the most common cause of megaloblastic anemia due to cobalamin deficiency in infancy in Western countries. More than 200 cases have been reported, with familial clusters in Finland, Norway, the Middle East, and North Africa. The patients secrete normal amounts of IF and gastric acid but are unable to absorb cobalamin. In Finland, impaired synthesis, processing, or ligand binding of cubilin due to inherited mutations is found. In Norway, mutation of the gene for *AMN* has been reported. Other tests of intestinal absorption are normal. Over 90% of these patients show nonspecific proteinuria, but renal function is otherwise normal and renal biopsy has not shown any consistent renal defect. A few have shown aminoaciduria and congenital renal abnormalities, such as duplication of the renal pelvis.

Tropical Sprue

Nearly all patients with acute and subacute tropical sprue show malabsorption of cobalamin; this may persist as the principal abnormality in the chronic form of the disease, when the patient may present with megaloblastic anemia or neuropathy due to cobalamin deficiency. Absorption of cobalamin usually improves after antibiotic therapy and, in the early stages, folic acid therapy.

Fish Tapeworm Infestation

The fish tapeworm (*Diphyllobothrium latum*) lives in the small intestine of humans and accumulates cobalamin from food, rendering the cobalamin unavailable for absorption. Individuals acquire the worm by eating raw or partly cooked fish. Infestation is common around the lakes of Scandinavia, Germany, Japan, North America, and Russia. Megaloblastic anemia or cobalamin neuropathy occurs only in those with a heavy infestation.

Gluten-Induced Enteropathy

Malabsorption of cobalamin occurs in ~30% of untreated patients (presumably those in whom the disease extends to the ileum). Cobalamin deficiency is not severe in these patients and is corrected with a gluten-free diet.

Severe Chronic Pancreatitis

In this condition, lack of trypsin is thought to cause dietary cobalamin attached to gastric non-IF (R) binder to be unavailable for absorption. It also has been proposed that in pancreatitis, the concentration of calcium ions in the ileum falls below the level needed to maintain normal cobalamin absorption.

HIV Infection

Serum cobalamin levels tend to fall in patients with HIV infection and are subnormal in 10–35% of those with AIDS. Malabsorption of cobalamin not corrected by IF has been shown in some, but not all, patients with subnormal serum cobalamin levels. Cobalamin deficiency sufficiently severe to cause megaloblastic anemia or neuropathy is rare.

Zollinger–Ellison Syndrome

Malabsorption of cobalamin has been reported in the Zollinger–Ellison syndrome. It is thought that there is a failure to release cobalamin from R-binding protein due to inactivation of pancreatic trypsin by high acidity, as well as interference with IF binding of cobalamin.

Radiotherapy

Both total-body irradiation and local radiotherapy to the ileum (e.g., as a complication of radiotherapy for carcinoma of the cervix) may cause malabsorption of cobalamin.

Graft-versus-Host Disease

This commonly affects the small intestine. Malabsorption of cobalamin due to abnormal gut flora, as well as damage to ileal mucosa, is common.

Drugs

The drugs that have been reported to cause malabsorption of cobalamin are listed in Table 105–4. Megaloblastic anemia due to these drugs is, however, rare.

ABNORMALITIES OF COBALAMIN METABOLISM

Congenital Transcobalamin II Deficiency or Abnormality

Infants with TC II deficiency usually present with megaloblastic anemia within a few weeks of birth. Serum cobalamin and folate levels are normal, but the anemia responds to massive (e.g., 1 mg three times weekly) injections of cobalamin. Some cases show neurologic complications. The protein may be present but functionally inert. Genetic abnormalities found include mutations of an intra-exonic cryptic splice site, extensive deletion, single nucleotide deletion, nonsense mutation, and an RNA editing defect. Malabsorption of cobalamin occurs in all cases, and serum immunoglobulins are usually reduced. Failure to institute adequate cobalamin therapy or treatment with folic acid may lead to neurologic damage.

Congenital Methylmalonic Acidemia and Aciduria

Infants with this abnormality are ill from birth with vomiting, failure to thrive, severe metabolic acidosis, ketosis, and mental retardation. Anemia, if present, is normocytic and normoblastic. The condition may be due to a functional defect in either mitochondrial methylmalonyl CoA mutase or its cofactor adocobalamin. Mutations in the methylmalonyl CoA mutase are not responsive, or only poorly responsive, to treatment with cobalamin. A proportion of infants with failure of adocobalamin synthesis respond to cobalamin in large doses. Some children have combined methylmalonic aciduria and homocy-stinuria due to defective formation of both cobalamin coenzymes. This usually presents in the first year of life with feeding difficulties, developmental delay, microcephaly, seizures, hypotonia, and megaloblastic anemia.

Acquired Abnormality of Cobalamin Metabolism: Nitrous Oxide Inhalation

Nitrous oxide irreversibly oxidizes methylcobalamin to an inactive precursor; this inactivates methionine synthase. Megaloblastic anemia has occurred in patients undergoing prolonged N_2O anesthesia (e.g., in intensive care units). A neuropathy resembling cobalamin neuropathy has been described in dentists and anesthetists who are exposed repeatedly to N_2O. Methylmalonic aciduria does not occur as adocobalamin is not inactivated by N_2O.

CAUSES OF FOLATE DEFICIENCY

(Table 105–5)

Table 105–5 Causes of Folate Deficiency

Dietary[a]
Particularly in: old age, infancy, poverty, alcoholism, chronic invalids, and the psychiatrically disturbed; may be associated with scurvy or kwashiorkor
Malabsorption
Major causes of deficiency
Tropical sprue, gluten-induced enteropathy in children and adults, and in association with dermatitis herpetiformis, specific malabsorption of folate, intestinal megaloblastosis caused by severe cobalamin or folate deficiency
Minor causes of deficiency
Extensive jejunal resection, Crohn's disease, partial gastrectomy, congestive heart failure, Whipple's disease, scleroderma, amyloid, diabetic enteropathy, systemic bacterial infection, lymphoma, salazopyrine
Excess utilization or loss
Physiologic
Pregnancy and lactation, prematurity
Pathologic
Hematologic diseases: chronic hemolytic anemias, sickle cell anemia, thalassemia major, myelofibrosis
Malignant diseases: carcinoma, lymphoma, leukemia, myeloma
Inflammatory diseases: tuberculosis, Crohn's disease, psoriasis, exfoliative dermatitis, malaria
Metabolic disease: homocystinuria
Excess urinary loss: congestive heart failure, active liver disease
Hemodialysis, peritoneal dialysis
Antifolate drugs[b]
Anticonvulsant drugs (phenytoin, primidone, barbiturates), sulphasalazine
Nitrofurantoin, tetracycline, antituberculosis (less well documented)
Mixed causes
Liver diseases, alcoholism, intensive care units

[a]In severely folate-deficient patients with causes other than those listed under Dietary, poor dietary intake is often present. [b]Drugs inhibiting dihydrofolate reductase are discussed in the text.

NUTRITIONAL

Dietary folate deficiency is common. Indeed, in most patients with folate deficiency a nutritional element is present.

Certain individuals are particularly prone to have diets containing inadequate amounts of folate (Table 105–5). In the United States and other countries where fortification of the diet with folic acid has been adopted, the prevalence of folate deficiency has dropped dramatically and is now almost restricted to high-risk groups with increased folate needs. Nutritional folate deficiency occurs in kwashiorkor and scurvy and in infants with repeated infections or those who are fed solely on goats' milk, which has a low folate content.

MALABSORPTION

Malabsorption of dietary folate occurs in tropical sprue and in gluten-induced enteropathy. In the rare congenital syndrome of selective malabsorption of folate due to mutation of the protein-coupled folate transporter (PCFT), there is an associated defect of folate transport into the cerebrospinal fluid, and these patients show megaloblastic anemia, which responds to physiologic doses of folic acid given parenterally but not orally. They also show mental retardation, convulsions, and other central nervous system abnormalities. Minor degrees of malabsorption may also occur after jejunal resection or partial gastrectomy, in Crohn's disease, and in systemic infections, but in these conditions, if severe deficiency occurs, it is usually largely due to poor nutrition. Malabsorption of folate has been described in patients receiving salazopyrine, cholestyramine, and triamterene.

EXCESS UTILIZATION OR LOSS

Pregnancy

Folate requirements are increased by 200–300 μg to ~400 μg daily in a normal pregnancy, partly because of transfer of the vitamin to the fetus but mainly because of increased folate catabolism due to cleavage of folate coenzymes in rapidly proliferating tissues. Megaloblastic anemia due to this deficiency is prevented by prophylactic folic acid therapy. It occurred in 0.5% of pregnancies in the UK and other Western countries before prophylaxis with folic acid, but the incidence is much higher in countries where the general nutritional status is poor.

Prematurity

A newborn infant, whether full term or premature, has higher serum and red cell folate concentrations than does an adult. However, a newborn infant's demand for folate has been estimated to be up to 10 times that of adults on a weight basis, and the neonatal folate level falls rapidly to the lowest values at about 6 weeks of age. The falls are steepest and are liable to reach subnormal levels in premature babies, a number of whom develop megaloblastic anemia responsive to folic acid at about 4–6 weeks of age. This occurs particularly in the smallest babies (<1500 g birth weight) and those who have feeding difficulties or infections or have undergone multiple exchange transfusions. In these babies, prophylactic folic acid should be given.

Hematologic Disorders

Folate deficiency frequently occurs in chronic hemolytic anemia, particularly in sickle cell disease, autoimmune hemolytic anemia, and congenital spherocytosis. In these and other conditions of increased cell turnover (e.g., myelofibrosis, malignancies), folate deficiency arises because it is not completely reutilized after performing coenzyme functions.

Inflammatory Conditions

Chronic inflammatory diseases such as tuberculosis, rheumatoid arthritis, Crohn's disease, psoriasis, exfoliative dermatitis, bacterial endocarditis, and chronic bacterial infections cause deficiency by reducing the appetite and increasing the demand for folate. Systemic infections also may cause malabsorption of folate. Severe deficiency is virtually confined to the patients with the most active disease and the poorest diet.

Homocystinuria

This is a rare metabolic defect in the conversion of homocysteine to cystathionine. Folate deficiency occurring in most of these patients may be due to excessive utilization because of compensatory increased conversion of homocysteine to methionine.

Long-Term Dialysis

As folate is only loosely bound to plasma proteins, it is easily removed from plasma by dialysis. In patients with anorexia, vomiting, infections, and hemolysis, folate stores are particularly likely to become depleted. Routine folate prophylaxis is now given.

Congestive Heart Failure, Liver Disease

Excess urinary folate losses of >100 μg per day may occur in some of these patients. The explanation appears to be release of folate from damaged liver cells.

ANTIFOLATE DRUGS

A large number of epileptics who are receiving long-term therapy with phenytoin or primidone, with or without barbiturates, develop low serum and red cell folate levels. The exact mechanism is unclear. Alcohol may also be a folate antagonist, as patients who are drinking spirits may develop megaloblastic anemia that will respond to normal quantities of dietary folate or to physiologic doses of folic acid only if alcohol is withdrawn. Macrocytosis of red cells is associated with chronic alcohol intake even when folate levels are normal. Inadequate folate intake is the major factor in the development of deficiency in spirit-drinking alcoholics. Beer is relatively folate-rich in some countries, depending on the technique used for brewing.

The drugs that inhibit DHF reductase include methotrexate, pyrimethamine, and trimethoprim. Methotrexate has the most powerful action against the human enzyme, whereas trimethoprim is most active against the bacterial enzyme and is likely to cause megaloblastic anemia only when used in conjunction with sulphamethoxazole in patients with preexisting folate or cobalamin deficiency. The activity of pyrimethamine is intermediate. The antidote to these drugs is folinic acid (5-formyl-THF).

CONGENITAL ABNORMALITIES OF FOLATE METABOLISM

Some infants with congenital defects of folate enzymes (e.g., cyclohydrolase or methionine synthase) have had megaloblastic anemia.

DIAGNOSIS OF COBALAMIN AND FOLATE DEFICIENCIES

The diagnosis of cobalamin or folate deficiency has traditionally depended on the recognition of the relevant abnormalities in the peripheral blood and analysis of the blood levels of the vitamins.

SERUM COBALAMIN

This is measured by an automated enzyme-linked immunosorbent assay (ELISA). Normal serum levels range from 118–148 pmol/L (160–200 ng/L) to ~738 pmol/L (1000 ng/L). In patients with megaloblastic anemia due to cobalamin deficiency, the level is usually <74 pmol/L (100 ng/L). In general, the more severe the deficiency, the lower the serum cobalamin level. In patients with spinal cord damage due to the deficiency, levels are very low even in the absence of anemia. Values between 74 and 148 pmol/L (100 and 200 ng/L) are regarded as borderline. They may occur, for instance, in pregnancy, in patients with megaloblastic anemia due to folate deficiency. They may also be due to heterozygous, homozygous, or compound heterozygous mutations of the gene *TCN1* that codes for haptocorrin (transcobalamin I). There is no clinical or rematologic abnormality. The serum cobalamin level is sufficiently robust, cost-effective, and most convenient to rule out cobalamin deficiency in the vast majority of patients suspected of having this problem.

SERUM METHYLMALONATE AND HOMOCYSTEINE

In patients with cobalamin deficiency sufficient to cause anemia or neuropathy, the serum MMA level is raised. Sensitive methods for measuring MMA and homocysteine in serum have been introduced and recommended for the early diagnosis of cobalamin deficiency, even in the absence of hematologic abnormalities or subnormal levels of serum cobalamin. Serum MMA levels fluctuate, however, in patients with renal failure. Mildly elevated serum MMA and/or homocysteine levels occur in up to 30% of apparently healthy volunteers, with serum cobalamin levels up to 258 pmol/L (350 ng/L) and normal serum folate levels; 15% of elderly subjects, even with cobalamin levels >258 pmol/L (>350 ng/L), have this pattern of raised metabolite levels. These findings bring into question the exact cutoff points for normal MMA and homocysteine levels. It is also unclear at present whether these mildly raised metabolite levels have clinical consequences.

Serum homocysteine is raised in both early cobalamin and folate deficiency but may be raised in other conditions, e.g., chronic renal disease, alcoholism, smoking, pyridoxine deficiency, hypothyroidism, and therapy with steroids, cyclosporine, and other drugs. Levels are also higher in serum than in plasma, in men than in premenopausal women, in women taking hormone replacement therapy or in oral contraceptive users, and in elderly persons and patients with several inborn errors of metabolism affecting enzymes in trans-sulfuration pathways of homocysteine metabolism. Thus, homocysteine levels are not used for diagnosis of cobalamin or folate deficiency.

OTHER TESTS

Studies of cobalamin absorption once were widely used, but difficulty in obtaining radioactive cobalamin and ensuring that IF preparations are free of viruses has made these tests obsolete. Tests to diagnose PA include serum gastrin, which is raised, and serum pepsinogen I, which is low in PA (90–92%) but also in other conditions. Tests for IF and parietal cell antibodies are also used as well as tests for individual intestinal diseases.

SERUM FOLATE

This is also measured by an ELISA technique. In most laboratories, the normal range is from 11 nmol/L (2 μg/L) to ~82 nmol/L (15 μg/L). The serum folate level is low in all folate-deficient patients. It also reflects recent diet. Because of this, serum folate may be low before there is hematologic or biochemical evidence of deficiency. Serum folate rises in severe cobalamin deficiency because of the block in conversion of MTHF to THF inside cells; raised levels have also been reported in the intestinal stagnant loop syndrome due to absorption of bacterially synthesized folate.

RED CELL FOLATE

The red cell folate assay is a valuable test of body folate stores. It is less affected than the serum assay by recent diet and traces of hemolysis. In normal adults, concentrations range from 880–3520 μmol/L (160–640 μg/L) of packed red cells. Subnormal levels occur in patients with megaloblastic anemia due to folate deficiency but also in nearly two-thirds of patients with severe cobalamin deficiency. False-normal results may occur if a folate-deficient patient has received a recent blood transfusion or if a patient has a raised reticulocyte count.

Treatment: Megaloblastic Anemia

It is usually possible to establish which of the two deficiencies, folate or cobalamin, is the cause of the anemia and to treat only with the appropriate vitamin. In patients who enter the hospital severely ill, however, it may be necessary to treat with both vitamins in large doses once blood samples have been taken for cobalamin and folate assays and a bone marrow biopsy has been performed (if deemed necessary). Transfusion is usually unnecessary and inadvisable. If it is essential, packed red cells should be given slowly, one or two units only, with the usual treatment for heart failure if present. Potassium supplements have been recommended to obviate the danger of the hypokalemia but are not necessary. Occasionally, an excessive rise in platelets occurs after 1–2 weeks of therapy. Antiplatelet therapy, e.g., aspirin, should be considered if the platelet count rises to >800 x 10^9/L.

COBALAMIN DEFICIENCY

It is usually necessary to treat patients who have developed cobalamin deficiency with lifelong regular cobalamin injections. In the UK, the form used is hydroxocobalamin; in the United States, cyanocobalamin. In a few instances, the underlying cause of cobalamin deficiency can be permanently corrected, e.g., fish tapeworm, tropical sprue, or an intestinal stagnant loop that is amenable to surgery. The indications for starting cobalamin therapy are a well-documented megaloblastic anemia or other hematologic abnormalities and neuropathy due to the deficiency. Patients with borderline serum cobalamin levels but no hematologic or other abnormality may be followed to make sure that the cobalamin deficiency does not progress (see below). If malabsorption of cobalamin or rises in serum MMA levels have been demonstrated, however, these patients also should be given regular maintenance cobalamin therapy. Cobalamin should be given routinely to all patients who have had a total gastrectomy or ileal resection. Patients who have undergone gastric reduction for control of obesity or who are receiving long-term treatment with proton pump inhibitors should be screened and, if necessary, given cobalamin replacement.

Replenishment of body stores should be complete with six 1000-μg IM injections of hydroxocobalamin given at 3- to 7-day intervals. More frequent doses are usually used in patients with cobalamin neuropathy, but there is no evidence that they produce a better response. Allergic reactions are rare and may require desensitization or antihistamine or glucocorticoid cover. For maintenance therapy, 1000 μg hydroxocobalamin IM once every 3 months is satisfactory. Because of the poorer retention of cyanocobalamin, protocols generally use higher and more frequent doses, e.g., 1000 μg IM, monthly, for maintenance treatment.

Because a small fraction of cobalamin can be absorbed passively through mucous membranes even when there is

complete failure of physiologic IF-dependent absorption, large daily oral doses (1000–2000 μg) of cyanocobalamin have been used in PA for replacement and maintenance of normal cobalamin status in, e.g., food malabsorption of cobalamin. Sublingual therapy has also been proposed for those in whom injections are difficult because of a bleeding tendency and who may not tolerate oral therapy. If oral therapy is used, it is important to monitor compliance, particularly with elderly, forgetful patients.

For treatment of patients with subnormal serum B_{12} levels with a normal MCV and no hypersegmentation of neutrophils, a negative IF antibody test in the absence of tests of B_{12} absorption is problematic. Some (perhaps 15%) cases may be due to TC I (HC) deficiency. Homocysteine and/or MMA measurements may help, but in the absence of these tests and with otherwise normal gastrointestinal function, repeat serum B_{12} assay after 6–12 months may help one decide whether to start cobalamin therapy.

FOLATE DEFICIENCY

Oral doses of 5–15 mg folic acid daily are satisfactory, as sufficient folate is absorbed from these extremely large doses even in patients with severe malabsorption. The length of time therapy must be continued depends on the underlying disease. It is customary to continue therapy for about 4 months, when all folate-deficient red cells will have been eliminated and replaced by new folate-replete populations.

Before large doses of folic acid are given, cobalamin deficiency must be excluded and, if present, corrected; otherwise cobalamin neuropathy may develop despite a response of the anemia of cobalamin deficiency to folate therapy. Studies in the United States, however, suggest that there is no increase in the proportion of individuals with low serum cobalamin levels and no anemia since food fortification with folic acid, but it is unknown if there has been a change in incidence of cobalamin neuropathy.

Long-term folic acid therapy is required when the underlying cause of the deficiency cannot be corrected and the deficiency is likely to recur, e.g., in chronic dialysis or hemolytic anemias. It may also be necessary in gluten-induced enteropathy that does not respond to a gluten-free diet. Where mild but chronic folate deficiency occurs, it is preferable to encourage improvement in the diet after correcting the deficiency with a short course of folic acid. In any patient receiving long-term folic acid therapy, it is important to measure the serum cobalamin level at regular (e.g., once-yearly) intervals to exclude the coincidental development of cobalamin deficiency.

Folinic Acid (5-Formyl-THF)

This is a stable form of fully reduced folate. It is given orally or parenterally to overcome the toxic effects of methotrexate or other DHF reductase inhibitors.

PROPHYLACTIC FOLIC ACID

In many countries, food is fortified with folic acid (in grain or flour) to prevent neural tube defects. It is also used in chronic dialysis patients and in parenteral feeds. Prophylactic folic acid has been used to reduce homocysteine levels to prevent cardiovascular disease, but further data are needed to assess the benefit for this and for cognitive function in the elderly.

Pregnancy

Folic acid, 400 μg daily, should be given as a supplement before and throughout pregnancy. In women who have had a previous fetus with a neural tube defect, 5 mg daily is recommended when pregnancy is contemplated and throughout the subsequent pregnancy.

Infancy and Childhood

The incidence of folate deficiency is so high in the smallest premature babies during the first 6 weeks of life that folic acid (e.g., 1 mg daily) should be given routinely to those weighing <1500 g at birth and to larger premature babies who require exchange transfusions or develop feeding difficulties, infections, or vomiting and diarrhea.

The World Health Organization currently recommends routine supplementation with iron and folic acid in children in countries where iron deficiency is common and child mortality, largely due to infectious diseases, is high. However, some studies suggest that in areas where malaria rates are high, this approach may increase the incidence of severe illness and death. Even where malaria is rare, there appears to be no survival benefit.

MEGALOBLASTIC ANEMIA NOT DUE TO COBALAMIN OR FOLATE DEFICIENCY OR ALTERED METABOLISM

This may occur with many antimetabolic drugs (e.g., hydroxy-urea, cytosine arabinoside, 6-mercaptopurine) that inhibit DNA replication. Antiviral nucleoside analogues used in treatment of HIV infection may also cause macrocytosis and megaloblastic marrow changes. In the rare disease orotic aciduria, two consecutive enzymes in purine synthesis are defective. The condition responds to therapy with uridine, which bypasses the block. In thiamine-responsive megaloblastic anemia, there is a genetic defect in the high-affinity thiamine transport (*SLC19A2*) gene. This causes defective RNA ribose synthesis through impaired activity of trans-ketolase, a thiamine-dependent enzyme in the pentose cycle. This leads to reduced nucleic acid production. It may be associated with diabetes mellitus and deafness and the presence of many ringed sideroblasts in the marrow. The explanation is unclear for megaloblastic changes in the marrow in some patients with acute myeloid leukemia and myelodysplasia.

FURTHER READINGS

Albert CM et al: Effect of folic acid and B vitamins on risk of cardiovascular events and total mortality among women at high risk for cardiovascular disease: A randomized trial. JAMA 299: 2027, 2008[PMID: 18460663]

Carmel R: How I treat cobalamin (vitamin B12) deficiency. Blood 12:2214, 2008

Dali-Youcef N et al: An update on cobalamin deficiency in adults. Q J Med 102:17, 2009[PMID: 18990719]

Durga J et al: Effect of 3-year folic acid supplementation on cognitive function in older adults in the FACIT trial: A randomized double-blind, controlled trial. Lancet 369:208, 2007[PMID: 17240287]

Lonn E et al: Homocysteine lowering with folic acid and B vitamins in vascular disease. N Engl J Med 354:1567, 2006[PMID: 16531613]

McMahon JA et al: A controlled trial of homocysteine lowering and cognitive performance. N Engl J Med 354:2764,

2006[PMID: 16807413]

Morris MS et al: Folate and vitamin B-12 status in relation to anemia, macrocytosis, and cognitive impairment in older Americans in the age of folic acid fortification. Am J Clin Nutr 85:193, 2007[PMID: 17209196]

Quadros EV: Advances in the understanding of cobalamin assimilation and metabolism. Br J Haematol 148:195, 2010[PMID: 19832808]

——— et al: The protein and the gene encoding the receptor for the cellular uptake of transcobalamin-bound cobalamin. Blood 113:186, 2008

Ulrich CM: Folate and cancer prevention—where to next? Cancer Epidemiol Biomarkers Prev 17:2226, 2008[PMID: 18768487]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEMOLYTIC ANEMIAS AND ANEMIA DUE TO ACUTE BLOOD LOSS: INTRODUCTION

Definitions

A finite life span is a distinct characteristic of red cells. Hence, a logical, time-honored classification of anemias is in three groups: (1) decreased production of red cells, (2) increased destruction of red cells, (3) acute blood loss. Decreased production is covered in Chaps. 103, 105, and 107; increased destruction and acute blood loss are covered in this chapter.

All patients who are anemic as a result of either increased destruction or acute blood loss have two important elements in common: the anemia results from overconsumption of red cells from the peripheral blood, yet the supply of cells from the bone marrow (in the absence of coexisting marrow disease) is usually increased, as reflected by a reticulocytosis. On the other hand, physical loss of red cells from the bloodstream—which in most cases also means physical loss *from* the body—is fundamentally different from destruction of red cells *within* the body. Therefore the clinical aspects and the pathophysiology of anemia in these two groups of patients are quite different, and they will be considered separately.

HEMOLYTIC ANEMIAS

With respect to primary etiology, anemias due to increased destruction of red cells, which we know as hemolytic anemias (HAs), may be *inherited* or *acquired* (Table 106–1). From the clinical point of view they may be more *acute* or more *chronic*, they may vary from mild to very severe, and the site of hemolysis may be predominantly *intravascular* or *extravascular*. With respect to mechanisms, HAs may be due to *intracorpuscular* causes or to *extracorpuscular* causes. But before reviewing the individual types of HA it is appropriate to consider what they have in common.

Table 106–1 Classification of Hemolytic Anemias*

	Intracorpuscular Defects	Extracorpuscular Factors
Hereditary	Hemoglobinopathies Enzymopathies Membrane-cytoskeletal defects	Familial (atypical) hemolytic uremic syndrome
Acquired	Paroxysmal nocturnal hemoglobinuria (PNH)	Mechanical destruction (microangiopathic) Toxic agents Drugs Infectious Autoimmune

*Hereditary causes correlate with intracorpuscular defects because these defects are due to inherited mutations. The one exception is PNH because the defect is due to an acquired somatic mutation. Similarly, acquired causes correlate with extracorpuscular factors because mostly these factors are exogenous. The one exception is familial hemolytic uremic syndrome (HUS; often referred to as atypical HUS) because here an inherited abnormality allows complement activation to be excessive, with bouts of production of membrane attack complex capable of destroying normal red cells.

General Clinical and Laboratory Features

The clinical presentation of a patient with anemia is greatly influenced in the first place by whether the onset is abrupt or gradual, and HAs are no exception. A patient with autoimmune HA or with favism may be a medical emergency, whereas a patient with mild hereditary spherocytosis or with cold agglutinin disease may be diagnosed after years. This is due in large measure to the remarkable ability of the body to adapt to anemia when it is slowly progressing (Chap. 57).

What differentiates HAs from other anemias is that the patient has signs and symptoms arising directly from hemolysis (Table 106–2). At the clinical level, the main sign is *jaundice*; in addition, the patient may report discoloration of the urine. In many cases of HA, the spleen is enlarged because it is a preferential site of hemolysis, and in some cases the liver may be enlarged as well. In all severe congenital forms of HA, there also may be skeletal changes due to overactivity of the bone marrow (although they are never as severe as they are in thalassemia).

Table 106–2 Some Common Features of Hemolytic Disorders

General examination	Jaundice, pallor
Other physical findings	Spleen may be enlarged; bossing of skull in severe congenital cases
Hemoglobin level	From normal to severely reduced
MCV, MCH	Usually increased
Reticulocytes	Increased
Bilirubin	Increased (mostly unconjugated)
LDH	Increased (up to 10x normal with intravascular hemolysis)
Haptoglobin	Reduced to absent (if hemolysis is part intravascular)

Abbreviations: LDH, lactate dehydrogenase; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume.

The laboratory features of HA are related to hemolysis per se and the erythropoietic response of the bone marrow. Hemolysis regularly produces in the serum an increase in unconjugated bilirubin and aspartate transaminase (AST); urobilinogen will be increased in both urine and stool. If hemolysis is mainly intravascular, the telltale sign is hemoglobinuria (often associated with hemosiderinuria); in the serum there is increased hemoglobin, lactate dehydrogenase (LDH) is increased, and haptoglobin is

reduced. In contrast, the bilirubin level may be normal or only mildly elevated. The main sign of the erythropoietic response by the bone marrow is an increase in reticulocytes (Table 106–2); a test all too often neglected in the initial workup of a patient with anemia. Usually the increase will be reflected in both the percentage of reticulocytes (the more commonly quoted figure) and the absolute reticulocyte count (the more definitive parameter). The increased number of reticulocytes is associated with an increased mean corpuscular volume (MCV) in the blood count. On the blood smear, this is reflected in the presence of macrocytes; there is also polychromasia and sometimes one sees nucleated red cells. In most cases, a bone marrow aspirate is not necessary in the diagnostic workup; if it is done, it will show erythroid hyperplasia. In practice, once a HA is suspected, specific tests will usually be required for a definitive diagnosis of a specific type of HA.

GENERAL PATHOPHYSIOLOGY

The mature red cell is the product of a developmental pathway that brings the phenomenon of differentiation to an extreme. An orderly sequence of events produces synchronous changes whereby the gradual accumulation of a huge amount of hemoglobin in the cytoplasm (to a final level of 340 g/L, i.e. about 5 mM) goes hand in hand with the gradual loss of cellular organelles and of biosynthetic abilities. In the end, the erythroid cell undergoes a process that has features of apoptosis, including nuclear pyknosis and actual loss of the nucleus. However, the final result is more altruistic than suicidal; the cytoplasmic body, instead of disintegrating, is now able to provide oxygen to all cells in the human organism for some remaining 120 days of the red cell "life" span.

As a result of this unique process of differentiation and maturation, intermediary metabolism is drastically curtailed in mature red cells (Fig. 106-1); for instance, cytochrome-mediated oxidative phosphorylation has been lost with the loss of mitochondria (through a process of physiologic autophagy); therefore, there is no backup to anaerobic glycolysis for the production of adenosine triphosphate (ATP). Also the capacity of making protein has been lost with the loss of ribosomes. This places the cell's limited metabolic apparatus at risk, because if any protein component deteriorates, it cannot be replaced, as it would be in most other cells, and in fact the activity of most enzymes gradually decreases as red cells age. Another consequence of the relative simplicity of red cells is that they have a very limited range of ways to manifest distress under hardship: in essence, any sort of metabolic failure will eventually lead either to structural damage to the membrane or to failure of the cation pump. In either case, the life span of the red cell is reduced, which is the definition of a *hemolytic disorder*. If the rate of red cell destruction exceeds the capacity of the bone marrow to produce more red cells, the hemolytic disorder will manifest as HA.

Figure 106-1

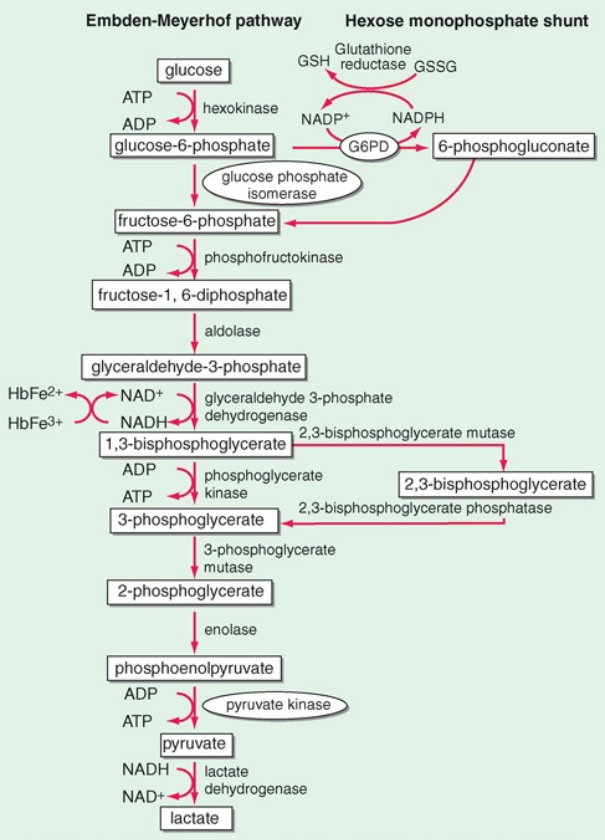

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

RBC metabolism. The Embden-Meyerhof pathway (glycolysis) generates ATP for energy and membrane maintenance. The generation of NADPH maintains hemoglobin in a reduced state. The hexose monophosphate shunt generates NADPH that is used to reduce glutathione, which protects the red cell against oxidant stress. Regulation of 2,3-bisphosphoglycerate levels is a critical determinant of oxygen affinity of hemoglobin. Enzyme deficiency states in order of prevalence: glucose-6-phosphate - dehydrogenase (G6PD) > pyruvate kinase > glucose-6-phosphate isomerase > rare deficiencies of other enzymes in the pathway. The more common enzyme deficiencies are encircled.

Thus, the essential pathophysiologic process common to all HAs is an increased red cell turnover. The gold standard for proving that the life span of red cells is reduced (compared to the normal value of about 120 days) is a *red cell survival* study, which can be carried out by labeling the red cells with ^{51}Cr and measuring residual radioactivity over several days or weeks; however, this classic test is now available in very few centers, and it is rarely necessary. If the hemolytic event is transient, it does not usually cause any long-term consequences, except for an increased requirement for erythropoietic factors, particularly folic acid. However, if hemolysis is recurrent or persistent, the increased bilirubin production favors the formation of gallstones. If a considerable proportion of hemolysis takes place in the spleen, as is often the case, splenomegaly may become increasingly a

feature, and hypersplenism may develop, with consequent neutropenia and/or thrombocytopenia.

The increased red cell turnover also has metabolic consequences. In normal subjects, the iron from effete red cells is very efficiently recycled by the body; however, with chronic intravascular hemolysis the persistent hemoglobinuria will cause considerable iron loss, needing replacement. With chronic extravascular hemolysis the opposite problem, iron overload, is more common, especially if the patient needs frequent blood transfusions. Chronic iron overload will cause secondary hemochromatosis: this will cause damage particularly to the liver, eventually leading to cirrhosis, and to the heart muscle, eventually causing heart failure.

COMPENSATED HEMOLYSIS VERSUS HA

Red cell destruction is a potent stimulus for erythropoiesis, which is mediated by erythropoietin (EPO) produced by the kidney. This mechanism is so effective that in many cases the increased output of red cells from the bone marrow can fully balance an increased destruction of red cells. In such cases we say that hemolysis is *compensated*. The pathophysiology of compensated hemolysis is similar to what we have just described, except there is no anemia. This notion is important from the diagnostic point of view, because a patient with a hemolytic condition, even an inherited one, may present without anemia. It is also important from the point of view of management, because compensated hemolysis may become "decompensated"—i.e., anemia may suddenly appear—in certain circumstances, for instance pregnancy, folate deficiency, or renal failure, interfering with adequate EPO production. Another general feature of chronic HAs is seen when any intercurrent condition, for instance an acute infection, depresses erythropoiesis. When this happens, in view of the increased rate of red cell turnover, the effect will be predictably much more marked than in a person who does not have hemolysis. The most dramatic example is infection by parvovirus B19, which may cause a rather precipitous fall in hemoglobin; an occurrence sometimes referred to as *aplastic crisis*.

Inherited Hemolytic Anemias

There are three essential components in the red cell: (1) hemoglobin, (2) the membrane-cytoskeleton complex, and (3) the metabolic machinery necessary to keep (1) and (2) in working order. Diseases caused by abnormalities of hemoglobin, or hemoglobinopathies, are covered in Chap. 104. Here we will deal with diseases of the other two components.

HEMOLYTIC ANEMIAS DUE TO ABNORMALITIES OF THE MEMBRANE-CYTOSKELETON COMPLEX

The detailed architecture of the red cell membrane is complex, but its basic design is relatively simple (Fig. 106-2). The lipid bilayer incorporates phospholipids and cholesterol, and it is spanned by a number of proteins that have their hydrophobic transmembrane domains embedded in the membrane. Most of these proteins have hydrophilic domains extending toward both the outside and the inside of the cell. Other proteins are tethered to the membrane through a glycosylphosphatidylinositol (GPI) anchor, and they have only an extracellular domain. These proteins are arranged roughly perpendicular to or lying across the membrane: they include ion channels, receptors for complement components, receptors for other ligands, and some of unknown function. The most abundant of these proteins are glycophorins and the so-called band 3, an anion transporter. The extracellular domains of many of these proteins are heavily glycosylated, and they carry antigenic determinants that correspond to blood groups. Underneath the membrane, and tangential to it, is a network of other proteins that make up the cytoskeleton: the main cytoskeletal protein is spectrin, the basic unit of which is a dimer of α-spectrin and β-spectrin. The membrane is physically linked to the cytoskeleton by a third set of proteins (including ankyrin and the so-called band 4.1 and band 4.2), which thus make these two structures intimately connected to each other.

Figure 106-2

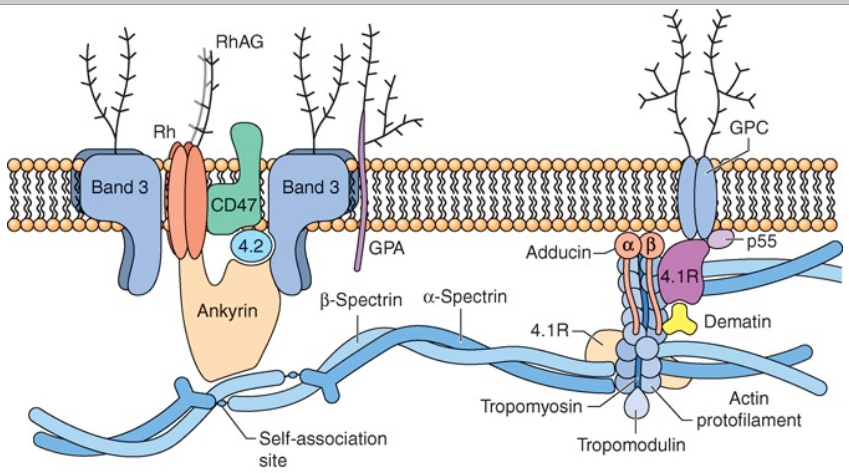

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of red cell membrane-cytoskeleton. (For explanation see text.) *(From N Young et al: Clinical Hematology. Copyright Elsevier, 2006; with permission.)*

The membrane-cytoskeleton complex is indeed so integrated that, not surprisingly, an abnormality of almost any of its components will be disturbing or disruptive, causing structural failure, which results ultimately in hemolysis. These abnormalities are almost invariably inherited mutations; thus, diseases of the membrane-cytoskeleton complex belong to the category of inherited HAs. Before the red cells lyse, they often exhibit more or less specific morphologic changes that alter the normal biconcave disk shape. Thus, the majority of the diseases in this group have been known for over a century as *hereditary spherocytosis* and *hereditary elliptocytosis*.

Hereditary Spherocytosis (HS)

This is a relatively common type of HA, with an estimated frequency of at least 1 in 5000. Its identification is credited to Minkowksy and Chauffard, who at the end of the nineteenth century reported families in which HS was inherited as an autosomal dominant condition (Fig 106-3A). From this seminal work, HS came to be defined as an inherited form of HA associated with the presence of spherocytes in the peripheral blood. In addition, in vitro studies revealed that the red cells were abnormally susceptible to lysis in hypotonic media: indeed, the presence of *osmotic fragility* became the main diagnostic test for HS. Today we know that HS, thus defined, is genetically heterogeneous; i.e., it can arise from a variety of mutations in one of several genes (Table 106–3). Whereas classically the inheritance of HS is autosomal dominant (with the patients being heterozygous), some severe forms are instead autosomal recessive (with the patient being homozygous).

Figure 106-3

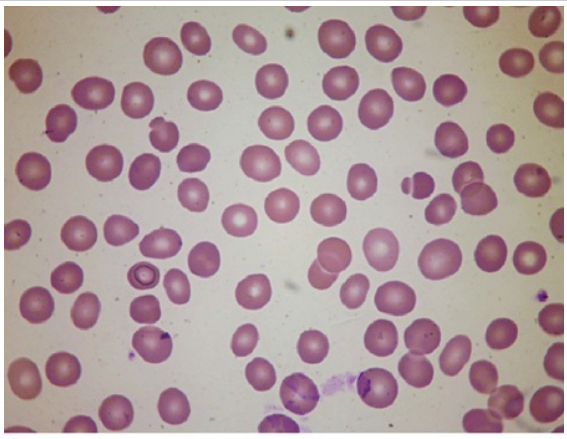

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

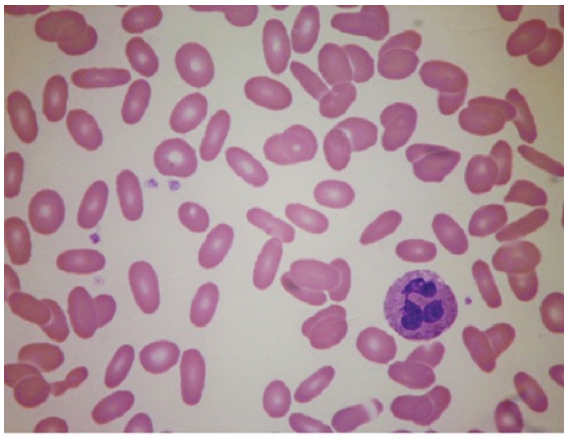

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

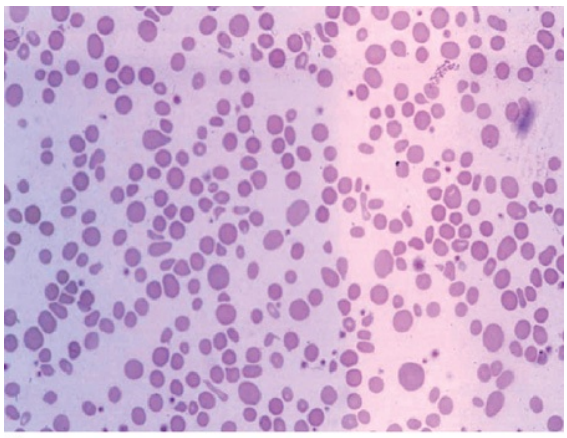

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Peripheral blood smear from patients with membrane-cytoskeleton abnormalities.*A.* Hereditary spherocytosis. ***B.*** Hereditary elliptocytosis, heterozygote. ***C.*** Elliptocytosis, with both alleles of the α-spectrin gene mutated.

Table 106–3 Inherited Diseases of the Red Cell Membrane-Cytoskeleton

Gene	Chromosomal Location	Protein Produced	Disease(s) with Certain Mutations (Inheritance)	Comments
SPTA1	1q22-q23	α-Spectrin	HS (recessive)	Rare
			HE (dominant)	Mutations of this gene account for about 65% of HE. More severe forms may be due to coexistence of an otherwise silent mutant allele
SPTB	14q23-q24.1	β-Spectrin	HS (dominant)	Rare

			HE (dominant)	Mutations of this gene account for about 30% of HE, including some severe forms
ANK1	8p11.2	Ankyrin	HS (dominant)	May account for majority of HS
SLC4A1	17q21	Band 3 (anion channel)	HS (dominant)	Mutations of this gene may account for about 25% of HS
			Southeast Asia ovalocytosis (dominant)	Polymorphic mutation (deletion of 9 amino acids); clinically asymptomatic; protective against *Plasmodium falciparum*
			Stomatocytosis	Certain specific missense mutations shift protein function from anion exchanger to cation conductance
EPB41	1p33-p34.2	Band 4.1	HE (dominant)	Mutations of this gene account for about 5% of HE: mostly with prominent morphology but no hemolysis in heterozygotes; severe hemolysis in homozygotes
EPB42	15q15-q21	Band 4.2	HS (recessive)	Mutations of this gene account for about 3% of HS.
RHAG	6p21.1-p11	Rhesus antigen	Chronic nonspherocytic hemolytic anemia	Very rare; associated with total loss of all Rh antigens

Abbreviations: HE, hereditary elliptocytosis; HS, hereditary spherocytosis.

Clinical Presentation and Diagnosis

The spectrum of clinical severity of HS is broad. Severe cases may present in infancy with severe anemia, whereas mild cases may present in young adults or even later in life. In women, HS is sometimes first diagnosed when anemia is investigated during pregnancy. The main clinical findings are jaundice, an enlarged spleen, and often gallstones; indeed, it is often the finding of gallstones in a young person that triggers diagnostic investigations.

The variability in clinical manifestations that is observed among patients with HS is largely due to the different underlying molecular lesions (Table 106–3). Not only are mutations of several genes involved, but individual mutations of the same gene can also give very different clinical manifestations. In milder cases hemolysis is often compensated (see above), and this may cause variation in time, even in the same patient, due to the fact that intercurrent conditions (e.g., infection) cause decompensation. The anemia is usually normocytic, with the characteristic morphology that gives the disease its name. A characteristic feature is an increase in mean corpuscular hemoglobin concentration (MCHC): this is almost the only condition in which an increased MCHC is seen.

When there is a family history (Fig. 106-3A) it is usually easy to suspect the diagnosis, but there may be no family history for at least two reasons. (1) The patient may have a de novo mutation, i.e., a mutation that has taken place in a germ cell of one of his or her parents or early after zygote formation. (2) The patient may have a recessive form of HS (Table 106–3). In most cases, the diagnosis can be made on the basis of red cell morphology and of a test for osmotic fragility, a modified version of which is called the "pink test." In some cases, a definitive diagnosis can be obtained only by molecular studies demonstrating a mutation in one of the genes underlying HS. This is usually carried out in laboratories with special expertise in this area.

TREATMENT: HEREDITARY SPHEROCYTOSIS

We don't have a causal treatment for HS; i.e., no way has yet been found to correct the basic defect in the membrane-cytoskeleton structure. However, it has been apparent for a long time that the spleen plays a special role in HS through a dual mechanism. On one hand, like in many other HAs, the spleen itself is a major site of destruction; on the other hand, transit through the splenic circulation makes the defective red cells more spherocytic and therefore accelerates their demise, even though lysis may take place elsewhere. For these reasons, splenectomy has long been regarded as a prime, almost obligatory therapeutic measure in HS. Therefore, current guidelines (not evidence-based) are as follows. (1) Avoid splenectomy in mild cases. (2) Delay splenectomy until at least 4 years of age, after the risk of severe sepsis has peaked. (3) Antipneumococcal vaccination before splenectomy is imperative, whereas penicillin prophylaxis postsplenectomy is controversial. (4) There is no doubt that HS patients often may require cholecystectomy, in which case the practice has been to also carry out a splenectomy at the same time. Today the decision regarding this combined surgery should not be regarded as automatic: cholecystectomy is usually done via the laparoscopic approach and splenectomy should be carried out if clinically indicated.

Hereditary Elliptocytosis (He)

HE is at least as heterogeneous as HS, both from the genetic point of view (Table 106–3) and from the clinical point of view. Again it is the shape of the red cells that gives the name to these conditions, but there is no direct correlation between the elliptocytic morphology and clinical severity. In fact, some mild or even asymptomatic cases may have nearly 100% elliptocytes, whereas in severe cases it is all sorts of bizarre poikilocytes that predominate. Clinical features and recommended management are similar to those outlined above for HS. Although the spleen may not have the specific role it has in HS, in severe cases splenectomy may be beneficial. The prevalence of HE causing clinical disease is similar to that of HS. However, an asymptomatic form, referred to as Southeast Asia ovalocytosis, has a frequency of up to 7% in certain populations, presumably as a result of malaria selection.

Disorders of Cation Transport

These rare conditions with autosomal dominant inheritance are characterized by increased intracellular sodium in red cells, with concomitant loss of potassium: indeed, they are sometimes discovered through the incidental finding, in a blood test, of high serum K^+ (*pseudohyperkalemia*). In patients from some families, the cation transport disturbance is associated with gain of water: as a result the red cells are overhydrated (low MCHC), and on a blood smear the normally round-shaped central pallor is replaced by a linear-shaped central pallor, which has earned this disorder the name *stomatocytosis*. In patients from other families, the red cells are instead dehydrated (high MCHC), and their consequent rigidity has earned this disorder the name *xerocytosis*. In these disorders, one would suspect that the primary defect may be in a cation transporter. In most cases this has not yet been demonstrated, but interestingly, certain missense mutations of the *SLC4A1* gene encoding band 3 (Table 106–3) give stomatocytosis. Hemolysis can vary from relatively mild to quite severe. From the practical point of view it is important to know that splenectomy is contraindicated, as it has been followed in a majority of cases by severe thromboembolic complications.

Enzyme Abnormalities

When there is an important defect in the membrane or in the cytoskeleton, hemolysis is a direct consequence of the fact that the very structure of the red cell is abnormal. Instead, when one of the enzymes is defective, the consequences will depend on the precise role of that enzyme in the metabolic machinery of the red cell, which, in first approximation, has two important functions: (1) to provide energy in the form of ATP, and (2) to prevent oxidative damage to hemoglobin and to other proteins.

Abnormalities of the Glycolytic Pathway

Since red cells, in the course of their differentiation, have sacrificed not only their nucleus and their ribsomes but also their mitochondria, they rely exclusively on the anaerobic portion of the glycolytic pathway for producing energy in the form of ATP. Most of the ATP is required by the red cell for cation transport against a concentration gradient across the membrane. If this fails, due to a defect of any of the enzymes of the glycolytic pathway, the result will be hemolytic disease (Table 106–4).

Table 106–4 Red Cell Ezyme Abnormalities Causing Hemolysis

	Enzyme (Acronym)	Chromosomal Location	Prevalence of Enzyme Deficiency (Rank)	Clinical Manifestations Extra-Red Cell	Comments
GLYCOLYTIC PATHWAY	Hexokinase (HK)	10q22	Very rare		Other isoenzymes known
	Glucose 6-phosphate isomerase (G6PI)	19q31.1	Rare (4)*	NM, CNS	
	Phosphofructokinase (PFK)	12q13	Very rare	Myopathy	
	Aldolase	16q22-24	Very rare		
	Triose phosphate isomerase (TPI)	12p13	Very rare	CNS (severe), NM	
	Glyceraldehyde 3-phosphate dehydrogenase (GAPD)	12p13.31-p13.1	Very rare	Myopathy	
	Diphosphoglycerate mutase (DPGM)	7q31-q34	Very rare		Erythrocytosis rather than hemolysis
	Phosphoglycerate kinase (PGK)	Xq13	Very rare	CNS, NM	May benefit from splenectomy
	Pyruvate kinase (PK)	1q21	Rare (2)*		May benefit from splenectomy
REDOX	Glucose 6-phosphate dehydrogenase (G6PD)	Xq28	Common (1)*	Very rarely granulocytes	In almost all cases only AHA from exogenous trigger
	Glutathione synthase	20q11.2	Very rare	CNS	
	γ-Glutamylcysteine synthase	6p12	Very rare	CNS	
	Cytochrome b5 reductase	22q13.31-qter	Rare	CNS	Methemoglobinemia rather than hemolysis
NUCLEOTIDE METABOLISM	Adenylate kinase (AK)	9q34.1	Very rare	CNS	
	Pyrimidine 5'-nucleotidase (P5N)	3q11-q12	Rare (3)*		May benefit from splenectomy

*The numbers from (1) to (4) indicate the ranking order of these enzymopathies in terms of frequency.

Abbreviations: AHA, acquired hemolytic anemia; CNS, central nervous system.

Pyruvate Kinase Deficiency

Abnormalities of the glycolytic pathway are all inherited and all rare. Among them, deficiency of pyruvate kinase (PK) is the less rare, with an estimated prevalence of the order of 1:10,000. The clinical picture is that of an HA that often presents in the newborn with neonatal jaundice; the jaundice persists, and it is usually associated with a very high reticulocytosis. The anemia is of variable severity; sometimes it is so severe as to require regular blood transfusion treatment; sometimes it is mild, bordering on a nearly compensated hemolytic disorder. As a result, the diagnosis may be delayed, and in some cases it is made in young adults; for instance, in a woman, during her first pregnancy, when the anemia may get worse. In part, the delay in diagnosis is due to the fact that the anemia is remarkably well tolerated, because the metabolic block at the last step in glycolysis causes an increase in bisphosphoglycerate (or DPG), a major effector of the hemoglobin-oxygen dissociation curve; thus, the oxygen delivery to the tissues is enhanced.

Treatment: Pyruvate Kinase Deficiency

The management of PK deficiency is mainly supportive. In view of the marked increase in red cell turnover, oral folic acid supplements should be given constantly. Blood transfusion should be used as necessary, and iron chelation may have to be added if the blood transfusion requirement is high enough to cause iron overload. In these patients, who have more severe disease, splenectomy may be beneficial. There is a single case report of curative treatment of PK deficiency by bone marrow transplantation from an HLA-identical PK-normal sibling. This seems a viable option for severe cases when a sibling donor is available.

Other Glycolytic Enzyme Abnormalities

All of these defects are rare to very rare (Table 106–4), and all cause HA with varying degrees of severity. It is not unusual for the presentation to be in the guise of severe neonatal jaundice, which may require exchange transfusion. If the anemia is less severe, it may present later in life, or it may even remain asymptomatic and be detected incidentally when a blood count is done for unrelated reasons. The spleen is often enlarged. When other systemic manifestations occur, they involve the central nervous system, sometimes entailing severe mental retardation (particularly in the case of triose phosphate isomerase deficiency), or the neuromuscular system, or both. The *diagnosis* of HA is usually not difficult, thanks to the triad of normo-macrocytic anemia, reticulocytosis, and hyperbilirubinemia. Enzymopathies should be considered in the differential diagnosis of any chronic Coombs-negative hemolytic anemia. In most cases of glycolytic enzymopathies, the morphologic abnormalities of red cells characteristically seen in membrane disorders are conspicuous by their absence. A definitive diagnosis can be made only by demonstrating the deficiency of an individual enzyme by quantitative assays carried out in only a few specialized laboratories. If a particular molecular abnormality is already known in the family, then of course one could test directly for that at the DNA level, bypassing the need for enzyme assays.

Abnormalities of Redox Metabolism

G6PD Deficiency

Glucose 6-phosphate dehydrogenase (G6PD) is a housekeeping enzyme critical in the redox metabolism of all aerobic cells

(Fig. 106-1). In red cells its role is even more critical, because it is the only source of NADPH that directly and via glutathione (GSH) defends these cells against oxidative stress. G6PD deficiency is a prime example of an HA due to interaction between an intracorpuscular cause and an extracorpuscular cause, because in the majority of cases, hemolysis is triggered by an exogenous agent. Although a decrease in G6PD activity is noted in most tissues of G6PD-deficient subjects, the decrease is less marked than in red cells, and it does not seem to have a clinical impact.

Genetic Considerations

The G6PD gene is X-linked, and this has important implications. First, as males have only one G6PD gene (i.e., they are hemizygous for this gene), they must be either normal or G6PD-deficient. By contrast, females, having two G6PD genes, can be either normal or deficient (homozygous), or intermediate (heterozygous). As a result of the phenomenon of X-chromosome inactivation, heterozygous females are genetic mosaics, with a highly variable ratio of G6PD-normal to G6PD-deficient cells and an equally variable degree of clinical expression: some heterozygotes can be just as affected as hemizygous males. The enzymatically active form of G6PD is either a dimer or a tetramer of a single protein subunit of 514 amino acids. G6PD- deficient subjects have been found invariably to have mutations in the coding region of the G6PD gene (Fig. 106-4). Almost all of some 150 different mutations known are single missense point mutations, entailing single amino acid replacements in the G6PD protein. In most cases, these mutations cause G6PD deficiency by decreasing the in vivo stability of the protein; thus, the physiologic decrease in G6PD activity that takes place with red cell aging is greatly accelerated. In some cases an amino acid replacement can also affect the catalytic function of the enzyme.

Figure 106-4

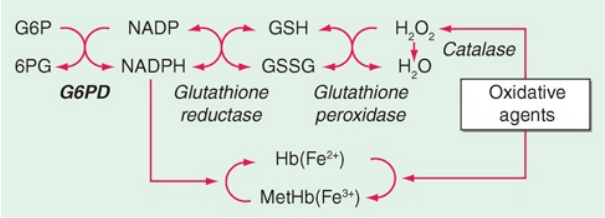

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of redox metabolism in the red cell. G6P, glucose 6-phosphate; 6PG, 6-phosphogluconate; G6PD, glucose 6-phosphate dehydrogenase; GSH, reduced glutathione; GSSG, oxidized glutathione; Hb, hemoglobin; MetHb, methemoglobin; NADP, nicotinamide adenine -dinucleotide phosphate; NADPH, reduced nicotinamide adenine dinucleotide phosphate.

Among these mutations, those underlying *chronic nonspherocytic hemolytic anemia* (CNSHA; see "Clinical Manifestations," below) are a discrete subset. This much more severe clinical phenotype can be ascribed in some cases to adverse qualitative changes (for instance, a decreased affinity for the substrate, glucose 6-phosphate) or simply to the fact that the enzyme deficit is more extreme because of a more severe instability of the enzyme. For instance, a cluster of mutations map at or near the dimer interface, and clearly they compromise severely the formation of the dimer.

Epidemiology

G6PD deficiency is widely distributed in tropical and subtropical parts of the world (Africa, Southern Europe, the Middle East, Southeast Asia, and Oceania) (Fig. 106-5) and wherever people from those areas have migrated. A conservative estimate is that at least 400 million people have a G6PD deficiency gene. In several of these areas the frequency of a G6PD deficiency gene may be as high as 20% or more. It would be quite extraordinary for a trait that causes significant pathology to spread widely and reach high frequencies in many populations without conferring some biologic advantage. Indeed, G6PD is one of the best characterized examples of genetic polymorphisms in the human species. Clinical field studies and in vitro experiments strongly support the view that G6PD deficiency has been selected by *Plasmodium falciparum* malaria, by virtue of the fact that it confers a relative resistance against this highly lethal infection. Whether this protective effect is exerted mainly in hemizygous males or in females heterozygous for G6PD deficiency is still not quite clear. Different G6PD variants underlie G6PD deficiency in different parts of the world. Some of the more widespread variants are G6PD Mediterranean on the shores of that sea, in the Middle East, and in India; G6PD A–in Africa and in Southern Europe; G6PD Vianchan and G6PD Mahidol in Southeast Asia; G6PD Canton in China; and G6PD Union worldwide. The heterogeneity of polymorphic G6PD variants is proof of their independent origin, and it supports the notion that they have been selected by a common environmental agent, in keeping with the concept of convergent evolution (Fig. 106-5).

Figure 106-5

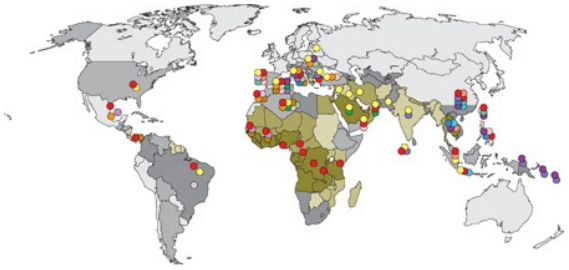

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Epidemiology of G6PD deficiency throughout the world. The different shadings indicate increasingly high levels of prevalence, up to about 20%; the different colored symbols indicate individual genetic variants of G6PD, each one having a different mutation. *[From L Luzzatto et al, in C Scriver et al (eds): The Metabolic & Molecular Bases of Inherited Disease, 8th ed. New York, McGraw-Hill, 2001.]*

Clinical Manifestations

The vast majority of people with G6PD deficiency remain clinically asymptomatic throughout their lifetime; however, all of them have an increased risk of developing neonatal jaundice (NNJ), and a risk of developing acute hemolytic anemia (AHA) when challenged by a number of oxidative agents. NNJ related to G6PD deficiency is very rarely present at birth. The peak incidence

of clinical onset is between day 2 and day 3, and in most cases the anemia is not severe. However, NNJ can be very severe in some G6PD-deficient babies, especially in association with prematurity, infection, and/or environmental factors (such as naphthalene-camphor balls used in babies' bedding and clothing), and the risk of severe NNJ is also increased by the coexistence of a monoallelic or biallelic mutation in the uridyl transferase gene (*UGT1A1*; the same mutations are associated with the Gilbert syndrome). If inadequately managed, NNJ associated with G6PD deficiency can produce kernicterus and permanent neurologic damage.

AHA can develop as a result of three types of triggers: (1) fava beans, (2) infections, and (3) drugs (Table 106–5). Typically, a hemolytic attack starts with malaise, weakness, and abdominal or lumbar pain. After an interval of several hours to 2–3 days, the patient develops jaundice and often dark urine, due to hemoglobinuria. The onset can be extremely abrupt, especially with favism in children. The anemia is from moderate to extremely severe. It is usually normocytic and normochromic, and it is due partly to intravascular hemolysis. Hence, it is associated with hemoglobinemia, hemoglobinuria, high LDH, and low or absent plasma haptoglobin. The blood film shows anisocytosis, polychromasia, and spherocytes (Fig. 106-6). The most typical feature is the presence of bizarre poikilocytes, with red cells that appear to have unevenly distributed hemoglobin ("hemighosts") and red cells that appear to have had parts of them bitten away ("bite cells" or "blister cells"). A classical test, now rarely carried out, is supravital staining with methyl violet that, if done promptly, reveals the presence of Heinz bodies, consisting of precipitates of denatured hemoglobin and regarded as a signature of oxidative damage to red cells (except for the rare occurrence of an unstable hemoglobin). LDH is high and so is the unconjugated bilirubin, indicating that there is also extravascular hemolysis. The most serious threat from AHA in adults is the development of acute renal failure (this is exceedingly rare in children). Once the threat of acute anemia is over, and in the absence of comorbidity, full recovery from AHA associated with G6PD deficiency is the rule.

Table 106–5 Drugs that Carry Risk of Clinical Hemolysis in Persons with G6PD Deficiency

	Definite Risk	Possible Risk	Doubtful Risk
Antimalarials	Primaquine	Chloroquine	Quinine
	Dapsone/chlorproguanil*		
Sulphonamides/sulphones	Sulfamethoxazole Others	Sulfasalazine Sulfadimidine	Sulfisoxazole Sulfadiazine
	Dapsone		
Antibacterial/antibiotics	Cotrimoxazole	Ciprofloxacin	Chloramphenicol
	Nalidixic acid	Norfloxacin	*p*-Aminosalicylic acid
	Nitrofurantoin		
Antipyretic/analgesics	Niridazole Acetanilide	Acetylsalicylic acid high dose (>3 g/d)	Acetylsalicylic acid (<3 g/d)
	Phenazopyridine		Acetaminophen
			Phenacetin
Other	Naphthalene	Vitamin K analogues	Doxorubicin
	Methylene blue	Ascorbic acid >1 g Rasburicase	Probenecid

*Marketed as Lapdap from 2003 to 2008.

Figure 106-6

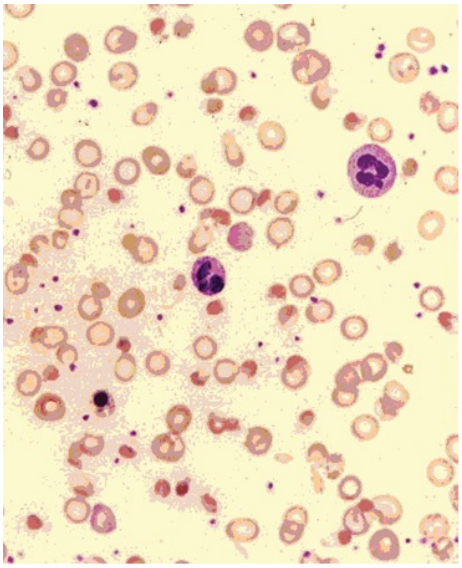

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Peripheral blood smear from a 5-year-old G6PD-deficient boy with acute favism.

A very small minority of subjects with G6PD deficiency have *chronic nonspherocytic hemolytic anemia* (CNSHA) of variable severity. The patient is always a male, usually with a history of NNJ, who may present with anemia, unexplained jaundice, or because of gallstones later in life. The spleen may be enlarged. The severity of anemia ranges in different patients from borderline to transfusion-dependent. The anemia is usually normo-macrocytic, with reticulocytosis. Bilirubin and LDH are

increased. Although hemolysis is, by definition, chronic in these patients, they are also vulnerable to acute oxidative damage, and therefore the same agents that can cause acute HA in people with the ordinary type of G6PD deficiency will cause severe exacerbations in people with the severe form of G6PD deficiency. In some cases of CNSHA, the deficiency of G6PD is so severe in granulocytes that it becomes rate-limiting for their oxidative burst, with consequent increased susceptibility to some bacterial infections.

Laboratory Diagnosis

The suspicion of G6PD deficiency can be confirmed by semiquantitative methods often referred to as screening tests, which are suitable for population studies and can correctly classify male subjects, in the steady state, as G6PD-normal or G6PD- deficient. However, in clinical practice, a diagnostic test is usually needed when the patient has had a hemolytic attack. This implies that the oldest, most G6PD-deficient red cells have been selectively destroyed, and young red cells, having higher G6PD activity, are being released into the circulation. Under these conditions, only a quantitative test can give a definitive result. In males, this test will identify normal hemizygotes and G6PD-deficient hemizygotes; among females, some heterozygotes will be missed, but those who are at most risk of hemolysis will be identified.

Treatment: G6PD Deficiency

The acute hemolytic anemia of G6PD deficiency is largely preventable by avoiding exposure to triggering factors of previously screened subjects. Of course, the practicability and cost-effectiveness of screening depends on the prevalence of G6PD deficiency in each community. Favism is entirely preventable in G6PD-deficient subjects by not eating fava beans. Drug-induced hemolysis can be prevented by testing for G6PD deficiency before prescribing; in most cases, one can use alternative drugs. When AHA develops and once its cause is recognized, in most cases no specific treatment is needed. However, if the anemia is severe, it may be a medical emergency, especially in children, requiring immediate action, including blood transfusion. This has been the case with an antimalarial drug combination containing dapsone (called Lapdap, introduced as recently as 2003) that has caused severe acute hemolytic episodes in children with malaria in several African countries; after a few years, it was taken off the market. If there is acute renal failure, hemodialysis may be necessary, but if there is no previous kidney disease, recovery is the rule. The management of NNJ associated with G6PD deficiency is no different from that of NNJ due to other causes.

In cases with CNSHA, if the anemia is not severe, regular folic acid supplements and regular hematologic surveillance will suffice. It will be important to avoid exposure to potentially hemolytic drugs, and blood transfusion may be indicated when exacerbations occur, mostly in concomitance with intercurrent infection. In rare patients, regular blood transfusions may be required, in which case appropriate iron chelation should be instituted. Unlike in hereditary spherocytosis, there is no evidence of selective red cell destruction in the spleen; however, in practice, splenectomy has proven beneficial in severe cases.

Other Abnormalities of the Redox System

As mentioned above, GSH is a key player in the defense against oxidative stress. Inherited defects of GSH metabolism are exceedingly rare, but each one of them can give rise to chronic HA (Table 106–4). A rare, peculiar, usually self-limited severe HA of the first month of life, called *infantile poikilocytosis*, may be associated with deficiency of glutathione peroxidase (GSHPx) due not to an inherited abnormality but to transient nutritional deficiency of selenium, an element essential for the activity of GSHPx.

Pyrimidine 5'-Nucleotidase (P5N) Deficiency

P5N is a key enzyme in the catabolism of nucleotides arising from the degradation of nucleic acids that takes place in the final stages of erythroid cell maturation. How exactly its deficiency causes HA is not well understood, but a highly distinctive feature of this condition is a morphologic abnormality of the red cells known as *basophilic stippling*. The condition is rare, but it probably ranks third in frequency among red cell enzyme defects (after G6PD deficiency and PK deficiency). The anemia is lifelong, of variable severity, and may benefit from splenectomy.

Familial (Atypical) Hemolytic Uremic Syndrome (aHUS)

This phrase is used to designate a group of rare disorders, mostly affecting children, characterized by microangiopathic HA with presence of fragmented erythrocytes in the peripheral blood smear, thrombocytopenia (usually mild), and acute renal failure. (The word *atypical* is part of the phrase because it is the HUS caused by infection with *Escherichia coli* producing the Shiga toxin that is regarded as typical). The genetic basis of aHUS has been elucidated only recently. Studies of more than 100 families have revealed that those family members who have developed HUS have mutations in any one of several genes encoding complement regulatory proteins: complement factor H (*CFH*), CD46 or membrane cofactor protein (*MCP*), complement factor I (*CFI*), complement component C3, complement factor B (*CFB*), and thrombomodulin. Thus, whereas all other inherited HAs are due to intrinsic red cell abnormalities, this group is unique in that hemolysis results from an inherited defect external to red cells (Table 106–1). Because the regulation of the complement cascade has considerable redundancy, in the steady state any of the above abnormalities can be tolerated. However, when an intercurrent infection or some other trigger activates complement through the alternative pathway, the deficiency of one of the complement regulators becomes critical. Endothelial cells get damaged, especially in the kidney, and at the same time and partly as a result of this, there will be brisk hemolysis [thus, the more common Shiga toxin–related HUS (Chap. 149) can be regarded as a phenocopy of aHUS]. Atypical HUS is a severe disease with up to 15% mortality in the acute phase and up to 50% of cases progressing to end-stage renal disease. Atypical HUS often undergoes spontaneous remission, and the best tested form of treatment is plasma exchange, which supplies the deficient complement regulator. Because the basis of aHUS is an inherited abnormality, it is not surprising that given exposure to an appropriate trigger, the syndrome will tend to recur: when it does, the prognosis is always serious. In some cases, kidney (and liver) transplantation has been carried out, but the role of these procedures is controversial.

Acquired Hemolytic Anemia

MECHANICAL DESTRUCTION OF RED CELLS

Although red cells are characterized by the remarkable deformability that enables them to squeeze through capillaries narrower than themselves for thousands of times in their lifetime, there are at least two situations in which they succumb to shear, if not to wear and tear. The result is intravascular hemolysis, resulting in hemoglobinuria. One situation is acute and self-inflicted, *march hemoglobinuria*. Why sometimes a marathon runner may develop this complication, whereas on another occasion this does not happen, we don't know (perhaps her or his footwear needs attention). A similar syndrome may develop after prolonged barefoot ritual dancing. The other situation is chronic and iatrogenic (it has been called *microangiopathic hemolytic anemia*); it takes place in patients with prosthetic heart valves, especially when paraprosthetic regurgitation is present. If the hemolysis consequent to mechanical trauma to the red cells is mild, and provided the supply of iron is adequate, it may be largely compensated. If more than mild anemia develops, reintervention to correct regurgitation may be required.

TOXIC AGENTS AND DRUGS

A number of chemicals with oxidative potential, whether medicinal or not, can cause hemolysis even in people who are not G6PD- deficient (see above). Examples are hyperbaric oxygen (or 100% oxygen), nitrates, chlorates, methylene blue, dapsone, cisplatin, and numerous aromatic (cyclic) compounds. Other chemicals may be hemolytic through a nonoxidative, largely unknown mechanism; examples are arsine, stibine, copper, and lead. The HA caused by lead poisoning is characterized by basophilic stippling. It is in fact a phenocopy of that seen in P5N deficiency (see above), suggesting it is mediated at least in part

by lead inhibiting this enzyme.

In these cases, hemolysis appears to be mediated by a direct chemical action on red cells. But drugs can cause hemolysis through at least two other mechanisms. (1) A drug can behave as a hapten and induce antibody production. In rare subjects this happens, for instance, with penicillin. Upon a subsequent exposure, red cells are caught, as innocent bystanders, in the reaction between penicillin and antipenicillin antibodies. Hemolysis will subside as soon as penicillin administration is stopped. (2) A drug can trigger, perhaps through mimicry, the production of an antibody against a red cell antigen. The best known example is methyldopa, an antihypertensive agent no longer in use, which in a small fraction of patients stimulated the production of the Rhesus antibody anti-e. In patients who have this antigen the anti-e is a true autoantibody, which would then cause an autoimmune HA (see below). Usually this would gradually subside once methyldopa was discontinued.

Severe intravascular hemolysis can be caused by the venom of certain snakes (cobras and vipers); and HA can also follow spider bites.

INFECTION

By far, the most frequent infectious cause of HA, in endemic areas, is malaria (Chap. 210). In other parts of the world, the most frequent cause is probably Shiga toxin–producing *Escherichia coli* O157:H7, now recognized as the main etiologic agent of the hemolytic-uremic syndrome, more common in children than in adults (Chap. 149). Life-threatening intravascular hemolysis, due to a toxin with lecithinase activity, occurs with *Clostridium perfringens* sepsis, particularly following open wounds, septic abortion, or as a disastrous accident due to a contaminated blood unit. Occasionally, HA is seen, especially in children, with sepsis or endocarditis from a variety of organisms.

AUTOIMMUNE HEMOLYTIC ANEMIA (AIHA)

Except for countries where malaria is endemic, AIHA is the most common form of *acquired hemolytic anemia*. In fact, not quite appropriately, the two phrases are sometimes used as synonymous.

Pathophysiology

AIHA is caused by an autoantibody directed against a red cell antigen, i.e., a molecule present on the surface of red cells. The autoantibody binds to the red cells. Once a red cell is coated by antibody, it will be destroyed by one or more mechanisms. In most cases the Fc portion of the antibody will be recognized by the Fc receptor of macrophages, and this will trigger erythrophagocytosis (Fig. 106-7). Thus, destruction of red cells will take place wherever macrophages are abundant, i.e., in the spleen, liver, and bone marrow. Because of the special anatomy of the spleen, it is particularly efficient in trapping antibody-coated red cells, and often this is the predominant site of red cell destruction. Although in severe cases even circulating monocytes can take part in this process, most of the phagocytosis-mediated red cell destruction takes place in the organs just mentioned, and it is therefore called *extravascular hemolysis*. In some cases, the nature of the antibody (usually an IgM antibody) is such that the antigen-antibody complex on the surface of red cells is able to activate complement (C). As a result, a large amount of membrane attack complex will form, and the red cells may be destroyed directly; this is known as *intravascular hemolysis*.

Figure 106-7

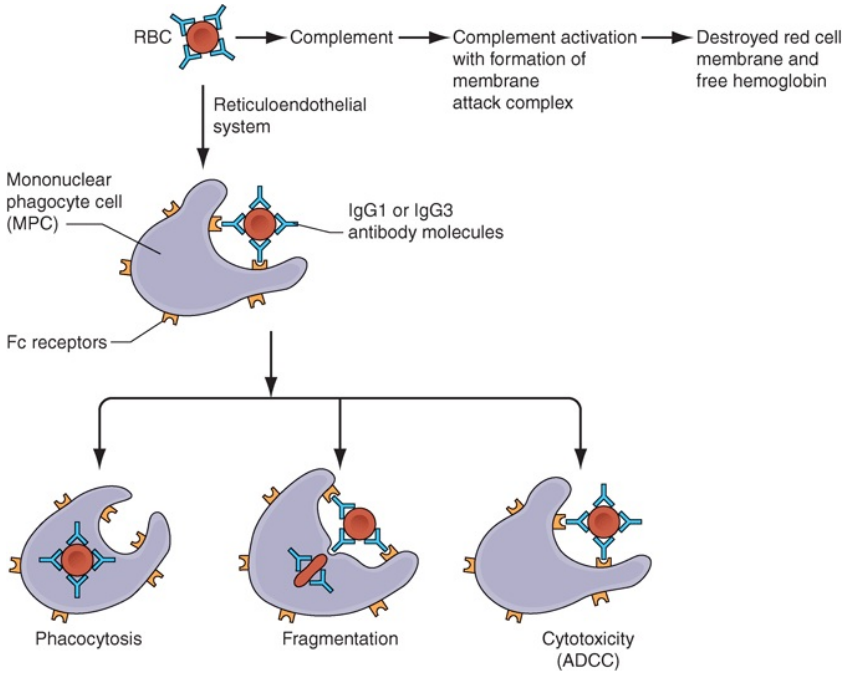

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanism of antibody-mediated immune destruction of red cells. *(From N Young et al: Clinical Hematology. Philadelphia, Elsevier, 2006; with permission.)*

Clinical Features

The onset of AIHA is very often abrupt and can be dramatic. The hemoglobin level can drop, within days, to as low as 4 g/dL; the massive red cell removal will produce jaundice; and sometimes the spleen is enlarged. When this triad is present, the suspicion of AIHA must be high. When hemolysis is (in part) intravascular, the telltale sign will be hemoglobinuria, which the patient may report or for which the physician must inquire/test. The diagnostic test for AIHA is the antiglobulin test worked out in 1945 by R. R. A. Coombs and known since by his name. The beauty of this test is that it directly detects the pathogenetic mediator of the disease, i.e., the presence of antibody on the red cells themselves. When the test is positive, it clinches the diagnosis, and when it is negative, the diagnosis is unlikely. However, the sensitivity of the Coombs test varies depending on the technology that is used, and in doubtful cases a repeat in a specialized lab is advisable; the term "Coombs-negative AIHA" is a

last resort. In some cases, the autoantibody has a defined identity: it may be specific for an antigen belonging to the Rhesus system (it is often anti-e). In many cases it is regarded as "unspecific" because it reacts with virtually all types of red cells.

As in autoimmune diseases in general, the real cause of AIHA remains obscure. However, from the clinical point of view, an important feature is that AIHA can appear to be isolated, or it can develop as part of a more general autoimmune disease, particularly systemic lupus erythematosus, of which sometimes it may be the first manifestation. Therefore, when AIHA is diagnosed, a full screen for autoimmune disease is imperative. In some cases, AIHA can be associated, on first presentation or subsequently, with autoimmune thrombocytopenia (Evans's syndrome).

Treatment: Autoimmune Hemolytic Anemia

Severe acute AIHA can be a medical emergency. The immediate treatment almost invariably includes transfusion of red cells. This may pose a special problem because if the antibody involved is unspecific, all the blood units cross-matched will be incompatible. In these cases it is often correct, paradoxically, to transfuse incompatible blood, the rationale being that the transfused red cells will be destroyed no less but no more than the patient's own red cells, but in the meantime the patient stays alive. Clearly, this rather unique situation requires good liaison and understanding between the clinical unit treating the patient and the blood transfusion/serology lab. Apart from emergency blood transfusion, the first-line treatment of AIHA is by using corticosteroids. In at least one-half of the cases, prednisone (1 mg/kg per day) will produce a remission promptly. Whereas some patients are then apparently cured, relapses are not uncommon. Although unfortunately most of the management of AIHA is not evidence-based, for patients who do not respond and for those who have relapsed (or who require more than 15 mg/d of prednisone to prevent relapse), it is highly recommended to consider a second-line treatment option, which might be either splenectomy or rituximab (anti-CD20). Splenectomy, although it does not cure the disease, can produce significant benefit by removing a major site of hemolysis, thus improving the anemia and/or reducing the need for other therapies (e.g., the dose of prednisone). Rituximab has emerged as a significant alternative to splenectomy because it can produce remissions in up to 80% of patients and it can be used repeatedly, even though progressive multifocal leukoencephalopathy is a dreaded if rare side effect. Azathioprine, cyclophosphamide, cyclosporine, and IV immunoglobulin have become third-line agents since the introduction of rituximab. In severe refractory cases, either auto- or allohematopoietic stem cell transplantation has been used, sometimes successfully.

Paroxysmal Cold Hemoglobinuria (PCH)

PCH is a rather rare form of AIHA occurring mostly in children, usually triggered by a viral infection, usually self-limited, and characterized by the involvement of the so-called Donath-Landsteiner antibody. In vitro this antibody has unique serologic features: it has anti-P specificity and it binds to red cells only at a low temperature (optimally at 4°C), but when the temperature is shifted to 37°C, lysis of red cells takes place in the presence of complement. Consequently, in vivo there is intravascular hemolysis, resulting in hemoglobinuria. Clinically, the differential diagnosis must include other causes of hemoglobinuria (Table 106–6), but the presence of the Donath-Landsteiner antibody will prove PCH. Active supportive treatment, including blood transfusion, is needed to control the anemia; subsequently, recovery is the rule.

Table 106–6 Diseases/Clinical Situations with Predominantly Intravascular Hemolysis

	Onset/Time Course	Main Mechanism	Appropriate Diagnostic Procedure	Comments
Mismatched blood transfusion	Abrupt	Nearly always ABO incompatibility	Repeat cross-match	
Paroxysmal nocturnal hemoglobinuria (PNH)	Chronic with acute exacerbations	Complement (C)-mediated destruction of CD59(–) red cells	Flow cytometry to display a CD59(–) red cell population	Exacerbations due to C activation through any pathway
Paroxysmal cold hemoglobinuria (PCH)	Acute	Immune lysis of normal red cells	Test for Donath-Landsteiner antibody	Often triggered by viral infection
Septicemia	Very acute	Exotoxins produced by *Clostridium perfringens*	Blood cultures	Other organisms may be responsible
Microangiopathic	Acute or chronic	Red cell fragmentation	Red cell morphology on blood smear	Different causes ranging from endothelial damage to hemangioma to leaky prosthetic heart valve
March hemoglobinuria	Abrupt	Mechanical destruction	Targeted history taking	
Favism	Acute	Destruction of older fraction of G6PD-deficient red cells	G6PD assay	Triggered by ingestion of large dish of fava beans; but trigger can be infection or drug instead

Cold Agglutinin Disease (CAD)

This designation is used for a form of chronic AIHA that usually affects the elderly and has special clinical and pathologic features. First, the term *cold* refers to the fact that the autoantibody involved reacts with red cells poorly or not at all at 37°C, whereas it reacts strongly at lower temperatures.[1] As a result, hemolysis is more prominent the more the body is exposed to the cold. The antibody is usually IgM with an anti-I specificity (the I antigen is present on the red cells of almost everybody), and it may have a very high titer (1:100,000 or more has been observed). Second, the antibody is produced by an expanded clone of B lymphocytes, and sometimes its concentration in the plasma is high enough to show up as a spike in plasma protein electrophoresis; i.e., as a monoclonal gammopathy. Third, since the antibody is IgM, CAD is related to Waldenström macroglobulinemia (WM) (Chap. 111), although in most cases the other clinical features of this disease are not present. Thus, CAD must be regarded as a form of WM, i.e., as a low-grade mature B cell lymphoma that manifests at an earlier stage precisely because the unique biologic properties of the IgM that it produces give the clinical picture of chronic HA.

[1] In the past this type of antibody was called a cold antibody, whereas the antibodies causing the more common form of AIHA were called warm antibodies.

In mild forms of CAD, avoidance of exposure to cold may be all that is needed to enable the patient to have a reasonably comfortable quality of life, but in more severe forms the management of CAD is not easy. Blood transfusion is not very effective because donor red cells are I-positive and will be rapidly removed. Immunosuppressive/cytotoxic treatment with azathioprine or cyclophosphamide can reduce the antibody titer, but clinical efficacy is limited and, in view of the chronic nature of the disease, the side effects may prove, in the long run, unacceptable. Unlike in AIHA, prednisone and splenectomy are ineffective. Plasma exchange is in theory a rational approach, but it is laborious and must be carried out at frequent intervals if it is to be beneficial.

Since the advent of rituximab, the picture has changed significantly for those 60% of patients with CAD who respond to this agent. Given the long clinical course of CAD, it remains to be seen with what periodicity rituximab will need to be administered.

Paroxysmal Nocturnal Hemoglobinuria (PNH)

PNH is an acquired chronic HA characterized by persistent intravascular hemolysis (Table 106–6) subject to recurrent exacerbations. In addition to hemolysis, there is often pancytopenia and a distinct tendency to venous thrombosis. This triad makes PNH a truly unique clinical condition. However, when not all of these three features are manifest on presentation, the diagnosis is often delayed, although it can be always made by appropriate laboratory investigations (see below).

PNH has about the same frequency in men and in women, and it is encountered in all populations throughout the world, but it is a rare disease. Its prevalence is estimated to be between 1 and 5 per million (it may be somewhat less rare in Southeast Asia and in the Far East). There is no evidence of inherited susceptibility. PNH has never been reported as a congenital disease, but it can present in small children or as late as in the seventies, although most patients are young adults.

Clinical Features

The patient may seek medical attention because, one morning, she or he has "passed blood instead of urine" (Fig. 106-8). This distressing or frightening event may be regarded as the classical presentation; however, more frequently, this symptom is not noticed or is suppressed. Indeed, the patient often presents simply as a problem in the differential diagnosis of *anemia*, whether symptomatic or discovered incidentally. Sometimes, the anemia is associated from the outset with neutropenia, or thrombocytopenia, or both, thus signaling an element of bone marrow failure (see below). Some patients may present with recurrent attacks of severe abdominal pain, defying a specific diagnosis and eventually found to be related to thrombosis. When thrombosis affects the hepatic veins, it may produce acute hepatomegaly and ascites, i.e., a full-fledged Budd-Chiari syndrome, which in the absence of liver disease ought to raise the suspicion of PNH.

Figure 106-8

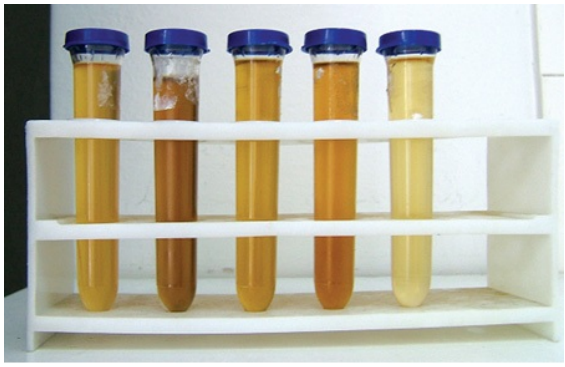

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Consecutive urine samples from a patient with paroxysmal nocturnal hemoglobinuria (PNH). The variation in the severity of hemoglobinuria within hours is probably unique to this condition.

The *natural history* of PNH can extend over decades. Without treatment, the median survival is estimated to be about 8–10 years. In the past, the most common cause of death has been venous thrombosis, followed by infection secondary to severe neutropenia and hemorrhage secondary to severe thrombocytopenia. PNH may evolve into aplastic anemia (AA), and PNH may manifest itself in patients who previously had AA. Rarely (estimated 1–2% of all cases), PNH may terminate in acute myeloid leukemia. On the other hand, full spontaneous recovery from PNH has been well documented, albeit rarely.

Laboratory Investigations and Diagnosis

The most consistent blood finding is anemia, which may range from mild to moderate to very severe. The anemia is usually normo-macrocytic, with unremarkable red cell morphology; if the MCV is high, it is usually largely accounted for by reticulocytosis, which may be quite marked (up to 20%, or up to 400,000/μL). The anemia may become microcytic if the patient is allowed to become iron-deficient as a result of chronic urinary blood loss through hemoglobinuria. Unconjugated bilirubin is mildly or moderately elevated, LDH is typically markedly elevated (values in the thousands are common), and haptoglobin is usually undetectable. All these findings make the diagnosis of HA compelling. Hemoglobinuria, the telltale sign of intravascular hemolysis (Table 106–6), may be overt in a random urine sample. If it is not, it may be helpful to obtain serial urine samples, since hemoglobinuria can vary dramatically from day to day, and even from hour to hour (Fig. 106-8). The bone marrow is usually cellular, with marked to massive erythroid hyperplasia, often with mild to moderate dyserythropoietic features (these do not justify confusing PNH with myelodysplastic syndrome). At some stage of the disease, the marrow may become hypocellular or even frankly aplastic (see below).

The definitive diagnosis of PNH must be based on the demonstration that a substantial proportion of the patient's red cells have an increased susceptibility to complement (C), due to the deficiency on their surface of proteins (particularly CD59 and CD55) that normally protect the red cells from activated C. The sucrose hemolysis test is unreliable, and the acidified serum (Ham) test is carried out in few labs. The gold standard today is flow cytometry, which can be carried out on granulocytes as well as on red cells. A bimodal distribution of cells, with a discrete population that is CD59-, CD55-, is diagnostic of PNH. Usually this population is at least 5% of the total in the case of red cells and at least 20% of the total in the case of granulocytes.

Pathophysiology

Hemolysis in PNH is due to an intrinsic abnormality of the red cell, which makes it exquisitely sensitive to activated C, whether it is activated through the alternative pathway or through an antigen-antibody reaction (Fig. 106-9). The former mechanism is mainly responsible for intravascular hemolysis in PNH. The latter mechanism explains why the hemolysis can be dramatically exacerbated in the course of a viral or bacterial infection. Hypersusceptibility to C is due to deficiency of several protective membrane proteins, of which CD59 is the most important because it hinders the insertion into the membrane of C9 polymers. The molecular basis for the deficiency of these proteins has been pinpointed not to a defect in any of the respective genes, but rather to the shortage of a unique glycolipid molecule, glycosylphosphatidyl-inositol (GPI), which, through a peptide bond, anchors these proteins to the surface membrane of cells. The shortage of GPI is due in turn to a mutation in an X-linked gene, called PIG-A, required for an early step in GPI biosynthesis. In virtually each patient, the PIG-A mutation is different. This is not surprising, since these mutations are not inherited: rather, each one takes place de novo in a hematopoietic stem cell (i.e., they are somatic mutations). As a result, the patient's marrow is a mosaic of mutant and nonmutant cells, and the peripheral blood

always contains both PNH cells and normal (non-PNH) cells. Thrombosis is one of the most immediately life-threatening complications of PNH and yet one of the least understood in its pathogenesis. It could be that deficiency of CD59 on the PNH platelet causes inappropriate platelet activation; however, other mechanisms are possible.

Figure 106-9

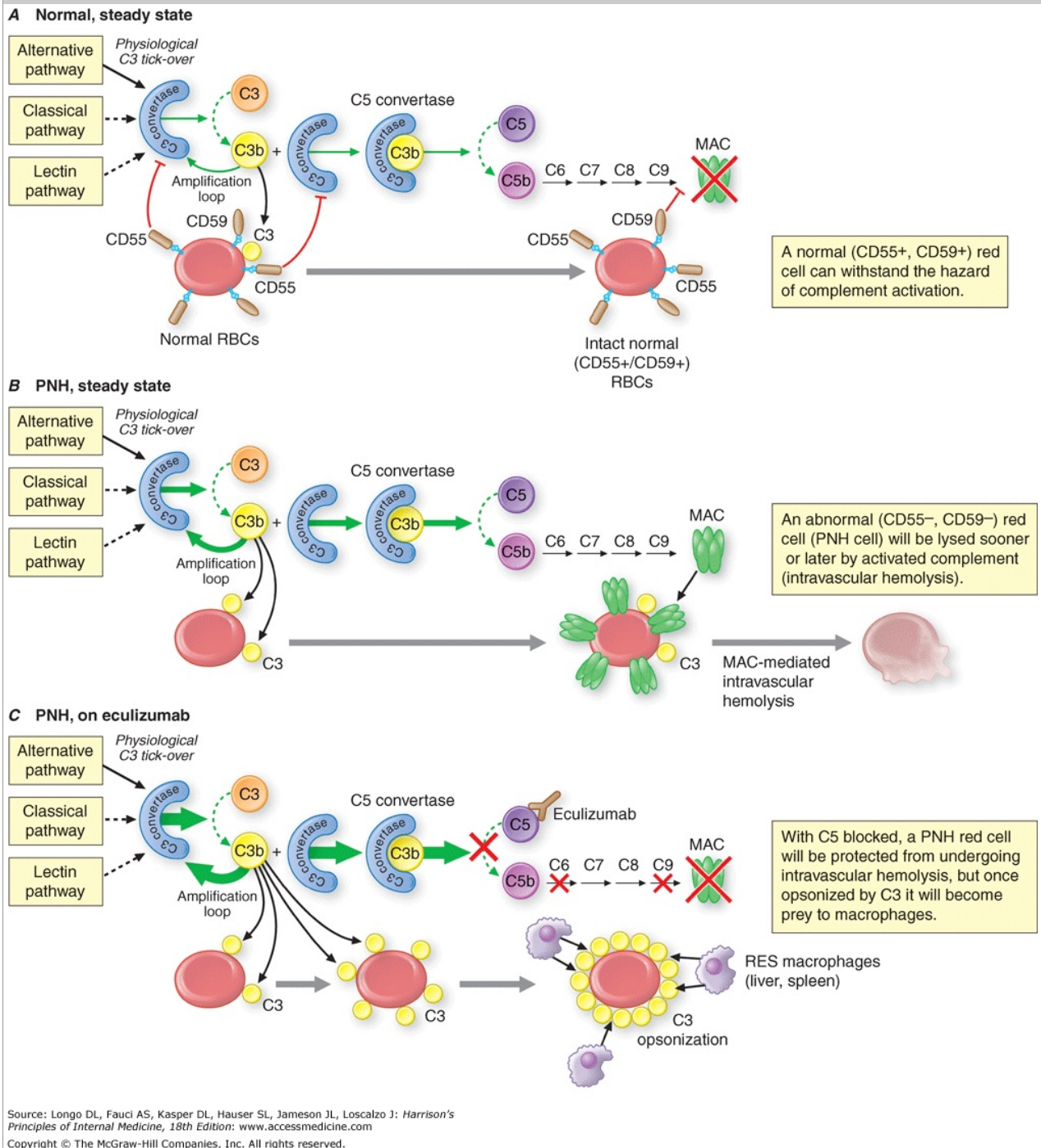

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The complement cascade and the fate of red cells.*A.* Normal red cells are protected from complement activation and subsequent hemolysis by CD55 and CD59. These two proteins, being GPI-linked, are missing from the surface of PNH red cells as a result of a somatic mutation of the X-linked *PIG-A* gene that encodes a protein required for an early step of the GPI molecule biosynthesis. ***B.*** In the steady state, PNH erythrocytes suffer from spontaneous (tick-over) complement activation, with consequent intravascular hemolysis through formation of the membrane -attack complex (MAC); when extra complement is activated through the classical pathway, an exacerbation of hemolysis will result. ***C.*** On eculizumab, PNH erythrocytes are protected from hemolysis from the inhibition of C5 cleavage; however, upstream complement activation may lead to C3 opsonization and possible extravascular hemolysis. GPI, glycosylphosphatidylinositol; PNH, paroxysmal nocturnal hemoglobinuria. *(From L Luzzatto, et al: Haematologica 95:523, 2010.)*

Bone Marrow Failure (BMF) and Relationship between PNH and Aplastic Anemia (AA)

It is not unusual that patients with firmly established PNH have a previous history of well-documented AA. On the other hand, sometimes a patient with PNH becomes less hemolytic and more pancytopenic and ultimately has the clinical picture of AA. Since AA is probably an organ-specific autoimmune disease, in which T cells cause damage to hematopoietic stem cells, the same may be true of PNH, with the specific proviso that the damage spares PNH stem cells. Skewing of the T cell repertoire in patients with PNH lends some support to this notion. In addition, there is evidence in mouse models that PNH stem cells do not expand when the rest of the bone marrow is normal, and by using high-sensitivity flow cytometry technology, very rare PNH cells harboring PIG-A mutations can be demonstrated in normal people. In view of these facts, it seems that an element of BMF in PNH is the rule rather than the exception. An extreme view is that PNH is a form of AA in which BMF is masked by the

massive expansion of the PNH clone that populates the patient's bone marrow. The mechanism whereby PNH stem cells escape the damage suffered by non-PNH stem cells is not yet known.

Treatment: Paroxysmal Nocturnal Hemoglobinuria

Unlike other acquired HAs, PNH may be a lifelong condition; standard care was formerly supportive treatment only, including transfusion of filtered red cells[2] whenever necessary, which, for some patients, means quite frequently. Folic acid supplements (at least 3 mg/d) are mandatory, and the serum iron should be checked periodically and iron supplements administered as appropriate. Long-term glucocorticoids are not indicated because there is no evidence that they have any effect on chronic hemolysis: in fact they are contraindicated because of their many dangerous side effects. A major advance in the management of PNH has been the development of a humanized monoclonal antibody, eculizumab, directed against the complement component C5. In an international, multicenter, placebo-controlled randomized trial of 87 patients (so far the only controlled therapeutic trial in PNH) who had been selected on grounds of having severe hemolysis making them transfusion- dependent, eculizumab proved effective and was licensed in 2007 (Fig. 106-10). By blocking the complement cascade downstream of C5, eculizumab abrogates complement-dependent intravascular hemolysis in all PNH patients, which in itself significantly improves their quality of life. One would expect that, as a result, the need for blood transfusion would be also abrogated, and this indeed is the case in about one-half of the patients, in many of whom there is also a rise in hemoglobin levels. In the remaining patients, the anemia remains sufficiently severe to require blood transfusion, apparently because of ongoing extravascular hemolysis of red cells opsonized by complement (C3) fragments. Based on its half-life, eculizumab must be administered intravenously every 14 days. The only form of treatment that currently can provide a definitive cure for PNH is allogeneic bone marrow transplantation (BMT). When an HLA-identical sibling is available, BMT should be offered to any young patient with severe PNH; the availability of eculizumab has probably decreased significantly the proportion of those who take up this option.

Figure 106-10

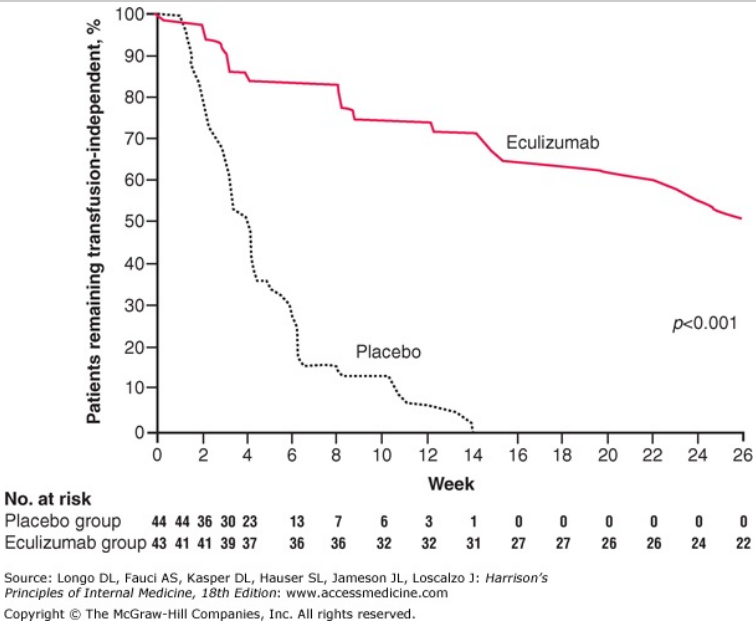

Therapeutic efficacy of an anti-C5 antibody on the anemia of paroxysmal nocturnal hemoglobinuria. *(From P Hillmen et al: N Engl J Med 355:1233, 2006; with permission.)*

For patients with the PNH-AA syndrome, immunosuppressive treatment with antilymphocyte globulin (ALG or ATG) and cyclosporine A may be indicated. Although no formal trial has ever been conducted, this approach has helped particularly to relieve severe thrombocytopenia and/or neutropenia in patients in whom these were the main problem(s). By contrast, there is often little immediate effect on hemolysis. Any patient who has had venous thrombosis or who has a genetically determined thrombophilic state in addition to PNH should be on regular anticoagulant prophylaxis.

[2]Now that filters with excellent retention of white cells are routinely used, the traditional washing of red cells, aiming to avoid white cell reactions triggering hemolysis, is no longer necessary and is wasteful.

ANEMIA DUE TO ACUTE BLOOD LOSS

Blood loss causes anemia by two main mechanisms. First, by the direct loss of red cells, and second, because if the loss of blood is protracted, it will gradually deplete the iron stores, eventually resulting in iron deficiency. The latter type of anemia is covered in Chap. 103. Here we are concerned with the former type, i.e., the *posthemorrhagic anemia*, which follows *acute* blood loss. This can be *external* (as after trauma, or obstetric hemorrhage) or *internal* (e.g., from bleeding in the gastrointestinal tract, rupture of the spleen, rupture of an ectopic pregnancy, subarachnoid hemorrhage). In any of these cases, i.e., after the sudden loss of a large amount of blood, there are three clinical/pathophysiologic stages. First, the dominant feature is hypovolemia, which poses a threat particularly to organs that normally have a high blood supply, like the brain and the kidneys; therefore, loss of consciousness and acute renal failure are major threats. It is important to note that at this stage an ordinary blood count will not show anemia, as the hemoglobin concentration is not affected. Second, as an emergency response, baroreceptors and stretch receptors will cause release of vasopressin and other peptides, and the body will shift fluid from the extravascular to the intravascular compartment, producing hemodilution; thus, the hypovolemia gradually converts to anemia. The degree of anemia will reflect the amount of blood lost: if after 3 days the hemoglobin is, say 7 g/dL, it means that about half of the entire blood had been lost. Third, provided bleeding does not continue, the bone marrow response will gradually ameliorate the anemia.

The diagnosis of acute posthemorrhagic anemia (APHA) is usually straightforward; although sometimes internal bleeding episodes—after a traumatic injury or otherwise—may not be immediately obvious, even when large. Whenever an abrupt fall in hemoglobin has taken place, whatever history is given by the patient, APHA should be suspected: supplementary history may have to be obtained by asking the appropriate questions, and appropriate investigations (e.g., a sonogram or an endoscopy) may have to be carried out.

Treatment: Anemia Due to Acute Blood Loss

With respect to treatment, a two-pronged approach is imperative. (1) In many cases the blood lost needs to be replaced promptly. Unlike with many chronic anemias, when finding and correcting the cause of the anemia is the first priority and blood transfusion may not be even necessary because the body is adapted to the anemia, with acute blood loss the reverse is true; i.e., since the body is not adapted to the anemia, blood transfusion takes priority. (2) While the emergency is being confronted, it is imperative to stop the hemorrhage and to eliminate its source.

A special type of APHA is blood loss during and immediately after surgery, which can be substantial (for instance, up to 2 L in the case of a radical prostatectomy). Of course, with elective surgical procedures, the patient's own stored blood may be available (through preoperative autologous blood donation), and in any case blood loss is carefully monitored. Since this blood loss is iatrogenic, ever more effort should be invested in optimizing transfusion management.

A Holy Grail of emergency medicine has been for a long time the idea of a blood substitute that would be universally available, suitable for all recipients, easy to store and to transport, safe, and as effective as blood itself. Two main paths have been pursued: (1) fluorocarbon synthetic chemicals that bind oxygen reversibly, and (2) artificially modified hemoglobins, known as hemoglobin-based oxygen carriers (HBOC). Although there are numerous anecdotal reports of the use of both approaches in humans, and although HBOCs have reached the stage of phase II–III clinical trials, no "blood substitute" has yet become standard treatment.

FURTHER READINGS

Chen JY et al: A review of blood substitutes: Examining the history, clinical trial results, and ethics of hemoglobin-based oxygen carriers. Clinics (Sao Paulo) 64:803, 2009[PMID: 19690667]

Dacie J: *The Haemolytic Anaemias.* London, Churchill Livingstone, 1985–1995

Grace RF, Lux SE: Disorders of the red cell membrane, in *Nathan & Oski's Hematology of Infancy and Childhood* , 7th ed, S Orkin et al (eds). Philadelphia, Saunders, 2009, pp. 659–837

Heier HE et al: Transfusion vs. alternative treatment modalities in acute bleeding: A systematic review. Acta Anaesthesiol Scand 509:20, 2006

Hillmen P et al: The complement inhibitor eculizumab in paroxysmal nocturnal hemoglobinuria. N Engl J Med 355:1233, 2006[PMID: 16990386]

Kaushansky K et al (eds): *Williams Hematology*, 8th ed. New York, McGraw Hill, 2010

Lechner K, Jager U: How I treat autoimmune hemolytic anemias in adults. Blood 116:1831, 2010[PMID: 20548093]

Luzzatto L: The rise and fall of the antimalarial Lapdap: A lesson in pharmacogenetics. Lancet 376:739, 2010[PMID: 20599264]

Luzzatto L, Araten D: Paroxysmal nocturnal hemoglobinuria, in *Clinical Hematology*, NS Young et al (eds). Philadelphia, Mosby, 2006, pp. 326–339

Noris M, Remuzzi G: Atypical hemolytic-uremic syndrome. N Engl J Med 361:1676, 2009[PMID: 19846853]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APLASTIC ANEMIA, MYELODYSPLASIA, AND RELATED BONE MARROW FAILURE SYNDROMES: INTRODUCTION

The hypoproliferative anemias are normochromic, normocytic, or macrocytic and are characterized by a low reticulocyte count. Deficient production of red blood cells (RBCs) occurs with marrow damage and dysfunction, which may be secondary to infection, inflammation, and cancer. Hypoproliferative anemia is also a prominent feature of hematologic diseases that are described as bone marrow failure states; these include aplastic anemia, myelodysplastic syndrome (MDS), pure red cell aplasia (PRCA), and myelophthisis. Anemia in these disorders is often not a solitary or even the major hematologic finding. More frequent in bone marrow failure is pancytopenia: anemia, leukopenia, and thrombocytopenia. Low blood counts in the marrow failure diseases result from deficient hematopoiesis, as distinguished from blood count depression due to peripheral destruction of red cells (hemolytic anemias), platelets [idiopathic thrombocytopenic purpura (ITP) or due to splenomegaly], and granulocytes (as in the immune leukopenias).

Hematopoietic failure syndromes are classified by dominant morphologic features of the bone marrow (Table 107–1). Although practical distinction among these syndromes usually is clear, they can occur secondary to other diseases, and some processes are so closely related that the diagnosis may be complex. Patients may seem to suffer from two or three related diseases simultaneously, or one diagnosis may appear to evolve into another. Many of these syndromes share an immune-mediated mechanism of marrow destruction and some element of genomic instability resulting in a higher rate of malignant transformation.

Table 107–1 Differential Diagnosis of Pancytopenia

Pancytopenia with Hypocellular Bone Marrow	
Acquired aplastic anemia	
Constitutional aplastic anemia (Fanconi's anemia, dyskeratosis congenita)	
Some myelodysplasia	
Rare aleukemic leukemia	
Some acute lymphoid leukemia	
Some lymphomas of bone marrow	
Pancytopenia with Cellular Bone Marrow	
Primary bone marrow diseases	Secondary to systemic diseases
Myelodysplasia	Systemic lupus erythematosus
Paroxysmal nocturnal hemoglobinuria	Hypersplenism
Myelofibrosis	B_{12}, folate deficiency
Some aleukemic leukemia	Overwhelming infection
Myelophthisis	Alcohol
Bone marrow lymphoma	Brucellosis
Hairy cell leukemia	Sarcoidosis
	Tuberculosis
	Leishmaniasis
Hypocellular Bone Marrow ± Cytopenia	
Q fever	
Legionnaires' disease	
Anorexia nervosa, starvation	
Mycobacterium	

It is important that the internist and general practitioner recognize the marrow failure syndromes, as their prognosis may be poor if the patient is untreated; effective therapies are often available but sufficiently complex in their choice and delivery so as to warrant the care of a hematologist or oncologist.

APLASTIC ANEMIA

Definition

Aplastic anemia is pancytopenia with bone marrow hypocellularity. Acquired aplastic anemia is distinguished from iatrogenic marrow aplasia, marrow hypocellularity after intensive cytotoxic chemotherapy for cancer. Aplastic anemia can also be constitutional: The genetic diseases Fanconi's anemia and dyskeratosis congenita, although frequently associated with typical physical anomalies and the development of pancytopenia early in life, can also present as marrow failure in normal-appearing adults. Acquired aplastic anemia is often stereotypical in its manifestations, with the abrupt onset of low blood counts in a previously well young adult; seronegative hepatitis or a course of an incriminated medical drug may precede the onset. The diagnosis in these instances is uncomplicated. Sometimes blood count depression is moderate or incomplete, resulting in anemia, leukopenia, and thrombocytopenia in some combination. Aplastic anemia is related to both paroxysmal nocturnal hemoglobinuria (PNH; Chap. 106) and to MDS, and in some cases a clear distinction among these disorders may not be possible.

Epidemiology

The incidence of acquired aplastic anemia in Europe and Israel is two cases per million persons annually. In Thailand and China, rates of five to seven per million have been established. In general, men and women are affected with equal frequency, but the age distribution is biphasic, with the major peak in the teens and twenties and a second rise in older adults.

Etiology

The origins of aplastic anemia have been inferred from several recurring clinical associations (Table 107–2); unfortunately, these relationships are not reliable in an individual patient and may not be etiologic. In addition, although most cases of aplastic anemia are idiopathic, little other than history separates these cases from those with a presumed etiology such as a drug exposure.

Table 107–2 Classification of Aplastic Anemia and Single Cytopenias

Acquired	Inherited
Aplastic Anemia	
Secondary	Fanconi's anemia
Radiation	Dyskeratosis congenita
Drugs and chemicals	Shwachman-Diamond syndrome
Regular effects	Reticular dysgenesis
Idiosyncratic reactions	Amegakaryocytic thrombocytopenia
Viruses	Familial aplastic anemias
Epstein-Barr virus (infectious mononucleosis)	Preleukemia (monosomy 7, etc.)
Hepatitis (non-A, non-B, non-C hepatitis)	Nonhematologic syndrome (Down, Dubowitz, Seckel)
Parvovirus B19 (transient aplastic crisis, PRCA)	
HIV-1 (AIDS)	
Immune diseases	
Eosinophilic fasciitis	
Hyperimmunoglobulinemia	
Thymoma/thymic carcinoma	
Graft-versus-host disease in immunodeficiency	
Paroxysmal nocturnal hemoglobinuria	
Pregnancy	
Idiopathic	
Cytopenias	
PRCA (see Table 107–4)	Congenital PRCA (Diamond-Blackfan anemia)
Neutropenia/agranulocytosis	

Idiopathic	Kostmann's syndrome
Drugs, toxins	Shwachman-Diamond syndrome
Pure white cell aplasia	Reticular dysgenesis
Thrombocytopenia	
Drugs, toxins	Amegakaryocytic thrombocytopenia
Idiopathic amegakaryocytic	Thrombocytopenia with absent radii

Abbreviation: PRCA, pure red cell aplasia.

RADIATION

Marrow aplasia is a major acute sequela of radiation. Radiation damages DNA; tissues dependent on active mitosis are particularly susceptible. Nuclear accidents can involve not only power plant workers but also employees of hospitals, laboratories, and industry (food sterilization, metal radiography, etc.), as well as innocents exposed to stolen, misplaced, or misused sources. Whereas the radiation dose can be approximated from the rate and degree of decline in blood counts, dosimetry by reconstruction of the exposure can help to estimate the patient's prognosis and also to protect medical personnel from contact with radioactive tissue and excreta. MDS and leukemia, but probably not aplastic anemia, are late effects of radiation.

CHEMICALS

Benzene is a notorious cause of bone marrow failure: epidemiologic, clinical, and laboratory data link benzene to aplastic anemia, acute leukemia, and blood and marrow abnormalities. For leukemia, incidence is correlated with cumulative exposure, but susceptibility must also be important, as only a minority of even heavily exposed workers develop myelotoxicity. The employment history is important, especially in industries where benzene is used for a secondary purpose, usually as a solvent. Benzene-related blood diseases have declined with regulation of industrial exposure. Although benzene is no longer generally available as a household solvent, exposure to its metabolites occurs in the normal diet and in the environment. The association between marrow failure and other chemicals is much less well substantiated.

DRUGS

(Table 107–3) Many chemotherapeutic drugs have marrow suppression as a major toxicity; effects are dose dependent and will occur in all recipients. In contrast, idiosyncratic reactions to a large and diverse group of drugs may lead to aplastic anemia without a clear dose-response relationship. These associations rested largely on accumulated case reports until a large international study in Europe in the 1980s quantitated drug relationships, especially for nonsteroidal analgesics, sulfonamides, thyrostatic drugs, some psychotropics, penicillamine, allopurinol, and gold. Association does not equal causation: A drug may have been used to treat the first symptoms of bone marrow failure (antibiotics for fever or the preceding viral illness) or provoked the first symptom of a preexisting disease (petechiae by nonsteroidal anti-inflammatory agents administered to the thrombocytopenic patient). In the context of total drug use, idiosyncratic reactions, although individually devastating, are rare events. Risk estimates are usually lower when determined in population-based studies; furthermore, the low absolute risk is also made more obvious: even a ten- or twenty-fold increase in risk translates, in a rare disease, to but a handful of drug-induced aplastic anemia cases among hundreds of thousands of exposed persons.

Table 107–3 Some Drugs and Chemicals Associated with Aplastic Anemia

Agents that regularly produce marrow depression as major toxicity in commonly employed doses or normal exposures:
Cytotoxic drugs used in cancer chemotherapy: *alkylating agents*, *antimetabolites*, *antimitotics*, some antibiotics
Agents that frequently but not inevitably produce marrow aplasia:
Benzene
Agents associated with aplastic anemia but with a relatively low probability:
Chloramphenicol
Insecticides
Antiprotozoals: *quinacrine* and chloroquine, mepacrine
Nonsteroidal anti-inflammatory drugs (including *phenylbutazone*,
indomethacin, ibuprofen, sulindac, aspirin)
Anticonvulsants (*hydantoins*, *carbamazepine*, phenacemide, felbamate)

Heavy metals (*gold*, arsenic, bismuth, mercury)
Sulfonamides: some antibiotics, antithyroid drugs (methimazole, methylthiouracil, propylthiouracil), antidiabetes drugs (tolbutamide, chlorpropamide), carbonic anhydrase inhibitors (acetazolamide and methazolamide)
Antihistamines (*cimetidine*, chlorpheniramine)
d-Penicillamine
Estrogens (in pregnancy and in high doses in animals)
Agents whose association with aplastic anemia is more tenuous:
Other antibiotics (streptomycin, tetracycline, methicillin,
mebendazole, trimethoprim/sulfamethoxazole, flucytosine)
Sedatives and tranquilizers (chlorpromazine, prochlorperazine,
piperacetazine, chlordiazepoxide, meprobamate, methyprylon)
Allopurinol
Methyldopa
Quinidine
Lithium
Guanidine
Potassium perchlorate
Thiocyanate
Carbimazole

Note: Terms set in italic show the most consistent association with aplastic anemia.

INFECTIONS

Hepatitis is the most common preceding infection, and posthepatitis marrow failure accounts for approximately 5% of etiologies in most series. Patients are usually young men who have recovered from a bout of liver inflammation 1 to 2 months earlier; the subsequent pancytopenia is very severe. The hepatitis is seronegative (non-A, non-B, non-C) and possibly due to an as yet undiscovered infectious agent. Fulminant liver failure in childhood also follows seronegative hepatitis, and marrow failure occurs at a high rate in these patients. Aplastic anemia can rarely follow infectious mononucleosis. Parvovirus B19, the cause of transient aplastic crisis in hemolytic anemias and of some PRCAs (see below), does not usually cause generalized bone marrow failure. Mild blood count depression is frequent in the course of many viral and bacterial infections but resolves with the infection.

IMMUNOLOGIC DISEASES

Aplasia is a major consequence and the inevitable cause of death in *transfusion-associated graft-versus-host disease* (GVHD) that can occur after infusion of nonirradiated blood products to an immunodeficient recipient. Aplastic anemia is strongly associated with the rare collagen vascular syndrome *eosinophilic fasciitis* that is characterized by painful induration of subcutaneous tissues (Chap. 323). Pancytopenia with marrow hypoplasia can also occur in systemic lupus erythematosus (SLE).

PREGNANCY

Aplastic anemia very rarely may occur and recur during pregnancy and resolve with delivery or with spontaneous or induced abortion.

PAROXYSMAL NOCTURNAL HEMOGLOBINURIA

An acquired mutation in the *PIG-A* gene in a hematopoietic stem cell is required for the development of PNH, but *PIG-A* mutations probably occur commonly in normal individuals. If the PIG-A mutant stem cell proliferates, the result is a clone of progeny deficient in glycosylphosphatidylinositol-linked cell surface membrane proteins (Chap. 106). Small clones of deficient cells can be detected by sensitive flow cytometry tests in approximately one-half of patients with aplastic anemia at the time of presentation [and PNH cells are also seen in MDS (see below)]. Functional studies of bone marrow

from PNH patients, even those with mainly hemolytic manifestations, show evidence of defective hematopoiesis. Patients with an initial clinical diagnosis of PNH, especially younger individuals, may later develop frank marrow aplasia and pancytopenia; patients with an initial diagnosis of aplastic anemia may suffer from hemolytic PNH years after recovery of blood counts.

CONSTITUTIONAL DISORDERS

Fanconi's anemia, an autosomal recessive disorder, manifests as congenital developmental anomalies, progressive pancytopenia, and an increased risk of malignancy. Chromosomes in Fanconi's anemia are peculiarly susceptible to DNA cross-linking agents, the basis for a diagnostic assay. Patients with Fanconi's anemia typically have short stature, café au lait spots, and anomalies involving the thumb, radius, and genitourinary tract. At least 12 different genetic defects (all but one with an identified gene) have been defined; the most common, type A Fanconi's anemia, is due to a mutation in *FANCA*. Most of the Fanconi's anemia gene products form a protein complex that activates FANCD2 by monoubiquitination to play a role in the cellular response to DNA damage and especially interstrand cross-linking.

Dyskeratosis congenita is characterized by mucous membrane leukoplasia, dystrophic nails, reticular hyperpigmentation, and the development of aplastic anemia in childhood. Dyskeratosis is due to mutations in genes of the telomere repair complex, which acts to maintain telomere length in replicating cells: The X-linked variety is due to mutations in the *DKC1* (*dyskerin*) gene; the more unusual autosomal dominant type is due to mutation in *TERC*, which encodes an RNA template, and *TERT*, which encodes the catalytic reverse transcriptase, telomerase. Mutations in TNF2, a component of the shelterin, proteins that bind the telomere DNA, also occur in dyskeratosis.

In Shwachman-Diamond syndrome, marrow failure is seen with pancreatic insufficiency and malabsorption; most patients have compound heterozygous mutations in *SBDS* that may affect marrow stroma function.

Mutations in *TERT*, *TERC*, *TNF2*, and *SBDS* also can occur in patients with apparently acquired aplastic anemia (TERT and TERC mutations also are etiologic in familial pulmonary fibrosis and in some hepatic cirrhosis).

PATHOPHYSIOLOGY

Bone marrow failure results from severe damage to the hematopoietic cell compartment. In aplastic anemia, replacement of the bone marrow by fat is apparent in the morphology of the biopsy specimen (Fig. 107-1) and MRI of the spine. Cells bearing the CD34 antigen, a marker of early hematopoietic cells, are greatly diminished, and in functional studies, committed and primitive progenitor cells are virtually absent; in vitro assays have suggested that the stem cell pool is reduced to ≤1% of normal in severe disease at the time of presentation.

Figure 107-1

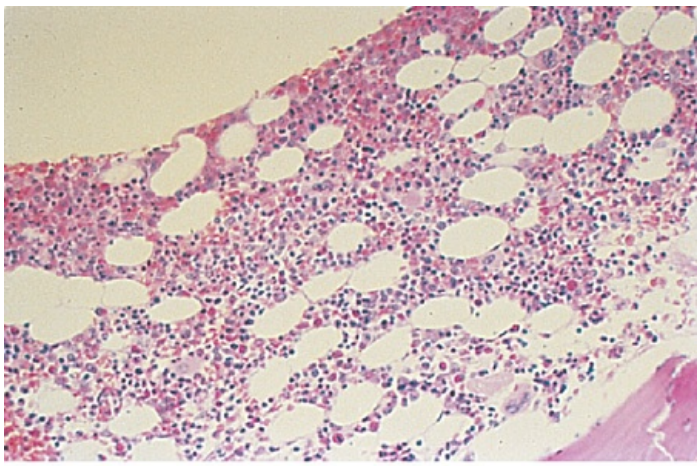

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

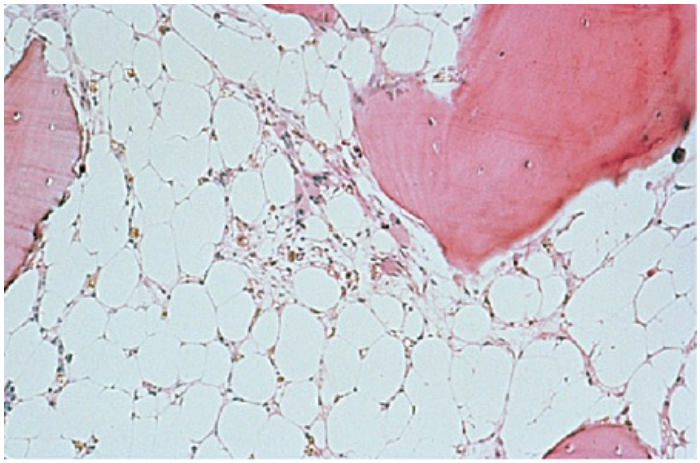

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

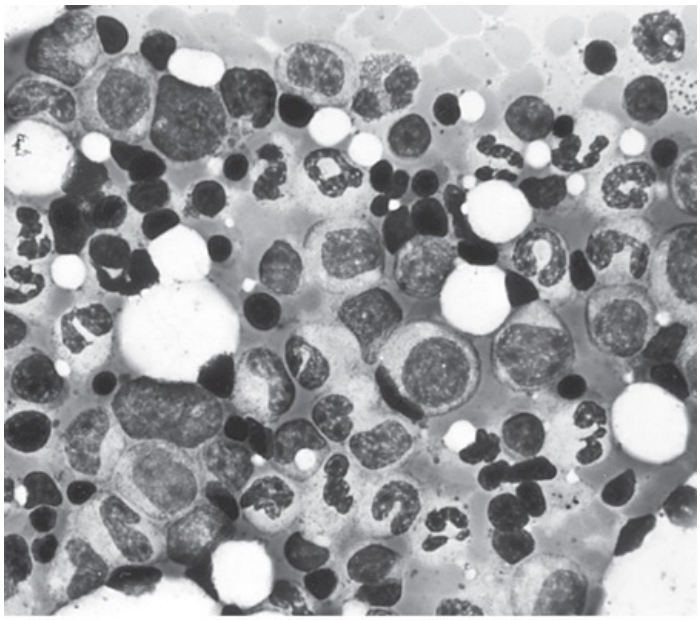

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

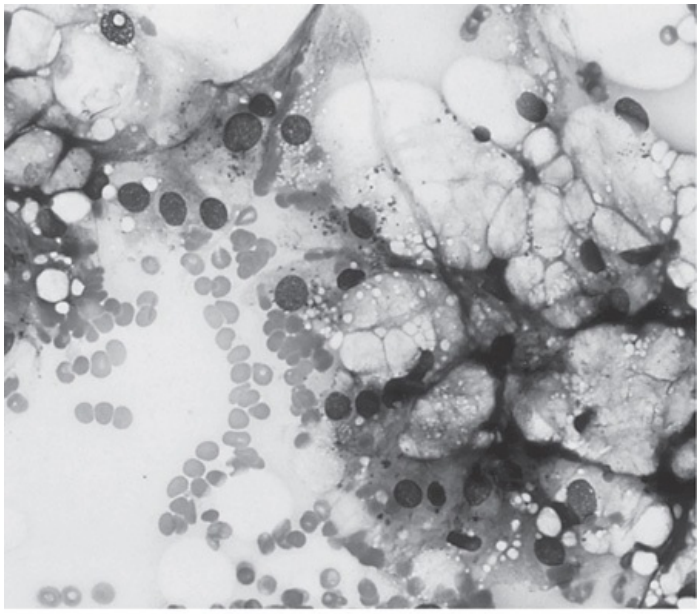

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Normal bone marrow biopsy. ***B.*** Normal bone marrow aspirate smear. The marrow is normally 30–70% cellular, and there is a heterogeneous mix of myeloid, erythroid, and lymphoid cells. ***C.*** Aplastic anemia biopsy. ***D.*** Marrow smear in aplastic anemia. The marrow shows replacement of hematopoietic tissue by fat and only residual stromal and lymphoid cells.

An intrinsic stem cell defect exists for the constitutional aplastic anemias: Cells from patients with Fanconi's anemia exhibit chromosome damage and death on exposure to certain chemical agents. Telomeres are short in some patients with aplastic anemia, due to heterozygous mutations in genes of the telomere repair complex. Variable penetrance means that *TERT* and *TERC* mutations represent risk factors for marrow failure, as family members with the same mutations may have normal or only slight hematologic abnormalities but more subtle evidence of (compensated) hematopoietic insufficiency.

Drug Injury

Extrinsic damage to the marrow follows massive physical or chemical insults such as high doses of radiation and toxic chemicals. For the more common idiosyncratic reaction to modest doses of medical drugs, altered drug metabolism has been invoked as a likely mechanism. The metabolic pathways of many drugs and chemicals, especially if they are polar and have limited water solubility, involve enzymatic degradation to highly reactive electrophilic compounds; these intermediates are toxic because of their propensity to bind to cellular macromolecules. For example, derivative hydroquinones and quinolones are responsible for benzene-induced tissue injury. Excessive generation of toxic intermediates or failure to detoxify the intermediates may be genetically determined and apparent only on specific drug challenge; the complexity and specificity of the pathways imply multiple susceptibility loci and would provide an explanation for the rarity of idiosyncratic drug reactions.

Immune-Mediated Injury

The recovery of marrow function in some patients prepared for bone marrow transplantation with antilymphocyte globulin (ALO) first suggested that aplastic anemia might be immune mediated. Consistent with this hypothesis was the frequent failure of simple bone marrow transplantation from a syngeneic twin, without conditioning cytotoxic chemotherapy, which also argued both *against* simple stem cell absence as the cause and *for* the presence of a host factor producing marrow failure. Laboratory data support an important role for the immune system in aplastic anemia. Blood and bone marrow cells of patients can suppress normal hematopoietic progenitor cell growth, and removal of T cells from aplastic anemia bone marrow improves colony formation in vitro. Increased numbers of activated cytotoxic T cell clones are observed in aplastic anemia patients and usually decline with successful immunosuppressive therapy; cytokine measurements show a T_H1 immune response [interferon γ (IFN γ) and tumor necrosis factor (TNF)]. Interferon and induce Fas expression on CD34 cells, leading to apoptotic cell death; localization of activated T cells to bone marrow and local production of their soluble factors are probably important in stem cell destruction.

Early immune system events in aplastic anemia are not well understood. An oligoclonal, T cell response implies an antigenic stimulus. Many different exogenous antigens appear capable of initiating a pathologic immune response, but at least some of the T cells may recognize true self-antigens. The

rarity of aplastic anemia despite common exposures (medicines, seronegative hepatitis) suggests that genetically determined features of the immune response can convert a normal physiologic response into a sustained abnormal autoimmune process, including polymorphisms in histocompatibility antigens, cytokinegenes, and genes that regulate T cell polarization and effector function.

Clinical Features

HISTORY

Aplastic anemia can appear with seeming abruptness or have a more insidious onset. Bleeding is the most common early symptom; a complaint of days to weeks of easy bruising, oozing from the gums, nose bleeds, heavy menstrual flow, and sometimes petechiae will have been noticed. With thrombocytopenia, massive hemorrhage is unusual, but small amounts of bleeding in the central nervous system can result in catastrophic intracranial or retinal hemorrhage. Symptoms of anemia are also frequent, including lassitude, weakness, shortness of breath, and a pounding sensation in the ears. Infection is an unusual first symptom in aplastic anemia (unlike in agranulocytosis, where pharyngitis, anorectal infection, or frank sepsis occur early). A striking feature of aplastic anemia is the restriction of symptoms to the hematologic system, and patients often feel and look remarkably well despite drastically reduced blood counts. Systemic complaints and weight loss should point to other etiologies of pancytopenia. Prior drug use, chemical exposure, and preceding viral illnesses must often be elicited with repeated questioning. A family history of hematologic diseases or blood abnormalities, and of pulmonary or liver fibrosis, may indicate a constitutional etiology of marrow failure.

PHYSICAL EXAMINATION

Petechiae and ecchymoses are typical, and retinal hemorrhages may be present. Pelvic and rectal examinations can often be deferred but, when performed, should be undertaken with great gentleness to avoid trauma; these will often show bleeding from the cervical os and blood in the stool. Pallor of the skin and mucous membranes is common except in the most acute cases or those already transfused. Infection on presentation is unusual but may occur if the patient has been symptomatic for a few weeks. Lymphadenopathy and splenomegaly are highly atypical of aplastic anemia. Café au lait spots and short stature suggest Fanconi's anemia; peculiar nails and leukoplakia suggest dyskeratosis congenita.

Laboratory Studies

BLOOD

The smear shows large erythrocytes and a paucity of platelets and granulocytes. Mean corpuscular volume (MCV) is commonly increased. Reticulocytes are absent or few, and lymphocyte numbers may be normal or reduced. The presence of immature myeloid forms suggests leukemia or MDS; nucleated RBCs suggest marrow fibrosis or tumor invasion; abnormal platelets suggest either peripheral destruction or MDS.

BONE MARROW

The bone marrow is usually readily aspirated but dilute on smear, and the fatty biopsy specimen may be grossly pale on withdrawal; a "dry tap" instead suggests fibrosis or myelophthisis. In severe aplasia the smear of the aspirated specimen shows only red cells, residual lymphocytes, and stromal cells; the biopsy (which should be >1 cm in length) is superior for determination of cellularity and shows mainly fat under the microscope, with hematopoietic cells occupying <25% of the marrow space; in the most serious cases the biopsy is virtually 100% fat. The correlation between marrow cellularity and disease severity is imperfect, in part because marrow cellularity declines physiologically with aging. Additionally, some patients with moderate disease by blood counts will have empty iliac crest biopsies, while "hot spots" of hematopoiesis may be seen in severe cases. If an iliac crest specimen is inadequate, cells may also be obtained by aspiration from the sternum. Residual hematopoietic cells should have normal morphology, except for mildly megaloblastic erythropoiesis; megakaryocytes are invariably greatly reduced and usually absent. Granulomas may indicate an infectious etiology of the marrow failure.

ANCILLARY STUDIES

Chromosome breakage studies of peripheral blood using diepoxybutane or mitomycin C should be performed on children and younger adults to exclude Fanconi's anemia. Very short telomere length (available commercially) strongly suggests the presence of a telomerase or shelterin mutation, which can be pursued by family studies and nucleotide sequencing. Chromosome studies of bone marrow cells are often revealing in MDS and should be negative in typical aplastic anemia. Flow cytometry offers a sensitive diagnostic test for PNH. Serologic studies may show evidence of viral infection, such as Epstein-Barr virus and HIV. Posthepatitis aplastic anemia is seronegative. The spleen size should be determined by CT scanning or ultrasound if the physical examination of the abdomen is unsatisfactory. MRI may be helpful to assess the fat content of a few vertebrae in order to distinguish aplasia from MDS.

DIAGNOSIS

The diagnosis of aplastic anemia is usually straightforward, based on the combination of pancytopenia with a fatty bone marrow. Aplastic anemia is a disease of the young and should be a leading diagnosis in the pancytopenic adolescent or young adult. When pancytopenia is secondary, the primary diagnosis is usually obvious from either history or physical examination: the massive spleen of alcoholic cirrhosis, the history of metastatic cancer or SLE, or miliary tuberculosis on chest radiograph (Table 107–1).

Diagnostic problems can occur with atypical presentations and among related hematologic diseases. Although pancytopenia is most common, some patients with bone marrow hypocellularity have depression of only one or two of three blood lines, with later progression to pancytopenia. The bone marrow in constitutional aplastic anemia is indistinguishable morphologically from the aspirate in acquired disease. The diagnosis can be suggested by family history, abnormal blood counts since childhood, or the presence of associated physical anomalies. Aplastic anemia may be difficult to distinguish from the hypocellular variety of MDS: MDS is favored by finding morphologic abnormalities, particularly of megakaryocytes and myeloid precursor cells, and typical cytogenetic abnormalities (see below).

PROGNOSIS

The natural history of severe aplastic anemia is rapid deterioration and death. Provision first of RBC and later of platelet transfusions and effective antibiotics are of some benefit, but few patients show spontaneous recovery. The major prognostic determinant is the blood count. Historically, severe disease was defined by the presence of two of three parameters: absolute neutrophil count <500/μL, platelet count <20,000/μL, and corrected reticulocyte count <1% (or absolute reticulocyte count <60,000/μL). In the era of effective immunosuppressive therapies, absolute numbers of reticulocytes (>25,000/uL) and lymphocytes (>1000/uL) may be a better predictor of response to treatment and long-term outcome.

Treatment: Aplastic Anemia

Severe acquired aplastic anemia can be cured by replacement of the absent hematopoietic cells (and the immune system) by stem cell transplant, or it can be ameliorated by suppression of the immune system to allow recovery of the patient's residual bone marrow function. Hematopoietic growth factors have limited usefulness, and glucocorticoids are of no value. Suspect exposures to drugs or chemicals should be discontinued; however, spontaneous recovery of severe blood count depression is rare, and a waiting period before beginning treatment may not be advisable unless the blood counts are only modestly depressed.

HEMATOPOIETIC STEM CELL TRANSPLANTATION

This is the best therapy for the younger patient with a fully histocompatible sibling donor (Chap. 114). Human leukocyte antigen (HLA) typing should be ordered as soon as the diagnosis of aplastic anemia is established in a child or younger adult. In transplant candidates, transfusion of blood from family members should be avoided so as to prevent sensitization to histocompatibility antigens, but limited numbers of blood products probably do not greatly affect outcome. For allogeneic transplant from fully matched siblings, long-term survival rates for children are approximately 90%. Transplant morbidity and mortality are increased among adults, due mainly to the higher risk of chronic GVHD and serious infections.

Most patients do not have a suitable sibling donor. Occasionally, a full phenotypic match can be found within the family and serve as well. Far more available are other alternative donors, either unrelated but histocompatible volunteers or closely but not perfectly matched family members. High resolution matching at HLA, as well as more effective conditioning regimens and GVHD prophylaxis, have led to improving survival rates in those patients who do proceed to alternative donor transplant, in some series approximating results with conventional sibling donors. Patients will be at risk for late complications, especially a higher rate of cancer, if radiation is used as a component of conditioning.

IMMUNOSUPPRESSION

The standard regimen of ATG in combination with cyclosporine induces hematologic recovery (independence from transfusion and a leukocyte count adequate to prevent infection) in 60–60% of patients. Children do especially well, while older adult patients often suffer complications due to the presence of comorbidities. An early robust hematologic response correlates with long-term survival. Improvement in granulocyte number is generally apparent within 2 months of treatment. Most recovered patients continue to have some degree of blood count depression, the MCV remains elevated, and the bone marrow cellularity returns toward normal very slowly if at all. Relapse (recurrent pancytopenia) is frequent, often occurring as cyclosporine is discontinued; most, but not all, patients respond to reinstitution of immunosuppression, but some responders become dependent on continued cyclosporine administration. Development of MDS, with typical marrow morphologic or cytogenetic abnormalities, occurs in approximately 15% of treated patients, usually but not invariably associated with a return of pancytopenia, and some patients develop leukemia. A laboratory diagnosis of PNH can generally be made at the time of presentation of aplastic anemia by flow cytometry; recovered patients may have frank hemolysis if the PNH clone expands. Bone marrow examinations

should be performed if there is an unfavorable change in blood counts.

Horse ATG and rabbit antilymphocyte globulin (ALG) are administered as intravenous infusions over 4 or 5 days, respectively. ATG binds to peripheral blood cells; therefore, platelet and granulocyte numbers may decrease further during active treatment. Serum sickness, a flulike illness with a characteristic cutaneous eruption and arthralgia, often develops approximately 10 days after initiating treatment. Methylprednisolone, 1 mg/kg per d for 2 weeks, can ameliorate the immune consequences of heterologous protein infusion. Excessive or extended glucocorticoid therapy is associated with avascular joint necrosis. Cyclosporine is administered orally at an initial high dose, with subsequent adjustment according to blood levels obtained every 2 weeks, rough levels should be between 150 and 200 ng/mL. The most important side effects are nephrotoxicity, hypertension, seizures, and opportunistic infections, especially *Pneumocystis carinii* (prophylactic treatment with monthly inhaled pentamidine is recommended).

Most patients with aplastic anemia lack a suitable marrow donor, and immunosuppression is the treatment of choice. Overall survival is equivalent with transplantation and immunosuppression. However, successful transplant cures marrow failure, whereas patients who recover adequate blood counts after immunosuppression remain at risk of relapse and malignant evolution. Because of excellent results in children and younger adults, allogeneic transplant should be performed if a suitable sibling donor is available. Increasing age and the severity of neutropenia are the most important factors weighing in the decision between transplant and immunosuppression in adults who have a matched family donor: Older patients do better with ATG and cyclosporine, whereas transplant is preferred if granulocytopenia is profound. Some patients may prefer immunosuppression; transplant is used for failure to recover blood counts or occurrence of late complications.

Outcomes following both transplant and immunosuppression have improved with time. High doses of cyclophosphamide, without stem cell rescue, have been reported to produce durable hematologic recovery, without relapse or evolution to MDS, but this treatment can produce sustained severe fatal neutropenia, and response is often delayed.

OTHER THERAPIES

The effectiveness of androgens has not been verified in controlled trials, but occasional patients will respond or even demonstrate blood count dependence on continued therapy. Sex hormones upregulate telomerase gene activity in vitro, possibly also their mechanism of action in improving marrow function. For patients with moderate disease or those with severe pancytopenia in whom immunosuppression has failed, a 3–4-month trial is appropriate. Hematopoietic growth factors (HGFs) are not recommended as initial therapy for severe aplastic anemia, and even their roles as adjuncts to immunosuppression are not clear.

SUPPORTIVE CARE

Meticulous medical attention is required so that the patient may survive to benefit from definitive therapy or, having failed treatment, to maintain a reasonable existence in the face of pancytopenia. First and most important, infection in the presence of severe neutropenia must be aggressively treated by prompt institution of parenteral, broad-spectrum antibiotics, usually ceftazidime or a combination of an aminoglycoside, cephalosporin, and semisynthetic penicillin. Therapy is empirical and must not await results of culture, although specific foci of infection such as oropharyngeal or anorectal abscesses, pneumonia, sinusitis, and typhlitis (necrotizing colitis) should be sought on physical examination and with radiographic studies. When indwelling plastic catheters become contaminated, vancomycin should be added. Persistent or recrudescent fever implies fungal disease: *Candida* and *Aspergillus* are common, especially after several courses of antibacterial antibiotics. A major reason for the improved prognosis in aplastic anemia has been the development of better antifungal drugs and the timely institution of such therapy when infection is suspected. Granulocyte transfusions using granulocyte colony-stimulating factor (G-CSF)—mobilized peripheral blood may be effective in the treatment of overwhelming or refractory infections. Hand washing, the single best method of preventing the spread of infection, remains a neglected practice. Nonabsorbed antibiotics for gut decontamination are poorly tolerated and not of proven value. Total reverse isolation does not reduce mortality from infections.

Both platelet and erythrocyte numbers can be maintained by transfusion. Alloimmunization historically limited the usefulness of platelet transfusions and is now minimized by several strategies, including use of single donors to reduce exposure and physical or chemical methods to diminish leukocytes in the product; HLA-matched platelets are often effective in patients refractory to random donor products. Inhibitors of fibrinolysis such as aminocaproic acid have not been shown to relieve mucosal oozing; the use of low-dose glucocorticoids to induce "vascular stability" is unproven and not recommended. Whether platelet transfusions are better used prophylactically or only as needed remains unclear. Any rational regimen of prophylaxis requires transfusions once or twice weekly to maintain the platelet count >10,000/μL (oozing from the gut, and presumably also from other vascular beds, increases precipitously at counts <5000/μL). Menstruation should be suppressed either by oral estrogens or nasal follicle-stimulating hormone/luteinizing hormone (FSH/LH) antagonists. Aspirin and other nonsteroidal anti-inflammatory agents inhibit platelet function and must be avoided.

Red blood cells should be transfused to maintain a normal level of activity, usually at a hemoglobin value of 70 g/L (90 g/L if there is underlying cardiac or pulmonary disease); a regimen of 2 units every 2 weeks will replace normal losses in a patient without a functioning bone marrow. In chronic anemia, the iron chelators, deferoxamine and deferasirox, should be added at approximately the fiftieth transfusion to avoid secondary hemochromatosis.

PURE RED CELL APLASIA

Other, more restricted forms of marrow failure occur, in which only a single circulating cell type is affected and the marrow shows corresponding absence or decreased numbers of specific precursor cells: aregenerative anemia as in PRCA (see below), thrombocytopenia with amegakaryocytosis (Chap. 115), and neutropenia without marrow myeloid cells in agranulocytosis (Chap. 60). In general, and in contrast to aplastic anemia and MDS, the unaffected lineages appear quantitatively and qualitatively normal. Agranulocytosis, the most frequent of these syndromes, is usually a complication of medical drug use (with agents similar to those related to aplastic anemia), either by a mechanism of direct chemical toxicity or by immune destruction. Agranulocytosis has an incidence similar to aplastic anemia but is especially frequent among older adults and in women. The syndrome should resolve with discontinuation of exposure, but significant mortality is attached to neutropenia in the older and often previously unwell patient. Both pure white cell aplasia (agranulocytosis without incriminating drug exposure) and amegakaryocytic thrombocytopenia are exceedingly rare and, like PRCA, appear to be due to destructive antibodies or lymphocytes and can respond to immunosuppressive therapies. In all the single lineage failure syndromes, progression to pancytopenia or leukemia is unusual.

Definition and Differential Diagnosis

PRCA is characterized by anemia, reticulocytopenia, and absent or rare erythroid precursor cells in the bone marrow. The classification of PRCA is shown in Table 107–4. In adults, PRCA is acquired. An identical syndrome can occur constitutionally: Diamond-Blackfan anemia, or congenital PRCA, is diagnosed at birth or in early childhood and often responds to glucocorticoid treatment; mutations in ribosomal RNA processing genes are etiologic. Temporary red cell failure occurs in transient aplastic crisis of hemolytic anemias due to acute parvovirus infection (Chap. 184) and in transient erythroblastopenia of childhood, which affects normal children.

Table 107–4 Classification of Pure Red Cell Aplasia

Self-limited
Transient erythroblastopenia of childhood
Transient aplastic crisis of hemolysis (acute B19 parvovirus infection)
Fetal red blood cell aplasia
Nonimmune hydrops fetalis (in utero B19 parvovirus infection)
Hereditary pure red cell aplasia
Congenital pure red cell aplasia (Diamond-Blackfan syndrome)
Acquired pure red cell aplasia
Thymoma and malignancy
Thymoma
Lymphoid malignancies (and more rarely other hematologic diseases)
Paraneoplastic to solid tumors
Connective tissue disorders with immunologic abnormalities
Systemic lupus erythematosus, juvenile rheumatoid arthritis, rheumatoid arthritis
Multiple endocrine gland insufficiency
Virus
Persistent B19 parvovirus, hepatitis, adult T cell leukemia virus, Epstein-Barr virus
Pregnancy
Drugs
Especially phenytoin, azathioprine, chloramphenicol, procainamide, isoniazid
Erythropoietin
Idiopathic

Clinical Associations and Etiology

PRCA has important associations with immune system diseases. A small minority of cases occur with a thymoma. More frequently, red cell aplasia can be the major manifestation of large granular lymphocytosis or may occur in chronic lymphocytic leukemia. Some patients may be

hypogammaglobulinemic. Infrequently (compared to agranulocytosis), PRCA can be due to an idiosyncratic drug reaction. Subcutaneous administration of erythropoietin (EPO) has provoked PRCA mediated by neutralizing antibodies.

Like aplastic anemia, PRCA results from diverse mechanisms. Antibodies to RBC precursors are frequently present in the blood, but T cell inhibition is probably the more common immune mechanism. Cytotoxic lymphocyte activity restricted by histocompatibility locus or specific for human T cell leukemia/lymphoma virus I—infected cells, as well as natural killer cell activity inhibitory of erythropoiesis, have been demonstrated in particularly well-studied individual cases.

Persistent Parvovirus B19 Infection

Chronic parvovirus infection is an important, treatable cause of PRCA. This common virus causes a benign exanthem of childhood (fifth disease) and a polyarthralgia/arthritis syndrome in adults. In patients with underlying hemolysis (or any condition that increases demand for RBC production), parvovirus infection can cause a transient aplastic crisis and an abrupt but temporary worsening of the anemia due to failed erythropoiesis. In normal individuals, acute infection is resolved by production of neutralizing antibodies to the virus, but in the setting of congenital, acquired, or iatrogenic immunodeficiency, persistent viral infection may occur. The bone marrow shows red cell aplasia and the presence of giant pronormoblasts (Fig. 107-2), which is the cytopathic sign of B19 parvovirus infection. Viral tropism for human erythroid progenitor cells is due to its use of erythrocyte P antigen as a cellular receptor for entry. Direct cytotoxicity of virus causes anemia if demands on erythrocyte production are high; in normal individuals, the temporary cessation of red cell production is not clinically apparent, and skin and joint symptoms are mediated by immune complex deposition.

Figure 107-2

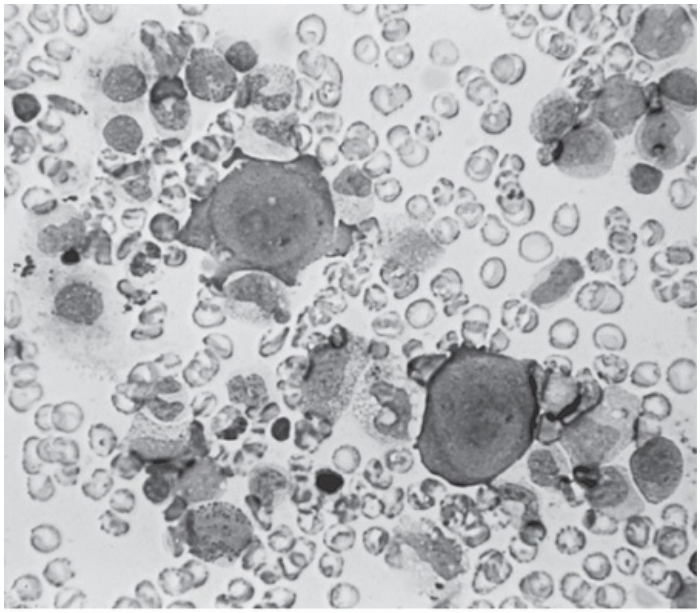

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

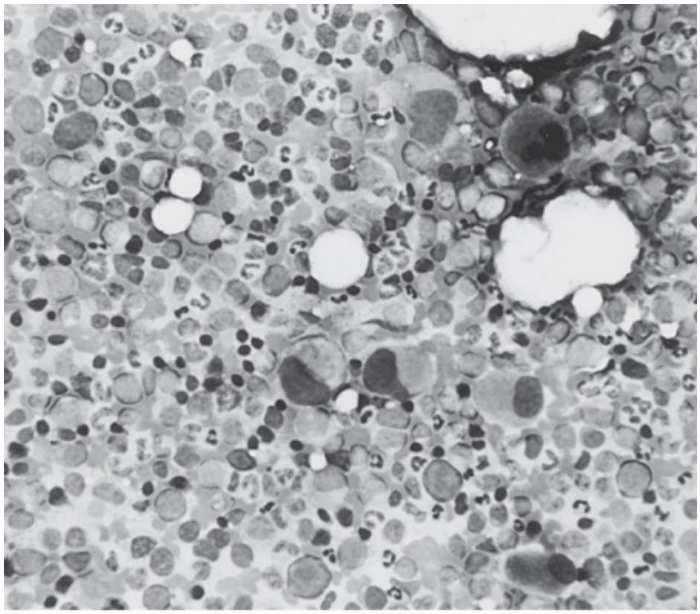

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

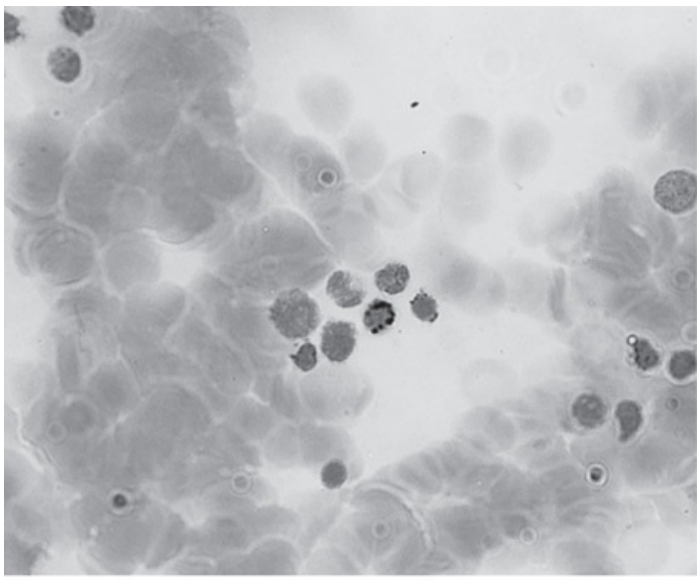

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

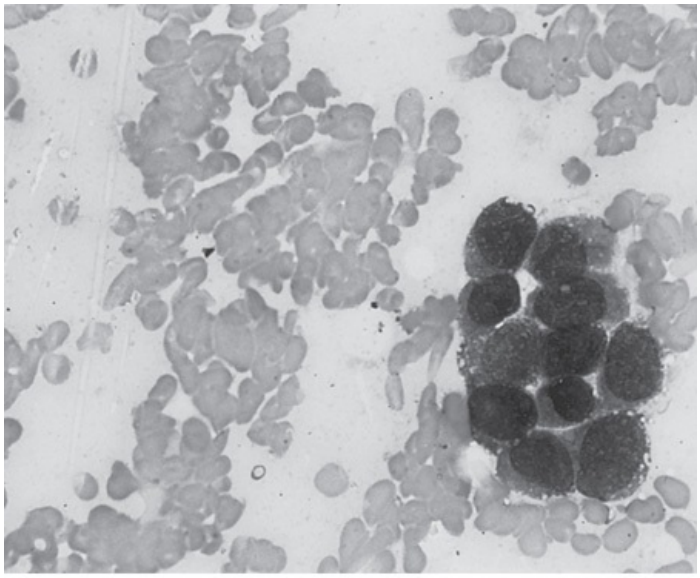

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathognomonic cells in marrow failure syndromes. ***A.*** Giant pronormoblast, the cytopathic effect of B19 parvovirus infection of the erythroid progenitor cell. ***B.*** Uninuclear megakaryocyte and microblastic erythroid precursors typical of the 5q–myelodysplasia syndrome. ***C.*** Ringed sideroblast showing perinuclear iron granules. ***D.*** Tumor cells present on a touch preparation made from the marrow biopsy of a patient with metastatic carcinoma.

Treatment: Pure Red Cell Aplasia

History, physical examination, and routine laboratory studies may disclose an underlying disease or a drug exposure. Thymoma should be sought by radiographic procedures. Tumor excision is indicated, but anemia does not necessarily improve with surgery. The diagnosis of parvovirus infection requires detection of viral DNA sequences in the blood (IgG and IgM antibodies are commonly absent). The presence of erythroid colonies has been considered predictive of response to immunosuppressive therapy in idiopathic PRCA.

Red cell aplasia is compatible with long-term survival with supportive care alone: a combination of erythrocyte transfusions and iron chelation. For persistent B19 parvovirus infection, almost all patients respond to intravenous immunoglobulin therapy (e.g., 0.4 g/kg daily for 5 days), although relapse and retreatment may be expected, especially in patients with AIDS. The majority of patients with idiopathic PRCA respond favorably to immunosuppression. Most first receive a course of glucocorticoids. Also effective are cyclosporine, ATG, azathioprine, cyclophosphamide, and the monoclonal antibody daclizumab, an antibody to the high affinity IL-2 receptor. PRCA developing on EPO therapy should be treated with immunosuppression and withdrawal of EPO.

MYELODYSPLASIA

Definition

Myelodysplasia or the MDSs are a heterogeneous group of hematologic disorders broadly characterized by cytopenias associated with a dysmorphic (or abnormal appearing) and usually cellular bone marrow, and by consequent ineffective blood cell production. A clinically useful nosology of these entities was first developed by the French-American-British Cooperative Group in 1983. Five entities were defined: refractory anemia (RA), refractory anemia with ringed sideroblasts (RARS), refractory anemia with excess blasts (RAEB), refractory anemia with excess blasts in transformation (RAEB-t), and chronic myelomonocytic leukemia (CMML). The World Health Organization (WHO) classification (2002) recognized that the distinction between RAEB-t and acute myeloid leukemia is arbitrary and groups them together as acute leukemia, that CMML behaves as a myeloproliferative disease, and separated refractory anemias with dysmorphic change restricted to erythroid lineage from those with multilineage changes. In a revision, specific categories for unilineage dysplasias were added (Table 107–5).

Table 107–5 World Health Organization (WHO) Classification of Myelodysplastic Syndromes/Neoplasms

Name	WHO Estimated Proportion of Patients with	Peripheral Blood: Key Features	Bone Marrow: Key Features

	MDS		
Refractory cytopenias with unilineage dysplasia (RCUD):			
Refractory anemia (RA)	10-20%	Anemia <1% of blasts	Unilineage erythroid dysplasia (in ≥10% of cells) <5% blasts
Refractory neutropenia (RN)	<1%	Neutropenia <1% blasts	Unilineage granulocytic dysplasia <5% blasts
Refractory thrombocytopenia (RT)	<1%	Thrombocytopenia <1% blasts	Unilineage megakaryocytic dysplasia <5% blasts
Refractory anemia with ring sideroblasts (RARS)	3-11%	Anemia No blasts	Unilineage erythroid dysplasia ≥15% of erythroid precursors are ring sideroblasts <5% blasts
Refractory cytopenias with multilineage dysplasia (RCMD)	30%	Cytopenia(s) <1% blasts No Auer rods	Multilineage dysplasia ± ring sideroblasts <5% blasts No Auer rods
Refractory anemia with excess blasts, Type 1 (RAEB-1)	40%	Cytopenia(s) <5% blasts No Auer rods	Unilineage or multilineage dysplasia
Refractory anemia with excess blasts, type 2 (RAEB-2)		Cytopenia(s) 5-19% blasts ± Auer rods	Unilineage or multilineage dysplasia 10-19% blasts ± Auer rods
MDS associated with isolated Del(5q) (Del(5q)	Uncommon	Anemia Normal or high platelet count <1% blasts	Isolated 5q31 chromosome deletion Anemia; hypolobated megakaryocytes <5% blasts
Childhood MDS, including refractory cytopenia of childhood (*provisional?*) (RCC)	<1 %	Pancytopenia	<5% marrow blasts for RCC Marrow usually hypocellular
MDS, unclassifiable (MDS-U)	?	Cytopenia ≤1% blasts	Does not fit other categories Dysplasia <5% blasts If no dysplasia, MDS-associated karyotype

Note: If peripheral blood blasts are 2–4%, the diagnosis is RAEB-1 even if marrow blasts are less than 5%. If Auer rods are present, the WHO considers the diagnosis RAEB-2 if the blast proportion is less than 20% (even if less than 10%), AML if at least 20% blasts. For all subtypes, peripheral blood monocytes are less than 1×10^9/L. Bicytopenia may be observed in RCUD subtypes, but pancytopenia with unilineage marrow dysplasia should be classified as MDS-U. Therapy-related MDS (t-MDS), whether due to alkylating agents, topoisomerase II (t-MDS/t-AML) in the WHO classification of AML and precursor lesions. The listing in this table excludes MDS/myeloproliferative neoplasm overlap categories, such as chronic myelomonocytic leukemia, juvenile myelomonocytic leukemia, and the provisional entity RARS with thrombocytosis.

Abbreviation: MDS, myelodysplastic syndrome.

The diagnosis of MDS may be a challenge, as sometimes subtle clinical and pathologic features must be distinguished in a usually older adult patient >70 years of age with comorbidities; furthermore, precise diagnostic categorization requires a hematopathologist knowledgeable in the latest classification scheme. Nonetheless, it is important that the internist and primary care physician be sufficiently familiar with MDS to expedite referral to a hematologist, both because many new therapies are now available to improve hematopoietic function and the judicious use of supportive care can improve the patient's quality of life.

Epidemiology

Idiopathic MDS is a disease of the older adult; the mean age at onset is older than 70 years. There is a slight male preponderance. MDS is a relatively common form of bone marrow failure, with reported incidence rates of 35 to >100 per million persons in the general population and 120 to >500 per million in the older adult. MDS is rare in children, but monocytic leukemia can be seen. Secondary or therapy-related MDS is not age related. Rates of MDS have increased over time, due to the recognition of the syndrome by physicians and the aging of the population.

Etiology and Pathophysiology

MDS is associated with environmental exposures such as radiation and benzene; other risk factors have been reported inconsistently. Secondary MDS occurs as a late toxicity of cancer treatment, usually with a combination of radiation and the radiomimetic alkylating agents such as busulfan,

nitrosourea, or procarbazine (with a latent period of 5–7 years) or the DNA topoisomerase inhibitors (2 years). Both acquired aplastic anemia following immunosuppressive treatment and Fanconi's anemia can evolve into MDS.

MDS is a clonal hematopoietic stem cell disorder leading to impaired cell proliferation and differentiation. Cytogenetic abnormalities are found in approximately one-half of patients, and some of the same specific lesions are also seen in frank leukemia; aneuploidy is more frequent than translocations. Both presenting and evolving hematologic manifestations result from the accumulation of multiple genetic lesions: loss of tumor suppressor genes, activating oncogene mutations, or other harmful alterations. Cytogenetic abnormalities are not random (loss of all or part of 5, 7, and 20, trisomy of 8) and may be related to etiology (11q23 following topoisomerase II inhibitors); CMML is often associated with t(5;12) that creates a chimeric *tel-PDGFβ* gene. The type and number of cytogenetic abnormalities strongly correlate with the probability of leukemic transformation and survival. Mutations of N-*ras* (an oncogene), *p53* and *IRF-1* (tumor suppressor genes), *Bcl-2* (an antiapoptotic gene), and others have been reported in some patients but likely occur late in the sequence leading to leukemic transformation. Apoptosis of marrow cells is increased in early stage and low-risk categories of MDS, presumably due to these acquired genetic alterations or possibly to an overlaid immune response. An immune pathophysiology has been suggested for trisomy 8 MDS, which often responds clinically to immunosuppressive therapy. Such patients have T cell activity directed to the cytogenetically aberrant clone. Sideroblastic anemia may be related to mutations in mitochondrial genes; ineffective erythropoiesis and disordered iron metabolism are the functional consequences of the genetic alterations.

Clinical Features

Anemia dominates the early course. Most symptomatic patients complain of the gradual onset of fatigue and weakness, dyspnea, and pallor, but at least one-half the patients are asymptomatic, and their MDS is discovered only incidentally on routine blood counts. Previous chemotherapy or radiation exposure is an important historic fact. Fever and weight loss should point to a myeloproliferative rather than myelodysplastic process. Children with Down syndrome are susceptible to MDS, and a family history may indicate a hereditary form of sideroblastic anemia or Fanconi's anemia.

The physical examination is remarkable for signs of anemia; approximately 20% of patients have splenomegaly. Some unusual skin lesions, including Sweet's syndrome (febrile neutrophilic dermatosis), occur with MDS. Autoimmune syndromes are not infrequent.

Laboratory Studies

Blood

Anemia is present in the majority of cases, either alone or as part of bi- or pancytopenia; isolated neutropenia or thrombocytopenia is more unusual. Macrocytosis is common, and the smear may be dimorphic with a distinctive population of large red blood cells. Platelets are also large and lack granules. In functional studies, they may show marked abnormalities, and patients may have bleeding symptoms despite seemingly adequate numbers. Neutrophils are hypogranulated; have hyposegmented, ringed, or abnormally segmented nuclei; contain Döhle bodies; and may be functionally deficient. Circulating myeloblasts usually correlate with marrow blast numbers, and their quantitation is important for classification and prognosis. The total white blood cell count (WBC) is usually normal or low, except in chronic myelomonocytic leukemia. As in aplastic anemia, MDS can be associated with a clonal population of PNH cells.

BONE MARROW

The bone marrow is usually normal or hypercellular, but in 20% of cases it is sufficiently hypocellular to be confused with aplasia. No single characteristic feature of marrow morphology distinguishes MDS, but the following are commonly observed: dyserythropoietic changes (especially nuclear abnormalities) and ringed sideroblasts in the erythroid lineage; hypogranulation and hyposegmentation in granulocytic precursors, with an increase in myeloblasts; and megakaryocytes showing reduced numbers or disorganized nuclei. Megaloblastic nuclei associated with defective hemoglobinization in the erythroid lineage are common. Prognosis strongly correlates with the proportion of marrow blasts. Cytogenetic analysis and fluorescent in situ hybridization can identify chromosomal abnormalities.

Differential Diagnosis

Deficiencies of vitamin B12 or folate should be excluded by appropriate blood tests; vitamin B6 deficiency can be assessed by a therapeutic trial of pyridoxine if the bone marrow shows ringed sideroblasts. Marrow dysplasia can be observed in acute viral infections, drug reactions, or chemical toxicity but should be transient. More difficult are the distinctions between hypocellular MDS and aplasia or between refractory anemia with excess blasts and early acute leukemia. The WHO considers the presence of 20% blasts in the marrow as the criterion that separates acute myeloid leukemia (AML) from MDS.

Prognosis

The median survival varies greatly from years for patients with 5q—or sideroblastic anemia to a few months in refractory anemia with excess blasts or severe pancytopenia associated with monosomy 7; an International Prognostic Scoring System (IPSS; Table 107–6) assists in making predictions. Most patients die as a result of complications of pancytopenia and not due to leukemic transformation; perhaps one-third will succumb to other diseases unrelated to their MDS. Precipitous worsening of pancytopenia, acquisition of new chromosomal abnormalities on serial cytogenetic determination, increase in the number of blasts, and marrow fibrosis are all poor prognostic indicators. The outlook in therapy-related MDS, regardless of type, is extremely poor, and most patients will progress within a few months to refractory AML.

Table 107–6 International Prognostic Scoring System (IPSS)

	Score Value				
Prognostic Variable	**0**	**0.5**	**1**	**1.5**	**2**
Bone marrow blasts (%)	<5%	5–10%		11–20%	21–30%
Karyotype[a]	Good	Intermediate	Poor		
Cytopenia[b] (lineages affected)	0 or 1	2 or 3			
Risk Group Scores	**Score**				
Low	0				
Intermediate-1	0.5–1				
Intermediate-2	1.5–2				
High	≥2.5				

[a]Good, normal, -Y, del(5q), del (20q); poor, complex (≥3 abnormalities) or chromosome 7 abnormalities; intermediate, all other abnormalities.

[b]Cytopenias defined as Hb <100 g/L, platelet count <100,000/μL, absolute neutrophil count <1500/μL.

Treatment: Myelodysplasia

Historically, the therapy of MDS has been unsatisfactory. Only stem cell transplantation offers cure. Survival rates of 50% at 3 years have been reported, but older patients are particularly prone to develop treatment-related mortality and morbidity. Results of transplant using matched unrelated donors are comparable, although most series contain younger and more highly selected cases. However, multiple new drugs have been approved for use in MDS. Several regimens appear to not only improve blood counts but to delay onset of leukemia and to improve survival. The choice of therapy and implementation of treatment are complicated and require hematologic expertise.

MDS has been regarded as particularly refractory to cytotoxic chemotherapy regimens but is probably no more resistant to effective treatment than AML in the older adult, in whom drug toxicity is often fatal and remissions, if achieved, are brief.

Low doses of cytotoxic drugs have been administered for their "differentiating" potential, and from this experience has emerged drug therapies based on pyrimidine analogues. Azacitidine is directly cytotoxic but also inhibits DNA methylation, thereby altering gene expression; however, demethylation status has not correlated well with clinical effects Azacitidine improves blood counts and also survival in a minority of MDS patients, compared to best supportive care. Azacitidine has been administered subcutaneously, daily for 7 days, at 4-week intervals, for at least four cycles before assessing for response. Decitabine is closely related to azacitidine and more potent. Similar to azacitidine, approximately 20% of patients show responses in blood counts, with a duration of response of almost a year. Activity may be higher in more advanced MDS subtypes. Decitabine is administered by continuous intravenous infusion, every 8 hours for 3 days, in repeating cycles. The optimal dose regimens for both azacitidine and decitabine are still being determined in clinical research protocols. The major toxicity of both azacitidine and decitabine is myelosuppression, leading to worsened blood counts. Other symptoms associated with cancer chemotherapy frequently occur.

Lenalidomide, a thalidomide derivative with a more favorable toxicity profile, is particularly effective in reversing anemia in MDS patients with 5q—syndrome; not only do a high proportion of these patients become transfusion independent with normal or near-normal hemoglobin levels, but their cytogenetics also become normal. The drug has many biologic activities, and it is unclear which is critical for clinical efficacy. Lenalidomide is administered orally. Most patients will improve within 3 months of initiating therapy. Toxicities include myelosuppression (worsening thrombocytopenia and neutropenia, necessitating blood count monitoring) and an increased risk of deep vein thrombosis and pulmonary embolism.

ATG and cyclosporine, as employed in aplastic anemia, also may produce sustained independence from transfusion and improve survival. Immunosuppression with ATG or, in more recent studies, the

anti-CD52 monoclonal antibody Campath is especially effective in younger MDS patients (younger than age 60 years) with more favorable IPSS scores and who bear the histocompatability antigen HLA-DR15.

HGFs can improve blood counts but, as in most other marrow failure states, have been most beneficial to patients with the least severe pancytopenia. G-CSF treatment alone failed to improve survival in a controlled trial. EPO alone or in combination with G-CSF can improve hemoglobin levels, but mainly in those with low serum EPO levels who have no or only a modest need for transfusions, and retrospective analysis suggests improved survival with treatment

The same principles of supportive care described for aplastic anemia apply to MDS. Despite improvements in drug therapy, many patients will be anemic for years. RBC transfusion support should be accompanied by iron chelation to prevent secondary hemochromatosis.

MYELOPHTHISIC ANEMIAS

Fibrosis of the bone marrow (see Fig. 103-2), usually accompanied by a characteristic blood smear picture called *leukoerythroblastosis*, can occur as a primary hematologic disease, called *myelofibrosis* or *myeloid metaplasia* (Chap. 108), and as a secondary process, called *myelophthisis*. Myelophthisis, or secondary myelofibrosis, is reactive. Fibrosis can be a response to invading tumor cells, usually an epithelial cancer of breast, lung, a prostate origin or neuroblastoma. Marrow fibrosis may occur with infection of mycobacteria (both *Mycobacterium tuberculosis* and *M. avium*), fungi, or HIV, and in sarcoidosis. Intracellular lipid deposition in Gaucher's disease and obliteration of the marrow space related to absence of osteoclast remodeling in congenital osteopetrosis also can produce fibrosis. Secondary myelofibrosis is a late consequence of radiation therapy or treatment with radiomimetic drugs. Usually the infectious or malignant underlying processes are obvious. Marrow fibrosis can also be a feature of a variety of hematologic syndromes, especially chronic myeloid leukemia, multiple myeloma, lymphomas, myeloma, and hairy cell leukemia.

The pathophysiology has three distinct features: proliferation of fibroblasts in the marrow space (myelofibrosis); the extension of hematopoiesis into the long bones and into extramedullary sites, usually the spleen, liver, and lymph nodes (myeloid metaplasia); and ineffective erythropoiesis. The etiology of the fibrosis is unknown but most likely involves dysregulated production of growth factors: Platelet-derived growth factor and transforming growth factor ß have been implicated. Abnormal regulation of other hematopoietins would lead to localization of blood-producing cells in nonhematopoietic tissues and uncoupling of the usually balanced processes of stem cell proliferation and differentiation. Myelofibrosis is remarkable for pancytopenia despite very large numbers of circulating hematopoietic progenitor cells.

Anemia is dominant in secondary myelofibrosis, usually normocytic and normochromic. The diagnosis is suggested by the characteristic leukoerythroblastic smear (see Fig. 108-1). Erythrocyte morphology is highly abnormal, with circulating nucleated RBCs, teardrops, and shape distortions. WBC numbers are often elevated, sometimes mimicking a leukemoid reaction, with circulating myelocytes, promyelocytes, and myeloblasts. Platelets may be abundant and are often of giant size. Inability to aspirate the bone marrow, the characteristic "dry tap," can allow a presumptive diagnosis in the appropriate setting before the biopsy is decalcified.

The course of secondary myelofibrosis is determined by its etiology, usually a metastatic tumor or an advanced hematologicmalignancy. Treatable causes must be excluded, especially tuberculosis and fungus. Transfusion support can relieve symptoms.

FURTHER READINGS

Ades L et al: Efficacy and safety of lenalidomide in intermediate-2 or high-risk myelodysplastic syndromes with 5q deletion: Results of a phase 2 study. Blood 113:3947, 2009[PMID: 18987358]

Bagby GC, Alter BP: Fanconi anemia. Semin Hematol 43:147, 2006[PMID: 16822457]

Calado RT, Young NS: Telomere diseases. N Engl J Med 361:2353, 2009[PMID: 20007561]

Estey E et al: Acute myeloid leukemia and myelodysplastic syndromes in older patients. J Clin Oncol 25:1908, 2007[PMID: 17488990]

Fenaux P et al: Efficacy of azacitidine compared with that of standard care regimens in the treatment of higher-risk myelodysplastic syndromes: A randomized, open-label, phase III trial. Lancet Oncol 10:223, 2009[PMID: 19230772]

Flygare J, Karlsson S: Diamond-Blackfan anemia: Erythropoiesis lost in translation. Blood 109:3152, 2007[PMID: 17164339]

Jadersten M et al: Erythropoietin and granulocyte-colony stimulating factor treatment associated with

improved survival in myelodysplastic syndromes. J Clin Oncol 26:3607, 2008[PMID: 18559873]

Ma X et al: Myelodysplastic syndromes: Incidence and survival in the United States. Cancer 109:1536, 2007[PMID: 17345612]

Young NS, Brown KE: Parvovirus B19. N Engl J Med 350:586, 2004[PMID: 14762186]

——— et al: Current concepts in the pathophysiology and treatment of aplastic anemia. Blood 108:2509, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

POLYCYTHEMIA VERA AND OTHER MYELOPROLIFERATIVE DISEASES: INTRODUCTION

The World Health Organization (WHO) classification of the chronic myeloproliferative diseases (MPDs) includes eight disorders, some of which are rare or poorly characterized (Table 108–1) but all of which share an origin in a multipotent hematopoietic progenitor cell; overproduction of one or more of the formed elements of the blood without significant dysplasia; a predilection to extramedullary hematopoiesis, myelofibrosis; and transformation at varying rates to acute leukemia. Within this broad classification, however, significant phenotypic heterogeneity exists. Some diseases such as chronic myelogenous leukemia (CML), chronic neutrophilic leukemia (CNL), and chronic eosinophilic leukemia (CEL) express primarily a myeloid phenotype, while in others such as polycythemia vera (PV), primary myelofibrosis (PMF), and essential thrombocytosis (ET), erythroid or megakaryocytic hyperplasia predominates. The latter three disorders, in contrast to the former three, also appear capable of transforming into each other.

Table 108–1 WHO Classification of Chronic Myeloproliferative Disorders

Chronic myelogenous leukemia, bcr-abl–positive
Chronic neutrophilic leukemia
Chronic eosinophilic leukemia, not otherwise specified
Polycythemia vera
Primary myelofibrosis
Essential thrombocytosis
Mastocytosis
Myeloproliferative neoplasms, unclassifiable

Such phenotypic heterogeneity has a genetic basis; CML is the consequence of the balanced translocation between chromosomes 9 and 22 [t(9;22)(q34;11)]; CNL has been associated with a t(15;19) translocation; and CEL occurs with a deletion or balanced translocations involving the *PDGFR* α gene. By contrast, to a greater or lesser extent, PV, PMF, and ET are characterized by expression of a *JAK2* mutation, V617F that causes constitutive activation of this tyrosine kinase that is essential for the function of the erythropoietin and thrombopoietin receptors but not the granulocyte colony-stimulating factor receptor. This essential distinction is also reflected in the natural history of CML, CNL, and CEL, which is usually measured in years, and their high rate of transformation into acute leukemia. By contrast, the natural history of PV, PMF, and ET is usually measured in decades, and transformation to acute leukemia is uncommon in the absence of exposure to mutagenic agents. This chapter, therefore, will focus only on PV, PMF, and ET, because their clinical overlap is substantial but their clinical courses are distinctly different.

Other chronic myeloproliferative disorders will be discussed in Chap. 109.

POLYCYTHEMIA VERA

PV is a clonal disorder involving a multipotent hematopoietic progenitor cell in which phenotypically normal red cells, granulocytes, and platelets accumulate in the absence of a recognizable physiologic stimulus. The most common of the chronic myeloproliferative disorders, PV occurs in 2 per 100,000 persons, sparing no adult age group and increasing with age to rates as high as 18/100,000. Familial transmission occurs but is infrequent and women predominate among sporadic cases.

Etiology

The etiology of PV is unknown. Although nonrandom chromosome abnormalities such as 20q and trisomy 8 and 9 have been documented in up to 30% of untreated PV patients, unlike CML, no consistent cytogenetic abnormality has been associated with the disorder. However, a mutation in the autoinhibitory, pseudokinase domain of the tyrosine kinase JAK2—that replaces valine with phenylalanine (V617F), causing constitutive activation of the kinase—appears to have a central role in the pathogenesis of PV.

JAK2 is a member of an evolutionarily well-conserved, nonreceptor tyrosine kinase family and serves

as the cognate tyrosine kinase for the erythropoietin and thrombopoietin receptors. It also functions as an obligate chaperone for these receptors in the Golgi apparatus and is responsible for their cell-surface expression. The conformational change induced in the erythropoietin and thrombopoietin receptors following binding to erythropoietin or thrombopoietin leads to JAK2 autophosphorylation, receptor phosphorylation, and phosphorylation of proteins involved in cell proliferation, differentiation, and resistance to apoptosis. Transgenic animals lacking *JAK2* die as embryos from severe anemia. Constitutive activation of JAK2 can explain the erythropoietin-independent erythroid colony formation, and the hypersensitivity of PV erythroid progenitor cells to erythropoietin and other hematopoietic growth factors, their resistance to apoptosis in vitro in the absence of erythropoietin, their rapid terminal differentiation, and their increase in Bcl-X_L expression, all of which are characteristic in PV.

Importantly, the *JAK2* gene is located on the short arm of chromosome 9, and loss of heterozygosity on chromosome 9p, due to mitotic recombination is the most common cytogenetic abnormality in PV. The segment of 9p involved contains the *JAK2* locus; loss of heterozygosity in this region leads to homozygosity for the mutant *JAK2* V617F. More than 90% of PV patients express this mutation, as do approximately 50% of PMF and ET patients. Homozygosity for the mutation occurs in approximately 30% of PV patients and 60% of PMF patients; homozygosity is rare in ET. Over time, a portion of PV *JAK2* V617F heterozygotes become homozygotes due to mitotic recombination, but usually not after 10 years of the disease. PV patients who do not express *JAK2* V617F are not clinically different from those who do, nor do *JAK2* V617F heterozygotes differ clinically from homozygotes. Interestingly, predisposition to acquire mutations in *JAK2* appears to be associated with a specific *JAK2* haplotype, GGCC. *JAK2* V617F is the basis for many of the phenotypic and biochemical characteristics of PV such as elevation of the leukocyte alkaline phosphatase (LAP) score; however, it cannot solely account for the entire PV phenotype and is probably not the initiating lesion in the three MPDs. First, some PV patients with the same phenotype and documented clonal disease lack this mutation. Second, ET and PMF patients have the same mutation but different clinical phenotypes. Third, familial PV can occur without the mutation, even when other members of the same family express it. Fourth, not all the cells of the malignant clone express *JAK2* V617F. Fifth, *JAK2* V617F has been observed in patients with long-standing idiopathic erythrocytosis. Sixth, in some patients, *JAK2* V617F appears to be acquired after another mutation. Finally, in some *JAK2* V617F-positive PV or ET patients, acute leukemia can occur in a *JAK2* V617F-negative progenitor cell. However, while *JAK2* V617F alone may not be sufficient to cause PV, it is essential for the transformation of ET to PV, though not for its transformation to PMF.

Clinical Features

Although splenomegaly may be the initial presenting sign in PV, most often the disorder is first recognized by the incidental discovery of a high hemoglobin or hematocrit. With the exception of aquagenic pruritus, no symptoms distinguish PV from other causes of erythrocytosis.

Uncontrolled erythrocytosis causes hyperviscosity, leading to neurologic symptoms such as vertigo, tinnitus, headache, visual disturbances, and transient ischemic attacks (TIAs). Systolic hypertension is also a feature of the red cell mass elevation. In some patients, venous or arterial thrombosis may be the presenting manifestation of PV. Any vessel can be affected; but cerebral, cardiac, or mesenteric vessels are most commonly involved. Intraabdominal venous thrombosis is particularly common in young women and may be catastrophic if a sudden and complete obstruction of the hepatic vein occurs. Indeed, PV should be suspected in any patient who develops hepatic vein thrombosis. Digital ischemia, easy bruising, epistaxis, acid-peptic disease, or gastrointestinal hemorrhage may occur due to vascular stasis or thrombocytosis. Erythema, burning, and pain in the extremities, a symptom complex known as erythromelalgia, is another complication of the thrombocytosis of PV due to increased platelet stickiness. Given the large turnover of hematopoietic cells, hyperuricemia with secondary gout, uric acid stones, and symptoms due to hypermetabolism can also complicate the disorder.

Diagnosis

When PV presents with erythrocytosis in combination with leukocytosis, thrombocytosis, or both, the diagnosis is apparent. However, when patients present with an elevated hemoglobin or hematocrit alone, or with thrombocytosis alone, the diagnostic evaluation is more complex because of the many diagnostic possibilities (Table 108–2). Furthermore, unless the hemoglobin level is ≥20 g/DL (hematocrit ≥60%), it is not possible to distinguish true erythrocytosis from disorders causing plasma volume contraction. Uniquely in PV, in contrast to other causes for true erythrocytosis, an expanded plasma volume can mask the elevated red cell mass; thus, red cell mass and plasma volume determinations are mandatory to establish the presence of an absolute erythrocytosis and to distinguish this from relative erythrocytosis due to a reduction in plasma volume alone (also known as *stress* or *spurious erythrocytosis* or *Gaisböck's syndrome*). This is true even with the finding of *JAK2* V617F mutation, because not every patient with PV expresses this mutation, while patients without PV do. Figure 57-18 illustrates a diagnostic algorithm for the evaluation of suspected erythrocytosis.

Table 108–2 Causes of Erythrocytosis

Relative Erythrocytosis	
Hemoconcentration secondary to dehydration, diuretics, ethanol abuse, androgens or tobacco abuse	
Absolute Erythrocytosis	
Hypoxia	**Tumors**
Carbon monoxide intoxication	Hypernephroma
High oxygen-affinity	Hepatoma
hemoglobin	Cerebellar hemangioblastoma
High altitude	Uterine myoma
Pulmonary disease	Adrenal tumors
Right to left cardiac or	Meningioma
vascular shunts	Pheochromocytoma
Sleep apnea syndrome	**Drugs**
Hepatopulmonary syndrome	Androgens
Renal Disease	Recombinant erythropoietin
Renal artery stenosis	**Familial (with normal hemoglobin function)**
Focal sclerosing or membranous glomerulonephritis	Erythropoietin receptor mutation
Postrenal transplantation	VHL mutations (Chuvash polycythemia)
Renal cysts	2,3-BPG mutation
Bartter's syndrome	**Polycythemia vera**

Abbreviations: 2,3-BPG, 2,3-bisphosphoglycerate; VHL, von Hippel-Lindau.

Once absolute erythrocytosis has been established, its cause must be determined. An elevated plasma erythropoietin level suggests either a hypoxic cause for erythrocytosis or autonomous erythropoietin production, in which case assessment of pulmonary function and an abdominal CT scan to evaluate renal and hepatic anatomy are appropriate. A normal erythropoietin level, however, does not exclude a secondary cause for erythrocytosis or PV. In PV, in contrast to hypoxic erythrocytosis, the arterial oxygen saturation is normal. However, a normal oxygen saturation does not exclude a high-affinity hemoglobin as a cause for erythrocytosis; documentation of previous hemoglobin levels and a family study are important in this regard.

Other laboratory studies that may aid in diagnosis include the red cell count, mean corpuscular volume, and red cell distribution width (RDW). Only three situations cause microcytic erythrocytosis: β-thalassemia trait, hypoxic erythrocytosis, and PV. With β-thalassemia trait the RDW is normal, whereas with hypoxic erythrocytosis and PV, the RDW is usually elevated due to iron deficiency. Today, an assay for *JAK2* V617F has superseded other tests for establishing the diagnosis of PV. Of course, in patients with associated acid-peptic disease, occult gastrointestinal bleeding may lead to presentation with hypochromic, microcytic anemia, masking the presence of PV.

A bone marrow aspirate and biopsy provide no specific diagnostic information since these may be normal or indistinguishable from ET or PMF, and unless there is a need to exclude some other disorder, these procedures need not be done. Although the presence of a cytogenetic abnormality such as trisomy 8 or 9 or 20q– in the setting of an expanded red cell mass supports a clonal etiology, no specific cytogenetic abnormality is associated with PV, and the absence of a cytogenetic marker does not exclude the diagnosis.

Complications

Many of the clinical complications of PV relate directly to the increase in blood viscosity associated with red cell mass elevation and indirectly to the increased turnover of red cells, leukocytes, and platelets with the attendant increase in uric acid and cytokine production. The latter appears to be responsible for constitutional symptoms, while peptic ulcer disease may be due to *Helicobacter pylori* and the pruritus associated with this disorder may be a consequence of basophil activation by *JAK2* V617F. A sudden increase in spleen size can be associated with splenic infarction. Myelofibrosis appears to be part of the natural history of the disease but is a reactive, reversible process that does not itself impede hematopoiesis and by itself has no prognostic significance. In some patients, however, the myelofibrosis is accompanied by significant extramedullary hematopoiesis, hepatosplenomegaly, and transfusion-dependent anemia, which are manifestations of stem cell failure. The organomegaly can cause significant mechanical discomfort, portal hypertension, and progressive cachexia. Although the incidence of acute nonlymphocytic leukemia is increased in PV, the incidence of acute leukemia in patients not exposed to chemotherapy or radiation is low, and the development of

leukemia is related to the development of extramedullary hematopoiesis, hepatosplenomegaly, and transfusion-dependent anemia or exposure to chemotherapy. Importantly, chemotherapy alone, including hydroxyurea, has been associated with acute leukemia that develops in *JAK2* V617F —negative stem cells. *Erythromelalgia* is a curious syndrome of unknown etiology associated with thrombocytosis, primarily involving the lower extremities and usually manifested by erythema, warmth, and pain of the affected appendage and occasionally digital infarction. It occurs with a variable frequency in MPD patients and is usually responsive to salicylates. Some of the central nervous system symptoms observed in patients with PV such as ocular migraine, appear to represent a variant of erythromelalgia.

If left uncontrolled, erythrocytosis can lead to thrombosis involving vital organs such as the liver, heart, brain, or lungs. Patients with massive splenomegaly are particularly prone to thrombotic events because the associated increase in plasma volume masks the true extent of the red cell mass elevation as measured by the hematocrit or hemoglobin level. A "normal" hematocrit or hemoglobin level in a PV patient with massive splenomegaly should be considered indicative of an elevated red cell mass until proven otherwise.

Treatment: Polycythemia Vera

PV is generally an indolent disorder, the clinical course of which is measured in decades, and its management should reflect its tempo. Thrombosis due to erythrocytosis is the most significant complication, and maintenance of the hemoglobin level at ≤140 g/L (14 g/dL; hematocrit <45%) in men and ≤120 g/L (12 g/dL; hematocrit <42%) in women is mandatory to avoid thrombotic complications. Phlebotomy serves initially to reduce hyperviscosity by bringing the red cell mass into the normal range. Periodic phlebotomies thereafter serve to maintain the red cell mass within the normal range and to induce a state of iron deficiency that prevents an accelerated reexpansion of the red cell mass. In most PV patients, once an iron-deficient state is achieved, phlebotomy is usually only required at 3-month intervals. Neither phlebotomy nor iron deficiency increases the platelet count relative to the effect of the disease itself, and thrombocytosis is not correlated with thrombosis in PV, in contrast to the strong correlation between erythrocytosis and thrombosis in this disease. The use of salicylates as a tonic against thrombosis in PV patients is not only potentially harmful if the red cell mass is not controlled by phlebotomy, but also an unproven remedy. Anticoagulants are only indicated when a thrombosis has occurred and can be difficult to monitor owing to the artifactual imbalance between the test tube anticoagulant and plasma that occurs when blood from these patients is assayed for prothrombin or partial thromboplastin activity if the red cell mass is substantially elevated. Asymptomatic hyperuricemia (<10 mg%) requires no therapy, but allopurinol should be administered to avoid further elevation of the uric acid when chemotherapy is employed to reduce splenomegaly or leukocytosis or to treat pruritus. Generalized pruritus intractable to antihistamines or antidepressants such as doxepin can be a major problem in PV; interferon α (IFN-α), psoralens with ultraviolet light in the A range (PUVA) therapy, and hydroxyurea are other methods of palliation. Asymptomatic thrombocytosis requires no therapy unless the platelet count is sufficiently high to cause an acquired form of von Willebrand's disease due to adsorption and proteolysis of high-molecular-weight von Willebrand factor (vWF) multimers by the expanded platelet mass. Symptomatic splenomegaly can be treated with IFN-α, although the drug can be associated with significant side effects when used chronically. Pegylated IFN-α produces complete remissions in PV patients, and its role in this disorder may be expanding. Anagrelide, a phosphodiesterase inhibitor, can reduce the platelet count and, if tolerated, is preferable to hydroxyurea because it lacks marrow toxicity and actually is protective against venous thrombosis. A reduction in platelet number may be necessary for the treatment of erythromelalgia or ocular migraine if salicylates are not effective or if the platelet count is sufficiently high to cause a hemorrhagic diathesis but only to the degree that symptoms are alleviated. Alkylating agents and radioactive sodium phosphate (^{32}P) are leukemogenic in PV, and their use should be avoided. If a cytotoxic agent must be used, hydroxyurea is preferred, but this drug does not prevent either thrombosis or myelofibrosis in this disorder, is itself leukemogenic, and should only be used for as short a time as possible. In some patients, massive splenomegaly unresponsive to reduction by therapy and associated with intractable weight loss will require splenectomy. In some patients with end-stage disease, pulmonary hypertension may develop due to fibrosis and extramedullary hematopoiesis. Allogeneic bone marrow transplantation may be curative in young patients. Several *JAK2* inhibitors are undergoing clinical trial; to date, these agents have been demonstrated to alleviate constitutional symptoms and to rapidly reduce spleen size without significant effects on blood counts or the *JAK2* V617F neutrophilallele burden, suggesting that they may at least have an important palliative role.

Most patients with PV can live long lives without functional impairment when their red cell mass is effectively managed with phlebotomy alone. Chemotherapy is never indicated to control the red cell mass unless venous access is inadequate.

PRIMARY MYELOFIBROSIS

Chronic PMF (other designations include idiopathic myelofibrosis, *agnogenic myeloid metaplasia* or

myelofibrosis with myeloid metaplasia) is a clonal disorder of a multipotent hematopoietic progenitor cell of unknown etiology characterized by marrow fibrosis, extramedullary hematopoiesis, and splenomegaly. PMF is the least common chronic MPD, and establishing this diagnosis in the absence of a specific clonal marker is difficult because myelofibrosis and splenomegaly are also features of both PV and CML. Furthermore, myelofibrosis and splenomegaly also occur in a variety of benign and malignant disorders (Table 108–3), many of which are amenable to specific therapies not effective in PMF. In contrast to the other chronic MPDs and so-called acute or malignant myelofibrosis, which can occur at any age, PMF primarily afflicts men in their sixth decade or later.

Table 108–3 Disorders Causing Myelofibrosis

Malignant	Nonmalignant
Acute leukemia (lymphocytic, myelogenous, megakaryocytic)	HIV infection
Chronic myelogenous leukemia	Hyperparathyroidism
Hairy cell leukemia	Renal osteodystrophy
Hodgkin's disease	Systemic lupus erythematosus
Idiopathic myelofibrosis	Tuberculosis
Lymphoma	Vitamin D deficiency
Multiple myeloma	Thorium dioxide exposure
Myelodysplasia	Gray platelet syndrome
Metastatic carcinoma	
Polycythemia vera	
Systemic mastocytosis	

Etiology

The etiology of PMF is unknown. Nonrandom chromosome abnormalities such as 9p, 20q– 13q– trisomy 8 or 9, or partial trisomy 1q are common, but no cytogenetic abnormality specific to the disease has been identified. *JAK2* V617F is present in approximately 50% of PMF patients and mutations in the thrombopoietin receptor *Mpl* occur in about 5%. The degree of myelofibrosis and the extent of extramedullary hematopoiesis are also not related. Fibrosis in this disorder is associated with overproduction of transforming growth factor β and tissue inhibitors of metalloproteinases, while osteosclerosis is associated with overproduction of osteoprotegerin, an osteoclast inhibitor. Marrow angiogenesis occurs due to increased production of vascular endothelial growth factor. Importantly, fibroblasts in PMF are polyclonal and not part of the neoplastic clone.

Clinical Features

No signs or symptoms are specific for PMF. Many patients are asymptomatic at presentation, and the disease is usually detected by the discovery of splenic enlargement and/or abnormal blood counts during a routine examination. However, in contrast to its companion MPDs, night sweats, fatigue, and weight loss are common presenting complaints. A blood smear will show the characteristic features of extramedullary hematopoiesis: teardrop-shaped red cells, nucleated red cells, myelocytes, and promyelocytes; myeloblasts may also be present (Fig. 108-1). Anemia, usually mild initially, is the rule, while the leukocyte and platelet counts are either normal or increased, but either can be depressed. Mild hepatomegaly may accompany the splenomegaly but is unusual in the absence of splenic enlargement; isolated lymphadenopathy should suggest another diagnosis. Both serum lactate dehydrogenase and alkaline phosphatase levels can be elevated. The LAP score can be low, normal, or high. Marrow is usually inaspirable due to the myelofibrosis (Fig. 108-2), and bone x-rays may reveal osteosclerosis. Exuberant extramedullary hematopoiesis can cause ascites; portal, pulmonary, or intracranial hypertension; intestinal or ureteral obstruction; pericardial tamponade; spinal cord compression; or skin nodules. Splenic enlargement can be sufficiently rapid to cause splenic infarction with fever and pleuritic chest pain. Hyperuricemia and secondary gout may ensue.

Figure 108-1

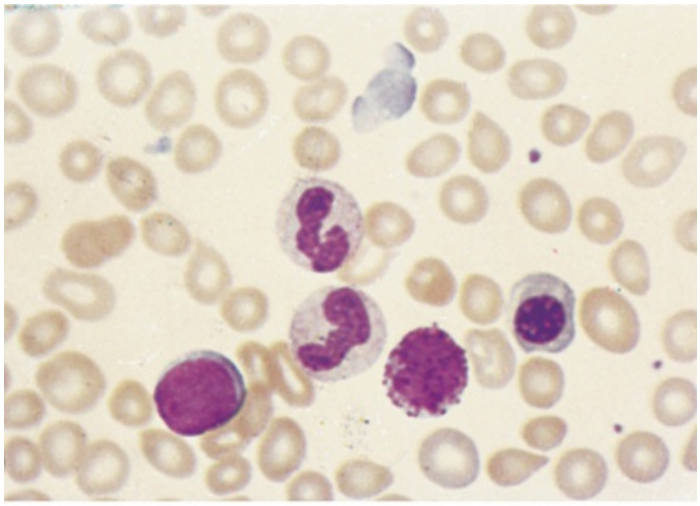

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Teardrop-shaped red blood cells indicative of membrane damage from passage through the spleen, a nucleated red blood cell, and immature myeloid cells indicative of extramedullary hematopoiesis are noted. This peripheral blood smear is related to any cause of extramedullary hematopoiesis.

Figure 108-2

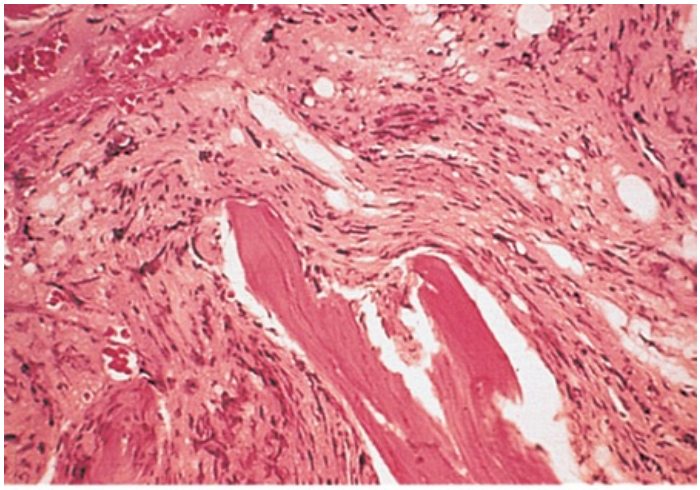

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

This marrow section shows the marrow cavity replaced by fibrous tissue composed of reticulin fibers and collagen. When this fibrosis is due to a primary hematologic process, it is called *myelofibrosis*. When the fibrosis is secondary to a tumor or a granulomatous process, it is called *myelophthisis*.

Diagnosis

While the clinical picture described above is characteristic of PMF, all of the clinical features described can also be observed in PV or CML. Massive splenomegaly commonly masks erythrocytosis in PV, and reports of intraabdominal thromboses in PMF most likely represent instances of unrecognized PV. In some patients with PMF, erythrocytosis has developed during the course of the disease. Furthermore, since many other disorders have features that overlap with PMF but respond to distinctly different therapies, the diagnosis of PMF is one of exclusion, which requires that the disorders listed in Table 108–3 be ruled out. A diagnostic algorithm has been proposed but does not distinguish one disease causing myeloid metaplasia from another.

The presence of teardrop-shaped red cells, nucleated red cells, myelocytes, and promyelocytes establishes the presence of extramedullary hematopoiesis, while the presence of leukocytosis, thrombocytosis with large and bizarre platelets, and circulating myelocytes suggests the presence of an MPD as opposed to a secondary form of myelofibrosis (Table 108–3). Marrow is usually not aspirable due to increased marrow reticulin, but marrow biopsy will reveal a hypercellular marrow with

trilineage hyperplasia and, in particular, increased numbers of megakaryocytes in clusters and with large, dysplastic nuclei. However, there are no characteristic morphologic abnormalities of the bone marrow that distinguish PMF from the other chronic MPDs. Splenomegaly due to extramedullary hematopoiesis may be sufficiently massive to cause portal hypertension and variceal formation. In some patients, exuberant extramedullary hematopoiesis can dominate the clinical picture. An intriguing feature of PMF is the occurrence of autoimmune abnormalities such as immune complexes, antinuclear antibodies, rheumatoid factor, or a positive Coombs' test. Whether these represent a host reaction to the disorder or are involved in its pathogenesis is unknown. Cytogenetic analysis of blood is useful both to exclude CML and for prognostic purposes, because complex karyotype abnormalities portend a poor prognosis in PMF. For unknown reasons, the number of circulating CD34+ cells is markedly increased in PMF (>15,000/μL) compared to the other chronic MPDs, unless they too develop myeloid metaplasia.

Importantly, approximately 50% of PMF patients, like patients with its companion myeloproliferative disorders PV and ET, express the *JAK2* V617F mutation, often as homozygotes. Such patients had a poorer survival in one retrospective study but not in another, where they were found to be older and to have higher hematocrits than those patients who were *JAK2* V617F—negative. Those patients with an *MPL* mutation tend to be more anemic than those without this mutation.

Complications

Survival in PMF varies according to specific clinical features (Table 108–4) but is shorter than in patients with PV or ET. The natural history of PMF is one of increasing marrow failure with transfusion-dependent anemia and increasing organomegaly due to extramedullary hematopoiesis. As with CML, PMF can evolve from a chronic phase to an accelerated phase with constitutional symptoms and increasing marrow failure. About 10% of patients spontaneously transform to an aggressive form of acute leukemia for which therapy is usually ineffective. Important prognostic factors for disease acceleration include anemia, leukocytosis, thrombocytopenia, the presence of circulating myeloblasts, older age, the presence of complex cytogenetic abnormalities, and constitutional symptoms such as unexplained fever, night sweats, or weight loss.

Table 108–4 Risk Stratification for Primary Myelofibrosis

Risk Factors		Frequency of Occurrence (%)	
Age >65 years		45	
Constitutional symptoms		26	
Hemoglobin <10 g/dL		35	
WBC >25 x 109/L		10	
Blood blasts >10%		36	
Risk Groups	**No. of Factors**	**Proportion of Patients (%)**	**Median Survival (years)**
Low	0	22	11
Intermediate-1	1	29	8
Intermediate-2	2	28	4
High	≥3	21	2

Source: From F Cervantes et al: Blood 113:2895, 2009.

Treatment: Primary Myelofibrosis

No specific therapy exists for PMF. Anemia may be due to gastrointestinal blood loss, may be exacerbated by folic acid deficiency, and in rare instances, pyridoxine therapy has been effective. However, anemia is more often due to ineffective erythropoiesis uncompensated by extramedullary hematopoiesis in the spleen and liver. Neither recombinant erythropoietin nor androgens such as Danazol have proved consistently effective as therapy for the anemia. Erythropoietin may worsen splenomegaly and will be ineffective if the serum erythropoietin level is >125 mU/L. A red cell splenic sequestration study can establish the presence of hypersplenism, for which splenectomy is indicated. Splenectomy may also be necessary if splenomegaly impairs alimentation and should be performed before cachexia sets in. In this situation, splenectomy should not be avoided because of concern over rebound thrombocytosis, loss of hematopoietic capacity, or compensatory hepatomegaly. However, for unexplained reasons, splenectomy increases the risk of blastic transformation. Splenic irradiation is, at best, temporarily palliative and associated with a significant risk of neutropenia, infection, and operative hemorrhage. Allopurinol can control significant hyperuricemia, and hydroxyurea has proved useful for controlling organomegaly in some patients. The role of IFN-α is still undefined; its side effects are more pronounced in the older individuals who are usually afflicted with this disorder and it should be used at lower doses. Glucocorticoids have been used to control constitutional symptoms and autoimmune complications and may ameliorate anemia alone, or in combination with low-dose thalidomide (50–100 mg/d); such therapy may also palliate splenomegaly. Allogeneic bone marrow

transplantation is the only curative treatment and should be considered in younger patients; reduced-intensity conditioning regimens may permit hematopoietic cell transplantation to be extended to older individuals. *JAK2* inhibitors have been effective in alleviating constitutional symptoms and splenomegaly in PMF patients and are currently used in phase III clinical trials. While their effects are reversible, these agents may offer a less toxic and more effective means of palliation in this disorder.

ESSENTIAL THROMBOCYTOSIS

Essential thrombocytosis (other designations include *essential thrombocythemia*, *idiopathic thrombocytosis*, *primary thrombocytosis*, and *hemorrhagic thrombocythemia*) is a clonal disorder of unknown etiology involving a multipotent hematopoietic progenitor cell manifested clinically by overproduction of platelets without a definable cause. ET is an uncommon disorder, with an incidence of 1—2/100,000 and a distinct female predominance. No clonal marker is available to consistently distinguish ET from the more common nonclonal, reactive forms of thrombocytosis (Table 108–5), making its diagnosis difficult. Once considered a disease of the elderly and responsible for significant morbidity due to hemorrhage or thrombosis, with the widespread use of electronic cell counters, it is now clear that ET can occur at any age in adults and often without symptoms or disturbances of hemostasis. There is an unexplained female predominance in contrast to PMF or the reactive forms of thrombocytosis where no sex difference exists. Because no specific clonal marker is available, clinical criteria have been proposed to distinguish ET from the other chronic MPDs, which may also present with thrombocytosis but have differing prognoses and therapies (Table 108–5). These criteria do not establish clonality; therefore, they are truly useful only in identifying disorders such as CML, PV, or myelodysplasia, which can masquerade as ET, as opposed to actually establishing the presence of ET. Furthermore, as with "idiopathic" erythrocytosis, nonclonal benign forms of thrombocytosis exist (such as hereditary overproduction of thrombopoietin) that are not widely recognized because we currently lack adequate diagnostic tools. Approximately 50% of ET patients carry the *JAK2* V617F mutation but its absence does not exclude the disorder.

Table 108–5 Causes of Thrombocytosis

Tissue inflammation: collagen vascular disease, inflammatory bowel disease	Hemorrhage
Malignancy	Iron deficiency anemia
Infection	Surgery
Myeloproliferative disorders: polycythemia vera, primary myelofibrosis, essential thrombocytosis, chronic myelogenous leukemia	Rebound: Correction of vitamin B12 or folate deficiency, post-ethanol abuse
Myelodysplastic disorders: 5q-syndrome, idiopathic refractory sideroblastic anemia	Hemolysis
Postsplenectomy or hyposplenism	Familial: Thrombopoietin overproduction, constitutive Mpl activation

Etiology

Megakaryocytopoiesis and platelet production depend upon thrombopoietin and its receptor *Mpl*. As in the case of early erythroid and myeloid progenitor cells, early megakaryocytic progenitors require the presence of interleukin 3 (IL-3) and stem cell factor for optimal proliferation in addition to thrombopoietin. Their subsequent development is also enhanced by the chemokine stromal cell-derived factor 1 (SDF-1). However, megakaryocyte maturation requires thrombopoietin.

Megakaryocytes are unique among hematopoietic progenitor cells because reduplication of their genome is endomitotic rather than mitotic. In the absence of thrombopoietin, endomitotic megakaryocytic reduplication and, by extension, the cytoplasmic development necessary for platelet production are impaired. Like erythropoietin, thrombopoietin is produced in both the liver and the kidneys, and an inverse correlation exists between the platelet count and plasma thrombopoietic activity. Like erythropoietin levels, plasma levels of thrombopoietin are controlled largely by the size of its progenitor cell pool. In contrast to erythropoietin, but like its myeloid counterparts, granulocyte- and granulocyte-macrophage colony-stimulating factors, thrombopoietin not only enhances the proliferation of its target cells but also enhances the reactivity of their end-stage product, the platelet. In addition to its role in thrombopoiesis, thrombopoietin also enhances the survival of multipotent hematopoietic stem cells.

The clonal nature of ET was established by analysis of glucose-6-phosphate dehydrogenase isoenzyme expression in patients hemizygous for this gene, by analysis of X-linked DNA polymorphisms in informative female patients, and by the expression in patients of nonrandom, though variable, cytogenetic abnormalities. Although thrombocytosis is its principal manifestation, like the other

chronic MPDs, a multipotent hematopoietic progenitor cell is involved in ET. Furthermore, a number of families have been described in which ET was inherited, in one instance as an autosomal dominant trait. In addition to ET, PMF and PV have also been observed in some kindreds.

Clinical Features

Clinically, ET is most often identified incidentally when a platelet count is obtained during the course of a routine medical evaluation. Occasionally, review of previous blood counts will reveal that an elevated platelet count was present but overlooked for manyyears. No symptoms or signs are specific for ET, but these patients can have hemorrhagic and thrombotic tendencies expressed as easy bruising for the former and microvascular occlusions for the latter such as erythromelalgia, ocular migraine, or TIAs. Physical examination is generally unremarkable except occasionally for mild splenomegaly. Massive splenomegaly is indicative of another MPD, in particular PV, PMF, or CML.

Anemia is unusual, but a mild neutrophilic leukocytosis is not. The blood smear is most remarkable for the number of platelets present, some of which may be very large. The large mass of circulating platelets may prevent the accurate measurement of serum potassium due to release of platelet potassium upon blood clotting. This type of hyperkalemia is a laboratory artifact and not associated with electrocardiographic abnormalities. Similarly, arterial oxygen measurements can be inaccurate unless thrombocythemic blood is collected on ice. The prothrombin and partial thromboplastin times are normal, while abnormalities of platelet function such as a prolonged bleeding time and impaired platelet aggregation can be present. However, in spite of much study, no platelet function abnormality is characteristic of ET, and no platelet function test predicts the risk of clinically significant bleeding or thrombosis.

The elevated platelet count may hinder marrow aspiration, but marrow biopsy usually reveals megakaryocyte hyperplasia and hypertrophy, as well as an overall increase in marrow cellularity. If marrow reticulin is increased, another diagnosis should be considered. The absence of stainable iron demands an explanation because iron deficiency alone can cause thrombocytosis, and absent marrow iron in the presence of marrow hypercellularity is a feature of PV.

Nonrandom cytogenetic abnormalities occur in ET but are uncommon, and no specific or consistent abnormality is notable, even those involving chromosomes 3 and 1, where the genes for thrombopoietin and its receptor *Mpl*, respectively, are located.

Diagnosis

Thrombocytosis is encountered in a broad variety of clinical disorders (Table 108–5) in many of which production of cytokines is increased. The absolute level of the platelet count is not a useful diagnostic aid for distinguishing between benign and clonal causes of thrombocytosis. About 50% of ET patients express the *JAK2* V617F mutation. When *JAK2* V617F is absent, cytogenetic evaluation is mandatory to determine if the thrombocytosis is due to CML or a myelodysplastic disorder such as the 5q– syndrome. Because the bcr-abl translocation can be present in the absence of the Ph chromosome, and because bcr-abl reverse transcriptasepolymerase chain reaction is associated with false-positive results, fluorescence in situ hybridization (FISH) analysis for bcr-abl is the preferred assay in patients with thrombocytosis in whom a cytogenetic study for the Ph chromosome is negative. Anemia and ringed sideroblasts are not features of ET, but they are features of idiopathic refractory sideroblastic anemia, and in some of these patients the thrombocytosis occurs in association with *JAK2* V617F expression. Massive splenomegaly should suggest the presence of another MPD, and in this setting a red cell mass determination should be performed because splenomegaly can mask the presence of erythrocytosis. Importantly, what appears to be ET can evolve into PV or PMF after a period of many years, revealing the true nature of the underlying MPD. There is sufficient overlap of the *JAK2* V617F neutrophilallele burden between ET and PV that this cannot be used as a distinguishing diagnostic feature; only a red cell mass and plasma volume determination can distinguish PV from ET and importantly in this regard, 64% of *JAK2* V617F—positive ET patients actually were found to have PV when red cell mass and plasma volume determinations were performed.

Complications

Perhaps no other condition in clinical medicine has caused otherwise astute physicians to intervene inappropriately more often than thrombocytosis, particularly if the platelet count is >1 x 10^6/μL. It is commonly believed that a high platelet count causes intravascular stasis and thrombosis; however, no controlled clinical study has ever established this association, and in patients younger than age 60 years, the incidence of thrombosis was not greater in patients with thrombocytosis than in age-matched controls.

To the contrary, very high platelet counts are associated primarily with hemorrhage due to acquired von Willebrand disease. This is not meant to imply that an elevated platelet count cannot cause symptoms in an ET patient, but rather that the focus should be on the patient, not the platelet count. For example, some of the most dramatic neurologic problems in ET are migraine-related and respond only to lowering of the platelet count, while other symptoms such as erythromelalgia respond simply to platelet cyclooxygenase 1 inhibitors such as aspirin or ibuprofen, without a reduction in platelet

number. Still others may represent an interaction between an atherosclerotic vascular system and a high platelet count, and others may have no relationship to the platelet count whatsoever. Recognition that PV can present with thrombocytosis alone as well as the discovery of previously unrecognized causes of hypercoagulability (Chap. 117) make the older literature on the complications of thrombocytosis unreliable.

Treatment: Essential Thrombocytosis

Survival of patients with ET is not different than for the general population. An elevated platelet count in an asymptomatic patient without cardiovascular risk factors requires no therapy. Indeed, before any therapy is initiated in a patient with thrombocytosis, the cause of symptoms must be clearly identified as due to the elevated platelet count. When the platelet count rises above 1 x 10^6/μL, a substantial quantity of high-molecular-weight von Willebrand multimers are removed from the circulation and destroyed by the enlarged platelet mass, resulting in an acquired form of von Willebrand's disease. This can be identified by a reduction in ristocetin cofactor activity. In this situation, aspirin could promote hemorrhage. Bleeding in this situation usually responds to ε-aminocaproic acid, which can be given prophylactically before and after elective surgery. Plateletpheresis is at best a temporary and inefficient remedy that is rarely required. Importantly, ET patients treated with ^{32}P or alkylating agents are at risk of developing acute leukemia without any proof of benefit; combining either therapy with hydroxyurea increases this risk. If platelet reduction is deemed necessary on the basis of symptoms refractory to salicylates alone, IFN-α, the quinazoline derivative, anagrelide, or hydroxyurea can be used to reduce the platelet count, but none of these is uniformly effective nor without significant side effects. Hydroxyurea and aspirin are more effective than anagrelide and aspirin for prevention of TIAs but not more effective for the prevention of other types of arterial thrombosis and are actually less effective for venous thrombosis. The effectiveness of hydroxyurea in preventing TIAs is because it is an NO donor. Normalizing the platelet count also does not prevent either arterial or venous thrombosis. The risk of gastrointestinal bleeding is also higher when aspirin is combined with anagrelide.

As more clinical experience is acquired, ET appears more benign than previously thought. Evolution to acute leukemia is more likely to be a consequence of therapy than of the disease itself. In managing patients with thrombocytosis, the physician's first obligation is to do no harm.

FURTHER READINGS

Alchalby H et al: Impact of JAK2V617F mutation status, allele burden, and clearance after allogeneic stem cell transplantation for myelofibrosis. Blood 116:3572, 2010[PMID: 20489052]

Buss DH et al: The incidence of thrombotic and hemorrhagic disorders in association with extreme thrombocytosis: An analysis of 129 cases. Am J Hematol 20:365, 1985[PMID: 3865532]

Cassinat B et al: Classification of myeloproliferative disorders in the JAK2 era: Is there a role for red cell mass? Leukemia 22:452, 2002

Cervantes F et al: New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. Blood 113:2895, 2009[PMID: 18988864]

Jones AV et al: JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. Nat Genet 41:446, 2009[PMID: 19363185]

Kiladjian JJ et al: Pegylated interferon-alfa-2a induces complete hematologic and molecular responses with low toxicity in polycythemia vera. Blood 112:3065, 2008[PMID: 18650451]

Reilly JT: Idiopathic myelofibrosis: Pathogenesis to treatment. Hematol Oncol 24:56, 2006[PMID: 16477581]

Spivak JL: Polycythemia vera: Myths, mechanisms, and management. Blood 100:4272, 2002[PMID: 12393615]

———: Narrative review: Thrombocytosis, polycythemia vera, and JAK2 mutations: The phenotypic mimicry of chronic myeloproliferation. Ann Intern Med 152:300, 2010

Verstovsek S et al: Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. N Engl J Med 363:1117, 2010[PMID: 20843246]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACUTE AND CHRONIC MYELOID LEUKEMIA: INTRODUCTION

The myeloid leukemias are a heterogeneous group of diseases characterized by infiltration of the blood, bone marrow, and other tissues by neoplastic cells of the hematopoietic system. In 2010 the estimated number of new myeloid leukemia cases in the United States was 17,200. These leukemias comprise a spectrum of malignancies that, untreated, range from rapidly fatal to slowly growing. Based on their untreated course, the myeloid leukemias have traditionally been designated acute or chronic.

ACUTE MYELOID LEUKEMIA

Incidence

The incidence of acute myeloid leukemia (AML) is ~3.5 per 100,000 people per year, and the age-adjusted incidence is higher in men than in women (4.3 vs 2.9). AML incidence increases with age; it is 1.7 in individuals aged <65 years and 15.9 in those aged >65 years. The median age at diagnosis is 67 years.

Etiology

Heredity, radiation, chemical and other occupational exposures, and drugs have been implicated in the development of AML. No direct evidence suggests a viral etiology.

HEREDITY

Certain syndromes with somatic cell chromosome aneuploidy, such as trisomy 21 noted in Down syndrome, are associated with an increased incidence of AML. Inherited diseases with defective DNA repair, e.g., Fanconi anemia, Bloom syndrome, and ataxia-telangiectasia, are also associated with AML. Congenital neutropenia (Kostmann syndrome) is a disease with mutations in the granulocyte colony-stimulating factor (G-CSF) receptor and, often, neutrophil elastase that may evolve into AML. Myeloproliferative syndromes may also evolve into AML (Chap. 108). Germ-line mutations of CCAAT/enhancer-binding protein α (*CEBPA*), runt-related transcription factor 1 (*RUNX1*), and tumor protein p53 (*TP53*) have also been associated with a higher predisposition to AML in some series.

RADIATION

High-dose radiation, like that experienced by survivors of the atomic bombs in Japan or nuclear reactor accidents, increase the risk of myeloid leukemias that peak 5–7 years after exposure. Therapeutic radiation alone seems to add little risk of AML but can increase the risk in people also exposed to alkylating agents.

CHEMICAL AND OTHER EXPOSURES

Exposure to benzene, a solvent used in the chemical, plastic, rubber, and pharmaceutical industries, is associated with an increased incidence of AML. Smoking and exposure to petroleum products, paint, embalming fluids, ethylene oxide, herbicides, and pesticides have also been associated with an increased risk of AML.

DRUGS

Anticancer drugs are the leading cause of therapy-associated AML. Alkylating agent–associated leukemias occur on average 4–6 years after exposure, and affected individuals have aberrations in chromosomes 5 and 7. Topoisomerase II inhibitor–associated leukemias occur 1–3 years after exposure, and affected individuals often have aberrations involving chromosome 11q23. Chloramphenicol, phenylbutazone, and, less commonly, chloroquine and methoxypsoralen can result in bone marrow failure that may evolve into AML.

Classification

The current categorization of AML uses the World Health Organization (WHO) classification (Table 109–1), which includes different biologically distinct groups based on clinical features and cytogenetic and molecular abnormalities in addition to morphology. In contrast to the previously used French-American-British (FAB) schema, the WHO classification places limited reliance on cytochemistry. Since some of the recent literature and some ongoing studies use the FAB classification, a description of this system is also provided in Table 109–1. A major difference between the WHO and FAB systems is the blast cutoff for a diagnosis of AML as opposed to myelodysplastic syndrome (MDS); it is 20% in the WHO classification and 30% in the FAB. AML with 20–30% blasts as defined by the WHO classification can benefit from therapies for MDS (such as decitabine or 5-azacytidine) that were approved by the U.S. Food and Drug Administration (FDA) based on trials using the FAB criteria. Selected components of the WHO classification are outlined below.

Table 109–1 AML Classification Systems

World Health Organization Classification[a]
AML with recurrent genetic abnormalities
AML with t(8;21)(q22;q22); *RUNX1-RUNX1T1*[b]
AML with inv(16)(pl3.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*[b]
Acute promyelocytic leukemia with t(15;17)(q22;q12); *PML-RARA*[b]
AML with t(9;11)(p22;q23); *MLLT3-MLL*
AML with t(6;9)(p23;q34); *DEK-NUP214*
AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *RPN1-EVI1*
AML (megakaryoblastic) with t(1;22)(p13;q13); *RBM15-MKL1*
Provisional entity: AML with mutated NPM1

Provisional entity: AML with mutated CEBPA
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML not otherwise specified
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic and monocytic leukemia
Acute erythroid leukemia
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Transient abnormal myelopoiesis
Myeloid leukemia associated with Down syndrome
Blastic plasmacytoid dendritic cell neoplasm
Acute leukemia of ambiguous lineage
Acute undifferentiated leukemia
Mixed phenotype acute leukemia with t(9;22)(q34;q11,20); *BCR-ABL11*
Mixed phenotype acute leukemia with t(v;11q23); *MLL* rearranged
Mixed phenotype acute leukemia, B/myeloid, NOS
Mixed phenotype acute leukemia, T/myeloid, NOS
Provisional entity: Natural killer (NK)-cell lymphoblastic leukemia/lymphoma
French-American-British (FAB) Classification[c]
MO: Minimally differentiated leukemia
MI: Myeloblastic leukemia without maturation
M2: Myeloblastic leukemia with maturation
M3: Hypergranular promyelocytic leukemia
M4: Myelomonocytic leukemia
M4Eo: Variant: Increase in abnormal marrow eosinophils
M5: Monocytic leukemia
M6: Erythroleukemia (DiGuglielmo's disease)
M7: Megakaryoblastic leukemia

[a] From SH Swerdlow et al (eds): *World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, IARC Press, 2008.

[b] Diagnosis is AML regardless of blast count.

[c] From JM Bennett et al: Ann Intern Med 103:620, 1985.

Abbreviation: AML, acute myeloid leukemia.

IMMUNOPHENOTYPE AND RELEVANCE TO THE WHO CLASSIFICATION

The immunophenotype of human leukemia cells can be studied by multiparameter flow cytometry after the cells are labeled with monoclonal antibodies to cell-surface antigens. This can be important for separating AML from acute lymphoblastic leukemia (ALL) and identifying some types of AML. For example, AML with minimal differentiation that is characterized by immature morphology and no lineage-specific cytochemical reactions may be diagnosed by flow-cytometric demonstration of the myeloid-specific antigens cluster designation (CD) 13 and/or 117. Similarly, acute megakaryoblastic leukemia can often be diagnosed only by expression of the platelet-specific antigens CD41 and/or CD61. While flow cytometry is useful, widely used, and in some cases essential for the diagnosis of AML, it is supportive only in establishing the different subtypes of AML through the WHO classification.

CLINICAL FEATURES AND RELEVANCE TO THE WHO CLASSIFICATION

The WHO classification considers clinical features in subdividing AML. For example, it identifies therapy-related AML as a separate entity that develops following prior therapy (e.g., alkylating agents, topoisomerase II inhibitors, ionizing radiation). It

also identifies AML with myelodysplasia-related changes based in part upon medical history of an antecedent MDS or myelodysplastic/myeloproliferative neoplasm. The clinical features likely contribute to the prognosis of AML and have therefore been included in the classification.

GENETIC FINDINGS AND RELEVANCE TO THE WHO CLASSIFICATION

The WHO classification is the first AML classification to incorporate genetic (chromosomal and molecular) information. Indeed, AML is first subclassified based on the presence or absence of specific recurrent genetic abnormalities. For example, AML FAB M3 is now designated *acute promyelocytic leukemia* (APL), based on the presence of either the t(15;17)(q22;q12) cytogenetic rearrangement or the *PML-RARA* fusion product of the translocation. A similar approach is taken with regard to core binding factor (CBF) AML that is now designated based on the presence of t(8;21)(q22;q22) or inv(16)(p13q22) or the respective fusion products *RUNX1-RUNX1T1* and *CBFB-MYH11*. Thus, the WHO classification separates recurrent cytogenetic and/or molecular types of AML and forces the clinician to take the appropriate steps to correctly identify the entity and thus tailor treatment(s) accordingly.

Chromosomal Analyses

Chromosomal analysis of the leukemic cell provides the most important pretreatment prognostic information in AML. The WHO classification incorporates cytogenetics in the AML classification by recognizing a category of AML with recurrent genetic abnormalities and a category of AML with myelodysplasia-related changes (Table 109–1). The latter category is diagnosed in part by AML with selected myelodysplasia-related cytogenetic abnormalities (e.g., complex karyotypes and unbalanced and balanced changes involving among others, chromosomes 5, 7, and 11). Only one cytogenetic abnormality has been invariably associated with specific morphologic features: t(l5;17)(q22;q12) with APL. Other chromosomal abnormalities have been associated primarily with one morphologic/immunophenotypic group, including inv(16)(p13q22) with AML with abnormal bone marrow eosinophils; t(8;21)(q22;q22) with slender Auer rods, expression of CD19 and increased normal eosinophils; and t(9;11)(p22;q23), and other translocations involving 11q23, with monocytic features. Recurring chromosomal abnormalities in AML may also be associated with specific clinical characteristics. More commonly associated with younger age are t(8;21) and t(l5;17), and with older age, del(5q) and del(7q). Myeloid sarcomas (see below) are associated with t(8;21), and disseminated intravascular coagulation (DIC) with t(15;17).

Molecular Classification

Molecular study of many recurring cytogenetic abnormalities has revealed genes that may be involved in leukemogenesis; this information is increasingly being incorporated into the WHO classification. For instance, t(15;17) results in the fusion gene *PML-RARA* that encodes a chimeric protein, promyelocytic leukemia (Pml)–retinoic acid receptor ∝ (Rar∝), which is formed by the fusion of the retinoic acid receptor ∝ (*RARA*) gene from chromosome 17, and the promyelocytic leukemia (*PML*) gene from chromosome 15. The *RARA* gene encodes a member of the nuclear hormone receptor family of transcription factors. After binding retinoic acid, *RARA* can promote expression of a variety of genes. The 15;17 translocation juxtaposes *PML* with *RARA* in a head-to-tail configuration that is under the transcriptional control of *PML*. Three different breakpoints in the *PML* gene lead to various fusion protein isoforms. The Pml-Rar∝ fusion protein tends to suppress gene transcription and blocks differentiation of the cells. Pharmacologic doses of the Rar∝ ligand, all-*trans*-retinoic acid (tretinoin), relieve the block and promote hematopoietic cell differentiation (see below). Similar examples of molecular subtypes of the disease included in the category of AML with recurrent genetic abnormalities are those characterized by the leukemogenic fusion genes *RUNX1-RUNX1T1*, *CBFB-MYH11*, *MLLT3-MLL*, and *DEK-NUP214* resulting, respectively, from t(8;21), inv(16), t(9;11), and t(6;9)(p23;q34).

Two new provisional entities defined by the presence of gene mutations, rather than macroscopic chromosomal abnormalities, have been recently added to the category of AML with recurrent genetic abnormalities: *AML with mutated nucleophosmin* (*NPM1*) and *AML with mutatedCEBPA*. AML with fms-related tyrosine kinase 3 (*FLT3*) mutations is not considered a distinct entity, although determining the presence of such mutations is recommended by WHO in patients with cytogenetically normal AML (CN-AML) because the relatively frequent *FLT3*-internal tandem duplication (ITD) carries a negative prognostic significance and therefore is clinically relevant (Table 109–2). *FLT3* encodes a tyrosine kinase receptor important in the development of myeloid and lymphoid lineages. Activating mutations of *FLT3* are present in ~30% of adult AML patients due to ITD in the juxtamembrane domain or mutations of the activating loop of the kinase. Continuous activation of the *FLT3*-encoded protein provides increased proliferation and antiapoptotic signals to the myeloid progenitor cell. *FLT3*-ITD, the more common of the *FLT3* mutations, occurs preferentially in patients with CN-AML. The importance of identifying *FLT3*-ITD at diagnosis relates to the fact that it not only is useful in prognostication but also may predict response to specific treatment such as the tyrosine kinase inhibitors that are being tested in clinical trials.

Table 109–2 Molecular Prognostic Markers in AML

Marker	Marker Location	Prognostic Impact
NPM1 mutation	5q35	Favorable
CEBPA mutation	19q13.1	Favorable
FLT3-ITD	13q12	Adverse
WT1 mutation	11p13	Adverse
KIT mutation	4q11-q12	Adverse
BAALC overexpression	8q22.3	Adverse
ERG overexpression	21q22.3	Adverse
MN1 overexpression	22q12.1	Adverse
EVI1 overexpression	3q26	Adverse

Abbreviations: AML, acute myeloid leukemia; ITD, internal tandem duplication.

Other molecular prognostic factors (Table 109–2) in AML include v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (KIT) mutations that are found in 25–30% of t(8;21) or inv(16) patients. Others include Wilms' tumor 1 (WT1) mutations found in 10–13% of CN-AML and overexpression of genes such as brain and acute leukemia, cytoplasmic (BAALC), ets erythroblastosis virus E26 oncogene homologue (avian) (erg), meningioma (disrupted in balanced translocation) 1 (MN1), and MDS1 and EVI1 complex locus (MECOM, also known as EVI1), which predict for poor outcome in CN-AML. The applicability of screening for these molecular aberrations to AML classification and clinical practice is being tested.

With progress in genomics technology including genomewide investigation of gene mutations and expression levels, additional aberrations are being discovered, underscoring the molecular heterogeneity of AML. The applicability of gene expression profiling to diagnosis and outcome prediction of cytogenetic and molecular subsets of AML patients and to clinical management of AML is under active investigation. MicroRNAs, naturally occurring noncoding RNAs, have been shown to regulate the

expression of proteins involved in hematopoietic differentiation and survival pathways by degradation or translation inhibition of target coding RNAs. Deregulated expression levels of microRNAs have been shown to associate with specific cytogenetic and molecular subsets of AML and predict outcome in CN-AML. Finally, massive parallel sequencing of the whole genome from AML patients' blasts is revealing previously unrecognized mutations of genes that are involved in metabolic pathways that have not been previously hypothesized to be disrupted in AML, such as mutations in the isocitrate dehydrogenase 1 (NADP+), soluble (*IDH1*) and isocitrate dehydrogenase 2 (NADP+), and mitochondrial (*IDH2*) genes.

It is likely that once the biologic and clinical significance of these emerging genetic aberrations is understood, AML will be primarily classified molecularly to define specific entities and stratify patients to a corresponding, optimal targeting therapy.

Clinical Presentation

SYMPTOMS

Patients with AML most often present with nonspecific symptoms that begin gradually or abruptly and are the consequence of anemia, leukocytosis, leukopenia or leukocyte dysfunction, or thrombocytopenia. Nearly half have had symptoms for ≤3 months before the leukemia was diagnosed.

Half mention fatigue as the first symptom, but most complain of fatigue or weakness at the time of diagnosis. Anorexia and weight loss are common. Fever with or without an identifiable infection is the initial symptom in ~10% of patients. Signs of abnormal hemostasis (bleeding, easy bruising) are noted first in 5% of patients. On occasion, bone pain, lymphadenopathy, nonspecific cough, headache, or diaphoresis is the presenting symptom.

Rarely patients may present with symptoms from a myeloid sarcoma that is a tumor mass consisting of myeloid blasts occurring at anatomic sites other than bone marrow. Sites involved are most commonly the skin, lymph node, gastrointestinal tract, soft tissue, and testis. This rare presentation, often characterized by chromosome aberrations [e.g., monosomy 7, trisomy 8, *MLL* rearrangement, inv(16), trisomy 4, t(8;21)] may precede or coincide with AML.

PHYSICAL FINDINGS

Fever, splenomegaly, hepatomegaly, lymphadenopathy, sternal tenderness, and evidence of infection and hemorrhage are often found at diagnosis. Significant gastrointestinal bleeding, intrapulmonary hemorrhage, or intracranial hemorrhage occur most often in APL. Bleeding associated with coagulopathy may also occur in monocytic AML and with extreme degrees of leukocytosis or thrombocytopenia in other morphologic subtypes. Retinal hemorrhages are detected in 15% of patients. Infiltration of the gingivae, skin, soft tissues, or the meninges with leukemic blasts at diagnosis is characteristic of the monocytic subtypes and those with 11q23 chromosomal abnormalities.

HEMATOLOGIC FINDINGS

Anemia is usually present at diagnosis and can be severe. The degree varies considerably, irrespective of other hematologic findings, splenomegaly, or duration of symptoms. The anemia is usually normocytic normochromic. Decreased erythropoiesis often results in a reduced reticulocyte count, and red blood cell (RBC) survival is decreased by accelerated destruction. Active blood loss also contributes to the anemia.

The median presenting leukocyte count is about 15,000/μL. Between 25 and 40% of patients have counts <5000/μL, and 20% have counts >100,000/μL. Fewer than 5% have no detectable leukemic cells in the blood. The morphology of the malignant cell varies in different subsets. In AML, the cytoplasm often contains primary (nonspecific) granules, and the nucleus shows fine, lacy chromatin with one or more nucleoli characteristic of immature cells. Abnormal rod-shaped granules called Auer rods are not uniformly present, but when they are, myeloid lineage is virtually certain (Fig. 109-1). Poor neutrophil function may be noted functionally by impaired phagocytosis and migration and morphologically by abnormal lobulation and deficient granulation.

Figure 109-1

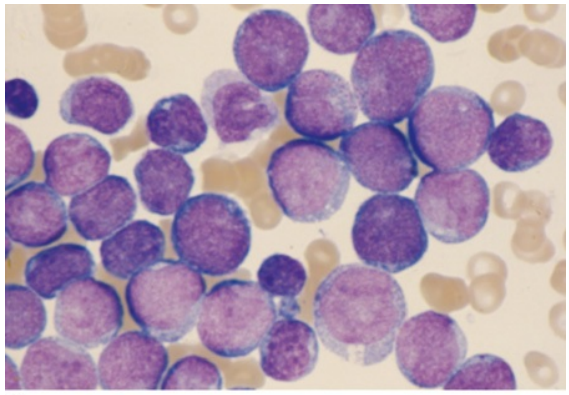

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

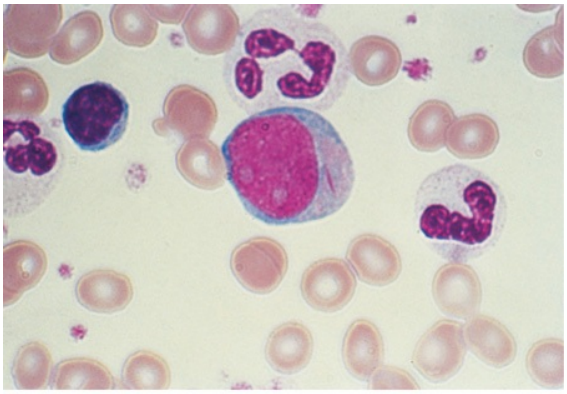

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

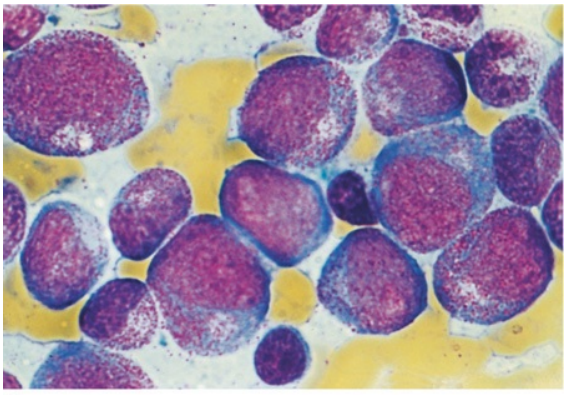

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

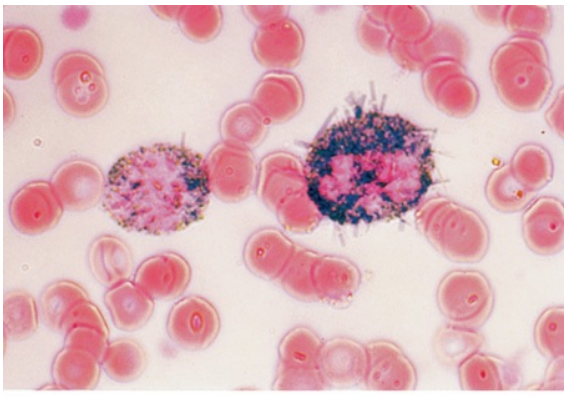

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Morphology of AML cells. ***A.*** Uniform population of primitive myeloblasts with immature chromatin, nucleoli in some cells, and primary cytoplasmic granules. ***B.*** Leukemic myeloblast containing an Auer rod. ***C.*** Promyelocytic leukemia cells with prominent cytoplasmic primary granules. ***D.*** Peroxidase stain shows dark blue color characteristic of peroxidase in granules in AML.

Platelet counts <100,000/μL are found at diagnosis in ~75% of patients, and about 25% have counts <25,000/μL. Both morphologic and functional platelet abnormalities can be observed, including large and bizarre shapes with abnormal granulation and inability of platelets to aggregate or adhere normally to one another.

PRETREATMENT EVALUATION

Once the diagnosis of AML is suspected, a rapid evaluation and initiation of appropriate therapy should follow (Table 109–3). In addition to clarifying the subtype of leukemia, initial studies should evaluate the overall functional integrity of the major organ systems, including the cardiovascular, pulmonary, hepatic, and renal systems. Factors that have prognostic significance, either for achieving complete remission (CR) or for predicting the duration of CR, should also be assessed before initiating treatment, including cytogenetics and molecular markers (at least *NMP1* and *CEBPA* mutations and *FLT3*-ITD in CN-AML). Leukemic cells should be obtained from all patients and cryopreserved for future use as new tests and therapeutics become available. All patients should be evaluated for infection.

Table 109–3 Initial Diagnostic Evaluation and Management of Adult Patients with AML

History
Increasing fatigue or decreased exercise tolerance (anemia)
Excess bleeding or bleeding from unusual sites (DIC, thrombocytopenia)

Fevers or recurrent infections (granulocytopenia)
Headache, vision changes, nonfocal neurologic abnormalities (CNS leukemia or bleed)
Early satiety (splenomegaly)
Family history of AML (Fanconi, Bloom, or Kostmann syndromes or ataxia-telangiectasia)
History of cancer (exposure to alkylating agents, radiation, topoisomerase II inhibitors)
Occupational exposures (radiation, benzene, petroleum products, paint, smoking, pesticides)
Physical Examination
Performance status (prognostic factor)
Ecchymosis and oozing from IV sites (DIC, possible acute promyelocytic leukemia)
Fever and tachycardia (signs of infection)
Papilledema, retinal infiltrates, cranial nerve abnormalities (CNS leukemia)
Poor dentition, dental abscesses
Gum hypertrophy (leukemic infiltration, most common in monocytic leukemia)
Skin infiltration or nodules (leukemia infiltration, most common in monocytic leukemia)
Lymphadenopathy, splenomegaly, hepatomegaly
Back pain, lower extremity weakness [spinal granulocytic sarcoma, most likely in t(8;21) patients]
Laboratory and Radiologic Studies
CBC with manual differential cell count
Chemistry tests (electrolytes, creatinine, BUN, calcium, phosphorus, uric acid, hepatic enzymes, bilirubin, LDH, amylase, lipase)
Clotting studies (prothrombin time, partial thromboplastin time, fibrinogen, D-dimer)
Viral serologies (CMV, HSV-1, varicella-zoster)
RBC type and screen
HLA typing for potential allogeneic HSCT
Bone marrow aspirate and biopsy (morphology, cytogenetics, flow cytometry, molecular studies for *NPM1* and *CEBPA* mutations and *FLT3*-ITD)
Cryopreservation of viable leukemia cells
Echocardiogram or heart scan
PA and lateral chest radiograph
Placement of central venous access device
Interventions for Specific Patients
Dental evaluation (for those with poor dentition)
Lumbar puncture (for those with symptoms of CNS involvement)
Screening spine MRI (for patients with back pain, lower extremity weakness, paresthesias)
Social work referral for patient and family psychosocial support
Counseling for All Patients
Provide patient with information regarding their disease, financial counseling, and support group contacts.

Abbreviations: AML, acute myeloid leukemia; BUN, blood urea nitrogen; CBC, complete blood count; CMV, cytomegalovirus; CNS, central nervous system; DIC, disseminated intravascular coagulation; HLA, human leukocyte antigen; HSCT, hematopoietic stem cell transplant; HSV, herpes simplex virus; LDH, lactate dehydrogenase; PA, posteroanterior; RBC, red blood (cell) count.

Most patients are anemic and thrombocytopenic at presentation. Replacement of the appropriate blood components, if necessary, should begin promptly. Because qualitative platelet dysfunction or the presence of an infection may increase the likelihood of bleeding, evidence of hemorrhage justifies the immediate use of platelet transfusion, even if the platelet count is only moderately decreased.

About 50% of patients have a mild to moderate elevation of serum uric acid at presentation. Only 10% have marked elevations, but renal precipitation of uric acid and the nephropathy that may result is a serious but uncommon complication. The initiation of chemotherapy may aggravate hyperuricemia, and patients are usually started immediately on allopurinol and hydration at diagnosis. Rasburicase (recombinant uric oxidase) is also useful for treating uric acid nephropathy and often can normalize the serum uric acid level within hours with a single dose of treatment. The presence of high concentrations of lysozyme, a marker for monocytic differentiation, may be etiologic in renal tubular dysfunction, which could worsen other renal problems that arise during the initial phases of therapy.

Prognostic Factors

Many factors influence the likelihood of entering CR, the length of CR, and the curability of AML. CR is defined after examination of both blood and bone marrow. The blood neutrophil count must be ≥1000/μL and the platelet count ≥100,000/μL. Hemoglobin concentration is not considered in determining CR. Circulating blasts should be absent. While rare blasts may be detected in the blood during marrow regeneration, they should disappear on successive studies. The bone marrow should contain <5% blasts, and Auer rods should be absent. Extramedullary leukemia should not be present.

For patients in morphologic CR, immunophenotyping to detect minute populations of blasts, reverse transcriptasepolymerase chain reaction (RT-PCR) to detect AML-associated molecular abnormalities, and either metaphase cytogenetics or interphase cytogenetics by fluorescence in situ hybridization (FISH) to detect AML-associated cytogenetic aberrations are currently being investigated to assess whether residual disease that has clinical significance is present following treatment. Detection of minimal residual disease may become a reliable discriminator between patients in CR who do or do not require additional and/or alternative therapies. In APL, detection of the *PML-RARA* fusion gene transcript by RT-PCR in bone marrow and/or blood during CR predicts relapse, and this assay is being routinely used in the clinic to anticipate clinical relapse and initiate timely salvage treatment. In other types of AML, the clinical relevance of minimal residual disease requires further investigation.

Age at diagnosis is among the most important risk factors. Advancing age is associated with a poorer prognosis, in part because of its influence on the patient's ability to survive induction therapy. Age also influences outcome because AML in older patients differs biologically. The leukemic cells in elderly patients more commonly express the multidrug resistance 1 (MDR1) efflux pump that conveys resistance to natural product–derived agents such as the anthracyclines (see below). With each successive

decade of age, a greater proportion of patients have more resistant disease. Chronic and intercurrent diseases impair tolerance to rigorous therapy; acute medical problems at diagnosis reduce the likelihood of survival. Performance status, independent of age, also influences ability to survive induction therapy and thus respond to treatment.

A prolonged symptomatic interval with cytopenias preceding diagnosis or a history of an antecedent hematologic disorder is another pretreatment clinical feature associated with a lower CR rate and shorter survival time. The CR rate is lower in patients who have had anemia, leukopenia, and/or thrombocytopenia for >3 months before the diagnosis of AML when compared to those without such a history. Responsiveness to chemotherapy declines as the duration of the antecedent disorder(s) increases. AML developing after treatment with cytotoxic agents for other malignancies is usually difficult to treat successfully.

A high presenting leukocyte count in some series is an independent prognostic factor for attaining a CR. Among patients with hyperleukocytosis (>100,000/μL), early central nervous system bleeding and pulmonary leukostasis contribute to poor outcome with initial therapy.

Chromosome findings at diagnosis are currently the most important independent prognostic factor. Patients with t(15;17) have a very good prognosis (approximately 85% cured), and those with t(8;21) and inv(16) a good prognosis (approximately 55% cured), while those with no cytogenetic abnormality have a moderately favorable outcome (approximately 40% cured). Patients with a complex karyotype, t(6;9), inv(3), or -7 have a very poor prognosis.

For those patients lacking prognostic cytogenetic abnormalities, such as those with CN-AML, outcome prediction utilizes molecular genetic abnormalities. *NPM1* mutations without concurrent presence of *FLT3*-ITD, and *CEBPA* mutations, especially if concurrently present in two different alleles, have been shown to predict favorable outcome, whereas *FLT3*-ITD predicts poor outcome. Given the prognostic importance of *NPM1* and *CEBPA* mutations and *FLT3*-ITD, molecular assessment of these genes at diagnosis have been incorporated in AML management guidelines by the National Comprehensive Cancer Network (NCCN) and the European Leukemia Net (ELN). Other molecular aberrations (Table 109–2) may in the future be utilized for prognostication.

In addition to pretreatment variables such as age, leukocyte count, and cytogenetics and/or molecular genetic aberrations, several treatment factors correlate with prognosis in AML, including, most importantly, achievement of CR. In addition, patients who achieve CR after one induction cycle have longer CR durations than those requiring multiple cycles.

Treatment: Acute Myeloid Leukemia

Treatment of the newly diagnosed patient with AML is usually divided into two phases, induction and postremission management (Fig. 109-2). The initial goal is to quickly induce CR. Once CR is obtained, further therapy must be used to prolong survival and achieve cure. The initial induction treatment and subsequent postremission therapy are often chosen based on the patient's age. Intensifying therapy with traditional chemotherapy agents such as cytarabine and anthracyclines in younger patients (<60 years) appears to increase the cure rate of AML. In older patients the benefit of intensive therapy is controversial; novel therapies are being pursued.

Figure 109-2

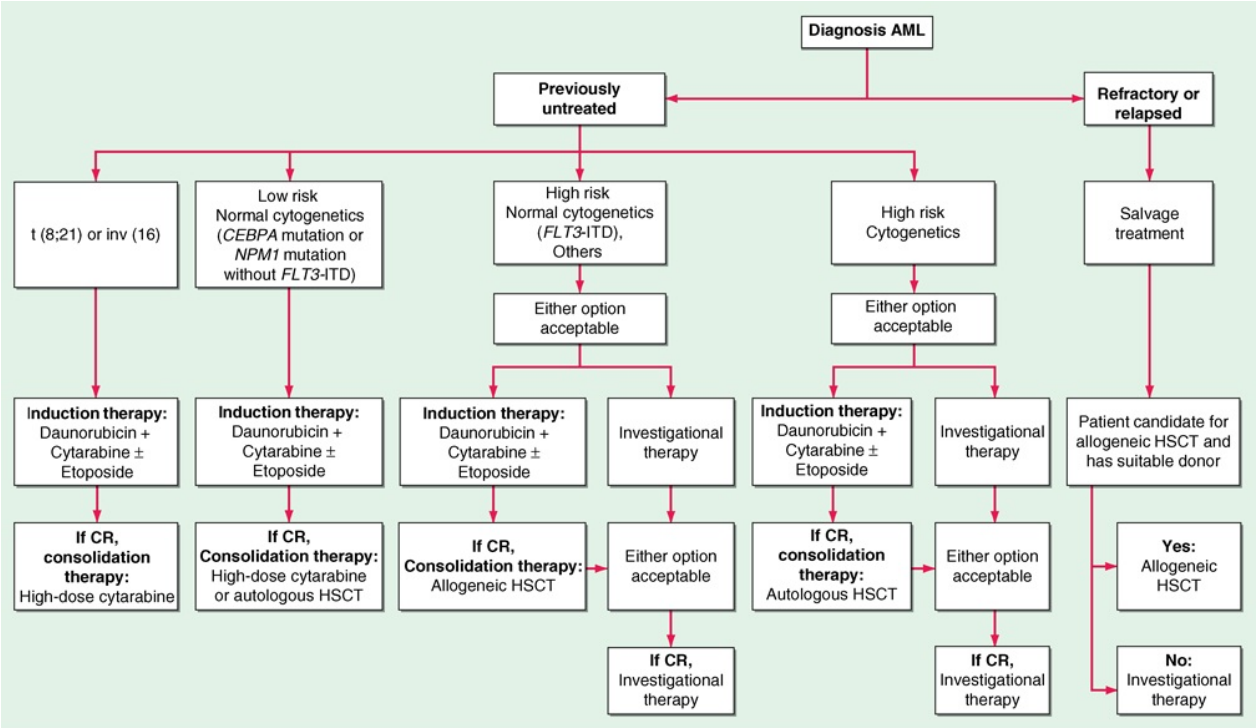

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flow chart for the therapy of newly diagnosed acute myeloid leukemia. For all forms of AML except acute promyelocytic leukemia (APL), standard therapy includes a 7-day continuous infusion of cytarabine (100–200 mg/m^2 per day) and a 3-day course of daunorubicin (60–90 mg/m^{2}per day) with or without 3 days of etoposide (only with daunorubicin 60 mg/m^2 per day) or novel therapies based on their predicted risk of relapse (i.e., risk-stratified therapy). Idarubicin (12–13 mg/m^2 per day) could be used in place of daunorubicin (not shown). Patients who achieve complete -remission undergo postremission consolidation therapy, including sequential courses of high-dose cytarabine, autologous hematopoietic stem cell transplant (HSCT), allogeneic HSCT, or novel therapies, based on their predicted risk of relapse (i.e., risk-stratified therapy). Patients with APL (see text for treatment) -usually receive tretinoin together with anthracycline-based chemotherapy for remission induction and then arsenic trioxide followed by consolidation with anthracycline-based chemotherapy and possibly maintenance with tretinoin. The role of cytarabine in APL induction and consolidation is controversial.

INDUCTION CHEMOTHERAPY

The most commonly used CR induction regimens (for patients other than those with APL) consist of combination chemotherapy with cytarabine and an anthracycline. Cytarabine is a cell cycle S-phase–specific antimetabolite that becomes phosphorylated intracellularly to an active triphosphate form that interferes with DNA synthesis. Anthracyclines are DNA intercalators. Their primary mode of action is thought to be inhibition of topoisomerase II, leading to DNA breaks. Cytarabine is usually administered as a continuous intravenous infusion for 7 days. Anthracycline therapy generally consists of daunorubicin intravenously on days 1, 2, and 3 (the 7 and 3 regimen). Treatment with idarubicin for 3 days in conjunction with cytarabine by 7-day continuous infusion is at least as effective as daunorubicin in younger patients. The addition of etoposide may improve the CR duration. When combined with cytarabine in a 7 and 3 regimen, a higher dose of anthracycline (i.e., daunorubicin 90 mg/m^2) improves outcome compared with a lower dose (i.e., daunorubicin 45 mg/m^2).

After induction chemotherapy, if persistence of leukemia is documented, the patient is usually re-treated with cytarabine and an anthracycline in doses similar to those given initially, but for 5 and 2 days, respectively. Our recommendation, however, is to consider changing therapy in this setting.

With the 7 and 3 cytarabine/daunorubicin regimen outlined above, 65–75% of adults with de novo AML younger than age 60 years achieve CR. Two-thirds achieve CR after a single course of therapy, and one-third require two courses. About 50% of patients who do not achieve CR have a drug-resistant leukemia, and 50% do not achieve CR because of fatal complications of bone marrow aplasia or impaired recovery of normal stem cells. Higher induction treatment–related mortality rate and frequency of resistant disease have been observed with increasing age and in patients with prior hematologic disorders (MDS or myeloproliferative syndromes) or chemotherapy treatment for another malignancy.

Patients who fail to attain CR after two induction courses should be treated with an allogeneic hematopoietic stem cell transplant (HSCT) if an appropriate donor exists. Whether achievement of cytoreduction of disease burden with a salvage treatment should be attempted before a patient with refractory disease after two induction courses can proceed to HSCT is controversial.

High-dose cytarabine-based regimens have high CR rates after a single cycle of therapy. When given in high doses, more cytarabine may enter the cells, saturate the cytarabine-inactivating enzymes, and increase the intracellular levels of 1-β-D-arabinofuranylcytosine-triphosphate, the active metabolite incorporated into DNA. Thus, higher doses of cytarabine may increase the inhibition of DNA synthesis and thereby overcome resistance to standard-dose cytarabine. In two randomized studies, high-dose cytarabine with an anthracycline produced CR rates similar to those achieved with standard 7 and 3 regimens. However, the CR duration was longer after high-dose cytarabine than after standard-dose cytarabine.

The hematologic toxicity of high-dose cytarabine-based induction regimens has typically been greater than that associated with 7 and 3 regimens. Toxicity with high-dose cytarabine includes myelosuppression, pulmonary toxicity, and significant and occasionally irreversible cerebellar toxicity. All patients treated with high-dose cytarabine must be closely monitored for cerebellar toxicity. Full cerebellar testing should be performed before each dose, and further high-dose cytarabine should be withheld if evidence of cerebellar toxicity develops. This toxicity occurs more commonly in patients with renal impairment and in those older than age 60 years. The increased toxicity observed with high-dose cytarabine has limited the use of this therapy in elderly AML patients.

Because of the negative impact of age on outcome when treatment with conventional chemotherapy is administered, clinical trials in elderly patients have focused upon new agents or alternative approaches such as reduced-intensity allogeneic HSCT. Among these, one promising therapy is decitabine, a nucleoside analogue that inhibits DNA methyltransferase, reverses aberrant DNA methylation, and subsequently induces transcription of otherwise silenced tumor suppressor genes in AML cells. Interestingly, this effect on inhibiting DNA methyltransferase occurs at a much lower dose than previously used with this agent to produce a cytotoxic effect in AML. Low-dose decitabine yields complete responses in older patients with AML, including those with unfavorable karyotypes. Other agents with relatively favorable toxicity profiles such as clofarabine have activity in older patients with AML.

SUPPORTIVE CARE

Measures geared to supporting patients through several weeks of granulocytopenia and thrombocytopenia are critical to the success of AML therapy. Patients with AML should be treated in centers expert in providing supportive measures.

Multilumen right atrial catheters should be inserted as soon as patients with newly diagnosed AML have been stabilized. They should be used thereafter for administration of intravenous medications and transfusions, as well as for blood drawing.

Adequate and prompt blood bank support is critical to therapy of AML. Platelet transfusions should be given as needed to maintain a platelet count ≥10,000/μL. The platelet count should be kept at higher levels in febrile patients and during episodes of active bleeding or DIC. Patients with poor posttransfusion platelet count increments may benefit from administration of platelets from human leukocyte antigen (HLA)–matched donors. RBC transfusions should be administered to keep the hemoglobin level >80 g/L (8 g/dL) in the absence of active bleeding, DIC, or congestive heart failure which require higher hemoglobin levels. Blood products leukodepleted by filtration should be used to avert or delay alloimmunization as well as febrile reactions. Blood products should also be irradiated to prevent transfusion associated graft-versus-host disease (GVHD). Cytomegalovirus (CMV)-negative blood products should be used for CMV-seronegative patients who are potential candidates for allogeneic HSCT. Leukodepleted products are also effective for these patients if CMV-negative products are not available.

Infectious complications remain the major cause of morbidity and death during induction and postremission chemotherapy for AML. Antibacterial (i.e., quinolones) and antifungal (e.g, fluconazole, posaconazole) prophylaxis in the absence of fever is likely to be beneficial. For patients who are herpes simplex virus–or varicella zoster–seropositive, antiviral prophylaxis should be initiated.

Fever develops in most patients with AML, but infections are documented in only half of febrile patients. Early initiation of empirical broad-spectrum antibacterial and antifungal antibiotics has significantly reduced the number of patients dying of infectious complications (Chap. 86). An antibiotic regimen adequate to treat gram-negative organisms should be instituted at the onset of fever in a granulocytopenic patient after clinical evaluation, including a detailed physical examination with inspection of the indwelling catheter exit site and a perirectal examination, as well as procurement of cultures and radiographs aimed at documenting the source of fever. Specific antibiotic regimens should be based on antibiotic sensitivity data obtained from the institution at which the patient is being treated. Acceptable regimens for empiric antibiotic therapy include monotherapy with imipenem-cilastin, meropenem, piperacillin/tazobactam, or an extended-spectrum antipseudomonal cephalosporin (cefepime or ceftazidime); an aminoglycoside in combination with an antipseudomonal penicillin (e.g., piperacillin); an aminoglycoside in combination with an extended-spectrum antipseudomonal cephalosporin; and ciprofloxacin in combination with an antipseudomonal penicillin. Aminoglycosides should be avoided if possible in patients with renal insufficiency. Empirical vancomycin should be initiated in neutropenic patients with catheter-related infections, blood cultures positive for gram-positive bacteria before final identification and susceptibility testing, hypotension or shock, and increased risk for viridans group streptococcal bacteremia.

Caspofungin (or similar echinocandin) or liposomal amphotericin B should be considered for antifungal treatment if fever persists 4–7 days following initiation of empiric antibiotic therapy in a patient who has received fluconazole prophylaxis. Voriconazole has also been shown to be equivalent in efficacy and less toxic than amphotericin B. Antibacterial and antifungal antibiotics should be continued until patients are no longer neutropenic, regardless of whether a specific source has been found for the fever.

Recombinant hematopoietic growth factors have been incorporated into clinical trials in AML. These trials have been designed to lower the infection rate after chemotherapy. Both G-CSF and granulocyte-macrophage colony-stimulating factor (GM-CSF) have reduced the median time to neutrophil recovery. This accelerated rate of neutrophil recovery, however, has not generally translated into significant reductions in infection rates or shortened hospitalizations. In most randomized studies, both G-CSF and GM-CSF have failed to improve the CR rate, disease-free survival, or overall survival. Although receptors for both G-CSF and GM-CSF are present on AML blasts, therapeutic efficacy is neither enhanced nor inhibited by these agents. The use of growth factors as supportive care for AML patients is controversial. We favor their use in elderly patients with complicated courses, those receiving intensive postremission regimens, patients with uncontrolled infections, or those participating in clinical trials.

TREATMENT OF PROMYELOCYTIC LEUKEMIA

Tretinoin is an oral drug that induces the differentiation of leukemic cells bearing the t(15;17). APL is responsive to cytarabine and daunorubicin, but about 10% of patients treated with these drugs die from DIC induced by the release of granule components by dying tumor cells. Tretinoin does not produce DIC but produces another complication called the APL differentiation syndrome. Occurring within the first 3 weeks of treatment, it is characterized by fever, fluid retention, dyspnea, chest pain, pulmonary infiltrates, pleural and pericardial effusions, and hypoxemia. The syndrome is related to adhesion of differentiated neoplastic cells to the pulmonary vasculature endothelium. Glucocorticoids, chemotherapy, and/or supportive measures can be effective for management of the APL differentiation syndrome. Temporary discontinuation of tretinoin is necessary in cases of severe APL differentiation syndrome (i.e., patients developing renal failure or requiring admission to the intensive care unit due to respiratory distress). The mortality rate of this syndrome is about 10%.

Tretinoin (45 mg/m^2 per day orally until remission is documented) plus concurrent anthracycline-based chemotherapy appears to be among the most effective treatments for APL leading to CR rates of 90–95%. The addition of cytarabine, although not demonstrated to increase the CR rate, seemingly decreases the risk for relapse. Following achievement of CR, patients should receive at least two cycles of anthracycline-based chemotherapy.

Given the progress made in APL resulting in high cure rates, the goals are to identify patients with very low risk of relapse where attempts are being made to decrease the amount of therapy administered and to identify patients at greatest risk of relapse in order to develop new approaches to increase cure.

Arsenic trioxide has significant antileukemic activity and is being explored as part of initial treatment in clinical trials of APL. In a randomized trial, arsenic trioxide improved outcome if utilized after achievement of CR and before consolidation therapy with anthracycline-based chemotherapy. Additionally, studies combining arsenic trioxide with tretinoin in the absence of chemotherapy are ongoing and preliminarily have shown promise in those patients "unfit" to receive chemotherapy. Furthermore, combinations of arsenic trioxide, tretinoin, and/or chemotherapy and/or gemtuzumab ozogamicin, a monoclonal CD33 antibody linked to the cytotoxic agent calicheamicin, have shown favorable response in high-risk APL patients (i.e., those presenting with a leukocyte count ≥10,000/μL) at diagnosis. Patients receiving arsenic trioxide are at risk of APL differentiation syndrome, especially when it is administered during induction or salvage treatment after disease relapse. In addition, arsenic trioxide may prolong the QT interval, increasing the risk of cardiac arrhythmias.

Assessment of residual disease by RT-PCR amplification of the t(l5;17) chimeric gene product *PML-RARA* following the final cycle of chemotherapy is an important step in the management of APL patients. Disappearance of the signal is associated with long-term disease-free survival; its persistence documented by two consecutive tests performed 2 weeks apart invariably predicts relapse. Sequential monitoring of RT-PCR for t(15;17) is now considered standard for postremission monitoring of APL.

Patients who continue in molecular remission may benefit from maintenance therapy with tretinoin. Patients in molecular, cytogenetic, or clinical relapse should be salvaged with arsenic trioxide; it produces meaningful responses in up to 85% of patients and can be followed by HSCT.

POSTREMISSION THERAPY

Induction of a durable first CR is critical to long-term disease-free survival in AML. However, without further therapy, virtually all patients experience relapse. Once relapse has occurred, AML is generally curable only by HSCT.

Postremission therapy is designed to eradicate residual leukemic cells to prevent relapse and prolong survival. Postremission therapy in AML is often based on age (younger than ages 55–65 years and older than ages 55–65 years). For younger patients, most studies include intensive chemotherapy and allogeneic or autologous HSCT. High-dose cytarabine is more effective than standard-dose cytarabine. The Cancer and Leukemia Group B (CALGB), for example, compared the duration of CR in patients randomly assigned postremission to four cycles of high (3 g/m^2, every 12 hours on days 1, 3, and 5), intermediate (400 mg/m^2 for 5 days by continuous infusion), or standard (100 mg/m^2 per day for 5 days by continuous infusion) doses of cytarabine. A dose-response effect for cytarabine in patients with AML who were aged ≤60 years was demonstrated. High-dose cytarabine significantly prolonged CR and increased the fraction cured in patients with favorable [t(8;21) and inv(16)] and normal cytogenetics, but it had no significant effect on patients with other abnormal karyotypes. For older patients, exploration of attenuated intensive therapy that includes either chemotherapy or reduced-intensity allogeneic HSCT has been pursued. Postremission therapy is a setting for introduction of new agents (Table 109–4).

Table 109–4 Selected New Agents under Study for the Treatment of Adults with AML

Class of Drugs	Examples of Agents in Class
Tyrosine kinase inhibitors	PKC412, MLN518, SU11248, CHIR-258, imatinib (STI571, Gleevec), dasatinib, AMN107
Demethylating agents	Decitabine, 5-azacytidine
Histone deacetylase inhibitors	Suberoylanilide hydroxamic acid (SAHA), MS275, LBH589, valproic acid
Heavy metals	Arsenic trioxide
Farnesyl transferase inhibitors	R115777, SCH66336
HSP-90 antagonists	17-allylaminogeldanamycin (17-AAG), DMAG, or derivatives
Cell cycle inhibitors	Flavopiridol, CYC202 (R-Roscovitine), SNS-032
Nucleoside analogues	Clofarabine, troxacitabine
Humanized antibodies	Anti-CD33 (SGN33), anti-KIR

Toxin-conjugated antibodies	Gemtuzumab ozogamicin
Proteasome inhibitors	Bortezomib
Aurora inhibitors	AZD1152, MLN-8237, AT9283
Immunomodulatory	Lenalidomide, IL-2, histamine dihydrochloride

Abbreviations: AML, acute myeloid leukemia; IL-2, interleukin.

Allogeneic HSCT is used in patients ages <70–75 years with an HLA-compatible donor who have high-risk cytogenetics. In patients with CN-AML and high-risk molecular features such as *FLT3*-ITD, allogeneic HSCT is best applied in the context of clinical trials, as the impact of aggressive therapy on outcome is unknown. Relapse following allogeneic HSCT occurs in only a small fraction of patients, but treatment-related toxicity is relatively high; complications include venoocclusive disease, GVHD, and infections. Autologous HSCT can be administered in young and older patients and uses the same preparative regimens. Patients subsequently receive their own stem cells collected while in remission. The toxicity is relatively low with autologous HSCT (5% mortality rate), but the relapse rate is higher than with allogeneic HSCT, due to the absence of the graft-versus-leukemia (GVL) effect seen with allogeneic HSCT and possible contamination of the autologous stem cells with residual tumor cells. Purging tumor from the autologous stem cells has not lowered the relapse rate with autologous HSCT.

Randomized trials comparing intensive chemotherapy and autologous and allogeneic HSCT have shown improved duration of remission with allogeneic HSCT compared to autologous HSCT or chemotherapy alone. However, overall survival is generally not different; the improved disease control with allogeneic HSCT is erased by the increase in fatal toxicity. While stem cells were previously harvested from the bone marrow, virtually all efforts currently collect these from the blood following mobilization regimens. Prognostic factors may help select patients in first CR for whom transplant is most effective.

Our approach includes allogeneic HSCT if feasible in first CR for patients with high-risk karyotypes (Fig. 109-2). Patients with CN-AML who have other poor risk factors (e.g., an antecedent hematologic disorder, or failure to attain remission with a single induction course) and patients lacking a favorable genotype (e.g., patients who do not have *CEBPA* mutations or *NPM1* mutations without *FLT3*-ITD) are also potential candidates. If a suitable HLA donor does not exist, investigational therapeutic approaches are considered. As *FLT3*-ITD can be targeted with emerging novel inhibitors, patients with this molecular abnormality should be considered for clinical trials with these agents whenever possible. New transplant strategies, including reduced-intensity HSCT, are being explored for consolidation of high-risk AML patients (Chap. 114). Patients with t(8;21) and inv(16) are treated with repetitive doses of high-dose cytarabine, which offers a high frequency of cure without the morbidity of transplant. In AML patients with t(8;21) and inv(16), those with *KIT* mutations, who have a worse prognosis, may be considered for novel investigational studies.

Autologous HSCT is generally applied to AML patients only in the context of a clinical trial or when the risk of repetitive intensive chemotherapy represents a higher risk than the autologous HSCT (e.g., in patients with severe platelet alloimmunization).

RELAPSE

Once relapse occurs, patients are rarely cured with further standard-dose chemotherapy. Patients eligible for allogeneic HSCT should receive transplants expeditiously at the first sign of relapse. Long-term disease-free survival is approximately the same (30–50%) with allogeneic HSCT in first relapse or in second remission. Autologous HSCT rescues about 20% of relapsed patients with AML who have chemosensitive disease. The most important factors predicting response at relapse are the length of the previous CR, whether initial CR was achieved with one or two courses of chemotherapy, and the type of postremission therapy.

Because of the poor outcome of patients in early first relapse (<12 months), it is justified (for patients without HLA-compatible donors) to explore innovative approaches, such as new drugs or immunotherapies (Table 109–4). Patients with longer first CR (>12 months) generally relapse with drug-sensitive disease and have a higher chance of attaining a CR. However, cure is uncommon, and treatment with novel approaches should be considered if allogeneic HSCT is not possible. New agents that may have clinical activity in AML are needed and many are being tested in clinical trials (Table 109–4).

For elderly patients (age >60 years) for whom clinical trials are not available, gemtuzumab ozogamicin is another alternative. The CR rate with this agent is ~30%. However, its effectiveness in early relapsing (<6 months) or refractory AML patients is limited, possibly due to calicheamicin being a potent MDR1 substrate. Toxicity, including myelosuppression, infusion toxicity, and venoocclusive disease, can be observed with gemtuzumab ozogamicin. Pretreatment with glucocorticoids can diminish many of the associated infusion reactions. Studies are examining this treatment in combination with chemotherapy for both young and older patients with previously untreated AML. This agent has been withdrawn from the U.S. market at the request of the U.S. Food and Drug Administration due to concerns about the product's safety and clinical benefit as shown in trials subsequent to those leading to its accelerated approval.

CHRONIC MYELOID LEUKEMIA

Incidence

The incidence of chronic myeloid leukemia (CML) is 1.5 per 100,000 people per year, and the age-adjusted incidence is higher in men than in women (1.9 vs 1.1). The incidence of CML increases slowly with age until the middle forties, when it starts to rise rapidly. The incidence of CML for females decreased slightly (1.8%) between 1994 and 2006 as compared to 1975–1994.

Definition

The diagnosis of CML is established by identifying a clonal expansion of a hematopoietic stem cell possessing a reciprocal translocation between chromosomes 9 and 22. This translocation results in the head-to-tail fusion of the breakpoint cluster region (*BCR*) gene on chromosome 22q11 with the *ABL1* (named after the abelson murine leukemia virus) gene located on chromosome 9q34. Untreated, the disease is characterized by the inevitable transition from a chronic phase to an accelerated phase and on to blast crisis in a median time of 4 years.

Etiology

No clear correlation with exposure to cytotoxic drugs has been found, and no evidence suggests a viral etiology. In the pre-imatinib era, cigarette smoking accelerated the progression to blast crisis and therefore adversely affected survival in CML. Atomic bomb survivors had an increased incidence; the development of a CML cell mass of 10,000/μL took 6.3 years. No increase in CML incidence was found in the survivors of the Chernobyl accident, suggesting that only large doses of radiation can induce CML.

Pathophysiology

The product of the fusion gene resulting from the t(9;22) plays a central role in the development of CML. This chimeric gene is transcribed into a hybrid *BCR-ABL1* mRNA in which exon 1 of *ABL1* is replaced by variable numbers of 5' *BCR* exons. Bcr-Abl fusion proteins, $p210^{BCR\text{-}ABL1}$, are produced that contain NH_2-terminal domains of Bcr and the COOH-terminal domains of

Abl. A rare breakpoint, occurring within the 3' region of the *BCR* gene, yields a fusion protein of 230 kDa, $p230^{BCR\text{-}ABL1}$. Bcr-Abl fusion proteins can transform hematopoietic progenitor cells in vitro. Furthermore, reconstituting lethally irradiated mice with bone marrow cells infected with retrovirus carrying the gene encoding the $p210^{BCR\text{-}ABL1}$ leads to the development of a myeloproliferative syndrome resembling CML in 50% of the mice. Specific antisense oligomers to the *BCR-ABL1* junction inhibit the growth of t(9;22)-positive leukemic cells without affecting normal colony formation.

The mechanism(s) by which $p210^{BCR\text{-}ABL1}$ promotes the transition from the benign state to the fully malignant one is still unclear. Messenger RNA for *BCR-ABL1* can occasionally be detected in normal individuals. However, attachment of the *BCR* sequences to *ABL1* results in three critical functional changes: (1) the Abl protein becomes constitutively active as a tyrosine kinase (TK) enzyme, activating downstream kinases that prevent apoptosis; (2) the DNA-protein-binding activity of Abl is attenuated; and (3) the binding of Abl to cytoskeletal actin microfilaments is enhanced.

DISEASE PROGRESSION

The events associated with transition to the acute phase, a common occurrence in the pre-imatinib era, were extensively studied. Chromosomal instability of the malignant clone resulting, for example, in the acquisition of an additional t(9;22), trisomy 8, or 17p- (*TP*53 loss) is a basic feature of CML. Acquisition of these additional genetic and/or molecular abnormalities is critical to the phenotypic transformation. Heterogeneous structural alterations of the *TP53* gene, as well as structural alterations and lack of protein production of the retinoblastoma 1 (*RB1*) gene and the catalytic component of telomerase, have been associated with disease progression in a subset of patients. Rare patients show alterations in the rat sarcoma viral oncogene homologue (*RAS*). Sporadic reports also document the presence of an altered v-myc myelocytomatosis viral oncogene homologue (avian) (*MYC*) gene. Progressive de novo DNA methylation at the *BCR-ABL1* locus and hypomethylation of the *LINE-1* retrotransposon promoter herald blastic transformation. Further, interleukin 1β may be involved in the progression of CML to the blastic phase. In addition, functional inactivation of the tumor suppressor protein phosphatase A2 may be required for blastic transformation. Finally, CML that develops resistance to imatinib is at an increased risk of progressing to accelerated/blast crisis. Multiple pathways to disease transformation exist, but the exact timing and relevance of each remain unclear.

Clinical Presentation

SYMPTOMS

The clinical onset of the chronic phase is generally insidious. Accordingly, some patients are diagnosed, while still asymptomatic, during health-screening tests; other patients present with fatigue, malaise, and weight loss or have symptoms resulting from splenic enlargement, such as early satiety and left upper quadrant pain or mass. Less common are features related to granulocyte or platelet dysfunction, such as infections, thrombosis, or bleeding. Occasionally, patients present with leukostatic manifestations due to severe leukocytosis or thrombosis such as vasoocclusive disease, cerebrovascular accidents, myocardial infarction, venous thrombosis, priapism, visual disturbances, and pulmonary insufficiency. Patients with $p230^{BCR\text{-}ABL1}$-positive CML have a more indolent course.

Progression of CML is associated with worsening symptoms. Unexplained fever, significant weight loss, increasing dose requirement of the drugs controlling the disease, bone and joint pain, bleeding, thrombosis, and infections suggest transformation into accelerated or blastic phases. Less than 10–15% of newly diagnosed patients present with accelerated disease or with de novo blastic phase CML.

PHYSICAL FINDINGS

Minimal to moderate splenomegaly is the most common physical finding; mild hepatomegaly is found occasionally. Persistent splenomegaly despite continued therapy is a sign of disease acceleration. Lymphadenopathy and myeloid sarcomas are unusual except late in the course of the disease; when they are present, the prognosis is poor.

HEMATOLOGIC FINDINGS

Elevated white blood (cell) counts (WBCs), with increases in both immature and mature granulocytes, are present at diagnosis. Usually <5% circulating blasts and <10% blasts and promyelocytes are noted, with the majority of cells being myelocytes, metamyelocytes, and band forms. Cycling of the counts may be observed in patients followed without treatment. Platelet counts are almost always elevated at diagnosis, and a mild degree of normocytic normochromic anemia is present. Leukocyte alkaline phosphatase is low in CML cells. Phagocytic functions are usually normal at diagnosis and remain normal during the chronic phase. Histamine production secondary to basophilia is increased in later stages, causing pruritus, diarrhea, and flushing.

At diagnosis, bone marrow cellularity is increased, with an increased myeloid-to-erythroid ratio. The marrow blast percentage is generally normal or slightly elevated. Marrow or blood basophilia, eosinophilia, and monocytosis may be present. While collagenfibrosis in the marrow is unusual at presentation, significant degrees of reticulin stain–measured fibrosis are noted in about half of the patients.

Disease acceleration is defined by the development of increasing degrees of anemia unaccounted for by bleeding or therapy; cytogenetic clonal evolution; or blood or marrow blasts between 10 and 20%, blood or marrow basophils ≥20%, or platelet count <100,000/μL. *Blast crisis* is defined as acute leukemia, with blood or marrow blasts ≥20%. Hyposegmented neutrophils may appear (Pelger-Huët anomaly). Blast cells can be classified as myeloid, lymphoid, erythroid, or undifferentiated, based on morphologic, cytochemical, and immunologic features. Occurrence of de novo blast crisis or following imatinib therapy is rare.

CHROMOSOMAL FINDINGS

The cytogenetic hallmark of CML, found in 90–95% of patients, is the t(9;22)(q34;q11.2). Originally, this was recognized by the presence of a shortened chromosome 22 (22q-), designated as the *Philadelphia chromosome*, that arises from the reciprocal t(9;22). Some patients may have complex translocations (designated as *variant translocations*) involving three, four, or five chromosomes (usually including chromosomes 9 and 22). However, the molecular consequences of these changes are similar to those resulting from the typical t(9;22). All patients should have evidence of the translocation molecularly or by cytogenetics or FISH to make a diagnosis of CML.

Prognostic Factors

The clinical outcome of patients with CML is variable. Before imatinib mesylate, death was expected in 10% of patients within 2 years and in about 20% yearly thereafter, and the median survival time was ~4 years. Therefore, several prognostic models that identify different risk groups in CML were developed. The most commonly used staging systems have been derived from multivariate analyses of prognostic factors. The *Sokal index* identified percentage of circulating blasts, spleen size, platelet count, age, and cytogenetic clonal evolution as the most important prognostic indicators. This system was developed based on chemotherapy-treated patients. The *Hasford system* was developed based on interferon (IFN) α–treated patients. It identified percentage of circulating blasts, spleen size, platelet count, age, and percentage of eosinophils and basophils as the most important prognostic indicators. This system differs from the Sokal index by ignoring clonal evolution and incorporating percentage of eosinophils and basophils. When applied to a data set of 272 patients treated with IFN-α, the Hasford system was

better than the Sokal score for predicting survival time; it identified more low-risk patients but left only a small number of cases in the high-risk group. Preliminary results suggest that both the Sokal and the Hasford systems are applicable to imatinib-treated patients.

Treatment: Chronic Myeloid Leukemia

The therapy of CML is changing rapidly because we have a proven curative treatment (allogeneic transplantation) that has significant toxicity and a new targeted treatment (imatinib) with outstanding outcome based on 8-year follow-up data. We recommend starting with TK inhibitors and reserving allogeneic transplantation for those who develop imatinib resistance.

At present, the goal of therapy in CML is to achieve prolonged, durable, nonneoplastic, nonclonal hematopoiesis, which entails the eradication of any residual cells containing the *BCR-ABL1* transcript. Hence, the goal is complete molecular remission and cure. A proposed imatinibtreatment algorithm for the newly diagnosed CML patient is presented in Table 109–5.

Table 109–5 Imatinib Treatment Milestones for Newly Diagnosed CML Patients

	NCCN[a]		ELN[b]	
Time Months	**Expected[c]**	**Failure[d]**	**Suboptimal[e]**	**Failure[d]**
3	Complete hematologicremission[f]	No complete hematologic remission	Minor cytogenetic remission	No cytogenetic remission; new mutations
6	Any cytogenetic remission	No cytogenetic remission	Partial cytogenetic remission	Minimal cytogenetic remission[g]; new mutations
12	Complete[h] or partial[i] cytogenetic remission	Minor[j] or no cytogenetic remission	Less than major molecular response	Less than partial cytogenetic remission; new mutations
18	Complete cytogenetic remission	Partial, minor, or no cytogenetic remission		
Anytime	Loss of previously achieved hematologic, cytogenetic, or molecular remission; new mutations[d]			

[a]National Comprehensive Cancer Network.

[b]European Leukemia Net.

[c]Denotes that at the indicated milestone, patients should stay on the same dose.

[d]Denotes that at the indicated milestones, for patients on 400 mg/d, one can either increase the dose to a maximum of 600–800 mg, as tolerated, or probably switch to another TK inhibitor.

[e]Denotes that the patients may still have substantial long-term benefit from continuing a specific treatment, but chances are reduced and therefore these patients may be eligible for alternative treatments.

[f]Complete hematologicremission, white blood cell count <10,000/μL, normal morphology, hemoglobin and platelet counts, and disappearance of splenomegaly.

[g]Minimal cytogenetic remission, 66–95% bone marrow metaphases with t(9;22).

[h]Complete cytogenetic remission, no bone marrow metaphases with t(9;22).

[i]Partial cytogenetic remission, 1–35% bone marrow metaphases with t(9;22).

[j]Minor cytogenetic remission, 36–85% bone marrow metaphases with t(9;22).

Abbreviation: CML, chronic myeloid leukemia.

IMATINIB MESYLATE

Imatinib mesylate (Gleevec) functions through competitive inhibition at the ATP-binding site of the Abl kinase in the inactive conformation, which leads to inhibition of tyrosine phosphorylation of proteins involved in Bcr-Abl signal transduction. It shows specificity for Bcr-Abl, the receptor for platelet-derived growth factor, and Kit TK. Imatinib induces apoptosis in cells expressing Bcr-Abl.

In newly diagnosed CML, imatinib (400 mg/d) is more effective than IFN-α and cytarabine. The complete hematologicremission rate of patients treated with imatinib was 95% compared to 56% in patients treated with IFN-α and cytarabine. Similarly, the complete cytogenetic remission rate at 18 months was 76% with imatinib compared to 15% with IFN-α and cytarabine. The rate of complete cytogenetic remission in imatinib-treated patients differed by Sokal score: the rate in those with low-risk disease was 89% compared with 82% for patients with intermediate-risk disease and 69% for those with high-risk disease.

All imatinib-treated patients who achieved major molecular remission (26%), defined as ≥3 log reduction in *BCR-ABL1* transcript level at 18 months compared to pretreatment level, were progression-free at 5 years. The progression-free survival (PFS) at 5 years for patients achieving complete cytogenetic remission but less pronounced molecular remission is 98%. The 5-year PFS for patients not achieving complete cytogenetic remission at 18 months was 87%. These results have led to a consensus that molecular responses can be used as a treatment goal in CML. Specific milestones have been developed for chronic-phase CML patients (Table 109–5). They differ between the Americans (NCCN) and the Europeans (ELN), with more strict milestones by the latter. For example, in the NCCN milestones, chronic-phase CML patients who do not achieve any cytogenetic remission following 6 months of imatinib should be offered other treatment approaches while in the ELN milestones, the same recommendation is offered following 3 months of imatinib treatment. We favor the ELN approach and expect the NCCN milestones to align with the ELN ones in the very near future.

Progression to accelerated/blastic phases of the disease was noted in 3% of patients treated with imatinib as compared to 8.5% of patients treated with IFN-α and cytarabine during the first year. Over time, the annual incidence of disease progression on imatinib decreased gradually to <1% during the fourth year and beyond, and no patient who achieved major molecular response by 12 months progressed to the accelerated/blastic phases of the disease.

Treatment is currently recommended for life unless patients are enrolled in a clinical trial with a specific question of treatment discontinuation. An early trial evaluating the effect of imatinib discontinuation after at least 2 years of complete molecular remission revealed molecular relapse in 6 of 12 patients. Interestingly, 6 of 10 patients who were treated with IFN-α before imatinib maintained molecular remission, while both patients who were not exposed to IFN-α relapsed. These results raised the hypothesis that IFN-α may have a protective effect against relapse, possibly by eradicating the leukemia-initiating cells. This

hypothesis is supported by the randomized trial comparing imatinib to imatinib plus IFN-α; preliminary results from this trial revealed better major molecular response for the combination although a significant number of patients discontinued IFN-α treatment during the first year due to toxicity. Finally, a recent IFN-α maintenance study, following imatinib discontinuation, demonstrated maintained molecular remission in 15 (75%) of 20 patients. IFN's mechanism of action in this situation is unclear.

Imatinib is administered orally. The main side effects are fluid retention, nausea, muscle cramps, diarrhea, and skin rashes. The management of these side effects is usually supportive. Myelosuppression is the most common hematologic side effect. Myelosuppression, while rare, may require holding drug and/or growth factor support. Doses <300 mg/d seem ineffective and may lead to development of resistance.

Four mechanisms of resistance to imatinib have been described to date. These are (1) gene amplification, (2) mutations at the kinase site, (3) enhanced expression of multidrug exporter proteins, and (4) alternative signaling pathways functionally compensating for the imatinib-sensitive mechanisms. All four mechanisms are being targeted in clinical trials.

BCR-ABL1 gene amplification and decreased intracellular imatinib concentrations are addressed by intensifying the therapy with higher (up to 800 mg/d) imatinib doses. Three randomized trials have been published so far. The first randomized study compared 400 mg/d to 800 mg/d in newly diagnosed CML patients and revealed improved major molecular responses at 3, 6, and 9 months but similar results at 12 months. A similar study comparing 600 mg/d to 800 mg/d showed a borderline benefit for the higher dose based on both cytogenetic and major molecular responses at 12 months, while a third study, concentrating only on high-risk (Sokal) patients failed to show any significant difference between 400 mg/d and 800 mg/d at 12 months. All these studies have too short follow-up to evaluate dosing effect on survival.

Mutations at the kinase domain occur in approximately half of imatinib-resistant chronic-phase cases and even more frequently in the more advanced phases of the disease. These mutations are being targeted by novel TK inhibitors that have a different conformation than imatinib, demonstrating activity against most imatinib-resistant mutations. Nilotinib (Tasigma), like imatinib, binds to the kinase domain in the inactive conformation. Dasatinib (Sprycel) binds to the kinase domain in the open conformation and also inhibits the SRC (sarcoma) family of kinases, addressing the last mechanism of resistance. CML with the T315I mutation is resistant to imatinib, nilotinib, and dasatinib. In addition, nilotinib is also resistant to E255K/V and Y253F/H while dasatinib is also resistant to X299L and F317L.

Dasatinib is approved by the FDA at a dose of 100 mg/day for the treatment of all stages of CML with resistance or intolerance to prior therapy, including imatinib. Nilotinib is approved by the FDA at a dose of 400 mg twice daily for the treatment of chronic- and accelerated-phase CML with resistance or intolerance to prior therapy, including imatinib. Both are oral agents, dasatinib is given once daily while nilotinib is given twice daily with food restrictions before and after dosing. Their toxicity profiles are similar to imatinib with small but significant differences. Dasatinib causes pleural effusions in 22% of patients, with 7% developing grade 3–4 toxicity. Nilotinib was associated with sudden death in 6 of approximately 550 CML patients. A suspected relationship to nilotinib was reported in two of these cases and led to a requirement for additional cardiac monitoring while using this drug. A randomized trial in chronic-phase imatinib-resistant CML patients showed superiority of switching to dasatinib over increasing imatinib to 800 mg/day. Finally, randomized trials have demonstrated that either nilotinib or dasatinib are more effective than imatinib as first-line treatment in newly diagnosed chronic-phase CML patients in time to complete hematologic and cytogenetic remission and major molecular response at 1 year and led to their approval for the first line setting. Similar results are likely with bosutinib, another Src and Abl TK inhibitor. These studies are expanding the armamentarium for newly diagnosed CML patients.

These new agents have already changed the treatment algorithm of CML. For example, patients who do not achieve any cytogenetic remission at 6 months (or 3 months by ELN) on imatinib are now offered dasatinib, nilotinib, or HSCT. IFN-α is FDA approved for CML but is only offered if all other options have failed.

The encouraging results with imatinib have led clinicians to offer it as first-line therapy for newly diagnosed CML patients, including those who otherwise would have benefited from transplant (e.g., young patients with a matched sibling donor). Prior exposure to imatinib does not affect transplant outcome. Similar data, in smaller series, were also described for dasatinib and nilotinib treatment before HSCT. However, delaying HSCT for high-risk patients (Sokal/Hasford criteria) may result in disease progression. HSCT after disease progression is associated with poorer outcome. Therefore, we recommend close monitoring of TK inhibitors response in these patients.

NEW AGENTS

Several new agents are now in development for CML with T315I and patients who fail all currently available TK inhibitors. These include omacetaxine, XL228, FTY720, AP24534, DCC-2036, PH-739358, and sorafenib (Table 109–6).

Table 109–6 Novel Agents for BCR-ABL with T315I and Patients Who Failed All Currently Available Tyrosine Kinase Inhibitors

Agent	Mechanism of Action
Omacetaxine (formerly known as homoharringtonine)	Protein translation inhibitor
XL228	Dual Src/Abl inhibitor with potential effect against T315I mutation
FTY720 (also known as fingolimod)	Activation of protein phosphatase 2A that is essential for *ABL1*-mediated leukemogenesis
AP24534	Pan-Bcr-Abl inhibitor that inhibits T315I
DCC-2036	Non-ATP-competitive Abl inhibition, avoids the steric clash with T315I mutation
PH-739358	Aurora kinase inhibitor that is also active against T315I mutation
Sorafenib	Raf kinase inhibitor that down regulates down stream Bcr-Abl targets

ALLOGENEIC HSCT

Allogeneic HSCT is complicated by early mortality owing to the transplant procedure. Outcome of HSCT depends on multiple factors, including (1) the patient (e.g., age and phase of disease); (2) the type of donor [e.g., syngeneic (monozygotic twins) or HLA-compatible allogeneic, related or unrelated]; (3) the preparative regimen (myeloablative or reduced-intensity); (4) GVHD; and (5) posttransplantation treatment.

Posttransplantation Treatment

Posttransplant *BCR-ABL1* transcript levels have served as early predictors for hematologicrelapse following HSCT. These should facilitate risk-adapted approaches with immunosuppression or TK inhibitor(s), or a combination of the two. Donor leukocyte infusions (without any preparative chemotherapy or GVHD prophylaxis) can induce hematologic and cytogenetic remissions in patients with CML who have relapsed after allogeneic HSCT but carry the risk of significant GVHD.

Imatinib can control CML that has recurred after allogeneic HSCT but is sometimes associated with myelosuppression and

recurrence of severe GVHD. Imatinib after allogeneic HSCT is being studied for prevention of relapse in patients with advanced disease at the time of transplantation (i.e., patients at high risk for relapse), patients undergoing reduced-intensity transplants, or patients with slow reduction of *BCR-ABL1* message following transplantation. Imatinib has also been combined with donor lymphocytes to induce rapid molecular remissions in CML patients with disease relapse after allogeneic HSCT. Of interest are studies with newer TK inhibitors following transplantation for imatinib-resistant CML.

INTERFERON

Before imatinib, when allogeneic HSCT was not feasible, IFN-α therapy was the treatment of choice. Only longer follow-up of patients treated with imatinib will prove whether IFN-α will still have a role in the treatment of CML. Its mode(s) of action in CML is still unknown.

CHEMOTHERAPY

Initial management of patients with chemotherapy is currently reserved for rapid lowering of WBCs, reduction of symptoms, and reversal of symptomatic splenomegaly. Hydroxyurea, a ribonucleotide reductase inhibitor, induces rapid disease control. The initial dose is 1–4 g/d; the dose should be halved with each 50% reduction of the leukocyte count. Unfortunately, cytogenetic remissions with hydroxyurea are uncommon. Busulphan, an alkylating agent that acts on early progenitor cells, has a more prolonged effect. However, we do not recommend its use because of its serious side effects, which include unexpected, and occasionally fatal, myelosuppression in 5–10% of patients; pulmonary, endocardial, and marrow fibrosis; and an Addison-like wasting syndrome.

AUTOLOGOUS HSCT

Autologous HSCT could potentially cure CML if cells are collected at complete molecular remission. However, since patients who achieve this degree of response do not relapse, this treatment modality has been abandoned by most groups.

LEUKAPHERESIS AND SPLENECTOMY

Intensive leukapheresis may control the blood counts in chronic-phase CML; however, it is expensive and cumbersome. It is useful in emergencies where leukostasis-related complications such as pulmonary failure or cerebrovascular accidents are likely. It may also have a role in the treatment of pregnant women, in whom it is important to avoid potentially teratogenic drugs.

Splenectomy was used in CML in the past because of the suggestion that evolution to the acute phase might occur in the spleen. However, this does not appear to be the case, and splenectomy is now reserved for symptomatic relief of painful splenomegaly unresponsive to imatinib or chemotherapy, or for significant anemia or thrombocytopenia associated with hypersplenism. Splenic radiation is used rarely to reduce the size of the spleen.

MINIMAL RESIDUAL DISEASE

The kinetics of *BCR-ABL1* transcript elimination is currently replacing qualitative detection of the *BCR-ABL1* message as an index of tumor burden, in spite of a lack of standard acceptable methodology. A consensus panel has proposed ways to harmonize the different methods and to use a conversion factor so that individual laboratories will be able to express *BCR-ABL1* transcript levels on an agreed upon scale.

Slow reduction of *BCR-ABL1* transcripts following HSCT correlates with the possibility of hematologicrelapse. However, the definition of "slow reduction" depends on the preparative regimen (reduced-intensity vs fully myeloablative) and the selection of time points to measure the transcript levels. While persistent RT-PCR positivity at 6 months was regarded as an indication for additional therapy in the past, current studies utilize periods between engraftment and day 100 for evaluating the clearance rate of *BCR-ABL1* transcripts and recommending additional therapies. Large trials with longer follow-up are needed to establish consensus guidelines.

The randomized trial of imatinib versus IFN-α and cytarabine (IRIS) was the first to establish the concept of $\log_{10}$ reduction of *BCR-ABL1* transcript from a standardized baseline for untreated patients. This measurement unit was developed instead of either the transcript numbers expressed per microgram of leukocyte RNA or the ratio of *BCR-ABL1* to a housekeeping gene on a log scale. In this randomized trial, patients who achieved ≥3 log reduction of *BCR-ABL1* message had an extremely low probability of relapse, with a median follow-up of 96 months.

These studies also established the value and convenience of using peripheral blood instead of bone marrow testing as a means to assess disease status in patients who achieve complete cytogenetic responses. However, one still needs to consider following CML patients in complete cytogenetic remission and at least major molecular remission with occasional cytogenetic bone marrow testing. This should be performed if they develop cytopenia late in the treatment course as such patients are at risk of developing cytogenetic aberrations, especially monosomy 7, in t(9;22)-negative cells and secondary MDS/AML. Other aberrations in the t(9;22)-negative cells are frequently transient, and their clinical significance is unclear. Development of secondary MDS/AML is rare.

TREATMENT OF BLAST CRISIS

Treatments for primary blast crisis, including imatinib, are generally ineffective. Only 52% of patients treated with imatinib achieved hematologicremission (21% complete hematologic remission), and the median overall survival was 6.6 months. Patients who achieve complete hematologic remission or whose disease returns to a second chronic phase should be considered for allogeneic HSCT. Other approaches include induction chemotherapy tailored to the phenotype of the blast cell followed by TK inhibitors, with or without additional chemotherapy and HSCT. Blast crisis following initial therapy with imatinib carries a dismal prognosis even if treated with dasatinib or nilotinib.

FURTHER READINGS

AML

Döhner H et al: Diagnosis and management of acute myeloid leukemia in adults: Recommendations from an international expert panel, on behalf of the European Leukemia Net. Blood 115:453, 2010[PMID: 21698843]

National Comprehensive Cancer Network: Acute myeloid leukemia. Clinical Practice Guidelines in Oncology, Version 2. 2010.*http://www.nccn.org/professionals/physician_gls/PDF/cml.pdf*

Sanz MA et al: Management of acute promyelocytic leukemia: Recommendations from an expert panel on behalf of the European Leukemia Net. Blood 113:1875, 2009[PMID: 18812465]

Vardiman JW et al: The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: Rationale and important changes. Blood 114:937, 2009[PMID: 19357394]

CML

Baccarani M et al: Chronic myeloid leukemia: An update of concepts and management recommendations of European Leukemia

Net. J Clin Oncol 27:6041, 2009[PMID: 19884523]

Irvine DA et al: Optimising chronic myeloid leukaemia therapy in the face of resistance to tyrosine kinase inhibitors—a synthesis of clinical and laboratory data. Blood Rev 24:1, 2010[PMID: 20005615]

Kantarjian H et al: Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med 362:2260, 2010[PMID: 20525995]

National Comprehensive Cancer Network: Chronic myelogenous leukemia. Clinical Practice Guidelines in Oncology, Version 2. 2010. *http://www.nccn.org/professionals/physician_gls/PDF/cml.pdf*

Saglio G et al: Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia. N Engl J Med 362:2251, 2010[PMID: 20525993]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MALIGNANCIES OF LYMPHOID CELLS: INTRODUCTION

Malignancies of lymphoid cells range from the most indolent to the most aggressive human malignancies. These cancers arise from cells of the immune system at different stages of differentiation, resulting in a wide range of morphologic, immunologic, and clinical findings. Insights on the normal immune system have allowed a better understanding of these sometimes confusing disorders.

Some malignancies of lymphoid cells almost always present as leukemia (i.e., primary involvement of bone marrow and blood), while others almost always present as lymphomas (i.e., solid tumors of the immune system). However, other malignancies of lymphoid cells can present as either leukemia or lymphoma. In addition, the clinical pattern can change over the course of the illness. This change is more often seen in a patient who seems to have a lymphoma and then develops the manifestations of leukemia over the course of the illness.

BIOLOGY OF LYMPHOID MALIGNANCIES: CONCEPTS OF THE WHO CLASSIFICATION OF LYMPHOID MALIGNANCIES

The classification of lymphoid cancers evolved steadily throughout the twentieth century. The distinction between leukemia and lymphoma was made early, and separate classification systems were developed for each. Leukemias were first divided into acute and chronic subtypes based on average survival. Chronic leukemias were easily subdivided into those of lymphoid or myeloid origin based on morphologic characteristics. However, a spectrum of diseases that were formerly all called *chronic lymphoid leukemia* has become apparent **(Table 110–1)**. The acute leukemias were usually malignancies of blast cells with few identifying characteristics. When cytochemical stains became available, it was possible to divide these objectively into myeloid malignancies and acute leukemias of lymphoid cells. Acute leukemias of lymphoid cells have been subdivided based on morphologic characteristics by the French-American-British (FAB) group **(Table 110–2)**. Using this system, lymphoid malignancies of small uniform blasts (e.g., typical childhood acute lymphoblastic leukemia) were called L1, lymphoid malignancies with larger and more variable size cells were called L2, and lymphoid malignancies of uniform cells with basophilic and sometimes vacuolated cytoplasm were called L3 (e.g., typical Burkitt's lymphoma cells). Acute leukemias of lymphoid cells have also been subdivided based on immunologic (i.e., T cell vs. B cell) and cytogenetic abnormalities (Table 110–2). Major cytogenetic subgroups include the t(9;22) (e.g., Philadelphia chromosome–positive acute lymphoblastic leukemia) and the t(8;14) found in the L3 or Burkitt's leukemia.

Table 110–1 Lymphoid Disorders that Can Present as "Chronic Leukemia" and Be Confused with Typical B Cell Chronic Lymphoid Leukemia

Follicular lymphoma	Prolymphocytic leukemia (B cell or
Splenic marginal zone lymphoma	T cell)
Nodal marginal zone lymphoma	Lymphoplasmacytic lymphoma
Mantle cell lymphoma	Sézary's syndrome
Hairy cell leukemia	Smoldering adult T cell leukemia/lymphoma

Table 110–2 Classification of Acute Lymphoid Leukemia (ALL)

Immunologic Subtype	% of Cases	FAB Subtype	Cytogenetic Abnormalities
Pre-B ALL	75	L1, L2	t(9;22), t(4;11), t(1;19)
T cell ALL	20	L1, L2	14q11 or 7q34
B cell ALL	5	L3	t(8;14), t(8;22), t(2;8)

Abbreviation: FAB, French-American-British classification.

Non-Hodgkin's lymphomas were separated from Hodgkin's disease by recognition of the Sternberg-Reed cells early in the twentieth century. The histologic classification for non-Hodgkin's lymphomas has been one of the most contentious issues in oncology. Imperfect morphologic systems were supplanted by imperfect immunologic systems, and poor reproducibility of diagnosis has hampered progress. In 1999, the World Health Organization (WHO) classification of lymphoid malignancies was devised through a process of consensus development among international leaders in hematopathology and clinical oncology. The WHO

classification takes into account morphologic, clinical, immunologic, and genetic information and attempts to divide non-Hodgkin's lymphomas and other lymphoid malignancies into clinical/pathologic entities that have clinical and therapeutic relevance. This system is presented in Table 110–3. This system is clinically relevant and has a higher degree of diagnostic accuracy than those used previously. The possibilities for subdividing lymphoid malignancies are extensive. However, Table 110–3 presents in bold those malignancies that occur in at least 1% of patients. Specific lymphoma subtypes will be dealt with in more detail below. Lymphomas associated with HIV infection are discussed in Chap. 189.

Table 110–3 WHO Classification of Lymphoid Malignancies

B Cell	T Cell	Hodgkin's Disease
Precursor B cell neoplasm	Precursor T cell neoplasm	Nodular lymphocyte-predominant Hodgkin's disease
Precursor B lymphoblastic leukemia/lymphoma (precursor B cell acute lymphoblastic leukemia)	**Precursor T lymphoblastic lymphoma/leukemia (precursor T cell acute lymphoblastic leukemia)**	
Mature (peripheral) B cell neoplasms	Mature (peripheral) T cell neoplasms	Classical Hodgkin's disease
B cell chronic lymphocytic leukemia/small lymphocytic lymphoma	T cell prolymphocytic leukemia	Nodular sclerosis Hodgkin's disease
B cell prolymphocytic leukemia	T cell granular lymphocytic leukemia	Lymphocyte-rich classic Hodgkin's disease
Lymphoplasmacytic lymphoma	Aggressive NK cell leukemia	Mixed-cellularity Hodgkin's disease
Splenic marginal zone B cell lymphoma (± villous lymphocytes)	Adult T cell lymphoma/leukemia (HTLV-I+)	Lymphocyte-depletion Hodgkin's disease
Hairy cell leukemia	Extranodal NK/T cell lymphoma, nasal type	
Plasma cell myeloma/plasmacytoma	Enteropathy-type T cell lymphoma	
Extranodal marginal zone B cell lymphoma of MALT type	Hepatosplenic $\gamma\delta$ T cell lymphoma	
Mantle cell lymphoma	Subcutaneous panniculitis-like T cell lymphoma	
Follicular lymphoma	**Mycosis fungoides/Sézary's syndrome**	
Nodal marginal zone B cell lymphoma (± monocytoid B cells)	Anaplastic large cell lymphoma, primary cutaneous type	
Diffuse large B cell lymphoma	**Peripheral T cell lymphoma, not otherwise specified (NOS)**	
Burkitt's lymphoma/Burkitt's cell leukemia	**Angioimmunoblastic T cell lymphoma**	
	Anaplastic large cell lymphoma, primary systemic type	

Note: Malignancies in bold occur in at least 1% of patients.

Abbreviations: HTLV, human T cell lymphotropic virus; MALT, mucosa-associated lymphoid tissue; NK, natural killer; WHO, World Health Organization.

Source: Adapted from Harris et al.

GENERAL ASPECTS OF LYMPHOID MALIGNANCIES

Etiology and Epidemiology

The relative frequency of the various lymphoid malignancies is shown in Fig. 110-1. Chronic lymphoid leukemia (CLL) is the most prevalent form of leukemia in Western countries. It occurs most frequently in older adults and is exceedingly rare in children. In 2010, 14,990 new cases were diagnosed in the United States, but because of the prolonged survival associated with this disorder, the total prevalence is many times higher. CLL is more common in men than in women and more common in whites than in blacks. This is an uncommon malignancy in Asia. The etiologic factors for typical CLL are unknown.

Figure 110-1

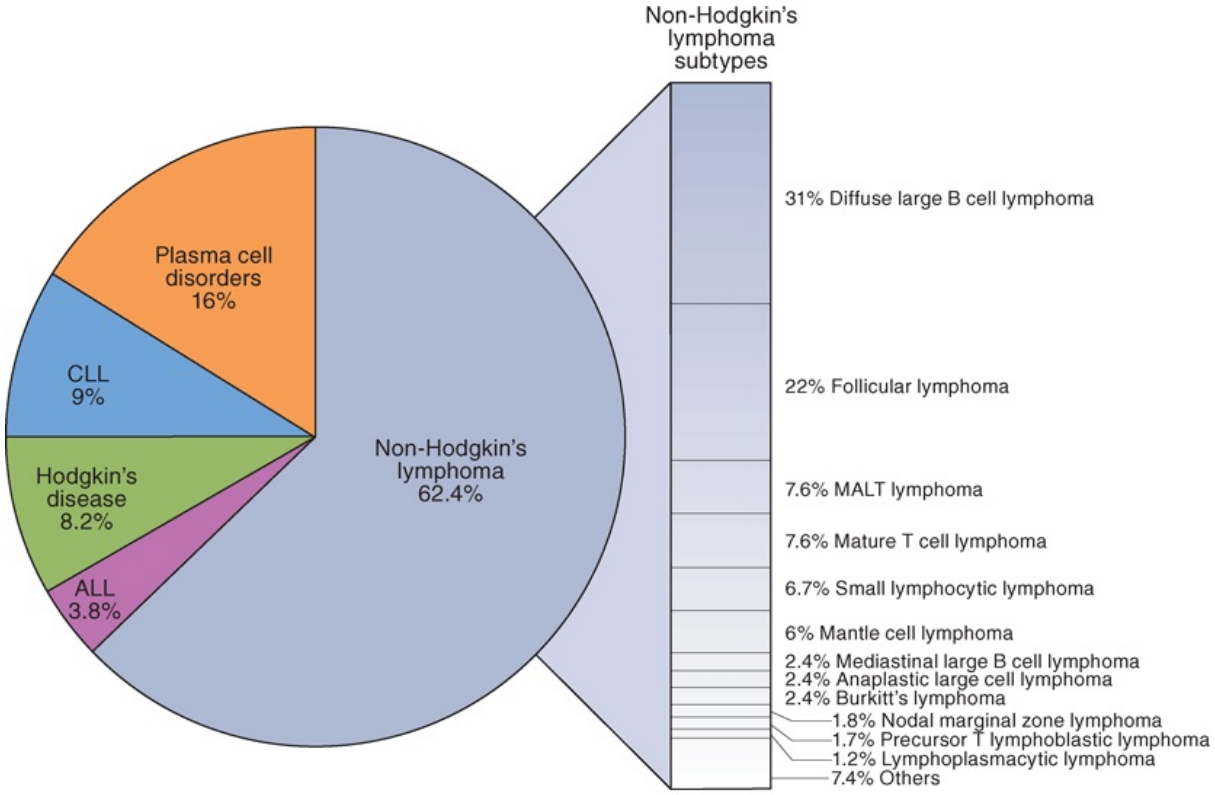

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relative frequency of lymphoid malignancies. ALL, acute lymphoid leukemia; CLL, chronic lymphoid leukemia; MALT, mucosa-associated lymphoid tissue.

In contrast to CLL, acute lymphoid leukemias (ALLs) are predominantly cancers of children and young adults. The L3 or Burkitt's leukemia occurring in children in developing countries seems to be associated with infection by the Epstein-Barr virus (EBV) in infancy. However, the explanation for the etiology of more common subtypes of ALL is much less certain. Childhood ALL occurs more often in higher socioeconomic subgroups. Children with trisomy 21 (Down syndrome) have an increased risk for childhood ALL as well as acute myeloid leukemia (AML). Exposure to high-energy radiation in early childhood increases the risk of developing T cell ALL.

The etiology of ALL in adults is also uncertain. ALL is unusual in middle-aged adults but increases in incidence in the elderly. However, AML is still much more common in older patients. Environmental exposures, including certain industrial exposures, exposure to agricultural chemicals, and smoking, might increase the risk of developing ALL as an adult. ALL was diagnosed in 5330 persons and AML in 12,330 persons in the United States in 2010.

The preponderance of evidence suggests that Hodgkin's disease is of B cell origin. The incidence of Hodgkin's disease appears fairly stable, with 8490 new cases diagnosed in 2010 in the United States. Hodgkin's disease is more common in whites than in blacks and more common in males than in females. A bimodal distribution of age at diagnosis has been observed, with one peak incidence occurring in patients in their twenties and the other in those in their eighties. Some of the late age peak may be attributed to confusion among entities with similar appearance such as anaplastic large cell lymphoma and T cell–rich B cell lymphoma. Patients in the younger age groups diagnosed in the United States largely have the nodular sclerosing subtype of Hodgkin's disease. Elderly patients, patients infected with HIV, and patients in Third World countries more commonly have mixed-cellularity Hodgkin's disease or lymphocyte-depleted Hodgkin's disease. Infection by HIV is a risk factor for developing Hodgkin's disease. In addition, an association between infection by EBV and Hodgkin's disease has been suggested. A monoclonal or oligoclonal proliferation of EBV-infected cells in 20–40% of the patients with Hodgkin's disease has led to proposals for this virus having an etiologic role in Hodgkin's disease. However, the matter is not settled definitively.

For unknown reasons, non-Hodgkin's lymphomas increased in frequency in the United States at the rate of 4% per year and increased 2–8% per year globally between 1950 and the late 1990s. The rate of increase in the past few years seems to be decreasing. About 65,540 new cases of non-Hodgkin's lymphoma were diagnosed in the United States in 2010 and nearly 360,000 cases worldwide. Non-Hodgkin's lymphomas are more frequent in the elderly and more frequent in men. Patients with both primary and secondary immunodeficiency states are predisposed to developing non-Hodgkin's lymphomas. These include patients with HIV infection; patients who have undergone organ transplantation; and patients with inherited immune deficiencies, the sicca syndrome, and rheumatoid arthritis.

The incidence of non-Hodgkin's lymphomas and the patterns of expression of the various subtypes differ geographically. T cell lymphomas are more common in Asia than in Western countries, while certain subtypes

of B cell lymphomas such as follicular lymphoma are more common in Western countries. A specific subtype of non-Hodgkin's lymphoma known as the angiocentric nasal T/natural killer (NK) cell lymphoma has a striking geographic occurrence, being most frequent in Southern Asia and parts of Latin America. Another subtype of non-Hodgkin's lymphoma associated with infection by human T cell lymphotropic virus (HTLV) I is seen particularly in southern Japan and the Caribbean (Chap. 188).

A number of environmental factors have been implicated in the occurrence of non-Hodgkin's lymphoma, including infectious agents, chemical exposures, and medical treatments. Several studies have demonstrated an association between exposure to agricultural chemicals and an increased incidence in non-Hodgkin's lymphoma. Patients treated for Hodgkin's disease can develop non-Hodgkin's lymphoma; it is unclear whether this is a consequence of the Hodgkin's disease or its treatment. However, a number of non-Hodgkin's lymphomas are associated with infectious agents **(Table 110–4)**. HTLV-I infects T cells and leads directly to the development of adult T cell lymphoma (ATL) in a small percentage of infected patients. The cumulative lifetime risk of developing lymphoma in an infected patient is 2.5%. The virus is transmitted by infected lymphocytes ingested by nursing babies of infected mothers, bloodborne transmission, or sexually. The median age of patients with ATL is ~56 years, emphasizing the long latency. HTLV-I is also the cause of tropical spastic paraparesis—a neurologic disorder that occurs somewhat more frequently than lymphoma and with shorter latency and usually from transfusion-transmitted virus (Chap. 188).

Table 110–4 Infectious Agents Associated with the Development of Lymphoid Malignancies

Infectious Agent	Lymphoid Malignancy
Epstein-Barr virus	Burkitt's lymphoma Post–organ transplant lymphoma Primary CNS diffuse large B cell lymphoma Hodgkin's disease Extranodal NK/T cell lymphoma, nasal type
HTLV-I	Adult T cell leukemia/lymphoma
HIV	Diffuse large B cell lymphoma Burkitt's lymphoma
Hepatitis C virus	Lymphoplasmacytic lymphoma
Helicobacter pylori	Gastric MALT lymphoma
Human herpesvirus 8	Primary effusion lymphoma Multicentric Castleman's disease

Abbreviations: CNS, central nervous system; HIV, human immunodeficiency virus; HTLV, human T cell lymphotropic virus; MALT, mucosa-associated lymphoid tissue; NK, natural killer.

EBV is associated with the development of Burkitt's lymphoma in Central Africa and the occurrence of aggressive non-Hodgkin's lymphomas in immunosuppressed patients in Western countries. The majority of primary central nervous system (CNS) lymphomas are associated with EBV. EBV infection is strongly associated with the occurrence of extranodal nasal T/NK cell lymphomas in Asia and South America. Infection with HIV predisposes to the development of aggressive, B cell non-Hodgkin's lymphoma. This may be through overexpression of interleukin 6 by infected macrophages. Infection of the stomach by the bacterium *Helicobacter pylori* induces the development of gastric MALT (mucosa-associated lymphoid tissue) lymphomas. This association is supported by evidence that patients treated with antibiotics to eradicate *H. pylori* have regression of their MALT lymphoma. The bacterium does not transform lymphocytes to produce the lymphoma; instead, a vigorous immune response is made to the bacterium, and the chronic antigenic stimulation leads to the neoplasia. MALT lymphomas of the skin may be related to *Borrelia* sp. infections, those of the eyes to *Chlamydophila psittaci*, and those of the small intestine to *Campylobacter jejuni*.

Chronic hepatitis C virus infection has been associated with the development of lymphoplasmacytic lymphoma. Human herpesvirus 8 is associated with primary effusion lymphoma in HIV-infected persons and multicentric Castleman's disease, a diffuse lymphadenopathy associated with systemic symptoms of fever, malaise, and weight loss.

In addition to infectious agents, a number of other diseases or exposures may predispose to developing lymphoma **(Table 110–5)**.

Table 110–5 Diseases or Exposures Associated with Increased Risk of Development of Malignant Lymphoma

Inherited immunodeficiency disease	Autoimmune disease
Klinefelter's syndrome	Sjögren's syndrome
Chédiak-Higashi syndrome	Celiac sprue

Ataxia-telangiectasia syndrome	Rheumatoid arthritis and systemic lupus erythematosus
Wiskott-Aldrich syndrome	Chemical or drug exposures
Common variable	Phenytoin
immunodeficiency disease	Dioxin, Phenoxy herbicides
Acquired immunodeficiency diseases	Radiation
Iatrogenic immunosuppression	Prior chemotherapy and radiation therapy
HIV-1 infection	
Acquired hypogammaglobulinemia	

Immunology

All lymphoid cells are derived from a common hematopoietic progenitor that gives rise to lymphoid, myeloid, erythroid, monocyte, and megakaryocyte lineages. Through the ordered and sequential activation of a series of transcription factors, the cell first becomes committed to the lymphoid lineage and then gives rise to B and T cells. About 75% of all lymphoid leukemias and 90% of all lymphomas are of B cell origin. A cell becomes committed to B cell development when it begins to rearrange its immunoglobulin genes. The sequence of cellular changes, including changes in cell-surface phenotype, that characterizes normal B cell development is shown in Fig. 110-2. A cell becomes committed to T cell differentiation upon migration to the thymus and rearrangement of T cell antigen receptor genes. The sequence of the events that characterize T cell development is depicted in Fig. 110-3.

Figure 110-2

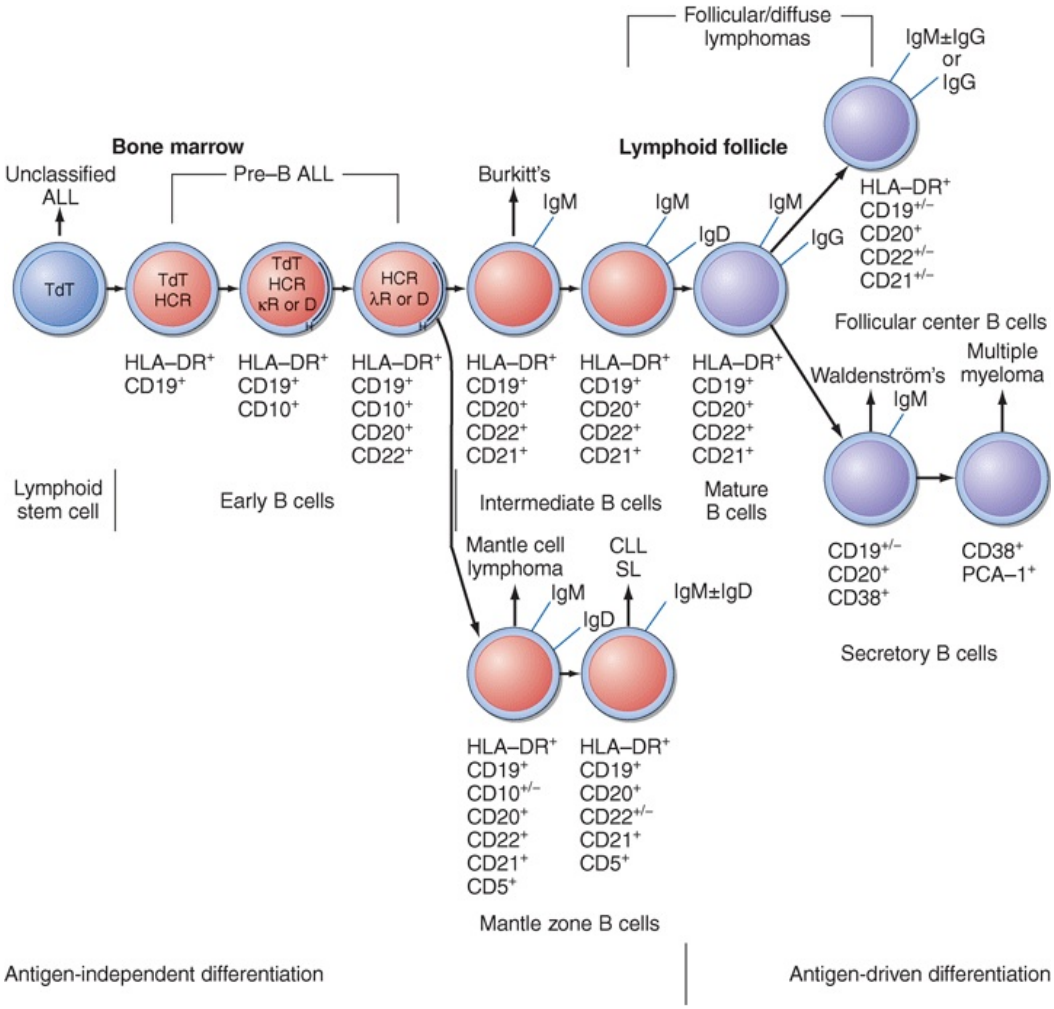

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathway of normal B cell differentiation and relationship to B cell lymphomas. HLA-DR, CD10, CD19, CD20, CD21, CD22, CD5, and CD38 are cell markers used to distinguish stages of development. Terminal transferase (TdT) is a cellular enzyme. Immunoglobulin heavy chain gene rearrangement (HCR) and light chain gene rearrangement or deletion (κR or D, λR or D) occur early in B cell development. The approximate normal stage of differentiation associated with particular lymphomas is shown. ALL, acute lymphoid leukemia; CLL, chronic lymphoid leukemia; SL, small lymphocytic lymphoma.

Figure 110-3

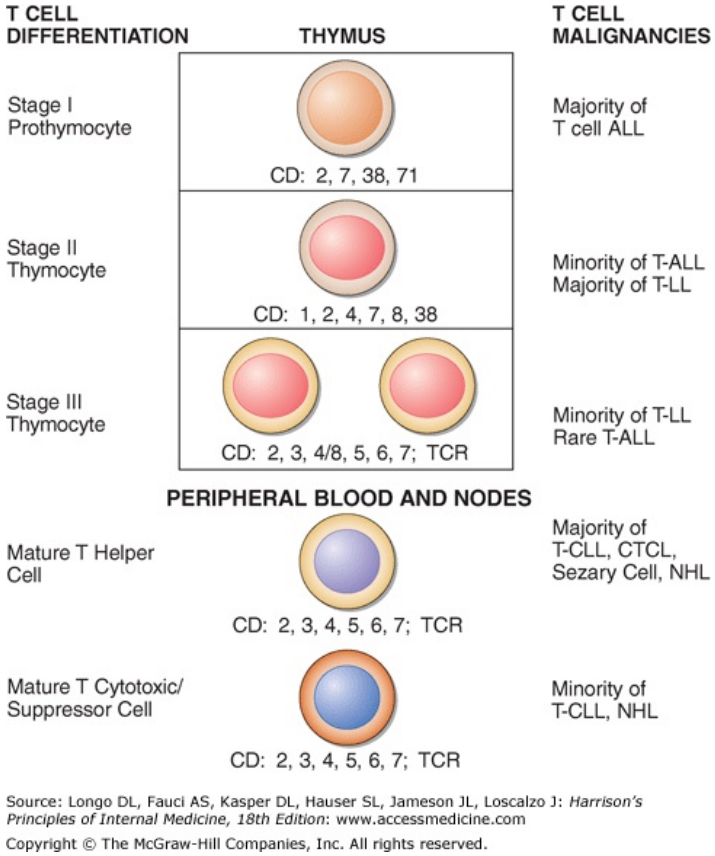

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathway of normal T cell differentiation and relationship to T cell lymphomas. CD1, CD2, CD3, CD4, CD5, CD6, CD7, CD8, CD38, and CD71 are cell markers used to distinguish stages of development. T cell antigen receptors (TCR) rearrange in the thymus, and mature T cells emigrate to nodes and peripheral blood. ALL, acute lymphoid leukemia; T-ALL, T cell ALL; T-LL, T cell lymphoblastic lymphoma; T-CLL, T cell chronic lymphoid leukemia; CTCL, cutaneous T cell lymphoma; NHL, non-Hodgkin's lymphoma.

Although lymphoid malignancies often retain the cell-surface phenotype of lymphoid cells at particular stages of differentiation, this information is of little consequence. The so-called stage of differentiation of a malignant lymphoma does not predict its natural history. For example, the clinically most aggressive lymphoid leukemia is Burkitt's leukemia, which has the phenotype of a mature follicle center IgM-bearing B cell. Leukemias bearing the immunologic cell-surface phenotype of more primitive cells (e.g., pre-B ALL, CD10+) are less aggressive and more amenable to curative therapy than the "more mature" appearing Burkitt's leukemia cells. Furthermore, the apparent stage of differentiation of the malignant cell does not reflect the stage at which the genetic lesions that gave rise to the malignancy developed. For example, follicular lymphoma has the cell-surface phenotype of a follicle center cell, but its characteristic chromosomal translocation, the t(14;18), which involves juxtaposition of the antiapoptotic *bcl-2* gene next to the immunoglobulin heavy chain gene (see below), had to develop early in ontogeny as an error in the process of immunoglobulin gene rearrangement. Why the subsequent steps that led to transformation became manifest in a cell of follicle center differentiation is not clear.

The major value of cell-surface phenotyping is to aid in the differential diagnosis of lymphoid tumors that appear similar by light microscopy. For example, benign follicular hyperplasia may resemble follicular lymphoma; however, the demonstration that all the cells bear the same immunoglobulin light chain isotype strongly suggests the mass is a clonal proliferation rather than a polyclonal response to an exogenous stimulus.

Malignancies of lymphoid cells are associated with recurring genetic abnormalities. While specific genetic abnormalities have not been identified for all subtypes of lymphoid malignancies, it is presumed that they exist. Genetic abnormalities can be identified at a variety of levels including gross chromosomal changes (i.e., translocations, additions, or deletions); rearrangement of specific genes that may or may not be apparent from cytogenetic studies; and overexpression, underexpression, or mutation of specific oncogenes. Altered expression or mutation of specific proteins is particularly important. Many lymphomas contain balanced chromosomal translocations involving the antigen receptor genes; immunoglobulin genes on chromosomes 2, 14, and 22 in B cells; and T cell antigen receptor genes on chromosomes 7 and 14 in T cells. The rearrangement of chromosome segments to generate mature antigen receptors must create a site of vulnerability to aberrant recombination. B cells are even more susceptible to acquiring mutations during their maturation in germinal centers; the generation of antibody of higher affinity requires the introduction of mutations into the variable region genes in the germinal centers. Other nonimmunoglobulin genes, e.g., *bcl-*

6, may acquire mutations as well.

In the case of diffuse large B cell lymphoma, the translocation t(14;18) occurs in ~30% of patients and leads to overexpression of the *bcl-2* gene found on chromosome 18. Some other patients without the translocation also overexpress the BCL-2 protein. This protein is involved in suppressing apoptosis—i.e., the mechanism of cell death most often induced by cytotoxic chemotherapeutic agents. A higher relapse rate has been observed in patients whose tumors overexpress the BCL-2 protein, but not in those patients whose lymphoma cells show only the translocation. Thus, particular genetic mechanisms have clinical ramifications.

Table 110–6 presents the best documented translocations and associated oncogenes for various subtypes of lymphoid malignancies. In some cases, such as the association of the t(14;18) in follicular lymphoma, the t(2;5) in anaplastic large T/null cell lymphoma, the t(8;14) in Burkitt's lymphoma, and the t(11;14) in mantle cell lymphoma, the great majority of tumors in patients with these diagnoses display these abnormalities. In other types of lymphoma where a minority of the patients have tumors expressing specific genetic abnormalities, the defects may have prognostic significance. No specific genetic abnormalities have been identified in Hodgkin's disease other than aneuploidy.

Table 110–6 Cytogenetic Translocation and Associated Oncogenes Often Seen in Lymphoid Malignancies

Disease	Cytogenetic Abnormality	Oncogene
CLL/small lymphocytic lymphoma	t(14;15)(q32;q13)	—
MALT lymphoma	t(11;18)(q21;q21)	API2/MALT, BCL-10
Precursor B cell acute lymphoid leukemia	t(9;22)(q34;q11) or variant t(4;11)(q21;q23)	BCR/ABL AF4, ALLI
Precursor acute lymphoid leukemia	t(9;22) t(1;19) t(17;19) t(5;14)	BCR, ABL E2A, PBX HLF, E2A HOX11L2, CTIP2
Mantle cell lymphoma	t(11;14)(q13;q32)	BCL-1, IgH
Follicular lymphoma	t(14;18)(q32;q21)	BCL-2, IgH
Diffuse large cell lymphoma Burkitt's lymphoma, Burkitt's leukemia	t(3;-)(q27;-)[a] t(17;-)(p13;-) t(8;-)(q24;-)[a]	BCL-6 p53 C-MYC
CD30+ Anaplastic large cell lymphoma	t(2;5)(p23;q35)	ALK
Lymphoplasmacytoid lymphoma	t(9;14)(p13;q32)	PAX5, IgH

[a]Numerous sites of translocation may be involved with these genes.

Abbreviations: CLL, chronic lymphoid leukemia; IgH, immunoglobulin heavy chain; MALT, mucosa-associated lymphoid tissue.

In typical B cell CLL, trisomy 12 conveys a poorer prognosis. In ALL in both adults and children, genetic abnormalities have important prognostic significance. Patients whose tumor cells display the t(9;22) and translocations involving the *MLL* gene on chromosome 11q23 have a much poorer outlook than patients who do not have these translocations. Other genetic abnormalities that occur frequently in adults with ALL include the t(4;11) and the t(8;14). The t(4;11) is associated with younger age, female predominance, high white cell counts, and L1 morphology. The t(8;14) is associated with older age, male predominance, frequent CNS involvement, and L3 morphology. Both are associated with a poor prognosis. In childhood ALL, hyperdiploidy has been shown to have a favorable prognosis.

Gene profiling using array technology allows the simultaneous assessment of the expression of thousands of genes. This technology provides the possibility to identify new genes with pathologic importance in lymphomas, the identification of patterns of gene expression with diagnostic and/or prognostic significance, and the identification of new therapeutic targets. Recognition of patterns of gene expression is complicated and requires sophisticated mathematical techniques. Early successes using this technology in lymphoma include the identification of previously unrecognized subtypes of diffuse large B cell lymphoma whose gene expression patterns resemble either those of follicular center B cells or activated peripheral blood B cells. Patients whose lymphomas have a germinal center B cell pattern of gene expression have a considerably better prognosis than those whose lymphomas have a pattern resembling activated peripheral blood B cells. This improved prognosis is independent of other known prognostic factors. Similar information is being generated in follicular lymphoma and mantle cell lymphoma. The challenge remains to provide information from such techniques in a clinically useful time frame.

Approach to the Patient: Lymphoid Cell Malignancies

Regardless of the type of lymphoid malignancy, the initial evaluation of the patient should include performance of a careful history and physical examination. These will help confirm the diagnosis, identify

those manifestations of the disease that might require prompt attention, and aid in the selection of further studies to optimally characterize the patient's status to allow the best choice of therapy. It is difficult to overemphasize the importance of a carefully done history and physical examination. They might provide observations that lead to reconsidering the diagnosis, provide hints at etiology, clarify the stage, and allow the physician to establish rapport with the patient that will make it possible to develop and carry out a therapeutic plan.

For patients with ALL, evaluation is usually completed after a complete blood count, chemistry studies reflecting major organ function, a bone marrow biopsy with genetic and immunologic studies, and a lumbar puncture. The latter is necessary to rule out occult CNS involvement. At this point, most patients would be ready to begin therapy. In ALL, prognosis is dependent upon the genetic characteristics of the tumor, the patient's age, the white cell count, and the patient's overall clinical status and major organ function.

In CLL, the patient evaluation should include a complete blood count, chemistry tests to measure major organ function, serum protein electrophoresis, and a bone marrow biopsy. However, some physicians believe that the diagnosis would not always require a bone marrow biopsy. Patients often have imaging studies of the chest and abdomen looking for pathologic lymphadenopathy. Patients with typical B cell CLL can be subdivided into three major prognostic groups. Those patients with only blood and bone marrow involvement by leukemia but no lymphadenopathy, organomegaly, or signs of bone marrow failure have the best prognosis. Those with lymphadenopathy and organomegaly have an intermediate prognosis, and patients with bone marrow failure, defined as hemoglobin <100 g/L (10 g/dL) or platelet count <100,000/μL, have the worst prognosis. The pathogenesis of the anemia or thrombocytopenia is important to discern. The prognosis is adversely affected when either or both of these abnormalities are due to progressive marrow infiltration and loss of productive marrow. However, either or both may be due to autoimmune phenomena or to hypersplenism that can develop during the course of the disease. These destructive mechanisms are usually completely reversible (glucocorticoids for autoimmune disease; splenectomy for hypersplenism) and do not influence disease prognosis.

Two popular staging systems have been developed to reflect these prognostic groupings **(Table 110–7)**. Patients with typical B cell CLL can have their course complicated by immunologic abnormalities, including autoimmune hemolytic anemia, autoimmune thrombocytopenia, and hypogammaglobulinemia. Patients with hypogammaglobulinemia benefit from regular (monthly) γ globulin administration. Because of expense, γ globulin is often withheld until the patient experiences a significant infection. These abnormalities do not have a clear prognostic significance and should not be used to assign a higher stage.

Table 110–7 Staging of Typical B Cell Lymphoid Leukemia

Stage	Clinical Features	Median Survival, Years
Rai System		
0: Low risk	Lymphocytosis only in blood and marrow	>10
I: Intermediate risk	Lymphocytosis + lymphadenopathy	>7
II: Intermediate risk	Lymphocytosis + lymphadenopathy + splenomegaly ± hepatomegaly	
III: High risk	Lymphocytosis + anemia	1.5
IV: High risk	Lymphocytosis + thrombocytopenia	
Binet System		
A	Fewer than three areas of clinical lymphadenopathy; no anemia or thrombocytopenia	>10
B	Three or more involved node areas; no anemia or thrombocytopenia	7
C	Hemoglobin ≤10 g/dL and/or platelets <100,000/μL	2

Two other features may be used to assess prognosis in B cell CLL, but neither has yet been incorporated into a staging classification. At least two subsets of CLL have been identified based on the cytoplasmic expression of ZAP-70; expression of this protein, which is usually expressed in T cells, identifies a subgroup with poorer prognosis. A less powerful subsetting tool is CD38 expression. CD38+ tumors tend to have a poorer prognosis than CD38– tumors.

The initial evaluation of a patient with Hodgkin's disease or non-Hodgkin's lymphoma is similar. In both situations, the determination of an accurate anatomic stage is an important part of the evaluation. Staging is done using the Ann Arbor staging system originally developed for Hodgkin's disease **(Table 110–8)**.

Table 110–8 The Ann Arbor Staging System for Hodgkin's Disease

Stage	Definition
I	Involvement of a single lymph node region or lymphoid structure (e.g., spleen, thymus, Waldeyer's ring)
II	Involvement of two or more lymph node regions on the same side of the diaphragm (the mediastinum is a single site; hilar lymph nodes should be considered "lateralized" and, when

	involved on both sides, constitute stage II disease)
III	Involvement of lymph node regions or lymphoid structures on both sides of the diaphragm
III_1	Subdiaphragmatic involvement limited to spleen, splenic hilar nodes, celiac nodes, or portal nodes
III_2	Subdiaphragmatic involvement includes paraaortic, iliac, or mesenteric nodes plus structures in III_1
IV	Involvement of extranodal site(s) beyond that designated as "E"
	More than one extranodal deposit at any location
	Any involvement of liver or bone marrow
A	No symptoms
B	Unexplained weight loss of >10% of the body weight during the 6 months before staging investigation
	Unexplained, persistent, or recurrent fever with temperatures >38°C during the previous month
	Recurrent drenching night sweats during the previous month
E	Localized, solitary involvement of extralymphatic tissue, excluding liver and bone marrow

Evaluation of patients with Hodgkin's disease will typically include a complete blood count; erythrocyte sedimentation rate; chemistry studies reflecting major organ function; CT scans of the chest, abdomen, and pelvis; and a bone marrow biopsy. Neither a positron emission tomography (PET) scan nor a gallium scan is absolutely necessary for primary staging, but one performed at the completion of therapy allows evaluation of persisting radiographic abnormalities, particularly the mediastinum. Knowing that the PET scan or gallium scan is abnormal before treatment can help in this assessment. In most cases, these studies will allow assignment of anatomic stage and the development of a therapeutic plan.

In patients with non-Hodgkin's lymphoma, the same evaluation described for patients with Hodgkin's disease is usually carried out. In addition, serum levels of lactate dehydrogenase (LDH) and β_2-microglobulin and serum protein electrophoresis are often included in the evaluation. Anatomic stage is assigned in the same manner as used for Hodgkin's disease. However, the prognosis of patients with non-Hodgkin's lymphoma is best assigned using the International Prognostic Index (IPI) **(Table 110–9)**. This is a powerful predictor of outcome in all subtypes of non-Hodgkin's lymphoma. Patients are assigned an IPI score based on the presence or absence of five adverse prognostic factors and may have none or all five of these adverse prognostic factors. **Figure 110-4** shows the prognostic significance of this score in 1300 patients with all types of non-Hodgkin's lymphoma. With the addition of rituximab to CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone), treatment outcomes have improved and the original IPI has lost some of its discrimination power. A revised IPI has been proposed that better predicts outcome of rituximab plus chemotherapy-based programs (Table 110–9). CT scans are routinely used in the evaluation of patients with all subtypes of non-Hodgkin's lymphoma, but PET and gallium scans are much more useful in aggressive subtypes such as diffuse large B cell lymphoma than in more indolent subtypes such as follicular lymphoma or small lymphocytic lymphoma. While the IPI does divide patients with follicular lymphoma into subsets with distinct prognoses, the distribution of such patients is skewed toward lower-risk categories. A follicular lymphoma–specific IPI (FLIPI) has been proposed that replaces performance status with hemoglobin level [<120 g/L (<12 g/dL)] and number of extranodal sites with number of nodal sites (more than four). Low risk (zero or one factor) was assigned to 36% of patients, intermediate risk (two factors) to 37%, and poor risk (more than two factors) to 27% of patients.

Table 110–9 International Prognostic Index for NHL

Five clinical risk factors:

- Age ≥ 60 years
- Serum lactate dehydrogenase levels elevated
- Performance status ≥2 (ECOG) or ≤ 70 (Karnofsky)
- Ann Arbor stage III or IV
- >1 site of extranodal involvement

Patients are assigned a number for each risk factor they have

Patients are grouped differently based upon the type of lymphoma

For diffuse large B cell lymphoma:

0, 1 factor = low risk:	35% of cases; 5-year survival, 73%
2 factors = low-intermediate risk:	27% of cases; 5-year survival, 51%
3 factors = high-intermediate risk:	22% of cases; 5-year survival, 43%
4, 5 factors = high risk:	16% of cases; 5-year survival, 26%

For diffuse large B cell lymphoma treated with R-CHOP:	
0 factor = very good:	10% of cases; 5-year survival, 94%
1, 2 factors = good:	45% of cases; 5-year survival, 79%
3, 4, 5 factors = poor:	45% of cases; 5-year survival, 55%

Abbreviations: ECOG, Eastern Cooperative Oncology Group; R-CHOP, rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone.

Figure 110-4

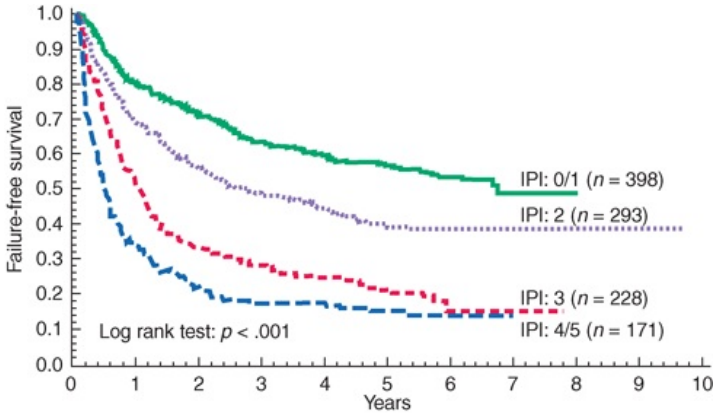

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship of International Prognostic Index (IPI) to survival. Kaplan-Meier survival curves for 1300 patients with various kinds of lymphoma stratified according to the IPI.

CLINICAL FEATURES, TREATMENT, AND PROGNOSIS OF SPECIFIC LYMPHOID MALIGNANCIES

Precursor Cell B Cell Neoplasms

PRECURSOR B CELL LYMPHOBLASTIC LEUKEMIA/LYMPHOMA

The most common cancer in childhood is B cell ALL. Although this disorder can also present as a lymphoma in either adults or children, presentation as lymphoma is rare.

The malignant cells in patients with precursor B cell lymphoblastic leukemia are most commonly of pre–B cell origin. Patients typically present with signs of bone marrow failure such as pallor, fatigue, bleeding, fever, and infection related to peripheral blood cytopenias. Peripheral blood counts regularly show anemia and thrombocytopenia but might show leukopenia, a normal leukocyte count, or leukocytosis based largely on the number of circulating malignant cells **(Fig. 110-5)**. Extramedullary sites of disease are frequently involved in patients who present with leukemia, including lymphadenopathy, hepato- or splenomegaly, CNS disease, testicular enlargement, and/or cutaneous infiltration.

Figure 110-5

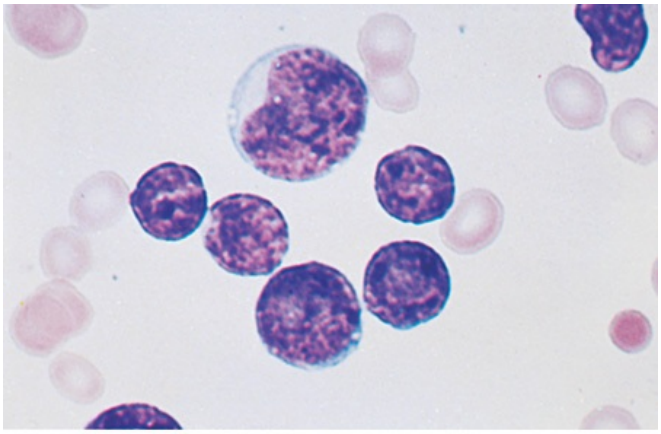

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute lymphoblastic leukemia. The cells are heterogeneous in size, have round or convoluted nuclei, high nuclear/cytoplasmic ratio, and absence of cytoplasmic granules.

The diagnosis is usually made by bone marrow biopsy, which shows infiltration by malignant lymphoblasts.

Demonstration of a pre–B cell immunophenotype (Fig. 110-2) and, often, characteristic cytogenetic abnormalities (Table 110–6) confirm the diagnosis. An adverse prognosis in patients with precursor B cell ALL is predicted by a very high white cell count, the presence of symptomatic CNS disease, and unfavorable cytogenetic abnormalities. For example, t(9;22), frequently found in adults with B cell ALL, has been associated with a very poor outlook. The bcr/abl kinase inhibitors have improved the prognosis.

TREATMENT: PRECURSOR B CELL LYMPHOBLASTIC LEUKEMIA

The treatment of patients with precursor B cell ALL involves remission induction with combination chemotherapy, a consolidation phase that includes administration of high-dose systemic therapy and treatment to eliminate disease in the CNS, and a period of continuing therapy to prevent relapse and effect cure. The overall cure rate in children is 90%, while ~50% of adults are long-term disease-free survivors. This reflects the high proportion of adverse cytogenetic abnormalities seen in adults with precursor B cell ALL.

Precursor B cell lymphoblastic lymphoma is a rare presentation of precursor B cell lymphoblastic malignancy. These patients often have a rapid transformation to leukemia and should be treated as though they had presented with leukemia. The few patients who present with the disease confined to lymph nodes have a high cure rate.

Mature (Peripheral) B Cell Neoplasms

B CELL CHRONIC LYMPHOID LEUKEMIA/SMALL LYMPHOCYTIC LYMPHOMA

B cell CLL/small lymphocytic lymphoma represents the most common lymphoid leukemia, and when presenting as a lymphoma, it accounts for ~7% of non-Hodgkin's lymphomas. Presentation can be as either leukemia or lymphoma. The major clinical characteristics of B cell CLL/small lymphocytic lymphoma are presented in Table 110–10.

Table 110–10 Clinical Characteristics of Patients with Common Types of Non-Hodgkin's Lymphomas (NHL)

Disease	Median Age, years	Frequency in Children	% Male	Stage I/II vs III/IV, %	B Symptoms, %	Bone Marrow Involvement, %	Gastrointestinal Tract Involvement, %	% Surviving 5 years
B cell chronic lymphocytic leukemia/small lymphocytic lymphoma	65	Rare	53	9 vs 91	33	72	3	51
Mantle cell lymphoma	63	Rare	74	20 vs 80	28	64	9	27
Extranodal marginal zone B cell lymphoma of MALT type	60	Rare	48	67 vs 33	19	14	50	74
Follicular lymphoma	59	Rare	42	33 vs 67	28	42	4	72
Diffuse large B cell lymphoma	64	~25% of childhood NHL	55	54 vs 46	33	16	18	46
Burkitt's lymphoma	31	~30% of childhood NHL	89	62 vs 38	22	33	11	45
Precursor T cell lymphoblastic lymphoma	28	~40% of childhood NHL	64	11 vs 89	21	50	4	26
Anaplastic large T/null cell lymphoma	34	Common	69	51 vs 49	53	13	9	77
Peripheral T cell non-Hodgkin's lymphoma	61	~5% of childhood NHL	55	20 vs 80	50	36	15	25

Abbreviation: MALT, mucosa-associated lymphoid tissue.

The diagnosis of typical B cell CLL is made when an increased number of circulating lymphocytes (i.e., >4 x 10^9/L and usually >10 x 10^9/L) is found **(Fig. 110-6)** that are monoclonal B cells expressing the CD5 antigen. Finding bone marrow infiltration by the same cells confirms the diagnosis. The peripheral blood smear in

such patients typically shows many "smudge" or "basket" cells, nuclear remnants of cells damaged by the physical shear stress of making the blood smear. If cytogenetic studies are performed, trisomy 12 is found in 25–30% of patients. Abnormalities in chromosome 13 are also seen.

Figure 110-6

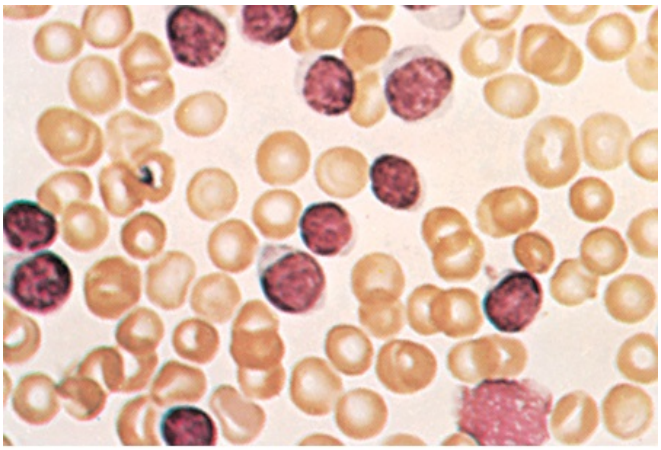

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic lymphocytic leukemia. The peripheral white blood cell count is high due to increased numbers of small, well–differentiated, normal-appearing lymphocytes. The leukemia lymphocytes are fragile, and substantial numbers of broken, smudged cells are usually also present on the blood smear.

If the primary presentation is lymphadenopathy and a lymph node biopsy is performed, pathologists usually have little difficulty in making the diagnosis of small lymphocytic lymphoma based on morphologic findings and immunophenotype. However, even in these patients, 70–75% will be found to have bone marrow involvement and circulating monoclonal B lymphocytes are often present.

The differential diagnosis of typical B cell CLL is extensive (Table 110–1). Immunophenotyping will eliminate the T cell disorders and can often help sort out other B cell malignancies. For example, only mantle cell lymphoma and typical B cell CLL are usually CD5 positive. Typical B cell small lymphocytic lymphoma can be confused with other B cell disorders, including lymphoplasmacytic lymphoma (i.e., the tissue manifestation of Waldenström's macroglobulinemia), nodal marginal zone B cell lymphoma, and mantle cell lymphoma. In addition, some small lymphocytic lymphomas have areas of large cells that can lead to confusion with diffuse large B cell lymphoma. An expert hematopathologist is vital for making this distinction.

Typical B cell CLL is often found incidentally when a complete blood count is done for another reason. However, complaints that might lead to the diagnosis include fatigue, frequent infections, and new lymphadenopathy. The diagnosis of typical B cell CLL should be considered in a patient presenting with an autoimmune hemolytic anemia or autoimmune thrombocytopenia. B cell CLL has also been associated with red cell aplasia. When this disorder presents as lymphoma, the most common abnormality is asymptomatic lymphadenopathy, with or without splenomegaly. The staging systems predict prognosis in patients with typical B cell CLL (Table 110–7). The evaluation of a new patient with typical B cell CLL/small lymphocytic lymphoma will include many of the studies **(Table 110–11)** that are used in patients with other non-Hodgkin's lymphomas. In addition, particular attention needs to be given to detecting immune abnormalities such as autoimmune hemolytic anemia, autoimmune thrombocytopenia, hypogammaglobulinemia, and red cell aplasia. Molecular analysis of immunoglobulin gene sequences in CLL has demonstrated that about half the patients have tumors expressing mutated immunoglobulin genes and half have tumors expressing unmutated or germ-line immunoglobulin sequences. Patients with unmutated immunoglobulins tend to have a more aggressive clinical course and are less responsive to therapy. Unfortunately, immunoglobulin gene sequencing is not routinely available. CD38 expression is said to be low in the better-prognosis patients expressing mutated immunoglobulin and high in poorer-prognosis patients expressing unmutated immunoglobulin, but this test has not been confirmed as a reliable means of distinguishing the two groups. ZAP-70 expression correlates with the presence of unmutated immunoglobulin genes, but the assay is not yet standardized and widely available.

Table 110–11 Staging Evaluation for Non-Hodgkin's Lymphoma

Physical examination
Documentation of B symptoms
Laboratory evaluation
Complete blood counts
Liver function tests
Uric acid

Calcium
Serum protein electrophoresis
Serum β_2-microglobulin
Chest radiograph
CT scan of abdomen, pelvis, and usually chest
Bone marrow biopsy
Lumbar puncture in lymphoblastic, Burkitt's, and diffuse large B cell lymphoma with positive marrow biopsy
Gallium scan (SPECT) or PET scan in large cell lymphoma

Abbreviations: PET, positron emission tomography; SPECT, single photon emission CT.

TREATMENT: B CELL CHRONIC LYMPHOID LEUKEMIA/SMALL LYMPHOCYTIC LYMPHOMA

Patients whose presentation is typical B cell CLL with no manifestations of the disease other than bone marrow involvement and lymphocytosis (i.e., Rai stage 0 and Binet stage A; Table 110–7) can be followed without specific therapy for their malignancy. These patients have a median survival >10 years, and some will never require therapy for this disorder. If the patient has an adequate number of circulating normal blood cells and is asymptomatic, many physicians would not initiate therapy for patients in the intermediate stage of the disease manifested by lymphadenopathy and/or hepatosplenomegaly. However, the median survival for these patients is ~7 years, and most will require treatment in the first few years of follow-up. Patients who present with bone marrow failure (i.e., Rai stage III or IV or Binet stage C) will require initial therapy in almost all cases. These patients have a serious disorder with a median survival of only 1.5 years. It must be remembered that immune manifestations of typical B cell CLL should be managed independently of specific antileukemia therapy. For example, glucocorticoid therapy for autoimmune cytopenias and γ globulin replacement for patients with hypogammaglobulinemia should be used whether or not antileukemia therapy is given.

Patients who present primarily with lymphoma and have a low IPI score have a 5-year survival of ~75%, but those with a high IPI score have a 5-year survival of <40% and are more likely to require early therapy.

The most common treatments for patients with typical B cell CLL/small lymphocytic lymphoma have been chlorambucil or fludarabine, alone or in combination. Chlorambucil can be administered orally with few immediate side effects, while fludarabine is administered IV and is associated with significant immune suppression. However, fludarabine is by far the more active agent and is the only drug associated with a significant incidence of complete remission. The combination of rituximab (375–500 mg/m^2 day 1), fludarabine (25 mg/m^2 days 2–4 on cycle 1 and 1–3 in subsequent cycles), and cyclophosphamide (250 mg/m^2 with fludarabine) achieves complete responses in 69% of patients, and those responses are associated with molecular remissions in half of the cases. Half the patients experience grade III or IV neutropenia. For young patients presenting with leukemia requiring therapy, regimens containing fludarabine are the treatment of choice. Because fludarabine is an effective second-line agent in patients with tumors unresponsive to chlorambucil, the latter agent is often chosen in elderly patients who require therapy. Bendamustine, an alkylating agent structurally related to nitrogen mustard, is highly effective and is vying with fludarabine as the primary treatment of choice. Patients who present with lymphoma (rather than leukemia) are also highly responsive to bendamustine and some patients will receive a combination chemotherapy regimen used in other lymphomas such as CVP (cyclophosphamide, vincristine, and prednisone) or CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone) plus rituximab. Alemtuzumab (anti-CD52) is an antibody with activity in the disease, but it kills both B and T cells and is associated with more immune compromise than rituximab. Young patients with this disease can be candidates for bone marrow transplantation. Allogeneic bone marrow transplantation can be curative but is associated with a significant treatment-related mortality rate. Mini-transplants using immunosuppressive rather than myeloablative doses of preparative drugs are being studied (Chap. 114). The use of autologous transplantation in patients with this disorder has been discouraging.

EXTRANODAL MARGINAL ZONE B CELL LYMPHOMA OF MALT TYPE

Extranodal marginal zone B cell lymphoma of MALT type (MALT lymphoma) makes up ~8% of non-Hodgkin's lymphomas. This small cell lymphoma presents in extranodal sites. It was previously considered a small lymphocytic lymphoma or sometimes a pseudolymphoma. The recognition that the gastric presentation of this lymphoma was associated with *H. pylori* infection was an important step in recognizing it as a separate entity. The clinical characteristics of MALT lymphoma are presented in Table 110–10.

The diagnosis of MALT lymphoma can be made accurately by an expert hematopathologist based on a characteristic pattern of infiltration of small lymphocytes that are monoclonal B cells and CD5 negative. In some cases, transformation to diffuse large B cell lymphoma occurs, and both diagnoses may be made in the same biopsy. The differential diagnosis includes benign lymphocytic infiltration of extranodal organs and other small cell B cell lymphomas.

MALT lymphoma may occur in the stomach, orbit, intestine, lung, thyroid, salivary gland, skin, soft tissues, bladder, kidney, and CNS. It may present as a new mass, be found on routine imaging studies, or be associated with local symptoms such as upper abdominal discomfort in gastric lymphoma. Most MALT lymphomas are gastric in origin. At least two genetic forms of gastric MALT exist: one (accounting for ~50%

of cases) characterized by t(11;18)(q21;q21) that juxtaposes the amino terminal of the *API2* gene with the carboxy terminal of the *MALT1* gene creating an API2/MALT1 fusion product, and the other characterized by multiple sites of genetic instability including trisomies of chromosomes 3, 7, 12, and 18. About 95% of gastric MALT lymphomas are associated with *H. pylori* infection, and those that are do not usually express t(11;18). The t(11;18) usually results in activation of NF-κB, which acts as a survival factor for the cells. Lymphomas with t(11;18) translocations are genetically stable and do not evolve to diffuse large B cell lymphoma. By contrast, t(11;18)-negative MALT lymphomas often acquire *BCL6* mutations and progress to aggressive histology lymphoma. MALT lymphomas are localized to the organ of origin in ~40% of cases and to the organ and regional lymph nodes in ~30% of patients. However, distant metastasis can occur—particularly with transformation to diffuse large B cell lymphoma. Many patients who develop this lymphoma will have an autoimmune or inflammatory process such as Sjögren's syndrome (salivary gland MALT), Hashimoto's thyroiditis (thyroid MALT), *Helicobacter* gastritis (gastric MALT), *C. psittaci* conjunctivitis (ocular MALT), or *Borrelia* skin infections (cutaneous MALT).

Evaluation of patients with MALT lymphoma follows the pattern (Table 110–11) for staging a patient with non-Hodgkin's lymphoma. In particular, patients with gastric lymphoma need to have studies performed to document the presence or absence of *H. pylori* infection. Endoscopic studies including ultrasound can help define the extent of gastric involvement. Most patients with MALT lymphoma have a good prognosis, with a 5-year survival of ~75%. In patients with a low IPI score, the 5-year survival is ~90%, while it drops to ~40% in patients with a high IPI score.

TREATMENT: MUCOSA-ASSOCIATED LYMPHOID TISSUE LYMPHOMA

MALT lymphoma is often localized. Patients with gastric MALT lymphomas who are infected with *H. pylori* can achieve remission in the 80% of cases with eradication of the infection. These remissions can be durable, but molecular evidence of persisting neoplasia is not infrequent. After *H. pylori* eradication, symptoms generally improve quickly but molecular evidence of persistent disease may be present for 12–18 months. Additional therapy is not indicated unless progressive disease is documented. Patients with more extensive disease or progressive disease are most often treated with single-agent chemotherapy such as chlorambucil. Combination regimens that include rituximab are also highly effective. Coexistent diffuse large B cell lymphoma must be treated with combination chemotherapy (see below). The additional acquired mutations that mediate the histologic progression also convey *Helicobacter* independence to the growth.

MANTLE CELL LYMPHOMA

Mantle cell lymphoma makes up ~6% of all non-Hodgkin's lymphomas. This lymphoma was previously placed in a number of other subtypes. Its existence was confirmed by the recognition that these lymphomas have a characteristic chromosomal translocation, t(11;14), between the immunoglobulin heavy chain gene on chromosome 14 and the *bcl-1* gene on chromosome 11, and regularly overexpress the BCL-1 protein, also known as cyclin D1. Table 110–10 shows the clinical characteristics of mantle cell lymphoma.

The diagnosis of mantle cell lymphoma can be made accurately by an expert hematopathologist. As with all subtypes of lymphoma, an adequate biopsy is important. The differential diagnosis of mantle cell lymphoma includes other small cell B cell lymphomas. In particular, mantle cell lymphoma and small lymphocytic lymphoma share a characteristic expression of CD5. Mantle cell lymphoma usually has a slightly indented nucleus.

The most common presentation of mantle cell lymphoma is with palpable lymphadenopathy, frequently accompanied by systemic symptoms. The median age is 63 years and men are affected four times as commonly as women. Approximately 70% of patients will be stage IV at the time of diagnosis, with frequent bone marrow and peripheral blood involvement. Of the extranodal organs that can be involved, gastrointestinal involvement is particularly important to recognize. Patients who present with lymphomatosis polyposis in the large intestine usually have mantle cell lymphoma. Table 110–11 outlines the evaluation of patients with mantle cell lymphoma. Patients who present with gastrointestinal tract involvement often have Waldeyer's ring involvement, and vice versa. The 5-year survival for all patients with mantle cell lymphoma is ~25%, with only occasional patients who present with a high IPI score surviving 5 years and ~50% of patients with a low IPI score surviving 5 years.

Treatment: Mantle Cell Lymphoma

Current therapies for mantle cell lymphoma are evolving. Patients with localized disease might be treated with combination chemotherapy followed by radiotherapy; however, these patients are exceedingly rare. For the usual presentation with disseminated disease, standard lymphoma treatments have been unsatisfactory, with the minority of patients achieving complete remission. Aggressive combination chemotherapy regimens followed by autologous or allogeneic bone marrow transplantation are frequently offered to younger patients. For the occasional elderly, asymptomatic patient, observation followed by single-agent chemotherapy might be the most practical approach. An intensive combination chemotherapy regimen originally used in the treatment of acute leukemia, HyperC-VAD (cyclophosphamide, vincristine, doxorubicin, dexamethasone, cytarabine, and methotrexate), in combination with rituximab, seems to be associated with better response rates—particularly in younger patients. Alternating two regimens, HyperC-VAD with rituximab added (R-HyperC-VAD) and rituximab plus high-dose methotrexate and cytarabine can achieve complete responses in >80% of patients and 8-year survival of 56%, comparable to regimens employing high-dose therapy and autologous hematopoietic stem cell transplantation. Bortezomib, temsirolimus, and bendamustine are single agents that induce transient partial responses in a minority of patients and are being added to primary combinations.

FOLLICULAR LYMPHOMA

Follicular lymphomas make up 22% of non-Hodgkin's lymphomas worldwide and at least 30% of non-Hodgkin's lymphomas diagnosed in the United States. This type of lymphoma can be diagnosed accurately on morphologic findings alone and has been the diagnosis in the majority of patients in therapeutic trials for "low-grade" lymphoma in the past. The clinical characteristics of follicular lymphoma are presented in Table 110–10.

Evaluation of an adequate biopsy by an expert hematopathologist is sufficient to make a diagnosis of follicular lymphoma. The tumor is composed of small cleaved and large cells in varying proportions organized in a follicular pattern of growth **(Fig. 110-7)**. Confirmation of B cell immunophenotype and the existence of the t(14;18) and abnormal expression of BCL-2 protein are confirmatory. The major differential diagnosis is between lymphoma and reactive follicular hyperplasia. The coexistence of diffuse large B cell lymphoma must be considered. Patients with follicular lymphoma are often subclassified into those with predominantly small cells, those with a mixture of small and large cells, and those with predominantly large cells. While this distinction cannot be made simply or very accurately, these subdivisions do have prognostic significance. Patients with follicular lymphoma with predominantly large cells have a higher proliferative fraction, progress more rapidly, and have a shorter overall survival with simple chemotherapy regimens.

Figure 110-7

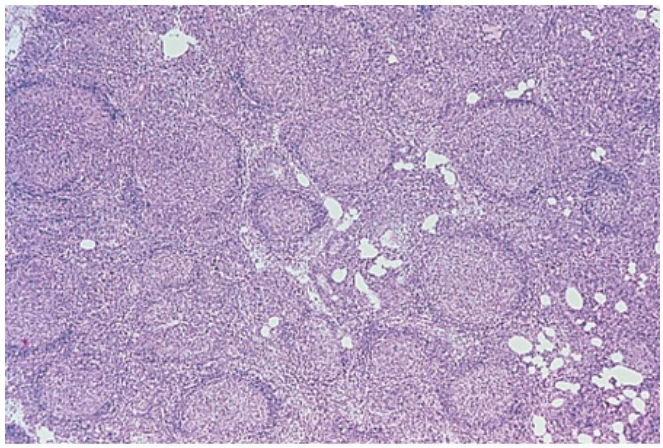

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Follicular lymphoma. The normal nodal architecture is -effaced by nodular expansions of tumor cells. Nodules vary in size and contain predominantly small lymphocytes with cleaved nuclei along with variable numbers of larger cells with vesicular chromatin and prominent nucleoli.

The most common presentation for follicular lymphoma is with new, painless lymphadenopathy. Multiple sites of lymphoid involvement are typical, and unusual sites such as epitrochlear nodes are sometimes seen. However, essentially any organ can be involved, and extranodal presentations do occur. Most patients do not have fevers, sweats, or weight loss, and an IPI score of 0 or 1 is found in ~50% of patients. Fewer than 10% of patients have a high (i.e., 4 or 5) IPI score. The staging evaluation for patients with follicular lymphoma should include the studies included in Table 110–11.

Treatment: Follicular Lymphoma

Follicular lymphoma is one of the malignancies most responsive to chemotherapy and radiotherapy. In addition, tumors in as many as 25% of the patients undergo spontaneous regression—usually transient—without therapy. In an asymptomatic patient, no initial treatment and watchful waiting can be an appropriate management strategy and is particularly likely to be adopted for older patients with advanced-stage disease. For patients who do require treatment, single-agent chlorambucil or cyclophosphamide or combination chemotherapy with CVP or CHOP are most frequently used. With adequate treatment, 50–75% of patients will achieve a complete remission. While most patients relapse (median response duration is ~2 years), at least 20% of complete responders will remain in remission for >10 years. For the rare patient (15%) with localized follicular lymphoma, involved field radiotherapy produces long-term disease-free survival in the majority.

A number of therapies have been shown to be active in the treatment of patients with follicular lymphoma. These include cytotoxic agents such as fludarabine, biologic agents such as interferon α, monoclonal antibodies with or without radionuclides, and lymphoma vaccines. In patients treated with a doxorubicin-containing combination chemotherapy regimen, interferon α given to patients in complete remission seems to prolong survival, but interferon toxicities can affect quality of life. The monoclonal antibody rituximab can cause objective responses in 35–50% of patients with relapsed follicular lymphoma, and radiolabeled antibodies appear to have response rates well in excess of 50%. The addition of rituximab to CHOP and other effective combination chemotherapy programs is beginning to show prolonged overall survival and a decreased risk of histologic progression. Complete remissions can be noted in 85% or more of patients treated with R-CHOP and median remission durations can exceed 6 or 7 years. Maintenance intermittent

rituximab therapy can prolong remissions even further, though it is not completely clear that overall survival is prolonged. Some trials with tumor vaccines have been encouraging. Both autologous and allogeneic hematopoietic stem cell transplantation yield high complete response rates in patients with relapsed follicular lymphoma, and long-term remissions can occur in 40% or more of patients.

Patients with follicular lymphoma with a predominance of large cells have a shorter survival when treated with single-agent chemotherapy but seem to benefit from receiving an anthracycline-containing combination chemotherapy regimen plus rituximab. When their disease is treated aggressively, the overall survival for such patients is no lower than for patients with other follicular lymphomas, and the failure-free survival is superior.

Patients with follicular lymphoma have a high rate of histologic transformation to diffuse large B cell lymphoma (5–7% per year). This is recognized ~40% of the time during the course of the illness by repeat biopsy and is present in almost all patients at autopsy. This transformation is usually heralded by rapid growth of lymph nodes—often localized—and the development of systemic symptoms such as fevers, sweats, and weight loss. Although these patients have a poor prognosis, aggressive combination chemotherapy regimens can sometimes cause a complete remission in the diffuse large B cell lymphoma, at times leaving the patient with persisting follicular lymphoma. With more frequent use of R-CHOP to treat follicular lymphoma at diagnosis, it appears that the rate of histologic progression is decreasing.

DIFFUSE LARGE B CELL LYMPHOMA

Diffuse large B cell lymphoma is the most common type of non-Hodgkin's lymphoma, representing approximately one-third of all cases. This lymphoma makes up the majority of cases in previous clinical trials of "aggressive" or "intermediate-grade" lymphoma. Table 110–10 shows the clinical characteristics of diffuse large B cell lymphoma.

The diagnosis of diffuse large B cell lymphoma can be made accurately by an expert hematopathologist **(Fig. 110-8)**. Cytogenetic and molecular genetic studies are not necessary for diagnosis, but some evidence has accumulated that patients whose tumors overexpress the BCL-2 protein might be more likely to relapse than others. Patients with prominent mediastinal involvement are sometimes diagnosed as a separate subgroup having primary mediastinal diffuse large B cell lymphoma. This latter group of patients has a younger median age (i.e., 37 years) and a female predominance (66%). Subtypes of diffuse large B cell lymphoma, including those with an immunoblastic subtype and tumors with extensive fibrosis, are recognized by pathologists but do not appear to have important independent prognostic significance.

Figure 110-8

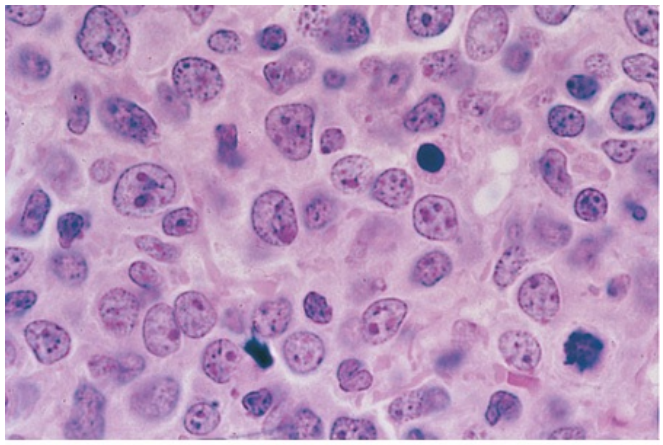

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diffuse large B cell lymphoma. The neoplastic cells are heterogeneous but predominantly large cells with vesicular chromatin and prominent nucleoli.

Diffuse large B cell lymphoma can present as either primary lymph node disease or at extranodal sites. More than 50% of patients will have some site of extranodal involvement at diagnosis, with the most common sites being the gastrointestinal tract and bone marrow, each being involved in 15–20% of patients. Essentially any organ can be involved, making a diagnostic biopsy imperative. For example, diffuse large B cell lymphoma of the pancreas has a much better prognosis than pancreatic carcinoma but would be missed without biopsy. Primary diffuse large B cell lymphoma of the brain is being diagnosed with increasing frequency. Other unusual subtypes of diffuse large B cell lymphoma such as pleural effusion lymphoma and intravascular lymphoma have been difficult to diagnose and associated with a very poor prognosis.

Table 110–11 shows the initial evaluation of patients with diffuse large B cell lymphoma. After a careful staging evaluation, ~50% of patients will be found to have stage I or II disease and ~50% will have widely disseminated lymphoma. Bone marrow biopsy shows involvement by lymphoma in ~15% of cases, with marrow involvement by small cells more frequent than by large cells.

Treatment: Diffuse Large B Cell Lymphoma

The initial treatment of all patients with diffuse large B cell lymphoma should be with a combination chemotherapy regimen. The most popular regimen in the United States is CHOP plus rituximab, although a variety of other anthracycline-containing combination chemotherapy regimens appear to be equally efficacious. Patients with stage I or nonbulky stage II disease can be effectively treated with three to four cycles of combination chemotherapy with or without subsequent involved field radiotherapy. The need for radiation therapy is unclear. Cure rates of 70–80% in stage II disease and 85–90% in stage I disease can be expected.

For patients with bulky stage II, stage III, or stage IV disease, six to eight cycles of CHOP plus rituximab are usually administered. A large randomized trial showed the superiority of CHOP combined with rituximab over CHOP alone in elderly patients. A frequent approach would be to administer four cycles of therapy and then reevaluate. If the patient has achieved a complete remission after four cycles, two more cycles of treatment might be given and then therapy discontinued. Using this approach, 70–80% of patients can be expected to achieve a complete remission, and 50–70% of complete responders will be cured. The chances for a favorable response to treatment are predicted by the IPI. In fact, the IPI was developed based on the outcome of patients with diffuse large B cell lymphoma treated with CHOP-like regimens. For the 35% of patients with a low IPI score of 0–1, the 5-year survival is >70%, while for the 20% of patients with a high IPI score of 4–5, the 5-year survival is ~20%. The addition of rituximab to CHOP has improved each of those numbers by ~15%. A number of other factors, including molecular features of the tumor, levels of circulating cytokines and soluble receptors, and other surrogate markers, have been shown to influence prognosis. However, they have not been validated as rigorously as the IPI and have not been uniformly applied clinically.

Because a number of patients with diffuse large B cell lymphoma are either initially refractory to therapy or relapse after apparently effective chemotherapy, 30–40% of patients will be candidates for salvage treatment at some point. Alternative combination chemotherapy regimens can induce complete remission in as many as 50% of these patients, but long-term disease-free survival is seen in ≤10%. Autologous bone marrow transplantation is superior to salvage chemotherapy at usual doses and leads to long-term disease-free survival in ~40% of patients whose lymphomas remain chemotherapy-sensitive after relapse.

BURKITT'S LYMPHOMA/LEUKEMIA

Burkitt's lymphoma/leukemia is a rare disease in adults in the United States, making up <1% of non-Hodgkin's lymphomas, but it makes up ~30% of childhood non-Hodgkin's lymphoma. Burkitt's leukemia, or L3 ALL, makes up a small proportion of childhood and adult acute leukemias. Table 110–10 shows the clinical features of Burkitt's lymphoma.

Burkitt's lymphoma can be diagnosed morphologically by an expert hematopathologist with a high degree of accuracy. The cells are homogeneous in size and shape **(Fig. 110-9)**. Demonstration of a very high proliferative fraction and the presence of the t(8;14) or one of its variants, t(2;8) (*c-myc* and the λ light chain gene) or t(8;22) (*c-myc* and the κ light chain gene), can be confirmatory. Burkitt's cell leukemia is recognized by the typical monotonous mass of medium-sized cells with round nuclei, multiple nucleoli, and basophilic cytoplasm with cytoplasmic vacuoles. Demonstration of surface expression of immunoglobulin and one of the above-noted cytogenetic abnormalities is confirmatory.

Figure 110-9

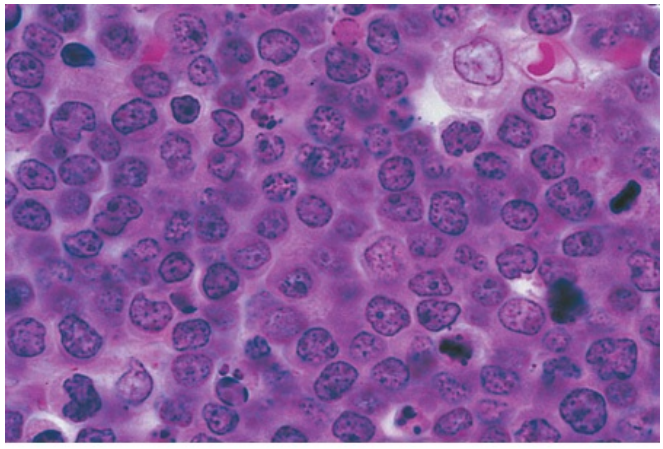

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Burkitt's lymphoma. The neoplastic cells are homogeneous, medium-sized B cells with frequent mitotic figures, a morphologic correlate of high growth fraction. Reactive macrophages are scattered through the tumor and their pale cytoplasm in a background of blue-staining tumor cells give the tumor a so-called starry sky appearance.

Three distinct clinical forms of Burkitt's lymphoma are recognized; endemic, sporadic, and immunodeficiency-associated. Endemic and sporadic Burkitt's lymphomas occur frequently in children in Africa, and the sporadic form in Western countries. Immunodeficiency-associated Burkitt's lymphoma is seen in patients with

HIV infection.

Pathologists sometimes have difficulty distinguishing between Burkitt's lymphoma and diffuse large B cell lymphoma. In the past, a separate subgroup of non-Hodgkin's lymphoma intermediate between the two was recognized. When tested, this subgroup could not be diagnosed accurately. Distinction between the two major types of B cell aggressive non-Hodgkin's lymphoma can sometimes be made based on the extremely high proliferative fraction seen in patients with Burkitt's lymphoma (i.e., essentially 100% of tumor cells are in cycle) caused by *c-myc* deregulation.

Most patients in the United States with Burkitt's lymphoma present with peripheral lymphadenopathy or an intraabdominal mass. The disease is rapidly progressive and has a propensity to metastasize to the CNS. Initial evaluation should always include an examination of cerebrospinal fluid to rule out metastasis in addition to the other staging evaluations noted in Table 110–11. Once the diagnosis of Burkitt's lymphoma is suspected, a diagnosis must be made promptly and staging evaluation must be accomplished expeditiously. This is the most rapidly progressive human tumor, and any delay in initiating therapy can adversely affect the patient's prognosis.

Treatment: Burkitt's Lymphoma

Treatment of Burkitt's lymphoma in both children and adults should begin within 48 h of diagnosis and involves the use of intensive combination chemotherapy regimens incorporating high doses of cyclophosphamide. Prophylactic therapy to the CNS is mandatory. Burkitt's lymphoma was one of the first cancers shown to be curable by chemotherapy. Today, cure can be expected in 70–80% of both children and young adults when effective therapy is administered precisely. Salvage therapy has been generally ineffective in patients failing the initial treatment, emphasizing the importance of the initial treatment approach.

OTHER B CELL LYMPHOID MALIGNANCIES

B cell prolymphocytic leukemia involves blood and marrow infiltration by large lymphocytes with prominent nucleoli. Patients typically have a high white cell count, splenomegaly, and minimal lymphadenopathy. The chances for a complete response to therapy are poor.

Hairy cell leukemia is a rare disease that presents predominantly in older males. Typical presentation involves pancytopenia, although occasional patients will have a leukemic presentation. Splenomegaly is usual. The malignant cells appear to have "hairy" projections on light and electron microscopy and show a characteristic staining pattern with tartrate-resistant acid phosphatase. Bone marrow is typically not able to be aspirated, and biopsy shows a pattern of fibrosis with diffuse infiltration by the malignant cells. Patients with this disorder are prone to unusual infections, including infection by *Mycobacterium avium intracellulare*, and to vasculitic syndromes. Hairy cell leukemia is responsive to chemotherapy with interferon α, pentostatin, or cladribine, with the latter being the usually preferred treatment. Clinical complete remissions with cladribine occur in the majority of patients, and long-term disease-free survival is frequent.

Splenic marginal zone lymphoma involves infiltration of the splenic white pulp by small, monoclonal B cells. This is a rare disorder that can present as leukemia as well as lymphoma. Definitive diagnosis is often made at splenectomy, which is also an effective therapy. This is an extremely indolent disorder, but when chemotherapy is required, the most usual treatment has been chlorambucil.

Lymphoplasmacytic lymphoma is the tissue manifestation of Waldenström's macroglobulinemia (Chap. 111). This type of lymphoma has been associated with chronic hepatitis C virus infection, and an etiologic association has been proposed. Patients typically present with lymphadenopathy, splenomegaly, bone marrow involvement, and occasionally peripheral blood involvement. The tumor cells do not express CD5. Patients often have a monoclonal IgM protein, high levels of which can dominate the clinical picture with the symptoms of hyperviscosity. Treatment of lymphoplasmacytic lymphoma can be aimed primarily at reducing the abnormal protein, if present, but will usually also involve chemotherapy. Chlorambucil, fludarabine, and cladribine have been utilized. The median 5-year survival for patients with this disorder is ~60%.

Nodal marginal zone lymphoma, also known as *monocytoid B cell lymphoma*, represents ~1% of non-Hodgkin's lymphomas. This lymphoma has a slight female predominance and presents with disseminated disease (i.e., stage III or IV) in 75% of patients. Approximately one-third of patients have bone marrow involvement, and a leukemic presentation occasionally occurs. The staging evaluation and therapy should use the same approach as used for patients with follicular lymphoma. Approximately 60% of the patients with nodal marginal zone lymphoma will survive 5 years after diagnosis.

Precursor T Cell Malignancies

PRECURSOR T CELL LYMPHOBLASTIC LEUKEMIA/LYMPHOMA

Precursor T cell malignancies can present either as ALL or as an aggressive lymphoma. These malignancies are more common in children and young adults, with males more frequently affected than females.

Precursor T cell ALL can present with bone marrow failure, although the severity of anemia, neutropenia, and thrombocytopenia is often less than in precursor B cell ALL. These patients sometimes have very high white cell counts, a mediastinal mass, lymphadenopathy, and hepatosplenomegaly. Precursor T cell lymphoblastic lymphoma is most often found in young men presenting with a large mediastinal mass and pleural effusions. Both presentations have a propensity to metastasize to the CNS, and CNS involvement is often present at diagnosis.

Treatment: Precursor T Cell Lymphoblastic Leukemia/Lymphoma

Children with precursor T cell ALL seem to benefit from very intensive remission induction and consolidation regimens. The majority of patients treated in this manner can be cured. Older children and young adults with precursor T cell lymphoblastic lymphoma are also often treated with "leukemia-like" regimens. Patients who present with localized disease have an excellent prognosis. However, advanced age is an adverse prognostic factor. Adults with precursor T cell lymphoblastic lymphoma who present with high LDH levels or bone marrow or CNS involvement are often offered bone marrow transplantation as part of their primary therapy.

Mature (Peripheral) T Cell Disorders

MYCOSIS FUNGOIDES

Mycosis fungoides is also known as *cutaneous T cell lymphoma*. This lymphoma is more often seen by dermatologists than internists. The median age of onset is in the midfifties, and the disease is more common in males and in blacks.

Mycosis fungoides is an indolent lymphoma with patients often having several years of eczematous or dermatitic skin lesions before the diagnosis is finally established. The skin lesions progress from patch stage to plaque stage to cutaneous tumors. Early in the disease, biopsies are often difficult to interpret, and the diagnosis may only become apparent by observing the patient over time. In advanced stages, the lymphoma can spread to lymph nodes and visceral organs. Patients with this lymphoma may develop generalized erythroderma and circulating tumor cells, called *Sézary's syndrome*.

Rare patients with localized early-stage mycosis fungoides can be cured with radiotherapy, often total-skin electron beam irradiation. More advanced disease has been treated with topical glucocorticoids, topical nitrogen mustard, phototherapy, psoralen with ultraviolet A (PUVA), extracorporeal photopheresis, retinoids (bexarotene), electron beam radiation, interferon, antibodies, fusion toxins, histone deacetylase inhibitors, and systemic cytotoxic therapy. Unfortunately, these treatments are palliative.

ADULT T CELL LYMPHOMA/LEUKEMIA

Adult T cell lymphoma/leukemia is one manifestation of infection by the HTLV-I retrovirus. Patients can be infected through transplacental transmission, mother's milk, blood transfusion, and by sexual transmission of the virus. Patients who acquire the virus from their mother through breast milk are most likely to develop lymphoma, but the risk is still only 2.5% and the latency averages 55 years. Nationwide testing for HTLV-I antibodies and the aggressive implementation of public health measures could theoretically lead to the disappearance of adult T cell lymphoma/leukemia. Tropical spastic paraparesis, another manifestation of HTLV-I infection (Chap. 188), occurs after a shorter latency (1–3 years) and is most common in individuals who acquire the virus during adulthood from transfusion or sex.

The diagnosis of adult T cell lymphoma/leukemia is made when an expert hematopathologist recognizes the typical morphologic picture, a T cell immunophenotype (i.e., CD4 positive), and the presence in serum of antibodies to HTLV-I. Examination of the peripheral blood will usually reveal characteristic, pleomorphic abnormal CD4-positive cells with indented nuclei, which have been called "flower" cells **(Fig. 110-10)**.

Figure 110-10

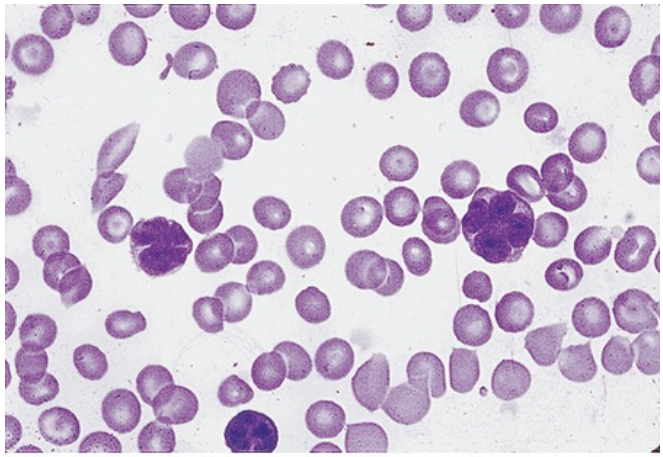

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adult T cell leukemia/lymphoma. Peripheral blood smear showing leukemia cells with typical "flower-shaped" nucleus.

A subset of patients have a smoldering clinical course and long survival, but most patients present with an aggressive disease manifested by lymphadenopathy, hepatosplenomegaly, skin infiltration, pulmonary infiltrates, hypercalcemia, lytic bone lesions, and elevated LDH levels. The skin lesions can be papules, plaques, tumors, and ulcerations. Lung lesions can be either tumor or opportunistic infection in light of the underlying immunodeficiency in the disease. Bone marrow involvement is not usually extensive, and anemia and thrombocytopenia are not usually prominent. Although treatment with combination chemotherapy

regimens can result in objective responses, true complete remissions are unusual, and the median survival of patients is ~7 months. A small phase II study reported a high response rate with interferon plus zidovudine and arsenic trioxide.

ANAPLASTIC LARGE T/NULL CELL LYMPHOMA

Anaplastic large T/null cell lymphoma was previously usually diagnosed as undifferentiated carcinoma or malignant histiocytosis. Discovery of the CD30 (Ki-1) antigen and the recognition that some patients with previously unclassified malignancies displayed this antigen led to the identification of a new type of lymphoma. Subsequently, discovery of the t(2;5) and the resultant frequent overexpression of the anaplastic lymphoma kinase (ALK) protein confirmed the existence of this entity. This lymphoma accounts for ~2% of all non-Hodgkin's lymphomas. Table 110–10 shows the clinical characteristics of patients with anaplastic large T/null cell lymphoma.

The diagnosis of anaplastic large T/null cell lymphoma is made when an expert hematopathologist recognizes the typical morphologic picture and a T cell or null cell immunophenotype with CD30 positivity. Documentation of the t(2;5) and/or overexpression of ALK protein confirm the diagnosis. Some diffuse large B cell lymphomas can also have an anaplastic appearance but have the same clinical course or response to therapy as other diffuse large B cell lymphomas.

Patients with anaplastic large T/null cell lymphoma are typically young (median age, 33 years) and male (~70%). Some 50% of patients present in stage I/II, and the remainder with more extensive disease. Systemic symptoms and elevated LDH levels are seen in about one-half of patients. Bone marrow and the gastrointestinal tract are rarely involved, but skin involvement is frequent. Some patients with disease confined to the skin have a different and more indolent disorder that has been termed *cutaneous anaplastic large T/null cell lymphoma* and might be related to lymphomatoid papulosis.

Treatment: Anaplastic Large T/Null Cell Lymphoma

Treatment regimens appropriate for other aggressive lymphomas, such as diffuse large B cell lymphoma, should be utilized in patients with anaplastic large T/null cell lymphoma, with the exception that the B cell –specific antibody, rituximab, is omitted. Surprisingly, given the anaplastic appearance, this disorder has the best survival rate of any aggressive lymphoma. The 5-year survival is >75%. While traditional prognostic factors such as the IPI predict treatment outcome, overexpression of the ALK protein is an important prognostic factor, with patients overexpressing this protein having a superior treatment outcome. The ALK inhibitor crizotinib appears highly active as well.

PERIPHERAL T CELL LYMPHOMA

The peripheral T cell lymphomas make up a heterogeneous morphologic group of aggressive neoplasms that share a mature T cell immunophenotype. They represent ~7% of all cases of non-Hodgkin's lymphoma. A number of distinct clinical syndromes are included in this group of disorders. Table 110–10 shows the clinical characteristics of patients with peripheral T cell lymphoma.

The diagnosis of peripheral T cell lymphoma, or any of its specific subtypes, requires an expert hematopathologist, an adequate biopsy, and immunophenotyping. Most peripheral T cell lymphomas are CD4+, but a few will be CD8+, both CD4+ and CD8+, or have an NK cell immunophenotype. No characteristic genetic abnormalities have yet been identified, but translocations involving the T cell antigen receptor genes on chromosomes 7 or 14 may be detected. The differential diagnosis of patients suspected of having peripheral T cell lymphoma includes reactive T cell infiltrative processes. In some cases, demonstration of a monoclonal T cell population using T cell receptor gene rearrangement studies will be required to make a diagnosis.

The initial evaluation of a patient with a peripheral T cell lymphoma should include the studies in Table 110–11 for staging patients with non-Hodgkin's lymphoma. Unfortunately, patients with peripheral T cell lymphoma usually present with adverse prognostic factors, with >80% of patients having an IPI score ≥2 and >30% having an IPI score ≥4. As this would predict, peripheral T cell lymphomas are associated with a poor outcome, and only 25% of the patients survive 5 years after diagnosis. Treatment regimens are the same as those used for diffuse large B cell lymphoma (omitting rituximab), but patients with peripheral T cell lymphoma have a poorer response to treatment. Because of this poor treatment outcome, hematopoietic stem cell transplantation is often considered early in the care of young patients.

A number of specific clinical syndromes are seen in the peripheral T cell lymphomas. *Angioimmunoblastic T cell lymphoma* is one of the more common subtypes, making up ~20% of T cell lymphomas. These patients typically present with generalized lymphadenopathy, fever, weight loss, skin rash, and polyclonal hypergammaglobulinemia. In some cases, it is difficult to separate patients with a reactive disorder from those with true lymphoma.

Extranodal T/NK cell lymphoma of nasal type has also been called *angiocentric lymphoma* and was previously termed *lethal midline granuloma*. This disorder is more frequent in Asia and South America than in the United States and Europe. EBV is thought to play an etiologic role. Although most frequent in the upper airway, it can involve other organs. The course is aggressive, and patients frequently have the hemophagocytic syndrome. When marrow and blood involvement occur, distinction between this disease and leukemia might be difficult. Some patients will respond to aggressive combination chemotherapy regimens, but the overall outlook is poor.

Enteropathy-type intestinal T cell lymphoma is a rare disorder that occurs in patients with untreated gluten-sensitive enteropathy. Patients are frequently wasted and sometimes present with intestinal perforation. The

prognosis is poor. *Hepatosplenic γδ T cell lymphoma* is a systemic illness that presents with sinusoidal infiltration of the liver, spleen, and bone marrow by malignant T cells. Tumor masses generally do not occur. The disease is associated with systemic symptoms and is often difficult to diagnose. Treatment outcome is poor. *Subcutaneous panniculitis-like T cell lymphoma* is a rare disorder that is often confused with panniculitis. Patients present with multiple subcutaneous nodules, which progress and can ulcerate. Hemophagocytic syndrome is common. Response to therapy is poor. The development of the hemophagocytic syndrome (profound anemia, ingestion of erythrocytes by monocytes and macrophages) in the course of any peripheral T cell lymphoma is generally associated with a fatal outcome.

Hodgkin's Disease

CLASSICAL HODGKIN'S DISEASE

Hodgkin's disease occurs in 8000 patients in the United States each year, and the disease does not appear to be increasing in frequency. Most patients present with palpable lymphadenopathy that is nontender; in most patients, these lymph nodes are in the neck, supraclavicular area, and axilla. More than half the patients will have mediastinal adenopathy at diagnosis, and this is sometimes the initial manifestation. Subdiaphragmatic presentation of Hodgkin's disease is unusual and more common in older males. One-third of patients present with fevers, night sweats, and/or weight loss—B symptoms in the Ann Arbor staging classification (Table 110–8). Occasionally, Hodgkin's disease can present as a fever of unknown origin. This is more common in older patients who are found to have mixed-cellularity Hodgkin's disease in an abdominal site. Rarely, the fevers persist for days to weeks, followed by afebrile intervals and then recurrence of the fever. This pattern is known as *Pel-Ebstein fever*. Hodgkin's disease can occasionally present with unusual manifestations. These include severe and unexplained itching, cutaneous disorders such as erythema nodosum and ichthyosiform atrophy, paraneoplastic cerebellar degeneration and other distant effects on the CNS, nephrotic syndrome, immune hemolytic anemia and thrombocytopenia, hypercalcemia, and pain in lymph nodes on alcohol ingestion.

The diagnosis of Hodgkin's disease is established by review of an adequate biopsy specimen by an expert hematopathologist. In the United States, most patients have nodular sclerosing Hodgkin's disease, with a minority of patients having mixed-cellularity Hodgkin's disease. Lymphocyte-predominant and lymphocyte-depleted Hodgkin's disease are rare. Mixed-cellularity Hodgkin's disease or lymphocyte-depletion Hodgkin's disease are seen more frequently in patients infected by HIV **(Fig. 110-11)**. Hodgkin's disease is a tumor characterized by rare neoplastic cells of B cell origin (immunoglobulin genes are rearranged but not expressed) in a tumor mass that is largely polyclonal inflammatory infiltrate, probably a reaction to cytokines produced by the tumor cells. The differential diagnosis of a lymph node biopsy suspicious for Hodgkin's disease includes inflammatory processes, mononucleosis, non-Hodgkin's lymphoma, phenytoin-induced adenopathy, and nonlymphomatous malignancies.

Figure 110-11

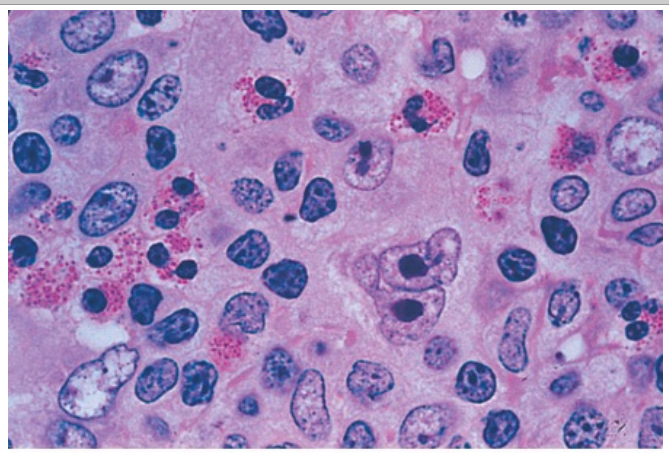

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mixed-cellularity Hodgkin's disease. A Reed-Sternberg cell is present near the center of the field; a large cell with a bilobed nucleus and prominent nucleoli giving an "owl's eyes" appearance. The majority of the cells are normal lymphocytes, neutrophils, and eosinophils that form a pleomorphic cellular infiltrate.

The staging evaluation for a patient with Hodgkin's disease would typically include a careful history and physical examination; complete blood count; erythrocyte sedimentation rate; serum chemistry studies including LDH; chest radiograph; CT scan of the chest, abdomen, and pelvis; and bone marrow biopsy. Many patients would also have a PET scan or a gallium scan. Although rarely utilized, a bipedal lymphangiogram can be helpful. PET and gallium scans are most useful to document remission. Staging laparotomies were once popular for most patients with Hodgkin's disease but are now done rarely because of an increased reliance on systemic rather than local therapy.

Treatment: Classical Hodgkin's Disease

Patients with localized Hodgkin's disease are cured >90% of the time. In patients with good prognostic

factors, extended-field radiotherapy has a high cure rate. Increasingly, patients with all stages of Hodgkin's disease are treated initially with chemotherapy. Patients with localized or good-prognosis disease receive a brief course of chemotherapy followed by radiotherapy to sites of node involvement. Patients with more extensive disease or those with B symptoms receive a complete course of chemotherapy. The most popular chemotherapy regimens used in Hodgkin's disease include doxorubicin, bleomycin, vinblastine, and dacarbazine (ABVD) and mechlorethamine, vincristine, procarbazine, and prednisone (MOPP), or combinations of the drugs in these two regimens. Today, most patients in the United States receive ABVD, but a weekly chemotherapy regimen administered for 12 weeks called *Stanford V* is becoming increasingly popular, but includes radiation therapy, which has been associated with life-threatening late toxicities such as premature coronary artery disease and second solid tumors. In Europe, a high-dose regimen called *BEACOPP* incorporating alkylating agents has become popular and might have a better response rate in very high risk patients. Long-term disease-free survival in patients with advanced disease can be achieved in >75% of patients who lack systemic symptoms and in 60–70% of patients with systemic symptoms.

Patients who relapse after primary therapy of Hodgkin's disease can frequently still be cured. Patients who relapse after initial treatment with only radiotherapy have an excellent outcome when treated with chemotherapy. Patients who relapse after an effective chemotherapy regimen are usually not curable with subsequent chemotherapy administered at standard doses. However, patients with a long initial remission can be an exception to this rule. Autologous bone marrow transplantation can cure half of patients who fail effective chemotherapy regimens.

Because of the very high cure rate in patients with Hodgkin's disease, long-term complications have become a major focus for clinical research. In fact, in some series of patients with early-stage disease, more patients died from late complications of therapy than from Hodgkin's disease itself. This is particularly true in patients with localized disease. The most serious late side effects include second malignancies and cardiac injury. Patients are at risk for the development of acute leukemia in the first 10 years after treatment with combination chemotherapy regimens that contain alkylating agents plus radiation therapy. The risk for development of acute leukemia appears to be greater after MOPP-like regimens than with ABVD. The risk of development of acute leukemia after treatment for Hodgkin's disease is also related to the number of exposures to potentially leukemogenic agents (i.e., multiple treatments after relapse) and the age of the patient being treated, with those aged >60 years at particularly high risk. The development of carcinomas as a complication of treatment for Hodgkin's disease has become a major problem. These tumors usually occur ≥10 years after treatment and are associated with use of radiotherapy. For this reason, young women treated with thoracic radiotherapy for Hodgkin's disease should institute screening mammograms 5–10 years after treatment, and all patients who receive thoracic radiotherapy for Hodgkin's disease should be discouraged from smoking. Thoracic radiation also accelerates coronary artery disease, and patients should be encouraged to minimize risk factors for coronary artery disease such as smoking and elevated cholesterol levels. Cervical radiation therapy increases the risk of carotid atherosclerosis and stroke.

A number of other late side effects from the treatment of Hodgkin's disease are well known. Patients who receive thoracic radiotherapy are at very high risk for the eventual development of hypothyroidism and should be observed for this complication; intermittent measurement of thyrotropin should be made to identify the condition before it becomes symptomatic. Lhermitte's syndrome occurs in ~15% of patients who receive thoracic radiotherapy. This syndrome is manifested by an "electric shock" sensation into the lower extremities on flexion of the neck. Infertility is a concern for all patients undergoing treatment for Hodgkin's disease. In both women and men, the risk of permanent infertility is age-related, with younger patients more likely to recover fertility. In addition, treatment with ABVD rather than MOPP increases the chances to retain fertility.

NODULAR LYMPHOCYTE-PREDOMINANT HODGKIN'S DISEASE

Nodular lymphocyte-predominant Hodgkin's disease is now recognized as an entity distinct from classical Hodgkin's disease. Previous classification systems recognized that biopsies from a subset of patients diagnosed as having Hodgkin's disease contained a predominance of small lymphocytes and rare Reed-Sternberg cells (Fig. 110-11). A subset of these patients have tumors with nodular growth pattern and a clinical course that varied from that of patients with classical Hodgkin's disease. This is an unusual clinical entity and represents <5% of cases of Hodgkin's disease.

Nodular lymphocyte-predominant Hodgkin's disease has a number of characteristics that suggest its relationship to non-Hodgkin's lymphoma. These include a clonal proliferation of B cells and a distinctive immunophenotype; tumor cells express J chain and display CD45 and epithelial membrane antigen (EMA) and do not express two markers normally found on Reed-Sternberg cells, CD30 and CD15. This lymphoma tends to have a chronic, relapsing course and sometimes transforms to diffuse large B cell lymphoma.

The treatment of patients with nodular lymphocyte-predominant Hodgkin's disease is controversial. Some clinicians favor no treatment and merely close follow-up. In the United States, most physicians will treat localized disease with radiotherapy and disseminated disease with regimens utilized for patients with classical Hodgkin's disease. Regardless of the therapy utilized, most series report a long-term survival of >80%.

LYMPHOMA-LIKE DISORDERS

The most common condition that pathologists and clinicians might confuse with lymphoma is reactive, atypical lymphoid hyperplasia. Patients might have localized or disseminated lymphadenopathy and might have the systemic symptoms characteristic of lymphoma. Underlying causes include a drug reaction to phenytoin or carbamazepine. Immune disorders such as rheumatoid arthritis and lupus erythematosus, viral infections such as cytomegalovirus and EBV, and bacterial infections such as cat-scratch disease may cause

adenopathy (Chap. 59). In the absence of a definitive diagnosis after initial biopsy, continued follow-up, further testing, and repeated biopsies, if necessary, are the appropriate approach rather than instituting therapy.

Specific conditions that can be confused with lymphoma include *Castleman's disease*, which can present with localized or disseminated lymphadenopathy; some patients have systemic symptoms. The disseminated form is often accompanied by anemia and polyclonal hypergammaglobulinemia, and the condition has been associated with overproduction of interleukin 6, possibly produced by human herpesvirus 8. Patients with localized disease can be treated effectively with local therapy, while the initial treatment for patients with disseminated disease is usually with systemic glucocorticoids. IL-6-directed therapy is being developed.

Sinus histiocytosis with massive lymphadenopathy (*Rosai-Dorfman's disease*) usually presents with bulky lymphadenopathy in children or young adults. The disease is usually nonprogressive and self-limited, but patients can manifest autoimmune hemolytic anemia.

Lymphomatoid papulosis is a cutaneous lymphoproliferative disorder that is often confused with anaplastic large cell lymphoma involving the skin. The cells of lymphomatoid papulosis are similar to those seen in lymphoma and stain for CD30, and T cell receptor gene rearrangements are sometimes seen. However, the condition is characterized by waxing and waning skin lesions that usually heal, leaving small scars. In the absence of effective communication between the clinician and the pathologist regarding the clinical course in the patient, this disease will be misdiagnosed. Since the clinical picture is usually benign, misdiagnosis is a serious mistake.

ACKNOWLEDGMENT

James Armitage was a coauthor of this chapter in prior editions, and substantial material from those editions has been included here.

FURTHER READINGS

Ansel SM, Armitage JO: Management of Hodgkin lymphoma. Mayo Clin Proc 81:419, 2006

Armitage JO: Early-stage Hodgkin's lymphoma. N Engl J Med 363:653, 2010[PMID: 20818856]

Brepoels L et al: PET and PET/CT for response evaluation in lymphoma: Current practice and developments. Leuk Lymphoma 48:270, 2007[PMID: 17325886]

Cabanillas F: Front-line therapy of diffuse large B-cell lymphoma. Curr Opin Oncol 22:642, 2010[PMID: 20811278]

Harris NL et al: World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: Report of the Clinical Advisory Committee Meeting, Airlie House, Virginia, November, 1997. J Clin Oncol 17:3835, 1999[PMID: 10577857]

Longo DL (ed): Hodgkin's lymphoma. Cancer J 15:114, 2009[PMID: 19390305]

Sehn LH et al: The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. Blood 109:1857, 2007[PMID: 17105812]

Stathis A et al: Treatment of gastric marginal zone lymphoma of MALT type. Expert Opin Pharmacother 11:2141, 2010[PMID: 20586708]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LESS COMMON HEMATOLOGIC MALIGNANCIES: INTRODUCTION

The most common lymphoid malignancies are discussed in Chap. 110, myeloid leukemias in Chap. 109, myelodysplastic syndromes in Chap. 107, and myeloproliferative syndromes in Chap. 108. This chapter will focus on the more unusual forms of hematologicmalignancy. The diseases discussed here are listed in **Table e21-1**. Each of these entities accounts for less than 1% of hematologic neoplasms.

Table e21-1 Unusual Lymphoid and Myeloid Malignancies
Lymphoid
Mature B cell neoplasms
B cell prolymphocytic leukemia
Splenic marginal zone lymphoma
Hairy cell leukemia
Nodal marginal zone B cell lymphoma
Mediastinal large B cell lymphoma
Intravascular large B cell lymphoma Primary effusion lymphoma
Lymphomatoid granulomatosis
Mature T cell and NK cell neoplasms
T cell prolymphocytic leukemia
T cell large granular lymphocytic leukemia
Aggressive NK cell leukemia
Extranodal NK/T cell lymphoma, nasal type
Enteropathy-type T cell lymphoma Hepatosplenic T cell lymphoma
Subcutaneous panniculitis-like T cell lymphoma
Blastic NK cell lymphoma
Primary cutaneous CD30+ T cell lymphoma
Angioimmunoblastic T cell lymphoma
Myeloid
Chronic neutrophilic leukemia
Chronic eosinophilic leukemia/hypereosinophilic syndrome
Histiocytic and Dendritic Cell Neoplasms
Histiocytic sarcoma
Langerhans cell histiocytosis
Langerhans cell sarcoma
Interdigitating dendritic cell sarcoma
Follicular dendritic cell sarcoma
Mast Cells
Mastocytosis Cutaneous mastocytosis
Systemic mastocytosis
Mast cell sarcoma
Extracutaneous mastocytoma

Lymphoid Malignancies

Precursor B cell and precursor T cell neoplasms are discussed in Chap. 110. All the lymphoid tumors discussed here are mature B cell or T cell, natural killer (NK) cell neoplasms.

Mature B Cell Neoplasms

B CELL PROLYMPHOCYTIC LEUKEMIA (B-PLL)

This is a malignancy of medium-sized (about twice the size of a normal small lymphocyte), round lymphocytes with a prominent nucleolus and light blue cytoplasm on Wright's stain. It dominantly affects the blood, bone marrow, and spleen and usually does not cause adenopathy. The median age of affected patients is 70 years and men are more often affected than women (male/female ratio is 1.6). This entity is distinct from chronic lymphoid leukemia (CLL) and does not develop as a consequence of that disease.

Clinical presentation is generally from symptoms of splenomegaly or incidental detection of an elevated WBC count. The clinical course can be rapid. The cells express surface IgM (with or without IgD) and typical B cell markers (CD19, CD20, CD22). CD23 is absent and about one-third of cases express CD5. The CD5 expression along with the presence of the t(11;14) translocation in 20% of cases leads to confusion in distinguishing B-PLL from the leukemic form of mantle cell lymphoma. No reliable criteria for the distinction have emerged. About half of patients have mutation or loss of p53 and deletions have been noted in 11q23 and 13q14. Nucleoside analogues like fludarabine and cladribine and combination chemotherapy (cyclophosphamide, doxorubicin, vincristine, prednisone, or CHOP) have produced responses. CHOP plus rituximab may be more effective than CHOP alone, but the disease is sufficiently rare that large series have not been reported. Splenectomy can produce palliation of symptoms but appears to have little or no impact on the course of the disease.

SPLENIC MARGINAL ZONE LYMPHOMA (SMZL)

This tumor of mainly small lymphocytes originates in the marginal zone of the spleen white pulp, grows to efface the germinal centers and mantle, and invades the red pulp. Splenic hilar nodes, bone marrow, and peripheral blood may be involved. The circulating tumor cells have short surface villi and are called villous lymphocytes. **Table e21-2** shows differences in tumor cells of a number of neoplasms of small lymphocytes that aid in the differential diagnosis. SMZL cells express surface immunoglobulin and CD20, but are negative for CD5, CD10, and CD103.

Table e21-2 Immunophenotype of Tumors of Small Lymphocytes

	CD5	CD20	CD43	CD10	CD103	sIg	cyclinD1
Follicular lymphoma	neg	pos	pos	pos	neg	pos	neg
Chronic lymphoid leukemia	pos	pos	pos	neg	neg	pos	neg
B cell prolymphocytic leukemia	pos	pos	pos	neg	neg	pos	pos
Mantle cell lymphoma	pos	pos	pos	neg	neg	pos	pos
Splenic marginal zone lymphoma	neg	pos	neg	neg	neg	pos	neg
Hairy cell leukemia	neg	pos	?	neg	pos	pos	neg

Abbreviations: neg, negative; pos, positive.

The median age of patients with SMZL is mid-fifties and men and women are equally represented. Patients present with incidental or symptomatic splenomegaly or incidental detection of lymphocytosis in the peripheral blood with villous lymphocytes. Autoimmune anemia or thrombocytopenia may be present. The immunoglobulin produced by these cells contains somatic mutations that reflect transit through a germinal center, and ongoing mutations suggest that the mutation machinery has remained active. About 40% of patients have either deletions or translocations involving 7q21, the site of the *CDK6* gene. The genetic lesions typically found in extranodal marginal zone lymphomas [e.g. trisomy 3 and t(11;18)] are uncommon in SMZL.

The clinical course of disease is generally indolent. Long remissions can be seen after splenectomy. A small fraction of patients undergo histologic progression to diffuse large B cell lymphoma with a concomitant change to a more aggressive natural history. Experience with combination chemotherapy in SMZL is limited.

HAIRY CELL LEUKEMIA

Hairy cell leukemia is a tumor of small lymphocytes with oval nuclei, abundant cytoplasm, and distinctive membrane projections (hairy cells). Patients have splenomegaly and diffuse bone marrow involvement. While some circulating cells are noted, the clinical picture is dominated by symptoms from the enlarged spleen and pancytopenia. The mechanism of the pancytopenia is not completely clear and may be mediated by both inhibitory cytokines and marrow replacement. The marrow has an increased level of reticulin fibers; indeed, hairy cell leukemia is a common cause of inability to aspirate bone marrow or so-called "dry tap" **(Table e21-3)**. Monocytopenia is profound and may explain a predisposition to atypical mycobacterial infection that is observed clinically. The tumor cells have

strong expression of CD22, CD25, and CD103; soluble CD25 level in serum is an excellent tumor marker for disease activity. The cells also express tartrate-resistant acid phosphatase. The immunoglobulin genes are rearranged and mutated, indicating the influence of a germinal center. No specific cytogenetic abnormality has been found, but most cases contain the activating BRAF mutation V600E.

Table e21-3 Differential Diagnosis of "Dry Tap"—Inability to Aspirate Bone Marrow

Dry taps occur in about 4% of attempts and are associated with:	
Metastatic carcinoma infiltration	17%
Chronic myeloid leukemia	15%
Myelofibrosis	14%
Hairy cell leukemia	10%
Acute leukemia	10%
Lymphomas, Hodgkin's disease	9%
Normal marrow	Rare

The median age of affected patients is mid-fifties and the male/female ratio is 5:1. Treatment options are numerous. Splenectomy is often associated with prolonged remission. Nucleosides including cladribine and deoxycoformycin are highly active but are also associated with further immunosuppression and can increase the risk of certain opportunistic infections. However, after brief courses of these agents, patients usually obtain very durable remissions during which immune function spontaneously recovers. Interferon α is also an effective therapy but is not as effective as nucleosides.

NODAL MARGINAL ZONE B CELL LYMPHOMA

This rare node-based disease bears an uncertain relationship with extranodal marginal zone lymphomas, which are often mucosa-associated and are called mucosa-associated lymphoid tissue or MALT lymphomas, and splenic marginal zone lymphomas. Patients may have localized or generalized adenopathy. The neoplastic cell is a marginal zone B cell with monocytoid features and has been called monocytoid B cell lymphoma in the past. Up to one-third of the patients may have extranodal involvement, and involvement of the lymph nodes can be secondary to the spread of a mucosal primary lesion. In authentic nodal primaries, the cytogenetic abnormalities associated with MALT lymphomas [trisomy 3 and t(11;18)] are very rare. The clinical course is indolent. Patients often respond to combination chemotherapy, though remissions have not been durable. Few patients have received CHOP plus rituximab, which is likely to be an effective approach to management.

MEDIASTINAL (THYMIC) LARGE B CELL LYMPHOMA

This entity was originally considered a subset of diffuse large B cell lymphoma; however, additional study has identified it as a distinct entity with its own characteristic clinical, genetic, and immunophenotypic features. This is a disease that can be bulky in size but usually remains confined to the mediastinum. It can be locally aggressive, including progressing to produce a superior vena cava obstruction syndrome or pericardial effusion. About one-third of patients develop pleural effusions and 5–10% can disseminate widely to kidney, adrenal, liver, skin, and even brain. The disease affects women more often than men (male/female ratio is 1:2–3), and the median age is 35–40 years.

The tumor is composed of sheets of large cells with abundant cytoplasm accompanied by variable, but often abundant, fibrosis. It is distinguished from nodular sclerosing Hodgkin's disease by the paucity of normal lymphoid cells and the absence of lacunar variants of Reed-Sternberg cells. However, more than one-third of the genes that are expressed to a greater extent in primary mediastinal large B cell lymphoma than in usual diffuse large B cell lymphoma are also overexpressed in Hodgkin's disease, suggesting a possible pathogenetic relationship between the two entities that affect the same anatomic site. Tumor cells may overexpress *MAL*. The genome of tumor cells is characterized by frequent chromosomal gains and losses. The tumor cells in mediastinal large B cell lymphoma express CD20 but surface immunoglobulin, and HLA class I and class II molecules may be absent or incompletely expressed. Expression of lower levels of class II HLA identifies a subset with poorer prognosis. The cells are CD5 and CD10 negative but may show light staining with anti-CD30. The cells are CD45 positive, unlike cells of classical Hodgkin's disease.

MACOP-B and R-CHOP are effective treatments, achieving 5-year survival of 75–87%. A role for mediastinal radiation therapy has not been definitively demonstrated but it is frequently used, especially in patients whose mediastinal area remains PET-avid after 4–6 cycles of chemotherapy.

INTRAVASCULAR LARGE B CELL LYMPHOMA

This is an extremely rare form of diffuse large B cell lymphoma characterized by the presence of lymphoma in the lumen of small vessels, particularly capillaries. It is also known as malignant angioendotheliomatosis or angiotropic large cell lymphoma. It is sufficiently rare that no consistent

picture has emerged to define a clinical syndrome or its epidemiologic and genetic features. It is thought to remain inside vessels because of a defect in adhesion molecules and homing mechanisms, an idea supported by scant data suggesting absence of expression of β-1 integrin and ICAM-1. Patients commonly present with symptoms of small vessel occlusion, skin lesions, or neurologic symptoms. The tumor cell clusters can promote thrombus formation. In general, the clinical course is aggressive and the disease is poorly responsive to therapy. Often a diagnosis is not made until very late in the course of the disease.

PRIMARY EFFUSION LYMPHOMA

This entity is another variant of diffuse large B cell lymphoma that presents with pleural effusions, usually without apparent tumor mass lesions. It is most common in the setting of immune deficiency disease, especially AIDS, and is caused by human herpes virus 8 (HHV-8)/Kaposi's sarcoma herpes virus (KSHV). It is also known as body cavity–based lymphoma. Some patients have been previously diagnosed with Kaposi's sarcoma. It can also occur in the absence of immunodeficiency in elderly men of Mediterranean heritage, similar to Kaposi's sarcoma but even less common.

The malignant effusions contain cells positive for HHV-8/KSHV and many are also co-infected with Epstein-Barr virus. The cells are large with large nuclei and prominent nucleoli that can be confused with Reed-Sternberg cells. The cells express CD20 and CD79a (immunoglobulin-signaling molecule), though they often do not express immunoglobulin. Some cases aberrantly express T cell markers such as CD3 or rearranged T cell receptor genes. No characteristic genetic lesions have been reported but gains in chromosome 12 and X material has been seen, similar to other HIV-associated lymphomas. The clinical course is generally characterized by rapid progression and death within 6 months.

LYMPHOMATOID GRANULOMATOSIS

This is an angiocentric, angiodestructive lymphoproliferative disease comprised by neoplastic Epstein-Barr virus–infected monoclonal B cells accompanied and outnumbered by a polyclonal reactive T cell infiltrate. The disease is graded based on histologic features such as cell number and atypia in the B cells. It is most often confused with extranodal NK–T cell lymphoma, nasal type, which can also be angiodestructive and is Epstein-Barr virus–related. The disease usually presents in adults (M>F) as a pulmonary infiltrate. Involvement is often entirely extranodal and can include kidney (32%), liver (29%), skin (25%), and brain (25%). The disease often but not always occurs in the setting of immune deficiency.

The disease can be remitting and relapsing in nature or can be rapidly progressive. The course is usually predicted by the histologic grade. The disease is highly responsive to combination chemotherapy and is curable in most cases. Some investigators have claimed that low-grade disease (grade I and II) can be treated with interferon α.

Mature T Cell and NK Cell Neoplasms

T CELL PROLYMPHOCYTIC LEUKEMIA

This is an aggressive leukemia of medium-sized prolymphocytes involving the blood, marrow, nodes, liver, spleen, and skin. It accounts for 1–2% of all small lymphocytic leukemias. Most patients present with elevated white blood cell count (often >100,000/μL), hepatosplenomegaly, and adenopathy. Skin involvement occurs in 20%. The diagnosis is made from peripheral blood smear, which shows cells about 25% larger than those in small lymphocytes, with cytoplasmic blebs and nuclei that may be indented. The cells express T cell markers like CD2, CD3, and CD7; two-thirds of patients have cells that are CD4+ and CD8–, and 25% have cells that are CD4+ and CD8+. T cell receptor β chains are clonally rearranged. In 80% of patients, inversion of chromosome 14 occurs between q11 and q32. Ten percent have t(14;14) translocations that bring the T cell receptor alpha/beta gene locus into juxtaposition with oncogenes *TCL1* and *TCL1b* at 14q32.1. Chromosome 8 abnormalities are also common. Deletions in the *ATM* gene are also noted.

The course of the disease is generally rapid, with median survival of about 12 months. Responses have been seen with the anti-CD52 antibody, nucleoside analogs, and CHOP chemotherapy. Small numbers of patients with T cell prolymphocytic leukemia have also been treated with high-dose therapy and allogeneic bone marrow transplantation after remission has been achieved with conventional-dose therapy.

T CELL LARGE GRANULAR LYMPHOCYTIC LEUKEMIA

T cell large granular lymphocytic leukemia (LGL leukemia) is characterized by increases in the number of LGLs in the peripheral blood (2000–20,000/μL) often accompanied by severe neutropenia, with or without concomitant anemia. Patients may have splenomegaly and frequently have evidence of systemic autoimmune disease, including rheumatoid arthritis, hypergammaglobulinemia, autoantibodies, and circulating immune complexes. Bone marrow involvement is mainly interstitial in pattern, with fewer than 50% lymphocytes on differential count. Usually the cells express CD3, T cell receptors, and CD8; NK-like variants may be CD3–. The leukemic cells often express Fas and Fas ligand.

The course of the disease is generally indolent and dominated by the neutropenia. Paradoxically,

immunosuppressive therapy with cyclosporine, methotrexate, or cyclophosphamide plus glucocorticoids can produce an increase in granulocyte counts. Nucleosides have been used anecdotally. Occasionally the disease can accelerate to a more aggressive clinical course.

AGGRESSIVE NK CELL LEUKEMIA

NK neoplasms are very rare and they may follow a range of clinical courses from very indolent to highly aggressive. They are more common in Asians than whites, and the cells frequently harbor a clonal Epstein-Barr virus episome. The peripheral blood white count is usually not greatly elevated, but abnormal large lymphoid cells with granular cytoplasm are noted. The aggressive form is characterized by symptoms of fever and laboratory abnormalities of pancytopenia. Hepatosplenomegaly is common; node involvement less so. Patients may have hemophagocytosis, coagulopathy, or multiorgan failure. Serum levels of Fas ligand are elevated.

The cells express CD2 and CD56 and do not have rearranged T cell receptor genes. Deletions involving chromosome 6 are common. The disease can be rapidly progressive. Some forms of NK neoplasms are more indolent. They tend to be discovered incidentally with LGL lymphocytosis and do not manifest the fever and hepatosplenomegaly characteristic of the aggressive leukemia. The cells are also CD2 and CD56 positive, but they do not contain clonal forms of Epstein-Barr virus and are not accompanied by pancytopenia or autoimmune disease.

EXTRANODAL NK/T CELL LYMPHOMA, NASAL TYPE

Like lymphomatoid granulomatosis, extranodal NK/T cell lymphoma tends to be an angiocentric and angiodestructive lesion, but the malignant cells are not B cells. In most cases, they are CD56+ Epstein-Barr virus–infected cells; occasionally they are CD56– Epstein-Barr virus–infected cytotoxic T cells. They are most commonly found in the nasal cavity. Historically, this illness was called lethal midline granuloma, polymorphic reticulosis, and angiocentric immunoproliferative lesion. This form of lymphoma is prevalent in Asia, Mexico, and Central and South America; it affects males more commonly than females. When it spreads beyond the nasal cavity, it may affect soft tissue, the gastrointestinal tract, or the testis. In some cases, hemophagocytic syndrome may influence the clinical picture. Patients may have B symptoms. Many of the systemic manifestations of disease are related to the production of cytokines by the tumor cells and the cells responding to their signals. Deletions and inversions of chromosome 6 are common.

Many patients with extranodal NK/T cell lymphoma, nasal type have excellent antitumor responses with combination chemotherapy regimens, particularly those with localized disease. Radiation therapy is often used after completion of chemotherapy. Four risk factors have been defined, including B symptoms, advanced stage, elevated LDH, and regional lymph node involvement. Patient survival is linked to the number of risk factors: 5-year survival is 81% for 0 risk factors, 64% for 1, 32% for 2, and 7% for 3 or 4. Combination regimens without anthracyclines have been touted as superior to CHOP, but data are sparse. High-dose therapy with stem cell transplantation has been used but its role is unclear.

ENTEROPATHY-TYPE T CELL LYMPHOMA

Enteropathy-type T cell lymphoma is a rare complication of long-standing celiac disease. It most commonly occurs in the jejunum or the ileum. In adults, the lymphoma may be diagnosed at the same time as celiac disease, but the suspicion is that the celiac disease was a longstanding precursor to the development of lymphoma. The tumor usually presents as multiple ulcerating mucosal masses, but may also produce a dominant exophytic mass or multiple ulcerations. The tumor expresses CD3 and CD7 nearly always and may or may not express CD8. The normal-appearing lymphocytes in the adjacent mucosa often have a similar phenotype to the tumor. Most patients have the HLAgenotype associated with celiac disease, HLA DQA1*0501 or DQB1*0201.

The prognosis of this form of lymphoma is typically (median survival is 7 months) poor but some patients have a good response to CHOP chemotherapy. Patients who respond can develop bowel perforation from responding tumor. If the tumor responds to treatment, recurrence may develop elsewhere in the celiac disease–affected small bowel.

HEPATOSPLENIC T CELL LYMPHOMA

Hepatosplenic T cell lymphoma is a malignancy derived from T cells expressing the gamma/delta T cell antigen receptor that affects mainly the liver and fills the sinusoids with medium-size lymphoid cells. When the spleen is involved, dominantly the red pulp is infiltrated. It is a disease of young people, especially young people with an underlying immunodeficiency or with an autoimmune disease that demands immunosuppressive therapy. The use of thiopurine and infliximab is particularly common in the history of patients with this disease. The cells are CD3+ and usually CD4- and CD8-negative. The cells may contain isochromosome 7q, often together with trisomy 8. The lymphoma has an aggressive natural history. Combination chemotherapy may induce remissions, but most patients relapse. Median survival is about 2 years. The tumor does not appear to respond to reversal of immunosuppressive therapy.

SUBCUTANEOUS PANNICULITIS-LIKE T CELL LYMPHOMA

Subcutaneous panniculitis-like T cell lymphoma involves multiple subcutaneous collections of neoplastic T cells that are usually cytotoxic cells in phenotype (i.e., contain perforin and granzyme B and express CD3 and CD8). The rearranged T cell receptor is usually alpha/beta-derived but occasionally the gamma/delta receptors are involved, particularly in the setting of immunosuppression. The cells are negative for Epstein-Barr virus. Patients may have a hemophagocytic syndrome in addition to the skin infiltration; fever and hepatosplenomegaly may also be present. Nodes are generally not involved. Patients frequently respond to combination chemotherapy, including CHOP. When the disease is progressive, the hemophagocytic syndrome can be a component of a fulminant downhill course. Effective therapy can reverse the hemophagocytic syndrome.

BLASTIC NK CELL LYMPHOMA

The neoplastic cells express NK cell markers, especially CD56, and are CD3 negative. They are large blastic-appearing cells and may produce a leukemia picture, but the dominant site of involvement is the skin. Morphologically, the cells are similar to the neoplastic cells in acute lymphoid and myeloid leukemia. No characteristic chromosomal abnormalities have been described. The clinical course is rapid and the disease is largely unresponsive to typical lymphoma treatments.

PRIMARY CUTANEOUS CD30+ T CELL LYMPHOMA

This tumor involves the skin and is composed of cells that appear similar to the cells of anaplastic T cell lymphoma. Among cutaneous T cell tumors, about 25% are CD30+ anaplastic lymphomas. If dissemination to lymph nodes occurs, it is difficult to distinguish between the cutaneous and systemic forms of the disease. The tumor cells are often CD4+, and the cells contain granules that are positive for granzyme B and perforin in 70% of cases. The typical t(2;5) of anaplastic T cell lymphoma is absent; indeed, its presence should prompt a closer look for systemic involvement and a switch to a diagnosis of anaplastic T cell lymphoma. This form of lymphoma has sporadically been noted as a rare complication of silicone on saline breast implants. Cutaneous CD30+ T cell lymphoma often responds to therapy. Radiation therapy can be effective and surgery can also produce long-term disease control. Five-year survival exceeds 90%.

ANGIOIMMUNOBLASTIC T CELL LYMPHOMA

Angioimmunoblastic T cell lymphoma is a systemic disease that accounts for about 15% of all T cell lymphomas. Patients frequently have fever, advanced stage, diffuse adenopathy, hepatosplenomegaly, skin rash, polyclonal hypergammaglobulinemia, and a wide range of autoantibodies including cold agglutinins, rheumatoid factor, and circulating immune complexes. Patients may have edema, arthritis, pleural effusions, and ascites. The nodes contain a polymorphous infiltrate of neoplastic T cells and nonneoplastic inflammatory cells together with proliferation of high endothelial venules and follicular dendritic cells. The most common chromosomal abnormalities are trisomy 3, trisomy 5, and an extra X chromosome. Aggressive combination chemotherapy can induce regressions. The underlying immune defects make conventional lymphoma treatments more likely to produce infectious complications.

MYELOID MALIGNANCIES

Chronic Neutrophilic Leukemia

Chronic neutrophilic leukemia is a rare myeloproliferative disorder that may be confused with the much more common chronic myeloid leukemia. Patients present with increased peripheral blood neutrophil counts (>25,000/μL) and hepatosplenomegaly, but unlike those with chronic myeloid leukemia, no Philadelphia chromosome or *BCR/ABL* fusion gene is detectable and immature forms comprise less than 10% of circulating white blood cells. The diagnosis is one of exclusion. One must rule out leukemoid reactions and other myeloproliferative and myelodysplastic syndromes. No morphologic dysplasia of myeloid precursors is present. Some patients with chronic neutrophilic leukemia appear to have an underlying plasma cell disorder but the relationship between the entities is not defined. In the vast majority of patients, cytogenetic analysis is normal. The disease course is variable from 1 to 20 years or more. A small number of patients may develop acute leukemia or myelodysplasia. Most have progressive marrow replacement with myeloid cells and crowding out of red cell and platelet precursors. Hydroxyurea can control the white blood cell count, but this is generally unnecessary even at counts >100,000/μL because, unlike the cells of acute leukemia, the cells of chronic neutrophilic leukemia are not invasive and not likely to cause leukostasis. Patients do not respond to splenectomy and no therapy has been shown to alter the natural history of the disease.

Chronic Eosinophilic Leukemia/Hypereosinophilic Syndrome

The diagnostic criteria for chronic eosinophilic leukemia/hypereosinophilic syndrome are provided in **Table e21-4**. The presence of eosinophilia is critical, defined as eosinophils ≥1500/μL in blood, increased marrow eosinophils, and myeloblasts <20% in blood or marrow for at least 6 months, in the absence of other symptoms requiring more immediate treatment. The eosinophilic disorders are much more common in men than women (9:1). Patients may be totally asymptomatic and have the eosinophilia detected incidentally in routine blood work or they may have any of a myriad of symptoms including fever, fatigue, cough, edema, shortness of breath, central nervous system dysfunction, muscle aches and pains, itching, abdominal pain, diarrhea, peripheral neuropathy, or rheumatologic

findings. The key diagnostic issue is distinguishing clonal eosinophilia, a neoplastic proliferation of eosinophils, from the many entities and drug exposures that can cause secondary eosinophilia. Because of the paucity of markers of eosinophil clonality, the diagnosis tends to be one of exclusion.

Table e21-4 Diagnosis of Chronic Eosinophilic Leukemia and Hypereosinophilic Syndrome
Required: Persistent eosinophilia ≥1500/μL in blood, increased marrow eosinophils, and myeloblasts <20% in blood or marrow.
1. Exclude all causes of reactive eosinophilia: allergy, parasites, infection, pulmonary disease (e.g., hypersensitivity pneumonitis, Loeffler's), and collagen vascular diseases
2. Exclude primary neoplasms associated with secondary eosinophilia: T cell lymphomas, Hodgkin's disease, acute lymphoid leukemia, mastocytosis
3. Exclude other primary myeloid neoplasms that may involve eosinophils: chronic myeloid leukemia, acute myeloid leukemia with inv(16) or t(16;16)(p13;q22), other myeloproliferative syndromes and myelodysplasia
4. Exclude T cell reaction with increased IL-5 or other cytokine production
If these entities have been excluded and no evidence documents a clonal myeloid disorder, the diagnosis is hypereosinophilic syndrome.
If these entities have been excluded and the myeloid cells show a clonal chromosome abnormality or some other evidence of clonality and blast cells are present in the peripheral blood (>2%) or are increased in the marrow (but <20%), the diagnosis is chronic eosinophilic leukemia.

The first condition to rule out is eosinophilia accompanying a *FIP1L1-PDGFRA*–associated myeloproliferative disorder. If a mutation is noted in the peripheral blood, patients are treated with imatinib, which inhibits the platelet-derived growth factor receptor activated in this condition. If mutation is absent, a bone marrow analysis is done with cytogenetics looking for a 5q33 (*PDGFRB*), 4q12 (*PDGFRA*), or 8p11.2 (*FGFR1*) translocation. These clonal eosinophilias are part of the clinical picture of myeloid malignancies involving these genes. The PDGF receptor abnormalities predict a favorable response to imatinib whereas the FGF receptor abnormality is associated with chemotherapy-refractory disease.

If these genetic lesions are absent, peripheral blood lymphocytes are studied for immunophenotype and T cell receptor gene rearrangements. The presence of clonal lymphocytes makes the diagnosis of lymphocyte variant hypereosinophilia and implies a cytokine-driven process. If the peripheral blood T cells are normal, one is left with chronic eosinophilic leukemia/hypereosinophilic syndrome as the diagnosis, and these two related entities are distinguished mainly by peripheral blood and bone marrow blast counts. If the peripheral blood has >2% blasts and the marrow has >5% blasts, chronic eosinophilic leukemia is the diagnosis; if the peripheral blood has <2% blasts and the marrow has <5% blasts, the diagnosis is hypereosinophilic syndrome.

The heart, lungs, and central nervous system are the organ systems often most affected by eosinophil-mediated tissue damage. Patients should have chest radiographs, echocardiography, and troponin level measured to assess lung and cardiac involvement. In the absence of abnormalities in the platelet-derived growth factor receptors, asymptomatic patients can be observed. When treatment is indicated based on symptoms, glucocorticoids are the initial treatment. Hydroxyurea, interferon α, cladribine, and cyclosporine have also been used. The anti-IL5 antibody mepolizumab is being tested. Responses are noted but are not durable. Anti-CD52 antibody (alemtuzumab) has also produced responses but is profoundly immunosuppressive.

HISTIOCYTIC AND DENDRITIC CELL NEOPLASMS

Tumors derived from histiocytes or dendritic cells are exceedingly rare. Through the past century, a number of disorders have been labeled initially as histiocytic disorders but, upon further study with newer analytic tools, the origin has been found to be nonhistiocytic; often, rare T cell disorders like anaplastic large cell lymphoma were initially thought to be derived from histiocytes. Histiocytes and macrophages are not cell types that routinely circulate and, therefore, neoplasms derived from them tend to produce localized tumor masses in the site of origin. The range of markers that characterize these cells is not as broad as those available for lymphocyte subsets. However, the four main types of macrophages and dendritic cells have some distinguishing features. Langerhans cells are bone marrow–derived and reside in the skin; their main function is to present antigen to T cells. They are MHC class II, Fc receptor, and S100 protein positive; they express CD4 and CD1a, and they are not phagocytosing cells. They contain distinctive morphologic features such as Birbeck granules, rod or tennis racket–shaped structures of uncertain function. Interdigitating dendritic cells are also bone marrow–derived antigen-presenting cells and they can be in any tissue. They are MHC class II and S100 positive but do not express other known markers. Follicular dendritic cells appear to be derived from a mesenchymal stem cell and reside in lymph node follicles; they present antigen to B cells. They

are CD21 and CD35 positive and CD68 and CD45 negative. Macrophages are also CD21 and CD38 positive, but they express CD68 and are phagocytic and express lysozyme.

Histiocytic Sarcoma

This is a tumor of histiocytes or macrophages that may present as a solitary mass with or without systemic symptoms of fever and weight loss. The tumor is composed of sheets of large cells effacing the tissue architecture. The cells resemble the cells of diffuse large B cell lymphoma but they do not express lymphocyte markers and are CD68, lysozyme, CD11c, and CD14 positive. The tumor is not highly responsive to treatment and the natural history is usually aggressive.

Langerhans Cell Histiocytosis

Langerhans cell histiocytosis is a disease of childhood that has been called histiocytosis X, Letterer-Siwe disease, Hand-Schuller-Christian disease, and eosinophilic granuloma. Patients with Langerhans cell histiocytosis are at increased risk of acute lymphoid leukemia or other lymphoid malignancy. Three clinical syndromes are recognized. Solitary bone lesions, particularly those involving the skull, femur, pelvis, or ribs, are frequently called eosinophilic granuloma. The presence of multiple lesions affecting a single tissue, most often bone, is called Hand-Schuller-Christian disease. The presence of multiple lesions in multiple organs including bones, skin, liver, spleen, and nodes is called Letterer-Siwe disease. The characteristic Birbeck granules are pathognomonic but only visible on electron microscopy. The clinical course tends to be inversely related to the number of lesions. Patients with solitary lesions may progress to multiple lesion and multiorgan involvement in 10% of cases; however, most patients respond to chemotherapy. Long-term survival is seen in 70–90% of cases.

Langerhans Cell Sarcoma

Langerhans cell sarcoma is distinguished from Langerhans cell histiocytosis by the presence of a high degree of cellular atypia; it can arise de novo or progress from prior Langerhans cell histiocytosis. The natural history is more aggressive, but the disease is responsive to treatment and the long-term survival is about 50%.

Interdigitating Dendritic Cell Sarcoma

Interdigitating dendritic cell sarcoma is a proliferation of spindle- to ovoid-shaped cells that usually present in lymph nodes but also can form skin nodules. The initial mass is usually asymptomatic but fatigue, fever, or night sweats may accompany the lesion. The cells express S100 and are negative for follicular dendritic cell markers like CD21 and CD35. The clinical course is variable. Localized disease is curable with local therapy.

Follicular Dendritic Cell Sarcoma

Follicular dendritic cell sarcoma is a tumor of follicular dendritic cells that originates in lymph nodes in about two-thirds of cases, usually cervical nodes. Extranodal sites may also be involved. The presentation is usually a slow-growing painless mass. The microscopic anatomy of the tumor is similar to that of the interdigitating dendritic cell tumors but the tumor cells express different markers (i.e., they are CD21 and CD35 positive and negative for CD1a). The tumor is typically indolent and can be controlled with surgery. There is no clear role for radiation therapy or chemotherapy.

MASTOCYTOSIS

Mastocytosis is a proliferation and accumulation of mast cells in one or more organ systems. In about 80% of cases, only the skin is involved. In the other 20%, the skin and at least one other organ system are involved.

Cutaneous Mastocytosis

Three major variants of cutaneous mastocytosis are described: (1) urticaria pigmentosa, the most common form, is a maculopapular pigmented rash affecting the papillarydermis; (2) diffuse cutaneous mastocytosis, occurring rarely but almost entirely in children, does not produce the maculopapular rash; instead, the skin is relatively smooth but may be red or thickened and on biopsy have infiltration in the papillary and reticular dermis with mast cells; (3) mastocytoma of skin, a single lesion, most often on the trunk or wrist, in which a tumor mass composed of mast cells forms.

Systemic Mastocytosis

Clinical manifestations of systemic mastocytosis may be mediated either by the infiltration of organs with mast cells or the release of mediators from the mast cells, including proteases, histamine, eicosanoids, or heparin. The signs and symptoms can be grouped into the following categories: (1) constitutional symptoms (fatigue, fever, weight loss, sweats); (2) skin manifestations of mast cell infiltration (pruritus, urticaria, rash, dermatographism); (3) mediator-related symptoms (abdominal pain, flushing, syncope, hypertension, headache, tachycardia, diarrhea); and (4) bone-related symptoms (fracture, pain, arthralgia). Patients may have splenomegaly, anemia, and either increases or decreases in platelet count and white blood cell count. Bone marrow involvement is common and can progress to crowd out normal hematopoietic elements. Eosinophilia can be seen to such a marked

degree that a primary eosinophilic disorder is suspected. Serum tryptase is a useful marker for mast cell mass. Tryptase levels >20 ng/mL indicate the presence of systemic mastocytosis. Tryptase levels tend to be <15 ng/mL in cutaneous mastocytosis.

In the Mayo Clinic series, 40% of patients with systemic mastocytosis had an associated myeloid neoplasm; in about 45% of these patients, the associated tumor was a myeloproliferative syndrome; in 29% it was chronic myelomonocytic leukemia, and in 23% it was a myelodysplastic syndrome. Eosinophilia was noted in about one-third of the patients with an associated myeloid neoplasm. Median survival for patients with systemic mastocytosis and another myeloid neoplasm was about 2 years.

In the absence of a myeloid neoplasm, systemic mastocytosis can have an indolent or an aggressive clinical course. Patients who follow a more indolent course do not have high levels of tryptase or bone marrow mastocytosis, no dysplasia, no hepatosplenomegaly, no skeletal involvement, normal blood counts, and no symptoms of malabsorption with weight loss. Such patients comprised 46% of the Mayo Clinic experience and had a median survival of 16+ years.

By contrast, about 12% of the Mayo Clinic patients with systemic mastocytosis had an aggressive course. They often had anemia and thrombocytopenia, B symptoms, and hepatosplenomegaly. Their median survival was about 3.5 years.

However, many of these patients were diagnosed and treated before it became widely known that the majority of patients with systemic mastocytosis have activating mutations of c-KIT, most notably *KIT*D816V. KIT is one of the kinases that are inhibited by imatinib, but this mutation is relatively resistant to its effects. The second- and third-generation inhibitors have not been tested. Interferon α produces response in about half of patients and the responses last about 1 year. Hydroxyurea may unpack the marrow sufficiently to restore hematopoiesis. Median responses last 2.5 years. Cladribine produced responses in 55% of patients, with responses lasting about a year, and is a reasonable first-line therapy.

Mast Cell Sarcoma/Leukemia

Mast cell sarcoma is very rare but consists of a destructive tumor mass composed of atypical-looking immature mast cells. They may appear de novo or in the setting of systemic mastocytosis as a solitary mass that is growing unusually fast compared with other involved sites. When the bone marrow becomes >50% mast cells, one may see circulating mast cells accounting for >10% of the white blood cell count. This finding permits a diagnosis of mast cell leukemia.

Extracutaneous Mastocytoma

These rare tumors of normal-appearing mast cells often present in the lung. The treatment experience is anecdotal.

FURTHER READINGS

Chandesris MO et al: Enteropathy-associated T-cell lymphoma: A review on clinical presentation, diagnosis, therapeutic strategies, and perspectives. Gastroenterol Clin Biol 34:590, 2010[PMID: 21050687]

Elliott MA et al: WHO-defined chronic neutrophilic leukemia: A long-term analysis of 12 cases and a critical review of the literature. Leukemia 19:313, 2005[PMID: 15549147]

Faris JE, LaCasce AS: Primary mediastinal large B-cell lymphoma. Clin Adv Hematol Oncol 7:125, 2009[PMID: 19367254]

Forconi F: Hairy cell leukaemia: Biological and clinical overview from immunogenetic insights. Hematol Oncol Nov 2, 2010 [Epub ahead of print]

Greer JP, Mosse CA: Natural killer cell neoplasms. Curr Hematol Malig Rep 4:245, 2009[PMID: 20425414]

Kotlyar DS et al: A systematic review of factors that contribute to hepatosplenic T-cell lymphoma in patients with inflammatory bowel disease. Clin Gastroenterol Hepatol 9:36, 2011 [PMID: 20888436]

Pardanani A, Tefferi A: Systemic mastocytosis in adults: A review on prognosis and treatment based on 342 Mayo Clinic patients and current literature. Curr Opin Hematol 17:125, 2010[PMID: 20075725]

Tefferi A et al: Hypereosinophilic syndrome and clonal eosinophilia: Point-of-care diagnostic algorithm and treatment update. Mayo Clin Proc 85:158, 2010[PMID: 20053713]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PLASMA CELL DISORDERS: INTRODUCTION

The *plasma cell disorders* are monoclonal neoplasms related to each other by virtue of their development from common progenitors in the B lymphocyte lineage. Multiple myeloma, Waldenström's macroglobulinemia, primary amyloidosis (Chap. 112), and the heavy chain diseases comprise this group and may be designated by a variety of synonyms such as *monoclonal gammopathies*, *paraproteinemias*, *plasma cell dyscrasias*, and *dysproteinemias*. Mature B lymphocytes destined to produce IgG bear surface immunoglobulin molecules of both M and G heavy chain isotypes with both isotypes having identical idiotypes (variable regions). Under normal circumstances, maturation to antibody-secreting plasma cells and their proliferation is stimulated by exposure to the antigen for which the surface immunoglobulin is specific; however, in the plasma cell disorders the control over this process is lost. The clinical manifestations of all the plasma cell disorders relate to the expansion of the neoplastic cells, to the secretion of cell products (immunoglobulin molecules or subunits, lymphokines), and to some extent to the host's response to the tumor. **Normal development of B lymphocytes is discussed in Chap. 314.**

There are three categories of structural variation among immunoglobulin molecules that form antigenic determinants, and these are used to classify immunoglobulins (Chap. 314). *Isotypes* are those determinants that distinguish among the main classes of antibodies of a given species and are the same in all normal individuals of that species. Therefore, isotypic determinants are, by definition, recognized by antibodies from a distinct species (heterologous sera) but not by antibodies from the same species (homologous sera). There are five heavy chain isotypes (M, G, A, D, E) and two light chain isotypes (κ, λ). *Allotypes* are distinct determinants that reflect regular small differences between individuals of the same species in the amino acid sequences of otherwise similar immunoglobulins. These differences are determined by allelic genes; by definition, they are detected by antibodies made in the same species. *Idiotypes* are the third category of antigenic determinants. They are unique to the molecules produced by a given clone of antibody-producing cells. Idiotypes are formed by the unique structure of the antigen-binding portion of the molecule.

Antibody molecules (Fig. 314-8) are composed of two heavy chains (~50,000 mol wt) and two light chains (~25,000 mol wt). Each chain has a constant portion (limited amino acid sequence variability) and a variable region (extensive sequence variability). The light and heavy chains are linked by disulfide bonds and are aligned so that their variable regions are adjacent to one another. This variable region forms the antigen recognition site of the antibody molecule; its unique structural features form a particular set of determinants, or idiotypes, that are reliable markers for a particular clone of cells because each antibody is formed and secreted by a single clone. Each chain is specified by distinct genes, synthesized separately, and assembled into an intact antibody molecule after translation. Because of the mechanics of the gene rearrangements necessary to specify the immunoglobulin variable regions (VDJ joining for the heavy chain, VJ joining for the light chain), a particular clone rearranges only one of the two chromosomes to produce an immunoglobulin molecule of only one light chain isotype and only one allotype (allelic exclusion). After exposure to antigen, the variable region may become associated with a new heavy chain isotype (class switch). Each clone of cells performs these sequential gene arrangements in a unique way. This results in each clone producing a unique immunoglobulin molecule. In most plasma cells, light chains are synthesized in slight excess, secreted as free light chains, and are cleared by the kidney, but <10 mg of such light chains is excreted per day.

Electrophoretic analysis permits separation of components of the serum proteins **(Fig. 111-1)**. The immunoglobulins move heterogeneously in an electric field and form a broad peak in the gamma region. The γ globulin region of the electrophoretic pattern is usually increased in the sera of patients with plasma cell tumors. There is a sharp spike in this region called an *M component* (M for monoclonal). Less commonly, the M component may appear in the β_2 or α_2 globulin region. The monoclonal antibody must be present at a concentration of at least 5 g/L (0.5 g/dL) to be accurately quantitated by this method. This corresponds to ~10^9 cells producing the antibody. Confirmation that such an M component is truly monoclonal and the type of immunoglobulin is determined by immunoelectrophoresis that reveals a single heavy and/or light chain type. Hence immunoelectrophoresis and electrophoresis provide qualitative and quantitative assessment of the M component, respectively. Once the presence of an M component has been confirmed, electrophoresis provides the more practical information for managing patients with monoclonal gammopathies. In a given patient, the amount of M component in the serum is a reliable measure of the tumor burden. This makes the M component an excellent tumor marker, yet it is not specific enough to be used to screen asymptomatic patients. In addition to the plasma cell disorders, M components may be detected in other lymphoid neoplasms such as chronic lymphocytic leukemia and lymphomas of B or T cell origin; nonlymphoid neoplasms such as chronic myeloid leukemia, breast cancer, and colon cancer; a variety of nonneoplastic conditions such as cirrhosis, sarcoidosis, parasitic diseases, Gaucher disease, and pyoderma gangrenosum; and a number of autoimmune conditions, including rheumatoid arthritis, myasthenia gravis, and cold agglutinin disease. At least two very rare skin diseases—lichen myxedematosus, or papular mucinosis, and necrobiotic xanthogranuloma—are associated with a monoclonal gammopathy. In papular mucinosis, highly cationic IgG is deposited in the dermis of patients. This organ specificity may reflect the specificity of the antibody for some antigenic component of the dermis. Necrobiotic xanthogranuloma is a histiocytic infiltration of the skin, usually of the face, that produces red or

yellow nodules that can enlarge to plaques. Some 10% progress to myeloma. Five percent of patients with sensory motor neuropathy are associated with monoclonal protein.

Figure 111-1

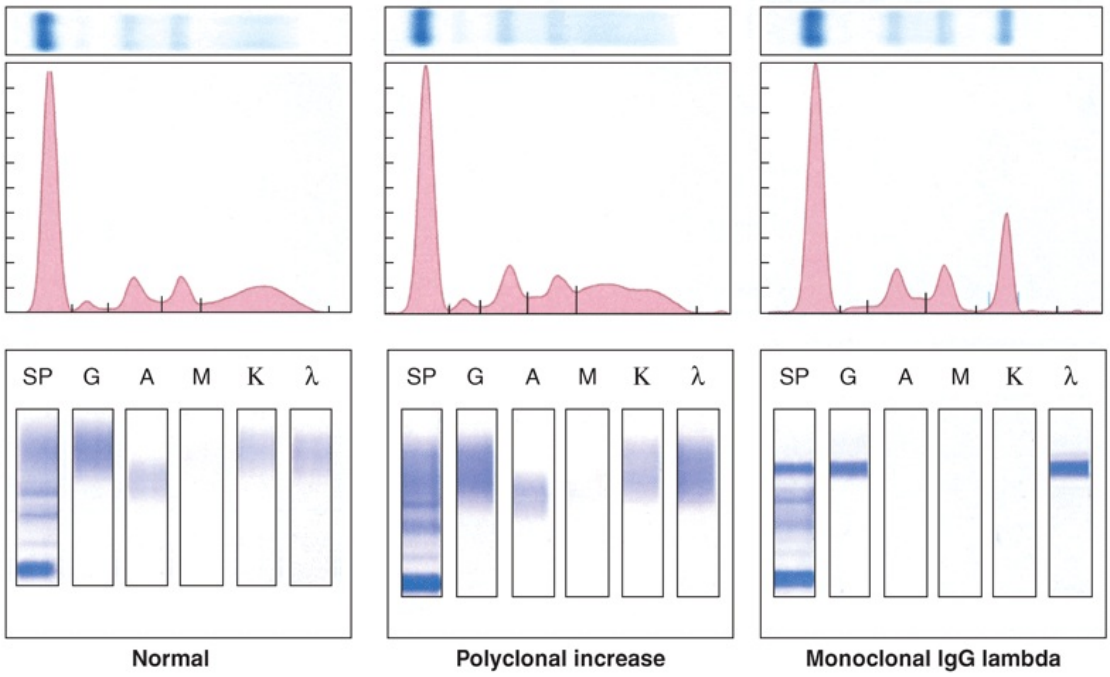

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Representative patterns of serum electrophoresis and immunofixation. The upper panels represent agarose gel, middle panels are the densitometric tracing of the gel, and lower panels are immunofixation patterns. Panel on the left illustrates the normal pattern of serum protein on electrophoresis. Since there are many different immunoglobulins in the serum, their differing mobilities in an electric field produce a broad peak. In conditions associated with increases in polyclonal immunoglobulin, the broad peak is more prominent (middle panel). In monoclonal gammopathies, the predominance of a product of a single cell produces a "church spire" sharp peak, usually in the γ globulin region (right panel). The immunofixation (lower panel) identifies the type of immunoglobulin. For example, normal and polyclonal increase in immunoglobulins produce no distinct bands; however, the right panel shows distinct bands in IgG and lambda protein lanes, confirming the presence of IgG lambda monoclonal protein. *(Courtesy of Dr. Neal I Lindeman; with permission.)*

The nature of the M component is variable in plasma cell disorders. It may be an intact antibody molecule of any heavy chain subclass, or it may be an altered antibody or fragment. Isolated light or heavy chains may be produced. In some plasma cell tumors such as extramedullary or solitary bone plasmacytomas, less than one-third of patients will have an M component. In ~20% of myelomas, only light chains are produced and in most cases are secreted in the urine as Bence Jones proteins. The frequency of myelomas of a particular heavy chain class is roughly proportional to the serum concentration, and therefore IgG myelomas are more common than IgA and IgD myelomas. In approximately 1% of patients with myeloma, biclonal or triclonal gammopathy is observed.

MULTIPLE MYELOMA

Definition

Multiple myeloma represents a malignant proliferation of plasma cells derived from a single clone. The tumor, its products, and the host response to it result in a number of organ dysfunctions and symptoms, including bone pain or fracture, renal failure, susceptibility to infection, anemia, hypercalcemia, and occasionally clotting abnormalities, neurologic symptoms, and manifestations of hyperviscosity.

Etiology

The cause of myeloma is not known. Myeloma occurred with increased frequency in those exposed to the radiation of nuclear warheads in World War II after a 20-year latency. Myeloma has been seen more commonly than expected among farmers, wood workers, leather workers, and those exposed to petroleum products. A variety of chromosomal alterations with prognostic significance has been found in patients with myeloma; 13q14 deletions, 17p13 deletions, and translocations t(11;14)(q13;q32) and t(4;14)(p16;q32) predominate, and evidence is strong that errors in switch recombination—the genetic mechanism to change antibody heavy chain isotype—participate in the transformation process. However, no common molecular pathogenetic pathway has yet emerged. The neoplastic event in myeloma may involve cells earlier in B cell differentiation than the plasma cell. Interleukin (IL)-6 may play a role in driving myeloma cell proliferation. It remains difficult to distinguish benign from malignant plasma cells on the basis of morphologic criteria in all but a few cases **(Fig. 111-2)**.

Figure 111-2

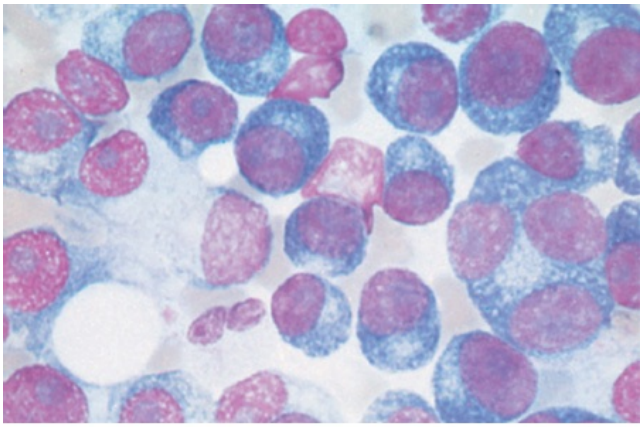

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple myeloma (marrow). The cells bear characteristic morphologic features of plasma cells, round or oval cells with an eccentric nucleus composed of coarsely clumped chromatin, a densely basophilic cytoplasm, and a perinuclear clear zone containing the Golgi apparatus. Binucleate and multinucleate malignant plasma cells can be seen.

Incidence and Prevalence

Estimated 20,180 new cases of myeloma were diagnosed in 2010, and 10,650 people died from the disease in the United States. Myeloma increases in incidence with age. The median age at diagnosis is 70 years; it is uncommon under age 40. Males are more commonly affected than females, and blacks have nearly twice the incidence of whites. Myeloma accounts for ~1% of all malignancies in whites and 2% in blacks, and 13% of all hematologic cancers in whites and 33% in blacks.

Global Considerations

The incidence of myeloma is highest in African Americans and Pacific islanders; intermediate in Europeans and North American whites; and lowest in developing countries including Asia. The higher incidence in more developed countries may result from the combination of a longer life expectancy and more frequent medical surveillance. Incidence of multiple myeloma in other ethnic groups including native Hawaiians, female Hispanics, American Indians from New Mexico, and Alaskan natives is higher relative to U.S. whites in the same geographic area. Chinese and Japanese populations have a lower incidence than whites. Immunoproliferative small intestinal disease with alpha heavy chain disease is most prevalent in the Mediterranean area. Despite these differences in prevalence, the characteristics, response to therapy, and prognosis of myeloma are similar worldwide.

Pathogenesis and Clinical Manifestations (Table 111–1)

Table 111–1 Clinical Features of Multiple Myeloma

Clinical Finding	Underlying Cause and Pathogenetic Mechanism
Hypercalcemia, osteoporosis, pathologic fractures, lytic bone lesions, bone pain	Tumor expansion, production of osteoclast activating factor by tumor cells, osteoblast inhibitory factors
Renal failure	Hypercalcemia, light chain deposition, amyloidosis, urate nephropathy, drug toxicity (nonsteroidal anti-inflammatory agents, bisphosphonates), contrast dye
Easy fatigue/anemia	Bone marrow infiltration, production of inhibitory factors, hemolysis, decreased red cell production, decreased erythropoietin levels
Recurrent infections	Hypogammaglobulinemia, low CD4 count, decreased neutrophil migration
Neurologic symptoms	Hyperviscosity, cryoglobulinemia, amyloid deposits, hypercalcemia, nerve compression, antineuronal antibody, POEMS syndrome, therapy-related toxicity
Nausea and vomiting	Renal failure, hypercalcemia
Bleeding/clotting disorder	Interference with clotting factors, antibody to clotting factors, amyloid damage of endothelium, platelet dysfunction, antibody coating of platelet, therapy-related hypercoagulable defects

Abbreviation: POEMS, polyneuropathy, organomegaly, endocrinopathy, multiple myeloma, and skin changes.

Multiple myeloma (MM) cells bind via cell-surface adhesion molecules to bone marrow stromal cells (BMSCs) and extracellular matrix (ECM), which triggers MM cell growth, survival, drug resistance, and migration in the bone marrow milieu **(Fig. 111-3)**. These effects are due both to direct MM cell–BMSC binding and to induction of various cytokines, including IL-6, insulin-like growth factor type I (IGF-I), vascular endothelial growth factor (VEGF), and stromal cell–derived growth factor (SDF)-1α. Growth, drug resistance, and migration are mediated

via Ras/Raf/mitogen-activated protein kinase, PI3-K/Akt, and protein kinase C signaling cascades, respectively.

Figure 111-3

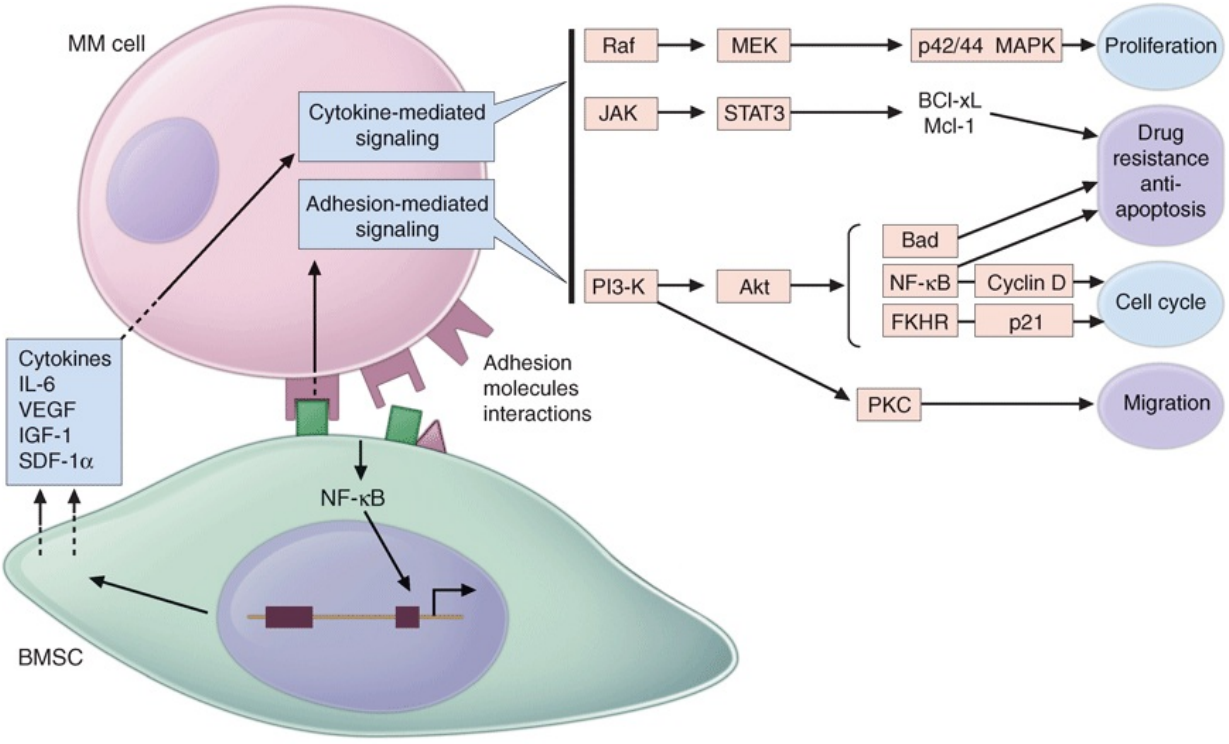

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis of multiple myeloma. Multiple myeloma cells interact with bone marrow stromal cells and extracellular matrix proteins via adhesion molecules, triggering adhesion-mediated signaling as well as cytokine production. This triggers cytokine-mediated signaling that provides growth, survival, and antiapoptotic effects as well as development of drug resistance. HSP, heparin sulfate proteoglycan.

Bone pain is the most common symptom in myeloma, affecting nearly 70% of patients. The pain usually involves the back and ribs, and unlike the pain of metastatic carcinoma, which often is worse at night, the pain of myeloma is precipitated by movement. Persistent localized pain in a patient with myeloma usually signifies a pathologic fracture. The bone lesions of myeloma are caused by the proliferation of tumor cells, activation of osteoclasts that destroy bone, and suppression of osteoblasts that form new bone. The increased osteoclast activity is mediated by osteoclast activating factors (OAF) made by the myeloma cells [OAF activity can be mediated by several cytokines, including IL-1, lymphotoxin, VEGF, receptor activator of NF-κB (RANK) ligand, macrophage inhibitory factor (MIP)-1α, and tumor necrosis factor (TNF)]. The bone lesions are lytic in nature and are rarely associated with osteoblastic new bone formation due to their suppression by dickhoff-1 (DKK-1) produced by myeloma cells. Therefore, radioisotopic bone scanning is less useful in diagnosis than is plain radiography. The bony lysis results in substantial mobilization of calcium from bone, and serious acute and chronic complications of hypercalcemia may dominate the clinical picture (see below). Localized bone lesions may expand to the point that mass lesions may be palpated, especially on the skull **(Fig. 111-4)**, clavicles, and sternum, and the collapse of vertebrae may lead to spinal cord compression.

Figure 111-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bony lesions in multiple myeloma. The skull demonstrates the typical "punched out" lesions characteristic of multiple myeloma. The lesion represents a purely osteolytic lesion with little or no osteoblastic activity. *(Courtesy of Dr. Geraldine Schechter; with permission.)*

The next most common clinical problem in patients with myeloma is susceptibility to bacterial infections. The most common infections are pneumonias and pyelonephritis, and the most frequent pathogens are *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Klebsiella pneumoniae* in the lungs and *Escherichia coli* and other gram-negative organisms in the urinary tract. In ~25% of patients, recurrent infections are the presenting features, and >75% of patients will have a serious infection at some time in their course. The susceptibility to infection has several contributing causes. First, patients with myeloma have diffuse hypogammaglobulinemia if the M component is excluded. The hypogammaglobulinemia is related to both decreased production and increased destruction of normal antibodies. Moreover, some patients generate a population of circulating regulatory cells in response to their myeloma that can suppress normal antibody synthesis. In the case of IgG myeloma, normal IgG antibodies are broken down more rapidly than normal because the catabolic rate for IgG antibodies varies directly with the serum concentration. The large M component results in fractional catabolic rates of 8–16% instead of the normal 2%. These patients have very poor antibody responses, especially to polysaccharide antigens such as those on bacterial cell walls. Most measures of T cell function in myeloma are normal, but a subset of CD4+ cells may be decreased. Granulocyte lysozyme content is low, and granulocyte migration is not as rapid as normal in patients with myeloma, probably the result of a tumor product. There are also a variety of abnormalities in complement functions in myeloma patients. All these factors contribute to the immune deficiency of these patients. Some commonly used therapeutic agents, e.g., dexamethasone, suppress immune responses and increase susceptibility to infection.

Renal failure occurs in nearly 25% of myeloma patients, and some renal pathology is noted in more than 50%. Many factors contribute to this. Hypercalcemia is the most common cause of renal failure. Glomerular deposits of amyloid, hyperuricemia, recurrent infections, frequent use of nonsteroidal anti-inflammatory agents for pain control, use of iodinated contrast dye for imaging, bisphosphonate use, and occasional infiltration of the kidney by myeloma cells all may contribute to renal dysfunction. However, tubular damage associated with the excretion of light chains is almost always present. Normally, light chains are filtered, reabsorbed in the tubules, and catabolized. With the increase in the amount of light chains presented to the tubule, the tubular cells become overloaded with these proteins, and tubular damage results either directly from light chain toxic effects or indirectly from the release of intracellular lysosomal enzymes. The earliest manifestation of this tubular damage is the adult Fanconi syndrome (a type 2 proximal renal tubular acidosis), with loss of glucose and amino acids, as well as defects in the ability of the kidney to acidify and concentrate the urine. The proteinuria is not accompanied by hypertension, and the protein is nearly all light chains. Generally, very little albumin is in the urine because glomerular function is usually normal. When the glomeruli are involved, nonselective proteinuria is also observed. Patients with myeloma also have a decreased anion gap [i.e., $Na^+ - (Cl^- + HCO_3^-)$] because the M component is cationic, resulting in retention of chloride. This is often accompanied by hyponatremia that is felt to be artificial (pseudohyponatremia) because each volume of serum has less water as a result of the increased protein. Renal dysfunction due to light chain deposition disease, light chain cast nephropathy, and amyloidosis is partially reversible with effective therapy. Myeloma patients are susceptible to developing acute renal failure if they become dehydrated.

Normocytic and normochromic anemia occurs in ~80% of myeloma patients. It is usually related to the replacement of normal marrow by expanding tumor cells, to the inhibition of hematopoiesis by factors made by the tumor, and to reduced production of erythropoietin by the kidney. In addition, mild hemolysis may contribute to the anemia. A larger than expected fraction of patients may have megaloblastic anemia due to either folate or vitamin B12 deficiency. Granulocytopenia and thrombocytopenia are very rare except when therapy-induced. Clotting abnormalities may be seen due to the failure of antibody-coated platelets to function properly or to the interaction of the M component with clotting factors I, II, V, VII, or VIII. Deep venous thrombosis is also observed with use of thalidomide or lenalidomide in combination with dexamethasone. Raynaud's phenomenon and

impaired circulation may result if the M component forms cryoglobulins, and hyperviscosity syndromes may develop depending on the physical properties of the M component (most common with IgM, IgG3, and IgA paraproteins). Hyperviscosity is defined on the basis of the relative viscosity of serum as compared with water. Normal relative serum viscosity is 1.8 (i.e., serum is normally almost twice as viscous as water). Symptoms of hyperviscosity occur at a level greater than 4 centipoise (cP), which is usually reached at paraprotein concentrations of ~40 g/L (4 g/dL) for IgM, 50 g/L (5 g/dL) for IgG3, and 70 g/L (7 g/dL) for IgA.

Although neurologic symptoms occur in a minority of patients, they may have many causes. Hypercalcemia may produce lethargy, weakness, depression, and confusion. Hyperviscosity may lead to headache, fatigue, visual disturbances, and retinopathy. Bony damage and collapse may lead to cord compression, radicular pain, and loss of bowel and bladder control. Infiltration of peripheral nerves by amyloid can be a cause of carpal tunnel syndrome and other sensorimotor mono- and polyneuropathies. Neuropathy associated with monoclonal gammopathy of undetermined significance (MGUS) and myeloma is more frequently sensory than motor neuropathy and is associated with IgM more than other isotypes. Sensory neuropathy is also a side effect of thalidomide and bortezomib therapy.

Many of the clinical features of myeloma, e.g., cord compression, pathologic fractures, hyperviscosity, sepsis, and hypercalcemia, can present as medical emergencies. Despite the widespread distribution of plasma cells in the body, tumor expansion is dominantly within bone and bone marrow and, for reasons unknown, rarely causes enlargement of spleen, lymph nodes, or gut-associated lymphatic tissue.

Diagnosis and Staging

The classic triad of myeloma is marrow plasmacytosis (>10%), lytic bone lesions, and a serum and/or urine M component. Bone marrow plasma cells are CD138+ and monoclonal. The most important differential diagnosis in patients with myeloma involves their separation from individuals with MGUS or smoldering multiple myeloma (SMM). MGUS are vastly more common than myeloma, occurring in 1% of the population older than age 50 years and in up to 10% of individuals older than age 75 years. The diagnostic criteria for MGUS, SMM, and myeloma are described in Table 111–2. When bone marrow cells are exposed to radioactive thymidine in order to quantitate dividing cells, patients with MGUS always have a labeling index <1%, whereas patients with myeloma always have a labeling index >1%. Although ~1% per year of patients with MGUS go on to develop myeloma, all myeloma is preceded by MGUS. Non-IgG subtype, abnormal kappa/lambda free light chain ratio, and serum M protein >15 g/L (1.5 g/dL) are associated with higher incidence of progression of MGUS to myeloma. The features responsible for higher risk of progression from smoldering myeloma to MM are bone marrow plasmacytosis >30%, abnormal kappa/lambda free light chain ratio, and serum M protein >30 g/L (3 g/dL). Typically, patients with MGUS and smoldering myeloma require no therapy. There are two important variants of myeloma, solitary bone plasmacytoma and extramedullary plasmacytoma. These lesions are associated with an M component in <30% of the cases, they may affect younger individuals, and both are associated with median survivals of ≥10 years. Solitary bone plasmacytoma is a single lytic bone lesion without marrow plasmacytosis. Extramedullary plasmacytomas usually involve the submucosal lymphoid tissue of the nasopharynx or paranasal sinuses without marrow plasmacytosis. Both tumors are highly responsive to local radiation therapy. If an M component is present, it should disappear after treatment. Solitary bone plasmacytomas may recur in other bony sites or evolve into myeloma. Extramedullary plasmacytomas rarely recur or progress.

Table 111–2 Diagnostic Criteria for Multiple Myeloma, Myeloma Variants, and Monoclonal Gammopathy of Undetermined Significance

Monoclonal Gammopathy of Undetermined Significance (MGUS)
M protein in serum <30 g/L
Bone marrow clonal plasma cells <10%
No evidence of other B cell proliferative disorders
No myeloma-related organ or tissue impairment (no end organ damage, including bone lesions)[a]
Asymptomatic Myeloma (Smoldering Myeloma)
M protein in serum ≥30 g/L *and/or*
Bone marrow clonal plasma cells ≥10%
No myeloma-related organ or tissue impairment (no end organ damage, including bone lesions)[a] or symptoms
Symptomatic Multiple Myeloma
M protein in serum and/or urine
Bone marrow (clonal) plasma cells[b] or plasmacytoma
Myeloma-related organ or tissue impairment (end organ damage, including bone lesions)
Nonsecretory Myeloma
No M protein in serum and/or urine with immunofixation
Bone marrow clonal plasmacytosis ≥10% or plasmacytoma
Myeloma-related organ or tissue impairment (end organ damage, including bone lesions)[a]
Solitary Plasmacytoma of Bone

No M protein in serum and/or urine[c]

Single area of bone destruction due to clonal plasma cells

Bone marrow not consistent with multiple myeloma

Normal skeletal survey (and MRI of spine and pelvis if done)

No related organ or tissue impairment (no end organ damage other than solitary bone lesion)[a]

[a]Myeloma-related organ or tissue impairment (end organ damage) (ROTI): Calcium levels increased: serum calcium >0.25 mmol/L above the upper limit of normal or >2.75 mmol/L; renal insufficiency: creatinine >173 mmol/L; anemia: hemoglobin 2 g/dL below the lower limit of normal or hemoglobin <10 g/dL; bone lesions: lytic lesions or osteoporosis with compression fractures (MRI or CT may clarify); other: symptomatic hyperviscosity, amyloidosis, recurrent bacterial infections (>2 episodes in 12 months).

[b]If flow cytometry is performed, most plasma cells (>90%) will show a "neoplastic" phenotype.

[c]A small M component may sometimes be present.

The clinical evaluation of patients with myeloma includes a careful physical examination searching for tender bones and masses. Chest and bone radiographs may reveal lytic lesions or diffuse osteopenia. MRI offers a sensitive means to document extent of bone marrow infiltration and cord or root compression in patients with pain syndromes. A complete blood count with differential may reveal anemia. Erythrocyte sedimentation rate is elevated. Rare patients (~2%) may have plasma cell leukemia with >2000 plasma cells/μL. This may be seen in disproportionate frequency in IgD (12%) and IgE (25%) myelomas. Serum calcium, urea nitrogen, creatinine, and uric acid levels may be elevated. Protein electrophoresis and measurement of serum immunoglobulins and free light chains are useful for detecting and characterizing M spikes, supplemented by immunoelectrophoresis, which is especially sensitive for identifying low concentrations of M components not detectable by protein electrophoresis. A 24-h urine specimen is necessary to quantitate Bence Jones protein excretion. Serum alkaline phosphatase is usually normal even with extensive bone involvement because of the absence of osteoblastic activity. It is also important to quantitate serum β_2-microglobulin (see below).

The serum M component will be IgG in 53% of patients, IgA in 25%, and IgD in 1%; 20% of patients will have only light chains in serum and urine. Dipsticks for detecting proteinuria are not reliable at identifying light chains, and the heat test for detecting Bence Jones protein is falsely negative in ~50% of patients with light chain myeloma. Fewer than 1% of patients have no identifiable M component; these patients usually have light chain myeloma in which renal catabolism has made the light chains undetectable in the urine. In most of these patients, light chains can now be detected by serum free light chain assay. IgD myeloma may also present as light chain myeloma. About two-thirds of patients with serum M components also have urinary light chains. The light chain isotype may have an impact on survival. Patients secreting lambda light chains have a significantly shorter overall survival than those secreting kappa light chains. Whether this is due to some genetically important determinant of cell proliferation or because lambda light chains are more likely to cause renal damage and form amyloid than are kappa light chains is unclear. The heavy chain isotype may have an impact on patient management as well. About half of patients with IgM paraproteins develop hyperviscosity compared with only 2–4% of patients with IgA and IgG M components. Among IgG myelomas, it is the IgG3 subclass that has the highest tendency to form both concentration- and temperature-dependent aggregates, leading to hyperviscosity and cold agglutination at lower serum concentrations.

The staging systems for patients with myeloma **(Table 111–3)** are functional systems for predicting survival and are based on a variety of clinical and laboratory tests, unlike the anatomic staging systems for solid tumors. The Durie-Salmon staging system used previously has been found not to predict prognosis after treatment with high-dose therapy or the novel targeted therapies that have emerged.

Table 111–3 International Staging System

	Stage	Median Survival, Months
β_2M < 3.5, alb ≥ 3.5	I (28%)	62
β_2M < 3.5, alb < 3.5 *or* β_2M = 3.5–5.5	II (39%)	44
β_2M > 5.5	III (33%)	29

Note: (#), % patients presenting at each stage.

Abbreviations: β_2M, serum β_2-microglobulin in mg/L; alb, serum albumin in g/Dl.

Serum β_2-microglobulin is the single most powerful predictor of survival and can substitute for staging. β_2-Microglobulin is a protein of 11,000 mol wt with homologies to the constant region of immunoglobulins that is the light chain of the class I major histocompatibility antigens (HLA-A, -B, -C) on the surface of every cell. Patients with β_2-microglobulin levels <0.004 g/L have a median survival of 43 months and those with levels >0.004 g/L only 12 months. Serum β_2-microglobulin and albumin levels are the basis for a three-stage International Staging System (ISS) (Table 111–3). It is also felt that once the diagnosis of myeloma is firm, histologic features of atypia may also exert an influence on prognosis. High labeling index and high levels of lactate dehydrogenase are also

associated with poor prognosis.

Other factors that may influence prognosis are the presence and number of cytogenetic abnormalities, hypodiploidy, chromosome 13q and 17p deletion, translocations t(4;14) and t(14;16); circulating plasma cells; performance status; as well as serum levels of soluble IL-6 receptor, C-reactive protein, hepatocyte growth factor, C-terminal cross-linked telopeptide of collagen I, transforming growth factor (TGF)-ß, and syndecan-1. Microarray profiling and comparative genomic hybridization have formed the basis for RNA- and DNA-based prognostic staging systems, respectively. The ISS system is the most widely used method of assessing prognosis (Table 111–3).

Treatment: Multiple Myeloma

About 10% of patients with myeloma will have an indolent course (smoldering myeloma) demonstrating only very slow progression of disease over many years. Such patients only require antitumor therapy when the disease becomes symptomatic with development of anemia, hypercalcemia, progressive lytic bone lesions, renal dysfunction, progressive rise in serum myeloma protein levels and/or Bence Jones proteinuria, or recurrent infections. Patients with solitary bone plasmacytomas and extramedullary plasmacytomas may be expected to enjoy prolonged disease-free survival after local radiation therapy to a dose of around 40 Gy. There is a low incidence of occult marrow involvement in patients with solitary bone plasmacytoma. Such patients are usually detected because their serum M component falls slowly or disappears initially only to return after a few months. These patients respond well to systemic therapy.

Patients with symptomatic and/or progressive myeloma require therapeutic intervention. In general, such therapy is of two sorts: systemic therapy to control the progression of myeloma and symptomatic supportive care to prevent serious morbidity from the complications of the disease. Therapy can significantly prolong survival and improve the quality of life for myeloma patients.

The initial standard treatment for newly diagnosed myeloma is dependent on whether or not the patient is a candidate for high-dose chemotherapy with autologous stem cell transplant.

In patients who are transplant candidates, alkylating agents such as melphalan should be avoided since they damage stem cells, leading to decreased ability to collect stem cells for autologous transplant. Newer agents combined with pulsed glucocorticoids have now become standard of care as induction therapy in newly diagnosed patients. Two phase II studies have combined thalidomide with dexamethasone as initial therapy for newly diagnosed multiple myeloma in transplant candidates and reported rapid responses in two-thirds of patients, while allowing for successful harvesting of peripheral blood stem cells for transplantation. A randomized phase III trial showed significantly higher response rates for thalidomide (200 mg PO qhs) plus dexamethasone (40 mg for 4 days every 2 weeks) compared to dexamethasone alone, setting the stage for use of this combination as standard therapy in newly diagnosed patients. Importantly, novel agents bortezomib, a proteasome inhibitor, and lenalidomide, an immunomodulatory derivative of thalidomide, have similarly been combined with dexamethasone and obtained high response rates (>80%) without compromising stem cell collection for transplantation. Their superior toxicity profile with improved efficacy has made them the preferred agents for induction therapy. Efforts to improve the fraction of patients responding and the degree of response have involved adding agents to the treatment program. Combination of lenalidomide, bortezomib, and dexamethasone achieves close to 100% response rate, and other similar three-drug combinations (bortezomib, thalidomide, and dexamethasone or bortezomib, cyclophosphamide, and dexamethasone) achieve >90% response rate. Initial therapy is continued until maximal cytoreduction.

In patients who are not transplant candidates, besides the options available for transplant candidates, therapy consisting of intermittent pulses of an alkylating agent, melphalan with prednisone, has been utilized. The usual doses of melphalan/prednisone (MP) are melphalan, 0.25 mg/kg per day, and prednisone, 1 mg/kg per day for 4 days. Doses may need adjustment due to unpredictable absorption and based on marrow tolerance. However, a number of studies have combined novel agents with MP combination and reported superior response and survival outcome. In patients >65 years old, combining thalidomide with MP (MPT) obtains higher response rates and overall survival compared to MP alone. Similarly, significantly improved response (71 vs 35%) and overall survival (3-year survival 72 vs 59%) were observed with combination of bortezomib with MP compared to MP alone. Lenalidomide added to MP followed by lenalidomide maintenance also prolonged progression-free survival compared to MP alone. These combinations of novel agents with MP also achieve high complete response rates (MPT ~ 15%; MPV ~ 30%, MPR ~ 20%, and MP ~ 2-4%). Patients responding to therapy generally have a prompt and gratifying reduction in bone pain, hypercalcemia, and anemia and often have fewer infections. Improvement in the serum M component may lag behind the symptomatic improvement. The fall in M component depends on the rate of tumor kill and the fractional catabolic rate of immunoglobulin, which in turn depends on the serum concentration (for IgG). Light chain excretion, with a functional half-life of ~6 h, may fall within the first week of treatment. Since urine light chain levels may relate to renal tubular function, they are not a reliable measure of tumor cell kill; however, improvements in serum free light chain measurement are often seen sooner. Although patients may not achieve complete remission, clinical responses may last long periods of time. The important feature of the level of the M protein is not how far or how fast it falls but the rate of its increase after therapy.

Randomized studies comparing standard-dose therapy to high-dose melphalan therapy (HDT) with hematopoietic stem cell support have shown that HDT can achieve high overall response rates and prolonged progression-free and overall survival; however, few, if any, patients are cured. Although complete responses are rare (<5%) with standard-dose chemotherapy, HDT achieves 25–40% complete responses. In randomized studies, HDT produced better median event-free survival in four of five studies, higher complete response rate in four of five trials, and better overall survival in three of five studies. A randomized study failed to show any significant difference in overall survival between early transplant after induction therapy versus delayed transplant at relapse. These data allow an option to delay transplant, especially with the availability of more

agents and combinations. Two successive HDTs (tandem transplants) are more effective than single HDT in the subset of patients who do not achieve a complete or very good partial response to the first transplant. Allogeneic transplants may also produce high response rates, but treatment-related mortality may be as high as 40%. Nonmyeloablative allogeneic transplantation is now under evaluation to reduce toxicity, while permitting an immune graft-vs-myeloma effect.

Oral prednisone maintenance therapy was effective in a single trial after standard-dose chemotherapy. Maintenance therapy prolongs remissions following standard-dose regimens as well as HDT. Thalidomide administered post-HDT prolongs relapse-free survival. Phase III studies have demonstrated improved outcome in patients receiving lenalidomide compared to placebo as maintenance therapy after HDT, and another phase III study showed prolonged progression-free survival after MP lenalidomide and lenalidomide maintenance therapy in nontransplant candidates.

Relapsed myeloma can be treated with novel agents including lenalidomide and/or bortezomib. These agents target not only the tumor cell but also the tumor cell–bone marrow interaction and the bone marrow milieu. These agents in combination with dexamethasone can achieve up to 60% partial responses and 10–15% complete responses in patients with relapsed disease. The combination of bortezomib and liposomal doxorubicin is active in relapsed myeloma. Thalidomide, if not used as initial therapy, can achieve responses in refractory cases. High-dose melphalan and stem cell transplant, if not used earlier, also have activity in patients with refractory disease.

The median overall survival of patients with myeloma is 7–8 years, with subsets of younger patients surviving more than 10 years. The major causes of death are progressive myeloma, renal failure, sepsis, or therapy-related myelodysplasia. Nearly a quarter of patients die of myocardial infarction, chronic lung disease, diabetes, or stroke—all intercurrent illnesses related more to the age of the patient group than to the tumor.

Supportive care directed at the anticipated complications of the disease may be as important as primary antitumor therapy. The hypercalcemia generally responds well to bisphosphonates, glucocorticoid therapy, hydration, and natriuresis. Calcitonin may add to the inhibitory effects of glucocorticoids on bone resorption. Bisphosphonates (e.g., pamidronate 90 mg or zoledronate 4 mg once a month) reduce osteoclastic bone resorption and preserve performance status and quality of life, decrease bone-related complications, and may also have antitumor effects. Osteonecrosis of the jaw and renal dysfunction can occur in a minority of cases. Treatments aimed at strengthening the skeleton, such as fluorides, calcium, and vitamin D, with or without androgens, have been suggested but are not of proven efficacy. Iatrogenic worsening of renal function may be prevented by maintaining a high fluid intake to prevent dehydration and to help excrete light chains and calcium. In the event of acute renal failure, plasmapheresis is ~10 times more effective at clearing light chains than peritoneal dialysis; however, its role in reversing renal failure remains controversial. Importantly, reducing the protein load by effective antitumor therapy with agents such as bortezomib may result in functional improvement. Urinary tract infections should be watched for and treated early. Plasmapheresis may be the treatment of choice for hyperviscosity syndromes. Although the pneumococcus is a dreaded pathogen in myeloma patients, pneumococcal polysaccharide vaccines may not elicit an antibody response. Prophylactic administration of IV γ globulin preparations is used in the setting of recurrent serious infections. Chronic oral antibiotic prophylaxis is probably not warranted. Patients developing neurologic symptoms in the lower extremities, severe localized back pain, or problems with bowel and bladder control may need emergency MRI and radiation therapy for cord compression. Most bone lesions respond to analgesics and chemotherapy, but certain painful lesions may respond most promptly to localized radiation. The anemia associated with myeloma may respond to erythropoietin along with hematinics (iron, folate, cobalamin). The pathogenesis of the anemia should be established and specific therapy instituted, where possible.

WALDENSTRÖM'S MACROGLOBULINEMIA

In 1948, Waldenström described a malignancy of lymphoplasmacytoid cells that secreted IgM. In contrast to myeloma, the disease was associated with lymphadenopathy and hepatosplenomegaly, but the major clinical manifestation was the hyperviscosity syndrome. The disease resembles the related diseases chronic lymphocytic leukemia, myeloma, and lymphocytic lymphoma. It originates from a post–germinal center B cell that has undergone somatic mutations and antigenic selection in the lymphoid follicle and has the characteristics of an IgM-bearing memory B cell. Waldenström's macroglobulinemia and IgM myeloma follow a similar clinical course, but therapeutic options are different. The diagnosis of IgM myeloma is usually reserved for patients with lytic bone lesions and predominant infiltration with CD138+ plasma cells in the bone marrow. Such patients are at greater risk of pathologic fractures than patients with Waldenström's macroglobulinemia.

The cause of macroglobulinemia is unknown. The disease is similar to myeloma in being slightly more common in men and occurring with increased incidence with age (median 64 years). There have been reports that the IgM in some patients with macroglobulinemia may have specificity for myelin-associated glycoprotein (MAG), a protein that has been associated with demyelinating disease of the peripheral nervous system and may be lost earlier and to a greater extent than the better known myelin basic protein in patients with multiple sclerosis. Sometimes patients with macroglobulinemia develop a peripheral neuropathy, and half of these patients are positive for anti-MAG antibody. The neuropathy may precede the appearance of the neoplasm. There is speculation that the whole process begins with a viral infection that may elicit an antibody response that cross-reacts with a normal tissue component.

Like myeloma, the disease involves the bone marrow, but unlike myeloma, it does not cause bone lesions or hypercalcemia. Bone marrow shows >10% infiltration with lymphoplasmacytic cells (surface IgM+, CD19+, CD20+, and CD22+, rarely CD5+, but CD10– and CD23–) with increase in number of mast cells. Like myeloma, M component is present in the serum in excess of 30 g/L (3 g/dL), but unlike myeloma, the size of the IgM paraprotein results in little renal excretion, and only ~20% of patients excrete light chains. Therefore, renal disease is not common. The light chain isotype is kappa in 80% of the cases. Patients present with weakness,

fatigue, and recurrent infections, similar to myeloma patients, but epistaxis, visual disturbances, and neurologic symptoms such as peripheral neuropathy, dizziness, headache, and transient paresis are much more common in macroglobulinemia. Physical examination reveals adenopathy and hepatosplenomegaly, and ophthalmoscopic examination may reveal vascular segmentation and dilation of the retinal veins characteristic of hyperviscosity states. Patients may have a normocytic, normochromic anemia, but rouleaux formation and a positive Coombs' test are much more common than in myeloma. Malignant lymphocytes are usually present in the peripheral blood. About 10% of macroglobulins are cryoglobulins. These are pure M components and are not the mixed cryoglobulins seen in rheumatoid arthritis and other autoimmune diseases. Mixed cryoglobulins are composed of IgM or IgA complexed with IgG, for which they are specific. In both cases, Raynaud's phenomenon and serious vascular symptoms precipitated by the cold may occur, but mixed cryoglobulins are not commonly associated with malignancy. Patients suspected of having a cryoglobulin based on history and physical examination should have their blood drawn into a warm syringe and delivered to the laboratory in a container of warm water to avoid errors in quantitating the cryoglobulin.

Treatment: Waldenström's Macroglobulinemia

Control of serious hyperviscosity symptoms such as an altered state of consciousness or paresis can be achieved acutely by plasmapheresis because 80% of the IgM paraprotein is intravascular. The median survival is ~50 months, similar to that of multiple myeloma. However, many individuals with Waldenström's macroglobulinemia have indolent disease that does not require therapy. Pretreatment parameters including older age, male sex, general symptoms, and cytopenias define a high-risk population. Fludarabine (25 mg/m^2 per day for 5 days every 4 weeks) or cladribine (0.1 mg/kg per day for 7 days every 4 weeks) are highly effective single agents. About 80% of patients respond to chemotherapy, and their median survival is >3 years. Rituximab (anti-CD20) can produce responses alone or combined with chemotherapy. As in multiple myeloma, the introduction of novel agents such as bortezomib, bendamustine, and lenalidomide have improved patient outcome.

POEMS SYNDROME

The features of this syndrome are *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *m*ultiple myeloma, and *s*kin changes (POEMS). Patients usually have a severe, progressive sensorimotor polyneuropathy associated with sclerotic bone lesions from myeloma. Polyneuropathy occurs in ~1.4% of myelomas, but the POEMS syndrome is only a rare subset of that group. Unlike typical myeloma, hepatomegaly and lymphadenopathy occur in about two-thirds of patients, and splenomegaly is seen in one-third. The lymphadenopathy frequently resembles Castleman's disease histologically, a condition that has been linked to IL-6 overproduction. The endocrine manifestations include amenorrhea in women and impotence and gynecomastia in men. Hyperprolactinemia due to loss of normal inhibitory control by the hypothalamus may be associated with other central nervous system manifestations such as papilledema and elevated cerebrospinal fluid pressure and protein. Type 2 diabetes mellitus occurs in about one-third of patients. Hypothyroidism and adrenal insufficiency are occasionally noted. Skin changes are diverse: hyperpigmentation, hypertrichosis, skin thickening, and digital clubbing. Other manifestations include peripheral edema, ascites, pleural effusions, fever, and thrombocytosis. Not all the components of POEMS syndrome may be present initially.

The pathogenesis of the disease is unclear, but high circulating levels of the proinflammatory cytokines IL-1, IL-6, VEGF, and TNF have been documented, and levels of the inhibitory cytokine TGF-β are lower than expected. Treatment of the myeloma may result in an improvement in the other disease manifestations.

Patients are often treated similarly to those with myeloma. Plasmapheresis does not appear to be of benefit in POEMS syndrome. Patients presenting with isolated sclerotic lesions may have resolution of neuropathic symptoms after local therapy for plasmacytoma with radiotherapy. Similar to multiple myeloma, novel agents as well as high-dose therapy with autologous stem cell transplant have been pursued in selected patients and have been associated with prolonged progression-free survival.

HEAVY CHAIN DISEASES

The heavy chain diseases are rare lymphoplasmacytic malignancies. Their clinical manifestations vary with the heavy chain isotype. Patients have absence of light chain and secrete a defective heavy chain that usually has an intact Fc fragment and a deletion in the Fd region. Gamma, alpha, and mu heavy chain diseases have been described, but no reports of delta or epsilon heavy chain diseases have appeared. Molecular biologic analysis of these tumors has revealed structural genetic defects that may account for the aberrant chain secreted.

Gamma Heavy Chain Disease (Franklin's Disease)

This disease affects individuals of widely different age groups and countries of origin. It is characterized by lymphadenopathy, fever, anemia, malaise, hepatosplenomegaly, and weakness. It is frequently associated with autoimmune diseases, especially rheumatoid arthritis. Its most distinctive symptom is palatal edema, resulting from involvement of nodes in Waldeyer's ring, and this may progress to produce respiratory compromise. The diagnosis depends on the demonstration of an anomalous serum M component [often <20 g/L (<2 g/dL)] that reacts with anti-IgG but not anti–light chain reagents. *The M component is typically present in both serum and urine*. Most of the paraproteins have been of the γ_1 subclass, but other subclasses have been seen. The patients may have thrombocytopenia, eosinophilia, and nondiagnostic bone marrow that may show increased numbers of lymphocytes or plasma cells that do not stain for light chain. Patients usually have a rapid downhill course and die of infection; however, some patients have survived 5 years with chemotherapy. Therapy is indicated when symptomatic and involves chemotherapeutic combinations used in low-grade lymphoma. Rituximab has also been reported to show efficacy.

Alpha Heavy Chain Disease (Seligmann's Disease)

This is the most common of the heavy chain diseases. It is closely related to a malignancy known as *Mediterranean lymphoma*, a disease that affects young persons in parts of the world where intestinal parasites are common, such as the Mediterranean, Asia, and South America. The disease is characterized by an infiltration of the lamina propria of the small intestine with lymphoplasmacytoid cells that secrete truncated alpha chains. Demonstrating alpha heavy chains is difficult because the alpha chains tend to polymerize and appear as a smear instead of a sharp peak on electrophoretic profiles. Despite the polymerization, hyperviscosity is not a common problem in alpha heavy chain disease. Without J chain–facilitated dimerization, viscosity does not increase dramatically. Light chains are absent from serum and urine. The patients present with chronic diarrhea, weight loss, and malabsorption and have extensive mesenteric and paraaortic adenopathy. Respiratory tract involvement occurs rarely. Patients may vary widely in their clinical course. Some may develop diffuse aggressive histologies of malignant lymphoma. Chemotherapy may produce long-term remissions. Rare patients appear to have responded to antibiotic therapy, raising the question of the etiologic role of antigenic stimulation, perhaps by some chronic intestinal infection. Chemotherapy plus antibiotics may be more effective than chemotherapy alone. Immunoproliferative small-intestinal disease (IPSID) is recognized as an infectious pathogen–associated human lymphoma that has association with *Campylobacter jejuni* . It involves mainly the proximal small intestine resulting in malabsorption, diarrhea, and abdominal pain. IPSID is associated with excessive plasma cell differentiation and produces truncated alpha heavy chain proteins lacking the light chains as well as the first constant domain. Early-stage IPSID responds to antibiotics (30–70% complete remission). Most untreated IPSID patients progress to lymphoplasmacytic and immunoblastic lymphoma. Patients not responding to antibiotic therapy are considered for treatment with combination chemotherapy used to treat low-grade lymphoma.

Mu Heavy Chain Disease

The secretion of isolated mu heavy chains into the serum appears to occur in a very rare subset of patients with chronic lymphocytic leukemia. The only features that may distinguish patients with mu heavy chain disease are the presence of vacuoles in the malignant lymphocytes and the excretion of kappa light chains in the urine. The diagnosis requires ultracentrifugation or gel filtration to confirm the nonreactivity of the paraprotein with the light chain reagents, because some intact macroglobulins fail to interact with these serums. The tumor cells seem to have a defect in the assembly of light and heavy chains, because they appear to contain both in their cytoplasm. There is no evidence that such patients should be treated differently from other patients with chronic lymphocytic leukemia (Chap. 110).

FURTHER READINGS

Dimopoulos M et al: Consensus recommendations for standard investigative workup: Report of the International Myeloma Workshop Consensus Panel 3. Blood 117:4701, 2011[PMID: 21292778]

Facon T et al: Frontline treatment in elderly patients with multiple myeloma. Semin Hematol 46:133, 2009[PMID: 19389497]

Harousseau JL, Moreau P: Autologous hematopoietic stem-cell transplantation for multiple myeloma. N Engl J Med 360:2645, 2009[PMID: 19535803]

Hideshima T et al: Understanding multiple myeloma pathogenesis in the bone marrow to identify new therapeutic targets. Nat Rev Cancer 7:585, 2007[PMID: 17646864]

The International Myeloma Working Group: Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: A report of the International Myeloma Working Group. Br J Haematol 121:749, 2003

Jagannath S et al: The current status and future of multiple myeloma in the clinic. Clin Lymphoma Myeloma Leuk 10:28, 2010[PMID: 20223727]

Kyle RA et al: Prevalence of monoclonal gammopathy of undetermined significance. N Engl J Med 354:1362, 2006[PMID: 16571879]

——— et al: Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma. N Engl J Med 356:2582, 2007

Munshi NC et al: Consensus recommendations for risk stratification in multiple myeloma: Report of the International Myeloma Workshop Consensus Panel 2. Blood 117:4696, 2011[PMID: 21292777]

Rajkumar SV et al: Front-line treatment in younger patients with multiple myeloma. Semin Hematol 46:118, 2009[PMID: 19389495]

Treon SP: How I treat Waldenström macroglobulinemia. Blood 114:2375, 2009[PMID: 19617573]

——— et al: Advances in the biology and treatment of Waldenström's macroglobulinemia: A report from the 5th International Workshop on Waldenström's Macroglobulinemia, Stockholm, Sweden. Clin Lymphoma Myeloma 9:10, 2009

Wahner-Roedler DL et al: Gamma-heavy chain disease: Review of 23 cases. Medicine 82:236, 2003[PMID: 12861101]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GENERAL PRINCIPLES

Amyloidosis is the term for diseases caused by the extracellular deposition of insoluble polymeric protein fibrils in tissues and organs. These diseases are a subset of a growing group of disorders attributed to misfolding of proteins. Among these are Alzheimer's disease and other neurodegenerative diseases, transmissible prion diseases, and genetic diseases caused by mutations that lead to misfolding, aggregation, and protein loss of function, such as certain of the cystic fibrosis mutations. Amyloid fibrils share a common β-pleated sheet structural conformation that confers unique staining properties. The term *amyloid* was coined by the pathologist Rudolf Virchow around 1854, who thought such deposits were cellulose-like under the microscope.

Amyloid diseases are defined by the biochemical nature of the protein in the fibril deposits and are classified according to whether they are systemic or localized, acquired or inherited, and by their clinical patterns (Table 112-1). The accepted nomenclature is *AX*, where *A* indicates amyloidosis and *X* represents the protein in the fibril. *AL* is amyloid composed of immunoglobulin light chains (LCs), and has been called *primary systemic amyloidosis*; it arises from a clonal B cell disorder and may be associated with myeloma or lymphoma. *AF* groups the *familial amyloidoses*, most commonly due to mutations in transthyretin, the transport protein for thyroid hormone and retinol-binding protein. *AA* amyloid is composed of the acute-phase reactant serum amyloid A protein and occurs in the setting of chronic inflammatory or infectious diseases and has been termed *secondary amyloidosis*. *Aβ$_2$M* is amyloid composed of β$_2$-microglobulin and occurs in individuals with end-stage renal disease (ESRD) of long duration. *Aβ* is the most common form of localized amyloidosis. Aβ is deposited in the brain in Alzheimer's disease and is derived from abnormal proteolytic processing of the amyloid precursor protein (APP).

Table 112-1 Amyloid Fibril Proteins and Their Clinical Syndromes

Term	Precursor	Clinical Syndrome	Clinical Involvement	
	Systemic Amyloidoses			
AL	Immunoglobulin light chain	Primary or myeloma associated[a]	Any	
AH	Immunoglobulin heavy chain	Primary or myeloma associated (rare)	Any	
AA	Serum amyloid A protein	Secondary; reactive[b]	Renal, any	
Aβ$_2$M	β$_2$-Microglobulin	Hemodialysis-associated	Synovial membrane, bone	
ATTR	Transthyretin	Familial (mutant) Senile systemic (wild type)	Cardiac, peripheral and autonomic nerves	
AApoAI	Apolipoprotein AI	Familial	Hepatic, renal	
AApoAII	Apolipoprotein AII	Familial	Renal	
AGel	Gelsolin	Familial	Corneas, cranial nerves, renal	
AFib	Fibrinogen Aα	Familial	Renal	
ALys	Lysozyme	Familial	Renal	
ALECT2	Leukocyte chemotactic factor 2	?	Renal	
	Localized Amyloidoses			
Aβ	Amyloid β protein	Alzheimer's disease;	Down syndrome	CNS
ACys	Cystatin C	Cerebral amyloid angiopathy	CNS, vascular	

APrP	Prion protein	Spongiform encephalopathies	CNS
AIAPP	Islet amyloid polypeptide (amylin)	Diabetes-associated	Pancreas
ACal	Calcitonin	Medullary carcinoma of the thyroid	Thyroid
AANF	Atrial natriuretic factor	Age-related	Cardiac atria
APro	Prolactin	Endocrinopathy	Pituitary

[a]Localized deposits can occur in skin, conjunctiva, urinary bladder, and tracheobronchial tree.

[b]Secondary to chronic inflammation or infection, or to a hereditary periodic fever syndrome, e.g., familial Mediterranean fever.

Diagnosis and treatment of the amyloidoses rest upon the pathologic diagnosis of amyloid deposits and immunohistochemical or biochemical identification of amyloid type (Fig. 112-1). In the systemic amyloidoses, the involved organs can be biopsied, but amyloid deposits may be found in any tissue of the body. Historically, blood vessels of the gingiva or rectal mucosa were examined, but the most easily accessible tissue, positive in more than 80% of patients with systemic amyloidosis, is fat. After local anesthesia, needle aspiration of fat from the abdominal wall can be expelled onto a slide and stained, avoiding even a minor surgical procedure. If this material is negative, biopsy of kidney, heart, liver, or gastrointestinal tract can be considered. The regular ß-sheet structure of amyloid deposits exhibits a unique green birefringence by polarized light microscopy when stained with Congo red dye; the 10-nm-diameter fibrils can be seen directly by electron microscopy of paraformaldehyde-fixed tissue. Once amyloid is found, the protein type must be determined, usually by immunohistochemistry, immunoelectron microscopy, or by extraction and biochemical analysis by mass spectrometry or other technique. Careful evaluation of the patient's history, physical findings, and clinical presentation, including age and ethnic origin, organ system involvement, underlying diseases, and family history, can provide helpful clues to the type of amyloid.

Figure 112-1

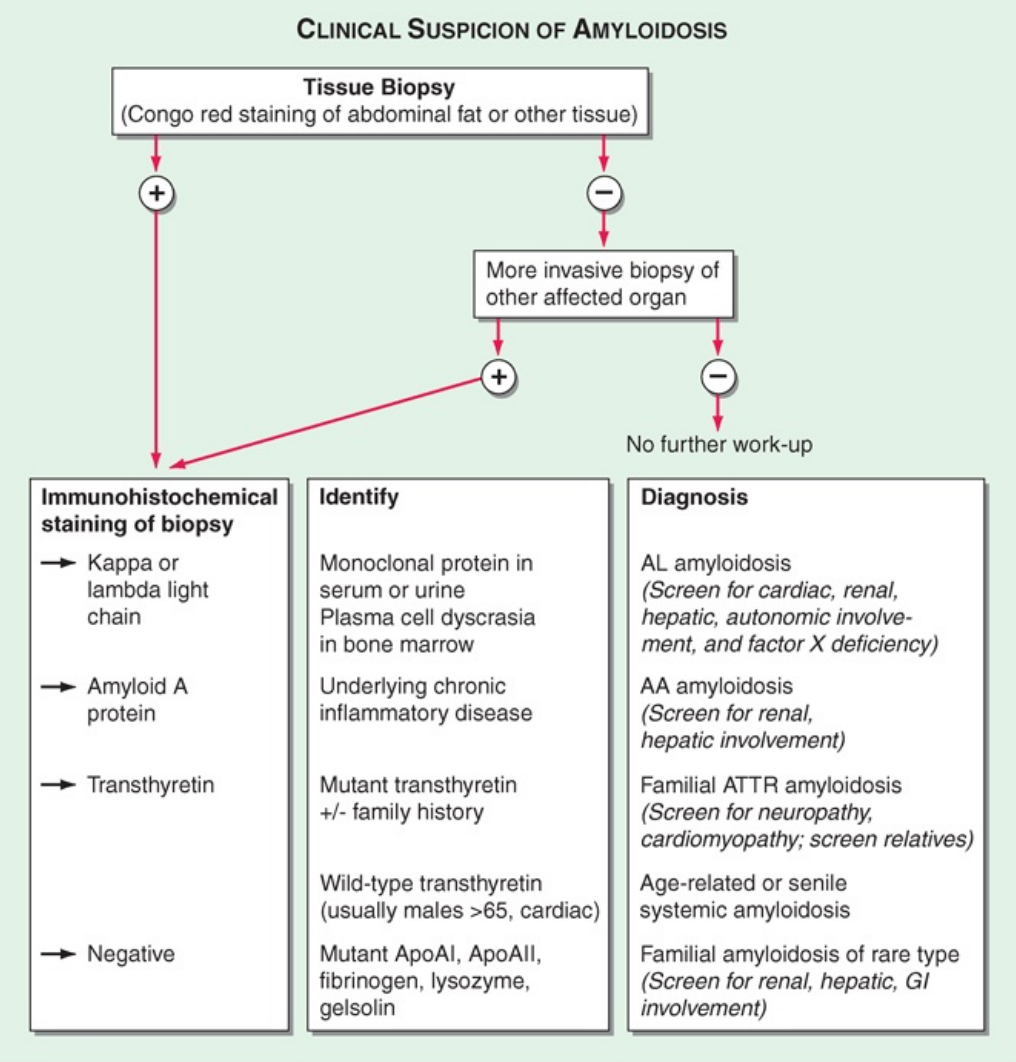

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the diagnosis of amyloidosis and determination of type: Clinical suspicion: unexplained nephropathy, cardiomyopathy, neuropathy, enteropathy, arthropathy, and macroglossia. ApoAI, apolipoprotein AI; ApoAII, apolipoprotein AII; GI, gastrointestinal.

The mechanisms of fibril formation and tissue toxicity remain controversial. Factors that contribute to fibrillogenesis include variant or unstable protein structure, extensive ß-sheet conformation of the precursor protein, proteolytic processing of the precursor protein, association with components of the serum or extracellular matrix (e.g., amyloid P-component, apolipoprotein E, or glycosaminoglycans), and local physical properties, including pH of the tissue. Monomeric proteins appear to go through an oligomeric aggregation step and then form higher order polymers. Once the polymers reach a critical size, they become insoluble and deposit in extracellular tissue sites as fibrils. These large macromolecular deposits interfere with organ function and, due to cellular uptake of oligomeric amyloid precursors, may be toxic to target cells.

The clinical syndromes of the amyloidoses are associated with relatively nonspecific alterations in routine laboratory tests. Blood counts are usually normal, although the erythrocyte sedimentation rate is frequently elevated. Patients with renal involvement will usually have proteinuria, which can be as much as 30 g/d, producing hypoalbuminemia that can be profound. Patients with cardiac involvement will often have elevation of brain natriuretic peptide (BNP), pro-BNP, and troponin. These can be useful for monitoring disease activity and have been proposed as prognostic factors; they can be falsely elevated in the presence of renal insufficiency. Patients with liver involvement, even when it is advanced, usually develop cholestasis with an elevated alkaline phosphatase but minimal elevation of the transaminases and preservation of synthetic function. In AL amyloidosis, endocrinopathies can occur, with laboratory testing demonstrating hypothyroidism, hypoadrenalism, or even hypopituitarism. None of these findings are specific for amyloidosis. Thus, a diagnosis of amyloidosis rests upon a tissue biopsy that, after Congo red staining, shows "apple-green" birefringence on polarization microscopy.

AL AMYLOIDOSIS

Etiology and Incidence

AL amyloidosis is most frequently caused by a clonal expansion of plasma cells in the bone marrow that secrete a monoclonal immunoglobulin LC that deposits as amyloid fibrils in tissues. It may be purely serendipitous whether the clonal plasma cells produce a LC that misfolds and leads to AL amyloidosis, or folds properly, allowing the cells to inexorably expand over time and develop into multiple myeloma (Chap. 111). It is also possible that the two processes have diverse molecular etiologies. AL amyloidosis can occur with multiple myeloma or other B lymphoproliferative diseases, including non-Hodgkin's lymphoma (Chap. 110) and Waldenström's macroglobulinemia (Chap. 111). AL amyloidosis is the most common type of systemic amyloidosis in North America. Its incidence has been estimated at 4.5 per 100,000; however, ascertainment continues to be inadequate, and the true incidence may be much higher. AL amyloidosis, like other plasma cell diseases, usually occurs after age 40 and is often rapidly progressive and fatal if untreated.

Pathology and Clinical Features of AL Amyloidosis

Amyloid deposits are usually widespread in AL amyloidosis and can be present in the interstitium of any organ outside of the central nervous system. The amyloid fibril deposits are composed of intact 23-kDa monoclonal Ig LCs or smaller fragments, 11–18 kDa in size, representing the variable (V) region alone, or the V region and a portion of the constant (C) region. Although all kappa and lambda LC subtypes have been identified in AL amyloid fibrils, lambda subtypes predominate. The lambda 6 subtype appears to have unique structural properties that predispose it to fibril formation, often in the kidney.

AL amyloidosis is usually a rapidly progressive disease that presents with a pleiotropic set of clinical syndromes, recognition of which is key to initiating appropriate workup. Nonspecific symptoms of fatigue and weight loss are common; however, the diagnosis is rarely considered until symptoms referable to a specific organ develop. The kidneys are the most frequently affected organ, in 70–70% of patients. Renal amyloidosis is usually manifested as proteinuria, often in the nephrotic range and associated with significant hypoalbuminemia, secondary hypercholesterolemia, and edema or anasarca. In some patients, tubular rather than glomerular deposition of amyloid can produce azotemia without significant proteinuria. The heart is the second most commonly affected organ, in 50–50% of patients, and the leading cause of mortality. Early on, the electrocardiogram may show low voltage in the limb leads, with a pseudo-infarct pattern. Eventually, the echocardiogram will display concentrically thickened ventricles and diastolic dysfunction, leading to a restrictive cardiomyopathy; systolic function is preserved until late in the disease. A "sparkly" appearance is usually not seen using modern high-resolution echocardiography equipment. Cardiac MRI can show an increased wall thickness and also a characteristic subendocardial enhancement with gadolinium. Nervous system symptoms include a peripheral sensory neuropathy and/or autonomic dysfunction with gastrointestinal motility disturbances (early satiety, diarrhea, constipation) and orthostatic hypotension. Macroglossia, with an enlarged, indented, or immobile tongue, is pathognomonic of AL amyloidosis but is seen only in ~10% of patients. Liver involvement causes cholestasis and hepatomegaly. The spleen is frequently involved, and there may be functional hyposplenism in the absence of significant splenomegaly. Many patients have "easy bruising" due to amyloid deposits in capillaries or to deficiency of clotting factor X, which can bind to amyloid fibrils; cutaneous ecchymoses appear, particularly around the eyes, giving the "raccoon-eye" sign. Other findings include nail dystrophy, alopecia, and amyloid arthropathy with thickening of synovial membranes in wrists and shoulders (Fig. 112-2). The presence of a multisystem illness or general fatigue along with any of these clinical syndromes should prompt a workup for amyloidosis.

Figure 112-2

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical signs of AL amyloidosis. ***A.*** Macroglossia. ***B.*** Periorbital ecchymoses. ***C.*** Fingernail dystrophy.

Diagnosis

Identification of the underlying B lymphoproliferative process and clonal LC is key to the diagnosis of

AL amyloidosis. The serum protein electrophoresis (SPEP) and urine protein electrophoresis (UPEP) are NOT useful screening tests if AL amyloidosis is suspected because the clonal LC or whole immunoglobulin, unlike in multiple myeloma, is often not present in sufficient quantity in the serum to produce a monoclonal "M-spike" or in the urine to cause LC (Bence Jones) proteinuria. However, more than 90% of patients have a serum or urine monoclonal LC or whole immunoglobulin that can be detected by immunofixation electrophoresis of serum (SIFE) or urine (UIFE) (Fig. 112-3A). Assaying for free immunoglobulin LCs circulating in the serum unbound to heavy chains using commercially available nephelometric (FreeLite©) assay demonstrates an elevation and abnormal free kappa:lambda ratio in more than 75% of patients. Examining the ratio as well as the absolute amount is essential, because in renal insufficiency LC clearance is reduced, and both types of LCs will be elevated. In addition, an increased percentage of plasma cells in the bone marrow, typically 5–5% of nucleated cells, is noted in about 90% of patients. Kappa or lambda clonality can be demonstrated by flow cytometry, immunohistochemical staining, or by in situ hybridization for LC mRNA (Fig. 112-3B).

Figure 112-3

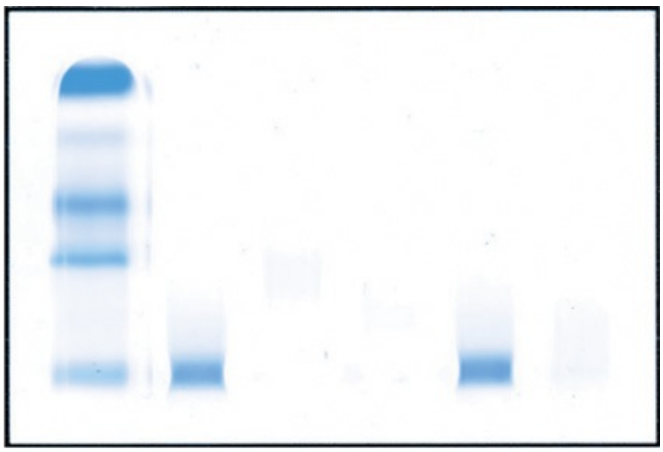

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

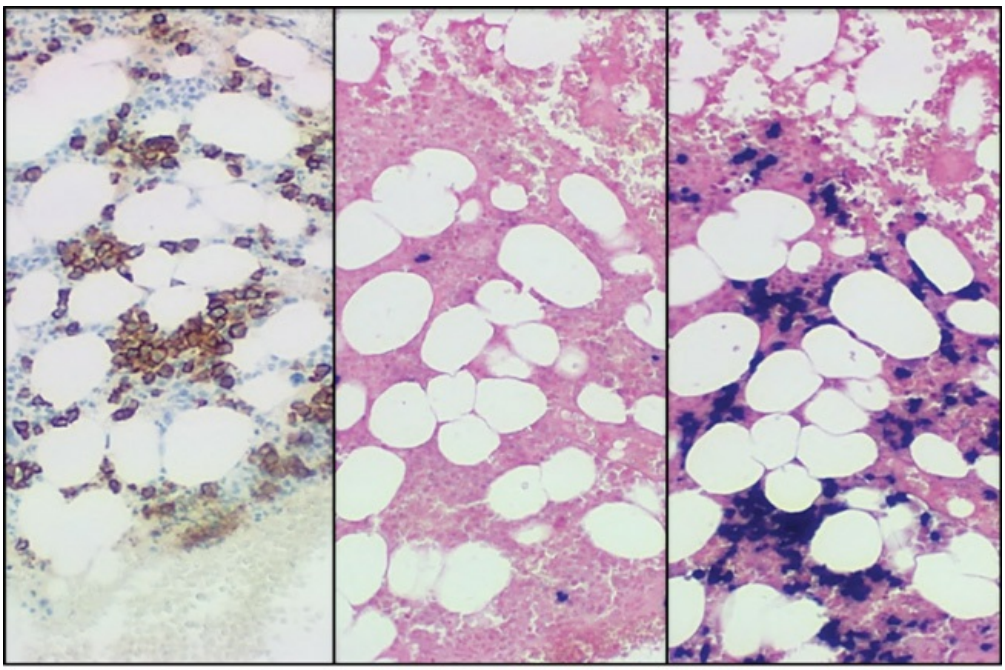

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Laboratory features of AL amyloidosis. ***A.*** Serum immunofixation electrophoresis reveals an IgG_{κ} monoclonal protein in this example; the serum protein electrophoresis is often normal. ***B.*** Bone marrow biopsy sections from another patient, stained with antibody to CD138 (syndecan, highly expressed on plasma cells) by immunohistochemistry (left panel). The middle and right panels are stained using in situ hybridization with fluorescein-tagged probes (Ventana Medical Systems) binding

stained using in situ hybridization with fluorescein-tagged probes (Ventana Medical Systems) binding to κ and λ mRNA respectively in plasma cells. *(Photomicrograph courtesy of C. O'Hara; with permission.)*

A monoclonal serum protein by itself is not diagnostic of amyloidosis, since monoclonal gammopathy of uncertain significance (MGUS) is common in older patients (Chap. 111). However, when MGUS is present in patients with biopsy-proven amyloidosis, the AL type should be strongly suspected. Similarly, patients thought to have "smoldering myeloma" because of modest elevation of bone marrow plasma cells should be screened for AL amyloidosis if they have evidence of organ dysfunction. Accurate typing is essential for appropriate treatment. Immunohistochemical staining of the amyloid deposits is useful if they bind one light chain antibody in preference to the other; some AL deposits bind many antisera nonspecifically. Immunoelectron microscopy is more reliable and mass-spectrometry-based microsequencing of small amounts of protein extracted from fibril deposits can also be done. In ambiguous cases, other forms of amyloidosis should be thoroughly excluded with appropriate genetic and other testing.

Treatment: AL Amyloidosis

Extensive multisystem involvement typifies AL amyloidosis, and median survival with no treatment is usually only about 1–2 years from the time of diagnosis. Current therapies target the clonal bone marrow plasma cells using approaches employed for multiple myeloma. Treatment with cyclic oral melphalan and prednisone can decrease the plasma cell burden but produces complete hematologicremission in only a few percent of patients and minimal organ responses and improvement in survival (median 2 years), and it is no longer widely used. The substitution of dexamethasone for prednisone produces a higher response rate and more durable remissions, although dexamethasone is not always well tolerated by patients with significant edema or cardiac disease. High-dose intravenous melphalan followed by autologous stem cell transplantation (HDM/SCT) produces complete hematologic responses in about 40% of treated patients, as measured by complete loss (CR) of clonal plasma cells in the bone marrow and disappearance of the monoclonal LC by IFE and assay for free LCs. Hematologic responses can be followed in the subsequent 6–12 months by improvement in organ function and quality of life. The CRs after HDM/SCT appear to be more durable than those seen in multiple myeloma, with remissions continuing in some patients beyond 15 years without additional treatment. Unfortunately, only about half of AL amyloidosis patients are eligible for such aggressive treatment, and even at specialized treatment centers, peritransplant mortality is higher than for other hematologic diseases because of impaired organ function. Amyloid cardiomyopathy, poor nutritional status, impaired performance status, and multiple-organ disease contribute to excess morbidity and mortality. The bleeding diathesis due to adsorption of clotting factor X to amyloid fibrils also confers high mortality during myelosuppressive therapy; however, this syndrome occurs in only a few percent of patients. The single randomized multicenter trial comparing oral melphalan and dexamethasone to HDM/SCT to date failed to show a benefit to dose-intensive treatment, although the transplant-related mortality in this study was very high.

For patients with impaired cardiac function or arrhythmias due to amyloid involvement of the myocardium, median survival is only about 6 months without treatment, and stem cell mobilization and high-dose chemotherapy are dangerous. In these patients, cardiac transplantation can be performed, followed by treatment with HDM/SCT to prevent amyloid deposition in the transplanted heart or other organs.

Recently, novel agents have been investigated for treatment of plasma cell diseases. The immunomodulators thalidomide and lenalidomide have activity; lenalidomide is well tolerated in doses lower than those used for myeloma and, in combination with dexamethasone, produces complete hematologic remissions and improvement in organ function. The proteasome inhibitor bortezomib has also been found to be effective in single- and multicenter trials. Combination therapy trials are now under development, and studies are examining the as yet unproven role of induction and maintenance treatment. Clinical trials are essential for improving therapy for this rare disease.

Supportive care is important for patients with any type of amyloidosis. For nephrotic syndrome, diuretics and supportive stockings can ameliorate edema; angiotensin-converting enzyme inhibitors should be used with caution and have not been shown to slow renal disease progression. Congestive heart failure due to amyloid cardiomyopathy is also best treated with diuretics; it is important to note that digitalis, calcium channel blockers, and beta blockers are relatively contraindicated as they can interact with amyloid fibrils and produce heart block and worsening heart failure. Amiodarone has been used for atrial and ventricular arrhythmias. Automatic implantable defibrillators have reduced effectiveness due to the thickened myocardium, but they can benefit some patients. Atrial ablation is an effective approach for atrial fibrillation. For conduction abnormalities, ventricular pacing may be indicated. Atrial contractile dysfunction is common in amyloid cardiomyopathy and is an indication for anticoagulation even in the absence of atrial fibrillation. Autonomic neuropathy can be treated with α agonists such as midodrine to support the blood pressure; gastrointestinal dysfunction may respond to motility or bulk agents. Nutritional supplementation, either orally or parenterally, is also important.

In localized AL, amyloid deposits can be produced by clonal plasma cells infiltrating local sites in the airways, bladder, skin, or lymph nodes (Table 112-1). Deposits may respond to surgical intervention or

radiation therapy; systemic treatment is generally not appropriate. Patients should be referred to a center familiar with management of these rare manifestations of amyloidosis.

AA AMYLOIDOSIS

Etiology and Incidence

AA amyloidosis can occur in association with almost any chronic inflammatory state [e.g., rheumatoid arthritis, inflammatory bowel disease, familial Mediterranean fever (Chap. 330) or other periodic fever syndromes] or chronic infections such as tuberculosis or subacute bacterial endocarditis. In the United States and Europe, AA amyloidosis has become less common, occurring in <2% of patients with these diseases, presumably because of advances in anti-inflammatory and antimicrobial therapies. It has also been described in association with Castleman's disease, and patients with AA amyloidosis should have CT scanning to look for such tumors, as well as serologic and microbiologic studies. AA amyloidosis can also be seen without any identifiable underlying disease. AA is the only type of systemic amyloidosis that occurs in children.

Pathology and Clinical Features

Deposits are more limited in AA amyloidosis than in AL amyloidosis; they usually begin in the kidneys. Hepatomegaly, splenomegaly, and autonomic neuropathy can also occur as the disease progresses; cardiomyopathy occurs albeit rarely. However, the symptoms and signs cannot be reliably distinguished from those of AL amyloidosis. AA amyloid fibrils are usually composed of an 8-kDa, 76-amino-acid *N*-terminal portion of a 12-kDa precursor protein, serum amyloid A (SAA). SAA is an acute-phase apoprotein synthesized in the liver and transported by high-density lipoprotein, HDL3, in the plasma. Several years of an underlying inflammatory disease causing chronic elevation of SAA usually precedes fibril formation, although infections can lead to AA deposition more rapidly.

Treatment: AA Amyloidosis

The primary therapy in AA amyloidosis is treatment of the underlying inflammatory or infectious disease. Treatment that suppresses or eliminates the inflammation or infection also decreases the SAA protein concentration. For familial Mediterranean fever, colchicine in a dose of 1.2–1.8 mg/d is the appropriate treatment. Colchicine has not been helpful for AA amyloidosis of other causes or for other amyloidoses. TNF and IL-1 antagonists can also be effective in syndromes related to cytokine elevation. For this disease, there is also a fibril-specific agent. Eprodisate was designed to interfere with the interaction of AA amyloid protein with glycosaminoglycans in tissues and prevent or disrupt fibril formation. This drug is well tolerated and delays progression of AA renal disease, regardless of the underlying inflammatory process. Eprodisate is awaiting FDA approval.

AF AMYLOIDOSIS

The familial amyloidoses are autosomal dominant diseases in which a variant plasma protein forms amyloid deposits, beginning in midlife. These diseases are rare, with an estimated incidence of <1 per 100,000 in the United States, although there are isolated areas of Portugal, Sweden, and Japan where founder effects have led to a much higher incidence of the disease. The most common form of AF is caused by mutation of the abundant plasma protein transthyretin (TTR, also known as *prealbumin*). More than 100 TTR mutations are known, and most are associated with ATTR amyloidosis. One variant, V122I, has a carrier frequency that may be as high as 4% in the African-American population and is associated with late-onset cardiac amyloidosis. The actual incidence and penetrance of disease in the African-American population is the subject of ongoing research, but it would be wise to consider this in the differential diagnosis of African-American patients who present with concentric cardiac hypertrophy and evidence of diastolic dysfunction, particularly in the absence of a history of hypertension. Even wild-type TTR can form fibrils, leading to so-called senile systemic amyloidosis (SSA) in older patients. It can be found in up to 25% of autopsies in patients older than age 80 years, and it can produce a clinical syndrome of amyloid cardiomyopathy that is similar to that occurring in younger patients carrying a mutant TTR. Other familial amyloidoses, caused by variant apolipoproteins AI or AII, gelsolin, fibrinogen Aα, or lysozyme, are reported in only a few families worldwide. New amyloidogenic serum proteins continue to be identified periodically, including recently the leukocyte chemotactic factor LECT2.

In ATTR and in other forms of familial amyloidosis, the variant structure of the precursor protein is the key factor in fibril formation. The role of aging is intriguing, since patients born with the variant proteins do not have clinically apparent disease until middle age, despite the lifelong presence of the abnormal protein. Further evidence of an age-related "trigger" is the occurrence of SSA in the elderly, caused by the deposition of fibrils derived from normal TTR.

Clinical Features and Diagnosis

AF amyloidosis has a variable presentation but is usually consistent within affected kindreds with the same mutant protein. A family history makes AF more likely, but many patients present sporadically with new mutations. ATTR usually presents as a syndrome of familial amyloidotic polyneuropathy or familial amyloidotic cardiomyopathy. Peripheral neuropathy usually begins as a lower-extremity sensory and

motor neuropathy and progresses to the upper extremities. Autonomic neuropathy is manifest by gastrointestinal symptoms of diarrhea with weight loss and orthostatic hypotension. Patients with TTR V30M, the most common mutation, have normal echocardiograms but may have conduction defects and require a pacemaker. Patients with TTR T60A and several other mutations have myocardial thickening similar to that caused by AL amyloidosis, although heart failure is less common and the prognosis is better. Vitreous opacities caused by amyloid deposits are pathognomonic of ATTR amyloidosis.

Typical syndromes associated with other forms of AF included renal amyloidosis with mutant fibrinogen, lysozyme, or apolipoproteins, or hepatic amyloidosis with apolipoprotein AI, and amyloidosis of cranial nerves and cornea with gelsolin.

Patients with AF amyloidosis can present with clinical syndromes that mimic those of patients with AL, and AF carriers can develop AL, or conversely, AF patients can develop a MGUS. Thus, it is important to screen for both plasma cell disorders and for mutations in some patients with amyloidosis. Variant TTR proteins can usually be detected by isoelectric focusing, but DNA sequencing is now standard for diagnosis of ATTR and the other AF mutations.

Treatment: ATTR Amyloidosis

Without intervention, survival after ATTR disease onset is 5–15 years. Orthotopic liver transplantation removes the major source of variant TTR production and replaces it with a source of normal TTR; it also arrests disease progression and leads to improvement in autonomic and peripheral neuropathy in some patients. Cardiomyopathy often does not improve, and in some patients it can worsen after liver transplantation, perhaps due to deposition of wild-type TTR as seen in SSA. Compounds have been identified that stabilize TTR in a nonpathogenic tetrameric conformation in vitro and are undergoing clinical testing in multicenter trials.

Aβ_2M AMYLOIDOSIS

Aβ_2M amyloidosis is composed of β_2-microglobulin, the invariant chain of class I human leukocyte antigens, and produces rheumatologic manifestations in patients on long-term hemodialysis. β_2-Microglobulin is excreted by the kidney, and levels become elevated in ESRD. The molecular mass of β_2M is 11.8 kDa, above the cutoff of some dialysis membranes. The incidence of this disease appears to be declining with newer high-flow dialysis techniques.

Aβ_2M amyloidosis usually presents with carpal tunnel syndrome, persistent joint effusions, spondyloarthropathy, or cystic bone lesions. Carpal tunnel syndrome is often the first symptom of disease. In the past, persistent joint effusions accompanied by mild discomfort were seen in up to 50% of patients on dialysis for more than 12 years. Involvement is bilateral, and large joints (shoulders, knees, wrists, and hips) are more frequently affected. The synovial fluid is noninflammatory, and β_2M amyloid can be found if the sediment is stained with Congo red. Although less common, visceral β_2M amyloid deposits do occasionally occur in the gastrointestinal tract, heart, tendons, and subcutaneous tissues of the buttocks. There is no specific therapy for Aβ_2M amyloidosis, but cessation of dialysis after renal allografting may lead to symptomatic improvement.

SUMMARY

A diagnosis of amyloidosis should be considered in patients with unexplained nephropathy, cardiomyopathy (particularly with diastolic dysfunction), neuropathy (either peripheral or autonomic), enteropathy, or the pathognomonic soft tissue findings of macroglossia or periorbital ecchymoses. Pathologic identification of amyloid fibrils can be made using Congo red staining of aspirated abdominal fat or of an involved organ biopsy specimen. Accurate typing using a combination of immunologic, biochemical, and genetic testing is essential to choosing the appropriate therapy (see algorithm for workup, Fig. 112-1). Tertiary referral centers can provide specialized diagnostic techniques and access to clinical trials for patients with these rare diseases.

FURTHER READINGS

Benson MD et al: Leukocyte chemotactic factor 2: A novel renal amyloid protein. Kidney Int 74:218, 2008[PMID: 18449172]

Connors L et al: Cardiac amyloidosis in African Americans: Comparison of clinical and laboratory features of transthyretin V122I amyloidosis and immunoglobulin light chain amyloidosis. Am Heart J 158:607, 2009

Dember L et al: Eprodisate for the treatment of renal disease in AA amyloidosis. N Engl J Med 356:2349, 2007

Dey BR et al: Cardiac transplantation followed by dose-intensive melphalan and autologous stem-cell

transplantation for light chain amyloidosis and heart failure. Transplantation 90:905, 2010[PMID: 20733534]

Merlini G, Bellotti V: Molecular mechanisms of amyloidosis. N Engl J Med 349:583, 2003

Sanchorawala V et al: Long-term outcome of patients with AL amyloidosis treated with high-dose melphalan and stem-cell transplantation. Blood 110:3561, 2007

Skinner M et al: High-dose melphalan and autologous stem-cell transplantation in patients with AL amyloidosis: An 8-year study. Ann Intern Med 140:85, 2004[PMID: 14734330]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BLOOD GROUP ANTIGENS AND ANTIBODIES

The study of red blood cell (RBC) antigens and antibodies forms the foundation of transfusion medicine. Serologic studies initially characterized these antigens, but now the molecular composition and structure of many are known. Antigens, either carbohydrate or protein, are assigned to a blood group system based on the structure and similarity of the determinant epitopes. Other cellular blood elements and plasma proteins are also antigenic and can result in *alloimmunization*, the production of antibodies directed against the blood group antigens of another individual. These antibodies are called *alloantibodies*.

Antibodies directed against RBC antigens may result from "natural" exposure, particularly to carbohydrates that mimic some blood group antigens. Those antibodies that occur via natural stimuli are usually produced by a T cell—independent response (thus, generating no memory) and are IgM isotype. *Autoantibodies* (antibodies against autologous blood group antigens) arise spontaneously or as the result of infectious sequelae (e.g., from *Mycoplasma pneumoniae*) and are also often IgM. These antibodies are often clinically insignificant due to their low affinity for antigen at body temperature. However, IgM antibodies can activate the complement cascade and result in hemolysis. Antibodies that result from allogeneic exposure, such as transfusion or pregnancy, are usually IgG. IgG antibodies commonly bind to antigen at warmer temperatures and may hemolyze RBCs. Unlike IgM antibodies, IgG antibodies can cross the placenta and bind fetal erythrocytes bearing the corresponding antigen, resulting in hemolytic disease of the newborn, or *hydrops fetalis*.

Alloimmunization to leukocytes, platelets, and plasma proteins may also result in transfusion complications such as fevers and urticaria but generally does not cause hemolysis. Assay for these other alloantibodies is not routinely performed; however, they may be detected using special assays.

ABO Antigens and Antibodies

The first blood group antigen system, recognized in 1900, was ABO, the most important in transfusion medicine. The major blood groups of this system are A, B, AB, and O. O type RBCs lack A or B antigens. These antigens are carbohydrates attached to a precursor backbone, may be found on the cellular membrane either as glycosphingolipids or glycoproteins, and are secreted into plasma and body fluids as glycoproteins. H substance is the immediate precursor on which the A and B antigens are added. This H substance is formed by the addition of fucose to the glycolipid or glycoprotein backbone. The subsequent addition of *N*-acetylgalactosamine creates the A antigen, while the addition of galactose produces the B antigen.

The genes that determine the A and B phenotypes are found on chromosome 9p and are expressed in a Mendelian codominant manner. The gene products are glycosyl transferases, which confer the enzymatic capability of attaching the specific antigenic carbohydrate. Individuals who lack the "A" and "B" transferases are phenotypically type "O," while those who inherit both transferases are type "AB." Rare individuals lack the H gene, which codes for fucose transferase, and cannot form H substance. These individuals are homozygous for the silent h allele (hh) and have Bombay phenotype (O_h).

The ABO blood group system is important because essentially all individuals produce antibodies to the ABH carbohydrate antigen that they lack. The naturally occurring anti-A and anti-B antibodies are termed *isoagglutinins*. Thus, type A individuals produce anti-B, while type B individuals make anti-A. Neither isoagglutinin is found in type AB individuals, while type O individuals produce both anti-A and anti-B. Thus, persons with type AB are "universal recipients" because they do not have antibodies against any ABO phenotype, while persons with type O blood can donate to essentially all recipients because their cells are not recognized by any ABO isoagglutinins. The rare individuals with Bombay phenotype produce antibodies to H substance (which is present on all red cells except those of hh phenotype) as well as to both A and B antigens and are therefore compatible only with other hh donors.

In most people, A and B antigens are secreted by the cells and are present in the circulation. Nonsecretors are susceptible to a variety of infections (e.g., *Candida albicans*, *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Haemophilus influenzae*) as many organisms may bind to polysaccharides on cells. Soluble blood group antigens may block this binding.

Rh System

The Rh system is the second most important blood group system in pretransfusion testing. The Rh

antigens are found on a 30- to 32-kDa RBC membrane protein that has no defined function. Although >40 different antigens in the Rh system have been described, five determinants account for the vast majority of phenotypes. The presence of the D antigen confers Rh "positivity," while persons who lack the D antigen are Rh negative. Two allelic antigen pairs, E/e and C/c, are also found on the Rh protein. The three Rh genes, E/e, D, and C/c, are arranged in tandem on chromosome 1 and inherited as a haplotype, i.e., cDE or Cde. Two haplotypes can result in the phenotypic expression of two to five Rh antigens.

The D antigen is a potent alloantigen. About 15% of individuals lack this antigen. Exposure of these Rh-negative people to even small amounts of Rh-positive cells, by either transfusion or pregnancy, can result in the production of anti-D alloantibody.

Other Blood Group Systems and Alloantibodies

More than 100 blood group systems are recognized, composed of more than 500 antigens. The presence or absence of certain antigens has been associated with various diseases and anomalies; antigens also act as receptors for infectious agents. Alloantibodies of importance in routine clinical practice are listed in Table 113–1.

Table 113–1 RBC Blood Group Systems and Alloantigens

Blood Group System	Antigen	Alloantibody	Clinical Significance
Rh (D, C/c, E/e)	RBC protein	IgG	HTR, HDN
Lewis (Le^a, *Leactivity="italic">*^{b})	Oligosaccharide	IgM/IgG	Rare HTR
Kell (K/k)	RBC protein	IgG	HTR, HDN
Duffy (Fy^a/Fy^b)	RBC protein	IgG	HTR, HDN
Kidd (Jk^a/Jk^b)	RBC protein	IgG	HTR (often delayed), HDN (mild)
I/i	Carbohydrate	IgM	None
MNSsU	RBC protein	IgM/IgG	Anti-M rare HDN, anti-S, -s, and -U HDN, HTR

Abbreviation: RBC, red blood cell; HDN, hemolytic disease of the newborn; HTR, hemolytic transfusion reaction.

Antibodies to *Lewis system* carbohydrate antigens are the most common cause of incompatibility during pretransfusion screening. The Lewis gene product is a fucosyl transferase and maps to chromosome 19. The antigen is not an integral membrane structure but is adsorbed to the RBC membrane from the plasma. Antibodies to Lewis antigens are usually IgM and cannot cross the placenta. Lewis antigens may be adsorbed onto tumor cells and may be targets of therapy.

I system antigens are also oligosaccharides related to H, A, B, and Le. I and i are not allelic pairs but are carbohydrate antigens that differ only in the extent of branching. The i antigen is an unbranched chain that is converted by the I gene product, a glycosyltransferase, into a branched chain. The branching process affects all the ABH antigens, which become progressively more branched in the first 2 years of life. Some patients with cold agglutinin disease or lymphomas can produce anti-I autoantibodies that cause RBC destruction. Occasional patients with mononucleosis or *Mycoplasma* pneumonia may develop cold agglutinins of either anti-I or anti-i specificity. Most adults lack i expression; thus, finding a donor for patients with anti-i is not difficult. Even though most adults express I antigen, binding is generally low at body temperature. Thus, administration of warm blood prevents isoagglutination.

The *P system* is another group of carbohydrate antigens controlled by specific glycosyltransferases. Its clinical significance is in rare cases of syphilis and viral infection that lead to paroxysmal cold hemoglobinuria. In these cases, an unusual autoantibody to P is produced that binds to RBCs in the cold and fixes complement upon warming. Antibodies with these biphasic properties are called *Donath-Landsteiner antibodies*. The P antigen is the cellular receptor of parvovirus B19 and also may be a receptor for *Escherichia coli* binding to urothelial cells.

The *MNSsU system* is regulated by genes on chromosome 4. M and N are determinants on glycophorin A, an RBC membrane protein, and S and s are determinants on glycophorin B. Anti-S and anti-s IgG antibodies may develop after pregnancy or transfusion and lead to hemolysis. Anti-U antibodies are rare but problematic; virtually every donor is incompatible because nearly all persons express U.

The *Kell* protein is very large (720 amino acids), and its secondary structure contains many different antigenic epitopes. The immunogenicity of Kell is third behind the ABO and Rh systems. The absence

of the Kell precursor protein (controlled by a gene on X) is associated with acanthocytosis, shortened RBC survival, and a progressive form of muscular dystrophy that includes cardiac defects. This rare condition is called the *McLeod phenotype*. The K_X gene is linked to the 91-kDa component of the NADPH-oxidase on the X chromosome, deletion or mutation of which accounts for about 60% of cases of chronic granulomatous disease.

The *Duffy* antigens are codominant alleles, Fy^a and Fy^b, that also serve as receptors for *Plasmodium vivax*. More than 70% of persons in malaria-endemic areas lack these antigens, probably from selective influences of the infection on the population.

The *Kidd* antigens, Jk^a and Jk^b, may elicit antibodies transiently. A delayed hemolytic transfusion reaction that occurs with blood tested as compatible is often related to delayed appearance of anti-Jk^a.

PRETRANSFUSION TESTING

Pretransfusion testing of a potential recipient consists of the "type and screen." The "forward type" determines the ABO and Rh phenotype of the recipient's RBC by using antisera directed against the A, B, and D antigens. The "reverse type" detects isoagglutinins in the patient's serum and should correlate with the ABO phenotype, or forward type.

The alloantibody screen identifies antibodies directed against other RBC antigens. The alloantibody screen is performed by mixing patient serum with type O RBCs that contain the major antigens of most blood group systems and whose extended phenotype is known. The specificity of the alloantibody is identified by correlating the presence or absence of antigen with the results of the agglutination.

Cross-matching is ordered when there is a high probability that the patient will require a packed RBC (PRBC) transfusion. Blood selected for cross-matching must be ABO compatible and lack antigens for which the patient has alloantibodies. Nonreactive cross-matching confirms the absence of any major incompatibility and reserves that unit for the patient.

In the case of Rh-negative patients, every attempt must be made to provide Rh-negative blood components to prevent alloimmunization to the D antigen. In an emergency, Rh-positive blood can be safely transfused to an Rh-negative patient who lacks anti-D; however, the recipient is likely to become alloimmunized and produce anti-D. Rh-negative women of childbearing age who are transfused with products containing Rh-positive RBCs should receive passive immunization with anti-D (RhoGam or WinRho) to reduce or prevent sensitization.

BLOOD COMPONENTS

Blood products intended for transfusion are routinely collected as whole blood (450 mL) in various anticoagulants. Most donated blood is processed into components: PRBCs, platelets, and fresh-frozen plasma (FFP) or cryoprecipitate (Table 113–2). Whole blood is first separated into PRBCs and platelet-rich plasma by slow centrifugation. The platelet-rich plasma is then centrifuged at high speed to yield one unit of random donor (RD) platelets and one unit of FFP. Cryoprecipitate is produced by thawing FFP to precipitate the plasma proteins, then separated by centrifugation.

Table 113–2 Characteristics of Selected Blood Components

Component	Volume, mL	Content	Clinical Response
PRBC	180–200	RBCs with variable leukocyte content and small amount of plasma	Increase hemoglobin 10 g/L and hematocrit 3%
Platelets	50–70	5.5×10^{10}/RD unit	Increase platelet count 5000 –10,000/μL
	200–400	≥3×10^{11}/SDAP product	CCI ≥10×10^9/L within 1 h and ≥7.5×10^9/L within 24 h posttransfusion
FFP	200–250	Plasma proteins—coagulation factors, proteins C and S, antithrombin	Increases coagulation factors about 2%
Cryoprecipitate	10–15	Cold-insoluble plasma proteins, fibrinogen, factor VIII, vWF	Topical fibrin glue, also 80 IU factor VIII

Abbreviation: CCI, corrected count increment; FFP, fresh-frozen plasma; PRBC, packed red blood cells; RBC, red blood cell; RD, random donor; SDAP, single-donor apheresis platelets; vWF, von Willebrand factor.

Apheresis technology is used for the collection of multiple units of platelets from a single donor. These

single-donor apheresis platelets (SDAP) contain the equivalent of at least six units of RD platelets and have fewer contaminating leukocytes than pooled RD platelets.

Plasma may also be collected by apheresis. Plasma derivatives such as albumin, intravenous immunoglobulin, antithrombin, and coagulation factor concentrates are prepared from pooled plasma from many donors and are treated to eliminate infectious agents.

Whole Blood

Whole blood provides both oxygen-carrying capacity and volume expansion. It is the ideal component for patients who have sustained acute hemorrhage of ≥25% total blood volume loss. Whole blood is stored at 4°C to maintain erythrocyte viability, but platelet dysfunction and degradation of some coagulation factors occurs. In addition, 2,3-bisphosphoglycerate levels fall over time, leading to an increase in the oxygen affinity of the hemoglobin and a decreased capacity to deliver oxygen to the tissues, a problem with all red cell storage. Fresh whole blood avoids these problems, but it is typically used only in emergency settings (i.e., military). Whole blood is not readily available, since it is routinely processed into components.

Packed Red Blood Cells

This product increases oxygen-carrying capacity in the anemic patient. Adequate oxygenation can be maintained with a hemoglobin content of 70 g/L in the normovolemic patient without cardiac disease; however, comorbid factors may necessitate transfusion at a higher threshold. The decision to transfuse should be guided by the clinical situation and not by an arbitrary laboratory value. In the critical care setting, liberal use of transfusions to maintain near-normal levels of hemoglobin has not proven advantageous. In most patients requiring transfusion, levels of hemoglobin of 100 g/L are sufficient to keep oxygen supply from being critically low.

PRBCs may be modified to prevent certain adverse reactions. The majority of cellular blood products are now leukocyte reduced and universal prestorage leukocyte reduction has been recommended. Prestorage filtration appears superior to bedside filtration as smaller amounts of cytokines are generated in the stored product. These PRBC units contain $<5 \times 10^6$ donor white blood cells (WBCs), and their use lowers the incidence of posttransfusion fever, cytomegalovirus (CMV) infections, and alloimmunization. Other theoretical benefits include less immunosuppression in the recipient and lower risk of infections. Plasma, which may cause allergic reactions, can be removed from cellular blood components by washing.

Platelets

Thrombocytopenia is a risk factor for hemorrhage, and platelet transfusion reduces the incidence of bleeding. The threshold for prophylactic platelet transfusion is 10,000/μL. In patients without fever or infections, a threshold of 5000/μL may be sufficient to prevent spontaneous hemorrhage. For invasive procedures, 50,000/μL platelets is the usual target level.

Platelets are given either as pools prepared RDs or as SDAPs from a single donor. In an unsensitized patient without increased platelet consumption [splenomegaly, fever, disseminated intravascular coagulation (DIC)], two units of transfused RD per square-meter body surface area (BSA) is anticipated to increase the platelet count by approximately 10,000/uL. Patients who have received multiple transfusions may be alloimmunized to many HLA- and platelet-specific antigens and have little or no increase in their posttransfusion platelet counts. Patients who may require multiple transfusions are best served by receiving SDAP and leukocyte-reduced components to lower the risk of alloimmunization.

Refractoriness to platelet transfusion may be evaluated using the corrected count increment (CCI):

$$\text{CCI} = \frac{\text{posttransfusion count (/μL)} - \text{pretransfusion count (/μL)}}{\text{number of platelets transfused} \times 10^{-11}} \times \text{BSA (m}^2\text{)}$$

where BSA is body surface area measured in square meters. The platelet count performed 1 h after the transfusion is acceptable if the CCI is 10×10^9/mL, and after 18—24 h an increment of 7.5×10^9/mL is expected. Patients who have suboptimal responses are likely to have received multiple transfusions and have antibodies directed against class I HLA antigens. Refractoriness can be investigated by detecting anti-HLA antibodies in the recipient's serum. Patients who are sensitized will often react with 100% of the lymphocytes used for the HLA-antibody screen, and HLA-matched SDAPs should be considered for those patients who require transfusion. Although ABO-identical HLA-matched SDAPs provide the best chance for increasing the platelet count, locating these products is difficult. Platelet cross-matching is available in some centers. Additional clinical causes for a low platelet CCI include fever, bleeding, splenomegaly, DIC, or medications in the recipient.

Fresh-Frozen Plasma

FFP contains stable coagulation factors and plasma proteins: fibrinogen, antithrombin, albumin, as well as proteins C and S. Indications for FFP include correction of coagulopathies, including the rapid reversal of warfarin; supplying deficient plasma proteins; and treatment of thrombotic thrombocytopenic purpura. FFP should not be routinely used to expand blood volume. FFP is an

acellular component and does not transmit intracellular infections, e.g., CMV. Patients who are IgA-deficient and require plasma support should receive FFP from IgA-deficient donors to prevent anaphylaxis (see below).

Cryoprecipitate

Cryoprecipitate is a source of fibrinogen, factor VIII, and von Willebrand factor (vWF). It is ideal for supplying fibrinogen to the volume-sensitive patient. When factor VIII concentrates are not available, cryoprecipitate may be used since each unit contains approximately 80 units of factor VIII. Cryoprecipitate may also supply vWF to patients with dysfunctional (type II) or absent (type III) von Willebrand disease.

Plasma Derivatives

Plasma from thousands of donors may be pooled to derive specific protein concentrates, including albumin, intravenous immunoglobulin, antithrombin, and coagulation factors. In addition, donors who have high-titer antibodies to specific agents or antigens provide hyperimmune globulins, such as anti-D (RhoGam, WinRho), and antisera to hepatitis B virus (HBV), varicella-zoster virus, CMV, and other infectious agents.

ADVERSE REACTIONS TO BLOOD TRANSFUSION

Adverse reactions to transfused blood components occur despite multiple tests, inspections, and checks. Fortunately, the most common reactions are not life threatening, although serious reactions can present with mild symptoms and signs. Some reactions can be reduced or prevented by modified (filtered, washed, or irradiated) blood components. When an adverse reaction is suspected, the transfusion should be stopped and reported to the blood bank for investigation.

Transfusion reactions may result from immune and nonimmune mechanisms. Immune-mediated reactions are often due to preformed donor or recipient antibody; however, cellular elements may also cause adverse effects. Nonimmune causes of reactions are due to the chemical and physical properties of the stored blood component and its additives.

Transfusion-transmitted viral infections are increasingly rare due to improved screening and testing. As the risk of viral infection is reduced, the relative risk of other reactions increases, such as hemolytic transfusion reactions and sepsis from bacterially contaminated components. Pretransfusion quality assurance improvements further increase the safety of transfusion therapy. Infections, like any adverse transfusion reaction, must be brought to the attention of the blood bank for appropriate studies (Table 113–3).

Table 113–3 Risks of Transfusion Complications

	Frequency, Episodes: Unit
Reactions	
Febrile (FNHTR)	• 1–4:100
Allergic	• 1–4:100
Delayed hemolytic	• 1:1000
TRALI	• 1:5000
Acute hemolytic	• 1:12,000
Fatal hemolytic	• 1:100,000
Anaphylactic	• 1:150,000
Infections[a]	
Hepatitis B	• 1:220,000
Hepatitis C	• 1:1,800,000
HIV-1, -2	• 1:2,300,000
HTLV-I and -II	• 1:2,993,000
Malaria	•1:4,000,000
Other complications	
RBC allosensitization	• 1:100
HLA allosensitization	• 1:10
Graft-versus-host disease	Rare

[a]Infectious agents rarely associated with transfusion, theoretically possible or of unknown risk include West Nile virus, hepatitis A virus, parvovirus B-19, *Babesia microti* (babesiosis), *Borrelia burgdorferi* (Lyme disease), *Anaplasma phagocytophilum* (human granulocytic ehrlichiosis), *Trypanosoma cruzi* (Chagas disease), *Treponema pallidum*, and human herpesvirus-8.

Abbreviations: FNHTR, febrile nonhemolytic transfusion reaction; TRALI, transfusion-related acute lung injury; HTLV, human T lymphotropic virus; RBC, red blood cell.

Immune-Mediated Reactions

ACUTE HEMOLYTIC TRANSFUSION REACTIONS

Immune-mediated hemolysis occurs when the recipient has preformed antibodies that lyse donor erythrocytes. The ABO isoagglutinins are responsible for the majority of these reactions, although alloantibodies directed against other RBC antigens, i.e., Rh, Kell, and Duffy, may result in hemolysis.

Acute hemolytic reactions may present with hypotension, tachypnea, tachycardia, fever, chills, hemoglobinemia, hemoglobinuria, chest and/or flank pain, and discomfort at the infusion site. Monitoring the patient's vital signs before and during the transfusion is important to identify reactions promptly. When acute hemolysis is suspected, the transfusion must be stopped immediately, intravenous access maintained, and the reaction reported to the blood bank. A correctly labeled posttransfusion blood sample and any untransfused blood should be sent to the blood bank for analysis. The laboratory evaluation for hemolysis includes the measurement of serum haptoglobin, lactate dehydrogenase (LDH), and indirect bilirubin levels.

The immune complexes that result in RBC lysis can cause renal dysfunction and failure. Diuresis should be induced with intravenous fluids and furosemide or mannitol. Tissue factor released from the lysed erythrocytes may initiate DIC. Coagulation studies including prothrombin time (PT), activated partial thromboplastin time (aPTT), fibrinogen, and platelet count should be monitored in patients with hemolytic reactions.

Errors at the patient's bedside, such as mislabeling the sample or transfusing the wrong patient, are responsible for the majority of these reactions. The blood bank investigation of these reactions includes examination of the pre- and posttransfusion samples for hemolysis and repeat typing of the patient samples; direct antiglobulin test (DAT), sometimes called the *direct Coombs test*, of the posttransfusion sample; repeating the cross-matching of the blood component; and checking all clerical records for errors. DAT detects the presence of antibody or complement bound to RBCs in vivo.

DELAYED HEMOLYTIC AND SEROLOGIC TRANSFUSION REACTIONS

Delayed hemolytic transfusion reactions (DHTRs) are not completely preventable. These reactions occur in patients previously sensitized to RBC alloantigens who have a negative alloantibody screen due to low antibody levels. When the patient is transfused with antigen-positive blood, an anamnestic response results in the early production of alloantibody that binds donor RBCs. The alloantibody is detectable 1–2 weeks following the transfusion, and the posttransfusion DAT may become positive due to circulating donor RBCs coated with antibody or complement. The transfused, alloantibody-coated erythrocytes are cleared by the reticuloendothelial system. These reactions are detected most commonly in the blood bank when a subsequent patient sample reveals a positive alloantibody screen or a new alloantibody in a recently transfused recipient.

No specific therapy is usually required, although additional RBC transfusions may be necessary. Delayed serologic transfusion reactions are similar to DHTR, as the DAT is positive and alloantibody is detected; however, RBC clearance is not increased.

FEBRILE NONHEMOLYTIC TRANSFUSION REACTION

The most frequent reaction associated with the transfusion of cellular blood components is a febrile nonhemolytic transfusion reaction (FNHTR). These reactions are characterized by chills and rigors and a ≥1°C rise in temperature. FNHTR is diagnosed when other causes of fever in the transfused patient are ruled out. Antibodies directed against donor leukocyte and HLA antigens may mediate these reactions; thus, multiply transfused patients and multiparous women are felt to be at increased risk. Although anti-HLA antibodies may be demonstrated in the recipient's serum, investigation is not routinely done because of the mild nature of most FNHTR. The use of leukocyte-reduced blood products may prevent or delay sensitization to leukocyte antigens and thereby reduce the incidence of these febrile episodes. Cytokines released from cells within stored blood components may mediate FNHTR; thus, leukoreduction before storage may prevent these reactions.

ALLERGIC REACTIONS

Urticarial reactions are related to plasma proteins found in transfused components. Mild reactions may be treated symptomatically by temporarily stopping the transfusion and administering antihistamines (diphenhydramine, 50 mg orally or intramuscularly). The transfusion may be completed after the signs and/or symptoms resolve. Patients with a history of allergic transfusion reaction should be premedicated with an antihistamine. Cellular components can be washed to remove residual plasma

for the extremely sensitized patient.

ANAPHYLACTIC REACTION

This severe reaction presents after transfusion of only a few milliliters of the blood component. Symptoms and signs include difficulty breathing, coughing, nausea and vomiting, hypotension, bronchospasm, loss of consciousness, respiratory arrest, and shock. Treatment includes stopping the transfusion, maintaining vascular access, and administering epinephrine (0.5–1 mL of 1:1000 dilution subcutaneously). Glucocorticoids may be required in severe cases.

Patients who are IgA-deficient, <1% of the population, may be sensitized to this Ig class and are at risk for anaphylactic reactions associated with plasma transfusion. Individuals with severe IgA deficiency should therefore receive only IgA-deficient plasma and washed cellular blood components. Patients who have anaphylactic or repeated allergic reactions to blood components should be tested for IgA deficiency.

GRAFT-VERSUS-HOST DISEASE

Graft-versus-host disease (GVHD) is a frequent complication of allogeneic stem cell transplantation, in which lymphocytes from the donor attack and cannot be eliminated by an immunodeficient host. Transfusion-related GVHD is mediated by donor T lymphocytes that recognize host HLA antigens as foreign and mount an immune response, which is manifested clinically by the development of fever, a characteristic cutaneous eruption, diarrhea, and liver function abnormalities. GVHD can also occur when blood components that contain viable T lymphocytes are transfused to immunodeficient recipients or to immunocompetent recipients who share HLA antigens with the donor (e.g., a family donor). In addition to the aforementioned clinical features of GVHD, transfusion-associated GVHD (TA-GVHD) is characterized by marrow aplasia and pancytopenia. TA-GVHD is highly resistant to treatment with immunosuppressive therapies, including glucocorticoids, cyclosporine, antithymocyte globulin, and ablative therapy followed by allogeneic bone marrow transplantation. Clinical manifestations appear at 8–10 days, and death occurs at 3–4 weeks posttransfusion.

TA-GVHD can be prevented by irradiation of cellular components (minimum of 2500 cGy) before transfusion to patients at risk. Patients at risk for TA-GVHD include fetuses receiving intrauterine transfusions, selected immunocompetent (e.g., lymphoma patients) or immunocompromised recipients, recipients of donor units known to be from a blood relative, and recipients who have undergone marrow transplantation. Directed donations by family members should be discouraged (they are not less likely to transmit infection); lacking other options, the blood products from family members should always be irradiated.

TRANSFUSION-RELATED ACUTE LUNG INJURY

Transfusion-related acute lung injury (TRALI) presents as acute respiratory distress, either during or within 6 h of transfusing the patient. The recipient develops symptoms of respiratory compromise and signs of noncardiogenic pulmonary edema, including bilateral interstitial infiltrates on chest x-ray. Treatment is supportive, and patients usually recover without sequelae. TRALI usually results from the transfusion of donor plasma that contains high-titer anti-HLA antibodies that bind recipient leukocytes. The leukocytes aggregate in the pulmonary vasculature and release mediators that increase capillary permeability. Testing the donor's plasma for anti-HLA antibodies can support this diagnosis. The implicated donors are frequently multiparous women, and transfusion of their plasma component should be avoided.

POSTTRANSFUSION PURPURA

This reaction presents as thrombocytopenia 7–10 days after platelet transfusion and occurs predominantly in women. Platelet-specific antibodies are found in the recipient's serum, and the most frequently recognized antigen is HPA-1a found on the platelet glycoprotein IIIa receptor. The delayed thrombocytopenia is due to the production of antibodies that react to both donor and recipient platelets. Additional platelet transfusions can worsen the thrombocytopenia and should be avoided. Treatment with intravenous immunoglobulin may neutralize the effector antibodies, or plasmapheresis can be used to remove the antibodies.

ALLOIMMUNIZATION

A recipient may become alloimmunized to a number of antigens on cellular blood elements and plasma proteins. Alloantibodies to RBC antigens are detected during pretransfusion testing, and their presence may delay finding antigen-negative cross-match-compatible products for transfusion. Women of childbearing age who are sensitized to certain RBC antigens (i.e., D, c, E, Kell, or Duffy) are at risk for bearing a fetus with hemolytic disease of the newborn. Matching for D antigen is the only pretransfusion selection test to prevent RBC alloimmunization.

Alloimmunization to antigens on leukocytes and platelets can result in refractoriness to platelet transfusions. Once alloimmunization has developed, HLA-compatible platelets from donors who share similar antigens with the recipient may be difficult to find. Hence, prudent transfusion practice is directed at preventing sensitization through the use of leukocyte-reduced cellular components, as well as limiting antigenic exposure by the judicious use of transfusions and use of SDAPs.

Nonimmunologic Reactions

FLUID OVERLOAD

Blood components are excellent volume expanders, and transfusion may quickly lead to volume overload. Monitoring the rate and volume of the transfusion and using a diuretic can minimize this problem.

HYPOTHERMIA

Refrigerated (4°C) or frozen (–18°C or below) blood components can result in hypothermia when rapidly infused. Cardiac dysrhythmias can result from exposing the sinoatrial node to cold fluid. Use of an in-line warmer will prevent this complication.

ELECTROLYTE TOXICITY

RBC leakage during storage increases the concentration of potassium in the unit. Neonates and patients in renal failure are at risk for hyperkalemia. Preventive measures, such as using fresh or washed RBCs, are warranted for neonatal transfusions because this complication can be fatal.

Citrate, commonly used to anticoagulate blood components, chelates calcium and thereby inhibits the coagulation cascade. Hypocalcemia, manifested by circumoral numbness and/or tingling sensation of the fingers and toes, may result from multiple rapid transfusions. Because citrate is quickly metabolized to bicarbonate, calcium infusion is seldom required in this setting. If calcium or any other intravenous infusion is necessary, it must be given through a separate line.

IRON OVERLOAD

Each unit of RBCs contains 200–250 mg of iron. Symptoms and signs of iron overload affecting endocrine, hepatic, and cardiac function are common after 100 units of RBCs have been transfused (total-body iron load of 20 g). Preventing this complication by using alternative therapies (e.g., erythropoietin) and judicious transfusion is preferable and cost effective. Chelating agents, such as deferoxamine and deferasirox, are available, but the response though is often suboptimal.

HYPOTENSIVE REACTIONS

Transient hypotension may be noted among transfused patients who take angiotensin-converting enzyme (ACE) inhibitors. Since blood products contain bradykinin that is normally degraded by ACE, patients on ACE inhibitors may have increased bradykinin levels that cause hypotension in the recipient. The blood pressure typically returns to normal without intervention.

IMMUNOMODULATION

Transfusion of allogeneic blood is immunosuppressive. Multiply transfused renal transplant recipients are less likely to reject the graft, and transfusion may result in poorer outcomes in cancer patients and increase the risk of infections. Transfusion-related immunomodulation is thought to be mediated by transfused leukocytes. Leukocyte-depleted cellular products may cause less immunosuppression, though controlled data have not been obtained and are unlikely to be obtained as the blood supply becomes universally leukocyte-depleted.

Infectious Complications

The blood supply is initially screened by selecting healthy donors without high-risk lifestyles, medical conditions, or exposure to transmissible pathogens, such as intravenous drug use or visiting malaria endemic areas. Multiple tests performed on donated blood to detect the presence of infectious agents using nucleic acid amplification testing (NAT) or evidence of prior infections by testing for antibodies to pathogens further reduce the risk of transfusion-acquired infections.

VIRAL INFECTIONS

Hepatitis C virus Blood donations are tested for antibodies to HCV and HCV RNA. The risk of acquiring HCV through transfusion is now calculated to be approximately 1 in 2,000,000 units. Infection with HCV may be asymptomatic or lead to chronic active hepatitis, cirrhosis, and liver failure.

Human immunodeficiency virus type 1 Donated blood is tested for antibodies to HIV-1, HIV-1 p24 antigen, and HIV RNA using NAT. Approximately a dozen seronegative donors have been shown to harbor HIV RNA. The risk of HIV-1 infection per transfusion episode is 1 in 2 million. Antibodies to HIV-2 are also measured in donated blood. No cases of HIV-2 infection have been reported in the United States since 1992.

Hepatitis B virus

Donated blood is screened for HBV using assays for hepatitis B surface antigen (HbsAg). NAT testing is not practical because of slow viral replication and lower levels of viremia. The risk of transfusion-associated HBV infection is several times greater than for HCV. Vaccination of individuals who require long-term transfusion therapy can prevent this complication.

Other hepatitis viruses

Hepatitis A virus is rarely transmitted by transfusion; infection is typically asymptomatic and does not

lead to chronic disease. Other transfusion-transmitted viruses—TTV, SEN-V, and GBV-C—do not cause chronic hepatitis or other disease states. Routine testing does not appear to be warranted.

West nile virus

Transfusion-transmitted WNV infections were documented in 2002. This RNA virus can be detected using NAT; routine screening began in 2003. WNV infections range in severity from asymptomatic to fatal, with the older population at greater risk.

Cytomegalovirus

This ubiquitous virus infects ≥50% of the general population and is transmitted by the infected "passenger" WBCs found in transfused PRBCs or platelet components. Cellular components that are leukocyte-reduced have a decreased risk of transmitting CMV, regardless of the serologic status of the donor. Groups at risk for CMV infections include immunosuppressed patients, CMV-seronegative transplant recipients, and neonates; these patients should receive leukocyte-depleted components or CMV seronegative products.

Human T lymphotropic virus (HTLV) type I

Assays to detect HTLV-I and -II are used to screen all donated blood. HTLV-I is associated with adult T cell leukemia/lymphoma and tropical spastic paraparesis in a small percentage of infected persons (Chap. 188). The risk of HTLV-I infection via transfusion is 1 in 641,000 transfusion episodes. HTLV-II is not clearly associated with any disease.

Parvovirus B-19

Blood components and pooled plasma products can transmit this virus, the etiologic agent of erythema infectiosum, or fifth disease, in children. Parvovirus B-19 shows tropism for erythroid precursors and inhibits both erythrocyte production and maturation. Pure red cell aplasia, presenting either as acute aplastic crisis or chronic anemia with shortened RBC survival, may occur in individuals with an underlying hematologic disease, such as sickle cell disease or thalassemia (Chap. 107). The fetus of a seronegative woman is at risk for developing hydrops from this virus.

BACTERIAL CONTAMINATION

The relative risk of transfusion-transmitted bacterial infection has increased as the absolute risk of viral infections has dramatically decreased.

Most bacteria do not grow well at cold temperatures; thus, PRBCs and FFP are not common sources of bacterial contamination. However, some gram-negative bacteria can grow at 1° to 6°C. *Yersinia*, *Pseudomonas*, *Serratia*, *Acinetobacter*, and *Escherichia* species have all been implicated in infections related to PRBC transfusion. Platelet concentrates, which are stored at room temperature, are more likely to contain skin contaminants such as gram-positive organisms, including coagulase-negative staphylococci. It is estimated that 1 in 1000—2000 platelet components is contaminated with bacteria. The risk of death due to transfusion-associated sepsis has been calculated at 1 in 17,000 for single-unit platelets derived from whole blood donation and 1 in 61,000 for apheresis product. Since 2004, blood banks have instituted methods to detect contaminated platelet components.

Recipients of transfusion contaminated with bacteria may develop fever and chills, which can progress to septic shock and DIC. These reactions may occur abruptly, within minutes of initiating the transfusion, or after several hours. The onset of symptoms and signs is often sudden and fulminant, which distinguishes bacterial contamination from an FNHTR. The reactions, particularly those related to gram-negative contaminants, are the result of infused endotoxins formed within the contaminated stored component.

When these reactions are suspected, the transfusion must be stopped immediately. Therapy is directed at reversing any signs of shock, and broad-spectrum antibiotics should be given. The blood bank should be notified to identify any clerical or serologic error. The blood component bag should be sent for culture and Gram stain.

Other Infectious Agents

Various parasites, including those causing malaria, babesiosis, and Chagas disease, can be transmitted by blood transfusion. Geographic migration and travel of donors shift the incidence of these rare infections. Other agents implicated in transfusion transmission include dengue, chikungunya virus, variant Creutzfeldt-Jakob disease, *Anaplasma phagocytophilum*, and yellow fever vaccine virus and the list will grow. Tests for some pathogens are available, such as *Trypanosoma cruzi*, but not universally required. These infections should be considered in the transfused patient in the appropriate clinical setting.

ALTERNATIVES TO TRANSFUSION

Alternatives to allogeneic blood transfusions that avoid homologous donor exposures with attendant immunologic and infectious risks remain attractive. Autologous blood is the best option when transfusion is anticipated. However, the cost-benefit ratio of autologous transfusion remains high. No

transfusion is a zero-risk event; clerical errors and bacterial contamination remain potential complications even with autologous transfusions. Additional methods of autologous transfusion in the surgical patient include preoperative hemodilution, recovery of shed blood from sterile surgical sites, and postoperative drainage collection. Directed or designated donation from friends and family of the potential recipient has not been safer than volunteer donor component transfusions. Such directed donations may in fact place the recipient at higher risk for complications such as GVHD and alloimmunization.

Granulocyte and granulocyte-macrophage colony-stimulating factors are clinically useful to hasten leukocyte recovery in patients with leukopenia related to high-dose chemotherapy. Erythropoietin stimulates erythrocyte production in patients with anemia of chronic renal failure and other conditions, thus avoiding or reducing the need for transfusion. This hormone can also stimulate erythropoiesis in the autologous donor to enable additional donation.

FURTHER READINGS

Brecher ME, Hay SN: Bacterial contamination of blood components. Clin Microbiol Rev 18:195, 2005[PMID: 15653826]

Chaiwat O, Lang JD: Early packed red blood cell transfusion and acute respiratory distress syndrome after trauma. Anesthesiology 110;351, 2009

Gerber DR: Transfusion of packed red blood cells in patient with ischemic heart disease. Crit Care Med 36:1068, 2008[PMID: 18379229]

Hendrickson JE, Hillyer CD: Noninfectious serious hazards for transfusion. Anesth Analg 108:759, 2009[PMID: 19224780]

Lange MM et al: Leucocyte depletion of perioperative blood transfusion does not affect long-term survival and recurrence in patients with gastrointestinal cancer. Br J Surg 96;734, 2009

Petersen LR, Busch MP: Transfusion-transmitted arboviruses. Vox Sang 98:495, 2010[PMID: 19951309]

Slichter SJ, Kaufman RM: Dose of prophylactic platelet transfusions and prevention of hemorrhage. N Engl J Med 362;600, 2010

Triulzi DJ: Transfusion-related acute lung injury: current concepts for the clinician. Anesth Analg 108;770, 2009

Vamvakas EC, Blajchman MA: Transfusion-related mortality: The ongoing risks of allogeneic blood transfusion and the available strategies for their prevention. Blood 113;3406, 2009

——— et al: *The Technical Manual*, 16th ed. Arlington, VA, American Association of Blood Banks, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEMATOPOIETIC CELL TRANSPLANTATION: INTRODUCTION

Bone marrow transplantation was the original term used to describe the collection and transplantation of hematopoietic stem cells, but with the demonstration that the peripheral blood and umbilical cord blood are also useful sources of stem cells, *hematopoietic cell transplantation* has become the preferred generic term for this process. The procedure is usually carried out for one of two purposes: (1) to replace an abnormal but nonmalignant lymphohematopoietic system with one from a normal donor or (2) to treat malignancy by allowing the administration of higher doses of myelosuppressive therapy than would otherwise be possible. The use of hematopoietic cell transplantation has been increasing, both because of its efficacy in selected diseases and because of increasing availability of donors. The Center for International Blood and Marrow Transplant Research (*http://www.cibmtr.org*) estimates that about 65,000 transplants are performed each year.

THE HEMATOPOIETIC STEM CELL

Several features of the hematopoietic stem cell make transplantation clinically feasible, including its remarkable regenerative capacity, its ability to home to the marrow space following intravenous injection, and the ability of the stem cell to be cryopreserved (Chap. 66). Transplantation of a single stem cell can replace the entire lymphohematopoietic system of an adult mouse. In humans, transplantation of a few percent of a donor's bone marrow volume regularly results in complete and sustained replacement of the recipient's entire lymphohematopoietic system, including all red cells, granulocytes, B and T lymphocytes, and platelets, as well as cells comprising the fixed macrophage population, including Kupffer cells of the liver, pulmonary alveolar macrophages, osteoclasts, Langerhans cells of the skin, and brain microglial cells. The ability of the hematopoietic stem cell to home to the marrow following intravenous injection is mediated, in part, by an interaction between stromal cell–derived factor 1 (SDF1) produced by marrow stromal cells and the alpha-chemokine receptor CXCR4 found on stem cells. Homing is also influenced by the interaction of cell-surface molecules, termed *selectins*, on bone marrow endothelial cells with ligands, termed *integrins*, on early hematopoietic cells. Human hematopoietic stem cells can survive freezing and thawing with little, if any, damage, making it possible to remove and store a portion of the patient's own bone marrow for later reinfusion following treatment of the patient with high-dose myelotoxic therapy.

CATEGORIES OF HEMATOPOIETIC CELL TRANSPLANTATION

Hematopoietic cell transplantation can be described according to the relationship between the patient and the donor and by the anatomic source of stem cells. In ~1% of cases, patients have identical twins who can serve as donors. With the use of syngeneic donors, there is no risk of graft-versus-host disease (GVHD) which often complicates allogeneic transplantation, and unlike the use of autologous marrow, there is no risk that the stem cells are contaminated with tumor cells.

Allogeneic transplantation involves a donor and a recipient who are not genetically identical. Following allogeneic transplantation, immune cells transplanted with the stem cells or developing from them can react against the patient, causing GVHD. Alternatively, if the immunosuppressive preparative regimen used to treat the patient before transplant is inadequate, immunocompetent cells of the patient can cause graft rejection. The risks of these complications are greatly influenced by the degree of matching between donor and recipient for antigens encoded by genes of the major histocompatibility complex.

The human leukocyte antigen (HLA) molecules are responsible for binding antigenic proteins and presenting them to T cells. The antigens presented by HLA molecules may derive from exogenous sources (e.g., during active infections) or may be endogenous proteins. If individuals are not HLA-matched, T cells from one individual will react strongly to the mismatched HLA, or "major antigens," of the second. Even if the individuals are HLA-matched, the T cells of the donor may react to differing endogenous or "minor antigens" presented by the HLA of the recipient. Reactions to minor antigens tend to be less vigorous. The genes of major relevance to transplantation include HLA-A, -B, -C, and -D; they are closely linked and therefore tend to be inherited as haplotypes, with only rare crossovers between them. Thus, the odds that any one full sibling will match a patient are one in four, and the probability that the patient has an HLA-identical sibling is $1 - (0.75)^n$, where n equals the number of siblings.

With current techniques, the risk of graft rejection is 1–3%, and the risk of severe, life-threatening

acute GVHD is ~15% following transplantation between HLA-identical siblings. The incidence of graft rejection and GVHD increases progressively with the use of family member donors mismatched for one, two, or three antigens. While survival following a one-antigen mismatched transplant is not markedly altered, survival following two- or three-antigen mismatched transplants is significantly reduced, and such transplants should be performed only as part of clinical trials.

Since the formation of the National Marrow Donor Program and other registries, it has become possible to identify HLA-matched unrelated donors for many patients. The genes encoding HLA antigens are highly polymorphic, and thus the odds of any two unrelated individuals being HLA-identical are extremely low, somewhat less than 1 in 10,000. However, by identifying and typing >14 million volunteer donors, HLA-matched donors can now be found for ~50% of patients for whom a search is initiated. It takes, on average, 3–4 months to complete a search and schedule and initiate an unrelated donor transplant. With improvements in HLA- typing and supportive care measures, survival following matched unrelated donor transplantation is essentially the same as that seen with HLA-matched siblings.

Autologous transplantation involves the removal and storage of the patient's own stem cells with subsequent reinfusion after the patient receives high-dose myeloablative therapy. Unlike allogeneic transplantation, there is no risk of GVHD or graft rejection with autologous transplantation. On the other hand, autologous transplantation lacks a graft-versus-tumor (GVT) effect, and the autologous stem cell product can be contaminated with tumor cells, which could lead to relapse. A variety of techniques have been developed to "purge" autologous products of tumor cells. Some use antibodies directed at tumor-associated antigens plus complement, antibodies linked to toxins, or antibodies conjugated to immunomagnetic beads. In vitro incubation with certain chemotherapeutic agents such as 4-hydroperoxycyclophosphamide and long-term culture of bone marrow have also been shown to diminish tumor cell numbers in stem cell products. Another technique is positive selection of stem cells using antibodies to CD34, with subsequent column adherence or flow techniques to select normal stem cells while leaving tumor cells behind. All these approaches can reduce the number of tumor cells from 1000- to 10,000-fold and are clinically feasible; however, no prospective randomized trials have yet shown that any of these approaches results in a decrease in relapse rates or improvements in disease-free or overall survival.

Bone marrow aspirated from the posterior and anterior iliac crests has traditionally been the source of hematopoietic stem cells for transplantation. Typically, anywhere from 1.5 to 5 x 10^8 nucleated marrow cells per kilogram are collected for allogeneic transplantation. Several studies have found improved survival in the settings of both matched sibling and unrelated transplantation by transplanting higher numbers of bone marrow cells.

Hematopoietic stem cells circulate in the peripheral blood but in very low concentrations. Following the administration of certain hematopoietic growth factors, including granulocyte colony-stimulating factor (G-CSF) or granulocyte-macrophage colony-stimulating factor (GM-CSF), and during recovery from intensive chemotherapy, the concentration of hematopoietic progenitor cells in blood, as measured either by colony-forming units or expression of the CD34 antigen, increases markedly. This has made it possible to harvest adequate numbers of stem cells from the peripheral blood for transplantation. Donors are typically treated with 4 or 5 days of hematopoietic growth factor, following which stem cells are collected in one or two 4-h pheresis sessions. In the autologous setting, transplantation of >2.5 x 10^6 CD34 cells per kilogram, a number that can be collected in most circumstances, leads to rapid and sustained engraftment in virtually all cases. In the 10–20% of patients who fail to mobilize sufficient CD34+ cells with growth factor alone, the addition of plerixafor, an antagonist of CXCR4, may be useful. Compared to the use of autologous marrow, use of peripheral blood stem cells results in more rapid hematopoietic recovery, with granulocytes recovering to 500/μL by day 12 and platelets recovering to 20,000/μL by day 14. While this more rapid recovery diminishes the morbidity rate of transplantation, no studies show improved survival.

Hesitation in studying the use of peripheral blood stem cells for allogeneic transplantation was because peripheral blood stem cell products contain as much as 1 log more T cells than are contained in the typical marrow harvest; in animal models, the incidence of GVHD is related to the number of T cells transplanted. Nonetheless, clinical trials have shown that the use of growth factor–mobilized peripheral blood stem cells from HLA-matched family members leads to faster engraftment without an increase in acute GVHD. Chronic GVHD may be increased with peripheral blood stem cells, but in trials conducted so far, this has been more than balanced by reductions in relapse rates and nonrelapse mortality rates, with the use of peripheral blood stem cells resulting in improved overall survival. Randomized trials are now evaluating the use of peripheral blood versus bone marrow for matched unrelated donor transplantation.

Umbilical cord blood contains a high concentration of hematopoietic progenitor cells, allowing for its use as a source of stem cells for transplantation. Cord blood transplantation from family members has been explored in the setting where the immediate need for transplantation precludes waiting the 9 or so months generally required for the baby to mature to the point of donating marrow. Use of cord blood results in slower engraftment and peripheral count recovery than seen with marrow but a low

incidence of GVHD, perhaps reflecting the low number of T cells in cord blood. Several banks have been developed to harvest and store cord blood for possible transplantation to unrelated patients from material that would otherwise be discarded. A summary of the first 562 unrelated cord blood transplants, facilitated by the New York Blood Center, reported engraftment in ~85% of patients but at a slower pace than seen with marrow. Severe GVHD was seen in 23% of patients. The risk of graft failure and transplant-related mortality were related to the dose of cord blood cells per kilogram, thus limiting the application of single cord blood transplantation for the treatment of larger adolescent and adult patients. Subsequent trials suggest that the use of double cord transplants diminishes the risk of graft failure and early mortality even though only one of the donors ultimately engrafts.

THE TRANSPLANT PREPARATIVE REGIMEN

The treatment regimen administered to patients immediately preceding transplantation is designed to eradicate the patient's underlying disease and, in the setting of allogeneic transplantation, immunosuppress the patient adequately to prevent rejection of the transplanted marrow. The appropriate regimen therefore depends on the disease setting and source of marrow. For example, when transplantation is performed to treat severe combined immunodeficiency and the donor is a histocompatible sibling, no treatment is needed because no host cells require eradication and the patient is already too immunoincompetent to reject the transplanted marrow. For aplastic anemia, there is no large population of cells to eradicate, and high-dose cyclophosphamide plus antithymocyte globulin are sufficient to immunosuppress the patient adequately to accept the marrow graft. In the setting of thalassemia and sickle cell anemia, high-dose busulfan is frequently added to cyclophosphamide in order to eradicate hyperplastic host hematopoiesis. A variety of different regimens have been developed to treat malignant diseases. Most of these regimens include agents that have high activity against the tumor in question at conventional doses and have myelosuppression as their predominant dose-limiting toxicity. Therefore, these regimens commonly include busulfan, cyclophosphamide, melphalan, thiotepa, carmustine, etoposide, and total-body irradiation in various combinations.

Although high-dose treatment regimens have typically been used in transplantation, the understanding that much of the antitumor effect of transplantation derives from an immunologically mediated GVT response has led investigators to ask if reduced-intensity conditioning regimens might be effective and more tolerable. Evidence for a GVT effect comes from studies showing that posttransplant relapse rates are lowest in patients who develop acute and chronic GVHD, higher in those without GVHD, and higher still in recipients of T cell–depleted allogeneic or syngeneic marrow. The demonstration that complete remissions can be obtained in many patients who have relapsed posttransplant by simply administering viable lymphocytes from the original donor further strengthens the argument for a potent GVT effect. Accordingly, a variety of less-intensive nonmyeloablative regimens have been studied, ranging in intensity from the very minimum required to achieve engraftment (e.g., fludarabine plus 200 cGy total-body irradiation) to regimens of more immediate intensity (e.g., fludarabine plus melphalan). Studies to date document that engraftment can be readily achieved with less toxicity than seen with conventional transplantation. Furthermore, the severity of acute GVHD appears to be decreased because less tissue damage is done by the lower doses of drugs in the preparative regimen. Complete sustained responses have been documented in many patients, particularly those with more indolent hematologicmalignancies. The role of reduced-intensity conditioning in any disease, however, has not been fully defined.

The Transplant Procedure

Marrow is usually collected from the donor's posterior and sometimes anterior iliac crests, with the donor under general or spinal anesthesia. Typically, 10–15 mL/kg of marrow is aspirated, placed in heparinized media, and filtered through 0.3- and 0.2-mm screens to remove fat and bony spicules. The collected marrow may undergo further processing depending on the clinical situation, such as the removal of red cells to prevent hemolysis in ABO-incompatible transplants, the removal of donor T cells to prevent GVHD, or attempts to remove possible contaminating tumor cells in autologous transplantation. Marrow donation is safe, with only very rare complications reported.

Peripheral blood stem cells are collected by leukapheresis after the donor has been treated with hematopoietic growth factors or, in the setting of autologous transplantation, sometimes after treatment with a combination of chemotherapy and growth factors. Stem cells for transplantation are generally infused through a large-bore central venous catheter. Such infusions are usually well tolerated, although occasionally patients develop fever, cough, or shortness of breath. These symptoms usually resolve with slowing of the infusion. When the stem cell product has been cryopreserved using dimethyl sulfoxide, patients more often experience short-lived nausea or vomiting due to the odor and taste of the cryoprotectant.

Engraftment

Peripheral blood counts usually reach their nadir several days to a week posttransplant as a consequence of the preparative regimen; then cells produced by the transplanted stem cells begin to appear in the peripheral blood. The rate of recovery depends on the source of stem cells, the use of

posttransplant growth factors, and the form of GVHD prophylaxis employed. If marrow is the source of stem cells, recovery to 100 granulocytes/μL occurs by day 16 and to 500/μL by day 22. Use of G-CSF–mobilized peripheral blood stem cells speeds the rate of recovery by ~1 week when compared to marrow, whereas engraftment following cord blood transplantation is typically delayed by ~1 week compared to marrow. Use of a myeloid growth factor (G-CSF or GM-CSF) posttransplant can accelerate recovery by 3–5 days, while use of methotrexate to prevent GVHD delays engraftment by a similar period. Following allogeneic transplantation, engraftment can be documented using fluorescence in situ hybridization of sex chromosomes if donor and recipient are sex-mismatched, HLA-typing if HLA-mismatched, or restriction fragment length polymorphism analysis if sex- and HLA-matched.

Complications Following Hematopoietic Cell Transplant

EARLY DIRECT CHEMORADIOTOXICITIES

The transplant preparative regimen may cause a spectrum of acute toxicities that vary according to intensity of the regimen and the specific agents used, but frequently results in nausea, vomiting, and mild skin erythema **(Fig. 114-1)**. Regimens that include high-dose cyclophosphamide can result in hemorrhagic cystitis, which can usually be prevented by bladder irrigation or with the sulfhydryl compound mercaptoethanesulfonate (MESNA); rarely, acute hemorrhagic carditis is seen. Most high-dose preparative regimens will result in oral mucositis, which typically develops 5–7 days posttransplant and often requires narcotic analgesia. Use of a patient-controlled analgesic pump provides the greatest patient satisfaction and results in a lower cumulative dose of narcotic. Keratinocyte growth factor (palifermin) can shorten the duration of mucositis by several days following autologous transplantation. Patients begin losing their hair 5–6 days posttransplant and by 1 week are usually profoundly pancytopenic.

Figure 114-1

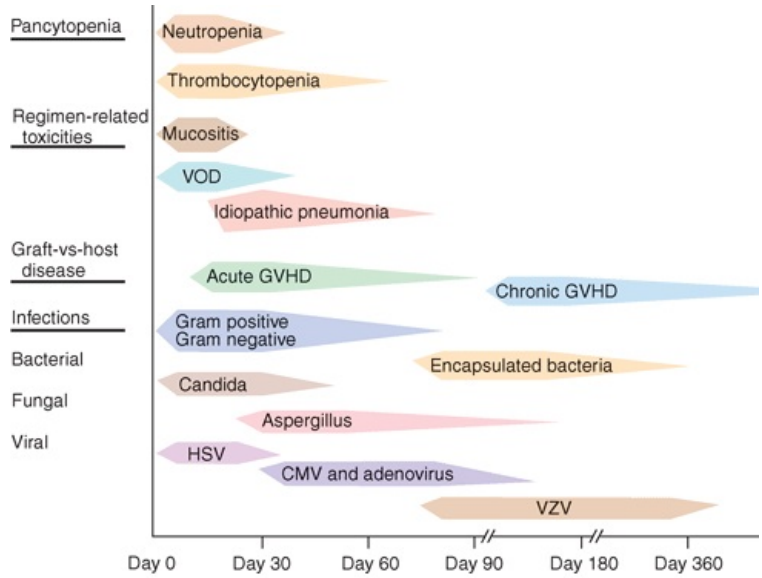

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Major syndromes complicating marrow transplantation. VOD, venoocclusive disease; GVHD, graft-versus-host disease; HSV, herpes simplex virus; CMV, cytomegalovirus; VZV, varicella-zoster virus. The size of the shaded area roughly reflects the risk of the complication.

Depending on the intensity of the conditioning regimen, 3–10% of patients will develop sinusoidal obstruction syndrome of the liver, a syndrome that results from direct cytotoxic injury to hepatic-venular and sinusoidal endothelium, with subsequent deposition of fibrin and the development of a local hypercoagulable state. This chain of events leads to the clinical symptoms of tender hepatomegaly, ascites, jaundice, and fluid retention. These symptoms can develop any time during the first month posttransplant, with the peak incidence at day 16. Predisposing factors include prior exposure to intensive chemotherapy, pretransplant hepatitis of any cause, and use of more intense conditioning regimens. The mortality rate of sinusoidal obstruction syndrome is ~30%, with progressive hepatic failure culminating in a terminal hepatorenal syndrome. Both thrombolytic and antithrombotic agents, such as tissue plasminogen activator, heparin, and prostaglandin E, have been studied as therapy, but none has proven of consistent major benefit in controlled trials, and all have significant toxicity. Early studies with defibrotide, a polydeoxyribonucleotide, seem encouraging.

Although most pneumonias developing posttransplant are caused by infectious agents, in ~5% of

patients a diffuse interstitial pneumonia will develop that is thought to be the result of direct toxicity of high-dose preparative regimens. Bronchoalveolar lavage typically shows alveolar hemorrhage, and biopsies are typically characterized by diffuse alveolar damage, although some cases may have a more clearly interstitial pattern. High-dose glucocorticoids or antitumor necrosis factor therapies are sometimes used as treatment, although randomized trials testing their utility have not been reported.

LATE DIRECT CHEMORADIOTOXICITIES

Late complications of the preparative regimen include decreased growth velocity in children and delayed development of secondary sex characteristics. These complications can be partly ameliorated with the use of appropriate growth and sex hormone replacement. Most men become azoospermic, and most postpubertal women will develop ovarian failure, which should be treated. Thyroid dysfunction, usually well compensated, is sometimes seen. Cataracts develop in 10–20% of patients and are most common in patients treated with total-body irradiation and those who receive glucocorticoid therapy posttransplant for treatment of GVHD. Aseptic necrosis of the femoral head is seen in 10% of patients and is particularly frequent in those receiving chronic glucocorticoid therapy. Both acute and late chemoradiotoxicities (except those due to glucocorticoids) are considerably less frequent in recipients of reduced- compared to high-dose preparative regimens.

GRAFT-VERSUS-HOST DISEASE

GVHD is the result of allogeneic T cells that are transferred with the donor's stem cell inoculum reacting with antigenic targets on host cells. GVHD developing within the first 3 months posttransplant is termed *acute GVHD*, while GVHD developing or persisting beyond 3 months posttransplant is termed *chronic GVHD*. Acute GVHD most often first becomes apparent 2–4 weeks posttransplant and is characterized by an erythematous maculopapular rash; persistent anorexia or diarrhea, or both; and by liver disease with increased serum levels of bilirubin, alanine and aspartate aminotransferase, and alkaline phosphatase. Since many conditions can mimic acute GVHD, diagnosis usually requires skin, liver, or endoscopic biopsy for confirmation. In all these organs, endothelial damage and lymphocytic infiltrates are seen. In skin, the epidermis and hair follicles are damaged; in liver, the small bile ducts show segmental disruption; and in intestines, destruction of the crypts and mucosal ulceration may be noted. A commonly used rating system for acute GVHD is shown in Table 114–1. Grade I acute GVHD is of little clinical significance, does not affect the likelihood of survival, and does not require treatment. In contrast, grades II to IV GVHD are associated with significant symptoms and a poorer probability of survival, and they require aggressive therapy. The incidence of acute GVHD is higher in recipients of stem cells from mismatched or unrelated donors, in older patients, and in patients unable to receive full doses of drugs used to prevent the disease.

Table 114–1 Clinical Staging and Grading of Acute Graft-versus-Host Disease

Clinical Stage	**Skin**	**Liver—Bilirubin, μmol/L (mg/dL)**	**Gut**
1	Rash <25% body surface	34–51 (2–3)	Diarrhea 500–1000 mL/d
2	Rash 25–50% body surface	51–103 (3–6)	Diarrhea 1000–1500 mL/d
3	Generalized erythroderma	103–257 (6–15)	Diarrhea >1500 mL/d
4	Desquamation and bullae	>257 (>15)	Ileus
Overall Clinical Grade	**Skin Stage**	**Liver Stage**	**Gut Stage**
I	1–2	0	0
II	1–3	1	1
III	1–3	2–3	2–3
IV	2–4	2–4	2–4

One general approach to the prevention of GVHD is the administration of immunosuppressive drugs early after transplant. Combinations of methotrexate and either cyclosporine or tacrolimus are among the most effective and widely used regimens. Prednisone, anti–T cell antibodies, mycophenolate mofetil, and other immunosuppressive agents have also been or are being studied in various combinations. A second general approach to GVHD prevention is removal of T cells from the stem cell inoculum. While effective in preventing GVHD, T cell depletion is associated with an increased incidence of graft failure and of tumor recurrence posttransplant; as yet, little evidence suggests that T-cell depletion improves cure rates in any specific setting.

Despite prophylaxis, significant acute GVHD will develop in ~30% of recipients of stem cells from matched siblings and in as many as 60% of those receiving stem cells from unrelated donors. The

disease is usually treated with glucocorticoids, additional immunosuppressants or monoclonal antibodies targeted against T cells or T cell subsets.

Between 20 and 50% of patients surviving >6 months after allogeneic transplantation will develop chronic GVHD. The disease is more common in older patients, in recipients of mismatched or unrelated stem cells, and in those with a preceding episode of acute GVHD. The disease resembles an autoimmune disorder with malar rash, sicca syndrome, arthritis, obliterative bronchiolitis, and bile duct degeneration and cholestasis. Single-agent prednisone or cyclosporine is standard treatment at present, although trials of other agents are under way. In most patients, chronic GVHD resolves, but it may require 1–3 years of immunosuppressive treatment before these agents can be withdrawn without the disease recurring. Because patients with chronic GVHD are susceptible to significant infection, they should receive prophylactic trimethoprim-sulfamethoxazole, and all suspected infections should be investigated and treated aggressively.

GRAFT FAILURE

While complete and sustained engraftment is usually seen posttransplant, occasionally marrow function either does not return or, after a brief period of engraftment, is lost. Graft failure after autologous transplantation can be the result of inadequate numbers of stem cells being transplanted, damage during ex vivo treatment or storage, or exposure of the patient to myelotoxic agents posttransplant. Infections with cytomegalovirus (CMV) or human herpesvirus type 6 have also been associated with loss of marrow function. Graft failure after allogeneic transplantation can also be due to immunologic rejection of the graft by immunocompetent host cells. Immunologically based graft rejection is more common following use of less-immunosuppressive preparative regimens, in recipients of T cell–depleted stem cell products, and in patients receiving grafts from HLA-mismatched donors or cord blood.

Treatment of graft failure usually involves removing all potentially myelotoxic agents from the patient's regimen and attempting a short trial of a myeloid growth factor. Persistence of lymphocytes of host origin in allogeneic transplant recipients with graft failure indicates immunologic rejection. Reinfusion of donor stem cells in such patients is usually unsuccessful unless preceded by a second immunosuppressive preparative regimen. Standard high-dose preparative regimens are generally tolerated poorly if administered within 100 days of a first transplant because of cumulative toxicities. However, use of regimens combining, for example, anti-CD3 antibodies with high-dose glucocorticoids, fludarabine plus low-dose total-body irradiation, or cyclophosphamide plus antithymocyte globulin, has been effective in some cases.

INFECTION

Posttransplant patients, particularly recipients of allogeneic transplantation, require unique approaches to the problem of infection. Early after transplantation, patients are profoundly neutropenic, and because the risk of bacterial infection is so great, most centers initiate antibiotic treatment once the granulocyte count falls to <500/μL. Fluconazole prophylaxis at a dose of 200–400 mg/kg per day reduces the risk of candidal infections. Patients seropositive for herpes simplex should receive acyclovir prophylaxis. One approach to infection prophylaxis is shown in Table 114–2. Despite these prophylactic measures, most patients will develop fever and signs of infection posttransplant. The management of patients who become febrile despite bacterial and fungal prophylaxis is a difficult challenge and is guided by individual aspects of the patient and by the institution's experience.

Table 114–2 Approach to Infection Prophylaxis in Allogeneic Transplant Recipients

Organism		Approach
Bacterial	Levofloxacin	750 mg PO or IV daily
Fungal	Fluconazole	400 mg PO qd to day 75 posttransplant
Pneumocystis carinii	Trimethoprim-sulfamethoxazole	1 double-strength tablet PO bid 2 days/week until day 180 or off immunosuppression
Viral		
Herpes simplex	Acyclovir	800 mg PO bid to day 30
Varicella-zoster	Acyclovir	800 mg PO bid to day 365
Cytomegalovirus	Ganciclovir	5 mg/kg IV bid for 7 days, then 5 (mg/kg)/d 5 days/week to day 100

The general problem of infection in the immunocompromised host is discussed in Chap. 132.

Once patients engraft, the incidence of bacterial infection diminishes; however, patients, particularly allogeneic transplant recipients, remain at significant risk of infection. During the period from engraftment until about 3 months posttransplant, the most common causes of infection are gram-positive bacteria, fungi (particularly *Aspergillus*), and viruses including CMV. CMV infection, which in

the past was frequently seen and often fatal, can be prevented in seronegative patients transplanted from seronegative donors by the use of either seronegative blood products or products from which the white blood cells have been removed. In seropositive patients or patients transplanted from seropositive donors, the use of ganciclovir, either as prophylaxis beginning at the time of engraftment or initiated when CMV first reactivates as evidenced by development of antigenemia or viremia, can significantly reduce the risk of CMV disease. Foscarnet is effective for some patients who develop CMV antigenemia or infection despite the use of ganciclovir or who cannot tolerate the drug.

Pneumocystis jiroveci pneumonia, once seen in 5–10% of patients, can be prevented by treating patients with oral trimethoprim-sulfamethoxazole for 1 week pretransplant and resuming the treatment once patients have engrafted.

The risk of infection diminishes considerably beyond 3 months after transplant unless chronic GVHD develops, requiring continuous immunosuppression. Most transplant centers recommend continuing trimethoprim-sulfamethoxazole prophylaxis while patients are receiving any immunosuppressive drugs and also recommend careful monitoring for late CMV reactivation. In addition, many centers recommend prophylaxis against varicella zoster, using acyclovir for 1 year posttransplant. Patients should be revaccinated against tetanus, diphtheria, haemophilus influenza, polio, and pneumococcal pneumonia starting at 12 months posttransplant and against measles, mumps, and rubella at 24 months.

TREATMENT OF SPECIFIC DISEASES USING HEMATOPOIETIC CELL TRANSPLANTATION

Treatment: Nonmalignant Diseases

IMMUNODEFICIENCY DISORDERS

By replacing abnormal stem cells with cells from a normal donor, hematopoietic cell transplantation can cure patients of a variety of immunodeficiency disorders including severe combined immunodeficiency, Wiskott-Aldrich syndrome, and Chédiak-Higashi syndrome. The widest experience has been with severe combined immunodeficiency disease, where cure rates of 90% can be expected with HLA-identical donors and success rates of 50–70% have been reported using haplotype-mismatched parents as donors **(Table 114–3)**.

Table 114–3 Estimated 5-Year Survival Rates Following Transplantation*

Disease	Allogeneic, %	Autologous, %
Severe combined immunodeficiency	90	N/A
Aplastic anemia	90	N/A
Thalassemia	90	N/A
Acute myeloid leukemia		
First remission	55–60	50
Second remission	40	30
Acute lymphocytic leukemia		
First remission	50	40
Second remission	40	30
Chronic myeloid leukemia		
Chronic phase	70	ID
Accelerated phase	40	ID
Blast crisis	15	ID
Chronic lymphocytic leukemia	50	ID
Myelodysplasia	45	ID
Multiple myeloma	30	35
Non-Hodgkin's lymphoma		
First relapse/second remission	40	40
Hodgkin's disease		
First relapse/second remission	40	50
Breast cancer		
High-risk stage II	N/A	70
Stage IV	N/A	15

*These estimates are generally based on data reported by the International Bone Marrow Transplant Registry. The analysis has not been reviewed by their Advisory Committee.

Abbreviations: N/A, not applicable; ID, insufficient data.

APLASTIC ANEMIA

Transplantation from matched siblings after a preparative regimen of high-dose cyclophosphamide and antithymocyte globulin can cure up to 90% of patients age <40 years with severe aplastic anemia. Results in older patients and in recipients of mismatched family member or unrelated marrow are less favorable; therefore, a trial of immunosuppressive therapy is generally recommended for such patients before considering transplantation. Transplantation is effective in all forms of aplastic anemia including, for example, the syndromes associated with paroxysmal nocturnal hemoglobinuria and Fanconi's anemia. Patients with Fanconi's anemia are abnormally sensitive to the toxic effects of alkylating agents and so less intensive preparative regimens must be used in their treatment (Chap. 107).

HEMOGLOBINOPATHIES

Marrow transplantation from an HLA-identical sibling following a preparative regimen of busulfan and cyclophosphamide can cure 70–90% of patients with thalassemia major. The best outcomes can be expected if patients are transplanted before they develop hepatomegaly or portal fibrosis and if they have been given adequate iron chelation therapy. Among such patients, the probabilities of 5-year survival and disease-free survival are 95 and 90%, respectively. Although prolonged survival can be achieved with aggressive chelation therapy, transplantation is the only curative treatment for thalassemia. Transplantation is being studied as a curative approach to patients with sickle cell anemia. Two-year survival and disease-free survival rates of 90 and 80%, respectively, have been reported following matched sibling transplantation. Decisions about patient selection and the timing of transplantation remain difficult, but transplantation represents a reasonable option for younger patients who suffer repeated crises or other significant complications and who have not responded to other interventions (Chap. 104).

OTHER NONMALIGNANT DISEASES

Theoretically, hematopoietic cell transplantation should be able to cure any disease that results from an inborn error of the lymphohematopoietic system. Transplantation has been used successfully to treat congenital disorders of white blood cells such as Kostmann's syndrome, chronic granulomatous disease, and leukocyte adhesion deficiency. Congenital anemias such as Blackfan-Diamond anemia can also be cured with transplantation. Infantile malignant osteopetrosis is due to an inability of the osteoclast to resorb bone, and since osteoclasts derive from the marrow, transplantation can cure this rare inherited disorder.

Hematopoietic cell transplantation has been used as treatment for a number of storage diseases caused by enzymatic deficiencies, such as Gaucher's disease, Hurler's syndrome, Hunter's syndrome, and infantile metachromatic leukodystrophy. Transplantation for these diseases has not been uniformly successful, but treatment early in the course of these diseases, before irreversible damage to extramedullary organs has occurred, increases the chance for success.

Transplantation is being explored as a treatment for severe acquired autoimmune disorders. These trials are based on studies demonstrating that transplantation can reverse autoimmune disorders in animal models and on the observation that occasional patients with coexisting autoimmune disorders and hematologicmalignancies have been cured of both with transplantation.

Treatment: Malignant Diseases

ACUTE LEUKEMIA

Allogeneic hematopoietic cell transplantation cures 15–20% of patients who do not achieve complete response from induction chemotherapy for acute myeloid leukemia (AML) and is the only form of therapy that can cure such patients. Cure rates of 30–35% are seen when patients are transplanted in second remission or in first relapse. The best results with allogeneic transplantation are achieved when applied during first remission, with disease-free survival rates averaging 55–60%. Meta-analyses of studies comparing matched related donor transplantation to chemotherapy for adult AML patients age <60 years show a survival advantage with transplantation. This advantage is greatest for those with unfavorable-risk AML and lost in those with favorable-risk disease. The role of autologous transplantation in the treatment of AML is less well defined. The rates of disease recurrence with autologous transplantation are higher than those seen after allogeneic transplantation, and cure rates are somewhat less.

Similar to patients with AML, adults with acute lymphocytic leukemia who do not achieve a complete response to induction chemotherapy can be cured in 15–20% of cases with immediate transplantation. Cure rates improve to 30–50% in second remission, and therefore transplantation can be recommended for adults who have persistent disease after induction chemotherapy or who have subsequently relapsed. Transplantation in first remission results in cure rates about 55%. Transplantation appears to offer a clear advantage over chemotherapy for patients with high-risk disease, such as those with Philadelphia chromosome–positive disease. Debate continues about whether adults with standard-risk disease should be transplanted in first remission or whether

transplantation should be reserved until relapse. Autologous transplantation is associated with a higher relapse rate but a somewhat lower risk of nonrelapse mortality when compared to allogeneic transplantation. There is no obvious role of autologous transplantation for ALL in first remission, and for second-remission patients, most experts recommend use of allogeneic stem cells if an appropriate donor is available.

CHRONIC LEUKEMIA

Allogeneic hematopoietic cell transplantation is the only therapy shown to cure a substantial portion of patients with chronic myeloid leukemia (CML). Five-year disease-free survival rates are 15–20% for patients transplanted for blast crisis, 25–50% for accelerated-phase patients, and 60–70% for chronic-phase patients, with cure rates as high as 80% at selected centers. However, with the availability of imatinib mesylate, a remarkably effective, relatively nontoxic oral agent, most physicians favor reserving transplantation for those who fail to achieve a complete cytogenetic response with imatinib, relapse after an initial response, or are intolerant of the drug (Chap. 109).

Allogeneic transplantation using high-dose preparative regimen was rarely used for chronic lymphocytic leukemia (CLL), in large part because of the chronic nature of the disease and because of the age profile of patients. In those cases where it was studied, complete remissions were achieved in the majority of patients, with disease-free survival rates of ~50% at 3 years, despite the advanced stage of the disease at the time of transplant. The marked antitumor effects have resulted in the increased use and study of allogeneic transplantation using reduced-intensity conditioning for the treatment of CLL.

MYELODYSPLASIA

Between 40 and 50% of patients with myelodysplasia appear to be cured with allogeneic transplantation. Results are better among younger patients and those with less advanced disease. However, some patients with myelodysplasia can live for extended periods without intervention, and so transplantation is generally recommended only for patients with disease categorized as intermediate risk I or greater according to the International Prognostic Scoring System (Chap. 107).

LYMPHOMA

Patients with disseminated intermediate- or high-grade non-Hodgkin's lymphoma who have not been cured by first-line chemotherapy and are transplanted in first relapse or second remission can still be cured in 40–50% of cases. This represents a clear advantage over results obtained with conventional-dose salvage chemotherapy. It is unsettled whether patients with high-risk disease benefit from transplantation in first remission. Most experts favor the use of autologous rather than allogeneic transplantation for patients with intermediate- or high-grade non-Hodgkin's lymphoma, because fewer complications occur with this approach and survival appears equivalent. For patients with recurrent disseminated indolent non-Hodgkin's lymphoma, autologous transplantation results in high response rates and improved progression-free survival compared to salvage chemotherapy. However, late relapses are seen after transplantation. The role of autologous transplantation in the initial treatment of patients is under study. Reduced-intensity conditioning regimens followed by allogeneic transplantation result in high response rates in patients with indolent lymphomas, but the exact role of this approach remains to be defined.

The role of transplantation in Hodgkin's disease is similar to that in intermediate- and high-grade non-Hodgkin's lymphoma. With transplantation, 5-year disease-free survival is 20–30% in patients who never achieve a first remission with standard chemotherapy and up to 70% for those transplanted in second remission. Transplantation has no defined role in first remission in Hodgkin's disease.

MYELOMA

Patients with myeloma who have progressed on first-line therapy can sometimes benefit from allogeneic or autologous transplantation. Autologous transplantation has been studied as part of the initial therapy of patients, and both disease-free survival and overall survival were improved with this approach in randomized trials. The use of autologous transplantation followed by nonmyeloablative allogeneic transplantation is the subject of ongoing research.

SOLID TUMORS

Among women with metastatic breast cancer, 15–20% disease-free survival rates at 3 years have been reported, with better results seen in younger patients who have responded completely to standard-dose therapy before undergoing transplantation. Randomized trials have not shown superior survival for patients treated for metastatic disease with high-dose chemotherapy plus stem cell support. Randomized trials evaluating transplantation as treatment for primary breast cancer have yielded mixed results. No role for autologous transplantation has been established in the treatment of breast cancer.

Patients with testicular cancer who have failed first-line chemotherapy have been treated with autologous transplantation; ~10–20% of such patients apparently have been cured with this approach.

The use of high-dose chemotherapy with autologous stem cell support is being studied for several

other solid tumors, including neuroblastoma and pediatric sarcomas. As in most other settings, the best results have been obtained in patients with limited amounts of disease and where the remaining tumor remains sensitive to conventional-dose chemotherapy. Few randomized trials of transplantation in these diseases have been completed.

Partial and complete responses have been reported following nonmyeloablative allogeneic transplantation for some solid tumors, most notably renal cell cancers. The GVT effect, well documented in the treatment of hematologicmalignancies, may apply to selected solid tumors under certain circumstances.

POSTTRANSPLANT RELAPSE

Patients who relapse following autologous transplantation sometimes respond to further chemotherapy and may be candidates for possible allogeneic transplantation, particularly if the remission following the initial autologous transplant was long. Several options are available for patients who relapse following allogeneic transplantation. Of particular interest are the response rates seen with infusion of unirradiated donor lymphocytes. Complete responses in as many as 75% of patients with chronic myeloid leukemia, 40% in myelodysplasia, 25% in AML, and 15% in myeloma have been reported. Major complications of donor lymphocyte infusions include transient myelosuppression and the development of GVHD. These complications depend on the number of donor lymphocytes given and the schedule of infusions, with less GVHD seen with lower dose, fractionated schedules.

FURTHER READINGS

Appelbaum FR: Hematopoietic-cell transplantation at 50. N Engl J Med 357:1472, 2007[PMID: 17928594]

Gyurkocza B et al: Allogeneic hematopoietic cell transplantation: The state of the art. Expert Rev Hematol 3:285, 2010[PMID: 20871781]

Koreth J et al: Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: A systematic review and meta-analysis of prospective clinical trials. JAMA 301(22):2349, 2009[PMID: 19509382]

Lee SJ et al: High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood 110:4576, 2007[PMID: 17785583]

Tomblyn M et al: Guidelines for preventing infectious complications among hematopoietic cell transplantation recipients: A global perspective. Biol Blood Marrow Transplant 15:1143, 2009[PMID: 19747629]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF PLATELETS AND VESSEL WALL: INTRODUCTION

Hemostasis is a dynamic process in which the platelet and the blood vessel wall play key roles. Platelets become activated upon adhesion to von Willebrand factor (VWF) and collagen in the exposed subendothelium after injury. Platelet activation is also mediated through shear forces imposed by blood flow itself, particularly in areas where the vessel wall is diseased, and is also affected by the inflammatory state of the endothelium. The activated platelet surface provides the major physiologic site for coagulation factor activation, which results in further platelet activation and fibrin formation. Genetic and acquired influences on the platelet and vessel wall, as well as on the coagulation and fibrinolytic systems, determine whether normal hemostasis, or bleeding or clotting symptoms, will result.

The Platelet

Platelets are released from the megakaryocyte, likely under the influence of flow in the capillary sinuses. The normal blood platelet count is 150,000–450,000/μL. The major regulator of platelet production is the hormone thrombopoietin (TPO), which is synthesized in the liver. Synthesis is increased with inflammation and specifically by interleukin 6. TPO binds to its receptor on platelets and megakaryocytes, by which it is removed from the circulation. Thus a reduction in platelet and megakaryocyte mass increases the level of TPO, which then stimulates platelet production. Platelets circulate with an average life span of 7 to 10 days. Approximately one-third of the platelets reside in the spleen, and this number increases in proportion to splenic size, although the platelet count rarely decreases to <40,000/μL as the spleen enlarges. Platelets are physiologically very active, but are anucleate, and thus have limited capacity to synthesize new proteins.

Normal vascular endothelium contributes to preventing thrombosis by inhibiting platelet function (Chap. 58). When vascular endothelium is injured, these inhibitory effects are overcome, and platelets adhere to the exposed intimal surface primarily through VWF, a large multimeric protein present in both plasma and in the extracellular matrix of the subendothelial vessel wall. Platelet adhesion results in the generation of intracellular signals that lead to activation of the platelet glycoprotein (Gp) IIb/IIIa ($\alpha_{IIb}\beta_3$) receptor and resultant platelet aggregation.

Activated platelets undergo release of their granule contents, which include nucleotides, adhesive proteins, growth factors, and procoagulants that serve to promote platelet aggregation and blood clot formation and influence the environment of the forming clot. During platelet aggregation, additional platelets are recruited to the site of injury, leading to the formation of an occlusive platelet thrombus. The platelet plug is stabilized by the fibrin mesh that develops simultaneously as the product of the coagulation cascade.

The Vessel Wall

Endothelial cells line the surface of the entire circulatory tree, totaling 1–6 x 10^{13} cells, enough to cover a surface area equivalent to about six tennis courts. The endothelium is physiologically active, controlling vascular permeability, flow of biologically active molecules and nutrients, blood cell interactions with the vessel wall, the inflammatory response, and angiogenesis.

The endothelium normally presents an antithrombotic surface (Chap. 58) but rapidly becomes prothrombotic when stimulated, which promotes coagulation, inhibits fibrinolysis, and activates platelets. In many cases, endothelium-derived vasodilators are also platelet inhibitors (e.g., nitric oxide) and, conversely, endothelium-derived vasoconstrictors (e.g., endothelin) can also be platelet activators. The net effect of vasodilation and inhibition of platelet function is to promote blood fluidity, whereas the net effect of vasoconstriction and platelet activation is to promote thrombosis. Thus, blood fluidity and hemostasis is regulated by the balance of antithrombotic/prothrombotic and vasodilatory/vasoconstrictor properties of endothelial cells.

DISORDERS OF PLATELETS

Thrombocytopenia

Thrombocytopenia results from one or more of three processes: (1) decreased bone marrow production; (2) sequestration, usually in an enlarged spleen; and/or (3) increased platelet destruction. Disorders of production may be either inherited or acquired. In evaluating a patient with thrombocytopenia, a key step is to review the peripheral blood smear and to first rule out

"pseudothrombocytopenia," particularly in a patient without an apparent cause for the thrombocytopenia. Pseudothrombocytopenia (Fig. 115-1B) is an in vitro artifact resulting from platelet agglutination via antibodies (usually IgG, but also IgM and IgA) when the calcium content is decreased by blood collection in ethylenediamine tetraacetic (EDTA) [the anticoagulant present in tubes (purple top) used to collect blood for complete blood counts (CBCs)]. If a low platelet count is obtained in EDTA-anticoagulated blood, a blood smear should be evaluated and a platelet count determined in blood collected into sodium citrate (blue top tube) or heparin (green top tube), or a smear of freshly obtained unanticoagulated blood, such as from a finger stick, can be examined.

Figure 115-1

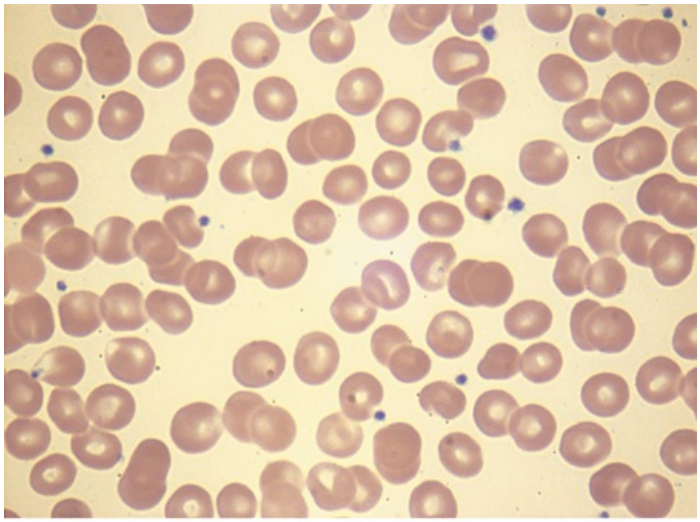

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

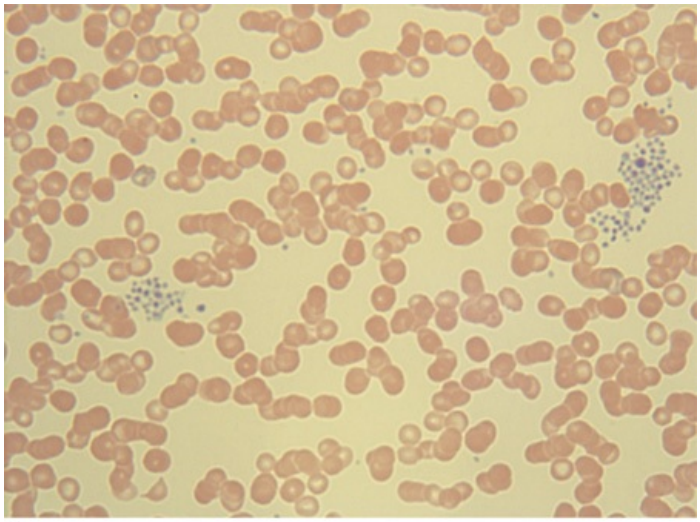

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

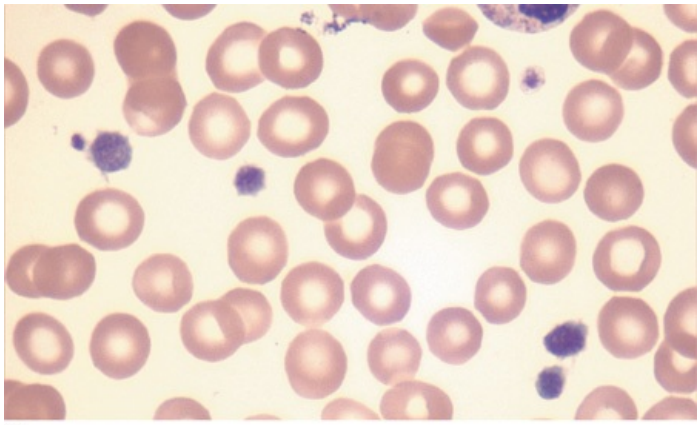

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

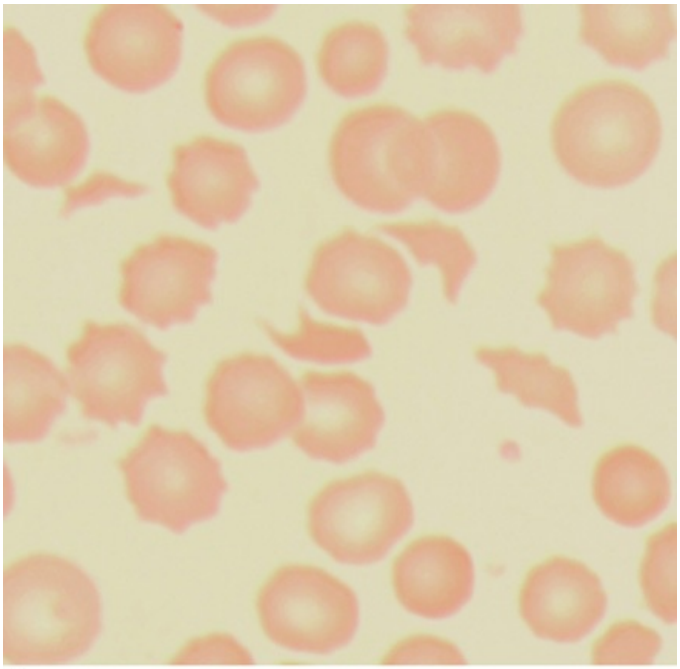

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Photomicrographs of peripheral blood smears: ***A.*** Normal peripheral blood. ***B.*** Platelet clumping in pseudothrombocytopenia. ***C.*** Abnormal large platelet in autosomal dominant macrothrombocytopenia. ***D.*** Schistocytes and decreased platelets in microangiopathic hemolytic anemia.

APPROACH TO THE PATIENT: THROMBOCYTOPENIA

The history and physical examination, results of the CBC, and review of the peripheral blood smear are all critical components in the initial evaluation of thrombocytopenic patients **(Fig. 115-2)**. The overall health of the patient and whether he or she is receiving drug treatment will influence the differential diagnosis. A healthy young adult with thrombocytopenia will have a much more limited differential diagnosis than an ill hospitalized patient who is receiving multiple medications. Except in unusual inherited disorders, decreased platelet production usually results from bone marrow disorders that also affect red blood cell (RBC) and/or white blood cell (WBC) production. Because myelodysplasia can present with isolated thrombocytopenia, the bone marrow should be examined in patients presenting with isolated thrombocytopenia who are older than 60 years of age. While inherited thrombocytopenia is rare, any prior platelet counts should be retrieved and a family history regarding thrombocytopenia obtained. A careful history of drug ingestion should be obtained, including nonprescription and herbal remedies, as drugs are the most common cause of thrombocytopenia.

Figure 115-2

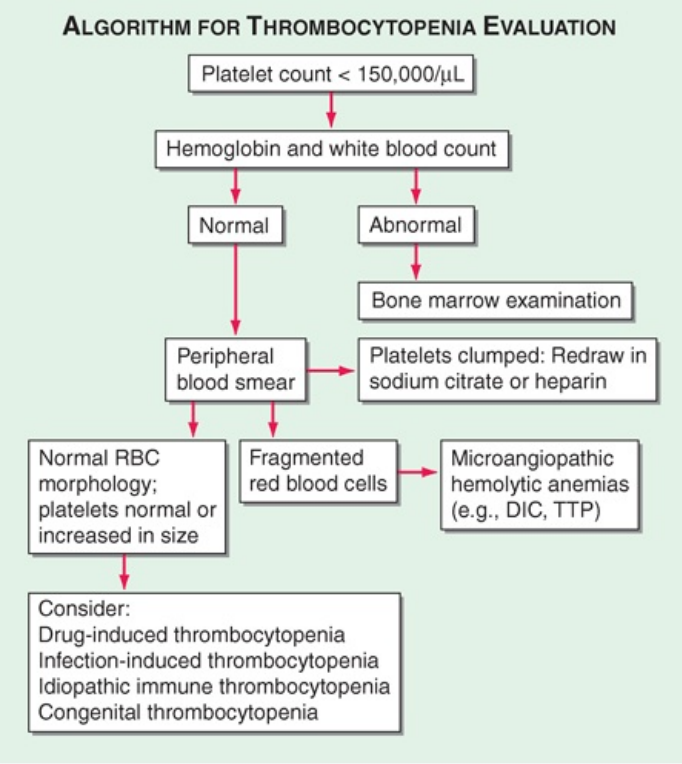

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for evaluating the thrombocytopenic patient.

The physical examination can document an enlarged spleen, evidence of chronic liver disease, and other underlying disorders. Mild to moderate splenomegaly may be difficult to appreciate in many individuals due to body habitus and/or obesity but can be easily assessed by abdominal ultrasound. A platelet count of approximately 5000–10,000 is required to maintain vascular integrity in the microcirculation. When the count is markedly decreased, petechiae first appear in areas of increased venous pressure, the ankles and feet in an ambulatory patient. Petechiae are pin-point, nonblanching hemorrhages and are usually a sign of a decreased platelet number and not platelet dysfunction. Wet purpura, blood blisters that form on the oral mucosa, are thought to denote an increased risk of life-threatening hemorrhage in the thrombocytopenic patient. Excessive bruising is seen in disorders of both platelet number and function.

INFECTION-INDUCED THROMBOCYTOPENIA

Many viral and bacterial infections result in thrombocytopenia and are the most common noniatrogenic cause of thrombocytopenia. This may or may not be associated with laboratory evidence of disseminated intravascular coagulation (DIC), which is most commonly seen in patients with systemic infections with gram-negative bacteria. Infections can affect both platelet production and platelet survival. In addition, immune mechanisms can be at work, as in infectious mononucleosis and early HIV infection. Late in HIV infection, pancytopenia and decreased and dysplastic platelet production is more common. Immune-mediated thrombocytopenia in children usually follows a viral infection and almost always resolves spontaneously. This association of infection with ITP is less clear in adults.

Bone marrow examination is often requested for evaluation of occult infections. A study evaluating the role of bone marrow examination in fever of unknown origin in HIV-infected patients found that for 86% of patients, the same diagnosis was established by less-invasive techniques, notably blood culture. In some instances, however, the diagnosis can be made earlier; thus, a bone marrow examination and culture is recommended when the diagnosis is needed urgently or when other, less invasive methods have been unsuccessful.

DRUG-INDUCED THROMBOCYTOPENIA

Many drugs have been associated with thrombocytopenia. A predictable decrease in platelet count occurs after treatment with many chemotherapeutic drugs due to bone marrow suppression (Chap. 85). Other, commonly used drugs that cause isolated thrombocytopenia are listed in Table 115–1, but all drugs should be suspect in a patient with thrombocytopenia without an apparent cause and should be stopped, or substituted, if possible. A helpful website, Platelets on the Internet (http://www.ouhsc.edu/platelets/index.html), lists drugs and supplements reported to have caused thrombocytopenia and the level of evidence supporting the association. Although not as well studied, herbal and over-the-counter preparations may also result in thrombocytopenia and should be

discontinued in patients who are thrombocytopenic.

Table 115–1 Drugs Reported as Definitely or Probably Causing Isolated Thrombocytopenia*

Abciximab	Ibuprofen
Acetaminophen	Iopanoic acid
Aminoglutethimide	Levamisole
Aminosalicylic acid	Linezolid
Amiodarone	Meclofenamate
Amphotericin B	Methicillin
Ampicillin	Methyldopa
Carbamazepine	Nalidixic acid
Chlorpropamide	Naproxen
Danazol	Oxyphenbutazone
Captopril	Phenytoin
Cimetidine	Piperacillin
Diatrizoate meglumine (Hypaque Meglumine®)	Procainamide
Diclofenac	Quinine
Digoxin	Quinidine
Dipyridamole	Rifampin
Eptifibatide	Simvastatin
Ethambutol	Sulfa-containing drugs
Famotidine	Tamoxifen
Fluconazole	Tirofiban
Furosemide	Trimethoprim/sulfamethoxazole
Glyburide	Valproic acid
Gold	Vancomycin
Hydrochlorothiazide	
Imipenem/Cilastatin	

*Reported in ≥2 patients

Source: Data from http://www.ouhsc.edu/platelets/index.html.

Classic drug-dependent antibodies are antibodies that react with specific platelet surface antigens, and result in thrombocytopenia only when the drug is present. Many drugs are capable of inducing these antibodies, but for some reason they are more common with quinine and sulfonamides. Drug-dependent antibody binding can be demonstrated by laboratory assays, showing antibody binding in the presence of, but not without, the drug present in the assay. The thrombocytopenia typically occurs after a period of initial exposure (median length 21 days), or upon reexposure, and usually resolves in 7–10 days after drug withdrawal. The thrombocytopenia caused by the platelet GpIIbIIIa inhibitory drugs, such as abciximab, differs in that it may occur within 24 h of initial exposure. This appears to be due to the presence of naturally occurring antibodies that cross-react with the drug bound to the platelet.

HEPARIN-INDUCED THROMBOCYTOPENIA

Drug-induced thrombocytopenia due to heparin differs from that seen with other drugs in two major ways. (1) The thrombocytopenia is not usually severe, with nadir counts rarely <20,000/μL. (2) Heparin-induced thrombocytopenia (HIT) is not associated with bleeding and, in fact, markedly increases the risk of thrombosis. HIT results from antibody formation to a complex of the platelet-specific protein platelet factor 4 (PF4) and heparin. The antiheparin/PF4 antibody can activate platelets through the FcγRIIa receptor and also activate monocytes and endothelial cells. Many patients exposed to heparin develop antibodies to heparin/PF4, but do not appear to have adverse consequences. A fraction of those who develop antibodies will develop HIT, and a portion of those (up to 50%) will develop thrombosis (HITT).

HIT can occur after exposure to low-molecular-weight heparin (LMWH) as well as unfractionated heparin (UFH), although it is about 10 times more common with the latter. Most patients develop HIT after exposure to heparin for 5–14 days (Fig. 115-3). It occurs before 5 days in those who were exposed to heparin in the prior few weeks or months (<~100 days) and have circulating antiheparin/PF4 antibodies. Rarely, thrombocytopenia and thrombosis begin several days after all heparin has been stopped (termed *delayed-onset HIT*). The 4 *T*'s have been recommended to be

used in a diagnostic algorithm for HIT: *t*hrombocytopenia, *t*iming of platelet count drop, *thrombosis* and other sequelae such as localized skin reactions, and o*t*her causes of thrombocytopenia not evident. A new scoring model based on broad expert opinion [the HIT Expert Probability (HEP) Score] has improved operating characteristics and should provide better utility as a scoring system.

Figure 115-3

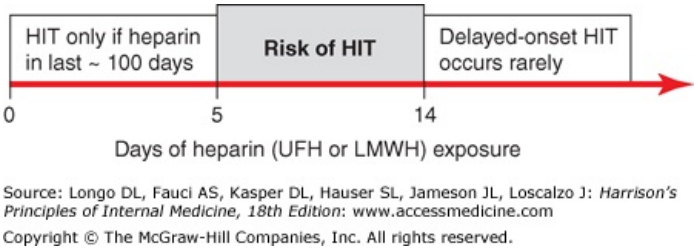

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Time course of heparin-induced thrombocytopenia (HIT) development after heparin exposure. The timing of development after heparin exposure is a critical factor in determining the likelihood of HIT in a patient. HIT occurs early after heparin exposure in the presence of preexisting heparin/platelet factor 4 (PF4) antibodies, which disappear from circulation by ~100 days following a prior exposure. Rarely, HIT may occur later after heparin exposure (termed delayed-onset HIT). In this setting, heparin/PF4 antibody testing is usually markedly positive. HIT can occur after exposure to either unfractionated (UFH) or low-molecular-weight heparin (LMWH).

Laboratory Testing for HIT

HIT (anti-heparin/PF4) antibodies can be detected using two types of assays. The most widely available is an enzyme-linked immunoassay (ELISA) with PF4/polyanion complex as the antigen. Since many patients develop antibodies but do not develop clinical HIT, the test has a low specificity for the diagnosis of HIT. This is especially true in patients who have undergone cardiopulmonary bypass surgery, where approximately 50% of patients develop these antibodies postoperatively. IgG-specific ELISAs increase specificity but may decrease sensitivity. The other assay is a platelet activation assay, which measures the ability of the patient's serum to activate platelets in the presence of heparin in a concentration-dependent manner. This test has lower sensitivity but higher specificity than the ELISA. However, HIT remains a clinical diagnosis.

Treatment: Heparin-Induced Thrombocytopenia

Early recognition is key in treatment of HIT, with prompt discontinuation of heparin and use of alternative anticoagulants. Thrombosis is a common complication of HIT, even after heparin discontinuation, and can occur in both the venous and arterial systems. Patients with higher anti-heparin/PF4 antibody titers have a higher risk of thrombosis. In patients diagnosed with HIT, imaging studies to evaluate the patient for thrombosis (at least lower extremity duplex Dopplers) are recommended. Patients requiring anticoagulation should be switched from heparin to an alternative anticoagulant. The direct thrombin inhibitors (DTIs) argatroban and lepirudin are effective in HITT. The DTI bivalirudin and the antithrombin-binding pentasaccharide fondaparinux are also effective but not yet approved by the U.S. Food and Drug Administration (FDA) for this indication. Danaparoid, a mixture of glycosoaminoglycans with anti-Xa activity, has been used extensively for the treatment of HITT; it is no longer available in the United States but is in other countries. HIT antibodies cross-react with LMWH, and these preparations should not be used in the treatment of HIT.

Because of the high rate of thrombosis in patients with HIT, anticoagulation should be strongly considered, even in the absence of thrombosis. In patients with thrombosis, patients can be transitioned to warfarin, with treatment usually for 3–6 months. In patients without thrombosis, the duration of anticoagulation needed is undefined. An increased risk of thrombosis is present for at least 1 month after diagnosis; however, most thromboses occur early, and whether thrombosis occurs later if the patient is initially anticoagulated is unknown. Options include continuing anticoagulation until a few days after platelet recovery or for one month. Introduction of warfarin alone in the setting of HIT or HITT may precipitate thrombosis, particularly venous gangrene, presumably due to clotting activation and severely reduced levels of proteins C and S. Warfarin therapy, if started, should be overlapped with a DTI or fonda parinux, and started after resolution of the thrombocytopenia and lessening of the prothrombotic state.

IMMUNE THROMBOCYTOPENIC PURPURA (ITP)

Immune thrombocytopenic purpura (ITP; also termed *idiopathic thrombocytopenic purpura*) is an acquired disorder in which there is immune-mediated destruction of platelets and possibly inhibition of platelet release from the megakaryocyte. In children, it is usually an acute disease, most commonly following an infection, and with a self-limited course. In adults, it usually runs a more chronic course. ITP is termed *secondary* if it is associated with an underlying disorder; autoimmune disorders, particularly systemic lupus erythematosus (SLE), and infections, such as HIV and hepatitis C, are common causes. The association of ITP with *Helicobacter pylori* infection is unclear.

ITP is characterized by mucocutaneous bleeding and a low, often very low, platelet count, with an otherwise normal peripheral blood cells and smear. Patients usually present either with ecchymoses and petechiae, or with thrombocytopenia incidentally found on a routine CBC. Mucocutaneous bleeding, such as oral mucosa, gastrointestinal, or heavy menstrual bleeding, may be present. Rarely, life-threatening, including central nervous system, bleeding can occur. Wet purpura (blood blisters in the mouth) and retinal hemorrhages may herald life-threatening bleeding.

Laboratory Testing in ITP

Laboratory testing for antibodies (serologic testing) is usually not helpful due to the low sensitivity and specificity of the current tests. Bone marrow examination can be reserved for older adults (usually >60 years) or those who have other signs or laboratory abnormalities not explained by ITP, or in patients who do not respond to initial therapy. The peripheral blood smear may show large platelets, with otherwise normal morphology. Depending on the bleeding history, iron deficiency anemia may be present.

Laboratory testing is performed to evaluate for secondary causes of ITP and should include testing for HIV infection and hepatitis C (and other infections if indicated); serologic testing for SLE, serum protein electrophoresis, and immunoglobulin levels to potentially detect hypogammaglobulinemia; selective testing for IgA deficiency or monoclonal gammopathies and, if anemia is present, direct antiglobulin testing (Coombs test) to rule out combined autoimmune hemolytic anemia with ITP (Evans syndrome).

Treatment: Immune Thrombocytopenic Purpura

The treatment of ITP utilizes drugs that decrease reticuloendothelial uptake of the antibody-bound platelet, decrease antibody production, and/or increase platelet production. The diagnosis of ITP does not necessarily mean that treatment must be instituted. Patients with platelet counts greater than 30,000/μL appear not to have increased mortality related to the thrombocytopenia.

Initial treatment in patients without significant bleeding symptoms, severe thrombocytopenia (<5000/μL), or signs of impending bleeding (such as retinal hemorrhage or large oral mucosal hemorrhages) can be instituted as an outpatient using single agents. Traditionally, this has been prednisone at 1 mg/kg, although Rh_0(D) immune globulin therapy (WinRho SDF), at 50–75 μg/kg, is also being used in this setting. Rh_0(D) immune globulin must be used only in Rh-positive patients as the mechanism of action is production of limited hemolysis, with antibody-coated cells "saturating" the Fc receptors, inhibiting Fc receptor function. Monitoring patients for 8 h postinfusion is now advised by the FDA because of the rare complication of severe intravascular hemolysis. Intravenous gamma globulin (IVIgG), which is pooled, primarily IgG antibodies, also blocks the Fc receptor system, but appears to work primarily through different mechanism(s). IVIgG has more efficacy than anti-Rh_0(D) in postsplenectomized patients. IVIgG is dosed at 2 g/kg total, given in divided doses over 2–5 days. Side effects are usually related to the volume of infusion and infrequently include aseptic meningitis and renal failure. All immunoglobulin preparations are derived from human plasma and undergo treatment for viral inactivation.

For patients with severe ITP and/or symptoms of bleeding, hospital admission and combined-modality therapy is given using high-dose glucocorticoids with IVIgG or anti-Rh_0D therapy, and, as needed, additional immunosuppressive agents. Rituximab, an anti-CD20 (B cell) antibody, has shown efficacy in the treatment of refractory ITP.

Splenectomy has been used for treatment of patients who relapse after glucocorticoids are tapered. Splenectomy remains an important treatment option; however, more patients than previously thought will go into a remission over time. Observation, if the platelet count is high enough, or intermittent treatment with anti-Rh_0(D) or IVIgG may be a reasonable approach to see if the ITP will resolve. Vaccination against encapsulated organisms (especially pneumococcus, but also meningococcus and *Haemophilus influenzae*, depending on patient age and potential exposure) is recommended before splenectomy. Accessory spleen(s) are a very rare cause of relapse.

Thrombopoietin receptor agonists are now available for the treatment of ITP. This approach stems from the finding that many patients with ITP do not have increased TPO levels, as was previously hypothesized. TPO levels reflect megakaryocyte mass, which is usually normal in ITP. TPO levels are not increased in the setting of platelet destruction. Two agents, one administered subcutaneously (romiplostim) and another orally (eltrombopag), have shown response in many patients with refractory ITP. Roles for these agents in ITP treatment are not fully defined but given the chronicity of treatment, they are generally reserved for patients with refractory disease

INHERITED THROMBOCYTOPENIA

Thrombocytopenia is rarely inherited, either as an isolated finding or as part of a syndrome, and may be inherited in an autosomal dominant, autosomal recessive, or X-linked pattern. Many forms of autosomal dominant thrombocytopenia are now known to be associated with mutations in the nonmuscle myosin heavy chain *MYH9* gene. Interestingly, these include the May-Hegglin anomaly, and Sebastian, Epstein's, and Fechtner syndromes, all of which have distinct distinguishing features. A

common feature of these disorders is large platelets (Fig. 115-1C). Autosomal recessive disorders include congenital amegakaryocytic thrombocytopenia, thrombocytopenia with absent radii, and Bernard Soulier syndrome. The latter is primarily a functional platelet disorder due to absence of GPIb-IX-V, the VWF adhesion receptor. X-linked disorders include Wiskott-Aldrich syndrome and a dyshematopoietic syndrome resulting from a mutation in GATA-1, an important transcriptional regulator of hematopoiesis.

Thrombotic Thrombocytopenic Purpura and Hemolytic Uremic Syndrome

Thrombotic thrombocytopenic microangiopathies are a group of disorders characterized by thrombocytopenia, a microangiopathic hemolytic anemia evident by fragmented RBCs (Fig. 115-1D) and laboratory evidence of hemolysis, and microvascular thrombosis. They include thrombotic thrombocytopenic purpura (TTP) and hemolytic uremic syndrome (HUS), as well as syndromes complicating bone marrow transplantation, certain medications and infections, pregnancy and vasculitis. In DIC, while thrombocytopenia and microangiopathy are seen, a coagulopathy predominates, with consumption of clotting factors and fibrinogen resulting in an elevated prothrombin time (PT), and often activated partial thromboplastin time (aPTT). The PT and aPTT are characteristically normal in TTP or HUS.

THROMBOTIC THROMBOCYTOPENIC PURPURA (TTP)

TTP and HUS were previously considered overlap syndromes. However, in the past few years the pathophysiology of inherited and idiopathic TTP has become better understood and clearly differs from HUS. TTP was first described in 1924 by Eli Moschcowitz and characterized by a pentad of findings that include microangiopathic hemolytic anemia, thrombocytopenia, renal failure, neurologic findings, and fever. The full-blown syndrome is less commonly seen now, probably due to earlier diagnosis. The introduction of treatment with plasma exchange markedly improved the prognosis in patients, with a decrease in mortality from 85–100% to 10–30%.

The pathogenesis of inherited (Upshaw-Schulman syndrome) and idiopathic TTP is related to a deficiency of, or antibodies to, the metalloprotease ADAMTS13, that cleaves VWF. VWF is normally secreted as ultra-large multimers, which are then cleaved by ADAMTS13. The persistence of ultra-large VWF molecules is thought to contribute to pathogenic platelet adhesion and aggregation **(Fig. 115-4)**. This defect alone, however, is not sufficient to result in TTP as individuals with a congenital absence of ADAMTS13 develop TTP only episodically. Additional provocative factors have not been defined. The level of ADAMTS13 activity, as well as antibodies, can now be detected by laboratory assays. However, assays with sufficient sensitivity and specificity to direct clinical management have yet to be clearly defined.

Figure 115-4

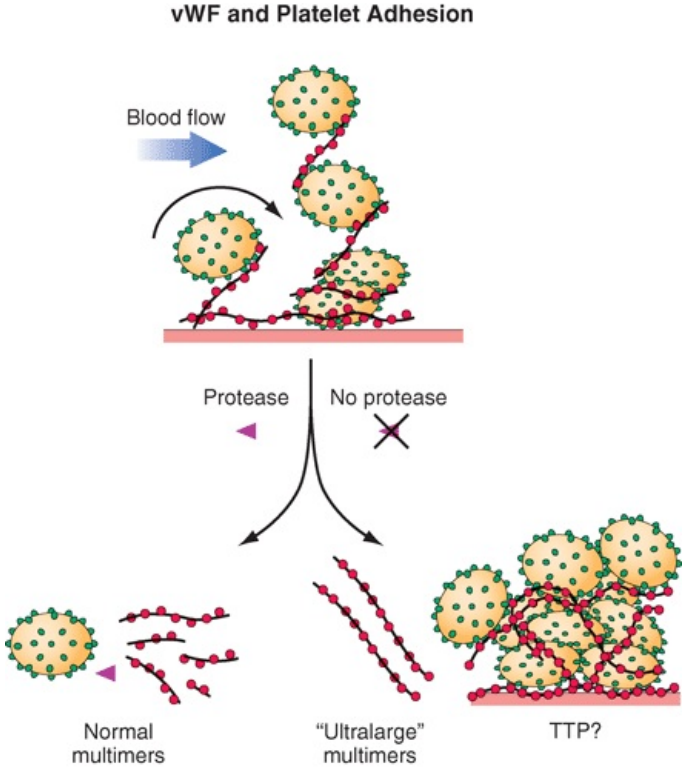

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis of thrombotic thrombocytopenic purpura (TTP). Normally the ultra-high-molecular-weight multimers of von Willebrand factor (VWF) produced by the endothelial cells are processed into smaller multimers by a plasma metalloproteinase called ADAMTS13. In TTP the activity of the protease is inhibited, and the ultra-high-molecular-weight multimers of VWF initiate platelet aggregation and thrombosis.

Idiopathic TTP appears to be more common in women than in men. No geographic or racial distribution has been defined. TTP is more common in patients with HIV infection and in pregnant women. TTP in pregnancy is not clearly related to ADAMTS13. Medication-related microangiopathic hemolytic anemia may be secondary to antibody formation (ticlopidine and possibly clopidogrel) or direct endothelial toxicity (cyclosporine, mitomycin C, tacrolimus, quinine), although this is not always so clear, and fear of withholding treatment, as well as lack of other treatment alternatives, results in broad application of plasma exchange. However, withdrawal, or reduction in dose, of endothelial toxic agents usually decreases the microangiopathy.

Treatment: Thrombotic Thrombocytopenic Purpura

TTP is a devastating disease if not diagnosed and treated promptly. In patients presenting with new thrombocytopenia, with or without evidence of renal insufficiency and other elements of classic TTP, laboratory data should be obtained to rule out DIC and to evaluate for evidence of microangiopathic hemolytic anemia. Findings to support the TTP diagnosis include an increased lactate dehydrogenase and indirect bilirubin, decreased haptoglobin, and increased reticulocyte count, with a negative direct antiglobulin test. The peripheral smear should be examined for evidence of schistocytes (Fig. 115-1D). Polychromasia is usually also present due to the increased number of young red blood cells, and nucleated RBCs are often present, which is thought to be due to infarction in the microcirculatory system of the bone marrow.

Plasma exchange remains the mainstay of treatment of TTP. ADAMTS 13 antibody-mediated TTP (idiopathic TTP) appears to respond best to plasma exchange. Plasma exchange is continued until the platelet count is normal and signs of hemolysis are resolved for at least 2 days. While never evaluated in clinical trial, the use of glucocorticoids seems a reasonable approach, but should only be used as an adjunct to plasma exchange. Additionally, other immunomodulatory therapies have been reported to be successful in refractory or relapsing TTP, including rituximab, vincristine, cyclophosphamide, and splenectomy. The role of rituximab in the treatment of this disorder needs to be defined. A significant relapse rate is noted, 25–45% within 30 days of initial "remission," and 12–40% with late relapses. Relapses may be more frequent in patients with severe ADAMTS 13 deficiency at presentation.

HEMOLYTIC UREMIC SYNDROME

HUS is a syndrome characterized by acute renal failure, microangiopathic hemolytic anemia, and thrombocytopenia. It is seen predominantly in children and in most cases is preceded by an episode of diarrhea, often hemorrhagic in nature. *Escherichia coli* O157:H7 is the most frequent, although not only, etiologic serotype. HUS not associated with diarrhea (termed *DHUS*) is more heterogeneous in presentation and course. Some children who develop DHUS have been found to have mutations in genes encoding factor H, a soluble complement regulator, and membrane cofactor protein that is mainly expressed in the kidney.

Treatment: Hemolytic Uremic Syndrome

Treatment of HUS is primarily supportive. In D^+HUS, many (~40%) children require at least some period of support with dialysis; however, the overall mortality is <5%. In D^-HUS, the mortality is higher, approximately 26%. Plasma infusion or plasma exchange has not been shown to alter the overall course. ADAMTS13 levels are generally reported to be normal in HUS, although occasionally they have been reported to be decreased. As ADAMTS13 assays improve, they may help in defining a subset that better fit a TTP diagnosis, and may respond to plasma exchange.

Thromobcytosis

Thrombocytosis is almost always due either to (1) iron deficiency; (2) inflammation, cancer, or infection (reactive thrombocytosis); or (3) an underlying myeloproliferative process [essential thrombocythemia or polycythemia vera) (Chap. 108)] or, rarely, the 5q- myelodysplastic process (Chap. 107). Patients presenting with an elevated platelet count should be evaluated for underlying inflammation or malignancy, and iron deficiency should be ruled out. Thrombocytosis in response to acute or chronic inflammation has not been associated with an increased thrombotic risk. In fact, patients with markedly elevated platelet counts (>1.5 million), usually seen in the setting of a myeloproliferative disorder, have an increased risk of bleeding. This appears to be due, at least in part, to acquired von Willebrand disease (VWD) due to platelet-VWF adhesion and removal.

Qualitative Disorders of Platelet Function

INHERITED DISORDERS OF PLATELET FUNCTION

Inherited platelet function disorders are thought to be relatively rare, although the prevalence of mild disorders of platelet function is unclear, in part because our testing for such disorders is suboptimal.

Rare qualitative disorders include the autosomal recessive disorders Glanzmann's thrombasthenia (absence of the platelet GpIIbIIIa receptor) and Bernard Soulier syndrome (absence of the platelet GpIb-IX-V receptor). Both are inherited in an autosomal recessive fashion and present with bleeding symptoms in childhood.

Platelet storage pool disorder (SPD) is the classic autosomal dominant qualitative platelet disorder. This results from abnormalities of platelet granule formation. It is also seen as a part of inherited disorders of granule formation, such as Hermansky-Pudlak syndrome. Bleeding symptoms in SPD are variable, but often are mild. The most common inherited disorders of platelet function are disorders that prevent normal secretion of granule content. Few of the abnormalities have been dissected at the molecular level but these are likely due to multiple abnormalities. They are usually described as *secretion defects*. Bleeding symptoms are usually mild in nature.

Treatment: Inherited Disorders of Platelet Dysfunction

Bleeding symptoms or prevention of bleeding in patients with severe platelet dysfunction frequently requires platelet transfusion. Care is taken to limit the risk of alloimmunization by limiting exposure, using apheresis leuko-depleted platelets for transfusion. Platelet disorders associated with milder bleeding symptoms frequently respond to desmopressin [1-deamino-8-d-arginine vasopressin (DDAVP)]. DDAVP increases plasma VWF and FVIII levels; it may also have a direct effect on platelet function. Particularly for mucosal bleeding symptoms, antifibrinolytic therapy (epsilon-aminocaproic acid or tranexamic acid) is used alone or in conjunction with DDAVP or platelet therapy.

ACQUIRED DISORDERS OF PLATELET FUNCTION

Acquired platelet dysfunction is common, usually due to medications, either intentionally as with antiplatelet therapy, or unintentionally as with high-dose penicillins. Acquired platelet dysfunction occurs in uremia. This is likely multifactorial but the resultant effect is defective adhesion and activation. The platelet defect is improved most by dialysis but may also be improved by increasing the hematocrit to 27–32%, giving DDAVP (0.3 μg/kg), or use of conjugated estrogens. Platelet dysfunction also occurs with cardiopulmonary bypass due to the effect of the artificial circuit on platelets, and bleeding symptoms respond to platelet transfusion. Platelet dysfunction seen with underlying hematologic disorders can result from nonspecific interference by circulating paraproteins or intrinsic platelet defects in myeloproliferative and myelodysplastic syndromes.

von Willebrand Disease

VWD vvVis the most common inherited bleeding disorder. Estimates from laboratory data suggest a prevalence of approximately 1%, but data based on symptomatic individuals suggest that it is closer to 0.1% of the population. VWF serves two roles: (1) as the major adhesion molecule that tethers the platelet to the exposed subendothelium; and (2) as the binding protein for FVIII, resulting in significant prolongation of the FVIII half-life in circulation. The platelet-adhesive function of VWF is critically dependent on the presence of large VWF multimers, while FVIII binding is not. Most of the symptoms of VWD are "platelet-like" except in more severe VWD when the FVIII is low enough to produce symptoms similar to those found in Factor VIII deficiency (hemophilia A).

VWD has been classified into three major types, with four subtypes of type 2 (Table 115–2; Fig. 115-5). By far the most common type of VWD is type 1 disease, with a parallel decrease in VWF protein, VWF function, and FVIII levels, accounting for at least 80% of cases. Patients have predominantly mucosal bleeding symptoms, although postoperative bleeding can also be seen. Bleeding symptoms are very uncommon in infancy and usually manifest later in childhood with excessive bruising and epistaxis. Since these symptoms occur commonly in childhood, the clinician should particularly note bruising at sites unlikely to be traumatized and/or prolonged epistaxis requiring medical attention. Menorrhagia is a common manifestation of VWD. Menstrual bleeding resulting in anemia should warrant an evaluation for VWD and, if negative, functional platelet disorders. Frequently, mild type 1 VWD first manifests with dental extractions, particularly wisdom tooth extraction, or tonsillectomy.

Table 115–2 Laboratory Diagnosis of von Willebrand Disease

Type	aPTT	VWF Antigen	VWF Activity	FVIII Activity	Multimer
1	Nl or ↑	↓	↓	↓	Normal distribution, decreased in quantity
2A	Nl or ↑	↓	↓ ↓	↓	Loss of high- and intermediate-MW multimers
2B[a]	Nl or ↑	↓	↓ ↓	↓	Loss of high-MW multimers
2M	Nl or ↑	↓	↓ ↓	↓	Normal distribution, decreased in quantity
2N	↑↑	Nl or ↓[b]	Nl or ↓[b]	↓↓	Normal distribution
3	↑↑	↓↓	↓↓	↓↓	Absent

[a]Usually also decreased platelet count.

[b]For type 2N, in the homozygous state, FVIII is very low; in the heterozygous state, only seen in conjunction with type 1 VWD.

Abbreviations: aPTT, activated partial thromboplastin time; F, Factor; MW, molecular weight; NI, normal; VWF, von Willebrand factor.

Figure 115-5

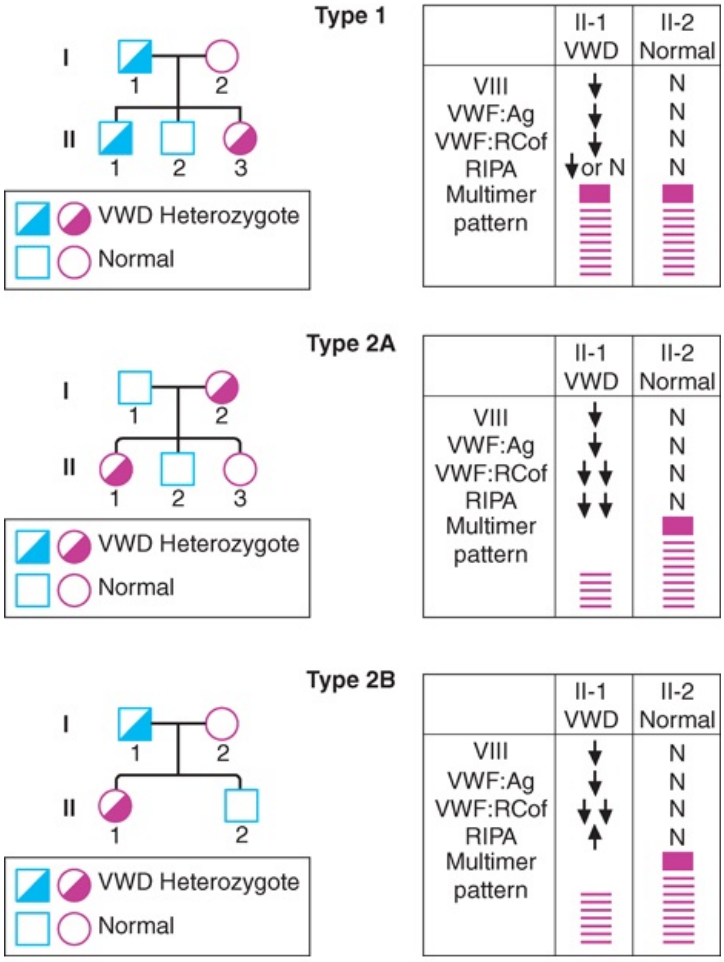

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pattern of inheritance and laboratory findings in von Willebrand disease. The assays of platelet function include a coagulation assay of Factor VIII bound and carried by von Willebrand factor (VWF), abbreviated as VIII; immunoassay of total VWF protein (VWF:Ag); bioassay of the ability of patient plasma to support ristocetin-induced agglutination of normal platelets (VWF:RCoF); and ristocetin-induced aggregation of patient platelets, abbreviated RIPA. The multimer pattern illustrates the protein bands present when plasma is electrophoresed in a polyacrylamide gel. The II-1 and II-2 columns refer to the phenotypes of the second-generation offspring.

Not all patients with low VWF levels have bleeding symptoms. Whether patients bleed or not will depend on the overall hemostatic balance they have inherited, along with environmental influences and the type of hemostatic challenges they experience. Although the inheritance of VWD is autosomal, many factors modulate both VWF levels and bleeding symptoms. These have not all been defined, but include blood type, thyroid hormone status, race, stress, exercise, and hormonal (both endogenous and exogenous) influences. Patients with type O blood have VWF protein levels of approximately one-half that of patients with AB blood type; and, in fact, the normal range for patients with type O blood overlaps that which has been considered diagnostic for VWD. A mildly decreased VWF level should perhaps be viewed more as a risk factor for bleeding than as an actual disease.

Patients with Type 2 VWD have functional defects; thus, the VWF antigen measurement is significantly higher than the test of function. For types 2A, 2B and 2M, VWF activity is decreased, measured as ristocetin cofactor or collagen-binding activity. In type 2A VWD, the impaired function is due either to increased susceptibility to cleavage by ADAMTS13, resulting in loss of intermediate and high-molecular-weight multimers, or to decreased secretion of these multimers by the cell. Type 2B VWD

results from gain of function mutations that result in increased spontaneous binding of VWF to platelets in circulation, with subsequent clearance of this complex by the reticuloendothelial system. The resulting VWF in the patients' plasma lacks the highest- molecular-weight multimers, and the platelet count is usually modestly reduced. Type 2M occurs as a consequence of a group of mutations that cause dysfunction of the molecule but do not affect multimer structure.

Type 2N VWD is due to mutations in VWF that preclude binding of FVIII. As FVIII is stabilized by binding to VWF, the FVIII in patients with type 2N VWD has a very short half-life, and the FVIII level is markedly decreased. This is sometimes termed *autosomal hemophilia*. Type 3 VWD, or severe VWD, describes patients with virtually no VWF protein and FVIII levels <10%. Patients experience mucosal and joint postoperative symptoms as well as other bleeding symptoms. Some patients with type 3 VWD, particularly those with large VWF gene deletions, are at risk of developing antibodies to infused VWF.

Acquired VWD is a rare disorder, most commonly seen in patients with underlying lymphoproliferative disorders, including monoclonal gammopathies of underdetermined significance (MGUS), multiple myeloma, and Waldenström's macroglobulinemia. It is seen most commonly in the setting of MGUS and should be suspected in patients, particularly elderly patients, with a new onset of severe mucosal bleeding symptoms. Laboratory evidence of acquired VWD is found in some patients with aortic valvular disease. Heyde's syndrome (aortic stenosis with gastrointestinal bleeding) is attributed to the presence of angiodysplasia of the gastrointestinal tract in patients with aortic stenosis. However, the shear stress on blood passing through the stenotic aortic valve appears to produce a change in VWF, making it susceptible to serum proteases. Consequently, large multimer forms are lost, leading to an acquired type 2 VWD, but return when the stenotic valve is replaced.

TREATMENT: VON WILLEBRAND DISEASE

The mainstay of treatment for type 1 VWD is 1-deamino-8-d-arginine vasopressin (DDAVP, or desmopressin), which results in release of VWF and FVIII from endothelial stores. DDAVP can be given intravenously or by a high concentration intranasal spray (1.5 mg/mL). The peak activity when given intravenously is approximately 30 minutes, while it is 2 h when given intranasally. The usual dose is 0.3 μg/kg intravenously or 2 squirts (1 in each nostril) for patients >50 kg (1 squirt for those <50 kg). It is recommended that patients with VWD be tested with DDAVP to assess their response before using it. In patients who respond well (increase in values of two- to fourfold), it can be used for procedures with minor to moderate risk of bleeding. Depending on the procedure, additional doses may be needed; it is usually given every 12–24 h. Less frequent dosing may result in less tachyphylaxis, which occurs when synthesis cannot compensate for the released stores. The major side effect of DDAVP is hyponatremia due to decreased free water clearance. This occurs most commonly in the very young and the very old, but fluid restriction should be advised for all patients for the 24 h following each dose.

Some patients with types 2A and 2M VWD respond to DDAVP such that it can be used for minor procedures. For the other subtypes, for type 3 disease, and for major procedures requiring longer periods of normal hemostasis, VWF replacement can be given. Virally inactivated VWF-containing factor concentrates are thought to be safer than cryoprecipitate as the replacement product.

Antifibrinolytic therapy using either ε-aminocaproic acid or tranexamic acid is an important therapy, either alone or in an adjunctive capacity, particularly for the prevention or treatment of mucosal bleeding. These agents are particularly useful in prophylaxis for dental procedures, with DDAVP for dental extractions and tonsillectomy, menorrhagia, and prostate procedures. It is contraindicated in the setting of upper urinary tract bleeding, due to the risk of ureteral obstruction.

DISORDERS OF THE VESSEL WALL

The vessel wall is an integral part of hemostasis, and separation of a fluid phase is artificial, particularly in disorders such as TTP or HIT that clearly involve the endothelium as well. Inflammation localized to the vessel wall, such as vasculitis, or inherited connective tissue disorders are abnormalities inherent to the vessel wall.

Metabolic and Inflammatory Disorders

Acute febrile illnesses may result in vascular damage. This can result from immune complexes containing viral antigens or the viruses themselves. Certain pathogens, such as the rickettsiae causing Rocky Mountain spotted fever, replicate in endothelial cells and damage them. Vascular purpura may occur in patients with polyclonal gammopathies but more commonly in those with monoclonal gammopathies, including Waldenström's macroglobulinemia, multiple myeloma, and cryoglobulinemia. Patients with mixed cryoglobulinemia develop a more extensive maculopapular rash due to immune complex–mediated damage to the vessel wall.

Patients with scurvy (vitamin C deficiency) develop painful episodes of perifollicular skin bleeding as well as more systemic bleeding symptoms. Vitamin C is needed to synthesize hydroxyproline, an essential constituent of collagen. Patients with Cushing's syndrome or on chronic glucocorticoid therapy develop skin bleeding and easy bruising due to atrophy of supporting connective tissue. A similar phenomenon is seen with aging, where following minor trauma, blood spreads superficially under the epidermis. This has been termed *senile purpura*. It is most common on skin that has been

previously damaged by sun exposure.

Henoch-Schönlein, or anaphylactoid, purpura is a distinct, self-limited type of vasculitis that occurs in children and young adults. Patients have an acute inflammatory reaction with IgA and complement components in capillaries, mesangial tissues, and small arterioles leading to increased vascular permeability and localized hemorrhage. The syndrome is often preceded by an upper respiratory infection, commonly with streptococcal pharyngitis, or is triggered by drug or food allergies. Patients develop a purpuric rash on the extensor surfaces of the arms and legs, usually accompanied by polyarthralgias or arthritis, abdominal pain, and hematuria from focal glomerulonephritis. All coagulation tests are normal but renal impairment may occur. Glucocorticoids can provide symptomatic relief but do not alter the course of the illness.

Inherited Disorders of the Vessel Wall

Patients with inherited disorders of the connective tissue matrix, such as Marfan's syndrome, Ehlers-Danlos syndrome, and pseudoxanthoma elasticum, frequently report easy bruising. Inherited vascular abnormalities can result in increased bleeding. This is notably seen in hereditary hemorrhagic telangiectasia (HHT, or Osler-Weber-Rendu disease), a disorder where abnormal telangiectatic capillaries result in frequent bleeding episodes, primarily from the nose and gastrointestinal tract. Arteriovenous malformation (AVM) in the lung, brain, and liver may also occur in HHT. The telangiectasia can often be visualized on the oral and nasal mucosa. Signs and symptoms develop over time. Expistaxis begins, on average, at the age of 12 and occurs in >95% of affected individuals by middle age. Two genes involved in the pathogenesis are *eng* (endoglin) on chromosome 9q33-34 (so-called HHT type 1), associated with pulmonary AVM in 40% of cases; and *alk1* (activin-receptor-like kinase 1*1*) on chromosome 12q13, associated with a much lower risk of pulmonary AVM.

ACKNOWLEDGMENT

Robert Handin, MD, contributed this chapter in the 16th edition and some materials from his chapter are included here.

FURTHER READINGS

Arepally GM, Ortel TL: Heparin-induced thrombocytopenia. Annu Rev Med 61:77, 2010[PMID: 20059332]

Aster RH et al: Drug-induced immune thrombocytopenia: Pathogenesis, diagnosis and management. J Thromb Haemost 7:911, 2009[PMID: 19344362]

Bussel JB: Traditional and new approaches to the management of immune thrombocytopenia: Issues of when and who to treat. Hematol Oncol Clin North Am 23:1329, 2009[PMID: 19932437]

Diagnosis, Evaluation and Management of von Willebrand Disease . NIH Publication #08-5832. National Heart, Lung and Blood Institute, 2007, www.nhlbi.nih.gov

Hayward CPM: Diagnostic approach to platelet function disorders. Trans Apher Sci 38:65, 2008[PMID: 18291723]

Israels SJ et al: Inherited disorders of platelet function and challenges to diagnosis of mucocutaneous bleeding. Haemophilia 16:152, 2010[PMID: 20590875]

Kremer Hovinga JA et al: Survival and relapse in patients with thrombotic thrombocytopenic purpura. Blood 115:1500, 2010

Rice TW, Wheeler AP: Coagulopathy in critically ill patients. Chest 136:1622, 2009[PMID: 19995764]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

COAGULATION DISORDERS: INTRODUCTION

Deficiencies of coagulation factors have been recognized for centuries. Patients with genetic deficiencies of plasma coagulation factors exhibit life-long recurrent bleeding episodes into joints, muscles, and closed spaces, either spontaneously or following an injury. The most common inherited factor deficiencies are the hemophilias, X-linked diseases caused by deficiency of factor (F) VIII (hemophilia A) or factor IX (FIX, hemophilia B). Rare congenital bleeding disorders due to deficiencies of other factors, including FII (prothrombin), FV, FVII, FX, FXI, FXIII, and fibrinogen are commonly inherited in an autosomal recessive manner **(Table 116–1)**. Advances in characterization of the molecular bases of clotting factor deficiencies have contributed to better understanding of the disease phenotypes and may eventually allow more targeted therapeutic approaches through the development of small molecules, recombinant proteins, or cell and gene-based therapies.

Table 116–1 Genetic and Laboratory Characteristics of Inherited Coagulation Disorders

Clotting Factor Deficiency	Inheritance	Prevalence in General Population	Laboratory Abnormality[a]			Minimum Hemostatic Levels	Treatment	Plasma Half-Life
			aPTT	PT	TT			
Fibrinogen	AR	1 in 1,000,000	+	+	+	100 mg/dL	Cryoprecipitate	2–4 d
Prothrombin	AR	1 in 2,000,000	+	+	–	20–30%	FFP/PCC	3–4 d
Factor V	AR	1 in 1,000,000	+/–	+/–	–	15–20%	FFP	36 h
Factor VII	AR	1 in 500,000	–	+	–	15–20%	FFP/PCC	4–6 h
Factor VIII	X-linked	1 in 5,000	+	–	–	30%	FVIII concentrates	8–12 h
Factor IX	X-linked	1 in 30,000	+	–	–	30%	FIX concentrates	18–24 h
Factor X	AR	1 in 1,000,000	+/–	+/–	–	15–20%	FFP/PCC	40–60 h
Factor XI	AR	1 in 1,000,000	+	–	–	15–20%	FFP	40–70 h
Factor XII	AR	ND	+	–	–	*b*	*b*	60 h
HK	AR	ND	+	–	–	*b*	*b*	150 h
Prekallikrein	AR	ND	+	–	–	*b*	*b*	35 h
Factor XIII	AR	1 in 2,000,000	–	–	+/–	2–5%	Cryoprecipi-tate	11–14 d

[a]Values within normal range (–) or prolonged (+).

[b]No risk for bleeding; treatment is not indicated.

Abbreviations: aPTT, activated partial thromboplastin time; AR, autosomal recessive; FFP, fresh-frozen plasma; HK, high-molecular-weight kininogen; ND, not determined; PCC, prothrombin complex concentrates; PT, prothrombin time; TT, thrombin time.

Commonly used tests of hemostasis provide the initial screening for clotting factor activity **(Fig. 116-1)**, and disease phenotype often correlates with the level of clotting activity. An isolated abnormal prothrombin time (PT) suggests FVII deficiency, whereas a prolonged activated partial thromboplastin time (aPTT) indicates most commonly hemophilia or FXI deficiency (Fig. 116-1). The prolongation of both PT and aPTT suggests deficiency of FV, FX, FII, or fibrinogen abnormalities. The addition of the missing factor at a range of doses to the subject's plasma will correct the abnormal clotting times; the result is expressed as a percentage of the activity observed in normal subjects.

Figure 116-1

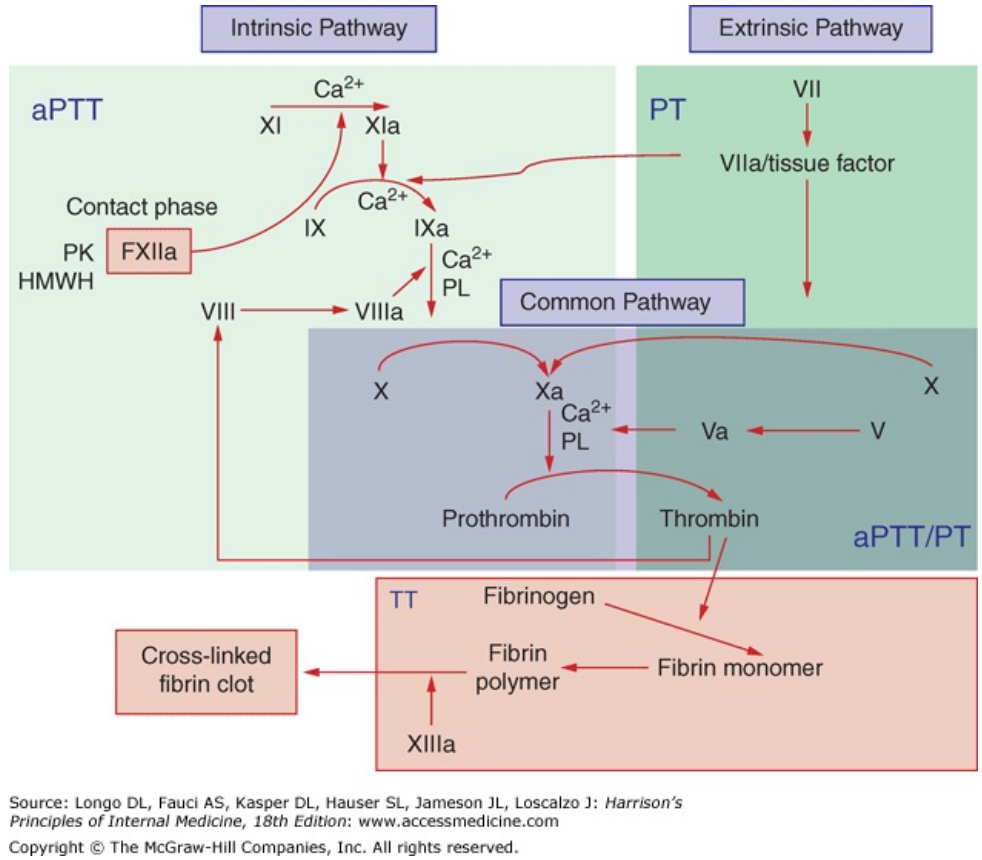

Coagulation cascade and laboratory assessment of clotting factor deficiency by activated partial prothrombin time (aPTT), prothrombin time (PT), and thrombin time (TT).

Acquired deficiencies of plasma coagulation factors are more frequent than congenital disorders; the most common disorders include hemorrhagic diathesis of liver disease, disseminated intravascular coagulation (DIC), and vitamin K deficiency. In these disorders, blood coagulation is hampered by the deficiency of more than one clotting factor, and the bleeding episodes are the result of perturbation of both primary (coagulation) and secondary (e.g., platelet and vessel wall interactions) hemostasis.

The development of antibodies to coagulation plasma proteins, clinically termed *inhibitors*, is a relatively rare disease that often affects hemophilia A or B and FXI-deficient patients on repetitive exposure to the missing protein to control bleeding episodes. Inhibitors also occur among subjects without genetic deficiency of clotting factors (e.g., in the postpartum setting as a manifestation of underlying autoimmune or neoplastic disease or idiopathically). Rare cases of inhibitors to thrombin or FV have been reported in patients receiving topical bovine thrombin preparation as a local hemostatic agent in complex surgeries. The diagnosis of inhibitors is based on the same tests as those used to diagnose inherited plasma coagulation factor deficiencies. However, the addition of the missing protein to the plasma of a subject with an inhibitor does not correct the abnormal aPTT and/or PT tests. This is the major laboratory difference between deficiencies and inhibitors. Additional tests are required to measure the specificity of the inhibitor and its titer.

The treatment of these bleeding disorders often requires replacement of the deficient protein using recombinant or purified plasma-derived products or fresh-frozen plasma (FFP). Therefore, it is imperative to arrive at a proper diagnosis to optimize patient care without unnecessary exposure to suboptimal treatment and the risks of bloodborne disease.

HEMOPHILIA

Pathogenesis and Clinical Manifestations

Hemophilia is an X-linked recessive hemorrhagic disease due to mutations in the *F8* gene (hemophilia A or classic hemophilia) or *F9* gene (hemophilia B). The disease affects 1 in 10,000 males worldwide, in all ethnic groups; hemophilia A represents 80% of all cases. Male subjects are clinically affected; women, who carry a single mutated gene, are generally asymptomatic. Family history of the disease is absent in ~30% of cases and in these cases, 80% of the mothers are carriers of the de novo mutated allele. More than 500 different mutations have been identified in the *F8* or *F9* genes of patients with hemophilia A or B, respectively. One of the most common hemophilia A mutations results from an inversion of the intron 22 sequence, and it is present in 40% of cases of severe hemophilia A. Advances in molecular diagnosis now permit precise identification of mutations, allowing accurate

diagnosis of women carriers of the hemophilia gene in affected families.

Clinically, hemophilia A and hemophilia B are indistinguishable. The disease phenotype correlates with the residual activity of FVIII or FIX and can be classified as severe (<1%), moderate (1–5%), or mild (6–30%). In the severe and moderate forms, the disease is characterized by bleeding into the joints (hemarthrosis), soft tissues, and muscles after minor trauma or even spontaneously. Patients with mild disease experience infrequent bleeding that is usually secondary to trauma. Among those with residual FVIII or FIX activity >25% of normal, the disease is discovered only by bleeding after major trauma or during routine presurgery laboratory tests. Typically, the global tests of coagulation show only an isolated prolongation of the aPTT assay. Patients with hemophilia have normal bleeding times and platelet counts. The diagnosis is made after specific determination of FVIII or FIX clotting activity.

Early in life, bleeding may present after circumcision or rarely as intracranial hemorrhages. The disease is more evident when children begin to walk or crawl. In the severe form, the most common bleeding manifestations are the recurrent hemarthroses, which can affect every joint but mainly affect knees, elbows, ankles, shoulders, and hips. Acute hemarthroses are painful, and clinical signs are local swelling and erythema. To avoid pain, the patient may adopt a fixed position, which leads eventually to muscle contractures. Very young children unable to communicate verbally show irritability and a lack of movement of the affected joint. Chronic hemarthroses are debilitating, with synovial thickening and synovitis in response to the intraarticular blood. After a joint has been damaged, recurrent bleeding episodes result in the clinically recognized "target joint," which then establishes a vicious cycle of bleeding, resulting in progressive joint deformity that in critical cases requires surgery as the only therapeutic option. Hematomas into the muscle of distal parts of the limbs may lead to external compression of arteries, veins, or nerves that can evolve to a compartment syndrome.

Bleeding into the oropharyngeal spaces, central nervous system (CNS), or the retroperitoneum is life threatening and requires immediate therapy. Retroperitoneal hemorrhages can accumulate large quantities of blood with formation of masses with calcification and inflammatory tissue reaction (pseudotumor syndrome) and also result in damage to the femoral nerve. Pseudotumors can also form in bones, especially long bones of the lower limbs. Hematuria is frequent among hemophilia patients, even in the absence of genitourinary pathology. It is often self-limited and may not require specific therapy.

Treatment: Hemophilia

Without treatment, severe hemophilia has a limited life expectancy. Advances in the blood fractionation industry during World War II resulted in the realization that plasma could be used to treat hemophilia, but the volumes required to achieve even modest elevation of circulating factor levels limit the utility of plasma infusion as an approach to disease management. The discovery in the 1960s that the cryoprecipitate fraction of plasma was enriched for FVIII, and the eventual purification of FVIII and FIX from plasma, led to the introduction of home infusion therapy with factor concentrates in the 1970s. The availability of factor concentrates resulted in a dramatic improvement in life expectancy and in quality of life for people with severe hemophilia. However, the contamination of the blood supply with hepatitis viruses and, subsequently HIV, resulted in widespread transmission of these bloodborne infections within the hemophilia population; complications of HIV and of hepatitis C are now the leading causes of death among U.S. adults with severe hemophilia. The introduction of viral inactivation steps in the preparation of plasma-derived products in the mid-1980s greatly reduced the risk of HIV and hepatitis, and the risks were further reduced by the successful production of recombinant FVIII and FIX proteins, both licensed in the 1990s. It is uncommon for hemophilic patients born after 1985 to have contracted either hepatitis or HIV, and for these individuals, life expectancy is in the range of age 65 years.

Factor replacement therapy for hemophilia can be provided either in response to a bleeding episode or as a prophylactic treatment. Primary prophylaxis is defined as a strategy for maintaining the missing clotting factor at levels ~1% or higher on a regular basis in order to prevent bleeds, especially the onset of hemarthroses. Hemophilic boys receiving regular infusions of FVIII (3 days/week) or FIX (2 days/week) can reach puberty without detectable joint abnormalities.

Prophylaxis has become gradually more common in young patients. The Centers for Disease Control and Prevention reported that 51% of children with severe hemophilia who are younger than age 6 years receive prophylaxis, increasing considerably from 33% in 1995. Although highly recommended, the high cost and difficulties in accessing peripheral veins in young patients and the potential infectious and thrombotic risks of long-term central vein catheters are important limiting factors for many patients.

General considerations regarding the treatment of bleeds in hemophilia include (1) the need to begin the treatment as soon as possible because symptoms often precede objective evidence of bleeding; because of the superior efficacy of early therapeutic intervention, classic symptoms of bleeding into the joint in a reliable patient, headaches, or automobile or other accidents require prompt replacement and further laboratory investigation; and (2) the need to avoid drugs that hamper platelet function, such as aspirin or aspirin-containing drugs; to control pain, drugs such as ibuprofen or propoxyphene are preferred.

Factor VIII and FIX are dosed in units. One unit is defined as amount of FVIII (100 ng/mL) or FIX (5 μg/mL) in 1 mL of normal plasma. One unit of FVIII per kilogram of body weight increases the plasma FVIII level by 2%. One can calculate the dose needed to increase FVIII levels to 100% in a 70-kg severe hemophilia patient (<1%) using the simple formula below. Thus, 3500 units of FVIII will raise the circulating level to 100%.

FVIII dose (IU) = Target FVIII levels – FVIII baseline levels x body weight (kg) x 0.5 unit/kg

The doses for FIX replacement are different from those for FVIII, because FIX recovery postinfusion is usually only 50% of the predicted value. Therefore, the formula for FIX replacement is

FIX dose (IU) = Target FIX levels – FIX baseline levels x body weight (kg) x 1 unit/kg

The FVIII half-life of 8–12 h requires injections twice a day to maintain therapeutic levels, whereas FIX half-life is longer, ~24 h, so that once-a-day injection is sufficient. In specific situations such as postsurgery, continuous infusion of factor may be desirable because of its safety in achieving sustained factor levels at a lower total cost.

Cryoprecipitate is enriched with FVIII protein (each bag contains ~80 IU of FVIII) and was commonly used for the treatment of hemophilia A decades ago; it is still in use in some developing countries, but because of the risk of bloodborne diseases, this product should be avoided in hemophilia patients when factor concentrates are available.

Mild bleeds such as uncomplicated hemarthroses or superficial hematomas require initial therapy with factor levels of 30–50%. Additional doses to maintain levels of 15–25% for 2 or 3 days are indicated for severe hemarthroses, especially when these episodes affect the "target joint." Large hematomas, or bleeds into deep muscles, require factor levels of 50% or even higher if the clinical symptoms do not improve, and factor replacement may be required for a period of 1 week or longer. The control of serious bleeds including those that affect the oropharyngeal spaces, CNS, and the retroperitoneum require sustained protein levels of 50–100% for 7–10 days. Prophylactic replacement for surgery is aimed at achieving normal factor levels (100%) for a period of 7–10 days; replacement can then be tapered depending on the extent of the surgical wounds. Oral surgery is associated with extensive tissue damage that usually requires factor replacement for 1–3 days coupled with oral antifibrinolytic drugs.

NONTRANFUSION THERAPY IN HEMOPHILIA

DDAVP (1-Amino-8-D-Arginine Vasopressin).

DDAVP is a synthetic vasopressin analog that causes a transient rise in FVIII and von Willebrand factor (vWF), but not FIX, through a mechanism involving release from endothelial cells. Patients with moderate or mild hemophilia A should be tested to determine if they respond to DDAVP before a therapeutic application. DDAVP at doses of 0.3 μg/kg body weight, over a 20-min period, is expected to raise FVIII levels by two- to threefold over baseline, peaking between 30 and 60 min postinfusion. DDAVP does not improve FVIII levels in severe hemophilia A patients, since there are no stores to release. Repeated dosing of DDAVP results in tachyphylaxis, since the mechanism is an increase in release rather than de novo synthesis of FVIII and vWF. More than three consecutive doses become ineffective and if further therapy is indicated, FVIII replacement is required to achieve hemostasis.

Antifibrinolytic Drugs.

Bleeding in the gums, gastrointestinal tract, and during oral surgery requires the use of oral antifibrinolytic drugs such as ε-amino caproic acid (EACA) or tranexamic acid to control local hemostasis. The duration of the treatment depending on the clinical indication is 1 week or longer. Tranexamic acid is given at doses of 25 mg/kg three to four times a day. EACA treatment requires a loading dose of 200 mg/kg (maximum of 10 g) followed by 100 mg/kg per dose (maximum 30 g/d) every 6 h. These drugs are not indicated to control hematuria because of the risk of formation of an occlusive clot in the lumen of genitourinary tract structures.

COMPLICATIONS

Inhibitor Formation.

The formation of alloantibodies to FVIII or FIX is currently the major complication of hemophilia treatment. The prevalence of inhibitors to FVIII is estimated to be between 5 and 10% of all cases and ~20% of severe hemophilia A patients. Inhibitors to FIX are detected in only 3–5% of all hemophilia B patients. The high-risk group for inhibitor formation includes severe deficiency (>80% of all cases of inhibitors), familial history of inhibitor, African descent, mutations in the FVIII or FIX gene resulting in deletion of large coding regions, or gross gene rearrangements. Inhibitors usually appear early in life, at a median of 2 years of age, and after 10 cumulative days of exposure. However, intensive replacement therapy such as for major surgery, intracranial bleeding, or trauma increases the risk of inhibitor formation for patients of all ages that requires close laboratory monitoring in the following weeks.

The clinical diagnosis of an inhibitor is suspected when patients do not respond to factor replacement at therapeutic doses. Inhibitors increase both morbidity and mortality in hemophilia. Because early

detection of an inhibitor is critical to a successful correction of the bleeding or to eradication of the antibody, most hemophilia centers perform annual screening for inhibitors. The laboratory test required to confirm the presence of an inhibitor is an aPTT with a mix (with normal plasma). In most hemophilia patients, a 1:1 mix with normal plasma completely corrects the aPTT. In inhibitor patients, the aPTT on a 1:1 mix is abnormally prolonged, because the inhibitor neutralizes the FVIII clotting activity of the normal plasma. The Bethesda assay uses a similar principle and defines the specificity of the inhibitor and its titer. The results are expressed in Bethesda units (BU), in which 1 BU is the amount of antibody that neutralizes 50% of the FVIII or FIX present in normal plasma after 2 h of incubation at 37°C. Clinically, inhibitor patients are classified as low responders or high responders, which provides guidelines for optimal therapy. Therapy for inhibitor patients has two goals, the control of acute bleeding episodes and the eradication of the inhibitor. For the control of bleeding episodes, low responders, those with titer <5 BU, respond well to high doses of human or porcine FVIII (50–100 U/kg), with minimal or no increase in the inhibitor titers. However, high-responder patients, those with initial inhibitor titer >10 BU or an anamnestic response in the antibody titer to >10 BU even if low titer initially, do not respond to FVIII or FIX concentrates. The control of bleeding episodes in high-responder patients can be achieved by using concentrates enriched for prothrombin, FVII, FIX, FX [prothrombin complex concentrates (PCCs) or activated PCCs], and more recently by recombinant activated factor VII (FVIIa) (Fig. 116-1). The rates of therapeutic success have been higher for FVIIa than for PCC or aPCC. For eradication of the inhibitory antibody, immunosuppression alone is not effective. The most effective strategy is the immune tolerance induction (ITI) based on daily infusion of missing protein until the inhibitor disappears, typically requiring periods longer than 1 year, with success rates in the range of 60%. The management of patients with severe hemophilia A and inhibitors resistant to ITI is challenging. The use of anti-CD20 monoclonal antibody (rituximab) combined with FVIII was thought to be effective. Although this therapy may reduce the inhibitor titers, sustained eradication is uncommon and may require two to three infusions weekly of FVIII concentrates.

Infectious Diseases.

Hepatitis C viral (HCV) infection is the major cause of morbidity and the second leading cause of death in hemophilia patients exposed to older clotting factor concentrates. The vast majority of young patients treated with plasma-derived products from 1970 to 1985 became infected with HCV. It has been estimated that >80% of patients older than 20 years of age are HCV antibody positive as of 2006. The comorbidity of the underlying liver disease in hemophilia patients is clear when these individuals require invasive procedures; correction of both genetic and acquired (secondary to liver disease) deficiencies may be needed. Infection with HIV also swept the population of patients using plasma-derived concentrates 2 decades ago. Co-infection of HCV and HIV, present in almost 50% of hemophilia patients, is an aggravating factor for the evolution of liver disease. The response to HCV antiviral therapy in hemophilia is restricted to <30% of patients and even poorer among those with both HCV and HIV infection. End-stage liver disease requiring organ transplantation may be curative for both the liver disease and for hemophilia.

Emerging Clinical Problems in Aging Hemophilia Patients.

There has been continuous improvement of the management of hemophilia since the increase in the population of adults living beyond middle age in the developing world. The life expectancy of a patient with severe hemophilia is only ~10 years shorter than the general male population. In patients with mild or moderate hemophilia, life expectancy is approaching that of the male population without coagulopathy. Elderly hemophilia patients have different problems compared to the younger generation; they have more severe arthropathy and chronic pain due to suboptimal treatment, and high rates of HCV and/or HIV infections.

Early data indicate that mortality from coronary artery disease is lower in hemophilia patients than the general male population. The underlying hypocoagulability probably provides a protective effect against thrombus formation, but it does not prevent the development of atherogenesis. Similar to the general population, these patients are exposed to cardiovascular risk factors such as age, obesity, and smoking. Moreover, physical inactivity, hypertension, and chronic renal disease are commonly observed in hemophilia patients. In HIV patients on combined antiretroviral therapy, there may be a further increase in the risk of cardiovascular disease. Therefore, these patients should be carefully considered for preventive and therapeutic approaches to minimize the risk of cardiovascular disease.

Excessive replacement therapy should be avoided, and it is prudent to slowly infuse factor concentrates. Continuous infusion of clotting factor is preferable to bolus dosing in patients with cardiovascular risk factors undergoing invasive procedures. The management of an acute ischemic event and coronary revascularization should include the collaboration of hematologists and internists. The early assumption that hemophilia would protect against occlusive vascular disease may change in this aging population.

Cancer is a common cause of mortality in aging hemophilia patients as they are at risk for HIV- and HCV-related malignancies. Hepatocellular carcinoma (HCC) is the most prevalent primary liver cancer and a common cause of death in HIV-negative patients. The recommendations for cancer screening for

the general population should be the same for age-matched hemophilia patients. Among those with high-risk HCV, a semiannual or annual ultrasound and α fetoprotein is recommended for HCC. Screening for urogenital neoplasm in the presence of hematuria or hematochezia may be delayed due to the underlying bleeding disease, thus preventing early intervention. Multidisciplinary interaction should facilitate the attempts to ensure optimal cancer prevention and treatment recommendations for those with hemophilia.

Management of Carriers of Hemophilia.

Usually hemophilia carriers, with factor levels of ~50% of normal, have not been considered to be at risk for bleeding. However, a wide range of values (22–116%) have been reported due to random inactivation of the X chromosomes (*lyonization*). Therefore, it is important to measure the factor level of carriers to recognize those at risk of bleeding and to optimize preoperative and postoperative management. During pregnancy, both FVIII and FIX levels increase gradually until delivery. FVIII levels increase approximately two- to threefold compared to nonpregnant women, whereas a FIX increase is less pronounced. After delivery, there is a rapid fall in the pregnancy-induced rise of maternal clotting factor levels. This represents an imminent risk of bleeding that can be prevented by infusion of factor concentrate to levels of 50–70% for 3 days in the setting of vaginal delivery and up to 5 days for cesarean section. In mild cases, the use of DDAVP and/or antifibrinolytic drugs is recommended.

Factor XI Deficiency

Factor XI is a zymogen of an active serine protease (FIXa) in the intrinsic pathway of blood coagulation that activates FIX (Fig. 116-1). There are two pathways for the formation of FXIa. In an aPTT-based assay, the protease is the result of activation by FXIIa in conjunction with high-molecular-weight kininogen and kallikrein. In vivo data suggest that thrombin is the physiologic activator of FXI. The generation of thrombin by the tissue-factor/factor VIIa pathway activates FXI on the platelet surface that contributes to additional thrombin generation after the clot has formed and thus augments resistance to fibrinolysis through a thrombin-activated fibrinolytic inhibitor (TAFI).

Factor XI deficiency is a rare bleeding disorder that occurs in the general population at a frequency of one in a million. However, the disease is highly prevalent among Ashkenazi and Iraqi Jewish populations, reaching a frequency of 6% as heterozygotes and 0.1 to 0.3% as homozygotes. More than 65 mutations in the FXI gene have been reported, whereas fewer mutations (two to three) are found among affected Jewish populations.

Normal FXI clotting activity levels range from 70 to 150 U/dL. In heterozygous patients with moderate deficiency, FXI ranges from 20 to 70 U/dL, whereas in homozygous or double heterozygote patients, FXI levels are <1–20 U/dL. Patients with FXI levels <10% of normal have a high risk of bleeding, but the disease phenotype does not always correlate with residual FXI clotting activity. A family history is indicative of the risk of bleeding in the propositus. Clinically, the presence of mucocutaneous hemorrhages such as bruises, gum bleeding, epistaxis, hematuria, and menorrhagia are common, especially following trauma. This hemorrhagic phenotype suggests that tissues rich in fibrinolytic activity are more susceptible to FXI deficiency. Postoperative bleeding is common but not always present, even among patients with very low FXI levels.

FXI replacement is indicated in patients with severe disease required to undergo a surgical procedure. A negative history of bleeding complications following invasive procedures does not exclude the possibility of an increased risk for hemorrhage.

TREATMENT: FACTOR XI DEFICIENCY

The treatment of FXI deficiency is based on the infusion of FFP at doses of 15 to 20 mL/kg to maintain trough levels ranging from 10 to 20%. Because FXI has a half-life of 40–70 h, the replacement therapy can be given on alternate days. The use of antifibrinolytic drugs is beneficial to control bleeds, with the exception of hematuria or bleeds in the bladder. The development of an FXI inhibitor was observed in 10% of severely FXI-deficient patients who received replacement therapy. Patients with severe FXI deficiency who develop inhibitors usually do not bleed spontaneously. However, bleeding following a surgical procedure or trauma can be severe. In these patients, FFP and FXI concentrates should be avoided. The use of PCC/aPCC or recombinant activated FVII has been effective.

RARE BLEEDING DISORDERS

Collectively, the inherited disorders resulting from deficiencies of clotting factors other than FVIII, FIX, and FXI (Table 116–1) represent a group of rare bleeding diseases. The bleeding symptoms in these patients vary from asymptomatic (dysfibrinogenemia or FVII deficiency) to life-threatening (FX or FXIII deficiency). There is no pathognomonic clinical manifestation that suggests one specific disease, but overall, in contrast to hemophilia, hemarthrosis is a rare event and bleeding in the mucosal tract or after umbilical cord clamping is common. Individuals heterozygous for plasma coagulation deficiencies are often asymptomatic. The laboratory assessment for the specific deficient factor following screening with general coagulation tests (Table 116–1) will define the diagnosis.

Replacement therapy using FFP or prothrombin complex concentrates (containing prothrombin, FVII, FIX, and FX) provides adequate hemostasis in response to bleeds or as prophylactic treatment. The

use of PCC should be carefully monitored and avoided in patients with underlying liver disease, or those at high risk for thrombosis because of the risk of disseminated intravascular coagulopathy.

Familial Multiple Coagulation Deficiencies

There are several bleeding disorders characterized by the inherited deficiency of more than one plasma coagulation factor. To date, the genetic defects in two of these diseases have been characterized and they provide new insights into the regulation of hemostasis by gene- encoding proteins outside blood coagulation.

COMBINED DEFICIENCY OF FV AND FVIII

Patients with combined FV and FVIII deficiency exhibit ~5% of residual clotting activity of each factor. Interestingly, the disease phenotype is a mild bleeding tendency, often following trauma. An underlying mutation has been identified in the endoplasmic reticulum/Golgi intermediate compartment (ERGIC-53) gene, a mannose-binding protein localized in the Golgi apparatus that functions as a chaperone for both FV and FVIII. In other families, mutations in the multiple coagulation factor deficiency 2 (MCFD2) gene have been defined; this gene encodes a protein that forms a Ca^{2+}–dependent complex with ERGIC-53 and provides cofactor activity in the intracellular mobilization of both FV and FVIII.

MULTIPLE DEFICIENCIES OF VITAMIN K–DEPENDENT COAGULATION FACTORS

Two enzymes involved in vitamin K metabolism have been associated with combined deficiency of all vitamin K–dependent proteins, including the procoagulant proteins prothrombin, VII, IX, and X and the anticoagulant proteins C and S. Vitamin K is a fat-soluble vitamin that is a cofactor for carboxylation of the gamma carbon of the glutamic acid residues in the vitamin K–dependent factors, a critical step for calcium and phospholipid binding of these proteins **(Fig. 116-2)**. The enzymes γ-glutamylcarboxylase and epoxide reductase are critical for the metabolism and regeneration of vitamin K. Mutations in the genes encoding the gamma-carboxylase (GGCX) or vitamin K epoxide reductase complex 1 (VKORC1) result in defective enzymes and thus in vitamin K–dependent factors with reduced activity, varying from 1 to 30% of normal. The disease phenotype is characterized by mild to severe bleeding episodes present from birth. Some patients respond to high doses of vitamin K. For severe bleeding, replacement therapy with FFP or PCC may be necessary for achieving full hemostatic control.

Figure 116-2

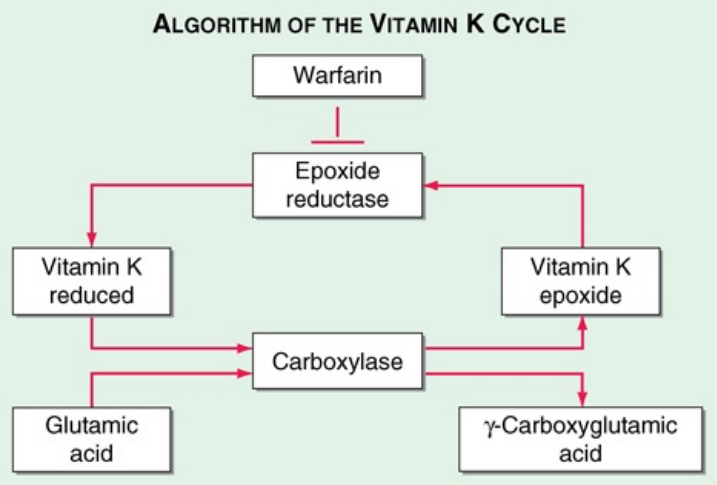

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The vitamin K cycle. Vitamin K is a cofactor for the formation of γ-carboxyglutamic acid residues on coagulation proteins. Vitamin K–dependent γ-glutamylcarboxylase, the enzyme that catalyzes the vitamin K epoxide reductase, regenerates reduced vitamin K. Warfarin blocks the action of the reductase and competitively inhibits the effects of vitamin K.

Disseminated Intravascular Coagulation

Disseminated intravascular coagulation (DIC) is a clinicopathologic syndrome characterized by widespread intravascular fibrin formation in response to excessive blood protease activity that overcomes the natural anticoagulant mechanisms. There are several underlying pathologies associated with DIC **(Table 116–2)**.

Table 116–2 Common Clinical Causes of Disseminated Intravascular Coagulation

Sepsis	Immunologic disorders

■ Bacterial: Staphylococci, streptococci, pneumococci, meningococci, gram-negative bacilli ■ Viral ■ Mycotic ■ Parasitic ■ Rickettsial	■ Acute hemolytic transfusion reaction ■ Organ or tissue transplant rejection ■ Graft-versus-host disease
Trauma and tissue injury	**Drugs**
■ Brain injury (gunshot) ■ Extensive burns ■ Fat embolism ■ Rhabdomyolysis	■ Fibrinolytic agents ■ Aprotinin ■ Warfarin (especially in neonates with protein C deficiency) ■ Prothrombin complex concentrates ■ Recreational drugs (amphetamines)
Vascular disorders	**Envenomation**
■ Giant hemangiomas (Kasabach-Merritt syndrome) ■ Large vessel aneurysms (e.g., aorta)	■ Snake ■ Insects
Obstetrical complications	**Liver disease**
■ Abruptio placentae ■ Amniotic-fluid embolism ■ Dead fetus syndrome ■ Septic abortion	■ Fulminant hepatic failure ■ Cirrhosis ■ Fatty liver of pregnancy
Cancer	**Miscellaneous**
■ Adenocarcinoma (prostate, pancreas, etc.) ■ Hematologicmalignancies (acute promyelocytic leukemia)	■ Shock ■ Respiratory distress syndrome ■ Massive transfusion

The most common causes are bacterial sepsis, malignant disorders such as solid tumors or acute promyelocytic leukemia, and obstetric causes. DIC is diagnosed in almost one-half of pregnant women with abruptio placentae, or with amniotic fluid embolism. Trauma, particularly to the brain, can also result in DIC. The exposure of blood to phospholipids from damaged tissue, hemolysis, and endothelial damage are all contributing factors to the development of DIC in this setting. Purpura fulminans is a severe form of DIC resulting from thrombosis of extensive areas of the skin; it affects predominantly young children following viral or bacterial infection, particularly those with inherited or acquired hypercoagulability due to deficiencies of the components of the protein C pathway. Neonateshomozygous for protein C deficiency also present high risk for purpura fulminans with or without thrombosis of large vessels.

The central mechanism of DIC is the uncontrolled generation of thrombin by exposure of the blood to pathologic levels of tissue factor **(Fig. 116-3)**. Simultaneous suppression of physiologic anticoagulant mechanisms and abnormal fibrinolysis further accelerate the process. Together, these abnormalities contribute to systemic fibrin deposition in small and midsize vessels. The duration and intensity of the fibrin deposition can compromise the blood supply of many organs, especially the lung, kidney, liver, and brain, with consequent organ failure. The sustained activation of coagulation results in consumption of clotting factors and platelets, which in turn leads to systemic bleeding. This is further aggravated by secondary hyperfibrinolysis. Studies in animals demonstrate that the fibrinolytic system is indeed suppressed at the time of maximal activation of coagulation. Interestingly, in patients with acute promyelocytic leukemia, a severe hyperfibrinolytic state often occurs in addition to the coagulation activation. The release of several proinflammatory cytokines such as interleukin-6 and tumor necrosis factor α play central roles in mediating the coagulation defects in DIC and symptoms associated with systemic inflammatory response syndrome (SIRS).

Figure 116-3

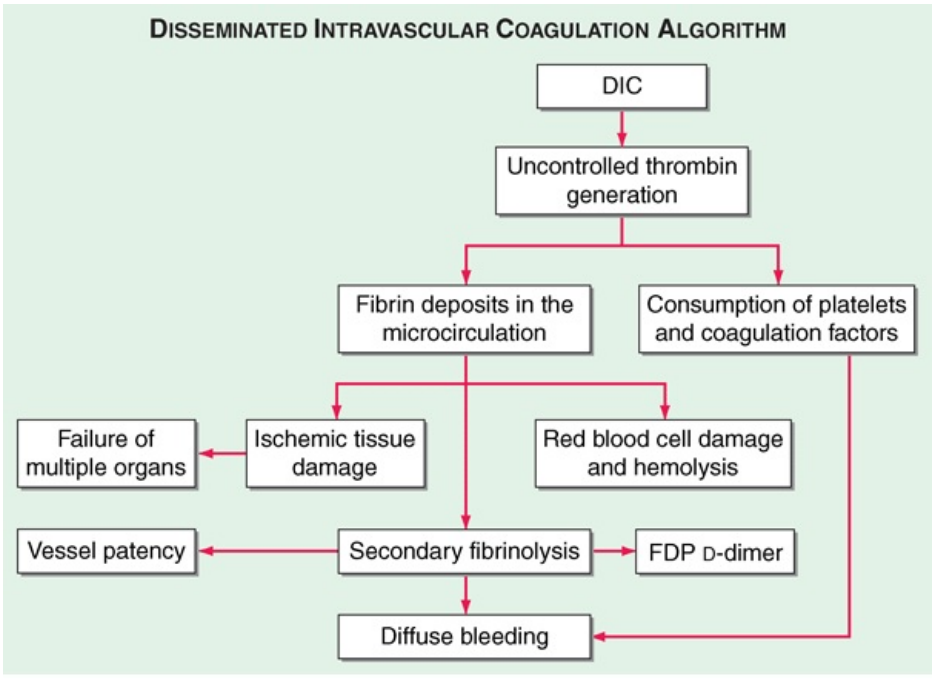

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The pathophysiology of disseminated intravascular coagulation (DIC). Interactions between coagulation and fibrinolytic pathways result in bleeding and thrombosis in the microcirculation in patients with DIC.

Clinical manifestations of DIC are related to the magnitude of the imbalance of hemostasis, to the underlying disease, or to both. The most common findings are bleeding ranging from oozing from venipuncture sites, petechiae, and ecchymoses to severe hemorrhage from the gastrointestinal tract, lung, or into the CNS. In chronic DIC, the bleeding symptoms are discrete and restricted to skin or mucosal surfaces. The hypercoagulability of DIC manifests as the occlusion of vessels in the microcirculation and resulting organ failure. Thrombosis of large vessels and cerebral embolism can also occur. Hemodynamic complications and shock are common among patients with acute DIC. The mortality ranges from 30 to >80% depending on the underlying disease, the severity of the DIC, and the age of the patient.

The diagnosis of clinically significant DIC is based on the presence of clinical and/or laboratory abnormalities of coagulation or thrombocytopenia. The laboratory diagnosis of DIC should prompt a search for the underlying disease if it is not already apparent. There is no single test that establishes the diagnosis of DIC. The laboratory investigation should include coagulation tests [aPTT, PT, thrombin time (TT)] and markers of fibrin degradation products (FDPs), in addition to platelet and red cell count and analysis of the blood smear. These tests should be repeated over a period of 6–8 hours because an initially mild abnormality can change dramatically in patients with severe DIC.

Common findings include the prolongation of PT and/or aPTT; platelet counts $\leq$100,000/μL^3, or a rapid decline in platelet numbers; the presence of schistocytes (fragmented red cells) in the blood smear; and elevated levels of FDP. The most sensitive test for DIC is the FDP level. DIC is an unlikely diagnosis in the presence of normal levels of FDP. The D-dimer test is more specific for detection of fibrin—but not fibrinogen—degradation products and indicates that the cross-linked fibrin has been digested by plasmin. Because fibrinogen has a prolonged half-life, plasma levels diminish acutely only in severe cases of DIC. High-grade DIC is also associated with levels of antithrombin III or plasminogen activity <60% of normal.

CHRONIC DIC

Low-grade, compensated DIC can occur in clinical situations including giant hemangioma, metastatic carcinoma, or the dead fetus syndrome. Plasma levels of FDP or d-dimers are elevated. aPTT, PT, and fibrinogen values are within the normal range or high. Mild thrombocytopenia or normal platelet counts are also common findings. Red cell fragmentation is often detected but at a lower degree than in acute DIC.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis between DIC and severe liver disease is challenging and requires serial measurements of the laboratory parameters of DIC. Patients with severe liver disease are at risk for bleeding and manifest laboratory features including thrombocytopenia (due to platelet sequestration, portal hypertension, or hypersplenism), decreased synthesis of coagulation factors and natural

anticoagulants, and elevated levels of FDP due to reduced hepatic clearance. However, in contrast to DIC, these laboratory parameters in liver disease do not change rapidly. Other important differential findings include the presence of portal hypertension or other clinical or laboratory evidence of an underlying liver disease.

Microangiopathic disorders such as thrombotic thrombocytopenic purpura present an acute clinical onset of illness accompanied by thrombocytopenia, red cell fragmentation, and multiorgan failure. However, there is no consumption of clotting factors or hyperfibrinolysis.

TREATMENT: DISSEMINATED INTRAVASCULAR COAGULATION

The morbidity and mortality associated with DIC are primarily related to the underlying disease rather than the complications of the DIC. The control or elimination of the underlying cause should therefore be the primary concern. Patients with severe DIC require control of hemodynamic parameters, respiratory support, and sometimes invasive surgical procedures. Attempts to treat DIC without accompanying treatment of the causative disease are likely to fail.

Management of Hemorrhagic Symptoms

The control of bleeding in DIC patients with marked thrombocytopenia (platelet counts <10,000 –20,000/μL^3) and low levels of coagulation factors will require replacement therapy. The PT (>1.5 times the normal) provides a good indicator of the severity of the clotting factor consumption. Replacement with FFP is indicated (1 unit of FFP increases most coagulation factors by 3% in an adult without DIC). Low levels of fibrinogen (<100 mg/dL) or brisk hyperfibrinolysis will require infusion of cryoprecipitate (plasma fraction enriched for fibrinogen, FVIII, and vWF). The replacement of 10 U of cryoprecipitate for every 2–3 U of FFP is sufficient to correct the hemostasis. The transfusion scheme must be adjusted according to the patient's clinical and laboratory evolution. Platelet concentrates at a dose of 1–2 U/10 kg body weight are sufficient for most DIC patients with severe thrombocytopenia.

Clotting factor concentrates are not recommended for control of bleeding in DIC because of the limited efficacy afforded by replacement of single factors (FVIII or FIX concentrates), and the high risk of products containing traces of aPCCs that further aggravate the disease.

Replacement of Coagulation or Fibrinolysis Inhibitors

Drugs to control coagulation such as heparin, ATIII concentrates, or antifibrinolytic drugs have all been tried in the treatment of DIC. Low doses of continuous infusion heparin (5–10 U/kg per h) may be effective in patients with low-grade DIC associated with solid tumor, acute promyelocytic leukemia, or in a setting with recognized thrombosis. Heparin is also indicated for the treatment of purpura fulminans during the surgical resection of giant hemangiomas and during removal of a dead fetus. In acute DIC, the use of heparin is likely to aggravate bleeding. To date, the use of heparin in patients with severe DIC has no proven survival benefit.

The use of antifibrinolytic drugs, EACA, or tranexamic acid, to prevent fibrin degradation by plasmin may reduce bleeding episodes in patients with DIC and confirmed hyperfibrinolysis. However, these drugs can increase the risk of thrombosis and concomitant use of heparin is indicated. Patients with acute promyelocytic leukemia or those with chronic DIC associated with giant hemangiomas are among the few patients who may benefit from this therapy.

The use of protein C concentrates to treat purpura fulminans associated with acquired protein C deficiency or meningococcemia has been proven efficacious. The results from the replacement of ATIII in early-phase studies are promising but require further study.

Vitamin K Deficiency

Vitamin K–dependent proteins are a heterogenous group, including clotting factor proteins and also proteins found in bone, lung, kidney, and placenta. Vitamin K mediates posttranslational modification of glutamate residues to γ-carboxylglutamate, a critical step for the activity of vitamin K–dependent proteins for calcium binding and proper assembly to phospholipid membranes (Fig. 116-2). Inherited deficiency of the functional activity of the enzymes involved in vitamin K metabolism, notably the GGCX or VKORC1 (see above), results in bleeding disorders. The amount of vitamin K in the diet is often limiting for the carboxylation reaction; thus recycling of the vitamin K is essential to maintain normal levels of vitamin K–dependent proteins. In adults, low dietary intake alone is seldom reason for severe vitamin K deficiency but may become common in association with the use of broad-spectrum antibiotics. Disease or surgical interventions that affect the ability of the intestinal tract to absorb vitamin K, either through anatomic alterations or by changing the fat content of bile salts and pancreatic juices in the proximal small bowel, can result in significant reduction of vitamin K levels. Chronic liver diseases such as primary biliary cirrhosis also deplete vitamin K stores. Neonatal vitamin K deficiency and the resulting hemorrhagic disease of the newborn have been almost entirely eliminated by routine administration of vitamin K to all neonates. Prolongation of PT values is the most common and earliest finding in vitamin K–deficient patients due to reduction in prothrombin, FVII, FIX, and FX levels. FVII has the shortest half-life among these factors that can prolong the PT before changes in the aPTT. Parenteral administration of vitamin K at a total dose of 10 mg is sufficient to restore normal levels of clotting factor within 8–10 h. In the presence of ongoing bleeding or a need for

immediate correction before an invasive procedure, replacement with FFP or PCC is required. The latter should be avoided in patients with severe underlying liver disorders due to high risk of thrombosis. The reversal of excessive anticoagulant therapy with warfarin or warfarin-like drugs can be achieved by minimal doses of vitamin K (1 mg orally or by intravenous injection) for asymptomatic patients. This strategy can diminish the risk of bleeding while maintaining therapeutic anticoagulation for an underlying prothrombotic state.

In patients with life-threatening bleeds, the use of recombinant factor VIIa in nonhemophilia patients on anticoagulant therapy has been shown to be effective at restoring hemostasis rapidly, allowing emergency surgical intervention. However, patients with underlying vascular disease, vascular trauma and other comorbidities are at risk for thromboembolic complications that affect both arterial and venous systems. Thus, the use of factor VIIa in this setting is limited to administration of low doses given for only a limited number of injections. Close monitoring for vascular complications is highly indicated.

Coagulation Disorders Associated with Liver Failure

The liver is central to hemostasis because it is the site of synthesis and clearance of most procoagulant and natural anticoagulant proteins and of essential components of the fibrinolytic system. Liver failure is associated with a high risk of bleeding due to deficient synthesis of procoagulant factors and enhanced fibrinolysis. Thrombocytopenia is common in patients with liver disease, and may be due to congestive splenomegaly (hypersplenism), or immune-mediated shortened platelet lifespan (primary biliary cirrhosis). In addition, several anatomic abnormalities secondary to underlying liver disease further promote the occurrence of hemorrhage **(Table 116–3)**. Dysfibrinogenemia is a relatively common finding in patients with liver disease due to impaired fibrin polymerization. The development of DIC concomitant to chronic liver disease is not uncommon and may enhance the risk for bleeding. Laboratory evaluation is mandatory for an optimal therapeutic strategy, either to control ongoing bleeding or to prepare patients with liver disease for invasive procedures. Typically, these patients present with prolonged PT, aPTT, and TT depending on the degree of liver damage, thrombocytopenia, and normal or slight increase of FDP. Fibrinogen levels are diminished only in fulminant hepatitis, decompensated cirrhosis or advanced liver disease, or in the presence of DIC. The presence of prolonged TT and normal fibrinogen and FDP levels suggest dysfibrinogenemia. FVIII levels are often normal or elevated in patients with liver failure, and decreased levels suggest superimposing DIC. Because FV is only synthesized in the hepatocyte and is not a vitamin K –dependent protein, reduced levels of FV may be an indicator of hepatocyte failure. Normal levels of FV and low levels of FVII suggest vitamin K deficiency. Vitamin K levels may be reduced in patients with liver failure due to compromised storage in hepatocellular disease, changes in bile acids or cholestasis that can diminish the absorption of vitamin K. Replacement of vitamin K may be desirable (10 mg given by slow intravenous injection) to improve hemostasis.

Table 116–3 Coagulation Disorders and Hemostasis in Liver Disease

Bleeding
Portal hypertension
Esophageal varices
Thrombocytopenia
Splenomegaly Chronic or acute DIC
Decreased synthesis of clotting factors
Hepatocyte failure
Vitamin K deficiency
Systemic fibrinolysis
DIC
Dysfibrinogenemia
Thrombosis
Decreased synthesis of coagulation inhibitors: protein C, protein S, antithrombin
Hepatocyte failure
Vitamin K deficiency (protein C, protein S)
Failure to clear activated coagulation proteins (DIC)
Dysfibrinogenemia
Iatrogenic: Transfusion of prothrombin complex concentrates
Antifibrinolytic agents: EACA, tranexamic acid

Abbreviations: DIC, disseminated intravascular coagulation; EACA, ε-aminocaproic acid.

Treatment with FFP is the most effective to correct hemostasis in patients with liver failure. Infusion of FFP (5–10 mL/kg; each bag contains ~200 mL) is sufficient to ensure 10–20% of normal levels of clotting factors but not correction of PT or aPTT. Even high doses of FFP (20 mL/kg) do not correct the clotting times in all patients. Monitoring for clinical symptoms and clotting times will determine if repeated doses are required 8–12 h after the first infusion. Platelet concentrates are indicated when platelet counts are <10,000–20,000/μL^3 to control an ongoing bleed or immediately before an invasive procedure if counts are <50,000/μL^3. Cryoprecipitate is indicated only when fibrinogen levels are less than 100 mg/mL; dosing is six bags for a 70-kg patient daily. Prothrombin complex concentrate infusion in patients with liver failure should be avoided due to the high risk of thrombotic complications. The safety of the use of antifibrinolytic drugs to control bleeding in patients with liver failure is not yet well defined and should be avoided.

LIVER DISEASE AND THROMBOEMBOLISM

The clinical bleeding phenotype of hemostasis in patients with stable liver disease is often mild or even asymptomatic. However, as the disease progresses, the hemostatic balance is less stable and more easily disturbed than in healthy individuals. Furthermore, the hemostatic balance is compromised by comorbid complications such as infections and renal failure **(Fig. 116-4)**. Based on the clinical bleeding complications in patients with cirrhosis and laboratory evidence of hypocoagulation such as a prolonged PT/aPTT, it has long been assumed that these patients are protected against thrombotic disease. Cumulative clinical experience, however, has demonstrated that these patients are at risk for thrombosis, especially those with advanced liver disease. Although hypercoagulability could explain the occurrence of venous thrombosis, according to Virchow's triad, hemodynamic changes and damaged vasculature may also be a contributing factor, and both processes may potentially also occur in patients with liver disease. Liver-related thrombosis, in particular, thrombosis of the portal and mesenteric veins, is common in patients with advanced cirrhosis. Hemodynamic changes such as decreased portal flow, and evidence that inherited thrombophilia may enhance the risk for portal vein thrombosis in patients with cirrhosis suggest that hypercoagulability may play a role as well. Patients with liver disease develop deep vein thrombosis and pulmonary embolism at appreciable rates (ranging from 0.5 to 1.9%). The implication of these findings is relevant to the erroneous exclusion of thrombosis in patients with advanced liver disease, even in the presence of prolongation of routine clotting times, and caution should be advised on overcorrection of these laboratory abnormalities.

Figure 116-4

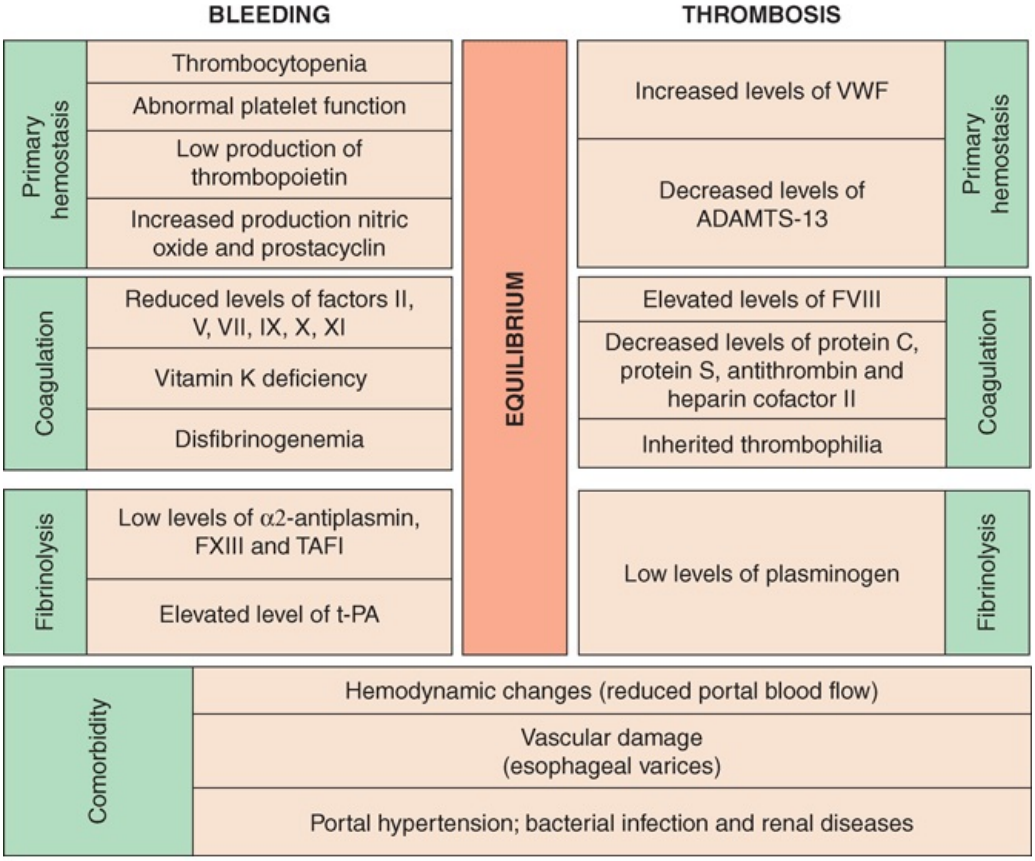

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Balance of hemostasis in liver disease. TAFI, thrombin-activated fibriolytic inhibitor; t-PA, tissue plasminogen activator; VWF, von Willebrand factor.

Acquired Inhibitors of Coagulation Factors

An acquired inhibitor is an immune-mediated disease characterized by the presence of an autoantibody against a specific clotting factor. FVIII is the most common target of antibody formation, but inhibitors to prothrombin, FV, FIX, FX, and FXI are also reported. The disease occurs predominantly in older adults (median age of 60 years), but occasionally in pregnant or postpartum women with no previous history of bleeding. In 50% of the patients with inhibitors, no underlying disease is identified at the time of diagnosis. In the remaining, the causes are automminune diseases, malignancies (lymphomas, prostate cancer), dermatologic diseases, and pregnancy. Bleeding episodes occur commonly in soft tissues, in the gastrointestinal or urinary tracts, and skin. In contrast to hemophilia, hemarthrosis is rare in these patients. Retroperitoneal hemorrhages and other life-threatening bleeding may appear suddenly. The overall mortality in untreated patients ranges from 8 to 22%, and most deaths occur within the first few weeks after presentation. The diagnosis is based on the prolonged aPTT with normal PT and TT. The aPTT remains prolonged after mixture of the test plasma with equal amounts of pooled normal plasma for 2 h at 37°C. The Bethesda assay using FVIII-deficient plasma as performed for inhibitor detection in hemophilia will confirm the diagnosis. Major bleeding is treated with high doses of human or porcine FVIII, PCC/PCCa, or recombinant FVIIa. High-dose intravenous gamma globulin, and anti-CD20 monoclonal antibody have been reported to be effective in patients with autoantibodies to FVIII. In contrast to hemophilia, inhibitors in nonhemophilia patients are sometimes responsive to prednisone alone or in association with cytotoxic therapy (e.g., cyclophosphamide).

Topical plasma-derived bovine and human thrombin are commonly used in the United States and worldwide. These effective hemostatic sealants are used during major surgery such as for cardiovascular, thoracic, neurologic, pelvic, and trauma indications, as well as in the setting of extensive burns. The development of antibody formation to the xenoantigen or its contaminant (bovine clotting protein) has the potential to show cross-reactivity with human clotting factors that may hamper their function and induce bleeding.

Clinical features of these antibodies include bleeding from a primary hemostastatic defect or coagulopathy that sometimes can be life threatening. The clinical diagnosis of these acquired coagulopathies is often complicated by the fact that the bleeding episodes may be detectable during or immediately following major surgery that could be assumed to be due to the procedure itself.

Notably, the risk of this complication is further increased by repeated exposure to topical thrombin preparations. Thus, a careful medical history of previous surgical interventions that may have occurred even decades earlier is critical to assessing risk.

The laboratory abnormalities are reflected by combined prolongation of the aPTT and PT that often fail to improve by transfusion of FFP and vitamin K. The abnormal laboratory tests cannot be corrected by mixing a test with equal parts of normal plasma that denotes the presence of inhibitory antibodies. The diagnosis of a specific antibody is obtained by the determination of the residual activity of human FV or other suspected human clotting factor. There are no commercially available assays specific for bovine thrombin coagulopathy.

There are no established treatment guidelines. Platelet transfusions have been utilized as a source of FV replacement for patients with FV inhibitors. Frequent injections of FFP and vitamin K supplementation may function as co-adjuvant rather than an effective treatment of the coagulopathy itself. Experience with recombinant FVIIa as a bypass agent is limited, and outcomes have been generally poor. Specific treatments to eradicate the antibodies based on immunosuppression with steroids, intravenous immunoglobulin, or serial plasmapheresis have been sporadically reported. Patients should be advised to avoid any topical thrombin sealant in the future.

Recently, novel plasma-derived and recombinant human thrombin preparations for topical hemostasis have been approved by the Food and Drug Administration. These preparations have demonstrated hemostatic efficacy with reduced immunogenicity compared to the first generation of bovine thrombin products.

The presence of lupus anticoagulant can be associated with venous or arterial thrombotic disease. However, bleeding has also been reported in lupus anticoagulant; it is due to the presence of antibodies to prothrombin, which results in hypoprothrombinemia. Both disorders show a prolonged PTT that do not correct on mixing. To distinguish acquired inhibitors from lupus anticoagulant, note that the dilute Russell's viper venom test and the hexagonal-phase phospholipids test will be negative in patients with an acquired inhibitor and positive in patients with lupus anticoagulants. Moreover, lupus anticoagulant interferes with the clotting activity of many factors (FVIII, FIX, FXII, FXI), whereas acquired inhibitors are specific to a single factor.

FURTHER READINGS

Caldwell SH et al: Coagulation disorders and hemostasis in liver disease: Pathophysiology and critical assessment of current management. Hepatology 44:1039, 2006[PMID: 17006940]

Franchini M, Mannucci PM: Co-morbidities and quality of life in elderly persons with haemophilia. Br J Haematol 148:522, 2010[PMID: 19958358]

Hoyer LW: Hemophilia A. N Engl J Med 330:39, 1994

Key NS, Negrier C: Coagulation factor concentrates: Past, present, and future. Lancet 370:439, 2007[PMID: 17679021]

Kessler CM, Ortel TL: Recent developments in topical thrombins. Thromb Haemost 101:15, 2009

Levi M, Opal SM: Coagulation abnormalities in critically ill patients. Critical Care 10:222, 2006[PMID: 16879728]

Mannucci PM et al: Recessively inherited coagulation disorders. Blood 104:1243, 2004[PMID: 15138162]

Stafford DW: The vitamin K cycle. J Thromb Haemost 3:1873, 2005[PMID: 16102054]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OVERVIEW OF THROMBOSIS

General Overview

Thrombosis is defined as "hemostasis in the wrong place,"[1] and it is a major cause of morbidity and mortality in a wide range of arterial and venous diseases and patient populations. In 2009 in the United States, an estimated 785,000 people had a new coronary thrombotic event and about 470,000 had a recurrent ischemic episode. Each year, approximately 795,000 people have a new or recurrent stroke. Annually, more than 200,000 new cases of venous thromboembolism are found; 30% of these individuals die within 30 days, with one-fifth having sudden death due to pulmonary embolism.

[1] Macfarlane RG. Haemostasis: Introduction. Brit Med Bull 33:183, 1977.

In the nondiseased state, physiologic hemostasis reflects a delicate interplay between factors that promote and inhibit blood clotting, favoring the former. This response is crucial as it prevents uncontrolled hemorrhage and exsanguination following injury. In specific settings, the same processes that regulate normal hemostasis can cause pathological thrombosis leading to arterial or venous occlusion. Importantly, many commonly used therapeutic interventions may also alter the thrombotic –hemostatic balance adversely.

Hemostasis and thrombosis primarily involve the interplay among three factors: the vessel wall, coagulation proteins, and platelets. Many prevalent acute vascular diseases are due to thrombus formation within a vessel, including myocardial infarction, thrombotic cerebrovascular events, and venous thrombosis. Although the end result is vessel occlusion and tissue ischemia, the pathophysiologic processes governing these pathologies have similarities as well as distinct differences. While many of the pathways regulating thrombus formation are similar to those that regulate hemostasis, the processes triggering thrombosis and, often, perpetuating the thrombus are distinct. In venous thrombosis, primary hypercoagulable states reflecting defects in the proteins governing coagulation and/or fibrinolysis or secondary hypercoagulable states involving abnormalities of blood vessels and blood flow lead to thrombosis. By contrast, arterial thrombosis is highly dependent upon the state of the vessel wall, the platelet, and factors related to blood flow.

ARTERIAL THROMBOSIS

Overview of Arterial Thrombosis

In arterial thrombosis, the platelet and abnormalities of the vessel wall typically play a key role in vessel occlusion. Arterial thrombus forms via a series of sequential steps in which platelets adhere to the vessel wall, additional platelets are recruited, and thrombin is activated. The regulation of platelet adhesion, activation, aggregation, and recruitment will be described in detail below. In addition, while the primary function of platelets is regulation of hemostasis, our understanding of their role in other processes, such as immunity and inflammation, continues to grow.

Arterial Thrombosis and Vascular Disease

Arterial thrombosis is a major cause of morbidity and mortality both in the United States and, increasingly, worldwide. Coronary heart disease is estimated to cause about 1 of every 5 deaths in the United States. In addition to the 785,000 Americans who will have a new coronary event, an additional 195,000 silent first myocardial infarctions are projected to occur annually. Each year, about 795,000 people experience a new or recurrent stroke, although not all are caused by thrombotic occlusion of the vessel. Approximately 610,000 strokes are first events and 185,000 are recurrent events; it is estimated that 1 of every 18 deaths in the United States is due to stroke.

The Platelet

Many processes in platelets have parallels with other cell types, such as the presence of specific receptors and signaling pathways; however, unlike most cells, platelets lack a nucleus and are unable to adapt to changing biologic settings by altered gene transcription. Platelets sustain limited protein synthetic capacity from megakaryocyte-derived mRNA. Most of the molecules needed to respond to various stimuli, however, are maintained in storage granules and membrane compartments.

Platelets are disc-shaped, very small, anucleate cells (1–5 μm in diameter) that circulate in the blood at concentrations of 200–400,000/μL, with an average lifespan of 7–10 days. Platelets are derived from megakaryocytes, polyploidal hematopoietic cells found in the bone marrow. The primary regulator of

platelet formation is thrombopoietin (TPO). The precise mechanism by which megakaryocytes produce and release fully formed platelets is unclear, but the process likely involves formation of proplatelets, pseudopod-like structures generated by the evagination of the cytoplasm from which platelets bud. Platelet granules are synthesized in megakaryocytes before thrombopoiesis and contain an array of prothrombotic, proinflammatory, and antimicrobial mediators. The two major types of platelet granules, alpha and dense, are distinguished by their size, abundance, and content. Alpha-granules contain soluble coagulation proteins, adhesion molecules, growth factors, integrins, cytokines, and inflammatory modulators. Platelet dense-granules are smaller than alpha-granules and less abundant. While alpha-granules contain proteins that may be more important in the inflammatory response, dense-granules contain high concentrations of small molecules, including ADP and serotonin, that influence platelet aggregation.

PLATELET ADHESION

(See Fig. 117-1) The formation of a thrombus is initiated by the adherence of platelets to the damaged vessel wall. Damage exposes subendothelial components responsible for triggering platelet reactivity, including collagen, von Willebrand factor, fibronectin, and other adhesive proteins, such as vitronectin and thrombospondin. The hemostatic response may vary, depending on the extent of damage, the specific proteins exposed, as well as flow conditions. Certain proteins are expressed on the platelet surface that subsequently regulate collagen-induced platelet adhesion, particularly under flow conditions, and include glycoprotein (GP) IV, GPVI, and the integrin $\alpha_2\beta_1$. The platelet GPIb-IX-V complex adhesive receptor is central both to platelet adhesion and to the initiation of platelet activation. Damage to the blood vessel wall exposes subendothelial von Willebrand factor and collagen to the circulating blood. The GPIb-IX-V complex binds to the exposed von Willebrand factor, causing platelets to adhere (Fig. 117-1). In addition, the engagement of the GPIb-IX-V complex with ligand induces signaling pathways that lead to platelet activation. von Willebrand factor–bound GPIb-IX-V promotes a calcium-dependent conformational change in the GPIIb/IIIa receptor, transforming it from an inactive low-affinity state to an active high-affinity receptor for fibrinogen.

Figure 117-1

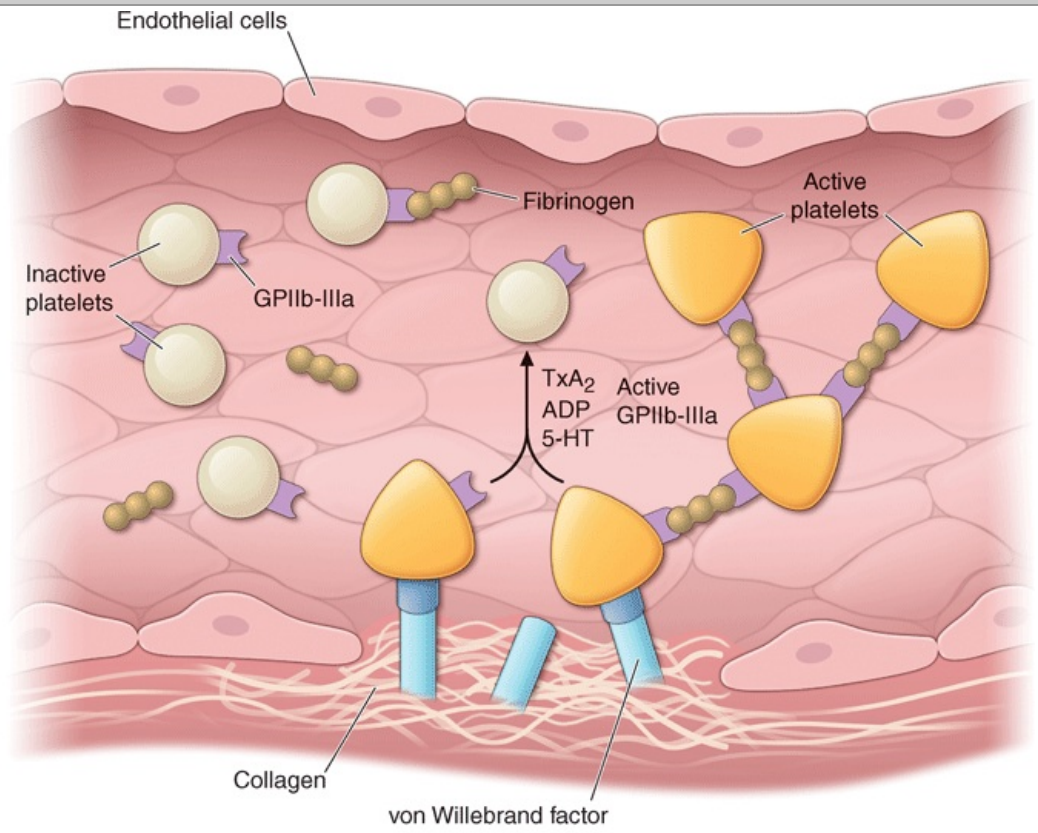

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Platelet activation and thrombosis. Platelets circulate in an inactive form in the vasculature. Damage to the endothelium and/or external stimuli activates platelets that adhere to the exposed subendothelial von Willebrand factor and collagen. This adhesion leads to activation of the platelet, shape change, and the synthesis and release of TxA_2, 5-HT, and ADP. Platelet stimuli causes conformational change in the platelet integrin glycoprotein IIb/IIIa receptor leading to the high-affinity binding of fibrinogen and the formation of a stable platelet thrombus (TxA_2, thromboxane; 5-HT,

serotonin).

PLATELET ACTIVATION

The activation of platelets is controlled by a variety of surface receptors that regulate various functions in the activation process. Platelet receptors are stimulated by a wide variety of agonists and adhesive proteins that result in variable degrees of activation. In general terms, the stimulation of platelet receptors triggers two specific processes: (1) activation of internal signaling pathways that lead to further platelet activation and granule release and (2) the capacity of the platelet to bind to other adhesive proteins/platelets. Both of these processes contribute to the formation of a thrombus.

Many families and subfamilies of receptors are found on platelets that regulate a variety of platelet functions. These include the seven transmembrane receptor family, which is the main agonist-stimulated receptor family. Several seven transmembrane receptors are found on platelets, including the ADP receptors, prostaglandin receptors, lipid receptors, and chemokine receptors. Receptors for thrombin comprise the major seven transmembrane receptors found on platelets. Among this last group, the first identified was the protease activation receptor 1 (PAR1). The PAR class of receptors has a distinct mechanism of activation that involves specific cleavage of the N-terminus of thrombin, which, in turn, acts as a ligand for the receptor. Other PAR receptors are present on platelets, including PAR2 (not activated by thrombin) and PAR4. Adenosine receptors are responsible for transduction of ADP-induced signaling events, which are initiated by the binding of ADP to purinergic receptors on the platelet surface. There are several distinct ADP receptors, classified as $P2X_1$, $P2Y_1$, and $P2Y_{12}$ (Fig. 117-2). The activation of both the $P2Y_{12}$ and $P2Y_1$ receptors is essential for ADP-induced platelet aggregation. The thienopyridine derivatives, clopidogrel and prasugrel, are clinically utilized inhibitors of ADP-induced platelet aggregation.

Figure 117-2

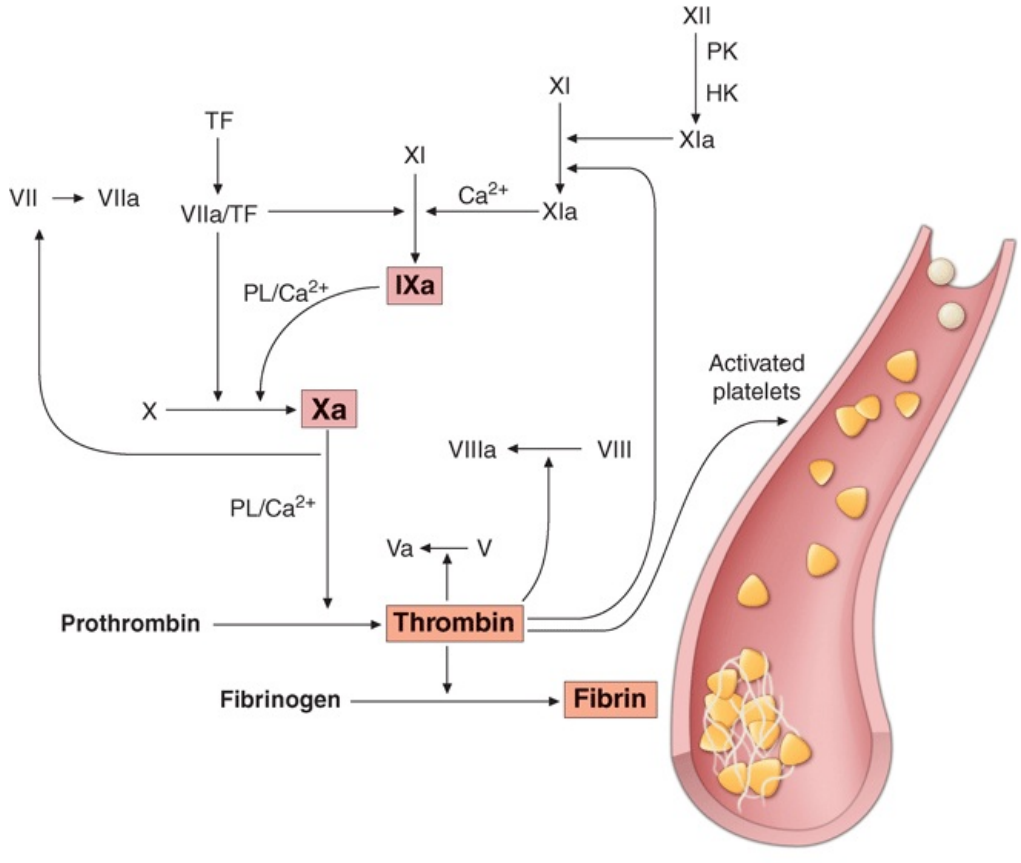

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Summary of the coagulation pathways. Specific coagulation factors ("a" indicates activated form) are responsible for the conversion of soluble plasma fibrinogen into insoluble fibrin. This process occurs via a series of linked reactions in which the enzymatically active product subsequently converts the downstream inactive protein into an active serine protease. In addition, the activation of thrombin leads to stimulation of platelets HK, high-molecular-weight kininogen; PK, prekallikrein; TF, tissue factor.

PLATELET AGGREGATION

Activation of platelets results in a rapid series of signal transduction events, including tyrosine kinase, serine/threonine kinase, and lipid kinase activation. In unstimulated platelets, the major platelet integrin GPIIb/IIIa is maintained in an inactive conformation and functions as a low-affinity adhesion receptor for fibrinogen. This integrin is unique as it is only expressed on platelets. After stimulation, the interaction between fibrinogen and GPIIb/IIIa forms intercellular connections between platelets leading to the formation of a platelet aggregate (Fig. 117-1). A calcium-sensitive conformational change in the extracellular domain of GPIIb/IIIa enables the high-affinity binding of soluble plasma fibrinogen as a result of a complex network of inside-out signaling events. The GPIIb/IIIa receptor serves as a bidirectional conduit with GPIIb/IIIa–mediated signaling (outside-in) occurring immediately after the binding of fibrinogen. This leads to additional intracellular signaling that further stabilizes the platelet aggregate and transforms platelet aggregation from a reversible to an irreversible process (inside-out).

The Role of Platelets and Thrombosis in Inflammation

Inflammation plays an important role during the acute thrombotic phase of acute coronary syndromes. Patients with acute coronary syndromes have not only increased interactions between platelets (homotypic aggregates), but also increased interactions between platelets and leukocytes (heterotypic aggregates) detectable in circulating blood. These latter aggregates form when platelets are activated and adhere to circulating leukocytes. Platelets bind via P-selectin (CD62P) expressed on the surface of activated platelets to the leukocyte receptor, P-selectin glycoprotein ligand 1 (PSGL-1). This association leads to increased expression of CD11b/CD18 (Mac-1) on leukocytes, which itself supports interactions with platelets partially via bivalent fibrinogen linking this integrin with its platelet surface counterpart, GPIIb/IIIa. Platelet surface P-selectin also induces the expression of tissue factor on monocytes, which promotes fibrin formation.

In addition to platelet–monocyte aggregates, the immunomodulator, soluble CD40 ligand (CD40L or CD154), also reflects a link between thrombosis and inflammation. The CD40 ligand is a trimeric transmembrane protein of the tumor necrosis factor family and, with its receptor CD40, is an important contributor to the inflammatory process leading both to thrombosis and atherosclerosis. While many immunologic and vascular cells have been found to express CD40 and/or CD40 ligand, in platelets, CD40 ligand is rapidly translocated to the surface after stimulation and is upregulated in the newly formed thrombus. The surface-expressed CD40 ligand is cleaved from the platelet to generate a soluble fragment (soluble CD40 ligand).

Links have also been established among platelets, infection, immunity, and inflammation. Bacterial infections are associated with a transient increase in the risk of acute thrombotic events, such as acute myocardial infarction and stroke. In addition, platelets contribute significantly to the pathophysiology and high mortality rates of sepsis. The expression, functionality, and signaling pathways of toll-like receptors (TLRs) have been established in platelets. Stimulation of platelet TLR2 directly activates the platelet's thrombotic and inflammatory responses, and live bacteria induce a proinflammatory response in platelets in a TLR2-dependent manner, suggesting a mechanism by which specific bacteria and bacterial components can directly activate platelet-dependent thrombosis.

Genetics of Arterial Thrombosis

Some studies have associated arterial thrombosis with genetic variants (Table 117–2A); however, in the area of genetic variability and platelet function, studies have primarily dealt with pharmacogenetics, the field of pharmacology dealing with the interindividual variability in drug response based on genetic determinants (Table 117–1). This focus has been driven by the wide variability among individuals in terms of response to antithrombotic drugs and the lack of a common explanation for this variance. The best described is the issue of "aspirin resistance," although heterogeneity for other antithrombotics (e.g., clopidogrel) has also been extensively examined. Primarily, platelet-dependent genetic determinants have been defined at the level of (a) drug effect, (b) drug compliance, and (c) drug metabolism. Many candidate platelet genes have been studied for their interaction with antiplatelet and antithrombotic agents.

Table 117–1 Genetic Variation and Pharmacogenetic Responses to Platelet Inhibitors

Potential Gene Altered	Target Therapeutic Class	Specific Drug
P2Y1 and *P2Y12CYP2C19, CYP3A4, CYP3A5*	ADP receptor inhibitors	Clopidogrel, prasugrel
COX1, COX2	Cyclooxygenase inhibitors	Aspirin
PIA1/A2 *INTB3, GPIbA*	Receptor inhibitors: Glycoprotein IIb-IIIa receptor inhibitors	Abciximab, eptifibatide, tirofiban

Table 117–2 Heritable Causes of Arterial and Venous Thrombosis

A. Arterial Thrombosis

Platelet Receptors
ß3 and α2 integrins
PlA2 polymorphism
Fc(gamma)RIIA
GPIV T13254C polymorphism
GPIb
Thrombin receptor PAR-1-5061 → D
Redox Enzymes
Plasma glutathione peroxidase
H2 promoter haplotype
Endothelial nitric oxide synthase
–786T/C, –922A/G, –1468T/A
Paraoxonase
–107T allele, 192R allele
Homocysteine
Cystathionine ß-synthase 833T → C
5,10-methylene tetrahydrofolate reductase (MTHFR) 677C → T
B. Venous Thrombosis
Procoagulant Proteins
Fibrinogen
–455G/A, –854G/A
Prothrombin (20210G → A)
Protein C Anticoagulant Pathway
Factor V Leiden: 1691G → A (Arg506Gln)
Thrombomodulin 1481C → T (Ala455Val)
Fibrinolytic Proteins with Known Polymorphisms
Tissue plasminogen activator (tPA)
7351C/T, 20 099T/C in exon 6, 27 445T/A in intron 10
Plasminogen activator inhibitor (PAI-1)
4G/5G insertion/deletion polymorphism at position –675
Homocysteine
Cystathionine ß-synthase 833T → C
5,10-methylene tetrahydrofolate reductase (MTHFR) 677C → T

Many patients have an inadequate response to the inhibitory effects of aspirin. Heritable factors contribute to the variability; however, ex vivo tests of residual platelet responsiveness after aspirin administration have not provided firm evidence for a pharmacogenetic interaction between aspirin and *COX1* or other relevant platelet receptors. As such, currently, there is no clinical indication for genotyping to optimize aspirin's antiplatelet efficiency. For the platelet P2Y12 receptor inhibitor clopidogrel, additional data suggest that genetics may affect the drug's responsiveness and utility. The responsible genetic variant appears not to be the expected P2Y12 receptor but an enzyme responsible for drug metabolism. Clopidogrel is a prodrug, and liver metabolism by specific cytochrome P450 enzymes is required for activation. The genes encoding the CYP-dependent oxidative steps are polymorphic, and carriers of specific alleles of the CYP2C19 and CYP3A4 loci have increased platelet aggregability. Increased platelet activity has also been specifically associated with the CYP2C19*2 allele, which causes loss of platelet function in select patients. As these are common genetic variants, this observation has been shown to be clinically relevant in large studies.

VENOUS THROMBOSIS

Overview of Venous Thrombosis

Coagulation is the process by which thrombin is activated and soluble plasma fibrinogen is converted into insoluble fibrin. These steps account for both normal hemostasis and the pathophysiologic processes influencing the development of venous thrombosis. The primary forms of venous thrombosis are deep-vein thrombosis (DVT) in the extremities and the subsequent embolization to the lungs (pulmonary embolism), referred to together as venous thromboembolic disease. Venous thrombosis occurs due to heritable causes (Table 117–2B) and acquired causes (Table 117–3).

Table 117–3 Acquired Causes of Venous Thrombosis
Surgery
Neurosurgery
Major abdominal surgery
Malignancy
Antiphospholipid syndrome
Other
Trauma
Pregnancy
Long-haul travel
Obesity
Oral contraceptives/hormone replacement
Myeloproliferative disorders Polycythemia vera

Deep Venous Thrombosis and Pulmonary Embolism

More than 200,000 new cases of venous thromboembolism occur each year. Of these cases, 30% die within 30 days and one-fifth suffer sudden death owing to pulmonary embolism; 30% go on to develop recurrent venous thromboembolism within 10 years. Data from the ARIC study reported a 9% 28-day fatality rate from deep venous thrombosis and a 15% fatality rate from pulmonary embolism. Pulmonary embolism in the setting of cancer has a 25% fatality rate. The mean incidence of first DVT in the general population is 5 per 10,000 person-years; the incidence is similar in males and females and increases dramatically with age from 2 to 3 per 10,000 person-years at 30–49 years of age to 20 at 70–79 years of age.

Overview of the Coagulation Cascade and Its Role in Venous Thrombosis

Coagulation is defined as the formation of fibrin by a series of linked enzymatic reactions in which each reaction product converts the subsequent inactive zymogen into an active serine protease (Fig. 117-2). This coordinated sequence is called the coagulation cascade and is a key mechanism for regulating hemostasis. Central to the function of the coagulation cascade is the principle of amplification: owing to a series of linked enzymatic reactions, a small stimulus can lead to much greater quantities of fibrin, the end product that prevents hemorrhage at the site of vascular injury.

The coagulation cascade is primarily initiated by vascular injury exposing tissue factor to blood components (Fig. 117-2). Tissue factor may also be found in bloodborne cell-derived microparticles and, under pathophysiologic conditions, in leukocytes or platelets. Plasma factor VII (FVII) is the ligand for and is activated (FVIIa) by binding to tissue factor exposed at the site of vessel damage. The binding of FVII/VIIa to tissue factor activates the downstream conversion of factor X (FX) to active FX (FXa). In an alternative reaction, the FVII/FVIIa–tissue factor complex initially converts FIX to FIXa, which then activates FX in conjunction with its cofactor factor VIII (FVIIIa). Factor Xa with its cofactor FVa converts prothrombin to thrombin, which then converts soluble plasma fibrinogen to insoluble fibrin, leading to clot or thrombus formation. Thrombin also activates FXIII to FXIIIa, a transglutaminase that covalently cross-links and stabilizes the fibrin clot.

Several antithrombotic factors also regulate coagulation; these include antithrombin, tissue factor pathway inhibitor (TFPI), heparin cofactor II, and protein C/protein S. Under normal conditions, these factors limit the production of thrombin to prevent the perpetuation of coagulation and thrombus formation. Typically, after the clot has caused occlusion at the damaged site and begins to expand toward adjacent uninjured vessel segments, the anticoagulant reactions governed by the normal endothelium become pivotal in limiting the extent of this hemostatically protective clot.

Risk Factors for Venous Thrombosis

The risk factors for venous thrombosis are primarily related to hypercoagulability, which can be genetic (Table 117–2) or acquired, or due to immobilization and venous stasis. Independent predictors for recurrence include increasing age, obesity, malignant neoplasm, and acute extremity paresis. Often, multiple risk factors are present in a single individual. Significant risk is incurred by major orthopedic, abdominal, or neurologic surgeries. Moderate risk is promoted by prolonged bedrest; certain types of cancer, pregnancy, hormone replacement therapy, or oral contraceptive use; and other sedentary conditions such as long-distance plane travel. It has been reported that the risk of developing a venous thromboembolic event doubles after air travel lasting 4 h, although the absolute risk remains low (1 in 6000). The relative risk of venous thromboembolism among pregnant or postpartum women is 4.3, and the overall incidence (absolute risk) is 199.7 per 100,000 woman-years.

Genetics of Venous Thrombosis

(See Table 117–2) Less common causes of venous thrombosis are those due to genetic variants. These abnormalities include loss-of-function mutations of endogenous anticoagulants as well as gain-of-function mutations of procoagulant proteins. Heterozygous antithrombin deficiency and homozygosity of the factor V Leiden mutation significantly increase the risk of venous thrombosis. While homozygousprotein C or protein S deficiencies are rare and may lead to fatal purpura fulminans, heterozygous deficiencies are associated with a moderate risk of thrombosis. Activated protein C impairs coagulation by proteolytic degradation of FVa. Patients resistant to the activity of activated protein C may have a point mutation in the FV gene located on chromosome 1, a mutant denoted factor V Leiden. Mildly increased risk has been attributed to elevated levels of procoagulant factors, as well as low levels of tissue factor pathway inhibitor. Polymorphisms of methylene tetrahydrofolate reductase as well as hyperhomocysteinemia have been shown to be independent risk factors for venous thrombosis, as well as arterial vascular disease; however, many of the initial descriptions of genetic variants and their associations with thromboembolism are being questioned in larger, more current studies.

Fibrinolysis and Thrombosis

Specific abnormalities in the fibrinolytic system have been associated with enhanced thrombosis. Factors such as elevated levels of tissue plasminogen activator (tPA) and plasminogen activator inhibitor type 1 (PAI-1) have been associated with decreased fibrinolytic activity and an increased risk of arterial thrombotic disease. Specific genetic variants have been associated with decreased fibrinolytic activity, including the 4G/5G insertion/deletion polymorphism in the (plasminogen activator type 1) PAI-1 gene. Additionally, the 311-bp Alu insertion/deletion in tPA's intron 8 has been associated with enhanced thrombosis; although genetic abnormalities have not been associated consistently with altered function or tPA levels, raising questions about the relevant pathophysiologic mechanism. Thrombin-activatable fibrinolysis inhibitor (TAFI) is a carboxypeptidase that regulates fibrinolysis; elevated plasma TAFI levels have been associated with an increased risk of both deep-vein thrombosis and cardiovascular disease.

The metabolic syndrome also is accompanied by altered fibrinolytic activity. This syndrome, which comprises abdominal fat (central obesity), altered glucose and insulin metabolism, dyslipidemia, and hypertension, has been associated with atherothrombosis. The mechanism for enhanced thrombosis appears to be due both to altered platelet function and to a procoagulant and hypofibrinolytic state. One of the most frequently documented prothrombotic abnormalities reported in this syndrome is an increase in plasma levels of PAI-1.

THE DISTINCTION BETWEEN ARTERIAL AND VENOUS THROMBOSIS

Although there is overlap, venous and arterial thrombosis are initiated differently and clot formation progresses by somewhat distinct pathways. In the setting of stasis or states of hypercoagulability, venous thrombosis is activated with the initiation of the coagulation cascade primarily due to exposure of tissue factor; this leads to the formation of thrombin and the subsequent conversion of fibrinogen to fibrin. In the artery, thrombin formation also occurs, but thrombosis is primarily promoted by the adhesion of platelets to an injured vessel and stimulated by exposed extracellular matrix (Figs. 117-1 and 117-2). There is wide variation in individual responses to vascular injury, an important determinant of which is the predisposition an individual has to arterial or venous thrombosis. This concept has been supported indirectly in prothrombotic animal models in which there is poor correlation between the propensity to develop venous versus arterial thrombosis.

Despite considerable progress in understanding the role of hypercoagulable states in venous thromboembolic disease, the contribution of hypercoagulability to arterial vascular disease is much less well understood. While specific thrombophilic conditions, such as factor V Leiden and the prothrombin G20210A mutation, are risk factors for DVT, pulmonary embolism, and other venous thromboembolic events, their contribution to arterial thrombosis is less well defined. In fact, to the contrary, many of these thrombophilic factors have not been found to be clinically important risk factors for arterial thrombotic events, such as acute coronary syndromes.

Clinically, although the pathophysiology is distinct, arterial and venous thrombosis do share common risk factors, including age, obesity, cigarette smoking, diabetes mellitus, arterial hypertension, hyperlipidemia, and metabolic syndrome. Select genetic variants, including those of the glutathione peroxidase gene, have also been associated with arterial and venous thrombo-occlusive disease. Importantly, arterial and venous thrombosis may both be triggered by pathophysiologic stimuli responsible for activating inflammatory and oxidative pathways.

The diagnosis and treatment of ischemic heart disease are discussed in Chap. 243. Stroke diagnosis and management are discussed in Chap. 275. The diagnosis and management of deep venous thrombosis and pulmonary embolus are discussed in Chap. 262.

ACKNOWLEDGMENT

The authors would like to thank Hannah Iafrati for her assistance with the figures.

FURTHER READINGS

Alexander K, Peterson ED: Managing the risks of antithrombotics and platelet inhibitors. Circulation 121:1960, 2010[PMID: 20439812]

Daví G, Patrono C: Platelet activation and atherothrombosis. N Engl J Med 13;357:2482, 2007[PMID: 21498593]

Freedman JE: Translational therapeutics at the platelet vascular interface: oxidative stress and platelets. Arterioscler Thromb Vasc Biol 28:s11, 2008

Furie B, Furie BC: Mechanisms of thrombus formation. N Engl J Med 28;359:938, 2008[PMID: 18753650]

Gurbel PA, Tantry US: Combination antithrombotic therapies. Circulation 121:569, 2010[PMID: 20124137]

Lloyd-Jones D et al. Heart disease and stroke statistics–2009 update. A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 119:e21, 2009

Tapson VF: Acute pulmonary embolism. N Engl J Med. 6;358:1037, 2008[PMID: 18322285]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ANTIPLATELET, ANTICOAGULANT, AND FIBRINOLYTIC DRUGS: INTRODUCTION

Arterial and venous thromboses are major causes of morbidity and mortality rates. Arterial thrombosis is the most common cause of acute myocardial infarction (MI), ischemic stroke, and limb gangrene, whereas deep-vein thrombosis (DVT) leads to pulmonary embolism (PE), which can be fatal, and to the postphlebitic syndrome. Most arterial thrombi are superimposed on disrupted atherosclerotic plaque because plaque rupture exposes thrombogenic material in the plaque core to the blood. This material then triggers platelet aggregation and fibrin formation, which results in the generation of a platelet-rich thrombus that can temporarily or permanently occlude blood flow. In contrast to arterial thrombi, venous thrombi rarely form at sites of obvious vascular disruption. Although they can develop after surgical trauma to veins or secondary to indwelling venous catheters, venous thrombi usually originate in the valve cusps of the deep veins of the calf or in the muscular sinuses, where they are triggered by stasis. Sluggish blood flow in these veins reduces the oxygen supply to the avascular valve cusps. Endothelial cells lining these valve cusps become activated and express adhesion molecules on their surface. Tissue factor–bearing leukocytes and microparticles adhere to these activated cells and induce coagulation. Local thrombus formation is exacerbated by reduced clearance of activated clotting factors as a result of impaired blood flow. If the thrombi extend into more proximal veins of the leg, thrombus fragments can dislodge, travel to the lungs, and produce a PE.

Arterial and venous thrombi are composed of platelets and fibrin, but the proportions differ. Arterial thrombi are rich in platelets because of the high shear in the injured arteries. In contrast, venous thrombi, which form under low shear conditions, contain relatively few platelets and are predominantly composed of fibrin and trapped red cells. Because of the predominance of platelets, arterial thrombi appear white, whereas venous thrombi are red in color, reflecting the trapped red cells.

Antithrombotic drugs are used for prevention and treatment of thrombosis. Targeting the components of thrombi, these agents include (1) antiplatelet drugs, (2) anticoagulants, and (3) fibrinolytic agents (Fig. 118-1). With the predominance of platelets in arterial thrombi, strategies to inhibit or treat arterial thrombosis focus mainly on antiplatelet agents, although, in the acute setting, they often include anticoagulants and fibrinolytic agents. Anticoagulants are the mainstay of prevention and treatment of venous thromboembolism because fibrin is the predominant component of venous thrombi. Antiplatelet drugs are less effective than anticoagulants in this setting because of the limited platelet content of venous thrombi. Fibrinolytic therapy is used in selected patients with venous thromboembolism. For example, patients with massive or submassive PE can benefit from systemic or catheter-directed fibrinolytic therapy. The latter can also be used as an adjunct to anticoagulants for treatment of patients with extensive iliofemoral-vein thrombosis.

Figure 118-1

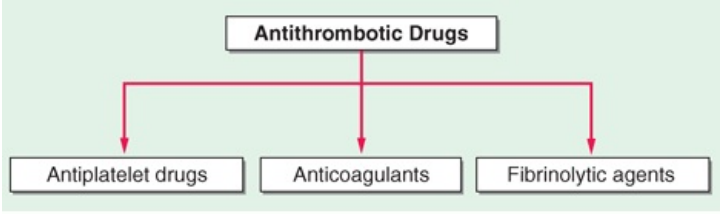

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Classification of antithrombotic drugs.

ANTIPLATELET DRUGS

Role of Platelets in Arterial Thrombosis

In healthy vasculature, circulating platelets are maintained in an inactive state by nitric oxide (NO) and prostacyclin released by endothelial cells lining the blood vessels. In addition, endothelial cells also express CD39 on their surface, a membrane-associated ecto-adenosine diphosphatase (ADPase) that degrades ADP released from activated platelets. When the vessel wall is damaged, release of these substances is impaired and subendothelial matrix is exposed. Platelets adhere to exposed collagen via

$\alpha_2\beta_1$ and glycoprotein (GP)V1 and to von Willebrand factor (vWF) via GPIbα and GPIIb/IIIa ($\alpha_{IIb}\beta_3$)—receptors that are constitutively expressed on the platelet surface. Adherent platelets undergo a change in shape, secrete ADP from their dense granules, and synthesize and release thromboxane A_2. Released ADP and thromboxane A_2, which are platelet agonists, activate ambient platelets and recruit them to the site of vascular injury (Fig. 118-2).

Figure 118-2

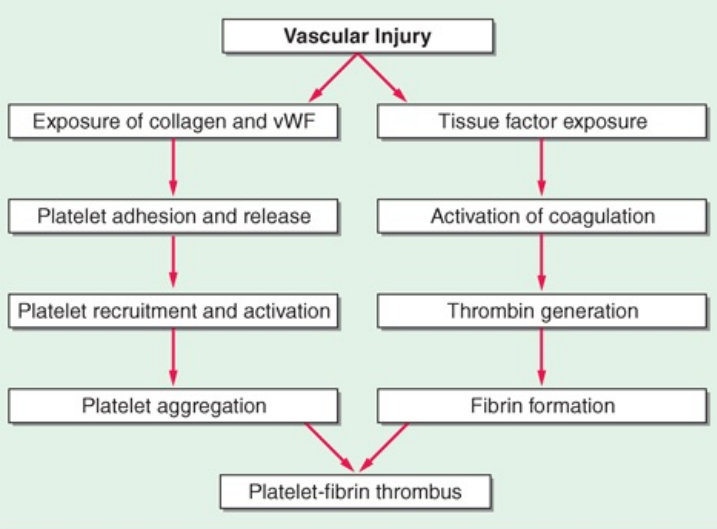

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coordinated role of platelets and the coagulation system in thrombogenesis. Vascular injury simultaneously triggers platelet activation and aggregation and activation of the coagulation system. Platelet activation is initiated by exposure of subendothelial collagen and von Willebrand factor (vWF), onto which platelets adhere. Adherent platelets become activated and release ADP and thromboxane A_2, platelet agonists that activate ambient platelets and recruit them to the site of injury. When platelets are activated, glycoprotein IIb/IIIa on their surface undergoes a conformational change that enables it to ligate fibrinogen and mediate platelet aggregation. Coagulation is triggered by tissue factor exposed at the site of injury. Tissue factor triggers thrombin generation. As a potent platelet agonist, thrombin amplifies platelet recruitment to the site of injury. Thrombin also converts fibrinogen to fibrin, and the fibrin strands then weave the platelet aggregates together to form a platelet/fibrin thrombus.

Disruption of the vessel wall also exposes tissue factor–expressing cells to the blood. Tissue factor initiates coagulation. Activated platelets potentiate coagulation by binding clotting factors and supporting the assembly of activation complexes that enhance thrombin generation. In addition to converting fibrinogen to fibrin, thrombin also serves as a potent platelet agonist and recruits more platelets to the site of vascular injury.

When platelets are activated, GPIIb/IIIa, the most abundant receptor on the platelet surface, undergoes a conformational change that enables it to bind fibrinogen and, under high shear conditions, vWF. Divalent fibrinogen or multivalent vWF molecules bridge adjacent platelets together to form platelet aggregates. Fibrin strands, generated through the action of thrombin, then weave these aggregates together to form a platelet/fibrin mesh.

Antiplatelet drugs target various steps in this process. The commonly used drugs include aspirin, thienopyridines (clopidogrel, prasugrel, and ticlopidine), dipyridamole, and GPIIb/IIIa antagonists.

Aspirin

The most widely used antiplatelet agent worldwide is aspirin. As a cheap and effective antiplatelet drug, aspirin serves as the foundation of most antiplatelet strategies.

Mechanism of Action

Aspirin produces its antithrombotic effect by irreversibly acetylating and inhibiting platelet cyclooxygenase (COX)-1 (Fig. 118-3), a critical enzyme in the biosynthesis of thromboxane A_2. At high doses (~1 g/d), aspirin also inhibits COX-2, an inducible COX isoform found in endothelial cells and inflammatory cells. In endothelial cells, COX-2 initiates the synthesis of prostacyclin, a potent vasodilator and inhibitor of platelet aggregation.

Figure 118-3

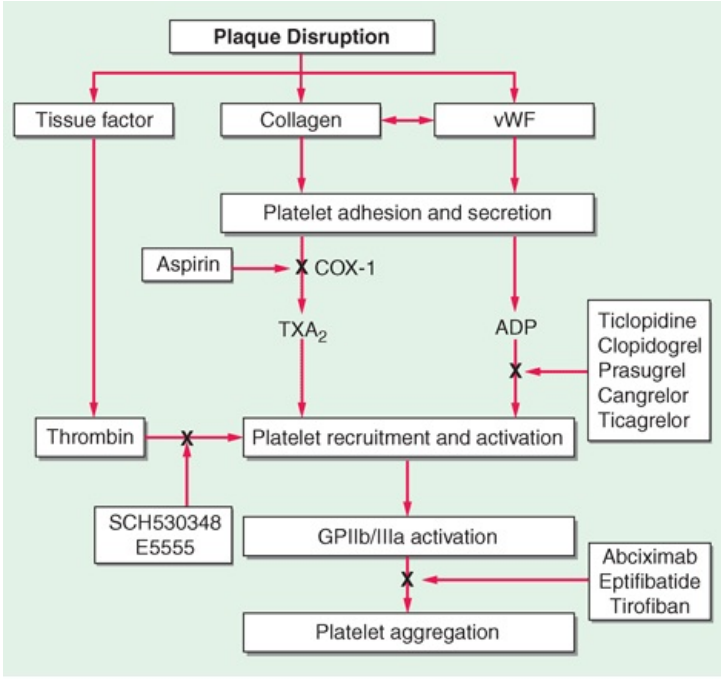

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Site of action of antiplatelet drugs. Aspirin inhibits thromboxane A_2 (TXA_2) synthesis by irreversibly acetylating cyclooxygenase-1 (COX-1). Reduced TXA_2 release attenuates platelet activation and recruitment to the site of vascular injury. Ticlopidine, clopidogrel, and prasugrel irreversibly block $P2Y_{12}$, a key ADP receptor on the platelet surface; cangrelor and ticagrelor are reversible inhibitors of $P2Y_{12}$. Abciximab, eptifibatide, and tirofiban inhibit the final common pathway of platelet aggregation by blocking fibrinogen and von Willebrand factor binding to activated glycoprotein (GP) IIb/IIIa. SCH530348 and E5555 inhibit thrombin-mediated platelet activation by targeting protease-activated receptor-1 (PAR-1), the major thrombin receptor on human platelets.

Indications

Aspirin is widely used for secondary prevention of cardiovascular events in patients with coronary artery, cerebrovascular, or peripheral vascular disease. Compared with placebo, aspirin produces a 25% reduction in the risk of cardiovascular death, MI, or stroke. Aspirin is also used for primary prevention in patients whose estimated annual risk of MI is >1%, a point where its benefits are likely to outweigh harms. This includes patients older than age 40 years with two or more major risk factors for cardiovascular disease or those older than age 50 years with one or more such risk factors. Aspirin is equally effective in men and women. In men, aspirin mainly reduces the risk of MI, while in women aspirin lowers the risk of stroke.

Dosages

Aspirin is usually administered at doses of 75–325 mg once daily. Higher doses of aspirin are not more effective than lower aspirin doses, and some analyses suggest reduced efficacy with higher doses. Because the side effects of aspirin are dose-related, daily aspirin doses of 75–100 mg are recommended for most indications. When rapid platelet inhibition is required, an initial aspirin dose of at least 160 mg should be given.

Side Effects

Most common side effects are gastrointestinal and range from dyspepsia to erosive gastritis or peptic ulcers with bleeding and perforation. These side effects are dose-related. Use of enteric-coated or buffered aspirin in place of plain aspirin does not eliminate the risk of gastrointestinal side effects. The overall risk of major bleeding with aspirin is 1–3% per year. The risk of bleeding is increased when aspirin is given in conjunction with anticoagulants, such as warfarin. When dual therapy is used, low-dose aspirin should be given (75–100 mg daily). Eradication of *Helicobacter pylori* infection and administration of proton pump inhibitors may reduce the risk of aspirin-induced gastrointestinal bleeding in patients with peptic ulcer disease.

Aspirin should not be administered to patients with a history of aspirin allergy characterized by bronchospasm. This problem occurs in ~0.3% of the general population but is more common in those with chronic urticaria or asthma, particularly in individuals with nasal polyps or chronic rhinitis. Hepatic and renal toxicity are observed with aspirin overdose.

Aspirin Resistance

Clinical aspirin resistance is defined as the failure of aspirin to protect patients from ischemic vascular events. This is not a helpful definition because it is made after the event occurs. Furthermore, it is not realistic to expect aspirin, which only blocks thromboxane A_2–induced platelet activation, to prevent all vascular events.

Aspirin resistance has also been described biochemically as failure of the drug to produce its expected inhibitory effects on tests of platelet function, such as thromboxane A_2 synthesis or arachidonic acid –induced platelet aggregation. However, the tests of platelet function used for diagnosis of biochemical aspirin resistance have not been well standardized. Furthermore, these tests are not proven to identify patients at risk of recurrent vascular events. In addition, resistance is not reversed by either giving higher doses of aspirin or adding other antiplatelet drugs. Thus, testing for aspirin resistance remains a research tool.

Thienopyridines

The thienopyridines include ticlopidine, clopidogrel, and prasugrel, drugs that target $P2Y_{12}$, a key ADP receptor on platelets.

Mechanism of Action

The thienopyridines are structurally related drugs that selectively inhibit ADP-induced platelet aggregation by irreversibly blocking $P2Y_{12}$ (Fig. 118-3). Ticlopidine and clopidogrel are prodrugs that require metabolism by the hepatic cytochrome P450 (CYP) enzyme system to acquire activity. Consequently, when given in usual doses, their onset of action is delayed for several days. Although prasugrel also is a prodrug that requires metabolic activation, its onset of action is more rapid than that of ticlopidine or clopidogrel, and prasugrel produces greater and more predictable inhibition of ADP-induced platelet aggregation. These characteristics reflect the rapid and complete absorption of prasugrel from the gut and its more efficient activation pathways. Whereas nearly all of the absorbed prasugrel undergoes metabolic activation in the liver, only 15% of absorbed clopidogrel is activated; the remainder is inactivated by esterases.

Indications

Like aspirin, ticlopidine is more effective than placebo at reducing the risk of cardiovascular death, MI, and stroke in patients with atherosclerotic disease. Because of its delayed onset of action, ticlopidine is not recommended in patients with acute MI. Ticlopidine was used routinely as an adjunct to aspirin after coronary artery stenting and as an aspirin substitute in those intolerant to aspirin. Because clopidogrel is more potent than ticlopidine and has a better safety profile, clopidogrel has replaced ticlopidine.

When compared with aspirin in patients with recent ischemic stroke, MI, or peripheral arterial disease, clopidogrel reduced the risk of cardiovascular death, MI, and stroke by 8.7%. Therefore, clopidogrel is more effective than aspirin but is also more expensive. In some patients, clopidogrel and aspirin are combined to capitalize on their capacity to block complementary pathways of platelet activation. For example, the combination of aspirin plus clopidogrel is recommended for at least 4 weeks after implantation of a bare metal stent in a coronary artery and longer in those with a drug-eluting stent. Concerns about late in-stent thrombosis with drug-eluting stents have led some experts to recommend long-term use of clopidogrel plus aspirin for this indication.

The combination of clopidogrel and aspirin is also effective in patients with unstable angina. Thus, in 12,562 such patients, the risk of cardiovascular death, MI, or stroke was 9.3% in those randomized to the combination of clopidogrel and aspirin and 11.4% in those given aspirin alone. This 20% relative risk reduction with combination therapy was highly statistically significant. However, combining clopidogrel with aspirin increases the risk of major bleeding to about 2% per year. This bleeding risk persists even if the daily dose of aspirin is ≤100 mg. Therefore, the combination of clopidogrel and aspirin should only be used when there is a clear benefit. For example, this combination has not proven to be superior to clopidogrel alone in patients with acute ischemic stroke or to aspirin alone for primary prevention in those at risk for cardiovascular events.

Prasugrel was compared with clopidogrel in 13,608 patients with acute coronary syndromes who were scheduled to undergo a percutaneous coronary intervention. The incidence of the primary efficacy endpoint, a composite of cardiovascular death, MI, and stroke, was significantly lower with prasugrel than with clopidogrel (9.9% and 12.1%, respectively), mainly reflecting a reduction in the incidence of nonfatal MI. The incidence of stent thrombosis also was significantly lower with prasugrel than with clopidogrel (1.1% and 2.4%, respectively). However, these advantages were at the expense of significantly higher rates of fatal bleeding (0.4% and 0.1%, respectively) and life-threatening bleeding (1.4% and 0.9%, respectively) with prasugrel. Because patients older than age 75 years and those with a history of prior stroke or transient ischemic attack have a particularly high risk of bleeding, prasugrel should generally be avoided in older patients, and the drug is contraindicated in those with a history of cerebrovascular disease. Caution is required if prasugrel is used in patients weighing less than 60 kg or in those with renal impairment.

Dosing

Ticlopidine is given twice daily at a dose of 250 mg. The more potent clopidogrel is given once daily at a dose of 75 mg. Loading doses of clopidogrel are given when rapid ADP receptor blockade is desired. For example, patients undergoing coronary stenting are often given a loading dose of 300 mg, which affects inhibition of ADP-induced platelet aggregation in about 6 h. Loading doses of 600 or 900 mg produce an even more rapid effect. After a loading dose of 60 mg, prasugrel is given once daily at a dose of 10 mg. Patients older than age 75 years or weighing less than 60 kg should receive a lower daily prasugrel dose of 5 mg.

Side Effects

The most common side effects of ticlopidine are gastrointestinal. More serious are the hematologic side effects, which include neutropenia, thrombocytopenia, and thrombotic thrombocytopenic purpura. These side effects usually occur within the first few months of starting treatment. Therefore, blood counts must be carefully monitored when initiating therapy with ticlopidine. Gastrointestinal and hematologic side effects are rare with clopidogrel and prasugrel.

Thienopyridine Resistance

The capacity of clopidogrel to inhibit ADP-induced platelet aggregation varies among subjects. This variability reflects, at least in part, genetic polymorphisms in the CYP isoenzymes involved in the metabolic activation of clopidogrel. Most important of these is *CYP2C19*. Clopidogrel-treated patients with the loss-of-function *CYP2C19*2* allele exhibit reduced platelet inhibition compared with those with the wild-type *CYP2C19*1* allele and experience a higher rate of cardiovascular events. This is important because estimates suggest that up to 25% of whites, 30% of African Americans, and 50% of Asians carry the loss-of-function allele, which would render them resistant to clopidogrel. Even patients with the reduced function *CYP2C19*3*, *4, or *5 alleles may derive less benefit from clopidogrel than those with the full-function *CYP2C19*1* allele. Concomitant administration of clopidogrel and proton pump inhibitors, which are inhibitors of *CYP2C19*, produces a small reduction in the inhibitory effects of clopidogrel on ADP-induced platelet aggregation. The extent to which this interaction increases the risk of cardiovascular events remains controversial.

In contrast to their effect on the metabolic activation of clopidogrel, *CYP2C19* polymorphisms appear to be less important determinants of the activation of prasugrel. Thus, no association was detected between the loss-of-function allele and decreased platelet inhibition or increased rate of cardiovascular events with prasugrel. The observation that genetic polymorphisms affecting clopidogrel absorption or metabolism influence clinical outcomes raises the possibilities that pharmacogenetic profiling may be useful to identify clopidogrel-resistant patients and that point-of-care assessment of the extent of clopidogrel-induced platelet inhibition may help detect patients at higher risk for subsequent cardiovascular events. It is unknown whether administration of higher doses of clopidogrel to such patients will overcome this resistance. Instead, prasugrel or newer $P2Y_{12}$ inhibitors may be better choices for these patients.

Dipyridamole

Dipyridamole is a relatively weak antiplatelet agent on its own, but an extended-release formulation of dipyridamole combined with low-dose aspirin, a preparation known as *Aggrenox*, is used for prevention of stroke in patients with transient ischemic attacks.

Mechanism of Action

By inhibiting phosphodiesterase, dipyridamole blocks the breakdown of cyclic AMP. Increased levels of cyclic AMP reduce intracellular calcium and inhibit platelet activation. Dipyridamole also blocks the uptake of adenosine by platelets and other cells. This produces a further increase in local cyclic AMP levels because the platelet adenosine A_2 receptor is coupled to adenylate cyclase (Fig. 118-4).

Figure 118-4

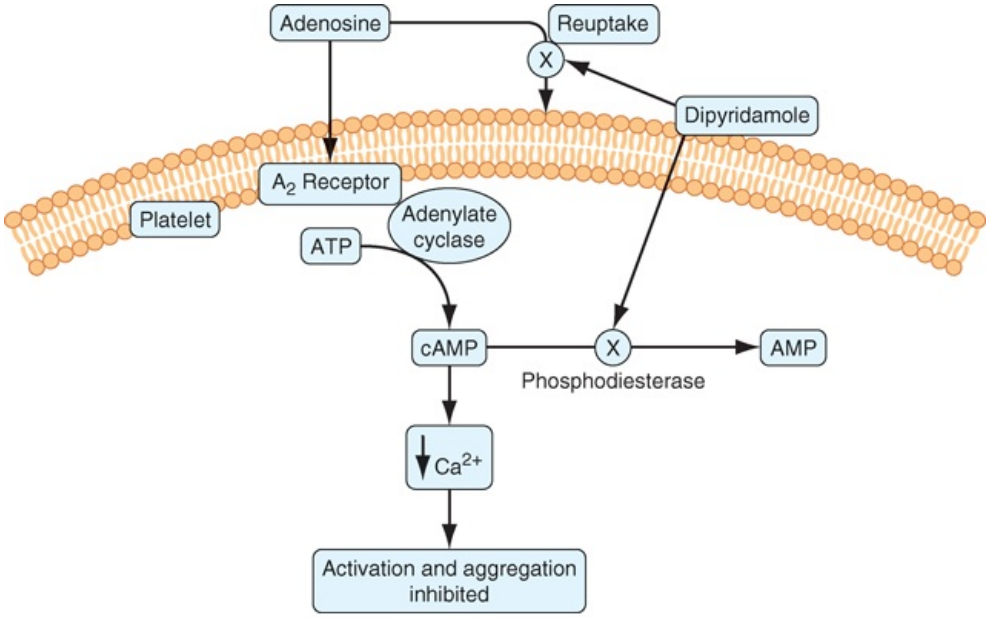

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanism of action of dipyridamole. Dipyridamole increases levels of cyclic AMP in platelets by (1) blocking the reuptake of adenosine and (2) inhibiting phosphodiesterase-mediated cyclic AMP degradation. By promoting calcium uptake, cyclic AMP reduces intracellular levels of calcium. This, in turn, inhibits platelet activation and aggregation.

Dosing

Aggrenox is given twice daily. Each capsule contains 200 mg of extended-release dipyridamole and 25 mg of aspirin.

Side Effects

Because dipyridamole has vasodilatory effects, it must be used with caution in patients with coronary artery disease. Gastrointestinal complaints, headache, facial flushing, dizziness, and hypotension can also occur. These symptoms often subside with continued use of the drug.

Indications

Dipyridamole plus aspirin was compared with aspirin or dipyridamole alone, or with placebo, in patients with an ischemic stroke or transient ischemic attack. The combination reduced the risk of stroke by 22.1% compared with aspirin and by 24.4% compared with dipyridamole. A second trial compared dipyridamole plus aspirin with aspirin alone for secondary prevention in patients with ischemic stroke. Vascular death, stroke, or MI occurred in 13% of patients given combination therapy and in 16% of those treated with aspirin alone. Based on these data, Aggrenox was often used for stroke prevention. Another trial randomized 20,332 patients with noncardioembolic ischemic stroke to either Aggrenox or clopidogrel. The primary efficacy endpoint of recurrent stroke occurred in 9.0% of those given Aggrenox and in 8.8% of patients treated with clopidogrel. Although this difference was not statistically significant, the study failed to meet the prespecified margin to claim noninferiority of Aggrenox relative to clopidogrel. These results have dampened enthusiasm for the use of Aggrenox.

Because of its vasodilatory effects and the paucity of data supporting the use of dipyridamole in patients with symptomatic coronary artery disease, Aggrenox should not be used for stroke prevention in such patients. Clopidogrel is a better choice in this setting.

GPIIb/IIIA Receptor Antagonists

As a class, parenteral GPIIb/IIIa receptor antagonists have an established niche in patients with acute coronary syndromes. The three agents in this class are abciximab, eptifibatide, and tirofiban.

Mechanism of Action

A member of the integrin family of adhesion receptors, GPIIb/IIIa is found on the surface of platelets and megakaryocytes. With about 80,000 copies per platelet, GPIIb/IIIa is the most abundant receptor. Consisting of a noncovalently linked heterodimer, GPIIb/IIIa is inactive on resting platelets. When platelets are activated, inside-outside signal transduction pathways trigger a conformational activation of the receptor. Once activated, GPIIb/IIIa binds adhesive molecules, such as fibrinogen and, under high shear conditions, vWF. Binding is mediated by the Arg-Gly-Asp (RGD) sequence found on the α chains of fibrinogen and on vWF, and by the Lys-Gly-Asp (KGD) sequence located within a unique dodecapeptide domain on the γ chains of fibrinogen. Once bound, fibrinogen and/or vWF bridge adjacent platelets together to induce platelet aggregation.

Although abciximab, eptifibatide, and tirofiban all target the GPIIb/IIIa receptor, they are structurally and pharmacologically distinct (Table 118-1). Abciximab is a Fab fragment of a humanized murine monoclonal antibody directed against the activated form of GPIIb/IIIa. Abciximab binds to the activated receptor with high affinity and blocks the binding of adhesive molecules. In contrast to abciximab, eptifibatide and tirofiban are synthetic small molecules. Eptifibatide is a cyclic heptapeptide that binds GPIIb/IIIa because it incorporates the KGD motif, whereas tirofiban is a nonpeptidic tyrosine derivative that acts as an RGD mimetic. Abciximab has a long half-life and can be detected on the surface of platelets for up to 2 weeks. Eptifibatide and tirofiban have shorter half-lives.

Table 118-1 Features of GPIIb/IIIA Antagonists

Feature	Abciximab	Eptifibatide	Tirofiban
Description	Fab fragment of humanized mouse monoclonal antibody	Cyclical KGD-containing heptapeptide	Nonpeptidic RGD mimetic
Specific for GPIIb/IIIa	No	Yes	Yes
Plasma half-life	Short (min)	Long (2.5 h)	Long (2.0 h)
Platelet-bound half-life	Long (days)	Short (s)	Short (s)
Renal clearance	No	Yes	Yes

In addition to targeting the GPIIb/IIIa receptor, abciximab also inhibits the closely related $\alpha_V\beta_3$ receptor, which binds vitronectin, and $\alpha_M\beta_2$, a leukocyte integrin. In contrast, eptifibatide and tirofiban are specific for GPIIb/IIIa. Inhibition of $\alpha_V\beta_3$ and $\alpha_M\beta_2$ may endow abciximab with anti-inflammatory and/or antiproliferative properties that extend beyond platelet inhibition.

Dosing

All of the GPIIb/IIIa antagonists are given as an IV bolus followed by an infusion. Because they are cleared by the kidneys, the doses of eptifibatide and tirofiban must be reduced in patients with renal insufficiency.

Side Effects

In addition to bleeding, thrombocytopenia is the most serious complication. Thrombocytopenia is immune-mediated and is caused by antibodies directed against neoantigens on GPIIb/IIIa that are exposed upon antagonist binding. With abciximab, thrombocytopenia occurs in up to 5% of patients. Thrombocytopenia is severe in ~1% of these individuals. Thrombocytopenia is less common with the other two agents, occurring in ~1% of patients.

Indications

Abciximab and eptifibatide are used in patients undergoing percutaneous coronary interventions, particularly those with acute MI. Tirofiban is used in high-risk patients with unstable angina. Eptifibatide also can be used for this indication.

New Antiplatelet Agents

New agents in advanced stages of development include cangrelor and ticagrelor, direct-acting reversible $P2Y_{12}$ antagonists, and SCH530348 (vorapaxar) and E5555 (atopaxar), orally active inhibitors of protease-activated receptor 1 (PAR-1), the major thrombin receptor on platelets (Fig. 118-3). Cangrelor is an adenosine analogue that binds reversibly to $P2Y_{12}$ and inhibits its activity. The drug has a half-life of 3–6 min and is given intravenously as a bolus followed by an infusion. When stopped, platelet function recovers within 60 min. Trials comparing cangrelor with placebo during percutaneous coronary interventions or comparing cangrelor with clopidogrel after such procedures revealed little or no advantages of cangrelor. Consequently, identification of a role for cangrelor requires additional studies.

Ticagrelor is an orally active, reversible inhibitor of $P2Y_{12}$. The drug is given twice daily and it not only has a more rapid onset and offset of action than clopidogrel but also produces greater and more predictable inhibition of ADP-induced platelet aggregation. When compared with clopidogrel in patients with acute coronary syndromes, ticagrelor produced a greater reduction in the primary efficacy endpoint, a composite of cardiovascular death, MI, and stroke at 1 year (9.8% and 11.7%, respectively; $p = .001$). This difference reflected a significantly greater reduction in cardiovascular death (4.0% and 5.1%, respectively; $p = .001$) and MI (5.8% and 6.9%, respectively; $p = .005$) with ticagrelor than with clopidogrel. Rates of stroke were similar with ticagrelor and clopidogrel (1.5% and 1.3%, respectively) and there were no differences in the rates of major bleeding. When minor bleeding was added to the major bleeding results, however, ticagrelor showed an increase relative to

clopidogrel (16.1% and 14.6%, respectively; $p = 0.008$). Ticagrelor also was superior to clopidogrel in the acute coronary syndrome patients who underwent percutaneous coronary interventions or aortocoronary bypass surgery. Although not yet licensed, ticagrelor is the first new antiplatelet drug to demonstrate a greater reduction in cardiovascular death than clopidogrel in patients with acute coronary syndromes.

SCH530348, an orally active inhibitor of PAR-1, is under investigation as an adjunct to aspirin or aspirin plus clopidogrel. Two large phase III trials are underway. E5555, a second oral PAR-1 antagonist, is earlier in development.

ANTICOAGULANTS

There are both parenteral and oral anticoagulants. Currently available parenteral anticoagulants include heparin, low-molecular-weight heparin (LMWH), and fondaparinux, a synthetic pentasaccharide. The only available oral anticoagulants are the vitamin K antagonists, of which warfarin is the agent most often used in North America.

Dabigatran etexilate, an oral thrombin inhibitor, and rivaroxaban, an oral Factor Xa inhibitor, are licensed in Europe and Canada for short-term thromboprophylaxis after elective hip or knee replacement surgery. Dabigatran etexilate was licensed in the United States and Canada as an alternative to warfarin for stroke prevention in patients with atrial fibrillation.

Parenteral Anticoagulants

HEPARIN

Heparin is a sulfated polysaccharide and is isolated from mammalian tissues rich in mast cells. Most commercial heparin is derived from porcine intestinal mucosa and is a polymer of alternating D-glucuronic acid and *N*-acetyl-d-glucosamine residues.

Mechanism of Action

Heparin acts as an anticoagulant by activating antithrombin (previously known as antithrombin III) and accelerating the rate at which antithrombin inhibits clotting enzymes, particularly thrombin and factor Xa. Antithrombin, the obligatory plasma cofactor for heparin, is a member of the serine protease inhibitor (serpin) superfamily. Synthesized in the liver and circulating in plasma at a concentration of 2.6 ± 0.4 μM, antithrombin acts as a suicide substrate for its target enzymes.

To activate antithrombin, heparin binds to the serpin via a unique pentasaccharide sequence that is found on one-third of the chains of commercial heparin (Fig. 118-5). The remainder of the heparin chains that lack this pentasaccharide sequence have little or no anticoagulant activity. Once bound to antithrombin, heparin induces a conformational change in the reactive center loop of antithrombin that renders it more readily accessible to its target proteases. This conformational change enhances the rate at which antithrombin inhibits Factor Xa by at least two orders of magnitude but has little effect on the rate of thrombin inhibition by antithrombin. To catalyze thrombin inhibition, heparin serves as a template that binds antithrombin and thrombin simultaneously. Formation of this ternary complex brings the enzyme in close apposition to the inhibitor, thereby promoting the formation of a stable covalent thrombin-antithrombin complex.

Figure 118-5

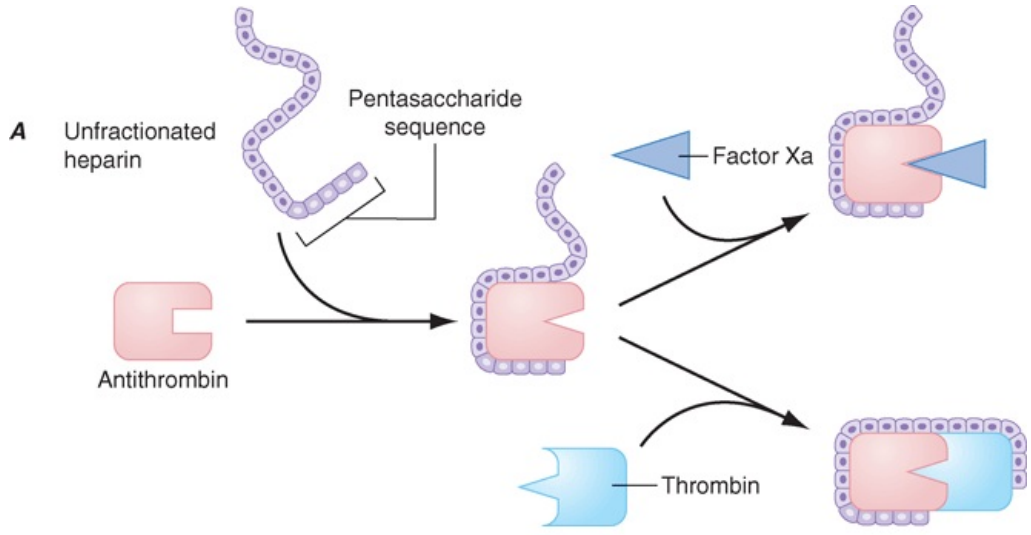

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

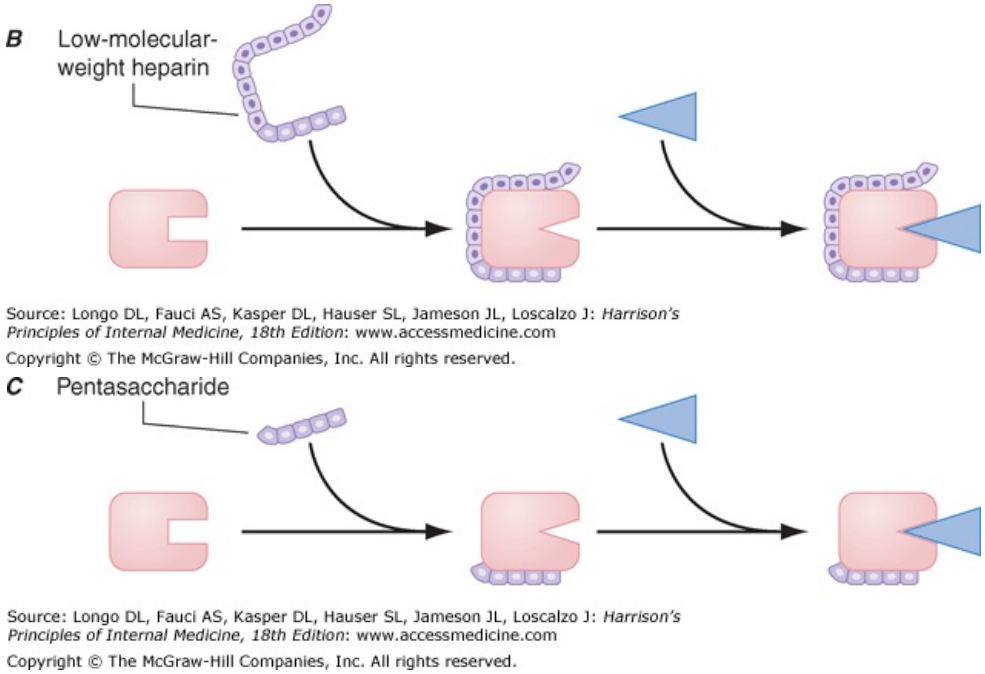

Mechanism of action of heparin, low-molecular-weight heparin (LMWH), and fondaparinux, a synthetic pentasaccharide. *A.* Heparin binds to antithrombin via its pentasaccharide sequence. This induces a conformational change in the reactive center loop of antithrombin that accelerates its interaction with factor Xa. To potentiate thrombin inhibition, heparin must simultaneously bind to antithrombin and thrombin. Only heparin chains composed of at least 18 saccharide units, which corresponds to a molecular weight of 5400, are of sufficient length to perform this bridging function. With a mean molecular weight of 15,000, all of the heparin chains are long enough to do this. ***B.*** LMWH has greater capacity to potentiate factor Xa inhibition by antithrombin than thrombin because, with a mean molecular weight of 4500–5000, at least half of the LMWH chains are too short to bridge antithrombin to thrombin. ***C.*** The pentasaccharide only accelerates factor Xa inhibition by antithrombin because the pentasaccharide is too short to bridge antithrombin to thrombin.

Only pentasaccharide-containing heparin chains composed of at least 18 saccharide units (which correspond to a molecular weight of 5400) are of sufficient length to bridge thrombin and antithrombin together. With a mean molecular weight of 15,000, and a range of 5000–30,000, almost all of the chains of unfractionated heparin are long enough to effect this bridging function. Consequently, by definition, heparin has equal capacity to promote the inhibition of thrombin and factor Xa by antithrombin and is assigned an anti-factor Xa to anti-factor IIa (thrombin) ratio of 1:1.

Heparin causes the release of tissue factor pathway inhibitor (TFPI) from the endothelium. A factor Xa –dependent inhibitor of tissue factor–bound factor VIIa, TFPI may contribute to the antithrombotic activity of heparin. Longer heparin chains induce the release of more TFPI than shorter chains.

Pharmacology

Heparin must be given parenterally. It is usually administered SC or by continuous IV infusion. When used for therapeutic purposes, the IV route is most often employed. If heparin is given SC for treatment of thrombosis, the dose of heparin must be high enough to overcome the limited bioavailability associated with this method of delivery.

In the circulation, heparin binds to the endothelium and to plasma proteins other than antithrombin. Heparin binding to endothelial cells explains its dose-dependent clearance. At low doses, the half-life of heparin is short because it binds rapidly to the endothelium. With higher doses of heparin, the half-life is longer because heparin is cleared more slowly once the endothelium is saturated. Clearance is mainly extrarenal; heparin binds to macrophages, which internalize and depolymerize the long heparin chains and secrete shorter chains back into the circulation. Because of its dose-dependent clearance mechanism, the plasma half-life of heparin ranges from 30 to 60 min with bolus IV doses of 25 and 100 U/kg, respectively.

Once heparin enters the circulation, it binds to plasma proteins other than antithrombin, a phenomenon that reduces its anticoagulant activity. Some of the heparin-binding proteins found in plasma are acute-phase reactants whose levels are elevated in ill patients. Others, such as high-molecular-weight multimers of vWF, are released from activated platelets or endothelial cells. Activated platelets also release platelet factor 4 (PF4), a highly cationic protein that binds heparin with high affinity. The large amounts of PF4 found in the vicinity of platelet-rich arterial thrombi can neutralize the anticoagulant activity of heparin. This phenomenon may attenuate heparin's capacity to suppress thrombus growth.

Because the levels of heparin-binding proteins in plasma vary from person to person, the anticoagulant response to fixed or weight-adjusted doses of heparin is unpredictable. Consequently,

coagulation monitoring is essential to ensure that a therapeutic response is obtained. This is particularly important when heparin is administered for treatment of established thrombosis because a subtherapeutic anticoagulant response may render patients at risk for recurrent thrombosis, whereas excessive anticoagulation increases the risk of bleeding.

Monitoring the Anticoagulant Effect

Heparin therapy can be monitored using the activated partial thromboplastin time (aPTT) or anti-factor Xa level. Although the aPTT is the test most often employed for this purpose, there are problems with this assay. aPTT reagents vary in their sensitivity to heparin, and the type of coagulometer used for testing can influence the results. Consequently, laboratories must establish a therapeutic aPTT range with each reagent-coagulometer combination by measuring the aPTT and anti-factor Xa level in plasma samples collected from heparin-treated patients. For most of the aPTT reagents and coagulometers in current use, therapeutic heparin levels are achieved with a two- to threefold prolongation of the aPTT.

Anti-factor Xa levels also can be used to monitor heparin therapy. With this test, therapeutic heparin levels range from 0.3 to 0.7 units/mL. Although this test is gaining in popularity, anti-factor Xa assays have yet to be standardized, and results can vary widely between laboratories.

Up to 25% of heparin-treated patients with venous thromboembolism require >35,000 units/d to achieve a therapeutic aPTT. These patients are considered heparin resistant. It is useful to measure anti-factor Xa levels in heparin-resistant patients because many will have a therapeutic anti-factor Xa level despite a subtherapeutic aPTT. This dissociation in test results occurs because elevated plasma levels of fibrinogen and factor VIII, both of which are acute-phase proteins, shorten the aPTT but have no effect on anti-factor Xa levels. Heparin therapy in patients who exhibit this phenomenon is best monitored using anti-factor Xa levels instead of the aPTT. Patients with congenital or acquired antithrombin deficiency and those with elevated levels of heparin-binding proteins may also need high doses of heparin to achieve a therapeutic aPTT or anti-factor Xa level. If there is good correlation between the aPTT and the anti-factor Xa levels, either test can be used to monitor heparin therapy.

Dosing

For prophylaxis, heparin is usually given in fixed doses of 5000 units SC two or three times daily. With these low doses, coagulation monitoring is unnecessary. In contrast, monitoring is essential when the drug is given in therapeutic doses. Fixed-dose or weight-based heparin nomograms are used to standardize heparin dosing and to shorten the time required to achieve a therapeutic anticoagulant response. At least two heparin nomograms have been validated in patients with venous thromboembolism and reduce the time required to achieve a therapeutic aPTT. Weight-adjusted heparin nomograms have also been evaluated in patients with acute coronary syndromes. After an IV heparin bolus of 5000 units or 70 units/kg, a heparin infusion rate of 12–15 units/kg per hour is usually administered. In contrast, weight-adjusted heparin nomograms for patients with venous thromboembolism use an initial bolus of 5000 units or 80 units/kg, followed by an infusion of 18 units/kg per hour. Thus, patients with venous thromboembolism appear to require higher doses of heparin to achieve a therapeutic aPTT than do patients with acute coronary syndromes. This may reflect differences in the thrombus burden. Heparin binds to fibrin, and the fibrin content of extensive deep-vein thrombi is greater than that of small coronary thrombi.

Heparin manufacturers in North America have traditionally measured heparin potency in USP units, with a unit defined as the concentration of heparin that prevents 1 mL of citrated sheep plasma from clotting for 1 h after calcium addition. In contrast, manufacturers in Europe measure heparin potency with anti-Xa assays using an international heparin standard for comparison. Because of problems with heparin contamination with oversulfated chondroitin sulfate, which the USP assay system does not detect, North American heparin manufacturers now use the anti-Xa assay to assess heparin potency. Although use of international units in place of USP units results in a 10% reduction in heparin doses, this change is unlikely to affect patient care because heparin has been dosed in international units in Europe for many years. Furthermore, heparin monitoring ensures a therapeutic anticoagulant response in high-risk situations, such as cardiopulmonary bypass surgery or percutaneous coronary intervention.

Limitations

Heparin has pharmacokinetic and biophysical limitations (Table 118-2). The pharmacokinetic limitations reflect heparin's propensity to bind in a pentasaccharide-independent fashion to cells and plasma proteins. Heparin binding to endothelial cells explains its dose-dependent clearance, whereas binding to plasma proteins results in a variable anticoagulant response and can lead to heparin resistance.

Table 118-2 Pharmacokinetic and Biophysical Limitations of Heparin

Limitations	Mechanism
Poor bioavailability at low doses	Binds to endothelial cells and macrophages

Dose-dependent clearance	Binds to macrophages
Variable anticoagulant response	Binds to plasma proteins whose levels vary from patient to patient
Reduced activity in the vicinity of platelet-rich thrombi	Neutralized by platelet factor 4 released from activated platelets
Limited activity against factor Xa incorporated in the prothrombinase complex and thrombin bound to fibrin	Reduced capacity of heparin-antithrombin complex to inhibit factor Xa bound to activated platelets and thrombin bound to fibrin

The biophysical limitations of heparin reflect the inability of the heparin-antithrombin complex to (1) inhibit factor Xa when it is incorporated into the prothrombinase complex, the complex that converts prothrombin to thrombin, and (2) to inhibit thrombin bound to fibrin. Consequently, factor Xa bound to activated platelets within platelet-rich thrombi has the potential to generate thrombin, even in the face of heparin. Once this thrombin binds to fibrin, it too is protected from inhibition by the heparin-antithrombin complex. Clot-associated thrombin can then trigger thrombus growth by locally activating platelets and amplifying its own generation through feedback activation of factors V, VIII, and XI. Further compounding the problem is the potential for heparin neutralization by the high concentrations of PF4 released from activated platelets within the platelet-rich thrombus.

Side Effects

The most common side effect of heparin is bleeding. Other complications include thrombocytopenia, osteoporosis, and elevated levels of transaminases.

Bleeding

The risk of heparin-induced bleeding increases with higher heparin doses. Concomitant administration of drugs that affect hemostasis, such as antiplatelet or fibrinolytic agents, increases the risk of bleeding, as does recent surgery or trauma. Heparin-treated patients with serious bleeding can be given protamine sulfate to neutralize the heparin. Protamine sulfate, a mixture of basic polypeptides isolated from salmon sperm, binds heparin with high affinity, and the resultant protamine-heparin complexes are then cleared. Typically, 1 mg of protamine sulfate neutralizes 100 units of heparin. Protamine sulfate is given IV. Anaphylactoid reactions to protamine sulfate can occur, and drug administration by slow IV infusion is recommended to reduce the risk.

Thrombocytopenia

Heparin can cause thrombocytopenia. Heparin-induced thrombocytopenia (HIT) is an antibody-mediated process that is triggered by antibodies directed against neoantigens on PF4 that are exposed when heparin binds to this protein. These antibodies, which are usually of the IgG isotype, bind simultaneously to the heparin-PF4 complex and to platelet Fc receptors. Such binding activates the platelets and generates platelet microparticles. Circulating microparticles are prothrombotic because they express anionic phospholipids on their surface and can bind clotting factors and promote thrombin generation.

The clinical features of HIT are illustrated in Table 118-3. Typically, HIT occurs 5–14 days after initiation of heparin therapy, but it can manifest earlier if the patient has received heparin within the past 3 months. It is rare for the platelet count to fall below 100,000/μL in patients with HIT, and even a 50% decrease in the platelet count from the pretreatment value should raise the suspicion of HIT in those receiving heparin. HIT is more common in surgical patients than in medical patients and, like many autoimmune disorders, occurs more frequently in females than in males.

Table 118-3 Features of Heparin-Induced Thrombocytopenia

Features	Details
Thrombocytopenia	Platelet count of ≤100,000/μL or a decrease in platelet count of ≥50%
Timing	Platelet count falls 5–10 days after starting heparin
Type of heparin	More common with unfractionated heparin than low-molecular-weight heparin
Type of patient	More common in surgical patients and patients with cancer than general medical patients. More common in women than in men
Thrombosis	Venous thrombosis more common than arterial thrombosis

HIT can be associated with thrombosis, either arterial or venous. Venous thrombosis, which manifests as DVT and/or PE, is more common than arterial thrombosis. Arterial thrombosis can manifest as ischemic stroke or acute MI. Rarely, platelet-rich thrombi in the distal aorta or iliac arteries can cause critical limb ischemia.

The diagnosis of HIT is established using enzyme-linked assays to detect antibodies against heparin-PF4 complexes or with platelet activation assays. Enzyme-linked assays are sensitive but can be positive in the absence of any clinical evidence of HIT. The most specific diagnostic test is the

serotonin release assay. This test is performed by quantifying serotonin release when washed platelets loaded with labeled serotonin are exposed to patient serum in the absence or presence of varying concentrations of heparin. If the patient serum contains the HIT antibody, heparin addition induces platelet activation and serotonin release.

Management of HIT is outlined in Table 118-4. Heparin should be stopped in patients with suspected or documented HIT, and an alternative anticoagulant should be administered to prevent or treat thrombosis. The agents most often used for this indication are parenteral direct thrombin inhibitors, such as lepirudin, argatroban, or bivalirudin, or factor Xa inhibitors, such as fondaparinux.

Table 118-4 Management of Heparin-Induced Thrombocytopenia

Stop all heparin
Give an alternative anticoagulant, such as lepirudin, argatroban, bivalirudin, or fondaparinux
Do not give platelet transfusions
Do not give warfarin until the platelet count returns to its baseline level. If warfarin is administered, give vitamin K to restore the INR to normal
Evaluate for thrombosis, particularly deep-vein thrombosis

Abbreviation: INR, International normalized ratio.

Patients with HIT, particularly those with associated thrombosis, often have evidence of increased thrombin generation that can lead to consumption of protein C. If these patients are given warfarin without a concomitant parenteral anticoagulant to inhibit thrombin or thrombin generation, the further decrease in protein C levels induced by the vitamin K antagonist can trigger skin necrosis. To avoid this problem, patients with HIT should be treated with a direct thrombin inhibitor or fondaparinux until the platelet count returns to normal levels. At this point, low-dose warfarin therapy can be introduced, and the thrombin inhibitor can be discontinued when the anticoagulant response to warfarin has been therapeutic for at least 2 days.

Osteoporosis

Treatment with therapeutic doses of heparin for >1 month can cause a reduction in bone density. This complication has been reported in up to 30% of patients given long-term heparin therapy, and symptomatic vertebral fractures occur in 2–3% of these individuals.

Heparin causes bone loss both by decreasing bone formation and by enhancing bone resorption. Thus, heparin affects the activity of both osteoblasts and osteoclasts.

Elevated Levels of Transaminases

Therapeutic doses of heparin frequently cause modest elevation in the serum levels of hepatic transaminases, without a concomitant increase in the level of bilirubin. The levels of transaminases rapidly return to normal when the drug is stopped. The mechanism of this phenomenon is unknown.

LOW-MOLECULAR-WEIGHT HEPARIN

Consisting of smaller fragments of heparin, LMWH is prepared from unfractionated heparin by controlled enzymatic or chemical depolymerization. The mean molecular weight of LMWH is 5000, one-third the mean molecular weight of unfractionated heparin. LMWH has advantages over heparin (Table 118-5) and has replaced heparin for most indications.

Table 118-5 Advantages of LMWH over Heparin

Advantage	Consequence
Better bioavailability and longer half-life after subcutaneous injection	Can be given subcutaneously once or twice daily for both prophylaxis and treatment
Dose-independent clearance	Simplified dosing
Predictable anticoagulant response	Coagulation monitoring is unnecessary in most patients
Lower risk of heparin-induced thrombocytopenia	Safer than heparin for short- or long-term administration
Lower risk of osteoporosis	Safer than heparin for extended administration

Abbreviation: LMWH, low-molecular-weight heparin.

Mechanism of Action

Like heparin, LMWH exerts its anticoagulant activity by activating antithrombin. With a mean molecular weight of 5000, which corresponds to about 17 saccharide units, at least half of the pentasaccharide-containing chains of LMWH are too short to bridge thrombin to antithrombin (Fig. 118-5). However, these chains retain the capacity to accelerate factor Xa inhibition by antithrombin because this activity

is largely the result of the conformational changes in antithrombin evoked by pentasaccharide binding. Consequently, LMWH catalyzes factor Xa inhibition by antithrombin more than thrombin inhibition. Depending on their unique molecular weight distributions, LMWH preparations have anti-factor Xa to anti-factor IIa ratios ranging from 2:1 to 4:1.

Pharmacology

Although usually given SC, LMWH also can be administered IV if a rapid anticoagulant response is needed. LMWH has pharmacokinetic advantages over heparin. These advantages reflect the fact that shorter heparin chains bind less avidly to endothelial cells, macrophages, and heparin-binding plasma proteins. Reduced binding to endothelial cells and macrophages eliminates the rapid, dose-dependent, and saturable mechanism of clearance that is a characteristic of unfractionated heparin. Instead, the clearance of LMWH is dose-independent and its plasma half-life is longer. Based on measurement of anti-factor Xa levels, LMWH has a plasma half-life of ~4 h. LMWH is cleared almost exclusively by the kidneys, and the drug can accumulate in patients with renal insufficiency.

LMWH exhibits about 90% bioavailability after SC injection. Because LMWH binds less avidly to heparin-binding proteins in plasma than heparin, LMWH produces a more predictable dose response, and resistance to LMWH is rare. With a longer half-life and more predictable anticoagulant response, LMWH can be given SC once or twice daily without coagulation monitoring, even when the drug is given in treatment doses. These properties render LMWH more convenient than unfractionated heparin. Capitalizing on this feature, studies in patients with venous thromboembolism have shown that home treatment with LMWH is as effective and safe as in-hospital treatment with continuous IV infusions of heparin. Outpatient treatment with LMWH streamlines care, reduces health care costs, and increases patient satisfaction.

Monitoring

In the majority of patients, LMWH does not require coagulation monitoring. If monitoring is necessary, anti-factor Xa levels must be measured because most LMWH preparations have little effect on the aPTT. Therapeutic anti-factor Xa levels with LMWH range from 0.5 to 1.2 units/mL when measured 3–4 h after drug administration. When LMWH is given in prophylactic doses, peak anti-Factor Xa levels of 0.2–0.5 units/mL are desirable.

Indications for LMWH monitoring include renal insufficiency and obesity. LMWH monitoring in patients with a creatinine clearance of ≤50 mL/min is advisable to ensure that there is no drug accumulation. Although weight-adjusted LMWH dosing appears to produce therapeutic anti-factor Xa levels in patients who are overweight, this approach has not been extensively evaluated in those with morbid obesity. It may also be advisable to monitor the anticoagulant activity of LMWH during pregnancy because dose requirements can change, particularly in the third trimester. Monitoring should also be considered in high-risk settings, such as in patients with mechanical heart valves who are given LMWH for prevention of valve thrombosis, and when LMWH is used in treatment doses in infants or children.

Dosing

The doses of LMWH recommended for prophylaxis or treatment vary depending on the LMWH preparation. For prophylaxis, once-daily SC doses of 4000–5000 units are often used, whereas doses of 2500–3000 units are given when the drug is administered twice daily. For treatment of venous thromboembolism, a dose of 150–200 units/kg is given if the drug is administered once daily. If a twice-daily regimen is employed, a dose of 100 units/kg is given. In patients with unstable angina, LMWH is given SC on a twice-daily basis at a dose of 100–120 units/kg.

Side Effects

The major complication of LMWH is bleeding. Meta-analyses suggest that the risk of major bleeding is lower with LMWH than with unfractionated heparin. HIT and osteoporosis are less common with LMWH than with unfractionated heparin.

Bleeding

Like the situation with heparin, bleeding with LMWH is more common in patients receiving concomitant therapy with antiplatelet or fibrinolytic drugs. Recent surgery, trauma, or underlying hemostatic defects also increase the risk of bleeding with LMWH.

Although protamine sulfate can be used as an antidote for LMWH, protamine sulfate incompletely neutralizes the anticoagulant activity of LMWH because it only binds the longer chains of LMWH. Because longer chains are responsible for catalysis of thrombin inhibition by antithrombin, protamine sulfate completely reverses the anti-factor IIa activity of LMWH. In contrast, protamine sulfate only partially reverses the anti-factor Xa activity of LMWH because the shorter pentasaccharide-containing chains of LMWH do not bind to protamine sulfate. Consequently, patients at high risk for bleeding may be more safely treated with continuous IV unfractionated heparin than with SC LMWH.

Thrombocytopenia

The risk of HIT is about fivefold lower with LMWH than with heparin. LMWH binds less avidly to

platelets and causes less PF4 release. Furthermore, with lower affinity for PF4 than heparin, LMWH is less likely to induce the conformational changes in PF4 that trigger the formation of HIT antibodies.

LMWH should not be used to treat HIT patients because most HIT antibodies exhibit cross-reactivity with LMWH. This in vitro cross-reactivity is not simply a laboratory phenomenon because there are case reports of thrombosis when HIT patients are treated with LMWH.

Osteoporosis

The risk of osteoporosis is lower with long-term LMWH than with heparin. For extended treatment, therefore, LMWH is a better choice than heparin because of the lower risk of osteoporosis and HIT.

FONDAPARINUX

A synthetic analogue of the antithrombin-binding pentasaccharide sequence, fondaparinux differs from LMWH in several ways (Table 118-6). Fondaparinux is licensed for thromboprophylaxis in general medical or surgical patients and in high-risk orthopedic patients and as an alternative to heparin or LMWH for initial treatment of patients with established venous thromboembolism. The drug is not yet licensed in the United States as an alternative for heparin or LMWH in patients with acute coronary syndromes.

Table 118-6 Comparison of LMWH and Fondaparinux

Features	LMWH	Fondaparinux
Number of saccharide units	15–17	5
Catalysis of factor Xa inhibition	Yes	Yes
Catalysis of thrombin inhibition	Yes	No
Bioavailability after subcutaneous administration (%)	90	100
Plasma half-life (h)	4	17
Renal excretion	Yes	Yes
Induces release of tissue factor pathway inhibitor	Yes	No
Neutralized by protamine sulfate	Partially	No

Mechanism of Action

As a synthetic analogue of the antithrombin-binding pentasaccharide sequence found in heparin and LMWH, fondaparinux has a molecular weight of 1728. Fondaparinux binds only to antithrombin (Fig. 118-5) and is too short to bridge thrombin to antithrombin. Consequently, fondaparinux catalyzes factor Xa inhibition by antithrombin and does not enhance the rate of thrombin inhibition.

Pharmacology

Fondaparinux exhibits complete bioavailability after SC injection. With no binding to endothelial cells or plasma proteins, the clearance of fondaparinux is dose independent and its plasma half-life is 17 h. The drug is given SC once daily. Because fondaparinux is cleared unchanged via the kidneys, it is contraindicated in patients with a creatinine clearance <30 mL/min and should be used with caution in those with a creatinine clearance <50 mL/min.

Fondaparinux produces a predictable anticoagulant response after administration in fixed doses because it does not bind to plasma proteins. The drug is given at a dose of 2.5 mg once daily for prevention of venous thromboembolism. For initial treatment of established venous thromboembolism, fondaparinux is given at a dose of 7.5 mg once daily. The dose can be reduced to 5 mg once daily for those weighing <50 kg and increased to 10 mg for those >100 kg. When given in these doses, fondaparinux is as effective as heparin or LMWH for initial treatment of patients with DVT or PE and produces similar rates of bleeding.

Fondaparinux is used at a dose of 2.5 mg once daily in patients with acute coronary syndromes. When this prophylactic dose of fondaparinux was compared with treatment doses of enoxaparin in patients with non-ST-segment elevation acute coronary syndromes, there was no difference in the rate of cardiovascular death, MI, or stroke at 9 days. However, the rate of major bleeding was 50% lower with fondaparinux than with enoxaparin, a difference that likely reflects the fact that the dose of fondaparinux was lower than that of enoxaparin. In acute coronary syndrome patients who require percutaneous coronary interventions, there is a risk of catheter thrombosis with fondaparinux, unless adjunctive heparin is given.

Side Effects

Fondaparinux does not cause HIT because it does not bind to PF4. In contrast to LMWH, there is no cross-reactivity of fondaparinux with HIT antibodies. Consequently, fondaparinux appears to be effective for treatment of HIT patients, although large clinical trials supporting its use are lacking.

The major side effect of fondaparinux is bleeding. There is no antidote for fondaparinux. Protamine

sulfate has no effect on the anticoagulant activity of fondaparinux because it fails to bind to the drug. Recombinant activated factor VII reverses the anticoagulant effects of fondaparinux in volunteers, but it is unknown whether this agent will control fondaparinux-induced bleeding.

PARENTERAL DIRECT THROMBIN INHIBITORS

Heparin and LMWH are indirect inhibitors of thrombin because their activity is mediated by antithrombin. In contrast, direct thrombin inhibitors do not require a plasma cofactor; instead, these agents bind directly to thrombin and block its interaction with its substrates. Approved parenteral direct thrombin inhibitors include lepirudin, argatroban, and bivalirudin (Table 118-7). Lepirudin and argatroban are licensed for treatment of patients with HIT, whereas bivalirudin is approved as an alternative to heparin in patients undergoing percutaneous coronary interventions, including those with HIT.

Table 118-7 Comparison of the Properties of Lepirudin, Bivalirudin, and Argatroban

	Lepirudin	Bivalirudin	Argatroban
Molecular mass	7000	1980	527
Site(s) of interaction with thrombin	Active site and exosite 1	Active site and exosite 1	Active site
Renal clearance	Yes	No	No
Hepatic metabolism	No	No	Yes
Plasma half-life (min)	60	25	45

Lepirudin

A recombinant form of hirudin, lepirudin is a bivalent direct thrombin inhibitor that interacts with both the active site and exosite 1, the substrate-binding site, on thrombin. For rapid anticoagulation, lepirudin is given by continuous IV infusion, but the drug can be given SC for thromboprophylaxis. Lepirudin has a plasma half-life of 60 min after IV infusion and is cleared by the kidneys. Consequently, lepirudin accumulates in patients with renal insufficiency. A high proportion of lepirudin-treated patients develop antibodies against the drug. Although these antibodies rarely cause problems, in a small subset of patients, they can delay lepirudin clearance and enhance its anticoagulant activity. Serious bleeding has been reported in some of these patients.

Lepirudin is usually monitored using the aPTT, and the dose is adjusted to maintain an aPTT that is 1.5–2.5 times the control. The aPTT is not an ideal test for monitoring lepirudin therapy because the clotting time plateaus with higher drug concentrations. Although the ecarin clotting time provides a better index of lepirudin dose than the aPTT, the ecarin clotting time has yet to be standardized.

Argatroban

A univalent inhibitor that targets the active site of thrombin, argatroban is metabolized in the liver. Consequently, this drug must be used with caution in patients with hepatic insufficiency. Argatroban is not cleared via the kidneys, so this drug is safer than lepirudin for HIT patients with renal insufficiency.

Argatroban is administered by continuous IV infusion and has a plasma half-life of ~45 min. The aPTT is used to monitor its anticoagulant effect, and the dose is adjusted to achieve an aPTT 1.5–3 times the baseline value, but not to exceed 100 s. Argatroban also prolongs the international normalized ratio (INR), a feature that can complicate the transitioning of patients to warfarin. This problem can be circumvented by using the levels of factor X to monitor warfarin in place of the INR. Alternatively, argatroban can be stopped for 2–3 h before INR determination.

Bivalirudin

A synthetic 20-amino-acid analogue of hirudin, bivalirudin is a divalent thrombin inhibitor. Thus, the N-terminal portion of bivalirudin interacts with the active site of thrombin, whereas its C-terminal tail binds to exosite 1, the substrate-binding domain on thrombin. Bivalirudin has a plasma half-life of 25 min, the shortest half-life of all the parenteral direct thrombin inhibitors. Bivalirudin is degraded by peptidases and is partially excreted via the kidneys. When given in high doses in the cardiac catheterization laboratory, the anticoagulant activity of bivalirudin is monitored using the activated clotting time. With lower doses, its activity can be assessed using the aPTT.

Studies comparing bivalirudin with heparin suggest that bivalirudin produces less bleeding. This feature plus its short half-life make bivalirudin an attractive alternative to heparin in patients undergoing percutaneous coronary interventions. Bivalirudin also has been used successfully in HIT patients who require percutaneous coronary interventions.

Oral Anticoagulants

Current oral anticoagulant practice dates back almost 60 years to when the vitamin K antagonists were discovered as a result of investigations into the cause of hemorrhagic disease in cattle. Characterized by a decrease in prothrombin levels, this disorder is caused by ingestion of hay containing spoiled

sweet clover. Hydroxycoumarin, which was isolated from bacterial contaminants in the hay, interferes with vitamin K metabolism, thereby causing a syndrome similar to vitamin K deficiency. Discovery of this compound provided the impetus for development of other vitamin K antagonists, including warfarin.

WARFARIN

A water-soluble vitamin K antagonist initially developed as a rodenticide, warfarin is the coumarin derivative most often prescribed in North America. Like other vitamin K antagonists, warfarin interferes with the synthesis of the vitamin K–dependent clotting proteins, which include prothrombin (factor II) and factors VII, IX, and X. The synthesis of the vitamin K–dependent anticoagulant proteins, proteins C and S, is also reduced by vitamin K antagonists.

Mechanism of Action

All of the vitamin K–dependent clotting factors possess glutamic acid residues at their N termini. A posttranslational modification adds a carboxyl group to the γ-carbon of these residues to generate γ-carboxyglutamic acid. This modification is essential for expression of the activity of these clotting factors because it permits their calcium-dependent binding to negatively charged phospholipid surfaces. The γ-carboxylation process is catalyzed by a vitamin K–dependent carboxylase. Thus, vitamin K from the diet is reduced to vitamin K hydroquinone by vitamin K reductase (Fig. 118-6). Vitamin K hydroquinone serves as a cofactor for the carboxylase enzyme, which in the presence of carbon dioxide replaces the hydrogen on the γ-carbon of glutamic acid residues with a carboxyl group. During this process, vitamin K hydroquinone is oxidized to vitamin K epoxide, which is then reduced to vitamin K by vitamin K epoxide reductase.

Figure 118-6

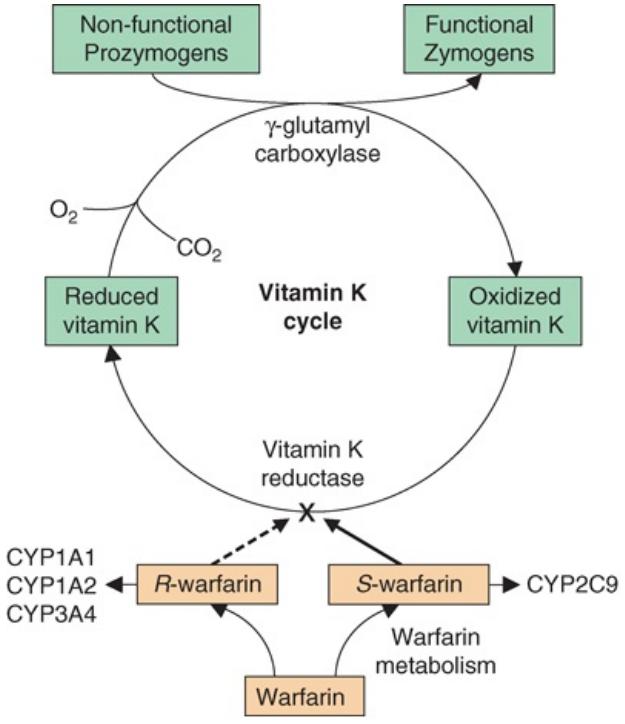

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanism of action of warfarin. A racemic mixture of *S*- and *R*-enantiomers, *S*-warfarin is most active. By blocking vitamin K epoxide reductase, warfarin inhibits the conversion of oxidized vitamin K into its reduced form. This inhibits vitamin K–dependent γ-carboxylation of factors II, VII, IX, and X because reduced vitamin K serves as a cofactor for a γ-glutamyl carboxylase that catalyzes the γ-carboxylation process, thereby converting prozymogens to zymogens capable of binding calcium and interacting with anionic phospholipid surfaces. S-warfarin is metabolized by *CYP2C9*. Common genetic polymorphisms in this enzyme can influence warfarin metabolism. Polymorphisms in the C1 subunit of vitamin K reductase (VKORC1) also can affect the susceptibility of the enzyme to warfarin-induced inhibition, thereby influencing warfarin dosage requirements.

Warfarin inhibits vitamin K epoxide reductase (VKOR), thereby blocking the γ-carboxylation process. This results in the synthesis of vitamin K–dependent clotting proteins that are only partially γ-carboxylated. Warfarin acts as an anticoagulant because these partially γ-carboxylated proteins have reduced or absent biologic activity. The onset of action of warfarin is delayed until the newly synthesized clotting factors with reduced activity gradually replace their fully active counterparts.

The antithrombotic effect of warfarin depends on a reduction in the functional levels of factor X and prothrombin, clotting factors that have half-lives of 24 and 72 h, respectively. Because of the delay in achieving an antithrombotic effect, initial treatment with warfarin is supported by concomitant administration of a rapidly acting parenteral anticoagulant, such as heparin, LMWH, or fondaparinux, in patients with established thrombosis or at high risk for thrombosis.

Pharmacology

Warfarin is a racemic mixture of *R* and *S* isomers. Warfarin is rapidly and almost completely absorbed from the gastrointestinal tract. Levels of warfarin in the blood peak about 90 min after drug administration. Racemic warfarin has a plasma half-life of 36–42 h, and more than 97% of circulating warfarin is bound to albumin. Only the small fraction of unbound warfarin is biologically active.

Warfarin accumulates in the liver where the two isomers are metabolized via distinct pathways. *CYP2C9* mediates oxidative metabolism of the more active *S* isomer (Fig. 118-6). Two relatively common variants, *CYP2C9*2* and *CYP2C9*3*, encode an enzyme with reduced activity. Patients with these variants require lower maintenance doses of warfarin. Approximately 25% of Caucasians have at least one variant allele of *CYP2C9*2* or *CYP2C9*3*, whereas those variant alleles are less common in African Americans and Asians (Table 118-8). Heterozygosity for *CYP2C9*2* or *CYP2C9*3* decreases the warfarin dose requirement by 20–30% relative to that required in subjects with the wild-type *CYP2C9*1/*1* alleles, whereas homozygosity for the *CYP2C9*2* or *CYP2C9*3* alleles reduces the warfarin dose requirement by 50–70%.

Table 118-8 Frequencies of *CYP2C9* Genotypes and *VKORC1* Haplotypes in Different Populations and Their Effect on Warfarin Dose Requirements

	Frequency, %			
Genotype/haploptye	**Caucasians**	**African Americans (A/A)**	**Asians (A)**	**Dose reduction compared with wild-type**
CYP2C9				
*1/*1	70	90	95	–
*1/*2	17	2	0	22
*1/*3	9	3	4	34
*2/*2	2	0	0	43
*2/*3	1	0	0	53
*3/*3	0	0	1	76
VKORC1				
Non-A/non-A	37	82	7	–
Non-A/A	45	12	30	26
A/A	18	6	63	50

Consistent with their decreased warfarin dose requirement, subjects with at least one *CYP2C9* variant allele are at increased risk for bleeding. Compared with individuals with no variant alleles, the relative risks for warfarin-associated bleeding in *CYP2C9*2* or *CYP2C9*3* carriers are 1.91 and 1.77, respectively.

Polymorphisms in *VKORC1* also can influence the anticoagulant response to warfarin. Several genetic variations of *VKORC1* are in strong linkage disequilibrium and have been designated as non-A haplotypes. *VKORC1* variants are more prevalent than variants of *CYP2C9*. Asians have the highest prevalence of *VKORC1* variants, followed by Caucasians and African Americans (Table 118-8). Polymorphisms in *VKORC1* likely explain 30% of the variability in warfarin dose requirements. Compared with *VKORC1* non-A/non-A homozygotes, the warfarin dose requirement decreases by 25 and 50% in A halotype heterozygotes and homozygotes, respectively. These findings prompted the Food and Drug Administration to amend the prescribing information for warfarin to indicate that lower initiation doses should be considered for patients with *CYP2C9* and *VKORC1* genetic variants. In addition to genotype data, other pertinent patient information has been incorporated into warfarin dosing algorithms. Although such algorithms help predict suitable warfarin doses, it remains unclear whether better dose identification improves patient outcome in terms of reducing hemorrhagic complications or recurrent thrombotic events.

In addition to genetic factors, the anticoagulant effect of warfarin is influenced by diet, drugs, and various disease states. Fluctuations in dietary vitamin K intake affect the activity of warfarin. A wide variety of drugs can alter absorption, clearance, or metabolism of warfarin. Because of the variability in the anticoagulant response to warfarin, coagulation monitoring is essential to ensure that a therapeutic response is obtained.

Monitoring

Warfarin therapy is most often monitored using the prothrombin time, a test that is sensitive to reductions in the levels of prothrombin, factor VII, and factor X. The test is performed by adding thromboplastin, a reagent that contains tissue factor, phospholipid, and calcium, to citrated plasma and determining the time to clot formation. Thromboplastins vary in their sensitivity to reductions in the levels of the vitamin K–dependent clotting factors. Thus, less sensitive thromboplastins will trigger the administration of higher doses of warfarin to achieve a target prothrombin time. This is problematic because higher doses of warfarin increase the risk of bleeding.

The INR was developed to circumvent many of the problems associated with the prothrombin time. To calculate the INR, the patient's prothrombin time is divided by the mean normal prothrombin time, and this ratio is then multiplied by the international sensitivity index (ISI), an index of the sensitivity of the thromboplastin used for prothrombin time determination to reductions in the levels of the vitamin K –dependent clotting factors. Highly sensitive thromboplastins have an ISI of 1.0. Most current thromboplastins have ISI values that range from 1.0 to 1.4.

Although the INR has helped to standardize anticoagulant practice, problems persist. The precision of INR determination varies depending on reagent-coagulometer combinations. This leads to variability in the INR results. Also complicating INR determination is unreliable reporting of the ISI by thromboplastin manufacturers. Furthermore, every laboratory must establish the mean normal prothrombin time with each new batch of thromboplastin reagent. To accomplish this, the prothrombin time must be measured in fresh plasma samples from at least 20 healthy volunteers using the same coagulometer that is used for patient samples.

For most indications, warfarin is administered in doses that produce a target INR of 2.0–3.0. An exception is patients with mechanical heart valves, where a target INR of 2.5–3.5 is recommended. Studies in atrial fibrillation demonstrate an increased risk of cardioembolic stroke when the INR falls to <1.7 and an increase in bleeding with INR values >4.5. These findings highlight the fact that vitamin K antagonists have a narrow therapeutic window. In support of this concept, a study in patients receiving long-term warfarin therapy for unprovoked venous thromboembolism demonstrated a higher rate of recurrent venous thromboembolism with a target INR of 1.5–1.9 compared with a target INR of 2.0 –3.0.

Dosing

Warfarin is usually started at a dose of 5–10 mg. Lower doses are used for patients with *CYP2C9* or *VKORC1* polymorphisms, which affect the pharmacodynamics or pharmacokinetics of warfarin and render patients more sensitive to the drug. The dose is then titrated to achieve the desired target INR. Because of its delayed onset of action, patients with established thrombosis or those at high risk for thrombosis are given concomitant treatment with a rapidly acting parenteral anticoagulant, such as heparin, LMWH, or fondaparinux. Initial prolongation of the INR reflects reduction in the functional levels of factor VII. Consequently, concomitant treatment with the parenteral anticoagulant should be continued until the INR has been therapeutic for at least 2 consecutive days. A minimum 5-day course of parenteral anticoagulation is recommended to ensure that the levels of prothrombin have been reduced into the therapeutic range with warfarin.

Because warfarin has a narrow therapeutic window, frequent coagulation monitoring is essential to ensure that a therapeutic anticoagulant response is obtained. Even patients with stable warfarin dose requirements should have their INR determined every 2–3 weeks. More frequent monitoring is necessary when new medications are introduced because so many drugs enhance or reduce the anticoagulant effects of warfarin.

Side Effects

Like all anticoagulants, the major side effect of warfarin is bleeding. A rare complication is skin necrosis. Warfarin crosses the placenta and can cause fetal abnormalities. Consequently, warfarin should not be used during pregnancy.

Bleeding

At least half of the bleeding complications with warfarin occur when the INR exceeds the therapeutic range. Bleeding complications may be mild, such as epistaxis or hematuria, or more severe, such as retroperitoneal or gastrointestinal bleeding. Life-threatening intracranial bleeding can also occur.

To minimize the risk of bleeding, the INR should be maintained in the therapeutic range. In asymptomatic patients whose INR is between 3.5 and 4.5, warfarin should be withheld until the INR returns to the therapeutic range. If the INR is >4.5, a therapeutic INR can be achieved more rapidly by administration of low doses of sublingual vitamin K. A vitamin K dose of 1 mg is usually adequate for patients with an INR between 4.9 and 9, whereas 2–3 mg can be used for those with an INR >9. Higher doses of vitamin K can be administered if more rapid reversal of the INR is required or if the INR is excessively high. Although vitamin K administration results in a more rapid reduction in the INR compared with simply holding the warfarin, there is no evidence that vitamin K administration reduces the risk of hemorrhage.

Patients with serious bleeding need more aggressive treatment. These patients should be given 10 mg of vitamin K by slow IV infusion. Additional vitamin K should be given until the INR is in the normal range. Treatment with vitamin K should be supplemented with fresh-frozen plasma as a source of the vitamin K–dependent clotting proteins. For life-threatening bleeds, or if patients cannot tolerate the volume load, prothrombin complex concentrates can be used.

Warfarin-treated patients who experience bleeding when their INR is in the therapeutic range require investigation into the cause of the bleeding. Those with gastrointestinal bleeding often have underlying peptic ulcer disease or a tumor. Similarly, investigation of hematuria or uterine bleeding in patients with a therapeutic INR may unmask a tumor of the genitourinary tract.

Skin Necrosis

A rare complication of warfarin, skin necrosis usually is seen 2–5 days after initiation of therapy. Well-demarcated erythematous lesions form on the thighs, buttocks, breasts, or toes. Typically, the center of the lesion becomes progressively necrotic. Examination of skin biopsies taken from the border of these lesions reveals thrombi in the microvasculature.

Warfarin-induced skin necrosis is seen in patients with congenital or acquired deficiencies of protein C or protein S. Initiation of warfarin therapy in these patients produces a precipitous fall in plasma levels of proteins C or S, thereby eliminating this important anticoagulant pathway before warfarin exerts an antithrombotic effect through lowering of the functional levels of factor X and prothrombin. The resultant procoagulant state triggers thrombosis. Why the thrombosis is localized to the microvasculature of fatty tissues is unclear.

Treatment involves discontinuation of warfarin and reversal with vitamin K, if needed. An alternative anticoagulant, such as heparin or LMWH, should be given in patients with thrombosis. Protein C concentrates or recombinant activated protein C can be given to protein C–deficient patients to accelerate healing of the skin lesions; fresh-frozen plasma may be of value for those with protein S deficiency. Occasionally, skin grafting is necessary when there is extensive skin loss.

Because of the potential for skin necrosis, patients with known protein C or protein S deficiency require overlapping treatment with a parenteral anticoagulant when initiating warfarin therapy. Warfarin should be started in low doses in these patients, and the parenteral anticoagulant should be continued until the INR is therapeutic for at least 2–3 consecutive days.

Pregnancy

Warfarin crosses the placenta and can cause fetal abnormalities or bleeding. The fetal abnormalities include a characteristic embryopathy, which consists of nasal hypoplasia and stippled epiphyses. The risk of embryopathy is highest if warfarin is given in the first trimester of pregnancy. Central nervous system abnormalities can also occur with exposure to warfarin at any time during pregnancy. Finally, maternal administration of warfarin produces an anticoagulant effect in the fetus that can cause bleeding. This is of particular concern at delivery when trauma to the head during passage through the birth canal can lead to intracranial bleeding. Because of these potential problems, warfarin is contraindicated in pregnancy, particularly in the first and third trimesters. Instead, heparin, LMWH, or fondaparinux can be given during pregnancy for prevention or treatment of thrombosis.

Warfarin does not pass into the breast milk. Consequently, warfarin can safely be given to nursing mothers.

Special Problems

Patients with a lupus anticoagulant or those who need urgent or elective surgery present special challenges. Although observational studies suggested that patients with thrombosis complicating the antiphospholipid antibody syndrome required higher intensity warfarin regimens to prevent recurrent thromboembolic events, two randomized trials showed that targeting an INR of 2.0–3.0 is as effective as higher intensity treatment and produces less bleeding. Monitoring warfarin therapy can be problematic in patients with antiphospholipid antibody syndrome if the lupus anticoagulant prolongs the baseline INR.

If patients receiving long-term warfarin treatment require an elective invasive procedure, warfarin can be stopped 5 days before the procedure to allow the INR to return to normal levels. Those at high risk for recurrent thrombosis can be bridged with once- or twice-daily SC injections of LMWH when the INR falls to <2.0. The last dose of LMWH should be given 12–24 h before the procedure, depending on whether LMWH is administered twice or once daily. After the procedure, treatment with warfarin can be restarted.

NEW ORAL ANTICOAGULANTS

New oral anticoagulants that target thrombin or factor Xa are under development. These drugs have a rapid onset of action and have half-lives that permit once- or twice-daily administration. Designed to produce a predictable level of anticoagulation, these new oral agents are given in fixed doses without routine coagulation monitoring. Therefore, these drugs are more convenient to administer than warfarin.

Dabigatran etexilate, an oral thrombin inhibitor, and rivaroxaban, an oral factor Xa inhibitor, are licensed in Europe and Canada for short-term thromboprophylaxis after elective hip or knee replacement surgery. Phase III trials with apixaban, another oral factor Xa inhibitor, also have been completed in patients undergoing major orthopedic surgery (Table 118-9).

Table 118-9 Comparison of the Features of New Oral Anticoagulants in Advanced Stages of Development

Features	Rivaroxaban	Apixaban	Dabigatran Etexilate
Target	Xa	Xa	IIa
Molecular weight	436	460	628
Prodrug	No	No	Yes
Bioavailability (%)	80	50	6
Time to peak (h)	3	3	2
Half-life (h)	9	9–14	12–17
Renal excretion (%)	65	25	80
Antidote	None	None	None

The RE-LY trial shows the promise of these new agents for long-term indications. This trial compared two different dose regimens of dabigatran etexilate (110 mg or 150 mg twice daily) with warfarin (dose-adjusted to achieve an INR between 2 and 3) for stroke prevention in 18,113 patients with nonvalvular atrial fibrillation. The annual rates of the primary efficacy outcome, stroke or systemic embolism, were 1.7% with warfarin, 1.5% with the lower dose dabigatran regimen, and 1.1% with the higher dose regimen. Thus, the lower dose dabigatran regimen was noninferior to warfarin, while the higher dose regimen was superior. Annual rates of major bleeding were 3.4% with warfarin compared with 2.7% and 3.1% with the lower and higher dose dabigatran regimens, respectively. Thus, the lower dose dabigatran regimen was associated with significantly less major bleeding than warfarin, while the rate of major bleeding with the higher dose regimen was not significantly different from that with warfarin. Rates of intracerebral bleeding were significantly lower with both doses of dabigatran than with warfarin, as were rates of life-threatening bleeding. There was no evidence of hepatotoxicity with dabigatran.

Based on the results of the RE-LY trial, dabigatran etexilate has been licensed in the United States and Canada for stroke prevention in patients with atrial fibrillation. The 150 mg twice daily dose of dabigatran is recommended for most patients. In the United States, a 75 mg twice daily dose is recommended for patients with a creatinine clearance of 30 to 50 mL/min, while in Canada, the 110 mg twice daily dose is recommended for those over the age of 80 years or for patients at high risk of bleeding. The drug is contraindicated in patients with a creatinine clearance less than 15 mL/min.

Dabigatran etexilate also was compared with warfarin in 2539 patients with acute venous thromboembolism. Patients were initially treated with heparin or LMWH and then randomized to a 6-month course of dabigatran (150 mg twice daily) or warfarin, which was dose-adjusted to achieve an INR of 2–3. The primary endpoint, a composite of recurrent venous thromboembolism or fatal pulmonary embolism, occurred in 2.4% of patients given dabigatran and in 2.1% of those treated with warfarin. Major bleeding occurred in 1.6 and 1.9% of patients given dabigatran and warfarin, respectively. Based on the results of this trial, unmonitored fixed-dose dabigatran appears to be noninferior to warfarin for treatment of patients with venous thromboembolism. Taken together with the results of the RE-LY trial, these findings suggest that the new oral anticoagulants will gradually replace warfarin.

FIBRINOLYTIC DRUGS

Role of Fibrinolytic Therapy

Fibrinolytic drugs can be used to degrade thrombi and are administered systemically or can be delivered via catheters directly into the substance of the thrombus. Systemic delivery is used for treatment of acute MI, acute ischemic stroke, and most cases of massive PE. The goal of therapy is to produce rapid thrombus dissolution, thereby restoring antegrade blood flow. In the coronary circulation, restoration of blood flow reduces morbidity and mortality rates by limiting myocardial damage, whereas in the cerebral circulation, rapid thrombus dissolution decreases the neuronal death and brain infarction that produce irreversible brain injury. For patients with massive PE, the goal of thrombolytic therapy is to restore pulmonary artery perfusion.

Peripheral arterial thrombi and thrombi in the proximal deep veins of the leg are most often treated using catheter-directed thrombolytic therapy. Catheters with multiple side holes can be utilized to enhance drug delivery. In some cases, intravascular devices that fragment and extract the thrombus are used to hasten treatment. These devices can be used alone or in conjunction with fibrinolytic

drugs.

Mechanism of Action

Currently approved fibrinolytic agents include streptokinase; acylated plasminogen streptokinase activator complex (anistreplase); urokinase; recombinant tissue-type plasminogen activator (rtPA), which is also known as alteplase or activase; and two recombinant derivatives of rtPA, tenecteplase and reteplase. All of these agents act by converting the proenzyme, plasminogen, to plasmin, the active enzyme (Fig. 118-7). Plasmin then degrades the fibrin matrix of thrombi and produces soluble fibrin degradation products.

Figure 118-7

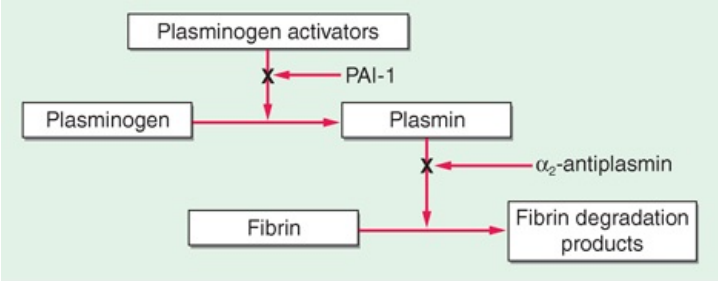

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The fibrinolytic system and its regulation. Plasminogen activators convert plasminogen to plasmin. Plasmin then degrades fibrin into soluble fibrin degradation products. The system is regulated at two levels. Type 1 plasminogen activator inhibitor (PAI-1) regulates the plasminogen activators, whereas α_2-antiplasmin serves as the major inhibitor of plasmin.

Endogenous fibrinolysis is regulated at two levels. Plasminogen activator inhibitors, particularly the type 1 form (PAI-1), prevent excessive plasminogen activation by regulating the activity of tPA and urokinase-type plasminogen activator (u-PA). Once plasmin is generated, it is regulated by plasmin inhibitors, the most important of which is α_2-antiplasmin. The plasma concentration of plasminogen is twofold higher than that of α_2-antiplasmin. Consequently, with pharmacologic doses of plasminogen activators, the concentration of plasmin that is generated can exceed that of α_2-antiplasmin. In addition to degrading fibrin, unregulated plasmin can also degrade fibrinogen and other clotting factors. This process, which is known as the *systemic lytic state*, reduces the hemostatic potential of the blood and increases the risk of bleeding.

The endogenous fibrinolytic system is geared to localize plasmin generation to the fibrin surface. Both plasminogen and tPA bind to fibrin to form a ternary complex that promotes efficient plasminogen activation. In contrast to free plasmin, plasmin generated on the fibrin surface is relatively protected from inactivation by α_2-antiplasmin, a feature that promotes fibrin dissolution. Furthermore, C-terminal lysine residues, exposed as plasmin degrades fibrin, serve as binding sites for additional plasminogen and tPA molecules. This creates a positive feedback that enhances plasmin generation. When used pharmacologically, the various plasminogen activators capitalize on these mechanisms to a lesser or greater extent.

Plasminogen activators that preferentially activate fibrin-bound plasminogen are considered fibrin-specific. In contrast, nonspecific plasminogen activators do not discriminate between fibrin-bound and circulating plasminogen. Activation of circulating plasminogen results in the generation of unopposed plasmin that can trigger the systemic lytic state. Alteplase and its derivatives are fibrin-specific plasminogen activators, whereas streptokinase, anistreplase, and urokinase are nonspecific agents.

Streptokinase

Unlike other plasminogen activators, streptokinase is not an enzyme and does not directly convert plasminogen to plasmin. Instead, streptokinase forms a 1:1 stoichiometric complex with plasminogen. Formation of this complex induces a conformational change in plasminogen that exposes its active site (Fig. 118-8). This conformationally altered plasminogen then converts additional plasminogen molecules to plasmin.

Figure 118-8

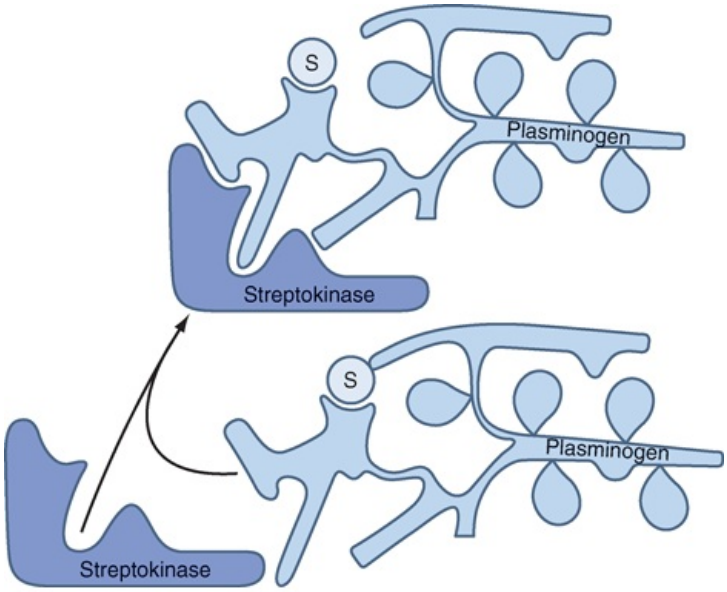

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanism of action of streptokinase. Streptokinase binds to plasminogen and induces a conformational change in plasminogen that exposes its active site. The streptokinase/plasmin(ogen) complex then serves as the activator of additional plasminogen molecules.

Streptokinase has no affinity for fibrin, and the streptokinase-plasminogen complex activates both free and fibrin-bound plasminogen. Activation of circulating plasminogen generates sufficient amounts of plasmin to overwhelm α_2-antiplasmin. Unopposed plasmin not only degrades fibrin in the occlusive thrombus but also induces a systemic lytic state.

When given systemically to patients with acute MI, streptokinase reduces mortality. For this indication, the drug is usually given as an IV infusion of 1.5 million units over 30–60 min. Patients who receive streptokinase can develop antibodies against the drug, as can patients with prior streptococcal injection. These antibodies can reduce the effectiveness of streptokinase.

Allergic reactions occur in ~5% of patients treated with streptokinase. These may manifest as a rash, fever, chills, and rigors. Although anaphylactic reactions can occur, these are rare. Transient hypotension is common with streptokinase and has been attributed to plasmin-mediated release of bradykinin from kininogen. The hypotension usually responds to leg elevation and administration of IV fluids and low doses of vasopressors, such as dopamine or norepinephrine.

Anistreplase

To generate this drug, streptokinase is combined with equimolar amounts of Lys-plasminogen, a plasmin-cleaved form of plasminogen with a Lys residue at its N terminus. The active site of Lys-plasminogen that is exposed upon combination with streptokinase is then masked with an anisoyl group. After IV infusion, the anisoyl group is slowly removed by deacylation, giving the complex a half-life of ~100 min. This allows drug administration via a single bolus infusion.

Although it is more convenient to administer, anistreplase offers few mechanistic advantages over streptokinase. Like streptokinase, anistreplase does not distinguish between fibrin-bound and circulating plasminogen. Consequently, it too produces a systemic lytic state. Likewise, allergic reactions and hypotension are just as frequent with anistreplase as they are with streptokinase.

When anistreplase was compared with alteplase in patients with acute MI, reperfusion was obtained more rapidly with alteplase than with anistreplase. Improved reperfusion was associated with a trend toward better clinical outcomes and reduced mortality rate with alteplase. These results and the high cost of anistreplase have dampened the enthusiasm for its use.

Urokinase

Urokinase is a two-chain serine protease derived from cultured fetal kidney cells with a molecular weight of 34,000. Urokinase converts plasminogen to plasmin directly by cleaving the Arg560-Val561 bond. Unlike streptokinase, urokinase is not immunogenic and allergic reactions are rare. Urokinase produces a systemic lytic state because it does not discriminate between fibrin-bound and circulating plasminogen.

Despite many years of use, urokinase has never been systemically evaluated for coronary thrombolysis. Instead, urokinase is often employed for catheter-directed lysis of thrombi in the deep veins or the peripheral arteries. Because of production problems, the availability of urokinase is

limited.

Alteplase

A recombinant form of single-chain tPA, alteplase has a molecular weight of 68,000. Alteplase is rapidly converted into its two-chain form by plasmin. Although single- and two-chain forms of tPA have equivalent activity in the presence of fibrin, in its absence, single-chain tPA has tenfold lower activity.

Alteplase consists of five discrete domains (Fig. 118-9); the N-terminal A chain of two-chain alteplase contains four of these domains. Residues 4 through 50 make up the finger domain, a region that resembles the finger domain of fibronectin; residues 50 through 87 are homologous with epidermal growth factor, whereas residues 92 through 173 and 180 through 261, which have homology to the kringle domains of plasminogen, are designated as the first and second kringle, respectively. The fifth alteplase domain is the protease domain; it is located on the C-terminal B chain of two-chain alteplase.

Figure 118-9

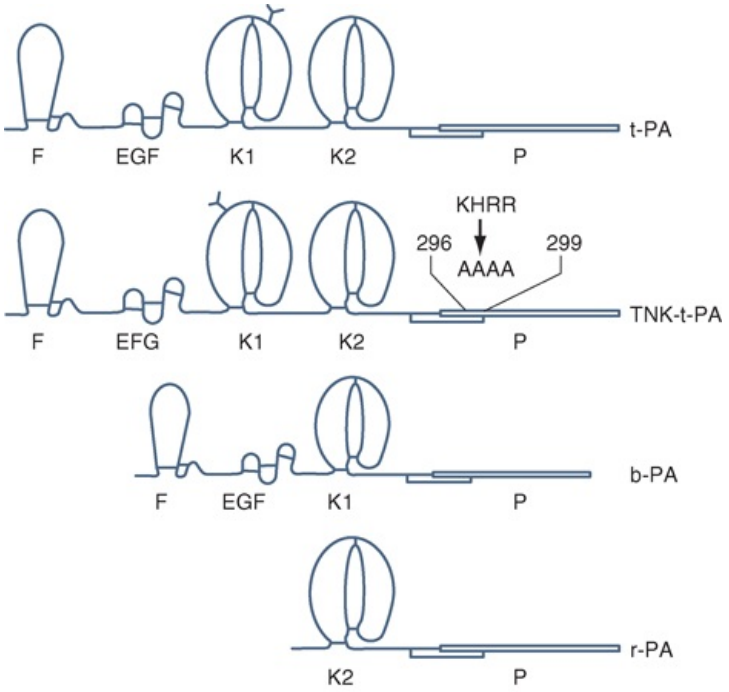

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Domain structures of alteplase (tPA), tenecteplase (TNK-tPA), desmoteplase (b-PA), and reteplase (r-PA). The finger (F), epidermal growth factor (EGF), first and second kringles (K1 and K2, respectively), and protease (P) domains are illustrated. The glycosylation site (Y) on K1 has been repositioned in tenecteplase to endow it with a longer half-life. In addition, a tetra-alanine substitution in the protease domain renders tenecteplase resistant to PAI-1 inhibition. Desmoteplase differs from alteplase and tenecteplase in that it lacks a K2 domain. Reteplase is a truncated variant that lacks the F, EGF, and K1 domains.

The interaction of alteplase with fibrin is mediated by the finger domain and, to a lesser extent, by the second kringle domain. The affinity of alteplase for fibrin is considerably higher than that for fibrinogen. Consequently, the catalytic efficiency of plasminogen activation by alteplase is two to three orders of magnitude higher in the presence of fibrin than in the presence of fibrinogen. This phenomenon helps to localize plasmin generation to the fibrin surface.

Although alteplase preferentially activates plasminogen in the presence of fibrin, alteplase is not as fibrin-selective as was first predicted. Its fibrin specificity is limited because like fibrin, (DD)E, the major soluble degradation product of cross-linked fibrin, binds alteplase and plasminogen with high affinity. Consequently, (DD)E is as potent as fibrin as a stimulator of plasminogen activation by alteplase. Whereas plasmin generated on the fibrin surface results in thrombolysis, plasmin generated on the surface of circulating (DD)E degrades fibrinogen. Fibrinogenolysis results in the accumulation of fragment X, a high-molecular-weight clottable fibrinogen degradation product. Incorporation of fragment X into hemostatic plugs formed at sites of vascular injury renders them susceptible to lysis. This phenomenon may contribute to alteplase-induced bleeding.

A trial comparing alteplase with streptokinase for treatment of patients with acute MI demonstrated significantly lower mortality with alteplase than with streptokinase, although the absolute difference was small. The greatest benefit was seen in patients age <75 years with anterior MI who presented <6

h after symptom onset.

For treatment of acute MI or acute ischemic stroke, alteplase is given as an IV infusion over 60–90 min. The total dose of alteplase usually ranges from 90 to 100 mg. Allergic reactions and hypotension are rare, and alteplase is not immunogenic.

Tenecteplase

Tenecteplase is a genetically engineered variant of tPA and was designed to have a longer half-life than tPA and to be resistant to inactivation by PAI-1. To prolong its half-life, a new glycosylation site was added to the first kringle domain (Fig. 118-9). Because addition of this extra carbohydrate side chain reduced fibrin affinity, the existing glycosylation site on the first kringle domain was removed. To render the molecule resistant to inhibition by PAI-1, a tetra-alanine substitution was introduced at residues 296–299 in the protease domain, the region responsible for the interaction of tPA with PAI-1.

Tenecteplase is more fibrin-specific than tPA. Although both agents bind to fibrin with similar affinity, the affinity of tenecteplase for (DD)E is significantly lower than that of tPA. Consequently, (DD)E does not stimulate systemic plasminogen activation by tenecteplase to the same extent as tPA. As a result, tenecteplase produces less fibrinogenolysis than tPA.

For coronary thrombolysis, tenecteplase is given as a single IV bolus. In a large phase III trial that enrolled >16,000 patients, the 30-day mortality rate with single-bolus tenecteplase was similar to that with accelerated-dose tPA. Although rates of intracranial hemorrhage were also similar with both treatments, patients given tenecteplase had fewer noncerebral bleeds and a reduced need for blood transfusions than those treated with tPA. The improved safety profile of tenecteplase likely reflects its enhanced fibrinspecificity.

Reteplase

Reteplase is a recombinant tPA derivative and is a single-chain variant that lacks the finger, epidermal growth factor, and first kringle domains (Fig. 118-9). This truncated derivative has a molecular weight of 39,000. Reteplase binds fibrin more weakly than tPA because it lacks the finger domain. Because it is produced in *Escherichia coli*, reteplase is not glycosylated. This endows it with a plasma half-life longer than that of tPA. Consequently, reteplase is given as two IV boluses, which are separated by 30 min. Clinical trials have demonstrated that reteplase is at least as effective as streptokinase for treatment of acute MI, but the agent is not superior to tPA.

New Fibrinolytic Agents

Several new drugs are under investigation. These include desmoteplase (Fig. 118-9), a recombinant form of the full-length plasminogen activator isolated from the saliva of the vampire bat, and alfimeprase, a truncated form of fibrolase, an enzyme isolated from the venom of the southern copperhead snake. Clinical studies with these agents have been disappointing. Desmoteplase, which is more fibrin-specific than tPA, was investigated for treatment of acute ischemic stroke. Patients presenting 3–9 h after symptom onset were randomized to one of two doses of desmoteplase or to placebo. Overall response rates were low and no different with desmoteplase from with placebo. Mortality rate was higher in the desmoteplase arms.

Alfimeprase is a metalloproteinase that degrades fibrin and fibrinogen in a plasmin-independent fashion. In the circulation, alfimeprase is inhibited by α_2-macroglobulin. Consequently, the drug must be delivered via a catheter directly into the thrombus. Studies of alfimeprase for treatment of peripheral arterial occlusion or for restoration of flow in blocked central venous catheters were stopped due to lack of efficacy. The disappointing results with desmoteplase and alfimeprase highlight the challenges of introducing new fibrinolytic drugs.

CONCLUSIONS AND FUTURE DIRECTIONS

Arterial and venous thromboses reflect a complex interplay among the vessel wall, platelets, the coagulation system, and the fibrinolytic pathways. Activation of coagulation also triggers inflammatory pathways that may contribute to thrombogenesis. A better understanding of the biochemistry of blood coagulation and advances in structure-based drug design have identified new targets and resulted in the development of novel antithrombotic drugs. Well-designed clinical trials have provided detailed information on which drugs to use and when to use them. Despite these advances, however, thromboembolic disorders remain a major cause of morbidity and mortality rates. Therefore, the search for better targets and more potent antiplatelet, anticoagulant, and fibrinolytic drugs continues.

FURTHER READINGS

Connolly SJ et al: Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 361:1139, 2009[PMID: 19717844]

Eikelboom JW, Weitz JI: New anticoagulants. Circulation 121:1523, 2010[PMID: 20368532]

Greinacher A: Heparin-induced thrombocytopenia. J Thromb Haemost 7:9, 2009[PMID: 19630757]

International Warfarin Pharmacogenetics Consortium: Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med 360:753, 2009

Mackman N: Triggers, targets and treatments of thrombosis. Nature 451:914, 2008[PMID: 18288180]

Mega JL et al: Cytochrome p-450 polymorphisms and response to clopidogrel. N Engl J Med 360:354, 2009[PMID: 19106084]

Rijken DC et al: New insights into the molecular mechanisms of the fibrinolytic system. J Thromb Haemost 7:4, 2009[PMID: 19017261]

Varenhorst C et al: Genetic variation of CYP2C19 affects both pharmacokinetic and pharmacodymamic responses to clopidogrel but not prasugrel in aspirin-treated patients with coronary artery disease. Eur Heart J 30:1744, 2009[PMID: 19429918]

Wallentin L et al: Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 361:1045, 2009[PMID: 19717846]

———: Effect of CYP2C19 and ABCB1 single nucleotide polymorphisms on outcomes of treatment with ticagrelor versus clopidogrel for acute coronary syndromes: A genetic substudy of the PLATO trial. Lancet 376:1320, 2010

Watson SP: Platelet activation by extracellular matrix proteins in haemostasis and thrombosis. Curr Pharm Des 15:1358, 2009[PMID: 19355974]

Wiviott SD et al for the TRITON–TIMI 38 Investigators: Prasugrel versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 357:2001, 2007[PMID: 17982182]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MOLECULAR MECHANISMS OF MICROBIAL PATHOGENESIS: INTRODUCTION

Over the past four decades, molecular studies of the pathogenesis of microorganisms have yielded an explosion of information about the various microbial and host molecules that contribute to the processes of infection and disease. These processes can be classified into several stages: microbial encounter with and entry into the host; microbial growth after entry; avoidance of innate host defenses; tissue invasion and tropism; tissue damage; and transmission to new hosts. *Virulence* is the measure of an organism's capacity to cause disease and is a function of the pathogenic factors elaborated by microbes. These factors promote *colonization* (the simple presence of potentially pathogenic microbes in or on a host), *infection* (attachment and growth of pathogens and avoidance of host defenses), and *disease* (often, but not always, the result of activities of secreted toxins or toxic metabolites). In addition, the host's inflammatory response to infection greatly contributes to disease and its attendant clinical signs and symptoms.

Microbial Entry and Adherence

ENTRY SITES

A microbial pathogen can potentially enter any part of a host organism. In general, the type of disease produced by a particular microbe is often a direct consequence of its route of entry into the body. The most common sites of entry are mucosal surfaces (the respiratory, alimentary, and urogenital tracts) and the skin. Ingestion, inhalation, and sexual contact are typical routes of microbial entry. Other portals of entry include sites of skin injury (cuts, bites, burns, trauma) along with injection via natural (i.e., vector-borne) or artificial (i.e., needle-stick injury) routes. A few pathogens, such as *Schistosoma* species, can penetrate unbroken skin. The conjunctiva can serve as an entry point for pathogens of the eye, which occasionally spread systemically from that site.

Microbial entry usually relies on the presence of specific factors needed for persistence and growth in a tissue. Fecal-oral spread via the alimentary tract requires a biologic profile consistent with survival in the varied environments of the gastrointestinal tract (including the low pH of the stomach and the high bile content of the intestine) as well as in contaminated food or water outside the host. Organisms that gain entry via the respiratory tract survive well in small moist droplets produced during sneezing and coughing. Pathogens that enter by venereal routes often survive best in the warm moist environment of the urogenital mucosa and have restricted host ranges (e.g., *Neisseria gonorrhoeae*, *Treponemapallidum*, and HIV).

The biology of microbes entering through the skin is highly varied. Some of these organisms can survive in a broad range of environments, such as the salivary glands or alimentary tracts of arthropod vectors, the mouths of larger animals, soil, and water. A complex biology allows protozoan parasites such as *Plasmodium*, *Leishmania*, and *Trypanosoma* spp. to undergo morphogenic changes that permit transmission of the organism to mammalian hosts during insect feeding for blood meals. Plasmodia are injected as infective sporozoites from the salivary glands during mosquito feeding. *Leishmania* parasites are regurgitated as promastigotes from the alimentary tract of sandflies and injected by bite into a susceptible host. Trypanosomes are first ingested from infected hosts by reduviid bugs; the pathogens then multiply in the gastrointestinal tract of the insects and are released in feces onto the host's skin during subsequent feedings. Most microbes that land directly on intact skin are destined to die, as survival on the skin or in hair follicles requires resistance to fatty acids, low pH, and other antimicrobial factors on the skin. Once it is damaged (and particularly if it becomes necrotic), the skin can be a major portal of entry and growth for pathogens and elaboration of their toxic products. Burn wound infections and tetanus are clear examples. After animal bites, pathogens resident in the animal's saliva gain access to the victim's tissues through the damaged skin. Rabies is the paradigm for this pathogenic process; rabies virus grows in striated muscle cells at the site of inoculation.

MICROBIAL ADHERENCE

Once in or on a host, most microbes must anchor themselves to a tissue or tissue factor; the possible exceptions are organisms that directly enter the bloodstream and multiply there. Specific ligands or adhesins for host receptors constitute a major area of study in the field of microbial pathogenesis. Adhesins comprise a wide range of surface structures, not only anchoring the microbe to a tissue and promoting cellular entry where appropriate but also eliciting host responses critical to the pathogenic process **(Table 120-1)**. Most microbes produce multiple adhesins specific for multiple host receptors. These adhesins are often redundant, are serologically variable, and act additively or synergistically with other microbial factors to promote microbial sticking to host tissues. In addition, some microbes adsorb host proteins onto their surface and utilize the natural host protein receptor for microbial binding and entry into target cells.

Table 120–1. Examples of Microbial Ligand-Receptor Interactions

Microorganism	Type of Microbial Ligand	Host Receptor
Viral Pathogens		
Influenza virus	Hemagglutinin	Sialic acid
Measles virus		
Vaccine strain	Hemagglutinin	CD46/moesin
Wild-type strains	Hemagglutinin	Signaling lymphocytic activation molecule (SLAM)
Human herpesvirus type 6	?	CD46
Herpes simplex virus	Glycoprotein C	Heparan sulfate
HIV	Surface glycoprotein	CD4 and chemokine receptors (CCR5 and CXCR4)
Epstein-Barr virus	Envelope protein	CD21 (CR2)

Adenovirus	Fiber protein	Coxsackie-adenovirus receptor (CAR)
Coxsackievirus	Viral coat proteins	CAR and major histocompatibility class I antigens
Bacterial Pathogens		
Neisseria spp.	Pili	Membrane co-factor protein (CD46)
Pseudomonas aeruginosa	Pili and flagella	Asialo-GM1
	Lipopolysaccharide	Cystic fibrosis transmembrane conductance regulator (CFTR)
Escherichia coli	Pili	Ceramides/mannose and digalactosyl residues
Streptococcus pyogenes	Hyaluronic acid capsule	CD44
Yersinia spp.	Invasin/accessory invasin locus	β_1 Integrins
Bordetella pertussis	Filamentous hemagglutinin	CR3
Legionella pneumophila	Adsorbed C3bi	CR3
Mycobacterium tuberculosis	Adsorbed C3bi	CR3; DC-SIGN[a]
Fungal Pathogens		
Blastomyces dermatitidis	WI-1	Possibly matrix proteins and integrins
Candida albicans	Int1p	Extracellular matrix proteins
Protozoal Pathogens		
Plasmodium vivax	Merozoite form	Duffy Fy antigen
Plasmodium falciparum	Erythrocyte-binding protein 175 (EBA-175)	Glycophorin A
Entamoeba histolytica	Surface lectin	*N*-Acetylglucosamine

[a] A novel dendritic cell–specific C-type lectin.

Viral Adhesins

All viral pathogens must bind to host cells, enter them, and replicate within them. Viral coat proteins serve as the ligands for cellular entry, and more than one ligand-receptor interaction may be needed; for example, HIV utilizes its envelope glycoprotein (gp) 120 to enter host cells by binding to both CD4 and one of two receptors for chemokines (designated CCR5 and CXCR4). Similarly, the measles virus H glycoprotein binds to both CD46 and the membrane-organizing protein moesin on host cells. The gB and gC proteins on herpes simplex virus bind to heparan sulfate, although this adherence is not essential for entry but rather serves to concentrate virions close to the cell surface; this step is followed by attachment to mammalian cells mediated by the viral gD protein, with subsequent formation of a homotrimer of viral gB protein or a heterodimer of viral gH and gL proteins that permits fusion of the viral envelope with the host cell membrane. Herpes simplex virus can use a number of eukaryotic cell surface receptors for entry, including the herpesvirus entry mediator (related to the tumor necrosis factor receptor), members of the immunoglobulin superfamily, the proteins nectin-1 and nectin-2, and modified heparan sulfate.

Bacterial Adhesins

Among the microbial adhesins studied in greatest detail are bacterial pili and flagella **(Fig. 120-1)**. *Pili* or *fimbriae* are commonly used by gram-negative bacteria for attachment to host cells and tissues; recent studies have identified similar factors produced by gram-positive organisms such as group B streptococci. In electron micrographs, these hairlike projections (up to several hundred per cell) may be confined to one end of the organism (polar pili) or distributed more evenly over the surface. An individual cell may have pili with a variety of functions. Most pili are made up of a major pilin protein subunit (molecular weight, 17,000–30,000) that polymerizes to form the pilus. Many strains of *Escherichia coli* isolated from urinary tract infections express mannose-binding type 1 pili, whose binding to integral membrane glycoproteins called *uroplakins* that coat the cells in the bladder epithelium is inhibited by d-mannose. Other strains produce the Pap (pyelonephritis-associated) or P pilus adhesin that mediates binding to digalactose (gal-gal) residues on globosides of the human P blood groups. Both of these types of pili have proteins located at the tips of the main pilus unit that are critical to the binding specificity of the whole pilus unit. Although immunization with the mannose-binding tip protein (FimH) of type 1 pili prevents experimental *E. coli* bladder infections in mice and monkeys, a human trial of this vaccine was not successful. *E. coli* cells causing diarrheal disease express pilus-like receptors for enterocytes on the small bowel, along with other receptors termed *colonization factors*.

Figure 120-1

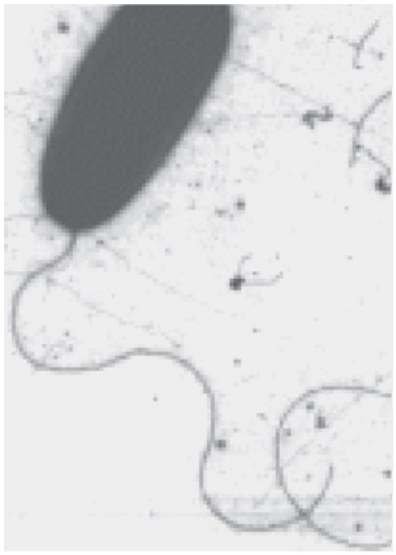

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

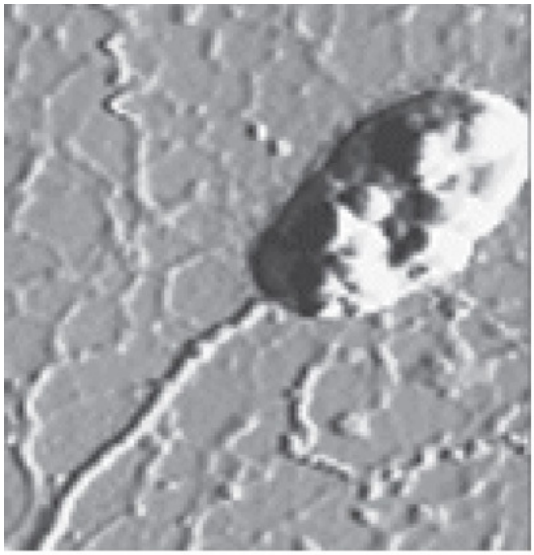

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

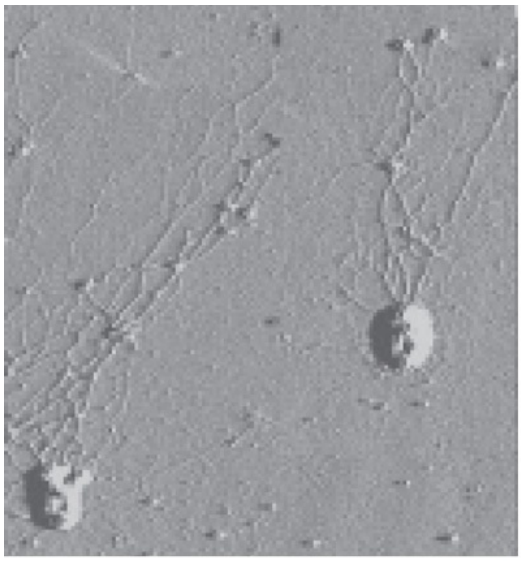

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bacterial surface structures. ***A*** and ***B***. Traditional electron micrographic images of fixed cells of *Pseudomonas aeruginosa*. Flagella **(*A*)** and pili **(*B*)** project out from the bacterial poles. ***C*** and ***D***. Atomic force microscopic image of live *P. aeruginosa* freshly planted onto a smooth mica surface. This technology reveals the fine, three-dimensional detail of the bacterial surface structures. *(Images courtesy of Drs. Martin Lee and Milan Bajmoczi, Harvard Medical School.)*

The type IV pilus, a common type of pilus found in *Neisseria* species, *Moraxella* species, *Vibrio cholerae*, *Legionella pneumophila, Salmonella enterica* serovar Typhi, enteropathogenic *E. coli*, and *Pseudomonas aeruginosa*, mediates adherence of organisms to target surfaces. Type IV pili tend to have a relatively conserved amino-terminal region and a more variable carboxyl-terminal region. For some species (e.g., *N.gonorrhoeae,Neisseriameningitidis*, and enteropathogenic *E. coli*), the pili are critical for attachment to mucosal epithelial cells. For others, such as *P. aeruginosa*, the pili only partially mediate the cells' adherence to host tissues. *V. cholerae* cells appear to use two different types of pili for intestinal colonization. Whereas interference with this stage of colonization would appear to be an effective antibacterial strategy, attempts to develop pilus-based vaccines for human diseases have not been highly successful to date.

Flagella are long appendages attached at either one or both ends of the bacterial cell (polar flagella) or distributed over the entire cell surface (peritrichous flagella). Flagella, like pili, are composed of a polymerized or aggregated basic protein. In flagella, the protein subunits form a tight helical structure and vary serologically with the species. Spirochetes such as *T. pallidum* and *Borrelia burgdorferi* have axial filaments similar to flagella running down the long axis of the center of the cell, and they "swim" by rotation around these filaments. Some bacteria can glide over a surface in the absence of obvious motility structures.

Other bacterial structures involved in adherence to host tissues include specific staphylococcal and streptococcal proteins that bind to human extracellular matrix proteins such as fibrin, fibronectin, fibrinogen, laminin, and collagen. Fibronectin appears to be a commonly used receptor for various pathogens; a particular amino acid sequence in fibronectin, Arg-Gly-Asp or RGD, is a critical target used by bacteria to bind to host tissues. Binding of a highly conserved *Staphylococcus aureus* surface protein, clumping factor A (ClfA), to fibrinogen has been implicated in many aspects of pathogenesis.

However, attempts to interrupt this interaction and prevent *S. aureus* sepsis in low-birth-weight infants by administering an intravenous IgG preparation derived from the plasma of individuals with high titers of antibody to ClfA failed to show efficacy in a clinical trial completed in April 2006. The conserved outer-core portion of the lipopolysaccharide (LPS) of *P. aeruginosa* mediates binding to the cystic fibrosis transmembrane conductance regulator (CFTR) on airway epithelial cells—an event that appears to be critical for normal host resistance to infection. A number of bacterial pathogens, including coagulase-negative staphylococci, *S. aureus*, and uropathogenic *E. coli* as well as *Yersinia pestis, Y. pseudotuberculosis*, *Y. enterocolitica, Bordetella* species, and *Acinetobacter baumannii* , express a surface polysaccharide composed of β-1-6-linked-poly-*N*-acetyl-d-glucosamine. One of its functions is to promote binding to materials used in catheters and other types of implanted devices. This polysaccharide may be a critical factor in the establishment of device-related infections by pathogens such as staphylococci and*E. coli*. High-powered imaging techniques (e.g., atomic force microscopy) have revealed that bacterial cells have a nonhomogeneous surface that is probably attributable to different concentrations of cell surface molecules, including microbial adhesins, at specific places on the cell surface (Fig. 120-1*D*).

Fungal Adhesins

Several fungal adhesins have been described that mediate colonization of epithelial surfaces, particularly adherence to structures like fibronectin, laminin, and collagen. The product of the *Candida albicansINT1* gene, Int1p, bears similarity to mammalian integrins that bind to extracellular matrix proteins. Transformation of normally nonadherent *Saccharomyces cerevisiae* with this gene allows these yeast cells to adhere to human epithelial cells. The agglutinin-like sequence (ALS) adhesins are large cell-surface glycoproteins mediating adherence of pathogenic *Candida* to host tissues. These adhesins possess a conserved three-domain structure composed of an N-terminal domain that mediates adherence to host tissue receptors, a central motif consisting of a number of repeats of a conserved sequence of 36 amino acids, and a C-terminal domain that varies in length and sequence and contains a glycosylphosphatidylinositol (GPI) anchor addition site that allows binding of the adhesin to the fungal cell wall. Variability in the number of central domains in different ALS proteins characterizes different adhesins with specificity for different host receptors. The ALS adhesins are expressed under certain environmental conditions—often associated with stress—and are crucial for pathogenesis of fungal infections.

For several fungal pathogens that initiate infections after inhalation of infectious material, the inoculum is ingested by alveolar macrophages, in which the fungal cells transform to pathogenic phenotypes. Like *C. albicans*, *Blastomyces dermatitidis* binds to CD11b/CD18 integrins as well as to CD14 on macrophages. *B. dermatitidis* produces a 120-kDa surface protein, designated WI-1, that mediates this adherence. The binding domain of WI-1 is homologous to the invasin protein of *Yersinia* that binds to the same type of host cell receptor. An unidentified factor on *Histoplasma capsulatum* also mediates binding of this fungal pathogen to the integrin surface proteins.

Eukaryotic Pathogen Adhesins

Eukaryotic parasites use complicated surface glycoproteins as adhesins, some of which are lectins (proteins that bind to specific carbohydrates on host cells). For example, *Plasmodium vivax*, one of five *Plasmodium* species causing malaria, binds (via Duffy-binding protein) to the Duffy blood group carbohydrate antigen Fy on erythrocytes. *Entamoeba histolytica*, the third leading cause of death from parasitic diseases, expresses two proteins that bind to the disaccharide galactose/N-acetyl galactosamine. Reports indicate that children with mucosal IgA antibody to one of these lectins are resistant to reinfection with virulent *E. histolytica*. A major surface glycoprotein (gp63) of *Leishmania* promastigotes is needed for these parasites to enter human macrophages—the principal target cell of infection. This glycoprotein promotes complement binding but inhibits complement lytic activity, allowing the parasite to use complement receptors for entry into macrophages; gp63 also binds to fibronectin receptors on macrophages. In addition, the pathogen can express a carbohydrate that mediates binding to host cells. Evidence suggests that, as part of hepatic granuloma formation, *Schistosoma mansoni* expresses a carbohydrate epitope related to the Lewis X blood group antigen that promotes adherence of helminthic eggs to vascular endothelial cells under inflammatory conditions.

HOST RECEPTORS

Host receptors are found both on target cells (such as epithelial cells lining mucosal surfaces) and within the mucus layer covering these cells. Microbial pathogens bind to a wide range of host receptors to establish infection (Table 120-1). Selective loss of host receptors for a pathogen may confer natural resistance to an otherwise susceptible population. For example, 70% of individuals in West Africa lack Fy antigens and are resistant to *P. vivax* infection. *S. enterica* serovar Typhi, the etiologic agent of typhoid fever, produces a pilus protein that binds to CFTR to enter the gastrointestinal submucosa after being ingested by enterocytes. As homozygous mutations in *CFTR* are the cause of the life-shortening disease cystic fibrosis, heterozygote carriers (e.g., 4–5% of individuals of European ancestry) may have had a selective advantage due to decreased susceptibility to typhoid fever. Genetic polymorphisms in CFTR besides those leading to cystic fibrosis have been associated with resistance to typhoid fever.

Numerous virus–target cell interactions have been described, and it is now clear that different viruses can use similar host cell receptors for entry. The list of certain and likely host receptors for viral pathogens is long. Among the host membrane components that can serve as receptors for viruses are sialic acids, gangliosides, glycosaminoglycans, integrins and other members of the immunoglobulin superfamily, histocompatibility antigens, and regulators and receptors for complement components. A notable example of the effect of host receptors on the pathogenesis of infection has emerged from studies comparing the binding of avian influenza A subtype H5N1 with that of influenza A strains expressing the H1 subtype of hemagglutinin. The H1 subtypes tend to be highly pathogenic and transmissible from human to human, and they bind to a receptor composed of two sugar molecules: sialic acid linked α;-2-6 to galactose. This receptor is highly expressed in the airway epithelium; when virus is shed from this surface, its transmission via coughing and aerosol droplets is facilitated. In contrast, the H5N1 avian influenza virus binds to sialic acid linked α;-2-3 to galactose, and this receptor is highly expressed in pneumocytes in the alveoli. Infection in the alveoli is thought to underlie the high mortality rate associated with avian influenza but also the low interhuman transmissibility of this strain, which is not readily transported to the airways from which it can be expelled by coughing.

Microbial Growth after Entry

Once established on a mucosal or skin site, pathogenic microbes must replicate before causing full-blown infection and disease. Within cells, viral particles release their nucleic acids, which may be directly translated into viral proteins (positive-strand RNA viruses), transcribed from a negative strand of RNA into a complementary mRNA (negative-strand RNA viruses), or transcribed into a complementary strand of DNA (retroviruses); for DNA viruses, mRNA may be transcribed directly from viral DNA, either in the cell nucleus or in the cytoplasm. To grow, bacteria must acquire specific

nutrients or synthesize them from precursors in host tissues. Many infectious processes are usually confined to specific epithelial surfaces—e.g., H1 subtype influenza to the respiratory mucosa, gonorrhea to the urogenital epithelium, shigellosis to the gastrointestinal epithelium. While there are multiple reasons for this specificity, one important consideration is the ability of these pathogens to obtain from these specific environments the nutrients needed for growth and survival.

Temperature restrictions also play a role in limiting certain pathogens to specific tissues. Rhinoviruses, a cause of the common cold, grow best at 33°C and replicate in cooler nasal tissues but not in the lung. Leprosy lesions due to *Mycobacterium leprae* are found in and on relatively cool body sites. Fungal pathogens that infect the skin, hair follicles, and nails (dermatophyte infections) remain confined to the cooler, exterior, keratinous layer of the epithelium.

A topic of major interest is the ability of many bacterial, fungal, and protozoal species to grow in multicellular masses referred to as *biofilms*. These masses are biochemically and morphologically quite distinct from the free-living individual cells referred to as *planktonic cells*. Growth in biofilms leads to altered microbial metabolism, production of extracellular virulence factors, and decreased susceptibility to biocides, antimicrobial agents, and host defense molecules and cells. *P. aeruginosa* growing on the bronchial mucosa during chronic infection, staphylococci and other pathogens growing on implanted medical devices, and dental pathogens growing on tooth surfaces to form plaques represent several examples of microbial biofilm growth associated with human disease. Many other pathogens can form biofilms during in vitro growth. It is increasingly accepted that this mode of growth contributes to microbial virulence and induction of disease and that biofilm formation can also be an important factor in microbial survival outside the host, promoting transmission to additional susceptible individuals.

Avoidance of Innate Host Defenses

As microbes have probably interacted with mucosal/epithelial surfaces since the emergence of multicellular organisms, it is not surprising that multicellular hosts have a variety of innate surface defense mechanisms that can sense when pathogens are present and contribute to their elimination. The skin is acidic and is bathed with fatty acids toxic to many microbes. Skin pathogens such as staphylococci must tolerate these adverse conditions. Mucosal surfaces are covered by a barrier composed of a thick mucus layer that entraps microbes and facilitates their transport out of the body by such processes as mucociliary clearance, coughing, and urination. Mucous secretions, saliva, and tears contain antibacterial factors such as lysozyme and antimicrobial peptides as well as antiviral factors such as interferons (IFNs). Gastric acidity is inimical to the survival of many ingested pathogens, and most mucosal surfaces—particularly the nasopharynx, the vaginal tract, and the gastrointestinal tract—contain a resident flora of commensal microbes that interfere with the ability of pathogens to colonize and infect a host.

Pathogens that survive these factors must still contend with host endocytic, phagocytic, and inflammatory responses as well as with host genetic factors that determine the degree to which a pathogen can survive and grow. The list of genes whose variants, usually by single-nucleotide polymorphisms, can affect host susceptibility and resistance to infection is rapidly expanding. A classic example is a 32-bp deletion in the gene for the HIV-1 co-receptor known as chemokine receptor 5 (CCR5), which, when present in the homozygous state, confers high-level resistance to HIV-1 infection. The growth of viral pathogens entering skin or mucosal epithelial cells can be limited by a variety of host genetic factors, including production of IFNs, modulation of receptors for viral entry, and age- and hormone-related susceptibility factors; by nutritional status; and even by personal habits such as smoking and exercise.

ENCOUNTERS WITH EPITHELIAL CELLS

Over the past decade, many bacterial pathogens have been shown to enter epithelial cells **(Fig. 120-2)**; the bacteria often use specialized surface structures that bind to receptors, with consequent internalization. However, the exact role and the importance of this process in infection and disease are not well defined for most of these pathogens. Bacterial entry into host epithelial cells is seen as a means for dissemination to adjacent or deeper tissues or as a route to sanctuary to avoid ingestion and killing by professional phagocytes. Epithelial cell entry appears, for instance, to be a critical aspect of dysentery induction by *Shigella*.

Figure 120-2

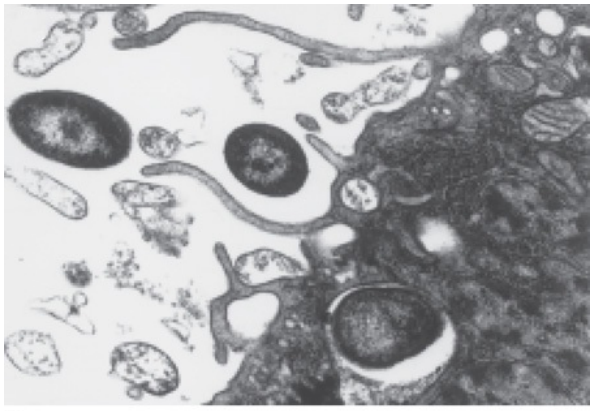

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

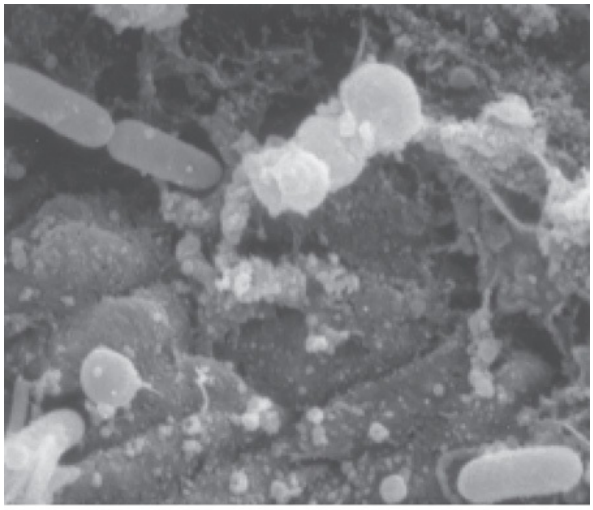

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Entry of bacteria into epithelial cells.*A.* Internalization of *P. aeruginosa* by cultured airway epithelial cells expressing wild-type cystic fibrosis transmembrane conductance regulator, the cell receptor for bacterial ingestion. ***B.*** Entry of *P. aeruginosa* into murine tracheal epithelial cells after murine infection by the intranasal route.

Curiously, the less virulent strains of many bacterial pathogens are more adept at entering epithelial cells than are more virulent strains; examples include pathogens that lack the surface polysaccharide capsule needed to cause serious disease. Thus, for *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Streptococcusagalactiae* (group B *Streptococcus*), and *Streptococcuspyogenes*, isogenic mutants or variants lacking capsules enter epithelial cells better than the wild-type, encapsulated parental forms that cause disseminated disease. These observations have led to the proposal that epithelial cell entry may be primarily a manifestation of host defense, resulting in bacterial clearance by both shedding of epithelial cells containing internalized bacteria and initiation of a protective and nonpathogenic inflammatory response. However, a possible consequence of this process could be the opening of a hole in the epithelium, potentially allowing uningested organisms to enter the submucosa. This scenario has been documented in murine *S. enterica* serovar Typhimurium infections and in experimental bladder infections with uropathogenic *E. coli*. In the latter system, bacterial pilus-mediated attachment to *uroplakins* induces exfoliation of the cells with attached bacteria. Subsequently, infection is produced by residual bacterial cells that invade the superficial bladder epithelium, where they can grow intracellularly into biofilm-like masses encased in an extracellular polysaccharide-rich matrix and surrounded by uroplakin. This mode of growth produces structures that have been referred to as *bacterial pods*. It is likely that at low bacterial inocula epithelial cell ingestion and subclinical inflammation are efficient means to eliminate pathogens, while at higher inocula a proportion of surviving bacterial cells enter the host tissue through the damaged mucosal surface and multiply, producing disease. Alternatively, failure of the appropriate epithelial cell response to a pathogen may allow the organism to survive on a mucosal surface where, if it avoids other host defenses, it can grow and cause a local infection. Along these lines, as noted above, *P. aeruginosa* is taken into epithelial cells by CFTR, a protein missing or nonfunctional in most severe cases of cystic fibrosis. The major clinical consequence of this disease is chronic airway-surface infection with *P. aeruginosa* in 80–90% of patients. The failure of airway epithelial cells to ingest and promote the removal of *P. aeruginosa* via a properly regulated inflammatory response has been proposed as a key component of the hypersusceptibility of these patients to chronic airway infection with this organism.

ENCOUNTERS WITH PHAGOCYTES

Phagocytosis and Inflammation

Phagocytosis of microbes is a major innate host defense that limits the growth and spread of pathogens. Phagocytes appear rapidly at sites of infection in conjunction with the initiation of inflammation. Ingestion of microbes by both tissue-fixed macrophages and migrating phagocytes probably accounts for the limited ability of most microbial agents to cause disease. A family of related molecules called *collectins*, *soluble defense collagens*, or *pattern-recognition molecules* are found in blood (mannose-binding lectins), in lung (surfactant proteins A and D), and most likely in other tissues as well and bind to carbohydrates on microbial surfaces to promote phagocyte clearance. Bacterial pathogens seem to be ingested principally by polymorphonuclear neutrophils, while eosinophils are frequently found at sites of infection by protozoan or multicellular parasites. Successful pathogens, by definition, must avoid being cleared by professional phagocytes. One of several antiphagocytic strategies employed by bacteria and by the fungal pathogen *Cryptococcus neoformans* is to elaborate large-molecular-weight surface polysaccharide antigens, often in the form of a capsule that coats the cell surface. Most pathogenic bacteria produce such antiphagocytic capsules. On occasion, proteins or polypeptides form capsule-like coatings for organisms such as group A streptococci and *Bacillus anthracis*.

As activation of local phagocytes in tissues is a key step in initiating inflammation and migration of additional phagocytes into infected sites, much attention has been paid to microbial factors that initiate inflammation. These are usually conserved factors critical to the microbes' survival and are referred to as *pathogen-associated molecular patterns* (PAMPs). Cellular responses to microbial encounters with phagocytes are governed largely by the structure of the microbial PAMPs that elicit inflammation, and detailed knowledge of these structures of bacterial pathogens has contributed greatly to our understanding of molecular mechanisms of microbial pathogenesis mediated by activation of host cell molecules such as Toll-like receptors (TLRs; **Fig. 120-3)**. One of the best-studied systems involves the interaction of LPS from gram-negative bacteria and the GPI-anchored membrane protein CD14 found on the surface of professional phagocytes, including migrating and tissue-fixed macrophages and polymorphonuclear neutrophils. A soluble form of CD14 is also found in plasma and on mucosal surfaces. A plasma protein, LPS-binding protein, transfers LPS to membrane-bound CD14 on myeloid cells and promotes binding of LPS to soluble CD14. Soluble CD14/LPS/LPS-binding protein complexes bind to many cell types and may be internalized to initiate cellular responses to microbial pathogens. It

has been shown that peptidoglycan and lipoteichoic acid from gram-positive bacteria and cell-surface products of mycobacteria and spirochetes can interact with CD14 (Fig. 120-3). Additional molecules, such as MD-2, also participate in the recognition of bacterial activators of inflammation.

Figure 120-3

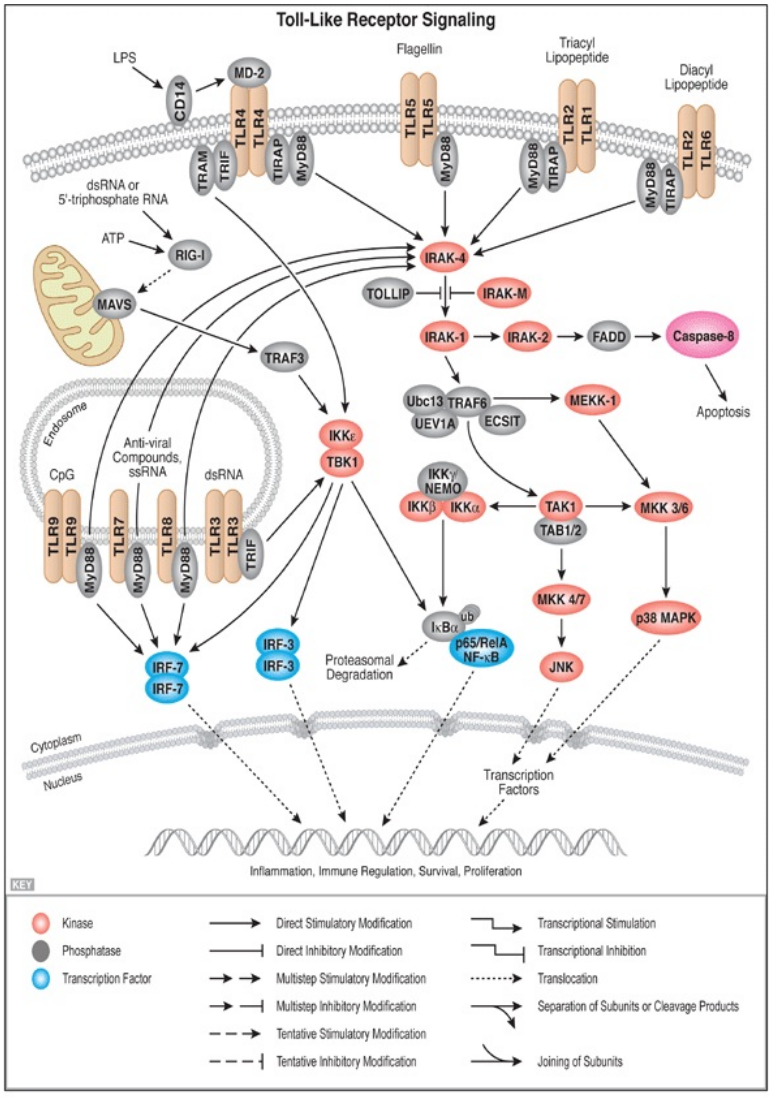

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cellular signaling pathways for production ofinflammatory cytokines in response to microbial products. Microbial cell-surface constituents interact with Toll-like receptors (TLRs), in some cases requiring additional factors such as MD-2, which facilitates the response to lipopolysaccharide (LPS) via TLR4. Although depicted as interacting with the TLRs on the cell surface, TLRs contain extracellular leucine-rich domains that become localized to the lumen of the phagosome upon uptake of bacterial cells. The internalized TLRs can bind to microbial products. The TLRs are oligomerized, usually forming homodimers, and then bind to the general adapter protein MyD88 via the C-terminal Toll/IL-1R (TIR) domains, which also bind to TIRAP (TIR domain-containing adapter protein), a molecule that participates in the transduction of signals from TLRs 1, 2, 4, and 6. The MyD88/TIRAP complex activates signal-transducing molecules such as IRAK-4 (IL-1Rc-associated kinase 4), which in turn activates IRAK-1. This activation can be blocked by IRAK-M and TOLLIP. IRAK-1 activates TRAF 6 (tumor necrosis factor receptor–associated factor 6), TAK-1 (transforming growth factor β–activating kinase 1), and TAB1/2 (TAK1-binding protein 1/2). This signaling complex associates with the ubiquitin-conjugating enzyme Ubc13 and the Ubc-like protein UEV1A to catalyze the formation of a polyubiquitin chain on TRAF6. Polyubiquitination of TRAF6 activates TAK1, which, along with TAB1/2 (a protein that binds to lysine residue 63 in polyubiquitin chains via a conserved zinc-finger domain), phosphorylates the inducible kinase complex: IKK-α, -β, and -γ . IKK-γ is also called NEMO [nuclear factor κB (NF-κB) essential modulator]. This large complex phosphorylates the inhibitory component of NF-κB, $I_\kappa B_\alpha$, resulting in release of $I_\kappa B_\alpha$ from NF-κB. Phosphorylated (PP) $I_\kappa B$ is then ubiquinated (ub) and degraded, and the two components of NF-κB, p50 or Rel and p65, translocate to the nucleus, where they bind to regulatory transcriptional sites on target genes, many of which encode inflammatory proteins. In addition to inducing NF-κB nuclear translocation, the TAK1/TAB1/2 complex activates MAP kinase transducers such as MKK 4/7 and MKK 3/6, which can lead to nuclear translocation of transcription factors such as AP1. TLR4 can also activate NF-κB nuclear translocation via the MyD88-independent TRIF (TIR-domain-containing adapter-inducing IFN-β) and TRAM (TRIF-related adapter molecule) cofactors. Intracellular TLRs 3, 7, 8, and 9 also use MyD88 and TRIF to activate IFN response factors 3 and 7 (IRF-3 and IRF-7), which also function as transcriptional factors in the nucleus. *[Pathway diagram reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com).]*

GPI-anchored receptors do not have intracellular signaling domains; therefore, it is the TLRs that transduce signals for cellular activation due to LPS binding. Binding of microbial factors to TLRs to activate signal transduction occurs in the phagosome—and not on the surface—of dendritic cells that have internalized the microbe. This binding is probably due to the release of the microbial surface factor from the cell in the environment of the phagosome, where the liberated factor

can bind to its cognate TLRs. TLRs initiate cellular activation through a series of signal-transducing molecules (Fig. 120-3) that lead to nuclear translocation of the transcription factor NF-κB, a master-switch for production of important inflammatory cytokines such as tumor necrosis factor β (TNF-β) and interleukin (IL) 1.

The initiation of inflammation can occur not only with LPS and peptidoglycan but also with viral particles and other microbial products such as polysaccharides, enzymes, and toxins. Bacterial flagella activate inflammation by binding of a conserved sequence to TLR5. Some pathogens (e.g., *Campylobacter jejuni, Helicobacter pylori,* and *Bartonella bacilliformis*) make flagella that lack this sequence and do not bind to TLR5; thus efficient host responses to infection are prevented. Bacteria also produce a high proportion of DNA molecules with unmethylated CpG residues that activate inflammation through TLR9. TLR3 recognizes double-stranded RNA, a pattern-recognition molecule produced by many viruses during their replicative cycle. TLR1 and TLR6 associate with TLR2 to promote recognition of acylated microbial proteins and peptides.

The myeloid differentiation factor 88 (MyD88) molecule and the Toll/IL-1R (TIR) domain-containing adapter protein (TIRAP) bind to the cytoplasmic domains of TLRs and also to receptors that are part of the IL-1 receptor families. Numerous studies have shown that MyD88/TIRAP-mediated transduction of signals from TLRs and other receptors is critical for innate resistance to infection, activating MAP-kinases and NF-κB and thereby leading to production of cytokines/chemokines. Mice lacking MyD88 are more susceptible than normal mice to infections with a broad range of pathogens. In one study, nine children homozygous for defective MyD88 genes had recurrent infections with *S. pneumoniae*, *S. aureus*, and *P. aeruginosa*—three bacterial species showing increased virulence in MyD88-deficient mice; however, unlike these mice, the MyD88-deficient children seemed to have no greater susceptibility to other bacteria, viruses, fungi, or parasites. Another component of the MyD88-dependent signaling pathway is a molecule known as IL-1 receptor–associated kinase 4 (IRAK-4). Individuals with a homozygous deficiency in genes encoding this protein are at increased risk for *S. pneumoniae* and *S. aureus* infections and, to some degree, for *P. aeruginosa* infections as well.

In addition to their role in MyD88-mediated signaling, some TLRs (e.g., TLR3 and TLR4) can activate signal transduction via a MyD88-independent pathway involving TIR domain-containing, adapter-inducing IFN-β (TRIF) and the TRIF-related adapter molecule (TRAM). Signaling through TRIF and TRAM activates the production of both NF-κB-dependent cytokines/chemokines and type 1 IFNs. The type 1 IFNs bind to the IFN-α receptor composed of two protein chains, IFNAR1 and IFNAR2. Humans produce three type 1 IFNs: IFN-α, IFN-β, and IFN-γ. These molecules activate another class of proteins known as the signal transducer and activator of transcription (STAT) complexes. The STAT factors are important in regulating immune system genes and thus play a critical role in responding to microbial infections.

Another intracellular complex of proteins found to be a major factor in the host cell response to infection is the inflammasome **(Fig. 120-4)**, where inflammatory cytokines IL-1 and Il-18 are changed from their precursor to active forms prior to secretion by the cysteineprotease caspase-1. Within the inflammasome are additional proteins that are members of the nucleotide binding and oligomerization domain (NOD)-like receptor (NLR) family. Like the TLRs, NOD proteins sense the presence of the conserved microbial factors released inside a cell. Recognition of these PAMPs by NLRs leads to caspase-1 activation and to secretion of active IL-1 and IL-18 by an unknown mechanism. Studies of mice indicate that as many as four inflammasomes with different components are formed: the IPAF inflammasome, the NALP1 inflammasome, the cryopyrin/NALP3 inflammasome, and an inflammasome triggered by *Francisella tularensis* infection (Fig. 120-4). The components depend on the type of stimulus driving inflammasome formation and activation.

Figure 120-4

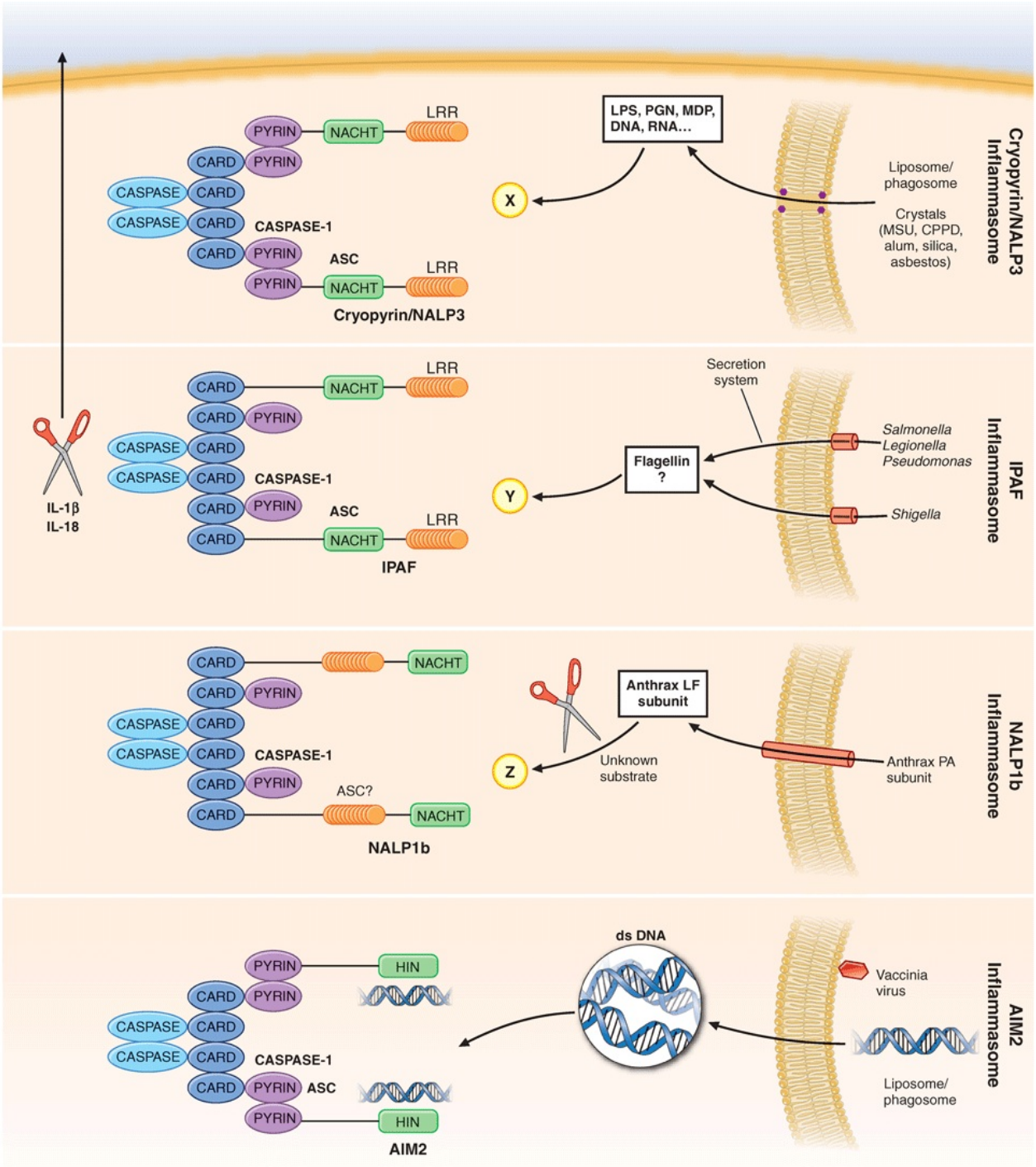

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The NOD-like receptor (NLR) proteins NALP1b, cryopyrin/NALP3, and IPAF and the HIN-200 protein AIM2 assemble a caspase-1-activating inflammasome complex in response to specific microbial or bacterial factors. The murine NALP1b inflammasome recognizes the cytosolic presence of anthrax lethal toxin. The cryopyrin/NALP3 inflammasome recognizes multiple pathogen-associated molecular patterns (PAMPs) in combination with ATP or nigericin as well as crystalline substances including MSU, silica, and asbestos particles. The IPAF inflammasome senses *Salmonella* and *Legionella* flagellin and a yet-unidentified *Shigella flexneri* compound, all of which access the cytosol through a type III or IV secretion system. Cytosolic PAMPs may trigger assembly of a particular inflammasome complex by causing modifications in unknown host factors (X, Y, Z) that are monitored by specific NLR proteins. In contrast, AIM2 directly binds dsDNA in the cytosol to induce caspase-1 activation. The CARD/pyrin-containing adapter protein ASC is essential for all these inflammasome complexes, although its role in the NALP1b inflammasome remains to be formally established. Once activated, caspase-1 processes IL-1β and IL-18 precursors into the mature cytokines, which are secreted through an unknown mechanism. *[Figure and legend from Lamkanfi M, Dixit VM (2009) The Inflammasomes. PLoS Pathog 5(12): e1000510. doi:10.1371/journal.ppat.1000510.*]

Additional Interactions of Microbial Pathogens and Phagocytes

Other ways that microbial pathogens avoid destruction by phagocytes include production of factors that are toxic to the phagocytes or that interfere with the chemotactic and ingestion function of phagocytes. Hemolysins, leukocidins, and the like are microbial proteins that can kill phagocytes that are attempting to ingest organisms elaborating these substances. For example, staphylococcal hemolysins inhibit macrophage chemotaxis and kill these phagocytes. Streptolysin O made

by *S. pyogenes* binds to cholesterol in phagocyte membranes and initiates a process of internal degranulation, with the release of normally granule-sequestered toxic components into the phagocyte's cytoplasm. *E. histolytica*, an intestinal protozoan that causes amebic dysentery, can disrupt phagocyte membranes after direct contact via the release of protozoal phospholipase A and pore-forming peptides.

Microbial Survival Inside Phagocytes

Many important microbial pathogens use a variety of strategies to survive inside phagocytes (particularly macrophages) after ingestion. Inhibition of fusion of the phagocytic vacuole (the phagosome) containing the ingested microbe with the lysosomal granules containing antimicrobial substances (the lysosome) allows *Mycobacterium tuberculosis*, *S. enterica* serovar Typhi, and *Toxoplasma gondii* to survive inside macrophages. Some organisms, such as *Listeria monocytogenes*, escape into the phagocyte's cytoplasm to grow and eventually spread to other cells. Resistance to killing within the macrophage and subsequent growth are critical to successful infection by herpes-type viruses, measles virus, poxviruses, *Salmonella*, *Yersinia*, *Legionella*, *Mycobacterium*, *Trypanosoma*, *Nocardia*, *Histoplasma*, *Toxoplasma*, and *Rickettsia*. *Salmonella* species use a master regulatory system—in which the *PhoP*/*PhoQ* genes control other genes—to enter and survive within cells, with intracellular survival entailing structural changes in the cell envelope LPS.

Tissue Invasion and Tissue Tropism

TISSUE INVASION

Most viral pathogens cause disease by growth at skin or mucosal entry sites, but some pathogens spread from the initial site to deeper tissues. Virus can spread via the nerves (rabies virus) or plasma (picornaviruses) or within migratory blood cells (poliovirus, Epstein-Barr virus, and many others). Specific viral genes determine where and how individual viral strains can spread.

Bacteria may invade deeper layers of mucosal tissue via intracellular uptake by epithelial cells, traversal of epithelial cell junctions, or penetration through denuded epithelial surfaces. Among virulent *Shigella* strains and invasive *E. coli*, outer-membrane proteins are critical to epithelial cell invasion and bacterial multiplication. *Neisseria* and *Haemophilus* species penetrate mucosal cells by poorly understood mechanisms before dissemination into the bloodstream. Staphylococci and streptococci elaborate a variety of extracellular enzymes, such as hyaluronidase, lipases, nucleases, and hemolysins, that are probably important in breaking down cellular and matrix structures and allowing the bacteria access to deeper tissues and blood. Organisms that colonize the gastrointestinal tract can often translocate through the mucosa into the blood and, under circumstances in which host defenses are inadequate, cause bacteremia. *Y. enterocolitica* can invade the mucosa through the activity of the invasin protein. Some bacteria (e.g., *Brucella*) can be carried from a mucosal site to a distant site by phagocytic cells (e.g., polymorphonuclear neutrophils) that ingest but fail to kill the bacteria.

Fungal pathogens almost always take advantage of host immunocompromise to spread hematogenously to deeper tissues. The AIDSepidemic has resoundingly illustrated this principle: the immunodeficiency of many HIV-infected patients permits the development of life-threatening fungal infections of the lung, blood, and brain. Other than the capsule of *C. neoformans*, specific fungal antigens involved in tissue invasion are not well characterized. Both fungal pathogens and protozoal pathogens (e.g., *Plasmodium* species and *E. histolytica*) undergo morphologic changes to spread within a host. Malarial parasites grow in liver cells as merozoites and are released into the blood to invade erythrocytes and become trophozoites. *E. histolytica* is found as both a cyst and a trophozoite in the intestinal lumen, through which this pathogen enters the host, but only the trophozoite form can spread systemically to cause amebic liver abscesses. Other protozoal pathogens, such as *T. gondii*, *Giardia lamblia*, and *Cryptosporidium*, also undergo extensive morphologic changes after initial infection to spread to other tissues.

TISSUE TROPISM

The propensity of certain microbes to cause disease by infecting specific tissues has been known since the early days of bacteriology, yet the molecular basis for this propensity is understood somewhat better for viral pathogens than for other agents of infectious disease. Specific receptor-ligand interactions clearly underlie the ability of certain viruses to enter cells within tissues and disrupt normal tissue function, but the mere presence of a receptor for a virus on a target tissue is not sufficient for tissue tropism. Factors in the cell, route of viral entry, viral capacity to penetrate into cells, viral genetic elements that regulate gene expression, and pathways of viral spread in a tissue all affect tissue tropism. Some viral genes are best transcribed in specific target cells, such as hepatitis B genes in liver cells and Epstein-Barr virus genes in B lymphocytes. The route of inoculation of poliovirus determines its neurotropism, although the molecular basis for this circumstance is not understood.

Compared with viral tissue tropism, the tissue tropism of bacterial and parasitic infections has not been as clearly elucidated, but studies of *Neisseria* species have provided insights. Both *N.gonorrhoeae,* which colonizes and infects the human genital tract, and *N.meningitidis*, which principally colonizes the human oropharynx but can spread to the brain, produce type IV pili (Tfp) that mediate adherence to host tissues. In the case of *N. gonorrhoeae*, the Tfp bind to a glucosamine-galactose-containing adhesin on the surface of cervical and urethral cells; in the case of *N. meningitidis*, the Tfp bind to cells in the human meninges in order to cross the blood-brain barrier. *N. meningitidis* expresses a capsular polysaccharide, while *N. gonorrhoeae* does not; however, there is no indication that this property plays a role in the different tissue tropisms displayed by these two bacterial species. *N. gonorrhoeae* can use cytidine monophosphate *N*-acetylneuraminic acid fromhost tissues to add *N*-acetylneuraminic acid (sialic acid) to its lipooligosaccharide O side chain, and this alteration appears to make the organism resistant to host defenses. Lactate, present at high levels on genital mucosal surfaces, stimulates sialylation of gonococcal lipooligosaccharide. Bacteria with sialic acid sugars in their capsules, such as *N. meningitidis*, *E. coli* K1, and group B streptococci, have a propensity to cause meningitis, but this generalization has many exceptions. For example, all recognized serotypes of group B streptococci contain sialic acid in their capsules, but only one serotype (III) is responsible for most cases of group B streptococcal meningitis. Moreover, both *H. influenzae* and *S. pneumoniae* can readily cause meningitis, but these organisms do not have sialic acid in their capsules.

Tissue Damage and Disease

Disease is a complex phenomenon resulting from tissue invasion and destruction, toxin elaboration, and host response. Viruses cause much of their damage by exerting a cytopathic effect on host cells and inhibiting host defenses. The growth of bacterial, fungal, and protozoal parasites in tissue, which may or may not be accompanied by toxin elaboration, can also compromise tissue function and lead to disease. For some bacterial and possibly some fungal pathogens, toxin production is one of the best-characterized molecular mechanisms of pathogenesis, while host factors such as IL-1, TNF-α, kinins, inflammatory proteins, products of complement activation, and mediators derived from arachidonic acid metabolites

(leukotrienes) and cellular degranulation (histamines) readily contribute to the severity of disease.

VIRAL DISEASE

Viral pathogens are well known to inhibit host immune responses by a variety of mechanisms. Immune responses can be affected by decreasing production of most major histocompatibility complex molecules (adenovirus E3 protein), by diminishing cytotoxic T cell recognition of virus-infected cells (Epstein-Barr virus EBNA1 antigen and cytomegalovirus IE protein), by producing virus-encoded complement receptor proteins that protect infected cells from complement-mediated lysis (herpesvirus and vaccinia virus), by making proteins that interfere with the action of IFN (influenza virus and poxvirus), and by elaborating superantigen-like proteins (mouse mammary tumor virus and related retroviruses and the rabies nucleocapsid). Superantigens activate large populations of T cells that express particular subsets of the T cell receptor β protein, causing massive cytokine release and subsequent host reactions. Another molecular mechanism of viral virulence involves the production of peptide growth factors for host cells, which disrupt normal cellular growth, proliferation, and differentiation. In addition, viral factors can bind to and interfere with the function of host receptors for signaling molecules. Modulation of cytokine production during viral infection can stimulate viral growth inside cells with receptors for the cytokine, and virus-encoded cytokine homologues (e.g., the Epstein-Barr virus BCRF1 protein, which is highly homologous to the immunoinhibitory IL-10 molecule) can potentially prevent immune-mediated clearance of viral particles. Viruses can cause disease in neural cells by interfering with levels of neurotransmitters without necessarily destroying the cells, or they may induce either programmed cell death (apoptosis) to destroy tissues or inhibitors of apoptosis to allow for prolonged viral infection of cells. For infection to spread, many viruses must be released from cells. In a newly identified function, viral protein U (Vpu) of HIV facilitates the release of virus, a process that is specific to certain cells. Mammalian cells produce a restriction factor involved in inhibiting the release of virus; for HIV, this factor is designated BST-2 (bone marrow stromal antigen 2)/HM1.24/CD317, or tetherin. Vpu of HIV interacts with tetherin, promoting release of infectious virus. Overall, disruption of normal cellular and tissue function due to viral infection, replication, and release promotes clinical disease.

BACTERIAL TOXINS

Among the first infectious diseases to be understood were those due to toxin-elaborating bacteria. Diphtheria, botulism, and tetanus toxins are responsible for the diseases associated with local infections due to *Corynebacterium diphtheriae*, *Clostridium botulinum*, and *Clostridium tetani*, respectively. Enterotoxins produced by *E. coli*, *Salmonella*, *Shigella*, *Staphylococcus*, and *V. cholerae* contribute to diarrheal disease caused by these organisms. Staphylococci, streptococci, *P. aeruginosa*, and *Bordetella* elaborate various toxins that cause or contribute to disease, including toxic shock syndrome toxin 1; erythrogenic toxin; exotoxins A, S, T and U; and pertussis toxin. A number of these toxins (e.g., cholera toxin, diphtheria toxin, pertussis toxin, *E. coli* heat-labile toxin, and *P. aeruginosa* exotoxin) have adenosine diphosphate ribosyl transferase activity; i.e., the toxins enzymatically catalyze the transfer of the adenosine diphosphate ribosyl portion of nicotinamide adenine diphosphate to target proteins and inactivate them. The staphylococcal enterotoxins, toxic shock syndrome toxin 1, and the streptococcal pyogenic exotoxins behave as superantigens, stimulating certain T cells to proliferate without processing of the protein toxin by antigen-presenting cells. Part of this process involves stimulation of the antigen-presenting cells to produce IL-1 and TNF-α, which have been implicated in many clinical features of diseases like toxic shock syndrome and scarlet fever. A number of gram-negative pathogens (*Salmonella*, *Yersinia*, and *P. aeruginosa*) can inject toxins directly into host target cells by means of a complex set of proteins referred to as the type III secretion system. Loss or inactivation of this virulence system usually greatly reduces the capacity of a bacterial pathogen to cause disease.

ENDOTOXIN

The lipid A portion of gram-negative LPS has potent biologic activities that cause many of the clinical manifestations of gram-negative bacterial sepsis, including fever, muscle proteolysis, uncontrolled intravascular coagulation, and shock. The effects of lipid A appear to be mediated by the production of potent cytokines due to LPS binding to CD14 and signal transduction via TLRs, particularly TLR4. Cytokines exhibit potent hypothermic activity through effects on the hypothalamus; they also increase vascular permeability, alter the activity of endothelial cells, and induce endothelial-cell procoagulant activity. Numerous therapeutic strategies aimed at neutralizing the effects of endotoxin are under investigation, but so far the results have been disappointing. One drug, activated protein C, was found to reduce mortality rates by ~20% during severe sepsis, a condition that can be induced by endotoxin release during gram-negative bacterial sepsis.

INVASION

Many diseases are caused primarily by pathogens growing in tissue sites that are normally sterile. Pneumococcal pneumonia is mostly attributable to the growth of *S. pneumoniae* in the lung and the attendant host inflammatory response, although specific factors that enhance this process (e.g., pneumolysin) may be responsible for some of the pathogenic potential of the pneumococcus. Disease that follows bacteremia and invasion of the meninges by meningitis-producing bacteria such as *N. meningitidis*, *H. influenzae*, *E. coli* K1, and group B streptococci appears to be due solely to the ability of these organisms to gain access to these tissues, multiply in them, and provoke cytokine production leading to tissue-damaging host inflammation.

Specific molecular mechanisms accounting for tissue invasion by fungal and protozoal pathogens are less well described. Except for studies pointing to factors like capsule and melanin production by *C. neoformans* and possibly levels of cell wall glucans in some pathogenic fungi, the molecular basis for fungal invasiveness is not well defined. Melanism has been shown to protect the fungal cell against death caused by phagocyte factors such as nitric oxide, superoxide, and hypochlorite. Morphogenic variation and production of proteases (e.g., the *Candida* aspartyl proteinase) have been implicated in fungal invasion of host tissues.

If pathogens are effectively to invade host tissues (particularly the blood), they must avoid the major host defenses represented by complement and phagocytic cells. Bacteria most often avoid these defenses through their surface polysaccharides—either capsular polysaccharides or long O-side-chain antigens characteristic of the smooth LPS of gram-negative bacteria. These molecules can prevent the activation and/or deposition of complement opsonins or can limit the access of phagocytic cells with receptors for complement opsonins to these molecules when they are deposited on the bacterial surface below the capsular layer. Another potential mechanism of microbial virulence is the ability of some organisms to present the capsule as an apparent self antigen through molecular mimicry. For example, the polysialic acid capsule of group B*N. meningitidis* is chemically identical to an oligosaccharide found on human brain cells.

Immunochemical studies of capsular polysaccharides have led to an appreciation of the tremendous chemical diversity that can result from the linking of a few monosaccharides. For example, three hexoses can link up in more than 300

different, potentially serologically distinct ways, while three amino acids have only six possible peptide combinations. Capsular polysaccharides have been used as effective vaccines against meningococcal meningitis as well as against pneumococcal and *H. influenzae* infections and may prove to be of value as vaccines against any organisms that express a nontoxic, immunogenic capsular polysaccharide. In addition, most encapsulated pathogens become virtually avirulent when capsule production is interrupted by genetic manipulation; this observation emphasizes the importance of this structure in pathogenesis.

HOST RESPONSE

The inflammatory response of the host is critical for interruption and resolution of the infectious process but also is often responsible for the signs and symptoms of disease. Infection promotes a complex series of host responses involving the complement, kinin, and coagulation pathways. The production of cytokines such as IL-1, IL-18, TNF-α, and other factors regulated in part by the NF-κB transcription factor leads to fever, muscle proteolysis, and other effects. An inability to kill or contain the microbe usually results in further damage due to the progression of inflammation and infection. For example, in many chronic infections, degranulation of host inflammatory cells can lead to release of host proteases, elastases, histamines, and other toxic substances that can degrade host tissues. Chronic inflammation in any tissue can lead to the destruction of that tissue and to clinical disease associated with loss of organ function, such as sterility from pelvic inflammatory disease caused by chronic infection with *N. gonorrhoeae*.

The nature of the host response elicited by the pathogen often determines the pathology of a particular infection. Local inflammation produces local tissue damage, while systemic inflammation, such as that seen during sepsis, can result in the signs and symptoms of septic shock. The severity of septic shock is associated with the degree of production of host effectors. Disease due to intracellular parasitism results from the formation of granulomas, wherein the host attempts to wall off the parasite inside a fibrotic lesion surrounded by fused epithelial cells that make up so-called multinucleated giant cells. A number of pathogens, particularly anaerobic bacteria, staphylococci, and streptococci, provoke the formation of an abscess, probably because of the presence of zwitterionic surface polysaccharides such as the capsular polysaccharide of *Bacteroides fragilis*. The outcome of an infection depends on the balance between an effective host response that eliminates a pathogen and an excessive inflammatory response that is associated with an inability to eliminate a pathogen and with the resultant tissue damage that leads to disease.

Transmission to New Hosts

As part of the pathogenic process, most microbes are shed from the host, often in a form infectious for susceptible individuals. However, the rate of transmissibility may not necessarily be high, even if the disease is severe in the infected individual, as these traits are not linked. Most pathogens exit via the same route by which they entered: respiratory pathogens by aerosols from sneezing or coughing or through salivary spread, gastrointestinal pathogens by fecal-oral spread, sexually transmitted diseases by venereal spread, and vector-borne organisms by either direct contact with the vector through a blood meal or indirect contact with organisms shed into environmental sources such as water. Microbial factors that specifically promote transmission are not well characterized. Respiratory shedding is facilitated by overproduction of mucous secretions, with consequently enhanced sneezing and coughing. Diarrheal toxins such as cholera toxin, *E. coli* heat-labile toxins, and *Shigella* toxins probably facilitate fecal-oral spread of microbial cells in the high volumes of diarrheal fluid produced during infection. The ability to produce phenotypic variants that resist hostile environmental factors (e.g., the highly resistant cysts of *E. histolytica* shed in feces) represents another mechanism of pathogenesis relevant to transmission. Blood parasites such as *Plasmodium* species change phenotype after ingestion by a mosquito—a prerequisite for the continued transmission of this pathogen. Venereally transmitted pathogens may undergo phenotypic variation due to the production of specific factors to facilitate transmission, but shedding of these pathogens into the environment does not result in the formation of infectious foci.

In summary, the molecular mechanisms used by pathogens to colonize, invade, infect, and disrupt the host are numerous and diverse. Each phase of the infectious process involves a variety of microbial and host factors interacting in a manner that can result in disease. Recognition of the coordinated genetic regulation of virulence factor elaboration when organisms move from their natural environment into the mammalian host emphasizes the complex nature of the host-parasite interaction. Fortunately, the need for diverse factors in successful infection and disease implies that a variety of therapeutic strategies may be developed to interrupt this process and thereby prevent and treat microbial infections.

FURTHER READINGS

Arana DM, Prieto D: The role of the cell wall in fungal pathogenesis. Microb Biotechnol 2:308, 2009 [PMID: 21261926]

Chai LY et al: Fungal strategies for overcoming host innate immune response. Med Mycol 47:227, 2009 [PMID: 18654922]

Diacovich L, Gorvel JP: Bacterial manipulation of innate immunity to promote infection. Nat Rev Microbiol 8:117, 2010 [PMID: 20075926]

Duan K et al: Chemical interactions between organisms in microbial communities. Contrib Microbiol 16:1, 2009 [PMID: 19494576]

Imberty A, Varrot A: Microbial recognition of human cell surface glycoconjugates. Curr Opin Struct Biol 18:567, 2008 [PMID: 18809496]

IWASAKI A, MEDZHITOV R: Regulation of adaptive immunity by the innate immune system. Science 327:291, 2010 [PMID: 20075244]

Kawai T, Akira S: The roles of TLRs, RLRs and NLRs in pathogen recognition. Int Immunol 21:317, 2009 [PMID: 19246554]

Kline KA et al: Bacterial adhesins in host-microbe interactions. Cell Host Microbe 5:580, 2009 [PMID: 19527885]

Stutz A et al: Inflammasomes: Too big to miss. J Clin Invest 119:3502, 2009 [PMID: 19955661]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE ACUTELY ILL INFECTED FEBRILE PATIENT: INTRODUCTION

The physician treating the acutely ill febrile patient must be able to recognize infections that require emergent attention. If such infections are not adequately evaluated and treated at initial presentation, the opportunity to alter an adverse outcome may be lost. In this chapter, the clinical presentations of and approach to patients with relatively common infectious disease emergencies are discussed. These infectious processes and their treatments are discussed in detail in other chapters.

Approach to the Patient Acute Febrile Illness

Before the history is elicited and a physical examination performed, an immediate assessment of the patient's general appearance can yield valuable information. The perceptive physician's subjective sense that a patient is septic or toxic often proves accurate. Visible agitation or anxiety in a febrile patient can be a harbinger of critical illness.

HISTORY

Presenting symptoms are frequently nonspecific. Detailed questions should be asked about the onset and duration of symptoms and about changes in severity or rate of progression over time. Host factors and comorbid conditions may enhance the risk of infection with certain organisms or of a more fulminant course than is usually seen. Lack of splenic function, alcoholism with significant liver disease, IV drug use, HIV infection, diabetes, malignancy, organ transplantation, and chemotherapy all predispose to specific infections and frequently to increased severity. The patient should be questioned about factors that might help identify a nidus for invasive infection, such as recent upper respiratory tract infections, influenza, or varicella; prior trauma; disruption of cutaneous barriers due to lacerations, burns, surgery, body piercing, or decubiti; and the presence of foreign bodies, such as nasal packing after rhinoplasty, tampons, or prosthetic joints. Travel, contact with pets or other animals, or activities that might result in tick or mosquito exposure can lead to diagnoses that would not otherwise be considered. Recent dietary intake, medication use, social or occupational contact with ill individuals, vaccination history, recent sexual contacts, and menstrual history may be relevant. A review of systems should focus on any neurologic signs or sensorium alterations, rashes or skin lesions, and focal pain or tenderness and should also include a general review of respiratory, gastrointestinal, or genitourinary symptoms.

PHYSICAL EXAMINATION

A complete physical examination should be performed, with special attention to several areas that are sometimes given short shrift in routine examinations. Assessment of the patient's general appearance and vital signs, skin and soft tissue examination, and the neurologic evaluation are of particular importance.

The patient may appear either anxious and agitated or lethargic and apathetic. Fever is usually present, although elderly patients and compromised hosts [e.g., patients who are uremic or cirrhotic and those who are taking glucocorticoids or nonsteroidal anti-inflammatory drugs (NSAIDs)] may be afebrile despite serious underlying infection. Measurement of blood pressure, heart rate, and respiratory rate helps determine the degree of hemodynamic and metabolic compromise. The patient's airway must be evaluated to rule out the risk of obstruction from an invasive oropharyngeal infection.

The etiologic diagnosis may become evident in the context of a thorough skin examination (Chap. 17). Petechial rashes are typically seen with meningococcemia or Rocky Mountain spotted fever (RMSF; see Fig. e7-16); erythroderma is associated with toxic shock syndrome (TSS) and drug fever. The soft tissue and muscle examination is critical. Areas of erythema or duskiness, edema, and tenderness may indicate underlying necrotizing fasciitis, myositis, or myonecrosis. The neurologic examination must include a careful assessment of mental status for signs of early encephalopathy. Evidence of nuchal rigidity or focal neurologic findings should be sought.

DIAGNOSTIC WORKUP

After a quick clinical assessment, diagnostic material should be obtained rapidly and antibiotic and supportive treatment begun. Blood (for cultures; baseline complete blood count with differential; measurement of serum electrolytes, blood urea nitrogen, serum creatinine, and serum glucose; and liver function tests) can be obtained at the time an IV line is placed and before antibiotics are administered. Three sets of blood cultures should be performed for patients with possible acute endocarditis. Asplenic patients should have a blood smear examined to confirm the presence of

Howell-Jolly bodies (indicating the absence of splenic function) and a buffy coat examined for bacteria; these patients can have >10^6 organisms per milliliter of blood (compared with 10^4/mL in patients with an intact spleen). Blood smears from patients at risk for severe parasitic disease, such as malaria or babesiosis (see Chap. e27), must be examined for the diagnosis and quantitation of parasitemia. Blood smears may also be diagnostic in ehrlichiosis and anaplasmosis.

Patients with possible meningitis should have cerebrospinal fluid (CSF) obtained before the initiation of antibiotic therapy. Focal findings, depressed mental status, or papilledema should be evaluated by brain imaging prior to lumbar puncture, which, in this setting, could initiate herniation. *Antibiotics should be administered before imaging but after blood for cultures has been drawn.* If CSF cultures are negative, blood cultures will provide the diagnosis in 50–70% of cases.

Focal abscesses necessitate immediate CT or MRI as part of an evaluation for surgical intervention. Other diagnostic procedures, such as wound cultures, should not delay the initiation of treatment for more than minutes. Once emergent evaluation, diagnostic procedures, and (if appropriate) surgical consultation (see below) have been completed, other laboratory tests can be conducted. Appropriate radiography, computed axial tomography, MRI, urinalysis, erythrocyte sedimentation rate (ESR) determination, and transthoracic or transesophageal echocardiography may all prove important.

Treatment: The Acutely Ill Patient

In the acutely ill patient, empirical antibiotic therapy is critical and should be administered without undue delay. Increased prevalence of antibiotic resistance in community-acquired bacteria must be considered when antibiotics are selected. **Table 121-1** lists first-line treatments for infections considered in this chapter. In addition to the rapid initiation of antibiotic therapy, several of these infections require urgent surgical attention. Neurosurgical evaluation for subdural empyema, otolaryngologic surgery for possible mucormycosis, and cardiothoracic surgery for critically ill patients with acute endocarditis are as important as antibiotic therapy. For infections such as necrotizing fasciitis and clostridial myonecrosis, rapid surgical intervention supersedes other diagnostic or therapeutic maneuvers.

Table 121–1. Empirical Treatment for Common Infectious Disease Emergencies

Clinical Syndrome	Possible Etiologies	Treatment	Comments	See Chap.
Sepsis Without a Clear Focus				
Septic shock	Pseudomonas spp., gram-negative enteric bacilli, Staphylococcus spp., Streptococcus spp.	Vancomycin (1 g q12h) ***plus*** Gentamicin (5 mg/kg per day) ***plus*** ***either*** Piperacillin/tazobactam (3.375 g q4h) ***or*** Cefepime (2 g q12h)	Adjust treatment when culture data become available. Drotrecogin alfa (activated)[a] or low-dose hydrocortisone and fludrocortisone[b] may improve outcome in patients with septic shock.	135, 136, 149, 152, 271
Overwhelming post splenectomy sepsis	Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis	Ceftriaxone (2 g q12h) ***plus*** Vancomycin (1 g q12h)	If a β-lactam-sensitive strain is identified, vancomycin can be discontinued.	271
Babesiosis	Babesia microti (U.S.), B. divergens (Europe)	***Either:*** Clindamycin (600 mg tid) ***plus*** Quinine (650 mg tid) ***or:*** Atovaquone (750 mg q12h) ***plus*** Azithromycin (500-mg loading dose, then 250 mg/d)	Atovaquone and azithromycin are as effective as clindamycin and quinine and are associated with fewer side effects. Treatment with doxycycline (100 mg bid[c]) for potential co-infection with Borrelia burgdorferi or Anaplasma spp. may be prudent.	208, 211
Sepsis with Skin Findings				
Meningococcemia	N. meningitidis	Penicillin (4 mU q4h) ***or*** Ceftriaxone (2 g q12h)	Consider protein C replacement in fulminant meningococcemia.	143

Rocky Mountain spotted fever (RMSF)	Rickettsia rickettsii	Doxycycline (100 mg bid)	If both meningococcemia and RMSF are being considered, use ceftriaxone (2 g q12h) plus doxycycline (100 mg bid[c]) or chloramphenicol alone (50–75 mg/kg per day in four divided doses).If RMSF is diagnosed, doxycycline is the proven superior agent.	174
Purpura fulminans	S. pneumoniae, H. influenzae, N. meningitidis	Ceftriaxone (2 g q12h) ***plus*** Vancomycin (1 g q12h)	If a β-lactam–sensitive strain is identified, vancomycin can be discontinued.	134, 143, 145, 271
Erythroderma: toxic shock syndrome	Group A Streptococcus, Staphylococcus aureus	Vancomycin (1 g q12h) ***plus*** Clindamycin (600 mg q8h)	If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 2 mU q4h; or oxacillin, 2 g q4h). The site of toxigenic bacteria should be debrided; IV immunoglobulin can be used in severe cases.[d]	135, 136
Sepsis with Soft Tissue Findings				
Necrotizing fasciitis	Group A Streptococcus, mixed aerobic/anaerobic flora, CA-MRSA[e]	Vancomycin (1 g q12h) ***plus*** Clindamycin (600 mg q8h) ***plus*** Gentamicin (5 mg/kg per day)	Urgent surgical evaluation is critical. If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 2 mU q4h; or oxacillin, 2 g q4h).	125, 135, 136
Clostridial myonecrosis	Clostridium perfringens	Penicillin (2 mU q4h) ***plus*** Clindamycin (600 mg q8h)	Urgent surgical evaluation is critical.	142
Neurologic Infections				
Bacterial meningitis	S. pneumoniae, N. meningitidis	Ceftriaxone (2 g q12h) ***plus*** Vancomycin (1 g q12h)	If a β-lactam–sensitive strain is identified, vancomycin can be discontinued. If the patient is >50 years old or has comorbid disease, add ampicillin (2 g q4h) for Listeria coverage. Dexamethasone (10 mg q6h x 4 days) improves outcome in adult patients with meningitis (especially pneumococcal) and cloudy CSF, positive CSF Gram's stain, or a CSF leukocyte count >1000/mL.	381
Brain abscess, suppurative intracranial infections	Streptococcus spp., Staphylococcus spp., anaerobes, gram-negative bacilli	Vancomycin (1 g q12h) ***plus*** Metronidazole (500 mg q8h) ***plus*** Ceftriaxone (2 g q12h)	Urgent surgical evaluation is critical. If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 4 mU q4h; or oxacillin, 2 g q4h).	381
Cerebral malaria	Plasmodium falciparum	Artesunate (2.4 mg/kg IV at 0, 12, and 24 h; then once daily)[f] ***or*** Quinine (IV loading dose of 20 mg salt/kg; then 10 mg/kg	Do not use glucocorticoids. Use IV quinidine if IV quinine is not available. During IV quinidine treatment, blood pressure and cardiac function should be	208, 210

		q8h) ***plus*** Doxycycline (100 mg IV q12h)	monitored continuously and blood glucose monitored periodically.	
Spinal epidural abscess	Staphylococcus spp., gram-negative bacilli	Vancomycin (1 g q12h) ***plus*** Ceftriaxone (2 g q24h)	Surgical evaluation is essential. If a penicillin- or oxacillin-sensitive strain is isolated, these agents are superior to vancomycin (penicillin, 4 mU q4h; or oxacillin, 2 g q4h).	377
Focal Infections				
Acute bacterial endocarditis	S. aureus, β-hemolytic streptococci, HACEK group,[g] Neisseria spp., S. pneumoniae	Ceftriaxone (2 g q12h) ***plus*** Vancomycin (1 g q12h)	Adjust treatment when culture data become available. Surgical evaluation is essential.	124

[a]Drotrecogin alfa (activated) is administered at a dose of 24 μg/kg per hour for 96 h. It has been approved for use in patients with severe sepsis and a high risk of death as defined by an Acute Physiology and Chronic Health Evaluation II (APACHE II) score of ≥25 and/or multiorgan failure.

[b]Hydrocortisone (50-mg IV bolus q6h) with fludrocortisone (50-μg tablet daily for 7 days) may improve outcomes of severe sepsis, particularly in the setting of relative adrenal insufficiency.

[c]Tetracyclines can be antagonistic in action to β-lactam agents. Adjust treatment as soon as the diagnosis is confirmed.

[d]The optimal dose of IV immunoglobulin has not been determined, but the median dose in observational studies is 2 g/kg (total dose administered over 1–5 days).

[e]Community-acquired methicillin-resistant *S. aureus*.

[f]In the United States, artesunate must be obtained by the Centers for Disease Control and Prevention. For patients diagnosed with severe malaria, full doses of parenteral antimalarial treatment should be started with whichever recommended antimalarial agent is first available.

[g]*Haemophilus aphrophilus, H. paraphrophilus, H. parainfluenzae, Aggregatibacter* (formerly *Actinobacillus*) *actinomycetemcomitans, Cardiobacterium hominis, Eikenella corrodens,* and *Kingella kingae*.

Adjunctive treatments may reduce morbidity and mortality rates and include dexamethasone for bacterial meningitis; IV immunoglobulin for TSS and necrotizing fasciitis caused by group A *Streptococcus*; low-dose hydrocortisone and fludrocortisone for septic shock; and drotrecogin alfa (activated), also known as recombinant human activated protein C, for meningococcemia and severe sepsis. Adjunctive therapies should usually be initiated within the first hours of treatment; however, dexamethasone for bacterial meningitis must be given before or at the time of the first dose of antibiotic.

SPECIFIC PRESENTATIONS

The infections considered below according to common clinical presentation can have rapidly catastrophic outcomes, and their immediate recognition and treatment can be life-saving. Recommended empirical therapeutic regimens are presented in Table 121-1.

Sepsis Without an Obvious Focus of Primary Infection

These patients initially have a brief prodrome of nonspecific symptoms and signs that progresses quickly to hemodynamic instability with hypotension, tachycardia, tachypnea, respiratory distress, and altered mental status. Disseminated intravascular coagulation (DIC) with clinical evidence of a hemorrhagic diathesis is a poor prognostic sign.

SEPTIC SHOCK

(See also Chap. 271) Patients with bacteremia leading to septic shock may have a primary site of infection (e.g., pneumonia, pyelonephritis, or cholangitis) that is not evident initially. Elderly patients with comorbid conditions, hosts compromised by malignancy and neutropenia, and patients who have recently undergone a surgical procedure or hospitalization are at increased risk for an adverse outcome. Gram-negative bacteremia with organisms such as *Pseudomonas aeruginosa* or *Escherichia coli* and gram-positive infection with organisms such as *Staphylococcus aureus,* including methicillin-resistant *S. aureus* (MRSA), or group A streptococci can present as intractable hypotension and

multiorgan failure. Treatment can usually be initiated empirically on the basis of the presentation, host factors (Table 271-3), and local patterns of bacterial resistance. Worse outcomes are evident when antimicrobial treatment is delayed or the pathogenic etiology ultimately proves to be nonsusceptible to the initial regimen. Broad-spectrum antimicrobial agents are therefore recommended. Adjunctive therapy with either drotrecogin alfa (activated) or glucocorticoids should be considered for patients with severe sepsis.

OVERWHELMING INFECTION IN ASPLENIC PATIENTS

(See also Chap. 271) Patients without splenic function are at risk for overwhelming bacterial sepsis. Asplenic adult patients succumb to sepsis at 58 times the rate of the general population. Most infections are thought to occur within the first 2 years after splenectomy, with a mortality rate of ~50%, but the increased risk persists throughout life. In asplenia, encapsulated bacteria cause the majority of infections. Adults, who are more likely to have antibody to these organisms, are at lower risk than children. *Streptococcus pneumoniae* is the most common isolate, causing 50–70% of cases, but the risk of infection with *Haemophilus influenzae* or *Neisseria meningitidis* is also high. Severe clinical manifestations of infections due to *E. coli*, *S. aureus*, group B streptococci, *P. aeruginosa*, *Capnocytophaga*, *Bordetella holmesii*, *Babesia*, and *Plasmodium* have been described.

BABESIOSIS

(See also Chap. 211) A history of recent travel to endemic areas raises the possibility of infection with *Babesia*. Between 1 and 4 weeks after a tick bite, the patient experiences chills, fatigue, anorexia, myalgia, arthralgia, shortness of breath, nausea, and headache; ecchymosis and/or petechiae are occasionally seen. The tick that most commonly transmits *Babesia*, *Ixodes scapularis*, also transmits *Borrelia burgdorferi* (the agent of Lyme disease) and *Anaplasma*; co-infection can occur, resulting in more severe disease. Infection with the European species *Babesia divergens* is more frequently fulminant than that due to the U.S. species *Babesia microti*. *B. divergens* causes a febrile syndrome with hemolysis, jaundice, hemoglobinemia, and renal failure and is associated with a mortality rate of >50%. Severe babesiosis is especially common in asplenic hosts but does occur in hosts with normal splenic function, particularly those who are >60 years of age and those with underlying immunosuppressive conditions such as HIV infection or malignancy. Complications include renal failure, acute respiratory failure, and DIC.

OTHER SEPSIS SYNDROMES

Tularemia (Chap. 158) is seen throughout the United States but occurs primarily in Arkansas, Missouri, South Dakota, and Oklahoma. This disease is associated with wild rabbit, tick, and tabanid fly contact. The typhoidal form can be associated with gram-negative septic shock and a mortality rate of >30%, especially in patients with underlying comorbid or immunosuppressive conditions. Plague occurs infrequently in the United States (Chap. 159), primarily after contact with ground squirrels, prairie dogs, or chipmunks, but is endemic in other parts of the world, with >90% of all cases occurring in Africa. The septic form is particularly rare and is associated with shock, multiorgan failure, and a 30% mortality rate. These infections should be considered in the appropriate epidemiologic setting. The Centers for Disease Control and Prevention lists *Francisella tularensis* and *Yersinia pestis* (the agents of tularemia and plague, respectively) along with *Bacillus anthracis* (the agent of anthrax) as important organisms that might be used for bioterrorism (Chap. 221).

SEPSIS WITH SKIN MANIFESTATIONS

(See also Chap. 17) Maculopapular rashes may reflect early meningococcal or rickettsial disease but are usually associated with nonemergent infections. Exanthems are usually viral. Primary HIV infection commonly presents with a rash that is typically maculopapular and involves the upper part of the body but can spread to the palms and soles. The patient is usually febrile and can have lymphadenopathy, severe headache, dysphagia, diarrhea, myalgias, and arthralgias. Recognition of this syndrome provides an opportunity to prevent transmission and to institute treatment and monitoring early on.

Petechial rashes caused by viruses are seldom associated with hypotension or a toxic appearance, although there can be exceptions (e.g., severe measles or arboviral infection). In other settings, petechial rashes require more urgent attention.

MENINGOCOCCEMIA

(See also Chap. 143) Almost three-quarters of patients with bacteremic *N. meningitidis* infection have a rash. Meningococcemia most often affects young children (i.e., those 6 months to 5 years old). In sub-Saharan Africa, the high prevalence of serogroup A meningococcal disease has been a threat to public health for more than a century. Thousands of deaths occur annually in this area, which is known as the "meningitis belt," and large epidemic waves occur approximately every 8–12 years. In the United States, sporadic cases and outbreaks occur in day-care centers, schools (grade school through college), and army barracks. Household contacts of index cases are at 400–800 times greater risk of disease than the general population. Patients may exhibit fever, headache, nausea, vomiting, myalgias, changes in mental status, and meningismus. However, the rapidly progressive form of disease is not usually associated with meningitis. The rash is initially pink, blanching, and

maculopapular, appearing on the trunk and extremities, but then becomes hemorrhagic, forming petechiae. Petechiae are first seen at the ankles, wrists, axillae, mucosal surfaces, and palpebral and bulbar conjunctiva, with subsequent spread to the lower extremities and trunk. A cluster of petechiae may be seen at pressure points—e.g., where a blood pressure cuff has been inflated. In rapidly progressive meningococcemia (10–20% of cases), the petechial rash quickly becomes purpuric (see Fig. 51-5), and patients develop DIC, multiorgan failure, and shock. Of these patients, 50–60% die, and survivors often require extensive debridement or amputation of gangrenous extremities. Hypotension with petechiae for <12 h is associated with significant mortality. Cyanosis, coma, oliguria, metabolic acidosis, and elevated partial thromboplastin time are also associated with a fatal outcome. Correction of protein C deficiency may improve outcome. Antibiotics given in the office by the primary care provider before hospital evaluation and admission may improve prognosis; this observation suggests that early initiation of treatment may be life-saving.

ROCKY MOUNTAIN SPOTTED FEVER

(See also Chap. 174) RMSF is a tickborne disease caused by *Rickettsia rickettsii* that occurs throughout North and South America. Up to 40% of patients do not report a history of a tick bite, but a history of travel or outdoor activity (e.g., camping in tick-infested areas) can often be ascertained. For the first 3 days, headache, fever, malaise, myalgias, nausea, vomiting, and anorexia are documented. By day 3, half of patients have skin findings. Blanching macules develop initially on the wrists and ankles and then spread over the legs and trunk. The lesions become hemorrhagic and are frequently petechial. The rash spreads to palms and soles later in the course. The centripetal spread is a classic feature of RMSF but occurs in a minority of patients. Moreover, 10–15% of patients with RMSF never develop a rash. The patient can be hypotensive and develop noncardiogenic pulmonary edema, confusion, lethargy, and encephalitis progressing to coma. The CSF contains 10–100 cells/μL, usually with a predominance of mononuclear cells. The CSF glucose level is often normal; the protein concentration may be slightly elevated. Renal and hepatic injury and bleeding secondary to vascular damage are noted. Untreated infection has a mortality rate of 20–30%.

Other rickettsial diseases cause significant morbidity and mortality worldwide. *Mediterranean spotted fever* caused by *Rickettsia conorii* is found in Africa, southwestern and south-central Asia, and southern Europe. Patients have fever, flu-like symptoms, and an inoculation eschar at the site of the tick bite. A maculopapular rash develops within 1–7 days, involving the palms and soles but sparing the face. Elderly patients or those with diabetes, alcoholism, uremia, or congestive heart failure are at risk for severe disease characterized by neurologic involvement, respiratory distress, and gangrene of the digits. Mortality rates associated with this severe form of disease approach 50%. *Epidemic typhus*, caused by *Rickettsia prowazekii*, is transmitted in louse-infested environments and emerges in conditions of extreme poverty, war, and natural disaster. Patients experience a sudden onset of high fevers, severe headache, cough, myalgias, and abdominal pain. A maculopapular rash develops (primarily on the trunk) in more than half of patients and can progress to petechiae and purpura. Serious signs include delirium, coma, seizures, noncardiogenic pulmonary edema, skin necrosis, and peripheral gangrene. Mortality rates approached 60% in the preantibiotic era and continue to exceed 10–15% in contemporary outbreaks. *Scrub typhus*, caused by *Orientia tsutsugamushi*—a separate genus in the family Rickettsiaceae—is transmitted by larval mites or chiggers and is one of the most common infections in southeastern Asia and the western Pacific. The organism is found in areas of heavy scrub vegetation (e.g., along riverbanks). Patients may have an inoculation eschar and may develop a maculopapular rash. Severe cases progress to pneumonia, meningoencephalitis, DIC, and renal failure. Mortality rates range from 1% to 35%.

If recognized in a timely fashion, rickettsial disease is very responsive to treatment. Doxycycline (100 mg twice daily for 3–14 days) is the treatment of choice for both adults and children. The newer macrolides and chloramphenicol may be suitable alternatives.

PURPURA FULMINANS

(See also Chaps. 143 and 271) Purpura fulminans is the cutaneous manifestation of DIC and presents as large ecchymotic areas and hemorrhagic bullae. Progression of petechiae to purpura, ecchymoses, and gangrene is associated with congestive heart failure, septic shock, acute renal failure, acidosis, hypoxia, hypotension, and death. Purpura fulminans has been associated primarily with *N. meningitidis* but, in splenectomized patients, may be associated with *S. pneumoniae* and *H. influenzae*. Several small studies have suggested that correction of the protein C deficiency evident in meningococcal purpura fulminans with drotrecogin alfa (activated) may dramatically improve outcome.

ECTHYMA GANGRENOSUM

Septic shock caused by *P. aeruginosa* or *Aeromonas hydrophila* can be associated with ecthyma gangrenosum (see Figs. 152-1 and e7-35): hemorrhagic vesicles surrounded by a rim of erythema with central necrosis and ulceration. These gram-negative bacteremias are most common among patients with neutropenia, extensive burns, and hypogammaglobulinemia.

OTHER EMERGENT INFECTIONS ASSOCIATED WITH RASH

Vibrio vulnificus and other noncholera *Vibrio* bacteremic infections (Chap. 156) can cause focal skin lesions and overwhelming sepsis in hosts with liver disease, diabetes, renal insufficiency, or other immunocompromising conditions. After ingestion of contaminated raw shellfish, there is a sudden onset of malaise, chills, fever, and hypotension. The patient develops bullous or hemorrhagic skin lesions, usually on the lower extremities, and 75% of patients have leg pain. The mortality rate can be as high as 50–60%, particularly when the patient presents with hypotension. Other infections, caused by agents such as *Aeromonas, Klebsiella*, and *E.coli*, can cause hemorrhagic bullae and death due to overwhelming sepsis in cirrhotic patients. *Capnocytophaga canimorsus* can cause septic shock in asplenic patients. Infection typically follows a dog bite. Patients present with fever, chills, myalgia, vomiting, diarrhea, dyspnea, confusion, and headache. Findings can include an exanthem or erythema multiforme (see Figs. 51-9 and e7-25), cyanotic mottling or peripheral cyanosis, petechiae, and ecchymosis. About 30% of patients with this fulminant form die of overwhelming sepsis and DIC, and survivors may require amputation because of gangrene.

ERYTHRODERMA

TSS (Chaps. 135 and 136) is usually associated with erythroderma. The patient presents with fever, malaise, myalgias, nausea, vomiting, diarrhea, and confusion. There is a sunburn-type rash that may be subtle and patchy but is usually diffuse and is found on the face, trunk, and extremities. Erythroderma, which desquamates after 1–2 weeks, is more common in *Staphylococcus*-associated than in *Streptococcus*-associated TSS. Hypotension develops rapidly—often within hours—after the onset of symptoms. Multiorgan failure is seen. Early renal failure may precede hypotension and distinguishes this syndrome from other septic shock syndromes. There may be no indication of a primary focal infection, although possible cutaneous or mucosal portals of entry for the organism can be ascertained when a careful history is taken. Colonization rather than overt infection of the vagina or a postoperative wound, for example, is typical with staphylococcal TSS, and the mucosal areas appear hyperemic but not infected. Streptococcal TSS is more often associated with skin or soft tissue infection (including necrotizing fasciitis), and patients are more likely to be bacteremic. The diagnosis of TSS is defined by the clinical criteria of fever, rash, hypotension, and multiorgan involvement. The mortality rate is 5% for menstruation-associated TSS, 10–15% for nonmenstrual TSS, and 30–70% for streptococcal TSS.

VIRAL HEMORRHAGIC FEVERS

Viral hemorrhagic fevers (Chaps. 196 and 197) are zoonotic illnesses caused by viruses that reside in either animal reservoirs or arthropod vectors. These diseases occur worldwide and are restricted to areas where the host species live. They are caused by four major groups of viruses: Arenaviridae (e.g., Lassa fever in Africa), Bunyaviridae (e.g., Rift Valley fever in Africa or hantavirus hemorrhagic fever with renal syndrome in Asia), Filoviridae (e.g., Ebola and Marburg virus infections in Africa), and Flaviviridae (e.g., yellow fever in Africa and South America and dengue in Asia, Africa, and the Americas). Lassa fever and Ebola and Marburg virus infections are also transmitted from person to person. The vectors for most viral fevers are found in rural areas; dengue and yellow fever are important exceptions. After a prodrome of fever, myalgias, and malaise, patients develop evidence of vascular damage, petechiae, and local hemorrhage. Shock, multifocal hemorrhaging, and neurologic signs (e.g., seizures or coma) predict a poor prognosis. Dengue (Chap. 196) is the most common arboviral disease worldwide. More than half a million cases of dengue hemorrhagic fever occur each year, with at least 12,000 deaths. Patients have a triad of symptoms: hemorrhagic manifestations, evidence of plasma leakage, and platelet counts <100,000/μL. Mortality rates are 10–20%. If dengue shock syndrome develops, mortality can reach 40%. Although supportive care to maintain blood pressure and intravascular volume with careful volume-replacement therapy is key, ribavirin also may be useful against Arenaviridae and Bunyaviridae.

SEPSIS WITH A SOFT TISSUE/MUSCLE PRIMARY FOCUS

See also Chap. 125.

NECROTIZING FASCIITIS

This infection is characterized by extensive necrosis of the subcutaneous tissue and fascia. It may arise at a site of minimal trauma or postoperative incision and may also be associated with recent varicella, childbirth, or muscle strain. The most common causes of necrotizing fasciitis are group A streptococci alone (Chap. 136) and a mixed facultative and anaerobic flora (Chap. 125). Diabetes mellitus, peripheral vascular disease, and IV drug use are associated risk factors. Physical findings are minimal compared with the severity of pain and the degree of fever. The examination is often unremarkable except for soft tissue edema and erythema. The infected area is red, hot, shiny, swollen, and exquisitely tender. In untreated infection, the overlying skin develops blue-gray patches after 36 h, and cutaneous bullae and necrosis develop after 3–5 days. Necrotizing fasciitis due to a mixed flora, but not that due to group A streptococci, can be associated with gas production. Without treatment, pain decreases because of thrombosis of the small blood vessels and destruction of the peripheral nerves—an ominous sign. The mortality rate is 15–34% overall, >70% in association with TSS, and nearly 100% without surgical intervention. Necrotizing fasciitis may also be due to *Clostridium perfringens* (Chap. 142); in this condition, the patient is extremely toxic and the mortality rate is high.

Within 48 h, rapid tissue invasion and systemic toxicity associated with hemolysis and death ensue. The distinction between this entity and clostridial myonecrosis is made by muscle biopsy. Necrotizing fasciitis caused by community-acquired MRSA has been reported. The MRSA-infected patients in one series required extensive surgical debridement, but there were no deaths.

CLOSTRIDIAL MYONECROSIS

(See also Chap. 142) Myonecrosis is often associated with trauma or surgery but can be spontaneous. The incubation period is usually 12–24 h long, and massive necrotizing gangrene develops within hours of onset. Systemic toxicity, shock, and death can occur within 12 h. The patient's pain and toxic appearance are out of proportion to physical findings. On examination, the patient is febrile, apathetic, tachycardic, and tachypneic and may express a feeling of impending doom. Hypotension and renal failure develop later, and hyperalertness is evident preterminally. The skin over the affected area is bronze-brown, mottled, and edematous. Bullous lesions with serosanguineous drainage and a mousy or sweet odor can develop. Crepitus can occur secondary to gas production in muscle tissue. The mortality rate is >65% with spontaneous myonecrosis, which is often associated with *Clostridium septicum* and underlying malignancy. The mortality rates associated with trunk and limb infection are 63% and 12%, respectively, and any delay in surgical treatment increases the risk of death.

Neurologic Infections with or Without Septic Shock

BACTERIAL MENINGITIS

(See also Chap. 381) Bacterial meningitis is one of the most common infectious disease emergencies involving the central nervous system. Although hosts with cell-mediated immune deficiency (including transplant recipients, diabetic patients, elderly patients, and cancer patients receiving certain chemotherapeutic agents) are at particular risk for *Listeria monocytogenes* meningitis, most cases in adults are due to *S. pneumoniae* (30–50%) and *N. meningitidis* (10–35%). The classic presentation of fever, meningismus, and altered mental status is seen in only one-half to two-thirds of patients. The elderly can present without fever or meningeal signs despite lethargy and confusion. Cerebral dysfunction is evidenced by confusion, delirium, and lethargy that can progress to coma. A fulminant presentation with sepsis and brain edema occurs in some cases; papilledema at presentation is unusual and suggests another diagnosis (e.g., an intracranial lesion). Focal signs, including cranial nerve palsies (IV, VI, VII), can be seen in 10–20% of cases; 50–70% of patients have bacteremia. A poor outcome is associated with coma, hypotension, meningitis due to *S. pneumoniae*, respiratory distress, a CSF glucose level of <0.6 mmol/L (<10 mg/dL), a CSF protein level of >2.5 g/L, a peripheral white blood cell count of <5000/μL, and a serum sodium level of <135 mmol/L. Rapid initiation of treatment is essential; the odds of an unfavorable outcome may increase by 30% for each hour that treatment is delayed.

SUPPURATIVE INTRACRANIAL INFECTIONS

(See also Chap. 381) In suppurative intracranial infections, rare intracranial lesions present along with sepsis and hemodynamic instability. Rapid recognition of the toxic patient with central neurologic signs is crucial to improvement of the dismal prognosis of these entities. *Subdural empyema* arises from the paranasal sinus in 60–70% of cases. Microaerophilic streptococci and staphylococci are the predominant etiologic organisms. The patient is toxic, with fever, headache, and nuchal rigidity. Of all patients, 75% have focal signs and 6–20% die. Despite improved survival rates, 15–44% of patients are left with permanent neurologic deficits. *Septic cavernous sinus thrombosis* follows a facial or sphenoid sinus infection; 70% of cases are due to staphylococci (including MRSA), and the remainder are due primarily to aerobic or anaerobic streptococci. A unilateral or retroorbital headache progresses to a toxic appearance and fever within days. Three-quarters of patients have unilateral periorbital edema that becomes bilateral and then progresses to ptosis, proptosis, ophthalmoplegia, and papilledema. The mortality rate is as high as 30%. *Septic thrombosis of the superior sagittal sinus* spreads from the ethmoid or maxillary sinuses and is caused by *S. pneumoniae*, other streptococci, and staphylococci. The fulminant course is characterized by headache, nausea, vomiting, rapid progression to confusion and coma, nuchal rigidity, and brainstem signs. If the sinus is totally thrombosed, the mortality rate exceeds 80%.

BRAIN ABSCESS

(See also Chap. 381) Brain abscess often occurs without systemic signs. Almost half of patients are afebrile, and presentations are more consistent with a space-occupying lesion in the brain; 70% of patients have headache and/or altered mental status, 50% have focal neurologic signs, and 25% have papilledema. Abscesses can present as single or multiple lesions resulting from contiguous foci or hematogenous infection, such as endocarditis. The infection progresses over several days from cerebritis to an abscess with a mature capsule. More than half of infections are polymicrobial, with an etiology consisting of aerobic bacteria (primarily streptococcal species) and anaerobes. Abscesses arising hematogenously are especially apt to rupture into the ventricular space, causing a sudden and severe deterioration in clinical status and high mortality. Otherwise, mortality is low but morbidity is high (30–55%). Patients presenting with stroke and a parameningeal infectious focus, such as sinusitis or otitis, may have a brain abscess, and physicians must maintain a high level of suspicion. Prognosis

worsens in patients with a fulminant course, delayed diagnosis, abscess rupture into the ventricles, multiple abscesses, or abnormal neurologic status at presentation.

INTRACRANIAL AND SPINAL EPIDURAL ABSCESSES

(See also Chap. 377) Spinal and intracranial epidural abscesses (SEAs and ICEAs) can result in permanent neurologic deficits, sepsis, and death. At-risk patients include those with diabetes mellitus; IV drug use; chronic alcohol abuse; recent spinal trauma, surgery, or epidural anesthesia; and other comorbid conditions, such as HIV infection. In the United States and Canada, where early treatment of otitis and sinusitis is typical, ICEA is rare but the number of cases of SEA is on the rise. In Africa and areas with limited access to health care, SEAs and ICEAs cause significant morbidity and mortality. ICEAs typically present as fever, mental status changes, and neck pain, while SEAs often present as fever, localized spinal tenderness, and back pain. ICEAs are typically polymicrobial, whereas SEAs are most often due to hematogenous seeding, with staphylococci the most common etiologic agent. Early diagnosis and treatment, which may include surgical drainage, minimize rates of mortality and permanent neurologic sequelae.

CEREBRAL MALARIA

(See also Chap. 210) This entity should be urgently considered if patients who have recently traveled to areas endemic for malaria present with a febrile illness and lethargy or other neurologic signs. Fulminant malaria is caused by *Plasmodium falciparum* and is associated with temperatures of >40°C (>104°F), hypotension, jaundice, adult respiratory distress syndrome, and bleeding. By definition, any patient with a change in mental status or repeated seizure in the setting of fulminant malaria has cerebral malaria. In adults, this nonspecific febrile illness progresses to coma over several days; occasionally, coma occurs within hours and death within 24 h. Nuchal rigidity and photophobia are rare. On physical examination, symmetric encephalopathy is typical, and upper motor neuron dysfunction with decorticate and decerebrate posturing can be seen in advanced disease. Unrecognized infection results in a 20–30% mortality rate.

OTHER FOCAL SYNDROMES WITH A FULMINANT COURSE

Infection at virtually any primary focus (e.g., osteomyelitis, pneumonia, pyelonephritis, or cholangitis) can result in bacteremia and sepsis. TSS has been associated with focal infections such as septic arthritis, peritonitis, sinusitis, and wound infection. Rapid clinical deterioration and death can be associated with destruction of the primary site of infection, as is seen in endocarditis and in necrotizing infections of the oropharynx (in which edema suddenly compromises the airway).

RHINOCEREBRAL MUCORMYCOSIS

(See also Chap. 205) Patients with diabetes or immunocompromising conditions are at risk for invasive rhinocerebral mucormycosis. Patients present with low-grade fever, dull sinus pain, diplopia, decreased mental status, decreased ocular motion, chemosis, proptosis, dusky or necrotic nasal turbinates, and necrotic hard-palate lesions that respect the midline. Without rapid recognition and intervention, the process continues on an inexorable invasive course, with high mortality.

ACUTE BACTERIAL ENDOCARDITIS

(See also Chap. 124) This entity presents with a much more aggressive course than subacuteendocarditis. Bacteria such as *S. aureus*,*S. pneumoniae*, *L. monocytogenes*, *Haemophilus* spp., and streptococci of groups A, B, and G attack native valves. Native-valve endocarditis caused by *S. aureus*, including MRSA, is increasing, particularly in health care settings. Mortality rates range from 10% to 40%. The host may have comorbid conditions such as underlying malignancy, diabetes mellitus, IV drug use, or alcoholism. The patient presents with fever, fatigue, and malaise <2 weeks after onset of infection. On physical examination, a changing murmur and congestive heart failure may be noted. Hemorrhagic macules on palms or soles (*Janeway lesions*) sometimes develop. Petechiae, Roth's spots, splinter hemorrhages, and splenomegaly are unusual. Rapid valvular destruction, particularly of the aortic valve, results in pulmonary edema and hypotension. Myocardial abscesses can form, eroding through the septum or into the conduction system and causing life-threatening arrhythmias or high-degree conduction block. Large friable vegetations can result in major arterial emboli, metastatic infection, or tissue infarction. Older patients with *S. aureus* endocarditis are especially likely to present with nonspecific symptoms—a circumstance that delays diagnosis and worsens prognosis. Rapid intervention is crucial for a successful outcome.

INHALATIONAL ANTHRAX

(See also Chap. 221) Inhalational anthrax, the most severe form of disease caused by *B. anthracis*, had not been reported in the United States for more than 25 years until the use of this organism as an agent of bioterrorism in 2001. Patients presented with malaise, fever, cough, nausea, drenching sweats, shortness of breath, and headache. Rhinorrhea was unusual. All patients had abnormal chest roentgenograms at presentation. Pulmonary infiltrates, mediastinal widening, and pleural effusions were the most common findings. Hemorrhagic meningitis was seen in 38% of these patients. Survival was more likely when antibiotics were given during the prodromal period and when multidrug regimens were used. In the absence of urgent intervention with antimicrobial agents and supportive care,

inhalational anthrax progresses rapidly to hypotension, cyanosis, and death.

AVIAN INFLUENZA (H5N1) INFECTION

(See also Chap. 187) Human cases of avian influenza have occurred primarily in Southeast Asia, particularly Vietnam. Avian influenza should be considered in patients with severe respiratory tract illness, particularly if they have been exposed to poultry. Patients present with high fever, an influenza-like illness, and lower respiratory tract symptoms; this illness can progress rapidly to bilateral pneumonia, acute respiratory distress syndrome, multiorgan failure, and death. Early antiviral treatment with neuraminidase inhibitors should be initiated along with aggressive supportive measures. Unlike avian influenza, for which human-to-human transmission has been rare so far, a novel swine-associated influenza A/H1N1 virus has spread rapidly throughout the world. Early in the pandemic, there has been a sudden increase in severe pneumonia affecting a younger population. Patients most at risk of severe disease are children <5 years of age, elderly persons, patients with underlying chronic conditions, and pregnant women.

HANTAVIRUS PULMONARY SYNDROME

(See also Chap. 196) Hantavirus pulmonary syndrome has been documented in the United States (primarily the southwestern states), Canada, and South America. Most cases occur in rural areas and are associated with exposure to rodents. Patients present with a nonspecific viral prodrome of fever, malaise, myalgias, nausea, vomiting, and dizziness that may progress to pulmonary edema and respiratory failure. Hantavirus pulmonary syndrome causes myocardial depression and increased pulmonary vascular permeability; therefore, careful fluid resuscitation and use of pressor agents are crucial. Aggressive cardiopulmonary support during the first few hours of illness can be life-saving. The early onset of thrombocytopenia may help distinguish this syndrome from other febrile illnesses in an appropriate epidemiologic setting.

CONCLUSION

Acutely ill febrile patients with the syndromes discussed in this chapter require close observation, aggressive supportive measures, and—in most cases—admission to intensive care units. The most important task of the physician is to distinguish these patients from other infected febrile patients whose illness will not progress to fulminant disease. The alert physician must recognize the acute infectious disease emergency and then proceed with appropriate urgency.

FURTHER READINGS

Dantas-Torres F: Rocky Mountain spotted fever. Lancet Infect Dis 7:724, 2007[PMID: 17961858]

Fitch MT, van de Beek D: Emergency diagnosis and treatment of adult meningitis. Lancet Infect Dis 7:191, 2007[PMID: 17317600]

Hasham S et al: Necrotising fasciitis. BMJ 330:830, 2005[PMID: 15817551]

Hayden FG et al: Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. N Engl J Med 362:1708, 2010[PMID: 20445182]

Lappin E, Ferguson AJ: Gram-positive toxic shock syndromes. Lancet Infect Dis 9:281, 2009[PMID: 19393958]

Nguyen HB et al: Severe sepsis and septic shock: Review of the literature and emergency department management guidelines. Ann Emerg Med 48:28, 2006[PMID: 16781920]

Pradilla G et al: Epidural abscess of the CNS. Lancet Neurol 8:292, 2009[PMID: 19233039]

Spelman D et al: Guidelines for the prevention of sepsis in asplenic and hyposplenic patients. Intern Med J 38:349, 2008[PMID: 18284463]

Stephens DS et al: Epidemic meningitis, meningococcemia, and *Neisseria meningitidis*. Lancet 369:2196, 2007[PMID: 17604802]

Talan DA et al (eds): Infectious Disease Emergencies. Infect Dis Clin North Am 22:1–187, 2008 [entire volume]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

IMMUNIZATION PRINCIPLES AND VACCINE USE: INTRODUCITON

Few medical interventions of the past century can rival the effect that immunization has had on longevity, economic savings, and quality of life. Seventeen diseases are now preventable through vaccines routinely administered to children and adults in the United States **(Table 122-1)**, and most vaccine-preventable diseases of childhood are at historically low levels **(Table 122-2)**. Health care providers deliver the vast majority of vaccines in the United States in the course of providing routine health services and therefore play an integral role in the nation's public health system.

Table 122–1. Diseases that Are Now Preventable with Vaccines Routinely Administered in the United States to Children and/or Adults

Condition	Target Population(s) for Routine Use
Pertussis	Children, adolescents, adults
Diphtheria	Children, adolescents, adults
Tetanus	Children, adolescents, adults
Poliomyelitis	Children
Measles	Children
Mumps	Children
Rubella, congenital rubella syndrome	Children
Hepatitis B	Children
Haemophilus influenzae type b infection	Children
Hepatitis A	Children
Influenza	Children, adolescents, adults
Varicella	Children
Invasive pneumococcal disease	Children, older adults
Meningococcal disease	Children, adolescents
Rotavirus infection	Infants
Human papillomavirus infection, cervical cancer	Adolescent girls and women
Zoster	Older adults

Vaccine Impact

DIRECT AND INDIRECT EFFECTS

Immunizations against specific infectious diseases protect individuals against infection and thereby prevent symptomatic illnesses. Specific vaccines may blunt the severity of clinical illness (e.g., rotavirus vaccines and severe gastroenteritis) or reduce complications (e.g., zoster vaccines and postherpetic neuralgia). Some immunizations also reduce transmission of infectious disease agents from immunized people to others, thereby reducing the impact of infection spread. This indirect impact is known as *herd immunity*. The level of immunization in a population that is required to achieve indirect protection of unimmunized people varies substantially with the specific vaccine.

Since childhood vaccines have become widely available in the United States, major declines in rates of vaccine-preventable diseases among both children and adults have become evident (Table 122-2). For example, vaccination of children <5 years of age against seven types of *Streptococcus pneumoniae* led to a >90% overall reduction in invasive disease caused by those types. A series of vaccines targeting 10 vaccine-preventable childhood diseases in a single birth cohort leads to prevention of 33,000 premature deaths and 14 million illnesses and saves $42 billion (U.S.): $9 billion in direct medical savings and $33 billion in indirect societal savings.

Table 122–2. Decline in Vaccine-Preventable Diseases in the United States Following Widespread Implementation of National Vaccine Recommendations

Condition	Annual No. of Prevaccine Cases (Average)	No. of Cases Reported in 2010[a]	Reduction (%) in Cases After Widespread Vaccination
Smallpox	29,005	0	100
Diphtheria	21,053	0	100
Measles	530,217	61	≥99
Mumps	162,344	2,528	98
Pertussis	200,752	21,291	89
Polio (paralytic)	16,316	0	100
Rubella	47,745	6	>99
Congenital rubella syndrome	152	0	100
Tetanus	580	8	99
Haemophilus influenzae type b infection	20,000	270[b]	99
Hepatitis A	117,333	11,049	91
Hepatitis B (acute)	66,232	11,269	83
Invasive pneumococcal infection: all ages	63,067	44,000[c]	30
Invasive pneumococcal infection: <5 years of age	16,069	4,167[c]	74
Varicella	4,085,120	449,363	89.0

[a] Except for cases of hepatitis A, hepatitis B, and pneumococcal infection, for which 2008 figures are shown.

[b] Includes 16 type b infections and 254 infections caused by unknown types (<5 years of age).

[c] Data are from the CDC's Active Bacterial Core Surveillance Report; *www.cdc.gov/abcs/survreports/spneu08.pdf.*

Source: Adapted from Roush et al., with permission.

CONTROL, ELIMINATION, AND ERADICATION OF VACCINE-PREVENTABLE DISEASES

Immunization programs are associated with the goals of controlling, eliminating, or eradicating a disease. *Control* of a vaccine-preventable disease reduces illness outcomes and often limits the disruptive impacts associated with outbreaks of disease in communities, schools, and institutions. Control programs can also reduce absences from work for ill persons and for parents caring for sick children, decrease absences from school, and limit health care utilization associated with treatment visits.

Elimination of a disease is a more demanding goal than control, usually requiring the reduction to zero of cases in a defined geographic area but sometimes defined as reduction in the indigenous sustained transmission of an infection in a geographic area. As of 2010, the United States had eliminated indigenous transmission of measles, rubella, poliomyelitis, and diphtheria. Importation of pathogens from other parts of the world continues to be important, and public health efforts are intended to react promptly to such cases and to limit forward spread of the infectious agent.

Eradication of a disease is achieved when its elimination can be sustained without ongoing interventions. The only vaccine-preventable disease that has been globally eradicated thus far is smallpox. Although smallpox vaccine is no longer given routinely, the disease has not naturally reemerged because all chains of human transmission were interrupted through earlier vaccination efforts and humans were the only natural reservoir of the virus. Currently, a major health initiative is targeting the global eradication of polio. Sustained transmission of polio has been eliminated from most nations but has never been interrupted in four countries: Afghanistan, India, Nigeria, and Pakistan. Detection of a case of disease that has been targeted for eradication or elimination is considered a sentinel event that could permit the infectious agent to become reestablished in the community or region. Hence, such episodes must be promptly reported to public health authorities.

OUTBREAK DETECTION AND CONTROL

Clusters of cases of a vaccine-preventable disease detected in an institution, a medical practice, or a community may signal important changes in the pathogen, vaccine, or environment. Several factors can give rise to increases in vaccine-preventable disease, including (1) low rates of immunization that result in an accumulation of susceptible people (e.g., measles resurgence among vaccination abstainers); (2) changes in the infectious agent that permit it to escape vaccine-induced protection (e.g., nonvaccine-type pneumococci); (3) waning of vaccine-induced immunity (e.g., pertussis among

adolescents and adults vaccinated in early childhood); and (4) point-source introductions of large inocula (e.g., food-borne exposure to hepatitis A virus). Reporting episodes of outbreak-prone diseases to public health authorities can facilitate recognition of clusters that require further interventions.

Public Health Reporting

Recognition of suspected cases of diseases targeted for elimination or eradication—along with other diseases that require urgent public health interventions, such as contact tracing, administration of chemo- or immunoprophylaxis, or epidemiologic investigation for common-source exposure)—is typically associated with special reporting requirements. Many diseases against which vaccines are routinely used, including measles, pertussis, *Haemophilus influenzae* invasive disease, and varicella, are nationally notifiable. Clinicians and laboratory staff have a responsibility to report some vaccine-preventable disease occurrences to local or state public health authorities according to specific case-definition criteria. All providers should be aware of state or city disease-reporting requirements and the best ways to contact public health authorities. A prompt response to vaccine-preventable disease outbreaks can greatly enhance the effectiveness of control measures.

Global Considerations

Several international health initiatives currently focus on reducing vaccine-preventable diseases in regions throughout the world. These efforts include improving access to new and underutilized vaccines, such as pneumococcal conjugate, rotavirus, human papillomavirus (HPV), and meningococcal A conjugate vaccines. The American Red Cross, the World Health Organization (WHO), the United Nations Foundation, the United Nations Children's Fund (UNICEF), and the Centers for Disease Control and Prevention (CDC) are partners in the Measles Initiative, which targeted reduction of worldwide measles deaths by 90% from 2000 to 2010. During 2000–2008, global measles mortality rates declined by 78%—i.e., from an estimated 733,000 deaths in 2000 to 164,000 deaths in 2008. Rotary International, UNICEF, the CDC, and the WHO are leading partners in the global eradication of polio, an endeavor that reduced the annual number of paralytic polio cases from 350,000 in 1988 to <2000 in 2009. The GAVI Alliance and the Bill and Melinda Gates Foundation have brought substantial momentum to global efforts to reduce vaccine-preventable diseases, expanding on earlier efforts by the WHO, UNICEF, and governments in developed and developing countries.

Enhancing Immunization in Adults

Although immunization has become a centerpiece of routine pediatric medical visits, it has not been as well integrated into routine health care visits for adults. This chapter focuses on immunization principles and vaccine use in adults.Accumulating evidence suggests that immunization coverage can be increased through efforts directed at consumer-, provider-, institution-, and system-level factors. The literature suggests that the application of multiple strategies is more effective at raising coverage rates than is the use of any single strategy.

Recommendations for Adult Immunizations

The CDC's Advisory Committee on Immunization Practices (ACIP) is the main source of recommendations for use of vaccines licensed by the U.S. Food and Drug Administration (FDA) for children and adults in the U.S. civilian population. The ACIP is a federal advisory committee that consists of 15 voting members (experts in fields associated with immunization) appointed by the Secretary of the U.S. Department of Health and Human Services; 8 ex officio members representing federal agencies; and 26 nonvoting representatives of various liaison organizations, including major medical societies and managed-care organizations. The ACIP recommendations are available at *www.cdc.gov/vaccines/pubs/ACIP-list.htm*. These recommendations are harmonized to the greatest extent possible with vaccine recommendations made by other organizations, including the American College of Obstetricians and Gynecologists, the American Academy of Family Physicians, and the American College of Physicians.

Adult Immunization Schedules

Immunization schedules for adults in the United States are updated annually and can be found online (*www.cdc.gov/vaccines/recs/schedules/adult-schedule.htm*). In January, the schedules are published in *American Family Physician*, the *Annals of Internal Medicine* , and *Morbidity and Mortality Weekly Report* (*www.cdc.gov/mmwr*). The adult immunization schedules for 2011 are summarized in Fig. 122-1. Additional information and specifications are contained in the footnotes to these schedules. In the time between annual publications, additions and changes to schedules are published as Notices to Readers in *Morbidity and Mortality Weekly Report* .

Figure 122-1

Recommended adult immunization schedule, by vaccine and age group

VACCINE ▼ / AGE GROUP ▶	19–26 years	27–49 years	50–59 years	60–64 years	≥65 years
Influenza[1,*]	1 dose annually				
Tetanus, diphtheria, pertussis (Td/Tdap)[2,*]	Substitute 1-time dose of Tdap for Td booster; then boost with Td every 10 yrs				Td booster every 10 yrs
Varicella[3,*]	2 doses				
Human papillomavirus (PHV)[4,*]	3 doses (females)				
Zoster[5]				1 dose	
Measles, mumps, rubella (MMR)[6,*]	1 or 2 doses		1 dose		
Pneumococcal (polysaccharide)[7,8]	1 or 2 doses				1 dose
Meningococcal[9,*]	1 or more doses				
Hepatitis A[10,*]	2 doses				
Hepatitis B[11,*]	3 doses				

*Covered by the Vaccine Injury Compensation Program.

For all persons in this category who meet the age requirements and who lack evidence of immunity (e.g., lack documentation of vaccination or have no evidence of previous infection)

Recommended if some other risk factor is present (e.g., based on medical, occupational, lifestyle, or other indications)

No recommendation

Report all clinically significant postvaccination reactions to the Vaccine Adverse Event Reporting System (VAERS). Reporting forms and instructions on filing a VAERS report are available at http://www.vaers.hhs.gov or by telephone, 800-822-7967.

Information on how to file a Vaccine Injury Compensation Program claim is available at http://www.hrsa.gov/vaccinecompensation or by telephone, 800-338-2382. Information about filing a claim for vaccine injury is available through the U.S. Court of Federal Claims, 717 Madison Place, N.W., Washington, D.C. 20005; telephone, 202-357-6400.

Additional information about the vaccines in this schedule, extent of available data, and contraindications for vaccination also is available at http://www.cdc.gov/vaccines or from the CDC-INFO Contact Center at 800-CDC- INFO (800-232-4636) in English and Spanish, 24 hours a day, 7 days a week.

Vaccines that might be indicated for adults based on medical and other indications

INDICATION ▶ / VACCINE ▼	Pregnancy	Immunocompromising conditions (excluding human immunodeficiency virus [HIV])[3,5,6,13]	HIV infection[3,6,12,13] CD4+ T lymphocyte count <200 cells/μL	HIV infection[3,6,12,13] CD4+ T lymphocyte count ≥200 cells/μL	Diabetes, heart disease, chronic lung disease, chronic alcoholism	Asplenia[12] (including elective splenectomy) and persistent complement component deficiencies	Chronic liver disease	Kidney failure, end-stage renal disease, receipt of hemodialysis	Healthcare personnel
Influenza[1,*]	1 dose TIV annually								1 dose TIV or LAIV annually
Tetanus, diphtheria, pertussis (Td/Tdap)[2,*]	Td	Substitute 1-time dose of Tdap for Td booster; then boost with Td every 10 yrs							
Varicella[3,*]	Contraindicated			2 doses					
Human papillomavirus (HPV)[4,*]		3 doses through age 26 yrs							
Zoster[5]	Contraindicated			1 dose					
Measles, mumps, rubella (MMR)[6,*]	Contraindicated			1 or 2 doses					
Pneumococcal (polysaccharide)[7,8]		1 or 2 doses							
Meningococcal[9,*]	1 or more doses								
Hepatitis A[10,*]	2 doses								
Hepatitis B[11,*]				3 doses					

*Covered by the Vaccine Injury Compensation Program.

For all persons in this category who meet the age requirements and who lack evidence of immunity (e.g., lack documentation of vaccination or have no evidence of previous infection)

Recommended if some other risk factor is present (e.g., on the basis of medical, occupational, lifestyle, or other indications)

No recommendation

These schedules indicate the recommended age groups and medical indications for which administration of currently licensed vaccines is commonly indicated for adults ages 19 years and older, as of February 4, 2011. For all vaccines being recommended on the adult immunization schedule, a vaccine series does not need to be restarted, regardless of the time that has elapsed between doses. Licensed combination vaccines may be used whenever any components of the combination are indicated and when the vaccine's other components are not contraindicated. For detailed recommendations on all vaccines, including those used primarily for travelers or that are issued during the year, consult the manufacturers' package inserts and the complete statements from the Advisory Committee on Immunization Practices (http:// www.cdc.gov/vaccines/pubs/acip-list.htm).

The recommendations in this schedule were approved by the Centers for Disease Control and Prevention's (CDC) Advisory Committee on Immunization Practices (ACIP), the American Academy of Family Physicians (AAFP), the American College of Obstetricians and Gynecologists (ACOG), and the American College of Physicians (ACP).

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
CENTERS FOR DISEASE CONTROL AND PREVENTION

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Recommended adult immunization schedules, United States, 2011. For complete statements by the Advisory Committee on Immunization Practices (ACIP), visit*www.cdc.gov/vaccines/pubs/ACIP-list.htm*.

1. Influenza vaccination Annual vaccination against influenza is recommended for all persons aged 6 months and older, including all adults. Healthy, nonpregnant adults aged less than 50 years without high-risk medical conditions can receive either intranasally administered live, attenuated influenza vaccine (FluMist), or inactivated vaccine. Other persons should receive the inactivated vaccine. Adults aged 65 years and older can receive the standard influenza vaccine or the high-dose (Fluzone) influenza vaccine. Additional information about influenza vaccination is available at *http://www.cdc.gov/vaccines/vpd-vac/flu/default.htm.*

2. Tetanus, diphtheria, and acellular pertussis (Td/Tdap) vaccination Administer a one-time dose of Tdap to adults aged less than 65 years who have not received Tdap previously or for whom vaccine status is unknown to replace one of the 10-year Td boosters, and as soon as feasible to all 1) postpartum women, 2) close contacts of infants younger than age 12 months (e.g., grandparents and child-care providers), and 3) healthcare personnel with direct patient contact. Adults aged 65 years and older who have not previously received Tdap and who have close contact with an infant aged less than 12 months also should be vaccinated. Other adults aged 65 years and older may receive Tdap. Tdap can be administered regardless of interval since the most recent tetanus or diphtheria-containing vaccine.

Adults with uncertain or incomplete history of completing a 3-dose primary vaccination series with Td-containing vaccines should begin or complete a primary vaccination series. For unvaccinated adults, administer the first 2 doses at least 4 weeks apart and the third dose 6–12 months after the second. If incompletely vaccinated (i.e., less than 3 doses), administer remaining doses. Substitute a one-time dose of Tdap for one of the doses of Td, either in the primary series or for the routine booster, whichever comes first.

If a woman is pregnant and received the most recent Td vaccination 10 or more years previously, administer Td during the second or third trimester. If the woman received the most recent Td vaccination less than 10 years previously, administer Tdap during the immediate postpartum period. At the clinician's discretion, Td may be deferred during pregnancy and Tdap substituted in the immediate postpartum period, or Tdap may be administered instead of Td to a pregnant woman after an informed discussion with the woman.

The ACIP statement for recommendations for administering Td as prophylaxis in wound management is available at *http://www.cdc.gov/vaccines/pubs/acip-list.htm.*

3. Varicella vaccination All adults without evidence of immunity to varicella should receive 2 doses of single-antigen varicella vaccine if not previously vaccinated or a second dose if they have received only 1 dose, unless they have a medical contraindication. Special consideration should be given to those who 1) have close contact with persons at high risk for severe disease (e.g., healthcare personnel and family contacts of persons with immunocompromising conditions) or 2) are at high risk for exposure or transmission (e.g., teachers; child-care employees; residents and staff members of institutional settings, including correctional institutions; college students; military personnel; adolescents and adults living in households with children; nonpregnant women of childbearing age; and international travelers).

Evidence of immunity to varicella in adults includes any of the following: 1) documentation of 2 doses of varicella vaccine at least 4 weeks apart; 2) U.S.-born before 1980 (although for healthcare personnel and pregnant women, birth before 1980 should not be considered evidence of immunity); 3) history of varicella based on diagnosis or verification of varicella by a healthcare provider (for a patient reporting a history of or having an atypical case, a mild case, or both, healthcare providers should seek either an epidemiologic link with a typical varicella case or to a laboratory-confirmed case or evidence of laboratory confirmation, if it was performed at the time of acute disease); 4) history of herpes zoster based on diagnosis or verification of herpes zoster by a healthcare provider; or 5) laboratory evidence of immunity or laboratory confirmation of disease.

Pregnant women should be assessed for evidence of varicella immunity. Women who do not have evidence of immunity should receive the first dose of varicella vaccine upon completion or termination of pregnancy and before discharge from the healthcare facility. The second dose should be administered 4–8 weeks after the first dose.

4. Human papillomavirus (HPV) vaccination HPV vaccination with either quadrivalent (HPV4) vaccine or bivalent vaccine (HPV2) is recommended for females at age 11 or 12 years and catch-up vaccination for females aged 13 through 26 years.

Ideally, vaccine should be administered before potential exposure to HPV through sexual activity; however, females who are sexually active should still be vaccinated consistent with age-based recommendations. Sexually active females who have not been infected with any of the four HPV vaccine types (types 6, 11, 16, and 18, all of which HPV4 prevents) or any of the two HPV vaccine types (types 16 and 18, both of which HPV2 prevents) receive the full benefit of the vaccination. Vaccination is less beneficial for females who have already been infected with one or more of the HPV vaccine types. HPV4 or HPV2 can be administered to persons with a history of genital warts, abnormal Papanicolaou test, or positive HPV DNA test, because these conditions are not evidence of previous infection with all vaccine HPV types.

HPV4 may be administered to males aged 9 through 26 years to reduce their likelihood of genital warts. HPV4 would be most effective when administered before exposure to HPV through sexual contact.

A complete series for either HPV4 or HPV2 consists of 3 doses. The second dose should be administered 1–2 months after the first dose; the third dose should be administered 6 months after the first dose.

Although HPV vaccination is not specifically recommended for persons with the medical indications described in Figure 2, "Vaccines that might be indicated for adults based on medical and other indications," it may be administered to these persons because the HPV vaccine is not a live-virus vaccine. However, the immune response and vaccine efficacy might be less for persons with the medical indications described in Figure 2 than in persons who do not have the medical indications described or who are immunocompetent.

5. Herpes zoster vaccination A single dose of zoster vaccine is recommended for adults aged 60 years and older regardless of whether they report a previous episode of herpes zoster. Persons with chronic medical conditions may be vaccinated unless their condition constitutes a contraindication.

6. Measles, mumps, rubella (MMR) vaccination Adults born before 1957 generally are considered immune to measles and mumps. All adults born in 1957 or later should have documentation of 1 or more doses of MMR vaccine unless they have a medical contraindication to the vaccine, laboratory evidence of immunity to each of the three diseases, or documentation of provider-diagnosed measles or mumps disease. For rubella, documentation of provider-diagnosed disease is not considered acceptable evidence of immunity.

Measles component: A second dose of MMR vaccine, administered a minimum of 28 days after the first dose, is recommended for adults who 1) have been recently exposed to measles or are in an outbreak setting; 2) are students in postsecondary educational institutions; 3) work in a healthcare facility; or 4) plan to travel internationally. Persons who received inactivated (killed) measles vaccine or measles vaccine of unknown type during 1963–1967 should be revaccinated with 2 doses of MMR vaccine.

Mumps component: A second dose of MMR vaccine, administered a minimum of 28 days after the first dose, is recommended for adults who 1) live in a community experiencing a mumps outbreak and are in an affected age group; 2) are students in postsecondary educational institutions; 3) work in a healthcare facility; or 4) plan to travel internationally. Persons vaccinated before 1979 with either killed mumps vaccine or mumps vaccine of unknown type who are at high risk for mumps infection (e.g. persons who are working in a healthcare facility) should be revaccinated with 2 doses of MMR vaccine.

Rubella component: For women of childbearing age, regardless of birth year, rubella immunity should be determined. If there is no evidence of immunity, women who are not pregnant should be vaccinated. Pregnant women who do not have evidence of immunity should receive MMR vaccine upon completion or termination of pregnancy and before discharge from the healthcare facility.

Healthcare personnel born before 1957: For unvaccinated healthcare personnel born before 1957 who lack laboratory evidence of measles, mumps, and/or rubella immunity or laboratory confirmation of disease, healthcare facilities should 1) consider routinely vaccinating personnel with 2 doses of MMR vaccine at the appropriate interval (for measles and mumps) and 1 dose of MMR vaccine (for rubella), and 2) recommend 2 doses of MMR vaccine at the appropriate interval during an outbreak of measles or mumps, and 1 dose during an outbreak of rubella. Complete information about evidence of immunity is available at *http://www.cdc.gov/vaccines/recs/provisional/default.htm*.

7. Pneumococcal polysaccharide (PPSV) vaccination Vaccinate all persons with the following indications:

Medical: Chronic lung disease (including asthma); chronic cardiovascular diseases; diabetes mellitus; chronic liver diseases; cirrhosis; chronic alcoholism; functional or anatomic asplenia (e.g., sickle cell disease or splenectomy [if elective splenectomy is planned, vaccinate at least 2 weeks before surgery]); immunocompromising conditions (including chronic renal failure or nephrotic syndrome); and cochlear implants and cerebrospinal fluid leaks. Vaccinate as close to HIV diagnosis as possible.

Other: Residents of nursing homes or long-term care facilities and persons who smoke cigarettes. Routine use of PPSV is not recommended for American Indians/Alaska Natives or persons aged less than 65 years unless they have underlying medical conditions that are PPSV indications. However, public health authorities may consider recommending PPSV for American Indians/Alaska Natives and persons aged 50 through 64 years who are living in areas where the risk for invasive pneumococcal disease is increased

8. Revaccination with PPSV One-time revaccination after 5 years is recommended for persons aged 19 through 64 years with chronic renal failure or nephrotic syndrome; functional or anatomic asplenia (e.g., sickle cell disease or splenectomy); and for persons with immunocompromising conditions. For persons aged 65 years and older, one-time revaccination is recommended if they were vaccinated 5 or more years previously and were aged less than 65 years at the time of primary vaccination.

9. Meningococcal vaccination Meningococcal vaccine should be administered to persons with the following indications:

Medical: A 2-dose series of meningococcal conjugate vaccine is recommended for adults with anatomic or functional asplenia, or persistent complement component deficiencies. Adults with HIV infection who are vaccinated should also receive a routine 2-dose series. The 2 doses should be administered at 0 and 2 months.

Other: A single dose of meningococcal vaccine is recommended for unvaccinated first-year college students living in dormitories; microbiologists routinely exposed to isolates of Neisseria meningitidis; military recruits; and persons who travel to or live in countries in which meningococcal disease is hyperendemic or epidemic (e.g., the "meningitis belt" of sub-Saharan Africa during the dry season [December through June]), particularly if their contact with local populations will be prolonged. Vaccination is required by the government of Saudi Arabia for all travelers to Mecca during the annual Hajj.

Meningococcal conjugate vaccine, quadrivalent (MCV4) is preferred for adults with any of the

meningococcal conjugate vaccine, quadrivalent (MCV4) is preferred for adults with any of the preceding indications who are aged 55 years and younger; meningococcal polysaccharide vaccine (MPSV4) is preferred for adults aged 56 years and older. Revaccination with MCV4 every 5 years is recommended for adults previously vaccinated with MCV4 or MPSV4 who remain at increased risk for infection (e.g., adults with anatomic or functional asplenia, or persistent complement component deficiencies).

10. Hepatitis A vaccination Vaccinate persons with any of the following indications and any person seeking protection from hepatitis A virus (HAV) infection:

Behavioral: Men who have sex with men and persons who use injection drugs.

Occupational: Persons working with HAV-infected primates or with HAV in a research laboratory setting.

Medical: Persons with chronic liver disease and persons who receive clotting factor concentrates.

Other: Persons traveling to or working in countries that have high or intermediate endemicity of hepatitis A (a list of countries is available at http://wwwn.cdc.gov/travel/contentdiseases.aspx).

Unvaccinated persons who anticipate close personal contact (e.g., household or regular babysitting) with an international adoptee during the first 60 days after arrival in the United States from a country with high or intermediate endemicity should be vaccinated. The first dose of the 2-dose hepatitis A vaccine series should be administered as soon as adoption is planned, ideally 2 or more weeks before the arrival of the adoptee.

Single-antigen vaccine formulations should be administered in a 2-dose schedule at either 0 and 6–12 months (Havrix), or 0 and 6–18 months (Vaqta). If the combined hepatitis A and hepatitis B vaccine (Twinrix) is used, administer 3 doses at 0, 1, and 6 months; alternatively, a 4-dose schedule may be used, administered on days 0, 7, and 21–30, followed by a booster dose at month 12.

11. Hepatitis B vaccination Vaccinate persons with any of the following indications and any person seeking protection from hepatitis B virus (HBV) infection:

Behavioral: Sexually active persons who are not in a long-term, mutually monogamous relationship (e.g., persons with more than one sex partner during the previous 6 months); persons seeking evaluation or treatment for a sexually transmitted disease (STD); current or recent injection-drug users; and men who have sex with men.

Occupational: Healthcare personnel and public-safety workers who are exposed to blood or other potentially infectious body fluids.

Medical: Persons with end-stage renal disease, including patients receiving hemodialysis; persons with HIV infection; and persons with chronic liver disease.

Other: Household contacts and sex partners of persons with chronic HBV infection; clients and staff members of institutions for persons with developmental disabilities; and international travelers to countries with high or intermediate prevalence of chronic HBV infection (a list of countries is available at *http://wwwn.cdc.gov/travel/contentdiseases.aspx*).

Hepatitis B vaccination is recommended for all adults in the following settings: STD treatment facilities; HIV testing and treatment facilities; facilities providing drug-abuse treatment and prevention services; healthcare settings targeting services to injection-drug users or men who have sex with men; correctional facilities; end-stage renal disease programs and facilities for chronic hemodialysis patients; and institutions and nonresidential day-care facilities for persons with developmental disabilities.

Administer missing doses to complete a 3-dose series of hepatitis B vaccine to those persons not vaccinated or not completely vaccinated. The second dose should be administered 1 month after the first dose; the third dose should be given at least 2 months after the second dose (and at least 4 months after the first dose). If the combined hepatitis A and hepatitis B vaccine (Twinrix) is used, administer 3 doses at 0, 1, and 6 months; alternatively, a 4-dose Twinrix schedule, administered on days 0, 7, and 21 to 30, followed by a booster dose at month 12 may be used.

Adult patients receiving hemodialysis or with other immunocompromising conditions should receive 1 dose of 40 μg/mL (Recombivax HB) administered on a 3-dose schedule or 2 doses of 20 μg/mL (Engerix-B) administered simultaneously on a 4-dose schedule at 0, 1, 2, and 6 months.

12. Selected conditions for which Haemophilus influenzae type b (Hib) vaccine may be used 1 dose of Hib vaccine should be considered for persons who have sickle cell disease, leukemia, or HIV infection, or who have had a splenectomy, if they have not previously received Hib vaccine.

13. Immunocompromising conditions Inactivated vaccines generally are acceptable (e.g., pneumococcal, meningococcal, influenza [inactivated influenza vaccine]) and live vaccines generally are avoided in persons with immune deficiencies or immunocompromising conditions. Information on specific conditions is available at *http://www.cdc.gov/vaccines/pubs/acip-list.htm*.

Immunization Practice Standards

Administering immunizations to adults involves a number of processes, such as deciding whom to vaccinate, assessing vaccinecontraindications and precautions, providing vaccine information statements (VISs), ensuring appropriate storage and handling of vaccines, administering vaccines, and maintaining vaccine records. In addition, provider reporting of adverse events that follow vaccination is an essential component of the vaccine safety monitoring system.

DECIDING WHOM TO VACCINATE

Every effort should be made to ensure that adults receive all indicated vaccines as expeditiously as

possible. When adults present for care, their immunization history should be assessed and recorded, and this information should be used to identify needed vaccinations according to the most current version of the adult immunization schedule. Decision-support tools incorporated into electronic health records can provide prompts for needed vaccinations. Standing orders, which are often used for routinely indicated vaccines (e.g., influenza and pneumococcal vaccines), permit a nurse or another approved licensed practitioner to administer vaccines without a specific physician order, thus lowering barriers to adult immunization.

ASSESSING CONTRAINDICATIONS AND REGULATIONS

Before vaccination, all patients should be screened for contraindications and precautions. A *contraindication* is a condition that is believed to substantially increase the risk of a serious adverse reaction to vaccination. A vaccine should not be administered when a contraindication is documented. For example, a history of an anaphylactic reaction to a dose of vaccine or to a vaccine component is a contraindication for further doses. A *precaution* is a condition that may increase the risk of an adverse event or that may compromise the ability of the vaccine to evoke immunity (e.g., administering measles vaccine to a person who has recently received a blood transfusion and may consequently have transient passive immunity to measles). Normally, a vaccine is not administered when a precaution is noted. However, situations may arise when the benefits of vaccination outweigh the estimated risk of an adverse event, and the provider may decide to vaccinate the patient despite the precaution.

In some cases, contraindications and precautions are temporary and may lead to mere deferral of vaccination until a later time. For example, moderate or severe febrile illnesses are generally considered transient precautions to vaccination and result in postponement of vaccine administration until the acute phase has resolved; thus the superimposition of adverse effects of vaccination on the underlying illness and the mistaken attribution of a manifestation of the underlying illness to the vaccine are avoided. Contraindications and precautions to vaccines licensed in the United States for use in civilian adults are summarized in **Table 122-3**. It is important to recognize conditions that are *not* contraindications in order not to miss opportunities for vaccination. For example, in most cases, mild acute illness (with or without low-grade fever), a history of a mild to moderate local reaction to a previous dose of the vaccine, and breast-feeding are not contraindications to vaccination.

Table 122–3. Contraindications and Precautions for Commonly Used Vaccines in Adults

Vaccine Formulation	Contraindications and Precautions
All vaccines	**Contraindication** Severe allergic reaction (e.g., anaphylaxis) after a previous vaccine dose or to a vaccine component **Precaution** Moderate or severe acute illness with or without fever; defer vaccination until illness resolves
Td	**Precautions** GBS within 6 weeks after a previous dose of TT-containing vaccine History of Arthus-type hypersensitivity reactions after a previous dose of TT-containing vaccine; defer vaccination until at least 10 years have elapsed since the last dose
Tdap	**Contraindication** History of encephalopathy (e.g., coma or prolonged seizures) not attributable to another identifiable cause within 7 days of administration of a vaccine with pertussis components, such as DTaP or Tdap **Precautions** GBS within 6 weeks after a previous dose of TT-containing vaccine Unstable neurologic condition (e.g., cerebrovascular events and acute encephalopathic conditions) History of Arthus-type hypersensitivity reactions after a previous dose of TT-containing and/or DT-containing vaccine, including MCV4; defer vaccination until at least 10 years have elapsed since the last dose Pregnancy
HPV	**Contraindication** History of immediate hypersensitivity to yeast (for Gardasil) **Precaution**

	Pregnancy. If a woman is found to be pregnant after initiation of the vaccination series, the remainder of the 3-dose regimen should be delayed until after completion of the pregnancy. If a vaccine dose has been administered during pregnancy, no intervention is needed. A vaccine-in-pregnancy registry has been established for Gardasil; patients and health care providers should report any exposure to quadrivalent HPV vaccine during pregnancy (telephone: 800-986-8999).
MMR	**Contraindications** History of immediate hypersensitivity reaction to gelatin[a] or neomycin Pregnancy Known severe immunodeficiency (e.g., hematologic and solid tumors; chemotherapy; congenital immunodeficiency; long-term immunosuppressive therapy; severe immunocompromise due to HIV infection) **Precaution** Recent (within 11 months) receipt of antibody-containing blood product
Varicella	**Contraindications** Pregnancy Known severe immunodeficiency History of immediate hypersensitivity reaction to gelatin[a] or neomycin **Precaution** Recent (within 11 months) receipt of antibody-containing blood product
Influenza, injectable, trivalent	**Contraindication** History of immediate hypersensitivity reaction to eggs[b] **Precautions** History of GBS within 6 weeks after a previous influenza vaccine dose Pregnancy is *not* a contraindication or precaution. This vaccine is recommended for women who will be pregnant during influenza season.
Influenza, live attenuated	**Contraindications** History of immediate hypersensitivity reaction to eggs[b] Age ≥50 years Pregnancy Immunosuppression, including that caused by medications or by HIV infection; known severe immunodeficiency (e.g., hematologic and solid tumors; chemotherapy; congenital immunodeficiency; long-term immunosuppressive therapy; severe immunocompromise due to HIV infection) Certain chronic medical conditions, such as diabetes mellitus; chronic pulmonary disease (including asthma); chronic cardiovascular disease (except hypertension); renal, hepatic, neurologic/neuromuscular, hematologic, or metabolic disorders Close contact with severely immunosuppressed persons who require a protected environment, such as isolation in a bone marrow transplantation unit Close contact with persons with lesser degrees of immunosuppression (e.g., persons receiving chemotherapy or radiation therapy who are not being cared for in a protective environment; persons with HIV infection) is *not* a contraindication or precaution. **Precaution** History of GBS within 6 weeks of a previous influenza vaccine dose
Pneumococcal polysaccharide	None
Hepatitis A	**Precaution** Pregnancy
Hepatitis B	**Contraindication** History of immediate hypersensitivity to yeast
Meningococcal conjugate	**Contraindications** Age >55 years (licensed for use only among persons 2–55 years of age)

	History of severe allergic reaction to dry natural rubber (latex) or to DT-containing vaccines **Precautions** History of GBS
Meningococcal polysaccharide	**Contraindication** History of severe allergic reaction to dry natural rubber (latex)
Zoster	**Contraindications** Age <60 years Pregnancy Known severe immunodeficiency History of immediate hypersensitivity reaction to gelatin[a] or neomycin

[a]Extreme caution must be exercised in administering MMR, varicella, or zoster vaccine to persons with a history of an anaphylactic reaction to gelatin or gelatin-containing products. Before administration, skin testing for sensitivity to gelatin can be considered. However, no specific protocols for this purpose have been published.

[b]Protocols have been published for safely administering influenza vaccine to persons with egg allergies. See references 222–224 in Fiore AE et al: MMWR 57:1, 2008.

Abbreviations: DT, diphtheria toxoid; GBS, Guillain-Barré syndrome; HPV, human papillomavirus; MMR, measles, mumps, and rubella; Td, tetanus and diphtheria toxoids; Tdap, tetanus and diphtheria toxoids and acellular pertussis; TT, tetanus toxoid.

History of Immediate Hypersensitivity to a Vaccine Component

A severe allergic reaction (e.g., anaphylaxis) to a previous dose of a vaccine or to one of its components is a contraindication to vaccination. While most vaccines have many components, substances to which individuals are most likely to have had a severe allergic reaction include egg protein, gelatin, and yeast. In addition, although natural rubber (latex) is not a vaccine component, some vaccines are supplied in vials or syringes that contain natural rubber. These vaccines can be identified by the product insert and should not be administered to persons who report a severe (anaphylactic) allergy to latex. The much more common local or contact hypersensitivity to latex is *not* a contraindication to administration of a vaccine supplied in a vial or syringe that contains latex. Vaccines that, as of April 2009, were sometimes supplied in a vial or syringe containing natural rubber included Havrix hepatitis A vaccine (syringe), Vaqta hepatitis A vaccine (vial and syringe), Engerix-B hepatitis B vaccine (syringe), Recombivax HB hepatitis B vaccine (vial), Boostrix Tdap vaccine (syringe), and Menomune meningococcal polysaccharide vaccine (vial).

Pregnancy

Live-virus vaccines are contraindicated during pregnancy because of the possibility that vaccine virus replication will cause congenital infection or have other adverse effects on the fetus. Most live-virus vaccines, including varicella vaccine, are not secreted in breast milk; therefore, breast-feeding is not a contraindication for live-virus or other vaccines. Pregnancy is not a contraindication to administration of inactivated vaccines, but most are avoided during pregnancy because relevant safety data are limited. The only vaccine routinely recommended for women in the United States who are or will be pregnant during influenza season is trivalent inactivated influenza vaccine. Some other vaccines, such as tetanus and diphtheria toxoid (Td) vaccine and tetanus and diphtheria toxoid and acellular pertussis (Tdap) vaccine, may be given to pregnant women in certain circumstances. Resurgence of pertussis in some areas has prompted greater use of Tdap in pregnancy.

Immunosuppression

Live-virus vaccines elicit an immune response due to replication of the attenuated (weakened) vaccine virus that is contained by the recipient's immune system. In persons with compromised immune function, enhanced replication of vaccine viruses is possible and could lead to disseminated infection with the vaccine virus. For this reason, live-virus vaccines are contraindicated for persons with severe immunosuppression, defined according to the specific vaccine on the basis—at least in part—of differences in the prevalence of conditions causing immunosuppression at the time of vaccine recommendation issuance. Severe immunosuppression may be caused by many disease conditions, including HIV infection and hematologic or generalized malignancy. In some of these conditions, all affected persons are severely immunocompromised. In others (e.g., HIV infection), the degree to which the immune system is compromised depends on the severity of the condition, which in turn depends on the stage of disease or treatment. Severe immunosuppression may also be due to therapy with immunosuppressive agents, including high-dose glucocorticoids. In this situation, the dose, duration, and route of administration may influence the degree of immunosuppression.

The definition of severe immunosuppression that is a contraindication to zoster vaccine—the most recently licensed live-virus vaccine for adults—may be used as a guide to conditions that are also contraindications to other live-virus vaccines. Recommendations state that zoster vaccine should not be administered to persons with primary or acquired immunodeficiency, including the following:

1. Persons with leukemia, lymphomas, or other malignantneoplasms affecting the bone marrow or lymphatic system. However, patients whose leukemia is in remission and who have not received chemotherapy (e.g., alkylating drugs or antimetabolites) or radiation therapy for at least 3 months can receive zoster vaccine.
2. Persons with AIDS or other clinical manifestations of HIV infection, including persons with CD4+ T lymphocyte counts of ≤200/μL or ≤15% of total lymphocytes.
3. Persons receiving immunosuppressive therapy, including high-dose glucocorticoids (≥20 mg of prednisone per day or the equivalent) for ≥2 weeks. Zoster vaccination should be deferred for at least 1 month after discontinuation of such therapy. Short-term glucocorticoid therapy (<14 days); low to moderate glucocorticoid dosage (<20 mg of prednisone per day or the equivalent); topically applied glucocorticoids (e.g., those applied directly to the nose or skin or inhaled); intraarticular, bursal, or tendon glucocorticoid injections; and long-term alternate-day treatment with low to moderate doses of short-acting systemic glucocorticoids are not considered sufficiently immunosuppressive to cause concerns about vaccine safety and should not preclude the administration of zoster vaccine. Low doses of methotrexate (≤0.4 mg/kg per week), azathioprine (≤3.0 mg/kg per day), or 6-mercaptopurine (≤1.5 mg/kg per day) for treatment of rheumatoid arthritis, psoriasis, polymyositis, sarcoidosis, inflammatory bowel disease, and other conditions likewise do not constitute a contraindication.
4. Persons with clinical or laboratory evidence of other unspecified cellular immunodeficiency. However, persons with impaired humoral immunity (e.g., hypogammaglobulinemia or dysgammaglobulinemia) can receive zoster vaccine.
5. Persons undergoing hematopoietic stem cell transplantation. The experience of these patients with varicella-zoster virus–containing vaccines (e.g., zoster vaccine) is limited. Physicians should assess the immune status of the recipient on a case-by-case basis to determine the relevant risks. If a decision is made to administer zoster vaccine, vaccination should take place no sooner than 24 months after transplantation.
6. Persons receiving recombinant human immune mediators and immune modulators, especially the anti–tumor necrosis factor agents adalimumab, infliximab, and etanercept. The safety and efficacy of zoster vaccine administered concurrently with these agents is unknown. If it is not possible to administer zoster vaccine to patients before initiation of therapy, physicians should assess immune status on a case-by-case basis to determine the relevant risks and benefits. Otherwise, vaccination should be deferred for at least 1 month after discontinuation of such therapy.

Vaccine Information Statements

A VIS is a one-page (two-sided) information sheet produced by the CDC that informs vaccine recipients (or their parents or legal representatives) about the benefits and risks of a vaccine. VISs are mandated by the National Childhood Vaccine Injury Act (NCVIA) of 1986 and—whether the vaccine recipient is a child or an adult—must be provided for any vaccine covered by the Vaccine Injury Compensation Program. As of June 2009, vaccines that are covered by the NCVIA and that are licensed for use in adults include Td, Tdap, hepatitis A, hepatitis B, HPV, inactivated influenza, live intranasal influenza, measles/mumps/rubella (MMR), meningococcal, polio, and varicella vaccines. When combination vaccines for which no separate VIS exists are given (e.g., hepatitis A and B combination vaccine), all relevant VISs should be provided. VISs also exist for some vaccines not covered by the NCVIA, such as pneumococcal polysaccharide, Japanese encephalitis, rabies, zoster, typhoid, and yellow fever vaccines. The use of these VISs is encouraged but is not mandated.

All current VISs are available on the internet at two websites: the CDC's Vaccines & Immunizations site *(www.cdc.gov/vaccines)* and the Immunization Action Coalition's site *(www.immunize.org/vis/)*. (The latter site also includes translations of the VISs.) VISs from these sites can be downloaded and printed.

Storage and Handling

Injectable vaccines are packaged in multidose vials, single-dose vials, or manufacturer-filled single-dose syringes. The live attenuated nasal-spray influenza vaccine is packaged in single-dose sprayers. Oral typhoid vaccine is packaged in capsules. Some vaccines, such as MMR, varicella, zoster, and meningococcal polysaccharide vaccines, come as lyophilized (freeze-dried) powders that must be reconstituted (i.e., mixed with a liquid diluent) before use. The lyophilized powder and the diluent come in separate vials. Diluents are not interchangeable but rather are specifically formulated for each type of vaccine; only the specific diluent provided by the manufacturer for each type of vaccine should be used. Once lyophilized vaccines have been reconstituted, their shelf-life is limited and they must be stored under appropriate temperature and light conditions. For example, varicella and zoster vaccines must be protected from light and administered within 30 minutes of reconstitution; MMR vaccine likewise must be protected from light but can be used up to 8 h after reconstitution. Single-dose vials of meningococcal polysaccharide vaccine must be used within 30 minutes of reconstitution, while

multidose vials must be used within 35 days.

Vaccines are stored either at refrigerator temperature (2–8°C) or at freezer temperature (–15°C or colder). In general, inactivated vaccines (e.g., inactivated influenza, pneumococcal polysaccharide, and meningococcal conjugate vaccines) are stored at refrigerator temperature, while vials of lyophilized-powder live-virus vaccines (e.g., varicella, zoster, and MMR vaccines) are stored at freezer temperature. Diluents for lyophilized vaccines may be stored at refrigerator or room temperature. Live attenuated influenza vaccine—a live-virus liquid formulation administered by nasal spray—is stored at refrigerator temperature.

To avoid temperature fluctuations, vaccines should be placed in the body of a refrigerator and not in the door, in vegetable bins, on the floor, next to the wall, or next to the freezer—locations where temperatures may differ significantly. Frequent opening of a refrigerator door to retrieve food items can adversely affect the internal temperature of the unit and damage vaccines; thus food and drink should not be stored in the same refrigerator as vaccines. Frozen vaccines must be stored in the body (not the door) of a freezer that has its own external door separate from the refrigerator. They should not be stored in small "dormitory-style" refrigerators. The temperature of refrigerators and freezers used for vaccine storage must be monitored and the temperature recorded at least twice a day. Ideally, continuous thermometers are used that measure and record temperature all day and all night.

Administration of Vaccines

Parenteral vaccines recommended for routine administration to adults in the United States are given by either the IM or the SC route. Most parenteral vaccines are given to adults by the IM route. Vaccines given by the SC route include live-virus vaccines such as varicella, zoster, and MMR vaccines as well as the inactivated meningococcal polysaccharide vaccine. The 23-valent pneumococcal polysaccharide vaccine may be given by either of these routes, but IM administration is preferred because it is associated with a lower risk of injection-site reactions.

Vaccines given to adults by the SC route are administered with a 5/8-inch needle into the upper outer-triceps area (Fig. 122-2). Vaccines administered to adults by the IM route are injected into the deltoid muscle (Fig. 122-2) with a needle whose length should be selected on the basis of the recipient's sex and weight to ensure adequate penetration into the muscle. Current guidelines indicate that, for men and women weighing <130 lbs (<60 kg), a 5/8-inch needle is sufficient; for women weighing 130–200 lbs (60–90 kg) and men weighing 130–260 lbs (60–118 kg), a 1- to 1.5-inch needle is needed; and for women weighing >200 lbs (>90 kg) and men weighing >260 lbs (>118 kg), a 1.5-inch needle is required.

Figure 122-2

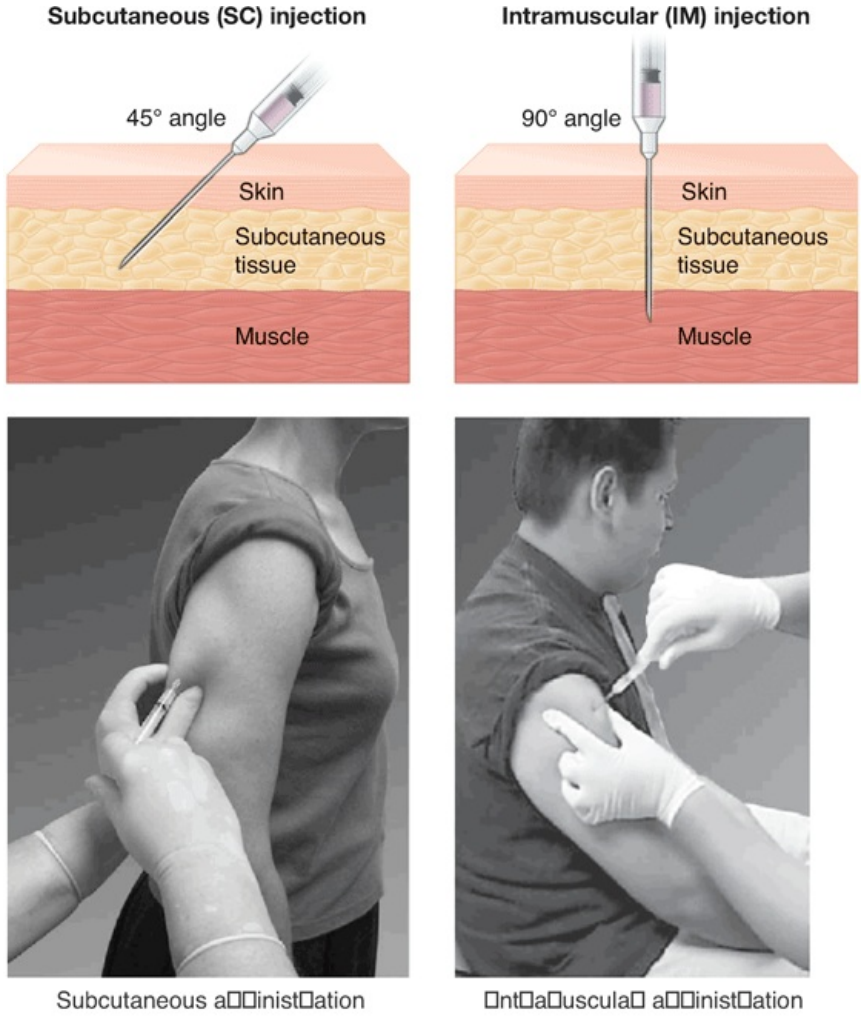

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Techniques for SC and IM administration of vaccines to adults. *(Adapted from materials provided by the Immunization Action Coalition; www.immunize.org.)*

Aspiration is the process of pulling back on the plunger of the syringe after skin penetration but prior to injection to ensure that the contents of the syringe are not injected into a blood vessel. Although this practice is advocated by some experts, aspiration is not required because of the lack of large blood vessels at the recommended vaccine injection sites.

Multiple vaccines can be administered at the same visit; indeed, administration of all needed vaccines at one visit is encouraged. Studies have shown that vaccines are as effective when administered simultaneously as they are individually, and simultaneous administration of multiple vaccines is not associated with an increased risk of adverse effects. If more than one vaccine must be administered in the same limb, the injection sites should be separated by 1–2 inches so that any local reactions can be differentiated. If a vaccine and an immune globulin preparation are administered simultaneously (e.g., Td vaccine and tetanus immune globulin), a separate anatomic site should be used for each injection.

For certain vaccines (e.g., HPV vaccine and hepatitis B vaccine), multiple doses are required for an adequate and persistent antibody response. The recommended vaccination schedule specifies the interval between doses. Many adults who receive the first dose in a multiple-dose vaccine series do not complete the series or do not receive subsequent doses within the recommended interval; in these circumstances, vaccine efficacy and/or the duration of protection may be compromised. Providers should implement recall systems that will prompt patients to return for subsequent doses in a vaccination series at the appropriate intervals. With the exception of oral typhoid vaccination, an interruption in the schedule does not require restarting of the entire series or the addition of extra doses.

Syncope may follow vaccination, especially with adolescents and young adults. Serious injuries, including head trauma and motor vehicle accidents, have occurred. Vaccine administration procedures that minimize the impact of postvaccination syncope should be used. To avoid trauma when syncope does occur, patients should be seated during vaccination. The ACIP recommends that vaccine providers strongly consider observing patients for 15 minutes after they are vaccinated. If syncope develops, patients should be observed until the symptoms resolve.

Anaphylaxis is a rare complication of vaccination. All facilities providing immunizations should have an emergency kit containing aqueous epinephrine for administration in the event of a systemic anaphylactic reaction.

Vaccine Safety Monitoring and Adverse Event Reporting

PRELICENSURE EVALUATIONS OF VACCINE SAFETY

Before vaccines are licensed by the FDA, they are evaluated in clinical trials with volunteers. These trials are conducted in three progressive phases. Phase 1 trials are small, usually involving fewer than 100 volunteers. Their purposes are to provide a basic evaluation of safety and to identify common adverse events. Phase 2 trials, which are larger and may involve several hundred participants, collect additional information on safety and are usually designed to evaluate immunogenicity as well. Data gained from phase 2 trials can be used to determine the composition of the vaccine, the number of doses required, and a profile of common adverse events. Vaccines that appear promising are evaluated in phase 3 trials, which typically involve several hundred to several thousand volunteers and are generally designed to demonstrate vaccine efficacy and provide additional information on vaccine safety.

POSTLICENSURE MONITORING OF VACCINE SAFETY

After licensure, a vaccine's safety is assessed by several mechanisms. The NCVIA of 1986 requires health care providers to report certain adverse events that follow vaccination of children. As a mechanism for that reporting, the Vaccine Adverse Event Reporting System (VAERS) was established in 1990 and is jointly managed by the CDC and the FDA. This safety surveillance system collects reports of adverse events associated with vaccines currently licensed in the United States. *Adverse events* are defined as health effects that occur after immunization and that may or may not be related to the vaccine. While VAERS was established in response to the NCVIA, any adverse event following vaccination—whether in a child or an adult, and whether or not it is believed to have been caused by vaccination—may be reported through VAERS. In 2008, VAERS received >25,000 reports of adverse events following vaccination. Of those, 9.5% were reportedly serious, causing disability, hospitalization, life-threatening illness, or death.

Anyone can file a VAERS report, including health care providers, manufacturers, and vaccine recipients or their parents or guardians. VAERS reports may be submitted online (*http://vaers.hhs.gov/esub/index*) or by completing a paper form requested online, by phone (800-822-7967), or by fax (877-721-0366). The VAERS form asks for the following information: the type of vaccine received; the timing of vaccination; the time of onset of the adverse event; and the recipient's current illnesses or medications, history of adverse events following vaccination, and demographic characteristics (e.g., age and gender). This information is entered into a database. The individual who reported the adverse event then receives a confirmation letter by mail with a VAERS identification number that can be used if additional information is submitted later. In selected cases of serious adverse reaction, the patient's recovery status may be followed up at 60 days and 1 year after vaccination. The FDA and the CDC have access to VAERS data and use this information to monitor vaccine safety and conduct research studies. VAERS data (minus personal information) are also available to the public.

While the VAERS provides useful information on vaccine safety, this passive reporting system has important limitations. One is that it only collects information about events following vaccination; it does not assess whether a given type of event occurs more often than expected after vaccination. A second is that event reporting is incomplete and is biased toward events that are believed to be more likely to be due to vaccination and that occur relatively soon after vaccination. To obtain more systematic information on adverse events occurring in both vaccinated and unvaccinated persons, the Vaccine Safety Datalink project was initiated in 1991. Directed by the CDC, this project includes eight managed-care organizations in the United States; member databases include information on immunizations, medical conditions, demographics, laboratory results, and medication prescriptions. The Department of Defense oversees a similar system monitoring the safety of immunizations among active-duty military personnel. In addition, postlicensure evaluations of vaccine safety may be conducted by the vaccine manufacturer. In fact, such evaluations are often required by the FDA as a condition of vaccine licensure.

MAINTENANCE OF VACCINE RECORDS

All vaccines administered should be fully documented in the patient's permanent medical record. Documentation should include the date of administration, the name or common abbreviation of the vaccine, the vaccine lot number and manufacturer, the administration site, the VIS edition, the date the VIS was provided, and the name of the person who administered the vaccine.

Consumer Access to and Demand for Immunization

By removing barriers to the consumer or patient, providers and health care institutions can improve vaccine use. Financial barriers have traditionally been important constraints, particularly among uninsured adults. Even for insured adults, out-of-pocket costs associated with newer, more expensive

adult vaccines (e.g., zoster vaccine) are an obstacle to be overcome. After influenza vaccine was included by Medicare for all beneficiaries in 1993, coverage among persons ≥65 years of age doubled (from ~30% in 1989 to >60% in 1997; **Fig. 122-3**). Other strategies that enhance patients' access to vaccination include extended office hours (e.g., evening and weekend hours) and scheduled vaccination-only clinics where waiting times are reduced. Provision of vaccines outside the "medical home" (e.g., through occupational clinics, universities, and retail settings) can expand access for adults who do not make medical visits frequently. Increasing proportions of nonelderly adults are being vaccinated in these settings.

Figure 122-3

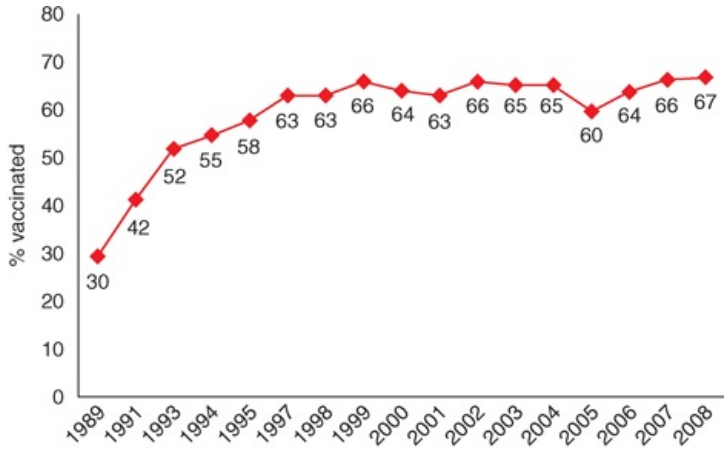

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Influenza vaccination coverage among adults ≥65 years of age, United States, 1989–2008. *(From http://www.cdc.gov/FLU/PROFESSIONALS/VACCINATION/pdf/NHIS89_08fluvaxtrendtab.pdf.)*

Health promotion efforts aimed at increasing the demand for immunization are common. Direct-to-consumer advertising by pharmaceutical companies has been used for some newer adolescent and adult vaccines. Efforts to raise consumer demand for vaccines have not increased immunization rates unless implemented in conjunction with other strategies that target strengthening of provider practices or reduction of consumer barriers. Attitudes and beliefs related to vaccination can be considerable impediments to consumer demand. Many adults view vaccines as important for children but are less familiar with vaccinations targeting disease prevention in adults. Several vaccines are recommended for adults with certain medical risk factors, but self-identification as a high-risk individual is relatively rare. Communication research suggests that many adults with chronic diseases may be more motivated to receive a vaccine by a desire to protect their family members rather than to reduce their own risk. Some vaccines are explicitly recommended for persons at relatively low risk of serious complications, with the goal of reducing the risk of transmission to higher-risk contacts. For example, for parents and caretakers of newborns, vaccinations against influenza and pertussis are recommended.

Strategies for Providers and Health Care Facilities

RECOMMENDATION FROM THE PROVIDER

Health care providers can have great influence on patients with regard to immunization. A recommendation from a doctor or nurse carries more weight than do recommendations from professional societies or endorsements by celebrities. Providers should be well informed about vaccine risks and benefits so that they can address patients' common concerns. The CDC, the American College of Physicians, and the American Academy of Family Physicians review and update the schedule for adult immunization on an annual basis and also have developed educational materials to facilitate provider–patient discussions about vaccination (*www.cdc.gov/vaccines*).

SYSTEM SUPPORTS

Medical offices can incorporate a variety of methods to ensure that providers consistently offer specific immunizations to patients with indications for specific vaccines. Decision-support tools have been incorporated into some electronic health records to alert the provider when specific vaccines are indicated. Manual or automated reminders and standing orders have been discussed (see "Deciding Whom to Vaccinate," above) and have consistently improved vaccination coverage in both office and hospital settings. Most clinicians' estimates of their own performance diverge from objective measurements of their patients' immunization coverage; quantitative assessment and feedback have been shown in pediatric practice to increase immunization performance significantly. Some health plans have instituted incentives for providers with high rates of immunization coverage. Specialty

providers, including obstetrician–gynecologists, may be the only providers serving high-risk patients with indications for selected vaccines (e.g., HPV, influenza, or pneumococcal polysaccharide vaccine).

IMMUNIZATION REQUIREMENTS

Vaccination against selected communicable diseases is required for attendance at many universities and colleges as well as for service in the U.S. military or in some occupational settings (e.g., child care, laboratory, veterinary, and health care). Immunizations are recommended and sometimes required for travel to certain countries (see Chap. 123).

VACCINATION OF HEALTH CARE STAFF

A particular area of focus for medical settings is vaccination of health care workers, including those with and without direct patient-care responsibilities. The Joint Commission (which accredits health care organizations), the CDC's Healthcare Infection Control Practices Advisory Committee, and the ACIP all recommend influenza vaccination of all health care personnel; new recommendations focus on requiring documentation of declination for providers that do not accept annual influenza vaccination. Some institutions and jurisdictions have added mandates on influenza vaccination of health care workers and have expanded on earlier requirements related to vaccination or proof of immunity for hepatitis B, measles, mumps, rubella, and varicella.

Vaccination in Nonmedical Settings

Rates of vaccination in medical offices are highest among young children and adults ≥65 years of age. People in these age groups make more office visits and are more likely to receive care in a consistent "medical home" than older children, adolescents, and nonelderly adults. Vaccination outside the medical home can expand access to those whose health care visits are limited and reduce the burden on busy clinical practices. In some locations, financial constraints related to inventory and storage requirements have led providers to stock few or no vaccines. Outside private office and hospital settings, vaccination may also occur at health department venues, workplaces, retail sites (including pharmacies and supermarkets), and schools or colleges.

When vaccines are given in nonmedical settings, it remains important for standards of immunization practice to be followed. Consumers should be provided with information on how to report adverse events (e.g., via provision of a VIS), and procedures should ensure that documentation of vaccine administration is forwarded to the primary care provider and the state or city public health immunization registry. Detailed documentation may be required for employment, school attendance, and travel. Personalized health records can help consumers keep track of their immunizations, and some occupational health clinics have incorporated automated immunization reports that help employees stay up-to-date with recommended vaccinations.

Performance Monitoring

Tracking of immunization coverage at national, state, institution, and practice levels can yield feedback to practitioners and programs and facilitate quality improvement. Healthcare Effectiveness Data and Information Set (HEDIS) measures related to adult immunization facilitate comparison of health plans. The CDC's National Immunization Survey and National Health Interview Survey provide selected information on immunization coverage among adults and track progress toward achievement of Healthy People 2020 targets for coverage among persons ≥65 years of age as well as among younger adults with conditions that increase risk. Influenza and pneumococcal vaccine coverage rates have been higher among persons ≥65 years of age (60–70%) than among high-risk 18- to 64-year-olds. Figures on state-specific immunization coverage with pneumococcal polysaccharide and influenza vaccines (as measured through the CDC's Behavioral Risk Factor Surveillance System) reveal substantial geographic variation in coverage. There are persistent disparities in adult immunization coverage rates between whites and racial and ethnic minorities. In contrast, racial and economic disparities in immunization of young children have been dramatically reduced during the past decade. Much of this progress is attributed to the Vaccines for Children Program, which since 1994 has entitled uninsured children to receive free vaccines. Approximately 70% of African-American and Hispanic children are eligible for this program.

Future Trends

Although most vaccines developed in the twentieth century targeted common acute infectious diseases of childhood, more recently developed vaccines prevent chronic conditions prevalent among adults. Hepatitis B vaccine prevents hepatitis B–related cirrhosis and hepatocellular carcinoma, zoster vaccine prevents shingles and postherpetic neuralgia, and HPV vaccine prevents some types of cervical cancer as well as genital warts and anogenital cancers. New targets of vaccine development and research may further broaden the definition of vaccine-preventable disease. Research is ongoing on vaccines to prevent insulin-dependent diabetes mellitus, nicotine addiction, and Alzheimer's disease. Expanding strategies for vaccine development are incorporating molecular approaches such as DNA, vector, and peptide vaccines. New technologies, such as the use of transdermal and other needle-less routes of administration, are being applied to vaccine delivery.

FURTHER READINGS

Bonhoeffer J et al: Guidelines for collection, analysis, and presentation of vaccine safety data in surveillance systems. Vaccine 27:2289, 2009[PMID: 19061929]

Centers for disease control and prevention: Progress toward interruption of wild poliovirus transmission worldwide, 2009. MMWR 59:545, 2010

Pickering LK et al: Immunization programs for infants, children, adolescents, and adults: Clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 49:817, 2009[PMID: 19659433]

Plotkin S et al (eds): *Vaccines*, 5th ed. Philadelphia, Saunders Elsevier, 2008

Poland GA et al: Standards for adult immunization practices. Am J Prev Med 25:144, 2003[PMID: 12880883]

Roush SW et al: Historical comparisons of morbidity and mortality for vaccine-preventable diseases in the United States. JAMA 298:2155, 2007[PMID: 18000199]

Schuchat A, Bell B: Monitoring the impact of vaccines post-licensure: New challenges, new opportunities. Expert Rev Vaccines 7:437, 2008[PMID: 18444891]

Smith JC et al: Immunization policy development in the United States: The role of the Advisory Committee on Immunization Practices. Ann Intern Med 150:45, 2009[PMID: 19124820]

Wharton M: Vaccine safety: Current systems and recent findings. Curr Opin Pediatr 22:88, 2010[PMID: 19952750]

Willis BC et al: Improving influenza, pneumococcal polysaccharide, and hepatitis B vaccination coverage among adults aged <65 years at high risk: A report on recommendations of the Task Force on Community Preventive Services. MMWR 54:1, 2005[PMID: 15800472]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEALTH RECOMMENDATIONS FOR INTERNATIONAL TRAVEL: INTRODUCTION

According to the World Tourism Organization, international tourist arrivals grew exponentially from 25 million in 1950 to >900 million in 2008. Not only are more people traveling; travelers are seeking more exotic and remote destinations. Travel from industrialized to developing regions has been increasing, with Asia and the Pacific, Africa, and the Middle East now emerging destinations. **Figure 123-1** summarizes the monthly incidence of health problems during travel in developing countries. Studies show that 50–75% of short-term travelers to the tropics or subtropics report some health impairment. Most of these health problems are minor: only 5% require medical attention, and <1% require hospitalization. Although infectious agents contribute substantially to morbidity among travelers, these pathogens account for only ~1% of deaths in this population. Cardiovascular disease and injuries are the most frequent causes of death among travelers from the United States, accounting for 49% and 22% of deaths, respectively. Age-specific rates of death due to cardiovascular disease are similar among travelers and nontravelers. In contrast, rates of death due to injury (the majority from motor vehicle, drowning, or aircraft accidents) are several times higher among travelers. If one excludes mortality due to cardiovascular disease and preexisting illness, motor vehicle accidents account for >40% of the remaining deaths.

Figure 123-1

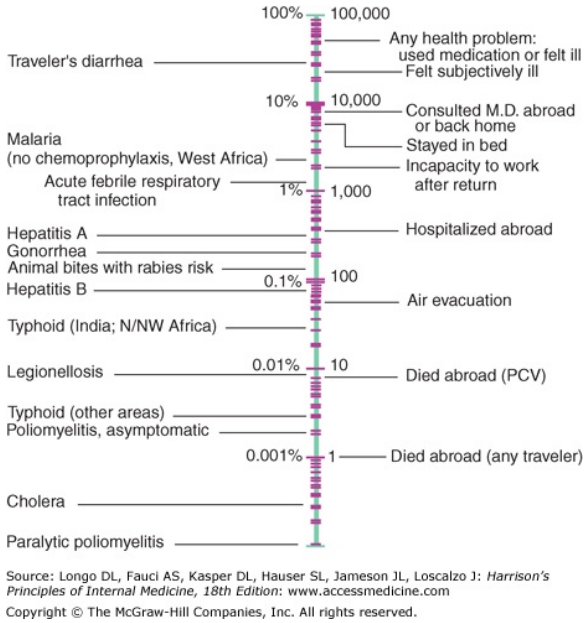

Incidence rate, per month, of health problems during a stay in developing countries. PCV, Peace Corps volunteer. *(From Steffen R, Lobel HO: Epidemiologic basis for the practice of travel medicine. J Wilderness Med 5:56, 1994. Reprinted with permission from Chapman and Hall, New York.)*

GENERAL ADVICE

Health maintenance recommendations are based not only on the traveler's destination but also on assessment of risk, which is determined by such variables as health status, specific itinerary, purpose of travel, season, and lifestyle during travel. Detailed information regarding country-specific risks and recommendations may be obtained from the Centers for Disease Control and Prevention (CDC) publication *Health Information for International Travel* (available at *wwwnc.cdc.gov/travel/*).

Fitness for travel is an issue of growing concern in view of the increased numbers of elderly and chronically ill individuals journeying to exotic destinations (see "Travel and Special Hosts," below). Since most commercial aircraft are pressurized to 2500 m (8000 ft) above sea level (corresponding to a Pa_{O2} of ~55 mmHg), individuals with serious cardiopulmonary problems or anemia should be evaluated before travel. In addition, those who have recently had surgery, a myocardial infarction, a cerebrovascular accident, or a deep vein thrombosis may be at high risk for adverse events during flight. A summary of current recommendations regarding fitness to fly has been published by the Aerospace Medical Association Air Transport Medicine Committee (*www.asma.org/publications/*). A pretravel health assessment may be advisable for individuals considering particularly adventurous recreational activities, such as mountain climbing and scuba diving.

Immunizations for Travel

Immunizations for travel fall into three broad categories: *routine* (childhood/adult boosters that are necessary regardless of travel), *required* (immunizations that are mandated by international regulations for entry into certain areas or for border crossings), and *recommended* (immunizations that are desirable because of travel-related risks). Required and recommended vaccines commonly given to travelers are listed in **Table 123-1**.

Table 123–1. Vaccines Commonly Used for Travel

Vaccine	Primary Series	Booster Interval
Cholera, live oral (CVD 103 - HgR)	1 dose	6 months
Hepatitis A (Havrix), 1440 enzyme immunoassay	2 doses, 6–12 months apart, IM	None required

U/mL		
Hepatitis A (VAQTA, AVAXIM, EPAXAL)	2 doses, 6–12 months apart, IM	None required
Hepatitis A/B combined (Twinrix)	3 doses at 0, 1, and 6–12 months *or* 0, 7, and 21 days plus booster at 1 year, IM	None required *except* 12 months (once only, for accelerated schedule)
Hepatitis B (Engerix B): accelerated schedule	3 doses at 0, 1, and 2 months *or* 0, 7, and 21 days plus booster at 1 year, IM	12 months, once only
Hepatitis B (Engerix B or Recombivax): standard schedule	3 doses at 0, 1, and 6 months, IM	None required
Immune globulin (hepatitis A prevention)	1 dose IM	Intervals of 3–5 months, depending on initial dose
Japanese encephalitis (JE-VAX)	3 doses, 1 week apart, SC	12–18 months (first booster), then 4 years
Japanese encephalitis (Ixiaro)	2 doses, 1 month apart, SC	Optimal booster schedule not yet determined
Meningococcus, quadrivalent [Menimmune (polysaccharide), Menactra, Menveo (conjugate)]	1 dose SC	>3 years (optimal booster schedule not yet determined)
Rabies (HDCV), rabies vaccine absorbed (RVA), or purified chick embryo cell vaccine (PCEC)	3 doses at 0, 7, and 21 or 28 days, IM	None required except with exposure
Typhoid Ty21a, oral live attenuated (Vivotif)	1 capsule every other day x 4 doses	5 years
Typhoid Vi capsular polysaccharide, injectable (Typhim Vi)	1 dose IM	2 years
Yellow fever	1 dose SC	10 years

ROUTINE IMMUNIZATIONS

Diphtheria, Tetanus, and Polio

Diphtheria (Chap. 138) continues to be a problem worldwide. Large outbreaks have occurred in countries that have reduced their public vaccination programs. Serologic surveys show that tetanus (Chap. 140) antitoxin is lacking in many North Americans, especially in women over the age of 50. The risk of polio (Chap. 191) to the international traveler is extremely low, and wild-type poliovirus has been eradicated from the Western Hemisphere and Europe. However, studies in the United States suggest that 12% of adult travelers are unprotected against at least one poliovirus serogroup. Foreign travel offers an ideal opportunity to have these immunizations updated. With the recent increase in pertussis among adults, the diphtheria–tetanus–acellular pertussis (Tdap) combination is now recommended for adults as a once-only replacement for the 10-year Td booster.

Measles

Measles (rubeola) continues to be a major cause of morbidity and death in the developing world (Chap. 192). Several outbreaks of measles in the United States have been linked to imported cases. The group at highest risk consists of persons born after 1956 and vaccinated before 1980, in many of whom primary vaccination failed.

Influenza

Influenza—possibly the most common vaccine-preventable infection in travelers—occurs year-round in the tropics and during the summer months in the Southern Hemisphere (coinciding with the winter months in the Northern Hemisphere). One prospective study showed that influenza developed in 1% of travelers to Southeast Asia per month of stay. Vaccination should be considered for all travelers to these regions, particularly those who are elderly or chronically ill. Travel-related influenza continues to occur during summer months in Alaska and the Northwest Territories of Canada among cruise-ship passengers and staff (Chap. 187). The speed of global spread of the pandemic H1N1 virus once again illustrates why influenza immunization is so important for travelers.

Pneumococcal Infection

Regardless of travel, pneumococcal vaccine should be administered routinely to the elderly and to persons at high risk of serious infection, including those with chronic heart, lung, or kidney disease and those who have been splenectomized or have sickle cell disease (Chap. 134).

REQUIRED IMMUNIZATIONS

Yellow Fever

Documentation of vaccination against yellow fever (Chap. 196) may be required as a condition of entry into or passage through countries of sub-Saharan Africa and equatorial South America, where the disease is endemic or epidemic, or for entry into countries at risk of having the infection introduced. This vaccine is given only by state-authorized yellow fever centers, and its administration must be documented on an official International Certificate of Vaccination. A registry of U.S. clinics that provide the vaccine is available from the CDC (*wwwnc.cdc.gov/travel/*). Recent data suggest that fewer than 50% of travelers entering areas endemic for yellow fever are immunized. Severe adverse events associated with this vaccine have recently increased in incidence. First-time vaccine recipients may present with a syndrome characterized as either neurotropic (1 case per 125,000 doses) or viscerotropic (1 case per 250,000 doses; among persons 60–69 years of age, 1 case per 100,000 doses; and among persons ≥70 years of age, 1 case per 40,000 doses). Immunosuppression and thymic disease increase the risk of these adverse events (*www.cdc.gov/vaccines/pubs/vis/downloads/vis-yf.pdf*).

Meningococcal Meningitis

Protection against meningitis (using one of the quadrivalent vaccines) is required for entry into Saudi Arabia during the Hajj (Chap. 143).

Influenza

Both seasonal and pandemic H1N1 vaccines (the latter, where available) were required for entry into Saudi Arabia during the Hajj in 2009.

RECOMMENDED IMMUNIZATIONS

Hepatitis A and B

Hepatitis A (Chap. 304) is one of the most common vaccine-preventable infections of travelers. The risk is six times greater for travelers who stray from the usual tourist routes. The mortality rate for hepatitis A increases with age, reaching almost 3% among individuals over age 50. Of the four hepatitis A vaccines currently available in North America (two in the United States), all are

interchangeable and have an efficacy rate of >95%.

Long-stay overseas workers appear to be at considerable risk for hepatitis B infection (Chap. 304). The recommendation that all travelers be immunized against hepatitis B before departure is supported by two studies showing that 17% of the assessed travelers who received health care abroad had some type of injection; according to the World Health Organization, nonsterile equipment is used for up to 75% of all injections given in the developing world. A 3-week accelerated schedule of the combined hepatitis A and B vaccine has been approved in the United States. Virtually all travelers to less developed countries should be immunized against hepatitis A and B.

Typhoid Fever

The attack rate for typhoid fever (Chap. 153) is 1 case per 30,000 travelers per month of travel to the developing world. However, the attack rates in India, Senegal, and North Africa are tenfold higher and are especially high among travelers to relatively remote destinations and among VFRs (immigrants and their families returning to their homelands to visit friends or relatives). Between 1999 and 2006 in the United States, 66% of imported cases involved the latter group. Both of the available vaccines—one oral (live) and the other injectable (polysaccharide)—have efficacy rates of ~70%. In some countries, a combined hepatitis A/typhoid vaccine is available.

Meningococcal Meningitis

Although the risk of meningococcal disease among travelers has not been quantified, it is likely to be higher among travelers who live with poor indigenous populations in overcrowded conditions (Chap. 143). Because of its enhanced ability to prevent nasal carriage (compared with the older polysaccharide vaccine), a quadrivalent conjugate vaccine is the product of choice for immunization of persons traveling to sub-Saharan Africa during the dry season or to areas of the world where there are epidemics. The vaccine, which protects against serogroups A, C, Y, and W-135, has an efficacy rate of >90%.

Japanese Encephalitis

The risk of Japanese encephalitis (Chap. 196), an infection transmitted by mosquitoes in rural Asia and Southeast Asia, is ~1 case per 5000 travelers per month of stay in an endemic area. Most symptomatic infections among U.S. residents have involved military personnel or their families. The vaccineefficacy rate is >90%. The vaccine is recommended for persons staying>1 month in rural endemic areas or for shorter periods if their activities (e.g., camping, bicycling, hiking) in these areas will increase exposure risk. A new Vero cell vaccine is now available in the United States.

Cholera

The risk of cholera (Chap. 156) is extremely low, with ~1 case per 500,000 journeys to endemic areas. Cholera vaccine, no longer available in the United States, was rarely recommended but was considered for aid and health care workers in refugee camps or in disaster-stricken/war-torn areas. A more effective oral cholera vaccine is available in other countries.

Rabies

Domestic animals, primarily dogs, are the major transmitters of rabies in developing countries (Chap. 195). Several studies have shown that the risk of rabies posed by a dog bite in an endemic area translates into 1–3.6 cases per 1000 travelers per month of stay. Countries where canine rabies is highly endemic include Mexico, the Philippines, Sri Lanka, India, Thailand, and Vietnam. The three vaccines available in the United States provide >90% protection. Rabies vaccine is recommended for long-stay travelers, particularly children, and persons who may be occupationally exposed to rabies in endemic areas; however, in a large-scale study, almost 50% of potential exposures occurred within the first month of travel. Even after receipt of a preexposure rabies vaccine series, two postexposure doses are required. Travelers who have had the preexposure series do not require rabies immune globulin (which is often unavailable in developing countries) if they are exposed to the disease.

Prevention of Malaria and Otherinsect-Borne Diseases

It is estimated that more than 30,000 American and European travelers develop malaria each year (Chap. 210). The risk to travelers is highest in Oceania and sub-Saharan Africa (estimated at 1:5 and 1:50 per month of stay, respectively, among persons not using chemoprophylaxis); intermediate in malarious areas on the Indian subcontinent and in Southeast Asia (1:250–1:1000 per month); and low in South and Central America (1:2500–1:10,000 per month). Of the more than 1000 cases of malaria reported annually in the United States, 90% of those due to *Plasmodium falciparum* occur in travelers returning or immigrating from Africa and Oceania. VFRs are at the highest risk of acquiring malaria. With the worldwide increase in chloroquine- and multidrug-resistant falciparum malaria, decisions about chemoprophylaxis have become more difficult. In addition, the spread of malaria due to primaquine- and chloroquine-resistant strains of *Plasmodium vivax* has added to the complexity of treatment, as has the recently described "monkey malaria" of humans, which is caused by *P. knowlesi*. The case-fatality rate of falciparum malaria in the United States is 4%; however, *in only one-third of patients who die is the diagnosis of malaria considered before death*.

Several studies indicate that fewer than 50% of travelers adhere to basic recommendations for malaria prevention. Keys to the prevention of malaria include both personal protection measures against mosquito bites (especially between dusk and dawn) and malaria chemoprophylaxis. The former measures include the use of DEET-containing insect repellents, permethrin-impregnated bed-nets and clothing, screened sleeping accommodations, and protective clothing. Thus, in regions where infections such as malaria are transmitted, DEET products (25–50%) are recommended, even for children and infants at birth. Studies suggest that concentrations of DEET above ~50% do not offer a marked increase in protection time against mosquitoes. The CDC also recommends picaridin (icaridin), oil of lemon eucalyptus (PMD, para-menthane-3,8-diol), and IR3535 (ethyl butylacetylaminopropionate). In general, higher concentrations of active ingredient provide longer duration of protection, regardless of the active ingredient. Personal protection measures also help prevent other insect-transmitted illnesses, such as dengue fever (Chap. 196). Over the past decade, the incidence of dengue has increased, particularly in the Caribbean region, Latin America, and Southeast Asia. Both dengue and chikungunya viruses are transmitted by an urban-dwelling mosquito that bites primarily at dawn and dusk.

Table 123-2 lists the currently recommended drugs of choice for prophylaxis of malaria, by destination.

Table 123–2. Malaria Chemosuppressive Regimens, According to Geographic Area[a]

Geographic Area	Drug of Choice	Alternatives
Central America (north of Panama), Iraq, Turkey, northern Argentina, and Paraguay	Chloroquine	Atovaquone-proguanil[b] Doxycycline Mefloquine Primaquine (except

		Honduras)
South America, including Haiti, Dominican Republic, and Panama but not northern Argentina or Paraguay; Asia, including Southeast Asia; Africa; and Oceania	Doxycycline Atovaquone-proguanil Mefloquine	
Thai-Myanmar and Thai-Cambodian borders and central Vietnam	Atovaquone-proguanil Doxycycline	

[a]See CDC's *Health Information for International Travel 2010* .

[b]Malarone.

Note: See also Chap. 210.

Prevention of Gastrointestinal Illness

Diarrhea, the leading cause of illness in travelers (Chap. 128), is usually a short-lived, self-limited condition; however, 40% of affected individuals need to alter their scheduled activities, and another 20% are confined to bed. The most important determinant of risk is the destination. Incidence rates per 2-week stay have been reported to be as low as 8% in industrialized countries and as high as 55% in parts of Africa, Central and South America, and Southeast Asia. Infants and young adults are at particularly high risk. Recent reviews suggested that there is little correlation between dietary indiscretions and the occurrence of travelers' diarrhea. Earlier studies of U.S. students in Mexico showed that eating meals in restaurants and cafeterias or consuming food from street vendors was associated with increased risk.

ETIOLOGY

(See also Table 128-3) The most frequently identified pathogens causing travelers' diarrhea are toxigenic *Escherichia coli* and enteroaggregative *E. coli* (Chap. 149), although in some parts of the world (notably northern Africa and Southeast Asia) *Campylobacter* infections (Chap. 155) appear to predominate. Other common causative organisms include *Salmonella* (Chap. 153), *Shigella* (Chap. 154), rotavirus (Chap. 190), and norovirus (Chap. 190). The latter virus has caused numerous outbreaks on cruise ships. Except for giardiasis (Chap. 215), parasitic infections are uncommon causes of travelers' diarrhea. A growing problem for travelers is the development of antibiotic resistance among many bacterial pathogens. Examples include strains of *Campylobacter* resistant to quinolones and strains of *E. coli*, *Shigella*, and *Salmonella* resistant to trimethoprim-sulfamethoxazole.

PRECAUTIONS

General food and water precautions include eating foods piping hot; avoiding foods that are raw, poorly cooked, or sold by street vendors; and drinking only boiled or commercially bottled beverages, particularly those that are carbonated. Heating kills diarrhea-causing organisms, whereas freezing does not; therefore, ice cubes made from unpurified water should be avoided. In spite of these recommendations, the literature has repeatedly documented dietary indiscretions by 98% of travelers within the first 72 h after arrival at their destination. The maxim "Boil it, cook it, peel it, or forget it!" is easy to remember but apparently difficult to follow.

SELF-TREATMENT

(See also Table 128-5) As travelers' diarrhea often occurs despite rigorous food and water precautions, travelers should carry medications for self-treatment. An antibiotic is useful in reducing the frequency of bowel movements and duration of illness in moderate to severe diarrhea. The standard regimen is a 3-day course of a quinolone taken twice daily (or, in the case of some newer formulations, once daily). However, studies have shown that a single double dose of a quinolone may be equally effective. For diarrhea acquired in areas such as Thailand, where >90% of *Campylobacter* infections are quinolone resistant, azithromycin may be a better alternative. Rifaximin, a poorly absorbed rifampin derivative, is highly effective against noninvasive bacterial pathogens such as toxigenic and enteroaggregative *E. coli*. The current approach to self-treatment of travelers' diarrhea is for the traveler to carry three once-daily doses of an antibiotic and to use as many doses as necessary to resolve the illness. If neither high fever nor blood in the stool accompanies the diarrhea, loperamide should be taken in combination with the antibiotic; studies have shown that this combination is more effective than an antibiotic alone.

PROPHYLAXIS

Prophylaxis of travelers' diarrhea with bismuth subsalicylate is widely used but only ~60% effective. For certain individuals (e.g., athletes, persons with a repeated history of travelers' diarrhea, and persons with chronic diseases), a single daily dose of a quinolone, azithromycin, or rifaximin during travel of <1 month's duration is 75–90% efficacious in preventing travelers' diarrhea. Probiotics have been only ~20% effective as prophylaxis. In Europe and Canada, an oral subunit cholera vaccine that cross-protects against enterotoxigenic *E. coli* has been shown to provide 30–50% protection against travelers' diarrhea.

ILLNESS AFTER RETURN

Although extremely common, acute travelers' diarrhea is usually self-limited or amenable to antibiotic therapy. Persistent bowel problems after the traveler returns home have a less well-defined etiology and may require medical attention from a specialist. Infectious agents (e.g., *Giardia lamblia*, *Cyclospora cayetanensis*, *Entamoeba histolytica*) appear to be responsible for only a small proportion of cases with persistent bowel symptoms. By far the most common causes of persistent diarrhea after travel are postinfectious sequelae such as lactose intolerance or irritable bowel syndrome. A meta-analysis showed that postinfectious irritable bowel syndrome lasting months to years may occur in as many as 4–13% of cases. When no infectious etiology can be identified, a trial of metronidazole therapy for presumed giardiasis, a strict lactose-free diet for 1 week, or a several-week trial of high-dose hydrophilic mucilloid (plus an osmotic laxative such as lactulose or PEG 3350 for persons with alternating diarrhea and constipation) relieves the symptoms of many patients.

Prevention of Other Travel-Related Problems

Travelers are at high risk for *sexually transmitted diseases* (Chap. 130). Surveys have shown that large numbers engage in casual sex, and there is a reluctance to use condoms consistently. An increasing number of travelers are being diagnosed with *schistosomiasis* (Chap. 219). Travelers should be cautioned to avoid bathing, swimming, or wading in freshwater lakes, streams, or rivers in parts of northeastern South America, the Caribbean, Africa, and Southeast Asia. Prevention of *travel-associated injury* depends mostly on common-sense precautions. Riding on motorcycles (especially without helmets) and in overcrowded public vehicles is not recommended; in developing countries, individuals should not travel by road in rural areas after dark. In addition to its association with motor vehicle accidents, excessive alcohol use has been a significant factor in drownings, assaults, and injuries. Travelers are cautioned to avoid walking barefoot because of the risk of hookworm and *Strongyloides*

infections (Chap. 217) and snakebites (Chap. 396).

The Traveler's Medical Kit

A traveler's medical kit is strongly advisable. The contents may vary widely, depending on the itinerary, duration of stay, style of travel, and local medical facilities. While many medications are available abroad (often over the counter), directions for their use may be nonexistent or in a foreign language, or a product may be outdated or counterfeit. For example, a multicountry study in Southeast Asia showed that a mean of 53% (range, 21–92%) of antimalarial products were counterfeit or contained inadequate amounts of active drug. In the medical kit, the short-term traveler should consider carrying an analgesic; an antidiarrheal agent and an antibiotic for self-treatment of travelers' diarrhea; antihistamines; a laxative; oral rehydration salts; a sunscreen with a skin-protection factor of at least 30; a DEET-containing or equivalent insect repellent for the skin; an insecticide for clothing (permethrin); and, if necessary, an antimalarial drug. To these medications, the long-stay traveler might add a broad-spectrum general-purpose antibiotic (levofloxacin or azithromycin), an antibacterial eye and skin ointment, and a topical antifungal cream. Regardless of the duration of travel, a first-aid kit containing such items as scissors, tweezers, and bandages should be considered. A practical approach to self-treatment of infections in the long-stay traveler who carries a once-daily dose of antibiotics (e.g., levofloxacin) is to use 3 tablets "below the waist" (bowel and bladder infections) and 6 tablets "above the waist" (skin and respiratory infections).

TRAVEL AND SPECIAL HOSTS

Pregnancy and Travel

(See also Chap. 7) A woman's medical history and itinerary, the quality of medical care at her destinations, and her degree of flexibility determine whether travel is wise during pregnancy. According to the American College of Obstetrics and Gynecology, the safest part of pregnancy in which to travel is between 18 and 24 weeks, when there is the least danger of spontaneous abortion or premature labor. Some obstetricians prefer that women stay within a few hundred miles of home after the 28th week of pregnancy in case problems arise. In general, however, healthy women may be advised that it is acceptable to travel.

Relative contraindications to international travel during pregnancy include a history of miscarriage, premature labor, incompetent cervix, or toxemia. General medical problems such as diabetes, heart failure, severe anemia, or a history of thromboembolic disease should also prompt the pregnant woman to postpone her travels. Finally, regions in which the pregnant woman and her fetus may be at excessive risk (e.g., those at high altitudes, those where live-virus vaccines are required, and those where multidrug-resistant malaria is endemic) are not ideal destinations during any trimester.

MALARIA

Malaria during pregnancy carries a significant risk of morbidity and death. Levels of parasitemia are highest and failure to clear the parasites after treatment is most frequent among primigravidae. Severe disease, with complications such as cerebral malaria, massive hemolysis, and renal failure, is especially likely in pregnancy. Fetal sequelae include spontaneous abortion, stillbirth, preterm delivery, and congenital infection.

ENTERIC INFECTIONS

Pregnant travelers must be extremely cautious regarding their food and beverage intake. Dehydration due to travelers' diarrhea can lead to inadequate placental blood flow. Infections such as toxoplasmosis, hepatitis E, and listeriosis can also cause serious sequelae in pregnancy.

The mainstay of therapy for travelers' diarrhea is rehydration. Loperamide may be used if necessary. For self-treatment, azithromycin may be the best option. Although quinolones are increasingly being used safely during pregnancy and rifaximin is poorly absorbed from the gastrointestinal tract, these drugs are not approved for this indication.

Because of the serious problems encountered when infants are given local foods and beverages, women are strongly encouraged to breast-feed when traveling with a neonate. A nursing mother with travelers' diarrhea should not stop breast-feeding but should increase her fluid intake.

AIR TRAVEL AND HIGH-ALTITUDE DESTINATIONS

Commercial air travel is not a risk to the healthy pregnant woman or to the fetus. The higher radiation levels reported at altitudes of >10,500 m (>35,000 ft) should pose no problem for the healthy pregnant traveler. Since each airline has a policy regarding pregnancy and flying, it is best to check with the specific carrier when booking reservations. Domestic air travel is usually permitted until the 36th week, whereas international air travel is generally curtailed after the 32nd week.

There are no known risks for pregnant women who travel to high-altitude destinations and stay for short periods. However, there are likewise no data on the safety of pregnant women at altitudes of >4500 m (15,000 ft).

The HIV-Infected Traveler

(See also Chap. 189) The HIV-infected traveler is at special risk of serious infections due to a number of pathogens that may be more prevalent at travel destinations than at home. However, the degree of risk depends primarily on the state of the immune system at the time of travel. For persons whose CD4+ T cell counts are normal or >500/μL, data suggest no greater risk during travel than for persons without HIV infection. Individuals with AIDS (CD4+ T cell counts of <200/μL) and others who are symptomatic need special counseling and should visit a travel medicine practitioner before departure, especially when traveling to the developing world.

Several countries now routinely deny entry to HIV-positive individuals for prolonged stay, even though these restrictions do not appear to decrease rates of transmission of the virus. In general, HIV testing is required for individuals who wish to stay abroad >3 months or who intend to work or study abroad. Some countries will accept an HIV serologic test done within 6 months of departure, whereas others will not accept a blood test done at any time in the traveler's home country. Border officials often have the authority to make inquiries of individuals entering a country and to check the medications they are carrying. If antiretroviral drugs are identified, the person may be barred from entering the country. Information on testing requirements for specific countries is available from consular offices but is subject to frequent change.

IMMUNIZATIONS

All of the HIV-infected traveler's routine immunizations should be up to date (Chap. 122). The response to immunization may be impaired at CD4+ T cell counts of <200/μL and in some cases at even higher counts. Thus HIV-infected persons should be vaccinated as early as possible to ensure adequate immune responses to all vaccines. For patients receiving antiretroviral therapy, at least 3 months must elapse before regenerated CD4+ T cells can be considered fully functional; therefore, vaccination of these patients should be delayed. However, when the risk of illness is high or the sequelae of illness are serious, immunization is recommended. In certain circumstances, it may be prudent to check the adequacy of the serum antibody response before departure.

Because of the increased risk of infections due to *Streptococcus pneumoniae* and other bacterial pathogens that cause pneumonia following influenza, pneumococcal polysaccharide and influenza vaccines should be administered. The estimated rates of response to influenza vaccine are >80% among persons with asymptomatic HIV infection and <50% among those with

AIDS.

In general, live attenuated vaccines are contraindicated for persons with immune dysfunction. Because measles (rubeola) can be a severe or lethal infection in HIV-positive patients, these patients should receive the measles vaccine (or the combination measles-mumps-rubella vaccine) unless the CD4+ T cell count is <200/μL. Between 18% and 58% of symptomatic HIV-infected vaccinees develop adequate antibody titers, and 50–100% of asymptomatic HIV-infected persons seroconvert.

It is recommended that the live yellow fever vaccine not be given to HIV-infected travelers. Although the potential adverse effects of a live vaccine in an HIV-infected individual are always a consideration, there appear to have been no reported cases of illness in those who have inadvertently received this vaccine. Nonetheless, if the CD4+ T cell count is <200/μL, an alternative itinerary that poses no risk of exposure to yellow fever is recommended. If the traveler is passing through or traveling to an area where the vaccine is required but the disease risk is low, a physician's waiver should be issued.

A transient increase in HIV viremia (lasting days to weeks) has been demonstrated in HIV-infected individuals following immunization against influenza, pneumococcal infection, and tetanus (Chap. 189). However, there is no evidence at this point that this transient increase is detrimental.

GASTROINTESTINAL ILLNESS

Decreased levels of gastric acid, abnormal gastrointestinal mucosal immunity, other complications of HIV infection, and medications taken by HIV-infected patients make travelers' diarrhea especially problematic in these individuals. Travelers' diarrhea is likely to occur more frequently, be more severe, be accompanied by bacteremia, and be more difficult to treat. *Cryptosporidium, Isospora belli*, and *Microsporidium* infections, although uncommon, are associated with increased morbidity and mortality rates in AIDS patients.

The HIV-infected traveler must be careful to consume only appropriately prepared foods and beverages and may benefit from antibiotic prophylaxis for travelers' diarrhea. Sulfonamides (as used to prevent pneumocystosis) are ineffective because of widespread resistance.

OTHER TRAVEL-RELATED INFECTIONS

Data are lacking on the severity of many vector-borne diseases in HIV-infected individuals. Malaria is especially severe in asplenic persons and in those with AIDS. The HIV load doubles during malaria, with subsidence in ~8–9 weeks; the significance of this increase in viral load is unknown.

Visceral leishmaniasis (Chap. 212) has been reported in numerous HIV-infected travelers. Diagnosis may be difficult, given that splenomegaly and hyperglobulinemia are often lacking and serologic results are frequently negative. Sandfly bites may be prevented by evening use of insect repellents.

Certain respiratory illnesses, such as histoplasmosis and coccidioidomycosis, cause greater morbidity and mortality among patients with AIDS. Although tuberculosis is common among HIV-infected persons (especially in developing countries), its acquisition by the short-term HIV-infected traveler has not been reported as a major problem. From a prospective study, it is estimated that for nonmedical travelers the risk of tuberculosis infection is ~3% per year of travel.

MEDICATIONS

Adverse events due to medications and drug interactions are common and raise complex issues for HIV-infected persons. Rates of cutaneous reaction (e.g., increased cutaneous sensitivity to sulfonamides) are unusually high among patients with AIDS. Since zidovudine is metabolized by hepatic glucuronidation, inhibitors of this process may elevate serum levels of the drug. Concomitant administration of the antimalarial drug mefloquine and the antiretroviral agent ritonavir may result in decreased plasma levels of ritonavir. In contrast, no significant influence of concomitant mefloquine administration on plasma levels of indinavir or nelfinavir was detected in two HIV-infected travelers. There is a strong theoretical concern that the antimalarial drugs lumefantrine (combined with artemisinin in Coartem) and halofantrine may interact with HIV protease inhibitors and nonnucleoside reverse transcriptase inhibitors since the latter are known to be potent inhibitors of cytochrome P450.

Chronic Illness, Disability, and Travel

Chronic health problems need not prevent travel, but special measures can make the journey safer and more comfortable.

HEART DISEASE

Cardiovascular events are the main cause of deaths among travelers and of in-flight emergencies on commercial aircraft. Extra supplies of all medications should be kept in carry-on luggage, along with a copy of a recent electrocardiogram and the name and telephone number of the traveler's physician at home. Pacemakers are not affected by airport security devices, although electronic telephone checks of pacemaker function cannot be transmitted by international satellites. Travelers with electronic defibrillators should carry a note to that effect and ask for hand screening. A traveler may benefit from supplemental oxygen; since oxygen delivery systems are not standard, supplementary oxygen should be ordered by the traveler's physician well before flight time. Travelers may benefit from aisle seating and should walk, perform stretching and flexing exercises, consider wearing support hose, and remain hydrated during the flight to prevent venous thrombosis and pulmonary embolism.

CHRONIC LUNG DISEASE

Chronic obstructive pulmonary disease is one of the most common diagnoses in patients who require emergency-department evaluation for symptoms occurring during airline flights. The best predictor of the development of in-flight problems is the sea-level Pa_{O2}. A Pa_{O2} of at least 72 mmHg corresponds to an in-flight arterial Pa_{O2} of ~55 mmHg when the cabin is pressurized to 2500 m (8000 ft). If the traveler's baseline Pa_{O2} is <72 mmHg, the provision of supplemental oxygen should be considered.

Contraindications to flight include active bronchospasm, lower respiratory infection, lower-limb deep vein phlebitis, pulmonary hypertension, and recent thoracic surgery (within the preceding 3 weeks) or pneumothorax. Decreased outdoor activity at the destination should be considered if air pollution is excessive.

DIABETES MELLITUS

Alterations in glucose control and changes in insulin requirements are common problems among patients with diabetes who travel. Changes in time zone, in the amount and timing of food intake, and in physical activity demand vigilant assessment of metabolic control. The traveler with diabetes should pack medication (including a bottle of regular insulin for emergencies), insulin syringes and needles, equipment and supplies for glucose monitoring, and snacks in carry-on luggage. Insulin is stable for ~3 months at room temperature but should be kept as cool as possible. The name and telephone number of the home physician and a card and bracelet listing the patient's medical problems and the type and dose of insulin used should accompany the traveler. In traveling eastward (e.g., from the United States to Europe), the morning insulin dose on arrival may need to be decreased. The blood glucose can then be checked during the day to determine whether additional insulin is required. For flights westward, with lengthening of the day, an additional dose of regular insulin may be required.

OTHER SPECIAL GROUPS

Other groups for whom special travel measures are encouraged include patients undergoing dialysis, those with transplants, and those with other disabilities. Up to 13% of travelers have some disability, but few advocacy groups and tour companies dedicate

themselves to this growing population. Medication interactions are a source of serious concern for these travelers, and appropriate medical information should be carried, along with the home physician's name and telephone number. Some travelers taking glucocorticoids carry stress doses in case they become ill. Immunization of these immunocompromised travelers may result in less than adequate protection. Thus the traveler and the physician must carefully consider which destinations are appropriate.

MEDICAL TOURISM

Travel for the purpose of obtaining health care abroad has recently received a great deal of attention in the medical literature and the media. According to the annual U.S. Department of Commerce in-flight survey, there were ~500,000 overseas trips during 2006 in which health treatment was at least one purpose of travel. Lower cost is usually cited as the motivation for this type of tourism, and an entire industry has flourished as a result of this phenomenon. However, the quality of facilities, assistance services, and care is neither uniform nor regulated; thus, in most instances, responsibility for assessing the suitability of an individual program or facility lies solely with the traveler. Persons considering this option must recognize that they are almost always at a disadvantage when being treated in a foreign country, particularly if there are complications. Concerns to be addressed include the quality of the health care facility and its staff; language and cultural differences that may impede accurate interpretation of both verbal and nonverbal communication; religious and ethical differences that may be encountered over issues such as efforts to preserve life and limb or care of the terminally ill; lack of familiarity with the local medical system; limited access of the care provider to the patient's medical history; the use of unfamiliar drugs and medicines; the relative difficulty of arranging follow-up care back in the United States; and the possibility that such follow-up care may be fraught with problems should there be complications. If serious issues arise, legal recourse may be difficult or impossible.

Patients planning to travel abroad to obtain health care, particularly when surgery is involved, should be immunized for hepatitis B and should consider having baseline hepatitis C and HIV tests preoperatively. Prevalence rates of hepatitis B and C and HIV infection vary considerably around the world and are generally higher in developing regions than in the United States and Western Europe. The latest information available on the safety of the blood supply outside the United States is the World Health Organization's Global Database on Blood Safety based on data from 2004–2005 (*www.who.int/bloodsafety/global_database/en*). Persons researching accreditation status of overseas facilities should note that, although these facilities may be part of a chain, they are surveyed and accredited individually. Accreditation resources include (1) the Joint Commission International (*www.jointcommissioninternational.org*), (2) the Australian Council for Healthcare Standards International (*www.achs.org.au/ACHSI*), and (3) the Canadian Council on Health Services (*www.cchsa.ca*).

PROBLEMS AFTER RETURN

The most common medical problems encountered by travelers after their return home are diarrhea, fever, respiratory illnesses, and skin diseases **(Fig. 123-2)**. Frequently ignored problems are fatigue and emotional stress, especially in long-stay travelers. The approach to diagnosis requires some knowledge of geographic medicine, in particular the epidemiology and clinical presentation of infectious disorders. A geographic history should focus on the traveler's exact itinerary, including dates of arrival and departure; exposure history (food indiscretions, drinking-water sources, freshwater contact, sexual activity, animal contact, insect bites); location and style of travel (urban vs. rural, first-class hotel accommodation vs. camping); immunization history; and use of antimalarial chemosuppression.

Figure 123-2

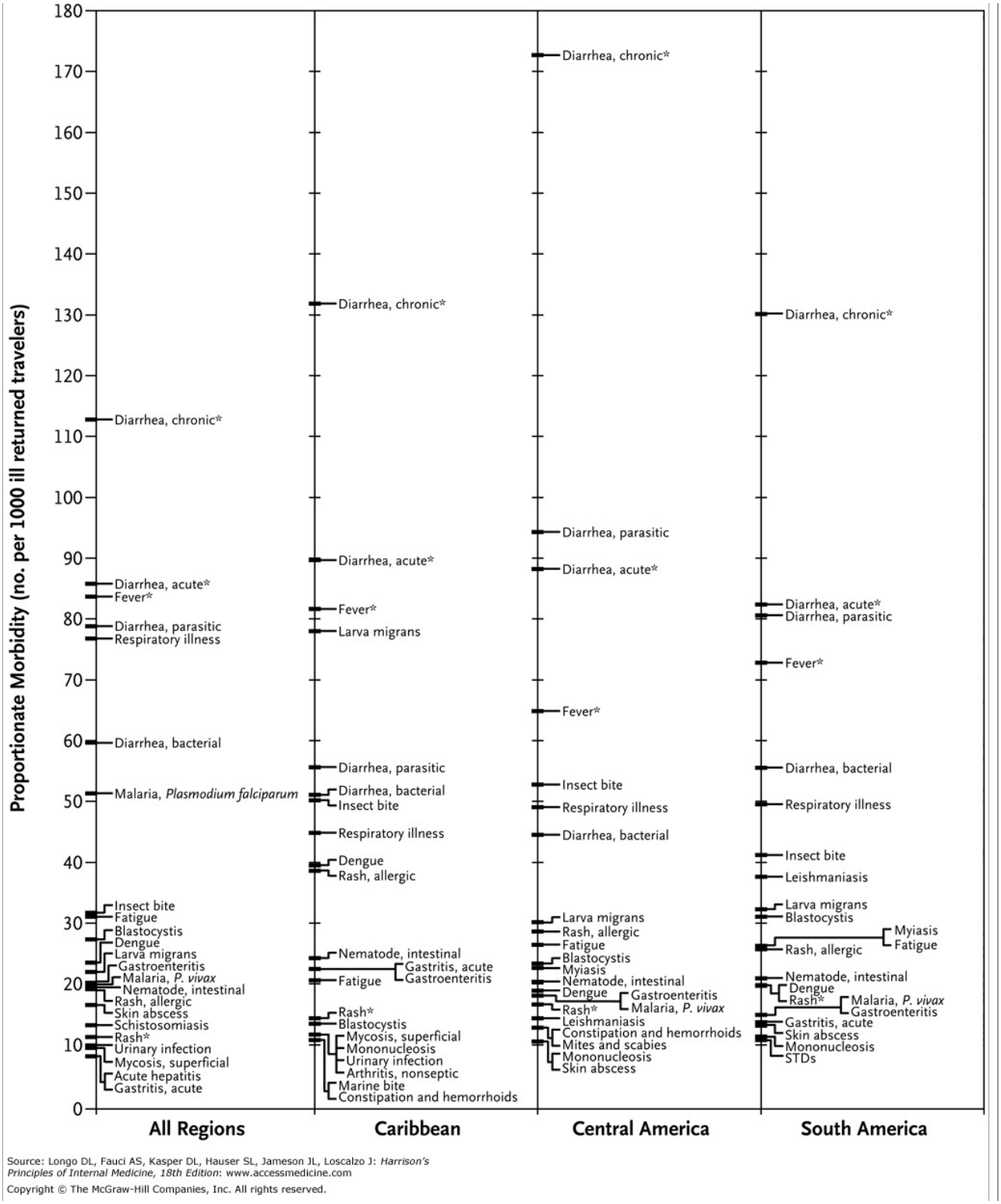

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

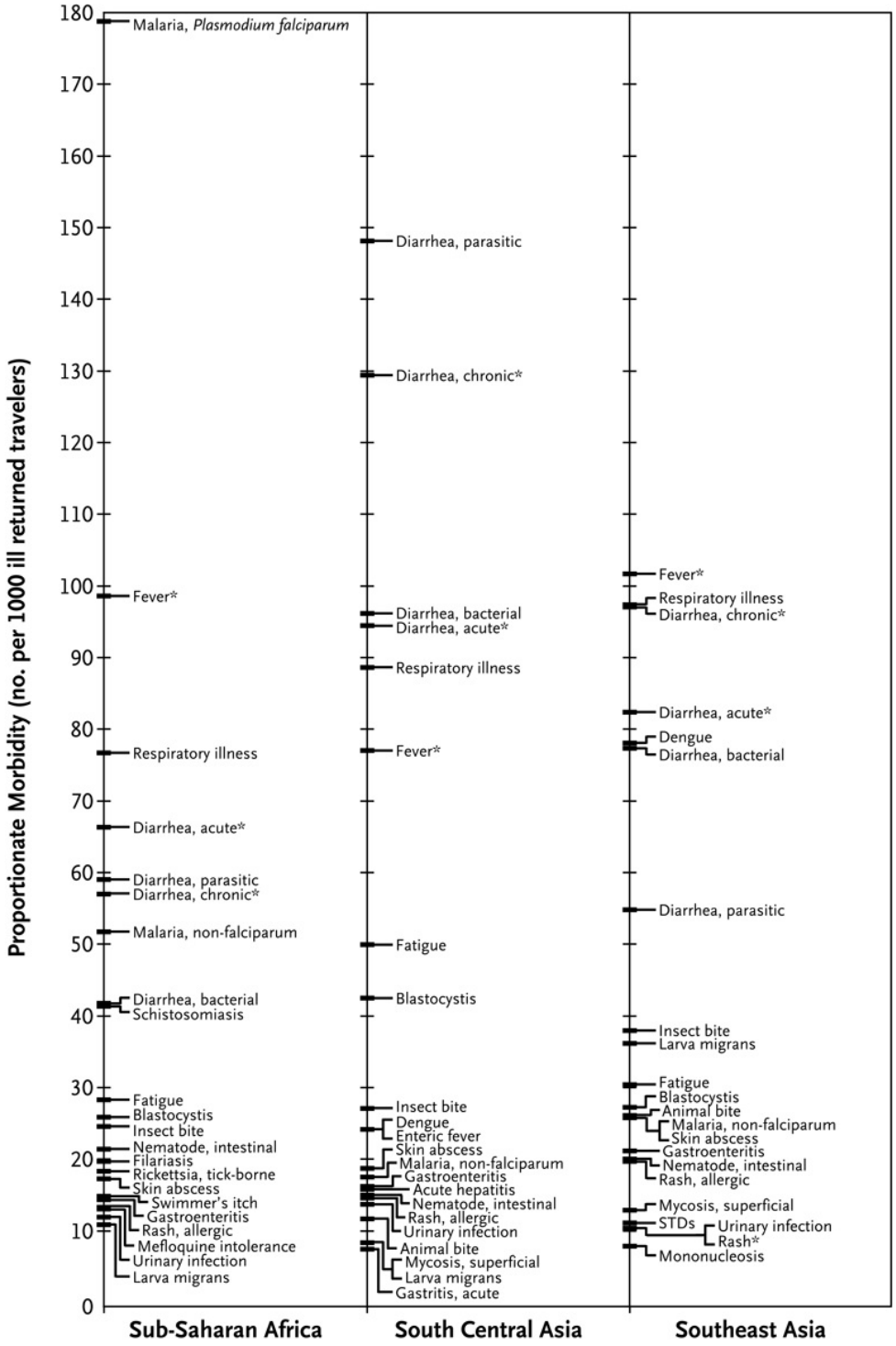

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Proportionate morbidity among ill travelers returning from the developing world, according to region of travel. The proportions (not incidence rates) are shown for each of the top 22 specific diagnoses among all ill returned travelers within each region. STDs, sexually transmitted diseases. Asterisks indicate syndromic diagnoses for which specific etiologies could not be assigned. *(Reprinted with permission from Freedman et al. © 2006 Massachusetts Medical Society.*)

Diarrhea

See "Prevention of Gastrointestinal Illness," above.

Fever

Fever in a traveler who has returned from a malarious area should be considered a medical emergency because death from *P. falciparum* malaria can follow an illness of only several days' duration. Although "fever from the tropics" does not always have a tropical cause, malaria should be the first diagnosis considered. The risk of *P. falciparum* malaria is highest among travelers returning from Africa or Oceania and among those who become symptomatic within the first 2 months after return. Other important causes of fever after travel include viral hepatitis (hepatitis A and E), typhoid fever, bacterial enteritis, arboviral infections (e.g., dengue fever), rickettsial infections (including tick and scrub typhus and Q fever), and—in rare instances—leptospirosis, acute HIV infection, and amebic liver abscess. A cooperative study by GeoSentinel (an emerging infectious disease surveillance group established by the CDC and the International Society of Travel Medicine) showed that, among 3907 febrile returned travelers, malaria was acquired most often from Africa, dengue from Southeast Asia and the Caribbean, typhoid

fever from southern Asia, and rickettsial infections (tick typhus) from southern Africa **(Table 123-3)**. In at least 25% of cases, no etiology can be found, and the fever resolves spontaneously. Clinicians should keep in mind that no present-day antimalarial agent guarantees protection from malaria and that some immunizations (notably, that against typhoid fever) are only partially protective.

Table 123–3. Etiology and Geographic Distribution of Systemic Febrile Illness in Returned Travelers (n = 3907)

	Percentage of Cases					
Etiology	**Carib**	**CAm**	**SAm**	**SSA**	**SCA**	**SEA**
Malaria	<1	13	13	**62**	14	13
Dengue	**23**	12	14	<1	14	**32**
Mononucleosis	7	7	8	1	2	3
Rickettsia	0	0	0	**6**	1	2
Salmonella	2	3	2	<1	**14**	3

Note: Carib, Caribbean; CAm, Central America; SAm, South America; SSA, sub-Saharan Africa; SCA, south-central Asia; SEA, Southeast Asia. Bold type is for emphasis only. ***Source:*** Revised from Table 2 in Freedman et al, 2006. Used with permission from the Massachusetts Medical Society.

When no specific diagnosis is forthcoming, the following investigations, where applicable, are suggested: complete blood count, liver function tests, thick/thin blood films or rapid diagnostic testing for malaria (repeated twice if necessary), urinalysis, urine and blood cultures (repeated once), chest x-ray, and collection of an acute-phase serum sample to be held for subsequent examination along with a paired convalescent-phase serum sample.

Skin Diseases

Pyodermas, sunburn, insect bites, skin ulcers, and cutaneous larva migrans are the most common skin conditions affecting travelers after their return home. In those with persistent skin ulcers, a diagnosis of cutaneous leishmaniasis, mycobacterial infection, or fungal infection should be considered. Careful, complete inspection of the skin is important in detecting the rickettsial eschar in a febrile patient or the central breathing hole in a "boil" due to myiasis.

Emerging Infectious Diseases

In recent years, travel and commerce have fostered the worldwide spread of HIV infection, led to the reemergence of cholera as a global health threat, and created considerable fear about the possible spread of severe acute respiratory syndrome (SARS) and avian influenza (H5N1). For travelers, there are more realistic concerns. One of the largest outbreaks of dengue fever ever documented is now raging in Latin America; chikungunya virus has spread rapidly from Africa to southern Asia and recently to southern Europe; schistosomiasis is being described in previously unaffected lakes in Africa; and antibiotic-resistant strains of sexually transmitted and enteric pathogens are emerging at an alarming rate in the developing world. In addition, concerns have been raised about the potential for bioterrorism involving not only standard strains of unusual agents but mutant strains as well. Time will tell whether travelers (as well as persons at home) will routinely be vaccinated against diseases such as anthrax and smallpox. As Nobel laureate Dr. Joshua Lederberg pointed out, "The microbe that felled one child in a distant continent yesterday can reach yours today and seed a global pandemic tomorrow." The vigilant clinician understands that the importance of a thorough travel history cannot be overemphasized.

FURTHER READINGS

Centers for Disease Control and Prevention: Health Information for International Travel 2010. Mosby, Philadelphia, 2009 (www.cdc.gov/travel)

Chen LH et al: Controversies and misconceptions in malaria chemoprophylaxis for travelers. JAMA 297:2251, 2007[PMID: 17519415]

DuPont HL et al: Expert review of the evidence base for self-therapy of travelers' diarrhea. J Travel Med 16:161, 2009[PMID: 19538576]

——— et al: Expert review of the evidence base for prevention of travelers' diarrhea. J Travel Med 16:149, 2009

Freedman DO: Malaria prevention in short-term travelers. N Engl J Med 359:603, 2008[PMID: 18687641]

——— et al (GeoSentinel Surveillance Network): Spectrum of disease and relation to place of exposure among ill returned travelers. N Engl J Med 354:119, 2006

Johnston V et al: Fever in returned travellers presenting in the United Kingdom: Recommendations for investigation and initial management. J Infect 59:1, 2009[PMID: 19595360]

Keystone JS et al: *Travel Medicine*. Mosby, Philadelphia, 2008

Ryan ET et al: Illness after international travel. N Engl J Med 347:505, 2002[PMID: 12181406]

Sherrard AW, McCarthy AE: Travel patterns and health risks for patients infected with HIV. Travel Med Infect Dis 7:291, 2009[PMID: 19747664]

Sohail MR, Fischer PR: Health risks to air travelers. Infect Dis Clin North Am 19:67, 2005[PMID: 15701547]

Wilson ME, Chen LH: Dermatologic infectious diseases in international travelers. Curr Infect Dis Rep 6:54, 2004[PMID: 14733850]

WEBSITES

Chronic renal failure: www.kidney.org

Diabetes: www.diabetesmonitor.com/other-14.htm

Dialysis: www.dialysisfinder.com

Disability: www.access-able.com

HIV: www.aegis.com

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LABORATORY DIAGNOSIS OF INFECTIOUS DISEASES: INTRODUCTION

The laboratory diagnosis of infection requires the demonstration—either direct or indirect—of viral, bacterial, fungal, or parasitic agents in tissues, fluids, or excreta of the host. Clinical microbiology laboratories are responsible for processing these specimens and also for determining the antibiotic susceptibility of bacterial and fungal pathogens. Traditionally, detection of pathogenic agents has relied largely on either the microscopic visualization of pathogens in clinical material or the growth of microorganisms in the laboratory. Identification generally is based on phenotypic characteristics such as fermentation profiles for bacteria, cytopathic effects in tissue culture for viral agents, and microscopic morphology for fungi and parasites. These techniques are reliable but are often time-consuming. Increasingly, the use of nucleic acid probes is becoming a standard method for detection, quantitation, and/or identification in the clinical microbiology laboratory, gradually replacing phenotypic characterization and microscopic visualization methods.

DETECTION METHODS

Reappraisal of the methods employed in the clinical microbiology laboratory has led to the development of strategies for detection of pathogenic agents through nonvisual biologic signal detection systems. Much of this methodology is based on the use of either electronic detection systems involving relatively inexpensive but sophisticated computers or nucleic acid probes directed at specific DNA or RNA targets. This chapter discusses both the methods that are currently available and those that are being developed.

Biologic Signals

A *biologic signal* is a material that can be reproducibly differentiated from other substances present in the same physical environment. Key issues in the use of a biologic (or electronic) signal are distinguishing it from background noise and translating it into meaningful information. Examples of biologic signals applicable to clinical microbiology include structural components of bacteria, fungi, and viruses; specific antigens; metabolic end products; unique DNA or RNA base sequences; enzymes; toxins or other proteins; and surface polysaccharides.

Detection Systems

A detector is used to sense a signal and discriminate between that signal and background noise. Detection systems range from the trained eyes of a technologist assessing morphologic variations to sensitive electronic instruments such as gas-liquid chromatographs coupled to computer systems for signal analysis. The sensitivity with which signals can be detected varies widely. It is essential to use a detection system that discerns small amounts of signal even when biologic background noise is present—i.e., that is both sensitive and specific. Common detection systems include immunofluorescence; chemiluminescence for DNA/RNA probes; flame ionization detection of short- or long-chain fatty acids; and detection of substrate utilization or end-product formation as color changes, of enzyme activity as a change in light absorbance, of turbidity changes as a measure of growth, of cytopathic effects in cell lines, and of particle agglutination as a measure of antigen presence.

Amplification

Amplification enhances the sensitivity with which weak signals can be detected. The most common microbiologic amplification technique is growth of a single bacterium into a discrete colony on an agar plate or into a suspension containing many identical organisms. The advantage of growth as an amplification method is that it requires only an appropriate growth medium; the disadvantage is the amount of time required. More rapid specific amplification of biologic signals can be achieved with techniques such as polymerase chain reaction (PCR), ligase chain reaction (LCR), and transcription-mediated amplification (TMA), all of which target the pathogen's DNA/RNA; enzyme immunoassays (EIAs, for antigens and antibodies); electronic amplification (for gas-liquid chromatography assays); antibody capture methods (for concentration and/or separation); and selective filtration or centrifugation. Although a variety of methods are available for the amplification and detection of biologic signals in research, thorough testing is required before these methods are validated as diagnostic assays.

DIRECT DETECTION

Microscopy

The field of microbiology has been defined largely by the development and use of the microscope. The examination of specimens by microscopic methods rapidly provides useful diagnostic information. Staining techniques permit organisms to be seen more clearly.

The simplest method for microscopic evaluation is the wet mount, which is used, for example, to examine cerebrospinal fluid (CSF) for the presence of *Cryptococcus neoformans*, with India ink as a background against which to visualize large-capsuled yeast cells. Wet mounts with dark-field illumination also are used to detect spirochetes from genital lesions and to reveal *Borrelia* or *Leptospira* in blood. Skin scrapings and hair samples can be examined with the use of either 10% KOH wet-mount preparations or the Calcofluor white method and ultraviolet illumination to detect fungal elements as fluorescing structures. Staining of wet mounts—e.g., with lactophenol cotton blue stain for fungal elements—often is used for morphologic identification. These techniques enhance signal detection and decrease the background, making it easier to identify specific fungal structures.

Staining

GRAM'S STAIN

Without staining, bacteria are difficult to see at the magnifications (400x to 1000x) used for their detection. Although simple one-step stains can be used, differential stains are more common. Gram's stain differentiates between organisms with thick peptidoglycan cell walls (gram-positive) and those with thin peptidoglycan cell walls and outer membranes that can be dissolved with alcohol or acetone (gram-negative). Cellular morphology and Gram's stain characteristics often can be used to categorize stained organisms into groups such as streptococci, staphylococci, and clostridia **(Fig. e22-1)**.

Figure e22-1

Gram-Negative Organisms

	GRx only	Oxidase +	Oxidase –	Fastidious	Anaerobic	Curved
Rod		*Pseudomonas* *Aeromonas* *Pasteurella* Others	Enterobacteriaceae Others	*Haemophilus* *Legionella* *Bordetella* *Brucella* *Francisella* Others	*Bacteroides* *Prevotella* *Fusobacterium* Others	*Vibrio* *Campylobacter*
Coccus	*Neisseria* *Branhamella*				*Veillonella* *Acidaminococcus* *Megasphaera*	

Gram-Positive Organisms

	Branching	Spores	Acid-Fast	Catalase +	Catalase –
Rod	*Nocardia* *Actinomyces* *Bifidobacterium*	*Clostridium* *Bacillus*	*Mycobacterium*	*Corynebacterium* *Listeria* Others	*Lactobacillus* Others
Coccus				*Staphylococcus* *Micrococcus* Others	*Streptococcus*

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Interpretation of Gram's stain.

Gram's stain is particularly useful for examining sputum for polymorphonuclear leukocytes (PMNs) and bacteria. Sputum specimens from immunocompetent patients with ≥25 PMNs and <10 epithelial cells per low-power field often provide clinically useful information. However, the presence in "sputum" samples of >10 epithelial cells per low-power field and of multiple bacterial types suggests contamination with oral microflora. Despite the difficulty of discriminating between normal microflora and pathogens, Gram's stain may prove useful for specimens from areas with a large resident microflora if a useful biologic marker (signal) is available. Gram's staining of vaginal swab specimens can be used to detect epithelial cells covered with gram-positive bacteria in the absence of lactobacilli and the presence of gram-negative rods—a scenario regarded as a sign of bacterial vaginosis. Similarly, examination of stained stool specimens for leukocytes is useful as a screening procedure before testing for *Clostridium difficile* toxin or other enteric pathogens.

The examination of CSF and joint, pleural, or peritoneal fluid with Gram's stain is useful for determining whether bacteria and/or PMNs are present. The sensitivity is such that $>10^4$ bacteria per milliliter should be detected. Centrifugation often is performed before staining to concentrate specimens thought to contain low numbers of organisms. The pellet is examined after staining. This simple method is particularly useful for examination of CSF for bacteria and white blood cells or examination of sputum for mycobacteria.

ACID-FAST STAIN

The acid-fast stain identifies organisms that retain carbol fuchsin dye after acid/organic solvent disruption (e.g., *Mycobacterium* spp.). Modifications of this procedure allow the differentiation of *Actinomyces* from *Nocardia* or other weakly (or partially) acid-fast organisms. The acid-fast stain is applied to sputum, other fluids, and tissue samples when acid-fast bacilli (AFB, e.g., *Mycobacterium* species) are suspected. Because few AFB may be detected in an entire smear, even when the specimen has been concentrated by centrifugation, identification of the pink/red AFB against the blue background of the counterstain requires a trained eye. An alternative method is the auramine-rhodamine combination fluorescent dye technique.

FLUOROCHROME STAINS

Fluorochrome stains such as acridine orange are used to identify white blood cells, yeasts, and bacteria in body fluids. Other specialized stains, such as Dappe's stain, may be used for the detection of mycoplasmas in cell cultures. Capsular, flagellar, and spore stains are used for identification or demonstration of characteristic structures.

IMMUNOFLUORESCENT STAINS

The *direct* immunofluorescent antibody technique uses antibody coupled to a fluorescent compound (e.g., fluorescein) and directed at a specific antigenic target to visualize organisms or subcellular structures. When samples are examined under appropriate conditions, the fluorescing compound absorbs ultraviolet light and reemits light at a higher wavelength that is visible to the human eye. In the *indirect* immunofluorescent antibody technique, an unlabeled (target) antibody binds a specific antigen. The specimen is then stained with fluorescein-labeled polyclonal antibody directed at the target antibody. Because each unlabeled target antibody attached to the appropriate antigen has multiple sites for attachment of the second antibody, the visual signal can be intensified (i.e., amplified). This form of staining is called *indirect* because a two-antibody system is used to generate the signal for detection of the antigen. Both direct and indirect methods detect viral antigens (e.g., cytomegalovirus, herpes simplex virus, and respiratory viruses) within cultured cells or clinical specimens as well as many difficult-to-grow bacterial agents (e.g., *Legionella pneumophila*) in clinical specimens.

Macroscopic Antigen Detection

Latex agglutination assays and EIAs are rapid and inexpensive methods for identifying organisms, extracellular toxins, and viral agents by means of protein and polysaccharide antigens. Such assays may be performed directly on clinical samples or after growth of organisms on agar plates or in viral cell cultures. The biologic signal in each case is the antigen to be detected. Monoclonal or polyclonal antibodies coupled to a reporter (such as latex particles or an enzyme) are used for detection of antibody-antigen binding reactions.

Techniques such as direct agglutination of bacterial cells with specific antibody are simple but relatively insensitive; latex agglutination and EIAs are more sensitive. Some cell-associated antigens, such as capsular polysaccharides and lipopolysaccharides, can be detected by agglutination of a suspension of bacterial cells when antibody is added; this method is useful for typing of the somatic antigens of *Shigella* and *Salmonella*. In systems such as EIAs, which employ antibodies coupled to an enzyme, an antigen-antibody reaction results in the conversion of a colorless substrate to a colored product. Because the coupling of an enzyme to the antibody can amplify a weak biologic signal, the sensitivity of such assays is often high. In each instance, the basis for antigen detection is antigen-antibody binding, with the detection system changed to accommodate the biologic signal. Most of these assays provide information about whether antigen is present but do not

quantify the antigen. EIAs are also useful for detecting bacterial toxins—e.g., *C. difficile* toxins A and B in stool.

Rapid and simple immunoassays for antigens of group A *Streptococcus*, influenza virus, and respiratory syncytial virus can be used in the clinical setting without a specialized diagnostic laboratory. Such tests are usually reasonably specific but may have only modest sensitivity.

DETECTION OF PATHOGENIC AGENTS BY CULTURE

Specimen Collection and Transport

To culture bacterial, fungal, or viral pathogens, an appropriate sample must be placed into the proper medium for growth (amplification). The success of efforts to identify a specific pathogen often depends on the collection and transport process coupled to a laboratory-processing algorithm suitable for the specific sample/agent. In some instances, it is better for specimens to be plated at the time of collection rather than first being transported to the laboratory (e.g., urethral swabs being cultured for *Neisseria gonorrhoeae* or sputum specimens for pneumococci). In general, the more rapidly a specimen is plated onto appropriate media, the better the chance is for isolating bacterial pathogens. Deep tissue or fluid (pus) samples are more likely to give useful culture results than are superficial swab specimens. Table e22-1 lists procedures for collection and transport of common specimens. Because there are many pathogen-specific paradigms for these procedures, it is important to seek advice from the microbiology laboratory when in doubt about a particular situation.

Table e22-1 Instructions for Collection and Transport of Specimens for Culture

Type of Culture (Synonyms)	Specimen	Minimal Volume	Container	Other Considerations
Blood				
Blood, routine (blood culture for aerobes, anaerobes, and yeasts)	Whole blood	10 mL in each of 2 bottles for adults and children; 5 mL, if possible, in aerobic bottles for infants; less for neonates	See below.[a]	See below.[b]
Blood for fungi *Mycobacterium* spp.	Whole blood	10 mL in each of 2 bottles, as for routine blood cultures, or in Isolator tube requested from laboratory	Same as for routine blood culture	Specify "hold for extended incubation," since fungal agents may require 4 weeks to grow.
Blood, Isolator (lysis centrifugation)	Whole blood	10 mL	Isolator tubes	Use mainly for isolation of fungi, *Mycobacterium*, and other fastidious aerobes and for elimination of antibiotics from cultured blood in which organisms are concentrated by centrifugation.
Respiratory tract				
Nose	Swab from nares	1 swab	Sterile culturette or similar transport system containing holding medium	Swabs made of calcium alginate may be used.
Throat	Swab of posterior pharynx, ulcerations, or areas of suspected purulence	1 swab	Sterile culturette or similar swab specimen collection system containing holding medium	See below.[c]
Sputum	Fresh sputum (not saliva)	2 mL	Commercially available sputum collection system or similar sterile container with screw cap	*Cause for rejection:* Care must be taken to ensure that the specimen is sputum and not saliva. Examination of Gram's stain, with number of epithelial cells and polymorphonuclear leukocytes (PMNs) noted, can be an important part of the evaluation process. Induced sputum specimens should not be rejected.
Bronchial aspirates	Transtracheal aspirate, bronchoscopy specimen, or bronchial aspirate	1 mL of aspirate or brush in transport medium	Sterile aspirate or bronchoscopy tube, bronchoscopy brush in a separate sterile container	Special precautions may be required, depending on diagnostic considerations (e.g., *Pneumocystis*).
Stool				
Stool for routine culture; stool for *Salmonella, Shigella*, and *Campylobacter*	Rectal swab or (preferably) fresh, randomly collected stool	1 g of stool or 2 rectal swabs	Plastic-coated cardboard cup or plastic cup with tight-fitting lid. Other leakproof containers are also acceptable.	If *Vibrio* spp. are suspected, the laboratory must be notified, and appropriate collection/transport methods should be used.
Stool for *Yersinia, Escherichia coli* O157	Fresh, randomly collected stool	1 g	Plastic-coated cardboard cup or plastic cup with tight-fitting lid	*Limitations:* Procedure requires enrichment techniques.
Stool for *Aeromonas* and *Plesiomonas*	Fresh, randomly collected stool	1 g	Plastic-coated cardboard cup or plastic cup with tight-fitting lid	*Limitations:* Stool should not be cultured for these organisms unless also cultured for other enteric pathogens.
Urogenital tract				

Urine	Clean-voided urine specimen or urine collected by catheter	0.5 mL	Sterile, leak-proof container with screw cap or special urine transfer tube	See below.[d]
Urogenital secretions	Vaginal or urethral secretions, cervical swabs, uterine fluid, prostatic fluid, etc.	1 swab or 0.5 mL of fluid	Vaginal and rectal swabs transported in Amies transport medium or similar holding medium for group B *Streptococcus;* direct inoculation preferred for *Neisseria gonorrhoeae*	Vaginal swab samples for "routine culture" should be discouraged whenever possible unless a particular pathogen is suspected. For detection of multiple organisms (e.g., group B *Streptococcus, Trichomonas, Chlamydia*, or *Candida* spp.), 1 swab per test should be obtained.
Body fluids, aspirates, and tissues				
Cerebrospinal fluid (lumbar puncture)	Spinal fluid	1 mL for routine cultures; ≥5 mL for *Mycobacterium*	Sterile tube with tight-fitting cap	Do not refrigerate; transfer to laboratory as soon as possible.
Body fluids	Aseptically aspirated body fluids	1 mL for routine cultures	Sterile tube with tight-fitting cap. Specimen may be left in syringe used for collection if the syringe is capped before transport.	For some body fluids (e.g., peritoneal lavage samples), increased volumes are helpful for isolation of small numbers of bacteria.
Biopsy and aspirated materials	Tissue removed at surgery, bone, anticoagulated bone marrow, biopsy samples, or other specimens from normally sterile areas	1 mL of fluid or a 1-g piece of tissue	Sterile "culturette"-type swab or similar transport system containing holding medium. Sterile bottle or jar should be used for tissue specimens.	Accurate identification of specimen and source is critical. Enough tissue should be collected for both microbiologic and histopathologic evaluations.
Wounds	Purulent material or abscess contents obtained from wound or abscess without contamination by normal microflora	2 swabs or 0.5 mL of aspirated pus	Culturette swab or similar transport system or sterile tube with tight-fitting screw cap. For simultaneous anaerobic cultures, send specimen in anaerobic transport device or closed syringe.	*Collection:* When possible, abscess contents or other fluids should be collected in a syringe (rather than with a swab) to provide an adequate sample volume and an anaerobic environment.
Special recommendations				
Fungi	Specimen types listed above may be used. When urine or sputum is cultured for fungi, a first morning specimen usually is preferred.	1 mL or as specified above for individual listing of specimens. Large volumes may be useful for urinary fungi.	Sterile, leakproof container with tight-fitting cap	*Collection:* Specimen should be transported to microbiology laboratory within 1 h of collection. Contamination with normal flora from skin, rectum, vaginal tract, or other body surfaces should be avoided.
Mycobacterium (acid-fast bacilli)	Sputum, tissue, urine, body fluids	10 mL of fluid or small piece of tissue. Swabs should not be used.	Sterile container with tight-fitting cap	Detection of *Mycobacterium* spp. is improved by use of concentration techniques. Smears and cultures of pleural, peritoneal, and pericardial fluids often have low yields. Multiple cultures from the same patient are encouraged. Culturing in liquid media shortens time to detection.
Legionella	Pleural fluid, lung biopsy, bronchoalveolar lavage fluid, bronchial/transbronchial biopsy. Rapid transport to laboratory is critical.	1 mL of fluid; any size tissue sample, although a 0.5-g sample should be obtained when possible	—	—
Anaerobic organisms	Aspirated specimens from abscesses or body fluids	1 mL of aspirated fluid, 1 g of tissue, or 2 swabs	An appropriate anaerobic transport device is required.[e]	Specimens cultured for obligate anaerobes should be cultured for facultative bacteria as well. Fluid or tissue is preferred to swabs.
Viruses[f]	Respiratory secretions, wash aspirates from respiratory tract, nasal swabs, blood samples (including buffy coats), vaginal and rectal swabs, swab specimens from suspicious skin lesions, stool samples (in some cases)	1 mL of fluid, 1 swab, or 1 g of stool in each appropriate transport medium	Fluid or stool samples in sterile containers or swab samples in viral culturette devices (kept on ice but not frozen) are generally suitable. Plasma samples and buffy coats in sterile collection tubes should be kept at 4–8°C. If specimens are to be shipped or kept for a long time, freezing at −80°C is usually adequate.	Most samples for culture are transported in holding medium containing antibiotics to prevent bacterial overgrowth and viral inactivation. Many specimens should be kept cool but not frozen, provided they are transported promptly to the laboratory. Procedures and transport media vary with the agent to be cultured and the duration of transport.

Note: It is absolutely essential that the microbiology laboratory be informed of the site of origin of the sample to be cultured and the infections that are suspected. This information determines the selection of culture media and the length of culture time.

[a]For samples from adults, two bottles (smaller for pediatric samples) should be used: one with dextrose phosphate, tryptic soy, or another appropriate broth and the other with thioglycollate or another broth containing reducing agents appropriate for isolation of obligate anaerobes. For children, from whom only limited volumes of blood can be obtained, only an aerobic culture should be done unless there is specific concern about anaerobic sepsis (e.g., with abdominal infections). For special situations

(e.g., suspected fungal infection, culture-negative endocarditis, or mycobacteremia), different blood collection systems may be used (Isolator systems; see table).

[b]Collection: An appropriate disinfecting technique should be used on both the bottle septum and the patient. Do not allow air bubbles to get into anaerobic broth bottles. Special considerations: There is no more important clinical microbiology test than the detection of bloodborne pathogens. The rapid identification of bacterial and fungal agents is a major determinant of patients' survival. Bacteria may be present in blood either continuously (as in endocarditis, overwhelming sepsis, and the early stages of salmonellosis and brucellosis) or intermittently (as in most other bacterial infections, in which bacteria are shed into the blood on a sporadic basis). Most blood culture systems employ two separate bottles containing broth medium: one that is vented in the laboratory for the growth of facultative and aerobic organisms and one that is maintained under anaerobic conditions. In cases of suspected continuous bacteremia/fungemia, two or three samples should be drawn before the start of therapy, with additional sets obtained if fastidious organisms are thought to be involved. For intermittent bacteremia, two or three samples should be obtained at least 1 h apart during the first 24 h.

[c]Normal microflora includes α-hemolytic streptococci, saprophytic Neisseria spp., diphtheroids, and Staphylococcus spp. Aerobic culture of the throat ("routine") includes screening for and identification of β-hemolytic Streptococcus spp. and other potentially pathogenic organisms. Although considered components of the normal microflora, organisms such as Staphylococcus aureus, Haemophilus influenzae, and Streptococcus pneumoniae will be identified by most laboratories, if requested. When Neisseria gonorrhoeae or Corynebacterium diphtheriae is suspected, a special culture request is recommended.

[d](1) Clean-voided specimens, midvoid specimens, and Foley or indwelling catheter specimens that yield 50,000 organisms/mL and from which no more than three species are isolated should have organisms identified. Neither indwelling catheter tips nor urine from the bag of a catheterized patient should be cultured. (2) Straight-catheterized, bladder-tap, and similar urine specimens should undergo a complete workup (identification and susceptibility testing) for all potentially pathogenic organisms regardless of colony count. (3) Certain clinical problems (e.g., acute dysuria in women) may warrant identification and susceptibility testing of isolates present at concentrations of <50,000 organisms/mL.

[e]Aspirated specimens in capped syringes or other transport devices designed to limit oxygen exposure are suitable for the cultivation of obligate anaerobes. A variety of commercially available transport devices may be used. Contamination of specimens with normal microflora from the skin, rectum, vaginal vault, or another body site should be avoided. Collection containers for aerobic culture (such as dry swabs) and inappropriate specimens (such as refrigerated samples; expectorated sputum; stool; gastric aspirates; and vaginal, throat, nose, and rectal swabs) should be rejected as unsuitable.

[f]Laboratories generally use diverse methods to detect viral agents, and the specific requirements for each specimen should be checked before a sample is sent.

Isolation of Bacterial Pathogens

Isolation of suspect pathogens from clinical material relies on the use of artificial media that support bacterial growth in vitro. Such media are composed of agar, which is not metabolized by bacteria; nutrients to support the growth of the species of interest; and sometimes substances to inhibit the growth of other bacteria. Broth is employed for growth (amplification) of organisms from specimens with few bacteria, such as peritoneal dialysis fluid, CSF, or samples in which anaerobes or other fastidious organisms may be present. The general use of liquid medium for all specimens is not worthwhile.

Two basic strategies are used to isolate pathogenic bacteria. The first is to employ enriched media that support the growth of any bacteria that may be present in a sample, such as blood or CSF, that contains no bacteria under normal conditions. Broths that allow the growth of small numbers of organisms may be subcultured to solid media when growth is detected. The second strategy is to use selective media to isolate (amplify) specific bacterial species from stool, genital tract secretions, or sputum —sites that contain many bacteria under normal conditions. Antimicrobial agents or other inhibitory substances are incorporated into the agar medium to inhibit growth of all but the bacteria of interest. After incubation, organisms that grow on such media are characterized further to determine whether they are pathogens. Selection for organisms that may be pathogens from the normal microflora shortens the time required for diagnosis (**Fig. e22-2**).

Figure e22-2

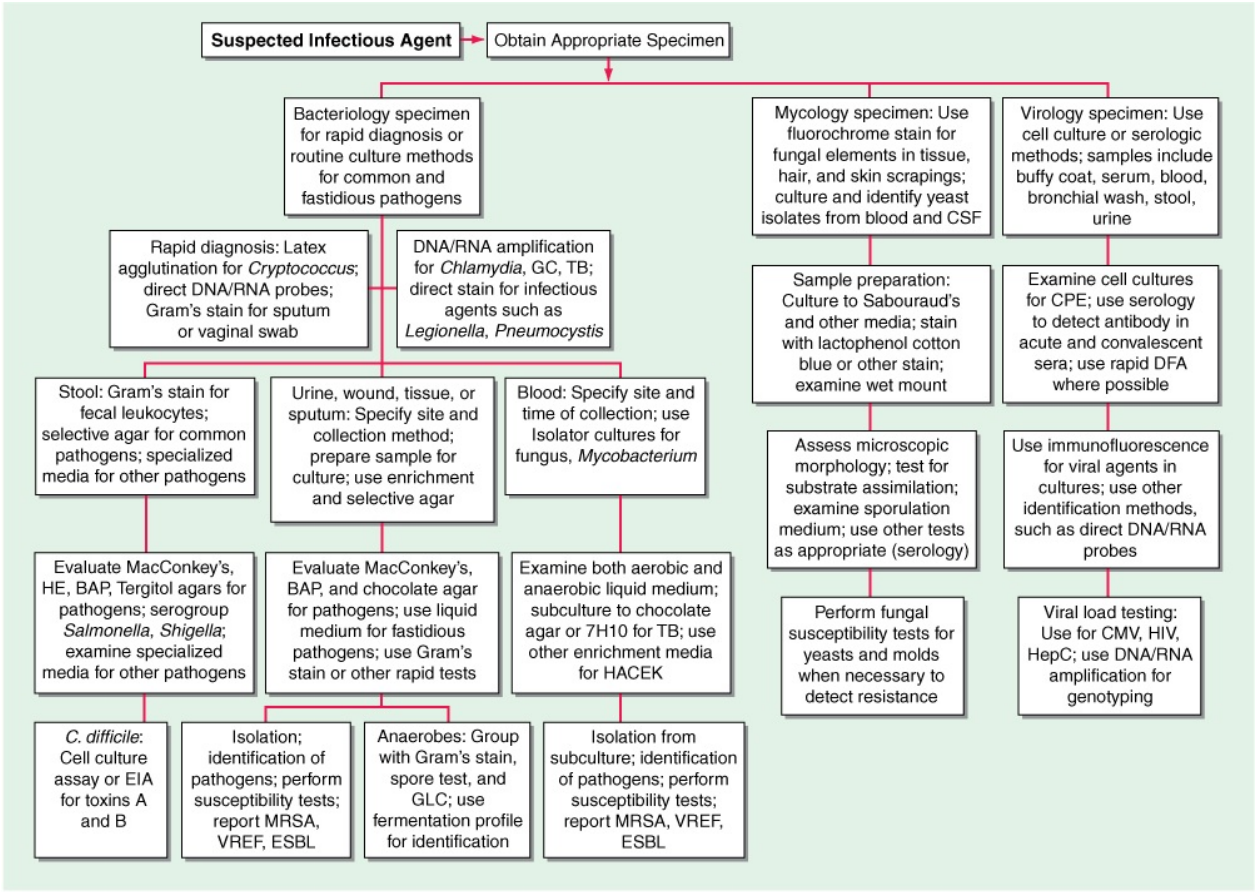

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Common specimen-processing algorithms used in clinical microbiology laboratories. BAP, blood agar plate; CMV, cytomegalovirus; CPE, cytopathic effects; CSF, cerebrospinal fluid; DFA, direct fluorescent antibody; EIA, enzyme immunoassay; ESBL, extended-spectrum β-lactamase; GBS, group B *Streptococcus*; GC, *Neisseria gonorrhoeae*; GLC, gas-liquid chromatography; HACEK, *Haemophilus aphrophilus/parainfluenzae/paraphrophilus*, *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella kingae*; HE, Hektoen enteric medium; HepC, hepatitis C virus; HIV, human immunodeficiency virus; MRSA, methicillin-resistant *Staphylococcus aureus*; TB, *Mycobacterium tuberculosis*; VREF, vancomycin-resistant *Enterococcus faecium*.

Isolation of Viral Agents

(See also Chap. 177) Pathogenic viral agents are often sought by culture when the presence of serum antibody is not a criterion for active infection, when serologic diagnosis is not practical, or when immunoassays have inadequate sensitivity. The biologic signal, virus, is amplified to a detectable level. Although a number of techniques for viral culture are available, an essential element is a monolayer of cultured mammalian cells sensitive to infection with the suspected virus. These cells serve as the amplification system by allowing the proliferation of viral particles. Virus may be detected by direct observation of the cultured cells for cytopathic effects or by immunofluorescent detection of viral antigens after incubation. Conventional viral culture is useful for detection of rapidly propagated agents, such as herpes simplex virus. Viruses that grow more slowly (e.g., cytomegalovirus and varicella-zoster virus) can be detected quickly by shell-vial culture, in which the specimen is centrifuged on a monolayer of cells that is then incubated for 1–2 days and finally is stained for viral antigens with fluorochrome-conjugated antibodies.

AUTOMATION OF MICROBIAL DETECTION IN BLOOD

The detection of microbial pathogens in blood is difficult because the number of organisms present in the sample is often low and the organisms' integrity and ability to replicate may be damaged by humoral defense mechanisms or antimicrobial agents. Over the years, systems that rely on the detection of gas (usually CO_2) produced by bacteria and yeasts in blood culture medium have allowed the automation of the detection procedure. The most common systems involve either (1) the measurement of gas pressure in the headspace to indicate bacterial gas production or consumption or (2) the use of reflectance optics, with a light-emitting diode and photodiode employed to detect a color change in a CO_2-sensitive indicator built into the bottom of the culture bottle. These systems measure CO_2 concentration as indicative of microbial growth. Such methods are no more sensitive than the human eye in detecting a positive culture; however, because the bottles in an automated system are monitored more frequently, a positive culture often is detected more rapidly than by manual techniques, and important information, including the result of Gram's stain and preliminary susceptibility assays, can be obtained sooner. One advantage of automated blood culture systems is that the bottles are scanned continuously in a noninvasive monitoring procedure, and thus the likelihood of laboratory contamination is decreased.

Several factors affect the yield of blood culture from bacteremic patients. Increasing the volume of blood tested increases the chance of a positive culture. An increase from 10 to 20 mL of blood increases the proportion of positive cultures by ~30%; however, this effect is less pronounced in patients with bacterial endocarditis. Obtaining multiple cultures (up to three per 24-h period) also increases the chance of detecting a bacterial pathogen. Prolonged culture and blind subculture for detection of most fastidious bacteria (e.g., *Haemophilus*, *Actinobacillus*, *Cardiobacterium*, *Eikenella*, and *Kingella* species) are not needed with automated blood culture systems.

Automated systems also have been applied to the detection of microbial growth from specimens other than blood, such as peritoneal and other normally sterile fluids. *Mycobacterium* species can be detected in certain automated systems if

appropriate liquid media are used for culture. Although automated blood culture systems are more sensitive than lysis-centrifugation methods (e.g., Isolator) for yeasts and most bacteria, lysis-centrifugation culture is recommended for filamentous fungi, *Histoplasma capsulatum*, and some fastidious bacteria (*Legionella* and *Bartonella*).

DETECTION OF PATHOGENIC AGENTS BY SEROLOGIC METHODS

Measurement of serum antibody provides an indirect marker for past or current infection with a specific viral agent or other pathogens, including *Brucella*, *Legionella*, *Rickettsia*, and *Helicobacter pylori* . The biologic signal is usually either IgM or IgG antibody directed at surface-expressed antigen. The detection systems include those used for bacterial antigens (agglutination reactions, immunofluorescence, and EIA) and unique systems such as hemolysis inhibition and complement fixation. Serologic methods generally fall into two categories: those that determine protective antibody levels and those that measure changing antibody titers during infection. Determination of an antibody response as a measure of current immunity is important in the case of viral agents for which there are vaccines, such as rubella virus and varicella-zoster virus; assays for this purpose normally use one or two dilutions of serum for a qualitative determination of protective antibody levels. Quantitative serologic assays to detect increases in antibody titers most often employ paired serum samples obtained at the onset of illness and 10 –14 days later (i.e., acute- and convalescent-phase samples). Since the incubation period before symptoms are noted may be long enough for an antibody response to occur, the demonstration of acute-phase antibody alone is often insufficient to establish the diagnosis of active infection as opposed to past exposure. In such circumstances, IgM may be useful as a measure of an early, acute-phase antibody response. A fourfold increase in total antibody titer or in EIA activity between the acute- and convalescent-phase samples is also regarded as evidence for active infection.

For certain viral agents, such as Epstein-Barr virus, the antibodies produced may be directed at different antigens during different phases of the infection. For this reason, most laboratories test for antibody directed at both viral capsid antigens and antigens associated with recently infected host cells to determine the stage of infection.

IDENTIFICATION METHODS

Once bacteria are isolated, characteristics that are readily detectable after growth on agar media (colony size, color, hemolytic reactions, odor, microscopic appearance) may suggest a species, but definitive identification requires additional tests. Identification methods include classic biochemical phenotyping, which is still the most common approach, and more sophisticated methods such as gas chromatography and nucleic acid tests.

Classic Phenotyping

Classic phenotypic identification of bacteria entails tests for protein or carbohydrate antigens, the production of specific enzymes, the ability to metabolize specific substrates and carbon sources (such as carbohydrates), or the production of certain metabolites. Rapid versions of some of these tests are available, and many common organisms can be identified on the first day of growth. Other organisms, particularly gram-negative bacteria, require more extensive testing, either manual or automated.

Automated systems allow rapid phenotypic identification of bacterial pathogens. Most of these systems are based on biotyping techniques in which isolates are grown on multiple substrates and the reaction pattern is compared with known patterns for various bacterial species. This procedure is relatively fast, and commercially available systems include miniaturized fermentation, coding to simplify recording of results, and probability calculations for the most likely pathogens. If the biotyping approach is automated and the reading process is coupled to computer-based data analysis, rapidly growing organisms (such as Enterobacteriaceae) can be identified within hours of detection on agar plates.

Several systems use preformed enzymes for even speedier identification (within 2–3 h). Those systems do not rely on bacterial growth per se to determine whether a substrate has been used. They employ a heavy inoculum in which specific bacterial enzymes are present in amounts sufficient to convert substrate to product rapidly. In addition, some systems use fluorogenic substrate/end-product detection methods to increase sensitivity (through signal amplification).

Gas-Liquid Chromatography

Gas-liquid chromatography often is used to detect metabolic end products of bacterial fermentations. One common application is identification of short-chain fatty acids produced by obligate anaerobes during glucose fermentation. Because the types and relative concentrations of volatile acids differ among the various genera and species that make up this group of organisms, such information serves as a metabolic "fingerprint" for a particular isolate.

Gas-liquid chromatography can be coupled to a sophisticated signal-analysis software system for identification and quantitation of long-chain fatty acids (LCFAs) in the outer membranes and cell walls of bacteria and fungi. For any particular species, the types and relative concentrations of LCFAs are distinctive enough to allow differentiation even from closely related species. An organism may be identified definitively within a few hours after detection of growth on appropriate media. LCFA analysis is one of the most advanced procedures currently available for phenotypic characterization.

Nucleic Acid Tests

Techniques for the detection and quantitation of specific DNA and RNA base sequences in clinical specimens have become powerful tools for the diagnosis of bacterial, viral, parasitic, and fungal infections. Nucleic acid tests are used for four purposes. First, they are used to detect, and sometimes to quantify, specific pathogens in clinical specimens. Second, such tests are used for identification of organisms (usually bacteria) that are difficult to identify by conventional methods. Third, nucleic acid tests are used to determine whether two or more isolates of the same pathogen are closely related (belonging to the same "clone" or "strain"). Fourth, these tests are used to predict the sensitivity of organisms (typically viruses) to chemotherapeutic agents. Current technology encompasses a wide array of methods for amplification and signal detection, some of which have been approved by the U.S. Food and Drug Administration (FDA) for clinical diagnosis.

Use of nucleic acid tests generally involves lysis of intact cells or viruses and denaturation of the DNA or RNA to render it single-stranded. Probe(s) or primer(s) complementary to the pathogen-specific target sequence may be hybridized to the target sequence in a solution or on a solid support, depending on the system employed. In situ hybridization of a probe to a target is also possible and allows the use of probes with agents present in tissue specimens. Once the probe(s) or primer(s) have been hybridized to the target (biologic signal), a variety of strategies may be employed to detect, amplify, and/or quantify the target-probe complex **(Fig. e22-3)**.

Figure e22-3

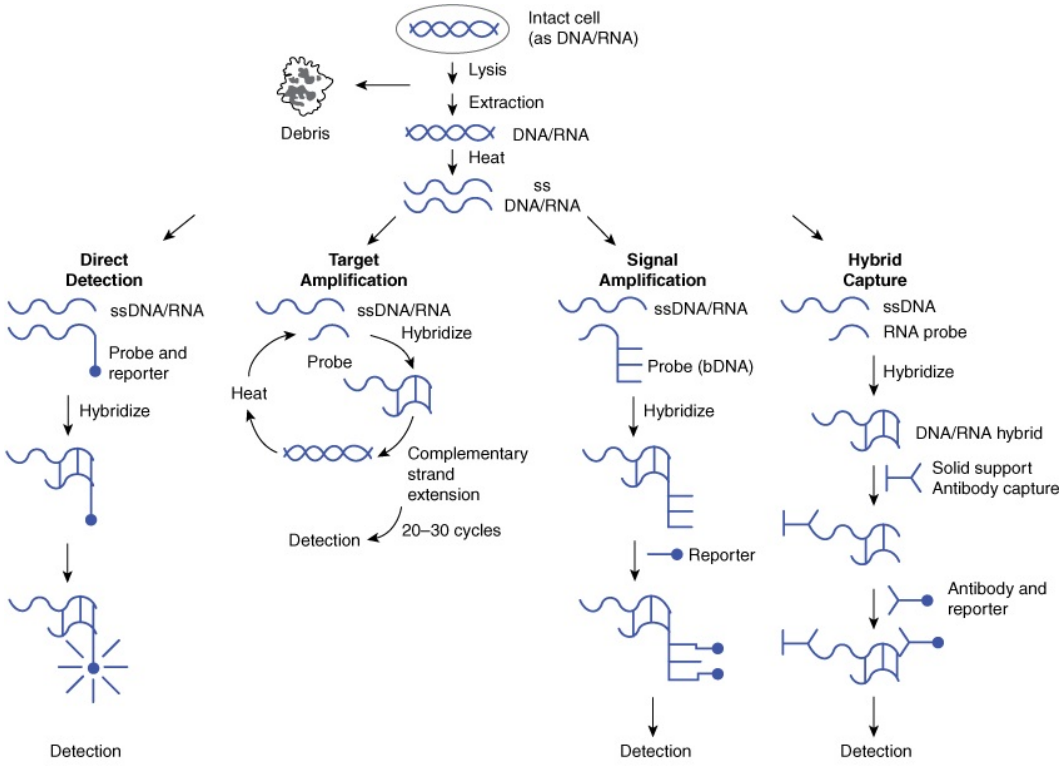

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Strategies for amplification and/or detection of a target-probe complex. DNA or RNA extracted from microorganisms is heated to create single-stranded (ss) DNA/RNA containing appropriate target sequences. These target sequences may be hybridized directly (direct detection) with probes attached to reporter molecules; they may be amplified by repetitive cycles of complementary strand extension (polymerase chain reaction) before attachment of a reporter probe; or the original target-probe signal may be amplified via hybridization with an additional probe containing multiple copies of a secondary reporter target sequence (branched-chain DNA, or bDNA). DNA/RNA hybrids also can be "captured" on a solid support (hybrid capture), with antibody directed at the DNA/RNA hybrids used to concentrate them and a second antibody coupled to a reporter molecule attached to the captured hybrid.

PROBES FOR DIRECT DETECTION OF PATHOGENS IN CLINICAL SPECIMENS

Nucleic acid probes are used for direct detection of pathogens in clinical specimens without amplification of the target strand of DNA or RNA. Such tests detect a relatively short sequence of bases specific for a particular pathogen on single-stranded DNA or RNA by hybridization of a complementary sequence of bases (probe) coupled to a "reporter" system that serves as the signal for detection. Nucleic acid probes are available commercially for direct detection of various bacterial and parasitic pathogens, including *Chlamydia trachomatis*, *N. gonorrhoeae*, and group A *Streptococcus*. A combined assay to detect and differentiate agents of vaginitis/vaginosis (*Gardnerella vaginalis*, *Trichomonas vaginalis*, and *Candida* species) also has been approved. An assortment of probes are available for confirming the identity of cultured pathogens, including some dimorphic molds, *Mycobacterium* species, and other bacteria (e.g., *Campylobacter* species, *Streptococcus* species, and *Staphylococcus aureus*). Probes for the direct detection of bacterial pathogens often are aimed at highly conserved 16S ribosomal RNA sequences, of which there are many more copies than there are of any single genomic DNA sequence in a bacterial cell. The sensitivity and specificity of probe assays for direct detection are comparable to those of more traditional assays, including EIA and culture.

In an alternative probe assay called *hybrid capture*, an RNA probe anneals to a DNA target, and the resulting DNA/RNA hybrid is captured on a solid support by antibody specific for DNA/RNA hybrids (concentration/amplification) and detected by chemiluminescent-labeled antibody specific for DNA/RNA hybrids. Hybrid capture assays are available for *C. trachomatis*, *N. gonorrhoeae*, cytomegalovirus, and human papillomavirus.

Many laboratories have developed their own probes for pathogens; however, unless a method-validation protocol for diagnostic testing has been performed, federal law in the United States restricts the use of such probes to research.

NUCLEIC ACID AMPLIFICATION TEST STRATEGIES

In theory, a single target nucleic acid sequence can be amplified to detectable levels. There are several strategies for nucleic acid amplification tests (NAATs), including PCR, LCR, strand displacement amplification, and self-sustaining sequence replication. In each case, exponential amplification of a pathogen-specific DNA or RNA sequence depends on primers that anneal to the target sequence. The amplified nucleic acid can be detected after the reaction is complete or (in *real-time* detection) as amplification proceeds. The sensitivity of NAATs is far greater than that of traditional assay methods such as culture. However, the care with which the assays are performed is important, because cross-contamination of clinical material with DNA or RNA from other sources (even at low levels) can cause false-positive results.

PCR, the first and still the most common NAAT, requires repeated heating of the DNA to separate the two complementary strands of the double helix, hybridization of a primer sequence to the appropriate target sequence, target amplification using PCR for complementary strand extension, and signal detection via a labeled probe. Methods for the monitoring of PCR after each amplification cycle—via either incorporation of fluorescent dyes into the DNA during primer extension or use of fluorescent probes capable of fluorescence resonance energy transfer—have decreased the period required to detect a specific target. An alternative NAAT employs transcription-mediated amplification, in which an RNA target sequence is converted to DNA, which then is exponentially transcribed into an RNA target. The advantage of this method is that only a single heating/annealing step is required for amplification. At present, amplification assays for *Mycobacterium tuberculosis*, *N. gonorrhoeae*, *C. trachomatis*, *Mycoplasma hominis*, group B *Streptococcus*, and methicillin-resistant *S. aureus* are on the market. Again, many laboratories have used commercially available *taq* polymerase, probe sequences, and analyte-specific

reagents to develop "in-house" assays for diagnostic use. Issues related to quality control, interpretation of results, sample processing, and regulatory requirements have slowed the commercial development of many diagnostic assay kits.

Identification of otherwise difficult-to-identify bacteria involves an initial amplification of a highly conserved region of 16S rDNA by PCR. Automated sequencing of several hundred bases is then performed, and the sequence information is compared with large databases containing sequence information for thousands of different organisms. Although 16S sequencing is not as rapid as other methods and is still relatively expensive for routine use in a clinical microbiology laboratory, it is becoming the definitive method for identification of unusual or difficult-to-cultivate organisms.

QUANTITATIVE NUCLEIC ACID TEST STRATEGIES

With the advent of newer therapeutic regimens for HIV-associated disease, cytomegalovirus infection, and hepatitis B and C virus infections, the response to therapy has been monitored by determining both genotype and "viral load" at various times after treatment initiation. Quantitative NAATs are available for HIV (PCR), cytomegalovirus (PCR), hepatitis B (PCR), and hepatitis C (PCR and TMA). Many laboratories have validated and perform quantitative assays for these and other pathogens (e.g., Epstein-Barr virus), using analyte-specific reagents for NAATs.

Branched-chain DNA (bDNA) testing is an alternative to NAATs for quantitative nucleic acid testing. In such testing, bDNA attaches to a site different from the target-binding sequence of the original probe. Chemiluminescent-labeled oligonucleotides can then bind to multiple repeating sequences on the bDNA. The amplified bDNA signal is detected by chemiluminescence. bDNA assays for viral load of HIV, hepatitis B virus, and hepatitis C virus have been approved by the FDA. The advantage of bDNA assays over PCR is that only a single heating/annealing step is required to hybridize the target-binding probe to the target sequence for amplification.

APPLICATION OF NUCLEIC ACID TESTS

In addition to the applications already discussed, nucleic acid tests are used to detect and identify difficult-to-grow or noncultivable bacterial pathogens such as *Mycobacterium*, *Legionella*, *Ehrlichia*, *Rickettsia*, *Babesia*, *Borrelia*, and *Tropheryma whippelii*. In addition, methods for rapid detection of agents of public health concern, such as *Franciscella tularensis*, *Bacillus anthracis*, smallpox virus, and *Yersinia pestis*, have been developed.

Nucleic acid tests also are used to determine how close the relationship is among different isolates of the same species of pathogen. The demonstration that bacteria of a single clone have infected multiple patients in the context of a possible means of transmission (e.g., a health care provider) offers confirmatory evidence for an outbreak. Pulsed-field gel electrophoresis remains the gold standard for bacterial strain analysis. This method involves the use of restriction enzymes that recognize rare sequences of nucleotides to digest bacterial DNA, resulting in large DNA fragments. These fragments are separated by gel electrophoresis with variable polarity of the electrophoretic current and then are visualized. Similar band patterns (i.e., differences in ≤3 bands) suggest that different bacterial isolates are closely related, or clonal. Simpler methods of strain typing include sequencing of single or multiple genes and PCR-based amplification of repetitive DNA sequences in the bacterial chromosome.

Future applications of nucleic acid testing probably will include the replacement of culture for identification of many pathogens with solid-state DNA/RNA chip technology, in which thousands of unique nucleic acid sequences can be detected on a single silicon chip.

Susceptibility Testing of Bacteria

A principal responsibility of the clinical microbiology laboratory is to determine which antimicrobial agents inhibit a specific bacterial isolate. Such testing is used for patient care and for monitoring of infection control problems, such as methicillin-resistant *S. aureus* or vancomycin-resistant *Enterococcus faecium*. Two approaches are useful. The first is a qualitative assessment of susceptibility, with responses categorized as susceptible, resistant, or intermediate. This approach can involve either the placement of paper disks containing antibiotics on an agar surface inoculated with the bacterial strain to be tested (Kirby-Bauer or disk/agar diffusion method), with measurement of the zones of growth inhibition after incubation, or the use of broth cultures containing a set concentration of antibiotic (breakpoint method). These methods have been calibrated carefully against quantitative methods and clinical experience with each antibiotic, and zones of inhibition and breakpoints have been calculated on a species-by-species basis.

The second approach is to inoculate the test strain of bacteria into a series of broth cultures (or agar plates) with increasing concentrations of antibiotic. The lowest concentration of antibiotic that inhibits visual microbial growth in this test system is known as the *minimal inhibitory concentration* (MIC). If tubes in which no growth is seen are subcultured, the minimal concentration of antibiotic required to kill 99.9% of the starting inoculum also can be determined (*minimal bactericidal concentration,* or MBC). The MIC value can be given a categorical interpretation of susceptible, resistant, or intermediate and so is more widely used than the MBC. Quantitative susceptibility testing by the microbroth dilution technique, a miniaturized version of the broth dilution technique using microwell plates, lends itself to automation and is used commonly in larger clinical laboratories.

A novel version of the disk/agar diffusion method employs a quantitative diffusion gradient, or epsilometer (E-test), and uses an absorbent strip with a known gradient of antibiotic concentrations along its length. When the strip is placed on the surface of an agar plate seeded with a bacterial strain to be tested, antibiotic diffuses into the medium, and bacterial growth is inhibited. The MIC is estimated as the lowest concentration that inhibits visible growth.

For some organisms, such as obligate anaerobes, routine susceptibility testing generally is not performed because of the difficulty of growing the organisms and the predictable sensitivity of most isolates to specific antibiotics.

Susceptibility Testing of Fungal Agents

With the advent of many new agents for treating yeasts and systemic fungal agents, the need for testing of individual isolates for susceptibility to specific antifungal agents has increased. In the past, few laboratories participated in such testing because of a lack of standard methods like those available for testing bacterial agents. However, several systems have been approved for antifungal susceptibility testing. These methods, which determine the minimal fungicidal concentration (MFC), are similar to the broth microdilution methods used to determine the MIC for bacteria. The E-test method is approved for testing the susceptibility of yeasts to fluconazole, itraconazole, and flucytosine, and disk diffusion can be used to test the susceptibility of *Candida* species to fluconazole and voriconazole. Methods for determining the MFC against fungal agents such as *Aspergillus* species are technically difficult, and most clinical laboratories refer requests for such testing to reference laboratories.

Antiviral Testing

See Chap. 177.

FURTHER READINGS

Aires de Sousa M, de Lencastre H: Bridges from hospitals to the laboratory: Genetic portraits of methicillin-resistant

Staphylococcus aureus clones. FEMS Immunol Med Microbiol 40:101, 2004

Baron EJ et al: Prolonged incubation and extensive subculturing do not increase recovery of clinically significant microorganisms from standard automated blood cultures. Clin Infect Dis 41:1677, 2005[PMID: 16267743]

Caliendo AM et al: Distinguishing cytomegalovirus (CMV) infection and disease with CMV nucleic acid assays. J Clin Microbiol 40:1581, 2002[PMID: 11980925]

Cockerhill FR et al: Optimal testing parameters for blood cultures. Clin Infect Dis 38:1724, 2004

Domiati-Saad R, Scheuermann RH: Nucleic acid testing for viral burden and viral genotyping. Clin Chim Acta 363:197, 2005[PMID: 16098959]

Loens K et al: Optimal sampling sites and methods for detection of pathogens possibly causing community-acquired lower respiratory tract infections. J Clin Microbiol 47:21, 2009[PMID: 19020070]

Magiorakos AP, Hadley S: Impact of real-time fungal susceptibility on clinical practices. Curr Opin Infect Dis 17:511, 2004[PMID: 15640704]

Pancholi P et al: Rapid detection of cytomegalovirus infection in transplant patients. Expert Rev Mol Diagn 4:231, 2004[PMID: 14995909]

Petrosino JF et al: Metagenomic pyrosequencing and microbial identification. Clin Chem 55:856, 2010

Pfaller MA et al: Clinical evaluation of a frozen commercially prepared microdilution panel for antifungal susceptibility testing of seven antifungal agents, including the new triazoles posaconazole, ravuconazole and voriconazole. J Clin Microbiol 40:1694, 2002[PMID: 11980944]

Servoss JC, Friedman LS: Serologic and molecular diagnosis of hepatitis B virus. Clin Liver Dis 8:267, 2004[PMID: 15481340]

Yeghiazarian T et al: Quantitation of human immunodeficiency virus type 1 RNA levels in plasma by using small-volume-format branched-DNA assays. J Clin Microbiol 36:2096, 1998[PMID: 9650972]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIVE ENDOCARDITIS: INTRODUCTION

The prototypic lesion of infective endocarditis, the *vegetation***(Fig. 124-1)**, is a mass of platelets, fibrin, microcolonies of microorganisms, and scant inflammatory cells. Infection most commonly involves heart valves (either native or prosthetic) but may also occur on the low-pressure side of a ventricular septal defect, on the mural endocardium where it is damaged by aberrant jets of blood or foreign bodies, or on intracardiac devices themselves. The analogous process involving arteriovenous shunts, arterioarterial shunts (patent ductus arteriosus), or a coarctation of the aorta is called *infective endarteritis*.

Figure 124-1

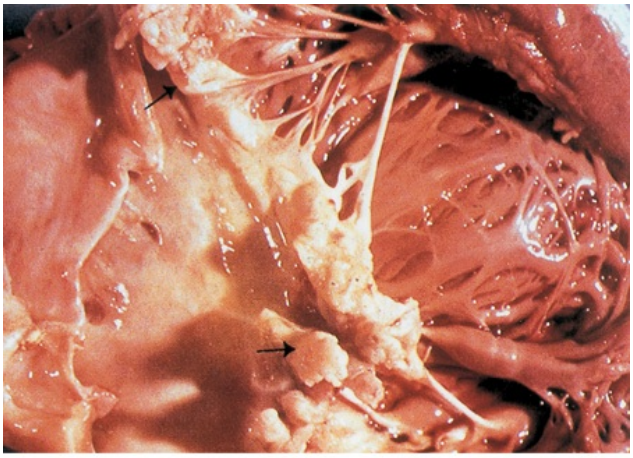

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vegetations (*arrows*) due to viridans streptococcal endocarditis involving the mitral valve.

Endocarditis may be classified according to the temporal evolution of disease, the site of infection, the cause of infection, or a predisposing risk factor such as injection drug use. While each classification criterion provides therapeutic and prognostic insight, none is sufficient alone. *Acute endocarditis* is a hectically febrile illness that rapidly damages cardiac structures, hematogenously seeds extracardiac sites, and, if untreated, progresses to death within weeks. *Subacute endocarditis* follows an indolent course; causes structural cardiac damage only slowly, if at all; rarely metastasizes; and is gradually progressive unless complicated by a major embolic event or ruptured mycotic aneurysm.

In developed countries, the incidence of endocarditis ranges from 2.6 to 7 cases per 100,000 population per year and has remained relatively stable during recent decades. While congenital heart diseases remain a constant predisposition, predisposing conditions in developed countries have shifted from chronic rheumatic heart disease (which remains a common predisposition in developing countries) to illicit IV drug use, degenerative valve disease, and intracardiac devices. The incidence of endocarditis is notably increased among the elderly. In developed countries, 30–35% of cases of native valve endocarditis (NVE) are associated with health care, and 16–30% of all cases of endocarditis involve prosthetic valves. The risk of prosthesis infection is greatest during the first 6–12 months after valve replacement; gradually declines to a low, stable rate thereafter; and is similar for mechanical and bioprosthetic devices.

Etiology

Although many species of bacteria and fungi cause sporadic episodes of endocarditis, a few bacterial species cause the majority of cases **(Tables 124–1)**. Because of their different portals of entry, the pathogens involved vary somewhat with the clinical types of endocarditis. The oral cavity, skin, and upper respiratory tract are the respective primary portals for the viridans streptococci, staphylococci, and HACEK organisms (*Haemophilus*, *Actinobacillus*, *Cardiobacterium*, *Eikenella*, and *Kingella*; *Haemophilus aphrophilus* and *Actinobacillus actinomycetemcomitans* have been reclassified into the genus *Aggregatibacter*). *Streptococcus gallolyticus* (formerly *S. bovis*) originates from the gastrointestinal tract, where it is associated with polyps and colonic tumors, and enterococci enter the bloodstream from the genitourinary tract. Health care–associated NVE, commonly caused by *Staphylococcus aureus*, coagulase-negative staphylococci (CoNS), and enterococci, has a nosocomial onset (55%) or a community onset (45%) in patients who have had extensive contact with the health care system over the preceding 90 days. Endocarditis complicates 6–25% of episodes of catheter-associated *S. aureus* bacteremia; the higher rates are detected by careful transesophageal echocardiography (TEE) screening (see "Echocardiography," below).

Tables 124–1 Organisms Causing Major Clinical Forms of Endocarditis

Organism	Percentage of Cases							
	Native Valve Endocarditis		Prosthetic Valve Endocarditis at Indicated Time of Onset (Months) after Valve Surgery			Endocarditis in Injection Drug Users		
	Community-Acquired (*n* =1718)	Health Care–Associated (*n* =788)	<2 (*n* = 144)	2–12 (*n* = 31)	>12 (n = 194)	Right-Sided (*n* = 346)	Left-Sided (*n* = 204)	Total (*n* = 675)[a]
Streptococci[b]	40	9	1	9	31	5	15	12
Pneumococci	2	—	—	—	—	—	—	—
Enterococci	9	13	8	12	11	2	24	9
Staphylococcus aureus	28	53[c]	22	12	18	77	23	57
Coagulase-negative staphylococci	5	12	33	32	11	—	—	—
Fastidious gram-negative coccobacilli (HACEK group)[d]	3	—	—	—	6	—	—	—
Gram-negative bacilli	1	2	13	3	6	5	13	7
Candida spp.	<1	2	8	12	1	—	12	4
Polymicrobial/miscellaneous	3	4	3	6	5	8	10	7
Diphtheroids	—	<1	6	—	3	—	—	0.1
Culture-negative	9	5	5	6	8	3	3	3

[a] The total number of cases is larger than the sum of right- and left-sided cases because the location of infection was not specified in some cases.

[b] Includes viridans streptococci; *Streptococcus gallolyticus*; other non–group A, groupable streptococci; and *Abiotrophia* spp. (nutritionally variant, pyridoxal-requiring streptococci).

[c] Methicillin resistance is common among these *S. aureus* strains.

[d] Includes *Haemophilus* spp., *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella* spp., and *Kingella* spp.

Note: Data are compiled from multiple studies.

Prosthetic valve endocarditis (PVE) arising within 2 months of valve surgery is generally nosocomial, the result of intraoperative contamination of the prosthesis or a bacteremic postoperative complication. This nosocomial origin is reflected in the primary microbial causes: *S. aureus*, CoNS, facultative gram-negative bacilli, diphtheroids, and fungi. The portals of entry and organisms causing cases beginning >12 months after surgery are similar to those in community-acquired NVE. PVE due to CoNS that presents 2–12 months after surgery often represents delayed-onset nosocomial infection. Regardless of the time of onset after surgery, at least 68–85% of CoNS strains that cause PVE are resistant to methicillin.

Transvenous pacemaker– or implanted defibrillator–associated endocarditis is usually nosocomial. The majority of episodes occur within weeks of implantation or generator change and are caused by *S. aureus* or CoNS, both of which are commonly resistant to methicillin.

Endocarditis occurring among injection drug users, especially that involving the tricuspid valve, is commonly caused by *S. aureus*, many strains of which are resistant to methicillin. Left-sided valve infections in addicts have a more varied etiology. In addition to the usual causes of endocarditis, these cases are caused by *Pseudomonas aeruginosa* and *Candida* species and sporadically by unusual organisms such as *Bacillus*, *Lactobacillus*, and *Corynebacterium* species. Polymicrobial endocarditis occurs among injection drug users. HIV infection in drug users does not significantly influence the causes of endocarditis.

From 5% to 15% of patients with endocarditis have negative blood cultures; in one-third to one-half of these cases, cultures are negative because of prior antibiotic exposure. The remainder of these patients are infected by fastidious organisms, such as nutritionally variant organisms (now designated *Granulicatella* and *Abiotrophia* species), HACEK organisms, *Coxiella burnetii*, and *Bartonella* species. Some fastidious organisms occur in characteristic geographic settings (e.g., *C. burnetii* and *Bartonella* species in Europe, *Brucella* species in the Middle East). *Tropheryma whipplei* causes an indolent, culture-negative, afebrile form of endocarditis.

Pathogenesis

The endothelium, unless damaged, is resistant to infection by most bacteria and to thrombus formation. Endothelial injury (e.g., at the site of impact of high-velocity blood jets or on the low-pressure side of a cardiac structural lesion) allows either direct infection by virulent organisms or the development of an uninfected platelet-fibrin thrombus—a condition called *nonbacterial thrombotic endocarditis* (NBTE). The thrombus subsequently serves as a site of bacterial attachment during transient bacteremia. The cardiac conditions most commonly resulting in NBTE are mitral regurgitation, aortic stenosis, aortic regurgitation, ventricular septal defects, and

complex congenital heart disease. NBTE also arises as a result of a hypercoagulable state; this phenomenon gives rise to the clinical entity of *marantic endocarditis* (uninfected vegetations seen in patients with malignancy and chronic diseases) and to bland vegetations complicating systemic lupus erythematosus and the antiphospholipid antibody syndrome.

Organisms that cause endocarditis generally enter the bloodstream from mucosal surfaces, the skin, or sites of focal infection. Except for more virulent bacteria (e.g., *S. aureus*) that can adhere directly to intact endothelium or exposed subendothelial tissue, microorganisms in the blood adhere at sites of NBTE. If resistant to the bactericidal activity of serum and the microbicidal peptides released locally by platelets, the organisms proliferate and induce platelet deposition and a procoagulant state at the site by eliciting tissue factor from the endothelium or, in the case of *S. aureus*, from monocytes as well. Fibrin deposition combines with platelet aggregation and microorganism proliferation to generate an infected vegetation. The organisms that commonly cause endocarditis have surface adhesin molecules, collectively called microbial surface components recognizing adhesin matrix molecules (MSCRAMMs), that mediate adherence to NBTE sites or injured endothelium. Fibronectin-binding proteins present on many gram-positive bacteria, clumping factor (a fibrinogen- and fibrin-binding surface protein) on *S. aureus*, and glucans or FimA (a member of the family of oral mucosal adhesins) on streptococci facilitate adherence. Fibronectin-binding proteins are required for *S. aureus* invasion of intact endothelium; thus these surface proteins may facilitate infection of previously normal valves. In the absence of host defenses, organisms enmeshed in the growing platelet-fibrin vegetation proliferate to form dense microcolonies. Organisms deep in vegetations are metabolically inactive (nongrowing) and relatively resistant to killing by antimicrobial agents. Proliferating surface organisms are shed into the bloodstream continuously.

The pathophysiologic consequences and clinical manifestations of endocarditis—other than constitutional symptoms, which probably result from cytokine production—arise from damage to intracardiac structures; embolization of vegetation fragments, leading to infection or infarction of remote tissues; hematogenous infection of sites during bacteremia; and tissue injury due to the deposition of circulating immune complexes or immune responses to deposited bacterial antigens.

Clinical Manifestations

The clinical syndrome of infective endocarditis is highly variable and spans a continuum between acute and subacute presentations. NVE (whether acquired in the community or in association with health care), PVE, and endocarditis due to injection drug use share clinical and laboratory manifestations **(Tables 124–2)**. The causative microorganism is primarily responsible for the temporal course of endocarditis. β-Hemolytic streptococci, *S. aureus*, and pneumococci typically result in an acute course, although *S. aureus* occasionally causes subacute disease. Endocarditis caused by *Staphylococcus lugdunensis* (a coagulase-negative species) or by enterococci may present acutely. Subacute endocarditis is typically caused by viridans streptococci, enterococci, CoNS, and the HACEK group. Endocarditis caused by *Bartonella* species, *T. whipplei,* or *C. burnetii* is exceptionally indolent.

Tables 124–2 Clinical and Laboratory Features of Infective Endocarditis

Feature	Frequency, %
Fever	80–90
Chills and sweats	40–75
Anorexia, weight loss, malaise	25–50
Myalgias, arthralgias	15–30
Back pain	7–15
Heart murmur	80–85
New/worsened regurgitant murmur	20–50
Arterial emboli	20–50
Splenomegaly	15–50
Clubbing	10–20
Neurologic manifestations	20–40
Peripheral manifestations (Osler's nodes, subungual hemorrhages, Janeway lesions, Roth's spots)	2–15
Petechiae	10–40
Laboratory manifestations	
Anemia	70–90
Leukocytosis	20–30
Microscopic hematuria	30–50
Elevated erythrocyte sedimentation rate	60–90
Elevated C-reactive protein level	>90
Rheumatoid factor	50
Circulating immune complexes	65–100
Decreased serum complement	5–40

The clinical features of endocarditis are nonspecific. However, these symptoms in a febrile patient with valvular abnormalities or a behavior pattern that predisposes to endocarditis (e.g., injection drug use) suggest the diagnosis, as do bacteremia with organisms that frequently cause endocarditis, otherwise-unexplained arterial

emboli, and progressive cardiac valvular incompetence. In patients with subacute presentations, fever is typically low-grade and rarely exceeds 39.4°C (103°F); in contrast, temperatures of 39.4°–40°C (103°–104°F) are often noted in acute endocarditis. Fever may be blunted or absent in patients who are elderly or severely debilitated or who have marked cardiac or renal failure.

CARDIAC MANIFESTATIONS

Although heart murmurs are usually indicative of the predisposing cardiac pathology rather than of endocarditis, valvular damage and ruptured chordae may result in new regurgitant murmurs. In acute endocarditis involving a normal valve, murmurs may be absent initially but ultimately are detected in 85% of cases. Congestive heart failure (CHF) develops in 30–40% of patients; it is usually a consequence of valvular dysfunction but occasionally is due to endocarditis-associated myocarditis or an intracardiac fistula. Heart failure due to aortic valve dysfunction progresses more rapidly than does that due to mitral valve dysfunction. Extension of infection beyond valve leaflets into adjacent annular or myocardial tissue results in perivalvular abscesses, which in turn may cause intracardiac fistulae with new murmurs. Abscesses may burrow from the aortic valve annulus through the epicardium, causing pericarditis, or into the upper ventricular septum, where they may interrupt the conduction system, leading to varying degrees of heart block. Perivalvular abscesses arising from the mitral valve rarely interrupt conduction pathways near the atrioventricular node or in the proximal bundle of His. Emboli to a coronary artery occur in 2% of patients and may result in myocardial infarction.

NONCARDIAC MANIFESTATIONS

The classic nonsuppurative peripheral manifestations of subacuteendocarditis are related to the duration of infection and, with early diagnosis and treatment, have become infrequent. In contrast, septic embolization mimicking some of these lesions (subungual hemorrhage, Osler's nodes) is common in patients with acute *S. aureus* endocarditis **(Fig. 124-2)**. Musculoskeletal pain usually remits promptly with treatment but must be distinguished from focal metastatic infections (e.g., spondylodiscitis), which may complicate 10–15% of cases. Hematogenously seeded focal infection is most often clinically evident in the skin, spleen, kidneys, skeletal system, and meninges. Arterial emboli are clinically apparent in up to 50% of patients. Endocarditis caused by *S. aureus,* vegetations >10 mm in diameter (as measured by echocardiography), and infection involving the mitral valve are independently associated with an increased risk of embolization. Emboli occurring late during or after effective therapy do not in themselves constitute evidence of failed antimicrobial treatment. Cerebrovascular emboli presenting as strokes or occasionally as encephalopathy complicate 15–35% of cases of endocarditis. One-half of these events precede the diagnosis of endocarditis. The frequency of stroke is 8 per 1000 patient-days during the week prior to diagnosis; the figure falls to 4.8 and 1.7 per 1000 patient-days during the first and second weeks of effective antimicrobial therapy, respectively. This decline exceeds that which can be attributed to change in vegetation size. Only 3% of strokes occur after 1 week of effective therapy. Other neurologic complications include aseptic or purulent meningitis, intracranial hemorrhage due to hemorrhagic infarcts or ruptured mycotic aneurysms, and seizures. (*Mycotic aneurysms* are focal dilations of arteries occurring at points in the artery wall that have been weakened by infection in the vasa vasorum or where septic emboli have lodged.) Microabscesses in brain and meninges occur commonly in *S. aureus* endocarditis; surgically drainable intracerebral abscesses are infrequent.

Figure 124-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Septic emboli with hemorrhage and infarction due to acute *Staphylococcus aureus* endocarditis. *(Used with permission of L. Baden.)*

Immune complex deposition on the glomerular basement membrane causes diffuse hypocomplementemic glomerulonephritis and renal dysfunction, which typically improve with effective antimicrobial therapy. Embolic renal infarcts cause flank pain and hematuria but rarely cause renal dysfunction.

MANIFESTATIONS OF SPECIFIC PREDISPOSING CONDITIONS

Almost 50% of endocarditis cases associated with injection drug use are limited to the tricuspid valve and present with fever but with faint or no murmur. In 75% of cases, septic emboli cause cough, pleuritic chest pain, nodular pulmonary infiltrates, or occasionally pyopneumothorax. Infection of the aortic or mitral valves on the left side of

the heart presents with the typical clinical features of endocarditis.

Health care–associated endocarditis has typical manifestations if it is not associated with a retained intracardiac device or masked by the symptoms of concurrent comorbid illness. Transvenous pacemaker– or implanted defibrillator–associated endocarditis may be associated with obvious or cryptic generator pocket infection and results in fever, minimal murmur, and pulmonary symptoms due to septic emboli.

Late-onset PVE presents with typical clinical features. In cases arising within 60 days of valve surgery (early onset), typical symptoms may be obscured by comorbidity associated with recent surgery. In both early-onset and more delayed presentations, paravalvular infection is common and often results in partial valve dehiscence, regurgitant murmurs, CHF, or disruption of the conduction system.

Diagnosis

THE DUKE CRITERIA

The diagnosis of infective endocarditis is established with certainty only when vegetations are examined histologically and microbiologically. Nevertheless, a highly sensitive and specific diagnostic schema—known as the *Duke criteria*—has been developed on the basis of clinical, laboratory, and echocardiographic findings **(Tables 124–3)**. Documentation of two major criteria, of one major criterion and three minor criteria, or of five minor criteria allows a clinical diagnosis of definite endocarditis. The diagnosis of endocarditis is rejected if an alternative diagnosis is established, if symptoms resolve and do not recur with ≤4 days of antibiotic therapy, or if surgery or autopsy after ≤4 days of antimicrobial therapy yields no histologic evidence of endocarditis. Illnesses not classified as definite endocarditis or rejected as such are considered cases of possible infective endocarditis when either one major criterion and one minor criterion or three minor criteria are fulfilled. Requiring the identification of clinical features of endocarditis for classification as possible infective endocarditis increases the specificity of the schema without significantly reducing its sensitivity.

Tables 124–3 The Duke Criteria for the Clinical Diagnosis of Infective Endocarditis[a]

Major Criteria
1. Positive blood culture
Typical microorganism for infective endocarditis from two separate blood cultures
Viridans streptococci, *Streptococcus gallolyticus*, HACEK group, *Staphylococcus aureus*, *or*
Community-acquired enterococci in the absence of a primary focus, *or*
Persistently positive blood culture, defined as recovery of a microorganism consistent with infective endocarditis from:
Blood cultures drawn >12 h apart; *or*
All of 3 or a majority of ≥4 separate blood cultures, with first and last drawn at least 1 h apart
Single positive blood culture for *Coxiella burnetii* or phase I IgG antibody titer of >1:800
2. Evidence of endocardial involvement
Positive echocardiogram[b]
Oscillating intracardiac mass on valve or supporting structures or in the path of regurgitant jets or in implanted material, in the absence of an alternative anatomic explanation, *or*
Abscess, *or*
New partial dehiscence of prosthetic valve, *or*
New valvular regurgitation (increase or change in preexisting murmur not sufficient)
Minor Criteria
1. Predisposition: predisposing heart condition or injection drug use
2. Fever ≥38.0°C (≥100.4°F)
3. Vascular phenomena: major arterial emboli, septic pulmonary infarcts, mycotic aneurysm, intracranial hemorrhage, conjunctival hemorrhages, Janeway lesions
4. Immunologic phenomena: glomerulonephritis, Osler's nodes, Roth's spots, rheumatoid factor
5. Microbiologic evidence: positive blood culture but not meeting major criterion as noted previously[c] or serologic evidence of active infection with organism consistent with infective endocarditis

[a] Definite endocarditis is defined by documentation of two major criteria, of one major criterion and three minor criteria, or of five minor criteria. See text for further details.

[b] Transesophageal echocardiography is recommended for assessing possible prosthetic valve endocarditis or complicated endocarditis.

[c] Excluding single positive cultures for coagulase-negative staphylococci and diphtheroids, which are common culture contaminants, and organisms that do not cause endocarditis frequently, such as gram-negative bacilli.

Note: HACEK, *Haemophilus* spp., *Aggregatibacter actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella corrodens*, *Kingella* spp.

Source: Adapted from Li et al, with permission from the University of Chicago Press.

The roles of bacteremia and echocardiographic findings in the diagnosis of endocarditis are emphasized in the Duke criteria. The requirement for multiple positive blood cultures over time is consistent with the continuous low-density bacteremia characteristic of endocarditis. Among patients with untreated endocarditis who ultimately have a positive blood culture, 95% of all blood cultures are positive. The diagnostic criteria attach significance to the species of organism isolated from blood cultures. To fulfill a major criterion, the isolation of an organism that causes both endocarditis and bacteremia in the absence of endocarditis (e.g., *S. aureus*, enterococci) must take place repeatedly (i.e., persistent bacteremia) and in the absence of a primary focus of infection. Organisms that rarely cause endocarditis but commonly contaminate blood cultures (e.g., diphtheroids, CoNS) must be isolated repeatedly if their isolation is to serve as a major criterion.

BLOOD CULTURES

Isolation of the causative microorganism from blood cultures is critical for diagnosis, determination of antimicrobial susceptibility, and planning of treatment. In the absence of prior antibiotic therapy, three 2-bottle blood culture sets, separated from one another by at least 1 h, should be obtained from different venipuncture sites over 24 h. If the cultures remain negative after 48–72 h, two or three additional blood culture sets should be obtained, and the laboratory should be consulted for advice regarding optimal culture techniques. Pending culture results, empirical antimicrobial therapy should be withheld initially from hemodynamically stable patients with suspected subacuteendocarditis, especially those who have received antibiotics within the preceding 2 weeks; thus, if necessary, additional blood culture sets can be obtained without the confounding effect of empirical treatment. Patients with acute endocarditis or with deteriorating hemodynamics who may require urgent surgery should be treated empirically immediately after three sets of blood cultures are obtained over several hours.

NON-BLOOD-CULTURE TESTS

Serologic tests can be used to implicate causally some organisms that are difficult to recover by blood culture: *Brucella*, *Bartonella*, *Legionella*, *Chlamydophila psittaci*, and *C. burnetii*. Pathogens can also be identified in vegetations by culture, microscopicexamination with special stains (i.e., the periodic acid–Schiff stain for *T. whipplei*), or direct fluorescence antibody techniques and by the use of polymerase chain reaction (PCR) to recover unique microbial DNA or 16S rRNA that, when sequenced, allows identification of organisms.

ECHOCARDIOGRAPHY

Echocardiography allows anatomic confirmation of infective endocarditis, sizing of vegetations, detection of intracardiac complications, and assessment of cardiac function **(Fig. 124-3)**. Transthoracic echocardiography (TTE) is noninvasive and exceptionally specific; however, it cannot image vegetations <2 mm in diameter, and in 20% of patients it is technically inadequate because of emphysema or body habitus. TTE detects vegetations in only 65% of patients with definite clinical endocarditis. Moreover, TTE is not adequate for evaluating prosthetic valves or detecting intracardiac complications. TEE is safe and detects vegetations in >90% of patients with definite endocarditis; nevertheless, initial studies may be false-negative in 6–18% of endocarditis patients. When endocarditis is likely, a negative TEE result does not exclude the diagnosis but rather warrants repetition of the study in 7–10 days. TEE is the optimal method for the diagnosis of PVE or the detection of myocardial abscess, valve perforation, or intracardiac fistulae.

Figure 124-3

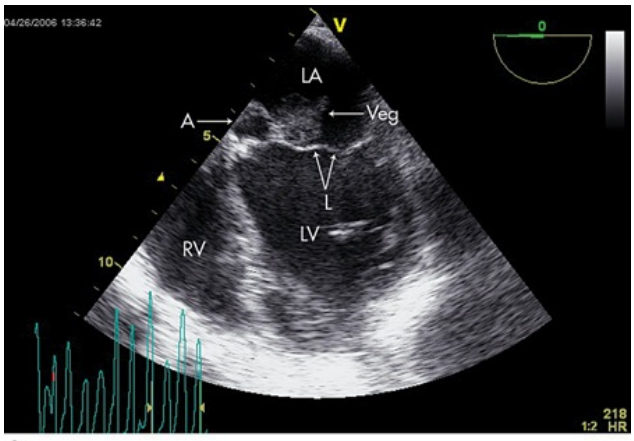

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

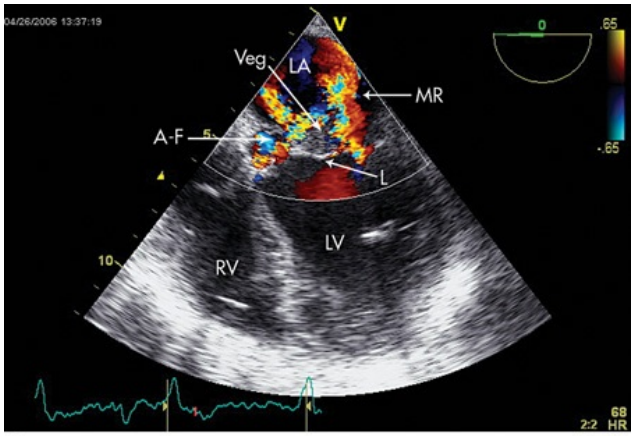

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Imaging of a mitral valve infected with *Staphylococcus aureus* by low-esophageal four-chamber-view transesophageal echocardiography (TEE). ***A.*** Two-dimensional echocardiogram showing a large vegetation with an adjacent echolucent abscess cavity. ***B.*** Color-flow Doppler image showing severe mitral regurgitation through both the abscess-fistula and the central valve orifice. A, abscess; A-F, abscess-fistula; L, valve leaflets; LA, left atrium; LV, left ventricle; MR, mitral central valve regurgitation; RV, right ventricle; veg, vegetation. *(With permission of Andrew Burger, MD)*

Experts favor echocardiographic evaluation of all patients with a clinical diagnosis of endocarditis; however, the test should not be used to screen patients with a low probability of endocarditis (e.g., patients with unexplained fever). An American Heart Association approach to the use of echocardiography for evaluation of patients with suspected endocarditis is illustrated in **Fig. 124-4**.

Figure 124-4

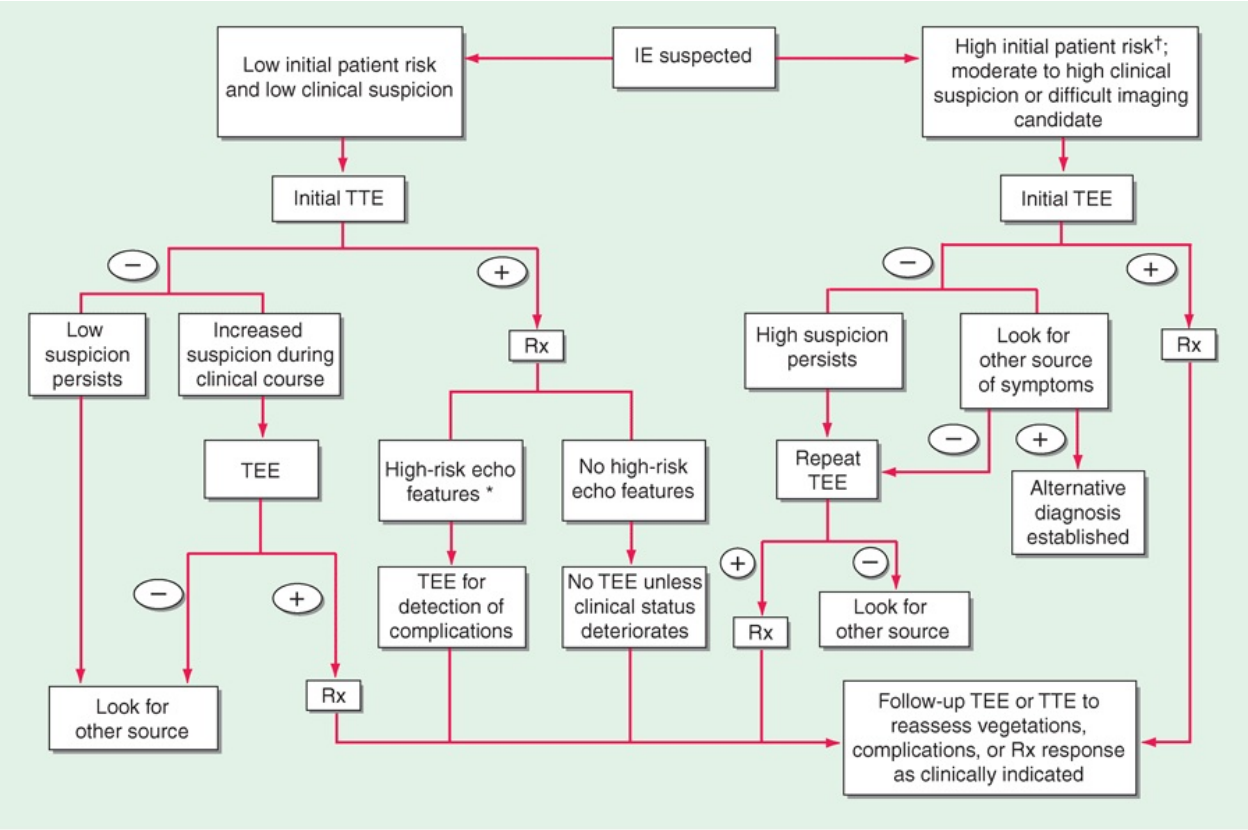

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The diagnostic use of transesophageal and transtracheal echocardiography (TEE and TTE, respectively). †High initial patient risk for endocarditis as listed in Tables 124–8 or evidence of intracardiac complications (new regurgitant murmur, new electrocardiographic conduction changes, or congestive heart failure). * High-risk echocardiographic features include large vegetations, valve insufficiency, paravalvular infection, or ventricular dysfunction. Rx indicates initiation of antibiotic therapy. *[Reproduced with permission*

from Diagnosis and Management of Infective Endocarditis and Its Complications (Circulation 1998; 98:2936-2948. © 1998 American Heart Association.)]

OTHER STUDIES

Many laboratory studies that are not diagnostic—i.e., complete blood count, creatinine determination, liver function tests, chest radiography, and electrocardiography—are nevertheless important in the management of patients with endocarditis. The erythrocyte sedimentation rate, C-reactive protein level, and circulating immune complex titer are commonly increased in endocarditis (Tables 124–2). Cardiac catheterization is useful primarily to assess coronary artery patency in older individuals who are to undergo surgery for endocarditis.

Treatment: Infective Endocarditis

ANTIMICROBIAL THERAPY

It is difficult to eradicate bacteria from the vegetation because local host defenses are deficient and because the largely nongrowing, metabolically inactive bacteria are less easily killed by antibiotics. To cure endocarditis, all bacteria in the vegetation must be killed; therefore, therapy must be bactericidal and prolonged. Antibiotics are generally given parenterally to achieve serum concentrations that, through passive diffusion, lead to effective concentrations in the depths of the vegetation. To select effective therapy requires knowledge of the susceptibility of the causative microorganisms. The decision to initiate treatment empirically must balance the need to establish a microbiologic diagnosis against the potential progression of disease or the need for urgent surgery (see "Blood Cultures," above). Simultaneous infection at other sites (such as meningitis), allergies, end-organ dysfunction, interactions with concomitant medications, and risks of adverse events must be considered in the selection of therapy.

Although given for several weeks longer, the regimens recommended for the treatment of endocarditis involving prosthetic valves (except for staphylococcal infections) are similar to those used to treat NVE **(Tables 124–4)**. Recommended doses and durations of therapy should be adhered to unless alterations are required by end-organ dysfunction or adverse events.

Tables 124–4 Antibiotic Treatment for Infective Endocarditis Caused by Common Organisms[a]

Organism	Drug (Dose, Duration)	Comments
Streptococci		
Penicillin-susceptble[b] streptococci, *S. gallolyticus*	• Penicillin G (2–3 mU IV q4h for 4 weeks) • Ceftriaxone (2 g/d IV as a single dose for 4 weeks) • Vancomycin[c] (15 mg/kg IV q12h for 4 weeks) • Penicillin G (2–3 mU IV q4h) or ceftriaxone (2 g IV qd) for 2 weeks *plus* Gentamicin[d] (3 mg/kg qd IV or IM, as a single dose[e] or divided into equal doses q8h for 2 weeks)	— Can use ceftriaxone in patients with nonimmediate penicillin allergy Use vancomycin in patients with severe or immediate β-lactam allergy Avoid 2-week regimen when risk of aminoglycoside toxicity is increased and in prosthetic valve or complicated endocarditis
Relatively penicillin-resistant[f] streptococci	• Penicillin G (4 mU IV q4h) or ceftriaxone (2 g IV qd) for 4 weeks *plus* Gentamicin[d] (3 mg/kg qd IV or IM, as a single dose[e] or divided into equal doses q8h for 2 weeks) • Vancomycin[c] as noted above for 4 weeks	Penicillin alone at this dose for 6 weeks or with gentamicin during initial 2 weeks preferred for prosthetic valve endocarditis caused by streptococci with penicillin MICs of ≤0.1 μg/mL —
Moderately penicillin-resistant[g] streptococci, nutritionally variant organisms, or	• Penicillin G (4–5 mU IV q4h) or ceftriaxone (2 g IV qd) for 6 weeks *plus* Gentamicin[d] (3	Preferred for prosthetic valve endocarditis caused by streptococci with penicillin MICs of >0.1 μg/mL

Gemella morbillorum	Gentamicin[d] (3 mg/kg qd IV or IM as a single dose[e] or divided into equal doses q8h for 6 weeks) • Vancomycin[c] as noted above for 4 weeks	—
Enterococci[h]		
	• Penicillin G (4–5 mU IV q4h) *plus* Gentamicin[d] (1 mg/kg IV q8h), both for 4–6 weeks • Ampicillin (2 g IV q4h) *plus* Gentamicin[d] (1 mg/kg IV q8h), both for 4–6 weeks • Vancomycin[c] (15 mg/kg IV q12h) *plus* Gentamicin[d] (1 mg/kg IV q8h), both for 4–6 weeks	Can use streptomycin (7.5 mg/kg q12h) in lieu of gentamicin if there is not high-level resistance to streptomycin — Use vancomycin plus gentamicin for penicillin-allergic patients, or desensitize to penicillin
Staphylococci		
Methicillin-susceptible, infecting native valves (no foreign devices)	• Nafcillin or oxacillin (2 g IV q4h for 4–6 weeks) • Cefazolin (2 g IV q8h for 4–6 weeks) • Vancomycin[c] (15 mg/kg IV q12h for 4–6 weeks)	Can use penicillin (4 mU q4h) if isolate is penicillin-susceptible (does not produce β-lactamase) Can use cefazolin regimen for patients with nonimmediate penicillin allergy Use vancomycin for patients with immediate (urticarial) or severe penicillin allergy
Methicillin-resistant, infecting native valves (no foreign devices)	• Vancomycin[c] (15 mg/kg IV q8–12h for 4–6 weeks)	No role for routine use of rifampin
Methicillin-susceptible, infecting prosthetic valves	• Nafcillin or oxacillin (2 g IV q4h for 6–8 weeks) *plus* Gentamicin[d] (1 mg/kg IM or IV q8h for 2 weeks) *plus* Rifampin[i] (300 mg PO q8h for 6–8 weeks)	Use gentamicin during initial 2 weeks; determine susceptibility to gentamicin before initiating rifampin (see text); if patient is highly allergic to penicillin, use regimen for methicillin-resistant staphylococci; if β-lactam allergy is of the minor, nonimmediate type, can substitute cefazolin for oxacillin/nafcillin
Methicillin-resistant, infecting prosthetic valves	• Vancomycin[c] (15 mg/kg IV q12h for 6–8 weeks) *plus* Gentamicin[d] (1 mg/kg IM or IV q8h for 2 weeks) *plus* Rifampin[i] (300 mg PO q8h for 6–8 weeks)	Use gentamicin during initial 2 weeks; determine gentamicin susceptibility before initiating rifampin (see text)
HACEK Organisms		
	• Ceftriaxone (2 g/d IV as a single dose for 4 weeks) • Ampicillin/sulbactam (3 g IV q6h for 4 weeks)	Can use another third-generation cephalosporin at comparable dosage —

[a] Doses are for adults with normal renal function. Doses of gentamicin, streptomycin, and vancomycin must be adjusted for reduced renal function. Ideal body weight is used to calculate doses of gentamicin and streptomycin

per kilogram (men = 50 kg + 2.3 kg per inch over 5 feet; women = 45.5 kg + 2.3 kg per inch over 5 feet).

b MIC, ≤0.1 μg/mL.

c Vancomycin dose is based on actual body weight. Adjust for trough level of 10–15 μg/mL for streptococcal and enterococcal infections and 15–20 μg/mL for staphylococcal infections.

d Aminoglycosides should not be administered as single daily doses for enterococcal endocarditis and should be introduced as part of the initial treatment. Target peak and trough serum concentrations of divided-dose gentamicin 1 h after a 20- to 30-min infusion or IM injection are ~3.5 μg/mL and ≤1 μg/mL, respectively; target peak and trough serum concentrations of streptomycin (timing as with gentamicin) are 20–35 μg/mL and <10 μg/mL, respectively.

e Netilmicin (4 mg/kg qd, as a single dose) can be used in lieu of gentamicin.

f MIC, >0.1 μg/mL and <0.5 μg/mL.

g MIC, ≥0.5 μg/mL and <8 μg/mL.

h Antimicrobial susceptibility must be evaluated; see text.

i Rifampin increases warfarin and dicumarol requirements for anticoagulation.

Organism-Specific Therapies

Streptococci

Optimal therapy for streptococcal endocarditis is based on the minimal inhibitory concentration (MIC) of penicillin for the causative isolate (Tables 124–4). The 2-week penicillin/gentamicin or ceftriaxone/gentamicin regimens should not be used to treat complicated NVE or PVE. The regimen recommended for relatively penicillin-resistant streptococci is advocated for treatment of group B, C, or G streptococcal endocarditis. Nutritionally variant organisms (*Granulicatella* or *Abiotrophia* species) and *Gemella morbillorum* are treated with the regimen for moderately penicillin-resistant streptococci, as is PVE caused by these organisms or by streptococci with a penicillin MIC of >0.1 μg/mL (Tables 124–4).

Enterococci

Enterococci are resistant to oxacillin, nafcillin, and the cephalosporins and are only inhibited—not killed—by penicillin, ampicillin, teicoplanin (not available in the United States), and vancomycin. To kill enterococci requires the synergistic interaction of a cell wall–active antibiotic (penicillin, ampicillin, vancomycin, or teicoplanin) that is effective at achievable serum concentrations and an aminoglycoside (gentamicin or streptomycin) to which the isolate does not exhibit high-level resistance. An isolate's resistance to cell wall–active agents or its ability to replicate in the presence of gentamicin at ≥500 μg/mL or streptomycin at 1000–2000 μg/mL—a phenomenon called *high-level aminoglycoside resistance*—indicates that the ineffective antimicrobial agent cannot participate in the interaction to produce killing. High-level resistance to gentamicin predicts that tobramycin, netilmicin, amikacin, and kanamycin also will be ineffective. In fact, even when enterococci are not highly resistant to gentamicin, it is difficult to predict the ability of these other aminoglycosides to participate in synergistic killing; consequently, they should not in general be used to treat enterococcal endocarditis. High concentrations of ampicillin plus ceftriaxone or cefotaxime, by expanded binding of penicillin-binding proteins, kill *E. faecalis* in vitro and in animal models of endocarditis.

Enterococci causing endocarditis must be tested for high-level resistance to streptomycin and gentamicin, β-lactamase production, and susceptibility to penicillin and ampicillin (MIC, <8 μg/mL) and to vancomycin (MIC, ≤4 μg/mL). If the isolate produces β-lactamase, ampicillin/sulbactam or vancomycin can be used as the cell wall–active component; if the penicillin/ampicillin MIC is ≥8 μg/mL, vancomycin can be considered; and if the vancomycin MIC is ≥8 μg/mL, penicillin or ampicillin can be considered. In the absence of high-level resistance, gentamicin or streptomycin should be used as the aminoglycoside (Tables 124–4). If there is high-level resistance to both these drugs, no aminoglycoside should be given; instead, an 8- to 12-week course of a single cell wall–active agent —or, for *E. faecalis*, high doses of ampicillin combined with ceftriaxone or cefotaxime—is suggested. If this alternative therapy fails or the isolate is resistant to all of the commonly used agents, surgical treatment is advised. The role of newer agents potentially active against multidrug-resistant enterococci [quinupristin/dalfopristin (*E. faecium* only), linezolid, and daptomycin] in the treatment of endocarditis has not been established. Although the dose of gentamicin used to achieve bactericidal synergy in treating enterococcal endocarditis is smaller than that used in standard therapy, nephrotoxicity is not uncommon during treatment for 4 –6 weeks. Regimens in which the aminoglycoside component is discontinued at2–3 weeks because of toxicity have been curative. Thus, discontinuation of the aminoglycoside is recommended when nephrotoxicity develops in patients who have responded satisfactorily to therapy. Alternatively, the ampicillin-ceftriaxone regimen can be used to treat *E. faecalis* endocarditis if nephrotoxicity develops or is exceptionally threatening.

Staphylococci

The regimens used to treat staphylococcal endocarditis (Tables 124–4) are based not on coagulase production but rather on the presence or absence of a prosthetic valve or foreign device, the native valve(s) involved, and the susceptibility of the isolate to penicillin, methicillin, and vancomycin. All staphylococci are considered penicillin-resistant until shown not to produce penicillinase. Similarly, methicillin resistance has become so prevalent among staphylococci that therapy should be initiated with a regimen for methicillin-resistant organisms and subsequently revised if the strain proves to be susceptible to methicillin. The addition of 3–5 days of gentamicin (if the isolate is susceptible) to a β-lactam antibiotic to enhance therapy for native mitral or aortic valve endocarditis has been optional. While the addition of gentamicin minimally hastens eradication of bacteremia, it does not improve survival rates, and even abbreviated gentamicin therapy may be associated with nephrotoxicity and thus is not recommended. Gentamicin generally is not added to the vancomycin regimen in this setting.

For treatment of endocarditis caused by methicillin-resistant *S. aureus* (MRSA), vancomycin dosing to achieve

trough concentrations of 15–20 μg/mL is recommended, with the recognition that this regimen may be associated with nephrotoxicity. Although resistance to vancomycin among staphylococci is rare, reduced vancomycin susceptibility among MRSA strains is increasingly encountered. Isolates with a vancomycin MIC of 4–16 μg/mL have intermediate susceptibility and are referred to as vancomycin-intermediate *S. aureus* (VISA). Isolates with an MIC of 2 μg/mL may harbor subpopulations with higher MICs. These isolates, called hetero-resistant VISA (hVISA), are not detectable by routine susceptibility testing. Because of the pharmacokinetics/pharmacodynamics of vancomycin, killing of MRSA with a vancomycin MIC of 2 μg/mL is unpredictable even with aggressive vancomycin dosing. Although not approved by the U.S. Food and Drug Administration, daptomycin [6 mg/kg (or, as some experts prefer, 8–10 mg/kg) IV once daily] has been recommended as an alternative to vancomycin, particularly for endocarditis caused by VISA, hVISA, and isolates with a vancomycin MIC of2 μg/mL. These isolates should be tested to document daptomycin susceptibility. Treatment of endocarditis in which bacteremia persists despite this therapy is beyond the scope of this chapter and requires consultation with an infectious disease specialist. The efficacy of linezolid for left-sided MRSA endocarditis has not been established.

Methicillin-susceptible *S. aureus* endocarditis that is uncomplicated and limited to the tricuspid or pulmonic valve—a condition occurring almost exclusively in injection drug users—can often be treated with a 2-week course that combines oxacillin or nafcillin (but not vancomycin) with gentamicin. Patients with prolonged fever (≥5 days) during therapy or multiple septic pulmonary emboli should receive standard therapy. Right-sided endocarditis caused by MRSA is treated for 4 weeks with a standard vancomycin regimen or with daptomycin (6 mg/kg as a single daily dose).

Staphylococcal PVE is treated for 6–8 weeks with a multidrug regimen. Rifampin is an essential component because it kills staphylococci that are adherent to foreign material in a biofilm. Two other agents (selected on the basis of susceptibility testing) are combined with rifampin to prevent in vivo emergence of resistance. Because many staphylococci (particularly MRSA and *S. epidermidis*) are resistant to gentamicin, susceptibility to gentamicin or an alternative agent should be established before rifampin treatment is begun. If the isolate is resistant to gentamicin, then another aminoglycoside, a fluoroquinolone (chosen on the basis of susceptibility), or another active agent should be substituted for gentamicin.

Other Organisms

In the absence of meningitis, endocarditis caused by *S. pneumoniae* with a penicillin MIC of ≤1 μg/mL can be treated with IV penicillin (4 million units every 4 h), ceftriaxone (2 g/d as a single dose), or cefotaxime (at a comparable dosage). Infection caused by pneumococcal strains with a penicillin MIC of ≥2 μg/mL should be treated with vancomycin. Until the strain's susceptibility to penicillin is established, therapy should consist of vancomycin plus ceftriaxone, especially if concurrent meningitis is suspected. *P. aeruginosa* endocarditis is treated with an antipseudomonal penicillin (ticarcillin or piperacillin) and high doses of tobramycin (8 mg/kg per day in three divided doses). Endocarditis caused by Enterobacteriaceae is treated with a potent β-lactam antibiotic plus an aminoglycoside. Corynebacterial endocarditis is treated with penicillin plus an aminoglycoside (if the organism is susceptible to the aminoglycoside) or with vancomycin, which is highly bactericidal for most strains. Therapy for *Candida* endocarditis consists of amphotericin B plus flucytosine and early surgery; long-term (if not indefinite) suppression with an oral azole is advised. Caspofungin treatment of *Candida* endocarditis has been effective in sporadic cases; nevertheless, the role of echinocandins in this setting has not been established.

Empirical Therapy

In the design and execution of therapy without culture data (i.e., before culture results are known or when cultures are negative), clinical clues (e.g., site of infection, patient's predispositions) as well as epidemiologic clues to etiology must be considered. Thus, empirical therapy for acute endocarditis in an injection drug user should cover MRSA and gram-negative bacilli. Treatment with vancomycin plus gentamicin, initiated immediately after blood is obtained for cultures, covers these as well as many other potential causes. Similarly, treatment of health care–associated endocarditis must cover MRSA. In the treatment of culture-negative episodes, marantic endocarditis must be excluded and fastidious organisms sought by serologic testing. In the absence of prior antibiotic therapy, it is unlikely that *S. aureus*, CoNS, or enterococcal infection will present with negative blood cultures; thus, in this situation, recommended empirical therapy targets not these organisms but rather nutritionally variant organisms, the HACEK group, and *Bartonella* species. Pending the availability of diagnostic data, blood culture–negative subacute NVE is treated either with ampicillin-sulbactam (12 g every 24 h) or with ceftriaxone plus gentamicin; doxycycline (100 mg twice daily) is added for *Bartonella* coverage. Vancomycin, gentamicin, cefepime, and rifampin should be used if prosthetic valves in place for ≥1 year are involved. Empirical therapy for infected prosthetic valves in place for ≤1 year is similar to that for culture-negative PVE. If negative cultures have been confounded by prior antibiotic administration, broader empirical therapy may be indicated, with particular attention to pathogens likely to be inhibited by the specific prior therapy.

Outpatient Antimicrobial Therapy

Fully compliant patients who have sterile blood cultures, no fever, and no clinical or echocardiographic findings that suggest an impending complication may complete therapy as outpatients. Careful follow-up and a stable home setting are necessary, as are predictable IV access and use of antimicrobial agents that are stable in solution.

Monitoring Antimicrobial Therapy

The serum bactericidal titer—the highest dilution of the patient's serum during therapy that kills 99.9% of the standard inoculum of the infecting organism—is no longer recommended for assessment of standard regimens. However, in the treatment of endocarditis caused by unusual organisms, this measurement may provide a patient-specific assessment of in vivo antibiotic effect. Serum concentrations of aminoglycosides and vancomycin should be monitored.

Antibiotic toxicities, including allergic reactions, occur in 25–40% of patients and commonly arise during the third week of therapy. Blood tests to detect renal, hepatic, and hematologic toxicity should be performed periodically.

Blood cultures should be repeated daily until sterile, rechecked if there is recrudescent fever, and performed again 4–6 weeks after therapy to document cure. Blood cultures become sterile within 2 days after the start of appropriate therapy when infection is caused by viridans streptococci, enterococci, or HACEK organisms. In *S. aureus* endocarditis, β-lactam therapy results in sterile cultures in 3–5 days, whereas with MRSA endocarditis positive cultures may persist for 7–9 days with vancomycin treatment. MRSA bacteremia persisting despite an adequate dosage of vancomycin may indicate infection due to a strain with reduced vancomycin susceptibility and therefore may point to a need for alternative therapy. When fever persists for 7 days despite appropriate antibiotic therapy, patients should be evaluated for paravalvular abscess, extracardiac abscesses (spleen, kidney), or complications (embolic events). Recrudescent fever raises the question of these complications but also of drug reactions or complications of hospitalization. Vegetations become smaller with effective therapy; however, 3 months after cure, 50% are unchanged and 25% are slightly larger.

SURGICAL TREATMENT

Intracardiac and central nervous system complications of endocarditis are important causes of morbidity and death. In some cases, effective treatment for these complications requires surgery. The indications for cardiac surgical treatment of endocarditis **(Tables 124–5)** have been derived from observational studies and expert opinion. The strength of individual indications vary; thus, the risks and benefits as well as the timing of surgery must be individualized **(Tables 124–6)**. From 25% to 40% of patients with left-sided endocarditis undergo cardiac surgery during active infection, with slightly higher surgery rates with PVE than with NVE. Clinical events resulting from intracardiac complications, which are most reliably detected by TEE, justify most surgery. In the absence of randomized trials to evaluate a survival benefit for surgical intervention, the effect of surgery has been assessed in studies comparing populations of medically and surgically treated patients matched for the necessity of surgery (indication), with adjustments for predictors of death (comorbidity) and time of the surgical intervention. Although study results vary, surgery for currently advised indications appears to convey a significant survival benefit (27–55%) that becomes apparent only with follow-up for ≥6 months after the intervention. During the initial weeks after surgery, mortality risk is actually increased (disease- plus surgery-related mortality). With less demanding surgical indications, this combined mortality risk may erode potential long-term benefits. Benefit is greatest for NVE complicated by heart failure or myocardial abscess and is less clear for PVE; this difference may reflect sample size in the relevant studies.

Tables 124–5 Indications for Cardiac Surgical Intervention in Patients with Endocarditis

Surgery required for optimal outcome
Moderate to severe congestive heart failure due to valve dysfunction
Partially dehisced unstable prosthetic valve
Persistent bacteremia despite optimal antimicrobial therapy
Lack of effective microbicidal therapy (e.g., fungal or *Brucella* endocarditis)
S. aureus prosthetic valve endocarditis with an intracardiac complication
Relapse of prosthetic valve endocarditis after optimal antimicrobial therapy
Surgery to be strongly considered for improved outcome[a]
Perivalvular extension of infection
Poorly responsive *S. aureus* endocarditis involving the aortic or mitral valve
Large (>10-mm diameter) hypermobile vegetations with increased risk of embolism
Persistent unexplained fever (≥10 days) in culture-negative native valve endocarditis
Poorly responsive or relapsed endocarditis due to highly antibiotic-resistant enterococci or gram-negative bacilli

[a] Surgery must be carefully considered; findings are often combined with other indications to prompt surgery.

Tables 124–6 Timing of Cardiac Surgical Intervention in Patients with Endocarditis

	Indication for Surgical Intervention	
Timing	**Strong Supporting Evidence**	**Conflicting Evidence, but Majority of Opinions Favor Surgery**
Emergent (same day)	Acute aortic regurgitation plus preclosure of mitral valve Sinus of Valsalva abscess ruptured into right heart Rupture into pericardial sac	
Urgent (within 1–2 days)	Valve obstruction by vegetation Unstable (dehisced) prosthesis Acute aortic or mitral regurgitation with heart failure (New York Heart Association class III or IV)	Major embolus plus persisting large vegetation (>10 mm in diameter)

	Septal perforation Perivalvular extension of infection with/without new electrocardiographic conduction system changes Lack of effective antibiotic therapy	
Elective (earlier usually preferred)	Progressive paravalvular prosthetic regurgitation Valve dysfunction plus persisting infection after ≥7–10 days of antimicrobial therapy Fungal (mold) endocarditis	Staphylococcal PVE Early PVE (≤2 months after valve surgery) Fungal endocarditis (*Candida* spp.) Antibiotic-resistant organisms

Note: PVE, prosthetic valve endocarditis.

Source: Adapted from L Olaison, G Pettersson: Infect Dis Clin North Am 16:453, 2002.

Congestive Heart Failure

Moderate to severe refractory CHF caused by new or worsening valve dysfunction is the major indication for cardiac surgical treatment of endocarditis. At 6 months of follow-up, patients with left-sided endocarditis and moderate to severe heart failure due to valve dysfunction who are treated only medically have a 50% mortality rate; the figure is 15% among matched patients who undergo surgery. The survival benefit with surgery is seen in both NVE and PVE. Surgery can relieve functional stenosis due to large vegetations or restore competence to damaged regurgitant valves by repair or replacement.

Perivalvular Infection

This complication, which is most common with aortic valve infection, occurs in 10–15% of native valve and 45–60% of prosthetic valve infections. It is suggested by persistent unexplained fever during appropriate therapy, new electrocardiographic conduction disturbances, and pericarditis. TEE with color Doppler is the test of choice to detect perivalvular abscesses (sensitivity, ≥85%). For optimal outcome, surgery is required, especially when fever persists, fistulae develop, prostheses are dehisced and unstable, and invasive infection relapses after appropriate treatment. Cardiac rhythm must be monitored since high-grade heart block may require insertion of a pacemaker.

Uncontrolled Infection

Continued positive blood cultures or otherwise-unexplained persistent fevers (in patients with either blood culture–positive or –negative endocarditis) despite optimal antibiotic therapy may reflect uncontrolled infection and may warrant surgery. Surgical treatment is also advised for endocarditis caused by organisms for which experience indicates that effective antimicrobial therapy is lacking (e.g., yeasts, fungi, *P. aeruginosa*, other highly resistant gram-negative bacilli, *Brucella* species, and probably *C. burnetii*).

Aureus *Endocarditis*

The mortality rate for *S. aureus* PVE exceeds 50% with medical treatment but is reduced to 25% with surgical treatment. In patients with intracardiac complications associated with *S. aureus* PVE, surgical treatment reduces the mortality rate twentyfold. Surgical treatment should be considered for patients with *S. aureus* native aortic or mitral valve infection who have TTE-demonstrable vegetations and remain septic during the initial week of therapy. Isolated tricuspid valve endocarditis, even with persistent fever, rarely requires surgery.

Prevention of Systemic Emboli

Death and persisting morbidity due to emboli are largely limited to patients suffering occlusion of cerebral or coronary arteries. Echocardiographic determination of vegetation size and anatomy, although predictive of patients at high risk of systemic emboli, does not identify those patients in whom the benefits of surgery to prevent emboli clearly exceed the risks of the surgical procedure. Net benefits from surgery to prevent emboli are most likely when other surgical benefits can be achieved simultaneously—e.g., repair of a moderately dysfunctional valve or debridement of a paravalvular abscess. Only 3.5% of patients undergo surgery solely to prevent systemic emboli. Valve repair avoiding insertion of a prosthesis makes the benefit-to-risk ratio of surgery to address vegetations more favorable.

Timing of Cardiac Surgery

In general, when indications for surgical treatment of infective endocarditis are identified, surgery should not be delayed simply to permit additional antibiotic therapy, since this course of action increases the risk of death (Tables 124–6). After 14 days of recommended antibiotic therapy, excised valves are culture-negative in 99% and 50% of patients with streptococcal and *S. aureus* endocarditis, respectively. Recrudescent endocarditis on a new implanted prosthetic valve follows surgery for active NVE and PVE in 2% and 6–15% of patients, respectively. These frequencies do not justify the risk of adverse outcome with delayed surgery, particularly in patients with severe heart failure, valve dysfunction, and staphylococcal infections. Delay is justified only when infection is controlled and CHF is resolved with medical therapy.

Among patients who have experienced a neurologic complication of endocarditis, further neurologic deterioration can occur as a consequence of cardiac surgery. The risk of neurologic deterioration is related to the type of neurologic complication and the interval between the complication and surgery. Whenever feasible, cardiac surgery should be delayed for 2–3 weeks after a nonhemorrhagic embolic infarction and for 4 weeks after a cerebral hemorrhage. A ruptured mycotic aneurysm should be treated before cardiac surgery.

Antibiotic Therapy after Cardiac Surgery

Bacteria visible in Gram-stained preparations of excised valves do not necessarily indicate a failure of antibiotic therapy. Organisms have been detected on Gram's stain—or their DNA has been detected by PCR—in excised

valves from 45% of patients who have successfully completed the recommended therapy for endocarditis. In only 7% of these patients are the organisms, most of which are unusual and antibiotic resistant, cultured from the valve. Despite the detection of organisms or their DNA, relapse of endocarditis after surgery is uncommon. Thus, when valve cultures are negative in uncomplicated NVE caused by susceptible organisms, the duration of preoperative plus postoperative treatment should equal the total duration of recommended therapy, with ~2 weeks of treatment administered after surgery. For endocarditis complicated by paravalvular abscess, partially treated PVE, or cases with culture-positive valves, a full course of therapy should be given postoperatively.

Extracardiac Complications

Splenic abscess develops in 3–5% of patients with endocarditis. Effective therapy requires either image-guided percutaneous drainage or splenectomy. Mycotic aneurysms occur in 2–15% of endocarditis patients; one-half of these cases involve the cerebral arteries and present as headaches, focal neurologic symptoms, or hemorrhage. Cerebral aneurysms should be monitored by angiography. Some will resolve with effective antimicrobial therapy, but those that persist, enlarge, or leak should be treated surgically if possible. Extracerebral aneurysms present as local pain, a mass, local ischemia, or bleeding; these aneurysms are treated by resection.

Outcome

Older age, severe comorbid conditions and diabetes, delayed diagnosis, involvement of prosthetic valves or the aortic valve, an invasive (*S. aureus*) or antibiotic-resistant (*P. aeruginosa*, yeast) pathogen, intracardiac and major neurologic complications, and an association with health care adversely affect outcome. Death and poor outcome often are related not to failure of antibiotic therapy but rather to the interactions of comorbidities and endocarditis-related end-organ complications. Overall survival rates for patients with NVE caused by viridans streptococci, HACEK organisms, or enterococci (susceptible to synergistic therapy) are 85–90%. For *S. aureus* NVE in patients who do not inject drugs, survival rates are 55–70%, whereas 85–90% of injection drug users survive this infection. PVE beginning within 2 months of valve replacement results in mortality rates of 40–50%, whereas rates are only 10–20% in later-onset cases.

Prevention

In the past, in an effort to prevent endocarditis (long a goal in clinical practice), expert committees have supported systemic antibiotic administration prior to many bacteremia-inducing procedures. In the absence of human trials, a reappraisal of the indirect evidence for antibiotic prophylaxis for endocarditis by the American Heart Association has culminated in guidelines that reverse prior recommendations and restrict prophylactic antibiotic use. At best, the benefit of antibiotic prophylaxis is minimal. Most endocarditis cases do not follow a procedure. In case-control studies, dental treatments—widely considered as predisposing to endocarditis—occur no more frequently before endocarditis than in matched controls. Furthermore, the frequency and magnitude of bacteremia associated with dental procedures and routine daily activities (e.g., tooth brushing and flossing) are similar. Because dental procedures are infrequent, exposure of cardiac structures to bacteremic oral-cavity organisms is notably greater from routine daily activities than from dental care. The relation of gastrointestinal and genitourinary procedures to subsequent endocarditis is more tenuous than that of dental procedures. In addition, cost-effectiveness and cost-benefit estimates suggest that antibiotic prophylaxis represents a poor use of resources.

Studies in animal models suggest that antibiotic prophylaxis may be effective. Thus it is possible that rare cases of endocarditis are prevented. Weighing the potential benefits, potential adverse events, and costs associated with antibiotic prophylaxis, the American Heart Association and the European Society of Cardiology now recommend prophylactic antibiotics **(Tables 124–7)** only for those patients at highest risk for severe morbidity or death from endocarditis **(Tables 124–8)**. Maintaining good dental hygiene is essential. Prophylaxis is recommended only when there is manipulation of gingival tissue or the periapical region of the teeth or perforation of the oral mucosa (including surgery on the respiratory tract). Prophylaxis is not advised for patients undergoing gastrointestinal or genitourinary tract procedures. High-risk patients should be treated before or when they undergo procedures on an infected genitourinary tract or on infected skin and soft tissue. The British Society for Antimicrobial Chemotherapy continues to recommend prophylaxis for at-risk patients undergoing selected gastrointestinal and genitourinary procedures. In contrast, the National Institute for Health and Clinical Excellence in the United Kingdom found no convincing evidence that antibiotic prophylaxis was cost effective and advised discontinuation of the practice (see *www.nice.org.uk/guidance/CG64*).

Tables 124–7 Antibiotic Regimens for Prophylaxis of Endocarditis in Adults with High-Risk Cardiac Lesions[a,b]

A. Standard oral regimen
1. Amoxicillin: 2 g PO 1 h before procedure
B. Inability to take oral medication
1. Ampicillin: 2 g IV or IM within 1 h before procedure
C. Penicillin allergy
1. Clarithromycin or azithromycin: 500 mg PO 1 h before procedure
2. Cephalexin[c]: 2 g PO 1 h before procedure
3. Clindamycin: 600 mg PO 1 h before procedure
D. Penicillin allergy, inability to take oral medication

1. Cefazolin[c] or ceftriaxone[c]: 1 g IV or IM 30 min before procedure
2. Clindamycin: 600 mg IV or IM 1 h before procedure

[a] Dosing for children: for amoxicillin, ampicillin, cephalexin, or cefadroxil, use 50 mg/kg PO; cefazolin, 25 mg/kg IV; clindamycin, 20 mg/kg PO, 25 mg/kg IV; clarithromycin, 15 mg/kg PO; and vancomycin, 20 mg/kg IV.

[b] For high-risk lesions, see Tables 124–8. Prophylaxis is not advised for other lesions.

[c] Do not use cephalosporins in patients with immediate hypersensitivity (urticaria, angioedema, anaphylaxis) to penicillin.

Source: W Wilson et al: Circulation, published online 4/19/2007.

Tables 124–8 High-Risk Cardiac Lesions for Which Endocarditis Prophylaxis Is Advised before Dental Procedures
Prosthetic heart valves
Prior endocarditis
Unrepaired cyanotic congenital heart disease, including palliative shunts or conduits
Completely repaired congenital heart defects during the 6 months after repair
Incompletely repaired congenital heart disease with residual defects adjacent to prosthetic material
Valvulopathy developing after cardiac transplantation

Source: W Wilson et al: Circulation, published online 4/19/2007.

FURTHER READINGS

Aksoy O et al: Early surgery in patients with infective endocarditis: A propensity score analysis. Clin Infect Dis 44:364, 2007[PMID: 17205442]

Baddour LM et al: Diagnosis, antimicrobial therapy, and management of complications. A statement for healthcare professionals from the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease, Council on Cardiovascular Disease in the Young, and the Councils on Clinical Cardiology, Stroke, and Cardiovascular Surgery and Anesthesia, American Heart Association. Circulation 111:e394, 2005

Bannay A et al: The impact of valve surgery on short- and long-term mortality in left-sided infective endocarditis: Do differences in methodological approaches explain previous conflicting results? Eur Heart J epub ahead of print, Feb 9, 2009 (http://eurheartj.oxfordjournal.org/content/early/2009/02/09/eurheartj.ehp008)

Benito N et al: Health care–associated native valve endocarditis: Importance of non-nosocomial acquisition. Ann Intern Med 150:586, 2009[PMID: 19414837]

Cosgrove SE et al: Initial low-dose gentamicin for *Staphylococcus aureus* bacteremia and endocarditis is nephrotoxic. Clin Infect Dis 48:713, 2009[PMID: 19207079]

Durack DT: Prevention of infective endocarditis, in *Principles and Practice of Infectious Diseases,* 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 1143–1151

Fowler VG Jr et al: Endocarditis and intravascular infections, in *Principles and Practice of Infectious Diseases,* 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 1067–1112

Habbib G et al: Guidelines on the prevention, diagnosis, and treatment of infective endocarditis (new version 2009). Eur Heart J 30:2369, 2009

Murdoch DR et al: Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century. Arch Intern Med 169:463, 2009[PMID: 19273776]

Rybak MJ et al: Vancomycin therapeutic guidelines: A summary of consensus recommendations from the Infectious Diseases Society of America, the American Society of Health-System Pharmacists, and the Society of Infectious Diseases Pharmacists. Clin Infect Dis 49:325, 2009[PMID: 19569969]

Thuny F et al: The timing of surgery influences mortality and morbidity in adults with severe complicated infective endocarditis: A propensity analysis. Eur Heart J epub ahead of print, March 26, 2009 (http://eurheartj.oxfordjournal.org/content/early/2009/03/26/eurheartj.ehp089)

Wilson W et al: Prevention of infective endocarditis. Guidelines from the American Heart Association. A guideline from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee, Council on Cardiovascular Disease in the Young, and the Council on Clinical Cardiology, Council on Cardiovascular Surgery and Anesthesia, and the Quality of Care and Outcomes Research Interdisciplinary Working Group. Circulation 116:1736, 2007[PMID: 17446442]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS OF THE SKIN, MUSCLES, AND SOFT TISSUES: INTRODUCTION

Anatomic Relationships: Clues to the Diagnosisof Soft Tissue Infections

Skin and soft tissue infections have been common human afflictions for centuries. However, between 2000 and 2004, hospital admissions for skin and soft tissue infections rose by 27%, a remarkable increase that was attributable largely to the emergence of the USA300 clone of methicillin-resistant *Staphylococcus aureus* (MRSA). This chapter provides an anatomic approach to understanding the types of soft tissue infections and the diverse microbes responsible.

Protection against infection of the epidermis depends on the mechanical barrier afforded by the stratum corneum, since the epidermis itself is devoid of blood vessels **(Fig. 125-1)**. Disruption of this layer by burns or bites, abrasions, foreign bodies, primary dermatologic disorders (e.g., herpes simplex, varicella, ecthyma gangrenosum), surgery, or vascular or pressure ulcer allows penetration of bacteria to the deeper structures. Similarly, the hair follicle can serve as a portal either for components of the normal flora (e.g., *Staphylococcus*) or for extrinsic bacteria (e.g., *Pseudomonas* in hot-tub folliculitis). Intracellular infection of the squamous epithelium with vesicle formation may arise from cutaneous inoculation, as in infection with herpes simplex virus (HSV) type 1; from the dermal capillary plexus, as in varicella and infections due to other viruses associated with viremia; or from cutaneous nerve roots, as in herpes zoster. Bacteria infecting the epidermis, such as *Streptococcus pyogenes*, may be translocated laterally to deeper structures via lymphatics, an event that results in the rapid superficial spread of erysipelas. Later, engorgement or obstruction of lymphatics causes flaccid edema of the epidermis, another characteristic of erysipelas.

Figure 125-1

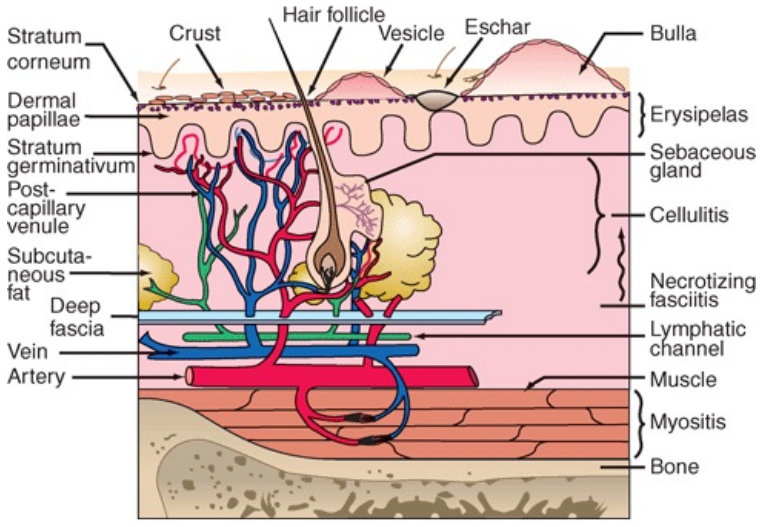

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Structural components of the skin and soft tissue, superficial infections, and infections of the deeper structures. The rich capillary network beneath the dermal papillae plays a key role in the localization of infection and in the development of the acute inflammatory reaction.

The rich plexus of capillaries beneath the dermal papillae provides nutrition to the stratum germinativum, and physiologic responses of this plexus produce important clinical signs and symptoms. For example, infective vasculitis of the plexus results in petechiae, Osler's nodes, Janeway lesions, and palpable purpura, which, if present, are important clues to the existence of endocarditis (Chap. 124). In addition, metastatic infection within this plexus can result in cutaneous manifestations of disseminated fungal infection (Chap. 203), gonococcal infection (Chap. 144), *Salmonella* infection (Chap. 153), *Pseudomonas* infection (i.e., ecthyma gangrenosum; Chap. 152), meningococcemia (Chap. 143), and staphylococcal infection (Chap. 135). The plexus also provides bacteria with access to the circulation, thereby facilitating local spread or bacteremia. The postcapillary venules of this

plexus are a major site of polymorphonuclear leukocyte sequestration, diapedesis, and chemotaxis to the site of cutaneous infection.

Exaggeration of these physiologic mechanisms by excessive levels of cytokines or bacterial toxins causes leukostasis, venous occlusion, and pitting edema. Edema with purple bullae, ecchymosis, and cutaneous anesthesia suggests loss of vascular integrity and necessitates exploration of the deeper structures for evidence of necrotizing fasciitis or myonecrosis. An early diagnosis requires a high level of suspicion in instances of unexplained fever and of pain and tenderness in the soft tissue, even in the absence of acute cutaneous inflammation.

Table 125–1 indicates the chapters in which the infections described below are discussed in greater detail. Many of these infections are illustrated in the chapters cited or in Chap. e7 (Atlas of Rashes Associated With Fever).

Table 125–1 Skin and Soft Tissue Infections

Lesion, Clinical Syndrome	Infectious Agent	Chapter(s)
Vesicles		
Smallpox	Variola virus	221
Chickenpox	Varicella-zoster virus	180
Shingles (herpes zoster)	Varicella-zoster virus	180
Cold sores, herpetic whitlow, herpes gladiatorum	Herpes simplex virus	179
Hand-foot-and-mouth disease	Coxsackievirus A16	191
Orf	Parapoxvirus	183
Molluscum contagiosum	Pox-like virus	183
Rickettsialpox	*Rickettsia akari*	174
Blistering distal dactylitis	*Staphylococcus aureus* or *Streptococcus pyogenes*	135, 136
Bullae		
Staphylococcal scalded-skin syndrome	*S. aureus*	135
Necrotizing fasciitis	*S. pyogenes, Clostridium* spp., mixed aerobes and anaerobes	136, 142, 164
Gas gangrene	*Clostridium* spp.	142
Halophilic vibrio	*Vibrio vulnificus*	156
Crusted lesions		
Bullous impetigo/ecthyma	*S. aureus*	135
Impetigo contagiosa	*S. pyogenes*	136
Ringworm	Superficial dermatophyte fungi	206
Sporotrichosis	*Sporothrix schenckii*	206
Histoplasmosis	*Histoplasma capsulatum*	199
Coccidioidomycosis	*Coccidioides immitis*	200
Blastomycosis	*Blastomyces dermatitidis*	201
Cutaneous leishmaniasis	*Leishmania* spp.	212
Cutaneous tuberculosis	*Mycobacterium tuberculosis*	165
Nocardiosis	*Nocardia asteroides*	162
Folliculitis		
Furunculosis	*S. aureus*	135
Hot-tub folliculitis	*Pseudomonas aeruginosa*	152
Swimmer's itch	*Schistosoma* spp.	219
Acne vulgaris	*Propionibacterium acnes*	52
Papular and nodular lesions		
Fish-tank or swimming-pool granuloma	*Mycobacterium marinum*	167
Creeping eruption (cutaneous larva migrans)	*Ancylostoma braziliense*	216
Dracunculiasis	*Dracunculus medinensis*	218
Cercarial dermatitis	*Schistosoma mansoni*	219
Verruca vulgaris	Human papillomaviruses 1, 2, 4	185

Condylomata acuminata (anogenital warts)	Human papillomaviruses 6, 11, 16, 18	185
Onchocerciasis nodule	*Onchocerca volvulus*	218
Cutaneous myiasis	*Dermatobia hominis*	397
Verruca peruana	*Bartonella bacilliformis*	160
Cat-scratch disease	*Bartonella henselae*	160
Lepromatous leprosy	*Mycobacterium leprae*	166
Secondary syphilis (papulosquamous and nodular lesions, condylomata lata)	*Treponema pallidum*	169
Tertiary syphilis (nodular gummatous lesions)	*T. pallidum*	169
Ulcers with or without eschars		
Anthrax	*Bacillus anthracis*	221
Ulceroglandular tularemia	*Francisella tularensis*	158, 221
Bubonic plague	*Yersinia pestis*	159, 221
Buruli ulcer	*Mycobacterium ulcerans*	167
Leprosy	*M. leprae*	166
Cutaneous tuberculosis	*M. tuberculosis*	165
Chancroid	*Haemophilus ducreyi*	145
Primary syphilis	*T. pallidum*	169
Erysipelas	*S. pyogenes*	136
Cellulitis	*Staphylococcus* spp., *Streptococcus* spp., various other bacteria	Various
Necrotizing fasciitis		
Streptococcal gangrene	*S. pyogenes*	136
Fournier's gangrene	Mixed aerobic and anaerobic bacteria	164
Staphylococcal necrotizing fasciitis	Methicillin-resistant *S. aureus*	135
Myositis and myonecrosis		
Pyomyositis	*S. aureus*	135
Streptococcal necrotizing myositis	*S. pyogenes*	136
Gas gangrene	*Clostridium* spp.	142
Nonclostridial (crepitant) myositis	Mixed aerobic and anaerobic bacteria	164
Synergistic nonclostridial anaerobic myonecrosis	Mixed aerobic and anaerobic bacteria	164

Infections Associated with Vesicles

(Table 125–1) Vesicle formation due to infection is caused by viral proliferation within the epidermis. In varicella and variola, viremia precedes the onset of a diffuse centripetal rash that progresses from macules to vesicles, then to pustules, and finally to scabs over the course of 1–2 weeks. Vesicles of varicella have a "dewdrop" appearance and develop in crops randomly about the trunk, extremities, and face over 3–4 days. Herpes zoster occurs in a single dermatome; the appearance of vesicles is preceded by pain for several days. Zoster may occur in persons of any age but is most common among immunosuppressed individuals and elderly patients, whereas most cases of varicella occur in young children. Vesicles due to HSV are found on or around the lips (HSV-1) or genitals (HSV-2) but may appear on the head and neck of young wrestlers (herpes gladiatorum) or on the digits of health care workers (herpetic whitlow). Recurrent herpes labialis (HSV-1) and herpes genitalis commonly follow primary infection. Coxsackievirus A16 characteristically causes vesicles on the hands, feet, and mouth of children. Orf is caused by a DNA virus related to smallpox virus and infects the fingers of individuals who work around goats and sheep. Molluscum contagiosum virus induces flaccid vesicles on the skin of healthy and immunocompromised individuals. Although variola (smallpox) in nature was eradicated as of 1977, recent terrorist events have renewed interest in this devastating infection (Chap. 221). Viremia beginning after an incubation period of 12 days is followed by a diffuse maculopapular rash, with rapid evolution to vesicles, pustules, and then scabs. Secondary cases can occur among close contacts.

Rickettsialpox begins after mite-bite inoculation of *Rickettsia akari* into the skin. A papule with a central vesicle evolves to form a 1- to 2.5-cm painless crusted black eschar with an erythematous halo and proximal adenopathy. While more common in the northeastern United States and the Ukraine in 1940–1950, rickettsialpox has recently been described in Ohio, Arizona, and Utah. Blistering dactylitis is a painful, vesicular, localized *S. aureus* or group A streptococcal infection of the pulps of the distal digits of the hands.

Infections Associated with Bullae

(Table 125–1) Staphylococcal scalded-skin syndrome (SSSS) in neonates is caused by a toxin (exfoliatin) from phage group II*S. aureus*. SSSS must be distinguished from toxic epidermal necrolysis (TEN), which occurs primarily in adults, is drug-induced, and is associated with a higher mortality rate. Punch biopsy with frozen section is useful in making this distinction since the cleavage plane is the stratum corneum in SSSS and the stratum germinativum in TEN (Fig. 125-1). Intravenous γ-globulin is a promising treatment for TEN. Necrotizing fasciitis and gas gangrene also induce bulla formation (see "Necrotizing Fasciitis," below). Halophilic vibrio infection can be as aggressive and fulminant as necrotizing fasciitis; a helpful clue in its diagnosis is a history of exposure to waters of the Gulf of Mexico or the Atlantic seaboard or (in a patient with cirrhosis) the ingestion of raw seafood. The etiologic organism (*Vibrio vulnificus*) is highly susceptible to tetracycline.

Infections Associated with Crusted Lesions

(Table 125–1) Impetigo contagiosa is caused by *S. pyogenes*, and bullous impetigo is due to *S. aureus*. Both skin lesions may have an early bullous stage but then appear as thick crusts with a golden-brown color. Epidemics of impetigo caused by MRSA have been reported. Streptococcal lesions are most common among children 2–5 years of age, and epidemics may occur in settings of poor hygiene, particularly among children in lower socioeconomic settings in tropical climates. It is important to recognize impetigo contagiosa because of its relationship to poststreptococcal glomerulonephritis. Rheumatic fever is not a complication of skin infection caused by*S. pyogenes*. Superficial dermatophyte infection (ringworm) can occur on any skin surface, and skin scrapings with KOH staining are diagnostic. Primary infections with dimorphic fungi such as *Blastomyces dermatitidis* and *Sporothrix schenckii* can initially present as crusted skin lesions resembling ringworm. Disseminated infection with *Coccidioides immitis* can also involve the skin, and biopsy and culture should be performed on crusted lesions in patients from endemic areas. Crusted nodular lesions caused by *Mycobacterium chelonei* have been described in HIV-seropositive patients. Treatment with clarithromycin looks promising.

Folliculitis

(Table 125–1) Hair follicles serve as portals for a number of bacteria, although *S. aureus* is the most common cause of localized folliculitis. Sebaceous glands empty into hair follicles and ducts and, if these portals are blocked, form sebaceous cysts that may resemble staphylococcal abscesses or may become secondarily infected. Infection of sweat glands (hidradenitis suppurativa) can also mimic infection of hair follicles, particularly in the axillae. Chronic folliculitis is uncommon except in acnevulgaris, where constituents of the normal flora (e.g., *Propionibacterium acnes*) may play a role.

Diffuse folliculitis occurs in two settings. Hot-tub folliculitis is caused by *Pseudomonas aeruginosa* in waters that are insufficiently chlorinated and maintained at temperatures of 37–40°C. Infection is usually self-limited, although bacteremia and shock have been reported. Swimmer's itch occurs when a skin surface is exposed to water infested with freshwater avian schistosomes. Warm water temperatures and alkaline pH are suitable for mollusks that serve as intermediate hosts between birds and humans. Free-swimming schistosomal cercariae readily penetrate human hair follicles or pores but quickly die and elicit a brisk allergic reaction, causing intense itching and erythema.

Papular and Nodular Lesions

(Table 125–1) Raised lesions of the skin occur in many different forms. *Mycobacterium marinum* infections of the skin may present as cellulitis or as raised erythematous nodules. Erythematous papules are early manifestations of cat-scratch disease (with lesions developing at the primary site of inoculation of *Bartonella henselae*) and bacillary angiomatosis (also caused by *B. henselae*). Raised serpiginous or linear eruptions are characteristic of cutaneous larva migrans, which is caused by burrowing larvae of dog or cat hookworms (*Ancylostoma braziliense*) and which humans acquire through contact with soil that has been contaminated with dog or cat feces. Similar burrowing raised lesions are present in dracunculiasis caused by migration of the adult female nematode *Dracunculus medinensis*. Nodules caused by *Onchocerca volvulus* measure 1–10 cm in diameter and occur mostly in persons bitten by *Simulium* flies in Africa. The nodules contain the adult worm encased in fibrous tissue. Migration of microfilariae into the eyes may result in blindness. Verruga peruana is caused by *Bartonella bacilliformis*, which is transmitted to humans by the sandfly *Phlebotomus*. This condition can take the form of single gigantic lesions (several centimeters in diameter) or multiple small lesions (several millimeters in diameter). Numerous subcutaneous nodules may also be present in cysticercosis caused by larvae of *Taenia solium*. Multiple erythematous papules develop in schistosomiasis; each represents a cercarial invasion site. Skin nodules as well as thickened subcutaneous tissue are prominent features of lepromatous leprosy. Large nodules or gummas are features of tertiary syphilis, whereas flat papulosquamous lesions are characteristic of secondary syphilis. Human papillomavirus may cause singular warts (verruca vulgaris) or multiple warts in the anogenital area (condylomata acuminata). The latter are major problems in HIV-infected individuals.

Ulcers with or Without Eschars

(Table 125–1) Cutaneousanthrax begins as a pruritic papule, which develops within days into an ulcer with surrounding vesicles and edema and then into an enlarging ulcer with a black eschar. Cutaneous anthrax may cause chronic nonhealing ulcers with an overlying dirty-gray membrane, although lesions may also mimic psoriasis, eczema, or impetigo. Ulceroglandular tularemia may have associated ulcerated skin lesions with painful regional adenopathy. Although buboes are the major cutaneous manifestation of plague, ulcers with eschars, papules, or pustules are also present in 25% of cases.

Mycobacterium ulcerans typically causes chronic skin ulcers on the extremities of individuals living in the tropics. *Mycobacterium leprae* may be associated with cutaneous ulcerations in patients with lepromatous leprosy related to Lucio's phenomenon, in which immune-mediated destruction of tissue bearing high concentrations of *M. leprae* bacilli occurs, usually several months after initiation of effective therapy. *Mycobacterium tuberculosis* may also cause ulcerations, papules, or erythematous macular lesions of the skin in both normal and immunocompromised patients.

Decubitus ulcers are due to tissue hypoxemia secondary to pressure-induced vascular insufficiency and may become secondarily infected with components of the skin and gastrointestinal flora, including anaerobes. Ulcerative lesions on the anterior shins may be due to pyoderma gangrenosum, which must be distinguished from similar lesions of infectious etiology by histologic evaluation of biopsy sites. Ulcerated lesions on the genitals may be either painful (chancroid) or painless (primary syphilis).

Erysipelas

(Table 125–1) Erysipelas is due to *S. pyogenes* and is characterized by an abrupt onset of fiery-red swelling of the face or extremities. The distinctive features of erysipelas are well-defined indurated margins, particularly along the nasolabial fold; rapid progression; and intense pain. Flaccid bullae may develop during the second or third day of illness, but extension to deeper soft tissues is rare. Treatment with penicillin is effective; swelling may progress despite appropriate treatment, although fever, pain, and the intense red color diminish. Desquamation of the involved skin occurs 5–10 days into the illness. Infants and elderly adults are most commonly afflicted, and the severity of systemic toxicity varies.

Cellulitis

(Table 125–1) Cellulitis is an acute inflammatory condition of the skin that is characterized by localized pain, erythema, swelling, and heat. It may be caused by indigenous flora colonizing the skin and appendages (e.g., *S. aureus* and *S. pyogenes*) or by a wide variety of exogenous bacteria. Because the exogenous bacteria involved in cellulitis occupy unique niches in nature, a thorough history (including epidemiologic data) provides important clues to etiology. When there is drainage, an open wound, or an obvious portal of entry, Gram's stain and culture provide a definitive diagnosis. In the absence of these findings, the bacterial etiology of cellulitis is difficult to establish, and in some cases staphylococcal and streptococcal cellulitis may have similar features. Even with needle aspiration of the leading edge or a punch biopsy of the cellulitis tissue itself, cultures are positive in only 20% of cases. This observation suggests that relatively low numbers of bacteria may cause cellulitis and that the expanding area of erythema within the skin may be a direct effect of extracellular toxins or of the soluble mediators of inflammation elicited by the host.

Bacteria may gain access to the epidermis through cracks in the skin, abrasions, cuts, burns, insect bites, surgical incisions, and IV catheters. Cellulitis caused by *S. aureus* spreads from a central localized infection, such as an abscess, folliculitis, or an infected foreign body (e.g., a splinter, a prosthetic device, or an IV catheter). MRSA is rapidly replacing methicillin-sensitive *S. aureus* (MSSA) as a cause of cellulitis in both inpatient and outpatient settings. Cellulitis caused by MSSA or MRSA is usually associated with a focal infection, such as a furuncle, a carbuncle, a surgical wound, or an abscess. In contrast, cellulitis due to *S. pyogenes* is a more rapidly spreading, diffuse process that is frequently associated with lymphangitis and fever. Recurrent streptococcal cellulitis of the lower extremities may be caused by organisms of group A, C, or G in association with chronic venous stasis or with saphenous venectomy for coronary artery bypass surgery. Streptococci also cause recurrent cellulitis among patients with chronic lymphedema resulting from elephantiasis, lymph node dissection, or Milroy's disease. Recurrent staphylococcal cutaneous infections are more common among individuals who have eosinophilia and elevated serum levels of IgE (Job's syndrome) and among nasal carriers of staphylococci. Cellulitis caused by *Streptococcus agalactiae* (group B *Streptococcus*) occurs primarily in elderly patients and those with diabetes mellitus or peripheral vascular disease. *Haemophilus influenzae* typically causes periorbital cellulitis in children in association with sinusitis, otitis media, or epiglottitis. It is unclear whether this form of cellulitis will (like meningitis) become less common as a result of the impressive efficacy of the *H. influenzae* type b vaccine.

Many other bacteria also cause cellulitis. It is fortunate that these organisms occur in such characteristic settings that a good history provides useful clues to the diagnosis. Cellulitis associated with cat bites and, to a lesser degree, with dog bites is commonly caused by *Pasteurella multocida*, although in the latter case *Staphylococcus intermedius* and *Capnocytophaga canimorsus* (formerly DF-2) must also be considered. Sites of cellulitis and abscesses associated with dog bites and human

bites also contain a variety of anaerobic organisms, including *Fusobacterium*, *Bacteroides*, aerobic and anaerobic streptococci, and *Eikenella corrodens*. *Pasteurella* is notoriously resistant to dicloxacillin and nafcillin but is sensitive to all other β-lactam antimicrobial agents as well as to quinolones, tetracycline, and erythromycin. Ampicillin/clavulanate, ampicillin/sulbactam, and cefoxitin are good choices for the treatment of animal or human bite infections. *Aeromonas hydrophila* causes aggressive cellulitis in tissues surrounding lacerations sustained in freshwater (lakes, rivers, and streams). This organism remains sensitive to aminoglycosides, fluoroquinolones, chloramphenicol, trimethoprim- sulfamethoxazole, and third-generation cephalosporins; it is resistant to ampicillin, however.

P. aeruginosa causes three types of soft tissue infection: ecthyma gangrenosum in neutropenic patients, hot-tub folliculitis, and cellulitis following penetrating injury. Most commonly, *P. aeruginosa* is introduced into the deep tissues when a person steps on a nail. Treatment includes surgical inspection and drainage, particularly if the injury also involves bone or joint capsule. Choices for empirical treatment while antimicrobial susceptibility data are awaited include an aminoglycoside, a third-generation cephalosporin (ceftazidime, cefoperazone, or cefotaxime), a semisynthetic penicillin (ticarcillin, mezlocillin, or piperacillin), or a fluoroquinolone (although drugs of the last class are not indicated for the treatment of children <13 years old).

Gram-negative bacillary cellulitis, including that due to *P. aeruginosa*, is most common among hospitalized, immunocompromised hosts. Cultures and sensitivity tests are critically important in this setting because of multidrug resistance (Chap. 152).

The gram-positive aerobic rod *Erysipelothrix rhusiopathiae* is most often associated with fish and domestic swine and causes cellulitis primarily in bone renderers and fishmongers. *E. rhusiopathiae* remains susceptible to most β-lactam antibiotics (including penicillin), erythromycin, clindamycin, tetracycline, and cephalosporins but is resistant to sulfonamides, chloramphenicol, and vancomycin. Its resistance to vancomycin, which is unusual among gram-positive bacteria, is of potential clinical significance since this agent is sometimes used in empirical therapy for skin infection. Fish food containing the water flea *Daphnia* is sometimes contaminated with *M. marinum*, which can cause cellulitis or granulomas on skin surfaces exposed to the water in aquariums or injured in swimming pools. Rifampin plus ethambutol has been an effective therapeutic combination in some cases, although no comprehensive studies have been undertaken. In addition, some strains of *M. marinum* are susceptible to tetracycline or to trimethoprim-sulfamethoxazole.

Necrotizing Fasciitis

(Table 125–1) Necrotizing fasciitis, formerly called streptococcal gangrene, may be associated with group A *Streptococcus* or mixed aerobic-anaerobic bacteria or may occur as part of gas gangrene caused by *Clostridium perfringens*. Strains of MRSA that produce the Panton-Valentine leukocidin (PVL) toxin have been reported to cause necrotizing fasciitis. Early diagnosis may be difficult when pain or unexplained fever is the only presenting manifestation. Swelling then develops and is followed by brawny edema and tenderness. With progression, dark-red induration of the epidermis appears, along with bullae filled with blue or purple fluid. Later the skin becomes friable and takes on a bluish, maroon, or black color. By this stage, thrombosis of blood vessels in the dermal papillae (Fig. 125-1) is extensive. Extension of infection to the level of the deep fascia causes this tissue to take on a brownish-gray appearance. Rapid spread occurs along fascial planes, through venous channels and lymphatics. Patients in the later stages are toxic and frequently manifest shock and multiorgan failure.

Necrotizing fasciitis caused by mixed aerobic-anaerobic bacteria begins with a breach in the integrity of a mucous membrane barrier, such as the mucosa of the gastrointestinal or genitourinary tract. The portal can be a malignancy, a diverticulum, a hemorrhoid, an anal fissure, or a urethral tear. Other predisposing factors include peripheral vascular disease, diabetes mellitus, surgery, and penetrating injury to the abdomen. Leakage into the perineal area results in a syndrome called *Fournier's gangrene*, characterized by massive swelling of the scrotum and penis with extension into the perineum or the abdominal wall and legs.

Necrotizing fasciitis caused by *S. pyogenes* has increased in frequency and severity since 1985. There are two distinct clinical presentations: those with no portal of entry and those with a defined portal of entry. Infections in the first category often begin deep at the site of a nonpenetrating minor trauma, such as a bruise or a muscle strain. Seeding of the site via transient bacteremia is likely, although most patients deny antecedent streptococcal infection. The affected patients present with only severe pain and fever. Late in the course, the classic signs of necrotizing fasciitis, such as purple (violaceous) bullae, skin sloughing, and progressive toxicity, develop. In infections of the second type, *S. pyogenes* may reach the deep fascia from a site of cutaneous infection or penetrating trauma. These patients have early signs of superficial skin infection with progression to necrotizing fasciitis. In either case, toxicity is severe, and renal impairment may precede the development of shock. In 20–40% of cases, myositis occurs concomitantly, and, as in gas gangrene (see below), serum creatine phosphokinase levels may be markedly elevated. Necrotizing fasciitis due to mixed aerobic-anaerobic bacteria may be associated with gas in deep tissue, but gas usually is not present when the cause is

S. pyogenes or MRSA. Prompt surgical exploration down to the deep fascia and muscle is essential. Necrotic tissue must be surgically removed, and Gram's staining and culture of excised tissue are useful in establishing whether group A streptococci, mixed aerobic-anaerobic bacteria, MRSA, or *Clostridium* species are present (see "Treatment," below).

Myositis and Myonecrosis

(Table 125–1) Muscle involvement can occur with viral infection (e.g., influenza, dengue, or coxsackievirus B infection) or parasitic invasion (e.g., trichinellosis, cysticercosis, or toxoplasmosis). Although myalgia can occur in most of these infections, severe muscle pain is the hallmark of pleurodynia (coxsackievirus B), trichinellosis, and bacterial infection. Acute rhabdomyolysis predictably occurs with clostridial and streptococcal myositis but may also be associated with influenza virus, echovirus, coxsackievirus, Epstein-Barr virus, and *Legionella* infections.

Pyomyositis is usually due to *S. aureus*, is common in tropical areas, and generally has no known portal of entry. Cases of pyomyositis caused by MRSA producing the PVL toxin have been described among children in the United States. Muscle infection begins at the exact site of blunt trauma or muscle strain. Infection remains localized, and shock does not develop unless organisms produce toxic shock syndrome toxin 1 or certain enterotoxins and the patient lacks antibodies to the toxin produced by the infecting organisms. In contrast, *S. pyogenes* may induce primary myositis (referred to as *streptococcal necrotizing myositis*) in association with severe systemic toxicity. Myonecrosis occurs concomitantly with necrotizing fasciitis in ~50% of cases. Both are part of the streptococcal toxic shock syndrome.

Gas gangrene usually follows severe penetrating injuries that result in interruption of the blood supply and introduction of soil into wounds. Such cases of traumatic gangrene are usually caused by the clostridial species *C. perfringens*, *C. septicum*, and *C. histolyticum*. Rarely, latent or recurrent gangrene can occur years after penetrating trauma; dormant spores that reside at the site of previous injury are most likely responsible. Spontaneous nontraumatic gangrene among patients with neutropenia, gastrointestinal malignancy, diverticulosis, or recent radiation therapy to the abdomen is caused by several clostridial species, of which *C. septicum* is the most commonly involved. The tolerance of this anaerobe to oxygen probably explains why it can initiate infection spontaneously in normal tissue anywhere in the body.

Gas gangrene of the uterus, especially that due to *C. sordellii,* historically occurred as a consequence of illegal or self-induced abortion and nowadays also follows spontaneous abortion, vaginal delivery, and cesarean section. *C. sordellii* has also been implicated in medically induced abortion. Postpartum *C. sordellii* infections in young, previously healthy women present as a unique clinical picture: little or no fever, lack of a purulent discharge, refractory hypotension, extensive peripheral edema and effusions, hemoconcentration, and a markedly elevated white blood cell count. The infection is almost uniformly fatal, with death ensuing rapidly.

Synergistic nonclostridial anaerobic myonecrosis, also known as necrotizing cutaneousmyositis and synergistic necrotizing cellulitis, is a variant of necrotizing fasciitis caused by mixed aerobic and anaerobic bacteria with the exclusion of clostridial organisms (see "Necrotizing Fasciitis," above).

Diagnosis

This chapter has emphasized the physical appearance and location of lesions within the soft tissues as important diagnostic clues. The temporal progression of the lesions as well as the patient's travel history, animal exposure or bite history, age, underlying disease status, and lifestyle are also crucial considerations in narrowing the differential diagnosis. However, even the astute clinician may find it challenging to diagnose all infections of the soft tissues by history and inspection alone. Soft tissue radiography, computed tomography **(Fig. 125-2)**, and magnetic resonance imaging may be useful in determining the depth of infection and should be performed when the patient has rapidly progressing lesions or evidence of a systemic inflammatory response syndrome. These tests are particularly valuable for defining a localized abscess or detecting gas in tissue. Unfortunately, they may reveal only soft tissue swelling and thus are not specific for fulminant infections such as necrotizing fasciitis or myonecrosis caused by group A *Streptococcus* (Fig. 125-2), where gas is not found in lesions.

Figure 125-2

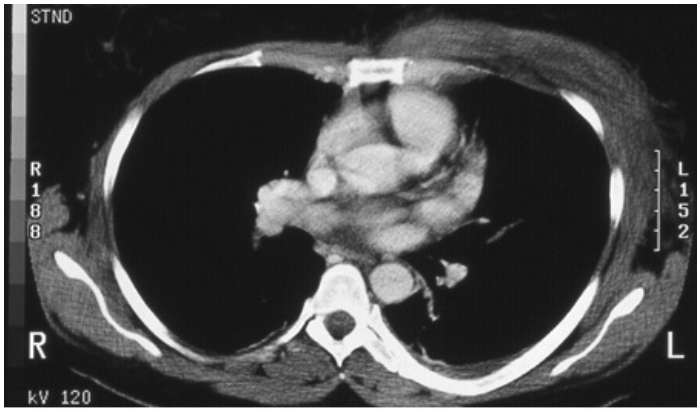

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomography showing edema and inflammation of the left chest wall in a patient with necrotizing fasciitis and myonecrosis caused by group A *Streptococcus*.

Aspiration of the leading edge or punch biopsy with frozen section may be helpful if the results of imaging tests are positive, but false-negative results occur in ∑80% of cases. There is some evidence that aspiration alone may be superior to injection and aspiration with normal saline. Frozen sections are especially useful in distinguishing SSSS from TEN and are quite valuable in cases of necrotizing fasciitis. Open surgical inspection, with debridement as indicated, is clearly the best way to determine the extent and severity of infection and to obtain material for Gram's staining and culture. Such an aggressive approach is important and may be lifesaving if undertaken early in the course of fulminant infections where there is evidence of systemic toxicity.

Treatment: Infections of the Skin, Muscles, and Soft Tissues

A full description of the treatment of all the clinical entities described herein is beyond the scope of this chapter. As a guide to the clinician in selecting appropriate treatment, the antimicrobial agents useful in the most common and the most fulminant cutaneous infections are listed in **Table 125–2**.

Table 125–2 Treatment of Common Infections of the Skin

Diagnosis/Condition	Primary Treatment	Alternative Treatment	See Also Chap(s).
Animal bite (prophylaxis or early infection)[a]	Amoxicillin/clavulanate,875/125 mg PO bid	Doxycycline, 100 mg PO bid	e24
Animal bite[a] (established infection)	Ampicillin/sulbactam, 1.5–3 g IV q6h	Clindamycin, 600–900 mg IV q8h, *plus* Ciprofloxacin, 400 mg IV q12h, *or* Cefoxitin, 2 g IV q6h	e24
Bacillary angiomatosis	Erythromycin, 500 mg PO qid	Doxycycline, 100 mg PO bid	160
Herpes simplex (primary genital)	Acyclovir, 400 mg PO tid for 10 days	Famciclovir, 250 mg PO tid for 5–10 days, *or* Valacyclovir, 1000 mg PO bid for 10 days	179
Herpes zoster (immunocompetent host >50 years of age)	Acyclovir, 800 mg PO 5 times daily for 7–10 days	Famciclovir, 500 mg PO tid for 7–10 days, *or* Valacyclovir, 1000 mg PO tid for 7 days	180
Cellulitis (staphylococcal or streptococcal[b,c])	Nafcillin or oxacillin, 2 g IV q4–6h	Cefazolin, 1–2 g q8h, *or* Ampicillin/sulbactam, 1.5–3 g IV q6h, *or* Erythromycin, 0.5–1 g IV q6h, *or* Clindamycin, 600–900 mg IV q8h	135, 136

MRSA skin infection[d]	Vancomycin, 1 g IV q12h	Linezolid, 600 mg IV q12h	135
Necrotizing fasciitis (group A streptococcal[b])	Clindamycin, 600–900 mg IV q6–8h, *plus* Penicillin G, 4 million units IV q4h	Clindamycin, 600–900 mg IV q6–8h, *plus* Cephalosporin (first- or second-generation)	136
Necrotizing fasciitis (mixed aerobes and anaerobes)	Ampicillin, 2 g IV q4h, *plus* Clindamycin, 600–900 mg IV q6–8h, *plus* Ciprofloxacin, 400 mg IV q6–8h	Vancomycin, 1 g IV q6h, *plus* Metronidazole, 500 mg IV q6h, *plus* Ciprofloxacin, 400 mg IV q6–8h	164
Gas gangrene	Clindamycin, 600–900 mg IV q6–8h, *plus* Penicillin G, 4 million units IV q4–6h	Clindamycin, 600–900 mg IV q6–8h, *plus* Cefoxitin, 2 g IV q6h	142

[a]*Pasteurella multocida*, a species commonly associated with both dog and cat bites, is resistant to cephalexin, dicloxacillin, clindamycin, and erythromycin. *Eikenella corrodens*, a bacterium commonly associated with human bites, is resistant to clindamycin, penicillinase-resistant penicillins, and metronidazole but is sensitive to trimethoprim-sulfamethoxazole and fluoroquinolones.

[b]The frequency of erythromycin resistance in group A *Streptococcus* is currently ∑5% in the United States but has reached 70–100% in some other countries. Most, but not all, erythromycin-resistant group A streptococci are susceptible to clindamycin. Approximately 90–95% of *Staphylococcusaureus* strains are sensitive to clindamycin.

[c]Severe hospital-acquired *S. aureus* infections or community-acquired *S. aureus* infections that are not responding to the ß-lactam antibiotics recommended in this table may be caused by methicillin-resistant strains, requiring a switch to vancomycin or linezolid.

[d]Some strains of methicillin-resistant *S. aureus* (MRSA) remain sensitive to tetracycline and trimethoprim-sulfamethoxazole. Daptomycin (4 mg/kg IV q24h) or tigecycline (100-mg loading dose followed by 50 mg IV q12h) are alternative treatments for MRSA.

Furuncles, carbuncles, and abscesses caused by MRSA and MSSA are common, and their treatment depends upon the size of the lesion. Furuncles <2.5 cm in diameter are usually treated with moist heat. Those that are larger (4.5 cm of erythema and induration) require surgical drainage, and the occurrence of these larger lesions in association with fever, chills, or leukocytosis requires both drainage and antibiotic treatment. A study in children demonstrated that surgical drainage of abscesses (mean diameter, 3.8 cm) was as effective when used alone as when combined with trimethoprim-sulfamethoxazole treatment. However, the rate of recurrence of new lesions was lower in the group undergoing both drainage and antibiotic treatment.

Early and aggressive surgical exploration is essential in cases of suspected necrotizing fasciitis, myositis, or gangrene in order to (1) visualize the deep structures, (2) remove necrotic tissue, (3) reduce compartment pressure, and (4) obtain suitable material for Gram's staining and for aerobic and anaerobic cultures. Appropriate empirical antibiotic treatment for mixed aerobic-anaerobic infections could consist of ampicillin/sulbactam, cefoxitin, or the following combination: (1) clindamycin (600–900 mg intravenously every 8 h) or metronidazole (500 mg every 6 h) plus (2) ampicillin or ampicillin/sulbactam (1.5-3 g intravenously every 6 h) plus (3) gentamicin (1–1.5 mg/kg every 8 h). Group A streptococcal and clostridial infection of the fascia and/or muscle carries a mortality rate of 20–50% with penicillin treatment. In experimental models of streptococcal and clostridial necrotizing fasciitis/myositis, clindamycin has exhibited markedly superior efficacy, but no comparative clinical trials have been performed. A retrospective study of children with invasive group A streptococcal infection demonstrated higher survival rates with clindamycin treatment than with ß-lactam antibiotic therapy. Hyperbaric oxygen treatment may also be useful in gas gangrene due to clostridial species. Antibiotic treatment should be continued until all signs of systemic toxicity have resolved, all devitalized tissue has been removed, and granulation tissue has developed (Chaps. 136, 142, and 164).

In summary, infections of the skin and soft tissues are diverse in presentation and severity and offer a great challenge to the clinician. This chapter provides an approach to diagnosis and understanding of the pathophysiologic mechanisms involved in these infections. More in-depth information is found in chapters on specific infections.

FURTHER READINGS

Bisno AI, Stevens DL: Streptococcal infections in skin and soft tissues. N Engl J Med 334:240,

1996[PMID: 8532002]

Breman JG, Henderson DA: Diagnosis and management of smallpox. N Engl J Med 346:1300, 2002[PMID: 11923491]

Duong M et al: Randomized, controlled trial of antibiotics in the management of community-acquired skin abscesses in the pediatric patient. Ann Emerg Med 55:401, 2010[PMID: 19409657]

Fridkin SK et al: Methicillin-resistant *Staphylococcus aureus* disease in three communities. N Engl J Med 352:1436, 2005[PMID: 15814879]

Miller LG et al: Necrotizing fasciitis caused by community-associated methicillin-resistant *Staphylococcus aureus* in Los Angeles. N Engl J Med 352:1445, 2005[PMID: 15814880]

Norrby-Teglund A, Stevens DL: Novel therapies in streptococcal toxic shock syndrome: Attenuation of virulence factor expression and modulation of host response. Curr Opin Infect Dis 11:285, 1998[PMID: 17033393]

Stevens DL: Streptococcal toxic shock syndrome associated with necrotizing fasciitis. Annu Rev Med 51:271, 2000

———: Necrotizing soft tissue infections. Curr Treat Opt Infect Dis 2:359, 2000

——— et al: Practice guidelines for the diagnosis and management of skin and soft-tissue infections.Clin Infect Dis 41:1373, 2005[PMID: 16231249]

Talan DA et al: Bacteriologic analysis of infected dog and cat bites. Emergency Medicine Animal Bite Infection Study Group. N Engl J Med 340:85, 1999[PMID: 9887159]

Zimbelman J et al: Improved outcome of clindamycin compared with beta-lactam antibiotic treatment for invasive *Streptococcus pyogenes* infection. Pediatr Infect Dis J 18:1096, 1999[PMID: 10608632]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OSTEOMYELITIS: INTRODUCTION

Osteomyelitis, an infection of bone that leads to tissue destruction and often to debility, can be caused by a wide variety of bacteria (including mycobacteria) and fungi and may be associated with viral infections. Its management must be individualized and depends on numerous factors, including the causative organism, the specific bone involved, vascular supply, nerve function, foreign bodies, recent injury, the physiologic status of the host, and associated comorbidities. The spectrum of the disease can range from extensive (e.g., tibial and vertebral osteomyelitis) to localized (e.g., bone invasion associated with a tooth abscess). Two major classification systems for osteomyelitis are used in making decisions about medical therapy and surgery. Lee and Waldvogel categorized cases as acute or chronic, hematogenous or contiguous, and with or without vascular compromise. The Cierny and Mader classification system for long-bone osteomyelitis encompasses the location and extent of the infection as well as a number of other factors.

Etiology (Table 126–1)

Table 126–1 Microorganisms that Cause Osteomyelitis

Organism	Comment
Frequently Encountered Bacteria	
Staphylococcus aureus	Most likely bacterial pathogen Aggressive, invasive Often metastatic foci with bacteremia Consider surgery early
Staphylococci other than *S. aureus* (coagulase-negative)	Usually associated with foreign material or implants Biofilm production
Streptococci	May spread rapidly through soft tissues
Enterobacteriaceae (*Escherichia coli*, *Klebsiella*, others)	Considerable variation in antibiotic susceptibility Increasing antibiotic resistance with overuse May become resistant to antibiotics during therapy
Pseudomonas aeruginosa	Increasingly resistant to antibiotics Frequent successor to other bacteria when initial therapy fails May be related to contamination
Unusual Organisms	
Anaerobic bacteria	Usually mixed with aerobic bacteria May be synergistic Survival dependent on devitalized tissue
Bartonella henselae	Associated with cat scratches and probably with fleas
Brucella species	Prominent in developing countries, especially with unpasteurized milk
Fungi	*Candida* the most likely genus Considerable variation in susceptibility, depending on species Surgery may be helpful if infection is invasive
Mycobacterium tuberculosis	May involve any bone Vertebral osteomyelitis common in some countries

Mycobacteria other than *M. tuberculosis*	Need special culture media to recover
Viruses	Associated with some viral infections, including varicella and variola

The foremost bacterial cause of osteomyelitis is *Staphylococcus aureus*. Gram-negative organisms such as *Pseudomonas aeruginosa* and *Escherichia coli*, coagulase-negative staphylococci, enterococci, and propionibacteria may also be involved. *Mycobacterium tuberculosis* is a common cause of osteomyelitis in countries with limited medical resources; other mycobacterial species that infect bone include *M. marinum*, *M.chelonei*, and *M. fortuitum*. Fungal etiologies include *Candida*, *Coccidioides*, *Histoplasma*, and *Aspergillus* species. Noninfectious pathogenic mechanisms that may cause disease mimicking osteomyelitis include avascular necrosis, rheumatoid diseases, neuropathy with chronic trauma, gout, and malignancies.

The precipitating event(s) for osteomyelitis vary greatly. The prosthetic joint implants and stabilization devices that are increasingly being used in orthopedic surgery are associated with complex infections. Trauma is also a common cause of infection, especially when a wound is involved and there is contamination of bone or surrounding tissue along with significant tissue damage or destruction. Even in the absence of an open wound or a compound fracture, damaged tissue and extravasated blood may slow the circulation, establishing a favorable medium for the growth of bacteria that may reach the area through low-level bacteremia from the peripheral venous circulation or from distal lymphatic channels. Bacteremia—whether due to endocarditis or due to seeding from other sites of infection (e.g., abscesses, boils, or vascular devices)—is also a frequent etiologic factor in osteomyelitis. Studies of *S. aureus* bacteremia indicate a rate of metastatic osteomyelitis approaching 28% if there is a prosthetic joint in place; *S. aureus* bacteremia can be complicated by the involvement of methicillin-resistant strains (MRSA), which are progressively replacing strains that are more susceptible to antibiotics. The overlapping circulations of the urinary tract and the spine may be a source of vertebral osteomyelitis due to urinary tract pathogens such as *E. coli* and *Klebsiella*. Additional predisposing factors include a poor arterial and venous supply, which may limit perfusion to bone to the point of an inadequate response and poor healing, even in patients with normal function. Host factors such as diabetes and its consequences contribute significantly to the development of osteomyelitis through impaired immunity with hyperglycemia, loss of sensation, vascular disease, and renal failure.

Epidemiology

In the United States, acute osteomyelitis affects ~0.1–1.8% of the otherwise healthy adult population. After a foot puncture, 30–40% of adults with diabetes develop osteomyelitis. In this country, there has been a major change in the profile of certain bacterial pathogens, with the emergence of MRSA strains over the last decade. MRSA has become a source of great concern in hospitals, especially after surgery. The morbidity and economic consequences appear to be greater for MRSA osteomyelitis than for osteomyelitis caused by methicillin-sensitive *S. aureus* strains. However, it is not clear that these poorer outcomes for MRSA are due to new or more destructive virulence factors. Rather, they may simply be the result of a delay in effective antimicrobial treatment.

The types and etiologies of osteomyelitis vary by region and with time. The United States has seen a rise in infections related to the increasing use of orthopedic surgery for correction of deformities and implantation of screws, pins, rods, plates, and prosthetic joints. With the aging of populations and the epidemics of obesity and diabetes in some countries, the frequency of these predisposing factors continues to increase, requiring adaptations in treatment approaches. Any type of instrumentation may lead to infection in a small proportion of cases. Osteomyelitis attributable to orthopedic devices and surgical interventions is considerably less common in countries with limited medical resources, where tuberculosis may be the dominant infection and brucellosis is not unusual. In many of these areas, agricultural injuries, industrial accidents, and war wounds are much more common than in wealthy countries, and the pathogens causing infection reflect those injuries. Osteomyelitis is more common in situations where wounds cannot promptly be debrided and repaired, microbiology laboratories are not readily available, and effective antimicrobial agents are in short supply.

Pathogenesis

The most common predisposing factor for osteomyelitis is an area of bone or contiguous surrounding tissue that is abnormal in terms of viability, blood supply, sensation, or edema. The damaged tissue not only compromises healthy circulation to the area but may slow the flow of venous blood and lymph, thereby providing nutrients to bacteria and fueling ongoing damage. Host factors such as poor nutrition and immunosuppression may also be relevant. Diabetes in adults poses the most significant risk. Diabetic neuropathy adds to the progression of osteomyelitis as the patient may be unaware of infection as it spreads into the bone; the consequences include thousands of amputations each year. Additional sources of immunosuppression, such as chemotherapy and treatment with glucocorticoids or tumor necrosis factor (TNF) inhibitors, also inhibit normal defense mechanisms and thus predispose to more frequent and serious infections whose symptoms are diminished because of reduced inflammatory responses.

The bacteria involved in osteomyelitis perpetuate themselves by elaborating toxins that further damage tissues, including bone. *S. aureus* is particularly adept in this respect; it colonizes the nasal area in about one-third of healthy individuals and can produce a wide variety of cytokines, enzymes, and toxins that destroy tissue and affect neutrophil response. Some *S. aureus* bacteria survive uptake into the phagocytic vacuoles of macrophages and continue to cause disease and recrudescence by persistently eluding the usual defense mechanisms. This capacity for "hibernation" and persistence may allow *S. aureus* to remain dormant for decades before infection erupts at the sites of old injuries (e.g., shrapnel or other penetrating wounds).

Coagulase-negative staphylococci are generally not as virulent as *S. aureus* but have been found to persist by producing a biofilm that protects them from the host and apparently allows them to exist for many years on prosthetic joints, with minimal symptoms. The extent to which other organisms use biofilm to their advantage is unclear, but biofilm production probably plays a significant role in osteomyelitis, especially the chronic forms.

Multiple bacteria may be recovered from cultures, especially when there is an entry wound. Decisions about which ones to target in antibiotic therapy are often difficult. Common skin-dwelling and colonizing microbes usually do not need to be treated, and overtreatment in fact results in unnecessary toxicity and increases antimicrobial resistance among the organisms that survive. Anaerobic bacteria can often be recovered and may play a synergistic role with usual or unusual pathogens; specific therapy is sometimes beneficial in these situations.

The intrinsic factors of organisms that are responsible for persistence and bone destruction have not yet been identified. However, there is probably strain-to-strain variation in virulence factors produced by particular clones, with some strains consequently much more virulent than others. The prevention of biofilm production merits investigation in this regard.

Approach to the Patient: Osteomyelitis

The best approach to the care of a patient with significant osteomyelitis is to assemble a team of providers who can work together in considering the microbiology of the infection and make sound decisions about antibiotic therapy and surgery. The most effective program will include evaluation and management of antibiotics, microbiology, pharmacology, glucose levels, vascular disease, neuropathy, and renal function, with close follow-up by a knowledgeable physician who is interested in leading the team in coordinating care.

When osteomyelitis is suspected, a careful, methodical approach is needed (see "Clinical Manifestations and Diagnosis," below). Patients should be educated about the significance of an infection that involves bone, especially if risk factors cannot be eliminated. Blood tests, cultures, standard radiography, scans, biopsies, and surgery may all be necessary for a clear diagnosis and full delineation of the pathogen. Collection of this baseline information can be very important in both early and late decision-making.

Initial evaluations for osteomyelitis must be aggressive, as the infection can progress rapidly in the absence of antibiotic therapy effective against the wide variety of potential pathogens. Inadequacies in cultures, surgery, or temporizing measures may greatly exacerbate the damage caused by the infection. Hospitalization may be indicated for rapid multispecialty evaluation, imaging, and stabilization of complex infections such as with a diabetic foot. Outpatient therapy may not be adequate for the teamwork and interventions needed. Early admission and procedures may actually shorten the length of hospital stay.

The physician should inform the patient about the value of all the necessary evaluations, the implications of surgery, and the possibility of a prolonged course of IV antibiotic therapy, whether in the hospital or at home. A patient's fear of amputation can lead to inordinate delays in seeking treatment that allow the infection to progress. Moreover, it is not unusual for a patient to refuse surgery and amputation even though such treatments will clearly increase the likelihood of a functional lifestyle. Therefore, it is best to prepare patients early on if there may be negative outcomes such as amputation and perhaps to set criteria and timelines for success or failure of therapy and interventions.

Clinical Manifestations and Diagnosis

Diagnosis of acute osteomyelitis within the first few weeks of onset is important and is usually relatively easy. If the diagnosis is missed, however, the symptoms may become chronic, with slow progression or a dormant phase of several years.

A thorough history and physical examination are the mainstays of evaluation for osteomyelitis. A clear pattern of pain, swelling, and possibly drainage after surgery or injury should raise suspicion, but such indicators may not all be present, even in a patient with neuropathy, compromised circulation, chronic edema, organ failure, diabetes, or other predisposing factors. Direct questions about previous injuries, infections, surgeries, or hardware implantation—even decades earlier—can yield information critical in guiding empirical antibiotic therapy and surgery. A history of injury is particularly important, even if the skin was not broken and there were no clinical signs of bacteremia. It is not unusual for a soft tissue injury to serve as a nidus of secondary bone infection, presumably seeded by low-level bacteremia

and often occurring without symptoms. Other sources of seeding may include boils, abscesses, cellulitis, or injection sites. A careful examination is essential in identifying additional predisposing factors and assessing the role of comorbidities such as neuropathy, arterial disease, venous insufficiency, and chronic trauma that can lead to severe accumulation of callus in insensate feet.

Careful consideration and assessment of disorders that may mimic or accompany osteomyelitis are essential. Arthritis, gout, ischemia, neuropathies, and recent surgery may be diagnosed when osteomyelitis is the real cause of symptoms on a cofactor. For example, chronic back pain may be attributed to degenerative arthritis, but there can be a substantial loss of neurologic function if the pain is actually due to diskitis with vertebral osteomyelitis.

Correctly diagnosing osteomyelitis early has crucial implications for later function, disability, treatment cost, and risk of a fatal outcome. A variety of tools must be used to definitively diagnose or conclusively rule out an infection. A standard x-ray is a good starting point that can reveal a variety of abnormalities (Fig. 126-1A) and may eliminate the need for further imaging studies. Bone loss, sequestra, periosteal elevation or swelling (which can develop early on), and shadows around foreign bodies are hallmarks of bone infection. However, these findings may also be found with other disorders, such as tumors, trauma, avascular necrosis, and gout. Standard two-dimensional images can be of limited value in assessing complex bones. The value of radiology may be limited by the time required for an infection to become apparent; actual dissolution or resorption of bone due to infection may not be apparent for several weeks or more.

Figure 126-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

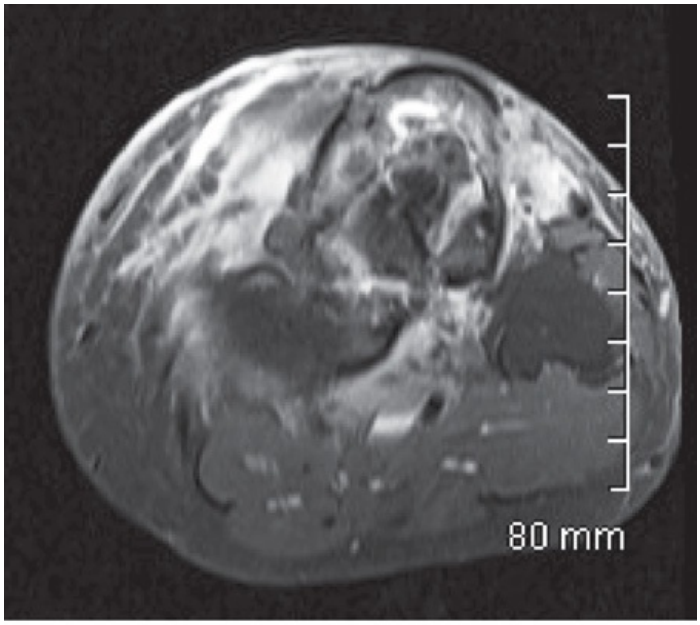

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

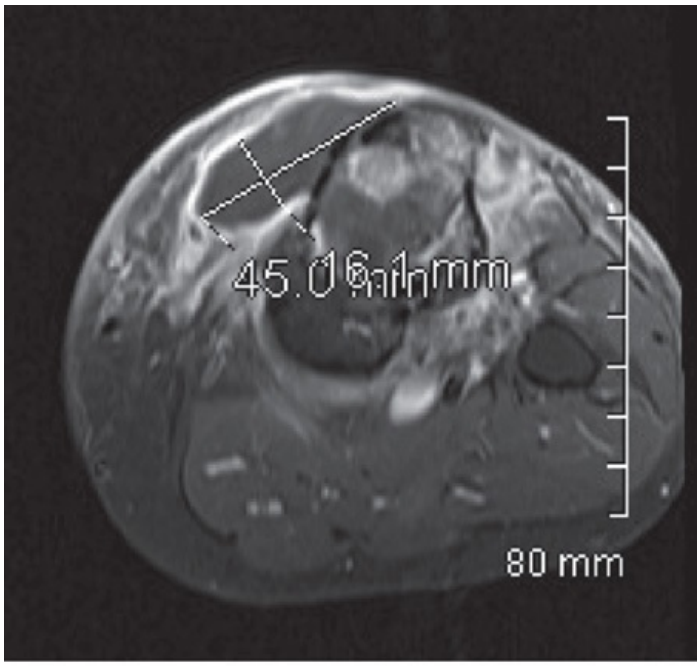

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

(***A***) Standard radiology image indicates infection with sclerosis of the proximal tibia and periosteal elevation and obvious bone destruction with an apparent cavity and the suggestion of a sequestrum in the proximal medial tibia. (***B, C***) Magnetic resonance images more clearly visualize the bone and soft tissue anatomy, confirming an extensive infection with destruction within the proximal tibia that has extended into the surrounding soft tissues and the joint as well as a ring of calcification most consistent with an abscess. (***D***) A longitudinal MRI shows the extent of longitudinal bone destruction and soft tissue involvement with contrast enhancement that suggests viable marrow from the middle to the distal tibial shaft.

Depending on the results of the initial x-ray, further investigations with invasive techniques may be appropriate. Collection of pus by needle aspiration through a clean area from a closed pocket not only documents bone infection but also permits recovery and evaluation of the pathogen(s). A culture of a wound swab may be of some value but is clearly less reliable in identifying the real culprit(s), which may be present in the bone but absent from its surface. Biopsy provides more accurate microbiologic information than needle aspiration and supplies tissue for pathology studies, which may be helpful. Some organisms that usually are not recovered (in a timely fashion or at all) by standard cultures may be rendered visible with special staining of tissue samples. Unfortunately, the size of the needle used for needle biopsy may not be appropriate for small bones of the hands or feet. Open surgical exploration, biopsy, and drainage, which can provide high-quality tissue samples for culture and pathology and offer a view of the infected bone and surrounding area, should also be considered. Necrotic tissue can be removed and circulation assessed with one procedure. Polymerase chain reaction and other sequencing technologies are increasingly being used to detect and identify specific organisms—and even to determine their susceptibilities—within hours instead of days or weeks. Information on specific strains of unusual organisms may be of value, especially in difficult cases.

Laboratory tests are useful in assessing osteomyelitis but usually do not yield specific information relevant to etiology or severity. Leukocytosis may be noted in acute infection but is less likely in chronic infection, which may also be associated with anemia. Determination of the erythrocyte sedimentation rate (ESR) is a simple, inexpensive aid to diagnosis; it serves as an indicator of response with *S. aureus* infections but is not as useful for gram-negative infections because the cytokines and inflammatory elements that result in elevations are different for gram-positive (*S. aureus*) than for gram-negative infections. C-reactive protein (CRP) measurement may be helpful, especially in the evaluation of children, but may not be as useful as an ESR determination in some cases. CRP changes occur earlier in response to bacterial infection. Both ESR and CRP determinations have significant limitations in multifactorial diseases, with elevated values reflecting conditions other than osteomyelitis. Additional laboratory tests for diseases associated with bone loss that may mimic or complicate osteomyelitis should include measurement of glucose levels and tests for renal failure, gout, vasculitis, and rheumatoid diseases.

Additional imaging studies may be of value if the diagnosis remains unclear. CT can delineate bone more clearly than standard radiography and offers three-dimensional displays that can be extremely useful in detecting abnormalities and devising a surgical approach. MRI (Fig. 126-1B–D) provides high-quality images of the soft tissue around the bone abnormality and may be essential in diagnosing

an epidural abscess related to vertebral osteomyelitis. Technetium and leukocyte isotope scans offer insight into the activity of the disease process and the affected site(s). Although these additional screening tools may be helpful in evaluation and decision-making, they may not be cost-effective.

Treatment: Osteomyelitis

Therapy for osteomyelitis is challenging because of the variety of causative organisms, the usual comorbidities, the need for a prolonged course and IV administration, the common physical limitations of the patient, and high costs. An aggressive therapeutic approach is warranted given the dire consequences of failure of medical therapy, which can include loss of limbs. The sooner the infection is diagnosed and treated, the better the outcome and the less damage done during delays in intervention. Antibiotic therapy should be used aggressively to stop disease progression and should be designed to avoid the development of resistant organisms. Early surgical intervention (e.g., debridement) can confirm the infection, identify and characterize the etiologic agent(s), and remove dead or devitalized tissue that may be providing bacteria with nutrients and allowing them to spread. A variety of antibiotics are available for most of the likely pathogens (**Table 126–2**), although the most common pathogen—*S. aureus*—continues to evolve mechanisms to elude these drugs. MRSA strains represent an increasing problem in both the hospital and the community. Staphylococci and Enterobacteriaceae resistant to even more antibiotics than MRSA appear to be evolving.

Table 126–2 Antibiotics for the Treatment of Osteomyelitis

Organism	Antimicrobial Agent	Dosing	Comments
Methicillin-susceptible *Staphylococcus aureus*	Oxacillin or nafcillin	2 g IV q6h	May be more active than cephalosporins More difficult than cephalosporins to administer for long periods
	Cephalosporins	Cefazolin: 2 g IV q8h Ceftriaxone: 1–2 g IV q24h	Ceftriaxone advantageous with OPAT
	Clindamycin[a]	600–900 mg IV q8h	Not well studied for osteomyelitis Oral form possible (300–600 mg oral q8h) Resistance significant and increasing Toxicity different from that of β-lactam antibiotics
Methicillin-resistant *S. aureus*	Vancomycin	15 mg/kg IV q12h	Strains with an MIC of ≥2 μg/mL may not respond well.
	Daptomycin[a]	4–6 mg/kg IV q24h	Promising, but concern about adverse effects with prolonged therapy
	Linezolid[a]	600 mg IV or PO q12h	Effectiveness and adverse effects with prolonged therapy unclear Bacteriostatic
Streptococci	Penicillin	5 mU IV q6h or 20 mU/d by continuous infusion	Not all streptococci are susceptible Ceftriaxone (1 g/d IV or IM) and ampicillin (12 g/d IV) are alternatives
Enterococci	Penicillin plus gentamicin	As above 5 mg/kg daily IV	If strain is susceptible
	Vancomycin	As above	If strain is susceptible
Enterobacteriaceae (*E. coli*, *Klebsiella*, other)	Ceftriaxone or another cephalosporin Ciprofloxacin	As above 400 mg IV q8–12h	If strain is susceptible 500–750 mg q8–12h if strain is susceptible
Pseudomonas aeruginosa	Ciprofloxacin	As above	Resistance may develop during therapy; if strain is resistant, drugs to consider include cefepime and ceftazidime

[a]Not approved for use in osteomyelitis by the U.S. Food and Drug Administration. ***Abbreviations:*** MIC, minimal inhibitory concentration; OPAT, outpatient parenteral antimicrobial therapy.

The most common targets for empirical antibiotic therapy are staphylococci, which are carried asymptomatically in and around the nares by nearly one-third of healthy people. The common β-lactam antibiotics provide excellent results against methicillin-sensitive *S. aureus* strains. Oxacillin and nafcillin are first-line agents but may elicit more adverse reactions than cephalosporins. Cefazolin is a reasonable alternative in the hospital, but ceftriaxone is preferred as an outpatient drug because it can be given (by the IV or IM route) only once a day.

MRSA strains have been controlled with vancomycin for many years, but this drug appears to be losing its effectiveness against these microbes. New antibiotics have been designed to fill this need, although their efficacy has not been documented. In an outpatient setting, vancomycin does not appear to be as effective against methicillin-susceptible staphylococcal osteomyelitis as oxacillin or ceftriaxone. Publications about the value of daptomycin for osteomyelitis are encouraging. Tigecycline is active against MRSA but is only bacteriostatic and does not yet have a well-established outcomes record. Telavancin may also be of value against vancomycin-resistant staphylococci but has not yet been adequately tested for bone infections.

Additional antimicrobial agents for use against staphylococcal infections include linezolid, which offers the advantage of both oral and IV formulations but is bacteriostatic and has not yet been well studied. Moreover, its use—although apparently less expensive than that of other parenteral drugs—is limited by its cost. Clindamycin can also be used as both an IV and an oral agent, although antimicrobial resistance is a growing problem. Rifampin, a potential adjunct to other antistaphylococcal agents, is highly active in vitro and can penetrate phagocytic vacuoles to reach staphylococci therein. Unfortunately, resistance develops rapidly if rifampin is used alone, and clinical outcomes are not always as good as anticipated. Other agents, such as aminoglycosides, folic acid inhibitors, and macrolides, may play a limited role; they generally are neither as effective nor as toxic as other available agents.

Fluoroquinolone antibiotics offer both IV and oral therapy options and are often included in the standard recommendation for treatment of many susceptible strains of Enterobacteriaceae and *Pseudomonas* species. Drugs of this class do, however, have some limitations in terms of emerging resistance (even during therapy) and may exert some adverse neuromuscular effects (e.g., tendon rupture and impaired healing) that may be particularly relevant to the prolonged courses of antibiotics usually needed to cure the infection. In general, fluoroquinolones should not be used to treat *S. aureus* infections because of these limitations and the availability of better-studied antibiotics.

The optimal route and duration of therapy for osteomyelitis remain controversial. The usual recommendations stem from a 1970 study in which cases of osteomyelitis were characterized and outcomes were evaluated in relation to the duration of IV therapy. Better outcomes appeared to be related to a course of ≥4 weeks in some types of infection. Even though the characteristics of the bacteria and the available antibiotics were quite different at that time, a 4- to 6-week course of IV therapy remains the standard and is the usual recommended minimum. This recommendation has been challenged in pediatric studies in light of increasing evidence that oral agents and shorter courses may be adequate. Because some of the active agents reach comparable levels when given by mouth, a switch from the recommended IV administration to oral therapy may be appropriate in some situations. The proper duration of antimicrobial therapy depends on a variety of factors, including the infecting organism, the bone involved, surgical procedures, and drug tolerance and safety. Prolonged courses may be justified by extensive disease, immunocompromise, poor clinical response, and vertebral osteomyelitis. Whether a bone infection has truly been cured becomes clear only over time; relapse is not uncommon and may occur years later, especially in patients with ongoing risk factors and comorbidities. The literature suggests that a 6-month follow-up period is adequate to determine the success of treatment. Patients should be followed for at least that long, even though antibiotics have been discontinued. The possibility of relapses and the potential for their prevention should not be overlooked.

Surgery is an important tool in the treatment of osteomyelitis, offering the benefits of direct observation, prompt removal of all devitalized tissue and bone, and drainage of the infection site. Nevertheless, it is not without risk, and loss of bone or other tissue may adversely affect function. In addition, because bone may regenerate to some degree when infection is eradicated, surgery is not always needed. Surgical approaches vary with the bone involved and the extent of disease. The Cierny-Mader classification system is helpful when three-dimensional imaging is done, and MRI may help determine the viability of bone or marrow. Residual dead spaces are a source of concern and may require tissue flaps and closure. Local antibiotics and impregnated cement or beads may be of value but not should not replace IV antibiotic therapy without further study. If surgery is performed and most or all of the infected bone is removed, a full 4- to 6-week course of IV therapy probably is not necessary. However, the precise duration that is required is not clear and most likely depends primarily on the other factors involved in individual cases. One week of IV therapy after surgery may be justified to ensure pathogen eradication and healing.

Outpatient parenteral antibiotic therapy (OPAT) is a valuable means of providing the long course of IV antibiotics that is considered the standard of care and has been proven efficacious over decades.

Despite potential risks outside the hospital that patients and their providers must consider, OPAT is safe and effective when properly managed and administered. This approach is conducive to a better quality of life in a familiar setting, is considered safer because of the lack of exposure to hospital-related infections (which affect ~1 patient in every 20 admitted), is much less expensive than treatment administered in the hospital, and generally facilitates recovery, often allowing the patient to return to work or resume other day-to-day activities during the treatment course.

Complications

The complications of osteomyelitis are numerous and are most commonly related to loss of full function of the bone or supporting tissues. Fractures are more likely with progressive disease. Local spread and dissemination of infection are also possible. Misdiagnosis is particularly likely when another disease is complicating the infection. In rare instances, chronic inflammation and infection may lead to malignant transformation into squamous cell carcinoma or sarcoma.

Prognosis

The outcomes of osteomyelitis vary tremendously depending on the bone involved, the predisposing factors, the underlying diseases, and the treatment provided. Standard guidelines cannot be applied uniformly; e.g., a case of mandible infection arising from a tooth abscess may be cured with an extraction alone, whereas a case of vertebral osteomyelitis may require a prolonged course of IV therapy as it cannot be approached surgically without neurologic sequelae. For large bones, the 4- to 6-week course of IV therapy still seems reasonable, although recent studies suggest that with some new antimicrobial agents a shorter course of IV therapy, possibly with an early switch to oral therapy, may be sufficient. Determining the outcome even of long-bone osteomyelitis is complicated by uncertainty as to the duration of follow-up needed. The actual outcome in terms of debility and limb salvage may be as dependent on underlying and complicating factors and care as it is on antibiotic therapy.

Prevention

Osteomyelitis can be prevented in some instances by better infection-control measures, especially before surgery. Both mupirocin and chlorhexidine are of proven value in preventing operative infections, which are an increasing cause of bone infections associated with implanted material. Prompt treatment of bacteremia and elimination of sources of infection (e.g., boils or folliculitis) before surgery and in other situations may prevent infections. Aggressive surgical management of injuries may also help avoid the constellation of factors that lead to bone infections.

Awareness of persistent sites of infection and reasonable attempts at eradication may promote prevention. Many persistent infections that do not initially impair function or cause pain are ignored by patients; an example is provided by the classic problem of diabetic foot infections, with ulcers that burrow into the soles of insensate feet and often reach bones. Likewise, sacral ulcers are often overlooked or ignored both by physicians and by patients with neurologic impairment. Attempts to eradicate or close entry wounds are critical and should be undertaken early on.

FURTHER READINGS

Byren I et al: Pharmacotherapy of diabetic foot osteomyelitis. Expert Opin Pharmacother 10:3033, 2009[PMID: 19954273]

Fowler VG Jr et al: *S. aureus* Endocarditis and Bacteremia Study Group. Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. N Engl J Med 355:653, 2006[PMID: 16914701]

Lesens O et al: Culture of per-wound bone specimen: A simplified approach for the medical management of diabetic foot osteomyelitis. Clin Microbiol Infect 17:285, 2011[PMID: 20156216]

Lew dp, waldvogel FA: Osteomyelitis. Lancet 364:369, 2004[PMID: 15276398]

Liu C et al: Clinical practice guidelines by the Infectious Diseases Society of America for the treatment of methicillin-resistant Staphylococcus aureus infections in adults and children. Clin Infect Dis 52:1, 2011

Nicolau DP, Stein GE: Therapeutic options for diabetic foot infections: A

review with an emphasis on tissue penetration characteristics. J Am Podiatr Med Assoc 100:52, 2010

Pääkkönen M et al: Sensitivity of erythrocyte sedimentation rate and C-reactive protein in childhood bone and joint infections. Clin Orthop Relat Res 468:861, 2010[PMID: 21698820]

Tice AD et al: Practice guidelines for outpatient parenteral antimicrobial therapy. IDSA guidelines. Clin Infect Dis 38:1651, 2004[PMID: 15227610]

——— et al: Outcomes of osteomyelitis among patients treated with outpatient parenteral antimicrobial therapy. Am J Med 114:723, 2003

Zimmerli W: Clinical practice. Vertebral osteomyelitis. N Engl J Med 362:1022, 2010[PMID: 20237348]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRAABDOMINAL INFECTIONS AND ABSCESSES: INTRODUCTION

Intraperitoneal infections generally arise because a normal anatomic barrier is disrupted. This disruption may occur when the appendix, a diverticulum, or an ulcer ruptures; when the bowel wall is weakened by ischemia, tumor, or inflammation (e.g., in inflammatory bowel disease); or with adjacent inflammatory processes, such as pancreatitis or pelvic inflammatory disease, in which enzymes (in the former case) or organisms (in the latter) may leak into the peritoneal cavity. Whatever the inciting event, once inflammation develops and organisms usually contained within the bowel or another organ enter the normally sterile peritoneal space, a predictable series of events takes place. Intraabdominal infections occur in two stages: peritonitis and—if the patient survives this stage and goes untreated—abscess formation. The types of microorganisms predominating in each stage of infection are responsible for the pathogenesis of disease.

PERITONITIS

Peritonitis is a life-threatening event that is often accompanied by bacteremia and sepsis syndrome (Chap. 271). The peritoneal cavity is large but is divided into compartments. The upper and lower peritoneal cavities are divided by the transverse mesocolon; the greater omentum extends from the transverse mesocolon and from the lower pole of the stomach to line the lower peritoneal cavity. The pancreas, duodenum, and ascending and descending colon are located in the anterior retroperitoneal space; the kidneys, ureters, and adrenals are found in the posterior retroperitoneal space. The other organs, including liver, stomach, gallbladder, spleen, jejunum, ileum, transverse and sigmoid colon, cecum, and appendix, are within the peritoneal cavity. The cavity is lined with a serous membrane that can serve as a conduit for fluids—a property exploited in peritoneal dialysis **(Fig. 127-1)**. A small amount of serous fluid is normally present in the peritoneal space, with a protein content (consisting mainly of albumin) of <30 g/L and <300 white blood cells (WBCs, generally mononuclear cells) per microliter. In bacterial infections, leukocyte recruitment into the infected peritoneal cavity consists of an early influx of polymorphonuclear leukocytes (PMNs) and a prolonged subsequent phase of mononuclear cell migration. The phenotype of the infiltrating leukocytes during the course of inflammation is regulated primarily by resident-cell chemokine synthesis.

Figure 127-1

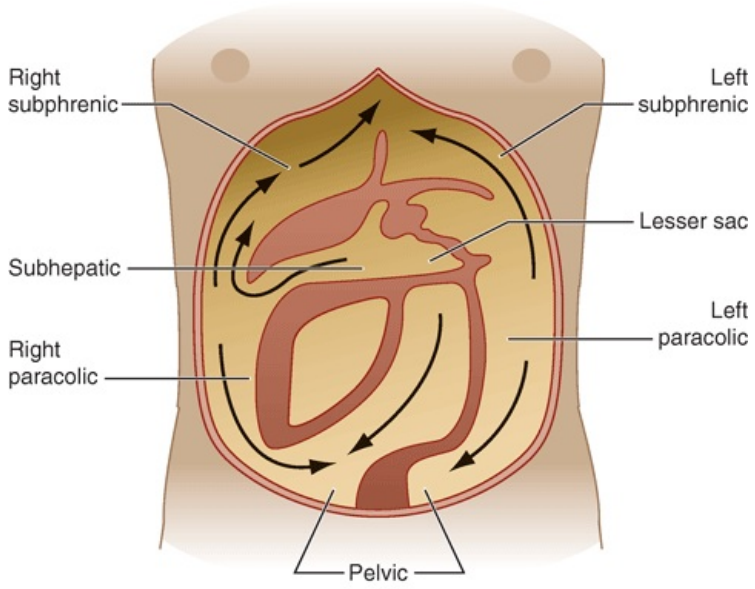

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of the intraperitoneal spaces, showing the circulation of fluid and potential areas for abscess formation. Some compartments collect fluid or pus more often than others. These compartments include the pelvis (the lowest portion), the subphrenic spaces on the right and left

compartments include the pelvis (the lowest portion), the subphrenic spaces on the right and left sides, and Morrison's pouch, which is a posterosuperior extension of the subhepatic spaces and is the lowest part of the paravertebral groove when a patient is recumbent. The falciform ligament separating the right and left subphrenic spaces appears to act as a barrier to the spread of infection; consequently, it is unusual to find bilateral subphrenic collections. *[Reprinted with permission from B Lorber (ed): Atlas of Infectious Diseases, vol VII: Intra-abdominal Infections, Hepatitis, and Gastroenteritis. Philadelphia, Current Medicine, 1996, p 1.13.]*

Primary (Spontaneous) Bacterial Peritonitis

Peritonitis is either primary (without an apparent source of contamination) or secondary. The types of organisms found and the clinical presentations of these two processes are different. In adults, primary bacterial peritonitis (PBP) occurs most commonly in conjunction with cirrhosis of the liver (frequently the result of alcoholism). However, the disease has been reported in adults with metastatic malignant disease, postnecrotic cirrhosis, chronic active hepatitis, acute viral hepatitis, congestive heart failure, systemic lupus erythematosus, and lymphedema as well as in patients with no underlying disease. Although PBP virtually always develops in patients with preexisting ascites, it is, in general, an uncommon event, occurring in ≤10% of cirrhotic patients. The cause of PBP has not been established definitively but is believed to involve hematogenous spread of organisms in a patient in whom a diseased liver and altered portal circulation result in a defect in the usual filtration function. Organisms multiply in ascites, a good medium for growth. The proteins of the complement cascade have been found in peritoneal fluid, with lower levels in cirrhotic patients than in patients with ascites of other etiologies. The opsonic and phagocytic properties of PMNs are diminished in patients with advanced liver disease.

The presentation of PBP differs from that of secondary peritonitis. The most common manifestation is fever, which is reported in up to 80% of patients. Ascites is found but virtually always predates infection. Abdominal pain, an acute onset of symptoms, and peritoneal irritation during physical examination can be helpful diagnostically, but the absence of any of these findings does not exclude this often-subtle diagnosis. Nonlocalizing symptoms (such as malaise, fatigue, or encephalopathy) without another clear etiology should also prompt consideration of PBP in a susceptible patient. It is vital to sample the peritoneal fluid of any cirrhotic patient with ascites and fever. The finding of >250 PMNs/μL is diagnostic for PBP, according to Conn (*http://jac.oxfordjournals.org/cgi/content/full/47/3/369*). This criterion does not apply to secondary peritonitis (see below). The microbiology of PBP is also distinctive. While enteric gram-negative bacilli such as *Escherichia coli* are most commonly encountered, gram-positive organisms such as streptococci, enterococci, or even pneumococci are sometimes found. In PBP, a single organism is typically isolated; anaerobes are found less frequently in PBP than in secondary peritonitis, in which a mixed flora including anaerobes is the rule. In fact, if PBP is suspected and multiple organisms including anaerobes are recovered from the peritoneal fluid, the diagnosis must be reconsidered and a source of secondary peritonitis sought.

The diagnosis of PBP is not easy. It depends on the exclusion of a primary intraabdominal source of infection. Contrast-enhanced CT is useful in identifying an intraabdominal source for infection. It may be difficult to recover organisms from cultures of peritoneal fluid, presumably because the burden of organisms is low. However, the yield can be improved if 10 mL of peritoneal fluid is placed directly into a blood culture bottle. Since bacteremia frequently accompanies PBP, blood should be cultured simultaneously. No specific radiographic studies are helpful in the diagnosis of PBP. A plain film of the abdomen would be expected to show ascites. Chest and abdominal radiography should be performed in patients with abdominal pain to exclude free air, which signals a perforation **(Fig. 127-2)**.

Figure 127-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pneumoperitoneum. Free air under the diaphragm on an upright chest film suggests the presence of a bowel perforation and associated peritonitis. *(Image courtesy of Dr. John Braver; with permission.)*

TREATMENT: PRIMARY BACTERIAL PERITONITIS

Treatment for PBP is directed at the isolate from blood or peritoneal fluid. Gram's staining of peritoneal fluid often gives negative results in PBP. Therefore, until culture results become available, therapy should cover gram-negative aerobic bacilli and gram-positive cocci. Third-generation cephalosporins such as cefotaxime (2 g q8h, administered IV) provide reasonable initial coverage in moderately ill patients. Broad-spectrum antibiotics, such as penicillin/β-lactamase inhibitor combinations (e.g., piperacillin/tazobactam, 3.375 g q6h IV for adults with normal renal function) or ceftriaxone (2 g q24h IV), are also options. Empirical coverage for anaerobes is not necessary. After the infecting organism is identified, therapy should be narrowed to target the specific pathogen. Patients with PBP usually respond within 72 h to appropriate antibiotic therapy. Antimicrobial treatment can be administered for as little as 5 days if rapid improvement occurs and blood cultures are negative, but a course of up to 2 weeks may be required for patients with bacteremia and for those whose improvement is slow. Persistence of WBCs in the ascitic fluid after therapy should prompt a search for additional diagnoses.

PREVENTION

Primary Prevention

One observational study raises the concern that proton pump inhibitor (PPI) therapy may increase the risk of PBP. No prospective studies have yet addressed whether avoidance of PPI therapy may prevent PBP.

Secondary Prevention

PBP has a high rate of recurrence. Up to 70% of patients experience a recurrence within 1 year. Antibiotic prophylaxis reduces this rate to <20% and improves short-term survival rates. Prophylactic regimens for adults with normal renal function include fluoroquinolones (ciprofloxacin, 750 mg weekly; norfloxacin, 400 mg/d) or trimethoprim-sulfamethoxazole (one double-strength tablet daily). However, long-term administration of broad-spectrum antibiotics in this setting has been shown to increase the risk of severe staphylococcal infections.

Secondary Peritonitis

Secondary peritonitis develops when bacteria contaminate the peritoneum as a result of spillage from an intraabdominal viscus. The organisms found almost always constitute a mixed flora in which facultative gram-negative bacilli and anaerobes predominate, especially when the contaminating source is colonic. Early in the course of infection, when the host response is directed toward containment of the infection, exudate containing fibrin and PMNs is found. Early death in this setting is attributable to gram-negative bacillary sepsis and to potent endotoxins circulating in the bloodstream (Chap. 271). Gram-negative bacilli, particularly *E. coli*, are common bloodstream isolates, but *Bacteroides fragilis* bacteremia also occurs. The severity of abdominal pain and the clinical course

depend on the inciting process. The organisms isolated from the peritoneum also vary with the source of the initial process and the normal flora at that site. Secondary peritonitis can result primarily from chemical irritation and/or bacterial contamination. For example, as long as the patient is not achlorhydric, a ruptured gastric ulcer will release low-pH gastric contents that will serve as a chemical irritant. The normal flora of the stomach comprises the same organisms found in the oropharynx (Chap. 164) but in lower numbers. Thus, the bacterial burden in a ruptured ulcer is negligible compared with that in a ruptured appendix. The normal flora of the colon below the ligament of Treitz contains ~10^{11} anaerobic organisms/g of feces but only 10^{8} aerobes/g; therefore, anaerobic species account for 99.9% of the bacteria. Leakage of colonic contents (pH 7–8) does not cause significant chemical peritonitis, but infection is intense because of the heavy bacterial load.

Depending on the inciting event, local symptoms may occur in secondary peritonitis—for example, epigastric pain from a ruptured gastric ulcer. In appendicitis (Chap. 300), the initial presenting symptoms are often vague, with periumbilical discomfort and nausea followed in a number of hours by pain more localized to the right lower quadrant. Unusual locations of the appendix (including a retrocecal position) can complicate this presentation further. Once infection has spread to the peritoneal cavity, pain increases, particularly with infection involving the parietal peritoneum, which is innervated extensively. Patients usually lie motionless, often with knees drawn up to avoid stretching the nerve fibers of the peritoneal cavity. Coughing and sneezing, which increase pressure within the peritoneal cavity, are associated with sharp pain. There may or may not be pain localized to the infected or diseased organ from which secondary peritonitis has arisen. Patients with secondary peritonitis generally have abnormal findings on abdominal examination, with marked voluntary and involuntary guarding of the anterior abdominal musculature. Later findings include tenderness, especially rebound tenderness. In addition, there may be localized findings in the area of the inciting event. In general, patients are febrile, with marked leukocytosis and a left shift of the WBCs to band forms.

While recovery of organisms from peritoneal fluid is easier in secondary than in primary peritonitis, a tap of the abdomen is rarely the procedure of choice in secondary peritonitis. An exception is in cases involving trauma, where the possibility of a hemoperitoneum may need to be excluded early. Emergent studies (such as abdominal CT) to find the source of peritoneal contamination should be undertaken if the patient is hemodynamically stable; unstable patients may require surgical intervention without prior imaging.

TREATMENT: SECONDARY PERITONITIS

Treatment for secondary peritonitis includes early administration of antibiotics aimed particularly at aerobic gram-negative bacilli and anaerobes (see below). Mild to moderate disease can be treated with many drugs covering these organisms, including broad-spectrum penicillin/β-lactamase inhibitor combinations (e.g., ticarcillin/clavulanate, 3.1 g q4–6h IV), cefoxitin (2 g q4–6h IV), or a combination of a fluoroquinolone (e.g., levofloxacin, 750 mg q24h IV) or a third-generation cephalosporin (e.g., ceftriaxone, 2 g q24h IV) plus metronidazole (500 mg q8h IV). Patients in intensive care units should receive imipenem (500 mg q6h IV), meropenem (1 g q8h IV), or combinations of drugs, such as ampicillin plus metronidazole plus ciprofloxacin. The role of enterococci and *Candida* spp. in mixed infections is controversial. Secondary peritonitis usually requires both surgical intervention to address the inciting process and antibiotics to treat early bacteremia, to decrease the incidence of abscess formation and wound infection, and to prevent distant spread of infection. While surgery is rarely indicated in PBP in adults, it may be life-saving in secondary peritonitis. Recombinant human activated protein C has been shown to reduce mortality rates among patients with severe sepsis and may benefit some patients with secondary peritonitis.

Peritonitis may develop as a complication of abdominal surgeries. These infections may be accompanied by localizing pain and/or nonlocalizing symptoms such as fever, malaise, anorexia, and toxicity. As a nosocomial infection, postoperative peritonitis may be associated with organisms such as staphylococci, components of the gram-negative hospital microflora, and the microbes that cause PBP and secondary peritonitis, as described above.

Peritonitis in Patients Undergoing CAPD

A third type of peritonitis arises in patients who are undergoing continuous ambulatory peritoneal dialysis (CAPD). Unlike PBP and secondary peritonitis, which are caused by endogenous bacteria, CAPD-associated peritonitis usually involves skin organisms. The pathogenesis of infection is similar to that of intravascular device–related infection, in which skin organisms migrate along the catheter, which both serves as an entry point and exerts the effects of a foreign body. Exit-site or tunnel infection may or may not accompany CAPD-associated peritonitis. Like PBP, CAPD-associated peritonitis is usually caused by a single organism. Peritonitis is, in fact, the most common reason for discontinuation of CAPD. Improvements in equipment design, especially the Y-set connector, have resulted in a decrease from one case of peritonitis per 9 months of CAPD to one case per 24 months.

The clinical presentation of CAPD peritonitis resembles that of secondary peritonitis in that diffuse pain and peritoneal signs are common. The dialysate is usually cloudy and contains >100 WBCs/μL, >50%

of which are neutrophils. The most common organisms are *Staphylococcus* spp., which accounted for ~45% of cases in one series. Historically, coagulase-negative staphylococcal species were identified most commonly in these infections, but more recently these isolates have been decreasing in frequency. *Staphylococcus aureus* is more often involved among patients who are nasal carriers of the organism than among those who are not, and this organism is the most common pathogen in overt exit-site infections. Gram-negative bacilli and fungi such as *Candida* spp. are also found. Vancomycin-resistant enterococci and vancomycin-intermediate *S. aureus* have been reported to produce peritonitis in CAPD patients. The finding of more than one organism in dialysate culture should prompt evaluation for secondary peritonitis. As with PBP, culture of dialysate fluid in blood culture bottles improves the yield. To facilitate diagnosis, several hundred milliliters of removed dialysis fluid should be concentrated by centrifugation before culture.

Treatment: CAPD Peritonitis

Empirical therapy for CAPD peritonitis should be directed at *S. aureus*, coagulase-negative *Staphylococcus*, and gram-negative bacilli until the results of cultures are available. Guidelines issued in 2005 suggest that agents should be chosen on the basis of local experience with resistant organisms. In some centers, a first-generation cephalosporin such as cefazolin (for gram-positive bacteria) and a fluoroquinolone or a third-generation cephalosporin such as ceftazidime (for gram-negative bacteria) may be reasonable; in areas with high rates of infection with methicillin-resistant *S. aureus*, vancomycin should be used instead of cefazolin, and gram-negative coverage may need to be broadened. Broad coverage including vancomycin should be particularly considered for toxic patients and for those with exit-site infections. Loading doses are administered intraperitoneally; doses depend on the dialysis method and the patient's renal function. Antibiotics are given either continuously (i.e., with each exchange) or intermittently (i.e., once daily, with the dose allowed to remain in the peritoneal cavity for at least 6 h). If the patient is severely ill, IV antibiotics should be added at doses appropriate for the patient's degree of renal failure. The clinical response to an empirical treatment regimen should be rapid; if the patient has not responded after 48–96 h of treatment, catheter removal should be considered.

Tuberculous Peritonitis

See Chap. 165.

INTRAABDOMINAL ABSCESSES

Intraperitoneal Abscesses

Abscess formation is common in untreated peritonitis if overt gram-negative sepsis either does not develop or develops but is not fatal. In experimental models of abscess formation, mixed aerobic and anaerobic organisms have been implanted intraperitoneally. Without therapy directed at anaerobes, animals develop intraabdominal abscesses. As in humans, these experimental abscesses may stud the peritoneal cavity, lie within the omentum or mesentery, or even develop on the surface of or within viscera such as the liver.

PATHOGENESIS AND IMMUNITY

There is often disagreement about whether an abscess represents a disease state or a host response. In a sense, it represents both: while an abscess is an infection in which viable infecting organisms and PMNs are contained in a fibrous capsule, it is also a process by which the host confines microbes to a limited space, thereby preventing further spread of infection. In any event, abscesses do cause significant symptoms, and patients with abscesses can be quite ill. Experimental work has helped to define both the host cells and the bacterial virulence factors responsible—most notably in the case of *B. fragilis*. This organism, although accounting for only 0.5% of the normal colonic flora, is the anaerobe most frequently isolated from intraabdominal infections, is especially prominent in abscesses, and is the most common anaerobic bloodstream isolate. On clinical grounds, therefore, *B. fragilis* appears to be uniquely virulent. Moreover, *B. fragilis* acts alone to cause abscesses in animal models of intraabdominal infection, whereas most other *Bacteroides* species must act synergistically with a facultative organism to induce abscess formation.

Of the several virulence factors identified in *B. fragilis*, one is critical: the capsular polysaccharide complex (CPC) found on the bacterial surface. The CPC comprises at least eight distinct surface polysaccharides. Structural analysis of these polysaccharides has shown an unusual motif of oppositely charged sugars. Polysaccharides having these *zwitterionic* characteristics, such as polysaccharide A (PSA), evoke a host response in the peritoneal cavity that localizes bacteria into abscesses. *B. fragilis* and PSA have been found to adhere to primary mesothelial cells in vitro; this adherence, in turn, stimulates the production of tumor necrosis factor α (TNF-α) and intercellular adhesion molecule 1 (ICAM-1) by peritoneal macrophages. Although abscesses characteristically contain PMNs, the process of abscess induction depends on the stimulation of T lymphocytes by these unique zwitterionic polysaccharides. The stimulated CD4+ T lymphocytes secrete leukoattractant cytokines and chemokines. The alternative pathway of complement and fibrinogen also

participate in abscess formation.

While antibodies to the CPC enhance bloodstream clearance of *B. fragilis*, CD4+ T cells are critical in immunity to abscesses. When administered subcutaneously, *B. fragilis* PSA has immunomodulatory characteristics and stimulates CD4+ T regulatory cells via an interleukin (IL) 2–dependent mechanism to produce IL-10. IL-10 downregulates the inflammatory response, thereby preventing abscess formation.

CLINICAL PRESENTATION

Of all intraabdominal abscesses, 74% are intraperitoneal or retroperitoneal and are not visceral. Most intraperitoneal abscesses result from fecal spillage from a colonic source, such as an inflamed appendix. Abscesses can also arise from other processes. They usually form within weeks of the development of peritonitis and may be found in a variety of locations—from omentum to mesentery, pelvis to psoas muscles, and subphrenic space to a visceral organ such as the liver, where they may develop either on the surface of the organ or within it. Periappendiceal and diverticular abscesses occur commonly. Diverticular abscesses are least likely to rupture. Infections of the female genital tract and pancreatitis are also among the more common causative events. When abscesses occur in the female genital tract—either as a primary infection (e.g., tuboovarian abscess) or as an infection extending into the pelvic cavity or peritoneum—*B. fragilis* figures prominently among the organisms isolated. *B. fragilis* is not found in large numbers in the normal vaginal flora. For example, it is encountered less commonly in pelvic inflammatory disease and endometritis without an associated abscess. In pancreatitis with leakage of damaging pancreatic enzymes, inflammation is prominent. Therefore, clinical findings such as fever, leukocytosis, and even abdominal pain do not distinguish pancreatitis itself from complications such as pancreatic pseudocyst, pancreatic abscess (Chap. 313), or intraabdominal collections of pus. Especially in cases of necrotizing pancreatitis, in which the incidence of local pancreatic infection may be as high as 30%, needle aspiration under CT guidance is performed to sample fluid for culture. Many centers prescribe preemptive antibiotics for patients with necrotizing pancreatitis. Imipenem is frequently used for this purpose since it reaches high tissue levels in the pancreas (although it is not unique in this regard). If needle aspiration yields infected fluid in the setting of acute necrotizing pancreatitis, most experts agree that surgery is superior to percutaneous drainage. Infected pseudocysts that occur remotely from acute pancreatitis are unlikely to be associated with significant amounts of necrotic tissue and may be treated with either surgical or percutaneous catheter drainage in conjunction with appropriate antibiotic therapy.

DIAGNOSIS

Scanning procedures have considerably facilitated the diagnosis of intraabdominal abscesses. Abdominal CT probably has the highest yield, although ultrasonography is particularly useful for the right upper quadrant, kidneys, and pelvis. Both indium-labeled WBCs and gallium tend to localize in abscesses and may be useful in finding a collection. Since gallium is taken up in the bowel, indium-labeled WBCs may have a slightly greater yield for abscesses near the bowel. Neither indium-labeled WBC nor gallium scans serve as a basis for a definitive diagnosis, however; both need to be followed by other, more specific studies, such as CT, if an area of possible abnormality is identified. Abscesses contiguous with or contained within diverticula are particularly difficult to diagnose with scanning procedures. Occasionally, a barium enema may detect a diverticular abscess not diagnosed by other procedures, although barium should not be injected if a perforation is suspected. If one study is negative, a second study sometimes reveals a collection. Although exploratory laparotomy has been less commonly used since the advent of CT, this procedure still must be undertaken on occasion if an abscess is strongly suspected on clinical grounds.

TREATMENT: INTRAPERITONEAL ABSCESSES

An algorithm for the management of patients with intraabdominal (including intraperitoneal) abscesses is presented in **Fig. 127-3**. The treatment of intraabdominal infections involves the determination of the initial focus of infection, the administration of broad-spectrum antibiotics targeting the organisms involved, and the performance of a drainage procedure if one or more definitive abscesses have formed. Antimicrobial therapy, in general, is adjunctive to drainage and/or surgical correction of an underlying lesion or process in intraabdominal abscesses. Unlike the intraabdominal abscesses resulting from most causes, for which drainage of some kind is generally required, abscesses associated with diverticulitis usually wall off locally after rupture of a diverticulum, so that surgical intervention is not routinely required.

Figure 127-3

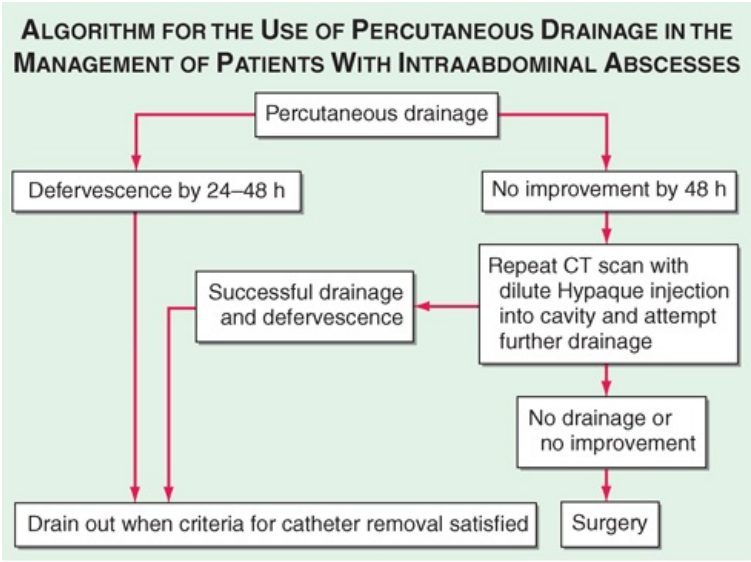

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the management of patients with intraabdominal abscesses using percutaneous drainage. Antimicrobial therapy should be administered concomitantly. *[Reprinted with permission from B Lorber (ed): Atlas of Infectious Diseases, vol VII: Intra-abdominal Infections, Hepatitis, and Gastroenteritis. Philadelphia, Current Medicine, 1996, p 1.30, as adapted from OD Rotstein, RL Simmons, in SL Gorbach et al (eds): Infectious Diseases. Philadelphia, Saunders, 1992, p 668.]*

A number of agents exhibit excellent activity against aerobic gram-negative bacilli. Since death in intraabdominal sepsis is linked to gram-negative bacteremia, empirical therapy for intraabdominal infection always needs to include adequate coverage of gram-negative aerobic, facultative, and anaerobic organisms. Even if anaerobes are not cultured from clinical specimens, they still must be covered by the therapeutic regimen. Empirical antibiotic therapy should be the same as that discussed above for secondary peritonitis.

Visceral Abscesses

LIVER ABSCESSES

The liver is the organ most subject to the development of abscesses. In one study of 540 intraabdominal abscesses, 26% were visceral. Liver abscesses made up 13% of the total number, or 48% of all visceral abscesses. Liver abscesses may be solitary or multiple; they may arise from hematogenous spread of bacteria or from local spread from contiguous sites of infection within the peritoneal cavity. In the past, appendicitis with rupture and subsequent spread of infection was the most common source for a liver abscess. Currently, associated disease of the biliary tract is most common. Pylephlebitis (suppurative thrombosis of the portal vein), usually arising from infection in the pelvis but sometimes from infection elsewhere in the peritoneal cavity, is another common source for bacterial seeding of the liver.

Fever is the most common presenting sign of liver abscess. Some patients, particularly those with associated disease of the biliary tract, have symptoms and signs localized to the right upper quadrant, including pain, guarding, punch tenderness, and even rebound tenderness. Nonspecific symptoms, such as chills, anorexia, weight loss, nausea, and vomiting, may also develop. Only 50% of patients with liver abscesses, however, have hepatomegaly, right-upper-quadrant tenderness, or jaundice; thus, one-half of patients have no symptoms or signs to direct attention to the liver. Fever of unknown origin (FUO) may be the only manifestation of liver abscess, especially in the elderly. Diagnostic studies of the abdomen, especially the right upper quadrant, should be a part of any FUO workup. The single most reliable laboratory finding is an elevated serum concentration of alkaline phosphatase, which is documented in 70% of patients with liver abscesses. Other tests of liver function may yield normal results, but 50% of patients have elevated serum levels of bilirubin, and 48% have elevated concentrations of aspartate aminotransferase. Other laboratory findings include leukocytosis in 77% of patients, anemia (usually normochromic, normocytic) in 50%, and hypoalbuminemia in 33%. Concomitant bacteremia is found in one-third to one-half of patients. A liver abscess is sometimes suggested by chest radiography, especially if a new elevation of the right hemidiaphragm is seen; other suggestive findings include a right basilar infiltrate and a right pleural effusion.

Imaging studies are the most reliable methods for diagnosing liver abscesses. These studies include ultrasonography, CT **(Fig. 127-4)**, indium-labeled WBC or gallium scan, and MRI. More than one such study may be required. Organisms recovered from liver abscesses vary with the source. In liver infection arising from the biliary tree, enteric gram-negative aerobic bacilli and enterococci are common

isolates. Unless previous surgery has been performed, anaerobes are not generally involved in liver abscesses arising from biliary infections. In contrast, in liver abscesses arising from pelvic and other intraperitoneal sources, a mixed flora including both aerobic and anaerobic species is common; *B. fragilis* is the species most frequently isolated. With hematogenous spread of infection, usually only a single organism is encountered; this species may be *S. aureus* or a streptococcal species such as *S. milleri*. Results of cultures obtained from drain sites are not reliable for defining the etiology of infections. Liver abscesses may also be caused by *Candida* spp.; such abscesses usually follow fungemia in patients receiving chemotherapy for cancer and often present when PMNs return after a period of neutropenia. Amebic liver abscesses are not an uncommon problem (Chap. 209). Amebic serologic testing gives positive results in >95% of cases; thus, a negative result helps to exclude this diagnosis.

Figure 127-4

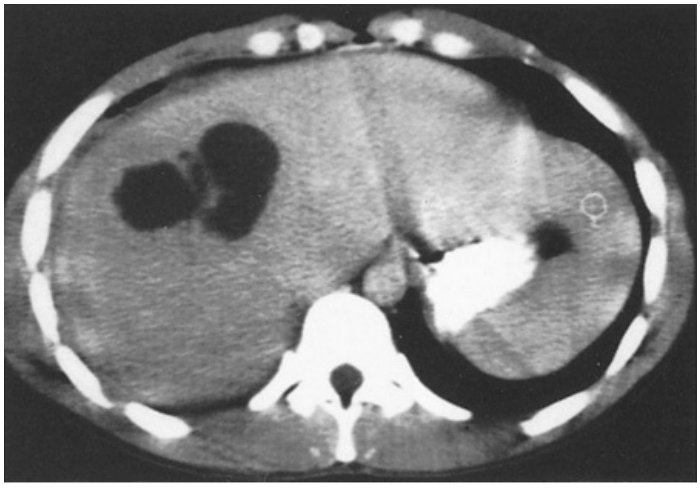

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multilocular liver abscess on CT scan. Multiple or multilocular abscesses are more common than solitary abscesses. *[Reprinted with permission from B Lorber (ed): Atlas of Infectious Diseases, Vol VII: Intra-abdominal Infections, Hepatitis, and Gastroenteritis. Philadelphia, Current Medicine, 1996, Fig. 1.22.]*

Treatment: Liver Abscesses

(Fig. 127-3) While drainage—either percutaneous (with a pigtail catheter kept in place) or surgical—is the mainstay of therapy for intraabdominal abscesses (including liver abscesses), there is growing interest in medical management alone for pyogenic liver abscesses. The drugs used for empirical therapy include the same ones used in intraabdominal sepsis and secondary bacterial peritonitis. Usually, blood cultures and a diagnostic aspirate of abscess contents should be obtained before the initiation of empirical therapy, with antibiotic choices adjusted when the results of Gram's staining and culture become available. Cases treated without definitive drainage generally require longer courses of antibiotic therapy. When percutaneous drainage was compared with open surgical drainage, the average length of hospital stay for the former was almost twice that for the latter, although both the time required for fever to resolve and the mortality rate were the same for the two procedures. The mortality rate was appreciable despite treatment, averaging 15%. Several factors predict the failure of percutaneous drainage and therefore may favor primary surgical intervention. These factors include the presence of multiple, sizable abscesses; viscous abscess contents that tend to plug the catheter; associated disease (e.g., disease of the biliary tract) requiring surgery; or the lack of a clinical response to percutaneous drainage in 4–7 days.

Treatment of candidal liver abscesses often entails initial administration of amphotericin B or liposomal amphotericin, with subsequent fluconazole therapy (Chap. 203). In some cases, therapy with fluconazole alone (6 mg/kg daily) may be used—e.g., in clinically stable patients whose infecting isolate is susceptible to this drug.

SPLENIC ABSCESSES

Splenic abscesses are much less common than liver abscesses. The incidence of splenic abscesses has ranged from 0.14% to 0.7% in various autopsy series. The clinical setting and the organisms isolated usually differ from those for liver abscesses. The degree of clinical suspicion for splenic abscess needs to be high, as this condition is frequently fatal if left untreated. Even in the most recently published series, diagnosis was made only at autopsy in 37% of cases. While splenic abscesses may arise occasionally from contiguous spread of infection or from direct trauma to the

spleen, hematogenous spread of infection is more common. Bacterial endocarditis is the most common associated infection (Chap. 124). Splenic abscesses can develop in patients who have received extensive immunosuppressive therapy (particularly those with malignancy involving the spleen) and in patients with hemoglobinopathies or other hematologic disorders (especially sickle cell anemia).

While ~50% of patients with splenic abscesses have abdominal pain, the pain is localized to the left upper quadrant in only one-half of these cases. Splenomegaly is found in ~50% of cases. Fever and leukocytosis are generally present; the development of fever preceded diagnosis by an average of 20 days in one series. Left-sided chest findings may include abnormalities to auscultation, and chest radiographic findings may include an infiltrate or a left-sided pleural effusion. CT scan of the abdomen has been the most sensitive diagnostic tool. Ultrasonography can yield the diagnosis but is less sensitive. Liver-spleen scan or gallium scan may also be useful. Streptococcal species are the most common bacterial isolates from splenic abscesses, followed by *S. aureus*—presumably reflecting the associated endocarditis. An increase in the prevalence of gram-negative aerobic isolates from splenic abscesses has been reported; these organisms often derive from a urinary tract focus, with associated bacteremia, or from another intraabdominal source. *Salmonella* species are seen fairly commonly, especially in patients with sickle cell hemoglobinopathy. Anaerobic species accounted for only 5% of isolates in the largest collected series, but the reporting of a number of "sterile abscesses" may indicate that optimal techniques for the isolation of anaerobes were not employed.

Treatment: Splenic Abscesses

Because of the high mortality figures reported for splenic abscesses, splenectomy with adjunctive antibiotics has traditionally been considered standard treatment and remains the best approach for complex, multilocular abscesses or multiple abscesses. However, percutaneous drainage has worked well for single, small (<3-cm) abscesses in some studies and may also be useful for patients with high surgical risk. Patients undergoing splenectomy should be vaccinated against encapsulated organisms (*Streptococcus pneumoniae*, *Haemophilus influenzae*, *Neisseria meningitidis*). The most important factor in successful treatment of splenic abscesses is early diagnosis.

PERINEPHRIC AND RENAL ABSCESSES

Perinephric and renal abscesses are not common: The former accounted for only ~0.02% of hospital admissions and the latter for ~0.2% in Altemeier's series of 540 intraabdominal abscesses. Before antibiotics became available, most renal and perinephric abscesses were hematogenous in origin, usually complicating prolonged bacteremia, with *S. aureus* most commonly recovered. Now, in contrast, >75% of perinephric and renal abscesses arise from a urinary tract infection. Infection ascends from the bladder to the kidney, with pyelonephritis occurring prior to abscess development. Bacteria may directly invade the renal parenchyma from medulla to cortex. Local vascular channels within the kidney may also facilitate the transport of organisms. Areas of abscess developing within the parenchyma may rupture into the perinephric space. The kidneys and adrenal glands are surrounded by a layer of perirenal fat that, in turn, is surrounded by Gerota's fascia, which extends superiorly to the diaphragm and inferiorly to the pelvic fat. Abscesses extending into the perinephric space may track through Gerota's fascia into the psoas or transversalis muscles, into the anterior peritoneal cavity, superiorly to the subdiaphragmatic space, or inferiorly to the pelvis. Of the risk factors that have been associated with the development of perinephric abscesses, the most important is concomitant nephrolithiasis obstructing urinary flow. Of patients with perinephric abscess, 20–60% have renal stones. Other structural abnormalities of the urinary tract, prior urologic surgery, trauma, and diabetes mellitus have also been identified as risk factors.

The organisms most frequently encountered in perinephric and renal abscesses are *E. coli*, *Proteus* spp., and *Klebsiella* spp. *E. coli*, the aerobic species most commonly found in the colonic flora, seems to have unique virulence properties in the urinary tract, including factors promoting adherence to uroepithelial cells. The urease of *Proteus* spp. splits urea, thereby creating a more alkaline and more hospitable environment for bacterial proliferation. *Proteus* spp. are frequently found in association with large struvite stones caused by the precipitation of magnesium ammonium sulfate in an alkaline environment. These stones serve as a nidus for recurrent urinary tract infection. While a single bacterial species is usually recovered from a perinephric or renal abscess, multiple species may also be found. If a urine culture is not contaminated with periurethral flora and is found to contain more than one organism, a perinephric abscess or renal abscess should be considered in the differential diagnosis. Urine cultures may also be polymicrobial in cases of bladder diverticulum.

Candida spp. can cause renal abscesses. This fungus may spread to the kidney hematogenously or by ascension from the bladder. The hallmark of the latter route of infection is ureteral obstruction with large fungal balls.

The presentation of perinephric and renal abscesses is quite nonspecific. Flank pain and abdominal pain are common. At least 50% of patients are febrile. Pain may be referred to the groin or leg, particularly with extension of infection. The diagnosis of perinephric abscess, like that of splenic abscess, is frequently delayed, and the mortality rate in some series is appreciable, although lower than in the past. Perinephric or renal abscess should be most seriously considered when a patient

presents with symptoms and signs of pyelonephritis and remains febrile after 4 or 5 days of treatment. Moreover, when a urine culture yields a polymicrobial flora, when a patient is known to have renal stones, or when fever and pyuria coexist with a sterile urine culture, these diagnoses should be entertained.

Renal ultrasonography and abdominal CT are the most useful diagnostic modalities. If a renal or perinephric abscess is diagnosed, nephrolithiasis should be excluded, especially when a high urinary pH suggests the presence of a urea-splitting organism.

Treatment: Perinephric and Renal Abscesses

Treatment for perinephric and renal abscesses, like that for other intraabdominal abscesses, includes drainage of pus and antibiotic therapy directed at the organism(s) recovered. For perinephric abscesses, percutaneous drainage is usually successful.

PSOAS ABSCESSES

The psoas muscle is another location in which abscesses are encountered. Psoas abscesses may arise from a hematogenous source, by contiguous spread from an intraabdominal or pelvic process, or by contiguous spread from nearby bony structures (e.g., vertebral bodies). Associated osteomyelitis due to spread from bone to muscle or from muscle to bone is common in psoas abscesses. When Pott's disease was common, *Mycobacterium tuberculosis* was a frequent cause of psoas abscess. Currently, either *S. aureus* or a mixture of enteric organisms including aerobic and anaerobic gram-negative bacilli is usually isolated from psoas abscesses in the United States. *S. aureus* is most likely to be isolated when a psoas abscess arises from hematogenous spread or a contiguous focus of osteomyelitis; a mixed enteric flora is the most likely etiology when the abscess has an intraabdominal or pelvic source. Patients with psoas abscesses frequently present with fever, lower abdominal or back pain, or pain referred to the hip or knee. CT is the most useful diagnostic technique.

Treatment: Psoas Abscesses

Treatment includes surgical drainage and the administration of an antibiotic regimen directed at the inciting organism(s).

PANCREATIC ABSCESSES

See Chap. 313.

ACKNOWLEDGMENT

The substantial contributions of Dori F. Zaleznik, MD, to this chapter in previous editions are gratefully acknowledged.

FURTHER READINGS

Bajaj JS et al: Association of proton pump inhibitor therapy with spontaneous bacterial peritonitis in cirrhotic patients with ascites. Am J Gastroenterol 104:1130, 2009[PMID: 19337238]

Barie PS et al: Benefit/risk of drotrecogin alpha [activated] in surgical patients with severe sepsis. Am J Surg 188:212, 2004[PMID: 15450822]

Meddings L et al: A population-based study of pyogenic liver abscesses in the United States: mortality, and temporal trends. Am J Gastroenterol 105:117, 2010[PMID: 19888200]

Navarro López V et al: GTI-SEMI Group: Microbiology and outcome of iliopsoas abscess in 124 patients. Medicine (Baltimore) 88:120, 2009

Piraino B et al: Peritoneal dialysis–related infections recommendations: 2005 update. Perit Dial Int 25:107, 2005[PMID: 15796137]

Saab S et al: Oral antibiotic prophylaxis reduces spontaneous bacterial peritonitis occurrence and improves short-term survival in cirrhosis: A meta-analysis. Am J Gastroenterol 104:993, 2009[PMID: 19277033]

Sifri CD, Madoff LC: Infections of the liver and biliary system, in *Principles and Practice of Infectious Diseases*, 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 1035–1044

Solomkin JS, Mazuski J: Intra-abdominal sepsis: Newer interventional and antimicrobial therapies. Infect Dis Clin North Am 23:593, 2009[PMID: 19665085]

Tzianabos AO et al: T cells activated by zwitterionic molecules prevent abscesses induced by pathogenic bacteria. J Biol Chem 275:6733, 2000[PMID: 10702228]

van Ruler O et al: Comparison of on-demand vs planned relaparotomy strategy in patients with severe peritonitis: A randomized trial. JAMA 298:865, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACUTE INFECTIOUS DIARRHEAL DISEASES AND BACTERIAL FOOD POISONING: INTRODUCTION

Ranging from a mild annoyance to a devastating dehydrating illness, acute diarrheal disease is a leading cause of illness globally, with an estimated 4.6 billion episodes worldwide per year. Diarrheal disease ranks second only to lower respiratory infection as the most common infectious cause of death worldwide. Among children <5 years old, diarrheal disease is a particularly important cause of death. Every year nearly 2 million children in this age group die of diarrheal disease; the majority of these young children are impoverished and live in resource-poor areas. By contributing to malnutrition and thereby reducing resistance to other infectious agents, diarrheal disease is also an indirect factor in a far greater burden of disease.

The wide range of clinical manifestations of acute gastrointestinal illnesses is matched by the wide variety of infectious agents involved, including viruses, bacteria, and parasitic pathogens **(Table 128–1)**. This chapter discusses factors that enable gastrointestinal pathogens to cause disease, reviews host defense mechanisms, and delineates an approach to the evaluation and treatment of patients presenting with acute diarrhea. Individual organisms causing acute gastrointestinal illnesses are discussed in detail in subsequent chapters.

Table 128–1 Gastrointestinal Pathogens Causing Acute Diarrhea

Mechanism	Location	Illness	Stool Findings	Examples of Pathogens Involved
Noninflammatory (enterotoxin)	Proximal small bowel	Watery diarrhea	No fecal leukocytes; mild or no increase in fecal lactoferrin	*Vibrio cholerae*, enterotoxigenic *Escherichia coli* (LT and/or ST), enteroaggregative *E. coli, Clostridium perfringens, Bacillus cereus, Staphylococcus aureus, Aeromonas hydrophila, Plesiomonas shigelloides*, rotavirus, norovirus, enteric adenoviruses, *Giardia lamblia, Cryptosporidium* spp., *Cyclospora* spp., microsporidia
Inflammatory (invasion or cytotoxin)	Colon or distal small bowel	Dysentery or inflammatory diarrhea	Fecal polymorphonuclear leukocytes; substantial increase in fecal lactoferrin	*Shigella* spp., *Salmonella* spp., *Campylobacter jejuni*, enterohemorrhagic *E. coli*, enteroinvasive *E. coli, Yersinia enterocolitica, Listeria monocytogenes, Vibrio parahaemolyticus, Clostridium difficile, A. hydrophila, P. shigelloides, Entamoeba histolytica, Klebsiella oxytoca*
Penetrating	Distal small bowel	Enteric fever	Fecal mononuclear leukocytes	*Salmonella typhi, Y. enterocolitica*

Abbreviations: LT, heat-labile enterotoxin; ST, heat-stable enterotoxin.

Source: After Steiner and Guerrant.

PATHOGENIC MECHANISMS

Enteric pathogens have developed a variety of tactics to overcome host defenses. Understanding the virulence factors employed by these organisms is important in the diagnosis and treatment of clinical disease.

INOCULUM SIZE

The number of microorganisms that must be ingested to cause disease varies considerably from species to species. For *Shigella*, enterohemorrhagic *Escherichia coli*, *Giardia lamblia*, or *Entamoeba*, as few as 10–100 bacteria or cysts can produce infection, while 10^5–10^8 *Vibrio cholerae* organisms

must be ingested orally to cause disease. The infective dose of *Salmonella* varies widely, depending on the species, host, and food vehicle. The ability of organisms to overcome host defenses has important implications for transmission; *Shigella*, enterohemorrhagic *E. coli*, *Entamoeba*, and *Giardia* can spread by person-to-person contact, whereas under some circumstances *Salmonella* may have to grow in food for several hours before reaching an effective infectious dose.

ADHERENCE

Many organisms must adhere to the gastrointestinal mucosa as an initial step in the pathogenic process; thus, organisms that can compete with the normal bowel flora and colonize the mucosa have an important advantage in causing disease. Specific cell-surface proteins involved in attachment of bacteria to intestinal cells are important virulence determinants. *V. cholerae*, for example, adheres to the brush border of small-intestinal enterocytes via specific surface adhesins, including the toxin-coregulated pilus and other accessory colonization factors. Enterotoxigenic *E. coli*, which causes watery diarrhea, produces an adherence protein called *colonization factor antigen* that is necessary for colonization of the upper small intestine by the organism prior to the production of enterotoxin. Enteropathogenic *E. coli*, an agent of diarrhea in young children, and enterohemorrhagic *E. coli*, which causes hemorrhagic colitis and the hemolytic-uremic syndrome, produce virulence determinants that allow these organisms to attach to and efface the brush border of the intestinal epithelium.

TOXIN PRODUCTION

The production of one or more exotoxins is important in the pathogenesis of numerous enteric organisms. Such toxins include *enterotoxins*, which cause watery diarrhea by acting directly on secretory mechanisms in the intestinal mucosa; *cytotoxins*, which cause destruction of mucosal cells and associated inflammatory diarrhea; and *neurotoxins*, which act directly on the central or peripheral nervous system.

The prototypical enterotoxin is cholera toxin, a heterodimeric protein composed of one A and five B subunits. The A subunit contains the enzymatic activity of the toxin, while the B subunit pentamer binds holotoxin to the enterocyte surface receptor, the ganglioside G_{M1}. After the binding of holotoxin, a fragment of the A subunit is translocated across the eukaryotic cell membrane into the cytoplasm, where it catalyzes the ADP-ribosylation of a GTP-binding protein and causes persistent activation of adenylate cyclase. The end result is an increase of cyclic AMP in the intestinal mucosa, which increases Cl^- secretion and decreases Na^+ absorption, leading to a loss of fluid and the production of diarrhea.

Enterotoxigenic strains of *E. coli* may produce a protein called *heat-labile enterotoxin* (LT) that is similar to cholera toxin and causes secretory diarrhea by the same mechanism. Alternatively, enterotoxigenic strains of *E. coli* may produce *heat-stable enterotoxin* (ST), one form of which causes diarrhea by activation of guanylate cyclase and elevation of intracellular cyclic GMP. Some enterotoxigenic strains of *E. coli* produce both LT and ST.

Bacterial cytotoxins, in contrast, destroy intestinal mucosal cells and produce the syndrome of dysentery, with bloody stools containing inflammatory cells. Enteric pathogens that produce such cytotoxins include *Shigella dysenteriae* type 1, *Vibrio parahaemolyticus*, and *Clostridium difficile*. *S. dysenteriae* type 1 and Shiga toxin–producing strains of *E. coli* produce potent cytotoxins and have been associated with outbreaks of hemorrhagic colitis and hemolytic-uremic syndrome.

Neurotoxins are usually produced by bacteria outside the host and therefore cause symptoms soon after ingestion. Included are the staphylococcal and *Bacillus cereus* toxins, which act on the central nervous system to produce vomiting.

INVASION

Dysentery may result not only from the production of cytotoxins but also from bacterial invasion and destruction of intestinal mucosal cells. Infections due to *Shigella* and enteroinvasive *E. coli* are characterized by the organisms' invasion of mucosal epithelial cells, intraepithelial multiplication, and subsequent spread to adjacent cells. *Salmonella* causes inflammatory diarrhea by invasion of the bowel mucosa but generally is not associated with the destruction of enterocytes or the full clinical syndrome of dysentery. *Salmonella typhi* and *Yersinia enterocolitica* can penetrate intact intestinal mucosa, multiply intracellularly in Peyer's patches and intestinal lymph nodes, and then disseminate through the bloodstream to cause enteric fever, a syndrome characterized by fever, headache, relative bradycardia, abdominal pain, splenomegaly, and leukopenia.

HOST DEFENSES

Given the enormous number of microorganisms ingested with every meal, the normal host must combat a constant influx of potential enteric pathogens. Studies of infections in patients with alterations in defense mechanisms have led to a greater understanding of the variety of ways in which the normal host can protect itself against disease.

NORMAL FLORA

The large numbers of bacteria that normally inhabit the intestine act as an important host defense by preventing colonization by potential enteric pathogens. Persons with fewer intestinal bacteria, such as infants who have not yet developed normal enteric colonization or patients receiving antibiotics, are at significantly greater risk of developing infections with enteric pathogens. The composition of the intestinal flora is as important as the number of organisms present. More than 99% of the normal colonic flora is made up of anaerobic bacteria, and the acidic pH and volatile fatty acids produced by these organisms appear to be critical elements in resistance to colonization.

GASTRIC ACID

The acidic pH of the stomach is an important barrier to enteric pathogens, and an increased frequency of infections due to *Salmonella*, *G. lamblia*, and a variety of helminths has been reported among patients who have undergone gastric surgery or are achlorhydric for some other reason. Neutralization of gastric acid with antacids, proton pump inhibitors, or H_2 blockers—a common practice in the management of hospitalized patients—similarly increases the risk of enteric colonization. In addition, some microorganisms can survive the extreme acidity of the gastric environment; rotavirus, for example, is highly stable to acidity.

INTESTINAL MOTILITY

Normal peristalsis is the major mechanism for clearance of bacteria from the proximal small intestine. When intestinal motility is impaired (e.g., by treatment with opiates or other antimotility drugs, anatomic abnormalities, or hypomotility states), the frequency of bacterial overgrowth and infection of the small bowel with enteric pathogens is increased. Some patients whose treatment for *Shigella* infection consists of diphenoxylate hydrochloride with atropine (Lomotil) experience prolonged fever and shedding of organisms, while patients treated with opiates for mild *Salmonella* gastroenteritis have a higher frequency of bacteremia than those not treated with opiates.

IMMUNITY

Both cellular immune responses and antibody production play important roles in protection from enteric infections. Humoral immunity to enteric pathogens consists of systemic IgG and IgM as well as secretory IgA. The mucosal immune system may be the first line of defense against many gastrointestinal pathogens. The binding of bacterial antigens to the luminal surface of M cells in the distal small bowel and the subsequent presentation of antigens to subepithelial lymphoid tissue lead to the proliferation of sensitized lymphocytes. These lymphocytes circulate and populate all of the mucosal tissues of the body as IgA-secreting plasma cells.

GENETIC DETERMINANTS

Host genetic variation influences susceptibility to diarrheal diseases. People with blood group O show increased susceptibility to disease due to *V. cholerae*, *Shigella*, *E. coli* O157, and norovirus. Polymorphisms in genes encoding inflammatory mediators have been associated with the outcome of infection with enteroaggregative *E. coli*, enterotoxin-producing *E. coli*, *Salmonella*, *C. difficile*, and *V. cholerae*.

Approach to the Patient: Infectious Diarrhea or Bacterial Food Poisoning

The approach to the patient with possible infectious diarrhea or bacterial food poisoning is shown in **Fig. 128-1**.

Figure 128-1

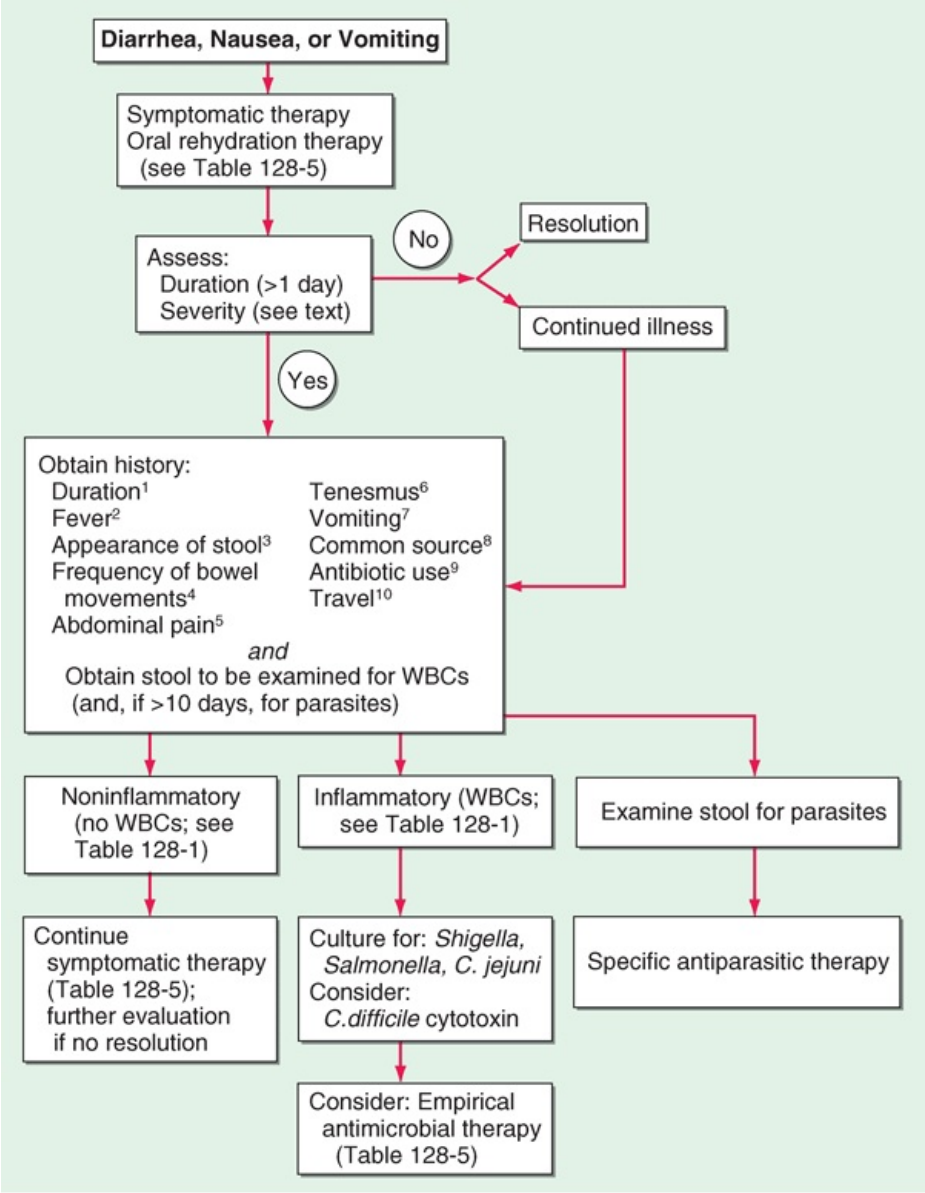

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical algorithm for the approach to patients with community-acquired infectious diarrhea or bacterial food poisoning. Key to superscripts: **1.** Diarrhea lasting >2 weeks is generally defined as chronic; in such cases, many of the causes of acute diarrhea are much less likely, and a new spectrum of causes needs to be considered. **2.** Fever often implies invasive disease, although fever and diarrhea may also result from infection outside the gastrointestinal tract, as in malaria. **3.** Stools that contain blood or mucus indicate ulceration of the large bowel. Bloody stools without fecal leukocytes should alert the laboratory to the possibility of infection with Shiga toxin–producing enterohemorrhagic *Escherichia coli*. Bulky white stools suggest a small-intestinal process that is causing malabsorption. Profuse "rice-water" stools suggest cholera or a similar toxigenic process. **4.** Frequent stools over a given period can provide the first warning of impending dehydration. **5.** Abdominal pain may be most severe in inflammatory processes like those due to *Shigella*, *Campylobacter*, and necrotizing toxins. Painful abdominal muscle cramps, caused by electrolyte loss, can develop in severe cases of cholera. Bloating is common in giardiasis. An appendicitis-like syndrome should prompt a culture for *Yersinia enterocolitica* with cold enrichment. **6.** Tenesmus (painful rectal spasms with a strong urge to defecate but little passage of stool) may be a feature of cases with proctitis, as in shigellosis or amebiasis. **7.** Vomiting implies an acute infection (e.g., a toxin-mediated illness or food poisoning) but can also be prominent in a variety of systemic illnesses (e.g., malaria) and in intestinal obstruction. **8.** Asking patients whether anyone else they know is sick is a more efficient means of identifying a common source than is constructing a list of recently eaten foods. If a common source seems likely, specific foods can be investigated. See text for a discussion of bacterial food poisoning. **9.** Current antibiotic therapy or a recent history of treatment suggests *Clostridium difficile* diarrhea (Chap. 129). Stop antibiotic treatment if possible and consider tests for *C. difficile* toxins. Antibiotic use may increase the risk of other infections, such as salmonellosis. **10.** See text (and Chap. 123) for a discussion of traveler's diarrhea. *(After Steiner and Guerrant; RL Guerrant, DA Bobak: N Engl J Med 325:327, 1991; with permission.)*

HISTORY

The answers to questions with high discriminating value can quickly narrow the range of potential causes of diarrhea and help determine whether treatment is needed. Important elements of the narrative history are detailed in Fig. 128-1.

PHYSICAL EXAMINATION

The examination of patients for signs of dehydration provides essential information about the severity of the diarrheal illness and the need for rapid therapy. Mild dehydration is indicated by thirst, dry mouth, decreased axillary sweat, decreased urine output, and slight weight loss. Signs of moderate dehydration include an orthostatic fall in blood pressure, skin tenting, and sunken eyes (or, in infants, a sunken fontanelle). Signs of severe dehydration include lethargy, obtundation, feeble pulse, hypotension, and frank shock.

DIAGNOSTIC APPROACH

After the severity of illness is assessed, the clinician must distinguish between *inflammatory* and *noninflammatory* disease. Using the history and epidemiologic features of the case as guides, the clinician can then rapidly evaluate the need for further efforts to define a specific etiology and for therapeutic intervention. Examination of a stool sample may supplement the narrative history. Grossly bloody or mucoid stool suggests an inflammatory process. A test for fecal leukocytes (preparation of a thin smear of stool on a glass slide, addition of a drop of methylene blue, and examination of the wet mount) can suggest inflammatory disease in patients with diarrhea, although the predictive value of this test is still debated. A test for fecal lactoferrin, which is a marker of fecal leukocytes, is more sensitive and is available in latex agglutination and enzyme-linked immunosorbent assay formats. Causes of acute infectious diarrhea, categorized as inflammatory and noninflammatory, are listed in Table 128–1.

POST-DIARRHEA COMPLICATIONS

Chronic complications may follow the resolution of an acute diarrheal episode. The clinician should inquire about prior diarrheal illness if the conditions listed in **Table 128–2** are observed.

Table 128–2 Post-Diarrhea Complications of Acute Infectious Diarrheal Illness

Complication	Comments
Chronic diarrhea • Lactase deficiency • Small-bowel bacterial overgrowth • Malabsorption syndromes (tropical and celiac sprue)	Occurs in ~1% of travelers with acute diarrhea • Protozoa account for ~1/3 of cases
Initial presentation or exacerbation of inflammatory bowel disease	May be precipitated by traveler's diarrhea
Irritable bowel syndrome	Occurs in ~10% of travelers with traveler's diarrhea
Reactive arthritis (formerly known as Reiter's syndrome)	Particularly likely after infection with invasive organisms (*Shigella, Salmonella, Campylobacter, Yersinia*)
Hemolytic-uremic syndrome (hemolytic anemia, thrombocytopenia, and renal failure)	Follows infection with Shiga toxin–producing bacteria (*Shigella dysenteriae* type 1 and enterohemorrhagic *Escherichia coli*)
Guillain-Barré syndrome	Particularly likely after *Campylobacter* infection

EPIDEMIOLOGY

TRAVEL HISTORY

Of the several million people who travel from temperate industrialized countries to tropical regions of Asia, Africa, and Central and South America each year, 20–50% experience a sudden onset of abdominal cramps, anorexia, and watery diarrhea; thus *traveler's diarrhea* is the most common travel-related infectious illness (Chap. 123). The time of onset is usually 3 days to 2 weeks after the traveler's arrival in a resource-poor area; most cases begin within the first 3–5 days. The illness is generally self-limited, lasting 1–5 days. The high rate of diarrhea among travelers to underdeveloped areas is related to the ingestion of contaminated food or water.

The organisms that cause traveler's diarrhea vary considerably with location **(Table 128–3)**, as does the pattern of antimicrobial resistance. In all areas, enterotoxigenic and enteroaggregative strains of *E. coli* are the most common isolates from persons with the classic secretory traveler's diarrhea syndrome. Infection with *Campylobacter jejuni* is especially common in areas of Asia.

Table 128–3 Causes of Traveler's Diarrhea

Etiologic Agent	Approximate Percentage of Cases	Comments
BACTERIA	**50–75**	
Enterotoxigenic *Escherichia coli*	10–45	Single most important agent
Enteroaggregative *E. coli*	5–35	Emerging enteric pathogen with worldwide distribution
Campylobacter jejuni	5–25	More common in Asia
Shigella	0–15	Major cause of dysentery
Salmonella	0–15	
Others	0–5	Including *Aeromonas, Plesiomonas*, and *Vibrio cholerae*
VIRUSES	**0–20**	
Norovirus	0–10	Associated with cruise ships
Rotavirus	0–5	Particularly common among children
PARASITES	**0–10**	
Giardia lamblia	0–5	Affects hikers and campers who drink from freshwater streams; contaminates water supplies in Russia
Cryptosporidium	0–5	Resistant to chlorine treatment
Entamoeba histolytica	<1	
Cyclospora	<1	
OTHER	**0–10**	
Acute food poisoning[a]	0–5	
No pathogen identified	10–50	

[a]For etiologic agents, see Table 128–4.

Source: After Hill et al.

LOCATION

Day-care centers have particularly high attack rates of enteric infections. Rotavirus is most common among children <2 years old, with attack rates of 75–100% among those exposed. *G. lamblia* is more common among older children, with somewhat lower attack rates. Other common organisms, often spread by fecal-oral contact, are *Shigella*, *C. jejuni*, and *Cryptosporidium*. A characteristic feature of infection among children attending day-care centers is the high rate of secondary cases among family members.

Similarly, hospitals are sites in which enteric infections are concentrated. Diarrhea is one of the most common manifestations of nosocomial infections. *C. difficile* is the predominant cause of nosocomial diarrhea among adults in the United States. *Klebsiella oxytoca* has been identified as a cause of antibiotic-associated hemorrhagic colitis. Viral pathogens, especially rotavirus, can spread rapidly in pediatric wards. Enteropathogenic *E. coli* has been associated with outbreaks of diarrhea in nurseries for newborns. One-third of elderly patients in chronic-care institutions develop a significant diarrheal illness each year; more than one-half of these cases are caused by cytotoxin-producing *C. difficile*. Antimicrobial therapy can predispose to pseudomembranous colitis by altering the normal colonic flora and allowing the multiplication of *C. difficile* (Chap. 129).

AGE

Globally, most morbidity and mortality from enteric pathogens involves children <5 years of age. Breast-fed infants are protected from contaminated food and water and derive some protection from maternal antibodies, but their risk of infection rises dramatically when they begin to eat solid foods. Exposure to rotavirus is universal, with most children experiencing their first infection in the first or second year of life. Older children and adults are more commonly infected with norovirus. Other organisms with higher attack rates among children than among adults include enterotoxigenic, enteropathogenic, and enterohemorrhagic *E. coli*; *Shigella*;*C. jejuni*; and *G. lamblia*.

HOST IMMUNE STATUS

Immunocompromised hosts are at elevated risk of acute and chronic infectious diarrhea. Individuals

with defects in cell-mediated immunity (including those with AIDS) are at particularly high risk of invasive enteropathies, including salmonellosis, listeriosis, and cryptosporidiosis. Individuals with hypogammaglobulinemia are at particular risk of *C. difficile* colitis and giardiasis. Patients with cancer are more likely to develop *C. difficile* infection as a result of chemotherapy and frequent hospitalizations. Infectious diarrhea can be life-threatening in immunocompromised hosts, with complications including bacteremia and metastatic seeding of infection. Furthermore, dehydration may compromise renal function and increase the toxicity of immunosuppressive drugs.

BACTERIAL FOOD POISONING

If the history and the stool examination indicate a noninflammatory etiology of diarrhea and there is evidence of a common-source outbreak, questions concerning the ingestion of specific foods and the time of onset of the diarrhea after a meal can provide clues to the bacterial cause of the illness. Potential causes of bacterial food poisoning are shown in **Table 128–4**.

Table 128–4 Bacterial Food Poisoning

Incubation Period, Organism	Symptoms	Common Food Sources		
1–6 h				
Staphylococcus aureus	Nausea, vomiting, diarrhea	Ham, poultry, potato or egg salad, mayonnaise, cream pastries		
Bacillus cereus			Nausea, vomiting, diarrhea	Fried rice
8–16 h				
Clostridium perfringens	Abdominal cramps, diarrhea (vomiting rare)	Beef, poultry, legumes, gravies		
B. cereus	Abdominal cramps, diarrhea (vomiting rare)	Meats, vegetables, dried beans, cereals		
>16 h				
Vibrio cholerae	Watery diarrhea	Shellfish, water		
Enterotoxigenic *Escherichia coli*	Watery diarrhea	Salads, cheese, meats, water		
Enterohemorrhagic *E. coli*	Bloody diarrhea	Ground beef, roast beef, salami, raw milk, raw vegetables, apple juice		
Salmonella spp.	Inflammatory diarrhea	Beef, poultry, eggs, dairy products		
Campylobacter jejuni	Inflammatory diarrhea	Poultry, raw milk		
Shigella spp.	Dysentery	Potato or egg salad, lettuce, raw vegetables		
Vibrio parahaemolyticus	Dysentery	Mollusks, crustaceans		

Bacterial disease caused by an enterotoxin elaborated outside the host, such as that due to *Staphylococcus aureus* or *B. cereus*, has the shortest incubation period (1–6 h) and generally lasts <12 h. Most cases of staphylococcal food poisoning are caused by contamination from infected human carriers. Staphylococci can multiply at a wide range of temperatures; thus, if food is left to cool slowly and remains at room temperature after cooking, the organisms will have the opportunity to form enterotoxin. Outbreaks following picnics where potato salad, mayonnaise, and cream pastries have been served offer classic examples of staphylococcal food poisoning. Diarrhea, nausea, vomiting, and abdominal cramping are common, while fever is less so.

B. cereus can produce either a syndrome with a short incubation period—the *emetic* form, mediated by a staphylococcal type of enterotoxin—or one with a longer incubation period (8–16 h)—the *diarrheal* form, caused by an enterotoxin resembling *E. coli* LT, in which diarrhea and abdominal cramps are characteristic but vomiting is uncommon. The emetic form of *B. cereus* food poisoning is associated with contaminated fried rice; the organism is common in uncooked rice, and its heat-resistant spores survive boiling. If cooked rice is not refrigerated, the spores can germinate and produce toxin. Frying before serving may not destroy the preformed, heat-stable toxin.

Food poisoning due to *Clostridium perfringens* also has a slightly longer incubation period (8–14 h) and results from the survival of heat-resistant spores in inadequately cooked meat, poultry, or legumes. After ingestion, toxin is produced in the intestinal tract, causing moderately severe abdominal

cramps and diarrhea; vomiting is rare, as is fever. The illness is self-limited, rarely lasting >24 h.

Not all food poisoning has a bacterial cause. Nonbacterial agents of short-incubation food poisoning include capsaicin, which is found in hot peppers, and a variety of toxins found in fish and shellfish (Chap. 396).

LABORATORY EVALUATION

Many cases of noninflammatory diarrhea are self-limited or can be treated empirically, and in these instances the clinician may not need to determine a specific etiology. Potentially pathogenic *E. coli* cannot be distinguished from normal fecal flora by routine culture, and tests to detect enterotoxins are not available in most clinical laboratories. In situations in which cholera is a concern, stool should be cultured on selective media such as thiosulfate–citrate–bile salts–sucrose (TCBS) or tellurite-taurocholate-gelatin (TTG) agar. A latex agglutination test has made the rapid detection of rotavirus in stool practical for many laboratories, while reverse-transcriptase polymerase chain reaction and specific antigen enzyme immunoassays have been developed for the identification of norovirus. Stool specimens should be examined by immunofluorescence-based rapid assays or (less sensitive) standard microscopy for *Giardia* cysts or *Cryptosporidium* if the level of clinical suspicion regarding the involvement of these organisms is high.

All patients with fever and evidence of inflammatory disease acquired outside the hospital should have stool cultured for *Salmonella*, *Shigella*, and *Campylobacter*. *Salmonella* and *Shigella* can be selected on MacConkey agar as non-lactose-fermenting (colorless) colonies or can be grown on *Salmonella-Shigella* agar or in selenite enrichment broth, both of which inhibit most organisms except these pathogens. Evaluation of nosocomial diarrhea should initially focus on *C. difficile*; stool culture for other pathogens in this setting has an extremely low yield and is not cost-effective. Toxins A and B produced by pathogenic strains of *C. difficile* can be detected by rapid enzyme immunoassays and latex agglutination tests (Chap. 129). Isolation of *C. jejuni* requires inoculation of fresh stool onto selective growth medium and incubation at 42°C in a microaerophilic atmosphere. In many laboratories in the United States, *E. coli* O157:H7 is among the most common pathogens isolated from visibly bloody stools. Strains of this enterohemorrhagic serotype can be identified in specialized laboratories by serotyping but also can be identified presumptively in hospital laboratories as lactose-fermenting, indole-positive colonies of sorbitol nonfermenters (white colonies) on sorbitol MacConkey plates. If the clinical presentation suggests the possibility of intestinal amebiasis, stool should be examined by a rapid antigen detection assay or by (less sensitive) microscopy.

Treatment: Infectious Diarrhea or Bacterial Food Poisoning

In many cases, a specific diagnosis is not necessary or not available to guide treatment. The clinician can proceed with the information obtained from the history, stool examination, and evaluation of dehydration severity. Empirical regimens for the treatment of traveler's diarrhea are listed in **Table 128–5**.

Table 128–5 Treatment of Traveler's Diarrhea on the Basis of Clinical Features[a]

Clinical Syndrome	Suggested Therapy
Watery diarrhea (no blood in stool, no fever), 1 or 2 unformed stools per day without distressing enteric symptoms	Oral fluids (oral rehydration solution, Pedialyte, Lytren, or flavored mineral water) and saltine crackers
Watery diarrhea (no blood in stool, no fever), 1 or 2 unformed stools per day with distressing enteric symptoms	Bismuth subsalicylate (for adults): 30 mL or 2 tablets (262 mg/tablet) every 30 min for 8 doses; or loperamide[b]: 4 mg initially followed by 2 mg after passage of each unformed stool, not to exceed 8 tablets (16 mg) per day (prescription dose) or 4 caplets (8 mg) per day (over-the-counter dose); drugs can be taken for 2 days
Watery diarrhea (no blood in stool, no distressing abdominal pain, no fever), >2 unformed stools per day	Antibacterial drug[c] plus (for adults) loperamide[b] (see dose above)
Dysentery (passage of bloody stools) or fever (>37.8°C)	Antibacterial drug[c]
Vomiting, minimal diarrhea	Bismuth subsalicylate (for adults; see dose above)
Diarrhea in infants (<2 years old)	Fluids and electrolytes (oral rehydration solution, Pedialyte, Lytren); continue feeding, especially with breast milk; seek medical attention for

	moderate dehydration, fever lasting >24 h, bloody stools, or diarrhea lasting more than several days

[a]All patients should take oral fluids (Pedialyte, Lytren, or flavored mineral water) plus saltine crackers. If diarrhea becomes moderate or severe, if fever persists, or if bloody stools or dehydration develops, the patient should seek medical attention.

[b]Loperamide should not be used by patients with fever or dysentery; its use may prolong diarrhea in patients with infection due to *Shigella* or other invasive organisms.

[c]The recommended antibacterial drugs are as follows:

Travel to high-risk country other than Thailand: *Adults:* (1) A fluoroquinolone such as ciprofloxacin, 750 mg as a single dose or 500 mg bid for 3 days; levofloxacin, 500 mg as a single dose or 500 mg qd for 3 days; or norfloxacin, 800 mg as a single dose or 400 mg bid for 3 days. (2) Azithromycin, 1000 mg as a single dose or 500 mg qd for 3 days. (3) Rifaximin, 200 mg tid or 400 mg bid for 3 days (not recommended for use in dysentery).

Children: Azithromycin, 10 mg/kg on day 1, 5 mg/kg on days 2 and 3 if diarrhea persists. Alternative agent: furazolidone, 7.5 mg/kg per day in four divided doses for 5 days.

Travel to Thailand (with risk of fluoroquinolone-resistant *Campylobacter*):

Adults: Azithromycin (at above dose for adults). Alternative agent: a fluoroquinolone (at above doses for adults).

Children: Same as for children traveling to other areas (see above).

Source: After Hill et al.

The mainstay of treatment is adequate rehydration. The treatment of cholera and other dehydrating diarrheal diseases was revolutionized by the promotion of oral rehydration solution (ORS), the efficacy of which depends on the fact that glucose-facilitated absorption of sodium and water in the small intestine remains intact in the presence of cholera toxin. The use of ORS has reduced mortality rates for cholera from >50% (in untreated cases) to <1%. A number of ORS formulas have been used. Initial preparations were based on the treatment of patients with cholera and included a solution containing 3.5 g of sodium chloride, 2.5 g of sodium bicarbonate, 1.5 g of potassium chloride, and 20 g of glucose (or 40 g of sucrose) per liter of water. Such a preparation can still be used for the treatment of severe cholera. Many causes of secretory diarrhea, however, are associated with less electrolyte loss than occurs in cholera; beginning in 2002, the World Health Organization recommended a "reduced-osmolarity/reduced-salt" ORS that is better tolerated and more effective than classic ORS. This preparation contains 2.6 g of sodium chloride, 2.9 g of trisodium citrate, 1.5 g of potassium chloride, and 13.5 g of glucose (or 27 g of sucrose) per liter of water. ORS formulations containing rice or cereal as the carbohydrate source may be even more effective than glucose-based solutions. Patients who are severely dehydrated or in whom vomiting precludes the use of oral therapy should receive IV solutions such as Ringer's lactate.

Although most secretory forms of traveler's diarrhea (usually due to enterotoxigenic or enteroaggregative *E. coli* or to *Campylobacter*) can be treated effectively with rehydration, bismuth subsalicylate, or antiperistaltic agents, antimicrobial agents can shorten the duration of illness from 3–4 days to 24–36 h. Changes in diet have not been shown to have an impact on the duration of illness, while the efficacy of probiotics continues to be debated. Most individuals who present with dysentery (bloody diarrhea and fever) should be treated empirically with an antimicrobial agent (e.g., a fluoroquinolone or a macrolide) pending microbiologic analysis of stool. Individuals with shigellosis should be treated with a 3- to 7-day course. Individuals with *Campylobacter* infection often benefit from antimicrobial treatment as well. Because of increasing resistance of *Campylobacter* to fluoroquinolones, especially in parts of Asia, a macrolide antibiotic such as erythromycin or azithromycin may be preferred for this infection.

Treatment of salmonellosis must be tailored to the individual patient. Since administration of antimicrobial agents often prolongs intestinal colonization with *Salmonella*, these drugs are usually reserved for individuals at high risk of complications from disseminated salmonellosis, such as young children, patients with prosthetic devices, elderly patients, and immunocompromised persons. Antimicrobial agents should not be administered to individuals (especially children) in whom enterohemorrhagic *E. coli* infection is suspected. Laboratory studies of enterohemorrhagic *E. coli* strains have demonstrated that a number of antibiotics induce replication of Shiga toxin–producing lambdoid bacteriophages, thereby significantly increasing toxin production by these strains. Clinical studies have supported these laboratory results, and antibiotics may increase by twentyfold the risk of hemolytic-uremic syndrome and renal failure during enterohemorrhagic *E. coli* infection. A clinical clue in the diagnosis of the latter infection is bloody diarrhea with low fever or none at all.

PROPHYLAXIS

Improvements in hygiene to limit fecal-oral spread of enteric pathogens will be necessary if the

prevalence of diarrheal diseases is to be significantly reduced in developing countries. Travelers can reduce their risk of diarrhea by eating only hot, freshly cooked food; by avoiding raw vegetables, salads, and unpeeled fruit; and by drinking only boiled or treated water and avoiding ice. Historically, few travelers to tourist destinations adhere to these dietary restrictions. Bismuth subsalicylate is an inexpensive agent for the prophylaxis of traveler's diarrhea; it is taken at a dosage of 2 tablets (525 mg) four times a day. Treatment appears to be effective and safe for up to 3 weeks, but adverse events such as temporary darkening of the tongue and tinnitus can occur. A meta-analysis suggests that probiotics may lessen the likelihood of traveler's diarrhea by ~15%. Prophylactic antimicrobial agents, although effective, are not generally recommended for the prevention of traveler's diarrhea except when travelers are immunosuppressed or have other underlying illnesses that place them at high risk for morbidity from gastrointestinal infection. The risk of side effects and the possibility of developing an infection with a drug-resistant organism or with more harmful, invasive bacteria make it more reasonable to institute an empirical short course of treatment if symptoms develop. If prophylaxis is indicated, the nonabsorbed antibiotic rifaximin can be considered for use in regions such as Latin America and Africa, where noninvasive *E. coli* predominates as the cause of traveler's diarrhea. Rifaximin is not effective against invasive enteropathogens.

The possibility of exerting a major impact on the worldwide morbidity and mortality associated with diarrheal diseases has led to intense efforts to develop effective vaccines against the common bacterial and viral enteric pathogens. An effective rotavirus vaccine is currently available. Vaccines against *S. typhi* and *V. cholerae* are also available, although the protection they offer is incomplete and/or short lived. At present, there is no effective commercially available vaccine against *Shigella*, enterotoxigenic *E. coli*, *Campylobacter*, nontyphoidal *Salmonella*, norovirus, or intestinal parasites.

ACKNOWLEDGMENTS

The substantial contributions of Joan R. Butterton, MD, an author of this chapter in previous editions, are gratefully acknowledged.

FURTHER READINGS

Boschi-Pinto C et al: Estimating child mortality due to diarrhea in developing countries. Bull WHO 86:710, 2008[PMID: 18797647]

Christopher PR et al: Antibiotic therapy for Shigella dysentry. Cochrane Databse Syst Rev August 4;(8):CD006784, 2010

Cohen SH et al: Clinical pratice guidelines for Clostridium Difficile infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the infectious Diseases Society of America (IDSA). Infect Control Hosp Epidemiol 31:431, 2010

Dupont HL: Clinical pratice. Bacterial diarrhea. N Engl J Med 361:1560, 2009

Hill DR et al: The practice of travel medicine: Guidelines by the Infectious Diseases Society of America. Clin Infect Dis 43:1499, 2006[PMID: 17109284]

——— et al: Management of travellers' diarrhea. BMJ 337:863, 2008

Hogenauer C et al: *Klebsiella oxytoca* as a causative organism of antibiotic-associated hemorrhagic colitis. N Engl J Med 355:2418, 2006[PMID: 17151365]

Sazawal S et al: Efficacy of probiotics in prevention of acute diarrhoea: A meta-analysis of masked, randomised, placebo-controlled trials. Lancet Infect Dis 6:374, 2006[PMID: 16728323]

Shah N et al: Global etiology of traveler's diarrhea: Systematic review from 1973 to the present. Am J Trop Med Hyg 80:609, 2009[PMID: 19346386]

Sodha SV et al: Foodborne disease, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 7th ed, GL Mandell et al (eds). Philadelphia, Churchill Livingstone, 2010, pp 1413–1427

Steiner TS, Guerrant RL: Principles and syndromes of enteric infection, in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 7th ed, GL Mandell et al (eds). Philadelphia, Churchill Livingstone, 2010, pp 1335–1351

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CLOSTRIDIUM DIFFICILE INFECTION, INCLUDING PSEUDOMEMBRANOUS COLITIS: INTRODUCTION

Definition

Clostridium difficile infection (CDI) is a unique colonic disease that is acquired almost exclusively in association with antimicrobial use and the consequent disruption of the normal colonic flora. The most commonly diagnosed diarrheal illness acquired in the hospital, CDI results from the ingestion of spores of *C. difficile* that vegetate, multiply, and secrete toxins, causing diarrhea and pseudomembranous colitis (PMC).

Etiology and Epidemiology

C. difficile is an obligately anaerobic, gram-positive, spore-forming bacillus whose spores are found widely in nature, particularly in the environment of hospitals and chronic-care facilities. CDI occurs most frequently in hospitals and nursing homes where the level of antimicrobial use is high and the environment is contaminated by *C. difficile* spores.

Clindamycin, ampicillin, and cephalosporins were the first antibiotics associated with CDI. The second- and third-generation cephalosporins, particularly cefotaxime, ceftriaxone, cefuroxime, and ceftazidime, are agents frequently responsible for this condition, and the fluoroquinolones (ciprofloxacin, levofloxacin, and moxifloxacin) are the most recent drug class to be implicated in hospital outbreaks. Penicillin/β-lactamase-inhibitor combinations such as ticarcillin/clavulanate and piperacillin/tazobactam pose significantly less risk. However, all antibiotics, including vancomycin and metronidazole (the agents most commonly used to treat CDI), have been found to carry a risk of subsequent CDI. Rare cases are reported in patients without prior antibiotic exposure.

C. difficile is acquired exogenously, most frequently in the hospital or nursing home, and is carried in the stool of symptomatic and asymptomatic patients. The rate of fecal colonization is often ≥20% among adult patients hospitalized for >1 week; in contrast, the rate is 1–3% among community residents. Community-onset CDI without recent hospitalization probably accounts for ≤10% of all cases. The risk of *C. difficile* acquisition increases in proportion to length of hospital stay. Asymptomatic fecal carriage of *C. difficile* in healthy neonates is very common, with rates often exceeding 50% during the first 6 months of life, but associated disease in this population is rare. Spores of *C. difficile* are found on environmental surfaces (where the organism can persist for months) and on the hands of hospital personnel who fail to practice good hand hygiene. Hospital epidemics of CDI have been attributed to a single *C. difficile* strain and to multiple strains present simultaneously. Other identified risk factors for CDI include older age, greater severity of underlying illness, gastrointestinal surgery, use of electronic rectal thermometers, enteral tube feeding, and antacid treatment. Use of proton pump inhibitors may be a risk factor, but this risk is probably modest, and no firm data have implicated these agents in patients who are not already receiving antibiotics.

Pathology and Pathogenesis

Spores of toxigenic *C. difficile* are ingested, survive gastric acidity, germinate in the small bowel, and colonize the lower intestinal tract, where they elaborate two large toxins: toxin A (an enterotoxin) and toxin B (a cytotoxin). These toxins initiate processes resulting in the disruption of epithelial-cell barrier function, diarrhea, and pseudomembrane formation. Toxin A is a potent neutrophil chemoattractant, and both toxins glucosylate the GTP-binding proteins of the Rho subfamily that regulate the actin cell cytoskeleton. Data from studies using molecular disruption of toxin genes in isogenic mutants suggest that toxin B is the essential virulence factor; this possibility, if confirmed, might account for the occurrence of clinical disease caused by toxin A–negative strains. Disruption of the cytoskeleton results in loss of cell shape, adherence, and tight junctions, with consequent fluid leakage. A third toxin, binary toxin CDT, was previously found in only ~6% of strains but is present in all isolates of the newly recognized epidemic strain (see "Global Considerations," below); this toxin is related to *C. perfringens* iota toxin. Its role in the pathogenesis of CDI has not yet been defined.

The pseudomembranes of PMC are confined to the colonic mucosa and initially appear as 1- to 2-mm whitish-yellow plaques. The intervening mucosa appears unremarkable, but, as the disease progresses, the pseudomembranes coalesce to form larger plaques and become confluent over the entire colon wall (Fig. 129-1). The whole colon is usually involved, but 10% of patients have rectal

sparing. Viewed microscopically, the pseudomembranes have a mucosal attachment point and contain necrotic leukocytes, fibrin, mucus, and cellular debris. The epithelium is eroded and necrotic in focal areas, with neutrophil infiltration of the mucosa.

Figure 129-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Autopsy specimen showing confluent pseudomembranes covering the cecum of a patient with pseudomembranous colitis. Note the sparing of the terminal ileum (*arrow*).

Patients colonized with *C. difficile* were initially thought to be at high risk for CDI. However, four prospective studies have shown that colonized patients actually have a decreased risk of subsequent CDI. At least three events are proposed as essential for the development of CDI (Fig. 129-2). Exposure to antimicrobial agents is the first event and establishes susceptibility to *C. difficile* infection. The second event is exposure to toxigenic *C. difficile*. Given that the majority of patients do not develop CDI after the first two events, a third event is clearly essential for its occurrence. Candidate third events include exposure to a *C. difficile* strain of particular virulence, exposure to antimicrobial agents especially likely to cause CDI, and an inadequate host immune response. The host anamnestic serum IgG antibody response to toxin A of *C. difficile* is the most likely third event that determines which patients develop diarrhea and which patients remain asymptomatic. The majority of humans first develop antibody to *C. difficile* toxins when colonized asymptomatically during the first year of life. Infants are thought not to develop symptomatic CDI because they lack suitable mucosal toxin receptors that develop later in life. In adulthood, serum levels of IgG antibody to toxin A increase more in response to infection in individuals who become asymptomatic carriers than in those who develop CDI. For persons who develop CDI, increasing levels of antitoxin A during treatment correlate with a lower risk of recurrence of CDI. A clinical trial using monoclonal antibodies to both toxin A and toxin B in addition to standard therapy showed rates of recurrence lower than those obtained with placebo plus standard therapy.

Figure 129-2

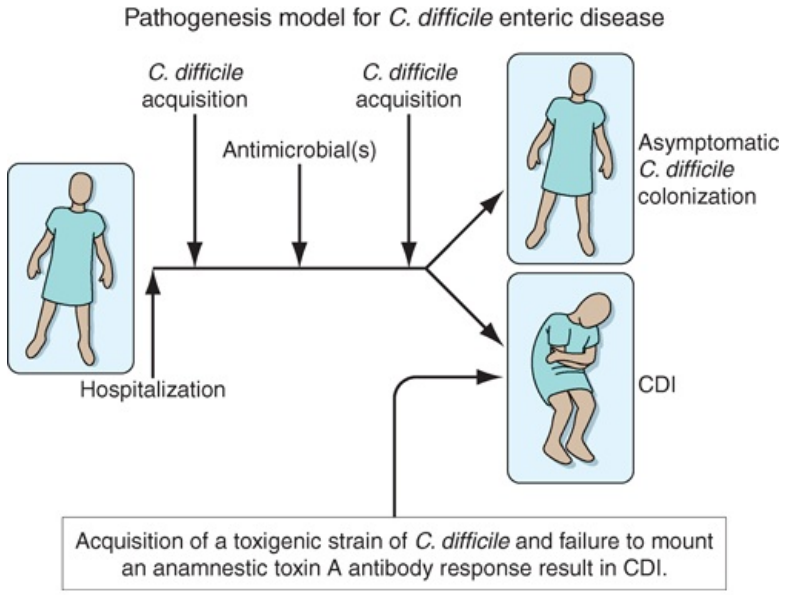

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis model for hospital-acquired *Clostridium difficile* infection (CDI). At least three events are integral to *C. difficile* pathogenesis. Exposure to antibiotics establishes susceptibility to infection. Once susceptible, the patient may acquire nontoxigenic (nonpathogenic) or toxigenic strains of *C. difficile* as a second event. Acquisition of toxigenic *C. difficile* may be followed by asymptomatic colonization or CDI, depending on one or more additional events, including an inadequate host anamnestic IgG response to *C. difficile* toxin A.

Global Considerations

Rates and severity of CDI in the United States, Canada, and Europe have increased markedly since the year 2000. Rates in U.S. hospitals tripled between 2000 and 2005. Hospitals in Montreal, Quebec, have reported rates four times higher than the 1997 baseline, with directly attributable mortality of 6.9% (increased from 1.5%). An epidemic strain, variously known as toxinotype III, REA type BI, PCR ribotype 027, and pulsed-field type NAP1, is thought to account for much of the increase in incidence and has been found in North America, Europe, and Asia. The epidemic organism is characterized by (1) an ability to produce 16–23 times as much toxin A and toxin B as control strains in vitro; (2) the presence of a third toxin (binary toxin CDT); and (3) high-level resistance to all fluoroquinolones. New strains have been and will probably continue to be implicated in outbreaks; their emergence may be explained in part by patterns of antibiotic use, particularly in hospitals.

Clinical Manifestations

Diarrhea is the most common manifestation caused by *C. difficile*. Stools are almost never grossly bloody and range from soft and unformed to watery or mucoid in consistency, with a characteristic odor. Patients may have as many as 20 bowel movements per day. Clinical and laboratory findings include fever in 28% of cases, abdominal pain in 22%, and leukocytosis in 50%. When adynamic ileus (which is seen on x-ray in ~20% of cases) results in cessation of stool passage, the diagnosis of CDI is frequently overlooked. A clue to the presence of unsuspected CDI in these patients is unexplained leukocytosis, with ≥15,000 white blood cells (WBCs)/μL. Such patients are at high risk for complications of *C. difficile* infection, particularly toxic megacolon and sepsis.

C. difficile diarrhea recurs after treatment in ~15–30% of cases, and this figure may be increasing. Recurrences may represent either relapses due to the same strain or reinfections with a new strain. Susceptibility to recurrence of clinical CDI is likely a result of continued disruption of the normal fecal flora caused by the antibiotic used to treat CDI.

Diagnosis

The diagnosis of CDI is based on a combination of clinical criteria: (1) diarrhea (≥3 unformed stools per 24 h for ≥2 days) with no other recognized cause plus (2) toxin A or B detected in the stool, toxin-producing *C. difficile* detected in the stool by polymerase chain reaction (PCR) or culture, or pseudomembranes seen in the colon. PMC is a more advanced form of CDI and is visualized at endoscopy in only ~50% of patients with diarrhea who have a positive stool culture and toxin assay for *C. difficile*(**Table 129–1**). Endoscopy is a rapid diagnostic tool in seriously ill patients with suspected PMC and an acute abdomen, but a negative result in this examination does not rule out CDI.

Table 129–1 Relative Sensitivity and Specificity of Diagnostic Tests for *Clostridium Difficile* Infection (CDI)

Type of Test	Relative Sensitivity[a]	Relative Specificity[a]	Comment
Stool culture for *C. difficile*	+ + + +	+ + +	Most sensitive test; specificity of + + + + if the *C. difficile* isolate tests positive for toxin; with clinical data, is diagnostic of CDI; turnaround time too slow for practical use
Cell culture cytotoxin test on stool	+ + +	+ + + +	With clinical data, is diagnostic of CDI; highly specific but not as sensitive as stool culture; slow turnaround time
Enzyme immunoassay for toxin A or toxins A and B in stool	+ + to + + +	+ + +	With clinical data, is diagnostic of CDI; rapid results, but not as sensitive as stool culture or cell culture cytotoxin test
Enzyme immunoassay for *C. difficile* common antigen in stool	+ + + to + + + +	+ + +	Detects glutamate dehydrogenase found in toxigenic and nontoxigenic strains of *C. difficile* and other stool organisms; more sensitive and less specific than enzyme immunoassay for toxins; rapid results
PCR for *C. difficile* toxin B gene in stool	+ + + +	+ + + +	Detects toxigenic *C. difficile* in stool; newly approved for clinical testing, but appears to be more sensitive than enzyme immunoassay toxin testing and at least as specific
Colonoscopy or sigmoidoscopy	+	+ + + +	Highly specific if pseudomembranes are seen; insensitive compared with other tests

[a]According to both clinical and test-based criteria.

Note: + + + +, >90%; + + +, 71–90%; + +, 51–70%; +, ~50%.

Despite the array of tests available for *C. difficile* and its toxins (Table 129–1), no single test has high sensitivity, high specificity, and rapid turnaround. Most laboratory tests for toxins, including enzyme immunoassays (EIAs), lack sensitivity. However, testing of multiple additional stool specimens is not recommended. PCR assays have now been approved for diagnostic testing and appear to be both rapid and sensitive while retaining high specificity. Empirical treatment is appropriate if CDI is strongly suspected on clinical grounds. Testing of asymptomatic patients is not recommended except for epidemiologic study purposes. In particular, so-called tests of cure following treatment are not recommended because many patients continue to harbor the organism and toxin after diarrhea has ceased and test results do not always predict recurrence of CDI. Thus these results should not be used to restrict placement of patients in long-term-care or nursing home facilities.

Treatment: *Clostridium Difficile* Infection

PRIMARY CDI

When possible, discontinuation of any ongoing antimicrobial administration is recommended as the first step in treatment of CDI. Earlier studies indicated that 15–23% of patients respond to this simple measure. However, with the advent of the current epidemic strain and the associated rapid clinical deterioration of some patients, prompt initiation of specific CDI treatment has become the standard. General treatment guidelines include hydration and the avoidance of antiperistaltic agents and opiates, which may mask symptoms and possibly worsen disease. Nevertheless, antiperistaltic agents have been used safely with vancomycin or metronidazole for mild to moderate CDI.

All drugs, particularly vancomycin, should be given orally if possible. When IV metronidazole is administered, fecal bactericidal drug concentrations are achieved during acute diarrhea, and CDI treatment has been successful; however, in the presence of adynamic ileus, IV metronidazole treatment of PMC has failed. In previous randomized trials, diarrhea response rates to oral therapy with vancomycin or metronidazole were ≥94%, but four recent observational studies found that response rates for metronidazole had declined to 62–78%. Although the mean time to resolution of diarrhea is 2–4 days, the response to metronidazole may be much slower. Treatment should not be deemed a failure until a drug has been given for at least 6 days. On the basis of data for shorter courses of vancomycin, it is recommended that metronidazole and vancomycin be given for at least 10 days, although no controlled comparisons are available. Metronidazole is not approved for this indication by the U.S. Food and Drug Administration (FDA), but most patients with mild to moderate illness respond to 500 mg given by mouth three times a day for 10 days; extension of the treatment period may be needed for slow responders. In addition to the reports of increases in metronidazole failures, a prospective, randomized, double-blind, placebo-controlled study has demonstrated the superiority of vancomycin over metronidazole for treatment of severe CDI. The severity assessment score in that study included age as well as laboratory parameters (elevated temperature, low albumin level, or elevated WBC count), documentation of PMC by endoscopy, or treatment of CDI in the

intensive care unit. Although a validated severity score is not yet available, it is important to initiate treatment with oral vancomycin for patients who appear seriously ill, particularly if they have a high WBC count (>15,000/μL) or a creatinine level that is ≥1.5 times higher than the premorbid value (Table 129–2). Small randomized trials of nitazoxanide, bacitracin, rifaximin, and fusidic acid for treatment of CDI have been conducted. While these drugs have not yet been extensively studied, shown to be superior, or approved by the FDA for this indication, they provide potential alternatives to vancomycin and metronidazole.

Table 129–2 Recommendations for the Treatment of *Clostridium Difficile* Infection (CDI)

Clinical Setting	Treatment(s)	Comments
Initial episode, mild to moderate	Oral metronidazole (500 mg tid x 10–14 d)	
Initial episode, severe	Oral vancomycin (125 mg qid x 10–14 d)	Indicators of severe disease may include leukocytosis (≥15,000 white blood cells/μL) and a creatinine level ≥ 1.5 times the premorbid value.
Initial episode, severe complicated or fulminant	Vancomycin (500 mg PO or via nasogastric tube) plus metronidazole (500 mg IV q8h) *plus consider* Rectal instillation of vancomycin (500 mg in 100 mL of normal saline as a retention enema q6–8h)	Severe complicated or fulminant CDI is defined as severe CDI with the addition of hypotension, shock, ileus, or toxic megacolon. The duration of treatment may need to be >2 weeks and is dictated by response. Consider using IV tigecycline (50 mg q12h after a 100-mg loading dose) in place of metronidazole.
First recurrence	Same as for initial episode	
Second recurrence	Vancomycin in tapered/pulsed regimen	Typical taper/pulse regimen: 125 mg qid x 10–14 d, then bid x 1 week, then daily x 1 week, then q2–3d for 2–8 weeks
Multiple recurrences	Consider the following options: • Repeat vancomycin taper/pulse • Vancomycin (500 mg qid x 10 d) plus *Saccharomyces boulardii* (500 mg bid x 28 d) • Vancomycin (125 mg qid x 10–14 d); then stop vancomycin and start rifaximin (400 mg bid x 2 weeks) • Nitazoxanide (500 mg bid x 10 d) • Fecal transplantation • IV immunoglobulin (400 mg/kg)	The only controlled study of treatment for recurrent CDI used *S. boulardii* and showed borderline significance compared with placebo.

RECURRENT CDI

Overall, ~15–30% of patients experience recurrences of CDI, either as relapses caused by the original organism or as reinfections following treatment. Recurrence rates are higher among patients ≥65 years old, those who continue to take antibiotics while being treated for CDI, and those who remain in the hospital after the initial episode of CDI. Patients who have a first recurrence of CDI have a high rate of second recurrence (33–65%). In the first recurrence, re-treatment with metronidazole is comparable to

treatment with vancomycin (Table 129–2). Recurrent disease, once thought to be relatively mild, has now been documented to pose a significant (11%) risk of serious complications (shock, megacolon, perforation, colectomy, or death within 30 days). There is no standard treatment for multiple recurrences, but long or repeated metronidazole courses should be avoided because of potential neurotoxicity. Approaches include the administration of vancomycin followed by the yeast *Saccharomyces boulardii*; the administration of vancomycin followed by a synthetic fecal bacterial enema; and the intentional colonization of the patient with a nontoxigenic strain of *C. difficile*. None of these biotherapeutic approaches has been approved by the FDA for use in the United States. Other strategies include (1) the use of vancomycin in tapering doses or with pulse dosing every other day for 2–8 weeks and (2) sequential treatment with vancomycin (125 mg four times daily for 10–14 days) followed by rifaximin (400 mg twice daily for 14 days). IV immunoglobulin, which has also been used with some success, presumably provides antibodies to *C. difficile* toxins.

SEVERE COMPLICATED OR FULMINANT CDI

Fulminant (rapidly progressive and severe) CDI presents the most difficult treatment challenge. Patients with fulminant disease often do not have diarrhea, and their illness mimics an acute surgical abdomen. Sepsis (hypotension, fever, tachycardia, leukocytosis) may result from severe CDI. An acute abdomen (with or without toxic megacolon) may include signs of obstruction, ileus, colon-wall thickening, and ascites on abdominal CT, often with peripheral-blood leukocytosis (≥20,000 WBCs/μL). With or without diarrhea, the differential diagnosis of an acute abdomen, sepsis, or toxic megacolon should include CDI if the patient has received antibiotics in the past 2 months. Cautious sigmoidoscopy or colonoscopy to visualize PMC and an abdominal CT examination are the best diagnostic tests in patients without diarrhea.

Medical management of fulminant CDI is suboptimal because of the difficulty of delivering metronidazole or vancomycin to the colon by the oral route in the presence of ileus (Table 129–2). The combination of vancomycin (given via nasogastric tube and by retention enema) plus IV metronidazole has been used with some success in uncontrolled studies, as has IV tigecycline in small-scale uncontrolled studies. Surgical colectomy may be life-saving if there is no response to medical management. If possible, colectomy should be performed before the serum lactate level reaches 5 mmol/L. The incidence of fulminant CDI requiring colectomy appears to be increasing in the evolving epidemic.

Prognosis

The mortality rate attributed to CDI, previously found to be 0.6–3.5%, has reached 6.9% in recent outbreaks and is progressively higher with increasing age. Most patients recover, but recurrences are common.

Prevention and Control

Strategies for the prevention of CDI are of two types: those aimed at preventing transmission of the organism to the patient and those aimed at reducing the risk of CDI if the organism is transmitted. Transmission of *C. difficile* in clinical practice has been prevented by gloving of personnel, elimination of the use of contaminated electronic thermometers, and use of hypochlorite (bleach) solution for environmental decontamination of patients' rooms. Hand hygiene is critical; hand washing is recommended in CDI outbreaks because alcohol hand gels are not sporicidal. CDI outbreaks have been best controlled by restricting the use of specific antibiotics, such as clindamycin and second- and third-generation cephalosporins. Outbreaks of CDI due to clindamycin-resistant strains have resolved promptly when clindamycin use is restricted.

FURTHER READINGS

Cohen SH et al: Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the Society of Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). Infect Control Hosp Epidemiol 31:431, 2010[PMID: 20307191]

Herpers BL et al: Intravenous tigecycline as adjunctive or alternative therapy for severe refractory *Clostridium difficile* infection. Clin Infect Dis 48:1732, 2009

Johnson S et al: Interruption of recurrent *Clostridium difficile*–associated diarrhea episodes by serial therapy with vancomycin and rifaximin. Clin Infect Dis 44:846, 2007

Kyne L et al: Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. N Engl J Med 342:390, 2000[PMID: 10666429]

Loo VG et al: A predominantly clonal multi-institutional outbreak of *Clostridium difficile*–associated diarrhea with high morbidity and mortality. N Engl J Med 353:2442, 2005

Lowy I et al: Treatment with monoclonal antibodies against *Clostridium difficile* toxins. N Engl J Med

362:197, 2010

Lyras D et al: Toxin B is essential for virulence of *Clostridium difficile.* Nature 458:1176, 2009

McDonald LC et al: *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996-2003. Emerg Infect Dis 12:409, 2006

——— et al: An epidemic, toxin gene–variant strain of *Clostridium difficile*. N Engl J Med 353:2433, 2005

Pepin J et al: The management and outcomes of a first recurrence of *Clostridium difficile* associated disease in Quebec. Clin Infect Dis 42:758, 2006

Zar FA et al: A comparison of vancomycin and metronidazole for the treatment of *Clostridium difficile*–associated diarrhea, stratified by disease severity. Clin Infect Dis 45:302, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CLASSIFICATION AND EPIDEMIOLOGY

Worldwide, most adults acquire at least one sexually transmitted infection (STI), and many remain at risk for complications. Each year, for example, an estimated 6.2 million persons in the United States acquire a new genital human papillomavirus (HPV) infection, and many of these individuals are at risk for genital neoplasias. Certain STIs, such as syphilis, gonorrhea, HIV infection, hepatitis B, and chancroid, are most concentrated within "core populations" characterized by high rates of partner change, multiple concurrent partners, or "dense," highly connected sexual networks—e.g., involving sex workers and their clients, some men who have sex with men (MSM), and persons involved in the use of illicit drugs, particularly crack cocaine and methamphetamine. Other STIs are distributed more evenly throughout societies. For example, chlamydial infections, genital infections with HPV, and genital herpes can spread widely, even in relatively low-risk populations.

In general, the product of three factors determines the initial rate of spread of any STI within a population: rate of sexual exposure of susceptible to infectious people, efficiency of transmission per exposure, and duration of infectivity of those infected. Accordingly, efforts to prevent and control STIs aim to decrease the rate of sexual exposure of susceptibles to infected persons (e.g., through individual counseling and efforts to change the norms of sexual behavior and through a variety of STI control efforts aimed at reducing the proportion of the population infected), to decrease the duration of infectivity (through early diagnosis and curative or suppressive treatment), and to decrease the efficiency of transmission (e.g., through promotion of condom use and safer sexual practices, through use of effective vaccines, and recently through male circumcision).

In all societies, STIs rank among the most common of all infectious diseases, with >30 infections now classified as predominantly sexually transmitted or as frequently sexually transmissible (Table 130–1). In developing countries, with three-quarters of the world's population and 90% of the world's STIs, factors such as population growth (especially in adolescent and young-adult age groups), rural-to-urban migration, wars, limited or no provision of reproductive health services for women, and poverty create exceptional vulnerability to disease resulting from unprotected sex. During the 1990s in China, Russia, the other states of the former Soviet Union, and South Africa, internal social structures changed rapidly as borders opened to the West, unleashing enormous new epidemics of HIV infection and other STIs. Despite advances in the provision of highly effective antiretroviral therapy worldwide, HIV remains the leading cause of death in some developing countries, and HPV and hepatitis B virus (HBV) remain important causes of cervical and hepatocellular carcinoma, respectively—two of the most common malignancies in the developing world. Sexually transmitted herpes simplex virus (HSV) infections now cause most genital ulcer disease throughout the world and an increasing proportion of cases of genital herpes in developing countries with generalized HIV epidemics, where the positive-feedback loop between HSV and HIV transmission is a growing, intractable problem. Despite this consistent link, randomized trials evaluating the efficacy of antiviral therapy in suppressing HSV in both HIV-uninfected and HIV-infected persons have not demonstrated a protective effect against acquisition or transmission of HIV. Globally, ~350 million new cases of five curable STIs—gonorrhea, chlamydial infection, syphilis, chancroid, and trichomoniasis—were reported annually in the mid-1990s. Up to 50% of women of reproductive age in developing countries have bacterial vaginosis (arguably acquired sexually). All five of these curable infections have been associated with increased risk of HIV transmission or acquisition.

Table 130–1 Sexually Transmitted and Sexually Transmissible Microorganisms

Bacteria	Viruses	Other[a]
Transmitted in Adults Predominantly by Sexual Intercourse		
Neisseria gonorrhoeae	HIV (types 1 and 2)	*Trichomonas vaginalis*
Chlamydia trachomatis	Human T-cell lymphotropic virus type I	*Phthirus pubis*
Treponema pallidum	Herpes simplex virus type 2	
Haemophilus ducreyi	Human papillomavirus (multiple genital genotypes)	
Klebsiella (Calymmatobacterium) granulomatis	Hepatitis B virus[b]	
Ureaplasma urealyticum		

Mycoplasma genitalium	Molluscum contagiosum virus	
Sexual Transmission Repeatedly Described but Not Well Defined or Not the Predominant Mode		
Mycoplasma hominis *Gardnerella vaginalis* and other vaginal bacteria Group B *Streptococcus* *Mobiluncus* spp. *Helicobacter cinaedi* *Helicobacter fennelliae*	Cytomegalovirus Human T-cell lymphotropic virus type II Hepatitis C virus (?) Hepatitis D virus Herpes simplex virus type 1 (?) Epstein-Barr virus Human herpesvirus type 8	*Candida albicans* *Sarcoptes scabiei*
Transmitted by Sexual Contact Involving Oral-Fecal Exposure; of Declining Importance in Men Who Have Sex with Men		
Shigella spp. *Campylobacter* spp.	Hepatitis A virus	*Giardia lamblia Entamoeba histolytica*

[a]Includes protozoa, ectoparasites, and fungi.

[b]Among U.S. patients for whom a risk factor can be ascertained, most hepatitis B virus infections are transmitted sexually.

In the United States, the prevalence of antibody to HSV-2 began to fall in the late 1990s, especially among adolescents and young adults; the decline is presumably due to delayed sexual debut, increased condom use, and lower rates of multiple (≥4) sex partners, as is well documented by the U.S. Youth Risk Behavior Surveillance System. The estimated annual incidence of HBV infection has also declined dramatically since the mid-1980s; this decrease is probably attributable more to adoption of safer sexual practices and reduced needle sharing among injection drug users than to use of hepatitis B vaccine, for which coverage among young adults (including those at high risk for this infection) initially was very limited. Genital HPV remains the most common sexually transmitted pathogen in this country, infecting 60% of a cohort of initially HPV-negative, sexually active Washington state college women within 5 years in a study conducted from 1990 to 2000. The scale-up of HPV vaccine coverage among young women promises to lower the incidence of infection with the HPV types included in the vaccines.

In industrialized countries, fear of HIV infection since the mid-1980s, coupled with widespread behavioral interventions and better-organized systems of care for the curable STIs, initially helped curb the transmission of the latter diseases. However, foci of hyperendemic transmission persist in the southeastern United States and in most large U.S. cities. Rates of gonorrhea and syphilis remain higher in the United States than in any other Western industrialized country.

In the United States, the Centers for Disease Control and Prevention (CDC) has compiled reported rates of STIs since 1941. The incidence of reported gonorrhea peaked at 468 cases per 100,000 population in the mid-1970s, fell to a low of 112 cases per 100,000 in 2004, and remained relatively unchanged through 2008. Because of increased testing and more sensitive tests, the incidence of reported *Chlamydia trachomatis* infection has been increasing steadily since reporting began in 1984, reaching an all-time peak of 401 cases per 100,000 in 2008. The incidence of primary and secondary syphilis per 100,000 peaked at 71 cases in 1946, fell rapidly to 3.9 cases in 1956, ranged from ~10 to 15 cases through 1987 (with markedly increased rates among MSM men and African Americans), and then fell to a nadir of 2.1 cases in 2000–2001 (with rates falling most rapidly among heterosexual African Americans). Unfortunately, since 1996, with the introduction of highly active antiretroviral therapy, the increased use of "serosorting" (i.e., the avoidance of unprotected sex with HIV-serodiscordant partners but not with HIV-seroconcordant partners, a strategy that provides no protection against STIs other than HIV infection), and an ongoing epidemic of methamphetamine use, gonorrhea, syphilis, and chlamydial infection have had a remarkable resurgence among MSM in North America and Europe, where outbreaks of a rare type of chlamydial infection (lymphogranuloma venereum; LGV) that had virtually disappeared during the AIDS era have occurred. These developments have resulted in a high degree of co-infection with HIV and other sexually transmitted pathogens (particularly syphilis and LGV), primarily among MSM.

MANAGEMENT OF COMMON SEXUALLY TRANSMITTED DISEASE (STD) SYNDROMES

Although other chapters discuss management of specific STIs, delineating treatment based on diagnosis of a specific infection, most patients are actually managed (at least initially) on the basis of presenting symptoms and signs and associated risk factors, even in industrialized countries. Table 130

–2 lists some of the most common clinical STD syndromes and their microbial etiologies. Strategies for their management are outlined below. **Chapters 188 and 189 address the management of infections with human retroviruses.**

Table 130–2 Major STD Syndromes and Sexually Transmitted Microbial Etiologies

Syndrome	ST Microbial Etiologies
AIDS	HIV types 1 and 2
Urethritis: males	*Neisseria gonorrhoeae, Chlamydia trachomatis, Mycoplasma genitalium, Ureaplasma urealyticum*(?subspecies *urealyticum), Trichomonas vaginalis,* HSV
Epididymitis	*C. trachomatis, N. gonorrhoeae*
Lower genital tract infections: females	
Cystitis/urethritis	*C. trachomatis, N. gonorrhoeae,*HSV
Mucopurulent cervicitis	*C. trachomatis, N. gonorrhoeae, M. genitalium*
Vulvitis	*Candida albicans,*HSV
Vulvovaginitis	*C. albicans, T. vaginalis*
Bacterial vaginosis (BV)	BV-associated bacteria (see text)
Acute pelvic inflammatory disease	*N. gonorrhoeae, C. trachomatis,* BV-associated bacteria, *M. genitalium,* group B streptococci
Infertility	*N. gonorrhoeae, C. trachomatis,* BV-associated bacteria
Ulcerative lesions of the genitalia	HSV-1, HSV-2, *Treponema pallidum, Haemophilus ducreyi, C. trachomatis*(LGV strains), *Klebsiella (Calymmatobacterium) granulomatis*
Complications of pregnancy/puerperium	Several agents implicated
Intestinal infections	
Proctitis	*C. trachomatis, N. gonorrhoeae,*HSV, *T. pallidum*
Proctocolitis or enterocolitis	*Campylobacter* spp., *Shigella* spp., *Entamoeba histolytica,* other enteric pathogens
Enteritis	*Giardia lamblia*
Acute arthritis with urogenital infection or viremia	*N. gonorrhoeae* (e.g., DGI), C. *trachomatis* (e.g., reactive arthritis), HBV
Genital and anal warts	HPV (30 genital types)
Mononucleosis syndrome	CMV, HIV, EBV
Hepatitis	Hepatitis viruses, *T. pallidum*, CMV, EBV
Neoplasias	
Squamous cell dysplasias and cancers of the cervix, anus, vulva, vagina, or penis	HPV (especially types 16, 18, 31, 45)
Kaposi's sarcoma, body-cavity lymphomas	HHV-8
T cell leukemia	HTLV-I
Hepatocellular carcinoma	HBV
Tropical spastic paraparesis	HTLV-I
Scabies	*Sarcoptes scabiei*
Pubic lice	*Pthirus pubis*

Abbreviations: CMV, cytomegalovirus; DGI, disseminated gonococcal infection; EBV, Epstein-Barr virus; HBV, hepatitis B virus; HHV-8, human herpesvirus type 8; HPV, human papillomavirus; HSV, herpes simplex virus; HTLV, human T cell lymphotropic virus; LGV, lymphogranuloma venereum.

STD care and management begin with risk assessment and proceed to clinical assessment, diagnostic testing or screening, treatment, and prevention. Indeed, the routine care of any patient begins with risk assessment (e.g., for risk of heart disease, cancer). STD/HIV risk assessment is important in primary care, urgent care, and emergency care settings as well as in specialty clinics providing adolescent, HIV/AIDS, prenatal, and family planning services. STD/HIV risk assessment guides detection and interpretation of symptoms that could reflect an STD; decisions on screening or prophylactic/preventive treatment; risk reduction counseling and intervention (e.g., hepatitis B

vaccination); and treatment of partners of patients with known infections. Consideration of routine demographic data (e.g., gender, age, area of residence) is a simple first step in STD/HIV risk assessment. For example, national guidelines strongly recommend routine screening of sexually active females ≤25 years of age for *C. trachomatis* infection. Table 130–3 provides a set of 10 STD/HIV risk-assessment questions that clinicians can pose verbally or that health care systems can adapt (with yes/no responses) into a routine self-administered questionnaire for use in clinics. The initial framing statement gives permission to discuss topics that may be perceived as sensitive or socially unacceptable by providers and patients alike.

Table 130–3 Ten-Question STD/HIV Risk Assessment

Framing Statement:
In order to provide the best care for you today and to understand your risk for certain infections, it is necessary for us to talk about your sexual behavior.
Screening Questions:
(1) Do you have any reason to think you might have a sexually transmitted infection? If so, what reason?
(2) For all adolescents <18 years old: Have you begun having any kind of sex yet?
STD History:
(3) Have you ever had any sexually transmitted infections or any genital infections? If so, which ones?
Sexual Preference:
(4) Have you had sex with men, women, or both?
Injection Drug Use:
(5) Have you ever injected yourself ("shot up") with drugs? (If yes, have you ever shared needles or injection equipment?)
(6) Have you ever had sex with a gay or bisexual man or with anyone who had ever injected drugs?
Characteristics of Partner(s):
(7) Has your sex partner(s) had any sexually transmitted infections? If so, which ones?
(8) Has your sex partner had other sex partners during the time you've been together?
STD Symptoms Checklist:
(9) Have you recently developed any of these symptoms?

For Men	For Women
(a) Discharge of pus (drip) from the penis	(a) Abnormal vaginal discharge (increased amount, abnormal odor, abnormal yellow color)
(b) Genital sores (ulcers) or rash	(b) Genital sores (ulcers), rash, or itching

Sexual Practices, Past 2 Months (for patients answering yes to any of the above questions, to guide examination and testing):
(10) Now I'd like to ask what parts of your body may have been sexually exposed to an STD (e.g., your penis, mouth, vagina, anus).
Query about Interest in STD Screening Tests (for patients answering no to all of the above questions):
(11) Would you like to be tested for HIV or any other STDs today? (If yes, clinician can explore which STD and why.)

Source: Adapted from JR Curtis, KK Holmes, in KK Holmes et al (eds): *Sexually Transmitted Diseases*, 4th ed. New York, McGraw-Hill, 2008.

Risk assessment is followed by clinical assessment (elicitation of information on specific current symptoms and signs of STDs). Confirmatory diagnostic tests (for persons with symptoms or signs) or screening tests (for those without symptoms or signs) may involve microscopic examination, culture, antigen detection tests, nucleic acid amplification tests (NAATs), or serology. Initial syndrome-based treatment should cover the most likely causes. For certain syndromes, results of rapid tests can narrow the spectrum of this initial therapy (e.g., saline microscopy of vaginal fluid for women with vaginal discharge, Gram's stain of urethral discharge for men with urethral discharge, rapid plasma reagin test for genital ulcer). After the institution of treatment, STD management proceeds to the "4 Cs" of prevention and control: *c*ontact tracing (see "Prevention and Control of STIs," below), ensuring *c*ompliance with therapy, and *c*ounseling on risk reduction, including *c*ondom promotion and provision.

Consistent with current guidelines, all adults should be screened for infection with HIV-1 at least once and more frequently if they are at risk for acquisition of this infection.

Urethritis in Men

Urethritis in men produces urethral discharge, dysuria, or both, usually without frequency of urination. Causes include *Neisseria gonorrhoeae*, *C. trachomatis*, *Mycoplasma genitalium*, *Ureaplasma urealyticum*, *Trichomonas vaginalis*, HSV, and adenovirus.

Until recently, *C. trachomatis* caused ~30–40% of cases of nongonococcal urethritis (NGU), particularly in heterosexual men; however, the proportion of cases due to this organism has probably declined in some populations served by effective chlamydial-control programs, and older men with urethritis appear less likely to have chlamydial infection. HSV and *T. vaginalis* each cause a small proportion of NGU cases in the United States. Recently, multiple studies have consistently implicated *M. genitalium* as a probable cause of many *Chlamydia*-negative cases. Fewer studies than in the past have implicated *Ureaplasma*; the ureaplasmas have been differentiated into *U. urealyticum* and *U. parvum*, and a few studies suggest that *U. urealyticum*—but not *U. parvum*—is associated with NGU. Coliform bacteria can cause urethritis in men who practice insertive anal intercourse. The initial diagnosis of urethritis in men currently includes specific tests only for *N. gonorrhoeae* and *C. trachomatis*; it does not yet include testing for *Mycoplasma* or *Urealyticum* species. The following summarizes the approach to the patient with suspected urethritis:

1. *Establish the presence of urethritis* . If proximal-to-distal "milking" of the urethra does not express a purulent or mucopurulent discharge, even after the patient has not voided for several hours (or preferably overnight), a Gram's-stained smear of overt discharge or of an anterior urethral specimen obtained by passage of a small urethrogenital swab 2–3 cm into the urethra usually reveals ≥5 neutrophils per 1000x field in areas containing cells; in gonococcal infection, such a smear usually reveals gram-negative intracellular diplococci as well. Alternatively, the centrifuged sediment of the first 20–30 mL of voided urine—ideally collected as the first morning specimen —can be examined for inflammatory cells, either by microscopy showing ≥10 leukocytes per high-power field or by the leukocyte esterase test. Patients with symptoms who lack objective evidence of urethritis may have functional rather than organic problems and generally do not benefit from repeated courses of antibiotics.
2. *Evaluate for complications or alternative diagnoses* . A brief history and examination will exclude epididymitis and systemic complications, such as disseminated gonococcal infection (DGI) and reactive arthritis. Although digital examination of the prostate gland seldom contributes to the evaluation of sexually active young men with urethritis, men with dysuria who lack evidence of urethritis as well as sexually inactive men with urethritis should undergo prostate palpation, urinalysis, and urine culture to exclude bacterial prostatitis and cystitis.
3. *Evaluate for gonococcal and chlamydial infection* . An absence of typical gram-negative diplococci on Gram's-stained smear of urethral exudate containing inflammatory cells warrants a preliminary diagnosis of NGU, as this test is 98% sensitive for the diagnosis of gonococcal urethral infection. However, an increasing proportion of men with symptoms and/or signs of urethritis are simultaneously assessed for infection with *N. gonorrhoeae* and *C. trachomatis* by "multiplex" NAATs of first-voided urine. The urine specimen tested should comprise the first 10–15 mL of the stream, and patients should not have voided for the prior 2 h, if possible. Culture or NAAT for *N. gonorrhoeae* may be positive when Gram's staining is negative; certain strains of *N. gonorrhoeae* can result in negative urethral Gram's stains in up to 30% of cases of urethritis. Results of tests for gonococcal and chlamydial infection predict the patient's prognosis (with greater risk for recurrent NGU if neither chlamydiae nor gonococci are found than if either is detected) and can guide both the counseling given to the patient and the management of the patient's sexual partner(s).
4. *Treat urethritis promptly while test results are pending* .

TREATMENT: URETHRITIS IN MEN

Table 130–4 summarizes the steps in management of sexually active men with urethral discharge and/or dysuria.

Table 130–4 Management of Urethral Discharge in Men

Usual Causes	Usual Initial Evaluation	
Chlamydia trachomatis *Neisseria gonorrhoeae* *Mycoplasma genitalium* *Ureaplasma urealyticum Trichomonas vaginalis* Herpes simplex virus		Demonstration of urethral discharge or pyuria Exclusion of local or systemic complications Urethral Gram's stain to confirm urethritis, detect gram-negative diplococci Test for *N. gonorrhoeae, C. trachomatis*

Initial Treatment for Patient and Partners		
Treat gonorrhea (unless excluded): Ceftriaxone, 250 mg IM; *or* Cefpodoxime, 400 mg PO; *or* Cefixime, 400 mg PO[a]	plus	**Treat chlamydial infection:** Azithromycin, 1 g PO; *or* Doxycycline, 100 mg bid PO for 7 days
Management of Recurrence		
Confirm objective evidence of urethritis. If patient was reexposed to untreated or new partner, repeat treatment of patient and partner.		
If patient was not reexposed, consider infection with *T. vaginalis*[b] or doxycycline-resistant *M. genitalium* or *Ureaplasma*, and consider treatment with metronidazole, azithromycin, or both.		

[a]Updates on the emergence of antimicrobial resistance in *N. gonorrhoeae* can be obtained from the Centers for Disease Control and Prevention at *http://www.cdc.gov/std*.

[b]In men, the diagnosis of *T. vaginalis* infection requires culture (or nucleic acid amplification test, where available) of early-morning first-voided urine sediment or of a urethral swab specimen obtained before voiding.

In practice, if Gram's stain does not reveal gonococci, urethritis is treated with a regimen effective for NGU, such as azithromycin or doxycycline. Both are effective, although azithromycin may give better results in *M. genitalium* infection. If gonococci are demonstrated by Gram's stain or if no diagnostic tests are performed to exclude gonorrhea definitively, treatment should include a single-dose regimen for gonorrhea (Chap. 144) plus azithromycin or doxycycline treatment for *C. trachomatis*, which frequently causes urethral co-infection in men with gonococcal urethritis. Sexual partners should ideally be tested for gonorrhea and chlamydial infection; regardless of whether they are tested for these infections, however, they should receive the same regimen given to the male index case. Patients with confirmed persistence or recurrence of urethritis after treatment should be re-treated with the initial regimen if they did not comply with the original treatment or were reexposed to an untreated partner. Otherwise, an intraurethral swab specimen and a first-voided urine sample should be tested for *T. vaginalis* (currently best done by culture, although NAATs appear to be more sensitive and are likely to become commercially available in the future). If compliance with initial treatment is confirmed and reexposure to an untreated sex partner is deemed unlikely, the recommended treatment is with metronidazole or tinidazole (2 g by mouth in a single dose) plus azithromycin (1 g by mouth in a single dose); the azithromycin component is especially important if this drug has not been given during initial therapy.

Epididymitis

Acute epididymitis, almost always unilateral, produces pain, swelling, and tenderness of the epididymis, with or without symptoms or signs of urethritis. This condition must be differentiated from testicular torsion, tumor, and trauma. Torsion, a surgical emergency, usually occurs in the second or third decade of life and produces a sudden onset of pain, elevation of the testicle within the scrotal sac, rotation of the epididymis from a posterior to an anterior position, and absence of blood flow on Doppler examination or ^{99m}Tc scan. Persistence of symptoms after a course of therapy for epididymitis suggests the possibility of testicular tumor or of a chronic granulomatous disease, such as tuberculosis. In sexually active men under age 35, acute epididymitis is caused most frequently by *C. trachomatis* and less commonly by *N. gonorrhoeae* and is usually associated with overt or subclinical urethritis. Acute epididymitis occurring in older men or following urinary tract instrumentation is usually caused by urinary pathogens. Similarly, epididymitis in men who have practiced insertive rectal intercourse is often caused by Enterobacteriaceae. These men usually have no urethritis but do have bacteriuria.

TREATMENT: EPIDIDYMITIS

Ceftriaxone (250 mg as a single dose IM) followed by doxycycline (100 mg by mouth twice daily for 10 days) constitutes effective treatment for epididymitis caused by *N. gonorrhoeae* or *C. trachomatis*. Fluoroquinolones are no longer recommended for treatment of gonorrhea in the United States because of the emergence of resistant strains of *N. gonorrhoeae,* especially (but not only) among MSM (Fig. 130-1). Oral levofloxacin (500 mg once daily for 10 days) is also effective for syndrome-based initial treatment of epididymitis when infection with Enterobacteriaceae is suspected; however, this regimen should be combined with effective therapy for possible gonococcal or chlamydial infection unless bacteriuria with Enterobacteriaceae is confirmed.

Figure 130-1

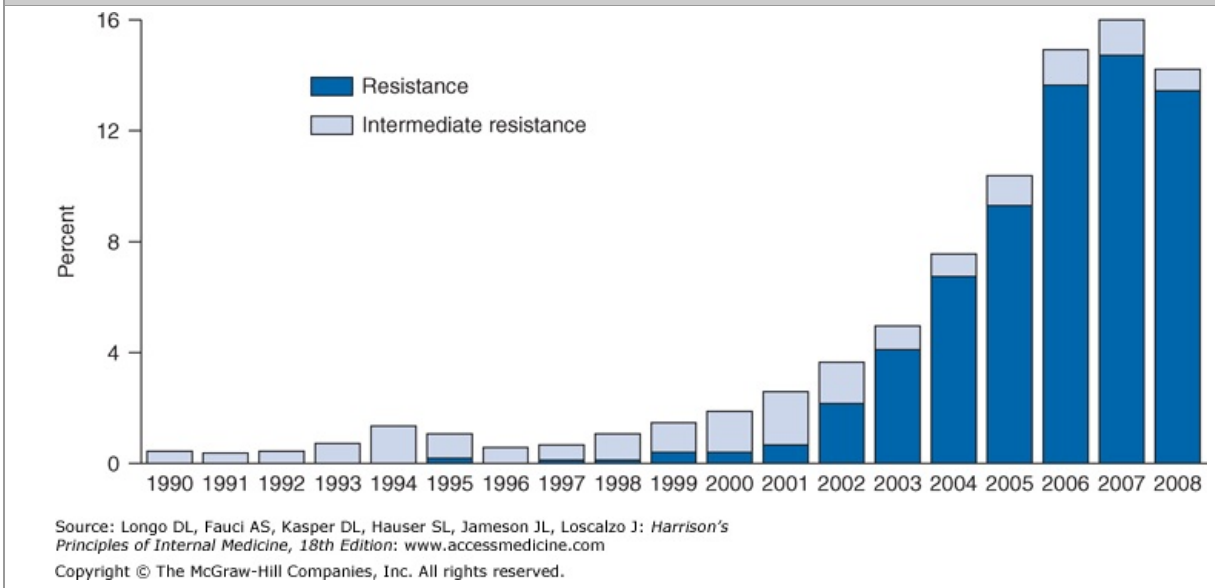

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Percentage of *N. gonorrhoeae* isolates with intermediate resistance or resistance to ciprofloxacin, by year: Gonococcal Isolate Surveillance Project, United States, 1990–2008. Intermediate resistance is defined by ciprofloxacin minimal inhibitory concentrations (MICs) of 0.125 –0.5 μg/mL and resistance by MICs of ≥1 μg/mL. *(From the Centers for Disease Control and Prevention: Sexually Transmitted Disease Surveillance, 2008. Atlanta, U.S. Department of Health and Human Services; November 2009.)*

Urethritis and the Urethral Syndrome in Women

C. trachomatis, *N. gonorrhoeae*, and occasionally HSV cause symptomatic urethritis—known as the urethral syndrome in women—that is characterized by "internal" dysuria (usually without urinary urgency or frequency), pyuria, and an absence of *Escherichia coli* and other uropathogens at counts of $\geq 10^2$/mL in urine. In contrast, the dysuria associated with vulvar herpes or vulvovaginal candidiasis (and perhaps with trichomoniasis) is often described as "external," being caused by painful contact of urine with the inflamed or ulcerated labia or introitus. Acute onset, association with urinary urgency or frequency, hematuria, or suprapubic bladder tenderness suggests bacterial cystitis. Among women with symptoms of acute bacterial cystitis, costovertebral pain and tenderness or fever suggests acute pyelonephritis. **The management of bacterial urinary tract infection (UTI) is discussed in Chap. 288.**

Signs of vulvovaginitis, coupled with symptoms of external dysuria, suggest vulvar infection (e.g., with HSV or *Candida albicans*). Among dysuric women without signs of vulvovaginitis, bacterial UTI must be differentiated from the urethral syndrome by assessment of risk, evaluation of the pattern of symptoms and signs, and specific microbiologic testing. An STI etiology of the urethral syndrome is suggested by young age, more than one current sexual partner, a new partner within the past month, a partner with urethritis, or coexisting mucopurulent cervicitis (see below). The finding of a single urinary pathogen, such as *E. coli* or *Staphylococcus saprophyticus*, at a concentration of $\geq 10^2$/mL in a properly collected specimen of midstream urine from a dysuric woman with pyuria indicates probable bacterial UTI, whereas pyuria with $<10^2$ conventional uropathogens per milliliter of urine ("sterile" pyuria) suggests acute urethral syndrome due to *C. trachomatis* or *N. gonorrhoeae*. Gonorrhea and chlamydial infection should be sought by specific tests (e.g., NAATs on the first 10 mL of voided urine). Among dysuric women with sterile pyuria caused by infection with *N. gonorrhoeae* or *C. trachomatis*, appropriate treatment alleviates dysuria.

Vulvovaginal Infections

ABNORMAL VAGINAL DISCHARGE

If directly questioned about vaginal discharge during routine health checkups, many women acknowledge having nonspecific symptoms of vaginal discharge that do not correlate with objective signs of inflammation or with actual infection. However, unsolicited reporting of abnormal vaginal discharge does suggest bacterial vaginosis or trichomoniasis. Specifically, an abnormally increased amount or an abnormal odor of the discharge is associated with one or both of these conditions. Cervical infection with *N. gonorrhoeae* or *C. trachomatis* does not often cause an increased amount or abnormal odor of discharge; however, when these pathogens cause cervicitis, they—like *T. vaginalis*—often result in an increased number of neutrophils in vaginal fluid that thus takes on a yellow color. Vulvar conditions such as genital herpes or vulvovaginal candidiasis can cause vulvar

pruritus, burning, irritation, or lesions as well as external dysuria (as urine passes over the inflamed vulva or areas of epithelial disruption) or vulvar dyspareunia.

Certain vulvovaginal infections may have serious sequelae. Trichomoniasis, bacterial vaginosis, and vulvovaginal candidiasis have all been associated with increased risk of acquisition of HIV infection. Vaginal trichomoniasis and bacterial vaginosis early in pregnancy independently predict premature onset of labor. Bacterial vaginosis can also lead to anaerobic bacterial infection of the endometrium and salpinges. Vaginitis may be an early and prominent feature of toxic shock syndrome, and recurrent or chronic vulvovaginal candidiasis develops with increased frequency among women with systemic illnesses, such as diabetes mellitus or HIV-related immunosuppression (although only a very small proportion of women with recurrent vulvovaginal candidiasis in industrialized countries actually have a serious predisposing illness).

Thus vulvovaginal symptoms or signs warrant careful evaluation, including pelvic examination, simple rapid diagnostic tests, and appropriate therapy specific for the anatomic site and type of infection. Unfortunately, a survey in the United States indicated that clinicians seldom perform the tests required to establish the cause of such symptoms. Further, comparison of telephone and office management of vulvovaginal symptoms has documented the inaccuracy of the former, and comparison of evaluations by nurse-midwives with those by physician-practitioners showed that the practitioners' clinical evaluations correlated poorly both with the nurses' evaluations and with diagnostic tests. The diagnosis and treatment of the three most common types of vaginal infection are summarized in Table 130–5.

Table 130–5 Diagnostic Features and Management of Vaginal Infection

Feature	Normal Vaginal Examination	Vulvovaginal Candidiasis	Trichomonal Vaginitis	Bacterial Vaginosis
Etiology	Uninfected; lactobacilli predominant	*Candida albicans*	*Trichomonas vaginalis*	Associated with *Gardnerella vaginalis*, various anaerobic and/or noncultured bacteria, and mycoplasmas
Typical symptoms	None	Vulvar itching and/or irritation	Profuse purulent discharge; vulvar itching	Malodorous, slightly increased discharge
Discharge				
Amount	Variable; usually scant	Scant	Often profuse	Moderate
Color[a]	Clear or translucent	White	White or yellow	White or gray
Consistency	Nonhomogeneous, floccular	Clumped; adherent plaques	Homogeneous	Homogeneous, low viscosity; uniformly coats vaginal walls
Inflammation of vulvar or vaginal epithelium	None	Erythema of vaginal epithelium, introitus; vulvar dermatitis, fissures common	Erythema of vaginal and vulvar epithelium; colpitis macularis	None
pH of vaginal fluid[b]	Usually ≤4.5	Usually ≤4.5	Usually ≥5	Usually >4.5
Amine ("fishy") odor with 10% KOH	None	None	May be present	Present
Microscopy[c]	Normal epithelial cells; lactobacilli predominant	Leukocytes, epithelial cells; mycelia or pseudomycelia in up to 80% of *C. albicans* culture-positive persons with typical symptoms	Leukocytes; motile trichomonads seen in 80–90% of symptomatic patients, less often in the absence of symptoms	Clue cells; few leukocytes; no lactobacilli or only a few outnumbered by profuse mixed microbiota, nearly always including *G. vaginalis* plus anaerobic species on Gram's stain (Nugent's score ≥7)
Other laboratory		Isolation of *Candida* spp.	Isolation of T. *vaginalis* or	

findings			positive NAAT[d]	
Usual treatment	None	Azole cream, tablet, or suppository—e.g., miconazole (100-mg vaginal suppository) or clotrimazole (100-mg vaginal tablet) once daily for 7 days Fluconazole, 150 mg orally (single dose)	Metronidazole or tinidazole, 2 g orally (single dose) Metronidazole, 500 mg PO bid for 7 days	Metronidazole, 500 mg PO bid for 7 days Metronidazole gel, 0.75%, one applicator (5 g) intravaginally once daily for 5 days Clindamycin, 2% cream, one full applicator vaginally each night for 7 days
Usual management of sexual partner	None	None; topical treatment if candidal dermatitis of penis is detected	Examination for STD; treatment with metronidazole, 2 g PO (single dose)	None

[a]Color of discharge is best determined by examination against the white background of a swab.

[b]A pH determination is not useful if blood is present.

[c]To detect fungal elements, vaginal fluid is digested with 10% KOH prior to microscopic examination; to examine for other features, fluid is mixed (1:1) with physiologic saline. Gram's stain is also excellent for detecting yeasts (less predictive of vulvovaginitis) and pseudomycelia or mycelin (strongly predictive of vulvovaginitis) and for distinguishing normal flora from the mixed flora seen in bacterial vaginosis, but it is less sensitive than the saline preparation for detection of *T. vaginalis.*

[d]NAAT, nucleic acid amplification test (where available).

Source: *The Practitioner's Handbook for the Management of Sexually Transmitted Diseases*accessed from *http://depts.washington.edu/nnptc/online_training/std_handbook/pdfs/ch6_vaginitis.pdf.*

Inspection of the vulva and perineum may reveal tender genital ulcerations or fissures (typically due to HSV infection or vulvovaginal candidiasis) or discharge visible at the introitus before insertion of a speculum (suggestive of bacterial vaginosis or trichomoniasis). Speculum examination permits the clinician to discern whether the discharge in fact looks abnormal and whether any abnormal discharge in the vagina emanates from the cervical os (mucoid and, if abnormal, yellow) or from the vagina (not mucoid, since the vaginal epithelium does not produce mucus). Symptoms or signs of abnormal vaginal discharge should prompt testing of vaginal fluid for pH, for a fishy odor when mixed with 10% KOH, and for certain microscopic features when mixed with saline (motile trichomonads and/or "clue cells") and with 10% KOH (pseudohyphae or hyphae indicative of vulvovaginal candidiasis). Additional objective laboratory tests useful for establishing the cause of abnormal vaginal discharge include rapid point-of-care tests for bacterial vaginosis, as described below, and a DNA probe test (the Affirm test) to detect *T. vaginalis* and *C. albicans* as well as the increased concentrations of *Gardnerella vaginalis* associated with bacterial vaginosis. Gram's staining of vaginal fluid can be used to score alterations in the vaginal microbiota but is employed primarily for research purposes and requires some familiarity with the morphotypes and scale involved.

Treatment: Vaginal Discharge

Patterns of treatment for vaginal discharge vary widely. In developing countries, where clinics or pharmacies often dispense treatment based on symptoms alone without examination or testing, oral treatment with metronidazole—particularly with a 7-day regimen—provides reasonable coverage against both trichomoniasis and bacterial vaginosis, the usual causes of symptoms of vaginal discharge; metronidazole treatment of sex partners prevents reinfection of women with trichomoniasis, even though it does not help prevent the recurrence of bacterial vaginosis. Guidelines for syndromic management promulgated by the World Health Organization suggest consideration of treatment for cervical infection and for trichomoniasis, bacterial vaginosis, and vulvovaginal candidiasis in women with symptoms of abnormal vaginal discharge. However, it is important to note that the majority of chlamydial and gonococcal cervical infections produce no symptoms.

In industrialized countries, clinicians treating symptoms and signs of abnormal vaginal discharge should, at a minimum, differentiate between bacterial vaginosis and trichomoniasis, because optimal management of patients and partners differs for these two conditions (as discussed briefly below).

VAGINAL TRICHOMONIASIS

(See also Chap. 215) Symptomatic trichomoniasis characteristically produces a profuse, yellow, purulent, homogeneous vaginal discharge and vulvar irritation, sometimes with visible inflammation of the vaginal and vulvar epithelium and petechial lesions on the cervix (the so-called strawberry cervix,

usually evident only by colposcopy). The pH of vaginal fluid—normally <4.7—usually rises to ≥5. In women with typical symptoms and signs of trichomoniasis, microscopic examination of vaginal discharge mixed with saline reveals motile trichomonads in most culture-positive cases. However, saline microscopy probably detects only one-half of all cases, and, especially in the absence of symptoms or signs, culture is usually required for detection of the organism. NAAT for *T. vaginalis* is as sensitive as or more sensitive than culture, and NAAT of urine has disclosed surprisingly high prevalences of this pathogen among men at several STD clinics in the United States. Treatment of asymptomatic as well as symptomatic cases reduces rates of transmission and prevents later development of symptoms.

Treatment: Vaginal Trichomoniasis

Only nitroimidazoles (e.g., metronidazole and tinidazole) consistently cure trichomoniasis. A single 2-g oral dose of metronidazole is effective and much less expensive than the alternatives. Tinidazole has a longer half-life than metronidazole, causes fewer gastrointestinal symptoms, and is especially useful in treating trichomoniasis that fails to respond to metronidazole. Treatment of sexual partners—facilitated by dispensing metronidazole to the female patient to give to her partner(s), with a warning about avoiding the concurrent use of alcohol—significantly reduces both the risk of reinfection and the reservoir of infection; treating the partner is the standard of care. Intravaginal treatment with 0.75% metronidazole gel is not reliable for vaginal trichomoniasis. Systemic use of metronidazole is recommended throughout pregnancy. In a large randomized trial, metronidazole treatment of trichomoniasis during pregnancy did not reduce—and in fact actually increased—the frequency of perinatal morbidity; thus routine screening of asymptomatic pregnant women for trichomoniasis is not recommended.

BACTERIAL VAGINOSIS

Bacterial vaginosis (formerly termed *nonspecific vaginitis*, *Haemophilusvaginitis*, *anaerobic vaginitis*, or *Gardnerella-associated vaginal discharge*) is a syndrome of uncertain etiology that is characterized by symptoms of vaginal malodor and a slightly to moderately increased white discharge, which appears homogeneous, is low in viscosity, and evenly coats the vaginal mucosa. Bacterial vaginosis has been associated with increased risk of acquiring several other genital infections, including those caused by HIV, *C. trachomatis*, and *N. gonorrhoeae*. Other risk factors include recent unprotected vaginal intercourse, having a female sex partner, and vaginal douching. Although bacteria associated with bacterial vaginosis have been detected under the foreskin of uncircumcised men, metronidazole treatment of male partners has not reduced the rate of recurrence among affected women.

Among women with bacterial vaginosis, culture of vaginal fluid has shown markedly increased prevalences and concentrations of *G. vaginalis*, *Mycoplasma hominis*, and several anaerobic bacteria [e.g., *Mobiluncus*, *Prevotella* (formerly *Bacteroides*), and some *Peptostreptococcus* species] as well as an absence of hydrogen peroxide–producing *Lactobacillus* spp., that constitute most of the normal vaginal microbiota and help protect against certain cervical and vaginal infections. Broad-range polymerase chain reaction (PCR) amplification of 16S rDNA in vaginal fluid, with subsequent identification of specific bacterial species by various methods, has documented an even greater and unexpected bacterial diversity, including several unique species not previously cultivated [e.g., three species in the order Clostridiales that appear to be specific for bacterial vaginosis and are associated with metronidazole treatment failure (Fig. 130-2)]. Also detected are DNA sequences related to *Atopobium vaginae*, an organism that is strongly associated with bacterial vaginosis, is resistant to metronidazole, and is also associated with recurrent bacterial vaginosis after metronidazole treatment. Other genera newly implicated in bacterial vaginosis include *Megasphaera*, *Leptotrichia*, *Eggerthella*, and *Dialister*.

Figure 130-2

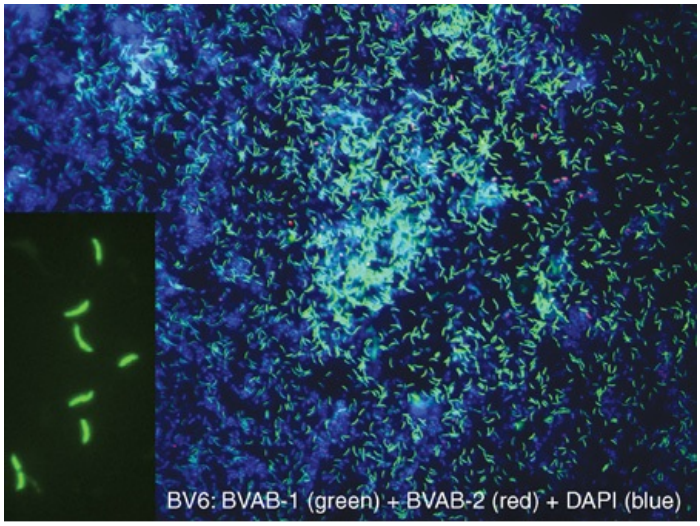

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Broad-range PCR amplification of 16S rDNA in vaginal fluid from a woman with bacterial vaginosis shows a field of bacteria hybridizing with probes for bacterial vaginosis–associated bacterium 1 (BVAB-1, visible as a thin, curved green rod) and for BVAB-2 (red). The inset shows that BVAB-1 has a morphology similar to that of *Mobiluncus* (curved rod). *(Reprinted with permission from DN Fredricks et al.)*

Bacterial vaginosis is conventionally diagnosed clinically with the Amsel criteria that include any three of the following four clinical abnormalities: (1) objective signs of increased white homogeneous vaginal discharge; (2) a vaginal discharge pH of >4.5; (3) liberation of a distinct fishy odor (attributable to volatile amines such as trimethylamine) immediately after vaginal secretions are mixed with a 10% solution of KOH; and (4) microscopic demonstration of "clue cells" (vaginal epithelial cells coated with coccobacillary organisms, which have a granular appearance and indistinct borders; Fig. 130-3) on a wet mount prepared by mixing vaginal secretions with normal saline in a ratio of ~1:1.

Figure 130-3

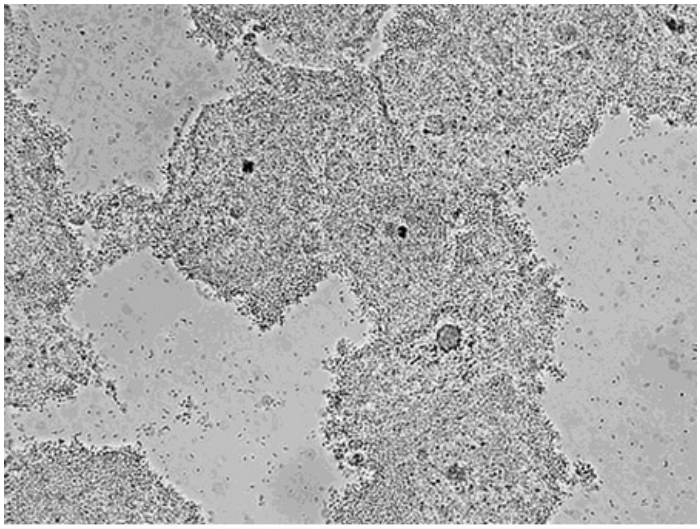

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Wet mount of vaginal fluid showing typical clue cells from a woman with bacterial vaginosis. Note the obscured epithelial cell margins and the granular appearance attributable to many adherent bacteria (x400). *(Photograph provided by Lorna K. Rabe, reprinted with permission from S Hillier et al, in KK Holmes et al (eds). Sexually Transmitted Diseases, 4th ed. New York, McGraw-Hill, 2008.)*

Treatment: Bacterial Vaginosis

The standard dosage of oral metronidazole for the treatment of bacterial vaginosis is 500 mg twice daily for 7 days. The single 2-g oral dose of metronidazole recommended for trichomoniasis produces

significantly lower short-term cure rates and should not be used. Intravaginal treatment with 2% clindamycin cream [one full applicator (5 g containing 100 mg of clindamycin phosphate) each night for 7 nights] or with 0.75% metronidazole gel [one full applicator (5 g containing 37.5 mg of metronidazole) twice daily for 5 days] is also approved for use in the United States and does not elicit systemic adverse reactions; the response to both of these treatments is similar to the response to oral metronidazole. Other alternatives include oral clindamycin (300 mg twice daily for 7 days), clindamycin ovules (100 g intravaginally once at bedtime for 3 days), and oral tinidazole (1 g daily for 5 days or 2 g daily for 3 days). Unfortunately, recurrence over the long term (i.e., several months later) is distressingly common after either oral or intravaginal treatment. A randomized trial comparing intravaginal gel containing 37.5 mg of metronidazole with a suppository containing 500 mg of metronidazole plus nystatin (the latter not marketed in the United States) showed significantly higher rates of recurrence with the 37.5-mg regimen; this result suggests that higher metronidazole dosages may be important in topical intravaginal therapy. Recurrences can be significantly lessened with the twice-weekly use of suppressive intravaginal metronidazole gel. As stated above, treatment of male partners with metronidazole does not prevent recurrence of bacterial vaginosis.

Efforts to replenish numbers of vaginal lactobacilli that produce hydrogen peroxide and probably sustain vaginal health have been largely unsuccessful. While one randomized trial of orally ingested lactobacilli found reduced rates of recurrent bacterial vaginosis, this result has not yet been either confirmed or refuted, and a randomized multicenter trial in the United States found no benefit of repeated intravaginal inoculation of a vaginal peroxide-producing *Lactobacillus* species following treatment of bacterial vaginosis with metronidazole. A meta-analysis of 18 studies concluded that bacterial vaginosis during pregnancy substantially increased the risk of preterm delivery and of spontaneous abortion. However, most studies of topical intravaginal treatment of bacterial vaginosis with clindamycin during pregnancy have not reduced adverse pregnancy outcomes. Numerous trials of oral metronidazole treatment during pregnancy have given inconsistent results, and a 2007 Cochrane review concluded that antenatal treatment of women with bacterial vaginosis—even those with previous preterm delivery—did not reduce the risk of preterm delivery. The U.S. Preventive Services Task Force thus recommends against routine screening of pregnant women for bacterial vaginosis.

VULVOVAGINAL PRURITUS, BURNING, OR IRRITATION

Vulvovaginal candidiasis produces vulvar pruritus, burning, or irritation, generally without symptoms of increased vaginal discharge or malodor. Genital herpes can produce similar symptoms, with lesions sometimes difficult to distinguish from the fissures and inflammation caused by candidiasis. Signs of vulvovaginal candidiasis include vulvar erythema, edema, fissures, and tenderness. With candidiasis, a white scanty vaginal discharge sometimes takes the form of white thrush-like plaques or cottage cheese–like curds adhering loosely to the vaginal mucosa. *C. albicans* accounts for nearly all cases of symptomatic vulvovaginal candidiasis, which probably arise from endogenous strains of *C. albicans* that have colonized the vagina or the intestinal tract. Complicated vulvovaginal candidiasis includes cases that recur four or more times per year; are unusually severe; are caused by non-*albicans Candida* spp.; or occur in women with uncontrolled diabetes, debilitation, immunosuppression, or pregnancy.

In addition to compatible clinical symptoms, the diagnosis of vulvovaginal candidiasis usually involves the demonstration of pseudohyphae or hyphae by microscopic examination of vaginal fluid mixed with saline or 10% KOH or subjected to Gram's staining. Microscopic examination is less sensitive than culture but correlates better with symptoms. Culture is typically reserved for cases that do not respond to standard first-line antimycotic agents and is undertaken to rule out imidazole or azole resistance (often associated with *Candida glabrata*) or before the initiation of suppressive antifungal therapy for recurrent disease.

Treatment: Vulvovaginal Pruritus, Burning, or Irritation

Symptoms and signs of vulvovaginal candidiasis warrant treatment, usually intravaginal administration of any of several imidazole antibiotics (e.g., miconazole or clotrimazole) for 3–7 days or of a single dose of oral fluconazole (Table 130–5). Over-the-counter marketing of such preparations has reduced the cost of care and made treatment more convenient for many women with recurrent yeast vulvovaginitis. However, most women who purchase these preparations do not have vulvovaginal candidiasis, while many do have other vaginal infections that require different treatment. Therefore, only women with classic symptoms of vulvar pruritus and a history of previous episodes of yeast vulvovaginitis documented by an experienced clinician should self-treat. Short-course topical intravaginal azole drugs are effective for the treatment of uncomplicated vulvovaginal candidiasis (e.g., clotrimazole, two 100-mg vaginal tablets daily for 3 days; or miconazole, a 1200-mg vaginal suppository as a single dose). Single-dose oral treatment with fluconazole (150 mg) is also effective and is preferred by many patients. Management of complicated cases (see above) and those that do not respond to the usual intravaginal or single-dose oral therapy often involves prolonged or periodic oral therapy; this situation is discussed extensively in the 2010 CDC STD treatment guidelines *(http://www.cdc.gov/std/treatment*). Treatment of sexual partners is not routinely indicated.

OTHER CAUSES OF VAGINAL DISCHARGE OR VAGINITIS

In the ulcerative vaginitis associated with staphylococcal toxic shock syndrome, *Staphylococcus aureus* should be promptly identified in vaginal fluid by Gram's stain and by culture. In desquamative inflammatory vaginitis, smears of vaginal fluid reveal neutrophils, massive vaginal epithelial-cell exfoliation with increased numbers of parabasal cells, and gram-positive cocci; this syndrome may respond to treatment with 2% clindamycin cream, often given in combination with topical steroid preparations for several weeks. Additional causes of vaginitis and vulvovaginal symptoms include retained foreign bodies (e.g., tampons), cervical caps, vaginal spermicides, vaginal antiseptic preparations or douches, vaginal epithelial atrophy (in postmenopausal women or during prolonged breast-feeding in the postpartum period), allergic reactions to latex condoms, vaginal aphthae associated with HIV infection or Behçt's syndrome, and vestibulitis (a poorly understood syndrome).

Mucopurulent Cervicitis

Mucopurulent cervicitis (MPC) refers to inflammation of the columnar epithelium and subepithelium of the endocervix and of any contiguous columnar epithelium that lies exposed in an ectopic position on the ectocervix. MPC in women represents the "silent partner" of urethritis in men, being equally common and often caused by the same agents (*N. gonorrhoeae*, *C. trachomatis*, or—as shown by case-control studies—*M. genitalium*); however, MPC is more difficult than urethritis to recognize. As the most common manifestation of these serious bacterial infections in women, MPC can be a harbinger or sign of upper genital tract infection, also known as *pelvic inflammatory disease* (PID; see below). In pregnant women, MPC can lead to obstetric complications. In a prospective study in Seattle of 167 consecutive patients with MPC [defined on the basis of yellow endocervical mucopus or ≥30 polymorphonuclear leukocytes (PMNs)/1000x microscopic field] who were seen at STD clinics during the 1980s, slightly more than one-third of cervicovaginal specimens tested for *C. trachomatis*, *N. gonorrhoeae*, *M. genitalium*, HSV, and *T. vaginalis* revealed no identifiable etiology (Fig. 130-4). More recently, a study in Baltimore using NAATs for these pathogens still failed to identify a microbiologic etiology in nearly one-half of the 133 women with MPC.

Figure 130-4

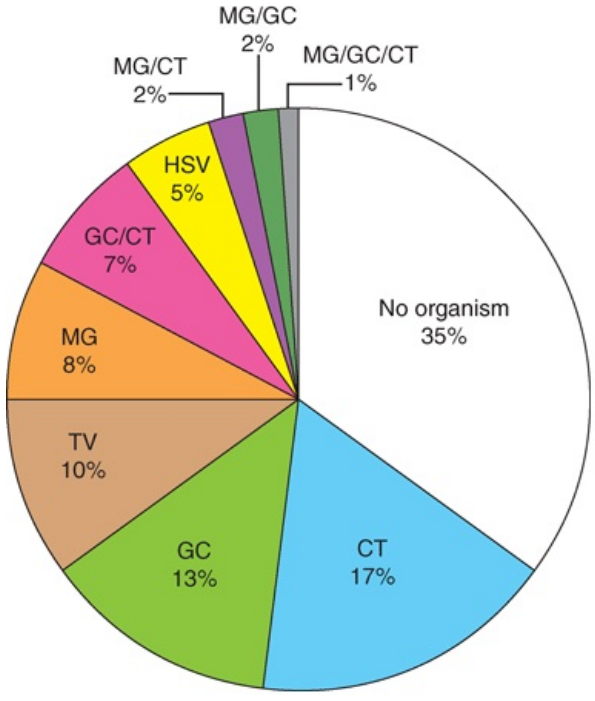

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Organisms detected among female STD clinic patients with mucopurulent cervicitis (*n* = 167). GC, gonococcus; CT, *Chlamydia trachomatis*; MG, *Mycoplasma genitalium*; TV, *Trichomonas vaginalis*; HSV, herpes simplex virus. (*Courtesy of Dr. Lisa Manhart; with permission* .)

The diagnosis of MPC rests on the detection of cardinal signs at the cervix, including yellow mucopurulent discharge from the cervical os, endocervical bleeding upon gentle swabbing, and edematous cervical ectopy (see below); the latter two findings are somewhat more common with MPC due to chlamydial infection, but signs alone do not allow a distinction between the causative pathogens. Unlike the endocervicitis produced by gonococcal or chlamydial infection, cervicitis caused by HSV produces ulcerative lesions on the stratified squamous epithelium of the ectocervix as well as on the columnar epithelium. Yellow cervical mucus on a white swab removed from the endocervix

indicates the presence of PMNs. Gram's staining may confirm their presence, although it adds relatively little to the diagnostic value of assessment for cervical signs. The presence of ≥20 PMNs/1000x microscopic field within strands of cervical mucus not contaminated by vaginal squamous epithelial cells or vaginal bacteria indicates endocervicitis. Detection of intracellular gram-negative diplococci in carefully collected endocervical mucus is quite specific but ≤50% sensitive for gonorrhea. Therefore, specific and sensitive tests for *N. gonorrhoeae* as well as for *C. trachomatis* (e.g., NAATs) are always indicated in the evaluation of MPC.

TREATMENT: MUCOPURULENT CERVICITIS

Although the above criteria for MPC are neither highly specific nor highly predictive of gonococcal or chlamydial infection in some settings, the 2010 CDC STD guidelines call for consideration of empirical treatment for MPC, pending test results, in most patients. Presumptive treatment with antibiotics active against *C. trachomatis* should be provided for women at increased risk for this common STI (risk factors: age <25 years, new or multiple sex partners, and unprotected sex), especially if follow-up cannot be ensured. Concurrent therapy for gonorrhea is indicated if the prevalence of this infection is substantial in the relevant patient population (e.g., young adults, a clinic with documented high prevalence). In this situation, therapy should include a single-dose regimen effective for gonorrhea plus treatment for chlamydial infection, as outlined in Table 130–4 for the treatment of urethritis. In settings where gonorrhea is much less common than chlamydial infection, initial therapy for chlamydial infection alone suffices, pending test results for gonorrhea. The etiology and potential benefit of treatment for endocervicitis not associated with gonorrhea or chlamydial infection have not been established. Although the antimicrobial susceptibility of *M. genitalium* is not yet well defined, the organism frequently persists after doxycycline therapy, and it currently seems reasonable to use azithromycin to treat possible *M. genitalium* infection in such cases. The sexual partner(s) of a woman with MPC should be examined and given a regimen similar to that chosen for the woman unless results of tests for gonorrhea or chlamydial infection in either partner warrant different therapy or no therapy.

Cervical Ectopy

Cervical ectopy, often mislabeled "cervical erosion," is easily confused with infectious endocervicitis. Ectopy represents the presence of the one-cell-thick columnar epithelium extending from the endocervix out onto the visible ectocervix. In ectopy, the cervical os may contain clear or slightly cloudy mucus but usually not yellow mucopus. Colposcopy shows intact epithelium. Normally found during adolescence and early adulthood, ectopy gradually recedes through the second and third decades of life, as squamous metaplasia replaces the ectopic columnar epithelium. Oral contraceptive use favors the persistence or reappearance of ectopy, while smoking apparently accelerates squamous metaplasia. Cauterization of ectopy is not warranted. Ectopy may render the cervix more susceptible to infection with *N. gonorrhoeae*, *C. trachomatis*, or HIV.

Pelvic Inflammatory Disease

The term *pelvic inflammatory disease* usually refers to infection that ascends from the cervix or vagina to involve the endometrium and/or fallopian tubes. Infection can extend beyond the reproductive tract to cause pelvic peritonitis, generalized peritonitis, perihepatitis, perisplenitis, or pelvic abscess. Rarely, infection not related to specific sexually transmitted pathogens extends secondarily to the pelvic organs (1) from adjacent foci of inflammation (e.g., appendicitis, regional ileitis, or diverticulitis) or bacterial vaginosis, (2) as a result of hematogenous dissemination (e.g., of tuberculosis or staphylococcal bacteremia), or (3) as a complication of certain tropical diseases (e.g., schistosomiasis). Intrauterine infection can be primary (spontaneously occurring and usually sexually transmitted) or secondary to invasive intrauterine surgical procedures [e.g., dilatation and curettage, termination of pregnancy, insertion of an intrauterine device (IUD), or hysterosalpingography] or to parturition.

ETIOLOGY

The agents most often implicated in acute PID include the primary causes of endocervicitis (e.g., *N. gonorrhoeae* and *C. trachomatis*) and organisms that can be regarded as components of an altered vaginal microbiota. In general, PID is most often caused by*N. gonorrhoeae* where there is a high incidence of gonorrhea—e.g., in inner-city populations in the United States. In case-control studies, *M. genitalium* has also been significantly associated with histopathologic diagnoses of endometritis and with salpingitis.

Anaerobic and facultative organisms (especially *Prevotella* species, peptostreptococci, *E. coli*, *Haemophilus influenzae*, and group B streptococci) as well as genital mycoplasmas have been isolated from the peritoneal fluid or fallopian tubes in a varying proportion (typically one-fourth to one-third) of women with PID studied in the United States. The difficulty of determining the exact microbial etiology of an individual case of PID—short of using invasive procedures for specimen collection—has implications for the approach to empirical antimicrobial treatment of this infection.

EPIDEMIOLOGY

In the United States, the estimated annual number of initial visits to physicians' offices for PID by

women 15–44 years of age fell from an average of 400,000 during the 1980s to 250,000 in 1999 and then to 104,000 in 2008. Hospitalizations for acute PID in the United States also declined steadily throughout the 1980s and early 1990s but have remained fairly constant at 70,000–100,000 per year since 1995. Important risk factors for acute PID include the presence of endocervical infection or bacterial vaginosis, a history of salpingitis or of recent vaginal douching, and recent insertion of an IUD. Certain other iatrogenic factors, such as dilatation and curettage or cesarean section, can increase the risk of PID, especially among women with endocervical gonococcal or chlamydial infection or bacterial vaginosis. Symptoms of *N. gonorrhoeae*–associated and *C. trachomatis*–associated PID often begin during or soon after the menstrual period; this timing suggests that menstruation is a risk factor for ascending infection from the cervix and vagina. Experimental inoculation of the fallopian tubes of nonhuman primates has shown that repeated exposure to *C. trachomatis* leads to the greatest degree of tissue inflammation and damage; thus, immunopathology probably contributes to the pathogenesis of chlamydial salpingitis. Women using oral contraceptives appear to be at decreased risk of symptomatic PID, and tubal sterilization reduces the risk of salpingitis by preventing intraluminal spread of infection into the tubes.

CLINICAL MANIFESTATIONS

Endometritis: A Clinical Pathologic Syndrome

A study of women with clinically suspected PID who were undergoing both endometrial biopsy and laparoscopy showed that those with endometritis alone differed from those who also had salpingitis in significantly less often having lower quadrant, adnexal, or cervical motion or abdominal rebound tenderness; fever; or elevated C-reactive protein levels. In addition, women with endometritis alone differed from those with neither endometritis nor salpingitis in more often having gonorrhea, chlamydial infection, and risk factors such as douching or IUD use. Thus, women with endometritis alone were intermediate between those with neither endometritis nor salpingitis and those with salpingitis with respect to risk factors, clinical manifestations, cervical infection prevalence, and elevated C-reactive protein level. Women with endometritis alone are at lower risk of subsequent tubal occlusion and resulting infertility than are those with salpingitis.

Salpingitis

Symptoms of nontuberculous salpingitis classically evolve from a yellow or malodorous vaginal discharge caused by MPC and/or bacterial vaginosis to midline abdominal pain and abnormal vaginal bleeding caused by endometritis and then to bilateral lower abdominal and pelvic pain caused by salpingitis, with nausea, vomiting, and increased abdominal tenderness if peritonitis develops.

The abdominal pain in nontuberculous salpingitis is usually described as dull or aching. In some cases, pain is lacking or atypical, but active inflammatory changes are found in the course of an unrelated evaluation or procedure, such as a laparoscopic evaluation for infertility. Abnormal uterine bleeding precedes or coincides with the onset of pain in ~40% of women with PID, symptoms of urethritis (dysuria) occur in 20%, and symptoms of proctitis (anorectal pain, tenesmus, and rectal discharge or bleeding) are occasionally seen in women with gonococcal or chlamydial infection.

Speculum examination shows evidence of MPC (yellow endocervical discharge, easily induced endocervical bleeding) in the majority of women with gonococcal or chlamydial PID. Cervical motion tenderness is produced by stretching of the adnexal attachments on the side toward which the cervix is pushed. Bimanual examination reveals uterine fundal tenderness due to endometritis and abnormal adnexal tenderness due to salpingitis that is usually, but not necessarily, bilateral. Adnexal swelling is palpable in about one-half of women with acute salpingitis, but evaluation of the adnexae in a patient with marked tenderness is not reliable. The initial temperature is >38°C in only about one-third of patients with acute salpingitis. Laboratory findings include elevation of the erythrocyte sedimentation rate (ESR) in 75% of patients with acute salpingitis and elevation of the peripheral white blood cell count in up to 60%.

Unlike nontuberculous salpingitis, genital tuberculosis often occurs in older women, many of whom are postmenopausal. Presenting symptoms include abnormal vaginal bleeding, pain (including dysmenorrhea), and infertility. About one-quarter of these women have had adnexal masses. Endometrial biopsy shows tuberculous granulomas and provides optimal specimens for culture.

Perihepatitis and Periappendicitis

Pleuritic upper abdominal pain and tenderness, usually localized to the right upper quadrant (RUQ), develop in 3–10% of women with acute PID. Symptoms of perihepatitis arise during or after the onset of symptoms of PID and may overshadow lower abdominal symptoms, thereby leading to a mistaken diagnosis of cholecystitis. In perhaps 5% of cases of acute salpingitis, early laparoscopy reveals perihepatic inflammation ranging from edema and erythema of the liver capsule to exudate with fibrinous adhesions between the visceral and parietal peritoneum. When treatment is delayed and laparoscopy is performed late, dense "violin-string" adhesions can be seen over the liver; chronic exertional or positional RUQ pain ensues when traction is placed on the adhesions. Although perihepatitis, also known as the *Fitz-Hugh–Curtis syndrome*, was for many years specifically attributed to gonococcal salpingitis, most cases are now attributed to chlamydial salpingitis. In patients with

chlamydial salpingitis, serum titers of microimmunofluorescent antibody to *C. trachomatis* are typically much higher when perihepatitis is present than when it is absent.

Physical findings include RUQ tenderness and usually include adnexal tenderness and cervicitis, even in patients whose symptoms do not suggest salpingitis. Results of liver function tests and RUQ ultrasonography are nearly always normal. The presence of MPC and pelvic tenderness in a young woman with subacute pleuritic RUQ pain and normal ultrasonography of the gallbladder points to a diagnosis of perihepatitis.

Periappendicitis (appendiceal serositis without involvement of the intestinal mucosa) has been found in ~5% of patients undergoing appendectomy for suspected appendicitis and can occur as a complication of gonococcal or chlamydial salpingitis.

Among women with salpingitis, HIV infection is associated with increased severity of salpingitis and with tuboovarian abscess requiring hospitalization and surgical drainage. Nonetheless, among women with HIV infection and salpingitis, the clinical response to conventional antimicrobial therapy (coupled with drainage of tuboovarian abscess, when found) has usually been satisfactory.

DIAGNOSIS

Treatment appropriate for PID must not be withheld from patients who have an equivocal diagnosis; it is better to err on the side of overdiagnosis and overtreatment. On the other hand, it is essential to differentiate between salpingitis and other pelvic pathology, particularly surgical emergencies such as appendicitis and ectopic pregnancy.

Nothing short of laparoscopy definitively identifies salpingitis, but routine laparoscopy to confirm suspected salpingitis is generally impractical. Most patients with acute PID have lower abdominal pain of <3 weeks' duration, pelvic tenderness on bimanual pelvic examination, and evidence of lower genital tract infection (e.g., MPC). Approximately 60% of such patients have salpingitis at laparoscopy, and perhaps 10–20% have endometritis alone. Among the patients with these findings, a rectal temperature >38°C, a palpable adnexal mass, and elevation of the ESR to >15 mm/h also raise the probability of salpingitis, which has been found at laparoscopy in 68% of patients with one of these additional findings, 90% of patients with two, and 96% of patients with three. However, only 17% of all patients with laparoscopy-confirmed salpingitis have had all three additional findings.

In a woman with pelvic pain and tenderness, increased numbers of PMNs (30 per 1000x microscopic field in strands of cervical mucus) or leukocytes outnumbering epithelial cells in vaginal fluid (in the absence of trichomonal vaginitis, that also produces PMNs in vaginal discharge) increase the predictive value of a clinical diagnosis of acute PID, as do onset with menses, history of recent abnormal menstrual bleeding, presence of an IUD, history of salpingitis, and sexual exposure to a male with urethritis. Appendicitis or another disorder of the gut is favored by the early onset of anorexia, nausea, or vomiting; the onset of pain later than day 14 of the menstrual cycle; or unilateral pain limited to the right or left lower quadrant. Whenever the diagnosis of PID is being considered, serum assays for human ß-chorionic gonadotropin should be performed; these tests are usually positive with ectopic pregnancy. Ultrasonography and MRI can be useful for the identification of tuboovarian or pelvic abscess. MRI of the tubes can also show increased tubal diameter, intratubal fluid, or tubal wall thickening in cases of salpingitis.

The primary and uncontested value of laparoscopy in women with lower abdominal pain is for the exclusion of other surgical problems. Some of the most common or serious problems that may be confused with salpingitis (e.g., acute appendicitis, ectopic pregnancy, corpus luteum bleeding, ovarian tumor) are unilateral. Unilateral pain or pelvic mass, although not incompatible with PID, is a strong indication for laparoscopy unless the clinical picture warrants laparotomy instead. Atypical clinical findings such as the absence of lower genital tract infection, a missed menstrual period, a positive pregnancy test, or failure to respond to appropriate therapy, are other common indications for laparoscopy. Endometrial biopsy is relatively sensitive and specific for the diagnosis of endometritis, which correlates well with the presence of salpingitis.

Endocervical swab specimens should be examined by NAATs for *N. gonorrhoeae* and *C. trachomatis*. At a minimum, vaginal fluid should be evaluated for the presence of PMNs, and endocervical secretions ideally should be assessed by Gram's staining for PMNs and gram-negative diplococci that indicate gonococcal infection. The clinical diagnosis of PID made by expert gynecologists is confirmed by laparoscopy or endometrial biopsy in ~90% of women who also have cultures positive for *N. gonorrhoeae* or *C. trachomatis*. Even among women with no symptoms suggestive of acute PID who were attending an STD clinic or a gynecology clinic in Pittsburgh, endometritis was significantly associated with endocervical gonorrhea or chlamydial infection or with bacterial vaginosis, being detected in 26%, 27%, and 15% of women with these conditions, respectively.

TREATMENT: PELVIC INFLAMMATORY DISEASE

The 2010 CDC guidelines recommend initiation of empirical treatment for PID in sexually active young women and other women at risk for PID if they are experiencing pelvic or lower abdominal pain, if no other cause for the pain can be identified, and if pelvic examination reveals one or more of the

following criteria for PID: cervical motion tenderness, uterine tenderness, or adnexal tenderness. Women with suspected PID can be treated as either outpatients or inpatients. In the multicenter Pelvic Inflammatory Disease Evaluation and Clinical Health (PEACH) trial, 831 women with mild to moderately severe symptoms and signs of PID were randomized to receive either inpatient treatment with IV cefoxitin and doxycycline or outpatient treatment with a single IM dose of cefoxitin plus oral doxycycline. Short-term clinical and microbiologic outcomes and long-term outcomes were equivalent in the two groups. Nonetheless, hospitalization should be considered when (1) the diagnosis is uncertain and surgical emergencies such as appendicitis and ectopic pregnancy cannot be excluded, (2) the patient is pregnant, (3) pelvic abscess is suspected, (4) severe illness or nausea and vomiting preclude outpatient management, (5) the patient has HIV infection, (6) the patient is assessed as unable to follow or tolerate an outpatient regimen, or (7) the patient has failed to respond to outpatient therapy. Some experts also prefer to hospitalize adolescents with PID for initial therapy, although younger women do as well as older women on outpatient therapy.

Recommended combination regimens for ambulatory or parenteral management of PID are presented in Table 130–6. Women managed as outpatients should receive a combined regimen with broad activity, such as ceftriaxone (to cover possible gonococcal infection) followed by doxycycline (to cover possible chlamydial infection). Metronidazole can be added, if tolerated, to enhance activity against anaerobes; this addition should be strongly considered if bacterial vaginosis is documented. Although few methodologically sound clinical trials (especially with prolonged follow-up) have been conducted, one meta-analysis suggested a benefit of providing good coverage against anaerobes.

Table 130–6 Combination Antimicrobial Regimens Recommended for Outpatient Treatment or for Parenteral Treatment of PID

Outpatient Regimens[a]	Parenteral Regimens
Ceftriaxone (250 mg IM once) *plus* Doxycycline (100 mg PO bid for 14 days) *plus*[b] Metronidazole (500 mg PO bid for 14 days)	Initiate parenteral therapy with either of the following regimens; continue parenteral therapy until 48 h after clinical improvement; then change to outpatient therapy, as described in the text. **Regimen A** Cefotetan (2 g IV q12h) *or* Cefoxitin (2 g IV q6h) *plus* Doxycycline (100 mg IV or PO q12h) **Regimen B** Clindamycin (900 mg IV q8h) *plus* Gentamicin (loading dose of 2 mg/kg IV or IM, then maintenance dose of 1.5 mg/kg q8h)

[a] See text for discussion of options in the patient who is intolerant of cephalosporins.

[b] The addition of metronidazole is recommended by some experts, particularly if bacterial vaginosis is present.

Source: Adapted from Centers for Disease Control and Prevention: MMWR Recomm Rep 59 (RR-12):1, 2010.

Neither doxycycline nor the fluoroquinolones provide reliable coverage for gonococcal infection today. Thus, adequate oral treatment of women with serious intolerance to cephalosporins is a challenge. If penicillins are an option, amoxicillin/clavulanic acid combined with doxycycline has effected short-term clinical response in one clinical trial. If fluoroquinolones are the only option and if the community prevalence and individual risk for gonorrhea are known to be low, oral levofloxacin (500 mg once daily) or ofloxacin (400 mg twice daily) for 14 days, with or without metronidazole, may be considered. In this case, it is imperative to perform a sensitive diagnostic test for gonorrhea (ideally, culture to test for antimicrobial susceptibility) before initiating therapy. For those women whose PID involves quinolone-resistant gonorrhea, treatment is uncertain but could include parenteral gentamicin or oral azithromycin, although the latter agent has not been studied for this purpose.

For hospitalized patients, the following two parenteral regimens have given nearly identical results in a multicenter randomized trial:

1. Doxycycline (100 mg twice daily, given IV or PO) plus cefotetan (2 g IV every 12 h) or cefoxitin (2 g IV every6 h): Administration of these drugs should be continued by the IV route for at least 48 h after the patient's condition improves and then followed with oral doxycycline (100 mg twice daily) to complete 14 days of therapy.
2. Clindamycin (900 mg IV every 8 h) plus gentamicin(2 mg/kg IV or IM, followed by 1.5 mg/kg every

8 h) in patients with normal renal function: Once-daily dosing of gentamicin (with combination of the total daily dose into a single daily dose) has not been evaluated in PID but has been efficacious in other serious infections and could be substituted. Treatment with these drugs should be continued for at least 48 h after the patient's condition improves and then followed with oral doxycycline (100 mg twice daily) or clindamycin (450 mg four times daily) to complete 14 days of therapy. In cases with tuboovarian abscess, clindamycin rather than doxycycline for continued therapy provides better coverage for anaerobic infection.

Follow-Up

Hospitalized patients should show substantial clinical improvement within 3–5 days. Women treated as outpatients should be clinically reevaluated within 72 h. A follow-up telephone survey of women seen in an emergency department and given a prescription for 10 days of oral doxycycline for PID found that 28% never filled the prescription and 41% stopped taking the medication early (after an average of 4.1 days), often because of persistent symptoms, lack of symptoms, or side effects. Women not responding favorably to ambulatory therapy should be hospitalized for parenteral therapy and further diagnostic evaluations, including a consideration of laparoscopy. Male sex partners should be evaluated and treated empirically for gonorrhea and chlamydial infection. After completion of treatment, tests for persistent or recurrent infection with *N. gonorrhoeae* or *C. trachomatis* should be performed if symptoms persist or recur or if the patient has not complied with therapy or has been reexposed to an untreated sex partner.

Surgery

Surgery is necessary for the treatment of salpingitis only in the face of life-threatening infection (such as rupture or threatened rupture of a tuboovarian abscess) or for drainage of an abscess. Conservative surgical procedures are usually sufficient. Pelvic abscesses can often be drained by posterior colpotomy, and peritoneal lavage can be used for generalized peritonitis.

PROGNOSIS

Late sequelae include infertility due to bilateral tubal occlusion, ectopic pregnancy due to tubal scarring without occlusion, chronic pelvic pain, and recurrent salpingitis. The overall postsalpingitis risk of infertility due to tubal occlusion in a large study in Sweden was 11% after one episode of salpingitis, 23% after two episodes, and 54% after three or more episodes. A University of Washington study found a sevenfold increase in the risk of ectopic pregnancy and an eightfold increase in the rate of hysterectomy after PID.

PREVENTION

A randomized controlled trial designed to determine whether selective screening for chlamydial infection reduces the risk of subsequent PID showed that women randomized to undergo screening had a 56% lower rate of PID over the following year than did women receiving the usual care without screening. This report helped prompt U.S. national guidelines for risk-based chlamydial screening of young women to reduce the incidence of PID and the prevalence of post-PID sequelae, while also reducing sexual transmission of *C. trachomatis*. The CDC and the U.S. Preventive Services Task Force recommend that sexually active women ≤25 years of age be screened for genital chlamydial infection annually. Despite this recommendation, screening coverage in many primary care settings remains low.

Ulcerative Genital or Perianal Lesions

Genital ulceration reflects a set of important STIs, most of which sharply increase the risk of sexual acquisition and shedding of HIV. In a 1996 study of genital ulcers in 10 of the U.S. cities with the highest rates of primary syphilis, PCR testing of ulcer specimens demonstrated HSV in 62% of patients, *Treponema pallidum* (the agent of chancroid) in 13%, and *Haemophilus ducreyi* (the agent of chancroid) in 12–20%. Today, genital herpes represents an even higher proportion of genital ulcers in the United States and other industrialized countries.

In Asia and Africa, chancroid (Fig. 130-5) was once considered the most common type of genital ulcer, followed in frequency by primary syphilis and then genital herpes (Fig. 130-6). With increased efforts to control chancroid and syphilis and widespread use of broad-spectrum antibiotics to treat STI-related syndromes, together with more frequent recurrences or persistence of genital herpes attributable to HIV infection, PCR testing of genital ulcers now clearly implicates genital herpes as the most common cause of genital ulceration in some developing countries. LGV caused by C. trachomatis; (Fig. 130-7) and donovanosis (granuloma inguinale, caused by Klebsiella granulomatis; see Fig. 161-1) continue to cause genital ulceration in developing countries. LGV virtually disappeared in industrialized countries during the first 20 years of the HIV pandemic, but outbreaks are again occurring in Europe (including the United Kingdom), in North America, and in Australia. In these outbreaks, LGV typically presents as proctitis in men who report unprotected receptive anal intercourse, very often in association with HIV and/or hepatitis C virus infection; the latter may be an acute infection acquired through the same exposure. Other causes of genital ulcers include (1) candidiasis and traumatized genital warts—both readily recognized; (2) lesions due to genital involvement by more widespread dermatoses; (3)

cutaneous manifestations of systemic diseases such as genital mucosal ulceration in Stevens-Johnson syndrome or Behçt's disease; (4) superinfections of lesions that may originally have been sexually acquired such as methicillin-resistant *S. aureus* complicating a genital ulcer due to HSV-2; and (5) localized drug reactions, such as the ulcers occasionally seen with topical paromomycin cream or boric acid preparations.

Figure 130-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chancroid: multiple, painful, punched-out ulcers with -undermined borders on the labia occurring after autoinoculation.

Figure 130-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Genital herpes. A relatively mild, superficial ulcer is typically seen in episodicoutbreaks. *(Courtesy of Michael Remington, University of Washington Virology Research Clinic.)*

Figure 130-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lymphogranuloma venereum: striking tender lymphadenopathy occurring at the femoral and inguinal lymph nodes, separated by a groove made by Poupart's ligament. This "sign-of-the-groove" is not considered specific for LGV; for example, lymphomas may present with this sign.

DIAGNOSIS

Although most genital ulcerations cannot be diagnosed confidently on clinical grounds alone, clinical findings and epidemiologic considerations (Table 130–7) can usually guide initial management (Table 130–8) pending results of further tests. Clinicians should order a rapid serologic test for syphilis in all cases of genital ulcer. To evaluate lesions except those highly characteristic of infection with HSV (i.e., those with herpetic vesicles), dark-field microscopy, direct immunofluorescence, and PCR for *T. pallidum* can be useful but are rarely available today in the United States. It is important to note that 30% of syphilitic chancres—the primary ulcer of syphilis—are associated with a nonreactive syphilis serology. All patients presenting with genital ulceration should be counseled and tested for HIV infection.

Table 130–7 Clinical Features of Genital Ulcers

Feature	Syphilis	Herpes	Chancroid	Lymphogranuloma Venereum	Donovanosis
Incubation period	9–90 days	2–7 days	1–14 days	3 days–6 weeks	1–4 weeks (up to 6 months)
Early primary lesions	Papule	Vesicle	Pustule	Papule, pustule, or vesicle	Papule
No. of lesions	Usually one	Multiple	Usually multiple, may coalesce	Usually one; often not detected, despite lymphadenopathy	Variable
Diameter	5–15 mm	1–2 mm	Variable	2–10 mm	Variable
Edges	Sharply demarcated, elevated, round, or oval	Erythematous	Undermined, ragged, irregular	Elevated, round, or oval	Elevated, irregular
Depth	Superficial or deep	Superficial	Excavated	Superficial or deep	Elevated
Base	Smooth, nonpurulent, relatively nonvascular	Serous, erythematous, nonvascular	Purulent, bleeds easily	Variable, nonvascular	Red and velvety, bleeds readily
Induration	Firm	None	Soft	Occasionally firm	Firm
Pain	Uncommon	Frequently tender	Usually very tender	Variable	Uncommon

Lymphadenopathy	Firm, nontender, bilateral	Firm, tender, often bilateral with initial episode	Tender, may suppurate, loculated, usually unilateral	Tender, may suppurate, loculated, usually unilateral	None; pseudobuboes

Source: From RM Ballard, in KK Holmes et al (eds): *Sexually Transmitted Diseases*, 4th ed. New York, McGraw-Hill, 2008.

Table 130–8 Initial Management of Genital or Perianal Ulcer

Usual causes
Herpes simplex virus (HSV)
Treponema pallidum(primary syphilis)
Haemophilus ducreyi (chancroid)
Usual initial laboratory evaluation
Dark-field exam (if available), direct FA, or PCR for *T. pallidum;* RPR, VDRL, or EIA test for syphilis (if negative but primary syphilis suspected, treat presumptively when indicated by epidemiologic and sexual risk assessment; repeat in 1 week); culture, direct FA, ELISA, or PCR for HSV; consider HSV-2-specific serology. In chancroid-endemic area: PCR or culture for *H. ducreyi*
Initial Treatment
Herpes confirmed or suspected (history or sign of vesicles):
Treat for genital herpes with acyclovir, valacyclovir, or famciclovir
Syphilis confirmed (dark-field, FA, or PCR showing *T. pallidum*, or RPR reactive):
Benzathine penicillin [2.4 million units IM once to patient, recent (e.g., within 3 months) seronegative partner(s), and all seropositive partners]
Chancroid confirmed or suspected (diagnostic test positive, or HSV and syphilis excluded, and persistent lesion):
Ciprofloxacin (500 mg PO as single dose) *or*
Ceftriaxone (250 mg IM as single dose) *or*
Azithromycin (1 g PO as single dose)

Abbreviations: FA, fluorescent antibody; PCR, polymerase chain reaction; RPR, rapid plasma reagin; EIA, enzyme immunoassay; ELISA, enzyme-linked immunosorbent assay; HSV, herpes simplex virus; VDRL, Venereal Disease Research Laboratory.

Typical vesicles or pustules or a cluster of painful ulcers preceded by vesiculopustular lesions suggests genital herpes. These typical clinical manifestations make detection of the virus optional; however, many patients want confirmation of the diagnosis, and differentiation of HSV-1 from HSV-2 has prognostic implications, since the latter causes more frequent genital recurrences.

Painless, nontender, indurated ulcers with firm, nontender inguinal adenopathy suggest primary syphilis. If results of dark-field examination and a rapid serologic test for syphilis are initially negative, presumptive therapy should be provided on the basis of the individual's risk. For example, with increasing rates of syphilis among MSM in the United States, most experts would not withhold therapy for this infection pending watchful waiting and/or subsequent detection of seroconversion. Repeated serologic testing for syphilis 1 or 2 weeks after treatment of seronegative primary syphilis usually demonstrates seroconversion.

"Atypical" or clinically trivial ulcers may be more common manifestations of genital herpes than classic vesiculopustular lesions. Specific tests for HSV in such lesions are therefore indicated (Chap. 179). Commercially available type-specific serologic tests for serum antibody to HSV-2 may give negative results, especially when patients present early with the initial episode of genital herpes or when HSV-1 is the cause of genital herpes (as is often the case today). Furthermore, a positive test for antibody to HSV-2 does not prove that the current lesions are herpetic, since nearly one-fifth of the general population of the United States (and no doubt a higher proportion of those at risk for other STIs) becomes seropositive for HSV-2 during early adulthood. Although even "type-specific" tests for HSV-2 that are commercially available in the United States are not 100% specific, a positive HSV-2 serology does enable the clinician to tell the patient that he or she has probably had genital herpes, should learn to recognize symptoms, should avoid sex during recurrences, and should consider use of condoms or suppressive antiviral therapy, both of which can reduce the risk of transmission to a sexual partner.

Demonstration of *H. ducreyi* by culture (or by PCR, when available) is most useful when ulcers are painful and purulent, especially if inguinal lymphadenopathy with fluctuance or overlying erythema is noted; if chancroid is prevalent in the community; or if the patient has recently had a sexual exposure

elsewhere in a chancroid-endemic area (e.g., a developing country). Enlarged, fluctuant lymph nodes should be aspirated for culture or PCR to detect *H. ducreyi* as well as for Gram's staining and culture to rule out the presence of other pyogenic bacteria.

When genital ulcers persist beyond the natural history of initial episodes of herpes (2–3 weeks) or of chancroid or syphilis (up to 6 weeks) and do not resolve with syndrome-based antimicrobial therapy, then—in addition to the usual tests for herpes, syphilis, and chancroid—biopsy is indicated to exclude donovanosis, carcinoma, and other nonvenereal dermatoses. If not performed previously, HIV serology should be standard, since chronic, persistent genital herpes is common in AIDS.

TREATMENT: ULCERATIVE GENITAL OR PERIANAL LESIONS

Immediate syndrome-based treatment for acute genital ulcerations (after collection of all necessary diagnostic specimens at the first visit) is often appropriate before all test results become available, because patients with typical initial or recurrent episodes of genital or anorectal herpes can benefit from prompt oral antiviral therapy (Chap. 179); because early treatment of sexually transmitted causes of genital ulcers decreases further transmission; and because many patients do not return for test results and treatment. A thorough assessment of the patient's sexual-risk profile and medical history is critical in determining the course of initial management. The patient who has risk factors consistent with exposure to syphilis (e.g., a male patient who reports sex with other men or who has HIV infection) should generally receive initial treatment for syphilis. Empirical therapy for chancroid should be considered if there has been an exposure in an area of the world where chancroid occurs or if regional lymph node suppuration is evident. In resource-poor settings lacking ready access to diagnostic tests, this approach to syndromic treatment for syphilis and chancroid has helped bring these two diseases under control. Finally, empirical antimicrobial therapy may be indicated if ulcers persist and the diagnosis remains unclear after a week of observation despite attempts to diagnose herpes, syphilis, and chancroid.

Proctitis, Proctocolitis, Enterocolitis, and Enteritis

Sexually acquired *proctitis*, with inflammation limited to the rectal mucosa (the distal 10–12 cm), results from direct rectal inoculation of typical STD pathogens. In contrast, inflammation extending from the rectum to the colon (*proctocolitis*), involving both the small and the large bowel (*enterocolitis*), or involving the small bowel alone (*enteritis*) can result from ingestion of typical intestinal pathogens through oral-anal exposure during sexual contact. Anorectal pain and mucopurulent, bloody rectal discharge suggest proctitis or protocolitis. Proctitis commonly produces tenesmus (causing frequent attempts to defecate, but not true diarrhea) and constipation, whereas proctocolitis and enterocolitis more often cause true diarrhea. In all three conditions, anoscopy usually shows mucosal exudate and easily induced mucosal bleeding (i.e., a positive "wipe test"), sometimes with petechiae or mucosal ulcers. Exudate should be sampled for Gram's staining and other microbiologic studies. Sigmoidoscopy or colonoscopy shows inflammation limited to the rectum in proctitis or disease extending at least up into the sigmoid colon in proctocolitis.

The AIDS era brought an extraordinary shift in the clinical and etiologic spectrum of intestinal infections among MSM. The number of cases of the acute intestinal STIs described above fell as high-risk sexual behaviors became less common in this group. At the same time, the number of AIDS-related opportunistic intestinal infections increased rapidly, many associated with chronic or recurrent symptoms. The incidence of these infections has since fallen with increasingly effective antiretroviral therapy. Two species initially isolated in association with intestinal symptoms in MSM are now known as *Helicobacter cinaedi* and *Helicobacter fennelliae*, and both have subsequently been isolated from the blood of HIV-infected men and other immunosuppressed persons, often in association with a syndrome of multifocal dermatitis and arthritis.

Acquisition of HSV, *N. gonorrhoeae*, or *C. trachomatis* (including LGV strains of *C. trachomatis*) during receptive anorectal intercourse causes most cases of infectious proctitis in women and MSM. Primary and secondary syphilis can also produce anal or anorectal lesions, with or without symptoms. Gonococcal or chlamydial proctitis typically involves the most distal rectal mucosa and the anal crypts and is clinically mild, without systemic manifestations. In contrast, primary proctitis due to HSV and proctocolitis due to the strains of *C. trachomatis* that cause LGV usually produce severe anorectal pain and often cause fever. Perianal ulcers and inguinal lymphadenopathy, most commonly due to HSV, can also occur in LGV or syphilis. Sacral nerve root radiculopathies, usually presenting as urinary retention, laxity of the anal sphincter, or constipation, may complicate primary herpetic proctitis. In LGV, rectal biopsy typically shows crypt abscesses, granulomas, and giant cells—findings resembling those in Crohn's disease; such findings should always prompt rectal culture and serology for LGV, which is a curable infection. Syphilis can also produce rectal granulomas, usually in association with infiltration by plasma cells or other mononuclear cells. Syphilis, LGV, and HSV infection involving the rectum can produce perirectal adenopathy that is sometimes mistaken for malignancy; syphilis, LGV, HSV infection, and chancroid involving the anus can produce inguinal adenopathy, because anal lymphatics drain to inguinal lymph nodes.

Diarrhea and abdominal bloating or cramping pain without anorectal symptoms and with normal

findings on anoscopy and sigmoidoscopy occur with inflammation of the small intestine (enteritis) or with proximal colitis. In MSM without HIV infection, enteritis is often attributable to *Giardia lamblia*. Sexually acquired proctocolitis is most often due to *Campylobacter* or *Shigella* species.

TREATMENT: PROCTITIS, PROCTOCOLITIS, ENTEROCOLITIS, AND ENTERITIS

Acute proctitis in persons who have practiced receptive anorectal intercourse is usually sexually acquired. Such patients should undergo anoscopy to detect rectal ulcers or vesicles and petechiae after swabbing of the rectal mucosa; to examine rectal exudates for PMNs and gram-negative diplococci; and to obtain rectal swab specimens for testing for rectal gonorrhea, chlamydial infection, herpes, and syphilis. Pending test results, patients with proctitis should receive empirical syndromic treatment—e.g., with ceftriaxone (a single IM dose of 125 mg for gonorrhea) plus doxycycline (100 mg by mouth twice daily for 7 days for possible chlamydial infection) plus treatment for herpes or syphilis if indicated.

PREVENTION AND CONTROL OF STIS

Prevention and control of STIs require the following:

1. Reduction of the average rate of sexual exposure to STIs through alteration of sexual risk behaviors and behavioral norms among both susceptible and infected persons in all population groups. The necessary changes include reduction in the total number of sexual partners and the number of concurrent sexual partners.
2. Reduction of the efficiency of transmission through the promotion of safer sexual practices, the use of condoms during casual or commercial sex, vaccination against HBV and HPV infection, male circumcision (which reduces risk of acquisition of HIV, chancroid, and perhaps other STIs), and a growing number of other approaches (e.g., early detection and treatment of other STIs to reduce the efficiency of sexual transmission of HIV). Longitudinal studies have shown that consistent condom use is associated with significant protection of both males and females against all STIs that have been examined, including HIV, HPV, and HSV infections as well as gonorrhea and chlamydial infection. The only exceptions are probably sexually transmitted *Pthirus pubis* and *Sarcoptes scabiei* infestations.
3. Shortening of the duration of infectivity of STIs through early detection and curative or suppressive treatment of patients and their sexual partners.

Financial and time constraints imposed by many clinical practices, along with the reluctance of some clinicians to ask questions about stigmatized sexual behaviors, often curtail screening and prevention services. As outlined in Fig. 130-8, the success of clinicians' efforts to detect and treat STIs depends in part on societal efforts to teach young people how to recognize symptoms of STIs; to motivate individuals with symptoms to seek care promptly; to educate persons who are at risk but have no symptoms about what tests they should undergo routinely; and to make high-quality, appropriate care accessible, affordable, and acceptable, especially to the young indigent patients most likely to acquire an STI.

Figure 130-8

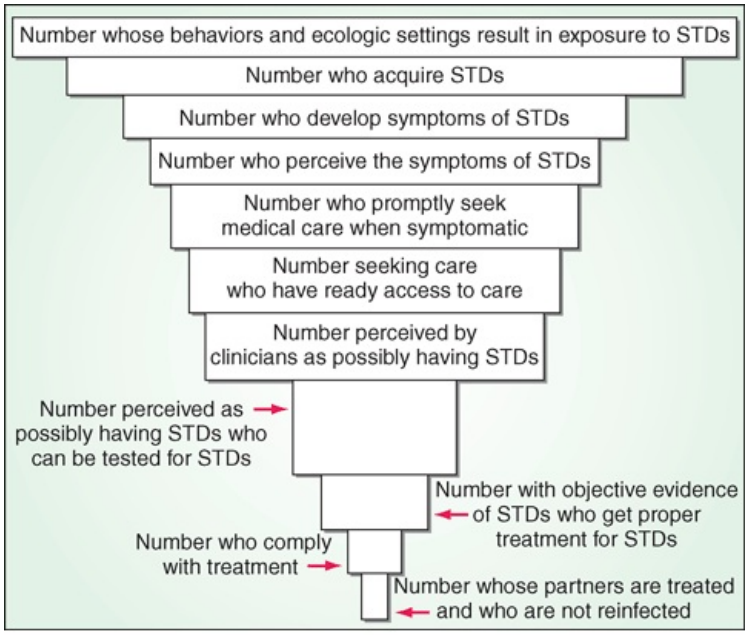

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Critical control points for preventive and clinical interventions against sexually transmitted diseases (STDs). [*Adapted from HT Waller and MA Piot: Bull World Health Organ 41:75, 1969 and 43:1, 1970; and from "Resource allocation model for public health planning—a case study of tuberculosis control," Bull World Health Organ 48 (Suppl), 1973.*]

Since many infected individuals develop no symptoms or fail to recognize and report symptoms, clinicians should routinely perform an STI risk assessment for teenagers and young adults as a guide to selective screening. As stated earlier, U.S. Preventive Services Task Force Guidelines recommend screening sexually active female patients ≤25 years of age for *C. trachomatis* whenever they present for health care (at least once a year); older women should be tested if they have more than one sexual partner, have begun a new sexual relationship since the previous test, or have another STI diagnosed. In women 25–29 years of age, chlamydial infection is uncommon but still may reach a prevalence of 3–5% in some settings; information on a sex partner's concurrency provided by women in this age group (i.e., whether a male partner has had another sex partner during the time they have been together) is helpful in identifying women at increased risk. In the United States, widespread selective screening of young women for cervical *C. trachomatis* infection in some regions has been associated with a 50–60% drop in prevalence, and such screening also protects the individual woman from PID. Sensitive urine-based genetic amplification tests permit expansion of screening to men, teenage boys, and girls in settings where examination is not planned or is impractical (e.g., during pre-participation sports examinations or during initial medical evaluation of adolescent girls). Vaginal swabs—collected either by the health care provider at a pelvic examination or by the woman herself—are highly sensitive and specific for the diagnosis of chlamydial and gonococcal infection; they are now the preferred type of specimen for screening and diagnosis of these infections.

Although gonorrhea is now substantially less common than chlamydial infection in industrialized countries, screening tests for *N. gonorrhoeae* are still appropriate for women and teenage girls attending STD clinics and for sexually active teens and young women from areas of high gonorrhea prevalence. Multiplex NAATs that combine screening for *N. gonorrhoeae* and *C. trachomatis* in a single low-cost assay now facilitate the prevention and control of both infections in populations at high risk.

All patients with newly detected STIs or at high risk for STIs according to routine risk assessment as well as all pregnant women should be encouraged to undergo serologic testing for syphilis and HIV infection, with appropriate HIV counseling before and after testing. Randomized trials have shown that risk-reduction counseling of patients with STIs significantly lowers subsequent risk of acquiring an STI; such counseling should now be considered a standard component of STI management. Preimmunization serologic testing for antibody to HBV is indicated for unvaccinated persons who are known to be at high risk, such as homosexually active men and injection drug users. In most young persons, however, it is more cost-effective to vaccinate against HBV without serologic screening. It is important to recognize that, while immunization against HBV has contributed to marked reductions in the incidence of infection with this virus, the majority of new cases that do occur are acquired through sex. In 2006, the Advisory Committee on Immunization Practices (ACIP) of the CDC recommended the following: (1) Universal hepatitis B vaccination should be implemented for all unvaccinated adults in settings in which a high proportion of adults have risk factors for HBV infection (e.g., STD clinics, HIV testing and treatment facilities, drug-abuse treatment and prevention settings, health care settings

targeting services to injection drug users or MSM, and correctional facilities). (2) In other primary care and specialty medical settings in which adults at risk for HBV infection receive care, health care providers should inform all patients about the health benefits of vaccination, the risk factors for HBV infection, and the persons for whom vaccination is recommended and should vaccinate adults who report risk factors for HBV infection as well as any adult who requests protection from HBV infection. To promote vaccination in all settings, health care providers should implement standing orders to identify adults recommended for hepatitis B vaccination, should administer hepatitis B vaccine as part of routine clinical services, should not require acknowledgment of an HBV infection risk factor for adult vaccination, and should use available reimbursement mechanisms to remove financial barriers to hepatitis B vaccination.

In 2007, the ACIP recommended routine immunization of 9- to 26-year-old girls and women with the quadrivalent HPV vaccine (against HPV types 6, 11, 16, and 18) approved by the U.S. Food and Drug Administration; the optimal age for recommended vaccination is 11–12 years because of the very high risk of HPV infection after sexual debut. In 2009, the ACIP added bivalent HPV vaccine (against types 6 and 11) as an option and expanded the groups in which immunization (with either quadrivalent or bivalent vaccine) is safe and effective to include boys and men 9–26 years old. HPV vaccines offering broader protection against additional oncogenic HPV types are anticipated.

Partner notification is the process of identifying and informing partners of infected patients about possible exposure to an STI and of examining, testing, and treating partners as appropriate. In a series of 22 reports concerning partner notification during the 1990s, index patients with gonorrhea or chlamydial infection named a mean of 0.75–1.6 partners, of whom one-fourth to one-third were infected; those with syphilis named 1.8–6.3 partners, with one-third to one-half infected; and those with HIV infection named 0.76–5.31 partners, with up to one-fourth infected. Persons who transmit infection or who have recently been infected and are still in the incubation period usually have no symptoms or only mild symptoms and seek medical attention only when notified of their exposure. Therefore, the clinician must encourage patients to participate in partner notification, must ensure that exposed persons are notified and treated, and must guarantee confidentiality to all involved. In the United States, local health departments often offer assistance in partner notification, treatment, and/or counseling. It seems both feasible and most useful to notify those partners exposed within the patient's likely period of infectiousness that is often considered the preceding 1 month for gonorrhea, 1–2 months for chlamydial infection, and up to 3 months for early syphilis.

Persons with a new-onset STI always have a *source* contact who gave them the infection; in addition, they may have a *secondary* (*spread* or *exposed*) contact with whom they had sex after becoming infected. The identification and treatment of these two types of contacts have different objectives. Treatment of the source contact (often a casual contact) benefits the community by preventing further transmission; treatment of the recently exposed secondary contact (typically a spouse or another steady sexual partner) prevents both the development of serious complications (such as PID) in the partner and reinfection of the index patient. A survey of a random sample of U.S. physicians found that most instructed patients to abstain from sex during treatment, to use condoms, and to inform their sex partners after being diagnosed with gonorrhea, chlamydial infection, or syphilis; physicians sometimes gave the patients drugs for their partners. However, follow-up of the partners by physicians was infrequent. A randomized trial compared patients' delivery of therapy to partners exposed to gonorrhea or chlamydial infection with conventional notification and advice to partners to seek evaluation for STD; patients' delivery of partners' therapy (PDPT), also known as *expedited partner therapy* (EPT), significantly reduced combined rates of reinfection of the index patient with *N. gonorrhoeae* or *C. trachomatis*. State-by-state variations in regulations governing this approach have not been well defined, but the 2010 CDC STD treatment guidelines and the EPT final report of 2006 (*http://www.cdc.gov/std/treatment/EPTFinalReport2006.pdf*) describe its potential use. Currently, EPT is commonly used by many practicing physicians. Its legal status varies by state, but EPT is now permissible in 22 states and potentially allowable in another 20. (Updated information on the legal status of EPT is available at *http://www.cdc.gov/std/ept*.)

In summary, clinicians and public health agencies share responsibility for the prevention and control of STIs. In the current health care environment, the role of primary care clinicians has become increasingly important in STI prevention as well as in diagnosis and treatment, and the resurgence of bacterial STIs like syphilis and LGV among MSM—particularly those co-infected with HIV—emphasizes the need for risk assessment and routine screening.

FURTHER READINGS

Celum CL et al: Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N Engl J Med 362:427, 2010[PMID: 20089951]

Centers for Disease Control and Prevention: Sexually transmitted diseases treatment guidelines, 2010. MMWR Recomm Rep 59 (RR-12):1, 2010. Available at *http://www.cdc.gov/std/treatment*

Fredricks DN et al: Molecular identification of bacteria associated with bacterial vaginosis. N Engl J

Med 353:1899, 2005[PMID: 16267321]

Holmes KK et al (eds): *Sexually Transmitted Diseases*, 4th ed. New York, McGraw-Hill, 2008

Manhart LE, Holmes KK: Randomized controlled trials of individual-level, population-level, and multilevel interventions for preventing sexually transmitted infections: What has worked? J Infect Dis 191(Suppl 1):S7, 2005

Markowitz LE et al: Quadrivalent human papillomavirus vaccine: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 56(RR-2):1, 2007. Updated information available at *http://www.cdc.gov/vaccines/recs/acip/*

——— et al: Seroprevalence of human papillomavirus vaccine types 6, 11, 16, and 18 in the United States: National Health and Nutrition Examination Survey 2003–2004. J Infect Dis 200:1059, 2009

Mast EE et al: A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP). Part II: Immunization of adults. MMWR Recomm Rep 55(RR-16):1, 2006

Morse SA et al: *Atlas of Sexually Transmitted Diseases* , 4th ed. Elsevier, London, 2010

U.S. Preventive Services Task Force: Behavioral counseling to prevent sexually transmitted infections: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 149:491–6, W95, 2008

World Health Organization: Sexually transmitted diseases diagnostics initiative. Geneva, WHO, 2001 (*http://www.who.int/std_diagnostics/*)

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

Infectious Complications of Burns: Introduction

The skin is an essential component of the nonspecific immune system, protecting the host from potential pathogens in the environment. Breaches in this protective barrier thus represent a form of immunocompromise that predisposes a patient to infection. Thermal burns may cause massive destruction of the integument as well as derangements in humoral and cellular immunity, permitting the development of infection caused by environmental opportunists and components of the host's skin flora.

Epidemiology

Over the last decade, the estimated incidence of burn injuries in the United States has declined steadily; however, >1 million burn injuries are brought to medical attention each year. Although many burn injuries are minor and require little or no intervention, 50,000 persons are hospitalized for these injuries, 60% of whom require intensive care in a specialized burn center and 20,000 of whom have major burns involving at least 25% of the total body surface area. The majority of burn patients are men. Infants account for ~10% of all reported cases. Scalds, structural fires, and flammable liquids and gases are the major causes of burns, but electrical, chemical, and smoking-related sources are also important. Burns predispose to infection by damaging the protective barrier function of the skin, thus facilitating the entry of pathogenic microorganisms, and by inducing systemic immunosuppression. It is therefore unsurprising that multiorgan failure and infectious complications are the major causes of morbidity and death in serious burn injuries. As many as 10,000 patients in the United States die of burn-related infections each year, and 6 of the top 10 complications recently identified by the American Burn Association's 10-year review are infectious. These 10 most common complications are pneumonia (4.6%), septicemia (2.7%), cellulitis/traumatic injury (2.6%), respiratory failure (2.5%), wound infection (2.2%), other infection (2.0%), renal failure (1.5%), line infection (1.4%), acute respiratory distress syndrome (1.2%), and arrhythmia (1.0) (www.ameriburn.org/2007NBRAnnualReport.pdf).

Pathophysiology

Loss of the cutaneous barrier facilitates entry of the patient's own flora and of organisms from the hospital environment into a burn wound. Initially, the wound is colonized with gram-positive bacteria from the surrounding tissue, but the number of bacteria grows rapidly beneath the burn eschar, reaching ~8.4 x 10^3 cfu/g on day 4 after the burn. The avascularity of the eschar, along with the impairment of local immune responses, favors further bacterial colonization and proliferation. By day 7, the wound is colonized with other microbes, including gram-positive bacteria, gram-negative bacteria, and yeasts derived from the gastrointestinal and upper respiratory flora. Invasive infection'localized and/or systemic—occurs when these bacteria penetrate viable tissue. In addition, a role for biofilms has been recognized in experimental animal models of burn-wound infection. (Biofilms are surface-associated communities of bacteria, often embedded in a matrix, that allow the microbes to persist and to resist the effects of host immunity and antimicrobial agents.)

Streptococci and staphylococci were the predominant causes of burn-wound infection in the preantibiotic era and are still important pathogens. With the advent of antimicrobial agents, *Pseudomonas aeruginosa* became a major problem in burn-wound management. Less common anaerobic bacteria typically are found in infections from electrical burns or when open wound dressings are used. The widespread use of topical and more effective antimicrobial drugs has resulted in a decline in bacterial wound infections and the emergence of fungi (particularly *Candida albicans*, *Aspergillus* species, and the agents of mucormycosis) as increasingly important pathogens in burn-wound patients. Herpes simplex virus has been found in burn wounds, especially those on the neck and face and those associated with inhalation injury. Autopsy reports on patients with severe thermal burns over the last decade have identified an association of *P. aeruginosa*, *Escherichia coli*, *Klebsiella pneumoniae*, and *Staphylococcus aureus* with death; this association is independent of the percentage of the total body surface area covered by burns, the percentage of burns that are full-thickness (as opposed to partial-thickness), inhalation injury, and day of death after a burn. In addition, members of the *Acinetobacter calcoaceticus-baumannii* complex are among the most common pathogens at some burn centers.

The cascade of events that follows a severe burn injury and that leads to multiorgan system failure and death is thought to represent a two-step process; i.e., the burn injury itself, with ensuing

hypovolemia and tissue hypoxia, is followed by invasive infection arising from large amounts of devitalized tissue. The frequency of infection parallels the extent and severity of the burn injury. Severe burn injuries cause a state of immunosuppression that affects innate and adaptive immune responses. The substantial impact of immunocompromise on infection is due to effects on both the cellular and the humoral arms of the immune system. For example, decreases in the number and activity of circulating helper T cells, increases in the number and activity of suppressor T cells, decreases in production and release of monocytes and macrophages, and diminution in levels of immunoglobulin follow major burns. Neutrophil and complement functions also have been shown to be impaired after burns. The increased levels of multiple cytokines detected in burn patients are compatible with the widely held belief that the inflammatory response becomes dysregulated in these individuals; bacterial cell products play a potent role in inducing proinflammatory mediators that contribute to this uncontrolled systemic inflammatory response. Increased permeability of the gut wall to bacteria and their components (e.g., endotoxin) also contributes to immune dysregulation and sepsis. Thus, a burn patient is predisposed to infection at remote sites (see below) as well as at the sites of burn injury. Another contributor to secondary immunosuppression after burn injuries is the endocrine system; increasing levels of vasopressin, aldosterone, cortisol, glucagon, growth hormone, catecholamines, and other hormones that directly affect lymphocyte proliferation, secretion of proinflammatory cytokines, natural killer cell activity, and suppressor T cells are seen.

Clinical Manifestations

Since clinical indications of wound infection are difficult to interpret, wounds must be monitored carefully for changes that may reflect infection. A margin of erythema frequently surrounds the sites of burns and by itself is not usually indicative of infection. Signs of infection include the conversion of a partial-thickness to a full-thickness burn, color changes (e.g., the appearance of a dark brown or black discoloration of the wound), the new appearance of erythema or violaceous edema in normal tissue at the wound margins, the sudden separation of the eschar from subcutaneous tissues, and the degeneration of the wound with the appearance of a new eschar.

Early surgical excision of devitalized tissue is now widely performed, and burn-wound infections can be classified in relation to the excision site as (1) burn-wound impetigo (infection characterized by loss of epithelium from a previously reepithelialized surface, as seen in a partial-thickness burn that is allowed to close by secondary intention, a grafted burn, or a healed skin donor site), (2) burn-related surgical wound infection (purulent infection of excised burn and donor sites that have not yet epithelialized, accompanied by positive cultures), (3) burn-wound cellulitis (extension of infection to surrounding healthy tissue; Fig. e23-1), and (4) invasive infection in unexcised burn wounds (infection that is secondary to a partial- or full-thickness burn wound and is manifested by separation of the eschar or by violaceous, dark brown, or black discoloration of the eschar; Fig. e23-2). The appearance of a green discoloration of the wound or subcutaneous fat (Fig. e23-3) or the development of ecthyma gangrenosum (see Fig. e7-35) at a remote site points to a diagnosis of invasive *P. aeruginosa* infection.

Figure e23-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cellulitis complicating a burn wound of the arm and demonstrating extension of the infection to adjacent healthy tissue. *(Courtesy of Dr. Robert L. Sheridan, Massachusetts General Hospital, Boston; with permission.)*

Figure e23-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A severe upper-extremity burn infected with *Pseudomonas aeruginosa*. The wound requires additional debridement. Note the dark brown to black discoloration of the eschar. *(Courtesy of Dr. Robert L. Sheridan, Massachusetts General Hospital, Boston; with permission.)*

Figure e23-3

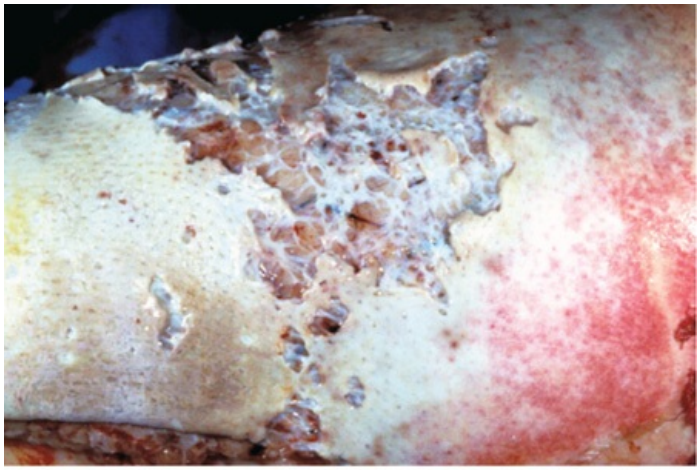

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A burn wound infected with *Pseudomonas aeruginosa*, with liquefaction of tissue. Note the green discoloration at the wound margins, which is suggestive of *Pseudomonas* infection. *(Courtesy of Dr. Robert L. Sheridan, Massachusetts General Hospital, Boston; with permission.)*

Changes in body temperature, hypotension, tachycardia, altered mentation, neutropenia or neutrophilia, thrombocytopenia, and renal failure may result from invasive burn wounds and sepsis. However, because profound alterations in homeostasis occur as a consequence of burns per se and because inflammation without infection is a normal component of these injuries, the assessment of these changes is complicated. Alterations in body temperature, for example, are attributable to thermoregulatory dysfunction; tachycardia and hyperventilation accompany the metabolic changes induced by extensive burn injury and are not necessarily indicative of bacterial sepsis.

In light of the difficulty of evaluating burn wounds solely on the basis of clinical observation and laboratory data, wound biopsies are necessary for definitive diagnosis of infection. The timing of these biopsies can be guided by clinical changes, but in some centers burn wounds are biopsied routinely at regular intervals. The biopsy specimen is examined for histologic evidence of bacterial invasion, and quantitative microbiologic cultures are performed. The presence of $>10^5$ viable bacteria per gram of tissue is highly suggestive of invasive infection and of a dramatically increased risk of sepsis. Histopathologic evidence of invasion of viable tissue and the presence of microorganisms in unburned

blood vessels and lymphatics are more definitive indicators of infection. A blood culture positive for the same organism seen in large quantities in biopsied tissue is a reliable indicator of burn sepsis. Surface cultures may provide some indication of the microorganisms present in the hospital environment but are not indicative of the etiology of infection. This noninvasive technique might be of use in determining the flora present in excised burn areas or in areas where the skin is too thin for biopsy (e.g., over the ears, eyes, or digits).

In addition to infection of the burn wound itself, a number of other infections due to the immunosuppression caused by extensive burns and the manipulations necessary for clinical care put burn patients at risk. Pneumonia, now the most common infectious complication among hospitalized burn patients, is most often nosocomially acquired via the respiratory route; among the risk factors associated with secondary pneumonia are inhalation injury, intubation, full-thickness chest wall burns, immobility, and uncontrolled wound sepsis with hematogenous spread. Septic pulmonary emboli also may occur. Suppurative thrombophlebitis may complicate the vascular catheterization necessary for fluid and nutritional support in burns. Endocarditis, urinary tract infection, bacterial chondritis (particularly in patients with burned ears), and intraabdominal infection also complicate serious burn injury.

Treatment: Burn Wound Infections

The ultimate goal of burn-wound management is closure and healing of the wound. Early surgical excision of burned tissue, with extensive debridement of necrotic tissue and grafting of skin or skin substitutes, greatly decreases mortality rates associated with severe burns. In addition, the four widely used topical antimicrobial agents—silver sulfadiazine cream, mafenide acetate cream, silver nitrate cream, and nanocrystalline silver dressings—dramatically decrease the bacterial burden of burn wounds and reduce the incidence of burn-wound infection; these agents are applied routinely to partial- and full-thickness burns. The bactericidal properties of silver are related to its effect on respiratory enzymes on bacterial cell walls; its interaction with structural proteins causes keratinocyte and fibroblast toxicity that can delay wound healing if silver-based compounds are used indiscriminately. All four agents are broadly active against many bacteria and some fungi and are useful before bacterial colonization is established. Silver sulfadiazine often is used initially, but its value can be limited by bacterial resistance, poor wound penetration, or toxicity (leukopenia). Mafenide acetate has broader activity against gram-negative bacteria. The cream penetrates eschars and thus can prevent or treat infection beneath them; its use without dressings allows regular examination of the wound area. The foremost disadvantages of mafenide acetate are that it can inhibit carbonic anhydrase, resulting in metabolic acidosis, and that it elicits hypersensitivity reactions in up to 7% of patients. This agent is used most often when gram-negative bacteria invade the burn wound and when treatment with silver sulfadiazine fails. The activity of mafenide acetate against gram-positive bacteria is limited. Nanocrystalline silver dressings provide broader antimicrobial coverage than does any other available topical preparation, exhibiting activity against methicillin-resistant *S. aureus* (MRSA) and vancomycin-resistant enterococci, moderate ability to penetrate eschars, and limited toxicity. In addition, this approach provides controlled and prolonged release of nanocrystalline silver into the wound, limiting the number of dressing changes and therefore reducing the risk of nosocomial infections as well as the cost of treatment. Mupirocin, a topical antimicrobial agent used to eradicate nasal colonization with MRSA, is being used increasingly in burn units where MRSA is prevalent. The efficacy of mupirocin in reducing burn-wound bacterial counts and preventing systemic infections is comparable to that of silver sulfadiazine.

In recent years, rates of fungal infection have increased in burn patients. When superficial fungal infection occurs, nystatin may be mixed with silver sulfadiazine or mafenide acetate as topical therapy. A small study found that nystatin powder (6 million units/g) was effective for treatment of superficial and deep burn-wound infections caused by *Aspergillus* or *Fusarium* species. In addition to these products, moisture-retention ointments with antimicrobial properties can promote rapid autolysis, debridement, and moist healing of partial-thickness wounds.

When invasive wound infection is diagnosed, topical therapy should be changed to mafenide acetate. Subeschar clysis (the direct instillation of an antibiotic, often piperacillin, into wound tissues under the eschar) is a useful adjunct to surgical and systemic antimicrobial therapy. Systemic treatment with antibiotics active against the pathogens present in the wound should be instituted. In the absence of culture data, treatment should be broad in spectrum, covering organisms commonly encountered in that particular burn unit. Such coverage usually is achieved with an antibiotic active against gram-positive pathogens (e.g., vancomycin, 1 g IV every 12 h) and with a drug active against *P. aeruginosa* and other gram-negative rods (e.g., ceftazidime, 2 g IV every 8 h). In penicillin-allergic patients, ciprofloxacin (400 mg IV every 12 h) may be substituted for ceftazidime. In settings where MRSA is not prevalent, oxacillin (2 g IV every 4 h) may be substituted for vancomycin. Patients with burn wounds frequently have alterations in metabolism and renal clearance mechanisms that mandate the monitoring of serum antibiotic levels; the levels achieved with standard doses are often subtherapeutic.

Treatment of infections caused by emerging resistant pathogens remains a challenge in the care of

burn patients. MRSA, resistant enterococci, multidrug-resistant gram-negative rods, and Enterobacteriaceae producing extended-spectrum β-lactamases have been associated with burn-wound infections and identified in burn-unit outbreaks. Strict infection-control practices (including microbiologic surveillance in burn units) and appropriate antimicrobial therapy remain important measures in reducing rates of infection due to resistant organisms.

In general, prophylactic systemic antibiotics have no role in the management of burn wounds and can, in fact, lead to colonization with resistant microorganisms. In some studies, antibiotic prophylaxis has been associated with increased secondary infections of the upper and lower respiratory tract and the urinary tract as well as with prolonged hospitalization. An exception involves cases requiring burn-wound manipulation. Since procedures such as debridement, excision, and grafting frequently result in bacteremia, prophylactic systemic antibiotics are administered at the time of wound manipulation; the specific agents used should be chosen on the basis of data obtained by wound culture or data on the hospital's resident flora.

The use of oral antibiotics for selective digestive decontamination (SDD) to decrease bacterial colonization and the risk of burn-wound infection is controversial and has not been widely adopted. In a randomized, double-blind, placebo-controlled trial in patients with burns involving >20% of the total body surface area, SDD was associated with reduced mortality rates in the burn intensive care unit and in the hospital and also with a reduced incidence of pneumonia. The effects of SDD on the normal anaerobic bowel flora must be taken into consideration before this approach is used.

All burn-injury patients should undergo tetanus booster immunization if they have completed primary immunization but have not received a booster dose in the last 5 years. Patients without prior immunization should receive tetanus immune globulin and undergo primary immunization.

Further Readings

D'Avignon LC et al: Contribution of bacterial and viral infections to attributable mortality in patients with severe burns: An autopsy series. Burns 36:773, 2010[PMID: 20074860]

De la Cal M et al: Survival benefit in critically ill burned patients receiving selective decontamination of the digestive tract: A randomized, placebo-controlled, double-blind trial. Ann Surg 241:424, 2005

Guggenheim M et al: Changes in bacterial isolates from burn wounds and their antibiograms: A 20-year study (1986–2005). Burns 35:553, 2009[PMID: 19167827]

Holmes GP et al: Guidelines for the prevention and treatment of B-virus infections in exposed persons: The B Virus Working Group. Clin Infect Dis 20:421, 1995[PMID: 7742451]

Kaye ET: Topical antibacterial agents. Infect Dis Clin North Am 14:321, 2000[PMID: 10829258]

McManus WF et al: Subeschar antibiotic infusion in the treatment of burn wound infection. J Trauma 20:1021, 1980[PMID: 7452744]

Murray CK: *Burns*, in G Mandell et al (eds). *Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases*, 7th ed. New York, Churchill Livingstone Elsevier, 2010, pp 3905–3909

Pruitt BJ et al: The changing epidemiology of infection in burn patients. World J Surg 16:57, 1992[PMID: 1290268]

Ressner RA et al: Outcomes of bacteremia in burn patients involved in combat operations overseas. J Am Coll Surg 206:439, 2008[PMID: 18308213]

Sheridan RL et al: Cutaneous herpetic infections complicating burns. Burns 26:621, 2000[PMID: 10925184]

Steer JA et al: Quantitative microbiology in the management of burn patients: I. Correlation between quantitative and qualitative burn wound biopsy culture and surface alginate swab culture. Burns 22:173, 1996[PMID: 8726252]

———: Quantitative microbiology in the management of burn patients: II. Relationship between bacterial counts obtained by burn wound biopsy culture and surface alginate swab culture, with clinical outcome following burn surgery and change of dressings. Burns 22:177, 1996

Youn YK et al: The role of mediators in the response to thermal injury. World J Surg 16:30, 1992[PMID: 1290263]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIOUS COMPLICATIONS OF BITES: INTRODUCTION

The skin is an essential component of the nonspecific immune system, protecting the host from potential pathogens in the environment. Breaches in this protective barrier thus represent a form of immunocompromise that predisposes the patient to infection. Bites and scratches from animals and humans allow the inoculation of microorganisms past the skin's protective barrier into deeper, susceptible host tissues.

Each year in the United States, millions of animal-bite wounds are sustained. The vast majority are inflicted by pet dogs and cats, which number >100 million; the annual incidence of dog and cat bites has been reported as 300 bites per 100,000 population. Other bite wounds are a consequence of encounters with animals in the wild or in occupational settings. While many of these wounds require minimal or no therapy, a significant number result in infection, which may be life-threatening. The microbiology of bite-wound infections in general reflects the oropharyngeal flora of the biting animal, although organisms from the soil, the skin of the animal and victim, and the animal's feces may also be involved.

Dog Bites

In the United States, dogs bite >4.7 million people each year and are responsible for 80% of all animal-bite wounds, an estimated 15–20% of which become infected. Each year, 800,000 Americans seek medical attention for dog bites; of those injured, 386,000 require treatment in an emergency department, with >1000 emergency department visits each day and about a dozen deaths per year. Most dog bites are provoked and are inflicted by the victim's pet or by a dog known to the victim. These bites frequently occur during efforts to break up a dogfight. Children are more likely than adults to sustain canine bites, with the highest incidence of 6 bites per 1000 population among boys 5–9 years old. Victims are more often male than female, and bites most often involve an upper extremity. Among children <4 years old, two-thirds of all these injuries involve the head or neck. Infection typically manifests 8–24 h after the bite as pain at the site of injury with cellulitis accompanied by purulent, sometimes foul-smelling discharge. Septic arthritis and osteomyelitis may develop if a canine tooth penetrates synovium or bone. Systemic manifestations (e.g., fever, lymphadenopathy, and lymphangitis) may also occur. The microbiology of dog-bite wound infections is usually mixed and includes β-hemolytic streptococci, *Pasteurella* species, *Staphylococcus* species [including methicillin-resistant *Staphylococcus aureus* (MRSA)], *Eikenella corrodens*, and *Capnocytophaga canimorsus*. Many wounds also include anaerobic bacteria such as *Actinomyces, Fusobacterium, Prevotella*, and *Porphyromonas* species.

While most infections resulting from dog-bite injuries are localized to the area of injury, many of the microorganisms involved are capable of causing systemic infection, including bacteremia, meningitis, brain abscess, endocarditis, and chorioamnionitis. These infections are particularly likely in hosts with edema or compromised lymphatic drainage in the involved extremity (e.g., after a bite on the arm in a woman who has undergone mastectomy) and in patients who are immunocompromised by medication or disease (e.g., glucocorticoid use, systemic lupus erythematosus, acute leukemia, or hepatic cirrhosis). In addition, dog bites and scratches may result in systemic illnesses such as rabies (Chap. 195) and tetanus (Chap. 140).

Infection with *C. canimorsus* following dog-bite wounds may result in fulminant sepsis, disseminated intravascular coagulation, and renal failure, particularly in hosts who have impaired hepatic function, who have undergone splenectomy, or who are immunosuppressed. This organism is a thin gram-negative rod that is difficult to culture on most solid media but grows in a variety of liquid media. The bacteria are occasionally seen within polymorphonuclear leukocytes on Wright-stained smears of peripheral blood from septic patients. Tularemia (Chap. 158) has also been reported to follow dog bites.

Cat Bites

Although less common than dog bites, cat bites and scratches result in infection in more than half of all cases. Because the cat's narrow, sharp canine teeth penetrate deeply into tissue, cat bites are more likely than dog bites to cause septic arthritis and osteomyelitis; the development of these conditions is particularly likely when punctures are located over or near a joint, especially in the hand. Women sustain cat bites more frequently than do men. These bites most often involve the hands and arms. Both bites and scratches from cats are prone to infection from organisms in the cat's oropharynx.

Pasteurella multocida, a normal component of the feline oral flora, is a small gram-negative coccobacillus implicated in the majority of cat-bite wound infections. Like that of dog-bite wound infections, however, the microflora of cat-bite wound infections is usually mixed. Other microorganisms causing infection after cat bites are similar to those causing dog-bite wound infections.

The same risk factors for systemic infection following dog-bite wounds apply to cat-bite wounds. *Pasteurella* infections tend to advance rapidly, often within hours, causing severe inflammation accompanied by purulent drainage; *Pasteurella* may also be spread by respiratory droplets from animals, resulting in pneumonia or bacteremia. Like dog-bite wounds, cat-bite wounds may result in the transmission of rabies or in the development of tetanus. Infection with *Bartonella henselae* causes cat-scratch disease (Chap. 160) and is an important late consequence of cat bites and scratches. Tularemia (Chap. 158) has also been reported to follow cat bites.

Other Animal Bites

Infections have been attributed to bites from many animal species. Often these bites are sustained as a consequence of occupational exposure (farmers, laboratory workers, veterinarians) or recreational exposure (hunters and trappers, wilderness campers, owners of exotic pets). Generally, the microflora of bite wounds reflects the oral flora of the biting animal. Most members of the cat family, including feral cats, harbor *P. multocida*. Bite wounds from aquatic animals such as alligators or piranhas may contain *Aeromonas hydrophila*. Venomous snakebites (Chap. 396) result in severe inflammatory responses and tissue necrosis—conditions that render these injuries prone to infection. The snake's oral flora includes many species of aerobes and anaerobes, such as *Pseudomonas aeruginosa*, *Proteus* species, *Staphylococcus epidermidis*, *Bacteroides fragilis*, and *Clostridium* species. Bites from nonhuman primates are highly susceptible to infection with pathogens similar to those isolated from human bites (see below). Bites from Old World monkeys (*Macaca*) may also result in the transmission of B virus (*Herpesvirus simiae*, cercopithecine herpesvirus), a cause of serious infection of the human central nervous system. Bites of seals, walruses, and polar bears may cause a chronic suppurative infection known as *seal finger*, which is probably due to one or more species of *Mycoplasma* colonizing these animals.

Small rodents, including rats, mice, and gerbils, as well as animals that prey on rodents may transmit *Streptobacillus moniliformis* (a microaerophilic, pleomorphic gram-negative rod) or *Spirillum minor* (a spirochete), which cause a clinical illness known as *rat-bite fever*. The vast majority of cases in the United States are streptobacillary, whereas *Spirillum* infection occurs mainly in Asia.

In the United States, the risk of rodent bites is usually greatest among laboratory workers or inhabitants of rodent-infested dwellings (particularly children). Rat-bite fever is distinguished from acute bite-wound infection by its typical manifestation after the initial wound has healed. Streptobacillary disease follows an incubation period of 3–10 days. Fever, chills, myalgias, headache, and severe migratory arthralgias are usually followed by a maculopapular rash, which characteristically involves the palms and soles and may become confluent or purpuric. Complications include endocarditis, myocarditis, meningitis, pneumonia, and abscesses in many organs. *Haverhill fever* is an *S. moniliformis* infection acquired from contaminated milk or drinking water and has similar manifestations. Streptobacillary rat-bite fever was frequently fatal in the preantibiotic era. The differential diagnosis includes Rocky Mountain spotted fever, Lyme disease, leptospirosis, and secondary syphilis. The diagnosis is made by direct observation of the causative organisms in tissue or blood, by culture of the organisms on enriched media, or by serologic testing with specific agglutinins.

Spirillum infection (referred to in Japan as *sodoku*) causes pain and purple swelling at the site of the initial bite, with associated lymphangitis and regional lymphadenopathy, after an incubation period of 1–4 weeks. The systemic illness includes fever, chills, and headache. The original lesion may eventually progress to an eschar. The infection is diagnosed by direct visualization of the spirochetes in blood or tissue or by animal inoculation.

Finally, NO-1 (CDC nonoxidizer group 1) is a bacterium associated with dog- and cat-bite wounds. Infections in which NO-1 has been isolated have tended to manifest locally (i.e., as abscess and cellulitis). These infections have occurred in healthy persons with no underlying illness and in some instances have progressed from localized to systemic illnesses. The phenotypic characteristics of NO-1 are similar to those of asaccharolytic *Acinetobacter* species; i.e., NO-1 is oxidase-, indole-, and urease-negative. To date, all strains identified have been shown to be susceptible to aminoglycosides, ß-lactam antibiotics, tetracyclines, quinolones, and sulfonamides.

Human Bites

Human bites may be self-inflicted; may be sustained by medical personnel caring for patients; or may take place during fights, domestic abuse, or sexual activity. Human-bite wounds become infected more frequently (~10–15% of the time) than do bites inflicted by other animals. These infections reflect the diverse oral microflora of humans, which includes multiple species of aerobic and anaerobic bacteria. Common aerobic isolates include viridans streptococci, *S. aureus*, *E. corrodens* (which is particularly common in clenched-fist injury; see below), and *Haemophilus influenzae*. Anaerobic species, including

Fusobacterium nucleatum and *Prevotella*, *Porphyromonas*, and *Peptostreptococcus* species, are isolated from 50% of human-bite wound infections; many of these isolates produce β-lactamases. The oral flora of hospitalized and debilitated patients often includes Enterobacteriaceae in addition to the usual organisms. Hepatitis B, hepatitis C, herpes simplex virus infection, syphilis, tuberculosis, actinomycosis, and tetanus have been reported to be transmitted by human bites; it is biologically possible to transmit HIV through human bites, although this event is quite unlikely.

Human bites are categorized as either "occlusional" injuries, which are inflicted by actual biting, or "clenched-fist" injuries, which are sustained when the fist of one individual strikes the teeth of another, causing traumatic laceration of the hand. For several reasons, clenched-fist injuries, which are more common than occlusional injuries, result in particularly serious infections. The deep spaces of the hand, including the bones, joints, and tendons, are frequently inoculated with organisms in the course of such injuries. The clenched position of the fist during injury, followed by extension of the hand, may further promote the introduction of bacteria as contaminated tendons retract beneath the skin's surface. Moreover, medical attention is often sought only after frank infection develops.

Approach to the Patient: Animal or Human Bites

A careful history should be elicited, including the type of biting animal, the type of attack (provoked or unprovoked), and the amount of time elapsed since injury. Local and regional public-health authorities should be contacted to determine whether an individual species could be rabid and/or to locate and observe the biting animal when rabies prophylaxis may be indicated (Chap. 195). Suspicious human-bite wounds should provoke careful questioning regarding domestic or child abuse. Details on antibiotic allergies, immunosuppression, splenectomy, liver disease, mastectomy, and immunization history should be obtained. The wound should be inspected carefully for evidence of infection, including redness, exudate, and foul odor. The type of wound (puncture, laceration, or scratch); the depth of penetration; and the possible involvement of joints, tendons, nerves, and bones should be assessed. It is often useful to include a diagram or photograph of the wound in the medical record. In addition, a general physical examination should be conducted and should include an assessment of vital signs as well as an evaluation for evidence of lymphangitis, lymphadenopathy, dermatologic lesions, and functional limitations. Injuries to the hand warrant consultation with a hand surgeon for the assessment of tendon, nerve, and muscular damage. Radiographs should be obtained when bone may have been penetrated or a tooth fragment may be present. Culture and Gram's staining of all infected wounds are essential; anaerobic cultures should be undertaken if abscesses, devitalized tissue, or foul-smelling exudate is present. A small-tipped swab may be used to culture deep punctures or small lacerations. It is also reasonable to culture samples from uninfected wounds due to bites inflicted by animals other than dogs and cats, since the microorganisms causing disease are less predictable in these cases. The white blood cell count should be determined and blood cultured if systemic infection is suspected.

TREATMENT: BITE-WOUND INFECTIONS

Wound Management

Wound closure is controversial in bite injuries. Many authorities prefer not to attempt primary closure of wounds that are or may become infected, preferring to irrigate these wounds copiously, debride devitalized tissue, remove foreign bodies, and approximate the wound edges. Delayed primary closure may be undertaken after the risk of infection is over. Small uninfected wounds may be allowed to close by secondary intention. Puncture wounds due to cat bites should be left unsutured because of the high rate at which they become infected. Facial wounds are usually sutured after thorough cleaning and irrigation because of the importance of a good cosmetic result in this area and because anatomic factors such as an excellent blood supply and the absence of dependent edema lessen the risk of infection.

Antibiotic Therapy

ESTABLISHED INFECTION

Antibiotics should be administered in all established bite-wound infections and should be chosen in light of the most likely potential pathogens, as indicated by the biting species and by Gram's stain and culture results (Table e24-1). For dog and cat bites, antibiotics should be effective against *S. aureus, Pasteurella* species, *C. canimorsus*, streptococci, and oral anaerobes. For human bites, agents with activity against *S. aureus, H. influenzae*, and β-lactamase-positive oral anaerobes should be used. The combination of an extended-spectrum penicillin with a β-lactamase inhibitor (amoxicillin/clavulanic acid, ticarcillin/clavulanic acid, ampicillin/sulbactam) appears to offer the most reliable coverage for these pathogens. Second-generation cephalosporins (cefuroxime, cefoxitin) also offer substantial coverage. The choice of antibiotics for penicillin-allergic patients (particularly those in whom immediate-type hypersensitivity makes the use of cephalosporins hazardous) is more difficult and is based primarily on in vitro sensitivity since data on clinical efficacy are inadequate. The combination of an antibiotic active against gram-positive cocci and anaerobes (such as clindamycin) with trimethoprim-sulfamethoxazole or a fluoroquinolone, which is active against many of the other potential pathogens, would appear reasonable. In vitro data suggest that azithromycin alone provides coverage against most commonly isolated bite-wound pathogens. As MRSA becomes more common in the community and evidence of its

transmission between humans and their animal contacts increases, empirical use of agents active against MRSA should be considered in high-risk situations while culture results are awaited.

Table e24-1 Management of Wound Infections Following Animal and Human Bites

Biting Species	Commonly Isolated Pathogens	Preferred Antibiotic(s)[a]	Alternative in Penicillin-Allergic Patient	Prophylaxis Advised for Early Uninfected Wounds	Other Considerations
Dog	*Staphylococcus aureus, Pasteurella multocida,* anaerobes, *Capnocytophaga canimorsus*	Amoxicillin/clavulanate (250-500 mg PO tid) or ampicillin/sulbactam (1.5-3.0 g IV q6h)	Clindamycin (150-300 mg PO qid) plus either TMP-SMX (1 DS tablet PO bid) or ciprofloxacin (500 mg PO bid)	Sometimes[b]	Consider rabies prophylaxis.
Cat	*P. multocida, S. aureus,* anaerobes	Amoxicillin/clavulanate or ampicillin/sulbactam, as above	Clindamycin plus TMP-SMX as above or a fluoroquinolone	Usually	Consider rabies prophylaxis. Carefully evaluate for joint/bone penetration.
Human, occlusional	Viridans streptococci, *S. aureus, Haemophilus influenzae,* anaerobes	Amoxicillin/clavulanate or ampicillin/sulbactam, as above	Erythromycin (500 mg PO qid) or a fluoroquinolone	Always	- [Au: Does the comment for clenched-fist injury also apply to occlusional injury?]
Human, clenched-fist	As for occlusional plus *Eikenella corrodens*	Ampicillin/sulbactam as above or imipenem (500 mg q6h)	Cefoxitin[c]	Always	Examine for tendon, nerve, or joint involvement.
Monkey	As for human bite	As for human bite	As for human bite	Always	For macaque monkeys, consider B virus prophylaxis with acyclovir.
Snake	*Pseudomonas aeruginosa, Proteus* spp., *Bacteroides fragilis, Clostridium* spp.	Ampicillin/sulbactam as above	Clindamycin plus TMP-SMX as above or a fluoroquinolone	Sometimes, especially with venomous snakes	Antivenin for venomous snake bite
Rodent	*Streptobacillus moniliformis, Leptospira* spp., *P. multocida*	Penicillin VK (500 mg PO qid)	Doxycycline (100 mg PO bid)	Sometimes	—

[a] Antibiotic choices should be based on culture data when available. These suggestions for empirical therapy need to be tailored to individual circumstances and local conditions. IV regimens should be used for hospitalized patients. A single IV dose of antibiotics may be given to patients who will be discharged after initial management.

[b] Prophylactic antibiotics are suggested for severe or extensive wounds, facial wounds, and crush injuries; when bone or joint may be involved; and when comorbidity is present (see text).

[c] May be hazardous in patients with immediate-type hypersensitivity reaction to penicillin.

Abbreviations: TMP-SMX, trimethoprim-sulfamethoxazole; DS, double-strength.

Antibiotics are generally given for 10–14 days, but the response to therapy must be carefully monitored. Failure to respond should prompt a consideration of diagnostic alternatives and surgical evaluation for possible drainage or debridement. Complications such as osteomyelitis or septic arthritis mandate a longer duration of therapy.

Management of *C. canimorsus* sepsis requires a 2-week course of IV penicillin G (2 million units IV every 4 h) and supportive measures. Alternative agents for the treatment of *C. canimorsus* infection

include cephalosporins and fluoroquinolones. Serious infection with *P. multocida* (e.g., pneumonia, sepsis, or meningitis) should also be treated with IV penicillin G. Alternative agents include second- or third-generation cephalosporins or ciprofloxacin.

Bites by venomous snakes (Chap. 396) may not require antibiotic treatment. Because it is often difficult to distinguish signs of infection from tissue damage caused by the envenomation, many authorities continue to recommend treatment directed against the snake's oral flora—i.e., the administration of broadly active agents such as ceftriaxone (1–2 g IV every 12–24 h) or ampicillin/sulbactam (1.5–3.0 g IV every 6 h).

Seal finger appears to respond to doxycycline (100 mg twice daily for a duration guided by the response to therapy).

PRESUMPTIVE OR PROPHYLACTIC THERAPY

The use of antibiotics for patients presenting early (within 8 h) after bite injury is controversial. Although symptomatic infection frequently will not yet have manifested at this point, many early wounds will harbor pathogens, and many will become infected. Studies of antibiotic prophylaxis for wound infections are limited and have often included only small numbers of cases in which various types of wounds have been managed according to various protocols. A meta-analysis of eight randomized trials of prophylactic antibiotics in patients with dog-bite wounds demonstrated a reduction in the rate of infection by 50% with prophylaxis. However, in the absence of sound clinical trials, many clinicians base the decision to treat bite wounds with empirical antibiotics on the species of the biting animal; the location, severity, and extent of the bite wound; and the existence of comorbid conditions in the host. All human- and monkey-bite wounds should be treated presumptively because of the high rate of infection. Most cat-bite wounds, particularly those involving the hand, should be treated. Other factors favoring treatment for bite wounds include severe injury, as in crush wounds; potential bone or joint involvement; involvement of the hands or genital region; host immunocompromise, including that due to liver disease or splenectomy; and prior mastectomy on the side of an involved upper extremity. When prophylactic antibiotics are administered, they are usually given for 3–5 days.

RABIES AND TETANUS PROPHYLAXIS

Rabies prophylaxis, consisting of both passive administration of rabies immune globulin (with as much of the dose as possible infiltrated into and around the wound) and active immunization with rabies vaccine, should be given in consultation with local and regional public health authorities for many wild-animal (and some domestic-animal) bites and scratches as well as for certain nonbite exposures (Chap. 195). Rabies is endemic in a variety of animals, including dogs and cats in many areas of the world. Many local health authorities require the reporting of all animal bites. A tetanus booster immunization should be given if the patient has undergone primary immunization but has not received a booster dose in the past 5 years. Patients who have not previously completed primary immunization should be immunized and should also receive tetanus immune globulin. Elevation of the site of injury is an important adjunct to antimicrobial therapy. Immobilization of the infected area, especially the hand, is also beneficial.

FURTHER READINGS

Baker AS et al: Isolation of Mycoplasma species from a patient with seal finger. Clin Infect Dis 27:1168, 1998 [PMID: 9827264]

Brook I: Management of human and animal bite wounds: An overview. Adv Skin Wound Care 18:197, 2005 [PMID: 15920371]

Centers for Disease Control and Prevention: Nonfatal dog bite–related injuries treated in hospital emergency departments—United States, 2001. MMWR Morb Mortal Wkly Rep 52:605, 2003

Cummings P: Antibiotics to prevent infection in patients with dog bite wounds: A meta-analysis of randomized trials. Ann Emerg Med 23:535, 1994 [PMID: 8135429]

Fallouji MA: Traumatic love bites. Br J Surg 77:100, 1990 [PMID: 2302496]

Fleisher GR: The management of bite wounds. N Engl J Med 340:138, 1999 [PMID: 9887167]

Goldstein EJ: Bite wounds and infection. Clin Infect Dis 14:633, 1992 [PMID: 1562653]

Kaiser RM et al: Clinical significance and epidemiology of NO-1, an unusual bacterium associated with dog and cat bites. Emerg Infect Dis 8:171, 2002 [PMID: 11897069]

Kullberg BJ et al: Purpura fulminans and symmetrical peripheral gangrene caused by *Capnocytophaga canimorsus* (formerly DF-2) septicemia—a complication of dog bite. Medicine (Baltimore) 70:287, 1991 [PMID: 1921703]

Oehler RL et al: Bite-related and septic syndromes caused by cats and dogs. Lancet Infect Dis 9:439, 2009 [PMID: 19555903]

Rankin S et al: Panton Valentine leukocidin (PVL) toxin positive MRSA strains isolated from companion animals. Vet Microbiol 108: 145, 2005 [PMID: 15917142]

Talan DA et al: Bacteriological analysis of infected dog and cat bites. N Engl J Med 340:85, 1999 [PMID: 9887159]

Tan JS: Human zoonotic infections transmitted by dogs and cats. Arch Intern Med 157:1933, 1977

Weber DJ et al: Infections resulting from animal bites. Infect Dis Clin North Am 5:663, 1991 [PMID: 1955705]

Weinberg AN, Branda JA: Case 31-2010. A 29-year-old woman with fever after a cat bite. N Engl J Med 363:1560, 2010 [PMID: 20942673]

Weiss HB et al: Incidence of dog bite injuries treated in emergency departments. JAMA 279:51, 1998 [PMID: 9424044]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS IN TRANSPLANT RECIPIENTS: INTRODUCTION

This chapter considers aspects of infection unique to patients receiving transplanted organs. The evaluation of infections in transplant recipients involves consideration of both the donor and the recipient of the transplanted organ. Two central issues are of paramount importance: (1) Infectious agents (particularly viruses, but also bacteria, fungi, and parasites) can be introduced into the recipient by the donor organ. (2) Treatment of the recipient with medicine to prevent rejection can suppress normal immune responses, greatly increasing susceptibility to infection. Thus, what might have been a latent or asymptomatic infection in an immunocompetent donor or in the recipient prior to therapy can become a life-threatening problem when the recipient becomes immunosuppressed. The pretransplantation evaluation of each patient should be guided by an analysis of both (1) what infections the recipient is currently harboring, since organisms that exist in a state of latency or dormancy before the procedure may cause fatal disease when the patient receives immunosuppressive treatment; and (2) what organisms are likely to be transmitted by the donor organ, particularly those to which the recipient may be naïve.

Pretransplantation Evaluation

THE DONOR

A variety of organisms have been transmitted by organ transplantation **(Table 132-1)**. Transmission of infections that may have been latent or not clinically apparent in the donor has resulted in the development of specific donor-screening protocols. Serologic studies should be ordered to detect viruses such as herpes simplex virus types 1 and 2 (HSV-1, HSV-2), varicella-zoster virus (VZV), cytomegalovirus (CMV), Epstein-Barr virus (EBV), and Kaposi's sarcoma–associated herpesvirus (KSHV) as well as hepatitis A, B, and C viruses and HIV. In addition, when relevant, donors should be screened for viruses such as West Nile virus, rabies virus, human T lymphotropic virus type I, and lymphocytic choriomeningitis virus as well as for parasites such as *Toxoplasma gondii*, *Strongyloides stercoralis*, *Schistosoma* species, and *Trypanosoma cruzi* (the latter particularly in Latin America). Clinicians caring for prospective organ donors should examine chest radiographs for evidence of granulomatous disease (e.g., caused by mycobacteria or fungi) and should perform skin testing or obtain blood for immune cell–based assays that detect active or latent *Mycobacterium tuberculosis* infection. Evaluation for syphilis should also be performed. An investigation of the donor's dietary habits (e.g., consumption of raw meat or fish or of unpasteurized dairy products), occupations or avocations (e.g., gardening or spelunking), and travel history (e.g., travel to areas with endemic fungi) is also indicated and may mandate additional testing (Table 132-1).

Table 132-1 Common Pathogens Transmitted by Organ Transplantation: Frequent Sites of Reactivation and Disease[a]

	Blood	Lungs	Heart	Brain	Liver/Spleen	Skin
Bacteria/Mycobacteria						
Mycobacterium tuberculosis	±	+			±	
Atypical mycobacteria	+	+				
Brucella spp.	+					
Viruses						
Cytomegalovirus[b]	+	+	±	±	+	
Epstein-Barr virus[c]	+	+	±	±	+	
Herpes simplex virus		±		±	±	+
Human herpesvirus type 6	+	±		±		+
Kaposi's sarcoma–associated herpesvirus	+	±			±	+
Hepatitis B and C viruses					+	
Rabies virus[d]				+		

West Nile virus	+			+		
Lymphocytic choriomeningitis virus	+			+	±	
Fungi						
Candida albicans	+	+			+	+
Histoplasma capsulatum	+	+			+	+
Cryptococcus neoformans	+	+		+	±	±
Parasites						
Toxoplasma gondii[e]		+	+	+		
Strongyloides stercoralis[f,g]		+				
Trypanosoma cruzi[g]			+			
Plasmodium falciparum[g]	+					
Schistosoma spp.					+	
Prion Diseases						
Creutzfeldt-Jakob disease (CJD)[h]				+		
Variant CJD/bovine spongiform encephalopathy[i]				+		

[a] +, well documented; ±, less common.

[b] Cytomegalovirus reactivation is prone to occur in the transplanted organ. The same may be true for Kaposi's sarcoma–associated herpesvirus.

[c] Epstein-Barr virus reactivation usually presents as an extranodal proliferation of transformed B cells and can be present either as a diffuse disease or as a mass lesion in a single organ. It can occur in the allograft.

[d] Rabies virus has been transmitted through corneal transplants. e T. gondii usually causes disease in the brain. In hematopoietic stem cell transplant recipients, acute pulmonary disease may also occur. Heart transplant recipients develop disease in the allograft.

[f] Strongyloides "hyperinfection" may present with pulmonary disease—often associated with gram-negative bacterial pneumonia.

[g] While transmission with organs has been described, it is unusual.

[h] CJD (sporadic and familial) has been transmitted with corneal transplants. Whether it can be transmitted with blood is not known.

[i] Variant CJD can be transmitted with transfused non-leukodepleted blood, posing a theoretical risk to transplant recipients.

THE RECIPIENT

It is expected that the recipient will have been even more comprehensively assessed than the donor. Additional studies recommended for the recipient include evaluation for acute respiratory virus and gastrointestinal pathogens in the immediate pretransplantation period. An important caveat is that, because of immune dysfunction resulting from chemotherapy or underlying chronic disease, serologic testing of the recipient may prove less reliable than usual.

THE DONOR CELLS/ORGAN

Careful attention to the sterility of the medium used to process the donor organ, combined with meticulous microbiologic evaluation, reduces rates of transmission of bacteria (or, rarely, yeasts) that may be present or grow in the organ culture medium. From 2% to >20% of donor kidneys are estimated to be contaminated with bacteria—in most cases, with the organisms that colonize the skin or grow in the tissue culture medium used to bathe the donor organ while it awaits implantation. The reported rate of bacterial contamination of transplanted stem cells (bone marrow, peripheral blood, cord blood) is as high as 17% but most commonly is ~1%. The use of enrichment columns and monoclonal antibody depletion procedures results in a higher incidence of contamination. In one series of patients receiving contaminated stem cells, 14% had fever or bacteremia, but none died. Results of cultures performed at the time of cryopreservation and at the time of thawing were helpful in guiding therapy for the recipient.

INFECTIONS IN HEMATOPOIETIC STEM CELL TRANSPLANT RECIPIENTS

Transplantation of hematopoietic stem cells (HSCs) from bone marrow or from peripheral or cord blood for cancer, immunodeficiency, or autoimmune disease results in a transient state of complete immunologic incompetence. Immediately after myeloablative chemotherapy and transplantation, both innate immune cells (phagocytes, natural killer cells) and adaptive immune cells (T and B cells) are absent, and the host is extremely susceptible to infection. The reconstitution that follows transplantation has been likened to maturation of the immune system in neonates. The analogy does not entirely predict infections seen in HSC transplant recipients, however, because the stem cells mature in an old host who has several latent infections already. The choice among the current variety of methods for obtaining stem cells is determined by availability and by the need to optimize the chances of cure for an individual recipient. One strategy is autologous HSC transplantation, in which the donor and the recipient are the same. After chemotherapy, stem cells are collected and are purged (ex vivo) of residual neoplastic populations. Allogeneic HSC transplantation has the advantage of providing a graft-versus-tumor effect. In this case, the recipient is matched to varying degrees for human leukocyte antigen (HLA) with a donor who may be related or unrelated. In some individuals, nonmyeloablative therapy (mini-allo transplantation) is used and permits recipient cells to persist for some time after transplantation while preserving the graft-versus-tumor effect and sparing the recipient myeloablative therapy. Cord-blood transplantation is increasingly utilized in adults; two independent cord-blood units are typically required for suitable neutrophil engraftment early after transplantation, even though only one of the units is likely to provide long-term engraftment. In each circumstance, a different balance is struck between the toxicity of conditioning therapy, the need for a maximal graft-versus-target effect, short-term and long-term infectious complications, and the risk of graft-versus-host disease (GVHD; acute versus chronic). The various approaches differ in terms of reconstitution speed, cell lineage, and likelihood of GVHD—all factors that can produce distinct effects on the risk of infection after transplantation **(Table 132-2).** Despite these caveats, most infections occur in a predictable time frame after transplantation **(Table 132-3)**.

Table 132-2 Risk of Infection, by Type of Hematopoietic Stem Cell Transplant

Type of Hematopoietic Stem Cell Transplant	Source of Stem Cells	Risk of Early Infection: Neutrophil Depletion	Risk of Late Infection: Impaired T and B Cell Function	Risk of Ongoing Infection: GVHD[a] and Iatrogenic Immunosuppression	Graft versus Tumor Effect
Autologous	Recipient (self)	High risk; neutrophil recovery sometimes prolonged	~ 1 year	Minimal to no risk of GVHD and late-onset severe infection	None (–)
Syngeneic (genetic twin)	Identical twin	Low risk; 1–2 weeks for recovery	~ 1 year	Minimal risk of GVHD and late-onset severe infection	+/–
Allogeneic related	Sibling	Low risk; 1–2 weeks for recovery	~ 1 year	Minimal to moderate risk of GVHD and late-onset severe infection	++
Allogeneic related	Child/parent (haploidentical)	Intermediate risk; 2–3 weeks for neutrophil recovery	1–2 years	Moderate risk of GVHD and late-onset severe infection	++++
Allogeneic unrelated adult	Unrelated donor	Intermediate risk; 2–3 weeks for neutrophil recovery	1–2 years	High risk of GVHD and late-onset severe infection	++++
Allogeneic unrelated cord blood	Unrelated cord blood units (x2)	Intermediate to high risk; neutrophil recovery sometimes prolonged	Prolonged	Minimal to moderate risk of GVHD and late-onset severe infection	++++
Allogeneic mini (nonmyeloablative)	Donor (transiently coexisting with recipient cells)	Low risk; neutrophil counts close to normal	1–2+ years	Variable risk of GVHD and late-onset severe infection[b]	++++ (but develops slowly)

[a] GVHD, graft-versus-host disease.

[b] Depending on the disparity of the match (major and minor histocompatibility antigens), GVHD may be severe or mild, the requirement for immunosuppression intense or minimal, and the risk of severe late

infections coordinate with the degree of immunosuppression.

Table 132-3 Common Sources of Infections after Hematopoietic Stem Cell Transplantation

	Period after Transplantation		
Infection Site	**Early (<1 Month)**	**Middle (1–4 Months)**	**Late (>6 Months)**
Disseminated	Aerobic bacteria (gram-negative, gram-positive)	*Nocardia, Candida, Aspergillus,* EBV	Encapsulated bacteria (*Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis*)
Skin and mucous membranes	HSV	HHV-6	VZV
Lungs	Aerobic bacteria (gram-negative, gram-positive), *Candida, Aspergillus*, other molds, HSV	CMV, seasonal respiratory viruses, *Pneumocystis, Toxoplasma*	*Pneumocystis, S. pneumoniae*
Gastrointestinal tract	*Clostridium difficile*	CMV, adenovirus	EBV, CMV
Kidney		BK virus, adenovirus	
Brain	HHV-6	HHV-6, *Toxoplasma*	*Toxoplasma*, JC virus (rare)
Bone marrow	HHV-6		

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus; HHV-6, human herpesvirus type 6; HSV, herpes simplex virus; VZV, varicella-zoster virus.

Bacterial Infections

In the first month after HSC transplantation, infectious complications are similar to those in granulocytopenic patients receiving chemotherapy for acute leukemia (Chap. 86). Because of the anticipated 1- to 4-week duration of neutropenia and the high rate of bacterial infection in this population, many centers give prophylactic antibiotics to patients upon initiation of myeloablative therapy. Quinolones decrease the incidence of gram-negative bacteremia among these patients. Bacterial infections are common in the first few days after HSC transplantation. The organisms involved are predominantly those found on the skin, mucosa, or IV catheters (*Staphylococcus aureus*, coagulase-negative staphylococci, streptococci) or aerobic bacteria that colonize the bowel (*Escherichia coli*, *Klebsiella*, *Pseudomonas*). *Bacillus cereus*, although rare, has emerged as a pathogen early after transplantation and can cause meningitis, which is unusual in these patients. Chemotherapy, use of broad-spectrum antibiotics, and delayed reconstitution of humoral immunity place HSC transplant patients at risk for diarrhea and colitis caused by *Clostridium difficile* overgrowth and toxin production.

Beyond the first few days of neutropenia, infections with nosocomial pathogens (e.g., vancomycin-resistant enterococci, *Stenotrophomonas maltophilia*, *Acinetobacter* species, and extended-spectrum ß-lactamase-producing gram-negative organisms) as well as with filamentous bacteria (e.g., *Nocardia* species) become more common. Vigilance is indicated, particularly for patients with a history of active or known latent tuberc ulosis, even when they have been appropriately pretreated. Episodes of bacteremia due to encapsulated organisms mark the late posttransplantation period (>6 months after HSC reconstitution); patients who have undergone splenectomy and those with persistent hypogammaglobulinemia are at particular risk.

Fungal Infections

Beyond the first week after transplantation, fungal infections become increasingly common, particularly among patients who have received broad-spectrum antibiotics. As in most granulocytopenic patients, *Candida* infections are most commonly seen in this setting. However, with increased use of prophylactic fluconazole, infections with resistant fungi—in particular, *Aspergillus* and other molds (*Fusarium*, *Scedosporium*, *Penicillium)*—have become more common, prompting some centers to replace fluconazole with agents such as micafungin, voriconazole, and even posaconazole. The role of antifungal prophylaxis with these different agents, in contrast to empirical treatment for suspected (based on positive ß-d-glucan assay or galactomannan antigen test) or documented infection, remains controversial (Chap. 86). In patients with GVHD who require prolonged or indefinite courses of glucocorticoids and other immunosuppressive agents [e.g., cyclosporine, tacrolimus (FK 506, Prograf), mycophenolate mofetil (Cellcept), rapamycin (sirolimus, Rapamune), antithymocyte globulin, or anti-CD52 antibody (alemtuzumab, Campath, an antilymphocyte and antimonocyte monoclonal antibody)], there is a high risk of fungal infection (usually with *Candida* or *Aspergillus*), even after engraftment

and resolution of neutropenia. These patients are also at high risk for reactivation of latent fungal infection (histoplasmosis, coccidioidomycosis, or blastomycosis) in areas where endemic fungi reside and after involvement in activities such as gardening or caving. Prolonged use of central venous catheters for parenteral nutrition (lipids) increases the risk of fungemia with *Malassezia*. Some centers administer prophylactic antifungal agents to these patients. Because of the high and prolonged risk of *Pneumocystisjiroveci* pneumonia (especially among patients being treated for hematologicmalignancies), most patients receive maintenance prophylaxis with trimethoprim-sulfamethoxazole (TMP-SMX) starting 1 month after engraftment and continuing for at least 1 year.

Parasitic Infections

The regimen just described for the fungal pathogen *Pneumocystis* may also protect patients seropositive for the parasite *T. gondii*, which can cause pneumonia, visceral disease (occasionally), and central nervous system (CNS) lesions (more commonly). The advantages of maintaining HSC transplant recipients on daily TMP-SMX for 1 year after transplantation include some protection against *Listeria monocytogenes* and nocardial disease as well as late infections with *Streptococcus pneumoniae* and *Haemophilus influenzae*, which stem from the inability of the immature immune system to respond to polysaccharide antigens.

With increasing international travel, parasitic diseases typically restricted to particular environmental niches may pose a risk of reactivation in certain patients after HSC transplantation. Thus, in recipients with an appropriate history who were not screened and/or treated before transplantation or in patients with recent exposures, evaluation for infection with *Strongyloides*, *Leishmania*, or various parasitic causes of diarrheal illness (*Giardia*, *Cryptosporidium*, microsporidia) may be warranted.

Viral Infections

HSC transplant recipients are susceptible to infection with a variety of viruses, including primary and reactivation syndromes caused by most human herpesviruses **(Table 132-4)** and acute infections caused by viruses that circulate in the community.

Table 132-4 Herpesvirus Syndromes of Transplant Recipients

Virus	Reactivation Disease
Herpes simplex virus type 1	Oral lesions Esophageal lesions Pneumonia (only in HSC transplant recipients) Hepatitis (rare)
Herpes simplex virus type 2	Anogenital lesions Hepatitis (rare)
Varicella-zoster virus	Zoster (can disseminate)
Cytomegalovirus	Associated with graft rejection Fever and malaise Bone marrow failure Pneumonitis Gastrointestinal disease
Epstein-Barr virus	B cell lymphoproliferative disease/lymphoma Oral hairy leukoplakia (rare)
Human herpesvirus type 6	Fever Delayed monocyte/platelet engraftment Encephalitis (controversial)
Human herpesvirus type 7	Undefined
Kaposi's sarcoma–associated virus	Kaposi's sarcoma Primary effusion lymphoma (rare) Multicentric Castleman's disease (rare) Marrow aplasia (rare)

Abbreviation: HSC, hematopoietic stem cell.

HERPES SIMPLEX VIRUS

Within the first 2 weeks after transplantation, most patients who are seropositive for HSV-1 excrete the virus from the oropharynx. The ability to isolate HSV declines with time. Administration of prophylactic acyclovir (or valacyclovir) to seropositive HSC transplant recipients has been shown to reduce mucositis and prevent HSV pneumonia (a rare condition reported almost exclusively in allogeneic HSC transplant recipients). Both esophagitis (usually due to HSV-1) and anogenital disease (commonly caused by HSV-2) may be prevented with acyclovir prophylaxis. **For further discussion, see Chap. 179.**

VARICELLA-ZOSTER VIRUS

Reactivation of VZV manifests as herpes zoster and may occur within the first month but more commonly occurs several months after transplantation. Reactivation rates are ~40% for allogeneic HSC transplant recipients and 25% for autologous recipients. Localized zoster can spread rapidly in an immunosuppressed patient. Fortunately, disseminated disease can usually be controlled with high doses of acyclovir. Because of frequent dissemination among patients with skin lesions, acyclovir is given prophylactically in some centers to prevent severe disease. Low doses of acyclovir (400 mg orally, three times daily) appear to be effective in preventing reactivation of VZV. However, acyclovir can also suppress the development of VZV-specific immunity. Thus, its administration for only 6 months after transplantation does not prevent zoster from occurring when treatment is stopped. Administration of low doses of acyclovir for an entire year after transplantation is effective and may eliminate most cases of posttransplantation zoster. **For further discussion, see Chap. 180.**

CYTOMEGALOVIRUS

The onset of CMV disease (interstitial pneumonia, bone marrow suppression, graft failure, hepatitis/colitis) usually begins 30–90 days after HSC transplantation, when the granulocyte count is adequate but immunologic reconstitution has not occurred. CMV disease rarely develops earlier than 14 days after transplantation and may become evident as late as 4 months after the procedure. It is of greatest concern in the second month after transplantation, particularly in allogeneic HSC transplant recipients. In cases in which the donor marrow is depleted of T cells (to prevent GVHD or eliminate a T cell tumor), the disease may be manifested earlier. The use of alemtuzumab to prevent GVHD in nonmyeloablative transplantation has been associated with an increase in CMV disease. Patients who receive ganciclovir for prophylaxis, preemptive treatment, or treatment (see below) may develop recurrent CMV infection even later than 4 months after transplantation, as treatment appears to delay the development of the normal immune response to CMV infection. Although CMV disease may present as isolated fever, granulocytopenia, thrombocytopenia, or gastrointestinal disease, the foremost cause of death from CMV infection in the setting of HSC transplantation is pneumonia.

With the standard use of CMV-negative or filtered blood products, primary CMV infection should be a major risk in allogeneic transplantation only when the donor is CMV-seropositive and the recipient is CMV-seronegative. Reactivation disease or superinfection with another strain from the donor is also common in CMV-positive recipients, and most seropositive patients who undergo HSC transplantation excrete CMV, with or without clinical findings. Serious CMV disease is much more common among allogeneic than autologous recipients and is often associated with GVHD. In addition to pneumonia and marrow suppression (and, less often, graft failure), manifestations of CMV disease in HSC transplant recipients include fever with or without arthralgias, myalgias, hepatitis, and esophagitis. CMV ulcerations occur in both the lower and the upper gastrointestinal tract, and it may be difficult to distinguish diarrhea due to GVHD from that due to CMV infection. The finding of CMV in the liver of a patient with GVHD does not necessarily mean that CMV is responsible for hepatic enzyme abnormalities. It is interesting that the ocular and neurologic manifestations of CMV infections, which are common in patients with AIDS, are uncommon in patients who develop disease after transplantation.

Management of CMV disease in HSC transplant recipients includes strategies directed at prophylaxis, preemptive therapy (suppression of silent replication), and treatment of disease. Prophylaxis results in a lower incidence of disease at the cost of treating many patients who otherwise would not require therapy. Because of the high fatality rate associated with CMV pneumonia in these patients and the difficulty of early diagnosis of CMV infection, prophylactic IV ganciclovir (or oral valganciclovir) has been used in some centers and has been shown to abort CMV disease during the period of maximal vulnerability (from engraftment to day 120 after transplantation). Ganciclovir also prevents HSV reactivation and reduces the risk of VZV reactivation; thus acyclovir prophylaxis should be discontinued when ganciclovir is administered. The foremost problem with the administration of ganciclovir relates to adverse effects, which include dose-related bone marrow suppression (thrombocytopenia, leukopenia, anemia, and pancytopenia). Because the frequency of CMV pneumonia is lower among autologous HSC transplant recipients (2–7%) than among allogeneic HSC transplant recipients (10–40%), prophylaxis in the former group will not become the rule until a less toxic oral antiviral agent becomes available.

Preemptive treatment of CMV—that is, initiation of therapy with drugs only after CMV is detected in blood, typically by a nucleic acid amplification test—is used at most centers. The preemptive approach

has supplanted prophylactic therapy, or treatment of all seropositive (recipient and/or donor) HSC transplants with an antiviral agent (typically ganciclovir), because of toxic drug side effects (e.g., neutropenia and bone marrow suppression). Quantitative viral load assays, which are not dependent on circulating leukocytes, have supplanted older antigen-based assays for CMV. A positive test (or increasing viral load) prompts the initiation of preemptive therapy with ganciclovir. Preemptive approaches that target patients who have polymerase chain reaction (PCR) evidence of CMV can still lead to unnecessary treatment of many individuals with drugs that have adverse effects on the basis of a laboratory test that is not highly predictive of disease; however, invasive disease, particularly in the form of pulmonary infection, is difficult to treat and is associated with high mortality rates. When prophylaxis or preemptive therapy is stopped, late manifestations of CMV replication may occur, although by then the HSC transplant patient is often equipped with improved graft function and is better able to combat disease.

CMV pneumonia in HSC transplant recipients (unlike that in other clinical settings) is often treated with both IV immunoglobulin (IVIg) and ganciclovir. In patients who cannot tolerate ganciclovir, foscarnet is a useful alternative, although it may produce nephrotoxicity and electrolyte imbalance. When neither ganciclovir nor foscarnet is clinically tolerated, cidofovir can be used; however, its efficacy is less well established, and its side effects include nephrotoxicity. Case reports have suggested that the immunosuppressive agent leflunomide may be active in this setting, but controlled studies are lacking. Transfusion of CMV-specific T cells from the donor has decreased viral load in a small series of patients; this result suggests that immunotherapy may play a role in the treatment of this disease in the future.

For further discussion, see Chap. 182.

HUMAN HERPESVIRUSES 6 AND 7

Human herpesvirus type 6 (HHV-6), the cause of roseola in children, is a ubiquitous herpesvirus that reactivates (as determined by quantitative plasma PCR) in ~50% of HSC transplant recipients 2–4 weeks after transplantation. Reactivation is more common among patients requiring glucocorticoids for GVHD and among those receiving second transplants. Reactivation of HHV-6, primarily type B, may be associated with delayed monocyte and platelet engraftment. Limbic encephalitis developing after transplantation has been associated with HHV-6 in cerebrospinal fluid (CSF). The causality of the association is not well defined: in several cases, plasma viremia was detected long before the onset of encephalitis. Nevertheless, most patients with encephalitis had very high viral loads in plasma at the time of CNS illness, and viral antigen has been detected in hippocampal astrocytes. HHV-6 DNA is sometimes found in lung samples after transplantation. However, its role in pneumonitis is unclear, as co-pathogens are frequently present. While HHV-6 is susceptible to foscarnet or cidofovir (and possibly to ganciclovir) in vitro, the efficacy of antiviral treatment has not been well studied. Little is known about the related herpesvirus HHV-7 or its role in posttransplantation infection. **For further discussion, see Chap. 182.**

EPSTEIN-BARR VIRUS

Primary EBV infection can be fatal to HSC transplant recipients; EBV reactivation can cause EBV–B cell lymphoproliferative disease (EBV-LPD), which may also be fatal to patients taking immunosuppressive drugs. Latent EBV infection of B cells leads to several interesting phenomena in HSC transplant recipients. The marrow ablation that occurs as part of the HSC transplantation procedure may sometimes eliminate latent EBV from the host. Infection can then be reacquired immediately after transplantation by transfer of infected donor B cells. Rarely, transplantation from a seronegative donor may result in cure. The recipient is then at risk for a second primary infection.

EBV-LPD can develop in the recipient's B cells (if any survive marrow ablation) but is more likely to be a consequence of outgrowth of infected donor cells. Both lytic replication and latent replication of EBV are more likely during immunosuppression (e.g., they are associated with GVHD and the use of antibodies to T cells). Although less likely in autologous transplantation, reactivation can occur in T cell –depleted autologous recipients (e.g., patients being given antibodies to T cells for the treatment of a T cell lymphoma with marrow depletion). EBV-LPD, which can become apparent as early as 1–3 months after engraftment, can cause high fevers and cervical adenopathy resembling the symptoms of infectious mononucleosis but more commonly presents as an extranodal mass. The incidence of EBV-LPD among allogeneic HSC transplant recipients is 0.6–1%, which contrasts with figures of ~5% for renal transplant recipients and up to 20% for cardiac transplant patients. In all cases, EBV-LPD is more likely to occur with high-dose, prolonged immunosuppression, especially that caused by the use of antibodies to T cells, glucocorticoids, and calcineurin inhibitors (e.g., cyclosporine, tacrolimus). Ganciclovir, administered to preempt CMV disease, may reduce EBV lytic replication and thereby diminish the pool of B cells that can become newly infected and give rise to LPD. Increasing evidence indicates that replacement of calcineurin inhibitors with m-Tor inhibitors (e.g., rapamycin) exerts an antiproliferative effect on EBV-infected B cells that decreases the likelihood of developing LPD or unrelated proliferative disorders associated with transplant-related immunosuppression.

PCR can be used to monitor EBV production after HSC transplantation. High or increasing viral loads predict an enhanced likelihood of developing EBV-LPD and should prompt rapid reduction of

immunosuppression and search for nodal or extra nodal disease. If reduction of immunosuppression does not have the desired effect, administration of a monoclonal antibody to CD20 (rituximab or others) for the treatment of B cell lymphomas that express this surface protein has elicited dramatic responses and currently constitutes first-line therapy for CD20-positive EBV-LPD. However, long-term suppression of new antibody responses accompanies therapy, and recurrences are not infrequent. Additional B cell–directed antibodies, including anti-CD22, are under study. The role of antiviral drugs is uncertain because no available agents have been documented to have activity against the different forms of latent EBV infection. Diminishing lytic replication and virion production in these patients would theoretically produce a statistical decrease in the frequency of latent disease by decreasing the number of virions available to cause additional infection. In case reports and small animal studies, ganciclovir and/or high-dose zidovudine (AZT), together with other agents, has been used to eradicate EBV-LPD and CNS lymphomas, another EBV-associated complication of transplantation. Both interferon and retinoic acid have been employed in the treatment of EBV-LPD, as has IVIg, but no large prospective studies have assessed the efficacy of any of these agents. Several additional drugs are undergoing preclinical evaluation. Standard chemotherapeutic regimens are used if disease persists after reduction of immunosuppressive agents and administration of antibodies. EBV-specific T cells generated from the donor have been used experimentally to prevent and to treat EBV-LPD in allogeneic recipients, and efforts are under way to increase the activity and specificity of ex vivo –generated T cells. **For further discussion, see Chap. 181.**

HUMAN HERPESVIRUS 8 (KSHV)

The EBV-related gammaherpesvirus KSHV, which is causally associated with Kaposi's sarcoma, primary effusion lymphoma, and multicentric Castleman's disease, has rarely resulted in disease in HSC transplant recipients, although some cases of virus-associated marrow aplasia have been reported in the peritransplantation period. The relatively low seroprevalence of KSHV in the population and the limited duration of profound T cell suppression after HSC transplantation provide a plausible explanation for the currently low incidence of KSHV disease compared with that in recipients of solid organ transplants and patients with HIV infection. **For further discussion, see Chap. 182.**

OTHER (NONHERPES) VIRUSES

The diagnosis of pneumonia in HSC transplant recipients poses special problems. Because patients have undergone treatment with multiple chemotherapeutic agents and sometimes irradiation, their differential diagnosis should include—in addition to bacterial and fungal pneumonia—CMV pneumonitis, pneumonia of other viral etiologies, parasitic pneumonia, diffuse alveolar hemorrhage, and chemical- or radiation-associated pneumonitis. Since fungi and viruses [e.g., influenza A and B viruses, respiratory syncytial virus (RSV), parainfluenza virus (types 1–4), adenovirus, enterovirus, bocavirus, human metapneumovirus, coronavirus, and rhinovirus (increasingly detected by multiplex PCR)] can also cause pneumonia in this setting, it is important to diagnose CMV specifically (see "Cytomegalovirus," above). *M. tuberculosis* has been an uncommon cause of pneumonia among HSC transplant recipients in Western countries (accounting for <0.1–0.2% of cases) but is common in Hong Kong (5.5%) and in countries where the prevalence of tuberculosis is high. The recipient's exposure history is clearly critical in an assessment of posttransplantation infections.

Both RSV and parainfluenza viruses, particularly type 3, can cause severe or even fatal pneumonia in HSC transplant recipients. Infections with both of these agents sometimes occur as disastrous nosocomial epidemics. Therapy with palivizumab or ribavirin for RSV infection remains controversial. Influenza also occurs in HSC transplant recipients and generally mirrors the presence of infection in the community. Progression to pneumonia is more common when infection occurs early after transplantation and when the recipient is lymphopenic. Several drugs are available for the treatment of influenza. Amantadine and rimantadine have limited effects, primarily reducing symptoms and shortening the duration of illness caused by sensitive strains of influenza A virus. The neuraminidase inhibitors oseltamivir (oral) and zanamivir (aerosolized) are active against both influenza A virus and influenza B virus and are a reasonable treatment option. Parenteral forms of neuraminidase inhibitors such as peramivir (intravenous) are undergoing clinical trials. Peramivir is currently available through the Centers for Disease Control and Prevention (CDC) for the treatment of severe H1N1 influenza. An important preventive measure is immunization of household members, hospital staff members, and other frequent contacts. Adenoviruses can be isolated from HSC transplant recipients at rates varying from 5% to ≤18%. Like CMV infection, adenovirus infection usually occurs in the first to third month after transplantation and is often asymptomatic, although pneumonia, hemorrhagic cystitis/nephritis, severe gastroenteritis with hemorrhage, and fatal disseminated infection have been reported. A role for cidofovir therapy has been suggested, but the efficacy of this agent is unproven in adenovirus infection.

Although diverse respiratory viruses can sometimes cause severe pneumonia and respiratory failure in HSC transplant recipients, mild or even asymptomatic infection may be more common. For example, rhinoviruses and coronaviruses are frequent co-pathogens in HSC transplant recipients; however, whether they independently contribute to significant pulmonary infection is not known. At present, the overall contribution to the burden of lower respiratory tract disease in HSC transplant recipients for the secondary group of viral respiratory pathogens listed above is unknown.

Infections with parvovirus B19 (presenting as anemia or occasionally as pancytopenia) and disseminated enteroviruses (sometimes fatal) can occur. Parvovirus B19 infection can be treated with IVIg (Chap. 184). Intranasal pleconaril, a capsid-binding agent, is being studied for the treatment of enterovirus (and rhinovirus) infection.

Rotaviruses are a common cause of gastroenteritis in HSC transplant recipients. The polyomavirus BK virus is found at high titers in the urine of patients who are profoundly immunosuppressed. BK viruria may be associated with hemorrhagic cystitis in these patients. Compared with the incidence among patients with impaired T cell function due to HIV infection, progressive multifocal leukoencephalopathy caused by the related JC virus is relatively rare among HSC transplant recipients (Chap. 381). When transmitted by mosquitoes or by blood transfusion, West Nile virus can cause encephalitis and death after HSC transplantation.

INFECTIONS IN SOLID ORGAN TRANSPLANT RECIPIENTS

Rates of morbidity and mortality among recipients of solid organ transplants (SOTs) are reduced by the use of effective antibiotics. The organisms that cause acute infections in recipients of SOTs are different from those that infect HSC transplant recipients because SOT recipients do not go through a period of neutropenia. As the transplantation procedure involves major surgery, however, SOT recipients are subject to infections at anastomotic sites and to wound infections. Compared with HSC transplant recipients, SOT patients are immunosuppressed for longer periods (often permanently). Thus they are susceptible to many of the same organisms as patients with chronically impaired T cell immunity (Chap. 86, especially Table 86-1). Moreover, the persistent HLA mismatch between recipient immune cells (e.g., effector T cells) and the donor organ (allograft) places the organ at permanently increased risk of infection.

During the early period (<1 month after transplantation; **Table 132-5**), infections are most commonly caused by extracellular bacteria (staphylococci, streptococci, enterococci, *E. coli*, other gram-negative organisms), which often originate in surgical wound or anastomotic sites. The type of transplant largely determines the spectrum of infection. In subsequent weeks, the consequences of the administration of agents that suppress cell-mediated immunity become apparent, and acquisition—or, more commonly, reactivation—of viruses, mycobacteria, endemic fungi, and parasites (from the recipient or from the transplanted organ) can occur. CMV infection is often a problem, particularly in the first 6 months after transplantation, and may present as severe systemic disease or as infection of the transplanted organ. HHV-6 reactivation (assessed by plasma PCR) occurs within the first 2–4 weeks after transplantation and may be associated with fever, leukopenia, and very rare cases of encephalitis. Data suggest that replication of HHV-6 and HHV-7 may exacerbate CMV-induced disease. CMV is associated not only with generalized immunosuppression but also with organ-specific, rejection-related syndromes: glomerulopathy in kidney transplant recipients, bronchiolitis obliterans in lung transplant recipients, vasculopathy in heart transplant recipients, and the vanishing bile duct syndrome in liver transplant recipients. A complex interplay between increased CMV replication and enhanced graft rejection is well established: elevated immunosuppression leads to increased CMV replication, which is associated with graft rejection. For this reason, considerable attention has been focused on the diagnosis, prophylaxis, and treatment of CMV infection in SOT recipients. Early transmission of West Nile virus to transplant recipients from a donated organ or transfused blood has been reported; however, the risk of West Nile acquisition (at least in transfused blood) has been reduced by implementation of screening procedures. In rare instances, rabies virus and lymphocytic choriomeningitis virus have also been transmitted in this setting; although accompanied by distinct clinical syndromes, both viral infections have resulted in fatal encephalitis. As screening for unusual viruses is not routine, only vigilant assessment of the prospective donor is likely to prevent the use of an infected organ.

Table 132-5 Common Infections after Solid Organ Transplantation, by Site of Infection

	Period after Transplantation		
Infected Site	**Early (<1 Month)**	**Middle (1–4 Months)**	**Late (>6 Months)**
Donor organ	Bacterial and fungal infections of the graft, anastomotic site, and surgical wound	CMV infection	EBV infection (may present in allograft organ)
Systemic	Bacteremia and candidemia (often resulting from central venous catheter colonization)	CMV infection (fever, bone marrow suppression)	CMV infection, especially in patients given early posttransplantation prophylaxis; EBV proliferative syndromes (may occur in donor organs)
Lung	Bacterial aspiration pneumonia with prevalent nosocomial organisms	*Pneumocystis infection*; CMV pneumonia (highest risk in lung	*Pneumocystis* infection; granulomatous lung diseases (nocardiae, reactivated fungal

	associated with intubation and sedation (highest risk in lung transplantation)	transplantation); *Aspergillus* infection (highest risk in lung transplantation)	and mycobacterial diseases)
Kidney	Bacterial and fungal (*Candida*) infections (cystitis, pyelonephritis) associated with urinary tract catheters (highest risk in kidney transplantation)	Renal transplantation: BK virus infection (associated with nephropathy); JC virus infection	Renal transplantation: bacteria (late urinary tract infections, usually not associated with bacteremia); BK virus (nephropathy, graft failure, generalized vasculopathy)
Liver and biliary tract	Cholangitis	CMV hepatitis	CMV hepatitis
Heart		*Toxoplasma gondii* infection (highest risk in heart transplantation)	*T. gondii* (highest risk in heart transplantation)
Gastrointestinal tract	Peritonitis, especially after liver transplantation	Colitis secondary to *Clostridium difficile* infection (risk can persist)	Colitis secondary to C. difficile infection (risk can persist)
Central nervous system		*Listeria* (meningitis); *T. gondii* infection	Listeria meningitis; Cryptococcus meningitis; Nocardia abscess; JC virus –associated PML

Abbreviations: CMV, cytomegalovirus, EBV, Epstein-Barr virus, PML, progressive multifocal leukoencephalopathy.

Beyond 6 months after transplantation, infections characteristic of patients with defects in cell-mediated immunity—e.g., infections with *Listeria*, *Nocardia*, *Rhodococcus*, mycobacteria, various fungi, and other intracellular pathogens—may be a problem. International patients and global travelers may experience reactivation of dormant infections with trypanosomes, *Leishmania*, *Plasmodium*, *Strongyloides*, and other parasites. Reactivation of latent *M. tuberculosis* infection, while rare in Western nations, is far more common among persons from developing countries. The recipient is typically the source, although reactivation and spread from the donor organ can occur. While pulmonary disease remains most common, atypical sites may be involved and mortality rates can be high (up to 30%). Elimination of these late infections will not be possible until the patient develops specific tolerance to the transplanted organ in the absence of drugs that lead to generalized immunosuppression. Meanwhile, vigilance, prophylaxis/preemptive therapy (when indicated), and rapid diagnosis and treatment of infections can be lifesaving in SOT recipients, who, unlike most HSC transplant recipients, continue to be immunosuppressed.

SOT recipients are susceptible to EBV-LPD from as early as 2 months to many years after transplantation. The prevalence of this complication is increased by potent and prolonged use of T cell –suppressive drugs. Decreasing the degree of immunosuppression may in some cases reverse the condition. Among SOT patients, those with heart and lung transplants—who receive the most intensive immunosuppressive regimens—are most likely to develop EBV-LPD, particularly in the lungs. Although the disease usually originates in recipient B cells, several cases of donor origin, particularly in the transplanted organ, have been noted. High organ-specific content of B lymphoid tissues (e.g., bronchial-associated lymphoid tissue in the lung), anatomic factors (e.g., lack of access of host T cells to the transplanted organ because of disturbed lymphatics), and differences in major histocompatibility loci between the host T cells and the organ (e.g., lack of cell migration or lack of effective T cell/macrophage cooperation) may result in defective elimination of EBV-infected B cells. SOT recipients are also highly susceptible to the development of Kaposi's sarcoma and, less frequently, to the B cell–proliferative disorders associated with KSHV, such as primary effusion lymphoma and multicentric Castleman's disease. Kaposi's sarcoma is 550–1000 times more common in SOT recipients than in the general population, can develop very rapidly after transplantation, and can also occur in the allograft. However, because the seroprevalence of KSHV is very low in Western countries, Kaposi's sarcoma is not often observed. Data suggest that a switch of immunosuppressive agents —from calcineurin inhibitors (cyclosporine, tacrolimus) to mTor pathway active agents (sirolimus, everolimus)—after adequate wound healing may significantly reduce the likelihood of developing Kaposi's sarcoma and perhaps of EBV-LPD and certain other posttransplantation malignancies.

Kidney Transplantation

(See Table 132-5)

EARLY INFECTIONS

Bacteria often cause infections that develop in the period immediately after kidney transplantation. There is a role for perioperative antibiotic prophylaxis, and many centers give cephalosporins to decrease the risk of postoperative complications. Urinary tract infections developing soon after

transplantation are usually related to anatomic alterations resulting from surgery. Such early infections may require prolonged treatment (e.g., 6 weeks of antibiotic administration for pyelonephritis). Urinary tract infections that occur >6 months after transplantation may be treated for shorter periods because they do not seem to be associated with the high rate of pyelonephritis or relapse seen with infections that occur in the first 3 months.

Daily prophylaxis with one double-strength tablet of TMP-SMX (800 mg of sulfamethoxazole; 160 mg of trimethoprim) for the first 4–6 months after transplantation decreases the incidence of early and middle-period infections (see below, Table 132-5, and **Table 132-6**).

Table 132-6 Prophylactic Regimens Commonly Used to Decrease Risk of Infection in Transplant Recipients

Risk Factor	Organism	Prophylactic Drug	Examination(s)*
Travel to or residence in area with known risk of endemic fungal infection	*Histoplasma, Blastomyces, Coccidioides*	Consider imidazoles based on clinical and laboratory assessment	Chest radiography, antigen testing, serology
Latent herpesviruses	HSV, VZV, CMV, EBV	Acyclovir after HSC transplantation to prevent HSV and VZV infection; ganciclovir for CMV (?EBV/?KSHV) in some settings	Serologic tests for HSV, VZV, CMV, HHV-6, EBV, KSHV; PCR
Latent fungi and parasites	*Pneumocystis jiroveci, Toxoplasma gondii*	Trimethoprim-sulfamethoxazole (dapsone or atovaquone)	Serologic test for *Toxoplasma*
History of exposure to active or latent tuberculosis	*Mycobacterium tuberculosis*	Isoniazid if recent conversion or positive chest imaging and/or no previous treatment	Chest imaging; TST and/or cell-based assay

*Serologic examination, TST (tuberculin skin test), and interferon assays may be less reliable after transplantation.

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus; HHV-6, human herpesvirus type 6; HSC, hematopoietic stem cell; HSV, herpes simplex virus; KSHV, Kaposi's sarcoma–associated herpesvirus; PCR, polymerase chain reaction; VZV, varicella-zoster virus.

MIDDLE-PERIOD INFECTIONS

Because of continuing immunosuppression, kidney transplant recipients are predisposed to lung infections characteristic of those in patients with T cell deficiency (i.e., infections with intracellular bacteria, mycobacteria, nocardiae, fungi, viruses, and parasites). A high mortality rate associated with *Legionella pneumophila* infection (Chap. 147) led to the closing of renal transplant units in hospitals with endemic legionellosis.

About 50% of all renal transplant recipients presenting with fever 1–4 months after transplantation have evidence of CMV disease; CMV itself accounts for the fever in more than two-thirds of cases and thus is the predominant pathogen during this period. CMV infection (Chap. 182) may also present as arthralgias, myalgias, or organ-specific symptoms. During this period, this infection may represent primary disease (in the case of a seronegative recipient of a kidney from a seropositive donor) or may represent reactivation disease or superinfection. Patients may have atypical lymphocytosis. Unlike immunocompetent patients, however, they rarely have lymphadenopathy or splenomegaly. Therefore, clinical suspicion and laboratory confirmation are necessary for diagnosis. The clinical syndrome may be accompanied by bone marrow suppression (particularly leukopenia). CMV also causes glomerulopathy and is associated with an increased incidence of other opportunistic infections. Because of the frequency and severity of disease, a considerable effort has been made to prevent and treat CMV infection in renal transplant recipients. An immune globulin preparation enriched with antibodies to CMV was used by many centers in the past in an effort to protect the group at highest risk for severe infection (seronegative recipients of seropositive kidneys). However, with the development of effective oral antiviral agents, CMV immune globulin is no longer used. Ganciclovir (valganciclovir) is beneficial when prophylaxis is indicated and for the treatment of serious CMV disease. The availability of valganciclovir has allowed most centers to move to oral prophylaxis for transplant recipients. Infection with the other herpesviruses may become evident within 6 months after transplantation or later. Early after transplantation, HSV may cause either oral or anogenital lesions that are usually responsive to acyclovir. Large ulcerating lesions in the anogenital area may lead to bladder and rectal dysfunction as well as predisposing to bacterial infection. VZV may cause fatal disseminated infection in nonimmune kidney transplant recipients, but in immune patients reactivation zoster usually does not disseminate outside the dermatome; thus disseminated VZV infection is a less fearsome complication in kidney transplantation than in HSC transplantation. HHV-6 reactivation may take place and (although usually asymptomatic) may be associated with fever, rash, marrow

suppression, or rarely encephalitis.

EBV disease is more serious; it may present as an extranodal proliferation of B cells that invade the CNS, nasopharynx, liver, small bowel, heart, and other organs, including the transplanted kidney. The disease is diagnosed by the finding of a mass of proliferating EBV-positive B cells. The incidence of EBV-LPD is higher among patients who acquire EBV infection from the donor and among patients given high doses of cyclosporine, tacrolimus, glucocorticoids, and anti–T cell antibodies. Disease may regress once immunocompetence is restored. KSHV infection can be transmitted with the donor kidney and result in development of Kaposi's sarcoma, although it more often represents reactivation of latent infection of the recipient. Kaposi's sarcoma often appears within 1 year after transplantation, although the range of onset is wide (1 month to ~20 years). Avoidance of immunosuppressive agents that inhibit calcineurin has been associated with less Kaposi's sarcoma, less EBV disease, and even less CMV replication. The use of rapamycin (sirolimus) has independently led to regression of Kaposi's sarcoma.

The papovaviruses BK virus and JC virus (polyomavirus hominis types 1 and 2) have been cultured from the urine of kidney transplant recipients (as they have from that of HSC transplant recipients) in the setting of profound immunosuppression. High levels of BK virus replication detected by PCR in urine and blood are predictive of pathology, especially in the setting of renal transplantation. JC virus may rarely cause similar disease in kidney transplantation. Urinary excretion of BK virus and BK viremia are associated with the development of ureteral strictures, polyomavirus-associated nephropathy (1–10% of renal transplant recipients), and (less commonly) generalized vasculopathy. Timely detection and early reduction of immunosuppression are critical and can reduce rates of graft loss related to polyomavirus-associated nephropathy from 90% to 10–30%. Therapeutic responses to IVIg, quinolones, leflunomide, and cidofovir have been reported, but the efficacy of these agents has not been substantiated through adequate clinical study. Most centers approach the problem by reducing immunosuppression in an effort to enhance host immunity and decrease viral titers. JC virus is associated with rare cases of progressive multifocal leukoencephalopathy. Adenoviruses may persist and cause hemorrhagic nephritis/cystitis with continued immunosuppression in these patients, but disseminated disease as seen in HSC transplant recipients is much less common.

Kidney transplant recipients are also subject to infections with other intracellular organisms. These patients may develop pulmonary infections with *Mycobacterium*, *Aspergillus*, and *Mucor* species as well as infections with other pathogens in which the T cell/macrophage axis plays an important role. *Listeria monocytogenes* is a common cause of bacteremia ≤1 month after renal transplantation and should be seriously considered in renal transplant recipients presenting with fever and headache. Kidney transplant recipients may develop *Salmonella* bacteremia, which can lead to endovascular infections and require prolonged therapy. Pulmonary infections with *Pneumocystis* are common unless the patient is maintained on TMP-SMX prophylaxis. *Nocardia* infection (Chap. 162) may present in the skin, bones, and lungs or in the CNS, where it usually takes the form of single or multiple brain abscesses. Nocardiosis generally occurs ≤1 month after transplantation and may follow immunosuppressive treatment for an episode of rejection. Pulmonary manifestations most commonly consist of localized disease with or without cavities, but the disease may be disseminated. The diagnosis is made by culture of the organism from sputum or from the involved nodule. As with *Pneumocystis*, prophylaxis with TMP-SMX is often efficacious in the prevention of disease.

Toxoplasmosis can occur in seropositive patients but is less common than in other transplant settings, usually developing in the first few months after kidney transplantation. Again, TMP-SMX is helpful in prevention. In endemic areas, histoplasmosis, coccidioidomycosis, and blastomycosis may cause pulmonary infiltrates or disseminated disease.

LATE INFECTIONS

Late infections (>6 months after kidney transplantation) may involve the CNS and include CMV retinitis as well as other CNS manifestations of CMV disease. Patients (particularly those whose immunosuppression has been increased) are at risk for subacutemeningitis due to *Cryptococcus neoformans*. Cryptococcal disease may present in an insidious manner (sometimes as a skin infection before the development of clear CNS findings). *Listeria* meningitis may have an acute presentation and requires prompt therapy to avoid a fatal outcome. TMP-SMX prophylaxis may reduce the frequency of Listeria infections.

Patients who continue to take glucocorticoids are predisposed to ongoing infection. "Transplant elbow," a recurrent bacterial infection in and around the elbow that is thought to result from a combination of poor tensile strength of the skin of steroid-treated patients and steroid-induced proximal myopathy, requires patients to push themselves up with their elbows to get out of chairs. Bouts of cellulitis (usually caused by *S. aureus*) recur until patients are provided with elbow protection.

Kidney transplant recipients are susceptible to invasive fungal infections, including those due to *Aspergillus* and *Rhizopus*, which may present as superficial lesions before dissemination. Mycobacterial infection (particularly that with *Mycobacterium marinum*) can be diagnosed by skin examination. Infection with *Prototheca wickerhamii* (an achlorophyllic alga) has been diagnosed by skin biopsy. Warts caused by human papillomaviruses (HPVs) are a late consequence of persistent

immunosuppression; imiquimod or other forms of local therapy are usually satisfactory. Merkel cell carcinoma, a rare and aggressive neuroendocrine skin tumor the frequency of which is increased fivefold in elderly SOT (especially kidney) recipients, has been linked to a novel polyoma virus (Merkel cell polyomavirus).

Notably, although BK virus replication and virus-associated disease can be detected far earlier, the median time to clinical diagnosis of polyomavirus-associated nephropathy is ~300 days, qualifying it as a late-onset disease. With establishment of better screening procedures (e.g., blood PCR), it is likely that this disease will be detected earlier (see "Middle-Period Infections," above).

Heart Transplantation

EARLY INFECTIONS

Sternal wound infection and mediastinitis are early complications of heart transplantation. An indolent course is common, with fever or a mildly elevated white blood cell count preceding the development of site tenderness or drainage. Clinical suspicion based on evidence of sternal instability and failure to heal may lead to the diagnosis. Common microbial residents of the skin (e.g., *S. aureus*, including methicillin-resistant strains, and *Staphylococcus epidermidis*) as well as gram-negative organisms (e.g., *Pseudomonas aeruginosa*) and fungi (e.g., *Candida*) are often involved. In rare cases, mediastinitis in heart transplant recipients can also be due to *Mycoplasma hominis* (Chap. 175). Since this organism requires an anaerobic environment for growth and may be difficult to see on conventional medium, the laboratory should be alerted that *M. hominis* infection is suspected. *M. hominis* mediastinitis has been cured with a combination of surgical debridement (sometimes requiring muscle-flap placement) and the administration of clindamycin and tetracycline. Organisms associated with mediastinitis may sometimes be cultured from pericardial fluid.

MIDDLE-PERIOD INFECTIONS

T. gondii (Chap. 214) residing in the heart of a seropositive donor may be transmitted to a seronegative recipient. Thus serologic screening for *T. gondii* infection is important before and in the months after cardiac transplantation. Rarely, active disease can be introduced at the time of transplantation. The overall incidence of toxoplasmosis is so high in the setting of heart transplantation that some prophylaxis is always warranted. Although alternatives are available, the most frequently used agent is TMP-SMX, which prevents infection with *Pneumocystis* as well as with *Nocardia* and several other bacterial pathogens. CMV also has been transmitted by heart transplantation. *Toxoplasma*, *Nocardia*, and *Aspergillus* can cause CNS infections. *L. monocytogenes* meningitis should be considered in heart transplant recipients with fever and headache.

CMV infection is associated with poor outcomes after heart transplantation. The virus is usually detected 1–2 months after transplantation, causes early signs and laboratory abnormalities (usually fever and atypical lymphocytosis or leukopenia and thrombocytopenia) at 2–3 months, and can produce severe disease (e.g., pneumonia) at 3–4 months. An interesting observation is that seropositive recipients usually develop viremia faster than patients whose primary CMV infection is a consequence of transplantation. Between 40% and 70% of patients develop symptomatic CMV disease in the form of (1) CMV pneumonia, the most likely form to be fatal; (2) CMV esophagitis and gastritis, sometimes accompanied by abdominal pain with or without ulcerations and bleeding; and (3) the CMV syndrome, consisting of CMV in the blood along with fever, leukopenia, thrombocytopenia, and hepatic enzyme abnormalities. Ganciclovir is efficacious in the treatment of CMV infection; prophylaxis with ganciclovir or possibly with other antiviral agents, as described for renal transplantation, may reduce the overall incidence of CMV-related disease.

LATE INFECTIONS

EBV infection usually presents as a lymphoma-like proliferation of B cells late after heart transplantation, particularly in patients maintained on intense immunosuppressive therapy. A subset of heart and heart-lung transplant recipients may develop early fulminant EBV-LPD (within 2 months). Treatment includes the reduction of immunosuppression (if possible), the use of glucocorticoid and calcineurin inhibitor–sparing regimens, and the consideration of therapy with anti–B cell antibodies (rituximab and possibly others). Immunomodulatory and antiviral agents continue to be studied. Ganciclovir prophylaxis for CMV disease may indirectly reduce the risk of EBV-LPD through reduced spread of replicating EBV to naïve B cells. Aggressive chemotherapy is a last resort, as discussed earlier for HSC transplant recipients. KSHV-associated disease, including Kaposi's sarcoma and primary effusion lymphoma, has been reported in heart transplant recipients. GVHD prophylaxis with sirolimus may decrease the risk of both rejection and outgrowth of KSHV-infected cells. Antitumor therapy is discussed in Chap. 85. Prophylaxis for *Pneumocystis* infection is required for these patients (see "Lung Transplantation, Late Infections," below).

Lung Transplantation

EARLY INFECTIONS

It is not surprising that lung transplant recipients are predisposed to the development of pneumonia. The combination of ischemia and the resulting mucosal damage, together with accompanying

denervation and lack of lymphatic drainage, probably contributes to the high rate of pneumonia (66% in one series). The prophylactic use of high doses of broad-spectrum antibiotics for the first 3–4 days after surgery may decrease the incidence of pneumonia. Gram-negative pathogens (Enterobacteriaceae and *Pseudomonas* species) are troublesome in the first 2 weeks after surgery (the period of maximal vulnerability). Pneumonia can also be caused by *Candida* (possibly as a result of colonization of the donor lung), *Aspergillus*, and *Cryptococcus*.

Mediastinitis may occur at an even higher rate among lung transplant recipients than among heart transplant recipients and most commonly develops within 2 weeks of surgery. In the absence of prophylaxis, pneumonitis due to CMV (which may be transmitted as a consequence of transplantation) usually presents between 2 weeks and 3 months after surgery, with primary disease occurring later than reactivation disease.

MIDDLE-PERIOD INFECTIONS

The incidence of CMV infection, either reactivated or primary, is 75–100% if either the donor or the recipient is seropositive for CMV. CMV-induced disease after solid organ transplantation appears to be most severe in recipients of lung and heart-lung transplants. Whether this severity relates to the mismatch in lung antigen presentation and host immune cells or is attributable to nonimmunologic factors is not known. More than half of lung transplant recipients with symptomatic CMV disease have pneumonia. Difficulty in distinguishing the radiographic picture of CMV infection from that of other infections or from organ rejection further complicates therapy. CMV can also cause bronchiolitis obliterans in lung transplants. The development of pneumonitis related to HSV has led to the prophylactic use of acyclovir. Such prophylaxis may also decrease rates of CMV disease, but ganciclovir is more active against CMV and is also active against HSV. The prophylaxis of CMV infection with IV ganciclovir—or increasingly with valganciclovir, the oral alternative—is recommended for lung transplant recipients. Antiviral alternatives are discussed in the earlier section on HSC transplantation. Although the overall incidence of serious disease is decreased during prophylaxis, late disease may occur when prophylaxis is stopped—a pattern observed increasingly in recent years. With recovery from peritransplantation complications and, in many cases, a decrease in immunosuppression, the recipient is often better equipped to combat late infection.

LATE INFECTIONS

The incidence of *Pneumocystis* infection (which may present with a paucity of findings) is high among lung and heart-lung transplant recipients. Some form of prophylaxis for *Pneumocystis* pneumonia is indicated in all organ transplant situations (Table 132-6). Prophylaxis with TMP-SMX for 12 months after transplantation may be sufficient to prevent *Pneumocystis* disease in patients whose immunosuppression is not increased.

As in other transplant recipients, infection with EBV may cause either a mononucleosis-like syndrome or EBV-LPD. The tendency of the B cell blasts to present in the lung appears to be greater after lung transplantation than after the transplantation of other organs, possibly because of a rich source of B cells in bronchial-associated lymphoid tissue. Reduction of immunosuppression and switching of regimens, as discussed in earlier sections, cause remission in some cases, but airway compression can be fatal and more rapid intervention may therefore become necessary. The approach to EBV-LPD is similar to that described in other sections.

Liver Transplantation

EARLY INFECTIONS

As in other transplantation settings, early bacterial infections are a major problem after liver transplantation. Many centers administer systemic broad-spectrum antibiotics for the first 24 h or sometimes longer after surgery, even in the absence of documented infection. However, despite prophylaxis, infectious complications are common and correlate with the duration of the surgical procedure and the type of biliary drainage. An operation lasting >12 h is associated with an increased likelihood of infection. Patients who have a choledochojejunostomy with drainage of the biliary duct to a Roux-en-Y jejunal bowel loop have more fungal infections than those whose bile is drained via a choledochocholedochostomy with anastomosis of the donor common bile duct to the recipient common bile duct.

Peritonitis and intraabdominal abscesses are common complications of liver transplantation. Bacterial peritonitis or localized abscesses may result from biliary leaks. Early leaks are even more common with live-donor liver transplants. Peritonitis in liver transplant recipients is often polymicrobial, commonly involving enterococci, aerobic gram-negative bacteria, staphylococci, anaerobes, *Candida*, or sometimes other invasive fungi. Only one-third of patients with intraabdominal abscesses have bacteremia. Abscesses within the first month after surgery may occur not only in and around the liver but also in the spleen, pericolic area, and pelvis. Treatment includes antibiotic administration and drainage as necessary.

Liver transplant patients have a high incidence of fungal infections, and the occurrence of fungal (often candidal) infection correlates with preoperative use of glucocorticoids, long duration of treatment with

antibacterial agents, and posttransplantation use of immunosuppressive agents. Many centers give fluconazole prophylactically in this setting.

MIDDLE-PERIOD INFECTIONS

The development of postsurgical biliary stricture predisposes patients to cholangitis. The incidence of strictures is increased in live-donor liver transplantation. Transplant recipients who develop cholangitis may have high spiking fevers and rigors but often lack the characteristic signs and symptoms of classic cholangitis, including abdominal pain and jaundice. Although these findings may suggest graft rejection, rejection is typically accompanied by marked elevation of liver function enzymes. In contrast, in cholangitis in transplant recipients, results of liver function tests (with the possible exception of alkaline phosphatase levels) are often within the normal range. Definitive diagnosis of cholangitis in liver transplant recipients requires demonstration of aggregated neutrophils in bile duct biopsy specimens. Unfortunately, invasive studies of the biliary tract (either T-tube cholangiography or endoscopic retrograde cholangiopancreatography) may themselves lead to cholangitis. For this reason, many clinicians recommend an empirical trial of therapy with antibiotics covering gram-negative organisms and anaerobes before these procedures are undertaken as well as antibiotic coverage if procedures are eventually performed.

Reactivation of viral hepatitis is a common complication of liver transplantation (Chap. 304). Recurrent hepatitis B and C infections, for which transplantation may be performed, are problematic. To prevent hepatitis B virus reinfection, prophylaxis with an optimal antiviral agent or combination of agents (lamivudine, adefovir, entecavir) and hepatitis B immune globulin is currently recommended, although the optimal dose, route, and duration of therapy remain controversial. Success in preventing reinfection with hepatitis B virus has increased in recent years; in contrast, reinfection of the graft with hepatitis C virus occurs in all patients, with a variable time frame. Studies of aggressive pretransplantation treatment of selected recipients with antiviral agents and prophylactic/preemptive regimens are ongoing. However, early initiation of treatment for histologically documented disease with a combination of ribavirin and pegylated interferon has produced sustained responses at rates in the range of 25–40%. Several protease and polymerase inhibitors that block production of hepatitis C virus as well as a monoclonal antibody to the virus are undergoing preclinical and clinical trials.

As in other transplantation settings, reactivation disease with herpesviruses is common (Table 132-4). Herpesviruses can be transmitted in donor organs. Although CMV hepatitis occurs in ~4% of liver transplant recipients, it is usually not so severe as to require retransplantation. Without prophylaxis, CMV disease develops in the majority of seronegative recipients of organs from CMV-positive donors, but fatality rates are lower among liver transplant recipients than among lung or heart-lung transplant recipients. Disease due to CMV can also be associated with the vanishing bile duct syndrome after liver transplantation. Patients respond to treatment with ganciclovir; prophylaxis with oral forms of ganciclovir or high-dose acyclovir may decrease the frequency of disease. A role for HHV-6 reactivation in early posttransplantation fever and leukopenia has been proposed, although the more severe sequelae described in HSC transplantation are unusual. HHV-6 and HHV-7 appear to exacerbate CMV disease in this setting. EBV-LPD after liver transplantation shows a propensity for involvement of the liver, and such disease may be of donor origin. See previous sections for discussion of EBV infections in solid organ transplantation.

Pancreas Transplantation

Pancreas transplantation is most frequently performed together with or after kidney transplantation, although it may be performed alone. Transplantation of the pancreas can be complicated by early bacterial and yeast infections. Most pancreatic transplants are drained into the bowel, with the remaining transplants drained into the bladder. A cuff of duodenum is used in the anastomosis between the pancreatic graft and either the gut or the bladder. Bowel drainage poses a risk of early intraabdominal and allograft infections with enteric bacteria and yeasts. These infections can result in loss of the graft. Bladder drainage causes a high rate of urinary tract infection and sterile cystitis; however, infection can usually be cured with appropriate antimicrobial agents. In both procedures, prophylactic antimicrobial agents are commonly used at the time of surgery. Aggressive immunosuppression is associated with late-onset systemic viral and fungal infections; thus many centers administer an antifungal drug and an antiviral agent (ganciclovir or a congener) for prophylaxis.

Issues related to the development of CMV infection, EBV-LPD, and infections with opportunistic pathogens in patients receiving a pancreatic transplant are similar to those in other SOT recipients.

Miscellaneous Infections in Solid Organ Transplantation

INDWELLING IV CATHETER INFECTIONS

The prolonged use of indwelling IV catheters for administration of medications, blood products, and nutrition is common in diverse transplantation settings and poses a risk of local and bloodstream infections. Significant insertion-site infection is most commonly caused by *S. aureus*. Bloodstream infection most frequently develops within a week of catheter placement or in patients who become neutropenic. Coagulase-negative staphylococci are the most common isolates from the blood.

For further discussion of differential diagnosis and therapeutic options, see Chap. 86.

TUBERCULOSIS

The incidence of tuberculosis within the first 12 months after solid organ transplantation is greater than that observed after HSC transplantation (0.23–0.79%) and ranges broadly worldwide (1.2–15%), reflecting the prevalence of tuberculosis in local populations. Lesions suggesting prior tuberculosis on chest radiograph, older age, diabetes, chronic liver disease, GVHD, and intense immunosuppression are predictive of tuberculosis reactivation and development of disseminated disease in a host with latent disease. Tuberculosis has rarely been transmitted from the donor organ. In contrast to the low mortality rate among HSC transplant recipients, mortality rates among SOT patients are reported to be as high as 30%. Vigilance is indicated, as the presentation of disease is often extrapulmonary (gastrointestinal, genitourinary, central nervous, endocrine, musculoskeletal, laryngeal) and atypical, sometimes manifesting as a fever of unknown origin. A careful history and a direct evaluation of both the recipient and the donor prior to transplantation are optimal. Skin testing of the recipient with purified protein derivative may be unreliable because of chronic disease and/or immunosuppression, but newer cell-based assays that measure interferon and/or cytokine production may prove more sensitive in the future. Isoniazid toxicity has not been a significant problem except in the setting of liver transplantation. Therefore, appropriate prophylaxis should proceed. An assessment of the need to treat latent disease should include careful consideration of the possibility of a false-negative test result. Pending final confirmation of suspected tuberculosis, aggressive multidrug treatment in accordance with the guidelines of the CDC, the Infectious Diseases Society of America, and the American Thoracic Society is indicated because of the high mortality rates among these patients. Altered drug metabolism (e.g., upon co-administration of rifampin and certain immunosuppressive agents) can be managed with careful monitoring of drug levels and appropriate dose adjustment. Close follow-up of hepatic enzymes is warranted, particularly during treatment with isoniazid, pyrazinamide, and/or rifampin. Drug-resistant tuberculosis is especially problematic in these individuals (Chap. 165).

VIRUS-ASSOCIATED MALIGNANCIES

In addition to malignancy associated with gammaherpesvirus infection (EBV, KSHV) and simple warts (HPV), other tumors that are virus-associated or suspected of being virus-associated are more likely to develop in transplant recipients, particularly those who require long-term immunosuppression, than in the general population. The interval to tumor development is usually >1 year. Transplant recipients develop nonmelanoma skin or lip cancers that, in contrast to de novo skin cancers, have a high ratio of squamous cells to basal cells. HPV may play a major role in these lesions. Cervical and vulvar carcinomas, quite clearly associated with HPV, develop with increased frequency in female transplant recipients. Among renal transplant recipients, rates of melanoma are modestly increased and rates of cancers of the kidney and bladder are increased.

VACCINATION OF TRANSPLANT RECIPIENTS

In addition to receiving antibiotic prophylaxis, transplant recipients should be vaccinated against likely pathogens **(Table 132-7)**. In the case of HSC transplant recipients, optimal responses cannot be achieved until after immune reconstitution, despite previous immunization of both donor and recipient. Recipients of an allogeneic HSC transplant must be reimmunized if they are to be protected against pathogens. The situation is less clear-cut in the case of autologous transplantation. T and B cells in the peripheral blood may reconstitute the immune response if they are transferred in adequate numbers. However, cancer patients (particularly those with Hodgkin's disease, in whom vaccination has been extensively studied) who are undergoing chemotherapy do not respond normally to immunization, and titers of antibodies to infectious agents fall more rapidly than in healthy individuals. Therefore, even immunosuppressed patients who have not undergone HSC transplantation may need booster vaccine injections. If memory cells are specifically eliminated as part of a stem cell "cleanup" procedure, it will be necessary to reimmunize the recipient with a new primary series. Optimal times for immunizations of different transplant populations are being evaluated. Yearly immunization of household and other contacts (including health care personnel) against influenza benefits the patient by preventing local spread.

Table 132-7 Vaccination of Hematopoietic Stem Cell (HSC) Transplant or Solid Organ Transplant (SOT) Recipients

	Type of Transplantation	
Vaccine	**HSC Transplant**	**SOT***
Streptococcus pneumoniae, Haemophilus influenzae, Neisseria meningitidis	Immunize after transplantation. See CDC recommendations. (For pneumococcus, a new primary series may be indicated.)	Immunize before transplantation. See CDC recommendations. (For pneumococcus, a booster with polysaccharide vaccine every 5 years may be recommended.)

Influenza	Vaccinate in the fall. Vaccinate close contacts.	Vaccinate in the fall. Vaccinate close contacts.
Polio	Administer inactivated vaccine.	Administer inactivated vaccine.
Measles/mumps/rubella	Immunize 24 months after transplantation if GVHD is absent.	Immunize before transplantation with attenuated vaccine.
Diphtheria, pertussis, tetanus	Reimmunize after transplantation with primary series, DTaP. See CDC recommendations.	Immunize or boost before transplantation with Tdap; give boosters at 10 years or as required.
Hepatitis B and A	Reimmunize after transplantation. See CDC recommendations.	Immunize before transplantation.
Human papillomavirus	Recommendations are pending.	Recommendations are pending.

*Immunizations should be given before solid organ transplantation whenever possible.

Abbreviations: CDC, Centers for Disease Control and Prevention; DTaP, full-level diphtheria and tetanus toxoids and acellular pertussis, adsorbed; GVHD, graft-versus-host disease; Tdap, tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis.

Note: Recommendations from the CDC should be checked regularly as they frequently change upon receipt of new clinical information and new formulations of specific vaccines.

In the absence of compelling data as to optimal timing, it is reasonable to administer the pneumococcal and *H. influenzae* typePa_{CO2} b conjugate vaccines to both autologous and allogeneic HSC transplant recipients beginning 12 months after transplantation. A series that includes both the 13-valent pneumococcal conjugate vaccine and the 23-valent Pneumovax is now recommended (following CDC guidelines). The pneumococcal and *H. influenzae* type b vaccines are particularly important for patients who have undergone splenectomy. The *Neisseria meningitidis* polysaccharide conjugate vaccine (Menactra or Menveo) is also recommended. In addition, diphtheria, tetanus, acellular pertussis, and inactivated polio vaccines can all be given at these same intervals (12 months and, as required, 24 months after transplantation). Some authorities recommend a new primary series for tetanus/diphtheria/pertussis and inactivated polio vaccine beginning 12 months after transplantation. Vaccination to prevent hepatitis B and hepatitis A (both killed vaccines) also seems advisable. Live-virus measles/mumps/rubella (MMR) vaccine can be given to autologous HSC transplant recipients 24 months after transplantation and to most allogeneic HSC transplant recipients at the same point if they are not receiving maintenance therapy with immunosuppressive drugs and do not have ongoing GVHD. The risk of spread from a household contact is lower for MMR vaccine than for polio vaccine. Neither patients nor their household contacts should be vaccinated with vaccinia unless they have been exposed to the smallpox virus. Among patients who have active GVHD and/or are taking high maintenance doses of glucocorticoids, it may be prudent to avoid all live-virus vaccines.

In the case of SOT recipients, administration of all the usual vaccines and of the indicated booster doses should be completed before immunosuppression, if possible, to maximize responses. For patients taking immunosuppressive agents, the administration of pneumococcal vaccine should be repeated every 5 years. No data are available for the meningococcal vaccine, but it is probably reasonable to administer it along with the pneumococcal vaccine. *H. influenzae* conjugate vaccine is safe and should be efficacious in this population; therefore, its administration before transplantation is recommended. Booster doses of this vaccine are not recommended for adults. SOT recipients who continue to receive immunosuppressive drugs should not receive live-virus vaccines. A person in this group who is exposed to measles should be given measles immune globulin. Similarly, an immunocompromised patient who is seronegative for varicella and who comes into contact with a person who has chickenpox should be given varicella-zoster immune globulin as soon as possible, certainly within 96 h; if this is not possible, the patient should be started immediately on a 10- to 14-day course of acyclovir therapy. Upon the discontinuation of treatment, clinical disease may still occur in a small number of patients; thus vigilance is indicated. Rapid re-treatment with acyclovir should limit the symptoms of disease. Household contacts of transplant recipients can receive live attenuated VZV vaccine, but vaccinees should avoid direct contact with the patient if a rash develops. Virus-like particle (VLP) vaccines have been licensed for the prevention of infection with several HPV serotypes most commonly implicated in cervical and anal carcinomas and in anogenital and laryngeal warts. VLP vaccines are not live; however, no information is yet available about their immunogenicity or efficacy in transplant recipients.

Immunocompromised patients who travel may benefit from some but not all vaccines (Chaps. 122 and 123). In general, these patients should receive any killed or inactivated vaccine preparation appropriate to the area they are visiting; this recommendation includes the vaccines for Japanese encephalitis, hepatitis A and B, poliomyelitis, meningococcal infection, and typhoid. The live typhoid vaccines are not recommended for use in most immunocompromised patients, but inactivated or purified polysaccharide typhoid vaccine can be used. Live yellow fever vaccine should not be

administered. On the other hand, primary immunization or boosting with the purified-protein hepatitis B vaccine is indicated if patients are likely to be exposed. Patients who will reside for >6 months in areas where hepatitis B is common (Africa, Southeast Asia, the Middle East, Eastern Europe, parts of South America, and the Caribbean) should receive hepatitis B vaccine. Inactivated hepatitis A vaccine should also be used in the appropriate setting (Chap. 122). A combined vaccine is now available that provides dual protection against hepatitis A and hepatitis B. If hepatitis A vaccine is not administered, travelers should consider receiving passive protection with immune globulin (the dose depending on the duration of travel in the high-risk area).

FURTHER READINGS

Aguado JM et al: Tuberculosis in solid-organ transplant recipients: Consensus statement of the Group for the Study of Infection in Transplant Recipients (GESTRA) of the Spanish Society of Infectious Diseases and Clinical Microbiology. Clin Infect Dis 48:1276, 2009[PMID: 19320593]

Allain J-P et al: Transfusion-transmitted infectious diseases. Biologicals 37:71, 2009[PMID: 19231236]

Jiang M et al: The role of polyomaviruses in human disease. Virology 384:266, 2009[PMID: 18995875]

Kotton CN: Zoonoses in solid-organ and hematopoietic stem cell transplant recipients. Clin Infect Dis 44:857, 2007[PMID: 17304461]

Neofytos D et al: Epidemiology and outcome of invasive fungal infection in adult hematopoietic stem cell transplant recipients: Analysis of Multicenter Prospective Antifungal Therapy (PATH) Alliance registry. Clin Infect Dis 48:265, 2009[PMID: 19115967]

Pappas PG et al: Invasive fungal infections among organ transplant recipients: Results of the Transplant-Associated Infection Surveillance Network (TRANSNET). Clin Infect Dis 50:1101, 2010[PMID: 20218876]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PNEUMOCOCCAL INFECTIONS: INTRODUCTION

In the late nineteenth century, pairs of micrococci were first recognized in the blood of rabbits injected with human saliva by both Louis Pasteur working in France and George Sternberg, an American army physician. The important role of these micrococci in human disease was not appreciated at that time. By 1886, when the organism was designated "pneumokokkus" and *Diplococcus pneumoniae*, the pneumococcus had been isolated by many independent investigators, and its role in the etiology of pneumonia was well known. In the 1930s, pneumonia was the third leading cause of death in the United States (after heart disease and cancer) and was responsible for ~7% of all deaths both in the United States and in Europe. While pneumonia was caused by a host of pathogens, lobar pneumonia—a pattern more likely to be caused by the pneumococcus—accounted for approximately one-half of all pneumonia deaths in the United States in 1929. In 1974, the organism was reclassified as *Streptococcus pneumoniae*.

Microbiology

ETIOLOGIC AGENT

Pneumococci are spherical gram-positive bacteria of the genus *Streptococcus*. Within this genus, cell division occurs along a single axis, and bacteria grow in chains or pairs—hence the name *Streptococcus*, from the Greek *streptos*, meaning "twisted," and *kokkos*, meaning "berry." At least 22 streptococcal species are recognized and are divided further into groups based on their hemolytic properties. *S. pneumoniae* belongs to the α-hemolytic group that characteristically produces a greenish color on blood agar because of the reduction of iron in hemoglobin **(Fig. 134-1)**. The bacteria are fastidious and grow best in 5% CO_2 but require a source of catalase (e.g., blood) for growth on agar plates, where they develop mucoid (smooth/shiny) colonies. Pneumococci without a capsule produce colonies with a rough surface. Unlike that of other α-hemolytic streptococci, their growth is inhibited in the presence of optochin (ethyl hydrocuprein hydrochloride), and they are bile soluble.

Figure 134-1

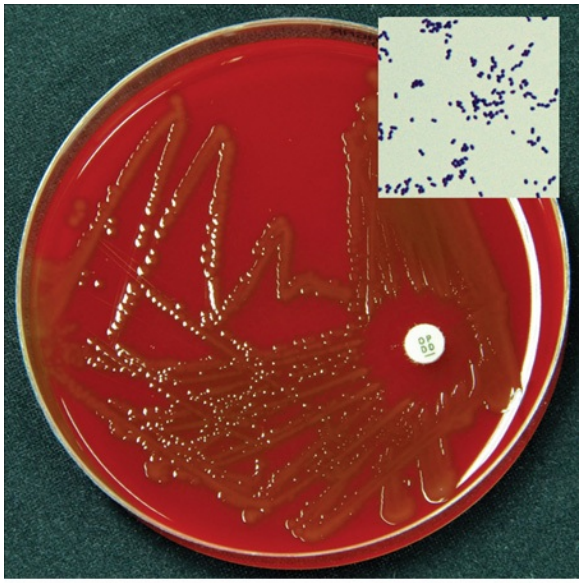

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pneumococci growing on blood agar, illustrating α hemolysis and optochin sensitivity (zone around optochin disk). *Inset:* Gram's stain, illustrating gram-positive diplococci. *(Photographs courtesy of Paul Turner, Shoklo Malaria Research Unit, Thailand.)*

In common with other gram-positive bacteria, pneumococci have a cell membrane beneath a cell wall, which in turn is covered by a polysaccharide capsule. Pneumococci are divided into serogroups or serotypes based on capsular polysaccharide structure, as distinguished with rabbit polyclonal antisera; capsules swell in the presence of specific antiserum (the Quellung reaction). The 91st and 92nd serotypes, 6C and 6D, have most recently been identified with monoclonal antibodies and by serologic, genetic, and biochemical means, respectively. Within the 92 serotypes there are 21 serogroups, each containing two to five serotypes with closely related capsules. The capsule protects the bacteria from phagocytosis by host cells in the absence of type-specific antibody and is arguably the most important determinant of pneumococcal virulence. Unencapsulated variants tend not to cause invasive disease.

Virulence Factors

Within the cytoplasm, cell membrane, and cell wall, many molecules that may play a role in pneumococcal pathogenesis and virulence have been identified **(Fig. 134-2)**. These proteins are often involved in direct interactions with host tissues or in concealment of the bacterial surface from host defense mechanisms. Pneumolysin is a secreted cytotoxin thought to result in cytolysis of cells and tissues, and LytA enhances pathogenesis. A number of cell wall proteins interfere with the complement pathway, thus inhibiting complement deposition and preventing lysis and/or opsonophagocytosis. The pneumococcal H

inhibitor (Hic) impedes the formation of C3 convertase, while pneumococcal surface protein C (PspC), also known as choline-binding protein A (CbpA), binds factor H and is thought to accelerate the breakdown of C3. PspA and CbpA inhibit the deposition of or degrade C3b. The numerous pneumococcal proteins thought to be involved in adhesion include the ubiquitous surface-anchored sialidase (neuraminidase) NanA, which cleaves sialic acid on host cells and proteins, and pneumococcal surface adhesin A (PsaA). Pili recently recognized by electron microscopy may also play an important role in binding to cells. Some of the antigens mentioned above are potential vaccine candidates (see "Prevention," below).

Figure 134-2

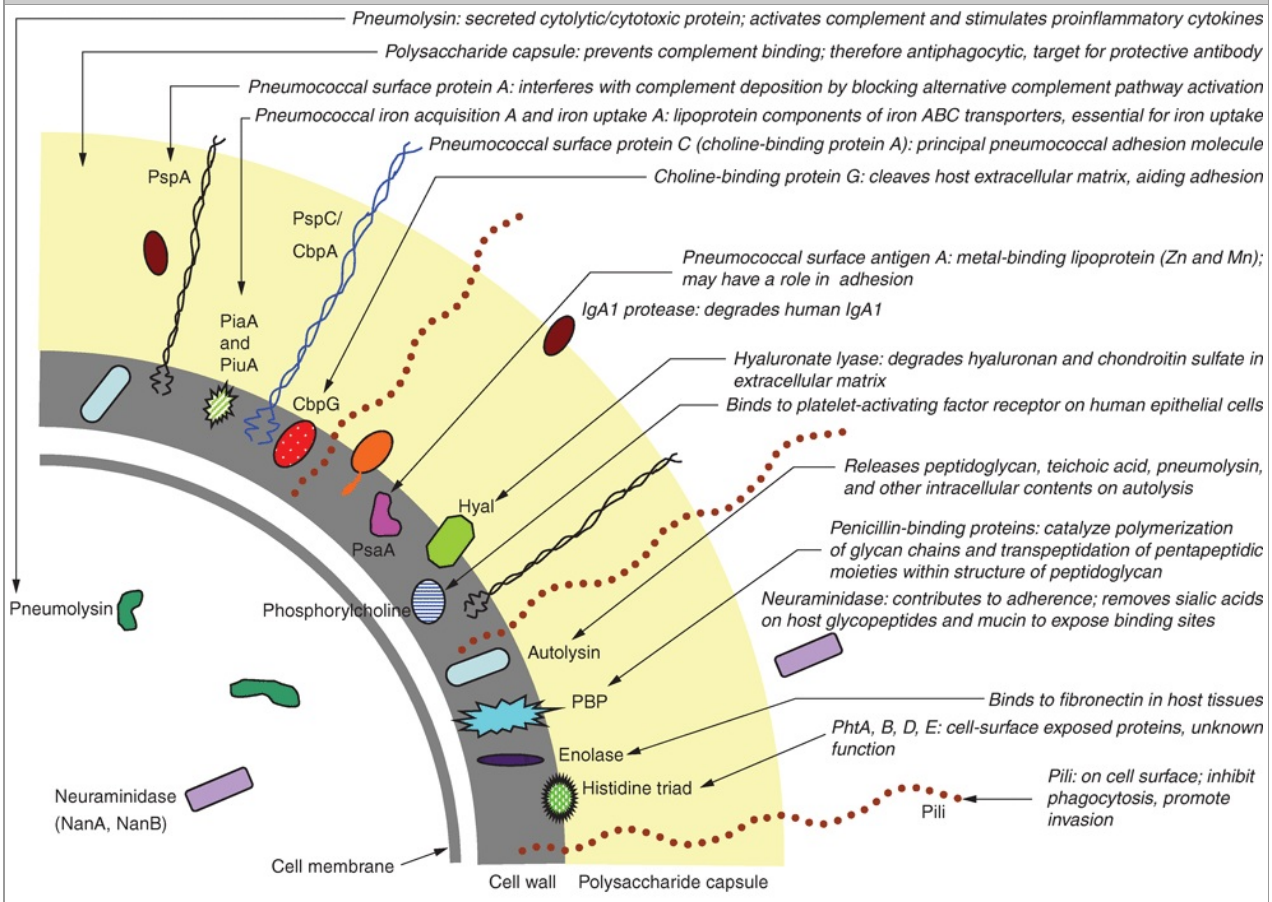

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of the pneumococcal cell surface, with key antigens and their roles highlighted.

Although the capsule surrounding the cell wall of *S. pneumoniae* is the basis for categorization by serotype, the behavior and pathogenic potential of a serotype may also be related to the genetic origin of the strain. Molecular typing is therefore of considerable interest. Initially, techniques such as pulsed-field gel electrophoresis were used to determine genetic relatedness; such techniques have been superseded by sequencing of housekeeping genes to define a clone (multilocus sequence typing, MLST). For *S. pneumoniae*, alleles at each of the loci *aroE*, *gdh*, *gki*, *recP*, *spi*, *xpt*, and *ddl* are sequenced and compared with all of the known alleles at that locus. Sequences identical to a known allele are assigned the same allele number, whereas those differing from any known allele—even at a single nucleotide site—are assigned new numbers. Software for assignment of alleles at each locus is available on the pneumococcal MLST web site (*http://spneumoniae.mlst.net/*), and the allelic profile of each isolate and its consequent sequence type are generated. With the advent of high-throughput and relatively inexpensive sequencing techniques, whole-genome sequencing will soon supersede MLST.

Epidemiology

Pneumococcal infections remain a significant global cause of morbidity and death, particularly among children and the elderly. Rapid and dramatic changes in the epidemiology of this disease during the past decade in several developed countries followed the licensure and routine childhood administration of pneumococcal polysaccharide–protein conjugate vaccine (PCV). With PCV introduction in developing and middle-income countries, additional profound changes in pneumococcal ecology and disease epidemiology are likely. The disease burden and serotype distribution in the PCV era may be different than expected because of concomitant secular trends in pneumococcal disease, the impact of antibiotic use on pneumococcal strain ecology, and surveillance system attributes that can themselves affect analysis of epidemiologic features.

Not all pneumococcal serotypes are equally likely to cause disease; serotype distribution varies by age, disease syndrome, and geography. Geographic differences may be driven by variation in the burden of disease rather than by true serotype distribution differences. Most data on serotype distribution are related to pediatric invasive pneumococcal disease (IPD, defined as infection of a normally sterile site); much less information on global distribution is available on disease in adults. Among children <5 years of age, five to seven serotypes cause >60% of IPD cases in most parts of the world, seven serotypes (1, 5, 6A, 6B, 14, 19F, and 23F) account for ~60% of cases in all areas of the world, but in any given region these seven serotypes may not all rank as the most common disease strains **(Fig. 134-3)**. Some serotypes (e.g., types 1 and 5) not only tend to cause disease in areas with a high disease burden but also cause waves of disease in lower-burden areas (e.g., Europe) or outbreaks (e.g., in military barracks; meningitis in sub-Saharan Africa). The broader range of serotypes causing

disease among adults than among children is apparent from a comparison of the coverage of existing multiserotype vaccines in different age groups. For example, data from the United States for 2006–2007 on the serotypes causing IPD indicated that a polysaccharide vaccine containing 23 serotypes (PPV23) would cover 84% of cases among children <5 years of age and 76% of those among persons 18–64 years of age but only 65% of those among persons ≥65 years of age.

Figure 134-3

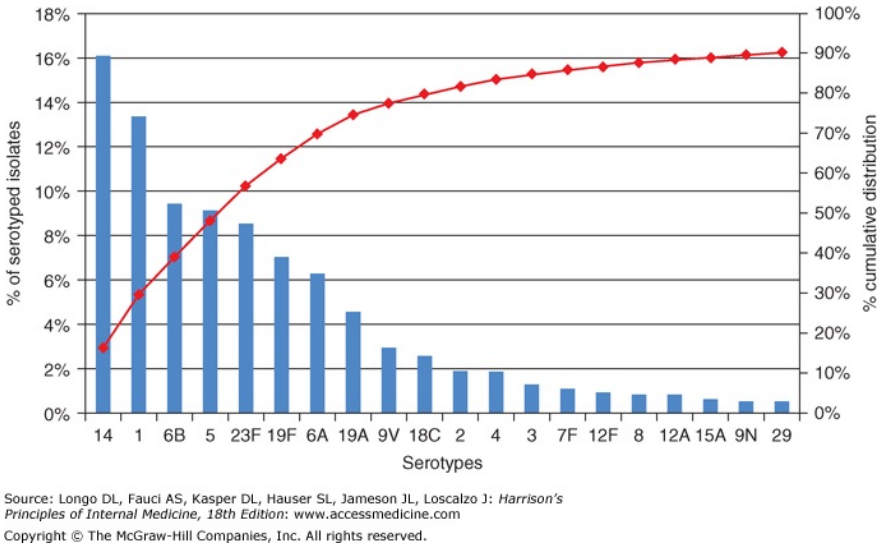

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Meta-analysis of available global serotype data, adjusted for regional disease incidence. The red line shows cumulative incidence, as indicated on the right-hand Y axis. *(Source: Global Serotype Project Report for the Pneumococcal Advance Market Commitment Target Product Profile; available at http://www.vaccinearmc.org/files/TPP_Codebook.pdf.)*

Pneumococci are intermittent inhabitants of the healthy human nasopharynx and are transmitted by respiratory droplets. In children, pneumococcal nasopharyngeal ecology varies by geographic region, socioeconomic status, climate, degree of crowding, and particularly intensity of exposure to other children, with children in day-care settings having higher rates of colonization. In developed-world settings, children serve as the major vectors of pneumococcal transmission. By 1 year of age, ~50% of children have had at least one episode of pneumococcal colonization. Cross-sectional prevalence data show rates of pneumococcal carriage ranging from 20% to 50% for children <5 years of age and from 5% to 15% for young and middle-aged adults; **Fig. 134-4** shows relevant data from the United Kingdom. Data on colonization rates among healthy elderly individuals are limited. In developing-world settings, pneumococcal acquisition occurs much earlier, sometimes within the first few days after birth, and nearly all infants have had at least one episode of colonization by 2 months of age. Cross-sectional studies show that up to the age of 5 years, 70–90% of children carry *S. pneumoniae* in the nasopharynx, and a significant proportion of adults (sometimes >40%) are also colonized. Their high rates of colonization make adults an important source of transmission and may affect community transmission dynamics.

Figure 134-4

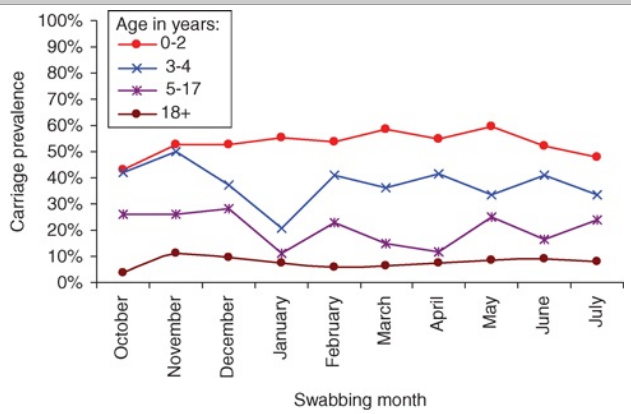

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prevalence of pneumococcal carriage in adults and children resident in the United Kingdom who had nasopharyngeal swabs collected monthly for 10 months (no seasonal trend; *t* test trend, >.05). *(Data adapted from D Goldblatt et al, 2005.)*

IPD develops when *S. pneumoniae* invades the bloodstream and seeds other organs or directly reaches the cerebrospinal fluid (CSF) by local extension. Pneumonia may follow aspiration of pneumococci, although only 10–30% of such cases are associated with a positive blood culture (and thus contribute to the measured burden of IPD). The dramatic variation of IPD rates with age is illustrated by data from the United States for 1998–1999, a period prior to PCV introduction. Rates of IPD were highest among children <2 years of age and among adults ≥65 years of age (188 and 60 cases per 100,000, respectively; **Fig. 134-5**). Since the introduction of PCV, IPD rates among infants and children in the United States have fallen by >75%, a decrease driven by the near elimination of vaccine-serotype IPD. A similar impact of PCV on vaccine-serotype IPD rates has been consistently observed in countries where PCV has been introduced into the routine pediatric

vaccination schedule. However, changes in the non-vaccine-serotype IPD rate in various countries have been heterogeneous; the interpretation of this heterogeneity is a complex issue. In the United States, Canada, and Australia, rates of non-vaccine-serotype IPD have increased but the magnitude of the increase is generally small relative to the substantial reductions in vaccine-serotype IPD. In contrast, in other settings (e.g., Alaska Native communities and the United Kingdom), the reduction in vaccine-serotype IPD has been offset by notable increases in rates of disease caused by non-vaccine serotypes. Explanations for the heterogeneity of findings include replacement disease resulting from vaccine pressure, changes in clinical case investigation, secular trends unrelated to PCV use, antibiotic pressure selecting for resistant organisms, changes in surveillance or reporting systems, rapidity of introduction, and inclusion of a catch-up campaign. The roles and relative importance of these hypothesized mechanisms in driving the observed non-vaccine-serotype IPD trends and in explaining the observed heterogeneity among populations are not yet fully understood.

Figure 134-5

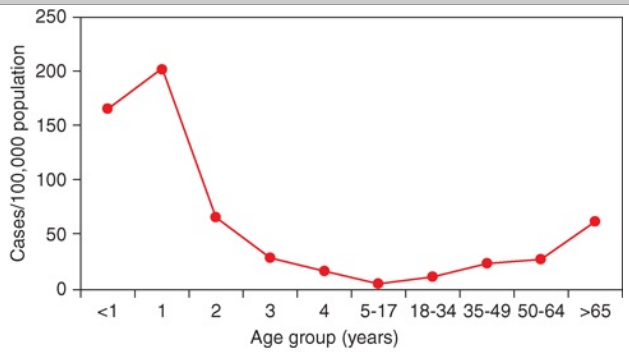

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rates of invasive pneumococcal disease before the introduction of pneumococcal conjugate vaccine, by age group: United States, 1998. *[Source: CDC, Active Bacterial Core Surveillance/Emerging Infectious Program Network, 2000. Data adapted from MMWR 49(RR-9), 2000.]*

Pneumonia is the most common of the serious pneumococcal disease syndromes and poses special challenges from a clinical and public health perspective. Most cases of pneumococcal pneumonia are not associated with bacteremia, and in these cases a definitive etiologic diagnosis is difficult. As a result, estimates of disease burden focus primarily on IPD rates and fail to include the major portion of the burden of serious pneumococcal disease. Among children, PCV trials designed to collect efficacy data on syndrome-based outcomes (e.g., radiographically confirmed pneumonia, clinically diagnosed pneumonia) have revealed the burden of culture-negative pneumococcal pneumonia.

The case-fatality ratios (CFRs) for pneumococcal pneumonia and IPD vary by age, underlying medical condition, and access to care. In addition, the CFR for pneumococcal pneumonia varies with the severity of disease at presentation (rather than according to whether or not the pneumonia episode is associated with bacteremia) and with the patient's age (from <5% among hospitalized patients 18–44 years old to >12% among those >65 years old, even when appropriate and timely management is available). Notably, the likelihood of death in the first 24 h of hospitalization did not change substantially with the introduction of antibiotics; this surprising observation highlights the fact that the pathophysiology of severe pneumococcal pneumonia among adults reflects a rapidly progressive cascade of events that often unfolds irrespective of antibiotic administration. Management in an intensive care unit can provide critical support for the patient through the acute period, with lower CFRs.

Rates of pneumococcal disease vary by season, with higher rates in colder than in warmer months in temperate climates; by gender, with males more often affected than females; and by risk group, with risk factors including underlying medical conditions, behavioral issues, and ethnic group. In the United States, some Native American populations (including Alaska natives) and African Americans have higher rates of disease than the general population; the increased risk is probably attributable to socioeconomic conditions and the prevalence of underlying risk factors for pneumococcal disease. Medical conditions that increase the risk of pneumococcal infection are listed in **Table 134–1**. Outbreaks of disease are well recognized in crowded settings with susceptible individuals, such as infant day-care facilities, military barracks, and nursing homes. Furthermore, there is a clear association between preceding viral respiratory disease (especially but not exclusively influenza) and risk of secondary pneumococcal infections. The significant role of pneumococcal pneumonia in the morbidity and mortality associated with seasonal and pandemic influenza is increasingly recognized.

Table 134–1 Clinical Risk Groups for Pneumococcal Infection

Clinical Risk Group	Examples
Asplenia or splenic dysfunction	Sickle cell disease, celiac disease
Chronic respiratory disease	Chronic obstructive pulmonary disease, bronchiectasis, cystic fibrosis, interstitial lung fibrosis, pneumoconiosis, bronchopulmonary dysplasia, aspiration risk, neuromuscular disease (e.g., cerebral palsy), severe asthma
Chronic heart disease	Ischemic heart disease, congenital heart disease, hypertension with cardiac complications, chronic heart failure
Chronic kidney disease	Nephrotic syndrome, chronic renal failure, renal transplantation
Chronic liver disease	Cirrhosis, biliary atresia, chronic hepatitis
Diabetes mellitus	Diabetes mellitus requiring insulin or oral hypoglycemic drugs
Immunocompromise/immunosuppression	HIV infection, common variable immunodeficiency, leukemia, lymphoma, Hodgkin's disease, multiple myeloma, generalized malignancy, chemotherapy, organ or bone marrow transplantation, systemic glucocorticoid treatment for >1 month at a dose equivalent to ≥20 mg/day (children, ≥1 mg/kg per day)
Cochlear implants	. . .
Cerebrospinal fluid leaks	. . .

Miscellaneous	Infancy and old age; prior hospitalization; alcoholism; malnutrition; cigarette smoking; day-care center attendance; residence in military training camps, prisons, homeless shelters

Note: Groups for whom pneumococcal vaccines are recommended by the Advisory Committee on Immunization Practices can be found at *http://www.cdc.gov/vaccines/recs/schedules/default.htm.*

Reduced pneumococcal susceptibility to penicillin was first noted in 1967, but not until the 1990s did reduced antibiotic susceptibility emerge as a significant clinical and public health issue, with an increasing prevalence of pneumococcal isolates resistant to single or multiple classes of antibiotics and a rising absolute magnitude of minimal inhibitory concentrations (MICs). Strains with reduced susceptibility to penicillin G, cefotaxime, ceftriaxone, macrolides, and other antibiotics are now found worldwide and account for a significant proportion of disease-causing strains in many locations, especially among children. Vancomycin resistance has not yet been observed in clinical pneumococcal strains. Lack of antimicrobial susceptibility is clearly related to a subset of serotypes, many of which disproportionately cause disease among children. The vicious cycle of antibiotic exposure, selection of resistant organisms in the nasopharynx, and transmission of these organisms within the community leading to difficult-to-treat infections and increased antibiotic exposure has been interrupted to some extent by the introduction and routine use of PCV. The clinical implications of pneumococcal antimicrobial nonsusceptibility are addressed below in the section on treatment.

Pathogenesis

Pneumococci colonize the human nasopharynx from an early age; colonization acquisition events are generally described as asymptomatic, but evidence exists to associate acquisition with mild respiratory symptoms, especially in the very young. From the nasopharynx, the bacteria spread either via the bloodstream to distant sites (e.g., brain, joint, bones, peritoneal cavity) or locally to mucosal surfaces where they can cause otitis media or pneumonia. Direct spread from the nasopharynx to the central nervous system (CNS) can occur in rare cases of skull base fracture, although most cases of pneumococcal meningitis are secondary to hematogenous spread. Pneumococci can cause disease in almost any organ or part of the body; however, otitis media, pneumonia, bacteremia, and meningitis are most common. Colonization is a relatively frequent event, yet disease is rare. In the nasopharynx, pneumococci survive in mucus secreted by epithelial cells, where they can avoid local immune factors such as leukocytes and complement. The mucus itself is a component of local defense mechanisms, and the flow of mucus (driven in part by cilia in what is known as the *mucociliary escalator*) effects mechanical clearance of pneumococci. While many colonization episodes are of short duration, longitudinal studies in adults and children have revealed persistent colonization with a specific serotype over many months. Colonization eventually results in the development of capsule-specific serum IgG, which is thought to play a role in mediating clearance of bacteria from the nasopharynx. IgG antibodies to surface exposed cell wall or secreted proteins also appear in the circulation in an age-dependent fashion or after colonization; the biological role of these antibodies is less clear. Recent acquisition of a new colonizing serotype is more likely to be associated with subsequent invasion, presumably as a result of the absence of type-specific immunity. Intercurrent viral infections make the host more susceptible to pneumococcal colonization, and pneumococcal disease in a colonized individual often follows perturbation of the nasopharyngeal mucosa by such infections. Local cytokine production after a viral infection is thought to upregulate adhesion factors in the respiratory epithelium, allowing pneumococci to adhere via a variety of surface adhesin molecules, including PsaA, PspA, CbpA, PspC, Hyl, pneumolysin, and the neuraminidases (Fig. 134-2). Adhesion coupled with inflammation induced by pneumococcal factors such as peptidoglycans and teichoic acids results in invasion. It is the inflammation induced by various bacterium-derived factors that is responsible for the pathology associated with pneumococcal infection. Cell wall–derived teichoic acids and peptidoglycans induce a variety of cytokines, including the proinflammatory cytokines interleukin (IL) 1, IL-6, and tumor necrosis factor (TNF), and activate complement via the alternative pathway. Polymorphonuclear leukocytes are thus attracted, and an intense inflammatory response is initiated. Pneumolysin is also important in local pathology, inducing proinflammatory cytokine production by local monocytes.

The pneumococcal capsule, consisting of polysaccharides with antiphagocytic properties due to resistance to the deposition of complement, plays an important role in pathogenesis. While most capsular types can cause human disease, certain capsular types are more commonly isolated from sites of infection. The reason for the dominance of some serotypes over others in IPD, as depicted in Fig. 134-3, is unclear.

Host Defense Mechanisms

Innate Immunity

As described above, intact respiratory epithelium and a host of nonspecific or innate immune factors (e.g., mucus, splenic function, complement, neutrophils, and macrophages) constitute the first line of defense against pneumococci. Physical factors such as the cough reflex and the mucociliary escalator are important in clearing bacteria from the lungs. Immunologic factors are critical as well: C-reactive protein (CRP) binds phosphorylcholine in the pneumococcal cell wall, inducing complement activation and leading to bacterial clearance; Toll-like receptor 2 (TLR2) recognizes both pneumococcal lipoteichoic acid and cell wall peptidoglycan; and in animal models, the absence of host TLR2 leads to more severe infection and impaired clearance of nasopharyngeal colonization. TLR4 appears to be necessary for the proinflammatory effect of pneumolysin on macrophages. The importance of TLR recognition is underlined by descriptions of an inherited deficiency of human IL-1 receptor–associated kinase 4 (IRAK-4) that manifests as an unusual susceptibility to infection with bacteria, including *S. pneumoniae*. IRAK-4 is essential for the normal functioning of several TLRs. Other factors that interfere with these nonspecific mechanisms (e.g., viral infections, cystic fibrosis, bronchiectasis, complement deficiency, and chronic obstructive pulmonary disease) all predispose to the development of pneumococcal pneumonia. Patients who lack a spleen or have abnormal splenic function (e.g., persons with sickle cell disease) are at high risk of developing overwhelming pneumococcal disease.

Acquired Immunity

Acquired immunity induced via contact following colonization or through cross-reactive antigens rests largely on the development of serum IgG antibody specific for the pneumococcal capsular polysaccharide. Nearly all polysaccharides are T cell–independent antigens; B cells can make antibodies to such antigens without T cell help. However, in children <1–2 years old, such B cell responses are poorly developed. This delayed ontogeny of capsule-specific IgG in young children is associated with susceptibility to pneumococcal infection (Fig. 134-5). The extremely high risk of pneumococcal infection in the absence of serum immunoglobulin (i.e., in conditions such as agammaglobulinemia) highlights the important role of capsular antibody in protection against disease. Each serotype's capsule is chemically distinct; thus immunity tends to be serotype specific, although some cross-immunity exists. For example, conjugate vaccine–induced antibodies to serotype 6B prevent infection due to serotype 6A. However, cross-protection against serotypes within serogroups is not universal; for instance, antibodies to serotype 19F do not appear to confer protection against disease caused by serotype 19A. Antibodies to surface exposed or secreted pneumococcal proteins (such as pneumolysin, PsaA, and PspA) also appear in the circulation

with increasing age of the host, but their functional significance remains unclear. Although data from murine models suggest that CD4+ T cells may play a role in preventing pneumococcal colonization and disease, these data have not yet been replicated in humans.

Approach to the Patient: Pneumococcal Infections

There is no pathognomonic presentation of pneumococcal disease; patients may present with a range of syndromes and with more than one clinical syndrome (e.g., pneumonia and meningitis). *S. pneumoniae* can infect nearly any body tissue, manifesting as disease ranging in severity from mild and self-limited to life-threatening. The differential diagnosis of common clinical syndromes such as pneumonia, otitis media, fever of unknown origin, and meningitis should always include pneumococcal infection. A microbiologically confirmed diagnosis is made in only a minority of pneumococcal cases since, in most circumstances (and especially in pneumonia and otitis media), fluid from the site of infection is not available for etiologic determination. Empirical therapy that includes appropriate treatment for *S. pneumoniae* is often indicated.

Algorithms for assessment and management of ill children have been developed for use in the developing world or in other settings where evaluation by a trained physician may not be feasible. Children who present with ominous signs such as inability to drink, convulsions, lethargy, and severe malnutrition are categorized as having very severe disease without further evaluation by the community health care worker, are given antibiotics, and are immediately referred to a hospital for diagnosis and management. Children who present with cough and tachypnea (the latter defined according to specific age strata) are further stratified into severity categories based on the presence or absence of lower chest wall indrawing and are managed accordingly with either antibiotics alone or antibiotics and referral to a hospital facility. Children with cough but no tachypnea are categorized as having a nonpneumonia respiratory illness.

Clinical Manifestations

The clinical manifestations of pneumococcal disease depend on the site of infection and the duration of illness. Clinical syndromes are classified as noninvasive (e.g., otitis media and nonbacteremic pneumonia) or invasive (e.g., bacteremic pneumonia). The pathogenesis of noninvasive illness involves contiguous spread from the nasopharynx or skin; invasive disease involves infection of a normally sterile body fluid or follows bacteremia.

Pneumonia

Pneumonia is the most common serious pneumococcal syndrome and is considered invasive when associated with a positive blood culture. Pneumococcal pneumonia can present as a mild community-acquired infection at one extreme and as a life-threatening disease requiring intubation and intensive support at the other.

PRESENTING MANIFESTATIONS

The presentation of pneumococcal pneumonia does not reliably distinguish it from pneumonia of other etiologies. In a subset of cases, pneumococcal pneumonia is recognized at the outset as associated with a viral upper respiratory infection and is characterized by the abrupt onset of cough and dyspnea accompanied by fever, shaking chills, and myalgias. The cough evolves from nonpurulent to productive of sputum that is purulent and sometimes tinged with blood. Patients may describe stabbing pleuritic chest pain and significant dyspnea indicating involvement of the parietal pleura. Among the elderly, the presenting clinical symptoms may be less specific, with confusion or malaise but without fever or cough. In such cases, a high index of suspicion is required because failure to treat pneumococcal pneumonia promptly in an elderly patient is likely to result in rapid evolution of the infection, with increased severity, morbidity, and risk of death.

FINDINGS ON PHYSICAL EXAMINATION

The clinical signs associated with pneumococcal pneumonia among adults include tachypnea (>30 breaths/min) and tachycardia, hypotension in severe cases, and fever in most cases (although not in all elderly patients). Respiratory signs are varied, including dullness to percussion in areas of the chest with significant consolidation, crackles on auscultation, reduced expansion of the chest in some cases as a result of splinting to reduce pain, bronchial breathing in a minority of cases, pleural rub in occasional cases, and cyanosis in cases with significant hypoxemia. Among infants with severe pneumonia, chest wall indrawing and nasal flaring are common. Nonrespiratory findings can include upper abdominal pain if the diaphragmatic pleura is involved as well as mental status changes, particularly confusion among elderly patients.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of pneumococcal pneumonia includes cardiac conditions such as myocardial infarction and heart failure with atypical pulmonary edema; pulmonary conditions such as atelectasis; and pneumonia caused by viral pathogens, mycoplasmas, *Haemophilus influenzae*, *Klebsiella pneumoniae*, *Staphylococcus aureus*, *Legionella*, or (in HIV-infected and otherwise immunocompromised hosts) *Pneumocystis*. In cases including abdominal symptoms, the differential diagnosis includes cholecystitis, appendicitis, perforated peptic ulcer disease, and subphrenic abscesses. The challenge in cases with abdominal symptoms is to remember to include pneumococcal pneumonia—a nonabdominal process—in the differential diagnosis.

DIAGNOSIS

Some authorities advocate treating uncomplicated, nonsevere, community-acquired pneumonia without determining the microbiologic etiology, given that this information is unlikely to alter clinical management. However, efforts to identify the cause of pneumonia are important when the disease is more severe and when the diagnosis of pneumonia is not clearly established. The gold standard for etiologic diagnosis of pneumococcal pneumonia is pathologic examination of lung tissue. In lieu of that procedure, evidence of an infiltrate on chest radiography warrants a diagnosis of pneumonia. However, cases of pneumonia without radiographic evidence do occur. An infiltrate can be absent either early in the course of the illness or with dehydration; upon rehydration, an infiltrate usually appears. The radiographic appearance of pneumococcal pneumonia is varied; it classically consists of lobar or segmental consolidation **(Fig. 134-6)** but in some cases is patchy. More than one lobe is involved in ~30% of cases. Consolidation may be associated with a small pleural effusion or empyema in complicated cases. In children, "round pneumonia," a distinctly spherical consolidation on chest radiography, is associated with a pneumococcal etiology. Round pneumonia is uncommon in adults. *S. pneumoniae* is not the only cause of such lesions; other causes, especially cancer, should be considered.

Figure 134-6

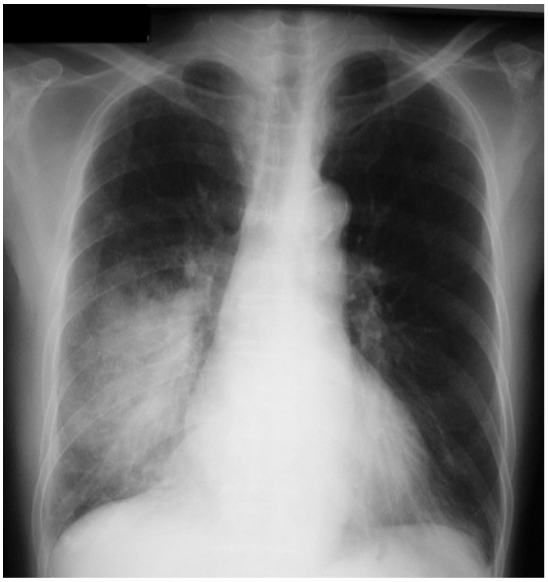

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiograph depicting classic lobar pneumonia in the right lower lobe of an elderly patient's lung.

Blood drawn from patients with suspected pneumococcal pneumonia can be used for supportive or definitive diagnostic tests. Blood cultures are positive for pneumococci in a minority (<30%) of cases of pneumococcal pneumonia. Nonspecific findings include an elevated polymorphonuclear leukocyte count (>15,000/μL in most cases and upward of 40,000/μL in some), leukopenia in <10% of cases (a poor prognostic sign associated with a fatal outcome), and elevated values in liver function tests (e.g., both conjugated and unconjugated hyperbilirubinemia). Anemia, low serum albumin levels, hyponatremia, and elevated serum creatinine levels are all found in ~20–30% of patients.

Urinary pneumococcal antigen assays have facilitated etiologic diagnosis. In adults, among whom the prevalence of pneumococcal nasopharyngeal colonization is relatively low, a positive pneumococcal urinary antigen test has a high predictive value. The same is not true for children, in whom a positive urinary antigen test can reflect the mere presence of *S. pneumoniae* in the nasopharynx.

Most cases of pneumococcal pneumonia are diagnosed by Gram's staining and culture of sputum. The utility of a sputum specimen is directly related to its quality and the patient's antibiotic treatment status.

COMPLICATIONS

Empyema is the most common focal complication of pneumococcal pneumonia, occurring in <5% of cases. When fluid in the pleural space is accompanied by fever and leukocytosis (even low-grade) after 4–5 days of appropriate antibiotic treatment for pneumococcal pneumonia, empyema should be considered. Parapneumonic effusions are more common than empyema, representing a self-limited inflammatory response to pneumonia. Pleural fluid with frank pus, bacteria (detected by microscopic examination), or a pH of ≤7.1 indicates empyema and demands aggressive and complete drainage, usually through chest tube insertion.

Meningitis

Pneumococcal meningitis typically presents as a pyogenic condition that is clinically indistinguishable from meningitis of other bacterial etiologies. Meningitis can be the primary presenting pneumococcal syndrome or a complication of other conditions such as skull fracture, otitis media, bacteremia, or mastoiditis. Now that *H. influenzae* type b vaccine is routinely used, *S. pneumoniae* and *Neisseria meningitidis* are the most common bacterial causes of meningitis in both adults and children. Pyogenic meningitis, including that due to *S. pneumoniae*, is associated clinically with findings that include severe, generalized, gradual-onset headache, fever, and nausea as well as specific CNS manifestations such as stiff neck, photophobia, seizures, and confusion. Clinical signs include a toxic appearance, altered consciousness, bradycardia, and hypertension indicative of increased intracranial pressure. A small proportion of adult patients have Kernig's or Brudzinski's sign or cranial nerve palsies (particularly of the 3rd and 6th cranial nerves).

A definitive diagnosis of pneumococcal meningitis rests on the examination of CSF for (1) evidence of turbidity (visual inspection); (2) elevated protein level, elevated white blood cell count, and reduced glucose concentration (quantitative measurement); and (3) specific identification of the etiologic agent (culture, Gram's staining, antigen testing, or PCR). A blood culture positive for *S. pneumoniae* in conjunction with clinical manifestations of meningitis is also considered confirmatory. Among adults, detection of pneumococcal antigen in urine is considered highly specific because of the low prevalence of nasopharyngeal colonization in this age group.

The mortality rate for pneumococcal meningitis is ~20%. In addition, up to 50% of survivors experience acute or chronic complications, including deafness, hydrocephalus, and mental retardation in children and diffuse brain swelling, subarachnoid bleeding, hydrocephalus, cerebrovascular complications, and hearing loss in adults.

Other Invasive Syndromes

S. pneumoniae can cause other invasive syndromes involving virtually any body site. These syndromes include primary bacteremia without other sites of infection (bacteremia without a source; occult bacteremia), osteomyelitis, septic arthritis, endocarditis, pericarditis, and peritonitis. The essential diagnostic approach is collection of fluid from the site of infection by sterile technique and examination by Gram's staining, culture, and—when relevant—capsular antigen assay or PCR. Hemolytic-uremic syndrome can complicate invasive pneumococcal disease.

Noninvasive Syndromes

The two major noninvasive syndromes caused by *S. pneumoniae* are sinusitis and otitis media; the latter is the most common pneumococcal syndrome and most often affects young children. The manifestations of otitis media include the acute

onset of severe pain, fever, deafness, and tinnitus, most frequently in the setting of a recent upper respiratory tract infection. Clinical signs include a red, swollen, often bulging tympanic membrane with reduced movement on insufflation or tympanography. Redness of the tympanic membrane is not sufficient for the diagnosis of otitis media.

Pneumococcal sinusitis is also a complication of upper respiratory tract infections and presents with facial pain, congestion, fever, and—in many cases—persistent nighttime cough. A definitive diagnosis is made by aspiration and culture of sinus material; however, presumptive treatment is most commonly initiated after application of a strict set of clinical diagnostic criteria.

Treatment: Pneumococcal Infections

Historically, the activity of penicillin against pneumococci made parenteral penicillin G the drug of choice for disease caused by susceptible organisms, including community-acquired pneumonia. For susceptible strains, penicillin G remains the most commonly used agent, with daily doses ranging from 50,000 U/kg for minor infections to 300,000 U/kg for meningitis. Other parenteral β-lactam drugs, such as ampicillin, cefotaxime, ceftriaxone, and cefuroxime, can be used against penicillin-susceptible strains but offer little advantage over penicillin. Macrolides and cephalosporins are alternatives for penicillin-allergic patients. While agents such as clindamycin, tetracycline, and trimethoprim-sulfamethoxazole exhibit some activity against pneumococci, resistance to these agents is frequently encountered in different parts of the world.

Penicillin-resistant pneumococci were first described in the mid-1960s, at which point tetracycline- and macrolide-resistant strains had already been reported. Multidrug-resistant strains were first described in the 1970s, but it was during the 1990s that pneumococcal drug resistance reached pandemicproportions. The use of antibiotics selects for resistant pneumococci, and strains resistant to β-lactam agents and to multiple drugs are now found all over the world. The emergence of high rates of macrolide and fluoroquinolone resistance has also been described.

The molecular basis of penicillin resistance in *S. pneumoniae* is the alteration of penicillin-binding protein (PBP) genes by transformation and horizontal transfer of DNA from related streptococcal species. Such alteration of PBPs results in lower affinity for penicillins. Depending on the specific PBP(s) and the number of PBPs altered, the level of resistance ranges from intermediate to high. For many years, penicillin susceptibility breakpoints have been defined by MICs as follows: susceptible, ≤0.06 μg/mL; intermediate, 0.12–1.0 μg/mL; and resistant, ≥2.0 μg/mL. However, in vitro results often were not predictive of the response of a patient to treatment for pneumococcal diseases other than meningitis. New recommendations have been based on the revised penicillin G breakpoints established in 2008 by the Clinical and Laboratory Standards Institute. For IV treatment of meningitis with at least 24 million units per day in 8 divided doses, the susceptibility breakpoint remains ≤0.06 μg/mL, and MICs of ≥0.12 μg/mL indicate resistance. For IV treatment of nonmeningeal infections with 12 million units per day in 6 divided doses, the breakpoints are ≤2 μg/mL for susceptible organisms, 4 μg/mL for intermediate organisms, and ≥ 8 μg/mL for resistant organisms; a dosage of 18–24 million units per day is recommended for strains with MICs in the intermediate category. The original breakpoints remain the same for oral treatment of nonmeningeal infections with penicillin V.

Although guidelines for antibiotic therapy should be driven in part by local patterns of resistance, guidelines from national organizations in many countries (e.g., the Infectious Diseases Society of America/American Thoracic Society, the British Thoracic Society, and the European Respiratory Society) lay out evidence-based approaches. The following guidelines for individual sepsis syndromes are based on those advocated by the American Academy of Pediatrics and published in the 2009 *Red Book*.

MENINGITIS LIKELY OR PROVEN TO BE DUE TO *S. PNEUMONIAE*

As a result of the increased prevalence of resistant pneumococci, first-line therapy for persons ≥1 month of age is a combination of vancomycin (adults, 30–60 mg/kg per day; infants and children, 60 mg/kg per day) and cefotaxime (adults, 8–12 g/d in 4–6 divided doses; children, 225–300 mg/kg per day in 1 dose or 2 divided doses) or ceftriaxone (adults, 4 g/d in 1 dose or 2 divided doses; children, 100 mg/kg per day in 1 dose or 2 divided doses). If children are hypersensitive to β-lactam agents (penicillins and cephalosporins), rifampin (adults, 600 mg/d; children, 20 mg/d in 1 dose or 2 divided doses) can be substituted for cefotaxime or ceftriaxone. A lumbar puncture should be considered after 48 h if the organism is not susceptible to penicillin and information on cephalosporin sensitivity is not yet available, if the patient's clinical condition does not improve or deteriorates, or if dexamethasone has been administered and may be compromising clinical evaluation. When antibiotic sensitivity data become available, treatment should be modified accordingly. If the isolate is sensitive to penicillin, vancomycin can be discontinued and penicillin can replace the cephalosporin, or cefotaxime or ceftriaxone can be continued alone. If the isolate displays any resistance to penicillin but is susceptible to the cephalosporins, vancomycin can be discontinued and cefotaxime or ceftriaxone continued. If the isolate exhibits any resistance to penicillin and is not susceptible to cefotaxime and ceftriaxone, vancomycin and high-dose cefotaxime or ceftriaxone can be continued; rifampin may be added as well if the isolate is susceptible and the patient's clinical condition is worsening, if the CSF remains positive for bacteria, or if the MIC of the cephalosporin in question against the infecting strain is high. Some physicians advocate the use of glucocorticoids in children >6 months old, but this recommendation remains controversial and is not universally considered the standard of care. Glucocorticoids significantly reduce rates of mortality, severe hearing loss, and neurologic sequelae in adults and should be administered to those with community-acquired bacterial meningitis. If dexamethasone is given to either adults or children, it should be administered before or in conjunction with the first antibiotic dose.

INVASIVE INFECTIONS (EXCLUDING MENINGITIS)

In previously well children with noncritical illness, antibiotic therapy with a recommended antibiotic should be instigated at the following dosages: penicillin G, 250,000–400,000 units/kg per day (in divided doses 4–6 h apart); cefotaxime, 75–100 mg/d (doses 8 h apart); or ceftriaxone, 50–75 mg/d (doses 12–24 h apart). For critically ill children, including those with myocarditis or multilobular pneumonia with hypoxia or hypotension, vancomycin may be added if the isolate may possibly be resistant to β-lactam drugs, with its use reviewed once susceptibility data become available. If the organism is resistant to β-lactam agents, therapy should be modified on the basis of clinical response and susceptibility to other antibiotics. Clindamycin or vancomycin can be used as a first-line agent for children with severe β-lactam hypersensitivity, but vancomycin should not be continued if the organism is shown to be sensitive to other non-β-lactam antibiotics.

For outpatient management, amoxicillin (1 g every 8 h) provides effective treatment for virtually all cases of pneumococcal pneumonia. Neither cephalosporins nor quinolones, which are far more expensive, offer any advantage over amoxicillin. Levofloxacin (500–750 mg/d as a single dose) and moxifloxacin (400 mg/d as a single dose) are also highly likely to be effective in the United States except in patients who come from closed populations where these drugs are used widely or who have themselves been treated recently with a quinolone. Clindamycin (600–1200 mg/d every 6 h) is effective in 90% of cases and azithromycin (500 mg on day 1 followed by 250–500 mg/d) or clarithromycin (500–750 mg/d as a single dose) in 80% of cases. Treatment failure resulting in bacteremic disease due to macrolide-resistant isolates has been amply documented in patients given azithromycin empirically. As noted above, rates of resistance to all these antibiotics are relatively low in some countries and much higher in others; high-dose amoxicillin remains the best option worldwide.

The optimal duration of treatment for pneumococcal pneumonia is uncertain, but its continuation for at least 5 days once the patient becomes afebrile appears to be a prudent approach. Cases with a second focus of infection (e.g., empyema or septic arthritis) require longer therapy.

ACUTE OTITIS MEDIA

Amoxicillin (80–90 mg/kg per day) is recommended for children with acute otitis media except in situations where observation and symptom-based treatment without antibiotics are advocated. These situations include nonsevere illness and an uncertain diagnosis in children 6 months to 2 years of age and nonsevere illness (even if the diagnosis seems certain) in children >2 years of age. Although the optimal duration of therapy has not been conclusively established, a 10-day course is recommended for younger children and for children with severe disease at any age. For children >6 years old who have mild or moderate disease, a course of 5–7 days is considered adequate. Patients whose illness fails to respond should be reassessed at 48–72 h. If acute otitis media is confirmed and antibiotic treatment has not been started, administration of amoxicillin should be commenced. If antibiotic therapy fails, a change is indicated. Failure to respond to second-line antibiotics as well indicates that myringotomy or tympanocentesis may need to be undertaken in order to obtain samples for culture.

The above recommendations can also be followed for the treatment of sinusitis. Detailed information on the further management of these conditions in children has been published by the American Academy of Pediatrics and the American Academy of Family Physicians.

Prevention

Measures to prevent pneumococcal disease include vaccination against *S. pneumoniae* and influenza viruses, reduction of comorbidities that increase the risk of pneumococcal disease, and prevention of antibiotic overuse, which fuels pneumococcal resistance.

Capsular Polysaccharide Vaccines

The 23-valent pneumococcal polysaccharide vaccine (PPV23), containing 25 μg of each capsular polysaccharide, has been licensed for use since 1983. Recommendations for its use vary by country. The U.S. Advisory Committee on Immunization Practices recommends PPV23 for all persons ≥65 years of age and for those 2–64 years of age who have underlying medical conditions that put them at increased risk for pneumococcal disease or severity (Table 134–1; see also *http://www.cdc.gov/vaccines/recs/schedules/default.htm*). Revaccination 5 years after the first dose is recommended for persons >2 years of age who have underlying medical conditions but not routinely for those whose only indication is an age of ≥65 years. PPV23 does not induce an anamnestic response, and antibody concentrations wane over time; thus revaccination is particularly important for individuals with conditions resulting in loss of antibody. Concerns about repeated revaccination have focused on safety (i.e., local reactions) and the induction of immune hyporesponsiveness. Neither the clinical relevance nor the biological basis of hyporesponsiveness is clear, but, given the possibility of its occurrence, more than one revaccination has not been recommended.

The effectiveness of PPV23 against IPD, pneumococcal pneumonia, all-cause pneumonia, and death is controversial, with wide variation in observations. The many published meta-analyses of PPV efficacy have often reached opposing conclusions with regard to a given clinical entity. Generally, observational studies cite greater effectiveness than do controlled clinical trials. The consensus is that PPV is effective against IPD but is less effective or ineffective against nonbacteremic pneumococcal pneumonia. However, published trials, observational studies, and meta-analyses contradict this view. Efficacy is often lower in the elderly and in immunodeficient patients whose condition is associated with reduced antibody responses to vaccines than in younger, healthier populations. When PPV is effective, the duration of protection following a single dose of vaccine is estimated to be ~5 years.

What is not disputed is that improved pneumococcal vaccines are needed for adults. Even in the setting of routine vaccination of infants (which indirectly protects adults from vaccine-serotype strains), disease caused by serotypes not represented in the vaccine continues to be responsible for a significant burden of disease among adults.

Polysaccharide–Protein Conjugate Vaccines

Infants and young children respond poorly to PPV, which contains T cell–independent antigens. Consequently, another class of pneumococcal vaccines, the PCVs, were developed specifically for infants and young children. The first product, a 7-valent PCV, was licensed in 2000 in the United States. As of 2010, three PCV products—containing 7, 10, and 13 serotypes, respectively—were commercially available. The serotypes included in these PCV formulations are important causes of IPD and antibiotic resistance among young children. Randomized controlled trials have demonstrated a high degree of efficacy of PCVs against vaccine-serotype IPD as well as efficacy against pneumonia, otitis media, nasopharyngeal colonization, and all-cause mortality. PCVs are recommended by the World Health Organization for inclusion in routine childhood immunization schedules worldwide, especially in countries with high infant mortality rates.

The United States was the first country to introduce PCV and therefore has the longest experience with its community-wide effects. The introduction of PCV in the United States has resulted in a >90% reduction in vaccine-serotype IPD among the whole population **(Fig. 134-7)**. This decline has been noted not only in those age groups immunized but also in adults and is attributable to the near elimination of vaccine-serotype nasopharyngeal colonization in immunized infants, which reduces spread to adults. This protection of unimmunized community members through vaccination of a subset of the community is termed *the indirect effect*. Increases in colonization with—and concomitantly in disease due to—non-vaccine-serotype strains (i.e., replacement colonization and disease) have been seen; however, the absolute rate increases in IPD caused by non-vaccine serotypes are generally small, especially relative to decreases in vaccine-serotype IPD (see "Epidemiology," above). Since vaccine-serotype strains are more commonly resistant to antibiotics than are non-vaccine serotypes, use of PCV has also resulted in dramatic declines in the proportion and absolute rates of drug-resistant pneumococcal disease. The recommendations of the Advisory Committee on Immunization Practices for the use of conjugate vaccines can be found at *http://www.cdc.gov/MMWR/pdf/wk/mm5909.pdf*. Recently, PCV has been shown to prevent pneumococcal infection in HIV-infected adults.

Figure 134-7

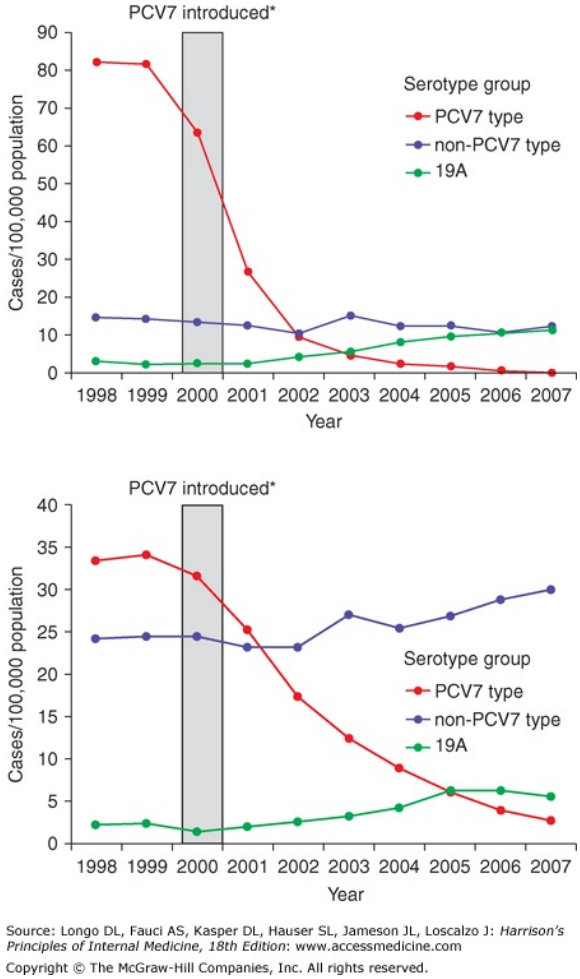

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Changes in invasive pneumococcal disease (IPD) incidence, by serotype group, among children <5 years old (*top*) and adults >65 years old (*bottom*), 1998–2007. *7-Valent pneumococcal conjugate vaccine (PCV7) was introduced in the United States for routine administration to infants and young children during the second half of 2000. (*Reprinted with permission from Pilishvili et al, 2010.*)

Other Prevention Strategies

Pneumococcal disease can also be averted through the prevention of illnesses that predispose individuals to pneumococcal infections. Relevant measures include influenza vaccination and improved management and control of diabetes, HIV infection, heart disease, and lung disease. Finally, the reduction of antibiotic misuse is a strategy for the prevention of pneumococcal disease in that antimicrobial resistance directly and indirectly perpetuates organism transmission and disease in the community.

FURTHER READINGS

French N et al: A trial of a 7-valent pneumococcal conjugate vaccine in HIV-infected adults. N Engl J Med 362:812, 2010

Goldblatt D et al: Antibody responses to nasopharyngeal carriage of *Streptococcus pneumoniae* in adults: A longitudinal household study. J Infect Dis 192:387, 2005

Huss A et al: Efficacy of pneumococcal vaccination in adults: A meta-analysis. CMAJ 180:48, 2009[PMID: 19124790]

Lucero MG et al: Pneumococcal conjugate vaccines for preventing vaccine-type invasive pneumococcal disease and X-ray defined pneumonia in children less than two years of age. Cochrane Database Syst Rev Oct 7(4):CD004977, 2009

Morens DM et al: Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: Implications for pandemic influenza preparedness. J Infect Dis 198:962, 2008[PMID: 18710327]

O'brien KL et al: Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: Global estimates. Lancet 374:893, 2009[PMID: 21698803]

Pilishvili T et al: Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. J Infect Dis 201:32, 2010[PMID: 19947881]

Siber GR et al (eds): *Pneumococcal Vaccines: The impact of conjugate vaccines* . Washington, DC, ASM Press, 2008

van der Poll T, Opal SM: Pathogenesis, treatment, and prevention of pneumococcal pneumonia. Lancet 374:1543, 2009

WEB SITES

American Academy of Pediatrics RED BOOK. The report of the Committee on Infectious Diseases: *http://aapredbook.aappublications.org*

Pneumococcal Regional Serotype Distribution for Pneumococcal AMC TPP:
http://www.vaccineamc.org/files/TPP_Codebook.pdf

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

STAPHYLOCOCCAL INFECTIONS: INTRODUCTION

Staphylococcus aureus, the most virulent of the many staphylococcal species, has demonstrated its versatility by remaining a major cause of morbidity and mortality despite the availability of numerous effective antistaphylococcal antibiotics. *S. aureus* is a pluripotent pathogen, causing disease through both toxin-mediated and non-toxin-mediated mechanisms. This organism is responsible for both nosocomial and community-based infections that range from relatively minor skin and soft tissue infections primarily to life-threatening systemic infections.

The "other" staphylococci, collectively designated *coagulase-negative staphylococci* (CoNS), are considerably less virulent than *S. aureus* but remain important pathogens in infections primarily associated with prosthetic devices.

MICROBIOLOGY AND TAXONOMY

Staphylococci, gram-positive cocci in the family Micrococcaceae, form grapelike clusters on Gram's stain (**Fig. 135-1**). These organisms are catalase-positive (unlike streptococcal species), nonmotile, aerobic, and facultatively anaerobic. They are capable of prolonged survival on environmental surfaces in varying conditions.

Figure 135-1

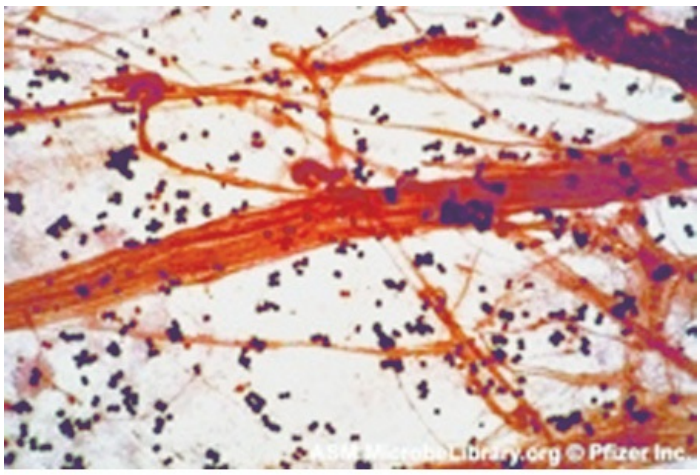

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gram's stain of *S. aureus* in a sputum sample with polymorphonuclear leukocytes. *(From ASM MicrobeLibrary.org.© Pfizer, Inc.)*

More than 30 staphylococcal species are pathogenic. A simple strategy for identification of the more clinically important species is outlined in **Fig. 135-2**. Automated diagnostic systems, kits for biochemical characterization, and DNA-based assays are available for species identification. With few exceptions, *S. aureus* is distinguished from other staphylococcal species by its production of coagulase, a surface enzyme that converts fibrinogen to fibrin. Latex kits designed to detect both protein A and clumping factor also distinguish *S. aureus* from other staphylococcal species. *S. aureus* ferments mannitol, is positive for protein A, and produces DNAse. On blood agar plates, *S. aureus* tends to form golden β-hemolytic colonies; in contrast, CoNS produce small white nonhemolytic colonies.

Figure 135-2

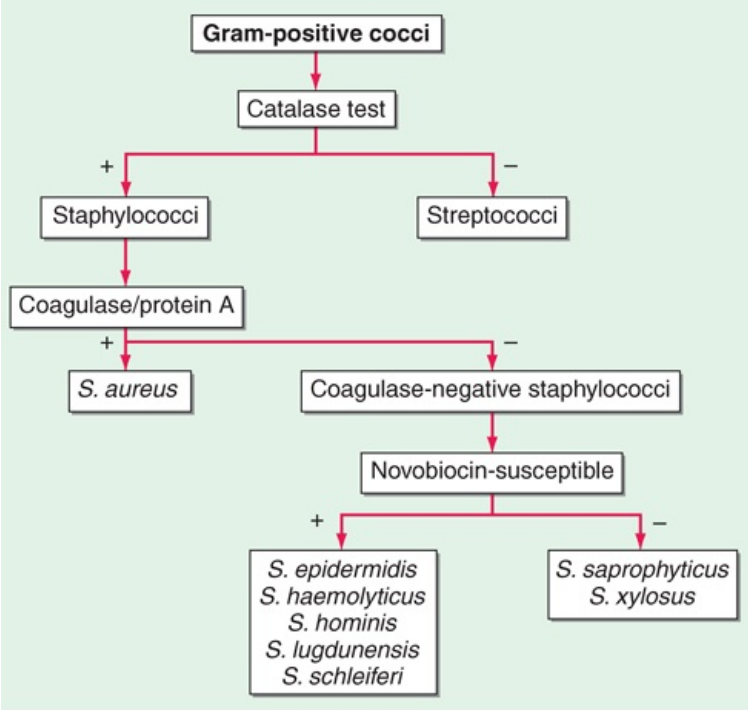

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Biochemical characterization of staphylococci: algorithm of biochemical tests used to discriminate among the clinically important staphylococci. Additional tests are necessary to identify all of the different species.

Determining whether multiple staphylococcal isolates from different patients are the same or different is often relevant when there is concern that a nosocomial outbreak is due to a common point source (e.g., a contaminated medical instrument). Molecular typing methods, such as pulsed-field gel electrophoresis and sequence-based techniques [e.g., staphylococcal protein A (*spa*) typing], have increasingly been used for this purpose.

S. AUREUS INFECTIONS

Epidemiology

S. aureus is a part of the normal human flora; ~25–50% of healthy persons may be persistently or transiently colonized. The rate of colonization is higher among insulin-dependent diabetics, HIV-infected patients, patients undergoing hemodialysis, and individuals with skin damage. The anterior nares are a frequent site of human colonization, although the skin (especially when damaged), vagina, axilla, perineum, and oropharynx may also be colonized. These colonization sites serve as a reservoir of strains for future infections, and persons colonized with *S. aureus* are at greater risk of subsequent infection than are noncolonized individuals.

Some diseases increase the risk of *S. aureus* infection; diabetes, for example, combines an increased rate of *S. aureus* colonization and the use of injectable insulin with the possibility of impaired leukocyte function. Individuals with congenital or acquired qualitative or quantitative defects of polymorphonuclear leukocytes (PMNs) are at increased risk of *S. aureus* infections; this group includes neutropenic patients (e.g., those receiving chemotherapeutic agents), those with chronic granulomatous disease, and those with Job's or Chédiak-Higashi syndrome. Other groups at risk include individuals with skin abnormalities and those with prosthetic devices.

Overall, *S. aureus* is a leading cause of nosocomial infections. It is the most common cause of surgical wound infections and is second only to CoNS as a cause of primary bacteremia. Increasingly, nosocomial isolates are resistant to multiple antibiotics. In the community, *S. aureus* remains an important cause of skin and soft tissue infections, respiratory infections, and (among injection drug users) infective endocarditis. The increasing prevalence of home infusion therapy is another cause of community-acquired staphylococcal infections.

Most individuals who develop *S. aureus* infections are infected with their own colonizing strains. However, *S. aureus* may also be acquired from other people or from environmental exposures. Transmission most frequently results from transient colonization of the hands of hospital personnel, who then transfer strains from one patient to another. Spread of staphylococci in aerosols of respiratory or nasal secretions from heavily colonized individuals has also been reported.

In the past 10 years, numerous outbreaks of community-based infection caused by methicillin-resistant *S. aureus* (MRSA) in individuals with no prior medical exposure have been reported. These outbreaks have taken place in both rural and urban settings in widely separated regions throughout the world. The reports document a dramatic change in the epidemiology of MRSA infections. The outbreaks have occurred among such diverse groups as children, prisoners, athletes, Native Americans, and drug users. Risk factors common to these outbreaks include poor hygienic conditions, close contact, contaminated material, and damaged skin. The community-associated infections have been caused by a limited number of MRSA strains. In the United States, strain USA300 (defined by pulsed-field gel electrophoresis) has been the predominant clone. While the majority of infections caused by this community-based clone of MRSA have involved the skin and soft tissue, 5–10% have been invasive. USA300 is also responsible for an increasing number of nosocomial infections. Of concern has been the apparent capacity of community-acquired MRSA (CA-MRSA) strains to cause serious disease in immunocompetent individuals.

Pathogenesis

GENERAL CONCEPTS

S. aureus is a pyogenic pathogen known for its capacity to induce abscess formation at sites of both local and metastatic infections. This classic pathologic response to *S. aureus* defines the framework within which the infection will progress. The bacteria elicit an inflammatory response characterized by an initial intense infiltration of PMNs and a subsequent infiltration of macrophages and fibroblasts. Either the host cellular response (including the deposition of fibrin and collagen) contains the infection, or infection spreads to the adjoining tissue or the bloodstream.

In toxin-mediated staphylococcal disease, infection is not invariably present. For example, once toxin has been elaborated into food, staphylococcal food poisoning can develop in the absence of viable bacteria. In staphylococcal toxic shock syndrome (TSS), conditions allowing toxin elaboration at colonization sites (e.g., the presence of a superabsorbent tampon) suffice for initiation of clinical illness.

THE *S. AUREUS* GENOME

The entire genome has been sequenced for numerous strains of *S. aureus*. Among the interesting revelations are (1) a high degree of nucleotide sequence similarity among the different strains; (2) acquisition of a relatively large amount of genetic information by horizontal transfer from other bacterial species; and (3) the presence of unique "pathogenicity" or "genomic" islands—mobile genetic elements that contain clusters of enterotoxin and exotoxin genes or antimicrobial resistance determinants. Among the genes in these islands are those carrying *mecA*, the gene responsible for methicillin resistance. Methicillin resistance–containing islands have been designated staphylococcal cassette chromosome *mecs* (SCC*mecs*) and range in size from ~20 to 60 kb. To date, eight SCC*mec*s have been identified. Types 1–3 are traditionally associated with nosocomial MRSA isolates, while types 4–6 have been associated with the epidemic CA-MRSA strains.

A limited number of MRSA clones have been responsible for most community and hospital-associated infections worldwide. A comparison of these strains with those from earlier outbreaks (e.g., the phage 80/81 strains from the 1950s) has revealed preservation of the nucleotide sequence over time. This observation suggests that these strains possess determinants that facilitate survival and spread.

REGULATION OF VIRULENCE GENE EXPRESSION

In both toxin-mediated and non-toxin-mediated diseases due to *S. aureus*, the expression of virulence determinants associated with infection depends on a series of regulatory genes [e.g., accessory gene regulator (*agr*) and staphylococcal accessory regulator (*sar*)] that coordinately control the expression of many virulence genes. The regulatory gene *agr* is part of a quorum-sensing signal transduction pathway that senses and responds to bacterial density. Staphylococcal surface proteins are synthesized during the bacterial exponential growth phase in vitro. In contrast, many secreted proteins, such as α toxin, the enterotoxins, and assorted enzymes, are released during the postexponential growth phase in response to transcription of the effector molecule of *agr*, RNAIII.

It has been hypothesized that these regulatory genes serve a similar function in vivo. Successful invasion requires the sequential expression of these different bacterial elements. Bacterial adhesins are needed to initiate colonization of host tissue surfaces. The subsequent release of various enzymes enables the colony to obtain nutritional support and permits bacteria to spread to adjacent tissues. Studies with strains in which these regulatory genes are inactivated show reduced virulence in several animal models of *S. aureus* infection.

PATHOGENESIS OF INVASIVE *S. AUREUS* INFECTION

Staphylococci are opportunists. For these organisms to invade the host and cause infection, some or all of the following steps are necessary: contamination and colonization of tissue surfaces, establishment of a localized infection, invasion, evasion of the host response, and metastatic spread. The initiation of staphylococcal infection requires a breach in cutaneous or mucosal barriers. Colonizing strains or strains transferred from other individuals are introduced into damaged skin, a

wound, or the bloodstream. Recurrences of *S. aureus* infections are common, apparently because of the capacity of these pathogens to survive, to persist in a quiescent state in various tissues, and then to cause recrudescent infections when suitable conditions arise.

S. Aureus Colonization of Body Surfaces

The anterior nares are a principal site of staphylococcal colonization in humans. Colonization appears to involve the attachment of *S. aureus* to keratinizedepithelial cells of the anterior nares. Other factors that may contribute to colonization include the influence of other resident nasal flora and their bacterial density, host factors, and nasal mucosal damage (e.g., that resulting from inhalational drug use). Other colonized body sites, such as damaged skin, the groin, and the oropharynx, may be particularly important reservoirs for CA-MRSA strains.

Inoculation and Colonization of Tissue Surfaces

Staphylococci may be introduced into tissue as a result of minor abrasions, administration of medications such as insulin, or establishment of IV access with catheters. After their introduction into a tissue site, bacteria replicate and colonize the host tissue surface. A family of structurally related *S. aureus* surface proteins referred to as MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) plays an important role as a mediator of adherence to these sites. By adhering to exposed matrix molecules (e.g., fibrinogen, fibronectin), MSCRAMMs such as clumping factor and collagen-binding protein enable the bacteria to colonize different tissue surfaces; these proteins contribute to the pathogenesis of invasive infections such as endocarditis and arthritis by facilitating the adherence of *S. aureus* to surfaces with exposed fibrinogen or collagen.

Although CoNS are classically known for their ability to elaborate biofilms and to colonize prosthetic devices, *S. aureus* also possesses the genes responsible for biofilm formation, such as the intercellular adhesion (*ica*) locus. Binding to these devices occurs in a stepwise fashion, involving staphylococcal adherence to serum constituents that have coated the device surface and subsequent biofilm elaboration. *S. aureus* is thus a frequent cause of biomedical-device infections.

Invasion

After colonization, staphylococci replicate at the initial site of infection, elaborating enzymes that include serine proteases, hyaluronidases, thermonucleases, and lipases. These enzymes facilitate bacterial survival and local spread across tissue surfaces, although their precise role in infections is not well defined. The lipases may facilitate survival in lipid-rich areas such as the hair follicles, where *S. aureus* infections are often initiated. The *S. aureus* toxin Panton-Valentine leukocidin is cytolytic to PMNs, macrophages, and monocytes. Strains elaborating this toxin have been epidemiologically linked with cutaneous and more serious infections caused by strains of CA-MRSA.

Constitutional findings may result from either localized or systemic infections. The staphylococcal cell wall—consisting of alternating N-acetyl muramic acid and N-acetyl glucosamine units in combination with an additional cell wall component, lipoteichoic acid—can initiate an inflammatory response that includes the sepsis syndrome. Staphylococcal α toxin, which causes pore formation in various eukaryotic cells, can also initiate an inflammatory response with findings suggestive of sepsis.

Evasion of Host Defense Mechanisms

Evasion of host defense mechanisms is critical to invasion. Staphylococci possess an antiphagocytic polysaccharide microcapsule. Most human *S. aureus* infections are due to capsular types 5 and 8. The *S. aureus* capsule also plays a role in the induction of abscess formation. The capsular polysaccharides are zwitterionic: they have both negative and positive charges—a feature that is critical to abscess formation. Protein A, an MSCRAMM unique to *S. aureus*, acts as an Fc receptor, binding the Fc portion of IgG subclasses 1, 2, and 4 and preventing opsonophagocytosis by PMNs. Both chemotaxis inhibitory protein of staphylococci (CHIPS, a secreted protein) and extracellular adherence protein (EAP, a surface protein) interfere with PMN migration to sites of infection.

An additional potential mechanism of *S. aureus* evasion is its capacity for intracellular survival. Both professional and nonprofessional phagocytes internalize staphylococci. Internalization by endothelial cells may provide a sanctuary that protects bacteria against the host's defenses. It also results in cellular changes, such as the expression of integrins and Fc receptors that may contribute to systemic manifestations of disease, including sepsis and vasculitis. The intracellular environment favors the phenotypic expression of *S. aureus* small-colony variants. These menadione- and hemin-auxotrophic mutants are generally deficient in α toxin and can persist within endothelial cells. Small-colony variants are often selected after aminoglycoside therapy and are more commonly found in sites of persistent infections (e.g., chronic bone infections) and in respiratory secretions from patients with cystic fibrosis. These variants may facilitate prolonged staphylococcal survival and enhance the likelihood of recurrences. Finally, *S. aureus* can survive within PMNs and may use these cells to spread and to seed other tissue sites.

Pathogenesis of Community-Acquired MRSA Infections

A number of different virulence determinants have been identified as contributing to the pathogenesis

of CA-MRSA infections. There is a strong epidemiologic association linking the presence of the gene for the Panton-Valentine leukocidin with skin and soft tissue infections as well as with invasive infections such as necrotizing pneumonia. Other determinants that may play a role in the pathogenesis of these infections include the arginine catabolic mobile element (ACME), a cluster of unique genes that may facilitate evasion of host defense mechanisms; phenol-soluble modulins, a family of cytolytic peptides; and alpha toxin.

HOST RESPONSE TO *S. AUREUS* INFECTION

The primary host response to *S. aureus* infection is the recruitment of PMNs. These cells are attracted to infection sites by bacterial components such as formylated peptides or peptidoglycan as well as by the cytokines tumor necrosis factor (TNF) and interleukins (ILs) 1 and 6, which are released by activated macrophages and endothelial cells.

Although most individuals have antistaphylococcal antibodies, it is not clear that the antibody levels are qualitatively or quantitatively sufficient to protect against infection. Although anticapsular and anti-MSCRAMM antibodies facilitate opsonization in vitro and have been protective against infection in several animal models, they have not yet successfully prevented staphylococcal infections in clinical trials.

PATHOGENESIS OF TOXIN-MEDIATED DISEASE

S. aureus produces three types of toxin: cytotoxins, pyrogenic toxin superantigens, and exfoliative toxins. Both epidemiologic data and studies in animals suggest that antitoxin antibodies are protective against illness in TSS, staphylococcal food poisoning, and staphylococcal scalded-skin syndrome (SSSS). Illness develops after toxin synthesis and absorption and the subsequent toxin-initiated host response.

Enterotoxin and Toxic Shock Syndrome Toxin 1 (TSST-1)

The pyrogenic toxin superantigens are a family of small-molecular-size, structurally similar proteins that are responsible for two diseases: TSS and food poisoning. TSS results from the ability of enterotoxins and TSST-1 to function as T cell mitogens. In the normal process of antigen presentation, the antigen is first processed within the cell, and peptides are then presented in the major histocompatibility complex (MHC) class II groove, initiating a measured T cell response. In contrast, enterotoxins bind directly to the invariant region of MHC—outside the MHC class II groove. The enterotoxins can then bind T cell receptors via the vβ chain, and this binding results in a dramatic overexpansion of T cell clones (up to 20% of the total T cell population). The consequence of this T cell expansion is a "cytokine storm," with the release of inflammatory mediators that include interferon γ, IL-1, IL-6, TNF-α, and TNF-β. The resulting multisystem disease produces a constellation of findings that mimic those in endotoxin shock; however, the pathogenic mechanisms differ. The release of endotoxin from the gastrointestinal tract may synergistically enhance the toxin's effects.

A different region of the enterotoxin molecule is responsible for the symptoms of food poisoning. The enterotoxins are heat stable and can survive conditions that kill the bacteria. Illness results from the ingestion of preformed toxin. As a result, the incubation period is short (1–6 h). The toxin stimulates the vagus nerve and the vomiting center of the brain. It also appears to stimulate intestinal peristaltic activity.

Exfoliative Toxins and the Staphylococcal Scalded-Skin Syndrome

The exfoliative toxins are responsible for SSSS. The toxins that produce disease in humans are of two serotypes: ETA and ETB. These toxins disrupt the desmosomes that link adjoining cells. Although the mechanism of this disruption remains uncertain, studies suggest that the toxins possess serine protease activity, which—through undefined mechanisms—triggers exfoliation. The result is a split in the epidermis at the granular level, and this event is responsible for the superficial desquamation of the skin that typifies this illness.

Diagnosis

Staphylococcal infections are readily diagnosed by Gram's stain (Fig. 135-1) and microscopic examination of abscess contents or of infected tissue. Routine culture of infected material usually yields positive results, and blood cultures are sometimes positive even when infections are localized to extravascular sites. Polymerase chain reaction (PCR)–based assays have been applied to the rapid diagnosis of *S. aureus* infection and are increasingly used in clinical microbiology laboratories. To date, serologic assays have not proved useful for the diagnosis of staphylococcal infections. Determining whether patients with documented *S. aureus* bacteremia also have infective endocarditis or a metastatic focus of infection remains a diagnostic challenge. Uniformly positive blood cultures suggest an endovascular infection such as endocarditis (see "Bacteremia, Sepsis, and Infective Endocarditis," below).

Clinical Syndromes

(Table 135–1)

Table 135–1 Common Illnesses Caused by *Staphylococcus Aureus*
Skin and Soft Tissue Infections
Folliculitis
Furuncle, carbuncle
Cellulitis
Impetigo
Mastitis
Surgical wound infections
Hidradenitis suppurativa
Musculoskeletal Infections
Septic arthritis
Osteomyelitis
Pyomyositis
Psoas abscess
Respiratory Tract Infections
Ventilator-associated or nosocomial pneumonia
Septic pulmonary emboli
Postviral pneumonia (e.g., influenza)
Empyema
Bacteremia and Its Complications
Sepsis, septic shock
Metastatic foci of infection (kidney, joints, bone, lung)
Infective endocarditis
Infective Endocarditis
Injection drug use–associated
Native-valve
Prosthetic-valve
Nosocomial
Device-Related Infections (e.g., intravascular catheters, prosthetic joints) **Toxin-Mediated Illnesses**
Toxic shock syndrome
Food poisoning
Staphylococcal scalded-skin syndrome
Invasive Infections Associated with Community-Acquired MRSA
Necrotizing fasciitis
Waterhouse-Friderichsen syndrome
Necrotizing pneumonia
Purpura fulminans

SKIN AND SOFT TISSUE INFECTIONS

S. aureus causes a variety of cutaneous infections, many of which can also be caused by group A streptococci or (less commonly) other streptococcal species. Common factors predisposing to *S. aureus* cutaneous infection include chronic skin conditions (e.g., eczema), skin damage (e.g., insect bites, minor trauma), injections (e.g., in diabetes, injection drug use), and poor personal hygiene. These infections are characterized by the formation of pus-containing blisters, which often begin in hair follicles and spread to adjoining tissues. *Folliculitis* is a superficial infection that involves the hair follicle, with a central area of purulence (pus) surrounded by induration and erythema. *Furuncles* (boils) are more extensive, painful lesions that tend to occur in hairy, moist regions of the body and extend from the hair follicle to become a true abscess with an area of central purulence. *Carbuncles* are most often located in the lower neck and are even more severe and painful, resulting from the coalescence of other lesions that extend to a deeper layer of the subcutaneous tissue. In general, furuncles and carbuncles are readily apparent, with pus often expressible or discharging from the abscess.

Mastitis develops in 1–3% of nursing mothers. This infection of the breast, which generally presents within 2–3 weeks after delivery, is characterized by findings that range from cellulitis to abscess formation. Systemic signs, such as fever and chills, are often present in more severe cases. Other cutaneous *S. aureus* infections include impetigo, cellulitis, and hidradenitis suppurativa (a recurrent follicular infection in regions with apocrine glands, such as the axilla). *S. aureus* is one of the most common causes of surgical wound infection.

MUSCULOSKELETAL INFECTIONS

S. aureus is among the most common causes of bone infections—both those resulting from hematogenous dissemination and those arising from contiguous spread from a soft tissue site. *Hematogenous osteomyelitis* in children most often involves the long bones. Infections present as fever and bone pain or with a child's reluctance to bear weight. The white blood cell count and erythrocyte sedimentation rate are often elevated. Blood cultures are positive in ~50% of cases. When necessary, bone biopsies for culture and histopathologic examination are usually diagnostic. Routine x-rays may be normal for up to 14 days after the onset of symptoms. ^{99m}Tc-phosphonate scanning often detects early evidence of infection. MRI is more sensitive than other techniques in establishing a radiologic diagnosis.

In adults, hematogenous osteomyelitis involving the long bones is less common. However, *vertebral osteomyelitis* is among the more common clinical presentations. Vertebral bone infections are most often seen in patients with endocarditis, those undergoing hemodialysis, diabetics, and injection drug users. These infections may present as intense back pain and fever but may also be clinically occult, presenting as chronic back pain and low-grade fever. *S. aureus* is the most common cause of epidural abscess, a complication that can result in neurologic compromise. Patients complain of difficulty voiding or walking and of radicular pain in addition to the symptoms associated with their osteomyelitis. Surgical intervention in this setting often constitutes a medical emergency. MRI most reliably establishes the diagnosis **(Fig. 135-3)**.

Figure 135-3

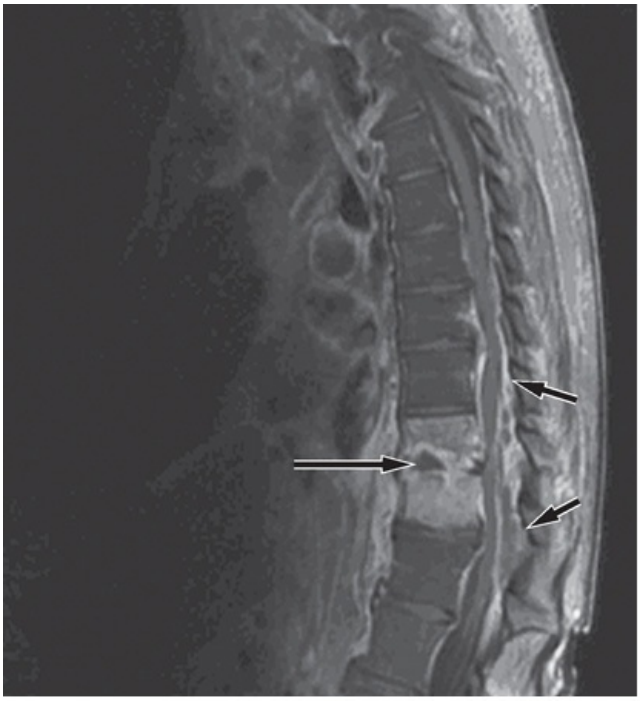

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***S. aureus* vertebral osteomyelitis and epidural abscess involving the thoracic disk -between T9 and T10.** Sagittal postcontrast MRI of the spine illustrates destruction of the T9–T10 intervertebral space with enhancement (*arrow*). There is impingement on the thoracic cord and an epidural collection extending from T9 through T11 (*short arrows*).

Bone infections that result from contiguous spread tend to develop from soft tissue infections, such as those associated with diabetic or vascular ulcers, surgery, or trauma. Exposure of bone, a draining fistulous tract, failure to heal, or continued drainage suggests involvement of underlying bone. Bone involvement is established by bone culture and histopathologic examination (revealing, for example, evidence of PMN infiltration). Contamination of culture material from adjacent tissue can make the diagnosis of osteomyelitis difficult in the absence of pathologic confirmation. In addition, it is sometimes hard to distinguish radiologically between osteomyelitis and overlying soft tissue infection with underlying osteitis.

In both children and adults, *S. aureus* is the most common cause of *septic arthritis* in native joints. This infection is rapidly progressive and may be associated with extensive joint destruction if left untreated. It presents as intense pain on motion of the affected joint, swelling, and fever. Aspiration of the joint reveals turbid fluid, with >50,000 PMNs/μL and gram-positive cocci in clusters on Gram's stain (Fig. 135-1). In adults, arthritis may result from trauma, surgery, or hematogenous dissemination. The most commonly involved joints include the knees, shoulders, hips, and phalanges. Infection frequently develops in joints previously damaged by osteoarthritis or rheumatoid arthritis. Iatrogenic infections resulting from aspiration or injection of agents into the joint also occur. In these settings, the patient experiences increased pain and swelling in the involved joint in association with fever.

Pyomyositis is an unusual infection of skeletal muscles that is seen primarily in tropical climates but also occurs in immunocompromised and HIV-infected patients. Pyomyositis presents as fever, swelling, and pain overlying the involved muscle. Aspiration of fluid from the involved tissue reveals pus. Although a history of trauma may be associated with the infection, its pathogenesis is poorly understood.

RESPIRATORY TRACT INFECTIONS

Respiratory tract infections caused by *S. aureus* occur in selected clinical settings. *S. aureus* is a cause of serious respiratory tract infections in newborns and infants; these infections present as shortness of breath, fever, and respiratory failure. Chest x-ray may reveal pneumatoceles (shaggy, thin-walled cavities). Pneumothorax and empyema are recognized complications of this infection.

In adults, nosocomial *S. aureus* pulmonary infections are commonly seen in intubated patients in intensive care units. Nasally colonized patients are at increased risk of these infections. The clinical presentation is no different from that encountered in pulmonary infections of other bacterial etiologies. Patients produce increased volumes of purulent sputum and develop respiratory distress, fever, and new pulmonary infiltrates. Distinguishing bacterial pneumonia from respiratory failure of other causes or new pulmonary infiltrates in critically ill patients is often difficult and relies on a constellation of clinical, radiologic, and laboratory findings.

Community-acquired respiratory tract infections due to *S. aureus* usually follow viral infections—most commonly influenza. Patients may present with fever, bloody sputum production, and midlung-field pneumatoceles or multiple, patchy pulmonary infiltrates **(Fig. 135-4)**. Diagnosis is made by sputum Gram's stain and culture. Blood cultures, although useful, are usually negative.

Figure 135-4

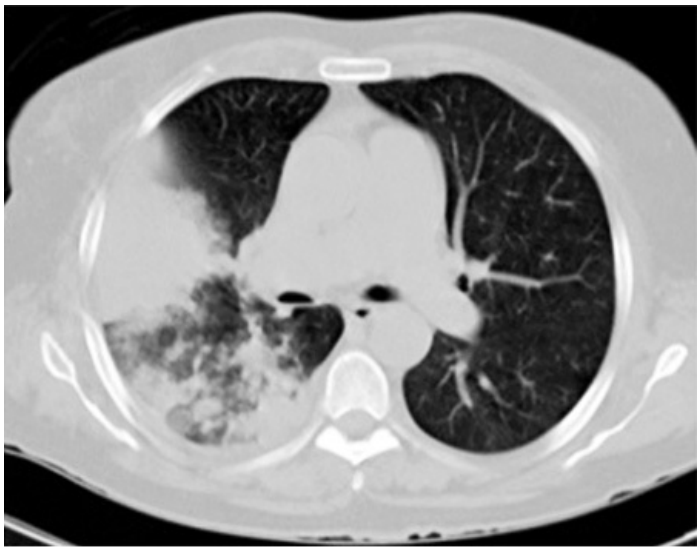

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT illustrating necrotizing pneumonia due to community-acquired MRSA in a diabetic woman who originally presented with a cutaneous abscess.

BACTEREMIA, SEPSIS, AND INFECTIVE ENDOCARDITIS

S. aureus bacteremia may be complicated by sepsis, endocarditis, vasculitis, or metastatic seeding (establishment of suppurative collections at other tissue sites). The frequency of metastatic seeding during bacteremia has been estimated to be as high as 31%. Among the more commonly seeded tissue sites are bones, joints, kidneys, and lungs.

Recognition of these complications by clinical and laboratory diagnostic methods alone is often difficult. Comorbid conditions that are frequently seen in association with *S. aureus* bacteremia and that increase the risk of complications include diabetes, HIV infection, and renal insufficiency. Other host factors associated with an increased risk of complications include presentation with community-acquired *S. aureus* bacteremia (except in injection drug users), lack of an identifiable primary focus of infection, and the presence of prosthetic devices or material.

Clinically, *S. aureus* sepsis presents in a manner similar to that documented for sepsis due to other bacteria. The well-described progression of hemodynamic changes—beginning with respiratory alkalosis and clinical findings of hypotension and fever—is commonly seen. The microbiologic diagnosis is established by positive blood cultures.

The overall incidence of *S. aureus* endocarditis has increased over the past 20 years. *S. aureus* is now the leading cause of endocarditis worldwide, accounting for 25–35% of cases. This increase is due, at least in part, to the increased use of intravascular devices; transesophageal echocardiography studies found an infective endocarditis incidence of 25% among patients with *S. aureus* bacteremia and intravascular catheters. Other factors associated with an increased risk of endocarditis are injection drug use, hemodialysis, the presence of intravascular prosthetic devices at the time of bacteremia, and immunosuppression. Despite the availability of effective antibiotics, mortality rates from these infections continue to range from 20 to 40%, depending on both the host and the nature of the infection. Complications of *S. aureus* endocarditis include cardiac valvular insufficiency, peripheral emboli, metastatic seeding, and central nervous system (CNS) involvement (e.g., mycotic aneurysms, embolic strokes).

S. aureus endocarditis is encountered in four clinical settings: (1) right-sided endocarditis in association with injection drug use, (2) left-sided native-valve endocarditis, (3) prosthetic-valve endocarditis, and (4) nosocomial endocarditis. In each of these settings, the diagnosis is established by recognition of clinical stigmata suggestive of endocarditis. These findings include cardiac manifestations, such as new or changing cardiac valvular murmurs; cutaneous evidence, such as vasculitic lesions, Osler's nodes, or Janeway lesions; evidence of right- or left-sided embolic disease; and a history suggesting a risk for *S. aureus* bacteremia. In the absence of antecedent antibiotic

therapy, blood cultures are almost uniformly positive. Transthoracic echocardiography, while less sensitive than transesophageal echocardiography, is less invasive and often establishes the presence of valvular vegetations. The Duke criteria (see Table 124-3) are now commonly used to help establish the likelihood of this diagnosis.

Acute right-sided tricuspid valvular *S. aureus* endocarditis is most often seen in injection drug users. The classic presentation includes a high fever, a toxic clinical appearance, pleuritic chest pain, and the production of purulent (sometimes bloody) sputum. Chest x-rays reveal evidence of septic pulmonary emboli (small, peripheral, circular lesions that may cavitate with time). A high percentage of affected patients have no history of antecedent valvular damage. At the outset of their illness, patients may present with fever alone, without cardiac or other localizing findings. As a result, a high index of clinical suspicion is essential for diagnosis.

Individuals with antecedent cardiac valvular damage more commonly present with left-sided native-valve endocarditis involving the previously affected valve. These patients tend to be older than those with right-sided endocarditis, their prognosis is worse, and their incidence of complications (including peripheral emboli, cardiac decompensation, and metastatic seeding) is higher.

S. aureus is one of the more common causes of prosthetic-valve endocarditis. This infection is especially fulminant in the early postoperative period and is associated with a high mortality rate. In most instances, medical therapy alone is not sufficient and urgent valve replacement is necessary. Patients are prone to develop valvular insufficiency or myocardial abscesses originating from the region of valve implantation.

The increased frequency of nosocomial endocarditis (15–30% of cases, depending on the series) reflects in part the increased use of intravascular devices. This form of endocarditis is most commonly caused by *S. aureus*. Because patients often are critically ill, are receiving antibiotics for various other indications, and have comorbid conditions, the diagnosis is often missed.

URINARY TRACT INFECTIONS

Urinary tract infections (UTIs) are infrequently caused by *S. aureus*. In contrast with that of most other urinary pathogens, the presence of *S. aureus* in the urine suggests hematogenous dissemination. Ascending *S. aureus* infections occasionally result from instrumentation of the genitourinary tract.

PROSTHETIC DEVICE–RELATED INFECTIONS

S. aureus accounts for a large proportion of prosthetic device–related infections. These infections often involve intravascular catheters, prosthetic valves, orthopedic devices, peritoneal catheters, pacemakers, left-ventricular-assist devices, and vascular grafts. In contrast with the more indolent presentation of CoNS infections, *S. aureus* device-related infections often present more acutely, with both localized and systemic manifestations. The latter infections also tend to progress more rapidly. It is relatively common for a pyogenic collection to be present at the device site. Aspiration of these collections and performance of blood cultures are important components in establishing a diagnosis. *S. aureus* infections tend to occur more commonly soon after implantation unless the device is used for access (e.g., intravascular or hemodialysis catheters). In the latter instance, infections can occur at any time. As in most prosthetic-device infections, successful therapy usually involves removal of the device. Left in place, the device is a potential nidus for either persistent or recurrent infections.

INFECTIONS ASSOCIATED WITH COMMUNITY-ACQUIRED MRSA

While the skin and soft tissues are the most common sites of infection associated with CA-MRSA, 5–10% of these infections are invasive and can even be life-threatening. The latter infections include necrotizing fasciitis, necrotizing pneumonia, and sepsis with Waterhouse-Friderichsen syndrome or purpura fulminans. These life-threatening infections reflect the increased virulence of MRSA strains.

TOXIN-MEDIATED DISEASES

Food Poisoning

S. aureus is among the most common causes of food-borne outbreaks of infection in the United States. Staphylococcal food poisoning results from the inoculation of toxin-producing *S. aureus* into food by colonized food handlers. Toxin is then elaborated in such growth-promoting food as custards, potato salad, or processed meats. Even if the bacteria are killed by warming, the heat-stable toxin is not destroyed. The onset of illness is rapid, occurring within 1–6 h of ingestion. The illness is characterized by nausea and vomiting, although diarrhea, hypotension, and dehydration may also occur. The differential diagnosis includes diarrhea of other etiologies, especially that caused by similar toxins (e.g., the toxins elaborated by *Bacillus cereus*). The rapidity of onset, the absence of fever, and the epidemic nature of the presentation (without 2° spread) arouse suspicion of staphylococcal food poisoning. Symptoms generally resolve within 8–10 h. The diagnosis can be established by the demonstration of bacteria or the documentation of enterotoxin in the implicated food. Treatment is entirely supportive.

Toxic Shock Syndrome

TSS gained attention in the early 1980s, when a nationwide outbreak occurred among young, otherwise healthy, menstruating women. Epidemiologic investigation demonstrated that these cases were associated with the use of a highly absorbent tampon that had recently been introduced to the market. Subsequent studies established the role of TSST-1 in these illnesses. Withdrawal of the tampon from the market resulted in a rapid decline in the incidence of this disease. However, menstrual and nonmenstrual cases continue to be reported.

The clinical presentation is similar in menstrual and nonmenstrual TSS, although the nature of the risk differs. Evidence of clinical *S. aureus* infection is not a prerequisite. TSS results from the elaboration of an enterotoxin or the structurally related enterotoxin-like TSST-1. More than 90% of menstrual cases are caused by TSST-1, whereas a high percentage of nonmenstrual cases are caused by enterotoxins. TSS begins with relatively nonspecific flulike symptoms. In menstrual cases, the onset usually comes 2 or 3 days after the start of menstruation. Patients present with fever, hypotension, and erythroderma of variable intensity. Mucosal involvement is common (e.g., conjunctival hyperemia). The illness can rapidly progress to symptoms that include vomiting, diarrhea, confusion, myalgias, and abdominal pain. These symptoms reflect the multisystemic nature of the disease, with involvement of the liver, kidneys, gastrointestinal tract, and/or CNS. Desquamation of the skin occurs during convalescence, usually 1–2 weeks after the onset of illness. Laboratory findings may include azotemia, leukocytosis, hypoalbuminemia, thrombocytopenia, and liver function abnormalities.

Diagnosis of TSS still depends on a constellation of findings rather than one specific finding and on a lack of evidence of other possible infections (e.g., Rocky Mountain spotted fever; **Table 135–2)**. Other diagnoses to be considered are drug toxicities, viral exanthems, sepsis, and Kawasaki disease. Illness occurs only in persons who lack antibody to TSST-1. Recurrences are possible if antibody fails to develop after the illness.

Table 135–2 Case Definition of *S. Aureus* Toxic Shock Syndrome

1. Fever: temperature of ≥38.9°C (≥102°F)
2. Hypotension: systolic blood pressure of ≤90 mmHg, or orthostatic hypotension (orthostatic drop in diastolic blood pressure by ≥15 mmHg, orthostatic syncope, or orthostatic dizziness)
3. Diffuse macular rash, with desquamation 1–2 weeks after onset (including the palms and soles)
4. Multisystem involvement
a. Hepatic: bilirubin or aminotransferase levels ≥2 times normal
b. Hematologic: platelet count ≤100,000/μL
c. Renal: blood urea nitrogen or serum creatinine level ≥2 times the normal upper limit
d. Mucous membranes: vaginal, oropharyngeal, or conjunctival hyperemia
e. Gastrointestinal: vomiting or diarrhea at onset of illness
f. Muscular: severe myalgias or serum creatine phosphokinase level ≥2 times the upper limit
g. Central nervous system: disorientation or alteration in consciousness without focal neurologic signs and in the absence of fever and hypotension
5. Negative serologic or other tests for measles, leptospirosis, and Rocky Mountain spotted fever as well as negative blood or cerebrospinal fluid cultures for organisms other than *S. aureus*

Source: M Wharton et al: Case definitions for public health surveillance. MMWR 39:1, 1990; with permission.

Staphylococcal Scalded-Skin Syndrome

SSSS most often affects newborns and children. The illness may vary from localized blister formation to exfoliation of much of the skin surface. The skin is usually fragile and often tender, with thin-walled, fluid-filled bullae. Gentle pressure results in rupture of the lesions, leaving denuded underlying skin (Nikolsky's sign; **Fig. 135-5**). The mucous membranes are usually spared. In more generalized infection, there are often constitutional symptoms, including fever, lethargy, and irritability with poor feeding. Significant amounts of fluid can be lost in more extensive cases. Illness usually follows localized infection at one of a number of possible sites. SSSS is much less common among adults but can follow infections caused by exfoliative toxin–producing strains.

Figure 135-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evidence of staphylococcal scalded-skin syndrome in a 6-year-old boy. Nikolsky's sign, with separation of the superficial layer of the outer epidermal layer, is visible. *(Reprinted with permission from LA Schenfeld et al: N Engl J Med 342:1178, 2000. © 2000 Massachusetts Medical Society. All rights reserved.)*

Prevention

Prevention of the spread of *S. aureus* infections in the hospital setting involves hand washing and careful attention to appropriate isolation procedures. Through careful screening for MRSA carriage and strict isolation practices, some Scandinavian countries have been remarkably successful at preventing the introduction and dissemination of MRSA in hospitals. Other countries, such as the United States and Great Britain, have been less successful.

The use of topical antimicrobial agents (e.g., mupirocin) to eliminate nasal colonization and/or chlorhexidine to eliminate cutaneous colonization with *S. aureus* and to prevent subsequent infection has been investigated in a number of clinical settings. Elimination of nasal carriage of *S. aureus* has reduced the incidence of infections among patients undergoing hemodialysis and peritoneal dialysis. Mupirocin effectively eliminates nasal colonization with *S. aureus.* An analysis of clinical trials suggests that there may also be a reduction in the incidence of postsurgical infections in those nasally colonized with *S. aureus*.

"Bundling" (the application of selected medical interventions in a sequence of prescribed steps) has reduced rates of nosocomial infections related to such procedures as the insertion of intravenous catheters, in which staphylococci are among the most common pathogens (see Table 131-3). A number of immunization strategies to prevent S. aureus infections—both active (e.g., capsular polysaccharide–protein conjugate vaccine) and passive (e.g., clumping factor antibody)—have been assessed. However, none has been successful for either prophylaxis or therapy.

COAGULASE-NEGATIVE STAPHYLOCOCCAL INFECTIONS

CoNS, although considerably less virulent than *S. aureus*, are among the most common causes of prosthetic-device infections. Approximately half of the identified CoNS species have been associated with human infections. Of these species, *S. epidermidis* is the most common human pathogen. This component of the normal human flora is found on the skin (where it is the most abundant bacterial species) as well as in the oropharynx and vagina. *S. saprophyticus*, a novobiocin-resistant species, is a pathogen in UTIs.

Pathogenesis

Among CoNS, *S. epidermidis* is the species most commonly associated with prosthetic-device infections. Infection is a two-step process, with initial adhesion to the device followed by colonization. *S. epidermidis* is uniquely adapted to colonize these devices by its capacity to elaborate the extracellular polysaccharide (glycocalyx or slime) that facilitates formation of a protective biofilm on the device surface.

Implanted prosthetic material is often coated with host serum or tissue constituents such as fibrinogen or fibronectin. These molecules serve as potential bridging ligands, facilitating initial bacterial attachment to the device surface. A number of surface-associated proteins, such as autolysin (AtlE), fibrinogen-binding protein, and accumulation-associated protein (AAP), may play a role in attachment to either modified or unmodified prosthetic surfaces. The polysaccharide intercellular adhesin facilitates subsequent staphylococcal colonization and accumulation on the device surface. In *S. epidermidis*, intercellular adhesin (ica) genes are more commonly found in strains associated with

device infections than in strains associated with colonization of mucosal surfaces. Biofilm appears to act as a barrier protecting bacteria from host defense mechanisms as well as from antibiotics, while providing a suitable environment for bacterial survival. Poly-γ-DL-glutamic acid is secreted by *S. epidermidis* and promotes protection against neutrophil phagocytosis.

Two additional staphylococcal species, *S. lugdunensis* and *S. schleiferi*, produce more serious infections (native-valve endocarditis and osteomyelitis) than do other CoNS. The basis for this enhanced virulence is not known, although both species appear to share more virulence determinants with *S. aureus* (e.g., clumping factor and lipase) than do other CoNS.

The capacity of *S. saprophyticus* to cause UTIs in young women appears to be related to its enhanced capacity to adhere to uroepithelial cells. A 160-kDa hemagglutinin/adhesin may contribute to this affinity.

Diagnosis

While the detection of CoNS at sites of infection or in the bloodstream is not difficult by standard microbiologic culture methods, interpretation of these results is frequently problematic. Since these organisms are present in large numbers on the skin, they often contaminate cultures. It has been estimated that only 10–25% of blood cultures positive for CoNS reflect true bacteremia. Similar problems arise with cultures of other sites. Among the clinical findings suggestive of true bacteremia are fever, evidence of local infection (e.g., erythema or purulent drainage at the IV catheter site), leukocytosis, and systemic signs of sepsis. Laboratory findings suggestive of true bacteremia include multiple isolations of the same strain (i.e., the same species with the same antibiogram or a closely related DNA fingerprint) from separate cultures, growth of the strain within 48 h, and bacterial growth in both aerobic and anaerobic bottles.

Clinical Syndromes

CoNS cause diverse prosthetic device–related infections, including those that involve prosthetic cardiac valves and joints, vascular grafts, intravascular devices, and CNS shunts. In all of these settings, the clinical presentation is similar. The signs of localized infection are often subtle, the rate of disease progression is slow, and the systemic findings are often limited. Signs of infection, such as purulent drainage, pain at the site, or loosening of prosthetic implants, are sometimes evident. Fever is frequently but not always present, and there may be mild leukocytosis. Acute-phase reactant levels, the erythrocyte sedimentation rate, and the C-reactive protein concentration may be elevated.

Infections that are not associated with prosthetic devices are infrequent, although native-valve endocarditis due to CoNS has accounted for ~5% of cases in some reviews. *S. lugdunensis* appears to be a more aggressive pathogen in this setting, causing greater mortality and rapid valvular destruction with abscess formation.

Treatment: Staphylococcal Infections

GENERAL PRINCIPLES OF THERAPY

Surgical incision and drainage of all suppurative collections constitute the most important therapeutic intervention for staphylococcal infections. The emergence of MRSA in the community has increased the importance of culturing all collections in order to identify pathogens and to determine antimicrobial susceptibility. Prosthetic-device infections are unlikely to be successfully managed unless the device is removed. In the limited number of situations in which removal is not possible or the infection is due to CoNS, an initial attempt at medical therapy without device removal may be warranted. Because of the well-recognized risk of complications associated with *S. aureus* bacteremia (e.g., endocarditis, metastatic foci of infection), therapy is generally prolonged (4–8 weeks) unless the patient is identified as being among the small percentage of individuals who are at low risk for complications.

DURATION OF ANTIMICROBIAL THERAPY

Debate continues regarding the duration of therapy for bacteremic *S. aureus* infections. Among the findings associated with an increased risk of complicated bacteremia are persistently positive blood cultures 48–96 h after institution of therapy, acquisition of the infection in the community, failure to remove a removable focus of infection (i.e., an intravascular catheter), and infection with cutaneous or embolic manifestations. For immunocompetent patients in whom short-course therapy is planned, transesophageal echocardiography to rule out endocarditis is warranted since neither clinical nor laboratory findings are adequate to detect cardiac involvement. In addition, an aggressive radiologic investigation to identify potential metastatic collections is often indicated. All symptomatic sites must be carefully evaluated.

CHOICE OF ANTIMICROBIAL AGENTS

The choice of antimicrobial agents to treat both coagulase-positive and coagulase-negative staphylococcal infections has become increasingly problematic because of the prevalence of multidrug-resistant strains. Staphylococcal resistance to most antibiotic families, including β-lactams, aminoglycosides, fluoroquinolones, and (to a lesser extent) glycopeptides, has increased. This trend is more apparent with CoNS: >80% of nosocomial isolates are resistant to methicillin, and these

methicillin-resistant strains are usually resistant to most other antibiotics as well. Because the selection of antimicrobial agents for *S. aureus* infections is similar to that for CoNS infections, treatment options for these pathogens are discussed together and are summarized in **Table 135–3**.

Table 135–3 Antimicrobial Therapy for Staphylococcal Infections[a]

Sensitivity/Resistance of Isolate	Drug of Choice	Alternative(s)	Comments
Parenteral Therapy for Serious Infections			
Sensitive to penicillin	Penicillin G (4 mU q4h)	Nafcillin or oxacillin (2 g q4h), cefazolin (2 g q8h), vancomycin (1 g q12h[b])	Fewer than 5% of isolates are sensitive to penicillin.
Sensitive to methicillin	Nafcillin or oxacillin (2 g q4h)	Cefazolin (2 g q8h[b]), vancomycin (15–20 mg/kg q8–12h[b])	Patients with penicillin allergy can be treated with a cephalosporin if the allergy does not involve an anaphylactic or accelerated reaction; desensitization to ß-lactams may be indicated in selected cases of serious infection when maximal bactericidal activity is needed (e.g., prosthetic valve endocarditis[d]). Type A ß-lactamase may rapidly hydrolyze cefazolin and reduce its efficacy in endocarditis. Vancomycin is a less effective option.
Resistant to methicillin	Vancomycin (15–20 mg/kg q8–12h[b])	Daptomycin (6 mg/kg q24h[b,c]) for bacteremia, endocarditis, and complicated skin infections; linezolid (600 mg q12h except: 400 mg q12h for uncomplicated skin infections); quinupristin/dalfopristin (7.5 mg/kg q8h)	Sensitivity testing is necessary before an alternative drug is used. Adjunctive drugs (those that should be used only in combination with other antimicrobial agents) include gentamicin (1 mg/kg q8h[b]), rifampin (300 mg PO q8h), and fusidic acid (500 mg q8h; not readily available in the United States). For some serious infections, higher doses of daptomycin have been used. Quinupristin/dalfopristin is bactericidal against methicillin-resistant isolates unless the strain is resistant to erythromycin or clindamycin. The efficacy of adjunctive therapy is not well established in many settings. Both linezolid and quinupristin/dalfopristin have had in vitro activity against most VISA and VRSA strains. See footnote for treatment of prosthetic-valve endocarditis.[d]
Resistant to methicillin with intermediate or complete resistance to vancomycin[e]	Uncertain	Same as for methicillin-resistant strains; check antibiotic susceptibilities	Same as for methicillin-resistant strains; check antibiotic susceptibilities
Not yet known (i.e., empirical therapy)	Vancomycin (15–20 mg/kg q8–12h[b])	–	Empirical therapy is given when the susceptibility of the isolate is not known. Vancomycin with or without an aminoglycoside is recommended for suspected community- or hospital-acquired *Staphylococcus aureus* infections because of the increased frequency of methicillin-resistant strains in the community.
Oral Therapy for Skin and Soft Tissue Infections			
Sensitive to methicillin	Dicloxacillin (500 mg qid), cephalexin (500 mg qid)	Minocycline or doxycycline (100 mg q12h[b]), TMP-SMX (1 or 2 ds tablets bid), clindamycin (300–450 mg/kg tid)	It is important to know the antibiotic susceptibility of isolates in the specific geographic region. All drainage should be cultured.

Resistant to methicillin	Clindamycin (300–450 mg/kg tid), TMP-SMX (1 or 2 ds tablets bid), minocycline or doxycycline (100 mg q12h[b]), linezolid (400–600 mg bid)		It is important to know the antibiotic susceptibility of isolates in the specific geographic region. All drainage should be cultured.

[a]Recommended dosages are for adults with normal renal and hepatic function.

[b]The dosage must be adjusted for patients with reduced creatinine clearance.

[c]Daptomycin cannot be used for pneumonia.

[d]For the treatment of prosthetic valve endocarditis, the addition of gentamicin (1 mg/kg q8h) and rifampin (300 mg PO q8h) is recommended, with adjustment of the gentamicin dosage if the creatinine clearance rate is reduced.

[e]Vancomycin-resistant *S. aureus* isolates from clinical infections have been reported.

Abbreviations: TMP-SMX, trimethoprim-sulfamethoxazole; VISA, vancomycin-intermediate *S. aureus;* VRSA, vancomycin-resistant *S. aureus.*

Source: Modified with permission from FD Lowy: N Engl J Med 339:520, 1998 (© 1998 Massachusetts Medical Society. All rights reserved.) and from DL Stevens et al: Clin Infect Dis 41:1373, 2006, and Med Lett 48:13, 2006.

As a result of the widespread dissemination of plasmids containing the enzyme penicillinase, few strains of staphylococci (≤5%) remain susceptible to penicillin. However, against susceptible strains, penicillin remains the drug of choice. Penicillin-resistant isolates are treated with semisynthetic penicillinase-resistant penicillins (SPRPs), such as oxacillin or nafcillin. Methicillin, the first of the SPRPs, is now used infrequently. Cephalosporins are alternative therapeutic agents for these infections. Second- and third-generation cephalosporins do not have a therapeutic advantage over first-generation cephalosporins for the treatment of staphylococcal infections. The carbapenems have excellent activity against methicillin-sensitive *S. aureus* but not against MRSA.

The isolation of MRSA was reported within 1 year of the introduction of methicillin. The prevalence of MRSA has since increased steadily. In many hospitals, 40–50% of *S. aureus* isolates are now resistant to methicillin. Resistance to methicillin indicates resistance to all SPRPs as well as to all cephalosporins. Production of a novel penicillin-binding protein (PBP 2a or 2') is responsible for methicillin resistance. This protein is synthesized by the *mecA* gene, which (as stated above) is part of a large mobile genetic element—a pathogenicity or genomic island—called SCC*mec*. It is hypothesized that this genetic material was acquired via horizontal transfer from a related staphylococcal species, such as *S. sciuri*. Phenotypic expression of methicillin resistance may be constitutive (i.e., expressed in all organisms in a population) or heterogeneous (i.e., displayed by only a proportion of the total organism population). Detection of methicillin resistance in the clinical microbiology laboratory can be difficult if the strain expresses heterogeneous resistance. Therefore, susceptibility studies are routinely performed at reduced temperatures (≤35°C for 24 h), with increased concentrations of salt in the medium to enhance the expression of resistance. In addition to PCR-based techniques, a number of rapid methods for the detection of methicillin resistance have been developed.

Vancomycin remains the drug of choice for the treatment of MRSA infections. Because it is less bactericidal than the β-lactams, it should be used only after careful consideration in patients with a history of β-lactam allergies. Three types of staphylococcal resistance to vancomycin have emerged. (1) Minimum inhibitory concentration (MIC) "creep" refers to the incremental increase in vancomycin MICs that has been detected in various geographic areas. Studies suggest that infections due to *S. aureus* strains with vancomycin MICs of >1 μg/mL may not respond as well to vancomycin therapy as those due to strains with MICs of <1 μg/mL. Some authorities (e.g., *The Medical Letter*) have recommended choosing an alternative agent in this setting. (2) In 1997, an *S. aureus* strain with reduced susceptibility to vancomycin (VISA) was reported from Japan. Subsequently, additional clinical isolates of VISA were reported. These strains were all resistant to methicillin and many other antimicrobial agents. The VISA strains appear to evolve (under vancomycin selective pressure) from strains that are susceptible to vancomycin but are heterogeneous, with a small proportion of the bacterial population expressing the resistance phenotype. The mechanism of VISA resistance is due to an abnormally large cell wall. Vancomycin is trapped by the abnormal peptidoglycan cross-linking and

is unable to gain access to its target site. (3) In 2002, the first clinical isolate of fully vancomycin-resistant *S. aureus* was reported. Resistance in this and six subsequently reported clinical isolates was due to the presence of *vanA*, the gene responsible for expression of vancomycin resistance in enterococci. This observation suggested that resistance was acquired as a result of horizontal conjugal transfer from a vancomycin-resistant strain of *Enterococcus faecalis*. Several patients had both MRSA and vancomycin-resistant enterococci cultured from infection sites. The *vanA* gene is responsible for the synthesis of the dipeptide D-Ala-D-Lac in place of D-Ala-D-Ala. Vancomycin cannot bind to the altered peptide.

Telavancin is a parenteral lipoglycopeptide derivative of vancomycin that was recently approved by the U.S. Food and Drug Administration for the treatment of complicated skin and soft tissue infections. This drug has two targets: the cell wall and the cell membrane. It remains active against VISA strains.

Daptomycin, a parenteral bactericidal agent with antistaphylococcal activity, is approved for the treatment of bacteremia (including right-sided endocarditis) and complicated skin infections. It is not effective in respiratory infections. This drug has a novel mechanism of action: it disrupts the cytoplasmic membrane. Staphylococcal resistance to daptomycin, sometimes developing during therapy, has been reported.

Linezolid—the first oxazolidinone—is bacteriostatic against staphylococci and offers the advantage of comparable bioavailability after oral or parenteral administration. Cross-resistance with other inhibitors of protein synthesis has not been detected. However, resistance to linezolid has been reported. Serious adverse reactions to linezolid include thrombocytopenia, occasional cases of neutropenia, and rare instances of peripheral neuropathy.

The parenteral streptogramin antibiotic quinupristin/dalfopristin displays bactericidal activity against all staphylococci, including VISA strains. This drug has been used successfully to treat serious MRSA infections. In cases of resistance to erythromycin or clindamycin, quinupristin/dalfopristin is bacteriostatic against staphylococci. There are limited data on the efficacy of either quinupristin/dalfopristin or linezolid for the treatment of infective endocarditis.

Although the quinolones are reasonably active against staphylococci in vitro, the frequency of staphylococcal resistance to these agents has increased progressively, especially among methicillin-resistant isolates. Of particular concern in MRSA is the possible emergence of quinolone resistance during therapy. Resistance to the quinolones is most commonly chromosomal and results from mutations of the topoisomerase IV or DNA gyrase genes, although multidrug efflux pumps may also contribute. While the newer quinolones exhibit increased in vitro activity against staphylococci, it is uncertain whether this increase translates into enhanced in vivo activity.

Tigecycline, a broad-spectrum minocycline analogue, has bacteriostatic activity against MRSA and is approved for use in skin and soft tissue infections as well as intraabdominal infections caused by *S. aureus*. Other antibiotics, such as minocycline and trimethoprim-sulfamethoxazole, have been used successfully to treat MRSA infections in cases of vancomycin toxicity or intolerance.

Combinations of antistaphylococcal agents are sometimes used to enhance bactericidal activity in the treatment of serious infections such as endocarditis or osteomyelitis. In selected instances (e.g., right-sided endocarditis), drug combinations are also used to shorten the duration of therapy. Among the antimicrobial agents used in combinations are rifampin, aminoglycosides (e.g., gentamicin), and fusidic acid (which is not readily available in the United States). While these agents are not effective singly because of the frequent emergence of resistance, they may be useful in combination with other agents because of their bactericidal activity against staphylococci. So far, however, clinical studies have not documented a therapeutic benefit, and recent reports have raised concern with regard to the potential nephrotoxicity of gentamicin and adverse drug reactions from the addition of rifampin.

ANTIMICROBIAL THERAPY FOR SELECTED SETTINGS

When necessary, the use of oral antistaphylococcal agents for uncomplicated skin and soft tissue infections is usually successful. For other infections, parenteral therapy is indicated.

S. aureus endocarditis is usually an acute, life-threatening infection. Thus prompt collection of blood for cultures must be followed immediately by empirical antimicrobial therapy. For life-threatening *S. aureus* native-valve endocarditis, many clinicians begin therapy with a 3- to 5-day course of a β-lactam and an aminoglycoside (gentamicin, 1 mg/kg IV every 8 h), although limited clinical data support this choice. If a MRSA strain is isolated, vancomycin (15–20 mg/kg every 8–12 h, given in equal doses up to a total of 2 g) is recommended. The vancomycin dose should be adjusted on the basis of trough vancomycin levels. Patients are generally treated for 4–6 weeks, with duration depending on whether there are complications. In prosthetic-valve endocarditis, surgery in addition to antibiotic therapy is often necessary. The combination of a β-lactam agent—or, if the isolate is β-lactam-resistant, vancomycin (30 mg/kg every 24 h, given in doses up to a total of 2 g)—with an aminoglycoside (gentamicin, 1 mg/kg IV every 8 h) and rifampin (300 mg orally or IV every 8 h) is recommended. This combination is used to avoid the possible emergence of rifampin resistance during therapy if only two drugs are used.

For hematogenous osteomyelitis or septic arthritis in children, a 4-week course of therapy is usually adequate. In adults, treatment is often more prolonged. For chronic forms of osteomyelitis, surgical debridement is necessary in combination with antimicrobial therapy. For joint infections, a critical component of therapy is the repeated aspiration or arthroscopy of the affected joint to prevent damage from leukocytes. The combination of rifampin with ciprofloxacin has been used successfully to treat prosthetic-joint infections, especially when the device cannot be removed. The efficacy of this combination may reflect enhanced activity against staphylococci in biofilms as well as the attainment of effective intracellular concentrations.

The choice of empirical therapy for staphylococcal infections depends in part on susceptibility data for the local geographic area. Increasingly, vancomycin (in combination with an aminoglycoside or rifampin for serious infections) is the drug of choice for both community- and hospital-acquired infections. The increase in CA-MRSA skin and soft tissue infections has drawn attention to the need for initiation of appropriate empirical therapy. Oral agents that have been effective against these isolates include clindamycin, trimethoprim-sulfamethoxazole, doxycycline, and linezolid.

THERAPY FOR TOXIC SHOCK SYNDROME

Supportive therapy with reversal of hypotension is the mainstay of therapy for TSS. Both fluids and pressors may be necessary. Tampons or other packing material should be promptly removed. The role of antibiotics is less clear. Some investigators recommend a combination of clindamycin and a semisynthetic penicillin or vancomycin (if the isolate is resistant to methicillin). Clindamycin is advocated because, as a protein synthesis inhibitor, it reduces toxin synthesis in vitro. Linezolid also appears to be effective as a toxin synthesis inhibitor. A semisynthetic penicillin or glycopeptide is suggested to eliminate any potential focus of infection as well as to eradicate persistent carriage that might increase the likelihood of recurrent illness. Anecdotal reports document the successful use of IV immunoglobulin to treat TSS. The role of glucocorticoids in the treatment of this disease is uncertain.

THERAPY FOR OTHER TOXIN-MEDIATED DISEASES

Therapy for staphylococcal food poisoning is entirely supportive. For SSSS, antistaphylococcal therapy targets the primary site of infection.

FURTHER READINGS

Bode LG et al: Preventing surgical-site infections in nasal carriers of Staphylococcus aureus. N Engl J med 362:9, 2010[PMID: 20054045]

Burton DC et al: Methicillin-resistant *Staphylococcus aureus* central line–associated bloodstream infections in US intensive care units, 1997–2007. JAMA 301:727, 2009[PMID: 19224749]

Deleo FR et al: Community-associated meticillin-resistant Staphylococcus aureus. Lancet 375:1557, 2010[PMID: 20206987]

Fowler VG Jr et al: Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. N Engl J Med 355:653, 2006[PMID: 16914701]

Grundmann H et al: Emergence and resurgence of meticillin-resistant *Staphylococcus aureus* as a public-health threat. Lancet 368:874, 2006[PMID: 16950365]

Klevens RM et al: Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States. JAMA 298:1763, 2007[PMID: 17940231]

Lowy FD: Antimicrobial resistance: The example of *Staphylococcus aureus*. J Clin Invest 111:1265, 2003[PMID: 12727914]

McCormick JK et al: Toxic shock syndrome and bacterial superantigens: An update. Annu Rev Microbiol 55:77, 2001[PMID: 11544350]

Moran GJ et al: Methicillin-resistant *S. aureus* infections among patients in the emergency department. N Engl J Med 355:666, 2006[PMID: 16914702]

Murdoch DR et al: Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century: The International Collaboration on Endocarditis–Prospective Cohort Study. Arch Intern Med 169:463, 2009[PMID: 19273776]

Otto M: *Staphylococcus epidermidis*—the "accidental" pathogen. Nat Rev Microbiol 7:555, 2009[PMID: 19609257]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

STREPTOCOCCAL INFECTIONS: INTRODUCTION

Many varieties of streptococci are found as part of the normal flora colonizing the human respiratory, gastrointestinal, and genitourinary tracts. Several species are important causes of human disease. Group A *Streptococcus* (GAS, *S. pyogenes*) is responsible for streptococcal pharyngitis, one of the most common bacterial infections of school-age children, and for the postinfectious syndromes of acute rheumatic fever (ARF) and poststreptococcal glomerulonephritis (PSGN). Group B *Streptococcus* (GBS, *S. agalactiae*) is the leading cause of bacterial sepsis and meningitis in newborns and a major cause of endometritis and fever in parturient women. Viridans streptococci are the most common cause of bacterial endocarditis. Enterococci, which are morphologically similar to streptococci, are now considered a separate genus on the basis of DNA homology studies. Thus, the species previously designated as *S. faecalis* and *S. faecium* have been renamed *Enterococcus faecalis* and *E. faecium*, respectively. The enterococci are discussed in Chap. 137.

Streptococci are gram-positive, spherical to ovoid bacteria that characteristically form chains when grown in liquid media. Most streptococci that cause human infections are facultative anaerobes, although some are strict anaerobes. Streptococci are relatively fastidious organisms, requiring enriched media for growth in the laboratory. Clinicians and clinical microbiologists identify streptococci by several classification systems, including hemolytic pattern, Lancefield group, species name, and common or trivial name. Many streptococci associated with human infection produce a zone of complete (β) hemolysis around the bacterial colony when cultured on blood agar. The β-hemolytic streptococci can be classified by the Lancefield system, a serologic grouping based on the reaction of specific antisera with bacterial cell-wall carbohydrate antigens. With rare exceptions, organisms belonging to Lancefield groups A, B, C, and G are all β-hemolytic, and each is associated with characteristic patterns of human infection. Other streptococci produce a zone of partial (α) hemolysis, often imparting a greenish appearance to the agar. These α-hemolytic streptococci are further identified by biochemical testing and include *S. pneumoniae* (Chap. 134), an important cause of pneumonia, meningitis, and other infections, and the several species referred to collectively as the *viridans streptococci*, which are part of the normal oral flora and are important agents of subacute bacterial endocarditis. Finally, some streptococci are nonhemolytic, a pattern sometimes called γ *hemolysis*. Among the organisms classified serologically as group D streptococci, the enterococci are classified as a distinct genus (Chap. 137). The classification of the major streptococcal groups causing human infections is outlined in **Table 136–1**.

Table 136–1 Classification of Streptococci

Lancefield Group	Representative Species	Hemolytic Pattern	Typical Infections
A	*S. pyogenes*	β	Pharyngitis, impetigo, cellulitis, scarlet fever
B	*S. agalactiae*	β	Neonatal sepsis and meningitis, puerperal infection, urinary tract infection, diabetic ulcer infection, endocarditis
C, G	*S. dysgalactiae* subsp. *equisimilis*	β	Cellulitis, bacteremia, endocarditis
D	Enterococci[a]: *E. faecalis; E. faecium*	Usually nonhemolytic	Urinary tract infection, nosocomial bacteremia, endocarditis
	Nonenterococci: *S. bovis*	Usually nonhemolytic	Bacteremia, endocarditis
Variable or nongroupable	Viridans streptococci: *S. sanguis; S. mitis*	α	Endocarditis, dental abscess, brain abscess
	Intermedius or *milleri* group: *S. intermedius, S. anginosus, S. constellatus*	Variable	Brain abscess, visceral abscess
	Anaerobic streptococci[b]:	Usually	Sinusitis, pneumonia, empyema, brain

	Peptostreptococcus magnus	nonhemolytic	abscess, liver abscess

[a]See Chap. 137.

[b]See Chap. 164.

GROUP A STREPTOCOCCI

Lancefield's group A consists of a single species, *S. pyogenes*. As its species name implies, this organism is associated with a variety of suppurative infections. In addition, GAS can trigger the postinfectious syndromes of ARF (which is uniquely associated with *S. pyogenes* infection; Chap. 322) and PSGN (Chap. 283).

Worldwide, GAS infections and their postinfectious sequelae (primarily ARF and rheumatic heart disease) account for an estimated 500,000 deaths per year. Although data are incomplete, the incidence of all forms of GAS infection and that of rheumatic heart disease are thought to be tenfold higher in resource-limited countries than in developed countries **(Fig. 136-1)**.

Figure 136-1

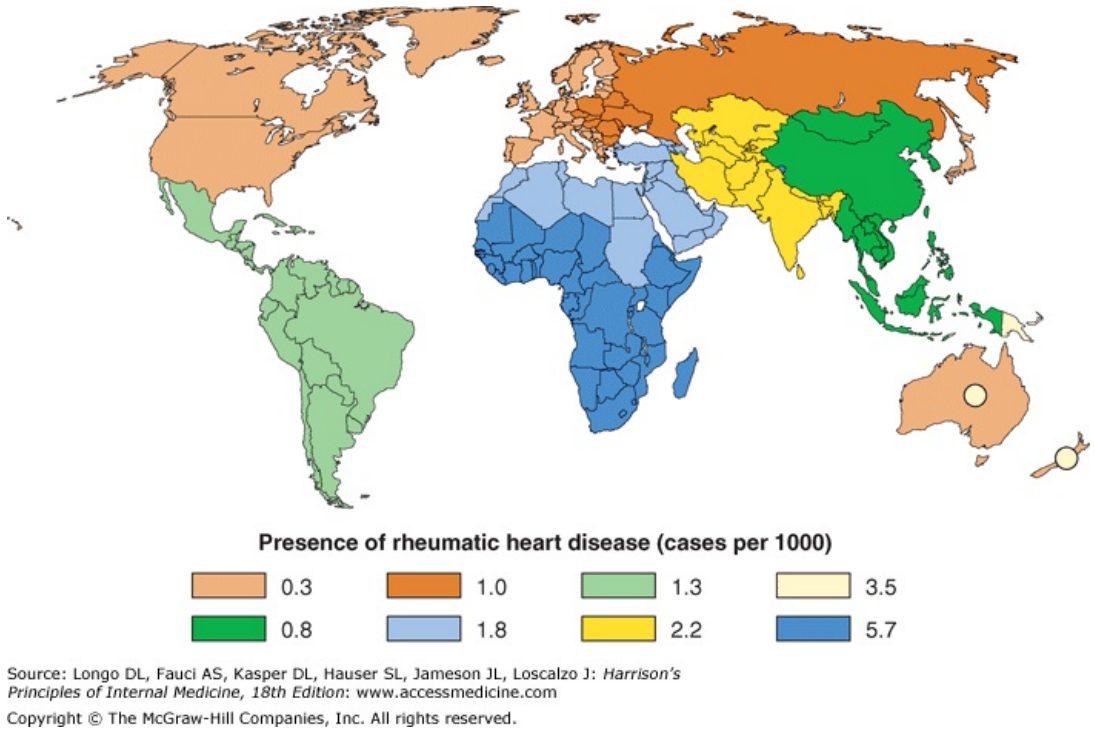

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prevalence of rheumatic heart disease in children 5–14 years old. The circles within Australia and New Zealand represent indigenous populations (and also Pacific Islanders in New Zealand). *(From Carapetis et al, 2005, with permission.)*

Pathogenesis

GAS elaborates a number of cell-surface components and extracellular products important in both the pathogenesis of infection and the human immune response. The cell wall contains a carbohydrate antigen that may be released by acid treatment. The reaction of such acid extracts with group A –specific antiserum is the basis for definitive identification of a streptococcal strain as *S. pyogenes*. The major surface protein of GAS is M protein, which occurs in more than 100 antigenically distinct types and is the basis for the serotyping of strains with specific antisera. The M protein molecules are fibrillar structures anchored in the cell wall of the organism that extend as hairlike projections away from the cell surface. The amino acid sequence of the distal or amino-terminal portion of the M protein molecule is quite variable, accounting for the antigenic variation of the different M types, while more proximal regions of the protein are relatively conserved. A newer technique for assignment of M type to GAS isolates uses the polymerase chain reaction to amplify the variable region of the *emm* gene, which encodes M protein. DNA sequence analysis of the amplified gene segment can be compared with an extensive database [developed at the Centers for Disease Control and Prevention (CDC)] for assignment of *emm* type. This method eliminates the need for typing sera, which are available in only a few reference laboratories. The presence of M protein on a GAS isolate correlates with its capacity to resist phagocytic killing in fresh human blood. This phenomenon appears to be due, at least in part, to the binding of plasma fibrinogen to M protein molecules on the streptococcal surface, which interferes with complement activation and deposition of opsonic complement fragments on the

bacterial cell. This resistance to phagocytosis may be overcome by M protein–specific antibodies; thus individuals with antibodies to a given M type acquired as a result of prior infection are protected against subsequent infection with organisms of the same M type but not against that with different M types.

GAS also elaborates, to varying degrees, a polysaccharide capsule composed of hyaluronic acid. The production of large amounts of capsule by certain strains imparts a characteristic mucoid appearance to the colonies. The capsular polysaccharide plays an important role in protecting GAS from ingestion and killing by phagocytes. In contrast to M protein, the hyaluronic acid capsule is a weak immunogen, and antibodies to hyaluronate have not been shown to be important in protective immunity. The presumed explanation is the apparent structural identity between streptococcal hyaluronic acid and the hyaluronic acid of mammalian connective tissues. The capsular polysaccharide may also play a role in GAS colonization of the pharynx by binding to CD44, a hyaluronic acid–binding protein expressed on human pharyngeal epithelial cells.

GAS produces a large number of extracellular products that may be important in local and systemic toxicity and in the spread of infection through tissues. These products include streptolysins S and O, toxins that damage cell membranes and account for the hemolysis produced by the organisms; streptokinase; DNAses; SpyCEP, a serine protease that cleaves and inactivates the chemoattractant cytokine interleukin 8, thereby inhibiting neutrophil recruitment to the site of infection; and several pyrogenic exotoxins. Previously known as erythrogenic toxins, the pyrogenic exotoxins cause the rash of scarlet fever. Since the mid-1980s, pyrogenic exotoxin–producing strains of GAS have been linked to unusually severe invasive infections, including necrotizing fasciitis and the streptococcal toxic shock syndrome (TSS). Several extracellular products stimulate specific antibody responses useful for serodiagnosis of recent streptococcal infection. Tests for these antibodies are used primarily for detection of preceding streptococcal infection in cases of suspected ARF or PSGN.

Clinical Manifestations

PHARYNGITIS

Although seen in patients of all ages, GAS pharyngitis is one of the most common bacterial infections of childhood, accounting for 20–40% of all cases of exudative pharyngitis in children; it is rare among those under the age of 3. Younger children may manifest streptococcal infection with a syndrome of fever, malaise, and lymphadenopathy without exudative pharyngitis. Infection is acquired through contact with another individual carrying the organism. Respiratory droplets are the usual mechanism of spread, although other routes, including food-borne outbreaks, have been well described. The incubation period is 1–4 days. Symptoms include sore throat, fever and chills, malaise, and sometimes abdominal complaints and vomiting, particularly in children. Both symptoms and signs are quite variable, ranging from mild throat discomfort with minimal physical findings to high fever and severe sore throat associated with intense erythema and swelling of the pharyngeal mucosa and the presence of purulent exudate over the posterior pharyngeal wall and tonsillar pillars. Enlarged, tender anterior cervical lymph nodes commonly accompany exudative pharyngitis.

The differential diagnosis of streptococcal pharyngitis includes the many other bacterial and viral etiologies **(Table 136–2)**. Streptococcal infection is an unlikely cause when symptoms and signs suggestive of viral infection are prominent (conjunctivitis, coryza, cough, hoarseness, or discrete ulcerative lesions of the buccal or pharyngeal mucosa). Because of the range of clinical presentations of streptococcal pharyngitis and the large number of other agents that can produce the same clinical picture, diagnosis of streptococcal pharyngitis on clinical grounds alone is not reliable. The throat culture remains the diagnostic gold standard. Culture of a throat specimen that is properly collected (i.e., by vigorous rubbing of a sterile swab over both tonsillar pillars) and properly processed is the most sensitive and specific means of definitive diagnosis. A rapid diagnostic kit for latex agglutination or enzyme immunoassay of swab specimens is a useful adjunct to throat culture. While precise figures on sensitivity and specificity vary, rapid diagnostic kits generally are >95% specific. Thus a positive result can be relied upon for definitive diagnosis and eliminates the need for throat culture. However, because rapid diagnostic tests are less sensitive than throat culture (relative sensitivity in comparative studies, 55–90%), a negative result should be confirmed by throat culture.

Table 136–2 Infectious Etiologies of Acute Pharyngitis

Organism	Associated Clinical Syndrome(s)
Viruses	
Rhinovirus	Common cold
Coronavirus	Common cold
Adenovirus	Pharyngoconjunctival fever
Influenza virus	Influenza
Parainfluenza virus	Cold, croup

Coxsackievirus	Herpangina, hand-foot-and-mouth disease
Herpes simplex virus	Gingivostomatitis (primary infection)
Epstein-Barr virus	Infectious mononucleosis
Cytomegalovirus	Mononucleosis-like syndrome
HIV	Acute (primary) infection syndrome
Bacteria	
Group A streptococci	Pharyngitis, scarlet fever
Group C or G streptococci	Pharyngitis
Mixed anaerobes	Vincent's angina
Arcanobacterium haemolyticum	Pharyngitis, scarlatiniform rash
Neisseria gonorrhoeae	Pharyngitis
Treponema pallidum	Secondary syphilis
Francisella tularensis	Pharyngeal tularemia
Corynebacterium diphtheriae	Diphtheria
Yersinia enterocolitica	Pharyngitis, enterocolitis
Yersinia pestis	Plague
Chlamydiae	
Chlamydophila pneumoniae	Bronchitis, pneumonia
Chlamydophila psittaci	Psittacosis
Mycoplasmas	
Mycoplasma pneumoniae	Bronchitis, pneumonia

Treatment: GAS Pharyngitis

In the usual course of uncomplicated streptococcal pharyngitis, symptoms resolve after 3–5 days. The course is shortened little by treatment, which is given primarily to prevent suppurative complications and ARF. Prevention of ARF depends on eradication of the organism from the pharynx, not simply on resolution of symptoms, and requires 10 days of penicillin treatment **(Table 136–3)**. A first-generation cephalosporin, such as cephalexin or cefadroxil, may be substituted for penicillin in cases of penicillin allergy if the nature of the allergy is not an immediate hypersensitivity reaction (anaphylaxis or urticaria) or another potentially life-threatening manifestation (e.g., severe rash and fever). Alternative agents are erythromycin and azithromycin. Azithromycin is more expensive but offers the advantages of better gastrointestinal tolerability, once-daily dosing, and a 5-day treatment course at a dose of 12 mg/kg once daily (maximum, 500 mg).

Table 136–3 Treatment of Group a Streptococcal Infections

Infection	Treatment[a]
Pharyngitis	Benzathine penicillin G, 1.2 mU IM; *or* penicillin V, 250 mg PO tid or 500 mg PO bid x 10 days (Children <27 kg: Benzathine penicillin G, 600,000 units IM; *or* penicillin V, 250 mg PO bid or tid x 10 days)
Impetigo Erysipelas/cellulitis	Same as pharyngitis Severe: Penicillin G, 1–2 mU IV q4h Mild to moderate: Procaine penicillin, 1.2 mU IM bid
Necrotizing fasciitis/myositis	Surgical debridement; *plus* penicillin G, 2–4 mU IV q4h; *plus* clindamycin,[b] 600–900 mg q8h
Pneumonia/empyema	Penicillin G, 2–4 mU IV q4h; *plus* drainage of empyema
Streptococcal toxic shock syndrome	Penicillin G, 2–4 mU IV q4h; *plus* clindamycin,[b] 600–900 mg q8h; *plus* IV immunoglobulin,[b] 2 g/kg as a single dose

[a]Penicillin allergy: A first-generation cephalosporin, such as cephalexin or cefadroxil, may be substituted for penicillin in cases of penicillin allergy if the nature of the allergy is not an immediate hypersensitivity reaction (anaphylaxis or urticaria) or another potentially life-threatening manifestation (e.g., severe rash and fever). Alternative agents for oral therapy are erythromycin (10 mg/kg PO qid, up to a maximum of 250 mg per dose) and azithromycin (a 5-day treatment course at a dose of 12 mg/kg

once daily, up to a maximum of 500 mg per day). Vancomycin is an alternative for parenteral therapy.

[b]Efficacy unproven, but recommended by several experts. See text for discussion.

Resistance to erythromycin and other macrolides is common among isolates from several countries, including Spain, Italy, Finland, Japan, and Korea. Macrolide resistance may be becoming more prevalent elsewhere with the increasing use of this class of antibiotics. In areas with resistance rates exceeding 5–10%, macrolides should be avoided unless results of susceptibility testing are known. Follow-up culture after treatment is no longer routinely recommended but may be warranted in selected cases, such as those involving patients or families with frequent streptococcal infections or those occurring in situations in which the risk of ARF is thought to be high (e.g., when cases of ARF have recently been reported in the community).

Complications

Suppurative complications of streptococcal pharyngitis have become uncommon with the widespread use of antibiotics for most symptomatic cases. These complications result from the spread of infection from the pharyngeal mucosa to deeper tissues by direct extension or by the hematogenous or lymphatic route and may include cervical lymphadenitis, peritonsillar or retropharyngeal abscess, sinusitis, otitis media, meningitis, bacteremia, endocarditis, and pneumonia. Local complications, such as peritonsillar or parapharyngeal abscess formation, should be considered in a patient with unusually severe or prolonged symptoms or localized pain associated with high fever and a toxic appearance. Nonsuppurative complications include ARF (Chap. 322) and PSGN (Chap. 283), both of which are thought to result from immune responses to streptococcal infection. Penicillin treatment of streptococcal pharyngitis has been shown to reduce the likelihood of ARF but not that of PSGN.

BACTERIOLOGIC TREATMENT FAILURE AND THE ASYMPTOMATIC CARRIER STATE

Surveillance cultures have shown that up to 20% of individuals in certain populations may have asymptomatic pharyngeal colonization with GAS. There are no definitive guidelines for management of these asymptomatic carriers or of asymptomatic patients who still have a positive throat culture after a full course of treatment for symptomatic pharyngitis. A reasonable course of action is to give a single 10-day course of penicillin for symptomatic pharyngitis and, if positive cultures persist, not to re-treat unless symptoms recur. Studies of the natural history of streptococcal carriage and infection have shown that the risk both of developing ARF and of transmitting infection to others is substantially lower among asymptomatic carriers than among individuals with symptomatic pharyngitis. Therefore, overly aggressive attempts to eradicate carriage probably are not justified under most circumstances. An exception is the situation in which an asymptomatic carrier is a potential source of infection to others. Outbreaks of food-borne infection and nosocomial puerperal infection have been traced to asymptomatic carriers who may harbor the organisms in the throat, vagina, or anus or on the skin.

Treatment: Asymptomatic Pharyngeal Colonization with Gas

When a carrier is transmitting infection to others, attempts to eradicate carriage are warranted. Data are limited on the best regimen to clear GAS after penicillin alone has failed. The combination of penicillin V (500 mg four times daily for 10 days) and rifampin (600 mg twice daily for the last 4 days) has been used to eliminate pharyngeal carriage. A 10-day course of oral vancomycin (250 mg four times daily) and rifampin (600 mg twice daily) has eradicated rectal colonization.

SCARLET FEVER

Scarlet fever consists of streptococcal infection, usually pharyngitis, accompanied by a characteristic rash **(Fig. 136-2)**. The rash arises from the effects of one of three toxins, currently designated streptococcal pyrogenic exotoxins A, B, and C and previously known as erythrogenic or scarlet fever toxins. In the past, scarlet fever was thought to reflect infection of an individual lacking toxin-specific immunity with a toxin-producing strain of GAS. Susceptibility to scarlet fever was correlated with results of the Dick test, in which a small amount of erythrogenic toxin injected intradermally produced local erythema in susceptible individuals but elicited no reaction in those with specific immunity. Subsequent studies have suggested that development of the scarlet fever rash may reflect a hypersensitivity reaction requiring prior exposure to the toxin. For reasons that are not clear, scarlet fever has become less common in recent years, although strains of GAS that produce pyrogenic exotoxins continue to be prevalent in the population. The symptoms of scarlet fever are the same as those of pharyngitis alone. The rash typically begins on the first or second day of illness over the upper trunk, spreading to involve the extremities but sparing the palms and soles. The rash is made up of minute papules, giving a characteristic "sandpaper" feel to the skin. Associated findings include circumoral pallor, "strawberry tongue" (enlarged papillae on a coated tongue, which later may become denuded), and accentuation of the rash in skinfolds (Pastia's lines). Subsidence of the rash in 6–9 days is followed after several days by desquamation of the palms and soles. The differential diagnosis of scarlet fever includes other causes of fever and generalized rash, such as measles and other viral exanthems, Kawasaki disease, toxic shock syndrome, and systemic allergic reactions (e.g., drug eruptions).

Figure 136-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scarlet fever exanthem. Finely punctate erythema has become confluent (scarlatiniform); petechiae can occur and have a linear configuration within the exanthem in body folds (Pastia's lines). *(From Fitzpatrick, Johnson, Wolff: Color Atlas and Synopsis of Clinical Dermatology, 4th ed, New York, McGraw-Hill, 2001, with permission.)*

SKIN AND SOFT TISSUE INFECTIONS

GAS—and occasionally other streptococcal species—causes a variety of infections involving the skin, subcutaneous tissues, muscles, and fascia. While several clinical syndromes offer a useful means for classification of these infections, not all cases fit exactly into one category. The classic syndromes are general guides to predicting the level of tissue involvement in a particular patient, the probable clinical course, and the likelihood that surgical intervention or aggressive life support will be required.

Impetigo (Pyoderma)

Impetigo, a superficial infection of the skin, is caused primarily by GAS and occasionally by other streptococci or *Staphylococcus aureus*. Impetigo is seen most often in young children, tends to occur during warmer months, and is more common in semitropical or tropical climates than in cooler regions. Infection is more common among children living under conditions of poor hygiene. Prospective studies have shown that colonization of unbroken skin with GAS precedes clinical infection. Minor trauma, such as a scratch or an insect bite, may then serve to inoculate organisms into the skin. Impetigo is best prevented, therefore, by attention to adequate hygiene. The usual sites of involvement are the face (particularly around the nose and mouth) and the legs, although lesions may occur at other locations. Individual lesions begin as red papules, which evolve quickly into vesicular and then pustular lesions that break down and coalesce to form characteristic honeycomb-like crusts **(Fig. 136-3)**. Lesions generally are not painful, and patients do not appear ill. Fever is not a feature of impetigo and, if present, suggests either infection extending to deeper tissues or another diagnosis. The classic presentation of impetigo usually poses little diagnostic difficulty. Cultures of impetiginous lesions often yield *S. aureus* as well as GAS. In almost all cases, streptococci are isolated initially and staphylococci appear later, presumably as secondary colonizing flora. In the past, penicillin was nearly always effective against these infections. However, an increasing frequency of penicillin treatment failure suggests that *S. aureus* may have become more prominent as a cause of impetigo. *Bullous impetigo* due to *S. aureus* is distinguished from typical streptococcal infection by more extensive, bullous lesions that break down and leave thin paper-like crusts instead of the thick amber crusts of streptococcal impetigo. Other skin lesions that may be confused with impetigo include herpetic lesions—either those of orolabial herpes simplex or those of chickenpox or zoster. Herpetic lesions can generally be distinguished by their appearance as more discrete, grouped vesicles and by a positive Tzanck test. In difficult cases, cultures of vesicular fluid should yield GAS in impetigo and the responsible virus in *Herpesvirus* infections.

Figure 136-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Impetigo contagiosa is a superficial streptococcal or *Staphylococcus aureus* infection consisting of honey-colored crusts and erythematous weeping erosions. Occasionally, bullous lesions may be seen. *(Courtesy of Mary Spraker, MD; with permission.*)

Treatment: Streptococcal Impetigo

Treatment of streptococcal impetigo is the same as that for streptococcal pharyngitis. In view of evidence that *S. aureus* has become a relatively frequent cause of impetigo, empirical regimens should cover both streptococci and *S. aureus*. For example, either dicloxacillin or cephalexin can be given at a dose of 250 mg four times daily for 10 days. Topical mupirocin ointment is also effective. Culture may be indicated to rule out methicillin-resistant *S. aureus*, especially if the response to empirical treatment is unsatisfactory. ARF is not a sequela to streptococcal skin infections, although PSGN may follow either skin or throat infection. The reason for this difference is not known. One hypothesis is that the immune response necessary for development of ARF occurs only after infection of the pharyngeal mucosa. In addition, the strains of GAS that cause pharyngitis are generally of different M protein types than those associated with skin infections; thus the strains that cause pharyngitis may have rheumatogenic potential, while the skin-infecting strains may not.

CELLULITIS

Inoculation of organisms into the skin may lead to *cellulitis*: infection involving the skin and subcutaneous tissues. The portal of entry may be a traumatic or surgical wound, an insect bite, or any other break in skin integrity. Often, no entry site is apparent. One form of streptococcal cellulitis, *erysipelas*, is characterized by a bright red appearance of the involved skin, which forms a plateau sharply demarcated from surrounding normal skin **(Fig. 136-4)**. The lesion is warm to the touch, may be tender, and appears shiny and swollen. The skin often has a *peau d'orange* texture, which is thought to reflect involvement of superficial lymphatics; superficial blebs or bullae may form, usually 2–3 days after onset. The lesion typically develops over a few hours and is associated with fever and chills. Erysipelas tends to occur on the malar area of the face (often with extension over the bridge of the nose to the contralateral malar region) and the lower extremities. After one episode, recurrence at the same site—sometimes years later—is not uncommon. Classic cases of erysipelas, with typical features, are almost always due to β-hemolytic streptococci, usually GAS and occasionally group C or G. Often, however, the appearance of streptococcal cellulitis is not sufficiently distinctive to permit a specific diagnosis on clinical grounds. The area involved may not be typical for erysipelas, the lesion may be less intensely red than usual and may fade into surrounding skin, and/or the patient may appear only mildly ill. In such cases, it is prudent to broaden the spectrum of empirical antimicrobial therapy to include other pathogens, particularly *S. aureus*, that can produce cellulitis with the same appearance. Staphylococcal infection should be suspected if cellulitis develops around a wound or an ulcer.

Figure 136-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erysipelas is a streptococcal infection of the superficial dermis and consists of well-demarcated, erythematous, edematous, warm plaques.

Streptococcal cellulitis tends to develop at anatomic sites in which normal lymphatic drainage has been disrupted, such as sites of prior cellulitis, the arm ipsilateral to a mastectomy and axillary lymph node dissection, a lower extremity previously involved in deep venous thrombosis or chronic lymphedema, or the leg from which a saphenous vein has been harvested for coronary artery bypass grafting. The organism may enter via a dermal breach some distance from the eventual site of clinical cellulitis. For example, some patients with recurrent leg cellulitis following saphenous vein removal stop having recurrent episodes only after treatment of tinea pedis on the affected extremity. Fissures in the skin presumably serve as a portal of entry for streptococci, which then produce infection more proximally in the leg at the site of previous injury. Streptococcal cellulitis may also involve recent surgical wounds. GAS is among the few bacterial pathogens that typically produce signs of wound infection and surrounding cellulitis within the first 24 h after surgery. These wound infections are usually associated with a thin exudate and may spread rapidly, either as cellulitis in the skin and subcutaneous tissue or as a deeper tissue infection (see below). Streptococcal wound infection or localized cellulitis may also be associated with *lymphangitis*, manifested by red streaks extending proximally along superficial lymphatics from the infection site.

Treatment: Streptococcal Cellulitis

See Table 136–3 and Chap. 125.

DEEP SOFT TISSUE INFECTIONS

Necrotizing fasciitis (*hemolytic streptococcal gangrene*) involves the superficial and/or deep fascia investing the muscles of an extremity or the trunk. The source of the infection is either the skin, with organisms introduced into tissue through trauma (sometimes trivial), or the bowel flora, with organisms released during abdominal surgery or from an occult enteric source, such as a diverticular or appendiceal abscess. The inoculation site may be inapparent and is often some distance from the site of clinical involvement; e.g., the introduction of organisms via minor trauma to the hand may be associated with clinical infection of the tissues overlying the shoulder or chest. Cases associated with the bowel flora are usually polymicrobial, involving a mixture of anaerobic bacteria (such as *Bacteroides fragilis* or anaerobic streptococci) and facultative organisms (usually gram-negative bacilli). Cases unrelated to contamination from bowel organisms are most commonly caused by GAS alone or in combination with other organisms (most often *S. aureus*). Overall, GAS is implicated in ~ 60% of cases of necrotizing fasciitis. The onset of symptoms is usually quite acute and is marked by severe pain at the site of involvement, malaise, fever, chills, and a toxic appearance. The physical findings, particularly early on, may not be striking, with only minimal erythema of the overlying skin. Pain and tenderness are usually severe. In contrast, in more superficial cellulitis, the skin appearance is more abnormal, but pain and tenderness are only mild or moderate. As the infection progresses (often over several hours), the severity and extent of symptoms worsen, and skin changes become more evident, with the appearance of dusky or mottled erythema and edema. The marked tenderness

of the involved area may evolve into anesthesia as the spreading inflammatory process produces infarction of cutaneous nerves.

Although myositis is more commonly due to *S. aureus* infection, GAS occasionally produces abscesses in skeletal muscles (*streptococcal myositis*), with little or no involvement of the surrounding fascia or overlying skin. The presentation is usually subacute, but a fulminant form has been described in association with severe systemic toxicity, bacteremia, and a high mortality rate. The fulminant form may reflect the same basic disease process seen in necrotizing fasciitis, but with the necrotizing inflammatory process extending into the muscles themselves rather than remaining limited to the fascial layers.

Treatment: Deep Soft Tissue Infections

Once necrotizing fasciitis is suspected, early surgical exploration is both diagnostically and therapeutically indicated. Surgery reveals necrosis and inflammatory fluid tracking along the fascial planes above and between muscle groups, without involvement of the muscles themselves. The process usually extends beyond the area of clinical involvement, and extensive debridement is required. Drainage and debridement are central to the management of necrotizing fasciitis; antibiotic treatment is a useful adjunct (Table 136–3), but surgery is life-saving. Treatment for streptococcal myositis consists of surgical drainage—usually by an open procedure that permits evaluation of the extent of infection and ensures adequate debridement of involved tissues—and high-dose penicillin (Table 136–3).

PNEUMONIA AND EMPYEMA

GAS is an occasional cause of pneumonia, generally in previously healthy individuals. The onset of symptoms may be abrupt or gradual. Pleuritic chest pain, fever, chills, and dyspnea are the characteristic manifestations. Cough is usually present but may not be prominent. Approximately one-half of patients with GAS pneumonia have an accompanying pleural effusion. In contrast to the sterile parapneumonic effusions typical of pneumococcal pneumonia, those complicating streptococcal pneumonia are almost always infected. The empyema fluid is usually visible by chest radiography on initial presentation, and its volume may increase rapidly. These pleural collections should be drained early, as they tend to become loculated rapidly, resulting in a chronic fibrotic reaction that may require thoracotomy for removal.

BACTEREMIA, PUERPERAL SEPSIS, AND STREPTOCOCCAL TOXIC SHOCK SYNDROME

GAS bacteremia is usually associated with an identifiable local infection. Bacteremia occurs rarely with otherwise uncomplicated pharyngitis, occasionally with cellulitis or pneumonia, and relatively frequently with necrotizing fasciitis. Bacteremia without an identified source raises the possibility of endocarditis, an occult abscess, or osteomyelitis. A variety of focal infections may arise secondarily from streptococcal bacteremia, including endocarditis, meningitis, septic arthritis, osteomyelitis, peritonitis, and visceral abscesses. GAS is occasionally implicated in infectious complications of childbirth, usually endometritis and associated bacteremia. In the preantibiotic era, puerperal sepsis was commonly caused by GAS; currently, it is more often caused by GBS. Several nosocomial outbreaks of puerperal GAS infection have been traced to an asymptomatic carrier, usually someone present at delivery. The site of carriage may be the skin, throat, anus, or vagina.

Beginning in the late 1980s, several reports described patients with GAS infections associated with shock and multisystem organ failure. This syndrome was called the streptococcal TSS because it shares certain features with staphylococcal TSS. In 1993, a case definition for streptococcal TSS was formulated **(Table 136–4)**. The general features of the illness include fever, hypotension, renal impairment, and respiratory distress syndrome. Various types of rash have been described, but rash usually does not develop. Laboratory abnormalities include a marked shift to the left in the white blood cell differential, with many immature granulocytes; hypocalcemia; hypoalbuminemia; and thrombocytopenia, which usually becomes more pronounced on the second or third day of illness. In contrast to patients with staphylococcal TSS, the majority with streptococcal TSS are bacteremic. The most common associated infection is a soft tissue infection—necrotizing fasciitis, myositis, or cellulitis—although a variety of other associated local infections have been described, including pneumonia, peritonitis, osteomyelitis, and myometritis. Streptococcal TSS is associated with a mortality rate of ≥ 30%, with most deaths secondary to shock and respiratory failure. Because of its rapidly progressive and lethal course, early recognition of the syndrome is essential. Patients should receive aggressive supportive care (fluid resuscitation, pressors, and mechanical ventilation) in addition to antimicrobial therapy and, in cases associated with necrotizing fasciitis, surgical debridement. Exactly why certain patients develop this fulminant syndrome is not known. Early studies of the streptococcal strains isolated from these patients demonstrated a strong association with the production of pyrogenic exotoxin A. This association has been inconsistent in subsequent case series. Pyrogenic exotoxin A and several other streptococcal exotoxins act as superantigens to trigger release of inflammatory cytokines from T lymphocytes. Fever, shock, and organ dysfunction in streptococcal TSS may reflect, in part, the systemic effects of superantigen-mediated cytokine release.

Table 136–4 Proposed Case Definition for the Streptococcal Toxic Shock Syndrome[a]

I. Isolation of group A streptococci (*Streptococcus pyogenes*)
A. From a normally sterile site
B. From a nonsterile site
II. Clinical signs of severity
A. Hypotension *and*
B. ≥2 of the following signs
1. Renal impairment
2. Coagulopathy
3. Liver function impairment
4. Adult respiratory distress syndrome
5. A generalized erythematous macular rash that may desquamate
6. Soft tissue necrosis, including necrotizing fasciitis or myositis; *or* gangrene

[a] An illness fulfilling criteria IA, IIA, and IIB is defined as a *definite* case. An illness fulfilling criteria IB, IIA, and IIB is defined as a *probable* case if no other etiology for the illness is identified.

Source: Modified from Working Group on Severe Streptococcal Infections: JAMA 269:390, 1993.

Treatment: Streptococcal Toxic Shock Syndrome

In light of the possible role of pyrogenic exotoxins or other streptococcal toxins in streptococcal TSS, treatment with clindamycin has been advocated by some authorities (Table 136–3), who argue that, through its direct action on protein synthesis, clindamycin is more effective in rapidly terminating toxin production than penicillin—a cell-wall agent. Support for this view comes from studies of an experimental model of streptococcal myositis, in which mice given clindamycin had a higher rate of survival than those given penicillin. Comparable data on the treatment of human infections are not available, although retrospective analysis has suggested a better outcome when patients with invasive soft tissue infection are treated with clindamycin rather than with cell wall–active antibiotics. Although clindamycin resistance in GAS is uncommon (<2% among U.S. isolates), it has been documented. Thus, if clindamycin is used for initial treatment of a critically ill patient, penicillin should be given as well until the antibiotic susceptibility of the streptococcal isolate is known. IV immunoglobulin has been used as adjunctive therapy for streptococcal TSS (Table 136–3). Pooled immunoglobulin preparations contain antibodies capable of neutralizing the effects of streptococcal toxins. Anecdotal reports and case series have suggested favorable clinical responses to IV immunoglobulin, but no adequately powered, prospective, controlled trials have been reported.

PREVENTION

No vaccine against GAS is commercially available. A formulation that consists of recombinant peptides containing epitopes of 26 M-protein types has undergone phase 1 and 2 testing in volunteers. Early results indicate that the vaccine is well tolerated and elicits type-specific antibody responses. Vaccines based on a conserved region of M protein or on a mixture of other conserved GAS protein antigens are in earlier stages of development.

Household contacts of individuals with invasive GAS infection (e.g., bacteremia, necrotizing fasciitis, or streptococcal TSS) are at greater risk of invasive infection than the general population. Asymptomatic pharyngeal colonization with GAS has been detected in up to 25% of persons with >4 h/d of same-room exposure to an index case. However, antibiotic prophylaxis is not routinely recommended for contacts of patients with invasive disease since such an approach (if effective) would require treatment of hundreds of contacts to prevent a single case.

STREPTOCOCCI OF GROUPS C AND G

Group C and group G streptococci are β-hemolytic bacteria that occasionally cause human infections similar to those caused by GAS. Strains that form small colonies on blood agar (<0.5 mm) are generally members of the *S. milleri* (*S. intermedius, S. anginosus*) group (see "Viridans Streptococci," below). Large-colony group C and G streptococci of human origin are now considered a single species, *S. dysgalactiae* subsp. *equisimilis*. They have been associated with pharyngitis, cellulitis and soft tissue infections, pneumonia, bacteremia, endocarditis, and septic arthritis. Puerperal sepsis, meningitis, epidural abscess, intraabdominal abscess, urinary tract infection, and neonatal sepsis have also been reported. Group C or G streptococcal bacteremia most often affects elderly or chronically ill patients and, in the absence of obvious local infection, is likely to reflect endocarditis. Septic arthritis, sometimes involving multiple joints, may complicate endocarditis or develop in its absence. Distinct streptococcal species of Lancefield group C cause infections in domesticated animals, especially horses and cattle; some human infections are acquired through contact with animals or consumption of unpasteurized milk. These zoonotic organisms include *S. equi* subsp. *zooepidemicus* and *S. equi*

subsp. *equi*.

Treatment: Group C or G Streptococcal Infection

Penicillin is the drug of choice for treatment of group C or G streptococcal infections. Antibiotic treatment is the same as for similar syndromes due to GAS (Table 136–3). Patients with bacteremia or septic arthritis should receive IV penicillin (2–4 mU every 4 h). All group C and G streptococci are sensitive to penicillin; nearly all are inhibited in vitro by concentrations of ≤0.03 μg/mL. Occasional isolates exhibit tolerance: although inhibited by low concentrations of penicillin, they are killed only by significantly higher concentrations. The clinical significance of tolerance is unknown. Because of the poor clinical response of some patients to penicillin alone, the addition of gentamicin (1 mg/kg every 8 h for patients with normal renal function) is recommended by some authorities for treatment of endocarditis or septic arthritis due to group C or G streptococci; however, combination therapy has not been shown to be superior to penicillin treatment alone. Patients with joint infections often require repeated aspiration or open drainage and debridement for cure; the response to treatment may be slow, particularly in debilitated patients and those with involvement of multiple joints. Infection of prosthetic joints almost always requires prosthesis removal in addition to antibiotic therapy.

GROUP B STREPTOCOCCI

Identified first as a cause of mastitis in cows, streptococci belonging to Lancefield's group B have since been recognized as a major cause of sepsis and meningitis in human neonates. GBS is also a frequent cause of peripartum fever in women and an occasional cause of serious infection in nonpregnant adults. Since the widespread institution of prenatal screening for GBS in the 1990s, the incidence of neonatal infection per 1000 live births has fallen from ~2–3 cases to ~0.8 case. During the same period, GBS infection in adults with underlying chronic illnesses has become more common; adults now account for a larger proportion of invasive GBS infections than do newborns. Lancefield group B consists of a single species, *S. agalactiae*, which is definitively identified with specific antiserum to the group B cell wall–associated carbohydrate antigen. A streptococcal isolate can be classified presumptively as GBS on the basis of biochemical tests, including hydrolysis of sodium hippurate (in which 99% of isolates are positive), hydrolysis of bile esculin (in which 99–100% are negative), bacitracin susceptibility (in which 92% are resistant), and production of CAMP factor (in which 98–100% are positive). CAMP factor is a phospholipase produced by GBS that causes synergistic hemolysis with β lysin produced by certain strains of *S. aureus*. Its presence can be demonstrated by cross-streaking of the test isolate and an appropriate staphylococcal strain on a blood agar plate. GBS organisms causing human infections are encapsulated by one of ten antigenically distinct polysaccharides. The capsular polysaccharide is an important virulence factor. Antibodies to the capsular polysaccharide afford protection against GBS of the same (but not of a different) capsular type.

Infection in Neonates

Two general types of GBS infection in infants are defined by the age of the patient at presentation. *Early-onset infections* occur within the first week of life, with a median age of 20 h at onset. Approximately half of these infants have signs of GBS disease at birth. The infection is acquired during or shortly before birth from the colonized maternal genital tract. Surveillance studies have shown that 5–40% of women are vaginal or rectal carriers of GBS. Approximately 50% of infants delivered vaginally by carrier mothers become colonized, although only 1–2% of those colonized develop clinically evident infection. Prematurity and maternal risk factors (prolonged labor, obstetric complications, and maternal fever) are often involved. The presentation of early-onset infection is the same as that of other forms of neonatal sepsis. Typical findings include respiratory distress, lethargy, and hypotension. Essentially all infants with early-onset disease are bacteremic, one-third to one-half have pneumonia and/or respiratory distress syndrome, and one-third have meningitis.

Late-onset infections occur in infants 1 week to 3 months old and, in rare instances, in older infants (mean age at onset, 3–4 weeks). The infecting organism may be acquired during delivery (as in early-onset cases) or during later contact with a colonized mother, nursery personnel, or another source. Meningitis is the most common manifestation of late-onset infection and in most cases is associated with a strain of capsular type III. Infants present with fever, lethargy or irritability, poor feeding, and seizures. The various other types of late-onset infection include bacteremia without an identified source, osteomyelitis, septic arthritis, and facial cellulitis associated with submandibular or preauricular adenitis.

Treatment: Group B Streptococcal Infection in Neonates

Penicillin is the agent of choice for all GBS infections. Empirical broad-spectrum therapy for suspected bacterial sepsis, consisting of ampicillin and gentamicin, is generally administered until culture results become available. If cultures yield GBS, many pediatricians continue to administer gentamicin, along with ampicillin or penicillin, for a few days until clinical improvement becomes evident. Infants with bacteremia or soft tissue infection should receive penicillin at a dosage of 200,000 units/kg per day in divided doses. For meningitis, infants ≥7 days of age should receive 250,000–450,00 units/kg per day

in three divided doses; infants ≤7 days of age should receive 450,000–500,000 units/kg per day in four divided doses. Meningitis should be treated for at least 14 days because of the risk of relapse with shorter courses.

PREVENTION

The incidence of GBS infection is unusually high among infants of women with risk factors: preterm delivery, early rupture of membranes (>24 h before delivery), prolonged labor, fever, or chorioamnionitis. Because the usual source of the organisms infecting a neonate is the mother's birth canal, efforts have been made to prevent GBS infections by the identification of high-risk carrier mothers and their treatment with various forms of antibiotic or immunoprophylaxis. Prophylactic administration of ampicillin or penicillin to such patients during delivery reduces the risk of infection in the newborn. This approach has been hampered by logistical difficulties in identifying colonized women before delivery; the results of vaginal cultures early in pregnancy are poor predictors of carrier status at delivery. The CDC recommends screening for anogenital colonization at 35–37 weeks of pregnancy by a swab culture of the lower vagina and anorectum; intrapartum chemoprophylaxis is recommended for culture-positive women and for women who, regardless of culture status, have previously given birth to an infant with GBS infection or have a history of GBS bacteriuria during pregnancy. Women whose culture status is unknown and who develop premature labor (<37 weeks), prolonged rupture of membranes (>18 h), or intrapartum fever should also receive intrapartum chemoprophylaxis. The recommended regimen for chemoprophylaxis is a loading dose of 5 million units of penicillin G followed by 2.5 million units every 4 h until delivery. Cefazolin is an alternative for women with a history of penicillin allergy who are thought not to be at high risk for anaphylaxis. For women with a history of immediate hypersensitivity, clindamycin or erythromycin may be substituted, but only if the colonizing isolate has been demonstrated to be susceptible. If susceptibility testing results are not available or indicate resistance, vancomycin should be used in this situation.

Treatment of all pregnant women who are colonized or have risk factors for neonatal infection will result in exposure of up to one-third of pregnant women and newborns to antibiotics, with the attendant risks of allergic reactions and selection for resistant organisms. Although still in the developmental stages, a GBS vaccine may ultimately offer a better solution to prevention. Because transplacental passage of maternal antibodies produces protective antibody levels in newborns, efforts are under way to develop a vaccine against GBS that can be given to childbearing-age women before or during pregnancy. Results of phase 1 clinical trials of GBS capsular polysaccharide–protein conjugate vaccines suggest that a multivalent conjugate vaccine would be safe and highly immunogenic.

Infection in Adults

The majority of GBS infections in otherwise healthy adults are related to pregnancy and parturition. Peripartum fever, the most common manifestation, is sometimes accompanied by symptoms and signs of endometritis or chorioamnionitis (abdominal distention and uterine or adnexal tenderness). Blood and vaginal swab cultures are often positive. Bacteremia is usually transitory but occasionally results in meningitis or endocarditis. Infections in adults that are not associated with the peripartum period generally involve individuals who are elderly or have an underlying chronic illness, such as diabetes mellitus or a malignancy. Among the infections that develop with some frequency in adults are cellulitis and soft tissue infection (including infected diabetic skin ulcers), urinary tract infection, pneumonia, endocarditis, and septic arthritis. Other reported infections include meningitis, osteomyelitis, and intraabdominal or pelvic abscesses. Relapse or recurrence of invasive infection weeks to months after a first episode is documented in ~4% of cases.

TREATMENT: GROUP B STREPTOCOCCAL INFECTION IN ADULTS

GBS is less sensitive to penicillin than GAS, requiring somewhat higher doses. Adults with serious localized infections (pneumonia, pyelonephritis, abscess) should receive doses of ~12 million units of penicillin G daily; patients with endocarditis or meningitis should receive 18–24 million units per day in divided doses. Vancomycin is an acceptable alternative for penicillin-allergic patients.

NONENTEROCOCCAL GROUP D STREPTOCOCCI

The main nonenterococcal group D streptococci that cause human infections belong to several species previously considered a single species, *S. bovis*. The organisms encompassed by *S. bovis* have recently been subdivided into four species: *S. gallolyticus, S. pasteurianus, S. infantarius,* and *S. lutetiensis. S. bovis* endocarditis is often associated with neoplasms of the gastrointestinal tract—most frequently, a colon carcinoma or polyp—but is also reported in association with other bowel lesions. When occult gastrointestinal lesions are carefully sought, abnormalities are found in >60% of patients with *S. bovis* endocarditis. In contrast to the enterococci, nonenterococcal group D streptococci like *S. bovis* are reliably killed by penicillin as a single agent, and penicillin is the agent of choice for *S. bovis* infections.

VIRIDANS AND OTHER STREPTOCOCCI

Viridans Streptococci

Consisting of multiple species of α-hemolytic streptococci, the viridans streptococci are a heterogeneous group of organisms that are important agents of bacterial endocarditis (Chap. 124). Several species of viridans streptococci, including *S. salivarius*, *S. mitis*, *S. sanguis*, and *S. mutans*, are part of the normal flora of the mouth, where they live in close association with the teeth and gingiva. Some species contribute to the development of dental caries.

Previously known as *S. morbillorum*, *Gemella morbillorum* has been placed in a separate genus, along with *G. haemolysans*, on the basis of genetic-relatedness studies. These species resemble viridans streptococci with respect to habitat in the human host and associated infections.

The transient viridans streptococcal bacteremia induced by eating, tooth-brushing, flossing, and other sources of minor trauma, together with adherence to biologic surfaces, is thought to account for the predilection of these organisms to cause endocarditis (see Fig. 124-1). Viridans streptococci are also isolated, often as part of a mixed flora, from sites of sinusitis, brain abscess, and liver abscess.

Viridans streptococcal bacteremia occurs relatively frequently in neutropenic patients, particularly after bone marrow transplantation or high-dose chemotherapy for cancer. Some of these patients develop a sepsis syndrome with high fever and shock. Risk factors for viridans streptococcal bacteremia include chemotherapy with high-dose cytosine arabinoside, prior treatment with trimethoprim-sulfamethoxazole or a fluoroquinolone, treatment with antacids or histamine antagonists, mucositis, and profound neutropenia.

The *S. milleri* group (also referred to as the *S. intermedius* or *S. anginosus* group) includes three species that cause human disease: *S. intermedius*, *S. anginosus*, and *S. constellatus*. These organisms are often considered viridans streptococci, although they differ somewhat from other viridans streptococci in both their hemolytic pattern (they may be α-, β-, or nonhemolytic) and the disease syndromes they cause. This group commonly produces suppurative infections, particularly abscesses of brain and abdominal viscera, and infections related to the oral cavity or respiratory tract, such as peritonsillar abscess, lung abscess, and empyema.

TREATMENT: INFECTION WITH VIRIDANS STREPTOCOCCI

Isolates from neutropenic patients with bacteremia are often resistant to penicillin; thus these patients should be treated presumptively with vancomycin until the results of susceptibility testing become available. Viridans streptococci isolated in other clinical settings usually are sensitive to penicillin.

Abiotrophia and *Granulicatella* Species (Nutritionally Variant Streptococci)

Occasional isolates cultured from the blood of patients with endocarditis fail to grow when subcultured on solid media. These *nutritionally variant streptococci* require supplemental thiol compounds or active forms of vitamin B6 (pyridoxal or pyridoxamine) for growth in the laboratory. The nutritionally variant streptococci are generally grouped with the viridans streptococci because they cause similar types of infections. However, they have been reclassified on the basis of 16S ribosomal RNA sequence comparisons into two separate genera: *Abiotrophia*, with a single species (*A. defectivus*), and *Granulicatella,* with three species associated with human infection (*G. adjacens, G. para-adjacens,* and *G. elegans*).

TREATMENT: INFECTION WITH NUTRITIONALLY VARIANT STREPTOCOCCI

Treatment failure and relapse appear to be more common in cases of endocarditis due to nutritionally variant streptococci than in those due to the usual viridans streptococci. Thus the addition of gentamicin (1 mg/kg every 8 h for patients with normal renal function) to the penicillin regimen is recommended for endocarditis due to the nutritionally variant organisms.

Other Streptococci

S. suis is an important pathogen in swine and has been reported to cause meningitis in humans, usually in individuals with occupational exposure to pigs. Strains of *S. suis* associated with human infections have generally reacted with Lancefield group R typing serum and sometimes with group D typing serum as well. Isolates may be α- or β-hemolytic and are sensitive to penicillin. *S. iniae*, a pathogen of fish, has been associated with infections in humans who have handled live or freshly killed fish. Cellulitis of the hand is the most common form of human infection, although bacteremia and endocarditis have been reported. *Anaerobic streptococci*, or *peptostreptococci*, are part of the normal flora of the oral cavity, bowel, and vagina. **Infections caused by the anaerobic streptococci are discussed in Chap. 164.**

FURTHER READINGS

Bisno AL, Stevens DL: Streptococcal infections of skin and soft tissues. N Engl J Med 334:240, 1996

Bruckner L, Gigliotti F: Viridans group streptococcal infections among children with cancer and the importance of emerging antibiotic resistance. Semin Pediatr Infect Dis 17:153, 2006

Carapetis JR et al: The global burden of group A streptococcal diseases. Lancet Infect Dis 5:685, 2005

Edwards MS, Baker CJ: Group B streptococcal infections in elderly adults. Clin Infect Dis 41:839, 2005

Gerber MA et al: Prevention of rheumatic fever and diagnosis and treatment of acute streptococcal pharyngitis: A scientific statement from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee of the Council on Cardiovascular Disease in the Young, the Interdisciplinary Council on Functional Genomics and Translational Biology, and the Interdisciplinary Council on Quality of Care and Outcomes Research: Endorsed by the American Academy of Pediatrics. Circulation 119:1541, 2009

Gibbs RS et al: Perinatal infections due to group B streptococci. Obstet Gynecol 104:1062, 2004

Kotloff K: The prospect of vaccination against group A beta-hemolytic streptococci. Curr Infect Dis Rep 10:192, 2008

Phares CR et al: Epidemiology of invasive group B streptococcal disease in the United States, 1999–2005. JAMA 299:2056, 2008

The Prevention of Invasive Group A Streptococcal Infections Workshop Participants: Prevention of invasive group A streptococcal disease among household contacts of case patients and among postpartum and postsurgical patients: Recommendations from the Centers for Disease Control and Prevention. Clin Infect Dis 35:950, 2002

Van Dyke MK et al: Evaluation of universal antenatal screening for group B streptococcus. N Engl J Med 360:2626, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ENTEROCOCCAL INFECTIONS: INTRODUCTION

Enterococci have been recognized as potential human pathogens for more than a century, but only in recent years have these organisms acquired prominence as causes of nosocomial infections. The ability of enterococci to survive and/or disseminate in the hospital environment and to acquire antibiotic resistance determinants makes the treatment of some enterococcal infections in critically ill patients a difficult challenge. Enterococci were first mentioned in the French literature in 1899; the "entérocoque" was found in the human gastrointestinal tract and was noted to have the potential to produce significant disease. Indeed, the first pathologic description of an enterococcal infection dates to the same year. A clinical isolate recovered from a patient who died as a consequence of endocarditis was initially designated *Micrococcus zymogenes*, was later named *Streptococcusfaecalis* subspecies *zymogenes*, and would now be classified as *Enterococcus faecalis*. The ability of this isolate to cause severe disease in both rabbits and mice illustrated its potential lethality in the appropriate settings.

ETIOLOGY

Enterococci are gram-positive organisms. In clinical specimens, they are usually observed as single cells, diplococci, or short chains (**Fig. 137-1**), although long chains are noted with some strains. Enterococci were originally classified as streptococci because organisms of the two genera share many morphologic and phenotypic characteristics, including a generally negative catalase reaction. Only DNA hybridization studies and then 16S rRNA sequencing clearly demonstrated that enterococci should be grouped as a genus distinct from the streptococci. Nonetheless, unlike the majority of streptococci, enterococci hydrolyze esculin in the presence of 40% bile salts and grow at high salt concentrations (6.5%) and at high temperatures (46°C). Enterococci are usually reported by the clinical laboratory to be nonhemolytic on the basis of their inability to lyse the ovine or bovine red blood cells (RBCs) commonly used in agar plates; however, some strains of *E. faecalis* do lyse RBCs from humans, horses, and rabbits. The majority of clinically relevant enterococcal species hydrolyze pyrrolidonyl-β-naphthylamide (PYR); this characteristic is helpful in differentiating enterococci from organisms of the *Streptococcus bovis* group (*S. gallolyticus* subsp. *gallolyticus*, *S. gallolyticus* subsp. *pasteurianus*, and *S. infantarius* subsp. *coli*) and from *Leuconostoc* species. Although at least 18 species of enterococci have been isolated from human infections, the overwhelming majority of cases are caused by two species: *E. faecalis* and *E. faecium*. Less frequently isolated species include *E. gallinarum*, *E. durans*, *E. hirae*, and *E. avium*.

Figure 137-1

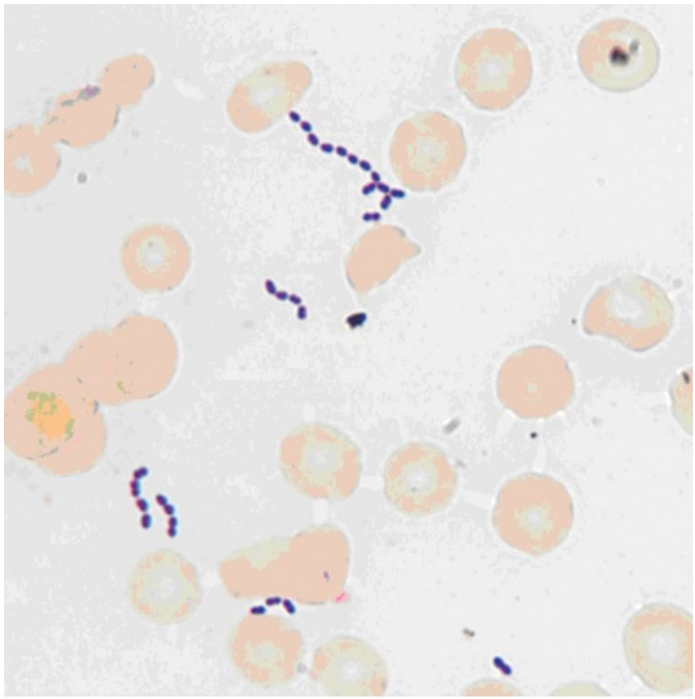

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gram's stain of cultured blood from a patient with enterococcal bacteremia. Oval gram-positive bacterial cells are arranged as diplococci and short chains. *(Courtesy of Audrey Wanger, PhD.)*

PATHOGENESIS

Enterococci are normal inhabitants of the large bowel of human adults, although they usually make up <1% of the culturable intestinal microflora. In the healthy human gastrointestinal tract, enterococci are typically symbionts that coexist with other bacteria; in fact, the utility of certain enterococcal strains as probiotics in the treatment of diarrhea suggests their possible role in maintaining the homeostatic equilibrium of the human bowel. Enterococci are intrinsically resistant to a variety of commonly used antibiotics; thus one of the most important factors that disrupts this equilibrium and promotes increased gastrointestinal colonization by enterococci is the administration of antimicrobial agents. In particular, antibiotics that are excreted in the bile and have broad-spectrum activity (i.e., certain cephalosporins that target anaerobes and gram-negative bacteria) are associated with the recovery of higher numbers of enterococci from feces. This increased colonization appears to be due not only to simple enterococcal replacement in a given "biological niche" after the eradication of competing components of the flora but also (at least in mice) to the suppression—upon reduction of the gram-negative microflora by antibiotics—of important immunologic signals (e.g., the lectin RegIIIγ) that help keep enterococcal counts low in the normal human bowel. Several studies have shown that higher levels of gastrointestinal colonization are a critical factor in the pathogenesis of enterococcal infections. However, the mechanisms by which enterococci successfully colonize the bowel and gain access to the lymphatics and/or bloodstream remain incompletely understood.

Several vertebrate, worm, and insect models have been developed to study the role of possible pathogenic determinants of both *E. faecalis* and *E. faecium*. Three main groups of factors may increase the ability of enterococci to colonize the gastrointestinal tract and/or cause disease. The first group, *enterococcal secreted factors*, are molecules released outside the bacterial cell that contribute to the process of infection; the best-studied of these molecules include enterococcal hemolysin/cytolysin and two enterococcal proteases (gelatinase and serine protease). Enterococcal cytolysin is a heterodimeric toxin produced by some strains of *E. faecalis* that is capable of lysing human RBCs as well as polymorphonuclear leukocytes and macrophages. The *E. faecalis* proteases GelE and SprE are thought to mediate virulence by several mechanisms, including the degradation of host tissues and the modification of critical components of the immune system. Mutants lacking the genes corresponding to these proteins are highly attenuated in experimental peritonitis, endocarditis, and endophthalmitis.

The second group of virulence factors, *enterococcal surface components* (e.g., adhesins), are thought to contribute to bacterial attachment to extracellular matrix molecules in the human host. Several molecules on the surface of enterococci have been characterized and shown to play a role in the pathogenesis of enterococcal infections. Among the characterized adhesins is aggregation substance of *E. faecalis*, which mediates the attachment of cells to each other, thereby facilitating conjugative plasmid exchange. Several lines of evidence indicate that aggregation substance and enterococcal

cytolysin act synergistically to increase the virulence potential of *E. faecalis* strains in experimental endocarditis. The *E. faecalis* surface protein (adhesin of collagen of *E. faecalis*, or Ace) and its *E. faecium* homologue (Acm) are microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) involved in bacterial attachment to host proteins such as collagen, fibronectin, and fibrinogen; both Ace and Acm are important in the pathogenesis of experimental endocarditis. Pili of gram-positive bacteria have been shown to be important mediators of attachment to and invasion of host tissues and are considered potential targets for immunotherapy. Both *E. faecalis* and *E. faecium* have surface pili. Mutants of *E. faecalis* lacking pili are attenuated in both experimental endocarditis and urinary tract infections (UTIs). Other surface proteins that share structural homology with MSCRAMMs and appear to play a role in enterococcal attachment to the host and in virulence include the *E. faecalis* enterococcal surface protein (Esp) and its *E. faecium* homologue (Esp_{fm}), the second collagen adhesin of *E. faecium* (Scm), the surface proteins of *E. faecium* (Fms), SgrA (which binds to components of the basal lamina), and EcbA (which binds to collagen type V). Additional surface components apparently associated with pathogenicity include polysaccharides, which are thought to interfere with phagocytosis of the organism by host immune cells. Some *E. faecalis* strains appear to harbor at least three distinct classes of capsular polysaccharide; some of these polysaccharides play a role in virulence and are potential targets for immunotherapy.

The third group of virulence factors have not been well characterized: the *E. faecalis* stress protein Gls24, which has been associated with enterococcal resistance to bile salts and appears to be important in the pathogenesis of endocarditis, and the hyl_{Efm}-containing plasmids of *E. faecium*, which are transferable between strains and increase colonization by *E. faecium*. In a mouse model of peritonitis, acquisition of these plasmids augmented the lethality of a commensal strain of *E. faecium*.

Genetics

The sequencing of bacterial genomes has increased our understanding of bacterial diversity, evolution, pathogenesis, and antibiotic resistance mechanisms. The genome sequences of >80 enterococcal strains are currently available, and some have been entirely closed and annotated. Sequence analysis has shown that the genetic diversity of enterococci is due mainly to the acquisition of mobile DNA (e.g., plasmids, transposons, and phages) and the recombination of "core" genomes. Furthermore, analysis of *E. faecium* indicates that this species harbors a malleable genome (the *accessory* genome) into which exogenous elements (including DNA from phages) are incorporated at substantial levels. This genomic information has provided new clues about enterococcal evolution from a commensal organism into an important nosocomial pathogen.

EPIDEMIOLOGY

According to the National Healthcare Safety Network of the Centers for Disease Control and Prevention, enterococci are the second most common organisms (after staphylococci) isolated from hospital-associated infections in the United States. Although *E. faecalis* remains the predominant species recovered from nosocomial infections, the isolation of *E. faecium* has increased substantially in the past 10–15 years. In fact, *E. faecium* is now almost as common as *E. faecalis* as an etiologic agent of hospital-associated infections. This point is important, since *E. faecium* is by far the most resistant and challenging enterococcal species to treat; indeed, >80% of *E. faecium* isolates recovered in U.S. hospitals are resistant to vancomycin, and >90% are resistant to ampicillin (historically the most effective β-lactam drug against enterococci). Resistance to vancomycin and ampicillin in *E. faecalis* isolates is much less common (~7% and ~4%, respectively).

The dynamics of enterococcal transmission and dissemination in the hospital environment have been extensively studied, with a focus on vancomycin-resistant enterococci (VRE). These studies have revealed that VRE colonization of the gastrointestinal tract is a critical step in the natural history of enterococcal disease and that a substantial proportion of patients colonized with VRE remain colonized for prolonged periods (sometimes >1 year) and are more likely to develop an *Enterococcus*-related illness (e.g., bacteremia) than are patients colonized with antibiotic-susceptible strains. The most important factors associated with VRE colonization and persistence in the gut include prolonged hospitalization; long courses of antibiotic therapy; hospitalization in long-term-care facilities, surgical units, and/or intensive care units; organ transplantation; renal failure (particularly in patients undergoing hemodialysis) and/or diabetes; high APACHE scores; and physical proximity to patients infected or colonized with VRE or to these patients' rooms. Once a patient becomes colonized with VRE, several key factors are involved in the organisms' dissemination in the hospital environment. VRE can survive exposure to heat and certain disinfectants and have been found on numerous inanimate objects in the hospital, including bed rails, medical equipment, doorknobs, gloves, telephones, and computer keyboards. Thus health care workers play a pivotal role in enterococcal transmission from patient to patient, and infection control measures are crucial in breaking the chain of transmission. Moreover, two meta-analyses have found that VRE infection increases the risk of death, independent of the patient's clinical status, over that among individuals infected with a glycopeptide-susceptible enterococcal strain.

Global Perspective

The epidemiology of enterococcal disease and the emergence of VRE have followed somewhat different trends in other parts of the world than in the United States. In Europe, the emergence of VRE in the mid-1980s was seen primarily in isolates recovered from animals and healthy humans rather than from hospitalized patients. The presence of VRE was associated with the use of the glycopeptide avoparcin as a growth promoter in animal feeds; this association prompted the European Union to ban the use of this compound in animal husbandry in 1996. However, after an initial decrease in the isolation of VRE from animals and humans, the prevalence of hospital-associated VRE infections has slowly increased in some European countries, with important regional differences. For example, rates of vancomycin resistance among *E. faecium* clinical isolates in Europe are highest in Greece, the United Kingdom, and Portugal (10–30%), whereas rates in Scandinavian countries and the Netherlands are <1%. These regional differences have been attributed in part to the implementation of aggressive policies of infection control in countries such as the Netherlands; these policies have kept the frequency of nosocomial methicillin-resistant *Staphylococcus aureus* (MRSA) and VRE very low. In spite of regional differences, rates of VRE continue to be much lower in most of Europe than in the United States. The reasons are not totally understood, although it has been postulated that the difference is related to the higher levels of human antibiotic use in the United States. Rates of enterococcal resistance to vancomycin in some Latin American countries are also lower (~4%) than those in the United States. Conversely, in Asia, rates of vancomycin resistance among enterococci appear to be similar to those in U.S. hospitals. Genetic analyses of vancomycin-resistant *E. faecium* in different parts of the world suggest that the nosocomial emergence and dissemination of these organisms globally is related to the success of unique clonal lineages (e.g., clonal cluster 17, or CC17) that was initially characterized by resistance to ampicillin and later acquired the genes responsible for resistance to vancomycin.

CLINICAL SYNDROMES

Urinary Tract Infections and Prostatitis

Enterococci are well-known agents of nosocomial UTI, the most common infection caused by these organisms (Chap. 288). Enterococcal UTIs are usually associated with indwelling catheterization, instrumentation, or anatomic abnormalities of the genitourinary tract, and it is often challenging to differentiate between true infection and colonization (particularly in patients with chronic indwelling catheters). The presence of leukocytes in the urine in conjunction with systemic manifestations (e.g., fever) or local signs and symptoms of infection with no other explanation and a positive urine culture ($>10^5$ CFU) suggests the diagnosis. Moreover, enterococcal UTIs often occur in critically ill patients whose comorbidities may obscure the diagnosis. In many cases, removal of the indwelling catheter may suffice without specific antimicrobial therapy. In rare circumstances, UTIs caused by enterococci may run a complicated course, with the development of pyelonephritis and perinephric abscesses that may be a portal of entry for bloodstream infections (see below). Enterococci are also known causes of chronic prostatitis, particularly in patients in whom the urinary tract has been manipulated surgically or endoscopically. These infections can be difficult to treat since the agents most potent against enterococci (i.e., aminopenicillins and glycopeptides) penetrate prostatic tissue poorly. Chronic prostatic infection can be a source of recurrent enterococcal bacteremia.

Bacteremia and Endocarditis

Bacteremia without endocarditis is one of the most common presentations of enterococcal disease. Intravascular catheters and other devices are commonly associated with these bacteremic episodes (Chap. 131). Other well-known sources of enterococcal bacteremia include the gastrointestinal and hepatobiliary tracts; pelvic and intraabdominal foci; and (less frequently) wound infections, UTIs, and bone infections. In the United States, enterococci rank second (after coagulase-negative staphylococci) as etiologic agents of central line–associated bacteremia. Patients with enterococcal bacteremia usually have comorbidities, have been in the hospital for prolonged periods, and/or have received several courses of antibiotics. Several studies indicate that the presence of *E. faecium* (as opposed to other enterococcal species) in the bloodstream may lead to worse outcomes and increased mortality rates; this finding may be related to the higher prevalence of vancomycin and ampicillin resistance in *E. faecium*, with the consequent reduction of therapeutic options. In many cases (usually when the gastrointestinal tract is the source), enterococcal bacteremia may be polymicrobial, with gram-negative organisms isolated at the same time. In addition, several cases have now been documented in which enterococcal bacteremia was associated with *Strongyloides stercoralis* hyperinfection syndrome in immunocompromised patients.

Enterococci are important causes of community- and health care–associated endocarditis, ranking second after staphylococci in the latter infections (Chap. 124). The presumed initial source of bacteremia leading to endocarditis is the gastrointestinal or genitourinary tract—e.g., in patients who have malignant and inflammatory conditions of the gut or have undergone procedures in which these tracts are manipulated. The affected patients tend to be male and elderly and to have other debilitating diseases and heart conditions. Both prosthetic and native valves can be involved; mitral and aortic

valves are affected most often. Community-associated endocarditis (usually caused by *E. faecalis*) also occurs in patients with no apparent risk factors or cardiac abnormalities. Endocarditis in women of childbearing age has been well described. The typical presentation of enterococcal endocarditis is a subacute course with fever, weight loss, malaise, and a cardiac murmur; typical stigmata of endocarditis (e.g., petechiae, Osler's nodes, Roth's spots) are found in only a minority of patients. Some atypical manifestations include arthralgias and signs and symptoms of metastatic disease (splenic abscesses, hiccups, pain in the left flank, pleural effusion, and spondylodiscitis). Embolic complications are variable and can affect the brain. Heart failure is a common complication of enterococcal endocarditis, and valve replacement may be critical in curing this infection, particularly when multidrug-resistant organisms or major complications are involved. The duration of therapy is usually 4–6 weeks, with more prolonged courses suggested for multidrug-resistant isolates in the absence of valvular replacement or with prolonged illness prior to treatment.

Meningitis

Enterococcal meningitis is an uncommon disease (accounting for only ~4% of meningitis cases) that is often associated with neurosurgical interventions and conditions such as shunts, central nervous system (CNS) trauma, and cerebrospinal fluid (CSF) leakage (Chap. 381). In some instances—usually in patients with a debilitating condition such as cardiovascular or congenital heart disease, chronic renal failure, malignancy, receipt of immunosuppressive therapy, or HIV/AIDS—presumed hematogenous seeding of the meninges is seen in infections such as endocarditis or bacteremia. Fever and changes in mental status are common, whereas overt meningeal signs are less so. CSF findings are consistent with bacterial infection—i.e., pleocytosis with a predominance of polymorphonuclear leukocytes (average, ~500/μL), an elevated protein level (usually >100 mg/dL), and a decreased glucose concentration (average, 28 mg/dL). Gram's staining yields a positive result in about half of cases, with a high rate of organism recovery from CSF cultures; the most common species isolated are *E. faecalis* and *E. faecium*. Complications include hydrocephalus, brain abscesses, and stroke; an association with *Strongyloides* hyperinfection has also been documented.

Intraabdominal, Pelvic, and Soft Tissue Infections

As mentioned earlier, enterococci are part of the commensal flora of the gastrointestinal tract and can produce spontaneous peritonitis in cirrhotic individuals and in patients undergoing chronic ambulatory peritoneal dialysis (Chap. 127). These organisms are commonly found (usually along with other bacteria, including enteric gram-negative species and anaerobes) in clinical samples from intraabdominal and pelvic collections. The presence of enterococci in intraabdominal infections is sometimes considered to be of low clinical relevance. Several studies have shown that the role of enterococci in intraabdominal infections originating in the community and involving previously healthy patients is minor, since surgery and broad-spectrum antimicrobial drugs that do not target enterococci are often sufficient in managing these infections successfully. In the last few decades, however, these organisms have become prominent as a cause of intraabdominal infections in hospitalized patients because of the emergence and spread of vancomycin resistance among enterococci and the increase in rates of nosocomial infections due to multidrug-resistant *E. faecium* isolates. In fact, several studies have now documented treatment failures due to enterococci, with consequently increased rates of postoperative complications and death in intraabdominal infections. Thus, anti-enterococcal therapy is recommended for nosocomial peritonitis in immunocompromised or severely ill patients who have had a prolonged hospital stay, have undergone multiple procedures, have persistent abdominal sepsis and collections, or have risk factors for the development of endocarditis (i.e., prosthetic or damaged heart valves). Conversely, treatment for enterococci in the first episode of intraabdominal infections originating in the community and affecting previously healthy patients with no important cardiac risk factors for endocarditis does not appear to be beneficial.

Enterococci are commonly isolated from soft tissue infections (Chap. 125), particularly those involving surgical wounds (Chap. 131). In fact, these organisms rank third as agents of nosocomial surgical-site infection, with *E. faecalis* the most frequently isolated species. The clinical relevance of enterococci in some of these infections—as in intraabdominal infections—is a matter of debate; differentiating between colonization and true infection can be challenging, although in some cases enterococci have been recovered from lung, liver, and skin abscesses. Diabetic foot and decubitus ulcers are often colonized with enterococci and may be the portal of entry for bone infections.

Other Infections

Enterococci are well-known causes of neonatal infections including sepsis (mostly late-onset), bacteremia, meningitis, pneumonia, and UTI. Outbreaks of enterococcal sepsis in neonatal units have been well documented. Risk factors for enterococcal disease in newborns include prematurity, low birth weight, indwelling devices, and abdominal surgery. Enterococci have also been described as etiologic agents of bone and joint infections including vertebral osteomyelitis, usually in patients with underlying conditions such as diabetes or endocarditis. Similarly, enterococci have been isolated from bone infections in patients who have undergone arthroplasty or reconstruction of fractures with the placement of hardware. Since enterococci can produce a biofilm that is likely to alter the efficacy of otherwise active anti-enterococcal agents, treatment of infections that involve foreign material is

challenging, and removal of the hardware may be necessary to eradicate the infection. Rare cases of enterococcal pneumonia, lung abscess, and spontaneous empyema have been reported.

Treatment: Enterococcal Infections

GENERAL PRINCIPLES

Enterococci are intrinsically resistant and/or tolerant to several antimicrobial agents [with *tolerance* defined as lack of killing by drug concentrations 16 times higher than the minimal inhibitory concentration (MIC)]. Monotherapy for endocarditis with a β-lactam antibiotic (to which many enterococci are tolerant) has produced disappointing results, with low cure rates at the end of therapy. However, the addition of an aminoglycoside to a cell wall–active agent (a β-lactam or a glycopeptide) increases cure rates and eradicates the organisms; moreover, this combination is synergistic and bactericidal in vitro. Therefore, combination therapy with a cell wall–active agent and an aminoglycoside is the standard of care for endovascular infections caused by enterococci. This synergistic effect can be explained, at least in part, by the increased penetration of the aminoglycoside into the bacterial cell, presumably as a result of cell wall alterations attributable to the β-lactam or glycopeptide. Nonetheless, attaining synergistic bactericidal activity in the treatment of severe enterococcal infections has become increasingly difficult because of the development of resistance to virtually all antibiotics available for this purpose.

The treatment of *E. faecalis* differs substantially from that of *E. faecium* (**Tables 137-1** and **137-2**), mainly because of differences in resistance profiles (see below); for example, resistance to ampicillin and vancomycin is rare in *E. faecalis*, whereas these antibiotics are only infrequently useful against current isolates of *E. faecium*. Moreover, as a consequence of the challenges and therapeutic limitations posed by the emergence of drug resistance, valve replacement may need to be considered in the treatment of endocarditis caused by multidrug-resistant enterococci. Less severe infections are often related to indwelling intravascular catheters; removal of the catheter increases the likelihood of enterococcal eradication of the organism by a short course of appropriate antimicrobial therapy.

Table 137–1 Suggested Regimens for the Management of Infections Caused by *Enterococcus Faecalis*

Clinical Syndrome	Ampicillin or Penicillin[a]	High-Level Resistance to Aminoglycosides[b]	Suggested Therapeutic Options[c]
Endovascular infections (includes endocarditis)	Susceptible	No	Ampicillin (12 g/d IV in divided doses q4h or by continuous infusion) *or* penicillin (18–30 million units/d IV in divided doses q4h or by continuous infusion) *plus* an aminoglycoside[d] Vancomycin (15–20 mg/kg per dose q8–12h, not to exceed 2 g per dose)[e] *plus* an aminoglycoside[d]
	Susceptible	Yes	Ampicillin (12 g/d IV in divided doses q4h) *plus* ceftriaxone (2 g q12h) *or* cefotaxime High-dose daptomycin[f] + another active agent[g] Vancomycin (15–20 mg/kg per dose q8–12h, not to exceed 2 g per dose)[e] Ampicillin *plus* imipenem
Nonendovascular bacteremia[h]	Susceptible	No	Ampicillin (12 g/d IV in divided doses q4h) *or* penicillin (18 mU/d IV in divided doses q4h)[i] Vancomycin (15–20 mg/kg per dose q8–12h, not to exceed 2 g per dose)[e,f]
	Susceptible	Yes	Ampicillin[j] (12 g/d IV in divided doses q4h) *or* penicillin Vancomycin (15–20 mg/kg per dose q8–12 h, not to exceed 2 g per dose)[e] High-dose daptomycin[f]
Meningitis	Susceptible	No	Ampicillin (20–24 g/d IV in divided doses q4h) *or* penicillin (24 mU/d IV in divided doses q4h) *plus*

			an aminoglycoside[k] Vancomycin (500–750 mg IV q6h)[e] *plus* an aminoglycoside[k] Linezolid
	Susceptible	Yes	Ampicillin (20–24 g/d IV in divided doses q4h) *or* penicillin (24 mU/d IV in divided doses q4h) *plus* ceftriaxone (2 g q12h) *or* cefotaxime Vancomycin (500–750 mg IV q6h)[e] Linezolid High-dose daptomycin[f] (*plus* intrathecal daptomycin) ± another active agent[g]
Urinary tract infections (uncomplicated)	Not applicable	Not applicable	Ampicillin (500 mg IV or PO q6h) Nitrofurantoin (100 mg PO q6h) Fosfomycin (3-g single dose PO)[l]

[a] In rare cases, β-lactamase-producing isolates may be found. Because these isolates are not detected by conventional MIC determination, additional tests (e.g., the nitrocefin disk test) are recommended for isolates from endocarditis. The use of ampicillin/sulbactam (12–24 g/d) is suggested in these cases.

[b] Determined by the clinical microbiology laboratory only for gentamicin or streptomycin as growth of enterococci in brain-heart infusion agar containing gentamicin (500 μg/mL) and streptomycin (2000 μg/mL). Resistance to one compound does not indicate resistance to the other, and the laboratory reports high-level resistance to each compound individually. In this table, *Yes* to high-level resistance to aminoglycosides indicates that the laboratory has reported resistance to synergism with both gentamicin and streptomycin.

[c] Author's first-choice regimen is underlined for each category.

[d] Gentamicin (1–1.5 mg/kg IV q8h) or streptomycin (15 mg/kg per day IV or IM in two divided doses).

[e] Vancomycin is recommended only as an alternative to β-lactam agents in case of allergy, toxicity, or inability to desensitize. Monitoring of vancomycin levels is recommended, although no data for enterococci are available; a regimen that achieves trough serum levels of 15–20 μg/mL is suggested by some experts, although high doses may predispose to renal toxicity. CSF concentrations may also be determined. Vancomycin-resistant strains of *E. faecalis* have been reported.

[f] Consider doses of 8–12 mg/kg per day (off-label use).

[g] Active agents may include ampicillin, a fluoroquinolone (which, if the isolate is susceptible, may be favored in meningitis), and tigecycline.

[h] In selected cases of catheter-associated bacteremia, removal of the catheter and a short course of therapy (~5–7 days) may be sufficient. A single positive blood culture that is likely to be associated with a catheter in a patient who is otherwise doing well may not require therapy after removal of the catheter.

[i] The addition of an aminoglycoside may be considered in severe infections.

[j] The addition of ceftriaxone (or cefotaxime) may be considered in severe infections.

[k] The addition of intrathecal or intraventricular therapy with gentamicin (2–10 mg/d) or vancomycin (10–20 mg/d), when the isolate is susceptible, has been suggested by some authorities for recalcitrant cases.

[l] Approved by the FDA only for uncomplicated urinary tract infections caused by vancomycin-susceptible *E. faecalis*.

Table 137–2 Suggested Regimens for the Management of Infections Caused by Vancomycin-Resistant *Enterococcus Faecium*

Clinical Syndrome	Ampicillin MIC (μg/mL)	High-Level Resistance to Aminoglycosides[a]	Suggested Therapeutic Options[b]

Endovascular infections (includes endocarditis)	≤64	No	High-dose ampicillin[c] *plus* an aminoglycoside[d] See regimens for MIC >64 μg/mL
	>64	No	High-dose daptomycin[e] *plus* an aminoglycoside[d] ± another active agent[f] Q/D[g] (22.5 mg/kg per day in divided doses q8h) ± another active agent[f] Linezolid[g] (600 mg IV q12h) ± another active agent[f]
	≤64	Yes	High-dose ampicillin[c] *plus* high-dose daptomycin[e] Q/D[g] (22.5 mg/kg per day in divided doses q8h) *plus* high-dose ampicillin[c] *or* doxycycline (100 mg IV q12h) with rifampin (300 mg PO q12h, if susceptible) High-dose ampicillin[c] *plus* imipenem-cilastatin (500 mg IV q6h)[h]
	>64	Yes	High-dose daptomycin[e] *plus* another active agent[f] Q/D[g] (22.5 mg/kg per day in divided doses q8h) *plus* doxycycline (100 mg IV q12h) with rifampin (300 mg PO q12h, if susceptible) Linezolid[g] (600 mg IV q12h) ± another active agent[f] High-dose ampicillin[c] ± an aminoglycoside[d]
Non endovascular bacteremia[i]	≤64	No	Q/D[g] ± another active agent[f]
	>64	No	Daptomycin[e] ± an aminoglycoside[d] Linezolid[g] ± another active agent[f] High-dose ampicillin[c] ± Q/D[g] (22.5 mg/kg per day in divided doses q8h)
	≤64	Yes	High-dose ampicillin[c] ± daptomycin[e] Linezolid[g] (600 mg IV q12h) ± another active agent[f] Q/D[g] (22.5 mg/kg per day in divided doses q8h) ± another active agent[f]
	>64	No	Daptomycin[e] ± another active agent[f] Linezolid[g] (600 mg IV q12h) ± another active agent[f]
Meningitis[j]	<16	No	High-dose ampicillin[c] *plus* gentamicin (5.1–7 mg/kg, single daily dose) *or* streptomycin (15 mg/kg, single daily dose)
	≥16	No	Linezolid ± another CSF-penetrating active agent[k] High-dose daptomycin[e] (plus intrathecal daptomycin[j]) *plus* gentamicin (5.1–7 mg/kg, single daily dose) *or* streptomycin (15 mg/kg, single daily dose)[j] ± another CSF-penetrating active agent[k]
	<16	Yes	High-dose ampicillin[c] *plus* high-dose daptomycin[e] (plus intrathecal daptomycin)[j] Linezolid ± another CSF-penetrating active agent[k]

			Q/D[j] (22.5 mg/kg per day in divided doses q8h) *plus* intrathecal Q/D *plus* high-dose ampicillin[c]
	≥16	Yes	Linezolid ± another CSF-penetrating active agent[k] High-dose daptomycin[e] ± another CSF-penetrating active agent[k] Q/D[j] (22.5 mg/kg per (plus intrathecal daptomycin) in divided doses q8h) *plus* intrathecal Q/D ± another CSF-penetrating active agent[k]
Urinary tract infections	≤512	Not applicable	Nitrofurantoin (100 mg PO q6h) Fosfomycin (3 g PO, one dose)[l] Ampicillin or amoxicillin (2 g IV or PO q4–6h)[m]

[a] *Yes* to high-level resistance to aminoglycosides indicates resistance to both gentamicin and streptomycin (the only two aminoglycosides recommended for the treatment of enterococcal infections). Resistance to one compound does not indicate resistance to the other.

[b] Author's first-choice regimen(s) are underlined for each category.

[c] Doses up to 30 g/d IV in divided doses q4h may be considered; clinical data on safety at these high doses are not available.

[d] Gentamicin (1–1.5 mg/kg IV q8h) or streptomycin (15 mg/kg per day IV or IM in two divided doses).

[e] Daptomycin at a dosage of 8–12 mg/kg once daily is suggested (off-label); close monitoring of creatine phosphokinase levels is recommended throughout therapy because of possible rhabdomyolysis.

[f] Agents with potential activity include doxycycline with rifampin or tigecycline (50 mg IV q12h after an initial loading dose of 100 mg IV) or fluoroquinolones (if the isolate is susceptible to each agent).

[g] Quinupristin-dalfopristin (Q/D) and linezolid are listed in the American Heart Association recommendations for the treatment of endocarditis caused by vancomycin- and ampicillin-resistant *E. faecium*.

[h] If the imipenem MIC is <32 μg/mL.

[i] In selected cases of catheter-associated bacteremia, removal of the catheter and a short course of therapy (~5–7 days) may be sufficient. A single positive blood culture that is likely to be associated with a catheter in a patient who is otherwise doing well may not require therapy after removal of the catheter.

[j] Intrathecal gentamicin (if no high-level resistance is detected; 2–10 mg/d) or Q/D (1–5 mg/d) has been used in combination with systemic therapy in refractory cases of postoperative meningitis. If Q/D is chosen, simultaneous systemic and intrathecal therapy is suggested. Intraventricular daptomycin has been used in two cases of meningitis.

[k] Fluoroquinolone antibiotics (e.g., moxifloxacin) and rifampin (if the isolate is susceptible to each agent) reach therapeutic levels in CSF.

[l] Approved by the FDA only for uncomplicated urinary tract infections caused by vancomycin-susceptible *E. faecalis*.

[m] Concentrations of amoxicillin in urine far exceed those in serum and can be potentially effective even against isolates with high MICs. Doses up to 12 g/d are suggested for isolates with MICs of ≥64 μg/mL.

Choice of Antimicrobial Agents

Among the β-lactam agents, the most active are the aminopenicillins (ampicillin, amoxicillin) and ureidopenicillins (i.e., piperacillins); next most active are penicillin G and imipenem. Against *E. faecium*, a combination of high-dose ampicillin (up to 30 g/d) and an aminoglycoside (Table 137–2) has been suggested even for ampicillin-resistant strains if the MIC is <64 μg/mL, since a plasma ampicillin concentration of >100 μg/mL can be achieved at high doses. The only two aminoglycosides recommended for synergistic therapy in severe enterococcal infections are gentamicin and streptomycin. The use of amikacin is discouraged, tobramycin should never be used against *E. faecium*, and aminoglycoside monotherapy is not effective. Vancomycin is an alternative to β-lactam drugs for the treatment of *E. faecalis* infections but is less useful against *E. faecium* because resistance is common. Cephalosporins (except ceftobiprole for E. faecalis) are inactive against

enterococci.

Linezolid and quinupristin/dalfopristin (Q/D) are approved by the U.S. Food and Drug Administration (FDA) for the treatment of some VRE infections (Table 137–2). Linezolid is not bactericidal, and its use in severe endovascular infections has produced mixed results; therefore, it is recommended only as an alternative to other agents. In addition, linezolid may cause significant toxicities (thrombocytopenia and peripheral neuropathy) when used in regimens given for >2 weeks. Nonetheless, this drug may play a role in the treatment of enterococcal meningitis and other CNS infections, although clinical data are limited. Q/D is not active against most *E. faecalis* isolates, and its in vivo efficacy against *E. faecium* may be compromised by resistance (see below). Adverse reactions to Q/D (including pain and inflammation at the infusion site and severe arthralgias and myalgias leading to discontinuation of treatment) are common. Therefore, this drug should be used with caution and probably combined with other agents (Table 137–2). The lipopeptide daptomycin is a bactericidal antibiotic with potent in vitro activity against all enterococci. Although daptomycin is not approved by the FDA for the treatment of VRE or *E. faecium* infections, it has been used alone (at high dosage) or in combination with other agents with apparent success against multidrug-resistant enterococcal infections (Tables 137-1 and 137-2). The main adverse reactions to daptomycin are elevated creatinine phosphokinase levels and eosinophilic pneumonitis. Daptomycin is not useful against pulmonary infections because the pulmonary surfactant inhibits its antibacterial activity. Although the glycylcycline tigecycline is active in vitro against all enterococci (regardless of the isolates' vancomycin susceptibility), its use as monotherapy for endovascular or severe enterococcal infections is not recommended because of low attainable blood levels. Telavancin, a lipoglycopeptide approved by the FDA for the treatment of skin and soft tissue infections, is active against vancomycin-susceptible enterococci but less so against VRE. Oritavancin, a compound of the same class that is active against VRE, is in the late stages of clinical development and may offer promise for the treatment of VRE infections in the future.

Antimicrobial Resistance

As mentioned above, resistance to ampicillin continues to be observed only infrequently in *E. faecalis*, although rare outbreaks caused by β-lactamase-producing isolates have occurred in the United States and Argentina. However, ampicillin resistance is common in *E. faecium*. The mechanism of ampicillin resistance in *E. faecium* is related to a penicillin-binding protein (PBP) designated PBP5, which is the target of β-lactam antibiotics. PBP5 exhibits lower affinity for ampicillin than other PBPs and can synthesize cell wall in the presence of this antibiotic, even when other PBPs are inhibited. Two common mechanisms of high-level ampicillin resistance (MIC, >64 μg/mL) in clinical strains are (1) mutations in the PBP5-encoding gene that further decrease the affinity of PBP5 for ampicillin and (2) hyperproduction of PBP5. These factors preclude the use of all β-lactam agents in the treatment of *E. faecium* infections.

Vancomycin is a glycopeptide antibiotic that inhibits cell wall peptidoglycan synthesis in susceptible enterococci and has been widely used against enterococcal infections in clinical practice when the use of penicillins is limited by resistance, allergy, or adverse reactions. This effect is mediated by binding of the antibiotic to peptidoglycan precursors (UDP-MurNAc-pentapeptides) upon their exit from the bacterial cytoplasm. The interaction of vancomycin with the peptidoglycan is specific and involves the last two d-alanine residues of the precursor. The first isolates of VRE were documented in 1986, and vancomycin resistance (particularly in *E. faecium*) has since increased considerably around the world. The mechanism involves the replacement of the last d-alanine residue of peptidoglycan precursors with d-lactate or d-serine, with consequent high- and low-level resistance, respectively. There is significant heterogeneity among isolates, but either substitution substantially decreases the affinity of vancomycin for the peptidoglycan; with the d-lactate substitution, the MIC is increased by up to one thousand-fold. Vancomycin-resistant organisms also produce enzymes that destroy the D-alanine–D-alanine-ending precursors, ensuring that additional binding sites for vancomycin are not available.

High-level resistance to aminoglycosides (of which gentamicin and streptomycin are the only two tested in clinical laboratories) abolishes the synergism observed between cell wall–active agents and the aminoglycoside. This important phenotype is routinely sought in isolates from serious infections (Tables 137-1 and 137-2). The laboratory reports high-level resistance as gentamicin and streptomycin MICs of >500 μg/mL and >2000 μg/mL, respectively (agar dilution method), or as "SYN-R" (resistance to synergism). Genes encoding aminoglycoside-modifying enzymes are usually the cause of high-level resistance to these compounds and are widely disseminated among enterococci, decreasing the options for the treatment of severe enterococcal infections. The aforementioned enterococcal resistance to newer antibiotics such as linezolid (usually due to mutations in the 23S rRNA genes), Q/D, daptomycin, and tigecycline further reduces therapeutic alternatives.

FURTHER READINGS

Arias CA, Murray BE: *Enterococcus* species, *Streptococcus bovis* group and *Leuconostoc* species, in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). Philadelphia, Churchill Livingstone Elsevier, 2010, pp 2643–2654

——— et al: Management of multidrug-resistant enterococcal infections. Clin Microbiol Infect 16:555, 2010

Brandl K et al: Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. Nature 455:804, 2008 [PMID: 18724361]

Facklam RR et al: History, taxonomy, biochemical characteristics and antibiotic susceptibility testing of enterococci, in *The Enterococci: Pathogenesis, Molecular Biology and Antibiotic Resistance* , MS Gilmore (ed). Washington, DC, ASM Press, 2002, pp 1–54

Nallapareddy SR et al: Endocarditis and biofilm associated pili of *Enterococcus faecalis*. J Clin Invest 116:2799, 2006 [PMID: 17016560]

Willems RJ, van Schaik W: Transition of *Enterococcus faecium* from commensal organism to nosocomial pathogen. Future Microbiol 4:1125, 2009 [PMID: 19895216]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIPHTHERIA

Diphtheria is a nasopharyngeal and skin infection caused by *Corynebacterium diphtheriae*. Toxigenic strains of *C. diphtheriae* produce a protein toxin that causes systemic toxicity, myocarditis, and polyneuropathy. The toxin is associated with the formation of pseudomembranes in the pharynx during respiratory diphtheria. While toxigenic strains most frequently cause pharyngeal diphtheria, nontoxigenic strains commonly cause cutaneous disease.

In the United States and Europe, diphtheria has been controlled in recent years with effective vaccination, although sporadic outbreaks have occurred. Diphtheria is still common in the Caribbean, Latin America, and the Indian subcontinent, where mass immunization programs are not enforced. Large epidemics have occurred in the independent states formerly encompassed by the Soviet Union. Additional outbreaks have been reported in Algeria, China, and Ecuador.

Etiology

C. diphtheriae is a gram-positive, unencapsulated, nonmotile, nonsporulating bacillus. *C. diphtheriae* organisms have a characteristic club-shaped bacillary appearance and typically form clusters of parallel rays (palisades) that are referred to as *Chinese characters*. In the specific laboratory media recommended for the cultivation of *C. diphtheriae*, tellurite, colistin, or nalidixic acid allows selective isolation of the organism in the presence of other autochthonous pharyngeal microbes. Human isolates of *C. diphtheriae* may display nontoxigenic (tox^-) or toxigenic (tox^+) phenotypes. Corynebacteriophage beta carries the structural gene (*tox*) encoding diphtheria toxin, and a family of closely related corynebacteriophages are responsible for toxigenic conversion of tox^- *C. diphtheriae* to the tox^+ phenotype. Moreover, lysogenic conversion from a nontoxigenic to a toxigenic phenotype has been shown to occur in situ. Growth of toxigenic strains of *C. diphtheriae* under iron-limiting conditions leads to the optimal expression of diphtheria toxin, and these conditions are believed to trigger *tox* expression and subsequent pathogenesis during human infection.

Epidemiology

C. diphtheriae is transmitted via the aerosol route, primarily during close contact. There are no significant reservoirs other than humans. The incubation period for respiratory diphtheria is 2–5 days; however, disease can develop as long as 10 days after exposure. Before the vaccine era, most individuals over the age of 10 were immune to *C. diphtheriae*; infants were protected by maternal IgG antibodies but became susceptible after ~6 months of age. Thus, the disease was seen primarily in children and nonimmune young adults. In temperate regions, respiratory diphtheria occurs year-round but is most common during winter months.

The development of diphtheria antitoxin and diphtheria toxoid vaccine led to the near-elimination of diphtheria in Western countries. The annual peak incidence rate was 191 cases per 100,000 population in the United States in 1921; in contrast, since 1980, the annual figure for the United States as a whole has been <5 cases. Nevertheless, pockets of colonization have persisted in North America, particularly in South Dakota, Ontario, and Washington state. Immunity induced by vaccination during childhood gradually decreases in adulthood. An estimated 30% of men 60–69 years old have antitoxin titers below the protective level. In addition to older age and lack of vaccination, risk factors for diphtheria outbreaks include alcoholism, low socioeconomic status, crowded living conditions, and Native American ethnic background. An outbreak that occurred in Seattle in 1972–1982 included 1100 cases, primarily manifesting as cutaneous disease. During the 1990s in the states of the former Soviet Union, a much larger diphtheria epidemic caused >150,000 cases and >5000 deaths. Clonally related toxigenic *C. diphtheriae* strains of the ET8 complex were associated with this outbreak. Given that the ET8 complex expressed a toxin against which the prevalent diphtheria toxoid vaccine was effective, the epidemic was attributed to failure of the public health infrastructure to effectively vaccinate the population. Beginning in 1998, the epidemic was controlled by mass vaccination programs. During the epidemic, the incidence rate was high among individuals from >15 years of age up to 50 years of age. Socioeconomic instability, migration, deteriorating public health programs, frequent vaccine shortages, delays in implementation of vaccination and of treatment in response to cases, and lack of public education and awareness were contributing factors in that outbreak.

Significant outbreaks of diphtheria and diphtheria-related mortality continue to be reported from many

developing countries, particularly in Africa and Asia. Statistics collected by the World Health Organization indicate the occurrence of ~7000 reported diphtheria cases in 2008 and ~5000 diphtheria deaths in 2004. Although ~82% of the global population has been adequately vaccinated, only 26% of countries have successfully vaccinated >80% of individuals in all districts.

Cutaneous diphtheria is usually a secondary infection that follows a primary skin lesion due to trauma, allergy, or autoimmunity. Most often, isolates from cases of cutaneous disease lack the *tox* gene and therefore do not express diphtheria toxin. In tropical regions, cutaneous diphtheria is more common than respiratory diphtheria. In contrast to respiratory disease, cutaneous diphtheria is not a reportable disease in United States.

Nontoxigenic strains of *C. diphtheriae* have also been associated with bacteremia and invasive disease in the urban poor in Vancouver, Canada, and with pharyngitis in Europe. Outbreaks have occurred among homosexual men and IV drug users.

Pathogenesis and Immunology

Diphtheria toxin, produced by toxigenic strains of *C. diphtheriae*, is the primary virulence factor in clinical disease. The toxin is synthesized in precursor form; is released as a 535-amino-acid, single-chain protein; and has an LD_{50} of ~100 ng/kg of body weight. The toxin is produced in the pseudomembranous lesion and is taken up into the bloodstream, through which it is distributed to all organ systems. Once bound to its cell surface receptor (a heparin-binding, epidermal growth factor–like precursor), the toxin is internalized by receptor-mediated endocytosis and enters the cytosol from an acidified early endosomal compartment. In vitro, the toxin may be separated into two chains after digestion with serine proteases: the N-terminal A fragment and the C-terminal B fragment. Delivery of the A fragment into the eukaryotic cell cytosol results in irreversible inhibition of protein synthesis by NAD+-dependent ADP ribosylation of elongation factor 2. The eventual result is the death of the cell.

In 1926, Ramon at the Institut Pasteur found that formalinization of diphtheria toxin resulted in the production of diphtheria toxoid, which was nontoxic but highly immunogenic. Subsequent studies showed that immunization with diphtheria toxoid elicited antibodies that neutralized the toxin and prevented most manifestations of diphtheria. In the 1930s, mass immunization of children and susceptible adults commenced in the United States and Europe.

Individuals with an antitoxin titer of >0.01 unit/mL are at low risk of diphtheria disease. In populations where a majority of individuals have protective antitoxin titers, the carrier rate for toxigenic strains of *C. diphtheriae* decreases and the overall risk of diphtheria among susceptible individuals is reduced. Nevertheless, individuals with nonprotective titers may contract diphtheria through either travel or exposure to individuals who have recently returned from regions where the disease is endemic.

Characteristic pathologic findings of diphtheria include mucosal ulcers with a pseudomembranous coating composed of an inner band of fibrin and a luminal band of neutrophils. Initially white and firmly adherent, in advanced diphtheria the pseudomembranes turn gray and even green or black as necrosis progresses. Mucosal ulcers result from toxin-induced necrosis of the epithelium accompanied by edema, hyperemia, and vascular congestion of the submucosal base. A fibrinosuppurative exudate from the ulcer develops into the pseudomembrane. Ulcers and pseudomembranes in severe respiratory diphtheria may extend from the pharynx into medium-sized bronchial airways. Expanding and sloughing membranes may result in fatal airway obstruction.

Approach to the Patient: Diphtheria

Although diphtheria is rare in the United States and other developed countries, this diagnosis should be considered in patients who have severe pharyngitis, particularly with difficulty swallowing, respiratory compromise, or signs of systemic disease including myocarditis or generalized weakness. In the differential diagnosis, the leading causes of pharyngitis that should be considered are respiratory viruses (rhinoviruses, influenza viruses, parainfluenza viruses, coronaviruses, and adenoviruses; ~25% of cases), group A streptococci (15–30%), group C streptococci (~5%), atypical bacteria such as *Mycoplasma pneumoniae* and *Chlamydophila pneumoniae* (15–20% in some series), and other viruses such as herpes simplex virus (~4%) and Epstein-Barr virus (EBV; <1% in infectious mononucleosis). Less common causes are acute HIV infection, infection with *Neisseria gonorrhoeae*, fusobacterial infection (e.g., Lemierre syndrome), and thrush due to *Candida albicans* or other *Candida* species. The presence of a pharyngeal pseudomembrane or an extensive exudate should prompt consideration of diphtheria (Fig. 138-1).

Figure 138-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Respiratory diphtheria due to toxigenic *C. diphtheriae* producing exudative pharyngitis in a 47-year-old woman with neck edema and a pseudomembrane extending from the uvula to the pharyngeal wall. The characteristic white pseudomembrane is caused by diphtheria toxin–mediated necrosis of the respiratory epithelial layer, producing fibri-nous coagulative exudate. Submucosal edema adds to airway narrowing. The pharyngitis is acute in onset, and respiratory obstruction from the pseudomembrane may occur in severe cases. Inoculation of pseudomembrane fragments or submembranous swabs onto Löffler's or tellurite selective medium reveals *C. diphtheriae.* *(Photograph by P. Strebel, MD, used by permission. From Kadirova R et al: J Infect Dis 181:S110, 2000.)*

Clinical Manifestations

RESPIRATORY DIPHTHERIA

The clinical diagnosis of diphtheria is based on the constellation of sore throat; adherent tonsillar, pharyngeal, or nasal pseudomembranous lesions; and low-grade fever. In addition, diagnosis requires the isolation of *C. diphtheriae* or the histopathologic isolation of compatible gram-positive organisms. The Centers for Disease Control and Prevention (CDC) recognizes confirmed respiratory diphtheria (laboratory proven or epidemiologically linked to a culture-confirmed case) and probable respiratory diphtheria (clinically compatible but not laboratory proven or epidemiologically linked). Carriers are defined as individuals who have positive cultures for *C. diphtheriae* and either are asymptomatic or have symptoms but lack pseudomembranes. Most patients seek medical care for initial manifestations of sore throat and fever. Occasionally, weakness, dysphagia, headache, and voice change are the initial manifestations. Neck edema and difficulty breathing are seen in more advanced cases and carry a poor prognosis.

The systemic manifestations of diphtheria stem from the effects of diphtheria toxin and include weakness as a result of neurotoxicity and cardiac arrhythmias or congestive heart failure due to myocarditis. The pseudomembranous lesion is most often located in the tonsillopharyngeal region. Less commonly, the lesions are detected in the larynx, nares, and trachea or bronchial passages. Large pseudomembranes are associated with severe disease and a poor prognosis. A few patients develop massive swelling of the tonsils and present with "bull-neck" diphtheria, which results from massive edema of the submandibular and paratracheal region and is further characterized by foul breath, thick speech, and stridorous breathing. The diphtheritic pseudomembrane is gray or whitish and sharply demarcated. Unlike the exudative lesion associated with streptococcal pharyngitis, the pseudomembrane in diphtheria is tightly adherent to the underlying tissues. Attempts to dislodge the membrane may cause bleeding. Hoarseness suggests laryngeal diphtheria, in which laryngoscopy may be diagnostically helpful.

CUTANEOUS DIPHTHERIA

This is a variable dermatosis most often characterized by punched-out ulcerative lesions with necrotic sloughing or pseudomembrane formation (Fig. 138-2). The diagnosis requires cultivation of *C. diphtheriae* from lesions, which most commonly occur on the extremities. Patients usually seek medical attention because of nonhealing or enlarging skin ulcers, which may be associated with a preexisting wound or dermatoses such as eczema, psoriasis, and venous stasis disease. The lesions rarely

exceed 5 cm.

Figure 138-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cutaneous diphtheria due to nontoxigenic *C. diphtheriae* on the lower extremity. *(From the Centers for Disease Control and Prevention.)*

OTHER CLINICAL MANIFESTATIONS

C. diphtheriae causes rare cases of endocarditis and septic arthritis, most often in patients with preexisting risk factors such as cardiac valvular disease, injection drug use, or cirrhosis.

Complications

Airway obstruction poses a significant early risk in patients presenting with advanced diphtheria. Pseudomembranes may slough and obstruct the airway or may advance to the larynx or into the tracheobronchial tree. Children are particularly prone to obstruction because of their small airways.

Polyneuropathy and myocarditis are late toxic manifestations of diphtheria. During the outbreak in the Kyrgyz Republic in 1995, myocarditis was seen in 22% and neuropathy in 5% of hospitalized patients. The mortality rate was 7% among patients with myocarditis as opposed to 2% among those without myocardial manifestations. The median time to death in hospitalized patients was 4.5 days. Myocarditis is typically associated with dysrhythmia of the conduction tract and dilated cardiomyopathy.

Neurologic manifestations may appear during the first or second week of illness, typically beginning with dysphagia and nasal dysarthria and progressing to other signs of cranial nerve involvement, including weakness of the tongue and facial numbness. Ciliary paralysis, which is typical, manifests as blurred vision due to paralysis of pupillary accommodation, with a preserved light reflex. Cranial neuropathy may be followed by respiratory and abdominal muscle weakness requiring artificial ventilation. Several weeks later—sometimes as cranial neuropathy is improving—a generalized sensorimotor polyneuropathy may appear, with prominent autonomic manifestations (including hypotension) in some cases. The clinical syndrome and the findings on lumbar puncture of raised levels of protein without pleocytosis in cerebrospinal fluid resemble Guillain-Barré syndrome (Chap. 385). Pathologically, diphtheria neuropathy is a noninflammatory demyelinating disorder mediated by the exotoxin. Gradual improvement is the rule in patients who survive the acute phase.

Other complications of diphtheria include pneumonia, renal failure, encephalitis, cerebral infarction, and pulmonary embolism. Serum sickness can result from treatment with diphtheria antitoxin (see "Diphtheria Treatment," below).

Diagnosis

The diagnosis of diphtheria is based on clinical signs and symptoms plus laboratory confirmation. Respiratory diphtheria should be considered in patients with sore throat, pharyngeal exudates, and fever. Other symptoms may include hoarseness, stridor, or palatal paralysis. The presence of a pseudomembrane should prompt consideration of diphtheria. Once a clinical diagnosis of diphtheria is made, diphtheria antitoxin should be administered as soon as possible.

Laboratory diagnosis is based either on cultivation of *C. diphtheriae* or toxigenic *C. ulcerans* from the site of infection or on the demonstration of local lesions with characteristic histopathology. *C. pseudodiphtheriticum*, a nontoxigenic organism, is a common component of the normal throat flora and does not pose a significant risk. Throat samples should be submitted to the laboratory for culture with

the notation that diphtheria is being considered. This information should prompt cultivation on special selective medium and subsequent biochemical testing to differentiate *C. diphtheriae* from other nasopharyngeal commensal corynebacteria. All laboratory isolates of *C. diphtheriae*, including nontoxigenic strains, should be submitted to the CDC.

A diagnosis of cutaneous diphtheria requires laboratory confirmation since the lesions are not characteristic and are clinically indistinguishable from other dermatoses. Diphtheritic ulcers occasionally—but not consistently—have a punched-out appearance (Fig. 138-2). Patients in whom cutaneous diphtheria is identified should have the nasopharynx cultured for *C. diphtheriae*. The laboratory media for cutaneous diphtheria are the same as those used for respiratory diphtheria: Löffler's or Tinsdale's selective medium in addition to nonselective medium such as blood agar. As has been mentioned, respiratory diphtheria remains a notifiable disease in the United States, whereas cutaneous diphtheria is not.

Treatment: Diphtheria

DIPHTHERIA ANTITOXIN

Prompt administration of diphtheria antitoxin is critical in the management of respiratory diphtheria. The antitoxin—a horse antiserum—is effective in reducing the extent of local disease as well as the risk of complications of myocarditis and neuropathy. Rapid institution of antitoxin therapy is also associated with a significant reduction in mortality risk. Because diphtheria antitoxin cannot neutralize cell-bound toxin, prompt initiation is important. This product, which is no longer made commercially in the United States, is available from the CDC under an investigational new drug protocol and may be obtained by calling the Emergency Operations center at 770-488-7100; the relevant Web site is *www.cdc.gov/vaccines/vpd-vac/diphtheria/dat/dat-main.htm*. The current protocol for the use of antitoxin includes a test dose to rule out immediate-type hypersensitivity. Patients who exhibit hypersensitivity require desensitization before a full therapeutic dose of antitoxin is administered.

ANTIMICROBIAL THERAPY

Antibiotics are used in the management of diphtheria primarily to prevent transmission to other susceptible contacts. Recommended options for the treatment of patients with respiratory diphtheria are as follows: (1) procaine penicillin G at a dosage of 600,000 units (for children, 12,500–25,000 U/kg) IM every 12 h until the patient can swallow comfortably, after which oral penicillin V is given at 125–250 mg four times daily to complete a 14-day course; or (2) erythromycin at a dosage of 500 mg IV every 6 h (for children, 40–50 mg/kg per day IV in two or four divided doses) until the patient can swallow comfortably, after which 500 mg is given by mouth four times daily to complete a 14-day course.

A clinical study in Vietnam found that penicillin was associated with a more rapid resolution of fever and a lower rate of bacterial resistance than erythromycin; however, relapses were more common with penicillin. Erythromycin therapy targets protein synthesis and thus offers the presumed benefit of stopping toxin synthesis more quickly than a cell wall–active β-lactam agent. Alternative agents for patients who are allergic to penicillin or cannot take erythromycin include rifampin and clindamycin. Eradication of *C. diphtheriae* should be documented at least 1 day after antimicrobial therapy is complete. A repeat throat culture 2 weeks later is recommended. For patients in whom the organism is not eradicated after a 14-day course of erythromycin or penicillin, an additional 10-day course followed by repeat culture is recommended.

Cutaneous diphtheria should be treated as described above for respiratory disease. Individuals infected with toxigenic strains should receive antitoxin. It is important to treat the underlying cause of the dermatoses in addition to the superinfection with *C. diphtheriae*.

Patients who recover from respiratory or cutaneous diphtheria should have antitoxin levels measured. If diphtheria antitoxin has been administered, this test should be performed 6 months later. Patients who recover from respiratory or cutaneous diphtheria should receive the appropriate vaccine (see "Prevention," below) to ensure the development of protective antibody titers, which does not occur in all cases.

MANAGEMENT

Patients in whom diphtheria is suspected should be hospitalized in respiratory isolation rooms, with close monitoring of cardiac and respiratory function. A cardiac workup is recommended to assess the possibility of myocarditis. In patients with extensive pseudomembranes, consultation with an anesthesiologist or an ear, nose, and throat specialist is recommended because of the possibility that tracheostomy or intubation will be required. In some settings, pseudomembranes can be removed surgically. Treatment with glucocorticoids has not been shown to reduce the risk of myocarditis or polyneuropathy.

Prognosis

Fatal pseudomembranous diphtheria typically occurs in patients with nonprotective antibody titers and in unimmunized patients. The pseudomembrane may increase in size from the time it is first noted. Risk

factors for death include bull-neck diphtheria; myocarditis with ventricular tachycardia; atrial fibrillation; complete heart block; an age of >60 years or <6 months; alcoholism; extensive pseudomembrane elongation; and laryngeal, tracheal, or bronchial involvement. Another important predictor of fatal outcome is the interval between local disease development and antitoxin administration. Cutaneous diphtheria has a low mortality rate and is rarely associated with myocarditis or peripheral neuropathy.

Prevention

VACCINATION

Sustained campaigns for vaccination of children and adequate boosting vaccination of adults are responsible for the exceedingly low incidence of diphtheria in most developed nations. At present, diphtheria toxoid vaccine is coadministered with tetanus (with or without acellular pertussis) vaccine. DTaP (full-level diphtheria and tetanus toxoids and acellular pertussis vaccine, adsorbed) is the currently recommended vaccine for children up to the age of 7; DTaP replaced DTP (diphtheria and tetanus toxoids and whole-cell pertussis vaccine) in 1997. Tdap is a tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccine formulated for adolescents and adults. Tdap was licensed for use in the United States in 2005 and is the recommended booster vaccine for children 11–12 years old and the recommended catch-up vaccine for children 7–10 and 13–18 years old. As of 2006, it is recommended that (1) adults 19–64 years old receive a single dose of Tdap if their last dose of Td (tetanus and reduced-dose diphtheria toxoids, adsorbed) was >10 years earlier and (2) intervals of <10 years be implemented for Tdap vaccination of health care workers, adults anticipating contact with infants, and adults not previously vaccinated for pertussis. Adults who have received acellular pertussis vaccines should continue to receive decennial Td booster vaccinations. **The vaccination schedule is detailed in Chap. 122.**

PROPHYLAXIS OF CONTACTS

Close contacts of diphtheria cases should undergo throat culture to determine whether they are carriers. After samples for throat culture are obtained, antimicrobial prophylaxis should be considered for all close contacts, even those who are culture-negative. The options are 7–10 days of oral erythromycin or one dose of IM benzathine penicillin G (1.2 million units for persons ≥6 years old or 600,000 units for children <6 years old).

Contacts of diphtheria cases who have an uncertain immunization status should receive the appropriate diphtheria toxoid–containing vaccine. Tdap (rather than Td) is now recommended as the booster vaccine of choice for adults who have not recently received an acellular pertussis–containing vaccine. Carriers of *C. diphtheriae* in the community should be treated and vaccinated when identified.

NONDIPHTHERIAL CORYNEBACTERIA AND RELATED SPECIES

Nondiphtherial corynebacteria, which are also referred to as *diphtheroids* or *coryneforms*, are a widely diverse collection of bacteria that are taxonomically lumped together on the basis of their 16S rDNA signature nucleotides. The diversity of this group is exemplified by the wide range in guanine-plus-cytosine content (45–70%). Although frequently considered colonizers or contaminants, the nondiphtherial corynebacteria have been associated with invasive disease, particularly in immunocompromised patients. Specifically, for example, these organisms have been implicated in bacteremia, endocarditis, and other serious infections, particularly in association with catheters and prosthetic devices. Patients infected with nondiphtherial corynebacteria usually have significant medical comorbidity or immunosuppression. Several of these organisms, including *C. jeikeium* and *C. urealyticum*, are associated with resistance to multiple antibiotics. The related organism *Rhodococcus equi* is associated with necrotizing pneumonia and granulomatous infection, particularly in immunocompromised individuals. Other related species that can cause infections in humans are *Actinomyces* (formerly *Corynebacterium*) *pyogenes* and *Arcanobacterium* (formerly *Corynebacterium*) *haemolyticum*.

Microbiology and Laboratory Diagnosis

These organisms are non-acid-fast, catalase-positive, aerobic or facultatively anaerobic bacilli. Their colonial morphologies vary widely; some species are small and α-hemolytic (similar to lactobacilli), whereas others form large white colonies (similar to yeasts). Many nondiphtherial coryneforms require special medium (e.g., Löffler's, Tinsdale's, or telluride medium) for growth.

Epidemiology

Humans are the natural reservoirs for several nondiphtherial coryneforms, including *C. xerosis*, *C. pseudodiphtheriticum*, *C. striatum*, *C. minutissimum*, *C. jeikeium*, *C. urealyticum*, and *A. haemolyticum*. Animal reservoirs are responsible for carriage of *A. pyogenes*, *C. ulcerans*, and *C. pseudotuberculosis*. Soil is the natural reservoir for *R. equi*.

C. pseudodiphtheriticum is part of the normal flora of the human pharynx and skin. *C. xerosis* is found on the skin, nasopharynx, and conjunctiva; *C. auris* in the external auditory canal; and *C. striatum* in the anterior nares and on the skin. *C. jeikeium* and *C. urealyticum* are found in the axilla, groin, and

perineum, particularly in hospitalized patients. *C. ulcerans* and *C. pseudotuberculosis* infections have been associated with the consumption of raw milk from infected cattle.

SPECIFIC NONDIPHTHERIAL CORYNEFORMS

C. ulcerans

This organism causes a diphtherialike illness and produces both diphtheria toxin and a dermonecrotic toxin. *C. ulcerans* is a commensal in horses and cattle and has been isolated from cow's milk. The organism causes exudative pharyngitis, primarily during summer months, in rural areas, and among individuals exposed to cattle. In contrast to diphtheria, *C. ulcerans* infection is considered a zoonosis, and pigs have been identified as a source of human infection; person-to-person transmission has not been established. Nevertheless, treatment with antitoxin and antibiotics should be initiated when respiratory *C. ulcerans* is identified, and a contact investigation (including throat cultures to determine the need for antimicrobial prophylaxis and vaccination with the appropriate diphtheria toxoid –containing vaccine for unimmunized human contacts) should be conducted. The organism grows on Löffler's, Tinsdale's, and telluride media as well as blood agar. In addition to exudative pharyngitis, cutaneous disease due to *C. ulcerans* has been reported. *C. ulcerans* is susceptible to a wide panel of antibiotics. Erythromycin and macrolides appear to be the first-line agents.

C. pseudotuberculosis (ovis)

Infections caused by *C. pseudotuberculosis* are rare and are reported almost exclusively from Australia. *C. pseudotuberculosis* causes suppurative granulomatous lymphadenitis and an eosinophilic pneumonia syndrome among individuals who handle horses, cattle, goats, and deer or who drink unpasteurized milk. The organism is an important veterinary pathogen, causing suppurative lymphadenitis, abscesses, and pneumonia, but is rarely a human pathogen. Successful treatment with erythromycin or tetracycline has been reported, with surgery also performed when indicated.

C. Jeikeium (Group JK)

After a 1976 survey of diseases caused by nondiphtherial corynebacteria, CDC group JK was recognized as an important opportunistic pathogen among neutropenic patients and later emerged in HIV-infected patients as a cause of AIDS-associated opportunistic infection. Accordingly, the organism was reclassified as a separate species, *C. jeikeium*. The predominant syndrome associated with *C. jeikeium* is sepsis, which can occur in conjunction with pneumonia, endocarditis, meningitis, osteomyelitis, or epidural abscess. Risk factors for *C. jeikeium* infection include hematologicmalignancy, neutropenia from comorbid conditions, prolonged hospitalization, exposure to multiple antibiotics, and skin disruption. There is evidence that *C. jeikeium* is part of the normal flora of the inguinal, axillary, genital, and perirectal areas in hospitalized patients.

Broad-spectrum antimicrobial therapy appears to select for colonization. Originally described in the United States, *C. jeikeium* has also been reported in Europe. The gram-positive coccobacilli, which slightly resemble streptococci, grow as small, gray to white, glistening, nonhemolytic colonies on blood agar. *C. jeikeium* lacks urease and nitrate reductase and does not ferment most carbohydrates. It is resistant to most antibiotics tested except for vancomycin. Effective therapy involves removal of the source of infection, be it a catheter, a prosthetic joint, or a prosthetic valve. There have been efforts to prevent *C. jeikeium* infection by use of antibacterial soap in the care of high-risk patients in intensive care settings.

C. Urealyticum (Group D2)

Identified as a urease-positive nondiphtherial *Corynebacterium* in 1972, *C. urealyticum* is an opportunistic cause of sepsis and urinary tract infection. This organism appears to be the etiologic agent of a severe urinary tract syndrome known as *alkaline-encrusted cystitis*: a chronic inflammatory bladder infection associated with deposition of ammonium magnesium phosphate on the surface and walls of ulcerating lesions in the bladder. Obstructive uropathy due to this organism has been reported in renal transplant recipients. In addition, *C. urealyticum* has been associated with pneumonia, peritonitis, endocarditis, osteomyelitis, and wound infection. It is similar to *C. jeikeium* in its resistance to most antibiotics except vancomycin, which has been used successfully in the treatment of severe infections.

C. minutissimum

Erythrasma is a cutaneous infection producing reddish-brown, macular, scaly, pruritic intertriginous patches. The dermatologic presentation under the Wood's lamp is of coral-red fluorescence. *C. minutissimum* appears to be a common cause of erythrasma, although there is evidence for a polymicrobial etiology in certain settings. In addition, this fluorescent microbe has been associated with bacteremia in patients with hematologicmalignancy. Erythrasma responds to topical erythromycin, clarithromycin, clindamycin, or fusidic acid, although more severe infections may require oral macrolide therapy.

OTHER NONDIPHTHERIAL CORYNEBACTERIA

C. xerosis is a human commensal found in the conjunctiva, nasopharynx, and skin. This nontoxigenic organism is occasionally identified as a source of invasive infection in immunocompromised or postoperative patients and prosthetic joint recipients. *C. striatum* is found in the anterior nares and on the skin, face, and upper torso of normal individuals. Also nontoxigenic, this organism has been associated with invasive opportunistic infections in severely ill or immunocompromised patients. *C. amycolatum* is a species isolated from human skin and is identified on the basis of a unique 16S ribosomal RNA sequence associated with opportunistic infection. *C. glucuronolyticum* is a nonlipophilic species that causes male genitourinary tract infections such as prostatitis and urethritis. These infections may be successfully treated with a wide variety of antibacterial agents, including ß-lactams, rifampin, aminoglycosides, or vancomycin; however, the organism appears to be resistant to fluoroquinolones, macrolides, and tetracyclines. *C. imitans* has been identified in Eastern Europe as a nontoxigenic cause of pharyngitis. *C. auris* has been isolated from children with otitis media and is susceptible to fluoroquinolones, rifampin, tetracycline, and vancomycin but resistant to penicillin G and variably susceptible to macrolides. *C. pseudodiphtheriticum* (*C. hofmannii*) is a nontoxigenic component of the normal human flora. Human infections—particularly endocarditis of either prosthetic or native valves and invasive pneumonia—have been identified only rarely. Although *C. pseudodiphtheriticum* may be isolated from the nasopharynx of patients with suspected diphtheria, it is part of the normal flora and does not produce diphtheria toxin. *C. propinquum*, a close relative of *C. pseudodiphtheriticum*, is part of CDC group ANF-3 and is isolated from human respiratory tract specimens and blood. *C. afermentans* subspecies *lipophilum* belongs to CDC group ANF-1 and has been isolated from human blood and abscess infections. *C. accolens* has been isolated from wound drainage, throat swabs, and sputum and is typically identified as a satellite of staphylococcal organisms; it has been associated with endocarditis. *C. bovis* is a veterinary commensal that has not been clearly identified as a cause of human disease. *C. aquaticum* is a water-associated organism that is occasionally isolated from patients using medical devices (e.g., for chronic ambulatory peritoneal dialysis or venous access).

RHODOCOCCUS

Rhodococcus species are phylogenetically related to the corynebacteria. These gram-positive coccobacilli have been associated with tuberculosis-like infections in humans with granulomatous pathology. Although *R. equi* is best known, other species have been identified, including *R.* (also *Gordonia*) *bronchialis*, *R.* (also *Tsukamurella*) *aurantiacus*, *R. luteus*, *R. erythropolis*, *R. rhodochrous*, and *R. rubropertinctus*. *R. equi* has been recognized as a cause of pneumonia in horses since the 1920s; it causes related infections in cattle, sheep, and swine. *R. equi* is found in soil as an environmental microbe. The organisms vary in length; appear as spherical to long, curved, clubbed rods; and produce large, irregular mucoid colonies. *R. equi* does not ferment carbohydrates or liquefy gelatin and is often acid-fast. An intracellular pathogen of macrophages, *R. equi* can cause granulomatous necrosis and caseation. The organism has been identified most commonly in pulmonary infections, but infections of brain, bone, and skin have also been reported. Most commonly, *R. equi* disease manifests as nodular cavitary pneumonia of the upper lobe—a picture similar to that seen in tuberculosis or nocardiosis. Most patients are immunocompromised, often with HIV infection. Subcutaneous nodular lesions have also been identified. The involvement of *R. equi* should be considered in any patient presenting with a tuberculosis-like syndrome.

Infection due to *R. equi* has been treated successfully with antibiotics that penetrate intracellularly, including macrolides, clindamycin, rifampin, trimethoprim-sulfamethoxazole, tigecycline, and linezolid. ß-Lactam antibiotics have not been useful. The organism is routinely susceptible to vancomycin, which is considered the drug of choice although there may be a role for oral therapies with bactericidal agents such as linezolid.

ACTINOMYCES PYOGENES

A cause of seasonal leg ulcers in humans in rural Thailand, *A. pyogenes* is a well-known pathogen of cattle, sheep, goats, and pigs. A few human cases of sepsis, endocarditis, septic arthritis, pneumonia, meningitis, and empyema have been reported. The agent is susceptible to ß-lactams, tetracycline, aminoglycosides, and fluoroquinolones.

ARCANOBACTERIUM HAEMOLYTICUM

A. haemolyticum was identified as an agent of wound infections in U.S. soldiers in the South Pacific during World War II. This organism appears to be a commensal of the human nasopharynx and skin but has been implicated as a cause of pharyngitis and chronic skin ulcers. In contrast to the much more common pharyngitis caused by *Streptococcus pyogenes*, *A. haemolyticum* pharyngitis is associated with a scarlatiniform rash on the trunk and proximal extremities in about half of cases; this illness is occasionally confused with toxic shock syndrome. Because *A. haemolyticum* pharyngitis primarily affects teenagers, it has been postulated that the rash-pharyngitis syndrome may represent copathogenicity or synergy with EBV or opportunistic secondary infection complicating EBV infection. *A. haemolyticum* has also been reported as a cause of bacteremia, soft tissue infection, osteomyelitis, and cavitary pneumonia, predominantly in the setting of underlying diabetes mellitus. The organism is

susceptible to β-lactams, macrolides, fluoroquinolones, clindamycin, vancomycin, and doxycycline. Penicillin resistance has been reported.

FURTHER READINGS

Centers for Disease Control and Prevention: Availability of diphtheria antitoxin through an investigational new drug protocol. MMWR Morb Mortal Wkly Rep 53:413, 2004

Dittmann S et al: Successful control of epidemic diphtheria in the states of the former Union of Soviet Socialist Republics: Lessons learned. J Infect Dis 181(Suppl 1):S10, 2000

Kretsinger K et al: Preventing tetanus, diphtheria, and pertussis among adults: Use of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccine; recommendations of the Advisory Committee on Immunization Practices (ACIP) and recommendation of ACIP, supported by the Healthcare Infection Control Practices Advisory Committee (HICPAC), for use of Tdap among health-care personnel. MMWR Recomm Rep 55(RR-17):1, 2006

López-Medrano F et al: Urinary tract infection due to *Corynebacterium urealyticum* in kidney transplant recipients: An underdiagnosed etiology for obstructive uropathy and graft dysfunction—results of a prospective cohort study. Clin Infect Dis 46:825, 2008[PMID: 21698810]

MacGregor RR: *Corynebacterium diphtheriae,* in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 2687–2694

Meyer DK, Reboli AC: Other coryneform bacteria and *Rhodococcus,* in *Principles and Practice of Infectious Diseases*, 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 2695–2706

Murphy TV et al: Prevention of pertussis, tetanus, and diphtheria among pregnant and postpartum women and their infants: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 57(RR-4): 1, 2008

Romney MG et al: Emergence of an invasive clone of nontoxigenic *Corynebacterium diphtheriae* in the urban poor population of Vancouver, Canada. J Clin Microbiol 44:1625, 2006[PMID: 16672385]

Schuhegger R et al: Pigs as source for toxigenic *Corynebacterium ulcerans*. Emerg Infect Dis 15:1314, 2009[PMID: 19751602]

Wagner KS et al: A review of the international issues surrounding the availability and demand for diphtheria antitoxin for therapeutic use. Vaccine 28:14, 2009[PMID: 19818425]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LISTERIA MONOCYTOGENES INFECTIONS: INTRODUCTION

Listeria monocytogenes is a food-borne pathogen that can cause serious infections, particularly in pregnant women and immunocompromised individuals. A ubiquitous saprophytic environmental bacterium, *L. monocytogenes* is also a facultative intracellular pathogen with a broad host range. Humans are probably accidental hosts for this microorganism. *L. monocytogenes* is of interest not only to clinicians but also to basic scientists as a model intracellular pathogen that is used to study basic mechanisms of microbial pathogenesis and host immunity.

Microbiology

L. monocytogenes is a facultatively anaerobic, nonsporulating, gram-positive rod that grows over a broad temperature range, including refrigeration temperatures. This organism is motile during growth at low temperatures but much less so at 37°C. The vast majority of cases of human listerial disease can be traced to serotypes 1/2a, 1/2b, and 4. *L. monocytogenes* is weakly β-hemolytic on blood agar, and (as detailed below) its β-hemolysin is an essential determinant of its pathogenicity.

Pathogenesis

Infections with *L. monocytogenes* follow ingestion of contaminated food that contains the bacteria at high concentrations. The conversion from environmental saprophyte to pathogen involves the coordinate regulation of bacterial determinants of pathogenesis that mediate entry into cells, intracellular growth, and cell-to-cell spread. Many of the organism's pathogenic strategies can be examined experimentally in tissue culture models of infection; such a model is presented in **Fig. 139–1**. Like other enteric pathogens, *L. monocytogenes* induces its own internalization by cells that are not normally phagocytic. Its entry into cells is mediated by host surface proteins classified as internalins. Internalin-mediated entry is important in the crossing of intestinal, blood-brain, and fetoplacental barriers, although how *L. monocytogenes* traffics from the intestine to the brain or fetus is only beginning to be investigated. In a pregnant guinea pig model of infection, *L. monocytogenes* was shown to traffic from maternal organs to the placenta; surprisingly, however, it also trafficked from the placenta back to maternal organs. These data are consistent with a model in which miscarriage can be viewed as a host defense strategy to eliminate a nidus of infection.

Figure 139-1

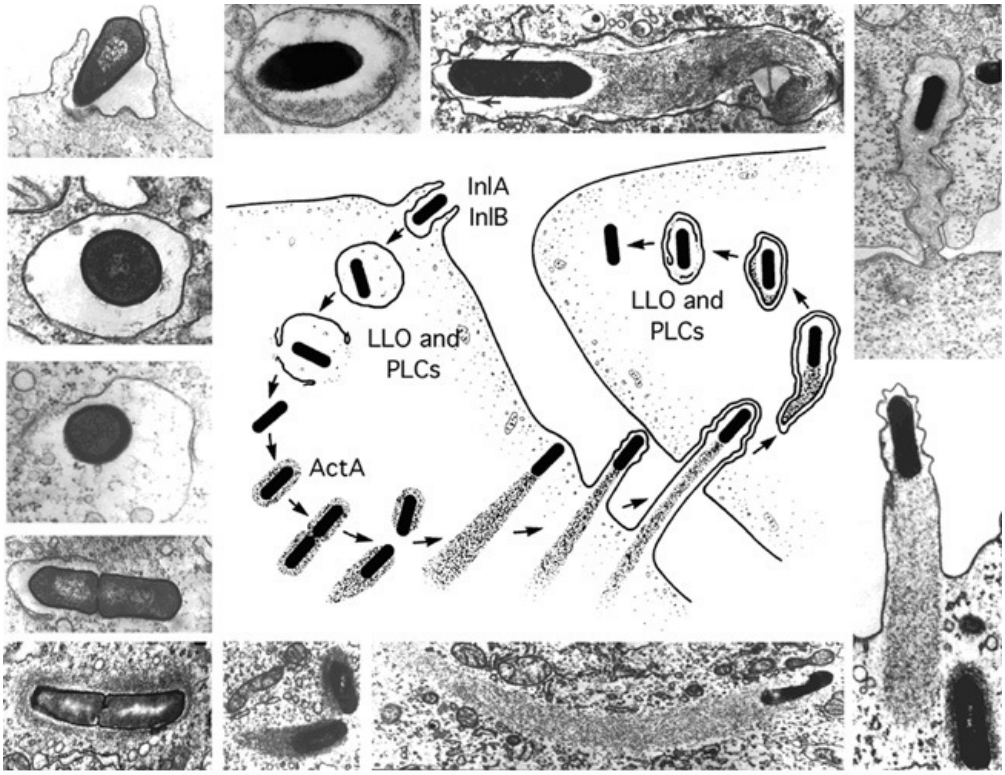

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stages in the intracellular life cycle of *Listeria monocytogenes.* The central diagram depicts cell entry, escape from a vacuole, actin nucleation, actin-based motility, and cell-to-cell spread. Surrounding the diagram are representative electron micrographs from which it was derived. ActA, surface protein mediating nucleation of host actin filaments to propel bacteria intra- and intercellularly; LLO, listeriolysin O; PLCs, phospholipases C; Inl, internalin. See text for further details. *(Adapted with permission from LG Tilney and DA Portnoy: J Cell Biol 109:1597, 1989. © Rockefeller University Press.)*

An essential determinant of the pathogenesis of *L. monocytogenes* is its β-hemolysin, listeriolysin O (LLO). LLO is a pore-forming, cholesterol-dependent cytolysin. (Related cytolysins include streptolysin O, pneumolysin, and perfringolysin O, all of which are produced by extracellular pathogens.) LLO is largely responsible for mediating the rupture of the phagosomal membrane that forms after phagocytosis of *L. monocytogenes*. LLO probably acts by inserting itself into an acidifying phagosome, thereby preventing the vesicle's maturation. In addition, LLO acts as a translocation pore for one or both of the *L. monocytogenes* phospholipases that also contribute to vacuolar lysis. LLO synthesis and activity are controlled at multiple levels to ensure that its lytic activity is limited to acidic vacuoles and does not affect the cytosol. Mutations in LLO that influence its synthesis, cytosolic half-life, or pH optimum cause premature toxicity to infected cells. There is an inverse relationship between toxicity and virulence—i.e., the more cytotoxic the strain, the less virulent it is in animals. This relationship may seem paradoxical, but, as an intracellular pathogen, *L. monocytogenes* benefits from leaving its host cell unharmed.

Shortly after exposure to the mammalian-cell cytosol, *L. monocytogenes* expresses a surface protein, ActA, that mediates the nucleation of host actin filaments to propel the bacteria intra- and intercellularly. ActA mimics host proteins of the Wiskott-Aldrich syndrome protein (WASP) family by promoting the actin nucleation properties of the Arp2/3 complex. Thus, *L. monocytogenes* can enter the cytosol of almost any eukaryotic cell or cell extract and can exploit a conserved and essential actin-based motility system. Other pathogens as diverse as certain *Shigella*, *Mycobacterium*, *Rickettsia*, and *Burkholderia* spp. use a related pathogenic strategy that allows cell-to-cell spread without exposure to the extracellular milieu.

Immune Response

The innate and acquired immune responses to *L. monocytogenes* have been studied extensively in mice. Shortly after IV injection, most bacteria are found in Kupffer cells in the liver, with some organisms in splenic dendritic cells and macrophages. Listeriae that survive the bactericidal activity of initially infected macrophages grow in the cytosol and spread from cell to cell. In the liver, the result is infection of hepatocytes. Neutrophils are crucial to host defense during the first 24 h of infection, while influx of activated macrophages from the bone marrow is critical subsequently. Mice that survive sublethal infection clear the infection within a week, with consequent sterile immunity. Studies with

knockout mice have been instrumental in dissecting the roles played by chemokines and cytokines during infection. For example, interferon γ and tumor necrosis factor (TNF) are essential in controlling infection. While innate immunity is sufficient to control infection, the acquired immune response is required for sterile immunity. Immunity is cell-mediated; antibody plays no measurable role. The critical effector cells are cytotoxic (CD8+) T cells that recognize and lyse infected cells, and the resulting extracellular bacteria are killed by circulating activated phagocytes. Animals that survive challenge with sublethal doses of bacteria become immune to subsequent infection. A hallmark of the *L. monocytogenes* model is that killed vaccines do not provide protective immunity. The explanation for this fundamental observation is multifactorial, involving the generation of appropriate cytokines and the compartmentalization of bacterial proteins for antigen processing and presentation. Because the organism has the capacity to induce a robust cell-mediated immune response, attenuated strains have been engineered to express foreign antigens and are undergoing clinical studies as therapeutic vaccines for cancer and infectious disease applications.

Epidemiology

L. monocytogenes usually enters the body via the gastrointestinal tract in foods. Listeriosis is most often sporadic, although outbreaks do occur. Recent annual incidences in the United States range from 2 to 9 cases per 1 million population. No epidemiologic or clinical evidence supports human-to-human transmission (other than vertical transmission from mother to fetus) or waterborne infection. In line with its survival and multiplication at refrigeration temperatures, *L. monocytogenes* is commonly found in processed and unprocessed foods of animal and plant origin, especially soft cheeses, delicatessen meats, hot dogs, milk, and cold salads. Because food supplies are increasingly centralized and normal hosts tolerate the organism well, outbreaks may not be immediately apparent; pulsed-field gel electrophoresis has proved useful in linking cases to specific foods. FoodNet, an active U.S. surveillance program, documented no significant change in the estimated incidence of listeriosis from 2005 through 2008. The U.S. Food and Drug Administration has a zero-tolerance policy for *L. monocytogenes* in ready-to-eat foods.

Diagnosis

Symptoms of listerial infection overlap greatly with those of other infectious diseases. Timely diagnosis requires that the illness be considered in groups at risk: pregnant women; elderly persons; neonates; individuals immunocompromised by organ transplants, cancer, or treatment with TNF antagonists or glucocorticoids; and patients with a variety of chronic medical conditions, including alcoholism, diabetes, renal disease, rheumatologic illness, and iron overload. Meningitis in older adults (especially with parenchymal brain involvement or subcortical brain abscess) should trigger consideration of *L. monocytogenes* infection. Listeriosis occasionally affects healthy, young, nonpregnant individuals. HIV-infected patients are at risk; however, listeriosis seems to be prevented by trimethoprim-sulfamethoxazole (TMP-SMX) prophylaxis targeting other AIDS-related infections. The diagnosis is typically made by culture of blood, cerebrospinal fluid (CSF), or amniotic fluid. *L. monocytogenes* may be confused with "diphtheroids" or pneumococci in gram-stained CSF or may be gram-variable and confused with *Haemophilus* spp. Serologic tests and polymerase chain reaction assays are not clinically useful diagnostic tools at present.

Clinical Manifestations

Listerial infections present as several clinical syndromes, of which meningitis and septicemia are most common. Monocytosis is seen in infected rabbits but is not a hallmark of human infection.

GASTROENTERITIS

Appreciated only since the outbreaks of the late 1980s, listerial gastroenteritis typically develops within 48 h of ingestion of a large inoculum of bacteria in contaminated foods such as milk, deli meats, and salads. Attack rates are high (50–100%). *L. monocytogenes* is neither sought nor found in routine fecal cultures, but its involvement should be considered in outbreaks when cultures for other likely pathogens are negative. Sporadic intestinal illness appears to be uncommon. Manifestations include fever, diarrhea, headache, and constitutional symptoms. The largest reported outbreak occurred in an Italian school system and included 1566 individuals; ~20% of patients were hospitalized, but only one person had a positive blood culture. Isolated gastrointestinal illness does not require antibiotic treatment. Surveillance studies show that 0.1–5% of healthy asymptomatic adults may have stool cultures positive for the organism.

BACTEREMIA

L. monocytogenes septicemia presents with fever, chills, and myalgias/arthralgias and cannot be differentiated from septicemia involving other organisms. Meningeal symptoms, focal neurologic findings, or mental status changes may suggest the diagnosis. Bacteremia is documented in 70–90% of cancer patients with listeriosis. A nonspecific flulike illness with fever is a common presentation in pregnant women. Endocarditis of prosthetic and native valves is an uncommon complication, with reported fatality rates of 35–50% in case series. A lumbar puncture is often prudent, although not necessary, in pregnant women without central nervous system (CNS) symptoms.

MENINGITIS

L. monocytogenes causes ~5–10% of all cases of community-acquired bacterial meningitis in adults in the United States. Case-fatality rates are reported to be 15–26% and do not appear to have changed over time. This diagnosis should be considered in all older or chronically ill adults with "aseptic" meningitis. The presentation is more frequently subacute (with illness developing over several days) than in meningitis of other bacterial etiologies, and nuchal rigidity and meningeal signs are less common. Photophobia is infrequent. Focal findings and seizures are common in some but not all series. The CSF profile in listerial meningitis most often shows white blood cell (WBC) counts in the range of 100–5000/μL (rarely higher); 75% of patients have WBC counts below 1000/μL, usually with a neutrophil predominance more modest than that in other bacterial meningitides. Low glucose levels and positive results on Gram's staining are found ~30–40% of the time. Hydrocephalus can occur.

MENINGOENCEPHALITIS AND FOCAL CNS INFECTION

L. monocytogenes can directly invade the brain parenchyma, producing either cerebritis or focal abscess. Approximately 10% of cases of CNS infection are macroscopic abscesses resulting from bacteremic seeding; the affected patients often have positive blood cultures. Concurrent meningitis can exist, but the CSF may appear normal. Abscesses can be misdiagnosed as metastatic or primary tumors and, in rare instances, occur in the cerebellum and the spinal cord. Invasion of the brainstem results in a characteristic severe rhombencephalitis, usually in otherwise healthy older adults. The presentation may be biphasic, with a prodrome of fever and headache followed by asymmetric cranial nerve deficits, cerebellar signs, and hemiparetic and hemisensory deficits. Respiratory failure can occur. The subacute course and the often minimally abnormal CSF findings may delay the diagnosis, which may be suggested by MRI images showing ring-enhancing lesions after gadolinium contrast and hyperintense lesions on diffusion-weighted imaging. MRI is superior to CT for the diagnosis of these infections.

INFECTION IN PREGNANT WOMEN AND NEONATES

Listeriosis in pregnancy is a severe and important infection. The usual presentation is a nonspecific acute or subacute febrile illness with myalgias, arthralgias, backache, and headache. Pregnant women with listeriosis are usually bacteremic. This syndrome should prompt blood cultures, especially in the absence of another reasonable explanation. Involvement of the CNS is rare in the absence of other risk factors. Preterm delivery is a common complication, and the diagnosis may be made only postpartum. As many as 70–90% of fetuses from infected women can become infected. Prepartum treatment of bacteremic women enhances the chances of delivery of a healthy infant. Women usually do well after delivery: maternal deaths are very rare, even when the diagnosis is made late in pregnancy or postpartum. Overall mortality rates for fetuses infected in utero approach 50% in some series; among live-born neonates treated with antibiotics, mortality rates are much lower (~20%). *Granulomatosis infantiseptica* is an overwhelming listerial fetal infection with miliary microabscesses and granulomas, most often in the skin, liver, and spleen. Less severe neonatal infection acquired in utero presents at birth. "Late-onset" neonatal illness typically develops ~10–30 days postpartum. Mothers of infants with late-onset disease are not ill .

Treatment: Infections Caused by *Listeria Monocytogenes*

No clinical trials have compared antimicrobial agents for the treatment of *L. monocytogenes* infections. Data obtained in studies conducted in vitro and in animals as well as observational clinical data indicate that ampicillin is the drug of choice, although penicillin is also highly active. Adults should receive IV ampicillin at high doses (2 g every 4 h), and many experts recommend the addition of gentamicin for synergy (1.0–1.7 mg/kg every 8 h); retrospective uncontrolled trials are not conclusive, but one study suggests that gentamicin may not help. TMP-SMX, given IV, is the best alternative for the penicillin-allergic patient (15–20 mg of TMP/kg per day in divided doses every 6–8 h). The dosages recommended cover CNS infection and bacteremia (see below for duration); dosages must be reduced for patients with renal insufficiency. One small nonrandomized study supports a combination of ampicillin and TMP-SMX. Case reports document success with vancomycin, imipenem, meropenem, linezolid, tetracycline, and macrolides, although there are also reports of clinical failure or disease development with some of these agents. Cephalosporins are *not* effective and should not be used. Neonates should receive ampicillin and gentamicin at doses based on weight.

The duration of therapy depends on the syndrome: 2 weeks for bacteremia, 3 weeks for meningitis, 6–8 weeks for brain abscess/encephalitis, and 4–6 weeks for endocarditis in both neonates and adults. Early-onset neonatal disease may be more severe and should be treated for >2 weeks.

Complications and Prognosis

Many individuals who are promptly diagnosed and treated recover fully, but permanent neurologic sequelae are common in patients with brain abscess or rhombencephalitis. Focal infections of visceral organs; the eye; the pleural, peritoneal, and pericardial spaces; and the bones and joints have all been reported. Of 100 live-born treated neonates in one series, 60% recovered fully, 24% died, and 13% had long-term neurologic or other complications.

Prevention

Healthy persons should take standard precautions to prevent food-borne illness: fully cooking meats, washing fresh vegetables, carefully cleaning utensils, and avoiding unpasteurized dairy products. In addition, persons at risk for listeriosis, including pregnant women, should avoid soft cheeses (although hard cheeses and yogurt are not problematic) and should avoid or thoroughly reheat ready-to-eat and delicatessen foods, even though the absolute risk they pose is relatively low.

FURTHER READINGS

Bakardjiev AI et al: *Listeria monocytogenes* traffics from maternal organs to the placenta and back. PLoS Pathog 2:e66, 2006

Brockstedt DG et al: Promises and challenges for the development of *Listeria monocytogenes*–based therapeutics. Expert Rev Vaccines 7:1069, 2008[PMID: 18767955]

Centers for Disease Control and Prevention: FoodNet—Foodborne Diseases Active Surveillance Network. Accessed at www.cdc.gov/foodnet/

Freitag NE et al: *Listeria monocytogenes*—from saprophyte to intracellular pathogen. Nat Rev Microbiol 9:623, 2009

Hamon M et al: *Listeria monocytogenes*: A multifaceted model. Nat Rev Microbiol 4:423, 2006[PMID: 16710323]

Mitja O et al: Predictors of mortality and impact of aminoglycosides on outcome in listeriosis in a retrospective cohort study. J Antimicrob Chemother 64:416, 2009[PMID: 19468027]

Mylonakis E et al: Listeriosis during pregnancy: A case series and review of 222 cases. Medicine (Baltimore) 81:260, 2002[PMID: 12169881]

Ooi ST, Lorber B: Gastroenteritis due to *Listeria monocytogenes*. Clin Infect Dis 40:1327, 2005[PMID: 15825036]

Pamer ER: Immune responses to *Listeria monocytogenes*. Nat Rev Immunol 4:812, 2004[PMID: 15459672]

Schnupf P, Portnoy DA: Listeriolysin: A phagosome-specific lysin. Microbes Infect 10:1176, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TETANUS: INTRODUCTION

Tetanus is an acute disease manifested by skeletal muscle spasm and autonomic nervous system disturbance. It is caused by a powerful neurotoxin produced by the bacterium *Clostridium tetani* and is completely preventable by vaccination. *C. tetani* is found throughout the world, and tetanus commonly occurs where the vaccination coverage rate is low. In developed countries, the disease is seen occasionally in individuals who are incompletely vaccinated. In any setting, established tetanus is a severe disease with a high mortality rate.

Definition

Tetanus is diagnosed on clinical grounds (sometimes with supportive laboratory confirmation of the presence of *C. tetani*; see "Diagnosis," below), and case definitions are often used to facilitate clinical and epidemiologic assessments. The Centers for Disease Control and Prevention (CDC) defines tetanus as "the acute onset of hypertonia or. . .painful muscular contractions (usually of the muscles of the jaw and neck) and generalized muscle spasms without other apparent medical cause." Neonatal tetanus is defined by the World Health Organization (WHO) as "an illness occurring in a child who has the normal ability to suck and cry in the first 2 days of life but who loses this ability between days 3 and 28 of life and becomes rigid and has spasms." Given the unique presentation of neonatal tetanus, the history generally permits accurate classification of the illness with a high degree of probability. Maternal tetanus is defined by the WHO as tetanus occurring during pregnancy or within 6 weeks after the conclusion of pregnancy (whether with birth, miscarriage, or abortion).

Etiology

C. tetani is an anaerobic, gram-positive, spore-forming rod whose spores are highly resilient and can survive readily in the environment throughout the world. Spores resist boiling and many disinfectants. In addition, *C. tetani* spores and bacilli survive in the intestinal systems of many animals, and fecal carriage is common. The spores or bacteria enter the body through abrasions, wounds, or (in the case of neonates) the umbilical stump. Once in a suitable anaerobic environment, the organisms grow, multiply, and release tetanus toxin, an exotoxin that enters the nervous system and causes disease. Very low concentrations of this highly potent toxin can result in tetanus (minimum lethal human dose, 2.5 ng/kg).

In ~20% of cases of tetanus, no puncture entry wound is found. Superficial abrasions to the limbs are the commonest infection sites in adults. Deeper infections (e.g., attributable to open fracture, abortion, or drug injection) are associated with more severe disease and worse outcomes. In neonates, infection of the umbilical stump can result from inadequate umbilical cord care; in some cultures, for example, the cord is cut with grass or animal dung is applied to the stump. Circumcision or ear-piercing can also result in neonatal tetanus.

Epidemiology

Reliable epidemiologic data on worldwide incidence are difficult to obtain, and tetanus is notoriously underreported. Studies have shown that in much of the world only 2–10% of tetanus cases are recorded. Estimates from the early 1980s suggested that tetanus caused >1 million deaths annually. As worldwide vaccination coverage has improved, the number of cases has fallen, particularly among children and neonates, who have been predominantly targeted in recent vaccination programs. In 2006, an estimated 290,000 people died of tetanus, mostly in Southeast Asia and Africa.

The elimination of maternal and neonatal tetanus is one goal of the WHO and its Expanded Programme on Immunization (EPI). An estimated 5% of maternal mortality in the 1990s was attributed to maternal tetanus. Tetanus in pregnant women and neonates is prevented by maternal immunization during pregnancy (see "Prevention," below), which is an integral component of the EPI. Although immunization coverage continues to increase, maternal and neonatal tetanus still represents an important global health burden, causing ~180,000 deaths per year.

Tetanus is a rare disease in the developed world. In 2007, a total of 28 cases were reported to the U.S. national surveillance system. Most cases occur in incompletely vaccinated or unvaccinated individuals. Persons >60 years of age are at greater risk of tetanus because antibody levels decrease over time. Injection drug users—particularly those injecting heroin subcutaneously ("skin-popping")—are increasingly recognized as a high-risk group. Between 1995 and 2000, 15–18% of U.S. tetanus infections occurred in injection drug users. In 2004, an outbreak of tetanus occurred in the United Kingdom, which had previously reported low rates among drug users. The reasons for this outbreak remain unclear but are thought to involve a combination of heroin contamination, skin-popping, and incomplete vaccination.

Pathogenesis

C. tetani produces two exotoxins: tetanolysin and tetanospasmin. Tetanolysin, which is related to the clostridial toxins and streptolysin, plays no role in the pathogenesis of the disease. Tetanospasmin, generally referred to as "tetanus toxin," is the neurotoxin that causes the manifestations of disease.

Toxin is transported by intra-axonal transport to motor nuclei of the cranial nerves or ventral horns of the spinal cord. Tetanus toxin is produced as a single 150-kDa protein that is cleaved to produce heavy (100-kDa) and light (50-kDa) chains linked by a disulfide bond and noncovalent forces. The carboxy terminal of the C chain binds to specific membrane components in presynaptic α-motor nerve terminals; evidence suggests binding to both polysialogangliosides and membrane proteins. This binding results in toxin internalization and uptake into the nerves. (Botulinum toxins enter the nervous system by a similar method but remain mostly at the neuromuscular junction and thus produce different clinical features.)

Once inside the neuron, the toxin enters a retrograde transport pathway, whereby it is transported proximally to the motor neuron body in what appears to be a highly specific process. Unlike other components of the endosomal contents, which undergo acidification following internalization, tetanus toxin is transported in a carefully regulated pH-neutral environment that prevents an acid-induced conformational change that would result in light-chain expulsion into the surrounding

cytosol.

The next stage in toxin trafficking is less clearly understood but involves tetanus toxin's escaping normal lysosomal degradation processes and undergoing translocation across the synapse to the GABA-ergic presynaptic inhibitory interneuron terminals. Here the light chain, which is a zinc-dependent endopeptidase, cleaves vesicle-associated monophosphate 2 (VAMP2, also known as synaptobrevin). This molecule is necessary for presynaptic binding and release of neurotransmitter; thus tetanus toxin prevents transmitter release and effectively blocks inhibitory interneuron discharge. The result is unregulated activity in the motor nervous system. Similar activity in the autonomic system accounts for the characteristic features of skeletal muscle spasm and autonomic system disturbance. The increased circulating catecholamine levels in severe tetanus are associated with cardiovascular complications.

Relatively little is known about the processes of recovery from tetanus. Recovery can take several weeks. Peripheral nerve sprouting is involved in recovery from botulism, and similar central nervous system sprouting may occur in tetanus. Other evidence suggests toxin degradation as a mechanism of recovery.

Approach to the Patient: Tetanus

The clinical manifestations of tetanus occur only after tetanus toxin has reached presynaptic inhibitory nerves. Once these effects become apparent, there may be little that can be done to affect disease progression. Management strategies aim to support vital functions until the effects of the toxin have worn off. Recent interest has focused on intrathecal methods of antitoxin administration to neutralize toxin within the central nervous system and limit disease progression (see "Treatment," below).

Clinical Manifestations

Tetanus produces a wide spectrum of clinical features that are broadly divided into generalized (including neonatal) and local. In the usually mild form of local tetanus, only isolated areas of the body are affected and only small areas of local muscle spasm may be apparent. If the cranial nerves are involved in localized cephalic tetanus, the phargyngeal or laryngeal muscles may spasm, with consequent aspiration or airway obstruction, and the prognosis may be poor. In the typical progression of generalized tetanus (Fig. 140-1), muscles of the face and jaw often are affected first, presumably because of the shorter distances toxin must travel up motor nerves to reach presynaptic terminals.

Figure 140-1

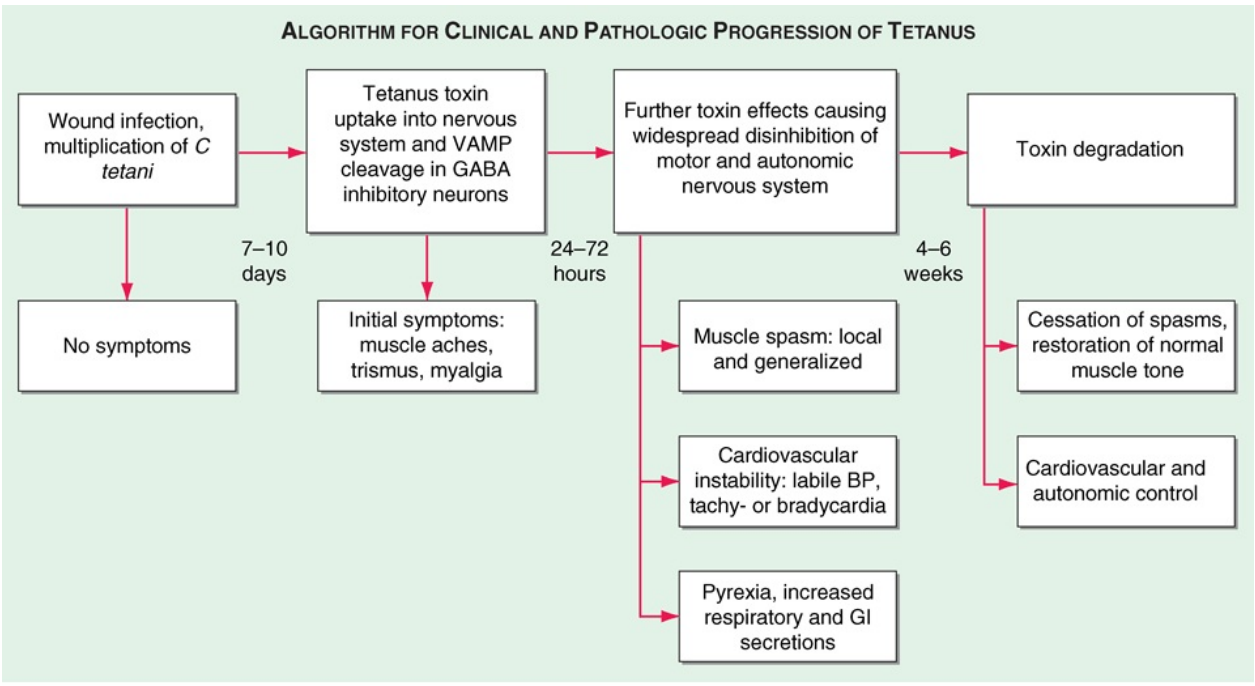

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical and pathologic progression of tetanus. BP, blood pressure; GABA, γ-aminobutyric acid; GI, gastrointestinal; VAMP, vesicle-associated monophosphate (synaptobrevin).

In assessing prognosis, the speed at which tetanus develops is important. The incubation period (time from wound to first symptom) and the period of onset (time from first symptom to first generalized spasm) are of particular significance; shorter times are associated with worse outcome. In neonatal tetanus, the younger the infant is when symptoms occur, the worse the prognosis.

The commonest initial symptoms are trismus (lockjaw), muscle pain and stiffness, back pain, and difficulty swallowing. In neonates, difficulty in feeding is the usual presentation. As the disease progresses, muscle spasm develops. Generalized muscle spasm can be very painful. Commonly, the laryngeal muscles are involved early or even in isolation. This is a life-threatening event as complete airway obstruction may ensue. Spasm of the respiratory muscles results in respiratory failure. Without ventilatory support, respiratory failure is the commonest cause of death in tetanus. Spasms strong enough to produce tendon avulsions and crush fractures have been reported, but this outcome is rare.

Autonomic disturbance is maximal during the second week of severe tetanus, and death due to cardiovascular events becomes the major risk. Blood pressure is usually labile, with rapid fluctuations from high to low accompanied by tachycardia. Episodes of bradycardia and heart block can also occur. Autonomic involvement is evidenced by gastrointestinal stasis, sweating, increased tracheal secretions, and acute (often high-output) renal failure.

Diagnosis

The diagnosis of tetanus is clinical; culture of *C. tetani* from a wound provides supportive evidence. The few conditions that mimic generalized tetanus include strychnine poisoning and dystonic reactions to antidopaminergic drugs. Abdominal muscle rigidity is characteristically continuous in tetanus but is episodic in the latter two conditions. Cephalic tetanus can be confused with other causes of trismus, such as oropharyngeal infection. Hypocalcemia and meningoencephalitis are included in the differential diagnosis of neonatal tetanus.

Treatment: Tetanus

If possible, the entry wound should be identified, cleaned, and debrided of necrotic material in order to remove anaerobic foci of infection and prevent further toxin production. Metronidazole (400 mg rectally or 500 mg IV every 6 h for 7 days) is the preferred antibiotic. An alternative is penicillin (100,000–200,000 IU/kg per day), although this drug theoretically may exacerbate spasms. Failure to remove pockets of ongoing infection may result in recurrent or prolonged tetanus.

Antitoxin should be given early in an attempt to deactivate any circulating tetanus toxin and prevent its uptake into the nervous system. Two preparations are available: human tetanus immune globulin (TIG) and equine antitoxin. TIG is the preparation of choice as it is less likely to be associated with anaphylactoid reactions. Standard therapy is 3000–6000 IU of TIG or 10,000–20,000 U of equine antitoxin as a single IM dose. However, there is evidence that intrathecal administration of TIG inhibits disease progression and leads to a better outcome. The results of a randomized controlled trial have been supported by a meta-analysis of trials involving both adults and neonates, with TIG doses of 50–1500 IU administered intrathecally.

Spasms are controlled by heavy sedation using benzodiazepines. Chlorpromazine or phenobarbital are commonly used worldwide, and IV magnesium sulfate has been used as a muscle relaxant. A significant problem with all these treatments is that the doses necessary to control spasms also cause respiratory depression; thus, in resource-limited settings without mechanical ventilators, controlling spasms while maintaining adequate ventilation is problematic, and respiratory failure is a common cause of death. In locations with ventilation equipment, severe spasms are best controlled with a combination of sedatives or magnesium and relatively short-acting, cardiovascularly inert, nondepolarizing neuromuscular blocking agents that allow titration against spasm intensity. Infusions of propofol have also been used successfully to control spasms and provide sedation.

It is important to establish a secure airway early in severe tetanus. Ideally, patients should be nursed in calm, quiet environments because light and noise can trigger spasms. Tracheal secretions are increased in tetanus, and dysphagia due to pharyngeal involvement combined with hyperactivity of laryngeal muscles makes endotracheal intubation difficult. Thus tracheostomy is the usual method of securing the airway in severe tetanus.

Cardiovascular instability in severe tetanus is notoriously difficult to treat. Rapid fluctuations in blood pressure and heart rate can occur. Cardiovascular stability is improved by increasing sedation with IV magnesium sulfate (plasma concentration, 2–4 mmol/L), morphine, or other sedatives. In addition, drugs acting specifically on the cardiovascular system (e.g., esmolol, calcium antagonists, and inotropes) may be required. Short-acting drugs that allow rapid titration are preferred; particular care should be taken when longer-acting β antagonists are administered, as their use has been associated with hypotensive cardiac arrest.

Complications arising from treatment are common and include thrombophlebitis associated with diazepam injection, ventilator-associated pneumonia, central-line infections, and septicemia. In some centers, prophylaxis against deep-vein thrombosis and thromboembolism is routine.

Recovery from tetanus may take 4–6 weeks. Patients must be given a full primary course of immunization as tetanus toxin is poorly immunogenic and the immune response following natural infection is inadequate.

Prognosis

Rapid development of tetanus is associated with more severe disease and poorer outcome; it is important to note time of onset and length of incubation period. More sophisticated modeling based on data from adults has revealed other important predictors of prognosis (Table 140–1). Few studies have formally addressed long-term outcomes of tetanus. However, it is generally accepted that recovery is typically complete unless periods of hypoventilation have been prolonged or other complications have ensued. Studies of children and neonates have suggested that neonates who have experienced prolonged periods of hypoxia may be at increased risk of learning disabilities, behavioral problems, and cerebral palsy.

Table 140–1 Factors Associated with a Poor Prognosis in Tetanus

Adult Tetanus	Neonatal Tetanus
Age >70 years Incubation period <7 days	Younger age, premature birth Incubation period <6 days
Short time from first symptom to admission Puerperal, IV, postsurgery, burn entry site	Delay in hospital admission Grass used to cut cord
Period of onset[a] <48 h	
Heart rate >140 bpm[b]	
Systolic blood pressure >140 mmHg[b]	
Severe disease or spasms[b]	
Temperature >38.5°C[b]	

[a]Time from first symptom to first generalized spasm.

[b]At hospital admission.

Prevention

Tetanus is prevented by good wound care and immunization (Chap. 122). In neonates, use of safe, clean delivery and cord-care practices as well as maternal vaccination are essential. Tetanus toxoid (TT) for vaccination is available in

various preparations: single-dose TT, TT with high- or low-dose diphtheria toxoid, or TT with diphtheria toxoid in combination with whole-cell/acellular pertussis, *Haemophilus influenzae* type b, hepatitis B, or polio vaccine.

The WHO guidelines for tetanus vaccination consist of a primary course of three doses in infancy, boosters at 4–7 and 12–15 years of age, and one booster in adulthood. In the United States, the CDC suggests an additional dose at 14–16 months and boosters every 10 years. "Catch-up" schedules recommend a three-dose primary course for unimmunized adolescents followed by two further doses. For persons who have received a complete primary course in childhood but no further boosters, two doses at least 4 weeks apart are recommended.

Standard WHO recommendations for prevention of maternal and neonatal tetanus call for administration of two doses of TT at least 4 weeks apart to the previously unimmunized pregnant woman. However, in high-risk areas, a more intensive approach has been successful, with all women of childbearing age receiving a primary course along with education on safe delivery and postnatal practices.

Individuals sustaining tetanus-prone wounds should be immunized if their vaccination status is incomplete or unknown or if their last booster was given >10 years earlier. Patients sustaining wounds not classified as clean or minor should also undergo passive immunization with TIG.

FURTHER READINGS

Advisory Committee on Immunization Practices: ACIP recommendations. MMWR Morb Mortal Wkly Rep 55:1, 2006

Attygalle D, Rodrigo N: New trends in the management of tetanus. Expert Rev Anti Infect Ther 2:73, 2004[PMID: 15482173]

Caleo M, Schiavo G: Central effects of tetanus and botulinum neurotoxins. Toxicon 54:593, 2009[PMID: 19264088]

Centers for Disease Control and Prevention: Health Information for International Travel 2010. Atlanta, CDC, 2010 (wwwnc.cdc.gov/travel/content/yellowbook/home-2010.aspx)

Galazka A et al: Tetanus, in *The Global Epidemiology of Infectious Diseases* , CJL Murray et al (eds). Geneva, World Health Organization, 2004, pp 151–199 (whqlibdoc.who.int/publications/2004/9241592303.pdf)

Kretsinger K, Srivastava P: Tetanus, in *Manual for the Surveillance of Vaccine-Preventable Diseases* , 4th ed. Atlanta, Centers for Disease Control and Prevention, 2008, Chap. 16

Miranda-Filho DDE B et al: Randomised controlled trial of tetanus treatment with antitetanus immunoglobulin by the intrathecal or intramuscular route. BMJ 328:615, 2004

Roper MH et al: Maternal and neonatal tetanus. Lancet 270:1906, 2007

Thwaites CL et al: Magnesium sulphate for treatment of severe tetanus: A randomised controlled trial. Lancet 368:1436, 2006[PMID: 17055945]

World Health Organization: WHO-recommended standards for surveillance of selected vaccine-preventable diseases. Geneva, WHO, Vaccines and Biologicals 2003 WHO/V&B/03.01 (www.who.int/immunization_monitoring/burden/routine_surveillance/en/)

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BOTULISM: INTRODUCTION

Botulinum toxin is the most toxic substance known. Botulism, a rare disease, occurs naturally as four syndromes: (1) food-borne illness due to ingestion of toxin in contaminated food; (2) wound infection due to wound colonization by toxigenic clostridia with in situ toxin production; (3) infant botulism due to colonization of the infant intestine by toxigenic clostridia with in situ toxin production; and (4) adult intestinal toxemia, a rare form of colonization with similarities to infant botulism. In addition to these recognized natural forms, botulism has been reported in association with injections of botulinum toxin for cosmetic or therapeutic purposes and after inhalation of aerosolized botulinum toxin. Botulism is caused by the toxin's inhibition of acetylcholine release at the neuromuscular junction through an enzymatic mechanism. All forms of botulism manifest as a distinct clinical syndrome of symmetric cranial nerve palsies followed by descending symmetric flaccid paralysis of voluntary muscles, which may progress to respiratory compromise and death. The mainstays of therapy are meticulous intensive care and timely treatment with antitoxin, which may limit the extent of paralysis. Rapid clinical diagnosis is critical for decisions about treatment.

Etiology and Pathogenesis

Botulinum toxin–producing clostridia are anaerobic gram-positive organisms that form subterminal spores and are ubiquitous in the environment. The hardy spores survive environmental conditions and ordinary cooking procedures. Toxin production, however, requires spore germination, which occurs only with a rare confluence of circumstances: an anaerobic atmosphere, a pH of >4.5, low salt and sugar concentrations, and temperatures of 4–120°C. Although commonly ingested, spores do not normally germinate in the intestine.

The various species of toxigenic clostridia—*C. botulinum* groups I, II, and III; *C. argentinense* (toxin type G); *C. baratii* (toxin type F); and *C. butyricum* (toxin type E)—can be differentiated on the basis of phenotypic characteristics, including specific biochemical properties and morphologic appearance on egg yolk agar. Strains of a given species can be distinguished by the antigenic specificity of the botulinum neurotoxin they produce; certain strains may produce more than one toxin serotype.

The seven identified toxin serotypes (A, B, C, D, E, F, and G) are antigenically distinct but structurally similar (~150-kDa zinc-endopeptidase proteins consisting of a 100-kDa heavy chain and a 50-kDa light chain). Whether ingested, inhaled, or produced in the intestine or a wound, botulinum neurotoxin enters the vascular system and is transported to peripheral cholinergic nerve terminals, including neuromuscular junctions, postganglionic parasympathetic nerve endings, and peripheral ganglia. The central nervous system probably is not involved. Steps in neurotoxin activity include (1) heavy-chain binding to nerve terminals, (2) internalization in endocytic vesicles, (3) translocation to cytosol, and (4) light-chain serotype-specific cleavage of one of several proteins involved in the release of the neurotransmitter acetylcholine. Inhibition of acetylcholine release by any of the seven toxin serotypes results in characteristic flaccid paralysis. Recovery follows sprouting of new nerve terminals.

Toxin serotypes A, B, E, and (rarely) F cause human disease. Toxin type A produces the most severe syndrome, with the greatest proportion of patients requiring mechanical ventilation. Toxin type B appears to cause milder disease than type A. Two cases of human illness due to toxin type C and one outbreak caused by toxin type D were reported more than 50 years ago. The reasons for the rarity of cases due to types C and D are not known; all four serotypes that affect humans produce botulism in experimental models. Toxin type E, most often associated with foods of aquatic origin, produces a syndrome of variable severity. The rare cases of illness caused by toxin type F are characterized by rapid progression to quadriplegia and respiratory failure but also by relatively rapid recovery.

Epidemiology

Person-to-person transmission of botulism has not been described.

Food-borne botulism is caused by consumption of foods contaminated with botulinum toxin. Most botulism cases are sporadic; outbreaks are typically small, involving two or three cases. The wide variation in reported rates of botulism by region and continent probably reflects differences both in true incidence and in diagnostic and reporting capacity. Worldwide, the highest incidence rate is reported from the Republic of Georgia and Armenia in the southern Caucasus region, where illness is associated with risky home-canning practices. In the United States during 1990–2000, the median number of food-borne cases per year was 23 (range, 17–43). The incidence rate was highest and the number of cases greatest in Alaska, with contaminated traditional Alaskan Native dishes implicated in

every instance. Outside Alaska, food was implicated in 75 of 101 botulinum intoxication events. Of the 68 events caused by homemade foods, 47 (69%) involved home-canned items. Of the 7 events caused by non-homemade foods, five (affecting 10 people) were caused by commercial foods, and two (affecting 25 people) were caused by restaurant-prepared foods. Severe outbreaks traced to commercially prepared foods have occurred in more recent years. In 2006, patients in the U.S. states of Georgia and Florida as well as in the Canadian province of Ontario developed severe botulism after consuming commercially produced carrot juice contaminated with a high level of toxin type A. Very high toxin levels were recorded in some of the patients' serum samples, with toxemia persisting for a record 25 days after illness onset in one patient. Outbreaks caused by contaminated commercial chili occurred in 2001 (15 cases in Texas) and 2007 (8 cases in Ohio, Indiana, and Texas).

Wound botulism is caused by contamination of wounds with *C. botulinum* spores, subsequent spore germination, and toxin production in the anaerobic milieu of an abscess. Since the early 1990s, cases in the United States have occurred almost exclusively in injection drug users. The typical patient is a 30- to 50-year-old resident of the western United States with a long history of black-tar heroin injection.

Infant botulism results from absorption of toxin producedin situ by toxigenic clostridia colonizing the intestine of children ≤1 year of age. Colonization is believed to occur because the normal bowel flora is not yet fully established; this theory is supported by studies in animals. Infant botulism is the most common form of the disease in the United States, with ~80–100 cases reported annually. Geographic "hot spots" are located in areas around Pennsylvania and California.

Adult intestinal toxemia botulism results from absorption of toxin produced in situ after rarely occurring intestinal colonization with toxigenic clostridia. Typically, patients have some anatomic or functional bowel abnormality or have recently used antibiotics that may protect normally fastidious *Clostridium* species from competing components of the bowel flora. Despite antitoxin treatment, protracted symptoms or relapse due to ongoing intraluminal production of toxin may be observed.

Iatrogenic botulism results from injection of toxin. Paralysis of variable severity has followed injection of licensed botulinum toxin products for treatment of conditions involving hypertonicity of large muscle groups. Injection of approved doses of licensed products for cosmetic purposes has not been associated with botulism. Botulism requiring ventilator support has also resulted from illegal injection of research-grade toxin.

BOTULISM AS A POTENTIAL WEAPON OF BIOTERRORISM

Botulinum toxin has been "weaponized" by governments and terrorist organizations. An attack might employ aerosolization of toxin or contamination of foods or beverages ranging in scope from small-scale tampering to contamination of a widely distributed food item. Initially, it might be difficult to differentiate a naturally occurring outbreak from an intentional attack. An unnatural event may be suggested by unusual relationships between patients and atypical exposure vehicles and toxin types. Epidemiologic features consistent with aerosol dissemination may include the victims' being in a common location (e.g., a building or public area) or being exposed to a common ventilation system, along with lack of a common food exposure. Unusual epidemiologic features of an outbreak of food-borne botulism might include implication of an unlikely commercial food product. Nevertheless, an unintentional food-borne botulism outbreak can likewise have unusual features, and an intentional event may have conventional features.

Clinical Manifestations

The distinctive clinical syndrome of botulism consists of symmetric cranial nerve palsies followed by symmetric descending flaccid paralysis that may progress to respiratory arrest and death. In *food-borne botulism*, the incubation period from ingestion of food containing botulinum toxin to onset of symptoms is usually 18–36 h but, depending on the toxin dose, can range from a few hours to several days. The extent of paralysis (from a few cranial nerves only to quadriplegia) also depends on the toxin dose. The illness ranges from a mild condition for which no medical advice is sought to severe disease that can result in death within 24 h. In *wound botulism* in injection drug users, the incubation period is difficult to establish because most patients inject drug several times daily. The clinical syndrome is indistinguishable from that of food-borne botulism except that gastrointestinal symptoms are typically absent. Often the abscess is a minor lesion, a furuncle, or cellulitic in appearance. The clinical presentation of *infant botulism* resembles that of adult forms of botulism, including inability to suck and swallow, weakened voice, ptosis, and floppy neck, sometimes with progression to generalized flaccidity and respiratory compromise.

Cranial nerve involvement, which almost always marks the onset of symptoms of botulism, usually produces diplopia, dysarthria, dysphonia, and/or dysphagia. Cranial nerve palsies are the presenting manifestations that typically cause patients to seek medical care; their absence or their onset after the appearance of other true neurologic symptoms makes botulism highly unlikely. Weakness progresses, often rapidly, from the head to involve the neck, arms, thorax, and legs; occasionally, weakness is asymmetric.

Cranial nerve palsies are characteristically followed by flaccid, descending, completely symmetric paralysis of voluntary muscles. Paresthesias have been reported and may represent secondary nerve compression from immobility due to paralysis. Paralysis of the diaphragm and accessory breathing muscles may result in respiratory compromise or arrest and death. Pharyngeal collapse secondary to cranial nerve paralysis may compromise the airway and require intubation in the absence of respiratory muscle compromise. Autonomic symptoms may include anhidrosis, with pronounced mucosal erythema and pain mimicking pharyngitis, and postural hypotension. In food-borne botulism, nausea, vomiting, and abdominal pain may precede or follow the onset of paralysis. Constipation due to paralytic ileus is nearly universal, and urinary retention is also common. Extraocular muscle paralysis manifests as blurred vision or diplopia and an inability to accommodate near vision. Ptosis and facial paralysis are frequent; the pupillary reflexes may be depressed, and fixed or dilated pupils are noted in half of patients. Dizziness, dry mouth, and very dry, occasionally sore throat are common. Vital signs are usually normal, but in some cases hypotension occurs. Patients are usually afebrile. The gag reflex may be suppressed, and deep tendon reflexes may be normal or may progressively disappear.

Patients usually exhibit no sensory or cognitive deficits. They are generally alert and oriented, but they may be drowsy, agitated, and anxious. Even when intubated, patients can respond to questions by moving their fingers or toes unless paralysis has affected the digits. Unfortunately, in some instances the severe ptosis, expressionless facies, and weak phonation of patients with botulism have been interpreted as signs of mental status changes from alcohol intoxication, drug overdose, encephalitis, or meningitis; in such cases, the consequences can be delayed diagnosis, prolonged paralysis, and complications. Because of skeletal muscle paralysis, patients experiencing respiratory distress may appear placid and detached even as they near respiratory arrest. Death in untreated botulism is usually due to airway obstruction from pharyngeal muscle paralysis and inadequate tidal volume resulting from paralysis of diaphragmatic and accessory respiratory muscles. Death can also result from nosocomial infections and other sequelae of long-term paralysis, hospitalization, and mechanical ventilatory support.

Toxin binding is irreversible, but nerve terminals do regenerate. In the United States, 95% of patients recover fully, but this process may take many months and often requires extended outpatient rehabilitation therapy.

Diagnosis

DIFFERENTIAL DIAGNOSIS

In the setting of an outbreak with multiple cases, the diagnosis readily suggests itself; that is, a cluster of two or more cases with compatible symptoms is essentially pathognomonic since other illnesses that resemble botulism do not cause outbreaks. In lone (sporadic) cases, however, the diagnosis is often missed. The differential diagnosis includes Guillain-Barré syndrome (GBS), myasthenia gravis, stroke syndromes, Eaton-Lambert syndrome, and tick paralysis. Less likely are poisoning by tetrodotoxin, shellfish, or a host of rarer agents and antimicrobial drug–associated paralysis. A thorough history and meticulous physical examination can effectively eliminate most alternative diagnoses. Brain imaging may help rule out rare nonlateralizing stroke syndromes.

GBS, a rare autoimmune demyelinating polyneuropathy that often follows an acute infection, presents most often as an ascending paralysis and never causes outbreaks. Occasional GBS cases present as the Miller Fisher variant, whose characteristic triad of ophthalmoplegia, ataxia, and areflexia is easily mistaken for the early descending paralysis of botulism. Protein levels in cerebrospinal fluid (CSF) are elevated in all forms of GBS; because this increase may be delayed until several days after symptom onset, an early lumbar puncture with a negative result may need to be repeated. In contrast, CSF findings are generally normal in botulism, although marginally elevated CSF protein concentrations have been reported in some patients with wound botulism. In experienced hands, electromyography may demonstrate findings consistent with GBS but not with botulism.

The edrophonium (Tensilon) test is sometimes of value in distinguishing botulism (usually a negative result; sometimes borderline positive) from myasthenia gravis (a positive result). A strongly positive edrophonium test, in either the presence or the absence of anti–acetylcholine receptor autoantibodies, confirms the diagnosis of myasthenia gravis.

In most cerebrovascular accidents, physical examination reveals asymmetry of paralysis and upper motor neuron signs. Brain imaging can reveal the rare basilar stroke that produces symmetric bulbar palsies. Eaton-Lambert syndrome usually manifests as proximal limb weakness in a patient already debilitated by cancer. Tick paralysis is a rare flaccid condition closely resembling botulism and caused by neurotoxins of certain ticks.

HISTORY AND LABORATORY CONFIRMATION

In suspected food-borne botulism, a 3- to 5-day food history should be obtained, with specific questions about home-canned, exotic, and unusual foods. A history of recent consumption of home-canned food substantially enhances the probability of food-borne botulism. The names of contacts who may have shared foods should be obtained early in case the patient's illness progresses to

respiratory failure. In wound botulism, material from abscesses should be collected in anaerobic culture tubes for testing at public health laboratories, and serum samples should be collected. Standard blood work and radiologic studies are not useful in diagnosing botulism.

Botulism in a symptomatic patient can be confirmed in the laboratory by demonstration of toxin in clinical specimens (serum, stool, sterile water or saline enema, gastric aspirates, wound material) or in samples of ingested foods. Isolation of toxigenic clostridia from stool also constitutes evidence of botulism; the organism is rarely isolated from the stools of asymptomatic persons. Wound cultures yielding the organism are highly suggestive in symptomatic cases. In the United States, testing is conducted only in public health laboratories staffed by experienced personnel. The universally accepted method for confirmation of botulism is the mouse bioassay. Neutralization of the paralytic effects in mice by a specific antitoxin provides evidence of that toxin serotype in the clinical sample. Mouse bioassay results may not be available for up to 48 h; accordingly, all decisions about clinical management, including the administration of botulinum antitoxin, must be based on the clinical presentation, with bioassay results serving as confirmation. The sensitivity of the bioassay varies inversely with the time elapsed between symptom onset and sample collection. A test may be negative even when a patient has botulism; however, if the clinical presentation is questionable and test results are negative, additional tests may be necessary to rule out other conditions. New tests for botulism are being developed but remain experimental. At this time, no alternative to the mouse bioassay exists for laboratory confirmation of botulism. In affected muscles, findings consistent with botulism include reduced amplitude of motor potentials and potentiation with rapid repetitive stimulation.

The earliest available serum sample (e.g., that obtained at admission) should be preserved for testing. Vomit or nasogastric tube secretions should be collected immediately. A stool sample should be collected by means of a sterile water enema if the patient is constipated or otherwise unable to produce a sample. All samples, including suspect foods, should be kept refrigerated—not frozen—pending shipment directions from public health officials. In general, ingested toxin is not demonstrable in serum later than 1 week after exposure, although detection 25 days after ingestion has been documented. Toxin and toxigenic clostridia can be detected in stool later in the course of illness, and the toxin is stable indefinitely in many food matrices. Diagnosis of adult intestinal toxemia botulism requires the demonstration of protracted excretion of organisms and toxin in the stool.

Treatment: Botulism

The cornerstones of treatment for botulism are meticulous intensive care and immediate administration of botulinum antitoxin. Persons of all ages (including infants) in whom botulism is suspected should be hospitalized immediately in an intensive care setting, with frequent monitoring of vital capacity and mechanical ventilation if required. Paralysis may last for weeks or months, and meticulous intensive care is required throughout this period of debilitation. The decision to administer botulinum antitoxin—the only specific treatment—must be based on a clinical diagnosis and cannot be postponed while laboratory confirmation is awaited. Botulinum antitoxin neutralizes only toxin molecules that have not yet bound to nerve endings; it cannot reverse existing paralysis. Thus antitoxin should be given early in the course of illness, ideally <24 h after symptom onset. Infant botulism is treated with a licensed human-origin antitoxin; although the survival rate in infant botulism is nearly 100% with or without antitoxin therapy, this treatment halves the median hospitalization period (from 6 to 3 weeks). Other forms of botulism are treated with equine-source antitoxin. Treatment with the licensed, non-despeciated, equine-origin antitoxins long used in the United States is associated with anaphylaxis, other hypersensitivity reactions, and serum sickness. A new heptavalent despeciated equine antitoxin has replaced the previously used non-despeciated equine antitoxin in this country.

In wound botulism, suspect wounds and abscesses should be cleaned, debrided, and drained promptly. *C. botulinum* is susceptible to penicillins and various other antimicrobial agents. The effectiveness of antimicrobial therapy for wound botulism has not been established, and such treatment should be guided by clinical judgment.

Although no case of person-to-person transmission of botulism has been reported and although levels of toxin in bodily fluids (including serum and stool) are insufficient to cause serious illness, absorption of botulinum toxin through mucosal surfaces, the eye, or nonintact skin sometimes causes local paralysis. Standard precautions should be exercised when evaluating and treating patients. As a precaution, persons exposed to bodily fluids or stool of botulism patients should be advised of the early signs of botulism and instructed to report for evaluation if these signs are noted. Standard hospital decontamination procedures should be used to clean medical equipment and supplies that have come into contact with a botulism patient.

Notification, Expert Consultation, and Antitoxin Provision

Every botulism case is a public health emergency. The clinician must report a suspected case on an emergency basis to the state health department, which will initiate an epidemiologic investigation and put the physician in contact with the Centers for Disease control and prevention (CDC) 24-h botulism consultancy service (Emergency Operations Center: 770-488-7100) or a locally available service. The CDC consultant will (1) review the case by telephone; (2) help arrange laboratory confirmation at appropriate testing facilities; and (3) arrange emergency shipment of antitoxin (adult cases only), which

in the United States is available exclusively from the CDC. In addition, the Infant Botulism Treatment and Prevention Program of the California Department of Public Health (510-231-7600) provides 24-h consultation and distributes antitoxin (licensed BabyBIG®) for the treatment of infant botulism cases. Except in cases involving infants who reside in California, laboratory testing requests must still be authorized by the state health department where the infant is located or by the CDC.

Prevention

No prophylaxis or licensed vaccine for botulism is available. An experimental vaccine for administration to laboratory workers is available from the CDC. Persons exposed to botulinum toxin should be evaluated by a physician and carefully observed for the development of symptoms. If symptoms do appear, the patient should be treated immediately with botulinum antitoxin as described above. If antitoxin supplies are limited, treatment will most likely benefit patients with symptoms that are progressive but have not yet progressed to respiratory failure.

FURTHER READINGS

Arnon SS et al: Botulism toxin as a biological weapon: Medical and public health management. JAMA 285:1059, 2001[PMID: 11209178]

——— et al: Human botulism immune globulin for the treatment of infant botulism. N Engl J Med 354:462, 2006

Centers for Disease Control and Prevention: Botulism in the United States, 1899–1996: Handbook for epidemiologists, clinicians and laboratory workers. Atlanta, CDC, 1998

Cherington M: Electrophysiologic methods as an aid in diagnosis of botulism: A review. Muscle Nerve 5:S28, 1982

Chertow DS et al: Botulism in four adults following cosmetic injections with an unlicensed, highly concentrated botulinum preparation. JAMA 296:2476, 2006[PMID: 17119144]

Gottlieb SL et al: Long-term outcomes of 217 botulism cases in the Republic of Georgia. Clin Infect Dis 45:174, 2007[PMID: 17578775]

Gupta A et al: Adult botulism type F in the United States. Neurology 65:1694, 2005[PMID: 16344510]

Passaro DJ et al: Wound botulism associated with black tar heroin among injecting drug users. JAMA 279:859, 1998[PMID: 9516001]

Simpson LL: Identification of the major steps in botulinum toxin action. Annu Rev Pharmacol Toxicol 44:167, 2004[PMID: 14744243]

Sobel J: Diagnosis and treatment of botulism: A century later, clinical suspicion remains the cornerstone. Clin Infect Dis 48:1674, 2009

Yu PA et al: Botulism, in Bacterial Infections of Humans, PS Brachman, E Abrutyn (eds). New York, Springer Science, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GAS GANGRENE AND OTHER CLOSTRIDIAL INFECTIONS: INTRODUCTION

The genus *Clostridium* encompasses more than 60 species that may be commensals of the gut microflora or may cause a variety of infections in humans and animals through the production of a plethora of proteinaceous exotoxins. *C. tetani* and *C. botulinum*, for example, cause specific clinical disease by elaborating single but highly potent toxins. In contrast, *C. perfringens* and *C. septicum* cause aggressive necrotizing infections that are attributable to multiple toxins, including bacterial proteases, phospholipases, and cytotoxins.

ETIOLOGIC AGENT

Vegetative cells of *Clostridium* species are pleomorphic, rod-shaped, and arranged singly or in short chains (**Fig. 142-1**); the cells have rounded or sometimes pointed ends. Although clostridia stain gram-positive in the early stages of growth, they may appear to be gram-negative or gram-variable later in the growth cycle or in infected tissue specimens. Most strains are motile by means of peritrichous flagella; *C. septicum* swarms on solid media. Nonmotile species include *C. perfringens*, *C. ramosum*, and *C. innocuum*. Most species are obligately anaerobic, although clostridial tolerance to oxygen varies widely; some species (e.g., *C. septicum*, *C. tertium*) will grow but will not sporulate in air.

Figure 142-1

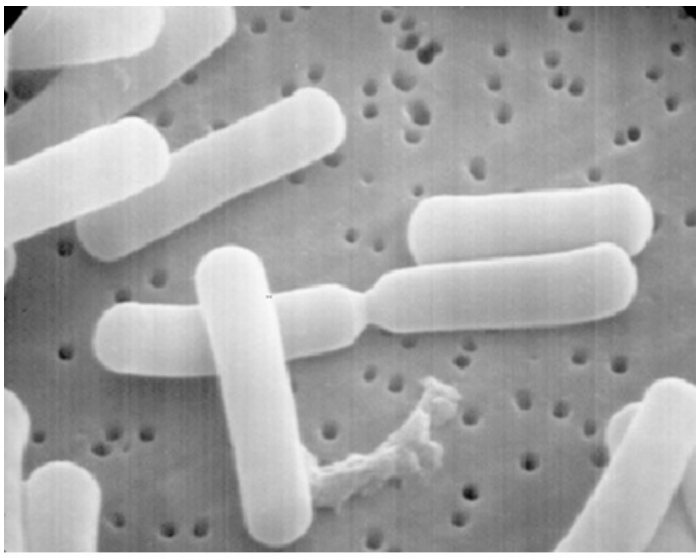

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scanning electron micrograph of *C. perfringens*.

Clostridia produce more protein toxins than any other bacterial genus, and more than 25 clostridial toxins lethal to mice have been identified. These proteins include neurotoxins, enterotoxins, cytotoxins, collagenases, permeases, necrotizing toxins, lipases, lecithinases, hemolysins, proteinases, hyaluronidases, DNases, ADP-ribosyltransferases, and neuraminidases. Botulinum andtetanus neurotoxins are the most potent toxins known, with lethal doses of 0.2–10 ng/kg for humans. Epsilon toxin, a 33-kDa protein produced by *C. perfringens* types B and D, causes edema and hemorrhage in the brain, heart, spinal cord, and kidneys of animals. It is among the most lethal of the clostridial toxins and is considered a potential agent of bioterrorism (Chap. 221). The genomic sequences of some pathogenic clostridia are now available and are likely to facilitate a comprehensive approach to understanding the virulence factors involved in clostridial pathogenesis.

EPIDEMIOLOGY AND TRANSMISSION

Clostridium species are widespread in nature, forming endospores that are commonly found in soil,

feces, sewage, and marine sediments. The ecology of *C. perfringens* in soil is greatly influenced by the degree and duration of animal husbandry in a given location and is relevant to the incidence of gas gangrene caused by contamination of war wounds with soil. For example, the incidence of clostridial gas gangrene is higher in agricultural regions of Europe than in the Sahara Desert of Africa. Similarly, the incidences of tetanus and food-borne botulism are clearly related to the presence of clostridial spores in soil, water, and many foods. Clostridia are present in large numbers in the indigenous microbiota of the intestinal tract of humans and animals, in the female genital tract, and on the oral mucosa. It should be noted that not all commensal clostridia are toxigenic.

Clostridial infections remain a serious public health concern worldwide. In developing nations, food poisoning, necrotizing enterocolitis, and gas gangrene are common because large portions of the population are poor and have little or no immediate access to health care. These infections remain prevalent in developed countries as well. Gas gangrene commonly follows knife or gunshot wounds or vehicular accidents or develops as a complication of surgery or gastrointestinal carcinoma. Severe clostridial infections have emerged as a health threat to injection drug users and to women undergoing childbirth or abortion. Historically, clostridial gas gangrene has been the scourge of the battlefield. The global political situation portends another possible scenario involving mass casualties of war or terrorism, with extensive injuries conducive to gas gangrene. Thus there is an ongoing need to develop novel strategies to prevent or attenuate the course of clostridial infections in both civilians and military personnel. Vaccination against exotoxins important in pathogenesis would be of great benefit in developing nations and could also be used safely in at-risk populations such as the elderly, patients with diabetes who may require lower-limb surgery due to trauma or poor circulation, and those undergoing intestinal surgery. Moreover, a hyperimmune globulin would be a valuable tool for prophylaxis in victims of acute traumatic injury or for attenuation of the spread of infection in patients with established gas gangrene.

CLINICAL SYNDROMES

Life-threatening clostridial infections range from intoxications (e.g., food poisoning, tetanus) to necrotizing enteritis/colitis, bacteremia, myonecrosis, and toxic shock syndrome (TSS). Tetanus and botulism are discussed in Chaps. 140 and 141, respectively. Colitis due to *C. difficile* is discussed in Chap. 129.

Clostridial Wound Contamination

Of open traumatic wounds, 30–80% are reportedly contaminated with clostridial species. In the absence of devitalized tissue, the presence of clostridia does not necessarily lead to infection. In traumatic injuries, clostridia are isolated with equal frequency from both suppurative and well-healing wounds. Thus, diagnosis and treatment of clostridial infection should be based on clinical signs and symptoms and not solely on bacteriologic findings.

Polymicrobial Infections Involving Clostridia

Clostridial species may be found in polymicrobial infections also involving microbial components of the indigenous flora. In these infections, clostridia often appear in association with non-spore-forming anaerobes and facultative or aerobic organisms. Head and neck infections, conjunctivitis, brain abscess, sinusitis, otitis, aspiration pneumonia, lung abscess, pleural empyema, cholecystitis, septic arthritis, and bone infections all may involve clostridia. These conditions are often associated with severe local inflammation but may lack the characteristic systemic signs of toxicity and rapid progression seen in other clostridial infections. In addition, clostridia are isolated from ~66% of intraabdominal infections in which the mucosal integrity of the bowel or respiratory system has been compromised. In this setting, *C. ramosum*, *C. perfringens*, and *C. bifermentans* are the most commonly isolated species. Their presence does not invariably lead to a poor outcome. Clostridia have been isolated from suppurative infections of the female genital tract (e.g., ovarian or pelvic abscess) and from diseased gallbladders. Although the most frequently isolated species is *C. perfringens*, gangrene is not typically observed; however, gas formation in the biliary system can lead to emphysematous cholecystitis, especially in diabetic patients. *C. perfringens* in association with mixed aerobic and anaerobic microbes can cause aggressive life-threatening type I necrotizing fasciitis or Fournier's gangrene.

The treatment of mixed aerobic/anaerobic infection of the abdomen, perineum, or gynecologic organs should be based on Gram's staining, culture, and antibiotic sensitivity information. Reasonable empirical treatment consists of ampicillin or ampicillin/sulbactam combined with either clindamycin or metronidazole (**Table 142-1**). Broader gram-negative coverage may be necessary if the patient has recently been hospitalized or treated with antibiotics. Such coverage can be obtained by substituting ticarcillin/clavulanic acid, piperacillin/sulbactam, or a penem antibiotic for ampicillin or by adding a fluoroquinolone or an aminoglycoside to the regimen.

Table 142-1 Treatment of Clostridial Infections

Condition	Antibiotic	Penicillin	Adiunctive Treatment/Note

	Treatment	Allergy	
Wound contamination	None	—	—
Polymicrobial anaerobic infections involving clostridia (e.g., abdominal wall, gynecologic)	Ampicillin (2 g IV q4h) *plus* Clindamycin (600–900 mg IV q6–8h) *plus* Ciprofloxacin (400 mg IV q6–8 h)	Vancomycin (1 g IV q12h) *plus* Metronidazole (500 mg IV q6h) *plus* Ciprofloxacin (400 mg IV q6–8h)	Empirical therapy should be initiated. Therapy should be based on Gram's stain and culture results and on sensitivity data when available. Add gram-negative coverage if indicated (see text).
Clostridial sepsis	Penicillin, 3–4 mU IV q4–6h *plus* Clindamycin (600–900 mg IV q6–8h)	Clindamycin alone *or* Metronidazole *or* Vancomycin as for polymicrobial anaerobic infections (see above)	Transient bacteremia without signs of systemic toxicity may be clinically insignificant.
Gas gangrene	Penicillin G (4 mU IV q4–6 h) *plus* Clindamycin (600–900 mg IV q6–8h)	Cefoxitin (2 g IV q6h) *plus* Clindamycin (600–900 mg IV q6–8h)	Emergent surgical exploration and thorough debridement are extremely important. Hyperbaric oxygen therapy may be considered after surgery and antibiotics have been initiated.

Enteric Clostridial Infections

C. perfringens type A is one of the most common bacterial causes of food-borne illness in the United States and Canada. The foods typically implicated include improperly cooked meat and meat products (e.g., gravy) in which residual spores germinate and proliferate during slow cooling or insufficient reheating. Illness results from the ingestion of food containing at least ~10^8 viable vegetative cells, which sporulate in the alkaline environment of the small intestine, producing *C. perfringens* enterotoxin in the process. The diarrhea that develops within 7–30 h of ingestion of contaminated food is generally mild and self-limiting; however, in the very young, the elderly, and the immunocompromised, symptoms are more severe and occasionally fatal. Enterotoxin-producing *C. perfringens* has been implicated as an etiologic agent of persistent diarrhea in elderly patients in nursing homes and tertiary-care institutions and has been considered to play a role in antibiotic-associated diarrhea without pseudomembranous colitis.

C. perfringens strains associated with food poisoning possess the gene (*cpe*) coding for enterotoxin, which acts by forming pores in host cell membranes. *C. perfringens* strains isolated from non-food-borne diseases, such as antibiotic-associated and sporadic diarrhea, carry *cpe* on a plasmid that may be transmitted to other strains. Several methods have been described for the detection *of C. perfringens* enterotoxin in feces, including cell culture assay (Vero cells), enzyme-linked immunosorbent assay, reversed-phase latex agglutination, and polymerase chain reaction (PCR) amplification of *cpe*. Each method has its advantages and limitations.

Enteritis necroticans (gas gangrene of the bowel) is a fulminating clinical illness characterized by extensive necrosis of the intestinal mucosa and wall. Cases can occur sporadically in adults or as epidemics in people of all ages. Enteritis necroticans is caused by α toxin– and β toxin–producing strains of *C. perfringens* type C; β toxin is located on a plasmid and is mainly responsible for pathogenesis. This life-threatening infection causes ischemic necrosis of the jejunum. In Papua New Guinea during the 1960s, enteritis necroticans (known in that locale as *pigbel*) was found to be the most common cause of death in children; it has been associated with pig feasts and occurs both sporadically and in outbreaks. Intramuscular immunization against the β toxin resulted in a decreased

incidence of the disease in Papua New Guinea. Enteritis necroticans has also been recognized in the United States, the United Kingdom, Germany ("*darmbrand*"), and other developed nations; especially affected are adults who are malnourished or who have diabetes, alcoholic liver disease, or neutropenia.

Necrotizing enterocolitis, a disease resembling enteritis necroticans but associated with *C. perfringens* type A, has been found in North America in previously healthy adults. It is also a serious gastrointestinal disease of low-birth-weight (premature) infants hospitalized in neonatal intensive care units. The etiology and pathogenesis of this disease have remained enigmatic for more than four decades. Pathologic similarities between necrotizing enterocolitis and enteritis necroticans include the pattern of small-bowel necrosis involving the submucosa, mucosa, and muscularis; the presence of gas dissecting the tissue planes; and the degree of inflammation. Enteritis necroticans most commonly involves the jejunum, whereas necrotizing enterocolitis affects the ileum and frequently the ileocecal valve. Both diseases may manifest as intestinal gas cysts, although this feature is more common in necrotizing enterocolitis. The sources of the gas, which contains hydrogen, methane, and carbon dioxide, are probably the fermentative activities of intestinal bacteria, including clostridia. Epidemiologic data support an important role for *C. perfringens* or other gas-producing microorganisms (e.g., *C. neonatale*, certain other clostridia, or *Klebsiella* species) in the pathogenesis of necrotizing enterocolitis.

Patients with suspected clostridial enteric infection should undergo nasogastric suction and receive IV fluids. Pyrantel is given by mouth, and the bowel is rested by fasting. Benzylpenicillin (1 mU) is given intravenously every 4 h, and the patient is observed for complications requiring surgery. Patients with mild cases recover without surgical intervention. If surgical indications are present (gas in the peritoneal cavity, absent bowel sounds, rebound tenderness, abdominal rigidity), however, the mortality rate ranges from 35% to 100%; fatal outcome is due in part to perforation of the intestine. As pigbel continues to be a common disease in Papua New Guinea, consideration should be given to the use of a *C. perfringens* type C toxoid vaccine in local areas. Two doses given 3–4 months apart are preventive.

Clostridial Bacteremia

Clostridium species are important causes of bloodstream infections. Molecular epidemiologic studies of anaerobic bacteremia have identified *C. perfringens* and *C. tertium* as the two most frequently isolated species; these organisms cause up to 79% and 5%, respectively, of clostridial bacteremias. Occasionally, *C. perfringens* bacteremia occurs in the absence of an identifiable infection at another site. When associated with myonecrosis, bacteremia has a grave prognosis.

C. septicum is also commonly associated with bacteremia. This species is isolated only rarely from the feces of healthy individuals but may be found in the normal appendix. More than 50% of patients whose blood cultures are positive for this organism have some gastrointestinal anomaly (e.g., diverticular disease) or underlying malignancy (e.g., carcinoma of the colon). In addition, a clinically important association of *C. septicum* bacteremia with neutropenia of any origin—and, more specifically, with neutropenic enterocolitis involving the terminal ileum or cecum—has been observed. Patients with diabetes mellitus, severe atherosclerotic cardiovascular disease, or anaerobic myonecrosis (gas gangrene) may also develop *C. septicum* bacteremia. *C. septicum* has been recovered from the bloodstream of cirrhotic patients, as have *C. perfringens*, *C. bifermentans*, and other clostridia. Infections of the bloodstream by *C. sordellii* and *C. perfringens* have been associated with TSS.

Bloodstream infection by *C. tertium*, either alone or in combination with *C. septicum* or *C. perfringens*, can be found in patients with serious underlying disease such as malignancy or acute pancreatitis, with or without neutropenic enterocolitis; the frequency has not been systematically studied. *C. tertium* may present special problems in terms of both identification and treatment. This organism may stain gram-negative; is aerotolerant; and is resistant to metronidazole, clindamycin, and cephalosporins.

Other clostridia from the *C. clostridioforme* group (including *C. clostridioforme*, *C. hathewayi*, and *C. bolteae*) can cause bacteremia.

The clinical importance of recognizing clostridial bacteremia—especially that due to *C. septicum*—and starting appropriate treatment immediately cannot be overemphasized. Patients with this condition usually are gravely ill, and infection may metastasize to distant anatomic sites, resulting in spontaneous myonecrosis (see next section). Alternative methods to identify these strains, such as PCR or other rapid diagnostic tests, are not currently available. Anaerobic blood cultures and Gram's stain interpretation remain the best diagnostic tests at this point.

Clostridial Skin and Soft-Tissue Infections

Histotoxic clostridial species such as *C. perfringens*, *C. histolyticum*, *C. septicum*, C. *novyi*, and *C. sordellii* cause aggressive necrotizing infections of the skin and soft tissues. These infections are attributable in part to the elaboration of bacterial proteases, phospholipases, and cytotoxins. Necrotizing clostridial soft-tissue infections are rapidly progressive and are characterized by marked

tissue destruction, gas in the tissues, and shock; they frequently end in death. Severe pain, crepitus, brawny induration with rapid progression to skin sloughing, violaceous bullae, and marked tachycardia are characteristics found in the majority of patients.

CLOSTRIDIAL MYONECROSIS (GAS GANGRENE)

Traumatic Gas Gangrene

C. perfringens myonecrosis (gas gangrene) is one of the most fulminant gram-positive bacterial infections of humans. Even with appropriate antibiotic therapy and management in an intensive care unit, tissue destruction can progress rapidly. Gas gangrene is accompanied by bacteremia, hypotension, and multiorgan failure and is invariably fatal if untreated. Gas gangrene is a true emergency and requires immediate surgical debridement.

The development of gas gangrene requires an anaerobic environment and contamination of a wound with spores or vegetative organisms. Devitalized tissue, foreign bodies, and ischemia reduce locally available oxygen levels and favor outgrowth of vegetative cells and spores. Thus conditions predisposing to traumatic gas gangrene include crush-type injury, laceration of large or medium-sized arteries, and open fractures of long bones that are contaminated with soil or bits of clothing containing the bacterial spores. Gas gangrene of the abdominal wall and flanks follows penetrating injuries such as knife or gunshot wounds that are sufficient to compromise intestinal integrity, with resultant leakage of the bowel contents into the soft tissues. Proximity to fecal sources of bacteria is a risk factor for cases following hip surgery, adrenaline injections into the buttocks, or amputation of the leg for ischemic vascular disease. In the last decade, cutaneous gas gangrene caused by *C. perfringens*, *C. novyi*, and *C. sordellii* has been described in the United States and northern Europe among persons injecting black-tar heroin subcutaneously.

The incubation period for traumatic gas gangrene can be as short as 6 h and is usually <4 days. The infection is characterized by the sudden onset of excruciating pain at the affected site and the rapid development of a foul-smelling wound containing a thin serosanguineous discharge and gas bubbles. Brawny edema and induration develop and give way to cutaneous blisters containing bluish to maroon-colored fluid. Such tissue later may become liquefied and slough. The margin between healthy and necrotic tissue often advances several inches per hour despite appropriate antibiotic therapy, and radical amputation remains the single best life-saving intervention. Shock and organ failure frequently accompany gas gangrene; when patients become bacteremic, the mortality rate exceeds 50%.

Diagnosis of traumatic gas gangrene is not difficult because the infection always begins at the site of significant trauma, is associated with gas in the tissue, and is rapidly progressive. Gram's staining of drainage or tissue biopsy is usually definitive, demonstrating large gram-positive (or gram-variable) rods, an absence of inflammatory cells, and widespread soft-tissue necrosis.

Spontaneous (Nontraumatic) Gas Gangrene

Spontaneous gas gangrene generally occurs via hematogenous seeding of normal muscle with histotoxic clostridia—principally *C. perfringens*, *C. septicum*, and *C. novyi* and occasionally *C. tertium*—from a gastrointestinal tract portal of entry (as in colonic malignancy, inflammatory bowel disease, diverticulitis, necrotizing enterocolitis, cecitis, or distal ileitis or after gastrointestinal surgery). These gastrointestinal pathologies permit bacterial access to the bloodstream; consequently, aerotolerant *C. septicum* can proliferate in normal tissues. Patients surviving bacteremia or spontaneous gangrene due to *C. septicum* should undergo aggressive diagnostic studies to rule out gastrointestinal pathology.

Additional predisposing host factors include leukemia, lympho-proliferative disorders, cancer chemotherapy, radiation therapy, and AIDS. Cyclic, congenital, or acquired neutropenia is also strongly associated with an increased incidence of spontaneous gas gangrene due to *C. septicum*; in such cases, necrotizing enterocolitis, cecitis, or distal ileitis is common, particularly among children.

The first symptom of spontaneous gas gangrene may be confusion followed by the abrupt onset of excruciating pain in the absence of trauma. These findings, along with fever, should heighten suspicion of spontaneous gas gangrene. However, because of the lack of an obvious portal of entry, the correct diagnosis is frequently delayed or missed. The infection is characterized by rapid progression of tissue destruction with demonstrable gas in the tissue (**Fig. 142-2**). Swelling increases and bullae filled with clear, cloudy, hemorrhagic, or purplish fluid appear. The surrounding skin has a purple hue, which may reflect vascular compromise resulting from the diffusion of bacterial toxins into surrounding tissues. Invasion of healthy tissue rapidly ensues, with quick progression to shock and multiple-organ failure. Mortality rates in this setting range from 67 to 100% among adults; among children, the mortality rate is 59%, with the majority of deaths occurring within 24 h of onset.

Figure 142-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiograph of patient with spontaneous gas gangrene due to *C. septicum*, demonstrating gas in the affected arm and shoulder.

Pathogenesis of Gas Gangrene

In traumatic gas gangrene, organisms are introduced into devitalized tissue. It is important to recognize that for *C. perfringens* and *C. novyi*, trauma must be sufficient to interrupt the blood supply and thereby to establish an optimal anaerobic environment for growth of these species. These conditions are not strictly required for the more aerotolerant species such as *C. septicum* and *C. tertium*, which can seed normal tissues from gastrointestinal lesions. Once introduced into an appropriate niche, the organisms proliferate locally and elaborate exotoxins.

The major *C. perfringens* extracellular toxins implicated in gas gangrene are α toxin and θ toxin. A lethal hemolysin that has both phospholipase C and sphingomyelinase activities, α toxin has been implicated as the major virulence factor of *C. perfringens*: immunization of mice with the C-terminal domain of α toxin provides protection against lethal challenge with *C. perfringens*, and isogenic α toxin–deficient mutant strains of *C. perfringens* are not lethal in a murine model of gas gangrene. It has been shown in experimental models that the severe pain, rapid progression, marked tissue destruction, and absence of neutrophils in *C. perfringens* gas gangrene are attributable in large part to α toxin–induced occlusion of blood vessels by heterotypic aggregates of platelets and neutrophils. The formation of these aggregates, which occurs within minutes, is largely mediated by α toxin's ability to activate the platelet adhesion molecule gpIIbIIIa (**Fig. 142-3**); the implication is that platelet glycoprotein inhibitors (e.g., eptifibatide, abciximab) may be therapeutic for maintaining tissue blood flow.

Figure 142-3

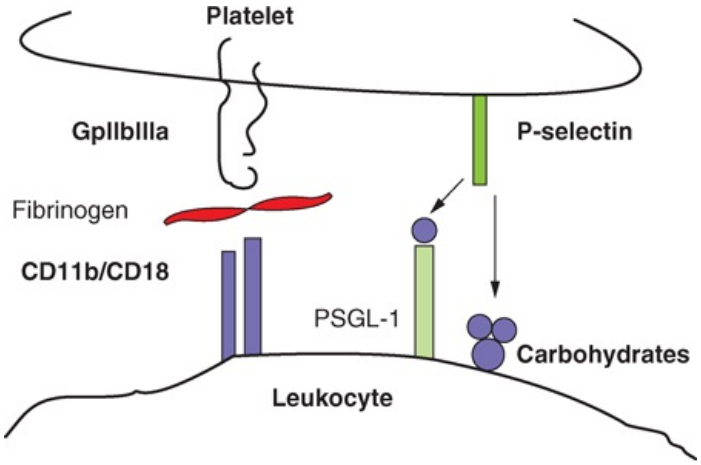

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic illustration of the molecular mechanisms of C. perfringens α toxin–induced platelet/neutrophil aggregates. Homotypic aggregates of platelets (not shown) and heterotypic aggregates of platelets and leukocytes are due to α *toxin–induced activation of the platelet fibrinogen receptor gpIIbIIIa and upregulation of leukocyte CD11b/CD18.* Binding of fibrinogen (*red*) bridges the connection between these adhesion molecules on adjacent cells. An auxiliary role for α toxin–induced upregulation of platelet P-selectin and its binding to leukocyte P-selectin glycoprotein ligand 1 (PSGL-1) or other leukocyte surface carbohydrates has also been demonstrated.

C. perfringens θ toxin (*perfringolysin*) is a member of the thiol-activated cytolysin family known as cholesterol-dependent cytolysins, which includes streptolysin O from group A *Streptococcus*, pneumolysin from *Streptococcus pneumoniae*, and several other toxins. Cholesterol-dependent cytolysins bind as oligomers to cholesterol in host cell membranes. At high concentrations, these toxins form ring-like pores resulting in cell lysis. At sublytic concentrations, θ toxin hyperactivates phagocytes and vascular endothelial cells.

Cardiovascular collapse and end-organ failure occur late in the course of *C. perfringens* gas gangrene and are largely attributable to both direct and indirect effects of α and θ toxins. In experimental models, θ toxin causes markedly reduced systemic vascular resistance but increased cardiac output (e.g., "warm shock"), probably via induction of endogenous mediators (e.g., prostacyclin, platelet-activating factor) that cause vasodilation. This effect is similar to that observed in gram-negative sepsis. In sharp contrast, α toxin directly suppresses myocardial contractility; the consequence is profound hypotension due to a sudden reduction in cardiac output. The roles of other endogenous mediators, such as cytokines (e.g., tumor necrosis factor, interleukin 1, interleukin 6) and vasodilators (e.g., bradykinin) have not been fully elucidated.

C. septicum produces four main toxins—α toxin (lethal, hemolytic, necrotizing activity), β toxin (DNase), γ toxin (hyaluronidase), and Δ" toxin (septicolysin, an oxygen-labile hemolysin)—as well as a protease and a neuraminidase. Unlike the α toxin of *C. perfringens*, that of *C. septicum* does not possess phospholipase activity. The mechanisms remain to be fully elucidated, but it is likely that each of these toxins contributes uniquely to *C. septicum* gas gangrene.

Treatment: Gas Gangrene

Patients with suspected gas gangrene (either traumatic or spontaneous) should undergo prompt surgical inspection of the infected site. Direct examination of a gram-stained smear of the involved tissues is of major importance. Characteristic histologic findings in clostridial gas gangrene include widespread tissue destruction, a paucity of leukocytes in infected tissues in conjunction with an accumulation of leukocytes in adjacent vessels (**Fig. 142-4**), and the presence of gram-positive rods (with or without spores). CT and MRI are invaluable for determining whether the infection is localized or is spreading along fascial planes, and needle aspiration or punch biopsy may provide an etiologic diagnosis in at least 20% of cases. However, these techniques should not replace surgical exploration, Gram's staining, and histopathologic examination. When spontaneous gas gangrene is suspected, blood should be cultured since bacteremia usually precedes cutaneous manifestations by several hours.

Figure 142-4

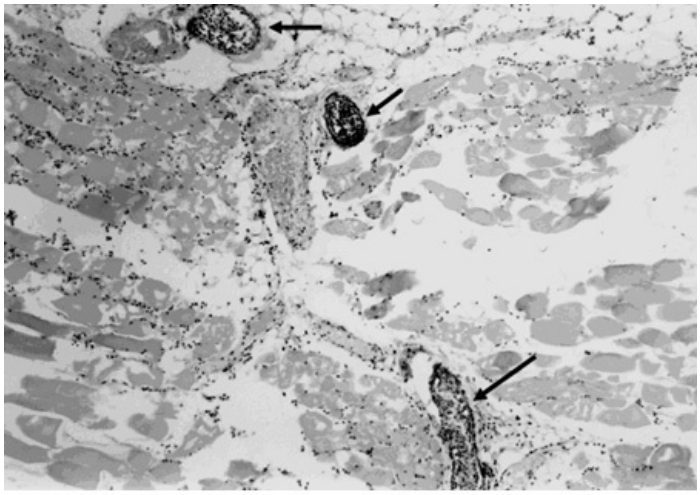

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histopathology of experimental gas gangrene due to *C. perfringens*, demonstrating widespread muscle necrosis, a paucity of leukocytes in infected tissues, and accumulation of leukocytes in adjacent vessels (*arrows*). These features are due to the effects of α and θ toxins on muscle cells, platelets, leukocytes, and endothelial cells.

For patients with evidence of clostridial gas gangrene, thorough emergent surgical debridement is of extreme importance. All devitalized tissue should be widely resected back to healthy viable muscle and skin so as to remove conditions that allow anaerobic organisms to continue proliferating. Closure of traumatic wounds or compound fractures should be delayed for 5–6 days until it is certain that these sites are free of infection.

Antibiotic treatment of traumatic or spontaneous gas gangrene (Table 142-1) consists of the administration of penicillin and clindamycin for 10–14 days. Penicillin is recommended on the basis of in vitro sensitivity data; clindamycin is recommended because of its superior efficacy over penicillin in animal models of *C. perfringens* gas gangrene and in some clinical reports. Controlled clinical trials comparing the efficacy of these agents in humans have not been performed. In the penicillin-allergic patient, clindamycin may be used alone. The superior efficacy of clindamycin is probably due to its ability to inhibit bacterial protein toxin production, its insensitivity to the size of the bacterial load or the stage of bacterial growth, and its ability to modulate the host's immune response.

C. tertium is resistant to penicillin, cephalosporins, and clindamycin. Appropriate antibiotic therapy for *C. tertium* infection is vancomycin (1 g every 12 h IV) or metronidazole (500 mg every 8 h IV).

The value of adjunctive treatment with hyperbaric oxygen (HBO) for gas gangrene remains controversial. Basic science studies suggest that HBO can inhibit the growth of *C. perfringens* but not that of the more aerotolerant *C. septicum*. In vitro, blood and macerated muscle inhibit the bactericidal potential of HBO. Numerous studies in animals demonstrate little efficacy of HBO alone, whereas antibiotics alone—especially those that inhibit bacterial protein synthesis—confer marked benefits. Addition of HBO to the therapeutic regimen provides some additional benefit, but only if surgery and antibiotic administration precede HBO treatment.

In conclusion, gas gangrene is a rapidly progressive infection whose outcome depends on prompt recognition, emergent surgery, and timely administration of antibiotics that inhibit toxin production. Gas gangrene associated with bacteremia probably represents a later stage of illness and is associated with the worst outcomes. Emergent surgical debridement is crucial to ensure survival, and ancillary procedures (e.g., CT or MRI) or transport to HBO units should not delay this intervention. Some trauma centers associated with HBO units may have special expertise in managing these aggressive infections, but proximity and speed of transfer must be carefully weighed against the need for haste.

Prognosis of Gas Gangrene

The prognosis for patients with gas gangrene is more favorable when the infection involves an extremity rather than the trunk or visceral organs, since debridement of the latter sites is more difficult. Gas gangrene is most likely to progress to shock and death in patients with associated bacteremia and intravascular hemolysis. Mortality rates are highest for patients in shock at the time of diagnosis. Mortality rates are relatively high among patients with spontaneous gas gangrene, especially that due to*C. septicum*. Survivors of gas gangrene may undergo multiple debridements and face long periods of hospitalization and rehabilitation.

Prevention of Gas Gangrene

Initial aggressive debridement of devitalized tissue can reduce the risk of gas gangrene in

contaminated deep wounds. Interventions to be avoided include prolonged application of tourniquets and surgical closure of traumatic wounds; patients with compound fractures are at significant risk for gas gangrene if the wound is closed surgically. Vaccination against α toxin is protective in experimental animal models of *C. perfringens* gas gangrene but has not been investigated in humans. In addition, as mentioned above, a hyperimmune globulin would represent a significant advance for prophylaxis in victims of acute traumatic injury or for attenuation of the spread of infection in patients with established gas gangrene.

TOXIC SHOCK SYNDROME

Clostridial infection of the endometrium, particularly that due to *C. sordellii*, can develop after gynecologic procedures, childbirth, or abortion (spontaneous or elective, surgical or medical) and, once established, proceeds rapidly to TSS and death. Systemic manifestations including edema, effusions, profound leukocytosis, and hemoconcentration are followed by the rapid onset of hypotension and multiple-organ failure. Elevation of the hematocrit to 75–80% and leukocytosis of 50,000–200,000 cells per microliter, with a left shift, are characteristic of *C. sordellii* infection. Pain may not be a prominent feature, and fever is typically absent. In one series, 18% of 45 cases of *C. sordellii* infection were associated with normal childbirth, 11% with medically induced abortion, and 0.4% with spontaneous abortion; the case-fatality rate was 100% in these groups. Of the infections in this series that were not related to gynecologic procedures or childbirth, 22% occurred in injection drug users, and 50% of these patients died. Other infections followed trauma or surgery (42%), mostly in healthy persons, and 53% of these patients died. Overall, the mortality rate was 69% (31 of45 cases). Of patients who succumbed, 85% died within 2–6 days of initial infection or following procedures.

Early diagnosis of *C. sordellii* infections often proves difficult for several reasons. First, the prevalence of these infections is low. Second, the initial symptoms are nonspecific and frankly misleading. Early in the course, the illness resembles any number of infectious diseases, including viral syndromes. Given these vague symptoms and an absence of fever, physicians usually do not aggressively pursue additional diagnostic tests. The absence of local evidence of infection and the lack of fever make early diagnosis of *C. sordellii* infection particularly problematic in patients who develop deep-seated infection following childbirth, therapeutic abortion, gastrointestinal surgery, or trauma. Such patients are frequently evaluated for pulmonary embolization, gastrointestinal bleeding, pyelonephritis, or cholecystitis. Unfortunately, such delays in diagnosis increase the risk of death, and, as in most necrotizing soft-tissue infections, patients are hypotensive with evidence of organ dysfunction by the time local signs and symptoms become apparent. In contrast, infection is more readily suspected in injection drug users presenting with local swelling, pain, and redness at injection sites; early recognition probably contributes to the lower mortality rates in this group.

Physicians should suspect *C. sordellii* infection in patients who present within 2–7 days after injury, surgery, drug injection, childbirth, or abortion and who complain of pain, nausea, vomiting, and diarrhea but are afebrile. There is little information regarding appropriate treatment for *C. sordellii* infections. In fact, the interval between onset of symptoms and death is often so short that there is little time to initiate empirical antimicrobial therapy. Indeed, anaerobic cultures of blood and wound aspirates are time-consuming, and many hospital laboratories do not routinely perform antimicrobial sensitivity testing on anaerobes. Antibiotic susceptibility data from older studies suggest that *C. sordellii*, like most clostridia, is susceptible to β-lactam antibiotics, clindamycin, tetracycline, and chloramphenicol but is resistant to aminoglycosides and sulfonamides. Antibiotics that suppress toxin synthesis (e.g., clindamycin) may possibly prove useful as therapeutic adjuncts since they are effective in necrotizing infections due to other toxin-producing gram-positive organisms.

OTHER CLOSTRIDIAL SKIN AND SOFT-TISSUE INFECTIONS

Crepitant cellulitis (also called anaerobic cellulitis) occurs principally in diabetic patients and characteristically involves subcutaneous tissues or retroperitoneal tissues, whereas the muscle and fascia are not involved. This infection can progress to fulminant systemic disease.

Cases of *C. histolyticum* infection with cellulitis, abscess formation, or endocarditis have also been documented in injection drug users. Endophthalmitis due to *C. sordellii* or *C. perfringens* has been described. *C. ramosum* is also isolated frequently from clinical specimens, including blood and intraabdominal and soft tissues. This species may be resistant to clindamycin and multiple cephalosporins.

FURTHER READINGS

Aldape MJ et al: *Clostridium sordellii* infection: Epidemiology, clinical findings, and current perspectives on diagnosis and treatment. Clin Infect Dis 43:1436, 2006[PMID: 17083018]

Bodey GP et al: Clostridial bacteremia in cancer patients. A 12-year experience. Cancer 67:1928, 1991[PMID: 2004306]

Bos J et al: Fatal necrotizing colitis following a foodborne outbreak of enterotoxigenic *Clostridium*

perfringens type A infection. Clin Infect Dis 40:e78, 2005

Bryant AE et al: Clostridial gas gangrene II: Phospholipase C–induced activation of platelet gpIIb/IIIa mediates vascular occlusion and myonecrosis in *C. perfringens* gas gangrene.J Infect Dis 182:808, 2000[PMID: 10950775]

Obladen M: Necrotizing enterocolitis—150 years of fruitless search for the cause. Neonatology 96:203, 2009[PMID: 19407465]

Sayeed S et al: Beta toxin is essential for the intestinal virulence of *Clostridium perfringens* type C disease isolate CN3685 in a rabbit ileal loop model. Mol Microbiol 67:15, 2008[PMID: 18078439]

Smith LDS, Williams BL: *The Pathogenic Anaerobic Bacteria,* 3rd ed. Springfield, IL: Charles C Thomas, 1984

Stevens DL et al: *Clostridium*, in *American Society of Microbiology's Manual of Clinical Microbiology*, 10th ed. Washington, DC, ASM Press, 2010

——— et al: Practice guidelines for the diagnosis and management of skin and soft-tissue infections. Clin Infect Dis 41:1373, 2005

Wang C et al: Hyperbaric oxygen for treating wounds: A systematic review of the literature. Arch Surg 138:272, 2003[PMID: 12611573]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MENINGOCOCCAL INFECTIONS: INTRODUCTION

Definition

Infection with *Neisseria meningitidis* most commonly manifests as asymptomatic colonization in the nasopharynx of healthy adolescents and adults. Invasive disease occurs rarely, usually presenting as either bacterial meningitis or meningococcal septicemia. Patients may also present with occult bacteremia, pneumonia, septic arthritis, conjunctivitis, and chronic meningococcemia.

Etiology and Microbiology

N. meningitidis is a gram-negative aerobic diplococcus that colonizes humans only and that causes disease after transmission to a susceptible individual. Several related organisms have been recognized, including the pathogen *N. gonorrhoeae* and the commensals *N. lactamica*, *N. flavescens*, *N. mucosa*, *N. sicca*, and *N. subflava* . *N. meningitidis* is a catalase- and oxidase-positive organism that utilizes glucose and maltose to produce acid.

Meningococci associated with invasive disease are usually encapsulated with polysaccharide, and the antigenic nature of the capsule determines an organism's serogroup **(Table 143-1)**. In total, 13 serogroups have been identified (A–D, X–Z, 29E, W135, H–J, and L), but just 5 serogroups—A, B, C, Y, and W135—account for the majority of cases of invasive disease. Acapsular meningococci are commonly isolated from the nasopharynx in studies of carriage; the lack of capsule often is a result of phase variation of capsule expression, but as many as 16% of isolates lack the genes for capsule synthesis and assembly. These "capsule-null" meningococci and those that express capsules other than A, B, C, Y, and W135 are only rarely associated with invasive disease and are most commonly identified in the nasopharynx of asymptomatic carriers.

Table 143-1 Structure of the Polysaccharide Capsule of Common Disease-Causing Meningococci

Meningococcal Serogroup	Chemical Structure of Oligosaccharide	Current Disease Epidemiology
A	2-Acetamido-2-deoxy-D-mannopyranosyl phosphate	Epidemic disease mainly in sub-Saharan Africa; sporadic cases worldwide
B	α-2,8-*N*-acetylneuraminic acid	Sporadic cases worldwide; propensity to cause hyperendemic disease
C	α-2,9-O-acetylneuraminic acid	Small outbreaks and sporadic disease
Y	4-*O*-α-D-glucopyranosyl-*N*-acetylneuraminic acid	Sporadic disease and occasional small institutional outbreaks
W135	4-*O*-α-D-galactopyranosyl-*N*-acetylneuraminic acid	Sporadic disease; outbreaks of disease associated with mass gatherings; epidemics in sub-Saharan Africa

Beneath the capsule, meningococci are surrounded by an outer phospholipid membrane containing lipopolysaccharide (LPS, endotoxin) and multiple outer-membrane proteins (**Figs. 143-1** and 143-2). Antigenic variability in porins expressed in the outer membrane defines the serotype (PorB) and serosubtype (PorA) of the organism, and structural differences in LPS determine the immunotype. Serologic methods for typing of meningococci are restricted by the limited availability of serologic reagents that can distinguish among the organisms' highly variable surface proteins. Where available, high-throughput antigen gene sequencing has superseded serology for meningococcal typing. A large database of antigen gene sequences for the outer-membrane proteins PorA, PorB, FetA, Opa, and factor H–binding protein is available online (www.neisseria.org). The number of specialized iron-regulated proteins found in the meningococcal outer membrane (e.g., FetA and transferrin-binding proteins) highlights the organisms' dependence on iron from human sources. A thin peptidoglycan cell wall separates the outer membrane from the cytoplasmic membrane.

Figure 143-1

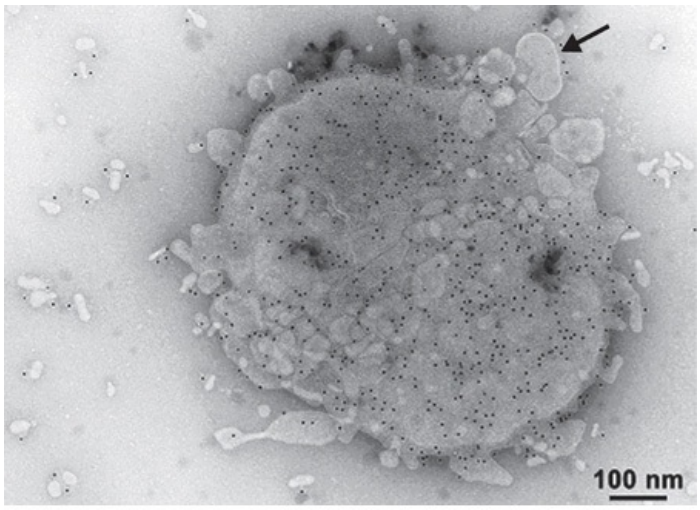

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Electron micrograph of *Neisseria meningitidis*. Black dots are gold-labeled polyclonal antibodies binding surface opacity proteins. Blebs of outer membrane can be seen being released from the bacterial surface (see arrow). *(Photo courtesy of D. Ferguson, Oxford University.)*

Figure 143-2

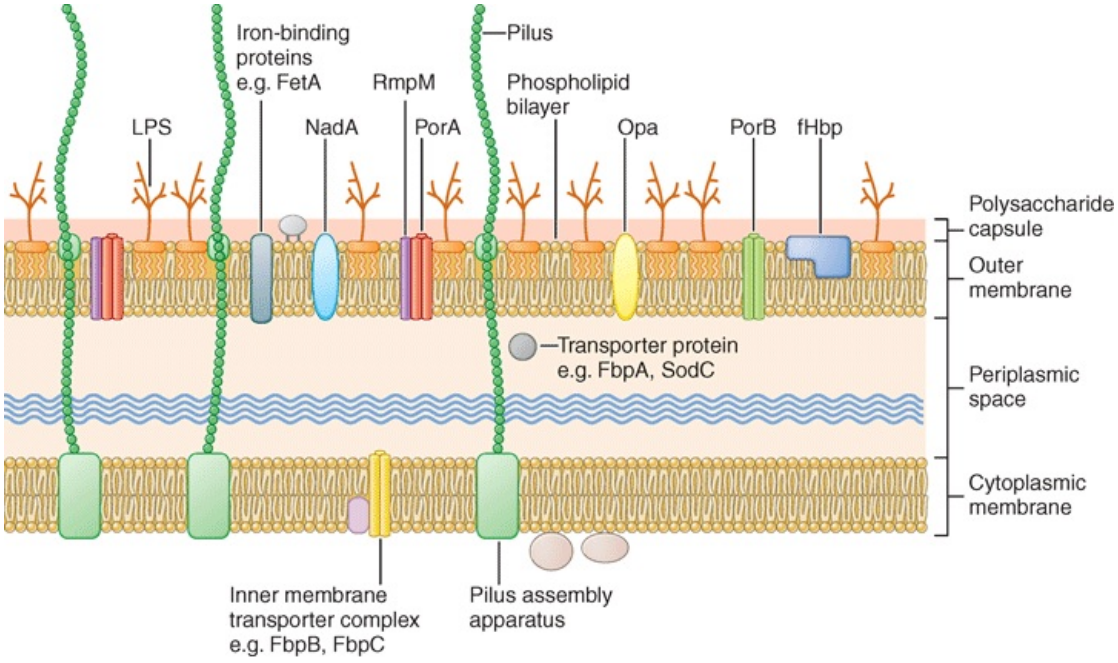

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cross-section through surface structures of *Neisseria meningitidis*. *(Reprinted with permission from Sadarangani and Pollard, 2010.)*

The structure of meningococcal populations involved in local and global spread has been studied with multilocus enzyme electrophoresis (MLEE), which characterizes isolates according to differences in the electrophoretic mobility of cytoplasmic enzymes. However, this technique has mostly been replaced by multilocus sequence typing (MLST), in which meningococci are characterized by sequence types assigned on the basis of sequences of internal fragments of seven housekeeping genes. The online MLST database currently includes more than 13,000 meningococcal isolates and 7600 unique sequence types (http://pubmlst.org/neisseria/). Seven hyperinvasive lineages of *N. meningitidis* have been identified and are responsible for the majority of cases of invasive meningococcal disease worldwide. The apparent genetic stability of these meningococcal clones over decades and during wide geographic spread indicates that they are well adapted to the nasopharyngeal environment of the

host and to efficient transmission.

The group B meningococcal genome is >2 megabases in length and contains 2158 coding regions. Many genes undergo phase variation that makes it possible to control their expression; this capacity is likely to be important in meningococcal adaptation to the host environment and evasion of the immune response. Meningococci can obtain DNA from their environment and can acquire new genes —including the capsular operon—such that a switch from one serogroup to another can occur.

Epidemiology

PATTERNS OF DISEASE

Up to 500,000 cases of meningococcal disease are thought to occur worldwide each year, and ~10% of the individuals affected die. There are several patterns of disease: epidemic, outbreak (small clusters of cases), hyperendemic, and sporadic or endemic.

Epidemics have continued since the original descriptions of meningococcal disease, especially affecting the sub-Saharan meningitis belt of Africa, where tens to hundreds of thousands of cases (caused mainly by serogroup A but also by serogroups W135 and X) may be reported over a season and rates may be as high as 1000 cases per 100,000 population. Serogroup A epidemics took place in Europe and North America after the First and Second World Wars, and serogroup A outbreaks have been documented over the past 30 years in New Zealand, China, Nepal, Mongolia, India, Pakistan, Poland, and Russia.

Clusters of cases occur where there is an opportunity for increased transmission—i.e., in (semi-)closed communities such as schools, colleges, universities, military training centers, and refugee camps. Recently, such clusters have been especially strongly linked with a particular clone (sequence type 11) that is mainly associated with the serogroup C capsule. Wider and more prolonged community outbreaks (hyperendemic disease) due to single clones of serogroup B meningococci account for ≥10 cases per 100,000. Regions affected in the past decade include the U.S. Pacific Northwest, New Zealand (both islands), and the province of Normandy in France.

Most countries now experience predominantly sporadic cases (0.3–5 cases per 100,000 population), with many different disease-causing clones involved and usually no clear epidemiologic link between one case and another. The disease rate and the distribution of meningococcal strains vary in different regions of the world and also in any one location over time. For example, in the United States, the rate of meningococcal disease fell from 1.2 cases per 100,000 population in 1997 to <0.4 case per 100,000 in 2007 (**Fig. 143-3**). Meningococcal disease in this country was previously dominated by serogroups B and C; however, serogroup Y emerged during the 1990s and became more common than serogroup C in 2007. In contrast, rates of disease in England and Wales rose to >5 cases per 100,000 during the 1990s because of an increase in cases caused by the ST11 serogroup C clone. As a result of a mass immunization program against serogroup C in 1999, almost all cases in the United Kingdom are now attributed to serogroup B (**Fig. 143-4**).

Figure 143-3

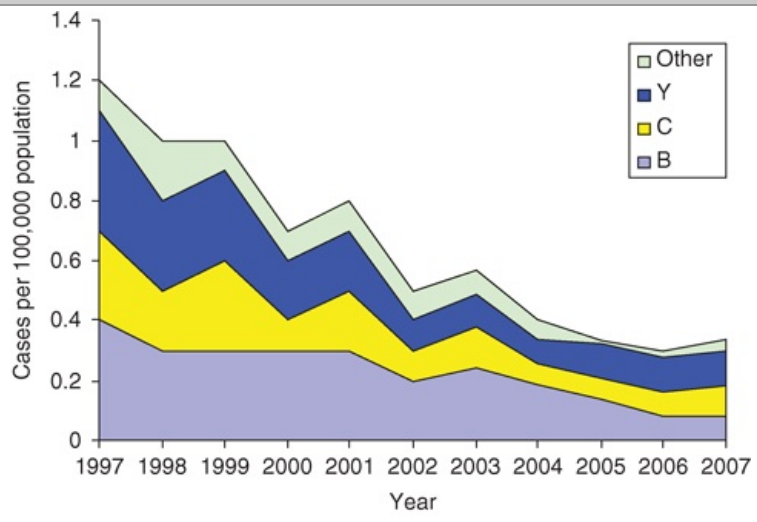

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Meningococcal disease in the United States over time. *(Adapted from ABC Surveillance data, Centers for Disease Control and Prevention; www.cdc.gov.)*

Figure 143-4

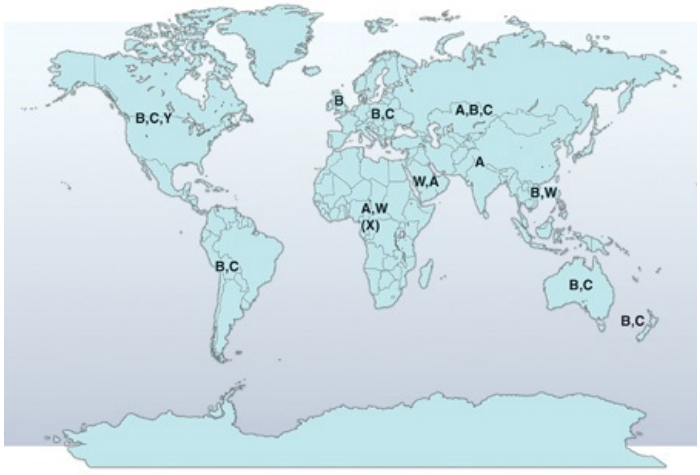

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Global distribution of meningococcal serogroups, 1999–2009.

FACTORS ASSOCIATED WITH DISEASE RISK AND SUSCEPTIBILITY

The principal determinant of disease susceptibility is age, with the peak incidence in the first year of life (**Fig. 143-5**). The susceptibility of the very young presumably results from an absence of specific adaptive immunity in combination with very close contact with colonized individuals, including parents. Compared with other age groups, infants appear to be particularly susceptible to serogroup B disease: >30% of serogroup B cases in the United States occur during the first year of life. In the early 1990s in North America, the median ages for patients with disease due to serogroups B, C, Y, and W135 were 6, 17, 24, and 33 years, respectively.

Figure 143-5

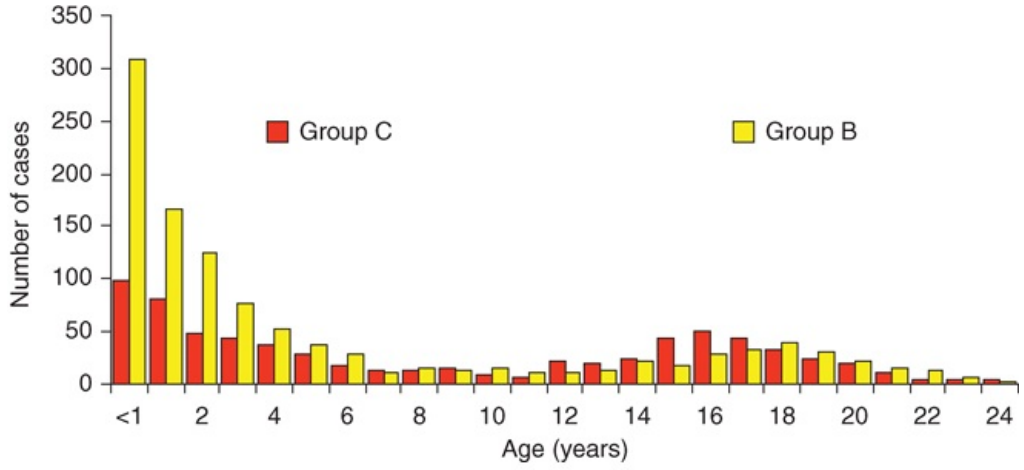

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Age distribution of serogroups B and C meningococcal disease in England and Wales, 1998/1999. *(Health Protection Agency, UK; www.hpa.org.uk.)*

After early childhood, a second peak of disease occurs in adolescents and young adults (15–25 years of age) in Europe and North America. It is thought that this peak relates to social behaviors and environmental exposures in this age group, as discussed below. Most cases of infection with *N. meningitidis* in developed countries today are sporadic, and the rarity of the disease suggests that individual susceptibility may be important. A number of factors probably contribute to individual susceptibility, including the host's genetic constitution, environment, and contact with a carrier or a case.

The best-documented genetic association with meningococcal disease is complement deficiency, chiefly of the terminal complement components (C5–9), properdin, or factor D; such a deficiency increases the risk of disease by up to 600-fold and may result in recurrent attacks. Complement

components are believed to be important for the bactericidal activity of serum, which is considered the principal mechanism of immunity against invasive meningococcal disease. However, when investigated, complement deficiency is found in only a very small proportion of individuals with meningococcal disease (0.3%). Conversely, 7–20% of persons whose disease is caused by the less common serogroups (W135, X, Y, Z, 29E) have a complement deficiency. Complement deficiency appears to be associated with serogroup B disease only rarely. Individuals with recurrences of meningococcal disease, particularly those caused by non-B serogroups, should be assessed for complement deficiency by measurement of total hemolytic complement activity. There is also limited evidence that hyposplenism (through reduction in phagocytic capacity) and hypogammaglobulinemia (through absence of specific antibody) increase the risk of meningococcal disease. Genetic studies have revealed various associations with disease susceptibility, including complement and mannose-binding lectin deficiency, single-nucleotide polymorphisms in Toll-like receptor (TLR) 4 and complement factor H, and variants of Fc gamma receptors.

Factors that increase the chance of a susceptible individual acquiring *N. meningitidis* via the respiratory route also increase the risk of meningococcal disease. Acquisition occurs through close contact with carriers as a result of overcrowding (e.g., in poor socioeconomic settings, in refugee camps, during the Hajj pilgrimage to Mecca, and during freshman-year residence in college dormitories) and certain social behaviors (e.g., attendance at bars and nightclubs, kissing). Secondary cases may occur in close contacts of an index case (e.g., household members and persons kissing the infected individual); the risk to these contacts may be as high as 1000 times the background rate in the population. Factors that damage the nasopharyngeal epithelium also increase the risk of both colonization with *N. meningitidis* and invasive disease. The most important of these factors are cigarette smoking (odds ratio, 4.1) and passive exposure to cigarette smoke. In addition, recent viral respiratory tract infection, infection with *Mycoplasma* species, and winter or the dry season have been associated with meningococcal disease; all of these factors presumably either increase the expression of adhesion molecules in the nasopharynx, thus enhancing meningococcal adhesion, or facilitate meningococcal invasion of the bloodstream.

Pathogenesis

N. meningitidis has evolved as an effective colonizer of the human nasopharynx, with asymptomatic infection rates of >25% described in some series of adolescents and young adults and among residents of crowded communities. Point-prevalence studies reveal widely divergent rates of carriage for different types of meningococci. This variation suggests that some types may be adapted to a short duration of carriage with frequent transmission to maintain the population, while others may be less efficiently transmitted but may overcome this disadvantage by colonizing for a long period. Despite the high rates of carriage among adolescents and young adults, only ~10% of adults carry meningococci, and colonization is very rare in early childhood. Many of the same factors that increase the risk of meningococcal disease also increase the risk of carriage, including smoking, crowding, and respiratory viral infection. Colonization of the nasopharynx involves a series of interactions of meningococcal adhesins (e.g., Opa proteins and pili) with their ligands on the epithelial mucosa. *N. meningitidis* produces an IgA1 protease that is likely to reduce interruption of colonization by mucosal IgA.

Colonization should be considered the normal state of meningococcal infection, with an increased risk of invasion the unfortunate consequence (for both host and organism) of adaptations of hyperinvasive meningococcal lineages. The meningococcal capsule is an important virulence factor: acapsular strains rarely cause invasive disease. The capsule provides resistance to phagocytosis and may be important in preventing desiccation during transmission between hosts. Antigenic diversity in surface structures and an ability to vary levels of their expression have probably evolved as important factors in maintaining meningococcal populations within and between individual hosts.

Invasion through the mucosa into the blood occurs rarely, usually within a few days of acquisition of an invasive strain by a susceptible individual. Only occasional cases of prolonged colonization prior to invasion have been documented. Once the organism is in the bloodstream, its growth may be limited if the individual is partially immune, although bacteremia may allow seeding of another site, such as the meninges or the joints. Alternatively, unchecked proliferation may continue, resulting in high bacterial counts in the circulation. During growth, meningococci release blebs of outer membrane (Fig. 143-1) containing outer-membrane proteins and endotoxin (LPS). Endotoxin binds cell-bound CD14 in association with TLR4 to initiate an inflammatory cascade with the release of high levels of various mediators, including tumor necrosis factor α, soluble tumor necrosis factor receptor, interleukin (IL) 1, IL-1 receptor antagonist, IL-1β, IL-6, IL-8, IL-10, plasminogen-activator inhibitor 1 (PAI-1), and leukemia inhibitory factor. Soluble CD14-bound endotoxin acts as a mediator of endothelial activation. The severity of meningococcal disease is related both to the levels of endotoxin in the blood and to the magnitude of the inflammatory response. The latter is determined to some extent by polymorphisms in the inflammatory response genes (and their inhibitors), and the release of the inflammatory cascade heralds the development of meningococcal septicemia (meningococcemia). Endothelial injury is central to many clinical features of meningococcemia, including increased vascular permeability, pathologic changes in vascular tone, loss of thromboresistance, intravascular coagulation, and myocardial dysfunction. Endothelial injury leads to increased vascular permeability (attributed to loss of

glycosaminoglycans and endothelial proteins), with subsequent gross proteinuria. Leakage of fluid and electrolytes into the tissues from capillaries leads to hypovolemia, tissue edema, and pulmonary edema. Initial compensation results in vasoconstriction and tachycardia, although cardiac output eventually falls. While resuscitation fluids may restore circulating volume, tissue edema will continue to increase, and, in the lung, the consequence may be respiratory failure.

Intravascular thrombosis (caused by activation of procoagulant pathways in association with upregulation of tissue factor on the endothelium) occurs in some patients with meningococcal disease and results in purpura fulminans and infarction of areas of skin or even of whole limbs. At the same time, multiple anticoagulant pathways are downregulated through loss of endothelial thrombomodulin and protein C receptors and decreases in levels of antithrombin III, protein C, protein S, and tissue factor pathway inhibitor. Thrombolysis is also profoundly impaired in meningococcal sepsis through the release of high levels of PAI-1.

Shock in meningococcal septicemia appears to be attributable to a combination of factors, including hypovolemia, which results from the capillary leak syndrome secondary to endothelial injury, and myocardial depression, which is driven by hypovolemia, hypoxia, metabolic derangements (e.g., hypocalcemia), and cytokines (e.g., IL-6). Decreased perfusion of tissues as a result of intravascular thrombosis, vasoconstriction, tissue edema, and reduced cardiac output in meningococcal septicemia can cause widespread organ dysfunction, including renal impairment and—later in the disease—a decreased level of consciousness due to central nervous system involvement.

Bacteria that reach the meninges cause a local inflammatory response, with release of a spectrum of cytokines similar to that seen in septicemia, that presents clinically as meningitis and is believed to determine the severity of neuronal injury. Local endothelial injury may result in cerebral edema and rapid onset of raised intracranial pressure in some cases.

Clinical Manifestations

As discussed above, the most common form of infection with *N. meningitidis* is asymptomatic carriage of the organism in the nasopharynx. Despite the location of infection in the upper airway, meningococcal pharyngitis is rarely reported; however, upper respiratory tract symptoms are common prior to presentation with invasive disease. It is not clear whether these symptoms relate to preceding viral infection (which may promote meningococcal acquisition) or to meningococcal acquisition itself. After acquiring the organism, susceptible individuals develop disease manifestations in 1–10 days (usually <4 days, although colonization for 11 weeks has been documented).

Along the spectrum of presentations of meningococcal disease, the most common clinical syndromes are meningitis and meningococcal septicemia. In fulminant cases, death may occur within hours of the first symptoms. Occult bacteremia is also recognized and, if untreated, progresses in two-thirds of cases to focal infection including meningitis or septicemia. Meningococcal disease may also present as pneumonia, pyogenic arthritis or osteomyelitis, purulent pericarditis, endophthalmitis, conjunctivitis, primary peritonitis, or (rarely) urethritis. Perhaps because it is difficult to diagnose, pneumococcal pneumonia is not commonly reported but is associated with serogroups Y, W135, and Z and appears most often to affect individuals >10 years of age.

RASH

A nonblanching rash (petechial or purpuric) develops in >80% of cases of meningococcal disease; however, the rash is often absent early in the illness. Usually initially blanching in nature (macules, maculopapules, or urticaria) and indistinguishable from more common viral rashes, the rash of meningococcal infection becomes petechial or frankly purpuric over the hours after onset. In the most severe cases, large purpuric lesions develop (purpura fulminans). Some patients (including those with overwhelming sepsis) may have no rash. While petechial rash and fever are important signs of meningococcal disease, fewer than 10% of children (and, in some clinical settings, fewer than 1% of patients) with this presentation are found to have meningococcal disease. Most patients presenting with a petechial or purpuric rash have a viral infection **(Table 143-2)**. The skin lesions exhibit widespread endothelial necrosis and occlusion of small vessels in the dermis and subcutaneous tissues, with a neutrophilic infiltrate.

Table 143-2 Common Causes of Petechial or Purpuric Rashes

Enteroviruses
Influenza and other respiratory viruses
Measles virus Epstein-Barr virus
Cytomegalovirus
Parvovirus
Deficiency of protein C or S (including postvaricella protein S deficiency)
Platelet disorders (e.g., idiopathic thrombocytopenic purpura, drug effects, bone marrow infiltration)

Henoch-Schönlein purpura, connective tissue disorders, trauma (including nonaccidental injuries in children)
Pneumococcal, streptococcal, staphylococcal, or gram-negative bacterial sepsis

MENINGITIS

Meningococcal meningitis commonly presents as nonspecific manifestations, including fever, vomiting, and (especially in infants and young children) irritability, and is indistinguishable from other forms of bacterial meningitis unless there is an associated petechial or purpuric rash, which occurs in two-thirds of cases. Headache is rarely reported in early childhood but is more common in later childhood and adulthood. When headache is present, the following features, in association with fever or a history of fever, are suggestive of bacterial meningitis: neck stiffness, photophobia, decreased level of consciousness, seizures or status epilepticus, and focal neurologic signs. Classic signs of meningitis, such as neck stiffness and photophobia, are often absent in infants and young children with bacterial meningitis.

While 30–50% of patients present with a meningitis syndrome alone, up to 40% of meningitis patients also present with some features of septicemia. Most deaths from meningococcal meningitis alone (i.e., without septicemia) are associated with raised intracranial pressure presenting as a reduced level of consciousness, relative bradycardia and hypertension, focal neurologic signs, abnormal posturing, and signs of brainstem involvement—e.g., unequal, dilated, or poorly reactive pupils; abnormal eye movement; and impaired corneal responses (Chap. 274).

SEPTICEMIA

Meningococcal septicemia alone accounts for up to 20% of cases of meningococcal disease. The condition may progress from early nonspecific symptoms to death within hours. Mortality rates among children with this syndrome have been high (25–40%), but aggressive management (as discussed below) may reduce the figure to <10%. Early symptoms are nonspecific and suggest an influenza-like illness with fever, headache, and myalgia accompanied by vomiting and abdominal pain. As discussed above, the rash, if present, may appear to be viral early in the course until petechiae or purpuric lesions develop. Purpura fulminans occurs in severe cases, with multiple large purpuric lesions and signs of peripheral ischemia. Surveys of patients have indicated that limb pain, pallor (including a mottled appearance and cyanosis), and cold hands and feet may be prominent. Shock is manifested by tachycardia, poor peripheral perfusion, tachypnea, and oliguria. Decreased cerebral perfusion leads to confusion, agitation, or decreased level of consciousness. With progressive shock, multiorgan failure ensues; hypotension is a late sign in children, who more commonly present with compensated shock. Poor outcome is associated with an absence of meningismus, hypotension, young age, coma, relatively low temperature (< 38°C), leukopenia, and thrombocytopenia. Spontaneous hemorrhage (pulmonary, gastric, or cerebral) may result from consumption of coagulation factors and thrombocytopenia.

CHRONIC MENINGOCOCCEMIA

Chronic meningococcemia, which is rarely recognized, presents as repeated episodes of petechial rash associated with fever, joint pain, features of arthritis, and splenomegaly that may progress to acute meningococcal septicemia if untreated. During the relapsing course, bacteremia characteristically clears without treatment and then recurs. The differential diagnosis includes bacterial endocarditis, acute rheumatic fever, Henoch-Schönlein purpura, infectious mononucleosis, disseminated gonococcal infection, and immune-mediated vasculitis. This condition has been associated with complement deficiencies in some cases and with inadequate sulfonamide therapy in others.

POSTMENINGOCOCCAL REACTIVE DISEASE

In a small proportion of patients, an immune complex disease develops ~4–10 days after the onset of meningococcal disease, with manifestations that include a maculopapular or vasculitic rash (2% of cases), arthritis (up to 8% of cases), iritis (1%), pericarditis, and/or polyserositis associated with fever. The immune complexes involve meningococcal polysaccharide antigen and result in immunoglobulin and complement deposition with an inflammatory infiltrate. These features resolve spontaneously without sequelae. It is important to recognize this condition since a new onset of fever and rash can lead to concerns about relapse of meningococcal disease and unnecessarily prolonged antibiotic treatment.

Diagnosis

Like other invasive bacterial infections, meningococcal disease may produce elevations of the white blood cell (WBC) count and of values for inflammatory markers (e.g., C-reactive protein and procalcitonin levels or the erythrocyte sedimentation rate). Values may be normal or low in rapidly progressive disease, and lack of these elevations does not exclude the diagnosis. However, in the presence of fever and a petechial rash, these elevations are suggestive of meningococcal disease. In patients with severe meningococcal septicemia, common laboratory findings include hypoglycemia,

acidosis, hypokalemia, hypocalcemia, hypomagnesemia, hypophosphatemia, anemia, and coagulopathy.

Although meningococcal disease is often diagnosed on clinical grounds, in suspected meningococcal meningitis or meningococcemia, blood should routinely be sent for culture to confirm the diagnosis and to facilitate public health investigations; blood cultures are positive in up to 75% of cases. Culture media containing sodium polyanethol sulfonate, which may inhibit meningococcal growth, should be avoided. Meningococcal viability is reduced if there is a delay in transport of the specimen to the microbiology laboratory for culture or in plating of cerebrospinal fluid (CSF) samples. In countries where treatment with antibiotics before hospitalization is recommended for meningococcal disease, the majority of clinically suspected cases are culture negative. Real-time polymerase chain reaction (PCR) analysis of whole-blood samples increases the diagnostic yield by >40%, and results obtained with this method may remain positive for several days after administration of antibiotics. Indeed, in the United Kingdom, more than half of clinically suspected cases are currently identified by PCR.

Unless contraindications exist (raised intracranial pressure, uncorrected shock, disordered coagulation, thrombocytopenia, respiratory insufficiency, local infection, ongoing convulsions), lumbar puncture should be undertaken to identify and confirm the etiology of suspected meningococcal meningitis, whose presentation cannot be distinguished from that of meningitis of other bacterial causes. Some authorities have recommended a CT brain scan prior to lumbar puncture because of the risk of cerebral herniation in patients with raised intracranial pressure. However, a normal CT scan is not uncommon in the presence of raised intracranial pressure in meningococcal meningitis, and the decision to perform a lumbar puncture should be made on clinical grounds. CSF features of meningococcal meningitis (elevated protein level and WBC count, decreased glucose level) are indistinguishable from those of other types of bacterial meningitis unless a gram-negative diplococcus is identified. (Gram's staining is up to 80% sensitive for meningococcal meningitis.) CSF should be submitted for culture (sensitivity, 90%) and (where available) PCR analysis. CSF antigen testing with latex agglutination is insensitive and should be replaced by molecular diagnosis when possible.

Lumbar puncture should generally be avoided in meningococcal septicemia, as positioning for the procedure may critically compromise the patient's circulation in the context of hypovolemic shock. Delayed lumbar puncture may still be useful when the diagnosis is uncertain, particularly if molecular technology is available.

In other types of focal infection, culture and PCR analysis of normally sterile body fluids (e.g., synovial fluid) may aid in the diagnosis. Although some authorities have recommended cultures of scrapings or aspirates from skin lesions, this procedure adds little to the diagnostic yield when compared with a combination of blood culture and PCR analysis. Urinary antigen testing is also insensitive, and serologic testing for meningococcal infection has not been adequately studied. Because *N. meningitidis* is a component of the normal human nasopharyngeal flora, identification of the organism on throat swabs has no diagnostic value.

Treatment: Meningococcal Infections

Death from meningococcal disease is associated most commonly with hypovolemic shock (meningococcemia) and occasionally with raised intracranial pressure (meningococcal meningitis). Therefore, management should focus on the treatment of these urgent clinical issues in addition to the administration of specific antibiotic therapy. Delayed recognition of meningococcal disease or its associated physiologic derangements, together with inadequate emergency management, is associated with poor outcome. Since the disease is rare, protocols for emergency management have been developed (see www.meningitis.org).

Airway patency may be compromised if the level of consciousness is depressed as a result of shock (impaired cerebral perfusion) or raised intracranial pressure; this situation may require intervention. In meningococcemia, pulmonary edema and pulmonary oligemia (presenting as hypoxia) require oxygen therapy or elective endotracheal intubation. In cases with shock, aggressive fluid resuscitation (with replacement of the circulating volume several times in severe cases) and inotropic support may be necessary to maintain cardiac output. If shock persists after volume resuscitation at 40 mL/kg, the risk of pulmonary edema is high, and elective intubation is recommended to improve oxygenation and decrease the work of breathing. Metabolic derangements including hypoglycemia, acidosis, hypokalemia, hypocalcemia, hypomagnesemia, hypophosphatemia, anemia, and coagulopathy should be anticipated and corrected. In the presence of raised intracranial pressure, management includes correction of coexistent shock and neurointensive care to maintain cerebral perfusion.

Empirical antibiotic therapy for suspected meningococcal disease consists of a third-generation cephalosporin such as ceftriaxone [75–100 mg/kg per day (maximum, 4 g/d) in one or two divided IV doses] or cefotaxime [200 mg/kg per day (maximum, 8 g/d) in four divided IV doses] to cover the various other (potentially penicillin-resistant) bacteria that may produce an indistinguishable clinical syndrome. Although unusual in most countries, reduced meningococcal sensitivity to penicillin (a minimal inhibitory concentration of 0.12–1.0 μg/mL) has been reported from Africa, the United Kingdom, Spain, Argentina, the United States, and Canada.

Both meningococcal meningitis and meningococcal septicemia are conventionally treated for 7 days, although courses of 3–5 days may be equally effective. Furthermore, a single dose of ceftriaxone or an oily suspension of chloramphenicol has been used successfully in resource-poor settings. No data are available to guide the duration of treatment for meningococcal infection at other foci (e.g., pneumonia, arthritis); antimicrobial therapy is usually continued until clinical and laboratory evidence of infection has resolved.

The use of glucocorticoids for adjunctive treatment of meningococcal meningitis remains controversial since no relevant studies have had sufficient power to determine true efficacy. One large study in adults did indicate a trend toward benefit, and in clinical practice a decision to use glucocorticoids usually precedes a definite diagnosis. Therapeutic doses of glucocorticoids are not recommended in meningococcal septicemia, but many intensivists recommend replacement glucocorticoid doses for patients who have refractory shock in association with impaired adrenal gland responsiveness. Various other adjunctive therapies for meningococcal disease have been considered, but few have been subjected to clinical trials and none can currently be recommended. An antibody to LPS (HA1A) failed to confer a demonstrable benefit. Recombinant bactericidal/permeability-increasing protein was tested in a study that had inadequate power to show an effect on mortality rates; however, there were trends toward lower mortality rates among patients who received a complete infusion, and this group also had fewer amputations, fewer blood-product transfusions, and a significantly improved functional outcome. Given that protein C concentrations are reduced in meningococcal disease, the use of activated protein C has been considered since a survival benefit was demonstrated in adult sepsis trials; however, trials in pediatric sepsis (of particular relevance for meningococcal disease) found no benefit and indicated a potential risk of bleeding complications with use of activated protein C.

The postmeningococcal immune-complex inflammatory syndrome has been treated with nonsteroidal anti-inflammatory agents until spontaneous resolution occurs.

Complications

About 10% of patients with meningococcal disease die despite the availability of antimicrobial therapy and other intensive medical interventions. The most common complication of meningococcal disease (10% of cases) is scarring after necrosis of purpuric skin lesions, for which skin grafting may be necessary. The lower limbs are most often affected; next in frequency are the upper limbs, the trunk, and the face. On average, 13% of the skin surface area is involved. Amputations are necessary in ~2% of survivors of meningococcal disease because of a loss of tissue viability after peripheral ischemia or compartment syndromes. Unless there is local infection, amputation should usually be delayed to allow the demarcation between viable and nonviable tissue to become apparent. Approximately 4% of patients with meningococcal disease suffer hearing loss, and 7% have neurologic complications. In one study, pain was reported by 21% of survivors. In some investigations, the rate of complications is higher for serogroup C disease (mostly associated with the ST11 clone) than for serogroup B disease. In patients with severe hypovolemic shock, renal perfusion may be impaired and prerenal failure is common, but permanent renal replacement therapy is rarely needed.

Several studies suggest adverse psychosocial outcomes after meningococcal disease, with reduced quality of life, lowered self-esteem, and poorer neurologic development, including increased rates of attention deficit/hyperactivity disorder and special educational needs. Other studies have not found evidence of such outcomes.

Prognosis

Several prognostic scoring systems have been developed to identify patients with meningococcal disease who are least likely to survive. Factors associated with a poorer prognosis are shock; young age (infancy), old age, and adolescence; coma; purpura fulminans; disseminated intravascular coagulation; thrombocytopenia; leukopenia; absence of meningitis; metabolic acidosis; low plasma concentrations of antithrombin and proteins S and C; high blood levels of PAI-1; and a low erythrocyte sedimentation rate or C-reactive protein level. The Glasgow Meningococcal Septicaemia Prognostic Score is probably the best-performing scoring system studied so far and may be clinically useful for severity assessment in meningococcal disease. However, scoring systems do not direct the clinician to specific interventions, and the priority in management should be recognition of compromised airways, breathing, or circulation and direct, urgent intervention. Most patients improve rapidly with appropriate antibiotics and supportive therapy. Fulminant meningococcemia is more likely to result in death or ischemic skin loss than is meningitis; optimal emergency management may reduce mortality rates among the most severely affected patients.

Prevention

Since mortality rates in meningococcal disease remain high despite improvements in intensive care management, immunization is the only rational approach to prevention on a population level. Secondary cases are common among household and "kissing" contacts of cases, and secondary prophylaxis with antibiotic therapy is widely recommended for these contacts (see below).

POLYSACCHARIDE VACCINES

Purified capsular polysaccharide has been used for immunization since the 1960s. Meningococcal polysaccharide vaccines are currently formulated as either bivalent (serogroups A and C) or quadrivalent (serogroups A, C, Y, and W135), with 50 μg of each polysaccharide per dose. Local reactions (erythema, induration, and tenderness) may occur in up to 40% of vaccinees, but serious adverse events (including febrile convulsions in young children) are very rarely reported. In adults, the vaccines are immunogenic, but immunity appears to be relatively short-lived (with antibody levels above baseline for only 2–10 years), and booster doses do not induce a further rise in antibody concentration. Indeed, a state of immunologic hyporesponsiveness has been widely reported to follow booster doses of plain polysaccharide vaccines. The repeating units of these vaccines cross-link B cell receptors to drive specific memory B cells to become plasma cells and produce antibody. Because polysaccharides are T cell–independent antigens, no memory B cells are produced after immunization, and the memory B cell pool is depleted such that fewer polysaccharide-specific cells are available to respond to a subsequent dose of vaccine **(Fig. 143-6)**. The clinical relevance of hyporesponsiveness is unknown. Plain polysaccharide vaccines generally are not immunogenic in early childhood, possibly because marginal-zone B cells are involved in polysaccharide responses and maturation of the splenic marginal zone is not complete until 18 months to 2 years of age. The efficacy of the meningococcal serogroup C component is >90% in young adults; no efficacy data are available for the serogroup Y and W135 polysaccharides in this age group.

Figure 143-6

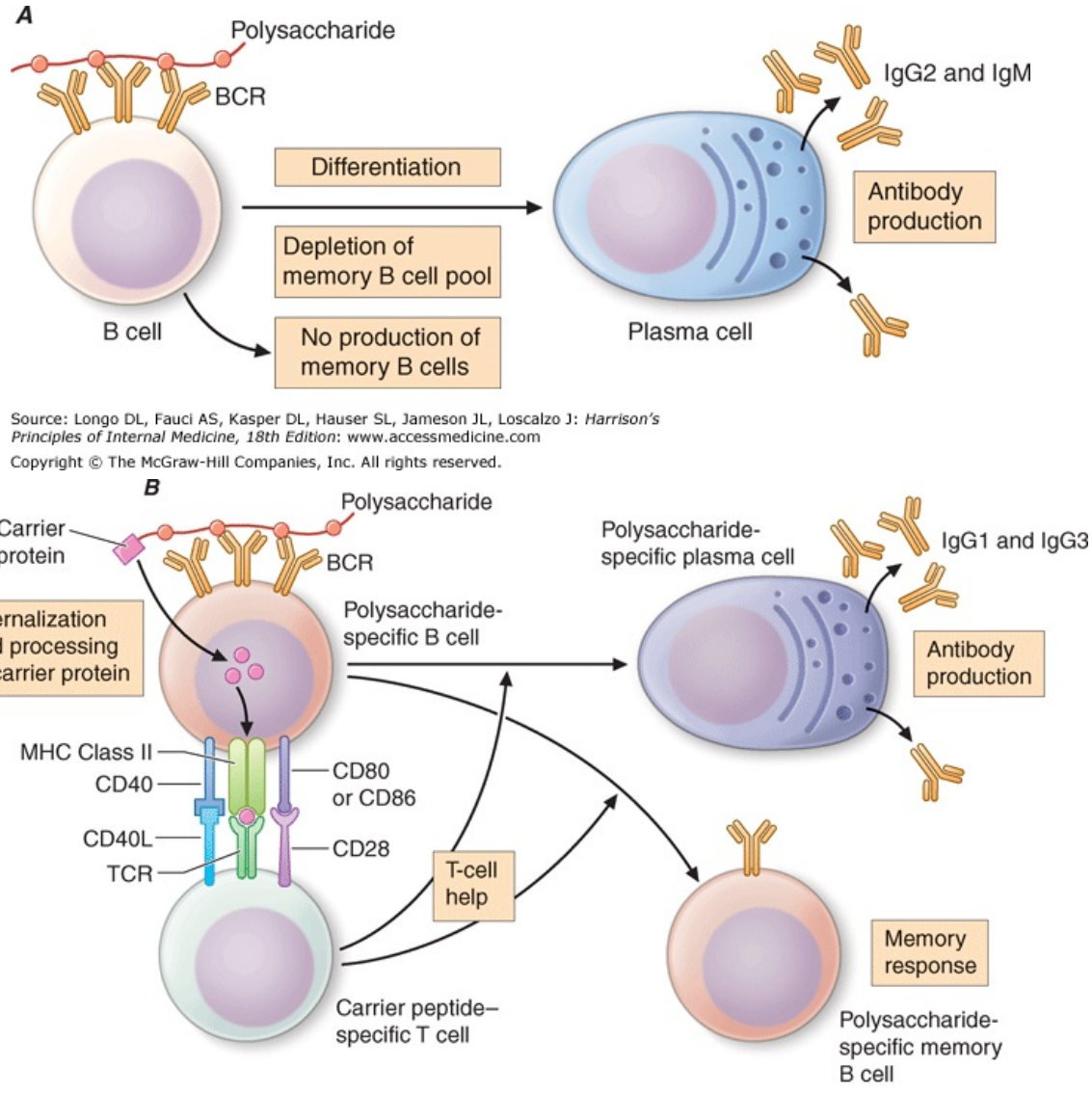

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Polysaccharides from the encapsulated bacteria that cause disease in early childhood stimulate B cells by cross-linking the BCR and driving the production of immunoglobulins. There is no production of memory B cells, and the B cell pool may be depleted by this process such that subsequent immune responses are decreased. ***B.*** The carrier protein from protein-polysaccharide conjugate vaccines is processed by the polysaccharide-specific B cell, and peptides are presented to carrier peptide

–specific T cells, with the consequent production of both plasma cells and memory B cells. BCR, B cell receptor; MHC, major histocompatibility complex; TCR, T cell receptor. *(Reprinted from Pollard et al, 2009.)*

Group A meningococcal polysaccharides are exceptional in that they are effective in preventing disease at all ages. Two doses administered 2–3 months apart to children 3–18 months of age or a single dose administered to older children or adults has a protective efficacy rate of >95%. The vaccine has been widely used in the control of meningococcal disease in the African meningitis belt. The duration of protection appears to be only 3–5 years.

There is no meningococcal serogroup B plain polysaccharide vaccine because ∝-2,8-*N*-acetylneuraminic acid is expressed on the surface of neural cells in the fetus such that the B polysaccharide is perceived as "self" and therefore is not immunogenic in humans.

CONJUGATE VACCINES

The poor immunogenicity of plain polysaccharide vaccines in infancy has been overcome by chemical conjugation of the polysaccharides to a carrier protein (CRM_{197}, tetanus toxoid, or diphtheria toxoid). Conjugates that contain monovalent serogroup C polysaccharide and quadrivalent vaccines with A, C, Y, and W135 polysaccharides have been developed, as have vaccines including various other antigen combinations. After immunization, peptides from the carrier protein are conventionally believed to be presented to peptide-specific T cells in association with major histocompatibility complex (MHC) class II molecules (some recent data suggesting that carrier protein peptide may actually be presented in association with an oligosaccharide and MHCII) by polysaccharide-specific B cells; the result is a T cell –dependent immune response that allows production of antibody and generation of an expanded B cell memory pool. Unlike responses to booster doses of plain polysaccharides, responses to booster doses of conjugate vaccines have the characteristics of memory responses. Indeed, conjugate vaccines overcome the hyporesponsiveness induced by plain polysaccharides by replenishing the memory pool. The reactogenicity of conjugate vaccines is similar to that of plain polysaccharide vaccines.

The first widespread use of serogroup C meningococcal conjugate vaccine (MenC) came in 1999 in the United Kingdom after a rise in serogroup C disease. A mass vaccination campaign involving all individuals <19 years of age was undertaken, and the number of cases fell from >1000 in 1999 to just 28 in 2006. The effectiveness of the immunization program was attributed both to direct protection of immunized persons and to reduced transmission of the organism in the population as a result of decreased rates of colonization among the immunized (herd immunity). Data on immunogenicity and effectiveness have shown that the duration of protection is short when the vaccine is administered in early childhood; thus booster doses are needed to maintain population immunity. In contrast, immunity after a dose of vaccine given in adolescence appears to be prolonged.

The first quadrivalent conjugate meningococcal vaccine containing A, C, Y, and W135 polysaccharides conjugated to diphtheria toxoid was initially recommended for all children >11 years of age in the United States in 2005. In 2007 the license was extended to high-risk children 2–10 years of age. In the same year, the vaccine was licensed in Canada for persons 2–55 years of age. Uptake was slow, but preliminary data suggest an efficacy rate of >80%. Limited data from the U.S. Vaccine Adverse Events Reporting System indicated that there might be a short-term increase in the risk of Guillain-Barré syndrome after immunization with the diphtheria conjugate vaccine; however, further investigation has not confirmed this finding. Another quadrivalent conjugate vaccine (CRM197 carrier protein) was licensed in 2010 in both Europe and North America, and a third vaccine is in late-stage development.

A monovalent serogroup A vaccine was developed and licensed in 2010 and is being rolled out in several countries in sub-saharan Africa. The goal for this vaccine is control of epidemic meningococcal disease in the African meningitis belt.

VACCINES BASED ON SUBCAPSULAR ANTIGENS

The lack of immunogenicity of the serogroup B capsule has led to the development of vaccines based on subcapsular antigens. Various surface components have been studied in early-phase clinical trials. Outer-membrane vesicles (OMVs) containing outer-membrane proteins, phospholipid, and lipopolysaccharide can be extracted from cultures of *N. meningitidis* by detergent treatment **(Fig. 143-7)**. OMVs prepared in this way were used in efficacy trials with a Norwegian outbreak strain and reduced the incidence of group B disease among 14- to 16-year-old schoolchildren by 53%. Similarly, OMV vaccines constructed from local outbreak strains in Cuba and New Zealand have had reported efficacy rates of >70%. These OMV vaccines appear to produce strain-specific immune responses, with only limited cross-protection, and are therefore best suited to clonal outbreaks (e.g., those in Cuba and New Zealand as well as others in Norway and the province of Normandy in France).

Figure 143-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Illustration of meningococcal outer-membrane vesicle containing outer-membrane structures.

Several purified surface proteins have been evaluated in phase 1 clinical trials but have not yet been developed further because of variability or poor immunogenicity (e.g., transferrin-binding proteins, neisserial surface protein A). Other vaccine candidates have been identified since sequencing of the meningococcal genome. A combination vaccine that includes the New Zealand OMV vaccine and three proteins (neisserial adhesin A, factor H–binding protein, and neisserial heparin-binding antigen) is immunogenic in infancy and was recently submitted for licensure. Finally, a highly immunogenic vaccine based on two variants of factor H–binding protein is undergoing clinical evaluation.

Management of Contacts

Close (household and kissing) contacts of individuals with meningococcal disease are at increased risk (up to 1000 times the rate for the general population) of developing secondary disease; a secondary case follows as many as 3% of sporadic cases. About one-fifth of secondary cases are actually co-primary cases—i.e., cases that occur soon after the primary case and in which transmission is presumed to have originated from the same third party. The rate of secondary cases is highest during the week after presentation of the index case. The risk falls rapidly but remains above baseline for up to 1 year after the index case; 30% of secondary cases occur in the first week, 20% in the second week, and most of the remainder over the next 6 weeks. In outbreaks of meningococcal disease, mass prophylaxis has been used; however, limited data support population intervention, and significant concerns have arisen about adverse events and the development of resistance. For these reasons, prophylaxis is usually restricted to (1) persons at greatest risk who are intimate and/or household contacts of the index case and (2) health care workers who have been directly exposed to respiratory secretions. In most cases, members of wider communities (e.g., at schools or colleges) are not offered prophylaxis.

The aim of prophylaxis is to eradicate colonization of close contacts with the strain that has caused invasive disease in the index case. Prophylaxis should be given to all contacts at the same time to avoid recolonization by meningococci transmitted from untreated contacts and should also be used as soon as possible to treat early disease in secondary cases. If the index patient is treated with an antibiotic that does not reliably clear colonization (e.g., penicillin), he or she should be given a prophylactic agent at the end of treatment to prevent relapse or onward transmission. Although rifampin has been most widely used and studied, it is not the optimal agent because it fails to eradicate carriage in 15–20% of cases, rates of adverse events have been high, compliance is affected by the need for four doses, and emerging resistance has been reported. Ceftriaxone as a single IM or IV injection is highly (97%) effective in carriage eradication and can be used at all ages and in pregnancy. Reduced susceptibility of isolates to ceftriaxone has occasionally been reported. Ciprofloxacin or ofloxacin is preferred in some countries; this agent is also highly effective and can be administered by mouth but is not recommended in pregnancy. Resistance to fluoroquinolones has been reported in some meningococci in North America, Europe, and Asia.

In documented serogroup A, C, Y, or W135 disease, contacts may be offered immunization (preferably

with a conjugate vaccine) in addition to chemoprophylaxis to provide protection beyond the duration of antibiotic therapy. Mass vaccination has been used successfully to control disease during outbreaks in closed communities (educational and military establishments) as well as during epidemics in open communities.

FURTHER READINGS

Caugant DA, Maiden MC: Meningococcal carriage and disease—population biology and evolution. Vaccine 27(Suppl 2):B64, 2009

Harrison LH et al: Global epidemiology of meningococcal disease. Vaccine 27:B51, 2009

Laforce FM et al: Epidemic meningitis due to group A *Neisseria meningitidis* in the African meningitis belt: A persistent problem with an imminent solution. Vaccine 27:B13, 2009

Milonovich LM: Meningococcemia: Epidemiology, pathophysiology, and management. J Pediatr Health Care 21:75, 2007[PMID: 17321906]

Pace D et al: Quadrivalent meningococcal conjugate vaccines. Vaccine 27:B30, 2009

Pollard AJ et al: Maintaining protection against invasive bacteria in childhood with protein –polysaccharide conjugate vaccines. Nat Rev Immunol 9:213, 2009[PMID: 19214194]

——— et al: Early management of meningococcal disease. Arch Dis Child 80:290, 1999

Sadarangani M, Pollard AJ: Serogroup B meningococcal vaccines—an unfinished story. Lancet Infect Dis 10:112, 2010[PMID: 20113980]

Stephens DS et al: Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. Lancet 369:2196, 2007[PMID: 17604802]

Wright V et al: Genetic polymorphisms in host response to meningococcal infection: The role of susceptibility and severity genes. Vaccine 27:B90, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GONOCOCCAL INFECTIONS: INTRODUCTION

Definition

Gonorrhea is a sexually transmitted infection (STI) of epithelium and commonly manifests as cervicitis, urethritis, proctitis, and conjunctivitis. If untreated, infections at these sites can lead to local complications such as endometritis, salpingitis, tuboovarian abscess, bartholinitis, peritonitis, and perihepatitis in female patients; periurethritis and epididymitis in male patients; and ophthalmia neonatorum in newborns. Disseminated gonococcemia is an uncommon event whose manifestations include skin lesions, tenosynovitis, arthritis, and (in rare cases) endocarditis or meningitis.

Microbiology

Neisseria gonorrhoeae is a gram-negative, nonmotile, non-spore-forming organism that grows singly and in pairs (i.e., as monococci and diplococci, respectively). Exclusively a human pathogen, the gonococcus contains, on average, three genome copies per coccal unit; this polyploidy permits a high level of antigenic variation and the survival of the organism in its host. Gonococci, like all other *Neisseria* species, are oxidase positive. They are distinguished from other neisseriae by their ability to grow on selective media and to utilize glucose but not maltose, sucrose, or lactose.

Epidemiology

The incidence of gonorrhea has declined significantly in the United States, but there were still ∑299,000 newly reported cases in 2008. Gonorrhea remains a major public health problem worldwide, is a significant cause of morbidity in developing countries, and may play a role in enhancing transmission of HIV.

Gonorrhea predominantly affects young, nonwhite, unmarried, less educated members of urban populations. The number of reported cases probably represents half of the true number of cases—a discrepancy resulting from underreporting, self-treatment, and nonspecific treatment without a laboratory-proven diagnosis. The number of reported cases of gonorrhea in the United States rose from ∑250,000 in the early 1960s to a high of 1.01 million in 1978. The recorded incidence of gonorrhea in modern times peaked in 1975, with 468 reported cases per 100,000 population in the United States. This peak was attributable to the interaction of several variables, including improved accuracy of diagnosis, changes in patterns of contraceptive use, and changes in sexual behavior. The incidence of the disease has since declined gradually and is currently estimated at 120 cases per 100,000, a figure that is still the highest among industrialized countries. A further decline in the overall incidence of gonorrhea in the United States over the past two decades may reflect increased condom use resulting from public health efforts to curtail HIV transmission. At present, the attack rate in the United States is highest among 15- to 19-year-old women and 20- to 24-year-old men; 40% of all reported cases occur in the preceding two groups together. From the standpoint of ethnicity, rates are highest among African Americans and lowest among persons of Asian or Pacific Island descent.

The incidence of gonorrhea is higher in developing countries than in industrialized nations. The exact incidence of any STI is difficult to ascertain in developing countries because of limited surveillance and variable diagnostic criteria. Studies in Africa have clearly demonstrated that nonulcerative STIs such as gonorrhea (in addition to ulcerative STIs) are an independent risk factor for the transmission of HIV (Chap. 189).

Gonorrhea is transmitted from males to females more efficiently than in the opposite direction. The rate of transmission to a woman during a single unprotected sexual encounter with an infected man is ∑40–60%. Oropharyngeal gonorrhea occurs in ∑20% of women who practice fellatio with infected partners. Transmission in either direction by cunnilingus is rare.

In any population, there exists a small minority of individuals who have high rates of new-partner acquisition. These "core-group members" or "high-frequency transmitters" are vital in sustaining STI transmission at the population level. Another instrumental factor in sustaining gonorrhea in the population is the large number of infected individuals who are asymptomatic or have minor symptoms that are ignored. These persons, unlike symptomatic individuals, may not cease sexual activity and therefore continue to transmit the infection. This situation underscores the importance of contact tracing and empirical treatment of the sex partners of index cases.

Pathogenesis, Immunology, and Antimicrobial Resistance

OUTER-MEMBRANE PROTEINS

Pili

Fresh clinical isolates of *N. gonorrhoeae* initially form piliated (fimbriated) colonies distinguishable on translucent agar. Pilus expression is rapidly switched off with unselected subculture because of rearrangements in pilus genes. This change is a basis for antigenic variation of gonococci. Piliated strains adhere better to cells derived from human mucosal surfaces and are more virulent in organ culture models and human inoculation experiments than nonpiliated variants. In a fallopian tube explant model, pili mediate gonococcal attachment to nonciliated columnar epithelial cells. This event initiates gonococcal phagocytosis and transport through these cells to intercellular spaces near the basement membrane or directly into the subepithelial tissue. Pili are also essential for genetic competence and transformation of *N. gonorrhoeae*, which permit horizontal transfer of genetic material between different gonococcal lineages in vivo.

Opacity-Associated Protein

Another gonococcal surface protein that is important in adherence to epithelial cells is opacity-associated protein (Opa, formerly called protein II). Opa contributes to intergonococcal adhesion, which is responsible for the opaque nature of gonococcal colonies on translucent agar and the organism's adherence to a variety of eukaryotic cells, including polymorphonuclear leukocytes (PMNs). Certain Opa variants promote invasion of epithelial cells, and this effect has been linked with the ability of Opa to bind vitronectin, glycosaminoglycans, and several members of the carcinoembryonic antigen –related cell adhesion molecule (CEACAM) receptor family. *N. gonorrhoeae* Opa proteins that bind CEACAM 1, which is expressed by primary CD4+ T lymphocytes, suppress the activation and proliferation of these lymphocytes. This phenomenon may serve to explain the transient decrease in CD4+ T lymphocyte counts associated with gonococcal infection.

Porin

Porin (previously designated protein I) is the most abundant gonococcal surface protein, accounting for >50% of the organism's total outer-membrane protein. Porin molecules exist as trimers that provide anion-transporting aqueous channels through the otherwise-hydrophobic outer membrane. Porin shows stable interstrain antigenic variation and forms the basis for gonococcal serotyping. Two main serotypes have been identified: PorB.1A strains are often associated with disseminated gonococcal infection (DGI), while PorB.1B strains usually cause local genital infections only. DGI strains are generally resistant to the killing action of normal human serum and do not incite a significant local inflammatory response; therefore, they may not cause symptoms at genital sites. These characteristics may be related to the ability of PorB.1A strains to bind to complement-inhibitory molecules, resulting in a diminished inflammatory response. Porin can translocate to the cytoplasmic membrane of host cells —a process that could initiate gonococcal endocytosis and invasion.

Other Outer-Membrane Proteins

Other notable outer-membrane proteins include H.8, a lipoprotein that is present in high concentration on the surface of all gonococcal strains and is an excellent target for antibody-based diagnostic testing. Transferrin-binding proteins (Tbp1 and Tbp2) and lactoferrin-binding protein are required for scavenging iron from transferrin and lactoferrin in vivo. Transferrin and iron have been shown to enhance the attachment of iron-deprived *N. gonorrhoeae* to human endometrial cells. IgA1 protease is produced by *N. gonorrhoeae* and may protect the organism from the action of mucosal IgA.

LIPOOLIGOSACCHARIDE

Gonococcal lipooligosaccharide (LOS) consists of a lipid A and a core oligosaccharide that lacks the repeating O-carbohydrate antigenic side chain seen in other gram-negative bacteria (Chap. 120). Gonococcal LOS possesses marked endotoxic activity and contributes to the local cytotoxic effect in a fallopian tube model. LOS core sugars undergo a high degree of phase variation under different conditions of growth; this variation reflects genetic regulation and expression of glycotransferase genes that dictate the carbohydrate structure of LOS. These phenotypic changes may affect interactions of *N. gonorrhoeae* with elements of the humoral immune system (antibodies and complement) and may also influence direct binding of organisms to both professional phagocytes and nonprofessional phagocytes (epithelial cells). For example, gonococci that are sialylated at their LOS sites bind complement factor H and inhibit the alternative pathway of complement. LOS sialylation may also decrease nonopsonic Opa-mediated association with neutrophils and inhibit the oxidative burst in PMNs. The unsialylated terminal lactosamine residue of LOS binds to an asialoglycoprotein receptor on male epithelial cells, which facilitates binding and subsequent gonococcal invasion of these cells. Moreover, LOS oligosaccharide structures can modulate host immune responses. For example, the terminal monosaccharide expressed by LOS determines the C-type lectin receptor on dendritic cells that is targeted by the bacteria. In turn, the specific C-type lectin receptor engaged influences whether a T_H1- or T_H2-type response is elicited; the latter response may be less favorable for clearance of gonococcal infection.

HOST FACTORS

In addition to gonococcal structures that interact with epithelial cells, host factors seem to be important in mediating entry of gonococci into nonphagocytic cells. Activation of phosphatidylcholine-specific phospholipase C and acidic sphingomyelinase by *N. gonorrhoeae*, which results in the release of diacylglycerol and ceramide, is a requirement for the entry of *N. gonorrhoeae* into epithelial cells. Ceramide accumulation within cells leads to apoptosis, which may disrupt epithelial integrity and facilitate entry of gonococci into subepithelial tissue. Release of chemotactic factors as a result of complement activation contributes to inflammation, as does the toxic effect of LOS in provoking the release of inflammatory cytokines.

The importance of humoral immunity in host defenses against neisserial infections is best illustrated by the predisposition of persons deficient in terminal complement components (C5 through C9) to recurrent bacteremic gonococcal infections and to recurrent meningococcal meningitis or meningococcemia. Gonococcal porin induces T cell–proliferative responses in persons with urogenital gonococcal disease. A significant increase in porin-specific interleukin (IL) 4–producing CD4+ as well as CD8+ T lymphocytes is seen in individuals with mucosal gonococcal disease. A portion of these lymphocytes that show a porin-specific T_H2-type response could traffic to mucosal surfaces and play a role in immune protection against the disease. Few data clearly indicate that protective immunity is acquired from a previous gonococcal infection, although bactericidal and opsonophagocytic antibodies to porin and LOS may offer partial protection. On the other hand, women who are infected and acquire high levels of antibody to another outer-membrane protein, Rmp (reduction modifiable protein, formerly called protein III), may be especially likely to become reinfected with *N. gonorrhoeae* because Rmp antibodies block the effect of bactericidal antibodies to porin and LOS. Rmp shows little, if any, interstrain antigenic variation; therefore, Rmp antibodies potentially may block antibody-mediated killing of all gonococci. The mechanism of blocking has not been fully characterized, but Rmp antibodies noncompetitively inhibit binding of porin and LOS antibodies because of the proximity of these structures in the gonococcal outer membrane. In male volunteers who have no history of gonorrhea, the net effect of these events may influence the outcome of experimental challenge with *N. gonorrhoeae*. Because Rmp bears extensive homology to enterobacterial OmpA and meningococcal class 4 proteins, it is possible that these blocking antibodies result from prior exposure to cross-reacting proteins from these species and also play a role in first-time infection with *N. gonorrhoeae*.

GONOCOCCAL RESISTANCE TO ANTIMICROBIAL AGENTS

It is no surprise that *N. gonorrhoeae*, with its remarkable capacity to alter its antigenic structure and adapt to changes in the microenvironment, has become resistant to numerous antibiotics. The first effective agents against gonorrhea were the sulfonamides, which were introduced in the 1930s and became ineffective within a decade. Penicillin was then employed as the drug of choice for the treatment of gonorrhea. By 1965, 42% of gonococcal isolates had developed low-level resistance to penicillin G. Resistance due to the production of penicillinase arose later.

Gonococci become fully resistant to antibiotics either by chromosomal mutations or by acquisition of R factors (plasmids). Two types of chromosomal mutations have been described. The first type, which is drug specific, is a single-step mutation leading to high-level resistance. The second type involves mutations at several chromosomal loci that combine to determine the level as well as the pattern of resistance. Strains with mutations in chromosomal genes were first observed in the late 1950s. As recently as 2004, chromosomal mutations accounted for resistance to penicillin, tetracycline, or both in ∑12% of strains surveyed in the United States.

ß-Lactamase (penicillinase)–producing strains of *N. gonorrhoeae* (PPNG) carrying plasmids with the Pc^r determinant had rapidly spread worldwide by the early 1980s. *N. gonorrhoeae* strains with plasmid-borne tetracycline resistance (TRNG) can mobilize some ß-lactamase plasmids, and PPNG and TRNG occur together, sometimes along with strains exhibiting chromosomally mediated resistance (CMRNG). Penicillin, ampicillin, and tetracycline are no longer reliable for the treatment of gonorrhea and should not be used.

Quinolone-containing regimens were also recommended for treatment of gonococcal infections; the fluoroquinolones offered the advantage of antichlamydial activity when administered for 7 days. However, quinolone-resistant *N. gonorrhoeae* (QRNG) appeared soon after these agents were first used to treat gonorrhea. QRNG is particularly common in the Pacific Islands (including Hawaii) and Asia, where, in certain areas, all gonococcal strains are now resistant to quinolones. At present, QRNG is also common in parts of Europe and the Middle East. In the United States, QRNG has been identified in midwestern and eastern areas as well as in states on the Pacific coast, where resistant strains were first seen. Alterations in DNA gyrase and topoisomerase IV have been implicated as mechanisms of fluoroquinolone resistance. Resistance to spectinomycin, which has been used in the past as an alternative agent, has been reported. Since this agent usually is not associated with resistance to other antibiotics, spectinomycin can be reserved for use against multiresistant strains of *N. gonorrhoeae*. Nevertheless, outbreaks caused by strains resistant to spectinomycin have been documented in Korea and England when the drug has been used for primary treatment of gonorrhea.

Third-generation cephalosporins have remained highly effective as single-dose therapy for gonorrhea

despite recent evidence that their minimal inhibitory concentrations (MICs) against strains of *N. gonorrhoeae* are increasing. Even though the MICs of ceftriaxone against certain strains may reach 0.015–0.125 μg/mL (higher than the MICs of 0.0001–0.008 μg/mL for fully susceptible strains), these levels are greatly exceeded in the blood, the urethra, and the cervix when the routinely recommended parenteral dose of ceftriaxone is administered. All *N. gonorrhoeae* strains with reduced susceptibility to ceftriaxone and cefixime (termed *cephalosporin intermediate/resistant strains*) contain (1) a mosaic *penA* allele encoding a penicillin-binding protein 2 (PBP 2) whose sequence differs in nearly 60 amino acids from that of wild-type PBP 2 and (2) additional genetic resistance determinants that are also required for high-level penicillin resistance.

Clinical Manifestations

GONOCOCCAL INFECTIONS IN MALES

Acute urethritis is the most common clinical manifestation of gonorrhea in males. The usual incubation period after exposure is 2–7 days, although the interval can be longer and some men remain asymptomatic. Strains of the PorB.1A serotype tend to cause a greater proportion of cases of mild and asymptomatic urethritis than do PorB.1B strains. Urethral discharge and dysuria, usually without urinary frequency or urgency, are the major symptoms. The discharge initially is scant and mucoid but becomes profuse and purulent within a day or two. Gram's stain of the urethral discharge may reveal PMNs and gram-negative intracellular monococci and diplococci **(Fig. 144-1)**. The clinical manifestations of gonococcal urethritis are usually more severe and overt than those of nongonococcal urethritis, including urethritis caused by *Chlamydia trachomatis* (Chap. 176); however, exceptions are common, and it is often impossible to differentiate the causes of urethritis on clinical grounds alone. The majority of cases of urethritis seen in the United States today are not caused by *N. gonorrhoeae* and/or *C. trachomatis*. Although a number of other organisms may be responsible, many cases do not have a specific etiologic agent identified.

Figure 144-1

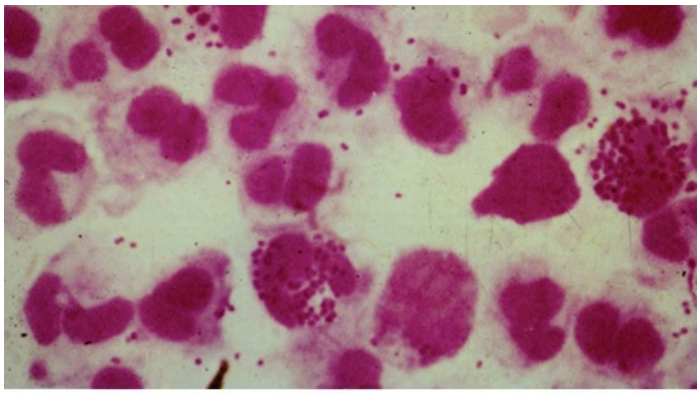

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gram's stain of urethral discharge from a male patient with gonorrhea shows gram-negative intracellular monococci and diplococci. *(From the Public Health Agency of Canada.)*

Most symptomatic men with gonorrhea seek treatment and cease to be infectious. The remaining men, who are largely asymptomatic, accumulate in number over time and constitute about two-thirds of all infected men at any point in time. Together with men incubating the organism (who shed the organism but are asymptomatic), they serve as the source of spread of infection. Before the antibiotic era, symptoms of urethritis persisted for ∑8 weeks. Epididymitis is now an uncommon complication, and gonococcal prostatitis occurs rarely, if at all. Other unusual local complications of gonococcal urethritis include edema of the penis due to dorsal lymphangitis or thrombophlebitis, submucous inflammatory "soft" infiltration of the urethral wall, periurethral abscess or fistula, inflammation or abscess of Cowper's gland, and seminal vesiculitis. Balanitis may develop in uncircumcised men.

GONOCOCCAL INFECTIONS IN FEMALES

Gonococcal Cervicitis

Mucopurulent cervicitis is the most common STI diagnosis in American women and may be caused by *N. gonorrhoeae*, *C. trachomatis*, and other organisms. Cervicitis may coexist with candidal or trichomonal vaginitis. *N. gonorrhoeae* primarily infects the columnar epithelium of the cervical os. Bartholin's glands occasionally become infected.

Women infected with *N. gonorrhoeae* usually develop symptoms. However, the women who either remain asymptomatic or have only minor symptoms may delay in seeking medical attention. These minor symptoms may include scant vaginal discharge issuing from the inflamed cervix (without vaginitis

or vaginosis per se) and dysuria (often without urgency or frequency) that may be associated with gonococcal urethritis. Although the incubation period of gonorrhea is less well defined in women than in men, symptoms usually develop within 10 days of infection and are more acute and intense than those of chlamydial cervicitis.

The physical examination may reveal a mucopurulent discharge (mucopus) issuing from the cervical os. Because Gram's stain is not sensitive for the diagnosis of gonorrhea in women, specimens should be submitted for culture or a nonculture assay (see below). Edematous and friable cervical ectopy as well as endocervical bleeding induced by gentle swabbing are more often seen in chlamydial infection. Gonococcal infection may extend deep enough to produce dyspareunia and lower abdominal or back pain. In such cases, it is imperative to consider a diagnosis of pelvic inflammatory disease (PID) and to administer treatment for that disease (Chaps. 130 and 176).

N. gonorrhoeae may be recovered from the urethra and rectum of women with cervicitis, but these are rarely the only infected sites. Urethritis in women may produce symptoms of internal dysuria, which is often attributed to "cystitis." Pyuria in the absence of bacteriuria seen on Gram's stain of unspun urine, accompanied by urine cultures that fail to yield $>10^5$ colonies of bacteria usually associated with urinary tract infection, signifies the possibility of urethritis due to *C. trachomatis*. Urethral infection with *N. gonorrhoeae* may also occur in this context, but in this instance urethral cultures are usually positive.

Gonococcal Vaginitis

The vaginal mucosa of healthy women is lined by stratified squamous epithelium and is rarely infected by *N. gonorrhoeae*. However, gonococcal vaginitis can occur in anestrogenic women (e.g., prepubertal girls and postmenopausal women), in whom the vaginal stratified squamous epithelium is often thinned down to the basilar layer, which can be infected by *N. gonorrhoeae*. The intense inflammation of the vagina makes the physical (speculum and bimanual) examination extremely painful. The vaginal mucosa is red and edematous, and an abundant purulent discharge is present. Infection in the urethra and in Skene's and Bartholin's glands often accompanies gonococcal vaginitis. Inflamed cervical erosion or abscesses in nabothian cysts may also occur. Coexisting cervicitis may result in pus in the cervical os.

ANORECTAL GONORRHEA

Because the female anatomy permits the spread of cervical exudate to the rectum, *N. gonorrhoeae* is sometimes recovered from the rectum of women with uncomplicated gonococcal cervicitis. The rectum is the sole site of infection in only 5% of women with gonorrhea. Such women are usually asymptomatic but occasionally have acute proctitis manifested by anorectal pain or pruritus, tenesmus, purulent rectal discharge, and rectal bleeding. Among men who have sex with men (MSM), the frequency of gonococcal infection, including rectal infection, fell by ≥90% throughout the United States in the early 1980s, but a resurgence of gonorrhea among MSM has been documented in several cities since the 1990s. Gonococcal isolates from the rectum of MSM tend to be more resistant to antimicrobial agents than are gonococcal isolates from other sites. Gonococcal isolates with a mutation in *mtrR* (multiple transferable resistance repressor) or in the promoter region of the gene that encodes for this transcriptional repressor develop increased resistance to antimicrobial hydrophobic agents such as bile acids and fatty acids in feces and thus are found with increased frequency in MSM. This situation may have been responsible for higher rates of failure of treatment for rectal gonorrhea with older regimens consisting of penicillin or tetracyclines.

PHARYNGEAL GONORRHEA

Pharyngeal gonorrhea is usually mild or asymptomatic, although symptomatic pharyngitis does occasionally occur with cervical lymphadenitis. The mode of acquisition is oral-genital sexual exposure, with fellatio being a more efficient means of transmission than cunnilingus. Most cases resolve spontaneously, and transmission from the pharynx to sexual contacts is rare. Pharyngeal infection almost always coexists with genital infection. Swabs from the pharynx should be plated directly onto gonococcal selective media. Pharyngeal colonization with *Neisseria meningitidis* needs to be differentiated from that with other *Neisseria* species.

OCULAR GONORRHEA IN ADULTS

Ocular gonorrhea in an adult usually results from autoinoculation from an infected genital site. As in genital infection, the manifestations range from severe to occasionally mild or asymptomatic disease. The variability in clinical manifestations may be attributable to differences in the ability of the infecting strain to elicit an inflammatory response. Infection may result in a markedly swollen eyelid, severe hyperemia and chemosis, and a profuse purulent discharge. The massively inflamed conjunctiva may be draped over the cornea and limbus. Lytic enzymes from the infiltrating PMNs occasionally cause corneal ulceration and rarely cause perforation.

Prompt recognition and treatment of this condition are of paramount importance. Gram's stain and culture of the purulent discharge establish the diagnosis. Genital cultures should also be performed.

GONORRHEA IN PREGNANT WOMEN, NEONATES, AND CHILDREN

Gonorrhea in pregnancy can have serious consequences for both the mother and the infant. Recognition of gonorrhea early in pregnancy also identifies a population at risk for other STIs, particularly chlamydial infection, syphilis, and trichomoniasis. The risks of salpingitis and PID —conditions associated with a high rate of fetal loss—are highest during the first trimester. Pharyngeal infection, most often asymptomatic, may be more common during pregnancy because of altered sexual practices. Prolonged rupture of the membranes, premature delivery, chorioamnionitis, funisitis (infection of the umbilical cord stump), and sepsis in the infant (with *N. gonorrhoeae* detected in the newborn's gastric aspirate during delivery) are common complications of maternal gonococcal infection at term. Other microorganisms and conditions, including *Mycoplasma hominis*, *Ureaplasma urealyticum*, *C. trachomatis*, and bacterial vaginosis (often accompanied by infection with *Trichomonas vaginalis*), have been associated with similar complications.

The most common form of gonorrhea in neonates is ophthalmia neonatorum, which results from exposure to infected cervical secretions during parturition. Ocular neonatal instillation of a prophylactic agent (e.g., 1% silver nitrate eyedrops or ophthalmic preparations containing erythromycin or tetracycline) prevents ophthalmia neonatorum but is not effective for its treatment, which requires systemic antibiotics. The clinical manifestations are acute and usually begin 2–5 days after birth. An initial nonspecific conjunctivitis with a serosanguineous discharge is followed by tense edema of the eyelids, chemosis, and a profuse, thick, purulent discharge. Corneal ulcerations that result in nebulae or perforation may lead to anterior synechiae, anterior staphyloma, panophthalmitis, and blindness. Infections described at other mucosal sites in infants, including vaginitis, rhinitis, and anorectal infection, are likely to be asymptomatic. Pharyngeal colonization has been demonstrated in 35% of infants with gonococcal ophthalmia, and coughing is the most prominent symptom in these cases. Septic arthritis (see below) is the most common manifestation of systemic infection or DGI in the newborn. The onset usually comes at 3–21 days of age, and polyarticular involvement is common. Sepsis, meningitis, and pneumonia are seen in rare instances.

Any STI in children beyond the neonatal period raises the possibility of sexual abuse. Gonococcal vulvovaginitis is the most common manifestation of gonococcal infection in children beyond infancy. Anorectal and pharyngeal infections are common in these children and are frequently asymptomatic. The urethra, Bartholin's and Skene's glands, and the upper genital tract are rarely involved. All children with gonococcal infection should also be evaluated for chlamydial infection, syphilis, and possibly HIV infection.

GONOCOCCAL ARTHRITIS (DGI)

DGI (gonococcal arthritis) results from gonococcal bacteremia. In the 1970s, DGI occurred in ∑0.5–3% of persons with untreated gonococcal mucosal infection. The lower incidence of DGI at present is probably attributable to a decline in the prevalence of particular strains that are likely to disseminate. DGI strains resist the bactericidal action of human serum and generally do not incite inflammation at genital sites, probably because of limited generation of chemotactic factors. Strains recovered from DGI cases in the 1970s were often of the PorB.1A serotype, were highly susceptible to penicillin, and had special growth requirements—including arginine, hypoxanthine, and uracil—that made the organism more fastidious and more difficult to isolate.

Menstruation is a risk factor for dissemination, and approximately two-thirds of cases of DGI are in women. In about half of affected women, symptoms of DGI begin within 7 days of onset of menses. Complement deficiencies, especially of the components involved in the assembly of the membrane attack complex (C5 through C9), predispose to neisserial bacteremia, and persons with more than one episode of DGI should be screened with an assay for total hemolytic complement activity.

The clinical manifestations of DGI have sometimes been classified into two stages: a bacteremic stage, which is less common today, and a joint-localized stage with suppurative arthritis. A clear-cut progression usually is not evident. Patients in the bacteremic stage have higher temperatures, and chills more frequently accompany their fever. Painful joints are common and often occur together with tenosynovitis and skin lesions. Polyarthralgias usually include the knees, elbows, and more distal joints; the axial skeleton is generally spared. Skin lesions are seen in ∑75% of patients and include papules and pustules, often with a hemorrhagic component **(Fig. 144-2**; see also Fig. e7-44). Other manifestations of noninfectious dermatitis, such as nodular lesions, urticaria, and erythema multiforme, have been described. These lesions are usually on the extremities and number between 5 and 40. The differential diagnosis of the bacteremic stage of DGI includes reactive arthritis, acute rheumatoid arthritis, sarcoidosis, erythema nodosum, drug-induced arthritis, and viral infections (e.g., hepatitis B and acute HIV infection). The distribution of joint symptoms in reactive arthritis differs from that in DGI **(Fig. 144-3)**, as do the skin and genital manifestations (Chap. 325).

Figure 144-2

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

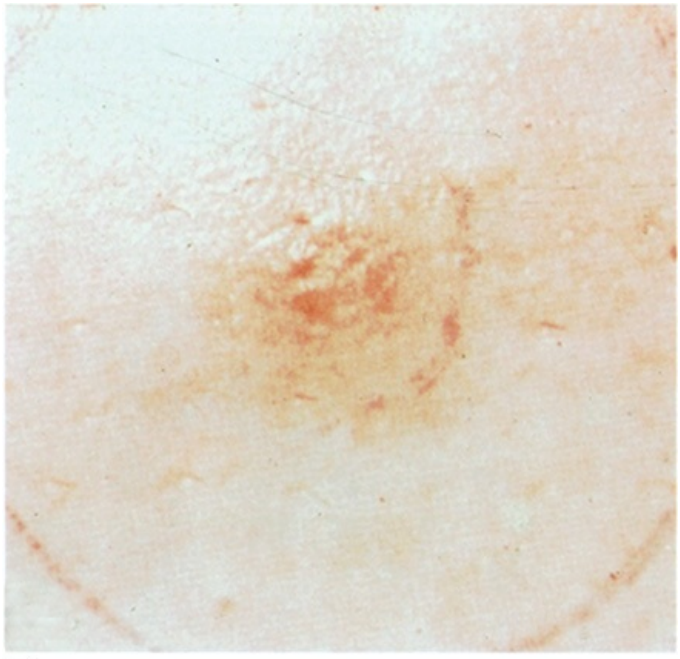

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

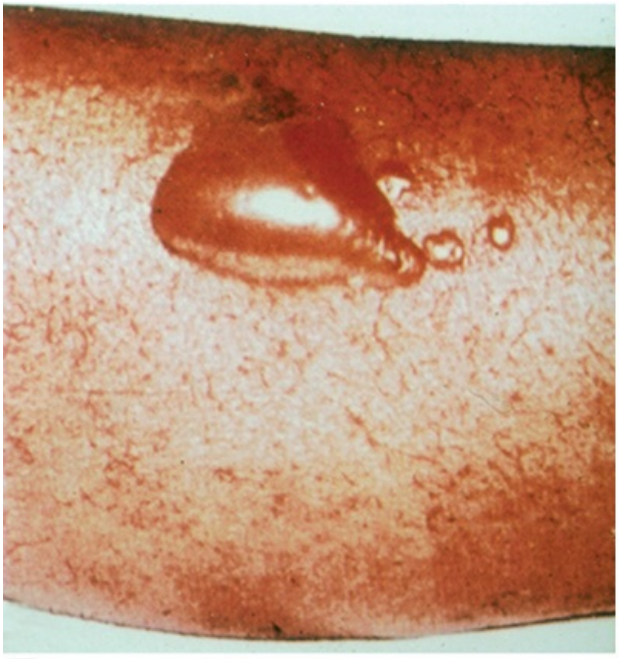

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Characteristic skin lesions in patients with proven gonococcal bacteremia. The lesions are in various stages of evolution. ***A.*** Very early petechia on finger. ***B.*** Early papular lesion, 7 mm in diameter, on lower leg. ***C.*** Pustule with central eschar resulting from early petechial lesion. ***D.*** Pustular lesion on finger. ***E.*** Mature lesion with central necrosis (black) on hemorrhagic base. ***F.*** Bullae on anterior tibial surface. *(Reprinted with permission from KK Holmes et al: Disseminated gonococcal infection. Ann Intern Med 74:979, 1971.)*

Figure 144-3

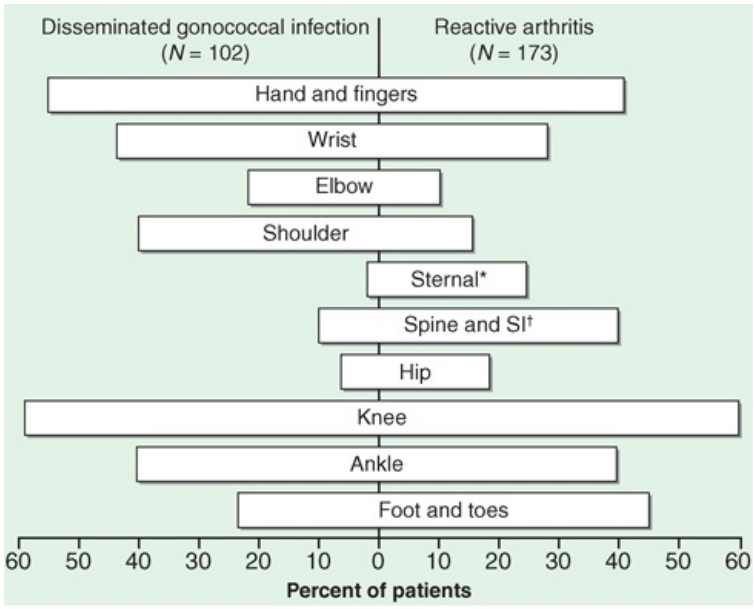

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Distributions of joints with arthritis in 102 patients with disseminated gonococcal infection and 173 patients with reactive arthritis. *Includes the sternoclavicular joints. †SI, sacroiliac joint. *(Reprinted with permission from M Kousa et al: Frequent association of chlamydial infection with Reiter's syndrome. Sex Transm Dis 5:57, 1978.)*

Suppurative arthritis involves one or two joints, most often the knees, wrists, ankles, and elbows (in decreasing order of frequency); other joints occasionally are involved. Most patients who develop gonococcal septic arthritis do so without prior polyarthralgias or skin lesions; in the absence of symptomatic genital infection, this disease cannot be distinguished from septic arthritis caused by other pathogens. The differential diagnosis of acute arthritis in young adults is discussed in Chap. 334. Rarely, osteomyelitis complicates septic arthritis involving small joints of the hand.

Gonococcal endocarditis, although rare today, was a relatively common complication of DGI in the preantibiotic era, accounting for about one-quarter of reported cases of endocarditis. Another unusual complication of DGI is meningitis.

GONOCOCCAL INFECTIONS IN HIV-INFECTED PERSONS

The association between gonorrhea and the acquisition of HIV has been demonstrated in several well-controlled studies, mainly in Kenya and Zaire. The nonulcerative STIs enhance the transmission of HIV by three- to fivefold, possibly because of increased viral shedding by persons with urethritis or cervicitis (Chap. 189). HIV has been detected by polymerase chain reaction (PCR) more commonly in ejaculates from HIV-positive men with gonococcal urethritis than in those from HIV-positive men with nongonococcal urethritis. PCR positivity diminishes by twofold after appropriate therapy for urethritis. Not only does gonorrhea enhance the transmission of HIV, it may also increase the individual's risk for acquisition of HIV. A proposed mechanism is the significantly greater number of CD4+ T lymphocytes and dendritic cells that can be infected by HIV in endocervical secretions of women with nonulcerative STIs than in those of women with ulcerative STIs.

Laboratory Diagnosis

A rapid diagnosis of gonococcal infection in men may be obtained by Gram's staining of urethral exudates (Fig. 144-1). The detection of gram-negative intracellular monococci and diplococci is usually highly specific and sensitive in diagnosing gonococcal urethritis in symptomatic males but is only ∑50% sensitive in diagnosing gonococcal cervicitis. Samples should be collected with Dacron or rayon swabs. Part of the sample should be inoculated onto a plate of modified Thayer-Martin or other gonococcal selective medium for culture. It is important to process all samples immediately because gonococci do not tolerate drying. If plates cannot be incubated immediately, they can be held safely for several hours at room temperature in candle extinction jars prior to incubation. If processing is to occur within 6 h, transport of specimens may be facilitated by the use of nonnutritive swab transport systems such as Stuart or Amies medium. For longer holding periods (e.g., when specimens for culture are to be mailed), culture media with self-contained CO_2-generating systems (such as the JEMBEC or Gono-Pak systems) may be used. Specimens should also be obtained for the diagnosis of chlamydial infection (Chap. 176).

PMNs are often seen in the endocervix on a Gram's stain, and an abnormally increased number (≥30 PMNs per field in five 1000x oil-immersion microscopic fields) establishes the presence of an

inflammatory discharge. Unfortunately, the presence or absence of gram-negative intracellular monococci or diplococci in cervical smears does not accurately predict which patients have gonorrhea, and the diagnosis in this setting should be made by culture or another suitable nonculture diagnostic method. The sensitivity of a single endocervical culture is ∑80–90%. If a history of rectal sex is elicited, a rectal wall swab (uncontaminated with feces) should be cultured. A presumptive diagnosis of gonorrhea cannot be made on the basis of gram-negative diplococci in smears from the pharynx, where other *Neisseria* species are components of the normal flora.

Increasingly, nucleic acid probe tests are being substituted for culture for the direct detection of *N. gonorrhoeae* in urogenital specimens. A common assay employs a nonisotopic chemiluminescent DNA probe that hybridizes specifically with gonococcal 16S ribosomal RNA; this assay is as sensitive as conventional culture techniques. A disadvantage of non-culture-based assays is that *N. gonorrhoeae* cannot be grown from the transport systems. Thus a culture-confirmatory test and formal antimicrobial susceptibility testing, if needed, cannot be performed. Nucleic acid amplification tests (NAATs), including Roche Amplicor, Gen-Probe APTIMA Combo2 (which also detects *Chlamydia*), and BD ProbeTec ET, offer an advantage: urine samples can be tested with a sensitivity similar to that obtained when urethral or cervical swab samples are assessed by culture and other non-NAATs.

Because of the legal implications, the preferred method for the diagnosis of gonococcal infection in children is a standardized culture. Two positive NAATs, each targeting a different nucleic acid sequence, may be substituted for culture of the cervix or the urethra as legal evidence of infection; however, cervical specimens are not recommended for prepubertal girls. Nonculture tests for gonococcal infection have not been approved by the U.S. Food and Drug Administration for use with specimens obtained from the pharynx and rectum of infected children. Cultures should be obtained from the pharynx and anus of both girls and boys, the vagina of girls, and the urethra of boys. For boys with a urethral discharge, a meatal specimen of the discharge is adequate for culture. Presumptive colonies of *N. gonorrhoeae* should be identified definitively by at least two independent methods.

Blood should be cultured in suspected cases of DGI. The use of Isolator blood culture tubes may enhance the yield. The probability of positive blood cultures decreases after 48 h of illness. Synovial fluid should be inoculated into blood culture broth medium and plated onto chocolate agar rather than selective medium because this fluid is not likely to be contaminated with commensal bacteria. Gonococci are infrequently recovered from early joint effusions containing <20,000 leukocytes/μL but may be recovered from effusions containing >80,000 leukocytes/μL. The organisms are seldom recovered from blood and synovial fluid of the same patient.

Treatment: Gonococcal Infections

Treatment failure can lead to continued transmission and the emergence of antibiotic resistance. The importance of adequate treatment with a regimen that the patient will adhere to cannot be overemphasized. Thus highly effective single-dose regimens have been developed for uncomplicated gonococcal infections. The 2010 treatment guidelines for gonococcal infections from the Centers for Disease Control and Prevention are summarized in **Table 144-1**; the recommendations for uncomplicated gonorrhea apply to HIV-infected as well as HIV-uninfected patients.

Table 144-1 Recommended Treatment for Gonococcal Infections: 2010 Guidelines of the Centers for Disease Control and Prevention

Diagnosis	Treatment of Choice[a]
Uncomplicated gonococcal infection of the cervix, urethra, pharynx[b], or rectum	
First-line regimens	Ceftriaxone (250 mg IM, single dose) *or* Cefixime (400 mg PO, single dose) *plus* Treatment for *Chlamydia* if chlamydial infection is not ruled out: Azithromycin (1 g PO, single dose) *or* Doxycycline (100 mg PO bid for 7 days)
Alternative regimens	Ceftizoxime (500 mg IM, single dose) *or* Cefotaxime (500 mg IM, single dose)

	or Spectinomycin (2 g IM, single dose)[c,d] *or* Cefotetan (1 g IM, single dose) plus probenecid (1 g PO, single dose)[c] *or* Cefoxitin (2 g IM, single dose) plus probenecid (1 g PO, single dose)[c]
Epididymitis	See Chap. 130
Pelvic inflammatory disease	See Chap. 130
Gonococcal conjunctivitis in an adult	Ceftriaxone (1 g IM, single dose)[e]
Ophthalmia neonatorum[f]	Ceftriaxone (25–50 mg/kg IV, single dose, not to exceed 125 mg)
Disseminated gonococcal infection[g]	
Initial therapy[h]	
Patient tolerant of ß-lactam drugs	Ceftriaxone (1 g IM or IV q24h; recommended) *or* Cefotaxime (1 g IV q8h) *or* Ceftizoxime (1 g IV q8h)
Patients allergic to ß-lactam drugs	Spectinomycin (2 g IM q12h)[d]
Continuation therapy	Cefixime (400 mg PO bid)
Meningitis or endocarditis	See text[i]

[a] True failure of treatment with a recommended regimen is rare and should prompt an evaluation for reinfection or consideration of an alternative diagnosis.

[b] Ceftriaxone is the only agent recommended for treatment of pharyngeal infection.

[c] Spectinomycin, cefotetan, and cefoxitin, which are alternative agents, currently are unavailable or in short supply in the United States.

[d] Spectinomycin may be ineffective for the treatment of pharyngeal gonorrhea.

[e] Plus lavage of the infected eye with saline solution (once).

[f] Prophylactic regimens are discussed in the text.

[g] Hospitalization is indicated if the diagnosis is uncertain, if the patient has frank arthritis with an effusion, or if the patient cannot be relied on to adhere to treatment.

[h] All initial regimens should be continued for 24–48 h after clinical improvement begins, at which time the switch may be made to one of the continuation regimens to complete a full week of antimicrobial treatment. Treatment for chlamydial infection (as above) should be given if this infection has not been ruled out. Fluoroquinolones may be an option if antimicrobial susceptibility can be documented by culture of the causative organism.

[i] Hospitalization is indicated to exclude suspected meningitis or endocarditis.

Single-dose regimens of the third-generation cephalosporins ceftriaxone (given IM) and cefixime (given orally) currently are the mainstays of therapy for uncomplicated gonococcal infection of the urethra, cervix, rectum, or pharynx and almost always result in an effective cure. Outside the United States, cefixime has been associated with rare treatment failures caused by strains of *N. gonorrhoeae* with elevated MICs of third-generation cephalosporins. Quinolone-containing regimens are no longer recommended in the United States as first-line treatment because of widespread resistance to these agents.

Because co-infection with *C. trachomatis* occurs frequently, initial treatment regimens must also incorporate an agent (e.g., azithromycin or doxycycline) that is effective against chlamydial infection. Pregnant women with gonorrhea, who should not take doxycycline, should receive concurrent treatment with a macrolide antibiotic for possible chlamydial infection. A single 1-g dose of azithromycin, which is effective therapy for uncomplicated chlamydial infections, results in an

unacceptably low cure rate (93%) for gonococcal infections and should not be used alone. A single 2-g dose of azithromycin, particularly in the extended-release microsphere formulation, delivers azithromycin to the lower gastrointestinal tract, thereby improving tolerability. Azithromycin is effective against sensitive strains, but this drug is expensive, causes gastrointestinal distress, and is not recommended for routine or first-line treatment of gonorrhea. Spectinomycin has been used as an alternative regimen for the treatment of uncomplicated gonococcal infections in penicillin-allergic persons outside the United States but is not currently available in this country. Of note, the limited effectiveness of spectinomycin for the treatment of pharyngeal infection reduces its utility in populations among whom such infection is common, such as MSM.

Persons with uncomplicated infections who receive a recommended regimen do not need a test of cure. Cultures for *N. gonorrhoeae* should be performed if symptoms persist after therapy with an established regimen, and any gonococci isolated should be tested for antimicrobial susceptibility.

Symptomatic gonococcal pharyngitis is more difficult to eradicate than genital infection. Persons who cannot tolerate cephalosporins and those in whom quinolones are contraindicated may be treated with spectinomycin if it is available, but this agent results in a cure rate of ≳52%. Persons given spectinomycin should have a pharyngeal sample cultured 3–5 days after treatment as a test of cure. A single 2-g dose of azithromycin may be used in areas where rates of resistance to azithromycin are low.

Treatments for gonococcal epididymitis and PID are discussed in Chap. 130. Ocular gonococcal infections in older children and adults should be managed with a single dose of ceftriaxone combined with saline irrigation of the conjunctivae (both undertaken expeditiously), and patients should undergo a careful ophthalmologic evaluation that includes a slit-lamp examination.

DGI may require higher dosages and longer durations of therapy (Table 144-1). Hospitalization is indicated if the diagnosis is uncertain, if the patient has localized joint disease that requires aspiration, or if the patient cannot be relied on to comply with treatment. Open drainage is necessary only occasionally—e.g., for management of hip infections that may be difficult to drain percutaneously. Nonsteroidal anti-inflammatory agents may be indicated to alleviate pain and hasten clinical improvement of affected joints. Gonococcal meningitis and endocarditis should be treated in the hospital with high-dose IV ceftriaxone (1–2 g every 12 h); therapy should continue for 10–14 days for meningitis and for at least 4 weeks for endocarditis. All persons who experience more than one episode of DGI should be evaluated for complement deficiency.

Prevention and Control

Condoms, if properly used, provide effective protection against the transmission and acquisition of gonorrhea as well as other infections that are transmitted to and from genital mucosal surfaces. Spermicidal preparations used with a diaphragm or cervical sponges impregnated with nonoxynol 9 offer some protection against gonorrhea and chlamydial infection. However, the frequent use of preparations that contain nonoxynol 9 is associated with mucosal disruption that paradoxically may enhance the risk of HIV infection in the event of exposure. All patients should be instructed to refer sex partners for evaluation and treatment. All sex partners of persons with gonorrhea should be evaluated and treated for *N. gonorrhoeae* and *C. trachomatis* infections if their last contact with the patient took place within 60 days before the onset of symptoms or the diagnosis of infection in the patient. If the patient's last sexual encounter was >60 days before onset of symptoms or diagnosis, the patient's most recent sex partner should be treated. Partner-delivered medications or prescriptions for medications to treat gonorrhea and chlamydial infection diminish the likelihood of reinfection (or relapse) in the infected patient. In states where it is legal, this approach is an option for partner management. Patients should be instructed to abstain from sexual intercourse until therapy is completed and until they and their sex partners no longer have symptoms. Greater emphasis must be placed on prevention by public health education, individual patient counseling, and behavior modification. Sexually active persons, especially adolescents, should be offered screening for STIs. For males, a NAAT on urine or a urethral swab may be used for screening. Preventing the spread of gonorrhea may help reduce the transmission of HIV. No effective vaccine for gonorrhea is yet available, but efforts to test several candidates are under way.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Dr. King K. Holmes and Dr. Stephen A. Morse to the chapter on this subject in earlier editions.

FURTHER READINGS

Centers for Disease Control and Prevention: Gonococcal Isolate Surveillance Project (GISP); *www.cdc.gov/std/GISP/*

———: Sexually transmitted diseases treatment guidelines, 2010. MMWR Recomm Rep 59(RR-12):49, 2010

Gaydos CA: Nucleic acid amplification tests for gonorrhea and *Chlamydia*: Practice and applications. Infect Dis Clin North Am 19:367, 2005[PMID: 15963877]

Golden MR et al: Effect of expedited treatment of sex partners on recurrent or persistent gonorrhea or chlamydial infections. N Engl J Med 352:676, 2005[PMID: 15716561]

Hook EW III, Holmes KK: Gonococcal infections. Ann Intern Med 102:229, 1985[PMID: 3917638]

Laga M et al: Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: Results from a cohort study. AIDS 7:95, 1993[PMID: 8442924]

O'Brien JP et al: Disseminated gonococcal infection: A prospective analysis of 49 patients and a review of pathophysiology and immune mechanisms. Medicine (Baltimore) 62:395, 1983[PMID: 21698803]

Workowski KA et al: Emerging antimicrobial resistance in *Neisseria gonorrhoeae*: Urgent need to strengthen prevention strategies. Ann Intern Med 148:606, 2008[PMID: 18413622]

Zhao S et al: Genetics of chromosomally mediated intermediate resistance to ceftriaxone and cefixime in *Neisseria gonorrhoeae*. Antimicrob Agents Chemother 53:3744, 2009[PMID: 19528266]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HAEMOPHILUS INFLUENZAE

Microbiology

Haemophilus influenzae was first recognized in 1892 by Pfeiffer, who erroneously concluded that the bacterium was the cause of influenza. The bacterium is a small (1- by 0.3-μm) gram-negative organism of variable shape; hence, it is often described as a pleomorphic coccobacillus. In clinical specimens such as cerebrospinal fluid (CSF) and sputum, it frequently stains only faintly with safranin and therefore can easily be overlooked.

H. influenzae grows both aerobically and anaerobically. Its aerobic growth requires two factors: hemin (X factor) and nicotinamide adenine dinucleotide (V factor). These requirements are used in the clinical laboratory to identify the bacterium. Caution must be used to distinguish *H. influenzae* from *H. haemolyticus*, a respiratory tract commensal that has identical growth requirements. *H. haemolyticus* has classically been distinguished from *H. influenzae* by hemolysis on horse blood agar. However, a significant proportion of isolates of *H. haemolyticus* have now been recognized as nonhemolytic. Analysis of 16S ribosomal sequences is one reliable method to distinguish these two species.

Six major serotypes of *H. influenzae* have been identified; designated *a* through *f*, they are based on antigenically distinct polysaccharide capsules. In addition, some strains lack a polysaccharide capsule and are referred to as *nontypable* strains. Type b and nontypable strains are the most relevant strains clinically **(Table 145-1)**, although encapsulated strains other than type b can cause disease. *H. influenzae* was the first free-living organism to have its entire genome sequenced.

Table 145-1 Characteristics of Type b and Nontypable Strains of *Haemophilus Influenzae*

Feature	Type b Strains	Nontypable Strains
Capsule	Ribosyl-ribitol phosphate	Unencapsulated
Pathogenesis	Invasive infections due to hemato-genous spread	Mucosal infections due to contiguous spread
Clinical manifestations	Meningitis and invasive infections in incompletely immunized infants and children	Otitis media in infants and children; lower respiratory tract infections in adults with chronic bronchitis
Evolutionary history	Basically clonal	Genetically diverse
Vaccine	Highly effective conjugate vaccines	None available; under development

The antigenically distinct type b capsule is a linear polymercomposed of ribosyl-ribitol phosphate. Strains of *H. influenzae* type b (Hib) cause disease primarily in infants and children <6 years of age. Nontypable strains are primarily mucosal pathogens but occasionally cause invasive disease.

Epidemiology and Transmission

H. influenzae, an exclusively human pathogen, is spread by airborne droplets or by direct contact with secretions or fomites. Colonization with nontypable *H. influenzae* is a dynamic process; new strains are acquired and other strains are replaced periodically.

The widespread use of Hib conjugate vaccines in many industrialized countries has resulted in striking decreases in the rate of nasopharyngeal colonization by Hib and in the incidence of Hib infection **(Fig. 145-1)**. However, the majority of the world's children remain unimmunized. Worldwide, invasive Hib disease occurs predominantly in unimmunized children and in those who have not completed the primary immunization series. Certain groups have a higher incidence of invasive Hib disease than the general population, including black children and Native American groups. Although this increased incidence has not yet been accounted for, several factors may be relevant, including age at exposure to the bacterium, socioeconomic conditions, and genetic differences.

Figure 145-1

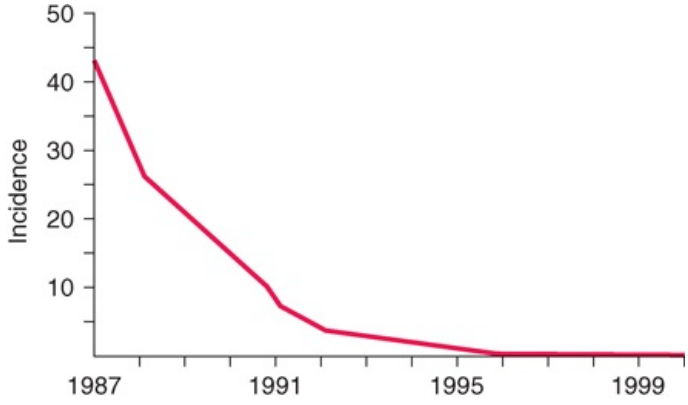

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated incidence (rate per 100,000) of invasive disease due to *Haemophilus influenzae* type b among children <5 years of age: 1987–2000. Fewer than 40 cases per year have been reported since 2000. *(Data from the Centers for Disease Control and Prevention.)*

Pathogenesis

Hib strains cause systemic disease by invasion and hematogenous spread from the respiratory tract to distant sites such as the meninges, bones, and joints. The type b polysaccharide capsule is an important virulence factor affecting the bacterium's ability to avoid opsonization and cause systemic disease.

Nontypable strains cause disease by local invasion of mucosal surfaces. Otitis media results when bacteria reach the middle ear by way of the eustachian tube. Adults with chronic bronchitis experience recurrent lower respiratory tract infection due to nontypable strains. In addition, persistent nontypable *H. influenzae* colonization of the lower airways of adults with chronic obstructive pulmonary disease (COPD) contributes to the airway inflammation that is a hallmark of the disease. Nontypable strains that cause infection in adults with COPD differ in pathogenic potential and genome content from strains that cause otitis media. The incidence of invasive disease caused by nontypable strains is low. Strains that cause invasive disease are genetically and phenotypically diverse.

Immune Response

Antibody to the capsule is important in protection from infection by Hib strains. The level of (maternally acquired) serum antibody to the capsular polysaccharide, which is a polymer of polyribitol ribose phosphate (PRP), declines from birth to 6 months of age and, in the absence of vaccination, remains low until ~2 or 3 years of age. The age at the antibody nadir correlates with that of the peak incidence of type b disease. Antibody to PRP then appears partly as a result of exposure to Hib or cross-reacting antigens. Systemic Hib disease is unusual after the age of 6 years because of the presence of protective antibody. Vaccines in which PRP is conjugated to protein carrier molecules have been developed and are now used widely. These vaccines generate an antibody response to PRP in infants and effectively prevent invasive infections in infants and children.

Since nontypable strains lack a capsule, the immune response to infection is directed at noncapsular antigens. These antigens have generated considerable interest as immune targets and potential vaccine components. The human immune response to nontypable strains appears to be strain-specific, accounting in part for the propensity of these strains to cause recurrent otitis media and recurrent exacerbations of chronic bronchitis in immunocompetent hosts.

Clinical Manifestations

HIB

The most serious manifestation of infection with Hib is *meningitis* (Chap. 381), which primarily affects children <2 years of age. The clinical manifestations of Hib meningitis are similar to those of meningitis caused by other bacterial pathogens. Fever and altered central nervous system function are the most common features at presentation. Nuchal rigidity may or may not be evident. Subdural effusion, the most common complication, is suspected when, despite 2 or 3 days of appropriate antibiotic therapy, the infant has seizures, hemiparesis, or continued obtundation. The overall mortality rate from Hib meningitis is ~5%, and the morbidity rate is high. Of survivors, 6% have permanent sensorineural hearing loss, and about one-fourth have a significant handicap of some type. If more subtle handicaps are sought, up to half of survivors are found to have some neurologic sequelae, such as partial hearing loss and delayed language development.

Epiglottitis (Chap. 31) is a life-threatening Hib infection involving cellulitis of the epiglottis and supraglottic tissues. It can lead to acute upper airway obstruction. Its unique epidemiologic features are its occurrence in an older age group (2–7 years old) than other Hib infections and its absence

among Navajo Indians and Alaskan Eskimos. Sore throat and fever rapidly progress to dysphagia, drooling, and airway obstruction. Epiglottitis also occurs in adults.

Cellulitis (Chap. 125) due to Hib occurs in young children. The most common location is on the head or neck, and the involved area sometimes takes on a characteristic bluish-red color. Most patients have bacteremia, and 10% have an additional focus of infection.

Hib causes *pneumonia* in infants. The infection is clinically indistinguishable from other types of bacterial pneumonia (e.g., pneumococcal pneumonia) except that Hib is more likely to involve the pleura.

Several less common invasive conditions can be important clinical manifestations of Hib infection in children. These include osteomyelitis, septic arthritis, pericarditis, orbital cellulitis, endophthalmitis, urinary tract infection, abscesses, and bacteremia without an identifiable focus.

NONTYPABLE *H. INFLUENZAE*

Nontypable *H. influenzae* is the most common bacterial cause of exacerbations of COPD; these exacerbations are characterized by increased cough, sputum production, and shortness of breath. Fever is low-grade, and no infiltrates are evident on chest x-ray. Nontypable strains also cause community-acquired bacterial pneumonia in adults, especially among patients with COPD or AIDS. The clinical features of *H. influenzae* pneumonia are similar to those of other types of bacterial pneumonia (including pneumococcal pneumonia).

Nontypable *H. influenzae* is one of the three most common causes of childhood otitis media (the other two being *Streptococcus pneumoniae* and *Moraxella catarrhalis*) (Chap. 31). Infants are febrile and irritable, while older children report ear pain. Symptoms of viral upper respiratory infection often precede otitis media. The diagnosis is made by pneumatic otoscopy. An etiologic diagnosis, although not routinely sought, can be established by tympanocentesis and culture of middle-ear fluid. Clinical features associated with *H. influenzae* otitis media include a history of recurrent episodes, treatment failure, concomitant conjunctivitis, bilateral otitis media, and recent antimicrobial therapy. The increasing use of pneumococcal polysaccharide conjugate vaccines in infants is resulting in a relative increase in the proportion of otitis media cases that are caused by *H. influenzae*.

Nontypable *H. influenzae* also causes puerperal sepsis and is an important cause of neonatalbacteremia. These nontypable strains, which are closely related to *H. haemolyticus*, tend to be of biotype IV and cause invasive disease after colonizing the female genital tract.

Nontypable *H. influenzae* causes sinusitis (Chap. 31) in adults and children. In addition, the bacterium is a less common cause of various invasive infections. These infections include empyema, adult epiglottitis, pericarditis, cellulitis, septic arthritis, osteomyelitis, endocarditis, cholecystitis, intraabdominal infections, urinary tract infections, mastoiditis, aortic graft infection, and bacteremia without a detectable focus. Most *H. influenzae* invasive infections in countries where Hib vaccines are used widely are caused by nontypable strains. Many patients with *H. influenzae* bacteremia have an underlying condition, such as HIV infection, cardiopulmonary disease, alcoholism, or cancer.

Diagnosis

The most reliable method for establishing a diagnosis of Hib infection is recovery of the organism in culture. The presence of gram-negative coccobacilli in Gram-stained CSF is strong evidence for Hib meningitis. Recovery of the organism from CSF confirms the diagnosis. Cultures of other normally sterile body fluids, such as blood, joint fluid, pleural fluid, pericardial fluid, and subdural effusion, are confirmatory in other infections.

Detection of PRP is an important adjunct to culture in rapid diagnosis of Hib meningitis. Immunoelectrophoresis, latex agglutination, coagglutination, and enzyme-linked immunosorbent assay are effective in detecting PRP. These assays are particularly helpful when patients have received prior antimicrobial therapy and thus are especially likely to have negative cultures.

Because nontypable *H. influenzae* is primarily a mucosal pathogen, it is a component of a mixed flora; thus etiologic diagnosis is challenging. Nontypable *H. influenzae* infection is strongly suggested by the predominance of gram-negative coccobacilli among abundant polymorphonuclear leukocytes in a Gram-stained sputum specimen from a patient in whom pneumonia is suspected. Although bacteremia is detectable in a small proportion of patients with pneumonia due to nontypable *H. influenzae*, most such patients have negative blood cultures.

A diagnosis of otitis media is based on the detection by pneumatic otoscopy of fluid in the middle ear. An etiologic diagnosis requires tympanocentesis but is not routinely sought. An invasive procedure is also required to determine the etiology of sinusitis; thus, treatment is often empirical once the diagnosis is suspected in light of clinical symptoms and sinus radiographs.

Treatment: Haemophilus Influenzae

Initial therapy for meningitis due to Hib should consist of a cephalosporin such as ceftriaxone or cefotaxime. For children, the dosage of ceftriaxone is 75–100 mg/kg daily given in two doses 12 h

apart. The pediatric dosage of cefotaxime is 200 mg/kg daily given in four doses 6 h apart. Adult dosages are 2 g every 12 h for ceftriaxone and 2 g every 4–6 h for cefotaxime. An alternative regimen for initial therapy is ampicillin (200–300 mg/kg daily in four divided doses) plus chloramphenicol (75 –100 mg/kg daily in four divided doses). Therapy should continue for a total of 1–2 weeks.

Administration of glucocorticoids to patients with Hib meningitis reduces the incidence of neurologic sequelae. The presumed mechanism is reduction of the inflammation induced by bacterial cell-wall mediators of inflammation when cells are killed by antimicrobial agents. Dexamethasone (0.6 mg/kg per day intravenously in four divided doses for 2 days) is recommended for the treatment of Hib meningitis in children >2 months of age.

Invasive infections other than meningitis are treated with the same antimicrobial agents. For epiglottitis, the dosage ofceftriaxone is 50 mg/kg daily, and the dosage of cefotaxime is150 mg/kg daily, given in three divided doses 8 h apart. Epiglottitis constitutes a medical emergency, and maintenance of an airway is critical. The duration of therapy is determined by the clinical response. A course of 1–2 weeks is usually appropriate.

Many infections caused by nontypable strains of *H. influenzae*, such as otitis media, sinusitis, and exacerbations of COPD, can be treated with oral antimicrobial agents. Approximately 20–35% of nontypable strains produce β-lactamase (with the exact proportion depending on geographic location), and these strains are resistant to ampicillin. Several agents have excellent activity against nontypable *H. influenzae*, including amoxicillin/clavulanic acid, various extended-spectrum cephalosporins, and the macrolides azithromycin and clarithromycin. Fluoroquinolones are highly active against *H. influenzae* and are useful in adults with exacerbations of COPD. However, fluoroquinolones are not currently recommended for the treatment of children or pregnant women because of possible effects on articular cartilage.

In addition to β-lactamase production, alteration of penicillin-binding proteins—a second mechanism of ampicillin resistance—has been detected in isolates of *H. influenzae*. Although rare in the United States, these β-lactamase– negative ampicillin-resistant strains are increasing in prevalence in Europe and Japan. Continued monitoring of the evolving antimicrobial susceptibility patterns of *H. influenzae* will be important.

Prevention

VACCINATION

(See also Chap. 122) Two conjugate vaccines that prevent invasive infections with Hib in infants and children are licensed in the United States. In addition to eliciting protective antibody, these vaccines prevent disease by reducing rates of pharyngeal colonization with Hib. The widespread use of conjugate vaccines has dramatically reduced the incidence of Hib disease in developed countries. Even though the manufacture of Hib vaccines is costly, vaccination is cost-effective. The Global Alliance for Vaccines and Immunizations has recognized the underuse of Hib conjugate vaccines. The disease burden has been reduced in developing countries that have implemented routine vaccination (e.g., The Gambia, Chile). An important obstacle to more widespread vaccination is the lack of data on the epidemiology and burden of Hib disease in many developing countries.

All children should be immunized with an Hib conjugate vaccine, receiving the first dose at ~2 months of age, the rest of the primary series at 2–6 months of age, and a booster dose at 12–15 months of age. Specific recommendations vary for the different conjugate vaccines. The reader is referred to the recommendations of the American Academy of Pediatrics (Chap. 122 and www.cispimmunize.org).

Currently, no vaccines are available for the prevention of disease caused by nontypable *H. influenzae*. However, a vaccine that contains a surface protein of *H. influenzae* conjugated to pneumococcal polysaccharides has shown partial efficacy in preventing *H. influenzae* otitis media. Additional progress in the development of vaccines against nontypable *H. influenzae* is anticipated.

CHEMOPROPHYLAXIS

The risk of secondary disease is greater than normal among household contacts of patients with Hib disease. Therefore, all children and adults (except pregnant women) in households with at least one incompletely immunized contact <4 years of age should receive prophylaxis with oral rifampin. When two or more cases of invasive Hib disease have occurred within 60 days at a child-care facility attended by incompletely vaccinated children, administration of rifampin to all attendees and personnel is indicated, as is recommended for household contacts. Chemoprophylaxis is not indicated in nursery and child-care contacts of a single index case. The reader is referred to the recommendations of the American Academy of Pediatrics.

HAEMOPHILUS DUCREYI

Haemophilus ducreyi is the etiologic agent of chancroid (Chap. 130), a sexually transmitted disease characterized by genital ulceration and inguinal adenitis. *H. ducreyi* poses a significant health problem in developing countries. In addition to being a cause of morbidity in itself, chancroid is associated with HIV infection because of the role played by genital ulceration in HIV transmission. Chancroid increases

both the efficiency of, transmission of, and the degree of susceptibility to HIV infection.

Microbiology

H. ducreyi is a highly fastidious coccobacillary gram-negative bacterium whose growth requires X factor (hemin). Although, in light of this requirement, the bacterium has been classified in the genus *Haemophilus*, DNA homology and chemotaxonomic studies have established substantial differences between *H. ducreyi* and other *Haemophilus* species. Taxonomic reclassification of the organism is likely in the future but awaits further study. Ulcers contain predominantly T cells. The fact that patients who have had chancroid may have repeated infections indicates that infection does not confer protection.

Epidemiology and Prevalence

Chancroid is a common cause of genital ulcers in developing countries. In the United States, several large outbreaks have occurred since 1981. Recurring epidemiologic themes have been apparent in these outbreaks: (1) transmission has been predominantly heterosexual; (2) males have outnumbered females by ratios of 3:1 to 25:1; (3) prostitutes have been important in transmission of the infection; and (4) chancroid has been strongly associated with illicit drug use. The annual number of cases reported in the United States has remained stable since 2000.

Clinical Manifestations

Infection is acquired as the result of a break in the epithelium during sexual contact with an infected individual. After an incubation period of 4–7 days, the initial lesion—a papule with surrounding erythema—appears. In 2 or 3 days, the papule evolves into a pustule, which spontaneously ruptures and forms a sharply circumscribed ulcer that is generally not indurated **(Fig. 145-2)**. The ulcers are painful and bleed easily; little or no inflammation of the surrounding skin is evident. Approximately half of patients develop enlarged, tender inguinal lymph nodes, which frequently become fluctuant and spontaneously rupture. Patients usually seek medical care after 1–3 weeks of painful symptoms.

Figure 145-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chancroid with characteristic penile ulcers and associated left inguinal adenitis (bubo).

The presentation of chancroid does not usually include all of the typical clinical features and is sometimes atypical. Multiple ulcers can coalesce to form giant ulcers. Ulcers can appear and then resolve, with inguinal adenitis (Fig. 145-2) and suppuration following 1–3 weeks later; this clinical picture can be confused with that of lymphogranuloma venereum (Chap. 176). Multiple small ulcers can resemble folliculitis. Other differential diagnostic considerations include the various infections causing genital ulceration, such as primary syphilis, secondary syphilis (condyloma latum), genital herpes, and donovanosis. In rare cases, chancroid lesions become secondarily infected with bacteria; the result is extensive inflammation.

Diagnosis

Clinical diagnosis of chancroid is often inaccurate, and laboratory confirmation should be attempted in suspected cases. An accurate diagnosis of chancroid relies on culture of *H. ducreyi* from the lesion. In addition, aspiration and culture of suppurative lymph nodes should be considered. Since the organism can be difficult to grow, the use of selective and supplemented media is necessary. A multiplex polymerase chain reaction assay has been developed for simultaneous amplification of DNA targets from *H. ducreyi*, *Treponema pallidum*, and herpes simplex virus types 1 and 2. When this assay becomes commercially available, it will be a useful diagnostic tool with which to identify the etiology of genital ulcers.

Treatment: *Haemophilus Ducreyi*

Treatment regimens recommended by the Centers for Disease Control and Prevention include (1) a single 1-g oral dose of azithromycin; (2) ceftriaxone (250 mg intramuscularly in a single dose); (3) ciprofloxacin (500 mg by mouth twice a day for 3 days); and (4) erythromycin base (500 mg by mouth three times a day for 7 days). Isolates from patients who do not respond promptly to treatment should be tested for antimicrobial resistance. In patients with HIV infection, healing may be slow and longer courses of treatment may be necessary. Clinical treatment failure in HIV-seropositive patients may reflect co-infection, especially with herpes simplex virus. Contacts of patients with chancroid should be identified and treated, whether or not symptoms are present, if they have had sexual contact with the patient during the 10 days preceding the patient's onset of symptoms.

MORAXELLA CATARRHALIS

Microbiology

M. catarrhalis is an unencapsulated gram-negative diplococcus the ecologic of which niche is the human respiratory tract. The organism was initially designated *Micrococcus catarrhalis*, but its name was changed to *Neisseria catarrhalis* in 1970 because of phenotypic similarities to commensal *Neisseria* species. On the basis of more rigorous analysis of genetic relatedness, *Moraxella catarrhalis* is now the widely accepted name for this species.

Epidemiology

Nasopharyngeal colonization by *M. catarrhalis* is common in infancy, with colonization rates ranging between 33% and 100% and depending on geographic location. Several factors probably account for this geographic variation, including living conditions, day-care attendance, hygiene, household smoking, and population genetics. The prevalence of colonization decreases steadily with age.

The widespread use of pneumococcal conjugate vaccines in some countries has resulted in alterations in patterns of nasopharyngeal colonization in resident populations. A relative increase in colonization by nonvaccine pneumococcal serotypes, nontypable *H. influenzae*, and *M. catarrhalis* has occurred. These changes in colonization patterns may be altering the distribution of pathogens of both otitis media and sinusitis in children.

Pathogenesis

M. catarrhalis causes mucosal infections of the respiratory tract. Strains exhibit substantial genetic diversity and differences in virulence properties. The species is composed of two distinct genetic lineages; the complement-resistant lineage is more strongly associated with virulence than are complement-sensitive strains.

The expression of several adhesin molecules with differing specificities for various host cell receptors reflects the importance of adherence to the respiratory epithelial surface in the pathogenesis of infection. *M. catarrhalis* invades multiple cell types. Its intracellular residence in lymphoid tissue provides a potential reservoir for persistence in the human respiratory tract.

Clinical Manifestations

In children, *M. catarrhalis* causes predominantly mucosal infections when the bacterium migrates from the nasopharynx to the middle ear or the sinuses (Chap. 31). The inciting event for both otitis media and sinusitis is often a preceding viral infection. Overall, cultures of middle-ear fluid obtained by tympanocentesis indicate that *M. catarrhalis* causes 15–20% of cases of acute otitis media. Acute otitis media caused by *M. catarrhalis* or nontypable *H. influenzae* is clinically milder than otitis media caused by *S. pneumoniae*, with less fever and a lower prevalence of a red bulging tympanic membrane. However, substantial overlap makes it impossible to predict etiology in an individual child on the basis of clinical features.

A small proportion of viral upper respiratory tract infections are complicated by bacterial sinusitis. Cultures of sinus puncture aspirates show that *M. catarrhalis* accounts for ~20% of cases of acute bacterial sinusitis in children and for a smaller proportion in adults.

M. catarrhalis is a common cause of exacerbations in adults with COPD. The bacterium has been overlooked in this clinical setting because it has long been considered to be a commensal and

because it is easily mistaken for commensal *Neisseria* species in cultures of respiratory secretions (see "Diagnosis," below). Several independent lines of evidence have established *M. catarrhalis* as a pathogen in COPD. These include (1) the demonstration of *M. catarrhalis* in the lower airways during exacerbations, (2) the association of exacerbation with acquisition of new strains, (3) elevations of inflammatory markers in association with *M. catarrhalis*, and (4) the development of specific immune responses following infection. *M. catarrhalis* is the second most common bacterial cause of COPD exacerbations (after *H. influenzae*), as shown in a 10-year prospective study; the distribution of new-strain acquisitions is shown in **Fig. 145-3**. Not included are culture-negative cases or cases from which a pathogen had been previously isolated. With the application of rigorous clinical criteria for defining the etiology of exacerbations (both culture-positive and culture-negative), ~10% of all exacerbations in the same study were caused by *M. catarrhalis*. The clinical features of an exacerbation due to *M. catarrhalis* are similar to those of exacerbations due to other bacterial pathogens, including *H. influenzae* and *S. pneumoniae*. The cardinal symptoms are cough with increased sputum production, sputum purulence, and dyspnea in comparison with baseline symptoms.

Figure 145-3

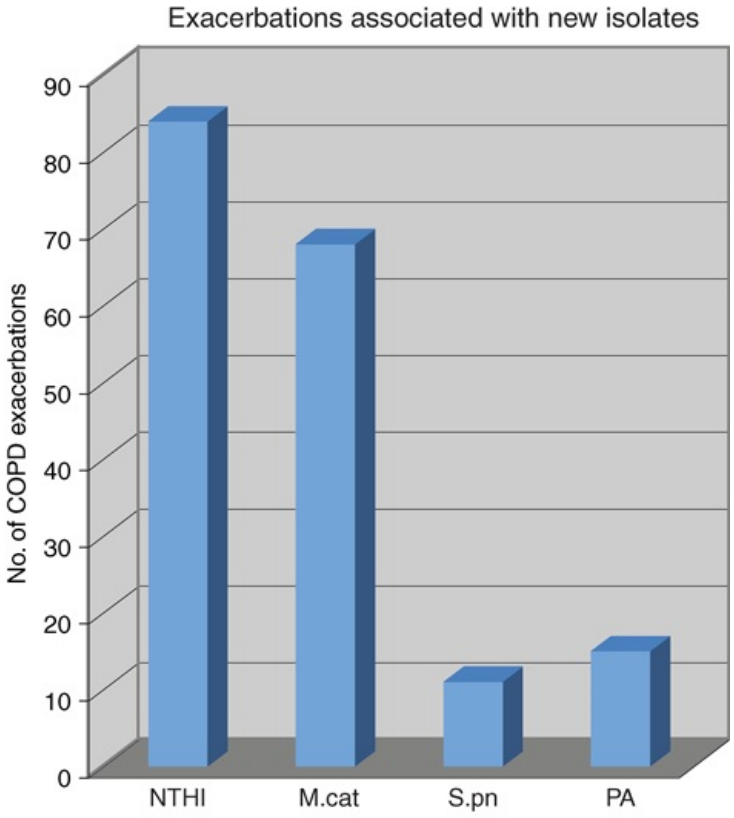

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cumulative results of a prospective study (1994–2004) of bacterial infection in COPD showing etiology of exacerbations. The numbers of exacerbations shown indicate the acquisition of a new strain simultaneous with clinical symptoms of an exacerbation. NTHI, nontypable *H. influenzae*; M.cat, *M. catarrhalis*; S.pn, *Streptococcus pneumoniae*; PA, *Pseudomonas aeruginosa*. *(Adapted from Murphy and Parameswaran, with permission.© 2009 Infectious Diseases Society of America.)*

Pneumonia due to *M. catarrhalis* occurs in the elderly, particularly in the setting of underlying cardiopulmonary disease, but is infrequent. Invasive infections, such as bacteremia, endocarditis, neonatalmeningitis, and septic arthritis, are rare.

Diagnosis

Tympanocentesis is required for etiologic diagnosis of otitis media, but this procedure is not performed routinely. Therefore, treatment of otitis media is generally empirical. Similarly, an etiologic diagnosis of sinusitis requires an invasive procedure and thus is usually not available to the clinician. Isolation of *M. catarrhalis* from an expectorated sputum sample from an adult experiencing clinical symptoms of an exacerbation is suggestive, but not diagnostic, of *M. catarrhalis* as the cause.

Upon culture, colonies of *M. catarrhalis* resemble commensal neisseriae that are part of the normal upper airway flora. As mentioned above, the difficulty in distinguishing colonies of *M. catarrhalis* from neisserial colonies in cultures of respiratory secretions explains in part why *M. catarrhalis* has been

overlooked as a pathogen. In contrast to these *Neisseria* species, *M. catarrhalis* colonies can be slid across the agar surface without disruption (the "hockey puck sign"). In addition, after 48 h of growth, *M. catarrhalis* colonies take on a pink color and tend to be larger than neisserial colonies. A variety of biochemical tests can distinguish M. catarrhalis from neisseriae. Kits that rely on these biochemical reactions are commercially available.

Treatment: *Moraxella Catarrhalis*

M. catarrhalis rapidly acquired β-lactamases during the 1970s and 1980s; antimicrobial susceptibility patterns have remained relatively stable since that time, with >90% of strains now producing β-lactamase and thus resistant to amoxicillin. Otitis media in children and exacerbations of COPD in adults are generally managed empirically with antimicrobial agents that are active against *S. pneumoniae, H. influenzae*, and *M. catarrhalis*. Most strains of *M. catarrhalis* are susceptible to amoxicillin/clavulanic acid, extended-spectrum cephalosporins, newermacrolides (azithromycin, clarithromycin), trimethoprim-sulfamethoxazole, and fluoroquinolones.

FURTHER READINGS

Broides A et al: Acute otitis media caused by Moraxella catarrhalis: Epidemiologic and clinical characteristics. Clin Infect Dis 49:1641, 2009[PMID: 19886799]

Dworkin MS et al: The changing epidemiology of invasive *Haemophilus influenzae* disease, especially in persons > or = 65 years old. Clin Infect Dis 44:810, 2007[PMID: 17304452]

Mawas F et al: Current progress with *Moraxella catarrhalis* antigens as vaccine candidates. Expert Rev Vaccines 8:77, 2009[PMID: 19093775]

Mohammed TT, Olumide YM: Chancroid and human immunodeficiency virus infection—a review. Int J Dermatol 47:1, 2008[PMID: 18173591]

Murphy TF et al: Nontypeable *Haemophilus influenzae* as a pathogen in children. Pediatr Infect Dis J 28:43, 2009[PMID: 19057458]

——— et al: *Haemophilus haemolyticus*: A human respiratory tract commensal to be distinguished from *Haemophilus influenzae*.J Infect Dis 195:81, 2007

Murphy TF, Parameswaran GI: *Moraxella catarrhalis*, a human respiratory tract pathogen. Clin Infect Dis 49:124, 2009[PMID: 19480579]

Prymula R et al: Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both *Streptococcus pneumoniae* and non-typable *Haemophilus influenzae*: A randomised double-blind efficacy study. Lancet 367:740, 2006[PMID: 16517274]

Sethi S, Murphy TF: Infection in the pathogenesis and course of chronic obstructive pulmonary disease. N Engl J Med 359:2355, 2008[PMID: 19038881]

Wirth T et al: The rise and spread of a new pathogen: Seroresistant *Moraxella catarrhalis*. Genome Res 17:1647, 2007[PMID: 17895425]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE HACEK GROUP

HACEK organisms are a group of fastidious, slow-growing, gram-negative bacteria the growth of which requires an atmosphere of carbon dioxide. Species belonging to this group include several ***H****aemophilus* species, ***A****ggregatibacter* (formerly ***A****ctinobacillus*) *actinomycetemcomitans*, ***C****ardiobacterium hominis*, ***E****ikenella corrodens*, and ***K****ingella kingae*. HACEK bacteria normally reside in the oral cavity and have been associated with local infections in the mouth. They are also known to cause severe systemic infections—most often bacterial endocarditis, which can develop on either native or prosthetic valves (Chap. 124).

HACEK Endocarditis

In large series, up to 3% of cases of infective endocarditis are attributable to HACEK organisms, most often *A. actinomycetemcomitans, Haemophilus* species, and *C. hominis.* Invasive infection typically occurs in patients with a history of cardiac valvular disease, often in the setting of a recent dental procedure, nasopharyngeal infection, or tongue piercing or scraping. The aortic and mitral valves are most commonly affected. The clinical course of HACEK endocarditis tends to be subacute; however, embolization is common. The overall prevalence of major emboli associated with HACEK endocarditis ranges from 28% to 71% in different series. On echocardiography, valvular vegetations are seen in up to 85% of patients. The vegetations are frequently large, although vegetation size has not been directly correlated with the risk of embolization. Cultures of blood from patients with suspected HACEK endocarditis may require up to 30 days to become positive, and the microbiology laboratory should be alerted when a HACEK organism is being considered. However, most cultures that ultimately yield a HACEK organism become positive within the first week, especially with improved culture systems such as BACTEC. In addition, polymerase chain reaction techniques (e.g., of cardiac valves) are facilitating the diagnosis of HACEK infections. Because of the organisms& slow growth, antimicrobial testing may be difficult, and β-lactamase production may not be detected. E-test methodology may increase the accuracy of susceptibility testing.

HAEMOPHILUS SPECIES

Haemophilus species are differentiated by their in vitro growth requirements for X factor (hemin) and V factor (nicotinamide adenine dinucleotide). *H. aphrophilus* requires only X factor for growth, while species designated *para-* require only V factor. *H. aphrophilus* and *H. parainfluenzae* are the *Haemophilus* species most commonly isolated from cases of HACEK endocarditis; *H. paraphrophilus* is less common. Of patients with HACEK endocarditis due to Haemophilus species, 60% have been ill for <2 months before presentation, and 19–50% develop congestive heart failure. Mortality rates as high as 30–50% were reported in older series; however, more recent studies have documented mortality rates of <5%. *H. aphrophilus* also causes invasive bone and joint infections, and *H. parainfluenzae* has been isolated from other infections, such as meningitis; brain, dental, and liver abscess; pneumonia; and septicemia.

AGGREGATIBACTER ACTINOMYCETEMCOMITANS

A. actinomycetemcomitans can be isolated from soft tissue infections and abscesses in association with *Actinomyces israelii*. Typically, patients who develop endocarditis with *A. actinomycetemcomitans* have severe periodontal disease or have recently undergone dental procedures in the setting of underlying cardiac valvular damage. The disease is insidious; patients may be sick for several months before diagnosis. Frequent complications include embolic phenomena, congestive heart failure, and renal failure. *A. actinomycetemcomitans* has been isolated from patients with brain abscess, meningitis, endophthalmitis, parotitis, osteomyelitis, urinary tract infection, pneumonia, and empyema, among other infections.

CARDIOBACTERIUM HOMINIS

C. hominis primarily causes endocarditis in patients with underlying valvular heart disease or with prosthetic valves. This organism most frequently affects the aortic valve. Many patients have signs and symptoms of long-standing infection before diagnosis, with evidence of arterial embolization, vasculitis, cerebrovascular accidents, immune complex glomerulonephritis, or arthritis at presentation. Embolization, mycotic aneurysms, and congestive heart failure are common complications. A second species, *C. valvarum*, has now been described in association with endocarditis.

EIKENELLA CORRODENS

E. corrodens is most frequently recovered from sites of infection in conjunction with other bacterial species. Clinical sources of *E. corrodens* include sites of human bite wounds (clenched-fist injuries), endocarditis, soft tissue infections, osteomyelitis, head and neck infections, respiratory infections, chorioamnionitis, gynecologic infections associated with intrauterine devices, meningitis and brain abscesses, and visceral abscesses.

KINGELLA KINGAE

Because of improved microbiologic methodology, isolation of *K. kingae* is increasingly common. Inoculation of clinical specimens (e.g., synovial fluid) into aerobic blood culture bottles enhances recovery of this organism. Specific real-time PCR studies of joint fluid can identify *K. kingae* in culture-negative cases. Invasive *K. kingae* infections with bacteremia are associated with upper respiratory tract infections and stomatitis in 80% of cases. Rates of oropharyngeal colonization with *K. kingae* are highest in the first 3 years of life, coinciding with increased incidence of skeletal infections due to this organism. *K. kingae* bacteremia can present with a petechial rash similar to that seen in *Neisseria meningitidis* sepsis.

Infective endocarditis, unlike other infections with *K. kingae*, occurs in older children and adults. The majority of patients have preexisting valvular disease. There is a high incidence of complications, including arterial emboli, cerebrovascular accidents, tricuspid insufficiency, and congestive heart failure with cardiovascular collapse.

Treatment: Endocarditis Caused by HACEK Organisms

See **Table 146-1**. Native-valve endocarditis should be treated for 4 weeks with antibiotics, whereas prosthetic-valve endocarditis requires 6 weeks of therapy. The cure rates for HACEK prosthetic-valve endocarditis appear to be high. Unlike prosthetic-valve endocarditis caused by other gram-negative organisms, HACEK endocarditis is often cured with antibiotic treatment alone—i.e., without surgical intervention.

Table 146-1 Treatment of Endocarditis Caused by HACEK Group Organisms[a]

Organism	Initial Therapy	Alternative Agents	Comments
Haemophilus species, *Aggregatibacter actinomycetemcomitans*	Ceftriaxone (2 g/d)	Ampicillin/sulbactam (3 g of ampicillin q6h) **or** fluoroquinolones[b]	Ampicillin ± an aminoglycoside can be used if the organism does not produce β-lactamase.[c]
Cardiobacterium hominis	Ceftriaxone (2 g/d)	Ampicillin/sulbactam (3 g of ampicillin q6h)	Penicillin (16–18 mU q4h) or ampicillin (2 g q4h) should be used if the organism is susceptible.
Eikenella corrodens	Ampicillin (2 g q4h)	Ceftriaxone (2 g/d) **or** fluoroquinolones[b]	The organism is typically resistant to clindamycin, metronidazole, and aminoglycosides.
Kingella kingae	Ceftriaxone (2 g/d) **or** ampicillin/sulbactam (3 g of ampicillin q6h)	Fluoroquinolones[b]	The prevalence of β-lactamase-producing strains is increasing. Efficacy for invasive infections is best demonstrated for first-line treatments.

[a]Susceptibility testing should be performed in all cases to guide therapy. See text for recommended durations of treatment.

[b]Fluoroquinolones are not recommended for treatment of children <18 years of age.

[c]European guidelines for endocarditis recommend the addition of gentamicin (3 mg/kg per day in 3 divided doses for 2–4 weeks).

OTHER GRAM-NEGATIVE BACTERIA

ACHROMOBACTER XYLOSOXIDANS

Achromobacter (previously *Alcaligenes*) *xylosoxidans* is probably part of the endogenous intestinal flora and has been isolated from a variety of water sources, including well water, IV fluids, and humidifiers. Immunocompromised hosts, including patients with cancer and postchemotherapy neutropenia, cirrhosis, chronic renal failure, and cystic fibrosis, are at increased risk. Nosocomial

outbreaks and pseudo-outbreaks of *A. xylosoxidans* infection have been attributed to contaminated fluids, and clinical illness has been associated with isolates from many sites, including blood (often in the setting of intravascular devices). Community-acquired *A. xylosoxidans* bacteremia usually occurs in the setting of pneumonia. Metastatic skin lesions are present in one-fifth of cases. The reported mortality rate is 67%—a figure similar to rates for other bacteremic gram-negative pneumonias.

Treatment: *Achromobacter Xylosoxidans* Infections

Treatment is based on in vitro susceptibility testing of all clinically relevant isolates. Imipenem, piperacillin-tazobactam, and trimethoprim-sulfamethoxazole (TMP-SMX) are typically the most active agents, but multidrug-resistant isolates sensitive only to colistin have been reported.

AEROMONAS SPECIES

More than 85% of *Aeromonas* infections are caused by *A. hydrophila*, *A. caviae*, and *A. veronii* biovar *sobria*. *Aeromonas* proliferates in potable water, freshwater, and soil. It remains controversial whether *Aeromonas* is a cause of bacterial gastroenteritis; asymptomatic colonization of the intestinal tract with *Aeromonas* occurs frequently. However, rare cases of hemolytic-uremic syndrome following bloody diarrhea have been shown to be secondary to the presence of *Aeromonas*.

Aeromonas causes sepsis and bacteremia in infants with multiple medical problems and in immunocompromised hosts, particularly those with cancer or hepatobiliary disease. *Aeromonas* infection and sepsis can occur in patients with trauma (including severe trauma with myonecrosis) and in burn patients exposed to *Aeromonas* by environmental (freshwater or soil) contamination of their wounds. Reported mortality rates range from 25% among immunocompromised adults with sepsis to >90% among patients with myonecrosis. *Aeromonas* can produce ecthyma gangrenosum (hemorrhagic vesicles surrounded by a rim of erythema with central necrosis and ulceration; see Fig. e7-35) resembling the lesions seen in *Pseudomonas aeruginosa* infection. *Aeromonas* causes nosocomial infections related to catheters, surgical incisions, or use of leeches. Other manifestations include necrotizing fasciitis, meningitis, peritonitis, pneumonia, and ocular infections.

Treatment: *Aeromonas* Infections

Aeromonas species are generally susceptible to fluoroquinolones (e.g., ciprofloxacin at a dosage of 500 mg every 12 h PO or 400 mg every 12 h IV), third-generation cephalosporins, carbapenems, and aminoglycosides. Because *Aeromonas* can produce various β-lactamases, including carbapenemases, susceptibility testing must be used to guide therapy. Antibiotic prophylaxis (e.g., with ciprofloxacin) is indicated when medicinal leeches are used.

CAPNOCYTOPHAGA SPECIES

This genus of fastidious, fusiform, gram-negative coccobacilli is facultatively anaerobic and requires an atmosphere enriched in carbon dioxide for optimal growth. *C. ochracea*, *C. gingivalis*, *C. haemolytica*, and *C. sputigena* have been associated with sepsis in immunocompromised hosts, particularly neutropenic patients with oral ulcerations. These species have been isolated from many other sites as well, usually as part of a polymicrobial infection. Most *Capnocytophaga* infections are contiguous with the oropharynx (e.g., periodontal disease, respiratory tract infections, cervical abscesses, and endophthalmitis).

C. canimorsus and *C. cynodegmi* are endogenous to the canine mouth (Chap. e24). Patients infected with these species frequently have a history of dog bites or of exposure to dogs without scratches or bites. Asplenia, glucocorticoid therapy, and alcohol abuse are predisposing conditions that can be associated with severe sepsis with shock and disseminated intravascular coagulation. Patients typically have a petechial rash that can progress from purpuric lesions to gangrene. Meningitis, endocarditis, cellulitis, osteomyelitis, and septic arthritis have also been associated with these organisms.

Treatment: *Capnocytophaga* Infections

Because of increasing β-lactamase production, a penicillin derivative plus a β-lactamase inhibitor —such as ampicillin/sulbactam (1.5–3.0 g of ampicillin every 6 h)—is currently recommended for empirical treatment of infections caused by *Capnocytophaga* species. If the isolate is known to be susceptible, infections with *C. canimorsus* should be treated with penicillin (12–18 million units every 4 h). *Capnocytophaga* is also susceptible to clindamycin (600–900 mg every 6–8 h). This regimen or ampicillin/sulbactam should be given prophylactically to asplenic patients who have sustained dog-bite injuries.

CHRYSEOBACTERIUM SPECIES

Chryseobacterium (formerly *Flavobacterium*) *meningosepticum* is an important cause of nosocomial infections, including outbreaks due to contaminated fluids (e.g., disinfectants and aerosolized antibiotics) and sporadic infections due to indwelling devices, feeding tubes, and other fluid-associated apparatuses. Nosocomial *C. meningosepticum* infection usually involves neonates or patients with underlying immunosuppression (e.g., related to malignancy). *C. meningosepticum* has been reported

to cause meningitis (primarily in neonates), pneumonia, sepsis, endocarditis, bacteremia, and soft tissue infections. *C. indologenes* has caused bacteremia, sepsis, and pneumonia, typically in immunocompromised patients with indwelling devices.

Treatment: *Chryseobacterium* Infections

Chryseobacteria are often susceptible to fluoroquinolones and TMP-SMX. They may be susceptible to β-lactam/β-lactamase inhibitor agents such as piperacillin-tazobactam but can possess extended-spectrum β-lactamases and metallo-β-lactamases. Susceptibility testing should be performed.

PASTEURELLA MULTOCIDA

P. multocida is a bipolar-staining, gram-negative coccobacillus that colonizes the respiratory and gastrointestinal tracts of domestic animals; oropharyngeal colonization rates are 70–90% in cats and 50–65% in dogs. *P. multocida* can be transmitted to humans through bites or scratches, via the respiratory tract from contact with contaminated dust or infectious droplets, or via deposition of the organism on injured skin or mucosal surfaces during licking. Most human infections affect skin and soft tissue; almost two-thirds of these infections are caused by cats. Patients at the extremes of age or with serious underlying disorders (e.g., cirrhosis, diabetes) are at increased risk for systemic manifestations, including meningitis, peritonitis, osteomyelitis and septic arthritis, endocarditis, and septic shock, but cases have also occurred in healthy individuals. If inhaled, *P. multocida* can cause acute respiratory tract infection, particularly in patients with underlying sinus and pulmonary disease.

Treatment: Pasteurella Multocida Infections

P. multocida is susceptible to penicillin, ampicillin, ampicillin/sulbactam, second- and third-generation cephalosporins, tetracyclines, and fluoroquinolones. β-lactamase-producing strains have been reported.

MISCELLANEOUS ORGANISMS

Rhizobium (formerly *Agrobacterium) radiobacter* has usually been associated with infection in the presence of medical devices, including intravascular catheter–related infections, prosthetic-joint and prosthetic-valve infections, and peritonitis caused by dialysis catheters. Most cases occur in immunocompromised hosts, especially individuals with malignancy or HIV infection. Strains are usually susceptible to fluoroquinolones, third- and fourth-generation cephalosporins, and carbapenems.

Shewanella putrefaciens and *S. algae* are ubiquitous organisms found primarily in seawater. Devitalized tissues can become colonized with *Shewanella* and serve as a nidus for systemic infection. *Shewanella* species cause skin and soft tissue infections, chronic ulcers of the lower extremities, ear infections, biliary tract infections, pneumonia, necrotizing fasciitis, bacteremia, and sepsis. A fulminant course is associated with cirrhosis, malignancy, or other severe underlying conditions. Organisms are often susceptible to fluoroquinolones, third- and fourth-generation cephalosporins, and aminoglycosides.

Chromobacterium violaceum has been responsible for life-threatening infections with severe sepsis and metastatic abscesses, particularly in children with defective neutrophil function (e.g., those with chronic granulomatous disease). *Ochrobactrum anthropi* causes infections related to central venous catheters in compromised hosts; other invasive infections have been described. Other organisms implicated in human infections include *Weeksella* species; various CDC groups, such as EF4 and Ve-2; *Flavimonas* species; *Sphingobacterium* species; *Protomonas* species; and *Oligella urethralis.* The reader is advised to consult subspecialty texts and references for further guidance on these organisms.

FURTHER READINGS

Das M et al: Infective endocarditis caused by HACEK microorganisms. Annu Rev Med 48:25, 1997[PMID: 9046942]

Dubnov-Raz G et al: Invasive *Kingella kingae* infections in children: Clinical and laboratory characteristics. Pediatrics 122:1305, 2008[PMID: 19047250]

Hung PP et al: *Chryseobacterium meningosepticum* infection: Antibiotic susceptibility and risk factors for mortality. J Microbiol Immunol Infect 41:137, 2008[PMID: 18473101]

Jolivet-Gougeon A et al: Antimicrobial treatment of *Capnocytophaga* infections. Int J Antimicrob Agents 29:367, 2007[PMID: 17250994]

Kristinsson G, Adam HM: *Pasteurella multocida* infections. Pediatr Rev 28:472, 2007[PMID: 18055649]

Lamy B et al: Prospective nationwide study of *Aeromonas* infections in France. J Clin Microbiol

47:1234, 2009[PMID: 19244464]

Oehler RL et al: Bite-related and septic syndromes caused by cats and dogs. Lancet Infect Dis 9:439, 2009[PMID: 19555903]

Teng SO et al: Complicated intra-abdominal infection caused by extended drug-resistant *Achromobacter xylosoxidans*. J Microbiol Immunol Infect 42:176, 2009[PMID: 19597652]

Udaka T et al: *Eikenella corrodens* in head and neck infections. J Infect 54:343, 2007[PMID: 16962664]

Yagupsky P et al: Dissemination of *Kingella kingae* in the community and long-term persistence of invasive clones. Pediatr Infect Dis J 28:707, 2009[PMID: 19593253]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LEGIONELLA INFECTIONS: INTRODUCTION

Legionellosis refers to the two clinical syndromes caused by bacteria of the genus *Legionella*. *Pontiac fever* is an acute, febrile, self-limited illness that has been serologically linked to *Legionella* species, whereas *Legionnaires' disease* is the designation for pneumonia caused by these species. Legionnaires' disease was first recognized in 1976, when an outbreak of pneumonia took place at a Philadelphia hotel during an American Legion convention.

Microbiology

The family Legionellaceae comprises more than 50 species with more than 70 serogroups. The species *L. pneumophila* causes 80–90% of human infections and includes at least 16 serogroups; serogroups 1, 4, and 6 are most commonly implicated in human infections. To date, 18 species other than *L. pneumophila* have been associated with human infections, among which *L. micdadei* (Pittsburgh pneumonia agent), *L. bozemanii*, *L. dumoffii*, and *L. longbeachae* are the most common. Members of the Legionellaceae are aerobic gram-negative bacilli that do not grow on routine microbiologic media. Buffered charcoal yeast extract (BCYE) agar is the medium used to grow *Legionella*.

Ecology and Transmission

The natural habitats for *L. pneumophila* are aquatic bodies, including lakes and streams. *L. longbeachae* has been isolated from natural soil and commercial potting soil. Legionellae can survive under a wide range of environmental conditions; for example, the organisms can live for years in refrigerated water samples. Natural bodies of water contain only small numbers of legionellae. However, once the organisms enter human-constructed aquatic reservoirs (such as drinking-water systems), they can grow and proliferate. Factors known to enhance colonization by and amplification of legionellae include warm temperatures (25°–42°C) and scale and sediment. *L. pneumophila* can form microcolonies within biofilms; its eradication from drinking-water systems requires disinfectants that can penetrate the biofilm. The presence of symbiotic microorganisms, including algae, amebas, ciliated protozoa, and other water-dwelling bacteria, promotes the growth of legionellae. The organisms can invade and multiply within free-living protozoa. Rainfall and humidity have been identified as environmental risk factors.

Sporadic community-acquired Legionnaires' disease has been linked to colonization of residential, hotel, and industrial water supplies. Drinking-water systems in hospitals and extended-care facilities have been linked to health care–associated Legionnaires' disease.

Cooling towers and evaporative condensers have been overestimated as sources of *Legionella*. Early investigations that implicated cooling towers antedated the discovery that the organism could also exist in drinking water. In many outbreaks attributed to cooling towers, cases of Legionnaires' disease continued to occur despite disinfection of the cooling towers; drinking water was the actual source. Koch's postulates have never been fulfilled for cooling tower–associated outbreaks as they have been for hospital-acquired Legionnaires' disease. Nevertheless, cooling towers have occasionally been identified in community-acquired outbreaks, including an outbreak in Murcia, Spain, in which several hundred suspected cases of Legionnaires' disease occurred over a 3-week period. As mentioned above, *L. longbeachae* infections have been linked to potting soil, but the mode of transmission remains to be clarified.

Multiple modes of transmission of *Legionella* to humans exist, including aerosolization, aspiration, and direct instillation into the lungs during respiratory tract manipulations. Aspiration is now known to be the predominant mode of transmission, but it is unclear whether *Legionella* enters the lungs via oropharyngeal colonization or directly via the drinking of contaminated water. Oropharyngeal colonization has been demonstrated in patients undergoing transplantation. Nasogastric tubes have been linked to hospital-acquired Legionnaires' disease; microaspiration of contaminated water was the hypothesized mode of transmission. Surgery with general anesthesia is a known risk factor that is consistent with aspiration. Especially compelling is the reported 30% incidence of postoperative Legionnaires' disease among patients undergoing head and neck surgery at a hospital with a contaminated water supply; aspiration is a recognized sequela in such cases. Studies of patients with hospital-acquired Legionnaires' disease have shown that these individuals underwent endotracheal intubation significantly more often and for a significantly longer duration than patients with hospital-

acquired pneumonia of other etiologies.

Aerosolization of *Legionella* by devices filled with tap water, including whirlpools, nebulizers, and humidifiers, has been implicated. An ultrasonic mist machine in the produce section of a grocery store was the source in a community outbreak. Pontiac fever has been linked to *Legionella*-containing aerosols from water-using machinery, a cooling tower, air-conditioners, and whirlpools.

Epidemiology

The incidence of Legionnaires' disease depends on the degree of contamination of the aquatic reservoir, the immune status of the persons exposed to water from that reservoir, the intensity of exposure, and the availability of specialized laboratory tests on which the correct diagnosis can be based. Numerous prospective studies have ranked *Legionella* among the top four microbial causes of community-acquired pneumonia, accounting for 2–9% of cases. (*Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Chlamydophila pneumoniae* are usually ranked first, second, and third, respectively.) On the basis of a multihospital study of community-acquired pneumonia in Ohio, the Centers for Disease Control and Prevention (CDC) estimated that as many as 18,000 cases of sporadic community-acquired Legionnaires' disease occur annually in the United States and that only 3% of these cases are correctly diagnosed. *Legionella* is responsible for 10–50% of cases of nosocomial pneumonia when a hospital's water system is colonized with the organisms. The incidence of hospital-acquired Legionnaires' disease depends on the degree of contamination of drinking water as defined by the rate of positivity of distal water sites (not as defined quantitatively by the number of colony-forming units per milliliter).

Risk factors for Legionnaires' disease include cigarette smoking; chronic lung disease; advanced age; prior hospitalization, with discharge within 10 days before onset of pneumonia symptoms; and immunosuppression. Immunosuppressive conditions that predispose to Legionnaires' disease include transplantation, HIV infection, and treatment with glucocorticoids or tumor necrosis factor α antagonists. However, in a large prospective study of community-acquired pneumonia, 28% of patients with Legionnaires' disease did not have these classic risk factors. Surgery is a prominent predisposing factor in hospital-acquired infection, with transplant recipients at highest risk. Hospital-acquired cases are now being recognized among neonates and immunosuppressed children.

Pontiac fever occurs in epidemics. The high attack rate (>90%) reflects airborne transmission.

Pathogenesis and Immunity

Legionella enters the lungs through aspiration or direct inhalation. Attachment to host cells is mediated by bacterial type IV pili, heat-shock proteins, a major outer-membrane protein, and complement. Because the organism possesses pili that mediate adherence to respiratory tract epithelial cells, conditions that impair mucociliary clearance, including cigarette smoking, lung disease, or alcoholism, predispose to Legionnaires' disease.

Both the innate and adaptive immune responses play a role in host defense. Toll-like receptors mediate recognition of *L. pneumophila* in alveolar macrophages and enhance early neutrophil recruitment to the site of infection. Alveolar macrophages phagocytose legionellae by a conventional or a coiling mechanism. The macrophage infectivity potentiation (MIP) surface protein enhances infection of the macrophages. After phagocytosis, *L. pneumophila* evades intracellular killing by inhibiting phagosome-lysosome fusion. Although many legionellae are killed, some proliferate intracellularly until the cells rupture; the bacteria are then phagocytosed again by newly recruited phagocytes, and the cycle begins anew. The role of neutrophils in immunity appears to be minimal: neutropenic patients are not predisposed to Legionnaires' disease. Although *L. pneumophila* is susceptible to oxygen-dependent microbiologic systems in vitro, it resists killing by neutrophils. The humoral immune system is active against *Legionella*. Type-specific IgM and IgG antibodies are measurable within weeks of infection. In vitro, antibodies promote killing of *Legionella* by phagocytes (neutrophils, monocytes, and alveolar macrophages). Immunized animals develop a specific antibody response, with subsequent resistance to *Legionella* challenge. However, antibodies neither enhance lysis by complement nor inhibit intracellular multiplication within phagocytes.

Some *L. pneumophila* strains are clearly more virulent than others, although the precise factors mediating virulence remain uncertain. For example, although multiple strains may colonize water-distribution systems, only a few cause disease in patients exposed to water from these systems. At least one surface epitope of *L. pneumophila* serogroup 1 is associated with virulence. Monoclonal antibody subtype mAb2 has been linked to virulence. *L. pneumophila* serogroup 6 is more commonly involved in hospital-acquired Legionnaires' disease and is more likely to be associated with a poor outcome.

The genome of *L. pneumophila* has been sequenced. A broad range of membrane transporters within the genome are thought to optimize the use of nutrients in water and soil.

Clinical and Laboratory Features

PONTIAC FEVER

Pontiac fever is an acute, self-limiting, flu-like illness with an incubation period of 24–48 h. Pneumonia does not develop. Malaise, fatigue, and myalgias are the most common symptoms, occurring in 97% of cases. Fever (usually with chills) develops in 80–90% of cases and headache in 80%. Other symptoms (seen in <50% of cases) include arthralgias, nausea, cough, abdominal pain, and diarrhea. Modest leukocytosis with a neutrophilic predominance is sometimes detected. Complete recovery occurs within a few days; antibiotic therapy is unnecessary. A few patients may experience lassitude for many weeks after recovery. The diagnosis is established by antibody seroconversion.

LEGIONNAIRES' DISEASE (PNEUMONIA)

Legionnaires' disease is often included in the differential diagnosis of "atypical pneumonia," along with pneumonia due to *C. pneumoniae*, *Chlamydophila psittaci*, *Mycoplasma pneumoniae*, *Coxiella burnetii*, and some viruses. The clinical similarities among these types of pneumonia include a relatively nonproductive cough and a low incidence of grossly purulent sputum. However, the clinical manifestations of Legionnaires' disease are usually more severe than those of most "atypical" pneumonias, and the course and prognosis of *Legionella* pneumonia more closely resemble those of bacteremic pneumococcal pneumonia than those of pneumonia due to other "atypical" pathogens. Patients with community-acquired Legionnaires' disease are significantly more likely than patients with pneumonia of other etiologies to be admitted to an intensive care unit on presentation.

The incubation period for Legionnaires' disease is usually 2–10 days, although longer incubation periods have been documented. The symptoms and signs may range from a mild cough and a slight fever to stupor with widespread pulmonary infiltrates and multisystem failure. Nonspecific symptoms—malaise, fatigue, anorexia, and headache—are seen early in the illness. Myalgias and arthralgias are uncommon but are prominent in a few patients. Upper respiratory symptoms, including coryza, are rare.

The mild cough of Legionnaires' disease is only slightly productive. Sometimes the sputum is streaked with blood. Chest pain—either pleuritic or nonpleuritic—can be a prominent feature and, when coupled with hemoptysis, can lead to an incorrect diagnosis of pulmonary embolism. Shortness of breath is reported by one-third to one-half of patients. Gastrointestinal difficulties are often pronounced; abdominal pain, nausea, and vomiting affect 10–20% of patients. Diarrhea (watery rather than bloody) is reported in 25–50% of cases. The most common neurologic abnormalities are confusion or changes in mental status; however, the multitudinous neurologic symptoms reported range from headache and lethargy to encephalopathy.

Patients with Legionnaires' disease virtually always have fever. Temperatures in excess of 40.5°C (104.9°F) were recorded in 20% of the cases in one series. Relative bradycardia has been overemphasized as a useful diagnostic finding; it occurs primarily in older patients with severe pneumonia. Chest examination reveals rales early in the course and evidence of consolidations as the disease progresses. Abdominal examination may reveal generalized or local tenderness.

Although the clinical manifestations often considered classic for Legionnaires' disease **(Table 147-1)** may suggest the diagnosis, prospective comparative studies have shown that clinical manifestations are generally nonspecific and that Legionnaires' disease is not readily distinguishable from pneumonia of other etiologies. In a review of 13 studies of community-acquired pneumonia, clinical manifestations that occurred significantly more often in Legionnaires' disease included diarrhea, neurologic findings (including confusion), and a temperature of >39°C. Hyponatremia, elevated values in liver function tests, and hematuria also occurred more frequently in Legionnaires' disease. Other laboratory abnormalities include creatine phosphokinase elevation, hypophosphatemia, serum creatinine elevation, and proteinuria.

Table 147-1 Clinical Clues Suggestive of Legionnaires' Disease

Diarrhea
High fever (>40°C; >104°F)
Numerous neutrophils but no organisms revealed by Gram's staining of respiratory secretions
Hyponatremia (serum sodium level <131 mg/dL)
Failure to respond to ß-lactam drugs (penicillins or cephalosporins) and aminoglycoside antibiotics
Occurrence of illness in an environment in which the poTable water supply is known to be contaminated with *Legionella*
Onset of symptoms within 10 days after discharge from the hospital

Sporadic cases of Legionnaires' disease appear to be more severe than outbreak-associated and hospital-acquired cases, presumably because their diagnosis is delayed. Results of the German CAPNETZ Study showed that, among cases of community-acquired *Legionella* pneumonia, ambulatory cases were as common as cases requiring hospitalization.

EXTRAPULMONARY LEGIONELLOSIS

Since the portal of entry for *Legionella* is the lung in virtually all cases, extrapulmonary manifestations usually result from bloodborne dissemination from the lung. *Legionella* has been identified in lymph nodes, spleen, liver, or kidneys in autopsied cases. The most common extrapulmonary site of legionellosis is the heart; numerous reports have described myocarditis, pericarditis, postcardiotomy syndrome, and prosthetic-valve endocarditis. Most cases have been hospital-acquired. In some patients without overt evidence of pneumonia, the organisms may gain entry through a postoperative sternal wound exposed to contaminated tap water or through a mediastinal-tube insertion site. Sinusitis, peritonitis, pyelonephritis, skin and soft tissue infection, septic arthritis, and pancreatitis have been seen predominantly in immunosuppressed patients.

Chest Radiography

Virtually all patients with Legionnaires' disease have abnormal chest radiographs showing pulmonary infiltrates at the time of clinical presentation. In a few cases of hospital-acquired disease, fever and respiratory tract symptoms have preceded the radiographic appearance of the infiltrate. Radiologic findings are nonspecific. Pleural effusion is evident in 28–63% of patients on hospital admission. In immunosuppressed patients, especially those receiving glucocorticoids, distinctive rounded nodular opacities may be seen; these lesions may expand and cavitate **(Fig. 147-1)**. Likewise, abscesses can occur in immunosuppressed hosts. The progression of infiltrates and pleural effusion on chest radiography despite appropriate antibiotic therapy within the first week is common, and radiographic improvement lags behind clinical improvement by several days. Complete clearing of infiltrates requires 1–4 months.

Figure 147-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiographic findings in a 52-year-old man who presented with pneumonia subsequently diagnosed as Legionnaires' disease. The patient was a cigarette smoker with chronic obstructive pulmonary disease and alcoholic cardiomyopathy; he had received glucocorticoids. *L. pneumophila* was identified by direct fluorescent antibody staining and culture of sputum. ***Left:*** Baseline chest radiograph showing long-standing cardiomegaly. ***Center:*** Admission chest radiograph showing new rounded opacities. ***Right:*** Chest radiograph taken 3 days after admission, during treatment with erythromycin.

Diagnosis

Given the nonspecific clinical manifestations of Legionnaires' disease and the high mortality rates for untreated Legionnaires' disease, the use of *Legionella* testing—especially the *Legionella* urinary antigen test—is recommended for all patients with community-acquired pneumonia, including patients with ambulatory pneumonia and hospitalized children. *Legionella* cultures should be made more widely available since the urinary antigen test can diagnose only *L. pneumophila* serogroup 1. Hospitals in which the drinking water is known to be colonized with *Legionella* species should have *Legionella* cultures routinely available for all patients with hospital-acquired pneumonia.

The diagnosis of Legionnaires' disease requires special microbiologic tests **(Table 147-2)**. The sensitivity of bronchoscopy specimens is similar to that of sputum samples for culture on selective media; if sputum is not available, bronchoscopy specimens may yield the organism. Bronchoalveolar lavage fluid gives higher yields than bronchial wash specimens. Thoracentesis should be performed if pleural effusion is found, and the fluid should be evaluated by direct fluorescent antibody (DFA) staining, culture, and the antigen assay designed for use with urine.

Table 147-2 Utility of Special Laboratory Tests for the Diagnosis of Legionnaires' Disease

Test	Sensitivity, %	Specificity, %
Culture		
Sputum[a]	80	100

Transtracheal aspirate	90	100
Direct fluorescent antibody staining of sputum	50–70	96–99
Urinary antigen testing[b]	70	100
Antibody serology[c]	40–60	96–99

[a]Use of multiple selective media with dyes.

[b]Serogroup 1 only.

[c]IgG and IgM testing of both acute- and convalescent-phase sera. A single titer of ≥1:256 is considered presumptive, while fourfold seroconversion is considered definitive.

STAINING

Gram's staining of material from normally sterile sites, such as pleural fluid or lung tissue, occasionally suggests the diagnosis; efforts to detect *Legionella* in sputum by Gram's staining typically reveal numerous leukocytes but no organisms. When they are visualized, the organisms appear as small, pleomorphic, faint, gram-negative bacilli. *L. micdadei* organisms can be detected as weakly or partially acid-fast bacilli in clinical specimens.

The DFA test is rapid and highly specific but is less sensitive than culture because large numbers of organisms are required for microscopic visualization. This test is more likely to be positive in advanced than in early disease.

CULTURE

The definitive method for diagnosis of *Legionella* infection is isolation of the organism from respiratory secretions, although culture for 3–5 days is required. Antibiotics added to the medium suppress the growth of competing flora from nonsterile sites, and dyes color the colonies and assist in identification. The use of multiple selective BCYE media is necessary for maximal sensitivity. When culture plates are overgrown with other microflora, pretreatment of the specimen with acid or heat can markedly improve the yield. *L. pneumophila* is often isolated from sputum that is not grossly or microscopically purulent; sputum containing more than 25 epithelial cells per high-power field (a finding that classically suggests contamination) may still yield *L. pneumophila*.

ANTIBODY DETECTION

Antibody testing of both acute- and convalescent-phase sera is necessary. A fourfold rise in titer is diagnostic; 12 weeks are often required for the detection of an antibody response. A single titer of 1:128 in a patient with pneumonia constitutes circumstantial evidence for Legionnaires' disease. Serology is of use primarily in epidemiologic studies. The specificity of serology for *Legionella* species other than *L. pneumophila* is uncertain; there is cross-reactivity with *Legionella* species and some gram-negative bacilli.

URINARY ANTIGEN

The assay for *Legionella* soluble antigen in urine is rapid, relatively inexpensive, easy to perform, second only to culture in terms of sensitivity, and highly specific. Several enzyme immunoassays and a rapid immunochromatographic assay are commercially available. The rapid immunochromatographic assay is relatively inexpensive and easy to perform. The urinary antigen test is available only for *L. pneumophila* serogroup 1, which causes ~80% of *Legionella* infections. Cross-reactivity with other *L. pneumophila* serogroups and other *Legionella* species has been detected in up to 22% of urine samples from patients with culture-proven cases. Antigen in urine is detectable 3 days after the onset of clinical disease and disappears over 2 months; positivity can be prolonged when patients receive glucocorticoids. The test is not affected by antibiotic administration.

MOLECULAR METHODS

DFA stains can identify a number of *Legionella* species. Both polyclonal and monoclonal antibody stains are commercially available. Although its application is currently limited to research investigations, polymerase chain reaction (PCR) with DNA probes is theoretically more sensitive and specific than other methods. A molecular probe is undergoing evaluation. PCR has proven somewhat useful in the identification of *Legionella* from environmental water specimens. In PCR (unlike culture), epidemiologic links cannot be made since the infecting pathogen is not available for molecular subtyping.

Treatment: *Legionella* Infection

Because *Legionella* is an intracellular pathogen, antibiotics that can attain high intracellular concentrations are most likely to be effective. The dosages for various drugs used in the treatment of *Legionella* infection are listed in **Table 147-3**.

Table 147-3 Antibiotic Therapy for *Legionella* Infection

Antimicrobial Agent	Dosage[a]
Macrolides	
Azithromycin	500 mg[b] PO or IV[c] q24h
Clarithromycin	500 mg PO or IV[c] q12h
Quinolones	
Levofloxacin	750 mg IV q24h 500 mg[b] PO q24h
Ciprofloxacin	400 mg IV q8h 750 mg PO q12h
Moxifloxacin	400 mg[b] PO q24h
Ketolide	
Telithromycin	800 mg PO q24h
Tetracyclines	
Doxycycline	100 mg[b] PO or IV q12h
Minocycline	100 mg[b] PO or IV q12h
Tetracycline	500 mg PO or IV q6h
Tigecycline	100-mg IV load, then 50 mg IV q12h
Others	
Trimethoprim-sulfamethoxazole	160/800 mg IV q8h
Rifampin[d]	160/800 mg PO q12h 300–600 mg PO or IV q12h

[a]Dosages are derived from clinical experience.

[b]The authors recommend doubling the first dose.

[c]The IV formulation is not available in some countries.

[d]Rifampin should be used only in combination with a macrolide or a quinolone.

The macrolides (especially azithromycin) and the respiratory quinolones are now the antibiotics of choice and are effective as monotherapy. Compared with erythromycin, the newer macrolides have superior in vitro activity, display greater intracellular activity, reach higher concentrations in respiratory secretions and lung tissue, and have fewer adverse effects. The pharmacokinetics of the newer macrolides and quinolones also allow once- or twice-daily dosing. Quinolones are the preferred antibiotics for transplant recipients because both macrolides and rifampin interact pharmacologically with cyclosporine and tacrolimus. Retrospective uncontrolled studies have shown that complications of pneumonia are fewer and clinical response is more rapid in patients receiving quinolones than in those receiving macrolides. Alternative agents include tetracycline and its analogues doxycycline and minocycline. Tigecycline is active in vitro but clinical experience is minimal. Anecdotal reports have described both successes and failures with trimethoprim-sulfamethoxazole, imipenem, and clindamycin. For severely ill patients with extensive pulmonary infiltrates, a two-drug combination of a newer macrolide or a quinolone with rifampin may be considered for initial treatment. Rifampin is highly active in vitro and in cell models, but its interaction with many other medications, including macrolides, is problematic. Initial therapy should be given by the IV route. A clinical response usually occurs within 3–5 days, after which oral therapy can be substituted. The total duration of therapy in the immunocompetent host is 10–14 days; a longer course (3 weeks) may be appropriate for immunosuppressed patients and those with advanced disease. For azithromycin, with its long half-life, a 5- to 10-day course is sufficient.

Pontiac fever requires only symptom-based treatment, not antimicrobial therapy.

Prognosis

Mortality rates for Legionnaires' disease vary with the patient's underlying disease and its severity, the patient's immune status, the severity of pneumonia, and the timing of administration of appropriate

antimicrobial therapy. Mortality rates are highest (80%) among immunosuppressed patients who do not receive appropriate antimicrobial therapy early in the course of illness. With appropriate and timely antibiotic treatment, mortality rates from community-acquired Legionnaires' disease among immunocompetent patients range from 0 to 11%; without treatment, the figure may be as high as 31%. In a study of survivors of an outbreak of community-acquired Legionnaires' disease, sequelae of fatigue, neurologic symptoms, and weakness were found in 63–75% of patients 17 months after receipt of antibiotics.

Prevention

Routine environmental culture of hospital water supplies is recommended as an approach to the prevention of hospital-acquired Legionnaires' disease. Guidelines mandating this proactive approach have been adopted throughout Europe and in several U.S. states. Positive cultures from the water supply mandate the use of specialized laboratory tests (especially culture on selective media and the urinary antigen test) for patients with hospital-acquired pneumonia. Studies have shown that neither a high degree of outward cleanliness of the water system nor routine application of maintenance measures decreases the frequency or intensity of *Legionella* contamination. Thus, engineering guidelines and building codes, although routinely advocated as preventive measures, have little impact on the presence of *Legionella*.

Disinfection of the drinking water supply is effective. Two methods have proved reliable and cost-effective. The superheat-and-flush method requires heating of the water so that the distal-outlet temperature is 70–80°C and flushing of the distal outlets with hot water for at least 30 min. This method is ideal for emergency situations. Commercial copper and silver ionization systems have proven effective in numerous hospitals. Chlorine dioxide is a promising modality. Tap water filters have been effective for high-risk patient areas, such as transplantation or intensive care units. Hyperchlorination is no longer recommended because of its expense, carcinogenicity, corrosive effects on piping, and unreliable efficacy.

FURTHER READINGS

Casati S et al: Compost facilities as a reservoir of *Legionella pneumophila* and other *Legionella* species. Clin Microbiol Infect 16:945, 2010[PMID: 19645762]

Greenberg D et al: Problem pathogens: Paediatric legionellosis—implications for improved diagnosis. Lancet Infect Dis 6:529, 2006[PMID: 16870531]

Hofmann A et al: Fulminant legionellosis in two patients treated with infliximab for Crohn's disease: Case series and literature review. Can J Gastroenterol 23:829, 2009[PMID: 20011736]

Pedro-Botet ML, Yu VL: Treatment strategies for Legionella infection. Expert Opin Pharmacother 10:1109, 2009[PMID: 19405787]

Ricketts KD et al: Weather patterns and Legionnaires' disease: A meteorological study. Epidemiol Infect 137:1003, 2009[PMID: 19017428]

Shin S, Roy CR: Host cell processes that influence the intracellular survival of *Legionella pneumophila*. Cell Microbiol 10:1209, 2008[PMID: 18363881]

Sopena N et al: Sporadic and epidemic community legionellosis: Two faces of the same illness. Eur Respir J 29:138, 2007[PMID: 17005576]

Squier CL et al: A proactive approach to prevention of healthcare-acquired Legionnaires' disease: The Allegheny County (Pittsburgh) experience. Am J Infect Control 33:360, 2005[PMID: 16061143]

Stout JE et al: Role of environmental surveillance in determining risk for hospital-acquired legionellosis: A national surveillance study with clinical correlations. Infect Control Hosp Epidemiol 28:818, 2007[PMID: 17564984]

Von Baum H et al: Community-acquired *Legionella* pneumonia: New insights from the German Competence Network for Community Acquired Pneumonia. Clin Infect Dis 46:1356, 2008

Yu VL, Lee TC: Neonatal legionellosis: The tip of the iceberg for pediatric hospital-acquired pneumonia. Pediatr Infect Dis 29:282, 2010[PMID: 20190615]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PERTUSSIS AND OTHER *BORDETELLA* INFECTIONS: INTRODUCTION

Pertussis is an acute infection of the respiratory tract caused by *Bordetella pertussis*. The name *pertussis* means "violent cough," which aptly describes the most consistent and prominent feature of the illness. The inspiratory sound made at the end of an episode of paroxysmal coughing gives rise to the common name for the illness, "whooping cough." However, this feature is variable: it is uncommon among infants ≥6 months of age and is frequently absent in older children and adults. The Chinese name for pertussis is "the 100-day cough," which accurately describes the clinical course of the illness. The identification of *B. pertussis* was first reported by Bordet and Gengou in 1906, and vaccines were produced over the following two decades.

Microbiology

Of the 10 identified species in the genus *Bordetella*, only three are of major medical significance. *B. pertussis* infects only humans and is the most important *Bordetella* species causing human disease. *B. parapertussis* causes an illness in humans that is similar to pertussis but is typically milder; co-infections with *B. parapertussis* and *B. pertussis* have been documented. *B. bronchiseptica* is an important pathogen of domestic animals that causes kennel cough in dogs, atrophic rhinitis and pneumonia in pigs, and pneumonia in cats. Both respiratory infection and opportunistic infection due to *B. bronchiseptica* are occasionally reported in humans. Two additional species, *B. hinzii* and *B. holmesii*, are unusual causes of bacteremia; both have been isolated from patients with sepsis, most often from those who are immunocompromised.

Bordetella species are gram-negative pleomorphic aerobic bacilli that share common genotypic characteristics. *B. pertussis* and *B. parapertussis* are the most similar of the species, but *B. parapertussis* does not express the gene coding for pertussis toxin. *B. pertussis* is a slow-growing fastidious organism that requires selective medium and forms small glistening bifurcated colonies. Suspicious colonies are presumptively identified as *B. pertussis* by direct fluorescent antibody testing or by agglutination with species-specific antiserum. *B. pertussis* is further differentiated from other *Bordetella* species by biochemical and motility characteristics.

B. pertussis produces a wide array of toxins and biologically active products that are important in its pathogenesis and in immunity. Most of these virulence factors are under the control of a single genetic locus that regulates their production, resulting in antigenic modulation and phase variation. Although these processes occur both in vitro and in vivo, their importance in the pathobiology of the organism is unknown; they may play a role in intracellular persistence and person-to-person spread. The organism's most important virulence factor is *pertussis toxin*, which is composed of a B oligomer –binding subunit and an enzymatically active A protomer that ADP-ribosylates a guanine nucleotide-binding regulatory protein (G protein) in target cells, producing a variety of biologic effects. Pertussis toxin has important mitogenic activity, affects the circulation of lymphocytes, and serves as an adhesin for bacterial binding to respiratory ciliated cells. Other important virulence factors and adhesins are *filamentous hemagglutinin*, a component of the cell wall, and *pertactin*, an outer-membrane protein. *Fimbriae*, bacterial appendages that play a role in bacterial attachment, are the major antigens against which agglutinating antibodies are directed. These agglutinating antibodies have historically been the primary means of serotyping *B. pertussis* strains. Other virulence factors include tracheal cytotoxin, which causes respiratory epithelial damage; adenylate cyclase toxin, which impairs host immune-cell function; dermonecrotic toxin, which may contribute to respiratory mucosal damage; and lipooligosaccharide, which has properties similar to those of other gram-negative bacterial endotoxins.

Epidemiology

Pertussis is a highly communicable disease, with attack rates of 80–100% among unimmunized household contacts and 20% within households in well-immunized populations. The infection has a worldwide distribution, with cyclical outbreaks every 3–5 years (a pattern that has persisted despite widespread immunization). Pertussis occurs in all months; however, in North America, its activity peaks in summer and autumn.

In developing countries, pertussis remains an important cause of infant morbidity and death. The reported incidence of pertussis worldwide has decreased as a result of improved vaccine coverage. However, coverage rates are still <50% in many developing nations **(Fig. 148-1)**; the World Health Organization (WHO) estimates that 90% of the burden of pertussis is in developing regions. In

addition, overreporting of immunization coverage and underreporting of disease result in substantial underestimation of the global burden of pertussis. The WHO estimates that there were 254,000 deaths from pertussis among children in 2004.

Figure 148-1

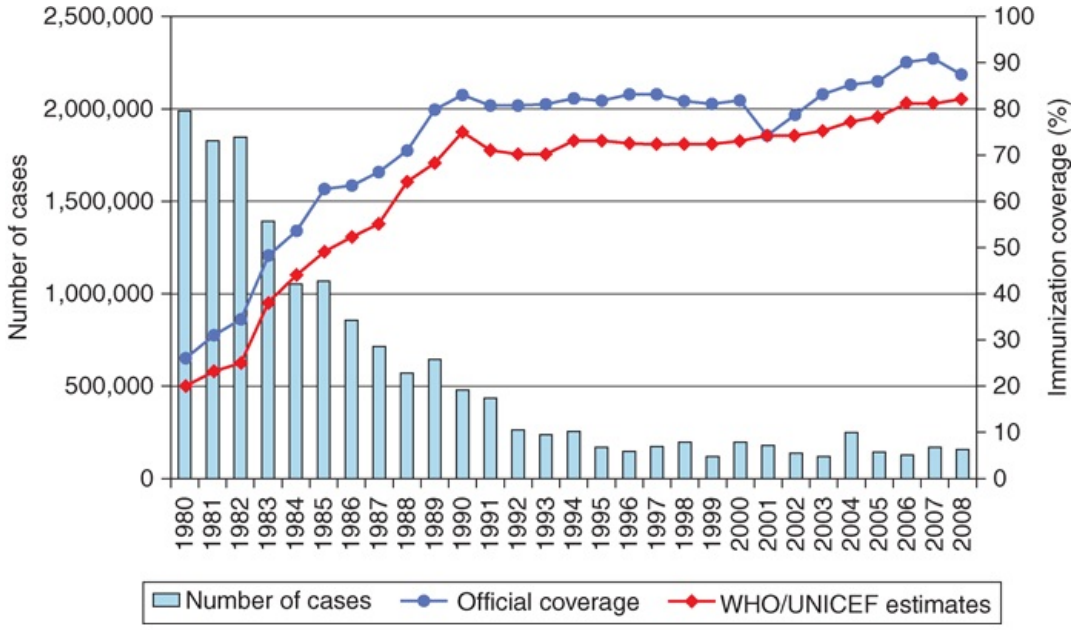

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Global annual reported pertussis incidence and rate of coverage with DTP3 (diphtheria toxoid, tetanus toxoid, and pertussis vaccine; 3 doses), 1980–2008. *(© World Health Organization, 2009. All rights reserved. From http://www.who.int/immunization_monitoring/diseases/pertussis/en/index.html.*

Before the institution of widespread immunization programs in the developed world, pertussis was one of the most common infectious causes of morbidity and death. In the United States before the 1940s, between 115,000 and 270,000 cases of pertussis were reported annually, with an average yearly rate of 150 cases per 100,000 population. With universal childhood immunization, the number of reported cases fell by >95%, and mortality rates decreased even more dramatically. Only 1010 cases of pertussis were reported in 1976 **(Fig. 148-2)**. After that historic low, rates of pertussis slowly increased, peaking at >25,000 cases annually in 2004 and 2005. In 2007, 10,454 cases of pertussis were reported in the United States.

Figure 148-2

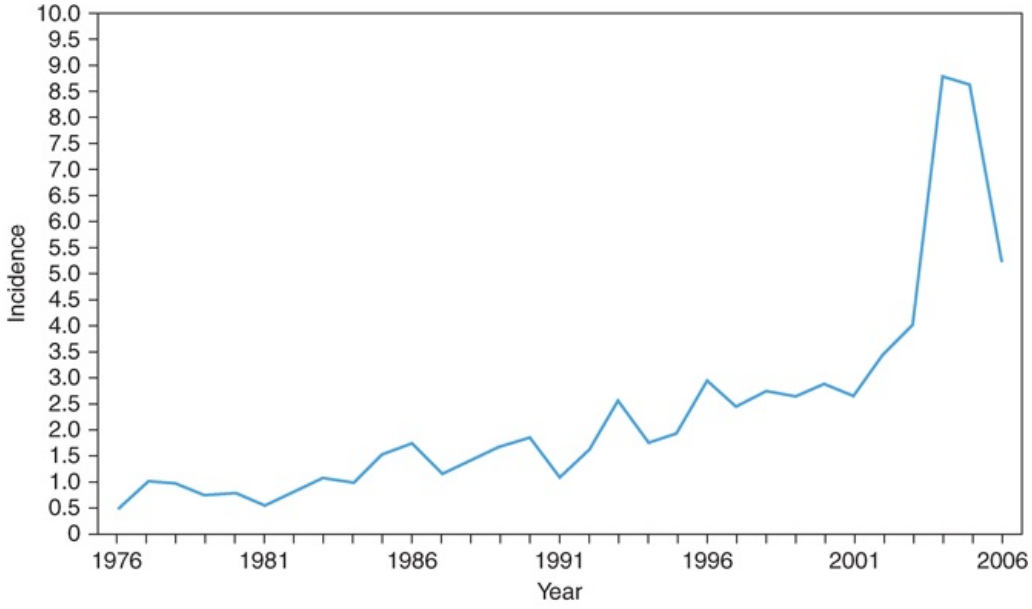

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pertussis incidence (per 100,000 population) by year—United States, 1976–2006. *[From the Centers for Disease Control and Prevention, MMWR Morb Mortal Wkly Rep 55(53):60, 2008.]*

Although thought of as a disease of childhood, pertussis can affect people of all ages and is increasingly being identified as a cause of prolonged coughing illness in adolescents and adults. In unimmunized populations, pertussis incidence peaks during the preschool years, and well over half of children have the disease before reaching adulthood. In highly immunized populations such as those in North America, the peak incidence is among infants <1 year of age who have not completed the three-dose primary immunization series. Recent trends, however, show an increasing incidence of pertussis among adolescents and adults. In the United States in 2007, although infants <6 months of age had the highest incidence of pertussis, most cases were reported in adolescents and adults. Moreover, the figures for adolescents and adults are likely to be underestimates because of a greater degree of underrecognition and underreporting in these age groups. A number of studies of prolonged coughing illness suggest that pertussis may be the etiologic agent in 12–30% of adults with cough that does not improve within 2 weeks. In one study of the efficacy of an acellular pertussis vaccine in adolescents and adults, the incidence of pertussis in the placebo group was 3.7–4.5 cases per 1000 person-years. Although this prospective cohort study yielded a lower estimate than the studies of cough illness, its results still translate to 600,000–800,000 cases of pertussis annually among adults in the United States. Severe morbidity and high mortality rates, however, are restricted almost entirely to infants. In Canada, there were 16 deaths from pertussis between 1991 and 2001; all those who died were infants ≥6 months of age. Although school-age children are the source of infection for most households, adults are the likely source for high-risk infants and may serve as the reservoir of infection between epidemic years.

Pathogenesis

Infection with *B. pertussis* is initiated by attachment of the organism to the ciliated epithelial cells of the nasopharynx. Attachment is mediated by surface adhesins (e.g., pertactin and filamentous hemagglutinin), which bind to the integrin family of cell-surface proteins, probably in conjunction with pertussis toxin. The role of fimbriae in adhesion and in maintenance of infection has not been fully delineated. At the site of attachment, the organism multiplies, producing a variety of other toxins that cause local mucosal damage (tracheal cytotoxin, dermonecrotic toxin). Impairment of host defense by *B. pertussis* is mediated by pertussis toxin and adenylate cyclase toxin. There is local cellular invasion, with intracellular bacterial persistence; however, systemic dissemination does not occur. Systemic manifestations (lymphocytosis) result from the effects of the toxins.

The pathogenesis of the clinical manifestations of pertussis is poorly understood. It is not known what causes the hallmark paroxysmal cough. A pivotal role for pertussis toxin has been proposed. Proponents of this position point to the efficacy of preventing clinical symptoms with a vaccine containing only pertussis toxoid. Detractors counter that pertussis toxin is not the critical factor because paroxysmal cough also occurs in patients infected with *B. parapertussis*, which does not produce pertussis toxin. It is thought that neurologic events in pertussis, such as seizures and encephalopathy, are due to hypoxia from coughing paroxysms or apnea rather than to the effects of specific bacterial products. *B. pertussis* pneumonia, which occurs in up to 10% of infants with pertussis, is usually a diffuse bilateral primary infection. In older children and adults with pertussis,

pneumonia is often due to secondary bacterial infection with streptococci or staphylococci.

Immunity

Both humoral and cell-mediated immunity are thought to be important in pertussis. Antibodies to pertussis toxin, filamentous hemagglutinin, pertactin, and fimbriae are all protective in animal models. Pertussis agglutinins were correlated with protection in early studies of whole-cell pertussis vaccines. Serologic correlates of protection conferred by acellular pertussis vaccines have not been established, although antibody to pertactin, fimbriae, and (to a lesser degree) pertussis toxin correlated best with protection in two efficacy trials. The duration of immunity after whole-cell pertussis vaccination is short-lived, with little protection remaining after 10–12 years. After a three-dose infant primary series of acellular pertussis vaccine, protection persists for at least 5–6 years; the duration of immunity after a four- or five-dose schedule is not yet known, but serologic and modeling studies suggest that a booster may be needed after 10 years. Although immunity after natural infection was thought to be lifelong, seroepidemiologic evidence demonstrates that it clearly is not and that subsequent episodes of clinical pertussis are prevented by intermittent subclinical infection.

Clinical Manifestations

Pertussis is a prolonged coughing illness with clinical manifestations that vary by age **(Table 148-1)**. Although not uncommon among adolescents and adults, classic pertussis is most often seen in preschool and school-age children. After an incubation period averaging 7–10 days, an illness develops that is indistinguishable from the common cold and is characterized by coryza, lacrimation, mild cough, low-grade fever, and malaise. After 1–2 weeks, this *catarrhal phase* evolves into the *paroxysmal phase*: the cough becomes more frequent and spasmodic with repetitive bursts of 5–10 coughs, often within a single expiration. Posttussive vomiting is frequent, with a mucous plug occasionally expelled at the end of an episode. The episode may be terminated by an audible whoop, which occurs upon rapid inspiration against a closed glottis at the end of a paroxysm. During a spasm, there may be impressive neck-vein distension, bulging eyes, tongue protrusion, and cyanosis. Paroxysms may be precipitated by noise, eating, or physical contact. Between attacks, the patient's appearance is normal but increasing fatigue is evident. The frequency of paroxysmal episodes varies widely, from several per hour to 5–10 per day. Episodes are often worse at night and interfere with sleep. Weight loss is not uncommon as a result of the illness's interference with eating. Most complications occur during the paroxysmal stage. Fever is uncommon and suggests bacterial superinfection.

Table 148-1 Clinical Features of Pertussis, by Age Group and Diagnostic Status

	Percentage of Patients		
	Adolescents and Adults		
Feature	**Laboratory Confirmation**	**No Laboratory Confirmation**	**Children**
Cough	95–100	95–100	95–100
Prolonged	60–80	60–80	60–95
Paroxysmal	60–90	50–90	80–95
Sleep-disturbing	50–80	50–80	90–100
Whoop	10–40	5–30	40–80
Posttussive vomiting	20–50	5–30	80–90

After 2–4 weeks, the coughing episodes become less frequent and less severe—changes heralding the onset of the *convalescent phase*. This phase can last 1–3 months and is characterized by gradual resolution of coughing episodes. For 6–12 months, intercurrent viral infections may be associated with a recrudescence of paroxysmal cough.

Not all individuals who develop pertussis have classic disease. The clinical manifestations in adolescents and adults are more often atypical. In a German study of pertussis in adults, more than two-thirds had paroxysmal cough and more than one-third had whoop. Adult illness in North America differs from this experience: the cough may be severe and prolonged but is less frequently paroxysmal, and a whoop is uncommon. Vomiting with cough is the best predictor of pertussis as the cause of prolonged cough in adults. Other predictive features are a cough at night and exposure to other individuals with a prolonged coughing illness.

Complications

Complications are frequently associated with pertussis and are more common among infants than among older children or adults. Subconjunctival hemorrhages, abdominal and inguinal hernias, pneumothoraces, and facial and truncal petechiae can result from increased intrathoracic pressure generated by severe fits of coughing. Weight loss can follow decreased caloric intake. In a series of >1100 children <2 years of age who were hospitalized with pertussis, 27.1% had apnea, 9.4% had

pneumonia, 2.6% had seizures, and 0.4% had encephalopathy; 10 children (0.9%) died. Pneumonia is reported in <5% of adolescents and adults and increases in frequency after 50 years of age. In contrast to the primary *B. pertussis* pneumonia that develops in infants, pneumonia in adolescents and adults with pertussis is usually caused by a secondary infection with encapsulated organisms such as *Streptococcus pneumoniae* or *Haemophilus influenzae*. Pneumothorax, severe weight loss, inguinal hernia, rib fracture, carotid artery aneurysm, and cough syncope have all been reported in adolescents and adults with pertussis.

Diagnosis

If the classic symptoms of pertussis are present, clinical diagnosis is not difficult. However, particularly in older children and adults, it is difficult to differentiate infections caused by *B. pertussis* and *B. parapertussis* from other respiratory tract infections on clinical grounds. Therefore, laboratory confirmation should be attempted in all cases. Lymphocytosis—an absolute lymphocyte count of $>10^8$–10^9/L—is common among young children (in whom it is unusual with other infections) but not among adolescents and adults. Culture of nasopharyngeal secretions remains the gold standard of diagnosis, although DNA detection by polymerase chain reaction (PCR) has replaced culture in many laboratories because of increased sensitivity and quicker results. The best specimen is collected by nasopharyngeal aspiration, in which a fine flexible plastic catheter attached to a 10-mL syringe is passed into the nasopharynx and withdrawn while gentle suction is applied. Since *B. pertussis* is highly sensitive to drying, secretions for culture should be inoculated without delay onto appropriate medium (Bordet-Gengou or Regan-Lowe), or the catheter should be flushed with a phosphate-buffered saline solution for culture and/or PCR. An alternative to the aspirate is a Dacron or rayon nasopharyngeal swab; again, inoculation of culture plates should be immediate or an appropriate transport medium (e.g., Regan-Lowe charcoal medium) should be used. Results of PCR can be available within hours; cultures become positive by day 5 of incubation. *B. pertussis* and *B. parapertussis* can be differentiated by agglutination with specific antisera or by direct immunofluorescence.

Nasopharyngeal cultures in untreated pertussis remain positive for a mean of 3 weeks after the onset of illness; these cultures become negative within 5 days of the institution of appropriate antimicrobial therapy. The duration of a positive PCR in untreated pertussis or after therapy is not known but exceeds that of positive cultures. Since much of the period during which the organism can be recovered from the nasopharynx falls into the catarrhal phase, when the etiology of the infection is not suspected, there is only a small window of opportunity for culture-proven diagnosis. Cultures from infants and young children are more frequently positive than those from older children and adults; this difference may reflect earlier presentation of the former age group for medical care. Direct fluorescent antibody tests of nasopharyngeal secretions for direct diagnosis may still be available in some laboratories but should not be used because of poor sensitivity and specificity. Pseudo-outbreaks of pertussis have been reported as a result of false-positive PCR results. Greater standardization of PCR methodology can alleviate this problem.

As a result of the difficulties with laboratory diagnosis of pertussis in adolescents, adults, and patients who have been symptomatic for >4 weeks, increasing attention is being given to serologic diagnosis. Enzyme immunoassays detecting IgA and IgG antibodies to pertussis toxin, filamentous hemagglutinin, pertactin, and fimbriae have been developed and assessed for reproducibility. Two- or fourfold increases in antibody titer are suggestive of pertussis, although cross-reactivity of some antigens (such as filamentous hemagglutinin and pertactin) among *Bordetella* species makes it difficult to depend diagnostically on seroconversion involving a single type of antibody. Late presentation for medical care and prior immunization also complicate serologic diagnosis because the first sample obtained may in fact be a convalescent-phase specimen. Criteria for serologic diagnosis based on comparison of results for a single serum specimen with established population values are gaining acceptance, and serologic measurement of antibody to pertussis toxin will probably become more widely standardized and available for diagnostic purposes.

Differential Diagnosis

A child presenting with paroxysmal cough, posttussive vomiting, and whoop is likely to have an infection caused by *B. pertussis* or *B. parapertussis*; lymphocytosis increases the likelihood of a *B. pertussis* etiology. Viruses such as respiratory syncytial virus and adenovirus have been isolated from patients with clinical pertussis but probably represent co-infection.

In adolescents and adults, who often do not have paroxysmal cough or whoop, the differential diagnosis of a prolonged coughing illness is more extensive. Pertussis should be suspected when any patient has a cough that does not improve within 14 days, a paroxysmal cough of any duration, a cough followed by vomiting (adolescents and adults), or any respiratory symptoms after contact with a laboratory-confirmed case of pertussis. Other etiologies to consider include infections caused by *Mycoplasma pneumoniae*, *Chlamydophila pneumoniae*, adenovirus, influenza virus, and other respiratory viruses. Use of angiotensin-converting enzyme (ACE) inhibitors, reactive airway disease, and gastroesophageal reflux disease are well-described noninfectious causes of prolonged cough in

adults.

Treatment: Pertussis

ANTIBIOTICS

The purpose of antibiotic therapy for pertussis is to eradicate the infecting bacteria from the nasopharynx; therapy does not substantially alter the clinical course unless given early in the catarrhal phase. Macrolide antibiotics are the drugs of choice for treatment of pertussis **(Table 148-2)**; macrolide-resistant *B. pertussis* strains have been reported but are rare. Trimethoprim-sulfamethoxazole is recommended as an alternative for individuals allergic to macrolides.

Table 148-2 Antimicrobial Therapy for Pertussis

Drug	Adult Daily Dose	Frequency	Duration (Days)	Comments
Erythromycin estolate	1–2 g	3 divided doses	7–14	Frequent gastrointestinal side effects
Clarithromycin	500 mg	2 divided doses	7	
Azithromycin	500 mg on day 1, 250 mg subsequently	1 daily dose	5	
Trimethoprim-sulfamethoxazole	160 mg of trimethoprim, 800 mg of sulfamethoxazole	2 divided doses	14	For patients allergic to macrolides; data on effectiveness limited

SUPPORTIVE CARE

Young infants have the highest rates of complication and death from pertussis; therefore, most infants (and older children with severe disease) should be hospitalized. A quiet environment may decrease the stimulation that can trigger paroxysmal episodes. Use of β-adrenergic agonists and/or glucocorticoids has been advocated by some authorities but has not been proven to be effective. Cough suppressants are not effective and play no role in the management of pertussis.

INFECTION CONTROL MEASURES

Hospitalized patients with pertussis should be placed in respiratory isolation, with the use of precautions appropriate for pathogens spread by large respiratory droplets. Isolation should continue for 5 days after initiation of erythromycin therapy or for 3 weeks (i.e., until nasopharyngeal cultures are consistently negative) when the patient cannot tolerate antimicrobial therapy.

Prevention

CHEMOPROPHYLAXIS

Because the risk of transmission of *B. pertussis* within households is high, chemoprophylaxis is widely recommended for household contacts of pertussis cases. The effectiveness of chemoprophylaxis, although unproven, is supported by several epidemiologic studies of institutional and community outbreaks. In the only randomized placebo-controlled study, erythromycin estolate (50 mg/kg per day in three divided doses; maximum dose, 1 g/d) was effective in reducing the incidence of bacteriologically confirmed pertussis by 67%; however, there was no decrease in the incidence of clinical disease. Despite these disappointing results, many authorities continue to recommend chemoprophylaxis, particularly in households with members at high risk of severe disease (children <1 year of age, pregnant women). Data are not available on use of the newer macrolides for chemoprophylaxis, but these drugs are commonly used because of their increased tolerability and their effectiveness.

IMMUNIZATION

(See also Chap. 122) The mainstay of pertussis prevention is active immunization. Pertussis vaccine, now available for >80 years, became widely used in North America after 1940; the reported number of pertussis cases has since fallen by >90%. Whole-cell pertussis vaccines are prepared through the heating, chemical inactivation, and purification of whole *B. pertussis* organisms. Although effective (average efficacy estimate, 85%; range for different products, 30–100%), whole-cell pertussis vaccines are associated with adverse events—both common (fever; injection-site pain, erythema, and swelling; irritability) and uncommon (febrile seizures, hypotonic hyporesponsive episodes). Alleged associations of whole-cell pertussis vaccine with encephalopathy, sudden infant death syndrome, and autism, although not substantiated, have spawned an active anti-immunization lobby. The development of acellular pertussis vaccines, which are effective but less reactogenic, has greatly alleviated concerns about the inclusion of pertussis vaccine in the combined infant immunization series.

Although whole-cell vaccines are still used extensively in developing regions of the world, acellular pertussis vaccines are used exclusively for childhood immunization in much of the developed world. In North America, acellular pertussis vaccines for children are given as a three-dose primary series at 2,

4, and 6 months of age, with a reinforcing dose at 15–18 months of age and a booster dose at 4–6 years of age.

Although a wide variety of acellular pertussis vaccines were developed, only a few are still widely marketed; all contain pertussis toxoid and filamentous hemagglutinin. One acellular pertussis vaccine also contains pertactin, and another contains pertactin and two types of fimbriae. In light of phase 3 efficacy studies, most experts have concluded that two-component acellular pertussis vaccines are more effective than monocomponent vaccines and that the addition of pertactin increases efficacy still more. The further addition of fimbriae appears to enhance protective efficacy against milder disease. In two studies, protection conferred by pertussis vaccines correlated best with the production of antibody to pertactin, fimbriae, and pertussis toxin.

Adult formulations of acellular pertussis vaccines have been shown to be safe, immunogenic, and efficacious in clinical trials in adolescents and adults and are now recommended for routine immunization of these groups in several countries, including the United States. In this country, adolescents should receive a dose of the adult-formulation diphtheria–tetanus–acellular pertussis vaccine at the preadolescent physician's visit, and all unvaccinated adults should receive a single dose of this combined vaccine. In addition, it is recommended that all health care workers in the United States be vaccinated against pertussis. Pertussis vaccine coverage among U.S. adolescents rose from 10.8% to 30.4% from 2006 to 2007. Further improvements in adolescent and adult vaccine coverage may permit better control of pertussis across the age spectrum, with collateral protection of infants too young to be immunized.

FURTHER READINGS

De Greeff SC et al: Pertussis disease burden in the household: How to protect young infants. Clin Infect Dis 50:1339, 2010

De Serres G et al: Morbidity of pertussis in adolescents and adults. J Infect Dis 182:174, 2000

Forsyth KD et al: Prevention of pertussis: Recommendations derived from the second Global Pertussis Initiative roundtable meeting. Vaccine 25:2634, 2007 [PMID: 17280745]

Halperin SA: The control of pertussis—2007 and beyond. N Engl J Med 356:110, 2007 [PMID: 17215528]

Kretsinger K et al: Preventing tetanus, diphtheria, and pertussis among adults: Use of tetanus toxoid, reduced diphtheria toxoid and acellular pertussis vaccine. Recommendations of the Advisory Committee on Immunization Practices (ACIP), supported by the Healthcare Infection Control Practices Advisory Committee (HICPAC), for use of Tdap among health-care personnel. MMWR Recomm Rep 55(RR-17):1, 2006

Lee GM et al: Pertussis in adolescents and adults: Should we vaccinate? Pediatrics 115:1675, 2005 [PMID: 15930232]

Mattoo S, Cherry JD: Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. Clin Microbiol Rev 18:326, 2005 [PMID: 15831828]

Ward JI et al: APERT Study Group. Efficacy of an acellular pertussis vaccine among adolescents and adults. N Engl J Med 353:1555, 2005 [PMID: 16221778]

Waters V et al: Outbreak of atypical pertussis detected by polymerase chain reaction in immunized preschool-aged children. Pediatr Infect Dis J 28:582, 2009 [PMID: 19561423]

Wendelboe AM et al: Infant Pertussis Study Group. Transmission of *Bordetella pertussis* to young infants. Pediatr Infect Dis J 26:293, 2007 [PMID: 17414390]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GENERAL FEATURES AND PRINCIPLES

The gram-negative enteric bacilli are common causes of a wide variety of infections involving diverse anatomic sites in both healthy and compromised hosts. Some members of this group have become increasingly resistant to antimicrobial treatment, and new infectious syndromes have emerged. Therefore, a thorough knowledge of clinical presentations and appropriate therapeutic choices is necessary for optimal outcomes.

Epidemiology

Escherichia coli, *Klebsiella*, *Proteus*, *Enterobacter*, *Serratia*, *Citrobacter*, *Morganella*, *Providencia*, and *Edwardsiella* are components of the normal animal and human colonic flora and/or of the flora of a variety of environmental habitats, including long-term-care facilities (LTCFs) and hospitals. As a result, except for certain pathotypes of intestinal pathogenic *E. coli*, these genera are global pathogens. In healthy humans, *E. coli* is the predominant species of gram-negative bacilli (GNB) in the colonic flora. GNB (primarily *E. coli*, *Klebsiella*, and *Proteus*) only transiently colonize the oropharynx and skin of healthy individuals. In contrast, in LTCF and hospital settings, a variety of GNB emerge as the dominant flora of both mucosal and skin surfaces, particularly in association with antimicrobial use, severe illness, and extended length of stay. This colonization may lead to subsequent infection; for example, oropharyngeal colonization may lead to pneumonia. In general, among adults, the incidence of infection due to these agents increases with age. Thus, as the mean age of the population increases, so will the number of these infections.

Structure and Function

GNB possess an extracytoplasmic outer membrane, a feature shared generally among gram-negative bacteria. This outer membrane consists of a lipid bilayer with associated proteins, lipoproteins, and polysaccharides [capsule, lipopolysaccharide (LPS)]. The outer membrane interfaces with the bacterial environment, including the human host. A variety of components of the outer membrane are critical determinants in pathogenesis and antimicrobial resistance.

Pathogenesis

Multiple bacterial virulence factors are required for the pathogenesis of infections caused by GNB. Possession of specialized virulence genes defines pathogens and enables them to infect the host efficiently. Hosts and their cognate pathogens have been co-adapting throughout evolutionary history, and it has been speculated that infection is just one point on the spectrum of evolved relationships between microbes and hosts. At one end of this spectrum is a commensal/symbiotic interaction (e.g., mitochondria—formerly bacteria—within eukaryotic cells); at the other end is a lethal outcome, producing a dead-end relationship (e.g., Ebola virus). During the host-pathogen "chess match" over time, various and redundant strategies have emerged in both the pathogens and their hosts that enable these partners to maintain their coexistence **(Table 149-1)**.

Table 149-1 Interactions of Extraintestinal Pathogenic *E. Coli* with the Human Host: A Paradigm for Extracellular, Extraintestinal Gram-Negative Bacterial Pathogens

Bacterial Goal	Host Obstacle	Bacterial Solution
Extraintestinal attachment	Flow of urine, mucociliary blanket	Multiple adhesins (e.g., type I, S, and F1C fimbriae; P pili)
Nutrient acquisition for growth	Nutrient sequestration (e.g., iron via intracellular storage and extracellular scavenging via lactoferrin and transferrin)	Cellular lysis (e.g., hemolysin), multiple mechanisms for competing for iron (e.g., siderophores) and other nutrients
Initial avoidance of host bactericidal activity	Complement, phagocytic cells, antimicrobial peptides	Capsular polysaccharide, lipopolysaccharide
Transmission	?	Irritant tissue damage resulting in increased excretion (e.g., toxins such as hemolysin)
Late avoidance	Acquired immunity (e.g., specific	?Cell entry, acquisition of antimicrobial

of host bactericidal activity	antibodies), treatment with antibiotics	resistance

Extraintestinal pathogenic *E. coli* (ExPEC) strains and the other genera discussed in this chapter cause infection outside the bowel. All are primarily extracellular pathogens and therefore share certain pathogenic features. Innate immunity (including the activities of complement, antimicrobial peptides, and professional phagocytes) and humoral immunity are the principal host defense components. Both susceptibility to and severity of infection are increased with dysfunction or deficiencies of these components (Chap. 119). In contrast, the virulence traits of intestinal pathogenic *E. coli*—i.e., the distinctive strains that can cause diarrheal disease—are for the most part different from those of ExPEC and other GNB that cause extraintestinal infections. This distinction reflects site-specific differences in host environments and defense mechanisms.

The virulence factors of extraintestinal pathogenic GNB subserve diverse functions. A given strain usually possesses multiple adhesins for binding to a variety of host cells (e.g., in *E. coli*: type 1, S, and F1C fimbriae; P pili). Nutrient acquisition (e.g., of iron via siderophores) requires many genes that are necessary but not sufficient for pathogenesis. The ability to resist the bactericidal activity of complement and phagocytes in the absence of antibody (e.g., as conferred by capsule or O antigen of LPS) is one of the defining traits of an extracellular pathogen. Tissue damage (e.g., as mediated by hemolysin in the case of *E. coli*) may facilitate spread within the host. Without doubt, many important virulence genes await identification, and our understanding of many aspects of the pathogenesis of infections due to GNB is in its infancy (Chap. 120).

The ability to induce septic shock is another defining feature of these genera. GNB are the most common causes of this potentially lethal syndrome. The lipid A moiety of LPS (via interaction with host Toll-like receptor 4) and probably also other bacterial factors stimulate a proinflammatory host response that, if overly exuberant, results in shock (Chap. 271).

Many antigenic variants (serotypes) exist in most genera of GNB. For example, there are >150 O-specific antigens and >80 capsular antigens in *E. coli*. This antigenic variability, which permits immune evasion and allows recurrent infection by different strains of the same species, has impeded vaccine development (Chap. 122).

Infectious Syndromes

Although certain strains of *E. coli* have evolved to be strictly intestinal pathogens, causing gastroenteritis by a variety of unique pathogenic mechanisms, extraintestinal infections are the predominant disease presentation caused by enteric GNB generally. Depending on both the host and the pathogen, nearly every organ or body cavity can be infected with GNB. *E. coli* and—to a lesser degree— *Klebsiella* and *Proteus* account for most extraintestinal infections due to GNB and are the most virulent pathogens within this group. However, the other genera are becoming increasingly important, particularly among LTCF residents and hospitalized patients. This expanding spectrum of disease-causing genera is due in large part to the intrinsic or acquired antimicrobial resistance of these organisms and the increasing number of individuals with alterations or disruptions of host defenses. The mortality rate is substantial in many GNB infections and correlates with the severity of illness. Especially problematic are pneumonia and bacteremia (arising from any source) when complicated by organ failure (severe sepsis) and/or shock, for which the associated mortality rates are 20–50%.

Diagnosis

Isolation of GNB from ordinarily sterile anatomic sites almost always implies infection, whereas their isolation from nonsterile sites, particularly from open soft-tissue wounds and the respiratory tract, requires clinical correlation to differentiate colonization from infection. Tentative laboratory identification based on lactose fermentation and indole production (described for each genus below), which usually is possible before final identification of the organism and determination of its antimicrobial susceptibilities, may guide empirical antimicrobial therapy.

Treatment: Infections Caused by Gram-Negative Enteric Bacilli

(See also Chap. 133) In this chapter, the Clinical Laboratory Standards Institute (CLSI) classification of cephalosporins will be used, according to which the previously designated first-, second-, third-, and fourth-generation cephalosporins will be designated cephalosporins I, II, III, and IV, respectively. Likewise, the terms *extended* or *expanded spectrum*, which have been used to describe third- and fourth-generation cephalosporins, will be avoided. Accumulating evidence indicates that initiation of appropriate empirical antimicrobial therapy early in the course of GNB infections (particularly serious infections) leads to improved outcomes. Familiarity with evolving patterns of antimicrobial resistance in enteric GNB is necessary in the selection of appropriate empirical therapy, particularly given the lag between published and real-time resistance rates and the ever-increasing prevalence of multidrug-resistant (MDR) GNB. However, if broad-spectrum treatment has been initiated and information on antimicrobial susceptibilities becomes available, it is just as important to use the most appropriate

narrower-spectrum agent. Such responsible antimicrobial stewardship will avoid unnecessary selection of and potential superinfection with resistant bacteria, may decrease costs, and will maximize the useful longevity of available antimicrobial agents. Likewise, it is important not to treat patients who are colonized but not infected. The antimicrobial resistance profiles of GNB vary by species, geographic location, regional antimicrobial use, and hospital site [e.g., intensive care units (ICUs) versus wards]. At present, the most reliably active agents against enteric GNB are the carbapenems (e.g., imipenem), the aminoglycoside amikacin, the cephalosporin IV cefepime, and piperacillin-tazobactam.

β-Lactamases, which inactivate β-lactam agents, are the most important mediators of resistance to these drugs in GNB. Decreased permeability and/or active efflux of β-lactam agents, although less common, may occur alone or in combination with β-lactamase-mediated resistance. *Broad-spectrum* β-lactamases, which mediate resistance to many penicillins and cephalosporins I, are frequently expressed in enteric GNB. These enzymes are inhibited by agents such as clavulanate. *Extended-spectrum* β-lactamases (ESBLs) confer resistance to the same drugs as broad-spectrum β-lactamases as well as to cephalosporins III, aztreonam, and (in some instances) cephalosporins IV. The prevalence of acquired ESBL-encoding genes via transferable plasmids is increasing in GNB worldwide, with rates varying greatly even among hospitals in a given region. To date, ESBLs are most prevalent in *Klebsiella pneumoniae*, *K. oxytoca*, and *E. coli* but also occur (and are probably underrecognized) in *Enterobacter*, *Citrobacter*, *Proteus*, *Serratia*, and other enteric GNB. At present, the regional prevalence of ESBL-producing GNB declines in rank order as follows: Latin America > Western Pacific > Europe > United States and Canada. ESBL-producing GNB were initially described in hospitals (ICUs > wards) and LTCFs. However, over the last decade, CTX-M ESBLs have been increasingly described in community-acquired strains. Hospital outbreaks due to ESBL-producing strains have been associated with extensive use of cephalosporins III, particularly ceftazidime. The carbapenems are the most reliably active β-lactam agents against ESBL-expressing strains. GNB that express ESBLs may also possess porin mutations that result in decreased uptake of cephalosporins and β-lactam/β-lactamase inhibitor combinations. Thus, ESBL-producing isolates should be considered resistant to all penicillins, cephalosporins, and aztreonam. Ceftobiprole, which is currently under review by the U.S. Food and Drug Administration (FDA), is a first-in-class cephalosporin with activity against methicillin-resistant *Staphylococcus aureus* and most Enterobacteriaceae; however, it has poor activity in vitro against ESBL-producing GNB. Oral options for the treatment of strains expressing CTX-M ESBLs are limited (see "Treatment of Extraintestinal *E. coli* Infections," below).

AmpC β-lactamases confer resistance to the same substrates as ESBLs plus the cephamycins (e.g., cefoxitin and cefotetan). AmpC enzymes resist inhibition by β-lactamase inhibitors. The presence of constitutive chromosomal AmpC β-lactamases in nearly all strains of *Enterobacter*, *Serratia*, *Citrobacter*, *Proteus vulgaris*, *Providencia*, and *Morganella* results in resistance to aminopenicillins, cefazolin, and cefoxitin. In addition, some strains of *E. coli*, *K. pneumoniae*, and other Enterobacteriaceae have acquired plasmids containing AmpC β-lactamase genes. The cephalosporin IV cefepime is stable to AmpC β-lactamases and is an appropriate treatment option if the concomitant presence of an ESBL can be excluded.

Carbapenemases (e.g., the IMP, VIM, and KPC families) confer resistance to the same drugs as ESBLs plus cephamycins and carbapenems. Linked resistance to fluoroquinolones and aminoglycosides is common. Occasionally, resistance to carbapenems is due to the possession of a β-lactamase plus decreased permeability. Unfortunately, carbapenemase-producing enteric GNB are becoming increasingly common, and infection with these strains is associated with elevated mortality rates. Automated susceptibility systems may be unreliable for detection of carba-penemases, particularly those conferring resistance to imipenem and meropenem. Resistance to ertapenem is the most sensitive marker for carbapenem resistance in automated systems. An elevated carbapenem minimal inhibitory concentration should prompt additional polymerase chain reaction (PCR) testing for resistance genes or a modified Hodge test; these methods are more reliable for detection of carbapenemase-producing strains. Tigecycline and the polymyxins exhibit the greatest in vitro activity against such strains. However, tigecycline reaches only low concentrations in serum and urine, a characteristic that warrants concern about its use in the treatment of bacteremia and urinary tract infection (UTI). Furthermore, emerging resistance to both of these agents poses the risk of onset of a post-antimicrobial era with respect to GNB.

Resistance to fluoroquinolones usually is due to alterations of the target site (DNA gyrase and/or topoisomerase IV), with or without decreased permeability, active efflux, or protection of the target site. Resistance to fluoroquino-lones is increasingly prevalent among GNB and is associated with resistance to other antimicrobial classes; for example, 20–80% of ESBL-producing enteric GNB are also resistant to fluoroquinolones. At present, fluoroquinolones should be considered unreliable as empirical therapy for infections due to GNB in critically ill patients.

Given the increasing prevalence of MDR GNB, it is reasonable—pending susceptibility results—to combine agents for empirical treatment of GNB infections in critically ill patients. Antimicrobial resistance may not always be identified by in vitro testing; therefore, it is important to assess the clinical response to treatment. Moreover, resistance may evolve during therapy (e.g., through stable

derepression of *AmpC* β-lactamases). In addition, drainage of abscesses and removal of infected foreign bodies are often required for cure. GNB are commonly involved in polymicrobial infections, in which the role of each individual pathogen is uncertain (Chap. 164). Although some GNB are more pathogenic than others, it is usually prudent, if possible, to design an antimicrobial regimen active against all of the GNB identified, since each is capable of pathogenicity in its own right. Finally, the possibility of a superinfection (e.g., *Clostridium difficile* colitis) must always be kept in mind.

Prevention

(See also Chap. 131) Diligent adherence to hand-hygiene protocols by health care personnel and avoidance of inappropriate antimicrobial use are key measures in preventing infection and the further development of antimicrobial resistance. Contact precautions should be implemented for patients colonized or infected with carbapenem-resistant (and perhaps other MDR) GNB. Likewise, avoidance of the use of indwelling devices (e.g., urinary and intravascular catheters, endotracheal tubes) and, when they are necessary, placement according to an appropriate protocol decrease infection risk. Positioning (e.g., head of bed at ≥30°) and good oral hygiene decrease the incidence of pneumonia in ventilated patients.

ESCHERICHIA COLI INFECTIONS

Commensal Strains

For the most part, commensal *E. coli* variants, which constitute the bulk of the normal facultative intestinal flora in most humans, confer benefits to the host (e.g., resistance to colonization with pathogenic organisms). These strains generally lack the specialized virulence traits that enable extraintestinal and intestinal pathogenic *E. coli* strains to cause disease outside and within the gastrointestinal tract, respectively. However, even commensal *E. coli* strains can be involved in extraintestinal infections in the presence of an aggravating factor, such as a foreign body (e.g., a urinary catheter), host compromise (e.g., local anatomic or functional abnormalities such as urinary or biliary tract obstruction or systemic immunocompromise), or an inoculum that is large or contains a mixture of bacterial species (e.g., fecal contamination of the peritoneal cavity).

Extraintestinal Pathogenic (ExPEC) Strains

The majority of *E. coli* isolates from symptomatic infections of the urinary tract, bloodstream, cerebrospinal fluid, respiratory tract, and peritoneum (spontaneous bacterial peritonitis) can be differentiated from commensal and intestinal pathogenic strains of *E. coli* by virtue of their distinctive virulence factor profiles (Tables 149-1 and 149-2) and phylogenetic background. ExPEC strains can also cause surgical wound infection, osteomyelitis, and myositis, but the number of cases evaluated to date is too small for a reliable assessment of proportions.

Table 149-2 Intestinal Pathogenic *E. Coli*

Pathotype[a]	Epidemiology	Clinical Syndrome[b]	Defining Molecular Trait	Responsible Genetic Element[c]
STEC/EHEC	Food, water, person-to-person; all ages, industrialized countries	Hemorrhagic colitis, hemolytic-uremic syndrome	Shiga toxin	Lambda-like Stx1- or Stx2-encoding bacteriophage
ETEC	Food, water; young children in and travelers to developing countries	Traveler's diarrhea	Heat-stable and -labile enterotoxins, colonization factors	Virulence plasmid(s)
EPEC	Person-to-person; young children and neonates in developing countries	Watery diarrhea, persistent diarrhea	Localized adherence, attaching and effacing lesion on intestinal epithelium	EPEC adherence factor plasmid pathogenicity island [locus for enterocyte effacement (LEE)]
EIEC	Food, water; children in and travelers to developing countries	Dysentery	Invasion of colonic epithelial cells, intracellular multiplication, cell-to-cell spread	Multiple genes contained primarily in a large virulence plasmid
EAEC	?Food, water; children in and travelers to developing ccountries; all ages, industrialized countries	Traveler's diarrhea, acute diarrhea, persistent	Aggregative/diffuse adherence, virulence factors regulated by AggR	Chromosomal or plasmid-associated adherence and toxin genes

		diarrhea		

[a]EAEC, enteroadherent *E. coli*; EHEC, enterohemorrhagic *E. coli*; EIEC, enteroinvasive *E. coli*; EPEC, enteropathogenic *E. coli*; ETEC, enterotoxigenic *E. coli*; STEC, Shiga toxin–producing *E. coli*.

[b]Classic syndromes; see text for details on disease spectrum.

[c]Pathogenesis involves multiple genes, including genes in addition to those listed.

Like commensal *E. coli* (but in contrast to intestinal pathogenic *E. coli*), ExPEC strains are often found in the intestinal flora of healthy individuals and do not cause gastroenteritis in humans. Although acquisition of an ExPEC strain by the host is a prerequisite for ExPEC infection, it is not the rate-limiting step, which instead is entry of an ExPEC strain from its site of colonization (e.g., the colon, vagina, or oropharynx) into a normally sterile extraintestinal site (e.g., the urinary tract, peritoneal cavity, or lungs). ExPEC strains have acquired genes encoding diverse extraintestinal virulence factors that enable the bacteria to cause infections outside the gastrointestinal tract in both normal and compromised hosts (Table 149–1). These virulence genes are, for the most part, distinct from those that enable intestinal pathogenic strains to cause diarrheal disease. All age groups, all types of hosts, and nearly all organs and anatomic sites are susceptible to infection by ExPEC. Even previously healthy hosts can become severely ill or die when infected with ExPEC; however, adverse outcomes are more common among hosts with comorbid illnesses and host defense abnormalities. *E. coli* is the most common enteric GNB to cause extraintestinal infection in ambulatory, LTCF, and hospital settings. The diversity and the medical and economic impact of ExPEC infections are evident from consideration of the following specific syndromes.

EXTRAINTESTINAL INFECTIOUS SYNDROMES

Urinary Tract Infection

The urinary tract is the site most frequently infected by ExPEC. An exceedingly common infection among ambulatory patients, UTI accounts for 1% of ambulatory care visits in the United States and is second only to lower respiratory tract infection among infections responsible for hospitalization. UTIs are best considered by clinical syndrome (e.g., uncomplicated cystitis, pyelonephritis, and catheter-associated UTIs) and within the context of specific hosts (e.g., premenopausal women, compromised hosts; Chap. 288). *E. coli* is the single most common pathogen for all UTI syndrome/host group combinations. Each year in the United States, *E. coli* causes 85–95% of an estimated 6–8 million episodes of uncomplicated cystitis in premenopausal women, with an estimated $1.6 billion in direct health care costs. Furthermore, 20% of women with an initial cystitis episode develop frequent recurrences (from 0.3 to >20 per year).

Uncomplicated cystitis, the most common acute UTI syndrome, is characterized by dysuria, urinary frequency, and suprapubic pain. Fever and/or back pain suggests progression to pyelonephritis. Even with appropriate treatment of pyelonephritis, fever may take 5–7 days to resolve completely. Persistently elevated or increasing fever and neutrophil counts should prompt evaluation for intrarenal or perinephric abscess and/or obstruction. Renal parenchymal damage and loss of renal function during pyelonephritis occur primarily with urinary obstruction. Pregnant women are at unusually high risk for developing pyelonephritis, which can adversely affect the outcome of pregnancy. As a result, prenatal screening for and treatment of asymptomatic bacteriuria are standard. Prostatic infection is a potential complication of UTI in men. The diagnosis and treatment of UTI, as detailed in Chap. 288, should be tailored to the individual host, the nature and site of infection, and local patterns of antimicrobial susceptibility.

Abdominal and Pelvic Infection

The abdomen/pelvis is the second most common site of extraintestinal infection due to *E. coli*. A wide variety of clinical syndromes occur in this location, including acute peritonitis secondary to fecal contamination, spontaneous bacterial peritonitis, dialysis-associated peritonitis, diverticulitis, appendicitis, intraperitoneal or visceral abscesses (hepatic, pancreatic, splenic), infected pancreatic pseudocysts, and septic cholangitis and/or cholecystitis. In intraabdominal infections, *E. coli* can be isolated either alone or (as often occurs) in combination with other facultative and/or anaerobic members of the intestinal flora (Chap. 127).

Pneumonia

E. coli is not usually considered a cause of pneumonia (Chap. 257). Indeed, enteric GNB account for only 2–5% of cases of community-acquired pneumonia (CAP), in part because these organisms only transiently colonize the oropharynx in a minority of healthy individuals. However, rates of oral colonization with *E. coli* and other GNB increase with severity of illness and antibiotic use. Consequently, GNB are a common cause of pneumonia among residents of LTCFs and are the most common cause (60–70% of cases) of hospital-acquired pneumonia (HAP) (Chap. 127), particularly among postoperative and ICU patients (e.g., ventilator-associated pneumonia). Pulmonary infection is usually acquired by small-volume aspiration but occasionally occurs via hematogenous spread, in

which case multifocal nodular infiltrates can be seen. Tissue necrosis, probably due to cytotoxins produced by GNB, is common. Despite significant institutional variation, *E. coli* is generally the third or fourth most commonly isolated GNB in hospital-acquired pneumonia, accounting for 5–8% of episodes in both U.S.-based and European-based studies. Regardless of the host, pneumonia due to enteric GNB is a serious disease, with high crude and attributable mortality rates (20–60% and 10–20%, respectively).

Meningitis

(See also Chap. 381) *E. coli* is one of the two leading causes of neonatalmeningitis, the other being group B *Streptococcus*. Most *E. coli* strains that cause neonatal meningitis possess the K1 capsular antigen and derive from a limited number of familiar meningitis-associated clonal groups. After the first month of life, *E. coli* meningitis is uncommon, occurring predominantly in the setting of disruption of the meninges from craniotomy or trauma or in the presence of cirrhosis. In patients with cirrhosis who develop meningitis, the meninges are presumably seeded as a result of poor hepatic clearance of portal vein bacteremia.

Cellulitis/Musculoskeletal Infection

E. coli contributes frequently to infection of decubitus ulcers and occasionally to infection of ulcers and wounds of the lower extremity in diabetic patients and other hosts with neurovascular compromise. Osteomyelitis secondary to contiguous spread can occur in these settings. *E. coli* also causes cellulitis or infections of burn sites or surgical wounds (accounting, in fact, for 10% of surgical site infections), particularly when the infection originates close to the perineum. Hematogenously acquired osteomyelitis, especially of vertebral bodies, is more commonly caused by *E. coli* than is generally appreciated; this organism accounts for up to 10% of cases in some series (Chap. 126). *E. coli* occasionally causes orthopedic device–associated infection or septic arthritis and rarely causes hematogenous myositis. Upper-leg myositis or fasciitis due to *E. coli* should prompt an evaluation for an abdominal source with contiguous spread.

Endovascular Infection

Despite being one of the most common causes of bacteremia, *E. coli* rarely seeds native heart valves. When the organism does seed native valves, it usually does so in the setting of prior valvular disease. *E. coli* infections of aneurysms and vascular grafts are quite uncommon.

Miscellaneous Infections

E. coli can cause infection in nearly every organ and anatomic site. It occasionally causes postoperative mediastinitis or complicated sinusitis and uncommonly causes endophthalmitis or brain abscess.

Bacteremia

E. coli bacteremia can arise from primary infection at any extraintestinal site. In addition, primary *E. coli* bacteremia can arise from percutaneous intravascular devices or transrectal prostate biopsy or from the increased intestinal mucosal permeability seen in neonates and in the settings of neutropenia and chemotherapy-induced mucositis, trauma, and burns. Roughly equal proportions of *E. coli* bacteremia cases originate in the community and in the hospital. In most studies, *E. coli* and *S. aureus* are the two most common blood isolates of clinical significance. *E. coli*, which accounts for 17–37% of cases, is the most frequent GNB blood isolate in the ambulatory setting and in most LTCF and hospital settings. Isolation of *E. coli* from the blood is almost always clinically significant and is typically accompanied by the sepsis syndrome, severe sepsis (sepsis-induced dysfunction of at least one organ or system), or septic shock (Chap. 271). Calculations based on conservative estimates for the incidence of severe sepsis (0.76/1000), the proportional contribution of E. coli to severe sepsis (17%), and a sepsis-associated mortality rate of 30% translate into an estimated 265,000 deaths annually in the world (2009 census data).

The urinary tract is the most common source of *E. coli* bacteremia, accounting for one-half to two-thirds of episodes. Bacteremia from a urinary tract source is particularly common in patients with pyelonephritis, urinary tract obstruction, or urinary instrumentation in the presence of infected urine. The abdomen is the second most common source, accounting for 25% of episodes. Although biliary obstruction (stones, tumor) and overt bowel disruption, which typically are readily apparent, are responsible for many of these cases, some abdominal sources (e.g., abscesses) are remarkably silent clinically and require identification via imaging studies (e.g., CT). Therefore, the physician should be cautious in designating the urinary tract as the source of *E. coli* bacteremia in the absence of characteristic signs and symptoms of UTI. Soft tissue, bone, pulmonary, and intravascular catheter infections are other sources of *E. coli* bacteremia.

DIAGNOSIS

Strains of *E. coli* that cause extraintestinal infections usually grow both aerobically and anaerobically within 24 h on standard diagnostic media and are easily identified by the clinical microbiology laboratory according to routine biochemical criteria. More than 90% of ExPEC strains are rapid lactose

fermenters and are indole positive.

Treatment: Extraintestinal *E. Coli* Infections

In the past, most *E. coli* isolates were highly susceptible to a broad range of antimicrobial agents. Unfortunately, this situation has changed, and, of the Enterobacteriaceae, *E. coli* is the species in which resistance is evolving most rapidly. In general, the high prevalence of resistance precludes empirical use of ampicillin and amoxicillin-clavulanate, even for community-acquired infections. The prevalence of resistance to cephalosporins I and trimethoprim-sulfamethoxazole (TMP-SMX) is increasing among community-acquired strains in the United States (with current rates of 10–40%) and is even higher outside North America. Until recently, TMP-SMX was the drug of choice for the treatment of uncomplicated cystitis in many locales. Although continued empirical use of TMP-SMX will predictably result in ever-diminishing cure rates, a wholesale switch to alternative agents (e.g., fluoroquinolones) will just as predictably accelerate the widespread emergence of resistance to these antimicrobial classes, as has already occurred in some areas. More than 90% of isolates that cause uncomplicated cystitis remain susceptible to nitrofurantoin and fosfomycin. The prevalence of resistance to fluoroquinolones has increased steadily over the last decade (e.g., from 5% to 20% in North America between 2002 and 2005 and from 8% to 25% among bacteremia isolates from the United Kingdom and Ireland between 2001 and 2006) and is even higher in other regions (Mexico, India). Prevalence figures are higher in settings where fluoroquinolone prophylaxis is used extensively (e.g., in patients with leukemia, transplant recipients, and patients with cirrhosis) and among isolates from LTCFs and hospitals. As for cephalosporin resistance, data from the U.S. National Healthcare Safety Network (NHSN) indicated that just 6% of *E. coli* device-associated/surgical site infections were due to strains resistant to cephalosporins III in 2006–2007. However, significantly higher rates have been reported outside North America; 54% of isolates assessed by the International Nosocomial Infection Control Consortium (INICC) in 2002–2007 were resistant. ESBL-containing strains are increasingly prevalent among both health care–associated (5–10%) and ambulatory isolates (region-dependent figures). An increasing number of reports describe *E. coli* strains causing community-acquired UTIs that contain CTX-M ESBLs. Data suggest that acquisition of CTX-M-containing and fluoroquinolone-resistant strains may result from consumption of meat products from food animals treated with cephalosporins III and IV and fluoroquinolones. Oral treatment options are limited with these strains; however, in vitro and limited clinical data indicate that fosfomycin and—for cystitis—nitrofurantoin appear to be viable options. Carbapenems (e.g., imipenem) and amikacin are the most predictably active agents overall, but carbapenemase-producing strains are on the rise (1–5% among health care–associated isolates). Tigecycline and polymyxin B have been used most frequently for these nearly panresistant isolates. Although this evolving antimicrobial resistance is a source of serious concern, of equal importance is the need to use the most appropriate narrower-spectrum agent whenever possible and to avoid treating colonized but uninfected patients so that the ever-escalating selection of increasingly resistant bacteria is not unnecessarily fueled.

Intestinal Pathogenic Strains

Certain strains of *E. coli* are capable of causing diarrheal disease. Other important intestinal pathogens are discussed in Chaps. 128, 142, 153, 154, 155, and 156. At least in the industrialized world, intestinal pathogenic strains of *E. coli* are rarely encountered in the fecal flora of healthy persons and instead appear to be essentially obligate pathogens. These strains have evolved a special ability to cause enteritis, enterocolitis, and colitis when ingested in sufficient quantities by a naive host. At least five distinct pathotypes of intestinal pathogenic *E. coli* exist: (1) Shiga toxin–producing *E. coli* (STEC)/enterohemorrhagic *E. coli* (EHEC), (2) enterotoxigenic *E. coli* (ETEC), (3) enteropathogenic *E. coli* (EPEC), (4) enteroinvasive *E. coli* (EIEC), and (5) enteroaggregative *E. coli* (EAEC). Diffusely adherent *E. coli* (DAEC) and cytodetaching *E. coli* are additional putative pathotypes. Transmission occurs predominantly via contaminated food and water for ETEC, STEC/EHEC, EIEC, and EAEC and by person-to-person spread for EPEC (and occasionally STEC/EHEC). Gastric acidity confers some protection against infection; therefore, persons with decreased stomach acid levels are especially susceptible. Humans are the major reservoir (except for STEC/EHEC, with regard to which bovines are the main concern); host range appears to be dictated by species-specific attachment factors. Although there is some overlap, each pathotype possesses a unique combination of virulence traits that results in a distinctive intestinal pathogenic mechanism (Table 149-2). These strains are largely incapable of causing disease outside the intestinal tract. Except in the cases of STEC/EHEC and EAEC, disease due to this group of pathogens occurs primarily in developing countries.

Shiga Toxin–Producing And Enterohemorrhagic *E. Coli*

STEC/EHEC strains constitute an emerging group of pathogens that can cause hemorrhagic colitis and the hemolytic-uremic syndrome (HUS). Several large outbreaks resulting from the consumption of fresh produce (e.g., lettuce, spinach, sprouts) and of undercooked ground beef have received significant attention in the media. O157:H7 is the most prominent serotype, but serogroups O6, O26, O55, O91, O103, O111, O113, and OX3 have also been associated with these syndromes. The ability of STEC/EHEC to produce Shiga toxin (Stx2 and/or Stx1) or related toxins is a critical factor in the

expression of clinical disease. *Shigella dysenteriae* strains that produce the closely related Shiga toxin Stx can cause the same syndrome. Stx2 and its Stx2C variant (which may be variably present in combination with Stx2 and/or Stx1) appear to be more important than Stx1 in the development of HUS. All Shiga toxins studied to date are multimers composing one enzymatically active A subunit and five identical B subunits that mediate binding to globosyl ceramides, which are membrane-associated glycolipids expressed on certain host cells. The Stx1 A subunit cleaves an adenine from the host cell's 28S rRNA, thereby irreversibly inhibiting ribosomal function, whereas the Stx2 A subunit inactivates *Bcl2*, inducing apoptosis.

Additional properties, such as acid tolerance and adherence, are necessary for full pathogenicity among STEC strains. Most disease-causing isolates possess the chromosomal locus for enterocyte effacement (LEE). This pathogenicity island was first described in EPEC strains and contains genes that mediate adherence to intestinal epithelial cells. EHEC strains make up the subgroup of STEC strains that possess stx_1 and/or stx_2 as well as LEE.

Domesticated ruminant animals, particularly cattle and young calves, serve as the major reservoir for STEC/EHEC. Ground beef—the most common food source of STEC/EHEC strains—is often contaminated during processing. Furthermore, manure from cattle or other animals (including that in the form of fertilizer) can contaminate produce (potatoes, lettuce, spinach, sprouts, fallen apples), and fecal runoff from this source can contaminate water systems. Petting zoos are another source of infection. It is estimated that $<10^2$ CFU of STEC/EHEC can cause disease. Therefore, not only can low levels of food or environmental contamination (e.g., in water swallowed while swimming) result in disease, but person-to-person transmission (e.g., at day-care centers and in institutions) is an important route for secondary spread. Laboratory-associated infections also take place. Illness due to this group of pathogens occurs both as outbreaks and as sporadic cases, with a peak incidence in the summer months.

In contrast to other intestinal pathotypes, STEC/EHEC causes infections more frequently in industrialized countries than in developing regions. O157:H7 strains are the fourth most commonly reported cause of bacterial diarrhea in the United States (after *Campylobacter*, *Salmonella*, and *Shigella*). Colonization of the colon and perhaps the ileum results in symptoms after an incubation period of 3 or 4 days. Colonic edema and an initial nonbloody secretory diarrhea may develop into the STEC/EHEC hallmark syndrome of grossly bloody diarrhea (as detected by history or examination) in >90% of cases. Significant abdominal pain and fecal leukocytes are common (70% of cases), whereas fever is not; absence of fever can incorrectly lead to consideration of noninfectious conditions (e.g., intussusception and inflammatory or ischemic bowel disease). Occasionally, infections caused by *C. difficile*, *K. oxytoca* (see "*Klebsiella* Infections," below), *Campylobacter*, and *Salmonella* present in a similar fashion. STEC/EHEC disease is usually self-limited, lasting 5–10 days. An uncommon but feared complication of this infection is HUS, which occurs 2–14 days after diarrhea in 2–8% of cases, most often affecting very young or elderly patients. It is estimated that >50% of all cases of HUS in the United States and 90% of HUS cases in children are caused by STEC/EHEC. This complication is probably mediated by the systemic translocation of Shiga toxins. Erythrocytes may serve as carriers of Stx to endothelial cells located in the small vessels of the kidney and brain. The subsequent development of thrombotic microangiopathy (perhaps with direct toxin-mediated effects on various non-endothelial cells) commonly produces some combination of fever, thrombocytopenia, renal failure, and encephalopathy. Although the mortality rate with dialysis support is <10%, residual renal and neurologic dysfunction may persist.

Enterotoxigenic *E. Coli*

In tropical or developing countries, ETEC is a major cause of endemic diarrhea. After weaning, children in these locales commonly experience several episodes of ETEC infection during the first 3 years of life. The incidence of disease diminishes with age, a pattern that correlates with the development of mucosal immunity to colonization factors (i.e., adhesins). In industrialized countries, infection usually follows travel to endemic areas, although occasional foodborne outbreaks occur. ETEC is the most common agent of traveler's diarrhea, causing 25–75% of cases. The incidence of infection may be decreased by prudent avoidance of potentially contaminated fluids and foods (Chap. 123). ETEC infection is uncommon in the United States, but outbreaks secondary to consumption of food products imported from endemic areas have occurred. A large inoculum (10^6–10^{10} CFU) is needed to produce disease. After ingestion of contaminated water or food (particularly items that are poorly cooked, unpeeled, or unrefrigerated), colonization factor–mediated intestinal adherence occurs over 12–72 h.

Disease is mediated primarily by a heat-labile toxin (LT-1) and/or a heat-stable toxin (STa) that causes net fluid secretion via activation of adenylate cyclase (LT-1) and/or guanylate cyclase (STa) in the jejunum and ileum. The result is watery diarrhea accompanied by cramps. LT-1 consists of an A and a B subunit and is structurally and functionally similar to cholera toxin. Strong binding of the B subunit to the GM_1 ganglioside on intestinal epithelial cells leads to the intracellular translocation of the A subunit, which functions as an ADP-ribosyltransferase. Mature STa is an 18- or 19-amino-acid

secreted peptide whose biologic activity is mediated by binding to the guanylate cyclase C found in the brush-border membrane of enterocytes; this binding results in increased intracellular concentrations of cyclic GMP. Characteristically absent in ETEC-mediated disease are histopathologic changes within the small bowel; mucus, blood, and inflammatory cells in stool; and fever. The disease spectrum ranges from a mild illness to a life-threatening cholera-like syndrome. Although symptoms are usually self-limited (typically lasting for 3 days), infection may result in significant morbidity and mortality (mostly from profound volume depletion) when access to health care or suitable rehydration fluids is limited and when small and/or undernourished children are affected.

Enteropathogenic *E. Coli*

EPEC causes disease primarily in young children, including neonates. The first *E. coli* pathotype recognized as an agent of diarrheal disease, EPEC was responsible for outbreaks of infantile diarrhea (including some outbreaks in hospital nurseries) in industrialized countries in the 1940s and 1950s. At present, EPEC infection is an uncommon cause of diarrhea in developed countries but is an important cause of diarrhea (both sporadic and epidemic) among infants in developing countries. Breast-feeding diminishes the incidence of EPEC infection. Rapid person-to-person spread may occur. Upon colonization of the small bowel, symptoms develop after a brief incubation period (1 or 2 days). Initial localized adherence leads to a characteristic effacement of microvilli, with the formation of cuplike, actin-rich pedestals. The actual mechanisms of diarrhea production are an area of ongoing investigation. Diarrheal stool often contains mucus but not blood. Although usually self-limited (lasting 5–15 days), EPEC diarrhea may persist for weeks.

Enteroinvasive *E. Coli*

EIEC, a relatively uncommon cause of diarrhea, is rarely identified in the United States, although a few food-related outbreaks have been described. In developing countries, sporadic disease is infrequently recognized in children and travelers. EIEC shares many genetic and clinical features with *Shigella*; however, unlike *Shigella*, EIEC produces disease only at a large inoculum (10^8–10^{10} CFU), with onset generally following an incubation period of 1–3 days. Initially, enterotoxins are believed to induce secretory small-bowel diarrhea. Subsequently, colonization and invasion of the colonic mucosa, followed by replication therein and cell-to-cell spread, result in the development of inflammatory colitis characterized by fever, abdominal pain, tenesmus, and scant stool containing mucus, blood, and inflammatory cells. Symptoms are usually self-limited (7–10 days).

Enteroaggregative and Diffusely Adherent *E. Coli*

EAEC has been described primarily in developing countries and in young children. However, recent studies indicate that it may be a relatively common cause of diarrhea in all age groups in industrialized countries. EAEC has also been recognized increasingly as an important cause of traveler's diarrhea. A large inoculum is required for infection, which manifests as watery and sometimes persistent diarrhea in both healthy and HIV-infected hosts. In vitro, the organisms exhibit a diffuse or "stacked-brick" pattern of adherence to epithelial cells. Virulence factors that probably are necessary for disease are regulated in part by the transcriptional activator AggR and include the aggregative adherence fimbriae (AAF/I-III); the Hda adhesin; surface protein dispersion; and the enterotoxins Pet, EAST-1, ShET1, and ShET2. Some strains of DAEC are capable of causing diarrheal disease, primarily in children 2–6 years of age in some developing countries, and may perhaps cause traveler's diarrhea. The Afa/Dr adhesins may contribute to the pathogenesis of infection.

Diagnosis

A practical approach to the evaluation of diarrhea is to distinguish noninflammatory from inflammatory cases (Chap. 128). ETEC, EPEC, and DAEC are uncommon causes of noninflammatory diarrhea in the United States; the incidence of EAEC infection in this country may be underrecognized. The diagnosis of these infections requires specialized assays (e.g., PCR-based tests for pathotype-specific genes) that are not routinely available and are rarely needed since the diseases are self-limited. ETEC causes the majority and EAEC a minority of cases of noninflammatory traveler's diarrhea. Definitive diagnosis generally is not necessary. Empirical antimicrobial (or symptom-based) treatment, along with rehydration therapy, is a reasonable approach. If diarrhea persists despite treatment, *Giardia* or *Cryptosporidium* (or, in immunocompromised hosts, certain other microbial agents) should be sought. The diagnosis of infection with EIEC, a rare cause of inflammatory diarrhea in the United States, also requires specialized assays. However, evaluation for STEC/EHEC infection, particularly when bloody diarrhea is reported or observed, is appropriate. Although the most common method currently used to detect STEC/EHEC is to screen for *E. coli* strains that do not ferment sorbitol, with subsequent serotyping for O157, testing for Shiga toxins or toxin genes is more sensitive, specific, and rapid. The latter approach offers the added advantage of detecting both non-O157 STEC/EHEC strains and sorbitol-fermenting strains of O157:H7, which otherwise are difficult to identify. DNA-based, enzyme-linked immunosorbent, and cytotoxicity assays are in various stages of development and are emerging as the diagnostic methods of choice.

Treatment: Intestinal E. *Coli* Infections

(See also Chap. 128) The mainstay of treatment for all diarrheal syndromes is replacement of water and electrolytes. The use of prophylactic antibiotics to prevent traveler's diarrhea generally should be discouraged, especially in light of high rates of antimicrobial resistance. However, in selected patients (e.g., those who cannot afford a brief illness or have an increased susceptibility to infection), the use of rifaximin, which is nonabsorbable and is well tolerated, is reasonable. When stools are free of mucus and blood, early patient-initiated treatment of traveler's diarrhea with a fluoroquinolone or azithromycin decreases the duration of illness, and the use of loperamide may halt symptoms within a few hours. Although dysentery caused by EIEC is self-limited, treatment hastens the resolution of symptoms, particularly in severe cases. In contrast, antimicrobial therapy for STEC/EHEC infection (the presence of which is suggested by grossly bloody diarrhea without fever) should be avoided, since antibiotics may increase the incidence of HUS (possibly via increased production/release of Stx).

KLEBSIELLA INFECTIONS

K. pneumoniae is the most important *Klebsiella* species from a medical standpoint, causing community-acquired, LTCF-acquired, and nosocomial infections. *K. oxytoca* is primarily a pathogen in LTCF and hospital settings. *Klebsiella* species are broadly prevalent in the environment and colonize mucosal surfaces of mammals. In healthy humans, the prevalence of *K. pneumoniae* colonization is 5–35% in the colon and 1–5% in the oropharynx; the skin is usually colonized only transiently. In LTCFs and hospitals, colonization with *K. oxytoca* also occurs, and carriage rates are substantial among both staff and patients. Person-to-person spread is the predominant mode of acquisition. Most *Klebsiella* infections due to "classic" *K. pneumoniae* (cKP) now occur in hospitals and LTCFs. The most common clinical syndromes due to cKP are pneumonia, UTI, abdominal infection, intravascular device infection, surgical site infection, soft tissue infection, and subsequent bacteremia. A critical feature of cKP as a successful health care–associated pathogen has been its evolution into an MDR gram-negative bacillus. MDR strains of cKP have caused a number of nosocomial infectionoutbreaks in ICUs and neonatal nurseries. Historically, cKP caused severe community-acquired pneumonia, primarily in alcoholics; this syndrome is still observed with some frequency in Africa and Asia but has become increasingly uncommon in the United States and Europe. cKP strains appear to be genomically distinct from hypervirulent *K. pneumoniae* (hvKP), an emerging pathogen that has been recognized increasingly over the past two decades (see "Abdominal Infection," below); it is possible, however, that cKP and hvKP strains have virulence factors in common. *K. pneumoniae* subspecies *rhinoscleromatis* is the causative agent of rhinoscleroma, a granulomatous mucosal upper respiratory infection that progresses slowly (over months or years) and causes necrosis and occasionally obstruction of the nasal passages. *K. pneumoniae* subspecies *ozaenae* has been implicated as a cause of chronic atrophic rhinitis and rarely of invasive disease in compromised hosts. These two *K. pneumoniae* subspecies are usually isolated from patients in tropical climates and are genomically distinct from both cKP and hvKP.

Infectious Syndromes

PNEUMONIA

K. pneumoniae accounts for only a small proportion of cases of CAP (Chap. 257); however, CAP due to *K. pneumoniae* is more common in Africa and Asia than in Europe and the United States. This infection occurs primarily in hosts with underlying conditions (e.g., alcoholism, diabetes, or chronic lung disease). Pulmonary infection is especially common among residents of LTCFs and hospitalized patients because of increased rates of oropharyngeal colonization. Mechanical ventilation is an important risk factor. As in all pneumonias due to enteric GNB, production of purulent sputum and evidence of airspace disease are typical. Presentation with earlier, less extensive infection is more common than the classically described lobar infiltrate with a bulging fissure. Pulmonary necrosis, pleural effusion, and empyema can occur with disease progression.

UTI

K. pneumoniae accounts for only 1–2% of UTI episodes among otherwise healthy adults but for 5–17% of episodes of complicated UTI, including infections associated with indwelling urinary catheters.

ABDOMINAL INFECTION

Klebsiella causes a spectrum of abdominal infections similar to that caused by *E. coli* but is less frequently isolated from these infections. The new hypervirulent variant of *K. pneumoniae* (hvKP) that has emerged over the past decade was initially reported from the Pacific Rim but was later described in the United States, Canada, Europe, and elsewhere. At first, hvKP infection was characterized and distinguished from traditional infections due to cKP by (1) presentation as community-acquired hepatic abscess, (2) occurrence in patients lacking a history of hepatobiliary disease, and (3) a propensity for metastatic spread to distant sites (e.g., eyes, central nervous system, lungs) in 11–80% of cases **(Fig. 149-1, left)**. More recently, this variant has been identified as the cause of a variety of serious extrahepatic abscesses/infections as well. The affected individuals often have diabetes mellitus and are of Asian extraction; however, nondiabetics and all ethnic groups can be affected. Not uncommonly,

hosts are young and healthy. Survivors with metastatic spread often suffer catastrophic morbidity, such as loss of vision and neurologic sequelae.

Figure 149-1

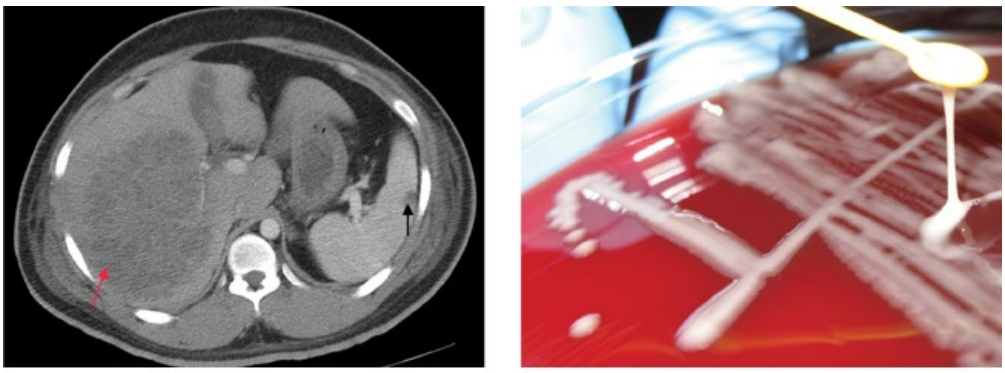

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

New hypervirulent variant of *K. pneumoniae* (hvKP). *Left*: An image from an abdominal CT scan of a previously healthy 24-year-old Vietnamese man shows a primary liver abscess (*red arrow*) with metastatic spread to the spleen (*black arrow*). *(Courtesy of Drs. Chiu-Bin Hsaio and Diana Pomakova.) Right*: A culture of the responsible hvKP strain was grown from the patient's blood and the abscess. A hypermucoviscous phenotype has been associated with hvKP strains that cause community-acquired primary liver abscess. This phenotype has been semiquantitatively defined by a positive "string test" (formation of a viscous string >5 mm long when bacterial colonies on an agar plate are stretched by an inoculation loop).

OTHER INFECTIONS

Klebsiella cellulitis or soft tissue infection most frequently affects devitalized tissue (e.g., decubitus and diabetic ulcers, burn sites) and immunocompromised hosts. *Klebsiella* causes some cases of surgical site infection, hematogenously derived endophthalmitis (especially in association with hepatic abscess), and nosocomial sinusitis in addition to occasional cases of osteomyelitis contiguous to soft tissue infection, nontropical myositis, and meningitis (both during the neonatal period and after neurosurgery). Cytotoxin-producing strains of *K. oxytoca* have been implicated as a cause of hemorrhagic (but *not* nonhemorrhagic) antibiotic-associated non–*C. difficile* colitis.

BACTEREMIA

Klebsiella infection at any site can produce bacteremia. Infections of the urinary tract, respiratory tract, and abdomen (especially hepatic abscess) each account for 15–30% of episodes of *Klebsiella* bacteremia. Intravascular device–related infections account for another 5–15% of episodes, and surgical site and miscellaneous infections account for the rest. *Klebsiella* is a cause of sepsis in neonates and of bacteremia in neutropenic patients. Like enteric GNB in general, *Klebsiella* rarely causes endocarditis or endovascular infection.

Diagnosis

Klebsiellae are readily isolated and identified in the laboratory. These organisms usually ferment lactose, although the subspecies *rhinoscleromatis* and *ozaenae* are nonfermenters and are indole negative. The new hypervirulent clinical variant most commonly is capsular serotype K1 or K2 and possesses a hypermucoviscous phenotype (Fig. 149-1, right).

Treatment: *Klebsiella* Infections

K. pneumoniae and *K. oxytoca* have largely similar antibiotic resistance profiles. These species are intrinsically resistant to ampicillin and ticarcillin, and nitrofurantoin is only poorly active against them. Data from the NHSN indicated that 24% of *K. pneumoniae* device-associated infections were due to strains resistant to cephalosporins III in 2006–2007. Even higher rates have been reported outside North America, with 68% of isolates from the INICC resistant in 2002–2007. This increasing resistance is mediated primarily by plasmid-encoded ESBLs. In addition, such plasmids usually encode resistance to aminoglycosides, tetracyclines, and TMP-SMX. Furthermore, isolates of *K. pneumoniae* that contain CTX-M ESBLs have been obtained from ambulatory patients with no recent health care contact (see *E. coli* for treatment options). Resistance to β-lactam/ β-lactamase inhibitor combinations and cephamycins independent of ESBL-encoding plasmids has also been described with increasing frequency, particularly in Latin America. The prevalence of fluoroquinolone resistance is 15–20% overall and is 50% among ESBL-containing strains. Given both the undesirability of treating the latter strains with penicillins or cephalosporins and the fluoroquinolone resistance often associated with ESBLs, empirical treatment of serious or health care–associated *Klebsiella* infections with amikacin or

carbapenems is prudent. Predictably, however, the ESBL-driven use of carba-penems has selected for strains of *Klebsiella* that possess carba-penemases, which confer resistance to the substrates of ESBLs as well as to cephamycins and carbapenems. In the United States, some strains of *Klebsiella* possess KPC-family carbapenemases on transferable plasmids and also exhibit resistance to fluoroquinolones and aminoglycosides. Treatment of infections due to strains that possess carbapenemases is highly challenging, and these strains are increasingly nearly panresistant. The optimal choice for therapy is unclear. Tigecycline, polymyxin B, and polymyxin E (colistin) are the most active agents in vitro and are used most frequently. However, resistance to these agents is already emerging, and strains of *Klebsiella* resistant to all known antimicrobial agents have been described in the United States and globally. At present, there are no treatment options for these strains. In treating isolates that are more "antimicrobial friendly," it is critical to use the most appropriate narrower-spectrum agent whenever possible.

PROTEUS INFECTIONS

P. mirabilis causes 90% of *Proteus* infections, which occur in the community, LTCFs, and hospitals. *P. vulgaris* and *P. penneri* are associated primarily with infections acquired in LTCFs or hospitals. *Proteus* species are part of the colonic flora of a wide variety of mammals, birds, fish, and reptiles. The ability of these GNB to generate histamine from contaminated fish has implicated them in the pathogenesis of scombroid (fish) poisoning (Chap. 396). *P. mirabilis* colonizes healthy humans (prevalence, 50%), whereas *P. vulgaris* and *P. penneri* are isolated primarily from individuals with underlying disease. The urinary tract is by far the most common site of *Proteus* infection, with adhesins, flagella, IgA protease, and urease representing the principal known urovirulence factors. *Proteus* less commonly causes infection at a variety of other extraintestinal sites.

Infectious Syndromes

UTI

Most *Proteus* infections arise from the urinary tract. *P. mirabilis* causes only 1–2% of cases of UTI in healthy women, and *Proteus* species collectively cause only 5% of cases of hospital-acquired UTI. However, *Proteus* is responsible for 10–15% of cases of complicated UTI, primarily those associated with catheterization; indeed, among UTI isolates from chronically catheterized patients, the prevalence of *Proteus* is 20–45%. This high prevalence is due in part to bacterial production of urease, which hydrolyzes urea to ammonia and results in alkalization of the urine. Alkalization of urine, in turn, leads to precipitation of organic and inorganic compounds, which contributes to formation of struvite and carbonate-apatite crystals, formation of biofilms on catheters, and/or development of frank calculi. *Proteus* becomes associated with the stones and biofilms; thereafter, it usually can be eradicated only by removal of the stones or the catheter. Over time, staghorn calculi may form within the renal pelvis and lead to obstruction and renal failure. Thus, urine samples with unexplained alkalinity should be cultured for *Proteus*, and identification of a *Proteus* species in urine should prompt consideration of an evaluation for urolithiasis.

OTHER INFECTIONS

Proteus occasionally causes pneumonia (primarily in LTCF residents or hospitalized patients), nosocomial sinusitis, intraabdominal abscesses, biliary tract infection, surgical site infection, soft tissue infection (especially decubitus and diabetic ulcers), and osteomyelitis (primarily contiguous); in rare cases, it causes nontropical myositis. In addition, *Proteus* uncommonly causes neonatalmeningitis, with the umbilicus commonly implicated as the source; this disease is often complicated by development of a cerebral abscess. Otogenic brain abscess also occurs.

BACTEREMIA

The majority of *Proteus* bacteremia episodes originate from the urinary tract; however, any of the less common sites of infection as well as intravascular devices are also potential sources. Endovascular infection is rare. *Proteus* species are occasional agents of sepsis in neonates and of bacteremia in neutropenic patients.

Diagnosis

Proteus is readily isolated and identified in the laboratory. Most strains are lactose negative, produce H_2S, and demonstrate characteristic swarming motility on agar plates. *P. mirabilis* is indole negative, whereas *P. vulgaris* and *P. penneri* are indole positive.

Treatment: *Proteus* Infections

P. mirabilis is usually susceptible to most antimicrobial agents except tetracycline, nitrofurantoin, polymyxin B, and tigecycline. Resistance to ampicillin and cephalosporins I has been acquired by 10 –50% of strains. Overall, 10–15% of *P. mirabilis* isolates are resistant to fluoroquinolones; 5% of isolates in the United States now produce ESBLs. Furthermore, isolates of *P. mirabilis* that contain CTX-M ESBLs have been obtained from ambulatory patients with no recent health care contact (see *E. coli* for treatment options). *P. vulgaris* and *P. penneri* exhibit more extensive drug resistance than

does *P. mirabilis*. Resistance to ampicillin and cephalosporins I is the rule, and 30–40% of isolates are resistant to fluoroquinolones. Derepression of an inducible chromosomal AmpC β-lactamase (not present in *P. mirabilis*) occurs in up to 30% of *P. vulgaris* isolates. Imipenem, cephalosporins IV (e.g., cefepime), amikacin, and TMP-SMX display excellent activity against *Proteus* species (90–100% of isolates susceptible).

ENTEROBACTER INFECTIONS

E. cloacae and *E. aerogenes* are responsible for most *Enterobacter* infections (65–75% and 15–25%, respectively); *E. sakazakii* (recently renamed Cronobacter sakazakii) and *E. gergoviae* are less commonly isolated (1% and <1% of *Enterobacter* isolates, respectively). *Enterobacter* species cause primarily hospital-acquired and other health care–related infections. The organisms are widely prevalent in foods, environmental sources (including equipment at health care facilities), and a variety of animals. Few healthy humans are colonized, but the percentage increases significantly with LTCF residence or hospitalization. Although colonization is an important prelude to infection, direct introduction via IV lines (e.g., contaminated IV fluids or pressure monitors) also occurs. Extensive antibiotic resistance has developed in *Enterobacter* species and probably has contributed to the emergence of the organisms as prominent nosocomial pathogens. Individuals who have previously received antibiotic treatment, have comorbid disease, and are being treated in ICUs are at greatest risk for infection. *Enterobacter* causes a spectrum of extraintestinal infections similar to that described for other GNB.

Infectious Syndromes

Pneumonia, UTI (particularly catheter-related), intravascular device–related infection, surgical site infection, and abdominal infection (primarily postoperative or related to devices such as biliary stents) are the most common syndromes encountered. Nosocomial sinusitis, meningitis related to neurosurgical procedures (including use of intracranial pressure monitors), osteomyelitis, and endophthalmitis after eye surgery are less frequent. *E. (C.) sakazakii* is associated with neonatal meningitis/sepsis (particularly in premature infants); contaminated formula has been implicated as a source of this infection, which is often complicated by brain abscess or ventriculitis. Bacteremia can result from infection at any anatomic site. In *Enterobacter* bacteremia of unclear origin, the contamination of IV fluids or medications, blood components or plasma derivatives, catheter-flushing fluids, pressure monitors, and dialysis equipment should be considered, particularly in an outbreak setting. *Enterobacter* can also cause bacteremia in neutropenic patients. *Enterobacter* endocarditis is rare, occurring primarily in association with illicit IV drug use or prosthetic valves.

Diagnosis

Enterobacter is readily isolated and identified in the laboratory. Most strains are lactose positive and indole negative.

Treatment: *Enterobacter* Infections

Significant antimicrobial resistance exists among *Enterobacter* strains. Ampicillin and cephalosporins I and II have little or no activity. The extensive use of cephalosporins III has resulted in the selection of strains that are derepressed for production of AmpC β-lactamase, which confers resistance to cephalosporins III, monobactams (e.g., aztreonam), and—in many cases—β-lactam/β-lactamase inhibitor combinations. Resistance may emerge during therapy; in one study, emergence of resistance was documented in 20% of patients. De novo resistance should be considered when clinical deterioration follows initial improvement, and cephalosporins III should be avoided in the treatment of serious *Enterobacter* infections. National Nosocomial Infections Surveillance System data for 2003 identified resistance to cephalosporins III in 31% of ICU isolates, and INICC data for 2002–2007 reported such resistance in 57% of isolates. Cefepime is stable in the presence of AmpC β-lactamases; thus, it is a suitable option for treatment of *Enterobacter* infections so long as no coexistent ESBL is present. However, the prevalence of ESBL production in *Enterobacter* species (particularly *E. cloacae*) has been increasing and is now 5–30%. Such strains, which are also resistant to cefepime, can be challenging to treat. Fortunately, in the United States, carbapenems, amikacin, and fluoroquinolones have generally retained excellent activity (90–99% of isolates susceptible). Although clinical experience is limited, tigecycline is highly active in vitro. Once again, it is critical to use the most appropriate narrower-spectrum agent whenever possible.

SERRATIA INFECTIONS

S. marcescens causes the majority (>90%) of *Serratia* infections; *S. liquefaciens, S. rubidaea, S. fonticola,* and *S. odorifera* are isolated occasionally. Serratiae are found primarily in the environment (including in health care institutions), particularly in moist settings. Although serratiae have been isolated from a variety of animals, healthy humans are rarely colonized. In LTCFs or hospitals, reservoirs for the organisms include the hands and fingernails of health care personnel, food, milk (on neonatal units), sinks, respiratory equipment, pressure monitors, IV solutions or parenteral medications (particularly those generated by compounding pharmacies), multiply accessed medication vials, blood

products (e.g., platelets), hand soaps and lotions, irrigation solutions, and even disinfectants. Infection results from either direct inoculation (e.g., via IV fluid) or colonization (primarily of the respiratory tract) and subsequent infection. Sporadic infection is most common, but epidemics (often involving MDR strains in adult and neonatal ICUs) and common-source outbreaks occasionally occur. The spectrum of extraintestinal infections caused by *Serratia* is similar to that for other GNB. *Serratia* species are usually considered as causative agents of health care–associated infection and account for 1–3% of hospital-acquired infections. However, population-based laboratory surveillance studies in Canada and Australia demonstrated that community-acquired infections occur more commonly than was previously appreciated.

Infectious Syndromes

The respiratory tract, the genitourinary tract, intravascular devices, and surgical wounds are the most common sites of *Serratia* infection and sources of *Serratia* bacteremia. Soft tissue infections (including myositis), osteomyelitis, abdominal and biliary tract infection (postprocedural), contact lens–associated keratitis, endophthalmitis, septic arthritis (primarily from intraarticular injections), and infusion-related bacteremias occur less commonly. Serratiae are uncommon causes of neonatal or postsurgical meningitis and of bacteremia in neutropenic patients. Endocarditis is rare.

Diagnosis

Serratiae are readily cultured and identified by the laboratory and are usually lactose and indole negative. Some *S. marcescens* strains and *S. rubidaea* are red-pigmented.

Treatment: *Serratia* Infections

Most *Serratia* strains (>80%) are resistant to ampicillin, cephalosporins I, nitrofurantoin, and polymyxin B. In general, >90% of *Serratia* isolates are susceptible to other antibiotics appropriate for use against GNB. Stable derepression of inducible chromosomal AmpC β-lactamases may be preexistent or may develop during therapy. Both in the United States and globally, the prevalence of ESBL-producing isolates is <5%.

CITROBACTER INFECTIONS

C. freundii and *C. koseri* cause most human *Citrobacter* infections, which are epidemiologically and clinically similar to *Enterobacter* infections. *Citrobacter* species are commonly present in water, food, soil, and certain animals. *Citrobacter* is part of the normal fecal flora in a minority of healthy humans, but colonization rates increase in LTCFs and hospitals—the settings in which nearly all *Citrobacter* infections occur. *Citrobacter* species account for 1–2% of nosocomial infections. The affected hosts are usually immunocompromised or have comorbid disease. *Citrobacter* causes extraintestinal infections similar to those described for other GNB.

Infectious Syndromes

The urinary tract accounts for 40–50% of *Citrobacter* infections. Less commonly involved sites include the biliary tree (particularly with stones or obstruction), the respiratory tract, surgical sites, soft tissue (e.g., decubitus ulcers), the peritoneum, and intravascular devices. Osteomyelitis (usually from a contiguous focus), neurosurgery-related infection, and myositis occur rarely. *Citrobacter* (particularly *C. koseri*) also uncommonly causes neonatalmeningitis, with brain abscess complicating 50–80% of cases. Bacteremia is most often due to UTI, biliary or abdominal infection, or intravascular device infection. *Citrobacter* occasionally causes bacteremia in neutropenic patients. Endocarditis and endovascular infections are rare.

Diagnosis

Citrobacter species are readily isolated and identified; 35–50% of isolates are lactose positive, and 100% are oxidase negative. *C. freundii* is indole negative, whereas *C. koseri* is indole positive.

Treatment: *Citrobacter* Infections

C. freundii is more extensively resistant to antibiotics than is *C. koseri*. Ampicillin and the cephalosporins I and II display poor activity. *Citrobacter* species possess inducible AmpC β-lactamases; stable derepression may be preexistent or may develop during therapy. Resistance to antipseudomonal penicillins, aztreonam, fluoroquinolones, gentamicin, and cephalosporins III is variable but increasing. The prevalence of ESBL-producing isolates is <5%. Carbapenems, amikacin, cefepime, tigecycline (with which there is limited clinical experience), ceftobiprole (FDA approval pending), fosfomycin (available in the United States only as an oral formulation), and polymyxins (the agents of last resort because of potential toxicities) are most active, with >90% of strains susceptible.

MORGANELLA AND *PROVIDENCIA* INFECTIONS

M. morganii, P. stuartii, and (less frequently) *P. rettgeri* are the members of their respective genera that cause human infections. In terms of epidemiologic associations, pathogenic properties, and clinical manifestations, these organisms are largely similar to *Proteus* species; however, *Morganella*

and *Providencia* occur more commonly among LTCF residents; to a lesser degree, they affect hospitalized patients.

Infectious Syndromes

These species are primarily urinary tract pathogens, causing UTIs that are most often associated with long-term (>30-day) catheterization. Such infections commonly lead to biofilm formation and catheter encrustation (sometimes causing catheter obstruction) or the development of struvite bladder or renal stones (sometimes causing renal obstruction and serving as foci for relapse). Other, less common infectious syndromes include surgical site infection, soft tissue infection (primarily involving decubitus and diabetic ulcers), burn site infection, pneumonia (particularly ventilator-associated), intravascular device infection, and intraabdominal infection. Rarely, the other extraintestinal infections described for GNB also occur. Bacteremia is uncommon; although any infected site can serve as the source, the urinary tract accounts for most cases, and the next most common sources are surgical, soft tissue, and hepatobiliary sites.

Diagnosis

M. morganii and *Providencia* are readily isolated and identified. Nearly all isolates are lactose negative and indole positive.

Treatment: *Morganella* and Providencia Infections

Morganella and *Providencia* may be extensively resistant to antibiotics. Most isolates are resistant to ampicillin, cephalosporins I, nitrofurantoin, fosfomycin, tigecycline, and polymyxin B; 40% are resistant to quinolones. *Morganella* and *Providencia* possess inducible AmpC β-lactamases; stable derepression may be preexistent or may develop during therapy. Resistance to antipseudomonal penicillins, aztreonam, gentamicin, TMP-SMX, and cephalosporins II and III is emerging but still variably prevalent. The β-lactamase inhibitor tazobactam increases susceptibility to β-lactam agents, but sulbactam and clavulanic acid do not. Imipenem, amikacin, and cefepime are the most active agents (>90% of isolates susceptible). Removal of a colonized catheter or stone is critical for eradication of UTI.

EDWARDSIELLA INFECTIONS

E. tarda is the only member of the genus *Edwardsiella* that is associated with human disease. This organism is found predominantly in freshwater and marine environments and the associated animal species. Human acquisition occurs primarily during interaction with these reservoirs. *E. tarda* infection is rare in the United States; recently reported cases are mostly from Southeast Asia. This pathogen shares clinical features with both *Salmonella* species (as a diarrheal pathogen; Chap. 153) and *Vibrio vulnificus* (as an extraintestinal pathogen; Chap. 156).

Infectious Syndromes

Gastroenteritis is the predominant infectious syndrome (50–80% of infections). Self-limiting watery diarrhea is most common, but severe colitis also occurs. The most common extraintestinal infection is wound infection due to direct inoculation, which is often associated with freshwater-, marine-, or snake-related injuries. Other infectious syndromes result from invasion of the gastrointestinal tract and subsequent bacteremia. Most afflicted hosts have comorbidities (e.g., hepatobiliary disease, iron overload, cancer, or diabetes mellitus). A primary bacteremic syndrome, sometimes complicated by meningitis, has a 40% case-fatality rate. Visceral (primarily hepatic) and intraperitoneal abscesses also occur.

Diagnosis

Although *E. tarda* can readily be isolated and identified, most laboratories do not routinely seek to identify it in stool samples. Production of hydrogen sulfide is a characteristic biochemical property.

Treatment: *Edwardsiella* Infections

E. tarda is susceptible to most antimicrobial agents appropriate for use against GNB. Gastroenteritis is generally self-limiting, but treatment with a fluoroquinolone may hasten resolution. In the setting of severe sepsis, fluoroquinolones, cephalosporins III and IV, carbapenems, and amikacin—either alone or in combination—are the safest choices pending susceptibility information.

INFECTIONS CAUSED BY MISCELLANEOUS GENERA

Species of *Hafnia*, *Kluyvera*, *Cedecea*, *Pantoea*, *Ewingella*, *Leclercia*, and *Photorhabdus* are occasionally isolated from diverse clinical specimens, including blood, sputum, cerebrospinal fluid, joint fluid, bile, and wounds. These organisms are rare and usually cause infection in a compromised host or in the setting of an invasive procedure or a foreign body.

FURTHER READINGS

Engel HJ et al: *Serratia* sp. bacteremia in Canberra, Australia: A population-based study over 10 years.

Eur J Clin Microbiol Infect Dis 28:821, 2009[PMID: 19194731]

Falagas ME et al: Fosfomycin: Use beyond urinary tract and gastrointestinal infections. Clin Infect Dis 46:1069, 2008[PMID: 18444827]

Flores J, okhuysen PC: Enteroaggregative *Escherichia coli* infection. Curr Opin Gastroenterol 25:8, 2009[PMID: 19114769]

Freeman JT et al: Emergence of extended-spectrum beta-lactamase-producing *Escherichia coli* in community hospitals throughout North Carolina: A harbinger of a wider problem in the United States? Clin Infect Dis 49:e30, 2009

Hidron AI et al: NHSN annual update: Antimicrobial-resistant pathogens associated with healthcare-associated infections: Annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006–2007. Infect Control Hosp Epidemiol 29:996, 2008[PMID: 18947320]

Hogenauer C et al: *Klebsiella oxytoca* as a causative organism of antibiotic-associated hemorrhagic colitis. N Engl J Med 355: 2418, 2006[PMID: 17151365]

Johnson JR et al: Molecular analysis of *Escherichia coli* from retail meats (2002–2004) from the United States National Antimicrobial Resistance Monitoring System. Clin Infect Dis 49:195, 2009[PMID: 19508167]

Rodriguez-Bano J et al: Community-onset bacteremia due to extended-spectrum β-lactamase-producing *Escherichia coli*: Risk factors and prognosis. Clin Infect Dis 50:40, 2010

Samonis G et al: *Citrobacter* infections in a general hospital: Characteristics and outcomes. Eur J Clin Microbiol Infect Dis 28:61, 200

Souli M et al: An outbreak of infection due to β-lactamase *Klebsiella pneumoniae* carbapenemase 2-producing *K. pneumoniae* in a Greek university hospital: Molecular characterization, epidemiology, and outcomes. Clin Infect Dis 50:364, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACINETOBACTER INFECTIONS: INTRODUCTION

Infections with bacteria of the genus *Acinetobacter* have become a significant problem worldwide. *Acinetobacter baumannii* is particularly formidable because of its propensity to acquire antibiotic resistance determinants. Outbreaks of infection caused by strains of *A. baumannii* resistant to multiple antibiotic classes, including carbapenems, are a serious concern in many specialized hospital units, including intensive care units (ICUs). The foremost implication of infection with carbapenem-resistant *A. baumannii* is the need to use "last-line" antibiotics such as colistin, polymyxin B, or tigecycline; these options have the potential to render these bacteria resistant to all available antibiotics.

Definition

Acinetobacter species are oxidase-negative, nonmotile, nonfermenting, short gram-negative bacilli that grow well at 37° C in aerobic conditions on a range of laboratory media (e.g., blood agar). Some species may not grow on MacConkey agar. Differentiation of *Acinetobacter* species is difficult with the means typically available to most clinical microbiology laboratories, including commercial semiautomated identification systems. DNA-DNA hybridization is a method used for speciation in reference laboratories. Identification of the most clinically relevant species, *A. baumannii*, by detection of the $bla_{OXA\text{-}51}$-like carbapenemase gene intrinsic to this species has been described.

Etiology

Widely distributed in nature, *Acinetobacter* species can be found in water, in soil, and on vegetables. *Acinetobacter* is a component of the human skin flora and is sometimes identified as a contaminant in blood samples collected for culture. Fecal carriage can be detected in both healthy and hospitalized individuals. Despite the ubiquity of some *Acinetobacter* species, the natural habitat of *A. baumannii* remains to be fully defined.

Epidemiology

A. baumannii infections have been diagnosed in patients on all inhabited continents. The vast majority of infections occur in hospitalized patients and other patients with significant health care contact. Outbreaks of carbapenem-resistant *A. baumannii* are particularly problematic. A significant issue is the introduction of carbapenem-resistant *A. baumannii* into hospitals as a result of medical transfers, especially from hospitals where the organism is highly endemic.

THE AMERICAS

In 1991 and 1992, outbreaks of carbapenem-resistant *A. baumannii* infection occurred in a hospital in New York City. Subsequently, numerous other hospitals in the United States and South America have had outbreaks of carbapenem-resistant *A. baumannii*. The incidence of infections with *A. baumannii* among military personnel from the United States and Canada has increased since 2002; 102 patients had bloodstream infections at facilities treating U.S. military personnel injured in Iraq or Afghanistan from January 1, 2002, through August 31, 2004. An epidemiologic investigation revealed that *A. baumannii* could be grown from environmental sites in field hospitals and that the environmental strains were closely related genotypically to clinical isolates. *A. baumannii* strains from injured military personnel from the United States and the United Kingdom were also genotypically related; this finding provided further evidence that *A. baumannii* was being acquired in field hospitals.

EUROPE

A. baumannii infections have posed a substantial clinical challenge in many parts of Europe since the early 1980s. Three clones (European clones I, II, and III) have been the predominant causes of *A. baumannii* infection in hospitals in Europe. Carbapenem resistance in *A. baumannii* is a significant issue in many European countries, most notably the United Kingdom, Greece, Italy, Spain, and Turkey.

ASIA, AUSTRALIA, THE MIDDLE EAST, AND AFRICA

Although surveillance data are sparse from many countries in these regions, problems with carbapenem-resistant *A. baumannii* abound. Community-acquired infections are well described in northern Australia and some parts of Asia. These infections may be more likely in men >45 years of age who have histories of cigarette smoking, alcoholism, diabetes mellitus, or chronic obstructive airway disease. Community-acquired strains are more susceptible to antimicrobial agents than are

hospital-acquired strains.

Pathogenesis

A. baumannii colonizes patients exposed to heavily contaminated hospital environments or to the hands of health care workers in these locations. Colonization of the upper airways in mechanically ventilated patients may lead to nosocomial pneumonia. Colonization of the skin may lead to central line–associated bloodstream infection, catheter-associated urinary tract infection (UTI), wound infection, or postneurosurgical meningitis. Throat carriage and microaspiration may be involved in the pathogenesis of community-acquired pneumonia due to *A. baumannii*.

Much less is known about the virulence mechanisms of and host responses to *A. baumannii* than about these aspects of other pathogenic gram-negative bacteria. Because of the emergence of multidrug-resistant strains, including those resistant to all available antibiotics, the impetus to study *A. baumannii* pathogenesis has grown. Novel targets for antibacterial drug development are desperately required, and drugs that have antivirulence mechanisms may provide new therapeutic options. Specific virulence mechanisms in *A. baumannii* include iron acquisition and transport systems; outer-membrane protein A (OmpA), which mediates mammalian cell adhesion, invasion, and cytotoxicity through mitochondrial damage and initiation of caspase-dependent apoptosis; lipopolysaccharide (LPS); and the ability to form biofilm on abiotic and biotic surfaces. Biofilm formation on abiotic surfaces is dependent on a pilus assembly system, which in turn is controlled by a traditional two-component regulatory system mediated by *bfmR*. Also important in biofilm formation are biofilm-associated protein; OmpA; the quorum-sensing gene *abaI*, which controls the secretion of 3-hydroxy-C_{12}-homoserine lactone; and the *pga* locus, which is essential for the production of the polysaccharide poly-β-1,6-*N*-acetylglucosamine.

New model systems for the study of *A. baumannii* infection, including both nonmammalian (invertebrate) and mammalian models, have been described. Furthermore, the use of *A. baumannii* transposon-generated mutant libraries to screen for mutants with attenuated growth in human biological fluids (serum and ascites fluid) has allowed the identification of new virulence mechanisms. These include phospholipase D; capsule production mediated by *ptk* and *epsA*; penicillin-binding protein 7/8 encoded by the *pbpG* gene; and a glycosyltransferase important for LPS biosynthesis encoded by the *lpsB* gene.

The LPS of *A. baumannii* appears to play a significant role in eliciting host responses. In studies with knockout mice, Toll-like receptor 4 and CD14 were shown to be important in host recognition, signaling, and cytokine production in response to *A. baumannii*. Humoral responses targeting iron-regulated outer-membrane proteins and the O-polysaccharide component of LPS have also been described.

Approach to the Patient: *Acinetobacter* Infection

Acinetobacter must be considered in the differential diagnosis of hospital-acquired pneumonia, central line–associated bloodstream infection, posttraumatic wound infection in military personnel returning from Iraq and Afghanistan, and postneurosurgical meningitis.

Clinical Manifestations

PNEUMONIA

It may be difficult to distinguish between upper-airway colonization with *A. baumannii* and hospital-acquired pneumonia. An estimated 5–10% of cases of ventilator-associated pneumonia are due to *A. baumannii*, although much regional variation exists. Typically, patients with *A. baumannii* ventilator-associated pneumonia have had a prolonged stay in ICUs; in outbreak situations, however, patients may acquire the infection within days of arrival in an ICU.

Community-acquired pneumonia due to *A. baumannii* has been described in tropical regions of Australia and Asia. The disease typically occurs during the "wet" season among people with a history of alcohol abuse. Infection may result in fulminant pneumonia requiring admission to an ICU, with a mortality rate of ~50%.

BLOODSTREAM INFECTION

Although *A. baumannii* accounts for only ~1–2% of nosocomial bloodstream infections, crude mortality rates from these infections may be as high as 40%. Sources of bloodstream infection are typically a central line or underlying pneumonia, UTI, or wound infection.

TRAUMATIC BATTLEFIELD AND OTHER WOUNDS

A. baumannii is a well-known pathogen in burn units. This organism is commonly isolated from wounds of combat casualties from Iraq or Afghanistan; it was the most commonly isolated organism in one assessment of combat victims with open tibial fractures but did not appear to contribute directly to persistent nonunion or the need for amputation.

MENINGITIS

A. baumannii may cause meningitis following neurosurgical procedures. Patients typically have an external ventricular drain in situ.

URINARY TRACT INFECTION

A. baumannii is an occasional cause of catheter-associated UTI. It is highly unusual for this organism to cause uncomplicated UTI in healthy women.

OTHER CLINICAL MANIFESTATIONS

A small number of case reports describe *Acinetobacter* prosthetic-valve endocarditis and endophthalmitis/keratitis. The latter is sometimes related to contact lens use or eye surgery.

Diagnosis

Acinetobacter infection should be suspected when plump coccobacilli are seen in Gram's-stained respiratory tract secretions, blood cultures, or cerebrospinal fluid. Sometimes the organisms are difficult to de-stain. Given their small size, they may be misidentified as either gram-negative or gram-positive cocci.

Treatment: *Acinetobacter* Infection Table 150-1

Treatment is hampered by the remarkable ability of *A. baumannii* to upregulate or acquire antibiotic resistance determinants. The most prominent example is that of β-lactamases, including those capable of inactivating carbapenems, cephalosporins, and penicillins. These enzymes, which include the OXA-type β-lactamases (e.g., OXA-23) and the metallo-β-lactamases, are typically resistant to β-lactamase inhibitors such as clavulanate or tazobactam. Plasmids harboring genes encoding these β-lactamases may also harbor genes encoding resistance to aminoglycosides and sulfur antibiotics. The end result is that carbapenem-resistant *A. baumannii* may become truly multidrug resistant.

Table 150-1 Treatment Options for *Acinetobacter* Infections

Antibiotic	Comments
Sulbactam	Intrinsic activity against *Acinetobacter*, not linked to β-lactamase inhibition
Trimethoprim-sulfamethoxazole	May be an option for urinary tract infection or wound infection
Meropenem	Widely used in ventilator-associated pneumonia, but carbapenem resistance is widespread
Amikacin	May be an option for carbapenem-resistant strains
Tigecycline	May be an option for carbapenem-resistant strains but inappropriate for urinary tract infection, bloodstream infection, or meningitis
Colistin or polymyxin B	May be an option for carbapenem-resistant strains, but pharmacokinetics not yet well understood

Selection of empirical antibiotic therapy when *A. baumannii* is suspected is challenging and must rely on a knowledge of local epidemiology. The interval from onset of infection to initiation of effective empirical therapy clearly influences outcome. Given the diversity of resistance mechanisms in *A. baumannii*, definitive therapy should be based on the results of antimicrobial susceptibility testing. Carbapenems (imipenem, meropenem, and doripenem but not ertapenem) have long been thought of as the agents of choice for serious *A. baumannii* infections. However, the clinical utility of carbapenems is increasingly jeopardized by the production of carbapenemases, as described above. Sulbactam may be an alternative to carbapenems. Unlike other β-lactamase inhibitors (e.g., clavulanic acid and tazobactam), sulbactam has intrinsic activity against *Acinetobacter*; this activity is mediated by the drug's binding to penicillin-binding protein 2 rather than its ability to inhibit β-lactamases. Sulbactam is commercially available in a combined formulation with either ampicillin or cefoperazone and may also be available as a single agent in some countries. Despite the absence of randomized clinical trials, sulbactam seems to be equivalent to carbapenems in clinical effectiveness against susceptible strains.

Therapy for carbapenem-resistant *A. baumannii* is particularly problematic. The only currently available choices are polymyxins (colistin and polymyxin B) or tigecycline. Neither option is perfect. Polymyxins may be nephrotoxic and neurotoxic. The optimal dose and schedule for administration of polymyxins to patients in vulnerable groups (e.g., those requiring renal replacement therapy) are unknown. Conventional doses of tigecycline may not result in serum concentrations adequate to treat bloodstream infections. Resistance of *A. baumannii* to tigecycline may develop during treatment with this drug. Clearly, new treatment options are needed for serious *A. baumannii* infections.

Complications and Prognosis

Given the propensity of *A. baumannii* to cause infections in seriously ill patients in ICUs, it is not surprising that *A. baumannii* infections are associated with high mortality rates. Thus a pertinent

question is whether *A. baumannii* infections are associated with high attributable mortality rates after the severity of illness is controlled for. A number of studies have addressed this issue but have had disparate results. Whether the discrepant results can be explained purely by methodologic differences is unknown at present.

Prevention

Multidrug-resistant *A. baumannii* clearly causes outbreaks of infection. In many outbreaks, just one or two strain types are found by molecular epidemiologic analysis. Even in endemic situations, a small number of strain types predominate. In the outbreaks in New York City, for example, two strain types accounted for >80% of carbapenem-resistant isolates. This "oligoclonality" plainly demonstrates the potential importance of infection control interventions in response to outbreaks of multidrug-resistant *A. baumannii* infection.

The hospital environment is an important reservoir of organisms capable of colonizing patients and causing infection. Environmental sources of *A. baumannii* include computer keyboards, glucometers, multidose medication vials, IV nutrition, inadequately sterilized reusable arterial pressure transducers, ventilator tubing, suction catheters, humidifiers, containers of distilled water, urine collection jugs, and moist bedding articles. Pulsatile-lavage wound treatment—a high-pressure irrigation system used to debride wounds—has been associated with an outbreak of *A. baumannii* infection.

Contaminated inanimate objects should be removed from the patient-care environment or subjected to enhanced environmental cleaning. Although contact-isolation procedures (use of gloves and gowns when dealing with colonized patients or their environment), accommodation of patients in single rooms, and improved hand hygiene are critical, attention to the patient-care environment may be the only measure that leads to control of outbreaks of *A. baumannii* infection.

FURTHER READINGS

Choi CH et al: Outer membrane protein 38 of *Acinetobacter baumannii* localizes to the mitochondria and induces apoptosis of epithelial cells. Cell Microbiol 7:1127, 2005[PMID: 16008580]

Doi Y et al: Extensively drug-resistant *Acinetobacter baumannii* . Emerg Infect Dis 15:980, 2009[PMID: 19523312]

Gaddy JA et al: Regulation of *Acinetobacter baumannii* biofilm formation. Future Microbiol 4:273, 2009[PMID: 19327114]

Guerrero DM et al: *Acinetobacter baumannii* –associated skin and soft tissue infections: Recognizing a broadening spectrum of disease. Surg Infect 11:49, 2010[PMID: 19788383]

Jacobs AC et al: Inactivation of phospholipase D diminishes *Acinetobacter baumannii* pathogenesis. Infect Immun 78:1952, 2010[PMID: 20194595]

Kim BN et al: Management of meningitis due to antibiotic-resistant *Acinetobacter* species. Lancet Infect Dis 9:245, 2009[PMID: 19324297]

Knapp S et al: Differential roles of CD14 and Toll-like receptors 4 and 2 in murine *Acinetobacter* pneumonia. Am J Respir Crit Care Med 173:122, 2006[PMID: 16210672]

Munoz-Price LS, Weinstein RA: *Acinetobacter* infection. N Engl J Med 358:1271, 2008[PMID: 18354105]

Peleg AY et al: *Acinetobacter baumannii* : Emergence of a successful pathogen. Clin Microbiol Rev 21:538, 2008[PMID: 18625687]

Perez F et al: Antibiotic resistance determinants in *Acinetobacter* spp and clinical outcomes in patients from a major military treatment facility. Am J Infect Control 38:63, 2010[PMID: 19783325]

Rodriguez-Bano J et al: Long-term control of hospital-wide, endemic multidrug-resistant *Acinetobacter baumannii* through a comprehensive "bundle" approach. Am J Infect Control 37:715, 2009[PMID: 19457584]

Smith MG et al: New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev 21:601, 2007[PMID: 17344419]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HELICOBACTER PYLORI INFECTIONS: INTRODUCTION

Definition

Helicobacter pylori colonizes the stomachs of ~50% of the world's human population throughout their lifetimes. Colonization with this organism is the main risk factor for peptic ulceration (Chap. 293) as well as for gastric adenocarcinoma and gastric MALT (mucosa-associated lymphoid tissue) lymphoma (Chap. 91). Treatment for *H. pylori* has revolutionized the management of peptic ulcer disease, providing a permanent cure in most cases. Such treatment also represents first-line therapy for patients with low-grade gastric MALT lymphoma. Treatment of *H. pylori* is of no benefit in the treatment of gastric adenocarcinoma, but prevention of *H. pylori* colonization could potentially prevent gastric malignancy and peptic ulceration. In contrast, increasing evidence indicates that lifelong *H. pylori* colonization may offer some protection against complications of gastroesophageal reflux disease (GERD), including esophageal adenocarcinoma. Recent research has focused on whether *H. pylori* colonization is a risk factor for some extragastric diseases and whether it is protective against some recently emergent medical problems, such as asthma and obesity.

Etiologic Agent

H. pylori is a gram-negative bacillus that has naturally colonized humans for at least 50,000 years—and probably throughout human evolution. It lives in gastric mucus, with a small proportion of the bacteria adherent to the mucosa and possibly a very small number of the organisms entering cells or penetrating the mucosa; its distribution is never systemic. Its spiral shape and flagella render *H. pylori* motile in the mucus environment. The organism has several acid-resistance mechanisms, most notably a highly expressed urease that catalyzes urea hydrolysis to produce buffering ammonia. *H. pylori* is microaerophilic (requiring low levels of oxygen), is slow-growing, and requires complex growth media in vitro. Publication of several complete genomic sequences of *H. pylori* since 1997 has led to significant advances in the understanding of the organism's biology.

A very small proportion of gastric *Helicobacter* infections are due to species other than *H. pylori*, possibly acquired as zoonoses. Whether these non-*pylori* gastric helicobacters cause disease remains controversial. In immunocompromised hosts, several nongastric (intestinal) *Helicobacter* species can cause disease with clinical features resembling those of *Campylobacter* infections; these species are covered in Chap. 155.

Epidemiology

The prevalence of *H. pylori* among adults is ~30% in the United States and other developed countries as opposed to >80% in most developing countries. In the United States, prevalence varies with age: ~50% of 60-year-old persons, ~20% of 30-year-old persons, and <10% of children are colonized. *H. pylori* is usually acquired in childhood. The age association is due mostly to a birth-cohort effect whereby current 60-year-olds were more commonly colonized as children than are current children. Spontaneous acquisition or loss of *H. pylori* in adulthood is uncommon. Other strong risk factors for *H. pylori* colonization are markers of crowding and maternal colonization. The low incidence among children in developed countries at present is due, at least in part, to decreased maternal colonization and increased use of antibiotics.

Humans are the only important reservoir of *H. pylori*. Children may acquire the organism from their parents (more often from the mother) or from other children. Whether transmission takes place more often by the fecal-oral or the oral-oral route is unknown, but *H. pylori* is easily cultured from vomitus and gastroesophageal refluxate and is less easily cultured from stool.

Pathology and Pathogenesis

H. pylori colonization induces a tissue response in the stomach, *chronic superficial gastritis*, which includes infiltration of the mucosa by both mononuclear and polymorphonuclear cells. (The term *gastritis* should be used specifically to describe histologic features; it has also been used to describe endoscopic appearances and even symptoms, which do not correlate with microscopic findings or even with the presence of *H. pylori*.) Although *H. pylori* is capable of numerous adaptations that prevent excessive stimulation of the immune system, colonization is accompanied by a considerable persistent immune response, including the production of both local and systemic antibodies as well as cell-mediated responses. However, these responses are ineffective in clearing the bacterium. This

inefficient clearing appears to be due in part to *H. pylori*'s downregulation of the immune system, which fosters its own persistence.

Most *H. pylori*–colonized persons do not develop clinical sequelae. That some persons develop overt disease whereas others do not is related to a combination of factors: bacterial strain differences, host susceptibility to disease, and environmental factors. Several *H. pylori* virulence factors are more common among strains that are associated with disease than among those that are not. The *cag* island is a group of genes that encodes a bacterial secretion system through which a specific protein, CagA, is translocated into epithelial cells. CagA affects host cell signal transduction, inducing proliferative, cytoskeletal, and inflammatory changes; a proportion of transgenic mice expressing CagA in the stomach develop gastric adenocarcinoma. The secretion system also translocates soluble components of the peptidoglycan cell wall into the gastric epithelial cell; these components are recognized by the intracellular emergency bacterial receptor Nod1, which stimulates a proinflammatory cytokine response resulting in enhanced gastric inflammation. Patients with peptic ulcer disease or gastric adenocarcinoma are more likely than persons without these conditions to be colonized by *cag*-positive strains. The secreted *H. pylori* protein VacA occurs in several forms. Strains with the more active forms are more commonly isolated from patients with peptic ulcer disease or gastric carcinoma than from persons without these conditions. Other bacterial factors that are associated with increased disease risk include adhesins, such as BabA and SabA, and incompletely characterized genes, such as *dupA*.

The best-characterized host determinants of disease are genetic polymorphisms leading to enhanced activation of the innate immune response, such as polymorphisms in cytokinegenes or genes encoding bacterial recognition proteins such as Toll-like receptors (TLRs). For example, colonized people with polymorphisms in the interleukin (IL) 1 gene that cause the production of large quantities of this cytokine in response to *H. pylori* infection are at increased risk of gastric adenocarcinoma. In addition, environmental cofactors are important in pathogenesis. Smoking increases the risks of ulcers and cancer in *H. pylori*–positive individuals. Diets high in salt and preserved foods increase cancer risk, whereas diets high in antioxidants and vitamin C are protective.

The pattern of gastric inflammation is associated with disease risk: antral-predominant gastritis is most closely linked with duodenal ulceration, whereas pangastritis is linked with gastric ulceration and adenocarcinoma. This difference probably explains why patients with duodenal ulceration are not at high risk of developing gastric adenocarcinoma later in life, despite being colonized by *H. pylori*.

How gastric colonization causes duodenal ulceration is now becoming clearer. *H. pylori*–induced inflammation diminishes the number of somatostatin-producing D cells. Since somatostatin inhibits gastrin release, gastrin levels are higher than in *H. pylori*–negative persons, and these higher levels lead to increased meal-stimulated acid secretion in the gastric corpus, which is only mildly inflamed in antral-predominant gastritis. How this increases duodenal ulcer risk remains controversial, but the increased acid secretion may contribute to the formation of the potentially protective gastric metaplasia found in the duodenum of duodenal ulcer patients. Gastric metaplasia in the duodenum may become colonized by *H. pylori* and subsequently inflamed and ulcerated.

The pathogenesis of gastric ulceration and that of gastric adenocarcinoma are less well understood, although both conditions arise in association with pan- or corpus-predominant gastritis. The hormonal changes described above still occur, but the inflammation in the gastric corpus means that it produces less acid (hypochlorhydria) despite hypergastrinemia. Gastric ulcers usually occur at the junction of antral and corpus-type mucosa, and this region is particularly inflamed. Gastric cancer probably stems from progressive DNA damage and the survival of abnormal epithelial cell clones. The DNA damage is thought to be due principally to reactive oxygen and nitrogen species arising from inflammatory cells and perhaps in relation to other bacteria that survive in a hypochlorhydric stomach. Longitudinal analyses of gastric biopsy specimens taken years apart from the same patient show that the common *intestinal* type of gastric adenocarcinoma follows stepwise changes from simple gastritis to gastric atrophy, intestinal metaplasia, and dysplasia. A second, *diffuse* type of gastric adenocarcinoma may arise directly from chronic gastritis alone.

Clinical Manifestations

Essentially all *H. pylori*–colonized persons have gastric tissue responses, but fewer than 15% develop associated illnesses such as peptic ulceration, gastric adenocarcinoma, or gastric lymphoma **(Fig. 151-1)**.

Figure 151-1

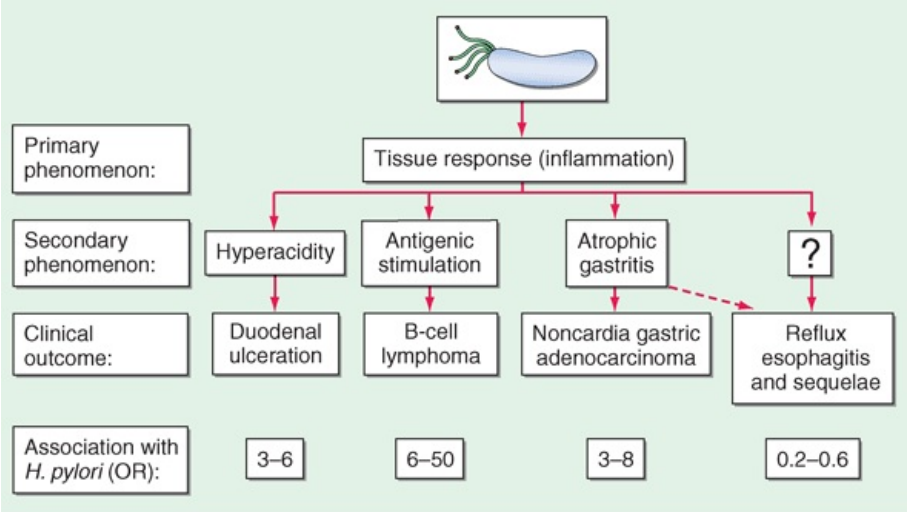

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic of the relationships between colonization with *Helicobacter pylori* and diseases of the upper gastrointestinal tract among persons in developed countries. Essentially all persons colonized with *H. pylori* develop a host response, which is generally termed chronic gastritis. The nature of the interaction of the host with the particular bacterial population determines the clinical outcome. *H. pylori* colonization increases the lifetime risk of peptic ulcer disease, noncardia gastric cancer, and B cell non-Hodgkin's gastric lymphoma [odds ratios (ORs) for all, >3]. In contrast, a growing body of evidence indicates that *H. pylori* colonization (especially with *cagA*$^+$ strains) protects against adenocarcinoma of the esophagus (and the sometimes related gastric cardia) and premalignant lesions such as Barrett's esophagus (OR, <1). While the incidences of peptic ulcer disease (cases not due to nonsteroidal anti-inflammatory drugs) and noncardia gastric cancer are declining in developed countries, the incidence of adenocarcinoma of the esophagus is rapidly increasing. *(Adapted from MJ Blaser: Hypothesis: The changing relationships of Helicobacter pylori and humans: Implications for health and disease. J Infect Dis 179:1523, 1999, with permission.)*

Worldwide, >80% of duodenal ulcers and >60% of gastric ulcers are related to *H. pylori* colonization (Chap. 293), although the proportion of ulcers due to aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs) is increasing, especially in developed countries. The main lines of evidence for an ulcer-promoting role for *H. pylori* are that (1) the presence of the organism is a risk factor for the development of ulcers, (2) non-NSAID-induced ulcers rarely develop in the absence of *H. pylori*, (3) eradication of *H. pylori* markedly reduces rates of ulcer relapse, and (4) experimental *H. pylori* infection of gerbils causes gastric ulceration.

Prospective nested case-control studies have shown that *H. pylori* colonization is a risk factor for adenocarcinomas of the distal (noncardia) stomach (Chap. 91). Long-term experimental infection of gerbils also may result in gastric adenocarcinoma. Moreover, the presence of *H. pylori* is strongly associated with primary gastric lymphoma, although this condition is much less common. Many low-grade gastric B cell lymphomas arising from MALT are driven by T cell proliferation, which in turn is driven by *H. pylori* antigen stimulation; *H. pylori* antigen–driven tumors may regress either fully or partially after *H. pylori* eradication but require careful long-term monitoring.

Many patients have upper gastrointestinal symptoms but have normal results in upper gastrointestinal endoscopy (so-called functional or nonulcer dyspepsia; Chap. 293). Because *H. pylori* is common, some of these patients will be colonized with the organism. *H. pylori* eradication leads to symptom resolution a little (7%) more commonly than does placebo treatment. Whether such patients have peptic ulcers in remission at the time of endoscopy or whether a small subgroup of patients with true functional dyspepsia respond to *H. pylori* treatment is unclear.

Much interest has focused on a possible protective role for *H. pylori* against GERD (Chap. 292), Barrett's esophagus (Chap. 292), and adenocarcinoma of the esophagus and gastric cardia (Chap. 91). The main lines of evidence for this role are (1) that there is a temporal relationship between a falling prevalence of gastric *H. pylori* colonization and a rising incidence of these conditions and (2) that, in most studies, the prevalence of *H. pylori* colonization (especially with proinflammatory *cagA*$^+$ strains) is significantly lower among patients with these esophageal diseases than among control subjects. The mechanism underlying this protective effect appears to include *H. pylori*–induced hypochlorhydria. Since, at the individual level, GERD symptoms may decrease, worsen, or remain unchanged after treatment targeting *H. pylori*, concerns about GERD should not affect decisions about *H. pylori* treatment when an indication exists.

H. pylori has an increasingly recognized role in other gastric pathologies. It may be one initial precipitant of autoimmune gastritis and pernicious anemia and also may predispose some patients to iron deficiency through occult blood loss and/or hypochlorhydria and reduced iron absorption. In addition, several extragastrointestinal pathologies have been linked with *H. pylori* colonization, although evidence of causality is less strong. Several small studies of *H. pylori* treatment in idiopathic thrombocytopenic purpura have described improvement in or even normalization of platelet counts. Potentially important but even more controversial associations are with ischemic heart disease and cerebrovascular disease. However, the strength of these latter associations is reduced if confounding factors are taken into account, and most authorities consider the associations to be noncausal. Recent studies have shown an inverse association of *cagA*$^+$ *H. pylori* with childhood-onset asthma, hay fever, and atopic disorders. Whether *H. pylori* status is merely a marker or is causally associated with protection against these diseases remains to be determined.

Diagnosis

Tests for the presence of *H. pylori* can be divided into two groups: invasive tests, which require upper gastrointestinal endoscopy and are based on the analysis of gastric biopsy specimens, and noninvasive tests **(Table 151-1)**. Endoscopy often is not performed in the initial management of young dyspeptic patients without "alarm" symptoms but is commonly used to exclude malignancy in older patients. If endoscopy is performed, the most convenient biopsy-based test is the biopsy urease test, in which one large or two small antral biopsy specimens are placed into a gel containing urea and an indicator. The presence of *H. pylori* urease leads to a pH alteration and therefore to a color change, which often occurs within minutes but can require up to 24 h. Histologic examination of biopsy specimens for *H. pylori* also is accurate, provided that a special stain (e.g., a modified Giemsa or silver stain) permitting optimal visualization of the organism is used. If biopsy specimens are obtained from both antrum and corpus, histologic study yields additional information, including the degree and pattern of inflammation, atrophy, metaplasia, and dysplasia. Microbiologic culture is most specific but may be insensitive because of difficulty with *H. pylori* isolation. Once the organism is cultured, its identity as *H. pylori* can be confirmed by its typical appearance on Gram's stain and its positive reactions in oxidase, catalase, and urease tests. Moreover, the organism's susceptibility to antibiotics can be determined, and this information can be clinically useful in difficult cases. The occasional biopsy specimens containing the less common non-*pylori* gastric helicobacters give only weakly positive results in the biopsy urease test. Positive identification of these bacteria requires visualization of the characteristic long, tight spirals in histologic sections.

Table 151-1 Tests Commonly Used to Detect *Helicobacter Pylori*

Test	Advantages	Disadvantages
Invasive (Based on Endoscopic Biopsy)		
Biopsy urease test	Quick, simple	Some commercial tests not fully sensitive before 24 h
Histology	May give additional histologic information	Sensitivity dependent on experience and use of special stains
Culture	Permits determination of antibiotic susceptibility	Sensitivity dependent on experience
Noninvasive		
Serology	Inexpensive and convenient; not affected by recent antibiotics or proton pump inhibitors to the same extent as breath and stool tests	Cannot be used for early follow-up after treatment; some commercial kits inaccurate, and all less accurate than breath test
^{13}C urea breath test	Inexpensive and simpler than endoscopy; useful for follow-up after treatment	Requires fasting; not as convenient as blood or stool tests
Stool antigen test	Inexpensive and convenient; useful for follow-up after treatment; may be useful in children	May be disliked by people from some cultures; may be slightly less accurate than urea breath test, particularly when used to assess treatment success

Noninvasive *H. pylori* testing is the norm if gastric cancer does not need to be excluded by endoscopy. The most consistently accurate test is the urea breath test. In this simple test, the patient drinks a solution of urea labeled with the nonradioactive isotope ^{13}C and then blows into a tube. If *H. pylori* urease is present, the urea is hydrolyzed and labeled carbon dioxide is detected in breath samples. The stool antigen test, another simple assay, is more convenient and potentially less expensive than

the urea breath test but has been slightly less accurate in some comparative studies. The simplest tests for ascertaining *H. pylori* status are serologic assays measuring specific IgG levels in serum by enzyme-linked immunosorbent assay or immunoblot. The best of these tests are as accurate as other diagnostic methods, but many commercial tests—especially rapid office tests—do not perform well.

The urea breath test, the stool antigen test, and biopsy-based tests can all be used to assess the success of treatment **(Fig. 151-2)**. However, because these tests are dependent on *H. pylori* load, their use <4 weeks after treatment may yield false-negative results. Furthermore, these tests are unreliable if performed within 4 weeks of intercurrent treatment with antibiotics or bismuth compounds or within 2 weeks of the discontinuation of proton pump inhibitor (PPI) treatment. In the assessment of treatment success, noninvasive tests are normally preferred; however, after gastric ulceration, endoscopy should be repeated to ensure healing and to exclude gastric carcinoma by further histologic sampling.

Figure 151-2

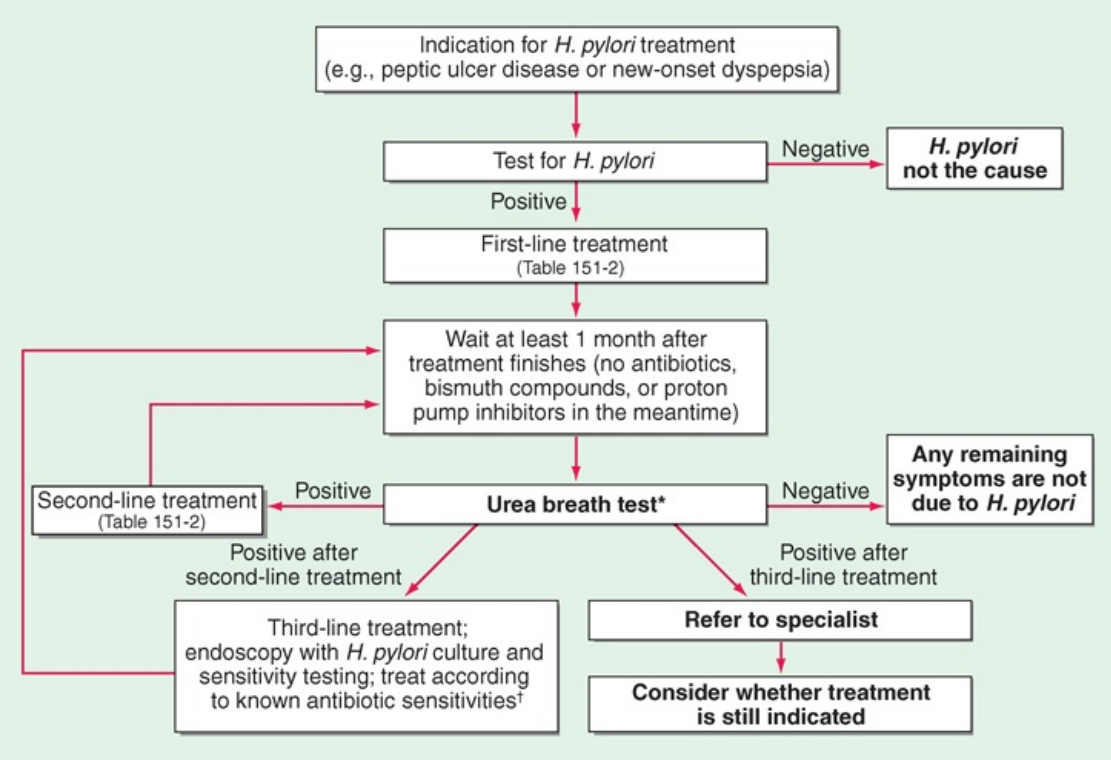

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the management of *Helicobacter pylori* infection. *Occasionally, an endoscopy and a biopsy-based test are used instead of a urea breath test in follow-up after treatment. The main indication for these invasive tests is gastric ulceration; in this condition, as opposed to duodenal ulceration, it is important to check healing and to exclude underlying gastric adenocarcinoma. †Some authorities now use empirical third-line regimens, several of which have been described.

Serologic tests are not used to monitor treatment success, as the gradual drop in titer of *H. pylori*–specific antibodies is too slow to be of practical use.

Treatment: *H. Pylori* Infections

The most clear-cut indications for treatment are *H. pylori*–related duodenal or gastric ulceration or low-grade gastric B cell lymphoma. *H. pylori* should be eradicated in patients with documented ulcer disease, whether or not the ulcers are currently active, to reduce the likelihood of relapse (Fig. 151-2). Many guidelines now recommend *H. pylori* eradication in uninvestigated simple dyspepsia following noninvasive diagnosis; others also recommend treatment in functional dyspepsia, in case the patient is one of the perhaps 7% (beyond placebo effects) to benefit from such treatment. Individuals with a strong family history of gastric cancer should be treated to eradicate *H. pylori* in the hope that their cancer risk will be reduced. Currently, widespread community screening for and treatment of *H. pylori* as primary prophylaxis for gastric cancer and peptic ulcers are not recommended, mainly because it is unclear whether treatment for *H. pylori* reduces the risk of cancer to that in persons who have never acquired the organism. The largest randomized controlled study to date (performed in China) showed no cancer risk reduction during the 7 years of follow-up, although a post hoc subgroup analysis

documented improvement in the group of participants who did not already have gastric atrophy or intestinal metaplasia. Other studies have found a reduced cancer risk after treatment, but the size of this effect in different populations remains unclear, and the results of further large-scale prospective interventional studies must be awaited. Other reasons for not treating *H. pylori* in asymptomatic populations at present include (1) the adverse side effects of the multiple-antibiotic regimens used (which are common and can be severe in rare cases); (2) antibiotic resistance, which may arise in *H. pylori* or other incidentally carried bacteria; (3) the anxiety that may arise in otherwise healthy people, especially if treatment is unsuccessful; and (4) the apparent existence of a subset of people who will develop GERD symptoms after treatment, although on average *H. pylori* treatment does not affect GERD symptoms or severity.

Although *H. pylori* is susceptible to a wide range of antibiotics in vitro, monotherapy is not usually successful, probably because of inadequate antibiotic delivery to the colonization niche. Failure of monotherapy has prompted the development of multidrug regimens, the most successful of which are triple and quadruple combinations. Initially these regimens produced *H. pylori* eradication rates of >90% in many trials; in recent years, however, resistance to key antibiotics has become more common, a trend leading to *H. pylori* eradication rates of only 75–80% for the most commonly used regimens. Current regimens consist of a PPI or ranitidine bismuth citrate and two or three antimicrobial agents given for 7–14 days (Table 151-2). Research on optimizing drug combinations to increase efficacy continues, and it is likely that guidelines will change as the field develops and as countries increasingly individualize treatment to suit local antibiotic resistance patterns and economic needs.

Table 151-2 Recommended Treatment Regimens for *Helicobacter Pylori*

Regimen (Duration)	Drug 1	Drug 2	Drug 3	Drug 4
Regimen 1: OCM (7–14 days)[a]	Omeprazole[b] (20 mg bid)	Clarithromycin (500 mg bid)	Metronidazole (500 mg bid)	—
Regimen 2: OCA (7–14 days)[a]	Omeprazole[b] (20 mg bid)	Clarithromycin (500 mg bid)	Amoxicillin (1 g bid)	—
Regimen 3: OBTM (14 days)[c]	Omeprazole[b] (20 mg bid)	Bismuth subsalicylate (2 tabs qid)	Tetracycline HCl (500 mg qid)	Metronidazole (500 mg tid)
Regimen 4[d]: sequential (5 days + 5 days)	Omeprazole[b] (20 mg bid)	Amoxicillin 1 g bid		
	Omeprazole[b] (20 mg bid)	Clarithromycin (500 mg bid)	Tinidazole (500 mg bid)	
Regimen 5[e]: OAL (10 days)	Omeprazole[b] (20 mg bid)	Amoxicillin (1 g bid)	Levofloxacin (500 mg qid)	

[a]Meta-analyses show that a 14-day course of therapy is slightly superior to a 7-day course. However, in populations where 7-day treatment is known to have very high success rates, this shorter course is still often used.

[b]Omeprazole may be replaced with any proton pump inhibitor at an equivalent dosage or, in regimens 1 and 2, with ranitidine bismuth citrate (400 mg).

[c]Data supporting this regimen come mainly from Europe and are based on the use of bismuth subcitrate and metronidazole (400 mg tid). This is the most commonly used second-line regimen.

[d]Data supporting this regimen come from Europe. Although the two 5-day courses of different drugs have usually been given sequentially, recent evidence suggests no added benefit from this approach. Thus 10 days of the four drugs combined may be as good and may aid compliance.

[e]Data supporting this second- or third-line regimen come from Europe. This regimen may be less effective where rates of quinolone use are high. Theoretically, it may also be wise to avoid it in populations where *Clostridium difficile* infection is common after broad-spectrum antibiotic use.

The two most important factors in successful *H. pylori* treatment are the patient's close compliance with the regimen and the use of drugs to which the patient's strain of *H. pylori* has not acquired resistance. Treatment failure following minor lapses in compliance is common and often leads to acquired resistance to metronidazole or clarithromycin. To stress the importance of compliance, written instructions should be given to the patient, and minor side effects of the regimen should be explained. Resistance to clarithromycin and, to a lesser extent, to metronidazole are of growing concern. Clarithromycin resistance is less prevalent but, if present, usually results in treatment failure. Strains of *H. pylori* that are apparently resistant to metronidazole are more common but still may be cleared by metronidazole-containing regimens, which have only slightly reduced efficacy. Assessment of antibiotic susceptibilities before treatment would be optimal but is not usually undertaken because endoscopy

and mucosal biopsy are necessary to obtain *H. pylori* for culture and because most microbiology laboratories are inexperienced in *H. pylori* culture. In the absence of susceptibility information, a history of the patient's (even distant) antibiotic use for other conditions should be obtained; use of the agent should then be avoided if possible, particularly in the case of clarithromycin (e.g., previous use for upper respiratory infection). If initial *H. pylori* treatment fails, one of two strategies may be used (Fig. 151-2). The more common approach is empirical re-treatment with another drug regimen, usually quadruple therapy (Table 151-2). The second approach is endoscopy, biopsy, and culture plus treatment based on documented antibiotic sensitivities. If re-treatment fails, susceptibility testing should ideally be performed, although empirical third-line therapies are often used.

Clearance of non-*pylori* gastric helicobacters can follow the use of bismuth compounds alone or of triple-drug regimens. However, in the absence of trials, it is unclear whether this outcome represents successful treatment or natural clearance of the bacterium.

Prevention

Carriage of *H. pylori* has considerable public health significance in developed countries, where it is associated with peptic ulcer disease and gastric adenocarcinoma, and in developing countries, where gastric adenocarcinoma may be an even more common cause of cancer death late in life. If mass prevention were contemplated, vaccination would be the most obvious method, and experimental immunization of animals has given promising results. However, given that *H. pylori* has co-evolved with its human host over millennia, preventing or eliminating colonization on a population basis may have distinct disadvantages. For example, lifelong absence of *H. pylori* is a risk factor for GERD complications, including esophageal adenocarcinoma. We have speculated that the disappearance of *H. pylori* may be associated with an increased risk of other emerging diseases reflecting aspects of the current Western lifestyle, such as asthma, obesity, and conceivably even type 2 diabetes mellitus.

FURTHER READINGS

Atherton JC, Blaser MJ: Co-adaptation of *Helicobacter pylori* and humans: Ancient history and modern implications. J Clin Invest 119:2475, 2009[PMID: 19729845]

Backert S, Selbach M: The role of type IV secretion in *Helicobacter pylori* pathogenesis. Cell Microbiol 10:1573, 2008[PMID: 18410539]

Chey WD et al: American College of Gastroenterology guidelines on the management of *Helicobacter pylori* infection. Am J Gastroenterol 102:1808, 2007[PMID: 17608775]

El-Omar EM et al: Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature 404:398, 2000[PMID: 10746728]

Linz B et al: An African origin for the intimate association between humans and *Helicobacter pylori* . Nature 445:915, 2007[PMID: 17287725]

Marshall BJ, Warren JR: Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1:1311, 1984[PMID: 6145023]

Ohnishi N et al: Transgenic expression of *Helicobacter pylori* CagA induces gastrointestinal and hematopoietic neoplasms in mouse. Proc Natl Acad Sci USA 105:1003, 2008[PMID: 18192401]

Polk DB, Peek RM Jr: Helicobacter pylori: Gastric cancer and beyond. Nat Rev Cancer 10:403, 2010[PMID: 20495574]

Wong BC et al: *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: A randomized controlled trial. JAMA 291:187, 2004[PMID: 14722144]

Wu C-Y et al: Early *Helicobacter pylori* eradication decreases risk of gastric cancer in patients with peptic ulcer disease. Gastroenterology 137:1641, 2009[PMID: 19664631]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS DUE TO *PSEUDOMONAS* SPECIES AND RELATED ORGANISMS: INTRODUCTION

The pseudomonads are a heterogeneous group of gram-negative bacteria that have in common an inability to ferment lactose. Formerly classified in the genus *Pseudomonas*, the members of this group are now assigned to three medically important genera—*Pseudomonas, Burkholderia*, and *Stenotrophomonas*—whose biologic behaviors encompass both similarities and marked differences and whose genetic repertoires differ in many respects. The pathogenicity of most pseudomonads is based on opportunism; the exceptions are the organisms that cause melioidosis (*B. pseudomallei*) and glanders (*B. mallei*), which can be considered primary pathogens.

P. aeruginosa, the major pathogen of the group, is a significant cause of infections in hospitalized patients and in patients with cystic fibrosis (CF; Chap. 259). Cytotoxic chemotherapy, mechanical ventilation, and broad-spectrum antibiotic therapy probably paved the way for increasing numbers of patients colonized and infected by this organism. Since the implementation of these advances in medical therapy, most conditions predisposing to *P. aeruginosa* infections have involved host compromise and/or broad-spectrum antibiotic use. The other members of the genus *Pseudomonas—P. putida, P. fluorescens*, and *P. stutzeri*—infect humans infrequently.

The genus *Burkholderia* comprises >40 species, of which *B. cepacia* is most frequently encountered in Western countries. Like *P. aeruginosa, B. cepacia* is both a nosocomial pathogen and a cause of infection in CF. The other medically important members of this genus are *B. pseudomallei* and *B. mallei,* which, as stated above, cause melioidosis and glanders, respectively.

The genus *Stenotrophomonas* contains one species of medical significance, *S. maltophilia* (previously classified in the genera *Pseudomonas* and *Xanthomonas*). This organism is strictly an opportunist that "overgrows" in the setting of potent broad-spectrum antibiotic use.

PSEUDOMONAS AERUGINOSA

Epidemiology

P. aeruginosa is found in most moist environments. Soil, plants, vegetables, tap water, and countertops can all be reservoirs for this microbe, as it has simple nutritional needs. Given the ubiquity of *P. aeruginosa*, contact with the organism obviously is not sufficient for colonization or infection. Clinical and experimental observations suggest that *P. aeruginosa* infection often occurs concomitantly with host defense compromise, mucosal trauma, physiologic derangement, and antibiotic-mediated suppression of normal flora. Thus, it comes as no surprise that the majority of *P. aeruginosa* infections occur in intensive care units (ICUs), where these factors frequently converge. It is believed that the organism is initially acquired from environmental sources, but patient-to-patient spread also occurs in clinics and families.

Burn patients once appeared to be unusually susceptible to *P. aeruginosa*. For example, in 1959–1963, *Pseudomonas* burn-wound sepsis was the principal cause of death in 60% of burned patients dying at the U.S. Army Institute of Surgical Research. For reasons that are unclear, *P. aeruginosa* infection in burns is no longer the major problem that it was during the 1950s and 1960s. Similarly, in the 1960s, *P. aeruginosa* appeared as a common pathogen in patients receiving cytotoxic chemotherapy at many institutions in the United States, but it subsequently diminished in importance. Despite this subsidence, *P. aeruginosa* remains one of the most feared pathogens in this population because of its high attributable mortality rate. In some parts of Asia and Latin America, *P. aeruginosa* continues to be the most common cause of gram-negative bacteremia in neutropenic patients.

In contrast to the trends for burn patients and neutropenic patients in the United States, the incidence of *P. aeruginosa* infections among patients with CF has not changed. *P. aeruginosa* remains the most common contributing factor to respiratory failure in CF and is responsible for the majority of deaths among CF patients.

Laboratory Features

P. aeruginosa is a nonfastidious, motile, gram-negative rod that grows on most common laboratory media, including blood and MacConkey agars. It is easily identified in the laboratory on primary-isolation agar plates by pigment production that confers a yellow to dark green or even bluish appearance. Colonies have a shiny "gun-metal" appearance and a characteristic fruity odor. Two of

the identifying biochemical characteristics of *P. aeruginosa* are an inability to ferment lactose on MacConkey agar and a positive reaction in the oxidase test. Most strains are identified on the basis of these readily detectable laboratory features even before extensive biochemical testing is done. Some isolates from CF patients are easily identified by their mucoid appearance, which is due to the production of large amounts of the mucoid exopolysaccharide or alginate.

Pathogenesis

Unraveling the mechanisms underlying disease caused by *P. aeruginosa* has proved challenging. Of the common gram-negative bacteria, no other species produces such a large number of putative virulence factors **(Table 152-1)**. Yet *P. aeruginosa* rarely initiates an infectious process in the absence of host injury or compromise, and few of its putative virulence factors have been shown definitively to be involved in disease in humans. Despite its metabolic versatility and possession of multiple colonizing factors, *P. aeruginosa* exhibits no competitive advantage over enteric bacteria in the human gut; neither is it a normal inhabitant of the human gastrointestinal tract, despite the host's continuous environmental exposure to the organism.

Table 152-1 Main Putative Virulence Factors of *Pseudomonas Aeruginosa*

Substance/Organelle	Function	Virulence in Animal Disease
Pili	Adhesion to cells	?
Flagella	Adhesion, motility, inflammation	Yes
Lipopolysaccharide	Antiphagocytic activity, inflammation	Yes
Type III secretion system	Cytotoxic activity (ExoU)	Yes
Proteases	Proteolytic activity, cytotoxicity	?
Phospholipases	Cytotoxicity	?
Exotoxin A	Cytotoxicity	?

VIRULENCE ATTRIBUTES INVOLVED IN ACUTE *P. AERUGINOSA* INFECTIONS

Motility and Colonization

A general tenet of bacterial pathogenesis is that most bacteria must adhere to surfaces or colonize a host niche in order to initiate disease. Most pathogens examined thus far possess adherence factors called *adhesins*. *P. aeruginosa* is no exception. Among its many adhesins are its pili, which demonstrate adhesive properties for a variety of cells and adhere best to injured cell surfaces. In the organism's flagellum, the flagellin molecule binds to cells, and the flagellar cap attaches to mucins through the recognition of glycan chains. Nonflagellated *P. aeruginosa* mutants are less virulent or avirulent in some but not all animal models; however, it is unclear whether this decreased virulence is due to the loss of adhesion or to the loss of other flagellar functions. Other *P. aeruginosa* adhesins include the outer core of the lipopolysaccharide (LPS) molecule, which binds to the cystic fibrosis transmembrane conductance regulator (CFTR) and aids in internalization of the organism, and the alginate coat of mucoid strains, which enhances adhesion to cells and mucins. In addition, membrane proteins and lectins have been proposed as colonization factors. It appears that the deletion of any given adhesin is not sufficient to abrogate the ability of *P. aeruginosa* to colonize surfaces.

Evasion of Host Defenses

The transition from bacterial colonization to disease requires the evasion of host defenses by a substantial number of bacteria. *P. aeruginosa* appears to be well equipped for evasion. Attached bacteria inject four known toxins (ExoS, ExoU, ExoT, and ExoY) via a type III secretion system that allows the bacteria to evade phagocytic cells either by cytotoxicity or by inhibition of phagocytosis. Mutants with defects in this system fail to disseminate in some animal models of infection. Secreted toxins such as exotoxin A and leukocidin have the potential to kill phagocytic cells, and multiple secreted proteases may degrade host effector molecules such as cytokines and chemokines that are released in response to infection.

Tissue Injury

Among gram-negative bacteria, *P. aeruginosa* probably produces the largest number of substances that are toxic to cells and thus may injure tissues. The toxins secreted by its type III secretion system are capable of tissue injury. However, their delivery requires the adherence of the organism to cells. Thus, the effects of these toxins are likely to be local or to depend on the presence of vast numbers of bacteria. On the other hand, diffusible toxins, secreted by the organism's type II secretion system, can act freely wherever they come into contact with cells. Exotoxin A, four different proteases, at least two phospholipases, rhamnolipids, pyocyanin, and hydrocyanic acid are all produced by *P. aeruginosa* and are all capable of inducing host injury.

Inflammatory Components

The inflammatory components of *P. aeruginosa* [e.g., the inflammatory responses to the lipid A component of LPSs and to flagellin, mediated through the Toll-like receptor (TLR) system (principally TLR4 and TLR5)] have been thought to represent the most important factor in disease causation. Although these inflammatory responses are required for successful defense against *P. aeruginosa* (i.e., in their absence, animals are defenseless against *P. aeruginosa* infection), florid responses are likely to result in disease. When the sepsis syndrome and septic shock develop in *P. aeruginosa* infection, they are probably the result of the host response to one or both of these substances, but injury to the lung by *Pseudomonas* toxins may also result in sepsis syndromes, possibly by causing cell death and the release of cellular components (e.g., heat-shock proteins) that may activate the TLR or another proinflammatory system.

CHRONIC *P. AERUGINOSA* INFECTIONS

Chronic infection due to *P. aeruginosa* occurs mainly in the lungs in the setting of structural pulmonary diseases. The classic example is CF; others include bronchiectasis and chronic relapsing panbronchiolitis, a disease seen in Japan and some Pacific Islands. Hallmarks of these illnesses are altered mucociliary clearance leading to mucus stasis and mucus accumulation in the lungs. There is probably a common factor that selects for *P. aeruginosa* colonization in these lung diseases—perhaps the adhesiveness of *P. aeruginosa* for mucus, a phenomenon that is not noted for most other common gram-negative bacteria, and/or the ability of *P. aeruginosa* to evade host defenses in mucus. Furthermore, *P. aeruginosa* seems to evolve in ways that allow its prolonged survival in the lung without an early fatal outcome for the host. The strains found in CF patients exhibit minimal production of virulence factors. Some strains even lose the ability to produce pili and flagella, and most become complement-sensitive because of the loss of the O side chain of their LPS molecules. An example of the impact of these changes is found in the organism's discontinuation of the production of flagellin (probably its most strongly proinflammatory molecule) when it encounters purulent mucus. This response probably dampens the host's response, allowing the organism to survive in mucus. *P. aeruginosa* is also believed to lose the ability to secrete many of its injectable toxins during growth in mucus. Although the alginate coat is thought to play a role in the organism's survival, alginate is not essential, as nonmucoid strains may also predominate for long periods. In short, virulence in chronic infections may be mediated mainly by the attenuated host inflammatory response, which injures the lungs over decades.

Clinical Manifestations

P. aeruginosa causes infections at almost all sites in the body but shows a rather strong predilection for the lungs. The infections encountered most commonly in hospitalized patients are described below.

BACTEREMIA

Crude mortality rates exceeding 50% have been reported among patients with *P. aeruginosa* bacteremia. Consequently, this clinical entity has been much feared, and its management has been attempted with the use of multiple antibiotics. Recent publications report attributable mortality rates of 28–44%, with the precise figure depending on the adequacy of treatment and the seriousness of the underlying disease. In the past, the patient with *P. aeruginosa* bacteremia classically was neutropenic or had a burn injury. Today, however, a minority of such patients have bacteremic *P. aeruginosa* infections. Rather, *P. aeruginosa* bacteremia is seen most often in patients on ICUs.

The clinical presentation of *P. aeruginosa* bacteremia rarely differs from that of sepsis in general (Chap. 270). Patients are usually febrile, but those who are most severely ill may be in shock or even hypothermic. The only point differentiating this entity from gram-negative sepsis of other causes may be the distinctive skin lesions (ecthyma gangrenosum) of *Pseudomonas* infection, which occur almost exclusively in markedly neutropenic patients and patients with AIDS. These small or large, painful, reddish, maculopapular lesions have a geographic margin; they are initially pink, then darken to purple, and finally become black and necrotic **(Fig. 152-1)**. Histopathologic studies indicate that the lesions are due to vascular invasion and are teeming with bacteria. Although similar lesions may occur in aspergillosis and mucormycosis, their presence suggests *P. aeruginosa* bacteremia as the most likely diagnosis.

Figure 152-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ecthyma gangrenosum in a neutropenic patient 3 days after onset.

Treatment: Bacteremia

(Table 152-2) Antimicrobial treatment of *P. aeruginosa* bacteremia has been controversial. Before 1971, the outcome of *Pseudomonas* bacteremia in febrile neutropenic patients treated with the available agents—gentamicin and the polymyxins—was dismal. Studies published around that time indicated that treatment with carbenicillin, with or without an aminoglycoside, significantly improved outcomes. Concurrently, several retrospective analyses suggested that the use of two agents that were synergistic against gram-negative pathogens in vitro resulted in better outcomes in neutropenic patients. Thus, combination therapy became the standard of care—first for *P. aeruginosa* bacteremia in febrile neutropenic patients and then for all *P. aeruginosa* infections in neutropenic or nonneutropenic patients.

Table 152-2 Antibiotic Treatment of Infections Due to *Pseudomonas Aeruginosa* and Related Species

Infection	Antibiotics and Dosages	Other Considerations
Bacteremia		
Nonneutropenic host	Monotherapy: Ceftazidime (2 g q8h IV) or cefepime (2 g q12h IV)Combination therapy: Piperacillin/tazobactam (3.375 g q4h IV) or imipenem (500 mg q6h IV) or meropenem (1 g q8h IV) or doripenem (500 mg q8h IV) *plus* Amikacin (7.5 mg/kg q12h or 15 mg/kg q24h IV)	Add an aminoglycoside for patients in shock and in regions or hospitals where rates of resistance to the primary β-lactam agents are high. Tobramycin may be used instead of amikacin (susceptibility permitting).
Neutropenic host	Cefepime (2 g q8h IV) or all other agents (except doripenem) in above dosages	
Endocarditis	Antibiotic regimens as for bacteremia for 6–8 weeks	Resistance during therapy is common. Surgery is required for relapse.
Pneumonia	Drugs and dosages as for bacteremia, except that the available carbapenems should not be the sole primary drugs because of high rates of resistance during therapy	IDSA guidelines recommend the addition of an aminoglycoside or ciprofloxacin. The duration of therapy is 10–14 days.
Bone infection, malignant otitis externa	Cefepime or ceftazidime at the same dosages as for bacteremia; aminoglycosides not a necessary component of therapy; ciprofloxacin (500–750 mg q12h PO) may be used	Duration of therapy varies with the drug used (e.g., 6 weeks for a β-lactam agent; at least 3 months for oral therapy except in puncture-wound osteomyelitis, for which the duration should be 2–4 weeks).
Central nervous system infection	Ceftazidime or cefepime (2 g q8h IV) or meropenem (1 g q8h IV)	Abscesses or other closed-space infections may require drainage. The duration of therapy is ≥2 weeks.
Eye infection		
Keratitis/ulcer	Topical therapy with tobramycin/ciprofloxacin/levofloxacin eyedrops	Use maximal strengths available or compounded by pharmacy.

Endophthalmitis	Ceftazidime or cefepime as for central nervous system infection *plus* Topical therapy	
Urinary tract infection	Ciprofloxacin (500 mg q12h PO) or levofloxacin (750 mg q24h) or any aminoglycoside (total daily dose given once daily)	Relapse may occur if an obstruction or a foreign body is present.
Multidrug-resistant *P. aeruginosa* infection	Colistin (100 mg q12h IV) for the shortest possible period to obtain a clinical response	Doses used have varied. Dosage adjustment is required in renal failure. Inhaled colistin may be added for pneumonia (100 mg q12h).
Stenotrophomonas maltophilia infection	TMP-SMX (1600/320 mg q12h IV for 14 days) Ticarcillin/clavulanate (3.1 g q4h IV for 14 days)	Resistance to all agents is increasing. Levofloxacin may be an alternative, but there is little published clinical experience with this agent.
Burkholderia cepacia infection	Meropenem (1 g q8h IV for 14 days) TMP-SMX (1600/320 mg q12h IV for 14 days)	Resistance to both agents is increasing. Do not use them in combination because of possible antagonism.
Melioidosis, glanders	Ceftazidime (2 g q6h for 2 weeks) or meropenem (1 g q8h for 2 weeks) or imipenem (500 mg q6h for 2 weeks) *followed by* TMP-SMX (1600/320 mg q12h PO for 3 months)	See "Further Readings" for more details on therapy and alternative agents.

Abbreviations: IDSA, Infectious Diseases Society of America; TMP-SMX, trimethoprim-sulfamethoxazole.

With the introduction of newer antipseudomonal drugs, a number of studies have revisited the choice between combination treatment and monotherapy for *Pseudomonas* bacteremia. Although the majority of experts still favor combination therapy, most of these observational studies indicate that a single modern antipseudomonal ß-lactam agent to which the isolate is sensitive is as efficacious as a combination. Even in patients at greatest risk of early death from *P. aeruginosa* bacteremia (i.e., those with fever and neutropenia), empirical antipseudomonal monotherapy is deemed to be as efficacious as empirical combination therapy by the practice guidelines of the Infectious Diseases Society of America. One firm conclusion is that monotherapy with an aminoglycoside is not optimal.

There are, of course, institutions and countries where rates of susceptibility of *P. aeruginosa* to first-line antibiotics are <80%. Thus, when a septic patient with a high probability of *P. aeruginosa* infection is encountered in such settings, empirical combination therapy should be administered until the pathogen is identified and susceptibility data become available. Thereafter, whether one or two agents should be continued remains a matter of individual preference.

ACUTE PNEUMONIA

Respiratory infections are the most common of all infections caused by *P. aeruginosa*. This organism appears first or second on most lists of the causes of ventilator-associated pneumonia (VAP). However, much debate centers on the actual role of *P. aeruginosa* in VAP. Many of the relevant data are based on cultures of sputum or endotracheal tube aspirates and may represent nonpathogenic colonization of the tracheobronchial tree, biofilms on the endotracheal tube, or simple tracheobronchitis.

Older reports of *P. aeruginosa* pneumonia described patients with an acute clinical syndrome of fever, chills, cough, and necrotizing pneumonia indistinguishable from other gram-negative bacterial pneumonias. The traditional accounts described a fulminant infection, with cyanosis, tachypnea, copious sputum, and systemic toxicity. Chest radiographs demonstrated bilateral pneumonia, often with nodular densities with or without cavities. This picture is now remarkably rare. Today, the typical patient is using a ventilator, has a slowly progressive infiltrate, and has been colonized with *P. aeruginosa* for days. While some cases may progress rapidly over 48–72 h, they are the exceptions. Nodular densities are not commonly seen. However, infiltrates may go on to necrosis. Necrotizing pneumonia has also been seen in the community (e.g., after inhalation of hot-tub water contaminated with *P. aeruginosa*). The typical patient has fever, leukocytosis, and purulent sputum, and the chest radiograph shows a new infiltrate or the expansion of a preexisting infiltrate. Chest examination generally detects rales or dullness. Of course, such findings are quite common among ventilated

patients in the ICU. A sputum Gram's stain showing mainly polymorphonuclear leukocytes (PMNs) in conjunction with a culture positive for *P. aeruginosa* in this setting suggests a diagnosis of acute *P. aeruginosa* pneumonia. There is no consensus about whether an invasive procedure (e.g., bronchoalveolar lavage or protected-brush sampling of the distal airways) is superior to tracheal aspiration to obtain samples for lung cultures in order to substantiate the occurrence of *P. aeruginosa* pneumonia and prevent antibiotic overuse.

Treatment: Acute Pneumonia

(Table 152-2) The results of therapy for *P. aeruginosa* pneumonia have been unsatisfactory. Reports suggest mortality rates of 40–80%, but how many of these deaths are attributable to underlying disease remains unknown. The drugs of choice for *P. aeruginosa* pneumonia are similar to those given for bacteremia. A potent antipseudomonal β-lactam drug is the mainstay of therapy. Failure rates were high when aminoglycosides were used as single agents, possibly because of their poor penetration into the airways and their binding to airway secretions. Thus a strong case cannot be made for the inclusion of the aminoglycoside component in regimens used against fully susceptible organisms, especially given the evidence that aminoglycosides are not optimally active in the lungs at concentrations normally reached after IV administration. Nonetheless, aminoglycosides are commonly used in clinical practice. Some experts suggest the combination of a β-lactam agent and an antipseudomonal fluoroquinolone instead when combination therapy is desired.

CHRONIC RESPIRATORY TRACT INFECTIONS

P. aeruginosa is responsible for chronic infections of the airways associated with a number of underlying or predisposing conditions—most commonly CF in Caucasian populations (Chap. 259). A state of chronic colonization beginning early in childhood is seen in some Asian populations with chronic or diffuse panbronchiolitis, a disease of unknown etiology. *P. aeruginosa* is one of the organisms that colonizes damaged bronchi in bronchiectasis, a disease secondary to multiple causes in which profound structural abnormalities of the airways result in mucus stasis.

Treatment: Chronic Respiratory Tract Infections

Optimal management of chronic *P. aeruginosa* lung infection has not been determined. Patients respond clinically to antipseudomonal therapy, but the organism is rarely eradicated. Since eradication is unlikely, the aim of treatment for chronic infection is to quell exacerbations of inflammation. The regimens used are similar to those used for pneumonia, but an aminoglycoside is almost always added because resistance is common in chronic disease. However, it may be appropriate to use an inhaled aminoglycoside preparation in order to maximize airway drug levels.

ENDOVASCULAR INFECTIONS

Infective endocarditis due to *P. aeruginosa* is a disease of IV drug users whose native valves are involved. This organism has also been reported to cause prosthetic valve endocarditis. Sites of prior native-valve injury due to the injection of foreign material such as talc or fibers probably serve as niduses for bacterial attachment to the heart valve. The manifestations of *P. aeruginosa* endocarditis resemble those of other forms of acute endocarditis in IV drug users except that the disease is more indolent than *Staphylococcus aureus* endocarditis. While most disease involves the right side of the heart, left-sided involvement is not rare and multivalvular disease is common. Fever is a common manifestation, as is pulmonary involvement (due to septic emboli to the lungs). Hence, patients may also experience chest pain and hemoptysis. Involvement of the left side of the heart may lead to signs of cardiac failure, systemic emboli, and local cardiac involvement with sinus of Valsalva abscesses and conduction defects. Skin manifestations are rare in this disease, and ecthyma gangrenosum is not seen. The diagnosis is based on positive blood cultures along with clinical signs of endocarditis.

Treatment: Endovascular Infections

(Table 152-2) It has been customary to use synergistic antibiotic combinations in treating *P. aeruginosa* endocarditis because of the development of resistance during therapy with a single antipseudomonal β-lactam agent. Which combination therapy is preferable is unclear, as all combinations have failed. Cases of *P. aeruginosa* endocarditis that relapse during or fail to respond to therapy are often caused by resistant organisms and may require surgical therapy. Other considerations for valve replacement are similar to those in other forms of endocarditis (Chap. 124).

BONE AND JOINT INFECTIONS

Although *P. aeruginosa* is an infrequent cause of bone and joint infections, *Pseudomonas* bacteremia or infective endocarditis caused by the injection of contaminated illicit drugs has been well documented to result in vertebral osteomyelitis and sternoclavicular joint arthritis. The clinical presentation of vertebral *P. aeruginosa* osteomyelitis is more indolent than that of staphylococcal osteomyelitis. The duration of symptoms in IV drug users with vertebral osteomyelitis due to *P. aeruginosa* varies from weeks to months. Fever is not uniformly present; when present, it tends to be low grade. There may be mild tenderness at the site of involvement. Blood cultures are usually negative unless there is concomitant endocarditis. The erythrocyte sedimentation rate (ESR) is

generally elevated. Vertebral osteomyelitis due to *P. aeruginosa* has also been reported in the elderly, in whom it originates from urinary tract infections (UTIs). The infection generally involves the lumbosacral area because of a shared venous drainage (Batson's plexus) between the lumbosacral spine and the pelvis. Sternoclavicular septic arthritis due to *P. aeruginosa* is seen almost exclusively in IV drug users. This disease may occur with or without endocarditis, and a primary site of infection often is not found. Plain radiographs show joint or bone involvement. Treatment of these forms of disease is generally successful.

Pseudomonas osteomyelitis of the foot most often follows puncture wounds through sneakers and mostly affects children. The main manifestation is pain in the foot, sometimes with superficial cellulitis around the puncture wound and tenderness on deep palpation of the wound. Multiple joints or bones of the foot may be involved. Systemic symptoms are generally absent, and blood cultures are usually negative. Radiographs may or may not be abnormal, but the bone scan is usually positive, as are MRI studies. Needle aspiration usually yields a diagnosis. Prompt surgery, with exploration of the nail puncture tract and debridement of the involved bones and cartilage, is generally recommended in addition to antibiotic therapy.

CENTRAL NERVOUS SYSTEM (CNS) INFECTIONS

CNS infections due to *P. aeruginosa* are relatively rare. Involvement of the CNS is almost always secondary to a surgical procedure or head trauma. The entity seen most often is postoperative or posttraumatic meningitis. Subdural or epidural infection occasionally results from contamination of these areas. Embolic disease arising from endocarditis in IV drug users and leading to brain abscesses has also been described. The cerebrospinal fluid (CSF) profile of *P. aeruginosa* meningitis is no different from that of pyogenic meningitis of any other etiology.

Treatment: Central Nervous System Infections

(Table 152-2) Treatment of *Pseudomonas* meningitis is difficult; little information has been published, and no controlled trials in humans have been undertaken. However, the general principles involved in the treatment of meningitis apply, including the need for high doses of bactericidal antibiotics to attain high drug levels in the CSF. The agent with which there is the most published experience in *P. aeruginosa* meningitis is ceftazidime, but other antipseudomonal β-lactam drugs that reach high CSF concentrations, such as cefepime and meropenem, have also been used successfully. Other forms of *P. aeruginosa* CNS infection, such as brain abscesses and epidural and subdural empyema, generally require surgical drainage in addition to antibiotic therapy.

EYE INFECTIONS

Eye infections due to *P. aeruginosa* occur mainly as a result of direct inoculation into the tissue during trauma or surface injury by contact lenses. Keratitis and corneal ulcers are the most common types of eye disease and are often associated with contact lenses (especially the extended-wear variety). Keratitis can be slowly or rapidly progressive, but the classic description is disease progressing over 48 h to involve the entire cornea, with opacification and sometimes perforation. *P. aeruginosa* keratitis should be considered a medical emergency because of the rapidity with which it can progress to loss of sight. *P. aeruginosa* endophthalmitis secondary to bacteremia is the most devastating of *P. aeruginosa* eye infections. The disease is fulminant, with severe pain, chemosis, decreased visual acuity, anterior uveitis, vitreous involvement, and panophthalmitis.

Treatment: Eye Infections

(Table 152-2) The usual therapy for keratitis is the administration of topical antibiotics. Therapy for endophthalmitis includes the use of high-dose local and systemic antibiotics (to achieve higher drug concentrations in the eye) and vitrectomy.

EAR INFECTIONS

P. aeruginosa infections of the ears vary from mild swimmer's ear to serious life-threatening infections with neurologic sequelae. Swimmer's ear is common among children and results from infection of moist macerated skin of the external ear canal. Most cases resolve with treatment, but some patients develop chronic drainage. Swimmer's ear is managed with topical antibiotic agents (otic solutions). The most serious form of *Pseudomonas* infection involving the ear has been given various names: two of these designations, malignant otitis externa and necrotizing otitis externa, are now used for the same entity. This disease was originally described in elderly diabetic patients, in whom the majority of cases still occur. However, it has also been described in patients with AIDS and in elderly patients without underlying diabetes or immunocompromise. The usual presenting symptoms are decreased hearing and ear pain, which may be severe and lancinating. The pinna is usually painful, and the external canal may be tender. The ear canal almost always shows signs of inflammation, with granulation tissue and exudate. Tenderness anterior to the tragus may extend as far as the temporomandibular joint and mastoid process. A small minority of patients have systemic symptoms. Patients in whom the diagnosis is made late may present with cranial nerve palsies or even with cavernous venous sinus thrombosis. The ESR is invariably elevated (≥100 mm/h). The diagnosis is made on clinical grounds in severe cases; however, the "gold standard" is a positive technetium-99 bone scan in a patient with otitis

externa due to *P. aeruginosa*. In diabetic patients, a positive bone scan constitutes presumptive evidence for this diagnosis and should prompt biopsy or empirical therapy.

Treatment: Ear Infections

(Table 152-2) Given the infection of the ear cartilage, sometimes with mastoid or petrous ridge involvement, patients with malignant (necrotizing) otitis externa are treated as for osteomyelitis.

URINARY TRACT INFECTIONS

UTIs due to *P. aeruginosa* generally occur as a complication of a foreign body in the urinary tract, an obstruction in the genitourinary system, or urinary tract instrumentation or surgery. However, UTIs caused by *P. aeruginosa* have been described in pediatric outpatients without stones or evident obstruction.

Treatment: Urinary Tract Infections

(Table 152-2) Most *P. aeruginosa* UTIs are considered complicated infections that must be treated longer than uncomplicated cystitis. In general, a 7- to 10-day course of treatment suffices, with up to 2 weeks of therapy in cases of pyelonephritis. Urinary catheters, stents, or stones should be removed to prevent relapse, which is common and may be due not to resistance but rather to factors such as a foreign body that has been left in place or an ongoing obstruction.

SKIN AND SOFT TISSUE INFECTIONS

Besides pyoderma gangrenosum in neutropenic patients, folliculitis and other papular or vesicular lesions due to *P. aeruginosa* have been extensively described and are collectively referred to as *dermatitis*. Multiple outbreaks have been linked to whirlpools, spas, and swimming pools. To prevent such outbreaks, the growth of *P. aeruginosa* in the home and in recreational environments must be controlled by proper chlorination of water. Most cases of hot-tub folliculitis are self-limited, requiring only the avoidance of exposure to the contaminated source of water.

Toe-web infections occur especially often in the tropics, and the "green nail syndrome" is caused by *P. aeruginosa* paronychia, which results from frequent submersion of the hands in water. In the latter entity, the green discoloration results from diffusion of pyocyanin into the nail bed. *P. aeruginosa* remains a prominent cause of burn wound infections in some parts of the world. The management of these infections is best left to specialists in burn wound care.

INFECTIONS IN FEBRILE NEUTROPENIC PATIENTS

In febrile neutropenia, *P. aeruginosa* has historically been the organism against which empirical coverage is always essential. In the 1960s and early 1970s, *P. aeruginosa* infection occurred commonly in febrile neutropenic patients, with high associated mortality rates. Although in Western countries these infections are now less common, their importance has not diminished because of persistently high mortality rates. In other parts of the world as well, *P. aeruginosa* continues to be a significant problem in febrile neutropenia, causing a larger proportion of infections in febrile neutropenic patients than any other single organism. For example, *P. aeruginosa* was responsible for 28% of documented infections in 499 febrile neutropenic patients in one study from the Indian subcontinent and for 31% of such infections in another. In a large study of infections in leukemia patients from Japan, *P. aeruginosa* was the most frequently documented cause of bacterial infection. In studies performed in North America, northern Europe, and Australia, the incidence of *P. aeruginosa* bacteremia in febrile neutropenia was quite variable. In a review of 97 reports published in 1987–1994, the incidence was reported to be 1–2.5% among febrile neutropenic patients given empirical therapy and 5–12% among microbiologically documented infections. The most common clinical syndromes encountered were bacteremia, pneumonia, and soft tissue infections manifesting mainly as ecthyma gangrenosum.

Treatment: Infections in Febrile Neutropenic Patients

(Table 152-2) Compared with rates three decades ago, improved rates of response to antibiotic therapy have been reported in many studies. A study of 127 patients demonstrated a reduction in the mortality rate from 71% to 25% with the introduction of ceftazidime and imipenem. Since neutrophils —the normal host defenses against this organism—are absent in febrile neutropenic patients, maximal doses of antipseudomonal ß-lactam antibiotics should be used for the management of *P. aeruginosa* bacteremia in this setting.

INFECTIONS IN PATIENTS WITH AIDS

Both community- and hospital-acquired *P. aeruginosa* infections were documented in patients with AIDS before the advent of antiretroviral therapy. Since the introduction of protease inhibitors, *P. aeruginosa* infections in AIDS patients have been seen less frequently but still occur, particularly in the form of sinusitis. The clinical presentation of *Pseudomonas* infection (especially pneumonia and bacteremia) in AIDS patients is remarkable in that, although the illness may appear not to be severe, the infection may nonetheless be fatal. Patients with bacteremia may have only a low-grade fever and may present with ecthyma gangrenosum. Bacteremia may herald underlying disease at another site

(often pneumonia or sinusitis). Pneumonia, with or without bacteremia, is perhaps the most common type of *P. aeruginosa* infection in AIDS patients. Patients with AIDS and *P. aeruginosa* pneumonia exhibit the classic clinical signs and symptoms of pneumonia, such as fever, productive cough, and chest pain. The infection may be lobar or multilobar and shows no predisposition for any particular location. The most striking feature is the high frequency of cavitary disease.

Treatment: Infections in Patients with AIDS

Therapy for any of these conditions in AIDS patients is no different from that in other patients. However, relapse is the rule unless the patient's CD4+ T cell count rises to >50/μL or suppressive antibiotic therapy is given. In attempts to achieve cures and prevent relapses, therapy tends to be more prolonged than in the case of an immunocompetent patient.

MULTIDRUG-RESISTANT INFECTIONS

(Table 152-2) *P. aeruginosa* is notorious for antibiotic resistance. During three decades, the impact of resistance was minimized by the rapid development of potent antipseudomonal agents. However, the situation has recently changed, with the worldwide selection of strains carrying determinants that mediate resistance to β-lactams, fluoroquinolones, and aminoglycosides. This situation has been compounded by the lack of development of new classes of antipseudomonal drugs for nearly two decades. Physicians now resort to drugs such as colistin and polymyxin, which were discarded decades ago. These alternative approaches to the management of multiresistant *P. aeruginosa* infections were first used some time ago in CF patients, who receive colistin (polymyxin E) IV and by aerosol despite its renal toxicity. Colistin is rapidly becoming the last-resort agent of choice, even in non-CF patients infected with multiresistant *P. aeruginosa*.

The clinical outcome of multidrug-resistant *P. aeruginosa* infections treated with colistin is difficult to judge from case reports, especially given the many drugs used in the complicated management of these patients. Although earlier reports described marginal efficacy and serious nephrotoxicity and neurotoxicity, recent reports have been more encouraging. Because colistin shows synergy with other antimicrobial agents in vitro, it may be possible to reduce the dosage—and thus the toxicity—of this drug when it is combined with drugs such as rifampin and β-lactams; however, no studies in humans or animals support this approach at this time.

OTHER PSEUDOMONADS

Stenotrophomonas Maltophilia

S. maltophilia is the only potential human pathogen among a genus of ubiquitous organisms found in the rhizosphere (i.e., the soil that surrounds the roots of plants). The organism is an opportunist that is acquired from the environment but is even more limited than *P. aeruginosa* in its ability to colonize patients or cause infections. Immunocompromise is not sufficient to permit these events; rather, major perturbations of the human flora are usually necessary for the establishment of *S. maltophilia*. Accordingly, most cases of human infection occur in the setting of very broad-spectrum antibiotic therapy with agents such as advanced cephalosporins and carbapenems, which eradicate the normal flora and other pathogens. The remarkable ability of *S. maltophilia* to resist virtually all classes of antibiotics is attributable to the possession of antibiotic efflux pumps and of two β-lactamases (L1 and L2) that mediate β-lactam resistance, including that to carbapenems. It is fortunate that the virulence of *S. maltophilia* appears to be limited. Although a serine protease is present in some strains, virulence is probably a result of the host's inflammatory response to components of the organism such as LPS and flagellin. *S. maltophilia* is most commonly found in the respiratory tract of ventilated patients, where the distinction between its roles as a colonizer and as a pathogen is often difficult to make. However, *S. maltophilia* does cause pneumonia and bacteremia in such patients, and these infections have led to septic shock. Also common is central venous line–associated infection (with or without bacteremia), which has been reported most often in patients with cancer. *S. maltophilia* is a rare cause of ecthyma gangrenosum in neutropenic patients. It has been isolated from ~5% of CF patients but is not believed to be a significant pathogen in this setting.

TREATMENT: S. *MALTOPHILIA* INFECTIONS

The intrinsic resistance of *S. maltophilia* to most antibiotics renders infection difficult to treat. The antibiotics to which it is most often (although not uniformly) susceptible are trimethoprim-sulfamethoxazole (TMP-SMX), ticarcillin/clavulanate, and levofloxacin (Table 152-2). Consequently, a combination of TMP-SMX and ticarcillin/clavulanate is recommended for initial therapy. Catheters must be removed in the treatment of bacteremia to hasten cure and prevent relapses. The treatment of VAP due to *S. maltophilia* is much more difficult than that of bacteremia, with the frequent development of resistance during therapy.

Burkholderia Cepacia

B. cepacia gained notoriety as the cause of a rapidly fatal syndrome of respiratory distress and septicemia (the "cepacia syndrome") in CF patients. Previously, it had been recognized as an antibiotic-resistant nosocomial pathogen (then designated *P. cepacia*) in ICU patients. Patients with

chronic granulomatous disease are also predisposed to *B. cepacia* lung disease. The organism has been reclassified into nine subgroups, only some of which are common in CF. *B. cepacia* is an environmental organism that inhabits moist environments and is found in the rhizosphere. This organism possesses multiple virulence factors that may play roles in disease as well as colonizing factors that are capable of binding to lung mucus—an ability that may explain the predilection of *B. cepacia* for the lungs in CF. *B. cepacia* secretes elastase and possesses components of an injectable toxin-secretion system like that of *P. aeruginosa*; its LPS is among the most potent of all LPSs in stimulating an inflammatory response in the lungs. Inflammation may be the major cause of the lung disease seen in the cepacia syndrome. The organism can penetrate epithelial surfaces by virtue of motility and inhibition of host innate immune defenses. Besides infecting the lungs in CF, *B. cepacia* appears as an airway colonizer during broad-spectrum antibiotic therapy and is a cause of VAP, catheter-associated infections, and wound infections.

TREATMENT: *B. CEPACIA* INFECTIONS

B. cepacia is intrinsically resistant to many antibiotics. Therefore, treatment must be tailored according to sensitivities. TMP-SMX, meropenem, and doxycycline are the most effective agents in vitro and may be started as first-line agents (Table 152-2). Some strains are susceptible to third-generation cephalosporins and fluoroquinolones, and these agents may be used against isolates known to be susceptible. Combination therapy for serious pulmonary infection (e.g., in CF) is suggested for multidrug-resistant strains; the combination of meropenem and TMP-SMX may be antagonistic, however. Resistance to all agents used has been reported during therapy.

Burkholderia Pseudomallei

B. pseudomallei is the causative agent of melioidosis, a disease of humans and animals that is geographically restricted to Southeast Asia and northern Australia, with occasional cases in countries such as India and China. This organism may be isolated from individuals returning directly from these endemic regions and from military personnel who have served in endemic regions and then returned home after stops in Europe. Symptoms of this illness may develop only at a later date because of the organism's ability to cause latent infections. *B. pseudomallei* is found in soil and water. Humans and animals are infected by inoculation, inhalation, or ingestion; only rarely is the organism transmitted from person to person. Humans are not colonized without being infected. Among the pseudomonads, *B. pseudomallei* is perhaps the most virulent. Host compromise is not an essential prerequisite for disease, although many patients have common underlying medical diseases (e.g., diabetes or renal failure). *B. pseudomallei* is a facultative intracellular organism whose replication in PMNs and macrophages may be aided by the possession of a polysaccharide capsule. The organism also possesses elements of a type III secretion system that plays a role in its intracellular survival. During infection, there is a florid inflammatory response whose role in disease is unclear.

B. pseudomallei causes a wide spectrum of disease, ranging from asymptomatic infection to abscesses, pneumonia, and disseminated disease. It is a significant cause of fatal community-acquired pneumonia and septicemia in endemic areas, with mortality rates as high as 44% reported in Thailand. Acute pulmonary infections are the most commonly diagnosed form of melioidosis. Pneumonia may be asymptomatic (with routine chest radiographs showing mainly upper-lobe infiltrates) or may present as severe necrotizing disease. *B. pseudomallei* also causes chronic pulmonary infections with systemic manifestations that mimic those of tuberculosis, including chronic cough, fever, hemoptysis, night sweats, and cavitary lung disease. Besides pneumonia, the other principal form of *B. pseudomallei* disease is skin ulceration with associated lymphangitis and regional lymphadenopathy. Spread from the lungs or skin, which is most often documented in debilitated individuals, gives rise to septicemic forms of melioidosis that carry a high mortality rate.

TREATMENT: *B. PSEUDOMALLEI* INFECTIONS

B. pseudomallei is susceptible to advanced penicillins and cephalosporins and to carbapenems (Table 152-2). Treatment is divided into two stages: an intensive 2-week phase of therapy with ceftazidime or a carbapenem followed by at least 12 weeks of oral TMP-SMX to eradicate the organism and prevent relapse. The recognition of this bacterium as a potential agent of biologic warfare has stimulated interest in the development of a vaccine.

Burkholderia Mallei

B. mallei causes the equine disease glanders in Africa, Asia, and South America. The organism was eradicated from Europe and North America decades ago. The last case seen in the United States occurred in 2001 in a laboratory worker; before that, *B. mallei* had last been seen in this country in 1949. In contrast to the other organisms discussed in this chapter, *B. mallei* is not an environmental organism and does not persist outside its equine hosts. Consequently, *B. mallei* infection is an occupational risk for handlers of horses, equine butchers, and veterinarians in areas of the world where it still exists. The polysaccharide capsule is a critical virulence determinant; diabetics are thought to be more susceptible to infection by this organism. The organism is transmitted from animals to humans by inoculation into the skin, where it causes local infection with nodules and lymphadenitis.

Regional lymphadenopathy is common. Respiratory secretions from infected horses are extremely infectious. Inhalation results in clinical signs of typical pneumonia but may also cause an acute febrile illness with ulceration of the trachea. The organism may disseminate from the skin or lungs to cause septicemia with signs of sepsis. The septicemic form is frequently associated with shock and a high mortality rate. The infection may also enter a chronic phase and present as disseminated abscesses. *B. mallei* infection may present as early as 1–2 days after inhalation or (in cutaneous disease) may not become evident for months.

TREATMENT: *B. MALLEI* INFECTIONS

The antibiotic susceptibility pattern of *B. mallei* is similar to that of *B. pseudomallei;* in addition, the organism is susceptible to the newer macrolides azithromycin and clarithromycin. *B. mallei* infection should be treated with the same drugs and for the same duration as melioidosis.

FURTHER READINGS

Chastre J et al: Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults: A randomized trial. JAMA 290:2588, 2003[PMID: 14625336]

Currie BJ: *Burkholderia pseudomallei* and *Burkholderia mallei*: Melioidosis and glanders, in *Principles and Practice of Infectious Diseases,* 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 2869–2879

Johnson MP, Ramphal R: Malignant external otitis: Report on therapy with ceftazidime and review of therapy and prognosis. Rev Infect Dis 12:173, 1990[PMID: 2184490]

Kallel H et al: Colistin as a salvage therapy for nosocomial infections caused by multidrug-resistant bacteria in the ICU. Int J Antimicrob Agents 28:366, 2006[PMID: 16971093]

Lodise TP Jr et al: Predictors of 30-day mortality among patients with *Pseudomonas aeruginosa* bloodstream infections: Impact of delayed appropriate antibiotic selection. Antimicrob Agents Chemother 51:3510, 2007[PMID: 17646415]

Maschmeyer G, Göbel UB: *Stenotrophomonas maltophilia* and *Burkholderia cepacia* complex, in *Principles and Practice of Infectious Diseases,* 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 2861–2868

Mendelson MH et al: *Pseudomonas aeruginosa* bacteremia in AIDS. Clin Infect Dis 18:886, 1994[PMID: 8086548]

Micek ST et al: *Pseudomonas aeruginosa* bloodstream infection: Importance of appropriate antibiotic therapy. Antimicrob Agents Chemother 49:1306, 2005[PMID: 15793102]

Obritsch MD et al: Nosocomial infections due to multidrug resistant *Pseudomonas aeruginosa*: Epidemiology and treatment options. Pharmacotherapy 25:1353, 2006

Pier GB, Ramphal R: *Pseudomonas aeruginosa*, in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 2835–2860

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SALMONELLOSIS: INTRODUCTION

Bacteria of the genus *Salmonella* are highly adapted for growth in both humans and animals and cause a wide spectrum of disease. The growth of serotypes *S. typhi* and *S. paratyphi* is restricted to human hosts, in whom these organisms cause enteric (typhoid) fever. The remaining serotypes (nontyphoidal *Salmonella*, or NTS) can colonize the gastrointestinal tracts of a broad range of animals, including mammals, reptiles, birds, and insects. More than 200 serotypes are pathogenic to humans, in whom they often cause gastroenteritis and can be associated with localized infections and/or bacteremia.

Etiology

This large genus of gram-negative bacilli within the family Enterobacteriaceae consists of two species: *S. enterica*, which contains six subspecies, and *S. bongori*. *S. enterica* subspecies I includes almost all the serotypes pathogenic for humans. According to the current *Salmonella* nomenclature system, the full taxonomic designation *S. enterica* subspecies *enterica* serotype *typhimurium* can be shortened to *Salmonella* serotype *typhimurium* or simply *S. typhimurium*.

Members of the seven *Salmonella* subspecies are classified into >2500 serotypes (serovars) according to the somatic O antigen [lipopolysaccharide (LPS) cell-wall components], the surface Vi antigen (restricted to *S. typhi* and *S. paratyphi* C), and the flagellar H antigen. For simplicity, most *Salmonella* serotypes are named for the city where they were identified, and the serotype is often used as the species designation.

Salmonellae are gram-negative, non-spore-forming, facultatively anaerobic bacilli that measure 2–3 by 0.4–0.6 μm. The initial identification of salmonellae in the clinical microbiology laboratory is based on growth characteristics. Salmonellae, like other Enterobacteriaceae, produce acid on glucose fermentation, reduce nitrates, and do not produce cytochrome oxidase. In addition, all salmonellae except *S. gallinarum-pullorum* are motile by means of peritrichous flagella, and all but *S. typhi* produce gas (H_2S) on sugar fermentation. Notably, only 1% of clinical isolates ferment lactose; a high level of suspicion must be maintained to detect these rare clinical lactose-fermenting isolates.

Although serotyping of all surface antigens can be used for formal identification, most laboratories perform a few simple agglutination reactions that define specific O-antigen serogroups, designated A, B, C_1, C_2, D, and E. Strains in these six serogroups cause ~99% of *Salmonella* infections in humans and other warm-blooded animals. Molecular typing methods, including pulsed-field gel electrophoresis and polymerase chain reaction (PCR) fingerprinting, are used in epidemiologic investigations to differentiate *Salmonella* strains of a common serotype.

Pathogenesis

All *Salmonella* infections begin with ingestion of organisms, most commonly in contaminated food or water. The infectious dose is 10^3–10^6 colony-forming units. Conditions that decrease either stomach acidity (an age of <1 year, antacid ingestion, or achlorhydric disease) or intestinal integrity (inflammatory bowel disease, prior gastrointestinal surgery, or alteration of the intestinal flora by antibiotic administration) increase susceptibility to *Salmonella* infection.

Once *S. typhi* and *S. paratyphi* reach the small intestine, they penetrate the mucus layer of the gut and traverse the intestinal layer through phagocytic microfold (M) cells that reside within Peyer's patches. Salmonellae can trigger the formation of membrane ruffles in normally nonphagocytic epithelial cells. These ruffles reach out and enclose adherent bacteria within large vesicles by a process referred to as *bacteria-mediated endocytosis* (BME). BME is dependent on the direct delivery of *Salmonella* proteins into the cytoplasm of epithelial cells by a specialized bacterial secretion system (*type III secretion*). These bacterial proteins mediate alterations in the actin cytoskeleton that are required for *Salmonella* uptake.

After crossing the epithelial layer of the small intestine, *S. typhi* and *S. paratyphi*, which cause enteric (typhoid) fever, are phagocytosed by macrophages. These salmonellae survive the antimicrobial environment of the macrophage by sensing environmental signals that trigger alterations in regulatory systems of the phagocytosed bacteria. For example, PhoP/PhoQ (the best-characterized regulatory system) triggers the expression of outer-membrane proteins and mediates modifications in LPS so that

the altered bacterial surface can resist microbicidal activities and potentially alter host cell signaling. In addition, salmonellae encode a second type III secretion system that directly delivers bacterial proteins across the phagosome membrane into the macrophage cytoplasm. This secretion system functions to remodel the *Salmonella*-containing vacuole, promoting bacterial survival and replication.

Once phagocytosed, typhoidal salmonellae disseminate throughout the body in macrophages via the lymphatics and colonize reticuloendothelial tissues (liver, spleen, lymph nodes, and bone marrow). Patients have relatively few or no signs and symptoms during this initial incubation stage. Signs and symptoms, including fever and abdominal pain, probably result from secretion of cytokines by macrophages and epithelial cells in response to bacterial products that are recognized by innate immune receptors when a critical number of organisms have replicated. Over time, the development of hepatosplenomegaly is likely to be related to the recruitment of mononuclear cells and the development of a specific acquired cell-mediated immune response to *S. typhi* colonization. The recruitment of additional mononuclear cells and lymphocytes to Peyer's patches during the several weeks after initial colonization/infection can result in marked enlargement and necrosis of the Peyer's patches, which may be mediated by bacterial products that promote cell death as well as the inflammatory response.

In contrast to enteric fever, which is characterized by an infiltration of mononuclear cells into the small-bowel mucosa, NTS gastroenteritis is characterized by massive polymorphonuclear leukocyte (PMN) infiltration into both the large- and small-bowel mucosa. This response appears to depend on the induction of interleukin (IL) 8, a strong neutrophil chemotactic factor, which is secreted by intestinal cells as a result of *Salmonella* colonization and translocation of bacterial proteins into host cell cytoplasm. The degranulation and release of toxic substances by neutrophils may result in damage to the intestinal mucosa, causing the inflammatory diarrhea observed with nontyphoidal gastroenteritis.

ENTERIC (TYPHOID) FEVER

Enteric (typhoid) fever is a systemic disease characterized by fever and abdominal pain and caused by dissemination of *S. typhi* or *S. paratyphi*. The disease was initially called *typhoid fever* because of its clinical similarity to typhus. However, in the early 1800s, typhoid fever was clearly defined pathologically as a unique illness on the basis of its association with enlarged Peyer's patches and mesenteric lymph nodes. In 1869, given the anatomic site of infection, the term *enteric fever* was proposed as an alternative designation to distinguish typhoid fever from typhus. However, to this day, the two designations are used interchangeably.

Epidemiology

In contrast to other *Salmonella* serotypes, the etiologic agents of enteric fever—*S. typhi* and *S. paratyphi* serotypes A, B, and C—have no known hosts other than humans. Most commonly, food-borne or waterborne transmission results from fecal contamination by ill or asymptomatic chronic carriers. Sexual transmission between male partners has been described. Health care workers occasionally acquire enteric fever after exposure to infected patients or during processing of clinical specimens and cultures.

With improvements in food handling and water/sewage treatment, enteric fever has become rare in developed nations. Worldwide, however, there are an estimated 22 million cases of enteric fever, with 200,000 deaths annually. The incidence is highest (>100 cases per 100,000 population per year) in south central and Southeast Asia; medium (10–100 cases per 100,000) in the rest of Asia, Africa, Latin America, and Oceania (excluding Australia and New Zealand); and low in other parts of the world **(Fig. 153-1**). A high incidence of enteric fever correlates with poor sanitation and lack of access to clean drinking water. In endemic regions, enteric fever is more common in urban than rural areas and among young children and adolescents. Risk factors include contaminated water or ice, flooding, food and drinks purchased from street vendors, raw fruits and vegetables grown in fields fertilized with sewage, ill household contacts, lack of hand washing and toilet access, and evidence of prior *Helicobacter pylori* infection (an association probably related to chronically reduced gastric acidity). It is estimated that there is one case of paratyphoid fever for every four cases of typhoid fever, but the incidence of infection associated with *S. paratyphi* A appears to be increasing, especially in India; this increase may be a result of vaccination for *S. typhi*.

Figure 153-1

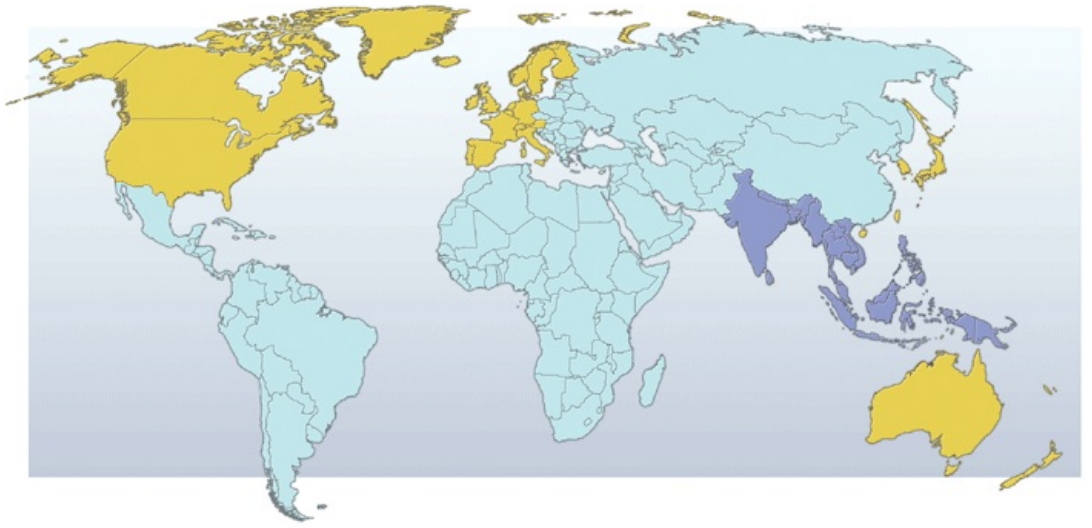

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Annual incidence of typhoid fever per 100,000 population. *(Adapted from Crump JA et al. The global burden of typhoid fever. Bull World Health Organ 82:346, 2004.)*

Multidrug-resistant (MDR) strains of *S. typhi* emerged in 1989 in China and Southeast Asia and have since disseminated widely. These strains contain plasmids encoding resistance to chloramphenicol, ampicillin, and trimethoprim—antibiotics long used to treat enteric fever. With the increased use of fluoroquinolones to treat MDR enteric fever in the 1990s, strains of *S. typhi* and *S. paratyphi* with reduced susceptibility to ciprofloxacin [minimal inhibitory concentration (MIC), 0.125–1 μg/mL] have emerged in the Indian subcontinent, southern Asia, and (most recently) sub-Saharan Africa and have been associated with clinical treatment failure. Testing of isolates for resistance to the first-generation quinolone nalidixic acid detects most but not all strains with reduced susceptibility to ciprofloxacin.

The incidence of enteric fever among U.S. travelers is estimated at 3–30 cases per 100,000. Of 1902 cases of *S. typhi*–associated enteric fever reported to the Centers for Disease Control and Prevention (CDC) in 1999–2006, 79% were associated with recent international travel, most commonly to India (47%), Pakistan (10%), Bangladesh (10%), Mexico (7%), and the Philippines (4%). Only 5% of travelers diagnosed with enteric fever had received *S. typhi* vaccine. Overall, 13% of *S. typhi* isolates in the United States were resistant to ampicillin, chloramphenicol, and trimethoprim-sulfamethoxazole (TMP-SMX), and the proportion of isolates resistant to nalidixic acid increased from 19% in 1999 to 58% in 2006. Infection with nalidixic acid-resistant (NAR) *S. typhi* was associated with travel to the Indian subcontinent. Of the 25–30% of reported cases of enteric fever in the United States that are domestically acquired, the majority are sporadic, but outbreaks linked to contaminated food products and previously unrecognized chronic carriers continue to occur.

Clinical Course

Enteric fever is a misnomer, in that the hallmark features of this disease—fever and abdominal pain—are variable. While fever is documented at presentation in >75% of cases, abdominal pain is reported in only 30–40%. Thus, a high index of suspicion for this potentially fatal systemic illness is necessary when a person presents with fever and a history of recent travel to a developing country.

The incubation period for *S. typhi* averages 10–14 days but ranges from 3–21 days, depending on the inoculum size and the host's health and immune status. The most prominent symptom is prolonged fever (38.8°–40.5°C; 101.8°–104.9°F), which can continue for up to 4 weeks if untreated. *S.* paratyphi A is thought to cause milder disease than *S.* typhi, with predominantly gastrointestinal symptoms. However, a prospective study of 669 consecutive cases of enteric fever in Kathmandu, Nepal, found that the infections were clinically indistinguishable. In this series, symptoms reported on initial medical evaluation included headache (80%), chills (35–45%), cough (30%), sweating (20–25%), myalgias (20%), malaise (10%), and arthralgia (2–4%). Gastrointestinal symptoms included anorexia (55%), abdominal pain (30–40%), nausea (18–24%), vomiting (18%), and diarrhea (22–28%) more commonly than constipation (13–16%). Physical findings included coated tongue (51–56%), splenomegaly (5–6%), and abdominal tenderness (4–5%).

Early physical findings of enteric fever include rash ("rose spots"; 30%), hepatosplenomegaly (3–6%), epistaxis, and relative bradycardia at the peak of high fever (<50%). Rose spots **(Fig. 153-2**; see also Fig. e7-9) make up a faint, salmon-colored, blanching, maculopapular rash located primarily on the trunk and chest. The rash is evident in ~30% of patients at the end of the first week and resolves

without a trace after 2–5 days. Patients can have two or three crops of lesions, and *Salmonella* can be cultured from punch biopsies of these lesions. The faintness of the rash makes it difficult to detect in highly pigmented patients.

Figure 153-2

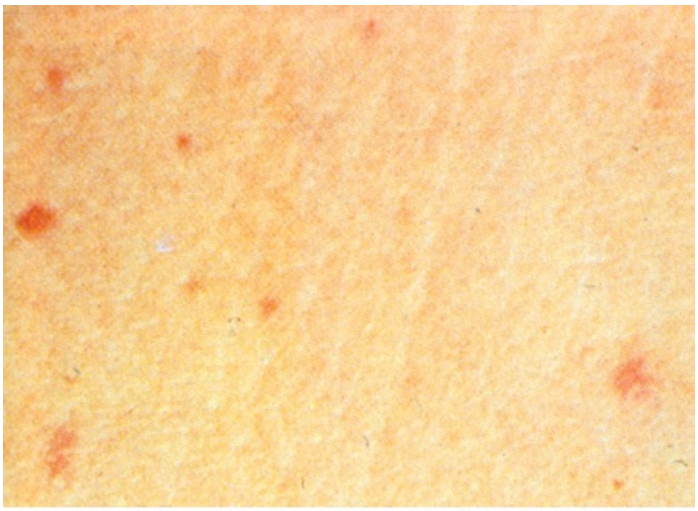

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

"Rose spots," the rash of enteric fever due to *S. typhi* or *S. paratyphi*.

The development of severe disease (which occurs in ~10–15% of patients) depends on host factors (immunosuppression, antacid therapy, previous exposure, and vaccination), strain virulence and inoculum, and choice of antibiotic therapy. Gastrointestinal bleeding (10–20%) and intestinal perforation (1–3%) most commonly occur in the third and fourth weeks of illness and result from hyperplasia, ulceration, and necrosis of the ileocecal Peyer's patches at the initial site of *Salmonella* infiltration. Both complications are life-threatening and require immediate fluid resuscitation and surgical intervention, with broadened antibiotic coverage for polymicrobial peritonitis (Chap. 127) and treatment of gastrointestinal hemorrhages, including bowel resection. Neurologic manifestations occur in 2–40% of patients and include meningitis, Guillain-Barré syndrome, neuritis, and neuropsychiatric symptoms (described as "muttering delirium" or "coma vigil"), with picking at bedclothes or imaginary objects.

Rare complications whose incidences are reduced by prompt antibiotic treatment include disseminated intravascular coagulation, hematophagocytic syndrome, pancreatitis, hepatic and splenic abscesses and granulomas, endocarditis, pericarditis, myocarditis, orchitis, hepatitis, glomerulonephritis, pyelonephritis and hemolytic-uremic syndrome, severe pneumonia, arthritis, osteomyelitis, and parotitis. Up to 10% of patients develop mild relapse, usually within 2–3 weeks of fever resolution and in association with the same strain type and susceptibility profile.

Up to 10% of untreated patients with typhoid fever excrete *S. typhi* in the feces for up to 3 months, and 1–4% develop chronic asymptomatic carriage, shedding *S. typhi* in either urine or stool for >1 year. Chronic carriage is more common among women, infants, and persons who have biliary abnormalities or concurrent bladder infection with *Schistosomahaematobium*. The anatomic abnormalities associated with the latter conditions presumably allow prolonged colonization.

Diagnosis

Since the clinical presentation of enteric fever is relatively nonspecific, the diagnosis needs to be considered in any febrile traveler returning from a developing region, especially the Indian subcontinent, the Philippines, or Latin America. Other diagnoses that should be considered in these travelers include malaria, hepatitis, bacterial enteritis, dengue fever, rickettsial infections, leptospirosis, amebic liver abscesses, and acute HIV infection (Chap. 123). Other than a positive culture, no specific laboratory test is diagnostic for enteric fever. In 15–25% of cases, leukopenia and neutropenia are detectable. Leukocytosis is more common among children, during the first 10 days of illness, and in cases complicated by intestinal perforation or secondary infection. Other nonspecific laboratory findings include moderately elevated liver function tests and muscle enzyme levels.

The definitive diagnosis of enteric fever requires the isolation of *S.* typhi or *S.* paratyphi from blood, bone marrow, other sterile sites, rose spots, stool, or intestinal secretions. The sensitivity of blood culture is only 40–80%, probably because of high rates of antibiotic use in endemic areas and the small quantities of *S. typhi* (i.e., <15 organisms/mL) typically present in the blood. Since almost all *S.*

typhi organisms in blood are associated with the mononuclear-cell/platelet fraction, centrifugation of blood and culture of the buffy coat can substantially reduce the time to isolation of the organism but do not increase sensitivity.

Bone marrow culture is 55–90% sensitive, and, unlike that of blood culture, its yield is not reduced by up to 5 days of prior antibiotic therapy. Culture of intestinal secretions (best obtained by a noninvasive duodenal string test) can be positive despite a negative bone marrow culture. If blood, bone marrow, and intestinal secretions are all cultured, the yield is >90%. Stool cultures, while negative in 60–70% of cases during the first week, can become positive during the third week of infection in untreated patients.

Several serologic tests, including the classic Widal test for "febrile agglutinins," are available. None of these tests is sufficiently sensitive or specific to replace culture-based methods for the diagnosis of enteric fever in developed countries. PCR and DNA probe assays to detect *S. typhi* in blood have been identified but have not yet been developed for clinical use.

Treatment: Enteric (Typhoid) Fever

Prompt administration of appropriate antibiotic therapy prevents severe complications of enteric fever and results in a case-fatality rate of <1%. The initial choice of antibiotics depends on the susceptibility of the *S. typhi* and *S. paratyphi* strains in the area of residence or travel (Table 153-1). For treatment of drug-susceptible typhoid fever, fluoroquinolones are the most effective class of agents, with cure rates of ~98% and relapse and fecal carriage rates of <2%. Experience is most extensive with ciprofloxacin. Short-course ofloxacin therapy is similarly successful against infection caused by nalidixic acid–susceptible strains. However, the increased incidence of NAR *S.* typhi in Asia, which is probably related to the widespread availability of fluoroquinolones over the counter, is now limiting the use of this drug class for empirical therapy. Patients infected with NAR *S.* typhi strains should be treated with ceftriaxone, azithromycin, or high-dose ciprofloxacin. High-dose fluoroquinolone therapy for 7 days for NAR enteric fever has been associated with delayed resolution of fever and high rates of fecal carriage during convalescence. For NAR strains, 10–14 days of high-dose ciprofloxacin is preferred.

Table 153-1 Antibiotic Therapy for Enteric Fever in Adults

Indication	Agent	Dosage (Route)	Duration, Days
Empirical Treatment			
	Ceftriaxone[a]	1–2 g/d (IV)	7–14
	Azithromycin	1 g/d (PO)	5
Fully Susceptible			
	Ciprofloxacin[b] (first line)	500 mg bid (PO) or 400 mg q12h (IV)	5–7
	Amoxicillin (second line)	1 g tid (PO) or 2 g q6h (IV)	14
	Chloramphenicol	25 mg/kg tid (PO or IV)	14–21
	Trimethoprim-sulfamethoxazole	160/800 mg bid (PO)	7–14
Multidrug-Resistant			
	Ciprofloxacin	500 mg bid (PO) or 400 mg q12h (IV)	5–7
	Ceftriaxone	2–3 g/d (IV)	7–14
	Azithromycin	1 g/d (PO)[c]	5
Nalidixic Acid–Resistant			
	Ceftriaxone	2–3 g/d (IV)	7–14
	Azithromycin	1 g/d (PO)	5
	High-dose ciprofloxacin	750 mg bid (PO) or 400 mg q8h (IV)	10–14

[a]Or another third-generation cephalosporin [e.g., cefotaxime, 2 g q8h (IV); or cefixime, 400 mg bid (PO)].

[b]Or ofloxacin, 400 mg bid (PO) for 2–5 days.

[c]Or 1 g on day 1 followed by 500 mg/d PO for 6 days.

Ceftriaxone, cefotaxime, and (oral) cefixime are effective for treatment of MDR enteric fever, including NAR and fluoroquinolone-resistant strains. These agents clear fever in ~1 week, with failure rates of ~5–10%, fecal carriage rates of <3%, and relapse rates of 3–6%. Oral azithromycin results in defervescence in 4–6 days, with rates of relapse and convalescent stool carriage of <3%. Against NAR strains, azithromycin is associated with lower rates of treatment failure and shorter durations of hospitalization than are fluoroquinolones. Despite efficient in vitro killing of *Salmonella*, first- and second-generation cephalosporins as well as aminoglycosides are ineffective in the treatment of clinical infections.

Most patients with uncomplicated enteric fever can be managed at home with oral antibiotics and antipyretics. Patients with persistent vomiting, diarrhea, and/or abdominal distension should be hospitalized and given supportive therapy as well as a parenteral third-generation cephalosporin or fluoroquinolone, depending on the susceptibility profile. Therapy should be administered for at least 10 days or for 5 days after fever resolution.

In a randomized, prospective, double-blind study of critically ill patients with enteric fever (i.e., those with shock and obtundation) in Indonesia in the early 1980s, the administration of dexamethasone (an initial dose of 3 mg/kg followed by eight doses of 1 mg/kg every 6 h) with chloramphenicol was associated with a substantially lower mortality rate than was treatment with chloramphenicol alone (10% vs 55%). Although this study has not been repeated in the "post-chloramphenicol era," severe enteric fever remains one of the few indications for glucocorticoid treatment of an acute bacterial infection.

The 1–5% of patients who develop chronic carriage of *Salmonella* can be treated for 4–6 weeks with an appropriate oral antibiotic. Treatment with oral amoxicillin, TMP-SMX, ciprofloxacin, or norfloxacin is ~80% effective in eradicating chronic carriage of susceptible organisms. However, in cases of anatomic abnormality (e.g., biliary or kidney stones), eradication often requires both antibiotic therapy and surgical correction.

Prevention and Control

Theoretically, it is possible to eliminate the salmonellae that cause enteric fever since they survive only in human hosts and are spread by contaminated food and water. However, given the high prevalence of the disease in developing countries that lack adequate sewage disposal and water treatment, this goal is currently unrealistic. Thus, travelers to developing countries should be advised to monitor their food and water intake carefully and to consider vaccination.

Two typhoid vaccines are commercially available: (1) Ty21a, an oral live attenuated *S. typhi* vaccine (given on days 1, 3, 5, and 7, with a booster every 5 years); and (2) Vi CPS, a parenteral vaccine consisting of purified Vi polysaccharide from the bacterial capsule (given in 1 dose, with a booster every 2 years). The old parenteral whole-cell typhoid/paratyphoid A and B vaccine is no longer licensed, largely because of significant side effects (see below). An acetone-killed whole-cell vaccine is available only for use by the U.S. military. The minimal age for vaccination is 6 years for Ty21a and 2 years for Vi CPS. Currently, there is no licensed vaccine for paratyphoid fever.

A large-scale meta-analysis of vaccine trials comparing whole-cell vaccine, Ty21a, and Vi CPS in populations in endemic areas indicates that, while all three vaccines are similarly effective for the first year, the 3-year cumulative efficacy of the whole-cell vaccine (73%) exceeds that of both Ty21a (51%) and Vi CPS (55%). In addition, the heat-killed whole-cell vaccine maintains its efficacy for 5 years, whereas Ty21a and Vi CPS maintain their efficacy for 4 and 2 years, respectively. However, the whole-cell vaccine is associated with a much higher incidence of side effects (especially fever: 16% vs 1–2%) than the other two vaccines.

Vi CPS typhoid vaccine is poorly immunogenic in children <5 years of age because of T cell–independent properties. In the recently developed Vi-rEPA vaccine, Vi is bound to a nontoxic recombinant protein that is identical to *Pseudomonas aeruginosa* exotoxin A. In 2- to 4-year-olds, two injections of Vi-rEPA induced higher T cell responses and higher levels of serum IgG antibody to Vi than did Vi CPS in 5- to 14-year-olds. In a two-dose trial in 2- to 5-year-old children in Vietnam, Vi-rEPA provided 91% efficacy at 27 months and 88% efficacy at 43 months and was very well tolerated. This vaccine is not yet commercially available in the United States. At least three new live vaccines are in clinical development and may prove more efficacious and longer-lasting than previous live vaccines.

Typhoid vaccine is not required for international travel, but it is recommended for travelers to areas where there is a moderate to high risk of exposure to *S. typhi*, especially those who are traveling to southern Asia and other developing regions of Asia, Africa, the Caribbean, and Central and South America and who will be exposed to potentially contaminated food and drink. Typhoid vaccine should be considered even for persons planning <2 weeks of travel to high-risk areas. In addition, laboratory workers who deal with *S. typhi* and household contacts of known *S. typhi* carriers should be vaccinated. Because the protective efficacy of vaccine can be overcome by the high inocula that are commonly encountered in food-borne exposures, immunization is an adjunct and not a substitute for avoiding high-risk foods and beverages. Immunization is not recommended for adults residing in typhoid-endemic areas or for the management of persons who may have been exposed in a common-

source outbreak.

Enteric fever is a notifiable disease in the United States. Individual health departments have their own guidelines for allowing ill or colonized food handlers or health care workers to return to their jobs. The reporting system enables public health departments to identify potential source patients and to treat chronic carriers in order to prevent further outbreaks. In addition, since 1–4% of patients with *S. typhi* infection become chronic carriers, it is important to monitor patients (especially child-care providers and food handlers) for chronic carriage and to treat this condition if indicated.

NONTYPHOIDAL SALMONELLOSIS

Epidemiology

In the United States, the incidence of NTS infection has doubled in the past 2 decades; the 2009 figure is ~14 million cases annually. In 2007, the incidence of NTS infection in this country was 14.9 per 100,000 persons—the highest rate among the 11 food-borne enteric pathogens under active surveillance. Five serotypes accounted for one-half of U.S. infections in 2007: *typhimurium* (19%), *enteritidis* (14%), Newport (9%), Javiana (5%), and Heidelberg (4%).

The incidence of nontyphoidal salmonellosis is highest during the rainy season in tropical climates and during the warmer months in temperate climates, coinciding with the peak in food-borne outbreaks. Rates of morbidity and mortality associated with NTS are highest among the elderly, infants, and immunocompromised individuals, including those with hemoglobinopathies, HIV infection, or infections that cause blockade of the reticuloendothelial system (e.g., bartonellosis, malaria, schistosomiasis, and histoplasmosis).

Unlike *S. typhi* and *S. paratyphi*, whose only reservoir is humans, NTS can be acquired from multiple animal reservoirs. Transmission is most commonly associated with animal food products, especially eggs, poultry, undercooked ground meat, dairy products, and fresh produce contaminated with animal waste.

S. enteritidis infection associated with chicken eggs emerged as a major cause of food-borne disease during the 1980s and 1990s. *S. enteritidis* infection of the ovaries and upper oviduct tissue of hens results in contamination of egg contents before shell deposition. Infection is spread to egg-laying hens from breeding flocks and through contact with rodents and manure. Of the 997 outbreaks of *S. enteritidis* with a confirmed source that were reported to the CDC in 1985–2003, 75% were associated with raw or undercooked eggs. After peaking at 3.9 cases per 100,000 U.S. population in 1995, the incidence of *S. enteritidis* infection declined substantially to 1.7 per 100,000 in 2003; this decrease probably reflected improved on-farm control measures, refrigeration, and education of consumers and food-service workers. Transmission via contaminated eggs can be prevented by cooking eggs until the yolk is solidified and through pasteurization of egg products.

Centralization of food processing and widespread food distribution have contributed to the increased incidence of NTS in developed countries. Manufactured foods to which recent *Salmonella* outbreaks have been traced include peanut butter; milk products, including infant formula; and various processed foods, including packaged breakfast cereal, salsa, frozen prepared meals, and snack foods. Large outbreaks have also been linked to fresh produce, including alfalfa sprouts, cantaloupe, fresh-squeezed orange juice, and tomatoes; these items become contaminated by manure or water at a single site and then are widely distributed.

An estimated 6% of sporadic *Salmonella* infections in the United States are attributed to contact with reptiles and amphibians, especially iguanas, snakes, turtles, and lizards. Reptile-associated *Salmonella* infection more commonly leads to hospitalization and more frequently involves infants than do other *Salmonella* infections. Other pets, including African hedgehogs, snakes, birds, rodents, baby chicks, ducklings, dogs, and cats, are also potential sources of NTS.

Increasing antibiotic resistance in NTS species is a global problem and has been linked to the widespread use of antimicrobial agents in food animals and especially in animal feed. In the early 1990s, *S. typhimurium* definitive phage type 104 (DT104), characterized by resistance to ≥5 antibiotics (ampicillin, chloramphenicol, streptomycin, sulfonamides, and tetracyclines; R-type ACSSuT), emerged worldwide. In 2005, resistance to at least ACSSuT was the most common MDR phenotype among NTS isolates in the United States. Acquisition is associated with exposure to ill farm animals and to various meat products, including uncooked or undercooked ground beef. Although probably no more virulent than susceptible *S. typhimurium* strains, DT104 strains are associated with an increased risk of bloodstream infection and hospitalization. NAR and trimethoprim-resistant DT104 strains are emerging, especially in the United Kingdom.

Because of increased resistance to conventional antibiotics such as ampicillin and TMP-SMX, extended-spectrum cephalosporins and fluoroquinolones have emerged as the agents of choice for the treatment of MDR NTS infections. In 2005, 2% of all NTS strains and 12.6% of *S.* Newport strains were resistant to ceftriaxone. Most ceftriaxone-resistant isolates were from children <18 years of age, in whom ceftriaxone is the antibiotic of choice for treatment of invasive NTS infection. These strains

contained plasmid-encoded AmpC β-lactamases that were probably acquired by horizontal genetic transfer from *Escherichia coli* strains in food-producing animals—an event linked to the widespread use of the veterinary cephalosporin ceftiofur.

Resistance to nalidixic acid and fluoroquinolones also has begun to emerge and is most commonly associated with point mutations in the DNA gyrase genes *gyr*A and *gyr*B. Nalidixic acid resistance is a good predictor of reduced susceptibility to clinically useful fluoroquinolones. From 1996–2005, the rate of NAR NTS isolates in the United States increased fivefold (from 0.5–2.4%). In Denmark, infection with NAR *S. typhimurium* DT104 has been linked to swine and associated with a threefold higher risk of invasive disease or death within 90 days. In Taiwan in 2000, a strain of ciprofloxacin-resistant (MIC, ≥4 mcg/mL) *S. choleraesuis* caused a large outbreak of invasive infections that was linked to the use of enrofloxacin in swine feed.

Clinical Manifestations

GASTROENTERITIS

Infection with NTS most often results in gastroenteritis indistinguishable from that caused by other enteric pathogens. Nausea, vomiting, and diarrhea occur 6–48 h after the ingestion of contaminated food or water. Patients often experience abdominal cramping and fever (38–39°C; 100.5–102.2°F). Diarrheal stools are usually loose, nonbloody, and of moderate volume. However, large-volume watery stools, bloody stools, or symptoms of dysentery may occur. Rarely, NTS causes pseudoappendicitis or an illness that mimics inflammatory bowel disease.

Gastroenteritis caused by NTS is usually self-limited. Diarrhea resolves within 3–7 days and fever within 72 h. Stool cultures remain positive for 4–5 weeks after infection and—in rare cases of chronic carriage (<1%)—for >1 year. Antibiotic treatment usually is not recommended and may prolong fecal carriage. Neonates, the elderly, and immunosuppressed patients (e.g., transplant recipients, HIV-infected persons) with NTS gastroenteritis are especially susceptible to dehydration and dissemination and may require hospitalization and antibiotic therapy. Acute NTS gastroenteritis was associated with a threefold increased risk of dyspepsia and irritable bowel syndrome at 1 year in a recent study from Spain.

BACTEREMIA AND ENDOVASCULAR INFECTIONS

Up to 8% of patients with NTS gastroenteritis develop bacteremia; of these, 5–10% develop localized infections. Bacteremia and metastatic infection are most common with *S.* choleraesuis and *S.* Dublin and among infants, the elderly, and immunocompromised patients. NTS endovascular infection should be suspected in high-grade or persistent bacteremia, especially with preexisting valvular heart disease, atherosclerotic vascular disease, prosthetic vascular graft, or aortic aneurysm. Arteritis should be suspected in elderly patients with prolonged fever and back, chest, or abdominal pain developing after an episode of gastroenteritis. Endocarditis and arteritis are rare (<1% of cases) but are associated with potentially fatal complications, including valve perforation, endomyocardial abscess, infected mural thrombus, pericarditis, mycotic aneurysms, aneurysm rupture, aortoenteric fistula, and vertebral osteomyelitis. In some areas of sub-Saharan Africa, NTS may be among the most common causes—or even the most common cause—of bacteremia in children. NTS bacteremia among these children is not associated with diarrhea and has been associated with nutritional status and HIV infection.

LOCALIZED INFECTIONS

Intraabdominal Infections

Intraabdominal infections due to NTS are rare and usually manifest as hepatic or splenic abscesses or as cholecystitis. Risk factors include hepatobiliary anatomic abnormalities (e.g., gallstones), abdominal malignancy, and sickle cell disease (especially with splenic abscesses). Eradication of the infection often requires surgical correction of abnormalities and percutaneous drainage of abscesses.

Central Nervous System Infections

NTS meningitis most commonly develops in infants 1–4 months of age. It often results in severe sequelae (including seizures, hydrocephalus, brain infarction, and mental retardation) with death in up to 60% of cases. Other rare central nervous system infections include ventriculitis, subdural empyema, and brain abscesses.

Pulmonary Infections

NTS pulmonary infections usually present as lobar pneumonia, and complications include lung abscess, empyema, and bronchopleural fistula formation. The majority of cases occur in patients with lung cancer, structural lung disease, sickle cell disease, or glucocorticoid use.

Urinary and Genital Tract Infections

Urinary tract infections caused by NTS present as either cystitis or pyelonephritis. Risk factors include malignancy, urolithiasis, structural abnormalities, HIV infection, and renal transplantation. NTS genital infections are rare and include ovarian and testicular abscesses, prostatitis, and epididymitis. Like other focal infections, both genital and urinary tract infections can be complicated by abscess

formation.

Bone, Joint, and Soft Tissue Infections

Salmonella osteomyelitis most commonly affects the femur, tibia, humerus, or lumbar vertebrae and is most often seen in association with sickle cell disease, hemoglobinopathies, or preexisting bone disease (e.g., fractures). Prolonged antibiotic treatment is recommended to decrease the risk of relapse and chronic osteomyelitis. Septic arthritis occurs in the same patient population as osteomyelitis and usually involves the knee, hip, or shoulder joints. Reactive arthritis (Reiter's syndrome) can follow NTS gastroenteritis and is seen most frequently in persons with the HLA-B27 histocompatibility antigen. NTS rarely can cause soft tissue infections, usually at sites of local trauma in immunosuppressed patients.

Diagnosis

The diagnosis of NTS infection is based on isolation of the organism from freshly passed stool or from blood or another ordinarily sterile body fluid. All salmonellae isolated in clinical laboratories should be sent to local public health departments for serotyping. Blood cultures should be done whenever a patient has prolonged or recurrent fever. Endovascular infection should be suspected if there is high-grade bacteremia (>50% of three or more positive blood cultures). Echocardiography, CT, and indium-labeled white cell scanning are used to identify localized infection. When another localized infection is suspected, joint fluid, abscess drainage, or cerebrospinal fluid should be cultured, as clinically indicated.

Treatment: Nontyphoidal Salmonellosis

Antibiotics should not be used routinely to treat uncomplicated NTS gastroenteritis. The symptoms are usually self-limited, and the duration of fever and diarrhea is not significantly decreased by antibiotic therapy. In addition, antibiotic treatment has been associated with increased rates of relapse, prolonged gastrointestinal carriage, and adverse drug reactions. Dehydration secondary to diarrhea should be treated with fluid and electrolyte replacement.

Preemptive antibiotic treatment (Table 153-2) should be considered for patients at increased risk for invasive NTS infection, including neonates (probably up to 3 months of age); persons >50 years of age with suspected atherosclerosis; and patients with immunosuppression, cardiac valvular or endovascular abnormalities, or significant joint disease. Treatment should consist of an oral or IV antibiotic administered for 48–72 h or until the patient becomes afebrile. Immunocompromised persons may require up to 7–14 days of therapy. The <1% of persons who develop chronic carriage of NTS should receive a prolonged antibiotic course, as described above for chronic carriage of *S. typhi*.

Table 153-2 Antibiotic Therapy for Nontyphoidal *Salmonella* Infection in Adults

Indication	Agent	Dosage (Route)	Duration, Days
Preemptive Treatment[a]			
	Ciprofloxacin[b]	500 mg bid (PO)	2–3
Severe Gastroenteritis[c]			
	Ciprofloxacin	500 mg bid (PO) or 400 mg q12h (IV)	3–7
	Trimethoprim-sulfamethoxazole	160/800 mg bid (PO)	
	Amoxicillin	1 g tid (PO)	
	Ceftriaxone	1–2 g/d (IV)	
Bacteremia			
	Ceftriaxone[d]	2 g/d (IV)	7–14
	Ciprofloxacin	400 mg q12h (IV), then 500 mg bid (PO)	
Endocarditis or Arteritis			
	Ceftriaxone	2 g/d (IV)	42
	Ciprofloxacin	400 mg q8h (IV), then 750 mg bid (PO)	
	Ampicillin	2 g q4h (IV)	
Meningitis			

	Ceftriaxone	2 g q12 h (IV)	14–21
	Ampicillin	2 g q4h (IV)	
Other Localized Infection			
	Ceftriaxone	2 g/d (IV)	14–28
	Ciprofloxacin	500 mg bid (PO) or 400 mg q12h (IV)	
	Ampicillin	2 g q6h (IV)	

[a]Consider for neonates; persons >50 years of age with possible atherosclerotic vascular disease; and patients with immunosuppression, endovascular graft, or joint prosthesis.

[b]Or ofloxacin, 400 mg bid (PO).

[c]Consider on an individualized basis for patients with severe diarrhea and high fever who require hospitalization.

[d]Or cefotaxime, 2 g q8h (IV).

Because of the increasing prevalence of antibiotic resistance, empirical therapy for life-threatening NTS bacteremia or focal NTS infection should include a third-generation cephalosporin or a fluoroquinolone (Table 153-2). If the bacteremia is low-grade (<50% of positive blood cultures), the patient should be treated for 7–14 days. Patients with HIV/AIDS and NTS bacteremia should receive 1–2 weeks of IV antibiotic therapy followed by 4 weeks of oral therapy with a fluoroquinolone. Patients whose infections relapse after this regimen should receive long-term suppressive therapy with a fluoroquinolone or TMP-SMX, as indicated by bacterial sensitivities.

If the patient has endocarditis or arteritis, treatment for 6 weeks with an IV β-lactam antibiotic (such as ceftriaxone or ampicillin) is indicated. IV ciprofloxacin followed by prolonged oral therapy is an option, but published experience is limited. Early surgical resection of infected aneurysms or other infected endovascular sites is recommended. Patients with infected prosthetic vascular grafts that cannot be resected have been maintained successfully on chronic suppressive oral therapy. For extraintestinal nonvascular infections, a 2- to 4-week course of antibiotic therapy (depending on the infection site) is usually recommended. In chronic osteomyelitis, abscess, or urinary or hepatobiliary infection associated with anatomic abnormalities, surgical resection or drainage may be required in addition to prolonged antibiotic therapy for eradication of infection.

Prevention and Control

Despite widespread efforts to prevent or reduce bacterial contamination of animal-derived food products and to improve food-safety education and training, recent declines in the incidence of NTS in the United States have been modest compared with those of other food-borne pathogens. This observation probably reflects the complex epidemiology of NTS. Identifying effective risk-reduction strategies requires monitoring of every step of food production, from handling of raw animal or plant products to preparation of finished foods. Contaminated food can be made safe for consumption by pasteurization, irradiation, or proper cooking. All cases of NTS infection should be reported to local public health departments, since tracking and monitoring of these cases can identify the source(s) of infection and help authorities anticipate large outbreaks. Lastly, the prudent use of antimicrobial agents in both humans and animals is needed to limit the emergence of MDR *Salmonella*.

FURTHER READINGS

Cohen JI et al: Extra-intestinal manifestations of *Salmonella* infections. Medicine 66:349, 1987[PMID: 3306260]

Glynn MK et al: Emergence of multidrug-resistant *Salmonella enterica* serotype *typhimurium* DT104 infections in the United States. N Engl J Med 338:1333, 1998[PMID: 9571252]

Haraga A et al: *Salmonella* interplay with host cells. Nat Rev Micobiol 6:53, 2008[PMID: 18026123]

Lin FY et al: The efficacy of a *Salmonella typhi* Vi conjugate vaccine in two- to five-year-old children. N Engl J Med 344:1263, 2001[PMID: 11320385]

Lynch MF et al: Typhoid fever in the United States, 1990–2006. JAMA 302:859, 2009[PMID: 19706859]

Maskey AP et al: *Salmonella* enteric serovar Paratyphi A and *S. enterica* serovar Typhi cause indistinguishable clinical syndromes in Kathmandu, Nepal. Clin Infect Dis 42:1247, 2006[PMID: 16586383]

Steinberg EB et al: Typhoid fever in travelers: Who should be targeted for prevention? Clin Infect Dis 39:186, 2004[PMID: 15307027]

Su LH et al: Antimicrobial resistance in nontyphoid *Salmonella* serotypes: A global challenge. Clin Infect Dis 39:546, 2004[PMID: 15356819]

Thaver D et al: Fluoroquinolones for treating typhoid and paratyphoid fever (enteric fever). Cochrane Database Syst Rev CD004530, 2008

Varma JK et al: Antimicrobial-resistant nontyphoidal *Salmonella* is associated with excess bloodstream infections and hospitalizations. J Infect Dis 191:554, 2005[PMID: 15655779]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SHIGELLOSIS: INTRODUCTION

The discovery of *Shigella* as the etiologic agent of dysentery—a clinical syndrome of fever, intestinal cramps, and frequent passage of small, bloody, mucopurulent stools—is attributed to the Japanese microbiologist Kiyoshi Shiga, who isolated the Shiga bacillus (now known as *Shigella dysenteriae* type 1) from patients' stools in 1897 during a large and devastating dysenteryepidemic. *Shigella* cannot be distinguished from *Escherichia coli* by DNA hybridization and remains a separate species only on historical and clinical grounds.

Definition

Shigella is a nonspore-forming, gram-negative bacterium that, unlike *E. coli*, is nonmotile and does not produce gas from sugars, decarboxylate lysine, or hydrolyze arginine. Some serovars produce indole, and occasional strains utilize sodium acetate. *S.dysenteriae*, *S. flexneri*, *S.boydii*, and *S. sonnei* (serogroups A, B, C, and D, respectively) can be differentiated on the basis of biochemical and serologic characteristics. Genome sequencing of *E. coli* K12, *S. flexneri* 2a, *S. sonnei*, *S. dysenteriae* type 1, and *S. boydii* has revealed that these species have ~93% of genes in common. The three major genomic "signatures" of *Shigella* are (1) a 215-kb virulence plasmid that carries most of the genes required for pathogenicity (particularly invasive capacity); (2) the lack or alteration of genetic sequences encoding products (e.g., lysine decarboxylase) that, if expressed, would attenuate pathogenicity; and (3) in *S. dysenteriae* type 1, the presence of genes encoding Shiga toxin, a potent cytotoxin.

Epidemiology

The human intestinal tract represents the major reservoir of *Shigella*, which is also found (albeit rarely) in the higher primates. Because excretion of shigellae is greatest in the acute phase of disease, the bacteria are transmitted most efficiently by the fecal-oral route via hand carriage; however, some outbreaks reflect food-borne or waterborne transmission. In impoverished areas, *Shigella* can be transmitted by flies. The high-level infectivity of *Shigella* is reflected by the very small inoculum required for experimental infection of volunteers [100 colony-forming units (CFU)], by the very high attack rates during outbreaks in day-care centers (33–73%), and by the high rates of secondary cases among family members of sick children (26–33%). Shigellosis can also be transmitted sexually.

Throughout history, *Shigella* epidemics have often occurred in settings of human crowding under conditions of poor hygiene—e.g., among soldiers in campaigning armies, inhabitants of besieged cities, groups on pilgrimages, and refugees in camps. Epidemics follow a cyclical pattern in areas such as the Indian subcontinent and sub-Saharan Africa. These devastating epidemics, which are most often caused by *S. dysenteriae* type 1, are characterized by high attack and mortality rates. In Bangladesh, for instance, an epidemic caused by *S. dysenteriae* type 1 was associated with a 42% increase in mortality rate among children 1–4 years of age. Apart from these epidemics, shigellosis is mostly an endemic disease, with 99% of cases occurring in the developing world and the highest prevalences in the most impoverished areas, where personal and general hygiene is below standard. *S. flexneri* isolates predominate in the least developed areas, whereas *S. sonnei* is more prevalent in economically emerging countries and in the industrialized world.

PREVALENCE IN THE DEVELOPING WORLD

In a review published under the auspices of the World Health Organization (WHO), the total annual number of cases in 1966–1997 was estimated at 165 million, and 69% of these cases occurred in children <5 years of age. In this review, the annual number of deaths was calculated to range between 500,000 and 1.1 million. More recent data (2000–2004) from six Asian countries indicate that even though the incidence of shigellosis remains stable, mortality rates associated with this disease may have decreased significantly, possibly as a result of improved nutritional status. However, extensive and essentially uncontrolled use of antibiotics, which may also account for declining mortality rates, has increased the rate of emergence of multidrug-resistant *Shigella* strains. An often-overlooked complication of shigellosis is the short- and long-term impairment of the nutritional status of infected children in endemic areas. Combined with anorexia, the exudative enteropathy resulting from mucosal abrasions contributes to rapid deterioration of the patient's nutritional status. Shigellosis is thus a major contributor to stunted growth among children in developing countries.

Peaking in incidence in the pediatric population, endemic shigellosis is rare in young and middle-aged

adults, probably because of naturally acquired immunity. Incidence then increases again in the elderly population.

PREVALENCE IN THE INDUSTRIALIZED WORLD

In pediatric populations, local outbreaks occur when proper and adapted hygiene policies are not implemented in group facilities like day-care centers and institutions for the mentally retarded. In adults, as in children, sporadic cases occur among travelers returning from endemic areas, and rare outbreaks of varying size can follow waterborne or food-borne infections.

Pathogenesis and Pathology

Shigella infection occurs essentially through oral contamination via direct fecal-oral transmission, the organism being poorly adapted to survive in the environment. Resistance to low-pH conditions allows shigellae to survive passage through the gastric barrier, an ability that may explain in part why a small inoculum (as few as 100 CFU) is sufficient to cause infection.

The watery diarrhea that usually precedes the dysenteric syndrome is attributable to active secretion and abnormal water reabsorption—a secretory effect at the jejunal level described in experimentally infected rhesus monkeys. This initial purge is probably due to the combined action of an enterotoxin (ShET-1) and mucosal inflammation. The dysenteric syndrome, manifested by bloody and mucopurulent stools, reflects invasion of the mucosa.

The pathogenesis of *Shigella* is essentially determined by a large virulence plasmid of 214 kb comprising ~100 genes, of which 25 encode a type III secretion system that inserts into the membrane of the host cell to allow effectors to transit from the bacterial cytoplasm to the host cell cytoplasm **(Fig. 154-1)**. Bacteria are thereby able to invade intestinal epithelial cells by inducing their own uptake after the initial crossing of the epithelial barrier through M cells (the specialized translocating epithelial cells in the follicle-associated epithelium that covers mucosal lymphoid nodules). The organisms induce apoptosis of subepithelial resident macrophages. Once inside the cytoplasm of intestinal epithelial cells, *Shigella* effectors trigger the cytoskeletal rearrangements necessary to direct uptake of the organism into the epithelial cell. The *Shigella*-containing vacuole is then quickly lysed, releasing bacteria into the cytosol.

Figure 154-1

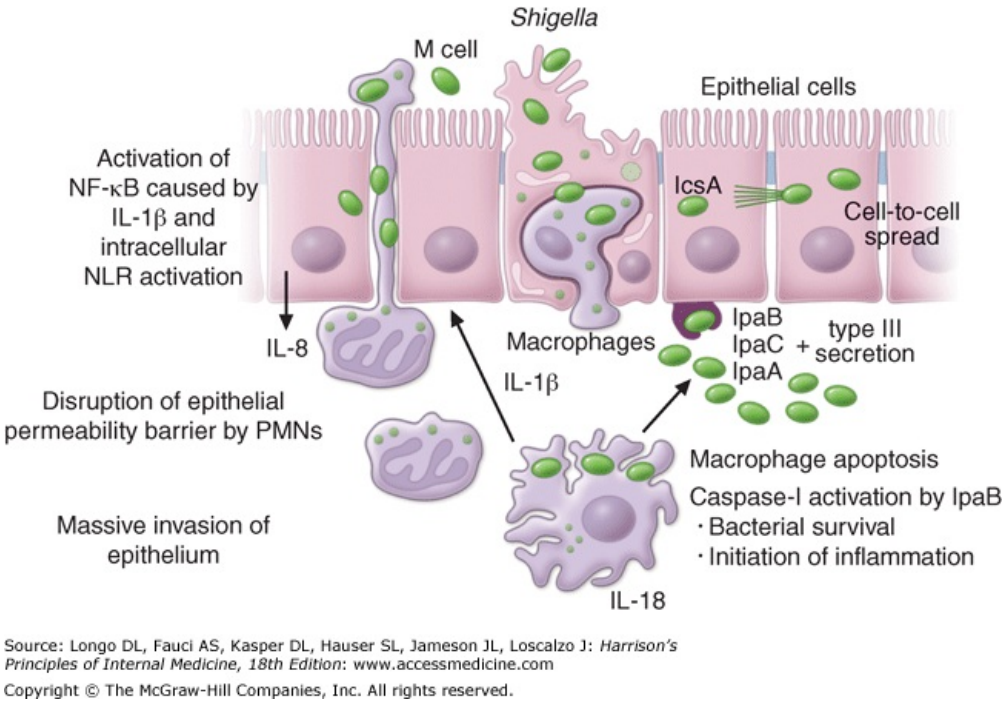

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Invasive strategy of *Shigella flexneri.* IL, interleukin; NF-κB, nuclear factor κB; NLR, NOD-like receptor; PMN, polymorphonuclear leukocyte.

Intracellular shigellae next use cytoskeletal components to propel themselves inside the infected cell; when the moving organism and the host cell membrane come into contact, cellular protrusions form and are engulfed by neighboring cells. This series of events permits bacterial cell-to-cell spread.

Cytokines released by a growing number of infected intestinal epithelial cells attract increased numbers of immune cells [particularly polymorphonuclear leukocytes (PMNs)] to the infected site, thus further destabilizing the epithelial barrier, exacerbating inflammation, and leading to the acute colitis that characterizes shigellosis. Evidence indicates that some type III secretion system–injected effectors

can control the extent of inflammation, thus facilitating bacterial survival.

Shiga toxin produced by *S. dysenteriae* type 1 increases disease severity. This toxin belongs to a group of A1-B5 protein toxins whose B subunit binds to the receptor globotriaosylceramide on the target cell surface and whose catalytic A subunit is internalized by receptor-mediated endocytosis and interacts with the subcellular machinery to inhibit protein synthesis by expressing RNA N-glycosidase activity on 28S ribosomal RNA. This process leads to inhibition of binding of the amino-acyl-tRNA to the 60S ribosomal subunit and thus to a general shutoff of cell protein biosynthesis. Shiga toxins are translocated from the bowel into the circulation. After binding of the toxins to target cells in the kidney, pathophysiologic alterations may result in hemolytic-uremic syndrome (HUS; see below).

Clinical Manifestations

The presentation and severity of shigellosis depend to some extent on the infecting serotype but even more on the age and the immunologic and nutritional status of the host. Poverty and poor standards of hygiene are strongly related to the number and severity of diarrheal episodes, especially in children <5 years old who have been weaned.

Shigellosis typically evolves through four phases: incubation, watery diarrhea, dysentery, and the postinfectious phase. The incubation period usually lasts 1–4 days but may be as long as 8 days. Typical initial manifestations are transient fever, limited watery diarrhea, malaise, and anorexia. Signs and symptoms may range from mild abdominal discomfort to severe cramps, diarrhea, fever, vomiting, and tenesmus. The manifestations are usually exacerbated in children, with temperatures up to 40°–41°C (104.0°–105.8°F) and more severe anorexia and watery diarrhea. This initial phase may represent the only clinical manifestation of shigellosis, especially in developed countries. Otherwise, dysentery follows within hours or days and is characterized by uninterrupted excretion of small volumes of bloody mucopurulent stools with increased tenesmus and abdominal cramps. At this stage, *Shigella* produces acute colitis involving mainly the distal colon and the rectum. Unlike most diarrheal syndromes, dysenteric syndromes rarely present with dehydration as a major feature. Endoscopy shows an edematous and hemorrhagic mucosa, with ulcerations and possibly overlying exudates resembling pseudomembranes. The extent of the lesions correlates with the number and frequency of stools and with the degree of protein loss by exudative mechanisms. Most episodes are self-limited and resolve without treatment in 1 week. With appropriate treatment, recovery takes place within a few days to a week, with no sequelae.

Acute life-threatening complications are seen most often in children <5 years of age (particularly those who are malnourished) and in elderly patients. Risk factors for death in a clinically severe case include nonbloody diarrhea, moderate to severe dehydration, bacteremia, absence of fever, abdominal tenderness, and rectal prolapse. Major complications are predominantly intestinal (e.g., toxic megacolon, intestinal perforations, rectal prolapse) or metabolic (e.g., hypoglycemia, hyponatremia, dehydration). Bacteremia is rare and is reported most frequently in severely malnourished and HIV-infected patients. Alterations of consciousness, including seizures, delirium, and coma, may occur, especially in children <5 years old, and are associated with a poor prognosis; fever and severe metabolic alterations are more often the major causes of altered consciousness than is meningitis or the Ekiri syndrome (toxic encephalopathy associated with bizarre posturing, cerebral edema, and fatty degeneration of viscera), which has been reported mostly in Japanese children. Pneumonia, vaginitis, and keratoconjunctivitis due to *Shigella* are rarely reported. In the absence of serious malnutrition, severe and very unusual clinical manifestations, such as meningitis, may be linked to genetic defects in innate immune functions [i.e., deficiency in interleukin 1 receptor–associated kinase 4 (IRAK-4)] and may require genetic investigation.

Two complications of particular importance are toxic megacolon and HUS. Toxic megacolon is a consequence of severe inflammation extending to the colonic smooth-muscle layer and causing paralysis and dilatation. The patient presents with abdominal distention and tenderness, with or without signs of localized or generalized peritonitis. The abdominal x-ray characteristically shows marked dilatation of the transverse colon (with the greatest distention in the ascending and descending segments); thumbprinting caused by mucosal inflammatory edema; and loss of the normal haustral pattern associated with pseudopolyps, often extending into the lumen. Pneumatosis coli is an occasional finding. If perforation occurs, radiographic signs of pneumoperitoneum may be apparent. Predisposing factors (e.g., hypokalemia and use of opioids, anticholinergics, loperamide, psyllium seeds, and antidepressants) should be investigated.

Shiga toxin produced by *S. dysenteriae* type 1 has been linked to HUS in developing countries but rarely in industrialized countries, where enterohemorrhagic *E. coli* (EHEC) predominates as the etiologic agent of this syndrome. HUS is an early complication that most often develops after several days of diarrhea. Clinical examination shows pallor, asthenia, and irritability and, in some cases, bleeding of the nose and gums, oliguria, and increasing edema. HUS is a nonimmune (Coombs test–negative) hemolytic anemia defined by a diagnostic triad: microangiopathic hemolytic anemia [hemoglobin level typically <80 g/L (<8 g/dL)], thrombocytopenia (mild to moderate in severity; typically <60,000 platelets/μL), and acute renal failure due to thrombosis of the glomerular capillaries (with markedly elevated creatinine levels). Anemia is severe, with fragmented red blood cells (schizocytes) in

the peripheral smear, high serum concentrations of lactate dehydrogenase and free circulating hemoglobin, and elevated reticulocyte counts. Acute renal failure occurs in 55–70% of cases; however, renal function recovers in most of these cases (up to 70% in various series). Leukemoid reactions, with leukocyte counts of 50,000/μL, are sometimes noted in association with HUS.

The postinfectious immunologic complication known as reactive arthritis can develop weeks or months after shigellosis, especially in patients expressing the histocompatibility antigen HLA-B27. About 3% of patients infected with *S. flexneri* later develop this syndrome, with arthritis, ocular inflammation, and urethritis—a condition that can last for months or years and can progress to difficult-to-treat chronic arthritis. Postinfectious arthropathy occurs only after infection with *S. flexneri* and not after infection with the other *Shigella* serotypes.

Laboratory Diagnosis

The differential diagnosis in patients with a dysenteric syndrome depends on the clinical and environmental context. In developing areas, infectious diarrhea caused by other invasive pathogenic bacteria (*Salmonella*, *Campylobacter jejuni* , *Clostridium difficile*, *Yersinia enterocolitica*) or parasites (*Entamoeba histolytica*) should be considered. Only bacteriologic and parasitologic examinations of stool can truly differentiate among these pathogens. A first flare of inflammatory bowel disease, such as Crohn's disease or ulcerative colitis (Chap. 295), should be considered in patients in industrialized countries. Despite similar symptoms, anamnesis discriminates between shigellosis, which usually follows recent travel in an endemic zone, and these other conditions.

Microscopic examination of stool smears shows the presence of erythrophagocytic trophozoites with very few PMNs in *E. histolytica* infection, whereas bacterial enteroinvasive infections (particularly shigellosis) are characterized by high PMN counts in each microscopic field. However, because shigellosis often manifests only as watery diarrhea, systematic attempts to isolate *Shigella* are necessary.

The "gold standard" for the diagnosis of *Shigella* infection remains the isolation and identification of the pathogen from fecal material. One major difficulty, particularly in endemic areas where laboratory facilities are not immediately available, is the fragility of *Shigella* and its common disappearance during transport, especially with rapid changes in temperature and pH. In the absence of a reliable enrichment medium, buffered glycerol saline or Cary-Blair medium can be used as a holding medium, but prompt inoculation onto isolation medium is essential. The probability of isolation is higher if the portion of stools that contains bloody and/or mucopurulent material is directly sampled. Rectal swabs can be used, as they offer the highest rate of successful isolation during the acute phase of disease. Blood cultures are positive in <5% of cases but should be done when a patient presents with a clinical picture of severe sepsis.

In addition to quick processing, the use of several media increases the likelihood of successful isolation: a nonselective medium such as bromocresol-purple agar lactose; a low-selectivity medium such as MacConkey or eosin-methylene blue; and a high-selectivity medium such as Hektoen, *Salmonella-Shigella*, or xylose-lysine-deoxycholate agar. After incubation on these media for 12–18 h at 37°C (98.6°F), shigellae appear as nonlactose-fermenting colonies that measure 0.5–1 mm in diameter and have a convex, translucent, smooth surface. Suspected colonies on nonselective or low-selectivity medium can be subcultured on a high-selectivity medium before being specifically identified or can be identified directly by standard commercial systems on the basis of four major characteristics: glucose positivity (usually without production of gas), lactose negativity, H_2S negativity, and lack of motility. The four *Shigella* serogroups (A–D) can then be differentiated by additional characteristics. This approach adds time and difficulty to the identification process; however, after presumptive diagnosis, the use of serologic methods (e.g., slide agglutination, with group- and then type-specific antisera) should be considered. Group-specific antisera are widely available; in contrast, because of the large number of serotypes and sub-serotypes, type-specific antisera are rare and more expensive and thus are often restricted to reference laboratories.

Treatment: Shigellosis

ANTIBIOTIC SUSCEPTIBILITY OF *SHIGELLA*

As an enteroinvasive disease, shigellosis requires antibiotic treatment. Since the mid-1960s, however, increasing resistance to multiple drugs has been a dominant factor in treatment decisions. Resistance rates are highly dependent on the geographic area. Clonal spread of particular strains and horizontal transfer of resistance determinants, particularly via plasmids and transposons, contribute to multidrug resistance. The current global status—i.e., high rates of resistance to classic first-line antibiotics such as amoxicillin—has led to a rapid switch to quinolones such as nalidixic acid. However, resistance to such early-generation quinolones has also emerged and spread quickly as a result of chromosomal mutations affecting DNA gyrase and topoisomerase IV; this resistance has necessitated the use of later-generation quinolones as first-line antibiotics in many areas. For instance, a review of the antibiotic resistance history of *Shigella* in India found that, after their introduction in the late 1980s, the second-generation quinolones norfloxacin, ciprofloxacin, and ofloxacin were highly effective in the

treatment of shigellosis, including cases caused by multidrug-resistant strains of *S. dysenteriae* type 1. However, investigations of subsequent outbreaks in India and Bangladesh detected resistance to norfloxacin, ciprofloxacin, and ofloxacin in 5% of isolates. The incidence of multidrug resistance parallels the widespread, uncontrolled use of antibiotics and calls for the rational use of effective drugs.

ANTIBIOTIC TREATMENT OF SHIGELLOSIS

(Table 154-1) Because of the ready transmissibility of *Shigella*, current public health recommendations in the United States are that every case be treated with antibiotics. Ciprofloxacin is recommended as first-line treatment. A number of other drugs have been tested and shown to be effective, including ceftriaxone, azithromycin, pivmecillinam, and some fifth-generation quinolones. While infections caused by non-*dysenteriae Shigella* in immunocompetent individuals are routinely treated with a 3-day course of antibiotics, it is recommended that *S. dysenteriae* type 1 infections be treated for 5 days and that *Shigella* infections in immunocompromised patients be treated for 7–10 days.

Table 154-1 Recommended Antimicrobial Therapy for Shigellosis

	Treatment Schedule		
Antimicrobial Agent	**Children**	**Adults**	**Limitations**
First line			
Ciprofloxacin	15 mg/kg	500 mg	
	2 times per day for 3 days, PO		
Second line			
Pivmecillinam	20 mg/kg	100 mg	Cost
	4 times per day for 5 days, PO		No pediatric formulation
			Frequent administration
			Resistance emerging
Ceftriaxone	50–100 mg/kg	–	Efficacy not validated
	Once a day IM for 2–5 days		Must be injected
Azithromycin	6–20 mg/kg	1–1.5 g	Cost
	Once a day for 1–5 days, PO		Efficacy not validated
			MIC near serum concentration Rapid emergence of resistance and spread to other bacteria

Source: WHO Library Cataloguing-in-Publication Data: Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1 (www.searo.who.int/LinkFiles/CAH_Publications_shigella.pdf).

Treatment for shigellosis must be adapted to the clinical context, with the recognition that the most fragile patients are children <5 years old, who represent two-thirds of all cases worldwide. There are few data on the use of quinolones in children, but *Shigella*-induced dysentery is a well-recognized indication for their use. The half-life of ciprofloxacin is longer in infants than in older individuals. The ciprofloxacin dose generally recommended for children is 30 mg/kg per d in two divided doses. Adults living in areas with high standards of hygiene are likely to develop milder, shorter-duration disease, whereas infants in endemic areas can develop severe, sometimes fatal dysentery. In the former setting, treatment will remain minimal and bacteriologic proof of infection will often come after symptoms have resolved; in the latter setting, antibiotic treatment and more aggressive measures, possibly including resuscitation, are often required.

REHYDRATION AND NUTRITION

Shigella infection rarely causes significant dehydration. Cases requiring aggressive rehydration (particularly in industrialized countries) are uncommon. In developing countries, malnutrition remains the primary indicator for diarrhea-related death, highlighting the importance of nutrition in early management. Rehydration should be oral unless the patient is comatose or presents in shock. Because of the improved effectiveness of reduced-osmolarity oral rehydration solution (especially for children with acute noncholera diarrhea), the WHO and UNICEF now recommend a standard solution of 245 mOsm/L (sodium, 75 mmol/L; chloride, 65 mmol/L; glucose (anhydrous), 75 mmol/L; potassium,

20 mmol/L; citrate, 10 mmol/L). In shigellosis, the coupled transport of sodium to glucose may be variably affected, but oral rehydration therapy remains the easiest and most efficient form of rehydration, especially in severe cases.

Nutrition should be started as soon as possible after completion of initial rehydration. Early refeeding is safe, well tolerated, and clinically beneficial. Because breast-feeding reduces diarrheal losses and the need for oral rehydration in infants, it should be maintained in the absence of contraindications (e.g., maternal HIV infection).

NONSPECIFIC, SYMPTOM-BASED THERAPY

Antimotility agents have been implicated in prolonged fever in volunteers with shigellosis. These agents are suspected of increasing the risk of toxic megacolon and are thought to have been responsible for HUS in children infected by EHEC strains. For safety reasons, it is better to avoid antimotility agents in bloody diarrhea.

TREATMENT OF COMPLICATIONS

There is no consensus regarding the best treatment for toxic megacolon. The patient should be assessed frequently by both medical and surgical teams. Anemia, dehydration, and electrolyte deficits (particularly hypokalemia) may aggravate colonic atony and should be actively treated. Nasogastric aspiration helps to deflate the colon. Parenteral nutrition has not been proven to be beneficial. Fever persisting beyond 48–72 h raises the possibility of local perforation or abscess. Most studies recommend colectomy if, after 48–72 h, colonic distention persists. However, some physicians recommend continuation of medical therapy for up to 7 days if the patient seems to be improving clinically despite persistent megacolon without free perforation. Intestinal perforation, either isolated or complicating toxic megacolon, requires surgical treatment and intensive medical support.

Rectal prolapse must be treated as soon as possible. With the health care provider using surgical gloves or a soft warm wet cloth and the patient in the knee-chest position, the prolapsed rectum is gently pushed back into place. If edema of the rectal mucosa is evident (rendering reintegration difficult), it can be osmotically reduced by applying gauze impregnated with a warm solution of saturated magnesium sulfate. Rectal prolapse often relapses but usually resolves along with the resolution of dysentery.

HUS must be treated by water restriction, including discontinuation of oral rehydration solution and potassium-rich alimentation. Hemofiltration is usually required.

Prevention

Hand washing after defecation or handling of children's feces and before handling of food is recommended. Stool decontamination (e.g., with sodium hypochlorite), together with a cleaning protocol for medical staff as well as for patients, has proven useful in limiting the spread of infection during *Shigella* outbreaks. Ideally, patients should have a negative stool culture before their infection is considered cured. Recurrences are rare if therapeutic and preventive measures are correctly implemented.

Although several live attenuated oral and subunit parenteral vaccine candidates have been produced and are undergoing clinical trials, no vaccine against shigellosis is currently available. Especially given the rapid progression of antibiotic resistance in *Shigella*, a vaccine is urgently needed.

FURTHER READINGS

Bennish ML, Wojtyniak BJ: Mortality due to shigellosis: Community and hospital data. Rev Infect Dis 13(Suppl 4):S245, 1991

Cossart P, Sansonetti PJ: Bacterial invasion: The paradigms of enteroinvasive pathogens. Science 304:242, 2004[PMID: 15073367]

Kotloff KL et al: Overview of live vaccine strategies against *Shigella*, in *New Generation Vaccines*, 3d ed, MM Levine et al (eds). London, Informa Healthcare, 2004, pp 723–735

——— et al: Global burden of *Shigella* infections: Implications for vaccine development and implementation of control strategies. Bull World Health Organ 77:651, 1999

Niyogi SK: Shigellosis. J Microbiol 43:133, 2005[PMID: 15880088]

Phalipon A, Sansonetti PJ: Shigella's ways of manipulating the host intestinal innate and adaptive immune system: A tool box for survival? Immunol Cell Biol 85:119, 2007 Traa BS et al: Antibiotics for the treatment of dysentery in children. Int J Epidemiol 39 (Suppl 1):i70, 2010

Von Seidlein L et al: A multicentre study of *Shigella* diarrhoea in six Asian countries: Disease burden, clinical manifestations, and microbiology. PLoS Med 3:e353, 2006

World Health Organization: Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1. WHO Library Cataloguing-in-Publication Data (*www.searo.who.int/LinkFiles/CAH_Publications_shigella.pdf*)

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS DUE TO *CAMPYLOBACTER* AND RELATED ORGANISMS: INTRODUCTION

Definition

Bacteria of the genus *Campylobacter* and of the related genera *Arcobacter* and *Helicobacter* (Chap. 151) cause a variety of inflammatory conditions. Although acute diarrheal illnesses are most common, these organisms may cause infections in virtually all parts of the body, especially in compromised hosts, and these infections may have late nonsuppurative sequelae. The designation *Campylobacter* comes from the Greek for "curved rod" and refers to the organism's vibrio-like morphology.

Etiology

Campylobacters are motile, non-spore-forming, curved, gram-negative rods. Originally known as *Vibrio fetus*, these bacilli were reclassified as a new genus in 1973, after their dissimilarity to other vibrios was recognized. More than 15 species have since been identified. These species are currently divided into three genera: *Campylobacter*, *Arcobacter*, and *Helicobacter*. Not all of the species are pathogens of humans. The human pathogens fall into two major groups: those that primarily cause diarrheal disease and those that cause extraintestinal infection. The principal diarrheal pathogen is *C. jejuni*, which accounts for 80–90% of all cases of recognized illness due to campylobacters and related genera. Other organisms that cause diarrheal disease include *C. coli*, *C. upsaliensis*, *C. lari*, *C. hyointestinalis*, *C. fetus*, *A. butzleri*, *A. cryaerophilus*, *H. cinaedi*, and *H. fennelliae*. The two *Helicobacter* species causing diarrheal disease, *H. cinaedi* and *H. fennelliae*, are intestinal rather than gastric organisms; in terms of the clinical features of the illnesses they cause, these species most closely resemble *Campylobacter* rather than *H. pylori* (Chap. 151) and thus are considered in this chapter.

The major species causing extraintestinal illnesses is *C. fetus*. However, any of the diarrheal agents listed above may cause systemic or localized infection as well, especially in compromised hosts. Neither aerobes nor strict anaerobes, these microaerophilic organisms are adapted for survival in the gastrointestinal mucous layer. This chapter focuses on *C. jejuni* and *C. fetus* as the major pathogens in and prototypes for their groups. The key features of infection are listed by species (excluding *C. jejuni*, described in detail in the text below) in Table 155-1.

Table 155-1 Clinical Features Associated with Infection Due to "Atypical" *Campylobacter* and Related Species Implicated as Causes of Human Illness

Species	Common Clinical Features	Less Common Clinical Features	Additional Information
Campylobacter coli	Fever, diarrhea, abdominal pain	Bacteremia[a]	Clinically indistinguishable from *C. jejuni*
Campylobacter fetus	Bacteremia,[a] sepsis, meningitis, vascular infections	Diarrhea, relapsing fevers	Not usually isolated from media containing cephalothin or incubated at 42°C
Campylobacter upsaliensis	Watery diarrhea, low-grade fever, abdominal pain	Bacteremia, abscesses	Difficult to isolate because of cephalothin susceptibility
Campylobacter lari	Abdominal pain, diarrhea	Colitis, appendicitis	Seagulls frequently colonized; organism often transmitted to humans via contaminated water
Campylobacter hyointestinalis	Watery or bloody diarrhea, vomiting, abdominal pain	Bacteremia	Causes proliferative enteritis in swine
Helicobacter fennelliae	Chronic mild diarrhea, abdominal cramps, proctitis	Bacteremia[a]	Best treated with fluoroquinolones
Helicobacter cinaedi	Chronic mild diarrhea, abdominal cramps,	Bacteremia[a]	Best treated with fluoroquinolones; identified in healthy hamsters

	proctitis		
Campylobacter jejuni subspecies *doylei*	Diarrhea	Chronic gastritis, bacteremia[b]	Uncertain role as human pathogen
Arcobacter cryaerophilus	Diarrhea	Bacteremia	Cultured under aerobic conditions
Arcobacter butzleri	Fever, diarrhea, abdominal pain, nausea	Bacteremia, appendicitis	Cultured under aerobic conditions; enzootic in nonhuman primates
Campylobacter sputorum	Pulmonary, perianal, groin, and axillary abscesses; diarrhea	Bacteremia	Three clinically relevant biovars: *sputorum, fecalis,* and *paraureolyticus*

[a]In immunocompromised hosts, especially HIV-infected persons.

[b]In children.

Source: Adapted from BM Allos, MJ Blaser: Clin Infect Dis 20:1092, 1995.

Epidemiology

Campylobacters are found in the gastrointestinal tract of many animals used for food (including poultry, cattle, sheep, and swine) and many household pets (including birds, dogs, and cats). These microorganisms usually do not cause illness in their animal hosts. In most cases, campylobacters are transmitted to humans in raw or undercooked food products or through direct contact with infected animals. In the United States and other developed countries, ingestion of contaminated poultry that has not been sufficiently cooked is the most common mode of acquisition (30–70% of cases). Other modes include ingestion of raw (unpasteurized) milk or untreated water, contact with infected household pets, travel to developing countries (campylobacters being among the leading causes of traveler's diarrhea; Chaps. 123 and 128), oral-anal sexual contact, and (occasionally) contact with an index case who is incontinent of stool (e.g., a baby).

Campylobacter infections are common. Several studies indicate that, in the United States, diarrheal disease due to campylobacters is more common than that due to *Salmonella* and *Shigella* combined. Infections occur throughout the year, but their incidence peaks during summer and early autumn. Persons of all ages are affected; however, attack rates for *C. jejuni* are highest among young children and young adults, while those for *C. fetus* are highest at the extremes of age. Systemic infections due to *C. fetus* (and to other *Campylobacter* and related species) are most common among compromised hosts. Persons at increased risk include those with AIDS, hypogammaglobulinemia, neoplasia, liver disease, diabetes mellitus, and generalized atherosclerosis as well as neonates and pregnant women. However, apparently healthy nonpregnant persons occasionally develop transient *Campylobacter* bacteremia as part of a gastrointestinal illness.

In contrast, in many developing countries, *C. jejuni* infections are hyperendemic, with the highest rates among children <2 years old. Infection rates fall with age, as does the illness-to-infection ratio. These observations suggest that frequent exposure to *C. jejuni* leads to the acquisition of immunity.

Pathology and Pathogenesis

C. jejuni infections may be subclinical, especially in hosts in developing countries who have had multiple prior infections and thus are partially immune. Symptomatic infections mostly occur within 2–4 days (range, 1–7 days) of exposure to the organism in food or water. The sites of tissue injury include the jejunum, ileum, and colon. Biopsies show an acute nonspecific inflammatory reaction, with neutrophils, monocytes, and eosinophils in the lamina propria, as well as damage to the epithelium, including loss of mucus, glandular degeneration, and crypt abscesses. Biopsy findings may be consistent with Crohn's disease or ulcerative colitis, but these "idiopathic" chronic inflammatory diseases should not be diagnosed unless infectious colitis, *specifically including* that due to infection with *Campylobacter* species and related organisms, has been ruled out.

The high frequency of *C. jejuni* infections and their severity and recurrence among hypogammaglobulinemic patients suggest that antibodies are important in protective immunity. The pathogenesis of infection is uncertain. Both the motility of the strain and its capacity to adhere to host tissues appear to favor disease, but classic enterotoxins and cytotoxins (although described and including cytolethal distending toxin, or CDT) appear not to play substantial roles in tissue injury or disease production. The organisms have been visualized within the epithelium, albeit in low numbers. The documentation of a significant tissue response and occasionally of *C. jejuni* bacteremia further suggests that tissue invasion is clinically significant, and in vitro studies are consistent with this pathogenetic feature.

The pathogenesis of *C. fetus* infections is better defined. Virtually all clinical isolates of *C. fetus* possess a proteinaceous capsule-like structure (an S-layer) that renders the organisms resistant to

complement-mediated killing and opsonization. As a result, *C. fetus* can cause bacteremia and can seed sites beyond the intestinal tract. The ability of the organism to switch the S-layer proteins expressed—a phenomenon that results in antigenic variability—may contribute to the chronicity and high rate of recurrence of *C. fetus* infections in compromised hosts.

Clinical Manifestations

The clinical features of infections due to *Campylobacter* and the related *Arcobacter* and intestinal *Helicobacter* species causing enteric disease appear to be highly similar. *C. jejuni* can be considered the prototype, in part because it is by far the most common enteric pathogen in the group. A prodrome of fever, headache, myalgia, and/or malaise often occurs 12–48 h before the onset of diarrheal symptoms. The most common signs and symptoms of the intestinal phase are diarrhea, abdominal pain, and fever. The degree of diarrhea varies from several loose stools to grossly bloody stools; most patients presenting for medical attention have ≥10 bowel movements on the worst day of illness. Abdominal pain usually consists of cramping and may be the most prominent symptom. Pain is usually generalized but may become localized; *C. jejuni* infection may cause pseudoappendicitis. Fever may be the only initial manifestation of *C. jejuni* infection, a situation mimicking the early stages of typhoid fever. Febrile young children may develop convulsions. *Campylobacter* enteritis is generally self-limited; however, symptoms persist for >1 week in 10–20% of patients seeking medical attention, and clinical relapses occur in 5–10% of such untreated patients. Studies of common-source epidemics indicate that milder illnesses or asymptomatic infections may commonly occur.

C. fetus may cause a diarrheal illness similar to that due to *C. jejuni*, especially in normal hosts. This organism also may cause either intermittent diarrhea or nonspecific abdominal pain without localizing signs. Sequelae are uncommon, and the outcome is benign. *C. fetus* may also cause a prolonged relapsing systemic illness (with fever, chills, and myalgias) that has no obvious primary source; this manifestation is especially common among compromised hosts. Secondary seeding of an organ (e.g., meninges, brain, bone, urinary tract, or soft tissue) complicates the course, which may be fulminant. *C. fetus* infections have a tropism for vascular sites: endocarditis, mycotic aneurysm, and septic thrombophlebitis may all occur. Infection during pregnancy often leads to fetal death. A variety of *Campylobacter* species and *H. cinaedi* can cause recurrent cellulitis with fever and bacteremia in immunocompromised hosts.

Complications

Except in infection with *C. fetus*, bacteremia is uncommon, developing most often in immunocompromised hosts and at the extremes of age. Three patterns of extraintestinal infection have been noted: (1) transient bacteremia in a normal host with enteritis (benign course, no specific treatment needed); (2) sustained bacteremia or focal infection in a normal host (bacteremia originating from enteritis, with patients responding well to antimicrobial therapy); and (3) sustained bacteremia or focal infection in a compromised host. Enteritis may not be clinically apparent. Antimicrobial therapy, possibly prolonged, is necessary for suppression or cure of the infection.

Campylobacter, *Arcobacter*, and intestinal *Helicobacter* infections in patients with AIDS or hypogammaglobulinemia may be severe, persistent, and extraintestinal; relapse after cessation of therapy is common. Hypogammaglobulinemic patients also may develop osteomyelitis and an erysipelas-like rash or cellulitis.

Local suppurative complications of infection include cholecystitis, pancreatitis, and cystitis; distant complications include meningitis, endocarditis, arthritis, peritonitis, cellulitis, and septic abortion. All these complications are rare, except in immunocompromised hosts. Hepatitis, interstitial nephritis, and the hemolytic-uremic syndrome occasionally complicate acute infection. Reactive arthritis and other rheumatologic complaints may develop several weeks after infection, especially in persons with the HLA-B27 phenotype. Guillain-Barré syndrome or its Miller Fisher (cranial polyneuropathy) variant follows *Campylobacter* infections uncommonly—i.e., in 1 of every 1000–2000 cases or, for certain *C. jejuni* serotypes (such as O19), in 1 of every 100–200 cases. Despite the low frequency of this complication, it is now estimated that *Campylobacter* infections, because of their high incidence, may trigger 20–40% of all cases of Guillain-Barré syndrome. Asymptomatic *Campylobacter* infection also may trigger this syndrome. Immunoproliferative small-intestinal disease (*alpha chain disease*), a form of lymphoma that originates in small-intestinal mucosa-associated lymphoid tissue, has been associated with *C. jejuni*; antimicrobial therapy has led to marked clinical improvement.

Diagnosis

In patients with *Campylobacter* enteritis, peripheral leukocyte counts reflect the severity of the inflammatory process. However, stools from nearly all *Campylobacter*-infected patients presenting for medical attention in the United States contain leukocytes or erythrocytes. Gram- or Wright-stained fecal smears should be examined in all suspected cases. When the diagnosis of *Campylobacter* enteritis is suspected on the basis of findings indicating inflammatory diarrhea (fever, fecal leukocytes), clinicians can ask the microbiology laboratory to attempt the visualization of organisms with characteristic vibrioid morphology by direct microscopic examination of stools with Gram's staining or

to use phase-contrast or dark-field microscopy to identify the organisms' characteristic "darting" motility. Confirmation of the diagnosis of *Campylobacter* infection is based on identification of an isolate from cultures of stool, blood, or another site. *Campylobacter*-specific media should be used to culture stools from all patients with inflammatory or bloody diarrhea. Since all *Campylobacter* species are fastidious, they will not be isolated unless selective media or other selective techniques are used. Not all media are equally useful for isolation of the broad array of campylobacters; therefore, failure to isolate campylobacters from stool does not entirely rule out their presence. The detection of the organisms in stool almost always implies infection; there is a brief period of postconvalescent fecal carriage and no obvious commensalism in humans. In contrast, *C. sputorum* and related organisms found in the oral cavity are commensals that only rarely have pathogenic significance. Because of the low levels of metabolic activity of *Campylobacter* species in standard blood culture media, *Campylobacter* bacteremia may be difficult to detect unless laboratorians check for low-positive results in quantitative assays.

Differential Diagnosis

The symptoms of *Campylobacter* enteritis are not sufficiently unusual to distinguish this illness from that due to *Salmonella*, *Shigella*, *Yersinia*, and other pathogens. The combination of fever and fecal leukocytes or erythrocytes is indicative of inflammatory diarrhea, and definitive diagnosis is based on culture or demonstration of the characteristic organisms on stained fecal smears. Similarly, extraintestinal *Campylobacter* illness is diagnosed by culture. Infection due to *Campylobacter* should be suspected in the setting of septic abortion, and that due to *C. fetus* should be suspected specifically in the setting of septic thrombophlebitis. It is important to reiterate that (1) the presentation of *Campylobacter* enteritis may mimic that of ulcerative colitis or Crohn's disease, (2) *Campylobacter* enteritis is much more common than either of the latter (especially among young adults), and (3) biopsy may not distinguish among these entities. Thus a diagnosis of inflammatory bowel disease should not be made until *Campylobacter* infection has been ruled out, especially in persons with a history of foreign travel, significant animal contact, immunodeficiency, or exposure incurring a high risk of transmission.

Treatment: *Campylobacter* Infection

Fluid and electrolyte replacement is central to the treatment of diarrheal illnesses (Chap. 128). Even among patients presenting for medical attention with *Campylobacter* enteritis, not all clearly benefit from specific antimicrobial therapy. Indications for therapy include high fever, bloody diarrhea, severe diarrhea, persistence for >1 week, and worsening of symptoms. A 5- to 7-day course of erythromycin (250 mg orally four times daily or—for children—30–50 mg/kg per day, in divided doses) is the regimen of choice. Both clinical trials and in vitro susceptibility testing indicate that other macrolides, including azithromycin (a 1- or 3-day regimen), also are useful therapeutic agents. An alternative regimen for adults is ciprofloxacin (500 mg orally twice daily) or another fluoroquinolone for 5–7 days, but resistance to this class of agents as well as to tetracyclines has been increasing. Patients infected with antibiotic-resistant strains are at increased risk of adverse outcomes. Use of antimotility agents, which may prolong the duration of symptoms and have been associated with toxic megacolon and with death, is not recommended.

For systemic infections, treatment with gentamicin (1.7 mg/kg IV every 8 h after a loading dose of 2 mg/kg), imipenem (500 mg IV every 6 h), or chloramphenicol (50 mg/kg IV each day in three or four divided doses) should be started empirically, but susceptibility testing should then be performed. Ciprofloxacin and amoxicillin/clavulanate are alternative agents for susceptible strains. In the absence of immunocompromise or endovascular infections, therapy should be administered for 14 days. For immunocompromised patients with systemic infections due to *C. fetus* and for patients with endovascular infections, prolonged therapy (for up to 4 weeks) is usually necessary. For recurrent infections in immunocompromised hosts, lifelong therapy/prophylaxis is sometimes necessary.

Prognosis

Nearly all patients recover fully from *Campylobacter* enteritis, either spontaneously or after antimicrobial therapy. Volume depletion probably contributes to the few deaths that are reported. As stated above, occasional patients develop reactive arthritis or Guillain-Barré syndrome or its variants. Systemic infection with *C. fetus* is much more often fatal than that due to related species; this higher mortality rate reflects in part the population affected. Prognosis depends on the rapidity with which appropriate therapy is begun. Otherwise-healthy hosts usually survive *C. fetus* infections without sequelae. Compromised hosts often have recurrent and/or life-threatening infections due to a variety of *Campylobacter* species.

FURTHER READINGS

Gradel KO et al: Increased short- and long-term risk of inflammatory bowel disease after *Salmonella* or *Campylobacter* gastroenteritis. Gastroenterology 137:495, 2009[PMID: 19361507]

——— et al: Increased risk of zoonotic *Salmonella* and *Campylobacter* gastroenteritis in patients with

haematological malignancies: A population-based study. Ann Hematol 88:761, 2009

Helms M et al: Adverse health events associated with antimicrobial drug resistance in *Campylobacter* species: A registry-based cohort study. J Infect Dis 191:1050, 2005[PMID: 15747238]

Lang DR et al (eds): Development of Guillain-Barré syndrome following *Campylobacter* infection. J Infect Dis 176:S91, 1997

Lecuit M et al: Immunoproliferative small intestinal disease associated with *Campylobacter jejuni* . N Engl J Med 350:239, 2004[PMID: 14724303]

Mead PS et al: Food-related illness and death in the United States. Emerg Infect Dis 5:607, 1999[PMID: 10511517]

Nachamkin I et al (eds): *Campylobacter jejuni* , 3rd ed. Washington, American Society for Microbiology, 2008

Nylund CM et al: Bacterial enteritis as a risk factor for childhood intussusception: A retrospective cohort study. J Pediatr 156:761, 2010[PMID: 20138300]

Smith KE et al: Quinolone-resistant *Campylobacter jejuni* infections in Minnesota, 1992–1998. Investigation Team. N Engl J Med 340:1525, 1999[PMID: 10332013]

Tribble DR et al: Traveler's diarrhea in Thailand: Randomized, double-blind trial comparing single-dose and 3-day azithromycin-based regimens with a 3-day levofloxacin regimen. Clin Infect Dis 44:338, 2007[PMID: 17205438]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHOLERA AND OTHER VIBRIOSES: INTRODUCTION

Members of the genus *Vibrio* cause a number of important infectious syndromes. Classic among them is cholera, a devastating diarrheal disease caused by *V. cholerae* that has been responsible for seven global pandemics and much suffering over the past two centuries. Epidemic cholera remains a significant public health concern in the developing world today. Other vibrioses caused by other *Vibrio* species include syndromes of diarrhea, soft tissue infection, or primary sepsis. All *Vibrio* species are highly motile, facultatively anaerobic, curved gram-negative rods with one or more flagella. In nature, vibrios most commonly reside in tidal rivers and bays under conditions of moderate salinity. They proliferate in the summer months when water temperatures exceed 20°C. As might be expected, the illnesses they cause also increase in frequency during the warm months.

CHOLERA

Definition

Cholera is an acute diarrheal disease that can, in a matter of hours, result in profound, rapidly progressive dehydration and death. Accordingly, cholera gravis (the severe form of cholera) is a much-feared disease, particularly in its epidemic presentation. Fortunately, prompt aggressive fluid repletion and supportive care can obviate the high mortality that cholera has historically wrought. While the term *cholera* has occasionally been applied to any severely dehydrating secretory diarrheal illness, whether infectious in etiology or not, it now refers to disease caused by *V. cholerae* serogroup O1 or O139—i.e., the serogroups with epidemic potential.

Microbiology and Epidemiology

The species *V. cholerae* is classified into more than 200 serogroups based on the carbohydrate determinants of their lipopolysaccharide (LPS) O antigens. Although some non-O1 *V. cholerae* serogroups (strains that do not agglutinate in antisera to the O1 group antigen) have occasionally caused sporadic outbreaks of diarrhea, serogroup O1 was, until the emergence of serogroup O139 in 1992, the exclusive cause of epidemic cholera. Two biotypes of *V. cholerae* O1, classical and El Tor, are distinguished. Each biotype is further subdivided into two serotypes, termed *Inaba* and *Ogawa*.

The natural habitat of *V. cholerae* is coastal salt water and brackish estuaries, where the organism lives in close relation to plankton. Humans become infected incidentally but, once infected, can act as vehicles for spread. Ingestion of water contaminated by human feces is the most common means of acquisition of *V. cholerae*. Consumption of contaminated food also can contribute to spread. There is no known animal reservoir. While the infectious dose is relatively high, it is markedly reduced in hypochlorhydric persons, in those using antacids, and when gastric acidity is buffered by a meal. Cholera is predominantly a pediatric disease in endemic areas, but it affects adults and children equally when newly introduced into a population. In endemic areas, the burden of disease is often greatest during "cholera seasons" associated with high temperatures, heavy rainfall, and flooding, but cholera can occur year-round. For unexplained reasons, susceptibility to cholera is significantly influenced by ABO blood group status; persons with type O blood are at greatest risk of severe disease if infected, while those with type AB are at least risk.

Cholera is native to the Ganges delta in the Indian subcontinent. Since 1817, seven global pandemics have occurred. The current (seventh) pandemic—the first due to the El Tor biotype—began in Indonesia in 1961 and spread throughout Asia as *V. cholerae* El Tor displaced the endemic classical biotype. In the early 1970s, El Tor cholera erupted in Africa, causing major epidemics before becoming a persistent endemic problem. Currently, >90% of cholera cases reported annually to the World Health Organization (WHO) are from Africa **(Fig. 156-1)**, but the true burden in Africa as well as in Asia is unknown since diagnosis is often syndromic and since many countries with endemic cholera do not report cholera to the WHO. It is possible that >3 million cases of cholera occur yearly (of which only ~200,000 are reported to the WHO), resulting in >100,000 deaths annually (of which <5000 are reported to the WHO).

Figure 156-1

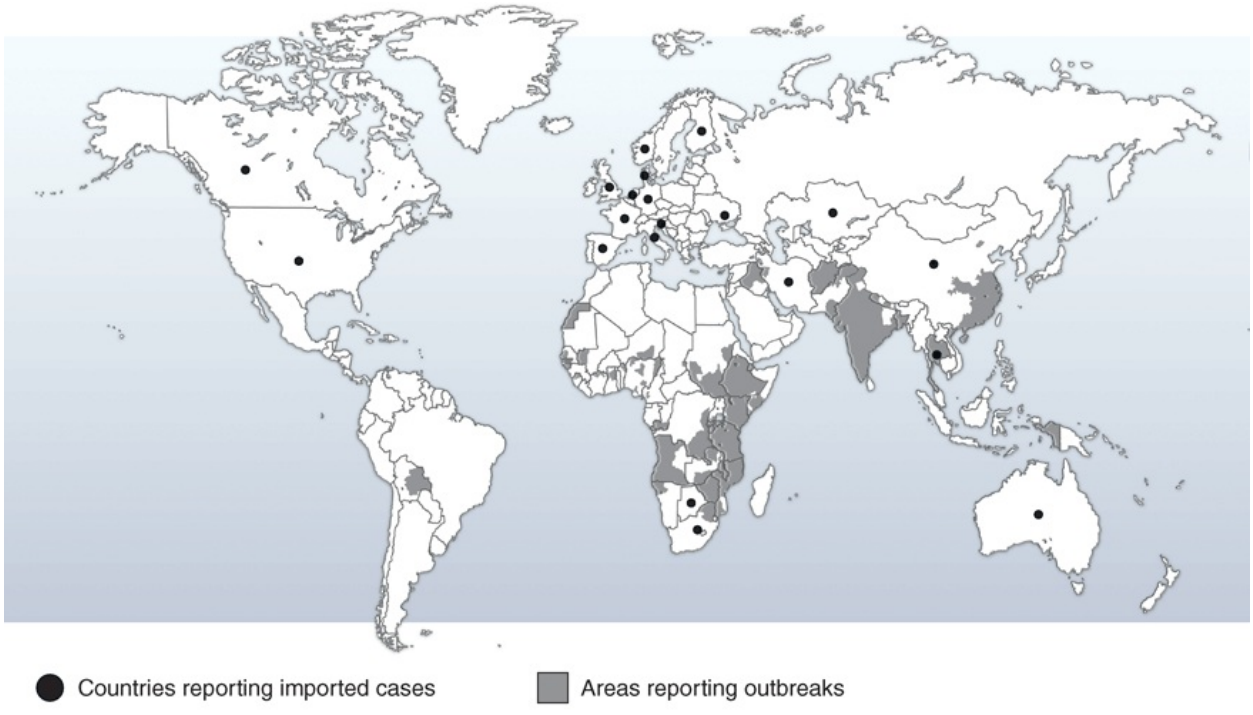

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

World distribution of cholera in 2009. *(Adapted from WHO: Cholera, 2009.)*

The recent history of cholera has been punctuated by severe outbreaks, especially among impoverished or displaced persons. Such outbreaks are often precipitated by war or other circumstances that lead to the breakdown of public health measures. Such was the case in the camps for Rwandan refugees set up in 1994 around Goma, Zaire; in 2008–2009 in Zimbabwe; and in 2010 in Haiti. Since 1973, sporadic endemic infections due to *V. cholerae* O1 strains related to the seventh-pandemic strain have been recognized along the U.S. Gulf Coast of Louisiana and Texas. These infections are typically associated with the consumption of contaminated, locally harvested shellfish. Occasionally, cases in U.S. locations remote from the Gulf Coast have been linked to shipped-in Gulf Coast seafood.

After a century without cholera in Latin America, the current cholera pandemic reached Central and South America in 1991. Following an initial explosive spread that affected millions **(Fig. 156-2)**, the burden of disease has markedly decreased in Latin America, although, as it did in Africa two decades earlier, the epidemic El Tor strain proved capable of establishing itself in inland fresh waters rather than in its classic niche of coastal salt waters. In 2010, cholera reappeared in Haiti after a century-long absence.

Figure 156-2

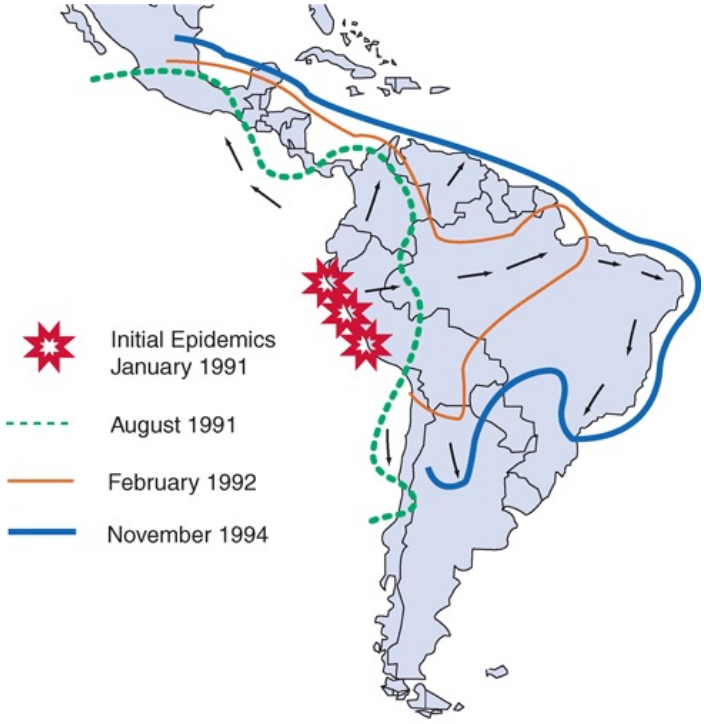

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spread of *Vibrio cholerae* O1 in the Americas, 1991–1994. *(Courtesy of Dr. Robert V. Tauxe, Centers for Disease Control and Prevention, Atlanta; with permission.)*

In October 1992, a large-scale outbreak of clinical cholera caused by a new serogroup, O139, occurred in southeastern India. The organism appears to be a derivative of El Tor O1 but has a distinct LPS and an immunologically related O-antigen polysaccharide capsule. (O1 organisms are not encapsulated.) After an initial spread across 11 Asian countries **(Fig. 156-3)**, *V. cholerae* O139 has once again been largely replaced by O1, although O139 still causes a minority of cases in some Asian countries. The clinical manifestations of disease caused by *V. cholerae* O139 are indistinguishable from those of O1 cholera. Immunity to one, however, is not protective against the other.

Figure 156-3

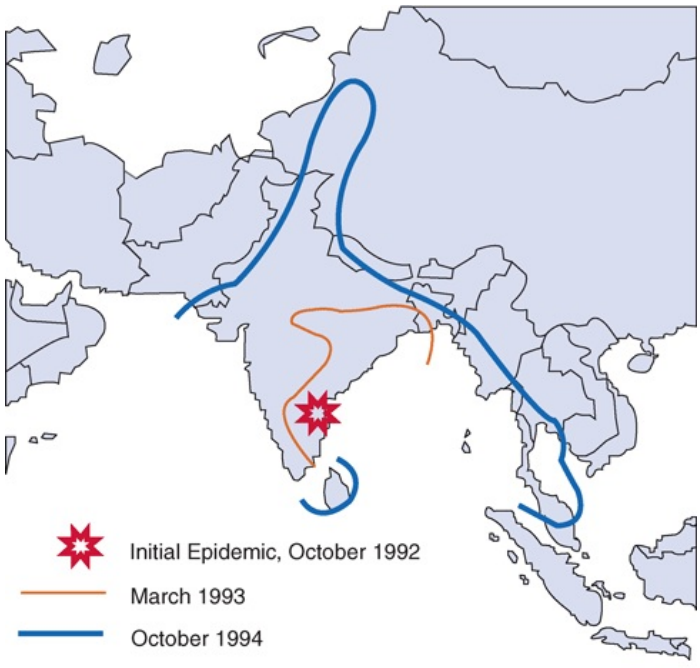

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spread of *Vibrio cholerae* O139 in the Indian subcontinent and elsewhere in Asia, 1992–1994.

(Courtesy of Dr. Robert V. Tauxe, CDC, Atlanta; with permission.)

Pathogenesis

In the final analysis, cholera is a toxin-mediated disease. The watery diarrhea characteristic of cholera is due to the action of cholera toxin, a potent protein enterotoxin elaborated by the organism in the small intestine. The toxin-coregulated pilus (TCP), so named because its synthesis is regulated in parallel with that of cholera toxin, is essential for *V. cholerae* to survive and multiply in (colonize) the small intestine. Cholera toxin, TCP, and several other virulence factors are coordinately regulated by ToxR. This protein modulates the expression of genes coding for virulence factors in response to environmental signals via a cascade of regulatory proteins. Additional regulatory processes, including bacterial responses to the density of the bacterial population (in a phenomenon known as *quorum sensing*), control the virulence of *V. cholerae*.

Once established in the human small bowel, the organism produces cholera toxin, which consists of a monomeric enzymatic moiety (the A subunit) and a pentameric binding moiety (the B subunit). The B pentamer binds to G_{M1} ganglioside, a glycolipid on the surface of epithelial cells that serves as the toxin receptor and makes possible the delivery of the A subunit to its cytosolic target. The activated A subunit (A_1) irreversibly transfers ADP-ribose from nicotinamide adenine dinucleotide to its specific target protein, the GTP-binding regulatory component of adenylate cyclase. The ADP-ribosylated G protein upregulates the activity of adenylate cyclase; the result is the intracellular accumulation of high levels of cyclic AMP. In intestinal epithelial cells, cyclic AMP inhibits the absorptive sodium transport system in villus cells and activates the secretory chloride transport system in crypt cells, and these events lead to the accumulation of sodium chloride in the intestinal lumen. Since water moves passively to maintain osmolality, isotonic fluid accumulates in the lumen. When the volume of that fluid exceeds the capacity of the rest of the gut to resorb it, watery diarrhea results. Unless the wasted fluid and electrolytes are adequately replaced, shock (due to profound dehydration) and acidosis (due to loss of bicarbonate) follow. Although perturbation of the adenylate cyclase pathway is the primary mechanism by which cholera toxin causes excess fluid secretion, cholera toxin also enhances intestinal secretion via prostaglandins and/or neural histamine receptors.

The *V. cholerae* genome comprises two circular chromosomes. Lateral gene transfer has played a key role in the evolution of epidemic *V. cholerae*. The genes encoding cholera toxin (*ctxAB*) are part of the genome of a bacteriophage, $CTX\Phi$. The receptor for this phage on the *V. cholerae* surface is the intestinal colonization factor TCP. Since *ctxAB* is part of a mobile genetic element ($CTX\Phi$), horizontal transfer of this bacteriophage may account for the emergence of new toxigenic *V. cholerae* serogroups. Many of the other genes important for *V. cholerae* pathogenicity, including the genes encoding the biosynthesis of TCP, those encoding accessory colonization factors, and those regulating virulence gene expression, are clustered together in the *V. cholerae* pathogenicity island. Similar clustering of virulence genes is found in other bacterial pathogens. It is believed that pathogenicity islands are acquired by horizontal gene transfer. *V. cholerae* O139 is probably derived from an El Tor O1 strain that acquired the genes for O139 O-antigen synthesis by horizontal gene transfer.

Clinical Manifestations

Individuals infected with *V. cholerae* O1 or O139 exhibit a range of clinical manifestations. Some individuals are asymptomatic or have only mild diarrhea; others present with the sudden onset of explosive and life-threatening diarrhea (*cholera gravis*). The reasons for the range in signs and symptoms of disease are incompletely understood but include the level of preexisting immunity, blood type, and nutritional status. In a nonimmune individual, after a 24- to 48-h incubation period, cholera characteristically begins with the sudden onset of painless watery diarrhea that may quickly become voluminous. Patients often vomit. In severe cases, volume loss can exceed 250 mL/kg in the first 24 h. If fluids and electrolytes are not replaced, hypovolemic shock and death may ensue. Fever is usually absent. Muscle cramps due to electrolyte disturbances are common. The stool has a characteristic appearance: a nonbilious, gray, slightly cloudy fluid with flecks of mucus, no blood, and a somewhat fishy, inoffensive odor. It has been called "rice-water" stool because of its resemblance to the water in which rice has been washed **(Fig. 156-4)**. Clinical symptoms parallel volume contraction: At losses of <5% of normal body weight, thirst develops; at 5–10%, postural hypotension, weakness, tachycardia, and decreased skin turgor are documented; and at >10%, oliguria, weak or absent pulses, sunken eyes (and, in infants, sunken fontanelles), wrinkled ("washerwoman") skin, somnolence, and coma are characteristic. Complications derive exclusively from the effects of volume and electrolyte depletion and include renal failure due to acute tubular necrosis. Thus, if the patient is adequately treated with fluid and electrolytes, complications are averted and the process is self-limited, resolving in a few days.

Figure 156-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rice water cholera stool. Note floating mucus and gray watery appearance. *(Courtesy of Dr. ASG Faruque, International Centre for Diarrhoeal Disease Research, Dhaka; with permission.)*

Laboratory data usually reveal an elevated hematocrit (due to hemoconcentration) in nonanemic patients; mild neutrophilic leukocytosis; elevated levels of blood urea nitrogen and creatinine consistent with prerenal azotemia; normal sodium, potassium, and chloride levels; a markedly reduced bicarbonate level (<15 mmol/L); and an elevated anion gap (due to increases in serum lactate, protein, and phosphate). Arterial pH is usually low (~7.2).

Diagnosis

The clinical suspicion of cholera can be confirmed by the identification of *V. cholerae* in stool; however, the organism must be specifically sought. With experience, it can be detected directly by dark-field microscopy on a wet mount of fresh stool, and its serotype can be discerned by immobilization with specific antiserum. Laboratory isolation of the organism requires the use of a selective medium such as taurocholate-tellurite-gelatin (TTG) agar or thiosulfate–citrate–bile salts–sucrose (TCBS) agar. If a delay in sample processing is expected, Carey-Blair transport medium and/or alkaline-peptone water-enrichment medium may be used as well. In endemic areas, there is little need for biochemical confirmation and characterization, although these tasks may be worthwhile in places where *V. cholerae* is an uncommon isolate. Standard microbiologic biochemical testing for Enterobacteriaceae will suffice for identification of *V. cholerae*. All vibrios are oxidase-positive. A point-of-care antigen-detection cholera dipstick assay is now commercially available for use in the field or where laboratory facilities are lacking.

Treatment: Cholera

Death from cholera is due to hypovolemic shock; thus treatment of individuals with cholera first and foremost requires fluid resuscitation and management. In light of the level of dehydration **(Table 156-1)** and the patient's age and weight, euvolemia should first be rapidly restored, and adequate hydration should then be maintained to replace ongoing fluid losses **(Table 156-2)**. Administration of oral rehydration solution (ORS) takes advantage of the hexose-Na^+ co-transport mechanism to move Na^+ across the gut mucosa together with an actively transported molecule such as glucose (or galactose). Cl^- and water follow. This transport mechanism remains intact even when cholera toxin is active. ORS may be made by adding safe water to prepackaged sachets containing salts and sugar or by adding 0.5 teaspoon of table salt (NaCl; 3.5 g) and 4 tablespoons of table sugar (glucose; 40 g) to 1 L of safe water. Potassium intake in bananas or green coconut water should be encouraged. A number of ORS formulations are available, and the WHO now recommends "low-osmolarity" ORS for treatment of individuals with dehydrating diarrhea of any cause **(Table 156-3)**. If available, rice-based ORS is considered superior to standard ORS in the treatment of cholera. ORS can be administered via a nasogastric tube to individuals who cannot ingest fluid; however, optimal management of individuals with severe dehydration includes the administration of IV fluid and electrolytes. Because profound acidosis (pH < 7.2) is common in this group, Ringer's lactate is the best choice among commercial products **(Table 156-4)**. It must be used with additional potassium supplements, preferably given by mouth. The total fluid deficit in severely dehydrated patients (>10% of body weight) can be replaced safely within the first 3–6 h of therapy, half within the first hour. Transient muscle cramps and tetany are common. Thereafter, oral therapy can usually be initiated, with the goal of maintaining fluid intake equal to fluid output. However, patients with continued large-volume diarrhea may require prolonged IV treatment to match gastrointestinal fluid losses. Severe hypokalemia can develop but will respond to potassium given either IV or orally. In the absence of adequate staff to monitor the patient's progress, the oral route of rehydration and potassium replacement is safer than

the IV route.

Table 156-1 Assessing the Degree of Dehydration in Patients with Cholera

Degree of Dehydration	Clinical Findings
None or mild, but diarrhea	Thirst in some cases; <5% loss of total body weight
Moderate	Thirst, postural hypotension, weakness, tachycardia, decreased skin turgor, dry mouth/tongue, no tears; 5–10% loss of total body weight
Severe	Unconsciousness, lethargy, or "floppiness"; weak or absent pulse; inability to drink; sunken eyes (and, in infants, sunken fontanelles); >10% loss of total body weight

Table 156-2 Treatment of Cholera, Based on Degree of Dehydration[a]

Degree of Dehydration, Patient's Age (Weight)	Treatment
None or Mild, but Diarrhea[b]	
<2 years	1/4–1/2 cup (50–100 mL) of ORS, to a maximum of 0.5 L/d
2–9 years	1/2–1 cup (100–200 mL) of ORS, to a maximum of 1 L/d
≥10 years	As much ORS as desired, to a maximum of 2 L/d
Moderate[b,c]	
<4 months (<5 kg)	200–400 mL of ORS
4–11 months (5–<8 kg)	400–600 mL of ORS
12–23 months (8–<11 kg)	600–800 mL of ORS
2–4 years (11–<16 kg)	800–1200 mL of ORS
5–14 years (16–<30 kg)	1200–2200 mL of ORS
≥15 years (≥30 kg)	2200–4000 mL of ORS
Severe[b]	
All ages and weights	IV fluid replacement with Ringer's lactate (or, if not available, normal saline): 100 mL/kg in first 3-h period (or first 6-h period for children <12 months old); start rapidly, then slow down; total of 200 mL/kg in first 24 h; continue until patient is awake, can ingest ORS, and no longer has a weak pulse

Note: Continue normal feeding during treatment.

[a]Adapted from World Health Organization: First steps for managing an outbreak of acute diarrhoea. Global Task Force on Cholera Control, 2009 (www.who.int/topics/cholera). ORS, oral rehydration solution.

[b]Reassess regularly; monitor stool and vomit output.

[c]Amounts of ORS listed should be given within the first 4 h.

Table 156-3 Composition of World Health Organization Reduced-Osmolarity Oral Rehydration Solution (ORS)[a,b]

Constituent	Concentration, mmol/L
Na^+	75
K^+	20
Cl^-	65
Citrate[c]	10
Glucose	75
Total osmolarity	245

[a]Contains (per package, to be added to 1 L of drinking water): NaCl, 2.6 g; $Na_3C_6H_5O_7 \cdot 2H_2O$, 2.9 g; KCl, 1.5 g; and glucose (anhydrous), 13.5 g.

[b]If prepackaged ORS is unavailable, a simple homemade alternative can be prepared by combining 3.5 g (~1/2 teaspoon) of NaCl with either 50 g of precooked rice cereal or 40 g (4 tablespoons) of Table sugar (sucrose) in 1 L of drinking water. In that case, potassium must be supplied separately (e.g., in orange juice or coconut water).

[c]10 mmol citrate per liter, which supplies 30 mmol HCO_3/L.

Table 156-4 Electrolyte Composition of Cholera Stool and of Intravenous Rehydration Solution

	Concentration, mmol/L			
Substance	**Na^+**	**K^+**	**Cl^-**	**Base**
Stool				
Adult	135	15	100	45
Child	100	25	90	30
Ringer's lactate	130	4[a]	109	28

[a]Potassium supplements, preferably administered by mouth, are required to replace the usual potassium losses from stool.

Although not necessary for cure, the use of an antibiotic to which the organism is susceptible diminishes the duration and volume of fluid loss and hastens clearance of the organism from the stool. The WHO recommends administration of antibiotics to cholera patients only if they are severely dehydrated, although wider use is often justifiable. Doxycycline (a single dose of 300 mg) or tetracycline (12.5 mg/kg four times a day for 3 days) may be effective in adults but is not recommended for children <8 years of age because of possible deposition in bone and developing teeth. Emerging drug resistance is an ever-present concern. For nonpregnant adults with cholera in areas where tetracycline resistance is prevalent, ciprofloxacin [either in a single dose (30 mg/kg, not to exceed a total dose of 1 g) or in a short course (15 mg/kg bid for 3 days, not to exceed a total daily dose of 1 g)], erythromycin (40–50 mg/kg daily in three divided doses for 3 days), or azithromycin (a single 1-g dose) may be a clinically effective substitute. Pregnant women and children are usually treated with erythromycin or azithromycin (10 mg/kg in children).

Prevention

Provision of safe water and facilities for sanitary disposal of feces, improved nutrition, and attention to food preparation and storage in the household can significantly reduce the incidence of cholera.

Much effort has been devoted to the development of an effective cholera vaccine over the past few decades, with a particular focus on oral vaccine strains. Traditional killed cholera vaccine given intramuscularly provides little protection to nonimmune subjects and predictably causes adverse effects, including pain at the injection site, malaise, and fever. The vaccine's limited efficacy is due, at least in part, to its failure to induce a local immune response at the intestinal mucosal surface.

Two types of oral cholera vaccines have been developed. The first is a killed whole-cell (WC) vaccine. Two formulations of the killed WC vaccine have been prepared: one that also contains the nontoxic B subunit of cholera toxin (WC/BS) and one composed solely of killed bacteria. In placebo-controlled field trials in Bangladesh, both of the killed vaccines offered significant protection from cholera for the first 6 months after vaccination, with protection rates of ~58% for WC vaccine and 85% for WC/BS vaccine. Protective efficacy rates for both vaccines declined to ~50% by 3 years after vaccine administration. Immunity was relatively sustained in persons vaccinated at an age of >5 years but was not well sustained in younger vaccinees. The WC/BS vaccine proved effective in a trial conducted in a sub-Saharan African population with a high prevalence of HIV infection. Trials of locally produced killed WC vaccines have yielded promising results in Vietnam and in Kolkata (Calcutta), India. Killed oral vaccines also confer herd protection to unvaccinated individuals living in proximity to vaccinated individuals. The WHO now recommends that vaccination against cholera be part of a larger response plan for populations at risk for epidemic cholera. The oral killed vaccines are available in Europe and Asia but (like other cholera vaccines) are not available in the United States.

The second type of cholera vaccine under development involves the use of oral live attenuated vaccine strains developed, for example, by the isolation or creation of mutants lacking the genes encoding cholera toxin. One such vaccine, CVD 103-HgR, was safe and immunogenic in phase 1 and 2 studies but afforded minimal protection in a large field trial in Indonesia. Other live attenuated vaccine candidate strains have been prepared from El Tor and O139 *V. cholerae* and are now undergoing clinical trials. The development of effective and safe cholera vaccines and strategies that provide long-lasting protective mucosal immunity, especially among malnourished, impoverished, and potentially HIV-infected adults and children (the individuals most at risk for cholera), is a priority. As mentioned above, no cholera vaccine is commercially available in the United States.

OTHER *VIBRIO* SPECIES

The genus *Vibrio* includes several human pathogens that do not cause cholera. Abundant in coastal waters throughout the world, noncholera vibrios can reach high concentrations in the tissues of filter-feeding mollusks.

As a result, human infection commonly follows the ingestion of seawater or of raw or undercooked shellfish **(Table 156-5)**. Most noncholera vibrios can be cultured on blood or MacConkey agar, which contains enough salt to support the growth of these halophilic species. In the microbiology laboratory, the species of noncholera vibrios are distinguished by standard biochemical tests. The most important of these organisms are *V. parahaemolyticus* and *V. vulnificus*.

Table 156-5 Features of Selected Noncholera Vibrioses

Organism	Vehicle or Activity	Host at Risk	Syndrome
V. parahaemolyticus	Shellfish, seawater	Normal	Gastroenteritis
	Seawater	Normal	Wound infection
Non-O1/O139 *V. cholerae*	Shellfish, travel	Normal	Gastroenteritis
	Seawater	Normal	Wound infection, otitis media
V. vulnificus	Shellfish	Immunosuppressed[a]	Sepsis, secondary cellulitis
	Seawater	Normal, immunosuppressed[a]	Wound infection, cellulitis
V. alginolyticus	Seawater	Normal	Wound infection, cellulitis, otitis
	Seawater	Burned, other immunosuppressed	Sepsis

[a]Especially with liver disease or hemochromatosis.

Source: Table 161-3 in *Harrisons Principles of Internal Medicine,* 14th edition.

The two major types of syndromes for which these species are responsible are gastrointestinal illness (due to *V. parahaemolyticus*, non-O1/O139 *V. cholerae*, *V. mimicus*, *V. fluvialis*, *V. hollisae*, and *V. furnissii*) and soft tissue infections (due to *V. vulnificus*, *V. alginolyticus*, and *V. damselae*). *V. vulnificus* is also a cause of primary sepsis in some compromised individuals.

Species Associated Primarily with Gastrointestinal Illness

PARAHAEMOLYTICUS

Widespread in marine environments, the halophilic *V. parahaemolyticus* causes food-borne enteritis worldwide. This species was originally implicated in enteritis in Japan in 1953, accounting for 24% of reported cases in one study—a rate that presumably was due to the common practice of eating raw seafood in that country. In the United States, common-source outbreaks of diarrhea caused by this organism have been linked to the consumption of undercooked or improperly handled seafood or of other foods contaminated by seawater. Since the mid-1990s, the incidence of *V. parahaemolyticus* infections has increased in several countries, including the United States. Serotypes O3:K6, O4:K68, and O1:K-untypable, which are genetically related to one another, account for this increase. The enteropathogenicity of *V. parahaemolyticus* is linked to its ability to cause hemolysis on Wagatsuma agar (i.e., the *Kanagawa phenomenon*). Although the mechanism by which the organism causes diarrhea remains unclear, the genome sequence of *V. parahaemolyticus* contains two type III secretion systems, which directly inject toxic bacterial proteins into host cells. *V. parahaemolyticus* should be considered a possible etiologic agent in all cases of diarrhea that can be linked epidemiologically to seafood consumption or to the sea itself.

Infections with *V. parahaemolyticus* can result in two distinct gastrointestinal presentations. The more common of the two presentations (including nearly all cases in North America) is characterized by watery diarrhea, usually occurring in conjunction with abdominal cramps, nausea, and vomiting and accompanied in ~25% of cases by fever and chills. After an incubation period of 4 h to 4 days, symptoms develop and persist for a median of 3 days. Dysentery, the less common presentation, is characterized by severe abdominal cramps, nausea, vomiting, and bloody or mucoid stools. *V. parahaemolyticus* also causes rare cases of wound infection and otitis and very rare cases of sepsis.

Most cases of *V. parahaemolyticus*–associated gastrointestinal illness, regardless of the presentation, are self-limited and require neither antimicrobial treatment nor hospitalization. Deaths are extremely rare among immunocompetent individuals. Severe infections are associated with underlying diseases, including diabetes, preexisting liver disease, iron-overload states, or immunosuppression. The occasional severe case should be treated with fluid replacement and antibiotics, as described above for cholera.

NON-O1/O139 (NONCHOLERA) V. CHOLERAE

The heterogeneous non-O1/O139 *V. cholerae* organisms cannot be distinguished from *V. cholerae* O1 or O139 by routine biochemical tests but do not agglutinate in O1 or O139 antiserum. Non-O1/O139 strains have caused several well-studied food-borne outbreaks of gastroenteritis and have also been responsible for sporadic cases of otitis media, wound infection, and bacteremia; although gastroenteritis outbreaks can occur, non-O1/O139 *V. cholerae* strains do not cause epidemics of cholera. Like other vibrios, non-O1/O139 *V. cholerae* organisms are widely distributed in marine environments. In most instances, recognized cases in the United States have been associated with the consumption of raw oysters or with recent travel, typically to Mexico. The broad clinical spectrum of diarrheal illness caused by these organisms is probably due to the group's heterogeneous virulence attributes.

In the United States, about half of all non-O1/O139 *V. cholerae* isolates are from stool samples. The typical

incubation period for gastroenteritis due to these organisms is <2 days, and the illness lasts for ~2–7 days. Patients' stools may be copious and watery or may be partly formed, less voluminous, and bloody or mucoid. Diarrhea can result in severe dehydration. Many cases include abdominal cramps, nausea, vomiting, and fever. Like those with cholera, patients who are seriously dehydrated should receive oral or IV fluids; the value of antibiotics is not clear.

Extraintestinal infections due to non-O1/O139 *V. cholerae* commonly follow occupational or recreational exposure to seawater. Around 10% of non-O1/O139 *V. cholerae* isolates come from cases of wound infection, 10% from cases of otitis media, and 20% from cases of bacteremia (which is particularly likely to develop in patients with liver disease). Extraintestinal infections should be treated with antibiotics. Information to guide antibiotic selection and dosing is limited, but most strains are sensitive in vitro to tetracycline, ciprofloxacin, and third-generation cephalosporins.

Species Associated Primarily with Soft Tissue Infection or Bacteremia (See Also Chap. 125)

V. VULNIFICUS

Infection with *V. vulnificus* is rare, but this organism is the most common cause of severe *Vibrio* infections in the United States. Like most vibrios, *V. vulnificus* proliferates in the warm summer months and requires a saline environment for growth. In this country, infections in humans typically occur in coastal states between May and October and most commonly affect men >40 years of age. *V. vulnificus* has been linked to two distinct syndromes: primary sepsis, which usually occurs in patients with underlying liver disease, and primary wound infection, which generally affects people without underlying disease. (*Vulnificus* is Latin for "wound maker.") Some authors have suggested that *V. vulnificus* also causes gastroenteritis independent of other clinical manifestations. *V. vulnificus* is endowed with a number of virulence attributes, including a capsule that confers resistance to phagocytosis and to the bactericidal activity of human serum as well as a cytolysin. Measured as the 50% lethal dose in mice, the organism's virulence is considerably increased under conditions of iron overload; this observation is consistent with the propensity of *V. vulnificus* to infect patients who have hemochromatosis.

Primary sepsis most often develops in patients who have cirrhosis or hemochromatosis. However, *V. vulnificus* bacteremia can also affect individuals who have hematopoietic disorders or chronic renal insufficiency, those who are using immunosuppressive medications or alcohol, or (in rare instances) those who have no known underlying disease. After a median incubation period of 16 h, the patient develops malaise, chills, fever, and prostration. One-third of patients develop hypotension, which is often apparent at admission. Cutaneous manifestations develop in most cases (usually within 36 h of onset) and characteristically involve the extremities (the lower more often than the upper). In a common sequence, erythematous patches are followed by ecchymoses, vesicles, and bullae. In fact, sepsis and hemorrhagic bullous skin lesions suggest the diagnosis in appropriate settings. Necrosis and sloughing may also be evident. Laboratory studies reveal leukopenia more often than leukocytosis, thrombocytopenia, or elevated levels of fibrin split products. *V. vulnificus* can be cultured from blood or cutaneous lesions. The mortality rate approaches 50%, with most deaths due to uncontrolled sepsis. Accordingly, prompt treatment is critical and should include empirical antibiotic administration, aggressive debridement, and general supportive care. *V. vulnificus* is sensitive in vitro to a number of antibiotics, including tetracycline, fluoroquinolones, and third-generation cephalosporins. Data from animal models suggest that either a fluoroquinolone or the combination of minocycline and cefotaxime should be used in the treatment of *V. vulnificus* septicemia.

V. vulnificus can infect either a fresh or an old wound that comes into contact with seawater; the patient may or may not have underlying disease. After a short incubation period (4 h to 4 days; mean, 12 h), the disease begins with swelling, erythema, and (in many cases) intense pain around the wound. These signs and symptoms are followed by cellulitis, which spreads rapidly and is sometimes accompanied by vesicular, bullous, or necrotic lesions. Metastatic events are uncommon. Most patients have a fever and leukocytosis. *V. vulnificus* can be cultured from skin lesions and occasionally from the blood. Prompt antibiotic therapy and debridement are usually curative.

V. ALGINOLYTICUS

First identified as a pathogen of humans in 1973, *V. alginolyticus* occasionally causes eye, ear, and wound infections. This species is the most salt-tolerant of the vibrios and can grow in salt concentrations of >10%. Most clinical isolates come from superinfected wounds that presumably become contaminated at the beach. Although its severity varies, *V. alginolyticus* infection tends not to be serious and generally responds well to antibiotic therapy and drainage. A few cases of otitis externa, otitis media, and conjunctivitis due to this pathogen have been described. Tetracycline treatment usually results in cure. *V. alginolyticus* is a rare cause of bacteremia in immunocompromised hosts.

ACKNOWLEDGMENT

The authors gratefully acknowledge the valuable contributions of Drs. Robert Deresiewicz and Gerald T. Keusch, coauthors of this chapter for previous editions.

FURTHER READINGS

Centers for disease control and prevention: Update on cholera—Haiti, Dominican Republic, and Florida, 2010. MMWR Morb Mortal Wkly Rep 59:1637, 2010

Chin CS et al: The origin of the Haitian cholera outbreak strain. N Engl J Med 364:33, 2011[PMID: 21142692]

Lucas Mes et al: Effectiveness of mass oral cholera vaccination in Beira, Mozambique. N Engl J Med 352:757, 2005[PMID: 15728808]

Ryan ET: The cholera pandemic, still with us after half a century: Time to rethink. PLoS Negl Trop Dis 5:e1003, 2011

Sack DA et al: Cholera. Lancet 363:223, 2004[PMID: 14738797]

Saha D et al: Single-dose azithromycin for the treatment of cholera in adults. N Engl J Med 354:2452, 2006[PMID: 16760445]

Sur D et al: Efficacy and safety of a modified killed-whole-cell oral cholera vaccine in India: An interim analysis of a cluster-randomised, double-blind, placebo-controlled trial. Lancet 374:1694, 2009[PMID: 19819004]

Vaccine Resource Library: Cholera Outbred Strains and Shigellosis (COTS) Program (http://www.path.org/vaccineresources/details.php?i=736)

World Health Organization: Cholera vaccines: WHO position paper. Wkly Epidemiol Rec 85:117, 2010

———: The Treatment of Diarrhoea: A Manual for Physicians and Other Senior Health Workers. Geneva, World Health Organization, 2005 (http://www.who.int/child_adolescent_health/documents/9241593180/en/index.html)

———: Cholera: global surveillance summary 2008. Wkly Epidemiol Rec 84:309, 2009 (http://www.who.int/wer/2009/wer8431/en/index.html)

———: "Cholera" and "First steps for managing an outbreak of acute diarrhoea" (www.who.int/topics/cholera/en/)

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BRUCELLOSIS: INTRODUCTION

Definition

Brucellosis is a bacterial zoonosis transmitted directly or indirectly to humans from infected animals, predominantly domesticated ruminants and swine. The disease is known colloquially as *undulant fever* because of its remittent character. Its distribution is worldwide apart from the few countries where it has been eradicated from the animal reservoir. Although brucellosis commonly presents as an acute febrile illness, its clinical manifestations vary widely, and definitive signs indicative of the diagnosis may be lacking. Thus the clinical diagnosis usually must be supported by the results of bacteriologic and/or serologic tests.

Etiologic Agents

Human brucellosis is caused by strains of *Brucella*, a bacterial genus that was previously suggested, on genetic grounds, to comprise a single species, *B. melitensis,* with a number of biologic variants exhibiting particular host preferences. This view was challenged on the basis of detailed differences in chromosomal structure and host preference. The traditional classification into nomen species is now favored both because of these differences and because this classification scheme closely reflects the epidemiologic patterns of the infection. The nomen system recognizes *B. melitensis*, which is the most common cause of symptomatic disease in humans and for which the main sources are sheep, goats, and camels; *B. abortus,* which is usually acquired from cattle or buffalo: *B. suis,* which generally is acquired from swine but has one variant enzootic in reindeer and caribou and another in rodents; and *B. canis*, which is acquired most often from dogs. *B. ovis,* which causes reproductive disease in sheep, and *B. neotomae,* which is specific for desert rodents, have not been clearly implicated in human disease. Other brucellae have been isolated from marine mammals, and two new nomen species, *B. ceti* sp. nov. and *B. pinnipedialis* sp. nov., have been proposed for these isolates; at least one case of laboratory-acquired human disease due to one of these proposed species has been described, and apparent cases of natural human infection have been reported. As infections in marine mammals seem widespread, more cases of zoonotic infection may be identified. Other newly proposed species include *B. microti* sp. nov. isolated from field voles and *B. inopinata* sp. nov. isolated from a patient with a breast implant. Moreover, it has become apparent that *Brucella* is closely related to the genus *Ochrobactrum*, which includes environmental bacteria sometimes associated with opportunistic infections.

All brucellae are small, gram-negative, unencapsulated, nonsporulating, nonmotile rods or coccobacilli. They grow aerobically on peptone-based medium incubated at 37°C; the growth of some types is improved by supplementary CO_2. In vivo, brucellae behave as facultative intracellular parasites. The organisms are sensitive to sunlight, ionizing radiation, and moderate heat; they are killed by boiling and pasteurization but are resistant to freezing and drying. Their resistance to drying renders brucellae stable in aerosol form, facilitating airborne transmission. The organisms can survive for up to 2 months in soft cheeses made from goat's or sheep's milk; for at least 6 weeks in dry soil contaminated with infected urine, vaginal discharge, or placental or fetal tissues; and for at least 6 months in damp soil or liquid manure kept under cool dark conditions. Brucellae are easily killed by a wide range of common disinfectants used under optimal conditions but are likely to be much more resistant at low temperatures or in the presence of heavy organic contamination.

Epidemiology

Brucellosis is a zoonosis whose occurrence is closely related to its prevalence in domesticated animals. The true global prevalence of human brucellosis is unknown because of the imprecision of diagnosis and the inadequacy of reporting and surveillance systems in many countries. Even in developed countries, the true incidence may be 10–20 times higher than the reported figures. Bovine brucellosis has been the target of control programs in many parts of the world and has been eradicated from the cattle populations of Australia, New Zealand, Bulgaria, Canada, Cyprus, Great Britain (including the Channel Islands), Japan, Luxembourg, Romania, the Scandinavian countries, Switzerland, and the Czech and Slovak Republics, among other nations. Its incidence has been reduced to a low level in the United States and most Western European countries, with a varied picture in other parts of the world. There is evidence of a resurgence in Eastern Europe following economic changes in recent years, and outbreaks have also occurred in Ireland. Efforts to eradicate *B. melitensis* infection from sheep and goat populations have been much less successful. These efforts

have relied heavily on vaccination programs, which have tended to fluctuate with changing economic and political conditions. In some countries (e.g., Israel), *B. melitensis* has caused serious outbreaks in cattle. Infections with *B. melitensis* still pose a major public health problem in Mediterranean countries; in western, central, and southern Asia; and in parts of Africa and South and Central America.

Human brucellosis is usually associated with occupational or domestic exposure to infected animals or their products. Farmers, shepherds, goatherds, veterinarians, and employees in slaughterhouses and meat-processing plants in endemic areas are occupationally exposed to infection. Family members of individuals involved in animal husbandry may be at risk, although it is often difficult to differentiate food-borne infection from environmental contamination under these circumstances. Laboratory workers who handle cultures or infected samples are also at risk. Travelers and urban residents usually acquire the infection through consumption of contaminated foods. In countries that have eradicated the disease, new cases are most commonly acquired abroad. Dairy products, especially soft cheeses, unpasteurized milk, and ice cream, are the most frequently implicated sources of infection; raw meat and bone marrow may be sources under exceptional circumstances. Infections acquired through cosmetic treatments using materials of fetal origin have been reported. Person-to-person transmission is extremely rare, as is transfer of infection by blood or tissue donation. Although brucellosis is a chronic intracellular infection, there is no evidence for increased prevalence or severity among individuals with HIV infection or with immunodeficiency or immunosuppression of other etiologies.

Brucellosis may be acquired by ingestion, inhalation, or mucosal or percutaneous exposure. Accidental injection of the live vaccine strains of *B. abortus* (19 and RB51) and *B. melitensis* (Rev 1) can cause disease. *B. melitensis* and *B. suis* have been developed as biological weapons by several countries and could be exploited for bioterrorism (Chap. 221). This possibility should be borne in mind in the event of sudden unexplained outbreaks.

Immunity and Pathogenesis

Exposure to brucellosis elicits both humoral and cell-mediated immune responses. The mechanisms of protective immunity against human brucellosis are presumed to be similar to those documented in laboratory animals. The response to infection and its outcome are influenced by the virulence, phase, and species of the infecting strain. Differences have been reported between *B. abortus* and *B. suis* in modes of cellular entry and subsequent compartmentalization and processing. Antibodies promote clearance of extracellular brucellae by bactericidal action and by facilitation of phagocytosis by polymorphonuclear and mononuclear phagocytes; however, antibodies alone cannot eradicate infection. Organisms taken up by macrophages and other cells can establish persistent intracellular infections. The key target cell is the macrophage, and bacterial mechanisms for suppressing intracellular killing and apoptosis result in very large intracellular populations. Opsonized bacteria are actively phagocytosed by neutrophilic granulocytes and by monocytes. In these and other cells, initial attachment takes place via specific receptors, including Fc, C3, fibronectin, and mannose-binding proteins. Opsonized—but not unopsonized—bacteria trigger an oxidative burst inside phagocytes. Unopsonized bacteria are internalized via similar receptors but at much lower efficiency. Smooth strains enter host cells via lipid rafts. Smooth lipopolysaccharide (LPS), β-cyclic glucan, and possibly an invasion-attachment protein (IalB) are involved in this process. Tumor necrosis factor α (TNF-α) produced early in the course of infection stimulates cytotoxic lymphocytes and activates macrophages, which can kill intracellular brucellae (probably mainly through production of reactive oxygen and nitrogen intermediates) and may clear infection. However, virulent *Brucella* cells can suppress the TNF-α response, and control of infection in this situation depends on macrophage activation and interferon γ (IFN-γ) responses. Cytokines such as interleukin (IL) 12 promote production of IFN-γ, which drives T_H1-type responses and stimulates macrophage activation. Inflammatory cytokines, including IL-4, IL-6, and IL-10, downregulate the protective response. As in other types of intracellular infection, it is assumed that initial replication of brucellae takes place within cells of the lymph nodes draining the point of entry. Subsequent hematogenous spread may result in chronic localizing infection at almost any site, although the reticuloendothelial system, musculoskeletal tissues, and genitourinary system are most frequently targeted. Both acute and chronic inflammatory responses develop in brucellosis, and the local tissue response may include granuloma formation with or without necrosis and caseation. Abscesses may also develop, especially in chronic localized infection.

The determinants of pathogenicity of *Brucella* have not been fully characterized, and the mechanisms underlying the manifestations of brucellosis are incompletely understood. The organism is a "stealth" pathogen whose survival strategy is centered on processes that avoid triggering innate immune responses and that permit survival within monocytic cells. The smooth *Brucella* LPS, which has an unusual O-chain and core-lipid composition, has relatively low endotoxin activity and plays a key role in pyrogenicity and in resistance to phagocytosis and serum killing in the nonimmune host. In addition, LPS is believed to play a key role in suppressing phagosome-lysosome fusion and diverting the internalized bacteria into vacuoles located in endoplasmic reticulum, where intracellular replication takes place. Specific exotoxins have not been isolated, but a type IV secretion system (VirB) that regulates intracellular survival and trafficking has been identified. In *B. abortus* this system can be activated extracellularly, but in *B. suis* it is activated (by low pH) only during intracellular growth.

Brucellae then produce acid-stable proteins that facilitate the organisms' survival in phagosomes and may enhance their resistance to reactive oxygen intermediates. A type III secretion system based on modified flagellar structures has also been identified. Virulent brucellae are resistant to defensins and produce a Cu-Zn superoxide dismutase that increases their resistance to reactive oxygen intermediates. A hemolysin-like protein may trigger the release of brucellae from infected cells.

Clinical Features

Brucellosis almost invariably causes fever, which may be associated with profuse sweats, especially at night. In endemic areas, brucellosis may be difficult to distinguish from the many other causes of fever. However, two features recognized in the nineteenth century distinguish brucellosis from other tropical fevers, such as typhoid and malaria: (1) Left untreated, the fever of brucellosis shows an undulating pattern that persists for weeks before the commencement of an afebrile period that may be followed by relapse. (2) The fever of brucellosis is associated with musculoskeletal symptoms and signs in about one-half of all patients.

The clinical syndromes caused by the different nomen species are similar, although *B. melitensis* tends to be associated with a more acute and aggressive presentation and *B. suis* with focal abscess induction. *B. abortus* infections may be more insidious in onset and more likely to become chronic. *B. canis* infections are reported to present frequently with acute gastrointestinal symptoms.

The incubation period varies from 1 week to several months, and the onset of fever and other symptoms may be abrupt or insidious. In addition to experiencing fever and sweats, patients become increasingly apathetic and fatigued; lose appetite and weight; and have nonspecific myalgia, headache, and chills. Overall, the presentation of brucellosis often fits one of three patterns: febrile illness that resembles typhoid but is less severe; fever and acute monoarthritis, typically of the hip or knee, in a young child; and long-lasting fever, misery, and low-back or hip pain in an older man. In an endemic area (e.g., much of the Middle East), a patient with fever and difficulty walking into the clinic would be regarded as having brucellosis until it was proved otherwise.

Diagnostic clues in the patient's history include travel to an endemic area, employment in a diagnostic microbiology laboratory, consumption of unpasteurized milk products (including soft cheeses), contact with animals, accidental inoculation with veterinary *Brucella* vaccines, and—in an endemic setting—a history of similar illness in the family (documented in almost 50% of cases). Focal features are present in the majority of patients. The most common are musculoskeletal pain and physical findings in the peripheral and axial skeleton (~40% of cases). Osteomyelitis more commonly involves the lumbar and low thoracic vertebrae than the cervical and high thoracic spine. Individual joints that are most commonly affected by septic arthritis are the knee, hip, sacroiliac, shoulder, and sternoclavicular joints; the pattern may be one of monoarthritis or polyarthritis. Osteomyelitis may also accompany septic arthritis.

In addition to the usual causes of vertebral osteomyelitis or septic arthritis, the most important differential diagnosis is tuberculosis. This point influences the therapeutic approach as well as the prognosis, given that several antimicrobial agents used to treat brucellosis are also used to treat tuberculosis. Septic arthritis in brucellosis progresses slowly, starting with small pericapsular erosions. In the vertebrae, anterior erosions of the superior end plate are typically the first features to become evident, with eventual involvement and sclerosis of the whole vertebra. Anterior osteophytes eventually develop, but vertebral destruction or impingement on the spinal cord is rare and usually suggests tuberculosis **(Table 157-1)**.

Table 157-1 Radiology of the Spine: Differentiation of Brucellosis from Tuberculosis

	Brucellosis	Tuberculosis
Site	Lumbar and others	Dorsolumbar
Vertebrae	Multiple or contiguous	Contiguous
Diskitis	Late	Early
Body	Intact until late	Morphology lost early
Canal compression	Rare	Common
Epiphysitis	Anterosuperior (Pom's sign)	General: upper and lower disk regions, central, subperiosteal
Osteophyte	Anterolateral (parrot beak)	Unusual
Deformity	Wedging uncommon	Anterior wedge, gibbus
Recovery	Sclerosis, whole body	Variable
Paravertebral abscess	Small, well-localized	Common and discrete loss, transverse process
Psoas abscess	Rare	More likely

Other systems may be involved in a manner that resembles typhoid. About one-quarter of patients have a dry cough, usually with few changes visible on the chest x-ray, although pneumonia, empyema, intrathoracic adenopathy, or lung abscess can occur. One-quarter of patients have hepatosplenomegaly, and 10–20% have significant lymphadenopathy; the differential diagnosis includes glandular fever–like illness such as that caused by Epstein-Barr virus, *Toxoplasma*, cytomegalovirus, HIV, or *Mycobacterium tuberculosis*. Up to 10% of men have acute epididymoorchitis, which must be distinguished from mumps and from surgical problems such as torsion. Prostatitis, inflammation of the seminal vesicles, salpingitis, and pyelonephritis all occur. There is an increased incidence of fetal loss among infected pregnant women, although teratogenicity has not been described and the tendency toward abortions is much less pronounced in humans than in farm animals.

Neurologic involvement is common, with depression and lethargy whose severity may not be fully appreciated by either the patient or the physician until after treatment. A small proportion of patients develop lymphocytic meningoencephalitis that mimics neurotuberculosis, atypical leptospirosis, or noninfectious conditions and that may be complicated by intracerebral abscess, a variety of cranial nerve deficits, or ruptured mycotic aneurysms.

Endocarditis occurs in ~1% of cases, most often affecting the aortic valve (natural or prosthetic). Any site in the body may be involved in metastatic abscess formation or inflammation; the female breast and the thyroid gland are affected particularly often. Nonspecific maculopapular rashes and other skin manifestations are uncommon and are rarely noticed by the patient even if they develop.

Diagnosis

Because the clinical picture of brucellosis is not distinctive, the diagnosis must be based on a history of potential exposure, a presentation consistent with the disease, and supporting laboratory findings. Results of routine biochemical assays are usually within normal limits, although serum levels of hepatic enzymes and bilirubin may be elevated. Peripheral leukocyte counts are usually normal or low, with relative lymphocytosis. Mild anemia may be documented. Thrombocytopenia and disseminated intravascular coagulation with raised levels of fibrinogen degradation products can develop. The erythrocyte sedimentation rate and C-reactive protein levels are often normal but may be raised.

In body fluids such as cerebrospinal fluid (CSF) or joint fluid, lymphocytosis and low glucose levels are the norm. Elevated CSF levels of adenosine deaminase cannot be used to distinguish tubercular meningitis, as they may also be found in brucellosis. Biopsied samples of tissues such as lymph node or liver may show noncaseating granulomas **(Fig. 157-1)** without acid/alcohol-fast bacilli. The radiologic features of bony disease develop late and are much more subtle than those of tuberculosis or septic arthritis of other etiologies, with less bone and joint destruction. Isotope scanning is more sensitive than plain x-ray and continues to give positive results long after successful treatment.

Figure 157-1

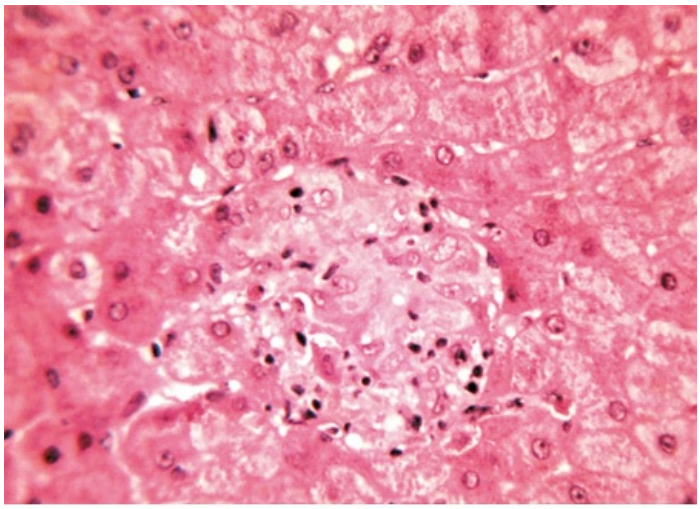

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Liver biopsy specimen from a patient with brucellosis shows a noncaseating granuloma. *[From Mandell's Atlas of Infectious Diseases, Vol II, in DL Stevens (ed): Skin, Soft Tissue, Bone and Joint Infections, Fig. 5-9; with permission.]*

Isolation of brucellae from blood, CSF, bone marrow, or joint fluid or from a tissue aspirate or biopsy sample is definitive, and attempts at isolation are usually successful in 50–70% of cases. Duplicate

cultures should be incubated for up to 6 weeks (in air and 10% CO_2, respectively). Concentration and lysis of buffy coat cells before culture may increase the isolation rate. Cultures in modern nonradiometric or similar signaling systems (e.g., Bactec) usually become positive within 7–10 days but should be maintained for at least 3 weeks before the results are declared negative. All cultures should be handled under containment conditions appropriate for dangerous pathogens. *Brucella* species may be misidentified as *Agrobacterium, Ochrobactrum*, or *Psychrobacter (Moraxella) phenylpyruvicus* by the gallery identification strips commonly used in the diagnostic laboratory.

The peripheral blood–based polymerase chain reaction has enormous potential to detect bacteremia, to predict relapse, and to exclude "chronic brucellosis." This method is probably more sensitive and is certainly quicker than blood culture, and it does not carry the attendant biohazard risk posed by culture. Nucleic acid amplification techniques are now quite widely used, although no single standardized procedure has been adopted. Primers for the spacer region between the genes encoding the 16S and 23S ribosomal RNAs (*rrs-rrl),* the outer-membrane protein Omp2*,* the insertion sequence *IS711*, and the protein BCSP31 are sensitive and specific. Blood and other tissues are the most suitable samples for analysis.

Serologic examination often provides the only positive laboratory findings in brucellosis. In acute infection, IgM antibodies appear early and are followed by IgG and IgA. All these antibodies are active in agglutination tests, whether performed by tube, plate, or microagglutination methods. The majority of patients have detectable agglutinins at this stage. As the disease progresses, IgM levels decline, and the avidity and subclass distribution of IgG and IgA change. The result is reduced or undetectable agglutinin titers. However, the antibodies are detectable by alternative tests, including the complement fixation test, Coomb's antiglobulin test, and enzyme-linked immunosorbent assay. There is no clear cutoff value for a diagnostic titer. Rather, serology results must be interpreted in the context of exposure history and clinical presentation. In endemic areas or in settings of potential occupational exposure, agglutinin titers of 1:320–1:640 or higher are considered diagnostic; in nonendemic areas, a titer of ≥1:160 is considered significant. Repetition of tests after 2–4 weeks may demonstrate a rising titer.

In most centers, the standard agglutination test (SAT) is still the mainstay of serologic diagnosis, although some investigators rely on the rose bengal test, which has not been fully validated for human diagnostic use. Dipstick assays for anti-*Brucella* IgM are useful for the diagnosis of acute infection but are less sensitive for infection with symptoms of several months' duration. In an endemic setting, >90% of patients with acute bacteremia have SAT titers of at least 1:320. Other screening tests are used in some centers.

Antibody to the *Brucella* LPS O chain—the dominant antigen—is detected by all the conventional tests that employ smooth *B. abortus* cells as antigen. Since *B. abortus* cross-reacts with *B. melitensis* and *B. suis*, there is no advantage in replicating the tests with these antigens. Cross-reactions also occur with the O chains of some other gram-negative bacteria, including *Escherichia coli* O157, *Francisella tularensis*, *Salmonella enterica* group N, *Stenotrophomonas maltophilia*, and *Vibrio cholerae*. Cross-reactions do not occur with the cell-surface antigens of rough *Brucella* strains such as *B. canis* or *B. ovis*; serologic tests for these nomen species must employ an antigen prepared from either one. The live *B. abortus* vaccine strain RB51 does not elicit responses in standard serologic testing. Most protein antigens are shared by all *Brucella* strains, and some are also common to *Ochrobactrum* species. Immunoblotting against protein extracts has been advocated as a differential test, but no validated procedure is yet available.

Treatment: Brucellosis

The broad aims of antimicrobial therapy are to treat and relieve the symptoms of current infection and to prevent relapse. Focal disease presentations may require specific intervention in addition to more prolonged and tailored antibiotic therapy. In addition, tuberculosis must always be excluded, or—to prevent the emergence of resistance—therapy must be tailored to specifically exclude drugs active against tuberculosis (e.g., rifampin used alone) or to include a full antituberculous regimen.

Early experience with streptomycin monotherapy showed that relapse was common; thus dual therapy with tetracyclines became the norm. This is still the most effective combination, but alternatives may be used, with the options depending on local or national policy about the use of rifampin for the treatment of nonmycobacterial infection. For the several antimicrobial agents that are active in vivo, efficacy can usually be predicted by in vitro testing. However, numerous Brucella strains show in vitro sensitivity to a whole range of antimicrobials that are therapeutically ineffective, including assorted β-lactams. Moreover, the use of fluoroquinolones remains controversial despite the good in vitro activity and white-cell penetration of most agents of this class. Low intravacuolar pH is probably a factor in the poor performance of these drugs.

For adults with acute nonfocal brucellosis (duration, <1 month), a 6-week course of therapy incorporating at least two antimicrobial agents is required. Complex or focal disease necessitates ≥3 months of therapy. Adherence to the therapeutic regimen is very important, and poor adherence underlies almost all cases of apparent treatment failure; such failure is rarely due to the emergence of

drug resistance, although increasing resistance to trimethoprim-sulfamethoxazole (TMP-SMX) has been reported at one center. There is good retrospective evidence that a 3-week course of two agents is as effective as a 6-week course for treatment and prevention of relapse in children, but this point has not yet been proven in prospective studies.

The gold standard for the treatment of brucellosis in adults is IM streptomycin (0.75–1 g daily for 14–21 days) together with doxycycline (100 mg twice daily for 6 weeks). In both clinical trials and observational studies, relapse follows such treatment in 5–10% of cases. The usual alternative regimen (and the current World Health Organization recommendation) is rifampin (600–900 mg/d) plus doxycycline (100 mg twice daily) for 6 weeks. The relapse/failure rate is ~10% in trial conditions but rises to >20% in many non-trial situations, possibly because doxycycline levels are reduced and clearance rates increased by concomitant rifampin administration. Patients who cannot tolerate or receive tetracyclines (children, pregnant women) can be given high-dose TMP-SMX instead (two or three standard-strength tablets twice daily for adults, depending on weight).

Increasing evidence supports the use of an aminoglycoside such as gentamicin (5–6 mg/kg per day for at least 2 weeks) instead of streptomycin. Shorter courses have been associated with high failure rates in adults. A 5- to 7-day course of therapy with gentamicin and a 3-week course of TMP-SMX may be adequate for children with uncomplicated disease, but prospective trials are still needed to support this recommendation. Early experience with fluoroquinolone monotherapy was disappointing, although it was suggested that ofloxacin or ciprofloxacin, given together with rifampin for 6 weeks, might be an acceptable alternative to the other 6-week regimens for adults. However, a meta-analysis did not support the use of fluoroquinolones in first-line treatment regimens, and these drugs are not recommended by an expert consensus group (the Ioannina Group) except in the context of well-designed clinical trials. A triple-drug regimen—doxycycline and rifampin combined with an initial course of an aminoglycoside—was superior to double-drug regimens in a meta-analysis. The triple-drug regimen should be considered for all patients with complicated disease and for those for whom treatment adherence is likely to be a problem.

Significant neurologic disease due to *Brucella* species requires prolonged treatment (i.e., for 3–6 months), usually with ceftriaxone supplementation of a standard regimen. *Brucella* endocarditis is treated with at least three drugs (an aminoglycoside, a tetracycline, and rifampin), and many experts add ceftriaxone and/or a fluoroquinolone to reduce the need for valve replacement. Treatment is usually given for at least 6 months, and clinical endpoints for its discontinuation are often difficult to define. Surgery is still required for the majority of cases of infection of prosthetic heart valves and prosthetic joints.

There is no evidence base to guide prophylaxis after exposure to *Brucella* organisms (e.g., in the laboratory), inadvertent immunization with live vaccine intended for use in animals, or exposure to deliberately released brucellae. Most authorities have recommended the administration of rifampin plus doxycycline for 3 weeks after a low-risk exposure (e.g., a nonspecific laboratory accident) and for 6 weeks after a major exposure to aerosol or injected material. However, such regimens are poorly tolerated, and doxycycline monotherapy of the same duration may be substituted. Rifampin should be omitted after exposure to vaccine strain RB51, which is resistant to rifampin but sensitive to doxycycline. After significant brucellosis exposure, expert consultation is advised for women who are (or may be) pregnant.

Prognosis and Follow-Up

Relapse occurs in up to 30% of poorly compliant patients. Thus patients should ideally be followed clinically for up to 2 years to detect relapse, which responds to a prolonged course of the same therapy used originally. The general well-being and the body weight of the patient are more useful guides than serology to lack of relapse. IgG antibody levels detected by the SAT and its variants can remain in the diagnostic range for >2 years after successful treatment. Complement fixation titers usually fall to normal within 1 year of cure. Immunity is not solid; patients can be reinfected after repeated exposures. Fewer than 1% of patients die of brucellosis. When the outcome is fatal, death is usually a consequence of cardiac involvement; more rarely, it results from severe neurologic disease. Despite the low mortality rate, recovery from brucellosis is slow, and the illness can cause prolonged inactivity, with domestic and economic consequences.

The existence of a prolonged chronic brucellosis state after successful treatment remains controversial. Evaluation of patients in whom this state is considered (often those with work-related exposure to brucellae) includes careful exclusion of malingering, nonspecific chronic fatigue syndromes, and other causes of excessive sweating, such as alcohol abuse and obesity. In the future, the availability of more sensitive assays to detect *Brucella* antigen or DNA may help to identify patients with ongoing infection.

Prevention

Vaccines based on live attenuated *Brucella* strains, such as *B. abortus* strain 19BA or 104M, have been used in some countries to protect high-risk populations but have displayed only short-term efficacy and high reactogenicity. Subunit vaccines have been developed but are of uncertain value

and cannot be recommended at present. Research in this area has been stimulated by interest in biodefense (Chap. 221) and may eventually yield new products, some of which may be based on the live attenuated WR 201 variant of *B. melitensis* strain 16M. The mainstay of veterinary prevention is a national commitment to testing and slaughter of infected herds/flocks (with compensation for owners), control of animal movement, and active immunization of animals. These measures are usually sufficient to control human disease as well. In their absence, pasteurization of all milk products before consumption is sufficient to prevent non-occupational animal-to-human transmission. All cases of brucellosis in animals and humans should be reported to the appropriate public health authorities.

FURTHER READINGS

Ariza J et al: Perspectives for the treatment of brucellosis in the 21st century: The Ioannina recommendations. PLoS Med 4:e317, 2007

Ashford DA et al: Adverse events in humans associated with accidental exposure to the livestock brucellosis vaccine RB51. Vaccine 22:3435, 2004[PMID: 15308369]

Barquero-Calvo E et al: *Brucella abortus* uses a stealth strategy to avoid activation of the innate immune system during the onset of infection. PLoS ONE 2:e631, 2007

Centers for Disease Control and Prevention: Laboratory-acquired brucellosis, Indiana and Minnesota, 2006. MMWR Morb Mortal Wkly Rep 57:39, 2008

Corbel MJ (ed): Brucellosis in humans and animals. Geneva, World Health Organization in collaboration with Food and Agriculture Organization and World Organization for Animal Health, 2006

———, Banai M: Genus *Brucella* Meyer and Shaw 1920,173AL, in *Bergey's Manual of Systematic Bacteriology*, 2nd ed, vol 2: *The Proteobacteria*, DJ Brenner et al (eds). New York, Springer, 2006, pp 370–386

Franco MP et al: Human brucellosis. Lancet Infect Dis 7:775, 2007[PMID: 18045560]

Pappas G et al: The new global map of human brucellosis. Lancet Infect Dis 6:91, 2006[PMID: 16439329]

Queipo-Ortuno MI et al: Usefulness of a quantitative real-time PCR assay using serum samples to discriminate between inactive, serologically positive and active human brucellosis. Clin Microbiol Infect 14:1128, 2008[PMID: 19046166]

Skalsky K et al: Treatment of human brucellosis: Systematic review and meta-analysis of randomised controlled trials. Br Med J 336:701, 2008[PMID: 18321957]

Whatmore AM: Current understanding of the genetic diversity of *Brucella*, an expanding genus of zoonotic pathogens. Infect Genet Evol 9:1168, 2009[PMID: 19628055]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TULAREMIA: INTRODUCTION

Tularemia is a zoonosis caused by *Francisella tularensis*. Humans of any age, sex, or race are universally susceptible to this systemic infection. Tularemia is primarily a disease of wild animals and persists in contaminated environments, ectoparasites, and animal carriers. Human infection is incidental and usually results from interaction with biting or blood-sucking insects, contact with wild or domestic animals, ingestion of contaminated water or food, or inhalation of infective aerosols. The illness is characterized by various clinical syndromes, the most common of which consists of an ulcerative lesion at the site of inoculation, with regional lymphadenopathy and lymphadenitis. Systemic manifestations, including pneumonia, typhoidal tularemia, meningitis, and fever without localizing findings, pose a greater diagnostic challenge.

Etiology and Epidemiology

Tularemia is common in Arkansas, Oklahoma, Missouri, and South Dakota; these states account for more than half of all reported cases in the United States. Small outbreaks in higher-risk populations (e.g., professional landscapers cutting up brush, mowing, and using a leaf blower) have been reported from the island of Martha's Vineyard in Massachusetts. Although the irregular distribution of cases of tularemia makes worldwide estimates difficult, increasing numbers of cases have been reported from the Scandinavian countries, Eastern Europe, and Siberia.

With rare exceptions, tularemia is the only disease produced by *F. tularensis*—a small (0.2 μm by 0.2–0.7 μm), gram-negative, pleomorphic, nonmotile, non-spore-forming bacillus. Bipolar staining results in a coccoid appearance. The organism is a thinly encapsulated, nonpiliated strict aerobe that invades host cells. In nature, *F. tularensis* is a hardy organism that persists for weeks or months in mud, water, and decaying animal carcasses. Dozens of biting and blood-sucking insects, especially ticks and tabanid flies, serve as vectors. Ticks and wild rabbits are the source for most human cases in endemic areas of the southeastern and Rocky Mountain states. In Utah, Nevada, and California, tabanid flies are the most common vectors. Animal reservoirs include wild rabbits, squirrels, birds, sheep, beavers, muskrats, and domestic dogs and cats. Person-to-person transmission is rare or nonexistent. Tularemia is more common among men than among women.

The four subspecies of *F. tularensis* are *tularensis*, *holarctica*, *novicida*, and *mediasiatica*. The first three of these subspecies are found in North America; in fact, subspecies *tularensis* has been isolated only in North America, where it accounts for >70% of cases of tularemia and produces more serious human disease than other subspecies (although, with treatment, the associated fatality rate is <2%). The progression of illness depends on the infecting strain's virulence, the inoculum size, the portal of entry, and the host's immune status. *F. tularensis* is a class A bioterrorism agent.

Ticks pass *F. tularensis* to their offspring transovarially. The organism is found in tick feces but not in large quantities in tick salivary glands. In the United States, the disease is carried by *Dermacentor andersoni* (Rocky Mountain wood tick), *D. variabilis* (American dog tick), *D. occidentalis* (Pacific Coast dog tick), and *Amblyomma americanum* (Lone Star tick). *F. tularensis* is transmitted frequently during blood meals taken by embedded ticks after hours of attachment. It is the taking of a blood meal through a fecally contaminated field that transmits the organism. Transmission by ticks and tabanid flies takes place mainly in the spring and summer. However, continued transmission in the winter by trapped or hunted animals has been documented.

Pathogenesis and Pathology

The most common portal of entry for human infection is through skin or mucous membranes, either directly—through the bite of ticks, other arthropods, or other animals—or via inapparent abrasions. Inhalation or ingestion of *F. tularensis* also can result in infection. *F. tularensis* is extremely infectious: Although $>10^8$ organisms are usually required to produce infection via the oral route (oropharyngeal or gastrointestinal tularemia), as few as 10 organisms can result in infection when injected into the skin (ulceroglandular/glandular tularemia) or inhaled (pulmonary tularemia). After inoculation into the skin, the organism multiplies locally; within 2–5 days (range, 1–10 days), it produces an erythematous, tender, or pruritic papule. The papule rapidly enlarges and forms an ulcer with a black base (chancriform lesion). The bacteria spread to regional lymph nodes, producing lymphadenopathy (buboes). All forms can lead to bacteremia with spread to distant organs, including the central nervous system.

Tularemia is characterized by mononuclear cell infiltration with pyogranulomatous pathology. The histopathologic findings can be quite similar to those in tuberculosis, although tularemia develops more rapidly. As a facultatively intracellular bacterium, *F. tularensis* can parasitize both phagocytic and nonphagocytic host cells and can survive intracellularly for prolonged periods. In the acute phase of infection, the primary organs affected (skin, lymph nodes, liver, and spleen) include areas of focal necrosis, which are initially surrounded by polymorphonuclear leukocytes (PMNs). Subsequently, granulomas form, with epithelioid cells, lymphocytes, and multinucleated giant cells surrounded by areas of necrosis. These areas may resemble caseation necrosis but later coalesce to form abscesses.

Conjunctival inoculation can result in infection of the eye, with regional lymph node enlargement (preauricular lymphadenopathy, Parinaud's complex). Aerosolization and inhalation or hematogenous spread of organisms can result in pneumonia. In the lung, an inflammatory reaction develops, including foci of alveolar necrosis and cell infiltration (initially polymorphonuclear and later mononuclear) with granulomas. Chest roentgenograms usually reveal bilateral patchy infiltrates rather than large areas of consolidation. Pleural effusions are common and may contain blood. Lymphadenopathy occurs in regions draining infected organs. Therefore, in pulmonary infection, mediastinal adenopathy may be evident, whereas patients with oropharyngeal tularemia develop cervical lymphadenopathy. In gastrointestinal or typhoidal tularemia, mesenteric lymphadenopathy may follow the ingestion of large numbers of organisms. (The term *typhoidal tularemia* may be used to describe severe bacteremic disease, irrespective of the mode of transmission or portal of entry.) Meningitis has been reported as a primary or secondary manifestation of bacteremia. Patients may also present with fever and no localizing signs.

Immunology

Although a complete and widely accepted understanding of the protective immune response to *F. tularensis* is lacking, significant advances in the study of natural and protective immunity have been made in recent years and may ultimately result in a vaccine candidate. The availability of attenuated *F. tularensis* strains developed through genetic manipulation is facilitating research that will expand our knowledge in this area.

A number of investigators have studied various models and proposed various hypotheses regarding the induction of protective immunity to *F. tularensis*. Although further research is needed, a synergy between humoral and cell-mediated immune (CMI) responses appears to be critical in inducing effective immune protection. Elucidation of the molecular mechanisms for the organism's evasion of the host response, pathogen-associated molecular patterns, and effective host immune protection has led to novel vaccination strategies tested in animal models. Antibodies to Fc receptors on antigen-presenting cells have been shown to be protective in animal models of pulmonary tularemia, resulting in both mucosal and CMI responses. This enhanced understanding of mucosal and serum antibodies in combination with a targeted CMI response holds great promise for future vaccine development.

Clinical Manifestations

Tularemia often starts with a sudden onset of fever, chills, headache, and generalized myalgias and arthralgias **(Table 158-1)**. This onset takes place when the organism penetrates the skin, is ingested, or is inhaled. An incubation period of 2–10 days is followed by the formation of an ulcer at the site of penetration, with local inflammation. The ulcer may persist for several months as organisms are transported via the lymphatics to the regional lymph nodes. These nodes enlarge and may become necrotic and suppurative. If the organism enters the bloodstream, widespread dissemination can result.

Table 158-1 Clinical Presentation of Tularemia

	Rate of Occurrence, %	
Sign or Symptom	**Children**	**Adults**
Lymphadenopathy	96	65
Fever (≥38.3°C or ≥101°F)	87	21
Ulcer/eschar/papule	45	51
Myalgias/arthralgias	39	2
Headache	9	5
Cough	9	5
Pharyngitis	43	—
Diarrhea	43	—

Source: Adapted from RF Jacobs, JP Narain: Pediatrics 76:818, 1985; with permission.

In the United States, most patients with tularemia (75–85%) acquire the infection by inoculation of the skin. In adults, the most common localized form is inguinal/femoral lymphadenopathy; in children, it is

cervical lymphadenopathy. About 20% of patients develop a generalized maculopapular rash, which occasionally becomes pustular. Erythema nodosum occurs infrequently. The clinical manifestations of tularemia have been divided into various syndromes, which are listed in **Table 158-2**.

Table 158-2 Clinical Syndromes of Tularemia

	Rate of Occurrence, %	
Syndrome	**Children**	**Adults**
Ulceroglandular	45	51
Glandular	25	12
Pulmonary (pneumonia)	14	18
Oropharyngeal	4	—
Oculoglandular	2	—
Typhoidal	2	12
Unclassified	6	11

Source: Adapted from RF Jacobs, JP Narain: Pediatrics 76:818, 1985; with permission.

ULCEROGLANDULAR/GLANDULAR TULAREMIA

These two forms of tularemia account for ~75–85% of cases. The predominant form in children involves cervical or posterior auricular lymphadenopathy and is usually related to tick bites on the head and neck. In adults, the most common form is inguinal/femoral lymphadenopathy resulting from insect and tick exposures on the lower limbs. In cases related to wild game, the usual portal of entry for *F. tularensis* is either an injury sustained while skinning or cleaning an animal carcass or a bite (usually on the hand). Epitrochlear lymphadenopathy/lymphadenitis is common in patients with bite-related injuries.

In ulceroglandular tularemia, the ulcer is erythematous, indurated, and nonhealing, with a punched-out appearance that lasts 1–3 weeks. The papule may begin as an erythematous lesion that is tender or pruritic; it evolves over several days into an ulcer with sharply demarcated edges and a yellow exudate. The ulcer gradually develops a black base, and simultaneously the regional lymph nodes become tender and severely enlarged **(Fig. 158-1)**. The affected lymph nodes may become fluctuant and drain spontaneously, but the condition usually resolves with effective treatment. Late suppuration of lymph nodes has been described in up to 25% of patients with ulceroglandular/glandular tularemia. Examination of material taken from these late fluctuant nodes after successful antimicrobial treatment reveals sterile necrotic tissue. In 5–10% of patients, the skin lesion may be inapparent, with lymphadenopathy plus systemic signs and symptoms the only physical findings (*glandular tularemia*). Conversely, a tick or deerfly bite on the trunk may result in an ulcer without evident lymphadenopathy.

Figure 158-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An 8-year-old boy with inguinal lymphadenitis and associated tick-bite site characteristic of ulceroglandular tularemia.

OCULOGLANDULAR TULAREMIA

In ~1% of patients, the portal of entry for *F. tularensis* is the conjunctiva, which the organism usually reaches through contact with contaminated fingers. The inflamed conjunctiva is painful, with numerous yellowish nodules and pinpoint ulcers. Purulent conjunctivitis with regional lymphadenopathy (preauricular, submandibular, or cervical) is evident. Because of debilitating pain, the patient may seek medical attention before regional lymphadenopathy develops. Painful preauricular lymphadenopathy is unique to tularemia and distinguishes it from tuberculosis, sporotrichosis, and syphilis. Corneal perforation may occur.

OROPHARYNGEAL AND GASTROINTESTINAL TULAREMIA

Rarely, tularemia follows ingestion of contaminated undercooked meat, oral inoculation of *F. tularensis* from the hands in association with the skinning and cleaning of animal carcasses, or consumption of contaminated food or water. Oral inoculation may result in acute, exudative, or membranous pharyngitis associated with cervical lymphadenopathy or in ulcerative intestinal lesions associated with mesenteric lymphadenopathy, diarrhea, abdominal pain, nausea, vomiting, and gastrointestinal bleeding. Infected tonsils become enlarged and develop a yellowish-white pseudomembrane, which can be confused with that of diphtheria. The clinical severity of gastrointestinal tularemia varies from mild, unexplained, persistent diarrhea with no other symptoms to a fulminant, fatal disease. In fatal cases, the extensive intestinal ulceration found at autopsy suggests an enormous inoculum.

PULMONARY TULAREMIA

Pneumonia due to *F. tularensis* presents as variable parenchymal infiltrates that are unresponsive to treatment with β-lactam antibiotics. Tularemia must be considered in the differential diagnosis of atypical pneumonia in a patient with a history of travel to an endemic area. The disease can result from inhalation of an infectious aerosol or can spread to the lungs and pleura via bacteremia. Inhalation-related pneumonia has been described in laboratory workers after exposure to contaminated materials and, if untreated, can be associated with a relatively high mortality rate. Exposure to *F. tularensis* in aerosols from live domestic animals or dead wildlife (including birds) has been reported to cause pneumonia. Hematogenous dissemination to the lungs occurs in 10–15% of cases of ulceroglandular tularemia and in about half of cases of typhoidal tularemia. Previously, tularemia pneumonia was thought to be a disease of older patients, but as many as 10–15% of children with clinical manifestations of tularemia have parenchymal infiltrates detected by chest roentgenography. Patients with pneumonia usually have a nonproductive cough and may have dyspnea or pleuritic chest pain. Roentgenograms of the chest usually reveal bilateral patchy infiltrates (described as ovoid or lobar densities), lobar parenchymal infiltrates, and cavitary lesions. Pleural effusions may have a predominance of mononuclear leukocytes or PMNs and sometimes red blood cells. Empyema may develop. Blood cultures may be positive for *F. tularensis*.

TYPHOIDAL TULAREMIA

The typhoidal presentation is now considered rare in the United States. The source of infection in typhoidal tularemia is usually associated with pharyngeal and/or gastrointestinal inoculation or bacteremic disease. Fever usually develops without apparent skin lesions or lymphadenopathy. Some patients have cervical and mesenteric lymphadenopathy. In the absence of a history of possible contact with a vector, diagnosis can be extremely difficult. Blood cultures may be positive and patients may present with classic sepsis or septic shock in this acute systemic form of the infection. Typhoidal tularemia is usually associated with a huge inoculum or with a preexisting compromising condition. High continuous fevers, signs of sepsis, and severe headache are common. The patient may be delirious and may develop prostration and shock. If presumptive antibiotic therapy in culture-negative cases does not include an aminoglycoside, the estimated mortality rate is relatively high.

OTHER MANIFESTATIONS

F. tularensis infection has been associated with meningitis, pericarditis, hepatitis, peritonitis, endocarditis, osteomyelitis, and sepsis and septic shock with rhabdomyolysis and acute renal failure. In cases of tularemia meningitis, a mean white blood cell count of 1788/μL, a predominantly mononuclear cell response (70–100%), a depressed glucose level, an elevated protein concentration, and a negative Gram's stain are typically found on examination of cerebrospinal fluid.

Differential Diagnosis

When patients in endemic areas present with fever, chronic ulcerative skin lesions, and large tender lymph nodes (Fig. 158-1), a diagnosis of tularemia should be made presumptively, and confirmatory diagnostic testing and appropriate therapy should be undertaken. When the possibility of tularemia is considered in a nonendemic area, an attempt should be made to identify contact with a potential animal vector. The level of suspicion should be especially high in hunters, trappers, game wardens, professional landscapers, veterinarians, laboratory workers, and individuals exposed to an insect or another animal vector. However, up to 40% of patients with tularemia have no known history of epidemiologic contact with an animal vector.

The characteristic presentation of ulceroglandular tularemia does not pose a diagnostic problem, but a less classic progression of regional lymphadenopathy or glandular tularemia must be differentiated from other diseases **(Table 158-3)**. The skin lesion of tularemia may resemble those seen in various other diseases but is generally accompanied by more impressive regional lymphadenopathy. In children, the differentiation of tularemia from cat-scratch disease is made more difficult by the chronic papulovesicular lesion associated with *Bartonella henselae* infection (Chap. 160). Oropharyngeal tularemia can resemble and must be differentiated from pharyngitis due to other bacteria or viruses. Pulmonary tularemia may resemble any atypical pneumonia. Typhoidal tularemia and tularemia meningitis may resemble a variety of other infections.

Table 158-3 Tularemia: Differential Diagnosis, by Clinical Disease Category

Glandular	Oropharyngeal	Typhoidal	Pulmonary
Pyogenic bacterial infection[a]	Group A streptococcal pharyngitis	Typhoid fever	*Mycoplasma pneumoniae* pneumonia
Nontuberculous mycobacterial infection	*Arcanobacterium haemolyticum* pharyngitis	Other *Salmonella* bacteremias	*Chlamydia pneumoniae* pneumonia
Sporotrichosis	Diphtheria	Rocky Mountain spotted fever	Psittacosis
Tuberculosis	Infectious mononucleosis	Human monocytotropic ehrlichiosis	*Legionella pneumophila* pneumonia
Syphilis	Various viral infections[b]	Human granulocytotropic anaplasmosis	Q fever
Anthrax		Infectious mononucleosis	Histoplasmosis
Rat-bite fever		Brucellosis	Blastomycosis
Scrub typhus		Toxoplasmosis	Coccidioidomycosis
Plague		Tuberculosis	Various viral infections[d]
Lymphogranuloma venereum		Sarcoidosis	
Cat-scratch disease		Malignancy[c]	

[a] *Staphylococcus aureus, Streptococcus pyogenes*.

[b] Adenovirus, enteroviruses, parainfluenza virus, influenza virus A and B, respiratory syncytial virus.

[c]Hematologic and reticuloendothelial malignancies.

[d]Influenza virus A and B, parainfluenza virus, respiratory syncytial virus, adenovirus, enteroviruses, hantavirus.

Laboratory Diagnosis

The diagnosis of tularemia is most frequently confirmed by agglutination testing. Microagglutination and tube agglutination are the techniques most commonly used to detect antibody to *F. tularensis*. In the standard tube agglutination test, a single titer of ≥1:160 is interpreted as a presumptive positive result. A fourfold increase in titer between paired serum samples collected 2–3 weeks apart is considered diagnostic. False-negative serologic responses are obtained early in infection; up to 30% of patients infected for 3 weeks have sera that test negative. Late in infection, titers into the thousands are common, and titers of 1:20–1:80 may persist for years. Enzyme-linked immunosorbent assays have proved useful for the detection of both antibodies and antigens.

Culture and isolation of *F. tularensis* are difficult. In one study, the organism was isolated in only 10% of more than 1000 human cases, 84% of which were confirmed by serology. The medium of choice is cysteine-glucose-blood agar. *F. tularensis* can be isolated directly from infected ulcer scrapings, lymph-node biopsy specimens, gastric washings, sputum, and blood cultures. Colonies are blue-gray, round, smooth, and slightly mucoid. On media containing blood, a small zone of α hemolysis usually surrounds the colony. Slide agglutination tests or direct fluorescent antibody tests with commercially available antisera can be applied directly to culture suspensions for identification. Most clinical laboratories will not attempt to culture *F. tularensis* because of the infectivity of the organism from the culture media and the consequent risk of a laboratory-acquired infection. Although tularemia is not spread from person to person, the organism can be inhaled from culture plates and infect unsuspecting laboratory workers. In most clinical laboratories, biosafety level 2 practices are recommended to handle clinical specimens thought to contain *F. tularensis*; however, biosafety level 3 conditions are required for procedures that produce aerosols or droplets during manipulation of cultures containing or possibly containing this organism.

A variety of polymerase chain reaction (PCR) methods have been used to detect *F. tularensis* DNA in many clinical specimens but mostly in ulceroglandular disease.. The majority of these methods target the genes encoding the outer-membrane proteins (e.g., *fopA* or *tul4*). A 16S rDNA sequence identification PCR may be helpful when the patient's clinical information does not lead the clinician to suspect a diagnosis of tularemia.

Treatment: Tularemia

Only aminoglycosides, tetracyclines, chloramphenicol, and rifampin are currently approved by the U.S. Food and Drug Administration for the treatment of tularemia. Gentamicin is considered the drug of choice for both adults and children. The dosage for adults is 5 mg/kg daily in two divided doses. The dosage for children is 2.5 mg/kg three times daily or 5 mg/kg twice daily. Gentamicin therapy is typically continued for 7–10 days; however, in mild to moderate cases of tularemia in which the patient becomes afebrile within the first 48–72 h of gentamicin treatment, a 5- to 7-day course has been successful.

If available, streptomycin given intramuscularly is also effective. The dosage for adults is 2 g/d in two divided doses. For children, the dosage is 30 mg/kg daily in two divided doses (maximal daily dose, 2 g). After a clinical response is demonstrated at 3–5 days, the dosage for children can be reduced to 10–15 mg/kg daily in two divided doses. The total duration of streptomycin therapy in both adults and children is usually 10 days. Unlike streptomycin and gentamicin, tobramycin is ineffective in the treatment of tularemia and should not be used.

Since doxycycline is bacteriostatic against *F. tularensis*, there is a risk of relapse if the patient is not treated for a long enough period. Therefore, if doxycycline is used, it should be given for at least 14 days. The lack of availability of chloramphenicol limits the utility of this agent as a viable treatment option. Fluoroquinolones—specifically, ciprofloxacin and levofloxacin—have been used with good outcomes to treat infections caused by subspecies *holarctica*, which is most often found in Europe. The lack of data on the efficacy of these agents against subspecies *tularensis* limits their use in North America at this time.

F. tularensis cannot be subjected to standardized antimicrobial susceptibility testing because the organism will not grow on the media used. A wide variety of antibiotics, including all β-lactam antibiotics and the newer cephalosporins, are ineffective for the treatment of tularemia. Several studies indicated that third-generation cephalosporins were active against *F. tularensis* in vitro, but clinical case reports suggested a nearly universal failure rate of ceftriaxone in pediatric patients with tularemia. Although in vitro data indicate that imipenem may be active, therapy with imipenem, sulfanilamides, and macrolides is not presently recommended because of the lack of relevant clinical data.

Virtually all strains of *F. tularensis* are susceptible to streptomycin and gentamicin. In successfully treated patients, defervescence usually occurs within 2 days, but skin lesions and lymph nodes may

take 1–2 weeks to heal. When therapy is not initiated within the first several days of illness, defervescence may be delayed. Relapses are uncommon with streptomycin or gentamicin therapy. Late lymph-node suppuration, however, occurs in ~40% of children, regardless of the treatment received. These nodes have typically been found to contain sterile necrotic tissue without evidence of active infection. Patients with fluctuant nodes should receive several days of antibiotic therapy before drainage to minimize the risk to hospital personnel.

Prognosis

If tularemia goes untreated, symptoms usually last 1–4 weeks but may continue for months. The mortality rate from severe untreated infection (including all cases of untreated pulmonary and typhoidal tularemia) can be as high as 30%. However, the overall mortality rate for untreated tularemia is <8%. With appropriate treatment, the mortality rate is <1%. Poor outcomes are often associated with long delays in diagnosis and treatment. Lifelong immunity usually follows tularemia.

Prevention

The prevention of tularemia is based on avoidance of exposure to biting and blood-sucking insects, especially ticks and deerflies. A wide range of approaches to vaccine development are being evaluated, but no vaccine against tularemia is yet licensed. Prophylaxis of tularemia has not proved effective in patients with embedded ticks or insect bites. However, in patients who are known to have been exposed to large quantities of organisms (e.g., in the laboratory) and who have incubating infection with *F. tularensis*, early treatment can prevent the development of significant clinical disease.

FURTHER READINGS

Barns SM et al: Detection of diverse new *Francisella*-like bacteria in environmental samples. Appl Environ Microbiol 71:5494, 2005[PMID: 16151142]

Centers for Disease Control and Prevention: Tularemia—Missouri, 2000–2007. MMWR Morb Mortal Wkly Rep 58:744, 2009

Eliasson H et al: The 2000 tularemia outbreak: A case-control study of risk factors in disease-endemic and emergent areas, Sweden. Emerg Infect Dis 8:956, 2002[PMID: 12194773]

Hofinger DM et al: Tularemic meningitis in the United States. Arch Neurol 66:523, 2009[PMID: 19364939]

IkÄheimo I et al: In vitro antibiotic susceptibility of *Francisella tularensis* isolated from humans and animals. J Antimicrob Chemother 46:287, 2000[PMID: 21677177]

Johansson A et al: In vitro susceptibility to quinolones of *Francisella tularensis* subspecies *tularensis*. Scand J Infect Dis 34:327, 2002[PMID: 12069013]

Jounio U et al: An outbreak of holarctica-type tularemia in pediatric patients. Pediatr Infect Dis J 29:160, 2010[PMID: 19918211]

Kirimanjeswara GS et al: Humoral and cell-mediated immunity to the intracellular pathogen *Francisella tularensis*. Immunol Rev 225:244, 2008[PMID: 18837786]

Oyston PCF et al: Tularaemia: Bioterrorism defense renews interest in *Francisella tularensis*. Nat Rev Microbiol 2:967, 2004[PMID: 15550942]

Sjostedt A: Tularemia: History, epidemiology, pathogen physiology, and clinical manifestations. Ann NY Acad Sci 1105:1, 2007[PMID: 17395726]

Staples JE et al: Epidemiology and molecular analysis of human tularemia, United States, 1964–2004. Emerg Infect Dis 12:1113, 2006[PMID: 16836829]

Tärvnik A et al: New approaches to diagnosis and therapy of tularemia. Ann NY Acad Sci 1105:378, 2007[PMID: 21698792]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PLAGUE AND OTHER *YERSINIA* INFECTIONS: INTRODUCTION

Plague

Plague is a systemic zoonosis caused by *Yersinia pestis*. It predominantly affects small rodents in rural areas of Africa, Asia, and the Americas and is usually transmitted to humans by an arthropod vector (the flea). Less often, infection follows contact with animal tissues or respiratory droplets. Plague is an acute febrile illness that is treatable with antimicrobial agents, but mortality rates among untreated patients are high. Patients can present with the bubonic, septicemic, or pneumonic form of the disease. Although there is concern among the general public about epidemic spread of plague by the respiratory route, this is not the usual route of plague transmission, and established infection-control measures for respiratory plague exist. However, the fatalities associated with plague and the capacity for infection via the respiratory tract mean that *Y. pestis* fits the profile of a potential agent of bioterrorism. Consequently, measures have been taken to restrict access to the organism, including legislation affecting diagnostic and research procedures in some countries (e.g., the United States).

Etiology

The genus *Yersinia* comprises gram-negative bacteria of the family Enterobacteriaceae (gamma proteobacteria). Overwhelming taxonomic evidence showing *Y. pestis* strains as a clonal group within *Y. pseudotuberculosis* suggests recent evolution from the latter organism—an enteric pathogen of mammals that is spread by the fecal-oral route and thus has a phenotype distinctly different from that of *Y. pestis*. When grown in vivo or at 37°C, *Y. pestis* forms an amorphous capsule made from a plasmid-specified fimbrial protein, Caf or fraction 1 (F1) antigen, which is an immunodiagnostic marker of infection.

Epidemiology

Human plague generally follows an outbreak in a host rodent population (epizootic). Mass deaths among the rodent primary hosts lead to a search by fleas for new hosts, with consequent incidental infection of other mammals. The precipitating cause for an epizootic may ultimately be related to climate or other environmental factors. The reservoir for *Y. pestis* causing enzootic plague in natural endemic foci between epizootics (i.e., when the organism may be difficult to detect in rodents or fleas) is a topic of ongoing research and may not be the same in all regions. The enzootic/epizootic pattern may be the result of complex dynamic interactions of host rodents that have different plague susceptibilities and different flea vectors; alternatively, an environmental reservoir may be important.

Global Features

In general, the enzootic areas for plague are lightly populated regions of Asia, Africa, and the Americas **(Fig. 159-1)**. Between 1989 and 2003, 38,359 cases of plague were reported to the World Health Organization (WHO) under the International Health Regulations then in force, which required national authorities to report all cases of plague in their jurisdiction and according to which plague was one of just three infectious diseases to be so reported. More than 80% of these cases were in Africa, and the percentage of all cases in Africa increased over this period; the majority of cases were reported from East Africa and the island of Madagascar. In 2007, the second edition of the International Health Regulations came into force, widening reporting to any disease event that could have a serious public health impact or could rapidly spread internationally. For plague, this requirement entails specific reporting of pneumonic plague or any suspected case of plague in an area not known to be endemic for plague. Recent outbreaks of pneumonic plague have been recorded in Uganda, the Democratic Republic of the Congo, and China.

Figure 159-1

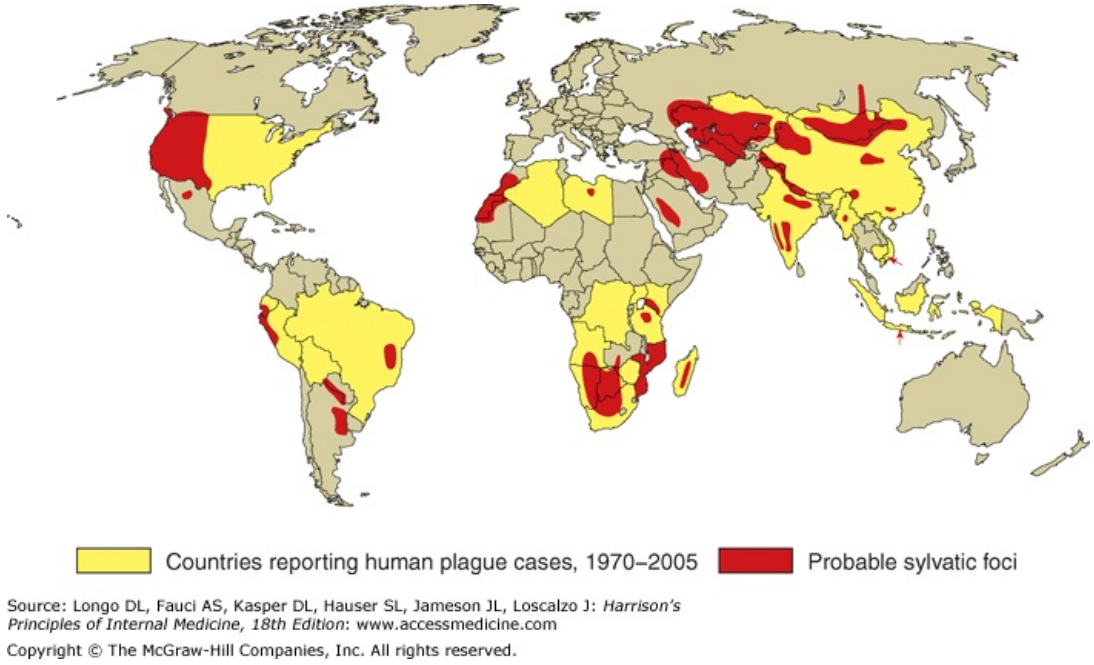

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approximate global distribution of *Yersinia pestis.* *(Compiled from WHO, CDC, and country sources. Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed, AS Fauci et al (eds). New York, McGraw-Hill, Chap. 152, 2008.)*

Plague was introduced into North America via the port of San Francisco in 1900 as part of the Third Pandemic, which spread around the world from Hong Kong. The disease is presently enzootic on the western side of the continent from southwestern Canada to Mexico. Most human cases in the United States occur in two regions: "Four Corners" (the junction point of New Mexico, Arizona, Colorado, and Utah), especially northern New Mexico, northern Arizona, and southern Colorado; and further west in California, southern Oregon, and western Nevada *(http://www.cdc.gov/ncidod/dvbid/plague/epi.htm)*. From 1990 to 2005, 107 cases of plague were reported in the United States, with a median of seven cases per year. Most cases occurred from May to October—the time of year when people are outdoors and rodents and their fleas are most plentiful. Infection is most often acquired by fleabite in peridomestic environments. Infection can also occur through the handling of living or dead small mammals (e.g., rabbits, hares, and prairie dogs) or wild carnivores (e.g., wildcats, coyotes, or mountain lions). Dogs and cats may bring plague-infected fleas into the home, and infected cats may transmit plague directly to humans by the respiratory route. The last recorded case of person-to-person transmission in the United States occurred in 1925.

Plague most often develops in areas with poor sanitary conditions and infestations of rats—in particular, the widely distributed roof rat *Rattus rattus* and the brown rat *R. norvegicus* (which serves as a laboratory model of plague). Rat control in warehouses and shipping facilities has been recognized as important in preventing the spread of plague since the early twentieth century and features in the current WHO International Health Regulations. Urban rodents acquire infection from wild rodents, and the proximity of the former to humans increases the risk of transmission. The oriental rat flea *Xenopsylla cheopis* is the most efficient vector for transmission of plague among rats and onward to humans in Asia, Africa, and South America.

Worldwide, bubonic plague is the predominant form reported (80–95% of suspected cases), with mortality rates of 10–20%. The mortality rate is higher (22%) in the small proportion of patients (10–20%) with primary septicemic plague (i.e., systemic *Y. pestis* sepsis with no bubo; see "Clinical Manifestations," below) and is highest with primary pulmonary plague; in this, the least common of the main plague presentations, mortality rate approaches 100% without antimicrobial treatment and is >50% even with such treatment. Rare outbreaks of pharyngeal plague following consumption of raw or undercooked camel or goat meat have been reported.

A total of 81 (76%) of the 107 plague cases reported in the United States from 1990 to 2005 were primary bubonic disease, 19 (18%) were primary septicemic disease, and 5 (5%) were primary pneumonic disease; 2 cases (2%) were not classified. Eleven cases (10%) were fatal.

Pathogenesis

As mentioned earlier, genetic evidence suggests that *Y. pestis* is a clone derived from the enteric pathogen *Y. pseudotuberculosis* in the recent evolutionary past (9000–40,000 years ago). The change from infection by the fecal-oral route to a two-stage life cycle, with alternate parasitization of arthropod and mammalian hosts, followed the acquisition of two plasmids—pFra and pPst—in conjunction with some adaptation of preexisting properties of the *Y. pseudotuberculosis* ancestor, including the

presence of a third plasmid, pYV. In the arthropod-parasitizing portion of its life cycle, *Y. pestis* multiplies and forms biofilm-embedded aggregates in the flea midgut after ingestion of a blood meal containing bacteria. In some fleas, biofilm-embedded bacteria eventually fill the proventriculus (a valve connecting the esophagus to the midgut) and block normal blood feeding. "Blocked" fleas die within a few days but in the interim make persistent efforts to feed, regurgitating esophageal contents and inoculating *Y. pestis* into each bite site. The ability of *Y. pestis* to colonize and multiply in the flea requires phospholipase D encoded by the *ymt* gene on the pFra plasmid, and biofilm synthesis requires the chromosomal *hms* locus shared with *Y. pseudotuberculosis*. Blockage takes days or weeks to come about after initial infection of the flea and is followed by the flea's death. Historically, blockage was thought to be required for efficient transmission, but many flea vectors (including *X. cheopis*) are, in fact, able to transmit plague in an early unblocked state or without blockage.

Y. pestis disseminates from the site of inoculation in the mammalian host in a process initially dependent on plasminogen activator Pla, which is encoded by the small pPst plasmid. This surface protease activates mammalian plasminogen, degrades complement, and adheres to the extracellular matrix component laminin. Pla is essential for the high-level virulence of *Y. pestis* in mice given an inoculum by subcutaneous or intradermal injection (laboratory proxies for fleabites) and for the development of primary pneumonic plague. When actual fleabite inoculation is used in mouse models, the fimbrial capsule-forming protein (Ca1 or fraction 1; F1 antigen) encoded on pFra increases the efficiency of transmission, and plasminogen activator is required for the formation of buboes. Because the antiphagocytic systems in *Y. pestis* are not fully operational at the time of inoculation into the mammalian host, the organism is taken up by macrophages from the inoculation site and transported to regional lymph nodes. After intracellular replication, *Y. pestis* switches to extracellular replication with full expression of its antiphagocytic systems: the type III secretion machines and their effectors encoded by pYV as well as the F1 capsule. Overproduction of the type III secretion substrate and translocation protein LcrV exerts an anti-inflammatory effect, reducing host immune responses. Likewise, *Y. pestis* lipopolysaccharide is modified to minimize stimulation of host Toll-like receptor 4, thereby reducing protective host inflammatory responses during peripheral infection and prolonging host survival with high-grade bacteremia—an effect that probably enhances the pathogen's subsequent transmission by fleabite.

Replication of *Y. pestis* in a regional lymph node results in the local swelling of the lymph node and periglandular region known as a *bubo*. On histology, the node is found to be hemorrhagic or necrotic, with thrombosed blood vessels, and the lymphoid cells and normal architecture are replaced by large numbers of bacteria and fibrin. Periglandular tissues are inflamed and also contain large numbers of bacteria in a serosanguineous, gelatinous exudate.

Continued spread through the lymphatic vessels to contiguous lymph nodes produces second-order primary buboes. Infection is initially contained in the infected regional lymph nodes, although transient bacteremia can be detected. As the infection progresses, spread via efferent lymphatics to the thoracic duct produces high-grade bacteremia. Hematogenous spread to the spleen, liver, and secondary buboes follows, with subsequent uncontrolled septicemia, endotoxic shock, and disseminated intravascular coagulation leading to death. In some patients, this septicemic phase occurs without obvious prior bubo development or lung disease (septicemic plague). Hematogenous spread to the lungs results in secondary plague pneumonia, with bacteria initially more prominent in the interstitium than in the air spaces (the reverse being the case in primary plague pneumonia). Hematogenous spread to other organs, including the meninges, can occur.

Clinical Manifestations

BUBONIC PLAGUE

After an incubation period of 2–6 days, the onset of bubonic plague is sudden and is characterized by fever (>38°C), malaise, myalgia, dizziness, and increasing pain due to progressive lymphadenitis in the regional lymph nodes near the fleabite or other inoculation site. Lymphadenitis manifests as a tense, tender swelling (bubo) that, when palpated, has a boggy consistency with an underlying hard core. Generally, there is one painful and erythematous bubo with surrounding periganglionic edema. The bubo is most commonly inguinal but can also be crural, axillary **(Fig. 159-2)**, cervical, or submaxillary, depending on the site of the bite. Abdominal pain from intraabdominal node involvement can occur without other visible signs. Children are most likely to present with cervical or axillary buboes.

Figure 159-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Plague patient in the southwestern United States with a left axillary bubo and an unusual plague ulcer and eschar at the site of the infective flea bite. *(Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed, AS Fauci et al (eds). New York, McGraw-Hill, Chap. 152, 2008.)*

The differential diagnosis includes acute focal lymphadenopathy of other etiologies, such as streptococcal or staphylococcal infection, tularemia, cat-scratch disease, tick typhus, infectious mononucleosis, or lymphatic filariasis. These infections do not progress as rapidly, are not as painful, and are associated with visible cellulitis or ascending lymphangitis—both of which are absent in plague.

Without treatment, *Y. pestis* dissemination occurs and causes serious illness, including pneumonia (secondary pneumonic plague) and meningitis. Secondary pneumonic plague can be the source of person-to-person transmission of respiratory infection by productive cough (droplet infection), with the consequent development of primary plague pneumonia. Appropriate treatment of bubonic plague results in fever resolution within 2–5 days, but buboes may remain enlarged for >1 week after initial treatment and can become fluctuant.

PRIMARY SEPTICEMIC PLAGUE

A minority (10–25%) of infections with *Y. pestis* present as gram-negative septicemia (hypotension, shock) without preceding lymphadenopathy. Septicemic plague occurs in all age groups, but persons older than age 40 years are at elevated risk. The term *septicemic plague* can be confusing since most patients with buboes have detectable bacteremia at some stage, with or without systemic signs of sepsis. In laboratory experiments, however, septicemic disease without histologic changes in lymph nodes is seen in a minority of mice infected via fleabites.

PNEUMONIC PLAGUE

Primary pneumonic plague results from inhalation of infectious bacteria in droplets expelled from another person or an animal with primary or secondary plague pneumonia. This syndrome has a short incubation period, averaging from a few hours to 2–3 days (range, 1–7 days), and is characterized by a sudden onset of fever, headache, myalgia, weakness, nausea, vomiting, and dizziness. Respiratory signs—cough, dyspnea, chest pain, and sputum production with hemoptysis—typically arise after 24 h. Progression of initial segmental pneumonitis to lobar pneumonia and then to bilateral lung involvement may occur **(Fig. 159-3)**. The possible release of aerosolized *Y. pestis* bacteria in a bioterrorist attack, manifesting as an outbreak of primary pneumonic plague in nonendemic regions or in an urban setting where plague is rarely seen, has been a source of public health concern. Secondary pneumonic plague is a consequence of bacteremia occurring in ~10–15% of patients with bubonic plague. Bilateral alveolar infiltrates are seen on chest X-ray, and diffuse interstitial pneumonitis with scanty sputum production is typical.

Figure 159-3

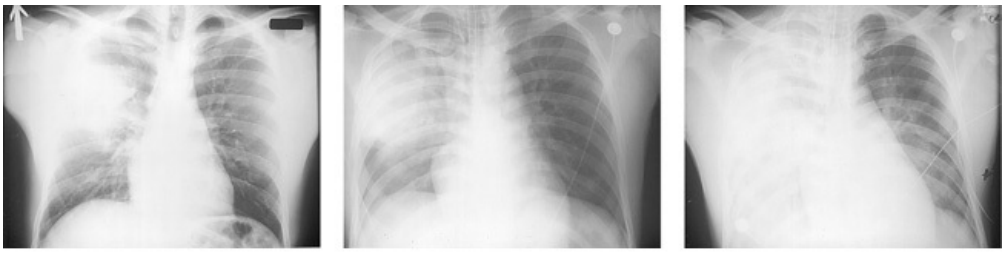

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sequential chest radiographs of a patient with fatal primary plague pneumonia. ***Left:*** Upright posteroanterior film taken at admission to hospital emergency department on third day of illness, showing segmental consolidation of right upper lobe. ***Center:*** Portable anteroposterior film taken 8 h after admission, showing extension of pneumonia to right middle and right lower lobes. ***Right:*** Portable anteroposterior film taken 13 h after admission (when patient had clinical adult respiratory distress syndrome), showing diffuse infiltration throughout right lung and patchy infiltration of left lower lung. A cavity later developed at the site of initial right-upper-lobe consolidation. *(Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed. AS Fauci et al (eds). New York, McGraw-Hill, Chap. 152, 2008.)*

MENINGITIS

Meningeal plague is uncommon, occurring in ≥6% of plague cases reported in the United States. Presentation with headache and fever typically occurs >1 week after the onset of bubonic or septicemic plague and may be associated with suboptimal antimicrobial therapy (delayed therapy, penicillin administration, or low-dose tetracycline treatment) and cervical or axillary buboes.

PHARYNGITIS

Symptomatic plaguepharyngitis can follow the consumption of contaminated meat from an animal dying of plague or contact with persons or animals with pneumonic plague. This condition can resemble tonsillitis, with peritonsillar abscess and cervical lymphadenopathy. Asymptomatic pharyngeal carriage of *Y. pestis* can also occur in close contacts of patients with pneumonic plague.

Laboratory Diagnosis

Because of the scarcity of laboratory facilities in regions where human *Y. pestis* infection is most common, and because of the potential significance of *Y. pestis* isolation in a nonendemic area or an area from which human plague has been absent for many years, the WHO recommends an initial presumptive diagnosis followed by reference laboratory confirmation **(Table 159-1)**. In the United States, comprehensive national diagnostic facilities for plague have been in place since a federal Laboratory Response Network (LRN; *www.bt.cdc.gov/lrn/*) was set up in 1999 to detect possible use of biological terrorism agents, including *Y. pestis*. Routine diagnostic clinical microbiology laboratories that are included in this network as sentinel-level laboratories use joint protocols from the Centers for Disease Control and Prevention (CDC) and the American Society for Microbiology to identify suspected *Y. pestis* isolates and to refer these specimens to LRN reference laboratories for confirmatory tests. *Y. pestis* is designated a "select agent" under the Public Health Security and Bioterrorism Preparedness and Response Act of 2002; the provisions of this act and of the Patriot Act of 2001 apply to all U.S. laboratories and individuals working with *Y. pestis*. Details of the applicable regulations are available from the CDC.

Table 159-1 World Health Organization Case Definitions of Plague

Suspected case	Compatible clinical presentation ***and*** Consistent epidemiologic features, such as exposure to infected animals or humans and/or evidence of fleabites and/or residence in or travel to a known endemic focus within the previous 10 days
Presumptive case	Meeting the definition of a suspected case ***plus***: **Putative new or reemerging focus**: ≥2 of the following tests positive • Microscopy: gram-negative coccobacilli in material from bubo, blood, or sputum; bipolar appearance on Wayson or Wright-Giemsa staining • F1 antigen detected in bubo aspirate, blood, or sputum

	• A single anti-F1 serology without evidence of previous *Y. pestis* infection or immunization • PCR detection of *Y. pestis* in bubo aspirate, blood, or sputum **Known endemic focus**: ≥1 of the following tests positive • Microscopic evidence of gram-negative or bipolar (Wayson, Wright-Giemsa) coccobacilli from bubo, blood, or sputum sample • A single anti-F1 serology without evidence of previous plague infection or immunization • F1 antigen detected in bubo aspirate, blood, or sputum • PCR detection of *Y. pestis* in bubo aspirate, blood, or sputum
Confirmed case	Meeting the definition of a suspected case ***plus***: • Identification of an isolate from a clinical sample as *Y. pestis* (colonial morphology and 2 of the 4 following tests positive: phage lysis of cultures at 20–25°C and 37°C; F1 antigen detection; PCR; *Y. pestis* biochemical profile) *or* • A fourfold rise in anti-F1 antibody titer in paired serum samples *or* • In endemic areas when no other confirmatory test can be performed, a positive rapid diagnostic test with immunochromatography to detect F1 antigen

Source: Interregional Meeting on Prevention and Control of Plague, Antananarivo, Madagascar, 7–11 April 2006 *(www.who.int/entity/csr/resources/publications/WHO_HSE_EPR_2008_3w.pdf).*

Yersinia species are gram-negative coccobacilli (short rods with rounded ends) 1–3 μm in length and 0.5–0.8 μm in diameter. *Y. pestis* in particular appears bipolar (with a "closed safety pin" appearance) and pleomorphic when stained with a polychromatic stain (Wayson or Wright-Giemsa; **Fig. 159-4**). Its lack of motility distinguishes *Y. pestis* from other *Yersinia* species, which are motile at 25°C and nonmotile at 37°C. Transport medium (e.g., Cary-Blair medium) preserves the viability of *Y. pestis* if transport is delayed.

Figure 159-4

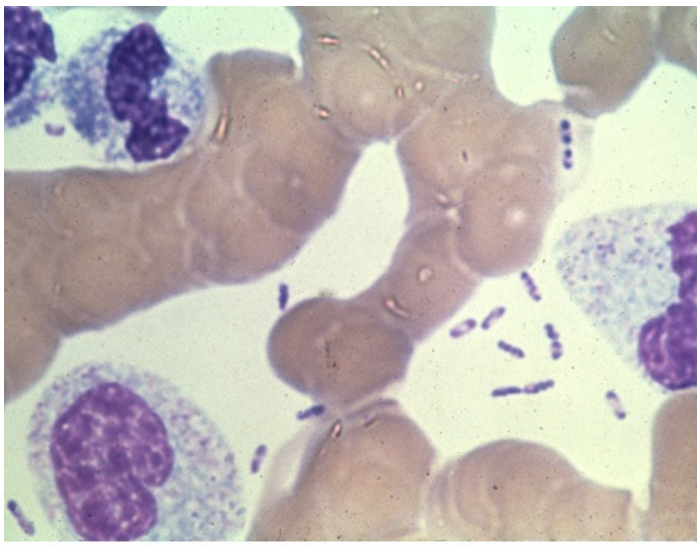

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Peripheral blood smear from a patient with fatal plague septicemia and shock, showing characteristic bipolar-staining *Yersinia pestis* bacilli (Wright's stain, oil immersion). *(Reprinted with permission from DT Dennis, GL Campbell: Plague and other Yersinia infections, in Harrison's Principles of Internal Medicine, 17th ed, AS Fauci et al (eds). New York, McGraw-Hill, Chap. 152, 2008.)*

The appropriate specimens for diagnosis of bubonic, pneumonic, and septicemic plague are bubo aspirate, bronchoalveolar lavage fluid or sputum, and blood, respectively. Culture of postmortem organ

biopsy samples can also be diagnostic. A bubo aspirate is obtained by injection of 1 mL of sterile normal saline into a bubo under local anesthetic and aspiration of a small amount of (usually blood-stained) fluid. Gram's staining of these specimens may reveal gram-negative rods, which are shown by Wayson or Wright-Giemsa staining to be bipolar. These bacteria may even be visible in direct blood smears in septicemic plague (Fig. 159-4); this finding indicates very high numbers of circulating bacteria and a poor prognosis.

Y. pestis grows on nutrient agar and other standard laboratory media but forms smaller colonies than do other Enterobacteriaceae. Specimens should be inoculated onto nutrient-rich media such as sheep blood agar (SBA), into nutrient-rich broth such as brain-heart infusion broth, and onto selective agar such as MacConkey or eosin methylene blue (EMB) agar. *Yersinia*-specific CIN [cefsulodin, triclosan (Irgasan), novobiocin] agar can be useful for culture of contaminated specimens, such as sputum. Blood should be cultured in a standard blood culture system. The optimal growth temperature is <37°C (25–29°C), with pinpoint colonies only on SBA at 24 h. Slower growth occurs at 37°C. *Y. pestis* is oxidase-negative, catalase-positive, urea-negative, indole-negative, and lactose-negative. Automated biochemical identification systems can misidentify *Y. pestis* as *Y. pseudotuberculosis* or other bacterial species.

Reference laboratory tests for definitive identification of isolates include direct immunofluorescence for F1 antigen; specific polymerase chain reaction (PCR) for targets such as F1 antigen, the pesticin gene, and the plasminogen activator gene; and specific bacteriophage lysis. PCR can also be applied to diagnostic specimens, as can direct immunofluorescence for F1 antigen (produced in large amounts by *Y. pestis*) by slide microscopy. An immunochromatographic test strip for F1 antigen detection by monoclonal antibodies in clinical specimens has been devised in Madagascar. This method is effective for both laboratory and near-patient use but is not yet commercially available. Several other rapid diagnostic kits for possible bioterrorism pathogens, including *Y. pestis*, have been described in recent years, but none is widely used for primary or reference laboratory identification.

In the absence of other positive laboratory diagnostic tests, a retrospective serologic diagnosis may be made on the basis of rising titers of hemagglutinating antibody to F1 antigen. Enzyme-linked immunosorbent assays (ELISAs) for IgG and IgM antibodies to F1 antigen are also available.

The white blood cell (WBC) count is generally raised (to 10,000–20,000/μL) in plague, with neutrophilic leukocytosis and a left shift (numerous immature neutrophils); in some cases, however, the WBC count is normal or leukopenia develops. WBC counts are occasionally very high, especially in children (>100,000/μL). Levels of fibrinogen degradation products are elevated in a majority of patients, but platelet counts are usually normal or low-normal. However, disseminated intravascular coagulation, with low platelet counts, prolonged prothrombin times, reduced fibrinogen, and elevated fibrinogen degradation product levels, occurs in a significant minority of patients.

Treatment: Plague

Guidelines for the treatment of plague are given in **Table 159-2**. A 10-day course of antimicrobial therapy is recommended. Streptomycin has historically been the parenteral treatment of choice for plague and is approved for this indication by the U.S. Food and Drug Administration (FDA). Although not yet approved by the FDA for plague, gentamicin has proven safe and effective in clinical trials in Tanzania and Madagascar and in retrospective reviewed cases in the United States. In view of streptomycin's adverse-reaction profile and limited availability, some experts now recommend gentamicin over streptomycin. Likewise, while systemic chloramphenicol therapy is available in the resource-poor countries primarily affected by plague, it is less likely to be available or used in high-income countries because of its adverse-effect profile. Tetracyclines are also effective and can be given by mouth but are not recommended for children under age 7 years because of tooth discoloration. Doxycycline is the tetracycline of choice; at an oral dosage of 100 mg twice daily, this drug was as effective as intramuscular gentamicin (2.5 mg/kg twice daily) in a trial in Tanzania.

Table 159-2 Guidelines for the Treatment of Plague

Drug	Daily Dose	Interval, h	Route
Gentamicin			
Adult	5 mg/kg[a]	24	IM/IV
	3–5 mg/kg	8 (2-mg/kg loading dose followed by 1.7 mg/kg tid, then by 1 mg/kg tid as soon as clinically indicated)	IM/IV
Child	5 mg/kg[a]	24	IM/IV
	7.5 mg/kg	8 (2.5 mg/kg tid)	IM/IV

Streptomycin			
Adult	2 g	12	IM
Child	30 mg/kg	12	IM
Doxycycline			
Adult	200 mg	12 or 24	PO/IV
Child ≥8 years	4.4 mg/kg	12 or 24	PO/IV
Tetracycline			
Adult	2 g	6	PO/IV
Child ≥8 years	25–50 mg/kg	6	PO/IV
Chloramphenicol			
Adult	50 mg/kg	6	PO/IV
Child ≥1 year	50 mg/kg	6	PO/IV

[a]The dose must be adjusted in cases of reduced renal function. There are no published trial data for once-daily gentamicin as therapy for plague in adults or children, but this regimen is efficacious in sepsis of other gram-negative etiologies and has been successful in a recent outbreak of pneumonic plague in the Democratic Republic of the Congo. Neonates up to 1 week of age and premature infants should receive 2.5 mg/kg IV bid.

Source: Inglesby et al, 2000.

Although *Y. pestis* is sensitive to β-lactam drugs in vitro and these drugs have been efficacious against plague in some animal models, the response to penicillins has been poor in some clinical cases; thus β-lactams and macrolides are not generally recommended as first-line therapy. Chloramphenicol, alone or in combination, is recommended for some focal complications of plague (e.g., meningitis, endophthalmitis, myocarditis) because of its tissue penetration properties. Fluoroquinolones, which have been effective in vitro and in animal models, are recommended in guidelines for possible bioterrorism-associated pneumonic plague and are increasingly used in therapy, although the only human efficacy data available so far are from a case report. Animal and in vitro studies suggest that fluoroquinolones at doses used in systemic gram-negative sepsis should be effective as therapy for plague: e.g., ciprofloxacin (400 mg twice daily IV, 500 mg twice daily by mouth), levofloxacin (500 mg/d IV or by mouth), ofloxacin (400 mg twice daily IV or by mouth), or moxifloxacin (400 mg/d IV or by mouth).

Prevention

In endemic areas, the control of plague in humans is based on reduction of the likelihood of being bitten by infected fleas or exposed to infected droplets from either humans or animals with plague pneumonia. In the United States, residence and outdoor activity in rural areas of western states where epizootics occur are the main risk factors for infection. To assess potential risks to humans in specific areas, surveillance for *Y. pestis* infection among animal plague hosts and vectors is carried out regularly as well as in response to observed animal die-offs. Personal protective measures include avoidance of areas where a plague epizootic has been identified and publicized (e.g., by warning signs or closure of campsites). Sick or dead animals should not be handled by the general public. Hunters and zoologists should wear gloves when handling wild animal carcasses in endemic areas. General measures to avoid rodent fleabite during outdoor activity are appropriate and include the use of insect repellant, insecticide, and protective clothing. General measures to reduce peridomestic and occupational human contact with rodents are advised and include rodent-proofing of buildings and food waste stores and removal of potential rodent habitats (e.g., woodpiles and junk heaps). Flea control by insecticide treatment of wild rodents is an effective means of minimizing human contact with plague if an epizootic is identified in an area close to human habitation. Any attempt to reduce rodent numbers must be preceded by flea suppression to reduce the migration of infected fleas to human hosts.

Patients in whom pneumonic plague is suspected should be managed in isolation, with droplet precautions observed until pneumonia is excluded or effective antimicrobial therapy has been given for 48 h. Review of the literature published before the advent of antimicrobial agents suggests that the main infective risk is posed by patients in the final stages of disease who are coughing up sputum with plentiful visible blood and/or pus. Cotton and gauze masks were protective in these circumstances. Current surgical masks capable of barrier protection against droplets, including large respiratory particles, are considered protective; a particulate respirator (e.g., N95 or greater) is not required.

ANTIMICROBIAL PROPHYLAXIS

Postexposure antimicrobial prophylaxis lasting 7 days is recommended following household, hospital, or other close contact with persons with untreated pneumonic plague. (*Close contact* is defined as contact with a patient at <2 m.) Doxycycline is probably the first choice for prophylaxis **(Table 159-3)**.

Table 159-3 Guidelines for Plague Prophylaxis

Drug	Daily Dose	Interval, h	Route
Doxycycline			
Adult	200 mg	12 or 24	PO
Child ≥8 y	If weight is ≥45 kg, give adult dosage; if <45 kg, give 2.2 mg/kg PO bid (maximum, 200 mg/d)	12	PO
Tetracycline			
Adult	1–2 g	6 or 12	PO
Child ≥8 y	25–50 mg/kg	6 or 12	PO
Ciprofloxacin			
Adult	1 g	12	PO
Child	40 mg/kg	12	PO
Trimethoprim-sulfamethoxazole			
Adult	320 mg[a]	12	PO
Child ≥2 months	8 mg/kg[a]	12	PO

[a]Trimethoprim component.

Source: DT Dennis, GL Campbell: Plague and other *Yersinia* infections, in *Harrison's Principles of Internal Medicine,* 17th ed, AS Fauci et al (eds). New York, McGraw-Hill, Chap. 152, 2008.

IMMUNIZATION

Studies with candidate plague vaccines in animal models show that neutralizing antibody provides protection against exposure but that cell-mediated immunity is critical for protection and clearance of *Y. pestis* from the host. A killed whole-cell vaccine used in humans required multiple doses, caused significant local and systemic reactions, and was not protective against pneumonic plague; this vaccine is not currently available in the United States. A live attenuated vaccine based on strain EV76 is still used in countries of the former Soviet Union but has significant side effects. Most research to date has focused on a subunit vaccine of recombinant F1 (rF1) and V (rV) proteins produced in *Escherichia coli*, combined either as a fusion protein or as a mixture, purified, and adsorbed to aluminum hydroxide for injection. This combination protects mice against both bubonic and pneumonic plague and has been evaluated in phase 2 clinical trials. Phase 3 trials are planned, but special ethical considerations with controlled clinical studies involving plague in humans make field efficacy studies unlikely. In the United States, the FDA is therefore prepared to assess this vaccine for human use under the Animal Rule, using efficacy data and other results from animal studies *(www.fda.gov/BiologicsBloodVaccines/ScienceResearch/BiologicsResearchAreas/ucm127288.htm)*. Future developments may include noninvasive vaccines (with intranasal administration or inhalation of dry powder incorporating these or other antigens) and mucosal delivery of microencapsulated proteins. Other antigens that could be added to this type of subunit vaccine are being investigated. Providing impetus for exploration of these antigens are (1) the recovery of F1-negative *Y. pestis* strains from natural sources and (2) the observation that F1 antigen is not required for virulence in primate models of pneumonic plague.

Yersiniosis

Yersiniosis is a zoonotic infection with an enteropathogenic *Yersinia* species, usually *Y. enterocolitica* or *Y. pseudotuberculosis*. The usual hosts for these organisms are pigs and other wild and domestic animals; humans are usually infected by the oral route, and outbreaks from contaminated food occur. Yersiniosis is most common in childhood and in colder climates. Patients present with abdominal pain and sometimes with diarrhea (which may not occur in up to 50% of cases). *Y. enterocolitica* is more closely associated with terminal ileitis and *Y. pseudotuberculosis* with mesenteric adenitis, but both organisms may cause mesenteric adenitis and symptoms of abdominal pain and tenderness that result in pseudoappendicitis, with the surgical removal of a normal appendix. Diagnosis is based on culture of the organism or convalescent serology. *Y. pseudotuberculosis* and some rarer strains of *Y. enterocolitica* are especially likely to cause systemic infection, which is also more likely in patients with diabetes or iron overload. Systemic sepsis is treatable with antimicrobial agents, but postinfective arthropathy responds poorly to such therapy. Ten other *Yersinia* species are now recognized, but all

lack the virulence plasmid pYV common to *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* and are generally considered to be, at most, opportunistic pathogens of humans (*Y. aldovae*, *Y. bercovieri*, *Y. frederiksenii*, *Y. intermedia*, *Y. kristensenii*, *Y. massiliensis*, *Y. mollaretii*, *Y. rohdei*, *Y. similis*, and *Y. ruckeri*).

Epidemiology

Y. ENTEROCOLITICA

Y. enterocolitica is found worldwide and has been isolated from a wide variety of wild and domestic animals and environmental samples, including samples of food and water. Strains are differentiated by combined biochemical reactions (biovar) and serogroup. Most clinical infections are associated with serogroups O:3, O:9, and O:5,27, with a declining number of O:8 infections in North America. Yersiniosis, mostly due to *Y. enterocolitica*, is the third commonest zoonosis reported in Europe; most reports come from northern Europe, especially Germany and Scandinavia. The incidence is highest among children; children under the age of 4 years are more likely to present with diarrhea than are older children. Abdominal pain with mesenteric adenitis and terminal ileitis is more prominent among older children and adults. Septicemia is more likely in patients with preexisting conditions such as diabetes mellitus, liver disease, any condition involving iron overload (including thalassemia and hemochromatosis), advanced age, malignancy, or HIV/AIDS. As in enteritis of other bacterial etiologies, postinfective complications such as reactive arthritis occur mainly in individuals who are HLA-B27 positive. Erythema nodosum (Fig. e7-40) following *Yersinia* infection is not associated with HLA-B27 and is more common among women than among men.

Consumption or preparation of raw pork products (such as chitterlings) and some processed pork products is strongly linked with infection because a high percentage of pigs carry pathogenic *Y. enterocolitica* strains. Outbreaks of *Y. enterocolitica* infection have been associated with consumption of milk (pasteurized, unpasteurized, and chocolate-flavored) and various foods contaminated with springwater. Person-to-person transmission is suspected in a few cases (e.g., in nosocomial and familial outbreaks) but is much less likely with *Y. enterocolitica* than with other causes of gastrointestinal infection, such as *Salmonella*. A multivariate analysis indicates that contact with companion animals is a risk factor for Y. enterocolitica infection among children in Sweden, and low-level colonization of dogs and cats with *Y. enterocolitica* has been reported. Transfusion-associated septicemia due to *Y. enterocolitica*, while recognized as a rare but frequently fatal event for nearly 30 years, has been difficult to eradicate.

Y. PSEUDOTUBERCULOSIS

Y. pseudotuberculosis is less frequently reported as a cause of human disease than *Y. enterocolitica*, and infection with *Y. pseudotuberculosis* is more likely to present as fever and abdominal pain due to mesenteric lymphadenitis. This organism is associated with wild mammals (rodents, rabbits, and deer), birds, and domestic pigs. Strains are differentiated by combined biochemical reactions (biovar) and serogroup. Although outbreaks are generally rare, several have recently occurred in Finland and have been associated with lettuce and raw carrots.

Pathogenesis

The usual route of infection is oral. Studies with both *Y. enterocolitica* and *Y. pseudotuberculosis* in animal models suggest that initial replication in the small intestine is followed by invasion of Peyer's patches of the distal ileum via M cells, with onward spread to mesenteric lymph nodes. The liver and spleen can also be involved after oral infection. The characteristic histologic appearance of enteropathogenic *Yersinia* after invasion of host tissues is as extracellular microabscesses surrounded by an epithelioid granulomatous lesion.

Experiments involving oral infection of mice with tagged *Y. enterocolitica* show that only a very small proportion of bacteria in the gut invade tissues. Individual bacterial clones from an orally inoculated pool give rise to each microabscess in a Peyer's patch, and the host restricts the invasion of previously infected Peyer's patches. A prior model positing progressive bacterial spread from Peyer's patches and mesenteric lymph nodes to the liver and spleen appears to be inaccurate: spread of individually tagged clones of *Y. pseudotuberculosis* to the liver and spleen of mice occurs independently of regional lymph node colonization and in mice lacking Peyer's patches.

Invasion requires the expression of several nonfimbrial adhesins, such as invasin (Inv) and—in *Y. pseudotuberculosis*—*Yersinia* adhesin A (YadA). Inv interacts directly with β1 integrins, which are expressed on the apical surfaces of M cells but not enterocytes. YadA of *Y. pseudotuberculosis* interacts with extracellular matrix proteins such as collagen and fibronectin to facilitate host cell integrin association and invasion. YadA of *Y. enterocolitica* lacks a crucial N-terminal region and binds collagen and laminin but not fibronectin and does not cause invasion. Inv is chromosomally encoded, whereas YadA is encoded on the virulence plasmid pYV. YadA helps to confer serum resistance by binding host complement regulators such as factor H and C4-binding protein. Another chromosomal gene, *ail* (attachment and invasion locus), encodes the extracellular protein Ail, which also confers serum resistance by binding these complement regulators.

By binding to host cell surfaces, YadA allows targeting of immune effector cells by the pYV plasmid–encoded type III secretion system (injectisome). As a consequence, the host's innate immune response is altered; toxins (*Yersinia* outer proteins, or Yops) are injected into host macrophages, neutrophils, and dendritic cells, affecting signal transduction pathways, resulting in reduced phagocytosis and inhibited production of reactive oxygen species by neutrophils, and triggering apoptosis of macrophages. Other factors functional in invasive disease include yersiniabactin (Ybt), a siderophore produced by some strains of *Y. pseudotuberculosis* and *Y. enterocolitica* as well as other Enterobacteriaceae. Yersiniabactin allows bacteria to access iron from saturated lactoferrin during infection and reduces production of reactive oxygen species by innate immune effector cells, thereby decreasing bacterial killing.

Clinical Manifestations

Self-limiting diarrhea is the most common reported presentation in infection with pathogenic *Y. enterocolitica*, especially in children under the age of 4, who form the single largest group in most case series. Blood may be detected in diarrheal stool. Older children and adults are more likely than younger children to present with abdominal pain, which can be localized to the right iliac fossa—a situation that often leads to laparotomy for presumed appendicitis (pseudoappendicitis). Appendectomy is not indicated for *Yersinia* infection causing pseudoappendicitis. Thickening of the terminal ileum and cecum is seen on endoscopy and ultrasound, with elevated round or oval lesions that may overlie Peyer's patches. Mesenteric lymph nodes are enlarged. Ulcerations of the mucosa are noted on endoscopy. Gastrointestinal complications include granulomatous appendicitis, a chronic inflammatory condition affecting the appendix that is responsible for ≥2% of cases of appendicitis; *Yersinia* is involved in a minority of cases. *Y. enterocolitica* infection can present as acute pharyngitis with or without other gastrointestinal symptoms. Fatal *Y. enterocolitica* pharyngitis has been recorded. Mycotic aneurysm can follow *Y. enterocolitica* bacteremia, as can focal infection (abscess) in many other sites and body compartments (liver, spleen, kidney, bone, meninges, endocardium).

In all age groups, *Y. pseudotuberculosis* is more likely to present as abdominal pain and fever than as diarrhea. A superantigenic toxin—*Y. pseudotuberculosis* mitogen (YPM)—is produced by strains seen in eastern Russia in association with Far Eastern scarlet-like fever, a childhood illness with desquamating rash, arthralgia, and toxic shock. A similar illness is recognized in Japan (Izumi fever) and Korea. Similarities have been noted with Kawasaki's disease, the idiopathic acute systematic vasculitis of childhood. There is an epidemiologic link between exposure of populations to superantigen-positive *Y. pseudotuberculosis* and an elevated incidence of Kawasaki's disease.

Y. enterocolitica or *Y. pseudotuberculosis* septicemia presents as a severe illness with fever and leukocytosis, often without localizing features, and is significantly associated with predisposing conditions such as diabetes mellitus, liver disease, and iron overload. Hemochromatosis combines several of these risk factors. Administration of iron chelators like desferrioxamine, which provide iron accessible to *Yersinia* (and have an inhibitory effect on neutrophil function), may result in *Yersinia* septicemia in patients with iron overload who presumably have an otherwise mild gastrointestinal infection. HIV/AIDS has been associated with *Y. pseudotuberculosis* septicemia. The unusual phenomenon of transfusion-associated septicemia is linked to the ability of *Y. enterocolitica* to multiply at refrigerator temperature (psychrotrophy). Typically, the transfused unit has been stored for >20 days, and it is believed that small numbers of yersiniae from an apparently healthy donor with subclinical bacteremia are amplified to very high numbers by growth inside the bag at ≥4°C, with consequent septic shock after transfusion. A method for preventing this very rare event (i.e., a range of 1 case in 500,000 to 1 case in several million transfused units in countries such as the United States and France) without unacceptable restriction in the blood supply has not yet been devised.

Postinfective Phenomena

As in other invasive intestinal infections (salmonellosis, shigellosis), reactive arthritis (articular arthritis of multiple joints developing within 2–4 weeks of a preceding infection) occurs as a result of autoimmune activity initiated by the deposition of bacterial components (not viable bacteria) in joints in combination with the immune response to invading bacteria. The majority of individuals affected by reactive arthritis due to *Yersinia* are HLA-B27 positive. Myocarditis with electrocardiographic ST-segment abnormalities may occur with *Yersinia*-associated reactive arthritis. Most *Yersinia*-associated cases follow *Y. enterocolitica* infection (presumably because it is more common than infection with other species), but *Y. pseudotuberculosis*–associated reactive arthritis is also well documented in Finland, where sporadic and outbreak infections with *Y. pseudotuberculosis* are more common than in other countries. Of infected individuals identified in a recent *Y. pseudotuberculosis* serotype O:3 outbreak in Finland, 12% developed reactive arthritis affecting the small joints of the hands and feet, knees, ankles, and shoulders and lasting >6 months in most cases. Erythema nodosum (Fig. e7-40) occurs after *Yersinia* infection (more commonly in women) with no evidence of HLA-B27 linkage.

There is a long-standing association between antithyroid and anti-*Yersinia* antibodies. Antibody evidence of prior *Y. enterocolitica* infection in Graves' disease and increased levels of antithyroid

antibody in patients with *Y. enterocolitica* antibodies were first noted in the 1970s. *Y. enterocolitica* contains a thyroid-stimulating hormone (TSH)–binding site that is recognized by anti-TSH antibodies from Graves' disease patients. Raised titers of antibodies to *Y. enterocolitica* whole cells and Yops have been found in some series of Graves' disease patients but not in others. One Danish study of twins found no evidence of an association between asymptomatic *Yersinia* infection (as evidenced by anti-Yop antibody titers) and antithyroid antibodies in euthyroid individuals, while another Danish study of twins with and without Graves' disease found that increased anti-Yop antibody titers were associated with Graves' disease. It remains unclear whether this cross-reactivity is significant in the etiology of Graves' disease.

Laboratory Diagnosis

Standard laboratory culture methods can be used to isolate enteropathogenic *Yersinia* species from sterile samples, including blood and cerebrospinal fluid. Culture on specific selective media (CIN agar), with or without preenrichment in broth or phosphate-buffered saline at either 4°C or 16°C, is the basis of most schema for isolation of yersiniae from stool or other nonsterile samples. Outside known high-incidence areas, specific culture may be carried out by laboratories only upon request. Virulence plasmid–negative strains of *Y. enterocolitica* can be isolated from cultures of stool from asymptomatic individuals, especially after cold enrichment. These strains usually differ in biotype (typically biovar 1a) from virulence plasmid–possessing strains; although some display apparent pathogenicity in a mouse model, virulence plasmid–negative strains are not commonly accepted as human pathogens. Because of the frequency with which the virulence plasmid is lost on laboratory subculture, combined biochemical identification (with biotyping according to a standard schema) and serologic identification are usually required to interpret the significance of an isolate of *Y. enterocolitica* from a nonsterile site. Most pathogenic *Y. enterocolitica* strains currently isolated from humans are of serogroup O:3/biovar 4 or serogroup O:9/biovar 2; this pattern holds even in the United States, where serogroup O:8/biovar 1B strains were previously predominant. Many self-validated multiplex PCR screens for detection of *Y. enterocolitica* in clinical samples—and rather more for its detection in food—have been described, but none of these assays is widely used outside its originating laboratory. A standard for PCR detection in food samples is being prepared by the International Organization for Standardization.

Agglutinating or ELISA antibody titers to specific O-antigen types are used in the retrospective diagnosis of both *Y. enterocolitica* and *Y. pseudotuberculosis* infections. IgA and IgG antibodies persist in patients with reactive arthritis. Serologic cross-reactions between *Y. enterocolitica* serogroup O:9 and *Brucella* are due to the similarity of their lipopolysaccharide structures. Multiple assays are required to cover even the predominant serogroups (*Y. enterocolitica* O:3, O5,27, and O:9; *Y. pseudotuberculosis* O:1a, O:1b, and O:3), and these assays are generally available only in reference laboratories. ELISA and western blot tests for antibodies to Yops, which are expressed by all pathogenic strains of *Y. enterocolitica* and *Y. pseudotuberculosis*, are also available; most of the positivity in these assays probably relates to previous infection with *Y. enterocolitica*.

Treatment: Yersiniosis

Most cases of diarrhea caused by enteropathogenic *Yersinia* are self-limiting. Data from clinical trials do not support antimicrobial treatment for adults or children with *Y. enterocolitica* diarrhea. Systemic infections with bacteremia or focal infections outside the gastrointestinal tract generally require antimicrobial therapy. Infants <3 months of age with documented *Y. enterocolitica* infection may require antimicrobial treatment because of the increased likelihood of bacteremia in this age group. *Y. enterocolitica* strains nearly always express β-lactamases. Because of the relative rarity of systemic Y. enterocolitica infection, there are no clinical trial data to guide antimicrobial choice or to suggest the optimal dose and duration of therapy. On the basis of retrospective case series and in vitro sensitivity data, fluoroquinolone therapy is effective for bacteremia in adults; for example, ciprofloxacin is given at a typical dose of 500 mg twice daily by mouth or 400 mg twice daily IV for at least 2 weeks (longer if positive blood cultures persist). A third-generation cephalosporin is an alternative—e.g., cefotaxime (typical dose, 6–8 g/d in 3 or 4 divided doses). In children, third-generation cephalosporins are effective; for example, cefotaxime is given to children ≥1month of age at a typical dose of 75–100 mg/kg per day in 3 or 4 divided doses, with an increase to 150–200 mg/kg per day in severe cases (maximal daily dose, 8–10 g). Amoxicillin and amoxicillin/clavulanate have shown poor efficacy in case series. Trimethoprim-sulfamethoxazole, gentamicin, and imipenem are all active in vitro. *Y. pseudotuberculosis* strains do not express β-lactamase but are intrinsically resistant to polymyxin. Because human infection with *Y. pseudotuberculosis* is less common than that with *Y. enterocolitica*, less case information is available; however, studies in mice suggest that ampicillin is ineffective. Drugs similar to those used against *Y. enterocolitica* should be used. The best results have been obtained with a quinolone.

Some trials of treatment for reactive arthritis (with a large proportion of cases due to *Yersinia*) found that 3 months of oral ciprofloxacin therapy did not affect outcome. One trial in which the same therapy was given specifically for *Y. enterocolitica*–reactive arthritis found that, while outcome indeed was not affected, there was a trend toward faster remission of symptoms in the treated group. Follow-up 4–7

years after initial antibiotic treatment of reactive arthritis (predominantly following *Salmonella* and *Yersinia* infections) demonstrated apparent efficacy in the prevention of chronic arthritis in HLA-B27-positive individuals. A trial showing that azithromycin therapy did not affect outcome in reactive arthritis included cases believed to follow yersiniosis, although no breakdown of cases was provided. A Cochrane review evaluating the use of antibiotics for reactive arthritis is in progress.

Prevention and Control

Current control measures are similar to those used against other enteric pathogens like *Salmonella* and *Campylobacter*, which colonize the intestine of food animals. The focus is on safe handling and processing of food. No vaccine is effective in preventing intestinal colonization of food animals by enteropathogenic *Yersinia*. Consumption of food made from raw pork (which is popular in Germany and Belgium) should be discouraged at present because it is not possible to eliminate contamination with the enteropathogenic *Yersinia* strains found worldwide in pigs. Exposure of infants to raw pig intestine during domestic preparation of chitterlings is inadvisable. Modification of abattoir technique in Scandinavian countries from the 1990s onward included the removal of pig intestines in a closed plastic bag; levels of carcass contamination with *Y. enterocolitica* were reduced, but such contamination was not eliminated. Experimental pig herds free of pathogenic *Y. enterocolitica* O:3 (and also of *Salmonella*, *Toxoplasma*, and *Trichinella*) have been established in Norway and may be commercialized in the future because of their enhanced safety. In the food industry, vigilance is required because of the potential for large outbreaks if small numbers of enteropathogenic yersiniae contaminate any ready-to-eat food whose safe preservation is based on refrigeration before consumption.

The rare phenomenon of contamination of blood for transfusion has proved impossible to eradicate. However, leukodepletion is now practiced in most blood transfusion centers, primarily to prevent nonhemolytic febrile transfusion reactions and alloimmunization against HLA antigens. This measure reduces but does not eliminate the risk of *Yersinia* blood contamination.

Notification of yersiniosis is now obligatory in some countries.

FURTHER READINGS

Abdel-Haq NM et al: *Yersinia enterocolitica* infection in children. Pediatr Infect Dis J 19:954, 2000[PMID: 11055595]

Boqvist S et al: Sources of sporadic *Yersinia enterocolitica* infection in children in Sweden, 2004: A case-control study. Epidemiol Infect 137:897, 2009[PMID: 18789174]

Gage KL, Kosoy MY: Natural history of plague: Perspectives from more than a century of research. Annu Rev Entomol 50:505, 2005[PMID: 15471529]

Inglesby TV et al: Plague as a biological weapon: Medical and public health management. Working Group on Civilian Biodefense. JAMA 283:2281, 2000[PMID: 10807389]

Kool JL: Risk of person-to-person transmission of pneumonic plague. Clin Infect Dis 40:1166, 2005[PMID: 15791518]

Long C et al: *Yersinia pseudotuberculosis* and *Y. enterocolitica* infections, FoodNet, 1996–2007. Emerg Infect Dis 16:566, 2010[PMID: 20202449]

Perdikogianni C et al: *Yersinia enterocolitica* infection mimicking surgical conditions. Pediatr Surg Int 22:589, 2006[PMID: 16770604]

Prentice MB, Rahalison L: Plague. Lancet 369:1196, 2007[PMID: 17416264]

Williamson ED: Plague. Vaccine 27(Suppl 4):D56, 2009

World Health Organization: International Meeting on Preventing and Controlling Plague: The old calamity still has a future. Wkly Epidemiol Rec 28:278, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BARTONELLA INFECTIONS, INCLUDING CAT-SCRATCH DISEASE: INTRODUCTION

Bartonella species are fastidious, facultative intracellular, slow-growing, gram-negative bacteria that cause a broad spectrum of diseases in humans. This genus includes at least 27 distinct species or subspecies, of which at least 13 have been recognized as confirmed or potential human pathogens; *B. bacilliformis*, *B. quintana*, and *B. henselae* are most commonly identified (Table 160-1). Most *Bartonella* species have successfully adapted to survival in specific domestic or wild mammals. Prolonged intraerythrocytic infection in these animals creates a reservoir for human infections. *B. bacilliformis* and *B. quintana*, which are not zoonotic, are exceptions to this rule. Arthropod vectors are often involved. Isolation and characterization of *Bartonella* species are difficult and require special techniques. Clinical presentation generally depends on both the infecting *Bartonella* species and the immune status of the infected individual. *Bartonella* species are susceptible to many antibiotics in vitro; however, clinical responses to therapy and studies in animal models suggest that the minimal inhibitory concentrations of many antimicrobial agents correlate poorly with the drugs' in vivo efficacies in patients with *Bartonella* infections.

Table 160-1 *Bartonella* Species Known or Suspected to Be Human Pathogens

***Bartonella* Species**[a]	**Disease**	**Reservoir Host**[b]	**Arthropod Vector**
B. henselae	Cat-scratch disease, bacillary angiomatosis, bacillary peliosis, bacteremia, endocarditis	Cats, other felines	Cat fleas (*Ctenocephalides felis*): associated with cat-to-cat, but not with cat-to-human, transmission
B. quintana	Trench fever, chronic bacteremia, bacillary angiomatosis, endocarditis	Humans	Human body lice (*Pediculus humanus corporis*)
B. bacilliformis	Bartonellosis (Carrion's disease)	Humans	Sandflies (*Lutzomyia verrucarum*)
B. elizabethae	Endocarditis	Rats, dogs	Unknown
B. grahamii	Retinitis	Mice, voles	Fleas
B. vinsonii subsp. *arupensis*	Endocarditis	Mice	Ticks
B. vinsonii subsp. *berkhoffii*	Endocarditis	Domestic dogs, coyotes, gray foxes	Ticks
B. washoensis	Myocarditis, meningitis	Squirrels, possibly other rodents	Fleas
B. alsatica	Endocarditis	Rabbits	Unknown
B. koehlerae	Endocarditis	Cats	Unknown
B. clarridgeiae	Possibly cat-scratch disease	Cats	Unknown
B. rochalimae	Bacteremia, fever, splenomegaly	Unknown	Possibly fleas
B. tamiae	Bacteremia, fever, myalgia, rash	Unknown	Unknown

[a]Many other *Bartonella* species exist but are not recognized as human pathogens.

[b]Animals are implicated when existing evidence supports their infection with *Bartonella* species. Data supporting animal-to-human transmission may be lacking.

CAT-SCRATCH DISEASE

Definition and Etiology

Usually a self-limited illness, cat-scratch disease (CSD) has two general clinical presentations. *Typical* CSD, the more common, is characterized by subacute regional lymphadenopathy; *atypical* CSD is the collective designation for numerous extranodal manifestations involving various organs. *B. henselae* is the principal etiologic agent of CSD. Rare cases have been associated with *Afipia felis* and *B. quintana*; *B. clarridgeiae* may occasionally be involved as well.

Epidemiology

CSD occurs worldwide, favoring warm and humid climates. In temperate climates, incidence peaks during fall and winter; in the tropics, disease occurs year-round. Adults are affected nearly as frequently as children. Intrafamilial clustering is rare, and person-to-person transmission does not occur. Apparently healthy cats constitute the major reservoir of *B. henselae*, and cat fleas (*Ctenocephalides felis*) may be responsible for cat-to-cat transmission. CSD usually follows contact with cats (especially kittens), but other animals (e.g., dogs) have been implicated as possible reservoirs in rare instances. In the United States, the estimated disease incidence is ~10 cases per 100,000 population. About 10% of patients are hospitalized.

Pathogenesis

Inoculation of *B. henselae*, possibly via contaminated flea feces, usually results from a cat scratch or bite. Exposure to mucous membranes or conjunctivae via droplets or licking may possibly be involved as well. With lymphatic drainage to one or more regional lymph nodes in immunocompetent hosts, a T_H1 response can result in necrotizing granulomatous lymphadenitis. Dendritic cells, along with their associated chemokines, play a role in the host inflammatory response and granuloma formation.

Clinical Manifestations and Prognosis

Of patients with CSD, 85–90% have typical disease. The primary lesion, a small (0.3- to 1-cm) painless erythematous papule or pustule, develops at the inoculation site (usually the site of a scratch or a bite) within days to 2 weeks in about two-thirds of patients (Fig. 160-1***A***, ***B***). Lymphadenopathy develops ≥1–3 weeks after cat contact. The affected lymph node(s) are enlarged and usually painful, sometimes have overlying erythema, and suppurate in 10–15% of cases (Fig. 160-1***C***, ***D***, and ***E***). Axillary/epitrochlear nodes are most commonly involved; next in frequency are head/neck nodes and then inguinal/femoral nodes. Approximately 50% of patients have fever, malaise, and anorexia. A smaller proportion experience weight loss and night sweats mimicking the presentation of lymphoma. Fever is usually low-grade but infrequently rises to ≥39°C. Resolution is slow, requiring weeks (for fever, pain, and accompanying signs and symptoms) to months (for node shrinkage).

Figure 160-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

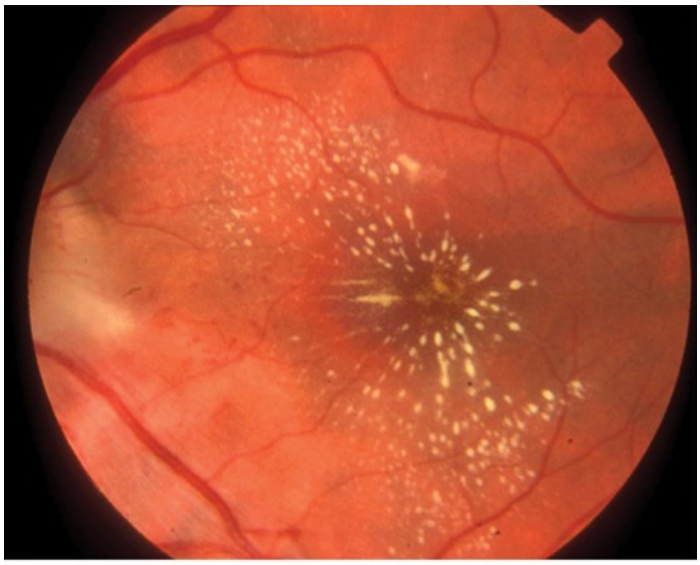

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Manifestations of cat-scratch disease. ***A:*** Primary inoculation lesion. Axillary and epitrochlear lymphadenitis appeared 2 weeks later. ***B:*** Primary inoculation lesion. Submental lymphadenitis appeared 10 days later. ***C:*** Axillary lymphadenopathy of 2 weeks' duration. The overlying skin appears normal. ***D:*** Cervical lymphadenopathy of 6 weeks' duration. The overlying skin is red. Thick, odorless pus (12 mL) was aspirated. ***E:*** Preauricular lymphadenopathy. ***F:*** Left-eye neuroretinitis. Note papilledema and stellate macular exudates ("macular star").

Atypical CSD occurs in 10–15% of patients as extranodal or complicated disease in the absence or presence of lymphadenopathy. Atypical disease includes Parinaud's oculoglandular syndrome (granulomatous conjunctivitis with ipsilateral preauricular lymphadenitis; Fig. 160-1, ***E***), granulomatous hepatitis/splenitis, neuroretinitis (often presenting as unilateral deterioration of vision; Fig. 160-1***F***), and other ophthalmologic manifestations. In addition, neurologic involvement (encephalopathy, seizures, myelitis, radiculitis, cerebellitis, facial and other cranial or peripheral palsies), fever of unknown origin (FUO), debilitating myalgia, arthritis or arthralgia (affecting mostly women >20 years old), osteomyelitis (including multifocal disease), tendinitis, neuralgia, and dermatologic manifestations [including erythema nodosum (see Fig. e7-40), sometimes accompanying arthropathy] occur. Other manifestations and syndromes [pneumonitis, pleural effusion, idiopathic thrombocytopenic purpura, Henoch-Schönlein purpura, erythema multiforme (see Fig. e7-25), hypercalcemia, glomerulonephritis, myocarditis] have also been associated with CSD. In elderly patients (>60 years old), lymphadenopathy is most often absent, but encephalitis and FUO are more common than in younger patients. In immunocompetent individuals, CSD—whether typical or atypical—usually resolves without treatment and without sequelae. Lifelong immunity is the rule.

Diagnosis

Routine laboratory tests usually yield normal or nonspecific results. Histopathology initially shows

lymphoid hyperplasia and later demonstrates stellate granulomata with necrosis, coalescing microabscesses, and occasional multinucleated giant cells, findings which, although nonspecific, may narrow the differential diagnosis. Serologic testing (immunofluorescence or enzyme immunoassay) is the most commonly used laboratory diagnostic approach, with variable sensitivity and specificity. Seroconversion may take a few weeks. Other tests are of low sensitivity (culture, Warthin-Starry silver staining), of low specificity (cytology, histopathology), or of limited availability in routine diagnostic laboratories (PCR, immunohistochemistry). PCR of lymph node tissue, pus, or the primary inoculation lesion is highly sensitive and specific and is particularly useful for definitive and rapid diagnosis in seronegative patients.

Approach to the Patient: Cat-Scratch Disease

A history of cat contact, a primary inoculation lesion, and regional lymphadenopathy are highly suggestive of CSD. A characteristic clinical course and corroborative laboratory tests make the diagnosis very likely. Conversely, when acute- and convalescent-phase sera are negative (as is the case in 10–20% of CSD patients), when spontaneous regression of lymph node size does not occur, and particularly when constitutional symptoms persist, malignancy must be ruled out. Pyogenic lymphadenitis, mycobacterial infection, brucellosis, syphilis, tularemia, plague, toxoplasmosis, sporotrichosis, and histoplasmosis should also be considered. In clinically suspected CSD in a seronegative individual, fine-needle aspiration may be adequate and PCR can confirm the diagnosis. When data are less supportive of CSD, lymph node biopsy rather than fine-needle aspiration is preferred. In seronegative CSD patients with lymphadenopathy and severe complications (e.g., encephalitis or neuroretinitis), early biopsy is important to establish a specific diagnosis.

Treatment: Cat-Scratch Disease

(Table 160-2) Treatment regimens are based on only minimal data. Suppurative nodes should be drained by large-bore needle aspiration and not by incision and drainage in order to avoid chronic draining tracts. Immunocompromised patients must always be treated with systemic antimicrobials.

Table 160-2 Antimicrobial Therapy for Disease Caused by *Bartonella* Species in Adults

Disease	Antimicrobial Therapy
Typical cat-scratch disease	Not routinely indicated; for patients with extensive lymphadenopathy, consider azithromycin (500 mg PO on day 1, then 250 mg PO qd for 4 days)
Cat-scratch disease retinitis	Doxycycline (100 mg PO bid) *plus* rifampin (300 mg PO bid) for 4–6 weeks
Other atypical cat-scratch disease manifestations[a]	As per retinitis; treatment duration should be individualized
Trench fever or chronic bacteremia with *B. quintana*	Gentamicin (3 mg/kg IV qd for 14 days) *plus* doxycycline (200 mg PO qd or 100 mg PO bid for 6 weeks)
Suspected *Bartonella* endocarditis	Gentamicin[b] (1 mg/kg IV q8h for ≥14 days) *plus* doxycycline (100 mg PO/IV bid for 6 weeks[c]) *plus* ceftriaxone (2 g IV qd for 6 weeks)
Confirmed *Bartonella* endocarditis	As for suspected *Bartonella* endocarditis *minus* ceftriaxone
Bacillary angiomatosis	Erythromycin[d] (500 mg PO qid for 3 months) *or* Doxycycline (100 mg PO bid for 3 months)
Bacillary peliosis Bartonellosis (Carrion's disease)	Erythromycin[d] (500 mg PO qid for 4 months) *or* Doxycycline (100 mg PO bid for 4 months)
Oroya fever	Chloramphenicol (500 mg PO/IV qid for 14 days) *plus* another antibiotic (β-lactam preferred) *or* Ciprofloxacin (500 mg PO bid for 10 days)
Verruga peruana	Rifampin (10 mg/kg PO qd, to a maximum of 600 mg, for 14 days)

	or Streptomycin (15–20 mg/kg IM qd for 10 days)

[a]Data on treatment efficacy for encephalitis and hepatosplenic CSD are lacking. Therapy similar to that given for retinitis is reasonable.

[b]Some experts recommend gentamicin at 3 mg/kg IV qd. If gentamicin is contraindicated, rifampin (300 mg PO bid) can be added to doxycycline for documented *Bartonella* endocarditis.

[c]Some experts recommend extending oral doxycycline therapy for 3–6 months.

[d]Other macrolides are probably effective and may be substituted for erythromycin or doxycycline.

Source: Recommendations are modified from Rolain et al, 2004.

Prevention

Avoiding cats (especially kittens) and instituting flea control are options for immunocompromised patients and for patients with valvular heart disease.

TRENCH FEVER AND CHRONIC BACTEREMIA

Definition and Etiology

Trench fever, also known as *5-day fever* or *quintan fever*, is a febrile illness caused by *B. quintana*. It was first described as an epidemic in the trenches of World War I and recently reemerged as chronic bacteremia seen most often in homeless people (also referred to as *urban* or *contemporary trench fever*).

Epidemiology

In addition to epidemics during World Wars I and II, sporadic outbreaks of trench fever have been reported in many regions of the world. The human body louse (*Pediculus humanuscorporis*) has been identified as the vector and humans as the only known reservoir. After a hiatus of several decades during which trench fever was almost forgotten, small clusters of cases of *B. quintana* chronic bacteremia were reported sporadically, primarily from the United States and France, in HIV-uninfected homeless people. Alcoholism and louse infestation were identified as risk factors.

Clinical Manifestations

The typical incubation period is 15–25 days (range, 3–38 days). "Classical" trench fever, as described in 1919, ranges from a mild febrile illness to a recurrent or protracted and debilitating disease. Onset may be abrupt or preceded by a prodrome of several days. Fever is often periodic, lasting 4–5 days with 5-day (range, 3- to 8-day) intervals between episodes. Other symptoms and signs include headache, back and limb pain, profuse sweating, shivering, myalgia, arthralgia, splenomegaly, a maculopapular rash in occasional cases, and nuchal rigidity in some cases. Untreated, the disease usually lasts 4–6 weeks. Death is rare. The clinical spectrum of *B. quintana* bacteremia in homeless people ranges from asymptomatic infection to a febrile illness with headache, severe leg pain, and thrombocytopenia. Endocarditis sometimes develops.

Diagnosis

Definitive diagnosis requires isolation of *B. quintana* by blood culture. Some patients have positive blood cultures for several weeks. Patients with acute trench fever typically develop significant titers of antibody to *Bartonella*, whereas those with chronic *B. quintana* bacteremia may be seronegative. Patients with high titers of IgG antibodies should be evaluated for endocarditis. In epidemics, trench fever should be differentiated from epidemic louse-borne typhus and relapsing fever, which occur under similar conditions and share many features.

Treatment: Bacteremia

(Table 160-2) In a small, randomized, placebo-controlled trial involving homeless people with *B. quintana* bacteremia, therapy with gentamicin and doxycycline was superior to administration of placebo in eradicating bacteremia. Treatment of bacteremia is important even in clinically mild cases to prevent endocarditis. Optimal therapy for trench fever without documented bacteremia is uncertain.

BARTONELLA ENDOCARDITIS

Definition and Etiology

Coxiella burnetii (Chap. 174) and *Bartonella* species are the most common pathogens in culture-negative endocarditis (Chap. 124). In France, for example, *Bartonella* species were identified as the etiologic agents in 28% of 348 cases of culture-negative endocarditis. Prevalence, however, varies by geographic location and epidemiologic setting. In addition to *B. quintana* and *B. henselae* (the most common *Bartonella* species implicated in endocarditis, with the former more commonly involved than

the latter), other *Bartonella* species have reportedly caused rare cases (Table 160-1).

Epidemiology

Bartonella endocarditis has been reported worldwide. Most patients are adults; more are male than female. Risk factors associated with *B. quintana* endocarditis include homelessness, alcoholism, and body louse infestation; however, individuals with no risk factors have had *Bartonella* endocarditis diagnosed as well. *B. henselae* endocarditis is associated with exposure to cats. Most cases involve native rather than prosthetic valves; the aortic valve accounts for ~60% of cases. Patients with *B. henselae* endocarditis usually have preexisting valvulopathy, whereas *B. quintana* often infects normal valves.

Clinical Manifestations

Clinical manifestations are usually characteristic of subacuteendocarditis of any etiology. However, a substantial number of patients have a prolonged, minimally febrile or even afebrile indolent illness, with mild nonspecific symptoms lasting weeks or months before the diagnosis is made. Initial echocardiography may not show vegetations. Acute, aggressive disease is rare.

Diagnosis

Blood cultures, even with use of special techniques (lysis centrifugation or EDTA-containing tubes), are positive in only ~25% of cases—mostly those caused by *B. quintana* and only rarely those caused by *B. henselae*. Prolonged incubation of cultures (up to 6 weeks) is required. Serologic tests—either immunofluorescence or enzyme immunoassay—usually demonstrate high-titer IgG antibodies to *Bartonella*. Because of cross-antigenicity, serology does not distinguish between *B. quintana* and *B. henselae* and may also be low-titer cross-reactive with other pathogens, such as *C. burnetii* and *Chlamydophila* species. Identification of *Bartonella* to the species level is usually accomplished by application of PCR-based methods to valve tissue.

Treatment: *Bartonella* Endocarditis

(Table 160-2) For patients with culture-negative endocarditis suspected to be due to *Bartonella* species, empirical treatment consists of gentamicin, doxycycline, and ceftriaxone; the major role of ceftriaxone in this regimen is to adequately treat other potential causes of culture-negative endocarditis, including members of the HACEK group. Once a diagnosis of *Bartonella* endocarditis has been established, ceftriaxone is discontinued. Aminoglycosides, the only antibiotics known to be bactericidal against *Bartonella*, should be included in the regimen for ≥2 weeks. Indications for valvular surgery are the same as in subacute endocarditis due to other pathogens; however, the proportion of patients who undergo surgery (~60%) is high, probably as a consequence of delayed diagnosis.

BACILLARY ANGIOMATOSIS AND PELIOSIS

Definition and Etiology

Bacillary angiomatosis (sometimes called *bacillary epithelioid angiomatosis* or *epithelioid angiomatosis*) is a disease of severely immunocompromised patients, is caused by *B. henselae* or *B. quintana*, and is characterized by neovascular proliferative lesions involving the skin and other organs. Both species cause cutaneous lesions; hepatosplenic lesions are caused only by *B. henselae*, while subcutaneous and lytic bone lesions are more frequently associated with *B. quintana*. Bacillary peliosis is a closely related angioproliferative disorder caused by *B. henselae* and involving primarily the liver (peliosis hepatis) but also the spleen and lymph nodes. Bacillary peliosis is characterized by blood-filled cystic structures whose size ranges from microscopic to several millimeters.

Epidemiology

Bacillary angiomatosis and bacillary peliosis occur primarily in HIV-infected persons (Chap. 189) with CD4+ T cell counts <100/μL but also affect other immunosuppressed patients and, in rare instances, immunocompetent patients. The previously reported incidence of ~1 case per 1000 HIV-infected persons is now lower; the recent decrease is most likely attributable to effective antiretroviral therapy and the routine use of rifabutin and macrolides to prevent *Mycobacterium avium* complex infection in AIDS patients. Contact with cats or cat fleas elevates the risk of *B. henselae* infection. Risk factors for *B. quintana* infection are low income, homelessness, and body louse infestation.

Clinical Manifestations

Bacillary angiomatosis presents most commonly as one or more cutaneous lesions that are not painful and that may be tan, red, or purple in color. Subcutaneous masses or nodules, superficial ulcerated plaques (Fig. 160-2), and verrucous growths are also seen. Nodular forms resemble those seen in fungal or mycobacterial infections. Subcutaneous nodules are often tender. Painful osseous lesions, most often involving long bones, may underlie cutaneous lesions and occasionally develop in their absence. In rare cases, other organs are involved in bacillary angiomatosis. Patients usually have constitutional symptoms, including fever, chills, malaise, headache, anorexia, weight loss, and night sweats. In osseous disease, lytic lesions are generally seen on radiography, and technetium scan

shows focal uptake. The differential diagnosis of cutaneous bacillary angiomatosis includes Kaposi's sarcoma, pyogenic granuloma, subcutaneous tumors, and verruga peruana. In bacillary peliosis, hypodense hepatic areas are usually evident on imaging. In patients with advanced immunodeficiency, *B. henselae* and *B. quintana* are important causes of FUO. Intermittent bacteremia with positive blood cultures can occur with or without endocarditis.

Figure 160-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nodular lesion of bacillary angiomatosis with superficial ulceration in an AIDS patient with advanced immunodeficiency. *[Reprinted with permission from DH Spach and E Darby: Bartonella Infections, Including Cat-Scratch Disease, in Harrison's Principles of Internal Medicine, 17th ed, AF Fauci et al (eds). New York, McGraw-Hill, 2008, p 989.]*

Pathology

Bacillary angiomatosis consists of lobular proliferations of small blood vessels lined by enlarged endothelial cells interspersed with mixed infiltrates of neutrophils and lymphocytes, with predominance of the former. Histologic examination of organs with bacillary peliosis reveals small blood-filled cystic lesions partially lined by endothelial cells that can be several millimeters in size. Peliotic lesions are surrounded by fibromyxoid stroma containing inflammatory cells, dilated capillaries, and clumps of granular material. Warthin-Starry silver staining of bacillary angiomatosis and peliosis lesions reveals clusters of bacilli. Cultures are usually negative.

Diagnosis

Bacillary angiomatosis and bacillary peliosis are diagnosed on histologic grounds. Blood cultures may be positive.

Treatment: Bacillary Angiomatosis and Peliosis

(Table 160-2) Prolonged therapy with a macrolide or doxycycline is recommended for both bacillary angiomatosis and bacillary peliosis.

Prevention

Control of cat-flea infestation and avoidance of cat scratches (for prevention of *B. henselae*) and avoidance and treatment of body louse infestation (for prevention of *B. quintana*) are reasonable strategies for HIV-infected persons. Primary prophylaxis is not recommended, but suppressive therapy with a macrolide or doxycycline is indicated in HIV-infected patients with bacillary angiomatosis or bacillary peliosis until CD4+ T cell counts are >200/μL. Relapse may necessitate lifelong suppressive therapy in individual cases.

BARTONELLOSIS (CARRIÓN'S DISEASE)

Definition and Etiology

Bartonellosis is a biphasic disease caused by *B. bacilliformis*. *Oroya fever* is the initial, bacteremic, systemic form, and *verruga peruana* is its late-onset, eruptive manifestation.

Epidemiology and Prevention

Infection is endemic to the geographically restricted Andes valleys of Peru, Ecuador, and Colombia (~ 500–3200 m above sea level). Sporadic epidemics occur. The disease is transmitted by the

phlebotomine sandfly *Lutzomyia verrucarum*. Humans are the only known reservoir of *B. bacilliformis*. Sandfly control measures (e.g., insecticides) as well as personal protection measures (e.g., repellents, screening, bednets) may decrease the risk of infection.

Pathogenesis

After inoculation by the sandfly, bacteria invade the blood vessel endothelium and proliferate; the reticuloendothelial system and various organs may also be involved. Upon re-entry into blood vessels, *B. bacilliformis* invades, replicates, and ultimately destroys erythrocytes, with consequent massive hemolysis and sudden, severe anemia. Microvascular thrombosis results in end-organ ischemia. Survivors sometimes develop cutaneous hemangiomatous lesions characterized by various inflammatory cells, endothelial proliferation, and the presence of *B. bacilliformis*.

Clinical Manifestations

The incubation period is 3 weeks (range, 2–14 weeks). Oroya fever may present as a nonspecific bacteremic febrile illness without anemia or as an acute, severe hemolytic anemia with hepatomegaly and jaundice of rapid onset leading to vascular collapse and clouded sensorium. Myalgia, arthralgia, lymphadenopathy, and abdominal pain may develop. Temperature is elevated but not extremely so; high fever may suggest intercurrent infection. Subclinical asymptomatic infection also occurs. In verruga peruana, red, hemangioma-like, cutaneous vascular lesions of various sizes appear either weeks to months after systemic illness or with no previous suggestive history. These lesions persist for months up to 1 year. Mucosal and internal lesions may also develop.

Diagnosis and Approach to the Patient

Systemic illness (with or without anemia) or the development of cutaneous lesions in a person who has been to an endemic area raises the possibility of *B. bacilliformis* infection. Severe anemia with exuberant reticulocytosis—and sometimes thrombocytopenia—can occur. In systemic illness, Giemsa-stained blood films show typical intraerythrocytic bacilli, and blood and bone marrow cultures are positive. Serologic assays may be helpful. Biopsy may be required to confirm the diagnosis of verruga peruana. Differential diagnosis includes the spectrum of coendemic systemic febrile illnesses (e.g., typhoid fever, malaria, brucellosis) as well as diseases producing cutaneous vascular lesions (e.g., hemangiomata, bacillary angiomatosis, Kaposi's sarcoma).

Treatment: Bartonellosis

(Table 160-2) Antibiotic therapy for systemic *B. bacilliformis* infection usually results in rapid defervescence. Additionalantibiotic treatment of intercurrent infection (particularlysalmonellosis) is often required. Blood transfusion may be necessary. Treatment of verruga peruana usually is not required, although large lesions or those interfering with function may require excision. Patients with numerous lesions, especially lesions that have been present for only a short period, may respond well to antibiotic therapy.

Complications and Prognosis

Mortality rates associated with Oroya fever have been reported to be as high as 40% without treatment but are considerably lower (~10%) with treatment. Complications such as bacterial superinfection and neurologic and cardiac manifestations occur frequently. Generalized massive edema (anasarca) and petechiae are associated with poor outcome. Permanent immunity usually develops.

FURTHER READINGS

Centers for Disease Control and Prevention: Guidelines for prevention and treatment of opportunistic infections in HIV-infected adults and adolescents. MMWR Recomm Rep 58:1, 2009

Florin TA et al: Beyond cat scratch disease: Widening spectrum of *Bartonella henselae* infection. Pediatrics 121:e1413, 2008

Fournier PE et al: Epidemiologic and clinical characteristics of *Bartonella quintana* and *Bartonella henselae* endocarditis: A study of 48 patients. Medicine (Baltimore) 80:245, 2001[PMID: 11470985]

Koehler JE et al: Molecular epidemiology of *Bartonella* infections in patients with bacillary angiomatosis-peliosis. N Engl J Med 337:1876, 1997[PMID: 9407154]

Maguina C et al: Bartonellosis (CarriÓn's disease) in the modern era. Clin Infect Dis 33:772, 2001[PMID: 11512081]

Rolain JM et al: Recommendations for treatment of human infections caused by *Bartonella* species. Antimicrob Agents Chemother 48:1921, 2004[PMID: 15155180]

Slater LN, Welch DF: *Bartonella* including cat scratch disease, in *Principles and Practice of Infectious Diseases*, 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 2995

–3009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DONOVANOSIS: INTRODUCTION

Donovanosis is a chronic, progressive bacterial infection that usually involves the genital region. The condition is generally regarded as a sexually transmitted infection of low infectivity. This infection has been known by many other names, the most common being *granuloma inguinale*.

Etiology

The causative organism has been reclassified as *Klebsiella granulomatis comb nov* on the basis of phylogenetic analysis, although there is ongoing debate about this decision. Some authorities consider the original nomenclature (*Calymmatobacterium granulomatis*), which is based on analysis of 16S rRNA gene sequences, to be more appropriate.

Donovanosis was first described in Calcutta in 1882, and the causative organism was recognized by Charles Donovan in Madras in 1905. He identified the characteristic Donovan bodies, measuring 1.5 x 0.7 μm, in macrophages and the stratum malpighii. The organism was not reproducibly cultured until the mid-1990s, when its isolation in peripheral blood monocytes and human epithelial cell lines was reported.

Epidemiology

Donovanosis has an unusual geographic distribution that includes Papua New Guinea, parts of southern Africa, India, French Guyana, Brazil, and aboriginal communities in Australia. In Australia, donovanosis has virtually been eliminated through a sustained program backed by strong political commitment and resources at the primary health care level. Although few cases are now reported in the United States, donovanosis was once prevalent in this country, with 5000–10,000 cases recorded in 1947. The largest epidemic recorded was in Dutch South Guinea, where 10,000 cases were identified in a population of 15,000 (the Marind-anim people) between 1922 and 1952.

Donovanosis is associated with poor hygiene and is more common in lower socioeconomic groups than in those who are better off and in men than in women. Infection in sexual partners of index cases occurs to a limited extent. Donovanosis is a risk factor for HIV infection (Chap. 189).

Globally, the incidence of donovanosis has decreased significantly in recent times. This decline probably reflects a greater focus on effective management of genital ulcers because of their role in facilitating HIV transmission.

Clinical Features

A lesion starts as a papule or subcutaneous nodule that later ulcerates after trauma. The incubation period is uncertain, but experimental infections in humans indicate that it lasts ~50 days. Four types of lesions have been described: (1) the classic ulcerogranulomatous lesion (Fig. 161-1), a beefy red ulcer that bleeds readily when touched; (2) a hypertrophic or verrucous ulcer with a raised irregular edge; (3) a necrotic, offensive-smelling ulcer causing tissue destruction; and (4) a sclerotic or cicatricial lesion with fibrous and scar tissue.

Figure 161-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ulcerogranulomatous penile lesion of donovanosis, with some hypertrophic features.

The genitals are affected in 90% of patients and the inguinal region in 10%. The most common sites of infection are the prepuce, coronal sulcus, frenum, and glans in men and the labia minora and fourchette in women. Cervical lesions may mimic cervical carcinoma. In men, lesions are associated with lack of circumcision. Lymphadenitis is uncommon. Extragenital lesions occur in 6% of cases and may involve the lip, gums, cheek, palate, pharynx, larynx, and chest. Hematogenous spread of *K. granulomatis comb nov* to liver and bone has been reported. During pregnancy, lesions tend to develop more quickly and respond more slowly to treatment. Polyarthritis and osteomyelitis are rare complications. In newborn infants, donovanosis may present with ear infection. Cases in children have been attributed to sitting on the laps of infected adults. As the incidence of donovanosis has decreased, the number of unusual case reports has appeared to be increasing.

Complications include neoplastic changes, pseudo-elephantiasis, and stenosis of the urethra, vagina, or anus.

Diagnosis

A clinical diagnosis of donovanosis is made by an experienced practitioner on the basis of the lesion's appearance and usually has a high positive predictive value. The diagnosis is confirmed by microscopic identification of Donovan bodies (Fig. 161-2) in tissue smears. Preparation of a good-quality smear is important. If donovanosis is suspected on clinical grounds, the smear for Donovan bodies should be taken before swab samples to be tested for other causes of genital ulceration so that enough material can be collected from the ulcer. A swab should be rolled firmly over an ulcer previously cleaned with a dry swab to remove debris. Smears can be examined in a clinical setting by direct microscopy with a rapid Giemsa or Wright's stain. Alternatively, a piece of granulation tissue crushed and spread between two slides can be used. Donovan bodies can be seen in large, mononuclear (Pund) cells as gram-negative intracytoplasmic cysts filled with deeply staining bodies that may have a safety-pin appearance. These cysts eventually rupture and release the infective organisms. Histologic changes include chronic inflammation with infiltration of plasma cells and neutrophils. Epithelial changes include ulceration, microabscesses, and elongation of rete ridges.

Figure 161-2

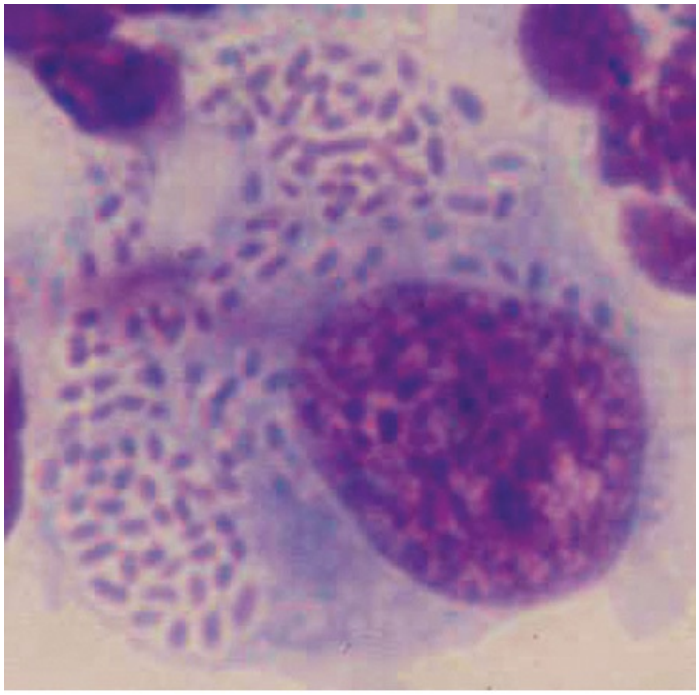

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pund cell stained by rapid Giemsa (RapiDiff) technique showing numerous Donovan bodies.

A diagnostic polymerase chain reaction (PCR) test was developed in light of the observation that two unique base changes in the *phoE* gene eliminate Hae111 restriction sites, enabling differentiation of *K. granulomatis comb nov* from related *Klebsiella* species. PCR analysis with a colorimetric detection system can now be used in routine diagnostic laboratories. A genital ulcer multiplex PCR that includes *K. granulomatis* has been developed. Serologic tests are only poorly specific and are not used currently.

The differential diagnosis includes primary syphilitic chancres, secondary syphilis (condylomata lata), chancroid, lymphogranuloma venereum, genital herpes, neoplasm, and amebiasis. Mixed infections are common. Histologic appearances should be distinguished from those of rhinoscleroma, leishmaniasis, and histoplasmosis.

Treatment: Donovanosis

Many patients with donovanosis present quite late with extensive ulceration. They may be embarrassed and have low self-esteem related to their disease. Reassurance that they have a treatable condition is important, as is the need to administer antibiotics and monitor patients for an adequate interval (see below). Epidemiologic treatment of sexual partners and advice about how to improve genital hygiene are recommended.

The recommended drug regimens for donovanosis are shown in **Table 161-1**. Gentamicin can be added if the response is slow. Ceftriaxone, chloramphenicol, and norfloxacin are also effective. Patients treated for 14 days should be monitored until lesions have healed completely. Those treated with azithromycin probably do not need such rigorous follow-up.

Table 161-1 Effective Antibiotics for the Treatment of Donovanosis

Antibiotic	Oral Dose
Azithromycin	1 g on day 1, then 500 mg daily for 7 days or 1 g weekly for 4 weeks
Trimethoprim-sulfamethoxazole	960 mg bid for 14 days
Doxycycline	100 mg bid for 14 days
Erythromycin	500 mg qid for 14 days (in pregnant women)
Tetracycline	500 mg qid for 14 days

Surgery may be indicated for very advanced lesions.

Control and Prevention

Donovanosis is probably the cause of genital ulceration that is most readily recognizable clinically. Donovanosis is now limited to a few specific locations, and its global eradication is a distinct

possibility.

FURTHER READINGS

Bowden FJ: Donovanosis in Australia: Going, going. . .. Sex Transm Infect 81:365, 2005[PMID: 16199732]

Carter J et al: Phylogenetic evidence for reclassification of *Calymmatobacterium granulomatis* as *Klebsiella granulomatis comb nov*. Int J Syst Bacteriol 49:1695, 1999[PMID: 10555350]

Kharsany AB et al: Phylogenetic analysis of *Calymmatobacterium granulomatis* based on 16S sequences. J Med Microbiol 48:841, 1999[PMID: 10482295]

Mackay IM et al: Detection and discrimination of herpes simplex viruses, *Haemophilus ducreyi*, *Treponema pallidum*, and *Calymmatobacterium (Klebsiella) granulomatis* from genital ulcers. Clin Infect Dis 42:1431, 2006[PMID: 16619156]

O'Farrell N: Donovanosis, in *Sexually Transmitted Diseases*, 4th ed, KK Holmes et al (eds). New York, McGraw-Hill, 2008, pp 700–708

——— et al: Risk factors for HIV-1 in heterosexual attenders at a sexually transmitted diseases clinic in Durban, South Africa. S Afr Med J 80:17, 1991

Rajam RV, Rangiah PN: Donovanosis, granuloma inguinale, granuloma venereum. Geneva, World Health Organization, 1954, pp 1–72

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NOCARDIOSIS: INTRODUCTION

Nocardia species are saprophytic aerobic actinomycetes and are common worldwide in soil, where they contribute to the decay of organic matter. More than 50 species have been identified, mostly on the basis of 16S rRNA gene sequences. More than 30 species have been associated with human disease. Until recently, isolates from the majority of cases of pneumonia and systemic disease were identified as *Nocardia asteroides*, but human disease involving *N. asteroides* proper is actually rare. Nocardiae are relatively inactive in standard biochemical tests, and speciation is difficult or impossible without molecular phylogenetic techniques. Most clinical laboratories cannot speciate isolates accurately and may identify them simply as *N. asteroides* or *Nocardia* species.

Nine species or species complexes are most commonly associated with human disease **(Table 162-1)**. Most systemic disease involves *N. cyriacigeorgica*, *N. farcinica*, *N. pseudobrasiliensis*, and species in the *N. transvalensis* and *N. nova* complexes. *N. brasiliensis* is usually associated with disease limited to the skin. Actinomycetoma—an indolent, slowly progressive disease of skin and underlying tissues with nodular swellings and draining sinuses—is often associated with *N. brasiliensis*, *N. otitidiscaviarum*, *N. transvalensis* complex strains, or other actinomycetes.

Table 162-1 *Nocardia* Species Most Commonly Associated with Human Disease and Their In Vitro Susceptibility Patterns

Species	Susceptible to	Resistant to
N. abscessus	Amikacin, amoxicillin/clavulanic acid, ampicillin, cefotaxime, ceftriaxone, gentamicin, linezolid, minocycline, sulfamethoxazole	Ciprofloxacin, clarithromycin, erythromycin, imipenem (v)[a]
N. brevicatena/paucivorans complex (*N. brevicatena, N. paucivorans, N. carnea*, others)	Amikacin, amoxicillin/clavulanic acid, ampicillin, cefotaxime, ceftriaxone, ciprofloxacin, linezolid, minocycline, tobramycin, sulfamethoxazole	Ciprofloxacin, clarithromycin, erythromycin, gentamicin, imipenem (v)
N. nova complex (*N. nova, N. veterana, N. africana, N. kruczakiae, N. elegans*, others)	Amikacin, ampicillin, ceftriaxone, clarithromycin, erythromycin, imipenem, linezolid, minocycline, sulfamethoxazole	Amoxicillin/clavulanic acid, ciprofloxacin, gentamicin
N. transvalensis complex (*N. blacklockiae, N. wallacei*, others)	Cefotaxime (v), ceftriaxone (v), ciprofloxacin, imipenem, linezolid, sulfamethoxazole	Amikacin, ampicillin, clarithromycin, erythromycin, gentamicin
N. farcinica	Amikacin, ciprofloxacin, imipenem, linezolid, sulfamethoxazole	Ampicillin, cefotaxime, ceftriaxone, clarithromycin, erythromycin, gentamicin, tobramycin
N. cyriacigeorgica	Amikacin, cefotaxime, ceftriaxone, imipenem, linezolid, minocycline (v), sulfamethoxazole	Amoxicillin/clavulanic acid, ampicillin (v), ciprofloxacin, erythromycin, gentamicin
N. brasiliensis	Amikacin, amoxicillin/clavulanic acid, cefotaxime, ceftriaxone, minocycline, sulfamethoxazole	Ampicillin, ciprofloxacin, clarithromycin, imipenem
N. pseudobrasiliensis	Amikacin, cefotaxime (v), ceftriaxone (v), ciprofloxacin, clarithromycin, sulfamethoxazole	Amoxicillin/clavulanic acid, ampicillin, imipenem, minocycline
N. otitidiscaviarum complex	Amikacin, ciprofloxacin, gentamicin, sulfamethoxazole	Amoxicillin/clavulanic acid, ampicillin, ceftriaxone, imipenem

[a] (v), variable.

Source: Adapted from Brown-Elliott et al, 2006.

Epidemiology

Nocardiosis occurs worldwide. The annual incidence has been estimated on three continents (North America, Europe, and Australia) and is ~0.375 cases per 100,000 persons. The disease is more common among adults than among children and among males than among females. Nearly all cases are sporadic, but outbreaks have been associated with contamination of the hospital environment, solutions, or drug injection equipment. Person-to-person spread is not well documented. There is no known seasonality.

More than 90% of cases of pulmonary or disseminated disease occur in people with a host defense defect. Most have deficient cell-mediated immunity, especially that associated with lymphoma, transplantation, glucocorticoid therapy, or AIDS. The incidence is ~140-fold greater among patients with AIDS and ~340-fold greater among bone marrow transplant recipients than in general populations. In AIDS, nocardiosis usually affects persons with <250 CD4+ T lymphocytes/μL. Nocardiosis has also been associated with pulmonary alveolar proteinosis, tuberculosis and other mycobacterial diseases, chronic granulomatous disease, interleukin 12 deficiency, and treatment with monoclonal antibodies to tumor necrosis factor. Any child with nocardiosis and no known cause of immunosuppression should undergo tests to determine the adequacy of the phagocytic respiratory burst.

Cases of actinomycetoma occur mainly in tropical and subtropical regions, especially those of Mexico, Central and South America, Africa, and India. The most important risk factor is frequent contact with soil or vegetable matter, especially in laborers.

Pathology and Pathogenesis

Pneumonia and disseminated disease are both thought to follow inhalation of fragmented bacterial mycelia. The characteristic histologic feature of nocardiosis is an abscess with extensive neutrophil infiltration and prominent necrosis. Granulation tissue usually surrounds the lesions, but extensive fibrosis or encapsulation is uncommon.

Actinomycetoma is characterized by suppurative inflammation with sinus tract formation. Granules —microcolonies composed of dense masses of bacterial filaments extending radially from a central core—are occasionally observed in histologic preparations. They are frequently found in discharges from lesions of actinomycetoma but almost never in discharges from lesions in other forms of nocardiosis. Infrequently, nocardiae and other indolent pathogens, including fungi or mycobacteria, are isolated from the same patient.

Nocardiae have evolved a number of properties that enable them to survive within phagocytes, including neutralization of oxidants, prevention of phagosome-lysosome fusion, and prevention of phagosome acidification. Neutrophils phagocytose the organisms and limit their growth but do not kill them efficiently. Cell-mediated immunity is important for definitive control and elimination of nocardiae.

Clinical Manifestations

RESPIRATORY TRACT DISEASE

Pneumonia, the most common form of nocardial disease in the respiratory tract, is typically subacute; symptoms have usually been present for days or weeks at presentation. The onset is occasionally more acute in immunosuppressed patients. Cough is prominent and produces small amounts of thick, purulent sputum that is not malodorous. Fever, anorexia, weight loss, and malaise are common; dyspnea, pleuritic pain, and hemoptysis are less common. Remissions and exacerbations over several weeks are frequent. Roentgenographic patterns vary, but some are highly suggestive of nocardial pneumonia. Infiltrates vary in size and are typically dense. Single or multiple nodules are common (**Figs. 162-1** and **162-2**), sometimes suggesting tumors or metastases. Infiltrates and nodules tend to cavitate (Fig. 162-2). Empyema is present in one-quarter of cases.

Figure 162-1

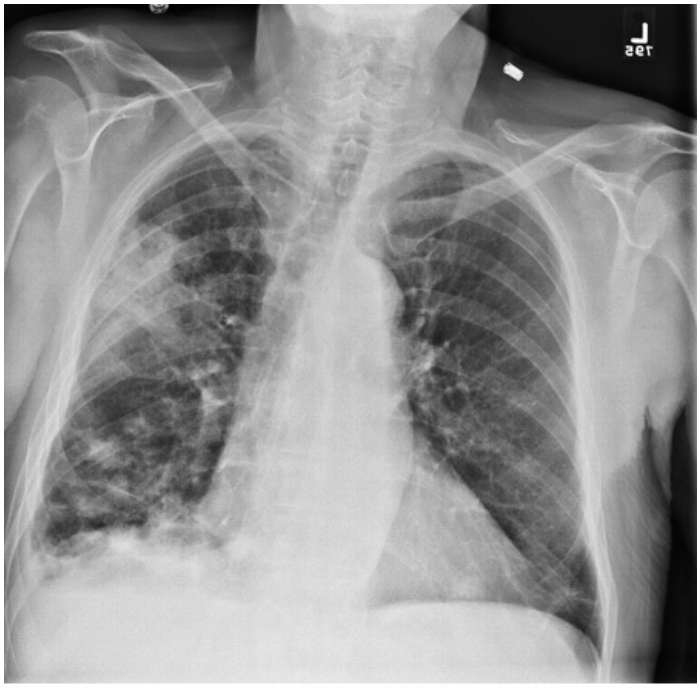

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nocardial pneumonia. A dense infiltrate with a possible cavity and several nodules are apparent in the right lung.

Figure 162-2

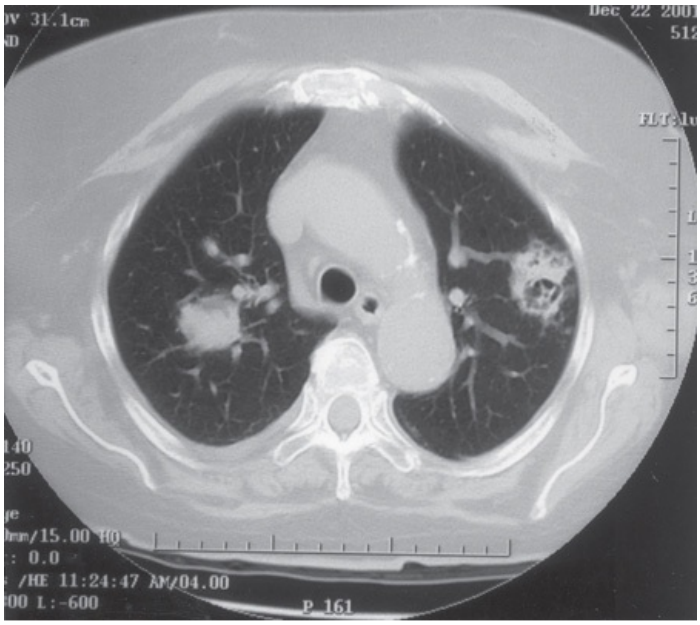

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nocardial pneumonia. A CT scan shows bilateral nodules, with cavitation in the nodule in the left lung.

Nocardiosis may spread directly from the lungs to adjacent tissues. Pericarditis, mediastinitis, and the superior vena cava syndrome have all been reported. Nocardial laryngitis, tracheitis, bronchitis, and sinusitis are much less common than pneumonia. In the major airways, disease often presents as a nodular or granulomatous mass. Nocardiae are sometimes isolated from respiratory secretions of persons without apparent nocardial disease, usually individuals who have underlying lung or airway abnormalities.

EXTRAPULMONARY DISEASE

In half of all cases of pulmonary nocardiosis, disease appears outside the lungs. In one-fifth of cases of disseminated disease, lung disease is not apparent. The most common site of dissemination is the brain. Other common sites include the skin and supporting structures, kidneys, bone, and muscle, but almost any organ can be involved. Peritonitis has been reported in patients undergoing peritoneal dialysis. Nocardiae have been recovered from blood in a few cases of pneumonia, disseminated disease, or central venous catheter infection. Nocardial endocarditis occurs rarely and can affect either native or prosthetic valves.

The typical manifestation of extrapulmonary dissemination is a subacute abscess. A minority of abscesses outside the lungs or central nervous system (CNS) form fistulas and discharge small amounts of pus. In CNS infections, brain abscesses are usually supratentorial, are often multiloculated, and may be single or multiple **(Fig. 162-3)**. Brain abscesses tend to burrow into the ventricles or extend out into the subarachnoid space. The symptoms and signs are somewhat more indolent than those of other types of bacterial brain abscess. Meningitis is uncommon and is usually due to spread from a nearby brain abscess. Nocardiae are not easily recovered from cerebrospinal fluid (CSF).

Figure 162-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nocardial abscesses in the right occipital lobe.

DISEASE FOLLOWING TRANSCUTANEOUS INOCULATION

Disease that follows transcutaneous nocardial inoculation usually takes one of three forms: cellulitis, lymphocutaneous syndrome, or actinomycetoma.

Cellulitis generally begins 1–3 weeks after a recognized breach of the skin, often with soil contamination. Subacute cellulitis, with pain, swelling, erythema, and warmth, develops over days to weeks. The lesions are usually firm and not fluctuant. Disease may progress to involve underlying muscles, tendons, bones, or joints. Dissemination is rare. *N. brasiliensis* and species in the *N. otitidiscaviarum* complex are most common in cellulitis cases.

Lymphocutaneous disease usually begins as a pyodermatous nodule at the site of inoculation, with central ulceration and purulent or honey-colored drainage. Subcutaneous nodules often appear along lymphatics that drain the primary lesion. Most cases of nocardial lymphocutaneous syndrome are associated with *N. brasiliensis*. Similar disease occurs with other pathogens, most notably *Sporothrix schenckii* (sporotrichosis, Chap. 206).

Actinomycetoma (**Fig. 162-4**) usually begins with a nodular swelling, sometimes at a site of local trauma. Lesions typically develop on the feet or hands but may involve the posterior part of the neck, the upper back, the head, and other sites. The nodule eventually breaks down, and a fistula appears, typically followed by others. The fistulas tend to come and go, with new ones forming as old ones disappear. The discharge is serous or purulent, may be bloody, and often contains 0.1- to 2-mm white

granules consisting of masses of mycelia. The lesions spread slowly along fascial planes to involve adjacent areas of skin, subcutaneous tissue, and bone. Over months or years, there may be extensive deformation of the affected part. Lesions involving soft tissues are only mildly painful; those affecting bones or joints are more so. Systemic symptoms are absent or minimal. Infection rarely disseminates from actinomycetoma, and lesions on the hands and feet usually cause only local disability. Lesions on the head, neck, and trunk can invade locally to involve deep organs, with consequent severe disability or death.

Figure 162-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Common features of nocardial actinomycetoma include swelling, multiple sinus tracts, and involvement of the foot. *(Image provided by Amor Khachemoune and Ronald O. Perelman, New York University School of Medicine.)*

EYE INFECTIONS

Nocardia species are uncommon causes of subacute keratitis, usually following eye trauma. Nocardial endophthalmitis can develop after eye surgery. In one series, nocardiae accounted for more than half of culture-proved cases of endophthalmitis after cataract surgery. Endophthalmitis can also occur during disseminated disease. Nocardial infection of lachrymal glands has been reported.

Diagnosis

The first step in diagnosis is examination of sputum or pus for crooked, branching, beaded, gram-positive filaments 1 μm wide and up to 50 μm long **(Fig. 162-5)**. Most nocardiae are acid-fast in direct smears if a weak acid is used for decolorization (e.g., in the modified Kinyoun, Ziehl-Neelsen, and Fite-Faraco methods). The organisms often take up silver stains. Recovery from specimens containing a mixed flora can be improved with selective media (colistin–nalidixic acid agar, modified Thayer-Martin agar, or buffered charcoal–yeast extract agar). Nocardiae grow well on most fungal and mycobacterial media, but procedures used for decontamination of specimens for mycobacterial culture can kill nocardiae and thus should not be used when these organisms are suspected. Nocardiae grow relatively slowly; colonies may take up to 2 weeks to appear and may not develop their characteristic appearance—white, yellow, or orange, with aerial mycelia and delicate, dichotomously branched substrate mycelia—for up to 4 weeks. Several blood culture systems support nocardial growth, although nocardiae may not be detected for up to 2 weeks. The growth of nocardiae is so different from that of more common pathogens that the laboratory should be alerted when nocardiosis is suspected in order to maximize the likelihood of isolation.

Figure 162-5

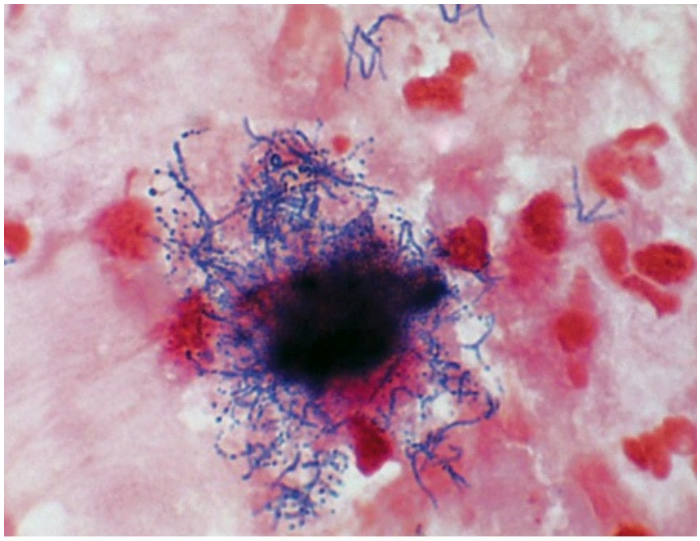

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gram-stained sputum from a patient with nocardial pneumonia. *(Image provided by Charles Cartwright and Susan Nelson, Hennepin County Medical Center, Minneapolis, MN.)*

In nocardial pneumonia, sputum smears are often negative. Unless the diagnosis can be made in smear-negative cases by sampling lesions in more accessible sites, bronchoscopy or lung aspiration is usually necessary. To evaluate the possibility of dissemination in patients with nocardial pneumonia, a careful history should be obtained and a thorough physical examination performed. Suggestive symptoms or signs should be pursued with further diagnostic tests. CT or MRI of the head, with and without contrast material, should be undertaken if signs or symptoms suggest brain involvement. Some authorities recommend brain imaging in all cases of pulmonary or disseminated disease. When clinically indicated, CSF or urine should be concentrated and then cultured. Actinomycetoma, eumycetoma (cases involving fungi; Chap. 206), and botryomycosis (cases involving cocci or bacilli, often Staphylococcus aureus) are difficult to distinguish clinically but are readily distinguished with microbiologic testing. Granules should be sought in any discharge. Suspect particles should be washed in saline, examined microscopically, and cultured. Granules in actinomycetoma cases are usually white, pale yellow, pink, or red. Viewed microscopically, they consist of tight masses of fine filaments (0.5–1 mm wide) radiating outward from a central core. Granules from eumycetoma cases are white, yellow, brown, black, or green. Under the microscope, they appear as masses of broader filaments (2–5 mm wide) encased in a matrix. Granules of botryomycosis consist of loose masses of cocci or bacilli. Organisms can also be seen in wound discharge or histologic specimens. The most reliable way to differentiate among the various organisms associated with mycetoma is by culture.

Isolation of nocardiae from sputum or blood occasionally represents colonization, transient infection, or contamination. In typical cases of respiratory tract colonization, Gram-stained specimens are negative and cultures are only intermittently positive. A positive sputum culture in an immunosuppressed patient usually reflects disease. When nocardiae are isolated from sputum of an immunocompetent patient without apparent nocardial disease, the patient should be observed carefully without treatment. A patient with a host-defense defect that increases the risk of nocardiosis should usually receive antimicrobial treatment.

Treatment: Nocardiosis

The Clinical Laboratory Standards Institute has approved a broth dilution antimicrobial susceptibility test protocol for use with human nocardial isolates. Procedures differ from those used with common human bacterial pathogens, and most clinical laboratories will not be sufficiently experienced with *Nocardia* to produce reliable results. Because nocardiosis is uncommon, data on the relation between susceptibility test results for specific drugs and clinical outcomes in patients treated with these drugs are meager. Empirical therapy with the drugs discussed below is recommended for newly diagnosed cases. When possible, and especially in severe cases or cases that do not improve promptly with empirical therapy, clinicians should arrange for susceptibility tests at a laboratory with experience in nocardial microbiology, such as the Mycobacteria/Nocardia Laboratory at the University of Texas Health Science Center (11937 US Highway 271, Tyler, TX 75708-3154; phone, 903-877-7685; fax, 903-877-7652).

Sulfonamides are the drugs of choice (Tables 162-1 and **162-2**). The combination of sulfamethoxazole (SMX) and trimethoprim (TMP) is probably equivalent to a sulfonamide alone; some authorities believe that the combination may in fact be more effective, but it also poses a modestly greater risk of hematologic toxicity. At the outset, 10–20 mg of TMP per kg and 50–100 mg of SMX per kg are given

each day in two divided doses. Later, daily doses can be decreased to as little as 5 mg/kg and 25 mg/kg, respectively. In persons with sulfonamide allergies, desensitization usually allows continuation of therapy with these effective and inexpensive drugs.

Table 162-2 Treatment Duration for Nocardiosis

Disease	Duration
Pulmonary or systemic	
Intact host defenses	6–12 months
Deficient host defenses	12 months[a]
CNS disease	12 months[b]
Cellulitis, lymphocutaneous syndrome	2 months
Osteomyelitis, arthritis, laryngitis, sinusitis	4 months
Actinomycetoma	6–12 months after clinical cure
Keratitis	Topical: until apparent cure Systemic: until 2–4 months after apparent cure

[a] In some patients with AIDS and CD_4+ T lymphocyte counts of <200/μL or with chronic granulomatous disease, therapy for pulmonary or systemic disease must be continued indefinitely.

[b] If all apparent CNS disease has been excised, the duration of therapy may be reduced to 6 months.

Clinical experience with other oral drugs is limited. Minocycline (100–200 mg twice a day) is often effective; other tetracyclines are usually less effective. Linezolid is active against all species in vitro and has been effective in a few clinical cases, but adverse effects are common with long-term use. Tigecycline appears to be active in vitro against some species, but no relevant clinical experience has been reported. Amoxicillin (875 mg) combined with clavulanic acid (125 mg), given twice a day, has been effective in some cases but should be avoided in cases involving strains of the *N. nova* complex, in which clavulanate induces β-lactamase production. Among quinolones, ciprofloxacin has been studied most often, but moxifloxacin and gemifloxacin now appear to be more active.

Amikacin, the best-established parenteral drug except in cases involving the *N. transvalensis* complex, is given in doses of 5–7.5 mg/kg every 12 h or 15 mg/kg every 24 h. Serum drug levels should be monitored during prolonged therapy in patients with diminished renal function and in the elderly. Cefotaxime, ceftriaxone, and imipenem are usually effective except as indicated in Table 162-1.

Patients with severe disease are initially treated with a combination including TMP-SMX, amikacin, and ceftriaxone or imipenem. Clinical improvement is usually noticeable after 1–2 weeks of therapy but may take longer, especially with CNS disease. After definite clinical improvement, therapy can be continued with a single drug (usually one that can be taken by mouth) in most cases. Some experts use two or more drugs for the entire course of therapy in some cases, but whether multiple drugs are better than a single agent is not known, and additional drugs increase the risk of toxicity. In patients with nocardiosis who need immunosuppressive therapy for an underlying disease or prevention of transplant rejection, immunosuppressive therapy should be continued.

Use of SMX and TMP in high-risk populations to prevent *Pneumocystis* disease or urinary tract infections appears to reduce but not eliminate the risk of nocardiosis. The incidence of nocardiosis is low enough that prophylaxis solely to prevent this disease is not recommended.

Surgical management of nocardial disease is similar to that of other bacterial diseases. Brain abscesses should be aspirated, drained, or excised if the diagnosis is unclear, if an abscess is large and accessible, or if an abscess fails to respond to chemotherapy. Small or inaccessible brain abscesses should be treated medically; clinical improvement should be noticeable within 1–2 weeks. Brain imaging should be repeated to document the resolution of lesions, although abatement on images often lags behind clinical improvement.

Antimicrobial therapy usually suffices for nocardial actinomycetoma. In deep or extensive cases, drainage or excision of heavily involved tissue may facilitate healing, but structure and function should be preserved whenever possible. Keratitis is treated with topical sulfonamide or amikacin drops plus a sulfonamide or an alternative drug given by mouth.

Nocardial infections tend to relapse (particularly in patients with chronic granulomatous disease), and long courses of antimicrobial therapy are necessary (Table 162-2). If disease is unusually extensive or if the response to therapy is slow, the recommendations in Table 162-2 should be exceeded.

With appropriate treatment, the mortality rate for pulmonary or disseminated nocardiosis outside the CNS should be <5%. CNS disease carries a higher mortality rate. Patients should be followed carefully for at least 6 months after therapy has ended.

FURTHER READINGS

Ameen M, Arenas R: Developments in the management of mycetomas. Clin Exp Dermatol 34:1, 2009[PMID: 19076788]

Brown-Elliott BA et al: Clinical and laboratory features of the *Nocardia* spp. based on current molecular taxonomy. Clin Microbiol Rev 19:259, 2006[PMID: 16614249]

Filice GA: Nocardiosis in persons with human immunodeficiency virus infection, transplant recipients, and large, geographically defined populations. J Lab Clin Med 145:156, 2005[PMID: 15871308]

Hearne CB et al: The gardener's cellulitis. Am J Med 122:27, 2009[PMID: 19114167]

Jodlowski TZ et al: Linezolid for the treatment of *Nocardia* spp. infections. Ann Pharmacother 41:1694, 2007[PMID: 17785610]

Khan BA et al: *Nocardia* infection in lung transplant recipients. Clin Transplant 22:562, 2008[PMID: 18435787]

Maraki S et al: Nocardial infection in Crete, Greece: Review of fifteen cases from 2003 to 2007. Scand J Infect Dis 41:122, 2009[PMID: 19101858]

Minero MV et al: Nocardiosis at the turn of the century. Medicine (Baltimore) 88:250, 2009[PMID: 19593231]

Pintado V et al: Infection with *Nocardia* species: Clinical spectrum of disease and species distribution in Madrid, Spain, 1978–2001. Infection 30:338, 2002[PMID: 12478321]

Poonyagariyagorn HK et al: Challenges in the diagnosis and management of *Nocardia* infections in lung transplant recipients. Transpl Infect Dis 10:403, 2008[PMID: 18823356]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACTINOMYCOSIS: INTRODUCTION

Actinomycosis is an indolent, slowly progressive infection caused by anaerobic or microaerophilic bacteria, primarily of the genus *Actinomyces*, that colonize the mouth, colon, and vagina. Mucosal disruption may lead to infection at virtually any site in the body. In vivo growth of actinomycetes usually results in the formation of characteristic clumps called *grains* or *sulfur granules*. The clinical presentations of actinomycosis are myriad. Common in the preantibiotic era, actinomycosis has diminished in incidence, as has its timely recognition. Actinomycosis has been called the most misdiagnosed disease, and it has been said that no disease is so often missed by experienced clinicians. Thus this entity remains a diagnostic challenge.

Three clinical presentations that should prompt consideration of this unique infection are (1) the combination of chronicity, progression across tissue boundaries, and mass-like features (mimicking malignancy, with which it is often confused); (2) the development of a sinus tract, which may spontaneously resolve and recur; and (3) a refractory or relapsing infection after a short course of therapy, since cure of established actinomycosis requires prolonged treatment. An awareness of the full spectrum of the disease, prompting clinical suspicion, will expedite its diagnosis and treatment and will minimize the unnecessary surgical interventions, morbidity, and mortality that are reported all too often.

Etiologic Agents

Actinomycosis is most commonly caused by *A. israelii. A. naeslundii, A. odontolyticus, A. viscosus, A. meyeri*, and *A. gerencseriae* are established but less common causes. Most if not all actinomycotic infections are polymicrobial. *Aggregatibacter (Actinobacillus) actinomycetemcomitans, Eikenella corrodens*, Enterobacteriaceae, and species of *Fusobacterium, Bacteroides, Capnocytophaga, Staphylococcus*, and *Streptococcus* are commonly isolated with actinomycetes in various combinations, depending on the site of infection. The contribution of these other species to the pathogenesis of actinomycosis is uncertain.

Comparative 16S rRNA gene sequencing has led to the identification of an ever-expanding list of *Actinomyces* species and to the reclassification of some actinomycetes as *Arcanobacterium*. Increasing data support the *Actinomyces* species *A. europaeus, A. neuii, A. radingae, A. graevenitzii, A. turicensis, A. cardiffensis, A. houstonensis, A. hongkongensis, A. lingnae,* and *A. funkei* as well as two former *Actinomyces* species now classified as *Arcanobacterium* (*A. pyogenes* and *A. bernardiae*) as additional causes of human actinomycosis.

Epidemiology

Actinomycosis has no geographic boundaries and occurs throughout life, with a peak incidence in the middle decades. Males have a threefold higher incidence than females, possibly because of poorer dental hygiene and/or more frequent trauma. Factors that have probably contributed to the decrease in actinomycosis incidence since the advent of antibiotics include improved dental hygiene and the initiation of antimicrobial treatment before the disease develops fully. Individuals who do not seek or have access to health care, those who have an intrauterine contraceptive device (IUCD) in place for a prolonged period (see "Pelvic Disease," below), and those who receive bisphosphonate treatment (see "Oral-Cervicofacial Disease," below) are probably at higher risk.

Pathogenesis and Pathology

The etiologic agents of actinomycosis are members of the normal oral flora and are often cultured from the bronchi, the gastrointestinal tract, and the female genital tract. The critical step in the development of actinomycosis is disruption of the mucosal barrier. Local infection may ensue. Once established, actinomycosis spreads contiguously in a slow progressive manner, ignoring tissue planes. Although acute inflammation may initially develop at the infection site, the hallmark of actinomycosis is the characteristic chronic, indolent phase manifested by lesions that usually appear as single or multiple indurations. Central necrosis consisting of neutrophils and sulfur granules develops and is virtually diagnostic. The fibrotic walls of the mass are typically described as "wooden." The responsible bacterial and/or host factors have not been identified. Over time, sinus tracts to the skin, adjacent organs, or bone may develop. In rare instances, distant hematogenous seeding may occur. As mentioned above, these unique features of actinomycosis mimic malignancy, with which it is often confused.

Foreign bodies appear to facilitate infection. This association most frequently involves IUCDs. Reports have described an association of actinomycosis with HIV infection; transplantation; treatment with infliximab, glucocorticoids, or bisphosphonates; and radio- or chemotherapy. Ulcerative mucosal infections (e.g., by herpes simplex virus or cytomegalovirus) may facilitate the development of actinomycosis.

Clinical Manifestations

ORAL-CERVICOFACIAL DISEASE

Actinomycosis occurs most frequently at an oral, cervical, or facial site, usually as a soft tissue swelling, abscess, or mass lesion that is often mistaken for a neoplasm. The angle of the jaw is generally involved, but a diagnosis of actinomycosis should be considered with any mass lesion or relapsing infection in the head and neck (Chap. 31). Radiation therapy and especially bisphosphonate treatment have been recognized as contributing to an increasing incidence of actinomycotic infection of the mandible and maxilla (**Fig. 163-1**). Otitis, sinusitis, and canaliculitis (most commonly due to *Propionibacterium propionicum)* also can develop. Pain, fever, and leukocytosis are variably reported. Contiguous extension to the cranium, cervical spine, or thorax is a potential sequela.

Figure 163-1

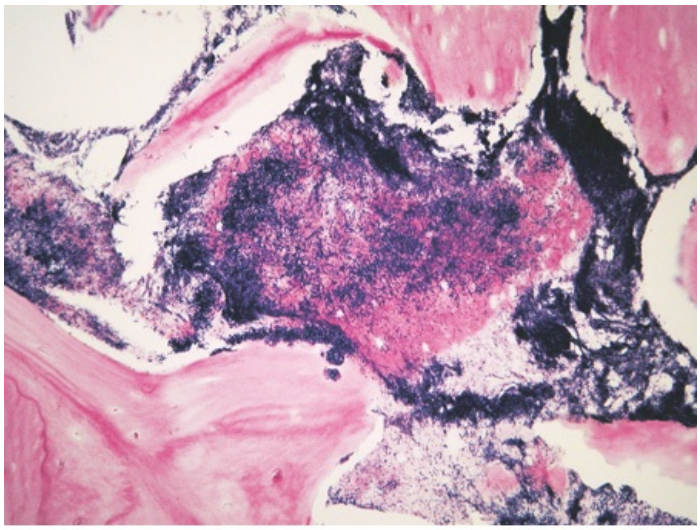

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bisphosphonate-associated maxillary osteomyelitis due to *A. viscosus*. A sulfur granule is seen within the bone. *(Reprinted with permission from NH Naik and TA Russo. © 2009 University of Chicago Press.)*

THORACIC DISEASE

Thoracic actinomycosis usually follows an indolent progressive course, with involvement of the pulmonary parenchyma and/or the pleural space. Chest pain, fever, and weight loss are common. A cough, when present, is variably productive. The usual radiographic finding is either a mass lesion or pneumonia. On CT, central areas of low attenuation and ringlike rim enhancement may be seen. Cavitary disease or hilar adenopathy may develop. More than 50% of cases include pleural thickening, effusion, or empyema (**Fig. 163-2**). Rarely, pulmonary nodules or endobronchial lesions occur. Pulmonary lesions suggestive of actinomycosis may cross fissures or pleura; may involve the mediastinum, contiguous bone, or chest wall; or may be associated with a sinus tract. In the absence of these findings, thoracic actinomycosis is usually mistaken for a neoplasm or for pneumonia due to more usual causes.

Figure 163-2

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

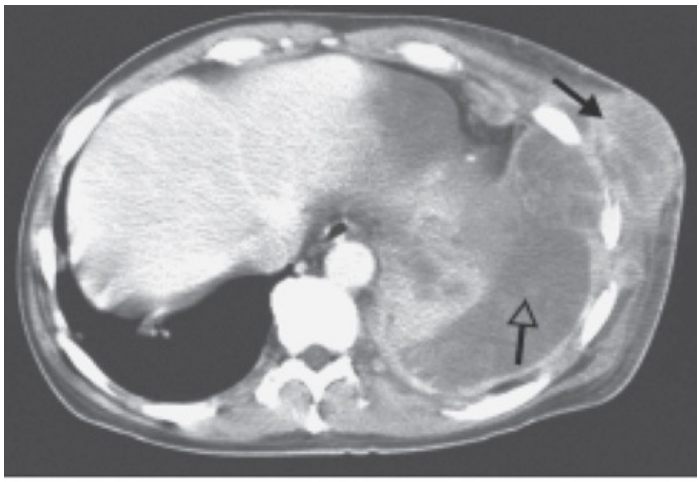

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thoracic actinomycosis. *A.* A chest wall mass from extension of pulmonary infection. *B.* Pulmonary infection is complicated by empyema (*open arrow*) and extension to the chest wall (*closed arrow*). *(Courtesy of Dr. C. B. Hsiao, Division of Infectious Diseases, Department of Medicine, State University of New York at Buffalo.)*

Mediastinal infection is uncommon, usually arising from thoracic extension but rarely resulting from perforation of the esophagus, from trauma, or from head and neck or abdominal disease. The structures within the mediastinum and the heart can be involved in various combinations; consequently, the possible presentations are diverse. Primary endocarditis and isolated disease of the breast have been described.

ABDOMINAL DISEASE

Abdominal actinomycosis poses a great diagnostic challenge. Months or years usually pass from the inciting event (e.g., appendicitis, diverticulitis, peptic ulcer disease, spillage of gall stones or bile during laparoscopic cholecystectomy, foreign-body perforation, bowel surgery, or ascension from IUCD-associated pelvic disease) to clinical recognition. Because of the flow of peritoneal fluid and/or the direct extension of primary disease, virtually any abdominal organ, region, or space can be involved. The disease usually presents as an abscess, a mass, or a mixed lesion that is often fixed to underlying tissue and mistaken for a tumor. On CT, enhancement is most often heterogeneous and adjacent bowel is thickened. Sinus tracts to the abdominal wall, to the perianal region, or between the bowel and other organs may develop and mimic inflammatory bowel disease (Chap. 295). Recurrent disease or a wound or fistula that fails to heal suggests actinomycosis.

Hepatic infection usually presents as one or more abscesses or masses (**Fig. 163-3**). Isolated disease presumably develops via hematogenous seeding from cryptic foci. Imaging and percutaneous techniques have resulted in improved diagnosis and treatment.

Figure 163-3

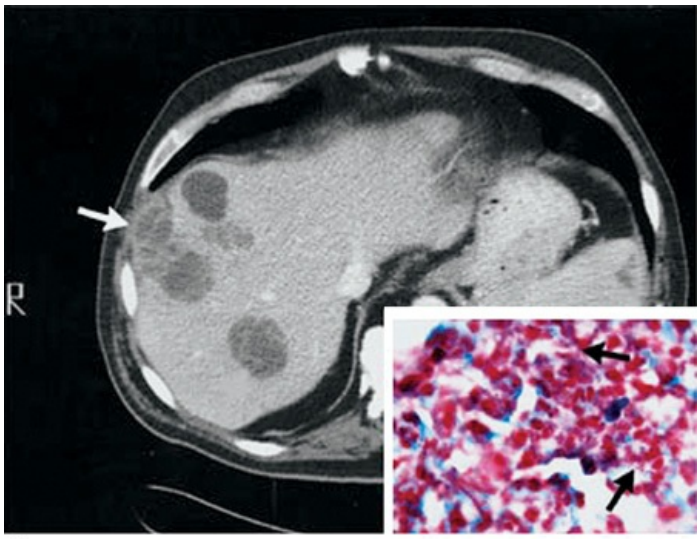

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hepatic-splenic actinomycosis. *A*. Computed tomogram showing multiple hepatic abscesses and a small splenic lesion due to ***A****. israelii.* Arrow indicates extension outside the liver. *Inset:* Gram's stain of abscess fluid demonstrating beaded filamentous gram-positive rods). ***B.*** Subsequent formation of a sinus tract. *(Reprinted with permission from Saad M: Actinomyces hepatic abscess with cutaneous fistula. N Engl J Med 353:e16, 2005. © 2005 Massachusetts Medical Society. All rights reserved.)*

All levels of the urogenital tract can be infected. Renal disease usually presents as pyelonephritis and/or renal and perinephric abscess. Bladder involvement, usually due to extension of pelvic disease, may result in ureteral obstruction or fistulas to bowel, skin, or uterus. *Actinomyces* can be detected in urine with appropriate stains and cultures.

PELVIC DISEASE

Actinomycotic involvement of the pelvis occurs most commonly in association with an IUCD. When an IUCD is in place or has recently been removed, pelvic symptoms should prompt consideration of actinomycosis. The risk, although not quantified, appears small. The disease rarely develops when the IUCD has been in place for <1 year, but the risk increases with time. Actinomycosis can also present months after IUCD removal. Symptoms are typically indolent; fever, weight loss, abdominal pain, and abnormal vaginal bleeding or discharge are the most common. The earliest stage of disease—often endometritis—commonly progresses to pelvic masses or a tuboovarian abscess (**Fig. 163-4**). Unfortunately, because the diagnosis is often delayed, a "frozen pelvis" mimicking malignancy or endometriosis can develop by the time of recognition.

Figure 163-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomogram showing pelvic actinomycosis associated with an intrauterine contraceptive device. The device is encased by endometrial fibrosis (*solid arrow*); also visible are paraendometrial fibrosis (*open triangular arrowhead*) and an area of suppuration (*open arrow*).

Identification of *Actinomyces*-like organisms (ALOs) on Papanicolaou-stained specimens, which occurs on average in 7% of women using an IUCD, has a low positive predictive value for a diagnosis of pelvic infection. Although the risk appears small, the consequences of infection are significant. Therefore, until more quantitative data become available, it seems prudent to remove the IUCD in the presence of symptoms that cannot be accounted for, regardless of whether ALOs are detected, and—if advanced disease is excluded—to initiate a 14-day course of empirical treatment for possible early pelvic actinomycosis. The detection of ALOs in the absence of symptoms warrants education of the patient and close follow-up but not removal of the IUCD unless a suitable contraceptive alternative is agreed on.

CENTRAL NERVOUS SYSTEM DISEASE

Actinomycosis of the central nervous system (CNS) is rare. Single or multiple brain abscesses are most common. An abscess usually appears on CT as a ring-enhancing lesion with a thick wall that may be irregular or nodular. Magnetic resonance perfusion and spectroscopy findings have also been described, as have meningitis, epidural or subdural space infection, and cavernous sinus syndrome.

MUSCULOSKELETAL AND SOFT TISSUE INFECTION

Actinomycotic infection of bone is usually due to adjacent soft-tissue infection but may be associated with trauma (e.g., fracture of the mandible), osteoradionecrosis and bisphosphonate osteonecrosis (limited to mandibular and maxillary bones), or hematogenous spread. Because of slow disease progression, new bone formation and bone destruction are seen concomitantly. Infection of an extremity is uncommon and is usually a result of trauma. Skin, subcutaneous tissue, muscle, and bone (with periostitis or acute or chronic osteomyelitis) are involved alone or in various combinations. Cutaneous sinus tracts frequently develop.

DISSEMINATED DISEASE

Hematogenous dissemination of disease from any location rarely results in multiple-organ involvement. The lungs and liver are most commonly affected, with the presentation of multiple nodules mimicking disseminated malignancy. The clinical presentation may be surprisingly indolent given the extent of disease.

Diagnosis

The diagnosis of actinomycosis is rarely considered. All too often, the first mention of actinomycosis is by the pathologist after extensive surgery. Since medical therapy alone is frequently sufficient for cure, the challenge for the clinician is to consider the possibility of actinomycosis, to diagnose it in the least invasive fashion, and to avoid unnecessary surgery. The clinical and radiographic presentations that suggest actinomycosis are discussed above. Of note, hypermetabolism has been demonstrated by positive emission tomography in actinomycotic disease. Aspirations and biopsies (with or without CT or ultrasound guidance) are being used successfully to obtain clinical material for diagnosis, although surgery may be required. The diagnosis is most commonly made by microscopic identification of sulfur granules (an in vivo matrix of bacteria, calcium phosphate, and host material) in pus or tissues. Occasionally, these granules are identified grossly from draining sinus tracts or pus. Although sulfur

granules are a defining characteristic of actinomycosis, granules are also found in mycetoma (Chaps. 162 and 206) and botryomycosis (a chronic suppurative bacterial infection of soft tissue or, in rare cases, visceral tissue that produces clumps of bacteria resembling granules). These entities can easily be differentiated from actinomycosis with appropriate histopathologic and microbiologic studies. Microbiologic identification of actinomycetes is often precluded by prior antimicrobial therapy or failure to perform appropriate microbiologic cultures. For optimal yield, the avoidance of even a single dose of antibiotics is mandatory. Primary isolation usually requires 5–7 days but may take as long as 2–4 weeks. Although not routinely used, 16S rRNA gene amplification and sequencing have been successfully applied to increase diagnostic sensitivity. Because actinomycetes are components of the normal oral and genital-tract flora, their identification in the absence of sulfur granules in sputum, bronchial washings, and cervicovaginal secretions is of little significance.

Treatment: Actinomycosis

Decisions about treatment are based on the collective clinical experience of the past 50 years. Actinomycosis requires prolonged treatment with high doses of antimicrobial agents. The need for intensive treatment is presumably due to the drugs' poor penetration of the thick-walled masses common in this infection and/or the sulfur granules themselves, which may represent a biofilm. Although therapy must be individualized, the IV administration of 18–24 million units of penicillin daily for 2–6 weeks, followed by oral therapy with penicillin or amoxicillin (total duration, 6–12 months), is a reasonable guideline for serious infections and bulky disease. Less extensive disease, particularly that involving the oral-cervicofacial region, may be cured with a shorter course. If therapy is extended beyond the resolution of measurable disease, the risk of relapse—a clinical hallmark of this infection—will be minimized; CT and MRI are generally the most sensitive and objective techniques by which to accomplish this goal. A similar approach is reasonable for immunocompromised patients, although refractory disease has been described in HIV-infected individuals. Suitable alternative antimicrobial agents and those deemed unreliable are listed in **Table 163-1**. Although the role played by "companion" microbes in actinomycosis is unclear, many isolates are pathogens in their own right, and a regimen covering these organisms during the initial treatment course is reasonable.

Table 163-1 Appropriate and Inappropriate Antibiotic Therapy for Actinomycosis [a]

Category	Agent
Extensive successful clinical experience[b]	Penicillin: 3–4 million units IV q4h Amoxicillin: 500 mg PO q6h Erythromycin: 500–1000 mg IV q6h or 500 mg PO q6h Tetracycline: 500 mg PO q6h Doxycycline: 100 mg IV or PO q12h Minocycline: 100 mg IV or PO q12h Clindamycin: 900 mg IV q8h or 300–450 mg PO q6h
Anecdotal successful clinical experience	Ceftriaxone[c] Ceftizoxime Imipenem-cilastatin Piperacillin-tazobactam
Agents that should be avoided	Metronidazole Aminoglycosides Oxacillin Dicloxacillin Cephalexin
Agents predicted to be efficacious on the basis of in vitro activity	Moxifloxacin Vancomycin Linezolid Quinupristin-dalfopristin Ertapenem[c] Azithromycin[c]

[a]Additional coverage for concomitant “companion” bacteria may be required.

[b]Controlled evaluations have not been performed. Dose and duration require individualization depending on the host, site, and extent of infection. As a general rule, a maximal parenteral antimicrobial dose for 2–6 weeks followed by oral therapy, for a total duration of 6–12 months, is required for serious infections and bulky disease, whereas a shorter course may suffice for less extensive disease, particularly in the oral-cervicofacial region. Monitoring the impact of therapy with CT or MRI is advisable when appropriate.

[c]This agent can be considered for at-home parenteral therapy.

Combined medical-surgical therapy is still advocated in some reports. However, an increasing body of literature now supports an initial attempt at cure with medical therapy alone, even in extensive disease. CT and MRI should be used to monitor the response to therapy. In most cases, either surgery can be avoided or a less extensive procedure can be used. This approach is particularly valuable in sparing critical organs, such as the bladder or the reproductive organs in women of child-bearing age. For a well-defined abscess, percutaneous drainage in combination with medical therapy is a reasonable approach. When a critical location is involved (e.g., the epidural space, the CNS) or when suitable medical therapy fails, surgical intervention may be appropriate.

FURTHER READINGS

Cohen RD et al: Pulmonary actinomycosis complicating infliximab therapy for Crohn's disease. Thorax 62:1013, 2007[PMID: 17965080]

Ho L et al: Actinomycosis mimicking anastomotic recurrent esophageal cancer on PET-CT. Clin Nucl Med 31:646, 2006[PMID: 16985378]

Joshi C et al: Pelvic actinomycosis: A rare entity presenting as a tubo-ovarian abscess. Arch Gynecol Obstet 281:305, 2010[PMID: 19529954]

Naik NH, Russo TA: Bisphosphonate related osteonecrosis of the jaw: The role of *Actinomyces*. Clin Infect Dis 49:1729, 2009[PMID: 19886792]

Russo TA: Agents of actinomycosis, in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). New York, Churchill Livingstone, 2010, pp 3209–3219

Vyas JM et al: Abdominal abscesses due to actinomycosis after laparoscopic cholecystectomy: Case reports and review. Clin Infect Dis 44:e1, 2007

Wang S et al: Actinomycotic brain infection: Registered diffusion, perfusion MR imaging and MR spectroscopy. Neuroradiology 48:346, 2006[PMID: 16614822]

Westhoff C: IUDs and colonization or infection with *Actinomyces*. Contraception 75:S48, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS DUE TO MIXED ANAEROBIC ORGANISMS: INTRODUCTION

Definitions

Anaerobic bacteria are organisms that require reduced oxygen tension for growth, failing to grow on the surface of solid media in 10% CO_2 in air. (In contrast, *microaerophilic bacteria* can grow in an atmosphere of 10% CO_2 in air or under anaerobic or aerobic conditions, although they grow best in the presence of only a small amount of atmospheric oxygen, and *facultative bacteria* can grow in the presence or absence of air.) This chapter describes infections caused by nonsporulating anaerobic bacteria. Most clinically relevant anaerobes, such as *Bacteroides fragilis, Prevotella melaninogenica*, and *Fusobacterium nucleatum*, are relatively aerotolerant. Although they can survive for sustained periods in the presence of up to 2–8% oxygen, generally they do not multiply in this environment. A far smaller number of pathogenic anaerobic bacteria (which are also part of the normal flora) die after brief contact with oxygen, even in low concentrations.

Most human mucocutaneous surfaces harbor a rich indigenous flora composed of aerobic and anaerobic bacteria. These surfaces are dominated by anaerobic bacteria, which often account for 99.0–99.9% of the culturable flora and range in concentration from 10^9/mL in saliva to 10^{12}/mL in gingival scrapings and the colon. Most of the normal anaerobic flora cannot be grown or characterized by current laboratory methods. The major reservoirs of these bacteria are the mouth, lower gastrointestinal tract, skin, and female genital tract **(Table 164-1)**. In the oral cavity, the ratio of anaerobic to aerobic bacteria ranges from 1:1 on the surface of a tooth to 1000:1 in the gingival crevices. Anaerobic bacteria are not found in appreciable numbers in the normal upper intestine until the distal ileum. In the colon, the proportion of anaerobes increases significantly, as does the overall bacterial count; for example, there are 10^{11}–10^{12} organisms per gram of stool, and >99% of these organisms are anaerobic, with an anaerobe-to-aerobe ratio of ~1000:1. In the female genital tract, there are ~10^9 organisms per milliliter of secretions, with an anaerobe-to-aerobe ratio of ~10:1.

Table 164-1 Anaerobic Human Flora: An Overview

Anatomic Site	Total Bacteria[a]	Anaerobic/Aerobic Ratio	Potential Pathogens
Oral cavity			
Saliva Tooth surface Gingival crevices	10^8–10^9 10^{10}–10^{11} 10^{11}–10^{12}	1:1 1:1 10^3:1	*Fusobacterium nucleatum, Prevotella melaninogenica, Prevotella oralis* group, *Bacteroides ureolyticus* group, *Peptostreptococcus* spp.
Gastrointestinal tract			
Stomach Jejunum/ileum Terminal ileum and colon	0–10^5 10^4–10^7 10^{11}–10^{12}	1:1 1:1 10^3:1	*Bacteroides* spp. (principally members of the *B. fragilis* group), *Prevotella* spp., *Clostridium* spp., *Peptostreptococcus* spp.
Female genital tract	10^7–10^9	10:1	*Peptostreptococcus* spp., *Bacteroides* spp., *Prevotella* bivia

[a] Per gram or milliliter.

Commensal anaerobes have been implicated as crucial mediators of physiologic, metabolic, and immunologic functions of the mammalian host. One of the most important roles that anaerobes serve as components of the normal colonic flora is colonization resistance, in which their presence effectively interferes with colonization by potentially pathogenic bacterial species through the depletion of oxygen

and nutrients, the production of enzymes and toxic end products, and the modulation of the host's intestinal innate immune response. *Bacteroides* and other intestinal bacteria ferment carbohydrates and produce volatile fatty acids that are reabsorbed and used by the host as an energy source. The anaerobic intestinal microflora is also responsible for the production of secreted products that promote human health (e.g., vitamin K and bile acids).

The anaerobic intestinal flora influences the development of an intact mucosa and of mucosa-associated lymphoid tissue. Colonization of germ-free mice with a single species, *Bacteroides thetaiotaomicron*, affects the expression of various host genes and corrects deficiencies of nutrient uptake, metabolism, angiogenesis, mucosal barrier function, and enteric nervous system development. The symbiosis factor polysaccharide A of *B. fragilis* influences the normal development and function of the mammalian immune system and protects mice against colitis in a model of inflammatory bowel disease.

Hundreds of species of anaerobic bacteria have been identified as part of the normal flora of humans. Despite the complex array of bacteria in the normal flora, relatively few species are isolated commonly from human infection. Anaerobic infections occur when the harmonious relationship between the host and the bacteria is disrupted. Any site in the body is susceptible to infection with these indigenous organisms when a mucosal barrier or the skin is compromised by surgery, trauma, tumor, ischemia, or necrosis, all of which can reduce local tissue redox potentials. Because the sites that are colonized by anaerobes contain many species of bacteria, disruption of anatomic barriers allows the penetration of many organisms, resulting in mixed infections involving multiple species of anaerobes combined with facultative or microaerophilic organisms. Such mixed infections are seen in the head and neck (chronic sinusitis, chronic otitis media, Ludwig's angina, and periodontal abscesses). Brain abscesses and subdural empyema are the most common anaerobic infections of the central nervous system (CNS). Anaerobes are responsible for pleuropulmonary diseases such as aspiration pneumonia, necrotizing pneumonia, lung abscess, and empyema. These organisms also play an important role in various intraabdominal infections, such as peritonitis and intraabdominal and hepatic abscesses (Chap. 127). They are isolated frequently in female genital tract infections, such as salpingitis, pelvic peritonitis, tuboovarian abscess, vulvovaginal abscess, septic abortion, and endometritis (Chap. 130). Anaerobic bacteria are also found often in bacteremia and in infections of the skin, soft tissues, and bones.

Etiology

The taxonomic classification of anaerobes is rapidly evolving, with frequent changes in nomenclature based on newly discovered relationships among bacterial species. Infections caused by anaerobic bacteria most frequently are due to more than one organism. These polymicrobial infections may be caused by one or several anaerobic species or by a combination of anaerobic organisms and microaerophilic or facultative bacteria acting synergistically. The major anaerobic gram-positive cocci that produce disease are *Peptostreptococcus* species; the major species of this genus that are involved in infections are *P. micros*, *P. magnus*, *P. asaccharolyticus*, *P. anaerobius*, and *P. prevotii*. Clostridia (Chap. 142) are spore-forming gram-positive rods that are isolated from wounds, abscesses, sites of abdominal infection, and blood. Gram-positive anaerobic non-spore-forming bacilli are uncommon as etiologic agents of human infection. *Propionibacterium acnes*, a component of the skin flora and a rare cause of foreign-body infections, is one of the few nonclostridial gram-positive rods associated with infections. The principal anaerobic gram-negative bacilli found in human infections are the *B. fragilis* group as well as *Fusobacterium*, *Prevotella*, and *Porphyromonas* species.

The most important potential anaerobic pathogens found in the upper airways and isolated from clinical specimens of oral and pleuropulmonary infections are the *Fusobacterium* species *F. necrophorum*, *F. nucleatum*, and *F. varium*; *P. melaninogenica*; the *Prevotella oralis* group; *Porphyromonas gingivalis*; *Porphyromonas asaccharolytica*; *Peptostreptococcus* species; and the *Bacteroides ureolyticus* group.

The *B. fragilis* group contains the anaerobic pathogens most frequently isolated from clinical infections. Members of this group are part of the normal bowel flora; they include several distinct species, such as *B. fragilis*, *B. thetaiotaomicron*, *B. vulgatus*, *B. uniformis*, *B. ovatus*, and *Parabacteroides distasonis*. *B. fragilis* is the most important clinical isolate, although it is isolated in lower numbers than some other *Bacteroides* species from cultures of commensal fecal flora.

In female genital tract infections, organisms normally colonizing the vagina (e.g., *Prevotella bivia* and *Prevotella disiens*) are the most common isolates. However, *B. fragilis* is not uncommon.

Pathogenesis

Anaerobic bacterial infections usually occur when an anatomic barrier is disrupted and constituents of the local flora enter a site that was previously sterile. Because of the specific growth requirements of anaerobic organisms and their presence as commensals on mucosal surfaces, conditions must arise that allow these organisms to penetrate mucosal barriers and enter tissue with a lowered oxidation-reduction potential. Therefore, tissue ischemia, trauma, surgery, perforated viscus, shock, and aspiration provide environments conducive to the proliferation of anaerobes. The introduction of many

bacterial species into otherwise-sterile sites leads to a polymicrobial infection in which certain organisms predominate. Three major factors are involved in the pathogenesis of anaerobic infections: bacterial synergy, bacterial virulence factors, and mechanisms of abscess formation. The ability of different anaerobic bacteria to act synergistically during polymicrobial infection contributes to the pathogenesis of anaerobic infections. It has been postulated that facultative organisms function in part to lower the oxidation-reduction potential in the microenvironment, allowing the propagation of obligate anaerobes. Anaerobes can produce compounds such as succinic acid and short-chain fatty acids that inhibit the ability of phagocytes to clear facultative organisms. In experimental models, facultative and obligate anaerobes synergistically potentiate abscess formation. Virulence factors associated with anaerobes typically confer the ability to evade host defenses, adhere to cell surfaces, produce toxins and/or enzymes, or display surface structures such as capsular polysaccharides and lipopolysaccharide (LPS) that contribute to pathogenic potential. The ability of an organism to adhere to host tissues is important to the establishment of infection. Some oral species adhere to the epithelium in the oral cavity. *P. melaninogenica* actually attaches to other microorganisms. *P. gingivalis*, a common isolate in periodontal disease, has fimbriae that facilitate attachment. Some *Bacteroides* strains appear to be piliated, a characteristic that may account for their ability to adhere.

The most extensively studied virulence factor of the nonsporulating anaerobes is the capsular polysaccharide complex of *B. fragilis*. This organism is unique among anaerobes in its potential for virulence during growth at normally sterile sites. Although it constitutes only 0.5–1% of the normal colonic flora, *B. fragilis* is the anaerobe most commonly isolated from intraabdominal infections and bacteremia. One polysaccharide of *B. fragilis*, polysaccharide A, has a unique zwitterionic motif of charged sugars that confers distinct biologic properties, such as the ability to promote abscess formation. Intraabdominal abscess induction is related to the capacity of this polysaccharide to stimulate the release of cytokines and chemokines—in particular, interleukin (IL) 8, IL-17, and tumor necrosis factor α (TNF-α)—from resident peritoneal cells. The release of cytokines and chemokines results in the chemotaxis of polymorphonuclear neutrophils (PMNs) into the peritoneum, where they adhere to mesothelial cells induced by TNF-α to upregulate their expression of intercellular adhesion molecule 1 (ICAM-1). PMNs adherent to ICAM-1-expressing cells probably represent the nidus for an abscess. Polysaccharide A also activates T cells to produce certain cytokines, including IL-17 and interferon γ, that are necessary for abscess formation. Furthermore, when the same polysaccharide is administered to experimental animals prophylactically or therapeutically, it confers protection against abscess induction after challenge with microorganisms capable of inducing abscesses. This protection is mediated by IL-10-producing T cells.

Anaerobic bacteria produce a number of exoproteins that can enhance the organisms' virulence. The collagenase produced by *P. gingivalis* may enhance tissue destruction. An enterotoxin has been identified in *B. fragilis* strains associated with diarrheal disease in animals and young children. Exotoxins produced by clostridial species, including botulinum toxins, tetanus toxin, *C. difficile* toxins A and B, and five toxins produced by *C. perfringens*, are among the most virulent bacterial toxins in mouse lethality assays. Anaerobic gram-negative bacteria such as *B. fragilis* possess LPSs (endotoxins) that are 100–1000 times less biologically potent than endotoxins associated with aerobic gram-negative bacteria. This relative biologic inactivity may account for the lower frequency of disseminated intravascular coagulation and purpura in *Bacteroides* bacteremia than in facultative and aerobic gram-negative bacillary bacteremia. An exception is the LPS from *Fusobacterium*, which may account for the severity of Lemierre's syndrome.

Approach to the Patient: Infections Due to Mixed Anaerobic Organisms

The physician must consider several points when approaching the patient with presumptive infection due to anaerobic bacteria.

1. Most of the organisms colonizing mucosal sites are harmless commensals; very few cause disease. When these organisms do cause disease, it often occurs in proximity to the mucosal site they colonize.
2. For anaerobes to cause tissue infection, they must spread beyond the normal mucosal barriers.
3. Conditions favoring the propagation of these bacteria, particularly a lowered oxidation-reduction potential, are necessary. These conditions exist at sites of trauma, tissue destruction, compromised vascular supply, and complications of preexisting infection, which produce necrosis.
4. There is a complex array of infecting flora. For example, as many as 12 types of organisms can be isolated from a suppurative site.
5. Anaerobic organisms tend to be found in abscess cavities or in necrotic tissue. The failure of an abscess to yield organisms on routine culture is a clue that the abscess is likely to contain anaerobic bacteria. Often smears of this "sterile pus" are found to be teeming with bacteria when Gram's stain is applied. Although some facultative organisms (e.g., *Staphylococcus aureus*) are also capable of causing abscesses, abscesses in organs or deeper body tissues should call to mind anaerobic infection.
6. Gas is found in many anaerobic infections of deep tissues but is not diagnostic because it can be

produced by aerobic bacteria as well.

7. Although a putrid-smelling infection site or discharge is considered diagnostic for anaerobic infection, this manifestation usually develops late in the course and is present in only 30–50% of cases.
8. Some species (the best example being the *B. fragilis* group) require specific therapy. However, many synergistic infections can be cured with antibiotics directed at some but not all of the organisms involved. Antibiotic therapy, combined with debridement and drainage, disrupts the interdependent relationship among the bacteria, and some species that are resistant to the antibiotic do not survive without the co-infecting organisms.
9. Manifestations of severe sepsis and disseminated intravascular coagulation are unusual in patients with purely anaerobic infection.

Epidemiology

Difficulties in the performance of appropriate cultures, contamination of cultures by components of the normal flora, and the lack of readily available, reliable culture techniques have made it impossible to obtain accurate data on incidence or prevalence. However, anaerobic infections are encountered frequently in hospitals with active surgical, trauma, and obstetric and gynecologic services. Depending on the institution, anaerobic bacteria account for 0.5–12% of all cases of bacteremia.

Clinical Manifestations

ANAEROBIC INFECTIONS OF THE MOUTH, HEAD, AND NECK

(See also Chap. 31) Anaerobic bacteria are commonly involved in infections of the mouth, head, and neck. The predominant isolates are components of the normal flora of the upper airways—mainly the *Bacteroides oralis* group, pigmented *Prevotella* species, *P. asaccharolytica*, *Fusobacterium* species, peptostreptococci, and microaerophilic streptococci.

Soft tissue infections of the oral-facial area may or may not be odontogenic. Odontogenic infections —primarily dental caries and periodontal disease (gingivitis and periodontitis)—are common and have both local consequences (especially tooth loss) and the potential for life-threatening spread to the deep fascial spaces of the head and neck. Infections of the mouth can arise from either a supragingival or a subgingival dental plaque composed of bacteria colonizing the tooth surface. Supragingival plaque formation begins with the adherence of gram-positive bacteria to the tooth surface. This form of plaque is influenced by salivary and dietary components, oral hygiene, and local host factors. Supragingival plaque can lead to dental caries and, with further invasion, to pulpitis (endodontic infection) that can further perforate the alveolar bone, causing periapical abscess. Subgingival plaque is associated with periodontal infections (e.g., gingivitis, periodontitis, and periodontal abscess) that can further disseminate to adjacent structures such as the mandible, causing osteomyelitis of the maxillary sinuses. Periodontitis may also result in spreading infection that can involve adjacent bone or soft tissues. In the healthy periodontium, the sparse microflora consists mainly of gram-positive organisms such as *Streptococcus sanguinis* and *Actinomyces* species. In the presence of gingivitis, there is a shift to a greater proportion of anaerobic gram-negative bacilli in the subgingival flora, with predominance of *Prevotella intermedia*. In well-established periodontitis, the complexity of the flora increases further. The predominant isolates are *P. gingivalis*, *P. intermedia*, *Aggregatibacter* (formerly *Actinobacillus*) *actinomycetemcomitans*, *Treponema denticola*, and *Tannerella forsythensis*.

NECROTIZING ULCERATIVE GINGIVITIS

Gingivitis may become a necrotizing infection (trench mouth, Vincent's stomatitis). The onset of disease is usually sudden and is associated with tender bleeding gums, foul breath, and a bad taste. The gingival mucosa, especially the papillae between the teeth, becomes ulcerated and may be covered by a gray exudate, which is removable with gentle pressure. Patients may become systemically ill, developing fever, cervical lymphadenopathy, and leukocytosis. Occasionally, ulcerative gingivitis can spread to the buccal mucosa, the teeth, and the mandible or maxilla, resulting in widespread destruction of bone and soft tissue. This infection is termed acute necrotizing ulcerative mucositis (cancrum oris, noma). It destroys tissue rapidly, causing the teeth to fall out and large areas of bone—or even the whole mandible—to be sloughed. A strong putrid odor is frequently detected, although the lesions are not painful. The gangrenous lesions eventually heal, leaving large disfiguring defects. This infection most commonly follows a debilitating illness or affects severely malnourished children. It has been known to complicate leukemia or to develop in individuals with a genetic deficiency of catalase.

ACUTE NECROTIZING INFECTIONS OF THE PHARYNX

These infections usually occur in association with ulcerative gingivitis. Symptoms include an extremely sore throat, foul breath, and a bad taste accompanied by fever and a sensation of choking. Examination of the pharynx demonstrates that the tonsillar pillars are swollen, red, ulcerated, and covered with a grayish membrane that peels easily. Lymphadenopathy and leukocytosis are common. The disease may last for only a few days or, if not treated, may persist for weeks. Lesions begin

unilaterally but may spread to the other side of the pharynx or the larynx. Aspiration of the infected material by the patient can result in lung abscesses.

PERIPHARYNGEAL SPACE INFECTIONS

These infections arise from the spread of organisms from the upper airways to potential spaces formed by the fascial planes of the head and neck. The etiology is typically polymicrobial and represents the normal flora of the mucosa of the originating site.

Peritonsillar abscess (*quinsy*) is a complication of acute tonsillitis caused mainly by a mixed flora containing anaerobes and group A *Streptococcus*. In submandibular space infection (*Ludwig's angina*), 80% of cases are caused by infection of the tissues surrounding the second and third molar teeth. This infection results in marked local swelling of tissues, with pain, trismus, and superior and posterior displacement of the tongue. Submandibular swelling of the neck can impair swallowing and cause respiratory obstruction. In some cases, tracheotomy may be life-saving. Cervicofacial actinomycosis (Chap. 163) is caused by a branching, gram-positive, non-spore-forming, strict/facultative anaerobe that is a part of the normal oral flora. This chronic disease is characterized by abscesses, draining sinus tracts, fistula, bone destruction, and fibrosis. It can easily be mistaken for malignancy or granulomatous disease. Actinomycosis less frequently involves the thorax, abdomen, pelvis, and CNS.

SINUSITIS AND OTITIS

Anaerobic bacteria have been implicated in chronic sinusitis but play little role in acute sinusitis. In several studies on chronic sinusitis, anaerobic bacteria were found in 12–93% of cases, depending on the method used to collect specimens. Predominant isolates were pigmented *Prevotella, Fusobacterium,* and *Peptostreptococcus* species. Aerobic gram-negative bacilli and *S. aureus* have also been implicated in chronic sinusitis. Polymicrobial infection is common and may be synergistic.

Anaerobic bacteria are much more easily implicated in chronic suppurative otitis media than in acute otitis media. Purulent exudate from chronically draining ears has been found to contain anaerobes, particularly *Bacteroides* species, in up to 50% of cases. *B. fragilis* has been isolated from up to 28% of patients with chronic otitis media.

COMPLICATIONS OF ANAEROBIC HEAD AND NECK INFECTIONS

Contiguous cranial spread of these infections may result in osteomyelitis of the skull or mandible or in intracranial infections such as brain abscess and subdural empyema. Caudal spread can produce mediastinitis or pleuropulmonary infection. Hematogenous complications may also result from anaerobic infections of the head and neck. Bacteremia, which occasionally is polymicrobial, can lead to endocarditis or other distant infections. Lemierre's syndrome, which has been uncommon in the antimicrobial era, is an acute oropharyngeal infection with secondary septic thrombophlebitis of the internal jugular vein and frequent metastasis, most commonly to the lung. *F. necrophorum* is the usual cause. This infection typically begins with pharyngitis, which is followed by local invasion in the lateral pharyngeal space with resultant internal jugular vein thrombophlebitis. A typical clinical triad seen in recent series is pharyngitis, a tender/swollen neck, and noncavitating pulmonary infiltrates.

CNS INFECTIONS

CNS infections associated with anaerobic bacteria are brain abscess (Chap. 381), epidural abscess, and subdural empyema. Anaerobic meningitis is rare and is usually related to parameningeal collection or shunt infection. If optimal bacteriologic techniques are employed, as many as 85% of brain abscesses yield anaerobic bacteria, which usually originate from otorhinolaryngeal infection. However, intraabdominal or pelvic infections can occasionally lead to bacteremia with an anaerobic organism that seeds the cerebral cortex. Commonly isolated are *Peptostreptococcus*, *Fusobacterium*, *Bacteroides*, *Prevotella*, *Propionibacterium, Eubacterium, Veillonella,* and *Actinomyces* species. Facultative or microaerophilic streptococci and coliforms are often part of a mixed infecting flora in brain abscesses.

PLEUROPULMONARY INFECTIONS

Anaerobic pleuropulmonary infections result from the aspiration of oropharyngeal contents, often in the context of an altered state of consciousness or an absent gag reflex. Four clinical syndromes are associated with anaerobic pleuropulmonary infection produced by aspiration: simple aspiration pneumonia, necrotizing pneumonia, lung abscess, and empyema. Many of these infections have an indolent course that may serve as a clinical clue differentiating them, for example, from pneumococcal pneumonia, which often presents with abrupt onset, shaking chills, and rapid progression.

Aspiration Pneumonitis

Bacterial aspiration pneumonitis must be distinguished from two other clinical syndromes associated with aspiration that are not of bacterial etiology. One syndrome results from aspiration of solids, usually food. Obstruction of major airways typically results in atelectasis and moderate nonspecific inflammation. Therapy consists of removal of the foreign body.

The second aspiration syndrome is more easily confused with bacterial aspiration. *Mendelson's syndrome,* a chemical pneumonitis, results from regurgitation of stomach contents and aspiration of

chemical material, usually acidic gastric juices. Pulmonary inflammation—including the destruction of the alveolar lining, with transudation of fluid into the alveolar space—occurs with remarkable rapidity. Typically this syndrome develops within hours, often following anesthesia when the gag reflex is depressed. The patient becomes tachypneic, hypoxic, and febrile. The leukocyte count may rise, and the chest x-ray may evolve suddenly from normal to a complete bilateral "whiteout" within 8–24 h. Sputum production is minimal. The pulmonary signs and symptoms can resolve quickly with symptom-based therapy or can culminate in respiratory failure, with the subsequent development of bacterial superinfection over a period of days. Antibiotic therapy is not indicated unless bacterial infection supervenes.

In contrast to these syndromes, bacterial aspiration pneumonia develops over a period of several days or weeks rather than hours. It is seen in patients who are hospitalized and have a depressed gag reflex, impaired swallowing, or a tracheal or nasogastric tube; elderly patients; and patients with transiently impaired consciousness in the wake of seizures, cerebrovascular accidents, or alcoholic blackouts. Patients who enter the hospital with this syndrome typically have been ill for several days and generally report low-grade fever, malaise, and sputum production. In some patients, weight loss and anemia reflect a more chronic process. Usually the history reveals factors predisposing to aspiration, such as alcohol overdose or residence in a nursing home. Examination sometimes yields evidence of periodontal disease. Sputum characteristically is not malodorous unless the process has been under way for at least a week. A mixed bacterial flora with many PMNs is evident on Gram's staining of sputum. Expectorated sputum is unreliable for anaerobic cultures because of inevitable contamination by normal oral flora. Reliable specimens for culture can be obtained by transtracheal or transthoracic aspiration—techniques that are rarely used at present. Culture of protected-brush specimens or bronchoalveolar lavage fluid obtained by bronchoscopy is controversial.

Chest x-rays show consolidation in dependent pulmonary segments: in the basilar segments of the lower lobes if the patient has aspirated while upright and in either the posterior segment of the upper lobe (usually on the right side) or the superior segment of the lower lobe if the patient has aspirated while supine. The organisms isolated from the lungs reflect the pharyngeal flora; pigmented and nonpigmented *Prevotella* species, *Peptostreptococcus* species, *Bacteroides* species, *Fusobacterium* species, and anaerobic cocci are the most common isolates. While most patients with aspiration pneumonia acquired in the community have a mixed infection caused by anaerobes and aerobic or microaerophilic streptococci, the patient who aspirates in the hospital may also have a mixed infection involving enteric gram-negative rods. In a study on the microbiology of severe aspiration pneumonia in institutionalized elderly patients, gram-negative bacilli were cultured in 49% of cases (with an anaerobe also recovered in 14% of this group), anaerobes in 16%, and *S. aureus* in 12%.

Necrotizing Pneumonitis

This form of anaerobic pneumonitis is characterized by numerous small abscesses that spread to involve several pulmonary segments. The process can be indolent or fulminating. This syndrome is less common than either aspiration pneumonia or lung abscess and includes features of both types of infection.

Anaerobic Lung Abscesses

These abscesses result from subacute anaerobic pulmonary infection. The clinical syndrome typically involves a history of constitutional signs and symptoms (including malaise, weight loss, fever, night sweats, and foul-smelling sputum), perhaps over a period of weeks (Chap. 257). Patients who develop lung abscesses characteristically have dental infection and periodontitis, but lung abscesses in edentulous patients have been reported. Abscess cavities may be single or multiple and generally occur in dependent pulmonary segments **(Fig. 164-1)**. Anaerobic abscesses must be distinguished from lesions associated with tuberculosis, neoplasia, and other conditions. Oral anaerobes predominate and are found in 60–80% of cases. There is also an important role for microaerophilic streptococci such as *S. milleri*. *S. aureus* and enteric gram-negative bacilli may be found as well. Septic pulmonary emboli may originate from intraabdominal or female genital tract infections and can produce anaerobic pneumonia and abscess.

Figure 164-1

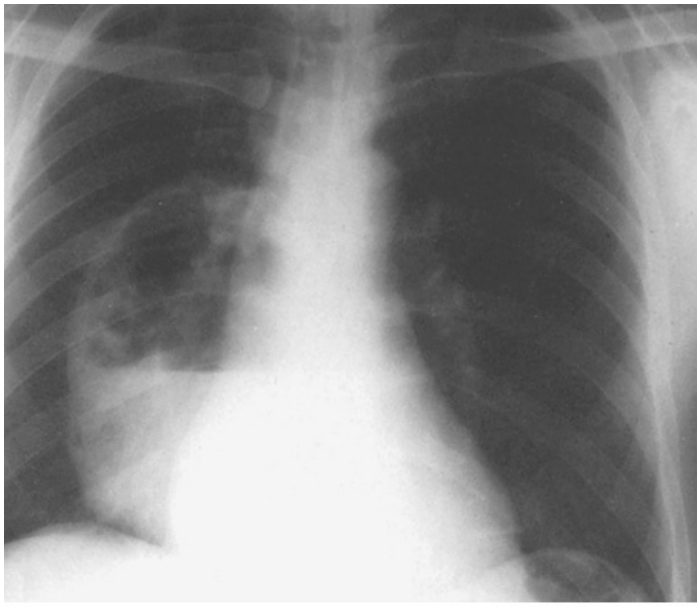

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiograph of right-lower-lobe lung abscess in a 60-year-old alcoholic patient. *[From GL Mandell (ed): Atlas of Infectious Diseases, Vol VI. Philadelphia, Current Medicine Inc, Churchill Livingstone, 1996; with permission.]*

Empyema

Empyema is a manifestation of long-standing anaerobic pulmonary infection. The clinical presentation, which includes foul-smelling sputum, resembles that of other anaerobic pulmonary infections. Patients may report pleuritic chest pain and marked chest-wall tenderness.

Empyema may be masked by overlying pneumonitis and should be considered especially in cases of persistent fever despite antibiotic therapy. Diligent physical examination and the use of ultrasound to localize a loculated empyema are important diagnostic tools. The collection of a foul-smelling exudate by thoracentesis is typical. Cultures of infected pleural fluid yield an average of 3.5 anaerobic and 0.6 facultative or aerobic bacterial species. Drainage is required. Defervescence, a return to a feeling of well-being, and resolution of the process may require several months.

Extension from a subdiaphragmatic infection may also result in anaerobic empyema.

INTRAABDOMINAL INFECTIONS

Intraabdominal infections—mainly peritonitis and abscesses—are usually polymicrobial and represent the normal intestinal (especially colonic) flora. These infections usually follow a breach in the mucosal barrier resulting from appendicitis, diverticulitis, neoplasm, inflammatory bowel disease, surgery, or trauma. On average, four to six bacterial species are isolated per specimen submitted to the microbiology laboratory, with a predominance of enteric aerobic/facultative gram-negative bacilli, anaerobes, and streptococci/enterococci. The most common isolates are *Escherichia coli* (found in ≥50% of patients) and *B. fragilis* (30–50%). Disease originating from proximal-bowel perforation reflects the flora of this site, with a predominance of aerobic and anaerobic gram-positive bacteria and *Candida*.

Enterotoxigenic *B. fragilis* has been associated with watery diarrhea in a few young children and adults. In case-control studies of children with undiagnosed diarrheal disease, enterotoxigenic *B. fragilis* was isolated from significantly more children with diarrhea than children in the control group. Neutropenic enterocolitis (typhlitis) has been associated with anaerobic infection of the cecum but—in the setting of neutropenia (Chap. 86)—may involve the entire bowel. Patients usually present with fever; abdominal pain, tenderness, and distention; and watery diarrhea. The bowel wall is edematous with hemorrhage and necrosis. The primary pathogen is thought by some authorities to be *Clostridium septicum*, but other clostridia and mixed anaerobes have also been implicated. More than 50% of patients developing early clinical signs can benefit from antibiotic therapy and bowel rest. Surgery is sometimes required to remove gangrenous bowel. **See Chap. 127 for a complete discussion of intraabdominal infections.**

PELVIC INFECTIONS

The vagina of a healthy woman is a major reservoir of anaerobic and aerobic bacteria. In the normal flora of the female genital tract, anaerobes outnumber aerobes by a ratio of ~10:1 and include anaerobic gram-positive cocci and *Bacteroides* species (Table 164-1). Anaerobes are isolated from

most women with genital tract infections that are not caused by a sexually transmitted pathogen. The major anaerobic pathogens are *B. fragilis*, *P. bivia*, *P. disiens*, *P. melaninogenica*, anaerobic cocci, and *Clostridium* species. Anaerobes are frequently encountered in Bartholin gland abscess, salpingitis, tuboovarian abscess, septic abortion, pyometra, endometritis, and postoperative wound infection, particularly following hysterectomy. These infections are often of mixed etiology, involving both anaerobes and coliforms; pure anaerobic infections without coliform or other facultative bacterial species occur more often in pelvic than in intraabdominal sites. Suppurative thrombophlebitis of the pelvic veins may complicate the infections and lead to repeated episodes of septic pulmonary emboli. **See Chap. 130 for a complete discussion of pelvic inflammatory disease.**

Anaerobic bacteria have been thought to be contributing factors in the etiology of bacterial vaginosis. This syndrome of unknown etiology is characterized by a profuse malodorous discharge and a change in the bacterial ecology that results in replacement of the *Lactobacillus*-dominated normal flora with an overgrowth of bacterial species including *Gardnerella vaginalis*, *Prevotella* species, *Mobiluncus* species, peptostreptococci, and genital mycoplasmas. A study based on 16S rRNA identification found other anaerobes that were predominant in cases but not in controls: *Atopobium, Leptotrichia, Megasphaera*, and *Eggerthella*. Pelvic infections due to *Actinomyces* species have been associated with the use of intrauterine devices (Chap. 163).

SKIN AND SOFT TISSUE INFECTIONS

Injury to skin, bone, or soft tissue by trauma, ischemia, or surgery creates a suitable environment for anaerobic infections. These infections are most frequently found in sites prone to contamination with feces or with upper airway secretions—e.g., wounds associated with intestinal surgery, decubitus ulcers, or human bites. Moreover, anaerobes have been isolated from cutaneous abscesses, rectal abscesses, and axillary sweat gland infections (hidradenitis suppurativa). Anaerobes are also frequently cultured from foot ulcers of diabetic patients. The deep soft-tissue infections associated with anaerobic bacteria are crepitant cellulitis, synergistic cellulitis, gangrene, and necrotizing fasciitis (Chaps. 125 and 142).

These soft tissue or skin infections are usually polymicrobial. A mean of 4.8 bacterial species are isolated, with an anaerobe-to-aerobe ratio of ~3:2. The most frequently isolated organisms include *Bacteroides*, *Peptostreptococcus*, *Clostridium*, *Enterococcus*, and *Proteus* species. The involvement of anaerobes in these types of infections is associated with a higher frequency of fever, foul-smelling lesions, gas in the tissues, and visible foot ulcer.

Anaerobic bacterial synergistic gangrene (*Meleney's gangrene*), a rare infection of the superficial fascia, is characterized by exquisite pain, redness, and swelling followed by induration. Erythema surrounds a central zone of necrosis. A granulating ulcer forms at the original center as necrosis and erythema extend outward. Symptoms are limited to pain; fever is not typical. These infections usually involve a combination of *Peptostreptococcus* species and *S. aureus*; the usual site of infection is an abdominal surgical wound or the area surrounding an ulcer on an extremity. Treatment includes surgical removal of necrotic tissue and antimicrobial administration.

Necrotizing fasciitis, a rapidly spreading destructive disease of the fascia, is usually attributed to group A streptococci (Chap. 136) but can also be a mixed infection involving anaerobes and aerobes, usually after surgeries and in patients with diabetes or peripheral vascular disease. The most frequently isolated anaerobes in these infections are *Peptostreptococcus* and *Bacteroides* species. Gas may be found in the tissues. Similarly, myonecrosis can be associated with mixed anaerobic infection. *Fournier's gangrene* consists of cellulitis involving the scrotum, perineum, and anterior abdominal wall, with mixed anaerobic organisms spreading along deep external fascial planes and causing extensive loss of skin.

BONE AND JOINT INFECTIONS

Although actinomycosis (Chap. 163) accounts on a worldwide basis for most anaerobic infections in bone, organisms including peptostreptococci or microaerophilic cocci, *Bacteroides* species, *Fusobacterium* species, and *Clostridium* species can also be involved. These infections frequently arise adjacent to soft tissue infections. Hematogenous seeding of bone is uncommon. *Prevotella* and *Porphyromonas* species are detected in infections involving the maxilla and mandible, whereas *Clostridium* species have been reported as anaerobic pathogens in cases of osteomyelitis of the long bones following fracture or trauma. Fusobacteria have been isolated in pure culture from sites of osteomyelitis adjacent to the perinasal sinuses. Peptostreptococci and microaerophilic cocci have been reported as significant pathogens in infections involving the skull, mastoid, and prosthetic implants placed in bone. In patients with osteomyelitis (Chap. 126), the most reliable culture specimen is a bone biopsy sample free of normal uninfected skin and subcutaneous tissue. In patients with anaerobic osteomyelitis, a mixed flora is frequently isolated from a bone biopsy specimen.

In cases of anaerobic septic arthritis, the most common isolates are *Fusobacterium* species. Most of the patients involved have uncontrolled peritonsillar infections progressing to septic cervical venous thrombophlebitis (Lemierre's syndrome) and resulting in hematogenous dissemination with a

predilection for the joints. Unlike anaerobic osteomyelitis, anaerobic pyoarthritis in most cases is not polymicrobial and may be acquired hematogenously. Anaerobes are important pathogens in infections involving prosthetic joints; in these infections, the causative organisms (such as *Peptostreptococcus* species and *P. acnes*) are part of the normal skin flora.

BACTEREMIA

Transient bacteremia is a well-known event in healthy individuals whose anatomic mucosal barriers have been injured (e.g., during dental extractions or dental scaling). These bacteremic episodes, which are often due to anaerobes, have no pathologic consequences. However, anaerobic bacteria are found in cultures of blood from clinically ill patients when proper culture techniques are used. Anaerobes have accounted for 2–5% of all bacteremias, depending on the institution. *B. fragilis* is the single most common anaerobic isolate from the bloodstream, accounting for 35–80% of anaerobic bacteremias. The rate decreased from the 1970s through the early 1990s. This change may be related to the administration of antibiotic prophylaxis before intestinal surgery, the earlier recognition of localized infections, and the empirical use of broad-spectrum antibiotics for presumed infection. However, anaerobic bacteremia may be reemerging. Comparing two periods (1993–1996 and 2001–2004), investigators at the Mayo Clinics found a 74% increase in the incidence of anaerobic bacteremias per 100,000 patient-days; this finding contrasts with a 45% decrease in incidence from 1977 to 1988 at the same institution.

Once the organism in the blood has been identified, both the portal of bloodstream entry and the underlying problem that probably led to seeding of the bloodstream can often be deduced from an understanding of the organism's normal site of residence. For example, mixed anaerobic bacteremia including *B. fragilis* usually implies colonic pathology with mucosal disruption from neoplasia, diverticulitis, or some other inflammatory lesion. The initial manifestations are determined by the portal of entry and reflect the localized condition. When bloodstream invasion occurs, patients can become extremely ill, with rigors and hectic fevers. The clinical picture may be quite similar to that seen in sepsis involving aerobic gram-negative bacilli. Although complications of anaerobic bacteremia (e.g., septic thrombophlebitis and septic shock) have been reported, their incidence in association with anaerobic bacteremia is low. Anaerobic bacteremia is potentially fatal and requires rapid diagnosis and appropriate therapy. The mortality rate appears to increase with the age of the patient (with reported rates of >66% among patients >60 years old), with the isolation of multiple species from the bloodstream, and with the failure to surgically remove a focus of infection. The attributable mortality rate for bacteremia associated with the *B. fragilis* group was examined in a matched case-control study. Patients with *B. fragilis*–group bacteremia had a significantly higher mortality rate (28% vs 8%), with an attributable mortality rate of 19.3% and a mortality risk ratio of 3.2.

ENDOCARDITIS AND PERICARDITIS

(See also Chap. 124) Endocarditis due to anaerobes is uncommon. However, anaerobic streptococci, which are often classified incorrectly, are responsible for this disease more frequently than is generally appreciated. Gram-negative anaerobes are unusual causes of endocarditis. Signs and symptoms of anaerobic endocarditis are similar to those of endocarditis due to facultative organisms. Mortality rates of 21–43% have been reported for anaerobic endocarditis.

Anaerobes, particularly *B. fragilis* and *Peptostreptococcus* species, are uncommonly found in infected pericardial fluids. Anaerobic pericarditis is associated with a mortality rate of >50%. Anaerobes can reach the pericardial space by hematogenous spread, by spread from a contiguous site of infection (e.g., heart or esophagus), or by direct inoculation arising from trauma or surgery.

Diagnosis

There are three critical steps in the diagnosis of anaerobic infection: (1) proper specimen collection; (2) rapid transport of the specimens to the microbiology laboratory, preferably in anaerobic transport media; and (3) proper handling of the specimens by the laboratory. Specimens must be collected by meticulous sampling of infected sites, with avoidance of contamination by the normal flora. When such contamination is likely, the specimen is unacceptable. Examples of specimens unacceptable for anaerobic culture include sputum collected by expectoration or nasal tracheal suction, bronchoscopy specimens, samples collected directly through the vaginal vault, urine collected by voiding, and feces. Specimens appropriate for anaerobic culture include sterile body fluids such as blood, pleural fluid, peritoneal fluid, cerebrospinal fluid, and aspirates or biopsies from normally sterile sites. As a general rule, liquid or tissue specimens are preferred; swab specimens should be avoided.

Because even brief exposure to oxygen may kill some anaerobic organisms and result in failure to isolate them in the laboratory, air must be expelled from the syringe used to aspirate the abscess cavity, and the needle must be capped with a sterile rubber stopper. It is also important to remember that prior antibiotic therapy reduces cultivability of these bacteria. Specimens can be injected into transport bottles containing a reduced medium or taken immediately in syringes to the laboratory for direct culture on anaerobic media. Delays in transport may lead to a failure to isolate anaerobes due to exposure to oxygen or overgrowth of facultative organisms, which may eliminate or obscure any anaerobes that are present. All clinical specimens from suspected anaerobic infections should be

Gram-stained and examined for organisms with characteristic morphology. It is not unusual for organisms to be observed on Gram's staining but not isolated in culture.

Because of the time and difficulty involved in the isolation of anaerobic bacteria, diagnosis of anaerobic infections must frequently be based on presumptive evidence. There are few clinical clues to the probable presence of anaerobic bacteria at infected sites. The involvement of certain sites with lowered oxidation-reduction potential (e.g., avascular necrotic tissues) and the presence of an abscess favor the diagnosis of an anaerobic infection. When infections occur in proximity to mucosal surfaces normally harboring an anaerobic flora, such as the gastrointestinal tract, female genital tract, or oropharynx, anaerobes should be considered as potential etiologic agents. A foul odor is often indicative of anaerobes, which produce certain organic acids as they proliferate in necrotic tissue. Although these odors are nearly pathognomonic for anaerobic infection, the absence of odor does not exclude an anaerobic etiology. Because anaerobes often coexist with other bacteria to cause mixed or synergistic infection, Gram's staining of exudate frequently reveals multiple morphotypes suggestive of anaerobes. Sometimes these organisms have morphologic characteristics associated with specific species.

The presence of gas in tissues is highly suggestive, but not diagnostic, of anaerobic infection. When cultures of obviously infected sites or purulent material yield no growth, streptococci only, or a single aerobic species (such as *E. coli*) and Gram's staining reveals a mixed flora, the involvement of anaerobes should be suspected; the implication is that the anaerobic microorganisms failed to grow because of inadequate transport and/or culture techniques. Failure of an infection to respond to antibiotics that are not active against anaerobes (e.g., aminoglycosides and—in some circumstances—penicillin, cephalosporins, or tetracyclines) suggests an anaerobic etiology.

Treatment: Anaerobic Infections

Successful therapy for anaerobic infections requires the administration of a combination of appropriate antibiotics, surgical resection, debridement of devitalized tissues, and drainage either surgically or percutaneously (guided by an imaging technique such as CT, MRI, or ultrasound). Any anatomic breach must be closed promptly, closed spaces drained, tissue compartments decompressed, and an adequate blood supply established. Abscess cavities should be drained as soon as fluctuation or localization occurs.

ANTIBIOTIC THERAPY AND RESISTANCE

Decisions about the treatment of anaerobic infections with antibiotics are usually based on known resistance patterns in certain species, on the likelihood of encountering a given species in the case at hand, and on Gram's stain findings. Antibiotics active against clinically relevant anaerobes can be grouped into four categories on the basis of their predicted activity **(Table 164-2)**. (Nearly all the drugs listed have toxic side effects, which are described in detail in Chap. 133.) In many infections, anaerobes are mixed with coliforms and other facultative organisms. The best therapeutic regimens, therefore, are usually those active against both aerobic and anaerobic bacteria. The choice of empirical antibiotics for the anaerobes in mixed infections can nearly always be made reliably, since patterns of antimicrobial susceptibility are usually predictable (Chap. 133 and Table 164-2).

Table 164-2 Antimicrobial Therapy for Infections Involving Commonly Encountered Anaerobic Gram-Negative Rods

Category 1 (<2% Resistance)	Category 2 (<15% Resistance)	Category 3 (Variable Resistance)	Category 4 (Resistance)
Carbapenems (imipenem, meropenem, doripenem) Metronidazole[a] β-Lactam/β-lactamase inhibitor combination (ampicillin/sulbactam, ticarcillin/clavulanic acid, piperacillin/tazobactam) Chloramphenicol[b]	Tigecycline High-dose antipseudomonal penicillins	Cephamycins Clindamycin Penicillin Cephalosporins Tetracycline Vancomycin Erythromycin Moxifloxacin	Aminoglycosides Monobactams Trimethoprim-sulfamethoxazole

[a] Usually needs to be given in combination with aerobic bacterial coverage. For infections originating below the diaphragm, aerobic gram-negative coverage is essential. For infections from an oral source, aerobic gram-positive coverage is added. Metronidazole also is not active against Actinomyces, Propionibacterium, or other gram-positive non-spore-forming bacilli (e.g., *Eubacterium, Bifidobacterium*) and is unreliable against peptostreptococci.

[b]Chloramphenicol is probably not as effective as other category 1 antimicrobial agents in treating anaerobic infections.

Antibiotic susceptibility testing of anaerobic bacteria has been difficult and controversial. Owing to the slow growth rate of many anaerobes, the lack of standardized testing methods and of clinically relevant standards for resistance, and the generally good results obtained with empirical therapy, there has been limited interest in testing these organisms for antibiotic susceptibility. However, one study of antibiotic-treated patients with *Bacteroides* isolates from blood found mortality rates of 45% among those whose isolates were deemed resistant to the agent used and 16% among those whose isolates were deemed sensitive. These figures suggest that in vitro susceptibility testing should be performed for *Bacteroides* isolates from hospitalized patients with bacteremia and that the results of this testing should guide treatment. In general, cure rates of >80% can be attained among *Bacteroides*-infected patients with appropriate antimicrobial therapy and drainage. Of the drugs active against most clinically relevant anaerobes, metronidazole, β-lactam/β-lactamase inhibitor combinations, and carbapenems are preferred.

Antibiotic resistance in anaerobic bacteria is an increasing problem. Resistance rates vary with the institution and the geographic region. In recent years, the activity of clindamycin, cefoxitin, cefotetan, and moxifloxacin has decreased against *B. fragilis* and related strains (*B. distasonis*, *B. ovatus*, *B. thetaiotaomicron*, *B. uniformis*, *B. vulgatus*). Nearly all organisms in the *B. fragilis* group (>97%) are resistant to penicillin G. Rates of resistance to β-lactam agents among anaerobes other than *Bacteroides* are lower but highly variable. β-Lactam/β-lactamase inhibitor combinations such as ampicillin/sulbactam, ticarcillin/clavulanic acid, and piperacillin/tazobactam are usually a good therapeutic option, but decreased susceptibility in up to 10% of *B. fragilis*–group isolates has been observed in a study from Belgium. Rates of resistance to the cephamycins (cefoxitin and cefotetan) have varied between 8% and 33% in different surveys. Metronidazole is active against gram-negative anaerobes, including the *B. fragilis* group; resistance is rare but has been reported. Resistance to metronidazole is more common among gram-positive anaerobes, including *P. acnes*, *Actinomyces* species, lactobacilli, and anaerobic streptococci. In the United States, rates of clindamycin resistance among isolates of the *B. fragilis* group increased from 3% in 1982 to 16% in 1996 and 26% in 2000, with figures as high as 44% in some series. Rates of resistance to clindamycin among non-*Bacteroides* anaerobes are much lower (<10%). Carbapenems (ertapenem, doripenem, meropenem, and imipenem) are equally active against anaerobes, with <1% of *B. fragilis* strains showing resistance. Tigecycline is active against some anaerobic bacteria, including *Peptostreptococcus*, *Propionibacterium*, *Prevotella*, *Fusobacterium*, and most *Bacteroides* species. Its efficacy for treatment of intraabdominal infections was comparable to that of imipenem in two phase 2 clinical trials. Low resistance rates (~4%) have been observed. High rates of resistance to moxifloxacin among *Bacteroides* and *Prevotella* species have been reported, ranging up to 32% in a recent survey from Greece.

If a patient fails to respond to one of the category 1 or category 2 drugs (Table 164-2), consideration should be given to alternative therapy and to determination of the resistance patterns among *Bacteroides* isolates. Although in vitro resistance of *Bacteroides* species to chloramphenicol has not been reported, this drug may not be as effective as other category 1 drugs.

INFECTIONS AT SPECIFIC SITES

In clinical situations, specific regimens must be tailored to the initial site of infection. The duration of therapy also depends on the infection site; the reader is referred to specific chapters on sites of infection for recommendations.

Infections above the diaphragm usually reflect the orodental flora, which does not include the *B. fragilis* group. β-Lactamase production has been reported in anaerobic strains that are usually isolated from infections originating above the diaphragm. Up to 60% of clinical isolates classified as *Prevotella* or *Porphyromonas* species, non–*B. fragilis* species of *Bacteroides*, or *Fusobacterium* species reportedly produce β-lactamase; thus all β-lactam drugs (penicillins and cephalosporins) are poor options. Because most of these infections have a mixed etiology that includes microaerophilic and aerobic streptococci, antibiotics that cover both aerobic and anaerobic bacteria are recommended. The recommended regimens include clindamycin, a β-lactam/β-lactamase inhibitor combination, or metronidazole in combination with a drug active against microaerophilic and aerobic streptococci.

Although many oral anaerobic infections and most cases of anaerobic pneumonia still respond to penicillin therapy, some infections due to oral organisms fail to respond to this drug, and in these cases the use of a drug that is effective against penicillin-resistant anaerobes is recommended (Table 164-2). Life-threatening infections involving the anaerobic flora of the mouth, such as space infections of the head and neck, should be treated empirically as if penicillin-resistant anaerobes are involved. Less serious infections involving the oral microflora can be treated with penicillin alone; metronidazole can be added (or clindamycin can be substituted) if the patient responds poorly to penicillin therapy. Bronchoscopy in lung abscess is indicated only to rule out airway obstruction and does not enhance drainage; in any event, it should be delayed until the antimicrobial regimen has begun to affect the

disease process so that the procedure does not spread the infection. Surgery is almost never indicated because of the danger of spilling the abscess contents into the lungs.

Chloramphenicol has been used successfully against anaerobic CNS infections at doses of 30–60 mg/kg per day, with the exact dose depending on the severity of illness. However, penicillin G and metronidazole also cross the blood-brain barrier and are bactericidal for many anaerobic organisms (Chap. 381).

Anaerobic infections arising below the diaphragm (e.g., colonic and intraabdominal infections) must be treated specifically with agents active against *Bacteroides* species (Table 164-2). In intraabdominal sepsis (Chap. 127), the use of antibiotics effective against penicillin-resistant anaerobes has clearly reduced the incidence of postoperative infections and serious infectious complications. Specifically, a drug from category 1 (Table 164-2) must be included for broad-spectrum coverage. Recommended doses for commonly used category 1 drugs are given in **Table 164-3**. Therapy for intraabdominal sepsis must also include drugs active against the gram-negative aerobic flora of the bowel. If the involvement of gram-positive bacteria such as enterococci is suspected, either ampicillin or vancomycin should be added. A meta-analysis of 40 randomized or quasi-randomized controlled trials of 16 antibiotic regimens for secondary peritonitis showed equivalent clinical success for all regimens.

Table 164-3 Doses and Schedules for Treatment of Serious Infections Due to Commonly Encountered Anaerobic Gram-Negative Rods

First-Line Therapy	Dose	Schedule[a]
Metronidazole[b]	500 mg	q6h
Ticarcillin/clavulanic acid	3.1 g	q4h
Piperacillin/tazobactam	3.375 g	q6h
Imipenem	0.5 g	q6h
Meropenem	1.0 g	q8h

[a] See disease-specific chapters for recommendations on duration of therapy.

[b] Should generally be used in conjunction with drugs active against aerobic or facultative organisms.

Note: All drugs are given by the IV route.

Cases of anaerobic osteomyelitis in which a mixed flora is isolated from a bone biopsy specimen should be treated with a regimen that covers all the isolates. When an anaerobic organism is recognized as a major or sole pathogen infecting a joint, the duration of treatment should be similar to that used for arthritis caused by aerobic bacteria (Chap. 334). Therapy includes the management of underlying disease states, the administration of appropriate antimicrobial agents, temporary joint immobilization, percutaneous drainage of effusions, and (usually) the removal of infected prostheses or internal fixation devices. Surgical drainage and debridement procedures such as sequestrectomy are essential for the removal of necrotic tissue that can sustain anaerobic infections.

The outcome of anaerobic bacteremia is significantly better in patients either initially given or switched to appropriate therapy based on known antibiotic susceptibilities.

FAILURE OF THERAPY

Anaerobic infections that fail to respond to treatment or that relapse should be reassessed. Consideration should be given to additional surgical drainage or debridement. Superinfections with resistant gram-negative facultative or aerobic bacteria should be ruled out. The possibility of drug resistance must be entertained; if resistance is involved, repeated cultures may yield the pathogenic organism.

SUPPORTIVE MEASURES

Other supportive measures in the management of anaerobic infections include careful attention to fluid and electrolyte balance (since extensive local edema may lead to hypoalbuminemia), hemodynamic support for septic shock, immobilization of infected extremities, maintenance of adequate nutrition during chronic infections by parenteral hyperalimentation, relief of pain, and anticoagulation with heparin for thrombophlebitis. For patients with severe anaerobic infections of soft tissues, hyperbaric oxygen therapy is advocated by some experts, but its value has not been proven in controlled trials.

FURTHER READINGS

Chung H, Kasper DL: Microbiota-stimulated immune mechanisms to maintain gut homeostasis. Curr Opin Immunol 22:455, 2010[PMID: 20656465]

Cohen-Poradosu R et al: Anaerobic infections: General concepts, in *Principles and Practice of Infectious Diseases*, 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, p 3083

Lassman B et al: Reemergence of anaerobic bacteremia. Clin Infect Dis 44:895, 2007

Mazmanian SK et al: The love-hate relationship between bacterial polysaccharides and the immune system. Nat Rev Immunol 6:849, 2006[PMID: 17024229]

Snydman DR et al: Lessons learned from the anaerobe survey: Historical perspective and review of the most recent data (2005-2007). Clin Infect Dis 50:S26, 2010

Solomkin JS et al: Diagnosis and management of complicated intra-abdominal infections in adults and children: Guidelines by the Surgical Infection Society and the Infectious Diseases Society of America. Clin Infect Dis 50:133, 2010[PMID: 20034345]

Wexler HM: *Bacteroides*: The good, the bad, and the nitty-gritty. Clin Microbiol Rev 20:593, 2007[PMID: 17934076]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TUBERCULOSIS: INTRODUCTION

Tuberculosis (TB), which is one of the oldest diseases known to affect humans and is likely to have existed in prehominids, is a major cause of death worldwide. This disease is caused by bacteria of the *Mycobacterium tuberculosis* complex and usually affects the lungs, although other organs are involved in up to one-third of cases. If properly treated, TB caused by drug-susceptible strains is curable in virtually all cases. If untreated, the disease may be fatal within 5 years in 50–65% of cases. Transmission usually takes place through the airborne spread of droplet nuclei produced by patients with infectious pulmonary TB.

ETIOLOGIC AGENT

Mycobacteria belong to the family Mycobacteriaceae and the order Actinomycetales. Of the pathogenic species belonging to the *M. tuberculosis* complex, the most common and important agent of human disease is *M. tuberculosis*. The complex includes *M. bovis* (the bovine tubercle bacillus —characteristically resistant to pyrazinamide, once an important cause of TB transmitted by unpasteurized milk, and currently the cause of a small percentage of cases worldwide), *M. caprae* (related to *M. bovis*), *M. africanum* (isolated from cases in West, Central, and East Africa), *M. microti* (the "vole" bacillus, a less virulent and rarely encountered organism), *M. pinnipedii* (a bacillus infecting seals and sea lions in the Southern Hemisphere and recently isolated from humans), and *M. canetti* (a rare isolate from East African cases that produces unusual smooth colonies on solid media and is considered closely related to a supposed progenitor type).

M. tuberculosis is a rod-shaped, nonspore-forming, thin aerobic bacterium measuring 0.5 μm by 3 μm. Mycobacteria, including *M. tuberculosis*, are often neutral on Gram's staining. However, once stained, the bacilli cannot be decolorized by acid alcohol; this characteristic justifies their classification as acid-fast bacilli (AFB; **Fig. 165-1**). Acid fastness is due mainly to the organisms' high content of mycolic acids, long-chain cross-linked fatty acids, and other cell-wall lipids. Microorganisms other than mycobacteria that display some acid fastness include species of *Nocardia* and *Rhodococcus*, *Legionella micdadei*, and the protozoa *Isospora* and *Cryptosporidium*. In the mycobacterial cell wall, lipids (e.g., mycolic acids) are linked to underlying arabinogalactan and peptidoglycan. This structure confers very low permeability of the cell wall, thus reducing the effectiveness of most antibiotics. Another molecule in the mycobacterial cell wall, lipoarabinomannan, is involved in the pathogen-host interaction and facilitates the survival of *M. tuberculosis* within macrophages. The complete genome sequence of *M. tuberculosis* comprises 4043 genes encoding 3993 proteins and 50 genes encoding RNAs; its high guanine-plus-cytosine content (65.6%) is indicative of an aerobic "lifestyle." A large proportion of genes are devoted to the production of enzymes involved in cell wall metabolism.

Figure 165-1

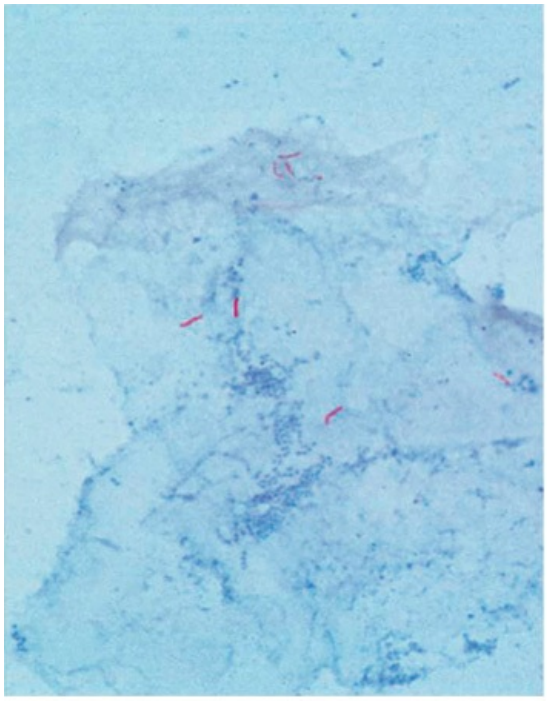

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acid-fast bacillus smear showing *M. tuberculosis* bacilli. *(Courtesy of the CDC, Atlanta.)*

EPIDEMIOLOGY

More than 5.8 million new cases of TB (all forms, both pulmonary and extrapulmonary) were reported to the World Health Organization (WHO) in 2009; 95% of cases were reported from developing countries. However, because of insufficient case detection and incomplete notification, reported cases represent only ~63% (range, 60–67%) of total estimated cases. The WHO estimated that 9.4 million (range, 8.9–9.9 million) new cases of TB occurred worldwide in 2009, 95% of them in developing countries of Asia (5.2 million), Africa (2.8 million), the Middle East (0.7 million), and Latin America (0.3 million). It is further estimated that 1.7 million (range, 1.5–1.9 million) deaths from TB, including 0.4 million among people living with HIV infection, occurred in 2008, 96% of them in developing countries. Estimates of TB incidence rates (per 100,000 population) and numbers of TB-related deaths in 2008 are depicted in Figs. 165-2 and 165-3, respectively. During the late 1980s and early 1990s, numbers of reported cases of TB increased in industrialized countries. These increases were related largely to immigration from countries with a high prevalence of TB; infection with HIV; social problems, such as increased urban poverty, homelessness, and drug abuse; and dismantling of TB services. During the past few years, numbers of reported cases have begun to decline again or stabilized in industrialized nations. In the United States, with the implementation of stronger control programs, the decrease resumed in 1993. In 2009, 11,540 cases of TB (3.8 cases per 100,000 population) were reported to the Centers for Disease Control and Prevention (CDC).

Figure 165-2

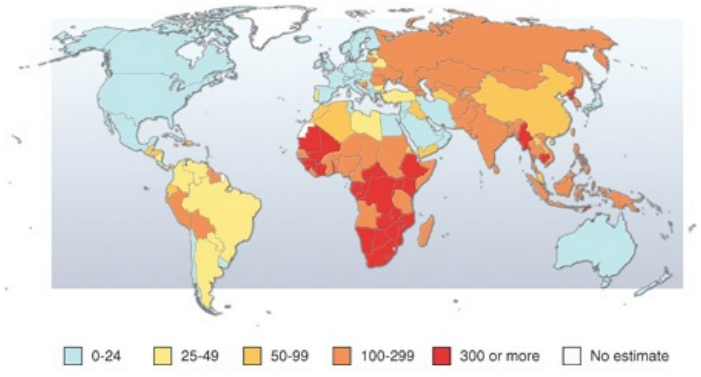

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated tuberculosis incidence rates (per 100,000 population) in 2008. The designations

employed and the presentation of material on this map do not imply the expression of any opinion whatsoever on the part of the WHO concerning the legal status of any country, territory, city, or area or of its authorities or concerning the delimitation of its frontiers or boundaries. White lines on maps represent approximate border lines for which there may not yet be full agreement. *(Courtesy of the Stop TB Department, WHO; with permission.)*

Figure 165-3

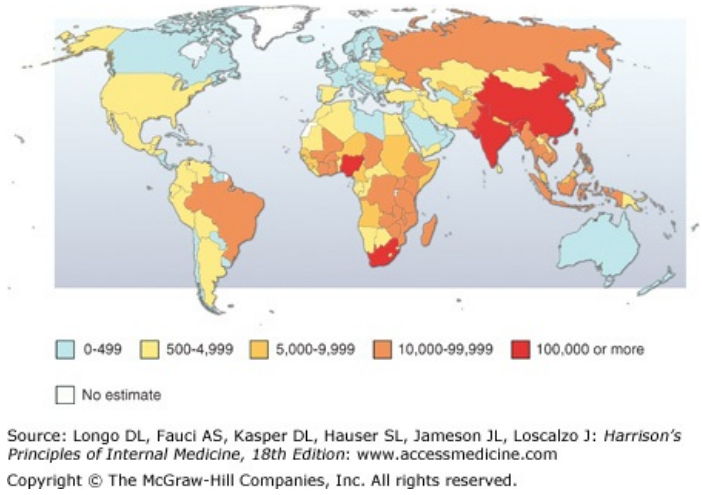

Estimated numbers of tuberculosis-related deaths in 2008. *(See disclaimer in* Fig. 165-2. *Courtesy of the Stop TB Department, WHO; with permission.)*

In the United States, TB is uncommon among young adults of European descent, who have only rarely been exposed to *M. tuberculosis* infection during recent decades. In contrast, because of a high risk of transmission in the past, the prevalence of *M. tuberculosis* infection is relatively high among elderly whites. Blacks, however, account for the highest proportion of cases (41.4% of 4499) among U.S.-born persons. TB in the United States is also a disease of adult members of the HIV-infected population, the foreign-born population (60% of all cases in 2009), and disadvantaged/marginalized populations. Overall, more TB cases were reported among Hispanics than among other ethnic groups; next in frequency were cases among Asians and blacks, with the highest rates per capita among Asians. Similarly, in Europe, TB has reemerged as an important public health problem, mainly as a result of cases among immigrants from high-prevalence countries and among marginalized populations. In many western European countries, there are currently more cases among foreign-born than native populations.

Recent data on global trends indicate that in 2009 TB incidence was stable or falling in most regions; this trend began in 2004 and appears to continue, with an average annual decline of <1% globally. This global decrease is due largely to a reduction (after a peak in 2004) in sub-Saharan Africa, where incidence had risen steeply since the 1980s as a result of the HIVepidemic and the weakness of health systems and services. In eastern Europe, incidence increased during the 1990s because of deterioration in socioeconomic conditions and the health care infrastructure; however, after peaking in 2001, incidence has since declined slowly.

Of the 9.4 million new cases estimated for 2009, 12% (1.1 million) were associated with HIV, and 80% of these HIV-associated cases occurred in Africa. An estimated 0.4 million deaths due to HIV-associated TB occurred in 2008. Furthermore, an estimated 440,000 cases of multidrug-resistant TB (MDR-TB), a form of disease caused by bacilli resistant at least to isoniazid and rifampin, may have emerged in 2008. At present, >90% of these cases are not identified because of a lack of culture and drug-susceptibility testing capacity in most settings worldwide. The independent states of the former Soviet Union have reported the highest rates of MDR-TB among new cases (up to 20% or even higher); several provinces of China follow, with peaks of 10%. Overall, 60% of all MDR-TB cases are in India, China, and the Russian Federation. Starting in 2006, 58 countries, including the United States, reported cases of extensively drug-resistant TB (XDR-TB), in which MDR-TB is compounded by additional resistance to the most powerful second-line anti-TB drugs (fluoroquinolones and at least one of the injectable drugs amikacin, kanamycin, and capreomycin). Probably ~10% of the MDR-TB cases worldwide are XDR-TB, but the vast majority of XDR cases remain undiagnosed.

From Exposure to Infection

M. tuberculosis is most commonly transmitted from a person with infectious pulmonary TB to others by droplet nuclei, which are aerosolized by coughing, sneezing, or speaking. The tiny droplets dry rapidly; the smallest (<5–10 μm in diameter) may remain suspended in the air for several hours and may reach the terminal air passages when inhaled. There may be as many as 3000 infectious nuclei per cough.

Other routes of transmission of tubercle bacilli (e.g., through the skin or the placenta) are uncommon and of no epidemiologic significance. The probability of contact with a person who has an infectious form of TB, the intimacy and duration of that contact, the degree of infectiousness of the case, and the shared environment in which the contact takes place are all important determinants of the likelihood of transmission. Several studies of close-contact situations have clearly demonstrated that TB patients whose sputum contains AFB visible by microscopy are the most likely to transmit the infection. The most infectious patients have cavitary pulmonary disease or, much less commonly, laryngeal TB and produce sputum containing as many as 10^5–10^7 AFB/mL. Patients with sputum smear–negative/culture-positive TB are less infectious, although they have been responsible for up to 20% of transmission in some studies in the United States, and those with culture-negative pulmonary TB and extrapulmonary TB are essentially noninfectious. Because persons with both HIV infection and TB are less likely to have cavitations, they may be less infectious than persons without HIV co-infection. Crowding in poorly ventilated rooms is one of the most important factors in the transmission of tubercle bacilli, since it increases the intensity of contact with a case.

In short, the risk of acquiring *M. tuberculosis* infection is determined mainly by exogenous factors. Because of delays in seeking care and in making a diagnosis, it is generally believed that, in high-prevalence settings, up to 20 contacts may be infected by each AFB-positive case before the index case is found to have TB.

From Infection to Disease

Unlike the risk of acquiring infection with *M. tuberculosis*, the risk of developing disease after being infected depends largely on endogenous factors, such as the individual's innate immunologic and nonimmunologic defenses and level of function of cell-mediated immunity (CMI). Clinical illness directly following infection is classified as *primary TB* and is common among children in the first few years of life and among immunocompromised persons. Although primary TB may be severe and disseminated, it is not generally associated with high-level transmissibility. When infection is acquired later in life, the chance is greater that the mature immune system will contain it at least temporarily. Dormant bacilli, however, may persist for years before reactivating to produce *secondary* (or *postprimary*) *TB*, which, because of frequent cavitation, is more often infectious than is primary disease. Overall, it is estimated that up to 10% of infected persons will eventually develop active TB in their lifetime, with half of them doing so during the first year after infection. The risk is much higher among HIV-infected persons. Reinfection of a previously infected individual, which is common in areas with high rates of TB transmission, may also favor the development of disease. At the height of the TB resurgence in the United States in the early 1990s, molecular typing and comparison of strains of *M. tuberculosis* suggested that up to one-third of cases of active TB in some inner-city communities were due to recent transmission rather than to reactivation of latent infection. Age is an important determinant of the risk of disease after infection. Among infected persons, the incidence of TB is highest during late adolescence and early adulthood; the reasons are unclear. The incidence among women peaks at 25–34 years of age. In this age group rates among women may be higher than those among men, while at older ages the opposite is true. The risk increases in the elderly, possibly because of waning immunity and comorbidity.

A variety of diseases and conditions favor the development of active TB **(Table 165-1)**. In absolute terms, the most potent risk factor for TB among infected individuals is clearly HIV co-infection, which suppresses cellular immunity. The risk that latent *M. tuberculosis* infection will proceed to active disease is directly related to the patient's degree of immunosuppression. In a study of HIV-infected, tuberculin skin test (TST)–positive persons, this risk varied from 2.6 to 13.3 cases per 100 person-years and increased as the CD4+ T cell count decreased.

Table 165-1 Risk Factors for Active Tuberculosis among Persons Who Have Been Infected with Tubercle Bacilli

Factor	Relative Risk/Odds[a]
Recent infection (<1 year)	12.9
Fibrotic lesions (spontaneously healed)	2–20
Comorbidity	
HIV infection	21–>30
Silicosis	30
Chronic renal failure/hemodialysis	10–25
Diabetes	2–4
IV drug use	10–30
Immunosuppressive treatment	10
Gastrectomy	2–5

Jejunoileal bypass	30–60
Posttransplantation period (renal, cardiac)	20–70
Tobacco smoking	2–3
Malnutrition and severe underweight	2

[a] Old infection = 1.

Natural History of Disease

Studies conducted in various countries before the advent of chemotherapy showed that untreated TB is often fatal. About one-third of patients died within 1 year after diagnosis, and more than 50% died within 5 years. The 5-year mortality rate among sputum smear–positive cases was 65%. Of the survivors at 5 years, ~60% had undergone spontaneous remission, while the remainder were still excreting tubercle bacilli. With effective, timely, and proper chemotherapy, patients have a very high chance of being cured. However, improper use of anti-TB drugs, while reducing mortality rates, may also result in large numbers of chronic infectious cases, often with drug-resistant bacilli.

PATHOGENESIS AND IMMUNITY

Infection and Macrophage Invasion

The interaction of *M. tuberculosis* with the human host begins when droplet nuclei containing microorganisms from infectious patients are inhaled. While the majority of inhaled bacilli are trapped in the upper airways and expelled by ciliated mucosal cells, a fraction (usually <10%) reach the alveoli. There, alveolar macrophages that have not yet been activated phagocytize the bacilli. Adhesion of mycobacteria to macrophages results largely from binding of the bacterial cell wall with a variety of macrophage cell-surface molecules, including complement receptors, the mannose receptor, the immunoglobulin GFcγ receptor, and type A scavenger receptors. Phagocytosis is enhanced by complement activation leading to opsonization of bacilli with C3 activation products such as C3b. After a phagosome forms, the survival of *M. tuberculosis* within it seems to depend on reduced acidification due to lack of accumulation of vesicular proton-adenosine triphosphatase. A complex series of events is probably generated by the bacterial cell-wall glycolipid lipoarabinomannan. This glycolipid inhibits the intracellular increase of Ca^{2+}. Thus the Ca^{2+}/calmodulin pathway (leading to phagosome-lysosome fusion) is impaired, and the bacilli may survive within the phagosomes. The *M. tuberculosis* phagosome has been found to inhibit the production of phosphatidylinositol 3-phosphate (PI3P). Normally, PI3P earmarks phagosomes for membrane sorting and maturation including phagolysosome formation, which would destroy the bacteria. Bacterial factors have also been found to block the newly identified host defense of autophagy, in which the cell sequesters the phagosome in a double-membrane vesicle (*autophagosome*) that is destined to fuse with lysosomes. If the bacilli are successful in arresting phagosome maturation, then replication begins and the macrophage eventually ruptures and releases its bacillary contents. Other uninfected phagocytic cells are then recruited to continue the infection cycle by ingesting dying macrophages and their bacillary content, thus in turn becoming infected themselves and expanding the infection.

Virulence of Tubercle Bacilli

Since the elucidation of the *M. tuberculosis* genome in 1998, large mutant collections have been generated, and many bacterial genes that contribute to *M. tuberculosis* virulence have been found. Different patterns of virulence defects have been defined in various animal models, predominantly mice but also guinea pigs, rabbits, and nonhuman primates. The *katG* gene encodes for a catalase/peroxidase enzyme that protects against oxidative stress and is required for isoniazid activation and subsequent bactericidal activity. Region of difference 1 (RD1) is a 9.5-kb locus that encodes two key small protein antigens—early secretory antigen-6 (ESAT-6) and culture filtrate protein-10 (CFP-10)—as well as a putative secretion apparatus that may facilitate their egress; the absence of this locus in the vaccine strain *M. bovis* bacille Calmette-Guérin (BCG) has been shown to be a key attenuating mutation. A recent observation in *Mycobacterium marinum*, the validity of which needs to be confirmed in *M. tuberculosis*, showed that a mutation in the RD1 virulence locus encoding the ESX1 secretion system impairs the capacity of apoptotic macrophages to recruit uninfected cells for further rounds of infection. The results are less replication and fewer new granulomas. Mutants lacking key enzymes of bacterial biosynthesis become auxotrophic for the missing substrate and are often totally unable to proliferate in animals; these include the *leuD* and *panCD* mutants, which require leucine and pantothenic acid, respectively. The isocitrate lyase gene *icl1* encodes a key step in the glyoxylate shunt that facilitates bacterial growth on fatty acid substrates; this gene is required for long-term persistence of *M. tuberculosis* infection in mice with chronic TB. *M. tuberculosis* mutants in regulatory genes such as sigma factor C and sigma factor H (*sigC* and *sigH*) are associated with normal bacterial growth in mice, but they fail to elicit full tissue pathology. Finally, the recently identified mycobacterial protein CarD (expressed by the *carD* gene) seems essential for the control of rRNA transcription that is required for replication and persistence in the host cell. Its loss exposes

mycobacteria to oxidative stress, starvation, DNA damage, and ultimately sensitivity to killing by a variety of host mutagens and defensive mechanisms.

Innate Resistance to Infection

Several observations suggest that genetic factors play a key role in innate nonimmune resistance to infection with *M. tuberculosis* and the development of disease. The existence of this resistance, which is polygenic in nature, is suggested by the differing degrees of susceptibility to TB in different populations. In mice, a gene called *Nramp1* (natural resistance–associated macrophage protein 1) plays a regulatory role in resistance/susceptibility to mycobacteria. The human homologue NRAMP1, which maps to chromosome 2q, may play a role in determining susceptibility to TB, as is suggested by a study among West Africans. Recent studies of mouse genetics identified a novel host resistance gene, *ipr1*, that is encoded within the *sst1* locus; *ipr1* encodes an interferon (IFN)-inducible nuclear protein that interacts with other nuclear proteins in macrophages primed with IFNs or infected by *M. tuberculosis*. In addition, polymorphisms in multiple genes, such as those encoding for various histocompatibility leukocyte antigen (HLA) alleles, IFN-γ, T cell growth factor β, interleukin (IL) 10, mannose-binding protein, IFN-γ receptor, Toll-like receptor 2, vitamin D receptor, and IL-1, have been associated with susceptibility to TB.

The Host Response and Granuloma Formation

In the initial stage of host-bacterium interaction, prior to the onset of an acquired CMI response, *M. tuberculosis* undergoes a period of extensive growth within naïve unactivated macrophages, and additional naïve macrophages are recruited to the early granuloma. Studies suggest that *M. tuberculosis* uses a specific virulence mechanism to subvert host cellular signaling and to elicit an early proinflammatory response that promotes granuloma expansion and bacterial growth during this key early phase. A recent study of *M. marinum* infection in zebrafish has delineated the likely molecular mechanism by which mycobacteria induce granuloma formation. The mycobacterial protein ESAT-6 induces secretion of matrix metalloproteinase 9 (MMP9) by nearby epithelial cells that are in contact with infected macrophages. MMP9 in turn stimulates recruitment of naïve macrophages, thus inducing granuloma maturation and bacterial growth. Disruption of MMP9 function results in reduced bacterial growth. Another study has shown that *M. tuberculosis*–derived cyclic AMP is secreted from the phagosome into host macrophages, subverting the cell's signal transduction pathways and stimulating an elevation in the secretion of tumor necrosis factor α (TNF-α) and further proinflammatory cell recruitment. Ultimately, the chemoattractants and bacterial products released during the repeated rounds of cell lysis and infection of newly arriving macrophages enable dendritic cells to access bacilli; these cells migrate to the draining lymph nodes and present mycobacterial antigens to T lymphocytes. At this point, the development of CMI and humoral immunity begins. These initial stages of infection are usually asymptomatic.

About 2–4 weeks after infection, two host responses to *M. tuberculosis* develop: a macrophage-activating CMI response and a tissue-damaging response. The *macrophage-activating response* is a T cell–mediated phenomenon resulting in the activation of macrophages that are capable of killing and digesting tubercle bacilli. The *tissue-damaging response* is the result of a delayed-type hypersensitivity (DTH) reaction to various bacillary antigens; it destroys unactivated macrophages that contain multiplying bacilli but also causes caseous necrosis of the involved tissues (see below). Although both of these responses can inhibit mycobacterial growth, it is the balance between the two that determines the form of TB that will develop subsequently.

With the development of specific immunity and the accumulation of large numbers of activated macrophages at the site of the primary lesion, granulomatous lesions (tubercles) are formed. These lesions consist of accumulations of lymphocytes and activated macrophages that evolve toward epithelioid and giant cell morphologies. Initially, the tissue-damaging response can limit mycobacterial growth within macrophages. As stated above, this response, mediated by various bacterial products, not only destroys macrophages but also produces early solid necrosis in the center of the tubercle. Although *M. tuberculosis* can survive, its growth is inhibited within this necrotic environment by low oxygen tension and low pH. At this point, some lesions may heal by fibrosis, with subsequent calcification, whereas inflammation and necrosis occur in other lesions. Some observations have challenged the traditional view that any encounter between mycobacteria and macrophages results in chronic infection. It is possible that an immune response capable of eradicating early infection may sometimes develop as a consequence, for instance, of disabling mutations in mycobacterial genomes rendering their replication ineffective.

Macrophage-Activating Response

CMI is critical at this early stage. In the majority of infected individuals, local macrophages are activated when bacillary antigens processed by macrophages stimulate T lymphocytes to release a variety of lymphokines. These activated macrophages aggregate around the lesion's center and effectively neutralize tubercle bacilli without causing further tissue destruction. In the central part of the lesion, the necrotic material resembles soft cheese (*caseous necrosis*)—a phenomenon that may also be observed in other conditions, such as neoplasms. Even when healing takes place, viable bacilli may

remain dormant within macrophages or in the necrotic material for many years. These "healed" lesions in the lung parenchyma and hilar lymph nodes may later undergo calcification.

Delayed-Type Hypersensitivity

In a minority of cases, the macrophage-activating response is weak, and mycobacterial growth can be inhibited only by intensified DTH reactions, which lead to lung tissue destruction. The lesion tends to enlarge further, and the surrounding tissue is progressively damaged. At the center of the lesion, the caseous material liquefies. Bronchial walls as well as blood vessels are invaded and destroyed, and cavities are formed. The liquefied caseous material, containing large numbers of bacilli, is drained through bronchi. Within the cavity, tubercle bacilli multiply, spill into the airways, and are discharged into the environment through expiratory maneuvers such as coughing and talking. In the early stages of infection, bacilli are usually transported by macrophages to regional lymph nodes, from which they gain access to the central venous return; from there they reseed the lungs and may also disseminate beyond the pulmonary vasculature throughout the body via the systemic circulation. The resulting extrapulmonary lesions may undergo the same evolution as those in the lungs, although most tend to heal. In young children with poor natural immunity, hematogenous dissemination may result in fatal miliary TB or tuberculous meningitis.

Role of Macrophages and Monocytes

While CMI confers partial protection against *M. tuberculosis*, humoral immunity plays a less well-defined role in protection (although evidence is accumulating on the existence of antibodies to lipoarabinomannan, which may prevent dissemination of infection in children). In the case of CMI, two types of cells are essential: macrophages, which directly phagocytize tubercle bacilli, and T cells (mainly CD4+ T lymphocytes), which induce protection through the production of cytokines, especially IFN-γ. After infection with *M. tuberculosis*, alveolar macrophages secrete various cytokines responsible for a number of events (e.g., the formation of granulomas) as well as systemic effects (e.g., fever and weight loss). Monocytes and macrophages attracted to the site are key components of the immune response. Their primary mechanism is probably related to production of nitric oxide, which has antimycobacterial activity and increases the synthesis of cytokines such as TNF-α and IL-1, which in turn regulate the release of reactive nitrogen intermediates. In addition, macrophages can undergo apoptosis—a defensive mechanism to prevent release of cytokines and bacilli via their sequestration in the apoptotic cell.

Role of T Lymphocytes

Alveolar macrophages, monocytes, and dendritic cells are also critical in processing and presenting antigens to T lymphocytes, primarily CD4+ and CD8+ T cells; the result is the activation and proliferation of CD4+ T lymphocytes, which are crucial to the host's defense against *M. tuberculosis*. Qualitative and quantitative defects of CD4+ T cells explain the inability of HIV-infected individuals to contain mycobacterial proliferation. Activated CD4+ T lymphocytes can differentiate into cytokine-producing T_H1 or T_H2 cells. T_H1 cells produce IFN-γ—an activator of macrophages and monocytes—and IL-2. T_H2 cells produce IL-4, IL-5, IL-10, and IL-13 and may also promote humoral immunity. The interplay of these various cytokines and their cross-regulation determine the host's response. The role of cytokines in promoting intracellular killing of mycobacteria, however, has not been entirely elucidated. IFN-γ may induce the generation of reactive nitrogen intermediates and regulate genes involved in bactericidal effects. TNF-α also seems to be important. Observations made originally in transgenic knockout mice and more recently in humans suggest that other T cell subsets, especially CD8+ T cells, may play an important role. CD8+ T cells have been associated with protective activities via cytotoxic responses and lysis of infected cells as well as with production of IFN-γ and TNF-α. Finally, natural killer cells act as co-regulators of CD8+ T cell lytic activities, and γ^{δ} T cells are increasingly thought to be involved in protective responses in humans.

Mycobacterial Lipids and Proteins

Lipids have been involved in mycobacterial recognition by the innate immune system, and lipoproteins (such as 19-kDa lipoprotein) have been proven to trigger potent signals through Toll-like receptors present in blood dendritic cells. *M. tuberculosis* possesses various protein antigens. Some are present in the cytoplasm and cell wall; others are secreted. That the latter are more important in eliciting a T lymphocyte response is suggested by experiments documenting the appearance of protective immunity in animals after immunization with live, protein-secreting mycobacteria. Among the antigens that may play a protective role are the 30-kDa (or 85B) and ESAT-6 antigens. Protective immunity is probably the result of reactivity to many different mycobacterial antigens.

Skin Test Reactivity

Coincident with the appearance of immunity, DTH to *M. tuberculosis* develops. This reactivity is the basis of the TST, which is used primarily for the detection of *M. tuberculosis* infection in persons without symptoms. The cellular mechanisms responsible for TST reactivity are related mainly to previously sensitized CD4+ T lymphocytes, which are attracted to the skin-test site. There, they proliferate and produce cytokines. While DTH is associated with protective immunity (TST-positive

persons being less susceptible to a new *M. tuberculosis* infection than TST-negative persons), it by no means guarantees protection against reactivation. In fact, cases of active TB are often accompanied by strongly positive skin-test reactions. There is also evidence of reinfection with a new strain of *M. tuberculosis* in patients previously treated for active disease. This evidence underscores the fact that previous latent or active TB may not confer fully protective immunity.

CLINICAL MANIFESTATIONS

TB is classified as pulmonary, extrapulmonary, or both. Before the advent of HIV infection, ~80% of all new cases of TB were limited to the lungs. However, up to two-thirds of HIV-infected patients with TB may have both pulmonary and extrapulmonary TB or extrapulmonary TB alone.

Pulmonary TB

Pulmonary TB can be conventionally categorized as primary or postprimary (adult-type, secondary). This distinction has been challenged by molecular evidence from TB-endemic areas indicating that a large percentage of cases of adult pulmonary TB result from recent infection (either primary infection or reinfection) and not from reactivation.

PRIMARY DISEASE

Primary pulmonary TB occurs soon after the initial infection with tubercle bacilli. It may be asymptomatic or present with fever and occasionally pleuritic chest pain. In areas of high TB transmission, this form of disease is often seen in children. Because most inspired air is distributed to the middle and lower lung zones, these areas of the lungs are most commonly involved in primary TB. The lesion forming after initial infection (the Ghon focus) is usually peripheral and accompanied by transient hilar or paratracheal lymphadenopathy, which may not be visible on standard chest radiography. Some patients develop erythema nodosum in the legs or phlyctenular conjunctivitis. In the majority of cases, the lesion heals spontaneously and only becomes evident as a small calcified nodule. Pleural reaction overlying a subpleural focus is also common. The Ghon focus, with or without overlying pleural reaction, thickening, and regional lymphadenopathy, is referred to as the *Ghon complex*.

In young children with immature CMI and in persons with impaired immunity (e.g., those with malnutrition or HIV infection), primary pulmonary TB may progress rapidly to clinical illness. The initial lesion increases in size and can evolve in different ways. Pleural effusion, which is found in up to two-thirds of cases, results from the penetration of bacilli into the pleural space from an adjacent subpleural focus. In severe cases, the primary site rapidly enlarges, its central portion undergoes necrosis, and cavitation develops (*progressive primary TB*). TB in young children is almost invariably accompanied by hilar or paratracheal lymphadenopathy due to the spread of bacilli from the lung parenchyma through lymphatic vessels. Enlarged lymph nodes may compress bronchi, causing total obstruction with distal collapse, partial obstruction with large-airway wheezing, or a ball-valve effect with segmental/lobar hyperinflation. Lymph nodes may also rupture into the airway with development of pneumonia, often including areas of necrosis and cavitation, distal to the obstruction. Bronchiectasis may develop in any segment/lobe damaged by progressive caseating pneumonia. Occult hematogenous dissemination commonly follows primary infection. However, in the absence of a sufficient acquired immune response, which usually contains the infection, disseminated or miliary disease may result (**Fig. 165-4**). Small granulomatous lesions develop in multiple organs and may cause locally progressive disease or result in tuberculous meningitis; this is a particular concern in very young children and immunocompromised persons (e.g., patients with HIV infection).

Figure 165-4

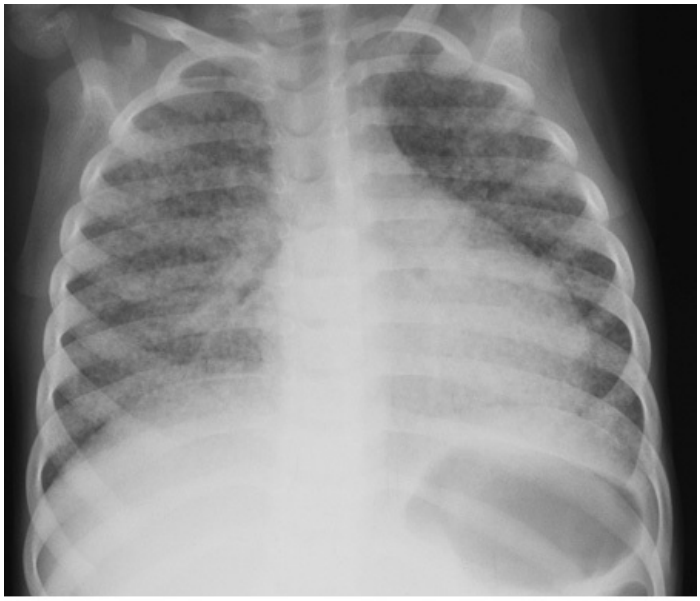

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiograph showing bilateral miliary (millet-sized) infiltrates in a child. *(Courtesy of Prof. Robert Gie, Department of Paediatrics and Child Health, Stellenbosch University, South Africa; with permission.)*

POSTPRIMARY (ADULT-TYPE) DISEASE

Also referred to as *reactivation* or *secondary TB*, postprimary TB is probably most accurately termed *adult-type TB*, since it may result from endogenous reactivation of distant latent infection or recent infection (primary infection or reinfection). It is usually localized to the apical and posterior segments of the upper lobes, where the substantially higher mean oxygen tension (compared with that in the lower zones) favors mycobacterial growth. The superior segments of the lower lobes are also more frequently involved. The extent of lung parenchymal involvement varies greatly, from small infiltrates to extensive cavitary disease. With cavity formation, liquefied necrotic contents are ultimately discharged into the airways and may undergo bronchogenic spread, resulting in satellite lesions within the lungs that may in turn undergo cavitation **(Figs. 165-5** and **165-6)**. Massive involvement of pulmonary segments or lobes, with coalescence of lesions, produces caseating pneumonia. While up to one-third of untreated patients reportedly succumb to severe pulmonary TB within a few months after onset (the classic "galloping consumption" of the past), others may undergo a process of spontaneous remission or proceed along a chronic, progressively debilitating course ("consumption" or *phthisis*). Under these circumstances, some pulmonary lesions become fibrotic and may later calcify, but cavities persist in other parts of the lungs. Individuals with such chronic disease continue to discharge tubercle bacilli into the environment. Most patients respond to treatment, with defervescence, decreasing cough, weight gain, and a general improvement in well-being within several weeks.

Figure 165-5

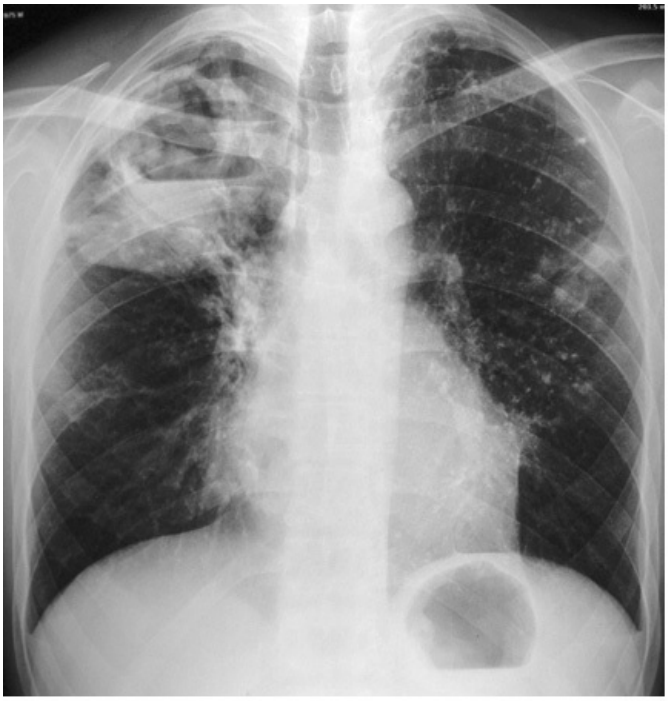

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiograph showing a right-upper-lobe infiltrate and a cavity with an air-fluid level in a patient with active tuberculosis. *(Courtesy of Dr. Andrea Gori, Department of Infectious Diseases, S. Paolo University Hospital, Milan, Italy; with permission.)*

Figure 165-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan showing a large cavity in the right lung of a patient with active tuberculosis. *(Courtesy of Dr. Enrico Girardi, National Institute for Infectious Diseases, Spallanzani Hospital, Rome, Italy; with permission.)*

Early in the course of disease, symptoms and signs are often nonspecific and insidious, consisting mainly of diurnal fever and night sweats due to defervescence, weight loss, anorexia, general malaise, and weakness. However, in up to 90% of cases, cough eventually develops—often initially nonproductive and limited to the morning and subsequently accompanied by the production of purulent sputum, sometimes with blood streaking. Hemoptysis develops in 20–30% of cases, and massive hemoptysis may ensue as a consequence of the erosion of a blood vessel in the wall of a cavity. Hemoptysis, however, may also result from rupture of a dilated vessel in a cavity (*Rasmussen's aneurysm*) or from aspergilloma formation in an old cavity. Pleuritic chest pain sometimes develops in

patients with subpleural parenchymal lesions or pleural disease. Extensive disease may produce dyspnea and, in rare instances, adult respiratory distress syndrome. Physical findings are of limited use in pulmonary TB. Many patients have no abnormalities detectable by chest examination, whereas others have detectable rales in the involved areas during inspiration, especially after coughing. Occasionally, rhonchi due to partial bronchial obstruction and classic amphoric breath sounds in areas with large cavities may be heard. Systemic features include fever (often low grade and intermittent) in up to 80% of cases and wasting. Absence of fever, however, does not exclude TB. In some cases, pallor and finger clubbing develop. The most common hematologic findings are mild anemia, leukocytosis, and thrombocytosis with a slightly elevated erythrocyte sedimentation rate and/or C-reactive protein level. None of these findings is consistent or sufficiently accurate for diagnostic purposes. Hyponatremia due to the syndrome of inappropriate secretion of antidiuretic hormone has also been reported.

Extrapulmonary TB

In order of frequency, the extrapulmonary sites most commonly involved in TB are the lymph nodes, pleura, genitourinary tract, bones and joints, meninges, peritoneum, and pericardium. However, virtually all organ systems may be affected. As a result of hematogenous dissemination in HIV-infected individuals, extrapulmonary TB is seen more commonly today than in the past.

LYMPH NODE TB (TUBERCULOUS LYMPHADENITIS)

The most common presentation of extrapulmonary TB in both HIV-seronegative and HIV-infected patients (35% in general and >40% of cases in the United States in recent series), lymph node disease is particularly frequent among HIV-infected patients and in children. In the United States, besides children, women (particularly non-Caucasians) seem to be especially susceptible. Once caused mainly by *M. bovis*, tuberculous lymphadenitis is today due largely to *M. tuberculosis*. Lymph node TB presents as painless swelling of the lymph nodes, most commonly at posterior cervical and supraclavicular sites (a condition historically referred to as *scrofula*). Lymph nodes are usually discrete in early disease but develop into a matted nontender mass over time and may result in a fistulous tract draining caseous material. Associated pulmonary disease is present in <50% of cases, and systemic symptoms are uncommon except in HIV-infected patients. The diagnosis is established by fine-needle aspiration biopsy (with a yield of up to 80%) or surgical excision biopsy. Bacteriologic confirmation is achieved in the vast majority of cases, granulomatous lesions with or without visible AFBs are typically seen, and cultures are positive in 70–80% of cases. Among HIV-infected patients, granulomas are less well organized and are frequently absent entirely, but bacterial loads are heavier than in HIV-seronegative patients, with higher yields from microscopy and culture. Differential diagnosis includes a variety of infectious conditions, neoplastic diseases such as lymphomas or metastatic carcinomas, and rare disorders like Kikuchi's disease (necrotizing histiocytic lymphadenitis), Kimura's disease, and Castleman's disease.

PLEURAL TB

Involvement of the pleura accounts for ~20% of extrapulmonary cases in the United States and elsewhere. Isolated pleural effusion usually reflects recent primary infection, and the collection of fluid in the pleural space represents a hypersensitivity response to mycobacterial antigens. Pleural disease may also result from contiguous parenchymal spread, as in many cases of pleurisy accompanying postprimary disease. Depending on the extent of reactivity, the effusion may be small, remain unnoticed, and resolve spontaneously or may be sufficiently large to cause symptoms such as fever, pleuritic chest pain, and dyspnea. Physical findings are those of pleural effusion: dullness to percussion and absence of breath sounds. A chest radiograph reveals the effusion and, in up to one-third of cases, also shows a parenchymal lesion. Thoracentesis is required to ascertain the nature of the effusion and to differentiate it from manifestations of other etiologies. The fluid is straw colored and at times hemorrhagic; it is an exudate with a protein concentration >50% of that in serum (usually ~4–6 g/dL), a normal to low glucose concentration, a pH of ~7.3 (occasionally <7.2), and detectable white blood cells (usually 500–6000/μL). Neutrophils may predominate in the early stage, but lymphocyte predominance is the typical finding later. Mesothelial cells are generally rare or absent. AFB are seen on direct smear in only 10–25% of cases, but cultures may be positive for *M. tuberculosis* in 25–75% of cases; positive cultures are more common among postprimary cases. Determination of the pleural concentration of adenosine deaminase (ADA) is a useful screening test: tuberculosis is virtually excluded if the value is very low. Lysozyme is also present in the pleural effusion. Measurement of IFN-γ, either directly or through stimulation of sensitized T cells with mycobacterial antigens, can be helpful. Needle biopsy of the pleura is often required for diagnosis and reveals granulomas and/or yields a positive culture in up to 80% of cases. This form of pleural TB responds rapidly to chemotherapy and may resolve spontaneously. Concurrent glucocorticoid administration may reduce the duration of fever and/or chest pain but is of not proven benefit.

Tuberculous empyema is a less common complication of pulmonary TB. It is usually the result of the rupture of a cavity, with spillage of a large number of organisms into the pleural space. This process may create a bronchopleural fistula with evident air in the pleural space. A chest radiograph shows hydropneumothorax with an air-fluid level. The pleural fluid is purulent and thick and contains large

numbers of lymphocytes. Acid-fast smears and mycobacterial cultures are often positive. Surgical drainage is usually required as an adjunct to chemotherapy. Tuberculous empyema may result in severe pleural fibrosis and restrictive lung disease. Removal of the thickened visceral pleura (decortication) is occasionally necessary to improve lung function.

TB OF THE UPPER AIRWAYS

Nearly always a complication of advanced cavitary pulmonary TB, TB of the upper airways may involve the larynx, pharynx, and epiglottis. Symptoms include hoarseness, dysphonia, and dysphagia in addition to chronic productive cough. Findings depend on the site of involvement, and ulcerations may be seen on laryngoscopy. Acid-fast smear of the sputum is often positive, but biopsy may be necessary in some cases to establish the diagnosis. Carcinoma of the larynx may have similar features but is usually painless.

GENITOURINARY TB

Genitourinary TB, which accounts for ~10–15% of all extrapulmonary cases in the United States and elsewhere, may involve any portion of the genitourinary tract. Local symptoms predominate, and up to 75% of patients have chest radiographic abnormalities suggesting previous or concomitant pulmonary disease. Urinary frequency, dysuria, nocturia, hematuria, and flank or abdominal pain are common presentations. However, patients may be asymptomatic and the disease discovered only after severe destructive lesions of the kidneys have developed. Urinalysis gives abnormal results in 90% of cases, revealing pyuria and hematuria. The documentation of culture-negative pyuria in acidic urine raises the suspicion of TB. IV pyelography, abdominal CT, or MRI **(Fig. 165-7)** may show deformities and obstructions, and calcifications and ureteral strictures are suggestive findings. Culture of three morning urine specimens yields a definitive diagnosis in nearly 90% of cases. Severe ureteral strictures may lead to hydronephrosis and renal damage. Genital TB is diagnosed more commonly in female than in male patients. In female patients, it affects the fallopian tubes and the endometrium and may cause infertility, pelvic pain, and menstrual abnormalities. Diagnosis requires biopsy or culture of specimens obtained by dilatation and curettage. In male patients, genital TB preferentially affects the epididymis, producing a slightly tender mass that may drain externally through a fistulous tract; orchitis and prostatitis may also develop. In almost half of cases of genitourinary TB, urinary tract disease is also present. Genitourinary TB responds well to chemotherapy.

Figure 165-7

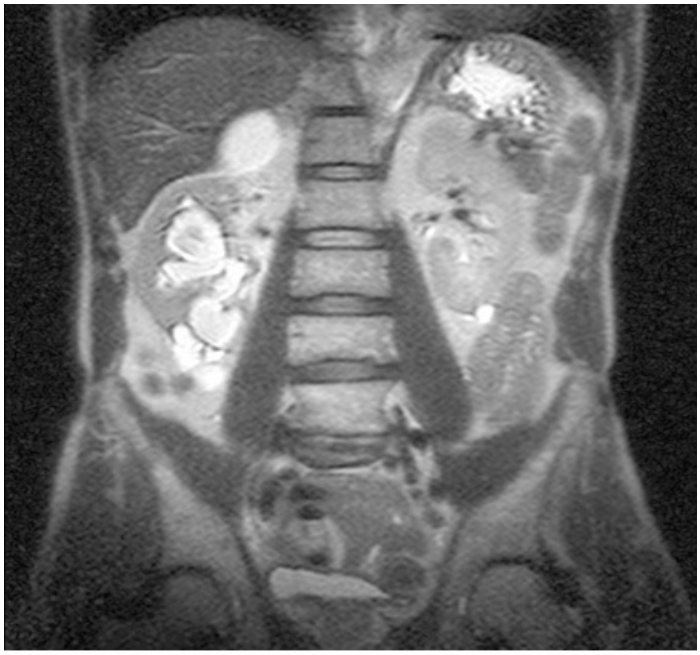

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI of culture-confirmed renal tuberculosis. T2-weighted coronary plane: coronal sections showing several renal lesions in both the cortical and the medullary tissues of the right kidney. *(Courtesy of Dr. Alberto Matteelli, Department of Infectious Diseases, University of Brescia, Italy; with permission.)*

SKELETAL TB

In the United States, TB of the bones and joints is responsible for ~10% of extrapulmonary cases. In bone and joint disease, pathogenesis is related to reactivation of hematogenous foci or to spread from

adjacent paravertebral lymph nodes. Weight-bearing joints (the spine in 40% of cases, the hips in 13%, and the knees in 10%) are most commonly affected. Spinal TB (Pott's disease or tuberculous spondylitis; **Fig. 165-8**) often involves two or more adjacent vertebral bodies. While the upper thoracic spine is the most common site of spinal TB in children, the lower thoracic and upper lumbar vertebrae are usually affected in adults. From the anterior superior or inferior angle of the vertebral body, the lesion slowly reaches the adjacent body, later affecting the intervertebral disk. With advanced disease, collapse of vertebral bodies results in kyphosis (*gibbus*). A paravertebral "cold" abscess may also form. In the upper spine, this abscess may track to and penetrate the chest wall, presenting as a soft tissue mass; in the lower spine, it may reach the inguinal ligaments or present as a psoas abscess. CT or MRI reveals the characteristic lesion and suggests its etiology. The differential diagnosis includes tumors and other infections. Pyogenic bacterial osteomyelitis, in particular, involves the disk very early and produces rapid sclerosis. Aspiration of the abscess or bone biopsy confirms the tuberculous etiology, as cultures are usually positive and histologic findings highly typical. A catastrophic complication of Pott's disease is paraplegia, which is usually due to an abscess or a lesion compressing the spinal cord. Paraparesis due to a large abscess is a medical emergency and requires rapid drainage. TB of the hip joints, usually involving the head of the femur, causes pain; TB of the knee produces pain and swelling. If the disease goes unrecognized, the joints may be destroyed. Diagnosis requires examination of the synovial fluid, which is thick in appearance, with a high protein concentration and a variable cell count. Although synovial fluid culture is positive in a high percentage of cases, synovial biopsy and tissue culture may be necessary to establish the diagnosis. Skeletal TB responds to chemotherapy, but severe cases may require surgery.

Figure 165-8

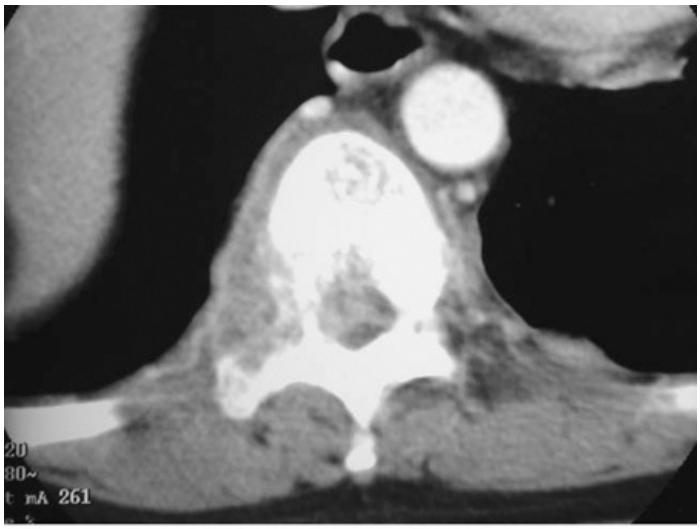

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan demonstrating destruction of the right pedicle of T10 due to Pott's disease. The patient, a 70-year-old Asian woman, presented with back pain and weight loss and had biopsy-proven tuberculosis. *(Courtesy of Charles L. Daley, MD, University of California, San Francisco; with permission.)*

TUBERCULOUS MENINGITIS AND TUBERCULOMA

TB of the central nervous system accounts for ~5% of extrapulmonary cases in the United States. It is seen most often in young children but also develops in adults, especially those infected with HIV. Tuberculous meningitis results from the hematogenous spread of primary or postprimary pulmonary TB or from the rupture of a subependymal tubercle into the subarachnoid space. In more than half of cases, evidence of old pulmonary lesions or a miliary pattern is found on chest radiography. The disease often presents subtly as headache and slight mental changes after a prodrome of weeks of low-grade fever, malaise, anorexia, and irritability. If not recognized, tuberculous meningitis may evolve acutely with severe headache, confusion, lethargy, altered sensorium, and neck rigidity. Typically, the disease evolves over 1–2 weeks, a course longer than that of bacterial meningitis. Since meningeal involvement is pronounced at the base of the brain, paresis of cranial nerves (ocular nerves in particular) is a frequent finding, and the involvement of cerebral arteries may produce focal ischemia. The ultimate evolution is toward coma, with hydrocephalus and intracranial hypertension.

Lumbar puncture is the cornerstone of diagnosis. In general, examination of cerebrospinal fluid (CSF) reveals a high leukocyte count (up to 1000/μL), usually with a predominance of lymphocytes but sometimes with a predominance of neutrophils in the early stage; a protein content of 1–8 g/L (100–800 mg/dL); and a low glucose concentration. However, any of these three parameters can be within

the normal range. AFB are seen on direct smear of CSF sediment in up to one-third of cases, but repeated lumbar punctures increase the yield. Culture of CSF is diagnostic in up to 80% of cases and remains the gold standard. Polymerase chain reaction (PCR) has a sensitivity of up to 80%, but rates of false-positivity reach 10%. Imaging studies (CT and MRI) may show hydrocephalus and abnormal enhancement of basal cisterns or ependyma. If unrecognized, tuberculous meningitis is uniformly fatal. This disease responds to chemotherapy; however, neurologic sequelae are documented in 25% of treated cases, in most of which the diagnosis has been delayed. Clinical trials have demonstrated that patients given adjunctive glucocorticoids may experience faster resolution of CSF abnormalities and elevated CSF pressure. In one study, adjunctive dexamethasone (0.4 mg/kg per day given IV and tapering by 0.1 mg/kg per week until the fourth week, when 0.1 mg/kg per day was administered; followed by 4 mg/d given by mouth and tapering by 1 mg per week until the fourth week, when 1 mg/d was administered) significantly enhanced the chances of survival among persons >14 years of age but did not reduce the frequency of neurologic sequelae.

Tuberculoma, an uncommon manifestation of central nervous system TB, presents as one or more space-occupying lesions and usually causes seizures and focal signs. CT or MRI reveals contrast-enhanced ring lesions, but biopsy is necessary to establish the diagnosis.

GASTROINTESTINAL TB

Gastrointestinal TB is uncommon, making up 3.5% of extrapulmonary cases in the United States. Various pathogenetic mechanisms are involved: swallowing of sputum with direct seeding, hematogenous spread, or (largely in developing areas) ingestion of milk from cows affected by bovine TB. Although any portion of the gastrointestinal tract may be affected, the terminal ileum and the cecum are the sites most commonly involved. Abdominal pain (at times similar to that associated with appendicitis) and swelling, obstruction, hematochezia, and a palpable mass in the abdomen are common findings at presentation. Fever, weight loss, anorexia, and night sweats are also common. With intestinal-wall involvement, ulcerations and fistulae may simulate Crohn's disease; the differential diagnosis with this entity is always difficult. Anal fistulae should prompt an evaluation for rectal TB. As surgery is required in most cases, the diagnosis can be established by histologic examination and culture of specimens obtained intraoperatively.

Tuberculous peritonitis follows either the direct spread of tubercle bacilli from ruptured lymph nodes and intraabdominal organs (e.g., genital TB in women) or hematogenous seeding. Nonspecific abdominal pain, fever, and ascites should raise the suspicion of tuberculous peritonitis. The coexistence of cirrhosis (Chap. 307) in patients with tuberculous peritonitis complicates the diagnosis. In tuberculous peritonitis, paracentesis reveals an exudative fluid with a high protein content and leukocytosis that is usually lymphocytic (although neutrophils occasionally predominate). The yield of direct smear and culture is relatively low; culture of a large volume of ascitic fluid can increase the yield, but peritoneal biopsy (with a specimen best obtained by laparoscopy) is often needed to establish the diagnosis.

PERICARDIAL TB (TUBERCULOUS PERICARDITIS)

Due either to direct extension from adjacent mediastinal or hilar lymph nodes or to hematogenous spread, pericardial TB has often been a disease of the elderly in countries with low TB prevalence. However, it also develops frequently in HIV-infected patients. Case-fatality rates are as high as 40% in some series. The onset may be subacute, although an acute presentation, with dyspnea, fever, dull retrosternal pain, and a pericardial friction rub, is possible. An effusion eventually develops in many cases; cardiovascular symptoms and signs of cardiac tamponade may ultimately appear (Chap. 239). In the presence of effusion, TB must be suspected if the patient belongs to a high-risk population (HIV-infected, originating in a high-prevalence country); if there is evidence of previous TB in other organs; or if echocardiography, CT, or MRI shows effusion and thickness across the pericardial space. A definitive diagnosis can be obtained by pericardiocentesis under echocardiographic guidance. The pericardial fluid must be submitted for biochemical, cytologic, and microbiologic study. The effusion is exudative in nature, with a high count of lymphocytes and monocytes. Hemorrhagic effusion is common. Direct smear examination is very rarely positive. Culture of pericardial fluid reveals *M. tuberculosis* in up to two-thirds of cases, while pericardial biopsy has a higher yield. High levels of ADA, lysozyme, and IFN-γ may suggest a tuberculous etiology. PCR may also be useful.

Without treatment, pericardial TB is usually fatal. Even with treatment, complications may develop, including chronic constrictive pericarditis with thickening of the pericardium, fibrosis, and sometimes calcification, which may be visible on a chest radiograph. Systematic reviews and meta-analyses show that adjunctive glucocorticoid treatment remains controversial with no conclusive evidence of benefits for all principal outcomes of pericarditis—i.e., no significant impact on resolution of effusion, no significant difference in functional status after treatment, and no significant reduction in the frequency of development of constriction or death. However, in HIV-infected patients, glucocorticoids do improve functional status after treatment.

Caused by direct extension from the pericardium or through retrograde lymphatic extension from affected mediastinal lymph nodes, tuberculous myocarditis is an extremely rare disease. Usually it is fatal and is diagnosed post-mortem.

MILIARY OR DISSEMINATED TB

Miliary TB is due to hematogenous spread of tubercle bacilli. Although in children it is often the consequence of primary infection, in adults it may be due to either recent infection or reactivation of old disseminated foci. The lesions are usually yellowish granulomas 1–2 mm in diameter that resemble millet seeds (thus the term *miliary*, coined by nineteenth-century pathologists). Clinical manifestations are nonspecific and protean, depending on the predominant site of involvement. Fever, night sweats, anorexia, weakness, and weight loss are presenting symptoms in the majority of cases. At times, patients have a cough and other respiratory symptoms due to pulmonary involvement as well as abdominal symptoms. Physical findings include hepatomegaly, splenomegaly, and lymphadenopathy. Eye examination may reveal choroidal tubercles, which are pathognomonic of miliary TB, in up to 30% of cases. Meningismus occurs in <10% of cases. A high index of suspicion is required for the diagnosis of miliary TB. Frequently, chest radiography (Fig. 165-4) reveals a miliary reticulonodular pattern (more easily seen on underpenetrated film), although no radiographic abnormality may be evident early in the course and among HIV-infected patients. Other radiologic findings include large infiltrates, interstitial infiltrates (especially in HIV-infected patients), and pleural effusion. Sputum smear microscopy is negative in 80% of cases. Various hematologic abnormalities may be seen, including anemia with leukopenia, lymphopenia, neutrophilic leukocytosis and leukemoid reactions, and polycythemia. Disseminated intravascular coagulation has been reported. Elevation of alkaline phosphatase levels and other abnormal values in liver function tests are detected in patients with severe hepatic involvement. The TST may be negative in up to half of cases, but reactivity may be restored during chemotherapy. Bronchoalveolar lavage and transbronchial biopsy are more likely to provide bacteriologic confirmation, and granulomas are evident in liver or bone-marrow biopsy specimens from many patients. If it goes unrecognized, miliary TB is lethal; with proper early treatment, however, it is amenable to cure. Glucocorticoid therapy has not proved beneficial.

A rare presentation seen in the elderly is *cryptic miliary TB* that has a chronic course characterized by mild intermittent fever, anemia, and—ultimately—meningeal involvement preceding death. An acute septicemic form, *nonreactive miliary TB*, occurs very rarely and is due to massive hematogenous dissemination of tubercle bacilli. Pancytopenia is common in this form of disease, which is rapidly fatal. At postmortem examination, multiple necrotic but nongranulomatous ("nonreactive") lesions are detected.

LESS COMMON EXTRAPULMONARY FORMS

TB may cause chorioretinitis, uveitis, panophthalmitis, and painful hypersensitivity-related phlyctenular conjunctivitis. Tuberculous otitis is rare and presents as hearing loss, otorrhea, and tympanic membrane perforation. In the nasopharynx, TB may simulate granulomatosis with polyangiitis (Wegener's). Cutaneous manifestations of TB include primary infection due to direct inoculation, abscesses and chronic ulcers, scrofuloderma, lupus vulgaris (a smoldering disease with nodules, plaques, and fissures), miliary lesions, and erythema nodosum. Tuberculous mastitis results from retrograde lymphatic spread, often from the axillary lymph nodes. Adrenal TB is a manifestation of disseminated disease presenting rarely as adrenal insufficiency. Finally, congenital TB results from transplacental spread of tubercle bacilli to the fetus or from ingestion of contaminated amniotic fluid. This rare disease affects the liver, spleen, lymph nodes, and various other organs.

HIV-ASSOCIATED TB

(See also Chap. 189) TB is one of the most common diseases among HIV-infected persons worldwide and a major cause of death. In some African countries, the rate of HIV infection among TB patients reaches 70–80% in certain urban settings. A person with a positive TST who acquires HIV infection has a 3–13% annual risk of developing active TB. A new TB infection acquired by an HIV-infected individual may evolve to active disease in a matter of weeks rather than months or years. TB can appear at any stage of HIV infection, and its presentation varies with the stage. When CMI is only partially compromised, pulmonary TB presents in a typical manner (Figs. 165-4 and 165-5), with upper-lobe infiltrates and cavitation and without significant lymphadenopathy or pleural effusion. In late stages of HIV infection, a primary TB–like pattern, with diffuse interstitial or miliary infiltrates, little or no cavitation, and intrathoracic lymphadenopathy, is more common. However, these forms are becoming less common because of the expanded use of antiretroviral treatment (ART). Overall, sputum smears may be positive less frequently among TB patients with HIV infection than among those without; thus, the diagnosis of TB may be unusually difficult, especially in view of the variety of HIV-related pulmonary conditions mimicking TB. Extrapulmonary TB is common among HIV-infected patients. In various series, extrapulmonary TB—alone or in association with pulmonary disease—has been documented in 40–60% of all cases in HIV-co-infected individuals. The most common forms are lymphatic, disseminated, pleural, and pericardial. Mycobacteremia and meningitis are also frequent, particularly in advanced HIV disease. The diagnosis of TB in HIV-infected patients may be difficult not only because of the increased frequency of sputum-smear negativity (up to 40% in culture-proven pulmonary cases) but also because of atypical radiographic findings, a lack of classic granuloma formation in the late stages, and a negative TST. Delays in treatment may prove fatal.

Exacerbations in systemic or respiratory symptoms, signs, and laboratory or radiographic

manifestations of TB—termed the *immune reconstitution inflammatory syndrome* (IRIS)—have been associated with the administration of ART. Usually occurring 1–3 months after initiation of ART, IRIS is more common among patients with advanced immunosuppression and extrapulmonary TB. "Unmasking IRIS" may also develop after the initiation of ART in patients with undiagnosed subclinical TB. The presumed pathogenesis of IRIS is an immune response that is elicited by antigens released as bacilli are killed during effective chemotherapy and that is temporally associated with improving immune function. The first priority in the management of a possible case of IRIS is to ensure that the clinical syndrome does not represent a failure of TB treatment or the development of another infection. Mild paradoxical reactions can be managed with symptom-based treatment. Glucocorticoids have been used for more severe reactions, although their use in this setting has not been formally evaluated in clinical trials.

Recommendations for the prevention and treatment of TB in HIV-infected individuals are provided below.

DIAGNOSIS

The key to the diagnosis of TB is a high index of suspicion. Diagnosis is not difficult with a high-risk patient—e.g., a homeless alcoholic who presents with typical symptoms and a classic chest radiograph showing upper-lobe infiltrates with cavities (Fig. 165-5). On the other hand, the diagnosis can easily be missed in an elderly nursing home resident or a teenager with a focal infiltrate. Often, the diagnosis is first entertained when the chest radiograph of a patient being evaluated for respiratory symptoms is abnormal. If the patient has no complicating medical conditions that cause immunosuppression, the chest radiograph may show typical upper-lobe infiltrates with cavitation (Fig. 165-5). The longer the delay between the onset of symptoms and the diagnosis, the more likely is the finding of cavitary disease. In contrast, immunosuppressed patients, including those with HIV infection, may have "atypical" findings on chest radiography—e.g., lower-zone infiltrates without cavity formation.

AFB Microscopy

A presumptive diagnosis is commonly based on the finding of AFB on microscopic examination of a diagnostic specimen, such as a smear of expectorated sputum or of tissue (e.g., a lymph node biopsy). Although inexpensive, AFB microscopy has relatively low sensitivity (40–60%) in culture-confirmed cases of pulmonary TB. The traditional method—light microscopy of specimens stained with Ziehl-Neelsen basic fuchsin dyes—is nevertheless satisfactory, although time-consuming. Most modern laboratories processing large numbers of diagnostic specimens use auramine-rhodamine staining and fluorescence microscopy. Less expensive light-emitting diode (LED) fluorescence microscopes are now available and should, over time, replace conventional light and fluorescence microscopes, especially facilitating the use of this technology in developing countries. For patients with suspected pulmonary TB, it has been recommended that two or three sputum specimens, preferably collected early in the morning, should be submitted to the laboratory for AFB smear and mycobacterial culture. Recent reviews have emphasized that two specimens collected on the same visit may be as effective as three. If tissue is obtained, it is critical that the portion of the specimen intended for culture not be put in formaldehyde. The use of AFB microscopy on urine or gastric lavage fluid is limited by the presence of commensal mycobacteria that can cause false-positive results.

Mycobacterial Culture

Definitive diagnosis depends on the isolation and identification of *M. tuberculosis* from a clinical specimen or the identification of specific sequences of DNA in a nucleic acid amplification test (see below). Specimens may be inoculated onto egg- or agar-based medium (e.g., Löwenstein-Jensen or Middlebrook 7H10) and incubated at 37°C (under 5% CO_2 for Middlebrook medium). Because most species of mycobacteria, including *M. tuberculosis*, grow slowly, 4–8 weeks may be required before growth is detected. Although *M. tuberculosis* may be identified presumptively on the basis of growth time and colony pigmentation and morphology, a variety of biochemical tests have traditionally been used to speciate mycobacterial isolates. In modern, well-equipped laboratories, the use of liquid culture for isolation and species identification by molecular methods or high-pressure liquid chromatography of mycolic acids has replaced isolation on solid media and identification by biochemical tests. A low-cost, rapid immunochromatographic lateral flow assay based on detection of MTP64 antigen may also be used for species identification of *M. tuberculosis* complex in culture isolates. These new methods, which should be introduced rapidly in developing countries, have decreased the time required for bacteriologic confirmation of TB to 2–3 weeks.

Nucleic Acid Amplification

Several test systems based on amplification of mycobacterial nucleic acid are available. These systems permit the diagnosis of TB in as little as several hours, with high specificity and sensitivity approaching that of culture. These tests are most useful for the rapid confirmation of TB in persons with AFB-positive specimens but also have utility for the diagnosis of AFB-negative pulmonary and extrapulmonary TB. In settings where these tests are available, nucleic acid amplification testing should be performed on at least one respiratory specimen from patients being evaluated for suspected

pulmonary TB.

Drug Susceptibility Testing

The initial isolate of *M. tuberculosis* should be tested for susceptibility to isoniazid and rifampin to detect MDR-TB, particularly if one or more risk factors for drug resistance are identified or the patient either fails to respond to initial therapy or has a relapse after the completion of treatment (see "Treatment Failure and Relapse," below). In addition, expanded susceptibility testing for second-line anti-TB drugs (especially the fluoroquinolones and the injectable drugs) is mandatory when MDR-TB is found. Susceptibility testing may be conducted directly (with the clinical specimen) or indirectly (with mycobacterial cultures) on solid or liquid medium. Results are obtained rapidly by direct susceptibility testing on liquid medium, with an average reporting time of 3 weeks. With indirect testing on solid medium, results may be unavailable for ≥8 weeks. Molecular methods for the rapid identification of genetic mutations known to be associated with resistance to rifampin and isoniazid (such as the line probe assays) have been developed and are being widely implemented for screening patients at increased risk of drug-resistant TB. Until the capacity for molecular testing is developed, a few noncommercial culture and drug-susceptibility testing methods (e.g., microscopically observed drug susceptibility, nitrate reductase assays, and colorimetric redox indicator assays) may be useful in resource-limited settings. Their use is limited to national reference laboratories with proven proficiency and adequate external quality control.

Radiographic Procedures

As noted above, the initial suspicion of pulmonary TB is often based on abnormal chest radiographic findings in a patient with respiratory symptoms. Although the "classic" picture is that of upper-lobe disease with infiltrates and cavities (Fig. 165-5), virtually any radiographic pattern—from a normal film or a solitary pulmonary nodule to diffuse alveolar infiltrates in a patient with adult respiratory distress syndrome—may be seen. In the era of AIDS, no radiographic pattern can be considered pathognomonic. CT (Fig. 165-6) may be useful in interpreting questionable findings on plain chest radiography and may be helpful in diagnosing some forms of extrapulmonary TB (e.g., Pott's disease; Fig. 165-8). MRI is useful in the diagnosis of intracranial TB.

Additional Diagnostic Procedures

Other diagnostic tests may be used when pulmonary TB is suspected. Sputum induction by ultrasonic nebulization of hypertonic saline may be useful for patients who cannot produce a sputum specimen spontaneously. Frequently, patients with radiographic abnormalities that are consistent with other diagnoses (e.g., bronchogenic carcinoma) undergo fiberoptic bronchoscopy with bronchial brushings and endobronchial or transbronchial biopsy of the lesion. Bronchoalveolar lavage of a lung segment containing an abnormality may also be performed. In all cases, it is essential that specimens be submitted for AFB smear and mycobacterial culture. For the diagnosis of primary pulmonary TB in children, who often do not expectorate sputum, induced sputum specimens and specimens from early-morning gastric lavage may yield positive cultures.

Invasive diagnostic procedures are indicated for patients with suspected extrapulmonary TB. In addition to testing of specimens from involved sites (e.g., CSF for tuberculous meningitis, pleural fluid and biopsy samples for pleural disease), biopsy and culture of bone marrow and liver tissue have a good diagnostic yield in disseminated (miliary) TB, particularly in HIV-infected patients, who also have a high frequency of positive blood cultures. In some cases, cultures are negative but a clinical diagnosis of TB is supported by consistent epidemiologic evidence (e.g., a history of close contact with an infectious patient), a positive TST or IFN-γ release assay (IGRA; see below), and a compatible clinical and radiographic response to treatment. In the United States and other industrialized countries with low rates of TB, some patients with limited abnormalities on chest radiographs and sputum positive for AFB are infected with nontuberculous mycobacteria, most commonly organisms of the *M. avium* complex or *M. kansasii* (Chap. 167). Factors favoring the diagnosis of nontuberculous mycobacterial disease over TB include an absence of risk factors for TB, a negative TST or IGRA, and underlying chronic pulmonary disease.

Patients with HIV-associated TB pose several diagnostic problems (see "HIV-associated TB," above). Moreover, HIV-infected patients with sputum culture–positive, AFB-positive TB may present with a normal chest radiograph. With the advent of ART, the occurrence of disseminated *M. avium* complex disease that can be confused with TB has become much less common.

Serologic and Other Diagnostic Tests for Active TB

A number of serologic tests based on detection of antibodies to a variety of mycobacterial antigens are marketed in developing countries but not in the United States. Careful independent assessments of these tests suggest that they are not useful as diagnostic aids, especially in persons with a low probability of TB. Various methods aimed at detection of mycobacterial antigens in diagnostic specimens are being investigated but are limited at present by low sensitivity. Determinations of ADA and IFN-γ levels in pleural fluid may be useful as adjunct tests in the diagnosis of pleural TB; the utility of these tests in the diagnosis of other forms of extrapulmonary TB (e.g., pericardial, peritoneal, and

meningeal) is less clear.

Diagnosis of Latent *M. Tuberculosis* Infection

TUBERCULIN SKIN TESTING

In 1891, Robert Koch discovered that components of *M. tuberculosis* in a concentrated liquid culture medium, subsequently named "old tuberculin," were capable of eliciting a skin reaction when injected subcutaneously into patients with TB. In 1932, Seibert and Munday purified this product by ammonium sulfate precipitation to produce an active protein fraction known as *tuberculin purified protein derivative* (PPD). In 1941, PPD-S, developed by Seibert and Glenn, was chosen as the international standard. Later, the WHO and UNICEF sponsored large-scale production of a master batch of PPD (RT23) and made it available for general use. The greatest limitation of PPD is its lack of mycobacterial species specificity, a property due to the large number of proteins in this product that are highly conserved in the various species. In addition, subjectivity of the skin-reaction interpretation, deterioration of the product, and batch-to-batch variations limit the usefulness of PPD.

Skin testing with tuberculin-PPD (TST) is most widely used in screening for latent *M. tuberculosis* infection (LTBI). The test is of limited value in the diagnosis of active TB because of its relatively low sensitivity and specificity and its inability to discriminate between latent infection and active disease. False-negative reactions are common in immunosuppressed patients and in those with overwhelming TB. False-positive reactions may be caused by infections with nontuberculous mycobacteria (Chap. 167) and by BCG vaccination.

IFN-γ RELEASE ASSAYS

Two in vitro assays that measure T cell release of IFN-γ in response to stimulation with the highly TB-specific antigens ESAT-6 and CFP-10 are available. The T-SPOT.TB® (Oxford Immunotec, Oxford, UK) is an enzyme-linked immunospot (ELISpot) assay, and the QuantiFERON-TB Gold® (Cellestis Ltd., Carnegie, Australia) is a whole-blood enzyme-linked immunosorbent assay (ELISA) for measurement of IFN-γ. The QuantiFERON-TB Gold In-Tube assay, which facilitates blood collection and initial incubation, also contains another specific antigen, TB7.7.

IGRAs are more specific than the TST as a result of less cross-reactivity due to BCG vaccination and sensitization by nontuberculous mycobacteria. Although diagnostic sensitivity for LTBI cannot be directly estimated because of the lack of a gold standard, these tests have shown better correlation than the TST with exposure to *M. tuberculosis* in contact investigations in low-incidence settings. However, their performance in high TB- and/or HIV-burden settings has been much more varied. Although limited, direct comparative studies of the two assays in routine practice suggest that the ELISpot has a lower rate of indeterminate results and probably has a higher degree of diagnostic sensitivity than whole-blood ELISA. Other potential advantages of IGRAs include logistical convenience, the need for fewer patient visits to complete testing, the avoidance of somewhat subjective measurements such as skin induration, and the ability to perform serial testing without inducing the boosting phenomenon (a spurious TST conversion due to boosting of reactivity on subsequent TSTs among BCG-vaccinated persons and those infected with other mycobacteria). IGRAs require that blood be drawn from patients and delivered to the laboratory in a timely fashion. Because of high specificity and other potential advantages, IGRAs may replace the TST for LTBI diagnosis in low-incidence, high-income settings where cross-reactivity due to BCG might adversely impact the interpretation and utility of the TST.

A number of national guidelines on the use of IGRAs for LTBI testing have been issued. In the United States, an IGRA is preferred over the TST for most persons over the age of 5 years who are being screened for LTBI. However, for those at high risk of progression to active TB (e.g., HIV-infected persons), either test may be used, or both may be used to optimize sensitivity. Because of the paucity of data on IGRA testing in children, the TST is preferred for LTBI testing of children under age 5. In Canada and some European countries, a two-step approach for those with positive TSTs—i.e., initial TST followed by an IGRA—is recommended. However, a TST may boost an IGRA response if the interval between the two tests exceeds 3 days.

Treatment: Tuberculosis

The two aims of TB treatment are (1) to interrupt transmission by rendering patients noninfectious and (2) to prevent morbidity and death by curing patients with TB while preventing the emergence of drug resistance. Chemotherapy for TB became possible with the discovery of streptomycin in 1943. Randomized clinical trials clearly indicated that the administration of streptomycin to patients with chronic TB reduced mortality rates and led to cure in the majority of cases. However, monotherapy with streptomycin was frequently associated with the development of resistance to this drug and the attendant failure of treatment. With the introduction into clinical practice of para-aminosalicylic acid (PAS) and isoniazid, it became axiomatic in the early 1950s that cure of TB required the concomitant administration of at least two agents to which the organism was susceptible. Furthermore, early clinical trials demonstrated that a long period of treatment—i.e., 12–24 months—was required to prevent recurrence. The introduction of rifampin (rifampicin) in the early 1970s heralded the era of effective

short-course chemotherapy, with a treatment duration of <12 months. The discovery that pyrazinamide, which was first used in the 1950s, augmented the potency of isoniazid/rifampin regimens led to the use of a 6-month course of this triple-drug regimen as standard therapy.

DRUGS

Four major drugs are considered the first-line agents for the treatment of TB: isoniazid, rifampin, pyrazinamide, and ethambutol **(Table 165-2)**. These drugs are well absorbed after oral administration, with peak serum levels at 2–4 h and nearly complete elimination within 24 h. These agents are recommended on the basis of their bactericidal activity (i.e., their ability to rapidly reduce the number of viable organisms and render patients noninfectious), their sterilizing activity (i.e., their ability to kill all bacilli and thus sterilize the affected tissues, measured in terms of the ability to prevent relapses), and their low rate of induction of drug resistance. Rifapentine and rifabutin, two drugs related to rifampin, are also available in the United States and are useful for selected patients. **For a detailed discussion of the drugs used for the treatment of TB, see Chap. 168.**

Table 165-2 Recommended Dosage[a] for Initial Treatment of Tuberculosis in Adults[b]

	Dosage	
Drug	**Daily Dose**	**Thrice-Weekly Dose[c]**
Isoniazid	5 mg/kg, max 300 mg	10 mg/kg, max 900 mg
Rifampin	10 mg/kg, max 600 mg	10 mg/kg, max 600 mg
Pyrazinamide	25 mg/kg, max 2 g	35 mg/kg, max 3 g
Ethambutol[d]	15 mg/kg	30 mg/kg

[a] The duration of treatment with individual drugs varies by regimen, as detailed in Table 165-3.

[b] Dosages for children are similar, except that some authorities recommend higher doses of isoniazid (10–15 mg/kg daily; 20–30 mg/kg intermittent) and rifampin (10–20 mg/kg).

[c] Dosages for twice-weekly administration are the same for isoniazid and rifampin but are higher for pyrazinamide (50 mg/kg, with a maximum of 4 g/d) and ethambutol (40–50 mg/d).

[d] In certain settings, streptomycin (15 mg/kg daily, with a maximum dose of 1 g; or 25–30 mg/kg thrice weekly, with a maximum dose of 1.5 g) can replace ethambutol in the initial phase of treatment. However, streptomycin is no longer considered a first-line drug by the ATS, the IDSA, or the CDC.

Source: Based on recommendations of the American Thoracic Society, the Infectious Diseases Society of America, and the Centers for Disease Control and Prevention and of the World Health Organization.

Because of a lower degree of efficacy and a higher degree of intolerability and toxicity, six classes of second-line drugs are generally used only for the treatment of patients with TB resistant to first-line drugs. Included in this group are the injectable aminoglycosides streptomycin (formerly a first-line agent), kanamycin, and amikacin; the injectable polypeptide capreomycin; the oral agents ethionamide, cycloserine, and PAS; and the fluoroquinolone antibiotics. Of the quinolones, third-generation agents are preferred: levofloxacin, gatifloxacin (no longer marketed in the United States because of its severe toxicity), and moxifloxacin. Today amithiozone (thiacetazone) is used very rarely (mainly for MDR-TB) since it is associated with severe and sometimes even fatal skin reactions among HIV-infected patients. Other drugs of unproven efficacy that have been used in the treatment of patients with resistance to most of the first- and second-line agents include clofazimine, amoxicillin/clavulanic acid, clarithromycin, imipenem, and linezolid. Two novel drugs currently under clinical development—OPC-67683, a nitroimidazole; and TMC207, a diarylquinoline—are active against MDR-TB and offer promise in shortening the course of treatment required for drug-susceptible TB as well. Moxifloxacin and gatifloxacin (see above) are in late-phase clinical development as 4-month treatment-shortening regimens for drug-susceptible TB.

REGIMENS

Standard short-course regimens are divided into an initial, or bactericidal, phase and a continuation, or sterilizing, phase. During the initial phase, the majority of the tubercle bacilli are killed, symptoms resolve, and usually the patient becomes noninfectious. The continuation phase is required to eliminate persisting mycobacteria and prevent relapse. The treatment regimen of choice for virtually all forms of TB in adults consists of a 2-month initial phase of isoniazid, rifampin, pyrazinamide, and ethambutol followed by a 4-month continuation phase of isoniazid and rifampin **(Table 165-3)**. In children, most forms can be safely treated without ethambutol in the intensive phase. Treatment may be given daily throughout the course or intermittently (either three times weekly throughout the course or twice weekly after an initial phase of daily therapy, although the twice-weekly option is not recommended by the WHO). However, HIV-infected patients should receive their initial-phase regimen

daily. A continuation phase of once-weekly rifapentine and isoniazid is equally effective for HIV-seronegative patients with noncavitary pulmonary TB who have negative sputum cultures at 2 months. Intermittent treatment is especially useful for patients whose therapy can be directly observed (see below). Patients with cavitary pulmonary TB and delayed sputum-culture conversion (i.e., those who remain culture-positive at 2 months) should have the continuation phase extended by 3 months, for a total course of 9 months. For patients with sputum culture–negative pulmonary TB, the duration of treatment may be reduced to a total of 4 months. To prevent isoniazid-related neuropathy, pyridoxine (10–25 mg/d) should be added to the regimen given to persons at high risk of vitamin B6 deficiency (e.g., alcoholics; malnourished persons; pregnant and lactating women; and patients with conditions such as chronic renal failure, diabetes, and HIV infection, which are also associated with neuropathy). A full course of therapy (completion of treatment) is defined more accurately by the total number of doses taken than by the duration of treatment. Specific recommendations on the required numbers of doses for each of the various treatment regimens have been published jointly by the American Thoracic Society, the Infectious Diseases Society of America, and the CDC. In some developing countries where the ability to ensure compliance with treatment is limited, a continuation-phase regimen of daily isoniazid and ethambutol for 6 months has been used. However, this regimen is associated with a higher rate of relapse and failure, especially among HIV-infected patients, and is no longer recommended by the WHO.

Table 165-3 Recommended Antituberculosis Treatment Regimens

	Initial Phase		Continuation Phase	
Indication	**Duration, Months**	**Drugs**	**Duration, Months**	**Drugs**
New smear- or culture-positive cases	2	HRZE[a,b]	4	HR[a,c,d]
New culture-negative cases	2	HRZE[a]	4	HR[a]
Pregnancy	2	HRE[e]	7	HR
Relapses and treatment default (pending susceptibility testing)	3	HRZES[f]	5	HRE
Failures[g]	—	—	—	—
Resistance (or intolerance) to H	Throughout (6)	RZE[h]		
Resistance (or intolerance) to R	Throughout (12–18)	HZEQ[i]		
Resistance to H + R	Throughout (at least 20 months)	ZEQ + S (or another injectable agent[j])		
Resistance to all first-line drugs	Throughout (at least 20 months)	1 injectable agent[j] + 3 of these 4: ethionamide, cycloserine, Q, PAS		
Intolerance to Z	2	HRE	7	HR

[a] All drugs can be given daily or intermittently (three times weekly throughout). A twice-weekly regimen after 2–8 weeks of daily therapy during the initial phase is sometimes used, although it is not recommended by the WHO.

[b] Streptomycin can be used in place of ethambutol but is no longer considered to be a first-line drug by the ATS/IDSA/CDC.

[c] The continuation phase should be extended to 7 months for patients with cavitary pulmonary tuberculosis who remain sputum culture–positive after the initial phase of treatment.

[d] HIV-negative patients with noncavitary pulmonary tuberculosis who have negative sputum AFB smears after the initial phase of treatment can be given once-weekly rifapentine/isoniazid in the continuation phase.

[e] The 6-month regimen with pyrazinamide can probably be used safely during pregnancy and is recommended by the WHO and the International Union Against Tuberculosis and Lung Disease. If pyrazinamide is not included in the initial treatment regimen, the minimum duration of therapy is 9 months.

[f] Streptomycin should be discontinued after 2 months. Drug susceptibility results will determine the best regimen option.

[g] The regimen is tailored according to the results of drug susceptibility tests. The availability of rapid molecular methods to identify drug resistance allows initiation of a proper regimen at the start of treatment.

[h] A fluoroquinolone may strengthen the regimen for patients with extensive disease.

[i] Streptomycin for the initial 2 months may strengthen the regimen for patients with extensive disease.

[j] Amikacin, kanamycin, or capreomycin. All these agents should be used for at least 6 months and for 4 months after culture conversion. If susceptibility is confirmed, streptomycin could be used as the injectable agent.

Abbreviations: E, ethambutol; H, isoniazid; PAS, para-aminosalicylic acid; Q, a quinolone antibiotic; R, rifampin; S, streptomycin; Z, pyrazinamide.

Lack of adherence to treatment is recognized worldwide as the most important impediment to cure. Moreover, the tubercle bacilli infecting patients who do not adhere to the prescribed regimen are likely to become drug resistant. Both patient- and provider-related factors may affect compliance. Patient-related factors include a lack of belief that the illness is significant and/or that treatment will have a beneficial effect; the existence of concomitant medical conditions (notably substance abuse); lack of social support; and poverty, with attendant joblessness and homelessness. Provider-related factors that may promote compliance include the education and encouragement of patients, the offering of convenient clinic hours, and the provision of incentives and enablers such as meals and travel vouchers. In addition to specific measures addressing noncompliance, two other strategic approaches are used: direct observation of treatment and provision of fixed-drug-combination products. Because it is difficult to predict which patients will adhere to the recommended treatment, all patients should have their therapy directly supervised, especially during the initial phase. In the United States, personnel to supervise therapy are usually available through TB control programs of local public health departments. Supervision increases the proportion of patients completing treatment and greatly lessens the chances of relapse and acquired drug resistance. Fixed-drug-combination products (e.g., isoniazid/rifampin, isoniazid/rifampin/pyrazinamide, and isoniazid/rifampin/pyrazinamide/ethambutol) are available (except, in the United States, for the four-drug fixed drug combination) and are strongly recommended as a means of minimizing the likelihood of prescription error and of the development of drug resistance as the result of monotherapy. In some formulations of these combination products, the bioavailability of rifampin has been found to be substandard. In North America and Europe, regulatory authorities ensure that combination products are of good quality; however, this type of quality assurance cannot be assumed to be operative in less affluent countries. Alternative regimens for patients who exhibit drug intolerance or adverse reactions are listed in Table 165-3. However, severe side effects prompting discontinuation of any of the first-line drugs and use of these alternative regimens are uncommon.

MONITORING TREATMENT RESPONSE AND DRUG TOXICITY

Bacteriologic evaluation is essential in monitoring the response to treatment for TB. Patients with pulmonary disease should have their sputum examined monthly until cultures become negative. With the recommended regimen, >80% of patients will have negative sputum cultures at the end of the second month of treatment. By the end of the third month, virtually all patients should be culture-negative. In some patients, especially those with extensive cavitary disease and large numbers of organisms, AFB smear conversion may lag behind culture conversion. This phenomenon is presumably due to the expectoration and microscopic visualization of dead bacilli. As noted above, patients with cavitary disease in whom sputum culture conversion does not occur by 2 months require extended treatment. When a patient's sputum cultures remain positive at ≥3 months, treatment failure and drug resistance or poor adherence to the regimen should be suspected (see below). A sputum specimen should be collected by the end of treatment to document cure. If mycobacterial cultures are not practical, then monitoring by AFB smear examination should be undertaken at 2, 5, and 6 months. Patients whose smears remain positive at 2 months should undergo a repeat examination at 3 months. Smears that are positive after 3 months of treatment when the patient is known to be adherent are indicative of treatment failure and possible drug resistance. Therefore, drug susceptibility testing should be done. Bacteriologic monitoring of patients with extrapulmonary TB is more difficult and often is not feasible. In these cases, the response to treatment must be assessed clinically and radiographically.

Monitoring of the response during chemotherapy by serial chest radiographs is not recommended, as radiographic changes may lag behind bacteriologic response and are not highly sensitive. After the completion of treatment, neither sputum examination nor chest radiography is recommended for routine follow-up purposes. However, a chest radiograph obtained at the end of treatment may be useful for comparative purposes should the patient develop symptoms of recurrent TB months or years later. Patients should be instructed to report promptly for medical assessment should they develop any such symptoms.

During treatment, patients should be monitored for drug toxicity (Table 165-3). The most common adverse reaction of significance is hepatitis. Patients should be carefully educated about the signs and

symptoms of drug-induced hepatitis (e.g., dark urine, loss of appetite) and should be instructed to discontinue treatment promptly and see their health care provider should these symptoms occur. Although biochemical monitoring is not routinely recommended, all adult patients should undergo baseline assessment of liver function (e.g., measurement of serum levels of hepatic aminotransferases and serum bilirubin). Older patients, those with concomitant diseases, those with a history of hepatic disease (especially hepatitis C), and those using alcohol daily should be monitored especially closely (i.e., monthly), with repeated measurements of aminotransferases, during the initial phase of treatment. Up to 20% of patients have small increases in aspartate aminotransferase (up to three times the upper limit of normal) that are not accompanied by symptoms and are of no consequence. For patients with symptomatic hepatitis and those with marked (five- to sixfold) elevations in serum levels of aspartate aminotransferase, treatment should be stopped and drugs reintroduced one at a time after liver function has returned to normal. Hypersensitivity reactions usually require the discontinuation of all drugs and rechallenge to determine which agent is the culprit. Because of the variety of regimens available, it is usually not necessary—although it is possible—to desensitize patients. Hyperuricemia and arthralgia caused by pyrazinamide can usually be managed by the administration of acetylsalicylic acid; however, pyrazinamide treatment should be stopped if the patient develops gouty arthritis. Individuals who develop autoimmune thrombocytopenia secondary to rifampin therapy should not receive the drug thereafter. Similarly, the occurrence of optic neuritis with ethambutol is an indication for permanent discontinuation of this drug. Other common manifestations of drug intolerance, such as pruritus and gastrointestinal upset, can generally be managed without the interruption of therapy.

TREATMENT FAILURE AND RELAPSE

As stated above, treatment failure should be suspected when a patient's sputum smears and/or cultures remain positive after 3 months of treatment. In the management of such patients, it is imperative that the current isolate be tested for susceptibility to first- and second-line agents. Initial molecular testing for rifampin resistance should also be done if the technology is available. When the results of susceptibility testing are expected to become available within a few weeks, changes in the regimen can be postponed until that time. However, if the patient's clinical condition is deteriorating, an earlier change in regimen may be indicated. A cardinal rule in the latter situation is always to add more than one drug at a time to a failing regimen: at least two and preferably three drugs that have never been used and to which the bacilli are likely to be susceptible should be added. The patient may continue to take isoniazid and rifampin along with these new agents pending the results of susceptibility tests.

Patients who experience a recurrence after apparently successful treatment (relapses) are less likely to harbor drug-resistant strains (see below) than are patients in whom treatment has failed. However, if the regimen administered initially does not contain rifampin, the probability of isoniazid resistance is high. Acquired resistance is uncommon among strains from patients who relapse after completing a standard short-course regimen. However, it is prudent to begin the treatment of all patients who have relapsed with all four first-line drugs plus streptomycin, pending the results of susceptibility testing. In less affluent countries and other settings where facilities for culture and drug susceptibility testing are not yet routinely available, the WHO recommends that a standard regimen with all four first-line drugs plus streptomycin be used in all instances of relapse and treatment default. Patients with treatment failure should receive an empirical regimen, including second-line agents, based on their history of anti-TB treatment and the drug resistance patterns in the population (Table 165-3). Once drug susceptibility testing results are available, the regimen should be adjusted accordingly.

DRUG-RESISTANT TB

Strains of *M. tuberculosis* resistant to individual drugs arise by spontaneous point mutations in the mycobacterial genome that occur at low but predictable rates (10^{-7}–10^{-10} for the key drugs). Because there is no cross-resistance among the commonly used drugs, the probability that a strain will be resistant to two drugs is the product of the probabilities of resistance to each drug and thus is low. The development of drug-resistant TB is invariably the result of monotherapy—i.e., the failure of the health care provider to prescribe at least two drugs to which tubercle bacilli are susceptible or of the patient to take properly prescribed therapy. Drug-resistant TB may be either primary or acquired. Primary drug resistance is that which develops in a strain infecting a patient who has not previously been treated. Acquired resistance develops during treatment with an inappropriate regimen. In North America and western Europe, rates of primary resistance are generally low, and isoniazid resistance is most common. In the United States, while rates of primary isoniazid resistance have been stable at ~7–8%, the rate of primary MDR-TB has declined from 2.5% in 1993 to 1% since 2000. Resistance rates are higher among foreign-born and HIV-infected patients. As described above, worldwide, MDR-TB is an increasingly serious problem in some regions, especially in the states of the former Soviet Union and in other parts of Asia **(Fig. 165-9)**. Even more serious is the recently described occurrence of virtually untreatable XDR-TB due to MDR strains that are resistant to all fluoroquinolones and to at least one of three second-line injectable agents (amikacin, kanamycin, and capreomycin). Drug-resistant TB can be prevented by adherence to the principles of sound therapy: the inclusion of at least two bactericidal drugs to which the organism is susceptible, the use of fixed-drug-combination products, and the verification that patients complete the prescribed course.

Figure 165-9

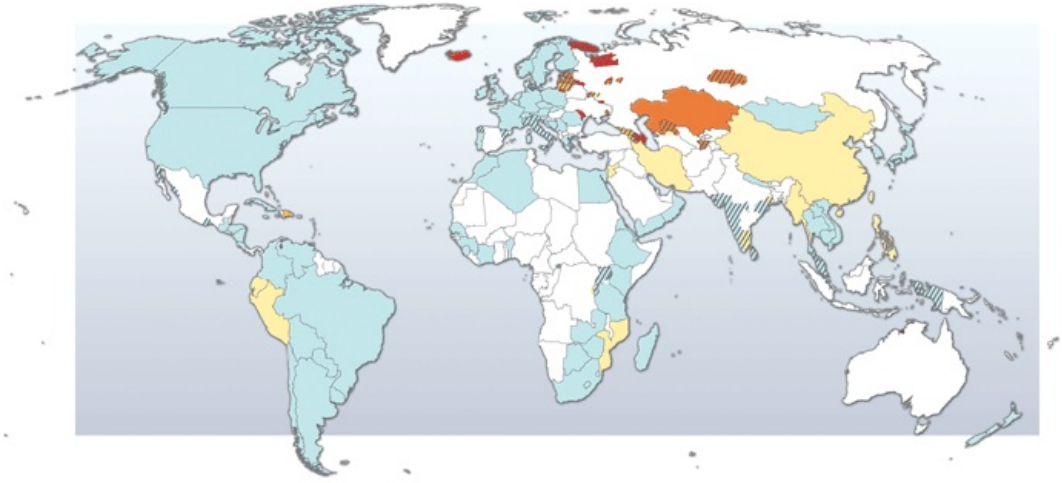

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Percentage of new tuberculosis cases with multidrug resistance in all countries surveyed by the WHO/Union Global Drug Resistance Surveillance Project during 1994–2008. *(See disclaimer in Fig. 165-2. Courtesy of the Stop TB Department, WHO; with permission.)*

Although the 6-month regimen described in Table 165-3 is generally effective for patients with initial isoniazid-resistant disease, it is prudent to include at least ethambutol and possibly pyrazinamide for the full 6 months. In such cases, isoniazid probably does not contribute to a successful outcome and could be omitted. For patients with extensive disease, a fluoroquinolone may be added. Patients whose isolates exhibit monoresistance to rifampin should receive a regimen containing isoniazid, pyrazinamide, ethambutol, and a fluoroquinolone for 12–18 months. MDR-TB is more difficult to manage than is disease caused by drug-susceptible organisms, especially because resistance to other first-line drugs besides isoniazid and rifampin is common. For treatment of TB due to strains resistant to isoniazid and rifampin, combinations of a fluoroquinolone, ethambutol, pyrazinamide, and streptomycin or, for strains resistant to streptomycin as well, another injectable agent (amikacin, kanamycin, or capreomycin) should be used. For patients with bacilli resistant to all of the first-line agents, cure may be attained with a combination of four second-line drugs, including one injectable agent (Table 165-3). Although the optimal duration of treatment is not known, a course of at least 20 months, is recommended. Patients with XDR-TB have fewer treatment options and a much poorer prognosis. However, observational studies have shown that aggressive management of cases comprising early drug-susceptibility testing, rational combination of at least five drugs, readjustment of the regimen, strict directly observed therapy, bacteriologic monitoring, and intensive patient support may result in cure rates of up to 60% and may avert deaths. **Table 165-4** describes how to manage patients with XDR-TB. For patients with localized disease and sufficient pulmonary reserve, lobectomy or pneumonectomy may be considered. Because the management of patients with MDR- and XDR-TB is complicated by both social and medical factors, care of these patients is ideally provided in specialized centers or, in their absence, in the context of programs with adequate resources and capacity.

Table 165-4 Management Guidelines for Patients with Documented or Strongly Suspected XDR-TB

1. Use any first-line oral agents that may be effective.
2. Use an injectable agent to which the strain is susceptible, and consider an extended duration of use (12 months or possibly the whole treatment period). If the strain is resistant to all injectable agents, use of an agent that the patient has not previously received is recommended.[a]
3. Use a later-generation fluoroquinolone, such as moxifloxacin.
4. Use all second-line oral agents (para-aminosalicylic acid, cycloserine, ethionamide, or prothionamide) that have not been used extensively in a previous regimen or any that are likely to be effective.
5. Use two or more of the following drugs of unclear role: clofazimine, amoxicillin/clavulanic acid, clarithromycin, imipenem, linezolid, thiacetazone.

6. Consider treatment with high-dose isoniazid if low-level resistance to this drug is documented.
7. Consider adjuvant surgery if there is localized disease.
8. Enforce strong infection-control measures.
9. Implement strict directly observed therapy and full adherence support as well as comprehensive bacteriologic and clinical monitoring.

[a] This recommendation is made because, while the reproducibility and reliability of susceptibility testing with injectable agents are good, there are few data on the correlation of clinical efficacy with test results. Options with XDR-TB are very limited, and some strains may be affected in vivo by an injectable agent even though they test resistant in vitro.

Source: Adapted from the World Health Organization, 2008.

HIV-ASSOCIATED TB

In general, the standard treatment regimens are equally efficacious in HIV-negative and HIV-positive patients. However, adverse drug effects may be more pronounced in HIV-infected patients. Three important considerations are relevant to TB treatment in HIV-infected patients: an increased frequency of paradoxical reactions, drug interactions between ART and rifamycins, and development of rifampin monoresistance with widely spaced intermittent treatment. IRIS—i.e., the exacerbation of symptoms and signs of TB—has been described above. All HIV-infected TB patients are candidates for ART, and the optimal timing for its initiation is as soon as possible and within the first 8 weeks of anti-TB therapy. Rifampin, a potent inducer of enzymes of the cytochrome P450 system, lowers serum levels of many HIV protease inhibitors and some nonnucleoside reverse transcriptase inhibitors—essential drugs used in ART. In such cases, rifabutin, which has much less enzyme-inducing activity, has been recommended in place of rifampin. However, dosage adjustment for rifabutin and/or the antiretroviral drugs may be necessary. Because recommendations are frequently updated, consultation of the CDC website is advised (www.cdc.gov/tb). Several clinical trials have found that patients with HIV-associated TB whose immunosuppression is advanced (CD4+ T cell counts of <100/μL) are prone to treatment failure and relapse with rifampin-resistant organisms when treated with "highly intermittent" (i.e., once- or twice-weekly) rifamycin-containing regimens. Consequently, it is recommended that these patients receive daily therapy for at least the initial phase.

SPECIAL CLINICAL SITUATIONS

Although comparative clinical trials of treatment for extrapulmonary TB are limited, the available evidence indicates that most forms of disease can be treated with the 6-month regimen recommended for patients with pulmonary disease. The American Academy of Pediatrics recommends that children with bone and joint TB, tuberculous meningitis, or miliary TB receive 9–12 months of treatment. Treatment for TB may be complicated by underlying medical problems that require special consideration. As a rule, patients with chronic renal failure should not receive aminoglycosides and should receive ethambutol only if serum drug levels can be monitored. Isoniazid, rifampin, and pyrazinamide may be given in the usual doses in cases of mild to moderate renal failure, but the dosages of isoniazid and pyrazinamide should be reduced for all patients with severe renal failure except those undergoing hemodialysis. Patients with hepatic disease pose a special problem because of the hepatotoxicity of isoniazid, rifampin, and pyrazinamide. Patients with severe hepatic disease may be treated with ethambutol, streptomycin, and possibly another drug (e.g., a fluoroquinolone); if required, isoniazid and rifampin may be administered under close supervision. The use of pyrazinamide by patients with liver failure should be avoided. Silicotuberculosis necessitates the extension of therapy by at least 2 months.

The regimen of choice for pregnant women (Table 165-3) is 9 months of treatment with isoniazid and rifampin supplemented by ethambutol for the first 2 months. Although the WHO has recommended routine use of pyrazinamide for pregnant women, this drug has not been recommended in the United States because of insufficient data documenting its safety in pregnancy. Streptomycin is contraindicated because it is known to cause eighth-cranial-nerve damage in the fetus. Treatment for TB is not a contraindication to breast-feeding; most of the drugs administered will be present in small quantities in breast milk, albeit at concentrations far too low to provide any therapeutic or prophylactic benefit to the child.

Medical consultation on difficult-to-manage cases is provided by the CDC Regional Training and Medical Consultation Centers (www.cdc.gov/tb/education/rtmc/).

PREVENTION

The best way to prevent TB is to diagnose and isolate infectious cases rapidly and to administer appropriate treatment until patients are rendered noninfectious (usually 2–4 weeks after the start of proper treatment) and the disease is cured. Additional strategies include BCG vaccination and treatment of persons with latent tuberculosis infection who are at high risk of developing active disease.

BCG Vaccination

BCG was derived from an attenuated strain of *M. bovis* and was first administered to humans in 1921. Many BCG vaccines are available worldwide; all are derived from the original strain, but the vaccines vary in efficacy, ranging from 80% to nil in randomized, placebo-controlled trials. A similar range of efficacy was found in recent observational studies (case-control, historic cohort, and cross-sectional) in areas where infants are vaccinated at birth. These studies also found higher rates of efficacy in the protection of infants and young children from relatively serious forms of TB, such as tuberculous meningitis and miliary TB. BCG vaccine is safe and rarely causes serious complications. The local tissue response begins 2–3 weeks after vaccination, with scar formation and healing within 3 months. Side effects—most commonly, ulceration at the vaccination site and regional lymphadenitis—occur in 1–10% of vaccinated persons. Some vaccine strains have caused osteomyelitis in ~1 case per million doses administered. Disseminated BCG infection ("BCGitis") and death have occurred in 1–10 cases per 10 million doses administered, although this problem is restricted almost exclusively to persons with impaired immunity, such as children with severe combined immunodeficiency syndrome or adults with HIV infection. BCG vaccination induces TST reactivity, which tends to wane with time. The presence or size of TST reactions after vaccination does not predict the degree of protection afforded.

BCG vaccine is recommended for routine use at birth in countries with high TB prevalence. However, because of the low risk of transmission of TB in the United States, the unreliable protection afforded by BCG, and its impact on the TST, the vaccine has never been recommended for general use in the United States. HIV-infected adults and children should not receive BCG vaccine. Moreover, infants whose HIV status is unknown but who have signs and symptoms consistent with HIV infection or who are born to HIV-infected mothers should not receive BCG.

Treatment: Latent Tuberculosis Infection

Treatment of selected persons with LTBI aims at preventing active disease. This intervention (also called *preventive chemotherapy* or *chemoprophylaxis*) is based on the results of a large number of randomized, placebo-controlled clinical trials demonstrating that a 6- to 12-month course of isoniazid reduces the risk of active TB in infected people by up to 90%. Analysis of available data indicates that the optimal duration of treatment is 9–10 months. In the absence of reinfection, the protective effect is believed to be lifelong. Clinical trials have shown that isoniazid reduces rates of TB among TST-positive persons with HIV infection. Studies in HIV-infected patients have also demonstrated the effectiveness of shorter courses of rifampin-based treatment.

Candidates for treatment of LTBI **(Table 165-5)** are identified by TST or IGRA of persons in defined high-risk groups. For skin testing, 5 tuberculin units of polysorbate-stabilized PPD should be injected intradermally into the volar surface of the forearm (Mantoux method). Multipuncture tests are not recommended. Reactions are read at 48–72 h as the transverse diameter (in millimeters) of induration; the diameter of erythema is not considered. In some persons, TST reactivity wanes with time but can be recalled by a second skin test administered ≥1 week after the first (i.e., two-step testing). For persons periodically undergoing the TST, such as health care workers and individuals admitted to long-term-care institutions, initial two-step testing may preclude subsequent misclassification of persons with boosted reactions as TST converters. The cutoff for a positive TST (and thus for treatment) is related both to the probability that the reaction represents true infection and to the likelihood that the individual, if truly infected, will develop TB (Table 165-5). Thus, positive reactions for close contacts of infectious cases, persons with HIV infection, persons receiving drugs that suppress the immune system, and previously untreated persons whose chest radiograph is consistent with healed TB are defined as an area of induration ≥5 mm in diameter. A 10-mm cutoff is used to define positive reactions in most other at-risk persons. For persons with a very low risk of developing TB if infected, a cutoff of 15 mm is used. (Except for employment purposes where longitudinal screening is anticipated, the TST is not indicated for these low-risk persons.) Treatment should be considered for persons from TB-endemic countries who have a history of BCG vaccination. A positive IGRA is based on the manufacturers' recommendations. For the ELISpot assay, there is an uncertainty zone (5–7 spots) for which epidemiologic and clinical factors guide the decision to implement treatment for LTBI. This approach has also been suggested for interpretation of results in the whole-blood assay that are close to the recommended cutoff for a positive test (0.35 IU of IFN-γ). Some TST- and IGRA-negative individuals are also candidates for treatment. Infants and children who have come into contact with infectious cases should be treated and should have a repeat skin test 2 or 3 months after contact ends. Those whose test results remain negative should discontinue treatment. HIV-infected persons who have been exposed to an infectious TB patient should receive treatment regardless of the TST result. Any HIV-infected candidate for LTBI treatment must be screened carefully to exclude active TB, which would necessitate full treatment.

Table 165-5 Tuberculin Reaction Size and Treatment of Latent *Mycobacterium tuberculosis* Infection

Risk Group	Tuberculin Reaction Size, mm

HIV-infected persons or persons receiving immunosuppressive therapy	≥5
Close contacts of tuberculosis patients	≥5[a]
Persons with fibrotic lesions on chest radiography	≥5
Recently infected persons (≤2 years)	≥10
Persons with high-risk medical conditions[b]	≥10
Low-risk persons[c]	≥15

[a] Tuberculin-negative contacts, especially children, should receive prophylaxis for 2–3 months after contact ends and should then undergo repeat TST. Those whose results remain negative should discontinue prophylaxis. HIV-infected contacts should receive a full course of treatment regardless of TST results.

[b] Includes diabetes mellitus, some hematologic and reticuloendothelial diseases, injection drug use (with HIV seronegativity), end-stage renal disease, and clinical situations associated with rapid weight loss.

[c] Except for employment purposes where longitudinal TST screening is anticipated, TST is not indicated for these low-risk persons. A decision to treat should be based on individual risk/benefit considerations.

Isoniazid is administered at a daily dose of 5 mg/kg (up to 300 mg/d) for 9 months **(Table 165-6)**. On the basis of cost-benefit analyses, a 6-month period of treatment has been recommended in the past and may be considered for HIV-negative adults with normal chest radiographs when financial considerations are important. When supervised treatment is desirable and feasible, isoniazid may be given at a dose of 15 mg/kg (up to 900 mg) twice weekly. An alternative regimen for adults is 4 months of daily rifampin. A 3-month regimen of isoniazid and rifampin is recommended in the United Kingdom for both adults and children. A previously recommended regimen of 2 months of rifampin and pyrazinamide has been associated with serious and fatal hepatotoxicity and now is generally not recommended. The rifampin regimen should be considered for persons who are likely to have been infected with an isoniazid-resistant strain. Pending the results of a large-scale study of LTBI treatment conducted by the CDC, it is possible that a regimen of isoniazid and rifapentine given once weekly for 12 weeks will also become an option. Furthermore, clinical trials are under way to assess the efficacy of long-term isoniazid administration (i.e., for at least 3 years). Isoniazid should not be given to persons with active liver disease. All persons at increased risk of hepatotoxicity (e.g., those abusing alcohol daily and those with a history of liver disease) should undergo baseline and then monthly assessment of liver function. All patients should be carefully educated about hepatitis and instructed to discontinue use of the drug immediately should any symptoms develop. Moreover, patients should be seen and questioned monthly during therapy about adverse reactions and should be given no more than 1 month's supply of drug at each visit.

Table 165-6 Revised Drug Regimens for Treatment of Latent Tuberculosis Infection (LTBI) in Adults

			Rating[b] (Evidence[c])	
Drug	**Interval and Duration**	**Comments[a]**	**HIV-Negative**	**HIV-Infected**
Isoniazid	Daily for 9 months[d,e]	In HIV-infected persons, isoniazid may be administered concurrently with nucleoside reverse transcriptase inhibitors, protease inhibitors, or NNRTIs.	A (II)	A (II)
	Twice weekly for 9 months[d,e]	DOT must be used with twice-weekly dosing.	B (II)	B (II)
	Daily for 6 months[e]	Regimen is not indicated for HIV-infected persons, those with fibrotic lesions on chest radiographs, or children.	B (I)	C (I)
	Twice weekly for 6 months[e]	DOT must be used with twice-weekly dosing.	B (II)	C (I)
Rifampin[f]	Daily for 4 months	Regimen is used for contacts of patients with isoniazid-resistant, rifampin-susceptible tuberculosis. In HIV-infected persons, most protease inhibitors and	B (II)	B (III)

		delavirdine should not be administered concurrently with rifampin. Rifabutin, with appropriate dose adjustments, can be used with protease inhibitors (saquinavir should be augmented with ritonavir) and NNRTIs (except delavirdine). Clinicians should consult web-based updates for the latest specific recommendations.		
Rifampin plus pyrazinamide	Daily for 2 months	Regimen generally should not be offered for treatment of LTBI in either HIV-infected or HIV-negative persons.	D (II)	D (II)
	Twice weekly for 2–3 months		D (III)	D (III)

[a] Interactions with HIV-related drugs are updated frequently and are available at *www.aidsinfo.nih.gov/guidelines*.

[b] Strength of the recommendation: A. Both strong evidence of efficacy and substantial clinical benefit support recommendation for use. Should always be offered. B. Moderate evidence for efficacy or strong evidence for efficacy but only limited clinical benefit supports recommendation for use. Should generally be offered. C. Evidence for efficacy is insufficient to support a recommendation for or against use, or evidence for efficacy might not outweigh adverse consequences (e.g., drug toxicity, drug interactions) or cost of the treatment or alternative approaches. Optional. D. Moderate evidence for lack of efficacy or for adverse outcome supports a recommendation against use. Should generally not be offered. E. Good evidence for lack of efficacy or for adverse outcome supports a recommendation against use. Should never be offered.

[c] Quality of evidence supporting the recommendation: I. Evidence from at least one properly randomized controlled trial. II. Evidence from at least one well-designed clinical trial without randomization, from cohort or case-controlled analytic studies (preferably from more than one center), from multiple time-series studies, or from dramatic results in uncontrolled experiments. III. Evidence from opinions of respected authorities based on clinical experience, descriptive studies, or reports of expert committees.

[d] Recommended regimen for persons aged <18 years.

[e] Recommended regimen for pregnant women.

[f] The substitution of rifapentine for rifampin is not recommended because rifapentine's safety and effectiveness have not been established for patients with LTBI.

Abbreviations: DOT, directly observed therapy; NNRTIs, nonnucleoside reverse transcriptase inhibitors.

Source: Adapted from CDC: Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR Recomm Rep 49:RR-6, 2000

It may be more difficult to ensure compliance when treating persons with latent infection than when treating those with active TB. If family members of active cases are being treated, compliance and monitoring may be easier. When feasible, twice-weekly supervised therapy may increase the likelihood of completion. As in active cases, the provision of incentives may also be helpful.

PRINCIPLES OF TB CONTROL

The highest priority in any TB control program is the prompt detection of cases and the provision of short-course chemotherapy to all TB patients under proper case-management conditions, including directly observed therapy. In addition, in low-prevalence countries with adequate resources (and increasingly in developing countries as well), screening of high-risk groups, such as immigrants from high-prevalence countries, migratory workers, prisoners, homeless individuals, substance abusers, and HIV-seropositive persons, is recommended. TST-positive high-risk persons should be treated for latent infection. Contact investigation is an important component of efficient TB control. In the United States and other countries worldwide, a great deal of attention has been given to the transmission of TB (particularly in association with HIV infection) in institutional settings such as hospitals, homeless shelters, and prisons. Measures to limit such transmission include respiratory isolation of persons with suspected TB until they are proven to be noninfectious (i.e., at least by sputum AFB smear negativity), proper ventilation in rooms of patients with infectious TB, use of ultraviolet irradiation in areas of increased risk of TB transmission, and periodic screening of personnel who may come into contact with known or unsuspected cases of TB. In the past, radiographic surveys, especially those conducted with portable equipment and miniature films, were advocated for case finding. Today, however, the prevalence of TB in industrialized countries is sufficiently low that "mass miniature radiography" is not cost-effective.

In high-prevalence countries, most TB control programs have made remarkable progress in reducing

morbidity and mortality during the past 15 years by adopting and implementing the DOTS strategy promoted by the WHO. Between 1995 and 2008, 36 million TB cases were cured and more than 6 million deaths averted compared with the pre-DOTS period. The DOTS approach consists of: (1) political commitment with increased and sustained financing; (2) case detection through quality-assured bacteriology (starting with microscopic examination of sputum from patients with cough of >2–3 weeks' duration, culture, and possibly drug susceptibility testing); (3) administration of standardized short-course chemotherapy, with direct supervision and patient support; (4) an effective drug supply and management system; and (5) a monitoring and evaluation system, with impact measurement (including assessment of treatment outcomes—e.g., cure, completion of treatment without bacteriologic proof of cure, death, treatment failure, and default—in all cases registered and notified). In 2006, the WHO indicated that, while DOTS remains the essential component of any control strategy, additional steps must be undertaken to reach the 2015 TB control targets set within the United Nations Millennium Development Goals. Thus, a new "Stop TB Strategy" with six components has been promoted: (1) Pursue high-quality DOTS expansion and enhancement. (2) Address HIV-associated TB, MDR-TB, and the needs of poor and vulnerable populations. (3) Contribute to health system strengthening. (4) Engage all care providers. (5) Empower people with TB and [their] communities. (6) Enable and promote research. As part of the fourth component, evidence-based International Standards for Tuberculosis Care, focused on diagnosis, treatment, and public health responsibilities, have recently been introduced for wide adoption by medical and professional societies, academic institutions, and all practitioners worldwide. Care and control of HIV-associated TB is particularly challenging in developing countries, since existing interventions require collaboration between HIV/AIDS and TB programs as well as standard services. While TB programs must test every patient for HIV in order to provide access to trimethoprim-sulfamethoxazole prophylaxis against common infections and ART, HIV/AIDS programs must regularly screen persons living with HIV/AIDS for active TB and provide treatment for LTBI. Early and active case detection is considered an important intervention not only among persons living with HIV/AIDS but also among other vulnerable populations, as it reduces transmission in a community and provides early effective care. For TB control efforts to succeed, programs must optimize their performance and include additional interventions as described. However, bold public health policies must be enforced to support work on TB control and care. These policies include free access to diagnosis and treatment, at least for the poorest patients; sound regulations to ensure drug quality; rational use of drugs by all care providers; laboratory networks equipped with the latest technology for rapid diagnosis; and airborne infection control in all facilities and congregate settings attended by TB patients, especially where HIV prevalence is high. Finally, elimination of TB will require control and attenuation of the multitude of risk factors (e.g., HIV, smoking, and diabetes) and socioeconomic determinants (e.g., extreme poverty, inadequate living conditions and bad housing, alcoholism, malnutrition, and indoor air pollution) with clear policies within the health sector and other sectors linked to human development and welfare.

FURTHER READINGS

Centers for Disease Control and Prevention: Control of tuberculosis in the United States: Recommendations from the American Thoracic Society, CDC, and the Infectious Diseases Society of America. MMWR Recomm Rep 54:RR-12, 2005

Lienhardt C et al: New drugs and new regimens for the treatment of tuberculosis: Review of the drug development pipeline and implications for national programmes. Curr Opin Pulmon Med 16:186, 2010[PMID: 20216421]

Lönnroth K et al: Tuberculosis control and elimination 2010–50: Cure, care, and social development. Lancet 375:1814, 2010[PMID: 21698810]

Menzies D et al: Effect of duration and intermittency of rifampin on tuberculosis treatment outcomes: A systematic review and meta-analysis. PLoS Medicine 6:e1000146, 2009

Pai M et al: Evidence-based tuberculosis diagnosis. PLoS Medicine 5:e156, 2008

Raviglione MC (ed): *Tuberculosis*, *The Essentials. Lung Biology in Health and Disease* , vol 237, 4th ed. New York, Informa Health Care USA, 2009

Tuberculosis Coalition for Technical Assistance: International standards for tuberculosis care (ISTC), 2d ed. The Hague, 2009

World Health Organization: Treatment of tuberculosis: Guidelines, 4th ed. Geneva, WHO, 2009 (www.who.int/tb/publications/2009/who_htm_tb_2009_420_beforeprint.pdf)

———: Guidelines for the programmatic management of drug-resistant tuberculosis. Emergency update 2008. Geneva, WHO, 2008

Wright A: Epidemiology of anti-tuberculosis drug resistance 2002–2007: An updated analysis of the Global Project on Anti-Tuberculosis Drug Resistance Surveillance. Lancet 373:1861, 2009[PMID: 19375159]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LEPROSY: INTRODUCTION

Leprosy, first described in ancient Indian texts from the sixth century b.c., is a nonfatal, chronic infectious disease caused by *Mycobacterium leprae,* the clinical manifestations of which are largely confined to the skin, peripheral nervous system, upper respiratory tract, eyes, and testes. The unique tropism of *M. leprae* for peripheral nerves (from large nerve trunks to microscopic dermal nerves) and certain immunologically mediated reactional states are the major causes of morbidity in leprosy. The propensity of the disease, when untreated, to result in characteristic deformities and the recognition in most cultures that the disease is communicable from person to person have resulted historically in a profound social stigma. Today, with early diagnosis and the institution of appropriate and effective antimicrobial therapy, patients can lead productive lives in the community, and deformities and other visible manifestations can largely be prevented.

Etiology

M. leprae is an obligate intracellular bacillus (0.3–1 μm wide and 1–8 μm long) that is confined to humans, armadillos in certain locales, and sphagnum moss. The organism is acid-fast, indistinguishable microscopically from other mycobacteria, and ideally detected in tissue sections by a modified Fite stain. Strain variability has been documented in this organism. *M. leprae* produces no known toxins and is well adapted to penetrate and reside within macrophages, yet it may survive outside the body for months. In untreated patients, only ~1% of *M. leprae* organisms are viable. The morphologic index (MI), a measure of the number of acid-fast bacilli (AFB) in skin scrapings that stain uniformly bright, correlates with viability. The bacteriologic index (BI), a logarithmic-scaled measure of the density of *M. leprae* in the dermis, may be as high as 4–6+ in untreated patients and falls by 1 unit per year during effective antimicrobial therapy; the rate of decrease is independent of the relative potency of therapy. A rising MI or BI suggests relapse and perhaps—if the patient is being treated—drug resistance. Drug resistance can be confirmed or excluded in the mouse model of leprosy, and resistance to dapsone and rifampin can be documented by the recognition of mutant genes. However, the availability of these technologies is extremely limited.

As a result of reductive evolution, almost half of the *M. leprae* genome contains nonfunctional genes; only 1605 genes encode for proteins, and 1439 genes are shared with *Mycobacterium tuberculosis*. In contrast, *M. tuberculosis* uses 91% of its genome to encode for 4000 proteins. Among the lost genes in *M. leprae* are those for catabolic and respiratory pathways; transport systems; purine, methionine, and glutamine synthesis; and nitrogen regulation. The genome of *M. leprae* provides a metabolic rationale for its obligate intracellular existence and reliance on host biochemical support, a template for targets of drug development, and ultimately a pathway to cultivation. The finding of strain variability among *M. leprae* isolates has provided a powerful tool with which to address anew the organism's epidemiology and pathobiology and to determine whether relapse represents reactivation or reinfection. The bacterium's complex cell wall contains large amounts of an *M. leprae*–specific phenolic glycolipid (PGL-1), which is detected in serologic tests. The unique trisaccharide of *M. leprae* binds to the basal lamina of Schwann cells; this interaction is probably relevant to the fact that *M. leprae* is the only bacterium to invade peripheral nerves.

Although it was the first bacterium to be etiologically associated with human disease, *M. leprae* remains one of the few bacterial species that still has not been cultivated on artificial medium or tissue culture. The multiplication of *M. leprae* in mouse footpads (albeit limited, with a doubling time of ~2 weeks) has provided a means to evaluate antimicrobial agents, monitor clinical trials, and screen vaccines. *M. leprae* grows best in cooler tissues (the skin, peripheral nerves, anterior chamber of the eye, upper respiratory tract, and testes), sparing warmer areas of the skin (the axilla, groin, scalp, and midline of the back).

Epidemiology

DEMOGRAPHICS

Leprosy is almost exclusively a disease of the developing world, affecting areas of Asia, Africa, Latin America, and the Pacific **(Fig. 166-1)**. While Africa has the highest disease prevalence, Asia has the most cases. More than 80% of the world's cases occur in a few countries: India, China, Myanmar, Indonesia, Brazil, Nigeria, Madagascar, and Nepal. Within endemic locales, the distribution of leprosy is quite uneven, with areas of high prevalence bordering on areas with little or no disease. In Brazil the

majority of cases occur in the Amazon basin and two western states, while in Mexico leprosy is mostly confined to the Pacific coast. Except as imported cases, leprosy is largely absent from the United States, Canada, and northwestern Europe. In the United States, ~4000 persons have leprosy and 100–200 new cases are reported annually, most of them in California, Texas, New York, and Hawaii among immigrants from Mexico, Southeast Asia, the Philippines, and the Caribbean. The comparative genomics of single-nucleotide polymorphisms support the likelihood that four distinct strains exist, having originated in East Africa or Central Asia. A mutation spread to Europe and subsequently underwent two separate mutations that were then followed by spread to West Africa and the Americas.

Figure 166-1

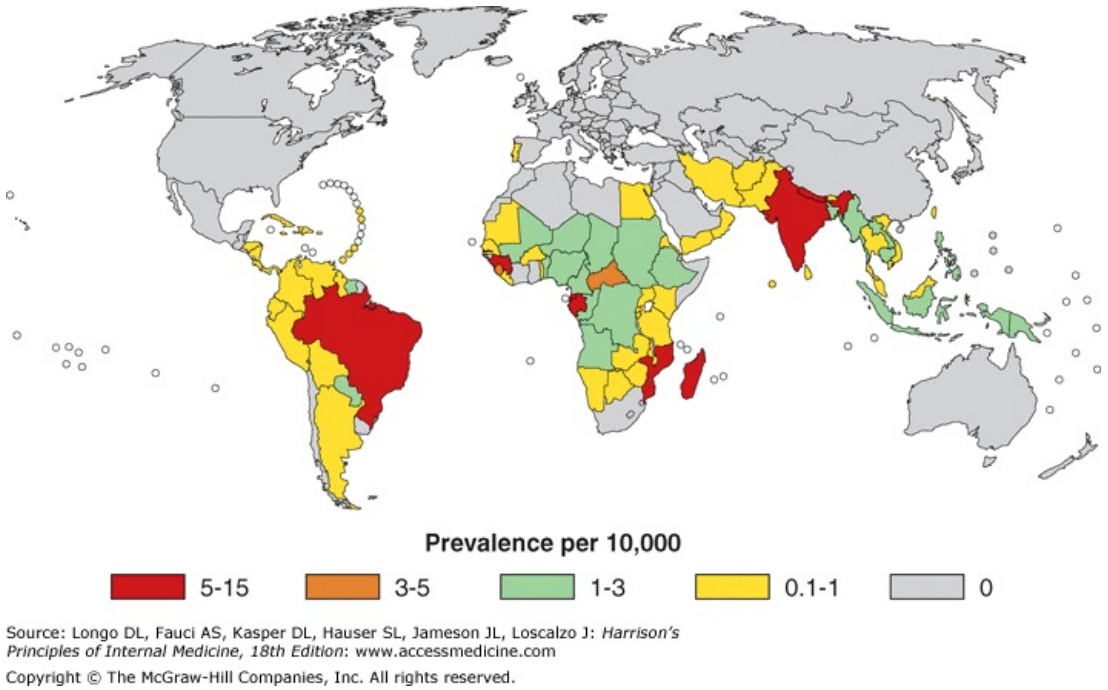

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated prevalence of leprosy at the turn of the millennium. Because data on leprosy prevalence in many endemic countries are unreliable, global prevalence is difficult to assess with any great degree of accuracy; however, it is not falling (see text). *(Courtesy of Patrick J. Brennan, PhD, with permission.)*

The global prevalence of leprosy is difficult to assess, given that many of the locales with high prevalence lack a significant medical or public health infrastructure. Estimates range from 0.6 to 8 million affected individuals. The lower estimate includes only persons who have not completed chemotherapy, excluding those who may be physically or psychologically damaged from leprosy and who may yet relapse or develop immune-mediated reactions. The higher figure includes patients whose infections probably are already cured and many who have no leprosy-related deformity or disability. Although the figures on the worldwide prevalence of leprosy are debatable, it is not falling; there are an estimated 600,000 new cases annually, 60% of them in India.

Leprosy is associated with poverty and rural residence. It appears not to be associated with AIDS, perhaps because of leprosy's long incubation period. Most individuals appear to be naturally immune to leprosy and do not develop disease manifestations after exposure. The time of peak onset is in the second and third decades of life. The most severe lepromatous form of leprosy is twice as common among men as among women and is rarely encountered in children. The frequency of the polar forms of leprosy in different countries varies widely and may in part be genetically determined; certain human leukocyte antigen (HLA) associations are known for both polar forms of leprosy (see below). Furthermore, variations in immunoregulatory genes are associated with an increased susceptibility to leprosy, particularly the multibacillary form. In India and Africa, 90% of cases are tuberculoid; in Southeast Asia, 50% are tuberculoid and 50% lepromatous; and in Mexico, 90% are lepromatous. (For definitions of disease types, see **Table 166-1** and "Clinical, Histologic, and Immunologic Spectrum," below.)

Table 166-1 Clinical, Bacteriologic, Pathologic, and Immunologic Spectrum of Leprosy

Feature	Tuberculoid (TT, BT) Leprosy	Borderline (BB, BL) Leprosy	Lepromatous (LL) Leprosy
Skin lesions	One or a few sharply defined annular	Intermediate between BT- and LL-type lesions;	Symmetric, poorly marginated, multiple infiltrated nodules and

	asymmetric macules or plaques with a tendency toward central clearing, elevated borders	ill-defined plaques with an occasional sharp margin; few or many in number	plaques or diffuse infiltration; xanthoma-like or dermatofibroma papules; leonine facies and eyebrow alopecia
Nerve lesions	Skin lesions anesthetic early; nerve near lesions sometimes enlarged; nerve abscesses most common in BT	Hypesthetic or anesthetic skin lesions; nerve trunk palsies, at times symmetric	Hypesthesia a late sign; nerve palsies variable; acral, distal, symmetric anesthesia common
Acid-fast bacilli (BI[a])	0–1+	3–5+	4–6+
Lymphocytes	2+	1+	0–1+
Macrophage differentiation	Epithelioid	Epithelioid in BB; usually undifferentiated, but may have foamy changes in BL	Foamy change the rule; may be undifferentiated in early lesions
Langhans giant cells	1–3+	—	—
Lepromin skin test	+++	—	—
Lymphocyte transformation test	Generally positive	1–10%	1–2%
CD4+/CD8+ T cell ratio in lesions	1.2	BB (NT); BL: 0.48	0.50
M. leprae PGL-1 antibodies	60%	85%	95%

[a]See text.

Abbreviations: BB, mid-borderline; BL, borderline lepromatous; BT, borderline tuberculoid; TT, polar tuberculoid; LL, polar lepromatous; BI, bacteriologic index; NT, not tested; PGL-1, phenolic glycolipid 1.

TRANSMISSION

The route of transmission of leprosy remains uncertain, and transmission routes may in fact be multiple. Nasal droplet infection, contact with infected soil, and even insect vectors have been considered the prime candidates. Aerosolized *M. leprae* can cause infection in immunosuppressed mice, and a sneeze from an untreated lepromatous patient may contain $>10^{10}$ AFB. Furthermore, both IgA antibody to *M. leprae* and genes of *M. leprae*—demonstrable by polymerase chain reaction (PCR)—have been found in the nose of individuals without signs of leprosy from endemic areas and in 19% of occupational contacts of lepromatous patients. Several lines of evidence implicate soil transmission. (1) In endemic countries such as India, leprosy is primarily a rural and not an urban disease. (2) *M. leprae* products reside in soil in endemic locales. (3) Direct dermal inoculation (e.g., during tattooing) may transmit *M. leprae*, and common sites of leprosy in children are the buttocks and thighs, suggesting that microinoculation of infected soil may transmit the disease. Evidence for insect vectors of leprosy includes the demonstration that bedbugs and mosquitoes in the vicinity of leprosaria regularly harbor *M. leprae* and that experimentally infected mosquitoes can transmit infection to mice. Skin-to-skin contact is generally not considered an important route of transmission.

In endemic countries, ~50% of leprosy patients have a history of intimate contact with an infected person (often a household member), while, for unknown reasons, leprosy patients in nonendemic locales can identify such contact only 10% of the time. Moreover, household contact with an infected lepromatous case carries an eventual risk of disease acquisition of ~10% in endemic areas as opposed to only 1% in nonendemic locales. Contact with a tuberculoid case carries a very low risk. Physicians and nurses caring for leprosy patients and the co-workers of these patients are not at risk for leprosy.

Although multilocus variable-number short-nucleotide tandem-repeat (VNTR) analyses have generally demonstrated considerable variability among isolates, highly similar and even identical VNTR results have been obtained with isolates from a limited number of families with multiple cases. Moreover, VNTR results have been similar for isolates within certain geographic locales and divergent for isolates within others. These findings suggest that genomic analyses may prove useful in the future for defining *M. leprae* transmission patterns.

M. leprae causes disease primarily in humans. However, in Texas and Louisiana, 15% of nine-banded armadillos are infected, and armadillo contact occasionally results in human disease. Armadillos

develop disseminated infection after IV inoculation of live *M. leprae*.

Clinical, Histologic, and Immunologic Spectrum

The incubation period prior to manifestation of clinical disease can vary between 2 and 40 years, although it is generally 5–7 years in duration. This long incubation period is probably, at least in part, a consequence of the extremely long doubling time for *M. leprae* (14 days in mice versus in vitro doubling times of 1 day and 20 min for *M. tuberculosis* and *Escherichia coli*, respectively). Leprosy presents as a spectrum of clinical manifestations that have bacteriologic, pathologic, and immunologic counterparts. The spectrum from polar tuberculoid (TT) to borderline tuberculoid (BT) to mid-borderline (BB, which is rarely encountered) to borderline lepromatous (BL) to polar lepromatous (LL) disease is associated with an evolution from asymmetric localized macules and plaques to nodular and indurated symmetric generalized skin manifestations, an increasing bacterial load, and loss of *M. leprae*–specific cellular immunity (Table 166-1). Distinguishing dermatopathologic characteristics include the number of lymphocytes, giant cells, and AFB as well as the nature of epithelioid cell differentiation. Where a patient presents on the clinical spectrum largely determines prognosis, complications, reactional states, and the intensity of antimicrobial therapy required.

TUBERCULOID LEPROSY

At the less severe end of the spectrum is tuberculoid leprosy, which encompasses TT and BT disease. In general, these forms of leprosy result in symptoms confined to the skin and peripheral nerves. The skin lesions of tuberculoid leprosy consist of one or a few hypopigmented macules or plaques **(Fig. 166-2)** that are sharply demarcated and hypesthetic, often have erythematous or raised borders, and are devoid of the normal skin organs (sweat glands and hair follicles) and thus are dry, scaly, and anhidrotic. AFB are generally absent or few in number. Tuberculoid leprosy patients may have asymmetric enlargement of one or a few peripheral nerves. Indeed, leprosy and certain rare hereditary neuropathies are the only human diseases associated with peripheral-nerve enlargement. Although any peripheral nerve may be enlarged (including small digital and supraclavicular nerves), those most commonly affected are the ulnar, posterior auricular, peroneal, and posterior tibial nerves, with associated hypesthesia and myopathy. TT leprosy is the most common form of the disease encountered in India and Africa but is virtually absent in Southeast Asia, where BT leprosy is frequent.

Figure 166-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tuberculoid (TT) leprosy: a well-defined, hypopigmented, anesthetic macule with anhidrosis and a raised granular margin *(arrowhead).*

In tuberculoid leprosy, T cells breach the perineurium, and destruction of Schwann cells and axons may be evident, resulting in fibrosis of the epineurium, replacement of the endoneurium with epithelial granulomas, and occasionally caseous necrosis. Such invasion and destruction of nerves in the dermis by T cells are pathognomonic for leprosy.

Circulating lymphocytes from patients with tuberculoid leprosy readily recognize *M. leprae* and its constituent proteins, patients have positive lepromin skin tests (see "Diagnosis," below), and—owing to a type 1 cytokine pattern in tuberculoid tissues—strong T cell and macrophage activation results in a localized infection. In tuberculoid leprosy tissue, there is a 2:1 predominance of helper CD4+ over CD8+ T lymphocytes. Tuberculoid tissues are rich in the mRNAs of the proinflammatory T_H1 family of cytokines: interleukin (IL) 2, interferon γ (IFN-γ), and IL-12; in contrast, IL-4, IL-5, and IL-10 mRNAs are scarce.

LEPROMATOUS LEPROSY

Lepromatous leprosy patients present with symmetrically distributed skin nodules **(Fig. 166-3)**, raised plaques, or diffuse dermal infiltration, which, when on the face, results in leonine facies. Late manifestations include loss of eyebrows (initially the lateral margins only) and eyelashes, pendulous earlobes, and dry scaling skin, particularly on the feet. In LL leprosy, bacilli are numerous in the skin (as many as 10^9/g), where they are often found in large clumps (*globi*), and in peripheral nerves, where they initially invade Schwann cells, resulting in foamy degenerative myelination and axonal degeneration and later in Wallerian degeneration. In addition, bacilli are plentiful in circulating blood and in all organ systems except the lungs and the central nervous system. Nevertheless, patients are afebrile, and there is no evidence of major organ system dysfunction. Found almost exclusively in western Mexico and the Caribbean is a form of lepromatous leprosy without visible skin lesions but with diffuse dermal infiltration and a demonstrably thickened dermis, termed *diffuse lepromatosis*. In lepromatous leprosy, nerve enlargement and damage tend to be symmetric, result from actual bacillary invasion, and are more insidious but ultimately more extensive than in tuberculoid leprosy. Patients with LL leprosy have acral, distal, symmetric peripheral neuropathy and a tendency toward symmetric nerve-trunk enlargement. They may also have signs and symptoms related to involvement of the upper respiratory tract, the anterior chamber of the eye, and the testes.

Figure 166-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lepromatous (LL) leprosy: advanced nodular lesions.

In untreated LL patients, lymphocytes regularly fail to recognize either *M. leprae* or its protein constituents, and lepromin skin tests are negative (see "Diagnosis," below). This loss of protective cellular immunity appears to be antigen-specific, as patients are not unusually susceptible to opportunistic infections, cancer, or AIDS and maintain delayed-type hypersensitivity to *Candida*, *Trichophyton*, mumps, tetanus toxoid, and even purified protein derivative of tuberculin. At times, *M. leprae*–specific anergy is reversible with effective chemotherapy. In LL tissues, there is a 2:1 ratio of CD8+ to CD4+ T lymphocytes. LL patients have a predominant T_H2 response and hyperglobulinemia, and LL tissues demonstrate a T_H2 cytokine profile, being rich in mRNAs for IL-4, IL-5, and IL-10 and poor in those for IL-2, IFN-γ, and IL-12. It appears that cytokines mediate a protective tissue response in leprosy, as injection of IFN-γ or IL-2 into lepromatous lesions causes a loss of AFB and histopathologic conversion toward a tuberculoid pattern. Macrophages of lepromatous leprosy patients appear to be functionally intact; circulating monocytes exhibit normal microbicidal function and responsiveness to IFN-γ.

REACTIONAL STATES

Lepra reactions comprise several common immunologically mediated inflammatory states that cause considerable morbidity. Some of these reactions precede diagnosis and the institution of effective

antimicrobial therapy; indeed, these reactions may precipitate presentation for medical attention and diagnosis. Other reactions occur after the initiation of appropriate chemotherapy and may cause patients to perceive that their leprosy is worsening and to lose confidence in conventional therapy. Only by warning patients of the potential for these reactions and describing their manifestations can physicians treating leprosy patients ensure continued credibility.

Type 1 Lepra Reactions (Downgrading and Reversal Reactions)

Type 1 lepra reactions occur in almost half of patients with borderline forms of leprosy but not in patients with pure lepromatous disease. Manifestations include classic signs of inflammation within previously involved macules, papules, and plaques and, on occasion, the appearance of new skin lesions, neuritis, and (less commonly) fever—generally low-grade. The nerve trunk most commonly involved in this process is the ulnar nerve at the elbow, which may be painful and exquisitely tender. If patients with affected nerves are not treated promptly with glucocorticoids (see below), irreversible nerve damage may result in as little as 24 h. The most dramatic manifestation is footdrop, which occurs when the peroneal nerve is involved.

When type 1 lepra reactions precede the initiation of appropriate antimicrobial therapy, they are termed *downgrading reactions*, and the case becomes histologically more lepromatous; when they occur after the initiation of therapy, they are termed *reversal reactions*, and the case becomes more tuberculoid. Reversal reactions often occur in the first months or years after the initiation of therapy but may also develop several years thereafter.

Edema is the most characteristic microscopic feature of type 1 lepra lesions, whose diagnosis is primarily clinical. Reversal reactions are typified by a T_H1 cytokine profile, with an influx of CD4+ T helper cells and increased levels of IFN-γ and IL-2. In addition, type 1 reactions are associated with large numbers of T cells bearing γ/δ receptors—a unique feature of leprosy.

Type 2 Lepra Reactions: Erythema Nodosum Leprosum

Erythema nodosum leprosum (ENL) **(Fig. 166-4)** occurs exclusively in patients near the lepromatous end of the leprosy spectrum (BL-LL), affecting nearly 50% of this group. Although ENL may precede leprosy diagnosis and initiation of therapy (sometimes, in fact, prompting the diagnosis), in 90% of cases it follows the institution of chemotherapy, generally within 2 years. The most common features of ENL are crops of painful erythematous papules that resolve spontaneously in a few days to a week but may recur; malaise; and fever that can be profound. However, patients may also experience symptoms of neuritis, lymphadenitis, uveitis, orchitis, and glomerulonephritis and may develop anemia, leukocytosis, and abnormal liver function tests (particularly increased aminotransferase levels). Individual patients may have either a single bout of ENL or chronic recurrent manifestations. Bouts may be either mild or severe and generalized; in rare instances, ENL results in death. Skin biopsy of ENL papules reveals vasculitis or panniculitis, sometimes with many lymphocytes but characteristically with polymorphonuclear leukocytes as well.

Figure 166-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Moderately severe skin lesions of erythema nodosum leprosum (ENL), some with pustulation and ulceration.

Elevated levels of circulating tumor necrosis factor (TNF) have been demonstrated in ENL; thus, TNF may play a central role in the pathobiology of this syndrome. ENL is thought to be a consequence of

immune complex deposition, given its T_H2 cytokine profile and its high levels of IL-6 and IL-8. However, in ENL tissue, the presence of HLA-DR framework antigen of epidermal cells—considered a marker for a delayed-type hypersensitivity response—and evidence of higher levels of IL-2 and IFN-γ than are usually seen in polar lepromatous disease suggest an alternative mechanism.

Lucio's Phenomenon

Lucio's phenomenon is an unusual reaction seen exclusively in patients from the Caribbean and Mexico who have the diffuse lepromatosis form of lepromatous leprosy, most often those who are untreated. Patients with this reaction develop recurrent crops of large, sharply marginated, ulcerative lesions—particularly on the lower extremities—that may be generalized and, when so, are frequently fatal as a result of secondary infection and consequent septic bacteremia. Histologically, the lesions are characterized by ischemic necrosis of the epidermis and superficial dermis, heavy parasitism of endothelial cells with AFB, and endothelial proliferation and thrombus formation in the larger vessels of the deeper dermis. Like ENL, the Lucio phenomenon is probably mediated by immune complexes.

COMPLICATIONS

The Extremities

Complications of the extremities in leprosy patients are primarily a consequence of neuropathy leading to insensitivity and myopathy. Insensitivity affects fine touch, pain, and heat receptors but generally spares position and vibration appreciation. The most commonly affected nerve trunk is the ulnar nerve at the elbow, whose involvement results in clawing of the fourth and fifth fingers, loss of dorsal interosseous musculature in the affected hand, and loss of sensation in these distributions. Median nerve involvement in leprosy impairs thumb opposition and grasp; radial nerve dysfunction, although rare in leprosy, leads to wristdrop. Tendon transfers can restore hand function but should not be performed until 6 months after the initiation of antimicrobial therapy and the conclusion of episodes of acute neuritis.

Plantar ulceration, particularly at the metatarsal heads, is probably the most frequent complication of leprous neuropathy. Therapy requires careful debridement; administration of appropriate antibiotics; avoidance of weight-bearing until ulcerations are healed, with slowly progressive ambulation thereafter; and wearing of special shoes to prevent recurrence.

Footdrop as a result of peroneal nerve palsy should be treated with a simple nonmetallic brace within the shoe or with surgical correction attained by tendon transfers. Although uncommon, Charcot's joints, particularly of the foot and ankle, may result from leprosy.

The loss of distal digits in leprosy is a consequence of insensitivity, trauma, secondary infection, and —in lepromatous patients—a poorly understood and sometimes profound osteolytic process. Conscientious protection of the extremities during cooking and work and the early institution of therapy have substantially reduced the frequency and severity of distal digit loss in recent times.

The Nose

In lepromatous leprosy, bacillary invasion of the nasal mucosa can result in chronic nasal congestion and epistaxis. Saline nose drops may relieve these symptoms. Long-untreated LL leprosy may further result in destruction of the nasal cartilage, with consequent saddle-nose deformity or anosmia (more common in the preantibiotic era than at present). Nasal reconstructive procedures can ameliorate significant cosmetic defects.

The Eye

Owing to cranial nerve palsies, lagophthalmos and corneal insensitivity may complicate leprosy, resulting in trauma, secondary infection, and (without treatment) corneal ulcerations and opacities. For patients with these conditions, eyedrops during the day and ointments at night provide some protection from such consequences. Furthermore, in LL leprosy, the anterior chamber of the eye is invaded by bacilli, and ENL may result in uveitis, with consequent cataracts and glaucoma. Thus leprosy is a major cause of blindness in the developing world. Slit-lamp evaluation of LL patients often reveals "corneal beading," representing globi of *M. leprae*.

The Testes

M. leprae invades the testes, while ENL may cause orchitis. Thus males with lepromatous leprosy often manifest mild to severe testicular dysfunction, with an elevation of luteinizing and follicle-stimulating hormones, decreased testosterone, and aspermia or hypospermia in 85% of LL patients but in only 25% of BL patients. LL patients may become impotent and infertile. Impotence is sometimes responsive to testosterone replacement.

Amyloidosis

Secondary amyloidosis is a complication of LL leprosy and ENL that is encountered infrequently in the antibiotic era. This complication may result in abnormalities of hepatic and particularly renal function.

Nerve Abscesses

Patients with various forms of leprosy, but particularly those with the BT form, may develop abscesses

of nerves (most commonly the ulnar) with an adjacent cellulitic appearance of the skin. In such conditions, the affected nerve is swollen and exquisitely tender. Although glucocorticoids may reduce signs of inflammation, rapid surgical decompression is necessary to prevent irreversible sequelae.

Diagnosis

Leprosy most commonly presents with both characteristic skin lesions and skin histopathology. Thus the disease should be suspected when a patient from an endemic area has suggestive skin lesions or peripheral neuropathy. The diagnosis should be confirmed by histopathology. In tuberculoid leprosy, lesional areas—preferably the advancing edge—must be biopsied because normal-appearing skin does not have pathologic features. In lepromatous leprosy, nodules, plaques, and indurated areas are optimal biopsy sites, but biopsies of normal-appearing skin are also generally diagnostic. Lepromatous leprosy is associated with diffuse hyperglobulinemia, which may result in false-positive serologic tests (e.g., Venereal Disease Research Laboratory, rheumatoid arthritis, and antinuclear antibody tests) and therefore may cause diagnostic confusion. On occasion, tuberculoid lesions may not (1) appear typical, (2) be hypesthetic, and (3) contain granulomas but only nonspecific lymphocytic infiltrates. In such instances, two of these three characteristics are considered sufficient for a diagnosis. It is preferable to overdiagnose leprosy rather than to allow a patient to remain untreated.

IgM antibodies to PGL-1 are found in 95% of patients with untreated lepromatous leprosy; the titer decreases with effective therapy. However, in tuberculoid leprosy—the form of disease most often associated with diagnostic uncertainty owing to the absence or paucity of AFB—patients have significant antibodies to PGL-1 only 60% of the time; moreover, in endemic locales, exposed individuals without clinical leprosy may harbor antibodies to PGL-1. Thus PGL-1 serology is of little diagnostic utility in tuberculoid leprosy. Heat-killed *M. leprae* (lepromin) has been used as a skin test reagent. It generally elicits a reaction in tuberculoid leprosy patients, may do so in individuals without leprosy, and gives negative results in lepromatous leprosy patients; consequently, it is likewise of little diagnostic value. Unfortunately, PCR of the skin for *M. leprae*, although positive in LL and BL leprosy, yields negative results in 50% of tuberculoid leprosy cases, again offering little diagnostic assistance.

Included in the differential diagnosis of lesions that resemble leprosy are sarcoidosis, leishmaniasis, lupus vulgaris, dermatofibroma, histiocytoma, lymphoma, syphilis, yaws, granuloma annulare, and various other disorders causing hypopigmentation (notably pityriasis alba, tinea, and vitiligo). Sarcoidosis may result in perineural inflammation, but actual granuloma formation within dermal nerves is pathognomonic for leprosy. In lepromatous leprosy, sputum specimens may be loaded with AFB—a finding that can be inappropriately interpreted as representing pulmonary tuberculosis.

Treatment: Leprosy

ANTIMICROBIAL THERAPY

Active Agents

Established agents used to treat leprosy include dapsone (50–100 mg/d), clofazimine (50–100 mg/d, 100 mg three times weekly, or 300 mg monthly), and rifampin (600 mg daily or monthly; see "Choice of Regimens," below). Of these drugs, only rifampin is bactericidal. The sulfones (folate antagonists), the foremost of which is dapsone, were the first antimicrobial agents found to be effective for the treatment of leprosy and are still the mainstay of therapy. With sulfone treatment, skin lesions resolve and numbers of viable bacilli in the skin are reduced. Although primarily bacteriostatic, dapsone monotherapy results in only a 2.5% resistance-related relapse rate; after ≥18 years of therapy and subsequent discontinuation, only another 10% of patients relapse, developing new, usually asymptomatic, shiny, "histoid" nodules. Dapsone is generally safe and inexpensive. Individuals with glucose-6-phosphate dehydrogenase deficiency who are treated with dapsone may develop severe hemolysis; those without this deficiency also have reduced red cell survival and a hemoglobin decrease averaging 1 g/dL. Dapsone's usefulness is limited occasionally by allergic dermatitis and rarely by the sulfone syndrome (including high fever, anemia, exfoliative dermatitis, and a mononucleosis-type blood picture). It must be remembered that rifampin induces microsomal enzymes, necessitating increased doses of medications such as glucocorticoids and oral birth control regimens. Clofazimine is often cosmetically unacceptable to light-skinned leprosy patients because it causes a red-black skin discoloration that accumulates, particularly in lesional areas, and makes the patient's diagnosis obvious to members of the community.

Other antimicrobial agents active against *M. leprae* in animal models and at the usual daily doses used in clinical trials include ethionamide/prothionamide; the aminoglycosides streptomycin, kanamycin, and amikacin (but not gentamicin or tobramycin); minocycline; clarithromycin; and several fluoroquinolones, particularly ofloxacin. Next to rifampin, minocycline, clarithromycin, and ofloxacin appear to be most bactericidal for *M. leprae*, but these drugs have not been used extensively in leprosy control programs. Most recently, rifapentine and moxifloxacin have been found to be especially potent against *M. leprae* in mice. In a clinical trial in lepromatous leprosy, moxifloxacin was profoundly bactericidal, matched in potency only by rifampin.

Choice of Regimens

Antimicrobial therapy for leprosy must be individualized, depending on the clinical/pathologic form of the disease encountered. Tuberculoid leprosy, which is associated with a low bacterial burden and a protective cellular immune response, is the easiest form to treat and can be cured reliably with a finite course of chemotherapy. In contrast, lepromatous leprosy may have a higher bacillary load than any other human bacterial disease, and the absence of a salutary T cell repertoire requires prolonged or even lifelong chemotherapy. Hence, careful classification of disease prior to therapy is important.

In developed countries, clinical experience with leprosy classification is limited; fortunately, however, the resources needed for skin biopsy are highly accessible and pathologic interpretation is readily available. In developing countries, clinical expertise is greater but is now waning substantially as the care of leprosy patients is integrated into general health services. In addition, access to dermatopathology services is often limited. In such instances, skin smears may prove useful, but in many locales access to the resources needed for their preparation and interpretation may also be unavailable. Use of skin smears is no longer encouraged by the World Health Organization (WHO) and is often replaced by mere counting of lesions, which, together with the lack of histopathology, may negatively affect decisions about chemotherapy, increase the potential for reactions, and worsen the ultimate prognosis. A reasoned approach to the treatment of leprosy is confounded by these and several other issues:

1. Even without therapy, TT leprosy may heal spontaneously, and prolonged dapsone monotherapy (even for LL leprosy) is generally curative in 80% of cases.
2. In tuberculoid disease, it is common for no bacilli to be found in the skin prior to therapy, and thus there is no objective measure of therapeutic success. Furthermore, despite adequate treatment, TT and particularly BT lesions often resolve little or incompletely, while relapse and late type 1 lepra reactions can be difficult to distinguish.
3. LL leprosy patients commonly harbor viable persistent *M. leprae* organisms after prolonged intensive therapy; the propensity of these organisms to initiate clinical relapse is unclear. Because relapse in LL patients after discontinuation of rifampin-containing regimens usually begins only after 7–10 years, follow-up over the very long term is necessary to assess ultimate clinical outcomes.
4. Even though primary dapsone resistance is exceedingly rare and multidrug therapy is generally recommended (at least for lepromatous leprosy), there is a paucity of information from experimental animals and clinical trials on the optimal combination of antimicrobial agents, dosing schedule, or duration of therapy.

In 1982, the WHO made recommendations for "the chemotherapy of leprosy for control programs." These recommendations came on the heels of the demonstration of the relative success of long-term dapsone monotherapy and in the context of concerns about dapsone resistance. Other complicating considerations included the limited resources available for leprosy care in the very areas where it is most prevalent and the frustration and discouragement of patients and program managers with the previous requirement for lifelong therapy for many leprosy patients. Thus, for the first time, the WHO delineated a finite duration of therapy for all forms of leprosy and—given the prohibitive cost of daily rifampin treatment in developing countries—encouraged the monthly administration of this agent as part of a multidrug regimen. Over the ensuing years, the WHO recommendations have been broadly implemented, and the duration of therapy required, particularly for lepromatous leprosy, has been progressively shortened. For treatment purposes, the WHO classifies patients as *paucibacillary* or *multibacillary*. Previously, patients without demonstrable AFB in the dermis were classified as paucibacillary and those with AFB as multibacillary. Currently, in light of the perceived unreliability of skin smears in the field, patients are classified as multibacillary if they have six or more skin lesions and as paucibacillary if they have fewer. (Unfortunately, this classification method has been found wanting, as some patients near the lepromatous pole have only one or a few skin lesions.) The WHO recommends that paucibacillary adults be treated with 100 mg of dapsone daily and 600 mg of rifampin monthly (supervised) for 6 months **(Table 166-2)**. For patients with single-lesion paucibacillary leprosy, the WHO recommends as an alternative a single dose of rifampin (600 mg), ofloxacin (400 mg), and minocycline (100 mg). Multibacillary adults should be treated with 100 mg of dapsone plus 50 mg of clofazimine daily (unsupervised) and with 600 mg of rifampin plus 300 mg of clofazimine monthly (supervised). Originally, the WHO recommended that lepromatous patients be treated for 2 years or until smears became negative (generally in ~5 years); subsequently, the acceptable course was reduced to 1 year—a change that remains especially controversial in the absence of supporting clinical trials.

Table 166-2 Antimicrobial Regimens Recommended for the Treatment of Leprosy in Adults

Form of Leprosy	More Intensive Regimen	WHO Recommended Regimen (1982)
Tuberculoid (paucibacillary)	Dapsone (100 mg/d) for 5 years	Dapsone (100 mg/d, unsupervised) *plus* rifampin (600 mg/month, supervised) for 6 months

Lepromatous (multibacillary)	Rifampin (600 mg/d) for 3 years *plus* dapsone (100 mg/d) indefinitely	Dapsone (100 mg/d) *plus* clofazimine (50 mg/d), unsupervised; *and* rifampin (600 mg) *plus* clofazimine (300 mg) monthly (supervised) for 1–2 years

Note: See text for discussion and comparison of WHO recommendations and more intensive approach as well as alternative WHO regimen for single-lesion paucibacillary leprosy.

Several factors have caused many authorities to question the WHO recommendations and to favor a more intensive approach. Among these factors are—for multibacillary patients—a high (double-digit) relapse rate in three locales (reaching 20–40% in one locale, with the rate directly related to the initial bacterial burden) and—for paucibacillary patients—demonstrable lesional activity for years in fully half of patients after the completion of therapy. The more intensive approach (Table 166-2) calls for tuberculoid leprosy to be treated with dapsone (100 mg/d) for 5 years and for lepromatous leprosy to be treated with rifampin (600 mg/d) for 3 years and with dapsone (100 mg/d) throughout life.

With effective antimicrobial therapy, new skin lesions and signs and symptoms of peripheral neuropathy cease appearing. Nodules and plaques of lepromatous leprosy noticeably flatten in 1–2 months and resolve in 1 year or a few years, while tuberculoid skin lesions may disappear, improve, or remain relatively unchanged. Although the peripheral neuropathy of leprosy may improve somewhat in the first few months of therapy, rarely is it significantly alleviated by treatment.

Given the recent findings that moxifloxacin, like rifampin, is profoundly bactericidal in leprosy patients and that short-course chemotherapy for tuberculosis is possible only when two or more bactericidal agents are used, a moxifloxacin/rifampin-based regimen including either minocycline or clarithromycin appears promising; such a regimen may prove to be more reliably curative than WHO-recommended multidrug therapy for lepromatous leprosy and may allow a considerably shorter course of treatment.

THERAPY FOR REACTIONS

Type 1

Type 1 lepra reactions are best treated with glucocorticoids (e.g., prednisone, initially at doses of 40–60 mg/d). As the inflammation subsides, the glucocorticoid dose can be tapered, but steroid therapy must be continued for at least 3–6 months lest recurrence supervene. Because of the myriad toxicities of prolonged glucocorticoid therapy, the indications for its initiation are strictly limited to lesions whose intense inflammation poses a threat of ulceration; lesions at cosmetically important sites, such as the face; and cases in which neuritis is present. Mild to moderate lepra reactions that do not meet these criteria should be tolerated and glucocorticoid treatment withheld. Thalidomide is ineffective against type 1 lepra reactions. Clofazimine (200–300 mg/d) is of questionable benefit but in any event is far less efficacious than glucocorticoids.

Type 2

Treatment of ENL must be individualized. If ENL is mild (i.e., without fever or other organ involvement, with occasional crops of only a few skin papules), it may be treated with antipyretics alone. However, in cases with many skin lesions, fever, malaise, and other tissue involvement, brief courses (1–2 weeks) of glucocorticoids (initially 40–60 mg/d) are often effective. With or without therapy, individual inflamed papules last for >1 week. Successful therapy is defined by the cessation of skin lesion development and the disappearance of other systemic signs and symptoms. If, despite two courses of glucocorticoid therapy, ENL appears to be recurring and persisting, treatment with thalidomide (100–300 mg nightly) should be initiated, with the dose depending on the initial severity of the reaction. Because even a single dose of thalidomide administered early in pregnancy may result in severe birth defects, including phocomelia, the use of this drug in the United States for the treatment of fertile female patients is tightly regulated and requires informed consent, prior pregnancy testing, and maintenance of birth control measures. Although the mechanism of thalidomide's dramatic action against ENL is not entirely clear, the drug's efficacy is probably attributable to its reduction of TNF levels and IgM synthesis and its slowing of polymorphonuclear leukocyte migration. After the reaction is controlled, lower doses of thalidomide (50–200 mg nightly) are effective in preventing relapses of ENL. Clofazimine in high doses (300 mg nightly) has some efficacy against ENL, but its use permits only a modest reduction of the glucocorticoid dose necessary for ENL control.

Lucio's Phenomenon

Neither glucocorticoids nor thalidomide is effective against this syndrome. Optimal wound care and therapy for bacteremia are indicated. Ulcers tend to be chronic and heal poorly. In severe cases, exchange transfusion may prove useful.

Prevention and Control

Vaccination at birth with bacille Calmette-Guérin (BCG) has proved variably effective in preventing leprosy: the results have ranged from total inefficacy to 80% efficacy. The addition of heat-killed *M. leprae* to BCG does not increase the effectiveness of vaccine. Because whole mycobacteria contain large amounts of lipids and carbohydrates that have proved in vitro to be immunosuppressive for lymphocytes and macrophages, *M. leprae* proteins may prove to be superior vaccines. Data from a

mouse model support this possibility.

Chemoprophylaxis with dapsone may reduce the number of cases of tuberculoid leprosy but not of lepromatous leprosy and hence is not recommended, even for household contacts. Because leprosy transmission appears to require close prolonged household contact, hospitalized patients need not be isolated.

In 1992, the WHO—on the basis of that organization's treatment recommendations—launched a landmark campaign to eliminate leprosy as a public health problem by the year 2000 (goal, <1 case per 10,000 population). The campaign mobilized and energized nongovernmental organizations and national health services to treat leprosy with multiple drugs and to clean up outdated registries. In these respects, the effort has proven hugely successful, with >6 million patients completing therapy. However, the target of leprosy elimination has not yet been reached. In fact, the success of the WHO campaign in reducing the number of cases worldwide has been largely attributable to the redefinition of what constitutes a case of leprosy. Formerly calculated by disease prevalence, the case count is now limited to those not yet treated with multiple drugs. In each of the 23 countries with the largest number of leprosy cases, the annual incidence of leprosy is stable or actually rising. Furthermore, after the completion of therapy, when a patient is no longer considered to represent a "case," half of all patients continue to manifest disease activity for years; relapse rates (at least for multibacillary patients) are unacceptably high; disabilities and deformities go unchecked; and the social stigma of the disease persists.

During most of the twentieth century, nongovernmental organizations, particularly Christian missionaries, provided a medical infrastructure devoted to the care and treatment of leprosy patients —the envy of those with other medical priorities in the developing world. With the public perception that leprosy is near eradication, resources for patient care are rapidly being diverted, and the burden of patient care is being transferred to nonexistent or overloaded national health services and to health workers who lack the tools and skills needed for disease diagnosis, classification, and nuanced therapy (particularly in cases of reactional neuritis). Thus the prerequisites for a salutary outcome increasingly go unmet.

FURTHER READINGS

Cole ST et al: Massive gene decay in the leprosy bacillus. Nature 409:1007, 2001[PMID: 11234002]

Fajardo TT et al: A comparative clinical trial in multibacillary leprosy with long-term relapse rates of four different multidrug regimens. Am J Trop Med Hyg 81:330, 2009[PMID: 19635893]

Gelber RH et al: The relapse rate in MB leprosy patients treated with 2-years of WHO-MDT is not low. Int J Lepr Other Mycobact Dis 72:493, 2004[PMID: 15755209]

Lockwood D: Leprosy elimination—a virtual phenomenon or a reality? BMJ 324:1516, 2002[PMID: 12077045]

Modlin RL, Rea TH: Immunology of leprosy granulomas. Springer Semin Immunopathol 10:359, 1998

Monet M et al: On the origin of leprosy. Science 308:1040, 2005

Pardillo FE et al: Powerful bactericidal activity of moxifloxacin in human leprosy.

Antimicrob Agents Chemother 52:3113, 2008

Ridley DS: Histological classification and the immunological spectrum of leprosy. Bull World Health Organ 51:451, 1974[PMID: 4549496]

WHO Expert Committee on Leprosy: Seventh Report. WHO Tech Rep Ser No. 874. Geneva, World Health Organization, 1998

Zhang FR et al: Genomewide association study of leprosy. N Engl J Med 361:2609, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NONTUBERCULOUS MYCOBACTERIAL INFECTIONS: INTRODUCTION

Several terms—nontuberculous mycobacteria (NTM), atypical mycobacteria, mycobacteria other than tuberculosis, and environmental mycobacteria—all refer to mycobacteria other than *Mycobacterium tuberculosis*, its close relatives (*M. bovis, M. caprae, M. africanum, M. pinnipedii, M. canetti*), and *M. leprae*. The number of identified species of NTM is growing and will continue to do so because of the use of DNA sequence typing for speciation. The number of known species currently exceeds 150. NTM are highly adaptable and can inhabit hostile environments, including industrial solvents.

Epidemiology

NTM are ubiquitous in soil and water. Specific organisms have recurring niches, such as *M. simiae* in certain aquifers, *M. fortuitum* in pedicure baths, and *M. immunogenum* in metalworking fluids. Most NTM cause disease in humans only rarely unless some aspect of host defense is impaired, as in bronchiectasis, or breached, as by inoculation (e.g., liposuction, trauma). There are no known instances of human-to-human transmission of NTM. Because infections due to NTM are rarely reported to health agencies and because their identification is sometimes problematic, reliable data on incidence and prevalence are lacking. Disseminated disease denotes significant immune dysfunction (e.g., advanced HIV infection), whereas pulmonary disease, which is much more common, is highly associated with pulmonary epithelial defects but not with systemic immunodeficiency.

In the United States, the incidence and prevalence of pulmonary infection with NTM, mostly in association with bronchiectasis (Chap. 258), have for many years been several-fold higher than the corresponding figures for tuberculosis, and rates of the former are increasing among the elderly. Among patients with cystic fibrosis, who often have bronchiectasis, rates of clinical infection with NTM range from 3% to 15%, with even higher rates among older patients. Although NTM may be recovered from the sputa of many individuals, it is critical to differentiate active disease from commensal harboring of the organisms. A scheme to help with the proper diagnosis of pulmonary infection caused by NTM has been developed by the American Thoracic Society and is widely used. The bulk of nontuberculous mycobacterial disease in North America is due to *M. kansasii*, organisms of the *M. avium* complex (MAC), and *M. abscessus*.

In Europe, Asia, and Australia, the distribution of NTM in clinical specimens is roughly similar to that in North America, with MAC species and rapidly growing organisms such as *M. abscessus* encountered frequently. *M. xenopi* and *M. malmoense* are especially prominent in northern Europe. *M. ulcerans* causes the distinct clinical entity Buruli ulcer, which occurs throughout tropical zones, especially in western Africa. *M. marinum* is a common cause of cutaneous and tendon infections in coastal regions and among individuals exposed to fish tanks or swimming pools.

The true international epidemiology of infections due to NTM is hard to determine since the isolation of these organisms often is not reported and speciation often is not performed. The increasing ease of identification and speciation of these organisms should have a major impact on the description of their international epidemiology in the next few years.

Pathobiology

Because exposure to NTM is essentially universal and disease is rare, it can be assumed that normal host defenses against these organisms must be strong and that otherwise healthy individuals in whom significant disease develops are highly likely to have specific susceptibility factors that permit NTM to become established, multiply, and cause disease. At the advent of HIV infection, CD4+ T lymphocytes were recognized as key effector cells against NTM; the development of disseminated MAC disease was highly correlated with a decline in CD4+ T lymphocyte numbers. Such a decrease has also been implicated in disseminated MAC infection in patients with idiopathic CD4+ T lymphocytopenia. Potent inhibitors of tumor necrosis factor α (TNF-α), such as infliximab, adalimumab, certolizumab, and etanercept, can neutralize this critical cytokine. The occasional result is severe mycobacterial or fungal infection; these associations indicate that TNF-α is a crucial element in mycobacterial control. However, in cases without the above risk factors, much of the genetic basis of susceptibility to disseminated infection with NTM is accounted for by specific mutations in the interferon γ (IFN-γ)/interleukin 12 (IL-12) synthesis and response pathways.

Mycobacteria are typically phagocytosed by macrophages, which respond with the production of IL-12, a heterodimer composed of IL-12p35 and IL-12p40 moieties that together make up IL-12p70. IL-12

activates T lymphocytes and natural killer cells through binding to its receptor (composed of IL-12Rβ1 and IL-12Rβ2/IL-23R), with consequent phosphorylation of STAT4. IL-12 stimulation of STAT4 leads to secretion of IFN-γ, which activates neutrophils and macrophages to produce reactive oxidants, increase expression of the major histocompatibility complex and Fc receptors, and concentrate certain antibiotics intracellularly. Signaling by IFN-γ through its receptor (composed of IFN-γR1 and IFN-γR2) leads to phosphorylation of STAT1, which in turn regulates IFN-γ-responsive genes, such as those coding for IL-12 and TNF-α. TNF-α signals through its own receptor via a downstream complex containing the nuclear factor κB (NFκB) essential modulator (NEMO). Therefore, the positive feedback loop between IFN-γ and IL-12/IL-23 drives the immune response to mycobacteria and other intracellular infections. These genes are known to be the critical ones in the pathway of mycobacterial control: specific Mendelian mutations have been identified in IFN-γR1, IFN-γR2, STAT1, IL-12A, IL-12Rβ1, and NEMO **(Fig. 167-1)**. Despite the identification of genes associated with disseminated disease, only ~ 50% of cases of disseminated nontuberculous mycobacterial infections that are not associated with HIV infection have a genetic diagnosis; the implication is that more mycobacterial susceptibility genes and pathways remain to be identified.

Figure 167-1

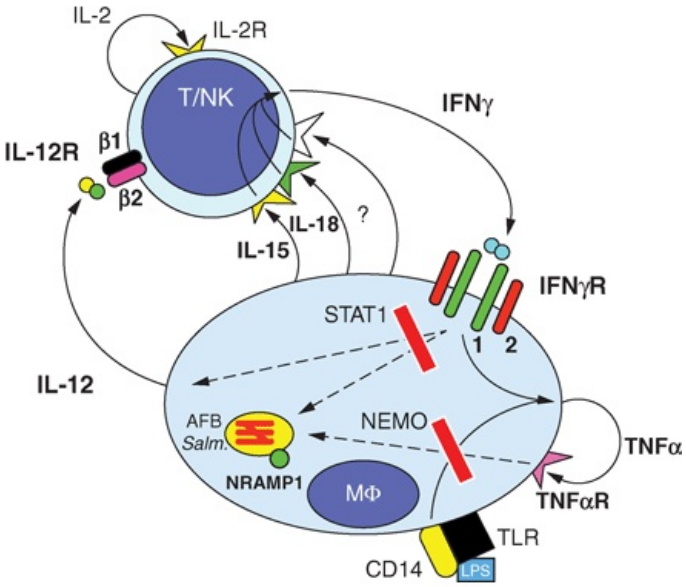

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cytokine interactions of infected macrophages (Mφ) with T and natural killer (NK) lymphocytes. Infection of macrophages by mycobacteria (AFB) leads to the release of heterodimeric interleukin 12 (IL-12). IL-12 acts on its receptor complex, with consequent STAT4 activation and production of homodimeric interferon γ (IFNγ). IFNγ acts through its receptor to activate STAT1, to stimulate the production of tumor necrosis factor α(TNFα), and to kill intracellular organisms such as mycobacteria, salmo-nellae, and some fungi. Homotrimeric TNFα acts through its receptor and requires nuclear factor κB essential modulator (NEMO) to activate nuclear factor κB, which also contributes to the killing of intracellular bacteria. Both IFNγ and TNFα lead to upregulation of IL-12. TNFα-blocking antibodies work either by blocking the ligand (infliximab, adalimumab, certolizumab) or by providing soluble receptor (etanercept). Mutations in IFNγR1, IFNγR2, IL-12p40, IL-12Rβ1, STAT1, and NEMO have been associated with a predisposition to mycobacterial infections. Other cytokines, such as IL-15 and IL-18, also contribute to IFNγ production. Signaling through the Toll-like receptor (TLR) complex and CD14 also upregulates TNFα production. LPS, lipopolysaccharide; NRAMP1, natural resistance-associated macrophage protein 1.

In contrast to the recognized genes and mechanisms associated with disseminated nontuberculous mycobacterial infection, the best-recognized underlying condition for pulmonary infection with NTM is bronchiectasis (Chap. 258). Most of the well-characterized forms of bronchiectasis, including cystic fibrosis, primary ciliary dyskinesia, STAT3-deficient hyper-IgE syndrome, and idiopathic bronchiectasis, have high rates of association with nontuberculous mycobacterial infection. The precise mechanism by which bronchiectasis predisposes to locally destructive but not systemic involvement is unknown.

Unlike disseminated or pulmonary infection, "hot-tub lung" represents pulmonary hypersensitivity to NTM—most commonly MAC organisms—growing in underchlorinated, often indoor hot tubs.

Clinical Manifestations

DISSEMINATED DISEASE

Disseminated MAC or *M. kansasii* infections in patients with advanced HIV infection are now

uncommon in North America because of effective antimycobacterial prophylaxis and improved treatment of HIV infection. When such mycobacterial disease was common, the portal of entry was the bowel, with spread to bone marrow and the bloodstream. Surprisingly, disseminated infections with rapidly growing NTM (e.g., *M. abscessus*, *M. fortuitum*) are very rare in HIV-infected patients, even those with very advanced HIV infection. Because these organisms are of low intrinsic virulence and disseminate only in conjunction with impaired immunity, disseminated disease can be indolent and progressive over weeks to months. Typical manifestations of malaise, fever, and weight loss are often accompanied by organomegaly, lymphadenopathy, and anemia. Since special cultures or stains are required to identify the organisms, the most critical step in diagnosis is to suspect infection with NTM. Blood cultures may be negative, but involved organs typically have significant organism burdens, sometimes with a grossly impaired granulomatous response. In a child, disseminated involvement (i.e., involvement of two or more organs) without an underlying iatrogenic cause should prompt an investigation of the IFN-γ/IL-12 pathway. Recessive mutations in IFN-γR1 and IFN-γR2 typically lead to severe infection with NTM. In contrast, dominant negative mutations in IFN-γR1, which lead to overaccumulation of a defective interfering mutant receptor on the cell surface, inhibit normal IFN-γ signaling and thus lead to nontuberculous mycobacterial osteomyelitis. Dominant negative mutations in STAT1 and recessive mutations in IL-12Rβ1 can have variable phenotypes consistent with their residual capacities for IFN-γ synthesis and response. Male patients who have disseminated nontuberculous mycobacterial infections along with conical, peg, or missing teeth and an abnormal hair pattern should be evaluated for defects in the pathway that activates NFκB through NEMO. These patients may have associated immune globulin defects as well. A recently recognized group of patients that often develops disseminated infections with rapidly growing NTM (predominantly *M. abscessus*) as well as other opportunistic infections has high-titer neutralizing autoantibodies to IFN-γ. Thus far, this syndrome has been reported most frequently in East Asian female patients.

IV catheters can become infected with NTM, usually as a consequence of contaminated water. *M. abscessus* and *M. fortuitum* sometimes infect deep indwelling lines as well as fluids used in eye surgery, subcutaneous injections, and local anesthetics. Infected catheters should be removed.

PULMONARY DISEASE

Lung disease is by far the most common form of nontuberculous mycobacterial infection in North America and the rest of the industrialized world. The clinical presentation typically consists of months or years of throat clearing, nagging cough, and slowly progressive fatigue. Patients will often have seen physicians multiple times and received symptom-based or transient therapy before the diagnosis is entertained and samples are sent for mycobacterial stains and cultures. Because not all patients can produce sputum, bronchoscopy may be required for diagnosis. The typical lag between onset of symptoms and diagnosis is ~5 years in older women. Predisposing factors include underlying lung diseases such as bronchiectasis (Chap. 258), pneumoconiosis (Chap. 256), chronic obstructive pulmonary disease (Chap. 260), primary ciliary dyskinesia (Chap. 258), alpha-1 antitrypsin deficiency (Chap. 309), and cystic fibrosis (Chap. 259). Bronchiectasis and nontuberculous mycobacterial infection often coexist and progress in tandem. This situation makes causality difficult to determine in a given index case, but bronchiectasis is certainly among the most critical predisposing factors that are exacerbated by infection.

MAC organisms are the most common causes of pulmonary nontuberculous mycobacterial infection in North America, but rates vary somewhat by region. MAC infection most commonly develops during the sixth or seventh decade of life in women who have had months or years of nagging intermittent cough and fatigue, with or without sputum production or chest pain. The constellation of pulmonary disease due to NTM in a tall and thin woman who may have chest wall abnormalities is often referred to as Lady Windermere's syndrome, after an Oscar Wilde character of the same name. In fact, pulmonary MAC infection does afflict older nonsmoking white women more than men, with onset at ~60 years. Patients tend to be taller and thinner than the general population, with high rates of scoliosis, mitral valve prolapse, and pectus anomalies. Whereas male smokers with upper-lobe cavitary disease tend to carry the same single strain of MAC indefinitely, nonsmoking females with nodular bronchiectasis tend to carry several strains of MAC simultaneously, with changes over the course of their disease.

M. kansasii can cause a clinical syndrome that strongly resembles tuberculosis, consisting of hemoptysis, chest pain, and cavitary lung disease. The rapidly growing NTM, such as *M. abscessus*, have been associated with esophageal motility disorders such as achalasia. Patients with pulmonary alveolar proteinosis are prone to pulmonary nontuberculous mycobacterial and *Nocardia* infections; the underlying mechanism may be inhibition of alveolar macrophage function due to the autoantibodies to granulocyte-macrophage colony-stimulating factor found in these patients.

CERVICAL LYMPH NODES

The most common form of nontuberculous mycobacterial infection among young children in North America is isolated cervical lymphadenopathy, most frequently caused by MAC organisms but also by other NTM. The cervical swelling is typically firm and relatively painless, with a paucity of systemic signs. Since the differential diagnosis of painless adenopathy includes malignancy, many children have infection with NTM diagnosed inadvertently at biopsy; cultures and special stains may not have

been requested because mycobacterial disease was not ranked high in the differential. Local fistulae usually resolve completely with resection and/or antibiotic therapy. Likewise, the entity of isolated pediatric intrathoracic nontuberculous mycobacterial infection, which is probably related to cervical lymph node infection, is usually mistaken for cancer. In neither isolated cervical nor isolated intrathoracic infections with NTM have children with underlying immune defects been identified, nor do the affected children go on to develop other opportunistic infections.

SKIN AND SOFT TISSUE DISEASE

Cutaneous involvement with NTM usually requires a break in the skin for introduction of the bacteria. Pedicure bath–associated infection with *M. fortuitum* is more likely if skin abrasion (e.g., during leg shaving) has occurred just before the pedicure. Outbreaks of skin infection are often caused by rapidly growing NTM (especially *M. abscessus*, *M. fortuitum*, and *M. chelonae*) acquired via skin contamination from surgical instruments (especially in cosmetic surgery), injections, and other procedures. These infections are typically accompanied by painful, erythematous, draining subcutaneous nodules, usually without associated fever or systemic symptoms.

M. marinum lives in many water sources and can be acquired from fish tanks, swimming pools, barnacles, and fish scales. This organism typically causes papules or ulcers ("fish-tank granuloma"), but the infection can progress to tendonitis with significant impairment of manual dexterity. Lesions appear days to weeks after inoculation of organisms by a typically minor trauma (e.g., incurred during the cleaning of boats or the handling of fish). Tender nodules due to *M. marinum* can advance up the arm in a pattern also seen with *Sporothrix schenckii* (*sporotricoid spread*). The typical carpal tendon involvement may be the first presenting manifestation and may lead to surgical exploration or steroid injection. The index of suspicion must be high for *M. marinum* infections to ensure that proper specimens obtained during procedures are sent for culture.

M. ulcerans, another waterborne skin pathogen, is found mainly in the tropics, especially in tropical areas of Africa. Infection follows skin trauma or insect bites that allow admission to contaminated water. The skin lesions are typically painless, clean ulcers that slough and can cause osteomyelitis. The toxin mycolactone accounts for the modest host inflammatory response and the painless ulcerations.

Diagnosis

NTM can be detected on acid-fast or fluorochrome smears of sputum or other body fluids. When the organism burden is high, the organisms may appear as gram-positive beaded rods, but this finding is unreliable. (In contrast, nocardiae may appear as gram-positive and beaded but filamentous bacteria.) Again, the requisite and most sensitive step in the diagnosis of any mycobacterial disease is to think of including it in the differential. In almost all laboratories, mycobacterial sample processing, staining, and culture are conducted separately from routine bacteriologic tests; thus many infections go undiagnosed because of the physician's failure to request the appropriate test. In addition, mycobacteria usually require separate blood culture media. NTM are broadly differentiated into rapidly growing (<7 days) and slowly growing (≥7 days) forms. Because *M. tuberculosis* typically takes ≥2 weeks to grow, many laboratories refuse to consider culture results final until 6 weeks have elapsed. Newer techniques using liquid culture media permit more rapid isolation of mycobacteria from specimens than is possible with traditional media. Species more readily detected with incubation at 30°C include *M. marinum*, *M. haemophilum*, and *M. ulcerans*. *M. haemophilum* prefers iron supplementation or blood, while *M. genavense* requires supplemented medium with the additive mycobactin J. Bacterial formation of pigment in light conditions (photochromogenicity) or dark conditions (scotochromogenicity) or a lack of bacterial pigment formation (nonchromogenicity) has been used to help categorize NTM. In contrast to NTM, *M. tuberculosis* is beige, rough, dry, and flat. Current identification schemes can reliably use biochemical, nucleic acid, or cell wall composition, as assessed by high-performance liquid chromatography or mass spectrometry, for speciation. With the remarkable decline in U.S. cases of tuberculosis over recent decades, NTM have become the mycobacteria most commonly isolated from humans in North America. However, not all isolations of NTM, especially from the lung, reflect pathology and require treatment. Whereas identification of an organism in a blood or organ biopsy specimen in a compatible clinical setting is diagnostic, the American Thoracic Society recommends that pulmonary infection due to NTM be diagnosed only when disease is clearly demonstrable—i.e., in an appropriate clinical and radiographic setting (nodules, bronchiectasis, cavities) and with repeated isolation of NTM from expectorated sputum or recovery of NTM from bronchoscopy or biopsy specimens. Given the large number of species of NTM and the importance of accurate diagnosis for the implementation of proper therapy, identification of these organisms is ideally taken to the species level.

The purified protein derivative (PPD) of tuberculin is delivered intradermally to evoke a memory T cell response to mycobacterial antigens. This test is variously referred to as the PPD test, the tuberculin skin test, and the Mantoux test, among other designations. Unfortunately, the cutaneous immune response to these tuberculosis-derived filtrate proteins does not differentiate well between infection with NTM and that with *M. tuberculosis*. Since intermediate reactions (~10 mm) to PPD in latent

tuberculosis and nontuberculous mycobacterial infections can overlap significantly, the progressive decline in active tuberculosis in the United States means that NTM probably account for increasing proportions of PPD reactivity. In addition, bacille Calmette-Guérin (BCG) can cause some degree of cross-reactivity, posing problems of interpretation for patients who have received BCG vaccine. Assays to measure the elaboration of IFN-γ in response to the relatively tuberculosis-specific proteins ESAT6 and CFP10 form the basis for IFN-γ-release assays (IGRAs). These assays can be performed with whole blood or on membranes. It is important to note that *M. marinum*, *M. kansasii*, and *M. szulgai* also have ESAT6 and CFP10 and may cause false-positive reactions in IGRAs. Despite cross-reactivity with NTM, large PPD reactions (>15 mm) most commonly signify tuberculosis.

Isolation of NTM from blood specimens is clear evidence of disease. Whereas rapidly growing mycobacteria may proliferate in routine blood culture media, slow-growing NTM typically do not; thus it is imperative to suspect the diagnosis and to use the correct bottles for cultures. Isolation of NTM from a biopsy specimen constitutes strong evidence for infection, but cases of laboratory contamination do occur. Identification of organisms on stained sections of biopsy material confirms the authenticity of the culture. Certain NTM require lower incubation temperatures (*M. genavense*) or special additives (*M. haemophilum*) for growth. Some NTM (e.g., *M. tilburgii*) remain noncultivable but can be identified molecularly in clinical samples.

The radiographic appearance of nontuberculous mycobacterial disease in the lung depends on the underlying disease, the severity of the infection, and the imaging modality used. The advent and increase in the use of CT has allowed the identification of characteristic changes that are highly consistent with nontuberculous mycobacterial infection, such as the "tree-in-bud" pattern of bronchiolar inflammation **(Fig. 167-2)**. Involvement of the lingual and right-middle lobes is commonly seen on chest CT but is difficult to appreciate on plain film. Severe bronchiectasis and cavity formation are common in more advanced disease. Isolation of NTM from respiratory samples can be confusing. *M. gordonae* is often recovered from respiratory samples but is not usually seen on smear and is almost never a pathogen. Patients with bronchiectasis occasionally have NTM recovered from sputum culture with a negative smear. The American Thoracic Society has developed guidelines for the diagnosis of infection with MAC, *M. abscessus*, and *M. kansasii*. A positive diagnosis requires the growth of NTM from two of three sputum samples, regardless of smear findings; a positive bronchoscopic alveolar sample, regardless of smear findings; or a pulmonary parenchymabiopsy sample with granulomatous inflammation or mycobacteria found on section and NTM on culture. These guidelines probably apply to other NTM as well.

Figure 167-2

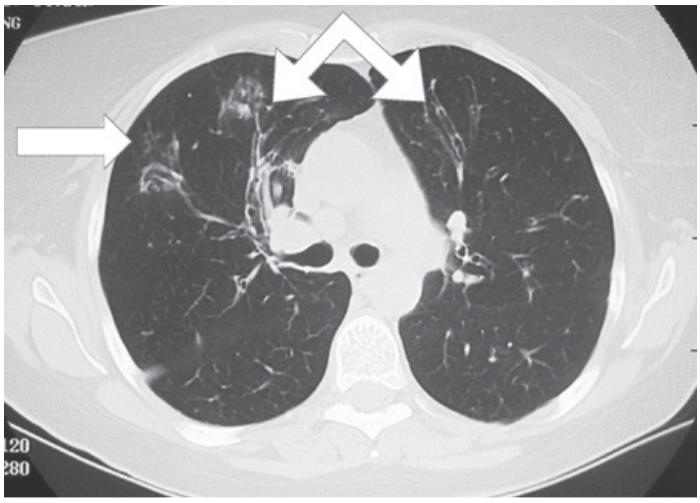

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest CT of a patient with pulmonary MAC infection. Arrows indicate the "tree-in-bud" pattern of bronchiolar inflammation (peripheral right lung) and bronchiectasis (central right and left lungs).

While many laboratories use DNA probes to identify *M. tuberculosis*, MAC, *M. gordonae*, and *M. kansasii*, speciation of NTM helps determine the antimycobacterial therapy to be used. Only testing of MAC organisms for susceptibility to clarithromycin and of *M. kansasii* for susceptibility to rifampin is indicated; few data support other in vitro susceptibility tests, attractive though they appear. MAC isolates that have not been exposed to macrolides are almost always susceptible. NTM that have persisted beyond a course of antimicrobial therapy are often tested for antibiotic susceptibility, but the value and meaning of these tests are undetermined.

Prevention

Prophylaxis of MAC disease in patients infected with HIV is started when the CD4+ T lymphocyte count falls to <50/μL. Azithromycin (1200 mg weekly), clarithromycin (1000 mg daily), or rifabutin (300 mg daily) is effective. Macrolide prophylaxis in immunodeficient patients who are susceptible to NTM (e.g., those with defects in the IFN-γ/IL-12 axis) has not been prospectively validated but seems prudent.

Treatment: Nontuberculous Mycobacteria

NTM cause chronic infections that evolve relatively slowly over a period of weeks to years. Therefore, it is rarely necessary to initiate treatment on an emergent basis before the diagnosis is clear and the infecting species is known. Treatment of NTM is complex, often poorly tolerated, and potentially toxic. Just as in tuberculosis, inadequate single-drug therapy is almost always associated with the emergence of antimicrobial resistance and relapse.

MAC infection often requires multidrug therapy, the foundation of which is a macrolide (clarithromycin or azithromycin), ethambutol, and a rifamycin (rifampin or rifabutin). For disseminated nontuberculous mycobacterial disease in HIV-infected patients, the use of rifamycins poses special problems—i.e., rifamycin interactions with protease inhibitors. For pulmonary MAC disease, thrice-weekly administration of a macrolide, a rifamycin, and ethambutol has been successful. Therapy is prolonged, generally continuing for 12 months after culture conversion; typically, a course lasts for at least 18 months. Other drugs with activity against MAC organisms include IV and aerosolized aminoglycosides, fluoroquinolones, and clofazimine. In elderly patients, rifabutin can exert significant toxicity. However, with only modest efforts, most antimycobacterial regimens are well tolerated by most patients. Resection of cavitary lesions or severely bronchiectatic segments has been advocated for some patients, especially those with macrolide-resistant infections. The success of therapy for pulmonary MAC infections depends on whether disease is nodular or cavitary and on whether it is early or advanced, ranging from 20% to 80%.

M. kansasii lung disease is similar to tuberculosis in many ways and is also effectively treated with isoniazid (300 mg/d), rifampin (600 mg/d), and ethambutol (15 mg/kg per day). Other drugs with very high-level activity against *M. kansasii* include clarithromycin, fluoroquinolones, and aminoglycosides. Treatment should continue until cultures have been negative for at least 1 year. In most instances, *M. kansasii* infection is easily cured.

Rapidly growing mycobacteria pose special therapeutic problems. Extrapulmonary disease in an immunocompetent host is usually due to inoculation (e.g., via surgery, injections, or trauma) or to line infection and is often treated successfully with a macrolide and another drug (with the choice based on in vitro susceptibility), along with removal of the offending focus. In contrast, pulmonary disease, especially that caused by *M. abscessus*, is extremely difficult to cure. Repeated courses of treatment are usually effective in reducing the infectious burden and symptoms. Therapy generally includes a macrolide along with an IV-administered agent such as amikacin, a carbapenem, cefoxitin, or tigecycline. Other oral agents (used according to in vitro susceptibility testing and tolerance) include fluoroquinolones, doxycycline, and linezolid. Because nontuberculous mycobacterial infections are chronic, care must be taken in the long-term use of drugs with neurotoxicities, such as linezolid and ethambutol. Prophylactic pyridoxine has been suggested in these cases. Durations of therapy for *M. abscessus* lung disease are difficult to predict since so many cases are chronic and require intermittent therapy. Expert consultation and management are strongly recommended.

Once recognized, *M. marinum* infection is highly responsive to antimicrobial therapy and is cured relatively easily with any combination of a macrolide, ethambutol, and a rifamycin. Therapy should be continued for 1–2 months after clinical resolution of isolated soft tissue disease; tendon and bone involvement may require longer courses in light of clinical evolution. Other drugs with activity against *M. marinum* include sulfonamides, trimethoprim-sulfamethoxazole, doxycycline, and minocycline.

Treatment of the other NTM is less well defined, but macrolides and aminoglycosides are usually effective, with other agents added as indicated. Expert consultation is strongly encouraged for difficult or unusual infections due to NTM.

Prognosis

The outcomes of nontuberculous mycobacterial infections are closely tied to the underlying condition (e.g., IFN-γ/IL-12 pathway defect, cystic fibrosis) and can range from recovery to death. With no or inadequate treatment, symptoms and signs can be debilitating, including persistent cough, fever, anorexia, and severe lung destruction. With treatment, patients typically regain strength and energy. The optimal duration of therapy when NTM persist in sputum is unknown, but treatment in this situation can be prolonged.

FURTHER READINGS

Browne SK, Holland SM: Anticytokine autoantibodies in infectious diseases: Pathogenesis and mechanisms. Lancet Infect Dis 10:875, 2010[PMID: 21109174]

Daley CL, Griffith DE: Pulmonary non-tuberculous mycobacterial infections. Int J Tuberc Lung Dis 14:665, 2010[PMID: 20487602]

Falkinham JO III: Surrounded by mycobacteria: Nontuberculous mycobacteria in the human environment. J Appl Microbiol 107:356, 2009[PMID: 19228258]

Griffith DE et al: American Thoracic Society: Diagnosis, treatment and prevention of nontuberculous mycobacterial diseases. Am J Respir Crit Care Med 175:367, 2007[PMID: 17277290]

Kim RD et al: Pulmonary nontuberculous mycobacterial disease: Prospective study of a distinct preexisting syndrome. Am J Respir Crit Care Med 178:1066, 2008[PMID: 18703788]

Prevots DR et al: Nontuberculous mycobacterial lung disease prevalence at four integrated health care delivery systems. Am J Respir Crit Care Med 182:970, 2010[PMID: 20538958]

Sologuren I et al: Partial recessive IFN-γR1 deficiency: Genetic, immunological and clinical features of 14 patients from 11 kindreds. Hum Mol Genet 20:1509, 2011[PMID: 21266457]

Changing epidemiology of pulmonary nontuberculous mycobacteria infections. Emerg Infect Dis 16:1576, 2010

van de Vosse E et al: Genetic deficiencies of innate immune signalling in human infectious disease. Lancet Infect Dis 9:688, 2009

Wallis RS: Infectious complications of tumor necrosis factor blockade. Curr Opin Infect Dis 22:403, 2009[PMID: 19491672]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ANTIMYCOBACTERIAL AGENTS: INTRODUCTION

Agents used for the treatment of mycobacterial infections, including tuberculosis (TB), leprosy (Hansen's disease), and infections due to nontuberculous mycobacteria (NTM), are administered in multiple-drug regimens for prolonged courses. Currently, more than 150 species of mycobacteria have been identified, the majority of which do not cause disease in humans. While the incidence of disease caused by *M. tuberculosis* has been declining in the United States, TB remains a leading cause of morbidity and mortality in developing countries—particularly in sub-Saharan Africa, where the HIVepidemic rages. Not only effective drug regimens are needed; without a well-organized infrastructure for diagnosis and treatment of TB, therapeutic and control efforts are severely hampered. Infections with NTM have gained in clinical prominence in the United States and other developed countries. These largely environmental organisms often establish infection in immunocompromised patients or in persons with structural lung disease.

TUBERCULOSIS

General Principles

The earliest recorded human case of TB dates back 9000 years. Early treatment modalities, such as bloodletting, were replaced by sanatorium regimens in the late 19th century. The discovery of streptomycin in 1943 launched the era of antibiotic treatment for TB. Over subsequent decades, the discovery of additional agents and the use of multiple-drug regimens allowed progressive shortening of the treatment course from years to as little as 6 months with the regimen for drug-susceptible TB. Latent TB infection (LTBI) and active TB disease are diagnosed by history, physical examination, tuberculin skin test, interferon γ release assay, radiographic imaging, and/or mycobacterial cultures. LTBI is treated with either isoniazid (9 months) or rifampin (4 months) **(Table 168-1)**.

Table 168-1 Regimens for the Treatment of Latent Tuberculosis Infection (LTBI) in Adults

Regimen	Schedule	Duration	Comments
Isoniazid	300 mg daily (5 mg/kg)	9 months	Supplement with pyridoxine (25–50 mg daily).
	Alternative: 900 mg twice weekly (15 mg/kg)		Twice-weekly regimens require directly observed therapy.
Rifampin	600 mg daily (10 mg/kg)	4 months	Broader efficacy studies are needed.
Isoniazid plus rifapentine[a]	900 mg weekly + 900 mg weekly (15 mg/kg)	4 months	Weekly regimens require directly observed therapy. Supplement with pyridoxine (25–50 mg daily).

[a]Under investigation.

For active or suspected TB disease, clinical factors, including HIV co-infection, symptom duration, radiographic appearance, and public health concerns about TB transmission, drive diagnostic testing and treatment initiation. Multiple-drug regimens are used for the treatment of TB disease **(Table 168-2)**. Initially, an intensive phase consisting of four drugs—isoniazid, rifampin, pyrazinamide, and ethambutol given for 2 months—is followed by a continuation phase of isoniazid and rifampin for 4 months, for a total treatment duration of 6 months. The continuation phase is extended to 7 months (for a total treatment duration of 9 months) if the 2-month course of pyrazinamide is not completed or, for patients with cavitary pulmonary TB, if sputum cultures remain positive beyond 2 months of treatment (delayed culture conversion).

Table 168-2 Simplified Approach to Treatment of Active Tuberculosis in Adults

Culture Results	Intensive Phase	Continuation Phase	Extension of Total Treatment
Culture	HRZE for 2 months,	HR for 4 months,	To 9 months, if 2 months of Z is not

positive	daily or intermittent (with dose adjustment)	daily or 5 d/wk *or* HR for 4 months, intermittent (with dose adjustment)	completed or culture conversion is prolonged[a] and/or cavitation is documented
Culture negative	HRZE for 2 months	2 months	To 6 months, if patient is infected with HIV
Resistant to H	RZE or S (Q[b]) for 6 months	. . .	Prolonged culture conversion, cavitation
Resistant to R	HZEQ[b] (IA[c]) for 2 months	HEQ(S) for 10–16 months	Prolonged culture conversion, delayed response
Resistant to HR[d]	ZEQ[b](IA[c]) ± alternative agents[e] for 18–24 months	. . .	Prolonged culture conversion

[a] Beyond 2 months.

[b] Moxifloxacin and levofloxacin are the preferred fluoroquinolones; ciprofloxacin should be avoided.

[c] Injectable agents: streptomycin, amikacin, kanamycin, and capreomycin.

[d] Management of multidrug-resistant TB should be performed by or in close consultation with an expert TB clinician. Surgical management should be considered.

[e] Alternative agents: cycloserine, ethionamide, para-aminosalicylic acid, clarithromycin, linezolid, and amoxicillin-clavulanate.

Abbreviations: E, ethambutol; H, isoniazid; IA, injectable agent; Q, fluoroquinolone; R, rifampin; S, streptomycin; Z, pyrazinamide.

Treatment of TB in individuals co-infected with HIV poses significant challenges, but some progress is being made. Recent data show improved survival when antiretroviral therapy (ART) is initiated early in TB therapy. Interactions of rifampin with protease inhibitors or nonnucleotide reverse transcriptase inhibitors are significant and require close monitoring and dose adjustments. The TB immune reconstitution inflammatory syndrome (IRIS) may appear as early as 1 week after initiation of ART and manifests as paradoxical worsening or unmasking of existing TB infection. Conservative management consists of continued administration of ART and TB medications; however, severe or debilitating IRIS has been anecdotally treated with varying doses of glucocorticoids. Intermittent therapy in patients co-infected with HIV and *M. tuberculosis* has been associated with low plasma levels of several key TB drugs and with higher rates of treatment failure or relapse; therefore, intermittent twice-weekly therapy for TB in HIV-co-infected individuals is not recommended.

Adherence to medications is critical in achieving a cure with antimycobacterial therapy. Consequently, directly observed therapy (DOT) by trained staff, either in the clinic or at home, is recommended to ensure adherence. In addition, monthly dispensing of TB medicines is recommended, since monthly clinical monitoring for hepatotoxicity due to these medications is essential for all patients. Discontinuation of suspected offending agents at the onset of hepatitis symptoms reduces the risk of progression to fatal hepatitis. Clinical monitoring includes at least monthly assessment for symptoms and signs (nausea, vomiting, abdominal discomfort, and unexplained fatigue) and signs (jaundice, dark urine, light stools, diffuse pruritus) of hepatotoxicity, although the latter represent comparatively late manifestations (Table 168-3). The presence of such symptoms and signs mandates provisional discontinuation of potentially hepatotoxic agents. Biochemical testing of at least serum alanine aminotransferase and total bilirubin levels and exclusion of other causes of these abnormalities are also indicated. For patients with active TB, monthly mycobacterial cultures of sputum are recommended until it is certain that the organisms have been cleared and the patient has responded to therapy or until no sputum is available for culture.

Table 168-3 Monitoring and Clinical Management of Tuberculosis Treatment in Adults

Drug	Assessment	Management
LTBI Treatment		
With hepatic risk factors[b], check ALT and bilirubin at baseline. If ALT is ≥3 x ULN or total bilirubin is >2, defer treatment and reevaluate.		
Isoniazid	Determine whether hepatic risk factors are present. If so,	If ALT is 5 x ULN (or 3 x ULN with symptoms)[c] or if bilirubin reaches jaundice levels (usually >2 x ULN),

	obtain baseline and periodic ALT and bilirubin values.	interrupt treatment. With normalization, consider an alternative agent.
Rifampin	Same	Same
TB Treatment		
Check ALT, bilirubin, platelets, creatinine, and hepatitis panel on all patients at baseline. If hepatic risk factors are present, check ALT and bilirubin monthly.		
Isoniazid	If ALT is >5 x ULN (or >3 x ULN with hepatitis symptoms)[c]	Obtain history of alcohol consumption and concomitant drugs. In most instances, discontinue H, Z, R, and other hepatotoxic drugs. Consider alternative agents. Obtain viral hepatitis serologies. Rechallenge: With normalization of liver enzymes, sequentially reintroduce R and then H. With no recurrence of hepatotoxicity, do not resume Z. Alternative rechallenge protocols have been proposed.
Rifampin	If primary elevation is in bilirubin and alkaline phosphatase, more likely rifampin	Discontinue R if bilirubin reaches jaundice levels (usually >2 x ULN). May try to reintroduce. If not tolerated, may substitute Q.
Ethambutol	Decrease in visual acuity or color vision or appearance on monthly screening	Discontinue ethambutol and repeat ocular exam. Peripheral neuropathy may be a precursor of ocular toxicity; if it occurs, consider repeat ocular exam.
Pyrazinamide	If ALT is >5 x ULN (or >3 x ULN with symptoms)[c]	Same as for H
Fluoroquinolone	If QT_C prolongation is discovered incidentally on ECG	Check audiometry and at least BUN and creatinine monthly.
Aminoglycoside	Abnormal results on audiometry testing, BUN, creatinine, electrolytes at baseline or on monthly check	Discontinue aminoglycoside if not MDR-TB. As appropriate, assess renal function, correct electrolytes, or seek ENT consultation.

[a] All regimens require monthly clinical monitoring.

[b] Hepatic risk factors: chronic alcohol use, viral hepatitis, preexisting liver disease, pregnancy or 3 months postpartum, hepatotoxic medications.

[c] Relevant manifestations include nausea, vomiting, abdominal pain, jaundice, or unexplained fatigue.

Abbreviations: ALT, alanine aminotransferase; BUN, blood urea nitrogen; H, isoniazid; Q, fluoroquinolone; R, rifampin; Z, pyrazinamide; Q, fluoroquinolone; ULN, upper limit of normal; MDR-TB, multidrug-resistant tuberculosis; ENT, ear, nose, and throat.

If significant clinical improvement does not occur or the patient's condition deteriorates over the course of therapy, possibilities include treatment failure due to nonadherence to therapy, poor medication absorption, or the development of resistance. For patients co-infected with HIV and M. tuberculosis, immune reconstitution inflammatory syndrome (IRIS) is a possibility and is a diagnosis of exclusion. or mycobacterial resistance should be suspected. Drug susceptibility testing should be repeated at this point. If resistance is documented or strongly suspected, at least two efficacious drugs to which the isolate is susceptible or which the patient has not already taken should be added to the therapeutic regimen.

Multidrug-resistant tuberculosis (MDR-TB) is defined as disease caused by a strain of *M. tuberculosis* that is resistant to both isoniazid and rifampin—the most efficacious of the first-line TB drugs. The risk of MDR-TB is elevated in patients presenting from geographic areas in which ≥5% of incident TB is MDR-TB and in patients previously treated for TB. Treatment regimens for MDR-TB generally include a late-generation fluoroquinolone and an injectable second-line agent (such as capreomycin, amikacin, or kanamycin). Regimens of at least five drugs are recommended for the treatment of MDR-TB. Both standardized and optimized/customized regimens are in use around the world. Extensively drug-resistant tuberculosis (XDR-TB) is defined as MDR-TB with additional resistance to any fluoroquinolone and at least one of the second-line injectable agents. Treatment of XDR-TB is individualized on the basis of extended antimicrobial susceptibility testing. Therapeutic regimens for either MDR-TB or XDR-TB should be constructed with input from clinicians experienced in the management of TB.

First-Line Antituberculosis Drugs

ISONIAZID

Isoniazid is a critical drug for treatment of both TB disease and LTBI. Isoniazid has excellent bactericidal activity against both intracellular and extracellular, actively dividing *M. tuberculosis*. This drug is bacteriostatic against slowly dividing organisms. In treatment of LTBI, isoniazid is considered the first-line agent because it is generally well tolerated, has well-established efficacy, and is inexpensive. In this setting, the drug is taken daily or intermittently (i.e., twice weekly) as DOT for 9 months. The 9-month course is more efficacious than the 6-month course (75–90% vs. ≳65%), but extension of treatment to 12 months is not likely to provide further protection. A 6-month course of daily or intermittent isoniazid is considered second-line, but acceptable, therapy.

For treatment of TB, isoniazid is used in combination with other agents to ensure killing of both actively dividing *M. tuberculosis* and slowly growing "persister" organisms. Unless the organism is resistant, the standard regimen includes isoniazid, rifampin, ethambutol, and pyrazinamide (Table 168-2). Isoniazid is often given together with 25–50 mg of pyridoxine daily to prevent drug-related peripheral neuropathy.

Mechanism of Action

Isoniazid is a prodrug activated by the mycobacterial KatG catalase/peroxidase; isoniazid is coupled with reduced nicotinamide adenine dinucleotide (NADH). The resulting isonicotinic acyl-NADH complex blocks the mycobacterial ketoenoylreductase known as InhA, binding to its substrate and inhibiting fatty acid synthase and ultimately mycolic acid synthesis. Mycolic acids are essential requirements for the mycobacterial cell wall. KatG activation of isoniazid also results in the release of free radicals that have antimycobacterial activity, including nitric oxide.

The minimal inhibitory concentrations (MICs) of isoniazid for wild-type (untreated) susceptible strains are <0.1 μg/mL for *M. tuberculosis* and 0.5–2 μg/mL for *M. kansasii*.

Pharmacology

Isoniazid is the hydrazide of isonicotinic acid, a small, water-soluble molecule. The usual adult oral daily dose of 300 mg results in peak serum levels of 3–5 μg/mL within 30 min to 2 h after ingestion —well in excess of the MICs for most susceptible strains of *M. tuberculosis*. Both oral and IM preparations of isoniazid achieve good levels in the body, although antacids and high-carbohydrate meals may interfere with oral absorption. Isoniazid diffuses well throughout the body, reaching therapeutic concentrations in body cavities and fluids, with concentrations in cerebrospinal fluid (CSF) comparable to those in serum.

Isoniazid is metabolized in the liver via acetylation by *N*-acetyltransferase 2 (NAT2) and hydrolysis. Both fast- and slow-acetylation phenotypes occur; patients who are "fast acetylators" may have lower serum levels of isoniazid, whereas slow acetylators may have higher levels and experience more toxicity. Satisfactory isoniazid levels are attained in the majority of homozygous fast NAT2 acetylators given a dose of 6 mg/kg and in the majority of homozygous slow acetylators given only 3 mg/kg. Genotyping is increasingly being used to characterize isoniazid-related pharmacogenomic responses.

Isoniazid's interactions with other drugs are due primarily to its inhibition of the cytochrome P450 system. Among the drugs with significant isoniazid interactions are warfarin, carbamazepine, benzodiazepines, acetaminophen, clopidogrel, maraviroc, dronedarone, salmeterol, tamoxifen, eplerenone, and phenytoin.

Dosing

The recommended daily dose for the treatment of TB in the United States is 5 mg/kg for adults and 10 –20 mg/kg for children, with a maximal daily dose of 300 mg for both. For intermittent therapy in adults (usually twice per week), the dose is 15 mg/kg with a maximal daily dose of 900 mg. Isoniazid does not require dosage adjustment in patients with renal disease.

Resistance

Although isoniazid is, along with rifampin, the mainstay of TB treatment regimens, ~7% of clinical *M. tuberculosis* isolates in the United States are resistant. Rates of primary isoniazid resistance among untreated patients are significantly higher in many populations born outside the United States. Four separate pathways for isoniazid resistance have been elucidated. Most strains have amino acid changes in either the catalase-peroxidase gene (*katG*) or the mycobacterial ketoenoylreductase gene (*inhA*). Less frequently, alterations in *kasA*, the gene for an enzyme involved in mycolic acid elongation, and loss of NADH dehydrogenase 2 activity confer isoniazid resistance.

Adverse Effects

Although isoniazid is generally well tolerated, drug-induced liver injury and peripheral neuropathy are significant adverse effects associated with this agent. Isoniazid may cause asymptomatic transient elevation of aminotransferase levels (often termed *hepatic adaptation*) in up to 20% of recipients. Other adverse reactions include rash (2%), fever (1.2%), anemia, acne, arthritic symptoms, a systemic

lupus erythematosus–like syndrome, optic atrophy, seizures, and psychiatric symptoms. Symptomatic hepatitis occurs in fewer than 0.1% of persons treated with isoniazid alone for LTBI, and fulminant hepatitis with hepatic failure occurs in fewer than 0.01%. Isoniazid-associated hepatitis is idiosyncratic, but its incidence increases with age, with daily alcohol consumption, and in women who are within 3 months postpartum.

In patients who have liver disorders or HIV infection, who are pregnant or in the 3-month postpartum period, who have a history of liver disease (e.g., hepatitis B or C, alcoholic hepatitis, or cirrhosis), who use alcohol regularly, who have multiple medical problems, or who have other risk factors for chronic liver disease, the risks and benefits of treatment for LTBI should be weighed. If treatment is undertaken, these patients should have serum concentrations of alanine aminotransferase (ALT) determined at baseline. Routine baseline hepatic ALT testing based solely on an age of >35 years is optional and depends on individual concerns. Monthly biochemical monitoring during isoniazid treatment is indicated for patients whose baseline liver function tests yield abnormal results and for persons at risk for hepatic disease, including the groups just mentioned. Guidelines recommend that isoniazid be discontinued in the presence of hepatitis symptoms or jaundice and an ALT level three times the upper limit of normal or in the absence of symptoms with an ALT level five times the upper limit of normal (Table 168-3).

Peripheral neuropathy associated with isoniazid occurs in up to 2% of patients given 5 mg/kg. Isoniazid appears to interfere with pyridoxine (vitamin B6) metabolism. The risk of isoniazid-related neurotoxicity is greatest for patients with preexisting disorders that also pose a risk of neuropathy, such as HIV infection; for those with diabetes mellitus, alcohol abuse, or malnutrition; and for those simultaneously receiving other potentially neuropathic medications, such as stavudine. These patients should be given prophylactic pyridoxine (25–50 mg/d).

RIFAMPIN

Rifampin is a semisynthetic derivative of *Amycolatopsis rifamycinica* (formerly known as *Streptomyces mediterranei*). The most active antimycobacterial agent available, rifampin is the keystone of first-line treatment for TB. Introduced in 1968, this drug eventually permitted dramatic shortening of TB treatment. Rifampin has bactericidal activity against both dividing and nondividing *M. tuberculosis*, with sterilizing activity. The drug is also active against an array of other organisms, including some gram-positive and gram-negative bacteria, *Legionella*, *M. kansasii*, and *M. marinum*.

Rifampin, administered for 4 months, is also an alternative agent to isoniazid for the treatment of LTBI, although efficacy data are scant at this time. A 3-month course of rifampin alone has been found to be similar in efficacy to a 6-month course of isoniazid. Although the 4-month regimen of rifampin has not yet been compared with 9 months of isoniazid, randomized safety and tolerability studies indicate that rates of discontinuation, hepatotoxicity, and adverse reaction are lower with the former than with the latter. The rates of treatment completion are also higher for the 4-month rifampin regimen.

Mechanism of Action

Rifampin exerts both intracellular and extracellular bactericidal activity. Like other rifamycins, rifampin specifically binds to and inhibits mycobacterial DNA-dependent RNA polymerase, blocking RNA synthesis. Susceptible strains of *M. tuberculosis* as well as *M. kansasii* and *M. marinum* are inhibited by rifampin concentrations of 1 μg/mL.

Pharmacology

Rifampin is a fat-soluble, complex macrocyclic molecule readily absorbed after oral administration. Serum levels of 10–20 μg/mL are achieved 2.5 h after the usual adult oral dose of 10 mg/kg (given without food). Rifampin has a half-life of 1.5–5 h. The drug distributes well throughout most body tissues, including CSF. Rifampin turns body fluids such as urine, saliva, sputum, and tears a reddish-orange color—an effect that offers a simple means of assessing patients' adherence to this medication. Rifampin is excreted primarily through the bile and enters the enterohepatic circulation; <30% of a dose is renally excreted.

As a potent inducer of the hepatic cytochrome P450 system, rifampin can decrease the half-life of some drugs, such as digoxin, warfarin, phenytoin, prednisone, cyclosporine, methadone, oral contraceptives, clarithromycin, azole antifungal agents, quinidine, and antiretroviral protease inhibitors and nonnucleoside reverse transcriptase inhibitors. The Centers for Disease Control and Prevention has issued guidelines for the management of drug interactions during treatment of HIV and *M. tuberculosis* co-infection (www.cdc.gov/tb/).

Dosing

The daily dosage of rifampin is 10 mg/kg for adults and 10–20 mg/kg for children, with a maximum of 600 mg/d for both. The drug is given once daily, twice weekly, or three times weekly. No adjustments of dose or frequency are necessary in patients with renal insufficiency.

Resistance

Resistance to rifampin in *M. tuberculosis*, *M. leprae*, and other organisms is the consequence of

spontaneous, mostly missense point mutations in a core region of the bacterial gene coding for the β subunit of RNA polymerase (*rpoB*). RNA polymerase altered in this manner is no longer subject to inhibition by rifampin. Most rapidly and slowly growing NTM harbor intrinsic resistance to rifampin, for which the mechanism has yet to be determined.

Adverse Effects

Adverse events associated with rifampin are infrequent and generally mild. Hepatotoxicity due to rifampin alone is uncommon in the absence of preexisting liver disease and often consists of isolated hyperbilirubinemia rather than aminotransferase elevation. Other adverse reactions include rash, pruritus, gastrointestinal symptoms, and pancytopenia. Rarely, a hypersensitivity reaction may occur with intermittent therapy, manifesting as fever, chills, malaise, rash, and—in some instances—renal and hepatic failure.

ETHAMBUTOL

Ethambutol is a bacteriostatic antimycobacterial agent first synthesized in 1961. A component of the standard first-line regimen, ethambutol provides synergy with the other drugs in the regimen and is generally well tolerated. Susceptible species include *M. tuberculosis*, *M. marinum*, *M. kansasii*, and organisms of the *M. avium* complex (MAC); however, among first-line drugs, ethambutol is the least potent against *M. tuberculosis*. This agent is also used in combination with other agents in the continuation phase of treatment in patients who cannot tolerate isoniazid or rifampin or who are infected with organisms resistant to either of the latter drugs.

Mechanism of Action

Ethambutol is bacteriostatic against *M. tuberculosis*. Its primary mechanism of action is the inhibition of the arabinosyltransferases involved in cell wall synthesis, which probably inhibits the formation of arabinogalactan and lipoarabinomannan.

Pharmacology and Dosing

From a single dose of ethambutol, 75–80% is absorbed within 2–4 h of administration. Serum levels peak at 2–4 μg/mL after the standard adult daily dose of 15 mg/kg. Ethambutol is well distributed throughout the body except in the CSF; a dosage of 25 mg/kg is necessary for attainment of a CSF level half of that in serum. For intermittent therapy, the dosage is 50 mg/kg twice weekly. To prevent toxicity, the dosage must be lowered and the frequency of administration reduced for patients with renal insufficiency.

Adverse Effects

Ethambutol is usually well tolerated and has no significant interactions with other drugs. Optic neuritis, the most serious adverse effect reported, typically presents as reduced visual acuity, central scotoma, and loss of the ability to see green (or, less commonly, red). The cause of this neuritis is unknown, but it may be due to an effect of ethambutol on the amacrine and bipolar cells of the retina. Symptoms typically develop several months after initiation of therapy, but ocular toxicity soon after initiation of ethambutol has been described. The risk of ocular toxicity is dose dependent, with occurrence in 1–5% of patients, and can be increased by renal insufficiency. The routine use of ethambutol in younger children is not recommended, as monitoring for visual complications can be difficult. If drug-resistant TB is suspected, ethambutol can be used in children.

All patients starting therapy with ethambutol should have a baseline test for visual acuity, visual fields, and color vision and should undergo an examination of the optic fundus. Visual acuity and color vision should be monitored monthly or less often as needed. Cessation of ethambutol in response to early symptoms of ocular toxicity usually results in reversal of the deficit within several months. Recovery of all visual function may take up to 1 year. In the elderly and in patients in whom symptoms are not recognized early, deficits may be permanent. Some experts think that supplementation with hydroxycobalamin (vitamin B12) is beneficial for patients with ethambutol-related ocular toxicity.

Other adverse effects of ethambutol are rare. Peripheral sensory neuropathy occurs in rare instances.

Resistance

Ethambutol resistance in *M. tuberculosis* and NTM is associated primarily with missense mutations in the *embB* gene that encodes for arabinosyltransferase. Mutations have been found in resistant strains at codon 306 in 50–70% of cases. Mutations at *embB*306 can cause significantly increased MICs of ethambutol, resulting in clinical resistance.

PYRAZINAMIDE

A nicotinamide analog, pyrazinamide is an important bactericidal drug used in the initial phase of TB treatment. Its administration for the first 2 months of therapy with rifampin and isoniazid allows treatment duration to be shortened from 9 months to 6 months and decreases rates of relapse.

Mechanism of Action

Pyrazinamide's antimycobacterial activity is essentially limited to *M. tuberculosis*. The drug is more

active against slowly replicating organisms than against actively replicating organisms. Pyrazinamide is a prodrug that is converted by the mycobacterial pyrimidase to the active form, pyrazinoic acid (POA). This agent is active only in acidic environments (pH <6.0), as are found within phagocytes or granulomas. The exact mechanism of action of POA is unclear, but fatty acid synthetase I may be the primary target in *M. tuberculosis*.

Pharmacology and Dosing

Pyrazinamide is well absorbed after oral administration, with peak serum concentrations of 20–60 μg/mL at 1–2 h after ingestion of the recommended adult daily dose of 15–30 mg/kg (maximum, 2 g/d). It distributes well to various body compartments, including CSF, and is an important component of treatment for tuberculous meningitis. The serum half-life of the drug is 9–11 h with normal renal and hepatic function. Pyrazinamide is metabolized in the liver to POA, 5-hydroxypyrazinamide, and 5-hydroxy-POA. A high proportion of pyrazinamide and its metabolites (~70%) is excreted in the urine. The dosage must be adjusted according to the level of renal function in patients with reduced creatinine clearance.

Adverse Effects

At the higher dosages used previously, hepatotoxicity was seen in as many as 15% of patients treated with pyrazinamide. However, at the currently recommended dosages, hepatotoxicity now occurs less commonly when this drug is administered with isoniazid and rifampin during the treatment of TB. Older age, active liver disease, HIV infection, and low albumin levels may increase the risk of hepatotoxicity. The use of pyrazinamide with rifampin for the treatment of LTBI is no longer recommended because of unacceptable rates of hepatotoxicity and death in this setting. Hyperuricemia is a common adverse effect of pyrazinamide therapy that usually can be managed conservatively. Clinical gout is rare.

Although pyrazinamide is recommended by international tuberculosis organizations for routine use in pregnancy, it is not recommended in the United States because of inadequate teratogenicity data.

Resistance

The basis of pyrazinamide resistance in *M. tuberculosis* is a mutation in the *pncA* gene coding for pyrazinamidase, the enzyme that converts the prodrug to active POA. Resistance to pyrazinamide is associated with loss of pyrazinamidase activity, which prevents conversion of pyrazinamide to POA. Of pyrazinamide-resistant *M. tuberculosis* isolates, 72–98% have mutations in *pncA*. Conventional methods of testing for susceptibility to pyrazinamide may produce both false-negative and false-positive results because the high-acidity environment required for the drug's activation also inhibits the growth of *M. tuberculosis*. There is some controversy as to the clinical significance of in vitro pyrazinamide resistance.

First-Line Supplemental Drugs

RIFABUTIN

Rifabutin, a semisynthetic derivative of rifamycin S, inhibits mycobacterial DNA-dependent RNA polymerase. Although rifabutin is active in vitro against some strains of rifampin-resistant *M. tuberculosis*, its clinical utility in this situation is not clear. Rifabutin is more active than rifampin against MAC organisms and other NTM in vitro.

Rifabutin is recommended in place of rifampin for the treatment of HIV-co-infected individuals who are taking protease inhibitors or nonnucleoside reverse transcriptase inhibitors, particularly nevirapine. Rifabutin's effect on hepatic enzyme induction is less pronounced than that of rifampin. Protease inhibitors may cause significant increases in rifabutin levels through inhibition of hepatic metabolism.

Pharmacology

Like rifampin, rifabutin is lipophilic and is absorbed rapidly after oral administration, reaching peak serum levels 2–4 h after ingestion. Rifabutin distributes best to tissues, reaching levels 5–10 times higher than those in plasma. Unlike rifampin, rifabutin and its metabolites are partially cleared by the hepatic microsomal system. Rifabutin's slow clearance results in a mean serum half-life of 45 h—much longer than the 3- to 5-h half-life of rifampin. Clarithromycin (but not azithromycin) and fluconazole appear to increase rifabutin levels by inhibiting hepatic metabolism.

Adverse Effects

Rifabutin is generally well tolerated, with adverse effects occurring at higher doses. The most common adverse events are gastrointestinal; other reactions include rash, headache, asthenia, chest pain, myalgia, and insomnia. Less common adverse reactions include fever, chills, a flulike syndrome, anterior uveitis, hepatitis, *Clostridium difficile*–associated diarrhea, a diffuse polymyalgia syndrome, and yellow skin discoloration ("pseudo-jaundice"). Laboratory abnormalities include neutropenia, leukopenia, thrombocytopenia, and increased levels of liver enzymes.

Resistance

Resistance to rifabutin is mediated by some mutations in *rpoB*.

RIFAPENTINE

Rifapentine is a semisynthetic cyclopentyl rifamycin, sharing a mechanism of action with rifampin. Rifapentine is lipophilic and has a prolonged half-life that permits weekly or twice-weekly dosing (although dosing and frequency of administration are still being actively studied). Because of higher rates of relapse, rifapentine is not approved for administration to patients with HIV disease. It is being studied for treatment of LTBI in a weekly combination regimen with isoniazid given for 3 months (versus the standard 9-month regimen).

Pharmacology

Rifapentine's good absorption is improved when the drug is taken with food. After oral administration, rifapentine reaches peak serum concentrations in 5–6 h and achieves a steady state in 10 days. The half-life of rifapentine and its active metabolite, 25-desacetyl rifapentine, is ~13 h. The administered dose is excreted via the liver (70%).

Adverse Effects

The adverse-effects profile of rifapentine is similar to that of other rifamycins. Rifapentine is teratogenic in animal models and is relatively contraindicated in pregnancy.

Resistance

Rifapentine resistance is mediated by mutations in *rpoB*. Mutations that cause resistance to rifampin also cause resistance to rifapentine.

STREPTOMYCIN

Streptomycin was the first antimycobacterial agent used for the treatment of TB. Derived from *Streptomyces griseus*, streptomycin is bactericidal against dividing *M. tuberculosis* organisms but has only low-level early bactericidal activity. This drug is administered only by the IM and IV routes. In developed nations, streptomycin is used infrequently because of its toxicity, the inconvenience of injections, and drug resistance. In developing countries, however, streptomycin is used because of its low cost.

Mechanism of Action

Streptomycin inhibits protein synthesis by binding at a site on the 30S mycobacterial ribosome.

Pharmacology and Dosing

Serum levels of streptomycin peak at 25–45 μg/mL after a 1-g dose. This agent penetrates poorly into the CSF, reaching levels that are only 20% of serum levels. The usual daily dose of streptomycin (given IM either daily or 5 days per week) is 15 mg/kg for adults and 20–40 mg/kg for children, with a maximum of 1 g/d for both. For patients ≥60 years of age, 10 mg/kg is the recommended daily dose, with a maximum of 750 mg/d. Because streptomycin is eliminated almost exclusively by the kidneys, its use in patients with renal impairment should be avoided or implemented with caution, with lower doses and less frequent administration.

Adverse Effects

Adverse reactions occur frequently with streptomycin (10–20% of patients). Ototoxicity (primarily vestibulotoxicity), neuropathy, and renal toxicity are the most common and the most serious. Renal toxicity, usually manifested as nonoliguric renal failure, is less common with streptomycin than with other frequently used aminoglycosides, such as gentamicin. Manifestations of vestibular toxicity include loss of balance, vertigo, and tinnitus. Patients receiving streptomycin must be monitored carefully for these adverse effects, undergoing audiometry at baseline and monthly thereafter.

Resistance

Spontaneous mutations conferring resistance to streptomycin are relatively common, occurring in 1 in 10^6 organisms. In the two-thirds of streptomycin-resistant *M. tuberculosis* strains exhibiting high-level resistance, mutations have been identified in one of two genes: a 16S rRNA gene (*rrs*) or the gene encoding ribosomal protein S12 (*rpsL*). Both targets are believed to be involved in streptomycin ribosomal binding. However, low-level resistance, which is seen in about one-third of resistant isolates, has no associated resistance mutation. A gene (*gidB*) that confers low-level resistance to streptomycin has recently been identified. Strains of *M. tuberculosis* resistant to streptomycin generally are not cross-resistant to capreomycin or amikacin.

Second-Line Antituberculosis Drugs

Second-line antituberculosis agents are indicated for treatment of drug-resistant TB, for patients who are intolerant or allergic to first-line agents, and when first-line supplemental agents are unavailable.

FLUOROQUINOLONES

Fluoroquinolones inhibit mycobacterial DNA gyrase and topoisomerase IV, preventing cell replication and protein synthesis, and are bactericidal. The later-generation fluoroquinolones moxifloxacin and levofloxacin are the most active against *M. tuberculosis* and are being investigated for their potential to

shorten the course of treatment for TB. Gatifloxacin, although also being assessed for shortening of treatment duration, causes significant dysglycemia. Ciprofloxacin is no longer recommended for the treatment of TB because of poor efficacy.

The fluoroquinolones are well absorbed orally, achieve high serum levels, and distribute well into body tissues and fluids. Their absorption is decreased by co-ingestion with products containing multivalent cations, such as antacids. Adverse effects are relatively infrequent (0.5–10% of patients) and include gastrointestinal intolerance, rashes, dizziness, and headache. Most studies of fluoroquinolone side effects have been based on relatively short-term administration for bacterial infections, but trials have now shown the relative safety and tolerability of fluoroquinolones administered for months during TB treatment in adults. The potential to prolong the QT_C interval leading to cardiac arrhythmias has been a source of concern with fluoroquinolones. QT_C prolongation requiring cessation of treatment is rare. There is increasing interest in the use of fluoroquinolones in children, which has traditionally been avoided because of the risks of tendon rupture and cartilage damage, as the benefits in treatment of drug-resistant TB may outweigh the risks.

Mycobacterial resistance can develop rapidly when a fluoroquinolone is inadvertently administered alone. Empirical fluoroquinolone therapy for presumed community-acquired pneumonia is associated with increased resistance in *M. tuberculosis*. Mutations in the genes encoding for DNA gyrase (*gyrA* and *gyrB*) are implicated in many but not all cases of clinical resistance to fluoroquinolones.

CAPREOMYCIN

Capreomycin, a cyclic peptide antibiotic derived from *Streptomyces capreolus*, is an important second-line agent used for treatment of MDR-TB, particularly when additional resistance to aminoglycosides is documented. Capreomycin is administered by the IM route; an inhaled preparation is under study. A dose of 15 mg/kg per day is given five to seven times per week (maximal daily dose, 1 g) and results in peak blood levels of 20–40 μg/mL. The dosage may be reduced to 1 g two or three times per week 2–4 months after mycobacterial cultures become negative. For individuals ≥60 years of age, the dose should be reduced to 10 mg/kg per day (maximal daily dose, 750 mg). For patients with renal insufficiency, the drug should be given intermittently and at lower dosage (12–15 mg/kg two or three times per week). A minimal duration of 3 months is recommended for MDR-TB treatment. Penetration of capreomycin into the CSF is believed to be poor.

The mechanism of capreomycin's action is not well understood but involves interference with the mycobacterial ribosome and inhibition of protein synthesis. Resistance to capreomycin is associated with mutations that inactivate a ribosomal methylase (TlyA) or that encode genes for the 16S ribosomal subunit (*rrs*). Cross-resistance to kanamycin and amikacin is common. However, some strains that are resistant to streptomycin, kanamycin, and amikacin generally remain susceptible to capreomycin.

Adverse effects of capreomycin are relatively common. Significant hypokalemia and hypomagnesemia as well as oto- and renal toxicity have been reported.

AMIKACIN AND KANAMYCIN

Amikacin and kanamycin are aminoglycosides that exert mycobactericidal activity by binding to the 16S ribosomal subunit. The spectrum of antibiotic activity for amikacin and kanamycin includes *M. tuberculosis*, several NTM species, and aerobic gram-negative and gram-positive bacteria. Although amikacin is highly active against *M. tuberculosis*, it is used only infrequently because of its significant side effects. The usual daily adult dosage of both amikacin and kanamycin is 15–30 mg/kg given IM or IV (maximal daily dose, 1 g), with a reduction to 10 mg/kg for patients ≥60 years old. For patients with renal insufficiency, the dose and frequency should be reduced (12–15 mg/kg two or three times per week). Mycobacterial resistance is due to mutations in the genes encoding the16S ribosomal RNA gene. Cross-resistance among kanamycin, amikacin, and capreomycin is common. Isolates resistant to streptomycin are frequently susceptible to amikacin or kanamycin. Adverse effects of amikacin include ototoxicity (in up to 10% of recipients, with auditory dysfunction occurring more commonly than vestibulotoxicity), nephrotoxicity, and neurotoxicity. Kanamycin has a similar side-effects profile, but adverse reactions are thought to be less frequent and less severe.

ETHIONAMIDE

Ethionamide is a derivative of isonicotinic acid. Its mechanism of action is through inhibition of the *inhA* gene product enoyl–acyl carrier protein (acp) reductase, which is involved in mycolic acid synthesis. Ethionamide is bacteriostatic against metabolically active *M. tuberculosis* and some NTM. It is used in the treatment of drug-resistant TB, but its use is limited by severe gastrointestinal reactions (including abdominal pain, nausea, and vomiting) as well as significant central and peripheral neurologic side effects, reversible hepatitis (in ~5% of recipients), hypersensitivity reactions, and hypothyroidism. Ethionamide should be taken with food to reduce gastrointestinal effects and with pyridoxine (50–100 mg/d) to limit neuropathic side effects.

PARA-AMINOSALICYLIC ACID

Para-aminosalicylic acid (PAS, 4-aminosalicylic acid) is an oral agent used in the treatment of MDR-

and XDR-TB. Its bacteriostatic activity is due to inhibition of folate synthesis and of iron uptake. PAS has relatively little activity as an anti-tuberculous agent, with a high level nausea, vomiting, and diarrhea. PAS may cause hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency. The drug should be taken with acidic foods to improve absorption. Enteric-coated PAS granules (4 g orally every 8 h) appear to be better tolerated than other formulations and produce higher therapeutic blood levels. PAS has a short half-life (1 h), and 80% of the dose is excreted in the urine.

CYCLOSERINE

Cycloserine is an analog of the amino acid d-alanine and prevents cell wall synthesis. It inhibits the action of enzymes, including alanine racemase, that are involved in the production of peptidoglycans. Cycloserine is active against a range of bacteria, including *M. tuberculosis*. Mechanisms of mycobacterial resistance are not well understood, but overexpression of alanine racemase can confer resistance in *M. smegmatis*. Cycloserine is well absorbed after oral administration and is widely distributed throughout body fluids, including CSF. The usual adult dosage is 250 mg two or three times per day. Serious potential side effects include seizures and psychosis (with suicide in some cases), peripheral neuropathy, headache, somnolence, and allergic reactions. Drug levels are monitored to achieve optimal dosing and to reduce the risk of adverse effects, especially in patients with renal failure. Cycloserine should be administered as DOT only with caution and with support from experienced TB physicians to patients with epilepsy, active alcohol abuse, severe renal insufficiency, or a history of depression or psychosis.

Newer Antituberculosis Drugs in Clinical Trials

LINEZOLID

Linezolid is an oxazolidinone used primarily for the treatment of drug-resistant gram-positive infections. However, this drug is active in vitro against *M. tuberculosis* and NTM. Several case series have suggested that linezolid may help clear organisms relatively rapidly when included in a regimen for the treatment of complex MDR- and XDR-TB. Linezolid's mechanism of action is disruption of protein synthesis by binding to the 50S bacterial ribosome. Linezolid has nearly 100% oral bioavailability, with good penetration into tissues and fluids, including CSF. Clinical drug resistance to linezolid has been reported, but the mechanism is unclear. Adverse effects may include optic and peripheral neuropathy, pancytopenia, and lactic acidosis. Linezolid is a weak MAO inhibitor and can be associated with the serotonin syndrome when given concomitantly with serotonergic drugs (primarily antidepressants such as selective serotonin reuptake inhibitors). Prolonged administration for TB and other mycobacterial infections may be associated with an increased rate of side effects.

TMC207

TMC207 is a new diarylquinoline with a novel mechanism of action: inhibition of the mycobacterial ATP synthetase proton pump. TMC207 is bactericidal for drug-susceptible and MDR strains of *M. tuberculosis*. Resistance has been reported and is due to point mutations in the gene coding for the ATP synthetase proton pump. A phase 2 randomized controlled clinical trial demonstrated substantial improvement in rates of 2-month culture conversion, with improved clearance of mycobacterial cultures, for MDR-TB patients. This drug is metabolized by the hepatic cytochrome CYP3A4. Rifampin lowers TMC207 levels by 50%, and protease inhibitors also interact significantly with this drug. The oral bioavailability of TMC207 appears to be excellent. The dosage is 400 mg/d for the first 2 weeks and then 200 mg thrice weekly. The elimination half-life is long (>14 days). Adverse effects are reported to be minimal, with nausea and slight prolongation of the QT_C interval.

OPC-67683 AND PA 824

The prodrugs OPC-67683 and PA 824 are novel nitro-dihydro-imidazooxazole derivatives whose antimycobacterial activity is attributable to inhibition of mycolic acid biosynthesis. Early clinical trials of these compounds are ongoing.

NONTUBERCULOUS MYCOBACTERIA

More than 150 species of NTM have been identified. Only a minority of these environmental organisms, which extensive are found in soil and water, are important human pathogens. NTM cause extensive disease primarily in persons with preexisting pulmonary disease or immunocompromise but can cause nodular/bronchiectatic disease in otherwise seemingly normal hosts. They are also important causes of infections in surgical settings. Two major classes of NTM are the slow-growing and rapidly growing species. Subcultures of the latter grow within 1 week. The growth characteristics of NTM have diagnostic, thapeutic, and prognostic implications. The rate of growth can provide useful preliminary information within a specific clinical context, in that growth within 2 -3 weeks is much more likely to indicate an NTM than M. tuberculosis. When NTM do grow from cultures, colonization should be distinguished from active disease in order to optimize the risk and benefit of prolonged treatment with multiple medications. Significant clinical manifestations and/or sputum radiographic evidence of progressive disease consistent with NTM as well as either reproducible sputum culture results or a single positive culture are required to diagnose NTM pulmonary disease, according to the recommendations of the American Thoracic Society and the Infectious Diseases Society of America.

Isolation of NTM from blood or from an infected-appearing extrapulmonary site, such as soft tissue or bone, is usually indicative of disseminated or local NTM infection (see Chap. 167). Treatment of NTM disease is prolonged and requires multiple medications. Side effects of the regimens employed are common, and intermittent therapy is often used to mitigate these adverse events. Treatment regimens depend on the NTM species, the extent or type of disease, and—to some degree—drug susceptibility test results. The nodular bronchiectatic form of MAC infection is generally treated three times per week, whereas fibrocavitary or disseminated MAC infection is treated daily.

Therapeutic Considerations for Specific NTM

***M. AVIUM* COMPLEX**

Among the NTM, MAC organisms most commonly cause human disease. In immunocompetent hosts, MAC species are most often found in conjunction with underlying significant lung disease, such as chronic obstructive pulmonary disease or bronchiectasis. For patients with nodular or bronchiectatic MAC lung disease, a initial regimen consisting of clarithromycin or azithromycin, is testing or rifabutin, and ethambutol is (i.e., with cultures persistently positive for NTM). given three times per week. Routine initial testing for macrolide resistance is recommended, as is testing at 6 months in a failing regimens (i.e., with cultures persistently positive for NTM).

In immunocompromised individuals, disseminated MAC infection is generally treated with clarithromycin, ethambutol, and rifabutin. Azithromycin may be substituted in patients unable to tolerate clarithromycin. Amikacin and fluoroquinolones are often used in salvage regimens. Treatment for disseminated MAC infection in AIDS patients may be lifelong in the absence of immune reconstitution. At least 12 months of MAC therapy and 6 months of effective immune reconstitution may be adequate.

Mycobacterium Kansasii

M. kansasii is the second most common NTM causing human disease. It is also the second most common cause of NTM pulmonary disease in the United States, where most commonly reported in southeast region. *M. kansasii* infection can be treated with isoniazid, rifampin, and ethambutol; therapy continues for 12 months after culture conversion. Rifampin-resistant *M. kansasii* has been treated with clarithromycin, trimethoprim-sulfamethoxazole, and streptomycin.

RAPIDLY GROWING MYCOBACTERIA

Rapidly growing mycobacteria causing human disease include *M. abscessus*, *M. fortuitum*, and *M. chelonae*. Treatment of these mycobacteria is complex and should be undertaken with input from experienced clinicians. Testing for macrolide resistance is recommended. However, in rapidly growing mycobacteria, an inducible *erm* gene may confer in vivo macrolide resistance to isolates that are susceptible in vitro.

Mycobacterium Marinum

M. marinum is an NTM found in salt water and freshwater, including swimming pools and fish tanks. It is a cause of localized soft tissue infections, which may require surgical management. Combination regimens include clarithromycin and either ethambutol or rifampin. Other agents with activity against *M. marinum* include doxycycline, minocycline, and trimethoprim-sulfamethoxazole.

Drugs for the Treatment of NTM

CLARITHROMYCIN

Clarithromycin is a macrolide antibiotic with broad activity against many gram-positive and gram-negative bacteria as well as NTM. This drug is active against MAC organisms and many other NTM species, inhibiting protein synthesis by binding to the 50S mycobacterial ribosomal subunit. NTM resistance to macrolides is probably caused by overexpression of the gene *ermB*, with consequent methylation of the binding site. Clarithromycin is well absorbed orally and distributes well to tissues. It is cleared both hepatically and renally; the dosage should be reduced in renal insufficiency. Clarithromycin is a substrate for and inhibits cytochrome 3A4 and should not be administered with cisapride, pimozide, or terfenadine, as cardiac arrhythmias may occur. Numerous drugs interact with clarithromycin through the CYP3A4 metabolic pathway. Rifampin lowers clarithromycin levels; conversely, rifampin levels are increased by clarithromycin. However, the clinical relevance of this interaction does not appear to be prominent.

For patients with nodular/bronchiectatic MAC infection, the dosage of clarithromycin is 500 mg, given morning and evening, three times a week. For the treatment of fibrocavitary or severe nodular/bronchiectatic MAC infection, a dose of 500–1000 mg is given daily. Disseminated MAC infection is treated with 1000 mg daily. Clarithromycin is used in combination regimens that typically include ethambutol and a rifamycin in order to avoid the development of macrolide resistance. Adverse effects include frequent gastrointestinal intolerance, hepatotoxicity, headache, rash, and rare instances of hypoglycemia. Clarithromycin is contraindicated during pregnancy because of its teratogenicity in animal models.

AZITHROMYCIN

Azithromycin is a derivative of erythromycin. Although technically an azalide and not a macrolide, it works similarly to macrolides, inhibiting protein synthesis through binding to the 50S ribosomal subunit. Resistance to azithromycin is almost always associated with complete cross-resistance to clarithromycin. Azithromycin is well absorbed orally, with good tissue penetration and a prolonged half-life (~48 h). The usual dosage for treatment of MAC infection is 250 mg daily or 500 mg three times per week. Azithromycin is used in combination with other agents to avoid development of resistance. For prophylaxis against disseminated MAC infection in immunocompromised individuals, a dose of 1200 mg once per week is given. Because azithromycin is not metabolized by cytochrome P450, it interacts with few drugs. Adjustment of the dosage on the basis of renal function is not necessary.

CEFOXITIN

Cefoxitin is a second-generation parenteral cephalosporin with activity against rapidly growing NTM, particularly *M. abscessus, M. marinum*, and *M. chelonae*. Its mechanism of action against NTM is unknown but may involve inactivation of cell wall synthesis enzymes. High doses are used for treatment of NTM: 200 mg/kg IV three or four times per day, with a maximal daily dose of 12 g. The half-life of cefoxitin is ~1 h, with primarily renal clearance that requires adjustment in renal insufficiency. Adverse effects are uncommon but include gastrointestinal manifestations, rash, eosinophilia, fever, and neutropenia.

CONCLUSION

Treatment of mycobacterial infections requires multiple-drug regimens that often exert significant side effects with the potential to limit tolerability. The prolonged duration of treatment has vastly improved results over those obtained in decades past, but drugs and regimens that will shorten treatment duration and limit adverse drug effects and interactions are needed.

FURTHER READINGS

Abdool Karim SS et al: Timing of initiation of antiretroviral drugs during tuberculosis therapy. N Engl J Med 372:697, 2010

American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: Treatment of tuberculosis. Am J Respir Crit Care Med 167:603, 2003

Crofton J, Mitchison DA: Streptomycin resistance in pulmonary tuberculosis. BMJ 2:1009, 1948 [PMID: 18100441]

Diacon AH et al: The diarylquinoline TMC207 for multidrug-resistant tuberculosis. N Engl J Med 360:2397, 2009 [PMID: 19494215]

Griffith D et al: An official ATS/IDSA statement: Diagnosis, treatment, and prevention of non-tuberculous mycobacterial diseases. Am J Respir Crit Care Med 175:367, 2007 [PMID: 17277290]

Menzies D et al: Adverse events with 4 months of rifampin therapy or 9 months of isoniazid therapy for latent tuberculosis infection: A randomized trial. Ann Intern Med 149:689, 2008 [PMID: 19017587]

Mitnick C et al: Comprehensive treatment of extensively drug-resistant tuberculosis. N Engl J Med 359:563, 2008 [PMID: 18687637]

Saukkonen JJ et al: An official ATS statement: Hepatotoxicity of antituberculosis therapy. Am J Respir Crit Care Med 174:935, 2006 [PMID: 17021358]

World Health Organization: Anti-tuberculosis drug resistance in the world. Report No. 4. Geneva, WHO, 2008

Yew WW et al: Outcomes of patients with multidrug-resistant pulmonary tuberculosis treated with ofloxacin/levofloxacin-containing regimens. Chest 117:744, 2000 [PMID: 10713001]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION

Syphilis, a chronic systemic infection caused by *Treponema pallidum* subspecies *pallidum*, is usually sexually transmitted and is characterized by episodes of active disease interrupted by periods of latency. After an incubation period averaging 2–6 weeks, a primary lesion appears, often associated with regional lymphadenopathy. The secondary stage, associated with generalized mucocutaneous lesions and generalized lymphadenopathy, is followed by a latent period of subclinical infection lasting years or decades. Central nervous system (CNS) involvement may occur early in infection and may be symptomatic or asymptomatic. In about one-third of untreated cases, the tertiary stage appears, characterized by progressive destructive mucocutaneous, musculoskeletal, or parenchymal lesions; aortitis; or late CNS manifestations.

ETIOLOGY

The Spirochaetales include four genera that are pathogenic for humans and for a variety of other animals: *Leptospira* species, which cause leptospirosis (Chap. 171); *Borrelia* species, which cause relapsing fever and Lyme disease (Chaps. 172 and 173); *Brachyspira* species, which cause intestinal infections; and *Treponema* species, which cause the diseases known collectively as treponematoses (see also Chap. 170). The *Treponema* species include *T. pallidum* subspecies *pallidum*, which causes venereal syphilis; *T. pallidum* subspecies *pertenue*, which causes yaws; *T. pallidum* subspecies *endemicum*, which causes endemic syphilis or bejel; and *T. carateum*, which causes pinta. Until recently, the subspecies were distinguished primarily by the clinical syndromes they produce. Researchers have now identified molecular signatures that can differentiate the three subspecies of *T. pallidum* by culture-independent methods based on polymerase chain reaction (PCR). Other *Treponema* species found in the human mouth, genital mucosa, and gastrointestinal tract have been associated with disease (e.g., periodontitis), but their role as primary etiologic agents is unclear.

T. pallidum subspecies *pallidum* (referred to hereafter as *T. pallidum*), a thin spiral organism, has a cell body surrounded by a trilaminar cytoplasmic membrane, a delicate peptidoglycan layer providing some structural rigidity, and a lipid-rich outer membrane containing relatively few integral membrane proteins. Endoflagella wind around the cell body in the periplasmic space and are responsible for motility.

T. pallidum cannot be cultured in vitro, and little was known about its metabolism until the genome was sequenced in 1998. This spirochete possesses severely limited metabolic capabilities, lacking the genes required for de novo synthesis of most amino acids, nucleotides, and lipids. In addition, *T. pallidum* lacks genes encoding the enzymes of the Krebs cycle and oxidative phosphorylation. To compensate, the organism contains numerous genes predicted to encode transporters of amino acids, carbohydrates, and cations. In addition, genome analyses and other studies have revealed the existence of a 12-member gene family (*tpr*) that bears similarities to variable outer-membrane antigens of other spirochetes. One member, TprK, has discrete variable (V) regions that undergo antigenic variation during infection, probably as a mechanism for immune evasion.

The only known natural host for *T. pallidum* is the human. *T. pallidum* can infect many mammals, but only humans, higher apes, and a few laboratory animals regularly develop syphilitic lesions. Virulent strains of *T. pallidum* are grown in rabbits.

TRANSMISSION AND EPIDEMIOLOGY

Nearly all cases of syphilis are acquired by sexual contact with infectious lesions [i.e., the chancre, mucous patch, skin rash, or condylomata lata (see Fig. e7-20)]. Less common modes of transmission include nonsexual personal contact, infection in utero, blood transfusion, and organ transplantation.

Syphilis in the United States

With the advent of penicillin therapy, the total number of cases of syphilis reported annually in the United States declined significantly to a low of 31,575 in 2000—a 95% decrease from 1943—with <6000 reported cases of primary and secondary syphilis. Since 2000, the number of cases of infectious primary and secondary syphilis (a better indicator of disease activity) has more than doubled, with 13,500 cases reported in 2008. These cases have particularly affected men who have sex with men (MSM), many of whom are co-infected with HIV. This outbreak among MSM is occurring throughout North America. Increases in the number of cases among women in the United States in

recent years indicate that heterosexual transmission is becoming more common. Surveillance of the number of new cases of primary and secondary syphilis has revealed multiple cycles of 7–10 years, which have been attributed to herd immunity in at-risk populations. A recent re-analysis of the data, however, fails to support this conclusion and proposes alternative explanations for the periodic rise and fall of infectious syphilis cases, including changing sexual behaviors and control efforts.

The populations at highest risk for acquiring syphilis have changed over time, with outbreaks among MSM in the late 1970s and early 1980s as well as at present. The epidemic that peaked in 1990 predominantly involved African-American heterosexual men and women and occurred largely in urban areas, where infectious syphilis was correlated significantly with the exchange of sex for crack cocaine. Although the rate of primary and secandary syphilis among African Americans declined from 1996 through 2003, the rate has nearly doubled since then and remains higher than rates for other racial/ethnic groups.

The incidence of congenital syphilis roughly parallels that of infectious syphilis in females. In 2008, 431 cases in infants <1 year of age were reported. The case definition for congenital syphilis was broadened in 1989 and now includes all live or stillborn infants delivered to women with untreated or inadequately treated syphilis.

One-third to one-half of individuals named as sexual contacts of persons with infectious syphilis become infected. Many will have already developed manifestations of syphilis when they are first seen, and ~30% of asymptomatic contacts examined within 30 days of exposure actually have incubating infection and will later develop infectious syphilis if not treated. Thus, identification and treatment of all recently exposed sexual contacts continue to be important aspects of syphilis control.

Global Syphilis

Syphilis remains a significant health problem globally; the number of new infections is estimated at nearly 12 million per year. The regions that are most affected include sub-Saharan Africa, South America, China, and Southeast Asia **(Fig. 169-1)**. During the past decade, the number of reported cases in China has increased 10-fold, and higher rates have been reported among MSM in many European countries. Worldwide, congenital syphilis has been reported to account for up to 50% of stillbirths, and between 500,000 and 1.5 million cases of congenital syphilis are estimated to occur annually.

Figure 169-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated annual new cases of syphilis among adults, 1999. *(Courtesy of the World Health Organization.)*

NATURAL COURSE AND PATHOGENESIS OF UNTREATED SYPHILIS

T. pallidum rapidly penetrates intact mucous membranes or microscopic abrasions in skin and, within a few hours, enters the lymphatics and blood to produce systemic infection and metastatic foci long before the appearance of a primary lesion. Blood from a patient with incubating or early syphilis is infectious. The generation time of *T. pallidum* during early active disease in vivo is estimated to be ~30 h, and the incubation period of syphilis is inversely proportional to the number of organisms inoculated. The 50% infectious dose for intradermal inoculation in humans has been calculated to be 57 organisms, and the treponeme concentration generally reaches 10^7/g of tissue before a clinical lesion appears. The median incubation period in humans (~21 days) suggests an average inoculum of 500

–1000 infectious organisms for naturally acquired disease; the incubation period rarely exceeds 6 weeks.

The primary lesion appears at the site of inoculation, usually persists for 4–6 weeks, and then heals spontaneously. Histopathologic examination shows perivascular infiltration, chiefly by CD4+ and CD8+ T lymphocytes, plasma cells, and macrophages, with capillary endothelial proliferation and subsequent obliteration of small blood vessels. The cellular infiltration displays a T_H1-type cytokine profile consistent with the activation of macrophages. Phagocytosis of opsonized organisms by activated macrophages ultimately causes their destruction, resulting in spontaneous resolution of the chancre.

The generalized parenchymal, constitutional, and mucocutaneous manifestations of secondary syphilis usually appear ~6–8 weeks after the chancre heals. Approximately 15% of patients with secondary syphilis still have persisting or healing chancres, and the stages may overlap more frequently in persons with concurrent HIV infection. In other patients, secondary lesions may appear several months after the chancre has healed, and some patients may enter the latent stage without ever recognizing secondary lesions. The histopathologic features of secondary maculopapular skin lesions include hyperkeratosis of the epidermis, capillary proliferation with endothelial swelling in the superficial corium, dermal papillae with transmigration of polymorphonuclear leukocytes, and—in the deeper corium—perivascular infiltration by CD8+ T lymphocytes, CD4+ T lymphocytes, macrophages, and plasma cells. Treponemes are found in many tissues, including the aqueous humor of the eye and the cerebrospinal fluid (CSF). Invasion of the CNS by *T. pallidum* occurs during the first weeks or months of infection, and CSF abnormalities are detected in as many as 40% of patients during the secondary stage. Clinical hepatitis and immune complex–induced glomerulonephritis are relatively rare but recognized manifestations of secondary syphilis; liver function tests may yield abnormal results in up to one-quarter of patients with early syphilis. Generalized nontender lymphadenopathy is noted in 85% of patients with secondary syphilis. The paradoxical appearance of secondary manifestations despite high titers of antibody (including immobilizing antibody) to *T. pallidum* may result from antigenic variation or changes in expression of surface antigens. Secondary lesions subside within 2–6 weeks, and the infection enters the latent stage, which is detectable only by serologic testing. In the preantibiotic era, up to 25% of untreated patients experienced at least one generalized or localized mucocutaneous relapse, usually during the first year. Therefore, identification and examination of sexual contacts are most important for patients with syphilis of <1 year's duration.

About one-third of patients with untreated latent syphilis developed clinically apparent tertiary disease in the preantibiotic era. In industrialized countries today, specific treatment for early and latent syphilis and coincidental therapy have nearly eliminated tertiary disease except for cases of neurosyphilis in HIV-infected persons. In the past, the most common types of tertiary disease were the gumma (a usually benign granulomatous lesion), cardiovascular syphilis (usually involving the vasa vasorum of the ascending aorta and resulting in aneurysm), and late symptomatic neurosyphilis (tabes dorsalis and paresis). Asymptomatic CNS involvement, however, is still demonstrable in up to 25% of patients with late latent syphilis. The factors that contribute to the development and progression of tertiary disease are unknown.

The course of untreated syphilis was studied retrospectively in a group of nearly 2000 patients with primary or secondary disease diagnosed clinically (the Oslo Study, 1891–1951) and was assessed prospectively in 431 African-American men with seropositive latent syphilis of ≥3 years' duration (the notorious Tuskegee Study, 1932–1972). In the Oslo Study, 24% of patients developed relapsing secondary lesions within 4 years, and 28% eventually developed one or more manifestations of tertiary syphilis. Cardiovascular syphilis, including aortitis, was detected in 10% of patients; 7% of patients developed symptomatic neurosyphilis, and 16% developed benign tertiary gummatous syphilis. Syphilis was the primary cause of death in 15% of men and 8% of women. Cardiovascular syphilis was documented in 35% of men and 22% of women who eventually came to autopsy. In general, serious late complications were nearly twice as common among men as among women.

The Tuskegee Study showed that the death rate among untreated African-American men with syphilis (25–50 years old) was 17% higher than the rate among uninfected subjects and that 30% of all deaths were attributable to cardiovascular or, to a lesser extent, CNS syphilis. Anatomic evidence of aortitis was found in 40–60% of autopsied subjects with syphilis (vs. 15% of control subjects), whereas CNS syphilis was found in only 4%. Rates of hypertension were also higher among the infected subjects. The ethical issues eventually raised by this study, begun in the preantibiotic era but continuing into the early 1970s, had a major influence on the development of current guidelines for human medical experimentation, and the history of the study may still contribute to a reluctance of some African Americans to participate as subjects in clinical research.

CLINICAL MANIFESTATIONS

PRIMARY SYPHILIS

The typical primary chancre usually begins as a single painless papule that rapidly becomes eroded and usually becomes indurated, with a characteristic cartilaginous consistency on palpation of the edge and base of the ulcer. Multiple primary lesions are seen in a minority of patients. In heterosexual

men the chancre is usually located on the penis **(Fig. 169-2)**, whereas in homosexual men it may be found in the anal canal or rectum, in the mouth, or on the external genitalia. In women, common primary sites are the cervix and labia. Consequently, primary syphilis goes unrecognized in women and homosexual men more often than in heterosexual men.

Figure 169-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary syphilis with a firm, nontender chancre.

Atypical primary lesions are common. The clinical appearance depends on the number of treponemes inoculated and on the immunologic status of the patient. A large inoculum produces a dark-field-positive ulcerative lesion in nonimmune volunteers but may produce a small dark-field-negative papule, an asymptomatic but seropositive latent infection, or no response at all in some individuals with a history of syphilis. A small inoculum may produce only a papular lesion, even in nonimmune individuals. Therefore, syphilis should be considered even in the evaluation of trivial or atypical dark-field-negative genital lesions. The genital lesions that most commonly must be differentiated from those of primary syphilis include those caused by herpes simplex virus infection (Chap. 179), chancroid (Chap. 145), traumatic injury, and donovanosis (Chap. 161). Regional (usually inguinal) lymphadenopathy accompanies the primary syphilitic lesion, appearing within 1 week of lesion onset. The nodes are firm, nonsuppurative, and painless. Inguinal lymphadenopathy is bilateral and may occur with anal as well as with external genital chancres. The chancre generally heals within 4–6 weeks (range, 2–12 weeks), but lymphadenopathy may persist for months.

SECONDARY SYPHILIS

The protean manifestations of the secondary stage usually include mucocutaneous lesions and generalized nontender lymphadenopathy. The healing primary chancre may still be present in ~15% of cases, and the stages may overlap more frequently in persons with concurrent HIV infection. The skin rash consists of macular, papular, papulosquamous, and occasionally pustular syphilides; often more than one form is present simultaneously. The eruption may be very subtle, and 25% of patients with a discernible rash may be unaware that they have dermatologic manifestations. Initial lesions are pale red or pink, nonpruritic, discrete macules distributed on the trunk and proximal extremities; these macules progress to papular lesions that are distributed widely and that frequently involve the palms and soles **(Fig. 169-3; see also Figs. e7-18 and e7-19)**. Rarely, severe necrotic lesions (*lues maligna*) may appear; they are more commonly reported in HIV-infected individuals. Involvement of the hair follicles may result in patchy alopecia of the scalp hair, eyebrows, or beard in up to 5% of cases.

Figure 169-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Secondary syphilis.*Left*: Maculopapular truncal eruption. *Middle*: Papules on the palms. *Right*: Papules on the soles. *(Courtesy of Jill McKenzie and Christina Marra.)*

In warm, moist, intertriginous areas (commonly the perianal region, vulva, and scrotum), papules can enlarge to produce broad, moist, pink or gray-white, highly infectious lesions [*condylomata lata* (see Fig. e7-20)] in 10% of patients with secondary syphilis. Superficial mucosal erosions (*mucous patches*) occur in 10–15% of patients and commonly involve the oral or genital mucosa. The typical mucous patch is a painless silver-gray erosion surrounded by a red periphery.

Constitutional symptoms that may accompany or precede secondary syphilis include sore throat (15–30%), fever (5–8%), weight loss (2–20%), malaise (25%), anorexia (2–10%), headache (10%), and meningismus (5%). *Acute meningitis* occurs in only 1–2% of cases, but CSF cell and protein concentrations are increased in up to 40% of cases, and *T. pallidum* has been recovered from CSF during primary and secondary syphilis in 30% of cases; the latter finding is often but not always associated with other CSF abnormalities.

Less common complications of secondary syphilis include hepatitis, nephropathy, gastrointestinal involvement (hypertrophic gastritis, patchy proctitis, or a rectosigmoid mass), arthritis, and periostitis. Ocular findings that suggest secondary syphilis include pupillary abnormalities and optic neuritis as well as the classic iritis or uveitis. The diagnosis of secondary syphilis is often considered in affected patients only after they fail to respond to steroid therapy. Anterior uveitis has been reported in 5–10% of patients with secondary syphilis, and *T. pallidum* has been demonstrated in aqueous humor from such patients. Hepatic involvement is common in syphilis; although usually asymptomatic, up to 25% of patients may have abnormal liver function tests. Frank *syphilitic hepatitis* may be seen. *Renal involvement* usually results from immune complex deposition and produces proteinuria associated with an acute nephrotic syndrome. Like those of primary syphilis, the manifestations of the secondary stage resolve spontaneously, usually within 1–6 months.

LATENT SYPHILIS

Positive serologic tests for syphilis, together with a normal CSF examination and the absence of clinical manifestations of syphilis, indicate a diagnosis of latent syphilis in an untreated person. The diagnosis is often suspected on the basis of a history of primary or secondary lesions, a history of exposure to syphilis, or the delivery of an infant with congenital syphilis. A previous negative serologic test or a history of lesions or exposure may help establish the duration of latent infection, which is an important factor in the selection of appropriate therapy. *Early latent* syphilis is limited to the first year after infection, whereas *late latent* syphilis is defined as that of ≥1 year's (or unknown) duration. *T. pallidum* may still seed the bloodstream intermittently during the latent stage, and pregnant women with latent syphilis may infect the fetus in utero. Moreover, syphilis has been transmitted through blood transfusion or organ donation from patients with latent syphilis. It was previously thought that untreated late latent syphilis had three possible outcomes: (1) persistent lifelong infection; (2) development of late syphilis; or (3) spontaneous cure, with reversion of serologic tests to negative. It is now apparent, however, that the more sensitive treponemal antibody tests rarely, if ever, become negative without treatment. Although progression to clinically evident late syphilis is very rare today, the occurrence of spontaneous cure is in doubt.

INVOLVEMENT OF THE CNS

Traditionally, neurosyphilis has been considered a late manifestation of syphilis, but this view is inaccurate. CNS syphilis represents a continuum encompassing early invasion (usually within the first weeks or months of infection), months to years of asymptomatic involvement, and, in some cases, development of early or late neurologic manifestations.

Asymptomatic Neurosyphilis

The diagnosis of asymptomatic neurosyphilis is made in patients who lack neurologic symptoms and signs but who have CSF abnormalities including mononuclear pleocytosis, increased protein concentrations, or CSF reactivity in the Venereal Disease Research Laboratory test. CSF abnormalities are demonstrated in up to 40% of cases of primary or secondary syphilis and in 25% of

cases of latent syphilis. *T. pallidum* has been recovered by CSF inoculation into rabbits from up to 30% of patients with primary or secondary syphilis but rarely from those with latent syphilis. The presence of *T. pallidum* in CSF is often associated with other CSF abnormalities, but organisms can be recovered from patients with otherwise-normal CSF. Although the prognostic implications of these findings in early syphilis are uncertain, it may be appropriate to conclude that even patients with early syphilis who have such findings do indeed have asymptomatic neurosyphilis and should be treated for neurosyphilis; such treatment is particularly important in patients with concurrent HIV infection. Before the advent of penicillin, the risk of development of clinical neurosyphilis in untreated asymptomatic persons was roughly proportional to the intensity of CSF changes, with the overall cumulative probability of progression to clinical neurosyphilis ~20% in the first 10 years but increasing with time. Most experts agree that neurosyphilis is more common in HIV-infected persons, while immunocompetent patients with untreated latent syphilis and normal CSF probably run a very low risk of subsequent neurosyphilis. In several recent studies, neurosyphilis was associated with a rapid plasma reagin titer of ≥1:32, regardless of clinical stage or HIV infection status.

Symptomatic Neurosyphilis

The major clinical categories of symptomatic neurosyphilis include meningeal, meningovascular, and parenchymatous syphilis. The last category includes general paresis and tabes dorsalis. The onset of symptoms usually occurs <1 year after infection for meningeal syphilis, up to 10 years after infection for meningovascular syphilis, at ~20 years for general paresis, and at 25–30 years for tabes dorsalis. However, symptomatic neurosyphilis, particularly in the antibiotic era, often presents not as a classic picture but rather as mixed and subtle or incomplete syndromes.

Meningeal syphilis may present as headache, nausea, vomiting, neck stiffness, cranial nerve involvement, seizures, and changes in mental status. This condition may be concurrent with or may follow the secondary stage. Patients presenting with uveitis, iritis, or hearing loss often have meningeal syphilis, but these clinical findings can also be seen in patients with normal CSF.

Meningovascular syphilis reflects meningitis together with inflammatory vasculitis of small, medium, or large vessels. The most common presentation is a stroke syndrome involving the middle cerebral artery of a relatively young adult. However, unlike the usual thrombotic or embolic stroke syndrome of sudden onset, meningovascular syphilis often becomes manifest after a subacute encephalitic prodrome (with headaches, vertigo, insomnia, and psychological abnormalities), which is followed by a gradually progressive vascular syndrome.

The manifestations of *general paresis* reflect widespread late parenchymal damage and include abnormalities corresponding to the mnemonic *paresis:p*ersonality, *a*ffect, *r*eflexes (hyperactive), *e*ye (e.g., Argyll Robertson pupils), *s*ensorium (illusions, delusions, hallucinations), *i*ntellect (a decrease in recent memory and in the capacity for orientation, calculations, judgment, and insight), and *s*peech. *Tabes dorsalis* is a late manifestation of syphilis that presents as symptoms and signs of demyelination of the posterior columns, dorsal roots, and dorsal root ganglia. Symptoms include ataxic wide-based gait and foot drop; paresthesia; bladder disturbances; impotence; areflexia; and loss of positional, deep-pain, and temperature sensations. Trophic joint degeneration (Charcot's joints) and perforating ulceration of the feet can result from loss of pain sensation. The small, irregular Argyll Robertson pupil, a feature of both tabes dorsalis and paresis, reacts to accommodation but not to light. *Optic atrophy* also occurs frequently in association with tabes.

OTHER MANIFESTATIONS OF LATE SYPHILIS

The slowly progressive inflammatory disease leading to tertiary disease begins early during infection, although these manifestations may not become clinically apparent for years or decades. Early syphilitic aortitis becomes evident soon after secondary lesions subside, and treponemes that trigger the development of gummas may have seeded the tissue years earlier.

Cardiovascular Syphilis

Cardiovascular manifestations, usually appearing 10–40 years after infection, are attributable to endarteritis obliterans of the vasa vasorum, which provide the blood supply to large vessels; *T. pallidum* DNA has been detected by PCR in aortic tissue. Cardiovascular involvement results in uncomplicated aortitis, aortic regurgitation, saccular aneurysm (usually of the ascending aorta), or coronary ostial stenosis. In the preantibiotic era, symptomatic cardiovascular complications developed in ~10% of persons with late untreated syphilis, although syphilitic aortitis was demonstrated at autopsy in about one-half of African-American men with untreated syphilis. Today, this form of late syphilis is rarely seen in the developed world. Linear calcification of the ascending aorta on chest x-ray films suggests asymptomatic syphilitic aortitis, as arteriosclerosis seldom produces this sign. Syphilitic aneurysms—usually saccular, occasionally fusiform—do not lead to dissection. Only 1 in 10 aortic aneurysms of syphilitic origin involves the abdominal aorta.

Late Benign Syphilis (Gumma)

Gummas are usually solitary lesions ranging from microscopic to several centimeters in diameter. Histologic examination shows a granulomatous inflammation, with a central area of necrosis due to

endarteritis obliterans. Although rarely demonstrated microscopically, *T. pallidum* has been detected by PCR or recovered from these lesions, and penicillin treatment results in rapid resolution, confirming the treponemal stimulus for the inflammation. Common sites include the skin and skeletal system; however, any organ (including the brain) may be involved. Gummas of the skin produce indolent, painless, indurated nodular or ulcerative lesions that may resemble other chronic granulomatous conditions, including tuberculosis, sarcoidosis, leprosy, and deep fungal infections. Skeletal gummas most frequently involve the long bones, although any bone may be affected. Upper respiratory gummas can lead to perforation of the nasal septum or palate.

CONGENITAL SYPHILIS

Transmission of *T. pallidum* across the placenta from a syphilitic woman to her fetus may occur at any stage of pregnancy, but fetal damage generally does not occur until after the fourth month of gestation, when fetal immunologic competence begins to develop. This timing suggests that the pathogenesis of congenital syphilis, like that of adult syphilis, depends on the host immune response rather than on a direct toxic effect of *T. pallidum*. The risk of fetal infection during untreated early maternal syphilis is ~75–95%, decreasing to ~35% for maternal syphilis of >2 years' duration. Adequate treatment of the woman before the 16th week of pregnancy should prevent fetal damage, and treatment before the third trimester should adequately treat the infected fetus. Untreated maternal infection may result in a rate of fetal loss of up to 40% (with stillbirth more common than abortion because of the late onset of fetal pathology), prematurity, neonatal death, or nonfatal congenital syphilis. Among infants born alive, only fulminant congenital syphilis is clinically apparent at birth, and these babies have a very poor prognosis. The most common clinical problem is the healthy-appearing baby born to a mother with a positive serologic test. Routine serologic testing in early pregnancy is considered cost-effective in virtually all populations, even in areas with a low prenatal prevalence of syphilis. Low-tech point-of-care tests are being developed to facilitate antenatal testing in resource-poor settings. Where the prevalence of syphilis is high or when the patient is at high risk of re-infection, serologic testing should be repeated in the third trimester and at delivery. Neonatal congenital syphilis must be differentiated from other generalized congenital infections, including rubella, cytomegalovirus or herpes simplex virus infection, and toxoplasmosis, as well as from erythroblastosis fetalis.

The manifestations of congenital syphilis include (1) early manifestations, which appear within the first 2 years of life (often at 2–10 weeks of age), are infectious, and resemble the manifestations of secondary syphilis in the adult; (2) late manifestations, which appear after 2 years and are noninfectious; and (3) residual stigmata. The earliest manifestations of congenital syphilis (appearing 2–6 weeks after birth) include rhinitis, or "snuffles" (23%); mucocutaneous lesions (35–41%); bone changes (61%), including osteochondritis, osteitis, and periostitis detectable by x-ray examination of long bones; hepatosplenomegaly (50%); lymphadenopathy (32%); anemia (34%); jaundice (30%); thrombocytopenia; and leukocytosis. CNS invasion by *T. pallidum* is detectable in 22% of infected neonates. Neonatal death is usually due to pulmonary hemorrhage, secondary bacterial infection, or severe hepatitis.

Late congenital syphilis (untreated after 2 years of age) is subclinical in 60% of cases; the clinical spectrum in the remainder of cases may include interstitial keratitis (which occurs at 5–25 years of age), eighth-nerve deafness, and recurrent arthropathy. Bilateral knee effusions are known as *Clutton's joints*. Asymptomatic neurosyphilis is present in about one-third of untreated patients, and clinical neurosyphilis occurs in one-quarter of untreated individuals >6 years old. Gummatous periostitis occurs at 5–20 years of age and, as in nonvenereal endemic syphilis, tends to cause destructive lesions of the palate and nasal septum.

Classic stigmata include *Hutchinson's teeth* (centrally notched, widely spaced, peg-shaped upper central incisors), "mulberry" molars (sixth-year molars with multiple, poorly developed cusps), saddle nose, and saber shins.

LABORATORY EXAMINATIONS

DEMONSTRATION OF THE ORGANISM

T. pallidum cannot be detected by culture. Historically, dark-field microscopy and immunofluorescence antibody staining have been used to identify this spirochete in samples from moist lesions such as chancres or condylomata lata, but these tests are rarely available today outside of research laboratories. More sensitive PCR tests have been developed but are not commercially available, although some laboratories perform in-house PCR testing.

T. pallidum can be found in tissue with appropriate silver stains, but these results should be interpreted with caution because artifacts resembling *T. pallidum* are often seen. Tissue treponemes can be demonstrated more reliably in research laboratories by PCR or by immunofluorescence or immunohistochemical methods using specific monoclonal or polyclonal antibodies to *T. pallidum*.

SEROLOGIC TESTS FOR SYPHILIS

There are two types of serologic test for syphilis: nontreponemal and treponemal. Both are reactive in persons with any treponemal infection, including yaws, pinta, and endemic syphilis.

The most widely used nontreponemal antibody tests for syphilis are the rapid plasma reagin (RPR) and Venereal Disease Research Laboratory (VDRL) tests, which measure IgG and IgM directed against a cardiolipin-lecithin-cholesterol antigen complex. The RPR test is easier to perform and uses unheated serum; it is the test of choice for rapid serologic diagnosis in a clinical setting and can be automated. The VDRL test remains the standard for examining CSF. The RPR and VDRL tests are recommended for screening or for quantitation of serum antibody. The titer reflects disease activity, rising during the evolution of early syphilis and often exceeding 1:32 in secondary syphilis. After therapy for early syphilis, a persistent fall by fourfold or more (e.g., a decline from 1:32 to 1:8) is considered an adequate response. VDRL titers do not correspond directly to RPR titers, and sequential quantitative testing (as for response to therapy) must employ a single test. As will be discussed (see "Evaluation for Neurosyphilis," below), the RPR titer may be useful in determining which patients will benefit from CSF examination.

Treponemal tests measure antibodies to native or recombinant *T. pallidum* antigens and include the fluorescent treponemal antibody–absorbed (FTA-ABS) test and the *T. pallidum* particle agglutination (TPPA) test, both of which are more sensitive for primary syphilis than the previously used hemagglutination tests. The *T. pallidum* hemagglutination (TPHA) test is widely used in Europe but is not available in the United States. When used to confirm positive nontreponemal test results, treponemal tests have a very high positive predictive value for diagnosis of syphilis. In a screening setting, however, these tests give false-positive results at rates as high as 1–2%. Treponemal tests are likely to remain reactive even after adequate treatment and cannot differentiate past from current *T. pallidum* infection.

Treponemal immunochromatographic strip (ICS) tests and enzyme immunoassays (EIAs), based largely on reactivity to recombinant antigens, have also been developed. Treponemal EIAs have been approved as confirmatory tests and, because of their ease of automation, are now used for screening purposes by some large laboratories. Because treponemal tests cannot distinguish between current and treated syphilis or may be falsely reactive, clinicians may be uncertain about how to interpret reactive EIA screening results. **Figure 169-4** provides a suggested algorithm for management of such cases.

Considerable interest has recently been focused on point-of-care ICS tests that can be used in the field or in resource-poor settings. These treponemal tests are not yet approved for use in the United States but have been assessed in antenatal clinics in a number of developing countries.

Figure 169-4

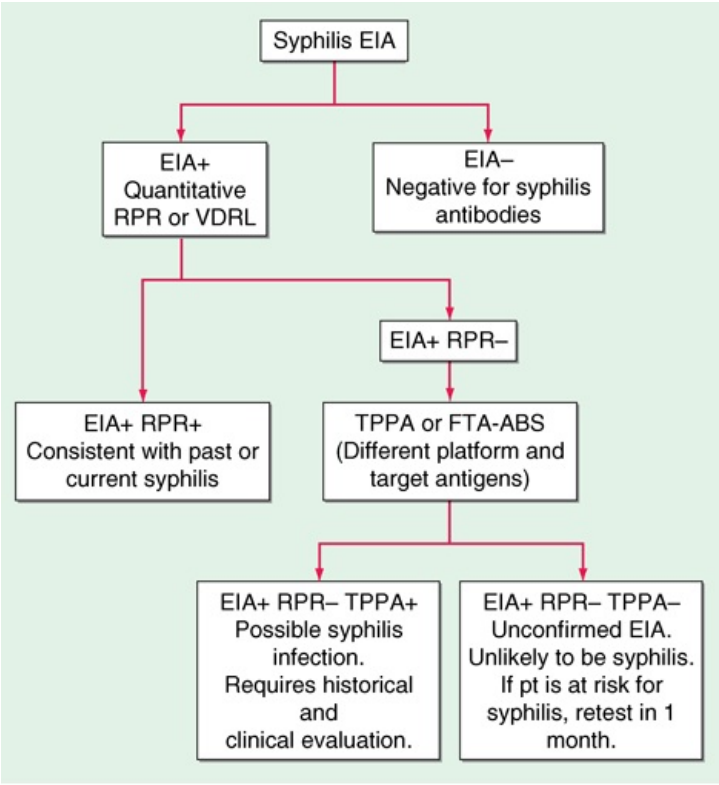

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for interpretation of results from syphilis enzyme immunoassays (EIAs) used for

screening. RPR, rapid plasma reagin; VDRL, Venereal Disease Research Laboratory; TPPA, *T. pallidum* particle agglutination; FTA-ABS, fluorescent treponemal antibody–absorbed. *(Based on the 2010 Sexually Transmitted Diseases Treatment Guidelines from the Centers for Disease Control and Prevention.)*

Both nontreponemal and treponemal tests may be nonreactive in early primary syphilis, although treponemal tests are slightly more sensitive (85–90%) during this stage than nontreponemal tests (~80%). All tests are reactive during secondary syphilis. (Fewer than 1% of patients with high titers have a nontreponemal test that is nonreactive or weakly reactive with undiluted serum but is reactive with diluted serum—the *prozone phenomenon*.) VDRL and RPR sensitivity and titers may decline in untreated persons with late latent or late syphilis, but treponemal tests remain sensitive in these stages. Whereas nontreponemal test titers will decline or the tests will become nonreactive after therapy for early syphilis, treponemal tests often remain reactive after therapy and are not helpful in determining the infection status of persons with past syphilis.

For practical purposes, most clinicians need to be familiar with three uses of serologic tests for syphilis: (1) screening or diagnosis (RPR or VDRL), (2) quantitative measurement of antibody to assess clinical syphilis activity or to monitor response to therapy (RPR or VDRL), and (3) confirmation of a syphilis diagnosis in a patient with a reactive RPR or VDRL test (FTA-ABS, TPPA, EIA). Studies have not demonstrated the utility of IgM testing for adult syphilis. While IgM titers appear to decline after therapy, the presence or absence of specific IgM does not strictly correlate with *T. pallidum* infection. Moreover, no commercially available IgM test is recommended for evaluation of infants with suspected congenital syphilis.

FALSE-POSITIVE SEROLOGIC TESTS FOR SYPHILIS

The lipid antigens of nontreponemal tests are similar to those found in human tissues, and the tests may be reactive (usually with titers ≤1:8) in persons without treponemal infection. Among patients being screened for syphilis because of risk factors, clinical suspicion, or history of exposure, ~1% of reactive tests are falsely positive. Modern VDRL and RPR tests are highly specific, and false-positive reactions are largely limited to persons with autoimmune conditions or injection drug use. The prevalence of false-positive results increases with advancing age, approaching 10% among persons >70 years old. In a patient with a false-positive nontreponemal test, syphilis is excluded by a nonreactive treponemal test.

EVALUATION FOR NEUROSYPHILIS

Involvement of the CNS is detected by examination of CSF for pleocytosis (>5 white blood cells/μL), increased protein concentration (>45 mg/dL), or VDRL reactivity. Elevated CSF cell counts and protein concentrations are not specific for neurosyphilis and may be confounded by HIV co-infection. Because CSF pleocytosis may also be due to HIV, some studies have suggested using a CSF white-cell cutoff of 20 cells/μL as diagnostic of neurosyphilis in HIV-infected patients with syphilis. The CSF VDRL test is highly specific and, when reactive, is considered diagnostic of neurosyphilis; however, this test is insensitive and may be nonreactive even in cases of symptomatic neurosyphilis. The FTA-ABS test on CSF is reactive far more often than the VDRL test on CSF in all stages of syphilis, but reactivity may reflect passive transfer of serum antibody into the CSF. A nonreactive FTA-ABS test on CSF, however, may be used to rule out asymptomatic neurosyphilis. A recent study demonstrated the utility of measuring CXCL13 in CSF to distinguish between neurosyphilis and HIV-related CSF abnormalities.

Clearly, all *T. pallidum*–infected patients who have signs or symptoms consistent with neurologic disease (e.g., meningitis, hearing loss) or ophthalmic disease (e.g., uveitis, iritis) should have a CSF examination, regardless of disease stage. The appropriate management of asymptomatic persons is less clear. Lumbar puncture on all asymptomatic patients with untreated syphilis is impractical and unnecessary. Because standard therapy with penicillin G benzathine fails to result in treponemicidal drug levels in CSF, it is important to identify those persons at higher risk for having or developing neurosyphilis so that appropriate therapy may be given. Large-scale prospective studies have now provided evidence-based guidelines for determining which syphilis patients may benefit most from CSF examination for evidence of neurosyphilis. Specifically, patients with RPR titers of ≥1:32 are at higher risk of having neurosyphilis (11-fold and 6-fold higher in HIV-infected and HIV-uninfected persons, respectively), as are HIV-infected patients with CD4+ T cell counts of ≤350/μL. Current recommendations for CSF examination are shown in **Table 169-1**.

Table 169-1 Indications for CSF Examination in Adults with All Stages of Syphilis

All Patients
Signs or symptoms of nervous system involvement [e.g., meningitis, hearing loss, cranial nerve dysfunction, altered mental status, ophthalmic disease (e.g., uveitis, iritis, pupillary abnormalities), ataxia, loss of vibration sense], *or*
RPR or VDRL titer ≥1:32, *or*

Suspected treatment failure
Additional Indications in HIV-Infected Persons
CD4+ T cell count ≤350/μL, *or* CSF examination is recommended by some experts for all HIV-infected persons.

Source: Adapted from the 2010 Sexually Transmitted Diseases Treatment Guidelines from the Centers for Disease Control and Prevention.

EVALUATION OF HIV-INFECTED PATIENTS FOR SYPHILIS

Because persons at highest risk for syphilis are also at increased risk for HIV infection, these two infections frequently coexist. There is evidence that syphilis and other genital ulcer diseases may be important risk factors in the acquisition and transmission of HIV infection. Some manifestations of syphilis may be altered in patients with concurrent HIV infection, and multiple cases of neurologic relapse after standard therapy have been reported in these patients. *T. pallidum* has been isolated from the CSF of several patients (with and without concurrent HIV infection) after penicillin G benzathine therapy for early syphilis.

Persons with newly diagnosed HIV infection should be tested for syphilis; conversely, all patients with newly diagnosed syphilis should be tested for HIV infection. Some authorities, persuaded by reports of persistent *T. pallidum* in CSF of HIV-infected persons after standard therapy for early syphilis, recommend CSF examination for evidence of neurosyphilis for all co-infected patients, regardless of the stage of syphilis, with treatment for neurosyphilis if CSF abnormalities are found. Others, on the basis of their own clinical experience, believe that standard therapy—without CSF examination—is sufficient for all cases of early syphilis in HIV-infected patients without neurologic signs or symptoms. As described above, RPR titer and CD4+ T cell count can be used to identify patients at higher risk of neurosyphilis for lumbar puncture, although some cases of neurosyphilis will not be identified by these criteria. Table 169-1 summarizes guidelines suggested by published studies. Serologic testing after treatment is important for all patients with syphilis, particularly for those also infected with HIV.

Treatment: Syphilis

TREATMENT OF ACQUIRED SYPHILIS

The CDC's 2010 guidelines for the treatment of syphilis are summarized in **Table 169-2** and are discussed below. Penicillin G is the drug of choice for all stages of syphilis. *T. pallidum* is killed by very low concentrations of penicillin G, although a long period of exposure to penicillin is required because of the unusually slow rate of multiplication of the organism. The efficacy of penicillin against syphilis remains undiminished after60 years of use, and there is no evidence of penicillin resistance in *T. pallidum*. Other antibiotics effective in syphilis include the tetracyclines and the cephalosporins. Aminoglycosides and spectinomycin inhibit *T. pallidum* only in very large doses, and the sulfonamides and the quinolones are inactive. Azithromycin has shown significant promise as an effective oral agent against *T. pallidum*; however, strains harboring 23S rRNA mutations that confer macrolide resistance are widespread; such strains represent >50% of recent isolates from Seattle and San Francisco and have now been identified in multiple North American and European sites. Macrolide resistance mutations have been identified in nearly all samples reported from China. In contrast, three studies conducted in Africa (Uganda, Tanzania, and Madagascar) have documented the clinical efficacy of azithromycin and, for the subset of samples examined, have found no molecular evidence of the mutation. In short, the prevalence of resistant strains varies widely by geographic location, and routine treatment of syphilis with azithromycin is not recommended. In all cases, careful follow-up of any patient treated for syphilis with azithromycin must be ensured.

Table 169-2 Recommendations for the Treatment of Syphilis[a]

Stage of Syphilis	Patients without Penicillin Allergy	Patients with Confirmed Penicillin Allergy[b]
Primary, secondary, or early latent	*CSF normal or not examined:*Penicillin G benzathine (single dose of 2.4 mU IM) *CSF abnormal:* Treat as neurosyphilis	*CSF normal or not examined:*Tetracycline HCl (500 mg PO qid) or doxycycline (100 mg PO bid) for 2 weeks *CSF abnormal:* Treat as neurosyphilis
Late latent (or latent of uncertain duration), cardiovascular, or benign tertiary	*CSF normal or not examined:*Penicillin G benzathine (2.4 mU IM weekly for 3 weeks)	*CSF normal and patient not infected with HIV:* Tetracycline HCl (500 mg PO qid) or

tertiary	*CSF abnormal:* Treat as neurosyphilis	doxycycline (100 mg PO bid) for 4 weeks *CSF normal and patient infected with HIV:* Desensitization and treatment with penicillin if compliance cannot be ensured *CSF abnormal:* Treat as neurosyphilis
Neurosyphilis (asymptomatic or symptomatic)	Aqueous crystalline penicillin G (18–24 mU/d IV, given as 3–4 mU q4h or continuous infusion) for 10–14 days *or* Aqueous procaine penicillin G (2.4 mU/d IM) plus oral probenecid (500 mg qid), both for 10–14 days	Desensitization and treatment with penicillin[c]
Syphilis in pregnancy	According to stage	Desensitization and treatment with penicillin

[a]See Table 169-1 and text for indications for CSF examination.

[b] Because of the documented presence of macrolide resistance in many T. pallidum strains in North America, Europe, and China, azithromycin should be used with caution only when treatment with penicillin or doxycycline is not feasible. Azithromycin should not be used for men who have sex with men or for pregnant women.

[c] Limited data suggest that ceftriaxone (2 g/d either IM or IV for 10–14 days) can be used; however, cross-reactivity between penicillin and ceftriaxone is possible.

[c]***Abbreviations:*** CSF, cerebrospinal fluid; mU, million units.

Source: Based on the 2010 Sexually Transmitted Diseases Treatment Guidelines from the Centers for Disease Control and Prevention.

Early Syphilis Patients and Their Contacts

Penicillin G benzathine is the most widely used agent for the treatment of early syphilis; a single dose of 2.4 million units is recommended. Preventive treatment is also recommended for individuals who have been exposed to infectious syphilis within the previous 3 months. *The regimens recommended for prevention are the same as those recommended for early syphilis*. Penicillin G benzathine cures >95% of cases of early syphilis, although clinical relapse can follow treatment, particularly in patients with concurrent HIV infection. Because the risk of neurologic relapse may be higher in HIV-infected patients, CSF examination is recommended in HIV-seropositive individuals with syphilis of any stage, particularly those with a serum RPR titer of ≥1:32 or a CD4+ T cell count of ≤350/μL. Therapy appropriate for neurosyphilis should be given if there is any evidence of CNS disease.

Late Latent and Late Syphilis

If CSF abnormalities are found, the patient should be treated for neurosyphilis. If CSF is normal, the recommended treatment is penicillin G benzathine (7.2 million units total; Table 169-2). The clinical response to treatment for benign tertiary syphilis is usually impressive. However, responses to therapy for cardiovascular syphilis are not dramatic because aortic aneurysm and aortic regurgitation cannot be reversed by antibiotics.

Penicillin-Allergic Patients

For penicillin-allergic patients with syphilis, a 2-week (early syphilis) or 4-week (late or late latent syphilis) course of therapy with doxycycline or tetracycline is recommended. These regimens appear to be effective in early syphilis but have not been tested for late or late latent syphilis, and compliance may be problematic. Limited studies suggest that ceftriaxone (1 g/d, given IM or IV, for 8–10 days) is effective for early syphilis. These nonpenicillin regimens have not been carefully evaluated in HIV-infected individuals and should be used with caution. If compliance and follow-up cannot be ensured, penicillin-allergic HIV-infected persons with late latent or late syphilis should be desensitized and treated with penicillin.

Neurosyphilis

Penicillin G benzathine, given in total doses of up to 7.2 million units, does not produce detectable concentrations of penicillin G in CSF and should not be used for treatment of neurosyphilis. Asymptomatic neurosyphilis may relapse into symptomatic disease following treatment with benzathine penicillin, and the risk of relapse may be higher in HIV-infected patients. Both symptomatic and

asymptomatic neurosyphilis should be treated with aqueous penicillin. Administration either of IV aqueous crystalline penicillin G or of aqueous procaine penicillin G plus probenecid in recommended doses is thought to ensure treponemicidal concentrations of penicillin G in CSF. The clinical response to penicillin therapy for meningeal syphilis is dramatic, but treatment of neurosyphilis with existing parenchymal damage may only arrest disease progression. Neurologic relapse has been reported after high-dose IV penicillin therapy for neurosyphilis in an HIV-infected patient. No alternative therapies have been studied, but careful follow-up is essential, and re-treatment is warranted in such patients. No data suggest that additional therapy (e.g., penicillin G benzathine for 3 weeks) is beneficial after treatment for neurosyphilis.

The use of antibiotics other than penicillin G for the treatment of neurosyphilis has not been studied, although very limited data suggest that ceftriaxone may be used. In patients with penicillin allergy demonstrated by skin testing, desensitization and treatment with penicillin are recommended.

Management of Syphilis in Pregnancy

Every pregnant woman should undergo a nontreponemal test at her first prenatal visit and, if at high risk of exposure, again in the third trimester and at delivery. In the untreated pregnant patient with presumed syphilis, expeditious treatment appropriate to the stage of the disease is essential. Patients should be warned of the risk of a Jarisch-Herxheimer reaction, which may be associated with mild premature contractions but rarely results in premature delivery.

Penicillin is the only recommended agent for the treatment of syphilis in pregnancy. If the patient has a documented penicillin allergy, desensitization and penicillin therapy should be undertaken according to the CDC's 2010 guidelines. After treatment, a quantitative nontreponemal test should be repeated monthly throughout pregnancy to assess therapeutic efficacy. Treated women whose antibody titers rise by fourfold or whose titers do not decrease by fourfold over a 3-month period should be re-treated.

EVALUATION AND MANAGEMENT OF CONGENITAL SYPHILIS

Whether or not they are infected, newborn infants of mothers with reactive serologic tests may themselves have reactive tests because of transplacental transfer of maternal IgG antibody. For asymptomatic infants born to women treated adequately with penicillin during the first or second trimester of pregnancy, monthly quantitative nontreponemal tests may be performed to monitor for appropriate reduction in antibody titers. Rising or persistent titers indicate infection, and the infant should be treated. Detection of neonatal IgM antibody may be useful, but no commercially available test is currently recommended.

An infant should be treated at birth if the treatment status of the seropositive mother is unknown; if the mother has received inadequate or nonpenicillin therapy or has received penicillin therapy in the third trimester; or if the infant may be difficult to follow. The CSF should be examined to obtain baseline values before treatment. Penicillin is the only recommended drug for the treatment of syphilis in infants. Specific recommendations for the treatment of infants and older children are included in the CDC's 2010 treatment guidelines.

JARISCH-HERXHEIMER REACTION

A dramatic though usually mild reaction consisting of fever, chills, myalgias, headache, tachycardia, increased respiratory rate, increased circulating neutrophil count, and vasodilation with mild hypotension may follow the initiation of treatment for syphilis. This reaction is thought to be a response to lipoproteins released by dying *T. pallidum* organisms. The Jarisch-Herxheimer reaction occurs in ~ 50% of patients with primary syphilis, 90% of those with secondary syphilis, and a lower proportion of persons with later-stage disease. Defervescence takes place within 12–24 h. In patients with secondary syphilis, erythema and edema of the mucocutaneous lesions may increase. Patients should be warned to expect such symptoms, which can be managed with symptom-based treatment. Steroid and other anti-inflammatory therapy is not required for this mild transient reaction.

FOLLOW-UP EVALUATION OF RESPONSES TO THERAPY

Efficacy of treatment should be assessed by clinical evaluation and monitoring of the quantitative VDRL or RPR titer for a fourfold decline (e.g., from 1:32 to 1:8). Patients with primary or secondary syphilis should be examined 6 and 12 months after treatment and persons with latent or late syphilis at 6, 12, and 24 months. More frequent clinical and serologic examination (3, 6, 9, 12, and 24 months) is recommended for patients concurrently infected with HIV, regardless of the stage of syphilis.

After successful treatment of seropositive first-episode primary or secondary syphilis, the VDRL or RPR titer progressively declines, becoming negative by 12 months in 40–75% of seropositive primary cases and in 20–40% of secondary cases. Patients with HIV infection or a history of prior syphilis are less likely to become nonreactive in the VDRL or RPR test. Rates of decline of serologic titers appear to be slower and serologically defined treatment failures more common among HIV-infected patients than among those without HIV co-infection; however, effective antiretroviral therapy may reduce these differences. Re-treatment should be considered if serologic responses are not adequate or if clinical signs persist or recur. Because it is difficult to differentiate treatment failure from reinfection, the CSF should be examined, with treatment for neurosyphilis if CSF is abnormal and treatment for late latent

syphilis if CSF is normal. Patients treated for late latent syphilis frequently have low initial VDRL or RPR titers and may not have a fourfold decline after therapy with penicillin. In such patients, re-treatment is not warranted unless the titer rises or signs and symptoms of syphilis appear. Because treponemal tests may remain positive despite treatment for seropositive syphilis, these tests are not useful in following the response to therapy.

The activity of neurosyphilis (symptomatic or asymptomatic) correlates best with CSF pleocytosis, and this measure provides the most sensitive index of response to treatment. Repeat CSF examinations should be performed every 6 months until the cell count is normal. An elevated CSF cell count falls to normal in 3–12 months in adequately treated HIV-uninfected patients. The persistence of mild pleocytosis in HIV-infected patients may be due to the presence of HIV in CSF; this scenario may be difficult to distinguish from treatment failure. Elevated levels of CSF protein fall more slowly, and the CSF VDRL titer declines gradually over several years. In patients treated for neurosyphilis, a fourfold reduction in serum RPR titer has been positively correlated with normalization of CSF abnormalities; this correlation is stronger in HIV-uninfected patients and in HIV-infected patients receiving effective antiretroviral therapy.

IMMUNITY TO SYPHILIS

The rate of development of acquired resistance to *T. pallidum* after natural or experimental infection is related to the size of the antigenic stimulus, which depends on both the size of the infecting inoculum and the duration of infection before treatment. Both humoral and cellular responses are considered to be of major importance in immunity and in the healing of early lesions. Cellular infiltration, predominantly by T lymphocytes and macrophages, produces a T_H1 cytokine milieu consistent with the clearance of organisms by activated macrophages. Specific antibody enhances phagocytosis and is required for macrophage-mediated killing of *T. pallidum*. Recent studies demonstrate antigenic variation of the TprK protein, which may lead to persistence of infection and determine susceptibility to reinfection with another strain. Comparative genomic studies have revealed some sequence variations among *T. pallidum* strains. Strains can be differentiated by molecular typing methods, and a possible correlation between molecular type and clinical manifestations is being examined.

FURTHER READINGS

Breban R et al: Is there any evidence that syphilis epidemics cycle? Lancet Infect Dis 8:577, 2008[PMID: 18718441]

Centers for Disease Control and Prevention: Sexually transmitted diseases treatment guidelines, 2010. MMWR Recommend Rep 59(RR-12): 1–110, 2010

Ghanem KG et al: Lumbar puncture in HIV-infected patients with syphilis and no neurologic symptoms. Clin Infect Dis 48:816, 2009[PMID: 19187028]

LaFond RE, Lukehart SA: Biological basis for syphilis. Clin Microbiol Rev 19:29, 2006[PMID: 16418521]

Marra CM et al: Cerebrospinal fluid abnormalities in patients with syphilis: Association with clinical and laboratory features. J Infect Dis 189:369, 2004[PMID: 14745693]

——— et al: Normalization of serum rapid plasma reagin titer predicts normalization of cerebrospinal fluid and clinical abnormalities after treatment of neurosyphilis. Clin Infect Dis 47:893, 2008

——— et al: CXCL13 as a cerebrospinal fluid marker for neurosyphilis in HIV-infected patients with syphilis. Sex Transm Dis 37:283, 2010

Stamm LV: Global challenge of antibiotic-resistant *Treponema pallidum*. Antimicrob Agents Chemother 54:583, 2010[PMID: 19805553]

Tucker JD et al: Lues maligna in early HIV infection: Case report and review of the literature. Sex Transm Dis 36:512, 2009[PMID: 19455078]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ENDEMIC TREPONEMATOSES: INTRODUCTION

The endemic, or nonvenereal, treponematoses are bacterial infections caused by close relatives of *Treponema pallidum* subspecies *pallidum*, the etiologic agent of venereal syphilis (Chap. 169). Yaws, pinta, and endemic syphilis are traditionally distinguished from venereal syphilis by mode of transmission, age of acquisition, geographic distribution, and clinical features. These infections are limited to rural areas of developing nations and are seen in developed countries only among recent immigrants from endemic regions. Our "knowledge" about the endemic treponematoses is based on observations by health care workers who have visited endemic areas; virtually no well-designed studies of the natural history, diagnosis, or treatment of these infections have been conducted. The treponemal infections are compared and contrasted in **Table 170-1**.

Table 170-1 Comparison of the Treponemes and Associated Diseases

Feature	Venereal Syphilis	Yaws	Endemic Syphilis	Pinta
Organism	*T. pallidum subsp. pallidum*	*T. pallidum subsp. pertenue*	*T. pallidum subsp. endemicum*	*T. carateum*
Modes of transmission	Sexual, transplacental	Skin-to-skin	Household contacts: mouth-to-mouth or via shared drinking/ eating utensils	Skin-to-skin
Usual age of acquisition	Adulthood or in utero	Early childhood	Early childhood	Late childhood
Primary lesion	Cutaneous ulcer (chancre)	Papilloma, often ulcerative	Rarely seen	Nonulcerating papule with satellites, pruritic
Location	Genital, oral, anal	Extremities	Oral	Extremities, face
Secondary lesions	Mucocutaneous lesions; condylomata lata	Cutaneous papulosquamous lesions; osteoperiostitis	Florid mucocutaneous lesions (mucous patch, split papule, condyloma latum); osteoperiostitis	Pintides, pigmented, pruritic
Infectious relapses	~25%	Common	Unknown	None
Late complications	Gummas, cardiovascular and CNS involvement[a]	Destructive gummas of skin, bone, cartilage	Destructive gummas of skin, bone, cartilage	Nondestructive, dyschromic, achromic macules

[a] CNS involvement in the endemic treponematoses has been postulated by some investigators (see text).

Epidemiology

The endemic treponematoses are chronic diseases transmitted by direct contact during childhood and, like syphilis, can cause severe late manifestations years after initial infection. In a World Health Organization (WHO)–sponsored mass eradication campaign from 1952 to 1969, more than 160 million people in Africa, Asia, and South America were examined for treponemal infections, and more than 50 million cases, contacts, and latent infections were treated. This campaign reduced the prevalence of active yaws from >20% to <1% in many areas. In recent decades, lack of focused surveillance and diversion of resources have resulted in documented resurgence of these infections in some regions. The estimated geographic distribution of the endemic treponematoses in the 1990s is shown in **Fig. 170-1**. The most recent WHO estimate (1997) suggested that there are 460,000 new cases per year and a prevalence of 2.5 million infected persons; during the subsequent decade, an increased incidence was documented in some countries. Areas of resurgent yaws morbidity include West Africa (Ivory Coast, Ghana, Togo, Benin), the Central African Republic, Nigeria, and rural Democratic

Republic of Congo. The prevalence of endemic syphilis is estimated to be >10% in some regions of Ghana, Mali, Niger, Burkina Faso, and Senegal. In Asia and the Pacific Islands, reports suggest active outbreaks of yaws in Indonesia, Papua New Guinea, East Timor, Vanuatu, Laos, and Kampuchea. India actively renewed its focus on yaws eradication in 1996 and has reported no new cases since 2003. In the Americas, foci of yaws are thought to persist in Haiti and other Caribbean islands, Peru, Colombia, Ecuador, Brazil, Guyana, and Surinam. Pinta is limited to Central America and northern South America, where it is found rarely and only in remote villages. Evidence of yaws-like disease and seroreactivity in wild gorillas and baboons in Africa has led to speculation that there may be an animal reservoir for yaws, although strains recently obtained from humans and nonhuman primates have not been subjected to molecular comparison. A single strain isolated from a baboon in 1966 contains several identified genetic differences from available yaws isolates from humans.

Figure 170-1

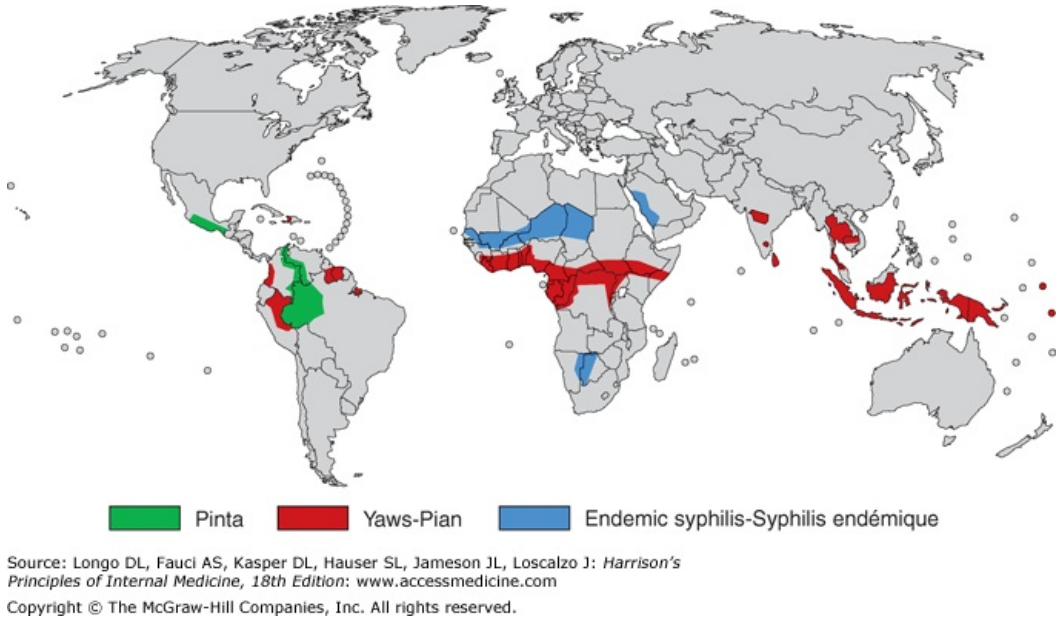

Geographic distribution of endemic treponematoses in the 1990s. *(Courtesy of the World Health Organization; www.who.int/yaws/epidemiology/Map_yaws_90s.jpg.)*

Microbiology

The etiologic agents of the endemic treponematoses are *T. pallidum* subspecies *pertenue* (yaws), *T. pallidum* subspecies *endemicum* (endemic syphilis), and *T. carateum* (pinta). These little-studied organisms are morphologically identical to *T. pallidum* subspecies *pallidum*, and no definitive antigenic differences among them have been identified to date. A controversy has existed about whether the pathogenic treponemes are truly different organisms. Three of the four organisms are classified as subspecies of *T. pallidum*; the fourth (*T. carateum*) remains a separate species simply because no organisms have been available for genetic studies. A number of genetic loci distinguish the agents of venereal and nonvenereal treponemal infections, and molecular signatures (assessed by polymerase chain reaction amplification of *tpr* genes and restriction digestion) can differentiate the individual agents of venereal syphilis, yaws, and bejel. Whether these genetic differences are related to the distinct clinical courses of these diseases has not been determined.

Clinical Features

All of the treponemal infections are chronic and are characterized by defined disease stages, with a localized primary lesion, disseminated secondary lesions, periods of latency, and possible late lesions. Primary and secondary stages are more frequently overlapping in yaws and endemic syphilis than in venereal syphilis, and the late manifestations of pinta are very mild relative to the destructive lesions of the other treponematoses. The current preference is to divide the clinical course of the endemic treponematoses into "early" and "late" stages.

The major clinical distinctions made between venereal syphilis and the nonvenereal infections are the apparent lack of congenital transmission and of central nervous system (CNS) involvement in the nonvenereal infections. It is not known whether these distinctions are entirely accurate. Because of the high degree of genetic relatedness among the organisms, there is little biological reason to think that *T. pallidum* subspecies *endemicum* and *T. pallidum* subspecies *pertenue* would be unable to cross the blood-brain barrier or to invade the placenta. These organisms are like *T. pallidum* subspecies *pallidum* in that they can disseminate from the site of primary infection and can persist for decades. The lack of recognized congenital infection may be due to the fact that childhood infections are often in

the latent stage (low bacterial load) before girls reach sexual maturity. Neurologic involvement may go unrecognized because of the lack of trained medical personnel in endemic regions, the delay of many years between infection and possible CNS manifestations, or a low rate of symptomatic CNS disease. Some published evidence supports congenital transmission as well as cardiovascular, ophthalmologic, and CNS involvement in yaws. Although the reported studies have been small, have failed to control for other causes of CNS abnormalities, have not included specific treponemal serologic tests, and have not analyzed the response to therapy, it may be erroneous to accept unquestioningly the frequently repeated belief that these organisms fail to cause such manifestations.

YAWS

Also known as *pian*, *framboesia*, or *bouba*, yaws is characterized by the development of one or several primary lesions ("mother yaw") followed by multiple disseminated skin lesions. All early skin lesions are infectious and may persist for many months; cutaneousrelapses are common during the first 5 years. Late manifestations, affecting 10% of untreated persons, are destructive and can involve skin, bone, and joints.

The infection is transmitted by direct contact with infectious lesions, often during play or group sleeping, and may be enhanced by disruption of the skin by insect bites or abrasions. After an average of 3–4 weeks, the first lesion begins as a papule—usually on an extremity—and then enlarges (particularly during moist warm weather) to become papillomatous or "raspberry-like" (thus the name "framboesia") **(Fig. 170-2, left)**. Regional lymphadenopathy develops, and the lesion usually heals within 6 months; dissemination is thought to occur during the early weeks of infection. A generalized secondary eruption, accompanied by generalized lymphadenopathy, appears either concurrent with or following the primary lesion, may take several forms (macular, papular, or papillomatous), and may become secondarily infected with other bacteria. Painful papillomatous lesions on the soles of the feet result in a crablike gait ("crab yaws"), and periostitis may result in nocturnal bone pain and polydactylitis. Late yaws is manifested by gummas of the skin and long bone, hyperkeratoses of the palms and soles, osteitis and periostitis, and hydrarthrosis. The late gummatous lesions are characteristically extensive. Destruction of the nose, maxilla, palate, and pharynx is termed *gangosa* and is similar to the destructive lesions seen in leprosy and leishmaniasis.

Figure 170-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical manifestations of endemic treponematoses. ***Left:*** Papillomatous primary lesion of yaws. ***Center:*** Split papules of early endemic syphilis. ***Right:*** Pigmented macules of pinta. *[Photos published with permission from Professor H. Assé, Côte d'Ivoire (left) and from PL Perine et al: Handbook of Endemic Treponematoses, Geneva, World Health Organization, 1984 (center and right).]*

ENDEMIC SYPHILIS

The early lesions of endemic syphilis (*bejel*, *siti*, *dichuchwa*, *njovera*, *skerljevo*) are localized primarily to mucocutaneous and mucosal surfaces. The infection is reported to be transmitted by direct contact, by kissing, or by sharing drinking and eating utensils. A role for insects in transmission has been suggested but is unproven. The initial lesion, usually an intraoral papule **(Fig. 170-2, center)**, often goes unrecognized and is followed by mucous patches on the oral mucosa and mucocutaneous lesions resembling the condylomata lata of secondary syphilis. This eruption may last for months or even years, and treponemes can readily be demonstrated in early lesions. Periostitis and regional lymphadenopathy are common. After a variable period of latency, late manifestations may appear, including osseous and cutaneousgummas. Destructive gummas, osteitis, and gangosa are more common in endemic syphilis than in late yaws.

PINTA

Pinta (*mal del pinto*, *carate*, *azul*, *purupuru*) is the most benign of the treponemal infections. This disease has three stages that are characterized by marked changes in skin color (**Fig. 170-2, right**), but pinta does not appear to cause destructive lesions or to involve other tissues. The initial papule is most often located on the extremities or face and is pruritic. After one to many months of infection, numerous disseminated secondary lesions (*pintides*) appear. These lesions are initially red but become deeply pigmented, ultimately turning a dark slate blue. The secondary lesions are infectious

and highly pruritic and may persist for years. Late pigmented lesions are called *dyschromic macules* and contain treponemes. Over time, most pigmented lesions show varying degrees of depigmentation, becoming brown and eventually white and giving the skin a mottled appearance. White achromic lesions are characteristic of the late stage.

Diagnosis

Diagnosis of the endemic treponematoses is based on clinical manifestations and, when available, dark-field microscopy and serologic testing. The same serologic tests that are used for venereal syphilis (Chap. 169) become reactive during all treponemal infections. Although several targets have been evaluated for specific serodiagnosis, to date there is no test that can discriminate among the differentinfections. The nonvenereal treponemal infections should be considered in the evaluation of a reactive syphilis serology in any person who has emigrated from an endemic area.

Treatment: Endemic Treponematoses

The WHO-recommended therapy for patients and their contacts is benzathine penicillin (1.2 million units IM for adults; 600,000 units for children <10 years old). This dose is half of that recommended for early venereal syphilis, and no controlled efficacy studies have been conducted. Definitive evidence of resistance to penicillin is lacking, although relapsing lesions have been reported after penicillin treatment in Papua New Guinea. Limited data suggest the efficacy of tetracycline for treatment of yaws, but no data exist for other endemic treponematoses. Solely on the basis of experience with venereal syphilis, it is thought that doxycycline or tetracycline (at doses appropriate for syphilis; Chap. 169) are alternatives for patients allergic to penicillin. Macrolide resistance mutations have been identified in *T. pallidum* subspecies *pallidum*, but no data are available on possible mutations in the nonvenereal treponemes. A Jarisch-Herxheimer reaction (Chap. 169) may follow treatment of endemic treponematoses. Nontreponemal serologic titers [in the Venereal Disease Research Laboratory (VDRL) slide test or the rapid plasma reagin (RPR) test] usually decline after effective therapy, but patients may not become seronegative.

Control

Because of lack of ongoing surveillance for the endemic treponematoses, these potentially destructive diseases are not recognized by public health decision-makers and control efforts are rarely undertaken, even though penicillin therapy is inexpensive and effective. There is concern that, as HIV spreads throughout developing countries, it may markedly affect the manifestations and transmission of the endemic treponematoses.

FURTHER READINGS

Antal GM et al: The endemic treponematoses. Microbes Infect 4:83, 2002[PMID: 11825779]

Aseidu K et al: Yaws eradication: Past efforts and future perspectives. Bull WHO 86:499, 2008

Centurion-Lara A et al: Molecular differentiation of *Treponema pallidum* subspecies. J Clin Microbiol 44:3377, 2006[PMID: 16954278]

Gerstl S et al: Prevalence study of yaws in the Democratic Republic of Congo using the lot quality assurance sampling method. PLoS ONE 4:e6338, 2009

Rinaldi A: Yaws: A second (and maybe last?) chance for eradication. PLoS Negl Trop Dis 2:e275, 2008

Satter E, Tokarz Va: Secondary yaws: An endemic treponemal infection. Ped Dermatol 27:364, 2010[PMID: 20653853]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LEPTOSPIROSIS: INTRODUCTION

Leptospirosis is a globally important zoonotic disease caused by spirochetes of the genus *Leptospira* (**Fig. 171-1**). In 1885, Adolf Weil described the clinical hallmarks of this disease as an acute process characterized by splenomegaly, jaundice, and nephritis. With time, the designation *Weil's disease* came to signify severe leptospirosis characterized by diverse clinical findings, particularly fever, jaundice, acute renal injury, refractory shock, and hemorrhage (especially pulmonary hemorrhage). The global burden of leptospirosis is hard to quantify because of the difficulties encountered in its clinical diagnosis and the lack of efficient confirmatory laboratory testing, which limits public health reporting.

Figure 171-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transmission electron micrograph of *Leptospira interrogans* serovar Icterohaemorrhagiae.

Etiologic Agent

The genus *Leptospira* (order Spirochetales, family Leptospiraceae) constitutes the most ancient lineage of spirochetes pathogenic for humans and the only spirochetes that can live both in animals and free in the environment. This genus includes 20 named species, 9 of which are classified as pathogenic, 5 as intermediately pathogenic, and 6 as nonpathogenic (saprophytic) based on molecular phylogenetic analysis (**Fig. 171-2**). Of the pathogenic and intermediate *Leptospira* species, more than 250 serovars—classified on the basis of agglutination testing with specific antisera—cause disease in humans and animals. New species and serovars continue to be discovered. Although all species, serovars, and strains are morphologically identical, leptospires are described by serovar for clinical and epidemiologic reasons.

Figure 171-2

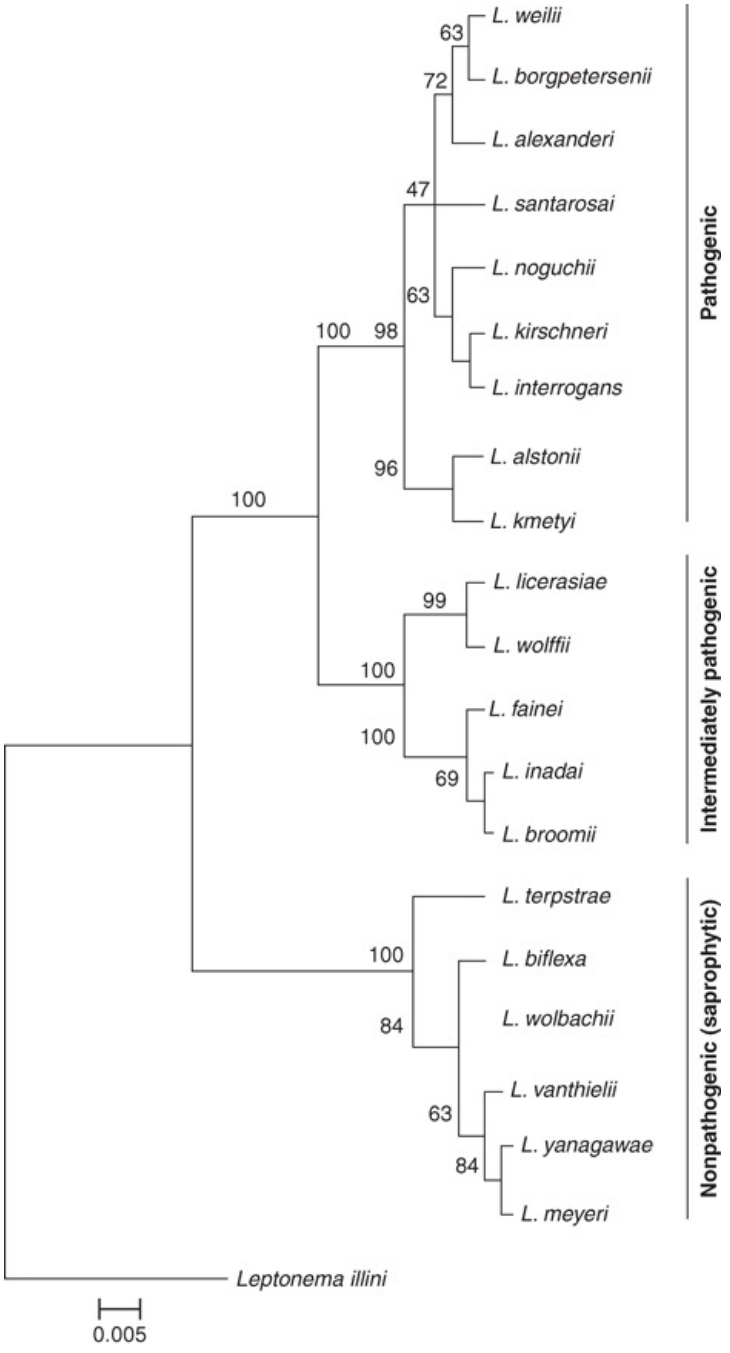

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Differentiation of pathogenic, intermediately pathogenic, and nonpathogenic (saprophytic) *Leptospira* species*based on molecular phylogenetic analysis using the 16S rRNA gene. Scale bar indicates rate of nucleotide substitution per base pair.*

The dimensions and motility of leptospires (~0.1 x 6–20 μm) allow them to pass through filters used to sterilize culture medium. Leptospires are tightly and regularly coiled, with characteristic hooked ends (hence the species name *interrogans*), and are highly motile, spinning around their longitudinal axis and darting back and forth. The organisms cannot be seen by direct light microscopy. To visualize the spirochetes directly in culture or in clinical specimens, dark-field or phase-contrast microscopy must be used. Small protein strands that appear motile by Brownian movement can easily be confused with leptospires. In tissues, leptospires can be visualized by silver impregnation (i.e., Warthin-Starry staining), immunohistochemistry, or immunofluorescence microscopy.

Leptospires are difficult to isolate in pure culture from clinical specimens such as blood, urine, and cerebrospinal fluid (CSF), although certain species and serovars (e.g., *L. interrogans* serovar Copenhageni) are grown more easily than others. The organisms have peculiar nutritional requirements, particularly their inability to ferment glucose and their apparently exclusive use of long-chain fatty acids to generate energy and metabolites for cell division and growth. Standard leptospiral culture medium [Ellinghausen–McCullough–Johnson–Harris (EMJH)] contains oleic acid polymers

(Tween60 and Tween40) as fatty acid sources. These spirochetes do not grow in medium typically used in automated blood culture systems but can be recovered if specimens are subcultured within~1 week onto EMJH, Stuart, Fletcher's, or Korthoff's medium. EMJH, the standard for isolation of *Leptospira* from clinical specimens, is a liquid polysorbate-Tween medium to which 0.1% agar is added (sometimes supplemented with antibiotics to prevent growth of contaminants). The primary isolation of *Leptospira* requires the presence of a solid phase in the medium, which is provided by the agar particles. Cultures are maintained in the dark at 28–30°C and are examined at weekly intervals by dark-field microscopy for up to 3 months. After some weeks, florid growth sometimes produces a dense ring of organisms—Dinger's ring—just under the surface of the medium.

Epidemiology

Leptospirosis is a zoonotic disease. Human-to-human transmission does not occur. Although more than 100 different mammals can be infected, the most important sources of transmission to humans are rats, dogs, cattle, and pigs. Rats do not become ill from leptospiral infection, but dogs often develop severe disease similar to that in humans; infection can cause reproductive failure in cattle and pigs. Even when vaccinated, asymptomatic dogs, cattle, and pigs can be leptospiruric and thus transmit infection to humans. Classic (but not exclusive) serovar-animal associations include Icterohaemorrhagiae and Copenhageni in domestic rats (*Rattus norvegicus, R. rattus*), Grippotyphosa in opossums and raccoons (emerging in the United States and Canada in the absence of a vaccine that covers this serovar), Canicola in dogs, Hardjo in cattle and buffalo, and Pomona in pigs.

Patterns of leptospirosis transmission are characterized as epidemic, endemic, and sporadic. Factors that facilitate human infection are those that bring susceptible persons into indirect contact with contaminated animal urine through surface waters, moist soil, or other wet environments or into direct contact with urine and other excreta (e.g., products of parturition, placenta) of infected animals. In recent years, fewer occupation-related cases and more cases related to environmental exposure have been seen. Seasonal rains and seasonal flooding are the most important factors in the occurrence of epidemic leptospirosis. Tropical humid environments, poor sanitation leading to rodent infestation, and uncontrolled dog populations are important for endemic transmission. Sporadic leptospirosis is associated with human contact with contaminated environments in various settings: on the job (veterinary, sewer, and slaughterhouse workers), in poorly hygienic inner-city alleys and slums, during adventure travel and other non-work-related outdoor activities, and during military training exercises in endemic regions.

Reliable data on the incidence of leptospirosis and on rates of associated morbidity and mortality remain scant and are generally drawn from biased hospital-based series or from governmental registries including passively reported serologic results. In the United States, leptospirosis was removed from the list of notifiable diseases in the 1990s. The ~50–100 cases passively reported annually to the Centers for Disease Control and Prevention (CDC) clearly represent an underestimate; the majority of these cases are from Hawaii, and others are sporadically acquired in inner-city settings or in association with environments such as farms, lakes, and adventure-sport locales. In large urban centers in Brazil that are subject to seasonal flooding (e.g., São Paulo, Rio de Janeiro, and Salvador), tens of thousands of cases are estimated to occur annually. Prospective, population-based cohort studies in Salvador indicate that 5% of slum dwellers are infected annually and that some people are reinfected. Case-fatality rates among hospitalized patients in São Paulo can be as high as ~20% despite state-of-the-art intensive care unit and supportive care. In the Peruvian Amazon, ~30–50% of patients with acute undifferentiated fever have been identified as having leptospirosis; the disease is severe in a small minority of such cases (<2%), and these severe cases are often associated with urban acquisition of infection. Men are affected more often by clinical disease than are women.

Infection by *Leptospira* does not occur via inhalation, and leptospirosis is a rare cause of laboratory-acquired infection. Laboratory strains usually used for serologic diagnosis have been serially passaged for long periods and usually have lost their virulence.

GLOBAL FEATURES

Leptospirosis affects urban and rural populations in industrialized and developing countries alike. The highest burden of disease falls upon those at lower socioeconomic levels whose activities of daily living bring them into contact with surface waters contaminated with the urine of animals carrying the infection. High rates of endemic leptospirosis, associated with both mild and severe disease, are found throughout tropical regions. Torrential seasonal flooding in areas of high population density (e.g., in urban slums in Brazil, India, and Thailand) is the major risk factor for epidemic severe disease. In 2009, outbreaks after typhoons in the Philippines affected large populations, prompting the Ministry of Health to provide antimicrobial chemoprophylaxis to millions of people; the efficacy of this intervention remains unknown.

Military training, outdoor athletic activities, and adventure travel have led to recognized outbreaks and sporadic cases of leptospirosis. Now-classic examples include the frequent occurrence of leptospirosis in U.S. soldiers undergoing jungle training in Panama (in which context the first clinical trial of antibiotic prophylaxis was carried out), whitewater rafters in Costa Rica, and almost half of the participants in

Eco-Challenge Sabah in Borneo, Malaysia, in 2000.

Sporadic leptospirosis, which is likely to be common but underrecognized in urban and rural settings, is generally identified when the manifestations of disease are severe and the index of suspicion is high.

Pathogenesis

Leptospires infect humans through the mucosa (usually conjunctival and possibly oral or tonsillar) or through macerated, punctured, or abraded skin. The organisms resist innate immune defenses (e.g., complement), proliferate in the bloodstream or extracellularly within organs, and then disseminate hematogenously to all organs. The incubation period averages 5–14 days (range, 2–30 days); leptospires can be isolated from blood during the first 3–10 days of clinical illness (**Fig. 171-3**). As antibodies develop, leptospires disappear from the blood but persist in various organs, including brain (the meninges and possibly other sites), liver, lung, heart, and kidney. The life cycle is completed as leptospires traverse the interstitial spaces of the kidney, penetrate the basement membrane of the proximal renal tubules, cross through proximal renal tubuleepithelial cells, and become adherent to the proximal renal tubular brush border, whence they are excreted in the urine. In humans, as in other mammalian hosts, chronic and persistent renal colonization can last for weeks or years, with unknown pathophysiologic consequences.

Figure 171-3

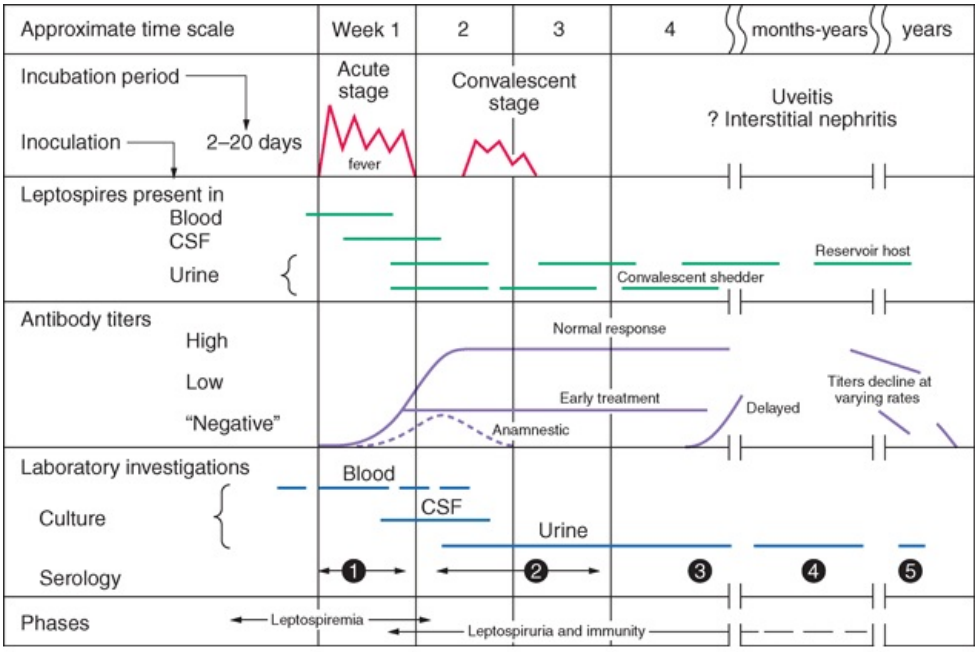

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Biphasic nature of leptospirosis and relevant investigations at different stages of disease. Specimens 1 and 2 for serology are acute-phase serum samples; specimen 3 is a convalescent-phase serum sample that may facilitate detection of a delayed immune response; and specimens 4 and 5 are follow-up serum samples that can provide epidemiologic information, such as the presumptive infecting serogroup. *[Reprinted as adapted by Levitt (from Turner LH: Leptospirosis. BMJ 1:231, 1969) with permission from the American Society for Microbiology and the BMJ Publishing Group.*]

Although severe human disease due to a wide variety of leptospires has been reported, some of the leptospires involved are thought to be more intrinsically pathogenic than others. Rat-associated *L. interrogans* serovars Icterohaemorrhagiae and Copenhageni are mostly commonly associated with Weil's disease; jaundice, renal failure, shock, and hemorrhage due to other species and serovars have been reported as well. Specific molecular determinants and virulencemechanisms responsible for disease manifestations have not been identified. An unusual lipid A structure renders the leptospiral lipopolysaccharide of low endotoxic potential in experimental systems. Multiple in vitro studies have shown that leptospires and their extracts cause cellular toxicity; however, the biochemical nature of damage to host cells and the underlying mechanisms remain unclear.

Pathologic findings are organ specific. Acute and chronic inflammation within the kidney is associated with acute tubular necrosis and interstitial nephritis. Autopsy studies have revealed abnormal regulation of fluid and electrolyte transporters—including the endogenous sodium/hydrogen exchanger isoform 3 (NHE 3), aquaporins 1 and 2, α-Na+K+ATPase, and sodium–potassium–chloride

cotransporter (NKCC2 isoform)—in both the presence and the absence of acute tubular necrosis. Primary injury of the proximal convoluted tubules is the primary renal pathophysiologic lesion in acute leptospirosis, with secondary increased distal tubular potassium excretion, hypokalemia, and polyuria.

Hepatic histopathology in fatal cases is associated with disruption of cellular cohesion, plugging of bile canaliculi, occasional acute inflammatory infiltrates, and focal periportal cellular necrosis and steatosis; widespread hepatocellular necrosis is not found. Cases of severe pulmonary hemorrhage syndrome that come to autopsy are characterized by the absence of inflammation, the paucity of organisms visible by silver or immunohistochemical staining, and grossly obvious frank hemorrhage (**Fig. 171-4**). Ultrastructural features of the lung in the few fatal cases reported include swelling, increased pinocytotic vesicles, and giant dense bodies in the cytoplasm of epithelial cells; these abnormalities are limited to hemorrhagic areas, and the intercellular junctions are preserved. Platelets are adherent to activated but abnormal-appearing endothelial cells. Septal capillary lesions seem to be causally related to death, leading to pulmonary hemorrhage; in both animal models and human cases, immunoglobulin and complement deposition have been demonstrated in lung tissue involved in leptospiral pulmonary hemorrhage.

Figure 171-4

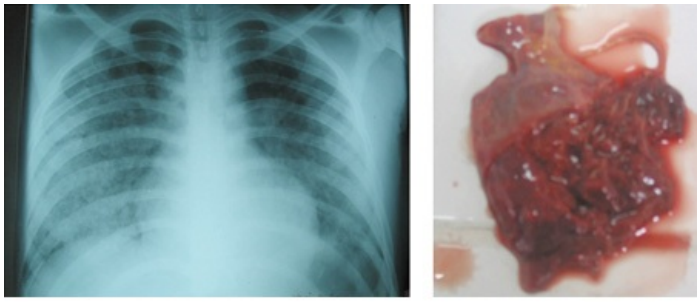

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe pulmonary hemorrhage in leptospirosis. ***Left:*** Chest x-ray. ***Right:*** Gross appearance of right lower lobe of lung at autopsy. This patient, a 15-year-old in the Peruvian Amazonian city of Iquitos, died several days after presentation with acute illness, jaundice, and hemoptysis. Blood culture yielded *Leptospira interrogans* serovar Copenhageni/Icterohaemorrhagiae. (*Adapted with permission from Segura et al, 2005. ©2005 by the Infectious Diseases Society of America.*)

The relation of disseminated intravascular coagulation to leptospirosis has long been debated. The prothrombin and activated partial thromboplastin times are not necessarily elevated in severe leptospirosis, and fibrinogen levels are typically elevated. Thrombocytopenia is characteristic, probably reflecting platelet consumption in the activated endothelial surface; platelet counts are lower in severe than in mild leptospirosis. Nonetheless, proteolytic products of fibrinogen (e.g., d-dimers), thrombin–antithrombin III complexes, and prothrombin fragment 1,2 have been found to be elevated in cases of leptospirosis in Thailand (but with no difference between severe and nonsevere cases), a finding indicating pathologic activation of the coagulation system in leptospirosis.

In the heart, pericardial and endocardial hemorrhage, disruption of myocardial fiber organization, and scattered myocyte necrosis (accompanied, grossly, by dilation of both right and left ventricles) are pathological lesions associated with severe leptospirosis.

Despite the traditional view that leptospirosis is characterized by vasculitis, formal demonstration of inflammatory infiltrates within any blood vessel has not been shown to be involved in the pathogenesis of this disease. A more likely possibility is that leptospires induce endothelial cell dysfunction with organ dysfunction and systemic disease, but this hypothesis remains to be validated.

Clinical Manifestations

The clinical expression of infection by *Leptospira*, which is related to diverse focal organ dysfunction, includes subclinical infection, an undifferentiated febrile illness, and Weil's disease—the most severe form. Leptospirosis is classically described as biphasic (Fig. 171-3). Acute fever in the initial leptospiremic phase lasts for 3–10 days, during which period the organism may be cultured from blood. In a later immune phase, fever is not responsive to antibiotic therapy but leptospires can be isolated from urine. Unlike milder forms, Weil's disease may also be monophasic and fulminant.

Physical examination may include any of the following findings, none of which is pathognomonic for leptospirosis: conjunctival suffusion (dilated conjunctival blood vessels in the absence of discharge); pharyngeal erythema without exudate; muscle tenderness; rales on lung auscultation or dullness on chest percussion over areas of pleural hemorrhage; rash (which may be macular, maculopapular, erythematous, petechial, or ecchymotic); jaundice; meningismus; and hypo- or areflexia, particularly in

the legs.

The natural course of mild uncomplicated leptospirosis usually ends in spontaneous resolution within 7–10 days without sequelae, but the difficulties of rapid diagnosis do not permit the initiation of specific antimicrobial therapy. Biomarkers to predict progression to severe disease are not available. Some patients experience a return of fever, headache, and other systemic symptoms after 3–10 days (the immune phase) in association with the clearance of leptospires from the blood and the appearance of antibodies; this phase of illness does not respond to antibiotic therapy.

Weil's disease is characterized by variable combinations of jaundice, acute kidney injury, hypotension, and hemorrhage—mostcommonly involving the lungs (Fig. 171-4) but also potentially affecting the gastrointestinal tract, retroperitoneum, pericardium, and brain. Other syndromes include aseptic meningitis, uveitis, cholecystitis, acute abdomen, and pancreatitis (with hypo- or hyperglycemia). Jaundice is not associated with fulminant hepatic necrosis or hepatocellular damage, but rather with abnormal laboratory values (see "Diagnosis," below). The liver can be enlarged and tender; splenomegaly is reported in a minority of cases. Acute kidney injury manifests after several days of illness and can be nonoliguric or oliguric, with serum electrolyte abnormalities reflecting proximal renal tubular dysfunction. Hypokalemia and hypomagnesemia are common in nonoliguric renal failure; hypomagnesemia can cause severe muscle weakness. Hypotension is associated with acute tubular necrosis and oliguria, requiring fluid resuscitation, sometimes pressors, and hemodialysis. Renal function typically returns to normal in survivors of severe disease. Severe pulmonary hemorrhage in leptospirosis is a clinical problem wherever the disease is endemic, manifesting with cough, chest pain, and hemoptysis but without purulent sputum.

Cardiac involvement is commonly reflected on the electrocardiogram as nonspecific ST and T wave changes but also asright-bundle-branch block and right- and/or left-sided ventricular dilation indicating myocarditis. Skeletal muscle involvement manifests as severe myalgia, typically of the legs (especially the calves) but also of the abdominal muscles (mimicking acute abdomen); these symptoms are associated with a moderately elevated serum concentration of creatine kinase that, by itself, is not sufficient to result in acute kidney injury. Skin abnormalities include petechiae and ecchymosis as well as macular and maculopapular rash. Neurologic findings include aseptic meningitis (in which CSF pleocytosis can range from a few cells to >1000 cells/μL, with a polymorphonuclear cell predominance) and hypo- or areflexia, especially of the legs.

Diagnosis

Leptospirosis should be suspected on the basis of an appropriate exposure history combined with any of the infection's protean manifestations. A high index of suspicion prompting elicitation of a detailed exposure history is critical and guides confirmatory testing (see below). Leptospiral infection is usually associated with obvious exposure events that go beyond casual activities such as simply having been on a farm or in an urban alley. An infected person usually has been immersed in or has had mucosal or percutaneous exposure to contaminated animal urine. Leptospiral infection resulting from the bite of a rat or another animal is rare.

Biochemical, hematologic, and urinalysis findings in acute leptospirosis are not specific, but certain patterns suggest the diagnosis. In the context of an appropriate exposure history and in the absence of a more likely explanation, classic Weil's disease is suggested by elevated levels of blood urea nitrogen and serum creatinine in conjunction with mixed conjugated and unconjugated hyperbilirubinemia with aminotransferase elevation to less than five times the upper limit of normal. In all forms of leptospirosis (not just Weil's disease), a variety of abnormalities can occur. Urinalysis may show abnormalities of the sediment (leukocytes, erythrocytes, hyaline and granular cases). Elevation of the noncardiac isoform of creatine kinase may indicate skeletal muscle damage. Troponin levels indicative of myocarditis have not been adequately studied in leptospirosis. Hematologic abnormalities are variable but common: leukocytosis (typical in severe disease), leukopenia, hemolytic anemia, mild to moderate anemia, and thrombocytopenia.

On chest radiography, the appearance of the lungs varies. Alveolar infiltrates predominate and are associated with hemoptysis but not with purulent sputum. Other findings include diffuse interstitial infiltrate patterns corresponding to hyaline membrane disease (acute respiratory distress syndrome) and small nodular infiltrates and pleural-based densities representing hemorrhage.

The confirmation of leptospirosis requires laboratory testing. Definitive diagnosis rests on demonstrating the presence of the organism by culture isolation, detection of nucleic acids or antigen in body fluids, or immunohistochemical visualization in tissue. Direct examination of urine or blood by dark-field microscopy has the potential to provide a rapid diagnosis but is not recommended because of complicating artifacts. Leptospiral cultures do not become positive for weeks and therefore cannot guide clinical care. Polymerase chain reaction–based assays have been used in research laboratories to detect leptospiral DNA but are not clinically available. Moreover, a negative result does not rule out the diagnosis: the assays' sensitivity is insufficient when the level of bacteremia is below a detectable threshold, and inhibitors in blood and urine can interfere with these tests.

Serologic assays are the diagnostic mainstay in leptospirosis. In the United States, the gold standard

—the microscopic agglutination test (MAT)—is performed only at the CDC; likewise, in other countries, performance of the MAT is generally limited to reference laboratories. The MAT entails growth of a battery of serovars representing the 26 leptospiral serogroups, incubation of a standard quantity of leptospires with the patient's serum on a microtiter plate, and detection of agglutination by dark-field microscopy. The highest dilution of serum that yields significant (50%) agglutination is reported as the titer. Although antibody titers are reported by serovar, a positive MAT result reflects the presence only of Leptospira-specific antibodies and cannot be used to precisely identify the infecting serovar because one serovar may induce antibodies that cross-react with other serovars. When patients have a high pretest probability of leptospirosis, a single antibody titer >1:200 is considered strong evidence of infection; however, in regions where leptospirosis transmission and subclinical disease are common, higher titers are generally required for a confident diagnosis because of long-lasting antibodies after a previous infection. Because the MAT is generally negative in the first 7–10 days after the onset of infection, paired acute- and convalescent-phase serum samples are preferred to document seroconversion or a fourfold rise in titer.

Other serologic tests for leptospirosis (e.g., enzyme-linked immunosorbent assay, indirect hemagglutination, dot-blot, and lateral flow) are based on solid-phase assays; some of these tests are commercially available. Such assays use the saprophytic (non-disease-associated) leptospire *L. biflexa* as antigen. *L. biflexa* has lipopolysaccharide epitopes in common with pathogenic *Leptospira* species. While useful in some regions of the world, *L. biflexa*–based tests are relatively insensitive (because of regional variation of leptospires) and nonspecific (because of previous exposure) and must be interpreted with caution.

Leptospires can be cultured from blood and CSF during the first 7–10 days of illness and from urine beginning in the second week. Cultures usually become positive after 2–4 weeks (range, 1 week to 6 months). Urine cultures can remain positive for months or years despite antibiotic therapy. Because leptospires can remain viable for as long as ~10 days at room temperature, specimens can be shipped to reference laboratories in anticoagulated blood (heparin, citrate, EDTA). Although leptospires do not grow in medium used in automated blood culture detection systems, inoculated bottles can be subcultured into leptospiral culture medium within 1 week after inoculation for attempted isolation.

Differential Diagnosis

The differential diagnosis of acute leptospirosis is broad, reflecting its diverse clinical presentations and depending on the patient's travel and exposure history and geographic region of presentation. When fever and severe myalgia predominate, influenza is often considered, although the absence of coryza, sore throat, and cough is not consistent with this diagnosis. Other important possibilities include malaria, rickettsial diseases, arboviral infections (e.g., dengue and chikungunya), typhoid fever, hantavirus infection (hemorrhagic fever with renal syndrome or hantavirus cardiopulmonary syndrome), and viral hepatitis.

Treatment: Leptospirosis

Although the value of antimicrobial treatment (**Table 171-1**) has not been proven in clinical trials, its prompt initiation probably shortens the course of severe leptospirosis and prevents the progression of mild disease. Despite debates about efficacy, antimicrobial drugs (typically penicillin, ceftriaxone, or cefotaxime) should be used to treat severe later-stage leptospirosis. Mild leptospirosis often is not specifically identified and typically resolves without antibiotic treatment. If clinical suspicion is high or the diagnosis is suggested or confirmed by laboratory findings in an appropriate context (e.g., clinical presentation, exposure history), mild disease should be treated with oral antibiotics—in particular, doxycycline, especially where rickettsial infections (including scrub typhus) are coendemic. Solid data from studies of animals indicate that oral azithromycin is also likely to be useful in mild leptospirosis. Like many acute bacterial diseases manifesting as multiorgan system dysfunction, severe leptospirosis frequently requires empirical initiation of broad-spectrumparenteral therapy before the diagnosis can be confirmed.

Table 171-1 Treatment and Chemoprophylaxis of Leptospirosis in Adults [a]

Indication	Regimen
Treatment	
Mild leptospirosis	Doxycycline (100 mg PO bid) *or* Amoxicillin (500 mg PO tid) *or* Ampicillin (500 mg PO tid)
Moderate/severe leptospirosis	Penicillin (1.5 million units IV or IM q6h) *or* Ceftriaxone (1 g/d IV) *or* Cefotaxime (1 g IV q6h)

Chemoprophylaxis[b]	
	Doxycycline (200 mg PO once a week) *or* Azithromycin (250 mg PO once or twice a week)

[a] All regimens are given for 7 days.

[b] The efficacy of doxycycline prophylaxis in endemic or epidemic settings remains unclear. Experiments in animal models and a cost-effectiveness model indicate that azithromycin has a number of characteristics that may make it efficacious in treatment and prophylaxis, but clinical trials have not been performed.

In rare instances, acute decompensation after the initiation of antimicrobial therapy occurs in association with a Jarisch-Herxheimer response and should be managed supportively. Fresh-frozen plasma, plasmapheresis, glucocorticoids, and activated protein C have no demonstrated role in the treatment of leptospirosis. Anecdotal reports in which inhaled nitric oxide was successfully used by leptospirosis patients with severe pulmonary involvement must be confirmed before such treatment can be recommended.

Prognosis

The severity of illness in terms of pulmonary and renal dysfunction is the most important determinant of prognosis. Advanced age, clinically evident pulmonary involvement, elevated serum creatinine level, oliguria, and thrombocytopenia are associated with a poor prognosis; liver dysfunction in acute leptospirosis has not been confirmed to be an independent risk factor for death. Chronic alcoholism seems to be associated with severe disease. Reported mortality rates among hospitalized patients have varied from <5% to >20%.

Leptospirosis is generally considered to leave no permanent sequelae, although renal dysfunction, as manifested by electrolyte imbalances, may persist for days or weeks after acute illness resolves. Severe pulmonary hemorrhage and liver disease are not known to lead to persistent or progressive organ dysfunction. Some authorities have suggested that neuropsychiatric disturbance may follow severe disease; such conclusions must be assessed in a prospective clinical investigation.

Prevention

No vaccine is available for human leptospirosis. Preventive strategies, including prophylaxis with doxycycline, have been variably effective in different settings. Antibiotic prophylaxis can be considered for anticipated short-term, well-defined exposures, such as those incurred during military training or specific adventure travel (with, for example, fresh-water swimming). Long-term antibiotic prophylaxis has not been shown to be effective in preventing infection in high-transmission endemic settings. General sanitation approaches (e.g., rodent control) and avoidance of swimming in potentially contaminated places (e.g., for recreational use) are recommended, but these measures are difficult to apply consistently.

FURTHER READINGS

Araujo ER et al: Acute kidney injury in human leptospirosis: An immunohistochemical study with pathophysiological correlation. Virchows Arch 456:367, 2010[PMID: 20217429]

Bharti AR et al: Leptospirosis: A zoonotic disease of global importance. Lancet Infect Dis 3:757, 2003[PMID: 14652202]

Faine S et al (eds): *Leptospira and Leptospirosis.* Melbourne, MedScience, 1999

Ganoza CA et al: Determining risk for severe leptospirosis by molecular analysis of environmental surface waters for pathogenic *Leptospira*. PLoS Med 3:E308, 2006

Gouveia EL et al: Leptospirosis-associated severe pulmonary hemorrhagic syndrome, Salvador, Brazil. Emerg Infect Dis 14:505, 2008[PMID: 18325275]

Ko AI et al: *Leptospira*: The dawn of the molecular genetics era for an emerging zoonotic pathogen. Nat Rev Microbiol 7:736, 2009[PMID: 19756012]

——— et al: Urban epidemic of severe leptospirosis in Brazil. Lancet 354:820, 1999

Medeiros FD et al: Leptospirosis-associated disturbances of blood vessels, lungs and hemostasis. Acta Trop 115:155, 2010

Segura E et al: Clinical spectrum of pulmonary involvement in leptospirosis in an endemic region, with

quantification of leptospiral burden. Clin Infect Dis 40:343, 2005[PMID: 15668855]

Vinetz JM et al: Sporadic urban leptospirosis. Ann Intern Med 125:794, 1996[PMID: 8928985]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

RELAPSING FEVER: INTRODUCTION

Relapsing fever is an illness characterized by recurring episodes of fever and nonspecific symptoms (e.g., headache, myalgia, arthralgia, shaking chills, and abdominal symptoms) after infection with one of several species of *Borrelia*.

GLOBAL FEATURES

In North America, relapsing fever (a zoonosis) is transmitted by the bite of an *Ornithodoros* tick. In many other parts of the world, including Africa and Asia, relapsing fever is endemic and occurs after the bite of a tick or the human body louse (*Pediculus humanus*). Tick-borne relapsing fever (TBRF) is also reported from countries in the Middle East, including Israel, Iran, and Jordan. Louse-borne relapsing fever (LBRF) is occasionally imported into the United States by a traveler. TBRF is rarely fatal in North America, where it is most often sporadic; in some African countries (e.g., Senegal and Tanzania), TBRF is a more significant bacterial infection, causing morbidity and death. Conditions that favor infestation with *P. humanus*, such as living in refugee camps or other stressful situations in which many people are crowded together without access to good hygiene and nutrition, have led to large outbreaks of LBRF with substantial rates of morbidity and death; thus, LBRF is also known as *epidemic relapsing fever* .

ETIOLOGY AND PATHOGENESIS

The borreliae are helical or wavy motile spirochetes whose length ranges from 3 to 25 μm and whose width is usually 0.2–0.3 μm. In fixed Wright-stained differential smears, the organisms appear as loose coils (**Fig. 172-1**). Borreliae are transmitted to humans by exposure to the bite of an infected *Ornithodoros* tick (TBRF) or to the hemolymph of an infected human body louse, which may be found on clothing (LBRF). For louse-borne disease, it is not the louse's bite that causes transmission; rather, spirochetes are introduced when the louse is crushed (e.g., by scratching) and the insect's infected hemolymph is released and contaminates abraded or normal skin and mucous membranes. Relapsing fever results when variation in borrelial surface antigens leads to repeated bacteremia and stimulation of the immune system by each new antigen. Each time the organism changes its surface antigens, thus evading the immune system, another febrile response occurs. LBRF is caused by *Borrelia recurrentis*, whereas TBRF is caused by a variety of *Borrelia* species whose names sometimes correspond to their tick vectors. For example, *B. hermsii* is transmitted by the tick *O. hermsi*, and *B. turicatae* is transmitted by *O. turicata* (**Fig. 172-2**).

Figure 172-1

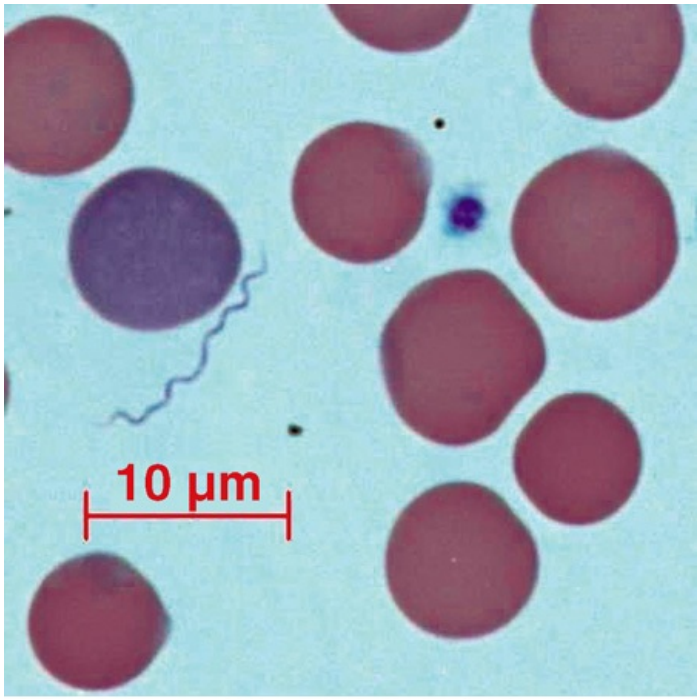

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Photomicrograph of tick-borne relapsing fever spirochete (*Borrelia hermsii*) in a Wright-Giemsa–stained peripheral blood film. *[Reprinted with permission from Dennis DT: Relapsing fever. in Harrison's Principles of Internal Medicine, 17th ed, AF Fauci et al (eds). New York, McGraw-Hill, 2008, p 1054.]*

Figure 172-2

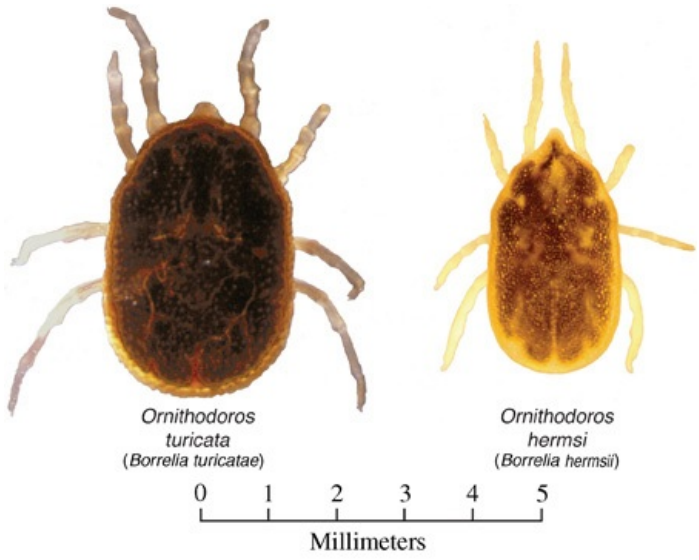

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ornithodoros turicata* and *O. hermsi, two of the many species of blood-feeding soft ticks responsible for transmitting tick-borne relapsing fever.

(Reprinted from Dworkin et al., 2008, with permission from Elsevier.)

EPIDEMIOLOGY

TBRF is endemic in the western United States, southern British Columbia, the plateau regions of Mexico, Central and South America, the Mediterranean, Central Asia, and throughout much of Africa. In the United States, it typically is not reported farther east than Montana, Colorado, New Mexico, and Texas, although cases have been acquired in Oklahoma, Kansas, and Ohio and one case has been

reported from Wyoming. A relapsing fever spirochete has been described as far east as Florida, although human infections have not yet been reported in that state. In the United States, exposure sites typically are forested areas at various elevations in mountainous regions (the Cascade, Rocky Mountain, San Bernardino, and Sierra Nevada ranges) and limestone caves in central Texas. Caves may likewise be an important source of TBRF in other areas of the world, such as Israel and Jordan. Houses, cabins, and cowsheds have been implicated as sources of infection because of tick-infested rodent nesting. The most common vector in the United States, *O. hermsi,* is often found in coniferous forests at elevations of 1500–8000 ft, where it feeds primarily on ground squirrels, tree squirrels, and chipmunks dwelling near freshwater lakes that attract humans who may live in or rent nearby cabins. The disease tends to be most common where humans come into contact with diurnal rodents and their ticks. Only 13 counties have accounted for ~50% of all U.S. cases. Surveillance of TBRF is not performed in all states where the disease is endemic, and substantial underreporting is likely where the disease is reportable. Therefore, the precise distribution of disease is not known.

Many cases of TBRF in West Africa have been attributed to infection with *B. crocidurae* transmitted by the tick *O. sonrai*. In Senegal, this disease has been identified as the most common bacterial infection causing febrile illness. Thus this infection is an important factor in the differential diagnosis of suspected malaria in West Africa, where LBRF has not been described in many decades. Co-infection with *Plasmodium* species was reported in more than one-third of blood films from Senegalese TBRF patients. In eastern sub-Saharan Africa, *B. duttonii* is more prevalent. TBRF has been detected (albeit less commonly) in northern Africa (e.g., in Morocco), where *B. hispanica* and *B. crocidurae* have been identified.

The epidemiology of LBRF is not as well characterized as that of TBRF, probably in part because of the higher prevalence of the former in regions with relatively few resources for communicable disease surveillance. Historically, LBRF has been described in North America and Europe, but it is now only uncommonly reported in these regions. LBRF is relatively well described in East Africa. Outbreaks have been reported in Sudan and Ethiopia. Reports of disease in the highlands of Ethiopia have included many documented cases, despite a recent decline; in that country, more cases have occurred in male than in female patients. Seasonality has not been reported consistently.

Approach to the Patient: Relapsing Fever

Recurring febrile episodes in a patient who lives in or has recently traveled to a geographic location where relapsing fever is known to occur should lead to consideration of relapsing fever. Diagnosis has sometimes been delayed because of a focus on symptoms other than the recurrence of fever, such as diarrhea, thrombocytopenia, and cranial nerve palsy. Failure to diagnose relapsing fever promptly can lead to prolonged (untreated) illness and can incur excessive (preventable) medical costs. One patient may present with illness resembling meningitis, another with illness resembling influenza, and others with a febrile gastrointestinal illness or no physical findings. In all these instances, the patient may undergo a variety of unnecessary invasive and noninvasive tests. In addition, some patients may have more than one diagnosis (e.g., LBRF and typhus); therefore, other local tick-borne or louse-borne diseases should be kept in mind during evaluation.

CLINICAL MANIFESTATIONS

The mean incubation period is 7 days for TBRF (range, 4–18 days or longer) and 8 days for LBRF (range, 5–15 days; sometimes a shorter period in North Africa). Regardless of the tick or louse vector, the clinical manifestations of relapsing fever are similar, although not identical. The signs and symptoms documented in a large series of cases of TBRF are listed with their frequencies in **Table 172-1**. Alteration of sensorium, abdominal pain, and vomiting are common. Diarrhea may develop in 25% of cases. Jaundice; central nervous system (CNS) involvement; petechiae on the trunk, extremities, and mucous membranes; epistaxis; and blood-tinged sputum are more likely in LBRF. Uncommon manifestations of relapsing fever include iritis, acute respiratory distress syndrome, uveitis, iridocyclitis, myocarditis, and splenic rupture. Cranial nerve palsy and other neurologic manifestations are often reversible. Neurologic findings may occur in 10–30% of cases and are more common in LBRF. These findings may include signs of meningitis with or without cerebrospinal fluid abnormalities, seizure, focal deficits, hemiplegia, paraplegia, paresthesias, psychosis, hallucinations, and delirium. Certain species of tick-borne *Borrelia* have been reported with particular frequency in cases with neurologic complications (*B. duttonii* and *B. turicatae*).

Table 172-1 Manifestations of Tick-Borne Relapsing Fever Acquired in the Northwestern United States and Southwestern British Columbia

Sign or Symptom	%	Sign or Symptom	%
Headache	94	Photophobia	25
Myalgia	92	Neck pain	24

Chills	88	Rash	18
Nausea	76	Dysuria	13
Arthralgia	73	Jaundice	10
Vomiting	71	Hepatomegaly	10
Abdominal pain	44	Splenomegaly	6
Confusion	38	Conjunctival injection	5
Dry cough	27	Eschar	2
Eye pain	26	Meningitis	2
Diarrhea	25	Nuchal rigidity	2
Dizziness	25		

Source: From a review of 182 cases reported during 1980–1995 (Dworkin et al. ©1998 Clinical Infectious Diseases).

The average duration of the first episode of TBRF is 3 days (range, 12 h to 17 days), and the episode terminates in a crisis. In contrast, the average duration of the first episode of LBRF is 5.5 days (range, 4–10 days). Subsequent relapsing febrile episodes are typically of shorter duration. The average time between the first episode and the first relapse is 7 days for TBRF and 9 days for LBRF. During afebrile intervals, the patient may have symptoms (e.g., malaise) or may feel well.

The differential diagnosis of infectious diseases causing fevers that may relapse or have biphasic patterns includes but is not limited to Colorado tick fever, yellow fever, dengue fever, lymphocytic choriomeningitis, brucellosis, malaria, leptospirosis, chronic meningococcemia, rat-bite fever, and infection with echovirus 9 or *Bartonella* species. The many other diagnoses that may overlap with relapsing fever in terms of other manifestations include typhus and typhoid fever. A history of travel, place of residence, and animal exposures is useful when patients have these fever patterns.

DIAGNOSIS

DETECTION AND ISOLATION OF SPIROCHETES

During asymptomatic intervals, relapsing fever borreliae are undetectable in the bloodstream by microscopy. Laboratory confirmation is made by the detection or isolation of spirochetes from blood during a febrile episode. Spirochetal counts may be high in the blood. Typically, an average of five organisms per oil-immersion field are observed in routine differential fixed smears of blood obtained from patients during the acute febrile phase of illness. A thin smear or a thick drop of blood is applied to a standard glass microscope slide, stained with Wright or Giemsa, and examined with a bright-field microscope at 1000x with oil immersion. Spirochetes also may be visualized by direct or indirect immunofluorescent staining and fluorescence microscopy. A dark-field microscope may be used to observe spirochetes in the blood. However, microscopic observation of spirochetes is relatively insensitive. Quantitative buffy coat analysis is an alternative method. Other available methods are most often used in research settings. Polymerase chain reaction (PCR) and monoclonal antibody can be used to determine the species of *Borrelia*. For laboratories with PCR technology and expertise, this method is more sensitive than microscopy.

SEROLOGY

Serologic confirmation of TBRF is demonstrated by a fourfold rise in antibody titer between acute- and convalescent-phase serum samples or by the diagnostic reactivity of a single convalescent-phase serum sample. However, serology may not be reliable because of lack of standardization. Patients infected previously with other species of spirochetes may have false-positive reactions in the enzyme-linked immunosorbent assay (ELISA) and the indirect fluorescence antibody (IFA) assay. The most reliable method—the recombinant GlpQ assay—is not widely available. Therefore, blood samples are often screened with ELISA or IFA, and, if the result is positive (e.g., an IFA titer of 1:128–1:256 or higher), an immunoblot can be performed to determine the pattern of reactivity. This procedure has led to recognition of patients erroneously diagnosed with Lyme disease who actually have TBRF.

Treatment: Relapsing Fever

Treatment options for adults with relapsing fever are summarized in **Fig. 172-3**. Relapsing fever spirochetes are commonly sensitive to antibiotics such as doxycycline and erythromycin. Single-dose therapy is generally recommended for LBRF, while a 7-day (or 10-day) course is usually used for TBRF. Insufficient information is available on the efficacy of single-dose therapy for TBRF.

Figure 172-3

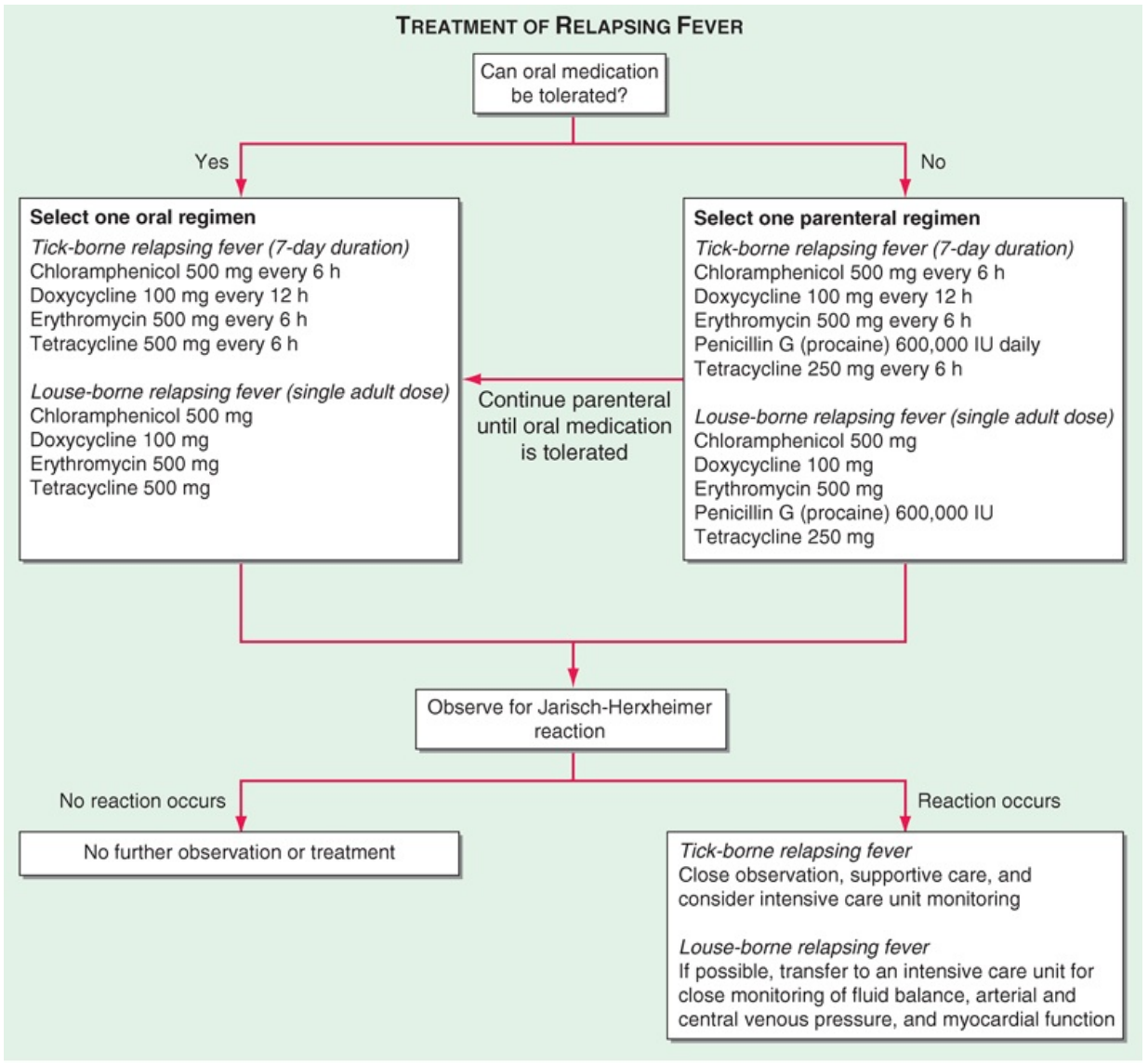

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment of relapsing fever and approach to the Jarisch-Herxheimer reaction.

Lumbar puncture should be considered when signs of meningitis or encephalitis are present; cerebrospinal fluid evidence of these neurologic conditions suggests the need for treatment with an IV antibiotic regimen. This issue is especially relevant in TBRF due to certain species such as *B. duttonii* in Africa and *B. turicatae* in the southwestern United States because these species are more prone to invade the CNS and can reenter the bloodstream after antibiotic treatment if a regimen with good CNS penetration is not used.

Children <8 years of age and pregnant women with relapsing fever should be treated with penicillin or erythromycin. When a Jarisch-Herxheimer reaction occurs (and its occurrence is unpredictable), it may be milder in children than in adults. Monitoring of patients for this reaction for the first 12 h after the first dose of antibiotic is recommended (see below).

COMPLICATIONS

Moderate to severe thrombocytopenia, although not associated with a fatal outcome, is a typical finding in acute TBRF. Bleeding complications, such as epistaxis, purpura, hemoptysis, hematemesis, bloody diarrhea, hematuria, subarachnoid and cerebral hemorrhages, splenic rupture, and retinal hemorrhage, are more common with LBRF. Although death from TBRF in North America is rare, this infection has been associated with complications during pregnancy, including spontaneous abortion, premature birth, or neonatal death.

The Jarisch-Herxheimer reaction is an acute exacerbation of symptoms that may occur on initial treatment of relapsing fever with an effective antibiotic. During this reaction, spirochetes disappear

rapidly from the circulation and there is massive cytokine release. The likelihood of a Jarisch-Herxheimer reaction is not reliably predictable in TBRF; the reaction is common in LBRF treated with tetracycline. Treatment with penicillin rapidly alters the morphology of the dividing spirochetes, making them susceptible to phagocytosis. Symptoms of the Jarisch-Herxheimer reaction often include hypotension, tachycardia, chills, rigors, diaphoresis, and marked elevation of body temperature. The reaction typically begins within 1–4 h of the first dose of antibiotic, and the symptoms may be extremely severe. When possible, patients with LBRF who develop the Jarisch-Herxheimer reaction should be transferred to an intensive care unit for close monitoring of fluid balance, arterial and central venous pressure, and myocardial function. Death (most often secondary to cardiovascular collapse) has been reported as a complication of the reaction in patients with LBRF. Patients with TBRF and the Jarisch-Herxheimer reaction also should be monitored closely; however, death from the Jarisch-Herxheimer reaction from relapsing fever acquired in North America has not been reported despite the severity of this reaction. An opioid partial agonist, meptazinol, has been shown to reduce the severity of symptoms, but this treatment has not been thoroughly studied.

PROGNOSIS

Death is more likely if relapsing fever is acquired from a louse rather than a tick. Nutritional status may play a significant role in outcome. LBRF often occurs in settings of famine or overcrowding, where nutrition may be poor and additional diseases may complicate the diagnosis or the disease course. Among treated individuals, the fatality rate is 5% for LBRF and much lower for TBRF. Two *Borrelia* species associated with a relatively high fatality rate from relapsing fever are *B. recurrentis* (louse-borne) and *B. duttonii* (tick-borne).

PREVENTION

Prevention of TBRF includes the avoidance of rodent- and tick-infested dwellings as well as infested natural sites, such as animal burrows or caves. Environmental health specialists from local health departments and pest removal services may be consulted about rodent-proofing of homes and vacation cabins and reduction of rodent habitat around homes (e.g., by removal of rodent nesting material from walls, ceilings, and floors). Chemical treatment of rodent-infested areas should be undertaken only by pest control specialists. Contact with ticks and potential animal hosts should occur only while gloves are worn, given that TBRF has been contracted through skin contact with contaminated blood (e.g., in laboratory accidents and blood transfusions). Wearing clothing that protects the skin (e.g., long pants and long-sleeved shirts) and applying insect repellents such as DEET and permethrin to exposed skin and clothing, respectively, can prevent transmission of disease by hard ticks and possibly by soft ticks in some settings (e.g., in caves or—as a partial measure of protection—during sleep). Protection while sleeping in a potentially infested dwelling may best be provided by topical repellents. Sleeping on the floor or on a bed positioned directly against a wall should be avoided.

To prevent LBRF, it is necessary to prevent louse infestation by promoting personal hygiene (e.g., bathing) and systematic delousing (e.g., application of permethrin to clothing). Laundering or disposal of infested clothing and bedding is another important measure. Control of epidemics may involve widespread antibiotic use.

TBRF is reportable through local health departments in some states where the disease is endemic. (State health departments maintain a list of reportable diseases.) Reporting is encouraged for cases diagnosed in these states; surveillance data derived from reported cases are vital for an enhanced understanding of this tick-borne disease. Reporting of cases of TBRF may result in prevention and education of the public and physicians about the epidemiology and clinical presentation of the disease.

FURTHER READINGS

Cadavid D, Barbour AG: Neuroborreliosis during relapsing fever: Review of the clinical manifestations, pathology, and treatment of infections in humans and experimental animals. Clin Infect Dis 26:151, 1998 [PMID: 9455525]

Dworkin MS et al: Tick-borne relapsing fever. Infect Dis Clin North Am 22:449, 2008 [PMID: 18755384]

——— et al: Tick-borne relapsing fever in the northwestern United States and southwestern Canada. Clin Infect Dis 26:122, 1998

Griffin GE: Cytokines involved in human septic shock—the model of the Jarisch-Herxheimer reaction. J Antimicrob Chemother 41(Suppl A):25, 1998 Lundqvist J et al: Concomitant infection decreases the malaria burden but escalates relapsing fever borreliosis. Infect Immun 78:1924, 2010

Ramos JM et al: Results of a 10-year survey of louse-borne relapsing fever in southern Ethiopia: A decline in endemicity. Ann Trop Med Parasitol 102:467, 2008 [PMID: 18577339]

Schwan TG et al: Phylogenetic analysis of the spirochetes *Borrelia parkeri* and *Borrelia turicatae* and the potential for tick-borne relapsing fever in Florida. J Clin Microbiol 43:3851, 2005 [PMID: 16081922]

——— et al: Borrelia, in *Manual of Clinical Microbiology* , 7th ed, PR Murray et al (eds). Washington, DC, American Society for Microbiology, 1999

Vial L et al: Incidence of tick-borne relapsing fever in west Africa: Longitudinal study. Lancet 368:37, 2006 [PMID: 16815378]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LYME BORRELIOSIS: INTRODUCTION

Definition

Lyme borreliosis is caused by a spirochete, *Borrelia burgdorferi sensu lato*, that is transmitted by ticks of the *Ixodes ricinus* complex. The infection usually begins with a characteristic expanding skin lesion, erythema migrans (EM; stage 1, localized infection). After several days or weeks, the spirochete may spread to many different sites (stage 2, disseminated infection). Possible manifestations of disseminated infection include secondary annular skin lesions, meningitis, cranial neuritis, radiculoneuritis, peripheral neuritis, carditis, atrioventricular nodal block, or migratory musculoskeletal pain. Months or years later (usually after periods of latent infection), intermittent or persistent arthritis, chronic encephalopathy or polyneuropathy, or acrodermatitis may develop (stage 3, persistent infection). Most patients experience early symptoms of the illness during the summer, but the infection may not become symptomatic until it progresses to stage 2 or 3.

Lyme disease was recognized as a separate entity in 1976 because of geographic clustering of children in Lyme, Connecticut, who were thought to have juvenile rheumatoid arthritis. It became apparent that Lyme disease was a multisystemic illness that affected primarily the skin, nervous system, heart, and joints. Epidemiologic studies of patients with EM implicated certain *Ixodes* ticks as vectors of the disease. Early in the twentieth century, EM had been described in Europe and attributed to *I. ricinus* tick bites. In 1982, a previously unrecognized spirochete, now called *Borrelia burgdorferi*, was recovered from *Ixodes scapularis* ticks and then from patients with Lyme disease. The entity is now called Lyme disease or Lyme borreliosis.

Etiologic Agent

B. burgdorferi, the causative agent of Lyme disease, is a fastidious microaerophilic bacterium. The spirochete's genome is quite small (~1.5 Mb) and consists of a highly unusual linear chromosome of 950 kb as well as 17–21 linear and circular plasmids. The most remarkable aspect of the *B. burgdorferi* genome is that there are sequences for more than 100 known or predicted lipoproteins—a larger number than in any other organism. The spirochete has few proteins with biosynthetic activity and depends on its host for most of its nutritional requirements. It has no sequences for recognizable toxins.

Currently, 13 closely related borrelial species are collectively referred to as *Borrelia burgdorferi sensu lato* (i.e., *B. burgdorferi* in the general sense). The human infection Lyme borreliosis is caused primarily by three pathogenic genospecies: *B. burgdorferi sensu stricto* (*B. burgdorferi* in the strict sense, hereafter referred to as *B. burgdorferi*), *Borrelia garinii*, and *Borrelia afzelii*. *B. burgdorferi* is the sole cause of the infection in the United States; all three genospecies are found in Europe, and the latter two species occur in Asia.

Strains of *B.burgdorferi* have been subdivided according to several typing schemes, including one based on sequence variation of outer-surface protein C (OspC) and a second based on differences in the 16S–23S rRNA intergenic spacer region (RST or IGS). From these typing systems, it is apparent that strains of *B. burgdorferi* differ in pathogenicity. OspC type A (RST1) strains seem to be particularly virulent and may have played a role in the emergence of Lyme disease in epidemic form in the late twentieth century.

Epidemiology

The 13 known genospecies of *B. burgdorferi sensu lato* live in nature in enzootic cycles involving 14 species of ticks that are part of the *I. ricinus* complex. *I. scapularis* (Fig. 397-1) is the principal vector in the eastern United States from Maine to Georgia and in the midwestern states of Wisconsin, Minnesota, and Michigan. *I. pacificus* is the vector in the western states of California and Oregon. The disease is acquired throughout Europe (from Great Britain to Scandinavia to European Russia), where *I. ricinus* is the vector, and in Asian Russia, China, and Japan, where *I. persulcatus* is the vector. These ticks may transmit other diseases as well. In the United States, *I. scapularis* also transmits babesiosis and human anaplasmosis; in Europe and Asia, *I. ricinus* and *I. persulcatus* also transmit tick-borne encephalitis.

Ticks of the *I. ricinus* complex have larval, nymphal, and adult stages. They require a blood meal at each stage. The risk of infection in a given area depends largely on the density of these ticks as well

as their feeding habits and animal hosts, which have evolved differently in different locations. For *I. scapularis* in the northeastern United States, the white-footed mouse and certain other rodents are the preferred hosts of the immature larvae and nymphs. It is critical that both of the tick's immature stages feed on the same host because the life cycle of the spirochete depends on horizontal transmission: in early summer from infected nymphs to mice and in late summer from infected mice to larvae, which then molt to become the infected nymphs that will begin the cycle again the following year. It is the tiny nymphal tick that is primarily responsible for transmission of the disease to humans during the early summer months. White-tailed deer, which are not involved in the life cycle of the spirochete, are the preferred host for the adult stage of *I. scapularis* and seem to be critical to the tick's survival.

Lyme disease is now the most common vector-borne infection in the United States and Europe. Since surveillance was begun by the Centers for Disease Control and Prevention (CDC) in 1982, the number of cases in the United States has increased dramatically. More than 25,000 new cases are now reported each summer. In Europe, the highest reported frequencies of the disease are in the middle of the continent and in Scandinavia.

Pathogenesis and Immunity

To maintain its complex enzootic cycle, *B. burgdorferi* must adapt to two markedly different environments: the tick and the mammalian host. The spirochete expresses outer-surface protein A (OspA) in the midgut of the tick, whereas OspC is upregulated as the organism travels to the tick's salivary gland. There, OspC binds a tick salivary-gland protein (Salp15), which is required for infection of the mammalian host. The tick must usually be attached for at least 24 h for transmission of *B. burgdorferi*.

After injection into the human skin, *B. burgdorferi* may migrate outward, producing EM, and may spread hematogenously or in the lymph to other organs. The only known virulence factors of *B. burgdorferi* are surface proteins that allow the spirochete to attach to mammalian proteins, integrins, glycosaminoglycans, or glycoproteins. For example, spread through the skin and other tissue matrices may be facilitated by the binding of human plasminogen and its activators to the surface of the spirochete. Some *Borrelia* strains bind complement regulator–acquiring surface proteins (FHL-1/reconectin, or factor H), which help to protect spirochetes from complement-mediated lysis. Dissemination of the organism in the blood is facilitated by binding to the fibrinogen receptor on activated platelets ($\alpha_{IIb}\beta_3$) and the vitronectin receptor ($\alpha_v\beta_3$) on endothelial cells. As the name indicates, spirochetal decorin-binding proteins A and B bind decorin, a glycosaminoglycan on collagen fibrils; this binding may explain why the organism is commonly aligned with collagen fibrils in the extracellular matrix in the heart, nervous system, or joints.

To control and eradicate *B. burgdorferi*, the host mounts both innate and adaptive immune responses, resulting in macrophage- and antibody-mediated killing of the spirochete. As part of the innate immune response, complement may lyse the spirochete in the skin. Chemokines released by constituent cells in the skin lead to the recruitment of neutrophils and macrophages; the latter release potent proinflammatory cytokines. The purpose of the adaptive immune response appears to be the production of specific antibodies, which opsonize the organism—a step necessary for optimal spirochetal killing. Studies with protein arrays expressing ~1400 *B. burgdorferi* proteins detected antibody responses to a total of 89 spirochetal proteins (primarily outer-surface lipoproteins) in a population of patients with Lyme arthritis. Histologic examination of all affected tissues reveals an infiltration of lymphocytes, macrophages, and plasma cells with some degree of vascular damage (including mild vasculitis or hypervascular occlusion). These findings suggest that the spirochete may have been present in or around blood vessels.

In enzootic infection, *B. burgdorferi* spirochetes must survive this immune assault only during the summer months before returning to larval ticks to begin the cycle again the following year. In contrast, infection of humans is a dead-end event for the spirochete. Within several weeks or months, innate and adaptive immune mechanisms—even without antibiotic treatment—control widely disseminated infection, and generalized systemic symptoms wane. However, without antibiotic therapy, spirochetes may survive in localized niches for several more years. For example, *B.burgdorferi* infection in the United States may cause persistent arthritis or, in rare cases, subtle encephalopathy or polyneuropathy. Thus, immune mechanisms seem to succeed eventually in the near or total eradication of *B. burgdorferi* from selected niches, including the joints or nervous system.

Clinical Manifestations

EARLY INFECTION: STAGE 1 (LOCALIZED INFECTION)

Because of the small size of nymphal ixodid ticks, most patients do not remember the preceding tick bite. After an incubation period of 3–32 days, EM, which occurs at the site of the tick bite, usually begins as a red macule or papule that expands slowly to form a large annular lesion **(Fig. 173-1)**. As the lesion increases in size, it often develops a bright red outer border and partial central clearing. The center of the lesion sometimes becomes intensely erythematous and indurated, vesicular, or necrotic. In other instances, the expanding lesion remains an even, intense red; several red rings are found

within an outside ring; or the central area turns blue before the lesion clears. Although EM can be located anywhere, the thigh, groin, and axilla are particularly common sites. The lesion is warm but not often painful. Approximately 20% of patients do not exhibit this characteristic skin manifestation.

Figure 173-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A classic erythema migrans lesion (9 cm in diameter) is shown near the right axilla. The lesion has partial central clearing, a bright red outer border, and a target center. *(Courtesy of Vijay K. Sikand, MD; with permission.)*

EARLY INFECTION: STAGE 2 (DISSEMINATED INFECTION)

In cases in the United States, *B. burgdorferi* often spreads hematogenously to many sites within days or weeks after the onset of EM. In these cases, patients may develop secondary annular skin lesions similar in appearance to the initial lesion. Skin involvement is commonly accompanied by severe headache, mild stiffness of the neck, fever, chills, migratory musculoskeletal pain, arthralgias, and profound malaise and fatigue. Less common manifestations include generalized lymphadenopathy or splenomegaly, hepatitis, sore throat, nonproductive cough, conjunctivitis, iritis, or testicular swelling. Except for fatigue and lethargy, which are often constant, the early signs and symptoms of Lyme disease are typically intermittent and changing. Even in untreated patients, the early symptoms usually become less severe or disappear within several weeks. In ~15% of patients, the infection presents with these nonspecific systemic symptoms.

Symptoms suggestive of meningeal irritation may develop early in Lyme disease when EM is present but usually are not associated with cerebrospinal fluid (CSF) pleocytosis or an objective neurologic deficit. After several weeks or months, ~15% of untreated patients develop frank neurologic abnormalities, including meningitis, subtle encephalitic signs, cranial neuritis (including bilateral facial palsy), motor or sensory radiculoneuropathy, peripheral neuropathy, mononeuritis multiplex, cerebellar ataxia, or myelitis—alone or in various combinations. In the United States, the usual pattern consists of fluctuating symptoms of meningitis accompanied by facial palsy and peripheral radiculoneuropathy. Lymphocytic pleocytosis (~100 cells/μL) is found in CSF, often along with elevated protein levels and normal or slightly low glucose concentrations. In Europe and Asia, the first neurologic sign is characteristically radicular pain, which is followed by the development of CSF pleocytosis (called meningopolyneuritis, or *Bannwarth's syndrome*); meningeal or encephalitic signs are frequently absent. In children, the optic nerve may be affected because of inflammation or increased intracranial pressure, which may lead to blindness. These early neurologic abnormalities usually resolve completely within months, but in rare cases chronic neurologic disease may occur later.

Within several weeks after the onset of illness, ~8% of patients develop cardiac involvement. The most common abnormality is a fluctuating degree of atrioventricular block (first-degree, Wenckebach, or complete heart block). Some patients have more diffuse cardiac involvement, including electrocardiographic changes indicative of acute myopericarditis, left ventricular dysfunction evident on radionuclide scans, or (in rare cases) cardiomegaly or pancarditis. Cardiac involvement usually lasts for only a few weeks but may recur. Chronic cardiomyopathy caused by *B. burgdorferi* has been reported in Europe.

During this stage, musculoskeletal pain is common. The typical pattern consists of migratory pain in joints, tendons, bursae, muscles, or bones (usually without joint swelling) lasting for hours or days and affecting one or two locations at a time.

LATE INFECTION: STAGE 3 (PERSISTENT INFECTION)

Months after the onset of infection, ~60% of patients in the United States who have received no

antibiotic treatment develop frank arthritis. The typical pattern comprises intermittent attacks of oligoarticular arthritis in large joints (especially the knees), lasting for weeks or months in a given joint. A few small joints or periarticular sites may also be affected, primarily during early attacks. The number of patients who continue to have recurrent attacks decreases each year. However, in a small percentage of cases, involvement of large joints—usually one or both knees—is persistent and may lead to erosion of cartilage and bone.

White cell counts in joint fluid range from 500 to 110,000/μL (average, 25,000/μL); most of these cells are polymorphonuclear leukocytes. Tests for rheumatoid factor or antinuclear antibodies usually give negative results. Examination of synovial biopsy samples reveals fibrin deposits, villous hypertrophy, vascular proliferation, microangiopathic lesions, and a heavy infiltration of lymphocytes and plasma cells.

Although most patients with Lyme arthritis respond well to antibiotic therapy, a small percentage in the northeastern United States have persistent arthritis for months or even for several years after the near or total eradication of spirochetes from the joints by antibiotic therapy. Compared with antibiotic-responsive patients, those with antibiotic-refractory arthritis are more often infected with RST1 strains of *B. burgdorferi*; have a higher frequency of certain class II major histocompatibility complex molecules (particularly HLA-DRBI*0401 or -*0101 molecules) that bind an epitope of OspA ($OspA_{163-175}$); and often exhibit T cell recognition of this epitope. In addition, these patients have significantly higher levels of proinflammatory chemokines and cytokines in joint fluid (especially CXCL9 and interferon γ) than do antibiotic-responsive patients; these higher levels persist during the postantibiotic period, when polymerase chain reaction (PCR) results for *B. burgdorferi* DNA are uniformly negative. It has been postulated that, in these genetically susceptible individuals, *B. burgdorferi* may trigger localized, tissue-specific autoimmunity within the proinflammatory milieu of the joints.

Although rare, chronic neurologic involvement may also become apparent from months to several years after the onset of infection, sometimes following long periods of latent infection. The most common form of chronic central nervous system involvement is subtle encephalopathy affecting memory, mood, or sleep, and the most common form of peripheral neuropathy is an axonal polyneuropathy manifested as either distal paresthesia or spinal radicular pain. Patients with encephalopathy frequently have evidence of memory impairment in neuropsychological tests and abnormal results in CSF analyses. In cases of polyneuropathy, electromyography generally shows extensive abnormalities of proximal and distal nerve segments. Encephalomyelitis or leukoencephalitis, a rare manifestation of Lyme borreliosis associated primarily with *B. garinii* infection in Europe, is a severe neurologic disorder that may include spastic paraparesis, upper motor-neuron bladder dysfunction, and, rarely, lesions in the periventricular white matter.

Acrodermatitis chronica atrophicans, the late skin manifestation of Lyme borreliosis, has been associated primarily with *B. afzelii* infection in Europe and Asia. It has been observed especially often in elderly women. The skin lesions, which are usually found on the acral surface of an arm or leg, begin insidiously with reddish-violaceous discoloration; they become sclerotic or atrophic over a period of years.

The basic patterns of Lyme borreliosis are similar worldwide, but there are regional variations, primarily between the illness found in North America, which is caused exclusively by *B. burgdorferi*, and that found in Europe, which is caused primarily by *B. afzelii* and *B. garinii*. With each of the *Borrelia* species, the infection usually begins with EM. However, *B. burgdorferi* often disseminates widely; it is particularly arthritogenic, and it may cause antibiotic-refractory arthritis. *B. garinii* typically disseminates less widely, but it is especially neurotropic and may cause borrelial encephalomyelitis. *B. afzelii* often infects only the skin but may persist in that site, where it may cause several different dermatoborrelioses, including acrodermatitis chronica atrophicans.

POST–LYME SYNDROME (CHRONIC LYME DISEASE)

Despite resolution of the objective manifestations of the infection with antibiotic therapy, a small percentage of patients have pain, neurocognitive manifestations, or fatigue symptoms for months or years afterward. This syndrome is similar to or indistinguishable from chronic fatigue syndrome (Chap. 389) and fibromyalgia (Chap. 335). Compared with symptoms of active Lyme disease, post-Lyme symptoms tend to be more generalized or disabling. They include marked fatigue, severe headache, diffuse musculoskeletal pain, multiple symmetric tender points in characteristic locations, pain and stiffness in many joints, diffuse paresthesias, difficulty with concentration, and sleep disturbances. Patients with this condition lack evidence of joint inflammation, have normal neurologic test results, and may exhibit anxiety and depression. In contrast, late manifestations of Lyme disease, including arthritis, encephalopathy, and neuropathy, are usually associated with minimal systemic symptoms. Currently, no evidence indicates that persistent subjective symptoms after recommended courses of antibiotic therapy are caused by active infection.

Diagnosis

The culture of *B. burgdorferi* in Barbour-Stoenner-Kelly (BSK) medium permits definitive diagnosis, but this method has been used primarily in research studies. Moreover, with a few exceptions, positive cultures have been obtained only early in the illness—particularly from biopsy samples of EM skin lesions, less often from plasma samples, and occasionally from CSF samples. Later in the infection, PCR is greatly superior to culture for the detection of *B. burgdorferi* DNA in joint fluid—the major use for PCR testing in Lyme disease. However, the sensitivity of PCR determinations in CSF from patients with neuroborreliosis has been much lower. There seems to be little if any role for PCR in the detection of *B. burgdorferi* DNA in blood or urine samples. Moreover, this procedure must be carefully controlled to prevent contamination.

Because of the problems associated with direct detection of *B. burgdorferi*, Lyme disease is usually diagnosed by the recognition of a characteristic clinical picture with serologic confirmation. Although serologic testing may yield negative results during the first several weeks of infection, most patients have a positive antibody response to *B. burgdorferi* after that time. The limitation of serologic tests is that they do not clearly distinguish between active and inactive infection. Patients with previous Lyme disease—particularly in cases progressing to late stages—often remain seropositive for years, even after adequate antibiotic treatment. In addition, ~10% of patients are seropositive because of asymptomatic infection. If these individuals subsequently develop another illness, the positive serologic test for Lyme disease may cause diagnostic confusion. According to an algorithm published by the American College of Physicians (Table 173-1), serologic testing for Lyme disease is recommended only for patients with at least an intermediate pretest probability of Lyme disease, such as those with oligoarticular arthritis. It should not be used as a screening procedure in patients with pain or fatigue syndromes. In such patients, the probability of a false-positive serologic result is higher than that of a true-positive result.

Table 173-1 Algorithm for Testing for and Treating Lyme Disease

Pretest Probability	Example	Recommendation
High	Patients with erythema migrans	Empirical antibiotic treatment without serologic testing
Intermediate	Patients with oligoarticular arthritis	Serologic testing and antibiotic treatment if test results are positive
Low	Patients with nonspecific symptoms (myalgias, arthralgias, fatigue)	Neither serologic testing nor antibiotic treatment

Source: Adapted from the recommendations of the American College of Physicians (G Nichol et al: Ann Intern Med 128:37, 1998, with permission).

For serologic analysis of Lyme disease in the United States, the CDC recommends a two-step approach in which samples are first tested by enzyme-linked immunosorbent assay (ELISA) and equivocal or positive results are then tested by western blotting. During the first month of infection, both IgM and IgG responses to the spirochete should be determined, preferably in both acute- and convalescent-phase serum samples. Approximately 20–30% of patients have a positive response detectable in acute-phase samples, whereas ~70–80% have a positive response during convalescence (2–4 weeks later). After 4–8 weeks of infection (by which time most patients with active Lyme disease have disseminated infection), the sensitivity and specificity of the IgG response to the spirochete are both very high—in the range of 99%—as determined by the two-test approach of ELISA and western blot. At this point and thereafter, a single test (that for IgG) is usually sufficient. In persons with illness of >2 months' duration, a positive IgM test result alone is likely to be false-positive and therefore should not be used to support the diagnosis.

According to current criteria adopted by the CDC, an IgM western blot is considered positive if two of the following three bands are present: 23, 39, and 41 kDa. However, the combination of the 23- and 41-kDa bands may still represent a false-positive result. Misuse or misinterpretation of IgM blots has been a factor in the incorrect diagnosis of Lyme disease in patients with other illnesses. An IgG blot is considered positive if 5 of the following 10 bands are present: 18, 23, 28, 30, 39, 41, 45, 58, 66, and 93 kDa. In European cases, there is less expansion of the antibody response, and no single set of criteria for the interpretation of immunoblots results in high levels of sensitivity and specificity in all countries.

The most promising second-generation serologic test is the C6 peptide IgGELISA, which employs a 26-mer of the sixth invariant region of the VlsE lipoprotein of *B. burgdorferi*. The results achieved with this test are similar to those obtained with the standard two-test approach (sonicate IgM and IgG ELISA and western blot). The principal advantage of the C6 peptide ELISA is the early detection of an IgG response, which renders an IgM test unnecessary. However, not all patients with late Lyme disease have a response to the C6 peptide, and this test is not quite as specific as sonicate western blot. Thus, at present, a two-test approach that includes western blot is still recommended. Like sonicate

test responses, the response to the VlsE peptide may persist for months or years after successful antibiotic treatment; therefore, persistence of antibody to VlsE cannot be equated with spirochetal persistence in Lyme disease.

Differential Diagnosis

Classic EM is a slowly expanding erythema, often with partial central clearing. If the lesion expands little, it may represent the red papule of an uninfected tick bite. If the lesion expands rapidly, it may represent cellulitis (e.g., streptococcal cellulitis) or an allergic reaction, perhaps to tick saliva. Patients with secondary annular lesions may be thought to have erythema multiforme, but neither the development of blistering mucosal lesions nor the involvement of the palms or soles is a feature of *B. burgdorferi* infection. In the southeastern United States, an EM-like skin lesion, sometimes with mild systemic symptoms, may be associated with *Amblyomma americanum* tick bites. However, the cause of this Southern tick-associated rash illness (STARI) has not yet been identified.

In the United States, *I. scapularis* ticks may transmit not only *B. burgdorferi* but also *Babesia microti*, a red blood cell parasite (Chap. 211), or *Anaplasma phagocytophilum*, the agent of human granulocytotropic anaplasmosis (formerly human granulocytotropic ehrlichiosis; Chap. 174). Although babesiosis and anaplasmosis are most often asymptomatic, infection with any of these three agents may cause nonspecific systemic symptoms, and co-infected patients may have more severe or persistent symptoms than patients infected with a single agent. Standard blood counts may yield clues regarding the presence of co-infection. Anaplasmosis may cause leukopenia or thrombocytopenia, and babesiosis may cause thrombocytopenia or (in severe cases) hemolytic anemia. IgM serologic responses may confuse the diagnosis. For example, *A. phagocytophilum* may elicit a positive IgM response to *B. burgdorferi*. The frequency of co-infection in different studies has been variable. In one prospective study, 4% of patients with EM had evidence of co-infection.

Facial palsy caused by *B. burgdorferi*, which occurs in the early disseminated phase of the infection (often in July, August, or September), is usually recognized by its association with EM. However, in rare cases, facial palsy without EM may be the presenting manifestation of Lyme disease. In such cases, both the IgM and the IgG responses to the spirochete are usually positive. The most common infectious agents that cause facial palsy are herpes simplex virus type 1 (Bell's palsy; Chap. 179) and varicella-zoster virus (Ramsay Hunt syndrome; Chap. 180).

Later in the infection, oligoarticular Lyme arthritis most resembles reactive arthritis in an adult or the pauciarticular form of juvenile idiopathic arthritis in a child. Patients with Lyme arthritis usually have the strongest IgG antibody responses seen in Lyme borreliosis, with reactivity to many spirochetal proteins.

The most common problem in diagnosis is to mistake Lyme disease for chronic fatigue syndrome (Chap. 389) or fibromyalgia (Chap. 335). This difficulty is compounded by the fact that a small percentage of patients do in fact develop these chronic pain or fatigue syndromes in association with or soon after Lyme disease. Moreover, a counterculture has emerged that ascribes pain and fatigue syndromes to chronic Lyme disease when there is little or no evidence of *B. burgdorferi* infection. In such cases, the term *chronic Lyme disease*, which is equated with chronic *B. burgdorferi* infection, is a misnomer, and the use of prolonged, dangerous, and expensive antibiotic treatment is not warranted.

Treatment: Lyme Borreliosis

ANTIBIOTIC TREATMENT

As outlined in the algorithm in Fig. 173-2, the various manifestations of Lyme disease can usually be treated successfully with orally administered antibiotics; the exceptions are objective neurologic abnormalities and third-degree atrioventricular heart block, which are generally treated with IV antibiotics. For early Lyme disease, doxycycline is effective and can be administered to men and nonpregnant women. An advantage of this regimen is that it is also effective against *A. phagocytophilum*, which is transmitted by the same tick that transmits the Lyme disease agent. Amoxicillin, cefuroxime axetil, and erythromycin or its congeners are second-, third-, and fourth-choice alternatives, respectively. In children, amoxicillin is effective (not more than 2 g/d); in cases of penicillin allergy, cefuroxime axetil or erythromycin may be used. In contrast to second- or third-generation cephalosporin antibiotics, first-generation cephalosporins, such as cephalexin, are not effective. For patients with infection localized to the skin, a 14-day course of therapy is generally sufficient; in contrast, for patients with disseminated infection, a 21-day course is recommended. Approximately 15% of patients experience a Jarisch-Herxheimer-like reaction during the first 24 h of therapy. In multicenter studies, >90% of patients whose early Lyme disease was treated with these regimens had satisfactory outcomes. Although some patients reported symptoms after treatment, objective evidence of persistent infection or relapse was rare, and re-treatment was usually unnecessary.

Figure 173-2

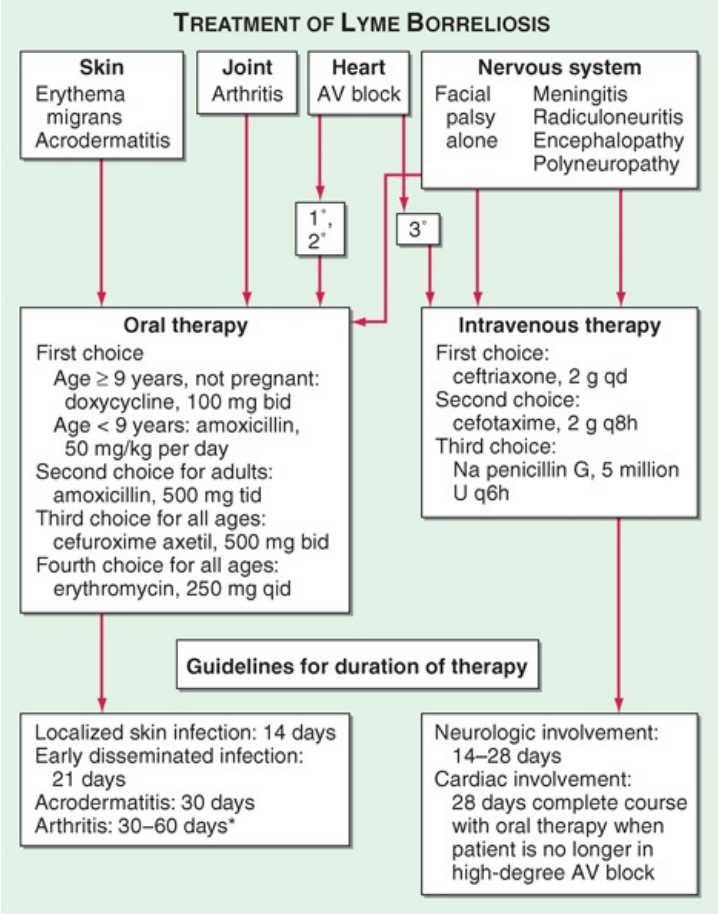

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the treatment of the various acute or chronic manifestations of Lyme borreliosis. AV, atrioventricular. *For Lyme arthritis, IV ceftriaxone (2 g given once a day for 14–28 days) is also effective and is necessary for a small percentage of patients; however, compared with oral treatment, this regimen is less convenient to administer, has more side effects, and is more expensive.

Oral administration of doxycycline or amoxicillin for 30 days is recommended for the initial treatment of Lyme arthritis in patients who do not have concomitant neurologic involvement. Among patients with arthritis who do not respond to oral antibiotics, re-treatment with IV ceftriaxone for 28 days is appropriate. In patients with arthritis in whom—despite a negative PCR result for *B. burgdorferi* DNA in joint fluid—joint inflammation persists for months or even several years after both oral and IV antibiotics, treatment with anti-inflammatory agents or synovectomy may be successful.

In the United States, parenteral antibiotic therapy is usually used for objective neurologic abnormalities (with the possible exception of facial palsy alone). Patients with neurologic involvement are most commonly treated with IV ceftriaxone for 14–28 days, but IV cefotaxime or IV penicillin G for the same duration may also be effective. In Europe, similar results have been achieved with oral doxycycline and IV antibiotics in the treatment of acute neuroborreliosis. In patients with high-degree atrioventricular block or a PR interval of >0.3 s, IV therapy for at least part of the course and cardiac monitoring are recommended, but the insertion of a permanent pacemaker is not necessary.

It is unclear how and whether asymptomatic infection should be treated, but patients with such infection are often given a course of oral antibiotics. Because maternal-fetal transmission of *B. burgdorferi* seems to occur rarely (if at all), standard therapy for the manifestations of the illness is recommended for pregnant women. Long-term persistence of *B. burgdorferi* has not been documented in any large series of patients after treatment with currently recommended regimens. Although an occasional patient requires a second course of antibiotics, there is no indication for multiple, repeated antibiotic courses in the treatment of Lyme disease.

CHRONIC LYME DISEASE

After appropriately treated Lyme disease, a small percentage of patients continue to have subjective symptoms, primarily musculoskeletal pain, neurocognitive difficulties, or fatigue. This *chronic Lyme disease* or *post–Lyme syndrome* is a disabling condition that is similar to chronic fatigue syndrome or fibromyalgia. In a large study, one group of patients with post–Lyme syndrome received IV ceftriaxone for 30 days followed by oral doxycycline for 60 days, while another group received IV and oral placebo

preparations for the same durations. No significant differences were found between groups in the numbers of patients reporting that their symptoms had improved, become worse, or stayed the same. Such patients are best treated for the relief of symptoms rather than with prolonged courses of antibiotics.

PROPHYLAXIS AFTER A TICK BITE

The risk of infection with *B. burgdorferi* after a recognized tick bite is so low that antibiotic prophylaxis is not routinely indicated. However, if an attached, engorged *I. scapularis* nymph is found or if follow-up is anticipated to be difficult, a single 200-mg dose of doxycycline, which usually prevents Lyme disease when given within 72 h after the tick bite, may be administered.

Prognosis

The response to treatment is best early in the disease. Later treatment of Lyme borreliosis is still effective, but the period of convalescence may be longer. Eventually, most patients recover with minimal or no residual deficits.

Reinfection

Reinfection may occur after EM when patients are treated with antimicrobial agents. In such cases, the immune response is not adequate to provide protection from subsequent infection. However, patients who develop an expanded immune response to the spirochete over a period of months (e.g., those with Lyme arthritis) have protective immunity for a period of years and do not acquire the infection again.

Prevention

Protective measures for the prevention of Lyme disease may include the avoidance of tick-infested areas, the use of repellents and acaricides, tick checks, and modification of landscapes in or near residential areas. Although a vaccine for Lyme disease used to be available, the manufacturer has discontinued its production. Therefore, no vaccine is now commercially available for the prevention of this infection.

FURTHER READINGS

Feder HM jr et al: A critical appraisal of "chronic lyme disease." N Engl J Med 357:1422, 2007[PMID: 17914043]

Klempner MS et al: Two controlled trials of antibiotic treatment in patients with persistent symptoms and a history of Lyme disease. N Engl J Med 345:85, 2001[PMID: 11450676]

Nadelman RB et al: Prophylaxis with single-dose doxycycline for the prevention of Lyme disease after *Ixodes scapularis* tick bite. N Engl J Med 345:79, 2001[PMID: 11450675]

Steere AC: Lyme disease. N Engl J Med 345:115, 2001[PMID: 11450660]

———, Angelis SM: Therapy for Lyme arthritis: Strategies for the treatment of antibiotic-refractory arthritis. Arthritis Rheum 54:3079, 2006

——— et al: Prospective study of serologic tests for Lyme disease. Clin Infect Dis 47:188, 2008

——— et al: The emergence of Lyme disease. J Clin Invest 113:1093, 2004

Wormser GP et al: The clinical assessment, treatment and prevention of Lyme disease, anaplasmosis, and babesiosis: Clinical practice guidelines of the Infectious Diseases Society of America. Clin Infect Dis 43:1089, 2006[PMID: 17029130]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

RICKETTSIAL DISEASES: INTRODUCTION

The rickettsiae are a heterogeneous group of small, obligately intracellular, gram-negative coccobacilli and short bacilli, most of which are transmitted by a tick, mite, flea, or louse vector. Except in the case of louse-borne typhus, humans are incidental hosts. Among rickettsiae, *Coxiella burnetii, Rickettsia prowazekii,* and *R. typhi* have the well-documented ability to survive for an extended period outside the reservoir or vector and to be extremely infectious: inhalation of a single *Coxiella* microorganism can cause pneumonia. High infectivity and severe illness after inhalation make *R. prowazekii, R. rickettsii, R. typhi, R. conorii,* and *C. burnetii* bioterrorism threats.

Clinical infections with rickettsiae can be classified according to (1) the taxonomy and diverse microbial characteristics of the agents, which belong to six genera (*Rickettsia, Orientia, Ehrlichia, Anaplasma, Neorickettsia,* and *Coxiella*); (2) epidemiology; or (3) clinical manifestations. The clinical manifestations of all the acute presentations are similar during the first 5 days: fever, headache, and myalgias with or without nausea, vomiting, and cough. As the course progresses, clinical manifestations—including occurrence of a macular, maculopapular, or vesicular rash; eschar; pneumonitis; and meningoencephalitis—vary from one disease to another. Given the 14 etiologic agents with varied mechanisms of transmission, geographic distributions, and associated disease manifestations, the consideration of rickettsial diseases as a single entity poses complex challenges **(Table 174-1)**.

Table 174-1 Features of Selected Rickettsial Infections

Disease	Organism	Transmission	Geographic Range	Incubation Period, Days	Duration, Days	Rash, %	Eschar, %	Lymphadenopathy[a]
Rocky Mountain spotted fever	*Rickettsia rickettsii*	Tick bite: *Dermacentor andersoni, D. variabilis*	United States	2–14	10–20	90	<1	+
		Amblyomma cajennense, A. aureolatum	Central/South America					
		Rhipicephalus sanguineus	Mexico, Brazil, United States					
Mediterranean spotted fever	*R. conorii*	Tick bite: *R. sanguineus, R. pumilio*	Southern Europe, Africa, Middle East, Central Asia	5–7	7–14	97	50	+
African tick-bite fever	*R. africae*	Tick bite: *A. hebraeum, A. variegatum*	Sub-Saharan Africa, West Indies	4–10	?	50	90	++++
Maculatum disease	*R. parkeri*	*A. maculatum*	United States, South America	2–10	?	88	94	++
Rickettsialpox	*R. akari*	Mite bite: *Liponyssoides sanguineus*	United States, Ukraine, Turkey, Mexico, Croatia	10–17	3–11	100	90	+++
Tick-borne lymphadenopathy	*R. slovaca*	Tick bite: *Dermacentor marginatus, D. reticularis*	Europe	7–9	17–180	5	100	++++
Flea-borne spotted fever	*R. felis*	Flea (mechanism undetermined): *Ctenocephalides felis*	Worldwide	8–16	8–16	80	15	—
Epidemic typhus	*R. prowazekii*	Louse feces: *Pediculus humanus corporis,* fleas and lice of flying squirrels, or recrudescence	Worldwide	7–14	10–18	80	None	—
Murine typhus	*R. typhi*	Flea feces: *Xenopsylla cheopis, C. felis,* others	Worldwide	8–16	9–18	80	None	—
Human monocytotropic ehrlichiosis	*Ehrlichia chaffeensis*	Tick bite: *Amblyomma americanum, D. variabilis*	United States	1–21	3–21	26	None	++
Ewingii ehrlichiosis	*E. ewingii*	Tick bite: *A. americanum*	United States				None	
Human granulocytotropic anaplasmosis	*Anaplasma phagocytophilum*	Tick bite: *Ixodes scapularis, I. ricinus, I. pacificus, I.*	United States, Europe, Asia	4–8	3–14	Rare	None	—

		persulcatus						
Scrub typhus	*Orientia tsutsugamushi*	Mite bite: *Leptotrombidium deliense,* others	Asia, Australia, New Guinea, Pacific Islands	9–18	6–21	50	35	+++
Q fever	*Coxiella burnetii*	Inhalation of aerosols of infected parturition material (sheep, dogs, others), ingestion of infected milk or milk products	Worldwide	3–30	5–57	<1	None	—

[a]++++, severe; +++, marked; ++, moderate; +, present in a small portion of cases; —, not a noted feature.

Establishing the etiologic diagnosis of rickettsioses is very difficult during the acute stage of illness, and definitive diagnosis usually requires the examination of paired serum samples after convalescence. Heightened clinical suspicion is based on epidemiologic data, history of exposure to vectors or reservoir animals, travel to endemic locations, clinical manifestations (sometimes including rash or eschar), and characteristic laboratory findings [including thrombocytopenia, normal or low white blood cell (WBC) counts, elevated hepatic enzyme levels, and hyponatremia]. Such suspicion should prompt empirical treatment. Doxycycline is the drug of choice for most of these infections. Only one agent, *C. burnetii*, has been documented to cause chronic illness. One other, *R. prowazekii*, causes recrudescent illness (Brill-Zinsser disease) when latent infection is reactivated years after resolution of the acute illness.

Rickettsial infections dominated by fever may resolve without further clinical evolution. However, after nonspecific early manifestations, the illnesses can also evolve along one or more of several principal clinical lines: (1) development of a macular or maculopapular rash; (2) development of an eschar at the site of tick or mite feeding; (3) development of a vesicular rash (often in rickettsialpox and African tick-bite fever); (4) development of pneumonitis with chest radiographic opacities and/or rales [Q fever and severe cases of Rocky Mountain spotted fever (RMSF), Mediterranean spotted fever (MSF), louse-borne typhus, human monocytotropic ehrlichiosis (HME), human granulocytotropic anaplasmosis (HGA), scrub typhus, and murine typhus]; (5) development of meningoencephalitis [louse-borne typhus and severe cases of RMSF, scrub typhus, HME, murine typhus, MSF, and (rarely) Q fever]; and (6) progressive hypotension and multiorgan failure as seen with sepsis or toxic shock syndrome (RMSF, MSF, louse-borne typhus, murine typhus, scrub typhus, HME, and HGA).

Epidemiologic clues to the transmission of a particular pathogen include (1) environmental exposure to ticks, fleas, or mites during the season of activity of the vector species for the disease in the appropriate geographic region (spotted fever and typhus rickettsioses, scrub typhus, ehrlichioses, anaplasmosis); (2) travel to or residence in an endemic geographic region during the incubation period (Table 174-1); (3) exposure to parturient ruminants, cats, and dogs (Q fever); (4) exposure to flying squirrels (*R. prowazekii* infection); and (5) history of previous louse-borne typhus (recrudescent typhus).

Clinical laboratory findings, such as thrombocytopenia (particularly in spotted fever and typhus rickettsioses, ehrlichioses, anaplasmosis, and scrub typhus), normal or low WBC counts, mild to moderate serum elevations of hepatic aminotransferases, and hyponatremia suggest some common pathophysiologic mechanisms.

Application of these clinical, epidemiologic, and laboratory principles requires consideration of a rickettsial diagnosis and knowledge of the individual diseases.

TICK-, MITE-, LOUSE-, AND FLEA-BORNE RICKETTSIOSES

These diseases, caused by organisms of the genera *Rickettsia* and *Orientia* in the family Rickettsiaceae, result from endothelial infection and increased vascular permeability. Pathogenic rickettsial species are very closely related, have small genomes (as a result of reductive evolution, which eliminated many genes for biosynthesis of intracellularly available molecules), and are traditionally separated into typhus and spotted fever groups on the basis of lipopolysaccharide antigens. Some diseases and their agents (e.g., *R. africae, R. parkeri*, and *R. sibirica*) are too similar to require separate descriptions. Indeed, the similarities among MSF [*R. conorii* (all strains)] and *R. massiliae*, North Asian tick typhus (*R. sibirica*), Japanese spotted fever (*R. japonica*), and Flinders Island spotted fever (*R. honei*) far outweigh the minor variations. The Rickettsiaceae that cause life-threatening infections are, in order of decreasing case-fatality rate, *R.rickettsii* (RMSF); *R. prowazekii* (louse-borne typhus); *Orientia tsutsugamushi* (scrub typhus); *R. conorii* (MSF); *R. typhi* (murine typhus); and, in rare cases, other spotted fever–group organisms. Some agents (e.g., *R. parkeri*, *R. africae*, *R. akari*, *R. slovaca*, *R. honei*, *R. felis*, *R. massiliae, R. helvetica, R. heilongjiangensis, R. aeschlimannii*, and *R. monacensis*) have never been documented to cause a fatal illness.

Rocky Mountain Spotted Fever

RMSF occurs in 47 states (with the highest prevalence in the south-central and southeastern states) as well as in Canada, Mexico, and Central and South America. The infection is transmitted by *Dermacentor variabilis*, the American dog tick, in the eastern two-thirds of the United States and California; by *D. andersoni*, the Rocky Mountain wood tick, in the western United States; by *Rhipicephalus sanguineus* in Mexico, Arizona, and probably Brazil; and by *Amblyomma cajennense* in Central and South America. Maintained principally by transovarian transmission from one generation of ticks to the next, *R. rickettsii* can be acquired by uninfected ticks through the ingestion of a blood meal from rickettsemic small mammals.

Humans become infected during tick season (in the Northern Hemisphere, from May to September), although some cases occur in winter. The mortality rate was 20–25% in the preantibiotic era and remains at ~3–5% principally because of delayed diagnosis and treatment. The case-fatality ratio increases with each decade of life above age 20.

PATHOGENESIS

R. rickettsii organisms are inoculated into the dermis along with secretions of the tick—s salivary glands after ≥6 h of feeding. The rickettsiae spread lymphohematogenously throughout the body and infect numerous foci of contiguous endothelial cells. The dose-dependent incubation period is ~1 week (range, 2–14 days). Occlusive thrombosis and ischemic necrosis are not the fundamental pathologic basis for tissue and organ injury. Instead, increased vascular permeability, with resulting edema, hypovolemia, and ischemia, is responsible. Consumption of platelets results in thrombocytopenia in 32–52% of patients, but disseminated intravascular coagulation with hypofibrinogenemia is rare. Activation of platelets, generation of thrombin, and activation of the fibrinolytic system all appear to be homeostatic physiologic responses to endothelial injury.

CLINICAL MANIFESTATIONS

Early in the illness, when medical attention usually is first sought, RMSF is difficult to distinguish from many self-limiting viral illnesses. Fever, headache, malaise, myalgia, nausea, vomiting, and anorexia are the most common symptoms during the first 3 days. The patient becomes progressively more ill as vascular infection and injury advance. In one large series, only one-third of patients were diagnosed with presumptive RMSF early in the clinical course and treated appropriately as outpatients. In the tertiary-

care setting, RMSF is all too often recognized only when late severe manifestations, developing at the end of the first week or during the second week of illness in patients without appropriate treatment, prompt return to a physician or hospital and admission to an intensive care unit.

The progressive nature of the infection is clearly manifested in the skin. Rash is evident in only 14% of patients on the first day of illness and in only 49% during the first 3 days. Macules (1–5 mm) appear first on the wrists and ankles and then on the remainder of the extremities and the trunk. Later, more severe vascular damage results in frank hemorrhage at the center of the maculopapule, producing a petechia that does not disappear upon compression **(Fig. 174-1)**. This sequence of events is sometimes delayed or aborted by effective treatment. However, the rash is a variable manifestation, appearing on day 6 or later in 20% of cases and not appearing at all in 9–16% of cases. Petechiae occur in 41–59% of cases, appearing on or after day 6 in 74% of cases that manifest a rash. Involvement of the palms and soles, often considered diagnostically important, usually develops relatively late in the course (after day 5 in 43% of cases) and does not develop at all in 18–64% of cases.

Figure 174-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Top: **Petechial lesions of Rocky Mountain spotted fever** on the lower legs and soles of a young, previously healthy patient. ***Bottom:*** **Close-up of lesions** from the same patient. *(Photos courtesy of Dr. Lindsey Baden; with permission.)*

Hypovolemia leads to prerenal azotemia and (in 17% of cases) hypotension. Infection of the pulmonary microcirculation leads to noncardiogenic pulmonary edema; 12% of patients have severe respiratory disease, and 8% require mechanical ventilation. Cardiac involvement manifests as dysrhythmia in 7–16% of cases.

Besides respiratory failure, central nervous system (CNS) involvement is the other important determinant of the outcome of RMSF. Encephalitis, presenting as confusion or lethargy, is apparent in 26–28% of cases. Progressively severe encephalitis manifests as stupor or delirium in 21–26% of cases, ataxia in 18%, coma in 10%, and seizures in 8%. Numerous focal neurologic deficits have been reported. Meningoencephalitis results in cerebrospinal fluid (CSF) pleocytosis in 34–38% of cases; usually there are 10–100 cells/μL and a mononuclear predominance, but occasionally there are >100 cells/μL and a polymorphonuclear predominance. The CSF protein concentration is increased in 30–35% of cases, but the CSF glucose concentration is usually normal.

Renal failure, often reversible with rehydration, is caused by acute tubular necrosis in severe cases with shock. Hepatic injury with increased serum aminotransferase concentrations (38% of cases) is due to focal death of individual hepatocytes without hepatic failure. Jaundice is recognized in 9% of cases and an elevated serum bilirubin concentration in 18–30%.

Life-threatening bleeding is rare. Anemia develops in 30% of cases and is severe enough to require transfusions in 11%. Blood is detected in the stools or vomitus of 10% of patients, and death has followed massive upper gastrointestinal hemorrhage.

Other characteristic clinical laboratory findings include increased plasma levels of proteins of the acute-phase response (C-reactive protein, fibrinogen, ferritin, and others), hypoalbuminemia, and hyponatremia (in 56% of cases) due to the appropriate secretion of antidiuretic hormone in response to the hypovolemic state. Myositis occurs occasionally, with marked elevations in serum creatine kinase levels and multifocal rhabdomyonecrosis. Ocular involvement includes conjunctivitis in 30% of cases and retinal vein engorgement, flame hemorrhages, arterial occlusion, and papilledema with normal CSF pressure in some instances.

In untreated cases, the patient usually dies 8–15 days after onset. A rare presentation, fulminant RMSF, is fatal within 5 days after onset. This fulminant presentation is seen most often in male black patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency and may be related to an undefined effect of hemolysis on the rickettsial infection. Although survivors of RMSF usually return to their previous state of health, permanent sequelae, including neurologic deficits and gangrene necessitating amputation of extremities, may follow severe illness.

DIAGNOSIS

The diagnosis of RMSF during the acute stage is more difficult than is generally appreciated. The most important epidemiologic factor is a history of exposure to a potentially tick-infested environment within the 12 days preceding disease onset during a season of possible tick activity. However, only 60% of patients actually recall being bitten by a tick during the incubation period.

The differential diagnosis for early clinical manifestations of RMSF (fever, headache, and myalgia without a rash) includes influenza, enteroviral infection, infectious mononucleosis, viral hepatitis, leptospirosis, typhoid fever, gram-negative or gram-positive bacterial sepsis, HME, HGA, murine typhus, sylvatic flying-squirrel typhus, and rickettsialpox. Enterocolitis may be suggested by nausea, vomiting, and abdominal pain; prominence of abdominal tenderness has resulted in exploratory laparotomy. CNS involvement may masquerade as bacterial or viral meningoencephalitis. Cough, pulmonary signs, and chest radiographic opacities may lead to a diagnostic consideration of bronchitis or pneumonia.

At presentation during the first 3 days of illness, only 3% of patients exhibit the classic triad of fever, rash, and history of tick exposure. When a rash appears, a diagnosis of RMSF should certainly be considered. However, many illnesses considered in the differential diagnosis may also be associated with a rash, including rubeola, rubella, meningococcemia, disseminated gonococcal infection, secondary syphilis, toxic shock syndrome, drug hypersensitivity, idiopathic thrombocytopenic purpura, thrombotic

thrombocytopenic purpura, Kawasaki syndrome, and immune complex vasculitis. Conversely, any person in an endemic area with a provisional diagnosis of one of the above illnesses may have RMSF. Thus, if a viral infection is suspected during RMSF season in an endemic area, it should always be kept in mind that RMSF can mimic viral infection early in the course; if the illness worsens over the next couple of days after initial presentation, the patient should return for reevaluation.

The most common serologic test for confirmation of the diagnosis is the indirect immunofluorescence assay. Not until 7–10 days after onset is a diagnostic titer of ≥1:64 usually detectable. The sensitivity and specificity of the indirect immunofluorescence assay are 94–100% and 100%, respectively. It is important to understand that serologic tests for RMSF are usually negative at the time of presentation for medical care and that treatment should not be delayed while a positive serologic result is awaited.

The only diagnostic test that is useful during the acute illness is immunohistologic examination of a cutaneousbiopsy sample from a rash lesion for *R. rickettsii*. Examination of a 3-mm punch biopsy from such a lesion is 70% sensitive and 100% specific. The sensitivity of polymerase chain reaction (PCR) amplification and detection of *R. rickettsii* DNA in peripheral blood is improving. However, although rickettsiae are present in large quantities in heavily infected foci of endothelial cells, there are relatively low quantities in the circulation. Cultivation of rickettsiae in cell culture is feasible but is seldom undertaken because of biohazard concerns. The recent dramatic increase in the reported incidence of RMSF correlates with the use of single-titer spotted fever–group cross-reactive enzyme immunoassay serology, with which few cases are specifically determined to be caused by *R. rickettsii*.

TREATMENT: ROCKY MOUNTAIN SPOTTED FEVER

The drug of choice for the treatment of both children and adults with RMSF is doxycycline, except when the patient is pregnant or allergic to this drug (see below). Because of the severity of RMSF, immediate empirical administration of doxycycline should be strongly considered for any patient with a consistent clinical presentation in the appropriate epidemiologic setting. Doxycycline is administered orally (or, in the presence of coma or vomiting, intravenously) at 200 mg/d in two divided doses. For children with suspected RMSF, up to five courses of doxycycline may be administered with minimal risk of dental staining. Other regimens include oral tetracycline (25–50 mg/kg per day) in four divided doses. Treatment with chloramphenicol, a less effective drug, is advised only for patients who are pregnant or allergic to doxycycline. The antirickettsial drug should be administered until the patient has been afebrile and improving clinically for 2–3 days. β-Lactam antibiotics, erythromycin, and aminoglycosides have no role in the treatment of RMSF, and sulfa-containing drugs are likely to exacerbate this infection. There is little clinical experience with fluoroquinolones, clarithromycin, and azithromycin, which are not recommended. The most seriously ill patients are managed in intensive care units, with careful administration of fluids to achieve optimal tissue perfusion without precipitating noncardiogenic pulmonary edema. In some severely ill patients, hypoxemia requires intubation and mechanical ventilation; oliguric or anuric acute renal failure requires hemodialysis; seizures necessitate the use of antiseizure medication; anemia or severe hemorrhage necessitates transfusions of packed red blood cells; or bleeding with severe thrombocytopenia requires platelet transfusions. Heparin is not a useful component of treatment, and there is no evidence that glucocorticoids affect outcome.

PREVENTION

Avoidance of tick bites is the only available preventive approach. Use of protective clothing and tick repellents, inspection of the body once or twice a day, and removal of ticks before they inoculate rickettsiae reduce the risk of infection.

Mediterranean Spotted Fever (Boutonneuse Fever), African Tick-Bite Fever, and Other Tick-Borne Spotted Fevers

R. conorii is prevalent in southern Europe, Africa, and southwestern and south-central Asia. Regional names for the disease caused by this organism include Mediterranean spotted fever, Kenya tick typhus, Indian tick typhus, Israeli spotted fever, and Astrakhan spotted fever. The disease is characterized by high fever, rash, and—in most geographic locales—an inoculation eschar (*tâche noire*) at the site of the tick bite. A severe form of the disease (mortality rate, 50%) occurs in patients with diabetes, alcoholism, or heart failure.

African tick-bite fever, caused by *R. africae,* occurs in rural areas of sub-Saharan Africa and in the Caribbean islands and is transmitted by *Amblyomma hebraeum* and *A. variegatum* ticks. The average incubation period is 4–10 days. The mild illness consists of headache, fever, eschar, and regional lymphadenopathy. *Amblyomma* ticks often feed in groups, with the consequent development of multiple eschars. Rash may be vesicular, sparse, or absent altogether. Because of tourism in sub-Saharan Africa, African tick-bite fever is the most frequently imported rickettsiosis in Europe and North America. A similar disease caused by the very closely related *R. parkeri* is transmitted by *A. maculatum* in the United States and *A. triste* in South America.

R. japonica causes Japanese spotted fever, which also occurs in Korea. Similar diseases in northern Asia are caused by *R. sibirica* and *R. heilongjiangensis.* Queensland tick typhus due to *R. australis* is transmitted by *Ixodes holocyclus.* Flinders Island spotted fever, found on the island for which it is named as well as in Tasmania, mainland Australia, and southeastern Asia, is caused by *R. honei*. In Europe, patients infected with *R. slovaca* after a wintertime *Dermacentor* tick bite manifest an afebrile illness with an eschar (usually on the scalp) and painful regional lymphadenopathy.

DIAGNOSIS

Diagnosis of these tick-borne spotted fevers is based on clinical and epidemiologic findings and is confirmed by serology, immunohistochemical demonstration of rickettsiae in skin biopsy specimens, cell-culture isolation of rickettsiae, or PCR of skin biopsy or blood samples. The serologic identification of the etiologic species requires knowledge of all the potential agents as well as expensive, laborious cross-adsorption of the patient—s serum. In an endemic area, a possible diagnosis of one of these rickettsial spotted fevers should be considered when patients present with fever, rash, and/or a skin lesion consisting of a black necrotic area or a crust surrounded by erythema.

TREATMENT: TICK-BORNE SPOTTED FEVERS

Successful therapeutic agents include doxycycline (100 mg bid orally for 1–5 days), ciprofloxacin (750 mg bid orally for 5 days), and chloramphenicol (500 mg qid orally for 7–10 days). Pregnant patients may be treated with josamycin (3 g/d orally for 5 days). Data on the efficacy of treatment of mildly ill children with clarithromycin or azithromycin should not be extrapolated to adults or to patients with moderate or severe illness.

Rickettsialpox

R. akari infects mice and their mites (*Liponyssoides sanguineus*), which maintain the organisms by transovarian transmission.

EPIDEMIOLOGY

Rickettsialpox is recognized principally in New York City, but cases have also been reported in other urban and rural locations in the United States and in Ukraine, Croatia, Mexico, and Turkey. Investigation of eschars suspected of representing bioterrorism-associated cutaneousanthrax revealed that rickettsialpox occurs more frequently than previously realized.

CLINICAL MANIFESTATIONS

A papule forms at the site of the mite—s feeding, develops a central vesicle, and becomes a 1- to 2.5-cm painless black crusted eschar surrounded by an erythematous halo **(Fig. 174-2)**. Enlargement of the regional lymph nodes draining the eschar suggests initial lymphogenous spread. After an incubation period of 10–17 days, during which the eschar and regional lymphadenopathy frequently go unnoticed, onset is marked by malaise, chills, fever, headache, and myalgia. A macular rash appears 2–6 days after onset and evolves sequentially into papules, vesicles, and crusts that heal without scarring **(Fig. 174-3)**. The rash may remain macular or maculopapular. Some patients develop nausea, vomiting, abdominal pain, cough, conjunctivitis, or photophobia. If untreated, fever lasts 6–10 days.

Figure 174-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Eschar at the site of the mite bite in a patient with rickettsialpox. *(Reprinted from A Krusell et al: Emerg Infect Dis 8:727, 2002. Photo obtained by Dr. Kenneth Kaye.)*

Figure 174-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***Top:*Papulovesicular lesions on the trunk of the patient with rickettsialpox** shown in Fig. 174-2. ***Bottom:*Close-up of lesions** from the same patient. *(Reprinted from A Krusell et al: Emerg Infect Dis 8:727, 2002. Photos obtained by Dr. Kenneth Kaye.)*

DIAGNOSIS AND TREATMENT

Clinical, epidemiologic, and convalescent serologic data establish the diagnosis of a spotted fever–group rickettsiosis that is seldom pursued further. Doxycycline is the drug of choice for treatment.

Flea-Borne Spotted Fever

An emerging rickettsiosis caused by *R. felis* occurs worldwide. Maintained transovarially in the geographically widespread cat flea *Ctenocephalides felis*, the infection has been described as moderately severe, with fever, rash, headache, and CNS, gastrointestinal, and pulmonary symptoms.

Endemic Murine Typhus

EPIDEMIOLOGY

R. typhi is maintained in mammalian host/flea cycles, with rats (*Rattus rattus* and *R. norvegicus*) and the Oriental rat flea (*Xenopsylla cheopis*) as the classic zoonotic niche. Fleas acquire *R. typhi* from rickettsemic rats and carry the organism throughout their life span. Nonimmune rats and humans are infected when rickettsia-laden flea feces contaminates pruritic bite lesions; less frequently, the flea bite transmits the organisms. Transmission also may occur via inhalation of aerosolized rickettsiae from flea feces. Infected rats appear healthy, although they are rickettsemic for ~2 weeks.

Murine typhus occurs mainly in southern Texas and southern California, where the classic rat/flea cycle is absent and an opossum/cat flea (*C. felis*) cycle is prominent. Globally, endemic typhus occurs mainly in warm (often coastal) areas throughout the tropics and subtropics, where it is highly prevalent though often unrecognized. The incidence peaks from April through June in southern Texas and during the warm months of summer and early fall in other geographic locations. Patients seldom recall exposure to fleas, although exposure to animals such as cats, opossums, and rats is reported in nearly 40% of cases.

CLINICAL MANIFESTATIONS

The incubation period of experimental murine typhus averages 11 days (range, 8–16 days). Headache, myalgia, arthralgia, nausea, and malaise develop 1–3 days before onset of chills and fever. Nearly all patients experience nausea and vomiting early in the illness.

The duration of untreated illness averages 12 days (range, 9–18 days). Rash is present in only 13% of patients at presentation for medical care (usually ~4 days after onset of fever), appearing an average of 2 days later in half of the remaining patients and never appearing in the others. The initial macular rash is often detected by careful inspection of the axilla or the inner surface of the arm. Subsequently, the rash becomes maculopapular, involving the trunk more often than the extremities; it is seldom petechial and rarely involves the face, palms, or soles. A rash is detected in only 20% of patients with darkly pigmented skin.

Pulmonary involvement is frequently prominent; 35% of patients have a hacking, nonproductive cough, and 23% of patients who undergo chest radiography have pulmonary densities due to interstitial pneumonia, pulmonary edema, and pleural effusions. Bibasilar rales are the most common pulmonary sign. Less common clinical manifestations include abdominal pain, confusion, stupor, seizures, ataxia, coma, and jaundice. Clinical laboratory studies frequently reveal anemia and leukopenia early in the course, leukocytosis late in the course, thrombocytopenia, hyponatremia, hypoalbuminemia, mildly increased serum hepatic aminotransferases, and prerenal azotemia. Complications may include respiratory failure, hematemesis, cerebral hemorrhage, and hemolysis. Severe illness necessitates the admission of 10% of hospitalized patients to an intensive care unit. Greater severity is generally associated with old age, underlying disease, and treatment with a sulfonamide; the case-fatality rate is 1%. In a study of children with murine typhus, 50% suffered only nocturnal fevers, feeling well enough for active daytime play.

DIAGNOSIS AND TREATMENT

Cultivation, PCR, or cross-adsorption serologic studies of acute- and convalescent-phase sera can provide a specific diagnosis, and an immunohistochemical method for identification of typhus group-specific antigens has been developed. Nevertheless, most patients are treated empirically with doxycycline (100 mg bid orally for 7–15 days) on the basis of clinical suspicion. Ciprofloxacin provides an alternative if doxycycline is contraindicated. Serologic methods are usually used when laboratory confirmation of the diagnosis is sought.

Epidemic (Louse-Borne) Typhus

The human body louse (*Pediculus humanus corporis*) lives in clothing under poor hygienic conditions and usually in impoverished cold areas. Lice acquire *R. prowazekii* when they ingest blood from a rickettsemic patient. The rickettsiae multiply in the midgut epithelial cells of the louse and are shed in the louse—s feces. The infected louse leaves a febrile person and deposits infected feces on its subsequent host during its blood meal; the patient autoinoculates the organisms by scratching. The louse is killed by the rickettsiae and does not pass *R. prowazekii* to its offspring.

Epidemic typhus haunts regions afflicted by wars and disasters. An outbreak involved 100,000 people in refugee camps in Burundi in 1997. A small focus occurred in Russia in 1998; sporadic cases have been reported from Algeria, and frequent outbreaks have occurred in Peru. Eastern flying squirrels (*Glaucomys volans*) and their lice and fleas maintain *R. prowazekii* in a zoonotic cycle. The fleas transmit the infection sporadically to humans.

Brill-Zinsser disease is a recrudescent illness occurring years after acute epidemic typhus, probably as a result of waning immunity. *R. prowazekii* remains latent for years; its reactivation results in sporadic cases of disease in louse-free populations or in epidemics in louse-infested populations.

Rickettsiae are potential agents of bioterrorism (Chap. 221). Infections with *R. prowazekii* and *R. rickettsii* have high case-fatality ratios. These organisms cause difficult-to-diagnose diseases, are highly infectious when inhaled as aerosols, and have been selected for resistance to tetracycline or chloramphenicol in the laboratory.

CLINICAL MANIFESTATIONS

After an incubation period of ~1–2 weeks, the onset of illness is abrupt, with prostration, severe headache, and fever rising rapidly to 38.8°–40.0°C (102°–104°F). Cough is prominent, occurring in 70% of patients. Myalgias are usually severe. In the outbreak in Burundi, the disease was referred to as sutama ("crouching"), a designation reflecting the posture of patients attempting to alleviate the pain. A rash begins on the upper trunk, usually on the fifth day, and then becomes generalized, involving the entire body except the face, palms, and soles. Initially, this rash is macular; without treatment, it becomes maculopapular, petechial, and confluent. The rash often is not detected in black skin; 60% of African patients have spotless epidemic typhus. Photophobia, with considerable conjunctival injection and eye pain, is common. The tongue may be dry, brown, and furred. Confusion and coma are common. Skin necrosis and gangrene of the digits as well as interstitial pneumonia may occur in severe cases. Untreated disease is fatal in 7–40% of cases, with outcome depending primarily on the condition of the host. Patients with untreated infections develop renal insufficiency and multiorgan involvement in which neurologic manifestations are frequently prominent. Overall, 12% of patients with epidemic typhus have neurologic involvement. Infection associated with North American flying squirrels is a milder illness; whether this milder disease is due to host factors (e.g., better health status) or attenuated virulence is unknown.

DIAGNOSIS AND TREATMENT

Epidemic typhus is sometimes misdiagnosed as typhoid fever in tropical countries (Chap. 153). The means even for serologic studies are often unavailable in settings of louse-borne typhus. Epidemics may be recognized by the serologic or immunohistochemical diagnosis of a single case or by detection of *R. prowazekii* in a louse found on a patient. Cross-adsorption indirect fluorescent antibody (IFA) studies can distinguish *R. prowazekii* and *R. typhi* infections. Doxycycline (200 mg/d, given in two divided doses) is administered orally or—if the patient is comatose or vomiting—intravenously. Although under epidemic conditions a single 200-mg dose has proved effective, treatment is generally continued until 2–3 days after defervescence. Pregnant patients should be evaluated individually and treated with either chloramphenicol early in pregnancy or, if necessary, doxycycline late in pregnancy.

PREVENTION

Prevention of epidemic typhus involves control of body lice. Clothes should be changed regularly, and insecticides should be used every 6 weeks to control the louse population.

SCRUB TYPHUS

O. tsutsugamushi differs substantially from *Rickettsia* species both genetically and in terms of cell wall composition (i.e., it lacks lipopolysaccharide). *O. tsutsugamushi* is maintained by transovarian transmission in trombiculid mites. After hatching, infected larval mites (chiggers, the only stage that feeds on a host) inoculate organisms into the skin. Infected chiggers are found particularly in areas of heavy scrub vegetation during the wet season, when mites lay eggs.

Scrub typhus is endemic and reemerging in eastern and southern Asia, northern Australia, and islands of the western Pacific and Indian Oceans. Infections are prevalent in these regions; in some areas, >3% of the population is infected or reinfected each month. Immunity wanes over 1–3 years, and the organism exhibits remarkable antigenic diversity.

CLINICAL MANIFESTATIONS

Illness varies from mild and self-limiting to fatal. After an incubation period of 6–21 days, onset is characterized by fever, headache, myalgia, cough, and gastrointestinal symptoms. Some patients recover spontaneously after a few days. The classic case description includes an eschar where the chigger has fed, regional lymphadenopathy, and a maculopapular rash—signs that are seldom seen in indigenous patients. Fewer than 50% of Westerners develop an eschar, and fewer than 40% develop a rash (on day 4–6 of illness). Severe cases typically include encephalitis and interstitial pneumonia due to vascular injury. The case-fatality rate for untreated classic cases is 7% but would probably be lower if all mild cases were diagnosed.

DIAGNOSIS AND TREATMENT

Serologic assays (IFA, indirect immunoperoxidase, and enzyme immunoassays) are the mainstays of laboratory diagnosis, and PCR amplification of *Orientia* genes from eschars and blood is also effective. Patients are treated with doxycycline (100 mg bid orally for 7–15 days), azithromycin (500 mg orally for 3 days), or chloramphenicol (500 mg qid orally for 7–15 days). Some cases of scrub typhus in Thailand are caused by doxycycline- or chloramphenicol-resistant strains that are susceptible to azithromycin and rifampin.

EHRLICHIOSES AND ANAPLASMOSIS

Ehrlichioses are acute febrile infections caused by members of the family Anaplasmataceae, which is made up of obligately intracellular organisms comprised by four genera: *Ehrlichia*, *Anaplasma*, *Wolbachia*, and *Neorickettsia*. The bacteria reside in vertebrate reservoirs and target vacuoles of hematopoietic cells **(Fig. 174-4)**. Two *Ehrlichia* species and one *Anaplasma* species are transmitted by ticks to humans and cause infection that can be severe and prevalent. *E. chaffeensis*, the agent of HME, infects predominantly mononuclear phagocytes; *E. ewingii* and *A. phagocytophilum* infect neutrophils.

Figure 174-4

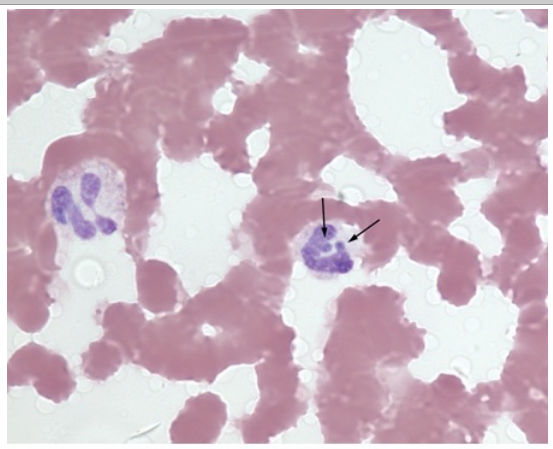

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Peripheral blood smear from a patient with human granulocytotropic anaplasmosis. A neutrophil contains two morulae (vacuoles filled with *A. phagocytophilum*). *(Photo courtesy of Dr. J. Stephen Dumler.)*

Ehrlichia and *Anaplasma* are maintained by horizontal tick-mammal-tick transmission, and humans are only inadvertently infected. Wolbachiae are associated with human filariasis, since they are important for filarial viability and pathogenicity; antibiotic treatment targeting wolbachiae is a strategy for filariasis control. Neorickettsiae parasitize flukes that in turn parasitize aquatic snails, fish, and insects. Only a single human neorickettsiosis has been described: sennetsu fever, an infectious mononucleosis–like illness that was first identified in 1953 and is probably due to the ingestion of raw fish containing *N. sennetsu*–infected flukes.

Human Monocytotropic Ehrlichiosis

EPIDEMIOLOGY

More than 5496 cases of *E. chaffeensis* infection had been reported to the Centers for Disease Control and Prevention (CDC) as of November 2009. However, active prospective surveillance has demonstrated an incidence as high as 414 cases per 100,000 population in some U.S. regions. Most *E. chaffeensis* infections are identified in the south-central, southeastern, and mid-Atlantic states, but cases have also been recognized in California, New York, and Minnesota. All stages of the Lone Star tick *(A. americanum)* feed on white-tailed deer—a major reservoir. Dogs and coyotes also serve as reservoirs and often lack clinical signs. Tick bites and exposures are frequently reported by patients in rural areas especially in May through July. The median age of HME patients is 53 years; however, severe and fatal infections in children are also well recognized. Of patients with HME, 61% are male.

CLINICAL MANIFESTATIONS

E. chaffeensis disseminates hematogenously from the dermal blood pool created by the feeding tick. After a median incubation period of 8 days, illness develops. Clinical manifestations are undifferentiated and include fever (96% of cases), headache (72%), myalgia (68%), and malaise (77%). Less frequently observed are nausea, vomiting, and diarrhea (25–57%); cough (28%); rash (26% overall, 6% at presentation); and confusion (20%). HME can be severe: 62% of patients with documented cases are hospitalized, and ~3% die. Severe manifestations include toxic shock–like or septic shock–like syndrome, adult respiratory distress syndrome, cardiac failure, hepatitis, meningoencephalitis, hemorrhage, and—in immunocompromised patients—overwhelming ehrlichial infection. Laboratory findings are valuable in the differential diagnosis of HME; 61% of patients have leukopenia (initially lymphopenia, later neutropenia), 73% have thrombocytopenia, and 84% have elevated serum levels of hepatic aminotransferases. Despite low blood cell counts, the bone marrow is hypercellular, and noncaseating granulomas may be present. Vasculitis is not a component of HME.

DIAGNOSIS

Because HME can be fatal, empirical antibiotic therapy based on clinical diagnosis is required. This diagnosis is suggested by fever with a known tick exposure during the preceding 3 weeks, thrombocytopenia and/or leukopenia, and increased serum aminotransferase levels. Morulae are demonstrated on <10% of peripheral-blood smears. Acute HME can be confirmed by PCR amplification of *E. chaffeensis* nucleic acids in blood obtained before the start of doxycycline therapy. Retrospective serodiagnosis requires a consistent clinical picture and a fourfold increase in *E. chaffeensis* antibody titer to ≥1:64 in paired sera obtained ~3 weeks apart. Separate specific diagnostic tests are necessary for HME and HGA.

Ewingii Ehrlichiosis

Ehrlichia ewingii, originally a neutrophilpathogen causing fever and lameness in dogs, resembles *E. chaffeensis* in its tick vector (*A. americanum*) and vertebrate reservoirs (white-tailed deer and dogs). *E. ewingii* illness is similar to but less severe than HME. The majority of cases have occurred in immunocompromised patients. No specific diagnostic test for ewingii ehrlichiosis is readily available.

TREATMENT: EHRLICHIOSES

Doxycycline is effective for HME and ewingii ehrlichiosis. Therapy with doxycycline (100 mg given orally or intravenously twice daily) or tetracycline (250–500 mg given orally every 6 h) lowers hospitalization rates and shortens fever duration. *E. chaffeensis* is not

susceptible to chloramphenicol in vitro, and the use of this drug is controversial. While a few reports document *E. chaffeensis* persistence in humans, this finding is rare; most infections are cured by short courses of doxycycline (continuing for 3–5 days after defervescence). Although poorly studied, rifampin may be suitable when doxycycline is contraindicated.

PREVENTION

HME and ewingii ehrlichiosis are prevented by the avoidance of ticks in endemic areas. The use of protective clothing and tick repellents, careful postexposure tick searches, and prompt removal of attached ticks probably diminish infection risk.

Human Granulocytotropic Anaplasmosis

EPIDEMIOLOGY

As of November 2009, 6218 cases of HGA had been reported to the CDC, most in the upper midwestern and northeastern United States; the geographic distribution is similar to that for Lyme disease because of the shared *I. scapularis* tick vector. White-footed mice, squirrels, and white-tailed deer in the United States and red deer in Europe are natural reservoirs for *A. phagocytophilum*. HGA incidence peaks in May through July, but the disease can occur throughout the year with exposure to *Ixodes* ticks. HGA often affects males (57%) and older persons (median age, 51 years).

CLINICAL MANIFESTATIONS

Seroprevalence rates are high in endemic regions; thus it seems likely that most individuals develop subclinical infections. The incubation period for HGA is 4–8 days, after which the disease manifests as fever (91% of cases), myalgia (77%), headache (77%), and malaise (94%). A minority of patients develop nausea, vomiting, or diarrhea (16–38%); cough (21%); or confusion (17%). Rash (6%) is almost invariably concurrent erythema migrans attributable to Lyme disease. Most patients develop thrombocytopenia (69%) and/or leukopenia (48%) with increased serum hepatic aminotransferase levels (71%).

Severe complications occur most often in the elderly and include adult respiratory distress syndrome, toxic shock–like syndrome, and life-threatening opportunistic infections. Meningoencephalitis has not been conclusively documented with HGA, but brachial plexopathy and demyelinating polyneuropathy are reported. For HGA, 7% of patients require intensive care, and the case-fatality rate is 0.5%. Neither vasculitis nor granulomas are components of HGA. While co-infections with *Borrelia burgdorferi* and *Babesia microti* [transmitted by the same tick vector(s)] occur, there is little evidence of comorbidity or persistence.

DIAGNOSIS

HGA should be included in the differential diagnosis of influenza-like illnesses during seasons with *Ixodes* tick activity (May through December), especially with tick bite or exposure. Concurrent thrombocytopenia, leukopenia, or elevation in serum alanine or aspartate aminotransferase further increases the likelihood of HGA. Many HGA patients develop Lyme disease antibodies in the absence of clinical findings consistent with that diagnosis. Thus, HGA should be considered in the differential diagnosis of atypical severe Lyme disease presentations. Peripheral-blood film examination for neutrophil morulae can yield a diagnosis in 20–75% of infections. PCR testing of blood from patients with active disease before doxycycline therapy is sensitive and specific. Serodiagnosis is retrospective, requiring a fourfold increase in *A. phagocytophilum* antibody titer (to ≥1:80) in paired serum samples obtained 1 month apart. Since seroprevalence is high in some regions, a single acute-phase titer should not be used for diagnosis.

TREATMENT: HUMAN GRANULOCYTOTROPIC ANAPLASMOSIS

No prospective studies of therapy for HGA have been conducted. However, doxycycline (100 mg by mouth twice daily) is effective. Rifampin therapy is associated with improvement of HGA in pregnant women and children. Most treated patients defervesce within 24–48 h.

PREVENTION

HGA prevention requires tick avoidance. Transmission can be documented as few as 4 h after a tick bite.

Q FEVER

The agent of Q fever is *C. burnetii*, a small intracellular microorganism that only recently was grown in cell-free medium. *C. burnetii*, a pleomorphic coccobacillus with a gram-negative cell wall, survives in harsh environments; it escapes intracellular killing in macrophages by inhibiting the final phagosome maturation step (cathepsin fusion) and has adapted to the acidic phagolysosome by producing superoxide dismutase. Infection with *C. burnetii* induces a range of immunomodulatory responses, from immunosuppression in chronic Q fever to the production of autoantibodies, particularly those to smooth muscle and cardiac muscle.

Q fever encompasses two broad clinical syndromes: acute and chronic infection. The host—s immune response (rather than the particular strain) most likely determines whether chronic Q fever develops. *C. burnetii* survives in monocytes from patients with chronic Q fever but not in monocytes from patients with acute Q fever or from uninfected subjects. Impairment of the bactericidal activity of the *C. burnetii*–infected monocyte is associated with overproduction of interleukin 10. The CD4+/CD8+ ratio is decreased in Q fever endocarditis. Very few organisms and a strong cellular response are observed in patients with acute Q fever, while many organisms and a moderate cellular response occur in chronic Q fever. Immune control of *C. burnetii* is T cell–dependent, but 80–90% of bone marrow aspirates obtained years after recovery from Q fever contain *C. burnetii* DNA.

EPIDEMIOLOGY

Q fever is a zoonosis. The primary sources of human infection are infected cattle, sheep, and goats. However, cats, rabbits, pigeons, and dogs have also served as sources for transmission of *C. burnetii* to humans. The wildlife reservoir is extensive and includes ticks, coyotes, gray foxes, skunks, raccoons, rabbits, deer, mice, bears, birds, and opossums. In female animals *C. burnetii* localizes to the uterus and mammary glands. Infection is reactivated during pregnancy and following radiotherapy in mouse models. High concentrations of *C. burnetii* are found in the placenta. At the time of parturition, the bacteria are released into the air, and infection follows inhalation of aerosolized organisms by a susceptible host. Windstorms can generate *C. burnetii* aerosols months after soil contamination during parturition. Individuals up to 18 km from the source have been infected. Because it is easily dispersed as an aerosol, *C. burnetii* is a potential agent of bioterrorism, with a high infectivity rate and pneumonia as the major manifestation.

Determining the source of an outbreak of Q fever can be challenging. An outbreak of Q fever at a horse-boarding ranch in Colorado in 2005 was due to spread of infection from two herds of goats that had been acquired by the owners. PCR testing confirmed the presence of *C. burnetii* in the soil and among the goats. Of 138 persons who lived within 1 mile of the ranch and who were also tested, 11 (8%) had evidence of *C. burnetii* infection, and 8 of these 11 individuals had had no direct contact with the ranch.

Persons at risk for Q fever include abattoir workers, veterinarians, farmers, and other individuals who have contact with infected animals, particularly newborn animals, or products of conception. The organism is shed in milk for weeks to months after parturition. The ingestion of contaminated milk in some geographic areas probably represents a major route of transmission to humans, although experimental evidence on this point is contradictory. In rare instances, human-to-human transmission has followed labor and childbirth in an infected woman, autopsy of an infected individual, or blood transfusion. Some evidence suggests that *C. burnetii* can be sexually transmitted among humans. Crushing an infected tick between the fingers has resulted in Q fever; the implication is that percutaneous transmission can occur.

Infections due to *C. burnetii* occur in most geographic locations except New Zealand and Antarctica. Thus Q fever can be associated with travel. The number of reported cases of Q fever in the United States ranges from 28 to 54 per year. More than 70% of these cases occur in males, and April through June is the most common time for acquisition. Q fever continues to be common in Australia,

with 30 cases per 1 million population per year. Cases among abattoir workers in Australia declined dramatically due to a vaccination program. An outbreak of Q fever began in the Netherlands in 2007, and by August 2009 more than 2000 cases had been reported. Pneumonia was a common manifestation in this outbreak, which appeared to be primarily due to transmission from infected goats.

The primary manifestations of acute Q fever differ geographically (e.g., pneumonia in Nova Scotia and granulomatous hepatitis in Marseille). These differences may reflect the route of infection (i.e., ingestion of contaminated milk for hepatitis and inhalation of contaminated aerosols for pneumonia).

Young age seems to be protective against infection with *C. burnetii*. In a large outbreak in Switzerland, symptomatic infection occurred five times more often among persons >15 years of age than among younger individuals. In many outbreaks, men are affected more commonly than women; the proposed explanation is that female hormones are partially protective.

CLINICAL MANIFESTATIONS

Acute Q Fever

After an incubation period of 3–30 days, 1070 patients with acute Q fever in southern France presented with hepatitis (40%), both pneumonia and hepatitis (20%), pneumonia (17%), isolated fever (14%), CNS involvement (2%), and pericarditis or myocarditis (1%). Acalculous cholecystitis, pancreatitis, lymphadenopathy, spontaneous rupture of the spleen, transient hypoplastic anemia, bone marrow necrosis, hemolytic anemia, histiocytic hemophagocytosis, optic neuritis, and erythema nodosum are less common manifestations.

The symptoms of acute Q fever are nonspecific; common among them are fever, extreme fatigue, photophobia, and severe headache that is frequently retro-orbital. Other symptoms include chills, sweats, nausea, vomiting, and diarrhea, each occurring in 5–20% of cases. Cough develops in about half of patients with Q fever pneumonia. Neurologic manifestations of acute Q fever are uncommon; however, in one outbreak in the United Kingdom, 23% of 102 patients had neurologic signs and symptoms as the major manifestation. A nonspecific rash may be evident in 4–18% of patients. The WBC count is usually normal. Thrombocytopenia occurs in ~25% of patients, and reactive thrombocytosis (with platelet counts exceeding $10^6/\mu L$) frequently develops during recovery. Chest radiography may show opacities similar to those seen in pneumonia caused by other pathogens, but multiple rounded opacities in patients in endemic areas suggest a diagnosis of Q fever pneumonia.

Acute Q fever occasionally complicates pregnancy. In one series, it resulted in premature birth in 35% of cases and in abortion or neonatal death in 43%. Neonatal death (previous or current) and lower infant birth weight are three times more likely among women seropositive for *C. burnetii*.

Post–Q Fever Fatigue Syndrome

Prolonged fatigue can follow Q fever and can be accompanied by a constellation of symptoms including headaches, sweats, arthralgia, myalgias, blurred vision, muscle fasciculations, and enlarged and painful lymph nodes. Long-term persistence of a noninfective, nonbiodegraded complex of *Coxiella* cell components with its antigens and specific lipopolysaccharide has been detected in the affected persons. Patients who develop this syndrome have a higher frequency of carriage of HLA-DRB1*11 and of the 2/2 genotype of the interferon γ intron 1 microsatellite.

Chronic Q Fever

Chronic Q fever almost always implies endocarditis and usually occurs in patients with previous valvular heart disease, immunosuppression, or chronic renal insufficiency. Fever is usually absent or low grade. Valvular vegetations are detected in only 12% of patients by transthoracic echocardiography, but the rate of detection is higher with transesophageal echocardiography. The vegetations in chronic Q fever endocarditis differ from those in bacterial endocarditis, manifesting as endothelium-covered nodules on the valves. A high index of suspicion is necessary for timely diagnosis. Patients with chronic Q fever are often ill for >1 year before the diagnosis is made. The disease should be suspected in all patients with culture-negative endocarditis. In addition, all patients with valvular heart disease and an unexplained purpuric eruption, renal insufficiency, stroke, and/or progressive heart failure should be tested for *C. burnetii* infection. Patients with chronic Q fever have hepatomegaly and/or splenomegaly, which, in combination with rheumatoid factor, elevated erythrocyte sedimentation rate, high C-reactive protein level, and/or increased γ-globulin concentrations (up to 60–70 g/L), suggests this diagnosis. Other manifestations of chronic Q fever include infection of vascular prostheses, aneurysms, and bone as well as chronic sternal wound infection. Unusual manifestations include chronic thrombocytopenia, mixed cryoglobulinemia, and livedo reticularis.

DIAGNOSIS

Isolation of *C. burnetii* from buffy-coat blood samples or tissue specimens by a shell-vial technique is easy but requires a biosafety level 3 laboratory. PCR detects *C. burnetii* DNA in tissue specimens, including paraffin-embedded samples. Serology is the most commonly used diagnostic tool. Indirect immunofluorescence is sensitive and specific and is the method of choice. Rheumatoid factor should be adsorbed from the specimen before testing. An IgG antibody titer of ≥1:800 to phase I antigen (i.e., naturally occurring *C. burnetii* with intact lipopolysaccharide) is suggestive of chronic Q fever; in chronic infection, the titer to phase I antigen is usually much higher than that to phase II antigen (i.e., *C. burnetii* that has truncated lipopolysaccharide associated with gene deletions during laboratory passages). The reverse is true in acute Q fever, in which a fourfold rise in titer may be demonstrated between acute- and convalescent-phase serum samples.

TREATMENT: Q FEVER

Treatment of acute Q fever with doxycycline (100 mg twice daily for 14 days) is usually successful. Quinolones are also effective. When Q fever is diagnosed during pregnancy, treatment with trimethoprim-sulfamethoxazole (TMP-SMX) is recommended for the duration of the pregnancy. One study showed no intrauterine fetal deaths and substantial reduction of obstetric complications in a group of Q fever patients treated with TMP-SMX.

The treatment of chronic Q fever is difficult and requires careful follow-up. Addition of hydroxychloroquine (to alkalinize the phagolysosome) renders doxycycline bactericidal against *C. burnetii,* and this combination is currently the favored regimen. Treatment with doxycycline (100 mg bid) and hydroxychloroquine (200 mg tid; plasma concentration maintained at 0.8–1.2 μg/mL) for 18 months is superior to a regimen of doxycycline and ofloxacin. Among 21 patients who received doxycycline and hydroxychloroquine, 1 died of a surgical complication, 2 were still being treated at the end of the study, 1 was still being evaluated, and 17 were cured. The mean duration of treatment was 31 months. In the ofloxacin and doxycycline group of 14 patients, 1 had died, 1 was still being treated, 7 had relapsed, and 5 had been cured by the end of the study. Optimal management of Q fever endocarditis entails determining the minimal inhibitory concentration (MIC) of doxycycline for the patient—s isolate and measuring serum doxycycline levels. A serum level–to–doxycycline MIC ratio of ≥1 is associated with a rapid decline in phase I antibodies with the doxycycline-hydroxychloroquine regimen. Patients treated with this regimen must be advised about photosensitivity and retinal toxicity risks. The doxycycline-hydroxychloroquine regimen was successful in one patient with HIV infection and Q fever endocarditis. The Jarisch-Herxheimer reaction occasionally complicates the treatment of chronic Q fever. Treatment of *C. burnetii*–infected aortic aneurysms is the same as that for Q fever endocarditis. Surgical intervention is often required.

If doxycycline-hydroxychloroquine cannot be used, the regimen chosen should include at least two antibiotics active against *C. burnetii*. Rifampin (300 mg once daily) combined with doxycycline (100 mg twice daily) or ciprofloxacin (750 mg twice daily) has been used successfully. The optimal duration of antibiotic therapy for chronic Q fever remains undetermined. At least 3 years of treatment, with discontinuation only if the phase I IgA antibody titer is ≤1:50 and the phase I IgG titer is ≤1:200, is recommended.

Patients with acute Q fever and lesions of native heart valves (e.g., bicuspid aortic valve), prosthetic valves, or prosthetic

intravascular material should undergo serologic monitoring every 4 months for 2 years. If the phase I IgG titer is >1:800, further investigation is warranted. Some authorities recommend that patients with valvulopathy and acute Q fever receive doxycycline and hydroxychloroquine to prevent chronic Q fever. For women who exhibit a serologic profile of chronic Q fever after childbirth, hydroxychloroquine and doxycycline should be given for 1 year.

Therapy with Biologic Modifying Agents

Interferon γ was successful in the treatment of a 3-year-old boy with prolonged fever, abdominal pain, and thrombocytopenia due to *C. burnetii* that had not been eradicated with conventional antibiotic therapy. Many patients with granulomatous hepatitis due to Q fever have a prolonged febrile illness that is unresponsive to antibiotics. For these individuals, treatment with prednisone (0.5 mg/kg) has resulted in defervescence within 2–15 days. After defervescence, the glucocorticoid dose is tapered over the next month.

PREVENTION

A whole-cell vaccine (Q-Vax) licensed in Australia effectively prevents Q fever in abattoir workers. Before administration of the vaccine, skin testing with intradermal diluted *C. burnetii* vaccine is performed, serologic testing is undertaken, and a history of possible Q fever is sought. Vaccine is given only to patients with no history of Q fever and negative results in serologic and skin testing.

Good animal-husbandry practices are important in preventing widespread contamination of the environment by *C. burnetii*. These practices include isolating aborting animals for up to 14 days, raising feed bunks to prevent contamination of feed by excreta, destroying aborted materials (by burning and burying fetal membranes and stillborn animals), and wearing masks and gloves when handling aborted materials. Only seronegative pregnant animals should be used in research settings, and only seronegative animals should be permitted in petting zoos.

During an outbreak of Q fever and for 4 weeks after it ceases, blood donations should not be accepted from individuals who live in the affected area.

Acknowledgment

The contributions of Didier Raoult, MD, to this chapter in previous editions are gratefully acknowledged.

FURTHER READINGS

Bakken JS, Dumler S: Human granulocytic anaplasmosis. Infect Dis Clin North Am 22:433, 2008[PMID: 18755383]

Bechah Y et al: Epidemic typhus. Lancet Infect Dis 8:417, 2008[PMID: 18582834]

Chapman AS et al: Cluster of sylvatic epidemic typhus cases associated with flying squirrels, 2004–2006. Emerg Infect Dis 15:1005, 2009[PMID: 19624912]

——— et al: Diagnosis and management of tickborne rickettsial diseases: Rocky Mountain spotted fever, ehrlichioses, and anaplasmosis—United States. MMWR Recomm Rep 55:1, 2006

De Sousa R et al: Host- and microbe-related risk factors for and pathophysiology of fatal Rickettsia conorii infection in Portuguese patients. J Infect Dis 198:576, 2008

Harris RJ et al: Long-term persistence of *Coxiella burnetii* in the host after primary Q fever. Epidemiol Infect 124:543, 2000[PMID: 10982079]

Hechemy KE et al (eds): *Century of Rickettsiology: Emerging, Reemerging Rickettsioses, Molecular Diagnostics, and Emerging Veterinary Rickettsioses*. Malden, MA, Blackwell Scientific, 2006

Holman RC et al: Analysis of risk factors for fatal Rocky Mountain spotted fever: Evidence for superiority of tetracyclines for therapy. J Infect Dis 184:1437, 2001[PMID: 11709786]

Mcquiston JH et al: National surveillance and the epidemiology of Q fever in the United States, 1978–2004. Am J Trop Med Hyg 75:36, 2006[PMID: 16837706]

Paddock CD et al: Rickettsia parkeri rickettsiosis and its clinical distinction from Rocky Mountain spotted fever. Clin Infect Dis 47:1188, 2008[PMID: 18808353]

Raoult D et al: Q fever 1985–1998. Clinical and epidemiologic features of 1,383 infections. Medicine (Baltimore) 79:109, 2000[PMID: 10771709]

Walker DH: Rickettsiae and rickettsial infections: The current state of knowledge. Clin Infect Dis 45(Suppl1):539, 2007

——— et al: Emerging and re-emerging tick-transmitted rickettsial and ehrlichial infections. Med Clin North Am 92:1345, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS DUE TO MYCOPLASMAS: INTRODUCTION

Mycoplasmas are prokaryotes of the class Mollicutes. Their size (150–350 nm) is closer to that of viruses than to that of bacteria. Unlike viruses, however, mycoplasmas grow in cell-free culture media; in fact, they are the smallest organisms capable of independent replication.

The entire genomes of many *Mycoplasma* species have been sequenced and have been found to be among the smallest of all prokaryotic genomes. Sequencing information for these genomes has helped define the minimal set of genes necessary for cellular life. The absence of genes related to the synthesis of amino acids, fatty acid metabolism, and cholesterol dictates the mycoplasmas— parasitic or saprophytic dependence on a host for exogenous nutrients and necessitates the use of complex fastidious media to culture these organisms. Mycoplasmas lack a cell wall and are bound only by a cell membrane. The absence of a cell wall explains the inactivity of β-lactam antibiotics (penicillins and cephalosporins) against infections caused by these organisms.

At least 13 *Mycoplasma* species, two *Acholeplasma* species, and two *Ureaplasma* species have been isolated from humans. Most of these species are thought to be normal inhabitants of oral and urogenital mucous membranes. Only four species—*M. pneumoniae*, *M. hominis*, *U. urealyticum*, and *U. parvum*—have been shown conclusively to be pathogenic in immunocompetent humans. *M. pneumoniae* primarily infects the respiratory tract, while *M. hominis*, *U. urealyticum*, and *U. parvum* are associated with a variety of genitourinary tract disorders and neonatal infections. Some data indicate that *M. genitalium* may be a cause of disease in humans. Other mycoplasmas may cause disease in immunocompromised persons.

MYCOPLASMA PNEUMONIAE

Pathogenicity

M. pneumoniae is generally thought to act as an extracellular pathogen. Although the organism has been shown to exist and replicate within human cells, it is not known whether these intracellular events contribute to the pathogenesis of disease. *M. pneumoniae* attaches to ciliated respiratory epithelial cells by means of a complex terminal organelle at the tip of one end of the organism. Cytoadherence is mediated by interactive adhesins and accessory proteins clustered on this organelle. After extracellular attachment, *M. pneumoniae* causes injury to host respiratory tissue. The mechanism of injury is thought to be mediated by the production of hydrogen peroxide and of a recently identified ADP-ribosylating and vacuolating cytotoxin of *M. pneumoniae* that has many similarities to pertussis toxin. Because mycoplasmas lack a cell wall, they also lack cell wall–derived stimulators of the innate immune system, such as lipopolysaccharide, lipoteichoic acid, and murein (peptidoglycan) fragments. However, lipoproteins from the mycoplasmal cell membrane appear to have inflammatory properties, probably acting through Toll-like receptors (primarily TLR2) on macrophages and other cells. Lung biopsy specimens from patients with *M. pneumoniae* respiratory tract infection reveal an inflammatory process involving the trachea, bronchioles, and peribronchial tissue, with a monocytic infiltrate coinciding with a luminal exudate of polymorphonuclear leukocytes.

Experimental evidence indicates that innate immunity provides most of the host—s defense against mycoplasmal infection in the lungs, whereas cellular immunity may actually play an immuno-pathogenic role, exacerbating mycoplasmal lung disease. Humoral immunity appears to provide protection against dissemination of *M. pneumoniae* infection; patients with humoral immunodeficiencies do not have more severe lung disease than do immunocompetent patients in the early stages of infection but more often develop disseminated infection resulting in syndromes such as arthritis, meningitis, and osteomyelitis. The immunity that follows severe *M. pneumoniae* infections is more protective and longer-lasting than that following mild infections. Genuine second attacks of *M. pneumoniae* pneumonia have been reported infrequently.

Epidemiology

M. pneumoniae infection occurs worldwide. It is likely that the incidence of upper respiratory illness due to *M. pneumoniae* is up to 20 times that of pneumonia caused by this organism. Infection is spread from one person to another by respiratory droplets expectorated during coughing and results in clinically apparent disease in an estimated 80% of cases. The incubation period for *M. pneumoniae* is 2–4 weeks; therefore, the time-course of infection in a specific population may be several weeks long.

Intrafamilial attack rates are as high as 84% among children and 41% among adults. Outbreaks of *M. pneumoniae* illness often occur in institutional settings such as military bases, boarding schools, and summer camps. Infections tend to be endemic, with sporadic epidemics every 4–7 years. There is no seasonal pattern.

Most significantly, *M. pneumoniae* is a major cause of community-acquired respiratory illness in both children and adults and is often grouped with *Chlamydophila pneumoniae* and *Legionella* species as being among the most important bacterial causes of "atypical" community-acquired pneumonia. For community-acquired pneumonia in adults, *M. pneumoniae* is the most frequently detected "atypical" organism. Analysis of 13 studies of community-acquired pneumonia published since 1995 (which included 6207 ambulatory and hospitalized adults) showed that the overall prevalence of *M. pneumoniae* was 22.7%; by comparison, the prevalence of *C. pneumoniae* was 11.7%, and that of *Legionella* species was 4.6%. *M. pneumoniae* pneumonia is also referred to as Eaton agent pneumonia (the organism having first been isolated in the early 1940s by Monroe Eaton), primary atypical pneumonia, and "walking" pneumonia.

Clinical Manifestations

UPPER RESPIRATORY TRACT INFECTIONS AND PNEUMONIA

Acute *M. pneumoniae* infections generally manifest as pharyngitis, tracheobronchitis, reactive airway disease/wheezing, or a nonspecific upper respiratory syndrome. Little evidence supports the commonly held belief that this organism is an important cause of otitis media, with or without bullous myringitis. Pneumonia develops in 3–13% of infected individuals; its onset is usually gradual, occurring over several days, but may be more abrupt. Although *Mycoplasma* pneumonia may begin with a sore throat, the most common presenting symptom is cough. The cough is typically nonproductive, but some patients produce sputum. Headache, malaise, chills, and fever are noted in the majority of patients.

On physical examination, wheezes or rales are detected in ~80% of patients with *M. pneumoniae* pneumonia. In many patients, however, pneumonia can be diagnosed only by chest radiography. The most common radiographic pattern is that of peribronchial pneumonia with thickened bronchial markings, streaks of interstitial infiltration, and areas of subsegmental atelectasis. Segmental or lobar consolidation is not uncommon. While clinically evident pleural effusions are uncommon, lateral decubitus views reveal that up to 20% of patients have pleural effusions.

Overall, the clinical presentation of pneumonia in an individual patient is not useful for differentiating *M. pneumoniae* pneumonia from other types of community-acquired pneumonia. The possibility of *M. pneumoniae* infection deserves particular consideration when community-acquired pneumonia fails to respond to treatment with a penicillin or a cephalosporin—antibiotics that are ineffective against mycoplasmas. Symptoms usually resolve within 2–3 weeks after the onset of illness. Although *M. pneumoniae* pneumonia is generally self-limited, appropriate antimicrobial therapy significantly shortens the duration of clinical illness. Infection uncommonly results in critical illness and only rarely in death. In some patients, long-term recurrent wheezing may follow the resolution of acute pneumonia. The significance of chronic infection, especially as it relates to asthma, is an area of active investigation.

EXTRAPULMONARY MANIFESTATIONS

An array of extrapulmonary manifestations may develop during *M. pneumoniae* infection. The most significant are neurologic, dermatologic, cardiac, rheumatologic, and hematologic in nature. Extrapulmonary manifestations can be a result of active infection (e.g., septic arthritis) or postinfectious autoimmune phenomena (e.g., Guillain-Barré syndrome). Overall, these manifestations are uncommon, given the frequency of *M. pneumoniae* infection. Notably, many patients with extrapulmonary *M. pneumoniae* disease do not have respiratory disease.

Skin eruptions described with *M. pneumoniae* infection include erythematous (macular or maculopapular), vesicular, bullous, petechial, and urticarial rashes. In some reports, 17% of patients with *M. pneumoniae* pneumonia have had an exanthem. Erythema multiforme major (Stevens-Johnson syndrome) is the most clinically significant skin eruption associated with *M. pneumoniae* infection; it appears to occur more commonly with *M. pneumoniae* than with other infectious agents.

A wide spectrum of neurologic manifestations has been reported with *M. pneumoniae* infection. The most common are meningoencephalitis, encephalitis, Guillain-Barré syndrome, and aseptic meningitis. *M. pneumoniae* has been implicated as a likely etiologic agent in 5–7% of cases of encephalitis. Other neurologic manifestations may include cranial neuropathy, acute psychosis, cerebellar ataxia, acute demyelinating encephalomyelitis, cerebrovascular thromboembolic events, and transverse myelitis. The roles of antimicrobial drugs, glucocorticoids, and IV immunoglobulin in the treatment of neurologic disease due to *M. pneumoniae* remain unknown.

Hematologic manifestations of *M. pneumoniae* infection include hemolytic anemia, aplastic anemia, cold agglutinins, disseminated intravascular coagulation, and hypercoagulopathy. When anemia does occur, it generally develops in the second or third week of illness.

In addition, hepatitis, glomerulonephritis, pancreatitis, myocarditis, pericarditis, rhabdomyolysis, and arthritis (septic and reactive) have been convincingly ascribed to *M. pneumoniae* infection. Septic arthritis has been described most commonly in hypogammaglobulinemic patients.

Diagnosis (Table 175-1)

Table 175-1 Diagnostic Tests for Respiratory *Mycoplasma Pneumoniae* Infection[a]

Test	Sensitivity, %	Specificity, %	Comment
Respiratory culture	≤60	100	—
Respiratory PCR	65–90	90–100	—
Serologic studies	55–100	55–100	Acute- and convalescent-phase serum samples are recommended.

[a]A combination of PCR and serology is suggested for routine diagnosis. If macrolide resistance is suspected, *M. pneumoniae* culture may prove useful, providing an isolate for susceptibility testing.

Abbreviation: PCR, polymerase chain reaction.

Clinical findings, nonmicrobiologic laboratory tests, and chest radiography are not useful for differentiating *M. pneumoniae* pneumonia from other types of community-acquired pneumonia. In addition, since *M. pneumoniae* lacks a cell wall, it is not visible on Gram—s stain. Although of historical interest, the measurement of cold agglutinin titers is no longer recommended for the diagnosis of *M. pneumoniae* infection because the findings are nonspecific and assays specific for *M. pneumoniae* are now available.

Acute *M. pneumoniae* infection can be diagnosed by polymerase chain reaction (PCR) detection of the organism in respiratory tract secretions or by isolation of the organism in culture. Oropharyngeal, nasopharyngeal, and pulmonary specimens are all acceptable for diagnosing *M. pneumoniae* pneumonia. Other bodily fluids, such cerebrospinal fluid, are acceptable for extrapulmonary infection. *M. pneumoniae* culture (which requires special media) is not recommended for routine diagnosis because the organism may take weeks to grow and is often difficult to isolate from clinical specimens. In contrast, PCR allows rapid, specific diagnosis earlier in the course of clinical illness.

The diagnosis can also be established by serologic tests for IgM and IgG antibodies to *M. pneumoniae* in paired (acute- and convalescent-phase) serum samples; enzyme-linked immunoassay is the recommended serologic method. An acute-phase sample alone is not adequate for diagnosis, as antibodies to *M. pneumoniae* may not develop until 2 weeks into the illness; therefore, it is important to test paired samples. In addition, IgM antibody to *M. pneumoniae* can persist for up to 1 year after acute infection. Thus its presence may indicate recent rather than acute infection.

The combination of PCR of respiratory tract secretions and serologic testing constitutes the most sensitive and rapid approach to the diagnosis of *M. pneumoniae* infection.

Treatment: *Mycoplasma Pneumoniae* Infections (Table 175-2)

Table 175-2 Antimicrobial Agents of Choice for Mycoplasma Infections [a]

Organism	Drug(s)
M. pneumoniae	Azithromycin, clarithromycin, erythromycin, doxycycline, levofloxacin, moxifloxacin, gemifloxacin (*not*ciprofloxacin)
U. urealyticum, U. parvum	Azithromycin, clarithromycin, erythromycin, doxycycline
M. hominis	Doxycycline, clindamycin
M. genitalium	Azithromycin

[a]Antimicrobial resistance has been reported for mycoplasmas, as described in the text.

Although in the majority of untreated cases symptoms resolve within 2–3 weeks without significant associated morbidity, *M. pneumoniae* pneumonia can be a serious illness that responds to appropriate antimicrobial therapy. Randomized, double-blind, placebo-controlled trials have demonstrated that antimicrobial treatment significantly decreases the duration of fever, cough, malaise, hospitalization, and radiologic abnormalities in *M. pneumoniae* pneumonia. Treatment options for acute *M. pneumoniae* infection include macrolides (e.g., oral azithromycin, 500 mg on day 1, then 250 mg/d on

days 2–5), tetracyclines (e.g., oral doxycycline, 100 mg twice daily for 10–14 days), and respiratory fluoroquinolones. However, ciprofloxacin and ofloxacin are *not* recommended because of their high minimal inhibitory concentrations against *M. pneumoniae* isolates and their poor performance in experimental studies. A 10- to 14-day course of therapy appears adequate.

In Japan and China, high levels of *M. pneumoniae* resistance to macrolides have been reported. In Europe and to a lesser degree in the United States, macrolide-resistant *M. pneumoniae* is emerging. If macrolide resistance is prominent in a geographic locale or is suspected, then a nonmacrolide antibiotic should be considered for treatment; in addition, culture of *M. pneumoniae* may prove useful in these instances, providing an isolate for susceptibility testing.

Clinical observations and experimental data suggest that the addition of glucocorticoids to an antibiotic regimen may be of value for the treatment of severe or refractory *M. pneumoniae* pneumonia. However, relevant clinical experience is still limited. Even though appropriate antibiotic therapy significantly reduces the duration of respiratory illness, it does not appear to shorten the duration of detection of *M. pneumoniae* by culture or PCR; therefore, a test of cure or eradication is not suggested.

UROGENITAL MYCOPLASMAS

(See Also Chap. 130)

Epidemiology

M. hominis, *M. genitalium*, *U. urealyticum*, and *U. parvum* can cause urogenital tract disease. The significance of isolation of these organisms in a variety of other syndromes is unknown and in some cases is being investigated. *M. fermentans* has not been shown convincingly to cause human disease.

While urogenital mycoplasmas may be transmitted to a fetus during passage through a colonized birth canal, sexual contact is the major mode of transmission, and the risk of colonization increases dramatically with increasing numbers of sexual partners. In asymptomatic women, these mycoplasmas may be found throughout the lower urogenital tract. The vagina yields the largest number of organisms; next most densely colonized are the periurethral area and the cervix. Ureaplasmas are isolated less often from urine than from the cervix, but *M. hominis* is found with approximately the same frequency at these two sites. Ureaplasmas are isolated from the vagina of 40–80% of sexually active, asymptomatic women and *M. hominis* from 21–70%. The two microorganisms are found concurrently in 31–60% of women. In men, colonization with each organism is less prevalent. Mycoplasmas have been isolated from urine, semen, and the distal urethra of asymptomatic men.

Clinical Manifestations

URETHRITIS, PYELONEPHRITIS, AND URINARY CALCULI

In many episodes of *Chlamydophila*-negative nongonococcal urethritis, ureaplasmas may be the causative agent. These organisms may also cause chronic voiding symptoms in women. The common presence of ureaplasmas in the urethra of asymptomatic men suggests either that only certain serovars are pathogenic or that predisposing factors, such as lack of immunity, must exist in persons who develop symptomatic infection. Alternatively, disease may develop only upon initial exposure to ureaplasmas. Ureaplasmas have been implicated in epididymitis. *M. genitalium* also appears to cause urethritis. *M. genitalium* and ureaplasmas do not have a known role in prostatitis. *M. hominis* does not appear to have a primary etiologic role in urethritis, epididymitis, or prostatitis.

Evidence suggests that *M. hominis* causes up to 5% of cases of acute pyelonephritis. Ureaplasmas have not been associated with this disease.

Ureaplasmas play a limited role in the production of urinary calculi. The frequency with which ureaplasmas reach the kidney, the predisposing factors that allow them to do so, and the relative frequency of urinary tract calculi induced by this organism (as compared with other organisms) are not known.

PELVIC INFLAMMATORY DISEASE

M. hominis can cause pelvic inflammatory disease. In most episodes, *M. hominis* occurs as part of a polymicrobial infection, but the organism may play an independent role in a limited number of cases. Some data also support an association of *M. genitalium* with pelvic inflammatory disease. Ureaplasmas are not thought to cause pelvic inflammatory disease.

POSTPARTUM AND POSTABORTAL INFECTION

Studies implicate *M. hominis* as the primary pathogen in ~5–10% of women who have postpartum or postabortal fever; ureaplasmas have been implicated to a lesser degree. These infections are generally self-limited; however, if symptoms persist, specific antimicrobial therapy should be given. Ureaplasmas also appear to play a role in occasional postcesarean wound infections.

NONUROGENITAL INFECTION

M. hominis rarely causes nonurogenital infections, such as brain abscess, wound infection, poststernotomy mediastinitis, endocarditis, and neonatalmeningitis. These infections are most common among immunocompromised and hypogammaglobulinemic patients. Ureaplasmas and *M. hominis* can cause septic arthritis in immunodeficient patients. Ureaplasmas probably cause neonatal pneumonitis; their significant role in the development of bronchopulmonary dysplasia, the chronic lung disease of premature infants, has been documented in a number of studies. It is unclear whether ureaplasmas and *M. hominis* cause infertility, spontaneous abortion, premature labor, low birth weight, or chorioamnionitis.

Diagnosis

Culture and PCR are both appropriate methods for the isolation of urogenital mycoplasmas. Culture of these organisms, however, requires special techniques and media that generally are available only at larger medical centers and reference laboratories. Serologic testing is not recommended for the clinical diagnosis of urogenital *mycoplasma* infections.

Treatment: Urogenital Mycoplasma Infections (Table 175-2)

Because colonization with urogenital mycoplasmas is common, it appears at present that their isolation from the urogenital tract in the absence of disease generally does not warrant treatment. Macrolides and doxycycline are considered the antimicrobial agents of choice for *Ureaplasma* infections. *Ureaplasma* resistance to macrolides, doxycycline, quinolones, and chloramphenicol has been reported. *M. hominis* is resistant to macrolides. Doxycycline is generally the drug of choice for *M. hominis* infections, although resistance has been reported. Clindamycin is also generally active against *M. hominis*. Quinolones are active in vitro against *M. hominis*. For *M. genitalium,* the agent of choice appears to be azithromycin; treatment failures have been reported with other macrolides as well as with quinolones.

FURTHER READINGS

Baka S et al: Prevalence of *Ureaplasma urealyticum* and *Mycoplasma hominis* in women with chronic urinary symptoms. Urology 74:62, 2009[PMID: 19371925]

Christie LJ et al: Pediatric encephalitis: What is the role of *Mycoplasma pneumoniae*? Pediatrics 120:305, 2007[PMID: 17671056]

Hardy RD et al: Analysis of pulmonary inflammation and function in the mouse and baboon after exposure to *Mycoplasma pneumoniae* CARDS toxin. PLoS One 4:e7562, 2009

Kannan TR, Baseman JB: ADP-ribosylating and vacuolating cytotoxin of *Mycoplasma pneumoniae* represents unique virulence determinant among bacterial pathogens. Proc Natl Acad Sci USA 103:6724, 2006[PMID: 16617115]

Li X et al: Emerging macrolide resistance in Mycoplasma pneumoniae in children: Detection and characterization of resistant isolates. Pediatr Infect Dis J 28:693, 2009[PMID: 19633515]

McCracken GH: Current status of antibiotic treatment for *Mycoplasma pneumoniae* infections. Pediatr Infect Dis 5:167, 1986[PMID: 3080736]

Mena LA et al: A randomized comparison of azithromycin and doxycycline for the treatment of *Mycoplasma genitalium*–positive urethritis in men. Clin Infect Dis 48:1649, 2009[PMID: 19438399]

Suzuki S et al: Clinical evaluation of macrolide-resistant *Mycoplasma pneumoniae*. Antimicrob Agents Chemother 50:709, 2006[PMID: 16436730]

Tagliabue C et al: The impact of steroids given with macrolide therapy on experimental Mycoplasma pneumoniae respiratory infection. J Infect Dis 198:1180, 2008[PMID: 18717637]

Waites KB et al: New insights into the pathogenesis and detection of Mycoplasma pneumoniae infections. Future Microbiol 3:635, 2008[PMID: 19072181]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHLAMYDIAL INFECTIONS: INTRODUCTION

Chlamydiae are obligate intracellular bacteria that cause a wide variety of diseases in humans and animals.

ETIOLOGIC AGENTS

The chlamydiae were originally classified as four species in the genus *Chlamydia*: *C. trachomatis*, *C. pneumoniae*, *C. psittaci*, and *C. pecorum* (the last species being found in ruminants). The *C. psittaci* group has been separated into three species: *C. psittaci*, *C. felis*, and *C. abortus.* The mouse pneumonitis strain (MoPn) is now classified as *C. muridarum*, and the guinea pig inclusion conjunctivitis strain (GPIC) is now designated *C. caviae*.

C. trachomatis is divided into two biovars: trachoma and LGV (lymphogranuloma venereum). The trachoma biovar causes two major types of disease in humans: ocular trachoma, the leading infectious cause of preventable blindness in the developing world; and urogenital infections, which are sexually or neonatally transmitted. The 18 serovars of *C. trachomatis* fall into three groups: the trachoma serovars A, B, Ba, and C; the oculogenital serovars D–K; and the LGV serovars L_1–L_3. Serovars can be distinguished by serologic typing with monoclonal antibodies or by molecular gene typing. However, serovar identification usually is not important clinically, since the antibiotic susceptibility pattern is the same for all three groups. The one exception applies when LGV is suspected on clinical grounds; in this situation, serovar determination is important because a longer treatment duration is required for LGV strains.

BIOLOGY, GROWTH CYCLE, AND PATHOGENESIS

Biology

During their intracellular growth, chlamydiae produce characteristic intracytoplasmic inclusions that can be visualized by direct fluorescent antibody (DFA) or Giemsa staining of infected clinical material, such as conjunctival scrapings or cervical or urethral epithelial cells. Chlamydiae are nonmotile, gram-negative, obligate intracellular bacteria that replicate within the cytoplasm of host cells, forming the characteristic membrane-bound inclusions that are the basis for some diagnostic tests. Originally considered to be large viruses, chlamydiae differ from viruses in possessing RNA and DNA as well as a cell wall that is quite similar in structure to the cell wall of typical gram-negative bacteria. However, chlamydiae lack peptidoglycan; their structural integrity depends on disulfide binding of outer-membrane proteins.

Growth Cycle

Among the defining characteristics of chlamydiae is a unique growth cycle that involves alternation between two highly specialized morphologic forms (**Figs. 176-1** and **176-2**): the elementary body (EB), which is the infectious form and is specifically adapted for extracellular survival, and the metabolically active and replicating reticulate body (RB), which is not infectious, is adapted for an intracellular environment, and does not survive well outside the host cell. The biphasic growth cycle begins with attachment of the EB (diameter, 0.25–0.35 μm) at specific sites on the surface of the host cell. The EB enters the cell through a process similar to receptor-mediated endocytosis and resides in an inclusion, where the entire growth cycle is completed. The chlamydiae prevent phagosome-lysosome fusion. The inclusion membrane is modified by insertion of chlamydial antigens. Once the EB has entered the cell, it reorganizes into an RB, which is larger (0.5–1 μm) and contains more RNA. After ~8 h, the RB starts to divide by binary fission. The intracytoplasmic, membrane-bound inclusion body containing the RBs increases in size as the RBs multiply. Approximately 18–24 h after infection of the cell, these RBs begin to become EBs by a reorganization or condensation process that is poorly understood. After rupture of the inclusion body, the EBs are released to initiate another cycle of infection.

Figure 176-1

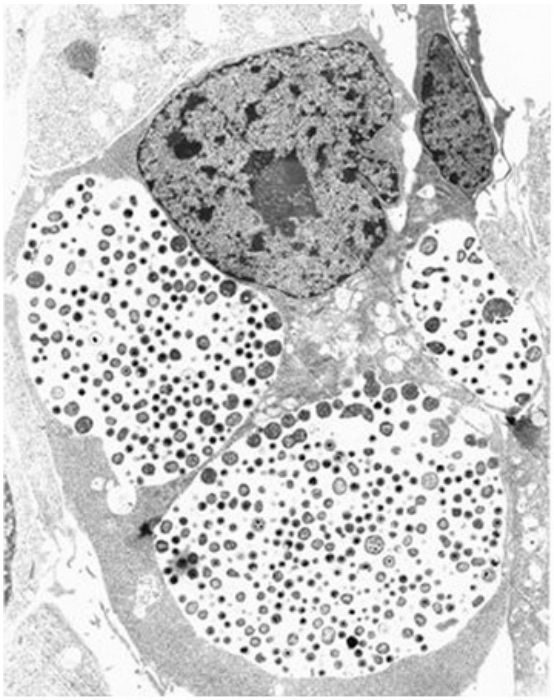

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chlamydial intracellular inclusions filled with smaller dense elementary bodies and larger reticulate bodies. *[Reprinted with permission from WE Stamm: Chlamydial infections, in Harrison—s Principles of Internal Medicine, 17th ed, AS Fauci et al (eds). New York, McGraw-Hill, 2008, p 1070.]*

Figure 176-2

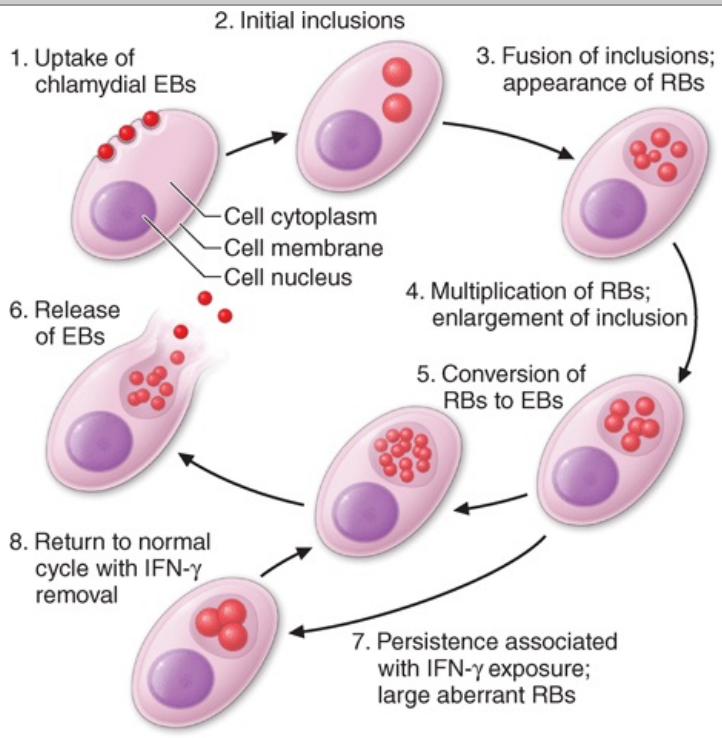

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chlamydial life cycle. EBs, elementary bodies; RBs, reticulate bodies; IFN-γ, interferon γ. *[Reprinted with permission from WE Stamm: Chlamydial infections, in Harrison—s Principles of Internal Medicine, 17th ed, AS Fauci et al (eds). New York, McGraw-Hill, 2008, p 1071.]*

Chlamydiae are susceptible to many broad-spectrum antibiotics and possess a number of enzymes, but they have a very restricted metabolic capacity. None of these metabolic reactions results in the production of energy. Chlamydiae have thus been considered to be energy parasites that use the ATP produced by the host cell for their own metabolic functions. Many aspects of chlamydial molecular biology are not well understood, but the sequencing of several chlamydial genomes and new proteomics research have provided researchers with many relevant tools for elucidating the biology of the life cycle.

Pathogenesis

Genital infections are mostly caused by *C. trachomatis* serovars D–K, with serovars D, E, and F involved most often. Molecular typing of the major outer-membrane protein gene (*omp1*) from which serovar differences arise has been used to demonstrate that polymorphisms can occur in isolates from patients who are exposed frequently to multiple infections, while less variation is observed in isolates from less sexually active populations. Polymorphisms in the major outer-membrane protein may provide antigenic variation, and the different forms allow persistence in the community because immunity to one is not protective against the others.

The trachoma biovar is essentially a parasite of squamocolumnar epithelial cells; the LGV biovar is more invasive and involves lymphoid cells. As is typical of chlamydiae, *C. trachomatis* strains are capable of causing chronic, clinically inapparent, asymptomatic infections. Because the duration of the chlamydial growth cycle is ~48–72 h, the incubation period of sexually transmitted chlamydial infections is relatively long—generally 1–3 weeks. *C. trachomatis* causes cell death as a result of its replicative cycle and can induce cell damage whenever it persists. However, few toxic effects are demonstrated, and cell death because of chlamydial replication is not sufficient to account for disease manifestations, the majority of which are due to immunopathologic mechanisms or nonspecific host responses to the organism or its byproducts.

In recent years, the entire genomes of various chlamydial species have been sequenced, the field of proteomics has become established, host innate immunity has been more precisely delineated, and innovative host cell–chlamydial interaction studies have been conducted. As a result, many insights have been gained into how chlamydiae adapt and replicate in their intracellular environment and produce disease. These insights into pathogenesis include information on the regulation of gene expression, protein localization, the type III secretion system, the roles of CD4+ and CD8+ T lymphocytes in the host response, and T lymphocyte trafficking.

The chlamydial heat-shock protein, which shares antigenic epitopes with similar proteins of other bacteria and with human heat-shock protein, may sensitize the host, and repeated infections may cause host cell damage. Persistent or recurrent chlamydial infections are associated with fibrosis, scarring, and complications following simple epithelial infections. A common endpoint of these late consequences is scarring of mucous membranes. Genital complications can lead to pelvic inflammatory disease (PID) and its late consequences of infertility, ectopic pregnancy, and chronic pelvic pain, while ocular infections may lead to blinding trachoma. High levels of antibody to human heat-shock protein have been associated with tubal factor infertility and ectopic pregnancy. Without adequate therapy, chlamydial infections may persist for several years, although symptoms—if present—usually abate.

Pathogenic mechanisms of *C. pneumoniae* have yet to be completely elucidated. The same is true for *C. psittaci,* except that this agent infects cells very efficiently and causes disease that may reflect direct cytopathic effects.

CHLAMYDIA TRACHOMATIS INFECTIONS

Genital Infections

SPECTRUM

Although chlamydiae cause a number of human diseases, localized lower genital tract infections caused by *C. trachomatis* and the sequelae of such infections are the most important in terms of medical and economic impact. Oculogenital infections due to *C. trachomatis* serovars D–K are transmitted during sexual contact or from mother to baby during childbirth and are associated with many syndromes, including cervicitis, salpingitis, acute urethral syndrome, endometritis, ectopic pregnancy, infertility, and PID in female patients; urethritis, proctitis, and epididymitis in male patients; and conjunctivitis and pneumonia in infants. Women bear the greatest burden of morbidity because of the serious sequelae of these infections. Untreated infections lead to PID, and multiple episodes of PID can lead to tubal factor infertility and chronic pelvic pain. Studies estimate that up to 80–90% of women and >50% of men with *C. trachomatis* genital infections lack symptoms; other patients have very mild symptoms. Thus a large reservoir of infected persons continues to transmit infection to sexual partners.

As their designations reflect, the LGV serovars (L_1, L_2, and L_3) cause LGV, an invasive sexually transmitted disease (STD) characterized by acute lymphadenitis with bubo formation and/or acute

hemorrhagic proctitis (see "Lymphogranuloma Venereum," below).

EPIDEMIOLOGY

C. trachomatis genital infections are global in distribution. The World Health Organization (WHO) estimates that >89 million cases occur annually worldwide. In the United States, these infections are the most commonly reported of all infectious diseases. In 2009, 1.24 million cases were reported to the U.S. Centers for Disease Control and Prevention (CDC); however, the CDC estimates that 2–3 million new cases occur per year, with substantial underreporting due to lack of screening in some populations. Rates of infection increase every year; higher rates among women than among men reflect the focus on expansion of screening programs in women during the past 20 years, the use of increasingly sensitive diagnostic tests, an increased emphasis on case reporting, and improvements in the information systems used for reporting. The CDC and other professional organizations recommend annual screening of all sexually active women ≤25 years of age as well as rescreening of previously infected individuals at 3 months. Young women have the highest infection rates; in 2008, the figures were 3275.8 and 3179.9 cases per 100,000 population at 15–19 and 20–24 years of age, respectively. Age-specific rates among men, while much lower than those among women, were highest in the 20- to 24-year-old age group, at 1056.1 cases per 100,000. In 2008, rates increased for all racial and ethnic groups; however, the rate among blacks was more than eight times higher than that among whites, with 1519.3 and 173.6 cases per 100,000, respectively. The rates among American Indian/Alaska Natives (808.8) and Latinos (510.4) were 4.7 and 2.9 times higher, respectively, than that among whites. These disparities are important reflections of health inequities in the United States.

The above statistics are based on case reporting. Studies based on screening surveys estimate that the U.S. prevalence of *C. trachomatis* cervical infection is 5% among asymptomatic female college students and prenatal patients, >10% for women seen in family planning clinics, and >20% for women seen in STD clinics. The prevalence of genital *C. trachomatis* infections varies substantially by geographic locale, with the highest rates in the southeastern United States. However, asymptomatic infections have been detected in >8–10% of young female military recruits from all parts of the country. The prevalence of *C. trachomatis* in the cervix of pregnant women is 5–10 times higher than that of *Neisseria gonorrhoeae*. The prevalence of genital infection with either agent is highest among women who are between the ages of 18 and 24, single, and non-Caucasian (e.g., African-American, Latina, Asian, Pacific Islander). Recurrent infections occur frequently in these same risk groups and are often acquired from untreated sexual partners. The use of oral contraception and the presence of cervical ectopy also confer an increased risk. The proportion of infections that are asymptomatic appears to be higher for *C. trachomatis* than for *N. gonorrhoeae*, and symptomatic *C. trachomatis* infections are clinically less severe. Mild or asymptomatic *C. trachomatis* infections of the fallopian tubes nonetheless cause ongoing tubal damage and infertility. The costs of *C. trachomatis* infections and their complications to the U.S. health care system are projected to be >$2.4 billion annually.

CLINICAL MANIFESTATIONS

Nongonococcal and Postgonococcal Urethritis

C. trachomatis is the most common cause of nongonococcal urethritis (NGU) and postgonococcal urethritis (PGU). The designation *PGU* refers to NGU developing in men 2–3 weeks after treatment of gonococcal urethritis with single doses of agents such as penicillin or cephalosporins, which lack antimicrobial activity against chlamydiae. Since current treatment regimens for gonorrhea have evolved and now include combination therapy with tetracycline, doxycycline, or azithromycin—all of which are effective against concomitant chlamydial infection—both the incidence of PGU and the causative role of *C. trachomatis* in this syndrome have declined.

In the United States, most of the estimated 2 million cases of acute urethritis are NGU, and *C. trachomatis* is implicated in 30–50% of these cases. The cause of most of the remaining cases of NGU is uncertain, but recent evidence suggests that *Ureaplasma urealyticum*, *Mycoplasma genitalium,Trichomonas vaginalis*, and herpes simplex virus (HSV) cause some cases. The rate of involvement of *C. trachomatis* in urethral infection ranges from 3–7% among asymptomatic men to 15–20% among symptomatic men attending STD clinics. One recent multisite study of men in Baltimore, Seattle, Denver, and San Francisco reported an overall chlamydial prevalence of 7% in urine samples assessed by nucleic acid amplification tests (NAATs). As in women, infection in men is age related, with young age as the greatest risk factor for chlamydial urethritis. The prevalence among men is highest at 20–24 years of age. In STD clinics, urethritis is usually less prevalent among men who have sex with men (MSM) than among heterosexual men and is almost always much more common among black men than among white men. One study reported prevalences of 19% and 9% among nonwhite and white heterosexual men, respectively.

NGU is diagnosed by documentation of a leukocytic urethral exudate and by exclusion of gonorrhea by Gram—s staining or culture. *C. trachomatis* urethritis is generally less severe than gonococcal urethritis, although in any individual patient these two forms of urethritis cannot reliably be differentiated solely on clinical grounds. Symptoms include urethral discharge (often whitish and mucoid rather than frankly purulent), dysuria, and urethral itching. Physical examination may reveal

meatal erythema and tenderness as well as a urethral exudate that is often demonstrable only by stripping of the urethra.

At least one-third of male patients with *C. trachomatis* urethral infection have no evident signs or symptoms of urethritis. The availability of NAATs for first-void urine specimens has facilitated broader-based testing for asymptomatic infection in male patients. As a result, asymptomatic chlamydial urethritis has been demonstrated in 5–10% of sexually active male adolescents screened at school-based clinics or community centers. Such patients generally have pyuria (≥15 leukocytes per 400x microscopic field in the sediment of first-void urine), a positive leukocyte esterase test, or an increased number of leukocytes on a Gram-stained smear prepared from a urogenital swab inserted 1–2 cm into the anterior urethra. To differentiate between true urethritis and functional symptoms in symptomatic patients or to make a presumptive diagnosis of *C. trachomatis* infection in high-risk but asymptomatic men (e.g., male patients in STD clinics, sex partners of women with nongonococcal salpingitis or mucopurulent cervicitis, fathers of children with inclusion conjunctivitis), the examination of an endourethral specimen for increased leukocytes is useful if specific diagnostic tests for chlamydiae are not available. Alternatively, urethritis can be assayed noninvasively by examination of a first-void urine sample for pyuria, either by microscopy or by the leukocyte esterase test. Urine (or a urethral swab) can also be tested directly for chlamydiae by DNA amplification methods, as described below (see "Detection Methods").

Epididymitis

Chlamydial urethritis may be followed by acute epididymitis, but this condition is rare, generally occurring in sexually active patients <35 years of age; in older men, epididymitis is usually associated with gram-negative bacterial infection and/or instrumentation procedures. It is estimated that 50–70% of cases of acute epididymitis are caused by *C. trachomatis*. The condition usually presents as unilateral scrotal pain with tenderness, swelling, and fever in a young man, often occurring in association with chlamydial urethritis. The illness may be mild enough to treat with oral antibiotics on an outpatient basis or severe enough to require hospitalization and parenteral therapy. Testicular torsion should be excluded promptly by radionuclide scan, Doppler flow study, or surgical exploration in a teenager or young adult who presents with acute unilateral testicular pain without urethritis. The possibility of testicular tumor or chronic infection (e.g., tuberculosis) should be excluded when a patient with unilateral intrascrotal pain and swelling does not respond to appropriate antimicrobial therapy.

Reactive Arthritis

Reactive arthritis (formerly known as Reiter—s syndrome) consists of conjunctivitis, urethritis (or, in female patients, cervicitis), arthritis, and characteristic mucocutaneous lesions. It may develop in 1–2% of cases of NGU and is thought to be the most common type of peripheral inflammatory arthritis in young men. *C. trachomatis* has been recovered from the urethra of 16–44% of patients with reactive arthritis and from 69% of men who have signs of urogenital inflammation at the time of examination. Antibodies to *C. trachomatis* have also been detected in 46–67% of patients with reactive arthritis, and *Chlamydia*-specific cell-mediated immunity has been documented in 72%. In addition, *C. trachomatis* has been isolated from synovial biopsy samples from 15 of 29 patients in a number of small series and from a smaller proportion of synovial fluid specimens. Chlamydial nucleic acids have been identified in synovial membranes and chlamydial EBs in joint fluid. The pathogenesis of reactive arthritis is unclear, but this condition probably represents an abnormal host response to a number of infectious agents, including those associated with bacterial gastroenteritis (e.g., *Salmonella*, *Shigella*, *Yersinia*, or *Campylobacter*) or to infection with *C. trachomatis* or *N. gonorrhoeae*. Since >80% of affected patients have the HLA-B27 phenotype and since other mucosal infections produce an identical syndrome, chlamydial infection is thought to initiate an aberrant hyperactive immune response that produces inflammation of the involved target organs in these genetically predisposed individuals. Evidence of exaggerated cell-mediated and humoral immune responses to chlamydial antigens in reactive arthritis supports this hypothesis. The finding of chlamydial EBs and DNA in joint fluid and synovial tissue from patients with reactive arthritis suggests that chlamydiae may actually spread from genital to joint tissues in these patients—perhaps in macrophages.

NGU is the initial manifestation of reactive arthritis in 80% of patients, typically occurring within 14 days after sexual exposure. The urethritis may be mild and may even go unnoticed by the patient. Similarly, gonococcal urethritis may precede reactive arthritis, but co-infection with an agent of NGU is difficult to rule out. The urethral discharge may be purulent or mucopurulent, and patients may or may not report dysuria. Accompanying prostatitis, usually asymptomatic, has been described. Arthritis usually begins ~4 weeks after the onset of urethritis but may develop sooner or, in a small percentage of cases, may actually precede urethritis. The knees are most frequently involved; next most commonly affected are the ankles and small joints of the feet. Sacroiliitis, either symmetrical or asymmetrical, is documented in two-thirds of patients. Mild bilateral conjunctivitis, iritis, keratitis, or uveitis is sometimes present but lasts for only a few days. Finally, dermatologic manifestations occur in up to 50% of patients. The initial lesions—usually papules with a central yellow spot—most often involve the soles and palms and, in ~25% of patients, eventually epithelialize and thicken to produce keratoderma blenorrhagicum. Circinate balanitis is usually painless and occurs in fewer than half of patients. The

initial episode of reactive arthritis usually lasts 2–6 months.

Proctitis

Primary anal or rectal infections with *C. trachomatis* have been described in women and MSM who practice anal intercourse. In these infections, rectal involvement is initially characterized by severe anorectal pain, a bloody mucopurulent discharge, and tenesmus. Oculogenital serovars D–K and LGV serovars L_1, L_2, and L_3 have been found to cause proctitis. The LGV serovars are far more invasive and cause much more severely symptomatic disease, including severe ulcerative proctocolitis that can be clinically confused with HSV proctitis. Histologically, LGV proctitis may resemble Crohn—s disease in that giant cell formation and granulomas are detected. In the United States and Europe, cases of LGV proctitis occur almost exclusively in MSM, many of whom are positive for HIV infection.

The less invasive non-LGV serovars of *C. trachomatis* cause mild proctitis. Many infected individuals are asymptomatic, and in these cases infection is diagnosed only by routine culture or NAAT of rectal swabs. The number of fecal leukocytes is usually abnormal in both asymptomatic and symptomatic cases. Sigmoidoscopy may yield normal findings or may reveal mild inflammatory changes or small erosions or follicles in the lower 10 cm of the rectum. Histologic examination of rectal biopsies generally shows anal crypts and prominent follicles as well as neutrophilic infiltration of the lamina propria. Chlamydial proctitis is best diagnosed by isolation of *C.trachomatis* from the rectum and documentation of a response to appropriate therapy. NAATs are reportedly more sensitive than culture for diagnosis and are also specific.

Mucopurulent Cervicitis

Although many women with chlamydial infections of the cervix have no symptoms, almost half generally have local signs of infection on examination. Cervicitis is usually characterized by the presence of a mucopurulent discharge, with >20 neutrophils per microscopic field visible in strands of cervical mucus in a thinly smeared, gram-stained preparation of endocervical exudate. Hypertrophic ectopy of the cervix may also be evident as an edematous area near the cervical os that is congested and bleeds easily on minor trauma (e.g., when a specimen is collected with a swab). A Papanicolaou smear shows increased numbers of neutrophils as well as a characteristic pattern of mononuclear inflammatory cells including plasma cells, transformed lymphocytes, and histiocytes. Cervical biopsy shows a predominantly mononuclear cell infiltrate of the subepithelial stroma. Clinical experience and collaborative studies indicate that a cutoff of >30 polymorphonuclear leukocytes (PMNs)/1000x field in a gram-stained smear of cervical mucus correlates best with chlamydial or gonococcal cervicitis.

Clinical recognition of chlamydial cervicitis depends on a high index of suspicion and careful cervical examination. No genital symptoms are specifically correlated with chlamydial cervical infection. The differential diagnosis of a mucopurulent discharge from the endocervical canal in a young, sexually active woman includes gonococcal endocervicitis, salpingitis, endometritis, and intrauterine contraceptive device–induced inflammation. Diagnosis of cervicitis is based on the presence of PMNs on a cervical swab as noted above; the presence of chlamydiae is confirmed by either culture or NAAT.

Pelvic Inflammatory Disease

Inflammation of sections of the fallopian tube is often referred to as salpingitis or PID. The proportion of acute salpingitis cases caused by *C. trachomatis* varies geographically and with the population studied. It has been estimated that *C. trachomatis* causes up to 50% of PID cases in the United States. PID occurs via ascending intraluminal spread of *C. trachomatis* or *N.gonorrhoeae* from the lower genital tract. Mucopurulent cervicitis is often followed by endometritis, endosalpingitis, and finally pelvic peritonitis. Evidence of mucopurulent cervicitis is often found in women with laparoscopically verified salpingitis. Similarly, endometritis, demonstrated by an endometrial biopsy showing plasma cell infiltration of the endometrial epithelium, is documented in most women with laparoscopy-verified chlamydial (or gonococcal) salpingitis. Chlamydial endometritis can also occur in the absence of clinical evidence of salpingitis. Histologic evidence of endometritis has been correlated with a syndrome consisting of vaginal bleeding, lower abdominal pain, and uterine tenderness in the absence of adnexal tenderness. Chlamydial salpingitis produces milder symptoms than gonococcal salpingitis and may be associated with less marked adnexal tenderness. Thus, mild adnexal or uterine tenderness in a sexually active woman with cervicitis suggests chlamydial PID.

Chronic untreated endometrial and tubal inflammation can result in tubal scarring, impaired tubal function, tubal occlusion, and infertility even among women who report no prior treatment for PID. *C. trachomatis* has been particularly implicated in "subclinical" PID on the basis of a lack of history of PID among *Chlamydia*-seropositive women with tubal damage and detection of chlamydial DNA or antigen among asymptomatic women with tubal infertility. These data suggest that the best method to prevent PID and its sequelae is surveillance and control of lower genital tract infections along with diagnosis and treatment of sex partners and prevention of reinfections. Promotion of early symptom recognition and health care presentation may reduce the frequency and severity of sequelae of PID.

Perihepatitis

The Fitz-Hugh–Curtis syndrome was originally described as a complication of gonococcal PID.

However, studies over the past several decades have suggested that chlamydial infection is more commonly associated with perihepatitis than is *N. gonorrhoeae*. Perihepatitis should be suspected in young, sexually active women who develop right-upper-quadrant pain, fever, or nausea. Evidence of salpingitis may or may not be found on examination. Frequently, perihepatitis is strongly associated with extensive tubal scarring, adhesions, and inflammation observed at laparoscopy, and high titers of antibody to the 57-kDa chlamydial heat-shock protein have been documented. Culture and/or serologic evidence of *C. trachomatis* is found in three-fourths of women with this syndrome.

Urethral Syndrome in Women

In the absence of infection with uropathogens such as coliforms or *Staphylococcus saprophyticus*, *C. trachomatis* is the pathogen most commonly isolated from college women with dysuria, frequency, and pyuria. Screening studies can recover *C. trachomatis* at both the cervix and the urethra; in up to 25% of infected women, the organism is isolated only from the urethra. The urethral syndrome in women consists of dysuria and frequency in conjunction with chlamydial urethritis, pyuria, and no bacteriuria or urinary pathogens. Although symptoms of the urethral syndrome may develop in some women with chlamydial infection, the majority of women attending STD clinics for urethral chlamydial infection do not have dysuria or frequency. Even in women with chlamydial urethritis causing the acute urethral syndrome, signs of urethritis such as urethral discharge, meatal redness, and swelling are uncommon. However, mucopurulent cervicitis in a woman presenting with dysuria and frequency strongly suggests *C. trachomatis* urethritis. Other correlates of chlamydial urethral syndrome include a duration of dysuria of >7–10 days, lack of hematuria, and lack of suprapubic tenderness. Abnormal urethral Gram—s stains showing >10 PMNs/1000x field in women with dysuria but without coliform bacteriuria support the diagnosis of chlamydial urethritis. Other possible diagnoses include gonococcal or trichomonal infection of the urethra.

Infection in Pregnancy and the Neonatal Period

Infections during pregnancy can be transmitted to infants during delivery. Approximately 20–30% of infants exposed to *C. trachomatis* in the birth canal develop conjunctivitis, and 10–15% subsequently develop pneumonia. Consequently, all newborn infants receive ocular prophylaxis at birth to prevent ophthalmia neonatorum. Without treatment, conjunctivitis usually develops at 5–19 days of life and often results in a profuse mucopurulent discharge. Roughly half of infected infants develop clinical evidence of inclusion conjunctivitis. However, it is impossible to differentiate chlamydial conjunctivitis from other forms of neonatal conjunctivitis (e.g., that due to *N. gonorrhoeae*, *Haemophilus influenzae*, *Streptococcus pneumoniae*, or HSV) on clinical grounds; thus laboratory diagnosis is required. Inclusions within epithelial cells are often detected in Giemsa-stained conjunctival smears, but these smears are considerably less sensitive than cultures or NAATs for chlamydiae. Gram-stained smears may show gonococci or occasional small gram-negative coccobacilli in *Haemophilus* conjunctivitis, but smears should be accompanied by cultures or NAATs for these agents.

C. trachomatis has also been isolated frequently and persistently from the nasopharynx, rectum, and vagina of infected infants—occasionally for >1 year in the absence of treatment. In some cases, otitis media results from perinatally acquired chlamydial infection. Pneumonia may develop in infants from 2 weeks to 4 months of age. *C. trachomatis* is estimated to cause 20–30% of pneumonia cases in infants <6 months of age. Epidemiologic studies have linked chlamydial pulmonary infection in infants with increased occurrence of subacute lung disease (bronchitis, asthma, wheezing) in later childhood.

Lymphogranuloma Venereum

C. trachomatis serovars L_1, L_2, and L_3 cause LGV, an invasive systemic STD. The peak incidence of LGV corresponds with the age of greatest sexual activity: the second and third decades of life. The worldwide incidence of LGV is falling, but the disease is still endemic and a major cause of morbidity in parts of Asia, Africa, South America, and the Caribbean. LGV is rare in industrialized countries; for more than a decade, the reported incidence in the United States has been only 0.1 case per 100,000 population. In the Bahamas, an apparent outbreak of LGV was described in association with a concurrent increase in heterosexual infection with HIV. Reports of outbreaks with the newly identified variant L_{2b} in Europe, Australia, and the United States indicate that LGV is becoming more prevalent among MSM. These cases have usually presented as hemorrhagic proctocolitis in HIV-positive men. More widespread use of NAATs for identification of rectal infections may have enhanced case recognition.

The frequency of infection following exposure is believed to be much lower for LGV than for gonorrhea and syphilis. Early manifestations are recognized more often in men than in women, who usually present with late complications. In the United States, where the reported male-to-female ratio of cases is 3.4:1, most cases involve MSM and persons returning from abroad (travelers, sailors, and military personnel).

LGV begins as a small painless papule that tends to ulcerate at the site of inoculation, often escaping attention. This primary lesion heals in a few days without scarring and, even when noticed, is usually recognized as LGV only in retrospect. LGV strains of *C. trachomatis* have occasionally been recovered

from genital ulcers and from the urethra of men and the endocervix of women who present with inguinal adenopathy; these areas may be the primary sites of infection in some cases. Proctitis is more common among people who practice receptive anal intercourse, and an elevated white blood cell count in anorectal smears may predict LGV in these patients. Ulcer formation may facilitate transmission of HIV infection and other sexually transmitted and blood-borne diseases.

As NAATs for *C. trachomatis* are being used more often, increasing numbers of cases of LGV proctitis are being recognized in MSM. Such patients present with anorectal pain and mucopurulent, bloody rectal discharge. Although these patients may report diarrhea, they are often referring not to diarrhea but rather to frequent, painful, unsuccessful attempts at defecation (tenesmus). Sigmoidoscopy reveals ulcerative proctitis or proctocolitis, with purulent exudate and mucosal bleeding. Histopathologic findings in the rectal mucosa include granulomas with giant cells, crypt abscesses, and extensive inflammation. These clinical, sigmoidoscopic, and histopathologic findings may closely resemble those of Crohn—s disease of the rectum.

The most common presenting picture in heterosexual men and women is the *inguinal syndrome*, which is characterized by painful inguinal lymphadenopathy beginning 2–6 weeks after presumed exposure; in rare instances, the onset comes after a few months. The inguinal adenopathy is unilateral in two-thirds of cases, and palpable enlargement of the iliac and femoral nodes is often evident on the same side as the enlarged inguinal nodes. The nodes are initially discrete, but progressive periadenitis results in a matted mass of nodes that becomes fluctuant and suppurative. The overlying skin becomes fixed, inflamed, and thin, and multiple draining fistulas finally develop. Extensive enlargement of chains of inguinal nodes above and below the inguinal ligament ("the sign of the groove") is not specific and, although not uncommon, is documented in only a minority of cases. On histologic examination, infected nodes are initially found to have characteristic small stellate abscesses surrounded by histiocytes. These abscesses coalesce to form large, necrotic, suppurative foci. Spontaneous healing usually takes place after several months; inguinal scars or granulomatous masses of various sizes persist for life. Massive pelvic lymphadenopathy may lead to exploratory laparotomy.

Constitutional symptoms are common during the stage of regional lymphadenopathy and, in cases of proctitis, may include fever, chills, headache, meningismus, anorexia, myalgias, and arthralgias. These findings in the presence of lymphadenopathy are sometimes mistakenly interpreted as malignant lymphoma. Other systemic complications are infrequent but include arthritis with sterile effusion, aseptic meningitis, meningoencephalitis, conjunctivitis, hepatitis, and erythema nodosum (Fig. e7-40). *C. trachomatis* has been recovered from the cerebrospinal fluid and in one case was isolated from the blood of a patient with severe constitutional symptoms—a result indicating dissemination of infection. Laboratory-acquired infections suspected of being due to the inhalation of aerosols have been associated with mediastinal lymphadenitis, pneumonitis, and pleural effusion.

Complications of untreated anorectal infection include perirectal abscess; anal fistulas; and rectovaginal, rectovesical, and ischiorectal fistulas. Secondary bacterial infection probably contributes to these complications. Rectal stricture is a late complication of anorectal infection and usually develops 2–6 cm from the anal orifice—i.e., at a site within reach on digital rectal examination. Proximal extension of the stricture for several centimeters may lead to a mistaken clinical and radiographic diagnosis of carcinoma. A small percentage of cases of LGV in men present as chronic progressive infiltrative, ulcerative, or fistular lesions of the penis, urethra, or scrotum. Associated lymphatic obstruction may produce elephantiasis. When urethral stricture occurs, it usually involves the posterior urethra and causes incontinence or difficulty with urination.

Diagnosis

DETECTION METHODS

Historically, chlamydiae were cultivated in the yolk sac of embryonated eggs. The organisms can be grown more easily in tissue culture, but cell culture—once considered the diagnostic gold standard—has been replaced by nonculture assays **(Table 176-1)**. In general, culture for chlamydiae in clinical specimens is now performed only in specialized laboratories. The first nonculture assays, such as DFA staining of clinical material and enzyme immunoassay (EIA), have been replaced by molecular tests that amplify the nucleic acids in clinical specimens. These NAATS are currently recommended by the CDC as the diagnostic assays of choice.

Table 176-1 Diagnostic Tests for Sexually Transmitted and Perinatal Chlamydia Trachomatis Infection

Infection	Suggestive Signs/Symptoms	Presumptive Diagnosis[a]	Confirmatory Test of Choice
Men			
NGU, PGU	Discharge, dysuria	Gram's stain with >4 neutrophils per oil-	Urine or ure-thral NAAT for *C. trachomatis*

		immersion field; no gonococci	
Epididymitis	Unilateral intrascrotal swelling, pain, tenderness; fever; NGU	Gram's stain with >4 neutrophils per oil-immersion field; no gonococci; urinalysis with pyuria	Urine or urethral NAAT for *C. trachomatis*
Women			
Cervicitis	Mucopurulent cervical discharge, bleeding and edema of the zone of cervical ectopy	Cervical Gram's stain with ≥20 neutrophils per oil-immersion field in cervical mucus	Urine, cervical, or vaginal NAAT for C. trachomatis
Salpingitis	Lower abdominal pain, cervical motion tenderness, adnexal tenderness or masses	*C. trachomatis* always potentially present in salpingitis	Urine, cervical, or vaginal NAAT for *C. trachomatis*
Urethritis	Dysuria and frequency without hematuria	MPC; sterile pyuria; negative routine urine culture	Urine or urethral NAAT for *C. trachomatis*
Adults of Either Sex			
Proctitis	Rectal pain, discharge, tenesmus, bleeding; history of receptive anorectal intercourse	Negative gonococcal culture and Gram's stain; at least 1 neutrophil per oil-immersion field in rectal Gram's stain	Rectal NAAT for *C. trachomatis* or culture
Reactive arthritis	NGU, arthritis, conjunctivitis, typical skin lesions	Gram's stain with >4 neutrophils per oil-immersion field; lack of gonococci indicative of NGU	Urine or urethral NAAT for *C. trachomatis*
LGV	Regional adenopathy, primary lesion, proctitis, systemic symptoms	None	Culture of LGV strain from node or rectum, occa-sionally from urethra or cervix; NAAT for *C. trachomatis* from these sites; LGV CF titer, ≥1:64; micro-IF titer, ≥1:512
Neonates			
Conjunctivitis	Purulent conjunctival discharge 6–18 days after delivery	Negative culture and Gram's stain for gonococci, *Haemophilus* spp., pneumococci, staphylococci	Conjunctival NAAT for *C. trachomatis*; FA-stained scraping of conjunctival material
Infant pneumonia	Afebrile, staccato cough, diffuse rales, bilateral hyperinflation, interstitial infiltrates	None	Chlamydial culture or NAAT of sputum, pharynx, eye, rectum; micro-IF antibody to *C. trachomatis*—fourfold change in IgG or IgM antibody titer

[a]A presumptive diagnosis of chlamydial infection is often made in the syndromes listed when gonococci are not found. A positive test for *Neisseria gonorrhoeae* does not exclude the involve-ment of *C. trachomatis*, which often is present in patients with gonorrhea.

Abbreviations: CF, complement-fixing; FA, fluorescent antibody; LGV, lymphogranuloma venereum; micro-IF, microimmunofluorescence; MPC, mucopurulent cervicitis; NAAT, nucleic acid amplification test; NGU, nongonococcal urethritis; PGU, postgonococcal urethritis.

Source: Reprinted with permission from WE Stamm: Chlamydial infections, in *Harrison's Principles of Internal Medicine*, 17th ed, AS Fauci et al (eds). New York, McGraw-Hill, 2008, p 1075.

CHOICE OF SPECIMEN

Cervical and urethral swabs have traditionally been used for the diagnosis of STDs in female and male patients, respectively. However, given the greatly increased sensitivity and specificity of NAATs, less invasive samples (e.g., urine for both sexes and vaginal swabs for women) can be used. For screening of asymptomatic women, the CDC now recommends that self-collected vaginal swabs, which are slightly more sensitive than urine, be used. Urine screening tests are often used in outreach screening programs, however. For symptomatic women undergoing a pelvic examination, cervical swab samples are desirable because they have slightly higher chlamydial counts. For male patients, a urine specimen

is the sample of choice.

ALTERNATIVE SPECIMEN TYPES

Ocular samples from babies and adults can be assessed by NAATs. However, since commercial NAATs for this purpose have not yet been approved by the U.S. Food and Drug Administration (FDA), laboratories must perform their own verification studies. Samples from rectal and pharyngeal sites have been used successfully to detect chlamydiae, but again, laboratories must verify test performance.

OTHER DIAGNOSTIC ISSUES

Because NAATs measure nucleic acids instead of live organisms, they should be used with caution as test-of-cure assays. Residual nucleic acid from cells rendered noninfective by antibiotics may continue to yield a positive result in NAATs until as long as 3 weeks after therapy, when viable organisms have actually been eradicated. Therefore, clinicians should not use NAATs for test of cure until after 3 weeks. The CDC currently does not recommend a test of cure after treatment for infection with *C. trachomatis*. However, because incidence studies have demonstrated that previous chlamydial infection increases the probability of becoming reinfected, the CDC does recommend that previously infected individuals be rescreened 3 months after treatment.

SEROLOGY

Serologic testing may be helpful in the diagnosis of LGV and neonatal pneumonia caused by *C. trachomatis.* The serologic test of choice is the microimmunofluorescence (MIF) test, in which high-titer purified EBs mixed with embryonated chicken yolk sac material are affixed to a glass microscope slide to which dilutions of sera are applied. After incubation and washing, fluorescein-conjugated IgG or IgM antibody is applied. The test is read with an epifluorescence microscope, with the highest dilution of serum producing visible fluorescence designated as the titer. The MIF test is not widely available and is highly labor intensive. Although the complement fixation (CF) test can also be used, it employs only lipopolysaccharide (LPS) as the antigen and therefore identifies the pathogen only to the genus level. Single-point titers of >1:64 support a diagnosis of LGV, in which it is difficult to demonstrate rising antibody titers—i.e., paired serum samples are difficult to obtain since, by its very nature, the disease results in the patient—s being seen by the physician after the acute stage. Any antibody titer of >1:16 is considered significant evidence of exposure to chlamydiae. However, serologic testing is not recommended for diagnosis of uncomplicated genital infections of the cervix, urethra, and lower genital tract or for *C. trachomatis* screening of asymptomatic individuals.

Treatment: *C. Trachomatis* Genital Infections

A 7-day course of tetracycline (500 mg four times daily), doxycycline (100 mg twice daily), erythromycin (500 mg four times daily), or a fluoroquinolone (ofloxacin, 300 mg twice daily; or levofloxacin, 500 mg/d) can be used for treatment of uncomplicated chlamydial infections. A single 1-g oral dose of azithromycin is as effective as a 7-day course of doxycycline for the treatment of uncomplicated genital *C. trachomatis* infections in adults. Azithromycin causes fewer adverse gastrointestinal reactions than do older macrolides such as erythromycin. The single-dose regimen of azithromycin has great appeal for the treatment of patients with uncomplicated chlamydial infection (especially those without symptoms and those with a likelihood of poor compliance) and of the sexual partners of infected patients. These advantages must be weighed against the considerably greater cost of azithromycin. Whenever possible, the single 1-g dose should be given as directly observed therapy. Although not approved by the FDA for use in pregnancy, this regimen appears to be safe and effective for this purpose. However, amoxicillin (500 mg three times daily for 7 days) can also be given to pregnant women. The fluoroquinolones are contraindicated in pregnancy. A 2-week course of treatment is recommended for complicated chlamydial infections (e.g., PID, epididymitis) and at least a 3-week course of doxycycline (100 mg orally twice daily) or erythromycin base (500 mg orally four times daily) for LGV. Failure of treatment with a tetracycline in genital infections usually indicates poor compliance or reinfection rather than involvement of a drug-resistant strain. To date, clinically significant drug resistance has not been observed in *C. trachomatis*.

Treatment or testing for chlamydiae should be considered among *N. gonorrhoeae*–infected patients because of the frequency of co-infection. Systemic treatment with erythromycin has been recommended for ophthalmia neonatorum and for *C. trachomatis* pneumonia in infants. For the treatment of adult inclusion conjunctivitis, a single 1-g dose of azithromycin was as effective as standard 10-day treatment with doxycycline. Recommended treatment regimens for both bubonic and anogenital LGV include tetracycline, doxycycline, or erythromycin for 21 days.

SEX PARTNERS

The continued high prevalence of chlamydial infections in most parts of the United States is due primarily to the failure to diagnose—and therefore treat—patients with symptomatic or asymptomatic infection and their sex partners. Urethral or cervical infection with *C. trachomatis* has been well documented in a high proportion of the sex partners of patients with NGU, epididymitis, reactive arthritis, salpingitis, and endocervicitis. If possible, confirmatory laboratory tests for chlamydiae should

be undertaken in these individuals, but even those without positive tests or evidence of clinical disease who have recently been exposed to proven or possible chlamydial infection (e.g., NGU) should be offered therapy. A novel approach is partner-delivered therapy, in which infected patients receive treatment and are also provided with single-dose azithromycin to give to their sex partner(s).

NEONATES AND INFANTS

In neonates with conjunctivitis or infants with pneumonia, erythromycin ethylsuccinate or estolate can be given orally at a dosage of 50 mg/kg per day, preferably in four divided doses, for 2 weeks. Careful attention must be given to compliance with therapy—a frequent problem. Relapses of eye infection are common after topical treatment with erythromycin or tetracycline ophthalmic ointment and may also follow oral erythromycin therapy. Thus follow-up cultures should be performed after treatment. Both parents should be examined for *C. trachomatis* infection and, if diagnostic testing is not readily available, should be treated with doxycycline or azithromycin.

PREVENTION

Since many chlamydial infections are asymptomatic, effective control and prevention must involve periodic screening of individuals at risk. Selective cost-effective screening criteria have been developed. Among women, young age (generally <25 years) is a critical risk factor for chlamydial infections in nearly all studies. Other risk factors include mucopurulent cervicitis; multiple, new, or symptomatic male sex partners; and lack of barrier contraceptive use. In some settings, screening based on young age may be as sensitive as criteria that incorporate behavioral and clinical measures. Another strategy is universal testing of all patients in high-prevalence clinic populations (e.g., STD clinics, juvenile detention facilities, and family planning clinics).

The effectiveness of selective screening in reducing the prevalence of chlamydial infection among women has been demonstrated in several studies. In the Pacific Northwest, where extensive screening began in family planning clinics in 1998 and in STD clinics in 1993, the prevalence declined from 10% in the 1980s to <5% in 2000. Similar trends have occurred in association with screening programs elsewhere. In addition, screening can effect a reduction in upper genital tract disease. In Seattle, women at a large health maintenance organization who were screened for chlamydial infection on a routine basis had a lower incidence of symptomatic PID than did women who received standard care and underwent more selective screening.

In settings with low to moderate prevalence, the prevalence at which selective screening becomes more cost-effective than universal screening must be defined. Most studies have concluded that universal screening is preferable in settings with a chlamydial prevalence of >3–7%. Depending on the criteria used, selective screening is likely to be more cost-effective when prevalence falls below 3%. Nearly all regions of the United States have now initiated screening programs, particularly in family planning and STD clinics. Along with single-dose therapy, the availability of highly sensitive and specific diagnostic NAATs using urine specimens and self-obtained vaginal swabs makes it feasible to mount an effective nationwide *Chlamydia* control program, with screening of high-risk individuals in traditional health-care settings and in novel outreach and community-based settings.

Trachoma

EPIDEMIOLOGY

Trachoma—a sequela of ocular disease in developing countries—continues to be a leading cause of preventable infectious blindness worldwide. The WHO estimates that ~6 million people have been blinded by trachoma and that ~1.3 million people in developing countries still suffer from preventable blindness due to trachoma; certainly hundreds of millions live in trachoma-endemic areas. Foci of trachoma persist in Australia, the South Pacific, and Latin America. Serovars A, B, Ba, and C are isolated from patients with clinical trachoma in areas of endemicity in developing countries in Africa, the Middle East, Asia, and South America.

The trachoma-hyperendemic areas of the world are in northern and sub-Saharan Africa, the Middle East, drier regions of the Indian subcontinent, and Southeast Asia. In hyperendemic areas, the prevalence of trachoma is essentially 100% by the second or third year of life. Active disease is most common among young children, who are the reservoir for trachoma. By adulthood, active infection is infrequent but sequelae result in blindness. In such areas, trachoma constitutes the major cause of blindness.

Trachoma is transmitted through contact with discharges from the eyes of infected patients. Transmission is most common under poor hygienic conditions and most often takes place between family members or between families with shared facilities. Flies can also transfer the mucopurulent ocular discharges, carrying the organisms on their legs from one person to another. The International Trachoma Initiative founded by the WHO in 1998 aims to eliminate blinding trachoma globally by 2020.

CLINICAL MANIFESTATIONS

Both endemic trachoma and adult inclusion conjunctivitis present initially as conjunctivitis characterized by small lymphoid follicles in the conjunctiva. In regions with hyperendemic classic blinding trachoma, the disease usually starts insidiously before the age of 2 years. Reinfection is common and probably

contributes to the pathogenesis of trachoma. Studies using polymerase chain reaction (PCR) or other NAATs indicate that chlamydial DNA is often present in the ocular secretions of patients with trachoma, even in the absence of positive cultures. Thus persistent infection may be more common than was previously thought.

The cornea becomes involved, with inflammatory leukocytic infiltrations and superficial vascularization (pannus formation). As the inflammation continues, conjunctival scarring eventually distorts the eyelids, causing them to turn inward so that the lashes constantly abrade the eyeball (trichiasis and entropion); eventually the corneal epithelium is abraded and may ulcerate, with subsequent corneal scarring and blindness. Destruction of the conjunctival goblet cells, lacrimal ducts, and lacrimal gland may produce a "dry-eye" syndrome, with resultant corneal opacity due to drying (xerosis) or secondary bacterial corneal ulcers.

Communities with blinding trachoma often experience seasonal epidemics of conjunctivitis due to *H. influenzae* that contribute to the intensity of the inflammatory process. In such areas, the active infectious process usually resolves spontaneously in affected persons at 10–15 years of age, but conjunctival scars continue to shrink, producing trichiasis and entropion with subsequent corneal scarring in adults. In areas with milder and less prevalent disease, the process may be much slower, with active disease continuing into adulthood; blindness is rare in these cases.

Eye infection with oculogenital *C. trachomatis* strains in sexually active young adults presents as an acute onset of unilateral follicular conjunctivitis and preauricular lymphadenopathy similar to that seen in acute conjunctivitis caused by adenovirus or HSV. If untreated, the disease may persist for 6 weeks to 2 years. It is frequently associated with corneal inflammation in the form of discrete opacities ("infiltrates"), punctate epithelial erosions, and minor degrees of superficial corneal vascularization. Very rarely, conjunctival scarring and eyelid distortion occur, particularly in patients treated for many months with topical glucocorticoids. Recurrent eye infections develop most often in patients whose sexual partners are not treated with antimicrobial agents.

DIAGNOSIS

The clinical diagnosis of classic trachoma can be made if two of the following signs are present: (1) lymphoid follicles on the upper tarsal conjunctiva; (2) typical conjunctival scarring; (3) vascular pannus; or (4) limbal follicles or their sequelae, Herbert—s pits. The clinical diagnosis of endemic trachoma should be confirmed by laboratory tests in children with relatively marked degrees of inflammation. Intracytoplasmic chlamydial inclusions are found in 10–60% of Giemsa-stained conjunctival smears in such populations, but chlamydial NAATs are more sensitive and are often positive when smears or cultures are negative. Follicular conjunctivitis in European or American adults living in trachomatous regions is rarely due to trachoma.

TREATMENT: TRACHOMA

Adult inclusion conjunctivitis responds well to treatment with the same regimens used in uncomplicated genital infections—namely, azithromycin (a 1-g single oral dose) or doxycycline (100 mg twice daily for 7 days). Simultaneous treatment of all sexual partners is necessary to prevent ocular reinfection and chlamydial genital disease. Topical antibiotic treatment is not required for patients who receive systemic antibiotics.

PSITTACOSIS

Psittacine birds and many other avian species act as natural reservoirs for *C. psittaci*–type organisms, common pathogens in domestic mammals and birds. The species *C. psittaci,* which now includes only avian strains, affects humans only as a zoonosis. (The other strains previously included in this species have been placed into different species that reflect the animals they infect: *C. abortus*, *C. muridarum*, *C. suis, C. felis*, and *C. caviae.*) Although all birds are susceptible, pet birds (parrots, parakeets, macaws, and cockatiels) and poultry (turkeys and ducks) are most frequently involved in transmission of *C. psittaci* to humans. Exposure is greatest in poultry-processing workers and in owners of pet birds. Infectious forms of the organisms are shed from both symptomatic and apparently healthy birds and may remain viable for several months. *C. psittaci* can be transmitted to humans by direct contact with infected birds or by inhalation of aerosols from avian nasal discharges and from infectious avian fecal or feather dust. Transmission from person to person has never been demonstrated.

The diagnosis is usually established serologically. Psittacosis in humans may present as acute primary atypical pneumonia (which can be fatal in up to 10% of untreated cases); as severe chronic pneumonia; or as a mild illness or asymptomatic infection in persons exposed to infected birds.

Epidemiology

Fewer than 50 confirmed cases of psittacosis are reported in the United States each year, although many more cases probably occur than are reported. Control of psittacosis depends on control of avian sources of infection. A pandemic of psittacosis was once stopped by banning shipment or importation of psittacine birds. Birds can receive prophylaxis in the form of a tetracycline-containing feed. Imported birds are currently quarantined for 30 days of treatment.

Clinical Manifestations

Typical symptoms include fever, chills, muscular aches and pains, severe headache, hepato- and/or splenomegaly, and gastrointestinal symptoms. Cardiac complications may involve endocarditis and myocarditis. Fatal cases were common in the preantibiotic era. As a result of quarantine of imported birds and improved veterinary-hygienic measures, outbreaks and sporadic cases of psittacosis are now rare. Severe pneumonia requiring management in an intensive care unit may develop. Endocarditis, hepatitis, and neurologic complications may occur, and fatal cases have been reported. The incubation period is usually 5–19 days but can last as long as 28 days.

Diagnosis

Previously, the most widely used serologic test for diagnosing chlamydial infections was the genus-specific CF test, in which assay of paired serum specimens often shows fourfold or greater increases in antibody titer. The CF test remains useful, but the gold standard of serologic tests is now the MIF test, which is not widely available (see section on diagnosis of *C. trachomatis* genital infection, above). Any antibody titer above 1:16 is considered significant evidence of exposure to chlamydiae, and a fourfold titer rise in paired sera in combination with a clinically compatible syndrome can be used to diagnose psittacosis. Some commercially available serologic tests based on measurement of antibodies to LPS can be useful when the clinical diagnosis is consistent with bird exposure; however, since these tests are reactive for all chlamydiae (i.e., all chlamydiae contain LPS), caution must be used in their interpretation.

Treatment: Psittacosis Reservoirs

The antibiotic of choice is tetracycline; the dosage for adults is 250 mg four times a day, continued for at least 3 weeks to avoid relapse. Severely ill patients may need cardiovascular and respiratory support. Erythromycin (500 mg four times a day by mouth) is an alternative therapy.

CHLAMYDIA PNEUMONIAE INFECTIONS

C. pneumoniae is a common cause of human respiratory diseases, such as pneumonia and bronchitis. This organism has been reported to account for as many as 10% of cases of community-acquired pneumonia, most of which are diagnosed by serology. Serologic studies have linked *C. pneumoniae* to atherosclerosis; isolation and PCR detection in cardiovascular tissues have also been reported. These findings suggest an expanded range of diseases and syndromes for *C. pneumoniae.* The role of *C. pneumoniae* in the etiology of atherosclerosis has been discussed since 1988, when Finnish researchers presented serologic evidence of an association of this organism with coronary heart disease and acute myocardial infarction. Subsequently, the organism was identified in atherosclerotic lesions by culture, PCR, immunohistochemistry, and transmission electron microscopy; however, discrepant study results (including those of animal studies) and failure of large-scale treatment studies have raised doubts as to the etiologic role of *C. pneumoniae* in atherosclerosis. Large-scale case-cohort studies have recently demonstrated some association of *C. pneumoniae* with lung cancer, as evaluated by serology.

Epidemiology

Primary infection occurs mainly in school-aged children and reinfection in adults. Seroprevalence rates of 40–70% show that *C. pneumoniae* is widespread in both industrialized and developing countries. Seropositivity usually is first detected at school age, and rates generally increase by ~10% per decade. About 50% of individuals have detectable antibody at 30 years of age, and most have detectable antibody by the eighth decade of life. Although serologic evidence suggests that *C. pneumoniae* may be associated with up to 10% of cases of community-acquired pneumonia, most of this evidence is based not on paired serum samples but rather on a single high IgG titer. Some doubt exists about the true prevalence and etiologic role of *C. pneumoniae* in atypical pneumonia, especially since reports of cross-reactivity have raised questions about the specificity of serology when only a single serum sample is used for diagnosis.

Pathogenesis

Little is known about the pathogenesis of *C. pneumoniae* infection. It begins in the upper respiratory tract and, in many persons, persists as a prolonged asymptomatic condition of the upper respiratory mucosal surfaces. However, evidence of replication within vascular endothelium and synovial membranes of joints shows that, in at least some individuals, the organism is transported to distant sites, perhaps within macrophages. A *C. pneumoniae* outer-membrane protein may induce host immune responses whose cross-reactivity with human proteins results in an autoimmune reaction.

As mentioned above, epidemiologic studies have demonstrated an association between serologic evidence of *C. pneumoniae* infection and atherosclerotic disease of the coronary and other arteries. In addition, *C. pneumoniae* has been identified in atherosclerotic plaques by electron microscopy, DNA hybridization, and immunocytochemistry. The organism has been recovered in culture from atheromatous plaque—a result indicating the presence of viable replicating bacteria in vessels.

Evidence from animal models supports the hypothesis that *C. pneumoniae* infection of the upper respiratory tract is followed by recovery of the organism from atheromatous lesions in the aorta and that the infection accelerates the process of atherosclerosis, especially in hypercholesterolemic animals. Antimicrobial treatment of the infected animals reverses the increased risk of atherosclerosis. In humans, two small trials in patients with unstable angina or recent myocardial infarction suggested that antibiotics reduce the likelihood of subsequent untoward cardiac events. However, larger-scale trials have not documented an effect of various antichlamydial regimens on the risk of these events.

Clinical Manifestations

C. pneumoniae was first reported as the etiologic agent of mild atypical pneumonia in military recruits and college students. The clinical spectrum of *C. pneumoniae* infection includes acute pharyngitis, sinusitis, bronchitis, and pneumonitis, primarily in young adults. The clinical manifestations of primary infection appear to be more severe and prolonged than those of reinfection. The pneumonitis of *C. pneumoniae* pneumonia resembles that of *Mycoplasma* pneumonia in that leukocytosis is frequently lacking and patients often have prominent antecedent upper respiratory tract symptoms, fever, nonproductive cough, mild to moderate illness, minimal findings on chest auscultation, and small segmental infiltrates on chest x-ray. In elderly patients, pneumonia due to *C. pneumoniae* can be especially severe and may necessitate hospitalization and respiratory support.

Chronic infection with *C. pneumoniae* has been reported among patients with chronic obstructive pulmonary disease and may also play a role in the natural history of asthma, including exacerbations. The clinical symptoms of respiratory infections caused by *C. pneumoniae* are nonspecific and do not differ from those caused by other agents of atypical pneumonia, such as *Mycoplasma pneumoniae*.

Diagnosis

Serology, PCR amplification, and culture can be used to diagnose *C. pneumoniae* infection. Serology has been the traditional method of diagnosing infection by *C. pneumoniae.* The gold standard serologic test is the MIF test (see section on diagnosis of *C. trachomatis* genital infection, above). Any antibody titer above 1:16 is considered significant evidence of exposure to chlamydiae. According to a CDC-sponsored expert working group, the diagnosis of acute *C. pneumoniae* infection requires demonstration of a fourfold rise in titer in paired serum samples. There are no official recommendations for diagnosis of chronic infections, although many research studies have used high titers of IgA as an indicator. The older CF tests and EIAs for LPS are not recommended, as they are not specific for *C. pneumoniae* but identify the chlamydiae only to the genus level. The organism is very difficult to grow in tissue culture but has been cultivated in HeLa cells, HEp-2 cells, and HL cells. Although NAATs are commercially available for *C. trachomatis,* only research-based PCR assays are available for *C. pneumoniae.*

Treatment: *C. Pneumoniae* Infections

Although few controlled trials of treatment have been reported, *C. pneumoniae* is inhibited in vitro by erythromycin, tetracycline, azithromycin, clarithromycin, gatifloxacin, and gemifloxacin. Recommended therapy consists of 2 g/d of either tetracycline or erythromycin for 10–14 days. Other macrolides (e.g., azithromycin) and some fluoroquinolones (e.g., levofloxacin and gatifloxacin) also appear to be effective.

Acknowledgment

The authors wish to acknowledge the late Walter E. Stamm, MD, for his significant contributions to the field of Chlamydia research. Dr. Stamm wrote the chapters on chlamydiae for previous editions of Harrison—s Principles of Internal Medicine , and we thank the editors for permission to reproduce Figs. 176-1 and 176-2 as well as Table 176-1 from his chapter in the 17th edition. Dr. Stamm died on December 14, 2009, and this chapter is dedicated to him.

FURTHER READINGS

Association of Public Health Laboratories: *Guidelines for the Laboratory Testing of STDs, 2009.* Silver Spring, MD, Association of Public Health Laboratories, 2009 (*www.aphl.org/aphlprograms/infectious/std/Pages/stdtestingguidelines.aspx)*

Bavoil PM, Wyrick PB: *Chlamydia: Genomics and Pathogenesis*. Norfolk, UK, Horizon Bioscience, 2006, 542 pp

Campbell LA et al: Chlamydial infections, in *Manual of Molecular and Clinical Laboratory Immunology*, 7th ed, B Detrick et al (eds). Washington, DC, ASM Press, 2006, pp 518–525

Centers for Disease Control and Prevention: *Sexually Transmitted Disease Surveillance, 2009.* Atlanta, U.S. Department of Health and Human Services, 2010 (www.cdc.gov/std/stats09/main.htm)

——— : Sexually transmitted diseases treatment guidelines, 2010. MMWR Recomm Rep 59(RR-12):1

–110, 2010

Gaydos CA: Chlamydiae, in *Clinical Virology Manual* , 4th ed, S Spector et al (eds). Washington, DC, ASM Press, 2009, pp 630–640

Kuypers J et al: Principles of laboratory diagnosis of STIs, in *Sexually Transmitted Diseases*, 4th ed, KK Holmes et al (eds). New York, McGraw-Hill, 2008, pp 937–948

Schachter J, Stephens RS: Biology of *Chlamydia trachomatis*, in *Sexually Transmitted Diseases*, 4th ed, KK Holmes et al (eds). New York, McGraw-Hill, 2008, pp 555–574

Scholes D et al: Prevention of pelvic inflammatory disease by screening for cervical chlamydial infection. N Engl J Med 334:1362, 1996[PMID: 8614421]

Taylor HR: *Trachoma: A Blinding Scourge from the Bronze Age to the Twenty-First Century* . East Melbourne, Victoria, Australia, Centre for Eye Research Australia/Haddington Press, 2008, 282 pp

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINING A VIRUS

Viruses consist of a nucleic acid surrounded by one or more proteins. Some viruses also have an outer-membrane envelope. Viruses are obligate intracellular parasites: they can replicate only within cells since their nucleic acids do not encode the many enzymes necessary for protein, carbohydrate, or lipid metabolism and for the generation of high-energy phosphates. Typically, viral nucleic acids encode proteins necessary for replicating and pack–aging the nucleic acids within the biochemical milieu of host cells.

Viruses differ from virusoids, viroids, and prions. *Virusoids* are nucleic acids that depend on helper viruses to package their nucleic acids into virus-like particles. *Viroids* are naked, cyclical, mostly double-strand, small RNAs that appear to be restricted to plants, spread from cell to cell, and are replicated by cellular RNA polymerase II. *Prions* (Chap. 383) are abnormal protein molecules that can spread, reproducing by changing the structure of their normal cellular protein counterparts. Prions have been implicated in neurodegenerative conditions such as Creutzfeldt-Jakob disease, Gerstmann-Straüssler disease, kuru, and human bovine spongiform encephalopathy ("mad cow disease").

VIRAL STRUCTURE

Viral genomes consist of (1) a single-strand or double-strand DNA, (2) a single-strand sense RNA, (3) a single-strand or segmented antisense RNA, or (4) a double-strand segmented RNA genome. The viral nucleic acid may encode only a few genes or more than 100. Sense-strand RNA genomes can be translated directly into protein, whereas antisense RNAs must be copied into translatable RNA. Sense and antisense genomes are also referred to as *positive-strand* and *negative-strand genomes*, respectively. Viral nucleic acid is usually associated with one or more virus-encoded nucleoproteins in the core of the viral particle. The viral nucleic acid and nucleoproteins are almost always enclosed in a protein shell called a *capsid*. Because of the limited genetic complexity of viruses, their capsids are usually composed of multimers of identical capsomeres. Capsomeres are in turn composed of one or a few proteins. Capsids have icosahedral or helical symmetry. Icosahedral structures approximate spheres but have two-, three-, and fivefold axes of symmetry, while helical structures have only a twofold axis of symmetry. The entire structural unit of nucleic acid, nucleoprotein(s), and capsid is called a *nucleocapsid*.

Many human viruses are composed simply of a core and a capsid. For these viruses, the outer surface of the capsid mediates contact with uninfected cells. Other viruses are more complex and have an outer lipid-containing envelope derived from virus-modified membranes of the infected cell. The piece of infected-cell membrane that becomes the viral envelope has usually been modified during infection by the insertion of virus-encoded glycoproteins, which mediate contact of enveloped viruses with uninfected cells. Matrix or tegument proteins fill the space between the nucleocapsid and the envelope in many enveloped viruses. In general, enveloped viruses are sensitive to lipid solvents and nonionic detergents that can dissolve the envelope, while viruses that consist only of nucleocapsids are somewhat resistant. A schematic diagram for large and complex herpesviruses is shown in **Fig. 177-1**. Prototypical pathogenic human viruses are listed in Table 177-1. The relative sizes and structures of typical pathogenic human viruses are shown in **Fig. 177-2**.

Figure 177-1

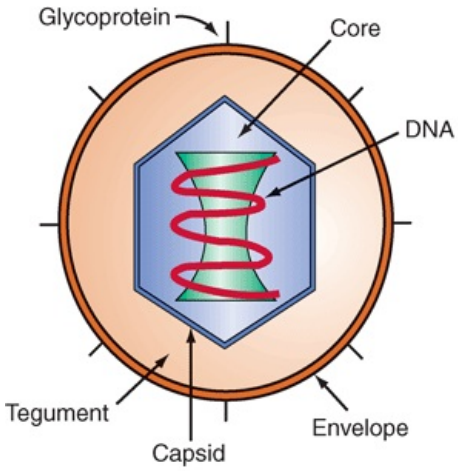

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of an enveloped herpesvirus with an icosahedral nucleocapsid. The approximate respective dimensions of the nucleocapsid and the enveloped particles are 110 and 180 nm. The capsid is composed of 162 capsomeres: 150 with sixfold and 12 with fivefold axes of symmetry.

Table 177-1 Virus Families Pathogenic for Humans

Family	Representative Viruses	Type of RNA/DNA	Lipid Envelope
RNA Viruses			
Picornaviridae	Poliovirus Coxsackievirus Echovirus Enterovirus Rhinovirus Hepatitis A virus	(+) RNA	No
Caliciviridae	Norwalk agent Hepatitis E virus	(+) RNA	No
Togaviridae	Rubella virus Eastern equine encephalitis virus Western equine encephalitis virus	(+) RNA	Yes
Flaviviridae	Yellow fever virus Dengue virus St. Louis encephalitis virus West Nile virus Hepatitis C virus Hepatitis G virus	(+) RNA	Yes
Coronaviridae Rhabdoviridae	Coronaviruses[a] Rabies virus Vesicular stomatitis virus	(+) RNA (–) RNA	Yes Yes
Filoviridae	Marburg virus Ebola virus	(–) RNA	Yes
Paramyxoviridae	Parainfluenza virus Respiratory syncytial virus	(–) RNA	Yes

	Newcastle disease virus Mumps virus Rubeola (measles) virus		
Orthomyxoviridae	Influenza A, B, and C viruses	(–) RNA, 8 segments	Yes
Bunyaviridae	Hantavirus California encephalitis virus Sandfly fever virus	(–) RNA, 3 circular segments	Yes
Arenaviridae	Lymphocytic choriomeningitis virus Lassa fever virus South American hemorrhagic fever virus	(–) RNA, 2 circular segments	Yes
Reoviridae	Rotavirus Reovirus Colorado tick fever virus	ds RNA, 10–12 segments	No
Retroviridae	Human T lymphotropic virustypes I and II Human immunodeficiency virustypes 1 and 2	(+) RNA, 2 identical segments	Yes
DNA Viruses			
Hepadnaviridae	Hepatitis B virus	ds DNA with ss portions	Yes
Parvoviridae	Parvovirus B19	ss DNA	No
Papovaviridae	Human papillomaviruses JC virus BK virus	ds DNA	No
Adenoviridae	Human adenoviruses	ds DNA	No
Herpesviridae	Herpes simplex virus types 1 and 2[b] Varicella-zoster virus[c] Epstein-Barr virus[d] Cytomegaloviruse Human herpesvirus 6 Human herpesvirus 7 Kaposi's sarcoma—associated herpesvirus[f]	ds DNA	Yes
Poxviridae	Variola (smallpox) virus Orf virus Molluscum contagiosum virus	ds DNA	Yes

[a] Including the coronavirus causing severe acute respiratory syndrome (SARS).

[b] Also called human herpesvirus (HHV) 1 and 2, respectively.

[c] Also called HHV-3.

[d] Also called HHV-4.

[e] Also called HHV-5.

[f] Also called HHV-8.

Abbreviations: ds, double-strand; ss, single-strand.

Figure 177-2

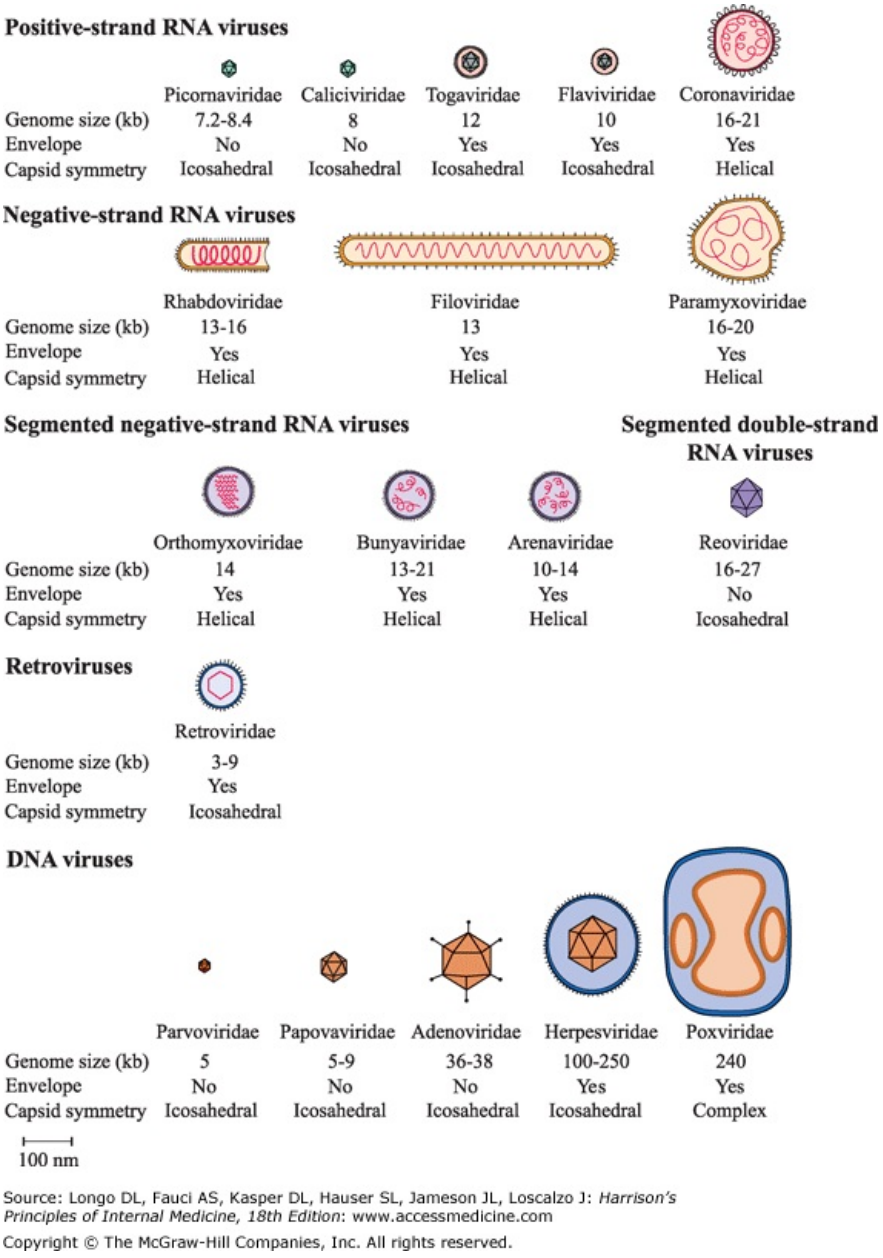

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagrams of the major virus families including species that infect humans. The viruses are grouped by genome type and are drawn approximately to scale. Prototype viruses of each family that cause human disease are listed in Table 177-1.

TAXONOMY OF PATHOGENIC HUMAN VIRUSES

As is apparent from Table 177-1 and Fig. 177-2, the classification of viruses into orders and families is based on nucleic acid composition, nucleocapsid size and symmetry, and presence or absence of an envelope. Viruses of a single family have similar types of genomes and are often morphologically indistinguishable in electron micrographs. Further subclassification into genera depends on similarities in epidemiology, biologic effects, and nucleic acid sequence.

Most human viruses have a common name related to their pathologic effects or the circumstances of their discovery. Formal species names have been assigned by the International Committee on Taxonomy of Viruses. The formal designation consists of the name of the host followed by the family or genus of the virus and a number. This dual terminology can cause confusion when viruses are referred to and referenced by either name—e.g., varicella-zoster virus (VZV) or human herpesvirus (HHV) 3.

VIRAL INFECTION IN VITRO

Stages of Viral Infection at the Cellular Level

VIRAL INTERACTIONS WITH THE CELL SURFACE AND CELL ENTRY

All viruses must overcome the barrier posed by the cell's plasma membrane in order to deliver their payload of nucleic acid into the cell. Infection is initiated by attachment of the virus to the cell surface. Various cellular proteins, carbohydrates, and lipids (e.g., heparan sulfate proteoglycans, sialic acids,

and lectins) can act as attachment factors that concentrate viruses at the cell surface through relatively weak or nonspecific interactions with viral surface proteins. Higher-affinity binding of viral surface proteins to specific cell-surface proteins, or receptors (see Table 120-1), is more critical for viral infection. Receptor binding is often augmented by interaction of viral surface proteins with other cell-surface proteins, or co-receptors, important for various aspects of virus entry. Receptors and co-receptors are important determinants of the cell types and species that a virus can infect. For example, the HIV envelope glycoprotein binds to the T cell surface protein CD4 and then engages one of several chemokine receptors that are co-receptors for the virus. Epstein-Barr virus (EBV) glycoprotein gp350 binds to the B lymphocyte complement receptor CD21 and then uses major histocompatibility complex (MHC) class II molecules as a co-receptor.

Viruses use different strategies to penetrate the cell membrane. Some enveloped viruses use membrane fusion to deliver their contents into the cytoplasm. In general, a trigger (e.g., receptor binding) induces a conformational change that allows the viral surface protein to extend into the cell membrane, bringing the virus and cell membrane into close proximity and thereby enabling fusion and formation of a pore through which the viral nucleocapsid can be delivered into the cytoplasm. Nonenveloped viruses and some enveloped viruses cannot use direct membrane fusion at the plasma membrane and are internalized by endocytosis. The low pH in endosomes can trigger viral membrane or capsid fusion with the endocytic membrane. Conformational changes in nonenveloped capsids can lead to endosomal membrane penetration and release of viral nucleic acid into the cytoplasm.

Influenza virus provides a well-studied example of the effect of low pH on viral penetration. Influenza hemagglutinin mediates adsorption, receptor aggregation, and endocytosis. In low-pH endosomes, changes in conformation of the hemagglutinin expose amphipathic domains that interact chemically with the cell membrane and initiate fusion of the virus and cell membranes. For influenza virus, the M2 membrane protein plays a key role in the uncoating of the viral envelope by providing an ion channel in the envelope.

The fusion of viral proteins with cell membranes is a crucial step in viral infection. The hydrophobic interactions required for fusion can be susceptible to chemical inhibition or blockade. The HIV envelope glycoprotein gp120 is associated with gp41 on the viral surface. Binding of HIV gp120 to CD4 and chemokine receptors results in a conformational change, allowing gp41 to initiate cell membrane fusion. Enfuvirtide is a small-peptide drug derived from gp41 that binds to gp41 and prevents the conformational change required for fusion. Maraviroc prevents virus entry by binding to CCR5, blocking interaction with gp120, and preventing fusion triggering.

VIRAL GENE EXPRESSION AND REPLICATION

After uncoating and release of viral nucleoprotein into the cytoplasm, the viral genome is transported to a site for expression and replication. In order to produce infectious progeny, viruses must (1) produce proteins necessary to replicate their nucleic acid, (2) produce structural proteins, and (3) assemble the nucleic acid and proteins into progeny virions. Different viruses use different strategies and gene repertoires to accomplish these goals. DNA viruses, except for poxviruses, replicate their nucleic acid and assemble into nucleocapsid complexes in the cell nucleus. RNA viruses, except for influenza viruses, transcribe and replicate their nucleic acid and assemble entirely in the cytoplasm. Thus, the replication strategies of DNA and RNA viruses are presented separately below. Positive-strand and negative-strand RNA viruses are discussed separately. Medically important viruses of each group are used for illustrative purposes.

Positive-Strand RNA Viruses

Medically important positive-strand RNA viruses include picornaviruses, flaviviruses, togaviruses, caliciviruses, and coronaviruses. Genomic RNA from positive-strand RNA viruses is released into the cytoplasm without associated enzymes. Cell ribosomes recognize and associate with an internal ribosome entry sequence in the viral RNA and translate a polyprotein. Protease components of the polyprotein cleave out the viral RNA polymerase and other viral proteins necessary for replication. Antigenomic RNA is then transcribed from the genomic RNA template. Positive-strand genomes and mRNAs are next transcribed from the antigenomic RNA by the viral RNA polymerase and are translated into capsid proteins. Genomic RNA is encapsidated in the cytoplasm as the infected cell undergoes lysis.

Negative-Strand RNA Viruses

Medically important negative-strand RNA viruses include rhabdoviruses, filoviruses, paramyxoviruses, orthomyxoviruses, and bunyaviruses. The genomes of negative-strand viruses are frequently segmented. Negative-strand RNA virus genomes are released into the cytoplasm with an associated RNA polymerase and one or more polymerase accessory proteins. The viral RNA polymerase transcribes messenger RNAs (mRNAs) as well as full-length antigenomic RNA, which is the template for genomic RNA replication. Viral mRNAs encode the viral RNA polymerase and accessory factors as well as viral structural proteins. Except for influenza virus, which transcribes its mRNAs and antigenomic RNAs in the cell's nucleus, negative-strand RNA viruses replicate entirely in the cytoplasm. All negative-strand RNA viruses, including influenza viruses, assemble in the cytoplasm.

Double-Strand Segmented RNA Viruses

Double-strand RNA viruses are taxonomically grouped in the family Reoviridae. The medically important viruses in this group are rotaviruses and Colorado tick fever virus. Reovirus genomes have 10–12 RNA segments. Reovirus particles contain an RNA polymerase complex. These viruses replicate and assemble in the cytoplasm.

DNA Viruses

Medically important DNA viruses include parvoviruses, papovaviruses [e.g., human papillomaviruses (HPVs) and polyomaviruses], adenoviruses, herpesviruses, and poxviruses. Most DNA virus genomes enter the cell's nucleus and are transcribed by cellular RNA polymerase II. For example, after receptor binding and fusion with plasma membranes or endocytic vesicle membranes, herpesvirus nucleocapsids are released into the cytoplasm along with tegument proteins. The nucleocapsid is transported along microtubules to a nuclear pore. Capsids then release DNA into the nucleus.

DNA virus transcription and mRNA processing depend on both viral and cellular proteins. For herpes simplex virus (HSV), a viral tegument protein enters the nucleus and activates immediate-early genes, the first genes expressed after infection. Transcription of immediate-early genes requires the viral tegument protein and cell transcription factors. HSV becomes nonreplicating, or latent, in neurons because essential cell transcription factors for viral immediate-early gene expression are docked in the cytoplasm in neurons. Heat shock or other cell stresses can cause these cell factors to enter the nucleus, activate viral gene expression, and initiate replication. This information explains HSV-1 latency in neurons and activation of lytic infection.

For adenoviruses and herpesviruses, immediate-early gene transcription results in expression of early proteins necessary for viral DNA replication. Viral DNA synthesis is required to turn on late gene expression and production of viral structural components. The HPVs, polyomaviruses, and parvoviruses are not dependent on transactivators encoded from the viral genome for early-gene transcription. Instead, their early genes have upstream enhancing elements that bind cell transcription factors. The early genes encode proteins that are necessary for viral DNA synthesis and late-gene transcription. DNA virus late genes encode structural proteins necessary for viral assembly and for viral egress from the infected cell. Late-gene transcription is continuously dependent on DNA replication. Therefore, inhibitors of DNA replication also stop late-gene transcription.

Each DNA virus family uses unique mechanisms for replicating its DNA. Adenovirus and herpesvirus DNAs are linear in the virion. Adenovirus DNA remains linear in infected cells and replicates as a linear genome, using an initiator protein-DNA complex. In contrast, herpesvirus DNA circularizes in the infected cell, and genomes replicate into linear concatemers through a "rolling-circle"mechanism. Full-length DNA genomes are cleaved and packaged into virus. Herpesviruses encode a DNA polymerase and at least six other viral proteins necessary for viral DNA replication. These viruses also encode enzymes that increase the deoxynucleotide triphosphate pools. HPV and polyomavirus DNAs are circular both within the virus and in infected cells. These genomes are reproduced by cellular DNA replication enzymes and remain circular through replication and packaging. HPV and polyomavirus early proteins are necessary for latent and lytic viral DNA replication. Early viral proteins stimulate cells to remain in cycle, facilitating viral DNA replication.

Parvoviruses have negative single-strand DNA genomes and are the smallest DNA viruses. Their genomes are half the size of the papovavirus genomes and include only two genes. The replication of autonomous parvoviruses, such as B19, depends on cellular DNA replication and requires the virus-encoded Rep protein. Other parvoviruses, such as adeno-associated virus (AAV), are not autonomous and require helper viruses of the adenovirus or herpesvirus family for their replication. AAV is being used as a potentially safe human gene therapy vector because its replication protein causes integration at a single chromosome site. The small genome size limits the range of proteins that can be expressed from AAV vectors.

Poxviruses are the largest DNA viruses. They are unique among DNA viruses in replicating and assembling in the cytoplasm. To accomplish cytoplasmic replication, poxviruses encode transcription factors, an RNA polymerase II orthologue, enzymes for RNA capping, enzymes for RNA polyadenylation, and enzymes for viral DNA synthesis. Poxvirus DNA also has a unique structure. The double-strand linear DNA is covalently linked at the ends; the packaged genome is therefore a covalently closed single-strand circle. In addition, there are inverted repeats at the ends of the linear DNA. During DNA replication, the genome is cleaved within the terminal inverted repeat, and the inverted repeats self-prime complementary-strand synthesis by the virus-encoded DNA polymerase. Like herpesviruses, poxviruses encode several enzymes that increase deoxynucleotide triphosphate precursor levels and thus facilitate viral DNA synthesis.

Viruses that Use Both RNA and DNA Genomes in Their Life Cycle

Retroviruses, including HIV, are RNA viruses that use a DNA intermediate to replicate their genomes; hepatitis B virus (HBV) is a DNA virus that uses an RNA intermediate to replicate its genome. Thus these viruses are not purely RNA or DNA viruses. Retroviruses are enveloped RNA viruses with two identical sense-strand genomes and associated reverse transcriptase and integrase enzymes.

Retroviruses differ from all other viruses in that they reverse-transcribe themselves into partially duplicated double-strand DNA copies and then routinely integrate into the host genome as part of their replication strategy. The fact that remnants and even complete copies of simple retroviral DNA are integrated into the human genome raises the possibility of replication-competent simple human retroviruses. However, replication has not been documented or associated with any disease. Integrated, replication-competent retroviral DNAs are also present in many animal species, such as pigs. These porcine retroviruses are a potential cause for concern in xenotransplantation because retroviral replication could cause disease in humans.

Cellular RNA polymerase II and transcription factors regulate transcription from the integrated provirus DNA genome. Some retroviruses also encode for regulators of transcription and RNA processing, such as Tax and Rex in human T lymphotropic virus (HTLV) types I and II. HIV-1 and HIV-2 have orthologous Tat and Rev genes as well as the additional accessory proteins Vpr, Vpu, and Vif, which are important for efficient infection and immune escape. Full-length proviral transcripts are made from a promoter in the viral terminal repeat and serve as both genomic RNAs that will be packaged in the nucleocapsids and differentially spliced mRNAs that encode for the viral Gag protein, polymerase/integrase protein, and envelope glycoprotein. The Gag protein includes a protease that cleaves it into several components, including a viral matrix protein that coats the viral RNA. Viral RNA polymerase/integrase, matrix protein, and cellular tRNA are key components of the viral nucleocapsid. The HIV reverse transcriptase, integrase, and Gag protease are important targets for inhibition of HIV replication.

HBV replication is unique in several respects. HBV has a partially double-strand DNA genome that is repaired to a fully double-strand circular DNA by the virion polymerase upon entry into an infected cell. Viral mRNAs are transcribed from the closed circular viral episome by cellular RNA polymerase II and are translated to produce viral proteins including core protein, surface antigen, and polymerase. In addition, a full-genome-length mRNA is packaged into viral core particles in the cytoplasm of infected cells as an intermediate for viral DNA replication. This RNA associates with the viral polymerase, which also has reverse transcriptase activity, to convert the full-length encapsidated RNA genome into partially double-strand DNA. HBV is believed to mature by budding through the cell's plasma membrane, which has been modified by the insertion of viral surface antigen protein.

VIRAL ASSEMBLY AND EGRESS

For most viruses, nucleic acid and structural protein synthesis is accompanied by the assembly of protein and nucleic acid complexes. The assembly and egress of mature infectious virus mark the end of the eclipse phase of infection, during which infectious virus cannot be recovered from the infected cell. Nucleic acids from RNA viruses and poxviruses assemble into nucleocapsids in the cytoplasm. For all DNA viruses except poxviruses, viral DNA assembles into nucleocapsids in the nucleus. In general, the capsid proteins of viruses with icosahedral nucleocapsids can self-assemble into densely packed and highly ordered capsid structures. Herpesviruses require an assemblin protein as a scaffold for capsid assembly. Viral nucleic acid then spools into the assembled capsid. For herpesviruses, a full unit of the viral DNA genome is packaged into the capsid, and a capsid-associated nuclease cleaves the viral DNA at both ends. In the case of viruses with helical nucleocapsids, the protein component appears to assemble around the nucleic acid, which contributes to capsid organization.

Viruses must egress from the infected cell and not bind back to their receptor(s) on the outer surface of the plasma membrane. Viruses can acquire envelopes from cytoplasmic membranes or by budding through the cell's plasma membrane. Excess viral membrane glycoproteins are synthesized to saturate cell receptors and facilitate separation of the virus from the infected cell. Some viruses encode membrane proteins with enzymatic activity for receptor destruction. Influenza virus, for example, encodes a glycoprotein with neuraminidase activity. Neuraminidase destroys sialic acid on the infected cell's plasma membrane so that newly released virus does not get stuck to the dying cell. Herpesvirus nucleocapsids acquire an initial envelope by assembling in the nucleus and then budding through the nuclear membrane into the endoplasmic reticular space. The initially enveloped herpesvirus is then de-enveloped and released from the cell either by exocytosis or by re-envelopment at the plasma membrane. Nonenveloped viruses depend on the death and dissolution of the infected cell for their release.

Fidelity of Viral Replication

Hundreds or thousands of progeny may be produced from a single virus-infected cell. Many particles partially assemble and never mature into virions. Many mature-appearing virions are imperfect and have only incomplete or nonfunctional genomes. Despite the inefficiency of assembly, a typical virus-infected cell releases 10–1000 infectious progeny. Some of these progeny may contain genomes that differ from those of the virus that infected the cell. Smaller, "defective" viral genomes have been noted with the replication of many RNA and DNA viruses. Virions with defective genomes can be produced in large numbers through packaging of incompletely synthesized nucleic acid. Adenovirus packaging is notoriously inefficient, and a high ratio of particle to infectious virus may limit the amount of recombinant adenovirus that can be administered for gene therapy since the immunogenicity of defective particles may contribute to adverse effects.

Changes in viral genomes can lead to mutant viruses of medical significance. In general, viral nucleic

acid replication is more error-prone than cellular nucleic acid replication. RNA polymerases and reverse transcriptases are significantly more error-prone than DNA polymerases. Mutations can also be introduced into the HIVgenome by APOBEC3G, a cellular protein that is packaged in the virion. APOBEC3G deaminates cytidine in the virion RNA to uridine. When reverse transcriptase subsequently uses the altered virion RNA as a template in the infected cell, a guanosine-to-adenosine mutation is introduced into the proviral DNA. Mutations resulting in less efficient viral growth, or fitness, may be detrimental to the virus. HIV-encoded Vif blocks APOBEC3G activity in the virion, inhibiting the debilitating effects of hypermutation on genetic integrity. Nevertheless, mutations resulting in evasion of the host immune response or resistance to antiviral drugs are preferentially selected in patients, with the consequent perpetuation of infection. Viral genomes can also be altered by recombination or reassortment between two related viruses in a single infected cell. While this occurrence is unusual under most circumstances of natural infection, the genome changes can be substantial and can significantly alter virulence or epidemiology. Reassortment of the avian or mammalian influenza A hemagglutiningene into a human influenza background can result in the emergence of new epidemic or pandemic influenza A strains.

Viral Genes Not Required for Viral Replication

Viruses frequently have genes encoding proteins that are not directly involved in replication or packaging of the viral nucleic acid, in virion assembly, or in regulation of the transcription of viral genes involved in those processes. Most of these proteins fall into five classes: (1) proteins that directly or indirectly alter cell growth; (2) proteins that inhibit cellular RNA or protein synthesis so that viral mRNA can be efficiently transcribed or translated; (3) proteins that promote cell survival or inhibit apoptosis so that progeny virus can mature and escape from the infected cell; (4) proteins that inhibit the host interferon response; and (5) proteins that downregulate host inflammatory or immune responses so that viral infection can proceed in an infected person to the extent consistent with the survival of the virus and its efficient transmission to a new host. More complex viruses of the poxvirus or herpesvirus family encode many proteins that serve these functions. Some of these viral proteins have motifs similar to those of cellular proteins, while others are quite novel. Virology has increasingly focused on these more sophisticated strategies evolved by viruses to permit the establishment of long-term infection in humans and other animals. These strategies often provide unique insights into the control of cell growth, cell survival, macromolecular synthesis, proteolytic processing, immune or inflammatory suppression, immune resistance, cytokine mimicry, or cytokine blockade.

MicroRNAs (miRNAs) are small noncoding RNAs that can regulate gene expression at the posttranscriptional level by targeting–and usually silencing–mRNAs. MiRNAs were initially discovered in plants and plant viruses, where they alter expression of cell defensins. Herpesviruses are especially rich in miRNAs; for example, at least 23 miRNAs have been identified in EBV and 11 in cytomegalovirus (CMV). Adenovirus and polyomavirus miRNAs have also been described. Increasing data indicate that animal viruses encode miRNAs to alter the growth and survival of host cells and the innate and acquired immune responses.

Host Range

The concept of host range was originally based on the cell types in which a virus replicates in tissue culture. For the most part, the host range is limited by specific cell-surface proteins required for viral adsorption or penetration—i.e., to the cell types that express receptors or co-receptors for a specific virus. Another common basis for host-range limitation is the degree of transcriptional activity from viral promoters in different cell types. Most DNA viruses depend not only on cellular RNA polymerase II and the basal components of the cellular transcription complex but also on activated components and transcriptional accessory factors, both of which differ among differentiated tissues, among cells at various phases of the cell cycle, and between resting and cycling cells. APOBEC3G, an important cell restriction factor for HIV infection, hypermutates viral RNA. The balance between HIV Vif and APOBEC3G is an important determinant of HIV-1 infection.

The importance of host range factors is illustrated by the effects of specific host determinants that limit the replication of influenza virus with avian or porcine hemagglutinins in humans. These viral proteins have adapted to bind avian or porcine sialic acids, and spread of avian or porcine influenza viruses in human populations is limited by their ability to infect human cells.

Viral Cytopathic Effects and Inhibitors of Apoptosis

The replication of almost all viruses has adverse effects on the infected cell, inhibiting cellular synthesis of DNA, RNA, or proteins through efficient competition for key substrates and enzymatic processes. These general inhibitory effects enable viruses to nonspecifically limit components of innate host resistance, such as interferon (IFN) production. Viruses can specifically inhibit host protein synthesis by attacking a component of the translational initiation complex–frequently, a component that is not required for efficient translation of viral RNAs. Poliovirus protease 2A, for example, cleaves a cellular component of the complex that ordinarily facilitates translation of cellular mRNAs by interacting with their cap structure. Poliovirus RNA is efficiently translated without a cap since it has an internal ribosome entry sequence. Influenza virus inhibits the processing of mRNA by snatching cap structures

from nascent cellular RNAs and using them as primers in the synthesis of viral mRNA. HSV has a virion tegument protein that inhibits cellular mRNA translation.

Apoptosis is the expected consequence of virus-induced inhibition of cellular macromolecular synthesis and viral nucleic acid replication. While the induction of apoptosis may be important for the release of some viruses (particularly nonenveloped viruses), many viruses have acquired genes or parts of genes that enable them to forestall infected-cell death. This delay increases the yield from viral replication. Adenoviruses and herpesviruses encode analogues of the cellular Bc12 protein, which blocks mitochondrial enhancement of proapoptotic stimuli. Poxviruses and some herpesviruses also encode caspase inhibitors. Many viruses, including HPVs and adenoviruses, encode proteins that inhibit p53 or its downstream proapoptotic effects.

VIRAL INFECTION IN VIVO

Transmission

The capsid and envelope of a virus protect the genome and enable efficient transmission of the virus from cell to cell and to new prospective hosts. Most common viral infections are spread by direct contact, by ingestion of contaminated water or food, or by inhalation of aerosolized particles. In all these situations, infection begins on an epithelial or mucosal surface and spreads along the mucosa and into deeper tissues. Infection may spread to cells that can enter blood vessels, lymphatics, or neural circuits. HBV, hepatitis C virus (HCV), HTLV, and HIV are dependent on transmission by parenteral inoculation. Insect vectors can mediate parenteral transfer of viruses that reach high titers in animal or human hosts.

Some viruses are transmitted only between humans. The dependence of smallpox and poliovirus infections on interhuman transmission makes it feasible to eliminate these viruses from human circulation by mass vaccination. Herpesviruses also survive by interhuman transmission but may be more difficult to eliminate because they establish persistent latent infection in humans and continuously reactivate to infect new and naïve generations.

Animals are also important reservoirs and vectors for transmission of viruses causing human disease. Arboviruses are parenterally transmitted from mammalian species to humans by mosquito vectors. Herpes B, monkeypox, rabies, and viral hemorrhagic fevers are other examples of zoonotic infections caused by direct contact with animals, animal tissues, or arthropod vectors.

Primary Infection

Initial viral infections usually last for several days or weeks. During this period, the concentration of virus at sites of infection rises and then falls, usually to unmeasurable levels. The rise and fall of viral replication at a given site depend on local innate immune responses and the access of systemic antibody and cell immune effectors to the virus. Typically, primary infections with enteroviruses, mumps virus, measles virus, rubella virus, rotavirus, influenza virus, AAV, adenovirus, HSV, and VZV are cleared from almost all sites within 3–4 weeks. Some viruses are especially proficient in altering or evading innate and acquired immune responses. Primary infection with AAV, EBV, or CMV can last for several months. Characteristically, primary infections due to HBV, HCV, hepatitis D virus (HDV), HIV, HPV, and molluscum contagiosum virus (MCV) extend beyond several weeks. For some of these viruses (e.g., HPV, HBV, HCV, HDV, and MCV), the manifestations of primary infection are almost indistinguishable from the persistent phase.

Disease manifestations usually arise as a consequence of viral replication, infected cell injury or death, and local inflammatory and innate immune responses. Disease severity may not necessarily correlate with the level of viral replication alone. For example, the clinical manifestations of intense primary infection with poliovirus, enterovirus, rabies virus, measles virus, mumps virus, or HSV at mucosal surfaces may be inapparent or relatively mild, whereas limited replication in neural cells can have dramatic consequences. Similarly, rubella virus or CMV infections in utero or neonatal HSV infections may have much more devastating effects than infections in adults.

Primary infections are cleared by nonspecific innate and specific adaptive immune responses. Thereafter, an immunocompetent host is usually immune to the disease manifestations of reinfection by the same virus. Immunity frequently does not prevent transient surface colonization on re-exposure, persistent colonization, or even limited deeper infection.

Persistent and Latent Infections

Relatively few viruses cause persistent or latent infections. HBV, HCV, rabies virus, measles virus, HIV, HTLV, HPV, HHVs, and MCV are notable exceptions. The mechanisms for persistent infection vary. HCV RNA polymerase and HIV reverse transcriptase are error prone and generate variant genomes. Genome variation can be sufficient to permit evasion of host immune responses, thereby allowing persistent infection. HIV is also directly immunosuppressive, depleting CD4+ T lymphocytes and compromising CD8+ cytotoxic T cell immune responsiveness. Moreover, HIV encodes the Nef protein, which downmodulates MHC class I expression, rendering HIV-infected cells partially resistant to immune CD8+ T cell lysis.

DNA viruses have low mutation rates. Their persistence in human populations usually depends on their ability to establish latent infection in some cells, to reactivate from latency, and then to replicate at epithelial surfaces. *Latency* is defined as a state of infection in which virus is not replicating, viral genes associated with lytic infection are not expressed, and infectious virus is not made. The complete viral genome is present and may be replicated by cellular DNA polymerase in conjunction with replication of the cell's genome. HPVs establish latent infection in basal epithelial cells. The latently infected basal cell replicates, along with the HPV episome, by using cellular DNA polymerase. Some of the progeny cells provide new latently infected basal cells, while others go on to squamous differentiation. Infected cells that differentiate to squamous cells become permissive for lytic viral infection. Herpesviruses establish latent infection in nonreplicating neural cells (HSV and VZV) or in replicating cells of hematopoietic lineages [EBV, CMV, HHV-6, HHV-7, and Kaposi's sarcoma—associated herpesvirus (KSHV, also known as HHV-8)]. In their latent stage, HPV and herpesvirus genomes are largely hidden from the normal immune response. Reactivated HPV and herpesvirus infections escape immediate and effective immune responses in highly immune hosts by inhibiting host innate immune and inflammatory responses. In addition, HPV, HSV, and VZV are somewhat protected because they replicate in the middle and upper layers of the squamous epithelium—sites not routinely visited by cells that mediate or amplify immune and inflammatory responses. HSV and CMV are also known to encode proteins that downregulate MHC class I expression and antigenic peptide presentation, enabling infected cells to escape recognition by and cytotoxic effects of CD8+ T lymphocytes.

Like other poxviruses, MCV cannot establish latent infection. This virus causes persistent infection in hypertrophic skin lesions that last for months or years. MCV encodes a chemokine homologue that probably blocks inflammatory responses, an MHC class I analogue that blocks cytotoxic T lymphocyte attack, and inhibitors of cell death that prolong infected cell viability.

Persistent Viral Infections and Cancer

Persistent viral infection is estimated to be the root cause of as many as 20% of human malignancies. Cancer is an accidental and highly unusual or long-term effect of oncogenic human virus infection. With most "oncogenic viruses," infection is a critical and ultimately determinative early step in carcinogenesis. Latent HPV infection can block cell death and cause cervical cells to proliferate. A virus-infected cell with an integrated HPV genome overexpressing E6 and E7 undergoes subsequent cellular genetic changes that enhance autonomous malignant cell growth.

Most hepatocellular carcinoma is believed to be caused by chronic inflammatory, immune, and regenerative responses to HBV or HCV infection. Epidemiologic data firmly link HBV and HCV infections to hepatocellular carcinoma. These infections elicit repetitive cycles of virus-induced liver injury followed by tissue repair and regeneration. Over decades, chronic virus infection, repetitive tissue regeneration, and acquired chromosomal changes can result in proliferative nodules. Further chromosomal mutations can lead to the degeneration of cells in a proliferating nodule into hepatocellular carcinoma. In rare instances, HBV DNA integrates into cellular DNA, promoting overexpression of a cell gene that can also contribute to oncogenesis.

Most cervical carcinoma is caused by persistent infection with "high-risk" HPV type 16 or 18. In contrast to HBV and HCV infections, which stimulate cell growth as a consequence of virus- induced cell death, HPV type 16 or 18 proteins E6 and E7 destroy p53 and pRB, respectively. Elimination of these key tumor-suppressive cell proteins increases cell growth, cell survival, and cell genome instability. However, like HBV and HCV infections, HPV infection alone is not sufficient for carcinogenesis. Cervical carcinoma is inevitably associated with persistent HPV infection and integration of the HPV genome into chromosomal DNA. Integrations that result in overexpression of E6 and E7 from HPV type 16 or 18 cause more profound changes in cell growth and survival and enable subsequent chromosomal changes that result in cervical carcinoma.

EBV is the most unusual oncogenic virus in that normal B cell infection results in latency with expression of viral proteins that can cause endless B lymphocyte growth. In almost all humans, strong CD4+ and CD8+ T cell immune responses to the antigenic EBV latent-infection nuclear proteins prevent uncontrolled B cell lymphoproliferation. However, when humans are severely immunosuppressed by transplantation-associated medication, HIV infection, or genetic immunodeficiencies, EBV-induced B cell malignancies can emerge.

EBV infection also plays a role in the long-term development of B lymphocyte and epithelial cellmalignancies. Persistent EBV infection with expression of an EBV latency-associated integral membrane protein (LMP1) in latently infected epithelial cells appears to be a critical early step in the evolution of anaplastic nasopharyngeal carcinoma, a common malignancy in populations in southern China and northern Africa. Genomic instability and chromosomal abnormalities also contribute to the development of EBV-associated nasopharyngeal carcinomas. EBV is an important cause of Hodgkin's lymphoma. High-level expression of LMP1 or LMP2 in Reed-Sternberg cells is a hallmark in up to 50% of Hodgkin's lymphoma cases. LMP1-induced nuclear factor κB (NF-κB) activity may prolong the survival of defective B cells that are normally eliminated by apoptosis, thereby allowing the acquisition of other genetic changes leading to malignant Reed-Sternberg cells.

The HTLV-I Tax and Rex proteins are critical to the initiation of cutaneous adult T cell lymphoma/leukemias that occur long after primary HTLV-I infection. Tax-induced NF-κB activation may contribute to cytokine production, infected cell survival, and eventual outgrowth of malignant cells.

Molecular data confirm the presence of KSHV DNA in all Kaposi's tumors, including those associated with HIV infection, transplantation, and familial transmission. KSHV infection is also etiologically implicated in pleural-effusion lymphomas and multicentric Castleman's disease, which are more common among HIV-infected than among HIV-uninfected people. KSHV can express a virus-encoded cyclin, an IFN regulatory factor, and a latency-associated nuclear antigen that are implicated in increased cell proliferation and survival.

Evidence supporting a causal role for virus infection in all of these malignancies includes (1) epidemiologic data, (2) the presence of viral DNA in all tumor cells, (3) the ability of the viruses to transform human cells in culture, (4) the results of in vitro cell culture—based assays that reveal transforming effects of specific viral genes on cell growth or survival, (5) pathologic data indicating the expression of transforming viral genes in premalignant or malignant cells in vivo, (6) the demonstration in animal models that these viral genes can cause malignant cell growth, and (7) the ability of virus-specific vaccines to reduce the incidence of virus-associated malignancy. Virus-related malignancies provide an opportunity to expand our understanding of the biologic mechanisms important in the development of cancer. They also offer unique opportunities to develop diagnostics, vaccines, or therapeutics that could prevent or specifically treat cancers associated with virus infection. Widespread immunization against hepatitis B has resulted in a decreased prevalence of HBV-associated hepatitis and will probably prevent most HBV-related liver cancers. An HPV vaccine can reduce rates of colonization with high-risk HPV strains and thereby decrease the risk of cervical cancer. The successful use of in vitro—expanded EBV-specific T cell populations to treat or prevent EBV-associated posttransplantation lymphoproliferative disease demonstrates the potential of immunoprevention or immunotherapy against virus-associated cancers.

Resistance to Viral Infections

Resistance to viral infections is initially provided by factors that are not virus-specific. Physical protection is afforded by the cornified layers of the skin and by mucous secretions that continuously sweep over mucosal surfaces. Once the first cell is infected, IFNs are induced and confer resistance to viral replication. Viral infection may also trigger the release of other cytokines from infected cells. These cytokines may be chemotactic to inflammatory and immune cells. Viral protein epitopes expressed on the cell surface in the context of MHC class I and II proteins can stimulate the expansion of T cell populations with receptors that can recognize the virus-encoded peptides. Cytokines and antigens released by virus-induced cell death further attract inflammatory cells, dendritic cells, granulocytes, natural killer (NK) cells, and B lymphocytes to sites of infection and to draining lymph nodes. IFNs and NK cells are particularly important in containing viral infection for the first several days. Granulocytes and macrophages are also important in the phagocytosis and degradation of viruses, especially after an initial antibody response.

By 7–10 days after infection, virus-specific antibody responses, virus-specific HLA class II—restricted CD4+ helper T lymphocyte responses, and virus-specific HLA class I—restricted CD8+ cytotoxic T lymphocyte responses develop. These responses, whose magnitude typically increases over the second and third weeks of infection, are important for rapid recovery. Also between the second and third weeks, the antibody type usually changes from IgM to IgG; IgG or IgA antibody can then be detected at infected mucosal surfaces. Antibody may directly neutralize virus by binding to its surface and preventing cell attachment or penetration. Complement can significantly enhance antibody-mediated virus neutralization. Antibody and complement can also lyse virus-infected cells that express viral membrane proteins on the cell surface. Cells infected with a replicating enveloped virus usually express the virus-envelope glycoproteins on the cell plasma membrane. Specific antibodies can bind to the glycoproteins, fix complement, and lyse the infected cell.

Antibody and CD4+/CD8+ T lymphocyte responses to virus infection can persist at high levels for several months after primary infection. Antibody-producing B lymphocytes and CD4+ or CD8+ T lymphocyte responses can persist in small numbers as memory cells and begin to proliferate rapidly in response to a second infection, providing an early barrier to reinfection with the same virus. Redevelopment of T cell immunity may take longer than secondary antibody responses, particularly when many years have elapsed between primary infection and re-exposure. However, persistent infections or frequent reactivations from latency can result in sustained high-level T cell responses. EBV and CMV typically induce high-level CD4+ and CD8+ T cell responses that are maintained for decades after primary infection.

Some viruses have genes that alter innate and acquired host defenses. Adenoviruses encode small RNAs that inhibit IFN-induced, protein kinase R (PKR)—mediated shutoff of infected-cell protein synthesis. Adenovirus E1A can also directly inhibit IFN-mediated changes in cell gene transcription. Moreover, adenovirus E3 proteins prevent tumor necrosis factor (TNF)—induced cytolysis and block HLA class I antigen synthesis by the infected cell. HSV ICP47 and CMV US11 also block class I antigen presentation. EBV encodes an interleukin (IL) 10 homologue that inhibits NK and T cell

responses. Vaccinia virus encodes a soluble receptor for IFN-α and binding proteins for IFN-γ, IL-1, IL-18, and TNF, which inhibit host innate and adaptive immune responses. Vaccinia virus also encodes a caspase inhibitor that inhibits the ability of CD8+ cytotoxic T cells to kill virus-infected cells. Some poxviruses and herpesviruses encode chemokine-binding proteins that inhibit cell inflammatory responses. The adoption of these strategies by viruses highlights the importance of the corresponding host resistance factors in containing viral infection and the importance of redundancy in host resistance.

The host inflammatory and immune responses to viral infection do not come without a price. These responses contribute to the symptoms, signs, and other pathophysiologic manifestations of viral infection. Inflammation at sites of viral infection can subvert an effective immune response and induce tissue death and dysfunction. Moreover, immune responses to viral infection could, in principle, result in immune attack upon cross-reactive epitopes on normal cells, with consequent autoimmunity.

Interferons

All human cells can synthesize IFN-α or IFN-β in response to viral infection. These IFN responses are usually induced by the presence of double-strand viral RNA, which can be made by both RNA and DNA viruses and sensed by double-strand RNA binding proteins (e.g., PKR and RIG-I) in the cell cytoplasm. IFN-γ is not closely related to IFN-α or IFN-β and is produced mainly by NK cells and by immune T lymphocytes responding to IL-12. IFN-α and -β bind to the IFN-α receptor, while IFN-γ binds to a different but related receptor. Both receptors signal through receptor-associated JAK kinases and other cytoplasmic proteins, including "STAT" proteins, which are tyrosine-phosphorylated by JAK kinases, translocate to the nucleus, and activate promoters for specific cell genes. Three types of antiviral effects are induced by IFN at the transcriptional level. The first effect is attributable to the induction of 2'-5' oligo(A) synthetases, which require double-strand RNA for their activation. Activated synthetase polymerizes oligo(A) and thereby activates RNAse L, which in turn degrades single-strand RNA. A second effect results from the induction of PKR, a serine and threonine kinase that is also activated by double-strand RNA. PKR phosphorylates and negatively regulates the translational initiation factor eIF2α, shutting down protein synthesis in the infected cell. A third effect is initiated through the induction of Mx proteins, a family of GTPases that is particularly important in inhibiting the replication of influenza virus and vesicular stomatitis virus. These IFN effects are mostly directed against the infected cell, causing virus and cell dysfunction and thereby limiting viral replication.

Diagnostic Virology

A wide variety of methods are used to diagnose viral infection. Serology and virus isolation in tissue culture remain important standards. Acute- and convalescent-phase sera with rising titers of antibody to virus-specific antigens and a shift from IgM to IgG antibodies are generally accepted as diagnostic of acute viral infection. Serologic diagnosis is based on a >4-fold rise in IgG antibody concentration when acute- and convalescent-phase serum samples are analyzed at the same time.

Immunofluorescence, hemadsorption, and hemagglutination assays for antiviral antibodies are labor-intensive and have been replaced by enzyme-linked immunosorbent assays (ELISAs), which generally use the specific viral proteins most frequently targeted by the antibody response. The proteins are purified from virus-infected cells or produced by recombinant DNA technology and are attached to a solid phase, where they can be incubated with serum, washed to eliminate nonspecific antibodies, and allowed to react with an enzyme-linked reagent to detect human IgG or IgM antibody specifically adhering to the viral antigen. The amount of antibody can then be quantitated by the intensity of a color reaction mediated by the linked enzyme. ELISAs can be sensitive and automated. Western blots can simultaneously confirm the presence of antibody to multiple specific viral proteins. The proteins are separated by size and transferred to an inert membrane, where they are incubated with serum antibodies. Western blots have an internal specificity control, since the level of reactivity for viral proteins can be compared with that for cellular proteins in the same sample. Western blots require individual evaluation and are inherently difficult to quantitate or automate.

Isolation of virus in tissue culture depends on infection and replication in susceptible cells. Growth of virus in cell cultures can frequently be identified by effects on cell morphology under light microscopy. For example, HSV produces a typical cytopathic effect in rabbit kidney cells within 3 days. Other viral cytopathic effects may not be as diagnostically distinctive. Identification usually requires confirmation by staining with virus-specific monoclonal antibodies. The efficiency and speed of virus identification can be enhanced by combining short-term culture with immune detection. In assays with "shell vials" of tissue culture cells growing on a coverslip, viral infection can be detected by staining with a monoclonal antibody to a specific viral protein expressed early in viral replication. Thus, virus-infected cells can be detected within hours or days of inoculation, whereas several rounds of infection would be required to produce visible cytopathic effects.

Isolation of virus in tissue culture also depends on the collection of specimens from appropriate sites and the rapid transport of these specimens in appropriate medium to the virology laboratory (Chap. e22). Rapid transport maintains viral viability and limits bacterial and fungal overgrowth. Enveloped viruses are generally more sensitive to freezing and thawing than nonenveloped viruses. The most appropriate site for culture depends on the pathogenesis of the virus in question. Nasopharyngeal,

tracheal, or endobronchial aspirates are most appropriate for the identification of respiratory viruses. Sputum cultures generally are less appropriate since bacterial contamination and viscosity threaten tissue-culture cell viability. Aspirates of vesicular fluid are useful for isolation of HSV and VZV. Nasopharyngeal aspirates and stool specimens may be useful when the patient has fever and a rash and an enteroviral infection is suspected. Adenoviruses can be cultured from the urine of patients with hemorrhagic cystitis. CMV can frequently be isolated from cultures of urine or buffy coat. Biopsy material can be effectively cultured when viruses infect major organs, as in HSV encephalitis or adenovirus pneumonia.

The isolation of a virus does not necessarily establish disease causality. Viruses can persistently or intermittently colonize normal human mucosal surfaces. Saliva can be positive for herpesviruses, and normal urine samples can be positive for CMV. Isolations from blood, cerebrospinal fluid (CSF), or tissue are more often diagnostic of significant viral infection.

Another method aimed at increasing the speed of viral diagnosis is direct testing for antigen or cytopathic effects. Virus-infected cells from the patient may be detected by staining with virus-specific monoclonal antibodies. For example, epithelial cells obtained by nasopharyngeal aspiration can be stained with a variety of specific monoclonal antibodies to identify the specific infecting respiratory virus.

Nucleic acid amplification techniques bring speed, sensitivity, and specificity to diagnostic virology. The ability to directly amplify minute amounts of viral nucleic acids in specimens means that detection no longer depends on viable virus and its replication. For example, amplification and detection of HSV nucleic acids in the CSF of patients with HSV encephalitis is a more sensitive detection method than culture of virus from CSF. The extreme sensitivity of these tests can be a problem, since subclinical infection or contamination can lead to false-positive results. Detection of viral nucleic acids does not necessarily indicate virus-induced disease.

Measurement of the amount of viral RNA or DNA in peripheral blood is an important means for determining whether a patient is at increased risk for virus-induced disease and for evaluating clinical responses to antiviral chemotherapy. Nucleic acid technologies for RNA quantification are routinely used in AIDS patients to evaluate responses to antiviral agents and to detect viral resistance or noncompliance with therapy. Viral-load measurements are also useful for evaluating the treatment of patients with HBV and HCV infections. Nucleic acid testing or direct staining with CMV-specific monoclonal antibodies to quantitate virus-infected cells in the peripheral blood (CMV antigenemia) is useful for identifying immunosuppressed patients who may be at risk for CMV-induced disease.

Drug Treatment for Viral Infections

Multiple steps in the life cycles of viruses can be effectively targeted by antiviral drugs (see also Chaps. 178 and 189). Nucleoside and nonnucleoside reverse transcriptase inhibitors prevent HIV provirus synthesis, while protease inhibitors block maturation of the HIV polyprotein after infection of the cell. Enfuvirtide is a small peptide derived from HIV gp41 that acts before cell infection by preventing a conformational change required for initial fusion of the virus with the cell membrane. Raltegravir is an integrase inhibitor that is approved for use with other anti-HIV drugs. Amantadine and rimantadine inhibit the influenza M2 protein, preventing release of viral RNA early during infection, whereas zanamivir and oseltamivir inhibit the influenza neuraminidase, which is necessary for the efficient release of mature virions from infected cells.

Viral genomes can evolve resistance to drugs by mutation and selection, by recombination with a drug-resistant virus, or (in the case of influenza virus and other segmented RNA viral genomes) by reassortment. The emergence of drug-resistant strains can limit the efficacy of antiviral therapy. As in antibacterial therapy, excessive and inappropriate use of antiviral therapy can select for the emergence of drug-resistant strains. HIV genotyping is a rapid method for identifying drug-resistant viruses. Resistance to reverse transcriptase or protease inhibitors has been associated with specific mutations in the reverse transcriptase or protease genes. Identification of these mutations by polymerase chain reaction amplification and nucleic acid sequencing can be clinically useful for determining which antiviral agents may still be effective. Drug resistance also can arise in herpesviruses but is a less common clinical problem.

Immunization for the Prevention of Viral Infections

Viral vaccines are among the outstanding accomplishments of medical science. Smallpox has been eradicated except as a potential weapon of biological warfare or bioterrorism (Chap. 221). Poliovirus eradication may soon follow. Measles can be contained or eliminated. Excess mortality due to influenza virus epidemics can be prevented, and the threat of influenza pandemics can be decreased by contemporary killed or live attenuated influenza vaccines. Mumps, rubella, and chickenpox are well controlled by childhood vaccination in the developed world. Reimmunization of mature adults can be used to control herpes zoster. New rotavirus vaccines can have a major impact on this leading cause of gastroenteritis and prominent cause of childhood death worldwide. Widespread HBV vaccination has dramatically lowered the frequency of acute and chronic hepatitis and is expected to lead to a dramatic decrease in the incidence of hepatocellular carcinoma. The HPV vaccine was the first

vaccine specifically licensed to prevent virus-induced cancer. Use of purified proteins, genetically engineered live-virus vaccines, and recombinant DNA—based strategies will make it possible to immunize against severe infections with other viruses. The development of effective HIV and HCV vaccines is complicated by the high mutation rate of viral RNA polymerase and reverse transcriptase, the population-based and individual divergence of HIV or HCV genomes, and repeated high-level exposure in some populations. Concerns about the use of smallpox and other viruses as weapons necessitate maintenance of immunity to agents that are not encountered naturally.

Viruses as Novel Therapeutic Tools or Agents

Viruses are being used experimentally to deliver biotherapeutic agents or novel vaccines. Foreign genes can be inserted into viral nucleic acids, and the recombinant virus vectors can be used to infect the patient or the patient's cells ex vivo. Retrovirus integration into the cell genome has been used to functionally replace the abnormal gene in T cells of patients with severe combined immunodeficiency, thereby restoring immune function. Recombinant adenovirus, AAV, and retroviruses are being explored for use in diseases due to single-gene defects, such as cystic fibrosis and hemophilia. Recombinant poxviruses, adenoviruses, and influenza viruses are also being used experimentally as vaccine vectors. Viral vectors are being tested experimentally for the expression of cytokines that can enhance immunity against tumor cells or for the expression of proteins that can increase the sensitivity of tumor cells to chemotherapy. HSV deficient for replication in resting cells is being used to selectively kill proliferating glioblastoma cells after injections into CNS tumors. For improved safety, nonreplicating viruses are frequently employed in clinical trials. Potential adverse events associated with virus-mediated gene transfer include the induction of inflammatory and antiviral immune responses. Instances of retrovirus-induced human malignances have raised concerns about the safety of retroviral gene therapy vectors.

FURTHER READINGS

Cullen Br: Viral and cellular messenger RNA targets of viral microRNAs. Nature 457:421, 2009[PMID: 19158788]

Future II Study Group: Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N Engl J Med 356:1915, 2007

Harrison Sc: Viral membrane fusion. Nat Struct Mol Biol 15:690, 2008[PMID: 18596815]

Knipe Dm et al (eds): *Fields Virology,* 5th ed. New York, Lippincott Williams & Wilkins, 2006

Malim Mh: APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos Trans R Soc Lond B Biol Sci 364:675, 2009[PMID: 19038776]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ANTIVIRAL CHEMOTHERAPY, EXCLUDING ANTIRETROVIRAL DRUGS: INTRODUCTION

The field of antiviral therapy—both the number of antiviral drugs and our understanding of their optimal use—historically has lagged behind that of antibacterial drug treatment, but significant progress has been made in recent years on new drugs for several viral infections. The development of antiviral drugs poses several challenges. Viruses replicate intracellularly and often employ host cell enzymes, macromolecules, and organelles for synthesis of viral particles. Therefore, useful antiviral compounds must discriminate between host and viral functions with a high degree of specificity; agents without such selectivity are likely to be too toxic for clinical use.

Significant progress has also been made in the development of laboratory assays to assist clinicians in the appropriate use of antiviral drugs. Phenotypic and genotypic assays for resistance to antiviral drugs are becoming more widely available, and correlations of laboratory results with clinical outcomes are being better defined. Of particular note has been the development of highly sensitive and specific methods that measure the concentration of virus in blood (*virus load*) and permit direct assessment of the antiviral effect of a given drug regimen in that host site. Virus load measurements have been useful in recognizing the risk of disease progression in patients with certain viral infections and in identifying patients for whom antiviral chemotherapy might be of greatest benefit. As with any in vitro laboratory test, results are highly dependent on (and likely to vary with) the laboratory techniques employed.

Information regarding the pharmacokinetics of some antiviral drugs, particularly in diverse clinical settings, is limited. Assays to measure the concentrations of these drugs, especially of their active moieties within cells, are primarily research procedures and are not widely available to clinicians. Thus, there are relatively few guidelines for adjusting dosages of antiviral agents to maximize antiviral activity and minimize toxicity. Consequently, clinical use of antiviral drugs must be accompanied by particular vigilance with regard to unanticipated adverse effects.

Like that of other infections, the course of viral infections is profoundly affected by an interplay of the pathogen with a complex set of host defenses. The presence or absence of preexisting immunity, the ability to mount humoral and/or cell-mediated immune responses, and the stimulation of innate immunity are important determinants of the outcome of viral infections. The state of the host's defenses needs to be considered when antiviral agents are used or evaluated.

As with any therapy, the optimal use of antiviral compounds requires a specific and timely diagnosis. For some viral infections, such as herpes zoster, the clinical manifestations are so characteristic that a diagnosis can be made on clinical grounds alone. For other viral infections, such as influenza A, epidemiologic information (e.g., the documentation of a community-wide outbreak) can be used to make a presumptive diagnosis with a high degree of accuracy. However, for most of the remaining viral infections, including herpes simplex encephalitis, cytomegaloviral infections other than retinitis, and enteroviral infections, diagnosis on clinical grounds alone cannot be accomplished with certainty. For such infections, rapid viral diagnostic techniques are of great importance. Considerable progress has also been made in recent years in the development of such tests, which are now widely available for a number of viral infections.

Despite these complexities, the efficacy of a number of antiviral compounds has been clearly established in rigorously conducted and controlled studies. As summarized in **Table 178-1**, this chapter reviews the antiviral drugs that are currently approved or are likely to be considered for approval in the near future for use against viral infections other than those caused by HIV. **Antiretroviral drugs are reviewed in Chap. 189.**

Table 178-1 Antiviral Chemotherapy and Chemoprophylaxis

Infection	Drug	Route	Dosage	Comment
Influenza A and B: Treatment	Oseltamivir	Oral	Adults: 75 mg bid x 5 d Children 1–12 years: 30–75 mg bid, depending	When started within 2 days of onset in uncomplicated disease, zanamivir and oseltamivir reduce symptom duration by 1.0–1.5 and 1.3 d, respectively. Their effectiveness in prevention or treatment of complications is unclear.

			on weight,a x 5 d	although some analyses suggest that oseltamivir may reduce the frequency of respiratory tract complications and hospitalizations. Oseltamivir's side effects of nausea and vomiting can be reduced in frequency by drug administration with food. Zanamivir may exacerbate bronchospasm in patients with asthma. Amantadine and rimantadine are not recommended for routine use unless antiviral susceptibilities are known because of widespread resistance in A/H3N2 viruses since 2005–2006 and in pandemic A/H1N1 viruses in 2009–2010. Their efficacy in treatment of uncomplicated disease caused by sensitive viruses has been similar to that of neuraminidase inhibitors.
	Zanamivir	Inhaled orally	Adults and children ≥7 years: 10 mg bid x 5 d	
	Amantadine[b]	Oral	Adults: 100 mg qd or bid x 5–7 d Children 1–9 years: 5 mg/kg per day (maximum, 150 mg/d) x 5–7 d	
	Rimantadine[b]	Oral	100 mg qd or bid x 5–7 d in adults	
Influenza A and B: Prophylaxis	Oseltamivir	Oral	Adults: 75 mg/d Children ≥1 year: 30–75 mg/d, depending on weight[a]	Prophylaxis must be continued for the duration of exposure and can be administered simultaneously with inactivated vaccine. Unless the sensitivity of isolates is known, neither amantadine nor rimantadine is currently recommended for prophylaxis or therapy.
	Zanamivir	Inhaled orally	Adults and children ≥5 years: 10 mg/d	
	Amantadine[b] or rimantadine[b]	Oral	Adults: 200 mg/d Children 1–9 years: 5 mg/kg per day (maximum, 150 mg/d)	
RSV infection	Ribavirin	Small-particle aerosol	Administered 12–18 h/dfrom reservoir containing 20 mg/mL x 3–6 d	Use of ribavirin is to be "considered" for treatment of infants and young children hospitalized with RSV pneumonia and bronchiolitis, according to the American Academy of Pediatrics.
CMV disease	Ganciclovir	IV	5 mg/kg bid x 14–21 d; then 5 mg/kg per day as maintenance dose	Ganciclovir, valganciclovir, foscarnet, and cidofovir are approved for treatment of CMV retinitis in patients with AIDS. They are also used for colitis, pneumonia, or "wasting" syndrome associated with CMV and for prevention of CMV disease in transplant recipients.
	Valganciclovir	Oral	900 mg bid x 21 d; then 900 mg/d as maintenance dose	Valganciclovir has largely supplanted oral ganciclovir and is frequently used in place of IV ganciclovir.
	Foscarnet	IV	60 mg/kg q8h x 14–21 d; then 90–120 mg/kg	Foscarnet is not myelosuppressive and is active against acyclovir- and ganciclovir-resistant

			per day as maintenance dose	herpesviruses.
	Cidofovir	IV	5 mg/kg once weekly x 2 weeks, then once every other week; given with probenecid and hydration	
	Fomivirsen	Intravitreal	330 mg on days 1 and 15 followed by 330 mg monthly as maintenance	Fomivirsen has reduced the rate of progression of CMV retinitis in patients in whom other regimens have failed or have not been well tolerated. The major form of toxicity is ocular inflammation.
Varicella: Immunocompetent host	Acyclovir	Oral	20 mg/kg (maximum, 800 mg) 4 or 5 times daily x 5 d	Treatment confers modest clinical benefit when administered within 24 h of rash onset.
	Valacyclovir	Oral	Children 2–18 years:20 mg/kg tid, not to exceed1 g tid, x5 d	
Varicella: Immunocompromised host	Acyclovir	IV	10 mg/kg q8h x 7 d	A change to oral valacyclovir can be considered once fever has subsided if there is no evidence of visceral involvement.
Herpes simplex encephalitis	Acyclovir	IV	10 mg/kg q8h x 14–21 d	Results are optimal when therapy is initiated early. Some authorities recommend treatment for 21 d to prevent relapses.
Neonatal herpes simplex	Acyclovir	IV	20 mg/kg q8h x 14–21 d	Serious morbidity is common despite therapy. Prolonged oral administration after initial IV therapy has been suggested because of long-term sequelae associated with cutaneous recurrences of HSV infection.
Genital herpes simplex: Primary (treatment)	Acyclovir	IV	5 mg/kg q8h x 5–10 d	The IV route is preferred for infections severe enough to warrant hospitalization or with neurologic complications.
		Oral	400 mg tid or 200 mg 5 times daily x 7–10 d or 800 mg tid x 2 d	The oral route is preferred for patients whose condition does not warrant hospitalization. Adequate hydration must be maintained.
		Topical	5% ointment; 4–6 applications daily x 7–10 d	Topical use—largely supplemented by oral therapy—may obviate systemic administration to pregnant women. Systemic symptoms and untreated areas are not affected.
	Valacyclovir	Oral	1 g bid x 7–10 d	Valacyclovir appears to be as effective as acyclovir but can be administered less frequently.
	Famciclovir	Oral	250 mg tid x	Famciclovir appears to be

			7–10 dc	similar in effectiveness to acyclovir.
Genital herpes simplex: Recurrent (treatment)	Acyclovir	Oral	400 mg tid or 800 mg bid x 5 d	The clinical effect is modest and is enhanced if therapy is initiated early. Treatment does not affect recurrence rates.
	Famciclovir	Oral	125 mg bid x 5 d or 1000 mg bid x 1 d	
	Valacyclovir	Oral	500 mg bid x 3 d or 1 g once a day x 5 d	
Genital herpes simplex: Recurrent (suppression)	Acyclovir	Oral	400 mg bid	Suppressive therapy is recommended only for patients with at least 6–10 recurrences per year. "Breakthrough" occasionally takes place, and asymptomatic shedding of virus occurs. The need for suppressive therapy should be reevaluated after 1 year. Suppression with valacyclovir reduces transmission of genital HSV among discordant couples.
	Valacyclovir	Oral	500–1000 mg/d	
	Famciclovir	Oral	250 mg bid	
Mucocutaneous herpes simplex in immunocompromised host: Treatment	Acyclovir	IV	5 mg/kg q8h x 7–14 d	The choice of the IV or oral route and the duration of therapy depend on the severity of infection and the patient's ability to take oral medication. Oral or IV treatment has supplanted topical therapy except for small, easily accessible lesions. Foscarnet is used for acyclovir-resistant viruses.
		Oral	400 mg 5 times daily x 10–14 d	
		Topical	5% ointment; 4–6 applications daily x 7 d or until healed	
	Valacyclovir	Oral	1 g tid x 7–10 dc	
	Famciclovir	Oral	500 mg bid x 7–10 dd	
Mucocutaneous herpes simplex in immunocompromised host: Prevention of recurrence during intense immunosuppression	Acyclovir	Oral	400 mg 2–5 times daily or 800 mg bid	Treatment is administered during periods when intense immunosuppression is expected—e.g., during antitumor chemotherapy or after transplantation—and is usually continued for 2–3 months.
		IV	5 mg/kg q12h	
	Valacyclovir	Oral	500 mg to 1 g bid or tid	
	Famciclovir	Oral	500 mg bid[c]	
Herpes simplex orolabialis (recurrent)[e]	Penciclovir	Topical	1.0% cream applied q2h during waking hours x 4 d	Treatment shortens healing time and symptom duration by 0.5–1.0 d (compared with placebo).
	Valacyclovir	Oral	2 g q12h x 1 d	Therapy begun at the earliest symptom reduces disease duration by 1 d.
	Famciclovirc	Oral	1500 mg once or 750 bid x 1 d	Therapy begun within 1 h of prodrome decreased time to healing by 1.8–2.2 d.
	Docosanolf	Topical	10% cream 5 times daily until healed	Application at initial symptoms reduces healing time by 1 d.
Herpes simplex keratitis	Trifluridine	Topical	1 drop of 1% ophthalmic solution q2h while awake	Therapy should be undertaken in consultation with an ophthalmologist.

			(maximum, 9 drops daily)	
	Vidarabine	Topical	0.5-in. ribbon of 3% ophthalmic ointment 5 times daily	
Herpes zoster: Immunocompetent host	Valacyclovir	Oral	1 g tid x 7 d	Valacyclovir may be more effective than acyclovir for pain relief; otherwise, it has a similar effect on cutaneous lesions and should be given within 72 h of rash onset.
	Famciclovir	Oral	500 mg q8h x 7 d	The duration of postherpetic neuralgia is shorter than with placebo. Famciclovir showed overall efficacy similar to that of acyclovir in a comparative trial. It should be given ≤72 h after rash onset.
	Acyclovir	Oral	800 mg 5 times daily x 7–10 d	Acyclovir causes faster resolution of skin lesions than placebo and provides some relief of acute symptoms if given within 72 h of rash onset. Combined with tapering doses of prednisone, acyclovir improves quality-of-life outcomes.
Herpes zoster: Immunocompromised host	Acyclovir	IV	10 mg/kg q8h x 7 d	Effectiveness in localized zoster is most marked when treatment is given early. Foscarnet may be used for acyclovir-resistant VZV infections.
		Oral	800 mg 5 times daily x 7 d	
	Famciclovir	Oral	500 mg tid x 10 d[c]	
Herpes zoster ophthalmicus	Acyclovir	Oral	600–800 mg 5 times daily x 10 d	Treatment reduces ocular complications, including ocular keratitis and uveitis.
	Valacyclovir	Oral	1 g tid x 7 d	
	Famciclovir	Oral	500 mg tid x 7 d	
Condyloma acuminatum	IFN-α2b	Intralesional	1 million units per wart (maximum of 5) thrice weekly x 3 weeks	Intralesional treatment frequently results in regression of warts, but lesions often recur. Parenteral administration may be useful if lesions are numerous.
	IFN-αn3	Intralesional	250,000 units per wart (maximum of 10) twice weekly x up to 8 weeks	
Chronic hepatitis B	IFN-α2b	SC	5 million units daily or 10 million units thrice weekly x 16–24 weeks	HBeAg and DNA are eliminated in 33–37% of cases. Histopathologic improvement is also seen.
	Pegylated IFN-α2a	SC	180 μg weekly x 48 weeks	ALT levels return to normal in 39% of patients, and histologic improvement occurs in 38%.
	Lamivudine	Oral	100 mg/d x 12–18	Lamivudine monotherapy is well tolerated and effective in

			months; 150 mg bid as part of therapy for HIV infection	reduction of HBV DNA levels, normalization of ALT levels, and improvement in histopathology. However, resistance develops in 24% of recipients when lamivudine is used as monotherapy for 1 year.
	Adefovir dipivoxil	Oral	10 mg/d x 48 weeks	A return of ALT levels to normal is documented in 48–72% of recipients and improved liver histopathology in 53–64%. Adefovir is effective in lamivudine-resistant hepatitis B. Renal function should be monitored.
	Entecavir	Oral	0.5 mg/d x 48 weeks (1 mg/d if HBV is resistant to lamivudine)	Normalization of ALT is seen in 68–78% of recipients and loss of HBeAg in 21%. Entecavir is active against lamivudine-resistant HBV.
	Telbivudine	Oral	600 mg/d x 52 weeks	HBV DNA is reduced by >5 log10 copies/mL along with normalization of ALT levels in 74–77% of patients and improved histopathology in 65–67%. Resistance develops in 9–22% of patients after 2 years of therapy. Elevated CPK levels and myopathy may occur.
	Tenofovir	Oral	300 mg/d x 48 weeks	ALT levels return to normal in 68–76% of patients, and liver histopathology improves in 72–74%. Resistance is uncommon with up to 2 years of therapy.
Chronic hepatitis C	IFN-α2a or IFN-α2b	SC	3 million units thrice weekly x 12–24 months	SVRs are noted in 20–30% of patients. Normalization of ALT levels and improvements in liver histopathology are also seen.
	IFN-α2b/ribavirin	SC (IFN)/oral (ribavirin)	3 million units thrice weekly (IFN)/1000–1200 mg daily (ribavirin) x 6–12 months	Combination therapy results in SVR in up to 40–50% of recipients.
	Pegylated IFN-α2b	SC	1.5 μg weekly x 48 weeks	The slower clearance of pegylated IFNs than of standard IFNs permits once-weekly administration. Pegylated formulations appear to be superior to standard IFNs in efficacy, both as monotherapy and in combination with ribavirin, and have largely supplanted standard IFNs in treatment of hepatitis C. SVRs were seen in 42–51% of patients infected with genotype 1 and in 76–82% of those infected with genotype 2 or 3.
	Pegylated IFN-α2a	SC	180 αg weekly x 48 weeks	
	Pegylated IFN-α2b/ribavirin	SC (IFN)/oral (ribavirin)	1.5 αg/kg weekly (IFN)/800–1400 mg daily (ribavirin) x 24–48 weeks	
	Pegylated IFN-α2a/ribavirin	SC (IFN)/oral (ribavirin)	180 αg weekly (IFN)/800–1200 mg daily (ribavirin) x 24–48 weeks	
	IFN-alfacon	SC	9–15 αg	Doses of 9 and 15 μg are

			thrice weekly x 6–12 months	equivalent to IFN-α2a and IFN-α2b doses of 3 million and 5 million units, respectively.
Chronic hepatitis D	IFN-α2a or IFN-α2b	SC	9 million units thrice weekly x 12 months	The overall efficacy and the optimal regimen and duration of therapy have not been established. Response rates have varied among studies.
	Pegylated IFN-α2b	SC	1.5 μg weekly x 48 weeks	
	Pegylated IFN-α2a	SC	180 μg weekly x 48 weeks	

[a]For detailed weight recommendations and for children ≥1 year of age, see *www.cdc.gov/flu/professionals/antivirals/summary-clinicians.htm*.

[b]Amantadine and rimantadine are not recommended for routine use because of widespread resistance in A/H3N2 and pandemic A/H1N1 viruses in 2009–2010. Their use may be considered if sensitivities become reestablished.

[c]Not approved for this indication by the U.S. Food and Drug Administration (FDA).

[d]Approved by the FDA for treatment of HIV-infected individuals.

[e]Acyclovir suspension (15 mg/kg PO to a maximum of 200 mg per dose) given for 7 d has been reported to be effective in treatment of primary herpetic gingivostomatitis in children.

[f]Active ingredient: benzyl alcohol. Available without prescription.

Abbreviations: ALT, alanine aminotransferase; CMV, cytomegalovirus; CPK, creatine phosphokinase; HBeAg, hepatitis B e antigen; HBV, hepatitis B virus; HSV, herpes simplex virus; IFN, interferon; RSV, respiratory syncytial virus; SVR, sustained virologic response; UV, ultraviolet; VZV, varicella-zoster virus.

ANTIVIRAL DRUGS ACTIVE AGAINST RESPIRATORY INFECTIONS

(Also See Chap. 187)

Zanamivir, Oseltamivir, and Peramivir

Zanamivir and oseltamivir are inhibitors of the influenza viral neuraminidaseenzyme, which is essential for release of the virus from infected cells and for its subsequent spread throughout the respiratory tract of the infected host. The enzyme cleaves terminal sialic acid residues and thus destroys the cellular receptors to which the viral hemagglutinin attaches. Zanamivir and oseltamivir are sialic acid transition-state analogues and are highly active and specific inhibitors of the neuraminidases of both influenza A and B viruses. The antineuraminidase activity of the two drugs is similar, although zanamivir has somewhat greater in vitro activity against influenza B. Both zanamivir and oseltamivir act through competitive and reversible inhibition of the active site of influenza A and B viral neuraminidases and have relatively little effect on mammalian cell enzymes.

Oseltamivir phosphate is an ethyl ester prodrug that is converted to oseltamivir carboxylate by esterases in the liver. Orally administered oseltamivir has a bioavailability of >60% and a plasma half-life of 7–9 h. The drug is excreted unmetabolized, primarily by the kidneys. Zanamivir has low oral bioavailability and is administered orally via a hand-held inhaler. By this route, ~15% of the dose is deposited in the lower respiratory tract, and low plasma levels of the drug are detected.

Orally inhaled zanamivir is generally well tolerated, although exacerbations of asthma may occur. The toxicities most frequently encountered with orally administered oseltamivir are nausea, gastrointestinal discomfort, and (less commonly) vomiting. Gastrointestinal discomfort is usually transient and is less likely if the drug is administered with food. Neuropsychiatric events (delirium, self-injury) have been reported in children who have been taking oseltamivir, primarily in Japan. An IV formulation of zanamivir is under development and is available from GlaxoSmith Kline as part of clinical trials.

Inhaled zanamivir and orally administered oseltamivir have been effective in the treatment of naturally occurring, uncomplicated influenza A or B in otherwise healthy adults. In placebo-controlled studies, illness has been shortened by 1.0–1.5 days of therapy with either of these drugs when treatment is administered within 2 days of onset. Pooled analyses of clinical studies of oseltamivir suggest that treatment may reduce the likelihood of hospitalizations and of certain respiratory tract complications associated with influenza (Chap. 187). Once-daily inhaled zanamivir or once-daily orally administered oseltamivir can provide prophylaxis against laboratory-documented influenza A– and influenza B —associated illness.

Resistance to the neuraminidase inhibitors may develop by changes in the viral neuraminidase

enzyme, by changes in the hemagglutinin that make it more resistant to the actions of the neuraminidase, or by both mechanisms. Isolates that are resistant to oseltamivir—most commonly through the H275Y mutation, which leads to a change from histidine to tyrosine at that residue in the neuraminidase—remain sensitive to zanamivir. Certain mutations impart resistance to both oseltamivir and zanamivir (e.g., I223R, which leads to a change from isoleucine to arginine). Since the mechanisms of action of the neuraminidase inhibitors differ from those of the adamantanes (see below), zanamivir and oseltamivir are active against strains of influenza A virus that are resistant to amantadine and rimantadine.

Appropriate use of antiviral agents against influenza viruses depends on a knowledge of the resistance patterns of circulating viruses. For example, in the 2008–2009 influenza season, the circulating influenza A/H3N2 viruses were sensitive to both oseltamivir and zanamivir, whereas the seasonal A/H1N1 viruses, although also sensitive to zanamivir, were resistant to oseltamivir. Moreover, the pandemic A/H1N1 viruses that circulated in 2009–2010 remained sensitive to zanamivir and oseltamivir, with a few exceptions in the latter case. Up-to-date information on resistance patterns to antiviral drugs is available from the Centers for Disease Control and Prevention (CDC) at *www.cdc.gov/flu*.

Zanamivir and oseltamivir have been approved by the U.S. Food and Drug Administration (FDA) for treatment of influenza in adults and in children (those ≥7 years old for zanamivir and those ≥1 year old for oseltamivir) who have been symptomatic for ≤2 days. Oseltamivir is approved for prophylaxis of influenza in individuals ≥1 year of age and zanamivir for those ≥5 years of age (Table 178-1). Guidelines for use of oseltamivir in children <1 year of age can be accessed through the CDC website, as noted in the footnote to Table 178-1.

Peramivir, an investigational neuraminidase inhibitor that can be administered intravenously to patients for whom such an intervention is considered necessary, is available as part of clinical trials through BioCryst Pharmaceuticals. Oseltamivir-resistant viruses generally exhibit reduced sensitivity to peramivir.

Amantadine and Rimantadine

Amantadine and the closely related compound rimantadine are primary symmetric amines that have antiviral activity limited to influenza A viruses. Amantadine and rimantadine have been shown to be efficacious in the prophylaxis and treatment of influenza A infections in humans for >45 years. High frequencies of resistance to these drugs were noted among influenza A/H3N2 viruses in the 2005–2006 influenza season and continued to be seen in 2008–2009. The pandemic A/H1N1 viruses that circulated in 2009–2010 were also resistant to amantadine and rimantadine. Therefore, these agents are no longer recommended for use unless the sensitivity of the individual influenza A isolate is known, in which case their use may be considered. Amantadine and rimantadine act through inhibition of the ion channel function of the influenza A M2 matrix protein, on which uncoating of the virus depends. A substitution of a single amino acid at critical sites in the M2 protein can result in a virus that is resistant to amantadine and rimantadine.

Amantadine and rimantadine have been shown to be effective in the prophylaxis of influenza A in large-scale studies of young adults and in less extensive studies of children and elderly persons. In such studies, efficacy rates of 55–80% in the prevention of influenza-like illness were noted, and even higher rates were reported when virus-specific attack rates were calculated. Amantadine and rimantadine have also been found to be effective in the treatment of influenza A infection in studies involving predominantly young adults and, to a lesser extent, children. Administration of these compounds within 24–72 h after the onset of illness has resulted in a reduction of the duration of signs and symptoms by ~50% compared to that in placebo recipients. The effect on signs and symptoms of illness is superior to that of commonly used antipyretic-analgesic agents. Only anecdotal reports are available concerning the efficacy of amantadine or rimantadine in the prevention or treatment of complications of influenza (e.g., pneumonia).

Amantadine and rimantadine are available only in oral formulations and are ordinarily administered to adults once or twice daily, with a dosage of 100–200 mg/d. Despite their structural similarities, the two compounds have different pharmacokinetics. Amantadine is not metabolized and is excreted almost entirely by the kidneys, with a half-life of 12–17 h and peak plasma concentrations of 0.4 μg/mL. In contrast, rimantadine is extensively metabolized to hydroxylated derivatives and has a half-life of 30 h. Only 30–40% of an orally administered dose of rimantadine is recovered in the urine. The peak plasma levels of rimantadine are approximately half those of amantadine, but rimantadine is concentrated in respiratory secretions to a greater extent than amantadine. For prophylaxis, the compounds must be administered daily for the period at risk (i.e., duration of the exposure). For therapy, amantadine or rimantadine is generally administered for 5–7 days.

Although these compounds are generally well tolerated, 5–10% of amantadine recipients experience mild central nervous system side effects consisting primarily of dizziness, anxiety, insomnia, and difficulty in concentrating. These effects are rapidly reversible upon cessation of the drug's administration. At a dose of 200 mg/d, rimantadine is better tolerated than amantadine; in a large-scale study of young adults, adverse effects were no more frequent among rimantadine recipients than

among placebo recipients. Seizures and worsening of congestive heart failure have also been reported in patients treated with amantadine, although a causal relationship has not been established. The dosage of amantadine should be reduced to 100 mg/d in patients with renal insufficiency [i.e., a creatinine clearance rate (Cr_{Cl}) of <50 mL/min] and in the elderly. A rimantadine dose of 100 mg/d should be used for patients with a Cr_{Cl} of <10 mL/min and for the elderly.

Ribavirin

Ribavirin is a synthetic nucleoside analogue that inhibits a wide range of RNA and DNA viruses. The mechanism of action of ribavirin is not completely defined and may be different for different groups of viruses. Ribavirin-5'-monophosphate blocks the conversion of inosine-5'-monophosphate to xanthosine –5'-monophosphate and interferes with the synthesis of guanine nucleotides as well as that of both RNA and DNA. Ribavirin-5'-monophosphate also inhibits capping of virus-specific messenger RNA in certain viral systems.

Ribavirin administered as a small-particle aerosol to young children hospitalized with RSV infection has been clinically beneficial and has improved oxygenation in some studies (7 of 11). Although ribavirin has been approved for treatment of infants hospitalized with respiratory syncytial virus (RSV) infection, the American Academy of Pediatrics has recommended that its use be considered on an individual basis rather than routinely in that setting. Aerosolized ribavirin has also been administered to older children and adults (including immunosuppressed patients) with severe RSV and parainfluenza virus infections and to older children and adults with influenza A or B infection, but the benefit of this treatment, if any, is unclear. In RSV infections in immunosuppressed patients, ribavirin is often given in combination with anti-RSV immunoglobulins.

Orally administered ribavirin has not been effective in the treatment of influenza A virus infections. IV or oral ribavirin has reduced mortality rates among patients with Lassa fever; it has been particularly effective in this regard when given within the first 6 days of illness. IV ribavirin has been reported to be of clinical benefit in the treatment of hemorrhagic fever with renal syndrome caused by Hantaan virus and as therapy for Argentinean hemorrhagic fever. Oral ribavirin has also been recommended for the treatment and prophylaxis of Congo-Crimean hemorrhagic fever. An open-label trial suggested that oral ribavirin may be beneficial in the treatment of Nipah virus encephalitis. Use of IV ribavirin in patients with hantavirus pulmonary syndrome in the United States has not been associated with clear-cut benefits. Oral administration of ribavirin reduces serum aminotransferase levels in patients with chronic hepatitis C virus (HCV) infection; since it appears not to reduce serum HCV RNA levels, the mechanism of this effect is unclear. The drug provides added benefit when given by mouth in doses of 800–1200 mg/d in combination with interferon (IFN) ∝2b or ∝2a (see below), and the ribavirin/IFN combination has been approved for the treatment of patients with chronic HCV infection. Large oral doses of ribavirin (800–1000 mg/d) have been associated with reversible hematopoietic toxicity. This effect has not been observed with aerosolized ribavirin, apparently because little drug is absorbed systemically. Aerosolized administration of ribavirin is generally well tolerated but occasionally is associated with bronchospasm, rash, or conjunctival irritation. It should be administered under close supervision—particularly in the setting of mechanical ventilation, where precipitation of the drug is possible. Health care workers exposed to the drug have experienced minor toxicity, including eye and respiratory tract irritation. Because ribavirin is mutagenic, teratogenic, and embryotoxic, its use is generally contraindicated in pregnancy. Its administration as an aerosol poses a risk to pregnant health care workers. Because clearance of ribavirin is primarily renal, dose reduction is required in the setting of significant renal dysfunction.

ANTIVIRAL DRUGS ACTIVE AGAINST HERPESVIRUS INFECTIONS

Acyclovir and Valacyclovir

Acyclovir is a highly potent and selective inhibitor of the replication of certain herpesviruses, including herpes simplex virus (HSV) types 1 and 2, varicella-zoster virus (VZV), and Epstein-Barr virus (EBV). It is relatively ineffective in the treatment of human cytomegalovirus (CMV) infections; however, some studies have indicated effectiveness in the prevention of CMV-associated disease in immunosuppressed patients. Valacyclovir, the l-valyl ester of acyclovir, is converted almost entirely to acyclovir by intestinal and hepatic hydrolysis after oral administration. Valacyclovir has pharmacokinetic advantages over orally administered acyclovir: it exhibits significantly greater oral bioavailability, results in higher blood levels, and can be given less frequently than acyclovir (two or three rather than five times daily).

The high degree of selectivity of acyclovir is related to its mechanism of action, which requires that the compound first be phosphorylated to acyclovir monophosphate. This phosphorylation occurs efficiently in herpesvirus-infected cells by means of a virus-coded thymidine kinase. In uninfected mammalian cells, little phosphorylation of acyclovir occurs, and the drug is therefore concentrated in herpesvirus-infected cells. Acyclovir monophosphate is subsequently converted by host cell kinases to a triphosphate that is a potent inhibitor of virus-induced DNA polymerase but has relatively little effect on host cell DNA polymerase. Acyclovir triphosphate can also be incorporated into viral DNA, with early chain termination.

Acyclovir is available in IV, oral, and topical forms, while valacyclovir is available in an oral formulation. IV acyclovir is effective in the treatment of mucocutaneous HSV infections in immunocompromised hosts, in whom it reduces time to healing, duration of pain, and virus shedding. When administered prophylactically during periods of intense immunosuppression (e.g., related to chemotherapy for leukemia or transplantation) and before the development of lesions, IV acyclovir reduces the frequency of HSV-associated disease. After prophylaxis is discontinued, HSV lesions recur. IV acyclovir is also effective in the treatment of HSV encephalitis.

Because VZV is generally less sensitive to acyclovir than is HSV, higher doses of acyclovir must be used to treat VZV infections. In immunocompromised patients with herpes zoster, IV acyclovir reduces the frequency of cutaneous dissemination and visceral complications and—in one comparative trial—was more effective than vidarabine. Acyclovir, administered at oral doses of 800 mg five times a day, had a modest beneficial effect on localized herpes zoster lesions in both immunocompromised and immunocompetent patients. Combination of acyclovir with a tapering regimen of prednisone appeared to be more effective than acyclovir alone in terms of quality-of-life outcomes in immunocompetent patients over age 50 with herpes zoster. A comparative study of acyclovir (800 mg PO five times daily) and valacyclovir (1 g PO three times daily) in immunocompetent patients with herpes zoster indicated that the latter drug may be more effective in eliciting the resolution of zoster-associated pain. Orally administered acyclovir (600 mg five times a day) reduced complications of herpes zoster ophthalmicus in a placebo-controlled trial.

In chickenpox, a modest overall clinical benefit is attained when oral acyclovir therapy is begun within 24 h of the onset of rash in otherwise healthy children (20 mg/kg, up to a maximum of 800 mg, four times a day) or adults (800 mg five times a day). IV acyclovir has also been reported to be effective in the treatment of immunocompromised children with chickenpox.

The most widespread use of acyclovir is in the treatment of genital HSV infections. IV or oral acyclovir or oral valacyclovir has shortened the duration of symptoms, reduced virus shedding, and accelerated healing when employed for the treatment of primary genital HSV infections. Oral acyclovir and valacyclovir have also had a modest effect in treatment of recurrent genital HSV infections. However, the failure of treatment of either primary or recurrent disease to reduce the frequency of subsequent recurrences has indicated that acyclovir is ineffective in eliminating latent infection. Chronic oral administration of acyclovir for ≥1–6 years or of valacyclovir for ≥1 year has reduced the frequency of recurrences markedly during therapy; once the drug is discontinued, lesions recur. In one study, suppressive therapy with valacyclovir (500 mg once daily for 8 months) reduced transmission of HSV-2 genital infections among discordant couples by 50%. A modest effect on herpes labialis (i.e., a reduction of disease duration by 1 day) was seen when valacyclovir was administered upon detection of the first symptom of a lesion at a dose of 2 g every 12 h for 1 day. In AIDS patients, chronic or intermittent administration of acyclovir has been associated with the development of HSV and VZV strains resistant to the action of the drug and with clinical failures. The most common mechanism of resistance is a deficiency of the virus-induced thymidine kinase. Patients with HSV or VZV infections resistant to acyclovir have frequently responded to foscarnet.

With the availability of the oral and IV forms, there are few indications for topical acyclovir, although treatment with this formulation has been modestly beneficial in primary genital HSV infections and in mucocutaneous HSV infections in immunocompromised hosts.

Overall, acyclovir is remarkably well tolerated and is generally free of toxicity. The most frequently encountered form of toxicity is renal dysfunction because of drug crystallization, particularly after rapid IV administration or with inadequate hydration. Central nervous system changes, including lethargy and tremors, are occasionally reported, primarily in immunosuppressed patients. However, whether these changes are related to acyclovir, to concurrent administration of other therapy, or to underlying infection remains unclear. Acyclovir is excreted primarily unmetabolized by the kidneys via both glomerular filtration and tubular secretion. Approximately 15% of a dose of acyclovir is metabolized to 9-[(carboxymethoxy)methyl]guanine or other minor metabolites. Reduction in dosage is indicated in patients with a Cr_{Cl} of <50 mL/min. The half-life of acyclovir is ~3 h in normal adults, and the peak plasma concentration after a 1-h infusion of a dose of 5 mg/kg is 9.8 μg/mL. Approximately 22% of an orally administered acyclovir dose is absorbed, and peak plasma concentrations of 0.3–0.9 μg/mL are attained after administration of a 200-mg dose. Acyclovir penetrates relatively well into the cerebrospinal fluid (CSF), with concentrations approaching half of those found in plasma.

Acyclovir causes chromosomal breakage at high doses, but its administration to pregnant women has not been associated with fetal abnormalities. Nonetheless, the potential risks and benefits of acyclovir should be carefully assessed before the drug is used in pregnancy.

Valacyclovir exhibits three to five times greater bioavailability than acyclovir. The concentration-time curve for valacyclovir, given as 1 g PO three times daily, is similar to that for acyclovir, given as 5 mg/kg IV every 8 h. The safety profiles of valacyclovir and acyclovir are similar, although thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome has been reported in immunocompromised patients who have received high doses of valacyclovir (8 g/d). Valacyclovir is approved for the treatment of herpes zoster, of initial and recurrent episodes of genital HSV infection, and of herpes

labialis in immunocompetent adults as well as for suppressive treatment of genital herpes. Although it has not been extensively studied in other clinical settings involving HSV or VZV infections, many consultants use valacyclovir rather than oral acyclovir in settings where only the latter has been approved because of valacyclovir's superior pharmacokinetics and more convenient dosing schedule.

Cidofovir

Cidofovir is a phosphonate nucleotide analogue of cytosine. Its major use is in CMV infections, particularly retinitis, but it is active against a broad range of herpesviruses, including HSV, human herpesvirus (HHV) type 6, HHV-8, and certain other DNA viruses such as polyomaviruses, papillomaviruses, adenoviruses, and poxviruses, including variola (smallpox) and vaccinia. Cidofovir does not require initial phosphorylation by virus-induced kinases; the drug is phosphorylated by host cell enzymes to cidofovir diphosphate, which is a competitive inhibitor of viral DNA polymerases and, to a lesser extent, of host cell DNA polymerases. Incorporation of cidofovir diphosphate slows or terminates nascent DNA chain elongation. Cidofovir is active against HSV isolates that are resistant to acyclovir because of absent or altered thymidine kinase and against CMV isolates that are resistant to ganciclovir because of UL97 phosphotransferase mutations. CMV isolates resistant to ganciclovir on the basis of UL54 mutations are usually resistant to cidofovir as well. Cidofovir is usually active against foscarnet-resistant CMV, although cross-resistance to foscarnet has also been described.

Cidofovir has poor oral availability and is administered intravenously. It is excreted primarily by the kidney and has a plasma half-life of 2.6 h. Cidofovir diphosphate's intracellular half-life of >48 h is the basis for the recommended dosing regimen of 5 mg/kg once a week for the initial 2 weeks and then 5 mg/kg every other week. The major toxic effect of cidofovir is proximal renal tubular injury, as manifested by elevated serum creatinine levels and proteinuria. The risk of nephrotoxicity can be reduced by vigorous saline hydration and by concomitant oral administration of probenecid. Neutropenia, rashes, and gastrointestinal tolerance may also occur.

IV cidofovir has been approved for the treatment of CMV retinitis in AIDS patients who are intolerant of ganciclovir or foscarnet or in whom those drugs have failed. In a controlled study, a maintenance dosage of 5 mg/kg per week administered to AIDS patients reduced the progression of CMV retinitis from that seen at 3 mg/kg. Intravitreal cidofovir has been used to treat CMV retinitis but has been associated with significant toxicity. IV cidofovir has been reported anecdotally to be effective for treatment of acyclovir-resistant mucocutaneous HSV infections. Likewise, topically administered cidofovir is reportedly beneficial against mucocutaneous HSV infections in HIV-infected patients. Anecdotal use of IV cidofovir has been described in disseminated adenoviral infections in immunosuppressed patients and in genitourinary infections with BK virus in renal transplant recipients; however, its efficacy, if any, in these circumstances is not established.

Fomivirsen

Fomivirsen is the first antisense oligonucleotide approved by the FDA for therapy in humans. This phosphorothioate oligonucleotide, 21 nucleotides in length, inhibits CMV replication through interaction with CMV messenger RNA. Fomivirsen is complementary to messenger transcripts of the major immediate early region 2 (IE2) of CMV, which codes for proteins regulating viral gene expression. In addition to its antisense mechanism of action, fomivirsen may exert activity against CMV through inhibition of viral adsorption to cells as well as direct inhibition of viral replication. Because of its different mechanism of action, fomivirsen is active against CMV isolates that are resistant to nucleoside or nucleotide analogues, such as ganciclovir, foscarnet, or cidofovir.

Fomivirsen has been approved for intravitreal administration in the treatment of CMV retinitis in AIDS patients who have failed to respond to other treatments or cannot tolerate them. Injections of 330 mg for two doses 2 weeks apart, followed by maintenance doses of 330 mg monthly, significantly reduce the rate of progression of CMV retinitis. The major toxicity is ocular inflammation, including vitritis and iritis, which usually responds to topically administered glucocorticoids.

Ganciclovir and Valganciclovir

An analogue of acyclovir, ganciclovir is active against HSV and VZV and is markedly more active than acyclovir against CMV. Ganciclovir triphosphate inhibits CMV DNA polymerase and can be incorporated into CMV DNA, whose elongation it eventually terminates. In HSV- and VZV-infected cells, ganciclovir is phosphorylated by virus-encoded thymidine kinases; in CMV-infected cells, it is phosphorylated by a viral kinase encoded by the UL97 gene. Ganciclovir triphosphate is present in tenfold higher concentrations in CMV-infected cells than in uninfected cells. Ganciclovir is approved for the treatment of CMV retinitis in immunosuppressed patients and for the prevention of CMV disease in transplant recipients. It is widely used for the treatment of other CMV-associated syndromes, including pneumonia, esophagogastrointestinal infections, hepatitis, and "wasting" illness.

Ganciclovir is available for IV or oral administration. Because its oral bioavailability is low (5–9%), relatively large doses (1 g three times daily) must be administered by this route. Oral ganciclovir has largely been supplanted by valganciclovir, which is the l-valyl ester of ganciclovir. Valganciclovir is well absorbed orally, with a bioavailability of 60%, and is rapidly hydrolyzed to ganciclovir in the intestine and liver. The area under the curve for a 900-mg dose of valganciclovir is equivalent to that for 5

mg/kg of IV ganciclovir, although peak serum concentrations are ~40% lower for valganciclovir. The serum half-life is 3.5 h after IV administration of ganciclovir and 4.0 h after PO administration of valganciclovir. Ganciclovir is excreted primarily by the kidneys in an unmetabolized form, and its dosage should be reduced in cases of renal failure. Ganciclovir therapy at the most commonly employed initial IV dosage—i.e., 5 mg/kg every 12 h for 14–21 days—can be changed to valganciclovir (900 mg PO twice daily) when the patient can tolerate oral therapy. The maintenance dose is 5 mg/kg IV daily or five times per week for ganciclovir and 900 mg by mouth once a day for valganciclovir. Dose adjustment in patients with renal dysfunction is required. Intraocular ganciclovir, given by either intravitreal injection or intraocular implantation, has also been used to treat CMV retinitis.

Ganciclovir is effective as prophylaxis against CMV-associated disease in organ and bone marrow transplant recipients. Oral ganciclovir administered prophylactically to AIDS patients with CD4+ T cell counts of <100/μL has provided protection against the development of CMV retinitis. However, the long-term benefits of this approach to prophylaxis in AIDS patients have not been established, and most experts do not recommend the use of oral ganciclovir for this purpose. As already mentioned, valganciclovir has supplanted oral ganciclovir in settings where oral prophylaxis or therapy is considered.

The administration of ganciclovir has been associated with profound bone marrow suppression, particularly neutropenia, which significantly limits the drug's use in many patients. Bone marrow toxicity is potentiated in the setting of renal dysfunction and when other bone marrow suppressants, such as zidovudine or mycophenolate mofetil, are used concomitantly.

Resistance has been noted in CMV isolates obtained after therapy with ganciclovir, especially in patients with AIDS. Such resistance may develop through a mutation in either the viral UL97 gene or the viral DNA polymerase. Ganciclovir-resistant isolates are usually sensitive to foscarnet (see below) or cidofovir (see above).

Famciclovir and Penciclovir

Famciclovir is the diacetyl 6-deoxyester of the guanosine analogue penciclovir. Famciclovir is well absorbed orally, has a bioavailability of 77%, and is rapidly converted to penciclovir by deacetylation and oxidation in the intestine and liver. Penciclovir's spectrum of activity and mechanism of action are similar to those of acyclovir. Thus, penciclovir usually is not active against acyclovir-resistant viruses. However, some acyclovir-resistant viruses with altered thymidine kinase or DNA polymerase substrate specificity may be sensitive to penciclovir. This drug is phosphorylated initially by a virus-encoded thymidine kinase and subsequently by cellular kinases to penciclovir triphosphate, which inhibits HSV-1, HSV-2, VZV, and EBV as well as hepatitis B virus (HBV). The serum half-life of penciclovir is 2 h, but the intracellular half-life of penciclovir triphosphate is 7–20 h—markedly longer than that of acyclovir triphosphate. The latter is the basis for the less frequent (twice-daily) dosing schedule for famciclovir than for acyclovir. Penciclovir is eliminated primarily in the urine by both glomerular filtration and tubular secretion. The usually recommended dosage interval should be adjusted for renal insufficiency.

Clinical trials involving immunocompetent adults with herpes zoster showed that famciclovir was superior to placebo in eliciting the resolution of skin lesions and virus shedding and in shortening the duration of postherpetic neuralgia; moreover, administered at 500 mg every 8 h, famciclovir was at least as effective as acyclovir administered at an oral dose of 800 mg five times daily. Famciclovir was also effective in the treatment of herpes zoster in immunosuppressed patients. Clinical trials have demonstrated its effectiveness in the suppression of genital HSV infections for up to 1 year and in the treatment of initial and recurrent episodes of genital herpes. Famciclovir is effective as therapy for mucocutaneous HSV infections in HIV-infected patients. Application of a 1% penciclovir cream reduces the duration of signs and symptoms of herpes labialis in immunocompetent patients (by 0.5–1 day) and has been approved for that purpose by the FDA. Famciclovir is generally well tolerated, with occasional headache, nausea, and diarrhea reported in frequencies similar to those among placebo recipients. The administration of high doses of famciclovir for 2 years was associated with an increased incidence of mammary adenocarcinomas in female rats, but the clinical significance of this effect is unknown.

Foscarnet

Foscarnet (phosphonoformic acid) is a pyrophosphate-containing compound that potently inhibits herpesviruses, including CMV. This drug inhibits DNA polymerases at the pyrophosphate binding site at concentrations that have relatively little effect on cellular polymerases. Foscarnet does not require phosphorylation to exert its antiviral activity and is therefore active against HSV and VZV isolates that are resistant to acyclovir because of deficiencies in -thymidine kinase as well as against most ganciclovir-resistant strains of CMV. Foscarnet also inhibits the reverse transcriptase of HIV and is active against HIV in vivo.

Foscarnet is poorly soluble and must be administered intravenously via an infusion pump in a dilute solution over 1–2 h. The plasma half-life of foscarnet is 3–5 h and increases with decreasing renal function, since the drug is eliminated primarily by the kidneys. It has been estimated that 10–28% of a dose may be deposited in bone, where it can persist for months. The most common initial dosage of

foscarnet–60 mg/kg every 8 h for 14–21 days—is followed by a maintenance dose of 90–120 mg/kg once a day.

Foscarnet is approved for the treatment of CMV retinitis in patients with AIDS and of acyclovir-resistant mucocutaneous HSV infections. In a comparative clinical trial, the drug appeared to be about as efficacious as ganciclovir against CMV retinitis but was associated with a longer survival period, possibly because of its activity against HIV. Intraocular foscarnet has been used to treat CMV retinitis. Foscarnet has also been employed to treat acyclovir-resistant HSV and VZV infections as well as ganciclovir-resistant CMV infections, although resistance to foscarnet has been reported in CMV isolates obtained during therapy. Foscarnet has also been used to treat HHV-6 infections in immunosuppressed patients.

The major form of toxicity associated with foscarnet is renal impairment. Thus renal function should be monitored closely, particularly during the initial phase of therapy. Since foscarnet binds divalent metal ions, hypocalcemia, hypomagnesemia, hypokalemia, and hypo- or hyperphosphatemia can develop. Saline hydration and slow infusion appear to protect the patient against nephrotoxicity and electrolyte disturbances. Although hematologic abnormalities have been documented (most commonly anemia), foscarnet is not generally myelosuppressive and can be administered concomitantly with myelosuppressive medications such as zidovudine.

Trifluridine

Trifluridine is a pyrimidine nucleoside active against HSV-1, HSV-2, and CMV. Trifluridine monophosphate irreversibly inhibits thymidylate synthetase, and trifluridine triphosphate inhibits viral and, to a lesser extent, cellular DNA polymerases. Because of systemic toxicity, its use is limited to topical therapy. Trifluridine is approved for treatment of HSV keratitis, against which trials have shown that it is more effective than topical idoxuridine but similar in efficacy to topical vidarabine. The drug has benefited some patients with HSV keratitis who have failed to respond to idoxuridine or vidarabine. Topical application of trifluridine to sites of acyclovir-resistant HSV mucocutaneous infections has also been beneficial in some cases.

Vidarabine

Vidarabine is a purine nucleoside analogue with activity against HSV-1, HSV-2, VZV, and EBV. Vidarabine inhibits viral DNA synthesis through its 5'-triphosphorylated metabolite, although its precise molecular mechanisms of action are not completely understood. IV-administered vidarabine has been shown to be effective in the treatment of herpes simplex encephalitis, mucocutaneous HSV infections, herpes zoster in immunocompromised patients, and neonatal HSV infections. Its use has been supplanted by that of IV acyclovir, which is more effective and easier to administer. Production of the IV preparation has been discontinued by the manufacturer, but vidarabine is available as an ophthalmic ointment, which is effective in the treatment of HSV keratitis.

ANTIVIRAL DRUGS ACTIVE AGAINST HEPATITIS VIRUSES

Lamivudine

Lamivudine is a pyrimidine nucleoside analogue that is used primarily in combination therapy against HIV infection (Chap. 189). Its activity against hepatitis B virus (HBV) is attributable to inhibition of the viral DNA polymerase. This drug has also been approved for the treatment of chronic HBV infection. At doses of 100 mg/d given for 1 year to patients positive for hepatitis B e antigen (HBeAg), lamivudine is well tolerated and results in suppression of HBV DNA levels, normalization of serum aminotransferase levels in 40–75% of patients, and reduction of hepatic inflammation and fibrosis in 50–60% of patients. Loss of HBeAg occurs in 30% of patients. Lamivudine also appears to be useful in the prevention or suppression of HBV infection associated with liver transplantation. Resistance to lamivudine develops in 24% of patients treated for 1 year and is associated with changes in the YMDD motif of HBV DNA polymerase. Because of the frequency of development of resistance, lamivudine has been largely supplanted by less-resistance-prone drugs for the treatment of HBV infection.

Adefovir Dipivoxil

Adefovir dipivoxil is the oral prodrug of adefovir, an acyclic nucleotide analogue of adenosine monophosphate that has activity against HBV, HIV, HSV, CMV, and poxviruses. It is phosphorylated by cellular kinases to the active triphosphate moiety, which is a competitive inhibitor of HBV DNA polymerase and results in chain termination after incorporation into nascent viral DNA. Adefovir is administered orally and is eliminated primarily by the kidneys, with a plasma half-life of 5–7.5 h. In clinical studies, therapy with adefovir at a dose of 10 mg/d for 48 weeks resulted in normalization of serum alanine aminotransferase (ALT) levels in 48–72% of patients and improved liver histology in 53–64%; it also resulted in a 3.5- to 3.9-$\log_{10}$ reduction in the number of HBV DNA copies per milliliter of plasma. Adefovir was effective in treatment-naïve patients as well as in those infected with lamivudine-resistant HBV. Resistance to adefovir appears to develop less readily than that to lamivudine, but adefovir resistance rates of 15–18% have been reported after 192 weeks of treatment and may reach 30% after 5 years. This agent is generally well tolerated. Significant nephrotoxicity attributable to adefovir is uncommon at the dose employed in the treatment of HBV infections (10 mg/d) but is a

treatment-limiting adverse effect at the higher doses used in therapy for HIV infections (30–120 mg/d). In any case, renal function should be monitored in patients taking adefovir, even at the lower dose. Adefovir is approved only for treatment of chronic HBV infection.

Tenofovir

Tenofovir disoproxil fumarate is a prodrug of tenofovir, a nucleotide analogue of adenosine monophosphate with activity against both retroviruses and hepadnaviruses. In both immunocompetent and immunocompromised patients (including those co-infected with HIV and HBV), tenofovir given at a dose of 300 mg/d for 48 weeks reduced HBV replication by 4.6–6 $\log_{10}$, normalized ALT levels in 68–76% of patients, and improved liver histopathology in 72–74% of patients. Resistance develops uncommonly during ≥2 years of therapy, and tenofovir is active against lamivudine-resistant HBV. The safety profile of tenofovir is similar to that of adefovir, but nephrotoxicity has not been encountered at the dose used for HBV therapy. Tenofovir is approved for the treatment of HIV and chronic HBV infections. **For a more detailed discussion of tenofovir, see Chap. 189.**

Entecavir

Entecavir is a cyclopentyl 2'-deoxyguanosine analogue that inhibits HBV through interaction of entecavir triphosphate with several HBV DNA polymerase functions. At a dose of 0.5 mg/d given for 48 weeks, entecavir reduced HBV DNA copies by 5.0–6.9 $\log_{10}$, normalized serum aminotransferase levels in 68–78% of patients, and improved histopathology in 70–72% of patients. Entecavir inhibits lamivudine-resistant viruses that have M550I or M550V/L526M mutations but only at serum concentrations 20- or 30-fold higher than those obtained with the 0.5-mg/d dose. Thus, higher doses of entecavir (1 mg/d) are recommended for the treatment of patients infected with lamivudine-resistant HBV. Development of resistance to entecavir is uncommon in treatment-naïve patients but does occur at unacceptably high rates (43% after 4 years) in patients previously infected with lamivudine-resistant virus. Entecavir-resistant strains appear to be sensitive to adefovir and tenofovir.

Entecavir is highly bioavailable but should be taken on an empty stomach since food interferes with its absorption. The drug is eliminated primarily in unchanged form by the kidneys, and its dosage should be adjusted for patients with Cr_{Cl} values of <50 mL/min. Overall, entecavir is well tolerated, with a safety profile similar to that of lamivudine. As with other anti-HBV treatments, exacerbation of hepatitis may occur when entecavir therapy is stopped. Entecavir is approved for treatment of chronic hepatitis B, including infection with lamivudine-resistant viruses, in adults. Entecavir has some activity against HIV-1 (median effective concentration, 0.026 to >10 μM) but should not be used as monotherapy in HIV-positive patients because of the potential for development of HIV resistance due to the M184V mutation.

Telbivudine

Telbivudine is a β-l enantiomer of thymidine and is a potent, selective inhibitor of HBV. Its active form is telbivudine triphosphate, which inhibits HBV DNA polymerase and causes chain termination but has little or no activity against human DNA polymerase. Administration of telbivudine at an oral dose of 600 mg/d for 52 weeks to patients with chronic hepatitis B resulted in reduction of HBV DNA by 5.2–6.4 $\log_{10}$ copies/mL along with normalization of ALT levels in 74–77% of recipients and improved histopathology in 65–67% of patients. Telbivudine-resistant HBV is generally cross-resistant with lamivudine-resistant virus but is usually susceptible to adefovir. After 2 years of therapy, resistance to telbivudine was noted in isolates from 22% of HBeAg-positive patients and in those from 9% of HBeAg-negative patients.

Orally administered telbivudine is rapidly absorbed; because it is eliminated primarily by the kidneys, its dosage should be reduced in patients with a Cr_{Cl} value of <50 mL/min. Telbivudine is generally well tolerated, but increases in serum levels of creatinine kinases as well as fatigue and myalgias have been observed. As with other anti-HBV drugs, hepatitis may be exacerbated in patients who discontinue telbivudine therapy. Telbivudine has been approved for the treatment of adults with chronic hepatitis B who have evidence of viral replication and either persistently elevated serum aminotransferase levels or histopathologically active disease, but it has not been widely used because of the frequency of development of resistance noted above.

INTERFERONS

IFNs are cytokines that exhibit a broad spectrum of antiviral activities as well as immunomodulating and antiproliferative properties. IFNs are not available for oral administration but must be given IM, SC, or IV. Early studies with human leukocyte IFN demonstrated an effect in the prophylaxis of experimentally induced rhinovirus infections in humans and in the treatment of VZV infections in immunosuppressed patients. DNA recombinant technology has made available highly purified α, β, and γ IFNs that have been evaluated in a variety of viral infections. Results of such trials have confirmed the effectiveness of intranasally administered IFN in the prophylaxis of rhinovirus infections, although its use has been associated with nasal mucosal irritation. Studies have also demonstrated a beneficial effect of intralesionally or systemically administered IFNs on genital warts. The effect of systemic

administration consists primarily of a reduction in the size of the warts, and this mode of therapy may be useful in persons who have numerous warts that cannot easily be treated by individual intralesional injections. However, lesions frequently recur after either intralesional or systemic IFN therapy is discontinued.

IFNs have undergone extensive study in the treatment of chronic HBV infection. The administration of standard IFN-α2b (5 million units daily or 10 million units three times a week for 16–24 weeks) to patients with stable chronic HBV infection resulted in loss of markers of HBV replication, such as HBeAg and HBV DNA, in 33–37% of cases; 8% of patients also became negative for hepatitis B surface antigen. In most patients who lose HBeAg and HBV DNA markers, serum aminotransferases return to normal levels, and both short- and long-term improvements in liver histopathology have been described. Predictors of a favorable response to standard IFN therapy include low pretherapy levels of HBV DNA, high pretherapy serum levels of ALT, a short duration of chronic HBV infection, and active inflammation in liver histopathology. Poor responses are seen in immunosuppressed patients, including those with HIV infection.

In pegylated IFNs, IFN alphas are linked to polyethylene glycol. This linkage results in slower absorption, decreased clearance, and more sustained serum concentrations, thereby permitting a more convenient, once-weekly dosing schedule; in many instances, pegylated IFN has supplanted standard IFN. After 48 weeks of treatment with 180 μg of pegylated IFN-α2a, HBV DNA was reduced by 4.1–4.5 $\log_{10}$ copies/mL, with normalization of serum ALT levels in 39% of patients and improved histology in 38%. Response rates were somewhat higher when lamivudine was administered with pegylated IFN-α2a. Adverse effects of IFN are common and include fever, chills, myalgia, fatigue, neurotoxicity (manifested primarily as somnolence, depression, anxiety, and confusion), and leukopenia. Autoantibodies (e.g., antithyroid antibodies) can also develop. IFN-α2b and pegylated IFN-α2a are approved for the treatment of patients with chronic hepatitis B. Data supporting the therapeutic efficacy of pegylated interferon-α2b in HBV infection have been published; the drug has not been approved for this indication in the United States but has been approved for treatment of chronic HBV infection in other countries.

Several IFN preparations, including IFN-α2a, IFN-α2b, IFN-alfacon-1, and IFN-αm1 (lymphoblastoid), have been studied as therapy for chronic HCV infections. A variety of monotherapy regimens have been studied, of which the most common for standard IFN is IFN-α2b or -α2a at 3 million units three times per week for 12–18 months. The addition of oral ribavirin to IFN-α2b—either as initial therapy or after failure of IFN therapy alone—results in significantly higher rates of sustained virologic and/or serum ALT responses (40–50%) than are obtained with monotherapy. Comparative studies indicate that pegylated IFN-α2b or -α2a therapy is more effective than standard IFN treatment against chronic HCV infection. The combination of SC pegylated IFN and oral ribavirin is more convenient and appears to be the most effective regimen for treatment of chronic hepatitis C. With this combination regimen, sustained virologic responses (SVRs) were seen in 42–51% of patients with genotype 1 infection and in 76–82% of patients with genotype 2 or 3 infection. Ribavirin appears to have a small antiviral effect in HCV infection but may also be working through an immunomodulatory effect in combination with IFN. Optimal results with ribavirin appear to be associated with weight-based dosing. Prognostic factors for a favorable response include an age of <40 years, a short duration of infection, low levels of HCV RNA, a lesser degree of liver histopathology, and infection with HCV genotypes other than 1. IFN-alfacon, a synthetic "consensus" α interferon, appears to produce response rates similar to those elicited by standard IFN-α2a or -α2b alone and is also approved in the United States for the treatment of chronic hepatitis C.

The efficacy of IFN-α treatment for chronic hepatitis D remains unestablished. Anecdotal reports suggested that doses ranging from 5 million units daily to 9 million units three times per week for 12 months elicit biochemical and virologic responses. Results from small controlled trials have been inconsistent, and observed responses have not generally been sustained. Limited experience has been published with the use of pegylated IFN-α2a or -α2b for treatment of hepatitis D, but some consultants prefer these agents for this indication because of their pharmacologic advantages over standard IFN.

PROTEASE INHIBITORS

This drug class is specifically designed to inhibit the 3/4A (NS3/4A) HCV protease. These agents resemble the HCV polypeptide and, when processed by the viral protease, form a covalent bond with the catalytic serine residues and block further activity. The most clinically advanced compound in this class is telaprevir. In initial phase 1 and 2 clinical studies, telaprevir monotherapy decreased the HCV load by 2–5 $\log_{10}$. Telaprevir in combination with IFN and ribavirin increased the SVR rate from ~40% to 60% when used as primary therapy for genotype 1 infections. For re-treatment of HCV-infected patients in whom prior IFN/ribavirin therapy had failed, the addition of telaprevir plus pegylated IFN/ribavirin increased the SVR rate to 51–53%. The combination of telaprevir/pegylated IFN and ribavirin is superior to this combination without ribavirin. Typically, an oral loading dose of 1125 mg is followed by 750 mg every 8 h orally for 12–24 weeks. Monotherapy is associated with the rapid emergence of antiviral resistance; substitutions are found in the NS3 protease, especially double

variants at positions V35M and R155K. Telaprevir therapy is associated with rashes in ~50% of patients; these eruptions are severe in ~5% of cases and often develop weeks after therapy has begun. Data suggest that telaprevir has the potential to increase the SVR rate and may shorten the overall duration of HCV therapy from 48 to 24 weeks when used in conjunction with pegylated IFN and ribavirin. As of this writing (February 2011), protease inhibitors have not been approved for treatment of HCV.

FURTHER READINGS

Aoki F et al: Antiviral drugs (other than antiretrovirals), in *Principles and Practice of Infectious Diseases*, 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 565 –610 Couch RB: Drug therapy: Prevention and treatment of influenza. N Engl J Med 343:1778, 2000

Crumpacker CS: Ganciclovir. N Engl J Med 335:721, 1996[PMID: 8786764]

Dolin R et al: A controlled trial of amantadine and rimantadine in the prophylaxis of influenza A infection. N Engl J Med 307:580, 1982[PMID: 7050702]

Field JJ, Hoofnagle JH: Mechanism of action of interferon and ribavirin in treatment of hepatitis C. Nature 436:967, 2005

Gish RG et al: Safety and antiviral activity of emtricitabine (FTC) for the treatment of chronic hepatitis B infection: A two-year study. J Hepatol 43:60, 2005[PMID: 15922478]

Hall CB et al: Aerosolized ribavirin treatment of infants with respiratory syncytial viral infection: A randomized double-blind study. N Engl J Med 308:1443, 1983[PMID: 6343860]

Lai CL et al: Entecavir versus lamivudine for patients with HBeAg-negative chronic hepatitis B. N Engl J Med 354:186, 2006

Lalezari JP et al: Randomized controlled study of the safety and efficacy of IV cidofovir for the treatment of relapsing cytomegalovirus retinitis in patients with AIDS. J AIDS 17:339, 1998[PMID: 9525435]

Lok AS et al: Management of hepatitis B: 2000—summary of a workshop. Gastroenterology 120:1828, 2001[PMID: 11375963]

Martin DF et al: A controlled trial of valganciclovir as induction therapy for cytomegalovirus retinitis. N Engl J Med 346:1119, 2002[PMID: 11948271]

National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C. September 12, 2002 (available at *www.niaid.nih.gov*)

Piret J, Boivin G: Resistance of herpes simplex viruses to nucleoside analogues: Mechanisms, prevalence, and management. Antimicrob Agents Chemother 55:459, 2011[PMID: 21078929]

Treanor JJ et al: Efficacy and safety in treating acute influenza: A randomized controlled trial. U.S. Oral Neuraminidase Study Group. JAMA 283:1016, 2000[PMID: 10697061]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HERPES SIMPLEX VIRUS INFECTIONS: INTRODUCTION

Definition

Herpes simplex viruses (HSV-1, HSV-2; *Herpesvirus hominis*) produce a variety of infections involving mucocutaneous surfaces, the central nervous system (CNS), and—on occasion—visceral organs. Prompt recognition and treatment reduce the morbidity and mortality rates associated with HSV infections.

Etiologic Agent

The genome of HSV is a linear, double-strand DNA molecule (molecular weight, ~100 x 10^6) that encodes >90 transcription units with 84 identified proteins. The genomic structures of the two HSV subtypes are similar. The overall genomic sequence homology between HSV-1 and HSV-2 is ~50%, while the proteome homology is >80%. The homologous sequences are distributed over the entire genome map, and most of the polypeptides specified by one viral type are antigenically related to polypeptides of the other viral type. Many type-specific regions unique to HSV-1 and HSV-2 proteins do exist, however, and a number of them appear to be important in host immunity. These type-specific regions have been used to develop serologic assays that distinguish between the two viral subtypes. Either restriction endonuclease analysis or sequencing of viral DNA can be used to distinguish between the two subtypes and among strains of each subtype. The variability of nucleotide sequences from clinical strains of HSV-1 and HSV-2 is such that HSV isolates obtained from two individuals can be differentiated by restriction enzyme patterns or genomic sequences. Moreover, epidemiologically related sources, such as sexual partners, mother-infant pairs, or persons involved in a common-source outbreak, can be inferred from such patterns.

The viral genome is packaged in a regular icosahedral protein shell (capsid) composed of 162 capsomeres (see Fig. 177-1). The outer covering of the virus is a lipid-containing membrane (envelope) acquired as the DNA-containing capsid buds through the inner nuclear membrane of the host cell. Between the capsid and lipid bilayer of the envelope is the tegument. Viral replication has both nuclear and cytoplasmic phases. Initial attachment to the cell membrane involves interactions of viral glycoproteins C and B with several cellular heparan sulfate–like surface receptors. Subsequently, viral glycoprotein D binds to cellular co-receptors that belong to the tumor necrosis factor receptor family of proteins, the immunoglobulin superfamily (nectin family), or both. The ubiquity of these receptors contributes to the wide host range of herpesviruses. Replication is highly regulated. After fusion and entry, the nucleocapsid enters the cytoplasm and several viral proteins are released from the virion. Some of these viral proteins shut off host protein synthesis (by increasing cellular RNA degradation), while others "turn on" the transcription of early genes of HSV replication. These early gene products, designated α *genes*, are required for synthesis of the subsequent polypeptide group, the β polypeptides, many of which are regulatory proteins and enzymes required for DNA replication. Most current antiviral drugs interfere with β proteins, such as viral DNA polymerase. The third (γ) class of HSV genes requires viral DNA replication for expression and constitutes most structural proteins specified by the virus.

After viral genome replication and structural protein synthesis, nucleocapsids are assembled in the cell's nucleus. Envelopment occurs as the nucleocapsids bud through the inner nuclear membrane into the perinuclear space. In some cells, viral replication in the nucleus forms two types of inclusion bodies: type A basophilic Feulgen-positive bodies that contain viral DNA and eosinophilic inclusion bodies that are devoid of viral nucleic acid or protein and represent a "scar" of viral infection. Enveloped virions are then transported via the endoplasmic reticulum and the Golgi apparatus to the cell surface.

Viral genomes are maintained by some neuronal cells in a repressed state called *latency*. Latency, which is associated with transcription of only a limited number of virus-encoded RNAs, accounts for the presence of viral DNA and RNA in neural tissue at times when infectious virus cannot be isolated. Maintenance and growth of neural cells from latently infected ganglia in tissue culture result in production of infectious virions (explantation) and in subsequent permissive infection of susceptible cells (cocultivation). Activation of the viral genome may then occur, resulting in *reactivation*, the normal pattern of regulated viral gene expression, replication, and release of HSV. The release of virions from the neuron follows a complex process of anterograde transport down the length of neuronal axons. In experimental animals, ultraviolet light, systemic and local immunosuppression, and trauma to the skin

or ganglia are associated with reactivation.

To date, three noncoding RNA latency-associated transcripts (LATs) are the only abundant transcripts in the nuclei of latently infected neurons. Deletion mutants of the genomic region that can become latent have been made, and the efficiency of their later reactivation is reduced. In addition, substitution of HSV-1 LATs for HSV-2 LATs induces an HSV-1 reactivation pattern. Thus, LATs appear to maintain —rather than establish—latency. HSV-1 LATs promote the survival of acutely infected neurons, perhaps by inhibiting apoptotic pathways. Highly expressed during latency, LAT-derived micro-RNA appears to silence expression of the key neurovirulence factor infected-cell protein 34.5 (ICP34.5) and to bind in an antisense configuration to ICP0 messenger RNA to prevent expression of this immediate-early protein, which is vital to HSV reactivation. Studies of individual neurons from cadaveric trigeminal ganglionic explants by microdissection and real-time polymerase chain reaction (PCR) revealed that many more neurons (2–11%) harbor HSV than would be predicted by in situ hybridization studies for LAT and that DNA copy number is similar in LAT-positive and LAT-negative neurons. These findings make it less clear what role LATs play in preventing reactivation. At present, the molecular mechanisms of HSV latency are not completely understood; CD8+ T cells have been found in ganglia of experimental animals and humans and appear to influence the process of reactivation, possibly by inducing antiviral factors such as interferon (IFN) γ. Strategies to interrupt or maintain latency in neurons are not available.

Pathogenesis

Exposure to HSV at mucosal surfaces or abraded skin sites permits entry of the virus into cells of the epidermis and dermis and initiation of viral replication therein. HSV infections are usually acquired subclinically. Whether clinical or subclinical, HSV acquisition is associated with sufficient viral replication to permit infection of either sensory or autonomic nerve endings. On entry into the neuronal cell, the virus—or, more likely, the nucleocapsid—is transported intra-axonally to the nerve cell bodies in ganglia. In humans, the transit interval from inoculation of virus in peripheral tissue to spread to the ganglia is unknown. During the initial phase of infection, viral replication occurs in ganglia and contiguous neural tissue. Virus then spreads to other mucocutaneous surfaces through centrifugal migration of infectious virions via peripheral sensory nerves. This mode of spread helps explain the large surface area involved, the high frequency of new lesions distant from the initial crop of vesicles that is characteristic in patients with primary genital or oral-labial HSV infection, and the ability to recover virus from neural tissue distant from neurons innervating the inoculation site. Contiguous spread of locally inoculated virus also may take place and allow further mucosal extension of disease. Recent studies have demonstrated HSV viremia—another mechanism for extension of infection throughout the body—in ~30–40% of persons with primary HSV-2 infection. Latent infection with both viral subtypes in both sensory and autonomic ganglia has been demonstrated. For HSV-1 infection, trigeminal ganglia are most commonly infected, although extension to the inferior and superior cervical ganglia also occurs. With genital infection, sacral nerve root ganglia (S2–S5) are most commonly affected.

After resolution of primary disease, infectious HSV can no longer be cultured from the ganglia; however, latent infection, as defined by the presence of viral DNA, persists in 2–11% of ganglionic cells in the anatomic region of the initial infection. The mechanism of reactivation from latency is unknown. Increasingly, studies indicate that host T cell responses at the ganglionic and peripheral mucosal level influence the frequency and severity of HSV reactivation. HSV-specific T cells have been recovered from peripheral nerve root ganglia. Many of these resident CD8+ T cells are juxtaposed with latently HSV-1-infected neurons in the trigeminal ganglia and can block reactivation with both IFN-γ release and granzyme B degradation of the immediate-early protein ICP4. In addition, there appears to be a latent viral load in the ganglia that correlates positively with the number of neurons infected and the rate of reactivation but inversely with the number of CD8+ cells present. It is not known whether reactivating stimuli transiently suppress these immune cells, independently upregulate transcription of lytic genes, or both. However, once virus reaches the dermal-epidermal junction, there are two possible outcomes: subclinical shedding or recurrence (the latter defined clinically by a skin blister and ulceration). Histologically, herpetic lesions involve a thin-walled vesicle or ulceration in the basal region, multinucleated cells that may include intranuclear inclusions, necrosis, and an acute inflammatory infection. Re-epithelialization occurs once viral replication is restricted, almost always in the absence of a scar.

Analysis of the DNA from sequential isolates of HSV or from isolates from multiple infected ganglia in any one individual has revealed similar, if not identical, restriction endonuclease or DNA sequence patterns in most persons. The finding of individual neurons infected with multiple strains of drug-susceptible and drug-resistant virus in severely immunosuppressed patients indicates that ganglia can be reseeded during chronic infection. As exposure to mucosal shedding is relatively common during a person's lifetime, current data suggest that exogenous infection with different strains of the same subtype, while possible, is uncommon.

Immunity

Host responses influence the acquisition of HSV disease, the severity of infection, resistance to the

development of latency, the maintenance of latency, and the frequency of recurrences. Both antibody-mediated and cell-mediated reactions are clinically important. Immunocompromised patients with defects in cell-mediated immunity experience more severe and more extensive HSV infections than those with deficits in humoral immunity, such as agammaglobulinemia. Experimental ablation of lymphocytes indicates that T cells play a major role in preventing lethal disseminated disease, although antibodies help reduce titers of virus in neural tissue. Some clinical manifestations of HSV appear to be related to the host immune response (e.g., stromal opacities associated with recurrent herpetic keratitis). The surface viral glycoproteins have been shown to be targets of antibodies that mediate neutralization and immune-mediated cytolysis (antibody-dependent cell-mediated cytotoxicity). Monoclonal antibodies specific for each of the known viral glycoproteins have, in experimental infections, conferred protection against subsequent neurologic disease or ganglionic latency. In humans, however, subunit glycoprotein vaccines have been only partially successful in reducing acquisition of infection. Multiple cell populations, including natural killer cells, macrophages, and a variety of T lymphocytes, play a role in host defenses against HSV infections, as do lymphokines generated by T lymphocytes. In animals, passive transfer of primed lymphocytes confers protection from subsequent challenge. Maximal protection usually requires the activation of multiple T cell subpopulations, including cytotoxic T cells and T cells responsible for delayed hypersensitivity. The latter cells may confer protection by the antigen-stimulated release of lymphokines (e.g., IFNs), which in turn have a direct antiviral effect and both activate and enhance a variety of specific and nonspecific effector cells. Increasing evidence suggests that HSV-specific CD8+ T cell responses are critical for clearance of virus from lesions. In addition, immunosuppressed patients with frequent and prolonged HSV lesions have fewer functional CD8+ T cells directed at HSV. The HSV virion contains a variety of genes that are directed at the inhibition of host responses. These include gene no. 12 (*US-12*), which can bind to the cellular transporter-activating protein TAP-1 and reduce the ability of this protein to bind HSV peptides to human leukocyte antigen (HLA) class I, thereby reducing recognition of viral proteins by cytotoxic T cells of the host. This effect can be overcome by the addition of IFN-γ, but this reversal requires 24–48 h; thus, the virus has time to replicate and invade other host cells. Entry of infectious HSV-1 and HSV-2 inhibits several signaling pathways of both CD4+ and CD8+ T cells, leading to their functional impairment in killing and influencing the spectrum of their cytokine secretion.

Recent studies suggest that the rate of HSV reactivation is far more frequent than previously recognized. PCR analysis of daily anogenital swab samples has shown that the virus is shed on a median of 25% of days by the 95% of patients who are positive for antibody to HSV-2 and who shed virus, with a wide range of interpatient variability (range, 2–75%). In studies with sampling performed every 6 h, 49% of genital reactivation episodes lasted <12 h and 29% lasted <6 h. Many of these short bursts of reactivation were associated with copy numbers thought to be high enough to cause transmission to susceptible sexual partners. These data suggest that peripheral immune control may dictate the likelihood and severity of recurrences as well as the frequency of subclinical shedding. There is a strong association between the magnitude of the CD8+ T lymphocyte response and the clearance of virus from genital lesions. The lack of this response, rather than a low CD4+ T lymphocyte count, also predicts frequent and severe HSV-2 recurrences in untreated as well as treated HIV-1-infected patients. HSV-2-specific CD8+ and CD4+ T cells appear to persist for prolonged periods (months) in genital skin previously involved in an HSV-2 reactivation. The location, effectiveness, and longevity of the T lymphocytes (and perhaps of other immune effector cells) may be important in the expression of disease and the likelihood of transmission over time.

Epidemiology

Seroepidemiologic studies have documented HSV infections worldwide. Serologic assays with whole-virus antigen preparations, such as complement fixation, neutralization, indirect immunofluorescence, passive hemagglutination, radioimmunoassay, and enzyme-linked immunosorbent assay, are useful for differentiating uninfected (seronegative) persons from those with past HSV-1 or HSV-2 infection, but they do not reliably distinguish between the two viral subtypes. Serologic assays that identify antibodies to type-specific surface proteins (epitopes) of the two viral subtypes have been developed and can distinguish reliably between the human antibody responses to HSV-1 and HSV-2. The most commonly used assays are those that measure antibodies to glycoprotein G of HSV-1 (gG1) and HSV-2 (gG2). A western blot assay that can detect several HSV type-specific proteins can also be used.

Infection with HSV-1 is acquired more frequently and earlier than infection with HSV-2. More than 90% of adults have antibodies to HSV-1 by the fifth decade of life. In populations of low socioeconomic status, most persons acquire HSV-1 infection before the third decade of life. Antibodies to HSV-2 are not detected routinely until puberty. Antibody prevalence rates correlate with past sexual activity and vary greatly among different population groups. There is some evidence that the prevalence of HSV-2 has decreased slightly over the past decade in the United States. Serosurveys indicate that 15–20% of the U.S. population has antibodies to HSV-2. In most routine obstetric and family planning clinics, 25% of women have HSV-2 antibodies, although only 10% of those who are seropositive for HSV-2 report a history of genital lesions. As many as 50% of heterosexual adults attending sexually transmitted disease clinics have antibodies to HSV-2.

A wide variety of serologic surveys have indicated a similar or even higher seroprevalence of HSV-2 in

most parts of Central America, South America, and Africa. There is an epidemiologic synergy between HSV-2 and HIV-1. HSV-2 infection is associated with a two- to fourfold increase in HIV-1 acquisition. In addition, HSV-2 is reactivated and transmitted more frequently in persons co-infected with HIV-1 and HSV-2 than in persons not infected with HIV-1. Thus, most areas of the world with a high HIV-1 prevalence also have a high HSV-2 prevalence. In Africa, HSV-2 seroprevalence has ranged from 40% to 70% in obstetric and other sexually experienced populations. Antibody prevalence rates average ~5–10% higher among women than among men.

Several studies suggest that many cases of "asymptomatic" genital HSV-2 infection are, in fact, simply unrecognized: when "asymptomatic" seropositive persons are shown pictures of genital lesions, >60% subsequently identify episodes of symptomatic reactivation. Most important, these asymptomatic seropositive persons with reactivation shed virus on mucosal surfaces almost as frequently as do those with symptomatic disease. The large reservoir of unidentified carriers of HSV-2 and the frequent asymptomatic reactivation of the virus from the genital tract have fostered the continued spread of genital herpes throughout the world. HSV-2 infection is an independent risk factor for the acquisition and transmission of infection with HIV-1. Among co-infected persons, HIV-1 virions can be shed from herpetic lesions of the genital region. This shedding may facilitate the spread of HIV through sexual contact. HSV-2 reactivation is associated with a localized persistent inflammatory response consisting of high concentrations of CCR5-enriched CD4+ T cells as well as inflammatory dendritic cells in the submucosa of the genital skin. These cells can support HIV infection and replication and hence are likely to account for the two- to threefold increase in HIV acquisition among persons with genital herpes. Unfortunately, antiviral therapy does not reduce this subclinical postreactivation inflammation, probably because of the inability of current antiviral agents to prevent the release of small amounts of HSV antigen into the genital mucosa.

HSV infections occur throughout the year. Transmission can result from contact with persons who have active ulcerative lesions or with persons who have no clinical manifestations of infection but who are shedding HSV from mucocutaneous surfaces. HSV reactivation on genital skin and mucosal surfaces is common. With once-daily sampling of immunocompetent adults, HSV-2 can be cultured from the genital tract on 2–10% of days tested, and HSV DNA can be detected on 20–30% of days by PCR. Corresponding figures for HSV-1 in oral secretions are similar. Rates of shedding are highest during the initial years after acquisition, with viral shedding occurring on as many as 30–50% of days during this period. Immunosuppressed patients shed HSV from mucosal sites at an even higher frequency (20–80% of days). With increased sampling frequency (e.g., four times daily), the rates of reactivation are two- to fourfold higher, with many episodes lasting <12 h. These high rates of mucocutaneous reactivation suggest that exposure to HSV from sexual or other close contact (kissing, sharing of glasses or silverware) is common and help explain the continuing spread and high seroprevalence of HSV infections worldwide. Reactivation rates vary widely among individuals. Among HIV-positive patients, a low CD4+ T cell count and a heavy viral load are associated with increased rates of HSV reactivation. Daily antiviral chemotherapy for HSV-2 infection can reduce shedding rates but does not eliminate shedding, as measured by PCR or culture.

Clinical Spectrum

HSV has been isolated from nearly all visceral and mucocutaneous sites. The clinical manifestations and course of HSV infection depend on the anatomic site involved, the age and immune status of the host, and the antigenic type of the virus. Primary HSV infections (i.e., first infections with either HSV-1 or HSV-2 in which the host lacks HSV antibodies in acute-phase serum) are frequently accompanied by systemic signs and symptoms. Compared with recurrent episodes, primary infections, which involve both mucosal and extramucosal sites, are characterized by a longer duration of symptoms and virus isolation from lesions. The incubation period ranges from 1 to 26 days (median, 6–8 days). Both viral subtypes can cause genital and oral-facial infections, and the infections caused by the two subtypes are clinically indistinguishable. However, the frequency of reactivation of infection is influenced by anatomic site and virus type. Genital HSV-2 infection is twice as likely to reactivate and recurs 8–10 times more frequently than genital HSV-1 infection. Conversely, oral-labial HSV-1 infection recurs more frequently than oral-labial HSV-2 infection. Asymptomatic shedding rates follow the same pattern.

ORAL-FACIAL INFECTIONS

Gingivostomatitis and pharyngitis are the most common clinical manifestations of first-episode HSV-1 infection, while recurrent herpes labialis is the most common clinical manifestation of reactivation HSV-1 infection. HSV pharyngitis and gingivostomatitis usually result from primary infection and are most commonly seen among children and young adults. Clinical symptoms and signs, which include fever, malaise, myalgias, inability to eat, irritability, and cervical adenopathy, may last 3–14 days. Lesions may involve the hard and soft palate, gingiva, tongue, lip, and facial area. HSV-1 or HSV-2 infection of the pharynx usually results in exudative or ulcerative lesions of the posterior pharynx and/or tonsillar pillars. Lesions of the tongue, buccal mucosa, or gingiva may occur later in the course in one-third of cases. Fever lasting 2–7 days and cervical adenopathy are common. It can be difficult to differentiate HSV pharyngitis clinically from bacterial pharyngitis, *Mycoplasma pneumoniae* infections, and pharyngeal ulcerations of noninfectious etiologies (e.g., Stevens-Johnson syndrome). No substantial

evidence suggests that reactivation of oral-labial HSV infection is associated with symptomatic recurrent pharyngitis.

Reactivation of HSV from the trigeminal ganglia may be associated with asymptomatic virus excretion in the saliva, development of intraoral mucosal ulcerations, or herpetic ulcerations on the vermilion border of the lip or external facial skin. About 50–70% of seropositive patients undergoing trigeminal nerve-root decompression and 10–15% of those undergoing dental extraction develop oral-labial HSV infection a median of 3 days after these procedures. Clinical differentiation of intraoral mucosal ulcerations due to HSV from aphthous, traumatic, or drug-induced ulcerations is difficult.

In immunosuppressed patients, HSV infection may extend into mucosal and deep cutaneous layers. Friability, necrosis, bleeding, severe pain, and inability to eat or drink may result. The lesions of HSV mucositis are clinically similar to mucosal lesions caused by cytotoxic drug therapy, trauma, or fungal or bacterial infections. Persistent ulcerative HSV infections are among the most common infections in patients with AIDS. HSV and *Candida* infections often occur concurrently. Systemic antiviral therapy speeds the rate of healing and relieves the pain of mucosal HSV infections in immunosuppressed patients. The frequency of HSV reactivation during the early phases of transplantation or induction chemotherapy is high (50–90%), and prophylactic systemic antiviral agents such as IV acyclovir and penciclovir or the oral congeners of these drugs are used to reduce reactivation rates. Patients with atopic eczema may also develop severe oral-facial HSV infections (*eczema herpeticum*), which may rapidly involve extensive areas of skin and occasionally disseminate to visceral organs. Extensive eczema herpeticum has resolved promptly with the administration of IV acyclovir. Erythema multiforme may also be associated with HSV infections (see Figs. 51-9 and e7-25); some evidence suggests that HSV infection is the precipitating event in ~75% of cases of cutaneous erythema multiforme. HSV antigen has been demonstrated both in circulatory immune complexes and in skin lesion biopsy samples from these cases. Patients with severe HSV-associated erythema multiforme are candidates for chronic suppressive oral antiviral therapy.

HSV-1 and varicella-zoster virus (VZV) have been implicated in the etiology of Bell's palsy (flaccid paralysis of the mandibular portion of the facial nerve). Some but not all trials have documented quicker resolution of facial paralysis with the prompt initiation of antiviral therapy, with or without glucocorticoids. However, other trials have shown little benefit. Thus there is no consensus on the relative value of antiviral drugs alone, glucocorticoids alone, and the two modalities combined for the treatment of Bell's palsy.

GENITAL INFECTIONS

First-episode primary genital herpes is characterized by fever, headache, malaise, and myalgias. Pain, itching, dysuria, vaginal and urethral discharge, and tender inguinal lymphadenopathy are the predominant local symptoms. Widely spaced bilateral lesions of the external genitalia are characteristic **(Fig. 179-1)**. Lesions may be present in varying stages, including vesicles, pustules, or painful erythematous ulcers. The cervix and urethra are involved in >80% of women with first-episode infections. First episodes of genital herpes in patients who have had prior HSV-1 infection are associated with systemic symptoms in a few patients and with faster healing than primary genital herpes. Subclinical DNAemia has been found in ~30% of cases of true primary genital herpes. The clinical courses of acute first-episode genital herpes are similar for HSV-1 and HSV-2 infection. However, the recurrence rates of genital disease differ with the viral subtype: the 12-month recurrence rates among patients with first-episode HSV-2 and HSV-1 infections are ~90% and ~55%, respectively (median number of recurrences, 4 and <1, respectively). Recurrence rates for genital HSV-2 infections vary greatly among individuals and over time within the same individual. HSV has been isolated from the urethra and urine of men and women without external genital lesions. A clear mucoid discharge and dysuria are characteristics of symptomatic HSV urethritis. HSV has been isolated from the urethra of 5% of women with the dysuria-frequency syndrome. Occasionally, HSV genital tract disease is manifested by endometritis and salpingitis in women and by prostatitis in men. About 15% of cases of HSV-2 acquisition are associated with nonlesional clinical syndromes, such as aseptic meningitis, cervicitis, or urethritis. A more complete discussion of the differential diagnosis of genital herpes is presented in Chap. 130.

Figure 179-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Genital herpes: primary vulvar infection. Multiple, extremely painful, punched-out, confluent, shallow ulcers on the edematous vulva and perineum. Micturition is often very painful. Associated inguinal lymphadenopathy is common. *(Reprinted with permission from K Wolff, RA Johnson, D Summond: Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology, 5th ed, New York, McGraw-Hill, 2005.)*

Both HSV-1 and HSV-2 can cause symptomatic or asymptomatic rectal and perianal infections. HSV proctitis is usually associated with rectal intercourse. However, subclinical perianal shedding of HSV is detected in women and men who report no rectal intercourse. This phenomenon is due to the establishment of latency in the sacral dermatome from prior genital tract infection, with subsequent reactivation in epithelial cells in the perianal region. Such reactivations are often subclinical. Symptoms of HSV proctitis include anorectal pain, anorectal discharge, tenesmus, and constipation. Sigmoidoscopy reveals ulcerative lesions of the distal 10 cm of the rectal mucosa. Rectal biopsies show mucosal ulceration, necrosis, polymorphonuclear and lymphocytic infiltration of the lamina propria, and (in occasional cases) multinucleated intranuclear inclusion–bearing cells. Perianal herpetic lesions are also found in immunosuppressed patients receiving cytotoxic therapy. Extensive perianal herpetic lesions and/or HSV proctitis is common among patients with HIV infection.

HERPETIC WHITLOW

Herpetic whitlow—HSV infection of the finger—may occur as a complication of primary oral or genital herpes by inoculation of virus through a break in the epidermal surface or by direct introduction of virus into the hand through occupational or some other type of exposure. Clinical signs and symptoms include abrupt-onset edema, erythema, and localized tenderness of the infected finger. Vesicular or pustular lesions of the fingertip that are indistinguishable from lesions of pyogenic bacterial infection are seen. Fever, lymphadenitis, and epitrochlear and axillary lymphadenopathy are common. The infection may recur. Prompt diagnosis (to avoid unnecessary and potentially exacerbating surgical therapy and/or transmission) is essential. Antiviral chemotherapy is usually recommended (see below).

HERPES GLADIATORUM

HSV may infect almost any area of skin. Mucocutaneous HSV infections of the thorax, ears, face, and hands have been described among wrestlers. Transmission of these infections is facilitated by trauma to the skin sustained during wrestling. Several recent outbreaks have illustrated the importance of prompt diagnosis and therapy to contain the spread of this infection.

EYE INFECTIONS

HSV infection of the eye is the most common cause of corneal blindness in the United States. HSV keratitis presents as an acute onset of pain, blurred vision, chemosis, conjunctivitis, and characteristic dendritic lesions of the cornea. Use of topical glucocorticoids may exacerbate symptoms and lead to involvement of deep structures of the eye. Debridement, topical antiviral treatment, and/or IFN therapy hastens healing. However, recurrences are common, and the deeper structures of the eye may sustain immunopathologic injury. Stromal keratitis due to HSV appears to be related to T cell–dependent destruction of deep corneal tissue. An HSV-1 epitope that is autoreactive with T cell–targeting corneal

antigens has been postulated to be a factor in this infection. Chorioretinitis, usually a manifestation of disseminated HSV infection, may occur in neonates or in patients with HIV infection. HSV and VZV can cause acute necrotizing retinitis as an uncommon but severe manifestation.

CENTRAL AND PERIPHERAL NERVOUS SYSTEM INFECTIONS

HSV accounts for 10–20% of all cases of sporadic viral encephalitis in the United States. The estimated incidence is ~2.3 cases per 1 million persons per year. Cases are distributed throughout the year, and the age distribution appears to be biphasic, with peaks at 5–30 and >50 years of age. HSV-1 causes >95% of cases.

The pathogenesis of HSV encephalitis varies. In children and young adults, primary HSV infection may result in encephalitis; presumably, exogenously acquired virus enters the CNS by neurotropic spread from the periphery via the olfactory bulb. However, most adults with HSV encephalitis have clinical or serologic evidence of mucocutaneous HSV-1 infection before the onset of CNS symptoms. In ~25% of the cases examined, the HSV-1 strains from the oropharynx and brain tissue of the same patient differ; thus some cases may result from reinfection with another strain of HSV-1 that reaches the CNS. Two theories have been proposed to explain the development of actively replicating HSV in localized areas of the CNS in persons whose ganglionic and CNS isolates are similar. Reactivation of latent HSV-1 infection in trigeminal or autonomic nerve roots may be associated with extension of virus into the CNS via nerves innervating the middle cranial fossa. HSV DNA has been demonstrated by DNA hybridization in brain tissue obtained at autopsy—even from healthy adults. Thus, reactivation of long-standing latent CNS infection may be another mechanism for the development of HSV encephalitis.

Recent studies have identified genetic polymorphisms in two separate genes among families with a high frequency of HSV encephalitis. Peripheral-blood mononuclear cells from these patients (predominantly children) appear to secrete reduced levels of IFN in response to HSV. These observations suggest that some cases of sporadic HSV encephalitis may be related to host genetic determinants.

The clinical hallmark of HSV encephalitis has been the acute onset of fever and focal neurologic symptoms and signs, especially in the temporal lobe **(Fig. 179-2)**. Clinical differentiation of HSV encephalitis from other viral encephalitides, focal infections, or noninfectious processes is difficult. Elevated cerebrospinal fluid (CSF) protein levels, leukocytosis (predominantly lymphocytes), and red blood cell counts due to hemorrhagic necrosis are common. While brain biopsy has been the gold standard for defining HSV encephalitis, a highly sensitive and specific PCR for detection of HSV DNA in CSF has largely replaced biopsy for defining CNS infection. Although titers of antibody to HSV in CSF and serum increase in most cases of HSV encephalitis, they rarely do so earlier than 10 days into the illness and therefore, while useful retrospectively, are generally not helpful in establishing an early clinical diagnosis. Demonstration of HSV antigen, HSV DNA, or HSV replication in brain tissue obtained by biopsy is highly sensitive and has a low complication rate; examination of such tissue also provides the best opportunity to identify alternative, potentially treatable causes of encephalitis. Antiviral chemotherapy with acyclovir reduces the rate of death from HSV encephalitis. Even with therapy, however, neurologic sequelae are common, especially among persons >50 years of age. Most authorities recommend the administration of IV acyclovir to patients with presumed HSV encephalitis until the diagnosis is confirmed or an alternative diagnosis is made. All confirmed cases should be treated with IV acyclovir (30 mg/kg per day in three divided doses for 14–21 days). After the completion of therapy, the clinical recurrence of encephalitis requiring more treatment has been reported. For this reason, some authorities prefer to treat initially for 21 days, and many continue therapy until HSV DNA has been eliminated from the CSF. Even with therapy, neurologic sequelae are common, especially among persons >35 years of age.

Figure 179-2

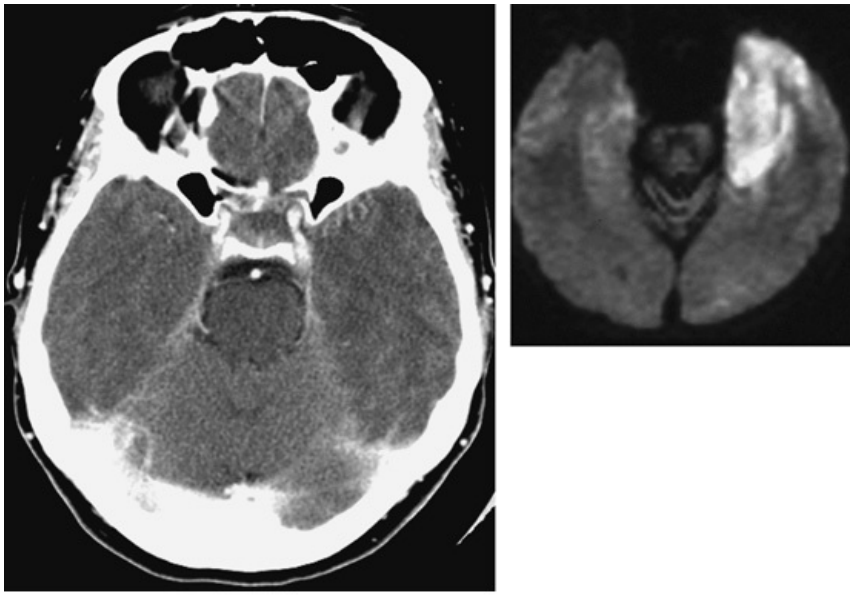

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT and diffusion-weighted MRI scans of the brain of a patient with left-temporal-lobe HSV encephalitis.

HSV DNA has been detected in CSF from 3–15% of persons presenting to the hospital with aseptic meningitis. HSV meningitis, which is usually seen in association with primary genital HSV infection, is an acute, self-limited disease manifested by headache, fever, and mild photophobia and lasting 2–7 days. Lymphocytic pleocytosis in the CSF is characteristic. Neurologic sequelae of HSV meningitis are rare. HSV is the most commonly identified cause of recurrent lymphocytic meningitis (Mollaret's meningitis). Demonstration of HSV antibodies in CSF or persistence of HSV DNA in CSF can establish the diagnosis. For persons with frequent recurrences of HSV meningitis, daily antiviral therapy has reduced the occurrence of such episodes.

Autonomic nervous system dysfunction, especially of the sacral region, has been reported in association with both HSV and VZV infections. Numbness, tingling of the buttocks or perineal areas, urinary retention, constipation, CSF pleocytosis, and (in males) impotence may occur. Symptoms appear to resolve slowly over days or weeks. Occasionally, hypoesthesia and/or weakness of the lower extremities persists for many months. Rarely, transverse myelitis, manifested by a rapidly progressive symmetric paralysis of the lower extremities or Guillain-Barré syndrome, follows HSV infection. Similarly, peripheral nervous system involvement (Bell's palsy) or cranial polyneuritis may be related to reactivation of HSV-1 infection. Transitory hypoesthesia of the area of skin innervated by the trigeminal nerve and vestibular system dysfunction (as measured by electronystagmography) are the predominant signs of disease. Whether antiviral chemotherapy can abort these signs or reduce their frequency and severity is not yet known.

VISCERAL INFECTIONS

HSV infection of visceral organs usually results from viremia, and multiple-organ involvement is common. Occasionally, however, the clinical manifestations of HSV infection involve only the esophagus, lung, or liver. HSV esophagitis may result from direct extension of oral-pharyngeal HSV infection into the esophagus or may occur de novo by reactivation and spread of HSV to the esophageal mucosa via the vagus nerve. The predominant symptoms of HSV esophagitis are odynophagia, dysphagia, substernal pain, and weight loss. Multiple oval ulcerations appear on an erythematous base with or without a patchy white pseudomembrane. The distal esophagus is most commonly involved. With extensive disease, diffuse friability may spread to the entire esophagus. Neither endoscopic nor barium examination can reliably differentiate HSV esophagitis from *Candida* esophagitis or from esophageal ulcerations due to thermal injury, radiation, or corrosives. Endoscopically obtained secretions for cytologic examination and culture or DNA detection by PCR provide the most useful material for diagnosis. Systemic antiviral chemotherapy usually reduces the severity and duration of symptoms and heals esophageal ulcerations.

HSV pneumonitis is uncommon except in severely immunosuppressed patients and may result from extension of herpetic tracheobronchitis into lung parenchyma. Focal necrotizing pneumonitis usually ensues. Hematogenous dissemination of virus from sites of oral or genital mucocutaneous disease may also occur, producing bilateral interstitial pneumonitis. Bacterial, fungal, and parasitic pathogens are commonly present in HSV pneumonitis. The mortality rate from untreated HSV pneumonia in immunosuppressed patients is high (>80%). HSV has also been isolated from the lower respiratory tract of persons with adult respiratory distress syndrome and prolonged intubation. Most authorities

believe that the presence of HSV in tracheal aspirates in such settings is due to reactivation of HSV in the tracheal region and localized tracheitis in persons with long-standing intubation. Such patients should be evaluated for extension of HSV infection into the lung parenchyma. Controlled trials evaluating the role of antiviral agents used against HSV in morbidity and mortality associated with acute respiratory distress syndrome have not been conducted. The role of lower respiratory tract HSV infection in overall rates of morbidity and mortality associated with these conditions is unclear. HSV is an uncommon cause of hepatitis in immunocompetent patients. HSV infection of the liver is associated with fever, abrupt elevations of bilirubin and serum aminotransferase levels, and leukopenia (<4000 white blood cells/μL). Disseminated intravascular coagulation may also develop.

Other reported complications of HSV infection include monarticular arthritis, adrenal necrosis, idiopathic thrombocytopenia, and glomerulonephritis. Disseminated HSV infection in immunocompetent patients is rare. In immunocompromised, burned, or malnourished patients, HSV occasionally disseminates to other visceral organs, such as the adrenal glands, pancreas, small and large intestines, and bone marrow. Rarely, primary HSV infection in pregnancy disseminates and may be associated with the death of both mother and fetus. This uncommon event is usually related to the acquisition of primary infection in the third trimester. Disseminated HSV infection is best detected by the presence of HSV DNA in plasma or blood.

NEONATAL HSV INFECTIONS

Of all HSV-infected populations, neonates (infants younger than 6 weeks) have the highest frequency of visceral and/or CNS infection. Without therapy, the overall rate of death from neonatal herpes is 65%; <10% of neonates with CNS infection develop normally. Although skin lesions are the most commonly recognized features of disease, many infants do not develop lesions at all or do so only well into the course of disease. Neonatal infection is usually acquired perinatally from contact with infected genital secretions at delivery. Congenitally infected infants have been reported. In most series, 30% of neonatal HSV infections are due to HSV-1 and 70% to HSV-2. The risk of developing neonatal HSV infection is 10 times higher for an infant born to a mother who has recently acquired HSV than for other infants. Neonatal HSV-1 infections may also be acquired through postnatal contact with immediate family members who have symptomatic or asymptomatic oral-labial HSV-1 infection or through nosocomial transmission within the hospital. All neonates with presumed herpes should be treated with IV acyclovir. Antiviral chemotherapy with high-dose IV acyclovir (60 mg/kg per day) has reduced the mortality rate from neonatal herpes to ~15%. However, rates of morbidity, especially among infants with HSV-2 infection involving the CNS, are still very high.

HSV IN PREGNANCY

In the United States, 22% of all pregnant women and 55% of non-Hispanic black pregnant women are seropositive for HSV-2. However, the risk of mother-to-child transmission of HSV in the perinatal period is highest when the infection is acquired near the time of labor—that is, in previously HSV-seronegative women. The clinical manifestations of recurrent genital herpes—including the frequency of subclinical versus clinical infection, duration of lesions, pain, and constitutional symptoms—are similar in pregnant and nonpregnant women. Recurrences increase in frequency over the course of pregnancy. However, when women are seropositive for HSV-2 at the outset of pregnancy, no effect on neonatal outcomes (including birth weight and gestational age) is seen. First-episode infections in pregnancy have more severe consequences for mother and infant. Maternal visceral dissemination during the third trimester occasionally occurs, as does premature birth or intrauterine growth retardation. The acquisition of primary disease in pregnancy, whether related to HSV-1 or HSV-2, carries the risk of transplacental transmission of virus to the neonate and can result in spontaneous abortion, although this outcome is relatively uncommon. Most authorities recommend antiviral treatment for newly acquired genital HSV infection during pregnancy with acyclovir (400 mg three times daily) or valacyclovir (500–1000 mg twice daily) administered for 7–10 days. However, the impact of this intervention on transmission is unknown. The high HSV-2 prevalence rate in pregnancy and the low incidence of neonatal disease (1 case per 6000–20,000 live births) indicate that only a few infants are at risk of acquiring HSV. Therefore, cesarean section is not warranted for all women with recurrent genital disease. Because intrapartum transmission of infection accounts for the majority of cases, only women who are shedding HSV at delivery need be considered for abdominal delivery. Several studies have shown no correlation between recurrence of viral shedding before delivery and viral shedding at term. Hence, weekly virologic monitoring and amniocentesis are not recommended.

The frequency of transmission from mother to infant is markedly higher among women who acquire HSV near term (30–50%) than among those in whom HSV-2 infection is reactivated at delivery (<1%). Although maternal antibody to HSV-2 is protective, antibody to HSV-1 offers little or no protection against neonatal HSV-2 infection. Primary genital infection with HSV-1 leads to a particularly high risk of transmission during pregnancy and accounts for an increasing proportion of neonatal HSV cases. Moreover, during reactivation, HSV-1 appears more transmissible to the neonate than HSV-2. Only 2% of women who are seropositive for HSV-2 have HSV-2 isolated from cervical secretions at delivery, and only 1% of infants exposed in this manner develop infection, presumably because of the protective effects of maternally transferred antibodies and perhaps lower viral titers during reactivation. Despite the low frequency of transmission of HSV in this setting, 30–50% of infants with neonatal HSV are

born to mothers with established genital herpes.

Isolation of HSV by cervicovaginal swab at the time of delivery is the greatest risk factor for intrapartum HSV transmission (relative risk = 346); however, culture-negative, PCR-positive cases of intrapartum transmission are well described. New acquisition of HSV [odds ratio (OR) = 49], isolation of HSV-1 versus HSV-2 (OR = 35), cervical versus vulvar HSV detection (OR = 15), use of fetal scalp electrodes (OR = 3.5), and young age confer further risk of transmission, whereas abdominal delivery is protective (OR = 0.14). Physical examination poorly predicts the absence of shedding, and PCR far exceeds culture in terms of sensitivity and speed. Therefore, PCR detection at the onset of labor should be used to aid clinical decision-making for women with HSV-2 antibody. Because cesarean section appears to be an effective means of reducing maternal-fetal transmission, patients with recurrent genital herpes should be encouraged to come to the hospital early at the time of delivery for careful examination of the external genitalia and cervix as well as collection of a swab sample for viral isolation. Women who have no evidence of lesions should have a vaginal delivery. The presence of active lesions on the cervix or external genitalia is an indication for abdominal delivery.

If first-episode exposure has occurred (e.g., if HSV serologies show that the mother is seronegative or if the mother is HSV-1- seropositive and the isolate at delivery is found to be HSV-2), many authorities would initiate antiviral therapy for the infant with IV acyclovir. At a minimum, samples for viral cultures and PCR should be obtained from the throat, nasopharynx, eyes, and rectum of these infants immediately and at 5- to 10-day intervals. Lethargy, skin lesions, or fever should be evaluated promptly. All infants from whom HSV is isolated 24 h after delivery should be treated with IV acyclovir at recommended doses.

Diagnosis

Both clinical and laboratory criteria are useful for diagnosing HSV infections. A clinical diagnosis can be made accurately when characteristic multiple vesicular lesions on an erythematous base are present. However, herpetic ulcerations may resemble skin ulcerations of other etiologies. Mucosal HSV infections may also present as urethritis or pharyngitis without cutaneous lesions. Thus, laboratory studies to confirm the diagnosis and to guide therapy are recommended. While staining of scrapings from the base of the lesions with Wright's, Giemsa's (Tzanck preparation), or Papanicolaou's stain to detect giant cells or intranuclear inclusions of *Herpesvirus* infection is a well-described procedure, few clinicians are skilled in these techniques, the sensitivity of staining is low (<30% for mucosal swabs), and these cytologic methods do not differentiate between HSV and VZV infections.

HSV infection is best confirmed in the laboratory by detection of virus, viral antigen, or viral DNA in scrapings from lesions. HSV DNA detection by PCR is the most sensitive laboratory technique for detecting mucosal or visceral HSV infections and should be utilized when available. HSV causes a discernible cytopathic effect in a variety of cell culture systems, and this effect can be identified within 48–96 h after inoculation. Spin-amplified culture with subsequent staining for HSV antigen has shortened the time needed to identify HSV to <24 h. The sensitivity of all detection methods depends on the stage of the lesions (with higher sensitivity for vesicular than for ulcerative lesions), on whether the patient has a first or a recurrent episode of the disease (with higher sensitivity in first than in recurrent episodes), and on whether the sample is from an immunosuppressed or an immunocompetent patient (with more antigen or DNA in immunosuppressed patients). Laboratory confirmation permits subtyping of the virus; information on subtype may be useful epidemiologically and may help to predict the frequency of reactivation after first-episode oral-labial or genital HSV infection.

Acute- and convalescent-phase serum can be useful in demonstrating seroconversion during primary HSV-1 or HSV-2 infection. However, few available tests report titers, and increases in index values do not reflect first episodes in all patients. Serologic assays based on type-specific proteins should be used to identify asymptomatic carriers of HSV-1 or HSV-2. No reliable IgM method for defining acute HSV infection is available.

Several studies have shown that persons with previously unrecognized HSV-2 infection can be taught to identify symptomatic reactivations. Individuals seropositive for HSV-2 should be told about the high frequency of subclinical reactivation in mucosal surfaces that are not visible to the eye (e.g., cervix, urethra, perianal skin) or in microscopic ulcerations that may not be clinically symptomatic. Transmission of infection during such episodes is well established. HSV-2-seropositive persons should be educated about the high likelihood of subclinical shedding and the role condoms (male or female) may play in reducing transmission. Antiviral therapy with valacyclovir (500 mg once daily) has been shown to reduce the transmission of HSV-2 between sexual partners.

Treatment: Herpes Simplex Virus Infections

Many aspects of mucocutaneous and visceral HSV infections are amenable to antiviral chemotherapy. For mucocutaneous infections, acyclovir and its congeners famciclovir and valacyclovir have been the mainstays of therapy. Several antiviral agents are available for topical use in HSV eye infections: idoxuridine, trifluorothymidine, topical vidarabine, and cidofovir. For HSV encephalitis and neonatal herpes, IV acyclovir is the treatment of choice.

All licensed antiviral agents for use against HSV inhibit the viral DNA polymerase. One class of drugs, typified by the drug acyclovir, is made up of substrates for the HSV enzyme thymidine kinase (TK). Acyclovir, ganciclovir, famciclovir, and valacyclovir are all selectively phosphorylated to the monophosphate form in virus-infected cells. Cellular enzymes convert the monophosphate form of the drug to the triphosphate, which is then incorporated into the viral DNA chain. Acyclovir is the agent most frequently used for the treatment of HSV infections and is available in IV, oral, and topical formulations. Valacyclovir, the valyl ester of acyclovir, offers greater bioavailability than acyclovir and thus can be administered less frequently. Famciclovir, the oral formulation of penciclovir, is clinically effective in the treatment of a variety of HSV-1 and HSV-2 infections. Ganciclovir is active against both HSV-1 and HSV-2; however, it is more toxic than acyclovir, valacyclovir, and famciclovir and generally is not recommended for the treatment of HSV infections. Anecdotal case reports suggest that ganciclovir may also be less effective than acyclovir for treatment of HSV infections. All three recommended compounds—acyclovir, valacyclovir, and famciclovir—have proved effective in shortening the duration of symptoms and lesions of mucocutaneous HSV infections in both immunocompromised and immunocompetent patients **(Table 179-1)**. IV and oral formulations prevent reactivation of HSV in seropositive immunocompromised patients during induction chemotherapy or in the period immediately after bone marrow or solid organ transplantation. Chronic daily suppressive therapy reduces the frequency of reactivation disease among patients with frequent genital or oral-labial herpes. Only valacyclovir has been subjected to clinical trials that demonstrated reduced transmission of HSV-2 infection between sexual partners. IV acyclovir (30 mg/kg per day, given as a 10-mg/kg infusion over 1 h at 8-h intervals) is effective in reducing rates of death and morbidity from HSV encephalitis. Early initiation of therapy is a critical factor in outcome. The major side effect associated with IV acyclovir is transient renal insufficiency, usually due to crystallization of the compound in the renal parenchyma. This adverse reaction can be avoided if the medication is given slowly over 1 h and the patient is well hydrated. Because CSF levels of acyclovir average only 30 –50% of plasma levels, the dosage of acyclovir used for treatment of CNS infection (30 mg/kg per day) is double that used for treatment of mucocutaneous or visceral disease (15 mg/kg per day). Even higher doses of IV acyclovir are used for neonatal HSV infection (60 mg/kg per day in three divided doses).

Table 179-1 Antiviral Chemotherapy for HSV Infection

I. Mucocutaneous HSV infections
A. Infections in immunosuppressed patients
1. Acute symptomatic first or recurrent episodes: IV acyclovir (5 mg/kg q8h) or oral acyclovir (400 mg qid), famciclovir (500 mg bid or tid), or valacyclovir (500 mg bid) is effective. Treatment duration may vary from 7 to 14 days.
2. Suppression of reactivation disease (genital or oral-labial): IV acyclovir (5 mg/kg q8h) or oral valacyclovir (500 mg bid) or acyclovir (400–800 mg 3–5 times per day) prevents recurrences during the 30-day period immediately after transplantation. Longer-term HSV suppression is often used for persons with continued immunosuppression. In bone marrow and renal transplant recipients, oral valacyclovir (2 g/d) is also effective in reducing cytomegalovirus infection. Oral valacyclovir at a dose of 4 g/d has been associated with thrombotic thrombocytopenic purpura after extended use in HIV-positive persons. In HIV-infected persons, oral acyclovir (400–800 mg bid), valacyclovir (500 mg bid), or famciclovir (500 mg bid) is effective in reducing clinical and subclinical reactivations of HSV-1 and HSV-2.
B. Infections in immunocompetent patients
1. Genital herpes
a. First episodes: Oral acyclovir (200 mg 5 times per day or 400 mg tid), valacyclovir (1 g bid), or famciclovir (250 mg bid) for 7–14 days is effective. IV acyclovir (5 mg/kg q8h for 5 days) is given for severe disease or neurologic complications such as aseptic meningitis.
b. Symptomatic recurrent genital herpes: Short-course (1- to 3-day) regimens are preferred because of low cost, likelihood of adherence, and convenience. Oral acyclovir (800 mg tid for 2 days), valacyclovir (500 mg bid for 3 days), or famciclovir (750 or 1000 mg bid for 1 day, a 1500-mg single dose, or 500 mg stat followed by 250 mg q12h for 3 days) effectively shortens lesion duration. Other options include oral acyclovir (200 mg 5 times per day), valacyclovir (500 mg bid), and famciclovir (125 mg bid for 5 days). *c. Suppression of recurrent genital herpes:* Oral acyclovir (400–800 mg bid) or valacyclovir (500 mg daily) is given. Patients with >9 episodes per year should take oral valacyclovir (1 g daily or 500 mg bid) or famciclovir (250 mg bid or 500 mg bid).
2. Oral-labial HSV infections
a. First episode: Oral acyclovir (200 mg) is given 4 or 5 times per day; an oral acyclovir suspension can be used (600 mg/m^2 qid). Oral famciclovir (250 mg bid) or valacyclovir (1 g bid) has been used clinically.

b. Recurrent episodes: If initiated at the onset of the prodrome, single-dose or 1-day therapy effectively reduces pain and speeds healing. Regimens include oral famciclovir (a 1500-mg single dose or 750 mg bid for 1 day) or valacyclovir (a 2-g single dose or 2 g bid for 1 day). Self-initiated therapy with 6-times-daily topical penciclovir cream effectively speeds healing of oral-labial HSV. Topical acyclovir cream has also been shown to speed healing.
c. Suppression of reactivation of oral-labial HSV: If started before exposure and continued for the duration of exposure (usually 5–10 days), oral acyclovir (400 mg bid) prevents reactivation of recurrent oral-labial HSV infection associated with severe sun exposure.
3. Surgical prophylaxis of oral or genital HSV infection: Several surgical procedures, such as laser skin resurfacing, trigeminal nerve-root decompression, and lumbar disk surgery, have been associated with HSV reactivation. IV acyclovir (3–5 mg/kg q8h) or oral acyclovir (800 mg bid), valacyclovir (500 mg bid), or famciclovir (250 mg bid) effectively reduces reactivation. Therapy should be initiated 48 h before surgery and continued for 3–7 days.
4. Herpetic whitlow: Oral acyclovir (200 mg; alternative: 400 mg tid) is given 5 times daily for 7–10 days.
5. HSV proctitis: Oral acyclovir (400 mg 5 times per day) is useful in shortening the course of infection. In immunosuppressed patients or in patients with severe infection, IV acyclovir (5 mg/kg q8h) may be useful.
6. Herpetic eye infections: In acute keratitis, topical trifluorothymidine, vidarabine, idoxuridine, acyclovir, penciclovir, and interferon are all beneficial. Debridement may be required. Topical steroids may worsen disease.
II. CNS HSV infections
A. HSV encephalitis: IV acyclovir (10 mg/kg q8h; 30 mg/kg per day) is given for 10 days or until HSV DNA is no longer detected in CSF.
B. HSV aseptic meningitis: No studies of systemic antiviral chemotherapy exist. If therapy is to be given, IV acyclovir (15–30 mg/kg per day) should be used.
C. Autonomic radiculopathy: No studies are available. Most authorities recommend a trial of IV acyclovir.
III. Neonatal HSV infections: Oral acyclovir (60 mg/kg per day, divided into 3 doses) is given. The recommended duration of treatment is 21 days. Monitoring for relapse should be undertaken, and some authorities recommend continued suppression with oral acyclovir suspension for 3–4 months.
IV. Visceral HSV infections
A. HSV esophagitis: IV acyclovir (15 mg/kg per day). In some patients with milder forms of immunosuppression, oral therapy with valacyclovir or famciclovir is effective.
B. HSV pneumonitis: No controlled studies exist. IV acyclovir (15 mg/kg per day) should be considered.
V. Disseminated HSV infections: No controlled studies exist. IV acyclovir (5 mg/kg q8h) should be tried. Adjustments for renal insufficiency may be needed. No definite evidence indicates that therapy will decrease the risk of death.
VI. Erythema multiforme associated with HSV: Anecdotal observations suggest that oral acyclovir (400 mg bid or tid) or valacyclovir (500 mg bid) will suppress erythema multiforme.
VII. Infections due to acyclovir-resistant HSV: IV foscarnet (40 mg/kg IV q8h) should be given until lesions heal. The optimal duration of therapy and the usefulness of its continuation to suppress lesions are unclear. Some patients may benefit from cutaneous application of trifluorothymidine or 5% cidofovir gel.

Increasingly, shorter courses of therapy are being used for treatment of recurrent mucocutaneous infection with HSV-1 or HSV-2 in immunocompetent patients. One-day courses of famciclovir and valacyclovir are clinically effective, more convenient, and generally less costly than longer courses of therapy (Table 179-1). These short-course regimens should be reserved for immunocompetent hosts.

SUPPRESSION OF MUCOCUTANEOUS HERPES

Recognition of the high frequency of subclinical reactivation provides a well-accepted rationale for the use of daily antiviral therapy to suppress reactivations of HSV, especially in persons with frequent clinical reactivations (e.g., those with recently acquired genital HSV infection). Immunosuppressed persons, including those with HIV infection, may also benefit from daily antiviral therapy. Recent studies have shown the efficacy of daily acyclovir and valacyclovir in reducing the frequency of HSV reactivations among HIV-positive persons. Regimens used include acyclovir (400–800 mg twice daily), famciclovir (500 mg twice daily), and valacyclovir (500 mg twice daily); valacyclovir at a dose of 4 g daily was associated with thrombotic thrombocytopenic purpura in one study of HIV-infected persons. In addition, daily treatment of HSV-2 reduces the titer of HIV RNA in plasma (0.5-log reduction) and in genital mucosa (0.33-log reduction).

REDUCED HSV TRANSMISSION TO SEXUAL PARTNERS

Once-daily valacyclovir (500 mg) has been shown to reduce transmission of HSV-2 between sexual

partners. Transmission rates are higher from males to females and among persons with frequent HSV-2 reactivation. Serologic screening can be used to identify at-risk couples. Daily valacyclovir appears to be more effective at reducing subclinical shedding than daily famciclovir.

ACYCLOVIR RESISTANCE

Acyclovir-resistant strains of HSV have been identified. Most of these strains have an altered substrate specificity for phosphorylating acyclovir. Thus, cross-resistance to famciclovir and valacyclovir is usually found. Occasionally, an isolate with altered TK specificity arises and is sensitive to famciclovir but not to acyclovir. In some patients infected with TK-deficient virus, higher doses of acyclovir are associated with clearing of lesions. In others, clinical disease progresses despite high-dose therapy. Almost all clinically significant acyclovir resistance has been seen in immunocompromised patients, and HSV-2 isolates are more often resistant than HSV-1 strains. A study by the Centers for Disease Control and Prevention indicated that ~5% of HSV-2 isolates from HIV-positive persons exhibit some degree of in vitro resistance to acyclovir. Of HSV-2 isolates from immunocompetent patients attending sexually transmitted disease clinics, <0.5% show reduced in vitro sensitivity to acyclovir. The lack of appreciable change in the frequency of detection of such isolates in the past 20 years probably reflects the reduced transmission of TK-deficient mutants. Isolation of HSV from lesions persisting despite adequate dosages and blood levels of acyclovir should raise the suspicion of acyclovir resistance. Therapy with the antiviral drug foscarnet is useful in acyclovir-resistant cases (Chap. 178). Because of its toxicity and cost, this drug is usually reserved for patients with extensive mucocutaneous infections. Cidofovir is a nucleotide analogue and exists as a phosphonate or monophosphate form. Most TK-deficient strains of HSV are sensitive to cidofovir. Cidofovir ointment speeds healing of acyclovir-resistant lesions. No well-controlled trials of systemic cidofovir have been reported. True TK-negative variants of HSV appear to have a reduced capacity to spread because of altered neurovirulence—a feature important in the relatively infrequent presence of such strains in immunocompetent populations, even with increasing use of antiviral drugs.

Prevention

The success of efforts to control HSV disease on a population basis through suppressive antiviral chemotherapy and/or educational programs will be limited. Barrier forms of contraception (especially condoms) decrease the likelihood of transmission of HSV infection, particularly during periods of asymptomatic viral excretion. When lesions are present, HSV infection may be transmitted by skin-to-skin contact despite the use of a condom. Nevertheless, the available data suggest that consistent condom use is an effective means of reducing the risk of genital HSV-2 transmission. Chronic daily antiviral therapy with valacyclovir can also be partially effective in reducing acquisition of HSV-2, especially among susceptible women. There are no comparative efficacy studies of valacyclovir versus condom use. Most authorities suggest both approaches. The need for a vaccine to prevent acquisition of HSV infection is great, especially in light of the role HSV-2 plays in enhancing the acquisition and transmission of HIV-1.

A substantial portion of neonatal HSV cases could be prevented by reducing the acquisition of HSV by women in the third trimester of pregnancy. Neonatal HSV infection can result from either the acquisition of maternal infection near term or the reactivation of infection at delivery in the already-infected mother. Thus strategies for reducing neonatal HSV are complex. Some authorities have recommended that antiviral therapy with acyclovir or valacyclovir be given to HSV-2-infected women in late pregnancy as a means of reducing reactivation of HSV-2 at term. Data are not available to support the efficacy of this approach. Moreover, the high treatment-to-prevention ratio makes this a dubious public health approach, even though it can reduce the frequency of HSV-associated cesarean delivery.

FURTHER READINGS

Abu-Raddad LJ et al: Genital herpes has played a more important role than any other sexually transmitted infection in driving HIV prevalence in Africa. PLoS ONE 3:e2230, 2008

Chilukuri S, Rosen T: Management of acyclovir-resistant herpes simplex virus. Dermatol Clin 21:311, 2003[PMID: 12757254]

Corey L, Wald A: Maternal and neonatal herpes simplex virus infections. N Engl J Med 361:1376, 2009[PMID: 19797284]

James SH et al: Antiviral therapy for herpesvirus central nervous system infections: Neonatal herpes simplex virus infection, herpes simplex encephalitis, and congenital cytomegalovirus infection. Antiviral Res 83:207, 2009[PMID: 19414035]

Mark KE et al: Rapidly cleared episodes of herpes simplex virus reactivation in immunocompetent adults. J Infect Dis 198:1141, 2008[PMID: 18783315]

Zhu J et al: Persistence of HIV-1 receptor-positive cells after HSV-2 reactivation is a potential mechanism for increased HIV-1 acquisition. Nat Med 15:886: 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VARICELLA-ZOSTER VIRUS INFECTIONS: INTRODUCTION

Definition

Varicella-zoster virus (VZV) causes two distinct clinical entities: varicella (chickenpox) and herpes zoster (shingles). Chickenpox, a ubiquitous and extremely contagious infection, is usually a benign illness of childhood characterized by an exanthematous vesicular rash. With reactivation of latent VZV (which is most common after the sixth decade of life), herpes zoster presents as a dermatomal vesicular rash, usually associated with severe pain.

Etiology

A clinical association between varicella and herpes zoster has been recognized for nearly 100 years. Early in the twentieth century, similarities in the histopathologic features of skin lesions resulting from varicella and herpes zoster were demonstrated. Viral isolates from patients with chickenpox and herpes zoster produced similar alterations in tissue culture—specifically, the appearance of eosinophilic intranuclear inclusions and multinucleated giant cells. These results suggested that the viruses were biologically similar. Restriction endonuclease analyses of viral DNA from a patient with chickenpox who subsequently developed herpes zoster verified the molecular identity of the two viruses responsible for these different clinical presentations.

VZV is a member of the family Herpesviridae, sharing with other members such structural characteristics as a lipid envelope surrounding a nucleocapsid with icosahedral symmetry, a total diameter of ~180–200 nm, and centrally located double-stranded DNA that is ~125,000 bp in length.

Pathogenesis and Pathology

PRIMARY INFECTION

Transmission occurs readily by the respiratory route; the subsequent localized replication of the virus at an undefined site (presumably the nasopharynx) leads to seeding of the reticuloendothelial system and ultimately to the development of viremia. Viremia in patients with chickenpox is reflected in the diffuse and scattered nature of the skin lesions and can be verified in selected cases by the recovery of VZV from the blood or routinely by the detection of viral DNA in either blood or lesions by polymerase chain reaction (PCR). Vesicles involve the corium and dermis, with degenerative changes characterized by ballooning, the presence of multinucleated giant cells, and eosinophilic intranuclear inclusions. Infection may involve localized blood vessels of the skin, resulting in necrosis and epidermal hemorrhage. With the evolution of disease, the vesicular fluid becomes cloudy because of the recruitment of polymorphonuclear leukocytes and the presence of degenerated cells and fibrin. Ultimately, the vesicles either rupture and release their fluid (which includes infectious virus) or are gradually reabsorbed.

RECURRENT INFECTION

The mechanism of reactivation of VZV that results in herpes zoster is unknown. Presumably, the virus infects dorsal root ganglia during chickenpox, where it remains latent until reactivated. Histopathologic examination of representative dorsal root ganglia during active herpes zoster demonstrates hemorrhage, edema, and lymphocytic infiltration.

Active replication of VZV in other organs, such as the lung or the brain, can occur during either chickenpox or herpes zoster but is uncommon in the immunocompetent host. Pulmonary involvement is characterized by interstitial pneumonitis, multinucleated giant cell formation, intranuclear inclusions, and pulmonary hemorrhage. Central nervous system (CNS) infection leads to histopathologic evidence of perivascular cuffing similar to that encountered in measles and other viral encephalitides. Focal hemorrhagic necrosis of the brain, characteristic of herpes simplex virus (HSV) encephalitis, is uncommon in VZV infection.

Epidemiology and Clinical Manifestations

CHICKENPOX

Humans are the only known reservoir for VZV. Chickenpox is highly contagious, with an attack rate of at least 90% among susceptible (seronegative) individuals. Persons of both sexes and all races are infected equally. The virus is endemic in the population at large; however, it becomes epidemic among susceptible individuals during seasonal peaks—namely, late winter and early spring in the temperate zone. Much of our knowledge of the disease's natural history and incidence predates the licensure of

the chickenpox vaccine in 1995. Historically, children 5–9 years old are most commonly affected and account for 50% of all cases. Most other cases involve children 1–4 and 10–14 years old. Approximately 10% of the population of the United States over the age of 15 is susceptible to infection. VZV vaccination during the second year of life has dramatically changed the epidemiology of infection, causing a significant decrease in the annualized incidence of chickenpox.

The incubation period of chickenpox ranges from 10–21 days but is usually 14–17 days. Secondary attack rates in susceptible siblings within a household are 70–90%. Patients are infectious ~48 h before onset of the vesicular rash, during the period of vesicle formation (which generally lasts 4–5 days), and until all vesicles are crusted.

Clinically, chickenpox presents with a rash, low-grade fever, and malaise, although a few patients develop a prodrome 1–2 days before onset of the exanthem. In the immunocompetent patient, chickenpox is usually a benign illness associated with lassitude and with body temperatures of 37.8°–39.4°C (100°–103°F) of 3–5 days' duration. The skin lesions—the hallmark of the infection—include maculopapules, vesicles, and scabs in various stages of evolution **(Fig. 180-1)**. These lesions, which evolve from maculopapules to vesicles over hours to days, appear on the trunk and face and rapidly spread to involve other areas of the body. Most are small and have an erythematous base with a diameter of 5–10 mm. Successive crops appear over a 2- to 4-day period. Lesions can also be found on the mucosa of the pharynx and/or the vagina. Their severity varies from one person to another. Some individuals have very few lesions, while others have as many as 2000. Younger children tend to have fewer vesicles than older individuals. Secondary and tertiary cases within families are associated with a relatively large number of vesicles. Immunocompromised patients—both children and adults, particularly those with leukemia—have lesions (often with a hemorrhagic base) that are more numerous and take longer to heal than those of immunocompetent patients. Immunocompromised individuals are also at greater risk for visceral complications, which occur in 30–50% of cases and are fatal 15% of the time in the absence of antiviral therapy.

Figure 180-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Varicella lesions at various stages of evolution: vesicles on an erythematous base, umbilical vesicles, and crusts.

The most common infectious complication of varicella is secondary bacterial superinfection of the skin, which is usually caused by *Streptococcus pyogenes* or *Staphylococcus aureus,* including strains that are methicillin-resistant. Skin infection results from excoriation of lesions after scratching. Gram's staining of skin lesions should help clarify the etiology of unusually erythematous and pustulated lesions.

The most common extracutaneous site of involvement in children is the CNS. The syndrome of acute cerebellar ataxia and meningeal inflammation generally appears ~21 days after onset of the rash and rarely develops in the preeruptive phase. The cerebrospinal fluid (CSF) contains lymphocytes and

elevated levels of protein. CNS involvement is a benign complication of VZV infection in children and generally does not require hospitalization. Aseptic meningitis, encephalitis, transverse myelitis, and Guillain-Barré syndrome can also occur. Reye's syndrome has been reported in children concomitantly treated with aspirin. Encephalitis is reported in 0.1–0.2% of children with chickenpox. Other than supportive care, no specific therapy (e.g., acyclovir administration) has proved efficacious for patients with CNS involvement.

Varicella pneumonia, the most serious complication following chickenpox, develops more commonly in adults (up to 20% of cases) than in children and is particularly severe in pregnant women. Pneumonia due to VZV usually has its onset 3–5 days into the illness and is associated with tachypnea, cough, dyspnea, and fever. Cyanosis, pleuritic chest pain, and hemoptysis are common. Roentgenographic evidence of disease consists of nodular infiltrates and interstitial pneumonitis. Resolution of pneumonitis parallels improvement of the skin rash; however, patients may have persistent fever and compromised pulmonary function for weeks.

Other complications of chickenpox include myocarditis, corneal lesions, nephritis, arthritis, bleeding diatheses, acute glomerulonephritis, and hepatitis. Hepatic involvement, distinct from Reye's syndrome and usually asymptomatic, is common in chickenpox and is generally characterized by elevated levels of liver enzymes, particularly aspartate and alanine aminotransferases.

Perinatal varicella is associated with a high mortality rate when maternal disease develops within 5 days before delivery or within 48 h thereafter. Because the newborn does not receive protective transplacental antibodies and has an immature immune system, the illness may be unusually severe. The reported mortality rate is as high as 30% in this group. *Congenital varicella*, with clinical manifestations of limb hypoplasia, cicatricial skin lesions, and microcephaly at birth, is extremely uncommon.

HERPES ZOSTER

Herpes zoster (shingles) is a sporadic disease that results from reactivation of latent VZV from dorsal root ganglia. Most patients with shingles have no history of recent exposure to other individuals with VZV infection. Herpes zoster occurs at all ages, but its incidence is highest (5–10 cases per 1000 persons) among individuals in the sixth decade of life and beyond. Data suggest that 1.2 million cases occur annually in the United States. Recurrent herpes zoster is exceedingly rare except in immunocompromised hosts, especially those with AIDS.

Herpes zoster is characterized by a unilateral vesicular dermatomal eruption, often associated with severe pain. The dermatomes from T3 to L3 are most frequently involved. If the ophthalmic branch of the trigeminal nerve is involved, *zoster ophthalmicus* results. The factors responsible for the reactivation of VZV are not known. In children, reactivation is usually benign; in adults, it can be debilitating because of pain. The onset of disease is heralded by pain within the dermatome, which may precede lesions by 48–72 h; an erythematous maculopapular rash evolves rapidly into vesicular lesions **(Fig. 180-2)**. In the normal host, these lesions may remain few in number and continue to form for only 3–5 days. The total duration of disease is generally 7–10 days; however, it may take as long as 2–4 weeks for the skin to return to normal. Patients with herpes zoster can transmit infection to seronegative individuals, with consequent chickenpox. In a few patients, characteristic localization of pain to a dermatome with serologic evidence of herpes zoster has been reported in the absence of skin lesions, an entity known as *zoster sine herpetica*. When branches of the trigeminal nerve are involved, lesions may appear on the face, in the mouth, in the eye, or on the tongue. *Zoster ophthalmicus* is usually a debilitating condition that can result in blindness in the absence of antiviral therapy. In the *Ramsay Hunt syndrome*, pain and vesicles appear in the external auditory canal, and patients lose their sense of taste in the anterior two-thirds of the tongue while developing ipsilateral facial palsy. The geniculate ganglion of the sensory branch of the facial nerve is involved.

Figure 180-2

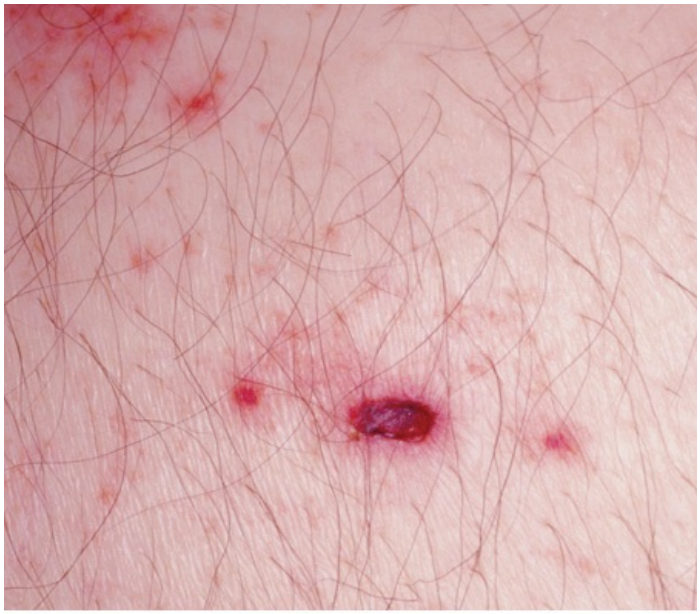

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Close-up of lesions of disseminated zoster. Note lesions at different stages of evolution, including pustules and crusting. *(Photo courtesy of Lindsey Baden; with permission .)*

In both normal and immunocompromised hosts, the most debilitating complication of herpes zoster is pain associated with acute neuritis and postherpetic neuralgia. Postherpetic neuralgia is uncommon in young individuals; however, at least 50% of zoster patients over age 50 report some degree of pain in the involved dermatome months after the resolution of cutaneous disease. Changes in sensation in the dermatome, resulting in either hypo- or hyperesthesia, are common.

CNS involvement may follow localized herpes zoster. Many patients without signs of meningeal irritation have CSF pleocytosis and moderately elevated levels of CSF protein. Symptomatic meningoencephalitis is characterized by headache, fever, photophobia, meningitis, and vomiting. A rare manifestation of CNS involvement is granulomatous angiitis with contralateral hemiplegia, which can be diagnosed by cerebral arteriography. Other neurologic manifestations include transverse myelitis with or without motor paralysis.

Like chickenpox, herpes zoster is more severe in immunocompromised than immunocompetent individuals. Lesions continue to form for >1 week, and scabbing is not complete in most cases until 3 weeks into the illness. Patients with Hodgkin's disease and non-Hodgkin's lymphoma are at greatest risk for progressive herpes zoster. Cutaneous dissemination **(Fig. 180-3)** develops in ~40% of these patients. Among patients with cutaneous dissemination, the risk of pneumonitis, meningoencephalitis, hepatitis, and other serious complications is increased by 5–10%. However, even in immunocompromised patients, disseminated zoster is rarely fatal.

Figure 180-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Herpes zoster in an HIV-infected patient is seen as hemorrhagic vesicles and pustules on an erythematous base grouped in a dermatomal distribution.

Recipients of hematopoietic stem cell transplants are at particularly high risk of VZV infection. Of all cases of posttransplantation VZV infection, 30% occur within 1 year (50% of these within 9 months); 45% of the patients involved have cutaneous or visceral dissemination. The mortality rate in this situation is 10%. Postherpetic neuralgia, scarring, and bacterial superinfection are especially common in VZV infections occurring within 9 months of transplantation. Among infected patients, concomitant graft-versus-host disease increases the chance of dissemination and/or death.

Differential Diagnosis

(See also Chap. e7) The diagnosis of chickenpox is not difficult. The characteristic rash and a history of recent exposure should lead to a prompt diagnosis. Other viral infections that can mimic chickenpox include disseminated HSV infection in patients with atopic dermatitis and the disseminated vesiculopapular lesions sometimes associated with coxsackievirus infection, echovirus infection, or atypical measles. However, these rashes are more commonly morbilliform with a hemorrhagic component rather than vesicular or vesiculopustular. Rickettsialpox (Chap. 174) can be confused with chickenpox; however, rickettsialpox can be distinguished easily by detection of the "herald spot" at the site of the mite bite and the development of a more pronounced headache. Serologic testing is also useful in differentiating rickettsialpox from varicella and can confirm susceptibility in adults unsure of their chickenpox history. Concern about smallpox has recently increased because of the threat of bioterrorism (Chap. 221). The lesions of smallpox are larger than those of chickenpox and are all at the same stage of evolution at any given time.

Unilateral vesicular lesions in a dermatomal pattern should lead rapidly to the diagnosis of herpes zoster, although the occurrence of shingles without a rash has been reported. Both HSV and coxsackievirus infections can cause dermatomal vesicular lesions. Supportive diagnostic virology and fluorescent staining of skin scrapings with monoclonal antibodies are helpful in ensuring the proper diagnosis. In the prodromal stage of herpes zoster, the diagnosis can be exceedingly difficult and may be made only after lesions have appeared or by retrospective serologic assessment.

Laboratory Findings

Unequivocal confirmation of the diagnosis is possible only through the isolation of VZV in susceptible tissue-culture cell lines, the demonstration of either seroconversion or a fourfold or greater rise in antibody titer between acute-phase and convalescent-phase serum specimens, or the detection of VZV DNA by PCR. A rapid impression can be obtained by a Tzanck smear, with scraping of the base of the lesions in an attempt to demonstrate multinucleated giant cells; however, the sensitivity of this method is low (~60%). PCR technology for the detection of viral DNA in vesicular fluid is available in a limited number of diagnostic laboratories. Direct immunofluorescent staining of cells from the lesion base or detection of viral antigens by other assays (such as the immunoperoxidase assay) is also useful, although these tests are not commercially available. The most frequently employed serologic

tools for assessing host response are the immunofluorescent detection of antibodies to VZV membrane antigens, the fluorescent antibody to membrane antigen (FAMA) test, immune adherence hemagglutination, and enzyme-linked immunosorbent assay (ELISA). The FAMA test and the ELISA appear to be most sensitive.

Treatment: Varicella-Zoster Virus Infections

Medical management of chickenpox in the immunologically normal host is directed toward the prevention of avoidable complications. Obviously, good hygiene includes daily bathing and soaks. Secondary bacterial infection of the skin can be avoided by meticulous skin care, particularly with close cropping of fingernails. Pruritus can be decreased with topical dressings or the administration of antipruritic drugs. Tepid water baths and wet compresses are better than drying lotions for the relief of itching. Administration of aspirin to children with chickenpox should be avoided because of the association of aspirin derivatives with the development of Reye's syndrome. Acyclovir (800 mg by mouth five times daily), valacyclovir (1 g three times daily), or famciclovir (250 mg three times daily) for 5–7 days is recommended for adolescents and adults with chickenpox of ≥24 h duration. (Valacyclovir is licensed for use in children and adolescents. Famciclovir is recommended but not licensed for varicella.) Likewise, acyclovir therapy may be of benefit to children <12 years of age if initiated early in the disease (<24 h) at a dose of 20 mg/kg every 6 h. The advantages (i.e., pharmacokinetics) of the second-generation agents valacyclovir and famciclovir are described in Chap. 178.

Aluminum acetate soaks for the management of herpes zoster can be both soothing and cleansing. Patients with herpes zoster benefit from oral antiviral therapy, as evidenced by accelerated healing of lesions and resolution of zoster-associated pain with acyclovir, valacyclovir, or famciclovir. Acyclovir, now off patent, is administered at a dosage of 800 mg five times daily for 7–10 days. Famciclovir, the prodrug of penciclovir, is at least as effective as acyclovir and perhaps more so; the dose is 500 mg by mouth three times daily for 7 days. Valacyclovir, the prodrug of acyclovir that is now off patent, accelerates healing and resolution of zoster-associated pain more promptly than acyclovir. The dose is 1 g by mouth three times daily for 5–7 days. Compared with acyclovir, both famciclovir and valacyclovir offer the advantage of less frequent administration.

In severely immunocompromised hosts (e.g., transplant recipients, patients with lymphoproliferative malignancies), both chickenpox and herpes zoster (including disseminated disease) should be treated, at least at the outset, with IV acyclovir, which reduces the occurrence of visceral complications but has no effect on healing of skin lesions or pain. The dose is 10 mg/kg every 8 h for 7 days. For low-risk immunocompromised hosts, oral therapy with valacyclovir or famciclovir appears beneficial. If medically feasible, it is desirable to decrease immunosuppressive treatment concomitant with the administration of IV acyclovir.

Patients with varicella pneumonia often require ventilatory support. Persons with zoster ophthalmicus should be referred immediately to an ophthalmologist. Therapy for this condition consists of the administration of analgesics for severe pain and the use of atropine. Acyclovir, valacyclovir, and famciclovir all accelerate healing. Decisions about the use of glucocorticoids should be made by the ophthalmologist.

The management of acute neuritis and/or postherpetic neuralgia can be particularly difficult. In addition to the judicious use of analgesics, ranging from nonnarcotics to narcotic derivatives, drugs such as gabapentin, pregabalin, amitriptyline hydrochloride, lidocaine (patches), and fluphenazine hydrochloride are reportedly beneficial for pain relief. In one study, glucocorticoid therapy administered early in the course of localized herpes zoster significantly accelerated such quality-of-life improvements as a return to usual activity and termination of analgesic medications. The dose of prednisone administered orally was 60 mg/d on days 1–7, 30 mg/d on days 8–14, and 15 mg/d on days 15–21. This regimen is appropriate only for relatively healthy elderly persons with moderate or severe pain at presentation. Patients with osteoporosis, diabetes mellitus, glycosuria, or hypertension may not be appropriate candidates. Glucocorticoids should not be used without concomitant antiviral therapy.

Prevention

Three methods are used for the prevention of VZV infections. First, a live attenuated varicella vaccine (Oka) is recommended for all children >1 year of age (up to 12 years of age) who have not had chickenpox and for adults known to be seronegative for VZV. Two doses are recommended for all children: the first at 12–15 months of age and the second at ~4–6 years of age. VZV-seronegative persons >13 years of age should receive two doses of vaccine at least 1 month apart. The vaccine is both safe and efficacious. Breakthrough cases are mild and may result in spread of the vaccine virus to susceptible contacts. The universal vaccination of children is resulting in a decreased incidence of chickenpox in sentinel communities. Furthermore, inactivation of the vaccine virus significantly decreases the occurrence of herpes zoster after hematopoietic stem-cell transplantation. In individuals >60 years of age, a VZV vaccine with 18 times the viral content of the Oka vaccine decreased the incidence of shingles by 51%, the burden of illness by 61%, and the incidence of postherpetic neuralgia by 66%. The Advisory Committee on Immunization Practices has therefore recommended that persons in this age group be offered this vaccine in order to reduce the frequency of shingles and

the severity of postherpetic neuralgia.

A second approach is to administer varicella-zoster immune globulin (VZIG) to individuals who are susceptible, are at high risk for developing complications of varicella, and have had a significant exposure. This product should be given within 96 h (preferably within 72 h) of the exposure. Indications for administration of VZIG appear in **Table 180–1**. VZIG is available under an investigational new drug protocol from FFF Enterprises (800-843-7477).

Table 180–1 Recommendations for VZIG Administration

Exposure Criteria
1. Exposure to person with chickenpox or zoster
a. Household: residence in the same household
b. Playmate: face-to-face indoor play
c. Hospital
Varicella: same 2- to 4-bed room or adjacent beds in large ward, face-to-face contact with infectious staff member or patient, visit by a person deemed contagious
Zoster: intimate contact (e.g., touching or hugging) with a person deemed contagious
d. Newborn infant: onset of varicella in the mother ≤5 days before delivery or ≤48 h after delivery; VZIG not indicated if the mother has zoster
2. Patient should receive VZIG as soon as possible but not >96 h after exposure
Candidates (Provided They Have Significant Exposure) Include:
1. Immunocompromised susceptible children without a history of varicella or varicella immunization
2. Susceptible pregnant women
3. Newborn infants whose mother had onset of chickenpox within 5 days before or within 48 h after delivery
4. Hospitalized premature infant (≥28 weeks of gestation) whose mother lacks a reliable history of chickenpox or serologic evidence of protection against varicella
5. Hospitalized premature infant (<28 weeks of gestation or ≤1000-g birth weight), regardless of maternal history of varicella or varicella-zoster virus serologic status

Lastly, antiviral therapy can be given as prophylaxis to individuals at high risk who are ineligible for vaccine or who are beyond the 96-h window after direct contact. While the initial studies have used acyclovir, similar benefit can be anticipated with either valacyclovir or famciclovir. Therapy is instituted 7 days after intense exposure. At this time, the host is midway into the incubation period. This approach significantly decreases disease severity, if not totally preventing disease.

FURTHER READINGS

Arvin A: Aging, immunity, and the varicella-zoster virus. N Engl J Med 352:2266, 2005[PMID: 15930416]

Gershon AA, Gershon MD: Perspectives on vaccines against varicella-zoster virus infections. Curr Top Microbiol Immuno 342:359, 2010[PMID: 20232192]

———: Varicella vaccine. N Engl J Med 356:2648, 2007

Gnann JW, Whitley RJ: Herpes zoster. N Engl J Med 347:340, 2002[PMID: 12151472]

Kimberlin DW, Whitley RJ: Varicella-zoster vaccine for the prevention of herpes zoster. N Engl J Med 356:1338, 2007[PMID: 17392303]

Nguyen HQ et al: Decline in mortality due to varicella after implementation of varicella vaccination in the United States. N Engl J Med 352:450, 2005[PMID: 15689583]

Oxman MN et al: A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults. N Engl J Med 352:2271, 2005[PMID: 15930418]

Seward JF et al: Contagiousness of varicella in vaccinated cases: A household contact study. JAMA 292:704, 2004[PMID: 15304467]

——— et al: Varicella disease after introduction of varicella vaccine in the United States, 1995–2000. JAMA 28:606, 2002

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EPSTEIN-BARR VIRUS INFECTIONS, INCLUDING INFECTIOUS MONONUCLEOSIS: INTRODUCTION

Definition

Epstein-Barr virus (EBV) is the cause of heterophile-positive infectious mononucleosis (IM), which is characterized by fever, sore throat, lymphadenopathy, and atypical lymphocytosis. EBV is also associated with several human tumors, including nasopharyngeal carcinoma, Burkitt's lymphoma, Hodgkin's disease, and (in patients with immunodeficiencies) B cell lymphoma. The virus is a member of the family Herpesviridae. The two types of EBV that are widely prevalent in nature are not distinguishable by conventional serologic tests.

Epidemiology

EBV infections occur worldwide. These infections are most common in early childhood, with a second peak during late adolescence. By adulthood, more than 90% of individuals have been infected and have antibodies to the virus. IM is usually a disease of young adults. In lower socioeconomic groups and in areas of the world with deficient standards of hygiene (e.g., developing regions), EBV tends to infect children at an early age, and IM is uncommon. In areas with higher standards of hygiene, infection with EBV is often delayed until adulthood, and IM is more prevalent.

EBV is spread by contact with oral secretions. The virus is frequently transmitted from asymptomatic adults to infants and among young adults by transfer of saliva during kissing. Transmission by less intimate contact is rare. EBV has been transmitted by blood transfusion and by bone marrow transplantation. More than 90% of asymptomatic seropositive individuals shed the virus in oropharyngeal secretions. Shedding is increased in immunocompromised patients and those with IM.

Pathogenesis

EBV is transmitted by salivary secretions. The virus infects the epithelium of the oropharynx and the salivary glands and is shed from these cells. While B cells may become infected after contact with epithelial cells, studies suggest that lymphocytes in the tonsillar crypts can be infected directly. The virus then spreads through the bloodstream. The proliferation and expansion of EBV-infected B cells along with reactive T cells during IM result in enlargement of lymphoid tissue. Polyclonal activation of B cells leads to the production of antibodies to host-cell and viral proteins. During the acute phase of IM, up to 1 in every 100 B cells in the peripheral blood is infected by EBV; after recovery, 1–50 in every 1 million B cells is infected. During IM, there is an inverted CD4+/CD8+ T cell ratio. The percentage of CD4+ T cells decreases, while there are large clonal expansions of CD8+ T cells; up to 40% of CD8+ T cells are directed against EBV antigens during acute infection. Memory B cells, not epithelial cells, are the reservoir for EBV in the body. When patients are treated with acyclovir, shedding of EBV from the oropharynx stops but the virus persists in B cells.

The EBV receptor (CD21) on the surface of B cells is also the receptor for the C3d component of complement. EBV infection of epithelial cells results in viral replication and production of virions. When B cells are infected by EBV in vitro, they become transformed and can proliferate indefinitely. During latent infection of B cells, only the EBV nuclear antigens (EBNAs), latent membrane proteins (LMPs), and small EBV RNAs are expressed in vitro. EBV-transformed B cells secrete immunoglobulin; only a small fraction of cells produce virus.

Cellular immunity is more important than humoral immunity in controlling EBV infection. In the initial phase of infection, suppressor T cells, natural killer cells, and nonspecific cytotoxic T cells are important in controlling the proliferation of EBV-infected B cells. Levels of markers of T cell activation and serum interferon (IFN) γ are elevated. Later in infection, human leukocyte antigen–restricted cytotoxic T cells that recognize EBNAs and LMPs and destroy EBV-infected cells are generated. Studies have shown that one of the late proteins expressed during EBV replication, *BCRF1*, is a homologue of interleukin 10 and can inhibit the production of IFN-γ by mononuclear cells in vitro.

If T cell immunity is compromised, EBV-infected B cells may begin to proliferate. When EBV is associated with lymphoma, virus-induced proliferation is but one step in a multistep process of neoplastic transformation. In many EBV-containing tumors, LMP-1 mimics members of the tumor necrosis factor receptor family (e.g., CD40), transmitting growth-proliferating signals.

Clinical Manifestations

SIGNS AND SYMPTOMS

Most EBV infections in infants and young children either are asymptomatic or present as mild pharyngitis with or without tonsillitis. In contrast, up to 75% of infections in adolescents present as IM. IM in the elderly presents relatively often as nonspecific symptoms, including prolonged fever, fatigue, myalgia, and malaise. In contrast, pharyngitis, lymphadenopathy, splenomegaly, and atypical lymphocytes are relatively rare in elderly patients.

The incubation period for IM in young adults is ~4–6 weeks. A prodrome of fatigue, malaise, and myalgia may last for 1–2 weeks before the onset of fever, sore throat, and lymphadenopathy. Fever is usually low-grade and is most common in the first 2 weeks of the illness; however, it may persist for >1 month. Common signs and symptoms are listed along with their frequencies in **Table 181–1**. Lymphadenopathy and pharyngitis are most prominent during the first 2 weeks of the illness, while splenomegaly is more prominent during the second and third weeks. Lymphadenopathy most often affects the posterior cervical nodes but may be generalized. Enlarged lymph nodes are frequently tender and symmetric but are not fixed in place. Pharyngitis, often the most prominent sign, can be accompanied by enlargement of the tonsils with an exudate resembling that of streptococcal pharyngitis. A morbilliform or papular rash, usually on the arms or trunk, develops in ~5% of cases **(Fig. 181-1)**. Most patients treated with ampicillin develop a macular rash; this rash is not predictive of future adverse reactions to penicillins. Erythema nodosum and erythema multiforme have also been described (Chap. 53). Most patients have symptoms for 2–4 weeks, but malaise and difficulty concentrating can persist for months.

Table 181–1 Signs and Symptoms of Infectious Mononucleosis

Manifestation	Median Percentage of Patients (Range)
Symptoms	
Sore throat	75 (50–87)
Malaise	47 (42–76)
Headache	38 (22–67)
Abdominal pain, nausea, or vomiting	17 (5–25)
Chills	10 (9–11)
Signs	
Lymphadenopathy	95 (83–100)
Fever	93 (60–100)
Pharyngitis or tonsillitis	82 (68–90)
Splenomegaly	51 (43–64)
Hepatomegaly	11 (6–15)
Rash	10 (0–25)
Periorbital edema	13 (2–34)
Palatal enanthem	7 (3–13)
Jaundice	5 (2–10)

Figure 181-1

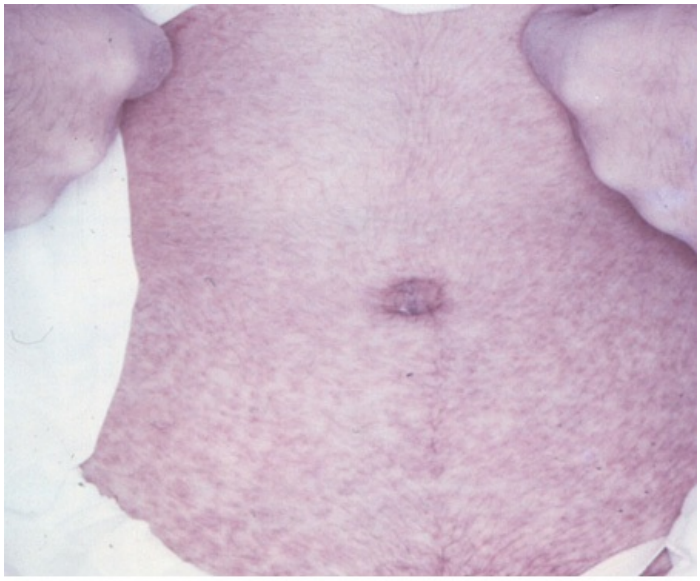

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rash in a patient with infectious mononucleosis due to Epstein-Barr virus.*(Courtesy of Maria Turner, MD; with permission.)*

LABORATORY FINDINGS

The white blood cell count is usually elevated and peaks at 10,000–20,000/μL during the second or third week of illness. Lymphocytosis is usually demonstrable, with >10% atypical lymphocytes. The latter cells are enlarged lymphocytes that have abundant cytoplasm, vacuoles, and indentations of the cell membrane **(Fig. 181-2)**. CD8+ cells predominate among the atypical lymphocytes. Low-grade neutropenia and thrombocytopenia are common during the first month of illness. Liver function is abnormal in >90% of cases. Serum levels of aminotransferases and alkaline phosphatase are usually mildly elevated. The serum concentration of bilirubin is elevated in ~40% of cases.

Figure 181-2

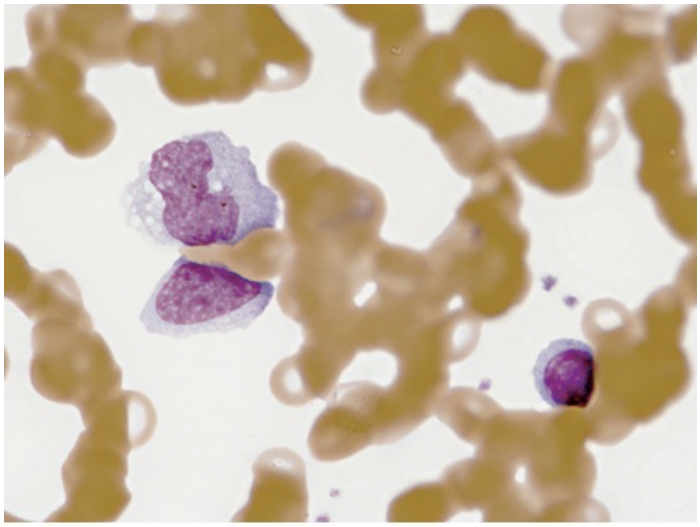

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atypical lymphocytes from a patient with infectious mononucleosis due to Epstein-Barr virus.

COMPLICATIONS

Most cases of IM are self-limited. Deaths are very rare and most often are due to central nervous system (CNS) complications, splenic rupture, upper airway obstruction, or bacterial superinfection.

When CNS complications develop, they usually do so during the first 2 weeks of EBV infection; in some patients, especially children, they are the only clinical manifestations of IM. Heterophile antibodies and atypical lymphocytes may be absent. Meningitis and encephalitis are the most common

neurologic abnormalities, and patients may present with headache, meningismus, or cerebellar ataxia. Acute hemiplegia and psychosis have also been described. The cerebrospinal fluid (CSF) contains mainly lymphocytes, with occasional atypical lymphocytes. Most cases resolve without neurologic sequelae. Acute EBV infection has also been associated with cranial nerve palsies (especially those involving cranial nerve VII), Guillain-Barré syndrome, acute transverse myelitis, and peripheral neuritis.

Autoimmune hemolytic anemia occurs in ~2% of cases during the first 2 weeks. In most cases, the anemia is Coombs-positive, with cold agglutinins directed against the red blood cell antigen. Most patients with hemolysis have mild anemia that lasts for 1–2 months, but some patients have severe disease with hemoglobinuria and jaundice. Nonspecific antibody responses may also include rheumatoid factor, antinuclear antibodies, anti–smooth muscle antibodies, antiplatelet antibodies, and cryoglobulins. IM has been associated with red-cell aplasia, severe granulocytopenia, thrombocytopenia, pancytopenia, and hemophagocytic lymphohistiocytosis. The spleen ruptures in <0.5% of cases. Splenic rupture is more common among male than female patients and may manifest as abdominal pain, referred shoulder pain, or hemodynamic compromise.

Hypertrophy of lymphoid tissue in the tonsils or adenoids can result in upper airway obstruction, as can inflammation and edema of the epiglottis, pharynx, or uvula. About 10% of patients with IM develop streptococcal pharyngitis after their initial sore throat resolves.

Other rare complications associated with acute EBV infection include hepatitis (which can be fulminant), myocarditis or pericarditis with electrocardiographic changes, pneumonia with pleural effusion, interstitial nephritis, genital ulcerations, and vasculitis.

EBV-ASSOCIATED DISEASES OTHER THAN IM

EBV-associated lymphoproliferative disease has been described in patients with congenital or acquired immunodeficiency, including those with severe combined immunodeficiency, patients with AIDS, and recipients of bone marrow or organ transplants who are receiving immunosuppressive drugs (especially cyclosporine). Proliferating EBV-infected B cells infiltrate lymph nodes and multiple organs, and patients present with fever and lymphadenopathy or gastrointestinal symptoms. Pathologic studies show B cell hyperplasia or poly- or monoclonal lymphoma. X-linked lymphoproliferative disease (XLPD) is a recessive disorder of young boys who have a normal response to childhood infections but develop fatal lymphoproliferative disorders after infection with EBV. The protein associated with most cases of this syndrome (SAP) binds to a protein that mediates interactions of B and T cells. Most patients with this syndrome die of acute IM. Others develop hypogammaglobulinemia, malignant B cell lymphomas, aplastic anemia, or agranulocytosis. Disease resembling XLPD has also been associated with mutations in the XIAP or ITK proteins. Moreover, IM has proved fatal to some patients with no obvious preexisting immune abnormality.

Oral hairy leukoplakia **(Fig. 181-3)** is an early manifestation of infection with HIV in adults (Chap. 189). Most patients present with raised, white corrugated lesions on the tongue (and occasionally on the buccal mucosa) that contain EBV DNA. Children infected with HIV can develop lymphoid interstitial pneumonitis; EBV DNA is often found in lung tissue from these patients.

Figure 181-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Oral hairy leukoplakia often presents as white plaques on the lateral surface of the tongue and is associated with Epstein-Barr virus infection.

Patients with chronic fatigue syndrome may have titers of antibody to EBV that are elevated but are not significantly different from those in healthy EBV-seropositive adults. While some patients have malaise and fatigue that persist for weeks or months after IM, persistent EBV infection is not a cause of chronic fatigue syndrome. Chronic active EBV infection is very rare and is distinct from chronic fatigue syndrome. The affected patients have an illness lasting >6 months, with elevated levels of EBV DNA in

the blood, very high titers of antibody to EBV, and evidence of organ involvement, including hepatosplenomegaly, lymphadenopathy, and pneumonitis, uveitis, or neurologic disease.

EBV is associated with several malignancies. About 15% of cases of Burkitt's lymphoma in the United States and ~90% of those in Africa are associated with EBV (Chap. 110). African patients with Burkitt's lymphoma have high levels of antibody to EBV, and their tumor tissue usually contains viral DNA. Malaria infection in Africa may impair cellular immunity to EBV and induce polyclonal B cell activation with an expansion of EBV-infected B cells. These changes may enhance the proliferation of B cells with elevated EBV DNA in the bloodstream, thereby increasing the likelihood of a *c-myc* translocation —the hallmark of Burkitt's lymphoma. EBV-containing Burkitt's lymphoma also occurs in patients with AIDS.

Anaplastic nasopharyngeal carcinoma is common in southern China and is uniformly associated with EBV; the affected tissues contain viral DNA and antigens. Patients with nasopharyngeal carcinoma often have elevated titers of antibody to EBV (Chap. 88). High levels of EBV plasma DNA before treatment or detectable levels of EBV DNA after radiation therapy correlate with lower rates of overall survival and relapse-free survival among patients with nasopharyngeal carcinoma.

EBV has been associated with Hodgkin's disease, especially the mixed-cellularity type (Chap. 110). Patients with Hodgkin's disease often have elevated titers of antibody to EBV. In about half of cases in the United States, viral DNA and antigens are found in Reed-Sternberg cells. The risk of EBV-positive Hodgkin's disease is significantly increased in young adults after EBV-seropositive IM. About 50% of non-Hodgkin's lymphomas in patients with AIDS are EBV-positive.

EBV is present in B cells of lesions from patients with lymphomatoid granulomatosis. In some cases, EBV DNA has been detected in tumors from immunocompetent patients with angiocentric nasal NK/T cell lymphoma, T cell lymphoma, gastric carcinoma, and CNS lymphoma. Studies have demonstrated viral DNA in leiomyosarcomas from AIDS patients and in smooth-muscle tumors from organ transplant recipients. Virtually all CNS lymphomas in AIDS patients are associated with EBV. Studies have found that a history of IM and higher levels of antibodies to EBV before the onset of disease is more common in persons with multiple sclerosis than in the general population, and additional research on a possible causal relationship is needed.

Diagnosis

SEROLOGIC TESTING

The heterophile test is used for the diagnosis of IM in children and adults **(Table 181–2)**. In the test for this antibody, human serum is absorbed with guinea pig kidney, and the heterophile titer is defined as the greatest serum dilution that agglutinates sheep, horse, or cow erythrocytes. The heterophile antibody does not interact with EBV proteins. A titer of ≥forty-fold is diagnostic of acute EBV infection in a patient who has symptoms compatible with IM and atypical lymphocytes. Tests for heterophile antibodies are positive in 40% of patients with IM during the first week of illness and in 80–90% during the third week. Therefore, repeated testing may be necessary, especially if the initial test is performed early. Tests usually remain positive for 3 months after the onset of illness, but heterophile antibodies can persist for up to 1 year. These antibodies usually are not detectable in children <5 years of age, in the elderly, or in patients presenting with symptoms not typical of IM. The commercially available monospot test for heterophile antibodies is somewhat more sensitive than the classic heterophile test. The monospot test is ~75% sensitive and ~90% specific compared with EBV-specific serologies. False-positive monospot results are more common among persons with connective tissue disease, lymphoma, viral hepatitis, and malaria.

Table 181–2 Serologic Features of EBV-Associated Diseases

	Result in Indicated Test					
		Anti-VCA		Anti-EA		
Condition	**Heterophile**	**IgM**	**IgG**	**EA-D**	**EA-R**	**Anti-EBNA**
Acute infectious mononucleosis	+	+	++	+	–	–
Convalescence	±	–	+	–	±	+
Past infection	–	–	+	–	–	+
Reactivation with immunodeficiency	–	–	++	+	+	±
Burkitt's lymphoma	–	–	+++	±	++	+
Nasopharyngeal carcinoma	–	–	+++	++	±	+

Abbreviations: EA, early antigen; EA-D antibody, antibody to early antigen in diffuse pattern in nucleus and cytoplasm of infected cells; EA-R antibody, antibody to early antigen restricted to the cytoplasm; EBNA, Epstein-Barr nuclear antigen; VCA, viral capsid antigen.

Source: Adapted from OkanoM et al: Clin Microbiol Rev 1:300, 1988.

EBV-specific antibody testing is used for patients with suspected acute EBV infection who lack heterophile antibodies and for patients with atypical infections (Table 181–2). Titers of IgM and IgG antibodies to viral capsid antigen (VCA) are elevated in the serum of more than 90% of patients at the onset of disease. IgM antibody to VCA is most useful for the diagnosis of acute IM because it is present at elevated titers only during the first 2–3 months of the disease; in contrast, IgG antibody to VCA is usually not useful for diagnosis of IM but is often used to assess past exposure to EBV because it persists for life. Seroconversion to EBNA positivity is also useful for the diagnosis of acute infection with EBV. Antibodies to EBNA become detectable relatively late (3–6 weeks after the onset of symptoms) in nearly all cases of acute EBV infection and persist for the lifetime of the patient. These antibodies may be lacking in immunodeficient patients and in those with chronic active EBV infection.

Titers of other antibodies may also be elevated in IM; however, these elevations are less useful for diagnosis. Antibodies to early antigens (EAs) are detectable 3–4 weeks after the onset of symptoms in patients with IM. About 70% of individuals with IM have EA-D antibodies during the illness; the presence of EA-D antibodies is especially likely in patients with relatively severe disease. These antibodies usually persist for only 3–6 months. Levels of EA-D antibodies are also elevated in patients with nasopharyngeal carcinoma or chronic active EBV infection. EA-R antibodies are only occasionally detected in patients with IM but are often found at elevated titers in patients with African Burkitt's lymphoma or chronic active EBV infection. IgA antibodies to EBV antigens have proved useful for the identification of patients with nasopharyngeal carcinoma and of persons at high risk for the disease.

OTHER STUDIES

Detection of EBV DNA, RNA, or proteins has been valuable in demonstrating the association of the virus with various malignancies. The polymerase chain reaction has been used to detect EBV DNA in the CSF of some AIDS patients with lymphomas and to monitor the amount of EBV DNA in the blood of patients with lymphoproliferative disease. Detection of high levels of EBV DNA in blood during the first few weeks of IM may be useful if serologic studies yield equivocal results. Culture of EBV from throat washings or blood is not helpful in the diagnosis of acute infection, since EBV commonly persists in the oropharynx and in B cells for the lifetime of the infected individual.

DIFFERENTIAL DIAGNOSIS

Whereas ~90% of cases of IM are due to EBV, 5–10% of cases are due to cytomegalovirus (CMV). CMV is the most common cause of heterophile-negative mononucleosis; less common causes of IM and differences from IM due to EBV are shown in **Table 181–3**.

Table 181–3 Differential Diagnosis of Infectious Mononucleosis

	Sign or Symptom				
Etiology	**Fever**	**Adenopathy**	**Sore Throat**	**Atypical Lymphocytes**	**Differences from EBV Mononucleosis**
EBV	+	+	+	+	—
CMV	+	±	±	+	Older age at presentation, longer duration of fever
HIV	+	+	+	±	Diffuse rash, oral/genital ulcers, aseptic meningitis
Toxoplasmosis	+	+	±	±	Less splenomegaly, exposure to cats or raw meat
HHV-6	+	+	+	+	Older age at presentation
Streptococcal pharyngitis	+	+	+	–	No splenomegaly, less fatigue
Viral hepatitis	+	±	–	±	Higher aminotransferase levels
Rubella	+	+	±	±	Maculopapular rash, no splenomegaly
Lymphoma	+	+	+	+	Fixed, nontender lymph nodes
Drugs[a]	+	+	–	±	Occurs at any age

[a] Most commonly phenytoin, carbamazepine, sulfonamides, or minocycline.CMV, cytomegalovirus; EBV, Epstein-Barr virus; HHV, human herpesvirus.

Treatment: EBV-Associated Disease

Therapy for IM consists of supportive measures, with rest and analgesia. Excessive physical activity during the first month should be avoided to reduce the possibility of splenic rupture, which necessitates splenectomy. Glucocorticoid therapy is not indicated for uncomplicated IM and in fact may predispose to bacterial superinfection. Prednisone (40–60 mg/d for 2–3 days, with subsequent tapering of the dose over 1–2 weeks) has been used for the prevention of airway obstruction in

patients with severe tonsillar hypertrophy, for autoimmune hemolytic anemia, for hemophagocytic lymphohistiocytosis, and for severe thrombocytopenia. Glucocorticoids have also been administered to a few selected patients with severe malaise and fever and to patients with severe CNS or cardiac disease.

Acyclovir has had no significant clinical impact on IM in controlled trials. In one study, the combination of acyclovir and prednisolone had no significant effect on the duration of symptoms of IM.

Acyclovir, at a dosage of 400–800 mg five times daily, has been effective for the treatment of oral hairy leukoplakia (despite common relapses). The posttransplantation EBV lymphoproliferative syndrome (Chap. 132) generally does not respond to antiviral therapy. When possible, therapy should be directed toward reduction of immunosuppression. Antibody to CD20 (rituximab) has been effective in some cases. Infusions of donor lymphocytes are often effective for stem cell transplant recipients, although graft-versus-host disease can occur. Infusions of EBV-specific cytotoxic T cells have been used to prevent EBV lymphoproliferative disease in high-risk settings as well as to treat the disease. IFN-α administration, cytotoxic chemotherapy, and radiation therapy (especially for CNS lesions) have also been used. Infusion of autologous EBV-specific cytotoxic T lymphocytes has shown promise in small studies of patients with nasopharyngeal carcinoma and Hodgkin's disease. Treatment of several cases of XLPD with antibody to CD20 resulted in a successful outcome of what otherwise would probably have been fatal acute EBV infection.

Prevention

The isolation of patients with IM is unnecessary. A vaccine directed against the major EBV glycoprotein reduced the frequency of IM but did not affect the rate of asymptomatic infection.

FURTHER READINGS

Choquet S et al: Efficacy and safety of rituximab in B-cell post-transplant lymphoproliferative disorders: Results of a prospective multicenter phase 2 study. Blood 107:3053, 2006[PMID: 16254143]

Cohen JI et al: Current understanding of the role of Epstein-Barr virus in lymphomagenesis and therapeutic approaches to EBV-associated lymphomas. Leuk Lymphoma 49(Suppl 1):27, 2008

———: Epstein-Barr virus infection. N Engl J Med 343:481, 2000

Gulley ML, Tang W: Using Epstein-Barr viral load assays to diagnose, monitor, and prevent posttransplant lymphoproliferative disorder. Clin Microbiol Rev 23:350, 2010[PMID: 20375356]

Haque T et al: Allogeneic cytotoxic T-cell therapy for EBV-positive posttransplantation lymphoproliferative disease: Results of a phase 2 multicenter clinical trial. Blood 110:1123, 2007[PMID: 17468341]

Landgren O et al: Risk factors for lymphoproliferative disorders after allogeneic hematopoietic cell transplantation. Blood 113:4992, 2009[PMID: 19264919]

Luzuriaga K, Sullivan JL: Infectious mononucleosis. N Engl J Med 362:1993, 2010[PMID: 20505178]

Milnone MC et al: Treatment of primary Epstein-Barr virus infection in patients with X-linked lymphoproliferative disease using B-cell-directed therapy. Blood 105:994, 2005

Njie R et al: The effects of acute malaria on Epstein-Barr virus (EBV) load and EBV-specific T cell immunity in Gambian children. J Infect Dis 199:31, 2009[PMID: 19032105]

Torre D, Tambini R: Acyclovir for treatment of infectious mononucleosis: A meta-analysis. Scand J Infect Dis 31:543, 1999[PMID: 10680982]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CYTOMEGALOVIRUS

Definition

Cytomegalovirus (CMV), which was initially isolated from patients with congenital cytomegalic inclusion disease, is now recognized as an important pathogen in all age groups. In addition to inducing severe birth defects, CMV causes a wide spectrum of disorders in older children and adults, ranging from an asymptomatic subclinical infection to a mononucleosis syndrome in healthyindividuals to disseminated disease in immunocompromised patients. Human CMV is one of several related species-specific viruses that cause similar diseases in various animals. All are associated with the production of characteristic enlarged cells—hence the name *cytomegalovirus*.

CMV, a β-herpesvirus, has double-strand DNA, four species of mRNA, a protein capsid, and a lipoprotein envelope. Like other herpesviruses, CMV demonstrates icosahedral symmetry, replicates in the cell nucleus, and can cause either a lytic and productive or a latent infection. CMV can be distinguished from other herpesviruses by certain biologic properties, such as host range and type of cytopathology. Viral replication is associated with the production of large intranuclear inclusions and smaller cytoplasmic inclusions. CMV appears to replicate in a variety of cell types in vivo; in tissue culture it grows preferentially in fibroblasts. Although there is little evidence that CMV is oncogenic in vivo, it does transform fibroblasts in rare instances, and genomic transforming fragments have been identified.

Epidemiology

CMV has a worldwide distribution. Of newborns in the United States, ~1% are infected with CMV; the percentages are higher in many less-developed countries. Communal living and poor personal hygiene facilitate early spread. Perinatal and early childhood infections are common. CMV may be present in breast milk, saliva, feces, and urine. Transmission has occurred among young children in day-care centers and has been traced from infected toddler to pregnant mother to developing fetus. When an infected child introduces CMV into a household, 50% of susceptible family members seroconvert within 6 months.

CMV is not readily spread by casual contact but rather requires repeated or prolonged intimate exposure for transmission. In late adolescence and young adulthood, CMV is often transmitted sexually, and asymptomatic carriage in semen or cervical secretions is common. Antibody to CMV is present at detectable levels in a high proportion of sexually active men and women, who may harbor several strains simultaneously. Transfusion of whole blood or certain blood products containing viable leukocytes may transmit CMV, with a frequency of 0.14–10% per unit transfused.

Once infected, an individual generally carries CMV for life. The infection usually remains silent. However, CMV reactivation syndromes develop frequently when T lymphocyte–mediated immunity is compromised—for example, after organ transplantation, in association with lymphoid neoplasms and certain acquired immunodeficiencies (in particular, HIV infection; Chap. 189), or in critically ill patients on intensive care units. Most primary CMV infections in organ transplant recipients (Chap. 132) result from transmission in the graft itself. In CMV-seropositive transplant recipients, infection results from reactivation of latent virus or, less commonly, from reinfection by a new strain. CMV infection may also be associated with diseases as diverse as coronary artery stenosis and malignant gliomas, but these associations require further validation.

Pathogenesis

Congenital CMV infection can result from either primary or reactivation infection of the mother. However, clinical disease in the fetus or newborn is related almost exclusively to primary maternal infection **(Table 182-1)**. The factors determining the severity of congenital infection are unknown; a deficient capacity to produce precipitating antibodies and to mount T cell responses to CMV is associated with relatively severe disease.

Table 182-1 CMV Disease in the Immunocompromised Host

Population	Risk Factors	Principal Syndromes	Treatment	Prevention
Fetus	Primary maternal	Cytomegalic	Ganciclovir for	Avoidance of exposure;

	infection/early pregnancy	inclusion disease	symptomatic neonates	possibly, maternal treatment with CMV immunoglobulin during pregnancy
Organ transplant recipient	Seropositivity of donor and/or recipient; immunosuppressive regimen; high degree of rejection	Febrile leukopenia; pneumonia; gastrointestinal disease	Ganciclovir or valganciclovir	Donor matching; prophylaxis or preemptive therapy with ganciclovir or valganciclovir
Bone marrow transplant recipient	Graft-vs.-host disease; older age of recipient; seropositive recipient; viremia	Pneumonia; gastrointestinal disease	Ganciclovir plus CMV immunoglobulin	Donor matching; prophylaxis or preemptive therapy with ganciclovir or valganciclovir
Person with AIDS	<100 CD4+ T cells/μL; CMV seropositivity	Retinitis; gastrointestinal disease; neurologic disease	Ganciclovir, valganciclovir, foscarnet, or cidofovir	Oral valganciclovir

Primary infection with CMV in late childhood or adulthood is often associated with a vigorous T lymphocyte response that may contribute to the development of a mononucleosis syndrome similar to that observed after infection with Epstein-Barr virus (Chap. 181). The hallmark of such infection is the appearance of atypical lymphocytes in the peripheral blood; these cells are predominantly activated CD8+ T lymphocytes. Polyclonal activation of B cells by CMV contributes to the development of rheumatoid factors and other autoantibodies during mononucleosis.

Once acquired, CMV persists indefinitely in host tissues. The sites of persistent infection probably include multiple cell types and various organs. Transmission via blood transfusion or organ transplantation is due to silent infections in these tissues. Autopsy studies suggest that salivary glands and bowel may be sites of latent infection.

If the host's T cell responses become compromised by disease or by iatrogenic immunosuppression, latent virus can be reactivated to cause a variety of syndromes. Chronic antigenic stimulation in the presence of immunosuppression (for example, after tissue transplantation) appears to be an ideal setting for CMV activation and CMV-induced disease. Certain particularly potent suppressants of T cell immunity (e.g., antithymocyte globulin) are associated with a high rate of clinical CMV syndromes, which may follow either primary or reactivation infection. CMV may itself contribute to further T lymphocyte hyporesponsiveness, which often precedes superinfection with other opportunistic pathogens, such as *Pneumocystis*. CMV and *Pneumocystis* are frequently found together in immunosuppressed patients with severe interstitial pneumonia.

Pathology

Cytomegalic cells in vivo (presumed to be infected epithelial cells) are two to four times larger than surrounding cells and often contain an 8- to 10-μm intranuclear inclusion that is eccentrically placed and is surrounded by a clear halo, producing an "owl's eye" appearance. Smaller granular cytoplasmic inclusions are demonstrated occasionally. Cytomegalic cells are found in a wide variety of organs, including the salivary gland, lung, liver, kidney, intestine, pancreas, adrenal gland, and central nervous system.

The cellular inflammatory response to infection consists of plasma cells, lymphocytes, and monocyte-macrophages. Granulomatous reactions occasionally develop, particularly in the liver. Immunopathologic reactions may contribute to CMV disease. Immune complexes have been detected in infected infants, sometimes in association with CMV-related glomerulopathies. Immune-complex glomerulopathy has also been observed in some CMV-infected patients after renal transplantation.

Clinical Manifestations

CONGENITAL CMV INFECTION

Fetal infections range from inapparent to severe and disseminated. Cytomegalic inclusion disease develops in ~5% of infected fetuses and is seen almost exclusively in infants born to mothers who develop primary infections during pregnancy. Petechiae, hepatosplenomegaly, and jaundice are the most common presenting features (60–80% of cases). Microcephaly with or without cerebral calcifications, intrauterine growth retardation, and prematurity are reported in 30–50% of cases. Inguinal hernias and chorioretinitis are less common. Laboratory abnormalities include elevated alanine aminotransferase levels in serum, thrombocytopenia, conjugated hyperbilirubinemia, hemolysis, and elevated protein levels in cerebrospinal fluid. The prognosis for severely infected infants is poor; the mortality rate is 20–30%, and few survivors escape intellectual or hearing difficulties later in childhood. The differential diagnosis of cytomegalic inclusion disease in infants includes syphilis, rubella, toxoplasmosis, infection with herpes simplex virus or enterovirus, and bacterial sepsis.

Most congenital CMV infections are clinically inapparent at birth. Of asymptomatically infected infants, 5–25% develop significant psychomotor, hearing, ocular, or dental abnormalities over the next several years.

PERINATAL CMV INFECTION

The newborn may acquire CMV at delivery by passage through an infected birth canal or by postnatal contact with infected breast milk or other maternal secretions. Of infants who are breast-fed for>1 month by seropositive mothers, 40–60% become infected. Iatrogenic transmission can result from neonatal blood transfusion; screening of blood products before transfusion into low-birth-weight seronegative infants or seronegative pregnant women decreases risk.

The great majority of infants infected at or after delivery remain asymptomatic. However, protracted interstitial pneumonitis has been associated with perinatally acquired CMV infection, particularly in premature infants, and occasionally has been accompanied by infection with *Chlamydia trachomatis*, *Pneumocystis*, or *Ureaplasma urealyticum*. Poor weight gain, adenopathy, rash, hepatitis, anemia, and atypical lymphocytosis may also be found, and CMV excretion often persists for months or years.

CMV MONONUCLEOSIS

The most common clinical manifestation of CMV infection in immunocompetent hosts beyond the neonatal period is a heterophile antibody–negative mononucleosis syndrome, which may develop spontaneously or follow transfusion of leukocyte-containing blood products. Although the syndrome occurs at all ages, it most often involves sexually active young adults. With incubation periods of 20 –60 days, the illness generally lasts for 2–6 weeks. Prolonged high fevers, sometimes with chills, profound fatigue, and malaise, characterize this disorder. Myalgias, headache, and splenomegaly are common, but in CMV (as opposed to Epstein-Barr virus) mononucleosis, exudative pharyngitis, and cervical lymphadenopathy are rare. Occasional patients develop rubelliform rashes, often after exposure to ampicillin or certain other antibiotics. Less common are interstitial or segmental pneumonia, myocarditis, pleuritis, arthritis, and encephalitis. In rare cases, Guillain-Barré syndrome complicates CMV mononucleosis. The characteristic laboratory abnormality is relative lymphocytosis in peripheral blood, with >10% atypical lymphocytes. Total leukocyte counts may be low, normal, or markedly elevated. Although significant jaundice is uncommon, serum aminotransferase and alkaline phosphatase levels are often moderately elevated. Heterophile antibodies are absent; however, transient immunologic abnormalities are common and may include the presence of cryoglobulins, rheumatoid factors, cold agglutinins, and antinuclear antibodies. Hemolytic anemia, thrombocytopenia, and granulocytopenia complicate recovery in rare instances.

Most patients recover without sequelae, although postviral asthenia may persist for months. The excretion of CMV in urine, genital secretions, and/or saliva often continues for months or years. Rarely, CMV infection is fatal in immunocompetent hosts; survivors can have recurrent episodes of fever and malaise, sometimes associated with autonomic nervous system dysfunction (e.g., attacks of sweating or flushing).

CMV INFECTION IN THE IMMUNOCOMPROMISED HOST

(Table 182-1) CMV appears to be the most common and important viral pathogen complicating organ transplantation (Chap. 132). In recipients of kidney, heart, lung, and liver transplants, CMV induces a variety of syndromes, including fever and leukopenia, hepatitis, pneumonitis, esophagitis, gastritis, colitis, and retinitis. CMV disease may be an independent risk factor for both graft loss and death. The period of maximal risk is between 1 and 4 months after transplantation, although retinitis may be a later complication. Disease likelihood and viral replication levels generally are greater after primary infection than after reactivation. In addition, molecular studies indicate that seropositive transplant recipients are susceptible to reinfection with donor-derived, genotypically variant CMV, and such infection often results in disease. Reactivation infection, although common, is less likely than primary infection to be important clinically. The risk of clinical disease is related to various factors, such as degree of immunosuppression; use of antibodies to T cell receptors; lack of utilization of anti-CMV prophylaxis; and co-infection with other pathogens. The transplanted organ is particularly vulnerable as a target for CMV infection; thus, there is a tendency for CMV hepatitis to follow liver transplantation and for CMV pneumonitis to follow lung transplantation.

CMV pneumonia occurs in 15–20% of bone marrow transplant recipients; the case-fatality rate is 84 –88%, although the risk of severe disease may be reduced by prophylaxis or preemptive therapy with antiviral drugs. The risk is greatest 5–13 weeks after transplantation, and identified risk factors include certain types of immunosuppressive therapy, acute graft-versus-host disease, older age, viremia, and pretransplantation seropositivity.

CMV is an important pathogen in patients with advanced HIV infection (Chap. 189), in whom it may cause retinitis or disseminated disease, particularly when peripheral-blood CD4+ T cell counts fall below 50–100/μL. As treatment for underlying HIV infection has improved, the incidence of serious CMV infections (e.g., retinitis) has decreased. However, during the first few weeks after institution of highly active antiretroviral therapy, acute flare-ups of CMV retinitis may occur secondary to an immune reconstitution inflammatory syndrome.

Syndromes produced by CMV in immunocompromised hosts often begin with prolonged fever, malaise, anorexia, fatigue, night sweats, and arthralgias or myalgias. Liver function abnormalities, leukopenia, thrombocytopenia, and atypical lymphocytosis may be observed during these episodes. The development of tachypnea, hypoxemia, and unproductive cough signals respiratory involvement. Radiologic examination of the lung often shows bilateral interstitial or reticulonodular infiltrates that begin in the periphery of the lower lobes and spread centrally and superiorly; localized segmental, nodular, or alveolar patterns are less common. The differential diagnosis includes *Pneumocystis* infection; other viral, bacterial, or fungal infections; pulmonary hemorrhage; and injury secondary to irradiation or to treatment with cytotoxic drugs.

Gastrointestinal CMV involvement may be localized or extensive and almost exclusively affects compromised hosts. Ulcers of the esophagus, stomach, small intestine, or colon may result in bleeding or perforation. CMV infection may lead to exacerbations of underlying ulcerative colitis. Hepatitis occurs frequently, particularly after liver transplantation, and acalculous cholecystitis and adrenalitis have been described.

CMV rarely causes meningoencephalitis in otherwise-healthy individuals. Two forms of CMV encephalitis are seen in patients with AIDS. One resembles HIV encephalitis and presents as progressive dementia; the other is a ventriculoencephalitis characterized by cranial-nerve deficits, nystagmus, disorientation, lethargy, and ventriculomegaly. In immunocompromised patients, CMV can also cause subacute progressive polyradiculopathy, which is often reversible if recognized and treated promptly.

CMV retinitis is an important cause of blindness in immunocompromised patients, particularly patients with advanced AIDS (Chap. 189). Early lesions consist of small, opaque, white areas of granular retinal necrosis that spread in a centrifugal manner and are later accompanied by hemorrhages, vessel sheathing, and retinal edema **(Fig. 182-1)**. CMV retinopathy must be distinguished from that due to other conditions, including toxoplasmosis, candidiasis, and herpes simplex virus infection.

Figure 182-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cytomegalovirus infection in a patient with AIDS may appear as an arcuate zone of retinitis with hemorrhages and optic disk swelling. Often CMV is confined to the retinal periphery, beyond view of the direct ophthalmoscope.

Fatal CMV infections are often associated with persistent viremia and the involvement of multiple organ systems. Progressive pulmonary infiltrates, pancytopenia, hyperamylasemia, and hypotension are characteristic features that are frequently found in conjunction with a terminal bacterial, fungal, or protozoan superinfection. Extensive adrenal necrosis with CMV inclusions is often documented at autopsy, as is CMV involvement of many other organs.

Diagnosis

The diagnosis of CMV infection usually cannot be made reliably on clinical grounds alone. Isolation of CMV or detection of its antigens or DNA in appropriate clinical specimens is the preferred approach. Virus excretion or viremia is readily detected by culture of appropriate specimens on human fibroblast monolayers. If CMV titers are high, as is common in congenital disseminated infection and in patients with AIDS, characteristic cytopathic effects may be detected within a few days. However, in some situations (e.g., CMV mononucleosis), viral titers are low, and cytopathic effects may take several weeks to appear. Many laboratories expedite diagnosis with an overnight tissue-culture method (shell

vial assay) involving centrifugation and an immunocytochemical detection technique employing monoclonal antibodies to an immediate-early CMV antigen. Isolation of virus from urine or saliva does not, by itself, constitute proof of acute infection, since excretion from these sites may continue for months or years after illness. Detection of viremia is a better predictor of acute infection.

Detection of CMV antigens (pp65) in peripheral-blood leukocytes or of CMV DNA in blood or tissues may hasten diagnosis. Such assays may yield a positive result several days earlier than culture methods. The most sensitive way to detect CMV in blood or other fluids may be by amplifying CMV DNA by polymerase chain reaction (PCR). PCR detection of CMV DNA in blood may predict the risk for disease progression, particularly in immunocompromised hosts, and PCR detection of CMV DNA in cerebrospinal fluid is useful in the diagnosis of CMV encephalitis or polyradiculopathy. However, considerable variation among different assays and different laboratories has been observed.

A variety of serologic assays detect increases in titers of antibody to CMV antigens. An increased antibody level may not be detectable for up to 4 weeks after primary infection, and titers often remain high for years after infection. For this reason, single-sample antibody determinations are of no value in assessing the acuteness of infection. Detection of CMV-specific IgM is sometimes useful in the diagnosis of recent or active infection; however, circulating rheumatoid factors may result in occasional false-positive IgM tests.

Treatment: Cytomegalovirus Infection

Several measures are useful for the prevention of CMV infection in high-risk patients. The use of blood from seronegative donors or of blood that has been frozen, thawed, and deglycerolized greatly decreases the rate of transfusion-associated transmission. Matching of organ or bone marrow transplants by CMV serology, with exclusive use of organs from seronegative donors in seronegative recipients, reduces rates of primary infection after transplantation. A CMV glycoprotein B vaccine reduced infections in a placebo-controlled trial among 464 CMV-seronegative women; this outcome raises the possibility that this experimental vaccine will reduce congenital infections, but further studies must validate this approach.

CMV immune or hyperimmune globulin has been reported (1) to reduce rates of CMV-associated syndromes and of fungal or parasitic superinfections among seronegative renal transplant recipients and (2) to prevent congenital CMV infection in infants of women with primary infection during pregnancy. Studies in bone marrow transplant recipients have produced conflicting results. Prophylactic acyclovir or valacyclovir may reduce rates of CMV infection and disease in certain seronegative renal transplant recipients, although neither drug is effective in the treatment of active CMV disease.

Ganciclovir is a guanosine derivative that has considerably more activity against CMV than its congener acyclovir. After intracellular conversion by a viral phosphotransferase encoded by CMV gene region UL97, ganciclovir triphosphate is a selective inhibitor of CMV DNA polymerase. Several clinical studies have indicated response rates of 70–90% among patients with AIDS who are given ganciclovir for the treatment of CMV retinitis or colitis. In severe infections (e.g., CMV pneumonia in bone marrow transplant recipients), ganciclovir is often combined with CMV immune globulin. Prophylactic or suppressive (preemptive) ganciclovir may be useful in high-risk bone marrow or organ transplant recipients (e.g., those who are CMV-seropositive before transplantation or who are CMV culture –positive afterward). In many patients with AIDS, persistently low CD4+ T cell counts, and CMV disease, clinical and virologic relapses occur promptly if treatment with ganciclovir is discontinued. Therefore, prolonged maintenance regimens are recommended for such patients. Resistance to ganciclovir is common among patients treated for >3 months and is usually related to mutations in the CMV UL97 gene.

Valganciclovir is an orally bioavailable prodrug that is rapidly metabolized to ganciclovir in intestinal tissues and the liver. Approximately 60–70% of an oral dose of valganciclovir is absorbed. An oral valganciclovir dose of 900 mg results in ganciclovir blood levels similar to those obtained with an IV ganciclovir dose of 5 mg/kg. Oral valganciclovir appears to be as effective as IV ganciclovir for both CMV induction and maintenance regimens. Furthermore, the adverse-event profiles and rates of resistance development for the two drugs are similar.

Ganciclovir or valganciclovir therapy for CMV retinitis consists of a 14- to 21-day induction course (5 mg/kg IV twice daily for ganciclovir or 900 mg twice daily for valganciclovir) followed by prolonged maintenance therapy. For parenteral maintenance, the ganciclovir dose is 5 mg/kg daily or 6 mg/kg 5 days per week; for oral maintenance, 900 mg of valganciclovir once daily is recommended. Peripheral-blood neutropenia develops in 16–29% of treated patients but may be ameliorated by granulocyte colony-stimulating factor or granulocyte-macrophage colony-stimulating factor. Discontinuation of maintenance therapy should be considered in patients with AIDS who, while receiving antiretroviral therapy, have a sustained (3- to 6-month) increase in CD4+ T cell counts to >100/μL.

For treatment of CMV retinitis, ganciclovir may also be administered via a slow-release pellet sutured into the eye. Although this intraocular device provides good local protection, contralateral eye disease and disseminated disease are not affected, and early retinal detachment is possible. A combination of

intraocular and systemic therapy may be better than the intraocular implant alone.

Foscarnet (sodium phosphonoformate) inhibits CMV DNA polymerase. Because this agent does not require phosphorylation to be active, it is also effective against most ganciclovir-resistant isolates. Foscarnet is less well tolerated than ganciclovir and causes considerable toxicity, including renal dysfunction, hypomagnesemia, hypokalemia, hypocalcemia, genital ulcers, dysuria, nausea, and paresthesia. Moreover, foscarnet administration requires the use of an infusion pump and close clinical monitoring. With aggressive hydration and dose adjustments for renal dysfunction, the toxicity of foscarnet can be reduced. The use of foscarnet should be avoided when a saline load cannot be tolerated (e.g., in cardiomyopathy). The approved induction regimen is 60 mg/kg every 8 h for 2 weeks, although 90 mg/kg every 12 h is equally effective and no more toxic. Maintenance infusions should deliver 90–120 mg/kg once daily. No oral preparation is available. Foscarnet-resistant virus may emerge during extended therapy.

Cidofovir is a nucleotide analogue with a long intracellular half-life that allows intermittent IV administration. Induction regimens of 5 mg/kg weekly for 2 weeks are followed by maintenance regimens of 3–5 mg/kg every 2 weeks. Cidofovir can cause severe nephrotoxicity through dose-dependent proximal tubular cell injury; however, this adverse effect can be tempered somewhat by saline hydration and probenecid.

It is not clear whether universal prophylaxis or preemptive therapy is the preferable approach in CMV-seropositiveimmunocompromised hosts. Both ganciclovir and valganciclovir have been used successfully for prophylaxis and preemptive therapy in transplant recipients. For patients with advanced HIV infection (CD4+ T cell counts of <50/μL), some authorities have advocated prophylaxis with oral ganciclovir or valganciclovir. However, side effects, lack of proven benefit, possible induction of viral resistance, and high cost have precluded the wide acceptance of this practice. Preemptive ganciclovir or valganciclovir therapy based on detection of CMV viremia by either antigenemia or PCR techniques is under study.

HUMAN HERPESVIRUS TYPES 6, 7, AND 8

Human herpesvirus (HHV) type 6 was first isolated in 1986 from peripheral-blood leukocytes of six persons with various lymphoproliferative disorders. The virus has a worldwide distribution, and two genetically distinct variants (HHV-6A and HHV-6B) are now recognized. HHV-6 appears to be transmitted by saliva and possibly by genital secretions.

Infection with HHV-6 frequently occurs during infancy as maternal antibody wanes. The peak age of acquisition is 9–21 months; by 24 months, seropositivity rates approach 80%. Older siblings appear to serve as a source of transmission. Congenital infection may also occur, and ~1% of newborns are infected with HHV-6; placental infection with HHV-6 has been described. Most postnatally infected children develop symptoms (fever, fussiness, and diarrhea). A minority develop exanthem subitum (roseola infantum; see Fig. e7-5), a common illness characterized by fever with subsequent rash. In addition, ~10–20% of febrile seizures without rash during infancy are caused by HHV-6. After initial infection, HHV-6 persists in peripheral-blood mononuclear cells as well as in the central nervous system, salivary glands, and female genital tract.

In older age groups, HHV-6 has been associated with mononucleosis syndromes; focal encephalitis; and (in immunocompromised hosts) pneumonitis, syncytial giant-cell hepatitis, and disseminated disease. In transplant recipients, HHV-6 infection may be associated with similar syndromes and with graft dysfunction. Acute HHV-6-associated limbic encephalitis has been reported in transplant recipients and is characterized by memory loss, confusion, seizures, hyponatremia, and abnormal electroencephalographic and magnetic resonance imaging results. High plasma loads of HHV-6 DNA in stem cell transplant recipients are associated with allelic-mismatched donors, use of glucocorticoids, delayed monocyte and platelet engraftment, development of limbic encephalitis, and increased all-cause mortality rates. Like many other viruses, HHV-6 has been implicated in the pathogenesis of multiple sclerosis, although further study is needed to distinguish between association and etiology.

HHV-7 was isolated in 1990 from T lymphocytes from the peripheral blood of a healthy 26-year-old man. The virus is frequently acquired during childhood, albeit at a later age than HHV-6. HHV-7 is commonly present in saliva, which is presumed to be the principal source of infection; breast milk can also carry the virus. Viremia can be associated with either primary or reactivation infection. The most common clinical manifestations of childhood HHV-7 infections are fever and seizures. Some children present with respiratory or gastrointestinal signs and symptoms. An association has been made between HHV-7 and pityriasis rosea, but evidence is insufficient to indicate a causal relationship.

HHV-6, HHV-7, and CMV infections may cluster in transplant recipients, making it difficult to sort out the roles of the various agents in individual clinical syndromes. HHV-6 and HHV-7 appear to be susceptible to ganciclovir and foscarnet, although definitive evidence of clinical response is lacking.

Unique herpesvirus-like DNA sequences were reported during 1994 and 1995 in tissues derived from Kaposi's sarcoma (KS) and body cavity–based lymphoma occurring in patients with AIDS. The virus from which these sequences were derived is designated HHV-8 or Kaposi's sarcoma–associated

herpesvirus (KSHV). HHV-8, which infects B lymphocytes, macrophages, and both endothelial and epithelial cells, appears to be causally related not only to KS but also to a subgroup of AIDS-related B cell body cavity–based lymphomas (primary effusion lymphomas) and to multicentric Castleman's disease, a lymphoproliferative disorder of B cells. The association of HHV-8 with several other diseases has been reported but not confirmed.

Unlike other herpesvirus infections, HHV-8 infection is much more common in some geographic areas (e.g., central and southern Africa) than in others (North America, Asia, northern Europe). In high-prevalence areas, infection occurs in childhood, seropositivity is associated with having a seropositive mother or (to a lesser extent) older sibling, and HHV-8 may be transmitted in saliva. In low-prevalence areas, infections typically occur in adults, probably with sexual transmission. Concurrent epidemics of HIV-1 and HHV-8 infections among certain populations (e.g., men who have sex with men) in the late 1970s and early 1980s appear to have resulted in the frequent association of AIDS and KS. Transmission of HHV-8 may also be associated with organ transplantation, injection drug use, and blood transfusion; however, transmission via blood transfusion in the United States appears to be rare or nonexistent.

Primary HHV-8 infection in immunocompetent children may manifest as fever and maculopapular rash. Among individuals with intact immunity, chronic asymptomatic infection is the rule, and neoplastic disorders generally develop only after subsequent immunocompromise. Immunocompromised persons with primary infection may present with fever, splenomegaly, lymphoid hyperplasia, pancytopenia, or rapid-onset KS. Quantitative analysis of HHV-8 DNA suggests a predominance of latently infected cells in KS lesions and frequent lytic replication in multicentric Castleman's disease.

Effective antiretroviral therapy for HIV-infected individuals has led to a marked reduction in rates of KS among persons dually infected with HHV-8 and HIV in resource-rich areas. HHV-8 itself is susceptible in vitro to ganciclovir, foscarnet, and cidofovir. A small randomized, double-blind, placebo-controlled, crossover trial suggested that oral valganciclovir administered once daily reduced HHV-8 replication. However, clinical benefits of valganciclovir or other drugs in HHV-8 infection have not yet been demonstrated.

FURTHER READINGS

Asahi-Ozaki Y et al: Quantitative analysis of Kaposi sarcoma–associated herpesvirus (KSHV) in KSHV-associated diseases. J Infect Dis 193:773, 2006[PMID: 16479510]

Cannon MJ et al: Lack of evidence for human herpesvirus-8 transmission via blood transfusion in a historical US cohort. J Infect Dis 199:1592, 2009[PMID: 19385734]

Caserta MT et al: Human herpesviruses (HHV)-6 and HHV-7 infections in pregnant women. J Infect Dis 196:1296, 2007[PMID: 17922393]

Casper C et al: Valganciclovir for suppression of human herpesvirus-8 replication: A randomized, double-blind, placebo-controlled, crossover trial. J Infect Dis 198:23, 2008[PMID: 18491970]

——— et al: Frequent and asymptomatic oropharyngeal shedding of human herpesvirus 8 among immunocompetent men. J Infect Dis 195:30, 2007

Mbulaiteye S et al: Molecular evidence for mother-to-child transmission of Kaposi sarcoma–associated herpesvirus in Uganda and K1 gene evolution within the host. J Infect Dis 193:1250, 2006[PMID: 16586362]

Nigro G et al: Passive immunization during pregnancy for congenital cytomegalovirus infection. N Engl J Med 353:1350, 2005[PMID: 16192480]

Pang XL et al: Interlaboratory comparison of cytomegalovirus viral load assays. Am J Transplant 9:258, 2009[PMID: 19178413]

Torres-Madriz G, Boucher HW: Perspectives in the treatment and prophylaxis of cytomegalovirus disease in solid-organ transplant recipients. Clin Infect Dis 47:702, 2008[PMID: 18652557]

Yao K et al: Detection of human herpesvirus-6 in cerebrospinal fluid of patients with encephalitis. Ann Neurol 65:235, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MOLLUSCUM CONTAGIOSUM, MONKEYPOX, AND OTHER POXVIRUS INFECTIONS: INTRODUCTION

The poxvirus family includes a large number of related DNA viruses that infect various vertebrate hosts. The poxviruses responsible for infections in humans, along with the main manifestations of these infections, are listed in **Table 183-1**. Infections with orthopoxviruses—e.g., smallpox (variola major) virus (Chap. 221) or the zoonotic monkeypox virus—can result in systemic, potentially lethal human disease. Other poxvirus infections cause primarily localized skin disease in humans.

Table 183-1 Poxviruses and Human Infections

Genus	Species	Geographic Location	Host Reservoir	Human Disease
Orthopoxvirus	Variola[a]	Extinct	Humans	Smallpox, systemic
	Monkeypox	Africa	Rodents	Smallpox-like, systemic
	Cowpox	Europe	Rodents	Local pox lesion, occasionally systemic
	Buffalopox	Indian subcontinent	Water buffalo	Local pox lesion, mild illness
	Cantagalo and Araçatuba	South America	Cattle	Local pox lesion, mild illness
	Vaccinia	—	—	Smallpox vaccine
Molluscipoxvirus	Molluscum contagiosum	Worldwide	Humans	Multiple cutaneous lesions (molluscum contagiosum)
Parapoxvirus	Orf	Worldwide	Sheep, goats	Local pox lesions (contagious pustular dermatitis)
	Pseudocowpox (paravaccinia)	Worldwide	Cattle	Local pox lesions (milker's nodule)
	Bovine papular stomatitis	Worldwide	Cattle	Local pox lesions
	Deerpox	Deer herds	Deer	Local pox lesions
	Sealpox	Seal colonies	Seals	Local pox lesions
Yatapoxvirus	Tanapox	Africa	Monkeys	Local pox lesions

[a]See Chap. 221.

MOLLUSCUM CONTAGIOSUM

Molluscum contagiosum virus is an obligate human pathogen that causes distinctive proliferative skin lesions. These lesions measure 2–5 mm in diameter and are pearly, flesh-colored, and umbilicated, with a characteristic dimple at the center **(Fig. 183-1)**. A relative lack of inflammation and necrosis distinguishes these proliferative lesions from other poxvirus lesions. Lesions may be found—singly or in clusters—anywhere on the body except on the palms and soles and may be associated with an eczematous rash.

Figure 183-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Molluscum contagiosum is a cutaneous poxvirus infection characterized by multiple umbilicated flesh-colored or hypopigmented papules.

Molluscum contagiosum is highly prevalent in children and is the most common human disease resulting from poxvirus infection. Swimming pools are a common vector for transmission. Atopy and compromise of skin integrity increase the risk of infection. Genital lesions are more frequent in adults, to whom the virus may be transmitted by sexual contact. The incubation period ranges from2 weeks to 6 months, with an average of 2–7 weeks. In most cases, the disease is self-limited and regresses spontaneously after 3–4 months in immunocompetent hosts. There are no systemic complications, but skin lesions may persist for 3–5 years. Molluscum contagiosum can be associated with immunosuppression and is frequently seen among HIV-infected patients (Chap. 189). The disease can be more generalized, severe, and persistent in AIDS patients than in other groups. Moreover, molluscum contagiosum can be exacerbated in the immune reconstitution inflammatory syndrome (IRIS) associated with the initiation of antiretroviral therapy.

The diagnosis of molluscum contagiosum is typically based on its clinical presentation and can be confirmed by histologic demonstration of the cytoplasmic eosinophilic inclusions (*molluscum bodies*) that are characteristic of poxvirus replication. Molluscum contagiosum virus cannot be propagated in vitro, but electron microscopy and molecular studies can be used for its identification.

There is no specific systemic treatment for molluscum contagiosum, but a variety of techniques for physical ablation have been used. Cidofovir displays in vitro activity against many poxviruses, and case reports suggest that parenteral or topical cidofovir may have some efficacy in the treatment of recalcitrant molluscum contagiosum in immunosuppressed hosts.

MONKEYPOX

Although monkeypox virus was named after the animal from which it was originally isolated, rodents are the primary viral reservoir. Human infections with monkeypox virus typically occur in Africa when humans come into direct contact with infected animals. Human-to-human propagation of monkeypox infection is rare. Human disease is characterized by a systemic illness and vesicular rash similar to those of variola. The clinical presentation of monkeypox can be confused with that of the more common varicella-zoster virus infection (Chap. 180). Compared with the lesions of this herpesvirus infection, monkeypox lesions tend to be more uniform (i.e., in the same stage of development), diffuse, and peripheral in distribution. Lymphadenopathy is a prominent feature of monkeypox infection.

The first outbreak of human monkeypox infection in the Western Hemisphere occurred during 2003, when more than 70 cases were reported in the midwestern United States. The outbreak was linked to contact with pet prairie dogs that had become infected while being housed with rodents imported from Ghana. Patients presented most frequently with fever, rash, and lymphadenopathy ~12 days after exposure. Nine patients were hospitalized, but there were no deaths. Smallpox vaccination can provide cross-reactive immunity to monkeypox infection; studies of people exposed in the outbreak detected subclinical infection in a few vaccinated individuals—an observation suggesting the possibility of long-term vaccine protection. The risk of human disease from animal orthopoxvirus infections may increase as smallpox immunity wanes in the general population and the popularity of exotic animals as household pets grows.

OTHER ZOONOTIC POXVIRUS INFECTIONS

Cowpox and buffalopox are rare zoonotic infections characterized by cutaneous poxlike lesions and mild systemic illness. Outbreaks of similar poxlike lesions among cattle and farm workers in Brazil have been due to Cantagalo and Araçatuba viruses, which are virtually identical to vaccinia virus and may

have become established in cattle during smallpox vaccination programs.

Parapoxviruses are widely scattered among animal species, but only a few are known to cause human disease via direct contact with infected animals. Parapoxviruses are antigenically distinct from orthopoxviruses and share no cross-immunity. *Tanapox* virus belongs to a separate, antigenically distinct genus and usually causes a single nodular lesion on the exposed area after contact with infected monkeys.

FURTHER READINGS

Gur I: The epidemiology of molluscum contagiosum in HIV-seropositive patients: A unique entity or insignificant finding? Int J STD AIDS 19:503, 2008[PMID: 18663032]

Hammarlund E et al: Multiple diagnostic techniques identify previously vaccinated individuals with protective immunity against monkeypox. Nat Med 11:1005, 2005[PMID: 16086024]

Lewis-Jones S: Zoonotic poxvirus infections in humans. Curr Opin Infect Dis 2:81, 2004

Van Der Wouden JC et al: Interventions for cutaneous molluscum contagiosum. Cochrane Database Syst Rev 4:CD004767, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PARVOVIRUS INFECTIONS: INTRODUCTION

Parvoviruses, members of the family Parvoviridae, are small (diameter, ~22 nm), nonenveloped, icosahedral-shaped viruses with a linear single-strand DNA genome of ~5000 nucleotides. These viruses are dependent on either rapidly dividing host cells or helper viruses for replication. At least four groups of parvoviruses infect humans: parvovirus B19 (B19V), dependoviruses (adeno-associated viruses; AAVs), PARV4/5 virus, and human bocaviruses (HBoVs). Human dependoviruses are nonpathogenic and will not be considered further in this chapter.

PARVOVIRUS B19

Definition

B19V is the type member of the genus *Erythrovirus*. On the basis of viral sequence, B19V is divided into three genotypes (designated 1, 2, and 3), but only a single B19V antigenic type has been described. Genotype 1 is predominant in most parts of the world; genotype 2 is rarely associated with active infection; and genotype 3 appears to predominate in parts of western Africa.

Epidemiology

B19V exclusively infects humans, and infection is endemic in virtually all parts of the world. Transmission occurs predominantly via the respiratory route and is followed by the onset of rash and arthralgia. By the age of 15 years, ~ 50% of children have detectable IgG; this figure rises to >90% among the elderly. In pregnant women, the estimated annual seroconversion rate is ~1%. Within households, secondary infection rates approach 50%.

Detection of high-titer B19V in blood is not unusual (see "Pathogenesis," below). Transmission can occur as a result of transfusion, most commonly of pooled components. To reduce the risk of transmission, plasma pools are screened by nucleic acid amplification technology, and high-titer pools are discarded. B19V is resistant to both heat and solvent-detergent inactivation.

Pathogenesis

B19V replicates primarily in erythroid progenitors. This specificity is due in part to the limited tissue distribution of the primary B19V receptor, blood group P antigen (globoside). Infection leads to high-titer viremia, with $>10^{12}$ virus particles (or IU)/mL detectable in the blood at the apex **(Fig. 184-1)**, and virus-induced cytotoxicity results in cessation of red cell production. In immunocompetent individuals, viremia and arrest of erythropoiesis are transient and resolve as the IgM and IgG antibody response is mounted. In individuals with normal erythropoiesis, there is only a minimal drop in hemoglobin levels; however, in those with increased erythropoiesis (especially with hemolytic anemia), this cessation of red cell production can induce a transient crisis with severe anemia (Fig. 184-1). Similarly, if an individual (or, after maternal infection, a fetus) does not mount a neutralizing antibody response and halt the lytic infection, erythroid production is compromised and chronic anemia develops (Fig. 184-1).

Figure 184-1

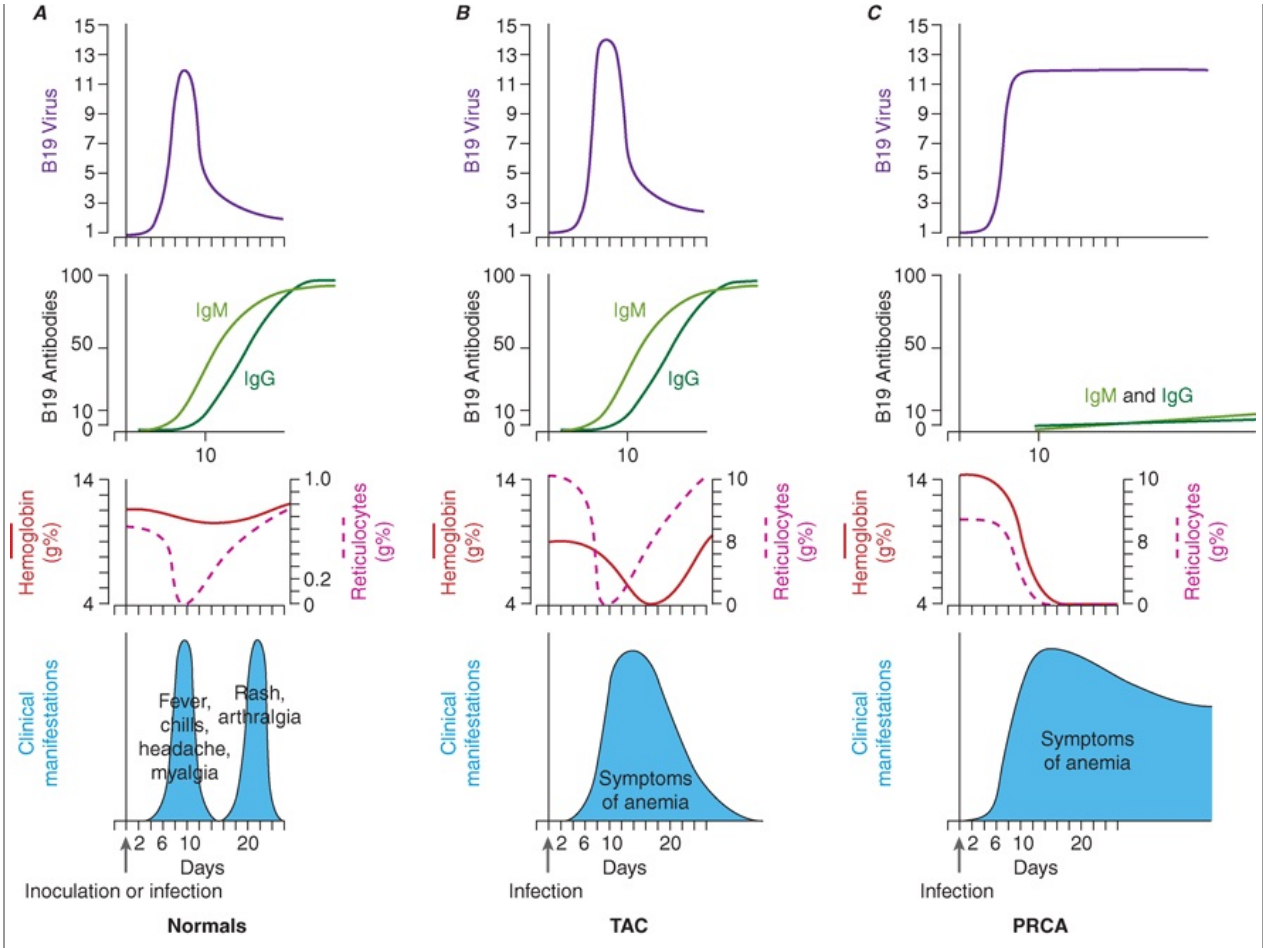

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic of the time course of parvovirus B19V infection in (*A*) normals (erythema infectiosum), (*B*) transient aplastic crisis (TAC), and (*C*) chronic anemia/pure red-cell aplasia (PRCA). *(Reprinted with permission from Young and Brown, 2004. © 2004 Massachusetts Medical Society. All rights reserved.)*

The immune-mediated phase of illness, which begins 2–3 weeks after infection as the IgM response peaks, manifests as the rash of fifth disease together with arthralgia and/or frank arthritis. Low-level B19V DNA can be detected by polymerase chain reaction (PCR) in blood and tissues for months to years after acute infection. The B19V receptor is found in a variety of other cells and tissues, including megakaryocytes, endothelial cells, placenta, myocardium, and liver. Infection of these tissues by B19V may be responsible for some of the more unusual presentations of the infection. Rare individuals who lack P antigen are naturally resistant to B19V infection.

Clinical Manifestations

ERYTHEMA INFECTIOSUM

Most B19V infections are asymptomatic or are associated with only a mild nonspecific illness. The main manifestation of symptomatic B19V infection is erythema infectiosum, also known as *fifth disease* or *slapped-cheek disease*(**Fig. 184-2**). Infection begins with a minor febrile prodrome ~7–10 days after exposure, and the classic facial rash develops several days later; after 2–3 days, the erythematous macular rash may spread to the extremities in a lacy reticular pattern. However, its intensity and distribution vary, and B19V-induced rash is difficult to distinguish from other viral exanthems. Adults typically do not exhibit the "slapped-cheek" phenomenon but present with arthralgia, with or without the macular rash.

Figure 184-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Young child with erythema infectiosum, or fifth disease, showing typical "slapped-cheek" appearance.

POLYARTHROPATHY SYNDROME

Although uncommon among children, arthropathy occurs in ~50% of adults and is more common among women than among men. The distribution of the affected joints is often symmetrical, with arthralgia affecting the small joints of the hands and occasionally the ankles, knees, and wrists. Resolution usually occurs within a few weeks, but recurring symptoms can continue for months. The illness may mimic rheumatoid arthritis, and rheumatoid factor can often be detected in serum. B19V infection may trigger rheumatoid disease in some patients and has been associated with juvenile idiopathic arthritis.

TRANSIENT APLASTIC CRISIS

Asymptomatic transient reticulocytopenia occurs in most individuals with B19V infection. However, in patients who depend on continual rapid production of red cells, infection can cause transient aplastic crisis (TAC). Affected individuals include those with hemolytic disorders, hemoglobinopathies, red cell enzymopathies, and autoimmune hemolytic anemias. Patients present with symptoms of severe anemia (sometimes life-threatening) and a low reticulocyte count, and bone marrow examination reveals an absence of erythroid precursors and characteristic giant pronormoblasts. As its name indicates, the illness is transient, and anemia resolves with the cessation of cytopathic infection in the erythroid progenitors.

PURE RED-CELL APLASIA/CHRONIC ANEMIA

Chronic B19V infection has been reported in a wide range of immunosuppressed patients, including those with congenital immunodeficiency, AIDS (Chap. 189), lymphoproliferative disorders (especially acute lymphocytic leukemia), and transplantation (Chap. 132). Patients have persistent anemia with reticulocytopenia, absent or low levels of B19V IgG, high titers of B19V DNA in serum, and—in many cases—scattered giant pronormoblasts in bone marrow. Rarely, nonerythroid hematologic lineages are also affected. Transient neutropenia, lymphopenia, and thrombocytopenia (including idiopathic thrombocytopenic purpura) have been observed. B19V occasionally causes a hemophagocytic syndrome.

A recent study in Papua New Guinea, where malaria is endemic, suggested that B19V infection plays a major role in the development of severe anemia. Further studies must determine whether B19V infection contributes to severe anemia in other malarial regions.

HYDROPS FETALIS

B19V infection during pregnancy can lead to hydrops fetalis and/or fetal loss. The risk of transplacental fetal infection is ~30%, and the risk of fetal loss (predominantly early in the second trimester) is ~9%. The risk of congenital infection is <1%. Although B19V does not appear to be teratogenic, anecdotal cases of eye damage and central nervous system (CNS) abnormalities have been reported. Cases of congenital anemia have also been described. B19V probably causes 10–20% of all cases of nonimmune hydrops.

UNUSUAL MANIFESTATIONS

B19V infection may rarely cause hepatitis, vasculitis, myocarditis, glomerulosclerosis, or meningitis. A variety of other cardiac manifestations, CNS diseases, and autoimmune infections have also been reported. However, B19V DNA can be detected by PCR for years in many tissues; this finding is of no known clinical significance, but its interpretation may cause confusion regarding B19V disease association.

Diagnosis

Diagnosis of B19V infection in immunocompetent individuals is generally based on detection of B19V IgM

antibodies **(Table 184-1)**. IgM can be detected at the time of rash in erythema infectiosum and by the third day of TAC in patients with hematologic disorders; these antibodies remain detectable for ~3 months. B19V IgG is detectable by the seventh day of illness and persists throughout life. Detection of B19V DNA should be used for the diagnosis of early TAC or chronic anemia. Although B19V levels fall rapidly with the development of the immune response, DNA can be detectable by PCR for months or even years after infection, even in healthy individuals; therefore, quantitative PCR should be used. In acute infection at the height of viremia, $>10^{12}$ B19V DNA IU/mL of serum can be detected; however, titers fall rapidly within 2 days. Patients with aplastic crisis or B19V-induced chronic anemia generally have $>10^{5}$ B19V DNA IU/mL.

Table 184-1 Diseases Associated with Human Parvovirus B19 Infection and Methods of Diagnosis

Disease	Host(s)	IgM	IgG	PCR	Quantitative PCR
Fifth disease	Healthy children	Positive	Positive	Positive	$>10^{3}$ IU/mL
Polyarthropathy syndrome	Healthy adults (more often women)	Positive within 3 months of onset	Positive	Positive	$>10^{3}$ IU/mL
Transient aplastic crisis	Patients with increased erythropoiesis	Negative/positive	Negative/positive	Positive	Often $>10^{12}$ IU/mL, but rapidly decreases
Persistent anemia/pure red-cell aplasia	Immunodeficient or immunocompetent patients	Negative/weakly positive	Negative/weakly positive	Positive	Often $>10^{12}$ IU/mL, but should be $>10^{6}$ in the absence of treatment
Hydrops fetalis/congenital anemia	Fetus (<20 weeks)	Negative/positive	Positive	Positive amniotic fluid or tissue	n/a

Abbreviations: IU, international units (1 IU equals ~1 genome); n/a, not applicable; PCR, polymerase chain reaction.

Treatment: Parvovirus B19 Infection

No antiviral drug effective against B19V is available, and treatment of B19V infection often targets symptoms only. TAC precipitated by B19V infection frequently necessitates symptom-based treatment with blood transfusions. In patients receiving chemotherapy, temporary cessation of treatment may result in an immune response and resolution. If this approach is unsuccessful or not applicable, commercial immune globulin (IVIg; Gammagard, Sandoglobulin) from healthy blood donors can cure or ameliorate persistent B19V infection in immunosuppressed patients. Generally, the dose used is 400 mg/kg daily for 5–10 days. Like patients with TAC, immunosuppressed patients with persistent B19V infection should be considered infectious. Administration of IVIg is not beneficial for erythema infectiosum or B19V-associated polyarthropathy. Intrauterine blood transfusion can prevent fetal loss in some cases of fetal hydrops.

Prevention

No vaccine has been approved for the prevention of B19V infection. A vaccine based on virus-like particles expressed in insect cells is under development; the results of phase 1 trials were promising.

PARV4/5

Definition

The PARV4 viral sequence was initially detected in a patient with an acute viral syndrome. Similar sequences, including the related PARV5 sequence, have been detected in pooled plasma collections. The DNA sequence of PARV4/5 is distinctly different from that of all other parvoviruses, and this virus cannot be classified within the current genera of Parvoviridae.

Epidemiology

Parv4 DNA is commonly found in plasma pools but at lower titers than those of B19V found before plasma pool screening. The higher levels of Parv4 DNA and IgG antibody in tissues (bone marrow and lymphoid tissue) and sera from IV drug users than in the corresponding specimens from control patients suggest that the virus is transmitted predominantly by parenteral means.

Clinical Manifestations

Parv4/5 infection has not been associated with any clinical disease to date.

HUMAN BOCAVIRUSES

Definition

Animal bocaviruses are associated with mild respiratory symptoms and enteritis in young animals. HBoV was originally identified in the respiratory tract of young children with lower respiratory tract infections. More recently, HBoV and the related viruses HBoV2 and HBoV3 have all been identified in human fecal samples.

Epidemiology

Seroepidemiologic studies with HBoV virus-like particles suggest that human bocavirus infection is common. Worldwide, most individuals are infected before the age of 5 years.

Clinical Manifestations

Although HBoV DNA is commonly found in respiratory secretions from children with acute respiratory infection, the role of this finding in pathogenesis is unknown, and HBoV sequences are often found in the presence of other pathogens. However, increasing evidence indicates that the virus is associated with wheezing in young children. The role of human bocaviruses in childhood gastroenteritis remains to be established.

FURTHER READINGS

Allander T et al: Cloning of a human parvovirus by molecular screening of respiratory tract samples. Proc Natl Acad Sci USA 102:12891, 2005[PMID: 16118271]

Arthur JL et al: A novel bocavirus associated with acute gastroenteritis in Australian children. PLoS Pathog 5:e1000391, 2009

Brown KE et al: Resistance to parvovirus B19 infection due to lack of virus receptor (erythrocyte P antigen). N Engl J Med 330:1192, 1994[PMID: 8139629]

——— et al: Erythrocyte P antigen: Cellular receptor for B19 parvovirus. Science 262:114, 1993

Fryer JF et al: Novel parvovirus and related variant in human plasma. Emerg Infect Dis 12:151, 2006[PMID: 16494735]

Kerr JR et al: *Parvoviruses*. London, Hodder Arnold, 2006

Kurtzman GJ et al: Chronic bone marrow failure due to persistent B19 parvovirus infection. N Engl J Med 317:287, 1987[PMID: 3037373]

Schildgen O et al: Human bocavirus: Passenger or pathogen in acute respiratory tract infections? Clin Microbiol Rev 21:291, 2008[PMID: 18400798]

Young NS, Brown KE: Parvovirus B19. N Engl J Med 350:586, 2004[PMID: 14762186]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HUMAN PAPILLOMAVIRUS INFECTIONS: INTRODUCTION

Definition

Human papillomaviruses (HPVs) selectively infect the epithelium of skin and mucous membranes. These infections may be asymptomatic, produce warts, or be associated with a variety of both benign and malignant neoplasias.

Etiologic Agent

Papillomaviruses constitute the *Papillomavirus* genus of the family Papillomaviridae. They are nonenveloped, measure 50–55 nm in diameter, have icosahedral capsids composed of 72 capsomeres, and contain a double-strand circular DNA genome of ~7900 base pairs. The genomic organization of all papillomaviruses is similar and consists of an early (E) region, a late (L) region, and a noncoding upstream regulatory region (URR). Oncogenic HPV types can immortalize human keratinocytes, and this activity has been mapped to products of early genes E6 and E7. E6 protein facilitates the degradation of the p53 tumor-suppressor protein, and E7 protein binds the retinoblastoma gene product and related proteins. The E1 and E2 proteins modulate viral DNA replication and regulate gene expression. The L1 gene codes for the major capsid protein, which makes up 80% of the virion mass. L2 codes for a minor capsid protein. Type-specific conformational antigenic determinants are located on the virion surface. Papillomavirus types are distinguished from one another by the degree of nucleic acid sequence homology. Distinct types share <90% of their DNA sequences in L1. More than 100 HPV types are recognized, and individual types are associated with specific clinical manifestations. For example, HPV-1 causes plantar warts, HPV-6 causes anogenital warts, and HPV-16 infection can produce cervical dysplasia and invasive cervical cancer. HPVs are species-specific and have not been propagated in routine tissue culture or in common experimental animals. However, some HPV types have been propagated in organotypic culture systems, and some have been produced in human tissues implanted in immunodeficient mice.

Epidemiology

There are few good studies of the incidence or prevalence of human warts in well-defined populations. Common warts (*verruca vulgaris*) are found in as many as 25% of some groups and are most prevalent among young children. Plantar warts (*verruca plantaris*) are also widely prevalent; they occur most often among adolescents and young adults. Anogenital warts (*condyloma acuminatum*) represent one of the most common sexually transmitted diseases in the United States. HPV infection of the uterine cervix produces the squamous cell abnormalities most frequently detected on Papanicolaou smears.

Most anogenital HPV infections are transmitted through direct contact with infectious lesions. However, lesion characteristics that are associated with transmission, including appearance, have not been defined, and individuals without obvious disease may transmit infection. Close personal contact is also assumed to play a role in the transmission of most cutaneous warts; the importance of fomites in this setting is not clear. Minor trauma at the site of inoculation may facilitate transmission. Recurrent respiratory papillomatosis in young children is an uncommon disease that is acquired from the infected maternal genital tract. In adults, orogenital sexual contact may transmit the disease.

A large body of epidemiologic and biologic data has established that some HPV infections cause cervical cancer. For example, >95% of cervical cancers contain HPV DNA of oncogenic (high-risk) types, such as 16, 18, 31, 33, and 45. HPV DNA is also present in the precursor lesions of cervical cancer (cervical intraepithelial neoplasias). Such lesions containing DNA of oncogenic types are more likely to progress than those associated with low-risk HPV types, such as 6 and 11. HPV DNA is transcribed in tumor tissues, and many epidemiologic studies have confirmed a strong relationship between HPV infection (with or without cofactors) and the development of cervical cancer. Definitive proof of the causative role of high-risk HPV types in the pathogenesis of high-grade cervical dysplasia has been provided by the results of recently conducted trials of HPV vaccines. However, it is important to realize that most cervical HPV infections, including those caused by high-risk types, are self-limited. Infection with high-risk HPV types has also been associated with squamous cell carcinomas and dysplasias of the penis, anus, vagina, and vulva. HPV infection may play a role in squamous cell carcinomas of the head and neck. In patients with *epidermodysplasia verruciformis* (see "Clinical Manifestations," below), squamous cell cancers develop frequently at sites infected with specific HPV types, including 5 and 8.

Clinical Manifestations

The clinical manifestations of HPV infection depend on the location of lesions and the type of virus. Common warts usually occur on the hands as flesh-colored to brown, exophytic, and hyperkeratotic papules. Plantar warts may be quite painful; they can be differentiated from calluses by paring of the surface to reveal thrombosed capillaries. Flat warts (*verruca plana*) are most common among children and occur on the face, neck, chest, and flexor surfaces of the forearms and legs.

Anogenital warts develop on the skin and mucosal surfaces of external genitalia and perianal areas (Fig. 185-1). Among circumcised men, warts are most commonly found on the penile shaft. Lesions frequently occur at the urethral meatus and may extend proximally. Receptive anal intercourse predisposes both men and women to the development of perianal warts, but such lesions occasionally develop without such a history. In women, warts appear first at the posterior introitus and on the adjacent labia. They then spread to other parts of the vulva and commonly involve the vagina and cervix. In both sexes, external warts suggest the presence of internal lesions; however, internal lesions may be present without external warts, particularly in women. The differential diagnosis of anogenital warts includes condylomata lata of secondary syphilis, molluscum contagiosum, hirsutoid papillomatosis (pearly penile papules), fibroepitheliomas, and a variety of benign and malignant mucocutaneous neoplasms. Respiratory papillomatosis in young children, which may be life-threatening, presents as hoarseness, stridor, or respiratory distress. The disease in adults is usually milder.

Figure 185-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anogenital warts are lesions produced by human papillomavirus and in this patient are seen as multiple verrucous papules coalescing into plaques.

Immunosuppressed patients, particularly those undergoing organ transplantation, often develop pityriasis versicolor–like lesions, from which DNA of several HPV types has been extracted. Occasionally, such lesions appear to undergo malignant transformation. Patients infected with HIV are often infected with uncommon HPV types, frequently have severe clinical manifestations of HPV infection, and are at high risk for cervical and anal dysplasia as well as for invasive cancer. HPV disease in patients with HIV infection may be associated with multiple HPV types, is difficult to treat, and often recurs (Chap. 189).

Epidermodysplasia verruciformis is a rare autosomal recessive disease characterized by an inability to control HPV infection. Patients are often infected with unique HPV types (i.e., types that affect only this group) and frequently develop cutaneous squamous cell malignancies, particularly in sun-exposed areas. The lesions resemble flat warts or macules similar to those of pityriasis versicolor.

The complications of warts include itching and occasionally bleeding. In rare cases, warts become secondarily infected with bacteria or fungi. Large masses of warts may cause mechanical problems,

such as obstruction of the birth canal or the urinary tract. Dysplasias of the uterine cervix are generally asymptomatic until frank carcinoma develops. Patients with anogenital HPV disease may develop serious psychological symptoms due to anxiety and depression over this condition.

Pathogenesis

The incubation period of HPV disease is usually 3–4 months (range, 1 month to 2 years). All types of squamous epithelium can be infected by HPV, and the gross and histologic appearances of individual lesions vary with the site of infection and the type of virus. The replication of HPV begins with infection of basal cells. As cellular differentiation proceeds, HPV DNA replicates and is transcribed. Ultimately, virions are assembled in the nucleus and released when keratinocytes are shed. This process is associated with proliferation of all epidermal layers except the basal layer and produces acanthosis, parakeratosis, and hyperkeratosis. Koilocytes—large round cells with pyknotic nuclei—appear in the granular layer. Histologically normal epithelium may contain HPV DNA, and residual DNA after treatment can be associated with recurrent disease.

Episomal HPV DNA is present in the nuclei of infected cells in benign lesions caused by HPV. However, in severe dysplasias and cancers, HPV DNA is generally integrated, with disruption of the E1/E2 open reading frames. This disruption leads to upregulation of E6 and E7 and subsequent interference with cellular tumor-suppressor proteins. Expression of E6 and E7 proteins of oncogenic HPV types is necessary for the development and maintenance of the transformed state in both cervical cancers and cell lines derived from these tumors.

Host defense responses to HPV infection remain incompletely understood. However, several studies of recently developed HPV vaccines have demonstrated that production of high titers of type-specific neutralizing antibodies by vaccinated individuals is associated with type-specific protection from HPV infection and disease. Because patients with defects in cell-mediated immune responses (including transplant recipients and patients with HIV infection) frequently develop severe HPV disease, such responses are probably important for the control of established virus replication and disease. Histologic studies demonstrating an epidermal lymphomonocytic infiltrate in resolving warts suggest that local immunity may be of particular importance in the resolution of disease. HPV infection also elicits a detectable serologic response in many patients. Using HPV virus-like particles (VLPs) as antigens, type-specific antibodies can be found in sera of about two-thirds of patients with anogenital infection. Antibodies to E-region proteins, most notably E7, have been detected among patients with cervical carcinoma.

Diagnosis

Most warts that are visible to the naked eye can be diagnosed correctly by history and physical examination alone. The use of a colposcope is invaluable in assessing vaginal and cervical lesions and is helpful in the diagnosis of oral and cutaneous HPV disease as well. Application of 3–5% solutions of acetic acid may aid in the visualization of lesions, although the sensitivity and specificity of this procedure are unknown. Papanicolaou smears prepared from cervical or anal scrapings often show cytologic evidence of HPV infection. Persistent or atypical lesions should be biopsied and examined by routine histologic methods. The most sensitive and specific methods of virologic diagnosis use techniques such as the polymerase chain reaction or the hybrid capture assay to detect HPV nucleic acids and to identify specific virus types. Such tests may be useful in the diagnosis and management of cervical HPV disease, although their utility may vary according to the prevalence of disease and the availability of traditional cytologic and histologic testing. Serologic techniques to diagnose HPV infection are not helpful in individual cases and are not widely available.

Treatment: Human Papillomavirus Infections

(Table 185-1) Decisions regarding the initiation of therapy should be made with the recognition that currently available modes of treatment are not completely effective and some have significant side effects. In addition, treatment may be expensive, and many HPV lesions resolve spontaneously. Frequently used therapies include cryosurgery, application of caustic agents, electrodesiccation, surgical excision, and ablation with a laser. Topical antimetabolites such as 5-fluorouracil have also been used. Both failure and recurrence have been well documented with all of these methods of treatment. Cryosurgery is the initial treatment of choice for condyloma acuminatum. Topically applied podophyllum preparations as well as podofilox may also be used. Various interferon preparations have been employed with modest success in the treatment of respiratory papillomatosis and condyloma acuminatum. A topically applied interferon inducer, imiquimod, is also of benefit in the treatment of condyloma acuminatum. The diagnosis and management of anogenital dysplasias and of internal anogenital warts require special skills and resources, and patients with such lesions should be referred to a qualified specialist.

Table 185-1 Treatment of External, Exophytic Anogenital Warts
I. Administered by provider
A. Cryotherapy with liquid nitrogen or cryoprobe weekly

B. Podophyllin resin, 10–25% weekly for up to 4 weeks
C. Trichloroacetic acid or bichloroacetic acid, 80–90% weekly
D. Surgical excision
E. Other regimens
1. Intralesionally administered interferon 2. Laser surgery
II. Administered by patient
A. Podofilox, 0.5% solution or gel twice daily for 3 days, followed by 4 days without therapy. This cycle may be repeated four times.
B. Imiquimod, 5% cream 3 times per week for up to 16 weeks

Source: Modified from Centers for Disease Control and Prevention: MMWR Recomm Rep 55(RR-11):1, 2006 *(www.cdc.gov/mmwr/preview/mmwrhtml/rr5511a1/htm?)*.

Prevention

Recently developed HPV VLP vaccines dramatically reduce rates of infection and disease produced by the HPV types in the vaccines. These products are directed against virus types that cause anogenital tract disease and are derived from expression of the major capsid protein (L1) gene in tissue culture. When expressed using appropriate vectors and tissue culture systems, L1 self-assembles into a VLP that cannot be distinguished morphologically or antigenically from its wild-type counterpart (Fig. 185-2) but that contains no viral nucleic acid. To date, one quadrivalent product (Gardasil, Merck) containing HPV types 6, 11, 16, and 18 and one bivalent product (Cervarix, GlaxoSmithKline) containing HPV types 16 and 18 have been licensed in the United States. HPV types 6 and 11 cause 90% of anogenital warts, whereas types 16 and 18 are responsible for 70% of cervical cancers. Both vaccines are highly immunogenic, as measured by serum antibody titers after vaccination. Vaccineefficacy has varied according to the immunologic and virologic characteristics of study populations at baseline and according to the endpoints evaluated. Among study participants who are shown at baseline not to be infected with a specific virus type contained in the vaccine and who adhere to the study protocol, rates of vaccine efficacy regularly exceed 90%, as measured by both infection and disease caused by that specific virus type. Study participants who are already infected at baseline with a specific virus type contained in the vaccine do not benefit from vaccination against that type but may benefit from vaccination against other virus types contained in the vaccine preparation. Thus available HPV vaccines have potent prophylactic effects but no therapeutic effects. The Advisory Committee on Immunization Practices (ACIP) of the Centers for Disease Control and Prevention has recommended that HPV vaccination be routinely offered to girls and young women 9–26 years of age. The quadrivalent vaccine has also been licensed in the United States for use in males; the ACIP has stated that this product may be used to prevent anogenital warts in boys and young men 9–26 years of age. Because 30% of cervical cancers are caused by HPV types not contained in the vaccines, no changes in cervical cancer screening programs are currently recommended. Barrier methods of contraception may also be helpful in preventing transmission of condyloma acuminatum and other anogenital HPV-associated diseases. Methods to prevent other HPV infections are limited to avoidance of contact with infectious lesions.

Figure 185-2

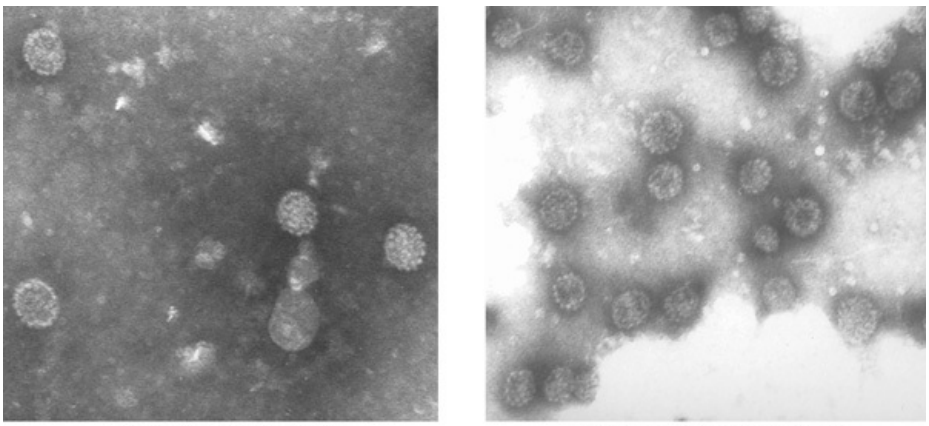

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

HPV-11 virus-like particles produced in insect cells (*right*) are morphologically and antigenically

indistinguishable from wild-type HPV-11 particles (*left*). *(Images courtesy of Drs. William Bonnez and Robert C. Rose; with permission.)*

FURTHER READINGS

Advisory Committee on Immunization Practices: Recommended adult immunization schedule: United States, 2010. Ann Intern Med 152:36, 2010

Bonnez W, Reichman RC: Papillomaviruses, in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). Churchill Livingstone, Philadelphia, 2010, pp 2035–2049

Centers for Disease Control and Prevention: Sexually transmitted diseases treatment guidelines, 2006. MMWR Recomm Rep 55:1, 2006 (*www.cdc.gov/mmwr/preview/mmwrhtml/rr5511a1.htm*)

D'souza G et al: Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med 356:1944, 2007[PMID: 21719359]

Future ii Study Group: Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N Engl J Med 356:1915, 2007

Garland SM et al: Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. N Engl J Med 356:1928, 2007[PMID: 17494926]

Markowitz LE et al: Seroprevalence of human papillomavirus types 6, 11, 16, and 18 in the United States: National Health and Nutrition Examination Survey 2003–2004. J Infect Dis 200:1059, 2009[PMID: 19719390]

Paavonen J et al: Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): Final analysis of a double-blind, randomised study in young women. Lancet 374:301, 2009[PMID: 19586656]

Schiffman M et al: Human papillomavirus and cervical cancer. Lancet 370:890, 2007[PMID: 17826171]

Sycuro LK et al: Persistence of genital human papillomavirus infection in a long-term follow-up study of female university students. J Infect Dis 198:971, 2008[PMID: 18694334]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

COMMON VIRAL RESPIRATORY INFECTIONS: INTRODUCTION

General Considerations

Acute viral respiratory illnesses are among the most common of human diseases, accounting for one-half or more of all acute illnesses. The incidence of acute respiratory disease in the United States is 3–5.6 cases per person per year. The rates are highest among children <1 year old (6.1–8.3 cases per year) and remain high until age 6, when a progressive decrease begins. Adults have 3–4 cases per person per year. Morbidity from acute respiratory illnesses accounts for 30–50% of time lost from work by adults and for 60–80% of time lost from school by children. The use of antibacterial agents to treat viral respiratory infections represents a major source of abuse of that category of drugs.

It has been estimated that two-thirds to three-fourths of cases of acute respiratory illnesses are caused by viruses. More than 200 antigenically distinct viruses from 10 genera have been reported to cause acute respiratory illness, and it is likely that additional agents will be described in the future. The vast majority of these viral infections involve the upper respiratory tract, but lower respiratory tract disease can also develop, particularly in younger age groups, in the elderly, and in certain epidemiologic settings.

The illnesses caused by respiratory viruses traditionally have been divided into multiple distinct syndromes, such as the "common cold," pharyngitis, croup (laryngotracheobronchitis), tracheitis, bronchiolitis, bronchitis, and pneumonia. Each of these general categories of illness has a certain epidemiologic and clinical profile; for example, croup occurs exclusively in very young children and has a characteristic clinical course. Some types of respiratory illness are more likely to be associated with certain viruses (e.g., the common cold with rhinoviruses), while others occupy characteristic epidemiologic niches (e.g., adenovirus infections in military recruits). The syndromes most commonly associated with infections with the major respiratory virus groups are summarized in **Table 186-1**. Most respiratory viruses clearly have the potential to cause more than one type of respiratory illness, and features of several types of illness may be found in the same patient. Moreover, the clinical illnesses induced by these viruses are rarely sufficiently distinctive to permit an etiologic diagnosis on clinical grounds alone, although the epidemiologic setting increases the likelihood that one group of viruses rather than another is involved. In general, laboratory methods must be relied on to establish a specific viral diagnosis.

Table 186-1 Illnesses Associated with Respiratory Viruses

	Frequency of Respiratory Syndromes		
Virus	**Most Frequent**	**Occasional**	**Infrequent**
Rhinoviruses	Common cold	Exacerbation of chronic bronchitis and asthma	Pneumonia in children
Coronaviruses[a]	Common cold	Exacerbation of chronic bronchitis and asthma	Pneumonia and bronchiolitis
Human respiratory syncytial virus	Pneumonia and bronchiolitis in young children	Common cold in adults	Pneumonia in elderly and immunosuppressed patients
Parainfluenza viruses	Croup and lower respiratory tract disease in young children	Pharyngitis and common cold	Tracheobronchitis in adults; lower respiratory tract disease in immunosuppressed patients
Adenoviruses	Common cold and pharyngitis in children	Outbreaks of acute respiratory disease in military recruits[b]	Pneumonia in children; lower respiratory tract and disseminated disease in immunosuppressed patients
Influenza A viruses	Influenza[c]	Pneumonia and excess mortality in high-risk patients	Pneumonia in healthy individuals
Influenza B viruses	Influenza[c]	Rhinitis or	Pneumonia

	Influenza[c]	pharyngitis alone	
Enteroviruses	Acute undifferentiated febrile illnesses[d]	Rhinitis or pharyngitis alone	Pneumonia
Herpes simplex viruses	Gingivostomatitis in children; pharyngotonsillitis in adults	Tracheitis and pneumonia in immunocompromised patients	Disseminated infection in immunocompromised patients
Human metapneumoviruses	Upper and lower respiratory tract disease in children	Upper respiratory tract illness in adults	Pneumonia in elderly and immunosuppressed patients

[a]SARS-associated coronavirus (SARS-CoV) caused epidemics of pneumonia from November 2002 to July 2003 (see text).

[b]Serotypes 4 and 7.

[c]Fever, cough, myalgia, malaise.

[d]May or may not have a respiratory component.

This chapter reviews viral infections caused by six of the major groups of respiratory viruses: rhinoviruses, coronaviruses, respiratory syncytial viruses, metapneumoviruses, parainfluenza viruses, and adenoviruses. The extraordinary outbreaks of lower respiratory tract disease associated with coronaviruses (severe acute respiratory syndrome, or SARS) in 2002–2003 are also discussed. **Influenza viruses, which are a major cause of death as well as morbidity, are reviewed in Chap. 187. Herpesviruses, which occasionally cause pharyngitis and which also cause lower respiratory tract disease in immunosuppressed patients, are reviewed in Chap. 179. Enteroviruses, which account for occasional respiratory illnesses during the summer months, are reviewed in Chap. 191.**

RHINOVIRUS INFECTIONS

Etiologic Agent

Rhinoviruses are members of the Picornaviridae family, small (15- to 30-nm) nonenveloped viruses that contain a single-stranded RNA genome and have been divided into three genetic species: HRV-A, HRV-B, and HRV-C. In contrast to other members of the picornavirus family, such as enteroviruses, rhinoviruses are acid-labile and are almost completely inactivated at pH ≤ 3. Rhinoviruses grow preferentially at 33°–34°C (the temperature of the human nasal passages) rather than at 37°C (the temperature of the lower respiratory tract). Of the 102 recognized serotypes of rhinovirus, 91 use intercellular adhesion molecule 1 (ICAM-1) as a cellular receptor and constitute the "major" receptor group, 10 use the low-density lipoprotein receptor (LDLR) and constitute the "minor" receptor group, and 1 uses decay-accelerating factor.

Epidemiology

Rhinoviruses are a prominent cause of the common cold and have been detected in up to 50% of common cold–like illnesses by tissue culture and polymerase chain reaction (PCR) techniques. Overall rates of rhinovirus infection are higher among infants and young children and decrease with increasing age. Rhinovirus infections occur throughout the year, with seasonal peaks in early fall and spring in temperate climates. These infections are most often introduced into families by preschool or grade-school children <6 years old. Of initial illnesses in family settings, 25–70% are followed by secondary cases, with the highest attack rates among the youngest siblings at home. Attack rates also increase with family size.

Rhinoviruses appear to spread through direct contact with infected secretions, usually respiratory droplets. In some studies of volunteers, transmission was most efficient by hand-to-hand contact, with subsequent self-inoculation of the conjunctival or nasal mucosa. Other studies demonstrated transmission by large- or small-particle aerosol. Virus can be recovered from plastic surfaces inoculated 1–3 h previously; this observation suggests that environmental surfaces contribute to transmission. In studies of married couples in which neither partner had detectable serum antibody, transmission was associated with prolonged contact (≥122 h) during a 7-day period. Transmission was infrequent unless (1) virus was recoverable from the donor's hands and nasal mucosa, (2) at least 1000 $TCID_{50}$ of virus was present in nasal washes from the donor, and (3) the donor was at least moderately symptomatic with the "cold." Despite anecdotal observations, exposure to cold temperatures, fatigue, and sleep deprivation have not been associated with increased rates of rhinovirus-induced illness in volunteers, although some studies have suggested that psychologically defined "stress" may contribute to development of symptoms.

Infection with rhinoviruses is worldwide in distribution. By adulthood, nearly all individuals have neutralizing antibodies to multiple serotypes, although the prevalence of antibody to any one serotype varies widely. Multiple serotypes circulate simultaneously, and generally no single serotype or group of serotypes has been more prevalent than the others.

Pathogenesis

Rhinoviruses infect cells through attachment to specific cellular receptors; as mentioned above, most serotypes attach to ICAM-1, while a few use LDLR. Relatively limited information is available on the histopathology and pathogenesis of acute rhinovirus infections in humans. Examination of biopsy specimens obtained during experimentally induced and naturally occurring illness indicates that the nasal mucosa is edematous, is often hyperemic, and—during acute illness—is covered by a mucoid discharge. There is a mild infiltrate with inflammatory cells, including neutrophils, lymphocytes, plasma cells, and eosinophils. Mucus-secreting glands in the submucosa appear hyperactive; the nasal turbinates are engorged, a condition that may lead to obstruction of nearby openings of sinus cavities. Several mediators—e.g., bradykinin; lysylbradykinin; prostaglandins; histamine; interleukins 1β, 6, and 8; and tumor necrosis factor α—have been linked to the development of signs and symptoms in rhinovirus-induced colds.

The incubation period for rhinovirus illness is short, generally 1–2 days. Virus shedding coincides with the onset of illness or may begin shortly before symptoms develop. The mechanisms of immunity to rhinovirus infection are not well worked out. In some studies, the presence of homotypic antibody has been associated with significantly reduced rates of subsequent infection and illness, but data conflict regarding the relative importance of serum and local antibody in protection from rhinovirus infection.

Clinical Manifestations

The most common clinical manifestations of rhinovirus infections are those of the common cold. Illness usually begins with rhinorrhea and sneezing accompanied by nasal congestion. The throat is frequently sore, and in some cases sore throat is the initial complaint. Systemic signs and symptoms, such as malaise and headache, are mild or absent, and fever is unusual. Illness generally lasts for 4–9 days and resolves spontaneously without sequelae. In children, bronchitis, bronchiolitis, and bronchopneumonia have been reported; nevertheless, it appears that rhinoviruses are not major causes of lower respiratory tract disease in children. Rhinoviruses may cause exacerbations of asthma and chronic pulmonary disease in adults. The vast majority of rhinovirus infections resolve without sequelae, but complications related to obstruction of the eustachian tubes or sinus ostia, including otitis media or acute sinusitis, can develop. In immunosuppressed patients, particularly bone marrow transplant recipients, severe and even fatal pneumonias have been associated with rhinovirus infections.

Diagnosis

Although rhinoviruses are the most frequently recognized cause of the common cold, similar illnesses are caused by a variety of other viruses, and a specific viral etiologic diagnosis cannot be made on clinical grounds alone. Rather, rhinovirus infection is diagnosed by isolation of the virus from nasal washes or nasal secretions in tissue culture. In practice, this procedure is rarely undertaken because of the benign, self-limited nature of the illness. In most settings, detection of rhinovirus RNA by PCR is more sensitive than that by tissue culture; however, PCR for rhinoviruses is largely a research procedure. Given the many serotypes of rhinovirus, diagnosis by serum antibody tests is currently impractical. Likewise, common laboratory tests, such as white blood cell count and erythrocyte sedimentation rate, are not helpful.

Treatment: Rhinovirus Infections

Because rhinovirus infections are generally mild and self-limited, treatment is not usually necessary. Therapy in the form of first-generation antihistamines and nonsteroidal anti-inflammatory drugs may be beneficial in patients with particularly pronounced symptoms, and an oral decongestant may be added if nasal obstruction is particularly troublesome. Reduction of activity is prudent in instances of significant discomfort or fatigability. Antibacterial agents should be used only if bacterial complications such as otitis media or sinusitis develop. Specific antiviral therapy is not available.

Prevention

Intranasal application of interferon sprays has been effective in the prophylaxis of rhinovirus infections but is also associated with local irritation of the nasal mucosa. Studies of prevention of rhinovirus infection by blocking of ICAM-1 or by drug binding to parts of the viral capsid (pleconaril) have yielded mixed results. Experimental vaccines to certain rhinovirus serotypes have been generated, but their usefulness is questionable because of the myriad serotypes and the uncertainty about mechanisms of immunity. Thorough hand washing, environmental decontamination, and protection against autoinoculation may help to reduce rates of transmission of infection.

CORONAVIRUS INFECTIONS

Etiologic Agent

Coronaviruses are pleomorphic, single-stranded RNA viruses that measure 100–160 nm in diameter. The name derives from the crownlike appearance produced by the club-shaped projections that stud the viral envelope. Coronaviruses infect a wide variety of animal species and have been divided into three antigenic and genetic groups. Before the emergence of the coronavirus associated with SARS (SARS-CoV), coronaviruses recognized as causes of infection in humans fell into groups 1 and 2, which include human isolates HCoV-229E and HCoV-OC43, respectively. SARS-CoV was at first believed to represent a novel group but now is considered to be a distantly related member of group 2. The SARS-CoV strains that have been fully sequenced have shown only minimal variation.

In general, human coronaviruses have been difficult to cultivate in vitro, and some strains grow only in human tracheal organ cultures rather than in tissue culture. SARS-CoV is an exception whose ready growth in African green monkey kidney (Vero E6) cells greatly facilitates its study.

Epidemiology

Generally, human coronavirus infections are present throughout the world. Seroprevalence studies of strains HCoV-229E and HCoV-OC43 have demonstrated that serum antibodies are acquired early in life and increase in prevalence with advancing age, so that >80% of adult populations have antibodies as measured by enzyme-linked immunosorbent assay (ELISA). Overall, coronaviruses account for 10–35% of common colds, depending on the season. Coronavirus infections appear to be particularly prevalent in late fall, winter, and early spring—times when rhinovirus infections are less common.

An extraordinary outbreak of the coronavirus-associated illness known as SARS occurred in 2002–2003. The outbreak apparently began in southern China and eventually resulted in 8096 recognized cases in 28 countries in Asia, Europe, and North and South America; ~90% of cases occurred in China and Hong Kong. The natural reservoir of SARS-CoV appeared to be the horseshoe bat, and the outbreak may have originated from human contact with infected semidomesticated animals such as the palm civet. In most cases, however, the infection was transmitted from human to human. Case-fatality rates varied among outbreaks, with an overall figure of ~9.5%. The disease appeared to be somewhat milder in cases in the United States and was clearly less severe among children. The outbreak ceased in 2003; 17 cases were detected in 2004, mostly in laboratory-associated settings, and no cases were reported in 2005–2009.

The mechanisms of transmission of SARS are incompletely understood. Clusters of cases suggest that spread may occur by both large and small aerosols and perhaps by the fecal-oral route as well. The outbreak of illness in a large apartment complex in Hong Kong suggested that environmental sources, such as sewage or water, may also play a role in transmission. Some ill individuals ("super-spreaders") appeared to be hyperinfectious and were capable of transmitting infection to 10–40 contacts, although most infections resulted in spread either to no one or to three or fewer individuals.

Pathogenesis

Coronaviruses that cause the common cold (e.g., strains HCoV-229E and HCoV-OC43) infect ciliated epithelial cells in the nasopharynx via the aminopeptidase N receptor (group 1) or a sialic acid receptor (group 2). Viral replication leads to damage of ciliated cells and induction of chemokines and interleukins, with consequent common-cold symptoms similar to those induced by rhinoviruses.

SARS-CoV infects cells of the respiratory tract via the angiotensin-converting enzyme 2 receptor. The result is a systemic illness in which virus is also found in the bloodstream, in the urine, and (for up to 2 months) in the stool. Virus persists in the respiratory tract for 2–3 weeks, and titers peak ~10 days after the onset of systemic illness. Pulmonary pathology consists of hyaline membrane formation, desquamation of pneumocytes in alveolar spaces, and an interstitial infiltrate made up of lymphocytes and mononuclear cells. Giant cells are frequently seen, and coronavirus particles have been detected in type II pneumocytes. Elevated levels of proinflammatory cytokines and chemokines have been detected in sera from patients with SARS.

Clinical Manifestations

After an incubation period that generally lasts 2–7 days (range, 1–14 days), SARS usually begins as a systemic illness marked by the onset of fever, which is often accompanied by malaise, headache, and myalgias and is followed in 1–2 days by a nonproductive cough and dyspnea. Approximately 25% of patients have diarrhea. Chest x-rays can show a variety of infiltrates, including patchy areas of consolidation—most frequently in peripheral and lower lung fields—or interstitial infiltrates, which can progress to diffuse involvement.

In severe cases, respiratory function may worsen during the second week of illness and progress to frank adult respiratory distress syndrome accompanied by multiorgan dysfunction. Risk factors for severe disease include an age of >50 years and comorbidities such as cardiovascular disease, diabetes, or hepatitis. Illness in pregnant women may be particularly severe, but SARS-CoV infection appears to be milder in children than in adults.

The clinical features of common colds caused by human coronaviruses are similar to those of illness caused by rhinoviruses. In studies of volunteers, the mean incubation period of colds induced by coronaviruses (3 days) is somewhat longer than that of illness caused by rhinoviruses, and the

duration of illness is somewhat shorter (mean, 6–7 days). In some studies, the amount of nasal discharge was greater in colds induced by coronaviruses than in those induced by rhinoviruses. Coronaviruses other than SARS-CoV have been recovered occasionally from infants with pneumonia and from military recruits with lower respiratory tract disease and have been associated with worsening of chronic bronchitis. Two novel coronaviruses, HCoV-NL63 (group 1) and HCoV-HKU1 (group 2), have been isolated from patients hospitalized with acute respiratory illness. Their overall role as causes of human respiratory disease remains to be determined.

Laboratory Findings and Diagnosis

Laboratory abnormalities in SARS include lymphopenia, which is present in ~50% of cases and which mostly affects CD4+ T cells but also involves CD8+ T cells and natural killer cells. Total white blood cell counts are normal or slightly low, and thrombocytopenia may develop as the illness progresses. Elevated serum levels of aminotransferases, creatine kinase, and lactate dehydrogenase have been reported.

A rapid diagnosis of SARS-CoV infection can be made by reverse-transcription PCR (RT-PCR) of respiratory tract samples and plasma early in illness and of urine and stool later on. SARS-CoV can also be grown from respiratory tract samples by inoculation into Vero E6 tissue culture cells, in which a cytopathic effect is seen within days. RT-PCR appears to be more sensitive than tissue culture, but only around one-third of cases are positive by PCR at initial presentation. Serum antibodies can be detected by ELISA or immunofluorescence, and nearly all patients develop detectable serum antibodies within 28 days after the onset of illness.

Laboratory diagnosis of coronavirus-induced colds is rarely required. Coronaviruses that cause those illnesses are frequently difficult to cultivate in vitro but can be detected in clinical samples by ELISA or immunofluorescence assays or by RT-PCR for viral RNA. These research procedures can be used to detect coronaviruses in unusual clinical settings.

Treatment: Coronavirus Infections

There is no specific therapy of established efficacy for SARS. Although ribavirin has frequently been used, it has little if any activity against SARS-CoV in vitro, and no beneficial effect on the course of illness has been demonstrated. Because of suggestions that immunopathology may contribute to the disease, glucocorticoids have also been widely used, but their benefit, if any, is likewise unestablished. Supportive care to maintain pulmonary and other organ-system functions remains the mainstay of therapy.

The approach to the treatment of common colds caused by coronaviruses is similar to that discussed above for rhinovirus-induced illnesses.

Prevention

The recognition of SARS led to a worldwide mobilization of public health resources to apply infection control practices to contain the disease. Case definitions were established, travel advisories were proposed, and quarantines were imposed in certain locales. As of this writing, no additional cases of SARS have been reported since 2004. However, it remains unknown whether the disappearance of cases is a result of control measures, whether it is part of a seasonal or otherwise unexplained epidemiologic pattern of SARS, or when or whether SARS might reemerge. The U.S. Centers for Disease Control and Prevention and the World Health Organization maintain recommendations for surveillance and assessment of potential cases of SARS (www.cdc.gov/ncidod/sars/). The frequent transmission of the disease to health care workers makes it mandatory that strict infection-control practices be employed by health care facilities to prevent airborne, droplet, and contact transmission from any suspected cases of SARS. Health care workers who enter areas in which patients with SARS may be present should don gowns, gloves, and eye and respiratory protective equipment (e.g., an N95 filtering facepiece respirator certified by the National Institute for Occupational Safety and Health).

Vaccines have been developed against several animal coronaviruses but not against known human coronaviruses. The emergence of SARS-CoV has stimulated interest in the development of vaccines against such agents.

HUMAN RESPIRATORY SYNCYTIAL VIRUS INFECTIONS

Etiologic Agent

Human respiratory syncytial virus (HRSV) is a member of the Paramyxoviridae family (genus *Pneumovirus*). An enveloped virus ~150–350 nm in diameter, HRSV is so named because its replication in vitro leads to the fusion of neighboring cells into large multinucleated syncytia. The single-stranded RNA genome codes for 11 virus-specific proteins. Viral RNA is contained in a helical nucleocapsid surrounded by a lipid envelope bearing two glycoproteins: the G protein, by which the virus attaches to cells, and the F (fusion) protein, which facilitates entry of the virus into the cell by fusing host and viral membranes. HRSV is considered to be of a single antigenic type, but two distinct subgroups (A and B) and multiple subtypes within each subgroup have now been described. Antigenic diversity is reflected by differences in the G protein, while the F protein is highly conserved. Both

antigenic groups can circulate simultaneously in outbreaks, although there are typically alternating patterns in which one subgroup predominates over 1- to 2-year periods.

Epidemiology

HRSV is a major respiratory pathogen of young children and the foremost cause of lower respiratory disease in infants. Infection with HRSV is seen throughout the world in annual epidemics that occur in late fall, winter, or spring and last up to 5 months. The virus is rarely encountered during the summer. Rates of illness are highest among infants 1–6 months of age, peaking at 2–3 months of age. The attack rates among susceptible infants and children are extraordinarily high, approaching 100% in settings such as day-care centers where large numbers of susceptible infants are present. By age 2, virtually all children will have been infected with HRSV. HRSV accounts for 20–25% of hospital admissions of young infants and children for pneumonia and for up to 75% of cases of bronchiolitis in this age group. It has been estimated that more than half of infants who are at risk will become infected during an HRSV epidemic.

In older children and adults, reinfection with HRSV is frequent but disease is milder than in infancy. A common cold–like syndrome is the illness most commonly associated with HRSV infection in adults. Severe lower respiratory tract disease with pneumonitis can occur in elderly (often institutionalized) adults and in patients with immunocompromising disorders or treatment, including recipients of stem cell and solid-organ transplants. HRSV is also an important nosocomial pathogen; during an outbreak, it can infect pediatric patients and up to 25–50% of the staff on pediatric wards. The spread of HRSV among families is efficient: up to 40% of siblings may become infected when the virus is introduced into the family setting.

HRSV is transmitted primarily by close contact with contaminated fingers or fomites and by self-inoculation of the conjunctiva or anterior nares. Virus may also be spread by coarse aerosols produced by coughing or sneezing, but it is inefficiently spread by fine-particle aerosols. The incubation period is ~4–6 days, and virus shedding may last for ≥2 weeks in children and for shorter periods in adults. In immunosuppressed patients, shedding can continue for weeks.

Pathogenesis

Little is known about the histopathology of minor HRSV infection. Severe bronchiolitis or pneumonia is characterized by necrosis of the bronchiolar epithelium and a peribronchiolar infiltrate of lymphocytes and mononuclear cells. Interalveolar thickening and filling of alveolar spaces with fluid can also be found. The correlates of protective immunity to HRSV are incompletely understood. Because reinfection occurs frequently and is often associated with illness, the immunity that develops after single episodes of infection clearly is not complete or long-lasting. However, the cumulative effect of multiple reinfections is to temper subsequent disease and to provide some temporary measure of protection against infection. Studies of experimentally induced disease in healthy volunteers indicate that the presence of nasal IgA neutralizing antibody correlates more closely with protection than does the presence of serum antibody. Studies in infants, however, suggest that maternally acquired antibody provides some protection from lower respiratory tract disease, although illness can be severe even in infants who have moderate levels of maternally derived serum antibody. The relatively severe disease observed in immunosuppressed patients and experimental animal models indicates that cell-mediated immunity is an important mechanism of host defense against HRSV. Evidence suggests that major histocompatibility class I–restricted cytotoxic T cells may be particularly important in this regard.

Clinical Manifestations

HRSV infection leads to a wide spectrum of respiratory illnesses. In infants, 25–40% of infections result in lower respiratory tract involvement, including pneumonia, bronchiolitis, and tracheobronchitis. In this age group, illness begins most frequently with rhinorrhea, low-grade fever, and mild systemic symptoms, often accompanied by cough and wheezing. Most patients recover gradually over 1–2 weeks. In more severe illness, tachypnea and dyspnea develop, and eventually frank hypoxia, cyanosis, and apnea can ensue. Physical examination may reveal diffuse wheezing, rhonchi, and rales. Chest radiography shows hyperexpansion, peribronchial thickening, and variable infiltrates ranging from diffuse interstitial infiltrates to segmental or lobar consolidation. Illness may be particularly severe in children born prematurely and in those with congenital cardiac disease, bronchopulmonary dysplasia, nephrotic syndrome, or immunosuppression. One study documented a 37% mortality rate among infants with HRSV pneumonia and congenital cardiac disease.

In adults, the most common symptoms of HRSV infection are those of the common cold, with rhinorrhea, sore throat, and cough. Illness is occasionally associated with moderate systemic symptoms such as malaise, headache, and fever. HRSV has also been reported to cause lower respiratory tract disease with fever in adults, including severe pneumonia in the elderly—particularly in nursing-home residents, among whom its impact can rival that of influenza. HRSV pneumonia can be a significant cause of morbidity and death among patients undergoing stem cell and solid-organ transplantation, in whom case-fatality rates of 20–80% have been reported. Sinusitis, otitis media, and worsening of chronic obstructive and reactive airway disease have also been associated with HRSV infection.

Laboratory Findings and Diagnosis

The diagnosis of HRSV infection can be suspected on the basis of a suggestive epidemiologic setting—that is, severe illness among infants during an outbreak of HRSV in the community. Infections in older children and adults cannot be differentiated with certainty from those caused by other respiratory viruses. The specific diagnosis is established by detection of HRSV in respiratory secretions, such as sputum, throat swabs, or nasopharyngeal washes. Virus can be isolated in tissue culture, but this method has been largely supplanted by rapid viral diagnostic techniques consisting of immunofluorescence or ELISA of nasopharyngeal washes, aspirates, and (less satisfactorily) nasopharyngeal swabs. With specimens from children, these techniques have sensitivities and specificities of 80–95%; they are somewhat less sensitive with specimens from adults. RT-PCR detection techniques have shown even higher rates of sensitivity and specificity, particularly in adults. Serologic diagnosis may be made by comparison of acute- and convalescent-phase serum specimens by ELISA or by neutralization or complement-fixation tests. These tests may be useful in older children and adults but are less sensitive in children <4 months of age.

Treatment: Human Respiratory Syncytial Virus Infections

Treatment of upper respiratory tract HRSV infection is aimed primarily at the alleviation of symptoms and is similar to that for other viral infections of the upper respiratory tract. For lower respiratory tract infections, respiratory therapy, including hydration, suctioning of secretions, and administration of humidified oxygen and antibronchospastic agents, is given as needed. In severe hypoxia, intubation and ventilatory assistance may be required. Studies of infants with HRSV infection who were given aerosolized ribavirin, a nucleoside analogue active in vitro against HRSV, demonstrated a modest beneficial effect on the resolution of lower respiratory tract illness, including alleviation of blood-gas abnormalities, in some studies. The American Academy of Pediatrics recommends that treatment with aerosolized ribavirin "may be considered" for infants who are severely ill or who are at high risk for complications of HRSV infection; included are premature infants and those with bronchopulmonary dysplasia, congenital heart disease, or immunosuppression. The efficacy of ribavirin against HRSV pneumonia in older children and adults, including those with immunosuppression, has not been established. No benefit has been found in the treatment of HRSV pneumonia with standard immunoglobulin; immunoglobulin with high titers of antibody to HRSV (RSVIg), which is no longer available; or chimeric mouse-human monoclonal IgG antibody to HRSV (palivizumab). Combined therapy with aerosolized ribavirin and palivizumab is being evaluated in immunosuppressed patients with HRSV pneumonia.

Prevention

Monthly administration of RSVIg (no longer available) or palivizumab has been approved as prophylaxis against HRSV for children <2 years of age who have bronchopulmonary dysplasia or cyanotic heart disease or who were born prematurely. Considerable interest exists in the development of vaccines against HRSV. Inactivated whole-virus vaccines have been ineffective; in one study, they actually potentiated disease in infants. Other approaches include immunization with purified F and G surface glycoproteins of HRSV or generation of stable, live attenuated virus vaccines. In settings such as pediatric wards where rates of transmission are high, barrier methods for the protection of hands and conjunctivae may be useful in reducing the spread of virus.

HUMAN METAPNEUMOVIRUS INFECTIONS

Etiologic Agent

Human metapneumovirus (HMPV) is a viral respiratory pathogen that has been assigned to the Paramyxoviridae family (genus *Metapneumovirus*). Its morphology and genomic organization are similar to those of avian metapneumoviruses, which are recognized respiratory pathogens of turkeys. HMPV particles may be spherical, filamentous, or pleomorphic in shape and measure 150–600 nm in diameter. Particles contain 15-nm projections from the surface that are similar in appearance to those of other Paramyxoviridae. The single-stranded RNA genome codes for nine proteins that, except for the absence of nonstructural proteins, generally correspond to those of HRSV. HMPV is of only one antigenic type; two closely related genotypes (A and B), four subgroups, and two sublineages have been described.

Epidemiology

HMPV infections are worldwide in distribution, are most frequent during the winter, and occur early in life, so that serum antibodies to the virus are present in nearly all children by the age of 5. HMPV infections have been detected in older age groups, including elderly adults, and in both immunocompetent and immunosuppressed hosts. This virus accounts for 1–5% of childhood upper respiratory tract infections and for 10–15% of respiratory tract illnesses requiring hospitalization of children. In addition, HMPV causes 2–4% of acute respiratory illnesses in ambulatory adults and elderly patients. HMPV has been detected in a few cases of SARS, but its role (if any) in these illnesses has not been established.

Clinical Manifestations

The spectrum of clinical illnesses associated with HMPV is similar to that associated with HRSV and includes both upper and lower respiratory tract illnesses, such as bronchiolitis, croup, and pneumonia. Reinfection with HMPV is common among older children and adults and has manifestations ranging from subclinical infections to common cold syndromes and occasionally pneumonia, which is seen primarily in elderly patients and those with cardiopulmonary diseases. Serious HMPV infections occur in immunocompromised patients, including those with neoplasia and hematopoietic stem cell transplants.

Diagnosis

HMPV can be detected in nasal aspirates and respiratory secretions by immunofluorescence, by PCR, or by growth in rhesus monkey kidney (LLC-MK2) tissue cultures. A serologic diagnosis can be made by ELISA, which uses HMPV-infected tissue culture lysates as sources of antigens.

Treatment: Human Metapneumovirus Infections

Treatment for HMPV infections is primarily supportive and symptom-based. Ribavirin is active against HMPV in vitro, but its efficacy in vivo is unknown.

Prevention

Vaccines against HMPV are in the early stages of development.

PARAINFLUENZA VIRUS INFECTIONS

Etiologic Agent

Parainfluenza viruses belong to the Paramyxoviridae family (genera *Respirovirus* and *Rubulavirus*). They are 150–200 nm in diameter, are enveloped, and contain a single-stranded RNA genome. The envelope is studded with two glycoproteins: one possesses both hemagglutinin and neuraminidase activity, and the other contains fusion activity. The viral RNA genome is enclosed in a helical nucleocapsid and codes for six structural and several accessory proteins. All five serotypes of parainfluenza virus (1, 2, 3, 4A, and 4B) share certain antigens with other members of the Paramyxoviridae family, including mumps and Newcastle disease viruses.

Epidemiology

Parainfluenza viruses are distributed throughout the world; infection with serotypes 4A and 4B has been reported less widely, probably because these types are more difficult than the other three to grow in tissue culture. Infection is acquired in early childhood; by 5 years of age, most children have antibodies to serotypes 1, 2, and 3. Types 1 and 2 cause epidemics during the fall, often occurring in an alternate-year pattern. Type 3 infection has been detected during all seasons of the year, but epidemics have occurred annually in the spring.

The contribution of parainfluenza infections to respiratory disease varies with both the location and the year. In studies conducted in the United States, parainfluenza virus infections have accounted for 4.3–22% of respiratory illnesses in children. In adults, parainfluenza infections are generally mild and account for <10% of respiratory illnesses. The major importance of parainfluenza viruses is as a cause of respiratory illness in young children, in whom they rank second only to HRSV as causes of lower respiratory tract illness. Parainfluenza virus type 1 is the most frequent cause of croup (laryngotracheobronchitis) in children, while serotype 2 causes similar, although generally less severe, disease. Type 3 is an important cause of bronchiolitis and pneumonia in infants, while illnesses associated with types 4A and 4B have generally been mild. Unlike types 1 and 2, type 3 frequently causes illness during the first month of life, when passively acquired maternal antibody is still present. Parainfluenza viruses are spread through infected respiratory secretions, primarily by person-to-person contact and/or by large droplets. The incubation period has varied from 3 to 6 days in experimental infections but may be somewhat shorter for naturally occurring disease in children.

Pathogenesis

Immunity to parainfluenza viruses is incompletely understood, but evidence suggests that immunity to infections with serotypes 1 and 2 is mediated by local IgA antibodies in the respiratory tract. Passively acquired serum neutralizing antibodies also confer some protection against infection with types 1, 2, and (to a lesser degree) 3. Studies in experimental animal models and in immunosuppressed patients suggest that T cell–mediated immunity may also be important in parainfluenza virus infections.

Clinical Manifestations

Parainfluenza virus infections occur most frequently among children, in whom initial infection with serotype 1, 2, or 3 is associated with an acute febrile illness in 50–80% of cases. Children may present with coryza, sore throat, hoarseness, and cough that may or may not be croupy. In severe croup, fever persists, with worsening coryza and sore throat. A brassy or barking cough may progress to frank stridor. Most children recover over the next 1 or 2 days, although progressive airway obstruction and hypoxia ensue occasionally. If bronchiolitis or pneumonia develops, progressive cough accompanied by wheezing, tachypnea, and intercostal retractions may occur. In this setting, sputum

production increases modestly. Physical examination shows nasopharyngeal discharge and oropharyngeal injection, along with rhonchi, wheezes, or coarse breath sounds. Chest x-rays can show air trapping and occasionally interstitial infiltrates.

In older children and adults, parainfluenza infections tend to be milder, presenting most frequently as a common cold or as hoarseness, with or without cough. Lower respiratory tract involvement in older children and adults is uncommon, but tracheobronchitis in adults has been reported. Severe, prolonged, and even fatal parainfluenza infection has been reported in children and adults with severe immunosuppression, including hematopoietic stem cell and solid-organ transplant recipients.

Laboratory Findings and Diagnosis

The clinical syndromes caused by parainfluenza viruses (with the possible exception of croup in young children) are not sufficiently distinctive to be diagnosed on clinical grounds alone. A specific diagnosis is established by detection of virus in respiratory tract secretions, throat swabs, or nasopharyngeal washings. Viral growth in tissue culture is detected either by hemagglutination or by a cytopathic effect. Rapid viral diagnosis may be made by identification of parainfluenza antigens in exfoliated cells from the respiratory tract with immunofluorescence or ELISA, although these techniques appear to be less sensitive than tissue culture. Highly specific and sensitive PCR assays have also been developed. Serologic diagnosis can be established by hemagglutination inhibition, complement-fixation, or neutralization tests of acute- and convalescent-phase specimens. However, since frequent heterotypic responses occur among the parainfluenza serotypes, the serotype causing illness often cannot be identified by serologic techniques alone.

Acute epiglottitis caused by *Haemophilus influenzae* type b must be differentiated from viral croup. Influenza A virus is also a common cause of croup during epidemic periods.

Treatment: Parainfluenza Virus Infections

For upper respiratory tract illness, symptoms can be treated as discussed for other viral respiratory tract illnesses. If complications such as sinusitis, otitis, or superimposed bacterial bronchitis develop, appropriate antibacterial antibiotics should be administered. Mild cases of croup should be treated with bed rest and moist air generated by vaporizers. More severe cases require hospitalization and close observation for the development of respiratory distress. If acute respiratory distress develops, humidified oxygen and intermittent racemic epinephrine are usually administered. Aerosolized or systemically administered glucocorticoids are beneficial; the latter have a more profound effect. No specific antiviral therapy is available, although ribavirin is active against parainfluenza viruses in vitro and anecdotal reports describe its use clinically, particularly in immunosuppressed patients.

Prevention

Vaccines against parainfluenza viruses are under development.

ADENOVIRUS INFECTIONS

Etiologic Agent

Adenoviruses are complex DNA viruses that measure 70–80 nm in diameter. Human adenoviruses belong to the genus *Mastadenovirus*, which includes 51 serotypes. Adenoviruses have a characteristic morphology consisting of an icosahedral shell composed of 20 equilateral triangular faces and 12 vertices. The protein coat (capsid) consists of hexon subunits with group-specific and type-specific antigenic determinants and penton subunits at each vertex primarily containing group-specific antigens. A fiber with a knob at the end projects from each penton; this fiber contains type-specific and some group-specific antigens. Human adenoviruses have been divided into six subgroups (A through F) on the basis of the homology of DNA genomes and other properties. The adenovirus genome is a linear double-stranded DNA that codes for structural and nonstructural polypeptides. The replicative cycle of adenovirus may result either in lytic infection of cells or in the establishment of a latent infection (primarily involving lymphoid cells). Some adenovirus types can induce oncogenic transformation, and tumor formation has been observed in rodents; however, despite intensive investigation, adenoviruses have not been associated with tumors in humans.

Epidemiology

Adenovirus infections most frequently affect infants and children. Infections occur throughout the year but are most common from fall to spring. Adenoviruses account for ~10% of acute respiratory infections in children but for <2% of respiratory illnesses in civilian adults. Nearly 100% of adults have serum antibody to multiple serotypes—a finding indicating that infection is common in childhood. Types 1, 2, 3, and 5 are the most common isolates from children. Certain adenovirus serotypes—particularly 4 and 7 but also 3, 14, and 21—are associated with outbreaks of acute respiratory disease in military recruits in winter and spring. Adenovirus infection can be transmitted by inhalation of aerosolized virus, by inoculation of virus into conjunctival sacs, and probably by the fecal-oral route as well. Type-specific antibody generally develops after infection and is associated with protection, albeit incomplete, against infection with the same serotype.

Clinical Manifestations

In children, adenoviruses cause a variety of clinical syndromes. The most common is an acute upper respiratory tract infection, with prominent rhinitis. On occasion, lower respiratory tract disease, including bronchiolitis and pneumonia, also develops. Adenoviruses, particularly types 3 and 7, cause pharyngoconjunctival fever, a characteristic acute febrile illness of children that occurs in outbreaks, most often in summer camps. The syndrome is marked by bilateral conjunctivitis in which the bulbar and palpebral conjunctivae have a granular appearance. Low-grade fever is frequently present for the first 3–5 days, and rhinitis, sore throat, and cervical adenopathy develop. The illness generally lasts for 1–2 weeks and resolves spontaneously. Febrile pharyngitis without conjunctivitis has also been associated with adenovirus infection. Adenoviruses have been isolated from cases of whooping cough with or without *Bordetella pertussis*; the significance of adenovirus in that disease is unknown.

In adults, the most frequently reported illness has been acute respiratory disease caused by adenovirus types 4 and 7 in military recruits. This illness is marked by a prominent sore throat and the gradual onset of fever, which often reaches 39°C (102.2°F) on the second or third day of illness. Cough is almost always present, and coryza and regional lymphadenopathy are frequently seen. Physical examination may show pharyngeal edema, injection, and tonsillar enlargement with little or no exudate. If pneumonia has developed, auscultation and x-ray of the chest may indicate areas of patchy infiltration.

Adenoviruses have been associated with a number of non–respiratory tract diseases, including acute diarrheal illness caused by types 40 and 41 in young children and hemorrhagic cystitis caused by types 11 and 21. Epidemic keratoconjunctivitis, caused most frequently by types 8, 19, and 37, has been associated with contaminated common sources such as ophthalmic solutions and roller towels. Adenoviruses have also been implicated in disseminated disease and pneumonia in immunosuppressed patients, including recipients of solid-organ or hematopoietic stem cell transplants. In hematopoietic stem cell transplant recipients, adenovirus infections have been manifested as pneumonia, hepatitis, nephritis, colitis, encephalitis, and hemorrhagic cystitis. In solid-organ transplant recipients, adenovirus infection may involve the organ transplanted (e.g., hepatitis in liver transplants, nephritis in renal transplants) but can disseminate to other organs as well. In patients with AIDS, high-numbered and intermediate adenovirus serotypes have been isolated, usually in the setting of low CD4+ T cell counts, but their isolation often has not been clearly linked to disease manifestations. Adenovirus nucleic acids have been detected in myocardial cells from patients with "idiopathic" myocardiopathies, and adenoviruses have been suggested as causative agents in some cases.

Laboratory Findings and Diagnosis

Adenovirus infection should be suspected in the epidemiologic setting of acute respiratory disease in military recruits and in certain of the clinical syndromes (such as pharyngoconjunctival fever or epidemic keratoconjunctivitis) in which outbreaks of characteristic illnesses occur. In most cases, however, illnesses caused by adenovirus infection cannot be differentiated from those caused by a number of other viral respiratory agents and *Mycoplasma pneumoniae*. A definitive diagnosis of adenovirus infection is established by detection of the virus in tissue culture (as evidenced by cytopathic changes) and by specific identification with immunofluorescence or other immunologic techniques. Rapid viral diagnosis can be established by immunofluorescence or ELISA of nasopharyngeal aspirates, conjunctival or respiratory secretions, urine, or stool. Highly sensitive and specific PCR assays and nucleic acid hybridization are also available. Adenovirus types 40 and 41, which have been associated with diarrheal disease in children, require special tissue-culture cells for isolation, and these serotypes are most commonly detected by direct ELISA of stool. Serum antibody rises can be demonstrated by complement-fixation or neutralization tests, ELISA, radioimmunoassay, or (for those adenoviruses that hemagglutinate red cells) hemagglutination inhibition tests.

Treatment: Adenovirus Infections

Only symptom-based treatment and supportive therapy are available for adenovirus infections, and clinically useful antiviral therapy has not been established. Ribavirin and cidofovir are active in vitro against certain adenoviruses. Retrospective studies and anecdotes describe the use of these agents in disseminated adenovirus infections, but definitive efficacy data from controlled studies are not available.

Prevention

Live vaccines have been developed against adenovirus types 4 and 7 and have been used to control illness among military recruits. These vaccines consist of live, unattenuated virus administered in enteric-coated capsules. Infection of the gastrointestinal tract with types 4 and 7 does not cause disease but stimulates local and systemic antibodies that are protective against subsequent acute respiratory disease due to those serotypes. This vaccine has not been produced since 1999, and outbreaks of acute respiratory illness caused by adenovirus types 4 and 7 have again emerged among military recruits. Therefore, a program to redevelop type 4 and 7 vaccines is under way. Adenoviruses are also being studied as live-virus vectors for the delivery of vaccine antigens and for gene therapy.

FURTHER READINGS

American Academy of Pediatrics : Diagnosis and management of bronchiolitis. Pediatrics 118:1774, 2006

Echavarría M: Adenoviruses in immunocompromised hosts. Clin Microbiol Rev 21:704, 2008

Graham BS et al: Respiratory syncytial virus immunobiology and pathogenesis. Virology 297:1, 2002

Nair H et al: Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: A systematic review and meta-analysis. Lancet 375:545, 2010

Peiris JS et al: Severe acute respiratory syndrome. Nat Med 10:S88, 2004

Peret T et al: Characterization of human metapneumoviruses isolated from patients in North America. J Infect Dis 185:1660, 2002

Pyrc K et al: Identification of new human coronaviruses. Expert Rev Anti Infect Ther 2:245, 2007

Turner RB: Rhinoviruses, in *Principles and Practice of Infectious Diseases* , 7th ed, GF Mandell et al (eds). Philadelphia, Elsevier, 2010, pp 2389–2398

Walsh EE et al*:* Human metapneumovirus infections in adults: Another piece of the puzzle. Arch Intern Med 168:2489, 2008

Wright PF: Parainfluenza viruses, in *Viral Infections of the Respiratory Tract* , R Dolin, PF Wright (eds). New York, Marcel Dekker, 1999

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFLUENZA: INTRODUCTION

Definition

Influenza is an acute respiratory illness caused by infection with influenza viruses. The illness affects the upper and/or lower respiratory tract and is often accompanied by systemic signs and symptoms such as fever, headache, myalgia, and weakness. Outbreaks of illness of variable extent and severity occur nearly every year. Such outbreaks result in significant morbidity rates in the general population and in increased mortality rates among certain high-risk patients, mainly as a result of pulmonary complications.

Etiologic Agent

Influenza viruses are members of the Orthomyxoviridae family, of which influenza A, B, and C viruses constitute three separate genera. The designation of influenza viruses as type A, B, or C is based on antigenic characteristics of the nucleoprotein (NP) and matrix (M) protein antigens. Influenza A viruses are further subdivided (subtyped) on the basis of the surface hemagglutinin (H) and neuraminidase (N) antigens (see below); individual strains are designated according to the site of origin, isolate number, year of isolation, and subtype—for example, influenza A/California/07/2009 (H1N1). Influenza A has 16 distinct H subtypes and 9 distinct N subtypes, of which only H1, H2, H3, N1, and N2 have been associated with epidemics of disease in humans. Influenza B and C viruses are similarly designated, but H and N antigens from these viruses do not receive subtype designations, since intratypic variations in influenza B antigens are less extensive than those in influenza A viruses and may not occur with influenza C virus.

Influenza A and B viruses are major human pathogens and the most extensively studied of the Orthomyxoviridae. Type A and type B viruses are morphologically similar. The virions are irregularly shaped spherical particles, measure 80–120 nm in diameter, and have a lipid envelope from the surface of which the H and N glycoproteins project (**Fig. 187-1**). The hemagglutinin is the site by which the virus binds to sialic acid cell receptors, whereas the neuraminidase degrades the receptor and plays a role in the release of the virus from infected cells after replication has taken place. Influenza viruses enter cells by receptor-mediated endocytosis, forming a virus-containing endosome. The viral hemagglutinin mediates fusion of the endosomal membrane with the virus envelope, and viral nucleocapsids are subsequently released into the cytoplasm. Immune responses to the H antigen are the major determinants of protection against infection with influenza virus, while those to the N antigen limit viral spread and contribute to reduction of the infection. The lipid envelope of influenza A virus also contains the M proteins M1 and M2, which are involved in stabilization of the lipid envelope and in virus assembly. The virion also contains the NP antigen, which is associated with the viral genome, as well as three polymerase (P) proteins that are essential for transcription and synthesis of viral RNA. Two nonstructural proteins function as an interferon antagonist and posttranscriptional regulator (NS1) and a nuclear export factor (NS2 or NEP).

Figure 187-1

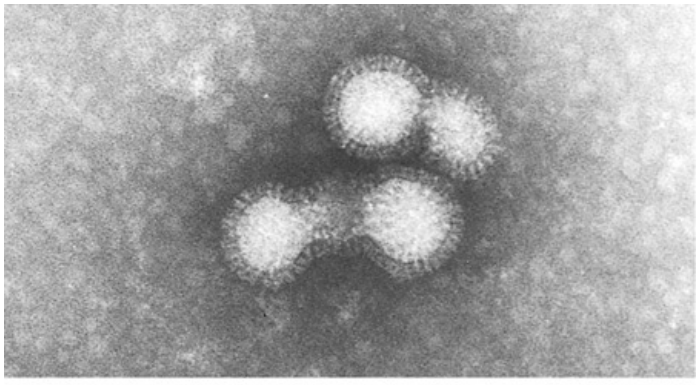

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An electron micrograph of influenza A virus (x40,000).

The genomes of influenza A and B viruses consist of eight single-strand RNA segments, which code for the structural and nonstructural proteins. Because the genome is segmented, the opportunity for gene reassortment during infection is high; reassortment often occurs during infection of cells with more than one influenza A virus.

Epidemiology

Influenza outbreaks are recorded virtually every year, although their extent and severity vary widely. Localized outbreaks take place at variable intervals, usually every 1–3 years. Global pandemics have occurred at variable intervals, but much less frequently than interpandemic outbreaks (**Table 187-1**). The most recent pandemic emerged in March of 2009 and was caused by an influenza A/H1N1 virus that rapidly spread worldwide over the next several months.

Table 187-1 Emergence of Antigenic Subtypes of Influenza A Virus Associated with Pandemic or Epidemic Disease

Years	Subtype	Extent of Outbreak
1889–1890	H2N8[a]	Severe pandemic
1900–1903	H3N8[a]	?Moderate epidemic
1918–1919	H1N1[b] (formerly HswN1)	Severe pandemic
1933–1935	H1N1[b] (formerly H0N1)	Mild epidemic
1946–1947	H1N1	Mild epidemic
1957–1958	H2N2	Severe pandemic
1968–1969	H3N2	Moderate pandemic
1977–1978[c]	H1N1	Mild pandemic
2009–2010[d]	H1N1	Pandemic

[a] As determined by retrospective serologic survey of individuals alive during those years ("seroarchaeology").

[b] Hemagglutinins formerly designated as Hsw and H0 are now classified as variants of H1.

[c] From this time until 2008–2009, viruses of the H1N1 and H3N2 subtypes circulated either in alternating years or concurrently.

[d] Novel influenza A/H1N1 emerged to cause this pandemic.

INFLUENZA A VIRUS

Antigenic Variation and Influenza Outbreaks and Pandemics

The most extensive and severe outbreaks of influenza are caused by influenza A viruses, in part because of the remarkable propensity of the H and N antigens of these viruses to undergo periodic antigenic variation. Major antigenic variations, called *antigenic shifts*, are seen only with influenza A viruses and may be associated with pandemics. Minor variations are called *antigenic drifts*. Antigenic variation may involve the hemagglutinin alone or both the hemagglutinin and the neuraminidase. An example of an antigenic shift involving both the hemagglutinin and the neuraminidase is that of 1957, when the predominant influenza A virus subtype shifted from H1N1 to H2N2; this shift resulted in a severe pandemic, with an estimated 70,000 excess deaths (i.e., deaths in excess of the number expected without an influenza epidemic) in the United States alone. In 1968, an antigenic shift involving only the hemagglutinin occurred (H2N2 to H3N2); the subsequent pandemic was less severe than that of 1957. In 1977, an H1N1 virus emerged and caused a pandemic that primarily affected younger individuals (i.e., those born after 1957). As can be seen in Table 187-1, H1N1 viruses circulated from 1918 to 1956; thus, individuals born prior to 1957 would be expected to have some degree of immunity to H1N1 viruses. The pandemic of 2009–2010 was caused by an A/H1N1 virus against which little immunity was present in the general population, although approximately one-third of individuals born before 1950 had some apparent immunity to related H1N1 strains.

During most outbreaks of influenza A, a single subtype has circulated at a time. However, since 1977, H1N1 and H3N2 viruses have circulated simultaneously, resulting in outbreaks of varying severity. In some outbreaks, influenza B viruses have also circulated simultaneously with influenza A viruses. In 2009–2010, the pandemic A/H1N1 virus appeared to circulate nearly exclusively.

Avian Influenza A Viruses

In 1997, human cases of influenza caused by avian influenza viruses (A/H5N1) were detected in Hong Kong during an extensive outbreak of influenza in poultry. Between that time and February 2010, 478 cases of avian influenza in humans were reported in Asia and the Middle East. Nearly all of these cases were associated with contact with infected poultry. Efficient person-to-person transmission has not been observed to date. Mortality rates have been high (60%), and clinical manifestations have differed somewhat from those associated with "typical" outbreaks of influenza (see below). Transmission of avian influenza A/H7N7 viruses from infected poultry to humans has been observed, including outbreaks in the Netherlands, which resulted predominantly in cases of conjunctivitis and some respiratory illnesses. Infection with avian A/H9N2 viruses along with mild respiratory illness has been reported in children in Hong Kong. Because of the absence of widespread immunity to the H5, H7, and H9 viruses, concern persists that avian-to-human transmission might also contribute to the emergence of pandemic strains.

The origin of actual pandemic influenza A virus strains has been partially elucidated with molecular virologic techniques. It appears that the pandemic strains of 1957 and 1968 resulted from a genetic reassortment between human viruses and avian viruses with novel surface glycoproteins (H2N2, H3). The pandemic A/H1N1 virus of 2009–2010 was a quadruple reassortant among swine influenza viruses that circulated in North America and Eurasia, an avian virus, and a human influenza virus. The influenza A/H1N1 virus responsible for the most severe pandemic of modern times (1918–1919) appears to have represented an adaptation of an avian virus to efficient infection of humans.

Features of Pandemic and Interpandemic Influenza A

Pandemics provide the most dramatic evidence of the impact of influenza A. However, illnesses occurring between pandemics (interpandemic disease) also account for extensive mortality and morbidity rates, albeit over a longer period. In the United States, influenza was associated with at least 19,000 excess deaths per season in 1976–1990 and with 36,000 excess deaths per season in 1990–1999. On average, there were 226,000 influenza-associated hospitalizations per year in this country in 1979–2001.

Influenza A viruses that circulate between pandemics demonstrate antigenic drifts in the H antigen. These antigenic drifts result from point mutations involving the RNA segment that codes for the hemagglutinin, which occur most frequently in five hypervariable regions. Epidemiologically significant strains—that is, those with the potential to cause widespread outbreaks—exhibit changes in amino acids in at least two of the major antigenic sites in the hemagglutinin molecule. Since two point mutations are unlikely to occur simultaneously, it is believed that antigenic drifts result from point mutations occurring sequentially during the spread of virus from person to person. Antigenic drifts have been reported nearly annually since 1977 for H1N1 viruses and since 1968 for H3N2 viruses.

Interpandemic influenza A outbreaks usually begin abruptly, peak over a 2- to 3-week period, generally last for 2–3 months, and often subside almost as rapidly as they began. In contrast, pandemic influenza may begin with rapid transmission at multiple locations, have high attack rates, and extend beyond the usual seasonality, with multiple waves of attack before or after the main outbreak. In interpandemic outbreaks, the first indication of influenza activity is an increase in the number of children with febrile respiratory illnesses who present for medical attention. This increase is followed by increases in rates of influenza-like illnesses among adults and eventually by an increase in hospital admissions for patients with pneumonia, worsening of congestive heart failure, and exacerbations of chronic pulmonary disease. Rates of absence from work and school also rise at this time. An increase in the number of deaths caused by pneumonia and influenza is generally a late observation in an outbreak. Attack rates have been highly variable from outbreak to outbreak in interpandemic influenza but most commonly are in the range of 10–20% of the general population.

While pandemic influenza may occur throughout the year, interpandemic influenza occurs almost exclusively during the winter months in the temperate zones of the Northern and Southern hemispheres. In those locations, it is highly unusual to detect influenza A virus at other times, although rises in serum antibody titer or even outbreaks have been noted rarely during warm-weather months. In contrast, influenza virus infections occur throughout the year in the tropics. Where or how influenza A viruses persist between outbreaks in temperate zones is unknown. It is possible that the viruses are maintained in the human population on a worldwide basis by person-to-person transmission and that large population clusters support a low level of interepidemic transmission. Alternatively, human strains may persist in animal reservoirs. Convincing evidence to support either explanation is not available. In the modern era, rapid transportation may contribute to the transmission of viruses among widespread geographic locales.

The factors that result in the inception and termination of outbreaks of influenza A are incompletely understood. A major determinant of the extent and severity of an outbreak is the level of immunity in the population at risk. With the emergence of an antigenically novel influenza virus to which little or no immunity is present in a community, extensive outbreaks may occur. When the absence of immunity is worldwide, epidemic disease may spread around the globe, resulting in a pandemic. Such pandemic waves can continue for several years, until immunity in the population reaches a high level. In the years following pandemic influenza, antigenic drifts among influenza viruses result in outbreaks of

variable severity in populations with high levels of immunity to the pandemic strain that circulated earlier. This situation persists until another antigenically novel pandemic strain emerges. On the other hand, outbreaks sometimes end despite the persistence of a large pool of susceptible individuals in the population. It has been suggested that certain influenza A viruses may be intrinsically less virulent and cause less severe disease than other variants, even in immunologically virgin subjects. If so, then other (undefined) factors besides the level of preexisting immunity must play a role in the epidemiology of influenza.

INFLUENZA B AND C VIRUSES

Influenza B virus causes outbreaks that are generally less extensive and are associated with less severe disease than those caused by influenza A virus. The hemagglutinin and neuraminidase of influenza B virus undergo less frequent and less extensive variation than those of influenza A viruses; this characteristic may account, in part, for the lesser extent of disease. Influenza B outbreaks are seen most frequently in schools and military camps, although outbreaks in institutions in which elderly individuals reside have also been noted on occasion. The most serious complication of influenza B virus infection is Reye's syndrome.

In contrast to influenza A and B viruses, influenza C virus appears to be a relatively minor cause of disease in humans. It has been associated with common cold–like symptoms and occasionally with lower respiratory tract illness. The widespread prevalence of serum antibody to this virus indicates that asymptomatic infection may be common.

INFLUENZA-ASSOCIATED MORBIDITY AND MORTALITY RATES

The morbidity and mortality rates caused by influenza outbreaks continue to be substantial. Most individuals who die in this setting have underlying diseases that place them at high risk for complications of influenza (**Table 187-2**). Excess annual hospitalizations for groups of adults and children with high-risk medical conditions ranged from 40 to 1900 per 100,000 during outbreaks of influenza in 1973–2004. The most prominent high-risk conditions are chronic cardiac and pulmonary diseases and old age. Mortality rates among individuals with chronic metabolic or renal diseases or certain immunosuppressive diseases have also been elevated, albeit lower than those among patients with chronic cardiopulmonary diseases. In the pandemic of 2009–2010, increased risk for severe disease was noted in children from birth to 4 years of age and in pregnant women. The morbidity rate attributable to influenza in the general population is considerable. It is estimated that interpandemic outbreaks of influenza currently incur annual economic costs of more than $87 billion in the United States. For pandemics, it is estimated that annual economic costs would range from $89.7 to $209.4 billion for attack rates of 15–35%.

Table 187-2 Persons at Higher Risk for Complications of Influenza

Children from birth to 4 years old
Pregnant women
Persons ≥65 years old
Children and adolescents (6 months to 18 years old) who are receiving long-term aspirin therapy and therefore may be at risk for developing Reye's syndrome after influenza
Adults and children who have chronic disorders of the pulmonary or cardiovascular system, including asthma
Adults and children who have chronic metabolic diseases (including diabetes mellitus), renal dysfunction, hemoglobinopathies, or immunodeficiency (including immunodeficiency caused by medications or by HIV)
Adults and children who have any condition that can compromise respiratory function or compromise the handling of respiratory secretions or can increase the risk of aspiration
Residents of nursing homes and other chronic-care facilities that house persons of any age who have chronic medical conditions

Pathogenesis and Immunity

The initial event in influenza is infection of the respiratory epithelium with influenza virus acquired from respiratory secretions of acutely infected individuals. In all likelihood, the virus is transmitted via aerosols generated by coughs and sneezes, although hand-to-hand contact, other personal contact, and even fomite transmission may take place. Experimental evidence suggests that infection by a small-particle aerosol (particle diameter, <10 μm) is more efficient than that by larger droplets. Initially, viral infection involves the ciliated columnar epithelial cells, but it may also involve other respiratory tract cells, including alveolar cells, mucous gland cells, and macrophages. In infected cells, virus replicates within 4–6 h, after which infectious virus is released to infect adjacent or nearby cells. In this way, infection spreads from a few foci to a large number of respiratory cells over several hours. In experimentally induced infection, the incubation period of illness has ranged from 18 to 72 h, depending on the size of the viral inoculum. Histopathologic study reveals degenerative changes,

including granulation, vacuolization, swelling, and pyknotic nuclei, in infected ciliated cells. The cells eventually become necrotic and desquamate; in some areas, previously columnar epithelium is replaced by flattened and metaplastic epithelial cells. The severity of illness is correlated with the quantity of virus shed in secretions; thus, the degree of viral replication itself may be an important factor in pathogenesis. Despite the frequent development of systemic signs and symptoms such as fever, headache, and myalgias, influenza virus has only rarely been detected in extrapulmonary sites (including the bloodstream). Evidence suggests that the pathogenesis of systemic symptoms in influenza may be related to the induction of certain cytokines, particularly tumor necrosis factor α, interferon α, interleukin 6, and interleukin 8, in respiratory secretions and in the bloodstream.

The host response to influenza infections involves a complex interplay of humoral antibody, local antibody, cell-mediated immunity, interferon, and other host defenses. Serum antibody responses, which can be detected by the second week after primary infection, are measured by a variety of techniques: hemagglutination inhibition (HI), complement fixation (CF), neutralization, enzyme-linked immunosorbent assay (ELISA), and antineuraminidase antibody assay. Antibodies to the hemagglutinin appear to be the most important mediators of immunity; in several studies, HI titers of ≥40 have been associated with protection from infection. Secretory antibodies produced in the respiratory tract are predominantly of the IgA class and also play a major role in protection against infection. Secretory antibody neutralization titers of ≥4 have also been associated with protection. A variety of cell-mediated immune responses, both antigen-specific and antigen-nonspecific, can be detected early after infection and depend on the prior immune status of the host. These responses include T cell proliferative, T cell cytotoxic, and natural killer cell activity. In humans, CD8+ human leukocyte antigen class I–restricted cytotoxic T lymphocytes (CTLs) are directed at conserved regions of internal proteins (NP, M, and P) as well as at the surface proteins H and N. Interferons can be detected in respiratory secretions shortly after the shedding of virus has begun, and rises in interferon titers coincide with decreases in virus shedding.

The host defense factors responsible for cessation of virus shedding and resolution of illness have not been defined specifically. Virus shedding generally stops within 2–5 days after symptoms first appear, at a time when serum and local antibody responses often are not detectable by conventional techniques (although antibody rises may be detected earlier by use of highly sensitive techniques, particularly in individuals with previous immunity to the virus). It has been suggested that interferon, cell-mediated immune responses, and/or nonspecific inflammatory responses all contribute to the resolution of illness. CTL responses may be particularly important in this regard.

Clinical Manifestations

Influenza has most frequently been described as an illness characterized by the abrupt onset of systemic symptoms, such as headache, feverishness, chills, myalgia, and malaise, as well as accompanying respiratory tract signs, particularly cough and sore throat. In many cases, the onset is so abrupt that patients can recall the precise time they became ill. However, the spectrum of clinical presentations is wide, ranging from a mild, afebrile respiratory illness similar to the common cold (with either a gradual or an abrupt onset) to severe prostration with relatively few respiratory signs and symptoms. In most of the cases that come to a physician's attention, the patient has a fever, with temperatures of 38°–41°C (100.4°–105.8°F). A rapid temperature rise within the first 24 h of illness is generally followed by gradual defervescence over 2–3 days, although, on occasion, fever may last as long as 1 week. Patients report a feverish feeling and chilliness, but true rigors are rare. Headache, either generalized or frontal, is often particularly troublesome. Myalgias may involve any part of the body but are most common in the legs and lumbosacral area. Arthralgias may also develop.

Respiratory symptoms often become more prominent as systemic symptoms subside. Many patients have a sore throat or persistent cough, which may last for ≥1 week and which is often accompanied by substernal discomfort. Ocular signs and symptoms include pain on motion of the eyes, photophobia, and burning of the eyes.

Physical findings are usually minimal in uncomplicated influenza. Early in the illness, the patient appears flushed and the skin is hot and dry, although diaphoresis and mottled extremities are sometimes evident, particularly in older patients. Examination of the pharynx may yield surprisingly unremarkable results despite a severe sore throat, but injection of the mucous membranes and postnasal discharge are apparent in some cases. Mild cervical lymphadenopathy may be noted, especially in younger individuals. The results of chest examination are largely negative in uncomplicated influenza, although rhonchi, wheezes, and scattered rales have been reported with variable frequency in different outbreaks. Frank dyspnea, hyperpnea, cyanosis, diffuse rales, and signs of consolidation are indicative of pulmonary complications. Patients with apparently uncomplicated influenza have been reported to have a variety of mild ventilatory defects and increased alveolar-capillary diffusion gradients; thus, subclinical pulmonary involvement may be more common than is appreciated.

In uncomplicated influenza, the acute illness generally resolves over 2–5 days, and most patients have largely recovered in 1 week, although cough may persist 1–2 weeks longer. In a significant minority (particularly the elderly), however, symptoms of weakness or lassitude (postinfluenza asthenia) may

persist for several weeks and may prove troublesome for persons who wish to resume their full level of activity promptly. The pathogenetic basis for this asthenia is unknown, although pulmonary function abnormalities may persist for several weeks after uncomplicated influenza.

Complications

Complications of influenza (Table 187-2) occur most frequently in patients >65 years old and in those with certain chronic disorders, including cardiac or pulmonary diseases, diabetes mellitus, hemoglobinopathies, renal dysfunction, and immunosuppression. Pregnancy in the second or third trimester predisposes to complications with influenza. Children <5 years old (especially infants) are also at high risk for complications.

PULMONARY COMPLICATIONS

Pneumonia

The most significant complication of influenza is pneumonia: "primary" influenza viral pneumonia, secondary bacterial pneumonia, or mixed viral and bacterial pneumonia.

Primary Influenza Viral Pneumonia

Primary influenza viral pneumonia is the least common but most severe of the pneumonic complications. It presents as acute influenza that does not resolve but instead progresses relentlessly, with persistent fever, dyspnea, and eventual cyanosis. Sputum production is generally scanty, but the sputum can contain blood. Few physical signs may be evident early in the illness. In more advanced cases, diffuse rales may be noted, and chest x-ray findings consistent with diffuse interstitial infiltrates and/or acute respiratory distress syndrome may be present. In such cases, arterial blood-gas determinations show marked hypoxia. Viral cultures of respiratory secretions and lung parenchyma, especially if samples are taken early in illness, yield high titers of virus. In fatal cases of primary viral pneumonia, histopathologic examination reveals a marked inflammatory reaction in the alveolar septa, with edema and infiltration by lymphocytes, macrophages, occasional plasma cells, and variable numbers of neutrophils. Fibrinthrombi in alveolar capillaries, along with necrosis and hemorrhage, have also been noted. Eosinophilic hyaline membranes can be found lining alveoli and alveolar ducts.

Primary influenza viral pneumonia has a predilection for individuals with cardiac disease, particularly those with mitral stenosis, but has also been reported in otherwise-healthy young adults as well as in older individuals with chronic pulmonary disorders. In some pandemics of influenza (notably those of 1918 and 1957), pregnancy increased the risk of primary influenza pneumonia. Subsequent epidemics of influenza have been associated with increased rates of hospitalization among pregnant women, which were also noted in the pandemic of 2009–2010.

Secondary Bacterial Pneumonia

Secondary bacterial pneumonia follows acute influenza. Improvement of the patient's condition over 2–3 days is followed by a reappearance of fever along with clinical signs and symptoms of bacterial pneumonia, including cough, production of purulent sputum, and physical and x-ray signs of consolidation. The most common bacterial pathogens in this setting are *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Haemophilus influenzae*—organisms that can colonize the nasopharynx and that cause infection in the wake of changes in bronchopulmonary defenses. The etiology can often be determined by Gram's staining and culture of an appropriately obtained sputum specimen. Secondary bacterial pneumonia occurs most frequently in high-risk individuals with chronic pulmonary and cardiac disease and in elderly individuals. Patients with secondary bacterial pneumonia often respond to antibiotic therapy when it is instituted promptly.

Mixed Viral and Bacterial Pneumonia

Perhaps the most common pneumonic complications during outbreaks of influenza have mixed features of viral and bacterial pneumonia. Patients may experience a gradual progression of their acute illness or may show transient improvement followed by clinical exacerbation, with eventual manifestation of the clinical features of bacterial pneumonia. Sputum cultures may contain both influenza A virus and one of the bacterial pathogens described above. Patchy infiltrates or areas of consolidation may be detected by physical examination and chest x-ray. Patients with mixed viral and bacterial pneumonia generally have less widespread involvement of the lung than those with primary viral pneumonia, and their bacterial infections may respond to appropriate antibacterial drugs. Mixed viral and bacterial pneumonia occurs primarily in patients with chronic cardiovascular and pulmonary diseases.

Other Pulmonary Complications

Other pulmonary complications associated with influenza include worsening of chronic obstructive pulmonary disease and exacerbation of chronic bronchitis and asthma. In children, influenza infection may present as croup. Sinusitis as well as otitis media (the latter occurring particularly often in children) may also be associated with influenza.

EXTRAPULMONARY COMPLICATIONS

In addition to the pulmonary complications of influenza, a number of extrapulmonary complications may

occur. These include *Reye's syndrome*, a serious complication in children that is associated with influenza B and to a lesser extent with influenza A virus infection as well as with varicella-zoster virus infection. An epidemiologic association between Reye's syndrome and aspirin therapy for the antecedent viral infection has been noted, and the syndrome's incidence has decreased markedly with widespread warnings regarding aspirin use by children with acute viral respiratory infections.

Myositis, rhabdomyolysis, and myoglobinuria are occasional complications of influenza infection. Although myalgias are exceedingly common in influenza, true myositis is rare. Patients with acute myositis have exquisite tenderness of the affected muscles, most commonly in the legs, and may not be able to tolerate even the slightest pressure, such as the touch of bedsheets. In the most severe cases, there is frank swelling and bogginess of muscles. Serum levels of creatine phosphokinase and aldolase are markedly elevated, and an occasional patient develops renal failure from myoglobinuria. The pathogenesis of influenza-associated myositis is also unclear, although the presence of influenza virus in affected muscles has been reported.

Myocarditis and pericarditis were reported in association with influenza virus infection during the 1918–1919 pandemic; these reports were based largely on histopathologic findings, and these complications have been reported only infrequently since that time. Electrocardiographic changes during acute influenza are common among patients who have cardiac disease but have been ascribed most often to exacerbations of the underlying cardiac disease rather than to direct involvement of the myocardium with influenza virus.

Central nervous system (CNS) diseases, including encephalitis, transverse myelitis, and Guillain-Barré syndrome, have been reported during influenza. The etiologic relationship of influenza virus to such CNS illnesses remains uncertain. Toxic shock syndrome associated with *S. aureus* or group A streptococcal infection following acute influenza infection has also been reported (Chaps. 135 and 136).

In addition to complications involving the specific organ systems described above, influenza outbreaks include a number of cases in which elderly and other high-risk individuals develop influenza and subsequently experience a gradual deterioration of underlying cardiovascular, pulmonary, or renal function—changes that occasionally are irreversible and lead to death. These deaths contribute to the overall excess mortality rate associated with influenza A outbreaks.

COMPLICATIONS OF AVIAN INFLUENZA

Cases of influenza caused by avian A/H5N1 virus are reportedly associated with high rates of pneumonia (>50%) and extrapulmonary manifestations such as diarrhea and CNS involvement. Deaths have been associated with multisystem dysfunction, including cardiac and renal failure.

Laboratory Findings and Diagnosis

During acute influenza, virus may be detected in throat swabs, nasopharyngeal swabs or washes, or sputum. The virus can be isolated by use of tissue culture—or, less commonly, chick embryos—within 48–72 h after inoculation. Most commonly, the laboratory diagnosis is established with rapid tests that detect viral antigens by means of immunologic or enzymatic techniques. The tests are relatively specific but are of variable sensitivity depending on the technique and the virus to be detected. Some rapid tests can distinguish between influenza A and B viruses, but detection of differences in hemagglutinin subtypes requires additional subtype-specific immunologic techniques. The most sensitive and specific in vitro test for influenza virus is reverse-transcriptase polymerase chain reaction; this test proved particularly important in detecting the 2009–2010 pandemic A/H1N1 viruses, for which some rapid antigen detection tests were poorly sensitive. Serologic methods for diagnosis require comparison of antibody titers in sera obtained during the acute illness with those in sera obtained 10–14 days after the onset of illness and are useful primarily in retrospect. Fourfold or greater titer rises as detected by HI or CF or significant rises as measured by ELISA are diagnostic of acute infection. Other laboratory tests generally are not helpful in the specific diagnosis of influenza virus infection. Leukocyte counts are variable, frequently being low early in illness and normal or slightly elevated later. Severe leukopenia has been described in overwhelming viral or bacterial infection, while leukocytosis with >15,000 cells/μL raises the suspicion of secondary bacterial infection.

Differential Diagnosis

During a community-wide outbreak, a clinical diagnosis of influenza can be made with a high degree of certainty in patients who present to a physician's office with the typical febrile respiratory illness described above. In the absence of an outbreak (i.e., in sporadic or isolated cases), influenza may be difficult to differentiate on clinical grounds alone from an acute respiratory illness caused by any of a variety of respiratory viruses or by *Mycoplasma pneumoniae*. Severe streptococcal pharyngitis or early bacterial pneumonia may mimic acute influenza, although bacterial pneumonias generally do not run a self-limited course. Purulent sputum in which a bacterial pathogen can be detected by Gram's staining is an important diagnostic feature in bacterial pneumonia.

Treatment: Influenza

Specific antiviral therapy is available for influenza (**Table 187-3**): the neuraminidase inhibitors

zanamivir, oseltamivir, and peramivir for both influenza A and influenza B and the adamantane agents amantadine and rimantadine for influenza A (Chap. 178). A 5-day course of oseltamivir or zanamivir reduces the duration of signs and symptoms of uncomplicated influenza by 1–1.5 days if treatment is started within 2 days of the onset of illness. Zanamivir may exacerbate bronchospasm in asthmatic patients, and oseltamivir has been associated with nausea and vomiting, whose frequency can be reduced by administration of the drug with food. Oseltamivir has also been associated with neuropsychiatric side effects in children. Peramivir, an investigational neuraminidase inhibitor that can be administered intravenously, is being evaluated in clinical trials, as is an intravenous form of zanamivir. Access to these medications can be sought through the FDA's Emergency Investigational New Drug (E-IND) application procedures.

Table 187-3 Antiviral Medications for Treatment and Prophylaxis of Influenza

	Age Group (Years)		
Antiviral Drug	**Children (≤12)**	**13–64**	**≥65**
Oseltamivir			
Treatment, influenza A and B	Age 1–12, dose varies by weight[a]	75 mg PO bid	75 mg PO bid
Prophylaxis, influenza A and B	Age 1–12, dose varies by weight[b]	75 PO qd	75 mg PO qd
Zanamivir			
Treatment, influenza A and B	Age 7–12, 10 mg bid by inhalation	10 mg bid by inhalation	10 mg bid by inhalation
Prophylaxis, influenza A and B	Age 5–12, 10 mg qd by inhalation	10 mg qd by inhalation	10 mg qd by inhalation
Amantadine[c]			
Treatment, influenza A	Age 1–9, 5 mg/kg in 2 divided doses, up to 150 mg/d	Age ≥10, 100 mg PO bid	≤100 mg/d
Prophylaxis, influenza A	Age 1–9, 5 mg/kg in 2 divided doses, up to 150 mg/d	Age ≥10, 100 mg PO bid	≤100 mg/d
Rimantadine[c]			
Treatment, influenza A	Not approved	100 mg PO bid	100–200 mg/d
Prophylaxis, influenza A	Age 1–9, 5 mg/kg in 2 divided doses, up to 150 mg/d	Age ≥10, 100 mg PO bid	100–200 mg/d

[a] <15 kg: 30 mg bid; >15–23 kg: 45 mg bid; >23–40 kg: 60 mg bid; >40 kg: 75 mg bid. For children <1 year of age, see www.cdc.gov/h1n1flu/recommendations.htm.

[b] <15 kg: 30 mg qd; >15–23 kg: 45 mg qd; >23–40 kg: 60 mg qd; >40 kg: 75 mg qd. For children <1 year of age, see www.cdc.gov/h1n1flu/recommendations.htm.

[c] Amantadine and rimantadine are not currently recommended (2009–2010) because of widespread resistance in influenza A viruses. Their use may be reconsidered if viral susceptibility is reestablished.

Amantadine or rimantadine treatment of illness caused by sensitive strains of influenza A virus similarly reduces the duration of symptoms of uncomplicated influenza by ~50% if begun within 48 h of onset of illness. Five to 10% of amantadine recipients experience mild CNS side effects, primarily jitteriness, anxiety, insomnia, or difficulty concentrating. These side effects disappear promptly upon cessation of therapy. Rimantadine appears to be equally efficacious and is associated with less frequent CNS side effects than is amantadine. In adults, the usual dose of amantadine or rimantadine is 200 mg/d for 3–7 days. Since both drugs are excreted via the kidney, the dose should be reduced to ≤100 mg/d in elderly patients and in patients with renal insufficiency.

The epidemiologic patterns of resistance to the influenza antiviral drugs are crucial elements in agent selection. Since 2005–2006, the vast majority of A/H3N2 viruses, including >90% of U.S. isolates, have been resistant to the adamantanes but have remained sensitive to neuraminidase inhibitors. In contrast, the seasonal A/H1N1 viruses that circulated in 2008–2009 remained sensitive to the adamantanes but were resistant to oseltamivir (although still sensitive to zanamivir). The pandemic A/H1N1 viruses that circulated in 2009–2010 were resistant to the adamantanes but sensitive to zanamivir and usually to oseltamivir; a few oseltamivir-resistant isolates were identified. Up-to-date information on patterns of resistance to influenza antiviral drugs is available through www.cdc.gov/flu.

Ribavirin is a nucleoside analogue with activity against influenza A and B viruses in vitro. It has been reported to be variably effective against influenza when administered as an aerosol but ineffective when administered orally. Its efficacy in the treatment of influenza A or B has not been established.

The therapeutic efficacy of antiviral compounds in influenza has been demonstrated primarily in studies of young adults with uncomplicated disease. The effectiveness of these drugs in the treatment or prevention of complications of influenza is unclear. Pooled analyses of observational investigations and some efficacy studies have suggested that treatment with oseltamivir may reduce the frequency of lower respiratory complications and hospitalization. Therapy for primary influenza pneumonia is directed at maintaining oxygenation and is most appropriately undertaken in an intensive care unit, with aggressive respiratory and hemodynamic support as needed.

Antibacterial drugs should be reserved for the treatment of bacterial complications of acute influenza, such as secondary bacterial pneumonia. The choice of antibiotics should be guided by Gram's staining and culture of appropriate specimens of respiratory secretions, such as sputum. If the etiology of a case of bacterial pneumonia is unclear from an examination of respiratory secretions, empirical antibiotics effective against the most common bacterial pathogens in this setting (*S. pneumoniae*, *S. aureus*, and *H. influenzae*) should be selected (Chaps. 134, 135, and 145).

For uncomplicated influenza in individuals at low risk for complications, symptom-based rather than antiviral therapy may be considered. Acetaminophen or nonsteroidal anti-inflammatory agents can be used for relief of headache, myalgia, and fever, but salicylates should be avoided in children <18 years of age because of the possible association with Reye's syndrome. Since cough is ordinarily self-limited, treatment with cough suppressants generally is not indicated; codeine-containing compounds may be employed if the cough is particularly troublesome. Patients should be advised to rest and maintain hydration during acute illness and to return to full activity only gradually after illness has resolved, especially if it has been severe.

Prophylaxis

The major public health measure for prevention of influenza is vaccination. Both inactivated (killed) and live attenuated vaccines are available and are generated from influenza A and B virus isolates that circulated in the previous influenza seasons and are anticipated to circulate in the upcoming season. For inactivated vaccines, 50–80% protection against influenza is expected if the vaccine virus and the currently circulating viruses are closely related. Available inactivated vaccines have been highly purified and are associated with few reactions. Up to 5% of individuals experience low-grade fever and mild systemic symptoms 8–24 h after vaccination, and up to one-third develop mild redness or tenderness at the vaccination site. Since the vaccine used in the United States and many other countries is produced in eggs, individuals with true hypersensitivity to egg products either should be desensitized or should not be vaccinated. Although the 1976 swine influenza vaccine appears to have been associated with an increased frequency of Guillain-Barré syndrome, influenza vaccines administered since 1976 generally have not been. Possible exceptions were noted during the 1992–1993 and 1993–1994 influenza seasons, when there may have been an excess risk of Guillain-Barré syndrome of slightly more than 1 case per million vaccine recipients. However, the overall health risk following influenza outweighs the potential risk associated with vaccination.

A live attenuated influenza vaccine administered by intranasal spray is available. The vaccine is generated by reassortment between currently circulating strains of influenza A and B virus and a cold-adapted, attenuated master strain. The cold-adapted vaccine is well tolerated and highly efficacious (>90% protective) in young children; in one study, it provided protection against a circulating influenza virus that had drifted antigenically away from the vaccine strain. Live attenuated vaccine is approved for use in healthy nonpregnant persons 2–49 years of age.

Historically, the U.S. Public Health Service has recommended influenza vaccination for certain groups at high risk for complications of influenza on the basis of age or underlying disease or for their close contacts (Table 187-2). While such individuals will continue to be the focus of vaccination programs, the recommendations have been progressively expanded. In 2009–2010, immunization of all children 6 months to 18 years of age was recommended; for 2010–2011, recommendations are for immunization of the entire population above the age of 6 months, including adults. This expanded recommendation reflects increased recognition of previously unappreciated risk factors, including obesity, postpartum conditions, and racial or ethnic influences, as well as an appreciation that more widespread use of vaccine is required for influenza control. Inactivated vaccines may be administered safely to immunocompromised patients. Influenza vaccination is not associated with exacerbations of chronic nervous-system diseases such as multiple sclerosis. Vaccine should be administered early in the autumn before influenza outbreaks occur and should then be given annually to maintain immunity against the most current influenza virus strains.

Although antiviral drugs provide chemoprophylaxis against influenza, their use for that purpose has been limited because of concern about current patterns and further development of resistance. Chemoprophylaxis with oseltamivir or zanamivir has been 84–89% efficacious against influenza A and B (Table 187-3). Chemoprophylaxis with amantadine or rimantadine is no longer recommended because of widespread resistance to these drugs. In earlier studies with sensitive viruses, prophylaxis with amantadine or rimantadine (100–200 mg/d) was 70–100% effective against illness associated with influenza A.

Chemoprophylaxis for healthy persons after community exposure generally is not recommended but may be considered for individuals at high risk of complications who have had close contact with an acutely ill person with influenza. During an outbreak, antiviral chemoprophylaxis can be administered simultaneously with inactivated vaccine, since the drugs do not interfere with an immune response to the vaccine. However, concurrent administration of chemoprophylaxis and live attenuated vaccine may interfere with the immune response to the latter. Antiviral drugs should not be administered until at least 2 weeks after administration of live vaccine, and administration of live vaccine should not begin until at least 48 h after antiviral drug administration has been stopped. Chemoprophylaxis may also be considered to control nosocomial outbreaks of influenza. For that purpose, prophylaxis should be instituted promptly when influenza activity is detected and must be continued daily for the duration of the outbreak.

FURTHER READINGS

Beigel JH et al: Avian influenza A (H5N1) infection in humans. N Engl J Med 353:1374, 2005[PMID: 16192482]

Belshe RB et al: The efficacy of live attenuated, cold adapted trivalent, intranasal influenza vaccine in children. N Engl J Med 38:1405, 1998

Dawood FS et al: Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med 360:2605, 2009[PMID: 19423869]

Dolin R: Interpandemic as well as pandemic disease. N Engl J Med 353:2535, 2005[PMID: 16354889]

Fiore AE et al: Prevention and control of influenza with vaccines. Recommendations of the Advisory Committee on Immunization Practices (ACIP), 2010. MMWR Rec Rep 59:1, 2010[PMID: 20689501]

Garten RJ et al: Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325:197, 2009[PMID: 19465683]

Jefferson T et al: Neuraminidase inhibitors for preventing and treating influenza in healthy adults: Systematic review and meta-analyses. Br Med J 339:b5106, 2009

Meltzer MI et al: The economic impact of pandemic influenza in the United States: Priorities for intervention. Emerg Infect Dis 5:659, 1999[PMID: 10511522]

MIST [Management of Influenza in the Southern Hemisphere Trialists] Study Group: Randomized trial of efficacy and safety of inhaled zanamivir in treatment of influenza A and B infections. Lancet 352:1871, 1998

Treanor JJ: Influenza virus, in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier, 2010, pp 2265–2288

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE HUMAN RETROVIRUSES: INTRODUCTION

The retroviruses, which make up a large family (Retroviridae), infect mainly vertebrates. These viruses have a unique replication cycle whereby their genetic information is encoded by RNA rather than DNA. Retroviruses contain an RNA-dependent DNA polymerase (a reverse transcriptase) that directs the synthesis of a DNA form of the viral genome after infection of a host cell. The designation *retrovirus* denotes that information in the form of RNA is transcribed into DNA in the host cell—a sequence that overturned a central dogma of molecular biology: that information passes unidirectionally from DNA to RNA to protein. The observation that RNA was the source of genetic information in the causative agents of certain animal tumors led to a number of paradigm-shifting biologic insights regarding not only the direction of genetic information passage but also the viral etiology of certain cancers and the concept of oncogenes as normal host genes scavenged and altered by a viral vector.

The family Retroviridae includes seven subfamilies **(Table 188-1).** Members of two of the families infect humans with pathologic consequences: the deltaretroviruses, of which human T cell lymphotropic virus (HTLV) type I is the most important in humans; and lentiviruses, of which HIV is the most important in humans.

Table 188-1 Classification of Retroviruses: The Family Retroviridae

Genus	Example(s)	Feature
Alpharetrovirus	Rous sarcoma virus	Contains *src* oncogene
Betaretrovirus	Mouse mammary tumor virus	Exogenous or endogenous
Gammaretrovirus	Abelson murine leukemia virus	Contains *abl* oncogene
Deltaretrovirus	HTLV-I	Causes T cell lymphoma and neurologic disease
Epsilonretrovirus	Walleye dermal sarcoma virus	
Lentivirus	HIV-1, 2	Causes AIDS
Spumavirus	Simian foamy virus	Not known to be pathogenic in humans

The wide variety of interactions of a retrovirus with its host range from completely benign events (e.g., silent carriage of endogenous retroviral sequences in the germ-line genome of many animal species) to rapidly fatal infections (e.g., exogenous infection with an oncogenic virus such as Rous sarcoma virus in chickens). The ability of retroviruses to acquire and alter the structure and function of host cell sequences has revolutionized our understanding of molecular carcinogenesis. The viruses can insert into the germ-line genome of the host cell and behave as a transposable or movable genetic element. They can activate or inactivate genes near the site of integration into the genome. They can rapidly alter their own genome by recombination and mutation under selective environmental stimuli.

Most human viral diseases occur as a consequence of tissue destruction either directly by the virus itself or indirectly by the host's response to the virus. Although these mechanisms are operative in retroviral infections, retroviruses have additional mechanisms of inducing disease, including the malignant transformation of an infected cell and the induction of an immunodeficiency state that renders the host susceptible to opportunistic diseases (infections and neoplasms; Chap. 189).

Structure and Life Cycle

All retroviruses are similar in structure, genome organization, and mode of replication. Retroviruses are 70–130 nm in diameter and have a lipid-containing envelope surrounding an icosahedral capsid with a dense inner core. The core contains two identical copies of the single-strand RNA genome. The RNA molecules are 8–10 kb long and are complexed with reverse transcriptase and tRNA. Other viral proteins, such as integrase, are also components of the virion particle. The RNA has features usually found in mRNA: a cap site at the 5' end of the molecule, which is important in the initiation of mRNA translation, and a polyadenylation site at the 3' end, which influences mRNA turnover (i.e., messages with shorter polyA tails turn over faster than messages with longer polyA tails). However, the retroviral RNA is not translated; instead it is transcribed into DNA. The DNA form of the retroviral genome is called a *provirus*.

The replication cycle of retroviruses proceeds in two phases **(Fig. 188-1)**. In the first phase, the virus

enters the cytoplasm after binding to one or more specific cell-surface receptors; the viral RNA and reverse transcriptase synthesize a double-strand DNA version of the RNA template; and the provirus moves into the nucleus and integrates into the host cell genome. This proviral integration is permanent. Although some animal retroviruses integrate into a single specific site of the host genome in every infected cell, the human retroviruses integrate randomly. This first phase of replication depends entirely on gene products in the virus. The second phase includes the synthesis and processing of viral genomes, mRNAs, and proteins using host cell machinery, often under the influence of viral gene products. Virions are assembled and released from the cell by budding from the membrane; host cell membrane proteins are frequently incorporated into the envelope of the virus. Proviral integration occurs during the S-phase of the cell cycle; thus, in general, nondividing cells are resistant to retroviral infection. Only the lentiviruses are able to infect nondividing cells. Once a host cell is infected, it is infected for the life of the cell.

Figure 188-1

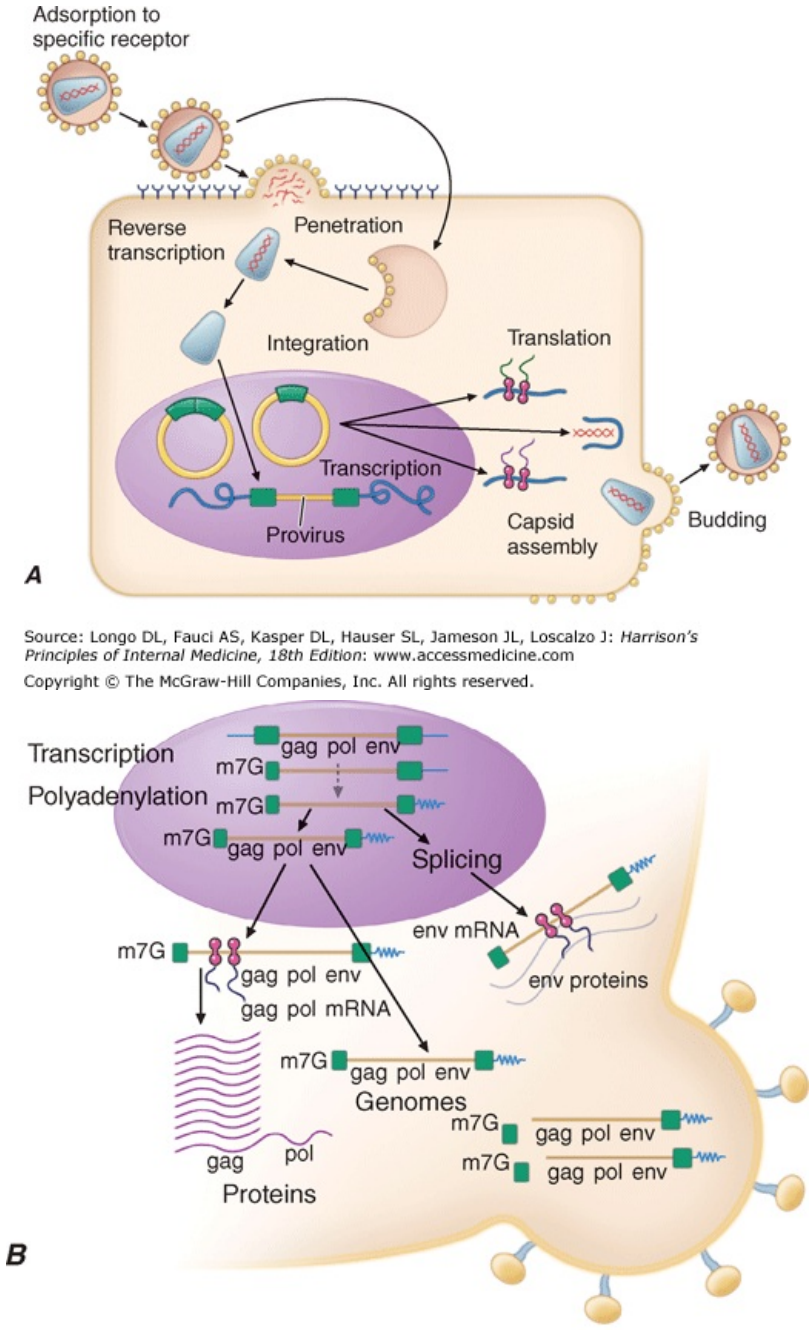

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The life cycle of retroviruses. ***A.*** Overview of virus replication. The retrovirus enters a target cell by binding to a specific cell-surface receptor; once the virus is internalized, its RNA is released from the nucleocapsid and is reverse-transcribed into proviral DNA. The provirus is inserted into the genome and then transcribed into RNA; the RNA is translated; and virions assemble and are extruded

from the cell membrane by budding. ***B.*** Overview of retroviral gene expression. The provirus is transcribed, capped, and polyadenylated. Viral RNA molecules then have one of three fates: they are exported to the cytoplasm, where they are packaged as the viral RNA in infectious viral particles; they are spliced to form the message for the envelope polyprotein; or they are translated into Gag and Pol proteins. Most of the messages for the Pol protein fail to initiate Pol translation because of a stop codon before its initiation; however, in a fraction of the messages, the stop codon is missed and the Pol proteins are translated. *[Modified from JM Coffin, in BN Fields, DM Knipe (eds): Fields Virology. New York, Raven, 1990; with permission.]*

Retroviral genomes include both coding and noncoding sequences **(Fig. 188-2)**. In general, noncoding sequences are important recognition signals for DNA or RNA synthesis or processing events and are located in the 5' and 3' terminal regions of the genome. All retroviral genomes are terminally redundant, containing identical sequences called *long terminal repeats* (LTRs). The ends of the retroviral RNA genome differ slightly in sequence from the integrated retroviral DNA. In the latter, the LTR sequences are repeated in both the 5' and the 3' terminus of the virus. The LTRs contain sequences involved in initiating the expression of the viral proteins, the integration of the provirus, and the polyadenylation of viral RNAs. The primer binding site, which is critical for the initiation of reverse transcription, and the viral packaging sequences are located outside the LTR sequences. The coding regions include the *gag* (group-specific antigen, core protein), *pol* (RNA-dependent DNA polymerase), and *env* (envelope) genes. The *gag* gene encodes a precursor polyprotein that is cleaved to form three to five capsid proteins; a fraction of the Gag precursor proteins also contain a protease responsible for cleaving the Gag and Pol polyproteins. A Gag-Pol polyprotein gives rise to the protease that is responsible for cleaving the Gag-Pol polyprotein. The *pol* gene encodes three proteins: the reverse transcriptase, the integrase, and the protease. The reverse transcriptase copies the viral RNA into the double-strand DNA provirus, which inserts itself into the host cell DNA via the action of integrase. The protease cleaves the Gag-Pol polyprotein into smaller protein products. The *env* gene encodes the envelope glycoproteins: one protein that binds to specific surface receptors and determines what cell types can be infected and a smaller transmembrane protein that anchors the complex to the envelope. **Fig. 188-3** shows how the retroviral gene products make up the virus structure.

Figure 188-2

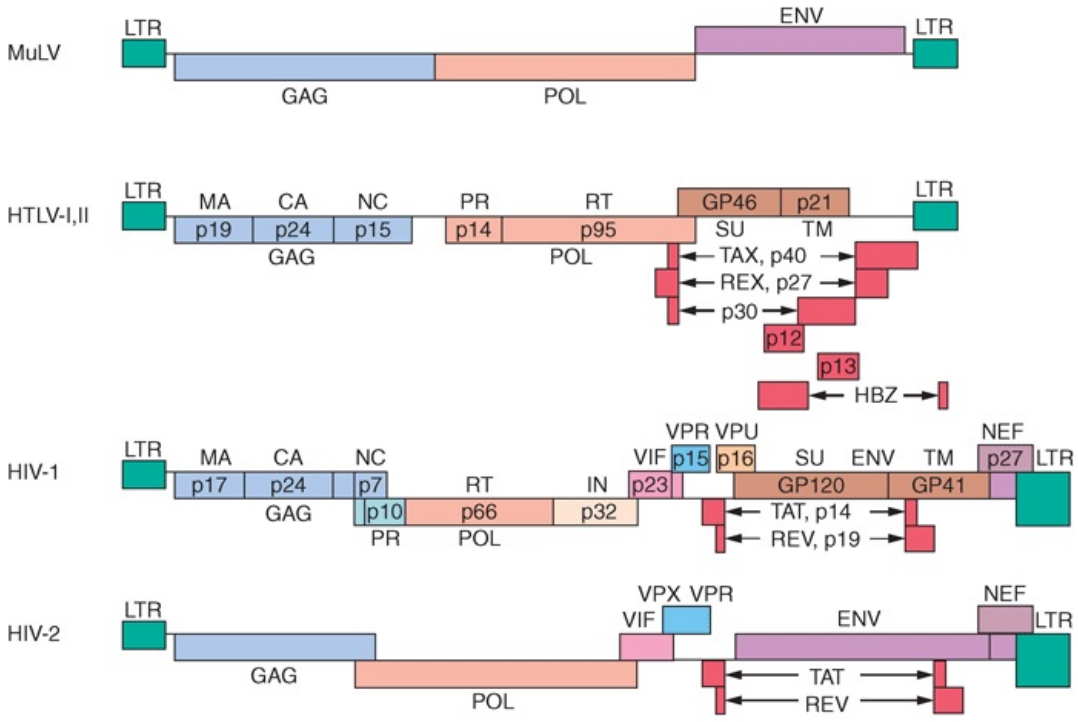

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Genomic structure of retroviruses. The murine leukemia virus MuLV has the typical three structural genes: *gag*, *pol*, and *env*. The *gag* region gives rise to three proteins: matrix (MA), capsid (CA), and nucleic acid–binding (NC) proteins. The *pol* region encodes both a protease (PR) responsible for cleaving the viral polyproteins and a reverse transcriptase (RT). In addition, HIV *pol* encodes an integrase (IN). The *env* region encodes a surface protein (SU) and a small transmembrane protein (TM). The human retroviruses have additional gene products translated in each of the three possible reading frames. HTLV-I and HTLV-II have *tax* and *rex* genes with exons on either side of the *env* gene. HIV-1 and HIV-2 have six accessory gene products: *tat*, *rev*, *vif*, *nef*, *vpr*,

and either *vpu* (in HIV-1) or *vpx* (in HIV-2). The genes for these proteins are located mainly between the *pol* and *env* genes. LTR, long terminal repeat; GP, glycoprotein; HBZ, HTLV-I basic leucine zipper domain-containing protein.

Figure 188-3

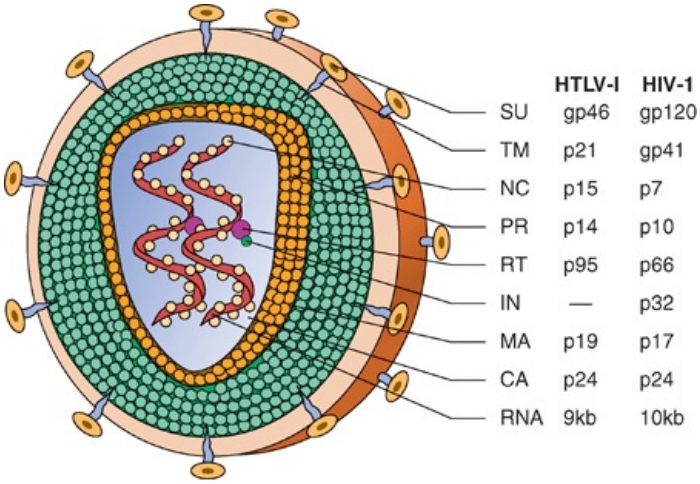

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic structure of human retroviruses. The surface glycoprotein (SU) is responsible for binding to receptors of host cells. The transmembrane protein (TM) anchors SU to the virus. NC is a nucleic acid–binding protein found in association with the viral RNA. A protease (PR) cleaves the polyproteins encoded by the *gag*, *pol*, and *env* genes into their functional components. RT is reverse transcriptase, and IN is an integrase present in some retroviruses (e.g., HIV-1) that facilitates insertion of the provirus into the host genome. The matrix protein (MA) is a Gag protein closely associated with the lipid of the envelope. The capsid protein (CA) forms the major internal structure of the virus, the core shell.

HTLVs have a region between *env* and the 3' LTR that encodes several proteins and transcripts in overlapping reading frames (Fig. 188-2). Tax is a 40-kDa protein that does not bind to DNA but induces the expression of host cell transcription factors that alter host cell gene expression and is capable of inducing cell transformation under certain circumstances. Rex is a 27-kDa protein that regulates the expression of viral mRNAs. Other transcripts from this region (p12, p13, p30) tend to restrict expression of viral genes and diminish the immunogenicity of infected cells. HBZ is a product of the complementary proviral DNA strand, and its protein interacts with many cellular transcription factors and signaling proteins. It stimulates proliferation of infected cells and is the only viral product universally expressed in HTLV-I-infected tumor cells. These proteins are produced from messages that are similar but that are spliced differently from overlapping but distinct exons.

The lentiviruses in general, and HIV-1 and -2 in particular, contain a larger genome than other pathogenic retroviruses. They contain an untranslated region between *pol* and *env* that encodes portions of several proteins, varying with the reading frame into which the mRNA is spliced. Tat is a 14-kDa protein that augments the expression of virus from the LTR. The Rev protein of HIV-1, similar to the Rex protein of HTLV, regulates RNA splicing and/or RNA transport. The Nef protein downregulates CD4, the cellular receptor for HIV; alters host T cell activation pathways; and enhances viral infectivity. The Vif protein is necessary for the proper assembly of the HIV nucleoprotein core in many types of cells; without Vif, proviral DNA is not efficiently produced in these infected cells. In addition, the Vif protein targets APOBEC (apolipoprotein B mRNA-editing enzyme catalytic polypeptide, a cytidine deaminase that mutates the viral sequence) for proteasomal degradation, thus blocking its virus-suppressing effect. Vpr, Vpu (HIV-1 only), and Vpx (HIV-2 only) are viral proteins encoded by translation of the same message in different reading frames. As noted above, oncogenic retroviruses depend on cell proliferation for their replication; lentiviruses can infect nondividing cells, largely through effects mediated by Vpr. Vpr facilitates transport of the provirus into the nucleus and can induce other cellular changes, such as G_2 growth arrest and differentiation of some target cells. Vpx is structurally related to Vpr, but its functions are not fully defined. Vpu promotes the degradation of CD4 in the endoplasmic reticulum and stimulates the release of virions from infected cells.

Retroviruses can be either exogenously acquired (by infection with an infected cell or a free virion capable of replication) or transmitted in the germ line as endogenous virus. Endogenous retroviruses are often replication defective. The human genome contains endogenous retroviral sequences, but there are no known replication-competent endogenous retroviruses in humans.

In general, viruses that contain only the *gag*, *pol*, and *env* genes either are not pathogenic or take a long time to induce disease; these observations indicate the importance of the other regulatory genes in viral disease pathogenesis. The pathogenesis of neoplastic transformation by retroviruses relies on the chance integration of the provirus at a spot in the genome resulting in the expression of a cellular gene (proto-oncogene) that becomes transforming by virtue of its unregulated expression. For example, avian leukosis virus causes B cell leukemia by inducing the expression of *myc*. Some retroviruses possess captured and altered cellular genes near their integration site, and these viral oncogenes can transform the infected host cell. Viruses that have oncogenes often have lost a portion of their genome that is required for replication. Such viruses need helper viruses to reproduce, a feature that may explain why these acute transforming retroviruses are rare in nature. All human retroviruses identified to date are exogenous and are not acutely transforming (i.e., they lack a transforming oncogene).

These remarkable properties of retroviruses have led to experimental efforts to use them as vectors to insert specific genes into particular cell types, a process known as *gene therapy* or *gene transfer*. The process could be used to repair a genetic defect or to introduce a new property that could be used therapeutically; for example, a gene (e.g., thymidine kinase) that would make a tumor cell susceptible to killing by a drug (e.g., ganciclovir) could be inserted. One source of concern about the use of retroviral vectors in humans is that replication-competent viruses might rescue endogenous retroviral replication, with unpredictable results. This concern is not merely hypothetical: the detection of proteins encoded by endogenous retroviral sequences on the surface of cancer cells implies that the genetic events leading to the cancer were able to activate the synthesis of these usually silent genes.

HUMAN T CELL LYMPHOTROPIC VIRUS

HTLV-I was isolated in 1980 from a T cell lymphoma cell line from a patient originally thought to have cutaneous T cell lymphoma. Later it became clear that the patient had a distinct form of lymphoma (originally reported in Japan) called *adult T cell leukemia/lymphoma* (ATL). Serologic data have determined that HTLV-I is the cause of at least two important diseases: ATL and tropical spastic paraparesis, also called *HTLV-I-associated myelopathy* (HAM). HTLV-I may also play a role in infective dermatitis, arthritis, uveitis, and Sjögren's syndrome.

Two years after the isolation of HTLV-I, HTLV-II was isolated from a patient with an unusual form of hairy cell leukemia that affected T cells. Epidemiologic studies of HTLV-II failed to reveal a consistent disease association.

Biology and Molecular Biology

Because the biology of HTLV-I and that of HTLV-II are similar, the following discussion will focus on HTLV-I.

Human glucose transporter protein 1 (GLUT-1) functions as a receptor for HTLV-1, probably acting together with neuropilin-1 (NRP1) and heparan sulfate proteoglycans. Generally, only T cells are productively infected, but infection of B cells and other cell types is occasionally detected. The most common outcome of HTLV-I infection is latent carriage of randomly integrated provirus in CD4+ T cells. HTLV-I does not contain an oncogene and does not insert into a unique site in the genome. Indeed, most infected cells express no viral gene products. The only viral gene product that is routinely expressed in tumor cells transformed by HTLV-I in vivo is *hbz*. The *tax* gene is thought to be critical to the transformation process but is not expressed in the tumor cells of many ATL patients, possibly because of the immunogenicity of *tax*-expressing cells. Cells transformed in vitro, by contrast, actively transcribe HTLV-I RNA and produce infectious virions. Most HTLV-I-transformed cell lines are the result of the infection of a normal host T cell in vitro. It is difficult to establish cell lines derived from authentic ATL cells.

Although *tax* does not itself bind to DNA, it does induce the expression of a wide range of host cell gene products, including transcription factors (especially c-rel/NF-κB, ets-1 and -2, and members of the fos/jun family), cytokines [e.g., interleukin (IL) 2, granulocyte-macrophage colony-stimulating factor, and tumor necrosis factor], and membrane proteins and receptors [major histocompatibility (MHC) molecules and IL-2 receptor α]. The genes activated by *tax* are generally controlled by transcription factors of the c-rel/NF-κB and cyclic AMP response element binding (CREB) protein families. It is unclear how this induction of host gene expression leads to neoplastic transformation; *tax* can interfere with G_1 and mitotic cell-cycle checkpoints, block apoptosis, inhibit DNA repair, and promote antigen-independent T cell proliferation. Induction of a cytokine-autocrine loop has been proposed; however, IL-2 is not the crucial cytokine. The involvement of IL-4, IL-7, and IL-15 has been proposed.

In light of the irregular expression of *tax* in ATL cells, it has been suggested that *tax* is important in the early phases of transformation but is not essential for the maintenance of the transformed state. The maintenance role is thought to be due to *hbz* expression. As is clear from the epidemiology of HTLV-I infection, transformation of an infected cell is a rare event and may depend on heterogeneous second, third, or fourth genetic hits. No consistent chromosomal abnormalities have been described in ATL; however, aneuploidy is common and individual cases with p53 mutations and translocations involving

the T cell receptor genes on chromosome 14 have been reported. *Tax* may repress certain DNA repair enzymes, permitting the accumulation of genetic damage that would normally be repaired. However, the molecular pathogenesis of HTLV-I-induced neoplasia is not fully understood.

Features of HTLV-I Infection

EPIDEMIOLOGY

HTLV-I infection is transmitted in at least three ways: from mother to child, especially via breast milk; through sexual activity, more commonly from men to women; and through the blood—via contaminated transfusions or contaminated needles. The virus is most commonly transmitted perinatally. Compared with HIV, which can be transmitted in cell-free form, HTLV-I is less infectious, and its transmission usually requires cell-to-cell contact.

HTLV-I is endemic in southwestern Japan and Okinawa, where >1 million persons are infected. Antibodies to HTLV-I are present in the serum of up to 35% of Okinawans, 10% of residents of the Japanese island of Kyushu, and <1% of persons in nonendemic regions of Japan. Despite this high prevalence of infection, only ~500 cases of ATL are diagnosed in this area each year. Clusters of infection have been noted in other areas of the Orient, such as Taiwan; in the Caribbean basin, including northeastern South America; in northwestern South America; in central and southern Africa; in Italy, Israel, Iran, and Papua New Guinea; in the Arctic; and in the southeastern part of the United States **(Fig. 188-4)**. An estimated 15–20 million persons have HTLV-I infection worldwide.

Figure 188-4

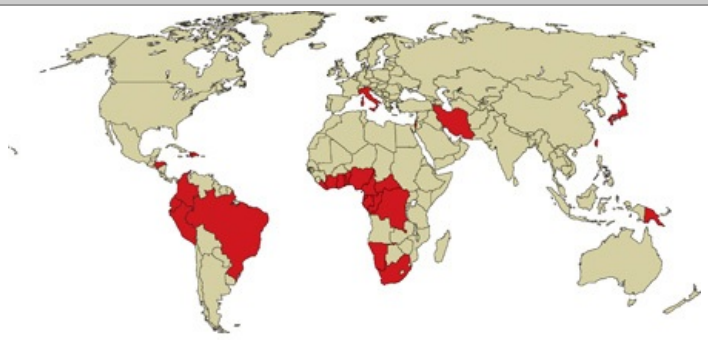

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Global distribution of HTLV-I infection. Countries with a prevalence of HTLV-I infection of 1–5% are shaded darkly. Note that the distribution of infected patients is not uniform in endemic countries. For example, the people of southwestern Japan and northeastern Brazil are more commonly affected than those in other regions of those countries.

Progressive spastic or ataxic myelopathy developing in an individual who is HTLV-I positive (i.e., who has serum antibodies to HTLV-I) may be due to direct infection of the nervous system with the virus, but destruction of the pyramidal tracts appears to involve HTLV-I-infected CD4+ T cells; a similar disorder may result from infection with HIV or HTLV-II. In rare instances, patients with HAM are seronegative but have detectable antibody to HTLV-I in the cerebrospinal fluid (CSF).

The cumulative lifetime risk of developing ATL is 3% among HTLV-I-infected patients, with a threefold greater risk among men than among women; a similar cumulative risk is projected for HAM (4%), but with women more commonly affected than men. The distribution of the two diseases overlaps the distribution of HTLV-I, with >95% of affected patients showing serologic evidence of HTLV-I infection. The latency period between infection and the emergence of disease is 20–30 years for ATL. For HAM, the median latency period is ~3.3 years (range, 4 months to 30 years). The development of ATL is rare among persons infected by blood products; however, ~20% of patients with HAM acquire HTLV-I from contaminated blood. ATL is more common among perinatally infected individuals, whereas HAM is more common among persons infected via sexual transmission.

ASSOCIATED DISEASES

ATL

Four clinical types of HTLV-I-induced neoplasia have been described: acute, lymphomatous, chronic, and smoldering. All of these tumors are monoclonal proliferations of CD4+ postthymic T cells with clonal proviral integrations and clonal T cell receptor gene rearrangements.

Acute ATL

About 60% of patients who develop malignancy have classic acute ATL, which is characterized by a short clinical prodrome (~2 weeks between the first symptoms and the diagnosis) and an aggressive natural history (median survival period, 6 months). The clinical picture is dominated by rapidly

progressive skin lesions, pulmonary involvement, hypercalcemia, and lymphocytosis with cells containing lobulated or "flower-shaped" nuclei (see Fig. 110-10). The malignant cells have monoclonal proviral integrations and express CD4, CD3, and CD25 (low-affinity IL-2 receptors) on their surface. Serum levels of CD25 can be used as a tumor marker. Anemia and thrombocytopenia are rare. The skin lesions may be difficult to distinguish from those in mycosis fungoides. Lytic bone lesions, which are common, do not contain tumor cells but rather are composed of osteolytic cells, usually without osteoblastic activity. Despite the leukemic picture, bone marrow involvement is patchy in most cases.

The hypercalcemia of ATL is multifactorial; the tumor cells produce osteoclast-activating factors (tumor necrosis factor α, IL-1, lymphotoxin) and can also produce a parathyroid hormone–like molecule. Affected patients have an underlying immunodeficiency that makes them susceptible to opportunistic infections similar to those seen in patients with AIDS (Chap. 189). The pathogenesis of the immunodeficiency is unclear. Pulmonary infiltrates in ATL patients reflect leukemic infiltration half the time and opportunistic infections with organisms such as *Pneumocystis* and other fungi the other half. Gastrointestinal symptoms are nearly always related to opportunistic infection. *Strongyloides stercoralis* is a gastrointestinal parasite that has a pattern of endemic distribution similar to that of HTLV-I. HTLV-I-infected persons also infected with this parasite may develop ATL more often or more rapidly than those without *Strongyloides* infections. Serum concentrations of lactate dehydrogenase and alkaline phosphatase are often elevated in ATL. About 10% of patients have leptomeningeal involvement leading to weakness, altered mental status, paresthesia, and/or headache. Unlike other forms of central nervous system (CNS) lymphoma, ATL may be accompanied by normal CSF protein levels. The diagnosis depends on finding ATL cells in the CSF (Chap. 110).

Lymphomatous ATL

The lymphomatous type of ATL occurs in ~20% of patients and is similar to the acute form in its natural history and clinical course, except that circulating abnormal cells are rare and lymphadenopathy is evident. The histology of the lymphoma is variable but does not influence the natural history. In general, the diagnosis is suspected on the basis of the patient's birthplace (see "Epidemiology," above) and the presence of skin lesions and hypercalcemia. The diagnosis is confirmed by the detection of antibodies to HTLV-I in serum.

Chronic ATL

Patients with the chronic form of ATL generally have normal serum levels of calcium and lactate dehydrogenase and no involvement of the CNS, bone, or gastrointestinal tract. The median duration of survival for these patients is 2 years. In some cases, chronic ATL progresses to the acute form of the disease.

Smoldering ATL

Fewer than 5% of patients have the smoldering form of ATL. In this form, the malignant cells have monoclonal proviral integration; <5% of peripheral blood cells exhibit typical morphologic abnormalities; hypercalcemia, adenopathy, and hepatosplenomegaly do not develop; the CNS, the bones, and the gastrointestinal tract are not involved; and skin lesions and pulmonary lesions may be present. The median survival period for this small subset of patients appears to be ≥5 years.

HAM (Tropical Spastic Paraparesis)

In contrast to ATL, in which there is a slight predominance of male patients, HAM affects female patients disproportionately. HAM resembles multiple sclerosis in certain ways (Chap. 380). The onset is insidious. Symptoms include weakness or stiffness in one or both legs, back pain, and urinary incontinence. Sensory changes are usually mild, but peripheral neuropathy may develop. The disease generally takes the form of slowly progressive and unremitting thoracic myelopathy; one-third of patients are bedridden within 10 years of diagnosis, and one-half are unable to walk unassisted by this point. Patients display spastic paraparesis or paraplegia with hyperreflexia, ankle clonus, and extensor plantar responses. Cognitive function is usually spared; cranial nerve abnormalities are unusual.

MRI reveals lesions in both the white matter and the paraventricular regions of the brain as well as in the spinal cord. Pathologic examination of the spinal cord shows symmetric degeneration of the lateral columns, including the corticospinal tracts; some cases involve the posterior columns as well. The spinal meninges and cord parenchyma contain an inflammatory infiltrate with myelin destruction.

HTLV-I is not usually found in cells of the CNS but may be detected in a small population of lymphocytes present in the CSF. In general, HTLV-I replication is greater in HAM than in ATL, and patients with HAM have a stronger immune response to the virus. Antibodies to HTLV-I are present in the serum and appear to be produced in the CSF of HAM patients, where titers are often higher than in the serum. The pathophysiology of HAM may involve the induction of autoimmune destruction of neural cells by T cells with specificity for viral components such as Tax or Env proteins. One theory is that susceptibility to HAM may be related to the presence of human leukocyte antigen (HLA)alleles capable of presenting viral antigens in a fashion that leads to autoimmunity. Insufficient data are available to confirm an HLA association. However, antibodies in the sera of HAM patients have been shown to bind a neuron-specific antigen [heteronuclear ribonuclear protein A1 (hnRNP A1)] and to

interfere with neurotransmission in vitro.

It is unclear what factors influence whether HTLV-I infection will cause disease and, if it does, whether it will induce a neoplasm (ATL) or an autoimmune disorder (HAM). Differences in viral strains, the susceptibility of particular MHChaplotypes, the route of HTLV-I infection, the viral load, and the nature of the HTLV-I-related immune response are putative factors, but few definitive data are available.

Other Putative HTLV-I-Related Diseases

In areas where HTLV-I is endemic, diverse inflammatory and autoimmune diseases have been attributed to the virus, including uveitis, dermatitis, pneumonitis, rheumatoid arthritis, and polymyositis. However, a causal relationship between HTLV-I and these illnesses has not been established.

PREVENTION

Women in endemic areas should not breast-feed their children, and blood donors should be screened for serum antibodies to HTLV-I. As in the prevention of HIV infection, the practice of safe sex and the avoidance of needle sharing are important.

Treatment: HTLV-I Infection

For the small number of patients who develop HTLV-I-related disease, therapies are not curative. In patients with the acute and lymphomatous types of ATL, the disease progresses rapidly. Hypercalcemia is generally controlled by glucocorticoid administration and cytotoxic therapy directed against the neoplasm. The tumor is highly responsive to combination chemotherapy that is employed against other forms of lymphoma; however, patients are susceptible to overwhelming bacterial and opportunistic infections, and ATL relapses within 4–10 months after remission in most cases. The combination of interferon α and zidovudine may extend survival. Because viral replication is not clearly associated with ATL progression, zidovudine is probably effective through its cytotoxic effects (as a chain-terminating thymidine analogue) rather than its antiviral effects. An experimental approach using an yttrium 90–labeled or toxin-conjugated antibody to the IL-2 receptor appears promising but is not widely available. Patients with the chronic or smoldering form of ATL may be managed with an expectant approach: treat any infections, and watch and wait for signs of progression to acute disease.

Patients with HAM may obtain some benefit from the use of glucocorticoids to reduce inflammation. Antiretroviral regimens have not been effective. In one study, danazol (200 mg three times daily) produced significant neurologic improvement in five of six treated patients, with resolution of urinary incontinence in two cases, decreased spasticity in three, and restoration of the ability to walk after confinement to a wheelchair in two. Physical therapy and rehabilitation are important components of management.

Features of HTLV-II Infection

EPIDEMIOLOGY

HTLV-II is endemic in certain Native American tribes and in Africa. It is generally considered to be a New World virus that was brought from Asia to the Americas 10,000–40,000 years ago during the migration of infected populations across the Bering land bridge. The mode of transmission of HTLV-II is probably the same as that of HTLV-I (see above). HTLV-II may be less readily transmitted sexually than HTLV-I.

Studies of large cohorts of injection drug users with serologic assays that reliably distinguish HTLV-I from HTLV-II indicated that the vast majority of HTLV-positive cohort members were infected with HTLV-II. The seroprevalence of HTLV in a cohort of 7841 injection drug users from drug treatment centers in Baltimore, Chicago, Los Angeles, New Jersey (Asbury Park and Trenton), New York City (Brooklyn and Harlem), Philadelphia, and San Antonio was 20.9%, with >97% of cases due to HTLV-II. The seroprevalence of HTLV-II was higher in the Southwest and the Midwest than in the Northeast. In contrast, the seroprevalence of HIV-1 was higher in the Northeast than in the Southwest or the Midwest. Approximately 3% of the cohort members were infected with both HTLV-II and HIV-1. The seroprevalence of HTLV-II increased linearly with age. Women were significantly more likely to be infected with HTLV-II than were men; the virus is thought to be more efficiently transmitted from male to female than from female to male.

ASSOCIATED DISEASES

Although HTLV-II was isolated from a patient with a T cell variant of hairy cell leukemia, this virus has not been consistently associated with a particular disease and in fact has been thought of as "a virus searching for a disease." However, evidence is accumulating that HTLV-II may play a role in certain neurologic, hematologic, and dermatologic diseases. These data require confirmation, particularly in light of the previous confusion regarding the relative prevalences of HTLV-I and HTLV-II among injection drug users.

PREVENTION

Avoidance of needle sharing, adherence to safe-sex practices, screening of blood (by assays for HTLV-I, which also detect HTLV-II), and avoidance of breast-feeding by infected women are important

principles in the prevention of spread of HTLV-II.

HUMAN IMMUNODEFICIENCY VIRUS

HIV-1 and HIV-2 are members of the lentivirus subfamily of Retroviridae and are the only lentiviruses known to infect humans. The lentiviruses are slow-acting by comparison with viruses that cause acute infection (e.g., influenza virus) but not by comparison with other retroviruses. The features of acute primary infection with HIV resemble those of more classic acute infections. The characteristic chronicity of HIV disease is consistent with the designation *lentivirus*. **For a detailed discussion of HIV, see Chap. 189.**

FURTHER READINGS

Boxus M, Willems L: Mechanisms of HTLV-1 persistence and transformation. Br J Cancer 101:1497, 2009[PMID: 19861996]

Lee SM et al: HTLV-1 induced molecular mimicry in neurological disease. Curr Top Microbiol Immunol 296:125, 2005[PMID: 16323422]

Legros S et al: Protein-protein interactions and gene expression regulation in HTLV-1-infected cells. Front Biosci 14:4138, 2009[PMID: 19273340]

Matsuoka M, Jeang KT: Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. Nat Rev Cancer 7:270, 2007[PMID: 17384582]

Mori N: Cell signaling modifiers for molecular targeted therapy in ATLL. Front Biosci 14:1479, 2009[PMID: 19273141]

Nyborg JK et al: The HTLV-1 Tax protein: Revealing mechanisms of transcriptional activation through histone acetylation and nucleosome disassembly. Biochim Biophys Acta 1799:266, 2010[PMID: 19782779]

Proietti FA et al: Global epidemiology of HTLV-I infection and associated diseases. Oncogene 24:6058, 2005[PMID: 16155612]

Tsukasaki K et al: Definition, prognostic factors, treatment, and response criteria of adult-T-cell leukemia-lymphoma: A proposal from an international consensus meeting. J Clin Oncol 27:453, 2008[PMID: 19064971]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HUMAN IMMUNODEFICIENCY VIRUS DISEASE: AIDS AND RELATED DISORDERS: INTRODUCTION

AIDS was first recognized in the United States in the summer of 1981, when the U.S. Centers for Disease Control and Prevention (CDC) reported the unexplained occurrence of *Pneumocystis jiroveci* (formerly *P. carinii*) pneumonia in five previously healthy homosexual men in Los Angeles and of Kaposi's sarcoma (KS) with or without *P. jiroveci* pneumonia in 26 previously healthy homosexual men in New York and Los Angeles. The disease was soon recognized in male and female injection drug users; in hemophiliacs and blood transfusion recipients; among female sexual partners of men with AIDS; and among infants born to mothers with AIDS or with a history of injection drug use. In 1983, human immunodeficiency virus (HIV) was isolated from a patient with lymphadenopathy, and by 1984 it was demonstrated clearly to be the causative agent of AIDS. In 1985, a sensitive enzyme-linked immunosorbent assay (ELISA) was developed; this led to an appreciation of the scope and evolution of the HIV epidemic at first in the United States and other developed nations and ultimately among developing nations throughout the world (see "HIV infection and AIDS Worldwide" below). The staggering worldwide evolution of the HIV pandemic has been matched by an explosion of information in the areas of HIV virology, pathogenesis (both immunologic and virologic), treatment of HIV disease, treatment and prophylaxis of the opportunistic diseases associated with HIV infection, prevention of infection, and vaccine development. The information flow related to HIV disease is enormous and continues to expand, and it has become almost impossible for the health care generalist to stay abreast of the literature. The purpose of this chapter is to present the most current information available on the scope of the epidemic; on its pathogenesis, treatment, and prevention; and on prospects for vaccine development. Above all, the aim is to provide a solid scientific basis and practical clinical guidelines for a state-of-the-art approach to the HIV-infected patient.

Definition

The current CDC classification system for HIV-infected adolescents and adults categorizes persons on the basis of clinical conditions associated with HIV infection and CD4+ T lymphocyte counts. The system is based on three ranges of CD4+ T lymphocyte counts and three clinical categories and is represented by a matrix of nine mutually exclusive categories **(Tables 189-1** and **189-2)**. Using this system, any HIV-infected individual with a CD4+ T cell count of <200/μL has AIDS by definition, regardless of the presence of symptoms or opportunistic diseases (Table 189-1). Once individuals have had a clinical condition in category B, their disease classification cannot be reverted back to category A, even if the condition resolves; the same holds true for category C in relation to category B.

Table 189-1 1993 Revised Classification System for HIV Infection and Expanded AIDS Surveillance Case Definition for Adolescents and Adults

	Clinical Categories		
CD4+ T Cell Categories	**A Asymptomatic, Acute (Primary) HIV or PGL**	**B Symptomatic, Not A or C Conditions**	**C AIDS-Indicator Conditions**
>500/μL	A1	B1	C1
200–499/μL	A2	B2	C2
<200/μL	A3	B3	C3

Abbreviation: PGL, progressive generalized lymphadenopathy.

Source: MMWR 42(No. RR-17), December 18, 1992.

Table 189-2 Clinical Categories of HIV Infection

Category A: Consists of one or more of the conditions listed below in an adolescent or adult (>13 years) with documented HIV infection. Conditions listed in categories B and C must not have occurred.
Asymptomatic HIV infection
Persistent generalized lymphadenopathy
Acute (primary) HIV infection with accompanying illness or history of acute HIV infection
Category B: Consists of symptomatic conditions in an HIV-infected adolescent
or adult that are not included among conditions listed in clinical category C and that meet at least one of the following criteria: (1) The conditions are attributed to HIV infection or are indicative of a defect in cell-mediated immunity; or (2) the conditions are considered by physicians to have a clinical course or to require management that is complicated by HIV infection. Examples include, but are not limited to, the following:
Bacillary angiomatosis
Candidiasis, oropharyngeal (thrush)
Candidiasis, vulvovaginal; persistent, frequent, or poorly responsive to therapy
Cervical dysplasia (moderate or severe)/cervical carcinoma in situ
Constitutional symptoms, such as fever (38.5°C) or diarrhea lasting >1 month
Hairy leukoplakia, oral
Herpes zoster (shingles), involving at least two distinct episodes or more than one dermatome
Idiopathic thrombocytopenic purpura
Listeriosis
Pelvic inflammatory disease, particularly if complicated by tuboovarian abscess
Peripheral neuropathy
Category C: Conditions listed in the AIDS surveillance case definition.
Candidiasis of bronchi, trachea, or lungs

Candidiasis, esophageal
Cervical cancer, invasive[a]
Coccidioidomycosis, disseminated or extrapulmonary
Cryptococcosis, extrapulmonary
Cryptosporidiosis, chronic intestinal (>1 month's duration)
Cytomegalovirus disease (other than liver, spleen, or nodes)
Cytomegalovirus retinitis (with loss of vision)
Encephalopathy, HIV-related
Herpes simplex: chronic ulcer(s) (>1 month's duration); or bronchitis, pneumonia, or esophagitis
Histoplasmosis, disseminated or extrapulmonary
Isosporiasis, chronic intestinal (>1 month's duration)
Kaposi's sarcoma
Lymphoma, Burkitt's (or equivalent term)
Lymphoma, primary, of brain
Mycobacterium avium complex or *M. kansasii*, disseminated or extrapulmonary
Mycobacterium tuberculosis, any site (pulmonary[a] or extrapulmonary)
Mycobacterium, other species or unidentified species, disseminated or extrapulmonary
Pneumocystis jiroveci pneumonia
Pneumonia, recurrent[a]
Progressive multifocal leukoencephalopathy
Salmonella septicemia, recurrent
Toxoplasmosis of brain
Wasting syndrome due to HIV

[a]Added in the 1993 expansion of the AIDS surveillance case definition.

Source: MMWR 42(No. RR-17), December 18, 1992.

The definition of AIDS is indeed complex and comprehensive and was established not for the practical care of patients, but for surveillance purposes. Thus, the clinician should not focus on whether the patient fulfills the strict definition of AIDS but should view HIV disease as a spectrum ranging from primary infection, with or without the acute syndrome, to the asymptomatic stage, to advanced stages associated with opportunistic diseases (see "Pathophysiology and Pathogenesis," below).

ETIOLOGIC AGENT

HIV is the etiologic agent of AIDS; it belongs to the family of human retroviruses (Retroviridae) and the subfamily of lentiviruses (Chap. 188). Nononcogenic lentiviruses cause disease in other animal species, including sheep, horses, goats, cattle, cats, and monkeys. The four retroviruses known to cause human disease belong to two distinct groups: the human T lymphotropic viruses (HTLV)-I and HTLV-II, which are transforming retroviruses; and the human immunodeficiency viruses, HIV-1 and HIV-2, which cause cytopathic effects either directly or indirectly (Chap. 188). The most common cause of HIV disease throughout the world, and certainly in the United States, is HIV-1, which comprises several subtypes with different geographic distributions (see "Molecular Heterogeneity of HIV-1," below). HIV-2 was first identified in 1986 in West African patients and was originally confined to West Africa. However, a number of cases that generally can be traced to West Africa or to sexual contacts with West Africans have been identified throughout the world. The currently defined groups of HIV-1 (M, N, O, P) and the HIV-2 groups A through G each are likely derived from a separate transfer to humans from a nonhuman primate reservoir. HIV-1 viruses likely came from chimpanzees and/or gorillas, and HIV-2 from sooty mangabeys. The AIDS pandemic is primarily caused by the HIV-1 M group viruses. Although HIV-1 group O and HIV-2 viruses have been found in numerous countries, including those in the developed world, they have caused much more localized epidemics. The taxonomic relationship between primate lentiviruses is shown in **Fig. 189-1**.

Figure 189-1

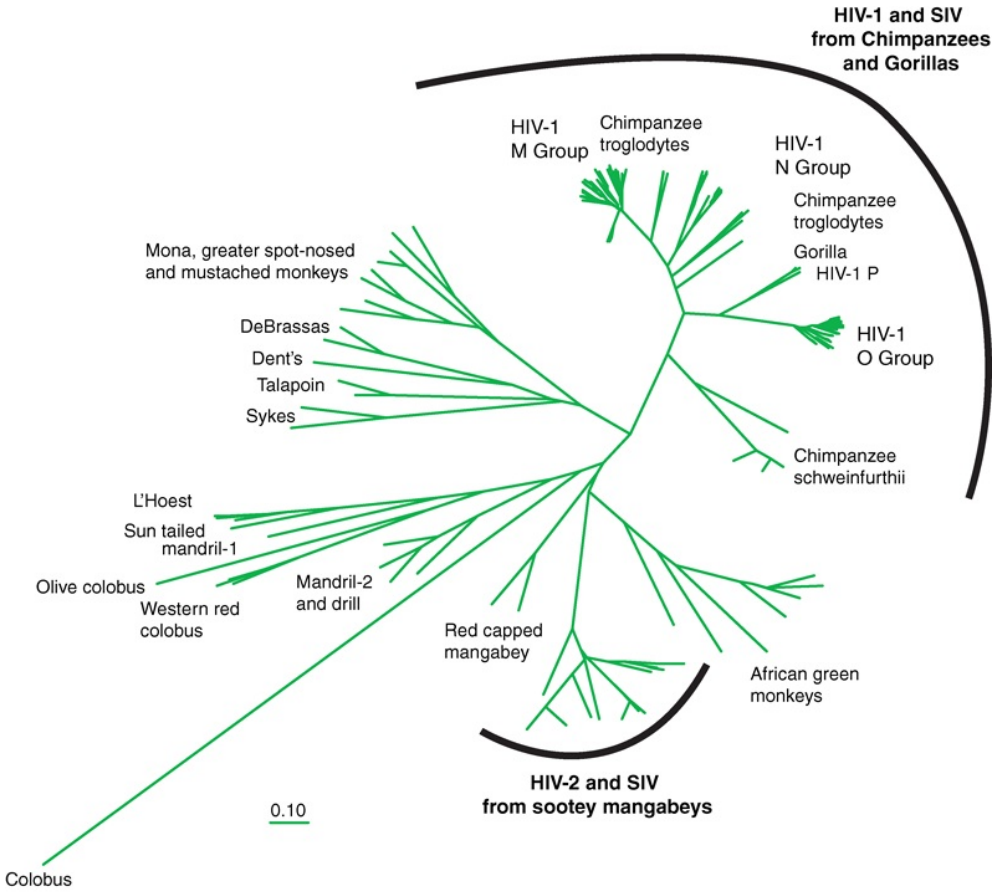

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A phylogenetic tree based on the complete genomes of primate immunodeficiency viruses. The scale (0.10) indicates a 10% difference at the nucleotide level. *(Prepared by Brian Foley, PhD, of the HIV Sequence Database, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; additional information at www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes.html.)*

Morphology of HIV

Electron microscopy shows that the HIV virion is an icosahedral structure **(Fig. 189-2)** containing numerous external spikes formed by the two major envelope proteins, the external gp120 and the transmembrane gp41. The virion buds from the surface of the infected cell and incorporates a variety of host proteins, including major histocompatibility complex (MHC) class I and II antigens (Chap. 315), into its lipid bilayer. The structure of HIV-1 is schematically diagrammed in **Fig. 189-2*B*** (Chap. 188).

Figure 189-2

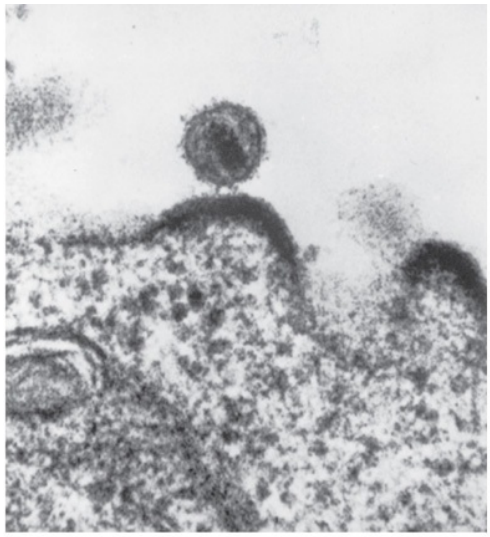

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

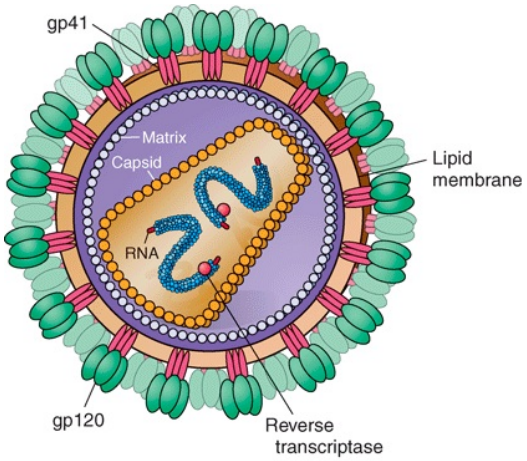

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

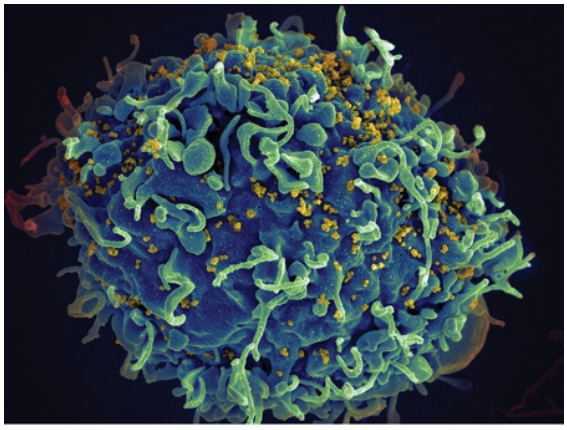

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A.* Electron micrograph of HIV.** Figure illustrates a typical virion following budding from the surface of a CD4+ T lymphocyte, together with two additional incomplete virions in the process of budding from the cell membrane. ***B. Structure of HIV-1, including the gp120 envelope, gp41 transmembrane components of the envelope, genomic RNA, enzymereverse transcriptase, p18(17) inner membrane (matrix), and p24 core protein (capsid). *(Copyright by George V. Kelvin). (Adapted from RC Gallo: Sci Am 256:46, 1987.)**C.*** Scanning electron micrograph of HIV-1 virions infecting a human CD4+ T lymphocyte. The original photograph was imaged at 8000x magnification. *(Courtesy of Elizabeth R. Fischer, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases; with permission.)*

Replication Cycle of HIV

HIV is an RNA virus whose hallmark is the reverse transcription of its genomic RNA to DNA by the enzyme *reverse transcriptase.* The replication cycle of HIV begins with the high-affinity binding of the gp120 protein via a portion of its V1 region near the N terminus to its receptor on the host cell surface, the CD4 molecule **(Fig. 189-3)**. The CD4 molecule is a 55-kDa protein found predominantly on a subset of T lymphocytes that are responsible for helper function in the immune system (Chap. 314). It is also expressed on the surface of monocytes/macrophages and dendritic/Langerhans cells. Once gp120 binds to CD4, the gp120 undergoes a conformational change that facilitates binding to one of two major co-receptors. The two major co-receptors for HIV-1 are CCR5 and CXCR4. Both receptors belong to the family of seven-transmembrane-domain G protein –coupled cellular receptors, and the use of one or the other or both receptors by the virus for entry into the cell is an important determinant of the cellular tropism of the virus. Certain dendritic cells express a diversity of C-type lectin receptors on their surface, one of which is called *DC-SIGN*, that also bind with high affinity to the HIV gp120 envelope protein, allowing the dendritic cell to facilitate the binding of virus to the CD4+ T cell upon engagement of dendritic cells with CD4+ T cells. Following binding of the envelope protein to the CD4 molecule associated with the above-mentioned conformational change in the viral envelope gp120, fusion with the host cell membrane occurs via the newly exposed gp41 molecule penetrating the plasma membrane of the target cell and then coiling upon itself to bring the virion and target cell together. Following fusion, the preintegration complex, composed of viral RNA and viral enzymes and surrounded by a capsid protein coat, is released into the cytoplasm of the target cell **(Fig. 189-4)**. As the preintegration complex traverses the cytoplasm to reach the nucleus (Fig. 189-3), the viral reverse transcriptase enzyme catalyzes the reverse transcription of the genomic RNA into DNA, and the protein coat opens to release the resulting double-strand proviral HIV-DNA. At this point in the replication cycle, the viral genome is vulnerable to cellular factors that can block the progression of infection. In particular, the cytoplasmic TRIM5-α protein in rhesus macaque cells blocks simian immunodeficiency virus (SIV) replication at a point shortly after the virus fuses with the host cell. Although the exact mechanisms of action of TRIM5-α remain unclear, the human form is inhibited by cyclophilin A and is not effective in restricting HIV replication in human cells. The recently described APOBEC family of cellular proteins also inhibits progression of virus infection after virus has entered the cell. APOBEC proteins bind to nascent reverse transcripts and deaminate viral cytidine, causing hypermutation of HIV genomes. It is still not clear whether viral replication is inhibited by: (1) the binding of APOBEC to the virus genome with subsequent accumulation of reverse transcripts, or (2) the hypermutations caused by the enzymatic deaminase activity of APOBEC proteins. HIV has evolved a powerful strategy to protect itself from APOBEC. The viral protein Vif targets APOBEC for proteasomal degradation.

Figure 189-3

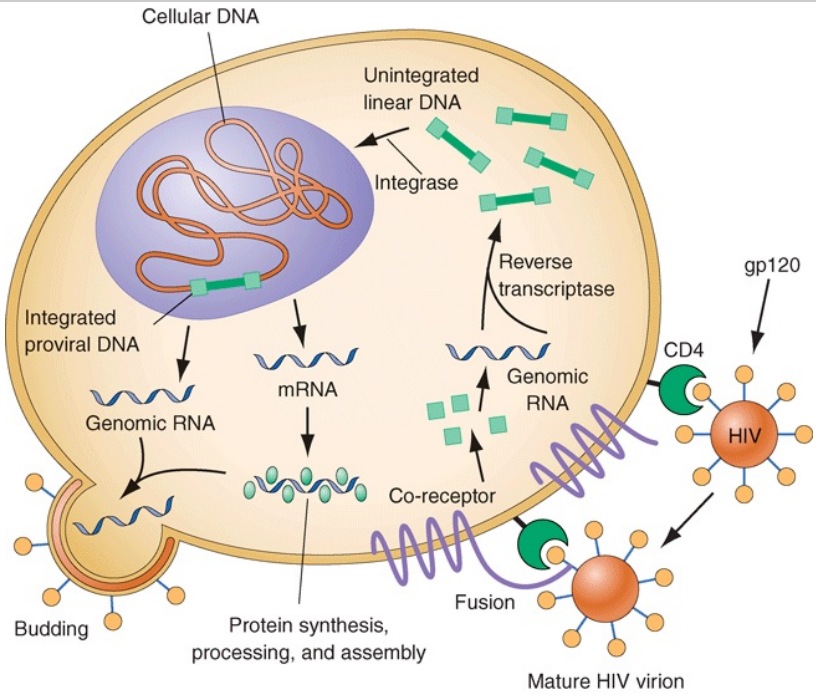

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The replication cycle of HIV. See text for description. *(Adapted from AS Fauci: Nature 384:529, 1996.)*

Figure 189-4

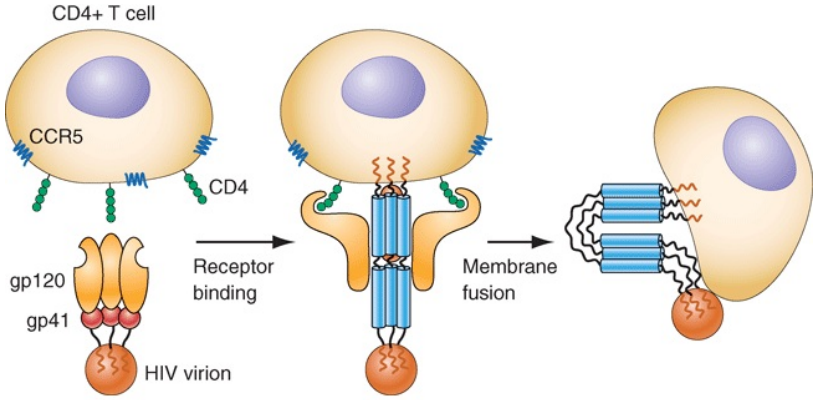

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Binding and fusion of HIV-1 with its target cell. HIV-1 binds to its target cell via the CD4 molecule, leading to a conformational change in the gp120 molecule that allows it to bind to the co-receptor CCR5 (for R5-using viruses). The virus then firmly attaches to the host cell membrane in a coiled-spring fashion via the newly exposed gp41 molecule. Virus-cell fusion occurs as the transitional intermediate of gp41 undergoes further changes to form a hairpin structure that draws the two membranes into close proximity (see text for details). *(Adapted from D Montefiori, JP Moore: Science 283:336, 1999; with permission.)*

With activation of the cell, the viral DNA accesses the nuclear pore and is exported from the cytoplasm to the nucleus, where it is integrated into the host cell chromosomes through the action of another virally encoded enzyme, *integrase*. HIV provirus (DNA) integrates into the nuclear DNA preferentially within introns of active genes and regional hotspots. This provirus may remain transcriptionally inactive (latent) or it may manifest varying levels of gene expression, up to active production of virus.

Cellular activation plays an important role in the replication cycle of HIV and is critical to the pathogenesis of HIV disease (see "pathogenesis and Pathophysiology," below). Following initial binding and internalization of virions into the target cell, incompletely reverse-transcribed DNA intermediates are labile in quiescent cells and do not integrate efficiently into the host cell genome unless cellular activation occurs shortly after infection. Furthermore, some degree of activation of the host cell is required for the initiation of transcription of the integrated proviral DNA into either genomic RNA or mRNA. This latter process may not necessarily be associated with the detectable expression of the classic cell-surface markers of activation. In this regard, activation of HIV expression from the latent state depends on the interaction of a number of cellular and viral factors. Following transcription, HIV mRNA is translated into proteins that undergo modification through glycosylation, myristoylation, phosphorylation, and cleavage. The viral particle is formed by the assembly of HIV proteins, enzymes, and genomic RNA at the plasma membrane of the cells. Budding of the progeny virion occurs through specialized regions in the lipid bilayer of the host cell membrane known as *lipid rafts*, where the core acquires its external envelope (Chap. 188). The virally encoded protease then catalyzes the cleavage of the gag-pol precursor to yield the mature virion. Progression through the virus replication cycle is profoundly influenced by a variety of viral regulatory gene products. Likewise, each point in the replication cycle of HIV is a real

or potential target for therapeutic intervention. Thus far, the reverse transcriptase, protease, and integrase enzymes as well as the process of virus–target cell binding and fusion have proved clinically to be susceptible to pharmacologic disruption.

HIV Genome

Figure 189-5 illustrates schematically the arrangement of the HIV genome. Like other retroviruses, HIV-1 has genes that encode the structural proteins of the virus: *gag* encodes the proteins that form the core of the virion (including p24 antigen); *pol* encodes the enzymes responsible for protease processing of viral proteins, reverse transcription, and integration; and *env* encodes the envelope glycoproteins. However, HIV-1 is more complex than other retroviruses, particularly those of the nonprimate group, in that it also contains at least six other genes (*tat*, *rev*, *nef*, *vif*, *vpr*, and *vpu*), which code for proteins involved in the modification of the host cell to enhance virus growth and the regulation of viral gene expression (Chap. 188). Several of these proteins are thought to play a role in the pathogenesis of HIV disease; their various functions are listed in Fig. 189-5. Flanking these genes are the long terminal repeats (LTRs), which contain regulatory elements involved in gene expression (Fig. 189-5). The major difference between the genomes of HIV-1 and HIV-2 is the fact that HIV-2 lacks the *vpu* gene and has a *vpx* gene not contained in HIV-1.

Figure 189-5

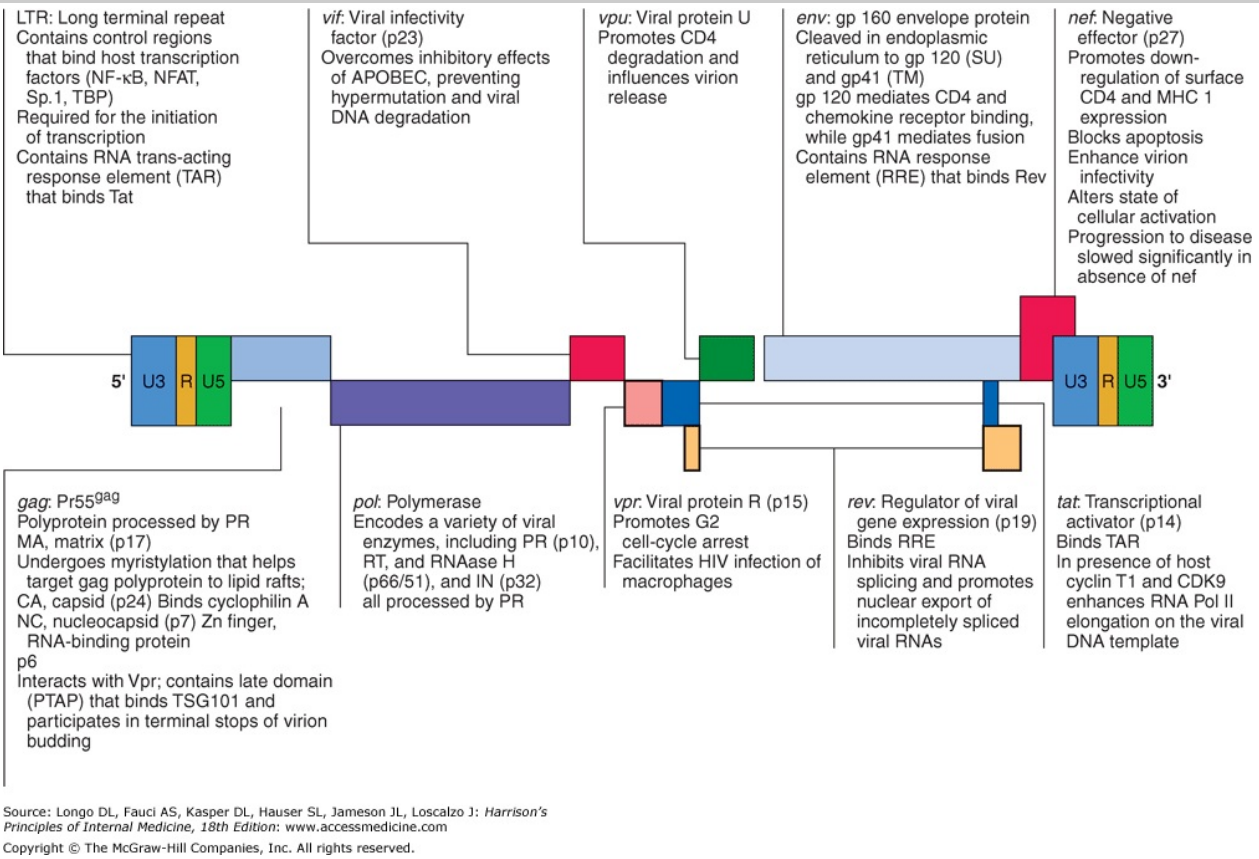

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Organization of the genome of the HIV provirus together with a summary description of its 9 genes encoding 15 proteins. *(Adapted from Greene and Peterlin.)*

Molecular Heterogeneity of HIV-1

Molecular analyses of HIV isolates reveal varying levels of sequence diversity over all regions of the viral genome. For example, the degree of difference in the coding sequences of the viral envelope protein ranges from a few percent (very close, among isolates from the same infected individual) to 50% (extreme diversity, among isolates from the different groups of HIV-1, M, N, O, and P). The changes tend to cluster in hypervariable regions. HIV can evolve by several means, including simple base substitution, insertions and deletions, recombination, and gain and loss of glycosylation sites. HIV sequence diversity arises directly from the limited fidelity of the reverse transcriptase. The balance of immune pressure and functional constraints on proteins influences the regional level of variation within proteins. For example, Envelope, which is exposed on the surface of the virion and is under immune selective pressure from both antibodies and cytolytic T lymphocytes, is extremely variable, with clusters of mutations in hypervariable domains. In contrast, reverse transcriptase, with important enzymatic functions, is relatively conserved, particularly around the active site. The extraordinary variability of HIV-1 is in marked contrast to the relative stability of HTLV-I and -II.

Four groups of HIV-1 have been defined. Group M (major) is responsible for most of the infections in the world. Group O (outlier) is a relatively rare viral form found originally in Cameroon, Gabon, and France. Group N was first identified in a Cameroonian woman with AIDS; very few group N isolates have been identified and sequenced. An additional human immunodeficiency virus, related to gorilla SIV and distinct from other HIV-1 groups, was identified in a Cameroonian woman in 2009 and proposed as group P.

Among primate lentiviruses, HIV-1 is most closely related to viruses isolated from chimpanzees and gorillas. The chimpanzee subspecies *Pan troglodytes troglodytes* has been established to be the natural reservoir of the HIV-1 M and N groups. The HIV-1 O group is most closely related to viruses found in Cameroonian gorillas. The M group comprises nine subtypes, or *clades*, designated A, B, C, D, F, G, H, J, and K, as well as a growing number of major and minor circulating recombinant forms (CRFs). CRFs are generated by infection of an individual with two subtypes that then recombine and create a virus with a selective advantage. These CRFs range from highly prevalent forms such as the AE virus, CRF01_AE, which is predominant in southeast Asia and often referred to simply as E, despite the fact that the parental E virus has never been found, and CRF02_AG from west and central Africa, to a large number of CRFs that are relatively rare. The subtypes and CRFs create the major lineages of the M group of HIV-1.

The picture was complicated somewhat when it was found that some subtypes are not equidistant from one another, while others contained sequences so diverse that they could not properly be considered to be the same subtype. Thus the term *sub-subtype* was introduced, and subtypes A and F are now subdivided into A1 and A2, and F1 and F2, respectively. It has also been argued that subtypes B and D are too close to be separate subtypes and should be considered sub-subtypes; it was decided, however, not to increase the confusion by renaming the clades **(Fig. 189-6)**.

Figure 189-6

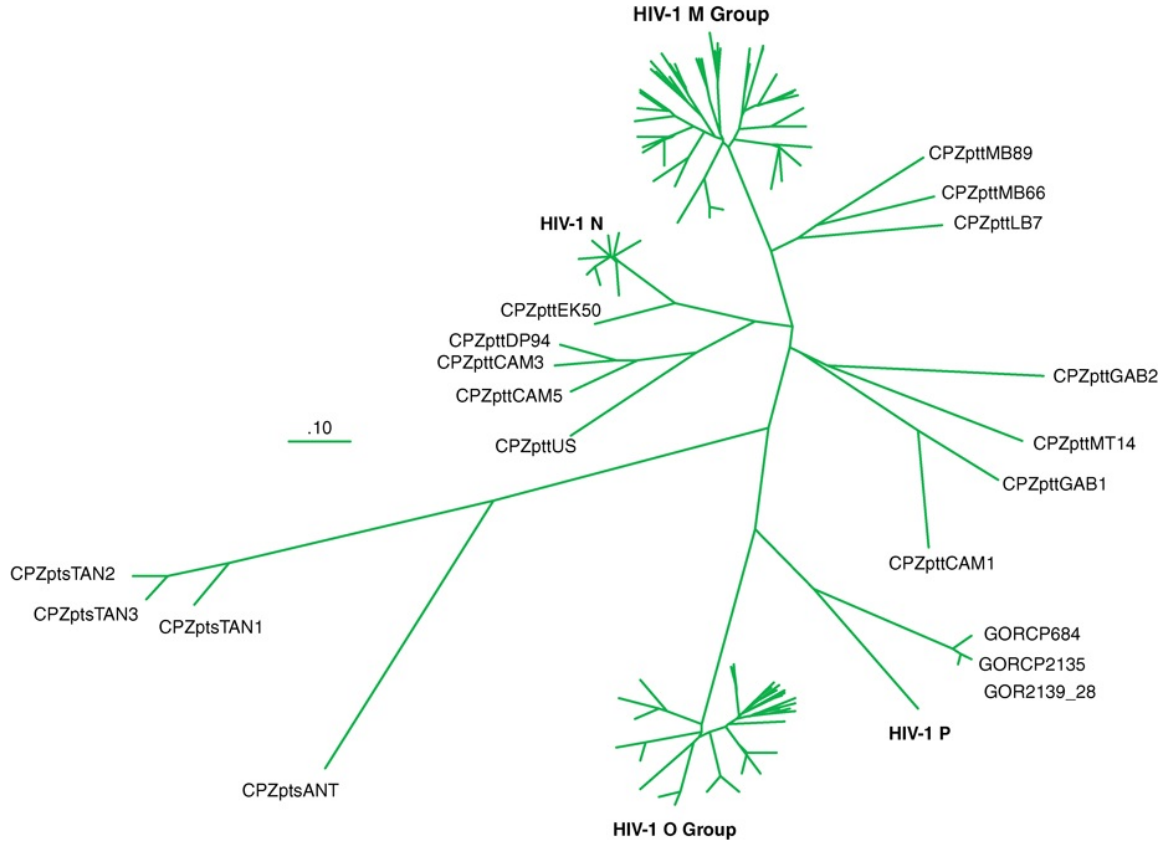

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Phylogenetic tree constructed from representative viral envelope sequences of the subtypes and CRF01 in HIV-1 group M, some isolates from groups N, O, and P (also human HIV-1), CPZ (chimpanzee), and gorilla (GOR). The scale bar indicates the genetic distances between the sequences. *(Prepared by Brian Foley, PhD, of the HIV Sequence Database, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.)*

Figure 189-7 schematically diagrams the worldwide distribution of HIV-1 subtypes by region. Seven strains have a global prevalence of >2.5% and account for the majority of HIV infections globally: HIV-1 subtypes A, B, C, D, G and two of the CRFs, CRF01_AE and CRF02_AG. Subtype C viruses (of the M group) are by far the most common form worldwide, accounting for ~ 50% of prevalent infections worldwide. In sub-Saharan Africa, home to approximately two-thirds of all individuals living with HIV/AIDS, the majority of infections are caused by subtype C, with smaller proportions of infections caused by subtype A, subtype G, CRF02_AG, and other subtypes and recombinants. In Asia, HIV-1 isolates of the CRF01_AE lineage and subtypes C and B predominate. CRF01_AE accounts for most infections in south and southeast Asia, while subtype C is prevalent in India (see "HIV Infection and AIDS Worldwide," below). Subtype B viruses are the overwhelmingly predominant viruses seen in the United States, Canada, certain countries in South America, western Europe, and Australia and account for 12–13% of global infections. It is thought that, purely by chance, subtype B was seeded into the United States in the late 1970s, thereby establishing an overwhelming founder effect. Many countries have co-circulating viral subtypes that are giving rise to new CRFs. Sequence analyses of HIV-1 isolates from infected individuals indicate that recombination among viruses of different clades likely occurs as a result of infection of an individual with viruses of more than one subtype, particularly in geographic areas where subtypes overlap.

Figure 189-7

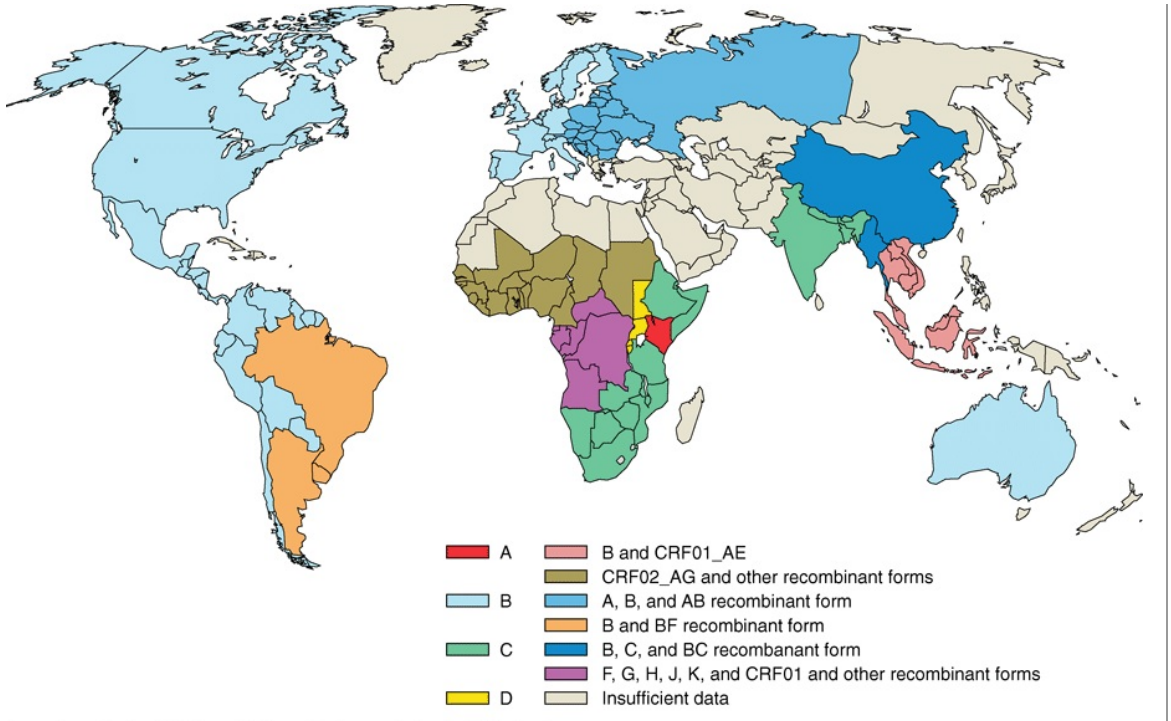

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Geographic distribution of HIV-1 subtypes and recombinant forms. *(Adapted from Taylor et al; with permission .)*

The extraordinary diversity of HIV, reflected by the presence of multiple subtypes, circulating recombinant forms, and continuous viral evolution, has implications for possible differential rates of disease progression, responses to therapy, and the development of resistance to antiretroviral drugs. This diversity is also a formidable obstacle to HIV vaccine development, as a broadly useful vaccine would need to induce protective responses against a wide range of viral strains.

TRANSMISSION

HIV is transmitted primarily by sexual contact (both heterosexual and male to male); by blood and blood products; and by infected mothers to infants intrapartum, perinatally, or via breast milk. After ~30 years of scrutiny, there is no evidence that HIV is transmitted by casual contact or that the virus can be spread by insects, such as by a mosquito bite.

Sexual Transmission

HIV infection is predominantly a sexually transmitted disease (STD) worldwide. By far the most common mode of infection, particularly in developing countries, is heterosexual transmission, although in many western countries a resurgence of male-to-male sexual transmission has occurred. Although a wide variety of factors including viral load and the presence of ulcerative genital diseases influence the efficiency of heterosexual transmission of HIV, such transmission is generally inefficient. In a pivotal study of heterosexual couples in the Rakai district of Uganda discordant for HIV infection (one partner was infected and the other was initially uninfected) the overall risk of HIV transmission rate was 0.12% per coital act in the absence of antiretroviral therapy. A meta-analysis of observational studies also found a low risk of heterosexual transmission in the absence of antiretrovirals: in high-income countries the estimated per-act rate was 0.04% for female-to-male transmission and 0.08% for male-to-female transmission. Rates for low-income-country female-to-male and male-to-female transmission were higher in this analysis (0.38% per act and 0.30% per act, respectively) in the absence of reported commercial sex exposure.

HIV has been demonstrated in seminal fluid both within infected mononuclear cells and in cell-free material. The virus appears to concentrate in the seminal fluid, particularly in situations where there are increased numbers of lymphocytes and monocytes in the fluid, as in genital inflammatory states such as urethritis and epididymitis, conditions closely associated with other STDs. The virus has also been demonstrated in cervical smears and vaginal fluid. There is an elevated risk of HIV transmission associated with unprotected receptive anal intercourse (URAI) among both men and women compared to the risk associated with receptive vaginal intercourse. Although data are limited, the per-act risk for HIV transmission via URAI was estimated to be ~ 1.4% for both men and women in a recent systemic review/meta-analysis. The risk of HIV acquisition associated with URAI is probably higher than that seen in penile-vaginal intercourse because only a thin, fragile rectal mucosal membrane separates the deposited semen from potentially susceptible cells in and beneath the mucosa, and trauma may be associated with anal intercourse. Anal douching and sexual practices that traumatize the rectal mucosa also increase the likelihood of infection. It is likely that anal intercourse provides at least two modalities of infection: (1) direct inoculation into blood in cases of traumatic tears in the mucosa; and (2) infection of susceptible target cells, such as Langerhans cells, in the mucosal layer in the absence of trauma. Insertive anal intercourse also confers an increased risk of HIV acquisition compared to insertive vaginal intercourse. Although the vaginal mucosa is several layers thicker than the rectal mucosa and less likely to be traumatized during intercourse, the virus can be transmitted to either partner through vaginal intercourse. Studies in the United States and Europe have found that male-to-female HIV transmission is usually more efficient than female-to-male transmission. The differences in reported transmission rates between men and women may be due in part to the prolonged exposure to infected seminal fluid of the vaginal and cervical mucosa, as well as the endometrium (when semen enters through the cervical os). By comparison, the penis and urethral orifice are exposed relatively briefly to infected vaginal fluid. Among various cofactors examined in studies of heterosexual HIV transmission, the presence of other STDs has been strongly associated with HIV transmission. In this regard, there is a close association between genital ulcerations and transmission, owing to both susceptibility to infection and infectivity. Infections with microorganisms such as *Treponema pallidum* (Chap. 169), *Haemophilus ducreyi* (Chap. 145), and herpes simplex virus (HSV; Chap. 179) are important causes of genital ulcerations linked to transmission of HIV. In addition, pathogens responsible for nonulcerative inflammatory STDs such as those caused by *Chlamydia trachomatis* (Chap. 176), *Neisseria gonorrhoeae* (Chap. 144), and *Trichomonas vaginalis* (Chap. 215) are also associated with an increased risk of transmission of HIV infection. Bacterial vaginosis, an infection related to sexual behavior, but not strictly an STD, may also be linked to an increased risk of transmission of HIV infection. Several studies suggest that treating other STDs and genital tract syndromes

may help prevent transmission of HIV. This effect is most prominent in populations in which the prevalence of HIV infection is relatively low.

The quantity of HIV-1 in plasma is a primary determinant of the risk of HIV-1 transmission. In a cohort of heterosexual couples in Uganda discordant for HIV infection and not receiving antiretroviral therapy, the mean serum HIV RNA level was significantly higher among HIV-infected subjects whose partners seroconverted than among those whose partners did not seroconvert. In fact, transmission was rare when the infected partner had a plasma level of <1700 copies of HIV RNA per milliliter, even when genital ulcer disease was present **(Fig 189-8)**. The rate of HIV transmission per coital act was highest during the early stage of HIV infection when plasma HIV RNA levels were high and in advanced disease as the viral set point increased.

Figure 189-8

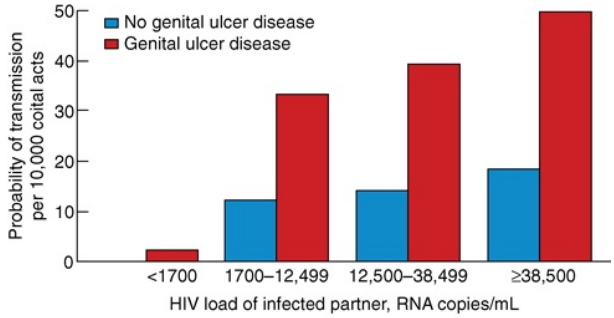

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Probability of HIV transmission per coital act among monogamous, heterosexual, HIV-serodiscordant couples in Uganda. *(From RH Gray et al.)*

Antiretroviral therapy dramatically reduces plasma viremia in most HIV-infected individuals (see "Treatment," below) and is associated with a reduction in risk of transmission. For example, in an analysis of ~3400 HIV serodiscordant heterosexual couples from 7 African countries, use of antiretroviral therapy by the infected person was accompanied by a 92% reduction in risk of HIV-1 transmission to the uninfected partner. Several studies also have suggested a beneficial effect of antiretroviral treatment at the community level.

A number of studies including large, randomized, controlled trials clearly have indicated that male *circumcision* is associated with a lower risk of HIV infection for heterosexual men. Studies are conflicting as to whether circumcision protects against HIV acquisition among men who have sex with men. The benefit of circumcision may be due to increased susceptibility of uncircumcised men to ulcerative STDs, as well as to other factors such as microtrauma to the foreskin and glans penis. In addition, the highly vascularized inner foreskin tissue contains a high density of Langerhans cells as well as increased numbers of CD4+ T cells, macrophages, and other cellular targets for HIV. Finally, the moist environment under the foreskin may promote the presence or persistence of microbial flora that, via inflammatory changes, may lead to even higher concentrations of target cells for HIV in the foreskin. In some studies the use of oral contraceptives was associated with an increase in incidence of HIV infection over and above that which might be expected by not using a condom for birth control. This phenomenon may be due to drug-induced changes in the cervical mucosa, rendering it more vulnerable to penetration by the virus. Adolescent girls might also be more susceptible to infection upon exposure due to the properties of an immature genital tract with increased cervical ectopy or exposed columnar epithelium.

Oral sex is a much less efficient mode of transmission of HIV than is anal intercourse or vaginal intercourse. A number of studies have reported that the incidence of transmission of infection by oral sex among couples discordant for HIV was extremely low. However, there have been reports of documented HIV transmission resulting solely from receptive fellatio and insertive cunnilingus. Therefore, the assumption that oral sex is completely safe is not warranted.

The association of alcohol consumption and illicit drug use with unsafe sexual behavior, both homosexual and heterosexual, leads to an increased risk of sexual transmission of HIV. Methamphetamine and other so-called club drugs (e.g., ecstasy, ketamine, and gamma hydroxybutyrate), sometimes taken in conjunction with PDE-5 inhibitors such as sildenafil (Viagra), tadalafil (Cialis), or vardenafil (Levitra), have been associated with risky sexual practices and increased risk of HIV infection, particularly among men who have sex with men.

Transmission by Blood and Blood Products

HIV can be transmitted to individuals who receive HIV-tainted blood transfusions, blood products, or transplanted tissue as well as to IDUs who are exposed to HIV while sharing injection paraphernalia such as needles, syringes, the water in which drugs are mixed, or the cotton through which drugs are filtered. Parenteral transmission of HIV during injection drug use does not require IV puncture; SC ("skin popping") or IM ("muscling") injections can transmit HIV as well, even though these behaviors are sometimes erroneously perceived as low-risk. Among IDUs, the risk of HIV infection increases with the duration of injection drug use; the frequency of needle sharing; the number of partners with whom paraphernalia are shared, particularly in the setting of "shooting galleries" where drugs are sold and large numbers of IDUs may share a limited number of "works"; comorbid psychiatric conditions such as antisocial personality disorder; the use of cocaine in injectable form or smoked as "crack"; and the use of injection drugs in a geographic location with a high prevalence of HIV infection, such as certain inner-city areas in the United States.

The first cases of AIDS among transfusion recipients and individuals with hemophilia or other clotting disorders were reported in 1982. The vast majority of HIV infections acquired via contaminated blood transfusions, blood components, or transplanted tissue in resource-rich countries occurred prior to the spring of 1985, when mandatory testing of donated blood for HIV-1 was initiated. It is estimated that >90% of individuals exposed to HIV-contaminated blood products become infected; unfortunately, in resource-poor countries, HIV continues to be transmitted by blood, blood products, and tissues due to the large number of blood donations that are inadequately screened for HIV. Transfusions of whole blood, packed red blood cells, platelets, leukocytes, and plasma are all capable of transmitting HIV infection. In contrast, hyperimmune gamma globulin, hepatitis B immune globulin, plasma-derived hepatitis B vaccine, and Rh_0 immune globulin have not been associated with transmission of HIV infection. The procedures involved in processing these products either inactivate or remove the virus.

Currently, in the United States and in most developed countries, the following measures have made the risk of transmission of HIV infection by transfused blood or blood products extremely small: the screening of blood donations for HIV antibodies, HIV p24 antigen, and HIV nucleic acid; the careful selection of potential blood donors with health history questionnaires to exclude

individuals with risk behavior; and opportunities for self-deferral and the screening out of HIV-negative individuals with serologic testing for infections that have shared risk factors with HIV, such as hepatitis B and C. The chance of infection of a hemophiliac via clotting factor concentrates has essentially been eliminated because of the added layer of safety resulting from heat treatment of the concentrates.

It is currently estimated that the risk of infection with HIV in the United States via transfused screened blood is approximately 1 in 1.5 million units. Therefore, among the ~16 million donations collected in the United States each year, there are about 11 infectious donations leading to approximately 20 HIV-positive blood components being released each year that could potentially infect recipients. Thus, despite the best efforts of science, one cannot completely eliminate the risk of transfusion-related transmission of HIV since current technology cannot detect HIV RNA for the first 10–15 days following infection due to the low levels of viremia. In this regard, 4 cases of transfusion-associated HIV infection attributable to infected blood that tested negative for HIV were reported in the United States in the period 2000–2008.

In other countries, there have been reports of sporadic breakdowns in routinely available screening procedures in which contaminated blood was allowed to be transfused, resulting in small clusters of patients becoming infected. For example, in China in the 1990s, a disturbingly large number of persons became infected by selling blood in situations where the collectors reused needles that were contaminated and, in some instances, mixed blood products from a number of individuals, separated the plasma, and reinfused mixed red blood cells back into the individual donors.

There have been no reported cases of transmission of HIV-2 in the United States via donated blood or tissues, and, currently, donated blood is screened for both HIV-1 and HIV-2. Transmission of HIV (both HIV-1 and HIV-2) by blood or blood products is still an ongoing threat in certain developing countries, particularly in sub-Saharan Africa, where routine screening of blood is not universally practiced.

Prior to the screening of donors, a small number of cases of transmission of HIV via semen used in artificial insemination and tissues used in organ transplantation were documented. At present, donors of such tissues are prescreened for HIV infection. With regard to HIV-serodiscordant couples (male, HIV-infected; female, HIV-uninfected) who wish to conceive a child, assisted reproductive techniques using sperm-washing to reduce the risk of HIV transmission have been successfully employed, with only one well-documented seroconversion in the uninfected female partner, reported in 1990.

Occupational Transmission of HIV: Health Care Workers, Laboratory Workers, and the Health Care Setting

There is a small but definite occupational risk of HIV transmission to health care workers and laboratory personnel and potentially others who work with HIV-containing materials, particularly when sharp objects are used. An estimated 600,000 to 800,000 health care workers are stuck with needles or other sharp medical instruments in the United States each year.

Exposures that place a health care worker at potential risk of HIV infection are percutaneous injuries (e.g., a needle stick or cut with a sharp object) or contact of mucous membrane or nonintact skin (e.g., exposed skin that is chapped, abraded, or afflicted with dermatitis) with blood, tissue, or other potentially infectious body fluids. Large, multi-institutional studies have indicated that the risk of HIV transmission following skin puncture from a needle or a sharp object that was contaminated with blood from a person with documented HIV infection is ~0.3% and after a mucous membrane exposure it is 0.09% (see "HIV and the Health Care Worker," below) if the injured and/or exposed person is not treated within 24 h with antiretroviral drugs. HIV transmission after nonintact skin exposure has been documented, but the average risk for transmission by this route has not been precisely determined; however, it is estimated to be less than the risk for mucous membrane exposure. Transmission of HIV through intact skin has not been documented. Currently, virtually all puncture wounds and mucous membrane exposures in health care workers involving blood from a patient with documented HIV infection are treated prophylactically with combination antiretroviral therapy (cART). This practice has dramatically reduced the occurrence of puncture-related transmissions of HIV to health care workers.

In addition to blood and visibly bloody body fluids, semen and vaginal secretions also are considered potentially infectious; however, they have not been implicated in occupational transmission from patients to health care workers. The following fluids also are considered potentially infectious: cerebrospinal fluid, synovial fluid, pleural fluid, peritoneal fluid, pericardial fluid, and amniotic fluid. The risk for transmission after exposure to fluids or tissues other than HIV-infected blood also has not been quantified, but it is probably considerably lower than the risk after blood exposures. Feces, nasal secretions, saliva, sputum, sweat, tears, urine, and vomitus are not considered potentially infectious for HIV unless they are visibly bloody. Rare cases of HIV transmission via human bites have been reported, but not in the setting of occupational exposure.

An increased risk for HIV infection following percutaneous exposures to HIV-infected blood is associated with exposures involving a relatively large quantity of blood, as in the case of a device visibly contaminated with the patient's blood, a procedure that involves a hollow-bore needle placed directly in a vein or artery, or a deep injury. Factors that might be associated with mucocutaneous transmission of HIV include exposure to an unusually large volume of blood, prolonged contact, and a potential portal of entry. In addition, the risk increases for exposures to blood from patients with advanced-stage disease or those patients in the acute stage of HIV infection, owing to the higher levels of HIV in the blood under those circumstances. The use of antiretroviral drugs as postexposure prophylaxis decreases the risk of infection compared with historic controls in occupationally exposed health care workers (see "HIV and the Health Care Worker," below). The risk of hepatitis B virus (HBV) infection following a similar type of exposure is ~6–30% in nonimmune individuals; if a susceptible worker is exposed to HBV, postexposure prophylaxis with hepatitis B immune globulin and initiation of HBV vaccine is >90% effective in preventing HBV infection. The risk of hepatitis C virus (HCV) infection following percutaneous injury is ~1.8% (Chap. 304).

Since the beginning of the HIV epidemic, there have been rare instances where transmission of infection from a health care worker to patients seemed highly probable. One notable cluster of infections involved an HIV-infected dentist in Florida who apparently infected as many as six of his patients, most likely through contaminated instruments. Despite these small number of documented cases, the risk of HIV transmission involving health care workers (infected or not) to patients is extremely low in developed countries—in fact, too low to be measured accurately. In this regard, several epidemiologic studies have been performed tracing thousands of patients of HIV-infected dentists, physicians, surgeons, obstetricians, and gynecologists, and no other cases of HIV transmission that could be linked to the health care providers were identified.

Breaches in infection control and the reuse of contaminated syringes have also resulted in the transmission of HIV from patient to patient in hospitals, nursing homes, and outpatient settings. For example, in the only report of HIV transmission from patient to patient during a surgical procedure, several patients in Australia apparently were infected by an HIV-negative general surgeon during routine outpatient surgery. Although the mechanism of transmission was not definitively identified, a failure on the part of the surgeon to sterilize instruments properly following prior surgery on an HIV-infected patient was considered a likely explanation for this outbreak. Three patients (two in hospitals in the United States and one in the Netherlands) undergoing nuclear medicine procedures were reported to have inadvertently received IV injections of blood or other material from patients infected with HIV. Hemodialysis centers have also been implicated in several reported HIV transmission incidents.

The most dramatic reports of HIV infection in the health care setting involved transmission of HIV to 8000–10,000 children in Romanian orphanages in the 1980s. Other large incidents occurred in hospitals in Russia and Libya in the late 1980s and late 1990s, respectively. Each of these incidents received considerable attention and likely was related to reuse of contaminated needles and/or administration of contaminated blood products. Finally, these very rare occurrences of transmission of HIV as

well as HBV and HCV to and from health care workers in the workplace underscore the importance of the use of universal precautions when caring for all patients (see below and Chap. 131).

Maternal-Fetal/Infant Transmission

HIV infection can be transmitted from an infected mother to her fetus during pregnancy, during delivery, or by breast-feeding. This remains an important form of transmission of HIV infection in certain developing countries, where the proportion of infected women to infected men is ~1:1. Virologic analyses of aborted fetuses indicate that HIV can be transmitted to the fetus during the first or second trimesters of pregnancy. However, maternal transmission to the fetus occurs most commonly in the perinatal period. Two studies performed in Rwanda and the former Zaire indicated that the relative proportions of mother-to-child transmissions were 23–30% before birth, 50–65% during birth, and 12–20% via breast-feeding.

In the absence of prophylactic antiretroviral therapy to the mother during pregnancy, labor, and delivery, and to the fetus following birth, the probability of transmission of HIV from mother to infant/fetus ranges from 15 to 25% in industrialized countries and from 25 to 35% in developing countries. These differences may relate to the adequacy of prenatal care as well as to the stage of HIV disease and the general health of the mother during pregnancy. Higher rates of transmission have been reported to be associated with many factors, the best documented of which is the presence of high maternal levels of plasma viremia. In one study of 552 singleton pregnancies in the United States, the rate of mother-to-baby transmission was 0% among women with <1000 copies of HIV RNA per milliliter of blood, 16.6% among women with 1000–10,000 copies/mL, 21.3% among women with 10,001–50,000 copies/mL, 30.9% among women with 50,001–100,000 copies/mL, and 40.6% among women with >100,000 copies/mL. However, there may not be a lower "threshold" below which transmission never occurs, since other studies have reported transmission by women with viral RNA levels <50 copies/mL. Low maternal CD4+ T cell counts also have been associated with higher rates of transmission; however, since low CD4+ T cell counts are often associated with high levels of plasma viremia, in one study using multivariate analysis including plasma viral load and CD4+ T cell count, only the level of plasma HIV RNA was significant. Increased mother-to-child transmission is also correlated with closer human leukocyte antigen (HLA) match between mother and child. A prolonged interval between membrane rupture and delivery is another well-documented risk factor for transmission. Other conditions that are potential risk factors, but that have not been consistently demonstrated, are the presence of chorioamnionitis at delivery; STDs during pregnancy; hard drug use during pregnancy; cigarette smoking; preterm delivery; and obstetric procedures such as amniocentesis, amnioscopy, fetal scalp electrodes, and episiotomy. In a seminal study conducted in the United States and France in the 1990s, zidovudine treatment of HIV-infected pregnant women from the beginning of the second trimester through delivery and of the infant for 6 weeks following birth dramatically decreased the rate of intrapartum and perinatal transmission of HIV infection from 22.6% in the untreated group to <5%. Today, the rate of mother-to-child transmission has fallen to 1% or less in pregnant women who are receiving combination antiretroviral therapy for their HIV infection. Such treatment, combined with cesarean section delivery, has rendered mother-to-child transmission of HIV an unusual event in the United States and other developed nations. In developed countries, current recommendations to reduce perinatal transmission of HIV include universal voluntary HIV testing and counseling of pregnant women, antiretroviral prophylaxis with one or more drugs in cases in which the mother does not require therapy for her HIV infection, combination therapy for women who do require therapy, obstetric management that attempts to minimize exposure of the infant to maternal blood and genital secretions, and avoidance of breast-feeding. It is recommended that the choice of antiretroviral therapy for pregnant women be based on the same considerations used for women who are not pregnant, with discussion of the recognized and unknown risks and benefits of such therapy during pregnancy (see below under "Treatment").

Certain studies have demonstrated that truncated regimens of zidovudine alone or in combination with lamivudine given to the mother during the last few weeks of pregnancy or even only during labor and delivery, and to the infant for a week or less, significantly reduced transmission to the infant compared with placebo. Short-course prophylactic antiretroviral regimens, such as a single dose of nevirapine given to the mother at the onset of labor and a single dose to the infant within 72 h of birth, are of particular relevance to low- to mid-income nations because of the low cost and the fact that in these regions perinatal care is often not available and pregnant women are often seen by a health care provider for the first time at or near the time of delivery. Given that cART is now increasingly available to individuals in developing countries due to the lower cost of drugs and programs that are making drugs available to these regions of the world, combinations of drugs are being used more frequently, where available, to treat HIV-infected pregnant women. This has had the effect of benefitting the women, blocking HIV transmission to the fetus, and protecting against subsequent transmission by breast-feeding.

Breast-feeding is an important modality of transmission of HIV infection in developing countries, particularly where mothers continue to breast-feed for prolonged periods. The risk factors for mother-to-child transmission of HIV via breast-feeding are not fully understood; factors that increase the likelihood of transmission include detectable levels of HIV in breast milk, the presence of mastitis, low maternal CD4+ T cell counts, and maternal vitamin A deficiency. The risk of HIV infection via breast-feeding is highest in the early months of breast-feeding. In addition, exclusive breast-feeding has been reported to carry a lower risk of HIV transmission than mixed feeding. Certainly in developed countries, breast-feeding by an infected mother should be avoided. However, there is disagreement regarding recommendations for breast-feeding in certain developing countries, where breast milk is the only source of adequate nutrition as well as immunity against potentially serious non-HIV infections for the infant. The optimal approach to prevent transmission by infected mothers who choose to breast-feed would be to provide continual treatment to the infected mother. This approach has become more feasible as cART becomes more widely available in developing countries. Despite progress in this regard, such therapy is currently available to only ~30–40% of persons in developing nations who require it.

Transmission by Other Body Fluids

Although HIV can be isolated typically in low titers from saliva of a small proportion of infected individuals, there is no convincing evidence that saliva can transmit HIV infection, either through kissing or through other exposures, such as occupationally to health care workers. Saliva contains endogenous antiviral factors; among these factors, HIV-specific immunoglobulins of IgA, IgG, and IgM isotypes are detected readily in salivary secretions of infected individuals. It has been suggested that large glycoproteins such as mucins and thrombospondin 1 sequester HIV into aggregates for clearance by the host. In addition, a number of soluble salivary factors inhibit HIV to various degrees in vitro, probably by targeting host cell receptors rather than the virus itself. Perhaps the best studied of these, secretory leukocyte protease inhibitor (SLPI), blocks HIV infection in several cell culture systems, and it is found in saliva at levels that approximate those required for inhibition of HIV in vitro. In this regard, higher salivary levels of SLPI in breast-fed infants were associated with a decreased risk of HIV transmission through breast milk. It has also been suggested that submandibular saliva reduces HIV infectivity by stripping gp120 from the surface of virions, and that saliva-mediated disruption and lysis of HIV-infected cells occurs because of the hypotonicity of oral secretions. There have been outlier cases of suspected transmission by saliva, but these have probably been blood-to-blood transmissions. Transmission of HIV by a human bite can occur but is a rare event. In addition, a most unusual form of HIV transmission from infected children to mothers in the former Soviet Union has been identified. In those cases, the children (infected through transfusion) were said to have bleeding sores in the mouth, and the mothers were said to have lacerations and abrasions on and around the nipples of the breast resulting from trauma from the children's teeth. Breast-feeding had been continued until the children were older than is usual in other developed countries.

Although virus can be identified, if not isolated, from virtually any body fluid, there is no evidence that HIV transmission can occur as a result of exposure to tears, sweat, or urine. However, there have been isolated cases of transmission of HIV infection

by body fluids that may or may not have been contaminated with blood. Most of these situations occurred in the setting of a close relative providing intensive nursing care for an HIV-infected person without observing universal precautions, underscoring the importance of adhering to such precautions in the handling of body fluids and wastes from HIV-infected individuals.

EPIDEMIOLOGY

HIV Infection and AIDS Worldwide

HIV infection/AIDS is a global pandemic, with cases reported from virtually every country. At the end of 2009, an estimated 33.3 million individuals were living with HIV infection according to the Joint United Nations Programme on HIV/AIDS (UNAIDS). More than 95% of people living with HIV/AIDS reside in low- and middle-income countries; ~50% are female, and 2.5 million are children <15 years. The global distribution of these cases is illustrated in **Fig. 189-9**. As illustrated in Fig. 189-10*A*, the estimated number of people living with HIV—i.e., the global prevalence—has increased approximately fourfold since 1990, reflecting the combined effects of continued high rates of new HIV infections and the beneficial (life-prolonging) impact of antiretroviral therapy.

Figure 189-9

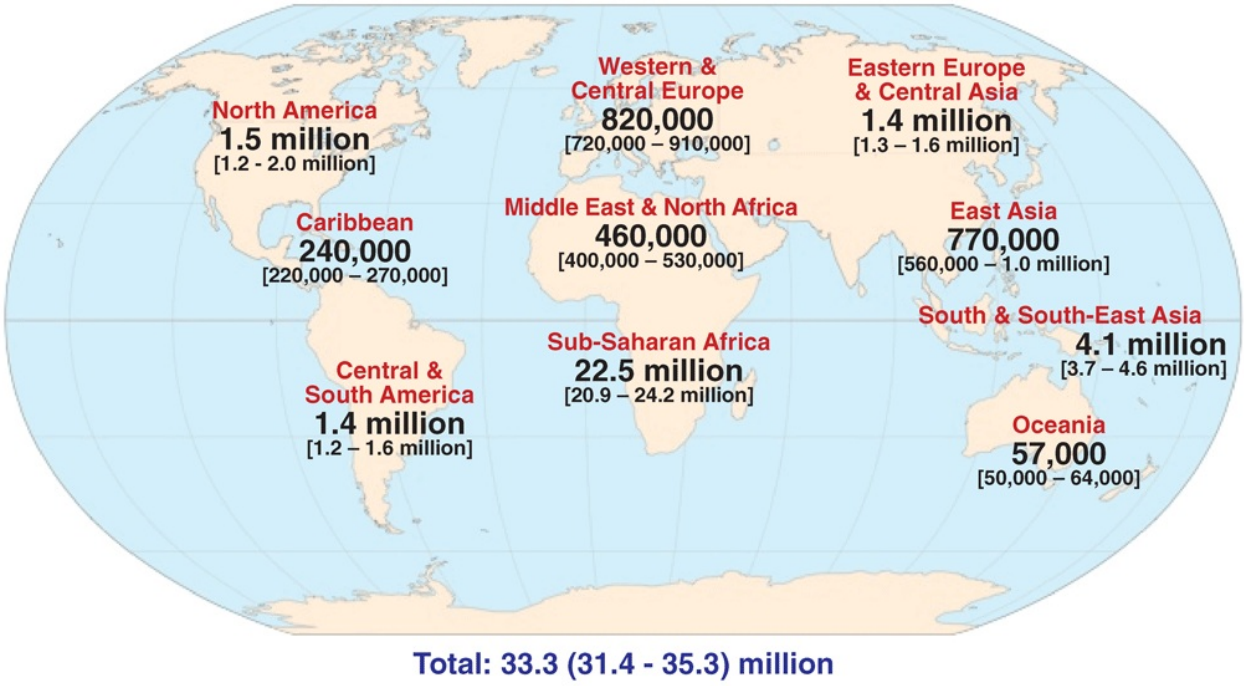

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated number of adults and children living with HIV infection as of December, 2009. Total: 33.3 (31.4–35.3) (31.1–35.8) million. *[From Joint United Nations Programme on HIV/AIDS (UNAIDS).]*

Figure 189-10

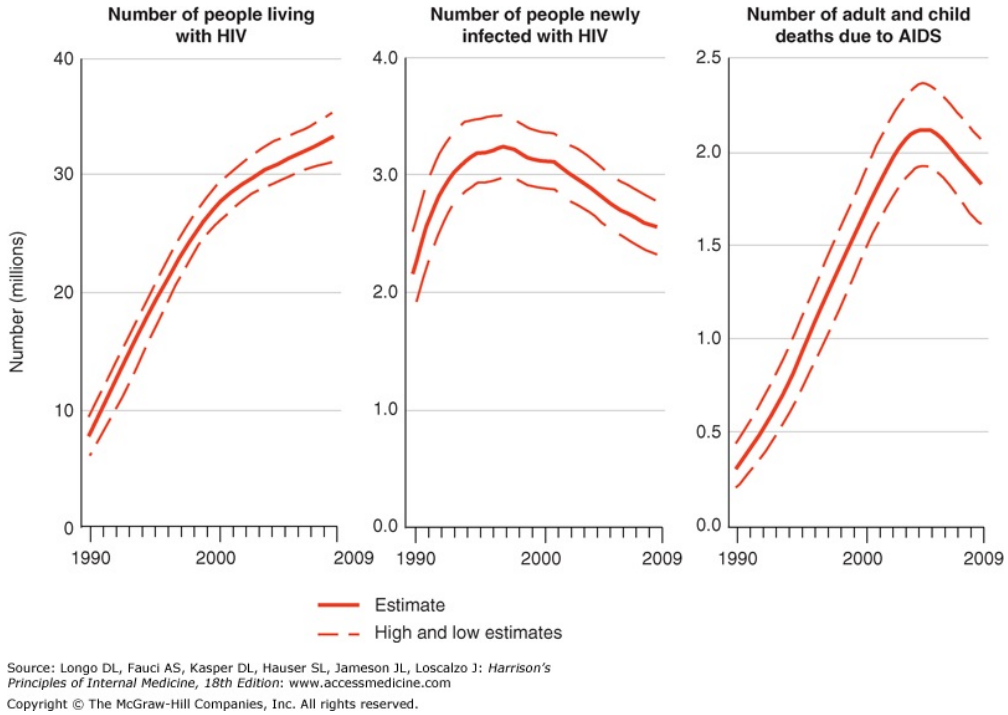

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Global HIV/AIDS epidemiologic estimates, 1990–2009. *A.* Number of people living with HIV. ***B.*** Number of people newly infected with HIV globally. ***C.*** Number of adult and child deaths due to AIDS. *(From UNAIDS.)*

In 2009, there were an estimated 2.6 million new cases of HIV infection worldwide, including 370,000 in children <15 years. UNAIDS estimates that the global spread of HIV peaked in 1997, when ~3.2 million new HIV infections occurred. In 2009, the estimated number of new HIV infections globally was approximately 21% lower than at the peak of the pandemic **(Fig. 189-10*B*)**. Recent reductions in global HIV incidence likely reflect natural trends in the pandemic as well as the results of prevention programs resulting in behavior change. In 2009, global AIDS deaths totaled 1.8 million (including 260,000 children <15 years). A rapid expansion of access to antiretroviral therapy likely has helped lower AIDS-related death rates in recent years **(Fig. 189-10*C*).** Since the beginning of the pandemic the cumulative total of AIDS deaths globally exceeds 25 million.

The HIV epidemic has occurred in "waves" in different regions of the world, each wave having somewhat different characteristics depending on the demographics of the country and region in question and the timing of the introduction of HIV into the population. Although the AIDS epidemic was first recognized in the United States and shortly thereafter in Western Europe, it very likely began in sub-Saharan Africa (see above), which has been particularly devastated by the epidemic. More than two-thirds of all people with HIV infection (~22.5 million) live in that region, even though sub-Saharan Africa is home to just 10–11% of the world's population (Fig. 189-9). Within the region, southern Africa is worst-affected. In each of the nine southern African countries, available seroprevalence data indicate that >10% of the adult population age 15–49 is HIV-infected. In addition, among high-risk individuals (e.g., commercial sex workers, patients attending STD clinics) who live in urban areas of sub-Saharan Africa, seroprevalence is now >50% in some countries. Sub-Saharan Africa's HIV regional epidemics vary significantly, with most appearing to have stabilized, although frequently at very high levels. Heterosexual exposure is the primary mode of HIV transmission in sub-Saharan Africa, with women and girls disproportionately affected, accounting for ~60 percent of all HIV infections in that region. In 2009, an estimated 460,000 people were living with HIV in the Middle East/North Africa region. Cases are largely concentrated among IDUs, men who have sex with men, and sex workers and their clients.

In east, south, and southeast Asia, an estimated 4.9 million people were living with HIV at the end of 2009. In this region of the world, national HIV prevalence is highest in southeast Asian countries, with wide variation in trends between different countries. Among countries in Asia, only Thailand has an adult seroprevalence rate of >1%. However, the populations of many Asian nations are so large (especially India and China) that even low infection and seroprevalence rates result in large numbers of people living with HIV. Although Asia's epidemic has been concentrated for some time among specific populations—sex workers and their clients, men who have sex with men, and IDUs—it is expanding to the heterosexual partners of those most at risk. While the regional epidemic appears to be stable overall, HIV prevalence has increased in certain countries such as Bangladesh and Pakistan.

The epidemic is expanding in Eastern Europe and Central Asia, where ~1.4 million people were living with HIV at the end of 2009. The Russian Federation and Ukraine account for the majority of HIV cases in the region; the Ukraine has an adult seroprevalence rate of 1.1%, the highest in all of Europe. Driven initially by injection drug use and increasingly by heterosexual transmission, the number of new infections in this region has increased dramatically over the past decade.

Approximately 1.6 million people are living with HIV/AIDS in Central and south America and the Caribbean. Brazil is home to the largest number of HIV-infected people in the region. However, the epidemic has been slowed in that country due to successful treatment and prevention efforts. Men who have sex with men account for the largest proportion of HIV infections in Central and south America. The Caribbean region has the highest regional adult seroprevalence rate after Africa, due in large part to the huge case load in Haiti. Heterosexual transmission, often tied to sex work, is the main driver of transmission in the region.

Approximately 2.4 million people are living with HIV/AIDS in North America, western and central Europe, and Oceania. The number of new infections among men who have sex with men has increased over the past decade in these mostly high-income areas, while rates of new infections among heterosexuals have stabilized and infections among IDUs have fallen.

AIDS in the United States

HIV/AIDS continues to have an extraordinary public health impact in the United States. As of January 1, 2010, an estimated 1,108,611 cases of AIDS had been diagnosed in the United States. Approximately 1.1 million individuals in the United States were living with HIV infection, ~21 percent of whom are unaware of their infection, according to recent analysis. Approximately two-thirds of those living with HIV/AIDS were nonwhite and nearly half (48%) were men who have sex with men. The estimated HIV seroprevalence rate among individuals age 13 years or older in the United States is ~0.5%.

The number of AIDS cases and deaths in the United States rose steadily through the 1980s; AIDS cases peaked in 1993 and

deaths in 1995 **(Fig. 189-11)**. Since then, the annual numbers of AIDS-related deaths in the United States have fallen ~70%. This trend is due to several factors, including the improved prophylaxis and treatment of opportunistic infections, the growing experience among the health professions in caring for HIV-infected individuals, improved access to health care, and a decrease in new infections due to saturational effects and prevention efforts. However, the most influential factor clearly has been the increased use of potent antiretroviral drugs, generally administered in a combination of three or four agents.

Figure 189-11

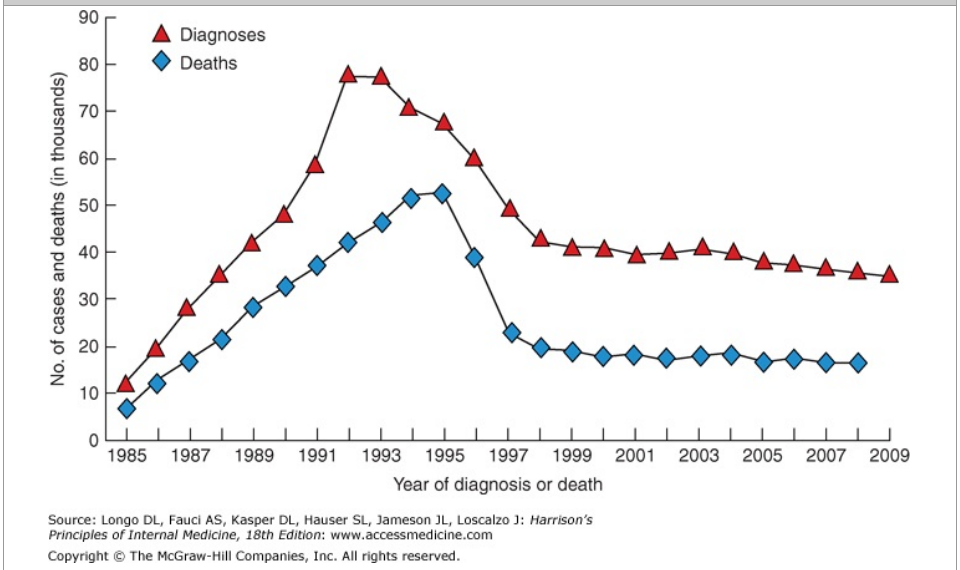

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated number of AIDS cases and AIDS deaths, United States, 1985–2009. *(From CDC.)*

An estimated 56,000 individuals are newly infected each year in the United States. This *HIV incidence* figure has remained stable for at least 15 years **(Fig. 189-12).** Among adults and adolescents newly diagnosed with HIV infection (regardless of AIDS status) in 2009, ~76% were men and ~24% were women **(Fig. 189-13)**. Of new HIV/AIDS *diagnoses* among men, ~75% were due to male-to-male sexual contact, ~14% to heterosexual contact, ~8% to injection drug use, and ~4% to a combination of male-to-male sexual contact and injection drug use. Of new HIV/AIDS diagnoses among women, ~85% were due to heterosexual contact and ~15% to injection drug use. It is important to note that the number of new diagnoses of HIV infection is not the same as the number of new infections with HIV (incidence): a person may be infected with HIV for years before being diagnosed.

Figure 189-12

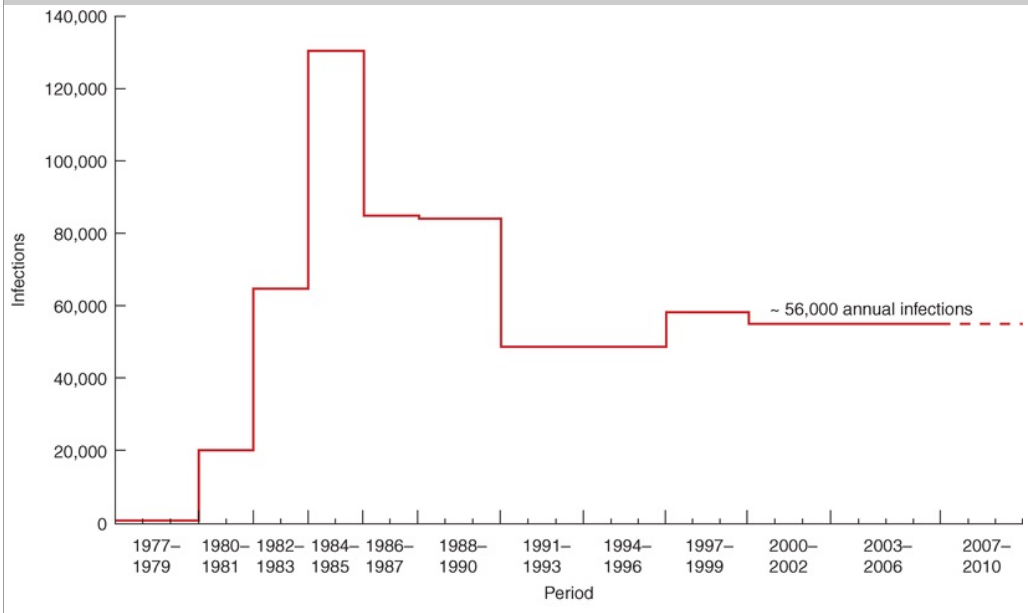

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated number of new HIV infections, United States. *(From Hall et al.)*

Figure 189-13

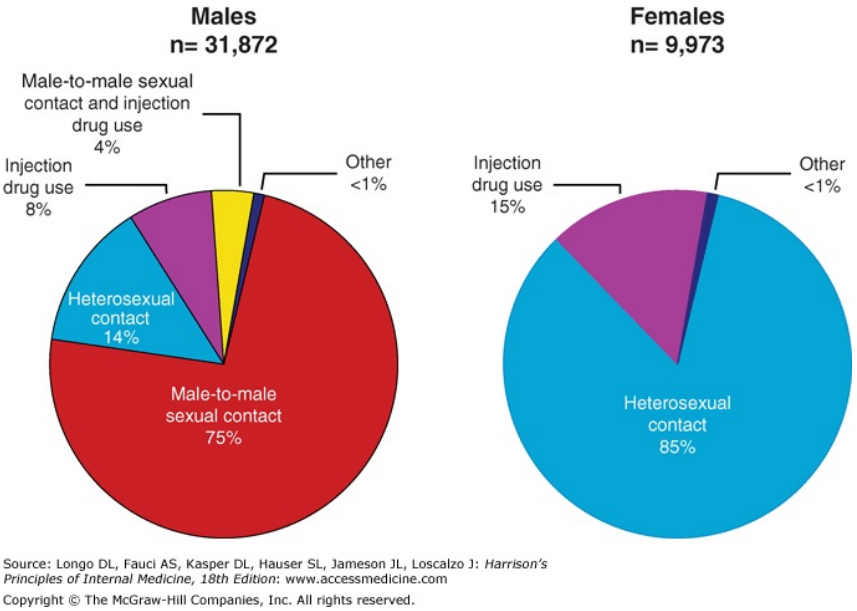

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transmission categories of adults and adolescents with HIV/AIDS diagnosed during 2009 in the United States. Estimates from 40 states with confidential, name-based HIV infection reporting. Data include persons with a diagnosis of HIV infection regardless of AIDS status at diagnosis. *(From CDC.)*

HIV transmission patterns in the United States have shifted over time **(Fig. 189-14).** When one looks at the totality of data collected from the beginning of the epidemic, ~48% of all AIDS cases are among men who have had sex with men. In 1985, male-to-male sexual contact accounted for an estimated 65% of all AIDS diagnoses; this proportion reached its lowest point in 1999 at ~40% of diagnoses. Since then, the percentage of AIDS diagnoses attributed to male-to-male sexual contact has increased; in 2009 this transmission category accounted for 48% of all AIDS diagnoses. The estimated percentage of AIDS diagnoses attributed to injection drug use increased from 20% to 31% during 1985–1994 and decreased since that time, accounting for 15% of diagnoses in 2009. The estimated percentage of AIDS diagnoses attributed to heterosexual contact increased from 3% in 1985 to 31% in 2009.

Figure 189-14

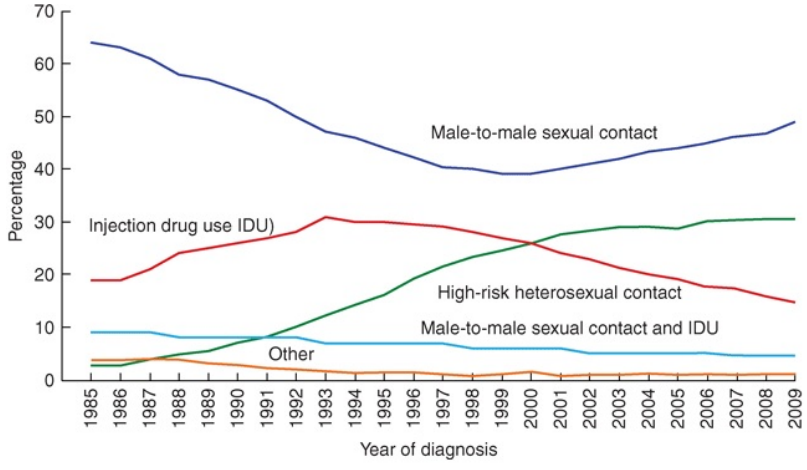

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated AIDS diagnoses among adults and adolescents, by transmission category and year of diagnosis, United States, 1985–2009. *(From CDC.)*

HIV infection and AIDS have disproportionately affected minority populations in the United States. Among those diagnosed with HIV (regardless of AIDS status) in 2009, 52% percent were blacks/African Americans, a group that constitutes only 12% of the U.S. population **(Fig. 189-15*A*)**. The estimated rate of new HIV diagnoses in 2009 by race/ethnicity per 100,000 population is shown in **Fig. 189-15*B***.

Figure 189-15

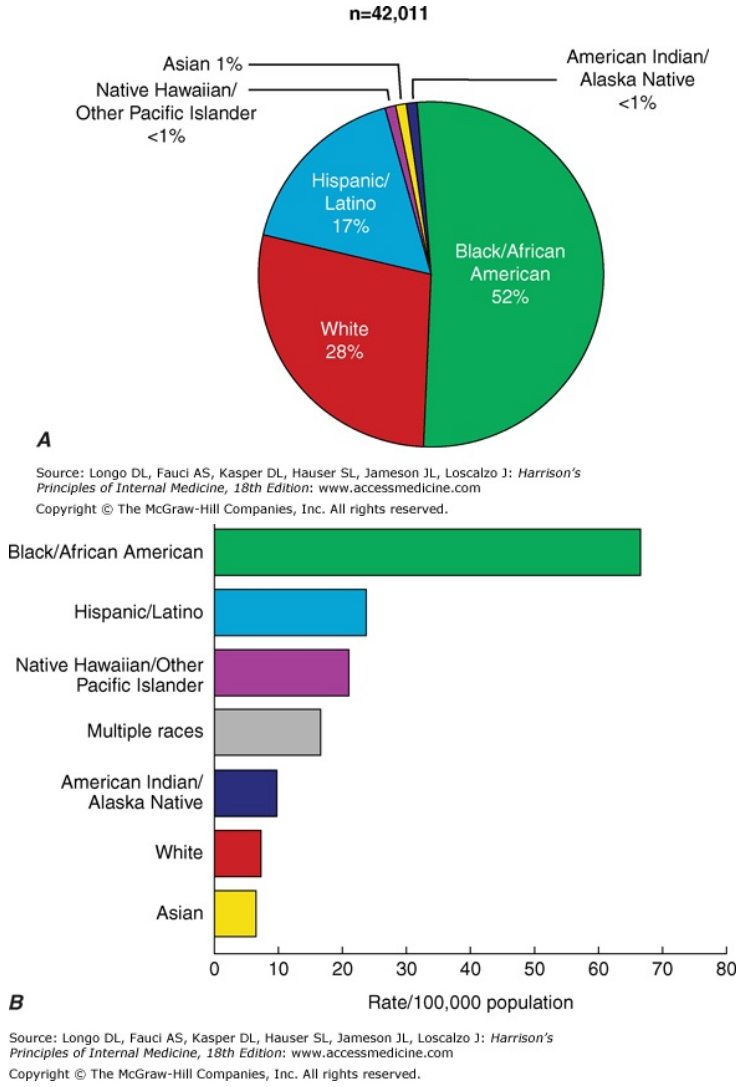

Race/ethnicity of persons (including children) with HIV/AIDS diagnosed during 2009 in the United States. ***A.*** Proportion of new infections by race/ethnicity. ***B.*** Rate of new infections by race/ethnicity (per 100,000 population). Estimates from 40 states with confidential, name-based HIV infection reporting. Data include persons with a diagnosis of HIV infection regardless of AIDS status at diagnosis. *(From CDC.)*

As of January 1, 2010, an estimated 9448 cases of AIDS in children <13 years old had been diagnosed in the Unites States, and ~59% of these individuals have died. Approximately 91% of these children were born to mothers who were HIV-infected or who were at risk for HIV infection; in the majority of those cases, the mother was either an IDU or the heterosexual partner of an IDU. The estimated number of AIDS cases diagnosed among children perinatally exposed to HIV peaked in 1992 and has decreased in recent years **(Fig. 189-16)**. The decline of these cases is likely associated with the implementation of guidelines for the universal counseling and voluntary HIV testing of pregnant women and the use of cART for pregnant women and newborn infants in order to prevent infection. Another contributing factor is the effective treatment of HIV infection in children who have become infected.

Figure 189-16

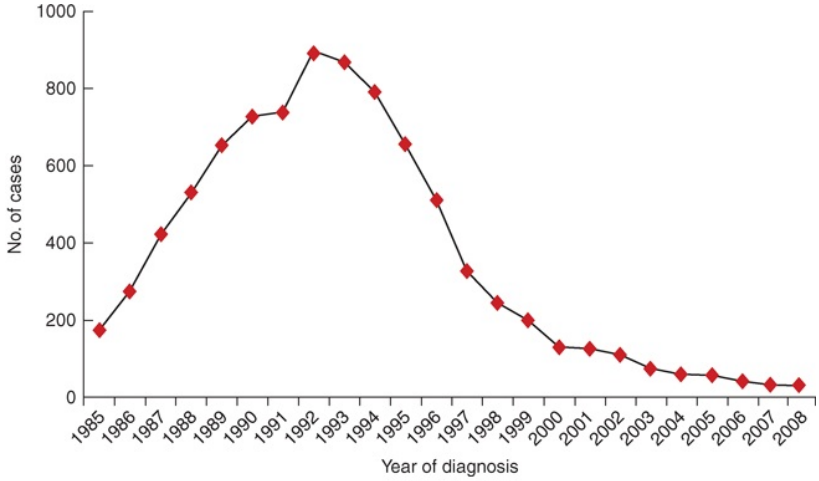

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estimated numbers of perinatally acquired AIDS cases in children by year of diagnosis, 1985–2009, United States *(From CDC.)*

Although the HIV/AIDS epidemic on the whole is plateauing in the United States, it is spreading rapidly among certain populations, stabilizing in others, and decreasing in others. Similar to other STDs, HIV infection will not spread homogeneously throughout the population of the United States. However, it is clear that anyone who practices high-risk behavior is at risk for HIV infection. In addition, recent increases in infections and AIDS cases among young men who have sex with men as well as the spread in pockets of poverty in both urban and rural regions (particularly among underserved minority populations in the southern United States with inadequate access to health care) testify that the epidemic of HIV infection in the United States remains a public health problem of major proportions.

PATHOPHYSIOLOGY AND PATHOGENESIS

The hallmark of HIV disease is a profound immunodeficiency resulting primarily from a progressive quantitative and qualitative deficiency of the subset of T lymphocytes referred to as *helper T cells* occurring in a setting of polyclonal immune activation. The *helper* subset of T cells is defined phenotypically by the presence on its surface of the CD4 molecule (Chap. 314), which serves as the primary cellular receptor for HIV. A co-receptor must also be present together with CD4 for efficient binding, fusion, and entry of HIV-1 into its target cells (Figs. 189-3 and 189-4). HIV uses two major co-receptors, CCR5 and CXCR4, for fusion and entry; these co-receptors are also the primary receptors for certain chemoattractive cytokines termed *chemokines* and belong to the seven-transmembrane-domain G protein–coupled family of receptors. A number of mechanisms responsible for cellular depletion and/or immune dysfunction of CD4+ T cells have been demonstrated in vitro; these include direct infection and destruction of these cells by HIV, as well as indirect effects such as immune clearance of infected cells, immune exhaustion due to aberrant cellular activation, and activation-induced cell death. Patients with CD4+ T cell levels below certain thresholds are at high risk of developing a variety of opportunistic diseases, particularly the infections and neoplasms that are AIDS-defining illnesses. Some features of AIDS, such as Kaposi's sarcoma and certain neurologic abnormalities, cannot be explained completely by the immunodeficiency caused by HIV infection, since these complications may occur prior to the development of severe immunologic impairment.

The combination of viral pathogenic and immunopathogenic events that occurs during the course of HIV disease from the moment of initial (primary) infection through the development of advanced-stage disease is complex and varied. It is important to appreciate that the pathogenic mechanisms of HIV disease are multifactorial and multiphasic and are different at different stages of the disease. Therefore, it is essential to consider the typical clinical course of an untreated HIV-infected individual in order to more fully appreciate these pathogenic events **(Fig. 189-17)**.

Figure 189-17

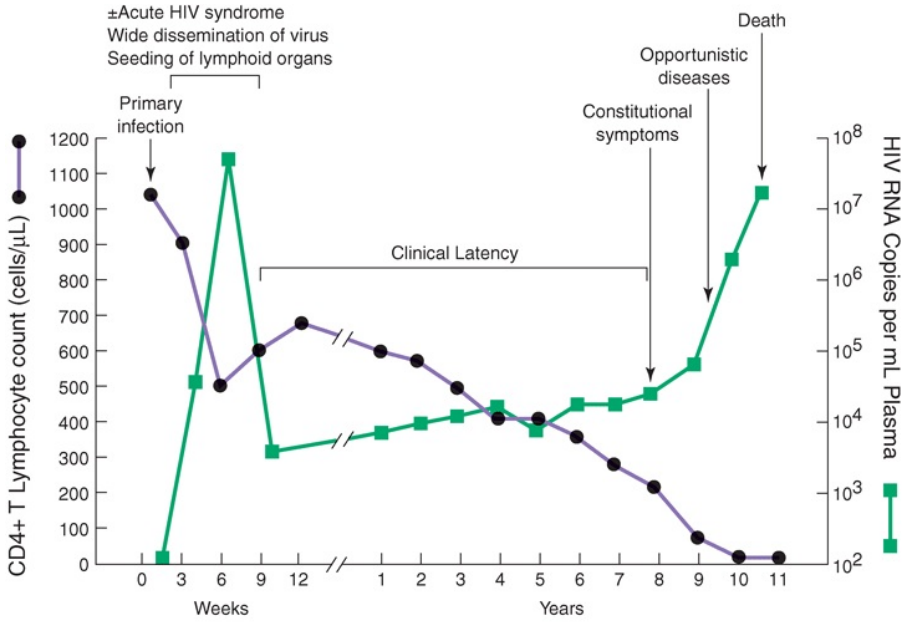

Typical course of an untreated HIV-infected individual. See text for detailed description. *(From G Pantaleo et al: N Engl J Med 328:327, 1993. Copyright 1993 Massachusetts Medical Society. All rights reserved.)*

Early Events in HIV Infection: Primary Infection and Initial Dissemination of Virus

Using mucosal transmission as a model, the earliest events (within hours) that occur following exposure of HIV to the mucosal surface determine whether an infection will be established as well as the subsequent course of events following infection. Although the mucosal barrier is relatively effective in limiting access of HIV to susceptible targets in the lamina propria, the virus can cross the barrier by transport on dendritic cells just beneath the surface or through microscopic rents in the mucosa. Significant disruptions in the mucosal barrier as seen in ulcerative genital disease facilitate viral entry and increase the efficiency of infection. Virus then seeks susceptible targets, which are primarily CD4+ T cells that are spatially dispersed in the mucosa. This spatial dispersion of targets provides a significant obstacle to the establishment of infection. Such obstacles account for the low efficiency of sexual transmission of HIV (see "Sexual Transmission," above). Both "partially" resting CD4+ T cells and activated CD4+ T cells serve as early amplifiers of infection. Resting CD4+ T cells are more abundant; however, activated CD4+ T cells produce larger amounts of virus. In order for infection to become established, the basic reproductive rate (R_0) must become greater or equal to 1, i.e., each infected cell would infect at least one other cell. As virus is produced within days to weeks, it is disseminated, first to the draining lymph nodes and then to other lymphoid compartments where it has easy access to dense concentrations of CD4+ T cell targets, allowing for a burst of high-level viremia **(Fig. 189-18).** An important lymphoid organ, the gut-associated lymphoid tissue (GALT), is a major target of HIV infection and the location where large numbers of CD4+ T cells (usually memory cells) are infected and depleted, both by direct viral effects and by activation-associated apoptosis. Once virus replication reaches this threshold and virus is widely disseminated, infection is firmly established and the process is irreversible. It is important to point out that the initial infection of susceptible cells may vary somewhat with the route of infection. Virus that enters directly into the bloodstream via infected blood or blood products (i.e., transfusions, use of contaminated needles for injection drugs, sharp-object injuries, maternal-to-fetal transmission either intrapartum or perinatally, or sexual intercourse where there is enough trauma to cause bleeding) is likely cleared from the circulation to the spleen and other lymphoid organs, where primary focal infections begin, followed by wider dissemination throughout other lymphoid tissues as described above.

Figure 189-18

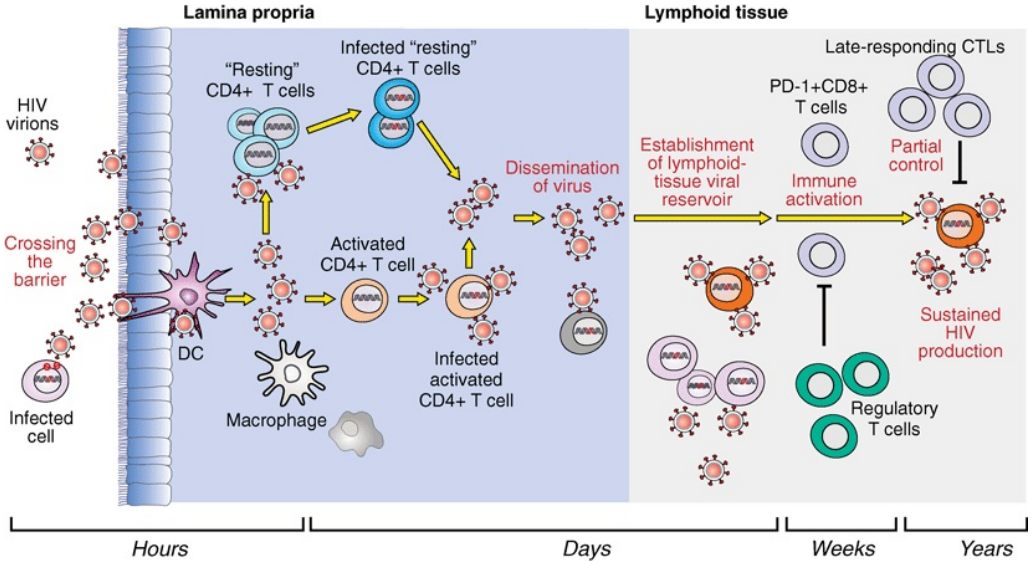

Summary of early events in HIV infection. See text for detailed description. CTLs, cytolytic T lymphocytes; HIV, human immunodeficiency virus. *(Adapted from Haase, 2005.)*

It has been demonstrated that sexual transmission of HIV is the result of a single infectious event and that a viral genetic bottleneck exists for transmission. In this regard, certain characteristics of the HIV envelope glycoprotein have a major influence on transmission, at least in subtype A and C viruses. Transmitting viruses, often referred to as "founder viruses," are usually underrepresented in the circulating viremia of the transmitting partner and are less-diverged viruses with signature sequences including shorter V1–V2 loop sequences and fewer predicted N-linked glycosylation sites relative to the major circulating variants. These viruses are usually sensitive to neutralization by antibody from the transmitting partner. Once replication proceeds in the newly infected partner, the founder virus diverges and accumulates glycosylation sites, becoming progressively more resistant to neutralization **(Fig. 189-19).**

Figure 189-19

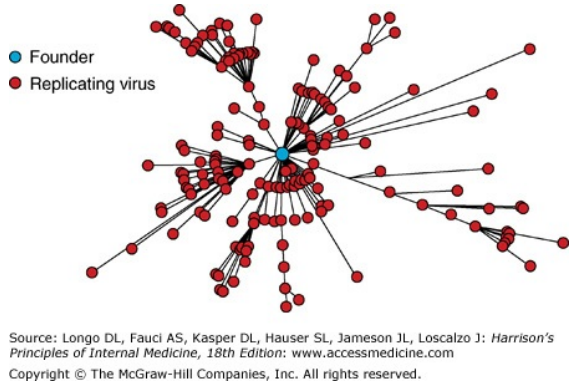

As HIV diverges from founder to chronically replicating virus, it accumulates N-linked glycosylation sites. See text for detailed description. *(Adapted from CA Derdeyn et al*: Science *303:2019, 2004; B Chohan et al:J Virol 79:6528, 2005; and BF Keele et al:Proc Natl Acad Sci USA 105:7552, 2008.)*

The acute burst of viremia and wide dissemination of virus in primary HIV infection may be associated with an *acute HIV syndrome,* which occurs to varying degrees in ~50% of individuals with primary infection. This syndrome is usually associated with high levels of viremia measured in millions of copies of HIV RNA per milliliter of plasma that last for several weeks. Acute mononucleosis-like symptoms are well correlated with the presence of viremia. Virtually all patients develop some degree of viremia during primary infection, which contributes to virus dissemination throughout the lymphoid tissue, even though they may remain asymptomatic or not recall experiencing symptoms. It appears that the initial level of plasma viremia in primary HIV infection does not necessarily determine the rate of disease progression; however, the set point of the level of steady-state plasma viremia after ~1 year does seem to correlate with the slope of disease progression in the untreated patient.

Establishment of Chronic and Persistent Infection

PERSISTENT VIRUS REPLICATION

HIV infection is unique among human viral infections. Despite the robust cellular and humoral immune responses that are mounted following primary infection (see "Immune Response to HIV," below), once infection has been established the virus succeeds in escaping immune-mediated clearance, paradoxically seems to thrive on immune activation, and is never eliminated completely from the body. Rather, a chronic infection develops and persists with varying degrees of continual virus replication in the untreated patient for a median of ~10 years before the patient becomes clinically ill (see "Advanced HIV Disease" below). It is this establishment of a chronic, persistent infection that is the hallmark of HIV disease. Throughout the often protracted course of chronic infection, virus replication can invariably be detected in untreated patients, both by highly sensitive assays for plasma viremia as well as by demonstration of cell-associated HIV RNA in immunocompetent cells (predominantly CD4+ T cells and macrophages) in the circulation and in lymphoid tissue. Recent studies using highly sensitive molecular techniques have demonstrated that even in certain patients in whom plasma viremia is suppressed to below 50 copies of HIV RNA/mL by cART,

there is a continual low level of virus replication. In other human viral infections, with very few exceptions, if the host survives, the virus is completely cleared from the body and a state of immunity against subsequent infection develops. HIV infection very rarely kills the host during primary infection. Certain viruses, such as HSV (Chap. 179), are not completely cleared from the body after infection but instead enter a latent state; in these cases, clinical latency is accompanied by microbiologic latency. This is not the case with HIV infection as described above. Chronicity associated with persistent virus replication can also be seen in certain cases of HBV and HCV infections (Chap. 306); however, in these infections the immune system is not a target of the virus.

EVASION OF IMMUNE SYSTEM CONTROL

Inherent to the establishment of chronicity of HIV infection is the ability of the virus to evade elimination and control by the immune system. There are a number of mechanisms whereby the virus accomplishes this evasion. Paramount among these is the establishment of a sustained level of replication associated with the generation of viral diversity via mutation and recombination, thus providing a means to evade control and elimination by the immune system. The selection of mutants that escape control by CD8+ cytolytic T lymphocytes (CTLs) is critical to the propagation and progression of HIV infection. The high rate of virus replication associated with inevitable mutations also contributes to the inability of neutralizing antibody to contain the virus quasispecies present in an individual at any given time. Extensive analyses of sequential HIV isolates and host responses have demonstrated that viral escape from B cell and CD8+ T cell epitopes occurs early after infection and allows the virus to stay one step ahead of effective immune responses. In addition, clones of CD8+ CTLs that expand greatly during primary HIV infection, and likely represent the high-affinity clones that would be expected to be most efficient in eliminating virus-infected cells, are no longer detectable after their initial burst of expansion. It is thought that the initially expanded clones may have been deleted or rendered dysfunctional owing to the overwhelming immune activation resulting from persistent viral replication, similar to the exhaustion of CD8+ CTLs that has been reported in the murine model of lymphocytic choriomeningitis virus (LCMV) infection. Recent studies have indicated that exhaustion of effector cells during prolonged immune activation is associated with expression of the programmed death (PD) 1 molecule (of the B7-CD28 family of molecules) on activated cells and its interaction with its ligands (L) PD-L1 and PD-L2 on antigen-presenting cells. This interaction results in a partially reversible signal for cell death and/or dysfunction. Another mechanism contributing to the evasion by HIV of immune system control is the downregulation of HLA class I molecules on the surface of HIV-infected cells by the Nef protein of HIV, resulting in the lack of ability of the CD8+ CTL to recognize and kill the infected target cell. Although this downregulation of HLA class I molecules would favor elimination of HIV-infected cells by natural killer (NK) cells, this latter mechanism does not seem to remove HIV-infected cells effectively (see below). The principal targets of neutralizing antibodies against HIV are the envelope proteins gp120 and gp41. HIV employs at least three mechanisms to evade neutralizing responses: hypervariability in the primary sequence of the envelope, extensive glycosylation of the envelope, and conformational masking of neutralizing epitopes.

CD4+ T cell help is essential for the integrity of antigen-specific immune responses, both humoral and cell-mediated. HIV preferentially infects activated CD4+ T cells including HIV-specific CD4+ T cells, and so this loss of viral-specific helper T cell responses has profound negative consequences for the immunologic control of HIV replication. Furthermore, this loss occurs early in the course of infection, and animal studies indicate that 40–70% of all memory CD4+ T cells in the GALT are eliminated during acute infection. Another potential means of escape of HIV-infected cells from elimination by CD8+ CTLs is the sequestration of infected cells in immunologically privileged sites such as the central nervous system (CNS).

Finally, the escape of HIV from elimination during primary infection allows the formation of a pool of latently infected cells that cannot be eliminated by virus-specific CTLs (see below). Thus, despite a potent immune response and the marked downregulation of virus replication following primary HIV infection, HIV succeeds in establishing a state of chronic infection with a variable degree of persistent virus replication. During this period most patients make the clinical transition from acute primary infection to variable periods of clinical latency or smoldering disease activity (see below).

RESERVOIRS OF HIV-INFECTED CELLS: OBSTACLES TO THE ERADICATION OF VIRUS

There exists in virtually all HIV-infected individuals a pool of latently infected, resting CD4+ T cells that serves as at least one component of the persistent reservoir of virus. Such cells manifest postintegration latency in that the HIV provirus integrates into the genome of the cell and can remain in this state until an activation signal drives the expression of HIV transcripts and ultimately replication-competent virus. This form of latency is to be distinguished from preintegration latency, in which HIV enters a resting CD4+ T cell and, in the absence of an activation signal, reverse transcription of the HIV genome occurs to a certain extent but the resulting proviral DNA fails to integrate into the host genome. This period of preintegration latency may last hours to days, and if no activation signal is delivered to the cell, the proviral DNA loses its capacity to initiate a productive infection. If these cells do become activated, reverse transcription proceeds to completion and the virus continues along its replication cycle (see above and **Fig. 189-20**). The pool of cells that are in the postintegration state of latency is established early during the course of primary HIV infection. Despite the suppression of plasma viremia to <50 copies of HIV RNA per milliliter by potent combinations of several antiretroviral drugs administered over several years, this pool of latently infected cells persists and can give rise to replication-competent virus. Modeling studies built on projections of decay curves have estimated that in such a setting of prolonged suppression, it would require 7–70 years for the pool of latently infected cells to be completely eliminated. Furthermore, the reservoir of latently infected cells is replenished during minor detectable rebounds or "blips" of virus replication that may occur intermittently, superimposed on the low levels of persistent virus replication that may remain below the limits of detection of current assays (see below) (Fig. 189-20), even in patients who for the most part are treated successfully. Reservoirs of HIV-infected cells, latent or otherwise, can exist in a number of compartments including the lymphoid tissue, peripheral blood, and the CNS (likely in cells of the monocyte/macrophage lineage) as well as in other unidentified locations. Over the past several years attempts have been made to eliminate HIV in the latent viral reservoir using agents that stimulate resting CD4+ T cells during the course of antiretroviral therapy; however, such attempts have been unsuccessful. Thus, this persistent reservoir of infected cells at various stages of latency and/or low levels of persistent virus replication are major obstacles to any goal of eradication of virus from infected individuals, despite the favorable clinical outcomes that have resulted from cART.

Figure 189-20

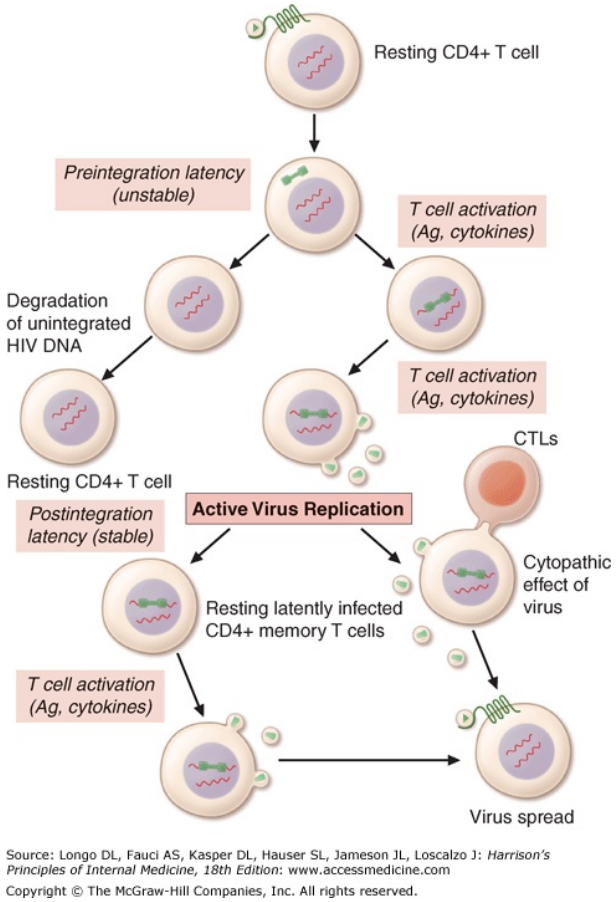

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Generation of latently infected, resting CD4+ T cells in HIV-infected individuals. See text for details. Ag, antigen; CTLs, cytolytic T lymphocytes. *(Courtesy of TW Chun; with permission.*)

VIRAL DYNAMICS

The dynamics of viral production and turnover have been quantified using mathematical modeling in the setting of the administration of reverse transcriptase and protease inhibitors to HIV-infected individuals in clinical studies. Treatment with these drugs resulted in a precipitous decline in the level of plasma viremia, which typically fell by well over 90% within 2 weeks. The number of CD4+ T cells in the blood increased concurrently, which suggested that the killing of CD4+ T cells was linked directly to the levels of replicating virus. However, a significant component of the early rise in CD4+ T cell numbers following the initiation of therapy may be due to the redistribution of cells into the peripheral blood from other body compartments as a consequence of therapy-related diminution in viremia-associated immune system activation. It was determined on the basis of modeling the kinetics of viral decline and the emergence of resistant mutants during therapy that 93–99% of the circulating virus originated from recently infected, rapidly turning over CD4+ T cells and that ~1–7% of circulating virus originated from longer-lived cells, likely monocytes/macrophages. A negligible amount of circulating virus originated from the pool of latently infected cells **(Fig. 189-21)**. It was also determined that the half-life of a circulating virion was ~30–60 min and that of productively infected cells was 1 day. Given the relatively steady level of plasma viremia and of infected cells, it appears that extremely large amounts of virus (~10^{10}–10^{11} virions) are produced and cleared from the circulation each day. In addition, data suggest that the minimal duration of the HIV-1 replication cycle in vivo is ~2 days. Other studies have demonstrated that the decrease in plasma viremia that results from cART correlates closely with a decrease in virus replication in lymph nodes, further confirming that lymphoid tissue is the main site of HIV replication and the main source of plasma viremia.

Figure 189-21

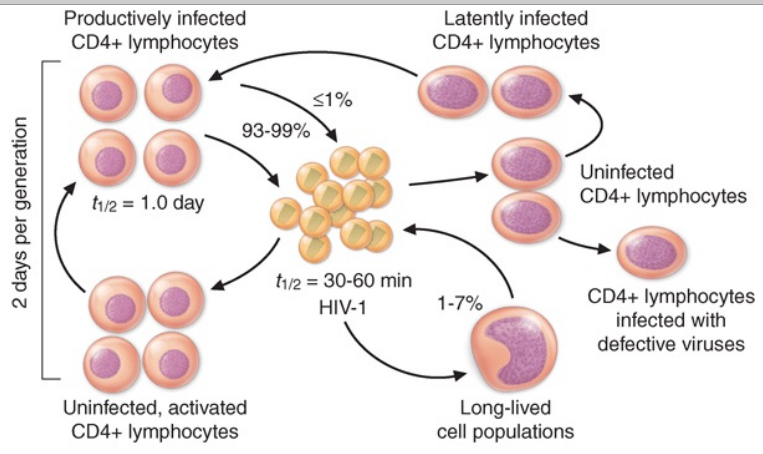

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dynamics of HIV infection in vivo. See text for detailed description. *(From AS Perelson et al: Science271:1582, 1996.)*

The level of steady-state viremia, called the viral *set point*, at ~1 year has important prognostic implications for the progression of HIV disease in the untreated patient. It has been demonstrated that as a group untreated HIV-infected individuals who have a low set point at 6 months to 1 year following infection progress to AIDS much more slowly than individuals whose set point is very high at that time **(Fig. 189-22)**.

Figure 189-22

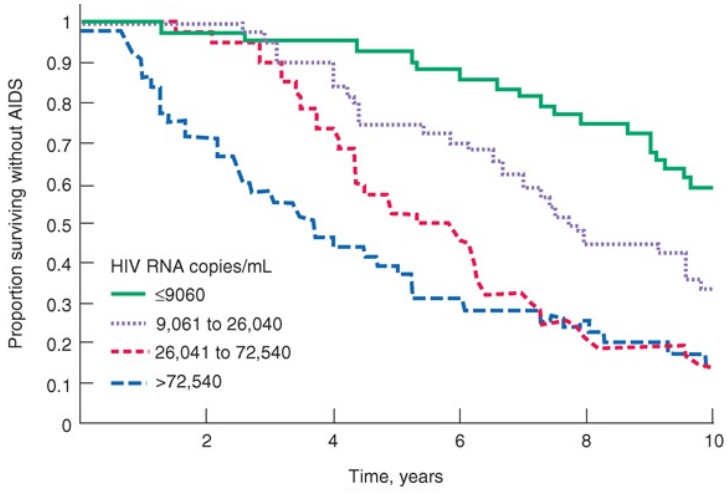

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship between levels of virus and rates of disease progression. Kaplan-Meier curves for AIDS-free survival stratified by baseline HIV-1 RNA categories (copies per milliliter). *(From Mellors et al.)*

CLINICAL LATENCY VERSUS MICROBIOLOGIC LATENCY

With the exception of long-term nonprogressors (see "Long-Term Survivors and Long-Term Nonprogessors," below), the level of CD4+ T cells in the blood decreases progressively in HIV-infected individuals. The decline in CD4+ T cells may be gradual or abrupt, the latter usually reflecting a significant spike in the level of plasma viremia. Most patients are relatively asymptomatic while this progressive decline is taking place (see below) and are often described as being in a state of *clinical latency*. However, this term is misleading; it does not mean disease latency, since progression, although slow in many cases, is generally relentless during this period. Furthermore, clinical latency should not be confused with microbiologic latency, since varying levels of virus replication inevitably occur during this period of clinical latency. Even in those rare patients who have <50 copies of HIV RNA per milliliter in the absence of therapy, there is virtually always some degree of ongoing virus replication.

Advanced HIV Disease

In untreated patients or in patients in whom therapy has not adequately controlled virus replication, after a variable period, usually measured in years, the CD4+ T cell count falls below a critical level (<200/μL) and the patient becomes highly susceptible to opportunistic disease (Fig. 189-17). For this reason, the CDC case definition of AIDS includes all HIV-infected individuals with CD4+ T cell counts below this level (Table 189-1). Patients may experience constitutional signs and symptoms or may develop an opportunistic disease abruptly without any prior symptoms, although the latter scenario is unusual. The depletion of CD4+ T cells continues to be progressive and unrelenting in this phase. It is not uncommon for CD4+ T cell counts in the untreated patient to drop as low as 10/μL or even to zero. In countries where cART and prophylaxis and treatment for opportunistic infections are readily accessible to such patients, survival is increased dramatically even in those patients with advanced HIV disease. In contrast, untreated patients who progress to this severest form of immunodeficiency usually succumb to opportunistic infections or neoplasms (see below).

Long-Term Survivors and Long-Term Nonprogressors

It is important to distinguish between the terms *long-term survivor* and *long-term nonprogressor*. Long-term nonprogressors are by definition long-term survivors; however, the reverse is not always true. Predictions from one study that antedated the availability of effective cART estimated that ~13% of homosexual/bisexual men who were infected at an early age may remain free of clinical AIDS for >20 years. Many of these individuals may have progressed in their degree of immune deficiency; however, they certainly survived for a considerable period of time. With the advent of effective antiretroviral therapy, the survival of HIV-infected individuals has dramatically increased. Early in the AIDS epidemic, prior to the availability of therapy, if a patient presented with a life-threatening opportunistic infection, the median survival was 26 weeks from the time of presentation. Currently, an HIV-infected 20-year-old individual in a high-income country who is appropriately treated with cART can expect to live at least 50 years according to mathematical model projections. In the face of cART, long-term survival is becoming commonplace. Definitions of long-term nonprogressors have varied considerably over the years, and so such individuals constitute a heterogeneous group. Long-term nonprogressors were first described in the 1990s. Originally, individuals were considered to be long-term nonprogressors if they had been infected with HIV for a long period (≥10 years), their CD4+ T cell counts were in the normal range, and they remained stable over years without receiving cART. Approximately 5–15% of HIV-infected individuals fell into this broader nonprogressor category. However, this group was rather heterogenous and over time a significant proportion of these individuals progressed and ultimately required therapy. From this broader group, a much smaller subgroup of "elite" controllers or nonprogressors was identified, and they constituted less than 1% of HIV-infected individuals. These elite controllers, by definition, have extremely low levels of plasma viremia and normal CD4+ T cell counts. It is noteworthy that certain of their HIV-specific immune responses are robust and clearly superior to those of HIV-infected progressors. In this group of elite controllers certain HLA class I haplotypes are overrepresented, particularly HLA-B57-01 and HLA-B27-05. Outside of the subgroup of elite controllers, a number of other genetic factors have been shown to be involved to a greater or lesser degree in the control of virus replication and thus in the rate of HIV disease progression (see "Genetic Factors in HIV Pathogenesis," below).

Lymphoid Organs and HIV Pathogenesis

Regardless of the portal of entry of HIV, lymphoid tissues are the major anatomic sites for the establishment and propagation of HIV infection. Despite the use of measurements of plasma viremia to determine the level of disease activity, virus replication occurs mainly in lymphoid tissue and not in blood; indeed, the level of plasma viremia directly reflects virus production in lymphoid tissue.

Some patients experience progressive generalized lymphadenopathy early in the course of the infection; others experience varying degrees of transient lymphadenopathy. Lymphadenopathy reflects the cellular activation and immune response to the virus in the lymphoid tissue, which is generally characterized by follicular or germinal center hyperplasia. Lymphoid tissue involvement is a common denominator of virtually all patients with HIV infection, even those without easily detectable lymphadenopathy.

Simultaneous examinations of lymph tissue and peripheral blood in patients and monkeys during various stages of HIV and SIV infection, respectively, have led to substantial insight into the pathogenesis of HIV disease. In most of the original human studies, peripheral lymph nodes have been used predominantly as the source of lymphoid tissue. More recent studies in monkeys and humans have focused on the GALT, where the earliest burst of virus replication occurs associated with marked depletion of CD4+ T cells. In detailed studies of peripheral lymph node tissue, using a combination of polymerase chain reaction (PCR) techniques for HIV DNA and HIV RNA in tissue and HIV RNA in plasma, in situ hybridization for HIV RNA, and light and electron microscopy, the following picture has emerged. During acute HIV infection resulting from mucosal transmission, virus replication progressively amplifies from scattered lymphoid cells in the lamina propria to draining lymphoid tissue, leading to high levels of plasma viremia. The GALT plays a major role in the amplification of virus replication, and virus is disseminated from replication in the GALT to peripheral lymphoid tissue. A profound degree of cellular activation occurs (see below) and is reflected in follicular or germinal center hyperplasia. At this time copious amounts of extracellular virions (both infectious and defective) are trapped on the processes of the follicular dendritic cells (FDCs) in the germinal centers of the lymph nodes. Virions that have bound complement components on their surfaces attach to the surface of FDCs via interactions with complement receptors and likely via Fc receptors that bind to antibodies that are attached to the virions. In situ hybridization reveals expression of virus in individual cells of the paracortical area and, to a lesser extent, the germinal center **(Fig. 189-23)**. The persistence of trapped virus after the transition from acute to chronic infection likely reflects a steady state whereby trapped virus turns over and is replaced by fresh virions that are continually produced. The trapped virus, either as whole virion or shed envelope, serves as a continual activator of CD4+ T cells, thus driving further virus replication.

Figure 189-23

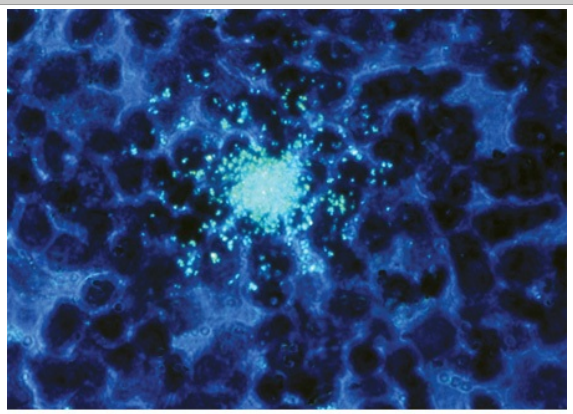

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

HIV in the lymph node of an HIV-infected individual. An individual cell infected with HIV shown expressing HIV RNA by in situ hybridization using a radiolabeled molecular probe. Original x500. *(Adapted from G Pantaleo, AS Fauci.)*

During early-stage HIV disease, the architecture of the germinal centers is generally preserved and may even be hyperplastic owing to in situ proliferation of cells (mostly B lymphocytes) and recruitment to the lymph nodes of a number of cell types (B cells, CD4+ and CD8+ T cells). Electron-microscopic studies have demonstrated a fine network of FDCs with many long, fingerlike processes that envelop virtually every lymphocyte in the germinal center. Extracellular virions can be seen attached to the processes, yet the FDCs appear to be relatively healthy. The trapping of antigen is a physiologically normal function for the FDCs, which present antigen to B cells and contribute to the generation of B cell memory. However, in the case of HIV, the trapped virions serve as a persistent source of cellular activation, resulting in the secretion of proinflammatory cytokines such as interleukin (IL) 1β, tumor necrosis factor (TNF) α, and IL-6, which can upregulate virus replication in infected cells (see below). Furthermore, although trapped virus is coated by neutralizing antibodies, it has been demonstrated that certain of these virions remain infectious for CD4+ T cells while attached to the processes of the FDCs. CD4+ T cells that migrate into the germinal center to provide help to B cells in the generation of an HIV-specific immune response are susceptible to infection by these trapped virions. Thus, in HIV infection, a normal physiologic function of the immune system that contributes to the clearance of virus, as well as to the generation of a specific immune response, can also have deleterious consequences.

As the disease progresses, the architecture of the germinal centers begins to show disruption. Electron microscopy reveals swollen organelles, and the FDCs begin to undergo cell death. The mechanisms of FDC death remain unclear; there is no indication by electron microscopy of copious virus replication or budding of virions off the cell in great quantities. This process of FDC death is accompanied by the deposition of collagen, leading to irreparable damage to the germinal centers. As the disease progresses to an advanced stage, there is complete disruption of the architecture of the germinal centers, accompanied by dissolution of the FDC network and massive dropout of FDCs. At this point, the lymph nodes are "burnt out." This destruction of lymphoid tissue compounds the immunodeficiency of HIV disease and contributes both to the inability to control HIV replication (leading usually to high levels of plasma viremia in the untreated or inadequately treated patient) and to the inability to mount adequate immune responses against opportunistic pathogens. The events from primary infection to the ultimate destruction of the immune system are illustrated in **Fig. 189-24**. Recently, nonhuman primate studies and some human studies have examined GALT at various stages of HIV disease. Within the GALT, the basal level of activation combined with virus-mediated cellular activation results in the infection and elimination of an estimated 50–90% of CD4+ T cells in the gut. The extent of this early damage to GALT, which constitutes a major component of lymphoid tissue in the body, may play a role in determining the potential for immunologic recovery of the memory cell subset.

Figure 189-24

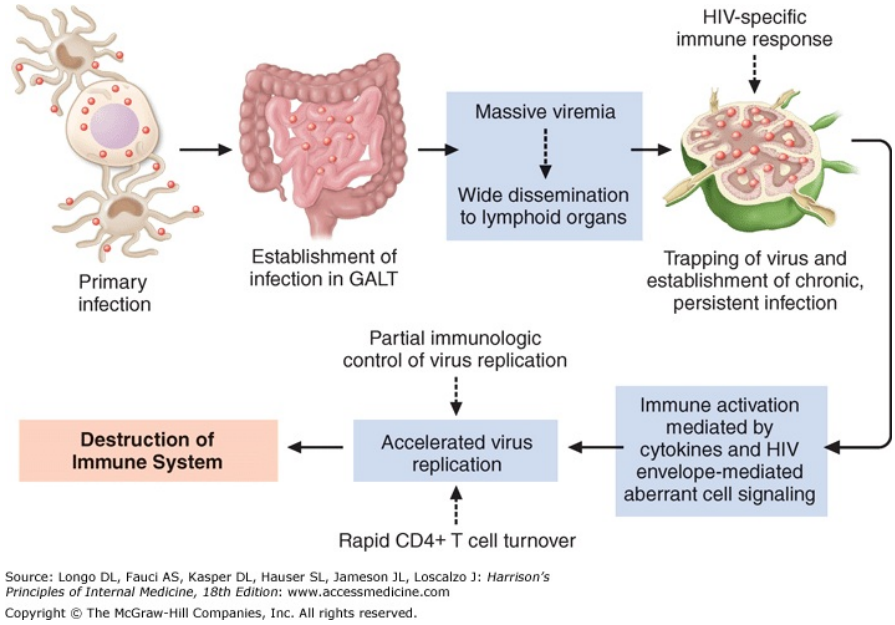

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Events that transpire from primary HIV infection through the establishment of chronic persistent infection to the ultimate destruction of the immune system. See text for details. CTLs, cytolytic T lymphocytes; GALT, gut-associated lymphoid tissue.

Immune Activation, Inflammation, and HIV Pathogenesis

Activation of the immune system and variable degrees of inflammation are essential components of any appropriate immune response to a foreign antigen. However, immune activation and inflammation, which can be considered aberrant in HIV-infected individuals, play a critical role in the pathogenesis of HIV disease and other chronic conditions associated with HIV disease. Immune activation and inflammation in the HIV-infected individual contribute substantially to (1) the replication of HIV, (2) the induction of immune dysfunction, and (3) the increased incidence of chronic conditions associated with persistent immune activation and inflammation **(Table 189-3)**.

Table 189-3 Conditions Associated with Persistent Immune Activation and Inflammation in Patients with HIV Infection

- Accelerated aging syndrome
- Bone fragility
- Cancers
- Cardiovascular disease
- Diabetes
- Kidney disease
- Liver disease
- Neurocognitive dysfunction

Induction of HIV Replication by Aberrant Immune Activation

The immune system is normally in a state of homeostasis, awaiting perturbation by foreign antigenic stimuli. Once the immune response deals with and clears the antigen, the system returns to relative quiescence (Chap. 314). This is generally not the case in HIV infection where, in the untreated patient, virus replication is invariably persistent with very few exceptions and immune activation is persistent. HIV replicates most efficiently in activated CD4+ T cells; in HIV infection, chronic activation provides the cell substrates necessary for persistent virus replication throughout the course of HIV disease, particularly in the untreated patient and to variable degrees even in certain patients receiving cART whose levels of plasma viremia are suppressed to below the level of detection by standard assays. From a virologic standpoint, although quiescent CD4+ T cells can be infected with HIV, reverse transcription, integration, and virus spread are much more efficient in activated cells. Furthermore, cellular activation induces expression of virus in cells latently infected with HIV. In essence, immune activation and inflammation provide the engine that drives HIV replication. In addition to endogenous factors such as cytokines, a number of exogenous factors such as other microbes that are associated with heightened cellular activation can enhance HIV replication and thus may have important effects on HIV pathogenesis. Co-infection in vivo or in vitro with a range of viruses, such as HSV types 1 and 2, cytomegalovirus (CMV), human herpesvirus (HHV) 6, Epstein-Barr virus (EBV), HBV, adenovirus, and HTLV-I have been shown to upregulate HIV expression. In addition, infestation with nematodes has been shown to be associated with a heightened state of immune activation that facilitates HIV replication; in certain studies deworming of the infected host has resulted in a decrease in plasma viremia. Two diseases of extraordinary global health significance, malaria and tuberculosis (TB), have been shown to increase HIV viral load in dually infected individuals. Globally, *Mycobacterium tuberculosis* is the most common opportunistic infection globally in HIV-infected individuals (Chap. 165). In addition to the fact that HIV-infected individuals are more likely to develop active TB after exposure, it has been demonstrated that active TB can accelerate the course of HIV infection. It has also been shown that levels of plasma viremia are greatly elevated in HIV-infected individuals with active TB who are not on cART, compared with pre-TB levels and levels of viremia after successful treatment of the active TB. The situation is similar in the interaction between HIV and malaria parasites (Chap. 210). Acute infection of HIV-infected individuals with *Plasmodium falciparum* increases HIV viral load, and the increased viral load is reversed by effective malaria treatment.

Microbial Translocation and Persistent Immune Activation

One proposed mechanism of persistent immune activation involves the disruption of the mucosal barrier in the gut due to HIV replication in and disruption of submucosal lymphoid tissue. As a result of this disruption, there is an increase in the products, particularly lipopolysaccharide (LPS), of bacteria that translocate from the bowel lumen through the damaged mucosa to the circulation, leading to persistent systemic immune activation and inflammation. This effect can persist even after the HIV viral load is brought to <50 copies/mL by cART. Depletion in the GALT of IL-17–producing T cells, which are responsible for defense against extracellular bacteria and fungi, is also thought to contribute to HIV pathogenesis.

Persistent Immune Activation and Inflammation Induce Immune Dysfunction

The activated state in HIV infection is reflected by hyperactivation of B cells leading to hypergammaglobulinemia; increased lymphocyte turnover; activation of monocytes; expression of activation markers on CD4+ and CD8+ T cells; increased activation-associated cellular apoptosis; lymph node hyperplasia, particularly early in the course of disease; increased secretion of proinflammatory cytokines, particularly IL-6; elevated levels of high-sensitivity C-reactive protein, fibrinogen, d-dimer, neopterin, β_2-microglobulin, acid-labile interferon, soluble (s) IL-2 receptors (R), sTNFR, sCD27, and sCD40L; and autoimmune phenomena (see "autoimmune phenomena," below). Even in the absence of direct infection of a target cell, HIV envelope proteins can interact with cellular receptors (CD4 molecules and chemokine receptors) to deliver potent activation signals resulting in calcium flux, the phosphorylation of certain proteins involved in signal transduction, co-localization of cytoplasmic proteins including those involved in cell trafficking, immune dysfunction, and, under certain circumstances, apoptosis. From an immunologic standpoint, chronic exposure of the immune system to a particular antigen over an extended period may ultimately lead to an inability to sustain an adequate immune response to the antigen in question. In many chronic viral infections, including HIV infection, persistent viremia is associated with "functional exhaustion" and apoptosis of virus-specific T cells. It has been demonstrated that this phenomenon may be mediated, at least in part, by the engagement of PD-1, which is highly expressed on the majority of HIV-specific T cells, with its ligands (PD-L1 and PD-L2) on antigen-presenting cells and epithelial cells, resulting in either T cell death or anergy. Furthermore, the ability of the immune system to respond to a broad spectrum of antigens may be compromised if immunocompetent cells are maintained in a state of chronic activation.

The deleterious effects of chronic immune activation on the progression of HIV disease are well established. As in most conditions of persistent antigen exposure, the host must maintain sufficient activation of antigen (HIV)-specific responses but must also prevent excessive activation and potential immune-mediated damage to tissues. Certain studies suggest that normal immunosuppressive mechanisms that act to keep hyperimmune activation in check, particularly CD4+, FoxP3+, CD25+ regulatory T cells (T-regs), may be dysfunctional or depleted in the context of advanced HIV disease.

Medical Conditions Associated with Persistent Immune Activation and Inflammation in HIV Disease

It has become clear as the survival of HIV-infected individuals has increased that a number of previously unrecognized medical complications are associated with HIV disease and that these complications relate to chronic immune activation and inflammation (Table 189-3). These complications can appear even after patients have experienced years of adequate control of viral replication (plasma viremia below detectable levels) for several years. Of particular note are endothelial cell dysfunction and its relationship to cardiovascular disease. Other chronic conditions that have been reported include bone fragility, certain cancers, persistent immune dysfunction, diabetes, kidney and liver disease, and neurocognitive dysfunction, thus presenting an overall picture of accelerated aging.

APOPTOSIS

Apoptosis is a form of programmed cell death that is a normal mechanism for the elimination of effete cells in organogenesis as well as in the cellular proliferation that occurs during a normal immune response (Chap. 314). Apoptosis is largely dependent on cellular activation, and the aberrant cellular activation associated with HIV disease is correlated with a heightened state of apoptosis. HIV can trigger both Fas-dependent and Fas-independent pathways of apoptosis. Mechanisms involved in this process include upregulation of Fas and Fas ligand, upregulation of caspase-1 and caspase-6, downregulation of the antiapoptotic Bcl-2 protein, and activation of cyclin-dependent kinases. Certain viral gene products have been associated with enhanced susceptibility to apoptosis; these include Env, Tat, and Vpr. In contrast, Nef has been shown to possess antiapoptotic properties. A number of studies, including those examining lymphoid tissue, have demonstrated that the rate of apoptosis is elevated in HIV infection and that apoptosis is seen in "bystander" cells such as CD8+ T cells and B cells as well as in uninfected CD4+ T cells. The intensity of apoptosis correlates with the general state of activation of the immune system and not with the stage of disease or with viral burden. It is likely that nonspecific apoptosis of immunocompetent cells related to immune activation contributes to the immune abnormalities in HIV disease.

AUTOIMMUNE PHENOMENA

The autoimmune phenomena that are common in HIV-infected individuals reflect, at least in part, chronic immune system activation as well as molecular mimicry by viral components. Although these phenomena usually occur in the absence of autoimmune disease, a wide spectrum of clinical manifestations that may be associated with autoimmunity have been described (see "Immunologic and Rheumatologic Diseases," below). Autoimmune phenomena include antibodies to lymphocytes and, less commonly, to platelets and neutrophils. Antiplatelet antibodies have some clinical relevance, in that they may contribute to the thrombocytopenia of HIV disease (see below). Antibodies to nuclear and cytoplasmic components of cells have been reported, as have antibodies to cardiolipin; CD4 molecules; CD43 molecules; C1q-A; variable regions of the T cell receptor α, β, and γ chains; Fas; denatured collagen; and IL-2. In addition, autoantibodies to a range of serum proteins, including albumin, immunoglobulin, and thyroglobulin, have been reported. There is antigenic cross-reactivity between HIV viral proteins (gp120 and gp41) and MHC class II determinants, and anti-MHC class II antibodies have been reported in HIV infection. These antibodies could potentially lead to the elimination of MHC class II–bearing cells via antibody-dependent cellular cytotoxicity (ADCC), although this has not been clearly demonstrated to occur (Chap. 314). In addition, regions of homology exist between HIV envelope glycoproteins and IL-2 as well as MHC class I molecules. The increased occurrence and/or exacerbation of certain autoimmune diseases have been reported in HIV infection; these diseases include psoriasis, idiopathic thrombocytopenic purpura, Graves' disease, antiphospholipid antibody syndrome, and primary biliary cirrhosis. With the widespread use of effective antiretroviral therapy, an *immune reconstitution inflammatory syndrome* (IRIS) has become increasingly common. IRIS is an autoimmune-like phenomenon characterized by a paradoxical deterioration of clinical condition, which is usually compartmentalized to a particular organ system, in individuals in whom cART has recently been initiated. It is associated with a decrease in viral load and at least partial recovery of immune competence, which is usually associated with increases in CD4+ T cell counts. The immunopathogenesis is felt to be related to an increase in immune response against the presence of residual antigens that are usually microbial and is commonly seen with underlying *Mycobacterium tuberculosis* and cryptococcosis. This syndrome is discussed in more detail below.

The Cytokine Network in HIV Pathogenesis

The immune system is homeostatically regulated by a complex network of immunoregulatory cytokines, which are pleiotropic and redundant and operate in an autocrine and paracrine manner. They are expressed continuously, even during periods of apparent quiescence of the immune system. On perturbation of the immune system by antigenic challenge, the expression of cytokines increases to varying degrees (Chap. 314). Cytokines that are important components of this immunoregulatory network have been demonstrated to play a major role in the regulation of HIV expression in vitro. Potent modulation of HIV expression

has been demonstrated either by manipulating endogenous cytokines or by adding exogenous cytokines to culture. Cytokines that induce or enhance HIV expression in one or more of these systems include IL-1, IL-2, IL-3, IL-6, IL-12, IL-18, TNF-α, TNF-β, macrophage colony-stimulating factor (M-CSF), and granulocyte-macrophage colony-stimulating factor (GM-CSF). IL-18 has also been shown to play a role in the development of the HIV-associated lipodystrophy syndrome. Among these cytokines, the most consistent and potent inducers of HIV expression are the *proinflammatory cytokines* TNF-α, IL-1β, and IL-6. Interferon (IFN) α and β as well as IL-32 suppress HIV replication, whereas transforming growth factor (TGF) β, IL-4, IL-10, and IFN-γ can either induce or suppress HIV expression, depending on the system involved. IL-27 suppresses HIV replication by inducing IFN-associated genes. The *CC-chemokines* RANTES (CCL5), macrophage inflammatory protein (MIP) 1α (CCL3), and MIP-1β (CCL4) (Chap. 314) inhibit infection by and spread of R5 HIV-1 strains, while *stromal cell–derived factor* (SDF) 1 inhibits infection by and spread of X4 strains (see below). The alpha defensin family of cytokines has been shown to inhibit both R5 and X4 viruses, and other soluble factors that have not yet been fully characterized have also been shown to suppress HIV replication.

The molecular mechanisms of HIV regulation are best understood for TNF-α, which activates NF-κB proteins that function as transcriptional activators of HIV expression. The HIV-inducing effect of IL-1β is thought to occur at the level of viral transcription in an NF-κB-independent manner. IL-6, GM-CSF, and IFN-γ regulate HIV expression mainly by posttranscriptional mechanisms. Elevated levels of TNF-α and IL-6 have been demonstrated in plasma and cerebrospinal fluid (CSF), and increased expression of TNF-α, IL-1β, IFN-γ, and IL-6 has been demonstrated in the lymph nodes of HIV-infected individuals. The mechanisms whereby the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4) inhibit infection of R5 strains of HIV or SDF-1 blocks X4 strains of HIV involve blocking of the binding of the virus to its co-receptors, the CC-chemokine receptor CCR5 and the CXC-chemokine receptor CXCR4, respectively. However, several CC-chemokines, including but not limited to CCL3, -4, and -5, induce intracellular signals that actually enhance infection by X4 strains of virus at both the entry and postentry levels. The mechanisms whereby other less well characterized factors inhibit HIV replication are not completely understood.

Blocking of endogenous HIV-inducing cytokines or addition of inhibitors of HIV-suppressor cytokines in cultures of peripheral blood and lymph node mononuclear cells from HIV-infected individuals has demonstrated that HIV replication is controlled tightly by endogenous cytokines that act synergistically and in an autocrine and paracrine manner, similar to their physiologic function in the regulation of the immune system. Indeed, the net level of virus replication in an HIV-infected individual reflects at least in part a balance between inductive and suppressive host factors, mediated mainly by cytokines. Finally, the secretion of certain proinflammatory and immunoregulatory cytokines is both a consequence of the aberrant immune activation associated with HIV infection and a mechanism of propagation of the process of aberrant cellular activation (see "Immune Activation, Inflammation, and HIV Pathogenesis," above).

Lymphocyte Turnover in HIV Infection

The immune systems of patients with HIV infection are characterized by a profound increase in lymphocyte turnover that is immediately reduced with effective cART. Studies utilizing in vivo or in vitro labeling of lymphocytes in the S-phase of the cell cycle have demonstrated a tight correlation between the degree of lymphocyte turnover and plasma levels of HIV RNA. This increase in turnover is seen in CD4+ and CD8+ T lymphocytes as well as B lymphocytes and can be observed in peripheral blood and lymphoid tissue. Mathematical models derived from these data suggest that one can view the lymphoid pool as consisting of dynamically distinct subpopulations of cells that are differentially affected by HIV infection. A major consequence of HIV infection appears to be a shift in cells from a more quiescent pool to a pool with a higher turnover rate. It is likely that a consequence of a higher rate of turnover is a higher rate of cell death. The role of the thymus in adult human T cell homeostasis and HIV pathogenesis is an area of controversy. While some data point to an important role for the thymus in maintaining T cell numbers and suggest that impairment of thymic function may be responsible for the declines in CD4+ T cells seen in the setting of HIV infection, other studies have concluded that the thymus plays a minor role in HIV pathogenesis. Among the data supporting an important role for the thymus are those that demonstrate an increase in the levels of T cell receptor excision circles (TRECs) following initiation of cART. TRECs are a byproduct of T cell development and represent episomal fragments of DNA that are excised during T cell receptor gene rearrangement (Chap. 314). Levels of TRECs will be the net result of changes in thymic output together with changes in T cell turnover. An increase in thymic output and/or a decrease in T cell turnover will lead to an increase in levels of TRECs. While it is clear that levels of TRECs increase following initiation of cART, it is not clear whether this is a consequence of increased thymic output or decreased T cell turnover.

The Role of Viral Receptors and Co-Receptors in HIV Pathogenesis

As mentioned above, HIV-1 utilizes two major co-receptors along with CD4 to bind to, fuse with, and enter target cells; these co-receptors are CCR5 and CXCR4, which are also receptors for certain endogenous chemokines. Strains of HIV that utilize CCR5 as a co-receptor are referred to as *R5 viruses*. Strains of HIV that utilize CXCR4 are referred to as *X4 viruses*. Many virus strains are *dual tropic* in that they utilize both CCR5 and CXCR4; these are referred to as *R5X4 viruses*.

The natural chemokine ligands for the major HIV co-receptors can readily block entry of HIV. For example, the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4), which are the natural ligands for CCR5, block entry of R5 viruses, whereas SDF-1, the natural ligand for CXCR4, blocks entry of X4 viruses. The mechanism of inhibition of viral entry is a steric inhibition of binding that is not dependent on signal transduction **(Fig. 189-25)**.

Figure 189-25

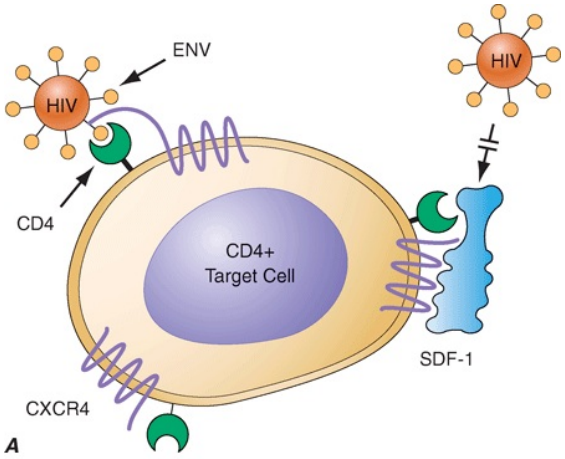

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

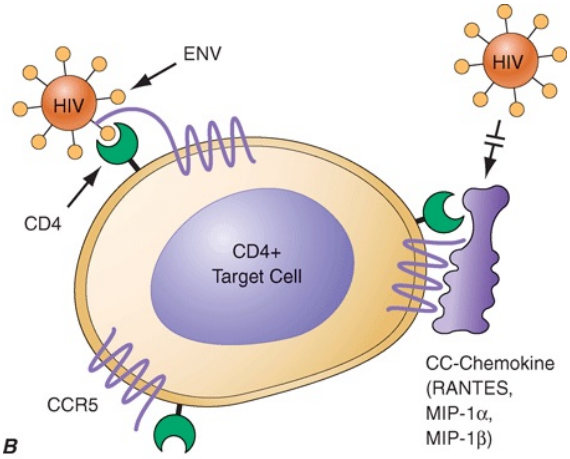

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Model for the role of co-receptors CXCR4 and CCR5 in the efficient binding and entry of X4 (*A*) and R5 (*B*) strains of HIV-1, respectively, into CD4+ target cells. Blocking of this initial event in the virus life cycle can be accomplished by inhibition of binding to the co-receptor by the normal ligand for the receptor in question. The ligand for CXCR4 is stromal cell–derived factor (SDF-1); the ligands for CCR5 are RANTES, MIP-1α, and MIP-1β.

The transmitting virus is almost invariably an R5 virus that predominates during the early stages of HIV disease. In ~40% of HIV-infected individuals, there is a transition to a predominantly X4 virus that is associated with a relatively rapid progression of disease. However, at least 60% of infected individuals progress in their disease while maintaining predominance of an R5 virus. It should be pointed out that clade C viruses, unlike other subgroups, almost never switch from CCR5 tropism to CXCR4 tropism; the reason for this difference is unclear.

The basis for the tropism of different envelope glycoproteins for either CCR5 or CXCR4 relates to the ability of the HIV envelope, including the third variable region (V3 loop) of gp120, to interact with these co-receptors. In this regard, binding of gp120 to CD4 induces a conformational change in gp120 that increases its affinity for CCR5 (see above). Finally, R5 viruses are more efficient in infecting monocytes/macrophages and microglial cells of the brain (see "Neuropathogenesis," below).

The Integrin α4β7 and Mucosal Transmission of HIV

A recently identified receptor for HIV has been reported; this receptor is not necessary for virus binding and fusion to its target CD4+ T cell or for virus replication, but it likely plays an important role in the transmission of HIV at mucosal surfaces such as the genital tract and gut. The integrin α4β7, which is the gut homing receptor for peripheral T cells, binds in its activated form to a specific tripeptide in the V2 loop of gp120 resulting in rapid activation of leukocyte function-associated antigen 1 (LFA-1), the central integrin in the establishment of virologic synapses, which facilitate efficient cell-to-cell spread of HIV. It has been demonstrated that $\alpha4\beta7^{high}$ CD4+ T cells are more susceptible to productive infection than are $\alpha4\beta7^{low\_neg}$ CD4+ T cells because this cellular subset is enriched with metabolically active CD4+ T cells that are $CCR5^{high}$. These cells are present at the gut and genital tract mucosal surfaces. Importantly, it has been demonstrated that the virus that is transmitted during sexual exposure binds much more efficiently to α4β7 than does the virus that diversifies from the transmitting virus over time by mutation, particularly involving the accumulation of glycosylation sites (see "Early Events in HIV Infection: Primary Infection and Initial Dissemination of Virus," above).

Cellular Targets of HIV

Although the CD4+ T lymphocytes and to a lesser extent CD4+ cells of monocyte lineage are the principal targets of HIV, virtually any cell that expresses the CD4 molecule together with co-receptor molecules (see above and below) can potentially be infected with HIV. Circulating dendritic cells have been reported to express low levels of CD4, and, depending on their stage of maturation, these cells can be infected with HIV. Epidermal Langerhans cells express CD4 and have been infected by HIV in vivo, although as has been shown in vivo for dendritic cells, FDCs, and B cells, these cells are more likely to bind and transfer virus to activated CD4+ T cells than to themselves be productively infected.

In vitro, HIV has been reported also to infect a wide range of cells and cell lines that express low levels of CD4, no detectable CD4, or only CD4 mRNA. However, since the only cells that have been shown unequivocally to be infected with HIV and to support replication of the virus are CD4+ T lymphocytes and cells of monocyte/macrophage lineage, the relevance of the in vitro

infection of these other cell types is questionable.

Of potentially important clinical relevance is the demonstration that thymic precursor cells, which were assumed to be negative for CD3, CD4, and CD8 molecules, actually do express low levels of CD4 and can be infected with HIV in vitro. In addition, human thymic epithelial cells transplanted into an immunodeficient mouse can be infected with HIV by direct inoculation of virus into the thymus. Since these cells may play a role in the normal regeneration of CD4+ T cells, it is possible that their infection and depletion contribute, at least in part, to the impaired ability of the CD4+ T cell pool to completely reconstitute itself in certain infected individuals in whom cART has suppressed viral replication to <50 copies of HIV RNA per milliliter (see below). In addition, CD34+ monocyte precursor cells have been shown to be infected in vivo in patients with advanced HIV disease. It is likely that these cells express low levels of CD4, and therefore it is not essential to invoke CD4-independent mechanisms to explain the infection.

Abnormalities of Mononuclear Cells

CD4+ T CELLS

The primary immunopathogenic lesion in HIV infection involves CD4+ T cells, and the range of CD4+ T cell abnormalities in advanced HIV infection is broad. The defects are both quantitative and qualitative and ultimately impact virtually every limb of the immune system, indicating the critical dependence of the integrity of the immune system on the inducer/helper function of CD4+ T cells. In advanced HIV disease, most of the observed immune defects can ultimately be explained by the quantitative depletion of CD4+ T cells. However, T cell dysfunction can be demonstrated in patients early in the course of infection, even when the CD4+ T cell count is in the low-normal range. The degree and spectrum of dysfunctions increase as the disease progresses. One of the first abnormalities to be detected is a defect in response to remote recall antigens, such as tetanus toxoid and influenza, at a time when mononuclear cells can still respond normally to mitogenic stimulation. Indeed, defects of central memory cells are a critical component of HIV immunopathogenesis. The progressive loss of antigen-specific CD4+ T cells has important implications for the control of HIV infection. In this regard, there is a correlation between the maintenance of HIV-specific CD4+ T cell proliferative responses and improved control of infection. Essentially every T cell function has been reported to be abnormal at some stage of HIV infection. Loss of polyfunctional HIV-specific CD4+ T cells, especially those that produce IL-2, occurs early in disease, whereas IFN-producing CD4+ T cells are maintained longer and do not correlate with control of HIV viremia. Loss of IL-2-producing polyfunctional CD4+ T cells is also associated with decreased capacity to upregulate CD40 ligand, which may contribute to the dysregulation of B cell function observed in HIV disease. Other abnormalities include impaired expression of IL-2 receptors, defective IL-2 production, reduced expression of the IL-7 receptor (CD127), and decreased proportion of CD4+ T cells that express CD28, a major co-stimulatory molecule necessary for the normal activation of T cells. Cells lacking expression of CD28 do not respond normally to activation signals and may express markers of terminal activation including HLA-DR, CD38, and CD45RO. As mentioned above ("Immune Activation, Inflammation, and HIV Pathogenesis"), a subset of CD4+ T cells referred to as *T regulatory cells*, or T-regs, may be involved in damping aberrant immune activation that propagates HIV replication. The presence of these T-reg cells correlates with lower viral loads and higher CD4+/CD8+ T cell ratios. A loss of this T-reg capability with advanced disease may be detrimental to the control of virus replication.

It is difficult to explain completely the profound immunodeficiency noted in HIV-infected individuals solely on the basis of direct infection and quantitative depletion of CD4+ T cells. This is particularly apparent during the early stages of HIV disease, when CD4+ T cell numbers may be only marginally decreased. In this regard, it is likely that CD4+ T cell dysfunction results from a combination of depletion of cells due to direct infection of the cell and a number of virus-related but indirect effects on the cell **(Table 189-4)**. Certain of these effects have been demonstrated by exposure of cells to virus in vitro and so their clinical relevance is not completely clear. However, it has been clearly demonstrated that patients with high levels of plasma viremia have a variety of subtle abnormalities of CD4+ T cell function, particularly involving aberrancies in signal transduction pathways. These abnormalities could be due either to aberrant activation induced by the cascade of cytokines that are expressed in viremic patients or to the direct effect of virus on the cell. In this regard, certain of these abnormalities can be reproduced by exposing CD4+ T cells of normal individuals to oligomeric HIV envelope proteins in vitro.

Table 189-4 Mechanisms of CD4+ T Cell Dysfunction and Depletion

Direct Mechanisms	Indirect Mechanisms
Loss of plasma membrane integrity due to viral budding	Aberrant intracellular signaling events
Accumulation of unintegrated viral DNA	Autoimmunity
Interference with cellular RNA processing	Innocent bystander killing of viral antigen–coated cells
Intracellular gp120-CD4 autofusion events	Apoptosis
Syncytia formation	Inhibition of lymphopoiesis
	Activation-induced cell death
	Elimination of HIV-infected cells by virus-specific immune responses

Humoral and cellular immune responses to HIV may contribute to protective immunity by eliminating virus and virus-infected cells (see "Immune Response to HIV," below). However, since the main targets of HIV infection are immunocompetent cells, these responses may contribute to immune cell depletion and immunologic dysfunction by eliminating both infected cells and "innocent bystander" cells. Soluble viral proteins, particularly gp120, can bind with high affinity to the CD4 molecules on uninfected T cells and monocytes; in addition, virus and/or viral proteins can bind to dendritic cells or FDCs. HIV-specific antibody can recognize these bound molecules and potentially collaborate in the elimination of the cells by ADCC.

HIV envelope glycoproteins gp120 and gp160 manifest high-affinity binding to CD4 as well as to various chemokine receptors.Intracellular signals transduced by gp120 through both CD4 and CCR5/CXCR4 have been associated with a number of immunopathogenic processes including anergy, apoptosis, and abnormalities of cell trafficking. The molecular mechanisms responsible for these abnormalities include dysregulation of the T cell receptor–phosphoinositide pathway, p56lck activation, phosphorylation of focal adhesion kinase, activation of the MAP kinase and ras signaling pathways, and downregulation of the co-stimulatory molecules CD40 ligand and CD80.

The inexorable decline in CD4+ T cell counts that occurs in most HIV-infected individuals may result in part from the inability of the immune system to regenerate over an extended period of time the rapidly turning over CD4+ T cell pool efficiently enough to compensate for both HIV-mediated and naturally occurring attrition of cells. In this regard, the degree and duration of decline of CD4+ T cells at the time of initiation of therapy is an important predictor of the restoration of these cells. A person who maintains a very low CD4+ T cell count for a considerable period of time before the initiation of ART almost invariably has an incomplete reconstitution of such cells. At least two major mechanisms may contribute to the failure of the CD4+ T cell pool to reconstitute itself adequately over the course of HIV infection. The first is the destruction of lymphoid precursor cells, including thymic and bone marrow progenitor cells; the other is the gradual disruption of the lymphoid tissue microenvironment, which is essential for

efficient regeneration of immunocompetent cells. Finally, during the advanced stages of CD4+ T lymphopenia, there are increased serum levels of the homeostatic cytokine IL-7. It was initially felt that this elevation was a homeostatic response to the lymphopenia; however, recent findings suggest that the increase in serum IL-7 was a result of reduced utilization of the cytokine related to the loss of cells expressing the IL-7 receptor, CD127, which serves as a normal physiologic regulator of IL-7 production.

CD8+ T CELLS

A relative CD8+ T lymphocytosis is generally associated with high levels of HIV plasma viremia and likely reflects dysregulated homeostasis associated with generalized immune activation. During the late stages of HIV infection, there may be a significant reduction in the numbers of CD8+ T cells despite the presence of high levels of viremia. HIV-specific CD8+ CTLs have been demonstrated in HIV-infected individuals early in the course of disease. The emergence of HIV escape mutants may ultimately evade these HIV-specific CD8+ T cells. However, as the disease progresses, the functional capability of these cells gradually decreases and may be lost entirely. The cause of this loss of cytolytic activity is unclear, although functional impairment is thought to be associated with the persistent nature of HIV infection and disease progression. In this regard, as chronic immune activation persists, CD8+ T cells assume an abnormal phenotype characterized by expression of activation markers such as HLA-DR and CD38 with an absence of expression of the IL-2 receptor (CD25), and a reduced expression of the IL-7 receptor (CD127). In addition, CD8+ T cells lacking CD28 expression are increased in HIV disease, reflecting a skewed expansion of a less differentiated CD8+ T cell subset. This skewing of subsets is also associated with diminished polyfunctionality, a qualitative difference that distinguishes nonprogressors from progressors. It has been reported that nonprogressors can also be distinguished from progressors by the maintenance in the former of a high proliferative capacity of their HIV-specific CD8+ T cells coupled to increases in perforin expression, characteristics that are markedly diminished in advanced HIV disease. It has been reported that the phenotype of CD8+ T cells in HIV-infected individuals may be of prognostic significance. Those individuals whose CD8+ T cells developed a phenotype of HLA-DR+/CD38– following seroconversion had stabilization of their CD4+ T cell counts, whereas those whose CD8+ T cells developed a phenotype of HLA-DR+/CD38+ had a more aggressive course and a poorer prognosis. In addition to the defects in HIV-specific CD8+ CTLs, functional defects in other MHC-restricted CTLs, such as those directed against influenza and CMV, have been demonstrated. CD8+ T cells secrete a variety of soluble factors that inhibit HIV replication including the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4) as well as one or more as yet poorly identified factors. The presence of high levels of HIV viremia in vivo as well as exposure of CD8+ T cells in vitro to HIV envelope, both of which are associated with aberrant immune activation, have been shown to be associated with a variety of cellular functional abnormalities. Furthermore, since the integrity of CD8+ T cell function depends in part on adequate inductive signals from CD4+ T cells, the defect in CD8+ CTLs is likely compounded by the quantitative loss and qualitative dysfunction of CD4+ T cells. Finally, certain cell surface negative regulatory molecules such as CTLA-4 and PD-1 are upregulated on activated T cells, and engagement of these molecules with their ligands may play a role in the exhaustion and death of CD8+, HIV-specific T cells.

B CELLS

The predominant defect in B cells from HIV-infected individuals is one of aberrant cellular activation, which is reflected by spontaneous proliferation and immunoglobulin secretion and by increased spontaneous secretion of TNF-α and IL-6. In addition, B cells from HIV viremic patients manifest a decreased capacity to mount a proliferative response to ligation of the B cell antigen receptor and other B cell stimuli in vitro, yet at the same time they are capable of robust differentiation in vivo as a result of HIV-induced immune activation. B cells from HIV-infected individuals manifest enhanced spontaneous secretion of immunoglobulins in vitro, a process that reflects their highly differentiated state in vivo. There is also an increased incidence of EBV-related B cell lymphomas in HIV-infected individuals that are likely due to combined effects of defective T cell immune surveillance and increased turnover that increases the risk of oncogenesis. Untransformed B cells cannot be infected with HIV, although HIV or its products can activate B cells directly. B cells from patients with high levels of viremia bind virions to their surface via the CD21 complement receptor. It is likely that in vivo activation of B cells by replication-competent or -defective virus as well as viral products during the viremic state accounts at least in part for the spontaneous activation of these cells noted ex vivo. B cell subpopulations from HIV-infected individuals undergo a number of changes over the course of HIV disease, including the attrition of resting memory B cells and replacement with several aberrant memory and differentiated B cell subpopulations that collectively express reduced levels of CD21 and either increased expression of activation markers or inhibitory receptors associated with functional exhaustion. The more activated and differentiated B cells are also responsible for increased secretion of immunoglobulins and increased susceptibility to Fas-mediated apoptosis. In more advanced disease, there is also the appearance of immature B cells associated with CD4+ T cell lymphopenia. Cognate B cell–CD4+ T cell interactions are abnormal in viremic HIV-infected individuals in that B cells respond poorly to CD4+ T cell help and CD4+ T cells receive inadequate co-stimulatory signals from activated B cells. In vivo, the aberrant activated state of B cells manifests itself by hypergammaglobulinemia and by the presence of circulating immune complexes and autoantibodies. HIV-infected individuals respond poorly to primary and secondary immunizations with protein and polysaccharide antigens. Using immunization with influenza vaccine, it has been demonstrated that there is a memory B cell defect in HIV-infected individuals, particularly those with high levels of HIV viremia. Taken together, these B cell defects are likely responsible in part for the decreased response to vaccinations and the increase in certain bacterial infections seen in advanced HIV disease in adults, as well as for the important role of bacterial infections in the morbidity and mortality rates of HIV-infected children, who cannot mount an adequate humoral response to common bacterial pathogens. The absolute number of circulating B cells may be depressed in HIV infection; this phenomenon likely reflects increased activation-induced apoptosis as well as a redistribution of cells out of the circulation and into the lymphoid tissue—phenomena that are associated with ongoing viral replication.

MONOCYTES/MACROPHAGES

Circulating monocytes are generally normal in number in HIV-infected individuals. Monocytes express the CD4 molecule and several co-receptors for HIV on their surface, including CCR5, CXCR4, and CCR3, and thus are targets of HIV infection. The degree of cytopathicity of HIV for cells of the monocyte lineage is low, and HIV can replicate extensively in cells of the monocyte lineage with relatively little cytopathic effect. Hence, monocyte-lineage cells may play a role in the dissemination of HIV in the body and can serve as reservoirs of HIV infection, thus representing an obstacle to the eradication of HIV by antiretroviral drugs. In vivo infection of circulating monocytes is difficult to demonstrate; however, infection of tissue macrophages and macrophage-lineage cells in the brain (infiltrating macrophages or resident microglial cells) and lung (pulmonary alveolar macrophages) can be demonstrated easily. Tissue macrophages are an important source of HIV during the inflammatory response associated with opportunistic infections. Infection of monocyte precursors in the bone marrow may directly or indirectly be responsible for certain of the hematologic abnormalities in HIV-infected individuals. A number of abnormalities of circulating monocytes have been reported in HIV-infected individuals, many of which may be related directly or indirectly to aberrant in vivo immune activation. In this regard, increased levels of lipopolysaccharide (LPS) are found in the sera of HIV-infected individuals due, at least in part, to translocation across the gut mucosal barrier (see above). LPS is a highly inflammatory bacterial product that preferentially binds to macrophages through CD14 and Toll-like receptors, resulting in cellular activation. Increased levels of soluble CD14 in plasma are associated with poor overall clinical outcomes. Monocyte/macrophage functional abnormalities in HIV disease include decreased secretion of IL-1 and IL-12; increased secretion of IL-10; defects in antigen presentation and induction of T cell responses due to decreased MHC class II expression; and abnormalities of Fc receptor function, C3 receptor–mediated clearance, oxidative burst responses, and certain cytotoxic functions such as ADCC, possibly related to low levels of expression

of Fc and complement receptors. Exposure of monocytes in vitro to viral proteins such as gp120 and Tat, as well as to certain cytokines, can cause abnormal activation, and this may play a role in cellular dysfunction.

DENDRITIC AND LANGERHANS CELLS

Dendritic cells (DCs) may play an important role in the initiation of HIV infection by virtue of the ability of HIV to bind to cell-surface C-type lectin receptors, particularly DC-SIGN (see above). This allows efficient presentation of virus to CD4+ T cell targets that become infected; complexes of infected CD4+ T cells and DCs provide an optimal microenvironment for virus replication. There has been considerable disagreement regarding the HIV infectibility and hence the depletion as well as the dysfunction of DCs themselves. The situation has recently been clarified by the recognition that DCs can be classified into myeloid (mDC) and plasmacytoid (pDC) subsets, leading to an appreciation of specific DC dysfunction in HIV disease. pDCs are an important component of the innate immune system and secrete large amounts of IFN-α in response to viral infections. The numbers of circulating pDCs are decreased in HIV infection through mechanisms that remain unclear. It has recently been demonstrated that HIV gp120 interacts directly with pDCs and interferes with TLR9 activation, resulting in a decreased ability of pDCs to secrete antiviral and inflammatory factors that play a role in immune responses against invading pathogens.

NATURAL KILLER CELLS

The role of NK cells is to provide immunosurveillance against virus-infected cells, certain tumor cells, and allogeneic cells (Chap. 314). There are no convincing data that HIV productively infects NK cells in vivo; however, functional abnormalities in NK cells have been observed throughout the course of HIV disease, and the severity of these abnormalities increases as disease progresses. In addition, it has been reported that HIV envelope induces aberrant signaling in NK cells that increases their susceptibility to apoptosis. HIV infection of target cells downregulates HLA-A and -B, but not HLA-C and -D molecules; this may explain in part the relative inability of NK cells to kill HIV-infected target cells. Most studies report that NK cells are normal in number; however, patients with high levels of virus replication manifest an abnormal representation of a functionally defective CD56–/CD16+ NK cell subset. This abnormal subset of NK cells manifests an increased expression of inhibitory NK cell receptors (iNKRs) and a substantial decrease in expression of natural cytotoxicity receptors (NCRs) and shows a markedly impaired lytic activity. The overrepresentation of this abnormal subset of NK cells may explain in part the observed defects in NK cell function in HIV-infected individuals. NK cells also serve as important sources of HIV-inhibitory CC-chemokines. NK cells isolated from HIV-infected individuals constitutively produce high levels of MIP-1α (CCL3), MIP-1β (CCL4), and RANTES (CCL5). In addition, high levels of these chemokines are seen when NK cells are stimulated with IL-2 or IL-15 or when CD16 is cross-linked or during the process of lytic killing of target cells. HIV-infected patients with high levels of plasma viremia manifest a decreased ability, compared with HIV-infected individuals who are aviremic, of their NK cells to block HIV replication in vitro in assays of both cell contact and supernatant-mediated suppression of virus. Finally, NK cell–dendritic cell interactions are important for normal immune function. NK cells and dendritic cells reciprocally modulate each other's activation and maturation. These interactions are markedly impaired in HIV-infected individuals with high levels of plasma viremia.

Genetic Factors in HIV Pathogenesis

Genetic association studies have served as a powerful means to identify host factors that influence HIV-AIDS pathogenesis in vivo. Polymorphisms in several genes have now been identified that influence several phenotypes relevant to HIV infection: risk of acquiring HIV, rates of disease progression, long-term nonprogression, spontaneous virologic control, and immunologic responses following initiation of ART. These include polymorphisms in genes in the MHC locus, chemokine receptors and chemokines, cytokines, and other host factors **(Table 189-5)**. Recent studies have capitalized on genome-wide association studies to identify novel genetic factors that influence HIV disease progression rates, and rapid progress is anticipated in this area. Moreover, in vitro genome-wide functional scanning using RNA interference techniques suggests that up to hundreds of host factors may be involved in the HIV replication life cycle. Theoretically, variations in all these genes could impact HIV susceptibility and/or disease progression. Below is a discussion of some representative genes.

Table 189-5 Host Genetic Factors that Influence Risk of HIV-1 Acquisition and Rates of HIV-1 Disease Progression

Gene[a]	Genetic Variation	Mechanisms[b]	Genetic Effect on HIV-AIDS[c]
Genes in MHC Locus			
HLA-B	B*27 and B*57	Presentation of specific immunogenic HIV antigens	Slow progression to AIDS; low viral load
	B*35Px	Restriction of specific immunogenic HIV peptide presentation	Fast progression to AIDS; high viral load
	HLA-Bw4	Providing ligands for activating KIR	Slow progression to AIDS
HLA class I allele	Homozygosity of HLA-A, B, C alleles	Reduced epitope recognition repertoire	Faster progression to AIDS; increased risk of mother-to-child transmission
	Shared donor-recipient HLA alleles	Preadaptation of HIV strains	Faster disease progression
	Rare HLA alleles	Limited adaptation of HIV strains; less frequent escape mutants	Protection against HIV infection
HLA extended haplotype	A1-B8-DR3-DQ2 (8.1)	Unknown; may influence immunologic hyperresponsiveness	Faster progression to AIDS
HLA-C	rs9264942-C	Increased expression of HLA-C; association may be due to linkage disequilibrium with HLA-B57	Decreased viral load set point
HCP5	rs2395029-G	Linkage disequilibrium with HLA-B*5701	Reduced viremia
ZNRD1	rs9261174-C	Possible interference in processing of HIV transcripts; influences ZNRD1 expression; linkage disequilibrium with *HLA-A10*	AIDS disease retardation
Chemokine Receptors			
CCR5	32-bp deletion in the ORF (δ32)	Truncated CCR5 protein	δ32/δ32: resistance to acquiring HIV infection δ32/wt: Delays AIDS onset; improves immune reconstitution during ART

	Promoter SNPs/haplotypes (HHA to HHG*2)	Altered CCR5 expression, e.g., HHE allele correlates with high CCR5 expression	HHE/HHE genotype associates with increased HIV/AIDS susceptibility
CCR2	Valine to isoleucine change (64 V→I)	Possibly due to linkage with polymorphism in *CCR5* promoter	Delayed AIDS onset
CX3CR1	SNPs in ORF (*249 V→ I*, *280 T → M*)	280M reduces receptor expression and binding of fractalkine, the CX3CR1 ligand	249I and 280M are associated with faster AIDS onset in some Caucasian cohorts; inconsistent effects were detected in other cohorts
DARC	African-specific promoter SNP (*46T→ C)*	–46C/C associates with low neutrophil counts; influences circulating chemokine levels; alters HIV binding to RBCs and transinfection of HIV-1	–46C/C: increased risk of acquiring HIV but slow HIV disease progression; the disease-retarding effects associated with –46C/C occur mainly in those HIV+ African Americans who are also leukopenic
Chemokines			
CCL3L, CCL4L	Gene copy number of *CCL3L* and *CCL4L*	High gene copies correlate with high CCL3L and CCL4L levels	Gene copy number lower than population median associates with increased HIV/AIDS susceptibility and reduced immune reconstitution during ART
CCL5	Promoter SNPs	Altered gene expression	Altered HIV-AIDS susceptibility
CCL2	Promoter SNP (–2578 T → G)	*–2578G* allele*:* increased CCL2 expression and monocyte recruitment	–2578G/G associates with increased risk of developing HIV-1-associated dementia and a rapid AIDS onset
Cytokines			
IL-6	Promoter SNP (–174 G → C)	–174C associates with altered IL-6 and CRP levels	Altered risk of KS development and variable recovery of CD4 cells during ART
IL-10	Promoter SNP	–592A results in decreased IL-10 levels	Increased HIV-AIDS susceptibility
Innate Immunity Genes			
MBL	Coding alleles (O)	Low plasma concentration and structural damage of MBL	Slow progression to AIDS in heterozygous subjects (A/O)
	X allele (promoter SNP –221)	Decreased levels of MBL	Faster progression to AIDS in homozygous X/X subjects
Apobec-3G	ORF SNP (186 H → R)	Reduced anti-HIV-1 activity	186R associates with rapid AIDS onset in African Americans
Others			
ApoE	E4, E3, E2 allele	E4 enhances HIV cell entry in vitro	ApoE4/E4 associates with rapid AIDS onset and dementia
Gene–Gene Interaction			
KIR+HLA	KIR3DS1 with HLA Bw4-80I + or 80I –	Altered NK cell activity required to kill HIV-infected cells	KIR3DS1 with HLA Bw4-80I +: delayed AIDS onset KIR3DS1 with HLA Bw4-80I –: rapid AIDS onset
CCL3L1 + CCR5	Low *CCL3L1* gene copies + detrimental *CCR5* genotypes	Low CCL3L1 and high CCR5 expression	Increased HIV/AIDS susceptibility and reduced immune reconstitution during ART

[a]Representative genes and polymorphisms and [b]possible mechanisms are listed. [c]Some of the associations are population specific and may display cohort-specific effects.

Note: Apobec, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; ApoE, apolipoprotein E; ART, antiretroviral therapy; CCL, CC ligand; CCL3L, CCL3-like; CCR5, CC chemokine receptor 5; CRP, C-reactive protein; DARC, Duffy antigen receptor for chemokines; HCP5, HLA class I histocompatibility antigen protein P5; HHE, human haplogroup E; HLA, human leukocyte antigen; IL, interleukin; KIR, killer cell immunoglobulin-like receptors; KS, Kaposi's sarcoma; MBL, mannose-binding lectin; MHC, major histocompatibility complex; ORF, open reading frame; SNP, single nucleotide polymorphism; VL, viral load; wt, wild-type; ZNRD1, zinc ribbon domain containing 1; +, present, –, absent.

Sources: Sunil K. Ahuja, MD, Weijing He, MD, and *www.hiv-pharmacogenomics.org*. Reviews for additional information: P An et al: Trends Genet 26:119, 2010; J Fellay: Antivir Ther 14:731, 2009; RA Kaslow et al: J Infect Dis 191:S68, 2005.

Researchers recently employed a genome-wide association strategy and identified polymorphisms within *HLA-B* (e.g., *HCP-5* gene) and *HLA-C* that explained approximately 15% of the variation in viral load among individuals during the asymptomatic period of infection. A number of mechanisms have been proposed whereby MHC-encoded molecules might predispose an individual either to rapid progression or to nonprogression to AIDS. These proposed mechanisms include the ability to present certain immunodominant HIV T helper or CTL epitopes, leading to a relatively protective immune response against HIV and hence to a slower rate of disease progression. In contrast, certain MHC class I or class II alleles might predispose an individual to an immunopathogenic response against viral epitopes in certain tissues, such as the CNS or lungs, or against certain HIV-infected cell types, such as macrophages or dendritic cells/Langerhans cells. In addition, certain rare MHC class I and class II alleles might facilitate rapid recognition of HIV-infected cells from the infecting partner in primary HIV infection and promote rejection of these cells by alloreactive responses. Similarly, common MHC alleles could lead to less effective removal of HIV-infected allogeneic cells. In this regard, it has been demonstrated that allele sharing at *HLA-B* locus is associated with increased risk of transmission of HIV infection between heterosexual Zambian couples discordant for HIV. It has also been demonstrated that *HLA* heterozygosity for class I loci (A, B, and C) is associated with a delayed onset of AIDS among HIV-infected individuals, whereas homozygosity for these loci is associated with a more rapid progression to AIDS and death. This observation is likely due to the fact that individuals who are heterozygous at *HLA* loci are able to present a greater variety of antigenic peptides to cytotoxic T lymphocytes than are homozygotes, resulting in a more effective immune response against a number of pathogens including HIV. Of particular note is the fact that the *HLA* class I alleles B*35 and Cw*04 were consistently associated with rapid development of AIDS. Other data have indicated that transporter associated with antigen-processing (*TAP*) genes play a role in determining the outcome of HIV infection. HLA profiles that reflect certain combinations of MHC-encoded *TAP* and class I and class II genes are strongly associated with different rates of progression to AIDS. It is noteworthy that the extended *HLA* haplotype 8.1 (A1-B8-DR3) has also been consistently correlated with a rapid decline in CD4+ T cells and development of HIV-related symptoms.

Recent genetic association studies have also highlighted the role for NK cells in HIV disease. A single nucleotide polymorphism (SNP) in the killer immunoglobulin-like receptor (*KIR*) gene was shown to be strongly associated with rapid progression to AIDS. However, when KIR3DS1 was present with HLA-Bw4-80I, the resultant phenotype was delayed progression to AIDS, even though this *HLA-B* allele alone has no effect on HIV disease progression. Furthermore, KIR3DS1/HLA-Bw4-80I-carrying individuals had a significantly reduced viral load, beginning early in the course of infection, and protection against opportunistic infections during the later stages of the disease. This observation points to the potential role of NK cells in the maintenance of the viral set point and strongly suggests that HLA-Bw4-80I serves as the ligand activating the KIR, resulting in the death of the target cell. These gene-gene interactions between *KIR* and MHC genes are illustrated in Table 189-5.

The most dramatic example of a genetic factor influencing HIV infection and/or pathogenesis relates to the gene that encodes for CC chemokine receptor 5 (CCR5), the major HIV co-receptor for cell entry. There are reports of rare individuals who have remained uninfected despite repetitive sexual exposure to HIV in high-risk situations (e.g., commercial sex workers). The peripheral blood mononuclear cells of two such individuals were found to be highly resistant to infection in vitro with R5 strains of HIV-1 but were readily infected with X4 strains. Genetic analysis revealed that these two individuals inherited a homozygous defect in the gene that encodes for CCR5. The defective *CCR5* allele contained a 32-bp deletion corresponding to the second extracellular loop of the receptor (*Δ32* allele). The encoded protein is severely truncated and is not expressed on the cell surface; it is therefore nonfunctional, explaining the refractoriness to infection with R5 strains of HIV-1. Population studies revealed that ~1% of the Caucasian population of western European ancestry possessed the homozygous defect for the *CCR5 Δ32* allele, and subjects with this genotype are highly resistant to HIV infection. A number of studies have found that the frequency of *CCR5 Δ32* allele was enriched in exposed, uninfected individuals of European descent. It is noteworthy that several individuals have been identified as homozygous for the *CCR5 Δ32* defect and in fact did become infected with HIV. These individuals were found to have an X4 strain of HIV that was associated in some cases with an accelerated disease course. X4 strain uses CXCR4 as the co-receptor for cell entry instead of CCR5. Up to 20% of individuals of European descent are heterozygous for the *CCR5 Δ32* allele and display partial resistance to acquiring HIV and a delayed disease course. Cohort studies of hundreds of DNA samples originating from western and central Africa and Far East Asia indicate that the *CCR5 Δ32* allele is either absent or extremely rare in these populations.

A number of SNPs in the *CCR5* promoter have been associated with varied rates of disease progression and altered risk of acquiring HIV. The promoter SNPs along the *CCR5 Δ32* and CCR2-V64I alleles define nine *CCR5* human haplogroups (HH) designated as HHA through HHE, HHF*1, HHF*2, HHG*1 and HHG*2 (Table 189-5). Studies have shown that homozygosity for the *CCR5* HHE haplotype is associated with an increased risk of acquiring HIV and progressing rapidly to AIDS. Pairing of the HHC and the *CCR5 Δ32*-containing HHG*2 haplotype is associated with a slower rate of disease progression and reduced risk of acquiring HIV. Heterozygosity for the *CCR2-64I* polymorphism associates with a slow rate of HIV disease course. This *CCR2-64I* allele–associated effect could be due to its linkage with SNPs in the *CCR5* promoter that are known to influence disease progression rates and/or due to dimerization of CXCR4 with the mutated *CCR2-64I*, resulting in a decreased expression of CXCR4 on the cell surface. Variations in the ligands of CCR5 (e.g., copy number of *CCL3L* genes) and CCR2 (e.g., *CCL2*) may also influence HIV-AIDS susceptibility.

Neuropathogenesis

While there has been a remarkable decrease in the incidence of HIV encephalopathy among those with access to treatment in the era of effective cART, HIV-infected individuals can still experience a variety of neurologic abnormalities due either to opportunistic infections and neoplasms or to direct effects of HIV or its products. With regard to the latter, HIV has been demonstrated in the brain and CSF of infected individuals with and without neuropsychiatric abnormalities. The main cell types that are infected in the brain in vivo are the perivascular macrophages and the microglial cells; monocytes that have already been infected in the blood can migrate into the brain, where they then reside as macrophages, or macrophages can be directly infected within the brain. The precise mechanisms whereby HIV enters the brain are unclear; however, they are thought to relate, at least in part, to the ability of virus-infected and immune-activated macrophages to induce adhesion molecules such as E-selectin and vascular cell adhesion molecule 1 (VCAM-1) on brain endothelium. Other studies have demonstrated that HIV gp120 enhances the expression of intercellular adhesion molecule 1 (ICAM-1) in glial cells; this effect may facilitate entry of HIV-infected cells into the CNS. Virus isolates from the brain are preferentially R5 strains as opposed to X4 strains; in this regard, HIV-infected individuals who are heterozygous for *CCR5-Δ32* appear to be relatively protected against the development of HIV encephalopathy compared with wild-type individuals. Distinct HIV envelope sequences are associated with the clinical expression of the AIDS dementia complex (see below). There is no convincing evidence that brain cells other than those of monocyte/macrophage lineage can be productively infected in vivo.

HIV-infected individuals may manifest white matter lesions as well as neuronal loss. Given the absence of evidence of HIV infection of neurons either in vivo or in vitro, it is highly unlikely that direct infection of these cells accounts for their loss. Rather, the HIV-mediated effects on neurons and oligodendrocytes are thought to involve indirect pathways whereby viral proteins, particularly gp120 and Tat, trigger the release of endogenous neurotoxins from macrophages and to a lesser extent from astrocytes. In addition, it has been demonstrated that both HIV-1 Nef and Tat can induce chemotaxis of leukocytes, including monocytes, into the CNS. Neurotoxins can be released from monocytes as a consequence of infection and/or immune activation. Monocyte-derived neurotoxic factors have been reported to kill neurons via the *N*-methyl-d-aspartate (NMDA) receptor. In addition, HIV gp120 shed by virus-infected monocytes could cause neurotoxicity by antagonizing the function of vasoactive intestinal peptide (VIP), by elevating intracellular calcium levels, and by decreasing nerve growth factor levels in the cerebral cortex. A variety of monocyte-derived cytokines can contribute directly or indirectly to the neurotoxic effects in HIV infection; these include TNF-α, IL-1, IL-6, TGF-β, IFN-γ, platelet-activating factor, and endothelin. Furthermore, among the CC-chemokines, elevated levels of monocyte chemotactic protein (MCP) 1 in the brain and CSF have been shown to correlate best with the presence and degree of HIV encephalopathy. In addition, infection and/or activation of monocyte-lineage cells can result in increased production of eicosanoids, quinolinic acid, nitric oxide, excitatory amino acids such as l-cysteine and glutamate, arachidonic acid, platelet activating factor, free radicals, TNF-α, and TGF-β, which may contribute to neurotoxicity. Astrocytes may play diverse roles in HIV neuropathogenesis. Reactive gliosis or astrocytosis has been demonstrated in the brains of HIV-infected individuals, and TNF-α and IL-6 have been shown to induce astrocyte proliferation. In addition, astrocyte-derived IL-6 can induce HIV expression in infected cells in vitro. Furthermore, it has been suggested that astrocytes may downregulate macrophage-produced neurotoxins. It has been reported that HIV-infected individuals with the E4 allele for apolipoprotein E (apo E) are at increased risk for AIDS encephalopathy and peripheral neuropathy. The likelihood that HIV or its products are involved in neuropathogenesis is supported by the observation that neuropsychiatric abnormalities may undergo remarkable and rapid improvement upon the initiation of cART.

It has also been suggested that the CNS may serve as a relatively sequestered site for a reservoir of latently infected cells that might be a barrier for the eradication of virus by cART (see "Reservoirs of HIV-Infected Cells: Obstacles to the Eradication of Virus," above).

Pathogenesis of Kaposi's Sarcoma

There are at least four distinct epidemiologic forms of KS: (1) the classic form that occurs in older men of predominantly

Mediterranean or eastern European Jewish backgrounds with no recognized contributing factors; (2) the equatorial African form that occurs in all ages, also without any recognized precipitating factors; (3) the form associated with organ transplantation and its attendant iatrogenic immunosuppressed state; and (4) the form associated with HIV-1 infection. In the latter two forms, KS is an opportunistic disease; in HIV-infected individuals, unlike typical opportunistic infections, its occurrence is not strictly related to the level of depression of CD4+ T cell counts. The pathogenesis of KS is complex; fundamentally, it is an angioproliferative disease that is not a true neoplastic sarcoma, at least not in its early stages. It is a manifestation of excessive proliferation of spindle cells that are believed to be of vascular origin and have features in common with endothelial and smooth-muscle cells. In HIV disease the development of KS is dependent on the interplay of a variety of factors including HIV-1 itself, human herpes virus 8 (HHV-8), immune activation, and cytokine secretion. A number of epidemiologic and virologic studies have clearly linked HHV-8, which is also referred to as *Kaposi's sarcoma–associated herpesvirus* (KSHV), to KS not only in HIV-infected individuals but also in individuals with the other forms of KS. HHV-8 is a γ-herpesvirus related to EBV and herpesvirus saimiri. It encodes a homologue to human IL-6 and, in addition to KS, has been implicated in the pathogenesis of body cavity lymphoma, multiple myeloma, and monoclonal gammopathy of undetermined significance. Sequences of HHV-8 are found universally in the lesions of KS, and patients with KS are virtually all seropositive for HHV-8. HHV-8 DNA sequences can be found in the B cells of 30–50% of patients with KS and 7% of patients with AIDS without clinically apparent KS.

Between 1 and 2% of eligible blood donors are positive for antibodies to HHV-8, while the prevalence of HHV-8 seropositivity in HIV-infected men is 30–35%. The prevalence of HHV-8 seropositivity in HIV-infected women is ~4%. This finding is reflective of the lower incidence of KS in women. It has been debated whether HHV-8 is actually the transforming agent in KS; the bulk of the cells in the tumor lesions of KS are not neoplastic cells. However, it has been demonstrated that endothelial cells can be transformed in vitro by HHV-8. In this regard, HHV-8 possesses a number of genes, including homologues of the IL-8 receptor, Bcl-2, and cyclin D, that can potentially transform the host cell. Despite the complexity of the pathogenic events associated with the development of KS in HIV-infected individuals, HHV-8 is the etiologic agent of this disease. The initiation and/or propagation of KS requires an activated state and is mediated, at least in part, by cytokines. A number of factors, including TNF-α, IL-1β, IL-6, GM-CSF, basic fibroblast growth factor, and oncostatin M, function in an autocrine and paracrine manner to sustain the growth and chemotaxis of the KS spindle cells. In this regard, KSHV-derived IL-6 has been demonstrated to induce proliferation of lymphoma cells and to inhibit the cytostatic effects of INF-α on KSHV-infected lymphoma cells.

IMMUNE RESPONSE TO HIV

As detailed above and below, following the initial burst of viremia during primary infection, HIV-infected individuals mount robust immune responses that in most cases substantially curtail the levels of plasma viremia and likely contribute to delaying the ultimate development of clinically apparent disease for a median of 10 years in untreated individuals. This immune response contains elements of both humoral and cell-mediated immunity involving both innate and adaptive immune responses **(Table 189-6**; **Fig. 189-26)**. It is directed against multiple antigenic determinants of the HIV virion as well as against viral proteins expressed on the surface of infected cells. Ironically, those CD4+ T cells with T cell receptors specific for HIV are theoretically those CD4+ T cells most likely to be activated—and thus to serve as early targets for productive HIV infection and the cell death or dysfunction associated with infection. Thus, an early consequence of HIV infection is interference with and decrease of the helper T cell population needed to generate an effective immune response.

Table 189-6 Elements of the Immune Response to HIV

Humoral immunity
Binding antibodies
Neutralizing antibodies
Type specific
Group specific
Antibodies participating in antibody-dependent cellular cytotoxicity (ADCC)
Protective
Pathogenic (bystander killing)
Enhancing antibodies
Complement
Cell-mediated immunity
Helper CD4+ T lymphocytes
Class I MHC–restricted cytotoxic CD8+ T lymphocytes
CD8+ T cell–mediated inhibition (noncytolytic)
ADCC
Natural killer cells

Abbreviation: MHC, major histocompatibility complex.

Figure 189-26

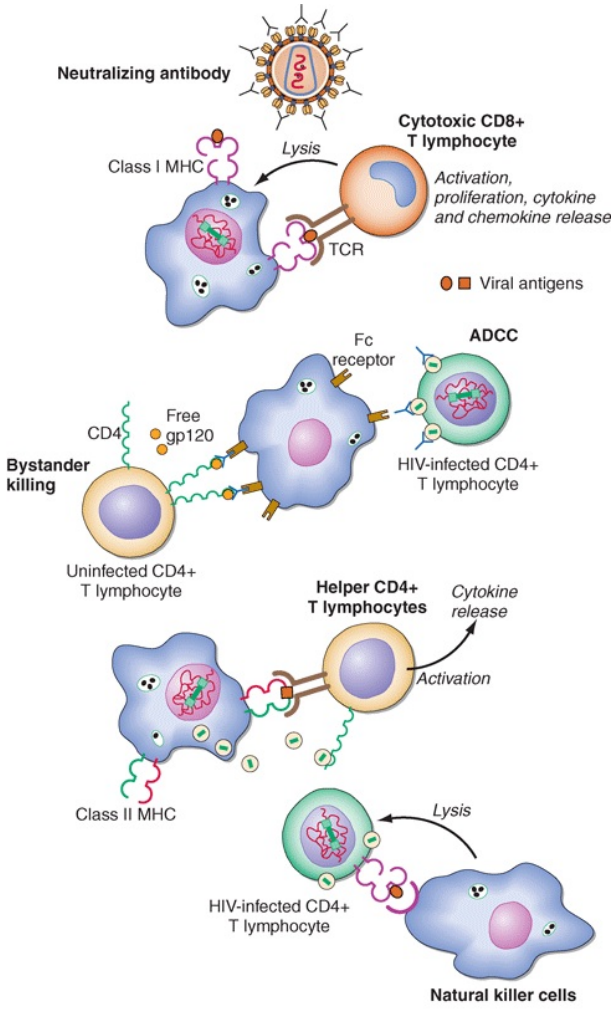

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of the different immunologic effector mechanisms thought to be active in the setting of HIV infection. Detailed descriptions are given in the text. ADCC, antibody-dependent cellular cytotoxicity; MHC, major histocompatibility complex; TCR, T cell receptor.

Although a great deal of investigation has been directed toward delineating and better understanding the components of this immune response, it remains unclear which immunologic effector mechanisms are most important in delaying progression of infection and which, if any, play a role in the pathogenesis of HIV disease. This lack of knowledge has also hampered the ability to develop an effective vaccine for HIV disease.

Humoral Immune Response

Antibodies to HIV usually appear within 3–6 weeks and almost invariably within 12 weeks of primary infection **(Fig. 189-27)**; rare exceptions are individuals who have defects in the ability to produce HIV-specific antibodies. Detection of these antibodies forms the basis of most diagnostic screening tests for HIV infection. The appearance of HIV-binding antibodies detected by ELISA and Western blot assays occurs prior to the appearance of neutralizing antibodies; the latter generally appear following the initial decreases in plasma viremia, which is more closely related to the appearance of HIV-specific CD8+ T lymphocytes. The first antibodies detected are those directed against the immunodominant region of the envelope gp41, followed by the appearance of antibodies to the structural or gag protein p24 and the gag precursor p55. Antibodies to p24 gag are followed by the appearance of antibodies to the outer envelope glycoprotein (gp120), the gag protein p17, and the products of the *pol* gene (p31 and p66). In addition, one may see antibodies to the low-molecular-weight regulatory proteins encoded by the HIV genes *vpr*, *vpu*, *vif*, *rev*, *tat*, and *nef*. On rare occasion, levels of HIV-specific antibodies may decline during treatment of acute HIV infection.

Figure 189-27

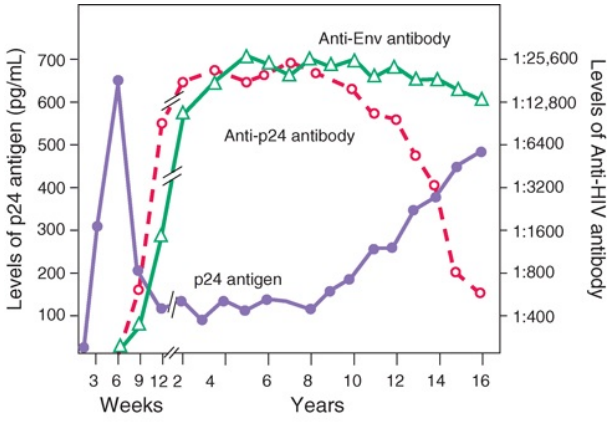

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship between antigenemia and the development of antibodies to HIV. Antibodies to HIV proteins are generally seen 6–12 weeks following infection and 3–6 weeks after the development of plasma viremia. Late in the course of illness, antibody levels to p24 decline, generally in association with a rising titer of p24 antigen.

While antibodies to multiple antigens of HIV are produced, the precise functional significance of these different antibodies is unclear. The only viral proteins that elicit neutralizing antibodies are the envelope proteins gp120 and gp41. Antibodies directed toward the envelope proteins of HIV have been characterized both as being protective and as possibly contributing to the pathogenesis of HIV disease. Among the protective antibodies are those that function to neutralize HIV directly and prevent the spread of infection to additional cells, as well as those that participate in ADCC. Within the first 6 months of infection, neutralizing antibodies appear; however, the virus quickly escapes these neutralizing antibodies. One of the principal mechanisms of immune escape is the addition of N-linked glycosylation sites. The added carbohydrate moieties interfere with envelope recognition by these initial antibodies. The hyperglycosylation of the envelope protein has been termed the *glycan shield*. Neutralizing antibodies appear to be of two forms, type-specific and group-specific. *Type-specific neutralizing antibodies* are generally directed to the V3 loop region. These antibodies appear early after infection and are generally directed toward linear epitopes within the V2 and V3 variable regions of gp120. They neutralize only viruses of a given strain and are present in low titer in most infected individuals. *Group-specific neutralizing antibodies* appear later in infection and are capable of neutralizing a wide variety of HIV isolates. At least two forms of group-specific antibodies have been identified: those directed toward the CD4 binding site (CD4bs) of gp120, and those binding to the membrane-proximal region of gp41 **(Fig. 189-28)**. The other major class of protective antibodies are those that participate in ADCC, which is actually a form of cell-mediated immunity (Chap. 314) in which NK cells that bear Fc receptors are armed with specific anti-HIV antibodies that bind to the NK cells via their Fc portion. These armed NK cells then bind to and destroy cells expressing HIV antigens. Antibodies to both gp120 and gp41 have been shown to participate in ADCC-mediated killing of HIV-infected cells. The levels of antienvelope antibodies capable of mediating ADCC are highest in the earlier stages of HIV infection. In vitro, IL-2 can augment ADCC-mediated killing.

Figure 189-28

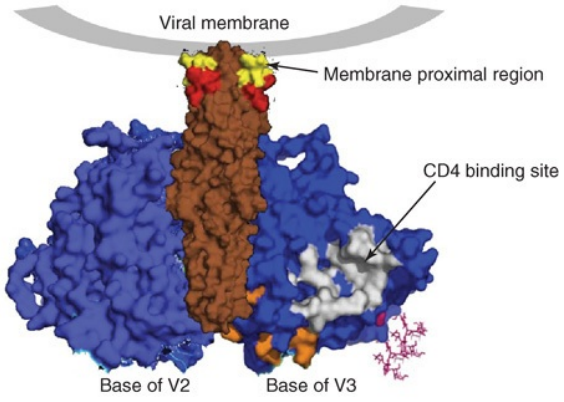

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Known targets of neutralizing antibodies against HIV-1. Group-specific antibody-binding regions include the membrane proximal region and the CD4 binding site. Type-specific antibody-binding regions include V2 and V3 loops. *(Adapted from DR Burton et al: Nat Immunol 5:233, 2004.)*

In addition to playing a role in host defense, HIV-specific antibodies have also been implicated in disease pathogenesis. Antibodies directed to gp41, when present in low titer, have been shown in vitro to be capable of facilitating infection of cells through an Fc receptor–mediated mechanism known as *antibody enhancement*. Thus, the same regions of the envelope protein of HIV that give rise to antibodies capable of mediating ADCC also elicit the production of antibodies that can facilitate infection of cells in vitro. In addition, it has been postulated that anti-gp120 antibodies that participate in the ADCC killing of HIV-infected cells might also kill uninfected CD4+ T cells if the uninfected cells had bound free gp120, a phenomenon referred to as *bystander killing* . One of the most primitive components of the humoral immune system is the complement system (Chap. 314). This element of innate immunity consists of ~30 proteins that are found circulating in blood or associated with cell membranes. While HIV alone is capable of directly activating the complement cascade, the resulting lysis is weak due to the presence of host cell regulatory proteins captured in the virion envelope during budding. It is possible that complement-opsonized HIV virions have increased infectivity in a manner analogous to antibody-mediated enhancement.

Cellular Immune Response

Given the fact that T cell–mediated immunity is known to play a major role in host defense against most viral infections (Chap. 314), it is generally thought to be an important component of the host immune response to HIV. T cell immunity can be divided into two major categories: that mediated by *helper/inducer CD4+ T cells* and that mediated by *cytotoxic/immunoregulatory CD8+ T cells*.

HIV-specific CD4+ T cells can be detected in the majority of HIV-infected patients through the use of flow cytometry to measure intracellular cytokine production in response to MHC class II tetramers pulsed with HIV peptides or through lymphocyte proliferation assays utilizing HIV antigens such as p24. These cells likely play a critical role in the orchestration of the immune response to HIV by providing help to HIV-specific B cells and CD8+ T cells. They may also be capable of directly killing HIV-infected cells. HIV-specific CD4+ T cells may be preferential targets of HIV infection by HIV-infected antigen-presenting cells during the generation of an immune response to HIV (Fig. 189-26). However, they also are likely to undergo clonal expansions in response to HIV antigens and thus survive as a population of cells. No clear correlations exist between levels of HIV-specific CD4+ T lymphocytes and plasma HIV RNA levels; however, in the setting of high viral loads, CD4+ T cell responses to HIV antigens appear to shift from one of proliferation and IL-2 production to one of IFN-γ production. Thus, while a reverse correlation exists between the level of p24-specific proliferation and levels of plasma HIV viremia, the nature of the causal relationship between these parameters is unclear.

MHC class I–restricted, HIV-specific CD8+ T cells have been identified in the peripheral blood of patients with HIV-1 infection. These cells include CTLs that produce perforins and T cells that can be induced by HIV antigens to express an array of cytokines such as IFN-γ, IL-2, and TNF-α. CTLs have been identified in the peripheral blood of patients within weeks of HIV infection and prior to the appearance of plasma virus. The selective pressure they exert on the evolution of the population of circulating viruses reflects their potential role in control of HIV infection. These CD8+ T lymphocytes, through their HIV-specific antigen receptors, bind to and cause the lytic destruction of target cells bearing autologous MHC class I molecules presenting HIV antigens. Two types of CTL activity can be demonstrated in the peripheral blood or lymph node mononuclear cells of HIV-infected individuals. The first type directly lyses appropriate target cells in culture without prior in vitro stimulation (*spontaneous CTL activity*). The other type of CTL activity reflects the *precursor frequency of CTLs* (CTLp); this type of CTL activity can be demonstrated by stimulation of CD8+ T cells in vitro with a mitogen such as phytohemagglutinin or anti-CD3 antibody.

In addition to CTLs, CD8+ T cells capable of being induced by HIV antigens to express cytokines such as IFN-γ also appear in the setting of HIV-1 infection. It is not clear whether these are the same or different effector pools compared with those cells mediating cytotoxicity; in addition, the relative roles of each in host defense against HIV are not fully understood. It does appear that these CD8+ T cells are driven to in vivo expansion by HIV antigen. There is a direct correlation between levels of CD8+ T cells capable of producing IFN-γ in response to HIV antigens and plasma levels of HIV-1 RNA. Thus, while these cells are clearly induced by HIV-1 infection, their overall ability to control infection remains unclear. Multiple HIV antigens, including Gag, Env, Pol, Tat, Rev, and Nef, can elicit CD8+ T cell responses. Among patients who control viral replication in the absence of antiretroviral drugs are a subset of patients referred to as elite nonprogressors (see "Long-Term Survivors and Long-Term Nonprogressors," above) whose peripheral blood contains a population of CD8+ T cells that undergo substantial proliferation and perforin expression in response to HIV antigens. It is possible that these cells play an important role in HIV-specific host defense.

At least three other forms of cell-mediated immunity to HIV have been described: CD8+ T cell–mediated suppression of HIV replication, ADCC, and NK cell activity. *CD8+ T cell–mediated suppression of HIV replication* refers to the ability of CD8+ T cells from an HIV-infected patient to inhibit the replication of HIV in tissue culture in a noncytolytic manner. There is no requirement for HLA compatibility between the CD8+ T cells and the HIV-infected cells. This effector mechanism is thus nonspecific and appears to be mediated by soluble factor(s) including the CC-chemokines RANTES (CCL5), MIP-1α (CCL3), and MIP-1β (CCL4). These CC-chemokines are potent suppressors of HIV replication and operate at least in part via blockade of the HIV co-receptor (*CCR5*) for R5 (macrophage-tropic) strains of HIV-1 (see above). *ADCC*, as described above in relation to humoral immunity, involves the killing of HIV-expressing cells by NK cells armed with specific antibodies directed against HIV antigens. Finally, *NK cells* alone have been shown to be capable of killing HIV-infected target cells in tissue culture. This primitive cytotoxic mechanism of host defense is directed toward nonspecific surveillance for neoplastic transformation and viral infection through recognition of altered class I MHC molecules.

DIAGNOSIS AND LABORATORY MONITORING OF HIV INFECTION

The establishment of HIV as the causative agent of AIDS and related syndromes early in 1984 was followed by the rapid development of sensitive screening tests for HIV infection. By March 1985, blood donors in the United States were routinely screened for antibodies to HIV. In 1996, blood banks in the United States added the p24 antigen capture assay to the screening process to help identify the rare infected individuals who were donating blood in the time (up to 3 months) between infection and the development of antibodies. In 2002, the ability to detect early infection with HIV was further enhanced by the licensure of nucleic acid testing (NAT) as a routine part of blood donor screening. These refinements decreased the interval between infection and detection (window period) from 22 days for antibody testing to 16 days with p24 antigen testing and subsequently to 12 days with nucleic acid testing. The development of sensitive assays for monitoring levels of plasma viremia ushered in a new era of being able to monitor the progression of HIV disease more closely. Utilization of these tests, coupled with the measurement of levels of CD4+ T lymphocytes in peripheral blood, is essential in the management of patients with HIV infection.

Diagnosis of HIV Infection

The CDC has recommended that screening for HIV infection be performed as a matter of routine health care. The diagnosis of HIV infection depends on the demonstration of antibodies to HIV and/or the direct detection of HIV or one of its components. As noted above, antibodies to HIV generally appear in the circulation 3–12 weeks following infection.

The standard blood screening test for HIV infection is the ELISA, also referred to as an *enzyme immunoassay* (EIA). This solid-phase assay is an extremely good screening test with a sensitivity of >99.5%. Most diagnostic laboratories use a commercial EIA kit that contains antigens from both HIV-1 and HIV-2 and thus are able to detect either. These kits use both natural and recombinant antigens and are continuously updated to increase their sensitivity to newly discovered species, such as group O viruses (Fig. 189-6). The fourth-generation EIA tests combine detection of antibodies to HIV with detection of the p24 antigen of HIV. EIA tests are generally scored as positive (highly reactive), negative (nonreactive), or indeterminate (partially reactive). While the EIA is an extremely sensitive test, it is not optimal with regard to specificity. This is particularly true in studies of low-risk individuals, such as volunteer blood donors. In this latter population, only 10% of EIA-positive individuals are subsequently confirmed to have HIV infection. Among the factors associated with false-positive EIA tests are antibodies to class II antigens (such as may be seen following pregnancy, blood transfusion, or transplantation), autoantibodies, hepatic disease, recent influenza vaccination, and acute viral infections. For these reasons, anyone suspected of having HIV infection based on a positive or inconclusive EIA result must have the result confirmed with a more specific assay such as the Western blot. One can estimate whether an individual has a recent infection with HIV-1 by comparing the results on a standard EIA test that will score positive for all infected individuals with the results on an assay modified to be less sensitive ("detuned assay") that will score positive only for individuals with established HIV infection. In rare instances, an HIV-infected individual treated early in the course of infection may revert to a negative EIA. This does *not* indicate clearing of infection; rather, it signifies levels of ongoing

exposure to virus insufficient to maintain a measurable antibody response. When these individuals have discontinued therapy, viruses and antibodies have reappeared.

The most commonly used confirmatory test is the Western blot **(Fig. 189-29)**. This assay takes advantage of the fact that multiple HIV antigens of different, well-characterized molecular weights elicit the production of specific antibodies. These antigens can be separated on the basis of molecular weight, and antibodies to each component can be detected as discrete bands on the Western blot. A negative Western blot is one in which no bands are present at molecular weights corresponding to HIV gene products. In a patient with a positive or indeterminate EIA and a negative Western blot, one can conclude with certainty that the EIA reactivity was a false positive. On the other hand, a Western blot demonstrating antibodies to products of all three of the major genes of HIV (*gag*, *pol*, and *env*) is conclusive evidence of infection with HIV. Criteria established by the U.S. Food and Drug Administration (FDA) in 1993 for a positive Western blot state that a result is considered positive if antibodies exist to two of the three HIV proteins: p24, gp41, and gp120/160. Using these criteria, ~10% of all blood donors deemed positive for HIV-1 infection lacked an antibody band to the *pol* gene product p31. Some 50% of these blood donors were subsequently found to be false positives. Thus, the absence of the p31 band should increase the suspicion that one may be dealing with a false-positive test result. In this setting it is prudent to obtain additional confirmation with an RNA-based test for HIV-1 and/or a follow-up Western blot. By definition, Western blot patterns of reactivity that do not fall into the positive or negative categories are considered "indeterminate." There are two possible explanations for an indeterminate Western blot result. The most likely explanation in a low-risk individual is that the patient being tested has antibodies that cross-react with one of the proteins of HIV. The most common patterns of cross-reactivity are antibodies that react with p24 and/or p55. The least likely explanation in this setting is that the individual is infected with HIV and is in the process of mounting a classic antibody response. In either instance, the Western blot should be repeated in 1 month to determine whether the indeterminate pattern is a pattern in evolution. In addition, one may attempt to confirm a diagnosis of HIV infection with the p24 antigen capture assay or one of the tests for HIV RNA (discussed below). While the Western blot is an excellent confirmatory test for HIV infection in patients with a positive or indeterminate EIA, it is a poor screening test. Among individuals with a negative EIA and PCR for HIV, 20–30% may show one or more bands on Western blot. While these bands are usually faint and represent cross-reactivity, their presence creates a situation in which other diagnostic modalities (such as DNA PCR, RNA PCR, the bDNA assay, or p24 antigen capture) must be employed to ensure that the bands do not indicate early HIV infection.

Figure 189-29

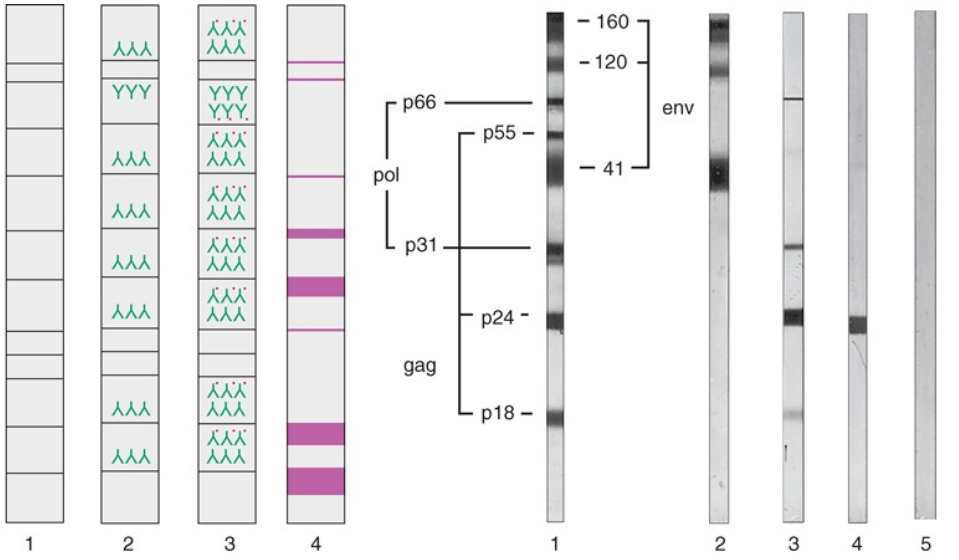

A
1. Virus digested: digest separated into components by molecular weight
2. Proteins transferred to filter paper: reaction with test serum
3. Enzyme-conjugated antihuman antibody added
4. Substrate added and color noted

B
1. Positive HIV-1 infection
2. gp 160 immunization
3. Indeterminate (HIV-2 infection)
4. Indeterminate (cross-reacting antibody to p24)
5. Negative

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Western blot assay for detection of antibodies to HIV. ***A***. Schematic representation of how a Western blot is performed. ***B.*** Examples of patterns of Western blot reactivity. In each instance the Western blot strip contains antigens to HIV-1. The serum from the patient immunized to the HIV-1 envelope gp160 contains only antibodies to the HIV-1 envelope proteins. The serum from the patient with HIV-2 infection cross-reacts with both *reverse transcriptase* and *gag* gene products of HIV-1.

A guideline for the use of these serologic tests in attempting to make a diagnosis of HIV infection is depicted in **Fig. 189-30**. In patients in whom HIV infection is suspected, the appropriate initial test is the EIA. If the result is negative, unless there is strong reason to suspect early HIV infection (as in a patient exposed within the previous 3 months), the diagnosis is ruled out and retesting should be performed only as clinically indicated. If the EIA is indeterminate or positive, the test should be repeated. If the repeat is negative on two occasions, one can assume that the initial positive reading was due to a technical error in the performance of the assay and that the patient is negative. If the repeat is indeterminate or positive, one should proceed to the HIV-1 Western blot. If the Western blot is positive, the diagnosis is HIV-1 infection. If the Western blot is negative, the EIA can be assumed to have been a false positive for HIV-1 and the diagnosis of HIV-1 infection is ruled out. It would also be prudent at this point to perform specific serologic testing for HIV-2 following the same type of algorithm. If the Western blot for HIV-1 is indeterminate, it should be repeated in 4–6 weeks; in addition, one may proceed to a p24 antigen capture assay, HIV-1 RNA assay, or HIV-1 DNA PCR and specific serologic testing for HIV-2. If the p24 and HIV RNA assays are negative and there is no progression in the Western blot, a diagnosis of HIV-1 is ruled out. If either the p24 or HIV-1 RNA assay is positive and/or the HIV-1 Western blot shows progression, a tentative diagnosis of HIV-1 infection can be made and later confirmed with a follow-up Western blot demonstrating a positive pattern. In addition to these standard laboratory-based assays for detecting antibodies to

HIV, a series of point-of-care tests are also available that can provide results in 1–60 minutes. Among the most popular of these is the OraQuick Rapid HIV-1 antibody test that can be run on blood, plasma, or saliva. The sensitivity and specificity of this test are each ~99%. While negative results from this test are adequate to rule out a diagnosis of HIV infection, a positive finding should be considered preliminary and confirmed with standard serologic testing, as described above.

Figure 189-30

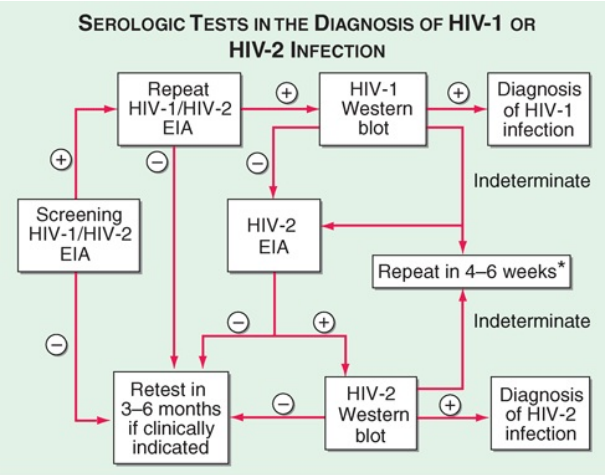

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the use of serologic tests in the diagnosis of HIV-1 or HIV-2 infection. *Stable indeterminate Western blot 4–6 weeks later makes HIV infection unlikely. However, it should be repeated twice at 3-month intervals to rule out HIV infection. Alternatively, one may test for HIV-1 p24 antigen or HIV RNA. EIA, enzyme immunoassay.

A variety of laboratory tests are available for the direct detection of HIV or its components **(Table 189-7; Fig. 189-31)**. These tests may be of considerable help in making a diagnosis of HIV infection when the Western blot results are indeterminate. In addition, the tests detecting levels of HIV RNA can be used to determine prognosis and to assess the response to antiretroviral therapies. The simplest of the direct detection tests is the *p24 antigen capture assay*. This is an EIA-type assay in which the solid phase consists of antibodies to the p24 antigen of HIV. It detects the viral protein p24 in the blood of HIV-infected individuals where it exists either as free antigen or complexed to anti-p24 antibodies. Overall, ~30% of individuals with untreated HIV infection have detectable levels of free p24 antigen. This increases to ~50% when samples are treated with a weak acid to dissociate antigen-antibody complexes. Throughout the course of HIV infection, an equilibrium exists between p24 antigen and anti-p24 antibodies. During the first few weeks of infection, before an immune response develops, there is a brisk rise in p24 antigen levels (Fig. 189-27). After the development of anti-p24 antibodies, these levels decline. Late in the course of infection, when circulating levels of virus are high, p24 antigen levels also increase, particularly when detected by techniques involving dissociation of antigen-antibody complexes. The p24 antigen capture assay has its greatest use as a screening test for HIV infection in patients suspected of having the acute HIV syndrome, as high levels of p24 antigen are present prior to the development of antibodies. Its use for routine blood donor screening for HIV infection has been replaced by use of nucleic acid testing. The ability to measure and monitor levels of HIV RNA in the plasma of patients with HIV infection has been of extraordinary value in furthering our understanding of the pathogenesis of HIV infection and in providing a diagnostic tool in settings where measurements of anti-HIV antibodies may be misleading, such as in acute infection and neonatal infection. Three assays are predominantly used for this purpose. They are reverse transcriptase PCR (*RT-PCR*; Amplicor); branched DNA (*bDNA*; VERSANT); and nucleic acid sequence–based amplification (*NASBA*; NucliSens). These tests are of value in making a diagnosis of HIV infection, in establishing initial prognosis, in determining the need for therapy, and for monitoring the effects of therapy. In addition to these three commercially available tests, the *DNA PCR* is also employed by research laboratories for making a diagnosis of HIV infection by amplifying HIV proviral DNA from peripheral blood mononuclear cells. The commercially available RNA detection tests have a sensitivity of 40–80 copies of HIV RNA per milliliter of plasma. Research laboratory–based RNA assays can detect as few as one HIV RNA copy per milliliter, while the DNA PCR tests can detect proviral DNA at a frequency of one copy per 10,000–100,000 cells. Thus, these tests are extremely sensitive. One frequent consequence of a high degree of sensitivity is some loss of specificity, and false-positive results have been reported with each of these techniques. For this reason, a positive EIA with a confirmatory Western blot remains the "gold standard" for a diagnosis of HIV infection, and the interpretation of other test results must be done with this in mind.

Table 189-7 Characteristics of Tests for Direct Detection of HIV

Test	Technique	Sensitivity[a]	Cost/Test[b]
Immune complex–dissociated p24 antigen capture assay	Measurement of levels of HIV-1 core protein in an EIA-based format following dissociation of antigen-antibody complexes by weak acid treatment	Positive in 50% of patients; detects down to 15 pg/mL of p24 protein	$1–2
HIV RNA by PCR	PCR amplification of cDNA generated from viral RNA (target amplification)	Reliable to 40 copies/mL of HIV RNA	$75–150
HIV RNA by bDNA	Measurement of levels of particle-associated HIV RNA in a nucleic acid capture assay employing signal amplification	Reliable to 50 copies/mL of HIV RNA	$75–150
HIV RNA by NASBA	Isothermic nucleic acid amplification with internal controls	Reliable to 80 copies/mL of HIV RNA	$75–150

[a]Sensitivity figures refer to those approved by the U.S. Food and Drug Administration.

[b]Prices may be lower in large-volume settings.

Abbreviations: bDNA, branched DNA; cDNA; complementary DNA;. EIA, enzyme immunoassay; NASBA, nucleic acid sequence

–based amplification; PCR, polymerase chain reaction.

Figure 189-31

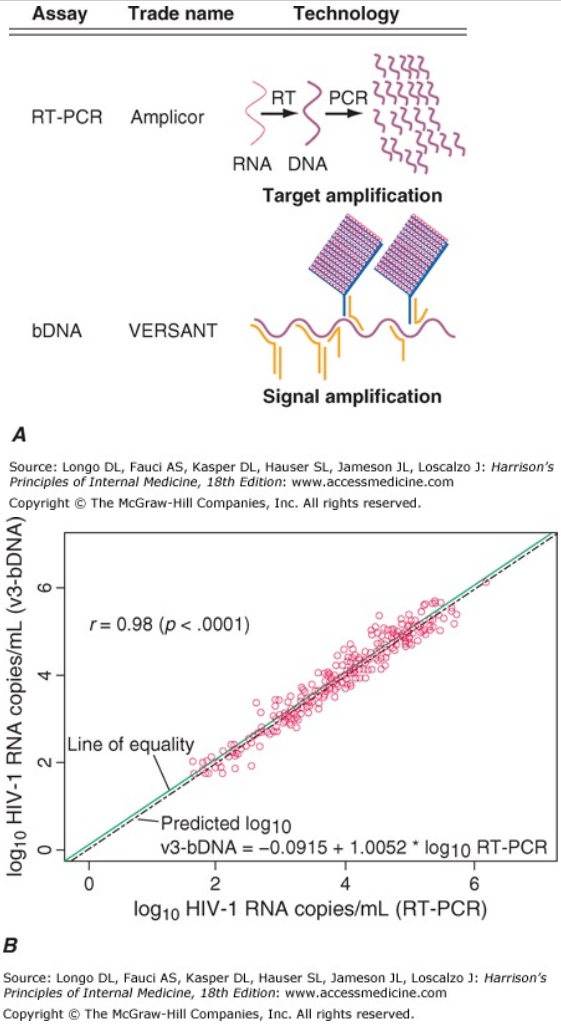

Comparison of RT-PCR and bDNA assays. ***A***. Schematic representation of reverse transcriptase–polymerase chain reaction (RT-PCR) and bDNA assays. See text for detailed description. ***B.*** Scatter plot of $\log_{10}$ v3-bDNA versus $\log_{10}$ RT-PCR with the line of equity (solid) and the fitted regression line (hatched). The equation for the fitted regression line is given in the lower-right-hand corner. There is good agreement between the two assays. v3, version 3 of the bDNA assay. *(Adapted from HC Highbarger et al: J Clin Microbiol 37:3612, 1999.)*

In the RT-PCR technique, following DNAse treatment, a cDNA copy is made of all RNA species present in plasma. Insofar as HIV is an RNA virus, this will result in the production of DNA copies of the HIV genome in amounts proportional to the amount of HIV RNA present in plasma. This cDNA is then amplified and characterized using standard PCR techniques, employing primer pairs that can distinguish genomic cDNA from messenger cDNA. The bDNA assay involves the use of a solid-phase nucleic acid capture system and signal amplification through successive nucleic acid hybridizations to detect small quantities of HIV RNA. Both tests can achieve a tenfold increase in sensitivity to 40–50 copies of HIV RNA per milliliter with a preconcentration step in which plasma undergoes ultracentrifugation to pellet the viral particles. The NASBA technique involves the isothermal amplification of a sequence within the gag region of HIV in the presence of internal standards and employs the production of multiple RNA copies through the action of T7-RNA polymerase. The resulting RNA species are quantitated through hybridization with a molecular beacon DNA probe that is quenched in the absence of hybridization. The lower limit of detection for the NucliSens assay is 80 copies/mL.

In addition to being a diagnostic and prognostic tool, RT-PCR and DNA-PCR are also useful for amplifying defined areas of the HIV genome for sequence analysis and have become an important technique for studies of sequence diversity and microbial resistance to antiretroviral agents. In patients with a positive or indeterminate EIA test and an indeterminate Western blot, and in patients in whom serologic testing may be unreliable (such as patients with hypogammaglobulinemia or advanced HIV disease), these tests for quantitating HIV RNA in plasma or detecting proviral DNA in peripheral blood mononuclear cells are valuable tools for making a diagnosis of HIV infection; however, they should be used for diagnosis only when standard serologic testing has failed to provide a definitive result.

Laboratory Monitoring of Patients with HIV Infection

The epidemic of HIV infection and AIDS has provided the clinician with new challenges for integrating clinical and laboratory data to effect optimal patient management. The close relationship between clinical manifestations of HIV infection and CD4+ T cell count has made measurement of CD4+ T cell numbers a routine part of the evaluation of HIV-infected individuals. The discovery of HIV as the cause of AIDS led to the development of sensitive tests that allow one to monitor the levels of HIV in the blood. Determinations of peripheral blood CD4+ T cell counts and measurements of the plasma levels of HIV RNA plasma provide a powerful set of tools for determining prognosis and monitoring response to therapy.

CD4+ T CELL COUNTS

The CD4+ T cell count is the laboratory test generally accepted as the best indicator of the immediate state of immunologic competence of the patient with HIV infection. This measurement, which can be made directly or calculated as the product of the

percent of CD4+ T cells (determined by flow cytometry) and the total lymphocyte count [determined by the white blood cell count (WBC) multiplied by the lymphocyte differential percent], has been shown to correlate very well with the level of immunologic competence. Patients with CD4+ T cell counts <200/μL are at high risk of disease from *P. jiroveci*, while patients with CD4+ T cell counts <50/μL are at high risk of disease from CMV, mycobacteria of the *M. avium* complex (MAC), and/or *T. gondii* **(Fig. 189-32)**. Patients with HIV infection should have CD4+ T cell measurements performed at the time of diagnosis and every 3–6 months thereafter. More frequent measurements should be made if a declining trend is noted. According to U.S. Department of Health and Human Services Guidelines, a CD4+ T cell count <500/μL is an indication for initiating cART, and a decline in CD4+ T cell count of >25% is an indication for considering a change in therapy. Once the CD4+ T cell count is <200/μL, patients should be placed on a regimen for *P. jiroveci* prophylaxis, and once the count is <50/μL, primary prophylaxis for MAC infection is indicated. As with any laboratory measurement, one may wish to obtain two determinations prior to any significant changes in patient management based on CD4+ T cell count alone. There are a handful of clinical situations in which the CD4+ T cell count may be misleading. Patients with HTLV-1/HIV co-infection may have elevated CD4+ T cell counts that do not accurately reflect their degree of immune competence. In patients with hypersplenism or those who have undergone splenectomy, and in patients receiving medications that suppress the bone marrow such as IFN-α, the CD4+ T cell percentage may be a more reliable indication of immune function than the CD4+ T cell count. A CD4+ T cell percent of 15 is comparable to a CD4+ T cell count of 200/μL.

Figure 189-32

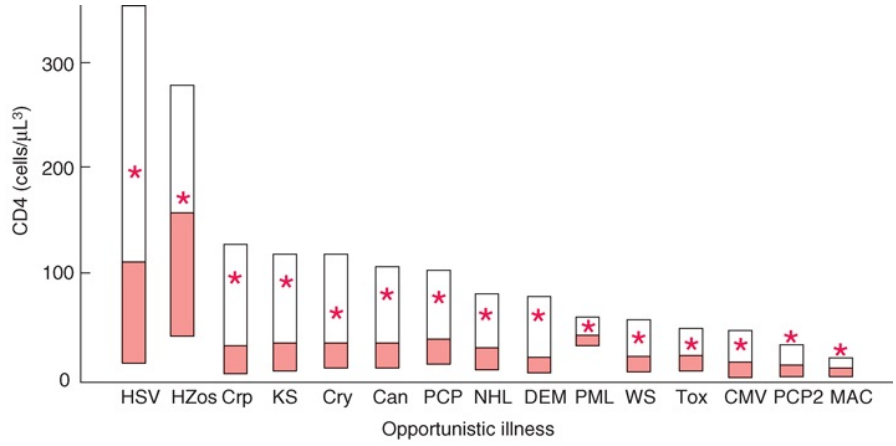

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship between CD4+ T cell counts and the development of opportunistic diseases. Boxplot of the median (line inside the box), first quartile (bottom of the box), third quartile (top of the box), and mean (asterisk) CD4+ lymphocyte count at the time of the development of opportunistic disease. Can, candidal esophagitis; CMV, cytomegalovirus infection; Crp, cryptosporidiosis; Cry, cryptococcal meningitis; DEM, AIDS dementia complex; HSV, herpes simplex virus infection; HZos, herpes zoster; KS, Kaposi's sarcoma; MAC, *Mycobacterium avium* complex bacteremia; NHL, non-Hodgkin's lymphoma; PCP, primary *Pneumocystis jiroveci* pneumonia; PCP2, secondary *P. jiroveci* pneumonia; PML, progressive multifocal leukoencephalopathy; Tox, *Toxoplasma gondii* encephalitis; WS, wasting syndrome. *(From RD Moore, RE Chaisson: Ann Intern Med 124:633, 1996.)*

HIV RNA DETERMINATIONS

Facilitated by highly sensitive techniques for the precise quantitation of small amounts of nucleic acids, the measurement of serum or plasma levels of HIV RNA has become an essential component in the monitoring of patients with HIV infection. As discussed under "Diagnosis of HIV Infection," above, the two most commonly used techniques are the RT-PCR assay and the bDNA assay. Both assays generate data in the form of number of copies of HIV RNA per milliliter of serum or plasma. Standard assays can detect as few as 40–50 copies of HIV RNA per milliliter of plasma, while research-based assays can detect down to one copy per milliliter. While it is common practice to describe levels of HIV RNA below these cut-offs as "undetectable," this is a term that should be avoided as it is imprecise and leaves the false impression that the level of virus is 0. By utilizing more sensitive, nested PCR techniques and by studying tissue levels of virus as well as plasma levels, HIV RNA can be detected in virtually every patient with HIV infection. One notable exception to this is a patient who undervent cytoreducion therapy followed by a bone marrow transplant from a ccr5Δ32 homozygous donor. Measurements of changes in HIV RNA levels over time have been of great value in delineating the relationship between levels of virus and rates of disease progression (Fig. 189-22), the rates of viral turnover, the relationship between immune system activation and viral replication, and the time to development of drug resistance. HIV RNA measurements are greatly influenced by the state of activation of the immune system and may fluctuate greatly in the setting of secondary infections or immunization. For these reasons, decisions based on HIV RNA levels should never be made on a single determination. Measurements of plasma HIV RNA levels should be made at the time of HIV diagnosis and every 3–6 months thereafter in the untreated patient. Following the initiation of therapy or any change in therapy, plasma HIV RNA levels should be monitored approximately every 4 weeks until the effectiveness of the therapeutic regimen is determined by the development of a new steady-state level of HIV RNA. In most instances of effective antiretroviral therapy the plasma level of HIV RNA will drop to <50 copies per milliliter within 6 months of the initiation of treatment. During therapy, levels of HIV RNA should be monitored every 3–4 months to evaluate the continuing effectiveness of therapy.

HIV RESISTANCE TESTING

The availability of multiple antiretroviral drugs as treatment options has generated a great deal of interest in the potential for measuring the sensitivity of an individual's HIV virus(es) to different antiretroviral agents. HIV resistance testing can be done through either genotypic or phenotypic measurements. In the genotypic assays, sequence analyses of the HIV genomes obtained from patients are compared with sequences of viruses with known antiretroviral resistance profiles. In the phenotypic assays, the in vivo growth of viral isolates obtained from the patient is compared to the growth of reference strains of the virus in the presence or absence of different antiretroviral drugs. A modification of this phenotypic approach utilizes a comparison of the enzymatic activities of the reverse transcriptase or protease genes obtained by molecular cloning of patients' isolates to the enzymatic activities of genes obtained from reference strains of HIV in the presence or absence of different drugs targeted to these genes. These tests are quite good in identifying those antiretroviral agents that have been utilized in the past and suggesting agents that may be of future value in a given patient. Drug resistance testing in the setting of virologic failure should be performed while the patient is still on the failing regimen because of the propensity for the pool of HIV quasispecies to rapidly

revert to wild-type in the absence of the selective pressures of cART. In the hands of experts, resistance testing enhances the short-term ability to decrease viral load by ~0.5 log compared with changing drugs merely on the basis of drug history. In addition to the use of resistance testing to help in the selection of new drugs in patients with virologic failure, it may also be of value in selecting an initial regimen for treatment of therapy-naïve individuals. This is particularly true in geographic areas with a high level of background resistance.

CO-RECEPTOR TROPISM ASSAYS

Following the licensure of maraviroc as the first CCR5 antagonist for the treatment of HIV infection (see below), it became necessary to be able to determine whether a patient's virus was likely to respond to this treatment. Patients tend to have CCR5-tropic virus early in the course of infection with a trend toward CXCR4 viruses later in disease. The antiretroviral agent maraviroc is effective only against CCR5-tropic viruses. Because the genotypic determinants of cellular tropism are poorly defined, a phenotypic assay is necessary to determine this property of HIV. Two commercial assays, the Trofile assay (Monogram Biosciences) and the Phenoscript assay (VIRalliance), are available to make this determination. These assays clone the envelope regions of the patient's virus into an indicator virus that is then used to infect target cells expressing either CCR5 or CXCR4 as their co-receptor. These assays take weeks to perform and are expensive. Although commercial genotypic assays of proviral DNA are available, their role as predictors of response to CCR5 inhibitor therapy is unclear.

OTHER TESTS

A variety of other laboratory tests have been studied as potential markers of HIV disease activity. Among these are quantitative culture of replication-competent HIV from plasma, peripheral blood mononuclear cells, or resting CD4+ T cells; circulating levels of β_2-microglobulin, soluble IL-2 receptor, IgA, acid-labile endogenous IFN, or TNF-α; and the presence or absence of activation markers such as CD38, HLA-DR, or PD-1 on CD8+ T cells. Nonspecific serologic markers of inflammation and/or coagulation such as IL-6, d-dimer, and sCD14 have been shown to have a high correlation with all-cause mortality **(Table 189-8)**. While these measurements have value as markers of disease activity and help to increase our understanding of the pathogenesis of HIV disease, they do not currently play a major role in the monitoring of patients with HIV infection.

Table 189-8 Association between High-Sensitivity CRP, IL-6, and D-Dimer with All-Cause Mortality in Patients with HIV Infection

	Unadjusted		Adjusted	
Marker	**Odds Ratio (Fourth/First)**	***p***	**Odds Ratio (Fourth/First)**	***p***
Hs-CRP	2.0	.05	2.8	.03
IL-6	8.3	<.0001	11.8	<.0001
D-dimer	12.4	<.0001	26.5	<.0001

Note: Hs-CRP, high-sensitivity C-reactive protein; IL-6, interleukin 6.

Source: From Kuller et al.

CLINICAL MANIFESTATIONS

The clinical consequences of HIV infection encompass a spectrum ranging from an acute syndrome associated with primary infection to a prolonged asymptomatic state to advanced disease. It is best to regard HIV disease as beginning at the time of primary infection and progressing through various stages. As mentioned above, active virus replication and progressive immunologic impairment occur throughout the course of HIV infection in most patients. With the exception of the rare, true, "elite" long-term nonprogressors (see "Long-Term Survivors and Long-Term Nonprogressors," above), HIV disease in untreated patients inexorably progresses even during the clinically latent stage. Since the mid-1990s, cART has had a major impact on preventing and reversing the progression of disease over extended periods of time in a substantial proportion of adequately treated patients.

The Acute HIV Syndrome

It is estimated that 50–70% of individuals with HIV infection experience an acute clinical syndrome ~3–6 weeks after primary infection **(Fig. 189-33)**. Varying degrees of clinical severity have been reported, and although it has been suggested that symptomatic seroconversion leading to the seeking of medical attention indicates an increased risk for an accelerated course of disease, there does not appear to be a correlation between the level of the initial burst of viremia in acute HIV infection and the subsequent course of disease. The typical clinical findings in the acute HIV syndrome are listed in **Table 189-9**; they occur along with a burst of plasma viremia. It has been reported that several symptoms of the acute HIV syndrome (fever, skin rash, pharyngitis, and myalgia) occur less frequently in those infected by injection drug use compared with those infected by sexual contact. The syndrome is typical of an acute viral syndrome and has been likened to acute infectious mononucleosis. Symptoms usually persist for one to several weeks and gradually subside as an immune response to HIV develops and the levels of plasma viremia decrease. Opportunistic infections have been reported during this stage of infection, reflecting the immunodeficiency that results from reduced numbers of CD4+ T cells and likely also from the dysfunction of CD4+ T cells owing to viral protein and endogenous cytokine-induced perturbations of cells (Table 189-4) associated with the extremely high levels of plasma viremia. A number of immunologic abnormalities accompany the acute HIV syndrome, including multiphasic perturbations of the numbers of circulating lymphocyte subsets. The number of total lymphocytes and T cell subsets (CD4+ and CD8+) are initially reduced. An inversion of the CD4+/CD8+ T cell ratio occurs later because of a rise in the number of CD8+ T cells. In fact, there may be a selective and transient expansion of CD8+ T cell subsets, as determined by T cell receptor analysis (see above). The total circulating CD8+ T cell count may remain elevated or return to normal; however, CD4+ T cell levels usually remain somewhat depressed, although there may be a slight rebound toward normal. Lymphadenopathy occurs in ~70% of individuals with primary HIV infection. Most patients recover spontaneously from this syndrome and many are left with only a mildly depressed CD4+ T cell count that remains stable for a variable period before beginning its progressive decline; in some individuals, the CD4+ T cell count returns to the normal range. Approximately 10% of patients manifest a fulminant course of immunologic and clinical deterioration after primary infection, even after the disappearance of initial symptoms. In most patients, primary infection with or without the acute syndrome is followed by a prolonged period of clinical latency or smoldering low disease activity. A small percentage of HIV-infected individuals treated with antiretroviral drugs during acute infection may revert to a negative EIA test during the time they remain on therapy. They rapidly re-seroconvert with the discontinuation of treatment.

Table 189-9 Clinical Findings in the Acute HIV Syndrome

General	Neurologic
Fever	Meningitis
Pharyngitis	Encephalitis

Lymphadenopathy	Peripheral neuropathy
Headache/retroorbital pain	Myelopathy
Arthralgias/myalgias	Dermatologic
Lethargy/malaise	Erythematous maculopapular rash
Anorexia/weight loss	Mucocutaneous ulceration
Nausea/vomiting/diarrhea	

Source: From B Tindall, DA Cooper: AIDS 5:1, 1991.

Figure 189-33

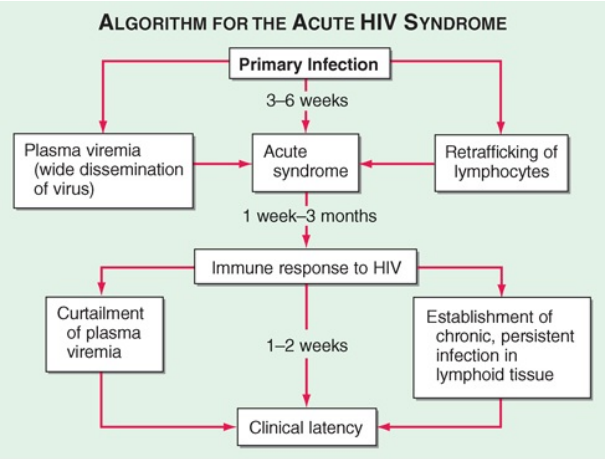

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The acute HIV syndrome. See text for detailed description. *(Adapted from G Pantaleo et al: N Engl J Med 328:327, 1993. Copyright 1993 Massachusetts Medical Society. All rights reserved.)*

The Asymptomatic Stage—Clinical Latency

Although the length of time from initial infection to the development of clinical disease varies greatly, the median time for untreated patients is ~10 years. As emphasized above, HIV disease with active virus replication is ongoing and progressive during this asymptomatic period. The rate of disease progression is directly correlated with HIV RNA levels. Patients with high levels of HIV RNA in plasma progress to symptomatic disease faster than do patients with low levels of HIV RNA (Fig. 189-22). Some patients referred to as *long-term nonprogressors* show little if any decline in CD4+ T cell counts over extended periods of time. These patients generally have extremely low levels of HIV RNA; a subset, referred to as *elite nonprogressors*, exhibits HIV RNA levels <50 copies per milliliter. Certain other patients remain entirely asymptomatic despite the fact that their CD4+ T cell counts show a steady progressive decline to extremely low levels. In these patients, the appearance of an opportunistic disease may be the first manifestation of HIV infection. During the asymptomatic period of HIV infection, the average rate of CD4+ T cell decline is ~50/μL per year. When the CD4+ T cell count falls to <200/μL, the resulting state of immunodeficiency is severe enough to place the patient at high risk for opportunistic infection and neoplasms and, hence, for clinically apparent disease.

Symptomatic Disease

Symptoms of HIV disease can appear at any time during the course of HIV infection. Generally speaking, the spectrum of illnesses that one observes changes as the CD4+ T cell count declines. The more severe and life-threatening complications of HIV infection occur in patients with CD4+ T cell counts <200/μL. A diagnosis of AIDS is made in anyone with HIV infection and a CD4+ T cell count <200/μL and in anyone with HIV infection who develops one of the HIV-associated diseases considered to be indicative of a severe defect in cell-mediated immunity (category C, Table 189-2). While the causative agents of the secondary infections are characteristically opportunistic organisms such as *P. jiroveci*, atypical mycobacteria, CMV, and other organisms that do not ordinarily cause disease in the absence of a compromised immune system, they also include common bacterial and mycobacterial pathogens. Following the widespread use of cART and implementation of guidelines for the prevention of opportunistic infections **(Table 189-10)**, the incidence of these secondary infections has decreased dramatically **(Fig. 189-34)**. Overall, the clinical spectrum of HIV disease is constantly changing as patients live longer and new and better approaches to treatment and prophylaxis are developed. In addition to the classic AIDS-defining illnesses, patients with HIV infection also have an increase in serious non-AIDS illnesses, including non-AIDS related cancers and, cardiovascular, renal and hepatic disease. Non-AIDS events dominate the disease burden for patients with HIV infection receiving cART (Table 189-3). Fewer than 50% of deaths among AIDS patients are as a direct result of an AIDS-defining illness. The physician providing care to a patient with HIV infection must be well versed in general internal medicine as well as HIV-related opportunistic diseases. In general, it should be stressed that a key element of treatment of symptomatic complications of HIV disease, whether they are primary or secondary, is achieving good control of HIV replication through the use of cART and instituting primary and secondary prophylaxis for opportunistic infections as indicated.

Table 189-10 NIH/CDC/IDSA 2009 Guidelines for the Prevention of Opportunistic Infections in Persons Infected with HIV

Pathogen	Indications	First Choice(s)	Alternatives
Recommended as Standard of Care for Primary and Secondary Prophylaxis			
Pneumocystis jiroveci	CD4+ T cell count <200/μL *or* Oropharyngeal candidiasis *or*	Trimethoprim/sulfamethoxazole (TMP/SMX), 1 DS tablet qd PO *or*	Dapsone 50 mg bid PO or 100 mg/d PO *or* Dapsone 50 mg/d PO +

	Prior bout of PCP		Pyrimethamine 50 mg/week PO + Leucovorin 25 mg/week PO
	May stop prophylaxis if CD4+ T cell count >200/μL for ≥3 months	TMP/SMX, 1 SS tablet qd PO	*or* (Dapsone 200 mg PO + Pyrimethamine 75 mg PO + Leucovorin 25 mg) weekly PO *or* Aerosolized pentamidine, 300 mg via Respirgard II nebulizer every month *or* Atovaquone 1500 mg/d PO *or* TMP/SMX 1 DS tablet 3x/week PO
Mycobacterium tuberculosis			
Isoniazid sensitive	Skin test >5 mm *or* Prior positive test without treatment *or* Close contact with case of active pulmonary TB	(Isoniazid 300 mg PO + Pyridoxine 50 mg PO) qd x 9 months *or* Isoniazid 900 mg PO twice weekly + Pyridoxine 50 mg PO daily x 9 months	Rifabutin 300 mg or rifampin 600 mg PO qd x 4 months
Isoniazid resistant	Same with high probability of exposure to isoniazid-resistant TB	(Rifabutin 300 mg or Rifampin 600 mg) PO qd x4 months	
Multidrug resistant	Same with high probability of exposure to multidrug-resistant TB	Consult local public health authorities	
Mycobacterium-avium complex	CD4+ T cell count <50/μL	Azithromycin 1200 mg weekly PO *or* Clarithromycin 500 mg bid PO	Rifabutin 300 mg/d PO *or* Azithromycin 600 mg twice weekly PO
	Prior documented disseminated disease May stop prophylaxis if CD4+ T cell count >100/μL for ≥3 months	Clarithromycin 500 mg bid PO + Ethambutol 15 (mg/kg)/d PO ± Rifabutin 300 mg/d PO	Azithromycin 500 mg/d PO + Ethambutol 15 (mg/kg)/d PO ± Rifabutin 300 mg/d PO
Toxoplasma gondii	TOXO IgG antibody positive and CD4+ T cell count <100/μL	TMP/SMX 1 DS tablet PO qd	TMP/SMX 1 DS 3x weekly PO *or* TMP/SMX, 1 SS PO daily *or* Dapsone 50 mg/d PO + Pyrimethamine 50 mg weekly PO + Leucovorin 25 mg weekly PO *or* (Dapsone 200 mg PO + Pyrimethamine 75 mg PO + Leucovorin 25 mg PO) weekly *or* (Atovaquone 1500 mg PO ± Pyrimethamine 25 mg PO + Leucovorin 10 mg PO) daily

	Prior toxoplasmic encephalitis and CD4+ T cell count <200/μL	Sulfadiazine 500–1000 mg qid PO + Pyrimethamine 25–50 mg/d PO + Leucovorin 10–25 mg/d PO	Clindamycin 600 mg q8h PO + Pyrimethamine 25–50 mg/d PO + Leucovorin 10–25 mg/d PO
	May stop prophylaxis if CD4+ T cell count >200/μL for ≥3 months		Atovaquone 750 mg PO q6–12 h ± Pyrimethamine 25 mg/d PO + Leucovorin 10 mg/d PO
Varicella zoster virus	Significant exposure to chickenpox or shingles in a patient with no history of immunization or prior exposure to either	Varicella zoster immune globulin, IM, within 96 h of exposure (1-800-843-74+77)	Acyclovir 800 mg PO 5 x 1 day for 5 days
Cryptococcus neoformans	Prior documented disease May stop prophylaxis if CD4+ T cell count >200/μL for 6 months and no evidence of active infection	Fluconazole 200 mg/d PO	Itraconazole 200 mg/d PO
Histoplasma capsulatum	Prior documented disease or CD4+ T cell count <150μL and high risk (endemic area or occupational exposure) May stop prophylaxis after 1 year if CD4+ T cell count >150/μL and patient on ARV therapy for ≥6 months	Itraconazole 200 mg bid PO	Fluconazole 800 mg/d PO
Coccidioides immitis	Prior documented disease *or* positive serology and CD4+ T cell count <250/μL if from a disease endemic area. (For this indication prophylaxis can be stopped if CD4+ T cell count ≥250 for 6 months.)	Fluconazole 400 mg/d PO	Itraconazole 200 mg bid PO
Penicillium marneffei	Prior documented disease May stop secondary prophylaxis in patients on ARV therapy with CD4+ T cell count >100/μL for ≥6 months	Itraconazole 200 mg/d PO	
Salmonella species	Prior reament bacteremia	Ciprofloxacin 500 mg bid PO for ≥6 months	
Bartonella	Prior infection May stop if CD4+ T cell count >200/μL for >3 months	Doxycycline 200 mg/d *or* Azithromycin 1200 mg weekly PO *or* Clarithromycin 500 mg bid PO	
Cytomegalovirus	Prior end-organ disease May stop prophylaxis if CD4+ T cell count >100/μL for 6 months and no evidence of active CMV disease Restart if prior retinitis and CD4+ T cells <100/μL	Valganciclovir 900 mg bid PO *or* Ganciclovir sustained-release implant q6–9 months + Valganciclovir 900 mg bid PO	Cidofovir 5 mg/kg every other week IV + Probenecid *or* Fomivirsen 330 μg intravitreal q2–4 week *or* Foscarnet 90–120 (mg/kg)/d IV
Immunizations Generally Recommended			
Hepatitis B virus	All susceptible (anti-HBc-and anti-HBs-negative) patients	Hepatitis B vaccine: 3 doses	
Hepatitis A virus	All susceptible (anti-HAV-negative) patients	Hepatitis A vaccine: 2 doses	
Influenza virus	All patients annually	Inactivated trivalent influenza virus vaccine 1 dose yearly	Oseltamivir 75 mg PO qd *or* Rimantadine or amantadine 100 mg PO bid (influenza A only)
Streptococcus pneumoniae	All patients, preferably before CD4+ T cell count ≤200/μL	Pneumococcal vaccine 0.5 mL IM x1 if CD4+ T cell count >200/μL Reimmunize patients initially immunized at a CD4+ T cell count <100/μL whose CD4+ T cell count then increases to >200/μL	
Human papillomavirus	All patients 9–26 years of age	HPV vaccine; 3 doses	

Recommended for Prevention of Severe or Frequent Recurrences			
Herpes simplex	Frequent/severe recurrences	Valacyclovir 500 mg bid PO *or* Acyclovir 400 mg bid PO *or* Famciclovir 500 mg bid PO	
Candida	Frequent/severe recurrences	Fluconazole 100–200 mg/d PO	Itraconazole solution 200 mg/d PO posacomagole 400 mg bid PO

Abbreviations: ARV, antiretroviral; bid, twice daily; DS, double-strength; PCP, *Pneumocystis jiroveci* pneumonia; PO, by mouth; SS, single-strength; TB, tuberculosis.

Figure 189-34

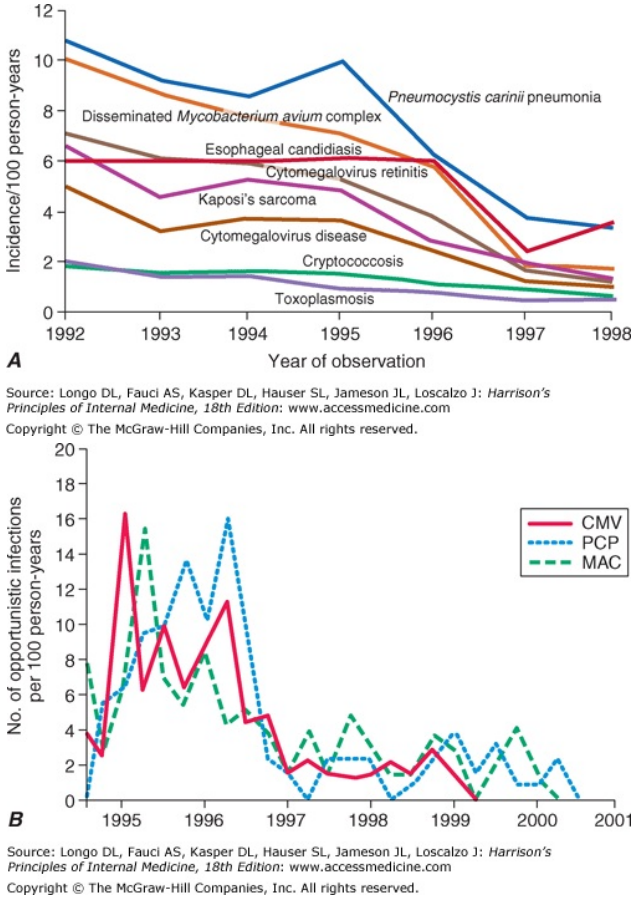

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Decrease in the incidence of opportunistic infections and Kaposi's sarcoma in HIV-infected individuals with CD4+ T cell counts <100/μL from 1992 through 1998. *[Adapted and updated from FJ Palella et al: N Engl J Med 338:853, 1998, and JE Kaplan et al: Clin Infect Dis 30(S1):S5, 2000, with permission.]* ***B.*** Quarterly incidence rates of cytomegalovirus (CMV), *Pneumocystis jiroveci* pneumonia (PCP), and *Mycobacterium avium* complex (MAC) from 1995 to 2001. *(From FJ Palella et al: AIDS 16:1617, 2002.)*

DISEASES OF THE RESPIRATORY SYSTEM

Acute bronchitis and sinusitis are prevalent during all stages of HIV infection. The most severe cases tend to occur in patients with lower CD4+ T cell counts. Sinusitis presents as fever, nasal congestion, and headache. The diagnosis is made by CT or MRI. The maxillary sinuses are most commonly involved; however, disease is also frequently seen in the ethmoid, sphenoid, and frontal sinuses. While some patients may improve without antibiotic therapy, radiographic improvement is quicker and more pronounced in patients who have received antimicrobial therapy. It is postulated that this high incidence of sinusitis results from an increased frequency of infection with encapsulated organisms such as *H. influenzae* and *Streptococcus pneumoniae*. In patients with low CD4+ T cell counts one may see mucormycosis infections of the sinuses. In contrast to the course of this infection in other patient populations, mucormycosis of the sinuses in patients with HIV infection may progress more slowly. In this setting aggressive, frequent local debridement in addition to local and systemic amphotericin B may result in effective treatment.

Pulmonary disease is one of the most frequent complications of HIV infection. The most common manifestation of pulmonary disease is pneumonia. Three of the 10 most common AIDS-defining illnesses are recurrent bacterial pneumonia, tuberculosis, and pneumonia due to the unicellular fungus *P. jiroveci*. Other major causes of pulmonary infiltrates include other mycobacterial infections, other fungal infections, nonspecific interstitial pneumonitis, KS, and lymphoma.

Bacterial pneumonia is seen with an increased frequency in patients with HIV infection, with 0.8–2.0 cases per 100 person-years. Patients with HIV infection are particularly prone to infections with encapsulated organisms. *S. pneumoniae* (Chap. 134) and *H. influenzae* (Chap. 145) are responsible for most cases of bacterial pneumonia in patients with AIDS. This may be a consequence of altered B cell function and/or defects in neutrophil function that may be secondary to HIV disease (see above). Pneumonias due to *S. aureus* (Chap. 135) and *P. aeruginosa* (Chap. 152) are also reported to occur with an increased

frequency in patients with HIV infection. *S. pneumoniae* (pneumococcal) infection may be the earliest serious infection to occur in patients with HIV disease. This can present as pneumonia, sinusitis, and/or bacteremia. Patients with untreated HIV infection have a sixfold increase in the incidence of pneumococcal pneumonia and a 100-fold increase in the incidence of pneumococcal bacteremia. Pneumococcal disease may be seen in patients with relatively intact immune systems. In one study, the baseline CD4+ T cell count at the time of a first episode of pneumococcal pneumonia was ~300/μL. Of interest is the fact that the inflammatory response to pneumococcal infection appears proportional to the CD4+ T cell count. Due to this high risk of pneumococcal disease, immunization with pneumococcal polysaccharide is one of the generally recommended prophylactic measures for patients with HIV infection. This is likely most effective if given while the CD4+ T cell count is >200/μL and, if given to patients with lower CD4+ T cell counts, should be repeated once the count has been above 200 for 6 months. Although clear guidelines do not exist, it also makes sense to repeat immunization every 5 years. The incidence of bacterial pneumonia is cut in half when patients quit smoking. *Pneumocystis* pneumonia (PCP), once the hallmark of AIDS, has dramatically declined in incidence following the development of effective prophylactic regimens and the widespread use of cART. It is, however, still the single most common cause of pneumonia in patients with HIV infection in the United States and can be identified as a likely etiologic agent in 25% of cases of pneumonia in patients with HIV infection, with an incidence in the range of 2–3 cases per 100 person-years. Approximately 50% of cases of HIV-associated PCP occur in patients who are unaware of their HIV status. The risk of PCP is greatest among those who have experienced a previous bout of PCP and those who have CD4+ T cell counts of <200/μL. Overall, 79% of patients with PCP have CD4+ T cell counts <100/μL and 95% of patients have CD4+ T cell counts <200/μL. Recurrent fever, night sweats, thrush, and unexplained weight loss are also associated with an increased incidence of PCP. For these reasons, it is strongly recommended that all patients with CD4+ T cell counts <200/μL (or a CD4 percentage <15) receive some form of PCP prophylaxis. The incidence of PCP is approaching zero in patients with known HIV infection receiving appropriate cART and prophylaxis. In the United States, primary PCP is now occurring at a median CD4+ T cell count of 36/μL, while secondary PCP is occurring at a median CD4+ T cell count of 10/μL. Patients with PCP generally present with fever and a cough that is usually nonproductive or productive of only scant amounts of white sputum. They may complain of a characteristic retrosternal chest pain that is worse on inspiration and is described as sharp or burning. HIV-associated PCP may have an indolent course characterized by weeks of vague symptoms and should be included in the differential diagnosis of fever, pulmonary complaints, or unexplained weight loss in any patient with HIV infection and <200 CD4+ T cells/μL. The most common finding on chest x-ray is either a normal film, if the disease is suspected early, or a faint bilateral interstitial infiltrate. The classic finding of a dense perihilar infiltrate is unusual in patients with AIDS. In patients with PCP who have been receiving aerosolized pentamidine for prophylaxis, one may see an x-ray picture of upper lobe cavitary disease, reminiscent of TB. Other less common findings on chest x-ray include lobar infiltrates and pleural effusions. Thin-section CT may demonstrate a patchy ground-glass appearance. Routine laboratory evaluation is usually of little help in the differential diagnosis of PCP. A mild leukocytosis is common, although this may not be obvious in patients with prior neutropenia. Elevation of lactate dehydrogenase is common. Arterial blood-gases may indicate hypoxemia with a decline in Pa_{O2} and an increase in the arterial-alveolar (a–A) gradient. Arterial blood-gas measurements not only aid in making the diagnosis of PCP but also provide important information for staging the severity of the disease and directing treatment (see below). A definitive diagnosis of PCP requires demonstration of the organism in samples obtained from induced sputum, bronchoalveolar lavage, transbronchial biopsy, or open-lung biopsy. PCR has been used to detect specific DNA sequences for *P. jiroveci* in clinical specimens where histologic examinations have failed to make a diagnosis.

In addition to pneumonia, a number of other clinical problems have been reported in HIV-infected patients as a result of infection with *P. jiroveci*. Otic involvement may be seen as a primary infection, presenting as a polypoid mass involving the external auditory canal. In patients receiving aerosolized pentamidine for prophylaxis against PCP, one may see a variety of extrapulmonary manifestations of *P. jiroveci*. These include ophthalmic lesions of the choroid, a necrotizing vasculitis that resembles Burger's disease, bone marrow hypoplasia, and intestinal obstruction. Other organs that have been involved include lymph nodes, spleen, liver, kidney, pancreas, pericardium, heart, thyroid, and adrenals. Organ infection may be associated with cystic lesions that may appear calcified on CT or ultrasound.

The standard treatment for PCP or disseminated pneumocystosis is trimethoprim/sulfamethoxazole (TMP/SMX). A high (20–85%) incidence of side effects, particularly skin rash and bone marrow suppression, is seen with TMP/SMX in patients with HIV infection. Alternative treatments for mild to moderate PCP include dapsone/trimethoprim, clindamycin/primaquine, and atovaquone. IV pentamidine is the treatment of choice for severe disease in the patient unable to tolerate TMP/SMX. For patients with a Pa_{O2} <70 mmHg or with an a–a gradient >35 mmHg, adjunct glucocorticoid therapy should be used in addition to specific antimicrobials. Overall, treatment should be continued for 21 days and followed by secondary prophylaxis. Prophylaxis for PCP is indicated for any HIV-infected individual who has experienced a prior bout of PCP, any patient with a CD4+ T cell count of <200/μL or a CD4 percentage <15, any patient with unexplained fever for >2 weeks, and any patient with a history of oropharyngeal candidiasis. The preferred regimen for prophylaxis is TMP/SMX, one double-strength tablet daily. This regimen also provides protection against toxoplasmosis and some bacterial respiratory pathogens. For patients who cannot tolerate TMP/SMX, alternatives for prophylaxis include dapsone plus pyrimethamine plus leucovorin, aerosolized pentamidine administered by the Respirgard II nebulizer, and atovaquone. Primary or secondary prophylaxis for PCP can be discontinued in those patients treated with cART who maintain good suppression of HIV (<50 copies per milliliter) and CD4+ T cell counts >200/μL for 3–6 months.

M. tuberculosis, once thought to be on its way to extinction in the United States, experienced a resurgence associated with the HIV epidemic (Chap. 165). Worldwide, approximately one-third of all AIDS-related deaths are associated with TB, and TB is the primary cause of death for 10–15% of patients with HIV infection. In the United States ~5% of AIDS patients have active TB. Patients with HIV infection are more likely to have active TB by a factor of 100 when compared with an HIV-negative population. For an asymptomatic HIV-negative person with a positive purified protein derivative (PPD) skin test, the risk of reactivation TB is around 1% per year. For the patient with untreated HIV infection, a positive PPD skin test, and no signs or symptoms of TB, the rate of reactivation TB is 7–10% per year. Untreated TB can accelerate the course of HIV infection. Levels of plasma HIV RNA increase in the setting of active TB and decline in the setting of successful TB treatment. Active TB is most common in patients 25–44 years of age, in African Americans and Hispanics, in patients in New York City and Miami, and in patients in developing countries. In these demographic groups, 20–70% of the new cases of active TB are in patients with HIV infection. The epidemic of TB embedded in the epidemic of HIV infection probably represents the greatest health risk to the general public and the health care profession associated with the HIV epidemic. In contrast to infection with atypical mycobacteria such as MAC, active TB often develops relatively early in the course of HIV infection and may be an early clinical sign of HIV disease. In one study, the median CD4+ T cell count at presentation of TB was 326/μL. The clinical manifestations of TB in HIV-infected patients are quite varied and generally show different patterns as a function of the CD4+ T cell count. In patients with relatively high CD4+ T cell counts, the typical pattern of pulmonary reactivation occurs: patients present with fever, cough, dyspnea on exertion, weight loss, night sweats, and a chest x-ray revealing cavitary apical disease of the upper lobes. In patients with lower CD4+ T cell counts, disseminated disease is more common. In these patients the chest x-ray may reveal diffuse or lower lobe bilateral reticulonodular infiltrates consistent with miliary spread, pleural effusions, and hilar and/or mediastinal adenopathy. Infection may be present in bone, brain, meninges, GI tract, lymph nodes (particularly cervical lymph nodes), and viscera. Some patients with advanced HIV infection and active TB may have no symptoms of illness, and thus screening for TB should be part of the initial evaluation of every patient with HIV infection. Approximately 60–80% of HIV-infected patients with TB have pulmonary disease,

and 30–40% have extrapulmonary disease. Respiratory isolation and a negative-pressure room should be used for patients in whom a diagnosis of pulmonary TB is being considered. This approach is critical to limit nosocomial and community spread of infection. Culture of the organism from an involved site provides a definitive diagnosis. Blood cultures are positive in 15% of patients. This figure is higher in patients with lower CD4 +T cell counts. In the setting of fulminant disease one cannot rely on the accuracy of a negative PPD skin test to rule out a diagnosis of TB. TB is one of the conditions associated with HIV infection for which cure is possible with appropriate therapy. Therapy for TB is generally the same in the HIV-infected patient as in the HIV-negative patient (Chap. 165). Due to the possibility of multidrug-resistant or extensively drug-resistant TB, drug susceptibility testing should be performed to guide therapy. Due to pharmacokinetic interactions, adjusted doses of rifabutin should be substituted for rifampin in patients receiving the HIV protease inhibitors or nonnucleoside reverse transcriptase inhibitors. Treatment is most effective in programs that involve directly observed therapy. Initiation of cART and/or anti-TB therapy may be associated with clinical deterioration due to immune reconstitution inflammatory syndrome (IRIS) reactions. These are most common in patients initiating both treatments at the same time, may occur as early as 1 week after initiation of therapy, and are seen more frequently in patients with advanced HIV disease. For these reasons it is often recommended that initiation of cART be delayed in antiretroviral-naïve patients until 2–8 weeks following the initiation of treatment for TB. Effective prevention of active TB can be a reality if the health care professional is aggressive in looking for evidence of latent or active TB by making sure that all patients with HIV infection receive a PPD skin test or evaluation with an IFN-γ release assay. Anergy testing is not of value in this setting. Since these tests rely on the host mounting an immune response to *M. tuberculosis*, patients with CD4+ T cell counts <200 cells/μL should be retested if their CD4+ T cell counts rise to persistently above 200. Patients at risk of continued exposure to TB should be tested annually. HIV-infected individuals with a skin-test reaction of >5 mm, those with a positive IFN-γ release assay, or those who are close household contacts of persons with active TB should receive treatment with 9 months of isoniazid and pyridoxine.

Atypical mycobacterial infections are also seen with an increased frequency in patients with HIV infection. Infections with at least 12 different mycobacteria have been reported, including *M. bovis* and representatives of all four Runyon groups. The most common atypical mycobacterial infection is with *M. avium* or *M. intracellulare* species—the *Mycobacterium avium* complex (MAC). Infections with MAC are seen mainly in patients in the United States and are rare in Africa. It has been suggested that prior infection with *M. tuberculosis* decreases the risk of MAC infection. MAC infections probably arise from organisms that are ubiquitous in the environment, including both soil and water. There is little evidence for person-to-person transmission of MAC infection. The presumed portals of entry are the respiratory and GI tract. MAC infection is a late complication of HIV infection, occurring predominantly in patients with CD4+ T cell counts of <50/μL. The average CD4+ T cell count at the time of diagnosis is 10/μL. The most common presentation is disseminated disease with fever, weight loss, and night sweats. At least 85% of patients with MAC infection are mycobacteremic, and large numbers of organisms can often be demonstrated on bone marrow biopsy. The chest x-ray is abnormal in ~25% of patients, with the most common pattern being that of a bilateral, lower lobe infiltrate suggestive of miliary spread. Alveolar or nodular infiltrates and hilar and/or mediastinal adenopathy can also occur. Other clinical findings include endobronchial lesions, abdominal pain, diarrhea, and lymphadenopathy. Anemia and elevated liver alkaline phosphatase are common. The diagnosis is made by the culture of blood or involved tissue. The finding of two consecutive sputum samples positive for MAC is highly suggestive of pulmonary infection. Cultures may take 2 weeks to turn positive. Therapy consists of a macrolide, usually clarithromycin, with ethambutol. Some physicians elect to add a third drug from among rifabutin, ciprofloxacin, or amikacin in patients with extensive disease. Therapy was generally for life; however, with the use of cART it is possible to discontinue therapy in patients with sustained suppression of HIV replication and CD4+ T cell counts >100/μL for 3–6 months. Primary prophylaxis for MAC is indicated in patients with HIV infection and CD4+ T cell counts <50/μL (Table 189-10). This may be discontinued in patients in whom cART induces a sustained suppression of viral replication and increases in CD4+ T cell counts to >100/μL for ≥3 months.

Rhodococcus equi is a gram-positive, pleomorphic, acid-fast, non-spore-forming bacillus that can cause pulmonary and/or disseminated infection in patients with advanced HIV infection. Fever and cough are the most common presenting signs. Radiographically one may see cavitary lesions and consolidation. Blood cultures are often positive. Treatment is based on antimicrobial sensitivity testing.

Fungal infections of the lung, in addition to PCP, can be seen in patients with AIDS. Patients with pulmonary cryptococcal disease present with fever, cough, dyspnea, and, in some cases, hemoptysis. A focal or diffuse interstitial infiltrate is seen on chest x-ray in >90% of patients. In addition, one may see lobar disease, cavitary disease, pleural effusions, and hilar or mediastinal adenopathy. More than half of patients are fungemic, and 90% of patients have concomitant CNS infection. *Coccidioides immitis* is a mold that is endemic in the southwest United States. It can cause a reactivation pulmonary syndrome in patients with HIV infection. Most patients with this condition will have CD4+ T cell counts <250/μL. Patients present with fever, weight loss, cough, and extensive, diffuse reticulonodular infiltrates on chest x-ray. One may also see nodules, cavities, pleural effusions, and hilar adenopathy. While serologic testing is of value in the immunocompetent host, serologies are negative in 25% of HIV-infected patients with coccidioidal infection. Invasive aspergillosis is not an AIDS-defining illness and is generally not seen in patients with AIDS in the absence of neutropenia or administration of glucocorticoids. When it does occur, *Aspergillus* infection may have an unusual presentation in the respiratory tract of patients with AIDS, where it gives the appearance of a pseudomembranous tracheobronchitis. Primary pulmonary infection of the lung may be seen with *histoplasmosis*. The most common pulmonary manifestation of histoplasmosis, however, is in the setting of disseminated disease, presumably due to reactivation. In this setting respiratory symptoms are usually minimal, with cough and dyspnea occurring in 10–30% of patients. The chest x-ray is abnormal in ~50% of patients, showing either a diffuse interstitial infiltrate or diffuse small nodules.

Two forms of *idiopathic interstitial pneumonia* have been identified in patients with HIV infection: lymphoid interstitial pneumonitis (LIP) and nonspecific interstitial pneumonitis (NIP). LIP, a common finding in children, is seen in about 1% of adult patients with untreated HIV infection. This disorder is characterized by a benign infiltrate of the lung and is thought to be part of the polyclonal activation of lymphocytes seen in the context of HIV and EBV infections. Transbronchial biopsy is diagnostic in 50% of the cases, with an open-lung biopsy required for diagnosis in the remainder of cases. This condition is generally self-limited and no specific treatment is necessary. Severe cases have been managed with brief courses of glucocorticoids. Although rarely a clinical problem since the use of cART, evidence of NIP may be seen in up to half of all patients with untreated HIV infection. Histologically, interstitial infiltrates of lymphocytes and plasma cells in a perivascular and peribronchial distribution are present. When symptomatic, patients present with fever and nonproductive cough occasionally accompanied by mild chest discomfort. Chest x-ray is usually normal or may reveal a faint interstitial pattern. Similar to LIP, NIP is a self-limited process for which no therapy is indicated other than appropriate management of the underlying HIV infection. HIV-related pulmonary arterial hypertension (HIV-PAH) is seen in ~0.5% of HIV-infected individuals. Patients may present with an array of symptoms including shortness of breath, fatigue, syncope, chest pain, and signs of right-sided heart failure. Chest x-ray reveals dilated pulmonary vessels and right-sided cardiomegaly with right ventricular hypertrophy seen on electrocardiogram. cART does not appear to be of clear benefit, and the prognosis is quite poor with a median survival in the range of 2 years.

Neoplastic diseases of the lung including KS and lymphoma are discussed below in the section on neoplastic diseases.

DISEASES OF THE CARDIOVASCULAR SYSTEM

Heart disease is a relatively common postmortem finding in HIV-infected patients (25–75% in autopsy series). The most common form of heart disease is coronary heart disease. In one large series the overall rate of myocardial infarction (MI) was 3.5/1000

patient-years, 28% of these events were fatal, and MI was responsible for 7% of all deaths in the cohort. In patients with HIV infection, cardiovascular disease may be associated with classic risk factors such as smoking, a direct consequence of HIV infection, or a complication of cART. Patients with HIV infection have higher levels of triglycerides, lower levels of high-density lipoprotein cholesterol, and a higher prevalence of smoking than cohorts of individuals without HIV infection. The finding that the rate of cardiovascular disease events was lower in patients on antiretroviral therapy than in those randomized to undergo a treatment interruption identified a clear association between HIV replication and risk of cardiovascular disease. In one study, a baseline CD4+ T cell count of <500/μL was found to be an independent risk factor for cardiovascular disease comparable in magnitude to that attributable to smoking. While the precise pathogenesis of this association remains unclear, it is likely related to the immune activation and increased propensity for coagulation seen as a consequence of HIV replication. Exposure to HIV protease inhibitors and certain reverse transcriptase inhibitors has been associated with increases in total cholesterol and/or risk of MI. Any increases in the risk of death from MI resulting from the use of certain antiretrovirals must be balanced against the marked increases in overall survival brought about by these drugs.

Another form of heart disease associated with HIV infection is a dilated cardiomyopathy associated with congestive heart failure (CHF) referred to as *HIV-associated cardiomyopathy*. This generally occurs as a late complication of HIV infection and, histologically, displays elements of myocarditis. For this reason some have advocated treatment with IV immunoglobulin (IVIg). HIV can be directly demonstrated in cardiac tissue in this setting, and there is debate over whether it plays a direct role in this condition. Patients present with typical findings of CHF including edema and shortness of breath. Patients with HIV infection may also develop cardiomyopathy as side effects of IFN-α or nucleoside analogue therapy. These are reversible once therapy is stopped. KS, cryptococcosis, Chagas' disease, and toxoplasmosis can involve the myocardium, leading to cardiomyopathy. In one series, most patients with HIV infection and a treatable myocarditis were found to have myocarditis associated with toxoplasmosis. Most of these patients also had evidence of CNS toxoplasmosis. Thus, MRI or double-dose contrast CT scan of the brain should be included in the workup of any patient with advanced HIV infection and cardiomyopathy.

A variety of other cardiovascular problems are found in patients with HIV infection. Pericardial effusions may be seen in the setting of advanced HIV infection. Predisposing factors include TB, CHF, mycobacterial infection, cryptococcal infection, pulmonary infection, lymphoma, and KS. While pericarditis is quite rare, in one series 5% of patients with HIV disease had pericardial effusions that were considered to be moderate or severe. Tamponade and death have occurred in association with pericardial KS, presumably owing to acute hemorrhage. Nonbacterial thrombotic endocarditis has been reported and should be considered in patients with unexplained embolic phenomena. Intravenous pentamidine, when given rapidly, can result in hypotension as a consequence of cardiovascular collapse.

DISEASES OF THE OROPHARYNX AND GASTROINTESTINAL SYSTEM

Oropharyngeal and GI diseases are common features of HIV infection. They are most frequently due to secondary infections. In addition, oral and GI lesions may occur with KS and lymphoma.

Oral lesions, including *thrush*, *hairy leukoplakia*, and *aphthous ulcers*(**Fig. 189-35)**, are particularly common in patients with untreated HIV infection. Thrush, due to *Candida* infection, and oral hairy leukoplakia, presumed due to EBV, are usually indicative of fairly advanced immunologic decline; they generally occur in patients with CD4+ T cell counts of <300/μL. In one study, 59% of patients with oral candidiasis went on to develop AIDS in the next year. Thrush appears as a white, cheesy exudate, often on an erythematous mucosa in the posterior oropharynx. While most commonly seen on the soft palate, early lesions are often found along the gingival border. The diagnosis is made by direct examination of a scraping for pseudohyphal elements. Culturing is of no diagnostic value, as patients with HIV infection may have a positive throat culture for *Candida* in the absence of thrush. Oral hairy leukoplakia presents as white, frondlike lesions, generally along the lateral borders of the tongue and sometimes on the adjacent buccal mucosa (Fig. 189-35). Despite its name, oral hairy leukoplakia is not considered a premalignant condition. Lesions are associated with florid replication of EBV. While usually more disconcerting as a sign of HIV-associated immunodeficiency than a clinical problem in need of treatment, severe cases have been reported to respond to topical podophyllin or systemic therapy with anti-herpesvirus agents. Aphthous ulcers of the posterior oropharynx are also seen with regularity in patients with HIV infection (Fig. 189-35). These lesions are of unknown etiology and can be quite painful and interfere with swallowing. Topical anesthetics provide immediate symptomatic relief of short duration. The fact that thalidomide is an effective treatment for this condition suggests that the pathogenesis may involve the action of tissue-destructive cytokines. Palatal, glossal, or gingival ulcers may also result from cryptococcal disease or histoplasmosis.

Figure 189-35

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

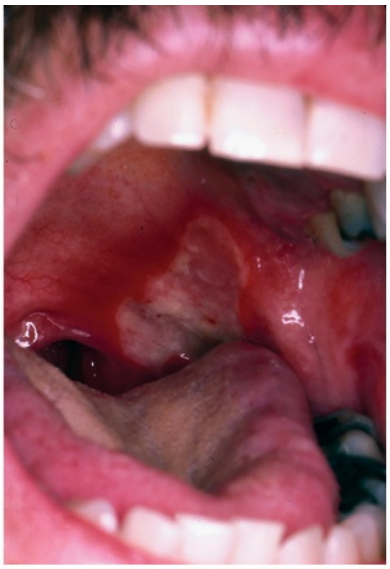

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

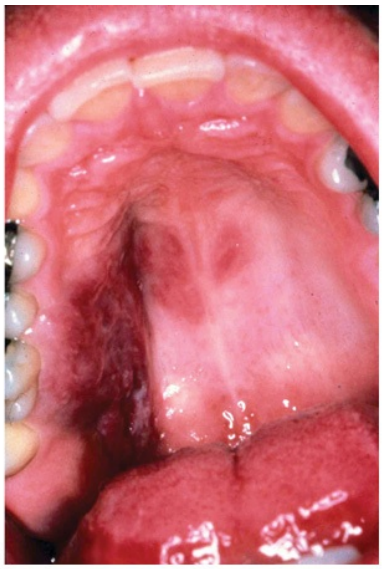

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Various oral lesions in HIV-infected individuals. ***A.*** Thrush. ***B.*** Hairy leukoplakia. ***C.*** Aphthous ulcer. ***D.*** Kaposi's sarcoma.

Esophagitis **(Fig. 189-36)** may present with odynophagia and retrosternal pain. Upper endoscopy is generally required to make an accurate diagnosis. Esophagitis may be due to *Candida*, CMV, or HSV. While CMV tends to be associated with a single large ulcer, HSV infection is more often associated with multiple small ulcers. The esophagus may also be the site of KS and lymphoma. Like the oral mucosa, the esophageal mucosa may have large, painful ulcers of unclear etiology that may respond to thalidomide. While achlorhydria is a common problem in patients with HIV infection, other gastric problems are generally rare. Among the neoplastic conditions involving the stomach are KS and lymphoma.

Figure 189-36

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Barium swallow of a patient with *Candida*esophagitis. The flow of barium along the mucosal surface is grossly irregular.

Infections of the small and large intestine leading to diarrhea, abdominal pain, and occasionally fever are among the most significant GI problems in HIV-infected patients. They include infections with bacteria, protozoa, and viruses.

Bacteria may be responsible for secondary infections of the GI tract. Infections with enteric pathogens such as *Salmonella*, *Shigella*, and *Campylobacter* are more common in homosexual men and are often more severe and more apt to relapse in patients with HIV infection. Patients with untreated HIV have approximately a 20-fold increased risk of infection with *S. typhimurium*. They may present with a variety of nonspecific symptoms including fever, anorexia, fatigue, and malaise of several weeks' duration. Diarrhea is common but may be absent. Diagnosis is made by culture of blood and stool. Long-term therapy with ciprofloxacin is the recommended treatment. HIV-infected patients also have an increased incidence of *S. typhi* infection in areas of the world where typhoid is a problem. *Shigella* spp., particularly *S. flexneri*, can cause severe intestinal disease in HIV-infected individuals. Up to 50% of patients will develop bacteremia. *Campylobacter* infections occur with an increased frequency in patients with HIV infection. While *C. jejuni* is the strain most frequently isolated, infections with many other strains have been reported. Patients usually present with crampy abdominal pain, fever, and bloody diarrhea. Infection may also present as proctitis. Stool examination reveals the presence of fecal leukocytes. Systemic infection can occur, with up to 10% of infected patients exhibiting bacteremia. Most strains are sensitive to erythromycin. Abdominal pain and diarrhea may be seen with MAC infection.

Fungal infections may also be a cause of diarrhea in patients with HIV infection. Histoplasmosis, coccidioidomycosis, and penicilliosis have all been identified as a cause of fever and diarrhea in patients with HIV infection. Peritonitis has been seen with *C. immitis*.

Cryptosporidia, microsporidia, and *Isospora belli* (Chap. 215) are the most common opportunistic protozoa that infect the GI tract and cause diarrhea in HIV-infected patients. Cryptosporidial infection may present in a variety of ways, ranging from a self-limited or intermittent diarrheal illness in patients in the early stages of HIV infection to a severe, life-threatening diarrhea in severely immunodeficient individuals. In patients with untreated HIV infection and CD4+ T cell counts of <300/μL, the incidence of cryptosporidiosis is ~1% per year. In 75% of cases the diarrhea is accompanied by crampy abdominal pain, and 25% of patients have nausea and/or vomiting. Cryptosporidia may also cause biliary tract disease in the HIV-infected patient, leading to cholecystitis with or without accompanying cholangitis and pancreatitis secondary to papillary stenosis. The diagnosis of cryptosporidial diarrhea is made by stool examination or biopsy of the small intestine. The diarrhea is noninflammatory, and the characteristic finding is the presence of oocysts that stain with acid-fast dyes. Therapy is predominantly supportive, and marked improvements have been reported in the setting of effective cART. Treatment with up to 2000 mg/d of nitazoxanide (NTZ) is associated with improvement in symptoms or a decrease in shedding of organisms in about half of patients. Its overall role in the management of this condition remains unclear. Patients can minimize their risk of developing cryptosporidiosis by avoiding contact with human and animal feces, by not drinking untreated water from lakes or rivers, and by not eating raw shellfish.

Microsporidia are small, unicellular, obligate intracellular parasites that reside in the cytoplasm of enteric cells (Chap. 215). The main species causing disease in humans is *Enterocytozoon bieneusi*. The clinical manifestations are similar to those described for cryptosporidia and include abdominal pain, malabsorption, diarrhea, and cholangitis. The small size of the organism may make it difficult to detect; however, with the use of chromotrope-based stains, organisms can be identified in stool samples by light microscopy. Definitive diagnosis generally depends on electron-microscopic examination of a stool specimen, intestinal aspirate, or intestinal biopsy specimen. In contrast to cryptosporidia, microsporidia have been noted in a variety of extraintestinal locations, including the eye, brain, sinuses, muscle, and liver, and they have been associated with conjunctivitis and hepatitis. The most effective way to deal with microsporia in a patient with HIV infection is to restore the immune system by treating the HIV infection with cART. Albendazole, 400 mg bid, has been reported to be of benefit in some patients.

I. belli is a coccidian parasite (Chap. 215) most commonly found as a cause of diarrhea in patients from tropical and subtropical regions. Its cysts appear in the stool as large, acid-fast structures that can be differentiated from those of cryptosporidia on the basis of size, shape, and number of sporocysts. The clinical syndromes of *Isospora* infection are identical to those caused by cryptosporidia. The important distinction is that infection with *Isospora* is generally relatively easy to treat with TMP/SMX. While relapses are common, a thrice-weekly regimen of TMP/SMX appears adequate to prevent recurrence.

CMV colitis was once seen as a consequence of advanced immunodeficiency in 5–10% of patients with AIDS. It is much less common with the advent of cART. CMV colitis presents as diarrhea, abdominal pain, weight loss, and anorexia. The diarrhea is

usually nonbloody, and the diagnosis is achieved through endoscopy and biopsy. Multiple mucosal ulcerations are seen at endoscopy, and biopsies reveal characteristic intranuclear and cytoplasmic inclusion bodies. Secondary bacteremias may result as a consequence of thinning of the bowel wall. Treatment is with either ganciclovir or foscarnet for 3–6 weeks. Relapses are common, and maintenance therapy is typically necessary in patients whose HIV infection is poorly controlled. Patients with CMV disease of the GI tract should be carefully monitored for evidence of CMV retinitis.

In addition to disease caused by specific secondary infections, patients with HIV infection may also experience a chronic diarrheal syndrome for which no etiologic agent other than HIV can be identified. This entity is referred to as *AIDS enteropathy* or *HIV enteropathy*. It is most likely a direct result of HIV infection in the GI tract. Histologic examination of the small bowel in these patients reveals low-grade mucosal atrophy with a decrease in mitotic figures, suggesting a hyporegenerative state. Patients often have decreased or absent small-bowel lactase and malabsorption with accompanying weight loss.

The initial evaluation of a patient with HIV infection and diarrhea should include a set of stool examinations, including culture, examination for ova and parasites, and examination for *Clostridium difficile* toxin. Approximately 50% of the time this workup will demonstrate infection with pathogenic bacteria, mycobacteria, or protozoa. If the initial stool examinations are negative, additional evaluation, including upper and/or lower endoscopy with biopsy, will yield a diagnosis of microsporidial or mycobacterial infection of the small intestine ~30% of the time. In patients for whom this diagnostic evaluation is nonrevealing, a presumptive diagnosis of HIV enteropathy can be made if the diarrhea has persisted for >1 month. An algorithm for the evaluation of diarrhea in patients with HIV infection is given in **Fig. 189-37.**

Figure 189-37

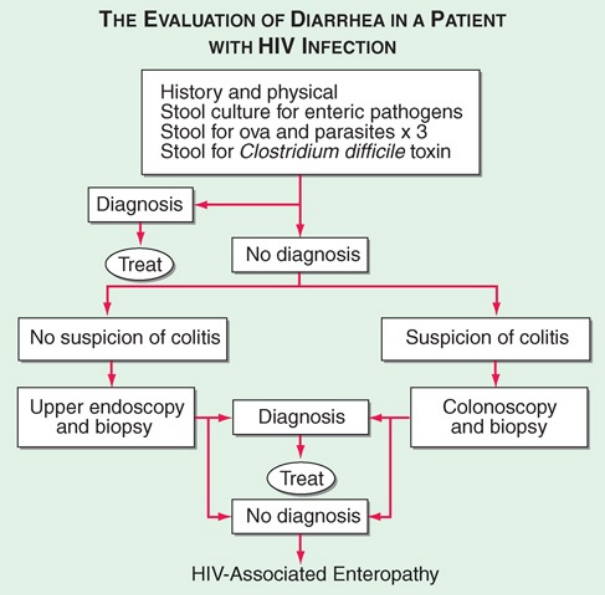

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the evaluation of diarrhea in a patient with HIV infection. HIV-associated enteropathy is a diagnosis of exclusion and can be made only after other, generally treatable, forms of diarrheal illness have been ruled out.

Rectal lesions are common in HIV-infected patients, particularly the perirectal ulcers and erosions due to the reactivation of HSV **(Fig. 189-38)**. These lesions may appear quite atypical, as denuded skin without vesicles, and they respond well to treatment with acyclovir, famciclovir, or foscarnet. Other rectal lesions encountered in patients with HIV infection include condylomata acuminata, KS, and intraepithelial neoplasia (see below).

Figure 189-38

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe, erosive perirectal herpes simplex in a patient with AIDS.

HEPATOBILIARY DISEASES

Diseases of the hepatobiliary system are a major problem in patients with HIV infection. It has been estimated that approximately one-third of the deaths of patients with HIV infection are in some way related to liver disease. While this is predominantly a reflection of the problems encountered in the setting of co-infection with hepatitis B or C, it is also a reflection of the hepatic injury, ranging from hepatic steatosis to hypersensitivity reactions to immune reconstitution, that can be seen in the context of cART.

The prevalence of co-infection with HIV and hepatitis viruses varies by geographic region. In the United States, ~90% of HIV-infected individuals have evidence of infection with HBV; 6–14% have chronic HBV infection; 5–50% of current or past patients are co-infected with HCV; and co-infection with hepatitis D, E, and/or G viruses is common. Among IV drug users with HIV infection, rates of HCV infection range from 70 to 95%. HIV infection has a significant impact on the course of hepatitis virus infection. It is associated with approximately a threefold increase in the development of persistent hepatitis B surface antigenemia. Patients infected with both HBV and HIV have decreased evidence of inflammatory liver disease. The presumption that this is due to the immunosuppressive effects of HIV infection is supported by the observations that this situation can be reversed, and one may see the development of more severe hepatitis following the initiation of effective cART. In studies of the impact of HIV on HBV infection, four- to tenfold increases in liver-related mortality rates have been noted in patients with HIV and active HBV infection compared to rates in patients with either infection alone. There is, however, only a slight increase in overall mortality rate in HIV-infected individuals who are also hepatitis B surface antigen (HBsAg)–positive. IFN-α is less successful as a treatment of HBV in patients with HIV co-infection. Lamivudine, emtricitabine, adefovir/tenofovir/entecavir, and telbivudine alone or in combination are useful in the treatment of hepatitis B in patients with HIV infection. It is important to remember that all the above-mentioned drugs also have activity against HIV and should not be used alone in patients with HIV infection, in order to avoid the emergence quasispecies of HIV resistant to these drugs. For this reason, the need to treat hepatitis B infection in a patient with HIV infection is an indication to treat HIV infection in that same patient, regardless of CD4+ T cell count. HCV infection is more severe in the patient with HIV infection; it does not appear to affect overall mortality rate in HIV-infected individuals when other variables such as age, baseline CD4+ T cell count, and use of cART are taken into account. In the setting of HIV and HCV co-infection, levels of HCV are approximately tenfold higher than in the HIV-negative patient with HCV infection and there is a tenfold increased risk of death due to liver disease in co-infected patients. Treatment for HCV infection consists of pegylated IFN-α and ribavirin with an array of experimental therapies currently in clinical trials. If a 2-log drop in levels of HCV RNA is not seen within 12 weeks, it is unlikely that therapy will be of value. Hepatitis A virus infection is not seen with an increased frequency in patients with HIV infection. It is recommended that all patients with HIV infection who have not experienced natural infection be immunized with hepatitis A and/or hepatitis B vaccines. Infection with hepatitis G virus, also known as GB virus C, is seen in ~50% of patients with HIV infection. For reasons that are currently unclear, there are data to suggest that patients with HIV infection co-infected with this virus have a decreased rate of progression to AIDS.

A variety of other infections may also involve the liver. Granulomatous hepatitis may be seen as a consequence of mycobacterial or fungal infections, particularly MAC infection. Hepatic masses may be seen in the context of TB, peliosis hepatis, or fungal infection. Among the fungal opportunistic infections, *C. immitis* and *Histoplasma capsulatum* are those most likely to involve the liver. Biliary tract disease in the form of papillary stenosis or sclerosing cholangitis has been reported in the context of cryptosporidiosis, CMV infection, and KS.

Many of the drugs used to treat HIV infection are metabolized by the liver and can cause liver injury. Fatal hepatic reactions have been reported with a wide array of antiretrovirals including nucleoside analogues, nonnucleoside analogues, and protease inhibitors. Nucleoside analogues work by inhibiting DNA synthesis. This can result in toxicity to mitochondria, which can lead to disturbances in oxidative metabolism. This may manifest as hepatic steatosis and, in severe cases, lactic acidosis and fulminant liver failure. It is important to be aware of this condition and to watch for it in patients with HIV infection receiving nucleoside analogues. It is reversible if diagnosed early and the offending agent(s) discontinued. Nevirapine has been associated with at times fatal fulminant and cholestatic hepatitis, hepatic necrosis, and hepatic failure. Indinavir may cause mild to moderate elevations in serum bilirubin in 10–15% of patients in a syndrome similar to Gilbert's syndrome. A similar pattern of hepatic injury may be seen with atazanavir. In the patient receiving cART with an unexplained increase in hepatic transaminases, strong consideration should be given to drug toxicity. *Pancreatic injury* is most commonly a consequence of drug toxicity, notably that secondary to pentamidine or dideoxynucleosides. While up to half of patients in some series have biochemical evidence of pancreatic injury, <5% of patients show any clinical evidence of pancreatitis that is not linked to a drug toxicity.

DISEASES OF THE KIDNEY AND GENITOURINARY TRACT

Diseases of the kidney or genitourinary tract may be a direct consequence of HIV infection, due to an opportunistic infection or neoplasm, or related to drug toxicity. Overall, microalbuminuria is seen in ~20% of untreated HIV-infected patients; significant

proteinuria is seen in closer to 2%. The presence of microalbuminuria has been associated with an increase in all-cause mortality rate. *HIV-associated nephropathy* (HIVAN) was first described in IDUs and was initially thought to be IDU nephropathy in patients with HIV infection; it is now recognized as a true direct complication of HIV infection. Although the majority of patients have CD4+ T cell counts <200/μL, HIV-associated nephropathy can be an early manifestation of HIV infection and is also seen in children. Over 90% of reported cases have been in African-American or Hispanic individuals; the disease is not only more prevalent in these populations but also more severe and is the third leading cause of end-stage renal failure among African Americans age 20–64 in the United States. Proteinuria is the hallmark of this disorder. Edema and hypertension are rare. Ultrasound examination reveals enlarged, hyperechogenic kidneys. A definitive diagnosis is obtained through renal biopsy. Histologically, focal segmental glomerulosclerosis is present in 80%, and mesangial proliferation in 10–15% of cases. Prior to effective antiretroviral therapy, this disease was characterized by relatively rapid progression to end-stage renal disease. Patients with HIV-associated nephropathy should be treated for their HIV infection regardless of CD4+ T cell count. Treatment with angiotensin-converting enzyme (ACE) inhibitors and/or prednisone, 60 mg/d, has also been reported to be of benefit in some cases. The incidence of this disease in patients receiving adequate cART has not been well defined; however, the impression is that it has decreased in frequency and severity. It is the leading cause of end-stage renal disease in patients with HIV infection.

Among the drugs commonly associated with renal damage in patients with HIV disease are pentamidine, amphotericin, adefovir, cidofovir, tenofovir, and foscarnet. TMP/SMX may compete for tubular secretion with creatinine and cause an increase in the serum creatinine level. Sulfadiazine may crystallize in the kidney and result in an easily reversible form of renal shutdown, while indinavir may form renal calculi. Adequate hydration is the mainstay of treatment and prevention for these latter two conditions.

Genitourinary tract infections are seen with a high frequency in patients with HIV infection; they present with skin lesions, dysuria, hematuria, and/or pyuria and are managed in the same fashion as in patients without HIV infection. Infections with HSV are covered below ("Dermatologic Diseases"). Infections with *T. pallidum*, the etiologic agent of *syphilis*, play an important role in the HIV epidemic. In HIV-negative individuals, genital syphilitic ulcers as well as the ulcers of chancroid are major predisposing factors for heterosexual transmission of HIV infection. While most HIV-infected individuals with syphilis have a typical presentation, a variety of formerly rare clinical problems may be encountered in the setting of dual infection. Among them are *lues maligna*, an ulcerating lesion of the skin due to a necrotizing vasculitis; unexplained fever; nephrotic syndrome; and neurosyphilis. The most common presentation of syphilis in the HIV-infected patient is that of *condylomata lata*, a form of secondary syphilis. Neurosyphilis may be asymptomatic or may present as acute meningitis, neuroretinitis, deafness, or stroke. The rate of neurosyphilis may be as high as 1% in patients with HIV infection, and one should consider a lumbar puncture to look for neurosyphilis in all patients with HIV infection and secondary syphilis. As a consequence of the immunologic abnormalities seen in the setting of HIV infection, diagnosis of syphilis through standard serologic testing may be challenging. On the one hand, a significant number of patients have false-positive Venereal Disease Research Laboratory (VDRL) tests due to polyclonal B cell activation. On the other hand, the development of a new positive VDRL may be delayed in patients with new infections, and the anti–fluorescent treponemal antibody (anti-FTA) test may be negative due to immunodeficiency. Thus, dark-field examination of appropriate specimens should be performed in any patient in whom syphilis is suspected, even if the patient has a negative VDRL. Similarly, any patient with a positive serum VDRL test, neurologic findings, and an abnormal spinal fluid examination should be considered to have neurosyphilis and treated accordingly, regardless of the CSF VDRL result. In any setting, patients treated for syphilis need to be carefully monitored to ensure adequate therapy. Approximately one-third of patients with HIV infection will experience a Jarisch-Herxheimer reaction upon initiation of therapy for syphilis.

Vulvovaginal candidiasis is a common problem in women with HIV infection. Symptoms include pruritus, discomfort, dyspareunia, and dysuria. Vulvar infection may present as a morbilliform rash that may extend to the thighs. Vaginal infection is usually associated with a white discharge, and plaques may be seen along an erythematous vaginal wall. Diagnosis is made by microscopic examination of the discharge for pseudohyphal elements in a 10% potassium hydroxide solution. Mild disease can be treated with topical therapy. More serious disease can be treated with fluconazole. Other causes of vaginitis include *Trichomonas* and mixed bacteria.

DISEASES OF THE ENDOCRINE SYSTEM AND METABOLIC DISORDERS

A variety of endocrine and metabolic disorders are seen in the context of HIV infection. These may be a direct consequence of HIV infection, secondary to opportunistic infections or neoplasms, or related to medication side effects. Between 33 and 75% of patients with HIV infection receiving cART develop a syndrome often referred to as *lipodystrophy*, consisting of elevations in plasma triglycerides, total cholesterol, and apolipoprotein B, as well as hyperinsulinemia and hyperglycemia. Many of the patients have been noted to have a characteristic set of body habitus changes associated with fat redistribution, consisting of truncal obesity coupled with peripheral wasting **(Fig. 189-39)**. Truncal obesity is apparent as an increase in abdominal girth related to increases in mesenteric fat, a dorsocervical fat pad ("buffalo hump") reminiscent of patients with Cushing's syndrome, and enlargement of the breasts. The peripheral wasting, or lipoatrophy, is particularly noticeable in the face and buttocks and by the prominence of the veins in the legs. These changes may develop at any time ranging from ~6 weeks to several years following the initiation of cART. Approximately 20% of the patients with HIV-associated lipodystrophy meet the criteria for the *metabolic syndrome* as defined by The International Diabetes Federation or The U.S. National Cholesterol Education Program Adult Treatment Panel III. The lipodystrophy syndrome has been reported in association with regimens containing a variety of different drugs, and while initially reported in the setting of protease inhibitor therapy, it appears that similar changes can also be induced by potent protease-sparing regimens. It has been suggested that the lipoatrophy changes are particularly severe in patients receiving the thymidine analogues stavudine and zidovudine. National Cholesterol Education Program (NCEP) guidelines should be followed in the management of these lipid abnormalities (Chap. 241). Due to concerns regarding drug interactions, the most commonly utilized lipid-lowering agents in this setting are gemfibrozil and atorvastatin.

Figure 189-39

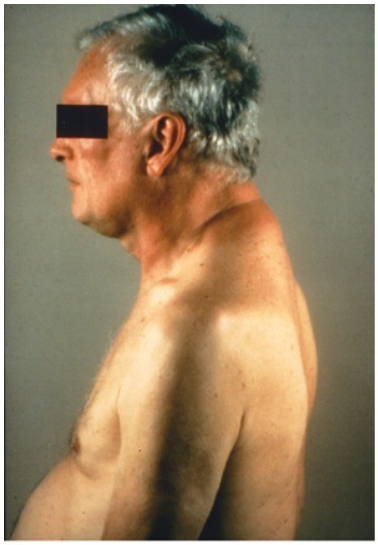

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

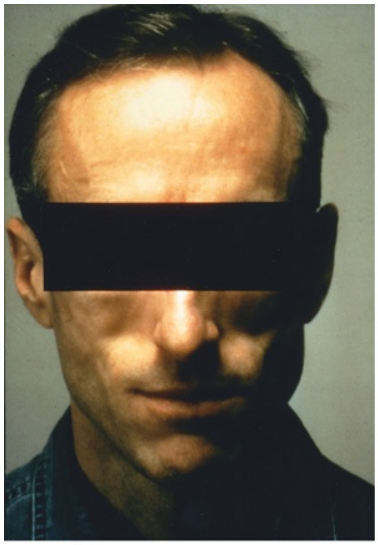

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Characteristics of lipodystrophy. ***A.*** Truncal obesity and buffalo hump. ***B.*** Facial wasting. ***C.*** Accumulation of intraabdominal fat on CT scan.

In addition to these abnormalities, patients with HIV infection treated with cART have been found to have an increased incidence of osteonecrosis or avascular necrosis of the hip and shoulders. In a study of asymptomatic patients, 4.4% were found to have evidence of osteonecrosis on MRI. This complication has been associated with the use of lipid-lowering agents, systemic

glucocorticoids, or testosterone; bodybuilding exercise; alcohol consumption; and the presence of anticardiolipin antibodies. Osteoporosis has been reported in 7% of women with HIV infection, with 41% of women demonstrating some degree of osteopenia. In addition, lactic acidosis is associated with cART. This is most commonly seen with nucleoside analogue reverse transcriptase inhibitors and can be fatal (see below).

Patients with advanced HIV disease may develop hyponatremia due to the syndrome of inappropriate antidiuretic hormone (vasopressin) secretion (SIADH) as a consequence of increased free-water intake and decreased free-water excretion. SIADH is usually seen in conjunction with pulmonary or CNS disease. Low serum sodium may also be due to adrenal insufficiency; a concomitant high serum potassium should alert one to this possibility. Hyperkalemia may be secondary to adrenal insufficiency; HIV nephropathy; or medications, particularly trimethoprim and pentamidine. Hypokalemia may be seen in the setting of tenofovir therapy. Adrenal gland disease may be due to mycobacterial infections, CMV disease, cryptococcal disease, histoplasmosis, or ketoconazole toxicity. Iatrogenic Cushing's syndrome with suppression of the hypothalamic-pituitary-adrenal axis may be seen with the use of local glucocorticoids (injected or inhaled) in patients receiving ritonavir. This is due to inhibition of the hepatic enzyme CYP3A4 by ritonavir leading to prolongation of the glucocorticoid half-life.

Thyroid function may be altered in 10–15% of patients with HIV infection. Both hypo- and hyperthyroidism may be seen. The predominant abnormality is subclinical hypothyroidism. In the setting of cART, up to 10% of patients have been noted to have elevated thyroid-stimulating hormone levels, suggesting that this may be a manifestation of immune reconstitution. Immune-reconstitution Graves' disease may occur as a late (9–48 months) complication of cART. In advanced HIV disease, infection of the thyroid gland may occur with opportunistic pathogens, including *P. jiroveci*, CMV, mycobacteria, *Toxoplasma gondii*, and *Cryptococcus neoformans*. These infections are generally associated with a nontender, diffuse enlargement of the thyroid gland. Thyroid function is usually normal. Diagnosis is made by fine-needle aspirate or open biopsy.

Depending on the severity of disease, HIV infection is associated with *hypogonadism* in 20–50% of men. While this is generally a complication of underlying illness, testicular dysfunction may also be a side effect of ganciclovir therapy. In some surveys, up to two-thirds of patients report decreased libido and one-third complain of erectile dysfunction. Androgen-replacement therapy should be considered in patients with symptomatic hypogonadism. HIV infection does not seem to have a significant effect on the menstrual cycle outside the setting of advanced disease.

IMMUNOLOGIC AND RHEUMATOLOGIC DISEASES

Immunologic and rheumatologic disorders are common in patients with HIV infection and range from excessive immediate-type hypersensitivity reactions (Chap. 317) to an increase in the incidence of reactive arthritis (Chap. 325) to conditions characterized by a diffuse infiltrative lymphocytosis. The occurrence of these phenomena is an apparent paradox in the setting of the profound immunodeficiency and immunosuppression that characterizes HIV infection and reflects the complex nature of the immune system and its regulatory mechanisms.

Drug allergies are the most significant allergic reactions occurring in HIV-infected patients and appear to become more common as the disease progresses. They occur in up to 65% of patients who receive therapy with TMP/SMX for PCP. In general, these drug reactions are characterized by erythematous, morbilliform eruptions that are pruritic, tend to coalesce, and are often associated with fever. Nonetheless, ~33% of patients can be maintained on the offending therapy, and thus these reactions are not an immediate indication to stop the drug. Anaphylaxis is extremely rare in patients with HIV infection, and patients who have a cutaneous reaction during a single course of therapy can still be considered candidates for future treatment or prophylaxis with the same agent. The one exception to this is the nucleoside analogue abacavir, where fatal hypersensitivity reactions have been reported with rechallenge. This hypersensitivity is strongly associated with the HLA-B5701 haplotype, and a hypersensitivity reaction to abacavir is an absolute contraindication to future therapy. For other agents, including TMP/SMX, desensitization regimens are moderately successful. While the mechanisms underlying these allergic-type reactions remain unknown, patients with HIV infection have been noted to have elevated IgE levels that increase as the CD4+ T cell count declines. The numerous examples of patients with multiple drug reactions suggest that a common pathway is involved.

HIV infection shares many similarities with a variety of autoimmune diseases, including a substantial polyclonal B cell activation that is associated with a high incidence of antiphospholipid antibodies, such as anticardiolipin antibodies, VDRL antibodies, and lupus-like anticoagulants. In addition, HIV-infected individuals have an increased incidence of antinuclear antibodies. Despite these serologic findings, there is no evidence that HIV-infected individuals have an increase in two of the more common autoimmune diseases, i.e., systemic lupus erythematosus and rheumatoid arthritis. In fact, it has been observed that these diseases may be somewhat ameliorated by the concomitant presence of HIV infection, suggesting that an intact CD4+ T cell limb of the immune response plays an integral role in the pathogenesis of these conditions. Similarly, there are anecdotal reports of patients with common variable immunodeficiency (Chap. 316), characterized by hypogammaglobulinemia, who have had a normalization of Ig levels following the development of HIV infection, suggesting a possible role for overactive CD4+ T cell immunity in certain forms of that syndrome. The one autoimmune disease that may occur with an increased frequency in patients with HIV infection is a variant of primary Sjögren's syndrome (Chap. 324). Patients with HIV infection may develop a syndrome consisting of parotid gland enlargement, dry eyes, and dry mouth that is associated with lymphocytic infiltrates of the salivary gland and lung. One also can see peripheral neuropathy, polymyositis, renal tubular acidosis, and hepatitis. In contrast to Sjögren's syndrome, in which the lymphocytic infiltrates are composed predominantly of CD4+ T cells, in patients with HIV infection the infiltrates are composed predominantly of CD8+ T cells. In addition, while patients with Sjögren's syndrome are mainly women who have autoantibodies to Ro and La and who frequently have HLA-DR3 or -B8 MHC haplotypes, HIV-infected individuals with this syndrome are usually African-American men who do not have anti-Ro or anti-La and who most often are HLA-DR5. This syndrome appears to be less common with the increased use of effective cART. The term *diffuse infiltrative lymphocytosis syndrome* (DILS) is used to describe this entity and to distinguish it from Sjögren's syndrome.

Approximately one-third of HIV-infected individuals experience arthralgias; furthermore, 5–10% are diagnosed as having some form of reactive arthritis, such as Reiter's syndrome or psoriatic arthritis as well as undifferentiated spondyloarthropathy (Chap. 325). These syndromes occur with increasing frequency as the competency of the immune system declines. This association may be related to an increase in the number of infections with organisms that may trigger a reactive arthritis with progressive immunodeficiency or to a loss of important regulatory T cells. Reactive arthritides in HIV-infected individuals generally respond well to standard treatment; however, therapy with methotrexate has been associated with an increase in the incidence of opportunistic infections and should be used with caution and only in severe cases.

HIV-infected individuals also experience a variety of joint problems without obvious cause that are referred to generically as *HIV-* or *AIDS-associated arthropathy*. This syndrome is characterized by subacute oligoarticular arthritis developing over a period of 1–6 weeks and lasting 6 weeks to 6 months. It generally involves the large joints, predominantly the knees and ankles, and is nonerosive with only a mild inflammatory response. X-rays of the joint are nonrevealing. Nonsteroidal anti-inflammatory drugs are only marginally helpful; however, relief has been noted with the use of intraarticular glucocorticoids. A second form of arthritis also thought to be secondary to HIV infection is called *painful articular syndrome*. This condition, reported as occurring in as many as 10% of AIDS patients, presents as an acute, severe, sharp pain in the affected joint. It affects primarily the knees, elbows, and shoulders; lasts 2–24 h; and may be severe enough to require narcotic analgesics. The cause of this arthropathy is unclear; however, it is thought to result from a direct effect of HIV on the joint. This condition is reminiscent of the fact that other lentiviruses, in particular the caprine arthritis-encephalitis virus, are capable of directly causing arthritis.

A variety of other immunologic or rheumatologic diseases have been reported in HIV-infected individuals, either de novo or in association with opportunistic infections or drugs. Using the criteria of widespread musculoskeletal pain of at least 3 months' duration and the presence of at least 11 of 18 possible tender points by digital palpation, 11% of an HIV-infected cohort containing 55% IDUs were diagnosed as having *fibromyalgia* (Chap. 335). While the incidence of frank arthritis was less in this population than in other studied populations that consisted predominantly of men who have sex with men, these data support the concept that there are musculoskeletal problems that occur as a direct result of HIV infection. In addition there have been reports of leukocytoclastic vasculitis in the setting of zidovudine therapy. CNS angiitis and polymyositis have also been reported in HIV-infected individuals. Septic arthritis is surprisingly rare, especially given the increased incidence of staphylococcal bacteremias seen in this population. When septic arthritis has been reported, it has usually been due to *Staphylococcus aureus*, systemic fungal infection with *C. neoformans*, *Sporothrix schenckii* , or *H. capsulatum* or to systemic mycobacterial infection with *M. tuberculosis*, *M. haemophilum*, *M. avium*, or *M. kansasii*.

As noted above, 4.4% of patients with HIV infection were found to have some evidence of osteonecrosis by MRI during systematic screening of asymptomatic patients. The percentage of patients with symptomatic osteonecrosis has been estimated to be as high as 1%. While this problem was first recognized in the setting of cART, it has been difficult to establish a cause-and-effect relationship. Alcohol consumption and a history of glucocorticoid use have been particularly associated with this condition in patients with HIV infection.

IMMUNE RECONSTITUTION INFLAMMATORY SYNDROME (IRIS)

Following the initiation of effective cART, a paradoxical worsening of preexisting, untreated, or partially treated opportunistic infections may be noted. One may also see exacerbations of pre-existing or the development of new autoimmune conditions following the initiation of antiretrovirals **(Table 189-11)**. IRIS related to a pre-existing infection is often referred to as immune reconstitution disease (IRD) to distinguish it from the autoimmune manifestations of IRIS. IRD is particularly common in patients with underlying untreated mycobacterial or fungal infections. IRIS is seen in anywhere from 10 to 30% of patients, depending on the clinical setting, and is most common in patients starting therapy with CD4+ T cell counts <50 cells/μL who have a precipitous drop in HIV RNA levels following the initiation of cART. Signs and symptoms may appear anywhere from 2 weeks to 2 years after the initiation of cART and can include localized lymphadenitis, prolonged fever, pulmonary infiltrates, increased intracranial pressure, uveitis, sarcoidosis, and Graves' disease. The clinical course can be protracted and severe cases can be fatal. The underlying mechanism appears to be related to a phenomenon similar to type IV hypersensitivity reactions and reflects the immediate improvements in immune function that occur as levels of HIV RNA drop and the immunosuppressive effects of HIV infection are controlled. In severe cases, the use of immunosuppressive drugs such as glucocorticoids may be required to blunt the inflammatory component of these reactions while specific antimicrobial therapy takes effect.

Table 189-11 Characteristics of Immune Reconstitution Inflammatory Syndrome (IRIS)

• Paradoxical worsening of clinical condition is seen following the initiation of antiretroviral therapy • Occurs weeks to months following the initiation of antiretroviral therapy
• Is most common in patients starting therapy with a CD4+ T cell count under 50/μL who experience a precipitous drop in viral load
• Is frequently seen in the setting of tuberculosis
• Can be fatal

DISEASES OF THE HEMATOPOIETIC SYSTEM

Disorders of the hematopoietic system including lymphadenopathy, anemia, leukopenia, and/or thrombocytopenia are common throughout the course of HIV infection and may be the direct result of HIV, manifestations of secondary infections and neoplasms, or side effects of therapy **(Table 189-12)**. Direct histologic examination and culture of lymph node or bone marrow tissue are often diagnostic. A significant percentage of bone marrow aspirates from patients with HIV infection have been reported to contain lymphoid aggregates, the precise significance of which is unknown. Initiation of cART will lead to reversal of most hematologic complications that are the direct result of HIV infection.

Table 189-12 Causes of Bone Marrow Suppression in Patients With HIV Infection

HIV infection	Medications
Mycobacterial infections	Zidovudine
Fungal infections	Dapsone
B19 parvovirus infection	Trimethoprim/sulfamethoxazole
Lymphoma	Pyrimethamine
	5-Flucytosine
	Ganciclovir
	Interferon α
	Trimetrexate
	Foscarnet

Some patients, otherwise asymptomatic, may develop *persistent generalized lymphadenopathy* as an early clinical manifestation of HIV infection. This condition is defined as the presence of enlarged lymph nodes (>1 cm) in two or more extrainguinal sites for >3 months without an obvious cause. The lymphadenopathy is due to marked follicular hyperplasia in the node in response to HIV infection. The nodes are generally discrete and freely movable. This feature of HIV disease may be seen at any point in the spectrum of immune dysfunction and is not associated with an increased likelihood of developing AIDS. Paradoxically, a loss in lymphadenopathy or a decrease in lymph node size outside the setting of cART may be a prognostic marker of disease progression. In patients with CD4+ T cell counts >200/μL, the differential diagnosis of lymphadenopathy includes KS, TB, Castleman's disease, and lymphoma. In patients with more advanced disease, lymphadenopathy may also be due to atypical mycobacterial infection, toxoplasmosis, systemic fungal infection, or bacillary angiomatosis. While indicated in patients with CD4+ T cell counts <200/μL, lymph node biopsy is not indicated in patients with early-stage disease unless there are signs and symptoms of systemic illness, such as fever and weight loss, or unless the nodes begin to enlarge, become fixed, or coalesce. Monoclonal gammopathy of unknown significance (MGUS) (Chap. 111), defined as the presence of a serum monoclonal IgG, IgA, or IgM in the absence of a clear cause, has been reported in 3% of patients with HIV infection. The overall clinical significance of this finding in patients with HIV infection is unclear, although it has been associated with other viral infections, non-Hodgkin's lymphoma, and plasma cell malignancy.

Anemia is the most common hematologic abnormality in HIV-infected patients and, in the absence of a specific treatable cause, is independently associated with a poor prognosis. While generally mild, anemia can be quite severe and require chronic blood transfusions. Among the specific reversible causes of anemia in the setting of HIV infection are drug toxicity, systemic fungal and mycobacterial infections, nutritional deficiencies, and parvovirus B19 infections. Zidovudine may block erythroid maturation prior to its effects on other marrow elements. A characteristic feature of zidovudine therapy is an elevated mean corpuscular volume (MCV). Another drug used in patients with HIV infection that has a selective effect on the erythroid series is dapsone. This drug can cause a serious hemolytic anemia in patients who are deficient in glucose-6-phosphate dehydrogenase and can create a functional anemia in others through induction of methemoglobinemia. Folate levels are usually normal in HIV-infected individuals; however, vitamin B12 levels may be depressed as a consequence of achlorhydria or malabsorption. True autoimmune hemolytic anemia is rare, although ~20% of patients with HIV infection may have a positive direct antiglobulin test as a consequence of polyclonal B cell activation. Infection with parvovirus B19 may also cause anemia. It is important to recognize this possibility given the fact that it responds well to treatment with IVIg. Erythropoietin levels in patients with HIV infection and anemia are generally lower than expected given the degree of anemia. Treatment with erythropoietin may result in an increase in hemoglobin levels. An exception to this is a subset of patients with zidovudine-associated anemia in whom erythropoietin levels may be quite high.

During the course of HIV infection, neutropenia may be seen in approximately half of patients. In most instances it is mild; however, it can be severe and can put patients at risk of spontaneous bacterial infections. This is most frequently seen in patients with severely advanced HIV disease and in patients receiving any of a number of potentially myelosuppressive therapies. In the setting of neutropenia, diseases that are not commonly seen in HIV-infected patients, such as aspergillosis or mucormycosis, may occur. Both granulocyte colony-stimulating factor (G-CSF) and GM-CSF increase neutrophil counts in patients with HIV infection regardless of the cause of the neutropenia. Earlier concerns about the potential of these agents to also increase levels of HIV were not confirmed in controlled clinical trials.

Thrombocytopenia may be an early consequence of HIV infection. Approximately 3% of patients with untreated HIV infection and CD4+ T cell counts ≥400/μL have platelet counts <150,000/μL. For untreated patients with CD4+ T cell counts <400/μL, this incidence increases to 10%. In patients receiving antiretrovirals, thrombocytopenia is associated with hepatitis C, cirrhosis, and ongoing high-level HIV replication. Thrombocytopenia is rarely a serious clinical problem in patients with HIV infection and generally responds well to successful cART. Clinically, it resembles the thrombocytopenia seen in patients with idiopathic thrombocytopenic purpura (Chap. 115). Immune complexes containing anti-gp120 antibodies and anti-anti-gp120 antibodies have been noted in the circulation and on the surface of platelets in patients with HIV infection. Patients with HIV infection have also been noted to have a platelet-specific antibody directed toward a 25-kDa component of the surface of the platelet. Other data suggest that the thrombocytopenia in patients with HIV infection may be due to a direct effect of HIV on megakaryocytes. Whatever the cause, it is very clear that the most effective medical approach to this problem has been the use of cART. For patients with platelet counts <20,000/μL a more aggressive approach combining IVIg or anti-Rh Ig for an immediate response with cART for a more lasting response is appropriate. Rituximab has been used with some success in otherwise refractory cases. Splenectomy is a rarely needed option and is reserved for patients refractory to medical management. Because of the risk of serious infection with encapsulated organisms, all patients with HIV infection about to undergo splenectomy should be immunized with pneumococcal polysaccharide. It should be noted that, in addition to causing an increase in the platelet count, removal of the spleen will result in an increase in the peripheral blood lymphocyte count, making CD4+ T cell counts unreliable markers of immunocompetence. In this setting, the clinician should rely on the CD4+ T cell percent for making diagnostic decisions with respect to the likelihood of opportunistic infections. A CD4+ T cell percent of 15 is approximately equivalent to a CD4+ T cell count of 200/μL. In patients with early HIV infection, thrombocytopenia has also been reported as a consequence of classic thrombotic thrombocytopenic purpura (Chap. 115). This clinical syndrome, consisting of fever, thrombocytopenia, hemolytic anemia, and neurologic and renal dysfunction, is a rare complication of early HIV infection. As in other settings, the appropriate management is the use of salicylates and plasma exchange. Other causes of thrombocytopenia include lymphoma, mycobacterial infections, and fungal infections.

The incidence of venous thromboembolic disease such as deep-vein thrombosis or pulmonary embolus is approximately 1% per year in patients with HIV infection. This is approximately 10 times higher than that seen in an age-matched population. Among the factors associated with clinical thrombosis are age over 45, history of an opportunistic infection, lower CD4 count, and estrogen use. Abnormalities of the coagulation cascade including decreased protein S activity, increases in factor VIII, anticardiolipin antibodies, or lupus-like anticoagulant have been reported in more than 50% of patients with HIV infection. The clinical significance of this increased propensity toward thromboembolic disease is likely reflected in the observation that elevations in d-dimer are strongly associated with all-cause mortality in patients with HIV infection (Table 189-8).

DERMATOLOGIC DISEASES

Dermatologic problems occur in >90% of patients with HIV infection. From the macular, roseola-like rash seen with the acute seroconversion syndrome to extensive end-stage KS, cutaneous manifestations of HIV disease can be seen throughout the course of HIV infection. Among the more common nonneoplastic problems are seborrheic dermatitis, folliculitis, and opportunistic infections. Extrapulmonary pneumocystosis may cause a necrotizing vasculitis. Neoplastic conditions are covered below.

Seborrheic dermatitis occurs in 3% of the general population and in up to 50% of patients with HIV infection. Seborrheic dermatitis increases in prevalence and severity as the CD4+ T cell count declines. In HIV-infected patients, seborrheic dermatitis may be aggravated by concomitant infection with *Pityrosporum*, a yeastlike fungus; use of topical antifungal agents has been recommended in cases refractory to standard topical treatment.

Folliculitis is among the most prevalent dermatologic disorders in patients with HIV infection and is seen in ~20% of patients. It is more common in patients with CD4+ T cell counts <200 cells/μL. Pruritic papular eruption is one of the most common pruritic rashes in patients with HIV infection. It appears as multiple papules on the face, trunk, and extensor surfaces and may improve with cART. *Eosinophilic pustular folliculitis* is a rare form of folliculitis that is seen with increased frequency in patients with HIV infection. It presents as multiple, urticarial perifollicular papules that may coalesce into plaquelike lesions. Skin biopsy reveals an eosinophilic infiltrate of the hair follicle, which in certain cases has been associated with the presence of a mite. Patients typically have an elevated serum IgE level and may respond to treatment with topical anthelmintics. Pruritus is a common symptom in patients with HIV infection and can lead to prurigo nodularis. Patients with HIV infection have also been reported to develop a severe form of *Norwegian scabies* with hyperkeratotic psoriasiform lesions.

Both *psoriasis* and *ichthyosis*, although they are not reported to be increased in frequency, may be particularly severe when they occur in patients with HIV infection. Preexisting psoriasis may become guttate in appearance and more refractory to treatment in the setting of HIV infection.

Reactivation herpes zoster (*shingles*) is seen in 10–20% of patients with HIV infection. This reactivation syndrome of varicella-zoster virus indicates a modest decline in immune function and may be the first indication of clinical immunodeficiency. In one series, patients who developed shingles did so an average of 5 years after HIV infection. In a cohort of patients with HIV infection and localized zoster, the subsequent rate of the development of AIDS was 1% per month. In that study, AIDS was more likely to develop if the outbreak of zoster was associated with severe pain, extensive skin involvement, or involvement of cranial or cervical dermatomes. The clinical manifestations of reactivation zoster in HIV-infected patients, although indicative of

immunologic compromise, are not as severe as those seen in other immunodeficient conditions. Thus, while lesions may extend over several dermatomes, involve the spinal cord, and/or be associated with frank cutaneous dissemination, visceral involvement has not been reported. In contrast to patients without a known underlying immunodeficiency state, patients with HIV infection tend to have recurrences of zoster with a relapse rate of ~20%. Valacyclovir, acyclovir or famciclovir is the treatment of choice. Foscarnet may be of value in patients with acyclovir-resistant virus.

Infection with *herpes simplex virus* in HIV-infected individuals is associated with recurrent orolabial, genital, and perianal lesions as part of recurrent reactivation syndromes (Chap. 179). As HIV disease progresses and the CD4+ T cell count declines, these infections become more frequent and severe. Lesions often appear as beefy red, are exquisitely painful, and have a tendency to occur high in the gluteal cleft (Fig. 189-38). Perirectal HSV may be associated with proctitis and anal fissures. HSV should be high in the differential diagnosis of any HIV-infected patient with a poorly healing, painful perirectal lesion. In addition to recurrent mucosal ulcers, recurrent HSV infection in the form of *herpetic whitlow* can be a problem in patients with HIV infection, presenting with painful vesicles or extensive cutaneous erosion. Valacyclovir, acyclovir or famciclovir is the treatment of choice in these settings. Of note is the fact that even subclinical reactivation of herpes simplex may be associated with increases in plasma HIV RNA levels.

Diffuse skin eruptions due to *Molluscum contagiosum* may be seen in patients with advanced HIV infection. These flesh-colored, umbilicated lesions may be treated with local therapy. They tend to regress with effective cART. Similarly, *condyloma acuminatum* lesions may be more severe and more widely distributed in patients with low CD4+ T cell counts. Imiquimod cream may be helpful in some cases. Atypical mycobacterial infections may present as erythematous cutaneous nodules, as may fungal infections, *Bartonella*, *Acanthamoeba*, and KS.

The skin of patients with HIV infection is often a target organ for drug reactions (Chap. 55). Although most skin reactions are mild and not necessarily an indication to discontinue therapy, patients may have particularly severe cutaneous reactions, including erythroderma, *Stevens-Johnson syndrome*, and toxic epidermal necrolysis, as a reaction to drugs—particularly sulfa drugs, the nonnucleoside reverse transcriptase inhibitors, abacavir, amprenavir, darunavir, fosamprenavir, and tipranavir. Similarly, patients with HIV infection are often quite photosensitive and burn easily following exposure to sunlight or as a side effect of radiation therapy (Chap. 56).

HIV infection and its treatment may be accompanied by cosmetic changes of the skin that are not of great clinical importance but may be troubling to patients. Yellowing of the nails and straightening of the hair, particularly in African-American patients, have been reported as a consequence of HIV infection. Zidovudine therapy has been associated with elongation of the eyelashes and the development of a bluish discoloration to the nails, again more common in African-American patients. Therapy with clofazimine may cause a yellow-orange discoloration of the skin and urine.

NEUROLOGIC DISEASES

Clinical disease of the nervous system accounts for a significant degree of morbidity in a high percentage of patients with HIV infection **(Table 189-13)**. The neurologic problems that occur in HIV-infected individuals may be either primary to the pathogenic processes of HIV infection or secondary to opportunistic infections or neoplasms (see above). Among the more frequent opportunistic diseases that involve the CNS are toxoplasmosis, cryptococcosis, progressive multifocal leukoencephalopathy, and primary CNS lymphoma. Other less common problems include mycobacterial infections; syphilis; and infection with CMV, HTLV-I, *Trypanosoma cruzi*, or *Acanthamoeba*. Overall, secondary diseases of the CNS have been reported to occur in approximately one-third of patients with AIDS. These data antedate the widespread use of cART, and this frequency is considerably lower in patients receiving effective antiretroviral drugs. Primary processes related to HIV infection of the nervous system are reminiscent of those seen with other lentiviruses, such as the Visna-Maedi virus of sheep.

Table 189-13 Neurologic Diseases in Patients with HIV Infection

Opportunistic infections	Myelopathy
Toxoplasmosis	Vacuolar myelopathy
Cryptococcosis	Pure sensory ataxia
Progressive multifocal leukoencephalopathy	Paresthesia/dysesthesia
Cytomegalovirus	Peripheral neuropathy
Syphilis	Acute inflammatory demyelinating polyneuropathy (Guillain-Barré syndrome)
Mycobacterium tuberculosis	Chronic inflammatory demyelinating polyneuropathy (CIDP)
HTLV-I infection	Mononeuritis multiplex
Amebiasis	Distal symmetric polyneuropathy
Neoplasms	Myopathy
Primary CNS lymphoma	
Kaposi's sarcoma	
Result of HIV-1 infection	
Aseptic meningitis	
HIV-associated neurocognitive disorders, including HIV encephalopathy/AIDS dementia complex	

Neurologic problems directly attributable to HIV occur throughout the course of infection and may be inflammatory, demyelinating, or degenerative in nature. The term *HIV-associated neurocognitive disorders* (HAND) is used to describe a spectrum of disorders that range from asymptomatic neurocognitive impairment (ANI) to minor neurocognitive disorder (MND) to clinically severe dementia. The most severe form, *HIV-associated dementia* (HAD), also referred to as the *AIDS dementia complex*, or *HIV encephalopathy*, is considered an AIDS-defining illness. Most HIV-infected patients have some neurologic problem during the course of their disease. Even in the setting of suppressive cART, approximately 50% of HIV-infected individuals can be shown to have mild to moderate neurocognitive impairment using sensitive neuropsychiatric testing. As noted in the section on pathogenesis, damage to the CNS may be a direct result of viral infection of the CNS macrophages or glial cells or may be secondary to the release of neurotoxins and potentially toxic cytokines such as IL-1β, TNF-α, IL-6, and TGF-β. It has been reported that HIV-infected individuals with the E4 allele for apo E are at increased risk for AIDS encephalopathy and peripheral neuropathy. Virtually all patients with HIV infection have some degree of nervous system involvement with the virus. This is evidenced by the fact that CSF findings are abnormal in ~90% of patients, even during the asymptomatic phase of HIV infection. CSF abnormalities include pleocytosis (50–65% of patients), detection of viral RNA (~75%), elevated CSF protein (35%), and evidence of intrathecal synthesis of anti-HIV antibodies (90%). It is important to point out that evidence of infection of

the CNS with HIV does not imply impairment of cognitive function. The neurologic function of an HIV-infected individual should be considered normal unless clinical signs and symptoms suggest otherwise.

Aseptic meningitis may be seen in any but the very late stages of HIV infection. In the setting of acute primary infection, patients may experience a syndrome of headache, photophobia, and meningismus. Rarely, an acute encephalopathy due to encephalitis may occur. Cranial nerve involvement may be seen, predominantly cranial nerve VII but occasionally V and/or VIII. CSF findings include a lymphocytic pleocytosis, elevated protein level, and normal glucose level. This syndrome, which cannot be clinically differentiated from other viral meningitides (Chap. 382), usually resolves spontaneously within 2–4 weeks; however, in some patients, signs and symptoms may become chronic. Aseptic meningitis may occur any time in the course of HIV infection; however, it is rare following the development of AIDS. This fact suggests that clinical aseptic meningitis in the context of HIV infection is an immune-mediated disease.

C. neoformans is the leading infectious cause of meningitis in patients with AIDS (Chap. 202). It is the initial AIDS-defining illness in ~2% of patients and generally occurs in patients with CD4+ T cell counts <100/μL. Cryptococcal meningitis is particularly common in untreated patients with AIDS in Africa, occurring in ~5% of patients. Most patients present with a picture of subacute meningoencephalitis with fever, nausea, vomiting, altered mental status, headache, and meningeal signs. The incidence of seizures and focal neurologic deficits is low. The CSF profile may be normal or may show only modest elevations in WBC or protein levels and decreases in glucose. The opening pressure in the CSF is usually elevated. In addition to meningitis, patients may develop cryptococcomas and cranial nerve involvement. Approximately one-third of patients also have pulmonary disease. Uncommon manifestations of cryptococcal infection include skin lesions that resemble *molluscum contagiosum*, lymphadenopathy, palatal and glossal ulcers, arthritis, gastroenteritis, myocarditis, and prostatitis. The prostate gland may serve as a reservoir for smoldering cryptococcal infection. The diagnosis of cryptococcal meningitis is made by identification of organisms in spinal fluid with india ink examination or by the detection of cryptococcal antigen. Blood cultures for fungus are often positive. A biopsy may be needed to make a diagnosis of CNS cryptococcoma. Treatment is with IV amphotericin B 0.7 mg/kg daily, or liposomal amphotericin 4–6 mg/kg daily, with flucytosine 25 mg/kg qid for at least 2 weeks and, if possible, until the CSF culture turns negative. This is followed by fluconazole 400 mg/d PO for 8 weeks, and then fluconazole 200 mg/d until the CD4+ T cell count has increased to >200 cells/μL for 6 months in response to cART. Repeated lumbar puncture may be required to manage increased intracranial pressure. Symptoms may recur with initiation of cART as an immune reconstitution syndrome (see above). Other fungi that may cause meningitis in patients with HIV infection are *C. immitis* and *H. capsulatum*. Meningoencephalitis has also been reported due to *Acanthamoeba* or *Naegleria*.

HIV-associated dementia consists of a constellation of signs and symptoms of CNS disease. While this is generally a late complication of HIV infection that progresses slowly over months, it can be seen in patients with CD4+ T cell counts >350 cells/μL. A major feature of this entity is the development of dementia, defined as a decline in cognitive ability from a previous level. It may present as impaired ability to concentrate, increased forgetfulness, difficulty reading, or increased difficulty performing complex tasks. Initially these symptoms may be indistinguishable from findings of situational depression or fatigue. In contrast to "cortical" dementia (such as Alzheimer's disease), aphasia, apraxia, and agnosia are uncommon, leading some investigators to classify HIV encephalopathy as a "subcortical dementia" characterized by defects in short-term memory and executive function (see below). In addition to dementia, patients with HIV encephalopathy may also have motor and behavioral abnormalities. Among the motor problems are unsteady gait, poor balance, tremor, and difficulty with rapid alternating movements. Increased tone and deep tendon reflexes may be found in patients with spinal cord involvement. Late stages may be complicated by bowel and/or bladder incontinence. Behavioral problems include apathy, irritability, and lack of initiative, with progression to a vegetative state in some instances. Some patients develop a state of agitation or mild mania. These changes usually occur without significant changes in level of alertness. This is in contrast to the finding of somnolence in patients with dementia due to toxic/metabolic encephalopathies.

HIV-associated dementia is the initial AIDS-defining illness in ~3% of patients with HIV infection and thus only rarely precedes clinical evidence of immunodeficiency. Clinically significant encephalopathy eventually develops in ~25% of untreated patients with AIDS. As immunologic function declines, the risk and severity of HIV-associated dementia increases. Autopsy series suggest that 80–90% of patients with HIV infection have histologic evidence of CNS involvement. Several classification schemes have been developed for grading HIV encephalopathy; a commonly used clinical staging system is outlined in **Table 189-14**.

Table 189-14 Clinical Staging of HIV Encephalopathy (AIDS Dementia Complex)

Stage	Definition
0 (Normal)	Normal mental and motor function.
0.5 (Equivocal/subclinical)	Absent, minimal, or equivocal symptoms without impairment of work or capacity to perform activities of daily living. Mild signs (snout response, slowed ocular or extremity movements) may be present. Gait and strength are normal.
1 (Mild)	Able to perform all but the more demanding aspects of work or activities of daily living but with unequivocal evidence (signs or symptoms that may include performance on neuropsychological testing) of functional, intellectual, or motor impairment. Can walk without assistance.
2 (Moderate)	Able to perform basic activities of self-care but cannot work or maintain the more demanding aspects of daily life. Ambulatory, but may require a single prop.
3 (Severe)	Major intellectual incapacity (cannot follow news or personal events, cannot sustain complex conversation, considerable slowing of all output) or motor disability (cannot walk unassisted, usually with slowing and clumsiness of arms as well).
4 (End-stage)	Nearly vegetative. Intellectual and social comprehension and output are at a rudimentary level. Nearly or absolutely mute. Paraparetic or paraplegic with urinary and fecal incontinence.

Source: Adapted from JJ Sidtis, RW Price: Neurology 40:197, 1990.

The precise cause of HIV-associated dementia remains unclear, although the condition is thought to be a result of a combination of direct effects of HIV on the CNS and associated immune activation. HIV has been found in the brains of patients with HIV encephalopathy by Southern blot, in situ hybridization, PCR, and electron microscopy. Multinucleated giant cells, macrophages, and microglial cells appear to be the main cell types harboring virus in the CNS. Histologically, the major changes are seen in the subcortical areas of the brain and include pallor and gliosis, multinucleated giant cell encephalitis, and vacuolar myelopathy. Less commonly, diffuse or focal spongiform changes occur in the white matter. Areas of the brain involved in motor, language, and judgment are most severely affected.

There are no specific criteria for a diagnosis of HIV-associated dementia, and this syndrome must be differentiated from a number of other diseases that affect the CNS of HIV-infected patients (Table 189-13). The diagnosis of dementia depends on demonstrating a decline in cognitive function. This can be accomplished objectively with the use of a Mini-Mental Status Examination (MMSE) in patients for whom prior scores are available. For this reason, it is advisable for all patients with a diagnosis of HIV infection to have a baseline MMSE. However, changes in MMSE scores may be absent in patients with mild

HIV encephalopathy. Imaging studies of the CNS, by either MRI or CT, often demonstrate evidence of cerebral atrophy **(Fig. 189-40)**. MRI may also reveal small areas of increased density on T2-weighted images. Lumbar puncture is an important element of the evaluation of patients with HIV infection and neurologic abnormalities. It is generally most helpful in ruling out or making a diagnosis of opportunistic infections. In HIV encephalopathy, patients may have the nonspecific findings of an increase in CSF cells and protein level. While HIV RNA can often be detected in the spinal fluid and HIV can be cultured from the CSF, this finding is not specific for HIV encephalopathy. There appears to be no correlation between the presence of HIV in the CSF and the presence of HIV encephalopathy. Elevated levels of macrophage chemoattractant protein (MCP-1), β_2-microglobulin, neopterin, and quinolinic acid (a metabolite of tryptophan reported to cause CNS injury) have been noted in the CSF of patients with HIV encephalopathy. These findings suggest that these factors as well as inflammatory cytokines may be involved in the pathogenesis of this syndrome.

Figure 189-40

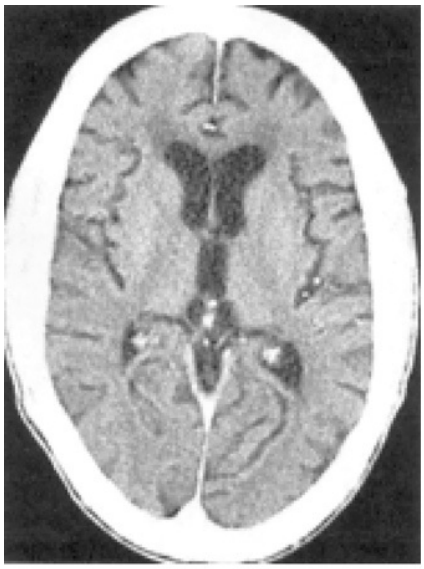

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

AIDS dementia complex. Postcontrast CT scan through the lateral ventricles of a 47-year-old man with AIDS, altered mental status, and dementia. The lateral and third ventricles and the cerebral sulci are abnormally prominent. Mild white matter hypodensity is also seen adjacent to the frontal horns of the lateral ventricles.

Combination antiretroviral therapy is of benefit in patients with HIV-associated dementia. Improvement in neuropsychiatric test scores has been noted for both adult and pediatric patients treated with antiretrovirals. The rapid improvement in cognitive function noted with the initiation of cART suggests that at least some component of this problem is quickly reversible, again supporting at least a partial role of soluble mediators in the pathogenesis. It should also be noted that these patients have an increased sensitivity to the side effects of neuroleptic drugs. The use of these drugs for symptomatic treatment is associated with an increased risk of extrapyramidal side effects; therefore, patients with HIV encephalopathy who receive these agents must be monitored carefully. It is felt by many physicians that the decrease in the prevalence of severe cases of HAND brought about by cART has resulted in an increase in the prevalence of milder forms of this disorder.

Seizures may be a consequence of opportunistic infections, neoplasms, or HIV encephalopathy **(Table 189-15)**. The seizure threshold is often lower than normal in patients with advanced HIV infection due to the frequent presence of electrolyte abnormalities. Seizures are seen in 15–40% of patients with cerebral toxoplasmosis, 15–35% of patients with primary CNS lymphoma, 8% of patients with cryptococcal meningitis, and 7–50% of patients with HIV encephalopathy. Seizures may also be seen in patients with CNS tuberculosis, aseptic meningitis, and progressive multifocal leukoencephalopathy. Seizures may be the presenting clinical symptom of HIV disease. In one study of 100 patients with HIV infection presenting with a first seizure, cerebral mass lesions were the most common cause, responsible for 32 of the 100 new-onset seizures. Of these 32 cases, 28 were due to toxoplasmosis and 4 to lymphoma. HIV encephalopathy accounted for an additional 24 new-onset seizures. Cryptococcal meningitis was the third most common diagnosis, responsible for 13 of the 100 seizures. In 23 cases, no cause could be found, and it is possible that these cases represent a subcategory of HIV encephalopathy. Of these 23 cases, 16 (70%) had two or more seizures, suggesting that anticonvulsant therapy is indicated in all patients with HIV infection and seizures unless a rapidly correctable cause is found. While phenytoin remains the initial treatment of choice, hypersensitivity reactions to this drug have been reported in >10% of patients with AIDS, and therefore the use of phenobarbital or valproic acid must be considered as alternatives. Due to a variety of drug-drug interactions between antiseizure medications and antiretrovirals, drug levels need to be monitored carefully.

Table 189-15 Causes of Seizures in Patients with HIV Infection

Disease	Overall Contribution to First Seizure, %	Fraction of Patients Who Have Seizures, %
HIV encephalopathy	24–47	7–50
Cerebral toxoplasmosis	28	15–40
Cryptococcal meningitis	13	8
Primary central nervous system lymphoma	4	15–30
Progressive multifocal leukoencephalopathy	1	

Source: From DM Holtzman et al: Am J Med 87:173, 1989.

Patients with HIV infection may present with *focal neurologic deficits* from a variety of causes. The most common causes are toxoplasmosis, progressive multifocal leukoencephalopathy, and CNS lymphoma. Other causes include cryptococcal infections (discussed above; also Chap. 202), stroke, and reactivation of Chagas' disease.

Toxoplasmosis has been one of the most common causes of secondary CNS infections in patients with AIDS, but its incidence is decreasing in the era of cART. It is most common in patients from the Caribbean and from France, where the seroprevalence of *T. gondii* is around 50%. This figure is closer to 15% in the United States. Toxoplasmosis is generally a late complication of HIV infection and usually occurs in patients with CD4+ T cell counts <200/μL. Cerebral toxoplasmosis is thought to represent a reactivation of latent tissue cysts. It is 10 times more common in patients with antibodies to the organism than in patients who are seronegative. Patients diagnosed with HIV infection should be screened for IgG antibodies to *T. gondii* during the time of their initial workup. Those who are seronegative should be counseled about ways to minimize the risk of primary infection including avoiding the consumption of undercooked meat and careful hand washing after contact with soil or changing the cat litter box. The most common clinical presentation of cerebral toxoplasmosis in patients with HIV infection is fever, headache, and focal neurologic deficits. Patients may present with seizure, hemiparesis, or aphasia as a manifestation of these focal deficits or with a picture more influenced by the accompanying cerebral edema and characterized by confusion, dementia, and lethargy, which can progress to coma. The diagnosis is usually suspected on the basis of MRI findings of multiple lesions in multiple locations, although in some cases only a single lesion is seen. Pathologically, these lesions generally exhibit inflammation and central necrosis and, as a result, demonstrate ring enhancement on contrast MRI **(Fig. 189-41)** or, if MRI is unavailable or contraindicated, on double-dose contrast CT. There is usually evidence of surrounding edema. In addition to toxoplasmosis, the differential diagnosis of single or multiple enhancing mass lesions in the HIV-infected patient includes primary CNS lymphoma and, less commonly, TB or fungal or bacterial abscesses. The definitive diagnostic procedure is brain biopsy. However, given the morbidity rate that can accompany this procedure, it is usually reserved for the patient who has failed 2–4 weeks of empiric therapy. If the patient is seronegative for *T. gondii*, the likelihood that a mass lesion is due to toxoplasmosis is <10%. In that setting, one may choose to be more aggressive and perform a brain biopsy sooner. Standard treatment is sulfadiazine and pyrimethamine with leucovorin as needed for a minimum of 4–6 weeks. Alternative therapeutic regimens include clindamycin in combination with pyrimethamine; atovaquone plus pyrimethamine; and azithromycin plus pyrimethamine plus rifabutin. Relapses are common, and it is recommended that patients with a history of prior toxoplasmic encephalitis receive maintenance therapy with sulfadiazine, pyrimethamine, and leucovorin as long as their CD4+ T cell counts remain <200 cells/μL. Patients with CD4+ T cell counts <100/μL and IgG antibody to *Toxoplasma* should receive primary prophylaxis for toxoplasmosis. Fortunately, the same daily regimen of a single double-strength tablet of TMP/SMX used for *P. jiroveci* prophylaxis provides adequate primary protection against toxoplasmosis. Secondary prophylaxis/maintenance therapy for toxoplasmosis may be discontinued in the setting of effective cART and increases in CD4+ T cell counts to >200/μL for 6 months.

Figure 189-41

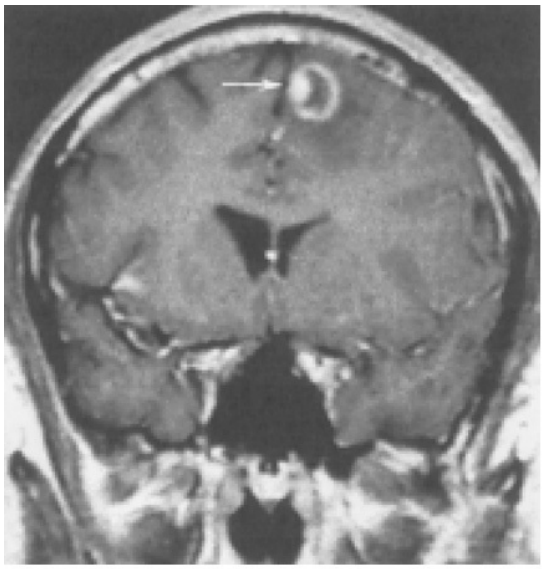

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Central nervous system toxoplasmosis. A coronal postcontrast T1-weighted MRI scan demonstrates a peripheral enhancing lesion in the left frontal lobe, associated with an eccentric nodular area of enhancement (*arrow*); this so-called eccentric target sign is typical of toxoplasmosis.

JC virus, a human polyomavirus that is the etiologic agent of *progressive multifocal leukoencephalopathy* (PML), is an important opportunistic pathogen in patients with AIDS (Chap. 381). While ~80% of the general adult population has antibodies to JC virus, indicative of prior infection, <10% of healthy adults show any evidence of ongoing viral replication. PML is the only known clinical manifestation of JC virus infection. It is a late manifestation of AIDS and is seen in ~4% of patients with AIDS. The lesions of PML begin as small foci of demyelination in subcortical white matter that eventually coalesce. The cerebral hemispheres, cerebellum, and brainstem may all be involved. Patients typically have a protracted course with multifocal neurologic deficits, with or without changes in mental status. Approximately 20% of patients experience seizures. Ataxia, hemiparesis, visual field defects, aphasia, and sensory defects may occur. Headache, fever, nausea, and vomiting are rarely seen. Their presence should suggest another diagnosis. MRI typically reveals multiple, nonenhancing white matter lesions that may coalesce and have a predilection for the occipital and parietal lobes. The lesions show signal hyperintensity on T2-weighted images and diminished signal on T1-weighted images. The measurement of JC virus DNA levels in CSF has a diagnostic sensitivity of 76% and a specificity of close to 100%. Prior to the availability of cART, the majority of patients with PML died within 3–6 months of the onset of symptoms. Paradoxical worsening of PML has been seen with initiation of cART as an immune reconstitution syndrome. There is no specific treatment for PML; however, a median survival of 2 years and survival of >15 years have been reported in patients with PML treated with cART for their HIV disease. Despite having a significant impact on survival, only ~50% of patients with HIV infection and PML show neurologic improvement with cART. Studies with

other antiviral agents such as cidofovir have failed to show clear benefit. Factors influencing a favorable prognosis for PML in the setting of HIV infection include a CD4+ T cell count >100/μL at baseline and the ability to maintain an HIV viral load of <500 copies per milliliter. Baseline HIV-1 viral load does not have independent predictive value of survival. PML is one of the few opportunistic infections that continues to occur with some frequency despite the widespread use of cART.

Reactivation American trypanosomiasis may present as acute meningoencephalitis with focal neurologic signs, fever, headache, vomiting, and seizures. Accompanying cardiac disease in the form of arrhythmias or heart failure should increase the index of suspicion. The presence of antibodies to *T. cruzi* supports the diagnosis. In South America, reactivation of *Chagas' disease* is considered to be an AIDS-defining condition and may be the initial AIDS-defining condition. The majority of cases occur in patients with CD4+ T cell counts <200 cells/μL. Lesions appear radiographically as single or multiple hypodense areas, typically with ring enhancement and edema. They are found predominantly in the subcortical areas, a feature that differentiates them from the deeper lesions of toxoplasmosis. *T. cruzi* amastigotes, or trypanosomes, can be identified from biopsy specimens or CSF. Other CSF findings include elevated protein and a mild (<100 cells/μL) lymphocytic pleocytosis. Organisms can also be identified by direct examination of the blood. Treatment consists of benzimidazole (2.5 mg/kg bid) or nifurtimox (2 mg/kg qid) for at least 60 days, followed by maintenance therapy for the duration of immunodeficiency with either drug at a dose of 5 mg/kg three times a week. As is the case with cerebral toxoplasmosis, successful therapy with antiretrovirals may allow discontinuation of therapy for Chagas' disease.

Stroke may occur in patients with HIV infection. In contrast to the other causes of focal neurologic deficits in patients with HIV infection, the symptoms of a stroke are sudden in onset. Patients with HIV infection have an increased prevalence of many classic risk factors associated with stroke, including smoking and diabetes. It also appears that HIV infection itself can lead to an increase in carotid artery stiffness. Among the secondary infectious diseases in patients with HIV infection that may be associated with stroke are vasculitis due to cerebral varicella zoster or neurosyphilis and septic embolism in association with fungal infection. Other elements of the differential diagnosis of stroke in the patient with HIV infection include atherosclerotic cerebral vascular disease, thrombotic thrombocytopenic purpura, and cocaine or amphetamine use.

Primary CNS lymphoma is discussed below in the section on neoplastic diseases.

Spinal cord disease, or myelopathy, is present in ~20% of patients with AIDS, often as part of HIV-associated neurocognitive disorder. In fact, 90% of the patients with HIV-associated myelopathy have some evidence of dementia, suggesting that similar pathologic processes may be responsible for both conditions. Three main types of spinal cord disease are seen in patients with AIDS. The first of these is a vacuolar myelopathy, as mentioned above. This condition is pathologically similar to subacute combined degeneration of the cord, such as that occurring with pernicious anemia. Although vitamin B12 deficiency can be seen in patients with AIDS as a primary complication of HIV infection, it does not appear to be responsible for the myelopathy seen in the majority of patients. Vacuolar myelopathy is characterized by a subacute onset and often presents with gait disturbances, predominantly ataxia and spasticity; it may progress to include bladder and bowel dysfunction. Physical findings include evidence of increased deep tendon reflexes and extensor plantar responses. The second form of spinal cord disease involves the dorsal columns and presents as a pure sensory ataxia. The third form is also sensory in nature and presents with paresthesias and dysesthesias of the lower extremities. In contrast to the cognitive problems seen in patients with HIV encephalopathy, these spinal cord syndromes do not respond well to antiretroviral drugs, and therapy is mainly supportive.

One important disease of the spinal cord that also involves the peripheral nerves is a *myelopathy* and *polyradiculopathy* seen in association with CMV infection. This entity is generally seen late in the course of HIV infection and is fulminant in onset, with lower extremity and sacral paresthesias, difficulty in walking, areflexia, ascending sensory loss, and urinary retention. The clinical course is rapidly progressive over a period of weeks. CSF examination reveals a predominantly neutrophilic pleocytosis, and CMV DNA can be detected by CSF PCR. Therapy with ganciclovir or foscarnet can lead to rapid improvement, and prompt initiation of foscarnet or ganciclovir therapy is important in minimizing the degree of permanent neurologic damage. Combination therapy with both drugs should be considered in patients who have been previously treated for CMV disease. Other diseases involving the spinal cord in patients with HIV infection include HTLV-I-associated myelopathy (HAM) (Chap. 188), neurosyphilis (Chap. 169), infection with herpes simplex (Chap. 179) or varicella-zoster (Chap. 180), TB (Chap. 165), and lymphoma (Chap. 110).

Peripheral neuropathies are common in patients with HIV infection. They occur at all stages of illness and take a variety of forms. Early in the course of HIV infection, an acute inflammatory demyelinating polyneuropathy resembling Guillain-Barré syndrome may occur (Chap. 385). In other patients, a progressive or relapsing-remitting inflammatory neuropathy resembling chronic inflammatory demyelinating polyneuropathy (CIDP) has been noted. Patients commonly present with progressive weakness, areflexia, and minimal sensory changes. CSF examination often reveals a mononuclear pleocytosis, and peripheral nerve biopsy demonstrates a perivascular infiltrate suggesting an autoimmune etiology. Plasma exchange or IVIg has been tried with variable success. Because of the immunosuppressive effects of glucocorticoids, they should be reserved for severe cases of CIDP refractory to other measures. Another autoimmune peripheral neuropathy seen in patients with AIDS is mononeuritis multiplex (Chaps. 385 and 326) due to a necrotizing arteritis of peripheral nerves. The most common peripheral neuropathy in patients with HIV infection is a *distal sensory polyneuropathy* (DSPN) also referred to as painful sensory neuropathy (HIV-SN), predominantly sensory neuropathy, or distal symmetric peripheral neuropathy. This condition may be a direct consequence of HIV infection or a side effect of dideoxynucleoside therapy. It is more common in taller individuals, older individuals, and those with lower CD4 counts. Two-thirds of patients with AIDS may be shown by electrophysiologic studies to have some evidence of peripheral nerve disease. Presenting symptoms are usually painful burning sensations in the feet and lower extremities. Findings on examination include a stocking-type sensory loss to pinprick, temperature, and touch sensation and a loss of ankle reflexes. Motor changes are mild and are usually limited to weakness of the intrinsic foot muscles. Response of this condition to antiretrovirals has been variable, perhaps because antiretrovirals are responsible for the problem in some instances. When due to dideoxynucleoside therapy, patients with lower extremity peripheral neuropathy may complain of a sensation that they are walking on ice. Other entities in the differential diagnosis of peripheral neuropathy include diabetes mellitus, vitamin B12 deficiency, and side effects from metronidazole or dapsone. For distal symmetric polyneuropathy that fails to resolve following the discontinuation of dideoxynucleosides, therapy is symptomatic; gabapentin, carbamazepine, tricyclics, or analgesics may be effective for dysesthesias. Treatment-naïve patients may respond to cART.

Myopathy may complicate the course of HIV infection; causes include HIV infection itself, zidovudine, and the generalized wasting syndrome. HIV-associated myopathy may range in severity from an asymptomatic elevation in creatine kinase levels to a subacute syndrome characterized by proximal muscle weakness and myalgias. Quite pronounced elevations in creatine kinase may occur in asymptomatic patients, particularly after exercise. The clinical significance of this as an isolated laboratory finding is unclear. A variety of both inflammatory and noninflammatory pathologic processes have been noted in patients with more severe myopathy, including myofiber necrosis with inflammatory cells, nemaline rod bodies, cytoplasmic bodies, and mitochondrial abnormalities. Profound muscle wasting, often with muscle pain, may be seen after prolonged zidovudine therapy. This toxic side effect of the drug is dose-dependent and is related to its ability to interfere with the function of mitochondrial polymerases. It is reversible following discontinuation of the drug. Red ragged fibers are a histologic hallmark of zidovudine-induced myopathy.

OPHTHALMOLOGIC DISEASES

Ophthalmologic problems occur in ~50% of patients with advanced HIV infection. The most common abnormal findings on funduscopic examination are cotton-wool spots. These are hard white spots that appear on the surface of the retina and often have an irregular edge. They represent areas of retinal ischemia secondary to microvascular disease. At times they are associated with small areas of hemorrhage and thus can be difficult to distinguish from CMV retinitis. In contrast to CMV retinitis, however, these lesions are not associated with visual loss and tend to remain stable or improve over time.

One of the most devastating consequences of HIV infection is CMV retinitis. Patients at high risk of CMV retinitis (CD4+ T cell count <100/μL) should undergo an ophthalmologic examination every 3–6 months. The majority of cases of CMV retinitis occur in patients with a CD4+ T cell count <50/μL. Prior to the availability of cART, this CMV reactivation syndrome was seen in 25–30% of patients with AIDS. CMV retinitis usually presents as a painless, progressive loss of vision. Patients may also complain of blurred vision, "floaters," and scintillations. The disease is usually bilateral, although typically it affects one eye more than the other. The diagnosis is made on clinical grounds by an experienced ophthalmologist. The characteristic retinal appearance is that of perivascular hemorrhage and exudate. In situations where the diagnosis is in doubt due to an atypical presentation or an unexpected lack of response to therapy, vitreous or aqueous humor sampling with molecular diagnostic techniques may be of value. CMV infection of the retina results in a necrotic inflammatory process, and the visual loss that develops is irreversible. CMV retinitis may be complicated by rhegmatogenous retinal detachment as a consequence of retinal atrophy in areas of prior inflammation. Therapy for CMV retinitis consists of oral valganciclovir, IV ganciclovir, or IV foscarnet, with cidofovir as an alternative. Combination therapy with ganciclovir and foscarnet has been shown to be slightly more effective than either ganciclovir or foscarnet alone in the patient with relapsed CMV retinitis. A 3-week induction course is followed by maintenance therapy with oral valganciclovir. If CMV disease is limited to the eye, a ganciclovir-releasing intraocular implant, periodic injections of the antisense nucleic acid preparation fomivirsen (no longer available in the United States), or intravitreal injections of ganciclovir or foscarnet may be considered; some choose to combine intraocular implants with oral valganciclovir. Intravitreal injections of cidofovir are generally avoided due to the increased risk of uveitis and hypotony. Maintenance therapy is continued until the CD4+ T cell count remains >100–150/μL for >6 months. The majority of patients with HIV infection and CMV disease develop some degree of uveitis with the initiation of cART. The etiology of this is unknown; however, it has been suggested that this may be due to the generation of an enhanced immune response to CMV as an IRIS (see above). In some instances this has required the use of topical glucocorticoids.

Both HSV and varicella zoster virus can cause a rapidly progressing, bilateral necrotizing retinitis referred to as the *acute retinal necrosis syndrome*, or *progressive outer retinal necrosis* (PORN). This syndrome, in contrast to CMV retinitis, is associated with pain, keratitis, and iritis. It is often associated with orolabial HSV or trigeminal zoster. Ophthalmologic examination reveals widespread pale gray peripheral lesions. This condition is often complicated by retinal detachment. It is important to recognize and treat this condition with IV acyclovir as quickly as possible to minimize the loss of vision.

Several other secondary infections may cause ocular problems in HIV-infected patients. *P. jiroveci* can cause a lesion of the choroid that may be detected as an incidental finding on ophthalmologic examination. These lesions are typically bilateral, are from half to twice the disc diameter in size, and appear as slightly elevated yellow-white plaques. They are usually asymptomatic and may be confused with cotton-wool spots. Chorioretinitis due to toxoplasmosis can be seen alone or, more commonly, in association with CNS toxoplasmosis. KS may involve the eyelid or conjunctiva, while lymphoma may involve the retina. Syphilis may lead to a uveitis that is highly associated with the presence of neurosyphilis.

ADDITIONAL DISSEMINATED INFECTIONS AND WASTING SYNDROME

Infections with species of the small, gram-negative, *Rickettsia*-like organism *Bartonella* (Chap. 160) are seen with increased frequency in patients with HIV infection. While it is not considered an AIDS-defining illness by the CDC, many experts view infection with *Bartonella* as indicative of a severe defect in cell-mediated immunity. It is usually seen in patients with CD4+ T cell counts <100/μL and is a significant cause of unexplained fever in patients with advanced HIV infection. Among the clinical manifestations of *Bartonella* infection are bacillary angiomatosis, cat-scratch disease, and trench fever. *Bacillary angiomatosis* is usually due to infection with *B. henselae* and is linked to exposure to flea-infested cats. It is characterized by a vascular proliferation that leads to a variety of skin lesions that have been confused with the skin lesions of KS. In contrast to the lesions of KS, the lesions of bacillary angiomatosis generally blanch, are painful, and typically occur in the setting of systemic symptoms. Infection can extend to the lymph nodes, liver (peliosis hepatis), spleen, bone, heart, CNS, respiratory tract, and GI tract. *Cat-scratch disease* is also due to *B. henselae* and generally begins with a papule at the site of inoculation. This is followed several weeks later by the development of regional adenopathy and malaise. Infection with *B. quintana* is transmitted by lice and has been associated with case reports of trench fever, endocarditis, adenopathy, and bacillary angiomatosis. The organism is quite difficult to culture, and diagnosis often relies on identifying the organism in biopsy specimens using the Warthin-Starry or similar stains. Treatment is with either doxycycline or erythromycin for at least 3 months.

Histoplasmosis is an opportunistic infection that is seen most frequently in patients in the Mississippi and Ohio River valleys, Puerto Rico, the Dominican Republic, and South America. These are all areas in which infection with *H. capsulatum* is endemic (Chap. 199). Because of this limited geographic distribution, the percentage of AIDS cases in the United States with histoplasmosis is only ~0.5. Histoplasmosis is generally a late manifestation of HIV infection; however, it may be the initial AIDS-defining condition. In one study, the median CD4+ T cell count for patients with histoplasmosis and AIDS was 33/μL. While disease due to *H. capsulatum* may present as a primary infection of the lung, disseminated disease, presumably due to reactivation, is the most common presentation in HIV-infected patients. Patients usually present with a 4- to 8-week history of fever and weight loss. Hepatosplenomegaly and lymphadenopathy are each seen in about 25% of patients. CNS disease, either meningitis or a mass lesion, is seen in 15% of patients. Bone marrow involvement is common, with thrombocytopenia, neutropenia, and anemia occurring in 33% of patients. Approximately 7% of patients have mucocutaneous lesions consisting of a maculopapular rash and skin or oral ulcers. Respiratory symptoms are usually mild, with chest x-ray showing a diffuse infiltrate or diffuse small nodules in ~50% of cases. Diagnosis is made by culturing the organisms from blood, bone marrow, or tissue or by detecting antigen in blood or urine. Treatment is typically with liposomal amphotericin B followed by maintenance therapy with oral itraconazole until the serum histoplasma antigen is <2 units, the patient has been on antiretrovirals for at least 6 months, and the CD4 count is >150 cells/μL. In the setting of mild infection, it may be appropriate to initiate therapy with itraconazole alone.

Following the spread of HIV infection to southeast Asia, disseminated infection with the fungus *Penicillium marneffei* was recognized as a complication of HIV infection and is considered an AIDS-defining condition in those parts of the world where it occurs. *P. marneffei* is the third most common AIDS-defining illness in Thailand, following TB and cryptococcosis. It is more frequently diagnosed in the rainy than the dry season. Clinical features include fever, generalized lymphadenopathy, hepatosplenomegaly, anemia, thrombocytopenia, and papular skin lesions with central umbilication. Treatment is with amphotericin B followed by itraconazole until the CD4+ T cell count is >100 cells/μL for at least 6 months.

Visceral leishmaniasis (Chap. 212) is recognized with increasing frequency in patients with HIV infection who live in or travel to areas endemic for this protozoal infection transmitted by sandflies. The clinical presentation is one of hepatosplenomegaly, fever, and hematologic abnormalities. Lymphadenopathy and other constitutional symptoms may be present. A chronic, relapsing course is seen in two-thirds of co-infected patients. Organisms can be isolated from cultures of bone marrow aspirates. Histologic stains may be negative, and antibody titers are of little help. Patients with HIV infection usually respond well initially to

standard therapy with amphotericin B or pentavalent antimony compounds. Eradication of the organism is difficult, however, and relapses are common.

Patients with HIV infection are at a slightly increased risk of clinical malaria. This is particularly true for patients from nonendemic areas with presumed primary infection and in patients with lower CD4+ T cell counts. HIV-positive individuals with CD4+ T cell counts <300 cells/μL have a poorer response to malaria treatment than others. Co-infection with malaria is associated with a modest increase in HIV viral load. The risk of malaria may be decreased with TMP/SMX prophylaxis.

Generalized wasting is an AIDS-defining condition; it is defined as involuntary weight loss of >10% associated with intermittent or constant fever and chronic diarrhea or fatigue lasting >30 days in the absence of a defined cause other than HIV infection. Prior to the widespread use of cART it was the initial AIDS-defining condition in ~10% of patients with AIDS in the United States and is an indication for initiation of cART. Generalized wasting is rarely seen today with the earlier initiation of antiretrovirals. A constant feature of this syndrome is severe muscle wasting with scattered myofiber degeneration and occasional evidence of myositis. Glucocorticoids may be of some benefit; however, this approach must be carefully weighed against the risk of compounding the immunodeficiency of HIV infection. Androgenic steroids, growth hormone, and total parenteral nutrition have been used as therapeutic interventions with variable success.

NEOPLASTIC DISEASES

The neoplastic diseases considered to be AIDS-defining conditions are Kaposi's sarcoma, non-Hodgkin's lymphoma, and invasive cervical carcinoma. In addition, there is also an increase in the incidence of a variety of non-AIDS-defining malignancies including Hodgkin's disease; multiple myeloma; leukemia; melanoma; and cervical, brain, testicular, oral, lung, gastric, liver, renal, and anal cancers. Since the introduction of potent cART, there has been a marked reduction in the incidence of KS (Fig. 189-34) and CNS lymphoma, such that the non-AIDS-defining malignancies now account for more morbidity and mortality in patients with HIV infection than the AIDS-defining malignancies. Rates of non-Hodgkin's lymphoma have declined as well; however, this decline has not been as dramatic as the decline in rates of KS. In contrast, cART has had little effect on human papillomavirus (HPV)-associated malignancies. As patients with HIV infection live longer, a wider array of cancers is seen in this population. While some may only reflect known risk factors (e.g., smoking, alcohol consumption, co-infection with other viruses such as hepatitis B) that are increased in patients with HIV infection, some may be a direct consequence of HIV and are clearly increased in patients with lower CD4+ T cell counts.

Kaposi's sarcoma is a multicentric neoplasm consisting of multiple vascular nodules appearing in the skin, mucous membranes, and viscera. The course ranges from indolent, with only minor skin or lymph node involvement, to fulminant, with extensive cutaneous and visceral involvement. In the initial period of the AIDS epidemic, KS was a prominent clinical feature of the first cases of AIDS, occurring in 79% of the patients diagnosed in 1981. By 1989 it was seen in only 25% of cases, by 1992 the number had decreased to 9%, and by 1997 the number was <1%. HHV-8 or KSHV has been strongly implicated as a viral cofactor in the pathogenesis of KS.

Clinically, KS has varied presentations and may be seen at any stage of HIV infection, even in the presence of a normal CD4+ T cell count. The initial lesion may be a small, raised reddish-purple nodule on the skin **(Fig. 189-42)**, a discoloration on the oral mucosa (Fig. 189-35*D*), or a swollen lymph node. Lesions often appear in sun-exposed areas, particularly the tip of the nose, and have a propensity to occur in areas of trauma (Koebner phenomenon). Because of the vascular nature of the tumors and the presence of extravasated red blood cells in the lesions, their colors range from reddish to purple to brown and often take the appearance of a bruise, with yellowish discoloration and tattooing. Lesions range in size from a few millimeters to several centimeters in diameter and may be either discrete or confluent. KS lesions most commonly appear as raised macules; however, they can also be papular, particularly in patients with higher CD4+ T cell counts. Confluent lesions may give rise to surrounding lymphedema and may be disfiguring when they involve the face and disabling when they involve the lower extremities or the surfaces of joints. Apart from skin, the lymph nodes, GI tract, and lung are the organ systems most commonly affected by KS. Lesions have been reported in virtually every organ, including the heart and the CNS. In contrast to most malignancies, in which lymph node involvement implies metastatic spread and a poor prognosis, lymph node involvement may be seen very early in KS and is of no special clinical significance. In fact, some patients may present with disease limited to the lymph nodes. These are generally patients with relatively intact immune function and thus the patients with the best prognosis. Pulmonary involvement with KS generally presents with shortness of breath. Some 80% of patients with pulmonary KS also have cutaneous lesions. The chest x-ray characteristically shows bilateral lower lobe infiltrates that obscure the margins of the mediastinum and diaphragm **(Fig. 189-43)**. Pleural effusions are seen in 70% of cases of pulmonary KS, a fact that is often helpful in the differential diagnosis. GI involvement is seen in 50% of patients with KS and usually takes one of two forms: (1) mucosal involvement, which may lead to bleeding that can be severe; these patients sometimes also develop symptoms of GI obstruction if lesions become large; and (2) biliary tract involvement. KS lesions may infiltrate the gallbladder and biliary tree, leading to a clinical picture of obstructive jaundice similar to that seen with sclerosing cholangitis. Several staging systems have been proposed for KS. One in common use was developed by the National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group; it distinguishes patients on the basis of tumor extent, immunologic function, and presence or absence of systemic disease **(Table 189-16)**.

Table 189-16 National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group TIS Staging System for Kaposi's Sarcoma

Parameter	Good Risk (Stage 0): All of the Following	Poor Risk (Stage 1): Any of the Following
Tumor (T)	Confined to skin and/or lymph nodes and/or minimal oral disease	Tumor-associated edema or ulceration Extensive oral lesions GI lesions Nonnodal visceral lesions
Immune system (I)	CD4+ T cell count ≥200/μL	CD4+ T cell count <200/μL
Systemic illness (S)	No B symptoms[a] Karnofsky performance status ≥70 No history of opportunistic infection, neurologic disease, lymphoma, or thrush	B symptoms[a] present Karnofsky performance status <70 History of opportunistic infection, neurologic disease, lymphoma, or thrush

[a]Defined as unexplained fever, night sweats, >10% involuntary weight loss, or diarrhea persisting for more than 2 weeks.

Figure 189-42

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Kaposi's sarcoma in three patients with AIDS demonstrating *(A)* periorbital edema and bruising; *(B)* classic truncal distribution of lesions; and *(C)* upper extremity lesions.

Figure 189-43

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest x-ray of a patient with AIDS and pulmonary Kaposi's sarcoma. The characteristic findings include dense bilateral lower lobe infiltrates obscuring the heart borders and pleural effusions.

A diagnosis of KS is based on biopsy of a suspicious lesion. Histologically one sees a proliferation of spindle cells and endothelial cells, extravasation of red blood cells, hemosiderin-laden macrophages, and, in early cases, an inflammatory cell infiltrate. Included in the differential diagnosis are lymphoma (particularly for oral lesions), bacillary angiomatosis, and cutaneous mycobacterial infections.

Management of KS **(Table 189-17)** should be carried out in consultation with an expert since definitive treatment guidelines do not exist. In the majority of cases, effective cART will go a long way in achieving control. Antiretroviral therapy has been associated with the spontaneous regression of KS lesions. Paradoxically, it has also been associated with the initial appearance of KS as a form of IRIS. For patients in whom tumor persists or in whom control of HIV replication is not possible, a variety of options exist. In some cases, lesions remain quite indolent, and many of these patients can be managed with no specific treatment. Fewer than 10% of AIDS patients with KS die as a consequence of their malignancy, and death from secondary infections is considerably more common. Thus, whenever possible one should avoid treatment regimens that may further suppress the immune system and increase susceptibility to opportunistic infections. Treatment is indicated under two main circumstances. The first is when a single lesion or a limited number of lesions are causing significant discomfort or cosmetic problems, such as with prominent facial lesions, lesions overlying a joint, or lesions in the oropharynx that interfere with swallowing or breathing. Under these circumstances, treatment with localized radiation, intralesional vinblastine, topical 9-*cis*-retinoic acid, or cryotherapy may be helpful. It should be noted that patients with HIV infection are particularly sensitive to the side effects of radiation therapy. This is especially true with respect to the development of radiation-induced mucositis; doses of radiation directed at mucosal surfaces, particularly in the head and neck region, should be adjusted accordingly. The use of systemic therapy, either IFN-α or chemotherapy, should be considered in patients with a large number of lesions or in patients with visceral involvement. The single most important determinant of response appears to be the CD4+ T cell count. This relationship between response rate and baseline CD4+ T cell count is particularly true for IFN-α. The response rate for patients with CD4+ T cell counts >600/μL is ~80%, while the response rate for patients with counts <150/μL is <10%. In contrast to the other systemic therapies, IFN-α provides an added advantage of having antiretroviral activity; thus, it may be the appropriate first choice for single-agent systemic therapy for early patients with disseminated disease. A variety of chemotherapeutic agents have also been shown to have activity against KS. Three of them, liposomal daunorubicin, liposomal doxorubicin, and paclitaxel, have been approved by the FDA for this indication. Liposomal daunorubicin is approved as first-line therapy for patients with advanced KS. It has fewer side effects than conventional chemotherapy. In contrast, liposomal doxorubicin and paclitaxel are approved only for KS patients who have failed standard chemotherapy. Response rates vary from 23 to 88%, appear to be comparable to what had been achieved earlier with combination chemotherapy regimens, and are greatly influenced by CD4+ T cell count.

Table 189-17 Management of AIDS-Associated Kaposi's Sarcoma

Observation and optimization of antiretroviral therapy
Single or limited number of lesions
Radiation Intralesional vinblastine
Cryotherapy
Extensive disease
Initial therapy
Interferon α (if CD4+ T cells >150/ L) Liposomal daunorubicin
Subsequent therapy
Liposomal doxorubicin
Paclitaxel
Combination chemotherapy with low-dose doxorubicin, bleomycin, and vinblastine (ABV)
Targeted radiation

Lymphomas occur with an increased frequency in patients with congenital or acquired T cell immunodeficiencies (Chap. 316). AIDS is no exception; at least 6% of all patients with AIDS develop lymphoma at some time during the course of their illness.

This is a 120-fold increase in incidence compared with the general population. In comparison to the situation with KS, primary CNS lymphoma, and most opportunistic infections, the incidence of AIDS-associated systemic lymphomas has not experienced as dramatic a decrease as a consequence of the widespread use of effective cART. Lymphoma occurs in all risk groups, with the highest incidence in patients with hemophilia and the lowest incidence in patients from the Caribbean or Africa with heterosexually acquired infection. Lymphoma is a late manifestation of HIV infection, generally occurring in patients with CD4+ T cell counts <200/μL. As HIV disease progresses, the risk of lymphoma increases. The attack rate for lymphoma increases exponentially with increasing duration of HIV infection and decreasing level of immunologic function. At 3 years following a diagnosis of HIV infection, the risk of lymphoma is 0.8% per year; by 8 years after infection, it is 2.6% per year. As individuals with HIV infection live longer as a consequence of improved cART and better treatment and prophylaxis of opportunistic infections, it is anticipated that the incidence of lymphomas may increase.

Three main categories of lymphoma are seen in patients with HIV infection: grade III or IV immunoblastic lymphoma, Burkitt's lymphoma, and primary CNS lymphoma. Approximately 90% of these lymphomas are B cell in phenotype; more than half contain EBV DNA. Some are associated with KSHV. These tumors may be either monoclonal or oligoclonal in nature and are probably in some way related to the pronounced polyclonal B cell activation seen in patients with AIDS.

Immunoblastic lymphomas account for ~60% of the cases of lymphoma in patients with AIDS. The majority of these are diffuse large B cell lymphomas (DLBCL). They are generally high grade and would have been classified as diffuse histiocytic lymphomas in earlier classification schemes. This tumor is more common in older patients, increasing in incidence from 0% in HIV-infected individuals <1 year old to >3% in those >50. Two variants of immunoblastic lymphoma that are seen primarily in HIV-infected patients are primary effusion lymphoma (PEL) and its solid variant, plasmacytic lymphoma of the oral cavity. PEL, also referred to as body cavity lymphoma, presents with lymphomatous pleural, pericardial, and/or peritoneal effusions in the absence of discrete nodal or extranodal masses. The tumor cells do not express surface markers for B cells or T cells and are felt to represent a preplasmacytic stage of differentiation. While both HHV-8 and EBV DNA sequences have been found in the genomes of the malignant cells from patients with body cavity lymphoma, KSHV is felt to be the driving force behind the oncogenesis (see above).

Small noncleaved cell lymphoma (*Burkitt's lymphoma*) accounts for ~20% of the cases of lymphoma in patients with AIDS. It is most frequent in patients 10–19 years old and usually demonstrates characteristic c-*myc* translocations from chromosome 8 to chromosomes 14 or 22. Burkitt's lymphoma is not commonly seen in the setting of immunodeficiency other than HIV-associated immunodeficiency, and the incidence of this particular tumor is more than 1000-fold higher in the setting of HIV infection than in the general population. In contrast to African Burkitt's lymphoma, where 97% of the cases contain EBV genome, only 50% of HIV-associated Burkitt's lymphomas are EBV-positive.

Primary CNS lymphoma accounts for ~20% of the cases of lymphoma in patients with HIV infection. In contrast to HIV-associated Burkitt's lymphoma, primary CNS lymphomas are usually positive for EBV. In one study, the incidence of Epstein-Barr positivity was 100%. This malignancy does not have a predilection for any particular age group. The median CD4+ T cell count at the time of diagnosis is ~50/μL. Thus, CNS lymphoma generally presents at a later stage of HIV infection than does systemic lymphoma. This may explain, at least in part, the poorer prognosis for this subset of patients.

The clinical presentation of lymphoma in patients with HIV infection is quite varied, ranging from focal seizures to rapidly growing mass lesions in the oral mucosa **(Fig. 189-44)** to persistent unexplained fever. At least 80% of patients present with extranodal disease, and a similar percentage have B-type symptoms of fever, night sweats, or weight loss. Virtually any site in the body may be involved. The most common extranodal site is the CNS, which is involved in approximately one-third of all patients with lymphoma. Approximately 60% of these cases are primary CNS lymphoma. Primary CNS lymphoma generally presents with focal neurologic deficits, including cranial nerve findings, headaches, and/or seizures. MRI or CT generally reveals a limited number (one to three) of 3- to 5-cm lesions **(Fig. 189-45)**. The lesions often show ring enhancement on contrast administration and may occur in any location. Contrast enhancement is usually less pronounced than that seen with toxoplasmosis. Locations that are most commonly involved with CNS lymphoma are deep in the white matter. The main diseases in the differential diagnosis are cerebral toxoplasmosis and cerebral Chagas' disease. In addition to the 20% of lymphomas in HIV-infected individuals that are primary CNS lymphomas, CNS disease is also seen in HIV-infected patients with systemic lymphoma. Approximately 20% of patients with systemic lymphoma have CNS disease in the form of leptomeningeal involvement. This fact underscores the importance of lumbar puncture in the staging evaluation of patients with systemic lymphoma.

Figure 189-44

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Immunoblastic lymphoma involving the hard palate of a patient with AIDS.

Figure 189-45

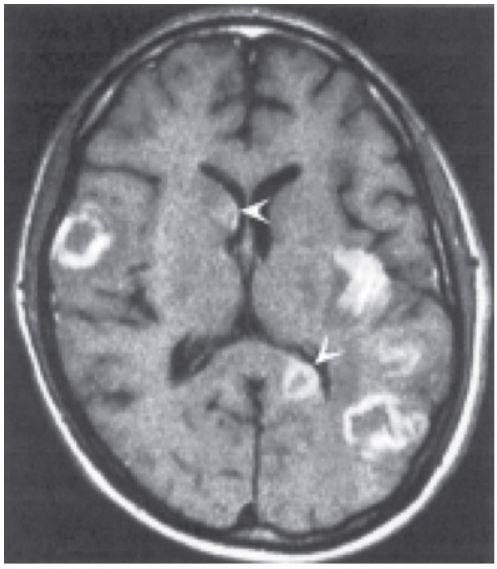

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Central nervous system lymphoma. Postcontrast T1-weighted MRI scan in a patient with AIDS, an altered mental status, and hemiparesis. Multiple enhancing lesions, some ring-enhancing, are present. The left sylvian lesion shows gyral and subcortical enhancement, and the lesions in the caudate and splenium (*arrowheads*) show enhancement of adjacent ependymal surfaces.

Systemic lymphoma is seen at earlier stages of HIV infection than primary CNS lymphoma. In one series the mean CD4+ T cell count was 189/μL. In addition to lymph node involvement, systemic lymphoma may commonly involve the GI tract, bone marrow, liver, and lung. GI tract involvement is seen in ~25% of patients. Any site in the GI tract may be involved, and patients may complain of difficulty swallowing or abdominal pain. The diagnosis is usually suspected on the basis of CT or MRI of the abdomen. Bone marrow involvement is seen in ~20% of patients and may lead to pancytopenia. Liver and lung involvement are each seen in ~10% of patients. Pulmonary disease may present as a mass lesion, multiple nodules, or an interstitial infiltrate.

Both conventional and unconventional approaches have been employed in an attempt to treat HIV-related lymphomas. Systemic lymphoma is generally treated by the oncologist with combination chemotherapy. Earlier disappointing figures are being replaced with more optimistic results for the treatment of systemic lymphoma following the availability of more effective cART and the use of rituximab in CD20+ tumors. While there is controversy regarding the use of antiretrovirals during chemotherapy, there is no question that their use overall in patients with HIV lymphoma has improved survival. As in most situations in patients with HIV disease, those with the higher CD4+ T cell counts tend to fare better. Response rates as high as 72% with a median survival of 33 months and disease-free intervals up to 9 years have been reported. Treatment of primary CNS lymphoma remains a significant challenge. Treatment is complicated by the fact that this illness usually occurs in patients with advanced HIV disease. Palliative measures such as radiation therapy provide some relief. The prognosis remains poor in this group, with a 2-year survival of 29%.

Multicentric Castleman's disease is a KSHV-associated lymphoproliferative disorder that is seen with an increased frequency in patients with HIV infection. While not a true malignancy, it shares many features with lymphoma including generalized lymphadenopathy, hepatosplenomegaly, and systemic symptoms of fever, fatigue, and weight loss. Pulmonary symptoms may be seen in ~50% of patients. KS is present in 75–82% of cases. Lymph node biopsies reveal a predominance of interfollicular plasma cells and/or germinal centers with vascularization and an "onionskin" (hyaline vascular) appearance. Prior to the availability of cART, HIV-infected patients with multicentric Castleman's disease had a 15-fold increased risk of developing non-Hodgkin's lymphoma compared with HIV-infected patients in general. Treatment typically involves chemotherapy. Anecdotal reports of success with rituximab suggest that more specific treatment may be successful, although in one series treatment with rituximab was associated with worsening of coexisting KS. The median survival of patients with treated multicentric Castleman's disease pre-cART was 14 months. This has increased to a 2-year survival of more than 90% in the era of cART.

Evidence of infection with *human papillomavirus* (HPV), associated with *intraepithelial dysplasia of the cervix* or *anus*, is approximately twice as common in HIV-infected individuals as in the general population and can lead to intraepithelial neoplasia and eventually invasive cancer. In a series of studies, HIV-infected men were examined for evidence of anal dysplasia, and Papanicolaou (Pap) smears were found to be abnormal in 20–80%. These changes tend to persist and are generally not affected by cART, raising the possibility of a subsequent transition to a more malignant condition. While the incidence of an abnormal Pap smear of the cervix is ~5% in otherwise healthy women, the incidence of abnormal cervical smears in women with HIV infection is 30–60%, and *invasive cervical cancer* is included as an AIDS-defining condition. Thus far, however, only small increases in the incidence of cervical or anal cancer have been seen as a consequence of HIV infection. However, given this high rate of dysplasia, a comprehensive gynecologic and rectal examination, including Pap smear, is indicated at the initial evaluation and 6 months later for all patients with HIV infection. If these examinations are negative at both time points, the patient should be followed with yearly evaluations. If an initial or repeat Pap smear shows evidence of severe inflammation with reactive squamous changes, the next Pap smear should be performed at 3 months. If, at any time, a Pap smear shows evidence of squamous intraepithelial lesions, colposcopic examination with biopsies as indicated should be performed. The 2-year survival rate for HIV-infected patients with invasive cervical cancer is 64% compared with 79% in non-HIV-infected patients. The most common HPV genotypes in the general population and the genotypes upon which current HPV vaccines are based are 16 and 18. This is not the case in the HIV-infected population, where other genotypes such as 56 and 53 predominate. This raises concerns as to the potential effectiveness of the current HPV vaccines for HIV-infected patients.

IDIOPATHIC CD4+ T LYMPHOCYTOPENIA

A syndrome was recognized in 1992 that was characterized by an absolute CD4+ T cell count of <300/μL or <20% of total T cells on a minimum of two occasions at least 6 weeks apart; no evidence of HIV-1, HIV-2, HTLV-I, or HTLV-II on testing; and the absence of any defined immunodeficiency or therapy associated with decreased levels of CD4+ T cells. By mid-1993, ~100 patients had been described. After extensive multicenter investigations, a series of reports were published in early 1993, which together allowed a number of conclusions. Idiopathic CD4+ lymphocytopenia (ICL) is a very rare syndrome, as determined by

studies of blood donors and cohorts of HIV-seronegative men who have sex with men. Cases were clearly identified as early as 1983 and were remarkably similar to the clinical features of ICL that had been identified decades earlier. The definition of ICL based on CD4+ T cell counts coincided with the ready availability of testing for CD4+ T cells in patients suspected of being immunodeficient. Although, as a result of immune deficiency, certain patients with ICL develop some of the opportunistic diseases (particularly cryptococcosis, nontuberculous mycobacterial infections, and cervical dysplasia) seen in HIV-infected patients, the syndrome is demographically, clinically, and immunologically unlike HIV infection and AIDS. Fewer than half of the reported ICL patients had risk factors for HIV infection, and there were wide geographic and age distributions. The fact that a significant proportion of patients did have risk factors probably reflects a selection bias, in that physicians who take care of HIV-infected patients are more likely to monitor CD4+ T cells. Approximately half of the patients are women, compared with approximately one-third among HIV-infected individuals in the United States. Many patients with ICL remained clinically stable, and their condition did not deteriorate progressively as is common with seriously immunodeficient HIV-infected patients. Approximately 15% of patients with ICL experience spontaneous reversal of the CD4+ T lymphocytopenia. Immunologic abnormalities in ICL are somewhat different from those of HIV infection. ICL patients often have increases in CD4+ T cell activation with decreases in CD8+ T cells and B cells. Furthermore, immunoglobulin levels are either normal or, more commonly, decreased in patients with ICL, compared with the usual hypergammaglobulinemia of HIV-infected individuals. Virologic studies of these patients have revealed no evidence of HIV-1, HIV-2, HTLV-I, or HTLV-II or of any other mononuclear cell–tropic virus. Furthermore, there has been no epidemiologic evidence to suggest that a transmissible microbe was involved. The cases of ICL have been widely dispersed, with no clustering. Close contacts and sexual partners who were studied were clinically well and were serologically, immunologically, and virologically negative for HIV. ICL is a heterogeneous syndrome, and it is highly likely that there is no common cause; however, there may be common causes among subgroups of patients that are currently unrecognized.

Patients who present with laboratory data consistent with ICL should be worked up for underlying diseases that could be responsible for the immune deficiency. If no underlying cause is detected, no specific therapy should be initiated. However, if opportunistic diseases occur, they should be treated appropriately (see above). Depending on the level of the CD4+ T cell count, patients should receive prophylaxis for the commonly encountered opportunistic infections.

Treatment: AIDS and Related Disorders

GENERAL PRINCIPLES OF PATIENT MANAGEMENT

The CDC guidelines call for the testing for HIV infection to be a part of routine medical care. It is recommended that the patient be informed of the intention to test, as is the case with other routine laboratory determinations, and be given the opportunity to "opt out." Such an approach is critical to the goal of identifying as many infected individuals as possible since ~21% of the >1 million individuals in the United States who are HIV-infected are not aware of their status. Under these circumstances of routine testing, although it is desirable, pretest counseling may not always be built into the testing process. However, no matter how well prepared a patient is for adversity, the discovery of a diagnosis of HIV infection is a devastating event. Thus, physicians should be sensitive to this fact and, where possible, execute some degree of pretest counseling to at least partially prepare the patient should the results demonstrate the presence of HIV infection. Following a diagnosis of HIV infection, the health care provider should be prepared to immediately activate support systems for the newly diagnosed patient. These should include an experienced social worker or nurse who can spend time talking to the person and ensuring that he or she is emotionally stable. Most communities have HIV support centers that can be of great help in these difficult situations.

The treatment of patients with HIV infection requires not only a comprehensive knowledge of the possible disease processes that may occur and up-to-date knowledge of and experience with cART, but also the ability to deal with the problems of a chronic, potentially life-threatening illness. A comprehensive knowledge of internal medicine is required to deal with the changing spectrum of illness associated with HIV infection. Great advances have been made in the treatment of patients with HIV infection. The appropriate use of potent cART and other treatment and prophylactic interventions are of critical importance in providing each patient with the best opportunity to live a long and healthy life despite the presence of HIV infection. In contrast to the earlier days of this epidemic, a diagnosis of HIV infection need no longer be equated with having an inevitably fatal disease. In addition to medical interventions, the health care provider has a responsibility to provide each patient with appropriate counseling and education concerning their disease as part of a comprehensive care plan. Patients must be educated about the potential transmissibility of their infection and about the fact that while health care providers may refer to levels of the virus as "undetectable," this is more a reflection of the sensitivity of the assay being used to measure the virus than a comment on the presence or absence of the virus. It is important for patients to be aware that the virus is still present and capable of being transmitted at all stages of HIV disease. Thus, there must be frank discussions concerning sexual practices and the sharing of syringes and other paraphernalia used in illicit drug use. The treating physician not only must be aware of the latest medications available for patients with HIV infection but also must educate patients concerning the natural history of their illness and listen and be sensitive to their fears and concerns. As with other diseases, therapeutic decisions should be made in consultation with the patient, when possible, and with the patient's proxy if the patient is incapable of making decisions. In this regard, it is recommended that all patients with HIV infection, and in particular those with CD4+ T cell counts <200/μL, designate a trusted individual with durable power of attorney to make medical decisions on their behalf, if necessary.

Following a diagnosis of HIV infection, there are several examinations and laboratory studies that should be performed to help determine the extent of disease and provide baseline standards for future reference **(Table 189-18)**. In addition to routine chemistry, fasting lipid profile, aspartate aminotransferase, alanine aminotransferase, total and direct bilirubin, fasting glucose and hematology screening panels, Pap smear, urinalysis, and chest x-ray, one should also obtain a CD4+ T cell count, two separate plasma HIV RNA levels, an HIV resistance test, a rapid plasma reagin or VDRL test, an anti-*Toxoplasma* antibody titer, and serologies for hepatitis A, B, and C. A PPD test should be done and an MMSE performed and recorded. A pregnancy test should be done in women in whom the drug efavirenz is being considered, and HLA-B5701 testing should be done in all patients in whom the drug abacavir is being considered. Patients should be immunized with pneumococcal polysaccharide, with annual influenza shots, and, if seronegative for these viruses, with hepatitis A and hepatitis B vaccines. The status of hepatitis C infection should be determined. In addition, patients should be counseled with regard to sexual practices and needle sharing, and counseling should be offered to those whom the patient knows or suspects may also be infected. Once these baseline activities are performed, short- and long-term medical management strategies should be developed based on the most recent information available and modified as new information becomes available. The field of HIV medicine is changing rapidly, and it is difficult to remain fully up to date. Fortunately there are a series of excellent sites on the Internet that are frequently updated, and they provide the most recent information on a variety of topics, including consensus panel reports on treatment **(Table 189-19)**.

Table 189-18 Initial Evaluation of the Patient with HIV Infection
History and physical examination
Routine chemistry and hematology
AST, ALT, direct and indirect bilirubin

Lipid profile and fasting glucose
CD4+ T lymphocyte count
Two plasma HIV RNA levels
HIV resistance testing
HLA-B5701 screening
RPR or VDRL test
Anti-*Toxoplasma* antibody titer
PPD skin test
Mini-Mental Status Examination
Serologies for hepatitis A, hepatitis B, and hepatitis C
Immunization with pneumococcal polysaccharide; influenza as indicated
Immunization with hepatitis A and hepatitis B if seronegative
Counseling regarding natural history and transmission
Help contacting others who might be infected

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; PPD, purified protein derivative; RPR, rapid plasma reagin; VDRL, Venereal Disease Research Laboratory.

Table 189-19 Resources Available on the World Wide Web on HIV Disease

http://aidsinfo.nih.gov	AIDSinfo, a service of the U.S. Department of Health and Human Services, posts federally approved treatment guidelines for HIV and AIDS; provides information on federally funded and privately funded clinical trials and CDC publications and data
www.cdcnpin.org	Updates on epidemiologic data and prevention information from the CDC

Note: CDC, Centers for Disease Control and Prevention.

ANTIRETROVIRAL THERAPY

Combination antiretroviral therapy (cART), also referred to as highly active antiretroviral therapy (HAART), is the cornerstone of management of patients with HIV infection. Following the initiation of widespread use of cART in the United States in 1995–1996, marked declines were noted in the incidence of most AIDS-defining conditions (Fig. 189-34). Suppression of HIV replication is an important component in prolonging life as well as in improving the quality of life in patients with HIV infection. Adequate suppression requires strict adherence to prescribed regimens of antiretroviral drugs. This has been facilitated by the coformulations of antiretrovirals and the development of once-daily regimens. Unfortunately, many of the most important questions related to the treatment of HIV disease currently lack definitive answers. Among them are the questions of when therapy should be started, what the best initial regimen is, when a given regimen should be changed, and what it should be changed to when a change is made. Notwithstanding these uncertainties, the physician and patient must come to a mutually agreeable plan based on the best available data. In an effort to facilitate this process, the U.S. Department of Health and Human Services makes available on the Internet (www.aidsinfo.nih.gov) a series of periodically updated guidelines, including "*Principles of Therapy of HIV Infection* ," "*Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents*," and "*Guidelines for the Prevention of Opportunistic Infections in Persons Infected with Human Immunodeficiency Virus*." At present, an extensive clinical trials network, involving both clinical investigators and patient advocates, is in place attempting to develop improved approaches to therapy. Consortia comprising representatives of academia, industry, independent foundations, and the federal government are involved in the process of drug development, including a wide-ranging series of clinical trials. As a result, new therapies and new therapeutic strategies are continually emerging. New drugs are often available through expanded access programs prior to official licensure. Given the complexity of this field, decisions regarding cART are best made in consultation with experts.

Currently available drugs for the treatment of HIV infection fall into four categories: those that inhibit the viral reverse transcriptaseenzyme (nucleoside and nucleotide reverse transcriptase inhibitors; nonnucleoside reverse transcriptase inhibitors), those that inhibit the viral protease enzyme (protease inhibitors), those that inhibit the viral integrase enzyme (integrase inhibitors), and those that interfere with viral entry (fusion inhibitors; CCR5 antagonists) **(Table 189-20**; **Fig. 189-46).**

Table 189-20 Antiretroviral Drugs Used in the Treatment of HIV Infection

Drug	Status	Indication	Dose in Combination	Supporting Data	Toxicity
Nucleoside or Neucleotide Reverse Transcriptase Inhibitors					
Zidovudine (AZT, azidothymidine, Retrovir, 3'azido-3'-deoxythymidine)	Licensed	Treatment of HIV infection in combination with other antiretroviral agents	200 mg q8h or 300 mg bid	19 vs 1 death in original placebo-controlled trial in 281 patients with AIDS or ARC	Anemia, granulocytopenia, myopathy, lactic acidosis, hepatomegaly with steatosis, headache, nausea, nail pigmentation, lipid abnormalities, lipoatrophy, hyperglycemia
		Prevention of maternal-fetal HIV transmission		In pregnant women with CD4+ T cell count ≥200/μL, AZT PO beginning at weeks 14–34 of gestation plus IV drug during labor and delivery plus PO AZT to infant for 6 weeks decreased transmission of HIV by 67.5% (from 25.5% to 8.3%), *n* = 363	
Didanosine	Licensed	For treatment	Buffered: Requires	Clinically superior to AZT as	Pancreatitis,

(Videx, Videx EC, ddI, dideoxyinosine, 2',3'-dideoxyinosine)		of HIV infection in combination with other antiretroviral agents	2 tablets to achieve adequate buffering of stomach acid; should be administered on an empty stomach ≥60 kg: 200 mg bid <60 kg: 125 mg bid Enteric coated: ≥60 kg: 400 mg qd < 60 kg: 250 mg qd	monotherapy in 913 patients with prior AZT therapy; clinically superior to AZT and comparable to AZT + ddI and AZT + ddC in 1067 AZT-naïve patients with CD4+ T cell counts of 200–500/μL	peripheral neuropathy, abnormalities on liver function tests, lactic acidosis, hepatomegaly with steatosis, optic neuritis, nausea, hyperglycemia
Zalcitabine (ddC, HIVID, 2'3'-dideoxycytidine)	Licensed Discontinued in 2006	In combination with other antiretroviral agents for the treatment of HIV infection	0.75 mg tid	Clinically inferior to AZT monotherapy as initial treatment; clinically as good as ddI in advanced patients intolerant to AZT; in combination with AZT, was clinically superior to AZT alone in patients with AIDS or CD4+ T cell count <350/μL	Peripheral neuropathy, pancreatitis, lactic acidosis, hepatomegaly with steatosis, oral ulcers
Stavudine (d4T, Zerit, 2'3'-didehydro-3'-dideoxythymidine)	Licensed	Treatment of HIV-infected patients in combination with other antiretroviral agents	≥60 kg: 40 mg bid <60 kg: 30 mg bid	Superior to AZT with respect to changes in CD4+ T cell counts in 359 patients who had received ≥24 weeks of AZT; following 12 weeks of randomization, the CD4+ T cell count had decreased in AZT-treated controls by a mean of 22/μL, while in stavudine-treated patients, it had increased by a mean of 22/μL	Peripheral neuropathy, pancreatitis, lactic acidosis, hepatomegaly with steatosis, ascending neuromuscular weakness, lipodystrophy, lipid abnormalities, hyperglycemia
Lamivudine (Epivir, 2'3'-dideoxy-3'-thiacytidine, 3TC)	Licensed	In combination with other antiretroviral agents for the treatment of HIV infection	150 mg bid 300 mg qd	In combination with AZT superior to AZT alone with respect to changes in CD4+ T cell counts in 495 patients who were zidovudine-naïve and 477 patients who were zidovudine-experienced; overall CD4+ T cell counts for the zidovudine group were at baseline by 24 weeks, while in the group treated with zidovudine plus lamivudine, they were 10–50 cells/μL above baseline; 54% decrease in progression to AIDS/death compared with AZT alone	Flare of hepatitis in HBV-coinfected patients who discontinue drug
Emtricitabine (FTC, Emtriva)	Licensed	In combination with other antiretroviral agents for the treatment of HIV infection	200 mg qd	Comparable to d4T in combination with ddI and efavirenz in 571 treatment-naïve patients; similar to 3TC in combination with AZT or d4T + NNRTI or PI in 440 patients doing well for ≥12 weeks on a 3TC regimen	Hepatotoxicity in HBV-coinfected patients who discontinue drug, skin discoloration
Abacavir (Ziagen)	Licensed	For treatment of HIV infection in combination with other antiretroviral agents	300 mg bid	Abacavir + AZT + 3TC equivalent to indinavir + AZT + 3TC with regard to viral load suppression (∑60% in each group with <400 HIV RNA copies/mL plasma) and CD4+ T cell increase (∑100/μL in each group) at 24 weeks	Hypersensitivity reaction In HLA-B5701+ individuals (can be fatal); fever, rash, nausea, vomiting, malaise or fatigue, and loss of appetite
Tenofovir (Viread)	Licensed	For use in combination with other antiretroviral agents when treatment is indicated	300 mg qd	Reduction of ∑0.6 log in HIV-1 RNA levels when added to background regimen in treatment-experienced patients	Renal osteomalacia, flare of hepatitis in HBV-coinfected patients who discontinue drug
Non-Nucleoside Reverse Transcriptase Inhibitors					
Delavirdine (Rescriptor)	Licensed	For use in combination with appropriate antiretrovirals when treatment is warranted	400 mg tid	Delavirdine + AZT superior to AZT alone with regard to viral load suppression at 52 weeks	Skin rash, abnormalities in liver function tests
Nevirapine (Viramune)	Licensed	In combination with other antiretroviral agents for treatment of progressive HIV infection	200 mg/d x 14 days then 200 mg bid *or* 400 mg extended release qd	Increase in CD4+ T cell count, decrease in HIV RNA when used in combination with nucleosides	Skin rash, hepatotoxicity
Efavirenz (Sustiva)	Licensed	For treatment of HIV infection	600 mg qhs	Efavirenz + AZT + 3TC comparable to indinavir + AZT +	Rash, dysphoria, elevated liver

		in combination with other antiretroviral agents		3TC with regard to viral load suppression (a higher percentage of the efavirenz group achieved viral load <50 copies/mL, but the discontinuation rate in the indinavir group was unexpectedly high, accounting for most treatment "failures"); CD4 cell increase (∑140/μL in each group) at 24 weeks	function tests, drowsiness, abnormal dreams, depression, lipid abnormalities, potentially teratogenic
Etravirine	Licensed	In combination with other antiretroviral agents in treatment-experienced patients whose HIV is resistant to nonnucleoside reverse transcriptase inhibitors and other antiretroviral medications	200 mg bid	Higher rates of HIV RNA suppression to <50 copies/mL (56% vs 39%); greater increases in CD4+ T cell count (89 vs 64 cells) compared to placebo when given in combination with an optimized background regimen	Rash, nausea, hypersensitivity reactions
Rilpivirine	Investigational	In combination	25 mg qd	In combination with dual nucleosides non-inferior to Efavirenz. Move virologic failures following rnitial suppression.	Nausea, dizziness, somnolence, vertigo, less CNS toxicity and rash than Efavirenz
Protease Inhibitors					
Saquinavir mesylate (Invirase—hard-gel capsule)	Licensed	In combination with other antiretroviral agents when therapy is warranted	1000 mg + 100 mg ritonavir bid	Increases in CD4+ T cell counts, reduction in HIV RNA most pronounced in combination therapy with ddC; 50% reduction in first AIDS-defining event or death in combination with ddC compared with either agent alone	Diarrhea, nausea, headaches, hyperglycemia, fat redistribution, lipid abnormalities, PR and QT interval prolongation
(Fortovase—soft-gel capsule)	Licensed Discontinued 2006	For use in combination with other antiretroviral agents when treatment is warranted	1200 mg tid	Reduction in the mortality rate and AIDS-defining events for patients who received hard-gel formulation in combination with ddC	Diarrhea, nausea, abdominal pain, headaches, hyperglycemia, fat redistribution, lipid abnormalities
Ritonavir (Norvir)	Licensed	In combination with other antiretroviral agents for treatment of HIV infection when treatment is warranted	600 mg bid (also used in lower doses as pharmaco-kinetic booster)	Reduction in the cumulative incidence of clinical progression or death from 34 to 17% in patients with CD4+ T cell count <100/μL treated for a median of 6 months	Nausea, abdominal pain, hyperglycemia, fat redistribution, lipid abnormalities, may alter levels of many other drugs, including saquinavir, paresthesias, hepatitis
Indinavir sulfate (Crixivan)	Licensed	For treatment of HIV infection in combination with other antiretroviral agents when antiretroviral treatment is warranted	800 mg q8h or 800 mg + 100 mg ritonavir bid or 1000 mg q8h when used with efavirenz or nevirapine	Increase in CD4+ T cell count by 100/μL and 2-log decrease in HIV RNA levels when given in combination with zidovudine and lamivudine; decrease of 50% in risk of progression to AIDS or death when given with zidovudine and lamivudine compared with zidovudine and lamivudine alone	Nephrolithiasis, indirect hyperbilirubinemia, hyperglycemia, fat redistribution, lipid abnormalities, nausea, transaminase elevations, headache
Nelfinavir mesylate (Viracept)	Licensed	For treatment of HIV infection in combination with other antiretroviral agents when antiretroviral therapy is warranted	750 mg tid or 1250 mg bid	2.0-log decline in HIV RNA when given in combination with stavudine	Diarrhea, loose stools, hyperglycemia, fat redistribution, lipid abnormalities, transaminase elevations
Amprenavir (Agenerase)	Licensed	In combination with other antiretroviral agents for treatment of HIV infection	Amprenavir: 1200 mg bid or 600 mg + 100 mg ritonavir bid or 1200 mg + 200 mg ritonavir qd	In treatment-naïve patients, amprenavir + AZT + 3TC superior to AZT + 3TC with regard to viral load suppression (53% vs 11% with <400 HIV RNA copies/mL plasma at 24 weeks); CD4+ T cell responses similar between treatment groups; in treatment-experienced patients, amprenavir + NRTIs similar to indinavir + NRTIs with regard to viral load suppression (43% vs	Nausea, vomiting, diarrhea, rash, oral paresthesias, elevated liver function tests, hyperglycemia, fat redistribution, lipid abnormalities, headache, nephrolithiasis
Fosamprenavir (Lexiva)	Licensed		Fosamprenavir: 1400 mg bid or 700 mg + 100 mg		

			ritonavir bid or 1400 mg + 200 mg ritonavir qd	53% with <400 HIV RNA copies/mL plasma at 24 weeks); CD4+ T cell responses superior in the indinavir + NRTIs group	
Lopinavir/ritonavir (Kaletra)	Licensed	For treatment of HIV infection in combination with other antiretroviral agents	400 mg/100 mg bid	In treatment-naïve patients, lopinavir/ritonavir + d4T + 3TC superior to nelfinavir + d4T + 3TC with regard to viral load suppression (79% vs 64% with <400 HIV RNA copies/mL at 40 weeks); CD4+ T cell increases similar in both groups	Diarrhea, hyperglycemia, fat redistribution, lipid abnormalities, nausea, pancreatitis, elevated liver function tests, PR and QT interval prolongations
Atazanavir (Reyataz)	Licensed	For treatment of HIV infection in combination with other antiretroviral agents	400 mg qd or 300 mg qd + ritonavir 100 mg qd when given with efavirenz	Comparable to efavirenz when given in combination with AZT + 3TC in a study of 810 treatment-naïve patients; comparable to nelfinavir when given in combination with d4T + 3TC in a study of 467 treatment-naïve patients	Hyperbilirubinemia, PR prolongation, nausea, vomiting, hyperglycemia, fat maldistribution, rash transaminase elevations
Tipranavir (Aptivus)	Licensed	In combination with 200 mg ritonavir for combination therapy in treatment-experienced adults	500 mg + 200 mg ritonavir twice daily	At 24 weeks, patients with prior extensive exposure to ARV therapy showed a –0.8 log change in HIV RNA levels and a 34-cell increase in CD4+ T cells compared with –0.25 log and 4 cells in the control arm; inferior to lopinavir/ritonavir in a randomized, controlled trial in naïve patients	Diarrhea, nausea, fatigue, headache, skin rash, hepatotoxicity, intracranial hemorrhage, hyperglycemia, lipid abnormalities, fat redistribution
Darunavir (Prezista)	Licensed	In combination with 100 mg ritonavir for combination therapy in treatment-experienced adults	600 mg + 100 mg ritonavir twice daily with food	At 24 weeks, patients with prior extensive exposure to antiretrovirals treated with a new combination including darunavir showed a –1.89 log change in HIV RNA levels and a 92-cell increase in CD4+ T cells compared with –0.48 log and 17 cells in the control arm	Diarrhea, nausea, headache, skin rash, hepatoxicity, hyperlipidemia, hyperglycemia
Entry Inhibitors					
Enfuvirtide (Fuzeon)	Licensed	In combination with other agents in treatment-experienced patients with evidence of HIV-1 replication despite ongoing antiretroviral therapy	90 mg SC bid	In treatment of experienced patients, superior to placebo when added to new optimized background (37% vs 16% with <400 HIV RNA copies/mL at 24 weeks; + 71 vs + 35 CD4+ T cells at 24 weeks)	Local injection reactions, hypersensitivity reactions, increased rate of bacterial pneumonia
Maraviroc (Selzentry)	Licensed	In combination with other antiretroviral agents in adults infected with only CCR5-tropic HIV-1	150–600 mg bid depending on concomitant medications (see text)	At 24 weeks, among 635 patients with CCR5-tropic virus and HIV-1 RNA >5000 copies/mL despite at least 6 months of prior therapy with at least 1 agent from 3 of the 4 antiretroviral drug classes, 61% of patients randomized to maraviroc achieved HIV RNA levels <400 copies/mL compared with 28% of patients randomized to placebo	Hepatotoxicity, nasopharyngitis, fever, cough, rash, abdominal pain, dizziness, musculoskeletal symptoms
Integrase Inhibitor					
Raltegravir (Isentress)	Licensed	In combination with other antiretroviral agents	400 mg bid	At 24 weeks, among 436 patients with 3-class drug resistance, 76% of patients randomized to receive raltegravir achieved HIV RNA levels <400 copies/mL compared with 41% of patients randomized to receive placebo	Nausea, headache, diarrhea, CPK elevation, muscle weakness and rhabdomyolysis
Elvitegravir	Investigational	In combination	150 mg qdt pharmacoenhancing agent cobicistat	Non-inferior to raltegravir in treatment-experienced patients	

Abbreviations: ARC, AIDS-related complex; NRTIs, nonnucleoside reverse transcriptase inhibitors.

Figure 189-46

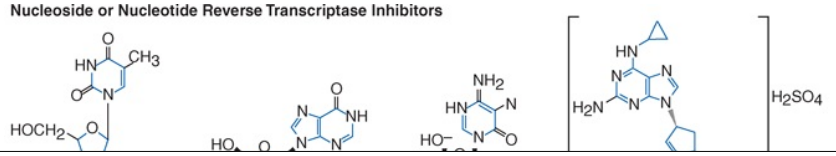

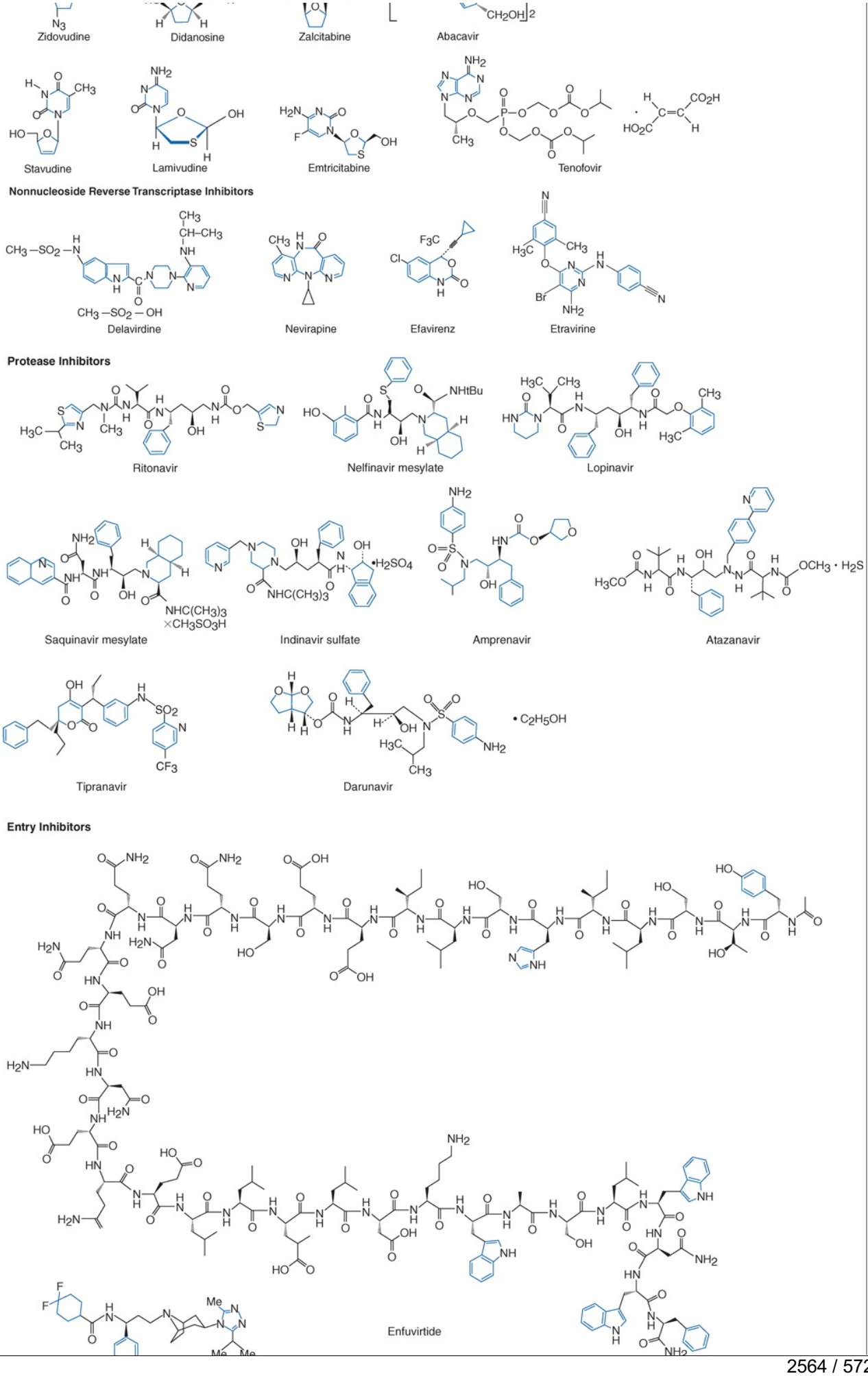
Zidovudine
Didanosine
Zalcitabine
Abacavir
Stavudine
Lamivudine
Emtricitabine
Tenofovir
Nonnucleoside Reverse Transcriptase Inhibitors
Delavirdine
Nevirapine
Efavirenz
Etravirine
Protease Inhibitors
Ritonavir
Nelfinavir mesylate
Lopinavir
Saquinavir mesylate
Indinavir sulfate
Amprenavir
Atazanavir
Tipranavir
Darunavir
Entry Inhibitors
Enfuvirtide

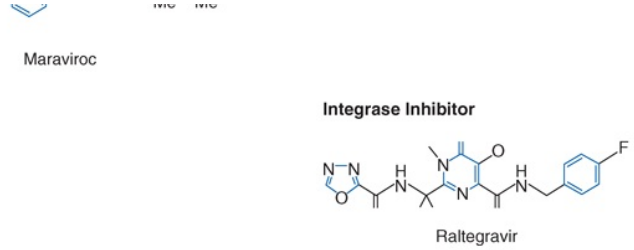

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Molecular structures of antiretroviral agents.

The FDA-approved reverse transcriptase inhibitors include the *nucleoside analogues* zidovudine, didanosine, zalcitabine, stavudine, lamivudine, abacavir, and emtricitabine; the *nucleotide analogue* tenofovir; and the *nonnucleoside reverse transcriptase inhibitors* nevirapine, delavirdine, efavirenz, and etravirine (Fig. 189-46; Table 189-20**)**. These represent the first class of drugs licensed for the treatment of HIV infection. They are indicated for this use as part of combination regimens. It should be stressed that none of these drugs should be used as monotherapy for HIV infection due to the relative ease with which drug resistance may develop under such circumstances. Thus, when lamivudine, emtricitabine, or tenofovir is used to treat hepatitis B infection in the setting of HIV infection, one should ensure that the patient is also on additional antiretroviral medication. The reverse transcriptase inhibitors block the HIV replication cycle at the point of RNA-dependent DNA synthesis, the reverse transcription step. While the nonnucleoside reverse transcriptase inhibitors are quite selective for the HIV-1 reverse transcriptase, the nucleoside and nucleotide analogues inhibit a variety of DNA polymerases in addition to those of the HIV-1 reverse transcriptase. For this reason, serious side effects are more varied with the nucleoside analogues and include mitochondrial damage that can lead to hepatic steatosis and lactic acidosis as well as peripheral neuropathy and pancreatitis. The use of either of the thymidine analogues zidovudine and stavudine has been associated with a syndrome of hyperlipidemia, glucose intolerance/insulin resistance, and fat redistribution often referred to as *lipodystrophy syndrome* (discussed in "Diseases of the Endocrine System and Metabolic Disorders," above).

Zidovudine (AZT; 3'-azido-2',3'-dideoxythymidine) was the first drug approved for the treatment of HIV infection and is the prototype nucleoside analogue. These compounds, in which the hydroxyl group in the 3' position of the ribose moiety is substituted with a hydrogen or other chemical group, act as DNA chain terminators owing to their inability to form a 3'–5' phosphodiester linkage with another nucleoside. They bind much more avidly to the active site of the RNA-dependent DNA polymerase of HIV (reverse transcriptase) than to the active site of mammalian cell DNA polymerases; this explains their selective effect on HIV replication. Zidovudine also has a relatively high avidity for the DNA polymerase γ of human mitochondria. This may contribute to the development of the fatty liver and myopathy sometimes observed in patients taking zidovudine. As with all the nucleoside analogues, the active form of zidovudine is the triphosphate, and the rate of phosphorylation, a thymidine kinase–dependent pathway, may be different in different cells. This may explain why zidovudine is more effective at inhibiting HIV replication in some cells than in others. The clinical benefit of zidovudine was clearly established in 1986 in a phase II, randomized, placebo-controlled trial in patients with advanced HIV disease. However, while treatment of patients with early stages of HIV infection with zidovudine monotherapy was associated with increases in CD4+ T cell count, it was not associated with a better overall outcome than waiting until later to treat. Subsequent trials established the ability of this drug to dramatically decrease the incidence of perinatal transmission of HIV from infected mother to infant. Eventually a series of studies demonstrated the superiority of cART regimens over zidovudine alone, and combination therapy (discussed below) remains the standard of treatment today. Among the side effects of zidovudine at the initiation of therapy are fatigue, malaise, nausea, and headache. These side effects often subside over time. Patients on zidovudine may develop a macrocytic anemia, neutropenia, myopathy, cardiomyopathy, and lactic acidosis associated with fatty infiltration of the liver. As with every antiretroviral drug, HIV has the ability to develop resistance to zidovudine. Zidovudine resistance has been reported to occur ~6 months following the initiation of zidovudine monotherapy. More recently, zidovudine-resistant viruses have been noted in patients with acute infection prior to the initiation of therapy, implying that zidovudine-resistant viruses can be transmitted from person to person. Resistance emerges more rapidly in late-stage patients, presumably as a consequence of a greater degree of viral replication and thus a greater opportunity for mutation. A variety of amino acid changes including substitutions, insertions, and deletions have been reported to confer zidovudine resistance **(Fig. 189-47)**. One combination preparation, Combivir, consists of zidovudine and lamivudine, while another, Trizivir, consists of zidovudine, lamivudine, and abacavir.

Figure 189-47

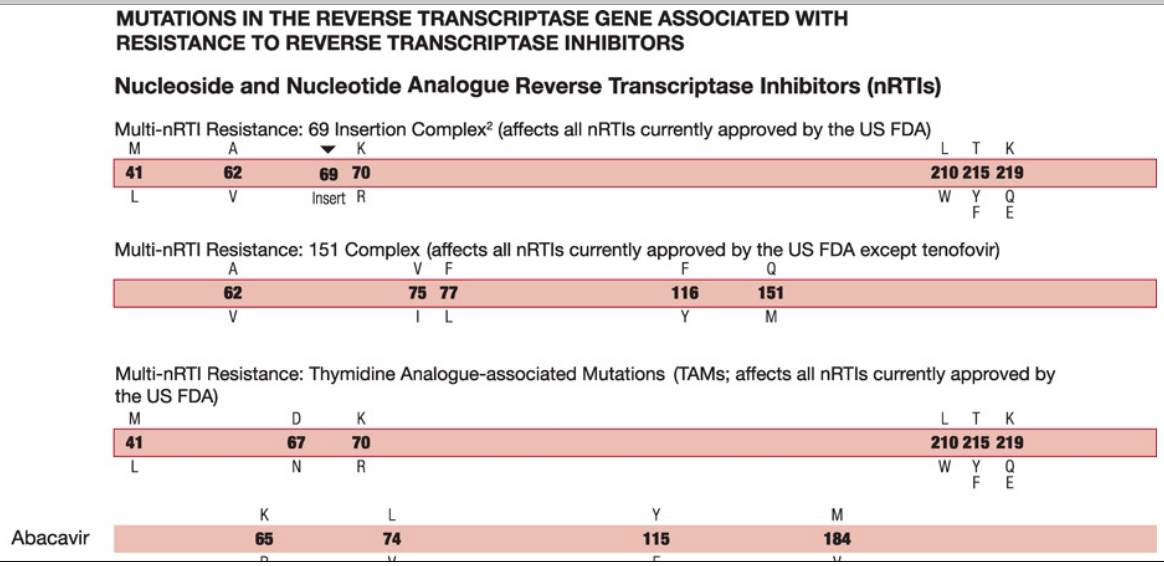

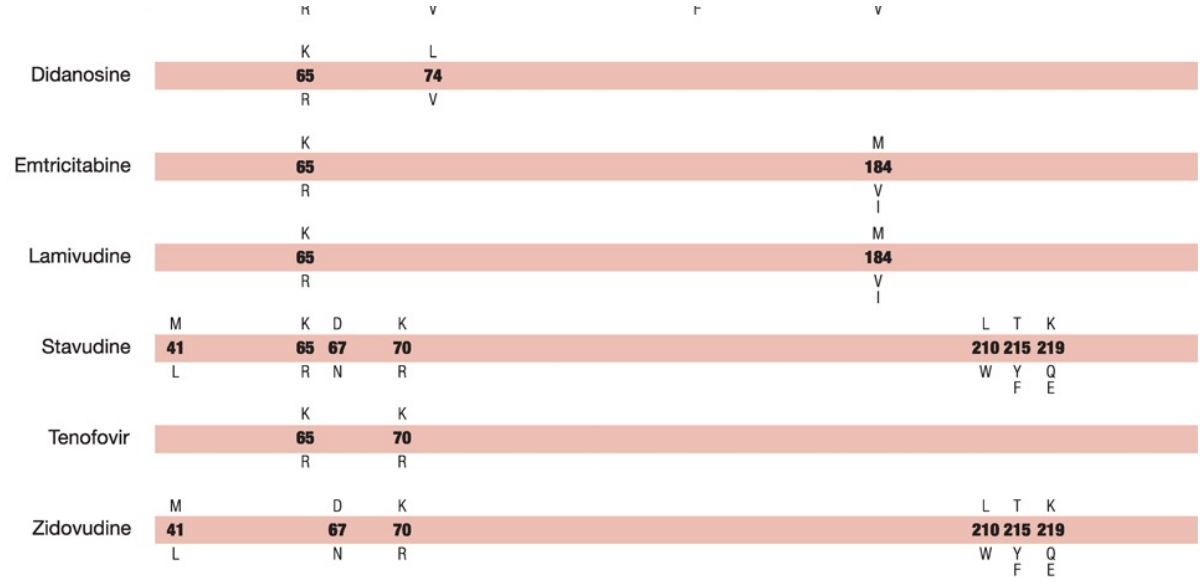

Nonnucleoside Analogue Reverse Transcriptase Inhibitors (NNRTIs)

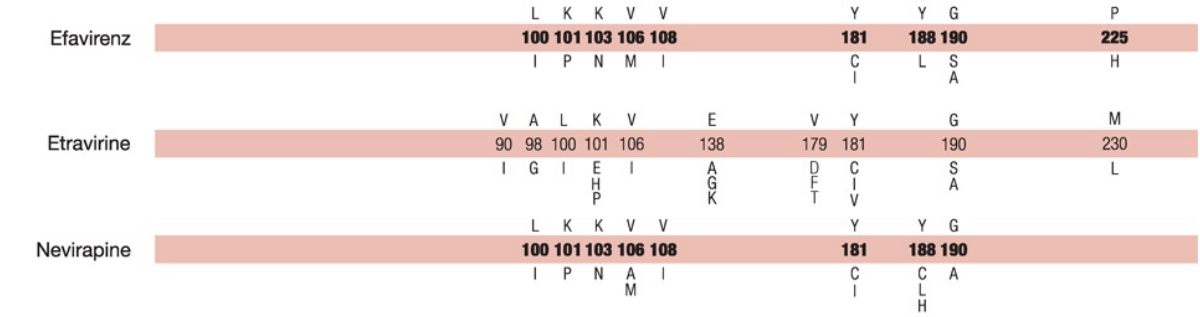

MUTATIONS IN THE PROTEASE GENE ASSOCIATED WITH RESISTANCE TO PROTEASE INHIBITORS

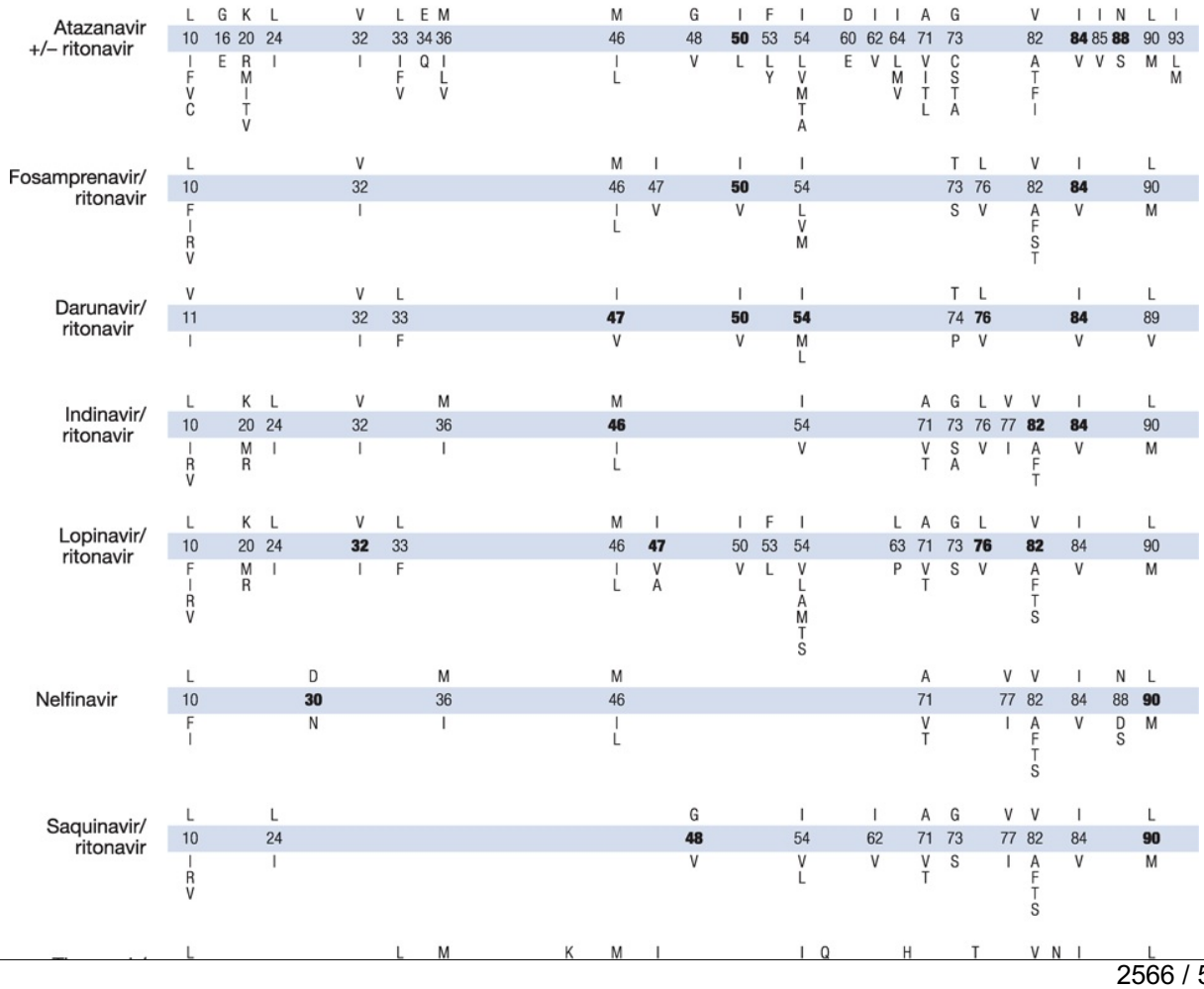

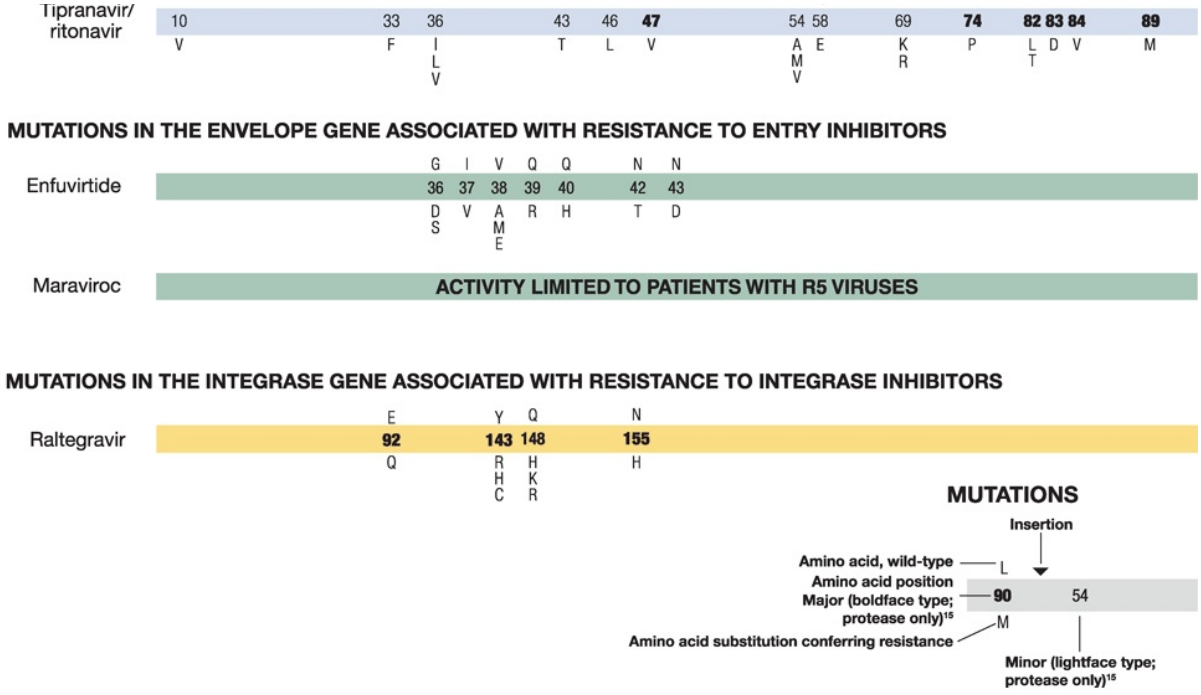

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Amino acid substitutions conferring resistance to antiretroviral drugs. For each amino acid residue, the letter above the bar indicates the amino acid associated with wild-type virus and the letter(s) below indicate the substitution(s) that confer viral resistance. The number shows the position of the mutation in the protein. Mutations selected by protease inhibitors in Gag cleavage sites are not listed. HR1 indicates first heptad repeat; NAMs indicates nRTI-associated mutations; nRTI indicates nucleoside reverse transcriptase inhibitor; NNRTI indicates nonnucleoside reverse transcriptase inhibitor; PI indicates protease inhibitor. Amino acid abbreviations: A, alanine; C, cysteine; D, aspartate; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine. *[Reprinted with permission from the International AIDS Society—USA. VA Johnson et al: Update of the Drug Resistance Mutations in HIV-1: December 2010. Topics in HIV Medicine 18:156–163, 2010. Updated information (and thorough explanatory notes) is available at www.iasusa.org.]*

Didanosine (ddI; 2',3'-dideoxyinosine) was the second drug licensed for the treatment of HIV infection, followed shortly thereafter by zalcitabine. Didanosine is metabolized to dideoxyadenosine in vivo. It is best absorbed on an empty stomach at a high pH. The toxicity profile of didanosine is quite different from that of zidovudine. The most common toxicity is a painful sensory peripheral neuropathy that occurs in ~30% of patients receiving >400 mg/d. It generally resolves with discontinuation of the drug and may not recur if the drug is resumed at a reduced dose. At higher doses than are currently used, one may see pancreatitis in ~10% of patients. Pancreatitis associated with didanosine therapy can be fatal. Didanosine should be discontinued if a patient experiences abdominal pain consistent with pancreatitis or if an elevated serum amylase or lipase level is found in association with an edematous pancreas on ultrasound. Didanosine is contraindicated in patients with a prior history of pancreatitis, regardless of etiology. A higher incidence of didanosine-associated toxicities has been seen when it is used in combination with stavudine, hydroxyurea, ribavirin, or tenofovir.

Zalcitabine (ddC; 2',3'-dideoxycytidine) is rarely used today in the management of patients with HIV infection and was discontinued from the U.S. market in 2006. Among the nucleoside analogues licensed for the treatment of HIV infection, it is probably the weakest. The main toxicities of ddC are peripheral neuropathy and pancreatitis.

Stavudine (d4T; 2',3'-didehydro-3'-deoxythymidine) was the fourth drug licensed for the treatment of HIV infection. Like zidovudine, stavudine is a thymidine analogue. These two drugs are antagonistic in vitro and in vivo and should not be given together. Stavudine has been associated with a higher incidence of mitochondrial toxicity than the other licensed nucleoside analogues. Peripheral neuropathy, lipoatrophy, lactic acidosis, and hepatic steatosis are the main toxicities of stavudine.

Lamivudine (3TC; 2',3'-dideoxy-3'-thiacytidine) is the fifth of the nucleoside analogues to be licensed in the United States. It is the negative enantiomer of a dideoxy analogue of cytidine. In actual practice, lamivudine or the closely related drug emtricitabine (see below) is a frequent element of many different combination regimens currently in use. These two drugs and the nucleotide reverse transcriptase inhibitor tenofovir (see below) also have activity against hepatitis B virus. For this reason flares of hepatitis may be seen in co-infected patients starting and/or or stopping any of these three agents due to the confounding issues of direct effects on hepatitis B, direct effects on HIV, and immune reconstitution (see above). To prevent the development of resistant strains of HIV, these drugs should never be used on their own for the treatment of hepatitis B in the patient with HIV infection. Lamivudine is available either alone or in coformulations including zidovudine and/or abacavir **(Table 189-21)**. One reason behind the excellent synergy seen between lamivudine and the other nucleoside analogues may be that strains of HIV resistant to lamivudine (M184V substitution) appear to have enhanced sensitivity to other nucleosides, and thus development of dual resistance is more difficult. In addition, there is a suggestion that 3TC-resistant strains of HIV may be less virulent and are less able to generate new mutants than are strains of HIV that are 3TC-sensitive. Lamivudine is among the best tolerated and least toxic of the nucleoside analogues.

Table 189-21 Combination Formulations of Antiretroviral Drugs

Name	Combination
Combivir	Zidovudine + lamivudine
Epzicom	Zidovudine + abacavir
Trizivir	Zidovudine + lamivudine + abacavir
Truvada	Tenofovir + emtricitabine
Atripla	Tenofovir + emtricitabine + efavirenz

Triomune[a]	Stavudine + lamivudine + nevirapine

[a]Not licensed in the United States.

Emtricitabine (FTC; 5-fluoro-1-(2R,5S)-[2-(hydroxymethyl)-1,3-oxathiolan-5-yl]cytosine) is the negative enantiomer of a thio analogue of cytidine with a fluorine in the 5 position. It is licensed for use in combination with other antiretroviral agents for treatment of HIV-1 infection in adults. Compared with lamivudine, it is similar in activity and has a longer half-life. It is available either alone or coformulated with tenofovir or tenofovir and efavirenz (Table 189-21). As with lamivudine, resistance to emtricitabine is associated with the M184V mutation in reverse transcriptase. Viruses showing the K65R mutation in reverse transcriptase may have reduced susceptibility to emtricitabine.

Abacavir {(1S,cis)-4-[2-amino-6-(cyclopropylamino)-9H-purin-9-yl]-2-cyclopentene-1-methanol sulfate (salt)(2:1)} is a synthetic carbocyclic analogue of the nucleoside guanosine. It is licensed to be used in combination with other antiretroviral agents for the treatment of HIV-1 infection. Hypersensitivity reactions that may occur with initial therapy or rechallenge have been reported in ~ 4% of patients treated with this drug, and patients developing signs or symptoms of hypersensitivity such as fever, skin rash, fatigue, and GI symptoms should discontinue the drug and not restart it. Fatal hypersensitivity reactions have been reported with rechallenge. Abacavir hypersensitivity occurs with a higher frequency in patients who are HLA-B5701-positive. It is recommended that patients be screened for HLA-B5701 prior to initiation of abacavir and that abacavir only be used as a last resort and with close monitoring in patients who are HLA-B5701-positive. Abacavir-resistant strains of HIV are typically also resistant to lamivudine, emtricitabine, didanosine, and zalcitabine. Abacavir is formulated alone as well as in combination with lamivudine or zidovudine and lamivudine.

Tenofovir disoproxil fumarate (9-[(R)-2-[[bis[[(isopropoxycarbonyl)oxy]methoxy]phosphinyl]methoxy]propyl]adenine fumarate (1:1)) is an acyclic nucleoside phosphonate diester analogue of adenosine monophosphate. It undergoes diester hydrolysis to form the nucleoside monophosphate tenofovir and is the first nucleotide analogue to be licensed for treatment of HIV infection. It is indicated in combination with other antiretroviral agents for the treatment of HIV-1 infection. HIV isolates with increased resistance typically express a K65R mutation in reverse transcriptase and a three- to fourfold reduction in sensitivity to tenofovir. Tenofovir is primarily eliminated by the kidneys, and renal impairment including a Fanconi-like syndrome with hypophosphatemia may occur. Tenofovir is contraindicated in patients with renal impairment. Coadministration with didanosine leads to a 60% increase in didanosine levels, and thus doses of didanosine need to be adjusted and patients monitored carefully if these two drugs are used in combination. In addition CD4+ T cell increases may be blunted in patients on this combination. Coadministration of tenofovir with atazanavir leads to a decrease in atazanavir levels, and thus low-dose ritonavir (see below) needs to be added when these drugs are used in combination. Tenofovir is available alone and coformulated with emtricitabine or emtricitabine and efavirenz.

Nevirapine, *delavirdine*, *efavirenz*,*etravirine*, and rilpivirine are nonnucleoside inhibitors of the HIV-1 reverse transcriptase. The first four are licensed for use in combination with nucleoside analogues for the treatment of HIV-infected adults while rilpivire is currently investigational. Coformulations that include efavirenz or nevirapine are available (Table 189-21). These agents inhibit reverse transcriptase by binding to regions of the enzyme outside the active site and causing conformational changes in the enzyme that render it inactive. Although these agents are active in the nanomolar range, they are also very selective for the reverse transcriptase of HIV-1, have no activity against HIV-2, and, when used as monotherapy, are associated with the rapid emergence of drug-resistant mutants (Table 189-20; Fig. 189-47). Efavirenz and rilpivirine are administered once a day, nevirapine and etravirine twice a day, and delavirdine three times a day. All are associated with the development of a maculopapular rash, generally seen within the first few weeks of therapy. While it is possible to treat through this rash, it is important to be sure that one is not dealing with a more severe eruption such as Stevens-Johnson syndrome by looking carefully for signs of mucosal involvement, significant fever, or painful lesions with desquamation. Severe, life-threatening, and in some cases fatal hepatotoxicity, including fulminant and cholestatic hepatitis, hepatic necrosis, and hepatic failure, have been reported in patients treated with nevirapine. There is a suggestion that this is more common in women with higher CD4+ T cell counts. Many patients treated with efavirenz note a feeling of light-headedness, dizziness, or out of sorts following the initiation of therapy. Some complain of vivid dreams. These symptoms tend to disappear after several weeks of therapy. Aside from difficulties with dreams, taking efavirenz at bedtime may minimize the side effects. Efavirenz may cause fetal harm when administered during the first trimester to a pregnant woman. Women of childbearing potential should undergo pregnancy testing prior to initiation of efavirenz. Efavirenz is commonly used in combination with two nucleoside analogues as part of initial treatment regimens. Etravirine is a diarylpyrimidine derivative currently licensed for treatment of HIV infection in combination with other agents. In contrast to the other nonnucleoside reverse transcriptase inhibitors, which all exhibit cross-resistance, etravirine may be active against strains of HIV that are resistant to other nonnucleoside reverse transcriptase inhibitors. Among its side effects are rash, headache, nausea, and diarrhea. Rilpivirine is effective across a broad range of NNRTI viruses, has comparable efficacy to efavirenz and fewer problems with skin rash or CNS disturbances.

The HIV-1 protease inhibitors (saquinavir, indinavir, ritonavir, nelfinavir, amprenavir, fosamprenavir, lopinavir/ritonavir, atazanavir, tipranavir, and darunavir) are a major part of the therapeutic armamentarium of antiretrovirals. When used as part of initial regimens in combination with reverse transcriptase inhibitors, these agents have been shown to be capable of suppressing levels of HIV replication to under 50 copies per milliliter in the majority of patients for a minimum of 5 years. As in the case of reverse transcriptase inhibitors, resistance to protease inhibitors can develop rapidly in the setting of monotherapy, and thus these agents should be used only as part of combination therapeutic regimens. A summary of known resistance mutations for protease inhibitors is shown in Fig. 189-47.

Saquinavir was the first of the HIV-1 protease inhibitors to be licensed. It is typically given with low doses of ritonavir to obtain therapeutic levels. Saquinavir is metabolized by the cytochrome P450 system in both the GI tract and the liver. Low-dose ritonavir results in inhibition of cytochrome P450. Thus, when both drugs are administered together there is an increase in saquinavir levels. This use of low doses of ritonavir to provide pharmacodynamic boosting of other agents is a common strategy in HIV therapy. Saquinavir is among the best-tolerated protease inhibitors.

Ritonavir was the first protease inhibitor for which clinical efficacy was demonstrated. In a study of 1090 patients with CD4+ T cell counts <100/μL who were randomized to receive either placebo or ritonavir in addition to any other licensed medications, patients receiving ritonavir had a reduction in the cumulative incidence of clinical progression or death from 34% to 17%. Mortality decreased from 10.1 to 5.8%. At full doses, ritonavir is poorly tolerated. Among the main side effects are nausea, diarrhea, abdominal pain, hyperlipidemia, and circumoral paresthesia. Ritonavir has a high affinity for several isoforms of cytochrome P450 (3A4, 2D6), and its use can result in large increases in the plasma concentrations of drugs metabolized by these pathways. Among the agents affected in this manner are most other protease inhibitors, macrolide antibiotics, R-warfarin, ondansetron, rifabutin, most calcium channel blockers, glucocorticoids, and some of the chemotherapeutic agents used to treat KS and/or lymphomas. In addition, ritonavir may increase the activity of glucuronyltransferases, thus decreasing the levels of drugs metabolized by this pathway. Overall, great care must be taken when prescribing additional drugs to patients taking protease inhibitors in general and ritonavir in particular. As mentioned above, the pharmacodynamic boosting property of ritonavir, seen with doses as low as 100–200 mg once or twice a day, is often used in the setting of cART for HIV infection to derive more convenient regimens. For example, when given with low-dose ritonavir, saquinavir and indinavir can be given on

twice-a-day schedules and taken with food.

Indinavir was the first protease inhibitor used in combination with dual nucleoside therapy. The combination of zidovudine, lamivudine, and indinavir was the first "triple combination" shown to have a profound effect on HIV replication. The main side effects of indinavir are nephrolithiasis (seen in 4% of patients) and asymptomatic indirect hyperbilirubinemia (seen in 10%). Indinavir is predominantly metabolized by the liver. The dose should be lowered in patients with cirrhosis. Levels of indinavir are decreased during concurrent therapy with rifabutin, efavirenz, or nevirapine and increased during concurrent therapy with ketoconazole, delavirdine, or ritonavir. Dosages should be modified appropriately in these circumstances. (Table 189-20).

Nelfinavir was approved in 1997 and *amprenavir* was approved in 1999 for the treatment of adult or pediatric HIV infection when cART is warranted. As with most of the antiretroviral agents licensed since 1999, these approvals were based on randomized, controlled trials that demonstrated decreases in plasma HIV RNA levels and increases in CD4+ T cell counts rather than clinical endpoints. Both nelfinavir and amprenavir have unique resistance profiles. Nelfinavir resistance is associated with a D30N substitution in the protease gene. Viruses harboring this single mutation retain sensitivity to other protease inhibitors. While it has been suggested that for this reason nelfinavir is a good initial protease inhibitor, enthusiasm for its use waned following 48-week clinical trials data demonstrating the virologic inferiority of nelfinavir to lopinavir/ritonavir, fosamprenavir, and efavirenz. Protease inhibitor resistance typically involves multiple amino acid substitutions and reduced susceptibility across the class. Amprenavir resistance is associated with a unique substitution at amino acid 50 (I50V). Nelfinavir and amprenavir are both associated with GI side effects. About 1% of patients receiving amprenavir have experienced severe and life-threatening skin reactions. An additional disadvantage of amprenavir is that the original formulation requires the patient to take 8 large capsules twice a day. Amprenavir has largely been replaced by fosamprenavir (see below).

Fosamprenavir was licensed in 2003 for the treatment of HIV infection in combination with other antiretroviral agents in adults. It is a prodrug of amprenavir that is rapidly converted to amprenavir by cellular phosphatases. It is supplied as a 700-mg tablet. The recommended dosage regimens are as follows: 1400 mg twice a day; 700 mg twice a day with ritonavir, 100 mg twice a day; or 1400 mg once a day with ritonavir, 200 mg once a day. As noted above, ritonavir-boosted fosamprenavir has been shown to have efficacy comparable to lopinavir/ritonavir and to efavirenz in combination regimens.

Lopinavir/ritonavir (Kaletra) is a fixed-dose combination of the protease inhibitors lopinavir (200 mg) and ritonavir (50 mg). It was licensed in 2000 for treatment of HIV-1 infection in adults and children in combination with other agents. A main advantage of this pill is that it combines the pharmacologic enhancement of low-dose ritonavir with a second protease inhibitor in a single capsule. In a randomized, controlled trial, this combination capsule was found to be superior to nelfinavir. Its main complications are GI upset and hyperlipidemia.

Atazanavir is an azapeptide inhibitor of the HIV-1 protease that was licensed in 2003. An advantage of atazanavir is that total cholesterol and triglyceride levels do not increase as much with atazanavir as with other protease inhibitors. This coupled with the fact that it can be given on a once-daily schedule has made atazanavir a popular component of initial treatment regimens. Atazanavir is associated with increases in serum bilirubin and prolongations of the ECG PR interval. Atazanavir-resistant isolates emerging in previously treatment-naïve individuals frequently harbor an I50L substitution. This mutation in some instances is associated with increased sensitivity to other protease inhibitors. Atazanavir requires an acidic gastric pH for absorption, and its use in combination with a proton pump inhibitor is contraindicated due to concerns about absorption. Atazanavir is an inhibitor of cytochrome P3A, and its use may be associated with increased levels of calcium channel blockers, macrolide antibiotics, HMG-CoA reductase inhibitors, and sildenafil. Levels of atazanavir are lower in the presence of tenofovir or efavirenz. In these settings, levels of atazanavir should be boosted with the use of low-dose ritonavir.

Tipranavir is a nonpeptidic HIVprotease inhibitor licensed in 2005. It is licensed for use in combination with 200 mg ritonavir and is indicated for cART of HIV-1 infection in treatment-experienced adults or in adults with evidence of HIV-1 strains resistant to multiple protease inhibitors. Tipranavir was found to be inferior to lopinavir/ritonavir in a randomized controlled trial in naïve patients. In that study, at lower doses it was virologically inferior, while at higher doses it exhibited a greater degree of hepatotoxicity. The main side effects of tipranavir are GI intolerance and skin rash; the latter is seen in ~10% of patients and may be related to the sulfonamide moiety in the molecule. Tipranavir coadministered with ritonavir has also been associated with reports of intracranial hemorrhage as well as reports of clinical hepatitis and hepatic decompensation, including some fatalities in both settings. The risk of hepatotoxicity is increased in patients with hepatitis B or C co-infection.

Darunavir is a nonpeptidic HIVprotease inhibitor initially licensed in 2006. It is indicated for coadministration with 100 mg of ritonavir and other antiretroviral agents for the treatment of HIV infection. In initial studies in treatment-experienced subjects, 46% of patients achieved a reduction in HIV RNA viral loads to <50 copies per milliliter. Studies in treatment-naïve patients demonstrated efficacy comparable to lopinavir/ritonavir-containing regimens. Skin rash, which may be severe, is seen in 7% of patients and may be related to the sulfonamide moiety contained in the molecule. GI intolerance and headache are the other most frequent side effects.

Entry inhibitors act by interfering with the binding of HIV to its receptor or co-receptor or by interfering with the process of fusion (see above). The first drug in this class to be licensed was the fusion inhibitor *enfuvirtide*, or T-20, followed by the CCR5 antagonist *maraviroc*. A variety of additional small molecules that bind to HIV-1 co-receptors are currently in clinical trials.

Enfuvirtide is a linear 36-amino-acid synthetic peptide with the *N* terminus acetylated and the *C* terminus a carboxamide. It is composed of naturally occurring l-amino acid residues and interferes with the fusion of the viral and cellular membranes by binding to the HR1 region in the gp41 subunit of the HIV-1 envelope. This binding interferes with the coil-coil interaction required to approximate the viral envelope and the host cell membrane during the process of viral fusion. Enfuvirtide was licensed in 2003 for treatment of HIV-1 infection in combination with other antiretroviral agents in treatment-experienced patients with ongoing viral replication despite antiretroviral therapy. Enfuvirtide is not active against HIV-2. Enfuvirtide resistant isolates of HIV exhibit amino acid changes in positions 36–45 of gp41. In two independent studies, patients who had persistent viremia despite prior treatment with agents from all three available classes of drugs were randomized to receive an individualized regimen (based on prior treatment history and resistance profile) with or without enfuvirtide. The change in plasma HIV-1 RNA from baseline was ~1 log greater (–1.53 vs. –0.68) in patients randomized to receive enfuvirtide. Among the drawbacks of this agent are the requirement for twice-a-day injection, the occurrence of injection site reactions in close to 100% of patients, and an increase in bacterial pneumonia in the enfuvirtide-treated patients compared with the control patients (4.68 vs 0.61 events per 100 patient-years) in the phase III studies.

Maraviroc is a CCR5 antagonist that interferes with HIV binding at the stage of co-receptor engagement. It was licensed in 2007 for treatment of HIV infection in combination with other agents in treatment-experienced patients infected with only CCR5-tropic (R5) virus resistant to multiple agents. The license was extended in 2009 to include treatment-naïve patients with R5 virus. A co-receptor tropism assay should be performed if one is considering the use of maraviroc to ensure that the potential patient is harboring R5 virus. In phase III trials of treatment-experienced patients randomized to receive optimal therapy plus maraviroc or placebo, 61% of patients randomized to maraviroc achieved HIV RNA levels <400 copies/mL compared with 28% of patients randomized to placebo. An allergic reaction–associated hepatotoxicity has been reported with maraviroc. Among the most common side effects of maraviroc are dizziness due to postural hypotension, cough, fever, colds, rash, muscle and joint pain, and stomach pain. Maraviroc is a substrate of CYP3A and Pgp, and the recommend dose varies depending on concomitant

medications. In combination with nucleoside analogues, tipranavir/ritonavir, enfuvirtide, and/or nevirapine, the dose is 300 mg twice daily. In the presence of CYP3A inhibitors, such as most protease inhibitors, the dose is 150 mg twice daily. In the presence of CYP3A inducers such as efavirenz, the dose is 600 mg twice daily.

The newest class of antiretroviral compounds is the *integrase inhibitors*. *Raltegravir* is an inhibitor of the viral enzyme integrase and the first of this class to be approved. Elvitegravir is currently in clinical trials. Raltegravir was approved in 2007 for treatment of HIV infection in combination with other agents in treatment-experienced patients, and the approval was extended in 2009 to include treatment-naïve patients. Raltegravir exhibits a wide range of activity against HIV-1 and HIV-2, including viruses with multiple resistance mutations to other classes of drugs. As with several other compounds, resistance to raltegravir comes at the expense of replicative fitness. In two phase III studies in which 436 patients with 3-class antiretroviral drug resistance were randomized to an optimized background regimen with raltegravir or placebo, 76% of patients receiving raltegravir achieved HIV RNA levels <400 copies/mL compared with 41% of patients randomized to the placebo arm. In contrast to many other antiretroviral drugs the side-effect profile of raltegravir is minimal, with similar side-effect profiles noted for the raltegravir and placebo groups.

PRINCIPLES OF THERAPY

The principles of therapy for HIV infection have been articulated by a panel sponsored by the U.S. Department of Health and Human Services as a working group of the NIH Office of AIDS Research Advisory Council. These principles are summarized in **Table 189-22**. As noted in these guidelines, cART of HIV infection does not lead to eradication or cure of HIV. Treatment decisions must take into account the fact that one is dealing with a chronic infection. While early therapy is generally the rule in infectious diseases, immediate treatment of every HIV-infected individual upon diagnosis may not be prudent, and therapeutic decisions must take into account the balance between risks and benefits. Patients initiating antiretroviral therapy must be willing to commit to life-long treatment and understand the importance of adherence to their prescribed regimen. The importance of adherence is illustrated by the observation that treatment interruption is associated with rapid increases in HIV RNA levels, rapid declines in CD4+ T cell counts, and an increased risk of clinical progression. While it seems reasonable to assume that the complications associated with cART could be minimized by regimens designed to minimize exposure to the drugs in question, all efforts to do so have paradoxically been associated with an increase in serious adverse events in the patients randomized to intermittent therapy, suggesting that some "non-AIDS-associated" serious adverse events such as heart attack and stroke may be linked to HIV replication. Thus, unless contraindicated for reasons of toxicity, patients started on cART should remain on cART.

Table 189-22 Principles of Therapy of HIV Infection

1. Ongoing HIV replication leads to immune system damage and progression to AIDS.
2. Plasma HIV RNA levels indicate the magnitude of HIV replication and the rate of CD4+ T cell destruction. CD4+ T cell counts indicate the current level of competence of the immune system.
3. Rates of disease progression differ among individuals, and treatment decisions should be individualized based on plasma HIV RNA levels and CD4+ T cell counts.
4. Maximal suppression of viral replication is a goal of therapy; the greater the suppression the less likely the appearance of drug-resistant quasispecies.
5. The most effective therapeutic strategies involve the simultaneous initiation of combinations of effective anti-HIV drugs with which the patient has not been previously treated and that are not cross-resistant with antiretroviral agents that the patient has already received.
6. The antiretroviral drugs used in combination regimens should be used according to optimum schedules and dosages.
7. The number of available drugs is limited. Any decisions on antiretroviral therapy have a long-term impact on future options for the patient.
8. Women should receive optimal antiretroviral therapy regardless of pregnancy status.
9. The same principles apply to children and adults. The treatment of HIV-infected children involves unique pharmacologic, virologic, and immunologic considerations.
10. Compliance is an important part of ensuring maximal effect from a given regimen. The simpler the regimen, the easier it is for the patient to be compliant.

Source: Modified from *Principles of Therapy of HIV Infection* , USPHS, and the Henry J. Kaiser Family Foundation.

At present, a reasonable course of action is to initiate cART in anyone with the acute HIV syndrome; all pregnant women; patients with an AIDS-defining illness; patients with HIV-associated nephropathy; patients with hepatitis B infection when treatment for hepatitis B is indicated, and patients with asymptomatic disease with CD4+ T cell counts <500/μL **(Table 189-23)**. Clinical trials are underway to determine the value of even earlier intervention, and some experts would place everyone with HIV infection on antiretroviral therapy. In addition, one may wish to administer a 6-week course of therapy to uninfected individuals immediately following a high-risk exposure to HIV. studies are underway to define the role of pre-exposure prophylaxis. For patients diagnosed with an opportunistic infection and HIV infection at the same time, one may consider a 2- to 4-week delay in the initiation of antiretroviral therapy during which time treatment is focused on the opportunistic infection. While not proven, it is postulated that this delay may decrease the severity of any subsequent immune reconstitution inflammatory syndrome by lowering the antigenic burden of the opportunistic infection.

Table 189-23 Indications for the Initiation of Antiretroviral Therapy in Patients with HIV Infection

I. Acute infection syndrome
II. Chronic infection
A. Symptomatic disease (including HIV-associated nephropathy)
B. Asymptomatic disease
1. CD4+ T cell count <500/μL[a]
2. Pregnancy
III. Postexposure prophylaxis

[a]This is an area of controversy. Some experts would treat everyone regardless of CD4+ T cell count.

Source: *Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents,* USPHS.

Once the decision has been made to initiate therapy, the health care provider must decide which drugs to use as the first regimen. The decision regarding choice of drugs not only will affect the immediate response to therapy but also will have implications regarding options for future therapeutic regimens. The initial regimen is usually the most effective insofar as the virus has yet to develop significant resistance. Given that patients can be infected with viruses that harbor drug resistance mutations, it is recommended that a viral genotype be done prior to the initiation of therapy to optimize the selection of antiretroviral agents. The three options for initial therapy most commonly in use today are three different three-drug regimens. The first regimen utilizes a nucleotide and a nucleoside analogue (tenofovir and emtricitabine) and a nonnucleoside reverse transcriptase inhibitor (efavirenz). The second regimen utilizes a ritonavir-boosted protease inhibitor (atazanavir or darunavir) in place of the nonnucleoside reverse transcriptase inhibitor. The third regimen utilizes an integrase inhibitor (raltegravir) in place of the nonnucleoside reverse transcriptase inhibitor. Unfortunately there are no clear data at present on which to base distinctions between these three approaches. Following the initiation of therapy one should expect a rapid, at least 1-log (tenfold) reduction in plasma HIV RNA levels within 1–2 months and then a slower decline in plasma HIV RNA levels to <50 copies per milliliter within 6 months. During this same time there should be a rise in the CD4+ T cell count of 100–150/μL that is also particularly brisk during the first month of therapy. Subsequently, one should anticipate a CD4+ T cell count increase of 50–100 cells/year until numbers approach normal. Many clinicians feel that failure to achieve these endpoints is an indication for a change in therapy. Other reasons for a change in therapy include a persistently declining CD4+ T cell count, a consistent increase in HIV RNA levels to >1000 copies/mL, clinical deterioration, or drug toxicity **(Table 189-24)**. As in the case of initiating therapy, changing therapy may have a lasting impact on future therapeutic options. When changing therapy because of treatment failure (clinical progression or worsening laboratory parameters), it is important to attempt to provide a regimen with at least two new active drugs. This decision can be guided by resistance testing (see below). In the patient in whom a change is made for reasons of drug toxicity, a simple replacement of one drug is reasonable. It should be stressed that in attempting to sort out a drug toxicity it may be advisable to hold all therapy for a period of time to distinguish between drug toxicity and disease progression. Drug toxicity will usually begin to show signs of reversal within 1–2 weeks. Prior to changing a treatment regimen because of drug failure, it is important to ensure that the patient has been adherent to the prescribed regimen. As in the case of initial therapy, the simpler the new therapeutic regimen, the easier it is for the patient to be compliant. Plasma HIV RNA levels and CD4+ T lymphocyte counts should be monitored every 3–6 months during therapy and more frequently if one is contemplating a change in regimen or immediately following a change in regimen.

Table 189-24 Indications for Changing Antiretroviral Therapy in Patients with HIV Infection[a]

Less than a 1-log drop in plasma HIV RNA by 4 weeks following the initiation of therapy
A reproducible significant increase (defined as threefold or greater) from the nadir of plasma HIV RNA level not attributable to intercurrent infection, vaccination, or test methodology
Persistently declining CD4+ T cell numbers
Clinical deterioration
Side effects

[a]Generally speaking, a change should involve the initiation of at least two drugs felt to be effective in the given patient. The exception to this is when change is being made to manage toxicity, in which case a single substitution is reasonable.

Source: *Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents* , USPHS.

In an attempt to determine an optimal therapeutic regimen for initial therapy or for a patient on a failing regimen, one may attempt to measure antiretroviral drug susceptibility through genotyping or phenotyping of HIV quasispecies and to determine adequacy of dosing through measurement of drug levels. Genotyping may be done through dideoxynucleotide sequencing, DNA chip hybridization, or line probe assays. Phenotypic assays typically measure the enzymatic activity of viral enzymes in the presence or absence of different concentrations of different drugs and have also been used to determine co-receptor tropism. These assays will generally detect quasispecies present at a frequency of ≥10%. It is generally recommended that resistance testing be used in selecting initial therapy in settings where the risk of transmission of resistant virus is high (such as the United States and Europe) and in determining new regimens for patients experiencing virologic failure while on therapy. Resistance testing may be of particular value in distinguishing drug-resistant virus from poor patient compliance. Due to the rapid rate at which drug-resistant viruses revert to wild-type, it is recommended that resistance testing performed in the setting of drug failure be carried out while the patient is still on the failing regimen. Measurement of plasma drug levels can also be used to tailor an individual treatment. The inhibitory quotient, defined as the trough blood level/IC_{50} of the patient's virus, is used by some to determine the adequacy of dosing of a given treatment regimen. Despite the best of efforts there will still be patients with ongoing high levels of HIV replication while receiving the best available therapy. These patients will receive benefit from remaining on antiretroviral therapy even though it is not fully suppressive.

In addition to the licensed medications discussed above, a large number of experimental agents are being evaluated as possible therapies for HIV infection. Therapeutic strategies are being developed that interfere with virtually every step of the replication cycle of the virus (Fig. 189-3). In addition, as more is discovered about the role of the immune system in controlling viral replication, additional strategies, generically referred to as "immune-based therapies," are being developed as a complement to antiviral therapy. Among the antiviral agents in early clinical trials are additional nucleoside and nucleotide analogues, protease inhibitors, fusion inhibitors, receptor and co-receptor antagonists, and integrase inhibitors as well as new antiviral strategies including antisense nucleic acids and maturation inhibitors. Among the immune-based therapies being evaluated are IFN-α, bone marrow transplantation, adoptive transfer of lymphocytes genetically modified to resist infection or enhance HIV-specific immunity, active immunotherapy with inactivated HIV or its components, IL-7, and IL-15.

HIV AND THE HEALTH CARE WORKER

Health care workers, especially those who deal with large numbers of HIV-infected patients, have a small but definite risk of becoming infected with HIV as a result of professional activities (see "Occupational Transmission of HIV: Health Care Workers, Laboratory Workers, and the Health Care Setting," above). The first case of HIV transmission from a patient to health care worker was reported in 1984. Occupational transmission of HIV has been reported in most countries, but the global surveillance data required to estimate the true frequency of this problem are not available.

In the United States between 1981 and 2006, 57 health care workers for whom case investigations were completed had documented seroconversions to HIV following occupational exposures. The routes of exposure resulting in infection were as follows: 48 percutaneous (puncture/cut injury); 5 mucocutaneous (mucous membrane and/or skin); 2 both percutaneous and mucocutaneous; and 2 of unknown route. Of the 57 health care personnel, 49 were exposed to HIV-infected blood; 3 to concentrated virus in a laboratory; 1 to visibly bloody fluid; and 4 to an unspecified fluid. The individuals with documented seroconversions included 19 laboratory workers (16 of whom were clinical laboratory workers), 24 nurses, 6 physicians, 2 surgical technicians, 1 dialysis technician, 1 respiratory therapist, 1 health aide, 1 embalmer/morgue technician, and 2

housekeeper/maintenance workers. In addition, at least 140 possible cases of occupationally acquired HIV infection have been reported among health care personnel in the United States. The number of these workers who actually acquired their infection through occupational exposures is not known. Taken together, data from several large studies suggest that the risk of HIV infection following a percutaneous exposure to HIV-contaminated blood is ~0.3%, and after a mucous membrane exposure, ~ 0.09%. Although episodes of HIV transmission after nonintact skin exposure have been documented, the average risk for transmission by this route has not been precisely quantified but is estimated to be less than the risk for mucous membrane exposures. The risk for transmission after exposure to fluids or tissues other than HIV-infected blood also has not been quantified but is probably considerably lower than for blood exposures. A seroprevalence survey of 3420 orthopedic surgeons, 75% of whom practiced in an area with a relatively high prevalence of HIV infection and 39% of whom reported percutaneous exposure to patient blood, usually through an accident involving a suture needle, failed to reveal any cases of possible occupational infection, suggesting that the risk of infection with a suture needle may be considerably less than that with a blood-drawing (hollow-bore) needle.

Most cases of health care worker seroconversion occur as a result of needle-stick injuries. When one considers the circumstances that result in needle-stick injuries, it is immediately obvious that adhering to the standard guidelines for dealing with sharp objects would result in a significant decrease in this type of accident. In one study, 27% of needle-stick injuries resulted from improper disposal of the needle (over half of these were due to recapping the needle), 23% occurred during attempts to start an IV line, 22% occurred during blood drawing, 16% were associated with an IM or SC injection, and 12% were associated with giving an IV infusion.

Clinicians should consider potential occupational exposures to HIV as urgent medical concerns to ensure timely postexposure management and possible administration of postexposure antiretroviral prophylaxis (PEP). Recommendations regarding PEP must take into account that a variety of circumstances determine the risk of transmission of HIV following occupational exposure. In this regard, several factors have been associated with an increased risk for occupational transmission of HIV infection, including deep injury, the presence of visible blood on the instrument causing the exposure, injury with a device that had been placed in the vein or artery of the source patient, terminal illness in the source patient, and lack of postexposure cART in the exposed health care worker. Other important considerations when considering PEP in the health care worker include known or suspected pregnancy or breast-feeding, the possibility of exposure to drug-resistant virus, and toxicities of PEP regimens. Regardless of the decision to use PEP, the wound should be cleansed immediately and antiseptic applied. If a decision is made to offer PEP, U.S. Public Health Service guidelines recommend (1) a combination of two nucleoside analogue reverse transcriptase inhibitors given for 4 weeks for less severe exposures, or (2) a combination of two nucleoside analogue reverse transcriptase inhibitors plus a third drug given for 4 weeks for more severe exposures. Most clinicians administer the latter regimen in all cases in which a decision is made to treat. Detailed guidelines are available from the *Updated U.S. Public Health Service Guidelines for the Management of Occupational Exposures to HIV and Recommendations for Postexposure Prophylaxis* (CDC, 2005). The report emphasizes the importance of adherence to PEP when it is indicated, follow-up of exposed workers to improve PEP adherences, monitoring for adverse events (including seroconversion), and expert consultation in the management of exposures.

For consultation on the treatment of occupational exposures to HIV and other bloodborne pathogens, the clinician managing the exposed patient can call the National Clinicians' Post-Exposure Prophylaxis Hotline (PEPline) at 888-448-4911. This service is available 24 hours a day at no charge. (Additional information on the Internet is available at *www.nccc.ucsf.edu*.) PEPline support may be especially useful in challenging situations, such as when drug-resistant HIV strains are suspected or the health care worker is pregnant.

Health care workers can minimize their risk of occupational HIV infection by following the CDC guidelines of July 1991, which include adherence to universal precautions, refraining from direct patient care if one has exudative lesions or weeping dermatitis, and disinfecting and sterilizing reusable devices employed in invasive procedures. The premise of universal precautions is that every specimen should be handled as if it came from someone infected with a bloodborne pathogen. All samples should be double-bagged, gloves should be worn when drawing blood, and spills should be immediately disinfected with bleach.

In attempting to put this small but definite risk to the health care worker in perspective, it is important to point out that ~200 health care workers die each year as a result of occupationally acquired hepatitis B infection. The tragedy in this instance is that these infections and deaths due to HBV could be greatly decreased by more extended use of the HBV vaccine. The risk of HBV infection following a needle-stick injury from a hepatitis antigen–positive patient is much higher than the risk of HIV infection (see "Transmission," above). There are multiple examples of needle-stick injuries where the patient was positive for both HBV and HIV and the health care worker became infected only with HBV. For these reasons, it is advisable, given the high prevalence of HBV infection in HIV-infected individuals, that all health care workers dealing with HIV-infected patients be immunized with the HBV vaccine.

TB is another infection common to HIV-infected patients that can be transmitted to the health care worker. For this reason, all health care workers should know their PPD status, have it checked yearly, and receive 6 months of isoniazid treatment if their skin test converts to positive. In addition, all patients in whom a diagnosis of TB is being entertained should be placed immediately in respiratory isolation, pending results of the diagnostic evaluation. The emergence of drug-resistant organisms, including the extensively drug-resistant TB strains that have been identified in Africa, has made TB an increasing problem for health care workers. This is particularly true for the health care worker with preexisting HIV infection.

One of the most charged issues ever to come between health care workers and patients is that of transmission of infection from HIV-infected health care workers to their patients. This is discussed under "Occupational Transmission of HIV: Health Care Workers, Laboratory Workers, and the Health Care Setting," above. Theoretically, the same universal precautions that are used to protect the health care worker from the HIV-infected patient will also protect the patient from the HIV-infected health care worker.

A PREVENTIVE VACCINE AGAINST HIV INFECTION

Given that human behavior, especially human sexual behavior, is extremely difficult to change, a critical modality for preventing the spread of HIV infection is the development of a safe and effective vaccine. Historically, vaccines have provided a safe, cost-effective, and efficient means of preventing illness, disability, and death from infectious diseases. Successful vaccines for the most part are predicated on the assumptions that the body can mount an adequate immune response to the microbe or virus in question during natural infection and that the vaccine will mimic the natural response to infection. Even with serious diseases, such as smallpox, poliomyelitis, measles, and influenza among others, the body in the vast majority of cases clears the infectious agent and provides protection, which is usually life-long against future exposure. Unfortunately, this is not the case with HIV infection since the natural immune response to HIV infection is unable to clear the virus from the body and cases of superinfection have been reported. Some of the factors that contribute to the problematic nature of development of a preventive HIV vaccine are the high mutability of the virus, the fact that the infection can be transmitted by cell-free or cell-associated virus, the fact that the HIV provirus integrates itself into the genome of the target cell and may remain in a latent form unexposed to the immune system, the likely need for the development of effective mucosal immunity, and the fact that it has been difficult to establish the precise correlates of protective immunity to HIV infection. Some HIV-infected individuals are long-term

nonprogressors, and a number of individuals have been exposed to HIV multiple times but remain uninfected; these facts suggest that there are elements of host defense or an HIV-specific immune response that have the potential to be protective. Early attempts to develop a vaccine with the envelope protein gp120 aimed at inducing neutralizing antibodies in humans were performed based on the induction of neutralizing antibodies in nonhuman primates. The significance of the laboratory assays were unknown at the time, and the elicited antisera failed to neutralize primary isolates of HIV cultured and tested in fresh peripheral blood mononuclear cells. In this regard, two phase 3 trials were undertaken in the United States and Thailand using soluble gp120, and the vaccines failed to protect human volunteers from HIV infection. A number of studies in monkeys using vaccines that induce predominantly cellular (T cell) immune responses have not protected the animals against infection but have lowered the initial burst of viremia following acute infection as well as decreased temporarily the viral set point. Since most sexually transmitted HIV infections occur when the transmitting partner is experiencing high levels of viremia such as during the acute phase of HIV infection or during the advanced stage of disease when the viral load is high, such a vaccine, which might limit the initial burst of viremia in primary infection and decrease the established viral set point, could have benefits for the individual as well as for their sexual partners. However, such a "T cell vaccine" failed in human clinical trials to lower either the initial burst of viremia or the viral set point after acquisition of infection. Recently, a vaccine using a poxvirus vector prime expressing various viral proteins followed by an envelope protein boost was tested in a 16,000-person clinical trial conducted in Thailand among predominantly low-prevalence heterosexuals. The vaccine provided the first positive, albeit very modest, signal ever reported in an HIV vaccine trial, showing 31% protection against acquisition of infection. Such a result is certainly not sufficient justification for clinical use of the vaccine, but it serves as an important first step in the direction of the development of a safe and effective vaccine against HIV infection. The next important step is to determine the correlate or correlates of immunity that provided the modest protection against infection.

PREVENTION

Education, counseling, and behavior modification are the cornerstones of an HIV prevention strategy. A major problem in the United States and elsewhere is that many infections are passed on by those who do not know that they are infected. Of the ~1.1 million persons in the United States who are HIV-infected, it is estimated that ~21% do not know their HIV status and thus may be putting others at risk by their own behavior. In this regard, the CDC has recently recommended that HIV testing become part of routine medical care and that all individuals between the ages of 13 and 64 years be informed of the testing and be tested without the need for written informed consent. The individual could "opt out" of testing, but if not, testing would be routinely administered. In addition to identifying individuals who might benefit from cART, information gathered from such an approach should serve as the basis for behavior-modification programs, both for infected individuals who may be unaware of their HIV status and who could infect others and for uninfected individuals practicing high-risk behavior. The practice of "safer sex" is the most effective way for sexually active uninfected individuals to avoid contracting HIV infection and for infected individuals to avoid spreading infection. Abstinence from sexual relations is the only absolute way to prevent sexual transmission of HIV infection. However, for many individuals this may not be feasible, and there are a number of relatively safe practices that can markedly decrease the chances of transmission of HIV infection. Partners engaged in monogamous sexual relationships who wish to be assured of safety should both be tested for HIV antibody. If both are negative, it must be understood that any divergence from monogamy puts both partners at risk; open discussion of the importance of honesty in such relationships should be encouraged. When the HIV status of either partner is not known, or when one partner is positive, there are a number of options. Use of condoms can markedly decrease the chance of HIV transmission. It should be remembered that condoms are not 100% effective in preventing transmission of HIV infection, and there is a ~10% failure rate of condoms used for contraceptive purposes. Most condom failures result from breakage or improper usage, such as not wearing the condom for the entire period of intercourse. Latex condoms are preferable, since virus has been shown to leak through natural skin condoms. Petroleum-based gels should never be used for lubrication of the condom, since they increase the likelihood of condom rupture. Some men who have sex with men practice fellatio as a "minimal risk" activity compared to anal intercourse. It should be emphasized that receptive fellatio is definitely not safe sex, and although the incidence of transmission via fellatio is considerably less than that of rectal or vaginal intercourse, there has been documentation of transmission of HIV where receptive fellatio was the only sexual act performed (see "Transmission," above). Topical microbicides for vaginal and anal use are being pursued actively as a means by which individuals could avoid infection when the insertive partner cannot be relied on to use a condom. In 2010, a topical microbicide composed of 1% tenofovir in a gel was demonstrated in a clinical trial in South Africa to be 39% effective in preventing acquisition of HIV infection in women engaging in vaginal intercourse. Three clinical trials of heterosexual men in South Africa, Uganda, and Kenya have shown that adult male circumcision results in a 50% to 65% reduction in HIV acquisition in the circumcised subject. Clearly, this approach has considerable potential as a preventive strategy for HIV infection and is currently being pursued, particularly in developing nations, as a component of HIV prevention. In 2010, a study of pre-exposure prophylaxis using two drugs (tenofovir plus emtricitabine; see "Treatment", above) on a daily basis in uninfected men who have sex with men and transgender women demonstrated a 44% efficacy. When participants had a high level of adherence to the regimen, the level of protection rose to 73%. Also in 2010, a study demonstrated that treatment of the infected partner in an HIV-discordant relationship with antiretroviral therapy provided a 92% protection to the uninfected partner.

The most effective way to prevent transmission of HIV infection among IDUs is to stop the use of injectable drugs. Unfortunately, that is extremely difficult to accomplish unless the individual enters a treatment program. For those who will not or cannot participate in a drug treatment program and who will continue to inject drugs, the avoidance of sharing of needles and other paraphernalia ("works") is the next best way to avoid transmission of infection. However, the cultural and social factors that contribute to the sharing of paraphernalia are complex and difficult to overcome. In addition, needles and syringes may be in short supply. Under these circumstances, paraphernalia should be cleaned after each usage with a virucidal solution, such as undiluted sodium hypochlorite (household bleach). Data from a number of studies have indicated that programs that provide sterile needles to addicts in exchange for used needles have resulted in a decrease in HIV transmission without increasing the use of injection drugs. It is important for IDUs to be tested for HIV infection and counseled to avoid transmission to their sexual partners. Secondary and tertiary spread of HIV infection by the heterosexual route within settings of a high level of injection drug use has increased greatly in the United States, particularly among African Americans. Studies are underway to determine the safety and efficacy of preexposure administration of antiretroviral drugs for the prevention of HIV infection.

Transmission of HIV via transfused blood or blood products has been decreased dramatically by a combination of screening of all blood donors for HIV infection by assays for both HIV antibody and nucleic acid and self-deferral of individuals at risk for HIV infection. In addition, clotting factor concentrates are heat-treated, essentially eliminating the risk to hemophiliacs who require these products. Autologous transfusions are preferable to transfusions from another individual. However, logistic constraints as well as the unpredictability of the need for most transfusions limit the feasibility of this approach. At present the risk of becoming HIV-infected from a contaminated blood transfusion is approximately 1 in 1.5 million donations in the United States.

Treatment of an HIV-infected mother with antiretroviral therapy during pregnancy and the infant during the first weeks following birth has proved very effective in dramatically decreasing mother-to-child transmission of HIV. In situations such as that seen in certain developing countries where pregnant women frequently present to a health care system during labor, administration of a short course (as little as a single dose of one drug) of antiretroviral therapy to the mother during labor and to the infant within 48 h of birth has also been successful in decreasing the incidence of mother-to-child transmission of HIV.

HIV can be transmitted via breast milk and colostrum. The avoidance of breast-feeding may not be practical in developing

countries, where nutritional concerns override the risk of HIV transmission. However, it is becoming appreciated that 5–15% of infants who were born of HIV-infected mothers and who were fortunate enough not to have been infected intrapartum or peripartum become infected via breast-feeding. Therefore, in developing countries, breast-feeding from an infected mother should be avoided if at all possible. Unfortunately, this is rarely the case, and given the disadvantages of withholding breast-feeding in developing countries, health authorities in most developing countries continue to recommend breast-feeding despite the potential for HIV transmission. In this regard, the most effective way to avoid mother-to-child transmission of HIV is to treat the infected mother throughout the entire pregnancy and to continue therapy during breast-feeding and beyond if the mother's clinical status warrants such treatment. Such an approach has become more feasible over the past few years as the availability of antiretroviral therapy in the developing world has increased as a result of various programs such as the President's Emergency Plan for AIDS Relief (PEPFAR) and the Global Fund to Fight AIDS, Tuberculosis and Malaria, among others. In developed countries such as the United States, where bottled formula and milk are readily accessible, breast-feeding is contraindicated when a mother is HIV-positive, even if she is receiving antiretroviral therapy.

It is clear that in order to control and ultimately end the AIDS pandemic, effective prevention is essential. There are a number of HIV preventive modalities that have been demonstrated to be effective in various target populations if properly implemented and adhered to, and several others that are showing promise in clinical trials. It is unlikely that major successes in HIV prevention will be achieved with a uni-dimensional approach. What will almost certainly be required are various versions of combination prevention strategies, depending on the target population.

FURTHER READINGS

Abdool Karim Q et al: Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. Science 329:1168, 2010

The antiretroviral therapy cohort collaboration: Life expectancy of individuals on combination antiretroviral therapy in high-income countries: A collaborative analysis of 14 cohort studies. Lancet 372:293, 2008

Arthos J et al: HIV-1 envelope protein binds to and signals through integrin α4β7, the gut mucosal homing receptor for peripheral T cells. Nat Immunol 9:301, 2008[PMID: 18264102]

Attia S et al: Sexual transmission of HIV according to viral load and antiretroviral therapy: Systematic review and meta-analysis. AIDS 23:1397, 2009[PMID: 19381076]

Baggaley RF et al: HIV transmission risk through anal intercourse: Systematic review, meta-analysis and implications for HIV prevention. Int J Epidemiol. 39:1048, 2010[PMID: 20406794]

——— et al: Systematic review of orogenital HIV-1 transmission probabilities. Int J Epidemiol 37:1255, 2008

Bailey RC et al: Male circumcision for HIV prevention in young men in Casuum, Kenya: A randomised controlled trial. Lancet 369:643, 2007[PMID: 17321310]

Boily MC et al: Heterosexual risk of HIV-1 infection per sexual act: Systematic review and meta-analysis of observational studies. Lancet Infect Dis. 9:118, 2009[PMID: 19179227]

Branson BM et al: Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR Recomm Rep 55:1, 2006[PMID: 16988643]

Brenchley JM et al: Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nat Med 12:1365, 2006[PMID: 17115046]

Brenner BG et al: High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis 195:951, 2007[PMID: 17330784]

Buchacz K et al: AIDS-defining opportunistic illnesses in US patients, 1994–2007: A cohort study. AIDS 24:1549, 2010[PMID: 20502317]

Carbone A et al: HIV-associated lymphomas and gamma-herpesviruses. Blood 113:1213, 2009[PMID: 18955561]

Centers for disease control and prevention: HIV transmission through transfusion—Missouri and Colorado, 2008. MMWR Recomm Rep 59:1335, 2010

———: HIV/AIDS Surveillance Report, 2009, 2011. Available at *www.cdc.gov/hiv/topics/surveillance/resources/reports/*

———: HIV prevalence estimates—United States, 2006. MMWR Recomm Rep 57:1073, 2008

———, Health resources and services administration, national institutes of health, hiv medicine association of the infectious diseases society of america: Incorporating HIV prevention into the medical care of persons living with HIV. MMWR Recomm Rep 52:1, 2003

Chohan B et al: Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. J Virol 79:6528, 2005[PMID: 15858037]

Chun TW et al: Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. AIDS 24:2803, 2010[PMID: 20962613]

Clerici M: Beyond IL-17: New cytokines in the pathogenesis of HIV infection. Curr Opin HIV AIDS 5:184, 2010[PMID: 20543598]

Corey L: Synergistic copathogens—HIV-1 and HSV-2. N Engl J Med 356:854, 2007[PMID: 17314346]

Das M et al: Decreases in community viral load are accompanied by reductions in new HIV infections in San Francisco. PLoS One 5:e11068, 2010

Department of health and human services panel on antiretroviral guidelines for adults and adolescents: *Guidelines for the Use*

of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents, January, 2011. Updates available at *www.aidsinfo.nih.gov*

Derdeyn CA et al: Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science 303:2019, 2004[PMID: 15044802]

Donnell D et al: Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: A prospective cohort analysis. Lancet 375:2092, 2010[PMID: 20537376]

Dorak MT et al: Transmission of HIV-1 and HLA-B allele-sharing within serodiscordant heterosexual Zambian couples. Lancet 363:2137, 2004[PMID: 15220037]

Dube MP, Sattler FR: Inflammation and complications of HIV disease. J Infect Dis 201:1783, 2010[PMID: 20446849]

Epstein JS, Holmberg JA: Progress in monitoring blood safety. Transfusion 50:1408, 2010[PMID: 20636529]

Este JA, Telenti A: HIV entry inhibitors. Lancet 370:81, 2007[PMID: 17617275]

Fauci AS: The AIDS epidemic—considerations for the 21st century. N Engl J Med 341:1046, 1999[PMID: 10502595]

———, Folkers GK: Investing to meet the scientific challenges of HIV/AIDS. Health Aff (Millwood) 28:1629, 2009

Fellay J et al: Host genetics and HIV-1: The final phase? PLoS Pathog 6:e1001033, 2010

Folkers GK, Fauci AS: Controlling and ultimately ending the HIV/AIDS pandemic: A feasible goal. JAMA 304:350, 2010[PMID: 20639573]

Freiberg M et al: The association between hepatitis C infection and prevalent cardiovascular disease among HIV-infected individuals. AIDS 21:193, 2007[PMID: 17197810]

French MA et al: Serum immune activation markers are persistently increased in patients with HIV infection after 6 years of antiretroviral therapy despite suppression of viral replication and reconstitution of $CD4^+$ T cells. J Infect Dis 200:1212, 2009[PMID: 19728788]

Ganczak M, Barss P: Nosocomial HIV infection: Epidemiology and prevention—a global perspective. AIDS Rev 10:47, 2008[PMID: 18385780]

Geretti AM: Epidemiology of antiretroviral drug resistance in drug-naive persons. Curr Opin Infect Dis 20:22, 2007[PMID: 17197878]

Getahun H et al: HIV infection-associated tuberculosis: The epidemiology and the response. Clin Infect Dis 50:S201, 2010

Grant PM et al: Risk factor analyses for immune reconstitution inflammatory syndrome in a randomized study of early vs. deferred ART during an opportunistic infection. PLoS One 5:e11416, 2010

Grant RM et al: Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 363:2587, 2010[PMID: 21091279]

Gray RH et al: Male circumcision for HIV prevention in men in Rakai, Uganda: A randomised trial. Lancet 369:657, 2007[PMID: 17321311]

——— et al: Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. Lancet 357:1149, 2001

Greene WC, Peterlin BM: Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nat Med 8:673, 2002[PMID: 12091904]

Grulich AE: Cancer: The effects of HIV and antiretroviral therapy, and implications for early antiretroviral therapy initiation. Curr Opin HIV AIDS 4:183, 2009[PMID: 19532048]

Gutierrez F et al: Osteonecrosis in patients infected with HIV: Clinical epidemiology and natural history in a large case series from Spain. J AIDS 42:286, 2006[PMID: 16763523]

Haase AT: Targeting early infection to prevent HIV-1 mucosal transmission. Nature 464:217, 2010[PMID: 20220840]

———: Perils at mucosal front lines for HIV and SIV and their hosts. Nat Rev Immunol 5:783, 2005

Hall HI et al: Estimation of HIV incidence in the United States. JAMA 300:520, 2008[PMID: 18677024]

Ho DD et al: Rapid turnover of plasma virions and CD4 lymphocytes in HIV infection. Nature 373:123, 1995[PMID: 7816094]

Joint united nations programme on HIV/AIDS (UNAIDS): Report on the global AIDS epidemic, 2008.

———: AIDS epidemic update, 2009.

Kaplan JE et al: Guidelines for prevention and treatment of opportunistic infections in HIV-infected adults and adolescents. Recommendations from the CDC, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. MMWR Recomm Rep 58:1, 2009. Updates available at *www.aidsinfo.nih.gov*

Keele BF et al: Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313:523, 2006[PMID: 16728595]

Kitahata MM et al: Effect of early versus deferred antiretroviral therapy for HIV on survival. N Engl J Med 360:1815, 2009[PMID: 19339714]

Kuller LH et al: Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. PLoS Med 5:e203, 2008

Malvestutto CD, Aberg JA: Coronary heart disease in people infected with HIV. Cleve Clin J Med 77:547, 2010[PMID: 20682517]

Mayer KH, Venkatesh KK: Antiretroviral therapy as HIV prevention: Status and prospects. Am J Public Health 100:1867, 2010[PMID: 20724682]

Mellors JW et al: Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science 272:1167, 1996[PMID: 8638160]

Migueles SA, Connors M: Long-term nonprogressive disease among untreated HIV-infected individuals: Clinical implications of understanding immune control of HIV. JAMA 304:194, 2010[PMID: 20628133]

Mofenson LM: Protecting the next generation—eliminating perinatal HIV-1 infection. N Engl J Med 362:2316, 2010[PMID: 20554987]

——— et al: Guidelines for the prevention and treatment of opportunistic infections among HIV-exposed and HIV-infected children: Recommendations from CDC, the National Institutes of Health, the HIV Medicine Association of the Infectious Diseases Society of America, the Pediatric Infectious Diseases Society, and the American Academy of Pediatrics. MMWR Recomm Rep 58:1, 2009. Updates available at *aidsinfo.nih.gov*

Moir S, Fauci AS: B cells in HIV infection and disease. Nat Rev Immunol 9:235, 2009[PMID: 19319142]

Montaner JS et al: Association of highly active antiretroviral therapy coverage, population viral load, and yearly new HIV diagnoses in British Columbia, Canada: A population-based study. Lancet 376: 532, 2010[PMID: 20638713]

Neuhaus J et al: Markers of inflammation, coagulation, and renal function are elevated in adults with HIV infection. J Infect Dis 201:1788, 2010[PMID: 20446848]

Office of national aids policy: National HIV/AIDS Strategy. Washington, DC, Office of National AIDS Policy, 2010. Available at *www.whitehouse.gov/onap*

Padian NS et al: Weighing the gold in the gold standard: Challenges in HIV prevention research. AIDS 24:621, 2010[PMID: 20179575]

Panel on antiretroviral therapy and medical management of hiv-infected children: Guidelines for the Use of Antiretroviral Agents in Pediatric HIV Infection. August 16, 2010. Updates available at *aidsinfo.nih.gov*

Panel on treatment of hiv-infected pregnant women and prevention of perinatal transmission: Recommendations for Use of Antiretroviral Drugs in Pregnant HIV-1-Infected Women for Maternal Health and Interventions to Reduce Perinatal HIV Transmission in the United States, May, 2010. Updates available at *aidsinfo.nih.gov*

Panlilio AL et al: Updated U.S. Public Health Service guidelines for the management of occupational exposures to HIV and recommendations for postexposure prophylaxis. MMWR Recomm Rep 54:1, 2005[PMID: 16195697]

Pantaleo G, Fauci AS: HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. Nature 362:355, 1993[PMID: 8455722]

Phillips AN et al: The role of HIV in serious diseases other than AIDS. AIDS 22:2409, 2008[PMID: 19005264]

Price JC, Thio CL: Liver disease in the human immunodeficiency virus-1-infected individual. Clin Gastroenterol Hepatol 8:1002, 2010[PMID: 20851211]

Rerks-Ngarm S et al: Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med 361:2209, 2009[PMID: 19843557]

Silvestri G et al: Understanding the benign nature of SIV infection in natural hosts. J Clin Invest 11:3148, 2007

Smith DK et al: Antiretroviral postexposure prophylaxis after sexual, injection-drug use, or other nonoccupational exposure to HIV in the United States: Recommendations from the U.S. Department of Health and Human Services. MMWR Recomm Rep. 54:1, 2005[PMID: 15660015]

Sterling TR et al: HIV infection-related tuberculosis: Clinical manifestations and treatment. Clin Infect Dis 50:S223, 2010

The Strategies for Management of Antiretroviral Therapy (SMART) Study Group: CD4+ count-guided interruption of antiretroviral treatment. N Engl J Med 355:2283, 2006

Taylor BS et al: The challenge of HIV-1 subtype diversity. N Engl J Med 358:1590, 2008[PMID: 18403767]

Thompson MA et al: Antiretroviral treatment of adult HIV infection: 2010 recommendations of the International AIDS Society-USA panel. JAMA 304:321, 2010[PMID: 20639566]

Trono D et al: HIV persistence and the prospect of long-term drug-free remissions for HIV-infected individuals. Science 329:174, 2010[PMID: 20616270]

U.S. Public health service: Updated U.S. Public Health Service Guidelines for the Management of Occupational Exposures to HBV, HCV, and HIV and Recommendations for Postexposure Prophylaxis. MMWR Recomm Rep 50:1, 2001. Updates available at *aidsinfo.nih.gov*

Van Sighem AI et al: Life expectancy of recently diagnosed asymptomatic HIV-infected patients approaches that of uninfected individuals. AIDS 24:1527, 2010

Wawer MJ et al: Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis 191:1403,

2005[PMID: 15809897]

Wei x et al: Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373:117, 1995[PMID: 7529365]

World Health Organization (WHO): Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants: Towards universal access. Recommendations for a public health approach, 2010

———: Antiretroviral therapy for HIV infection in adults and adolescents. Recommendations for a public health approach, 2010

———: Antiretroviral therapy for HIV infection in infants and children: Towards universal access. Recommendations for a public health approach, 2010

———: Priority interventions. HIV/AIDS prevention, treatment and care in the health sector, 2010

———, United Nations Children's Fund (UNICEF), UNAIDS: Towards universal access: Scaling up priority HIV/AIDS interventions in the health sector—progress report 2010

Zancanaro PCQ et al: Cutaneous manifestations of HIV in the era of highly active antiretroviral therapy: An institutional urban clinic experience. J Am Acad Dermatol 54:581, 2006[PMID: 16546578]

Zhou T et al: Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science 329:811, 2010[PMID: 20616231]

Zonios DI et al: Idiopathic CD4+ lymphocytopenia: Natural history and prognostic factors. Blood 112:287, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VIRAL GASTROENTERITIS: INTRODUCTION

Acute infectious gastroenteritis is a common illness that affects persons of all ages worldwide. It is a leading cause of mortality among children in developing countries, accounting for an estimated 1.8 million deaths each year, and is responsible for up to 10–12% of all hospitalizations among children in industrialized countries, including the United States. Elderly persons, especially those with debilitating health conditions, are also at risk of severe complications and death from acute gastroenteritis. Among healthy young adults, acute gastroenteritis is rarely fatal but incurs substantial medical and social costs, including those of time lost from work.

Several enteric viruses have been recognized as important etiologic agents of acute infectious gastroenteritis **(Table 190-1**, **Fig. 190-1)**. Although most viral gastroenteritis is caused by RNA viruses, the DNA viruses that are occasionally involved (e.g., adenovirus types 40 and 41) are included in this chapter. Illness caused by these viruses is characterized by the acute onset of vomiting and/or diarrhea, which may be accompanied by fever, nausea, abdominal cramps, anorexia, and malaise. As shown in **Table 190-2**, several features can help distinguish gastroenteritis caused by viruses from that caused by bacterial agents. However, the distinction based on clinical and epidemiologic parameters alone is often difficult, and laboratory tests may be required to confirm the diagnosis.

Table 190-1 Viral Causes of Gastroenteritis Among Humans

Virus	Family	Genome	Primary Age Group at Risk	Clinical Severity	Detection Assays
Group A rotavirus	Reoviridae	Double-strand segmented RNA	Children <5 years	+ + +	EM, EIA (commercial), PAGE, RT-PCR
Norovirus	Caliciviridae	Positive-sense single-strand RNA	All ages	+ +	EM, EIA, RT-PCR
Sapovirus	Caliciviridae	Positive-sense single-strand RNA	Children <5 years	+	EM, EIA, RT-PCR
Astrovirus	Astroviridae	Positive-sense single-strand RNA	Children <5 years	+	EM, EIA, RT-PCR
Adenovirus (types 40 and 41)	Adenoviridae	Double-strand DNA	Children <5 years	+/+ +	EM, EIA (commercial), PCR

Abbreviations: EIA, enzyme immunoassay; EM, electron microscopy; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; RT-PCR, reverse-transcription PCR.

Table 190-2 Characteristics of Gastroenteritis Caused by Viral and Bacterial Agents

Feature	Viral Gastroenteritis	Bacterial Gastroenteritis
Setting	Incidence similar in developing and developed countries	More common in settings with poor hygiene and sanitation
Infectious dose	Low (10–100 viral particles) for most agents	High (>10^5 bacteria) for *Escherichia coli, Salmonella, Vibrio;* medium (10^2–10^5 bacteria) for *Campylobacter jejuni;* low (10–100 bacteria) for *Shigella*
Seasonality	In temperate climates, winter seasonality for most agents; year-round occurrence in tropical areas	More common in summer or rainy months, particularly in developing countries with a high disease burden
Incubation period	1–3 days for most agents; can be shorter for norovirus	1–7 days for common agents (e.g., *Campylobacter, E. coli, Shigella, Salmonella);*a few hours for bacteria producing preformed toxins (e.g., *Staphylococcus aureus, Bacillus cereus?*)

Reservoir	Primarily humans	Depending on species, human (e.g., *Shigella, Salmonella?*), animal (e.g., *Campylobacter, Salmonella, E. coli?*), and water (e.g., *Vibrio?*) reservoirs exist.
Fever	Common with rotavirus and norovirus; uncommon with other agents	Common with agents causing inflammatory diarrhea (e.g., *Salmonella, Shigella?*)
Vomiting	Prominent and can be the only presenting feature, especially in children	Common with bacteria producing preformed toxins; less prominent in diarrhea due to other agents
Diarrhea	Common; nonbloody in almost all cases	Prominent and frequently bloody with agents causing inflammatory diarrhea
Duration	1–3 days for norovirus and sapovirus; 2–8 days for other viruses	1–2 days for bacteria producing preformed toxins; 2–8 days for most other bacteria
Diagnosis	This is often a diagnosis of exclusion in clinical practice. Commercial enzyme immunoassays are available for detection of rotavirus and adenovirus, but identification of other agents is limited to research and public health laboratories.	Fecal examination for leukocytes and blood is helpful in differential diagnosis. Culture of stool specimens, sometimes on special media, can identify several pathogens. Molecular techniques are useful epidemiologic tools but are not routinely used in most laboratories.
Treatment	Supportive therapy to maintain adequate hydration and nutrition should be given. Antibiotics and antimotility agents are contraindicated.	Supportive hydration therapy is adequate for most patients. Antibiotics are recommended for patients with dysentery caused by *Shigella* or *Vibrio cholerae* and for some patients with *Clostridium difficile* colitis.

Figure 190-1

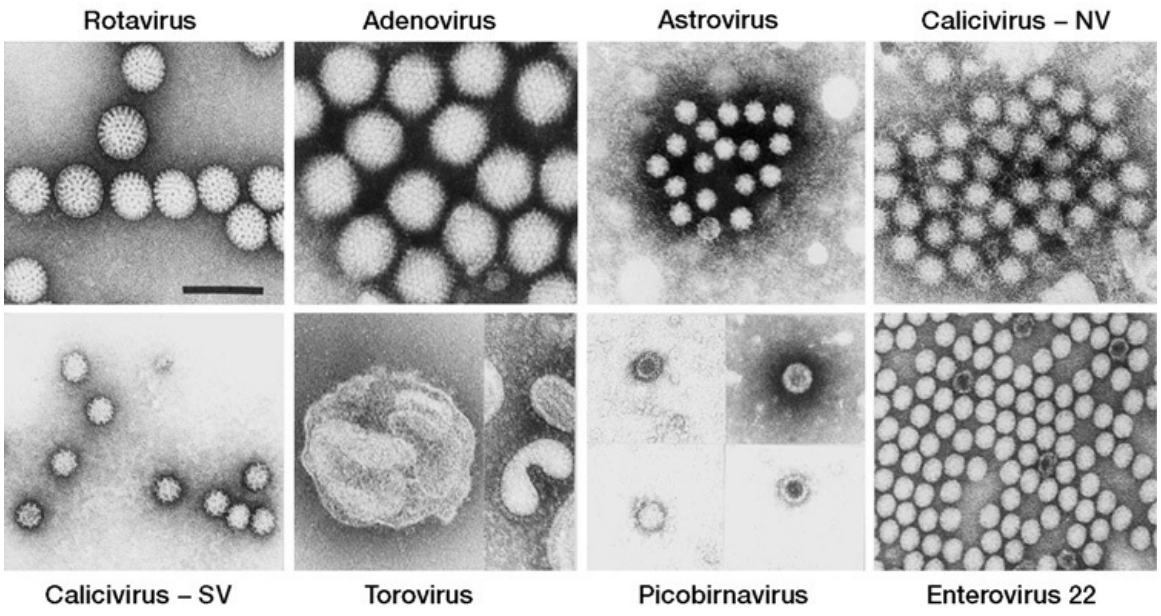

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Viral agents of gastroenteritis. NV, norovirus; SV, sapovirus.

Human Caliciviruses

ETIOLOGIC AGENT

The Norwalk virus is the prototype strain of a group of nonenveloped, small (27–40 nm), round, icosahedral viruses with relatively amorphous surface features on visualization by electron microscopy. These viruses have been difficult to classify because they have not been adapted to cell culture, they often are shed in low titers for only a few days, and no animal models are available. Molecular cloning and characterization have demonstrated that the viruses have a single, positive-strand RNA genome ~ 7.5 kb in length and that they possess a single virion-associated protein—similar to that of typical caliciviruses—with a molecular mass of 60 kDa. On the basis of these molecular characteristics, these viruses are presently classified in two genera belonging to the family Caliciviridae: the *noroviruses* and the *sapoviruses* (previously called Norwalk-like viruses and Sapporo-like viruses, respectively).

EPIDEMIOLOGY

Infections with the Norwalk and related human caliciviruses are common worldwide, and most adults have antibodies to these viruses. Antibody is acquired at an earlier age in developing countries—a pattern consistent with the presumed fecal-oral mode of transmission. Infections occur year-round, although, in temperate climates, a distinct increase has been noted in cold-weather months. Noroviruses may be the most common infectious agents of mild gastroenteritis in the community and affect all age groups, whereas sapoviruses primarily cause gastroenteritis in children. Noroviruses also cause traveler's diarrhea, and outbreaks have occurred among military personnel deployed to various parts of the world. The limited data available indicate that norovirus may be the second most common viral agent (after rotavirus) among young children and the most common agent among older children and adults. For example, in a comprehensive evaluation of eight enteric pathogens in patients with gastroenteritis in England, three-fourths of patients had at least one pathogen detected in fecal specimens, and noroviruses were the most prevalent, detected in 36% of patients and 18% of healthy controls. Noroviruses are also recognized as the major cause of epidemics of gastroenteritis worldwide. In the United States, >90% of outbreaks of nonbacterial gastroenteritis are caused by noroviruses.

Virus is transmitted predominantly by the fecal-oral route but is also present in vomitus. Because an inoculum with very few viruses can be infectious, transmission can occur by aerosolization, by contact with contaminated fomites, and by person-to-person contact. Viral shedding and infectivity are greatest during the acute illness, but challenge studies with Norwalk virus in volunteers indicate that viral antigen may be shed by asymptomatically infected persons and also by symptomatic persons before the onset of symptoms and for several weeks after the resolution of illness.

PATHOGENESIS

The exact sites and cellular receptors for attachment of viral particles have not been determined. Data suggest that carbohydrates that are similar to human histo-blood group antigens and are present on the gastroduodenal epithelium of individuals with the secretor phenotype may serve as ligands for the attachment of Norwalk virus. Additional studies must more fully elucidate norovirus-carbohydrate interactions, including potential strain-specific variations. After the infection of volunteers, reversible lesions are noted in the upper jejunum, with broadening and blunting of the villi, shortening of the microvilli, vacuolization of the lining epithelium, crypt hyperplasia, and infiltration of the lamina propria by polymorphonuclear neutrophils and lymphocytes. The lesions persist for at least 4 days after the resolution of symptoms and are associated with malabsorption of carbohydrates and fats and a decreased level of brush-border enzymes. Adenylate cyclase activity is not altered. No histopathologic changes are seen in the stomach or colon, but gastric motor function is delayed, and this alteration is believed to contribute to the nausea and vomiting that are typical of this illness.

CLINICAL MANIFESTATIONS

Gastroenteritis caused by Norwalk and related human caliciviruses has a sudden onset, following an average incubation period of 24 h (range, 12–72 h). The illness generally lasts 12–60 h and is characterized by one or more of the following symptoms: nausea, vomiting, abdominal cramps, and diarrhea. Vomiting is more prevalent among children, whereas a greater proportion of adults develop diarrhea. Constitutional symptoms are common, including headache, fever, chills, and myalgias. The stools are characteristically loose and watery, without blood, mucus, or leukocytes. White cell counts are generally normal; rarely, leukocytosis with relative lymphopenia may be observed. Death is a rare outcome and usually results from severe dehydration in vulnerable persons (e.g., elderly patients with debilitating health conditions).

IMMUNITY

Approximately 50% of persons challenged with Norwalk virus become ill and acquire short-term immunity against the infecting strain. Immunity to Norwalk virus appears to correlate inversely with level of antibody; i.e., persons with higher levels of preexisting antibody to Norwalk virus are more susceptible to illness. This observation suggests that some individuals have a genetic predisposition to illness. Specific ABO, Lewis, and secretor blood group phenotypes may influence susceptibility to norovirus infection.

DIAGNOSIS

Cloning and sequencing of the genomes of Norwalk and several other human caliciviruses have allowed the development of assays based on polymerase chain reaction (PCR) for detection of virus in stool and vomitus. Virus-like particles produced by expression of capsid proteins in a recombinant baculovirus vector have been used to develop enzyme immunoassays (EIAs) for detection of virus in stool or a serologic response to a specific viral antigen. These newer diagnostic techniques are considerably more sensitive than previous detection methods, such as electron microscopy, immune electron microscopy, and EIAs based on reagents derived from humans. However, no currently available single assay can detect all human caliciviruses because of their great genetic and antigenic diversity. In addition, the assays are still cumbersome and are available primarily in research laboratories, although they are increasingly being adopted by public health laboratories for routine

screening of fecal specimens from patients affected by outbreaks of gastroenteritis. Commercial EIA kits, which are available in some European countries and in Japan but not yet in the United States, have limited sensitivity and usefulness in clinical practice and are of greatest utility in outbreaks, in which many specimens are tested and only a few need be positive to identify norovirus as the cause.

TREATMENT: INFECTIONS WITH NORWALK AND RELATED HUMAN CALICIVIRUSES

The disease is self-limited, and oral rehydration therapy is generally adequate. If severe dehydration develops, IV fluid therapy is indicated. No specific antiviral therapy is available.

PREVENTION

Epidemic prevention relies on situation-specific measures, such as control of contamination of food and water, exclusion of ill food handlers, and reduction of person-to-person spread through good personal hygiene and disinfection of contaminated fomites. The role of immunoprophylaxis is not clear, given the lack of long-term immunity from natural disease, but efforts to develop norovirus vaccines are ongoing.

Rotavirus

ETIOLOGIC AGENT

Rotaviruses are members of the family Reoviridae. The viral genome consists of 11 segments of double-strand RNA that are enclosed in a triple-layered, nonenveloped, icosahedral capsid 75 nm in diameter. Viral protein 6 (VP6), the major structural protein, is the target of commercial immunoassays and determines the group specificity of rotaviruses. There are seven major groups of rotavirus (A through G); human illness is caused primarily by group A and, to a much lesser extent, by groups B and C. Two outer-capsid proteins, VP7 (G-protein) and VP4 (P-protein), determine serotype specificity, induce neutralizing antibodies, and form the basis for binary classification of rotaviruses (G and P types). The segmented genome of rotavirus allows genetic reassortment (i.e., exchange of genome segments between viruses) during co-infection—a property that may play a role in viral evolution and has been utilized in the development of reassortant animal-human rotavirus–based vaccines.

EPIDEMIOLOGY

Worldwide, nearly all children are infected with rotavirus by 3–5 years of age. Neonatal infections are common but are often asymptomatic or mild, presumably because of protection from maternal antibody or breast-feeding. First infections after 3 months of age are likely to be symptomatic, and the incidence of disease peaks among children 4–23 months of age. Reinfections are common, but the severity of disease decreases with each repeat infection. Therefore, severe rotavirus infections are relatively uncommon among older children and adults. Nevertheless, rotavirus can cause illness in parents and caretakers of children with rotavirus diarrhea, immunocompromised persons, travelers, and elderly individuals and should be considered in the differential diagnosis of gastroenteritis among adults.

In tropical settings, rotavirus disease occurs year-round, with less pronounced seasonal peaks than in temperate settings, where rotavirus disease occurs predominantly during the cooler fall and winter months. Before the introduction of rotavirus vaccine in the United States, the rotavirus season each year began in the Southwest during the autumn and early winter (October through December) and migrated across the continent, peaking in the Northeast during late winter and spring (March through May). The reasons for this characteristic pattern are not clear, but a recent study suggested a correlation with state-specific differences in birth rates, which could influence the rate of accumulation of susceptible infants after each rotavirus season. After the implementation of routine vaccination of U.S. infants against rotavirus in 2006, the onset of the 2007–2008 and 2008–2009 rotavirus seasons was delayed by 11 weeks and 6 weeks, respectively, and the seasons were shorter, lasting 14 and 17 weeks, respectively, in comparison with a median of 26 weeks in 2000–2006 **(Fig. 190-2)**. These changes in seasonal patterns of rotavirus activity were accompanied by declines in the number of detections of rotavirus by 64% and 60% in 2007–2008 and 2008–2009, respectively, from the figures for 2000–2006, as collected by a national network of sentinel laboratories.

Figure 190-2

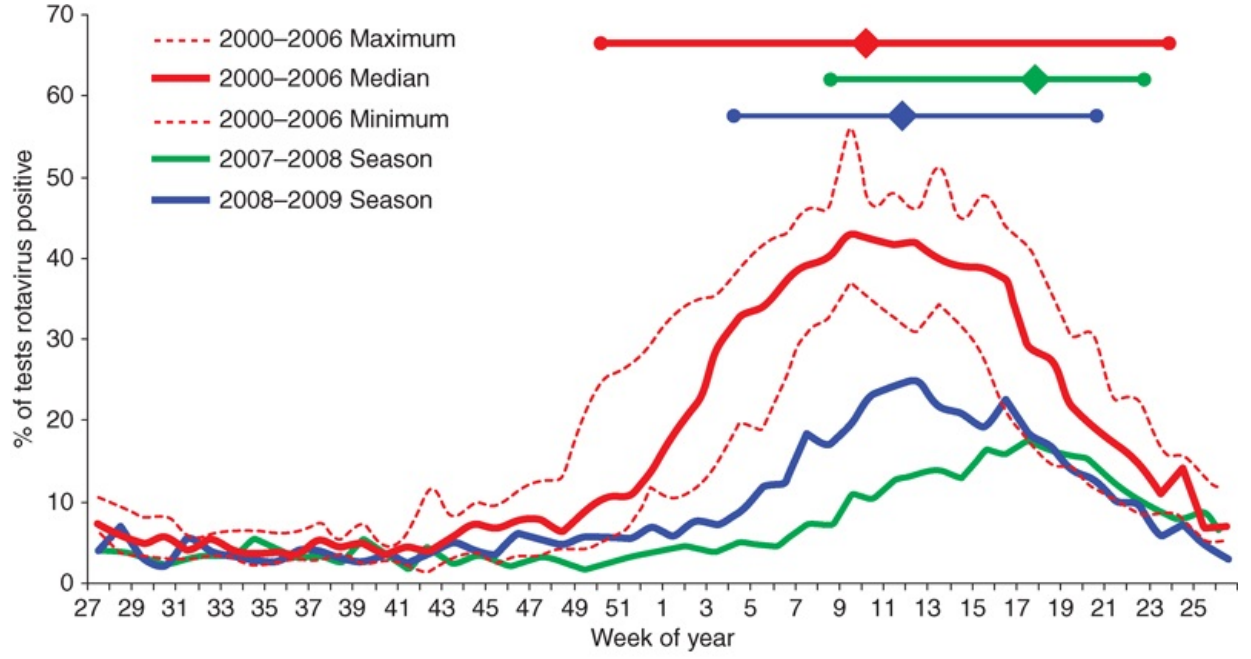

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The maximal or minimal percentage of rotavirus-positive tests for 2000–2006 may have occurred during any of the six baseline seasons. The onset of rotavirus season was defined as the first of two consecutive weeks during which the percentage of stool specimens testing positive for rotavirus was ≥10%, and the end of the season was defined as the last of two consecutive weeks during which the percentage of stool specimens testing positive for rotavirus was ≥10%. At the top right, the dots bracket the rotavirus season from onset to end, and the diamond indicates the peak week during each period. *(Adapted from Centers for Disease Control and Prevention, 2009.)*

During episodes of rotavirus-associated diarrhea, virus is shed in large quantities in stool (10^7–10^{12}/g). Viral shedding detectable by EIA usually subsides within 1 week but may persist for >30 days in immunocompromised individuals. Viral shedding may be detected for longer periods by sensitive molecular assays, such as PCR. The virus is transmitted predominantly through the fecal-oral route. Spread through respiratory secretions, person-to-person contact, or contaminated environmental surfaces has also been postulated to explain the rapid acquisition of antibody in the first 3 years of life, regardless of sanitary conditions.

At least 10 different G serotypes of group A rotavirus have been identified in humans, but only five types (G1 through G4 and G9) are common. While human rotavirus strains that possess a high degree of genetic homology with animal strains have been identified, animal-to-human transmission appears to be uncommon.

Group B rotaviruses have been associated with several large epidemics of severe gastroenteritis among adults in China since 1982 and have also been identified in India. Group C rotaviruses have been associated with a small proportion of pediatric gastroenteritis cases in several countries worldwide.

PATHOGENESIS

Rotaviruses infect and ultimately destroy mature enterocytes in the villous epithelium of the proximal small intestine. The loss of absorptive villous epithelium, coupled with the proliferation of secretory crypt cells, results in secretory diarrhea. Brush-border enzymes characteristic of differentiated cells are reduced, and this change leads to the accumulation of unmetabolized disaccharides and consequent osmotic diarrhea. Studies in mice indicate that a nonstructural rotavirus protein, NSP4, functions as an enterotoxin and contributes to secretory diarrhea by altering epithelial cell function and permeability. In addition, rotavirus may evoke fluid secretion through activation of the enteric nervous system in the intestinal wall. Recent data indicate that rotavirus antigenemia and viremia are common among children with acute rotavirus infection, although the antigen and RNA levels in serum are substantially lower than those in stool.

CLINICAL MANIFESTATIONS

The clinical spectrum of rotavirus infection ranges from subclinical infection to severe gastroenteritis leading to life-threatening dehydration. After an incubation period of 1–3 days, the illness has an abrupt onset, with vomiting frequently preceding the onset of diarrhea. Up to one-third of patients may have a temperature of >39°C. The stools are characteristically loose and watery and only infrequently contain red or white cells. Gastrointestinal symptoms generally resolve in 3–7 days.

Respiratory and neurologic features in children with rotavirus infection have been reported, but causal associations have not been proven. Moreover, rotavirus infection has been associated with a variety of other clinical conditions (e.g., sudden infant death syndrome, necrotizing enterocolitis, intussusception, Kawasaki's disease, and type 1 diabetes), but no causal relationship has been confirmed with any of these syndromes.

Rotavirus does not appear to be a major opportunistic pathogen in children with HIV infection. In severely immunodeficient children, rotavirus can cause protracted diarrhea with prolonged viral excretion and, in rare instances, can disseminate systemically. Persons who are immunosuppressed for bone marrow transplantation are also at risk for severe or even fatal rotavirus disease.

IMMUNITY

Protection against rotavirus disease is correlated with the presence of virus-specific secretory IgA antibodies in the intestine and, to some extent, the serum. Because virus-specific IgA production at the intestinal surface is short lived, complete protection against disease is only temporary. However, each infection and subsequent reinfection confers progressively greater immunity; thus severe disease is most common among young children with first or second infections. Immunologic memory is believed to be important in the attenuation of disease severity upon reinfection.

DIAGNOSIS

Illness caused by rotavirus is difficult to distinguish clinically from that caused by other enteric viruses. Because large quantities of virus are shed in feces, the diagnosis can usually be confirmed by a wide variety of commercially available EIAs or by techniques for detecting viral RNA, such as gel electrophoresis, probe hybridization, or PCR.

TREATMENT: ROTAVIRUS INFECTIONS

Rotavirus gastroenteritis can lead to severe dehydration. Thus appropriate treatment should be instituted early. Standard oral rehydration therapy is successful in most children who can take oral fluids, but IV fluid replacement may be required for patients who are severely dehydrated or are unable to tolerate oral therapy because of frequent vomiting. The therapeutic role of probiotics, bismuth subsalicylate, enkephalinase inhibitors, and nitazoxanide has been evaluated in clinical studies but is not clearly defined. Antibiotics and antimotility agents should be avoided. In immunocompromised children with chronic symptomatic rotavirus disease, orally administered immunoglobulins or colostrum may result in the resolution of symptoms, but the best choices regarding agents and their doses have not been well studied, and treatment decisions are often empirical.

PREVENTION

Efforts to develop rotavirus vaccines were pursued because it was apparent—given the similar rates in less-developed and industrialized nations—that improvements in hygiene and sanitation were unlikely to reduce disease incidence. The first rotavirus vaccine licensed in the United States in 1998 was withdrawn from the market within 1 year because it was linked with intussusception, a severe bowel obstruction.

In 2006, promising safety and efficacy results for two new rotavirus vaccines were reported from large clinical trials conducted in North America, Europe, and Latin America. Both vaccines are now recommended for routine immunization of all U.S. infants, and their use has rapidly led to a decline in rotavirus hospitalizations and emergency department visits at hospitals across the United States. In Mexico, a decline in deaths from childhood diarrhea following introduction of rotavirus vaccines has been documented. Furthermore, postmarketing surveillance information has not revealed an association of these vaccines with any serious adverse events (including intussusception), although a risk of low magnitude cannot be excluded on the basis of available data.

GLOBAL CONSIDERATIONS

Rotavirus is ubiquitous and infects nearly all children worldwide by 5 years of age. However, compared with rotavirus disease in industrialized countries, disease in developing countries occurs at a younger age, is less seasonal, and is more frequently caused by uncommon rotavirus strains. Moreover, because of suboptimal access to hydration therapy, rotavirus is a leading cause of diarrheal death among children in the developing world, with the highest mortality rates among children in sub-Saharan Africa and southern Asia **(Fig. 190-3).**

Figure 190-3

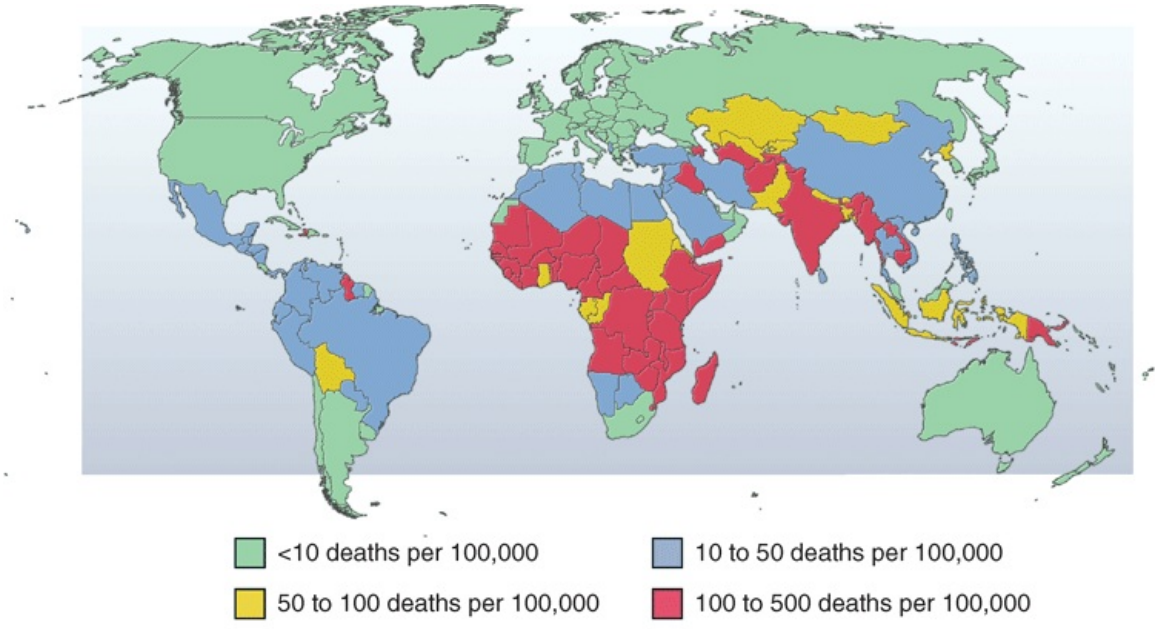

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Rotavirus mortality rates by country, per 100,000 children <5 years of age. *(Reproduced with permission from UD Parashar et al: J Infect Dis 200:S9, 2009.)*

The different epidemiology of rotavirus disease and the greater prevalence of co-infection with other enteric pathogens, of comorbidities, and of malnutrition in developing countries may adversely affect the performance of oral rotavirus vaccines, as is the case with oral vaccines against poliomyelitis, cholera, and typhoid in these regions. Therefore, evaluation of the efficacy of rotavirus vaccines in resource-poor settings of Africa and Asia was specifically recommended, and these trials have now been completed. As anticipated, the efficacy of rotavirus vaccines was moderate (50–75%) in these settings when compared with that in industrialized countries. Nevertheless, even a moderately efficacious rotavirus vaccine would be likely to have substantial public health benefits in these areas with a high disease burden. Given these considerations, in April 2009 the World Health Organization recommended the use of rotavirus vaccines in all countries worldwide.

Other Viral Agents of Gastroenteritis

Enteric *adenoviruses* of serotypes 40 and 41 belonging to subgroup F are 70- to 80-nm viruses with double-strand DNA that cause ~2–12% of all diarrhea episodes in young children. Unlike adenoviruses that cause respiratory illness, enteric adenoviruses are difficult to cultivate in cell lines, but they can be detected with commercially available EIAs.

Astroviruses, 28- to 30-nm viruses with a characteristic icosahedral structure, contain a positive-sense, single-strand RNA. At least seven serotypes have been identified, of which serotype 1 is most common. Astroviruses are primarily pediatric pathogens, causing ~2–10% of cases of mild to moderate gastroenteritis in children. The availability of simple immunoassays to detect virus in fecal specimens and of molecular methods to confirm and characterize strains will permit more comprehensive assessment of the etiologic role of these agents.

Toroviruses are 100- to 140-nm, enveloped, positive-strand RNA viruses that are recognized as causes of gastroenteritis in horses (Berne virus) and cattle (Breda virus). Their role as a cause of diarrhea in humans is still unclear, but studies from Canada have demonstrated associations between torovirus excretion and both nosocomial gastroenteritis and necrotizing enterocolitis in neonates. These associations require further evaluation.

Picobirnaviruses are small, bisegmented, double-strand RNA viruses that cause gastroenteritis in a variety of animals. Their role as primary causes of gastroenteritis in humans remains unclear, but several studies have found an association between picobirnaviruses and gastroenteritis in HIV-infected adults.

Several other viruses (e.g., enteroviruses, reoviruses, pestiviruses, and parvovirus B) have been identified in the feces of patients with diarrhea, but their etiologic role in gastroenteritis has not been proven. Diarrhea has also been noted as a manifestation of infection with recently recognized viruses that primarily cause severe respiratory illness: the severe acute respiratory syndrome–associated coronavirus (SARS-CoV), influenza A/H5N1 virus, and the current pandemic strain of influenza A/H1N1 virus.

FURTHER READINGS

Amar CF et al: Detection by PCR of eight groups of enteric pathogens in 4,627 fecal specimens: Re-examination of the English case-control Infectious Intestinal Disease Study (1993–1996). Eur J Clin Microbiol Infect Dis 26:311, 2007[PMID: 17447091]

Centers for Disease Control and Prevention: Reduction in rotavirus after vaccine introduction—United States, 2000–2009. MMWR Morb Mortal Wkly Rep 58:1146, 2009

Curns AT et al: Reduction in acute gastroenteritis hospitalizations among U.S. children after introduction of rotavirus vaccine: Analysis of hospital discharge data from 18 U.S. states. J Infect Dis 201:1617, 2010[PMID: 20402596]

Glass RI et al: Norovirus gastroenteritis. N Engl J Med 361:1776, 2009[PMID: 19864676]

Madhi SA et al: Effect of human rotavirus vaccine on severe diarrhea in African infants. N Engl J Med 362:289, 2010[PMID: 20107214]

Patel MM et al: Systematic literature review of role of noroviruses in sporadic gastroenteritis. Emerg Infect Dis 14:1224, 2008[PMID: 18680645]

Richardson V et al: Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. N Engl J Med 362:299, 2010[PMID: 20107215]

Pitzer VE et al: Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics. Science 325:274, 2009

Ruiz-Palacios G et al: Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. N Engl J Med 354:11, 2006[PMID: 16394298]

Vesikari T et al: Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. N Engl J Med 354:23, 2006[PMID: 16394299]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ENTEROVIRUSES

Classification and Characterization

Enteroviruses are so named because of their ability to multiply in the gastrointestinal tract. Despite their name, these viruses are not a prominent cause of gastroenteritis. Enteroviruses encompass 96 human serotypes: 3 serotypes of poliovirus, 21 serotypes of coxsackievirus A, 6 serotypes of coxsackievirus B, 28 serotypes of echovirus, enteroviruses 68–71, and 34 new enteroviruses (beginning with enterovirus 73) that have been identified by molecular techniques. Echoviruses 22 and 23 have been reclassified as parechoviruses 1 and 2; 12 additional human parechoviruses have been identified. These viruses cause disease similar to that caused by echoviruses. Enterovirus surveillance conducted in the United States by the Centers for Disease Control and Prevention (CDC) in 2007–2008 showed that the most common serotype, coxsackievirus B1, was followed in frequency by echoviruses 18, 9, and 6; together, these four viruses accounted for 52% of all isolates.

Human enteroviruses contain a single-stranded RNA genome surrounded by an icosahedral capsid comprising four viral proteins. These viruses have no lipid envelope and are stable in acidic environments, including the stomach. They are susceptible to chlorine-containing cleansers but resistant to inactivation by standard disinfectants (e.g., alcohol, detergents) and can persist for days at room temperature.

Pathogenesis and Immunity

Much of what is known about the pathogenesis of enteroviruses has been derived from studies of poliovirus infection. After ingestion, poliovirus is thought to infect epithelial cells in the mucosa of the gastrointestinal tract and then to spread to and replicate in the submucosal lymphoid tissue of the tonsils and Peyer's patches. The virus next spreads to the regional lymph nodes, a viremic phase ensues, and the virus replicates in organs of the reticuloendothelial system. In some cases, a second viremia occurs and the virus replicates further in various tissues, sometimes causing symptomatic disease.

It is uncertain whether poliovirus reaches the central nervous system (CNS) during viremia or whether it also spreads via peripheral nerves. Since viremia precedes the onset of neurologic disease in humans, it has been assumed that the virus enters the CNS via the bloodstream. The poliovirus receptor is a member of the immunoglobulin superfamily. Poliovirus infection is limited to primates, largely because their cells express the viral receptor. Studies demonstrating the poliovirus receptor in the end-plate region of muscle at the neuromuscular junction suggest that, if the virus enters the muscle during viremia, it could travel across the neuromuscular junction up the axon to the anterior horn cells. Studies of monkeys and of transgenic mice expressing the poliovirus receptor show that, after IM injection, poliovirus does not reach the spinal cord if the sciatic nerve is cut. Taken together, these findings suggest that poliovirus can spread directly from muscle to the CNS by neural pathways. Intercellular adhesion molecule 1 (ICAM-1) is a receptor for coxsackieviruses A13, A18, and A21; CAR for coxsackievirus B; VLA-2 integrin for echovirus types 1 and 8; CD55 for enterovirus 70 and some serotypes of coxsackievirus A and B and echovirus; and P-selectin glycoprotein ligand-1 and scavenger receptor B2 for enterovirus 71.

Poliovirus can usually be cultured from the blood 3–5 days after infection, before the development of neutralizing antibodies. While viral replication at secondary sites begins to slow 1 week after infection, it continues in the gastrointestinal tract. Poliovirus is shed from the oropharynx for up to 3 weeks after infection and from the gastrointestinal tract for as long as 12 weeks; hypogammaglobulinemic patients can shed poliovirus for >20 years. During replication in the gastrointestinal tract, attenuated oral poliovirus can mutate, reverting to a more neurovirulent phenotype within a few days; however, additional mutations are probably required for full neurovirulence.

Humoral and secretory immunity in the gastrointestinal tract is important for the control of enterovirus infections. Enteroviruses induce specific IgM, which usually persists for <6 months, and specific IgG, which persists for life. Capsid protein VP1 is the predominant target of neutralizing antibody, which generally confers lifelong protection against subsequent disease caused by the same serotype but does not prevent infection or virus shedding. Enteroviruses also induce cellular immunity whose significance is uncertain. Patients with impaired cellular immunity are not known to develop unusually severe disease when infected with enteroviruses. In contrast, the severe infections in patients with agammaglobulinemia emphasize the importance of humoral immunity in controlling enterovirus

infections. Disseminated enterovirus infections have occurred in hematopoietic cell transplant recipients. IgA antibodies are instrumental in reducing poliovirus replication in and shedding from the gastrointestinal tract. Breast milk contains IgA specific for enteroviruses and can protect humans from infection.

Epidemiology

Enteroviruses have a worldwide distribution. More than 50% of nonpoliovirus enterovirus infections and more than 90% of poliovirus infections are subclinical. When symptoms do develop, they are usually nonspecific and occur in conjunction with fever; only a minority of infections are associated with specific clinical syndromes. The incubation period for most enterovirus infections ranges from 2 to 14 days but usually is <1 week.

Enterovirus infection is more common in socioeconomically disadvantaged areas, especially in those where conditions are crowded and in tropical areas where hygiene is poor. Infection is most common among infants and young children; serious illness develops most often during the first few days of life and in older children and adults. In developing countries, where children are infected at an early age, poliovirus infection has less often been associated with paralysis; in countries with better hygiene, older children and adults are more likely to be seronegative, become infected, and develop paralysis. Passively acquired maternal antibody reduces the risk of symptomatic infection in neonates. Young children are the most frequent shedders of enteroviruses and are usually the index cases in family outbreaks. In temperate climates, enterovirus infections occur most often in the summer and fall; no seasonal pattern is apparent in the tropics.

Most enteroviruses are transmitted primarily by the fecal-oral or oral-oral route. Patients are most infectious shortly before and after the onset of symptomatic disease, when virus is present in the stool and throat. The ingestion of virus-contaminated food or water can also cause disease. Certain enteroviruses (such as enterovirus 70, which causes acute hemorrhagic conjunctivitis) can be transmitted by direct inoculation from the fingers to the eye. Airborne transmission is important for some viruses that cause respiratory tract disease, such as coxsackievirus A21. Enteroviruses can be transmitted across the placenta from mother to fetus, causing severe disease in the newborn. The transmission of enteroviruses through blood transfusions or insect bites has not been documented. Nosocomial spread of coxsackievirus and echovirus has taken place in hospital nurseries.

Clinical Features

POLIOVIRUS INFECTION

Most infections with poliovirus are asymptomatic. After an incubation period of 3–6 days, ~5% of patients present with a minor illness (abortive poliomyelitis) manifested by fever, malaise, sore throat, anorexia, myalgias, and headache. This condition usually resolves in 3 days. About 1% of patients present with aseptic meningitis (nonparalytic poliomyelitis). Examination of cerebrospinal fluid (CSF) reveals lymphocytic pleocytosis, a normal glucose level, and a normal or slightly elevated protein level; CSF polymorphonuclear leukocytes may be present early. In some patients, especially children, malaise and fever precede the onset of aseptic meningitis.

Paralytic Poliomyelitis

The least common presentation is that of paralytic disease. After one or several days, signs of aseptic meningitis are followed by severe back, neck, and muscle pain and by the rapid or gradual development of motor weakness. In some cases the disease appears to be biphasic, with aseptic meningitis followed first by apparent recovery but then (1–2 days later) by the return of fever and the development of paralysis; this form is more common among children than among adults. Weakness is generally asymmetric, is proximal more than distal, and may involve the legs (most commonly); the arms; or the abdominal, thoracic, or bulbar muscles. Paralysis develops during the febrile phase of the illness and usually does not progress after defervescence. Urinary retention may also occur. Examination reveals weakness, fasciculations, decreased muscle tone, and reduced or absent reflexes in affected areas. Transient hyperreflexia sometimes precedes the loss of reflexes. Patients frequently report sensory symptoms, but objective sensory testing usually yields normal results. Bulbar paralysis may lead to dysphagia, difficulty in handling secretions, or dysphonia. Respiratory insufficiency due to aspiration, involvement of the respiratory center in the medulla, or paralysis of the phrenic or intercostal nerves may develop, and severe medullary involvement may lead to circulatory collapse. Most patients with paralysis recover some function weeks to months after infection. About two-thirds of patients have residual neurologic sequelae.

Paralytic disease is more common among older individuals, pregnant women, and persons exercising strenuously or undergoing trauma at the time of CNS symptoms. Tonsillectomy predisposes to bulbar poliomyelitis, and IM injections increase the risk of paralysis in the involved limb(s).

Vaccine-Associated Poliomyelitis

Until recently, poliomyelitis due to live poliovirus vaccine occurred in the United States. The risk of developing poliomyelitis after oral vaccination is estimated at 1 case per 2.5 million doses. The risk is ~2000 times higher among immunodeficient persons, especially in persons with hypo- or

agammaglobulinemia. Before 1997, an average of eight cases of vaccine-associated poliomyelitis occurred—in both vaccinees and their contacts—in the United States each year. With the change in recommendations first to a sequential regimen of inactivated poliovirus vaccine (IPV) and oral poliovirus vaccine (OPV) in 1997 and then to an all-IPV regimen in 2000, the number of cases of vaccine-associated polio declined. From 1997 to 1999, six such cases were reported in the United States; no cases have been reported since 1999.

Postpolio Syndrome

The *postpolio syndrome* presents as a new onset of weakness, fatigue, fasciculations, and pain with additional atrophy of the muscle group involved during the initial paralytic disease 20–40 years earlier. The syndrome is more common among women and with increasing time after acute disease. The onset is usually insidious, and weakness occasionally extends to muscles that were not involved during the initial illness. The prognosis is generally good; progression to further weakness is usually slow, with plateau periods of 1–10 years. The postpolio syndrome is thought to be due to progressive dysfunction and loss of motor neurons that compensated for the neurons lost during the original infection and not to persistent or reactivated poliovirus infection.

OTHER ENTEROVIRUSES

An estimated 5–10 million cases of symptomatic disease due to enteroviruses other than poliovirus occur in the United States each year. Among neonates, enteroviruses are the most common cause of aseptic meningitis and nonspecific febrile illnesses. Certain clinical syndromes are more likely to be caused by certain serotypes **(Table 191-1)**.

Table 191-1 Manifestations Commonly Associated With Enterovirus Serotypes

	Serotype(s) of Indicated Virus	
Manifestation	**Coxsackievirus**	**Echovirus (E) and Enterovirus (Ent)**
Acute hemorrhagic conjunctivitis	A24	E70
Aseptic meningitis	A2, 4, 7, 9, 10; B1–5	E4, 6, 7, 9, 11, 13, 16, 18, 19, 30, 33; Ent70, 71
Encephalitis	A9; B1–5	E3, 4, 6, 7, 9, 11, 18, 25, 30; Ent71
Exanthem	A4, 5, 9, 10, 16; B1, 3–5	E4–7, 9, 11, 16–19, 25, 30; Ent71
Generalized disease of the newborn	B1–5	E4–6, 7, 9, 11, 14, 16, 18, 19
Hand-foot-and-mouth disease	A5, 7, 9, 10, 16; B1, 2, 5	Ent71
Herpangina	A1–10, 16, 22; B1–5	E6, 9, 11, 16, 17, 25, 30; Ent71
Myocarditis, pericarditis	A4, 9, 16; B1–5	E6, 9, 11, 22
Paralysis	A4, 7, 9; B1–5	E2–4, 6, 7, 9, 11, 18, 30; Ent70, 71
Pleurodynia	A1, 2, 4, 6, 9, 10, 16; B1–6	E1–3, 6, 7, 9, 11, 12, 14, 16, 19, 24, 25, 30
Pneumonia	A9, 16; B1–5	E6, 7, 9, 11, 12, 19, 20, 30; Ent68, 71

Nonspecific Febrile Illness (Summer Grippe)

The most common clinical manifestation of enterovirus infection is a nonspecific febrile illness. After an incubation period of 3–6 days, patients present with an acute onset of fever, malaise, and headache. Occasional cases are associated with upper respiratory symptoms, and some cases include nausea and vomiting. Symptoms often last for 3–4 days, and most cases resolve in a week. While infections with other respiratory viruses occur more often from late fall to early spring, enterovirus febrile illness frequently occurs in the summer and early fall.

Generalized Disease of the Newborn

Most serious enterovirus infections in infants develop during the first week of life, although severe disease can occur up to 3 months of age. Neonates often present with an illness resembling bacterial sepsis, with fever, irritability, and lethargy. Laboratory abnormalities include leukocytosis with a left shift, thrombocytopenia, elevated values in liver function tests, and CSF pleocytosis. The illness can be complicated by myocarditis and hypotension, fulminant hepatitis and disseminated intravascular coagulation, meningitis or meningoencephalitis, or pneumonia. It may be difficult to distinguish neonatal enterovirus infection from bacterial sepsis, although a history of a recent virus-like illness in the mother provides a clue.

Aseptic Meningitis and Encephalitis

In children and young adults, enteroviruses are the cause of up to 90% of cases of aseptic meningitis

in which an etiologic agent can be identified. Patients with aseptic meningitis typically present with an acute onset of fever, chills, headache, photophobia, and pain on eye movement. Nausea and vomiting are also common. Examination reveals meningismus without localizing neurologic signs; drowsiness or irritability may also be apparent. In some cases, a febrile illness may be reported that remits but returns several days later in conjunction with signs of meningitis. Other systemic manifestations may provide clues to an enteroviral cause, including diarrhea, myalgias, rash, pleurodynia, myocarditis, and herpangina. Examination of the CSF invariably reveals pleocytosis; the CSF cell count shows a shift from neutrophil to lymphocyte predominance within 1 day of presentation, and the total cell count does not exceed 1000/μL. The CSF glucose level is usually normal (in contrast to the low CSF glucose level in mumps) with a normal or slightly elevated protein concentration. Partially treated bacterial meningitis may be particularly difficult to exclude in some instances. Enteroviral meningitis is more frequent in summer and fall in temperate climates, while viral meningitis of other etiologies is more common in winter and spring. Symptoms ordinarily resolve within a week, although CSF abnormalities can persist for several weeks. Enteroviral meningitis is often more severe in adults than in children. Neurologic sequelae are rare, and most patients have an excellent prognosis.

Enteroviral encephalitis is much less common than enteroviral aseptic meningitis. Occasional highly inflammatory cases of enteroviral meningitis may be complicated by a mild form of encephalitis that is recognized on the basis of progressive lethargy, disorientation, and sometimes seizures. Less commonly, severe primary encephalitis may develop. An estimated 10–35% of cases of viral encephalitis are due to enteroviruses. Immunocompetent patients generally have a good prognosis.

Patients with hypogammaglobulinemia or agammaglobulinemia or severe combined immunodeficiency may develop chronic meningitis or encephalitis; about half of these patients have a dermatomyositis-like syndrome, with peripheral edema, rash, and myositis. They may also have chronic hepatitis. Patients may develop neurologic disease while receiving immunoglobulin replacement therapy. Echoviruses (especially echovirus 11) are the most common pathogens in this situation.

Paralytic disease due to enteroviruses other than poliovirus occurs sporadically and is usually less severe than poliomyelitis. Most cases are due to enterovirus 70 or 71 or to coxsackievirus A7 or A9. Guillain-Barré syndrome is also associated with enterovirus infection. While some studies have suggested a link between enteroviruses and the chronic fatigue syndrome, most recent studies have not demonstrated such an association.

Pleurodynia (Bornholm Disease)

Patients with pleurodynia present with an acute onset of fever and spasms of pleuritic chest or upper abdominal pain. Chest pain is more common in adults, and abdominal pain is more common in children. Paroxysms of severe, knifelike pain usually last 15–30 min and are associated with diaphoresis and tachypnea. Fever peaks within an hour after the onset of paroxysms and subsides when pain resolves. The involved muscles are tender to palpation, and a pleural rub may be detected. The white blood cell count and chest x-ray are usually normal. Most cases are due to coxsackievirus B and occur during epidemics. Symptoms resolve in a few days, and recurrences are rare. Treatment includes the administration of nonsteroidal anti-inflammatory agents or the application of heat to the affected muscles.

Myocarditis and Pericarditis

Enteroviruses are estimated to cause up to one-third of cases of acute myocarditis. Coxsackievirus B and its RNA have been detected in pericardial fluid and myocardial tissue in some cases of acute myocarditis and pericarditis. Most cases of enteroviral myocarditis or pericarditis occur in newborns, adolescents, or young adults. More than two-thirds of patients are male. Patients often present with an upper respiratory tract infection that is followed by fever, chest pain, dyspnea, arrhythmias, and occasionally heart failure. A pericardial friction rub is documented in half of cases, and the electrocardiogram shows ST-segment elevations or ST- and T-wave abnormalities. Serum levels of myocardial enzymes are often elevated. Neonates commonly have severe disease, while most older children and adults recover completely. Up to 10% of cases progress to chronic dilated cardiomyopathy. Chronic constrictive pericarditis may also be a sequela.

Exanthems

Enterovirus infection is the leading cause of exanthems in children in the summer and fall. While exanthems are associated with many enteroviruses, certain types have been linked to specific syndromes. Echoviruses 9 and 16 have frequently been associated with exanthem and fever. Rashes may be discrete or confluent, beginning on the face and spreading to the trunk and extremities. Echovirus 9 is the most common cause of a rubelliform (discrete) rash. Unlike the rash of rubella, the enteroviral rash occurs in the summer and is not associated with lymphadenopathy. Roseola-like rashes develop after defervescence, with macules and papules on the face and trunk. The Boston exanthem, caused by echovirus 16, is a roseola-like rash. A variety of other rashes have been associated with enteroviruses, including erythema multiforme (see Fig. e7-25) and vesicular, urticarial, petechial, or purpuric lesions. Enanthems also occur, including lesions that resemble the Koplik's spots seen with measles (see Fig. e7-2).

Hand-Foot-and-Mouth Disease

After an incubation period of 4–6 days, patients with hand-foot-and-mouth disease present with fever, anorexia, and malaise; these manifestations are followed by the development of sore throat and vesicles **(Fig. 191-1*A***; see also Fig. e7-23**)** on the buccal mucosa and often on the tongue and then by the appearance of tender vesicular lesions on the dorsum of the hands, sometimes with involvement of the palms. The vesicles may form bullae and quickly ulcerate. About one-third of patients also have lesions on the palate, uvula, or tonsillar pillars, and one-third have a rash on the feet (including the soles) or on the buttocks **(Fig. 191-1*B*)**. The disease is highly infectious, with attack rates of close to 100% among young children. The lesions usually resolve in 1 week. Most cases are due to coxsackievirus A16 or enterovirus 71.

Figure 191-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

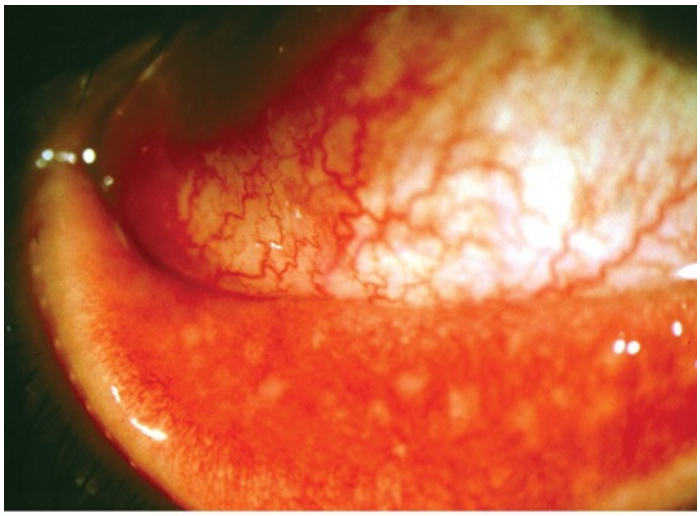

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Tender vesicles in the mouth of a patient with hand-foot-and-mouth disease. ***B.*** Vesicles over the Achilles tendon in a patient with hand-foot-and-mouth disease. ***C.*** Soft-palate lesions of herpangina due to coxsackievirus. ***D.*** Acute hemorrhagic conjunctivitis due to enterovirus 71. *(Images are reprinted with permission from Red Book 2009: Committee on Infectious Diseases, 28th ed. Used with permission of the American Academy of Pediatrics.)*

An epidemic of enterovirus 71 infection in Taiwan in 1998 resulted in thousands of cases of hand-foot-and-mouth disease or herpangina. Severe complications included CNS disease, myocarditis, and pulmonary hemorrhage. About 90% of those who died were children ≥5 years old, and death was associated with pulmonary edema or pulmonary hemorrhage. CNS disease included aseptic meningitis, flaccid paralysis (similar to that seen in poliomyelitis), or rhombencephalitis with myoclonus and tremor or ataxia. The mean age of patients with CNS complications was 2.5 years, and MRI in cases with encephalitis usually showed brain-stem lesions. Follow-up of children at 6 months showed persistent dysphagia, cranial nerve palsies, hypoventilation, limb weakness, and atrophy; at 3 years, persistent neurologic sequelae were documented, with delayed development and impaired cognitive function.

Herpangina

Herpangina is usually caused by coxsackievirus A and presents as acute-onset fever, sore throat, odynophagia, and grayish-white papulovesicular lesions on an erythematous base that ulcerate **(Fig. 191-1*C*)**. The lesions can persist for weeks; are present on the soft palate, anterior pillars of the tonsils, and uvula; and are concentrated in the posterior portion of the mouth. In contrast to herpes stomatitis, enteroviral herpangina is not associated with gingivitis. Acute lymphonodular pharyngitis associated with coxsackievirus A10 presents as white or yellow nodules surrounded by erythema in the posterior oropharynx. The lesions do not ulcerate.

Acute Hemorrhagic Conjunctivitis

Patients with acute hemorrhagic conjunctivitis present with an acute onset of severe eye pain, blurred vision, photophobia, and watery discharge from the eye. Examination reveals edema, chemosis, and

subconjunctival hemorrhage and often shows punctate keratitis and conjunctival follicles as well **(Fig. 191-1*D*)**. Preauricular adenopathy is often found. Epidemics and nosocomial spread have been associated with enterovirus 70 and coxsackievirus A24. Systemic symptoms, including headache and fever, develop in 20% of cases, and recovery is usually complete in 10 days. The sudden onset and short duration of the illness help to distinguish acute hemorrhagic conjunctivitis from other ocular infections, such as those due to adenovirus and *Chlamydia trachomatis*. Paralysis has been associated with some cases of acute hemorrhagic conjunctivitis due to enterovirus 70 during epidemics.

Other Manifestations

Enteroviruses are an infrequent cause of childhood pneumonia and the common cold. Coxsackievirus B has been isolated at autopsy from the pancreas of a few children presenting with type 1 diabetes mellitus; however, most attempts to isolate the virus have been unsuccessful. Other diseases that have been associated with enterovirus infection include parotitis, bronchitis, bronchiolitis, croup, infectious lymphocytosis, polymyositis, acute arthritis, and acute nephritis.

Diagnosis

Isolation of enterovirus in cell culture is the traditional diagnostic procedure. While cultures of stool, nasopharyngeal, or throat samples from patients with enterovirus diseases are often positive, isolation of the virus from these sites does not prove that it is directly associated with disease because these sites are frequently colonized for weeks in patients with subclinical infections. Isolation of virus from the throat is more likely to be associated with disease than isolation from the stool since virus is shed for shorter periods from the throat. Cultures of CSF, serum, fluid from body cavities, or tissues are positive less frequently, but a positive result is indicative of disease caused by enterovirus. In some cases, the virus is isolated only from the blood or only from the CSF; therefore, it is important to culture multiple sites. Cultures are more likely to be positive earlier than later in the course of infection. Most human enteroviruses can be detected within a week after inoculation of cell cultures. Cultures may be negative because of the presence of neutralizing antibody, lack of susceptibility of the cells used, or inappropriate handling of the specimen. Coxsackievirus A may require inoculation into special cell-culture lines or into suckling mice.

Identification of the enterovirus serotype is useful primarily for epidemiologic studies and, with a few exceptions, has little clinical utility. It is important to identify serious infections with enterovirus during epidemics and to distinguish the vaccine strain of poliovirus from the other enteroviruses in the throat or in the feces. Stool and throat samples for culture as well as acute- and convalescent-phase serum specimens should be obtained from all patients with suspected poliomyelitis. In the absence of a positive CSF culture, a positive culture of stool obtained within the first 2 weeks after the onset of symptoms is most often used to confirm the diagnosis of poliomyelitis. If poliovirus infection is suspected, two or more fecal and throat swab samples should be obtained at least 1 day apart and cultured for enterovirus as soon as possible. If poliovirus is isolated, it should be sent to the CDC for identification as either wild-type or vaccine virus.

The polymerase chain reaction (PCR) has been used to amplify viral nucleic acid from CSF, serum, urine, throat swabs, and tissues. A pan-enterovirus PCR assay can detect all human enteroviruses. With the proper controls, PCR of the CSF is highly sensitive (70–100%) and specific (>80%) and is more rapid than culture. PCR of the CSF is less likely to be positive when patients present ≥3 days after the onset of meningitis or with enterovirus 71 infection; in these cases, PCR of throat or rectal swabs—although less specific than PCR of CSF—should be considered.

PCR of serum is also highly sensitive and specific in the diagnosis of disseminated disease. PCR may be particularly helpful for the diagnosis and follow-up of enterovirus disease in immunodeficient patients receiving immunoglobulin therapy, whose viral cultures may be negative. Antigen detection is less sensitive than PCR.

Serologic diagnosis of enterovirus infection is limited by the large number of serotypes and the lack of a common antigen. Demonstration of seroconversion may be useful in rare cases for confirmation of culture results, but serologic testing is usually limited to epidemiologic studies. Serum should be collected and frozen soon after the onset of disease and again ~4 weeks later. Measurement of neutralizing titers is the most accurate method for antibody determination; measurement of complement-fixation titers is usually less sensitive. Titers of virus-specific IgM are elevated in both acute and chronic infection.

Treatment: Enterovirus Infections

Most enterovirus infections are mild and resolve spontaneously; however, intensive supportive care may be needed for cardiac, hepatic, or CNS disease. IV, intrathecal, or intraventricular immunoglobulin has been used with apparent success in some cases for the treatment of chronic enterovirus meningoencephalitis and dermatomyositis in patients with hypogammaglobulinemia or agammaglobulinemia. The disease may stabilize or resolve during therapy; however, some patients decline inexorably despite therapy. IV immunoglobulin often prevents severe enterovirus disease in these patients. IV administration of immunoglobulin with high titers of antibody to the infecting virus has

been used in some cases of life-threatening infection in neonates, who may not have maternally acquired antibody. In one trial involving neonates with enterovirus infections, immunoglobulin containing very high titers of antibody to the infecting virus reduced rates of viremia; however, the study was too small to show a substantial clinical benefit. The level of enteroviral antibodies varies with the immunoglobulin preparation. Although a phase 2 trial of pleconaril for severe neonatal enterovirus disease is in progress, the drug is no longer available on a compassionate-use basis. Glucocorticoids are contraindicated.

Good hand-washing practices and the use of gowns and gloves are important in limiting nosocomial transmission of enteroviruses during epidemics. Enteric precautions are indicated for 7 days after the onset of enterovirus infections.

Prevention and Eradication of Poliovirus

(See also Chap. 122) After a peak of 57,879 cases of poliomyelitis in the United States in 1952, the introduction of inactivated vaccine in 1955 and of oral vaccine in 1961 ultimately eradicated disease due to wild-type poliovirus in the Western Hemisphere. Such disease has not been documented in the United States since 1979, when cases occurred among religious groups who had declined immunization. In the Western Hemisphere, paralysis due to wild-type poliovirus was last documented in 1991.

In 1988, the World Health Organization adopted a resolution to eradicate poliomyelitis by the year 2000. From 1988 to 2001, the number of cases worldwide decreased by >99%, with fewer than 1000 confirmed cases reported in 2001. In 2002, however, there were ~1900 cases of polio, with ~1500 reported in India. Wild-type poliovirus type 2 has not been detected in the world since 1999. The Americas were certified free of indigenous wild-type poliovirus transmission in 1994, the Western Pacific Region in 2000, and the European Region in 2002. The total number of cases worldwide fell to a nadir of 498 in 2001. However, from 2002 to 2005, 21 countries previously free of polio reported cases imported from 6 polio-endemic countries. By 2006, polio transmission had been reduced in most of these 21 countries. In 2009, 1781 cases of polio were reported; 80% were from India, Nigeria, Pakistan, and Afghanistan, the only countries where polio remains endemic (**Table 191-2**). Polio is a source of concern for unimmunized or partially immunized travelers. Importation of poliovirus into 20 countries accounted for 20% of cases in 2009. The number of cases of wild-type polio remained relatively constant from 2005 through 2009, with 1315–1997 cases per year. Countries that reported cases of wild-type polio in 2010 but not in 2009 included Tajikistan, Senegal, and Nepal. Outbreaks of polio in Europe and North America have been traced to cases imported from the Indian subcontinent. Clearly, global eradication of polio is necessary to eliminate the risk of importation of wild-type virus. Outbreaks are thought to have been facilitated by suboptimal rates of vaccination, isolated pockets of unvaccinated children, poor sanitation and crowding, improper vaccine-storage conditions, and a reduced level of response to one of the serotypes in the vaccine. While the global eradication campaign has markedly reduced the number of cases of endemic polio, doubts have been raised as to whether eradication is a realistic goal given the large number of asymptomatic infections and the political instability in developing countries.

Table 191-2 Laboratory-Confirmed Cases of Poliomyelitis in 2009

Country	Type of Transmission	No. of Cases
India	Endemic	752[a]
Nigeria	Endemic	541[b]
Pakistan	Endemic	89
Chad	Imported	66
Sudan	Imported	45
Guinea	Imported	43[c]
Afghanistan	Endemic	38
Angola	Imported	29
Cote d'Ivoire	Imported	26
Others[d]	Imported	142
Others[e]	Vaccine-derived	10
Total		1781

[a]Of these cases, 11 were vaccine-derived.

[b]Of these cases, 153 were vaccine-derived.

[c]Of these cases, 1 was vaccine-derived.

[d]Benin, 20; Kenya, 19; Burkina Faso, 15; Niger, 15; Central African Republic, 14; Mauritania, 13; Liberia, 11; Sierra Leone, 11; Uganda, 8; Togo, 6; Cameroon, 3; Democratic Republic of the Congo, 3; Burundi, 2; Mali, 2.

[e]Democratic Republic of the Congo, 4; Somalia, 4; Ethiopia, 2.

Source: World Health Organization.

The occurrence of outbreaks of poliomyelitis due to circulating vaccine-derived poliovirus of all three types has been increasing, especially in areas with low vaccination rates. In Egypt, 32 cases of vaccine-derived polio occurred in 1983–1993; in the Dominican Republic and Haiti, 22 cases occurred in 2000–2001; in Indonesia, 46 cases were reported in 2005; in Nigeria, 292 cases occurred in 2005–2009; in the Democratic Republic of the Congo, 20 cases were reported in 2005–2009; and fewer cases have occurred in other countries. These OPV-derived viruses reverted to a more neurovirulent phenotype after undetected circulation (probably for >2 years). The epidemic in Hispaniola was rapidly terminated after intensive vaccination with OPV. In 2005, a case of vaccine-derived polio occurred in an unvaccinated U.S. woman returning from a visit to Central and South America. In the same year, an unvaccinated immunocompromised infant in Minnesota was found to be shedding vaccine-derived poliovirus; further investigation identified 4 of 22 infants in the same community who were shedding the virus. All 5 infants were asymptomatic. These outbreaks emphasize the need for maintaining high levels of vaccine coverage and continued surveillance for circulating virus.

IPV is used in most industrialized countries and OPV in most developing countries, including those in which polio still is or recently was endemic. After several doses of OPV alone, the seropositivity rate for individual poliovirus serotypes may still be suboptimal for children in developing countries; one or more supplemental doses of IPV can increase the rate of seropositivity for these serotypes. Against a given serotype, monovalent OPV containing only that serotype is more immunogenic than trivalent vaccine because of a lack of interference from other serotypes. While IM injections of other vaccines (live or attenuated) can be given concurrently with OPV, unnecessary IM injections should be avoided during the first month after vaccination because they increase the risk of vaccine-associated paralysis. Since 1988, an enhanced-potency inactivated poliovirus vaccine has been available in the United States.

OPV and IPV induce antibodies that persist for at least 5 years. Both vaccines induce IgG and IgA antibodies. Compared with recipients of IPV, recipients of OPV shed less virus and less frequently develop reinfection with wild-type virus after exposure to poliovirus. Although IPV is safe and efficacious, OPV offers the advantages of ease of administration, lower cost, and induction of intestinal immunity resulting in a reduction in the risk of community transmission of wild-type virus. Because of progress toward global eradication of polio (with a reduced risk of imported cases) and the continued occurrence of cases of vaccine-associated polio, an all-IPV regimen was recommended in 2000 for childhood poliovirus vaccination in the United States, with vaccine administration at 2, 4, and 6–18 months and 4–6 years of age. The risk of vaccine-associated polio should be discussed before OPV is administered. Recommendations for vaccination of adults are listed in **Table 191-3**.

Table 191-3 Recommendations for Poliovirus Vaccination of Adults

1. Most adults in the United States have been vaccinated during childhood and have little risk of exposure to wild-type virus in the United States. Immunization is recommended for those with a higher risk of exposure than the general population, including:
   a. travelers to areas where poliovirus is or may be epidemic or endemic;
   b. members of communities or population groups with disease caused by wild-type polioviruses;
   c. laboratory workers handling specimens that may contain wild-type polioviruses; and
   d. health care workers in close contact with patients who may be excreting wild-type polioviruses.
2. Three doses of IPV are recommended for adults who need to be immunized. The second dose should be given 1–2 months after the first dose; the third dose should be given 6–12 months after the second dose.
3. Adults who are at increased risk of exposure to wild-type poliovirus and who have previously completed primary immunization should receive a single dose of IPV. Adults who did not complete primary immunization should receive the remaining vaccinations with IPV.

Abbreviation: IPV, inactivated poliovirus vaccine.

Source: Modified from Pickering LK, ed. Red Book 2009: Committee on Infectious Diseases, 28th ed. Used with permission of the American Academy of Pediatrics.

There are concerns about discontinuing vaccination in the event that endemic spread of poliovirus is eliminated. Among the reasons for these concerns are that poliovirus is shed from some immunocompromised persons for >10 years, that vaccine-derived poliovirus can circulate and cause

disease, and that wild-type poliovirus is present in a large number of laboratories. A national survey began in October 2002 to encourage laboratories to dispose of all unneeded wild-type poliovirus materials and to identify laboratories that have wild-type poliovirus or specimens that may contain virus.

REOVIRUSES

Reoviruses are double-stranded RNA viruses encompassing three serotypes. Serologic studies indicate that most humans are infected with reoviruses during childhood. Most infections either are asymptomatic or cause very mild disease. One outbreak of reovirus infection in children resulted in minor upper respiratory tract symptoms. Reovirus is considered a rare cause of mild gastroenteritis in infants and children. Speculation regarding an association of reovirus type 3 with idiopathicneonatal hepatitis and extrahepatic biliary atresia is based on an elevated prevalence of antibody to reovirus among some of these patients and the detection of viral RNA by PCR in hepatobiliary tissues in some studies. Two new orthoreoviruses (Melaka and Kampar viruses) have been associated with fever and acute respiratory disease in Malaysia.

FURTHER READINGS

Alexander JP et al: Transmission of imported vaccine-derived poliovirus in an undervaccinated community in Minnesota. J Infect Dis 199:391, 2009[PMID: 19090774]

Arita I et al: Is polio eradication realistic? Science 312:852, 2006[PMID: 16690846]

Centers for Disease Control and Prevention: Progress toward interruption of wild poliovirus transmission worldwide, 2009. MMWR Morb Mortal Wkly Rep 59:545, 2010

———: Update on vaccine-derived polioviruses—worldwide, January 2008–June 2009. MMWR Morb Mortal Wkly Rep 58:1002, 2009

El-Sayed N et al: Monovalent type 1 oral poliovirus vaccine in newborns. N Engl J Med 359:1655, 2008[PMID: 18923170]

Kew O et al: Vaccine-derived polioviruses and the endgame strategy for global polio eradication. Annu Rev Microbiol 59:587, 2005[PMID: 16153180]

Kupilia L et al: Diagnosis of enteroviral meningitis by use of polymerase chain reaction of cerebrospinal fluid, stool, and serum specimens. Clin Infect Dis 40:982, 2005

Lee TC et al: Diseases caused by enterovirus 71 infection. Pediatr Infect Dis J 28:904, 2009[PMID: 20118685]

Perez-Velez CM et al: Outbreak of neurologic enterovirus type 71 disease: A diagnostic challenge. Clin Infect Dis 45:950, 2007[PMID: 17879907]

Wikswo MW: Increased activity of coxsackievirus B1 strains associated with severe disease among young infants in the United States, 2007–2008. Clin Infect Dis 49:e44, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MEASLES (RUBEOLA): INTRODUCTION

Definition

Measles is a highly contagious viral disease that is characterized by a prodromal illness of fever, cough, coryza, and conjunctivitis followed by the appearance of a generalized maculopapular rash. Before the widespread use of measles vaccines, it was estimated that measles caused between 5 million and 8 million deaths worldwide each year.

Global Considerations

Remarkable progress has been made in reducing global measles incidence and mortality rates through measles vaccination. In the Americas, intensive vaccination and surveillance efforts—based in part on the successful Pan American Health Organization strategy of periodic nationwide measles vaccination campaigns (supplementary immunization activities, or SIAs)—and high routine measles vaccine coverage have interrupted endemic transmission of measles virus. In the United States, high coverage with two doses of measles vaccine eliminated endemic measles virus transmission in 2000. More recently, progress has been made in reducing measles incidence and mortality rates in sub-Saharan Africa as a consequence of increasing routine measles vaccine coverage and provision of a second opportunity for measles vaccination through mass measles vaccination campaigns.

In 2003, the World Health Assembly endorsed a resolution urging member countries to reduce the number of deaths attributed to measles by 50% (compared with 1999 estimates) by the end of 2005. This target was met. Global measles mortality rates were further reduced in 2008; during that year, there were an estimated 164,000 deaths due to measles (uncertainty bounds: 115,000 and 222,000 deaths). These achievements attest to the enormous public-health significance of measles vaccination. The revised global goal, as stated in the Global Immunization Vision and Strategy 2006 –2015 of the World Health Organization and United Nations Children's Fund, is to reduce global measles deaths by 90% (compared with the estimated 757,000 deaths in 2000) by 2010.

Etiology

Measles virus is a spherical, nonsegmented, single-stranded, negative-sense RNA virus and a member of the *Morbillivirus* genus in the family of Paramyxoviridae. Measles was originally a zoonotic infection, arising from cross-species transmission from animals to humans by an ancestral morbillivirus ~10,000 years ago, when human populations attained sufficient size to sustain virus transmission. Although RNA viruses typically have high mutation rates, measles virus is considered to be an antigenically monotypic virus; i.e., the surface proteins responsible for inducing protective immunity have retained their antigenic structure across time and space. The public health significance of this stability is that measles vaccines developed decades ago from a single strain of measles virus remain protective worldwide. Measles virus is killed by ultraviolet light and heat, and attenuated measles vaccine viruses retain these characteristics, necessitating a cold chain for vaccine transport and storage.

Epidemiology

Measles virus is one of the most highly contagious directly transmitted pathogens. Outbreaks can occur in populations in which <10% of persons are susceptible. Chains of transmission are common among household contacts, school-age children, and health care workers. There are no latent or persistent measles virus infections that result in prolonged contagiousness, nor are there animal reservoirs for the virus. Thus, measles virus can be maintained in human populations only by an unbroken chain of acute infections, which requires a continuous supply of susceptible individuals. Newborns become susceptible to measles virus infection when passively acquired maternal antibody is lost and, when not vaccinated, account for the bulk of new susceptible individuals.

Endemic measles has a typical temporal pattern characterized by yearly seasonal epidemics superimposed on longer epidemic cycles of 2–5 years or more. In temperate climates, annual measles outbreaks typically occur in the late winter and early spring. These annual outbreaks are probably attributable to social networks facilitating transmission (e.g., congregation of children at school) and environmental factors favoring the viability and transmission of measles virus. Measles cases continue to occur during interepidemic periods in large populations, but at low incidence. The longer cycles occurring every several years result from the accumulation of susceptible persons over successive birth cohorts and the subsequent decline in the number of susceptibles following an outbreak.

Secondary attack rates in susceptible household and institutional contacts generally exceed 90%. The average age at which measles occurs depends on rates of contact with infected persons, protective maternal antibody decline, and vaccine coverage. In densely populated urban settings with low vaccination coverage, measles is a disease of infants and young children. The cumulative distribution can reach 50% by 1 year of age, with a significant proportion of children acquiring measles before 9 months—the age of routine vaccination in many countries, in line with the schedule recommended by the Expanded Programme on Immunization. As measles vaccine coverage increases or population density decreases, the age distribution shifts toward older children. In such situations, measles cases predominate in school-age children. Infants and young children, although susceptible if not protected by vaccination, are not exposed to measles virus at a rate sufficient to cause a large disease burden in this age group. As vaccination coverage increases further, the age distribution of cases may be shifted into adolescence and adulthood; this distribution is seen in measles outbreaks in the United States and necessitates targeted measles vaccination programs for these older age groups.

Persons with measles are infectious for several days before and after the onset of rash, when levels of measles virus in blood and body fluids are highest and when cough, coryza, and sneezing, which facilitate virus spread, are most severe. The contagiousness of measles before the onset of recognizable disease hinders the effectiveness of quarantine measures. Measles virus can be isolated from urine as late as 1 week after rash onset, and viral shedding by children with impaired cell-mediated immunity can be prolonged.

Medical settings are well-recognized sites of measles virus transmission. Children may present to health care facilities during the prodrome, when the diagnosis is not obvious although the child is infectious and is likely to infect susceptible contacts. Health care workers can acquire measles from infected children and transmit measles virus to others. Nosocomial transmission can be reduced by maintenance of a high index of clinical suspicion, use of appropriate isolation precautions when measles is suspected, administration of measles vaccine to susceptible children and health care workers, and documentation of health care workers' immunity to measles (i.e., proof of receipt of two doses of measles vaccine or detection of antibodies to measles virus).

As efforts at measles control are increasingly successful, public perceptions of the risk of measles as a disease diminish and are replaced by concerns about possible adverse events associated with measles vaccine. As a consequence, numerous measles outbreaks have occurred because of opposition to vaccination on religious or philosophical grounds or unfounded fears of serious adverse events (see "Active Immunization," below).

Pathogenesis

Measles virus is transmitted primarily by respiratory droplets over short distances and, less commonly, by small-particle aerosols that remain suspended in the air for long periods. Airborne transmission appears to be important in certain settings, including schools, physicians' offices, hospitals, and enclosed public places. The virus can be transmitted by direct contact with infected secretions but does not survive for long on fomites.

The incubation period for measles is ~10 days to fever onset and 14 days to rash onset. This period may be shorter in infants and longer (up to 3 weeks) in adults. Infection is initiated when measles virus is deposited on epithelial cells in the respiratory tract, oropharynx, or conjunctivae (Fig. 192-1A). During the first 2–4 days after infection, measles virus proliferates locally in the respiratory mucosa and spreads to draining lymph nodes. Virus then enters the bloodstream in infected leukocytes (primarily monocytes), producing the primary viremia that disseminates infection throughout the reticuloendothelial system. Further replication results in secondary viremia that begins 5–7 days after infection and disseminates measles virus throughout the body. Replication of measles virus in these target organs, together with the host's immune response, is responsible for the signs and symptoms of measles that occur 8–12 days after infection and mark the end of the incubation period.

Figure 192-1

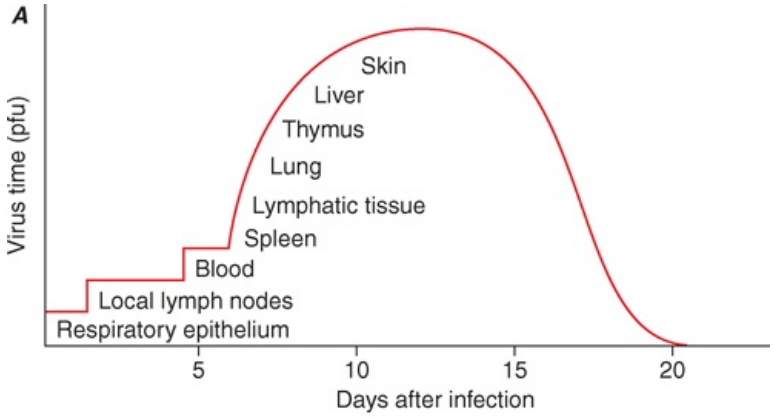

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

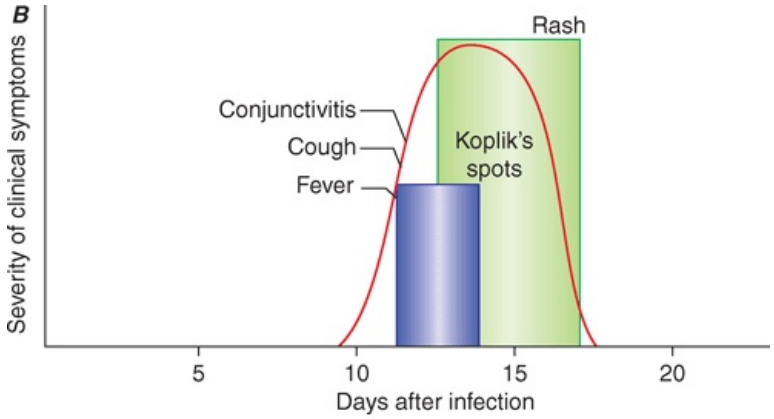

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

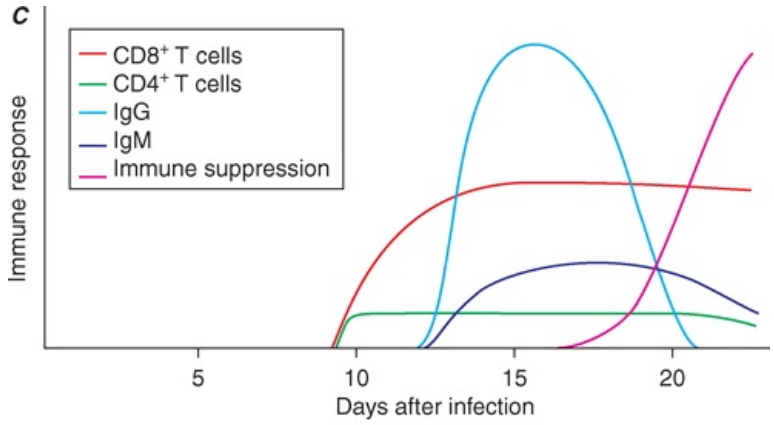

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Measles virus infection: pathogenesis, clinical features, and immune responses. *A*: Spread of measles virus, from initial infection of the respiratory tract through dissemination to the skin. ***B*:** Appearance of clinical signs and symptoms, including Koplik's spots and rash. ***C*:** Antibody and T cell responses to measles virus. The signs and symptoms of measles arise coincident with the host immune response. *(Source: WJ Moss, DE Griffin: Nat Rev Microbiol 4:900, 2006.)*

Immune Responses

Host immune responses to measles virus are essential for viral clearance, clinical recovery, and the establishment of long-term immunity (Fig. 192-1C). Early nonspecific (innate) immune responses during the prodromal phase include activation of natural killer (NK) cells and increased production of the antiviral proteins interferon (IFN) α and IFN-γ. The adaptive immune responses consist of measles virus–specific antibody and cellular responses. The protective efficacy of antibodies to measles virus is illustrated by the immunity conferred to infants from passively acquired maternal antibodies and the protection of exposed, susceptible individuals after administration of anti–measles virus immunoglobulin. The first measles virus–specific antibodies produced after infection are of the IgM subtype, with a subsequent switch to predominantly IgG1 and IgG4 isotypes. The IgM antibody

response is typically absent following reexposure or revaccination and serves as a marker of primary infection.

The importance of cellular immunity to measles virus is demonstrated by the ability of children with agammaglobulinemia (congenital inability to produce antibodies) to recover fully from measles and the contrasting picture for children with severe defects in T lymphocyte function, who often develop severe or fatal disease (Chap. 316). The initial predominant T_H1 response (characterized by IFN-γ) is essential for viral clearance, and the later T_H2 response (characterized by interleukin 4) promotes the development of measles virus–specific antibodies that are critical for protection against reinfection.

The duration of protective immunity following wild-type measles virus infection is generally thought to be lifelong. Immunologic memory to measles virus includes both continued production of measles virus –specific antibodies and circulation of measles virus–specific $CD4^+$ and $CD8^+$ T lymphocytes.

However, the intense immune responses induced by measles virus infection are paradoxically associated with depressed responses to unrelated (nonmeasles virus) antigens, which persist for several weeks to months beyond resolution of the acute illness. This state of immune suppression enhances susceptibility to secondary infections with bacteria and viruses that cause pneumonia and diarrhea and is responsible for a substantial proportion of measles-related morbidity and deaths. Delayed-type hypersensitivity responses to recall antigens, such as tuberculin, are suppressed, and cellular and humoral responses to new antigens are impaired. Reactivation of tuberculosis and remission of autoimmune diseases after measles have been described and are attributed to this state of immune suppression.

Approach to the Patient: Measles

Clinicians should consider measles in persons presenting with fever and generalized rash, particularly when measles virus is known to be circulating or the patient has a history of travel to endemic areas. Appropriate precautions need to be taken to prevent nosocomial transmission. The diagnosis requires laboratory confirmation except during large outbreaks in which an epidemiologic link to a confirmed case can be established. Care is largely supportive and consists of the administration of vitamin A and antibiotics (see "Treatment," below). Complications of measles, including secondary bacterial infections and encephalitis, may occur after acute illness and require careful monitoring, particularly in immunocompromised persons.

Clinical Manifestations

In most persons, the signs and symptoms of measles are highly characteristic (Fig. 192-1B). Fever and malaise beginning ~10 days after exposure are followed by cough, coryza, and conjunctivitis. These signs and symptoms increase in severity over 4 days. Koplik's spots (see Fig. e7-2) develop on the buccal mucosa ~2 days before the rash appears. The characteristic rash of measles (see Fig. e7-3) begins 2 weeks after infection, when the clinical manifestations are most severe, and signal the host's immune response to the replicating virus. Headache, abdominal pain, vomiting, diarrhea, and myalgia may be present.

Koplik's spots (Fig. e7-2) are pathognomonic of measles and consist of bluish white dots ~1 mm in diameter surrounded by erythema. The lesions appear first on the buccal mucosa opposite the lower molars but rapidly increase in number to involve the entire buccal mucosa. They fade with the onset of rash.

The rash of measles begins as erythematous macules behind the ears and on the neck and hairline. The rash progresses to involve the face, trunk, and arms (Fig. e7-3), with involvement of the legs and feet by the end of the second day. Areas of confluent rash appear on the trunk and extremities, and petechiae may be present. The rash fades slowly in the same order of progression as it appeared, usually beginning on the third or fourth day after onset. Resolution of the rash may be followed by desquamation.

Because the characteristic rash of measles is a consequence of the cellular immune response, it may not develop in persons with impaired cellular immunity (e.g., those with AIDS; Chap. 189). These persons have a high case-fatality rate and frequently develop giant-cell pneumonitis caused by measles virus. T lymphocyte defects due to causes other than HIV-1 infection (e.g., cancer chemotherapy) also are associated with increased severity of measles.

A severe atypical measles syndrome was observed in recipients of a formalin-inactivated measles vaccine (used in the United States from 1963 to 1967 and in Canada until 1970) who were subsequently exposed to wild-type measles virus. The atypical rash began on the palms and soles and spread centripetally to the proximal extremities and trunk, sparing the face. The rash was initially erythematous and maculopapular but frequently progressed to vesicular, petechial, or purpuric lesions (see Fig. e7-22).

Differential Diagnosis

The differential diagnosis of measles includes other causes of fever, rash, and conjunctivitis, including rubella, Kawasaki disease, infectious mononucleosis, roseola, scarlet fever, Rocky Mountain spotted

fever, enterovirus or adenovirus infection, and drug sensitivity. Rubella is a milder illness without cough and with distinctive lymphadenopathy. The rash of roseola (exanthem subitum) appears after fever has subsided. The atypical lymphocytosis in infectious mononucleosis contrasts with the leukopenia commonly observed in children with measles.

Diagnosis

Measles is readily diagnosed on clinical grounds by clinicians familiar with the disease, particularly during outbreaks. Koplik's spots (Fig. e7-2) are especially helpful because they appear early and are pathognomonic. Clinical diagnosis is more difficult (1) during the prodromal illness; (2) when the rash is attenuated by passively acquired antibodies or prior immunization; (3) when the rash is absent or delayed in immunocompromised or severely malnourished children with impaired cellular immunity; and (4) in regions where the incidence of measles is low and other pathogens are responsible for the majority of illnesses with fever and rash. The Centers for Disease Control and Prevention case definition for measles requires (1) a generalized maculopapular rash of at least 3 days' duration; (2) fever of at least 38.3^{O}C (101°F); and (3) cough, coryza, or conjunctivitis.

Serology is the most common method of laboratory diagnosis. The detection of measles virus–specific IgM in a single specimen of serum or oral fluid is considered diagnostic of acute infection, as is a fourfold or greater increase in measles virus–specific IgG antibody levels between acute- and convalescent-phase serum specimens. Primary infection in the immunocompetent host results in antibodies that are detectable within 1–3 days of rash onset and reach peak levels in 2–4 weeks. Measles virus–specific IgM antibodies may not be detectable until 4–5 days or more after rash onset and usually fall to undetectable levels within 4–8 weeks of rash onset.

Several methods for measurement of antibodies to measles virus are available. Neutralization tests are sensitive and specific, and the results are highly correlated with protective immunity; however, these tests require propagation of measles virus in cell culture and thus are expensive and laborious. Commercially available enzyme immunoassays are most frequently used. Measles also can be diagnosed by isolation of the virus in cell culture from respiratory secretions, nasopharyngeal or conjunctival swabs, blood, or urine. Direct detection of giant cells in respiratory secretions, urine, or tissue obtained by biopsy provides another method of diagnosis.

For detection of measles virus RNA by reverse-transcriptase polymerase chain reaction (RT-PCR) amplification of RNA extracted from clinical specimens, primers targeted to highly conserved regions of measles virus genes are used. Extremely sensitive and specific, RT-PCR assays may also permit identification and characterization of measles virus genotypes for molecular epidemiologic studies and can distinguish wild-type from vaccine virus strains.

Treatment: Measles

There is no specific antiviral therapy for measles. Treatment consists of general supportive measures, such as hydration and administration of antipyretic agents. Because secondary bacterial infections are a major cause of morbidity and death following measles, effective case management involves prompt antibiotic treatment for patients who have clinical evidence of bacterial infection, including pneumonia and otitis media. *Streptococcus pneumoniae* and *Haemophilus influenzae* type b are common causes of bacterial pneumonia following measles; vaccines against these pathogens probably lower the incidence of secondary bacterial infections following measles.

Vitamin A is effective for the treatment of measles and can markedly reduce rates of morbidity and mortality. The World Health Organization recommends administration of once-daily doses of 200,000 IU of vitamin A for 2 consecutive days to all children with measles who are ≥12 months of age. Lower doses are recommended for younger children: 100,000 IU per day for children 6–12 months of age and 50,000 IU per day for children <6 months old. A third dose is recommended 2–4 weeks later for children with evidence of vitamin A deficiency. While such deficiency is not a widely recognized problem in the United States, many American children with measles do, in fact, have low serum levels of vitamin A, and these children experience increased morbidity following measles. The Committee on Infectious Diseases of the American Academy of Pediatrics recommends that the administration of two consecutive daily doses of vitamin A be considered for children hospitalized with measles and its complications as well as for children with measles who are immunodeficient; who have ophthalmologic evidence of vitamin A deficiency, impaired intestinal absorption, or moderate to severe malnutrition; or who have recently immigrated from areas with high measles mortality rates. Parenteral and oral formulations of vitamin A are available.

Anecdotal reports have described the recovery of previously healthy pregnant and immunocompromised patients with measles pneumonia and of immunocompromised patients with measles encephalitis after treatment with aerosolized and IV ribavirin. However, the clinical benefits of ribavirin in persons with measles have not been conclusively demonstrated in clinical trials.

Complications

Most complications of measles involve the respiratory tract and include the effects of measles virus replication itself and secondary bacterial infections. Acute laryngotracheobronchitis (croup) can occur

during measles and may result in airway obstruction, particularly in young children. Giant-cell pneumonitis due to replication of measles virus in the lungs can develop in immunocompromised children, including those with HIV-1 infection. Many children with measles develop diarrhea, which contributes to malnutrition.

Most complications of measles result from secondary bacterial infections of the respiratory tract that are attributable to a state of immune suppression lasting for several weeks to months after acute measles. Otitis media and bronchopneumonia are most common and may be caused by *S. pneumoniae*, *H. influenzae* type b, or staphylococci. Recurrence of fever or failure of fever to subside with the rash suggests secondary bacterial infection.

Rare but serious complications of measles involve the central nervous system (CNS). Postmeasles encephalomyelitis complicates ~1 in 1000 cases, affecting mainly older children and adults. Encephalomyelitis occurs within 2 weeks of rash onset and is characterized by fever, seizures, and a variety of neurologic abnormalities. The finding of periventricular demyelination, the induction of immune responses to myelin basic protein, and the absence of measles virus in the brain suggest that postmeasles encephalomyelitis is an autoimmune disorder triggered by measles virus infection. Other CNS complications that occur months to years after acute infection are measles inclusion body encephalitis (MIBE) and subacute sclerosing panencephalitis (SSPE). In contrast to postmeasles encephalomyelitis, MIBE and SSPE are caused by persistent measles virus infection. MIBE is a rare but fatal complication that affects individuals with defective cellular immunity and typically occurs months after infection. SSPE is a slowly progressive disease characterized by seizures and progressive deterioration of cognitive and motor functions, with death occurring 5–15 years after measles virus infection. SSPE most often develops in persons infected with measles virus at <2 years of age.

Prognosis

Most persons with measles recover and develop long-term protective immunity to reinfection. Measles case-fatality proportions vary with the average age of infection, the nutritional and immunologic status of the population, measles vaccine coverage, and access to health care. Among previously vaccinated persons who do become infected, disease is less severe and mortality rates are significantly lower. In developed countries, <1 in 1000 children with measles die. In endemic areas of sub-Saharan Africa, the measles case-fatality proportion may be 5–10% or even higher. Measles is a major cause of childhood deaths in refugee camps and in internally displaced populations, where case-fatality proportions have been as high as 20–30%.

Prevention

PASSIVE IMMUNIZATION

Human immunoglobulin given shortly after exposure can attenuate the clinical course of measles. In immunocompetent persons, administration of immunoglobulin within 72 h of exposure usually prevents measles virus infection and almost always prevents clinical measles. Administered up to 6 days after exposure, immunoglobulin will still prevent or modify the disease. Prophylaxis with immunoglobulin is recommended for susceptible household and nosocomial contacts who are at risk of developing severe measles, particularly children <1 year of age, immunocompromised persons (including HIV-infected persons previously immunized with live attenuated measles vaccine), and pregnant women. Except for premature infants, children <6 months of age usually will be partially or completely protected by passively acquired maternal antibody. If measles is diagnosed in a mother, all unimmunized children in the household should receive immunoglobulin. The recommended dose is 0.25 mL/kg given intramuscularly. Immunocompromised persons should receive 0.5 mL/kg. The maximum total dose is 15 mL. IV immunoglobulin contains antibodies to measles virus; the usual dose of 100–400 mg/kg generally provides adequate prophylaxis for measles exposures occurring as long as 3 weeks or more after IV immunoglobulin administration.

ACTIVE IMMUNIZATION

The first live attenuated measles vaccine was developed by passage of the Edmonston strain in chick embryo fibroblasts to produce the Edmonston B virus, which was licensed in 1963 in the United States. Further passage of Edmonston B virus produced the more attenuated Schwarz vaccine that currently serves as the standard in much of the world. The Moraten ("more attenuated") strain, which was licensed in 1968 and is used in the United States, is genetically closely related to the Schwarz strain.

Lyophilized measles vaccines are relatively stable, but reconstituted vaccine rapidly loses potency. Live attenuated measles vaccines are inactivated by light and heat and lose about half their potency at 20°C and almost all their potency at 37°C within 1 h after reconstitution. Therefore, a cold chain must be maintained before and after reconstitution. Antibodies first appear 12–15 days after vaccination and peak at 1–3 months. Measles vaccines are often combined with other live attenuated virus vaccines, such as those for mumps and rubella (MMR) and for mumps, rubella, and varicella (MMR-V).

The recommended age of first vaccination varies from 6 to 15 months and represents a balance between the optimal age for seroconversion and the probability of acquiring measles before that age.

The proportions of children who develop protective levels of antibody after measles vaccination approximate 85% at 9 months of age and 95% at 12 months. Common childhood illnesses concomitant with vaccination may reduce the level of immune response, but such illness is not a valid reason to withhold vaccination. Measles vaccines have been well tolerated and immunogenic in HIV-1-infected children and adults, although antibody levels may wane. Because of the potential severity of wild-type measles virus infection in HIV-1-infected children, routine measles vaccination is recommended except for those who are severely immunocompromised. Measles vaccination is contraindicated in individuals with other severe deficiencies of cellular immunity because of the possibility of disease due to progressive pulmonary or CNS infection with the vaccine virus.

The duration of vaccine-induced immunity is at least several decades if not longer. Rates of secondary vaccine failure 10–15 years after immunization have been estimated at ~5% but are probably lower when the vaccination is given after 12 months of age. Decreasing antibody concentrations do not necessarily imply a complete loss of protective immunity: a secondary immune response usually develops after reexposure to measles virus, with a rapid rise in antibody titers in the absence of overt clinical disease.

Standard doses of currently licensed measles vaccines are safe for immunocompetent children and adults. Fever to 39.4°C (103°F) occurs in ~5% of seronegative vaccine recipients, and 2% of vaccine recipients develop a transient rash. Mild transient thrombocytopenia has been reported, with an incidence of ~1 case per 40,000 doses of MMR vaccine.

Since the publication of a report in 1998 hypothesizing that MMR vaccine may cause a syndrome of autism and intestinal inflammation, much public attention has focused on this purported association. The events that followed publication of this report led to diminished vaccine coverage in the United Kingdom and provide important lessons in the misinterpretation of epidemiologic evidence and the communication of scientific results to the public. The publication that incited the concern was a case series describing 12 children with a regressive developmental disorder and chronic enterocolitis; 9 of these children had autism. In 8 of the 12 cases, the parents associated onset of the developmental delay with MMR vaccination. This simple temporal association was misinterpreted and misrepresented as a possible causal relationship, first by the lead author of the study and then by elements of the media and the public. Subsequently, several comprehensive reviews and additional epidemiologic studies refuted evidence of a causal relationship between MMR vaccination and autism.

Prospects for Measles Eradication

Progress in global measles control has renewed discussion of measles eradication. In contrast to poliovirus eradication, the eradication of measles virus will not entail challenges posed by prolonged shedding of potentially virulent vaccine viruses and environmental viral reservoirs. However, in comparison with smallpox eradication, higher levels of population immunity will be necessary to interrupt measles virus transmission, more highly skilled health care workers will be required to administer measles vaccines, and containment through case detection and ring vaccination will be more difficult for measles virus because of infectivity before rash onset. New tools, such as aerosol administration of measles vaccines, will facilitate mass vaccination campaigns. Despite enormous progress, measles remains a leading vaccine-preventable cause of childhood mortality worldwide and continues to cause outbreaks in communities with low vaccination coverage rates in industrialized nations.

FURTHER READINGS

American Academy of Pediatrics: Measles, in *Red Book: 2009 Report of the Committee on Infectious Diseases*, 28th ed, LK Pickering et al (eds). Elk Grove Village, IL, American Academy of Pediatrics, 2009, pp 444–455

Destefano F, Thompson WW: MMR vaccine and autism: An update of the scientific evidence. Expert Rev Vaccines 3:19, 2004[PMID: 14761240]

Duke T, Mgone CS: Measles: Not just another viral exanthem. Lancet 361:763, 2003[PMID: 12620751]

Feikin DR et al: Individual and community risks of measles and pertussis associated with personal exemptions to immunization. JAMA 284:3145, 2000[PMID: 11135778]

Gerber JS, Offit PA: Vaccines and autism: A tale of shifting hypotheses. Clin Infect Dis 48:456, 2009[PMID: 19128068]

Griffin DE et al: Measles vaccines. Front Biosci 13:1352, 2008[PMID: 17981635]

Moss WJ: Measles control and the prospect of eradication. Curr Top Microbiol Immunol 330:173, 2009[PMID: 19203110]

Omer SB et al: Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. N Engl J Med 360:1981, 2009[PMID: 19420367]

Wolfson LJ et al: Has the 2005 measles mortality reduction goal been achieved? A natural history modelling study. Lancet 369:191, 2007[PMID: 17240285]

World Health Organization: Measles vaccines. Wkly Epidemiol Rec 84:349, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

RUBELLA (GERMAN MEASLES): INTRODUCTION

Rubella was historically viewed as a variant of measles or scarlet fever. Not until the mid-1900s was a separate viral agent for rubella isolated. After an epidemic of rubella in Australia in the early 1940s, the ophthalmologist Norman Gregg noticed the occurrence of congenital cataracts among infants whose mothers had reported rubella infection during early pregnancy, and congenital rubella syndrome (CRS) was first described.

Etiology

Rubella virus is a member of the *Togaviridae* family and the only member of the genus *Rubivirus*. This single-stranded RNA enveloped virus measures 50–70 nm in diameter. Its core protein is surrounded by a single-layer lipoprotein envelope with spike-like projections containing two glycoproteins, E1 and E2. There is only one antigenic type of rubella virus, and humans are its only known reservoir.

Pathogenesis and Pathology

Although the pathogenesis of postnatal (acquired) rubella has been well documented, data on pathology are limited because of the mildness of the disease. Rubella virus is spread from person to person via respiratory droplets. Primary implantation and replication in the nasopharynx are followed by spread to the lymph nodes. Subsequent viremia occurs, which in pregnant women often results in infection of the placenta. Placental virus replication may lead to infection of fetal organs. The pathology of CRS in the infected fetus is well defined, with almost all organs found to be infected; however, the pathogenesis of CRS is only poorly delineated. In tissue, infections with rubella virus have diverse effects, ranging from no obvious impact to cell destruction. The hallmark of fetal infection is chronicity, with persistence throughout fetal development in utero and for up to 1 year after birth.

Individuals with acquired rubella may shed virus from 7 days before rash onset to ~5–7 days thereafter. Both clinical and subclinical infections are considered contagious. Infants with CRS may shed large quantities of virus from bodily secretions, particularly from the throat and in the urine, up to 1 year of age. Outbreaks of rubella, including some in nosocomial settings, have originated with index cases of CRS. Thus only individuals immune to rubella should have contact with infants who have CRS or who are congenitally infected with rubella virus but are not showing signs of CRS.

Epidemiology

The largest recent rubella epidemic in the United States took place in 1964–1965, when an estimated 12.5 million cases occurred, resulting in ~20,000 cases of CRS. Since the introduction of the routine rubella vaccination program in the United States in 1969, the number of rubella cases reported each year has dropped by >99%; the rate of vaccination coverage with rubella-containing vaccine has been >90% among children 19–35 months old since 1995 and >95% for kindergarten and first-grade entrants since 1980. In 1989 a goal for the elimination of rubella and CRS in the United States was set, and in 2004 a panel of experts agreed unanimously that rubella was no longer an endemic disease in this country. The criteria used to document lack of endemic transmission included low disease incidence, high nationwide rubella antibody seroprevalence, outbreaks that were few and contained (i.e., small numbers of cases), and lack of endemic virus transmission (as assessed by genetic sequencing). In the United States, interruption of endemic transmission of rubella virus has been sustained since 2001.

Although rubella and CRS are no longer endemic in the United States, they remain important public health problems globally. The number of rubella cases reported worldwide in 1999 was ~900,000; this figure declined steadily to 165,000 in 2007. However, numbers of rubella cases are substantially underestimated because cases in many countries are identified through measles surveillance systems that are not specific for rubella. In developing countries, an estimated 110,000 cases of CRS occur during nonepidemic years.

Clinical Features

ACQUIRED RUBELLA

Acquired rubella is characterized by a generalized maculopapular rash that usually lasts for up to 3 days **(Fig. 193-1)**. Up to 50% of rubella virus infections may be subclinical or inapparent. The rash is usually mild and may be difficult to detect in persons with darker skin. In children, rash is usually the first sign of illness. However, in older children and adults, a 1- to 5-day prodrome often precedes the

rash and may include low-grade fever, malaise, and upper respiratory symptoms. The incubation period is 14 days (range, 12–23 days).

Lymphadenopathy, particularly occipital and postauricular, may be noted during the second week after exposure. Although acquired rubella is usually thought of as a benign disease, arthralgia and arthritis are common in infected adults, particularly women. Thrombocytopenia and encephalitis are less common complications.

Figure 193-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mild maculopapular rash of rubella in a child.

CONGENITAL RUBELLA SYNDROME

The most serious consequence of rubella virus infection can develop when a woman becomes infected during pregnancy, particularly during the first trimester. The resulting complications may include miscarriage, fetal death, premature delivery, or live birth with congenital defects. Infants infected with rubella virus in utero may have a myriad of physical defects **(Table 193-1)**, which most commonly relate to the eyes, ears, and heart. This constellation of severe birth defects is known as *congenital rubella syndrome*. In addition to permanent manifestations, there are a host of transient physical manifestations, including thrombocytopenia with purpura/petechiae (e.g., dermal erythropoiesis, "blueberry muffin syndrome"). Some infants may be born with congenital rubella virus infection but have no apparent signs or symptoms of CRS and are referred to as infants with congenital rubella infection only.

Table 193-1 Common Transient and Permanent Manifestations in Infants With Congenital Rubella Syndrome

Transient Manifestations	Permanent Manifestations
Hepatosplenomegaly	Hearing impairment/deafness
Interstitial pneumonitis	Congenital heart defects (patent ductus arteriosus, pulmonary arterial stenosis)
Thrombocytopenia with purpura/petechiae (e.g., dermal erythropoiesis, or "blueberry muffin syndrome") Hemolytic anemia Bony radiolucencies	Eye defects (cataracts, cloudy cornea, microphthalmos, pigmentary retinopathy, congenital glaucoma)
Intrauterine growth retardation	Microcephaly
Adenopathy	

Meningoencephalitis	Central nervous system sequelae (mental and motor delay, autism)

Diagnosis

ACQUIRED RUBELLA

Clinical diagnosis of acquired rubella is difficult because of the mimicry of many illnesses with rashes, the varied clinical presentations, and the high rates of subclinical and mild disease. Illnesses that may be similar to rubella in presentation include scarlet fever, roseola, toxoplasmosis, fifth disease, measles, and illnesses with suboccipital and postauricular lymphadenopathy. Thus laboratory documentation of rubella virus infection is considered the only reliable way to confirm acute disease.

Laboratory assessment of rubella infection is conducted by serologic and virologic methods. For acquired rubella, serologic diagnosis is most common and depends on the demonstration of IgM antibodies in an acute-phase serum specimen or a fourfold rise in IgG antibody titer between acute- and convalescent-phase specimens. The enzyme-linked immunosorbent assay IgM capture technique is considered most accurate for serologic diagnosis, but the indirect IgM assay is also acceptable. After rubella virus infection, IgM antibody may be detectable for up to 6 weeks. In case of a negative result for IgM in specimens taken earlier than day 5 after rash onset, serologic testing should be repeated. Although uncommon, reinfection with rubella virus is possible, and IgM antibodies may be present. To detect a rise in IgG antibody titer indicative of acute disease, the acute-phase serum specimen should be collected within 7–10 days after onset of illness and the convalescent-phase specimen ~14–21 days after the first specimen.

IgG avidity testing is used in conjunction with IgG testing. Low-avidity antibodies indicate recent infection. Mature (high-avidity) IgG antibodies most likely indicate an infection occurring at least 2 months previously. This test helps distinguish primary infection from reinfection.

Rubella virus can be isolated from the blood and nasopharynx during the prodromal period and for as long as 2 weeks after rash onset. However, as the secretion of virus in individuals with acquired rubella is maximal just before or up to 4 days after rash onset, this is the optimal time frame for collecting specimens for viral cultures. Rubella RNA detection by reverse-transcriptase polymerase chain reaction (RT-PCR) is a more recently developed technique for rubella diagnosis.

CONGENITAL RUBELLA SYNDROME

A clinical diagnosis of CRS is reasonable when an infant presents with a combination of cataracts, hearing impairment, and heart defects; this pattern is seen in ~10% of infants with CRS. However, as with acquired rubella, laboratory diagnosis of congenital infection is highly recommended, particularly because most features of the clinical presentation are nonspecific and may be associated with other intrauterine infections. Early diagnosis of CRS facilitates appropriate medical intervention for specific disabilities and prompts implementation of infection control measures.

Diagnostic tests used to confirm CRS include serologic assays and virus isolation. In an infant with congenital infection, serum IgM antibodies may be present for up to 1 year after birth. In some instances, IgM may not be detectable until 1 month of age; thus infants who have symptoms consistent with CRS but who test negative shortly after birth should be retested at 1 month. A rubella serum IgG titer persisting beyond the time expected after passive transfer of maternal IgG antibody (i.e., a rubella titer that does not decline at the expected rate of a twofold dilution per month) is another serologic criterion used to confirm CRS.

In congenital infection, rubella virus is isolated most commonly from throat swabs and less commonly from urine and cerebrospinal fluid. Infants with congenital rubella may excrete virus for up to 1 year, but specimens for virus isolation are most likely to be positive if obtained within the first 6 months after birth. Rubella virus in infants with CRS can also be detected by RT-PCR.

RUBELLA DIAGNOSIS IN PREGNANT WOMEN

In the United States, screening for rubella IgG antibodies is recommended as part of routine prenatal care. Pregnant women with a positive IgG antibody serologic test are considered immune. Susceptible pregnant women should be vaccinated postpartum.

A susceptible pregnant woman exposed to rubella virus should be tested for IgM antibodies and a fourfold rise in IgG antibody titer between acute- and convalescent-phase serum specimens to determine whether she was infected during pregnancy. Pregnant women with evidence of acute infection must be clinically monitored, and gestational age at the time of maternal infection must be determined to assess the possibility of risk to the fetus. Of women infected with rubella virus during the first 11 weeks of gestation, up to 90% deliver an infant with CRS; for maternal infection during the first 20 weeks of pregnancy, the CRS rate is 20%.

Treatment: Rubella

No specific therapy is available for rubella virus infection. Symptom-based treatment for various manifestations, such as fever and arthralgia, is appropriate. Immunoglobulin does not prevent rubella

virus infection after exposure and therefore is not recommended as routine postexposure prophylaxis. Although immunoglobulin may modify or suppress symptoms, it can create an unwarranted sense of security: infants with congenital rubella have been born to women who received immunoglobulin shortly after exposure. Administration of immunoglobulin should be considered only if a pregnant woman who has been exposed to rubella will not consider termination of pregnancy under any circumstances. In such cases, IM administration of 20 mL of immunoglobulin within 72 h of rubella exposure may reduce—but does not eliminate—the risk of rubella.

Prevention

After the isolation of rubella virus in the early 1960s and the occurrence of a devastating pandemic, a vaccine for rubella was developed and licensed in 1969. Currently, the majority of rubella-containing vaccines (RCVs) used worldwide are combined measles and rubella (MR) or measles, mumps, and rubella (MMR) formulations. A tetravalent measles, mumps, rubella, and varicella (MMRV) vaccine is available but is not widely used.

The public health burden of rubella infection is measured primarily through the resulting CRS cases. The 1964–1965 rubella epidemic in the United States encompassed >30,000 infections during pregnancy. CRS occurred in ~20,000 infants born alive, including >11,000 infants who were deaf, >3500 infants who were blind, and almost 2000 infants who were mentally retarded. The cost of the epidemic exceeded $1.5 billion. In 1982, it was estimated that the cost per child with CRS exceeded $200,000.

In most countries, there is little documented evidence to illuminate the epidemiology of CRS. Clusters of CRS cases have been reported in developing countries, and modeling studies have shown that, before the introduction of an immunization program, the incidence of CRS is 0.1–0.2 per 1000 live births during endemic periods and 1–4 per 1000 live births during epidemic periods. Where rubella virus is circulating and women of childbearing age are susceptible, CRS cases will continue to occur.

The most effective method of preventing acquired rubella and CRS is through vaccination with an RCV. One dose induces seroconversion in ≥95% of persons >1 year of age. Immunity is considered long-term and is probably lifelong. The most commonly used vaccine globally is derived from the RA27/3 virus strain. The current recommendation for routine rubella vaccination in the United States is a first dose of MMR vaccine at 12–15 months of age and a second dose at 4–6 years. Target groups for rubella vaccine in all countries include children >1 year of age, adolescents and adults without documented evidence of immunity, individuals in congregate settings (e.g., college students, military personnel, child care and health care workers), and susceptible women before and after pregnancy.

Because of the theoretical risk of transmission of live attenuated rubella vaccine virus to the developing fetus, women known to be pregnant should not receive an RCV. In addition, pregnancy should be avoided for 28 days after receipt of an RCV. In follow-up studies of 680 unknowingly pregnant women who received rubella vaccine, no infant was born with CRS. Receipt of an RCV during pregnancy is not ordinarily a reason to consider termination of the pregnancy.

As of 2008, 127 (66%) of the 193 WHO member countries recommended inclusion of an RCV in the routine childhood vaccination schedule **(Fig. 193-2)**. Vaccination coverage varies widely among the member countries, with the European and American Regions reporting coverage of >90%. Goals for control or elimination of rubella and CRS have been established in the American Region, the European Region, and the Western Pacific Region. The other three regions (Eastern Mediterranean, South East Asian, and African) have not yet set such goals.

Figure 193-2

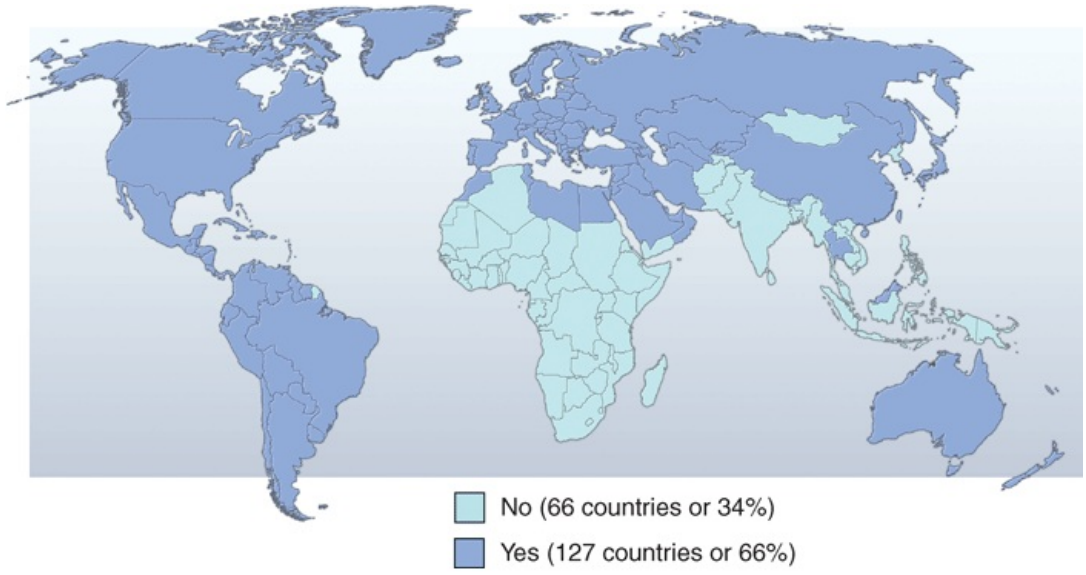

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Countries using rubella vaccine in National Immunization Schedule, 2008. *(From the World Health Organization.)*

FURTHER READINGS

Centers for Disease Control and Prevention: Control and prevention of rubella: Evaluation and management of suspected outbreaks, rubella in pregnant women, and surveillance for congenital rubella syndrome. MMWR Morb Mortal Wkly Rep 50:1, 2001

———: Notice to readers: Revised ACIP recommendation for avoiding pregnancy after receiving a rubella-containing vaccine. MMWR Morb Mortal Wkly Rep 50:1117, 2001

———: Rubella, in *Manual for the Surveillance of Vaccine-Preventable Diseases* , 4th ed, SW Roush et al (eds). Atlanta, Centers for Disease Control and Prevention, 2008, Chapter 14 (http://www.cdc.gov/vaccines/pubs/surv-manual/)

———: Rubella, in *Epidemiology and Prevention of Vaccine Preventable Diseases* , 11th ed, W Atkinson et al (eds). Washington, DC, Public Health Foundation, 2009, Chapter 18 (order information at http://www.cdc.gov/vaccines/pubs/pinkbook/default.htm)

Gershon A: Rubella (German measles), in *Harrison's Principles of Internal Medicine* , 17th ed, AS Fauci et al (eds). New York, McGraw Hill, 2009, pp 1217–1220

Papania M et al: Nosocomial measles, mumps, rubella, and other viral infections, in *Hospital Epidemiology and Infection Control*, 3rd ed, CG Mayhall (ed). Philadelphia, Lippincott Williams and Wilkins, 2004, pp 828–849

Plotkin SA, Reef S: Rubella vaccine, in *Vaccines*, SA Plotkin and WA Orenstein (eds). Philadelphia, Saunders, 2008,pp 733–771

World Health Organization: Rubella vaccines: WHO position paper. Wkly Epidemiol Rec 75:161, 2000

———: Rubella, module 11, in *The Immunological Basis for Immunization Series* . Geneva, WHO, 2009 (http://www.who.int/immunization/documents/ISBN9789241596848/en/index.html)

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MUMPS: INTRODUCTION

Definition

Mumps is an acute, systemic viral infection classically associated with swelling of one or both parotid glands.

Etiologic Agent

Mumps is caused by a paramyxovirus with a negative-strand nonsegmented RNA genome of 15,384 bases encoding nine proteins. The nucleoprotein, phosphoprotein, and polymerase protein participate in viral replication and, together with genomic RNA, form the ribonucleocapsid. The ribonucleocapsid is surrounded by a host-derived lipid bilayer envelope containing the viral hemagglutinin-neuraminidase (HN) and fusion (F) proteins, which are responsible for cell binding by and entry of the virus and are major targets of virus-neutralizing antibodies. The functions of the other virus proteins (small-hydrophobic, matrix, V, and I) are less well understood. The small-hydrophobic gene sequence is highly variable and forms the basis for the 13 genotypes (A through M) used mainly for molecular epidemiologic purposes.

Epidemiology

Mumps is endemic worldwide, with epidemics occurring every 3–5 years in unvaccinated populations. The estimated annual global incidence is 100–1000 cases per 100,000 population in countries without national mumps vaccination programs, where virtually the entire population has been infected by adulthood. Following the 1967 introduction of mumps vaccine in the United States, the reported number of cases declined; by 2001, this number had decreased from >150,000 to <300—a 99.8% reduction from prevaccine levels. In 2006, the United States experienced its largest mumps outbreak in more than 20 years, with 6584 reported cases. This outbreak was preceded by outbreaks in the United Kingdom (2004–2005) and followed by outbreaks in Canada. Compelling epidemiologic evidence links the genotype G virus to the outbreaks in all three countries. The majority of cases occurred in college students 18–23 years of age, most of whom had been vaccinated in early childhood. These outbreaks are probably the result of several coincident circumstances, including (1) situations promoting the spread of respiratory viruses among young adults (e.g., residence in college dormitories), (2) waning of vaccine immunity with time, (3) lack of endemically circulating wild-type virus to periodically boost vaccine-induced immune responses, and (4) continuing global epidemics of mumps (due either to lack of mumps vaccination programs or to low rates of mumps vaccination where such programs do exist). Whereas in the pre- and early postvaccine era mumps was historically a disease of childhood, the majority of U.S. cases now occur in previously vaccinated young adults.

Pathogenesis

Humans are the only natural hosts for mumps virus infection. The incubation period of mumps is ~19 days (range, 7–23 days). The virus is transmitted by the respiratory route via droplets, saliva, and fomites. Mumps virus is typically shed from 1 week before to 1 week after symptom onset, although this window appears to be narrower in vaccinated individuals. Persons are most contagious 1–2 days before onset of clinical symptoms. Primary replication occurs in the nasal mucosa or upper respiratory mucosal epithelium. Mononuclear cells and cells within regional lymph nodes can become infected; such infection facilitates the development of viremia and poses a risk for a wide array of acute inflammatory reactions. Classic sites of mumps virus replication include the salivary glands, testes, pancreas, ovaries, mammary glands, and central nervous system (CNS).

Little is known of the pathology of mumps since the disease is rarely fatal. The virus replicates well in glandular epithelium, but classic parotitis is not a necessary component of mumps infection. Affected glands contain perivascular and interstitial mononuclear cell infiltrates and exhibit hemorrhage with prominent edema. Necrosis of acinar and epithelial duct cells is evident in the salivary glands and in the germinal epithelium of the seminiferous tubules of the testes. The virus probably enters cerebrospinal fluid (CSF) through the choroid plexus or via transiting mononuclear cells during plasma viremia. Although relevant data are limited, typical mumps encephalitis appears to be secondary to respiratory spread and is probably a parainfectious process, as suggested by perivenous demyelination, perivascular mononuclear cell inflammation, and relative sparing of neurons. Although rare, presumed primary encephalitis has been associated with mumps virus isolation from brain tissue. Evidence of placental and intrauterine spread in pregnancy has been found in both early and late gestation.

Clinical Manifestations

Up to half of mumps virus infections are asymptomatic or lead to nonspecific respiratory symptoms. Inapparent infections are more common in adults than in children. The prodrome of mumps consists of low-grade fever, malaise, myalgia, headache, and anorexia. Mumps parotitis—acute-onset unilateral or bilateral swelling of the parotid or other salivary glands lasting >2 days without another apparent cause—develops in 70–90% of symptomatic infections, usually within 24 h of prodromal symptoms but sometimes as long as 1 week thereafter. Parotitis is generally bilateral, although the two sides may not be involved synchronously. Unilateral involvement is documented in about one-third of cases. Swelling of the parotid is accompanied by tenderness and obliteration of the space between the earlobe and the angle of the mandible **(Figs. 194-1** and **194-2)**. The patient frequently reports an earache and finds it difficult to eat, swallow, or talk. The orifice of Stensen's duct is commonly red and swollen. The submaxillary and sublingual glands are involved less often than the parotid gland and are almost never involved alone. Glandular swelling increases for a few days and then gradually subsides, disappearing within 1 week. Recurrent sialadenitis is a rare sequela of mumps parotitis. In ~6% of mumps cases, obstruction of lymphatic drainage secondary to bilateral salivary gland swelling may lead to presternal pitting edema, associated often with submandibular adenitis and rarely with the more life-threatening supraglottic edema.

Figure 194-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Child with mumps. Note the classic submandibular and preauricular enlargement of the parotid gland. *(From the Centers for Disease Control and Prevention.)*

Figure 194-2

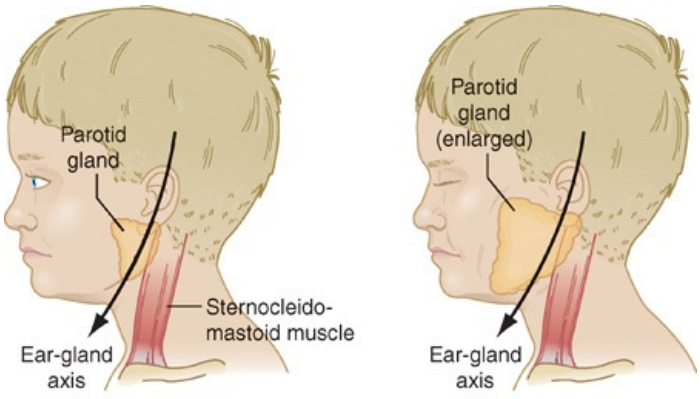

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic drawing of a parotid gland infected with mumps virus (*right*) compared with a normal gland (*left*). An enlarged cervical lymph node is usually posterior to the imaginary line. *(Reprinted*

with permission from Gershon A et al: Mumps, in Krugman's Infectious Diseases of Children, *11th ed. Philadelphia, Elsevier, 2004, p 392.)*

Epididymo-orchitis is the next most common manifestation of mumps, developing in 15–30% of cases in postpubertal males, with bilateral involvement in 10–30% of those cases. Orchitis, accompanied by fever, typically occurs during the first week of parotitis but can develop up to 6 weeks after parotitis or in its absence. The testis is painful and tender and can be enlarged to several times its normal size; this condition usually resolves within 1 week. Testicular atrophy develops in one-half of affected men. Sterility after mumps is rare, although subfertility is estimated to occur in 13% of cases of unilateral orchitis and in 30–87% of cases of bilateral orchitis. Oophoritis occurs in ~5% of women with mumps and may be associated with lower abdominal pain and vomiting but has only rarely been associated with sterility or premature menopause. Mumps infection in postpubertal women may also present with mastitis.

Documented CSF pleocytosis indicates that mumps virus invades the CNS in ~50% of cases; however, symptomatic CNS disease, typically in the form of aseptic meningitis, occurs in <10% of cases, with a male predominance. CNS symptoms of aseptic meningitis (e.g., stiff neck, headache, and drowsiness) appear ~5 days after parotitis and also occur often in the absence of parotid involvement. Within the first 24 h polymorphonuclear leukocytes may predominate in CSF (1000–2000 cells/μL), but by the second day nearly all the cells are lymphocytes. The glucose level in CSF may be low and the protein concentration high, a pattern reminiscent of bacterial meningitis. Mumps meningitis is a self-limited manifestation without significant risk of death or long-term sequelae. Cranial nerve palsies have occasionally led to permanent sequelae, particularly deafness. The reported incidence of mumps-associated hearing loss varies between 1 in 1000 and 1 in 100,000. In ~0.1% of infections, mumps virus may cause encephalitis, which presents as high fever with marked changes in the level of consciousness, seizures, and focal neurologic symptoms. Electroencephalographic abnormalities may be seen. Permanent sequelae are sometimes identified in survivors, and adult infections more commonly have poor outcomes than do pediatric infections. The mortality rate associated with mumps encephalitis is ~1.5%. Other CNS problems occasionally associated with mumps include cerebellar ataxia, facial palsy, transverse myelitis, hydrocephalus, Guillain-Barré syndrome, flaccid paralysis, and behavioral changes.

Mumps pancreatitis, which may present as abdominal pain, occurs in ~4% of infections but is difficult to diagnose because an elevated serum amylase level can be associated with either parotitis or pancreatitis. An etiologic association of mumps virus and juvenile diabetes mellitus remains controversial. Myocarditis and endocardial fibroelastosis are rare and self-limited but may represent severe complications of mumps infection; however, mumps-associated electrocardiographic abnormalities have been reported in up to 15% of cases. Other unusual complications include thyroiditis, nephritis, arthritis, hepatic disease, keratouveitis, and thrombocytopenic purpura. Abnormal renal function is common, but severe, life-threatening nephritis is rare. It remains at issue whether an excessive number of spontaneous abortions are associated with gestational mumps. Mumps in pregnancy does not appear to lead to premature birth, low birth weight, or fetal malformations.

Differential Diagnosis

During a mumps outbreak, the diagnosis is made easily in patients with parotitis and a history of recent exposure; however, when disease incidence is low, other causes of parotitis should be considered and laboratory testing is required for case confirmation. Infectious causes of parotitis include other viruses (e.g., HIV, coxsackievirus, parainfluenza virus type 3, influenza A virus, Epstein-Barr virus, adenovirus, parvovirus B19, lymphocytic choriomeningitis virus, human herpesvirus 6), gram-positive bacteria, atypical mycobacteria, and *Bartonella* species. Rarely, other gram-negative or anaerobic bacteria are associated with parotitis. Parotitis can also develop in the setting of sarcoidosis, Sjögren's syndrome, Mikulicz's syndrome, Parinaud's syndrome, uremia, diabetes mellitus, laundry starch ingestion, malnutrition, cirrhosis, and some drug treatments. Unilateral parotitis can be caused by ductal obstruction, cysts, and tumors. In the absence of parotitis or other salivary gland enlargement, symptoms of other visceral organ and/or CNS involvement may predominate, and a laboratory diagnosis is required. Other entities should be considered when manifestations consistent with mumps appear in organs other than the parotid. Testicular torsion may produce a painful scrotal mass resembling that seen in mumps orchitis. Other viruses (e.g., enteroviruses) may cause aseptic meningitis that is clinically indistinguishable from that due to mumps virus.

Laboratory Diagnosis

Laboratory diagnosis is based on detection of viral antigens or RNA or on serology. Viral antigens may be detected by mumps-specific immunofluorescent staining of clinical specimens either directly or, more commonly, after incubation of clinical samples with cell cultures. Most commonly, mumps virus is detected as viral RNA by reverse-transcription polymerase chain reaction (PCR), which is believed to be more rapid, sensitive, and specific than detection of live virus by immunofluorescence assays; however, false-negative results are not uncommon with either method. Virus is typically assayed in material obtained by throat swab, although it has been detected in CSF, urine, and seminal fluid. Despite the apparent frequency of viremia, mumps virus has only rarely been isolated from blood,

possibly because of the presence of specific antibodies. In urine samples, presumed PCR inhibitors reduce the relative value of RNA testing methods.

Serologic diagnosis of mumps (i.e., a positive IgM response or a significant increase in IgG titer in paired acute- and convalescent-phase sera) is typically obtained by enzyme-linked immunosorbent assay (ELISA). Serologic diagnoses are now of limited value: IgM is detected in <20% of cases in immunized individuals, and IgG titers in convalescent-phase sera may be only nominally greater than those in acute-phase sera. Thus, at present, the capacity of RNA or viral antigen detection to confirm cases is much greater than that of serology. Traditional and labor-intensive serologic tests such as complement fixation, hemagglutination inhibition, and virus neutralization are now performed only rarely. The main downside to replacement of these functional serologic assays with the more rapid ELISA method is the latter's detection of all virus-specific antibodies, including those that are nonneutralizing (i.e., nonprotective). Thus, an individual may be seropositive by ELISA but may lack protective levels of antibody. While there is a strong association between the presence of mumps virus neutralizing antibody and protection from disease, an absolute antibody titer predictive of serologic protection is lacking; in this respect, mumps differs from other respiratory infections, such as measles.

Prevention

Vaccination is the only practical control measure; in the United States, the cost-benefit ratios for mumps vaccination alone are >13 for direct costs (e.g., medical expenses) and >24 for societal costs (including productivity losses for patients and caregivers). Several mumps virus vaccines are used throughout the world; in the United States, only the live attenuated Jeryl Lynn strain is used. Current recommendations are that mumps vaccine be administered as part of the combined trivalent measles-mumps-rubella vaccine (MMR-II®) or the quadrivalent measles-mumps-rubella-varicella vaccine (ProQuad®). Monovalent vaccine (MumpsVax®) is not generally available.

Before administering mumps-containing vaccine, physicians should always consult the latest recommendations from the Advisory Committee on Immunization Practices (ACIP). Current recommendations for children specify two doses of mumps-containing vaccine: the first dose on or after the first birthday and the second dose administered no earlier than 1 month after the first. In the United States, children often receive the second dose between the ages of 4 and 6 years.

In 2009, the ACIP revised its recommendations for evidence of mumps immunity in health care personnel to include (1) documented administration of two doses of a preparation containing live mumps vaccine, (2) laboratory evidence of immunity or laboratory confirmation of disease, or (3) birth date before 1957. For unvaccinated health care personnel born before 1957 who lack laboratory evidence of mumps immunity or laboratory confirmation of mumps, health care facilities should consider two doses of MMR vaccine at the appropriate interval; during a mumps outbreak, vaccination of these individuals is recommended.

Mumps vaccine contains live attenuated virus. It is not recommended for pregnant women, for individuals who have had a life-threatening allergic reaction to components of the vaccine, or for people in settings of clinically significant primary or secondary immunosuppression. (For details, see the ACIP guidelines on the website of the Centers for Disease Control and Prevention.) Occasionally, febrile reactions and parotitis have been reported soon after mumps vaccination. Allergic reactions after vaccination (e.g., rash and pruritus) are uncommon and are usually mild and self-limited. More serious complications, such as aseptic meningitis, have been causally associated with certain vaccine strains but not with the Jeryl Lynn strain.

Immunity to mumps is associated with the development of neutralizing antibody, although a specific correlate of protection has not been established. Seroconversion occurs in ~95% of recipients of the Jeryl Lynn strain; however, vaccineefficacy is ~80% for one dose and 90% for two doses. Recent data indicate declining seropositivity rates with time since vaccination. Although it is generally accepted that mumps virus is serologically monotypic, antigenic differences between virus isolates have been detected. It is unclear whether such differences can lead to immune escape. The role of the cellular arm of the immune response is unclear, but there is evidence that it may help limit virus spread and complications.

Treatment: Mumps

Mumps is generally a benign, self-resolving illness. Therapy for parotitis and other clinical manifestations is symptom based and supportive. The administration of analgesics and the application of warm or cold compresses to the parotid area may be helpful. Testicular pain may be minimized by the local application of cold compresses and gentle support for the scrotum. Anesthetic blocks may also be used. Neither the administration of glucocorticoids nor incision of the tunica albuginea is of proven value in severe orchitis. Anecdotal information on a small number of patients with orchitis suggests that subcutaneous administration of interferon α2b may help preserve the organ and fertility. Lumbar puncture is occasionally performed to relieve headache associated with meningitis. Mumps immune globulin has not been consistently shown to be effective in preventing mumps and is not recommended for treatment or postexposure prophylaxis.

ACKNOWLEDGMENT

The authors thank and acknowledge Dr. Anne Gershon, the author of this chapter in earlier editions of this book.

FURTHER READINGS

Davidkin I et al: Etiology of mumps-like illnesses in children and adolescents vaccinated for measles, mumps, and rubella. J Infect Dis 191:719, 2005[PMID: 15688285]

Huang AS et al: Risk factors for mumps at a university with a large mumps outbreak. Public Health Rep 124:419, 2009[PMID: 19445418]

Hviid A et al: Mumps. Lancet 371:932, 2008[PMID: 18342688]

Kutty PK et al: Guidance for isolation precautions for mumps in the United States: A review of the scientific basis of policy change. Clin Infect Dis 50:1619, 2010[PMID: 20455692]

LeBaron CW et al: Persistence of mumps antibodies after 2 doses of measles-mumps-rubella vaccine. J Infect Dis 199:552, 2009[PMID: 19113988]

Marin M et al: Use of combination measles, mumps, rubella, and varicella vaccine: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 59:1, 2010[PMID: 20448530]

Masarani M et al: Mumps orchitis. J R Soc Med 99:573, 2006[PMID: 17082302]

Polgreen PM et al: The duration of mumps virus shedding after the onset of symptoms. Clin Infect Dis 46:1447, 2008[PMID: 18419451]

Reid F et al: Epidemiologic and diagnostic evaluation of a recent mumps outbreak using oral fluid samples. J Clin Virol 41:134, 2008[PMID: 18354822]

Rota JS et al: Investigation of a mumps outbreak among university students with two measles-mumps-rubella (MMR) vaccinations, Virginia, September–December 2006. J Med Virol 81:1819, 2009[PMID: 19697404]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

RABIES

Rabies is a rapidly progressive, acute infectious disease of the central nervous system (CNS) in humans and animals that is caused by infection with rabies virus. The infection is normally transmitted from animal vectors. Rabies has encephalitic and paralytic forms that progress to death.

Etiologic Agent

Rabies virus is a member of the family Rhabdoviridae. Two genera in this family, *Lyssavirus* and *Vesiculovirus*, contain species that cause human disease. Rabies virus is a lyssavirus that infects a broad range of animals and causes serious neurologic disease when transmitted to humans. This single-strand RNA virus has a nonsegmented, negative-sense (antisense) genome that consists of 11,932 nucleotides and encodes five proteins: nucleocapsid protein, phosphoprotein, matrix protein, glycoprotein, and a large polymerase protein. Rabies virus variants, which can be characterized by distinctive nucleotide sequences, are associated with specific animal reservoirs. Five other nonrabies virus species in the *Lyssavirus* genus have been reported to cause a clinical picture similar to rabies. Vesicular stomatitis virus, a vesiculovirus, causes vesiculation and ulceration in cattle, horses, and other animals and causes a self-limited, mild, systemic illness in humans (see "Other Rhabdoviruses," below).

Epidemiology

Rabies is a zoonotic infection that occurs in a variety of mammals throughout the world except in Antarctica and on some islands. Rabies virus is usually transmitted to humans by the bite of an infected animal. Canine rabies is endemic in many resource-poor and resource-limited countries and continues to be a threat to humans, particularly in Asia and Africa (see "Global Considerations," below); endemic canine rabies has been eliminated from the United States and most other resource-rich countries. Rabies is endemic in wildlife species, and a variety of animal reservoirs have been identified in different countries. Surveillance data from 2008 identified 6841 confirmed animal cases of rabies in the United States (including Puerto Rico). Only 7% of these cases were in domestic animals, including 294 cases in cats, 75 in dogs, and 59 in cattle. North American wildlife reservoirs, including bats, raccoons, skunks, and foxes, have endemic infection, with involvement of one or more rabies virus variants in each species **(Fig. 195-1).** "Spillover" of rabies to other wildlife species and to domestic animals occurs. Bat rabies virus variants are present in every state except Hawaii and are responsible for most indigenously acquired human rabies cases in the United States. Raccoon rabies is endemic along the entire eastern coast of the United States. Skunk rabies is present in the midwestern states, with another focus in California. Rabies in foxes occurs in Texas, New Mexico, Arizona, and Alaska.

Figure 195-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Distribution of the major rabies virus variants among wild terrestrial reservoirs in the United States and Puerto Rico, 2008. *(From JD Blanton et al: J Am Vet Med Assoc 235:676, 2009, Centers for Disease Control and Prevention.)*

Rabies virus variants isolated from humans or other mammalian species can be identified by reverse-transcription polymerase chain reaction (RT-PCR) amplification and sequencing or by characterization with monoclonal antibodies. These techniques are helpful in human cases with no known history of an exposure. Worldwide, most human rabies is transmitted from dogs in countries with endemic canine rabies and dog-to-dog transmission, and human cases can be imported by travelers returning from these regions. In North America, human disease is usually associated with transmission from bats; there may be no known history of bat bite or other bat exposure in these cases. Most human cases are due to a bat rabies virus variant associated with silver-haired and eastern pipistrelle bats. These are small bats whose bite may not be recognized, and the virus has adapted for replication at skin temperature and in cell types that are present in the skin.

Transmission from nonbite exposures is relatively uncommon. Aerosols generated in the laboratory or in caves containing millions of Brazilian free-tail bats have rarely caused human rabies. Transmission has resulted from corneal transplantation and recently from solid organ transplantation and from a vascular conduit (for a liver transplant) from undiagnosed donors with rabies in Texas and Germany. Human-to-human transmission is extremely rare, although theoretical concern about transmission to health care workers has prompted the implementation of barrier techniques to prevent exposures.

Pathogenesis

The incubation period of rabies (defined as the interval between exposure and the onset of clinical disease) is usually 20–90 days but in rare cases is as short as a few days or is >1 year. During most of the incubation period, rabies virus is thought to be present at or close to the site of inoculation **(Fig. 195-2)**. In muscles, the virus is known to bind to nicotinic acetylcholine receptors on postsynaptic membranes at neuromuscular junctions, but the exact details of viral entry into the skin and subcutaneous tissues have not yet been clarified. Rabies virus spreads centripetally along peripheral nerves toward the CNS at a rate up to ~250 mm/d via retrograde fast axonal transport to the spinal cord or brainstem. There is no well-documented evidence for hematogenous spread of rabies virus. Once the virus enters the CNS, it rapidly disseminates to other regions of the CNS via fast axonal transport along neuroanatomic connections. Neurons are prominently infected in rabies; infection of astrocytes is unusual. After CNS infection becomes established, there is centrifugal spread along sensory and autonomic nerves to other tissues, including the salivary glands, heart, adrenal glands, and skin. Rabies virus replicates in acinar cells of the salivary glands and is secreted in the saliva of rabid animals that serve as vectors of the disease.

Figure 195-2

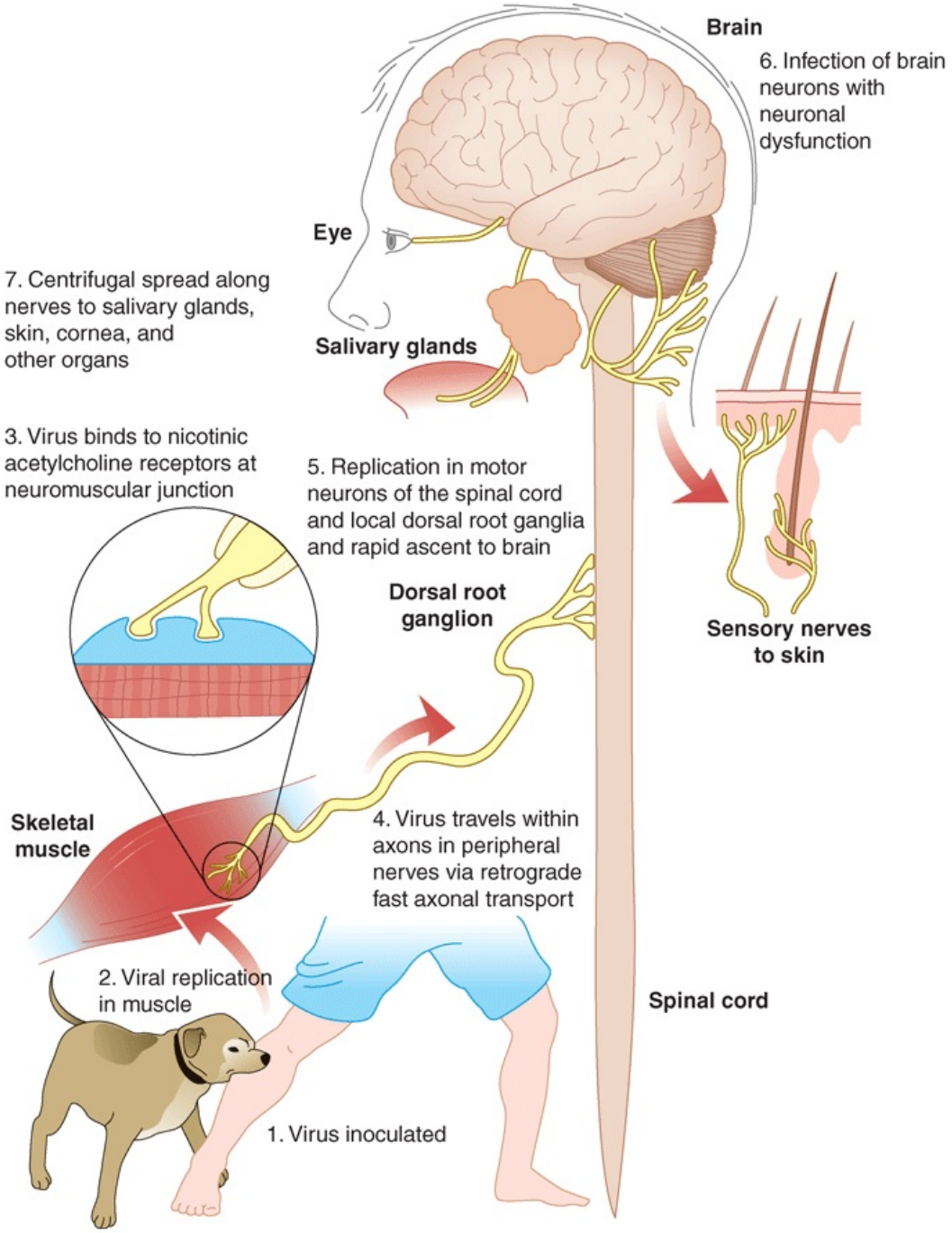

Schematic representation of the pathogenetic events following peripheral inoculation of rabies virus. *(Adapted from Jackson AC: Human disease, in Rabies, AC Jackson, WH Wunner (eds), San Diego, Academic Press, 2002, pp 219–244; with permission.)*

Pathologic studies show mild inflammatory changes in the CNS in rabies, with mononuclear inflammatory infiltration in the leptomeninges, perivascular regions, and parenchyma, including microglial nodules called *Babes nodules*. Degenerative neuronal changes usually are not prominent, and there is little evidence of neuronal death; neuronophagia is observed occasionally. The pathologic changes are surprisingly mild in light of the clinical severity and fatal outcome of the disease. The most characteristic pathologic finding in rabies is the *Negri body* **(Fig. 195-3)**. Negri bodies are eosinophilic cytoplasmic inclusions in brain neurons that are composed of rabies virus proteins and viral RNA. These inclusions occur in a minority of infected neurons, are commonly observed in Purkinje cells of the cerebellum and in pyramidal neurons of the hippocampus, and are less frequently seen in cortical and brainstem neurons. Negri bodies are not observed in all cases of rabies. The lack of prominent degenerative neuronal changes has led to the concept that neuronal dysfunction—rather than neuronal death—is responsible for clinical disease in rabies. The basis for behavioral changes, including the aggressive behavior of rabid animals, is not well understood.

Figure 195-3

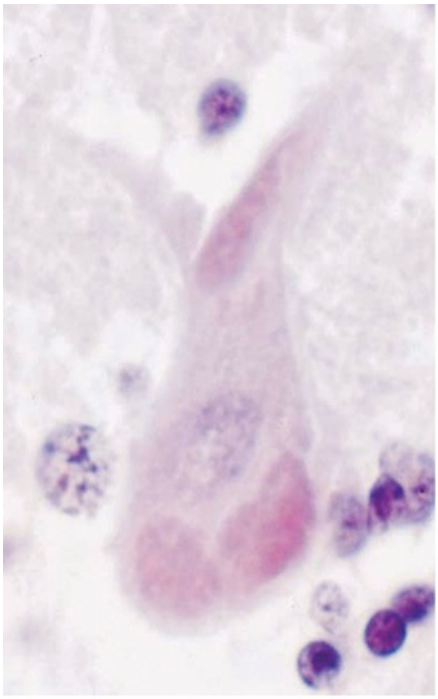

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three large Negri bodies in the cytoplasm of a cerebellar Purkinje cell from an 8-year-old boy who died of rabies after being bitten by a rabid dog in Mexico. *(From AC Jackson, E Lopez-Corella: N Engl J Med 335:568, 1996. © Massachusetts Medical Society.)*

Clinical Manifestations

In rabies, the emphasis must be on postexposure prophylaxis initiated before any symptoms or signs develop. Rabies should usually be suspected on the basis of the clinical presentation. The disease usually presents as an atypical encephalitis with relative preservation of consciousness. Rabies may be difficult to recognize late in the clinical course when progression to coma has occurred. A minority of patients present with acute flaccid paralysis. There are prodromal, acute neurologic, and comatose phases that usually progress to death despite aggressive therapy (**Table 195-1**).

Table 195-1 Clinical Stages of Rabies

Phase	Typical Duration	Symptoms and Signs
Incubation period	20–90 days	None
Prodrome	2–10 days	Fever, malaise, anorexia, nausea, vomiting, paresthesias, pain, or pruritus at the wound site
Acute neurologic disease		
Encephalitic (80%)	2–7 days	Anxiety, agitation, hyperactivity, bizarre behavior, hallucinations, autonomic dysfunction, hydrophobia
Paralytic (20%)	2–10 days	Flaccid paralysis in limb(s) progressing to quadriparesis with facial paralysis
Coma, death[a]	0–14 days	

[a]Recovery is rare.

Source: MAW Hattwick: Rabies virus, in *Principles and Practice of Infectious Diseases,* GL Mandell et al (eds). New York, Wiley, 1979, pp 1217–1228. Adapted with permission from Elsevier.

Prodromal Features

The earliest clinical features of rabies begin with nonspecific prodromal manifestations, including fever, malaise, headache, nausea, and vomiting. Anxiety or agitation may also occur. The earliest specific

neurologic symptoms of rabies include paresthesias, pain, or pruritus near the site of the exposure, which occurs in 50–80% of patients and strongly suggests rabies. The wound has usually healed by this point, and these symptoms probably reflect infection with associated inflammatory changes in local dorsal root or cranial sensory ganglia.

Encephalitic Rabies

Two acute neurologic forms of rabies are seen in humans: encephalitic (furious) in 80% and paralytic in 20%. Some of the manifestations of encephalitic rabies may be seen in other viral encephalitides as well. These features include fever, confusion, hallucinations, combativeness, and seizures. Autonomic dysfunction is common and may result in hypersalivation, gooseflesh, cardiac arrhythmia, and priapism. In encephalitic rabies, episodes of hyperexcitability are typically followed by periods of complete lucidity that become shorter as the disease progresses. Rabies encephalitis is distinguished by early brainstem involvement, which results in the classic features of hydrophobia (involuntary, painful contraction of the diaphragm and accessory respiratory, laryngeal, and pharyngeal muscles in response to swallowing liquids) and aerophobia (the same features caused by stimulation from a draft of air). These symptoms are probably due to dysfunction of infected brainstem neurons that normally inhibit inspiratory neurons near the nucleus ambiguus, resulting in exaggerated defense reflexes that protect the respiratory tract. The combination of hypersalivation and pharyngeal dysfunction is also responsible for the classic appearance of "foaming at the mouth" **(Fig. 195-4)**. Brainstem dysfunction progresses rapidly, and coma followed within days by death is the rule unless the course is prolonged by supportive measures. With such measures, late complications can include cardiac and/or respiratory failure, disturbances of water balance (syndrome of inappropriate antidiuretic hormone secretion or diabetes insipidus), noncardiogenic pulmonary edema, and gastrointestinal hemorrhage. Cardiac arrhythmias may be due to dysfunction affecting vital centers in the brainstem or to myocarditis. Multiple-organ failure is common in patients treated aggressively in critical care units.

Figure 195-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hydrophobic spasm of inspiratory muscles associated with terror in a patient with encephalitic (furious) rabies who is attempting to swallow water. *(Copyright DA Warrell, Oxford, UK; with permission.)*

Paralytic Rabies

About 20% of patients have paralytic rabies in which muscle weakness predominates and cardinal features of encephalitic rabies (hyperexcitability, hydrophobia, and aerophobia) are lacking. There is early and prominent flaccid muscle weakness, often beginning in the bitten extremity and spreading to produce quadriparesis and facial weakness. Sphincter involvement is common, sensory involvement is usually mild, and these cases are commonly misdiagnosed as Guillain-Barré syndrome. Patients with paralytic rabies generally survive a few days longer than those with encephalitic rabies, but multiple-organ failure nevertheless ensues.

Laboratory Investigations

Most routine laboratory tests in rabies yield normal results or show nonspecific abnormalities. Complete blood counts are usually normal. Examination of cerebrospinal fluid (CSF) often reveals mild mononuclear cell pleocytosis with a mildly elevated protein level. Severe pleocytosis (>1000 white cells/μL) is unusual and should prompt a search for an alternative diagnosis. CT head scans are usually normal in rabies. MRI brain scans may show signal abnormalities in the brainstem or other gray-matter areas, but these findings are variable and nonspecific. Electroencephalograms show only nonspecific abnormalities. Of course, important tests in suspected cases of rabies include those that may identify an alternative, potentially treatable diagnosis (see "Differential Diagnosis," below).

Diagnosis

In North America, a diagnosis of rabies often is not considered until relatively late in the clinical course, even with a typical clinical presentation. This diagnosis should be considered in patients presenting with acute atypical encephalitis or acute flaccid paralysis, including those in whom Guillain-Barré syndrome is suspected. The absence of an animal-bite history is common in North America. The lack of hydrophobia is not unusual in rabies. Once rabies is suspected, rabies-specific laboratory tests should be performed to confirm the diagnosis. Diagnostically useful specimens include serum, CSF, fresh saliva, skin biopsy samples from the neck, and brain tissue (rarely obtained before death). Because skin biopsy relies on the demonstration of rabies virus antigen in cutaneous nerves at the base of hair follicles, samples are usually taken from hairy skin at the nape of the neck. Corneal impression smears are of low diagnostic yield and are generally not performed. Negative antemortem rabies-specific laboratory tests never exclude a diagnosis of rabies, and tests may need to be repeated after an interval for diagnostic confirmation.

RABIES VIRUS–SPECIFIC ANTIBODIES

In a previously unimmunized patient, serum neutralizing antibodies to rabies virus are diagnostic. However, because rabies virus infects immunologically privileged neuronal tissues, serum antibodies may not develop until late in the disease. Antibodies may be detected within a few days after the onset of symptoms, but some patients die without detectable antibodies. The presence of rabies virus –specific antibodies in the CSF suggests rabies encephalitis, regardless of immunization status.

RT-PCR AMPLIFICATION

Detection of rabies virus RNA by RT-PCR is highly sensitive and specific. This technique can detect virus in fresh saliva samples, CSF, and skin and brain tissues. In addition, RT-PCR with genetic sequencing can distinguish among rabies virus variants, permitting identification of the probable source of an infection.

DIRECT FLUORESCENT ANTIBODY TESTING

Direct fluorescent antibody (DFA) testing with rabies virus antibodies conjugated to fluorescent dyes is highly sensitive and specific and can be performed quickly and applied to skin biopsies and brain tissue. In skin biopsies, rabies virus antigen may be detected in cutaneous nerves at the base of hair follicles.

Differential Diagnosis

The diagnosis of rabies may be difficult without a history of animal exposure, and no exposure to an animal (e.g., a bat) may be recalled. The presentation of rabies is usually quite different from that of acute viral encephalitis due to most other causes, including herpes simplex encephalitis and arboviral (e.g., West Nile) encephalitis. Early neurologic symptoms may occur at the site of the bite, and there may be early features of brainstem involvement with preservation of consciousness. Postinfectious (immune-mediated) encephalomyelitis may follow influenza, measles, mumps, and other infections; it may also occur as a sequela of immunization with rabies vaccine derived from neural tissues, which are used only in resource-limited and resource-poor countries. Rabies may present with unusual neuropsychiatric symptoms and may be misdiagnosed as a psychiatric disorder. Rabies hysteria may occur as a psychological response to the fear of rabies and is often characterized by a shorter incubation period than rabies, aggressive behavior, inability to communicate, and a long course with recovery.

As previously mentioned, paralytic rabies may mimic Guillain-Barré syndrome. In these cases, fever, bladder dysfunction, a normal sensory examination, and CSF pleocytosis favor a diagnosis of rabies. Conversely, Guillain-Barré syndrome may occur as a complication of rabies vaccination with a neural tissue–derived product (e.g., suckling mouse brain vaccine) and may be mistaken for paralytic rabies (i.e., vaccine failure).

Treatment: Rabies

There is no established treatment for rabies. There have been several recent treatment failures with the combination of antiviral drugs, ketamine, and therapeutic (induced) coma—measures that were used in a healthy survivor in whom antibodies to rabies virus were detected at presentation. Expert opinion should be sought before a course of experimental therapy is embarked upon. A palliative

approach may be appropriate for some patients.

Prognosis

Rabies is an almost uniformly fatal disease but is almost always preventable with appropriate postexposure therapy during the early incubation period (see below). There are seven well-documented cases of survival from rabies. All but one of these patients had received rabies vaccine before disease onset. The single survivor who had not received vaccine had neutralizing antibodies to rabies virus in serum and CSF at clinical presentation. Most patients with rabies die within several days of illness, despite aggressive care in a critical care unit.

Prevention

POSTEXPOSURE PROPHYLAXIS

Since there is no effective therapy for rabies, it is extremely important to prevent the disease after an animal exposure. **Figure 195-5** shows the steps involved in making decisions about rabies postexposure prophylaxis (PEP). On the basis of the history of the exposure and local epidemiologic information, the physician must decide whether initiation of PEP is warranted. Healthy dogs, cats, or ferrets may be confined and observed for 10 days. PEP is not necessary if the animal remains healthy. If the animal develops signs of rabies during the observation period, it should be euthanized immediately, and the head should be transported to the laboratory under refrigeration and examined for the presence of rabies virus by DFA testing and viral isolation using cell culture and/or mouse inoculation. Any animal other than a dog, cat, or ferret should be euthanized immediately and the head submitted for laboratory examination. In high-risk exposures and in areas where canine rabies is endemic, rabies prophylaxis should be initiated without waiting for laboratory results. If the laboratory results prove to be negative, it may safely be concluded that the animal's saliva did not contain rabies virus, and immunization should be discontinued. If an animal escapes after an exposure, it must be considered rabid, and PEP must be initiated unless information from public health officials indicates otherwise (i.e., there is no endemic rabies in the area). PEP may be warranted in situations where a person (e.g., a small child or a sleeping adult) is present in the same space as a bat and an unrecognized bite cannot be reliably excluded.

Figure 195-5

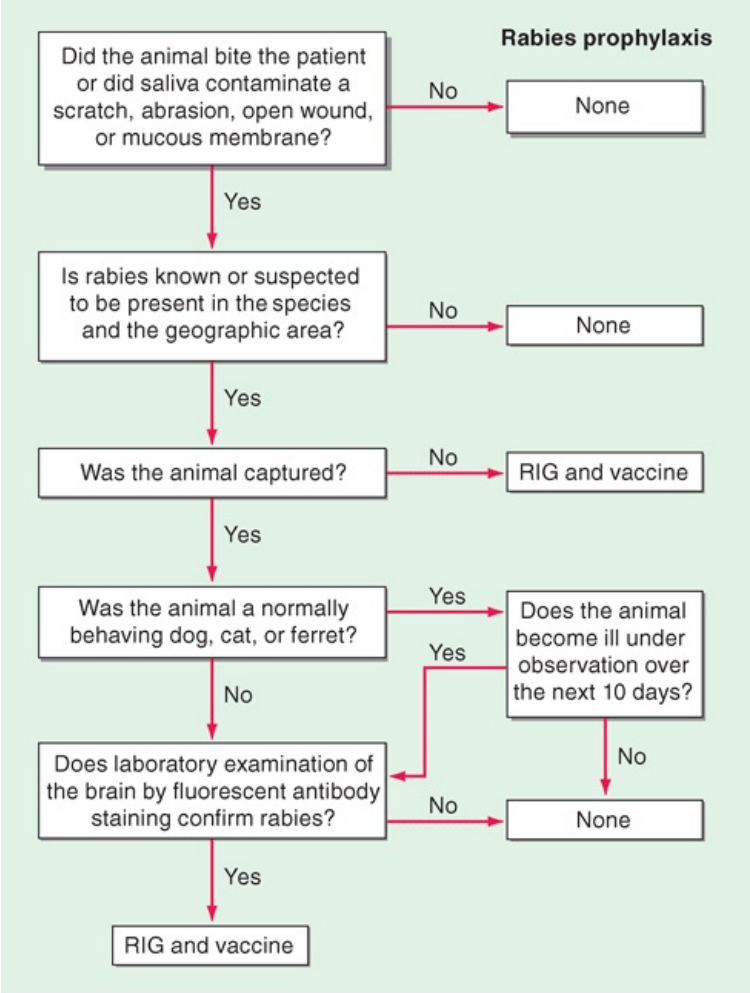

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for rabies postexposure prophylaxis. RIG, rabies immune globulin. *[From L Corey, in Harrison's Principles of Internal Medicine, 15th ed. E Braunwald et al (eds): New York, McGraw-Hill, 2001; adapted with permission.]*

PEP includes local wound care and both active and passive immunization. Local wound care is essential and may greatly decrease the risk of rabies virus infection. Wound care should not be delayed, even if the initiation of immunization is postponed pending the results of the 10-day observation period. All bite wounds and scratches should be washed thoroughly with soap and water. Devitalized tissues should be debrided, tetanus prophylaxis given, and antibiotic treatment initiated whenever indicated.

All previously unvaccinated persons (but not those who have previously been immunized) should be passively immunized with rabies immune globulin (RIG). If RIG is not immediately available, it should be administered no later than 7 days after the first vaccine dose. After day 7, endogenous antibodies are being produced, and passive immunization may actually be counterproductive. If anatomically feasible, the entire dose of RIG (20 IU/kg) should be infiltrated at the site of the bite; otherwise, any RIG remaining after infiltration of the bite site should be administered IM at a distant site. With multiple or large wounds, the RIG preparation may need to be diluted in order to obtain a sufficient volume for adequate infiltration of all wound sites. If the exposure involves a mucous membrane, the entire dose should be administered IM. Rabies vaccine and RIG should never be administered at the same site or with the same syringe. Commercially available RIG in the United States is purified from the serum of hyperimmunized human donors. These human RIG preparations are much better tolerated than are the equine-derived preparations still in use in some countries (see "Global Considerations," below). Serious adverse effects of human RIG are uncommon. Local pain and low-grade fever may occur.

Two purified inactivated rabies vaccines are available for rabies PEP in the United States. They are highly immunogenic and remarkably safe compared with earlier vaccines. Four 1-mL doses of rabies vaccine should be given IM in the deltoid area. (The anterolateral aspect of the thigh is also acceptable in children.) Gluteal injections, which may not always reach muscle, should not be given and have been associated with rare vaccine failures. Ideally, the first dose should be given as soon as possible after exposure; failing that, it should be given without further delay. The three additional doses should be given on days 3, 7, and 14; a fifth dose on day 28 is no longer recommended. Pregnancy is

not a contraindication for immunization. Glucocorticoids and other immunosuppressive medications may interfere with the development of active immunity and should not be administered during PEP unless they are essential. Routine measurement of serum neutralizing antibody titers is not required, but titers should be measured 2–4 weeks after immunization in immunocompromised persons. Local reactions (pain, erythema, edema, and pruritus) and mild systemic reactions (fever, myalgias, headache, and nausea) are common; anti-inflammatory and antipyretic medications may be used, but immunization should not be discontinued. Systemic allergic reactions are uncommon, but anaphylaxis does occur rarely and can be treated with epinephrine and antihistamines. The risk of rabies development should be carefully considered before the decision is made to discontinue vaccination because of an adverse reaction.

PREEXPOSURE RABIES VACCINATION

Preexposure rabies prophylaxis should be considered for people with an occupational or recreational risk of rabies exposures, including certain travelers to rabies-endemic areas. This primary schedule consists of three doses of rabies vaccine given on days 0, 7, and 21 or 28. Serum neutralizing antibody tests help determine the need for subsequent booster doses. When a previously immunized individual is exposed to rabies, two booster doses of vaccine should be administered on days 0 and 3. Wound care remains essential. As stated above, RIG should not be administered to previously vaccinated persons.

Global Considerations

Worldwide, endemic canine rabies is estimated to cause 55,000 human deaths annually. Most of these deaths occur in Asia and Africa, with rural populations and children most frequently affected. Most of the burden of rabies PEP is borne by people with the least resources. In Latin America, rabies control efforts in dogs have been quite successful in recent years. In Canada and Europe, epizootics of rabies in red foxes have been well controlled with the use of baits containing rabies vaccine. A similar approach is used in Canada to control raccoon rabies.

In addition to the rabies vaccines discussed above, vaccines grown in either primary cell lines (hamster or dog kidney) or continuous cell lines (Vero cells) are satisfactory and are available in many countries outside the United States. Less expensive vaccines derived from neural tissues have been used in developing countries; however, these vaccines are associated with serious neuroparalytic complications, including postinfectious encephalomyelitis and Guillain-Barré syndrome. The use of these vaccines should be discontinued as soon as possible, and progress has been made in this regard. Worldwide, >10 million individuals receive postexposure vaccination against rabies each year.

If human RIG is unavailable, purified equine RIG can be used in the same manner at a dose of 40 IU/kg. Before the administration of equine RIG, hypersensitivity should be assessed by intradermal testing with a 1:10 dilution. The incidence of anaphylactic reactions and serum sickness has been low with recent equine RIG products.

OTHER RHABDOVIRUSES

Other Lyssaviruses

A growing number of lyssaviruses other than rabies virus have been discovered to infect bat populations in Africa, Europe, and Australia. Four of these viruses have produced a very small number of cases of a human disease indistinguishable from rabies: European bat lyssaviruses 1 and 2, Australian bat lyssavirus, and the Duvenhage virus (in Africa). Mokola virus, a lyssavirus that has been isolated from shrews with an unknown reservoir species in Africa, may also produce human disease indistinguishable from rabies.

Vesicular Stomatitis Virus (VSV)

Vesicular stomatitis is a viral disease of cattle, horses, pigs, and some wild mammals. VSV is a member of the genus *Vesiculovirus* in the family Rhabdoviridae. Outbreaks of vesicular stomatitis in horses and cattle occur sporadically in the southwestern United States. The animal infection is associated with severe vesiculation and ulceration of oral tissues, teats, and feet and may be clinically indistinguishable from the more dangerous foot-and-mouth disease. Epidemics are usually seasonal, typically beginning in the late spring, and are probably due to arthropod vectors. Direct animal-to-animal spread can also occur, although the virus cannot penetrate intact skin. Transmission to humans usually results from direct contact with infected animals (particularly cattle) and occasionally follows laboratory exposure. In human disease, early conjunctivitis is followed by an acute influenza-like illness with fever, chills, nausea, vomiting, headache, retrobulbar pain, myalgias, substernal pain, malaise, pharyngitis, and lymphadenitis. Small vesicular lesions may be present on the buccal mucosa or on the fingers. Encephalitis is very rare. The illness usually lasts 3–6 days, with complete recovery. Subclinical infections are common. A serologic diagnosis can be made on the basis of a rise in titer of complement-fixing or neutralizing antibodies. Therapy is symptom-based.

FURTHER READINGS

Baer GM (ed): *The Natural History of Rabies* , 2nd ed. Boca Raton, CRC Press, 1991

Jackson AC: Update on rabies diagnosis and treatment. Curr Infect Dis Rep 11:296, 2009[PMID: 19545499]

———, Wunner WH (eds): *Rabies*, 2nd ed. London, Elsevier, 2007

——— et al: Management of rabies in humans. Clin Infect Dis 36:60, 2003

Letchworth GJ et al: Vesicular stomatitis. Vet J 157:239, 1999[PMID: 10328837]

Manning SE et al: Human rabies prevention—United States, 2008: Recommendations of the Advisory Committee on Immunization Practices. MMWR Recomm Rep 57(RR-3):1, 2008

Warrell MJ, Warrell DA: Rabies and other lyssavirus diseases. Lancet 363:959, 2004[PMID: 15043965]

Willoughby RE Jr et al: Survival after treatment of rabies with induction of coma. N Engl J Med 352:2508, 2005[PMID: 15958806]

World Health Organization: *WHO Expert Consultation on Rabies: First Report* . First Report ed. Geneva, WHO, 2005

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIONS CAUSED BY ARTHROPOD- AND RODENT-BORNE VIRUSES: INTRODUCTION

Some zoonotic viruses are transmitted in nature without regard to humans and only incidentally infect and produce disease in humans; in addition, a few agents are regularly spread among humans by arthropods. Most of these viruses either are maintained by arthropods or chronically infect rodents. Obviously, the mode of transmission is not a rational basis for taxonomic classification. Indeed, zoonotic viruses from at least seven families act as significant human pathogens **(Table 196-1)**. The virus families differ fundamentally from one another in terms of morphology, replication mechanisms, and genetics. Information on a virus's membership in a family or genus is enlightening with regard to maintenance strategies, sensitivity to antiviral agents, and some aspects of pathogenesis but does not necessarily predict which clinical syndromes (if any) the virus will cause in humans.

Table 196-1 Major Zoonotic Virus Families and Some Characteristics of Typical Members

Family	Genus or Group	Syndrome(s): Typical Viruses	Maintenance Strategy
Arenaviridae	Old World complex	FM, E: Lymphocytic choriomeningitis virus HF: Lassa fever virus	Chronic infection of rodents, often with persistent viremia; vertical transmission common
	New World or Tacaribe complex	HF: South American HF viruses (Machupo, Junin, Guanarito, Sabia)	Chronic infection of rodents, sometimes with persistent viremia; vertical infection may occur
Bunyaviridae	*Bunyavirus*	E: California serogroup viruses (La Crosse, Jamestown Canyon, California encephalitis) FM: Bunyamwera, group C, Tahyna viruses	Mosquito-vertebrate cycle; transovarial transmission in mosquito common
		FM: Oropouche virus	Transmitted by *Culicoides*
	Phlebovirus	FM: Sandfly fever, Toscana viruses FM: Punta Toro virus	Sandfly transmission between vertebrates, with prominent transovarial component in sandfly
		HF, FM, E: Rift Valley fever virus	Mosquito-vertebrate transmission, with transovarial component in mosquito
	Nairovirus	HF: Crimean-Congo HF virus	Tick-vertebrate, with transovarial transmission in tick
	Hantavirus	HF: Hantaan, Dobrava, Puumala viruses	Rodent reservoir; chronic virus shedding, but chronic viremia unknown
		HF: Sin Nombre and related hantaviruses	Sigmodontine rodent reservoir
Filoviridae[a]	*Ebolavirus, Marburgvirus*	HF: Marburg viruses, Ebola viruses (4 species)	Unknown
Flaviviridae	*Flavivirus* (mosquito-borne)	HF: Yellow fever virus FM, HF: Dengue viruses (4 serotypes) E: St. Louis, Japanese, West Nile, and Murray Valley encephalitis viruses; Rocio viruses	Mosquito-vertebrate; transovarial rare
	Flavivirus (tick-borne)	E: Central European tick-borne encephalitis, Russian spring-summer encephalitis, Powassan viruses HF: Omsk HF, Kyasanur Forest; Alkhurma disease viruses	Tick-vertebrate
Reoviridae	*Coltivirus*	FM, E: Colorado tick fever virus	Tick-vertebrate
	Orbivirus	FM, E: Orungo, Kemerovo viruses	Arthropod-vertebrate
Rhabdoviridae[b]	*Vesiculovirus*	FM: Vesicular stomatitis virus (Indiana, New Jersey); Chandipura, Piry viruses	Sandfly-vertebrate, with prominent transovarial component in sandfly

Togaviridae	*Alphavirus*	AR: Sindbis, chikungunya, Mayaro, Ross River, Barmah Forest viruses E: Eastern, western, and Venezuelan equine encephalitis viruses	Mosquito-vertebrate

[a]The Filoviridae are discussed in Chap. 197.

[b]The Rhabdoviridae are discussed in Chap. 195.

Note: Abbreviations refer to the disease syndrome most commonly associated with the virus: AR, arthritis, rash; E, encephalitis; FM, fever, myalgia; HF, hemorrhagic fever.

Families of Arthropod- and Rodent-Borne Viruses (Table 196-1)

THE ARENAVIRIDAE

The Arenaviridae are spherical, 110- to 130-nm particles that bud from the cell's plasma membrane and utilize ambisense RNA genomes with two segments for replication. There are two main phylogenetic branches of Arenaviridae: the Old World viruses, such as Lassa fever and lymphocytic choriomeningitis (LCM) viruses, and the New World viruses, including those causing the South American hemorrhagic fevers (HFs). Arenaviruses persist in nature by chronically infecting rodents with a striking one-virus–one-rodent species relationship. These rodent infections result in long-term virus excretion and perhaps in lifelong viremia; vertical infection is common with some arenaviruses. Humans become infected through the inhalation of aerosols containing arenaviruses, which are then deposited in the terminal air passages, and probably also through close contact with rodents and their excreta, which results in the contamination of mucous membranes or breaks in the skin.

THE BUNYAVIRIDAE

The family Bunyaviridae includes four medically significant genera. All of these spherical viruses have three negative-sense RNA segments maturing into 90- to 120-nm particles in the Golgi complex and exiting the cell by exocytosis. Viruses of the genus *Bunyavirus* are largely mosquito-borne and have a viremic vertebrate intermediate host; many are also transovarially transmitted in their specific mosquito host. One serologic group also uses biting midges as vectors. Sandflies or mosquitoes are the vectors for the genus *Phlebovirus* (named after phlebotomus fever or sandfly fever, the best-known disease associated with the genus), while ticks serve as vectors for the genus *Nairovirus*. Viruses of both of these genera are also associated with vertical transmission in the arthropod host and with horizontal spread through viremic vertebrate hosts. The genus *Hantavirus* is unique among the Bunyaviridae in that it is not transmitted by arthropods but is maintained in nature by rodent hosts that chronically shed virus. Like the arenaviruses, the hantaviruses usually display striking virus–rodent species specificity. Hantaviruses do not cause chronic viremia in their rodent hosts and are transmitted only horizontally from rodent to rodent.

OTHER FAMILIES

The Flaviviridae are positive-sense, single-strand RNA viruses that form particles of 40–50 nm in the endoplasmic reticulum. The flaviviruses discussed here are from the genus *Flavivirus* and make up two phylogenetically and antigenically distinct divisions transmitted among vertebrates by mosquitoes and ticks, respectively. The mosquito-borne viruses fall into phylogenetic groups that include yellow fever virus, the four dengue viruses, and encephalitis viruses, while the tick-borne group encompasses a geographically varied spectrum of species, some of which are responsible for encephalitis or for hemorrhagic disease with encephalitis. The Reoviridae are double-strand RNA viruses with multisegmented genomes. These 80-nm particles are the only viruses discussed in this chapter that do not have a lipid envelope and thus are insensitive to detergents. The Togaviridae have a single positive-strand RNA genome and bud particles of ~ 60–70 nm from the plasma membrane. The togaviruses discussed here are all members of the genus *Alphavirus* and are transmitted among vertebrates by mosquitoes in their natural cycle. **The Filoviridae and the Rhabdoviridae are discussed in Chaps. 197 and 195, respectively.**

Prominent Features of Arthropod- and Rodent-Borne Viruses

Although this chapter discusses the major features of selected arthropod- and rodent-borne viruses, it does not deal with >500 other distinct recognized zoonotic viruses, about one-fourth of which infect humans. Zoonotic viruses are undergoing genetic evolution, "new" zoonotic viruses are being discovered, and the epidemiology of zoonotic viruses is continuing to evolve through environmental changes affecting vectors, reservoirs, and humans. These zoonotic viruses are most numerous in the tropics but are also found in temperate and frigid climates. Their distribution and seasonal activity may be variable and often depend largely on ecologic conditions such as rainfall and temperature, which in turn affect the density of vectors and reservoirs and the development of infection therein.

MAINTENANCE AND TRANSMISSION

Arthropod-borne viruses infect their vectors after the ingestion of a blood meal from a viremic vertebrate. The vectors then develop chronic, systemic infection as the viruses penetrate the gut and spread throughout the body. The viruses eventually reach the salivary glands during a period that is referred to as *extrinsic incubation* and that typically lasts 1–3 weeks in mosquitoes. At this point, an arthropod is competent to continue the chain of transmission by infecting another vertebrate when a subsequent blood meal is taken. The arthropod generally is unharmed by the infection, and the natural vertebrate partner usually has only transient viremia with no overt disease. An alternative mechanism for virus maintenance in its arthropod host

is transovarial transmission, which is common among members of the family Bunyaviridae.

Rodent-borne viruses such as the hantaviruses and arenaviruses are maintained in nature by chronic infection transmitted between rodents. As in arthropod-borne virus cycles, there is usually a high degree of rodent–virus specificity, and there is no overt disease in the reservoir/vector.

EPIDEMIOLOGY

The distribution of arthropod- and rodent-borne viruses is restricted by the areas inhabited by their reservoir/vectors and provides an important clue in the differential diagnosis. **Table 196-2** shows the approximate geographic distribution of the most important of these viruses. Members of each family, each genus, and even each serologically related group usually occur in each area but may not be pathogenic in all areas or may not be a commonly recognized cause of disease in all areas and so may not be included in the table.

Table 196-2 Geographic Distribution of Some Important and Commonly Encountered Human Zoonotic Viral Diseases

Area	Arenaviridae	Bunyaviridae	Flaviviridae	Rhabdoviridae	Togaviridae
North America	Lymphocytic choriomeningitis	La Crosse, Jamestown Canyon, California encephalitis; hantavirus pulmonary syndrome	St. Louis, Powassan, West Nile encephalitis; dengue	Vesicular stomatitis	Eastern, western equine encephalitis
South America	Bolivian (Machupo, Chapare), Argentine, Venezuelan, and Brazilian HF; lymphocytic choriomeningitis	Oropouche, group C, Punta Toro infection; hantavirus pulmonary syndrome	Yellow fever, dengue, Rocio virus infection	Vesicular stomatitis, Piry virus infection	Mayaro virus infection, Venezuelan equine encephalitis
Europe	Lymphocytic choriomeningitis	Tahyna, Toscana, sandfly fever; HF with renal syndrome	West Nile, Central European tick-borne, Russian spring-summer encephalitis	—	Sindbis virus infection
Middle East	Alkhurma HF virus infection	Sandfly fever, Crimean-Congo HF	West Nile encephalitis, dengue	—	—
Eastern Asia	—	Sandfly fever; Hantaan, Seoul virus infection	Dengue; Japanese, Russian spring-summer encephalitis; Omsk HF	Chandipura virus infection	—
Southwestern Asia	—	Sandfly fever, Crimean-Congo HF	West Nile, Japanese encephalitis; dengue; Kyasanur Forest disease	—	Chikungunya virus infection
Southeast Asia	—	Seoul virus infection	Japanese encephalitis, dengue	—	Chikungunya virus infection
Africa	Lassa fever; Lujo virus infection	Bunyamwera virus infection, Rift Valley fever	Yellow fever, dengue	—	Sindbis, chikungunya virus infection
Australia	—	—	Murray Valley encephalitis, dengue	—	Ross River, Barmah Forest virus infection

Abbreviation: HF, hemorrhagic fever.

Most of these diseases are acquired in a rural setting; a few have urban vectors. Seoul, sandfly fever, and Oropouche viruses are examples of urban viruses, but the most notable are yellow fever, dengue, and chikungunya viruses. A history of mosquito bite has little diagnostic significance in the individual; a history of tick bite is more diagnostically specific. Rodent exposure is often reported by persons infected with an arenavirus or a hantavirus but again has little specificity. Indeed, aerosols may infect persons who have no recollection of having even seen rodents.

SYNDROMES

Human disease caused by arthropod- and rodent-borne viruses is often subclinical. The spectrum of possible responses to infection is wide, and our knowledge of the outcome of most of these infections is limited. The usual disease syndromes associated with these viruses have been grouped into four categories: fever and myalgia, arthritis and rash, encephalitis, and hemorrhagic fever. Although for the purposes of this discussion most viruses have been placed in a single group, the categories often overlap. For example, West Nile and Venezuelan equine encephalitis viruses are discussed as encephalitis viruses, but during epidemics many cases of milder febrile syndromes are recognized relative to less common cases of encephalitis. Similarly, Rift Valley fever virus is best known as a cause of HF, but the attack rates for febrile disease are far higher, and encephalitis is occasionally seen as well. LCM virus is classified as a cause of fever and myalgia because this syndrome is its most common disease manifestation and because, even when central nervous system (CNS) disease occurs, it is usually mild and is preceded by fever and myalgia. Dengue virus infection is considered as a cause of fever and myalgia (dengue fever) because this is by far the most common manifestation worldwide and is the syndrome most likely to be seen in the United States; however, dengue HF is also discussed in the HF section because of its complicated pathogenesis and importance in pediatric practice in certain areas of the world.

DIAGNOSIS

Laboratory diagnosis is required in any given case, although epidemics occasionally provide clinical and epidemiologic clues on which an educated guess as to etiology can be based. For most arthropod- and rodent-borne viruses, acute-phase serum samples (collected within 3 or 4 days of onset) have yielded isolates, and paired sera have been used to demonstrate rising antibody titers by a variety of tests. Intensive efforts to develop rapid tests for HF have resulted in an antigen-detection enzyme-linked immunosorbent assay (ELISA) and an IgM-capture ELISA that can provide a diagnosis based on a single serum sample within a few hours and are particularly useful in severe cases. More sensitive reverse-transcription polymerase chain reaction (RT-PCR) tests may yield diagnoses based on samples without detectable antigen and may also provide useful genetic information about the virus. Hantavirus infections differ from others discussed here in that severe acute disease is immunopathologic; patients present with serum IgM that serves as the basis for a sensitive and specific test.

At diagnosis, patients with encephalitis are generally no longer viremic or antigenemic and usually do not have virus in cerebrospinal fluid (CSF). In this situation, the value of serologic methods for IgM determination and RT-PCR is high. IgM capture is increasingly being used for the simultaneous testing of serum and CSF. IgGELISA or classic serology is useful in the evaluation of past exposure to the viruses, many of which circulate in areas with a minimal medical infrastructure and sometimes cause mild or subclinical infection.

The remainder of this chapter offers general descriptions of the broad syndromes caused by arthropod- and rodent-borne viruses. Most of the diseases under consideration have not been studied in detail with modern medical approaches; thus available data may be incomplete or biased.

FEVER AND MYALGIA

Fever and myalgia constitute the syndrome most commonly associated with zoonotic virus infection. Many of the numerous viruses belonging to the families listed in Table 196-1 probably cause this syndrome, but several viruses have been selected for inclusion in the table because of their prominent associations with the syndrome and their biomedical importance.

The syndrome typically begins with the abrupt onset of fever, chills, intense myalgia, and malaise. Patients may also report joint or muscle pains, but no true arthritis is detectable. Anorexia is characteristic and may be accompanied by nausea or even vomiting. Headache is common and may be severe, with photophobia and retroorbital pain. Physical findings are minimal and are usually confined to conjunctival injection with pain on palpation of muscles or the epigastrium. The duration of symptoms is quite variable but generally is 2–5 days, with a biphasic course in some instances. The spectrum of disease varies from subclinical to temporarily incapacitating.

Less constant findings include a maculopapular rash. Epistaxis may occur but does not necessarily indicate a bleeding diathesis. A minority of the cases caused by some viruses are known or suspected to include aseptic meningitis, but this diagnosis is difficult to make in remote areas, given the patients' photophobia and myalgia as well as the lack of opportunity to examine the CSF. Although pharyngitis may be noted or radiographic evidence of pulmonary infiltrates found in some cases, these viruses are not primary respiratory pathogens. The differential diagnosis includes anicteric leptospirosis, rickettsial diseases, and the early stages of other syndromes discussed in this chapter. These diseases are often described as "flulike," but the usual absence of cough and coryza makes influenza an unlikely confounder except at the earliest stages.

Complete recovery is generally the outcome in this syndrome, although prolonged asthenia and nonspecific symptoms have been described in some cases, particularly after infection with LCM or dengue virus. Treatment is supportive, with aspirin avoided because of the potential for exacerbated bleeding and Reye's syndrome. Efforts at prevention are best based on vector control, which, however, may be expensive or impossible. For mosquito control, destruction of breeding sites is generally the most economically and environmentally sound approach. Emerging technologies include mosquito genetic transformation and introduction of *Wolbachia* to limit multiplication rates. Measures taken by the individual to avoid the vector can be valuable. Avoiding the vector's habitat and times of peak activity, using screens or other barriers (e.g., permethrin-impregnated bed nets) to prevent the vector from entering dwellings, judiciously applying arthropod repellents such as diethyltoluamide (DEET) to the skin, and wearing permethrin-impregnated clothing are all possible approaches, depending on the vector and its habits.

Lymphocytic Choriomeningitis

LCM is transmitted from the common house mouse (*Mus musculus*) to humans by aerosols of excreta and secreta. LCM virus, an arenavirus, is maintained in the mouse mainly by vertical transmission from infected dams. The vertically infected mouse remains viremic for life, with high concentrations of virus in all tissues. Infected colonies of pet hamsters have also served as a link to humans. LCM virus is widely used in immunology laboratories as a model of T cell function and can silently infect cell cultures and passaged tumor lines, resulting in infections among scientists and animal caretakers. Patients with LCM may have a history of residence in rodent-infested housing or other exposure to rodents. An antibody prevalence of ~5–10% has been reported among adults from the United States, Argentina, and endemic areas of Germany.

LCM differs from the general syndrome of fever and myalgia in that its onset is gradual. Among the conditions occasionally associated with LCM are orchitis, transient alopecia, arthritis, pharyngitis, cough, and maculopapular rash. An estimated one-fourth of patients or fewer experience a febrile phase of 3–6 days and then, after a brief remission, develop renewed fever accompanied by severe headache, nausea and vomiting, and meningeal signs lasting for ~1 week. These patients virtually always recover fully, as do the uncommon patients with clear-cut signs of encephalitis. Recovery may be delayed by transient hydrocephalus.

During the initial febrile phase, leukopenia and thrombocytopenia are common and virus can usually be isolated from blood. During the CNS phase, virus may be found in the CSF, but antibodies are present in blood. The pathogenesis of LCM is thought to resemble that following direct intracranial inoculation of the virus into adult mice; the onset of the immune response leads to T cell–mediated immunopathologic meningitis. During the meningeal phase, CSF mononuclear-cell counts range from the hundreds to the low thousands per microliter, and hypoglycorrhachia is found in one-third of cases. The IgM-capture ELISA of serum and CSF is usually positive; RT-PCR assays have been developed for application to CSF. Recent infections transmitted by organ transplantation did not include evidence of an immune response, followed a fulminant course (not unlike that of Lassa fever), and required immunohistochemistry or RT-PCR for diagnosis.

Infection with LCM virus should be suspected in acutely ill febrile patients with marked leukopenia and thrombocytopenia. In cases of aseptic meningitis, any of the following should suggest LCM: well-marked febrile prodrome, adult age, autumn seasonality, low CSF glucose levels, or CSF mononuclear cell counts of >1000/μL.

In pregnant women, LCM virus infection may lead to fetal invasion with consequent congenital hydrocephalus and chorioretinitis. Since the maternal infection may be mild, consisting of only a short febrile illness, antibodies to the virus should be sought in both the mother and the fetus in suspicious circumstances, particularly TORCH-negative neonatal hydrocephalus. [TORCH is a battery of tests encompassing *t*oxoplasmosis, *o*ther conditions (congenital syphilis and viral infection), *r*ubella, *c*ytomegalovirus infection, and *h*erpes simplex virus infection.]

Bunyamwera Virus Infection

The mosquito-transmitted Bunyamwera serogroup viruses are found on every continent except Australia and Antarctica. Bunyamwera virus and its close relative Ilesha virus commonly cause febrile disease in Africa. Ngari virus, a reassortant of Bunyamwera virus, has recently been identified as an important human pathogen in Africa. Other related viruses are implicated in such disease in Southeast Asia (Batai virus), Europe (Calovo virus), and South America (Wyeomyia virus). In North America, Cache Valley virus has been implicated in febrile human disease and in rare instances of more serious systemic illness; the presence of serum antibodies to this virus may be associated with congenital malformations. In Central America, the closely related Fort Sherman virus causes the fever-myalgia syndrome.

Group C Virus Infection

The group C viruses include at least 11 agents transmitted by mosquitoes in neotropical forests. These agents are among the most common causes of arboviral infection in humans entering American jungles and cause acute febrile disease.

Tahyna Virus Infection

This California serogroup virus (see discussion of California encephalitis, below) occurs in central and western Europe, and related viruses are emerging in Russia. The significance of Tahyna virus in human health has been well studied only in the Czech and Slovak Republics; there, the virus was found to be a prominent cause of febrile disease, in some cases causing pharyngitis, pulmonary syndromes, and aseptic meningitis. The potential for arboviruses to be unexpectedly involved in such cases in areas of high mosquito prevalence needs to be kept in mind.

Oropouche Fever

Oropouche virus is transmitted in Central and South America by a biting midge, *Culicoides paraensis*, which often breeds to high density in cacao husks and other vegetable detritus found in towns and cities. Explosive epidemics involving thousands of cases have been reported from several towns in Brazil and Peru. Rash and aseptic meningitis have been detected in a number of cases.

Sandfly Fever

The sandfly *Phlebotomus papatasi* transmits sandfly fever. Female sandflies may be infected by the oral route as they take a blood meal and may transmit the virus to offspring when they lay their eggs after a second blood meal. This prominent transovarial pattern was the first to be recognized among dipterans and complicates virus control. A previous designation for sandfly fever, "3-day fever," instructively describes the brief, debilitating course associated with this essentially benign infection. There is neither a rash nor CNS

involvement, and complete recovery is the rule.

Sandfly fever is found in the circum-Mediterranean area, extending to the east through the Balkans into parts of China as well as into the Middle East and southwestern Asia. The vector is found in both rural and urban settings and is known for its small size, which enables it to penetrate standard mosquito screens and netting, and for its short flight range. Epidemics have been described in the wake of natural disasters and wars. In parts of Europe, sandfly populations and virus transmission were greatly reduced by the extensive residual spraying conducted after World War II to control malaria, and the incidence continues to be low. A common pattern of disease in endemic areas consists of high attack rates among travelers and military personnel and little or no disease in the local population, who are protected after childhood infection. More than 30 related phleboviruses are transmitted by sandflies and mosquitoes, but most are of unknown significance in terms of human health.

Toscana Virus Disease

Toscana virus is a *Phlebovirus* (family Bunyaviridae) transmitted primarily by the circum-Mediterranean sandfly *P. perniciosus*. The vertebrate amplifying host, if one exists, is unknown. Toscana virus infection is common during the summer among rural residents and vacationers, particularly in Italy, Spain, and Portugal; a number of cases have been identified in travelers returning to Germany and Scandinavia. The disease may manifest as an uncomplicated febrile illness but is often associated with aseptic meningitis, with virus isolated from the CSF.

Punta Toro Virus Disease

Of the several phleboviruses that are associated with New World sandflies and infect humans, Punta Toro virus is the best known. The disease caused by this virus is clinically similar to but epidemiologically different from that caused by the Naples or Sicilian sandfly fever viruses. Punta Toro virus infections are sporadic and are acquired in the tropical forest, where the vectors rest on tree buttresses. Epidemics have not been reported, but antibody prevalences among inhabitants of villages in the endemic areas indicate a cumulative lifetime exposure rate of >50%.

Dengue Fever

All four distinct dengue viruses (dengue 1–4) have *Aedes aegypti* as their principal vector, and all cause a similar clinical syndrome. In rare cases, second infection with a serotype of dengue virus different from that involved in the primary infection leads to dengue HF with severe shock (see below). Sporadic cases are seen in the settings of endemic transmission and epidemic disease. Year-round transmission between latitudes 25°N and 25°S has been established, and seasonal forays of the viruses to points as far north as Philadelphia are thought to have taken place in the United States. Dengue fever is seen in the Caribbean region, including Puerto Rico. With increasing spread of the vector mosquito throughout the tropics and subtropics, large areas of the world have become vulnerable to the introduction of dengue viruses, particularly through air travel by infected humans, and both dengue fever and the related dengue HF are becoming increasingly common. Conditions favorable to dengue transmission exist in Hawaii and the southern United States, and bursts of dengue fever activity are to be expected in this region, particularly along the Mexican border, where water may be stored in containers and *A. aegypti* numbers may therefore be greatest. This mosquito, which is also an efficient vector of the yellow fever and chikungunya viruses, typically breeds near human habitation, using relatively fresh water from sources such as water jars, vases, discarded containers, coconut husks, and old tires. *A. aegypti* usually inhabits dwellings and bites during the day. Closed habitations with air-conditioning inhibit transmission of many arboviruses, as has been particularly well illustrated by studies along the Texas-Mexico border. *Aedes albopictus* has now extended its range from Asia to the United States, the Indian Ocean, and parts of Europe. For example, *A. albopictus* has transmitted dengue virus in Hawaii, chikungunya virus in Italy, and chikungunya virus in the Indian Ocean (see below).

After an incubation period of 2–7 days, the typical patient experiences the sudden onset of fever, headache, retroorbital pain, and back pain along with the severe myalgia that gave rise to the colloquial designation "break-bone fever." There is often a macular rash on the first day as well as adenopathy, palatal vesicles, and scleral injection. The illness may last a week, with additional symptoms usually including anorexia, nausea or vomiting, marked cutaneous hypersensitivity, and—near the time of defervescence—a maculopapular rash beginning on the trunk and spreading to the extremities and the face. Epistaxis and scattered petechiae are often noted in uncomplicated dengue, and preexisting gastrointestinal lesions may bleed during the acute illness.

Laboratory findings include leukopenia, thrombocytopenia, and, in many cases, serum aminotransferase elevations. The diagnosis is made by IgM ELISA or paired serology during recovery or by antigen-detection ELISA or RT-PCR during the acute phase. Virus is readily isolated from blood in the acute phase if mosquito inoculation or mosquito cell culture is used.

Colorado Tick Fever

Several hundred cases of Colorado tick fever are reported annually in the United States. The infection is acquired between March and November through the bite of an infected *Dermacentor andersoni* tick in mountainous western regions at altitudes of 1200–3000 m (4000–10,000 ft). Small mammals serve as the amplifying host. The most common presentation consists of fever and myalgia; meningoencephalitis is not uncommon, and hemorrhagic disease, pericarditis, myocarditis, orchitis, and pulmonary presentations are also reported. Rash develops in a substantial minority of cases. The disease usually lasts 7–10 days and is often biphasic. The most important differential diagnostic considerations since the beginning of the twentieth

century have been Rocky Mountain spotted fever and tularemia. In Colorado, Colorado tick fever is much more common than Rocky Mountain spotted fever.

Infection of erythroblasts and other marrow cells by Colorado tick fever virus results in the appearance and persistence (for several weeks) of erythrocytes containing the virus. This feature, detected in smears stained by immunofluorescence, can be diagnostically helpful. The clinical laboratory detects leukopenia and thrombocytopenia.

Orbivirus Infection

The orbiviruses encompass many human and veterinary pathogens. For example, Orungo virus is widely transmitted by mosquitoes in tropical Africa and causes febrile disease in humans. The Kemerovo complex includes the Kemerovo, Lipovnik, and Tribec viruses of Russia and central Europe; these viruses are transmitted by ticks and are associated with febrile and neurologic disease.

ENCEPHALITIS

Arboviral encephalitis is a seasonal disease, commonly occurring in the warmer months. Its incidence varies markedly with time and place, depending on ecologic factors. The causative viruses differ substantially in terms of case-infection ratio (i.e., the ratio of clinical to subclinical infections), mortality rate, and residua **(Table 196-3)**. Humans are not an important amplifier of these viruses.

Table 196-3 Prominent Features of Arboviral Encephalitis

Virus	Natural Cycle	Incubation Period, Days	Annual No. of Cases	Case-to-Infection Ratio	Age of Cases	Case-Fatality Rate, %	Residua
La Crosse	*Aedes triseriatus*–chipmunk (transovarial component in mosquito also important)	~3–7	70 (U.S.)	<1:1000	<15 years	<0.5	Recurrent seizures in ~10%; severe deficits in rare cases; decreased school performance and behavioral change suspected in small proportion
St. Louis	*Culex tarsalis*, *C. pipiens*, *C. quinquefasciatus*–birds	4–21	85, with hundreds to thousands in epidemic years (U.S.)	<1:200	Milder cases in the young; more severe cases in adults >40 years old, particularly the elderly	7	Common in the elderly
Japanese	*Culex tritaeniorhynchus*–birds	5–15	>25,000	1:200–300	All ages; children in highly endemic areas	20–50	Common (approximately half of cases); may be severe
West Nile	*Culex* mosquitoes–birds	3–6	?	Very low	Mainly the elderly	5–10	Uncommon
Central European	*Ixodes ricinus*–rodents, insectivores	7–14	Thousands	1:12	All ages; milder in children	1–5	20%
Russian spring-summer	*I. persulcatus*–rodents, insectivores	7–14	Hundreds	—	All ages; milder in children	20	Approximately half of cases; often severe; limb-girdle paralysis
Powassan	*I. cookei*–wild mammals	~10	~1 (U.S.)	—	All ages; some predilection for children	~10	Common (approximately half of cases)
Eastern equine	*Culiseta melanura*–birds	~5–10	5 (U.S.)	1:40 (adult) 1:17 (child)	All ages; predilection for children	50–75	Common

Western equine	*Culex tarsalis*–birds	~5–10	~20 (U.S.)	1:1000 (adult) 1:50 (child) 1:1 (infant)	All ages; predilection for children <2 years old (increased mortality among elderly)	3–7	Common only among infants <1 year old
Venezuelan equine (epidemic)	Unknown (multiple mosquito species and horses in epidemics)	1–5	?	1:250 (adult) ~1:25 (child)	All ages; predilection for children	~10	—

All the viral encephalitides discussed in this section have a similar pathogenesis as far as is known. An infected arthropod ingests a blood meal from a human and infects the host. The initial period of viremia is thought to originate most commonly from the lymphoid system. Viremia leads to CNS invasion, presumably through infection of olfactory neuroepithelium with passage through the cribriform plate or through infection of brain capillaries and multifocal entry into the CNS. During the viremic phase, there may be little or no recognized disease except in the case of tick-borne flaviviral encephalitis, in which there may be a clearly delineated phase of fever and systemic illness. The disease process in the CNS arises partly from direct neuronal infection and subsequent damage and partly from edema, inflammation, and other indirect effects. The usual pathologic picture is one of focal necrosis of neurons, inflammatory glial nodules, and perivascular lymphoid cuffing; the severity and distribution of these abnormalities vary with the infecting virus. Involved areas display the "luxury perfusion" phenomenon, with normal or increased total blood flow and low oxygen extraction.

The typical patient presents with a prodrome of nonspecific constitutional symptoms, including fever, abdominal pain, vertigo, sore throat, and respiratory symptoms. Headache, meningeal signs, photophobia, and vomiting follow quickly. Involvement of deeper structures may be signaled by lethargy, somnolence, and intellectual deficit (as disclosed by the mental status examination or failure at serial 7 subtraction); more severely affected patients are obviously disoriented and may be comatose. Tremors, loss of abdominal reflexes, cranial nerve palsies, hemiparesis, monoparesis, difficulty in swallowing, and frontal lobe signs are all common. Spinal and motor neuron diseases are documented with West Nile and Japanese encephalitis viruses. Convulsions and focal signs may be evident early or may appear during the course of the disease. Some patients present with an abrupt onset of fever, convulsions, and other signs of CNS involvement. The results of human infection range from no significant symptoms through febrile headache to aseptic meningitis and finally to full-blown encephalitis; the proportions and severity of these manifestations vary with the infecting virus.

The acute encephalitis usually lasts from a few days to as long as 2–3 weeks, but recovery may be slow, with weeks or months required for the return of maximal recoupable function. Difficulty concentrating, fatigability, tremors, and personality changes are common during recovery. The acute illness requires management of a comatose patient who may have intracranial pressure elevations, inappropriate secretion of antidiuretic hormone, respiratory failure, and convulsions. There is no specific therapy for these viral encephalitides. The only practical preventive measures are vector management and personal protection against the arthropod transmitting the virus; for Japanese encephalitis or tick-borne encephalitis, vaccination should be considered in certain circumstances (see relevant sections below).

The diagnosis of arboviral encephalitis depends on the careful evaluation of a febrile patient with CNS disease, with rapid identification of treatable herpes simplex encephalitis, ruling out of brain abscess, exclusion of bacterial meningitis by serial CSF examination, and performance of laboratory studies to define the viral etiology. Leptospirosis, neurosyphilis, Lyme disease, cat-scratch fever, and newer viral encephalitides such as Nipah virus infection from Malaysia and southwestern Asia should be considered. The CSF examination usually shows a modest cell count—in the tens or hundreds or perhaps a few thousand. Early in the process, a significant proportion of these cells may be polymorphonuclear leukocytes, but usually there is a mononuclear cell predominance. CSF glucose levels are generally normal. There are exceptions to this pattern of findings. In eastern equine encephalitis, for example, polymorphonuclear leukocytes may predominate during the first 72 h of disease and hypoglycorrhachia may be detected. In LCM, lymphocyte counts may be in the thousands, and the glucose concentration may be diminished. Experience with imaging studies is still evolving; clearly, however, both CT and MRI may be normal, except for evidence of preexisting conditions, or sometimes may suggest diffuse edema. Several patients with eastern equine encephalitis have had focal abnormalities, and individuals with severe Japanese encephalitis have presented with bilateral thalamic lesions that have often been hemorrhagic. Electroencephalography usually shows diffuse abnormalities and is not directly helpful.

A humoral immune response is usually detectable at or near the onset of disease. Both serum and CSF should be examined for IgM antibodies. Virus generally cannot be isolated from blood or CSF, although Japanese encephalitis virus has been recovered from CSF in severe cases. RT-PCR analysis of CSF may yield positive results. Virus can be obtained from and viral antigen is present in brain tissue, although its distribution may be focal.

California, La Crosse, and Jamestown Canyon Virus Encephalitis

The isolation of California encephalitis virus established the California serogroup of viruses as a cause of encephalitis, and its use as a diagnostic antigen led to the description of many cases of "California encephalitis." In fact, however, this virus has been implicated in only a few cases of encephalitis, and the serologically related La Crosse virus is the major cause of encephalitis among viruses in the California serogroup. "California encephalitis" due to La Crosse virus infection is most commonly reported from the upper Midwest but is also found in other areas of the central and eastern United States, most often in West Virginia, Tennessee, North Carolina, and Georgia. The serogroup includes 13 other viruses, some of which may also be involved in human disease that is misattributed because of the complexity of the group's serology; these viruses include the Jamestown Canyon, snowshoe hare, Inkoo, and Trivittatus viruses, all of which have *Aedes* mosquitoes as their vector and all of which have a strong element of transovarial transmission in their natural cycles.

The mosquito vector of La Crosse virus is *A. triseriatus*. In addition to a prominent transovarial component of transmission, a mosquito can become infected through feeding on viremic chipmunks and other mammals as well as through venereal transmission from another mosquito. The mosquito breeds in sites such as tree holes and abandoned tires and bites during daylight hours. These habits correlate with the risk factors for human cases: recreation in forested areas, residence at the forest's edge, and the presence of abandoned tires around the home. Intensive environmental modification based on these findings has reduced the incidence of disease in a highly endemic area in the Midwest. Most cases occur from July through September. The Asian tiger mosquito, *A. albopictus*, efficiently transmits the virus to mice and also transmits the agent transovarially in the laboratory; this aggressive anthropophilic mosquito has the capacity to urbanize, and its possible impact on transmission to humans is of concern.

An antibody prevalence of ≳20% in endemic areas indicates that infection is common, but CNS disease has been recognized primarily in children <15 years of age. The illness varies from a picture of aseptic meningitis accompanied by confusion to severe and occasionally fatal encephalitis. Although there may be prodromal symptoms, the onset of CNS disease is sudden, with fever, headache, and lethargy often joined by nausea and vomiting, convulsions (in one-half of patients), and coma (in one-third of patients). Focal seizures, hemiparesis, tremor, aphasia, chorea, Babinski signs, and other evidence of significant neurologic dysfunction are common, but residua are not. Perhaps 10% of patients have recurrent seizures in the succeeding months. Other serious sequelae are rare, although a decrease in scholastic standing has been reported and mild personality change has occasionally been suggested. Treatment is supportive over a 1- to 2-week acute phase during which status epilepticus, cerebral edema, and inappropriate secretion of antidiuretic hormone are important concerns. Ribavirin has been used in severe cases, and a clinical trial of this drug is under way.

The blood leukocyte count is commonly elevated, sometimes reaching levels of 20,000/μL, and there is usually a left shift. CSF cell counts are typically 30–500/μL with a mononuclear cell predominance (although 25–90% of cells are polymorphonuclear in some cases). The protein level is normal or slightly increased, and the glucose level is normal. Specific virologic diagnosis based on IgM-capture assays of serum and CSF is efficient. The only human anatomic site from which virus has been isolated is the brain.

Jamestown Canyon virus has been implicated in several cases of encephalitis in adults; in these cases, the disease was usually associated with a significant respiratory illness at onset. Human infection with this virus has been documented in New York, Wisconsin, Ohio, Michigan, Ontario, and other areas of North America where the vector mosquito, *A. stimulans*, feeds on its main host, the white-tailed deer.

ST. Louis Encephalitis

St. Louis encephalitis virus is transmitted between *Culex* mosquitoes and birds. This virus causes low-level endemic infection among rural residents of the western and central United States, where *C. tarsalis* is the vector (see "Western Equine Encephalitis," below), but the more urbanized mosquito species *C. pipiens* and *C. quinquefasciatus* have been responsible for epidemics resulting in hundreds or even thousands of cases in cities of the central and eastern United States. Most cases occur in June through October. The urban mosquitoes breed in accumulations of stagnant water and sewage with high organic content and readily bite humans in and around houses at dusk. The elimination of open sewers and trash-filled drainage systems is expensive and may not be possible, but screening of houses and implementation of personal protective measures may be an effective approach for individuals. The rural vector is most active at dusk and outdoors; its bites can be avoided by modification of activities and use of repellents.

Disease severity increases with age: infections that result in aseptic meningitis or mild encephalitis are concentrated in children and young adults, while severe and fatal cases primarily affect the elderly. Infection rates are similar in all age groups; thus the greater susceptibility of older persons to disease is a biologic consequence of aging. The disease has an abrupt onset, sometimes following a prodrome, and begins with fever, lethargy, confusion, and headache. In addition, nuchal rigidity, hypotonia, hyperreflexia, myoclonus, and tremor are common. Severe cases can include cranial nerve palsies, hemiparesis, and convulsions. Patients often report dysuria and may have viral antigen in urine as well as pyuria. The overall mortality rate is generally ~7% but may reach 20% among patients over the age of 60. Recovery is slow. Emotional lability, difficulties in concentration and memory, asthenia, and tremor are commonly prolonged in older patients.

The CSF of patients with St. Louis encephalitis usually contains tens to hundreds of cells, with a lymphocytic predominance and a normal glucose level. Leukocytosis with a left shift is often documented.

Japanese Encephalitis

Japanese encephalitis virus is found throughout Asia, including far eastern Russia, Japan, China, India, Pakistan, and Southeast Asia, and causes occasional epidemics on western Pacific islands. The virus has

been detected in the Torres Strait islands, and a human encephalitis case has been identified on the nearby Australian mainland. This flavivirus is particularly common in areas where irrigated rice fields attract the natural avian vertebrate hosts and provide abundant breeding sites for mosquitoes such as *C. tritaeniorhynchus*, which transmit the virus to humans. Additional amplification by pigs, which suffer abortion, and horses, which develop encephalitis, may be significant as well. Vaccination of these additional amplifying hosts may reduce the transmission of the virus. An effective, formalin-inactivated vaccine purified from mouse brain is produced in Japan and licensed for human use in the United States. It is given on days 0, 7, and 30 or—with some sacrifice in serum neutralizing titer—on days 0, 7, and 14. Vaccination is indicated for summer travelers to rural Asia, where the risk of clinical disease may be 0.05–2.1/10,000 per week. The severe and often fatal disease reported in expatriates must be balanced against the 0.1–1% chance of a late systemic or cutaneous allergic reaction. These reactions are rarely fatal but may be severe and have been known to begin 1–9 days after vaccination, with associated pruritus, urticaria, and angioedema. Live attenuated vaccines are being used in China but are not recommended in the United States at this time.

West Nile Virus Infection

West Nile virus was initially described as being transmitted among wild birds by *Culex* mosquitoes in Africa, the Middle East, southern Europe, and Asia. It is a common cause of febrile disease without CNS involvement, but it occasionally causes aseptic meningitis and severe encephalitis; these serious infections are particularly common among the elderly. The febrile-myalgic syndrome caused by West Nile virus differs from many others by the frequent appearance of a maculopapular rash concentrated on the trunk and lymphadenopathy. Headache, ocular pain, sore throat, nausea and vomiting, and arthralgia (but not arthritis) are common accompaniments. In addition, the virus has been implicated in severe and fatal hepatic necrosis in Africa.

West Nile virus was introduced into New York City in 1999 and subsequently spread to other areas of the northeastern United States, causing >60 cases of aseptic meningitis or encephalitis among humans as well as die-offs among crows, exotic zoo birds, and other birds. The virus has continued to spread and is now found in almost all states as well as in Canada, Mexico, South America, and the Caribbean. *C. pipiens* remains the major vector in the northeastern United States, but several other *Culex* species are also involved, and blue jays compete with crows as amplifiers and lethal targets in other areas of the country. Annually, ~1000–3000 cases of encephalitis with ~100–300 deaths are reported in the United States. The ratio of CNS involvement to infection is thought to be ~1:100; the remainder of patients have subclinical infection or West Nile fever. Encephalitis, sequelae, and death are all more common among elderly, diabetic, and hypertensive patients and among patients with previous CNS insults. In addition to the more severe motor and cognitive sequelae, milder findings may include tremor, slight abnormalities in motor skills, and loss of executive functions. Intense clinical interest and the availability of laboratory diagnostic methods have made it possible to define a number of unusual clinical features, including chorioretinitis, flaccid paralysis with histologic lesions resembling poliomyelitis, and initial presentation with fever and focal neurologic deficits in the absence of diffuse encephalitis. Immunosuppressed patients may have fulminant courses or develop persistent CNS infection. Virus transmission through both transplantation and blood transfusion has necessitated screening of blood and organ donors by nucleic acid–based tests. Pregnant women may occasionally infect the fetus. All these "new" findings, particularly transfusion-transmitted disease and severe disease associated with transplantation, show what could be expected in the United States should other arboviruses be transmitted in North America with a high frequency or should surveillance in other areas of the world be more efficient and deeper.

West Nile virus falls into the same phylogenetic group of flaviviruses as St. Louis and Japanese encephalitis viruses, as do Murray Valley and Rocio viruses. The latter two viruses are both maintained in mosquitoes and birds and produce a clinical picture resembling that of Japanese encephalitis. Murray Valley virus has caused occasional epidemics and sporadic cases in Australia. Rocio virus caused recurrent epidemics in a focal area of Brazil in 1975–1977 and then virtually disappeared.

Central European Tick-Borne Encephalitis and Russian Spring-Summer Encephalitis

A spectrum of tick-borne flaviviruses has been identified across the Eurasian land mass. Many are known mainly as agricultural pathogens (e.g., louping ill virus in the United Kingdom). From Scandinavia to the Urals, central European tick-borne encephalitis is transmitted by *Ixodes ricinus*. Human cases occur between April and October, with a peak in June and July. A related and more virulent virus is that of Russian spring-summer encephalitis, which is associated with *I. persulcatus* and is distributed from Europe across the Urals to the Pacific Ocean. The ticks transmit the disease primarily in the spring and early summer, with a lower rate of transmission later in summer. Small mammals are the vertebrate amplifiers for both viruses. The risk varies by geographic area and can be highly localized within a given area; human cases usually follow outdoor activities or consumption of raw milk from infected goats or other infected animals.

After an incubation period of 7–14 days or perhaps longer, the central European viruses classically result in a febrile-myalgic phase that lasts for 2–4 days and is thought to correlate with viremia. A subsequent remission for several days is followed by the recurrence of fever and the onset of meningeal signs. The CNS phase varies from mild aseptic meningitis, which is more common among younger patients, to severe encephalitis with coma, convulsions, tremors, and motor signs lasting for 7–10 days before improvement begins. Spinal and medullary involvement can lead to typical limb-girdle paralysis and to respiratory paralysis. Most patients recover, only a minority with significant deficits. Infections with the Far Eastern viruses generally run a more abrupt course. The encephalitic syndrome caused by these viruses sometimes begins without a remission and has more severe manifestations than the European syndrome. The mortality rate is high, and major sequelae—most notably, lower motor neuron paralyses of the proximal muscles of the extremities, trunk, and

neck—are common.

In the early stage of the illness, virus may be isolated from the blood. In the CNS phase, IgM antibodies are detectable in serum and/or CSF. Thrombocytopenia sometimes develops during the initial febrile illness, which resembles the early hemorrhagic phase of some other tick-borne flaviviral infections, such as Kyasanur Forest disease. Other tick-borne flaviviruses are less common causes of encephalitis, including louping ill virus in the United Kingdom and Powassan virus.

There is no specific therapy for infection with these viruses. However, effective alum-adjuvanted, formalin-inactivated vaccines are produced in Austria, Germany, and Russia. Two doses of the Austrian vaccine separated by an interval of 1–3 months appear to be effective in the field, and antibody responses are similar when vaccine is given on days 0 and 14. Other vaccines have elicited similar neutralizing antibody titers. Since rare cases of postvaccination Guillain-Barré syndrome have been reported, vaccination should be reserved for persons likely to experience rural exposure in an endemic area during the season of transmission. Cross-neutralization for the central European and Far Eastern strains has been established, but there are no published field studies on cross-protection of formalin-inactivated vaccines. Because 0.2–4% of ticks in endemic areas may be infected, tick bites raise the issue of immunoglobulin prophylaxis. Prompt administration of high-titered specific preparations should probably be undertaken, although no controlled data are available to prove the efficacy of this measure. Immunoglobulin should not be administered late because of the risk of antibody-mediated enhancement.

Powassan Encephalitis

Powassan virus is a member of the tick-borne encephalitis virus complex and is transmitted by *I. cookei* among small mammals in eastern Canada and the United States, where it has been responsible for 20 recognized cases of human disease. Other ticks may transmit the virus in a wider geographic area, and there is some concern that *I. scapularis* (also called *I. dammini*), a competent vector in the laboratory, may become involved as it becomes more prominent in the United States. Patients with Powassan encephalitis (many of whom are children) present in May through December after outdoor exposure and an incubation period thought to be ~1 week. Powassan encephalitis is severe, and sequelae are common.

Eastern Equine Encephalitis

Eastern equine encephalitis is found primarily within endemic swampy foci along the eastern coast of the United States, with a few inland foci as far removed as Michigan. Human cases present from June through October, when the bird–*Culiseta* mosquito cycle spills over into other mosquito species such as *A. sollicitans* or *A. vexans*, which are more likely to bite mammals. There is concern over the potential role of the introduced anthropophilic mosquito species *A. albopictus*, which has been found to be naturally infected and is an effective vector in the laboratory. Horses are a common target for the virus; contact with unvaccinated horses may be associated with human disease, but horses probably do not play a significant role in amplification of the virus.

Eastern equine encephalitis is one of the most destructive of the arboviral conditions, with a brusque onset, rapid progression, high mortality rate, and frequent residua. This severity is reflected in the extensive necrotic lesions and polymorphonuclear infiltrates found at postmortem examination of the brain and the acute polymorphonuclear CSF pleocytosis often occurring during the first 1–3 days of disease. In addition, leukocytosis with a left shift is a common feature. A formalin-inactivated vaccine has been used to protect laboratory workers but is not generally available or applicable.

Western Equine Encephalitis

The primary maintenance cycle for western equine encephalitis virus in the United States is between *C. tarsalis* and birds, principally sparrows and finches. Equines and humans become infected, and both species suffer encephalitis without amplifying the virus in nature. St. Louis encephalitis is transmitted in a similar cycle in the same region but causes human disease about a month earlier than the period (July through October) in which western equine encephalitis virus is active. Large epidemics of western equine encephalitis took place in the western and central United States and Canada during the 1930s to 1950s, but in recent years the disease has been uncommon. There were 41 reported cases in the United States in 1987 but only 5 reported cases from 1988 to 2001. This decline in incidence may reflect in part the integrated approach to mosquito management that has been employed in irrigation projects and the increasing use of agricultural pesticides; it almost certainly reflects the increased tendency for humans to be indoors behind closed windows at dusk—the peak period of biting by the major vector.

Western equine encephalitis virus causes a typical diffuse viral encephalitis with an increased attack rate and increased morbidity rate among the young, particularly children <2 years old. In addition, mortality rates are high among the young and the very elderly. One-third of individuals who have convulsions during the acute illness have subsequent seizure activity. Infants <1 year old—particularly those in the first months of life—are at serious risk of motor and intellectual damage. Twice as many males as females develop clinical encephalitis after 5–9 years of age; this difference may be related to greater outdoor exposure of boys to the vector but is also likely to be due in part to biologic differences. A formalin-inactivated vaccine has been used to protect laboratory workers but is not generally available or applicable.

Venezuelan Equine Encephalitis

There are six known types of virus in the Venezuelan equine encephalitis complex. An important distinction is between the *epizootic* viruses (subtypes IAB and IC) and the *enzootic* viruses (subtypes ID to IF and types II to VI). The epizootic viruses have an unknown natural cycle but periodically cause extensive epidemics in equines and humans in the Americas. These epidemics rely on the high-level viremia in horses and mules

that results in the infection of several species of mosquitoes, which in turn infect humans and perpetuate virus transmission. Humans also have high-level viremia but probably are not important in virus transmission. Enzootic viruses are found primarily in humid tropical forest habitats and are maintained between *Culex* mosquitoes and rodents; these viruses cause human disease but are not pathogenic for horses and do not cause epizootics.

Epizootics of Venezuelan equine encephalitis occurred repeatedly in Venezuela, Colombia, Ecuador, Peru, and other South American countries at intervals of ≥10 years from the 1930s until 1969, when a massive epizootic spread throughout Central America and Mexico, reaching southern Texas in 1972. Genetic sequencing of the virus from the 1969–1972 outbreak suggested that it originated from residual "un-inactivated" virus in veterinary vaccines. The outbreak was terminated in Texas with the use of a live attenuated vaccine (TC-83) originally developed for human use by the U.S. Army; the epizootic virus was then used for further production of inactivated veterinary vaccines. No further epizootic disease was identified until 1995 and subsequently, when additional epizootics took place in Colombia, Venezuela, and Mexico. The viruses involved in these epizootics as well as previously epizootic subtype IC viruses have been shown to be close phylogenetic relatives of known enzootic subtype ID viruses. This finding suggests that active evolution and selection of epizootic viruses are under way in northern South America.

During epizootics, extensive human infection is the rule, with clinical disease in 10–60% of infected individuals. Most infections result in notable acute febrile disease, while relatively few result in encephalitis. A low rate of CNS invasion is supported by the absence of encephalitis among the many infections resulting from exposure to aerosols in the laboratory or from vaccine accidents. The most recent large epizootic of Venezuelan equine encephalitis occurred in Colombia and Venezuela in 1995; of the >85,000 clinical cases, 4% (with a higher proportion among children than adults) included neurologic symptoms and 300 ended in death.

Enzootic strains of Venezuelan equine encephalitis virus are common causes of acute febrile disease, particularly in areas such as the Florida Everglades and the humid Atlantic coast of Central America. Encephalitis has been documented only in the Florida infections; the three cases were caused by type II enzootic virus, also called *Everglades virus*. All three patients had preexisting cerebral disease. Extrapolation from the rate of genetic change suggests that Everglades virus may have been introduced into Florida <200 years ago and that it is most closely related to the ID subtypes that appear to have given evolutionary rise to the epizootic strains active in South America.

The prevention of epizootic Venezuelan equine encephalitis depends on vaccination of horses with the attenuated TC-83 vaccine or with an inactivated vaccine prepared from that strain. Humans can be protected with similar vaccines, but the use of such products is restricted to laboratory personnel because of reactogenicity and limited availability. In addition, wild-type virus and perhaps TC-83 vaccine may have some degree of fetal pathogenicity. Enzootic viruses are genetically and antigenically different from epizootic viruses, and protection against the former with vaccines prepared from the latter is relatively ineffective.

ARTHRITIS AND RASH

True arthritis is a common accompaniment of several viral diseases, such as rubella (caused by a non-alphavirus togavirus), parvovirus B19 infection, and hepatitis B; it is an occasional accompaniment of infection due to mumps virus, enteroviruses, herpesviruses, and adenoviruses. It is not generally appreciated that the alphaviruses are also common causes of arthritis. In fact, the alphaviruses discussed below all cause acute febrile diseases accompanied by the development of true arthritis and a maculopapular rash. Rheumatic involvement includes arthralgia alone, periarticular swelling, and (less commonly) joint effusions. Most of these diseases are less severe and have fewer articular manifestations in children than in adults. In temperate climates, these are summer diseases. No specific therapy or licensed vaccines exist.

Sindbis Virus Infection

Sindbis virus is transmitted among birds by mosquitoes. Infections with the northern European strains of this virus (which cause, for example, Pogosta disease in Finland, Karelian fever in the independent states of the former Soviet Union, and Ockelbo disease in Sweden) and with the genetically related southern African strains are particularly likely to result in the arthritis-rash syndrome. Exposure to a rural environment is commonly associated with this infection, which has an incubation period of <1 week.

The disease begins with rash and arthralgia. Constitutional symptoms are not marked, and fever is modest or lacking altogether. The rash, which lasts ~1 week, begins on the trunk, spreads to the extremities, and evolves from macules to papules that often vesiculate. The arthritis of this condition is multiarticular, migratory, and incapacitating, with resolution of the acute phase in a few days. Wrists, ankles, phalangeal joints, knees, elbows, and—to a much lesser extent—proximal and axial joints are involved. Persistence of joint pains and occasionally of arthritis is a major problem and may go on for months or even years despite a lack of deformity.

Chikungunya Virus Infection

It is likely that chikungunya virus ("that which bends up") is of African origin and is maintained among nonhuman primates on that continent by *Aedes* mosquitoes of the subgenus *Stegomyia* in a fashion similar to yellow fever virus. Like yellow fever virus, chikungunya virus is readily transmitted among humans in urban areas by *A. aegypti*. The *A. aegypti*–chikungunya virus transmission cycle has also been introduced into Asia, where it poses a prominent health problem. The disease is endemic in rural areas of Africa, and intermittent epidemics take place in towns and cities of Africa and Asia. In 2004, a massive epidemic in the Indian Ocean region began; it now appears to have been spread by travelers. *A. albopictus* was identified as

the major vector, and there were multiple exportations to temperate zones and to areas where *A. aegypti* is present. Chikungunya is one more reason (in addition to dengue and yellow fever) that *A. aegypti* and *A. albopictus* must be controlled.

Full-blown disease is most common among adults, in whom the clinical picture may be dramatic. The abrupt onset follows an incubation period of 2–3 days. Fever and severe arthralgia are accompanied by chills and constitutional symptoms such as headache, photophobia, conjunctival injection, anorexia, nausea, and abdominal pain. Migratory polyarthritis mainly affects the small joints of the hands, wrists, ankles, and feet, with lesser involvement of the larger joints. Rash may appear at the outset or several days into the illness; its development often coincides with defervescence, which takes place around day 2 or 3 of disease. The rash is most intense on the trunk and limbs and may desquamate. Petechiae are occasionally seen, and epistaxis is not uncommon, but this virus is not a regular cause of the HF syndrome, even in children. A few patients develop leukopenia. Elevated levels of aspartate aminotransferase (AST) and C-reactive protein have been described, as have mildly decreased platelet counts. Recovery may require weeks. Some older patients continue to experience stiffness, joint pain, and recurrent effusions for several years; this persistence may be especially common in HLA-B27 patients. An investigational live attenuated vaccine has been developed but requires additional testing. It appears to be headed for further development and commercial manufacture stimulated by the Indian Ocean outbreak.

A related virus, O'nyong-nyong, caused a major epidemic of arthritis and rash involving at least 2 million people as it moved across eastern and central Africa in the 1960s. After its mysterious emergence, the virus virtually disappeared, leaving only occasional evidence of its persistence in Kenya until a transient resurgence of epidemic activity in 1997.

Mayaro Fever

Mayaro virus is maintained in the forests of the Americas by *Haemagogus* mosquitoes and nonhuman primates. It causes a frequently endemic and sometimes epidemic infection of humans and appears to produce a syndrome resembling chikungunya virus infection.

Epidemic Polyarthritis (Ross River Virus Infection)

Ross River virus has caused epidemics of distinctive clinical disease in Australia since the beginning of the twentieth century and continues to be responsible for thousands of cases in rural and suburban areas annually. The virus is transmitted by *A. vigilax* and other mosquitoes, and its persistence is thought to involve transovarial transmission. No definitive vertebrate host has been identified, but several mammalian species, including wallabies, have been suggested. Endemic transmission has also been documented in New Guinea, and in 1979 the virus swept through the eastern Pacific Islands, causing hundreds of thousands of illnesses. The virus was carried from island to island by infected humans and was believed to have been transmitted among humans by *A. polynesiensis* and *A. aegypti*.

The incubation period is 7–11 days long, and the onset of illness is sudden, with joint pain usually ushering in the disease. The rash generally develops coincidentally or follows shortly but in some cases precedes joint pains by several days. Constitutional symptoms such as low-grade fever, asthenia, myalgia, headache, and nausea are not prominent and indeed are absent in many cases. Most patients are incapacitated for considerable periods by joint involvement, which interferes with sleeping, walking, and grasping. Wrist, ankle, metacarpophalangeal, interphalangeal, and knee joints are the most commonly involved, although toes, shoulders, and elbows may be affected with some frequency. Periarticular swelling and tenosynovitis are common, and one-third of patients have true arthritis. Only half of all arthritis patients can resume normal activities within 4 weeks, and 10% still must limit their activity at 3 months. Occasional patients are symptomatic for 1–3 years but without progressive arthropathy. Aspirin and nonsteroidal anti-inflammatory drugs are effective for the treatment of symptoms.

Clinical laboratory values are normal or variable in Ross River virus infection. Tests for rheumatoid factor and antinuclear antibodies are negative, and the erythrocyte sedimentation rate is acutely elevated. Joint fluid contains 1000–60,000 mononuclear cells/μL, and Ross River virus antigen is demonstrable in macrophages. IgM antibodies are valuable in the diagnosis of this infection, although they occasionally persist for years. The isolation of the virus from blood by mosquito inoculation or mosquito cell culture is possible early in the illness. Because of the great economic impact of annual epidemics in Australia, an inactivated vaccine is being developed and has been found to be protective in mice.

Perhaps because of the local interest in arboviruses in general and in Ross River virus in particular, other arthritogenic arboviruses have been identified in Australia, including Gan Gan virus, a member of the family Bunyaviridae; Kokobera virus, a flavivirus; and Barmah Forest virus, an alphavirus. The last virus is a common cause of infection and must be differentiated from Ross River virus by specific testing.

HEMORRHAGIC FEVERS

The viral HF syndrome is a constellation of findings based on vascular instability and decreased vascular integrity. An assault, direct or indirect, on the microvasculature leads to increased permeability and (particularly when platelet function is decreased) to actual disruption and local hemorrhage. Blood pressure is decreased, and in severe cases shock supervenes. Cutaneous flushing and conjunctival suffusion are examples of common, observable abnormalities in the control of local circulation. The hemorrhage is inconstant and is in most cases an indication of widespread vascular damage rather than a life-threatening loss of blood volume. Disseminated intravascular coagulation (DIC) is occasionally found in any severely ill patient with HF but is thought to occur regularly only in the early phases of HF with renal syndrome, Crimean-Congo HF, and perhaps some cases of filovirus HF. In some viral HF syndromes, specific organs may be

particularly impaired, such as the kidney in HF with renal syndrome, the lung in hantavirus pulmonary syndrome, or the liver in yellow fever, but in all these diseases the generalized circulatory disturbance is critically important.

The pathogenesis of HF is poorly understood and varies among the viruses regularly implicated in the syndrome, which number more than a dozen. In some cases direct damage to the vascular system or even to parenchymal cells of target organs is important, whereas in others soluble mediators are thought to play the major role. The acute phase in most cases of HF is associated with ongoing virus replication and viremia. Exceptions are the hantavirus diseases and dengue HF/dengue shock syndrome (DHF/DSS), in which the immune response plays a major pathogenic role.

The HF syndromes all begin with fever and myalgia, usually of abrupt onset. Within a few days the patient presents for medical attention because of increasing prostration that is often accompanied by severe headache, dizziness, photophobia, hyperesthesia, abdominal or chest pain, anorexia, nausea or vomiting, and other gastrointestinal disturbances. Initial examination often reveals only an acutely ill patient with conjunctival suffusion, tenderness to palpation of muscles or abdomen, and borderline hypotension or postural hypotension, perhaps with tachycardia. Petechiae (often best visualized in the axillae), flushing of the head and thorax, periorbital edema, and proteinuria are common. Levels of AST are usually elevated at presentation or within a day or two thereafter. Hemoconcentration from vascular leakage, which is usually evident, is most marked in hantavirus diseases and in DHF/DSS. The seriously ill patient progresses to more severe symptoms and develops shock and other findings typical of the causative virus. Shock, multifocal bleeding, and CNS involvement (encephalopathy, coma, convulsions) are all poor prognostic signs.

One of the major diagnostic clues is travel to an endemic area within the incubation period for a given syndrome **(Table 196-4)**. Except for Seoul, dengue, and yellow fever virus infections, which have urban vectors, travel to a rural setting is especially suggestive of a diagnosis of HF.

Table 196-4 Viral Hemorrhagic Fever (HF) Syndromes and Their Distribution

Disease	Incubation Period, Days	Case-Infection Ratio	Case-Fatality Rate, %	Geographic Range	Target Population
Lassa fever	5–16	Mild infections probably common	15	West Africa	All ages, both sexes
South American HF	7–14	Most infections (more than half) result in disease	15–30	Selected rural areas of Bolivia, Argentina, Venezuela, and Brazil	Bolivia: Men in countryside; all ages, both sexes in villages Argentina: All ages, both sexes; excess exposure and disease in men Venezuela: All ages, both sexes
Rift Valley fever	2–5	~1:100[a]	~50	Sub-Saharan Africa, Madagascar, Egypt	All ages, both sexes; more often diagnosed in men; preexisting liver disease may predispose
Crimean-Congo HF	3–12	≥1:5	15–30	Africa, Middle East, Turkey, Balkans, southern region of former Soviet Union, western China	All ages, both sexes; men more exposed in some settings
HF with renal syndrome	9–35	Hantaan, >1:1.25; Puumala, 1:20	Hantaan, 5–15; Puumala, <1	Worldwide, depending on rodent reservoir	Excess of male patients (partially due to greater exposure); mainly adults
Hantavirus pulmonary syndrome	~7–28	Very high	40–50	Americas	Excess of male patients due to some occupational exposure; mainly adults
Marburg or Ebola HF	3–16	High	25–90	Sub-Saharan Africa	All ages, both sexes; children less exposed
Yellow fever	3–6	1:2–1:20	20	Africa, South America	All ages, both sexes; adults more exposed in jungle setting; preexisting flavivirus immunity may cross-protect
Dengue HF/dengue shock syndrome	2–7	Nonimmune, 1:10,000; heterologous immune, 1:100	<1 with supportive treatment	Tropics and subtropics worldwide	Predominantly children; previous heterologous dengue infection predisposes to HF
Kyasanur Forest/Omsk	3–8	Variable	0.5–10	Mysore State, India/western Siberia	Variable

HF				

[a]Figure is for HF cases only. Most infections with Rift Valley fever virus result in fever and myalgia rather than HF.

Early recognition is important because of the need for virus-specific therapy and supportive measures, including prompt, atraumatic hospitalization; judicious fluid therapy that takes into account the patient's increased capillary permeability; administration of cardiotonic drugs; use of pressors to maintain blood pressure at levels that will support renal perfusion; treatment of the relatively common secondary bacterial infections; replacement of clotting factors and platelets as indicated; and the usual precautionary measures used in the treatment of patients with hemorrhagic diatheses. DIC should be treated only if clear laboratory evidence of its existence is found and if laboratory monitoring of therapy is feasible; there is no proven benefit of such therapy. The available evidence suggests that HF patients have a decreased cardiac output and will respond poorly to fluid loading as it is often practiced in the treatment of shock associated with bacterial sepsis. Specific therapy is available for several of the HF syndromes. In addition, several diseases considered in the differential diagnosis—malaria, shigellosis, typhoid fever, leptospirosis, relapsing fever, and rickettsial diseases—are treatable and potentially lethal. Strict barrier nursing and other precautions against infection of medical staff and visitors are indicated in HF except that due to hantaviruses, yellow fever, Rift Valley fever, and dengue.

Lassa Fever

Lassa virus is known to cause endemic and epidemic disease in Nigeria, Sierra Leone, Guinea, and Liberia, although it is probably more widely distributed in West Africa. This virus and its relatives exist elsewhere in Africa, but their health significance is unknown. Like other arenaviruses, Lassa virus is spread to humans by small-particle aerosols from chronically infected rodents and may also be acquired during the capture or eating of these animals. It can be transmitted by close person-to-person contact. The virus is often present in urine during convalescence and is suspected to be present in seminal fluid early in recovery. Nosocomial spread has occurred but is uncommon if proper sterile parenteral techniques are used. Individuals of all ages and both sexes are affected; the incidence of disease is highest in the dry season, but transmission takes place year-round. In countries where Lassa virus is endemic, Lassa fever can be a prominent cause of febrile disease. For example, in one hospital in Sierra Leone, laboratory-confirmed Lassa fever is consistently responsible for one-fifth of admissions to the medical wards. There are probably tens of thousands of Lassa fever cases annually in West Africa alone. New arenaviruses continue to be discovered, often without being thoroughly characterized. These arenaviruses include Chapare virus in Bolivia (distinct from Machupo virus) and Lujo virus in Zambia.

Among the HF agents, only the arenaviruses are typically associated with a gradual onset of illness. The average case of Lassa fever has a gradual onset that gives way to more severe constitutional symptoms and prostration. Bleeding is seen in only ~15–30% of cases. A maculopapular rash is often noted in light-skinned Lassa patients. Effusions are common, and male-dominant pericarditis may develop late. The fetal death rate is 92% in the last trimester, when the maternal mortality rate is also increased from the usual 15–30%; these figures suggest that interruption of the pregnancy of infected women should be considered. White blood cell counts are normal or slightly elevated, and platelet counts are normal or somewhat low. Deafness coincides with clinical improvement in ~20% of cases and is permanent and bilateral in some. Reinfection may occur but has not been associated with severe disease.

High-level viremia or a high serum concentration of AST statistically predicts a fatal outcome. Thus patients with an AST level of >150 IU/mL should be treated with IV ribavirin. This antiviral nucleoside analogue appears to be effective in reducing mortality rates from the levels documented among retrospective controls, and its only major side effect is reversible anemia that usually does not require transfusion. The drug should be given by slow IV infusion in a dose of 32 mg/kg; this dose should be followed by 16 mg/kg every 6 h for 4 days and then by 8 mg/kg every 8 h for 6 days.

South American HF Syndromes (Argentine, Bolivian, Venezuelan, and Brazilian)

These diseases are similar to one another clinically, but their epidemiology differs with the habits of their rodent reservoirs and the interactions of these animals with humans. Person-to-person or nosocomial transmission is rare but has occurred.

The basic disease resembles Lassa fever, with two marked differences. First, thrombocytopenia—often marked—is the rule, and bleeding is quite common. Second, CNS dysfunction is much more common than in Lassa fever and is often manifested by marked confusion, tremors of the upper extremities and tongue, and cerebellar signs. Some cases follow a predominantly neurologic course, with a poor prognosis. The clinical laboratory is helpful in diagnosis since thrombocytopenia, leukopenia, and proteinuria are typical findings.

Argentine HF is readily treated with convalescent-phase plasma given within the first 8 days of illness. In the absence of passive antibody therapy, IV ribavirin in the dose recommended for Lassa fever is likely to be effective in all the South American HF syndromes. The transmission of the disease from men convalescing from Argentine HF to their wives suggests the need for counseling of arenavirus HF patients concerning the avoidance of intimate contacts for several weeks after recovery. A safe, effective, live attenuated vaccine exists for Argentine HF; after vaccination of >250,000 high-risk persons in the endemic area, incidence decreased markedly. In experimental animals, this vaccine is cross-protective against the Bolivian HF virus.

Rift Valley Fever

The mosquito-borne Rift Valley fever virus is also a pathogen of domestic animals such as sheep, cattle, and

goats. It is maintained in nature by transovarial transmission in floodwater *Aedes* mosquitoes and presumably also has a vertebrate amplifier. Epizootics and epidemics occur when sheep or cattle become infected during particularly heavy rains; developing high-level viremia, these animals infect many species of mosquitoes. Remote sensing via satellite can detect the ecologic changes associated with high rainfall that predict the likelihood of Rift Valley fever transmission; it can also detect the special depressions from which the floodwater *Aedes* mosquito vectors emerge. In addition, the virus is infectious when transmitted by contact with blood or aerosols from domestic animals or their abortuses. The slaughtered meat is not infectious; anaerobic glycolysis in postmortem tissues results in an acidic environment that rapidly inactivates Bunyaviridae such as Rift Valley fever virus and Crimean-Congo HF virus. The natural range of Rift Valley fever virus is confined to sub-Saharan Africa, where its circulation is markedly enhanced by substantial rainfall such as that which occurred during the El Niño phenomenon of 1997; subsequent spread to the Arabian Peninsula caused epidemic disease in 2000. The virus has also been found in Madagascar and has been introduced into Egypt, where it caused major epidemics in 1977–1979, 1993, and subsequently. Neither person-to-person nor nosocomial transmission has been documented.

Rift Valley fever virus is unusual in that it causes several clinical syndromes. Most infections are manifested as the febrile-myalgic syndrome. A small proportion of infections result in HF with especially prominent liver involvement. Renal failure and probably DIC are also common features. Perhaps 10% of otherwise mild infections lead to retinal vasculitis; funduscopic examination reveals edema, hemorrhages, and infarction, and some patients have permanently impaired vision. A small proportion of cases (<1 in 200) are followed by typical viral encephalitis. One of the complicated syndromes does not appear to predispose to another.

There is no proven therapy for any of the syndromes described above. Both retinal disease and encephalitis occur after the acute febrile syndrome has ended and serum neutralizing antibody has developed—events suggesting that only supportive care need be given. Epidemic disease is best prevented by vaccination of livestock. The established ability of this virus to propagate after an introduction into Egypt suggests that other potentially receptive areas, including the United States, should have a response ready for such an eventuality. It seems likely that this disease, like Venezuelan equine encephalitis, can be controlled only with adequate stocks of an effective live attenuated vaccine, and there are no such global stocks. A formalin-inactivated vaccine confers immunity to humans, but quantities are limited and three injections are required; this vaccine is recommended for exposed laboratory workers and for veterinarians working in sub-Saharan Africa. A new live attenuated vaccine, MP-12, is being tested in humans and may soon become available for general use.

Crimean-Congo HF

This severe HF syndrome has a wide geographic distribution, potentially being found wherever ticks of the genus *Hyalomma* occur. The propensity of these ticks to feed on domestic livestock and certain wild mammals means that veterinary serosurveys are the most effective mechanism for the surveillance of virus circulation in a region. Human infection is acquired via a tick bite or during the crushing of infected ticks. Domestic animals do not become ill but do develop viremia; thus, there is danger of infection at the time of slaughter and for a brief interval thereafter (through contact with hides or carcasses). Cases have followed sheep shearing. An epidemic in South Africa was associated with slaughter of tick-infested ostriches. Nosocomial epidemics are common and are usually related to extensive blood exposure or needle sticks.

Although generally similar to other HF syndromes, Crimean-Congo HF causes extensive liver damage, resulting in jaundice in some cases. Clinical laboratory values indicate DIC and show elevations in AST, creatine phosphokinase, and bilirubin. Patients with fatal cases generally have more marked changes, even in the early days of illness, and also develop leukocytosis rather than leukopenia. In addition, thrombocytopenia is more marked and develops earlier in cases with a fatal outcome.

No controlled trials have been performed with IV ribavirin, but clinical experience and retrospective comparison of patients with ominous clinical laboratory values suggest that ribavirin is efficacious and should be given. No human or veterinary vaccines are recommended.

HF with Renal Syndrome

This disease, the first to be identified as an HF, is widely distributed over Europe and Asia; the major causative viruses and their rodent reservoirs on these two continents are Puumala virus (bank vole, *Clethrionomys glareolus*) and Hantaan virus (striped field mouse, *Apodemus agrarius*), respectively. Other potential causative viruses exist, including Dobrava virus (yellow-necked field mouse, *A. flavicollis*), which causes severe HF with renal syndrome in the Balkans. Seoul virus is associated with the Norway or sewer rat, *Rattus norvegicus*, and has a worldwide distribution through the migration of the rodent; it is associated with mild or moderate HF with renal syndrome in Asia, but in many areas of the world the human disease has been difficult to identify. Most cases occur in rural residents or vacationers; the exception is Seoul virus disease, which may be acquired in an urban or rural setting or from contaminated laboratory rat colonies. Classic Hantaan disease in Korea (Korean HF) and in rural China (epidemic HF) is most common in spring and fall and is related to rodent density and agricultural practices. Human infection is acquired primarily through aerosols of rodent urine, although virus is also present in saliva and feces. Patients with hantavirus diseases are not infectious. HF with renal syndrome is the most important form of HF today, with >100,000 cases of severe disease in Asia annually and milder Puumala infections numbering in the thousands as well.

Severe cases of HF with renal syndrome caused by Hantaan virus evolve in identifiable stages: the febrile stage with myalgia, lasting 3 or 4 days; the hypotensive stage, often associated with shock and lasting from a few hours to 48 h; the oliguric stage with renal failure, lasting 3–10 days; and the polyuric stage with diuresis and hyposthenuria.

The *febrile stage* is initiated by the abrupt onset of fever, headache, severe myalgia, thirst, anorexia, and often nausea and vomiting. Photophobia, retroorbital pain, and pain on ocular movement are common, and the vision may become blurred with ciliary body inflammation. Flushing over the face, the V area of the neck, and the back is characteristic, as are pharyngeal injection, periorbital edema, and conjunctival suffusion. Petechiae often develop in areas of pressure, the conjunctivae, and the axillae. Back pain and tenderness to percussion at the costovertebral angle reflect massive retroperitoneal edema. Laboratory evidence of mild to moderate DIC is present. Other laboratory findings include proteinuria and an active urinary sediment.

The *hypotensive stage* is ushered in by falling blood pressure and sometimes by shock. The relative bradycardia typical of the febrile phase is replaced by tachycardia. Kinin activation is marked. The rising hematocrit reflects increasing vascular leakage. Leukocytosis with a left shift develops, and thrombocytopenia continues. Atypical lymphocytes—which in fact are activated CD8+ (and, to a lesser extent, CD4+) T cells—circulate. Proteinuria is marked, and the urine's specific gravity falls to 1.010. The renal circulation is congested and compromised from local and systemic circulatory changes resulting in necrosis of tubules, particularly at the corticomedullary junction, and oliguria.

During the *oliguric stage*, hemorrhagic tendencies continue, probably in large part because of uremic bleeding defects. The oliguria persists for 3–10 days before the return of renal function marks the onset of the *polyuric stage*, which carries the danger of dehydration and electrolyte abnormalities.

Mild cases of HF with renal syndrome may be much less stereotypical. The presentation may include only fever, gastrointestinal abnormalities, and transient oliguria followed by hyposthenuria.

HF with renal syndrome should be suspected in patients with rural exposure in an endemic area. Prompt recognition of the disease permits rapid hospitalization and expectant management of shock and renal failure. Useful clinical laboratory parameters include leukocytosis, which may be leukemoid and is associated with a left shift; thrombocytopenia; and proteinuria. Mainstays of therapy are the management of shock, reliance on pressors, modest crystalloid infusion, IV use of human serum albumin, and treatment of renal failure with prompt dialysis for the usual indications. Hydration may result in pulmonary edema, and hypertension should be avoided because of the possibility of intracranial hemorrhage. Use of IV ribavirin has reduced mortality and morbidity rates in severe cases provided treatment is begun within the first 4 days of illness. The case-fatality ratio may be as high as 15% but with proper therapy should be <5%. Sequelae have not been definitively established, but there is a correlation in the United States between chronic hypertensive renal failure and the presence of antibodies to Seoul virus.

Infections with Puumala virus, the most common cause of HF with renal syndrome in Europe, result in a much attenuated picture but the same general presentation. The syndrome may be referred to by its former name, *nephropathia epidemica*. Bleeding manifestations are found in only 10% of cases, hypotension rather than shock is usually seen, and oliguria is present in only about half of patients. The dominant features may be fever, abdominal pain, proteinuria, mild oliguria, and sometimes blurred vision or glaucoma followed by polyuria and hyposthenuria in recovery. The mortality rate is <1%.

The diagnosis is readily made by IgM-capture ELISA, which should be positive at admission or within 24–48 h thereafter. The isolation of virus is difficult, but RT-PCR of a blood clot collected early in the clinical course or of tissues obtained postmortem will give positive results. Such testing is usually undertaken only if definitive identification of the infecting viral species is required or if molecular epidemiologic questions exist.

Hantavirus Pulmonary Syndrome

Hantavirus pulmonary syndrome was discovered in 1993, but retrospective identification of cases by immunohistochemistry (1978) and serology (1959) support the idea that it is a recently discovered rather than a truly new disease. The causative agents are hantaviruses of a distinct phylogenetic lineage that is associated with the rodent subfamily Sigmodontinae. Sin Nombre virus, which chronically infects the deer mouse (*Peromyscus maniculatus*), is the most important agent of hantavirus pulmonary syndrome in the United States. The disease is also caused by a Sin Nombre virus variant from the white-footed mouse (*P. leucopus*), by Black Creek Canal virus (*Sigmodon hispidus*, the cotton rat), and by Bayou virus (*Oryzomys palustris*, the rice rat). Several other related viruses cause the disease in South America, but Andes virus is unusual in that it, alone among hantaviruses, has been implicated in human-to-human transmission. The disease is linked to rodent exposure and particularly affects rural residents living in dwellings permeable to rodent entry or working at occupations that pose a risk of rodent exposure. Each rodent species has its own particular habits; in the case of the deer mouse, these behaviors include living in and around human habitation.

The disease begins with a prodrome of ~3–4 days (range, 1–11 days) comprising fever, myalgia, malaise, and often gastrointestinal disturbances such as nausea, vomiting, and abdominal pain. Dizziness is common and vertigo occasional. Severe prodromal symptoms bring some individuals to medical attention, but patients are usually recognized as the cardiopulmonary phase begins. Typically, there is slightly lowered blood pressure, tachycardia, tachypnea, mild hypoxemia, and early radiographic signs of pulmonary edema. Physical findings in the chest are often surprisingly scant. The conjunctival and cutaneous signs of vascular involvement seen in other types of HF are absent. During the next few hours, decompensation may progress rapidly to severe hypoxemia and respiratory failure. Most patients surviving the first 48 h of hospitalization are extubated and discharged within a few days, with no apparent residua.

Management during the first few hours after presentation is critical. The goal is to prevent severe hypoxemia by oxygen therapy, with intubation and intensive respiratory management if needed. During this period, hypotension and shock with increasing hematocrit invite aggressive fluid administration, but this intervention should be undertaken with great caution. Because of low cardiac output with myocardial depression and

increased pulmonary vascular permeability, shock should be managed expectantly with pressors and modest infusion of fluid guided by the pulmonary capillary wedge pressure. Mild cases can be managed by frequent monitoring and oxygen administration without intubation. Many patients require intubation to manage hypoxemia and also develop shock. Mortality rates remain at ~30–40% even with good management. The antiviral drug ribavirin inhibits the virus in vitro but did not have a marked effect on patients treated in an open-label study.

During the prodrome, the differential diagnosis of hantavirus pulmonary syndrome is difficult, but by the time of presentation or within 24 h thereafter, a number of diagnostically helpful clinical features become apparent. Cough is not usually present at the outset but may develop later. Interstitial edema is evident on the chest x-ray. Later, bilateral alveolar edema with a central distribution develops in the setting of a normal-sized heart; occasionally, the edema is initially unilateral. Pleural effusions are often seen. Thrombocytopenia, circulating atypical lymphocytes, and a left shift (often with leukocytosis) are almost always evident; thrombocytopenia is a particularly important early clue. Hemoconcentration, proteinuria, and hypoalbuminemia should also be sought. Although thrombocytopenia virtually always develops and prolongation of the partial thromboplastin time is the rule, clinical evidence for coagulopathy or laboratory indications of DIC are found in only a minority of cases, usually in severely ill patients. Patients with severe illness also have acidosis and elevated serum levels of lactate. Mildly increased values in renal function tests are common, but patients with severe cases often have markedly elevated concentrations of serum creatinine; some of the viruses other than Sin Nombre virus have been associated with more kidney involvement, but few such cases have been studied. The differential diagnosis includes abdominal surgical conditions and pyelonephritis as well as rickettsial disease, sepsis, meningococcemia, plague, tularemia, influenza, and relapsing fever.

A specific diagnosis is best made by IgM testing of acute-phase serum, which has yielded positive results even in the prodrome. Tests using a Sin Nombre virus antigen detect the related hantaviruses causing the pulmonary syndrome in the Americas. Occasionally, heterologous viruses will react only in the IgGELISA, but this finding is highly suspicious given the very low seroprevalence of these viruses in normal populations. RT-PCR is usually positive when used to test blood clots obtained in the first 7–9 days of illness as well as tissues; this test is useful in identifying the infecting virus in areas outside the home range of the deer mouse and in atypical cases.

Yellow Fever

Yellow fever virus caused major epidemics in the Americas, Africa, and Europe before the discovery of mosquito transmission in 1900 led to its control through attacks on its urban vector, *A. aegypti*. Only then was it found that a jungle cycle also existed in Africa, involving other *Aedes* mosquitoes and monkeys, and that colonization of the New World with *A. aegypti*, originally an African species, had established urban yellow fever as well as an independent sylvatic yellow fever cycle involving *Haemagogus* mosquitoes and New World monkeys in American jungles. Today, urban yellow fever transmission occurs only in some African cities, but the threat exists in the great cities of South America, where reinfestation by *A. aegypti* has taken place and dengue transmission by the same mosquito is common. As late as 1905, New Orleans suffered >3000 cases with 452 deaths from "yellow jack." Despite the existence of a highly effective and safe vaccine, several hundred jungle yellow fever cases occur annually in South America, and thousands of jungle and urban cases occur each year in Africa.

Yellow fever is a typical HF accompanied by prominent hepatic necrosis. A period of viremia, typically lasting 3 or 4 days, is followed by a period of "intoxication." During the latter phase in severe cases, the characteristic jaundice, hemorrhages, black vomit, anuria, and terminal delirium occur, perhaps related in part to extensive hepatic involvement. Blood leukocyte counts may be normal or reduced and are often high in terminal stages. Albuminuria is usually noted and may be marked; as renal function fails in terminal or severe cases, the level of blood urea nitrogen rises proportionately. Abnormalities detected in liver function tests range from modest elevations of AST levels in mild cases to severe derangement.

Urban yellow fever can be prevented by the control of *A. aegypti*. The continuing sylvatic cycle requires vaccination of all visitors to areas of potential transmission. With few exceptions, reactions to vaccine are minimal; immunity is provided within 10 days and lasts for at least 10 years. An egg allergy dictates caution in vaccine administration. Although there are no documented harmful effects of the vaccine on the fetus, pregnant women should be immunized only if they are definitely at risk of yellow fever exposure. Since vaccination has been associated with several cases of encephalitis in children <6 months of age, it should be delayed until after 12 months of age unless the risk of exposure is very high. Rare, serious, multisystemic adverse reactions (occasionally fatal) have been reported, particularly affecting the elderly; nevertheless, the number of deaths of unvaccinated travelers with yellow fever exceeds the number of deaths from vaccination, and a liberal vaccination policy for travelers to involved areas should be pursued. Timely information on changes in yellow fever distribution and yellow fever vaccine requirements can be obtained from the Centers for Disease Control and Prevention (*wwwnc.cdc.gov/travel*).

Dengue Hemorrhagic Fever/Dengue Shock Syndrome

A syndrome of HF noted in the 1950s among children in the Philippines and Southeast Asia was soon associated with dengue virus infections, particularly those occurring against a background of previous exposure to another dengue-virus serotype. The transient heterotypic protection after dengue virus infection is replaced within several weeks by the potential for heterotypic infection resulting in typical dengue fever (see above) or—uncommonly—in enhanced disease (secondary DHF/DSS). In rare instances, primary dengue infections lead to an HF syndrome, but much less is known about pathogenesis in this situation. In the past 20 years, *A. aegypti* has progressively reinvaded Latin America and other areas, and frequent travel

by infected individuals has introduced multiple strains of dengue virus from many geographic areas. Thus the pattern of hyperendemic transmission of multiple dengue serotypes has now been established in the Americas and the Caribbean and has led to the emergence of DHF/DSS as a major problem there as well. Millions of dengue infections, including many thousands of cases of DHF/DSS, occur annually. The severe syndrome is unlikely to be seen in U.S. citizens since few children have the dengue antibodies that can trigger the pathogenetic cascade when a second infection is acquired.

Macrophage/monocyte infection is central to the pathogenesis of dengue fever and to the origin of DHF/DSS. Previous infection with a heterologous dengue-virus serotype may result in the production of nonprotective antiviral antibodies that nevertheless bind to the virion's surface and through interaction with the Fc receptor focus secondary dengue viruses on the target cell, the result being enhanced infection. The host is also primed for a secondary antibody response when viral antigens are released and immune complexes lead to activation of the classic complement pathway, with consequent phlogistic effects. Cross-reactivity at the T cell level results in the release of physiologically active cytokines, including interferon γ and tumor necrosis factor α. The induction of vascular permeability and shock depends on multiple factors, including the following:

1. *Presence of enhancing and nonneutralizing antibodies* —Transplacental maternal antibody may be present in infants <9 months old, or antibody elicited by previous heterologous dengue infection may be present in older individuals. T cell reactivity is also intimately involved.
2. *Age*—Susceptibility to DHF/DSS drops considerably after 12 years of age.
3. *Sex*—Females are more often affected than males.
4. *Race*—Whites are more often affected than blacks.
5. *Nutritional status*—Malnutrition is protective.
6. *Sequence of infection*—For example, serotype 1 followed by serotype 2 seems to be more dangerous than serotype 4 followed by serotype 2.
7. *Infecting serotype*—Type 2 is apparently more dangerous than other serotypes.

In addition, there is considerable variation among strains of a given serotype, with Southeast Asian serotype 2 strains having more potential to cause DHF/DSS than others.

Dengue HF is identified by the detection of bleeding tendencies (tourniquet test, petechiae) or overt bleeding in the absence of underlying causes such as preexisting gastrointestinal lesions. Dengue shock syndrome, usually accompanied by hemorrhagic signs, is much more serious and results from increased vascular permeability leading to shock. In mild DHF/DSS, restlessness, lethargy, thrombocytopenia (<100,000/μL), and hemoconcentration are detected 2–5 days after the onset of typical dengue fever, usually at the time of defervescence. The maculopapular rash that often develops in dengue fever may also appear in DHF/DSS. In more severe cases, frank shock is apparent, with low pulse pressure, cyanosis, hepatomegaly, pleural effusions, ascites, and in some cases severe ecchymoses and gastrointestinal bleeding. The period of shock lasts only 1 or 2 days, and most patients respond promptly to close monitoring, oxygen administration, and infusion of crystalloid or—in severe cases—colloid. The case-fatality rates reported vary greatly with case ascertainment and the quality of treatment; however, most DHF/DSS patients respond well to supportive therapy, and the overall mortality rate at an experienced center in the tropics is probably as low as 1%.

A virologic diagnosis can be made by the usual means, although multiple flavivirus infections lead to a broad immune response to several members of the group, and this situation may result in a lack of virus specificity of the IgM and IgG immune responses. A secondary antibody response can be sought with tests against several flavivirus antigens to demonstrate the characteristic wide spectrum of reactivity.

The key to control of both dengue fever and DHF/DSS is the control of *A. aegypti*, which also reduces the risk of urban yellow fever and chikungunya virus circulation. Control efforts have been handicapped by the presence of nondegradable tires and long-lived plastic containers in trash repositories, insecticide resistance, urban poverty, and an inability of the public health community to mobilize the populace to respond to the need to eliminate mosquito breeding sites. Live attenuated dengue vaccines are in the late stages of development and have produced promising results in early tests. Whether vaccines can provide safe, durable immunity to an immunopathologic disease such as DHF/DSS in endemic areas is an issue that will have to be tested, but it is hoped that vaccination will reduce transmission to negligible levels.

Kyasanur Forest Disease and Omsk Hemorrhagic Fever

Kyasanur Forest virus and Omsk HF virus are geographically restricted, tick-borne flaviviruses that cause a syndrome of viral HF during a wave of viremia and that may also enter the CNS to cause subsequent viral encephalitis (see discussion of tick-borne encephalitis, above). There is no therapy for these infections, but an inactivated vaccine has been used in India against Kyasanur Forest disease. A new and related virus isolate has been obtained from butchers with HF in the Middle East; the implication is that there are more agents in this group.

Filovirus Hemorrhagic Fever

See Chap. 197.

FURTHER READINGS

Brackney MM et al: Epidemiology of Colorado tick fever in Montana, Utah, and Wyoming, 1995–2003. Vector Borne Zoonotic Dis 10:381, 2010[PMID: 19725767]

Bruno P et al: The protean manifestations of hemorrhagic fever with renal syndrome. A retrospective review of 26 cases from Korea. Ann Intern Med 113:385, 1990[PMID: 1974404]

Calisher CH: Medically important arboviruses of the United States and Canada. Clin Microbiol Rev 7:89, 1994[PMID: 8118792]

Centers for Disease Control and Prevention: Update: Management of patients with suspected viral hemorrhagic fever—United States. MMWR Morb Mortal Wkly Rep 44:475, 1995 (*www.cdc.gov/mmwr/preview/mmwrhtml/00038033.htm*)

Davis LE et al: North American encephalitis arboviruses. Neurol Clin 26:727, 2008[PMID: 18657724]

Depaquit J et al: Arthropod-borne viruses transmitted by phlebotomine sandflies in Europe: A review. Euro Surveill 15:19507, 2010[PMID: 20403307]

Deresiewicz RL et al: Clinical and neuroradiographic manifestations of eastern equine encephalitis. N Engl J Med 336:1867, 1997[PMID: 9197215]

Enria D et al: Arenaviruses, in *Tropical Infectious Diseases: Principles, Pathogens, & Practice*, RL Guerrant et al (eds). New York, Saunders, 1999, pp 1189–1212

Jonsson CB et al: A global perspective on hantavirus ecology, epidemiology, and disease. Clin Microbiol Rev 23:412, 2010[PMID: 20375360]

Peters CJ, Khan AS: Hantavirus pulmonary syndrome: The new American hemorrhagic fever. Clin Infect Dis 34:1224, 2002[PMID: 11941549]

Peterson LR, Hayes EB: West Nile virus in the Americas. Med Clin North Am 92:1307, 2008

Rivas F et al: Epidemic Venezuelan equine encephalitis in La Guajira, Colombia, 1995. J Infect Dis 175:828, 1997[PMID: 9086137]

Simpson SQ et al: Hantavirus pulmonary syndrome. Infect Dis Clin North Am 24:159, 2010[PMID: 20171551]

Solomon SR, Vaughn DW: Pathogenesis and clinical features of Japanese encephalitis and West Nile virus infections. Curr Top Microbiol Immunol 267:171, 2002[PMID: 12082989]

Wurtz R, Paleologos N: La Crosse encephalitis presenting like herpes simplex encephalitis in an immunocompromised adult. Clin Infect Dis 31:1113, 2000[PMID: 11049801]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EBOLA AND MARBURG VIRUSES: INTRODUCTION

Definition

Both Marburg virus and Ebola virus cause an acute febrile illness associated with a high mortality rate. This illness is characterized by multisystem involvement that begins with the abrupt onset of headache, myalgias, and fever and proceeds to prostration, rash, and shock and often to bleeding manifestations. Epidemics usually begin with a single case acquired from an unknown reservoir in nature (bats are suspected; see "Epidemiology," below) and spread mainly through close contact with sick persons or their body fluids, either at home or in the hospital.

Etiology

The family Filoviridae (Fig. 197-1) comprises two antigenically and genetically distinct genera: *Marburgvirus* and *Ebolavirus*. *Ebolavirus* has five readily distinguishable species named for their original sites of recognition: Zaire, Sudan, Côte d'Ivoire, Bundibugyo, and Reston. Except for the Reston virus, all the Filoviridae are African viruses that cause severe and often fatal disease in humans (Figs. 197-2 and 197-3). The Reston virus, which has been exported from the Philippines on several occasions, has caused fatal infections in monkeys but only subclinical infections in humans. Different strains of the five Ebola species, isolated over time and space, exhibit remarkable sequence conservation, indicating marked genetic stability in their selective niche.

Figure 197-1

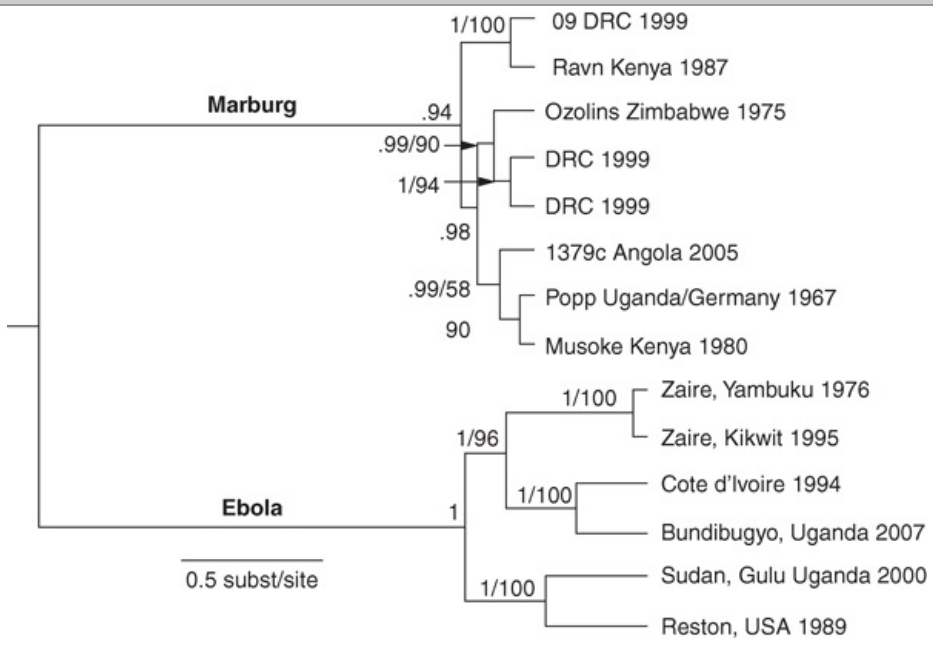

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Phylogenetic tree of filoviruses. *Marburgvirus* and *Ebolavirus* are seen to be two different genera. The genus *Ebolavirus* includes five distinct species. Note that the Yambuku and Kikwit Zaire viruses are virtually identical even though the epidemics for which they were responsible are separated by two decades and hundreds of kilometers. Virtually every virus sequenced from each of those two epidemics is identical over the part of the genome examined. This pattern is typical of that seen with single introductions followed by human-to-human passage via needle or close contact in an African hospital. In the *Marburgvirus* branch of the tree, there is one major clade with a slightly divergent group characterized by the Ravn 1987 Kenya isolate. All the viruses from the major Angola 2005 outbreak are represented by a single virus because the sequences in this human-to-human epidemic are virtually identical. However, in the outbreak occurring in the Democratic Republic of the Congo (DRC) in 1999 and resulting from multiple independent infections after cave entry, two viruses with slightly different phylogenies are represented within the major group, and there is even another virus within the Ravn subgroup. These sequences were selected from hundreds determined at the U.S.

Centers for Disease Control and Prevention and elsewhere. *(Adapted from Peters, 2010.)*

Figure 197-2

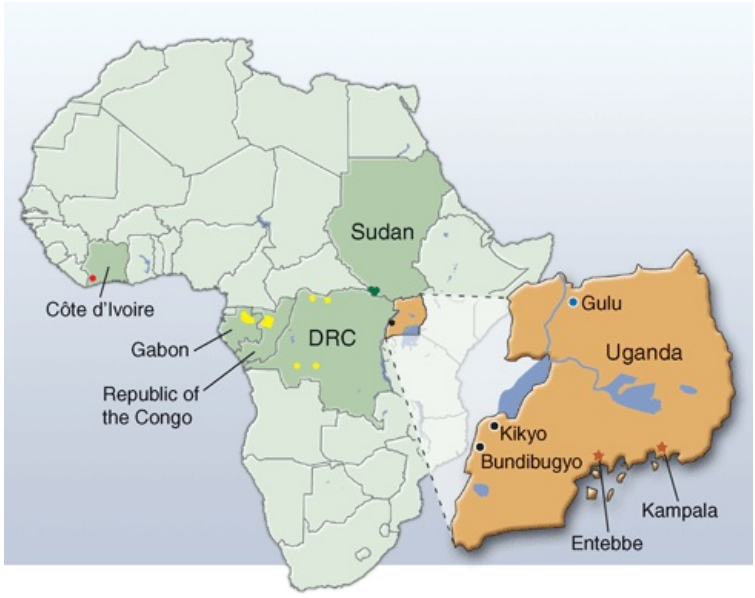

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left: **Geographic sites of *Ebolavirus* species identification,** as represented by dots (yellow, Zaire; green, Sudan; red, Côte d'Ivoire; black, Bundibugyo), in or adjacent to the Central African primary or secondary forest. Even *Ebolavirus* Côte d'Ivoire was isolated in the Tai forest reserve. ***Right:*****Amplified map of Uganda** shows the zone along the border of the Democratic Republic of the Congo (DRC, formerly Zaire) where the newest *Ebolavirus* species, Bundibugyo, was identified. Bundibugyo and the nearby town of Kikyo, which was also affected by this epidemic, are tourist destinations close to the Ugandan capital of Kampala. *(Adapted from Peters, 2010.)*

Figure 197-3

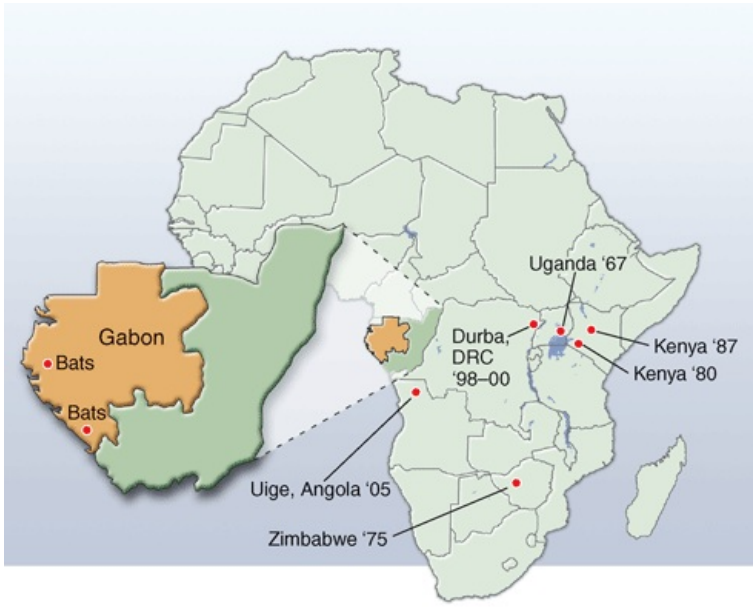

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Maps of the African continent and the country of Gabon (with adjacent Republic of the Congo) show the geographic distribution of *Marburgvirus* identification. Red dots indicate a case or an -epidemic. Uige, Angola, is the site of the largest Marburg epidemic (252 cases, 90% mortality rate). The Angolan strains differ by only 0–0.07% at the nucleotide level (Fig. 197-1). The

Durba outbreak lasted 3 years and was characterized by multiple introductions of virus into men entering a subterranean mine. Nine distinct lineages were detected, of which one was in the rather distant (21%) Ravn lineage. Red dots on the Gabon map indicate detection of virus in bats by PCR. *(Adapted from Peters, 2010.)*

Typical filovirus particles contain a single linear, negative-sense, single-stranded RNA arranged in a helical nucleocapsid. The virions are 790–970 nm in length; they may also appear in elongated, contorted forms (Fig. 197-4). The lipid envelope confers sensitivity to lipid solvents and common detergents. The viruses are largely destroyed by heat (60°C, 30 min) and by acidity but may persist for weeks in blood at room temperature. The glycoprotein self-associates to form the virion surface spikes, which presumably mediate attachment to cells and fusion. The glycoprotein's high sugar content may contribute to its low capacity to elicit effective neutralizing antibodies. A smaller form of the glycoprotein, bearing many of its antigenic determinants, is produced by in vitro–infected cells and is found in the circulation in human disease; it has been speculated that this circulating soluble protein may suppress the immune response to the virion surface protein or block antiviral effector mechanisms. Both Marburg virus and Ebola virus are biosafety level 4 pathogens because of their high associated mortality rate and aerosol infectivity.

Figure 197-4

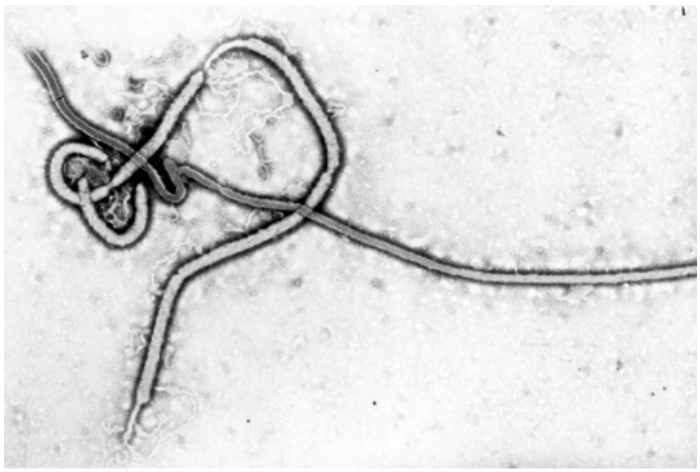

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ebola virions: diagnostic specimen from the first passage in Vero cells of a blood sample from a patient. Some of the filamentous (negatively stained) virions were fused together, end-to-end, giving the appearance of a "bowl of spaghetti." This image was from the first isolation and visualization of Ebola virus in 1970. *(Courtesy of Fredrick A. Murphy, MD, University of Texas Medical Branch, Galveston, Texas; with permission.)*

Epidemiology

Marburg virus was first identified in Germany in 1967, when infected African green monkeys (*Cercopithecus aethiops*) imported from Uganda transmitted the agent to workers in a vaccine laboratory. Of the 25 human cases acquired from monkeys, seven ended in death. The six secondary cases were associated with close contact or parenteral exposure. Secondary spread to the wife of one patient was documented, and virus was isolated from the husband's semen despite the presence of circulating serum antibodies. Isolated cases of Marburg virus infection were reported from eastern and southern Africa, with limited spread. Then, in 1999, repeated transmission of Marburg virus to workers in a gold mine in eastern Democratic Republic of the Congo (DRC; formerly Zaire) was studied. The secondary spread of the virus among patients' families was more extensive than previously noted, resembling that of Ebola virus and suggesting the importance of hygiene and proper barrier nursing in the epidemiology of these viruses in Africa. Finally, in 2004–2005, an alarming, massive Marburg virus epidemic, with >250 cases, occurred in Angola. The epidemiologic features resembled those of the Ebola virus epidemics described below, and the case-fatality rate was 90%. This high figure may have been due in part to poor conditions in African hospitals; however, the virus isolated in this epidemic was slightly different phylogenetically from other known strains and exhibited increased virulence in nonhuman primates.

Ebola virus first appeared in 1976, causing simultaneous epidemics of severe hemorrhagic fever (550 human cases) in Zaire and Sudan. Later, it was shown that different species of virus (with associated mortality rates of 90% and 50%, respectively) had caused the two epidemics. Both epidemics were associated with interhuman spread (particularly in the hospital setting) and the use of unsterilized

needles and syringes—a common practice in developing-country hospitals. The epidemics dwindled as the clinics were closed and as persons in the endemic area increasingly shunned affected persons and avoided traditional burial practices.

After an interval of apparent inactivity of almost 20 years, the Zaire Ebola virus recurred in a major epidemic (317 cases) in the DRC in 1995 and in smaller epidemics in Gabon in 1994–1996. Mortality rates were high (88% in the DRC), transmission to caregivers and others who had direct contact with body fluids was common, and poor hygiene in hospitals exacerbated spread. In the DRC epidemic, an index case was infected in Kikwit in January 1995. The epidemic smoldered until April, when intense nosocomial transmission forced closure of the hospitals; samples were finally sent to the laboratory for Ebola testing, which yielded positive results within a few hours. International assistance, with barrier nursing instruction and materials, was provided; nosocomial transmission ceased, hospitals reopened, and patients were segregated to prevent intrafamilial spread. The last case was reported in June 1995.

Separate emergences of Ebola virus (Zaire) were detected in Gabon in 1994–2003, usually in association with deep-forest exposure and subsequent familial and nosocomial transmission. Die-offs of nonhuman primates were sometimes documented, and Ebola infection was confirmed in at least some animals. In a 1996 episode, a physician exposed to Ebola-infected patients traveled to South Africa with a fever; a nurse who assisted in a cutdown on the physician developed Ebola hemorrhagic fever and died despite intensive care. The index patient was identified retrospectively on the basis of serum antibodies and virus isolation from semen. No additional cases arising from care of the primary or secondary case were detected, nor did any secondary cases follow care of an unsuspected Côte d'Ivoire Ebola case in Switzerland. Thus, distant transport of Ebola virus is an established risk, but limited nosocomial spread occurs under proper hygienic conditions.

After its first documented activity in 1976, the Sudan Ebola species returned in epidemic form to cause an indolent outbreak in Uganda in 2000–2001. This outbreak claimed the lives of 224 (53%) of 425 patients.

Reston Ebola virus was first seen in the United States in 1989, when it caused a fatal, highly transmissible disease among cynomolgus macaques imported from the Philippines and quarantined in Reston, Virginia, pending distribution to biomedical researchers. This and other appearances of the Reston virus have been traced to a single export facility in the Philippines, but no source in nature had been established until the discovery of this viral species in Philippine pigs. Occasional serologic evidence of human infection was found, but no cases of human disease were identified.

Epidemiologic studies (including a specific search in the Kikwit epidemic) have failed to yield evidence for an important role of airborne particles in human disease. This lack of epidemiologic evidence is surprising and seems to conflict with the viruses' classification as biosafety level 4 pathogens (which is based in large part on aerosol infectivity) and with formal laboratory assessments showing a high degree of aerosol infectivity for monkeys. Sick humans apparently do not usually generate sufficient amounts of infectious aerosols to pose a significant hazard to those around them.

Although numerous die-offs have been reported among chimpanzees and gorillas (some even threatening the viability of these endangered species), these animals (like humans) appear to be sentinels for virus activity. Speculation about the true reservoirs has centered on bats, and preliminary evidence indicates that bats may indeed be the reservoirs of filoviruses. This evidence includes the detection of antibodies and reverse-transcriptase polymerase chain reaction (RT-PCR) products in bats, the epidemiologic findings in subterranean gold mines in Durba (DRC) where Marburg transmission has occurred, and reported associations of human antibody production with the handling of bats. Recent isolation of Marburg virus from Egyptian fruit bats (*Rousettus aegyptiacus*) captured in Uganda in proximity to cases of human disease further supports bats as reservoirs, but the exact biologic relation and the natural cycle remain to be elucidated.

Pathology and Pathogenesis

In humans and in animal models, Ebola and Marburg viruses replicate well in virtually all cell types, including endothelial cells, macrophages, and parenchymal cells of multiple organs. In macaques, the earliest involvement—that of the mononuclear phagocyte system—is responsible for initiation of the disease process. In human disease and macaque models, upregulation of tissue factor and disseminated intravascular coagulation (DIC) are the inciting mechanisms. Viral replication is associated with cellular necrosis both in vivo and in vitro. Significant findings at the light-microscopic level include liver necrosis with Councilman bodies, intracellular inclusions that correlate with extensive collections of viral nucleocapsids, interstitial pneumonitis, cerebral glial nodules, and small infarcts. Antigen and virions are abundant in fibroblasts, interstitium, and (to a lesser extent) the appendages of the subcutaneous tissues in fatal cases; escape through small breaks in the skin or possibly through sweat glands may occur and, if so, may be correlated with the established epidemiologic risk of close contact with patients and the touching of the deceased. Inflammatory cells are not prominent, even in necrotic areas.

In addition to sustaining direct damage from viral infection, patients infected with Ebola virus (Zaire)

have high circulating levels of proinflammatory cytokines, which presumably contribute to the severity of the illness. In fact, the virus interacts intimately with the cellular cytokine system. It is resistant to the antiviral effects of interferon α, although this mediator is amply induced. Viral infection of endothelial cells selectively inhibits the expression of major histocompatibility complex class I molecules and blocks the induction of several genes by the interferons. In addition, glycoprotein expression inhibits αV integrin expression, an effect that leads to detachment and subsequent death of endothelial cells in vitro and that correlates with the limited inflammatory response evident in lesions.

Acute infection is associated with high levels of circulating virus and viral antigen. Clinical improvement takes place when viral titers decrease concomitant with the onset of a virus-specific immune response, as detected by enzyme-linked immunosorbent assay (ELISA) or fluorescent antibody testing. In fatal cases, there is usually little evidence of an antibody response, and there is extensive depletion of spleen and lymph nodes. Ebola Sudan virus amplification by PCR shows a correlation between serum viral RNA concentration and the likelihood of death. Recovery is apparently mediated by the cellular immune response: convalescent-phase plasma has little in vitro virus-neutralizing capacity and is not protective in humans or in passive transfer experiments in monkey and guinea pig models.

Clinical Manifestations

After an incubation period of ~7–10 days (range, 3–16 days), the patient abruptly develops fever, severe headache, malaise, myalgia, nausea, and vomiting. Continued fever is joined by diarrhea (often severe), chest pain (accompanied by cough), prostration, and depressed mentation. In light-skinned patients (and less often in dark-skinned individuals), a maculopapular rash appears around day 5–7 and is followed by desquamation. Bleeding may begin about this time and is apparent from any mucosal site and into the skin. In some epidemics, fewer than half of patients have had overt bleeding, and this manifestation has been absent even in some fatal cases. Additional findings include edema of the face, neck, and/or scrotum; hepatomegaly; flushing; conjunctival injection; and pharyngitis. Around 10–12 days after the onset of disease, the sustained fever may break, with improvement and eventual recovery of the patient. Recrudescence of fever may be associated with secondary bacterial infections or possibly with localized virus persistence. Late hepatitis, uveitis, and orchitis have been reported, with isolation of virus from semen or detection of PCR products in vaginal secretions for several weeks.

Laboratory Findings

Leukopenia is common early on; neutrophilia has its onset later. Platelet counts fall below (sometimes much below) 50,000/μL. Laboratory evidence of DIC is found, but its clinical significance and the need for therapy are controversial. Serum levels of alanine and aspartate aminotransferases (particularly the latter) rise progressively, and jaundice develops in some cases. The serum amylase level may be elevated, and this elevation may be associated with abdominal pain, suggesting pancreatitis. Proteinuria is usual; decreased kidney function is proportional to shock.

Diagnosis

Most patients acutely ill as a result of infection with Ebola or Marburg virus have high concentrations of virus in blood. Antigen-detection ELISA is a sensitive, robust diagnostic modality. Virus isolation and reverse-transcription PCR are also effective and provide additional sensitivity needed in some cases. Recovering patients develop IgM and IgG antibodies that are readily detected by ELISA. The indirect fluorescent antibody test with paired sera is an effective diagnostic tool in most acute cases but is extremely misleading in population-based serologic surveys for Ebola virus activity. Real-time PCR is extremely useful in detecting the need for quarantine or geographic spread. Skin biopsies are an extremely useful adjunct in postmortem diagnosis of infection with Ebola virus (and, to a lesser extent, Marburg virus) because of the presence of large amounts of viral antigen, the relatively low risk posed by sample collection, and the lack of cold-chain requirements for formalin-fixed tissues.

Treatment: Marburg and Ebola Virus Infections

No virus-specific therapy is available, and—given the extensive viral involvement in fatal cases—supportive treatment may not be as useful as was once hoped. However, studies in rhesus monkeys have shown improved survival among animals treated with an inhibitor of factor VIIa/tissue factor or with activated protein C; this effect demonstrates the importance of DIC in pathogenesis. In addition, direct intervention against viral replication with small interfering RNA (siRNA) is effective in postexposure prophylaxis against the highly virulent Zairespecies in macaques. Vigorous treatment of shock should take into account the likelihood of vascular leak in the pulmonary and systemic circulation and of myocardial functional compromise. The membrane fusion mechanism of Ebola virus resembles that of retroviruses, and the identification of "fusogenic" sequences suggests that inhibitors of cell entry may be developed. Despite the poor neutralizing capacity of polyclonal convalescent-phase sera, phage display of immunoglobulin mRNA from convalescent-phase bone marrow has yielded monoclonal antibodies that have in vitro neutralizing capacity and mediate protection in guinea pig models (but, unfortunately, not in the more sensitive monkey models).

Prevention

No vaccine or antiviral drug is currently available, but barrier nursing precautions in African hospitals can greatly decrease the spread of filoviruses beyond the index case and thus prevent epidemics of infection with these viruses and other agents as well. An adenovirus-vectored Ebola glycoprotein gene has proved protective in nonhuman primates and is undergoing phase 1 trials in humans. An experimental vesicular stomatitis virus–based vaccine has protected macaques when given both before and after infection with the Zaire Ebola virus.

FURTHER READINGS

Feldmann H et al: Proceedings of an international symposium on filoviruses. J Infect Dis 196(Suppl 2), 2007 [whole issue]

Geisbert TW et al: Treatment of Ebola virus infection with a recombinant inhibitor of factor VIIa/tissue factor: A study in rhesus monkeys. Lancet 362:1953, 2003[PMID: 14683653]

Peters CJ: Filoviridae: Marburg and Ebola virus hemorrhagic fevers, in GL Mandell et al (eds): *Principles and Practice of Infectious Diseases* , 7th ed. Philadelphia, Churchill Livingstone, 2010, pp 2259–2262

———, LeDuc JW: An introduction to Ebola: The virus and the disease. J Infect Dis 179(Suppl 1):ix, 1999

Sanchez A et al: Analysis of human peripheral blood samples from fatal and nonfatal cases of Ebola (Sudan) hemorrhagic fever: Cellular responses, virus load, and nitric oxide levels. J Virol 78:10370, 2004[PMID: 15367603]

Towner JS et al: *Marburgvirus* genomics and association with a large hemorrhagic fever outbreak in Angola. J Virol 80:6497, 2006[PMID: 16775337]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIAGNOSIS AND TREATMENT OF FUNGAL INFECTIONS: INTRODUCTION

Terminology and Microbiology

Traditionally, fungal infections have been classified into specific categories based on both anatomic location and epidemiology. The most common general anatomic categories are mucocutaneous and deep organ infection; the most common general epidemiologic categories are endemic and opportunistic. Although *mucocutaneous infections* can cause serious morbidity, they are rarely fatal. *Deep organ infections* also cause severe illness in many cases and, in contrast to mucocutaneous infections, are often fatal. The *endemic mycoses* (e.g., coccidioidomycosis) are infections caused by fungal organisms that are not part of the normal human microbial flora and are acquired from environmental sources. In contrast, *opportunistic mycoses* are caused by organisms (e.g., *Candida* and *Aspergillus*) that commonly are components of the normal human flora and whose ubiquity in nature renders them easily acquired by the immunocompromised host. Opportunistic fungi cause serious infections when the immunologic response of the host becomes ineffective, allowing the organisms to transition from harmless commensals to invasive pathogens. Frequently, the diminished effectiveness of the immune system is a result of advanced modern therapies that coincidentally either unbalance the host's microflora or directly interfere with immunologic responses. Endemic mycoses cause more severe illness in immunocompromised patients than in immunocompetent individuals.

Patients acquire deep organ infection with endemic fungi almost exclusively by inhalation. Cutaneous infections result either from hematogenous dissemination or, more often, from direct contact with soil —the natural reservoir for the vast majority of endemic mycoses. The dermatophytic fungi may be acquired by human-to-human transmission, but the majority of infections result from environmental contact. In contrast, the opportunistic fungus *Candida* invades the host from normal sites of colonization, usually the mucous membranes of the gastrointestinal tract. In general, innate immunity is the primary defense mechanism against fungi. Although antibodies are formed during many fungal infections (and even during commensalism), they generally do not constitute the primary mode of defense. Nevertheless, in selected infections, as discussed below, measurement of antibody titers may be a useful diagnostic test.

Three other terms frequently used in clinical discussions of fungal infections are *yeast*, *mold*, and *dimorphic fungus*. *Yeasts* are seen as rounded single cells or as budding organisms. *Candida* and *Cryptococcus* are traditionally classified as yeasts. *Molds* grow as filamentous forms called *hyphae* both at room temperature and in invaded tissue. *Aspergillus*, *Rhizopus* [the species that causes mucormycosis (zygomycosis)], and fungi commonly infecting the skin to cause ringworm and related cutaneous conditions are classified as molds. Variations occur within this classification of yeasts and molds. For instance, when *Candida* infects tissue, both yeasts and filamentous forms may occur (except with *C. glabrata*, which forms only yeasts in tissue); in contrast, *Cryptococcus* exists only in yeast form. *Dimorphic* is the term used to describe fungi that grow as yeasts or large spherical structures in tissue but as filamentous forms at room temperature in the environment. Classified in this group are the organisms causing blastomycosis, paracoccidioidomycosis, coccidioidomycosis, histoplasmosis, and sporotrichosis.

The incidence of nearly all fungal infections has risen substantially. Opportunistic infections have increased in frequency as a consequence of intentional immunosuppression in organ and stem cell transplantation and other diseases, the administration of cytotoxic chemotherapy for cancers, and the liberal use of antibacterial agents. The incidence of endemic mycoses has increased in geographic locations where there has been substantial population growth.

Diagnosis

The definitive diagnosis of any fungal infection requires histopathologic identification of the fungus invading tissue, accompanied by evidence of an inflammatory response. The identification of an inflammatory response has been especially important with regard to *Aspergillus* infection. *Aspergillus* is ubiquitous and can float from the air onto biopsy material. Therefore, in rare but important instances, this fungus is an ex vivo contaminant during processing of a specimen for microscopy, with a consequent incorrect diagnosis. The stains most commonly used to identify fungi are periodic acid –Schiff and Gomori methenamine silver. *Candida*, unlike other fungi, is visible on gram-stained tissue smears. Hematoxylin and eosin stain is not sufficient to identify *Candida* in tissue specimens. When positive, an India ink preparation of cerebrospinal fluid (CSF) is diagnostic for cryptococcosis. Most

laboratories now use calcofluor white staining coupled with fluorescent microscopy to identify fungi in fluid specimens.

Extensive investigations of the diagnosis of deep organ fungal infections have yielded a variety of tests with different degrees of specificity and sensitivity. The most reliable tests are the detection of antibody to *Coccidioides immitis* in serum and CSF; of *Histoplasma capsulatum* antigen in urine, serum, and CSF; and of cryptococcal polysaccharide antigen in serum and CSF. These tests have a general sensitivity and specificity of 90%; however, because there is variability among laboratories, testing on multiple occasions is advisable. The test for galactomannan has been used extensively in Europe and is now approved in the United States for diagnosis of aspergillosis. Sources of concern regarding galactomannan are the incidence of false-negative results and the need for multiple serial tests to reduce this incidence. The β-glucan test for *Candida* is also under evaluation but, like the galactomannan test, requires additional validation; this test has a negative predictive value of ~90%. Numerous polymerase chain reaction assays to detect antigens are in the developmental stages, as are nucleic acid hybridization techniques.

Of the fungal organisms, *Candida* is by far the most frequently recovered from blood. Although *Candida* species can be detected with any of the automated blood culture systems widely used at present, the lysis-centrifugation technique increases the sensitivity of blood cultures for *Candida* and for less common organisms (e.g., *H. capsulatum*). Lysis-centrifugation should be used when disseminated fungal infection is suspected.

Except in the cases of coccidioidomycosis, cryptococcosis, and histoplasmosis, there are no fully validated and widely used tests for serodiagnosis of disseminated fungal infection. Skin tests for the endemic mycoses are no longer available.

Treatment: Fungal Infections

This discussion is intended as a brief overview of general strategies for the use of antifungal agents in the treatment of fungal infections. Regimens, schedules, and strategies are detailed in the chapters on specific mycoses that follow in this section.

Since fungal organisms are eukaryotic cells that contain most of the same organelles (with many of the same physiologic functions) as human cells, the identification of drugs that selectively kill or inhibit fungi but are not toxic to human cells has been highly problematic. Far fewer antifungal than antibacterial agents have been introduced into clinical medicine.

AMPHOTERICIN B

The introduction of amphotericin B (AmB) in the late 1950s revolutionized the treatment of fungal infections in deep organs. Before AmB became available, cryptococcal meningitis and other disseminated fungal infections were nearly always fatal. For nearly a decade after AmB was introduced, it was the only effective agent for the treatment of life-threatening fungal infections. AmB remains the broadest-spectrum antifungal agent but carries several disadvantages, including significant nephrotoxicity, lack of an oral preparation, and unpleasant side effects (fever, chills, and nausea) during treatment. To circumvent nephrotoxicity and infusion side effects, lipid formulations of AmB were developed and have virtually replaced the original colloidal deoxycholate formulation in clinical use (although the older formulation is still available). The lipid formulations include liposomal AmB (L-AmB; 3–5 mg/kg per day) and AmB lipid complex (ABLC; 5 mg/kg per day). A third preparation, AmB colloidal dispersion (ABCD; 3–4 mg/kg per day), is rarely used because of the high incidence of side effects associated with infusion. (The doses listed are standard doses for adults with invasive infection.)

The lipid formulations of AmB have the disadvantage of being considerably more expensive than the deoxycholate formulation. Experience is still accumulating on the comparative efficacy, toxicity, and advantages of the different formulations for specific clinical fungal infections [e.g., central nervous system (CNS) infection]. Whether there is a clinically significant difference in these drugs with respect to CNS penetration or nephrotoxicity remains controversial. Despite these issues and despite the expense, the lipid formulations are now much more commonly used than AmB deoxycholate in developed countries. In developing countries, AmB deoxycholate is still preferred because of the expense of the lipid formulations.

AZOLES

This class of antifungal drugs offers important advantages over AmB: the azoles cause little or no nephrotoxicity and are available in oral preparations. Early azoles included ketoconazole and miconazole, which have been replaced by newer agents for the treatment of deep organ fungal infections. The azoles' mechanism of action is inhibition of ergosterol synthesis in the fungal cell wall. Unlike AmB, these drugs are considered fungistatic, not fungicidal.

Fluconazole

Since its introduction, fluconazole has played an extremely important role in the treatment of a wide variety of serious fungal infections. Its major advantages are the availability of both oral and IV

formulations, a long half-life, satisfactory penetration of most body fluids (including ocular fluid and CSF), and minimal toxicity (especially relative to that of AmB). Its disadvantages include (usually reversible) hepatotoxicity and—at high doses—alopecia, muscle weakness, and dry mouth with a metallic taste. Fluconazole is not effective for the treatment of aspergillosis, mucormycosis, or *Scedosporium apiospermum* infections. It is less effective than the newer azoles against *C. glabrata* and *C. krusei*.

Fluconazole has become the agent of choice for the treatment of coccidioidal meningitis, although relapses have followed therapy with this drug. In addition, fluconazole is useful for both consolidation and maintenance therapy for cryptococcal meningitis. This agent has been shown to be as efficacious as AmB in the treatment of candidemia. The effectiveness of fluconazole in candidemia and the drug's relatively minimal toxicity, in conjunction with the inadequacy of diagnostic tests for widespread hematogenously disseminated candidiasis, have led to a change in the paradigm for candidemia management. The standard of care is now to treat all candidemic patients with an antifungal agent and to change all their intravascular lines, if feasible, rather than merely to remove a singular suspect intravascular line and then observe the patient. The usual fluconazole regimen for treatment of candidemia is 400 mg/d given until 2 weeks after the last positive blood culture.

Fluconazole is considered effective as fungal prophylaxis in bone marrow transplant recipients and high-risk liver transplant patients. Its general use for prophylaxis in patients with leukemia, in AIDS patients with low CD4+ T cell counts, and in patients on surgical intensive care units remains controversial.

Voriconazole

Voriconazole, which is available in both oral and IV formulations, has a broader spectrum than fluconazole against *Candida* species (including *C. glabrata* and *C. krusei*) and is active against *Aspergillus*, *Scedosporium*, and *Fusarium*. It is generally considered the first-line drug of choice for treatment of aspergillosis. A few case reports have shown voriconazole to be effective in individual patients with coccidioidomycosis, blastomycosis, and histoplasmosis, but because of limited data this agent is not recommended for treatment of the endemic mycoses. Among the disadvantages of voriconazole (compared with fluconazole) are its more numerous interactions with many of the drugs used in patients predisposed to fungal infections. Hepatotoxicity, skin rashes (including photosensitivity), and visual disturbances are relatively common. Skin cancer surveillance is now recommended for patients taking voriconazole. Voriconazole is also considerably more expensive than fluconazole. Moreover, it is advisable to monitor voriconazole levels in certain patients since (1) this drug is completely metabolized in the liver by CYP2C9, CYP3A4, and CYP2C19; and (2) human genetic variability in CYP2C19 activity exists. Dosages should be reduced accordingly in those patients with liver failure. Dose adjustments for renal insufficiency are not necessary; however, because the IV formulation is prepared in cyclodextrin, it should not be given to patients with severe renal insufficiency.

Itraconazole

Itraconazole is available in IV and oral (capsule and suspension) formulations. Varying blood levels among patients taking oral itraconazole reflect a disadvantage compared with the other azoles. Itraconazole is the drug of choice for mild to moderate histoplasmosis and blastomycosis and has often been used for chronic mucocutaneous candidiasis. It has been approved by the U.S. Food and Drug Administration (FDA) for use in febrile neutropenic patients. Itraconazole has also proved useful for the treatment of chronic coccidioidomycosis, sporotrichosis, and *S. apiospermum* infection. The mucocutaneous and cutaneous fungal infections that have been treated successfully with itraconazole include oropharyngeal candidiasis (especially in AIDS patients), tinea versicolor, tinea capitis, and onychomycosis. Disadvantages of itraconazole include its poor penetration into CSF, the use of cyclodextrin in both the oral suspension and the IV preparation, the variable absorption of the drug in capsule form, and the need for monitoring of blood levels in patients taking capsules for disseminated mycoses. Reported cases of severe congestive heart failure in patients taking itraconazole have been a source of concern. Like the other azoles, itraconazole can cause hepatic toxicity.

Posaconazole

Posaconazole is approved by the FDA for prophylaxis of aspergillosis and candidiasis in patients at high risk for developing these infections because of severe immunocompromise. It has also been approved for the treatment of oropharyngeal candidiasis and has been evaluated for the treatment of zygomycosis, fusariosis, aspergillosis, cryptococcosis, and various other forms of candidal infection. The relevant studies of posaconazole in zygomycosis, fusariosis, and aspergillosis have examined salvage therapy. A study of >90 patients whose zygomycosis was refractory to other therapy yielded encouraging results. No trials of posaconazole for the treatment of candidemia have yet been reported. Case reports have described the drug's efficacy in coccidioidomycosis and histoplasmosis. Controlled trials have shown its effectiveness as a prophylactic agent in patients with acute leukemia and in bone marrow transplant recipients. In addition, posaconazole has been found to be effective against fluconazole-resistant *Candida* species. The results of a large-scale study of the use of posaconazole as salvage therapy for aspergillosis indicated that it is an alternative to other agents for

salvage therapy; however, that study predated the use of voriconazole and the echinocandins.

ECHINOCANDINS

The echinocandins, including the FDA-approved drugs caspofungin, anidulafungin, and micafungin, have added considerably to the antifungal armamentarium. All three of these agents inhibit β-1,3-glucan synthase, which is necessary for cell wall synthesis in fungi and is not a component of human cells. None of these agents is currently available in an oral formulation. The echinocandins are considered fungicidal for *Candida* and fungistatic for *Aspergillus*. Their greatest use to date is against candidal infections. They offer two advantages: broad-spectrum activity against all *Candida* species and relatively low toxicity. The minimal inhibitory concentrations (MICs) of all the echinocandins are highest against *C. parapsilosis*; it is not clear whether these higher MIC values represent less clinical effectiveness against this species. The echinocandins are among the safest antifungal agents.

In controlled trials, *caspofungin* has been at least as efficacious as AmB for the treatment of candidemia and invasive candidiasis and as efficacious as fluconazole for the treatment of candidal esophagitis. In addition, caspofungin has been efficacious as salvage therapy for aspergillosis. *Anidulafungin* has been approved by the FDA as therapy for candidemia in nonneutropenic patients and for *Candida* esophagitis, intraabdominal infection, and peritonitis. In controlled trials, anidulafungin has been shown to be noninferior and possibly superior to fluconazole against candidemia and invasive candidiasis. It is as efficacious as fluconazole against candidal esophagitis. When anidulafungin is used with cyclosporine, tacrolimus, or voriconazole, no dosage adjustment is required for either drug in the combination. *Micafungin* has been approved for the treatment of esophageal candidiasis and candidemia and for prophylaxis in patients receiving stem cell transplants. In a head-to-head trial, micafungin was noninferior to caspofungin for the treatment of candidemia. Studies thus far have shown that coadministration of micafungin and cyclosporine does not require dose adjustments for either drug. When micafungin is given with sirolimus, the AUC rises for sirolimus, usually necessitating a reduction in its dose. In open-label trials, favorable results have been obtained with micafungin for the treatment of deep-seated *Aspergillus* and *Candida* infections.

FLUCYTOSINE (5-FLUOROCYTOSINE)

The use of flucytosine has diminished as newer antifungal drugs have been developed. Flucytosine has a unique mechanism of action based on intrafungal conversion to 5-fluorouracil, which is toxic to the fungal cell. Development of resistance to the compound has limited its use as a single agent. Flucytosine is nearly always used in combination with AmB. Its good penetration into the CSF makes it attractive for use with AmB for treatment of cryptococcal meningitis. Flucytosine has also been recommended for the treatment of candidal meningitis in combination with AmB; comparative trials with AmB alone have not been done. Significant and frequent bone marrow depression is seen with flucytosine when this drug is used with AmB.

GRISEOFULVIN AND TERBINAFINE

Historically, griseofulvin has been useful primarily for ringworm infection. This agent is usually given for relatively long periods. Terbinafine has been used primarily for onychomycosis but also for ringworm. In comparative studies, terbinafine has been as effective as itraconazole and more effective than griseofulvin for both conditions.

TOPICAL ANTIFUNGAL AGENTS

A detailed discussion of the agents used for the treatment of cutaneous fungal infections and onychomycosis is beyond the scope of this chapter; the reader is referred to the dermatology literature. Many classes of compounds have been used to treat the common fungal infections of the skin. Among the azoles used are clotrimazole, econazole, miconazole, oxiconazole, sulconazole, ketoconazole, tioconazole, butoconazole, and terconazole. In general, topical treatment of vaginal candidiasis has been successful. Since there is considered to be little difference in the efficacy of the various vaginal preparations, the choice of agent is made by the physician and/or the patient on the basis of preference and availability. Fluconazole given orally at 150 mg has the advantage of not requiring repeated intravaginal application. Nystatin is a polyene that has been used for both oropharyngeal thrush and vaginal candidiasis. Useful agents in other classes include ciclopirox olamine, haloprogin, terbinafine, naftifine, tolnaftate, and undecylenic acid.

FURTHER READINGS

Chen SC et al: Antifungal therapy in invasive fungal infections. Curr opin Pharmacol 10:522, 2010[PMID: 20598943]

Cornely OA et al: Treatment outcome of invasive mould disease after sequential exposure to azoles and liposomal amphotericin B. J Antimicrob Chemother 65:114, 2010[PMID: 19887460]

Cronin S, Chandrasekar PH: Safety of triazole antifungal drugs in patients with cancer. J Antimicrob Chemother 65:410, 2010[PMID: 20035021]

Koo S et al: Prognostic features of galactomannan antigenemia in galactomannan-positive invasive aspergillosis. J Clin Microbiol 48:1255, 2010[PMID: 20147646]

Leventakos K et al: Fungal infections in leukemia patients: How do we prevent and treat them? Clin Infect Dis 50:405, 2010[PMID: 20047485]

Maertens J et al: Bronchoalveolar lavage fluid galactomannan for the diagnosis of invasive pulmonary aspergillosis in patients with hematologic diseases. Clin Infect Dis 49:1688, 2009[PMID: 19886801]

Mandell GL et al (eds): Mycoses, in Principles and Practice of Infectious Diseases, 7th ed. Elsevier Churchill Livingstone, Philadelphia, 2010, pp 3221–3391

Pappas G et al: Clinical practice guidelines for the management of candidiasis: 2009 update by the Infectious Diseases Society of America. Clin Infect Dis 48:503, 2009[PMID: 19191635]

Perfect JR et al: Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the Infectious Diseases Society of America. Clin Infect Dis 50:291, 2010[PMID: 20047480]

Racil Z et al: Difficulties in using 1,3-beta-d glucan as the screening test for the early diagnosis of invasive fungal diseases in patients with hematological malignancies—high frequency of false positive results and their analysis. J Med Microbol 59:1016, 2010[PMID: 20488937]

Vyas KS et al: Treatment of endemic mycoses. Expert Rev Respir Med 4:85, 2010[PMID: 20387295]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HISTOPLASMOSIS: INTRODUCTION

Etiology

Histoplasma capsulatum, a thermal dimorphic fungus, is the etiologic agent of histoplasmosis. In most endemic areas, *H. capsulatum* var. *capsulatum* is the causative agent; in Africa, *H. capsulatum* var. *duboisii* is also found. Mycelia—the naturally infectious form of *Histoplasma*—have a characteristic appearance, with microconidial and macroconidial forms. Microconidia are oval and are small enough (2–4 μm) to reach the terminal bronchioles and alveoli. Shortly after infecting the host, mycelia transform into the yeasts that are found inside macrophages and other phagocytes. The yeast forms are characteristically small (2–5 μm), with occasional narrow budding. In the laboratory, mycelia are best grown at room temperature, whereas yeasts are grown at 37°C on enriched media.

Epidemiology

Histoplasmosis is the most prevalent endemic mycosis in North America. Although this fungal disease has been reported throughout the world, its endemicity is particularly notable in certain parts of North, Central, and South America; Africa; and Asia. In the United States, the endemic areas spread over the Ohio and Mississippi river valleys. This pattern is related to the humid and acidic nature of the soil in these areas. Soil enriched with bird or bat droppings promotes the growth and sporulation of *Histoplasma*. Disruption of soil containing the organism leads to aerosolization of the microconidia and exposure of humans nearby. Activities associated with high-level exposure include spelunking, excavation, cleaning of chicken coops, demolition and remodeling of old buildings, and cutting of dead trees. Most cases seen outside of highly endemic areas represent imported disease—e.g., cases reported in Europe after travel to the Americas, Africa, or Asia.

Pathogenesis and Pathology

Infection follows inhalation of microconidia **(Fig. 199-1)**. Once they reach the alveolar spaces, microconidia are rapidly recognized and engulfed by alveolar macrophages. At this point, the microconidia transform into budding yeasts **(Fig. 199-2)**, a process that is integral to the pathogenesis of histoplasmosis and is dependent on the availability of calcium and iron inside the phagocytes. The yeasts are capable of growing and multiplying inside resting macrophages. Neutrophils and then lymphocytes are attracted to the site of infection. Before the development of cellular immunity, yeasts use the phagosomes as a vehicle for translocation to local draining lymph nodes, whence they spread hematogenously throughout the reticuloendothelial system. Adequate cellular immunity develops ~2 weeks after infection. T cells produce interferon γ to assist the macrophages in killing the organism and controlling the progression of disease. Interleukin 12 and tumor necrosis factor α (TNF-α) play an essential role in cellular immunity to *H. capsulatum*. In the immunocompetent host, macrophages, lymphocytes, and epithelial cells eventually organize and form granulomas that contain the organisms. These granulomas typically fibrose and calcify; calcified mediastinal lymph nodes and hepatosplenic calcifications are frequently found in healthy individuals from endemic areas. In immunocompetent hosts, infection with *H. capsulatum* confers some immunity to reinfection. In patients with impaired cellular immunity, the infection is not contained and can disseminate. Progressive disseminated histoplasmosis (PDH) can involve multiple organs, most commonly the bone marrow, spleen, liver **(Fig. 199-3)**, adrenal glands, and mucocutaneous membranes. Unlike latent tuberculosis, latent histoplasmosis rarely reactivates.

Figure 199-1

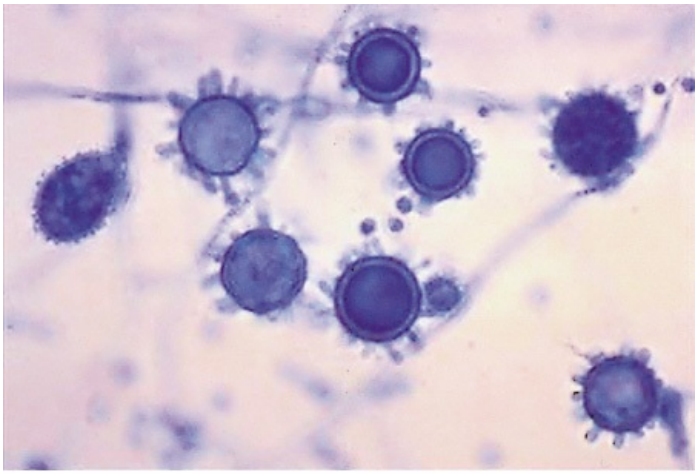

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spiked spherical conidia of *H. capsulatum* (lacto-phenolcotton blue stain).

Figure 199-2

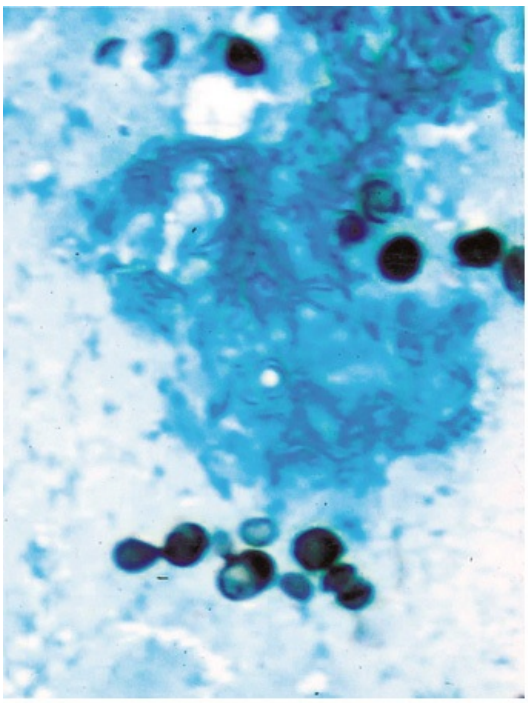

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Small (2–5 μm) narrow budding yeasts of *H. capsulatum* from bronchoalveolar lavage fluid (Grocott's methenamine silver stain).

Figure 199-3

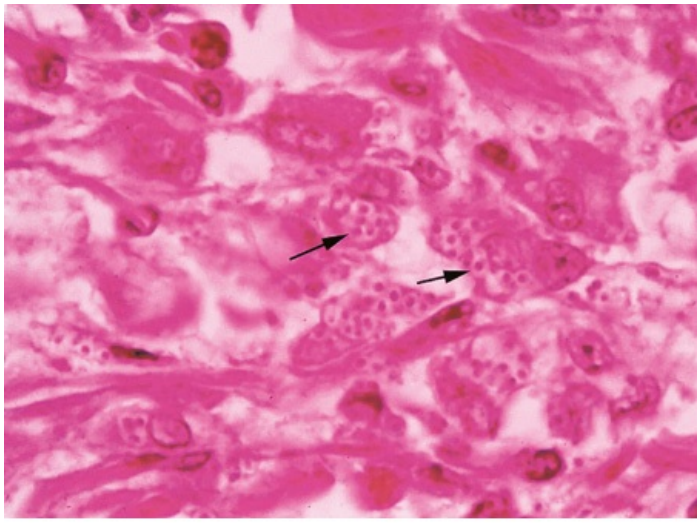

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Intracellular yeasts (*arrows*) of *H. capsulatum* in a liver biopsy specimen (hematoxylin and eosin stain).

Structural lung disease (e.g., emphysema) impairs the clearance of pulmonary histoplasmosis, and chronic pulmonary disease can result. This chronic process is characterized by progressive inflammation, tissue necrosis, and fibrosis mimicking cavitary tuberculosis.

Clinical Manifestations

The clinical spectrum of histoplasmosis ranges from asymptomatic infection to life-threatening illness. The attack rate and the extent and severity of the disease depend on the intensity of exposure, the immune status of the exposed individual, and the underlying lung architecture of the host.

In immunocompetent individuals with low-level exposure, most *Histoplasma* infections are either asymptomatic or mild and self-limited. Of adults residing in endemic areas, 50–80% have skin-test and/or radiographic evidence of previous infection without clinical manifestations. When symptoms do develop, they usually appear 1–4 weeks after exposure. Heavy exposure leads to a flulike illness with fever, chills, sweats, headache, myalgia, anorexia, cough, dyspnea, and chest pain. Chest radiographs usually show signs of pneumonitis with hilar or mediastinal adenopathy. Pulmonary infiltrates may be focal with light exposure or diffuse with heavy exposure. Rheumatologic symptoms of arthralgia or arthritis, often associated with erythema nodosum, occur in 5–10% of patients with acute histoplasmosis. Pericarditis may also develop. These manifestations represent inflammatory responses to the acute infection rather than its direct effects. Hilar or mediastinal lymph nodes may undergo necrosis and coalesce to form large mediastinal masses that can cause compression of great vessels, proximal airways, and the esophagus. These necrotic lymph nodes may also rupture and create fistulas between mediastinal structures (e.g., bronchoesophageal fistulas).

PDH is typically seen in immunocompromised individuals, who account for ~70% of cases. Common risk factors include AIDS (CD4+ T cell count, <200/μL), extremes of age, and the use of immunosuppressive medications such as prednisone, methotrexate, and anti-TNF-α agents. The spectrum of PDH ranges from an acute, rapidly fatal course—with diffuse interstitial or reticulonodular lung infiltrates causing respiratory failure, shock, coagulopathy, and multiorgan failure—to a more subacute course with a focal organ distribution. Common manifestations include fever and weight loss. Hepatosplenomegaly is also common. Other findings may include meningitis or focal brain lesions, ulcerations of the oral mucosa, gastrointestinal ulcerations, and adrenal insufficiency. Prompt recognition of this devastating illness is of paramount importance in patients with more severe manifestations or with underlying immunosuppression, especially AIDS (Chap. 189).

Chronic cavitary histoplasmosis is seen in smokers who have structural lung disease (e.g., bullous emphysema). This chronic illness is characterized by productive cough, dyspnea, low-grade fever, night sweats, and weight loss. Chest radiographs usually show upper-lobe infiltrates, cavitation, and pleural thickening—findings resembling those of tuberculosis. Without treatment, the course is slowly progressive.

Fibrosing mediastinitis is an uncommon and serious complication of histoplasmosis. In certain patients, acute infection is followed for unknown reasons by progressive fibrosis around the hilar and mediastinal lymph nodes. Involvement may be unilateral or bilateral; bilateral involvement carries a worse prognosis. Major manifestations include superior vena cava syndrome, obstruction of pulmonary vessels, and airway obstruction. Patients may experience recurrent pneumonia, hemoptysis, or respiratory failure. Fibrosing mediastinitis is fatal in up to one-third of cases.

In healed histoplasmosis, calcified mediastinal nodes or lung parenchyma may erode through the walls of the airways and cause hemoptysis. This condition is called *broncholithiasis*.

African histoplasmosis caused by *H. capsulatum* var. *duboisii* is clinically distinct and is characterized by frequent skin and bone involvement.

Diagnosis

Fungal culture remains the gold standard diagnostic test for histoplasmosis. However, culture results may not be known for up to 1 month, and cultures are often negative in less severe cases. Cultures are positive in ~75% of cases of PDH and chronic pulmonary histoplasmosis. Cultures of bronchoalveolar lavage (BAL) fluid are positive in about half of patients with acute pulmonary histoplasmosis causing diffuse infiltrates with hypoxemia. In PDH, the culture yield is highest for BAL fluid, bone marrow aspirate, and blood. Cultures of sputum or bronchial washings are usually positive in chronic pulmonary histoplasmosis. Cultures are typically negative, however, in other forms of histoplasmosis.

Fungal stains of cytopathology or biopsy materials showing structures resembling *Histoplasma* yeasts are helpful in the diagnosis of PDH, yielding positive results in about half of cases. Yeasts can be seen in BAL fluid (Fig. 199-2) from patients with diffuse pulmonary infiltrates, in bone marrow biopsy samples, and in biopsy specimens of other involved organs (e.g., the adrenal glands). Occasionally, yeasts are seen in blood smears from patients with severe PDH. However, staining artifacts and other fungal elements may be misidentified as *Histoplasma* yeasts.

The detection of *Histoplasma* antigen in body fluids is extremely useful in the diagnosis of PDH and acute diffuse pulmonary histoplasmosis. The sensitivity of this technique is >95% in patients with PDH and ~80% in patients with acute pulmonary histoplasmosis if both urine and serum are tested. Antigen can be detected in cerebrospinal fluid from patients with meningitis and in BAL fluid from those with pneumonia. Cross-reactivity occurs with African histoplasmosis, blastomycosis, coccidioidomycosis, paracoccidioidomycosis, and *Penicillium marneffei* infection.

Serologic tests, including immunodiffusion and complement fixation, are especially useful for the diagnosis of self-limited pulmonary histoplasmosis; however, at least 1 month is required for the production of antibodies after acute infection. A fourfold rise in antibody titer may be seen in patients with acute histoplasmosis. Serologic tests are also useful for the diagnosis of chronic pulmonary histoplasmosis. Limitations of serology, however, include insensitivity early in the course of infection and in immunosuppressed patients and the persistence of detectable antibody for several years after infection. Positive results from past infection may lead to a misdiagnosis of active histoplasmosis in a patient with another disease process.

Treatment: Histoplasmosis

Treatment recommendations for histoplasmosis are summarized in **Table 199-1**. Treatment is indicated for all patients with PDH or chronic pulmonary histoplasmosis as well as for symptomatic patients with acute pulmonary histoplasmosis causing diffuse infiltrates, especially with hypoxemia. In most cases of pulmonary histoplasmosis, treatment is not recommended because the degree of exposure is not heavy; the infection is asymptomatic or symptoms are mild, subacute, and not progressive; and the illness resolves without therapy.

Table 199-1 Recommendations for the Treatment of Histoplasmosis

Type of Histoplasmosis	Treatment Recommendations	Comments
Acute pulmonary, moderate to severe illness with diffuse infiltrates and/or hypoxemia	Lipid AmB (3–5 mg/kg per day) ± glucocorticoids for 1–2 weeks; then itraconazole (200 mg bid) for 12 weeks. Monitor renal and hepatic function.	Patients with mild cases usually recover without therapy, but itraconazole should be considered if the patient's condition has not improved after 1 month.
Chronic/cavitary pulmonary	Itraconazole (200 mg qd or bid) for at least 12 months. Monitor hepatic function.	Continue treatment until radiographic findings show no further improvement. Monitor for relapse after treatment is stopped.
Progressive disseminated	Lipid AmB (3–5 mg/kg per day) for 1–2 weeks; then itraconazole (200 mg bid) for at least 12 months. Monitor renal and hepatic function.	Liposomal AmB is preferred, but the AmB lipid complex may be used because of cost. Chronic maintenance therapy may be necessary if the degree of immunosuppression cannot be reduced.
Central nervous system	Liposomal AmB (5 mg/kg per day) for 4–6 weeks; then itraconazole (200 mg bid or tid) for at least 12 months. Monitor renal and hepatic -function.	A longer course of lipid AmB is recommended because of the high risk of relapse. Itraconazole should be continued until cerebrospinal fluid or CT abnormalities clear.

Abbreviation: AmB, amphotericin B.

The preferred treatments for histoplasmosis include the lipid formulations of amphotericin B in more severe cases and itraconazole in others. Liposomal amphotericin B has been more effective than the deoxycholate formulation for treatment of PDH in patients with AIDS. The deoxycholate formulation of amphotericin B is an alternative to a lipid formulation for patients at low risk for nephrotoxicity. Posaconazole, voriconazole, and fluconazole are alternatives for patients who cannot take itraconazole.

In severe cases requiring hospitalization, a lipid formulation of amphotericin B is followed by itraconazole. In patients with meningitis, a lipid formulation of amphotericin B should be given for 4–6 weeks before the switch to itraconazole. In immunosuppressed patients, the degree of immunosuppression should be reduced if possible, although immune reconstitution inflammatory syndrome (IRIS) may ensue. Antiretroviral treatment improves the outcome of PDH in patients with AIDS and is recommended; however, whether antiretroviral treatment should be delayed to avoid IRIS is unknown.

Blood levels of itraconazole should be monitored to ensure adequate drug exposure, with target concentrations of 2–10 μg/mL. Drug interactions should be carefully assessed: itraconazole not only is cleared by cytochrome P450 metabolism but also inhibits cytochrome P450. This profile causes interactions with many other medications.

The duration of treatment for acute pulmonary histoplasmosis is 6–12 weeks, while that for PDH and chronic pulmonary histoplasmosis is ≥1 year. Antigen levels in urine and serum should be monitored during and for at least 1 year after therapy for PDH. Stable or rising antigen levels suggest treatment failure or relapse.

Previously, lifelong itraconazole maintenance therapy was recommended for patients with AIDS once histoplasmosis was diagnosed. Today, however, maintenance therapy is not required for patients who respond well to antiretroviral therapy, with CD4+ T cell counts of at least 150/μL (preferably >250/μL); who complete at least 1 year of itraconazole therapy; and who exhibit neither clinical evidence of active histoplasmosis nor an antigenuria level of >4 ng/mL. Maintenance therapy also appears to be unnecessary in patients receiving immunosuppressive treatment if the degree of immunosuppression can be reduced through an approach similar to that used for patients with AIDS.

Fibrosing mediastinitis, which represents a chronic fibrotic reaction to past mediastinal histoplasmosis rather than an active infection, does not respond to antifungal therapy. While treatment is often prescribed for patients with pulmonary histoplasmosis who have not recovered within 1 month and for those with persistent mediastinal lymphadenopathy, the effectiveness of antifungal therapy in these situations is unknown.

FURTHER READINGS

Goldman M et al: Safety of discontinuation of maintenance therapy for disseminated histoplasmosis after immunologic response to antiretroviral therapy. Clin Infect Dis 38:1485, 2004 [PMID: 15156489]

Goodwin RA Jr, Des Prez RM: Histoplasmosis. Am Rev Respir Dis 117:929, 1978 [PMID: 655496]

Hage CA et al: Recognition, diagnosis, and treatment of histoplasmosis complicating tumor necrosis factor blocker therapy. Clin Infect Dis 50:85, 2010 [PMID: 19951231]

——— et al: Diagnosis of histoplasmosis by antigen detection in Bal fluid Chest 137:623, 2010

——— et al: Pulmonary histoplasmosis. Semin Respir Crit Care Med 29:151, 2008

Kauffman CA: Histoplasmosis: A clinical and laboratory update. Clin Microbiol Rev 20:115, 2007 [PMID: 17223625]

Swartzentruber S et al: Diagnosis of acute pulmonary histoplasmosis by antigen detection Clin Infect Dis 49: 1878, 2009 [PMID: 19911965]

Wheat LJ: Approach to the diagnosis of the endemic mycoses. Clin Chest Med 30:379, 2009 [PMID: 19375642]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

COCCIDIOIDOMYCOSIS: INTRODUCTION

Definition and Etiology

Coccidioidomycosis, commonly known as Valley Fever, is caused by dimorphic soil-dwelling fungi of the genus *Coccidioides*. Genetic analysis has demonstrated the existence of two species, *C. immitis* and *C. posadasii*. These species are indistinguishable with regard to the clinical disease they cause and their appearance on routine laboratory media. Thus, the organisms will be referred to simply as *Coccidioides* for the remainder of this chapter.

Epidemiology

Coccidioidomycosis is confined to the Western Hemisphere between the latitudes of 40°N and 40°S. In the United States, areas of high endemicity include the southern portion of the San Joaquin Valley of California and the south-central region of Arizona. However, infection may be acquired in other areas of the southwestern United States, including the southern coastal counties in California, southern Nevada, southwestern Utah, southern New Mexico, and western Texas, including the Rio Grande Valley. Outside the United States, coccidioidomycosis is endemic to northern Mexico as well as to localized regions of Central America. In South America, there are endemic foci in Colombia, Venezuela, northeastern Brazil, Paraguay, Bolivia, and north-central Argentina.

The risk of infection is increased by direct exposure to soil harboring *Coccidioides*. Because of difficulty in isolating *Coccidioides* from the soil, the precise characteristics of potentially infectious soil are not known. Several outbreaks of coccidioidomycosis have been associated with soil from archaeologic excavations of Amerindian sites both within and outside of the recognized endemic region. These cases often involved alluvial soils in regions of relative aridity with moderate temperature ranges. *Coccidioides* was isolated at depths of 2–20 cm below the surface.

In endemic areas, many cases of *Coccidioides* infection occur without obvious soil or dust exposure. Climatic factors appear to increase the infection rate in these regions. In particular, periods of aridity following rainy seasons have been associated with marked increases in the number of symptomatic cases. The number of cases of symptomatic coccidioidomycosis has increased dramatically in south-central Arizona, where most of the state's population resides. The factors causing this increase have not been fully elucidated; however, an influx of older individuals without prior coccidioidal infection into the region appears to be involved. Other variables, such as climate change, construction activity, and increased awareness and reporting, may also be factors. A similar increase in the incidence of symptomatic cases has recently been observed in the southern San Joaquin Valley of California.

Pathogenesis, Pathology, and Immune Response

On agar media and in the soil, *Coccidioides* organisms exist as filamentous molds. Within this mycelial structure, individual filaments (*hyphae*) elongate and branch, some growing upward. Alternating cells within the hyphae degenerate, leaving barrel-shaped viable elements called *arthroconidia*. Measuring ~2 by 5 μm, arthroconidia may become airborne for extended periods. Their small size allows them to evade initial mechanical mucosal defenses and reach deep into the bronchial tree, where infection is initiated in the nonimmune host.

Once in a susceptible host, the arthroconidia enlarge, become rounded, and develop internal septations. The resulting structures, called *spherules* (Fig. 200-1), may attain sizes of 200 μm and are unique to *Coccidioides*. The septations encompass uninuclear elements called *endospores*. Spherules may rupture and release packets of endospores that can themselves develop into spherules, thus propagating infection locally. If returned to artificial media or the soil, the fungus reverts to its mycelial stage.

Figure 200-1

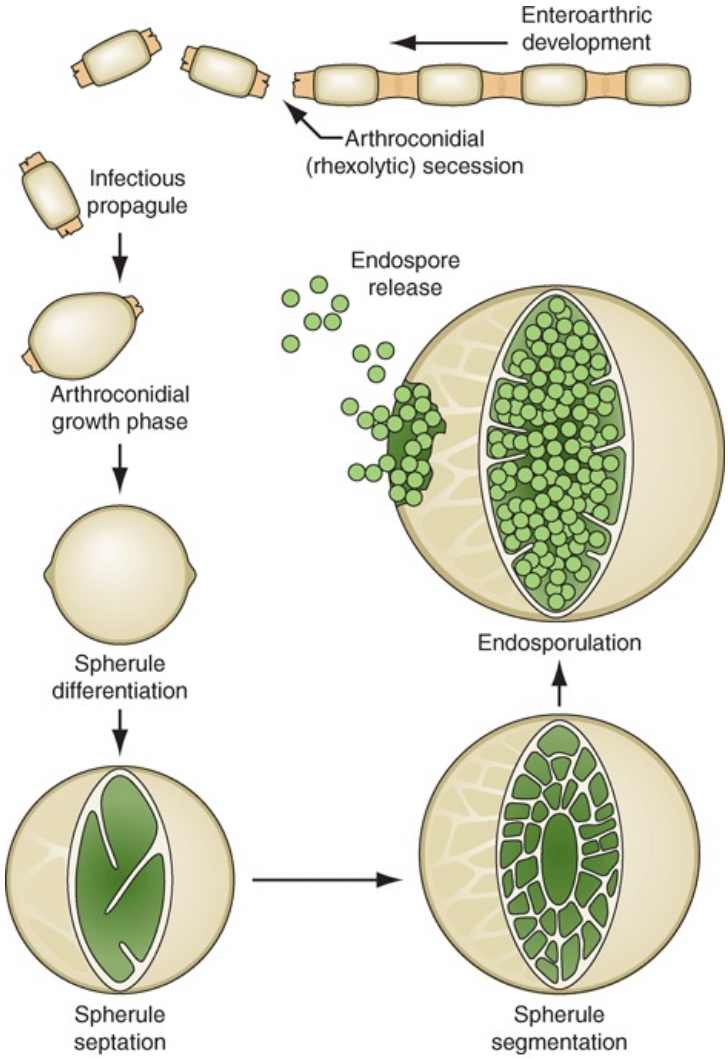

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Life cycle of *Coccidioides*. (*From TN Kirkland, J Fierer: Emerg Infect Dis 2:192, 1996* .)

Clinical observations and data from studies of animals strongly support the critical role of a robust cellular immune response in the host's control of coccidioidomycosis. Necrotizing granulomas containing spherules are typically identified in patients with resolved pulmonary infection. In disseminated disease, granulomas are generally poorly formed or do not develop at all, and a polymorphonuclear leukocyte response occurs frequently. In patients who are asymptomatic or in whom the initial pulmonary infection resolves, delayed-type hypersensitivity to coccidioidal antigens is routinely documented.

Clinical and Laboratory Manifestations

Coccidioidomycosis is protean in its manifestations. Among infected individuals, 60% are completely asymptomatic, and the remaining 40% have symptoms that are related principally to pulmonary infection, including fever, cough, and pleuritic chest pain. The risk of symptomatic illness increases with age. Coccidioidomycosis is commonly misdiagnosed as community-acquired bacterial pneumonia.

There are several cutaneous manifestations of primary pulmonary coccidioidomycosis. Toxic erythema consisting of a maculopapular rash has been noted in some cases. Erythema nodosum (see Fig. e7-40)—typically over the lower extremities—or erythema multiforme (see Fig. e7-25)—usually in a necklace distribution—may occur; these manifestations are seen particularly often in women. Arthralgias and arthritis may develop. The diagnosis of primary pulmonary coccidioidomycosis is suggested by a history of night sweats or profound fatigue as well as by peripheral-blood eosinophilia and hilar or mediastinal lymphadenopathy on chest radiography. While pleuritic chest pain is common, pleural effusions occur in fewer than 10% of cases. Such effusions are invariably associated with a pulmonary infiltrate on the same side. The cellular content of these effusions is mononuclear in nature; *Coccidioides* is rarely grown from effusions.

In most patients, primary pulmonary coccidioidomycosis usually resolves without sequelae in several weeks. However, a variety of pneumonic complications may ensue. Pulmonary nodules are residua of primary pneumonia. Generally single, frequently located in the upper lobes, and ≤4 cm in diameter,

nodules are often discovered on a routine chest radiograph in an asymptomatic patient. Calcification is uncommon. Coccidioidal pulmonary nodules can be difficult to distinguish radiographically from pulmonary malignancies. Like malignancies, coccidioidal nodules often enhance on positron emission tomography. However, routine CT often demonstrates multiple nodules in coccidioidomycosis. Biopsy is often required to distinguish between these two conditions.

Pulmonary cavities occur when a nodule extrudes its contents into the bronchus, resulting in a thin-walled shell. These cavities can be associated with persistent cough, hemoptysis, and pleuritic chest pain. Rarely, a cavity may rupture into the pleural space, causing pyopneumothorax. In such cases, patients present with acute dyspnea, and the chest radiograph reveals a collapsed lung with a pleural air-fluid level. Chronic or persistent pulmonary coccidioidomycosis manifests with prolonged symptoms of fever, cough, and weight loss and is radiographically associated with pulmonary scarring, fibrosis, and cavities. It occurs in fewer than 1% of patients, many of whom already have chronic lung disease of other etiologies.

In some cases, primary pneumonia presents as a diffuse reticulonodular pulmonary process (detected by plain chest radiography) in association with dyspnea and fever. Primary diffuse coccidioidal pneumonia may occur in settings of intense environmental exposure or profoundly suppressed cellular immunity (e.g., in patients with AIDS), with unrestrained fungal growth that is frequently associated with fungemia.

Clinical dissemination outside the thoracic cavity occurs in fewer than 1% of infected individuals. Dissemination is more likely to occur in male patients, particularly those of African-American or Filipino ancestry, and in persons with depressed cellular immunity, including patients with HIV infection and peripheral-blood CD4+ T cell counts of <250/μL; those receiving chronic glucocorticoid therapy; those with allogeneic solid-organ transplants; and those being treated with tumor necrosis factor α (TNF-α) antagonists. Women who acquire infection during the second or third trimester of pregnancy are also at risk for disseminated disease. Common sites for dissemination include the skin, bone, joints, soft tissues, and meninges. Dissemination may follow symptomatic or asymptomatic pulmonary infection and may involve only one site or multiple anatomic foci. When it occurs, clinical dissemination is usually evident within the first few months after primary pulmonary infection.

Meningitis, if untreated, is uniformly fatal. Patients usually present with a persistent headache, which is occasionally accompanied by lethargy and confusion. Nuchal rigidity, if present, is not severe. Examination of cerebrospinal fluid (CSF) demonstrates lymphocytic pleocytosis with profound hypoglycorrhachia and elevated protein levels. CSF eosinophilia is occasionally documented. With or without appropriate therapy, patients may develop hydrocephalus, which presents clinically as a marked decline in mental status, often with gait disturbances.

Diagnosis

As mentioned above, coccidioidomycosis is often misdiagnosed as community-acquired bacterial pneumonia. Serology plays an important role in establishing the diagnosis of coccidioidomycosis. Several techniques are available, including the traditional tube-precipitin (TP) and complement-fixation (CF) assays, immunodiffusion (IDTP and IDCF), and enzyme immunoassay (EIA) to detect IgM and IgG antibodies. TP and IgM antibodies are found in serum soon after infection and persist for weeks. They are not useful for gauging disease progression and are not found in the CSF. The CF and IgG antibodies occur later in the course of the disease and persist longer than TP and IgM antibodies. Rising CF titers are associated with clinical progression, and the presence of CF antibody in CSF is an indicator for coccidioidal meningitis. Antibodies disappear over time in persons whose clinical illness resolves.

Because of its commercial availability, the coccidioidal EIA is frequently used as a screening tool for coccidioidal serology. There has been concern that the IgM EIA is occasionally falsely positive. In addition, while the sensitivity and specificity of the IgG EIA appear to be high when compared with those of the CF and IDCF assays, the optical density obtained in the EIA does not correlate with the serologic titer of either of the latter tests.

Coccidioides grows within 3–7 days at 37°C on a variety of artificial media, including blood agar. Therefore, it is always useful to obtain samples of sputum or other respiratory fluids and tissues for culture in suspected cases of coccidioidomycosis. The clinical laboratory should be alerted to the possibility of this diagnosis, since *Coccidioides* can pose a significant hazard to laboratory workers if it is inadvertently inhaled. The organism can also be identified directly. While treatment of samples with potassium hydroxide is rarely fruitful in establishing the diagnosis, examination of sputum or other respiratory fluids after Papanicolaou or Gomori methenamine silver staining reveals spherules in a significant proportion of patients with pulmonary coccidioidomycosis. For fixed tissues (e.g., those obtained from biopsy specimens), spherules with surrounding inflammation can be demonstrated with hematoxylin-eosin or Gomori methenamine silver staining.

A commercially available test for coccidioidal antigenuria and antigenemia that appears to be useful for the diagnosis of coccidioidomycosis, particularly in immunosuppressed patients with severe or disseminated disease, has been developed. However, this test can yield false-positive results,

especially in cases of histoplasmosis and blastomycosis. Some laboratories offer genomic detection by polymerase chain reaction.

Treatment: Coccidioidomycosis

Currently, two main classes of antifungal agents are useful for the treatment of coccidioidomycosis (Table 200-1). While once routinely prescribed, amphotericin B in all its formulations is now reserved for only the most severe cases of dissemination and for intrathecal or intraventricular administration to patients with coccidioidal meningitis in whom triazole therapy has failed. The original formulation of amphotericin B, which is dispersed with deoxycholate, is usually administered intravenously in doses of 0.7–1.0 mg/kg either daily or three times per week. The newer lipid-based formulations —amphotericin B lipid complex (ABLC), amphotericin B colloidal dispersion (ABCD), and amphotericin B liposomal complex—appear to offer no therapeutic advantage over the deoxycholate formulation for the treatment of coccidioidomycosis but are associated with less renal toxicity. The lipid dispersions are administered intravenously at doses of 5 mg/kg daily or three times per week.

Table 200-1 Clinical Presentations of Coccidioidomycosis, Their Frequency, and Recommended Initial Therapy for the Immunocompetent Host

Clinical Presentation	Frequency, %	Recommended Therapy
Asymptomatic	60	None
Primary pneumonia (focal)	40	In most cases, none[a]
Diffuse pneumonia	<1	Amphotericin B followed by prolonged oral triazole therapy
Pulmonary sequelae	5	
Nodule	—	None
Cavity	—	In most cases, none[b]
Chronic pneumonia	—	Prolonged triazole therapy
Disseminated disease	≤1	
Skin, bone, joint, soft tissue	—	Prolonged triazole therapy[c]
Meningitis	—	Life-long triazole therapy[d]

[a]Treatment is indicated for hosts with depressed cellular immunity as well as for those with prolonged symptoms and signs of increased severity, including night sweats for >3 weeks, weight loss of >10%, a complement-fixation titer of >1:16, and extensive pulmonary involvement on chest radiography.

[b]Treatment (usually with the oral triazoles fluconazole and itraconazole) is recommended for persistent symptoms.

[c]In severe cases, some clinicians would use amphotericin B as initial therapy.

[d]Intraventricular or intrathecal amphotericin B is recommended in cases of triazole failure. Hydrocephalus may occur, requiring a CSF shunt.

Note: See text for dosages and durations.

Triazole antifungals are the principal drugs now used to treat most cases of coccidioidomycosis. Clinical trials have demonstrated the usefulness of both fluconazole and itraconazole, and evidence indicates that itraconazole may be more efficacious against bone and joint disease. Because of its demonstrated penetration into CSF, fluconazole is the azole of choice for the treatment of coccidioidal meningitis, but itraconazole is also effective. For both drugs, a minimal oral adult dosage of 400 mg/d should be used. The maximal dose of itraconazole is 200 mg three times daily, but higher doses of fluconazole may be given. Two newer triazole antifungals, posaconazole and voriconazole, are now available. However, given the relative paucity of clinical data, the high cost, and (particularly for voriconazole) the potential toxicity, these agents should be reserved for cases that remain recalcitrant when treated with fluconazole or itraconazole. High-dose triazole therapy may be teratogenic, particularly during the first trimester; thus, amphotericin B should be considered as therapy for coccidioidomycosis in pregnant women.

Most patients with focal primary pulmonary coccidioidomycosis require no therapy. Patients for whom antifungal therapy should be considered include those with underlying cellular immunodeficiencies and those with prolonged symptoms and signs of extensive disease. Specific criteria include symptoms

persisting for ≥2 months, night sweats occurring for >3 weeks, weight loss of >10%, a serum CF antibody titer of >1:16, and extensive pulmonary involvement apparent on chest radiography.

Diffuse pulmonary coccidioidomycosis represents a special situation. Because most patients with this form of disease are profoundly hypoxemic and critically ill, many clinicians favor beginning therapy with amphotericin B and switching to an oral triazole once clinical improvement occurs.

The nodules that may follow primary pulmonary coccidioidomycosis do not require treatment. As noted above, these nodules are not easily distinguished from pulmonary malignancies by means of radiographic imaging. Close clinical follow-up and biopsy may be required to distinguish between these two entities. Most pulmonary cavities do not require therapy. Antifungal treatment should be considered in patients with persistent cough, pleuritic chest pain, and hemoptysis. Occasionally, pulmonary coccidioidal cavities become secondarily infected. This development is usually manifested by an air-fluid level within the cavity. Bacterial flora or *Aspergillus* species are commonly involved, and therapy directed at these organisms should be considered. Surgery is rarely required except in cases of persistent hemoptysis or pyopneumothorax. For chronic pulmonary coccidioidomycosis, prolonged antifungal therapy—lasting for at least 1 year—is usually required, with monitoring of symptoms, radiographic changes, sputum cultures, and serologic titers.

Most cases of disseminated coccidioidomycosis require prolonged antifungal therapy. Duration of treatment is based on resolution of the signs and symptoms of the lesion in conjunction with a significant decline in serum CF antibody titer. Such therapy routinely is continued for at least several years. Relapse occurs in 15–30% of individuals once therapy is discontinued.

Coccidioidal meningitis poses a special challenge. While most patients with this form of disease respond to treatment with oral triazoles, 80% experience relapse when therapy is stopped. Thus, life-long therapy is recommended. In cases of triazole failure, intrathecal or intraventricular amphotericin B may be used. Installation requires considerable expertise and should be performed only by an experienced health care provider. Shunting of CSF in addition to appropriate antifungal therapy is required in cases of meningitis complicated by hydrocephalus. It is prudent to obtain expert consultation in all cases of coccidioidal meningitis.

Prevention

There are no proven methods to reduce the risk of acquiring coccidioidomycosis among residents of an endemic region. Avoidance of direct contact with uncultivated soil or with visible dust containing soil presumably reduces the risk. Targeted prophylactic antifungal therapy is appropriate in patients who have evidence of active or recent coccidioidomycosis and are about to undergo allogeneic solid-organ transplantation. Data on the use of antifungal agents for prophylaxis in other situations are scanty. However, most experts would administer triazole antifungal therapy to patients with a history of active coccidioidomycosis or a positive coccidioidal serology in whom therapy with TNF–α antagonists is being initiated.

FURTHER READINGS

Binnicker MJ et al: Detection of *Coccidioides* species in clinical specimens by real-time PCR. J Clin Microbiol 45:173, 2007 [PMID: 17108077]

Blair JE: Approach to the solid organ transplantation patient with latent infection and disease caused by *Coccidioides* species. Curr Opinion Infect Dis 21:415, 2008 [PMID: 18594295]

Durkin M et al: Detection of *Coccidioides* antigenemia following dissociation of immune complexes. Clin Vaccine Immunol 16:1453, 2009 [PMID: 19675225]

Drutz DJ, Catanzaro A: Coccidioidomycosis (parts I and II). Am Rev Respir Dis 117:559 and 727, 1978

Fisher MC et al: Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. Mycologia 94:73, 2002 [PMID: 21156479]

Galgiani JN et al: Coccidioidomycosis. Clin Infect Dis 41:1217, 2005 [PMID: 16206093]

——— et al: Comparison of oral fluconazole and itraconazole for progressive, nonmeningeal coccidioidomycosis. A randomized, double-blind trial. Mycoses Study Group. Ann Intern Med 133:676, 2000

Kuberski T et al: False-positive IgM serology in coccidioidomycosis. J Clin Microbiol 48:2047, 2010 [PMID: 20357210]

Massanant FY, Ampel NM: Coccidioidomycosis in patients with HIV-1 infection in the era of potent antiretroviral therapy. Clin Infect Dis 50:1, 2010

Valdivia L et al: Coccidioidomycosis as a common cause of community-acquired pneumonia. Emerg Infect Dis 12:958, 2006 [PMID: 16707052]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BLASTOMYCOSIS: INTRODUCTION

Blastomycosis is a systemic pyogranulomatous infection, involving primarily the lungs, that arises after inhalation of the conidia of *Blastomyces dermatitidis*. Pulmonary blastomycosis varies from an asymptomatic infection to acute or chronic pneumonia. Hematogenous dissemination occurs frequently. Extrapulmonary disease of the skin, bones, and genitourinary system is common, but almost any organ can be infected.

Etiologic Agent

B. dermatitidis is the asexual state of *Ajellomyces dermatitidis*. Two serotypes have been identified on the basis of the presence or absence of the A antigen. *B. dermatitidis* exhibits thermal dimorphism, growing as the mycelial phase at room temperature and as the yeast phase at 37°C. Primary isolation is most dependable for the mycelial phase incubated at 30°C. Definitive identification usually requires conversion to the yeast phase at 37°C or, more commonly, the use of nucleic acid amplification techniques (e.g., AccuProbe, Gen-Probe, San Diego, CA) that detect mycelial-phase growth. Yeast cells are usually 8–15 μm in diameter, have thick refractile cell walls, are multinucleate, and reproduce by a single, large, broad-based bud.

Epidemiology

Most cases of blastomycosis have been reported in North America. Endemic areas include the southeastern and south-central states bordering the Mississippi and Ohio river basins, the midwestern states and Canadian provinces bordering the Great Lakes, and a small area in New York and Canada along the St. Lawrence River. Outside North America, blastomycosis has been reported most frequently in Africa.

Early studies of endemic cases indicated that middle-aged men with outdoor occupations were at greatest risk. Reported outbreaks, however, do not suggest a predilection according to sex, age, race, occupation, or season. *B. dermatitidis* probably grows as microfoci in the warm, moist soil of wooded areas rich in organic debris. Exposure to soil, whether related to work or recreation, appears to be the common factor associated with infection.

Pathogenesis

After inhalation, the conidia of *B. dermatitidis* are susceptible to phagocytosis and killing in the lungs by polymorphonuclear leukocytes, monocytes, and alveolar macrophages. This phagocytic response represents innate immunity and probably explains the high frequency of asymptomatic infections in outbreaks. Conidia that escape phagocytosis rapidly convert to the yeast phase in tissue. The greater resistance of the thick-walled yeast form to phagocytosis and killing probably contributes to infection. This yeast-phase conversion also induces the expression of the 120-kDa glycoprotein BAD-1, which is an adhesin, an essential virulence factor, and the major epitope for humoral and cellular immunity. The primary acquired host defense against *B. dermatitidis* is cellular immunity mediated by antigen-specific T cells and lymphokine-activated macrophages.

Approach to the Patient: Blastomycosis

Whether acute or chronic, blastomycosis mimics many other disease processes. For example, acute pulmonary blastomycosis may present with signs and symptoms indistinguishable from those of bacterial pneumonia or influenza. Chronic pulmonary blastomycosis most commonly mimics malignancy or tuberculosis. Skin lesions are often misdiagnosed as basal cell or squamous cell carcinoma, pyoderma gangrenosum, or keratoacanthoma. Laryngeal lesions are frequently mistaken for squamous cell carcinoma. Thus, the clinician must maintain a high index of suspicion and perform a careful histologic evaluation of secretions or biopsy material from patients who live in or have visited regions endemic for blastomycosis.

Clinical Manifestations

Acute pulmonary infection is usually diagnosed in association with point-source outbreaks and is accompanied by the abrupt onset of fever, chills, pleuritic chest pain, arthralgias, and myalgias. Cough is initially nonproductive but frequently becomes purulent as disease progresses. Chest radiographs usually reveal alveolar infiltrates with consolidation. Pleural effusions and hilar adenopathy are uncommon. Most patients diagnosed with pulmonary blastomycosis have chronic indolent pneumonia with signs and symptoms of fever, weight loss, productive cough, and hemoptysis. The most common

radiologic findings are alveolar infiltrates with or without cavitation, mass lesions that mimic bronchogenic carcinoma, and fibronodular infiltrates. Respiratory failure (adult respiratory distress syndrome) associated with miliary disease or diffuse pulmonary infiltrates is more common among immunocompromised patients, especially those in the late stages of AIDS (Chap. 189). Mortality rates are ≥50% among these patients, and most deaths occur within the first few days of therapy.

Skin disease is the most common extrapulmonary manifestation of blastomycosis. Two types of skin lesions occur: verrucous (more common) and ulcerative. Osteomyelitis is associated with as many as one-fourth of *B. dermatitidis* infections. The vertebrae, pelvis, sacrum, skull, ribs, or long bones are most frequently involved. Patients with *B. dermatitidis* osteomyelitis often present with contiguous soft-tissue abscesses or chronic draining sinuses. In men, blastomycosis may involve the prostate and epididymis. Central nervous system (CNS) disease occurs in <5% of immunocompetent patients with blastomycosis. In AIDS patients, however, CNS disease has been reported in ~40% of cases, usually presenting as a brain abscess. Less common forms of CNS disease are cranial or spinal epidural abscess and meningitis.

Diagnosis

Definitive diagnosis of blastomycosis requires growth of the organism from sputum, pus, or biopsy material. A presumptive diagnosis is made by visualization of the characteristic broad-based budding yeast in clinical specimens. Serologic diagnosis of blastomycosis is of limited usefulness because of cross-reactivity with other fungal antigens.

A *Blastomyces* antigen assay that detects antigen in urine and serum is commercially available (Mira Vista Diagnostics, Indianapolis, IN). Antigen detection in urine appears to be more sensitive than serum antigen detection. This antigen test may be useful for monitoring of patients during therapy or for early detection of relapse. Molecular identification techniques, including DNA probe hybridization, are commercially available but are currently used only to supplement traditional methods of diagnosis.

Treatment: Blastomycosis

The Infectious Diseases Society of America has published guidelines for the treatment of blastomycosis. Selection of an appropriate therapeutic regimen must be based on the clinical form and severity of the disease, the immune status of the patient, and the toxicity of the antifungal agent (Table 201-1). Although spontaneous cures of acute pulmonary infection have been well documented, there are no criteria by which to distinguish patients whose disease will progress or disseminate. Thus, almost all patients with blastomycosis should be treated.

Table 201-1 Treatment of Blastomycosis

Disease	Primary Therapy	Alternative Therapy
Immunocompetent Patient/Life-Threatening Disease		
Pulmonary	Lipid AmB, 3–5 mg/kg qd, *or* Deoxycholate AmB, 0.7–1.0 mg/kg qd (total dose: 1.5–2.5 g)	Itraconazole, 200–400 mg/d (once patient's condition has stabilized)
Disseminated		
CNS	Lipid AmB, 3–5 mg/kg qd, *or* Deoxycholate AmB, 0.7–1.0 mg/kg qd (total dose: at least 2 g)	Fluconazole, 800 mg/d (if patient is intolerant to full course of AmB)
Non-CNS	Lipid AmB, 3–5 mg/kg qd, *or* Deoxycholate AmB, 0.7–1.0 mg/kg qd (total dose: 1.5–2.5 g)	Itraconazole, 200–400 mg/d (once patient's condition has stabilized)
Immunocompetent Patient/Non-Life-Threatening Disease		
Pulmonary or disseminated (non-CNS)	Itraconazole, 200–400 mg/d, or Lipid AmB, 3–5 mg/kg qd, *or* Deoxycholate AmB, 0.5–0.7 mg/kg qd (in patients intolerant to itraconazole or whose disease progresses despite therapy)	Fluconazole, 400–800 mg/d, *or* Ketoconazole, 400–800 mg/d

Immunocompromised Patient[a]		
All infections	Lipid AmB, 3–5 mg/kg qd, *or* Deoxycholate AmB, 0.7–1.0 mg/kg qd (total dose: 1.5–2.5 g)	Itraconazole, 200–400 mg/d (non-CNS disease, once clinically improved)

[a]Suppressive therapy with itraconazole may be considered for patients whose immunocompromised state continues. Fluconazole (800 mg/d) may be useful for patients who have CNS disease or cannot tolerate itraconazole.

Abbreviations: AmB, amphotericin B; CNS, central nervous system.

Itraconazole is the agent of choice for immunocompetent patients with mild to moderate pulmonary or non-CNS extrapulmonary disease. Therapy is continued for 6–12 months. Amphotericin B is the preferred initial treatment for patients who are severely immunocompromised, who have life-threatening disease or CNS disease, or whose disease progresses during treatment with itraconazole. Although not rigorously studied, lipid formulations of amphotericin B can provide an alternative for patients who cannot tolerate amphotericin B deoxycholate. Most patients with non-CNS disease whose clinical condition improves after an initial course of amphotericin B (usually 2 weeks in duration) can be switched to itraconazole to complete 6–12 months of therapy. Fluconazole, because of its excellent penetration of the CNS, may have a role in the treatment of patients with brain abscess or meningitis after an initial course of amphotericin B.

Voriconazole has been used successfully to treat refractory blastomycosis, blastomycosis in immunosuppressed patients, and—given its good CSF penetration—CNS blastomycosis. There are no data to support the use of posaconazole in human cases of blastomycosis. The echinocandins have variable activity against *B. dermatitidis* and have no place in the treatment of blastomycosis.

PROGNOSIS

Clinical and mycologic response rates are 90–95% among compliant immunocompetent patients given itraconazole for mild to moderate pulmonary and extrapulmonary disease without CNS involvement. Bone and joint disease usually requires 12 months of therapy. The <5% of infections that relapse after an initial course of itraconazole usually respond well to a second treatment course.

FURTHER READINGS

Bradsher RW: Blastomycosis, in *Clinical Mycology*, WE Dismukes et al (eds). New York, Oxford University Press, 2003, pp 299–310

Chapman SW et al: Blastomycosis. Infect Dis Clin North Am 17:21, 2003 [PMID: 12751259]

———, Sullivan DC: *Blastomyces dermatitidis*, in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). New York, Churchill Livingstone, 2010, pp 3319–3332

———, Sullivan DC: Diagnosis and treatment of blastomycosis, in *Diagnosis and Treatment of Human Mycoses*, D Hospental, M Rinaldi (eds). Totowa, NJ, Humana Press, 2007

——— et al: Clinical practice guidelines for the management of blastomycosis: 2008 update by the Infectious Diseases Society of America. Clin Infect Dis 46:1801, 2008 (updates: www.idsociety.org)

Deepe GS et al: Progress in vaccination for histoplasmosis and blastomycosis: Coping with cellular immunity. Med Mycol 43:381, 2005 [PMID: 16178365]

Light RB et al: Seasonal variations in the clinical presentation of pulmonary and extrapulmonary blastomycosis. Med Mycol 46:835, 2008

Mason AR et al: Cutaneous blastomycosis: A diagnostic challenge. Int J Dermatol 47:824, 2008 [PMID: 18717863]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CRYPTOCOCCOSIS: INTRODUCTION

Definition and Etiology

Cryptococcus, a genus of yeast-like fungi, is the etiologic agent of cryptococcosis. There are two species, *C. neoformans* and *C. gattii*, each of which can cause cryptococcosis in humans. *C. neoformans* occurs in two varieties known as *grubii* and *neoformans*, which correlate with serotypes A and D, respectively. *C. gattii* has not been divided into varieties but is also antigenically diverse, consisting of serotypes B and C. Most clinical microbiology laboratories do not routinely distinguish between *C. neoformans* and *C. gattii* or among varieties, but rather identify and report all isolates simply as *C. neoformans*.

Epidemiology

Cryptococcosis was first described in the 1890s but remained relatively rare until the mid-twentieth century, when advances in diagnosis and increases in the number of immunosuppressed individuals markedly raised its reported prevalence. The spectrum of disease caused by *Cryptococcus* species consists predominantly of meningoencephalitis and pneumonia, but skin and soft tissue infections also occur. Serologic studies have shown that, although evidence for cryptococcal *infection* is common among immunocompetent individuals, cryptococcal *disease* (cryptococcosis) is relatively rare in the absence of impaired immunity. Individuals at high risk for cryptococcosis include patients with hematologicmalignancies, recipients of solid organ transplants who require ongoing immunosuppressive therapy, persons whose medical conditions necessitate glucocorticoid therapy, and patients with advanced HIV infection and CD4+ T lymphocyte counts of <200/μL.

Since the onset of the HIVpandemic in the early 1980s, the overwhelming majority of cryptococcosis cases have occurred in patients with AIDS (Chap. 189). To understand the impact of HIV infection on the epidemiology of cryptococcosis, it is instructive to note that in the early 1990s there were >1000 cases of cryptococcal meningitis each year in New York City—a figure far exceeding that for all cases of bacterial meningitis. With the advent of effective antiretroviral therapy, the incidence of AIDS-related cryptococcosis has been sharply reduced among treated individuals; however, the disease remains distressingly common in regions where antiretroviral therapy is not readily available, such as Africa and Asia, where up to one-third of patients with AIDS have cryptococcosis. The global burden of cryptococcosis was recently estimated at ~1 million cases, with >600,000 deaths annually. Thus cryptococci are major human pathogens.

Cryptococcal infection is acquired from the environment. *C. neoformans* and *C. gattii* inhabit different ecologic niches. *C. neoformans* is frequently found in soils contaminated with avian excreta and can easily be recovered from shaded and humid soils contaminated with pigeon droppings. In contrast, *C. gattii* is not found in bird feces. Instead, it inhabits a variety of arboreal species, including several types of eucalyptus tree. *C. neoformans* strains are found throughout the world; however, var. *grubii* (serotype A) strains are far more common than var. *neoformans* (serotype D) strains among both clinical and environmental isolates. The geographic distribution of *C. gattii* was thought to be largely limited to tropical regions until an outbreak of cryptococcosis caused by a new serotype B strain began in Vancouver in 1999. This outbreak has extended into the United States, and *C. gattii* is now being encountered in several states in the Pacific Northwest. In addition to the different geographic distributions of the two cryptococcal species, individual susceptibility to these species affects epidemiology. Cryptococcosis caused by the *C. neoformans* varieties occurs mostly in individuals with AIDS (Chap. 189) and other forms of impaired immunity. In contrast, *C. gattii*–related disease is not associated with specific immune deficits and often occurs in immunocompetent individuals.

Pathogenesis

Cryptococcal infection is acquired by inhalation of aerosolized infectious particles. The exact nature of these particles is not known; the two leading candidate forms are small desiccated yeast cells and basidiospores. Little is known about the pathogenesis of initial infection. Serologic studies have shown that cryptococcal infection is acquired in childhood, but it is not known whether the initial infection is symptomatic. Given that cryptococcal infection is common while disease is rare, the consensus is that pulmonary defense mechanisms in immunologically intact individuals are highly effective at containing this fungus. It is not clear whether initial infection leads to a state of immunity or whether most individuals are subject throughout life to frequent and recurrent infections that resolve without clinical

disease. However, evidence indicates that some human cryptococcal infections lead to a state of latency in which viable organisms are harbored for prolonged periods, possibly in granulomas. Thus the inhalation of cryptococcal cells and/or spores can be followed by either clearance or establishment of the latent state. The consequences of prolonged harboring of cryptococcal cells in the lung are not known, but evidence from animal studies indicates that the organisms' prolonged presence could alter the immunologic milieu in the lung and predispose to allergic airway disease.

Cryptococcosis usually presents clinically as chronic meningoencephalitis. The mechanisms by which the fungus undergoes extrapulmonary dissemination and enters the central nervous system (CNS) remain poorly understood. The mechanism by which cryptococcal cells cross the blood-brain barrier is a subject of intensive study. Current evidence suggests either direct fungal-cell migration across the endothelium or fungal-cell carriage inside macrophages as "Trojan horse" invaders. *Cryptococcus* species have well-defined virulence factors that include the polysaccharide capsule, the ability to make melanin, and the elaboration of enzymes (e.g., phospholipase and urease) that enhance the survival of fungal cells in tissue. Among these virulence factors, the capsule and melanin production have been most extensively studied. The cryptococcal capsule is antiphagocytic, and the capsular polysaccharide has been associated with numerous deleterious effects on host immune function. Cryptococcal infections can elicit little or no tissue inflammatory response. The immune dysfunction seen in cryptococcosis has been attributed to the release of copious amounts of capsular polysaccharide into tissues, where it probably interferes with local immune responses (Fig. 202-1). In clinical practice, the capsular polysaccharide is the antigen that is measured as a diagnostic marker of cryptococcal infection.

Figure 202-1

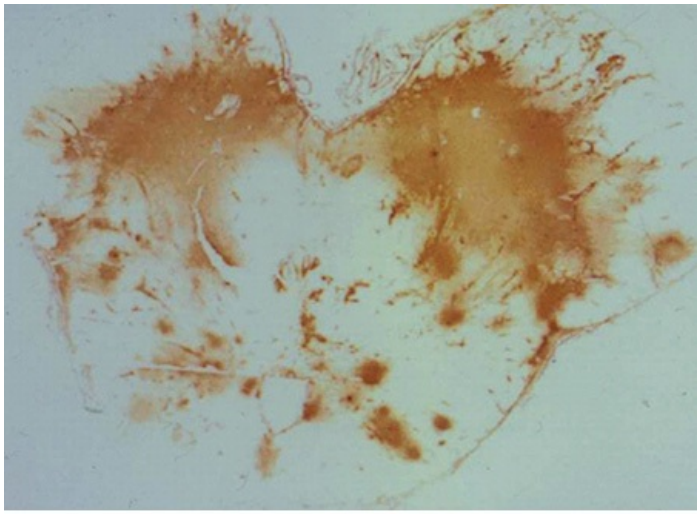

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cryptococcal antigen in human brain tissue, as revealed by immunohistochemical staining. Brown areas show polysaccharide deposits in the midbrain of a patient who died of cryptococcal meningitis. (*Reprinted with permission from SC Lee et al: Hum Pathol 27:839, 1996.*)

Approach to the Patient: Cryptococcosis

Cryptococcosis should be included in the differential diagnosis when any patient presents with findings suggestive of chronic meningitis. Concern about cryptococcosis is heightened by a history of headache and neurologic symptoms in a patient with an underlying immunosuppressive disorder or state that is associated with an increased incidence of cryptococcosis, such as advanced HIV infection or solid organ transplantation.

Clinical Manifestations

The clinical manifestations of cryptococcosis reflect the site of fungal infection. Although infection can affect any tissue or organ, the majority of cases that come to clinical attention involve the CNS and/or the lungs. CNS involvement usually presents as signs and symptoms of chronic meningitis, such as headache, fever, lethargy, sensory deficits, memory deficits, cranial nerve paresis, vision deficits, and meningismus. Cryptococcal meningitis differs from bacterial meningitis in that many *Cryptococcus*-infected patients present with symptoms of several weeks' duration. In addition, classic characteristics of meningeal irritation, such as meningismus, may be absent in cryptococcal meningitis. Indolent cases can present as subacutedementia. Meningeal cryptococcosis can lead to sudden catastrophic vision loss.

Pulmonary cryptococcosis usually presents as cough, increased sputum production, and chest pain.

Patients infected with *C. gattii* can present with granulomatous pulmonary masses known as *cryptococcomas*. Fever develops in a minority of cases. Like CNS disease, pulmonary cryptococcosis can follow an indolent course, and the majority of cases probably do not come to clinical attention. In fact, many cases are discovered incidentally during the workup of an abnormal chest radiograph obtained for other diagnostic purposes. Pulmonary cryptococcosis can be associated with antecedent diseases such as malignancy, diabetes, and tuberculosis.

Skin lesions are common in patients with disseminated cryptococcosis and can be highly variable, including papules, plaques, purpura, vesicles, tumor-like lesions, and rashes. The spectrum of cryptococcosis in HIV-infected patients is so varied and has changed so much since the advent of antiretroviral therapy that a distinction between HIV-related and HIV-unrelated cryptococcosis is no longer pertinent. In patients with AIDS and solid organ transplant recipients, the lesions of cutaneous cryptococcosis often resemble those of molluscum contagiosum (Fig. 202-2; Chap. 183).

Figure 202-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Disseminated fungal infection. A liver transplant recipient developed six cutaneous lesions similar to the one shown. Biopsy and serum antigen testing demonstrated *Cryptococcus*. Important features of the lesion include a benign-appearing fleshy papule with central umbilication resembling molluscum contagiosum. (*Photo courtesy of Dr. Lindsey Baden; with permission.*)

Diagnosis

A diagnosis of cryptococcosis requires the demonstration of yeast cells in normally sterile tissues. Visualization of the capsule of fungal cells in cerebrospinal fluid (CSF) mixed with india ink is a useful rapid diagnostic technique. Cryptococcal cells in india ink have a distinctive appearance because their capsules exclude ink particles. However, the CSF india ink examination may yield negative results in patients with a low fungal burden. This examination should be performed by a trained individual, since leukocytes and fat globules can sometimes be mistaken for fungal cells. Cultures of CSF and blood that are positive for cryptococcal cells are diagnostic for cryptococcosis. In cryptococcal meningitis, CSF examination usually reveals evidence of chronic meningitis with mononuclear cell pleocytosis and increased protein levels. A particularly useful test is cryptococcal antigen (CRAg) detection in CSF and blood. The assay is based on serologic detection of cryptococcal polysaccharide and is both sensitive and specific. A positive cryptococcal antigen test provides strong presumptive evidence for cryptococcosis; however, because the result is often negative in pulmonary cryptococcosis, the test is less useful in the diagnosis of pulmonary disease.

Treatment: Cryptococcosis

Both the site of infection and the immune status of the host must be considered in the selection of therapy for cryptococcosis. The disease has two general patterns of manifestation: (1) pulmonary cryptococcosis, with no evidence of extrapulmonary dissemination; and (2) extrapulmonary (systemic) cryptococcosis, with or without meningoencephalitis. Pulmonary cryptococcosis in an immunocompetent host sometimes resolves without therapy. However, given the propensity of

Cryptococcus species to disseminate from the lung, the inability to gauge the host's immune status precisely, and the availability of low-toxicity therapy in the form of fluconazole, the current recommendation is for pulmonary cryptococcosis in an immunocompetent individual to be treated with fluconazole (200–400 mg/d for 3–6 months). Extrapulmonary cryptococcosis without CNS involvement in an immunocompetent host can be treated with the same regimen, although amphotericin B (AmB; 0.5–1 mg/kg daily for 4–6 weeks) may be required for more severe cases. In general, extrapulmonary cryptococcosis without CNS involvement requires less intensive therapy—with the caveat that morbidity and death in cryptococcosis are associated with meningeal involvement. Thus the decision to categorize cryptococcosis as "extrapulmonary without CNS involvement" should be made only after careful evaluation of the CSF reveals no evidence of cryptococcal infection. For CNS involvement in a host without AIDS or obvious immune impairment, most authorities recommend initial therapy with AmB (0.5–1 mg/kg daily) during an induction phase, which is followed by prolonged therapy with fluconazole (400 mg/d) during a consolidation phase. For cryptococcal meningoencephalitis without a concomitant immunosuppressive condition, the recommended regimen is AmB (0.5–1 mg/kg) plus flucytosine (100 mg/kg) daily for 6–10 weeks. Alternatively, patients can be treated with AmB (0.5–1 mg/kg) plus flucytosine (100 mg/kg) daily for 2 weeks and then with fluconazole (400 mg/d) for at least 10 weeks. Patients with immunosuppression are treated with the same initial regimens except that consolidation therapy with fluconazole is given for a prolonged period to prevent relapse.

Cryptococcosis in patients with HIV infection always requires aggressive therapy and is considered incurable unless immune function improves. Consequently, therapy for cryptococcosis in the setting of AIDS has two phases: induction therapy (intended to reduce the fungal burden and alleviate symptoms) and lifelong maintenance therapy (to prevent a symptomatic clinical relapse). Pulmonary and extrapulmonary cryptococcosis without evidence of CNS involvement can be treated with fluconazole (200–400 mg/d). In patients who have more extensive disease, flucytosine (100 mg/kg per day) may be added to the fluconazole regimen for 10 weeks, with lifelong fluconazole maintenance therapy thereafter. For HIV-infected patients with evidence of CNS involvement, most authorities recommend induction therapy with AmB. An acceptable regimen is AmB (0.7–1 mg/kg) plus flucytosine (100 mg/kg) daily for 2 weeks followed by fluconazole (400 mg/d) for at least 10 weeks and then by lifelong maintenance therapy with fluconazole (200 mg/d). Fluconazole (400–800 mg/d) plus flucytosine (100 mg/kg per day) for 6–10 weeks followed by fluconazole (200 mg/d) as maintenance therapy can be used as an alternative. Newer triazoles like voriconazole are highly active against cryptococcal strains, but clinical experience with these agents in the treatment of cryptococcosis is limited. Lipid formulations of AmB can be substituted for AmB deoxycholate in patients with renal impairment. Neither caspofungin nor micafungin is effective against *Cryptococcus* species; consequently, neither drug has a role in the treatment of cryptococcosis. Cryptococcal meningoencephalitis is often associated with increased intracranial pressure, which is believed to be responsible for damage to the brain and cranial nerves. Appropriate management of CNS cryptococcosis requires careful attention to the management of intracranial pressure, including the reduction of pressure by repeated therapeutic lumbar puncture and the placement of shunts.

In HIV-infected patients with previously treated cryptococcosis who are receiving fluconazole maintenance therapy, it may be possible to discontinue antifungal drug treatment if antiretroviral therapy results in immunologic improvement. However, certain recipients of maintenance therapy who have a history of successfully treated cryptococcosis can develop a troublesome immune reconstitution syndrome when antiretroviral therapy produces a rebound in immunologic function.

PROGNOSIS AND COMPLICATIONS

Even with antifungal therapy, cryptococcosis is associated with high rates of morbidity and death. For the majority of patients with cryptococcosis, the most important prognostic factor is the extent and the duration of the underlying immunologic deficits that predisposed them to develop the disease. Therefore, cryptococcosis is often curable with antifungal therapy in individuals with no apparent immunologic dysfunction, but, in patients with severe immunosuppression (e.g., those with AIDS), the best that can be hoped for is that antifungal therapy will induce remission, which can then be maintained with lifelong suppressive therapy. Before the advent of antiretroviral therapy, the median overall survival period for AIDS patients with cryptococcosis was <1 year. Cryptococcosis in patients with underlying neoplastic disease has a particularly poor prognosis. For CNS cryptococcosis, poor prognostic markers are a positive CSF assay for yeast cells by initial india ink examination (evidence of a heavy fungal burden), high CSF pressure, low CSF glucose levels, low CSF pleocytosis (<2/μL), recovery of yeast cells from extraneural sites, absence of antibody to capsular polysaccharide, a CSF or serum cryptococcal antigen level of ≥1:32, and concomitant glucocorticoid therapy or hematologicmalignancy. A response to treatment does not guarantee cure since relapse of cryptococcosis is common even among patients with relatively intact immune systems. Complications of CNS cryptococcosis include cranial nerve deficits, vision loss, and cognitive impairment.

Prevention

No vaccine is available for cryptococcosis. In patients at high risk (e.g., those with advanced HIV infection and CD4+ T lymphocyte counts of <200/μL), primary prophylaxis with fluconazole (200 mg/d)

is effective in reducing the prevalence of disease. Since antiretroviral therapy raises the CD4+ T lymphocyte count, it constitutes an immunologic form of prophylaxis. However, cryptococcosis in the setting of immune reconstitution has been reported in patients with HIV infection and recipients of solid organ transplants.

FURTHER READINGS

Datta K et al: Spread of *Cryptococcus gattii* into the Pacific Northwest region of the United States. Emerg Infect Dis 15:1185, 2009 [PMID: 19757550]

Liliang P et al: Use of ventriculoperitoneal shunts to treat uncontrollable intracranial hypertension in patients who have cryptococcal meningitis without hydrocephalus. Clin Infect Dis 34:E64, 2002

Lortholary O et al: Incidence and risk factors of immune reconstitution inflammatory syndrome complicating HIV-associated cryptococcosis in France. AIDS 19:1043, 2005 [PMID: 15958835]

Masur H et al: Guidelines for preventing opportunistic infections among HIV-infected persons—2002. Ann Intern Med 137:435, 2002 [PMID: 12617574]

Park BJ et al: Estimation of the global burden of cryptococcal meningitis among persons living with HIV/AIDS. AIDS 23:525, 2009 [PMID: 19182676]

Perfect JR et al: Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the Infectious Diseases Society of America. Clin Infect Dis 50:291, 2010 [PMID: 20047480]

Saag MS et al: Practice guidelines for the management of cryptococcal disease. Clin Infect Dis 30:710, 2000 [PMID: 10770733]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CANDIDIASIS: INTRODUCTION

The genus *Candida* encompasses more than 150 species, only a few of which cause disease in humans. With rare exceptions, the human pathogens are *C. albicans, C. guilliermondii, C. krusei, C. parapsilosis, C. tropicalis, C. kefyr, C. lusitaniae, C. dubliniensis,* and *C. glabrata*. Ubiquitous in nature, these organisms are found on inanimate objects, in foods, and on animals and are normal commensals of humans. They inhabit the gastrointestinal tract (including the mouth and oropharynx), the female genital tract, and the skin. Although cases of candidiasis have been described since antiquity in debilitated patients, the advent of *Candida* species as common human pathogens dates to the introduction of modern therapeutic approaches that suppress normal host defense mechanisms. Of these relatively recent advances, the most important is the use of antibacterial agents that alter the normal human microbial flora and allow nonbacterial species to become more prevalent in the commensal flora. With the introduction of antifungal agents, the causes of *Candida* infections shifted from an almost complete dominance of *C. albicans* to the common involvement of *C. glabrata* and the other species listed above. The non-albicans species now account for approximately half of all cases of candidemia and hematogenously disseminated candidiasis. Recognition of this change is clinically important, since the various species differ in susceptibility to the newer antifungal agents. In developed countries, where medical therapeutics are commonly used, *Candida* species are now among the most common nosocomial pathogens. In the United States, these species are the fourth most common isolates from the blood of hospitalized patients.

Candida is a small, thin-walled, ovoid yeast that measures 4–6 μm in diameter and reproduces by budding. Organisms of this genus occur in three forms in tissue: blastospores, pseudohyphae, and hyphae. *Candida* grows readily on simple medium; lysis centrifugation enhances its recovery from blood. Species are identified by biochemical testing (currently with automated devices) or on special agar.

Pathogenesis

In the most serious form of *Candida* infection, the organisms disseminate hematogenously and form microabscesses and small macroabscesses in major organs. Although the exact mechanism is not known, *Candida* probably enters the bloodstream from mucosal surfaces after growing to large numbers as a consequence of bacterial suppression by antibacterial drugs; alternatively, in some instances, the organism may enter from the skin. A change from the blastospore stage to the pseudohyphal and hyphal stages is generally considered integral to the organism's penetration into tissue. However, *C. glabrata* can cause extensive infection even though it does not transform into pseudohyphae or hyphae. Numerous reviews of cases of hematogenously disseminated candidiasis have identified the following predisposing factors or conditions: antibacterial agents, indwelling intravascular catheters, hyperalimentation fluids, indwelling urinary catheters, parenteral glucocorticoids, respirators, neutropenia, abdominal and thoracic surgery, cytotoxic chemotherapy, and immunosuppressive agents for organ transplantation. Patients with severe burns, low-birth-weight neonates, and persons using illicit IV drugs are also susceptible. HIV-infected patients with low CD4+ T cell counts and patients with diabetes are susceptible to mucocutaneous infection, which may eventually develop into the disseminated form when other predisposing factors are encountered. Women who receive antibacterial agents may develop vaginal candidiasis.

Innate immunity is the most important defense mechanism against hematogenously disseminated candidiasis, and the neutrophil is the most important component of this defense. Although many immunocompetent individuals have antibodies to *Candida*, the role of these antibodies in defense against the organism is not clear.

Clinical Manifestations

MUCOCUTANEOUS CANDIDIASIS

Thrush is characterized by white, adherent, painless, discrete or confluent patches in the mouth, tongue, or esophagus, occasionally with fissuring at the corners of the mouth. This form of *Candida* disease may also occur at points of contact with dentures. Organisms are identifiable in gram-stained scrapings from lesions. The occurrence of thrush in a young, otherwise healthy-appearing person should prompt an investigation for underlying HIV infection. More commonly, thrush is seen as a nonspecific manifestation of severe debilitating illness. Vulvovaginal candidiasis is accompanied by

pruritus, pain, and vaginal discharge that is usually thin but may contain whitish "curds" in severe cases.

Other *Candida* skin infections include *paronychia*, a painful swelling at the nail-skin interface; onychomycosis, a fungal nail infection rarely caused by this genus; *intertrigo*, an erythematous irritation with redness and pustules in the skin folds; *balanitis*, an erythematous-pustular infection of the glans penis; *erosio interdigitalis blastomycetica*, an infection between the digits of the hands or toes; *folliculitis*, with pustules developing most frequently in the area of the beard; *perianal candidiasis*, a pruritic, erythematous, pustular infection surrounding the anus; and *diaper rash*, a common erythematous-pustular perineal infection in infants. *Generalized disseminated cutaneous candidiasis*, another form of infection that occurs primarily in infants, is characterized by widespread eruptions over the trunk, thorax, and extremities. The diagnostic macronodular lesions of hematogenously disseminated candidiasis (Fig. 203-1) indicate a high probability of dissemination to multiple organs as well as the skin. While the lesions are seen predominantly in immunocompromised patients treated with cytotoxic drugs, they may also develop in patients without neutropenia.

Figure 203-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Macronodular skin lesions associated with hematogenously disseminated candidiasis. *Candida* organisms are usually but not always visible on histopathologic examination. The fungi grow when a portion of the biopsied specimen is cultured. Therefore, for optimal identification, both histopathology and culture should be performed. (*Image courtesy of Dr. Noah Craft and the Victor Newcomer collection at UCLA, archived by Logical Images, Inc.; with permission.*)

Chronic mucocutaneous candidiasis is a heterogeneous infection of the hair, nails, skin, and mucous membranes that persists despite intermittent therapy. The onset of disease usually comes in infancy or within the first two decades of life but in rare cases can come in later life. The condition may be mild and limited to a specific area of the skin or nails, or it may take a severely disfiguring form (*Candida* granuloma) characterized by exophytic outgrowths on the skin. The condition is usually associated with specific immunologic dysfunction; most frequently reported is a failure of T lymphocytes to proliferate or to stimulate cytokines in response to stimulation by *Candida* antigens in vitro. Approximately half of patients have associated endocrine abnormalities that together are designated the *a*utoimmune *p*oly*e*ndocrinopathy–*c*andidiasis–*e*ctodermal *d*ystrophy (APECED) syndrome. This syndrome is due to mutations in the autoimmune regulator (*AIRE*) gene and is most prevalent among Finns, Iranian Jews, Sardinians, northern Italians, and Swedes. Conditions that usually follow the onset of the disease include hypoparathyroidism, adrenal insufficiency, autoimmune thyroiditis, Graves' disease, chronic active hepatitis, alopecia, juvenile-onset pernicious anemia, malabsorption, and primary hypogonadism. In addition, dental enamel dysplasia, vitiligo, pitted nail dystrophy, and calcification of the tympanic membranes may occur. Patients with chronic mucocutaneous candidiasis rarely develop hematogenously disseminated candidiasis, probably because their neutrophil function remains intact.

DEEPLY INVASIVE CANDIDIASIS

Deeply invasive *Candida* infections may or may not be due to hematogenous seeding. Deep esophageal infection may result from penetration by organisms from superficial esophageal erosions; joint or deep wound infection from contiguous spread of organisms from the skin; kidney infection from catheter-initiated spread of organisms through the urinary tract; infection of intraabdominal organs and

the peritoneum from perforation of the gastrointestinal tract; and gallbladder infection from retrograde migration of organisms from the gastrointestinal tract into the biliary drainage system.

However, far more commonly, deeply invasive candidiasis is a result of hematogenous seeding of various organs as a complication of candidemia. Once the organism gains access to the intravascular compartment (either from the gastrointestinal tract or, less often, from the skin through the site of an indwelling intravascular catheter), it may spread hematogenously to a variety of deep organs. The brain, chorioretina (Fig. 203-2), heart, and kidneys are most commonly infected and the liver and spleen less commonly so (most often in neutropenic patients). In fact, nearly any organ can become involved, including the endocrine glands, pancreas, heart valves (native or prosthetic), skeletal muscle, joints (native or prosthetic), bone, and meninges. *Candida* organisms may also spread hematogenously to the skin and cause classic macronodular lesions (Fig. 203-1). Frequently, painful muscular involvement is also evident beneath the area of affected skin. Chorioretinal involvement and skin involvement are highly significant, since both findings are associated with a very high probability of abscess formation in multiple deep organs as a result of generalized hematogenous seeding. Ocular involvement (Fig. 203-2) may require specific treatment (e.g., partial vitrectomy or intraocular injection of antifungal agents) to prevent permanent blindness. An ocular examination is indicated for all patients with candidemia, whether or not they have ocular manifestations.

Figure 203-2

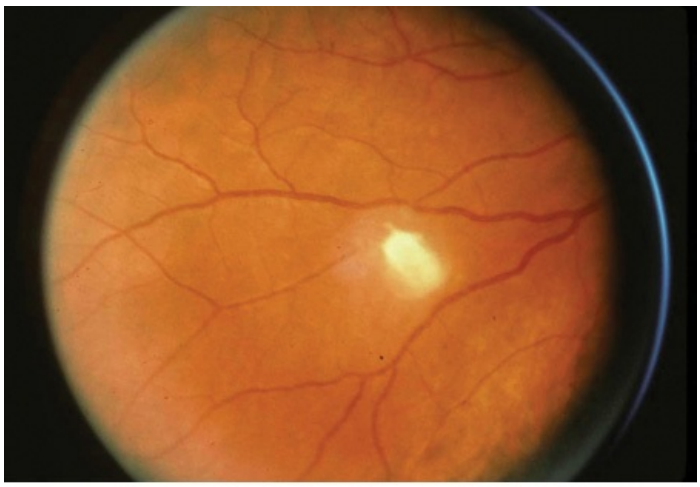

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hematogenous *Candida* endophthalmitis. A classic off-white lesion projecting from the chorioretina into the vitreous causes the surrounding haze. The lesion is composed primarily of inflammatory cells rather than organisms. Lesions of this type may progress to cause extensive vitreal inflammation and eventual loss of the eye. Partial vitrectomy, combined with IV and possibly intravitreal antifungal therapy, may be helpful in controlling the lesions. (*Image courtesy of Dr. Gary Holland; with permission.*)

DIAGNOSIS

The diagnosis of *Candida* infection is established by visualization of pseudohyphae or hyphae on wet mount (saline and 10% KOH), tissue Gram's stain, periodic acid–Schiff stain, or methenamine silver stain in the presence of inflammation. Absence of organisms on hematoxylin-eosin staining does not reliably exclude *Candida* infection. The most challenging aspect of diagnosis is determining which patients with *Candida* isolates have hematogenously disseminated candidiasis. For instance, recovery of *Candida* from sputum, urine, or peritoneal catheters may indicate mere colonization rather than deep-seated infection, and *Candida* isolation from the blood of patients with indwelling intravascular catheters may reflect inconsequential seeding of the blood from or growth of the organisms on the catheter. Despite extensive research into both antigen and antibody detection systems, there is currently no widely available and validated diagnostic test to distinguish patients with inconsequential seeding of the blood from those whose positive blood cultures represent hematogenous dissemination to multiple organs. Many studies are under way to establish the utility of the β-glucan test; at present, its greatest utility is its negative predictive value (~90%). Meanwhile, the presence of ocular or macronodular skin lesions is highly suggestive of widespread infection of multiple deep organs.

Treatment: *Candida* Infections

MUCOCUTANEOUS *CANDIDA* INFECTION

The treatment of mucocutaneous candidiasis is summarized in Table 203-1.

Table 203-1 Treatment of Mucocutaneous Candidal Infections

Disease	Preferred Treatment	Alternatives
Cutaneous	Topical azole	Topical nystatin
Vulvovaginal	Oral fluconazole (150 mg) or azole cream or suppository	Nystatin suppository
Thrush	Clotrimazole troches	Nystatin
Esophageal	Fluconazole tablets (100–200 mg/d) or itraconazole solution (200 mg/d)	Caspofungin, micafungin, or amphotericin B

CANDIDEMIA AND SUSPECTED HEMATOGENOUSLY DISSEMINATED CANDIDIASIS

All patients with candidemia are now treated with a systemic antifungal agent. A certain percentage of patients, including many of those who have candidemia associated with an indwelling intravascular catheter, probably have "benign" candidemia rather than deep-organ seeding. However, because there is no reliable way to distinguish benign candidemia from deep-organ infection, and because antifungal drugs less toxic than amphotericin B are available, it has become the standard of practice to treat all patients with candidemia, whether or not there is clinical evidence of deep-organ involvement. In addition, if an indwelling intravascular catheter may be involved, it is best to remove or replace the device whenever possible.

The drugs used for the treatment of candidemia and suspected disseminated candidiasis are listed in Table 203-2. Various lipid formulations of amphotericin B, three echinocandins, and the azoles fluconazole and voriconazole are used; no agent within a given class has been clearly identified as superior to the others. Most institutions choose an agent from each class on the basis of their own specific microbial epidemiology, strategies to minimize toxicities, and cost considerations. Unless azole resistance is considered likely, fluconazole is the agent of choice for the treatment of candidemia and suspected disseminated candidiasis in nonneutropenic, hemodynamically stable patients. Initial treatment in the context of likely azole resistance depends, as mentioned above, on the epidemiology of the individual hospital. For example, certain hospitals have a high rate of recovery of *C. glabrata*, while others do not. At institutions where non-*albicans Candida* spp. are frequently recovered, therapy with an echinocandin is typically started while the results of sensitivity testing are awaited. For hemodynamically unstable or neutropenic patients, initial treatment with broader-spectrum agents is desirable; these drugs include polyenes, echinocandins, or later-generation azoles such as voriconazole. Once the clinical response has been assessed and the pathogen specifically identified, the regimen can be altered accordingly. At present, the vast majority of *C. albicans* isolates are sensitive to fluconazole. Isolates of *C. glabrata* and *C. krusei* are less sensitive to fluconazole and more sensitive to polyenes and echinocandins. *C. parapsilosis* is less sensitive to echinocandins in vitro, although the clinical significance of this finding is not known.

Table 203-2 Agents for the Treatment of Disseminated Candidiasis

Agent	Route of Administration	Dose[a]	Comment
Amphotericin B deoxycholate	IV only	0.5–1.0 mg/kg daily	Being replaced by lipid formulations
Amphotericin B lipid formulations			Not FDA approved as primary therapy, but used commonly because less toxic than amphotericin B deoxycholate
Liposomal (AmBiSome, Abelcet)	IV only	3.0–5.0 mg/kg daily	
Lipid complex (ABLC)	IV only	3.0–5.0 mg/kg daily	
Colloidal dispersion (ABCD)	IV only	3.0–5.0 mg/kg daily	Associated with frequent infusion reactions
Azoles			
Fluconazole	IV and oral	400 mg/d	Most commonly used
Voriconazole	IV and oral	400 mg/d	Multiple drug interactions Approved for candidemia in nonneutropenic patients
Echinocandins			Broad spectrum against *Candida* species; approved for

			disseminated candidiasis
Caspofungin	IV only	50 mg/d	
Anidulafungin	IV only	100 mg/d	
Micafungin	IV only	100 mg/d	

[a]See Pappas et al. (2009) for loading doses and adjustments in renal failure. The recommended duration of therapy is 2 weeks beyond the last positive blood cultures and resolution of signs and symptoms of infection.

Note: Although ketoconazole is approved for the treatment of disseminated candidiasis, it has been replaced by the newer agents listed in this table. Posaconazole has been approved for prophylaxis in neutropenic patients and for oropharyngeal candidiasis. FDA, U.S. Food and Drug Administration.

Some generalizations about the management of specific *Candida* infections are possible. Recovery of *Candida* from sputum is almost never indicative of underlying pulmonary candidiasis and does not by itself warrant antifungal treatment. Similarly, *Candida* in the urine of a patient with an indwelling bladder catheter may represent colonization only rather than bladder or kidney infection; however, the threshold for systemic treatment is lower in severely ill patients in this category since it is not possible to distinguish colonization from lower or upper urinary tract infection. If the isolate is *C. albicans*, most clinicians use oral fluconazole rather than a bladder washout with amphotericin, which was more commonly used in the past. Caspofungin has been used with success; although they are poorly excreted into the urine, echinocandins may be an option, especially for non-*albicans* isolates. The doses and duration are the same as for disseminated candidiasis. The significance of the recovery of *Candida* from abdominal drains in postoperative patients is also unclear, but again the threshold for treatment is generally low because most of the affected patients have been subjected to factors predisposing to disseminated candidiasis.

Removal of the infected valve and long-term antifungal therapy constitute appropriate treatment for *Candida*endocarditis. Although definitive studies are not available, patients usually are treated for weeks with a systemic antifungal agent (Table 203-2) and then given chronic suppressive therapy for months or years (and sometimes indefinitely) with an oral azole (usually fluconazole at 400–800 mg/d).

Hematogenous *Candida* endophthalmitis is a special problem requiring ophthalmologic consultation. In lesions that are expanding or that threaten the macula, an IV polyene combined with flucytosine (25 mg/kg four times daily) has been the regimen of choice. However, as more data on the azoles and echinocandins become available, new strategies involving these agents are developing. Of paramount importance is the decision to perform a partial vitrectomy. This procedure debulks the infection and can preserve sight, which may otherwise be lost as a result of vitreal scarring. All patients with candidemia should undergo ophthalmologic examination because of the relatively high frequency of this ocular complication. Not only can this examination detect a developing eye lesion early in its course; in addition, identification of a lesion signifies a probability of ~90% of deep-organ abscesses and may prompt prolongation of therapy for candidemia beyond the recommended 2 weeks after the last positive blood culture.

Although the basis for the consensus is a very small data set, the recommended treatment for *Candida*meningitis is a polyene (Table 203-2) plus flucytosine (25 mg/kg four times daily). Successful treatment of *Candida*-infected prosthetic material (e.g., an artificial joint) nearly always requires removal of the infected material followed by long-term administration of an antifungal agent selected on the basis of the isolate's sensitivity and the logistics of administration.

Prophylaxis

The use of antifungal agents to prevent *Candida* infections has been controversial, but some general principles have emerged. Most centers administer prophylactic fluconazole (400 mg/d) to recipients of allogeneic stem cell transplants. High-risk liver transplant recipients are also given fluconazole prophylaxis in most centers. The use of prophylaxis for neutropenic patients has varied considerably from center to center; most centers that elect to give prophylaxis to this population use either fluconazole (200–400 mg/d) or a lipid formulation of amphotericin B (AmBiSome, 1–2 mg/d). Caspofungin (50 mg/d) has also been recommended. Some centers have used itraconazole suspension (200 mg/d). Posaconazole (200 mg three times daily) has also been approved by the FDA for prophylaxis in neutropenic patients.

Prophylaxis is sometimes given to surgical patients at very high risk. The widespread use of prophylaxis for nearly all patients in general surgical or medical intensive care units is not—and should not be—a common practice for three reasons: (1) the incidence of disseminated candidiasis is relatively low, (2) the cost-benefit ratio is suboptimal, and (3) increased resistance with widespread prophylaxis is a valid concern.

Prophylaxis for oropharyngeal or esophageal candidiasis in HIV-infected patients is not recommended unless there are frequent recurrences.

FURTHER READINGS

Anaissie EJ et al (eds): *Clinical Mycology*, 2nd ed. Elsevier Churchill Livingstone, Philadelphia, 2009

Edwards JE Jr: Candidiasis, in *Principles and Practice of Infectious Diseases* , 7th ed, GL Mandell et al (eds). Philadelphia, Elsevier Churchill Livingstone, 2010, pp 3225–3240

Hassan I et al: Excess mortality, length of stay and cost attributable to candidaemia. J Infect 59:360, 2009 [PMID: 19744519]

Kauffman CA: Clinical efficacy of new antifungal agents. Curr Opin Microbiol 9:1, 2006

Magill SS: The epidemiology of *Candida* colonization and invasive candidiasis in a surgical intensive care unit where fluconazole prophylaxis is utilized: Follow-up to a randomized clinical trial. Ann Surg 249:657, 2009 [PMID: 19300221]

Maschmeyer G: The changing epidemiology of invasive fungal infections: New threats. Int J Antimicrob Agents 27:3, 2006 [PMID: 16707250]

Moudgal V, Sobel JL: Antifungals to treat *Candida albicans*. Expert Opin Pharmacother 11:2037, 2010 [PMID: 20536294]

Ostrosky-Zeichner L et al: Multicenter clinical evaluation of the (1→3) beta-D-glucan assay as an aid to diagnosis of fungal infections in humans. Clin Infect Dis 41:654, 2005 [PMID: 16080087]

Pappas PG et al: Clinical practice guidelines for the management of candidiasis: 2009 update by the Infectious Diseases Society of America. Clin Infect Dis 48:503, 2009 [PMID: 19191635]

Ruhnke M: Epidemiology of *Candida albicans* infections and role of non-*Candida-albicans* yeasts. Curr Drug Targets 7:495, 2006 [PMID: 16611037]

Sobel JD et al: Caspofungin in the treatment of symptomatic candiduria. Clin Infect Dis 44:e46, 2007

———: *Candida* urinary tract infections—epidemiology. Clin Infect Dis 52(Suppl 6);S433, 2011 [additional relevant articles appear in this issue]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ASPERGILLOSIS: INTRODUCTION

Aspergillosis is the collective term used to describe all disease entities caused by any one of ~35 pathogenic and allergenic species of Aspergillus. Only those species that grow at 37°C can cause invasive infection, although some species without this capability can cause allergic syndromes. A. fumigatus is responsible for most cases of invasive aspergillosis, almost all cases of chronic aspergillosis, and most allergic syndromes. A. flavus is more prevalent in some hospitals and causes a higher proportion of cases of sinus and cutaneous infections and keratitis than A. fumigatus. A. niger can cause invasive infection but more commonly colonizes the respiratory tract and causes external otitis. A. terreus causes only invasive disease, usually with a poor prognosis. A. nidulans occasionally causes invasive infection, primarily in patients with chronic granulomatous disease.

Epidemiology and Ecology

Aspergillus has a worldwide distribution, most commonly growing in decomposing plant materials (i.e., compost) and in bedding. This hyaline (nonpigmented), septate, branching mold produces vast numbers of conidia (spores) on stalks above the surface of mycelial growth. Aspergilli are found in indoor and outdoor air, on surfaces, and in water from surface reservoirs. Daily exposures vary from a few to many millions of conidia; the latter high numbers of conidia are encountered in hay barns and other very dusty environments. The required size of the infecting inoculum is uncertain; however, only intense exposures (e.g., during construction work, handling of moldy bark or hay, or composting) are sufficient to cause disease in healthy immunocompetent individuals. Allergic syndromes may be exacerbated by continuous antigenic exposure arising from sinus or airway colonization or from nail infection. High-efficiency particulate air (HEPA) filtration is often protective against infection; thus HEPA filters should be installed and monitored for efficiency in operating rooms and in hospital environments that house high-risk patients.

The incubation period of invasive aspergillosis after exposure is highly variable, extending in documented cases from 2 to 90 days. Thus community-acquired acquisition of an infecting strain frequently manifests as invasive infection during hospitalization, although nosocomial acquisition is also common. Outbreaks usually are directly related to a contaminated air source in the hospital.

Risk Factors and Pathogenesis

The primary risk factors for invasive aspergillosis are profound neutropenia and glucocorticoid use; risk increases with longer duration of these conditions. Higher doses of glucocorticoids increase the risk of both acquisition of invasive aspergillosis and death from the infection. Neutrophil and/or phagocyte dysfunction is also an important risk factor, as evidenced by aspergillosis in chronic granulomatous disease, advanced HIV infection, and relapsed leukemia. An increasing incidence of invasive aspergillosis in medical intensive care units suggests that, in patients who are not immunocompromised, temporary abrogation of protective responses as a result of glucocorticoid use or a general anti-inflammatory state is a significant risk factor. Many patients have some evidence of prior pulmonary disease—typically, a history of pneumonia or chronic obstructive pulmonary disease. Glucocorticoid use does not appear to predispose to invasive Aspergillus sinusitis but probably increases the risk of dissemination after pulmonary infection. Anti–tumor necrosis factor therapy also carries an increased risk of infection.

Patients with chronic pulmonary aspergillosis have a wide spectrum of underlying pulmonary disease, often tuberculosis or sarcoidosis. Patients are immunocompetent except for some cytokine regulation defects, most of which are consistent with an inability to mount an inflammatory immune (TH1-like) response. Glucocorticoids accelerate disease progression.

Allergic bronchopulmonary aspergillosis (ABPA) is associated with polymorphisms of interleukin (IL) 4Ra, IL-10, and SPA2 genes (and others) and with heterozygosity of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. These associations suggest a strong genetic basis for the development of a TH2-like and "allergic" response to A. fumigatus.

CD4+CD25+ T (Treg) cells also appear to be pivotal in determining the disease phenotype. Remarkably, high-dose glucocorticoid treatment for exacerbations of ABPA almost never leads to invasive aspergillosis.

Clinical Features and Approach to the Patient (Table 204-1)

Table 204-1 Major Manifestations of Aspergillosis

	Type of Disease			
Organ	**Invasive (Acute and Subacute)**	**Chronic**	**Saprophytic**	**Allergic**
Lung	Angioinvasive in neutropenia, non-angioinvasive, granulomatous	Chronic cavitary, chronic fibrosing	Aspergilloma (single), airway colonization	Allergic bronchopulmonary, severe asthma with fungal sensitization, extrinsic allergic alveolitis
Sinus	Acute invasive	Chronic invasive, chronic granulomatous	Maxillary fungal ball	Allergic fungal sinusitis, eosinophilic fungal rhinosinusitis
Brain	Abscess, hemorrhagic infarction, meningitis	Granulomatous, meningitis	None	None
Skin	Acute disseminated, locally invasive (trauma, burns, IV access)	External otitis, onychomycosis	None	None
Heart	Endocarditis (native or prosthetic), pericarditis	None	None	None
Eye	Keratitis, endophthalmitis	None	None	None described

INVASIVE PULMONARY ASPERGILLOSIS

Both the frequency of invasive disease and the pace of its progression increase with greater degrees of immunocompromise (Fig. 204-1). Invasive aspergillosis is arbitrarily divided into acute and subacute forms that have courses of ≤1 month and 1–3 months, respectively. More than 80% of cases of invasive aspergillosis involve the lungs. The most common clinical features are no symptoms at all, fever, cough (sometimes productive), nondescript chest discomfort, trivial hemoptysis, and shortness of breath. Although the fever often responds to glucocorticoids, the disease progresses. The keys to early diagnosis in at-risk patients are a high index of suspicion, screening for circulating antigen (in leukemia), and urgent CT of the thorax.

Figure 204-1

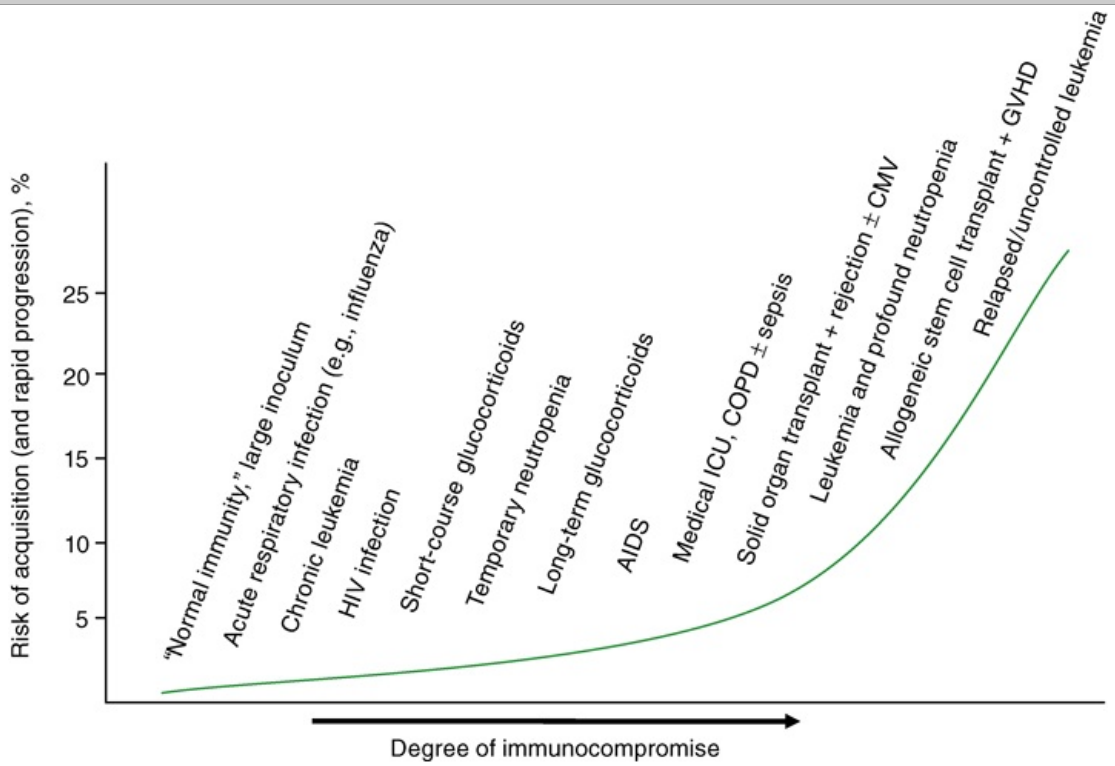

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Invasive aspergillosis: conditions placing patients at elevated risk of acquisition and relatively rapid progression. CMV, cytomegalovirus; COPD, chronic obstructive pulmonary disease; GVHD, graft-

versus-host disease; ICU, intensive care unit.

INVASIVE SINUSITIS

The sinuses are involved in 5–10% of cases of invasive aspergillosis, especially in patients with leukemia and recipients of hematopoietic stem cell transplants. In addition to fever, the most common features are nasal or facial discomfort, blocked nose, and nasal discharge (sometimes bloody). Endoscopic examination of the nose reveals pale, dusky or necrotic-looking tissue in any location. CT or MRI of the sinuses is essential but does not distinguish invasive Aspergillus sinusitis from preexisting allergic or bacterial sinusitis early in the disease process.

TRACHEOBRONCHITIS

Occasionally, only the airways are infected by Aspergillus. The resulting manifestations range from acute or chronic bronchitis to ulcerative or pseudomembranous tracheobronchitis. These entities are particularly common among lung transplant recipients. Obstruction with mucous plugs occurs in normal individuals, persons with ABPA, and immunocompromised patients.

DISSEMINATED ASPERGILLOSIS

In the most severely immunocompromised patients, Aspergillus disseminates from the lungs to multiple organs—most often to the brain but also to the skin, thyroid, bone, kidney, liver, gastrointestinal tract, eye (endophthalmitis), and heart valve. Aside from cutaneous lesions, the most common features are gradual clinical deterioration over 1–3 days, with low-grade fever and features of mild sepsis, and nonspecific abnormalities in laboratory tests. In most cases, at least one localization becomes apparent before death occurs. Blood cultures are almost always negative.

CEREBRAL ASPERGILLOSIS

Hematogenous dissemination to the brain is a devastating complication of invasive aspergillosis. Single or multiple lesions may develop. In acute disease, hemorrhagic infarction is most typical, and cerebral abscess is common. Rarer manifestations include meningitis, mycotic aneurysm, and cerebral granuloma. Local spread from cranial sinuses also occurs. Postoperative infection occurs rarely and is exacerbated by glucocorticoids, which are often given after neurosurgery. The presentation can be either acute or subacute, with mood changes, focal signs, seizures, and decline in mental status. Cerebral granuloma can mimic a primary or secondary tumor. MRI is the most useful immediate investigation; unenhanced CT of the brain is usually nonspecific, and contrast is often contraindicated because of poor renal function.

ENDOCARDITIS

Most cases of Aspergillus endocarditis are prosthetic valve infections resulting from contamination during surgery. Native valve disease is reported, especially as a feature of disseminated infection and in persons using illicit IV drugs. Culture-negative endocarditis with large vegetations is the most common presentation, but embolectomy reveals the diagnosis in a few cases.

CUTANEOUS ASPERGILLOSIS

Dissemination of Aspergillus occasionally results in cutaneous features, usually an erythematous or purplish nontender area that progresses to a necrotic eschar. Direct invasion of the skin occurs in neutropenic patients at the site of IV catheter insertion and in burn patients. Rapidly progressive local aspergillosis of the skin and underlying tissue may follow trauma, and wounds may become infected with Aspergillus (especially A. flavus) after surgery.

Chronic Pulmonary Aspergillosis

The hallmark of chronic cavitary pulmonary aspergillosis (also called semi-invasive aspergillosis, chronic necrotizing aspergillosis, or complex aspergilloma) (Fig. 204-2) is one or more pulmonary cavities expanding over a period of months or years in association with pulmonary symptoms and systemic manifestations such as fatigue and weight loss. (Pulmonary aspergillosis developing over <3 months is better classified as subacute invasive aspergillosis.) Often mistaken initially for tuberculosis, almost all cases occur in patients with prior pulmonary disease (e.g., tuberculosis, atypical mycobacterial infection, sarcoidosis, rheumatoid lung disease, pneumothorax, bullae) or lung surgery. The onset is insidious, and systemic features may be more prominent than pulmonary symptoms. Cavities may have a fluid level or a well-formed fungal ball, but pericavitary infiltrates and multiple cavities—with or without pleural thickening—are typical. IgG antibodies (usually precipitating) to Aspergillus are almost always detectable in blood. Some patients have concurrent infections—even without a fungal ball—with atypical mycobacteria and/or other bacterial pathogens. If untreated, chronic pulmonary aspergillosis typically progresses (sometimes relatively rapidly) to unilateral or upper-lobe fibrosis. This end-stage entity is termed chronic fibrosing pulmonary aspergillosis.

Figure 204-2

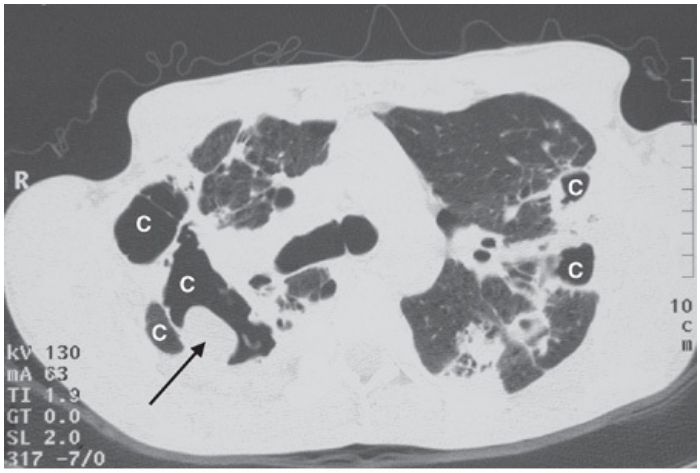

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan image of the chest in a patient with longstanding bilateral chronic cavitary pulmonary aspergillosis. He had a prior history of several bilateral pneumothoraces and required bilateral pleurodesis (1990). CT scan then demonstrated multiple bullae, and sputum cultures grew *A. fumigatus*. The patient had initially weakly and later strongly positive serum Aspergillus antibody tests (precipitins). This scan (2003) shows a mixture of thick- and thin-walled cavities in both lungs (each marked with C), with a probable fungal ball (*black arrow*) protruding into the large cavity on the patient's right side (*R*). There is also considerable pleural thickening bilaterally.

ASPERGILLOMA

Aspergilloma (fungal ball) occurs in up to 20% of residual pulmonary cavities ≥2.5 cm in diameter. Signs and symptoms associated with single (simple) aspergillomas are minor, including a cough (sometimes productive), hemoptysis, wheezing, and mild fatigue. More significant signs and symptoms are associated with chronic cavitary pulmonary aspergillosis and should be treated as such. The vast majority of fungal balls are caused by A. fumigatus, but A. niger has been implicated, particularly in diabetic patients; aspergillomas due to A. niger can lead to oxalosis with renal dysfunction. The most significant complication of aspergilloma is life-threatening hemoptysis, which may be the presenting manifestation. Some fungal balls resolve spontaneously, but the cavity may still be infected.

CHRONIC SINUSITIS

Three entities are subsumed under this broad label: sinus aspergilloma, chronic invasive sinusitis, and chronic granulomatous sinusitis. Sinus aspergilloma is limited to the maxillary sinus and consists of a chronic saprophytic entity in which the sinus cavity is filled with a fungal ball. This form of disease is associated with prior upper-jaw root canal work and chronic (bacterial) sinusitis. About 90% of CT scans show focal hyperattenuation related to concretions; on MRI scans, the T2-weighted signal is decreased, whereas it is increased in bacterial sinusitis. Removal of the fungal ball is curative. No tissue invasion is demonstrable histologically or radiologically.

In contrast, chronic invasive sinusitis is a slowly destructive process that most commonly affects the ethmoid and sphenoid sinuses but can involve any sinus. Patients are usually but not always immunocompromised to some degree (e.g., as a result of diabetes or HIV infection). Imaging of the cranial sinuses shows opacification of one or more sinuses, local bone destruction, and invasion of local structures. The differential diagnosis is wide, as numerous other fungi may cause a similar disease and sphenoid sinusitis is often caused by bacteria. Apart from a history of chronic nasal discharge and blockage, loss of the sense of smell, and persistent headache, the usual presenting features are related to local involvement of critical structures. The orbital apex syndrome (blindness and proptosis) is characteristic. Facial swelling, cavernous sinus thrombosis, carotid artery occlusion, pituitary fossa, and brain and skull base invasion have been described.

Chronic granulomatous sinusitis due to Aspergillus is most commonly seen in the Middle East and India and is often caused by A. flavus. It typically presents late, with facial swelling and unilateral proptosis. The prominent granulomatous reaction histologically distinguishes this disease from chronic invasive sinusitis, in which tissue necrosis with a low-grade mixed-cell infiltrate is typical.

ALLERGIC BRONCHOPULMONARY ASPERGILLOSIS

In almost all cases, ABPA represents a hypersensitivity reaction to A. fumigatus; rare cases are due to other aspergilli and other fungi. ABPA occurs in ~1% of patients with asthma and in up to 15% of adults with cystic fibrosis; occasional cases are reported in patients with neither of the latter. Episodes of bronchial obstruction with mucous plugs leading to coughing fits, "pneumonia," consolidation, and breathlessness are typical. Many patients report coughing up thick sputum casts, usually brown or clear. Eosinophilia commonly develops before systemic glucocorticoids are given. The cardinal

diagnostic tests include an elevated serum level of total IgE (usually >1000 IU/mL), a positive skin-prick test to A. fumigatus extract, or detection of Aspergillus-specific IgE and IgG (precipitating) antibodies. Central bronchiectasis is characteristic, but patients may present before it becomes apparent.

SEVERE ASTHMA WITH FUNGAL SENSITIZATION

Many adults with severe asthma do not fulfill the criteria for ABPA and yet are allergic to fungi. Although A. fumigatus is a common allergen, numerous other fungi (e.g., Cladosporium and Alternaria spp.) are implicated by skin-prick testing and/or specific IgE radioallergosorbent testing.

ALLERGIC SINUSITIS

Like the lungs, the sinuses manifest allergic responses to Aspergillus and other fungi. The affected patients present with chronic (i.e., perennial) sinusitis typically requiring multiple courses of antibiotics that are of only limited benefit. Many of these patients have nasal polyps, and all have congested nasal mucosae and sinuses full of mucoid material. The histologic hallmark of allergic fungal sinusitis is local eosinophilia and Charcot-Leyden crystals (the breakdown products of eosinophils). Removal of abnormal mucus and polyps, with local and occasionally systemic administration of glucocorticoids, usually leads to resolution. Persistent or recurrent signs and symptoms may require more extensive surgery (ethmoidectomy) and possibly local antifungal therapy.

SUPERFICIAL ASPERGILLOSIS

Aspergillus can cause keratitis and otitis externa. The former may be difficult to diagnose early enough to save the patient's sight. Treatment requires local surgical debridement as well as both systemic and topical antifungal therapy. Otitis externa usually resolves with debridement and local application of antifungal agents.

Diagnosis

Several techniques are required to establish the diagnosis of any form of aspergillosis with confidence. Patients with acute invasive aspergillosis have a relatively heavy load of fungus in the affected organ; thus culture, molecular diagnosis, antigen detection, and histopathology usually confirm the diagnosis. However, the pace of progression leaves only a narrow window for making the diagnosis without losing the patient, and some invasive procedures are not possible because of coagulopathy, respiratory compromise, and other factors. Currently, ~40% of cases of invasive aspergillosis are missed clinically and are diagnosed only at autopsy. Histologic examination of affected tissue reveals either infarction, with invasion of blood vessels by many fungal hyphae, or acute necrosis, with limited inflammation and hyphae. Aspergillus hyphae are hyaline, narrow, and septate, with branching at 45°; no yeast forms are present in infected tissue. Hyphae can be seen in cytology or microscopy preparations, which therefore provide a rapid means of presumptive diagnosis.

Culture is important in confirming the diagnosis, given that multiple other (rarer) fungi can mimic Aspergillus spp. histologically. Bacterial agar is less sensitive than fungal media for culture. Thus, if physicians do not request fungal culture, the diagnosis may be missed. Culture may be falsely positive (e.g., in patients whose airways are colonized by Aspergillus) or falsely negative. Only 10–30% of patients with invasive aspergillosis have a positive culture at any time. Molecular diagnostic techniques are faster and much more sensitive than culture of respiratory samples and blood.

The Aspergillus antigen test relies on detection of galactomannan release from Aspergillus spp. during growth. Antigen testing in high-risk patients is best done prospectively, as positive results usually precede clinical or radiologic features by several days. Antigen testing may be falsely positive in patients receiving certain β-lactam/β-lactamase inhibitor antibiotic combinations. Antigen and molecular testing on bronchoalveolar lavage fluid and cerebrospinal fluid are useful if performed before antifungal therapy has been given for more than a few days. The sensitivity of antigen detection is reduced by antifungal prophylaxis.

Definitive confirmation of the diagnosis requires (1) a positive culture of a sample taken directly from an ordinarily sterile site (e.g., a brain abscess) or (2) positive results of both histologic testing and culture of a sample taken from an affected organ (e.g., sinuses or skin). Most diagnoses of invasive aspergillosis are inferred from fewer data, including the presence of the halo sign on a high-resolution thoracic CT scan, in which a localized ground-glass appearance representing hemorrhagic infarction surrounds a nodule. While a halo sign may be produced by other fungi, Aspergillus spp. are by far the most common cause. Halo signs are present for ~7 days early in the course of infection in neutropenic patients and are a good prognostic feature. Thick CT sections can give the false appearance of a halo sign, as can other technical factors. Other common radiologic features of invasive pulmonary aspergillosis include pleural-based infarction or cavitation.

For chronic invasive aspergillosis, Aspergillus antibody testing is invaluable although relatively imprecise. Titers fall with successful therapy. Cultures are infrequently positive. Some patients with chronic pulmonary aspergillosis also have elevated titers of total serum IgE and Aspergillus-specific IgE.

ABPA and severe asthma with fungal sensitization are diagnosed serologically with elevated total and

specific serum IgE levels and with skin-prick tests. Allergic Aspergillus sinusitis is usually diagnosed histologically, although precipitating antibodies in blood may also be useful.

Treatment: Aspergillosis

Antifungal drugs active against Aspergillus include voriconazole, itraconazole, posaconazole, caspofungin, micafungin, and amphotericin B. Initial IV administration is preferred for acute invasive aspergillosis and oral administration for all other disease that requires antifungal therapy. Current recommendations are shown in Table 204-2. Voriconazole is the preferred agent for invasive aspergillosis; caspofungin, posaconazole, and lipid-associated amphotericin B are second-line agents. Amphotericin B is not active against A. terreus or A. nidulans. An infectious disease consultation is advised for patients with invasive disease, given the complexity of management. It is not clear whether combination therapy for acute invasive aspergillosis is beneficial, but it is widely used for very ill patients and for those with a poor prognosis. Commonly used combinations include an azole with either caspofungin or micafungin. The interactions of voriconazole and itraconazole with many drugs must be considered before these agents are prescribed. In addition, the plasma concentrations of both drugs vary substantially from one patient to another, and many authorities recommend monitoring to ensure that drug concentrations are adequate but not excessive. The duration of therapy for invasive aspergillosis varies from ~3 months to several years, depending on the patient's immune status and response to therapy. Relapse occurs if the response is suboptimal and immune reconstitution is not complete.

Table 204-2 Treatment of Aspergillosis

Indication	Primary Treatment	Evidence Level[a]	Precautions	Secondary Treatment	Comments
Invasive[b]	Voriconazole	AI	Drug interactions (especially with rifampin), renal failure (IV only)	AmB, caspofungin, posaconazole, micafungin	As primary therapy, voriconazole carries 20% more responses than AmB. If azole prophylaxis fails, it is unclear whether a class change is required for therapy.
Prophylaxis	Posaconazole, itraconazole solution	AI	Diarrhea and vomiting with itraconazole, vincristine interaction	Micafungin, aerosolized AmB	Some centers monitor plasma levels of itraconazole and posaconazole.
ABPA	Itraconazole	AI	Some glucocorticoid interactions, including with inhaled formulations	Voriconazole, posaconazole	Long-term therapy is helpful in most patients. No evidence indicates whether therapy modifies progression to bronchiectasis/fibrosis.
Single aspergilloma	Surgery	BII	Multicavity disease: poor outcome of surgery; medical therapy preferable	Itraconazole, voriconazole, intracavity AmB	Single large cavities with an aspergilloma are best resected.
Chronic pulmonary[b]	Itraconazole, voriconazole	BII	Poor absorption of capsules with proton pump inhibitors or H_2 blockers	Posaconazole, IV AmB, IV micafungin	Resistance may emerge during treatment, especially if plasma drug levels are subtherapeutic.

[a]Evidence levels are those used in treatment guidelines [Walsh TJ et al: Treatment of aspergillosis: Clinical practice guidelines of the Infectious Diseases Society of America (IDSA). Clin Infect Dis 46:327, 2008].

[b]An infectious disease consultation is appropriate for these patients.

Note: The oral dose is usually 200 mg bid for voriconazole and itraconazole and 400 mg bid for posaconazole. The IV dose of voriconazole is 6 mg/kg twice at 12-h intervals (loading doses) followed by 4 mg/kg q12h. Plasma monitoring is helpful in optimizing the dosage. Caspofungin is given as a single loading dose of 70 mg and then at 50 mg/d; some authorities use 70 mg/d for patients weighing >80 kg, and lower doses are required with hepatic dysfunction. Micafungin is given as 50 mg/d for

prophylaxis and as at least 150 mg/d for treatment; this drug is not yet approved by the U.S. Food and Drug Administration (FDA) for this indication. AmB deoxycholate is given at a daily dose of 1 mg/kg if tolerated. Several strategies are available for minimizing renal dysfunction. Lipid-associated AmB is given at 3 mg/kg (AmBisome) or 5 mg/kg (Abelcet). Different regimens are available for aerosolized AmB, but none is FDA approved. Other considerations that may alter dose selection or route include age; concomitant medications; renal, hepatic, or intestinal dysfunction; and drug tolerability. AmB, amphotericin B.

Itraconazole is the preferred oral agent for chronic and allergic forms of aspergillosis. Voriconazole or posaconazole can be substituted when failure, emergence of resistance, or adverse events occur. An itraconazole dose of 200 mg twice daily is recommended, with monitoring of drug concentrations in the blood. Chronic cavitary pulmonary aspergillosis probably requires life-long therapy, whereas the duration of treatment for other forms of chronic and allergic aspergillosis requires case-by-case evaluation.

Resistance to one or more azoles, although uncommon, may develop during long-term treatment, and a positive culture during antifungal therapy is an indication for susceptibility testing. Glucocorticoids should be used with caution in chronic cavitary pulmonary aspergillosis.

Surgical treatment is important in several forms of aspergillosis, including maxillary fungal ball and single aspergillomas, in which surgery is curative; invasive aspergillosis involving bone, heart valve, sinuses, and proximal areas of the lung; brain abscess; keratitis; and endophthalmitis. In allergic fungal sinusitis, removal of abnormal mucus and polyps, with local and occasionally systemic glucocorticoid treatment, usually leads to resolution. Persistent or recurrent signs and symptoms may require more extensive surgery (ethmoidectomy) and possibly local antifungal therapy. Surgery is problematic in chronic pulmonary aspergillosis, usually resulting in serious complications. Bronchial artery embolization is preferred for problematic hemoptysis.

PROPHYLAXIS

In situations in which moderate or high risk is predicted (e.g., after induction therapy for acute myeloid leukemia), the need for antifungal prophylaxis for superficial and systemic candidiasis and for invasive aspergillosis is generally accepted. Fluconazole is commonly used in these situations but has no activity against Aspergillus spp. Itraconazole capsules are ineffective, and itraconazole solution offers only modest efficacy. Posaconazole solution is more effective. Some data support the use of IV micafungin. No prophylactic regimen is completely successful.

Outcome

Invasive aspergillosis is curable if immune reconstitution occurs, whereas allergic and chronic forms are not. The mortality rate for invasive aspergillosis is ~50% if the infection is treated but is 100% if the diagnosis is missed. Cerebral aspergillosis, Aspergillus endocarditis, and bilateral extensive invasive pulmonary aspergillosis have very poor outcomes, as does invasive infection in persons with late-stage AIDS or relapsed uncontrolled leukemia and in recipients of allogeneic hematopoietic stem cell transplants.

FURTHER READINGS

Agarwal R et al: Aspergillus hypersensitivity and allergic bronchopulmonary aspergillosis in patients with bronchial asthma: Systematic review and meta-analysis. Int J Tuberc Lung Dis 13:936, 2009[PMID: 19723372]

Denning DW et al: Randomized controlled trial of oral antifungal treatment for severe asthma with fungal sensitization (SAFS), the FAST study. Am J Respir Crit Care Med 179:11, 2009[PMID: 18948425]

Guinea J et al: Pulmonary aspergillosis in patients with chronic obstructive pulmonary disease: Incidence, risk factors, and outcome. Clin Microbiol Infect 16:870, 2010[PMID: 19906275]

Herbrecht R et al: Voriconazole versus amphotericin B for primary therapy of invasive aspergillosis. N Engl J Med 347:408, 2002[PMID: 12167683]

Hope WW et al: Laboratory diagnosis of invasive aspergillosis. Lancet Infect Dis 9:609, 2005

——— et al: The invasive and saprophytic syndromes due to Aspergillus spp. Med Mycol 43:S207, 2005

Meersseman W et al: Invasive aspergillosis in the intensive care unit. Clin Infect Dis 45:205, 2007[PMID: 17578780]

O'Gorman CM et al: Discovery of a sexual cycle in the opportunistic fungal pathogen Aspergillus fumigatus. Nature 457:471, 2009[PMID: 19043401]

Verweij PE et al: Azole-resistance in Aspergillus: Proposed nomenclature and breakpoints. Drug Resist Updates 12:141, 2009[PMID: 19879181]

Viscoli C et al: An EORTC phase II study of caspofungin as first-line therapy of invasive aspergillosis in haematological patients. J Antimicrob Chemother 64:1274, 2009[PMID: 19841031]

Walsh TJ et al: Treatment of aspergillosis: Clinical practice guidelines of the Infectious Diseases Society of America (IDSA). Clin Infect Dis 46:327, 2008[PMID: 18177225]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MUCORMYCOSIS: INTRODUCTION

Mucormycosis represents a group of life-threatening infections caused by fungi of the order Mucorales. Recent reclassification has abolished the class Zygomycetes and placed the order Mucorales in the subphylum Mucoromycotina. Therefore, infection caused by the Mucorales is most accurately referred to as mucormycosis, although the term *zygomycosis* may still be used by some sources. Mucormycosis is highly invasive and relentlessly progressive, resulting in higher rates of morbidity and mortality (>40%) than many other infections. A high index of suspicion is critical for diagnosis, and early initiation of therapy—often before confirmation of the diagnosis—is necessary to optimize outcomes.

Etiology

Fungi of the order Mucorales belong to six families, all of which can cause mucormycosis. Among the Mucorales, *Rhizopus oryzae* (in the family Mucoraceae) is by far the most common cause of infection. Less frequently isolated species of the Mucoraceae family that cause a similar spectrum of infections include *Rhizopus microsporus*, *Rhizomucor pusillus*, *Mycocladus corymbifer* (formerly *Absidia corymbifera*), *Apophysomyces elegans*, and *Mucor* species (which, despite its name, is a rare cause of mucormycosis). Increasing numbers of cases of mucormycosis due to infection with *Cunninghamella* species (family Cunninghamellaceae) have also been reported. Rare case reports have demonstrated the ability of fungi in the remaining families of the Mucorales to cause mucormycosis.

Pathogenesis

The Mucorales are ubiquitous environmental fungi to which humans are constantly exposed. These fungi cause infection primarily in patients with diabetes or defects in phagocytic function (e.g., associated with neutropenia or glucocorticoid treatment). Patients with elevated levels of free iron, which supports fungal growth in serum and tissues, are likewise at increased risk for mucormycosis. In iron-overloaded patients with end-stage renal failure, treatment with deferoxamine predisposes to the development of rapidly fatal disseminated mucormycosis; this agent, an iron chelator for the human host, serves as a fungal siderophore, directly delivering iron to the Mucorales. Furthermore, patients with diabetic ketoacidosis (DKA) are at high risk of developing rhinocerebral mucormycosis. The acidosis causes dissociation of iron from sequestering proteins in serum, resulting in enhanced fungal survival and virulence. It is likely that hyperglycemia during DKA also contributes to the risk of mucormycosis through its association with poorly characterized defects in phagocytic function.

Epidemiology

Mucormycosis typically occurs in patients with diabetes mellitus, solid organ or hematopoietic stem cell transplantation (HSCT), prolonged neutropenia, or malignancy. In patients undergoing HSCT, mucormycosis develops at least as commonly during nonneutropenic as during neutropenic periods, probably because of glucocorticoid treatment of graft-versus-host disease. Mucormycosis can occur as isolated cutaneous or subcutaneous infection in immunologically normal individuals after traumatic implantation of soil or vegetation, after maceration of the skin by a moist surface, or in nosocomial settings via direct access through intravenous catheters or subcutaneous injections.

Patients receiving antifungal prophylaxis with either itraconazole or voriconazole may be at increased risk of mucormycosis. These patients typically present with disseminated mucormycosis, the most lethal form of disease. Breakthrough mucormycosis has been described repeatedly in patients receiving posaconazole or echinocandin prophylaxis.

Clinical Manifestations

Mucormycosis can be divided into at least six clinical categories based on clinical presentation and the involvement of a particular anatomic site: rhinocerebral, pulmonary, cutaneous, gastrointestinal, disseminated, and miscellaneous. These categories of invasive mucormycosis tend to affect patients with specific defects in host defense. For example, patients with DKA typically develop the rhinocerebral form and much more rarely develop pulmonary or disseminated disease. In contrast, pulmonary mucormycosis occurs most commonly in leukemic patients who are receiving chemotherapy and in patients undergoing HSCT.

Rhinocerebral mucormycosis continues to be the most common form of the disease. Most cases occur in patients with diabetes, although such cases (probably due to glucocorticoid use) are increasingly

being described in the transplantation setting. The initial symptoms of rhinocerebral mucormycosis are nonspecific and include eye or facial pain and facial numbness followed by the onset of conjunctival suffusion, blurry vision, and soft tissue swelling. Fever may be absent in up to half of cases, while white blood cell counts are typically elevated as long as the patient has functioning bone marrow. If untreated, infection usually spreads from the ethmoid sinus to the orbit, resulting in compromise of extraocular muscle function and proptosis, typically with chemosis. Onset of signs and symptoms in the contralateral eye, with resulting bilateral proptosis, chemosis, vision loss, and ophthalmoplegia, is ominous and suggests the development of cavernous sinus thrombosis.

Upon visual inspection, infected tissue may appear to be normal during the earliest stages of fungal spread and then progresses through an erythematous phase, with or without edema, before the onset of a violaceous appearance and finally the development of a black necrotic eschar. Infection can sometimes extend from the sinuses into the mouth and produce painful necrotic ulcerations of the hard palate, but this is a late finding that suggests extensive, well-established infection.

Pulmonary mucormycosis is the second most common manifestation. Symptoms include dyspnea, cough, and chest pain; fever is often but not invariably present. Angioinvasion results in necrosis, cavitation, and/or hemoptysis. Lobar consolidation, isolated masses, nodular disease, cavities, or wedge-shaped infarcts may be seen on chest radiography. High-resolution chest CT is the best method for determining the extent of pulmonary mucormycosis and may demonstrate evidence of infection before it is seen on the chest x-ray. In the setting of cancer, where mucormycosis may be difficult to differentiate from aspergillosis, the presence of ≥10 pulmonary nodules, pleural effusion, or concomitant sinusitis makes mucormycosis more likely. It is critical to distinguish mucormycosis from aspergillosis as rapidly as possible, as treatments for these infections differ. Indeed, voriconazole—the first-line treatment for aspergillosis—exacerbates mucormycosis in mouse and fly models.

Cutaneous mucormycosis may result from external implantation of the fungus or conversely from hematogenous dissemination. External implantation–related infection has been described in the setting of soil exposure from trauma (e.g., in a motor vehicle accident), penetrating injury with plant material (e.g., a thorn), injections of medications (e.g., insulin), catheter insertion, contaminated surgical dressings, and use of tape to secure endotracheal tubes. Cutaneous disease can be highly invasive, penetrating into muscle, fascia, and even bone. In mucormycosis, necrotizing fasciitis carries a mortality rate approaching 80%. Necrotic cutaneous lesions in the setting of hematogenous dissemination are also associated with an extremely high mortality rate. However, with prompt, aggressive surgical debridement, isolated cutaneous mucormycosis has a favorable prognosis and a low mortality rate.

Gastrointestinal mucormycosis has occurred in premature neonates in association with disseminated disease and necrotizing enterocolitis; more rarely, it has been described in adults with neutropenia or other immunocompromising conditions. In addition, gastrointestinal disease has been reported as a nosocomial process following administration of medications mixed with contaminated wooden applicator sticks. Nonspecific abdominal pain and distention associated with nausea and vomiting are the most common symptoms. Gastrointestinal bleeding is common, and fungating masses may be seen in the stomach at endoscopy. The disease may progress to visceral perforation, with extremely high mortality rates.

Hematogenously disseminated mucormycosis may originate from any primary site of infection. The most common site of dissemination is the brain, but metastatic lesions may also be found in any other organ. The mortality rate associated with dissemination to the brain approaches 100%. Even without central nervous system (CNS) involvement, mortality rates for disseminated mucormycosis exceed 90%. Miscellaneous forms of mucormycosis may affect any body site, including bones, mediastinum, trachea, kidneys, and (in association with dialysis) peritoneum.

Diagnosis

A high index of suspicion is required for diagnosis of mucormycosis. Unfortunately, autopsy series have shown that up to half of cases are diagnosed only post-mortem. Because the Mucorales are environmental isolates, definitive diagnosis requires a positive culture from a sterile site (e.g., a needle aspirate, a tissue biopsy specimen, or pleural fluid) or histopathologic evidence of invasive mucormycosis. A probable diagnosis of mucormycosis can be established by culture from a nonsterile site (e.g., sputum or bronchoalveolar lavage) when a patient has appropriate risk factors as well as clinical and radiographic evidence of disease. However, given the urgency of administering therapy early, the patient should be treated while confirmation of the diagnosis is awaited.

Biopsy with histopathologic examination remains the most sensitive and specific modality for definitive diagnosis (**Fig. 205-1**). Biopsy reveals characteristic wide (≥6- to 30-μm), thick-walled, ribbon-like, aseptate hyphal elements that branch at right angles. Other fungi, including *Aspergillus*, *Fusarium*, and *Scedosporium* species, have septae, are thinner, and branch at acute angles. Because artificial septae may result from folding of tissue during processing (which may also alter the appearance of the angle of branching), the width and the ribbon-like form of the fungus are the most reliable features distinguishing mucormycosis. The Mucorales are visualized most effectively with periodic acid–Schiff or methenamine silver stain or, if the organism burden is high, with hematoxylin and eosin. While

histopathology can identify the Mucorales, species can be identified only by culture. PCR is being investigated as a diagnostic tool for mucormycosis but is not yet approved by the U.S. Food and Drug Administration (FDA) for this purpose and is not generally available.

Figure 205-1

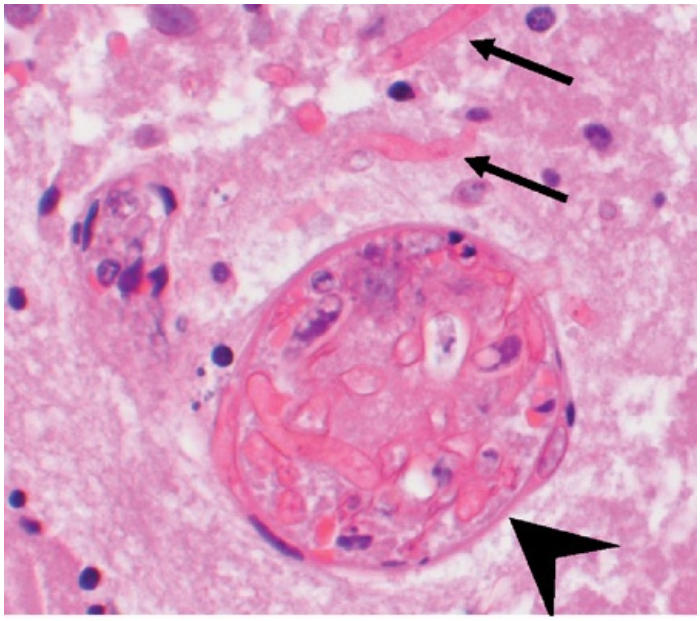

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

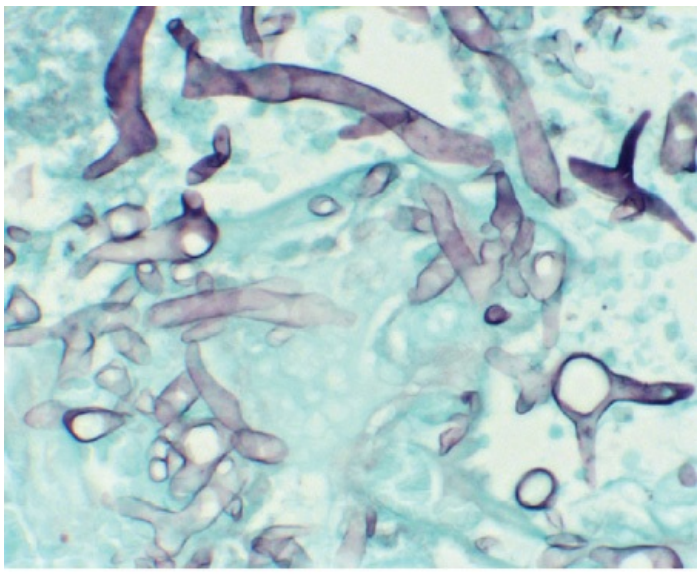

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histopathology sections of *Rhizopus oryzae* in infected brain. **A.** Broad, ribbon-like, nonseptate hyphae in the parenchyma (*arrows*) and a thrombosed blood vessel with extensive intravascular hyphae (*arrowhead*) (hematoxylin and eosin). ***B***. Extensive, broad, ribbon-like hyphae invading the parenchyma (Gomori methenamine silver).

Unfortunately, cultures are positive in fewer than half of cases of mucormycosis. Nevertheless, the Mucorales are not fastidious organisms and tend to grow quickly (i.e., within 48 h) on culture media. The likely explanation for the low sensitivity of culture is that the Mucorales form long filamentous structures that are killed by tissue homogenization—the standard method for preparing tissue cultures in the clinical microbiology laboratory. Thus the laboratory should be advised when a diagnosis of mucormycosis is suspected, and the tissue should be cut into sections and placed in the center of culture dishes rather than homogenized.

Imaging techniques often yield subtle findings that underestimate the extent of disease. For example, the most common finding on CT or MRI of the head or sinuses of a patient with rhino-orbital mucormycosis is sinusitis that is indistinguishable from bacterial sinusitis. It is also common to detect no abnormalities in sinus bones despite clinical evidence of progressive disease. MRI is more sensitive (~80%) for detecting orbital and CNS disease than is CT. High-risk patients should always undergo endoscopy and/or surgical exploration, with biopsy of the areas of suspected infection. If mucormycosis is suspected, initial empirical therapy with a polyene antifungal agent should be initiated while the diagnosis is being confirmed.

Differential Diagnosis

Other mold infections, including aspergillosis, scedosporiosis, fusariosis, and infections caused by the dematiaceous fungi (brown-pigmented soil organisms), can cause clinical syndromes identical to mucormycosis. Histopathologic examination usually allows distinction of the Mucorales from these other organisms, and a positive culture permits definitive species identification. It is important to distinguish the Mucorales from these other fungi, as the preferred antifungal treatments differ (i.e., polyenes for the Mucorales vs. expanded-spectrum triazoles for most septate molds). The entomophthoromycoses caused by *Basidiobolus* and *Conidiobolus* (fungi formerly grouped with the Mucorales in the class Zygomycetes) can also cause identical clinical syndromes. These fungi may appear similar to the Mucorales on histopathology and can be reliably distinguished from the latter only by culture.

In a patient with sinusitis and proptosis, orbital cellulitis and cavernous sinus thrombosis caused by bacterial pathogens (most commonly *Staphylococcus aureus*, but also streptococcal and gram-negative species) must be excluded. *Klebsiella rhinoscleromatis* is a rare cause of an indolent facial rhinoscleroma syndrome that may appear similar to mucormycosis. Finally, the Tolosa-Hunt syndrome causes painful ophthalmoplegia, ptosis, headache, and cavernous sinus inflammation; biopsies and clinical follow-up may be needed to distinguish the Tolosa-Hunt syndrome from mucormycosis by the lack of progression of the former entity.

Treatment: Mucormycosis

GENERAL PRINCIPLES

The successful treatment of mucormycosis requires four steps: (1) early diagnosis; (2) reversal of underlying predisposing risk factors, if possible; (3) surgical debridement; and (4) prompt antifungal therapy. Early diagnosis of mucormycosis is critical, since early initiation of therapy is associated with improved outcomes. It is also crucial to reverse (or prevent) underlying defects in host defense during treatment (e.g., by stopping or reducing the dosage of immunosuppressive medications or by rapidly restoring euglycemia and normal acid-base status). Finally, it is advisable to avoid administration of iron to patients with active mucormycosis, as iron exacerbates infection in animal models.

Blood vessel thrombosis and resulting tissue necrosis during mucormycosis can result in poor penetration of antifungal agents to the site of infection. Therefore, debridement of necrotic tissues may be critical for complete eradication of disease. Surgery has been found (by logistic regression and in multiple case series) to be an independent variable for favorable outcome in patients with mucormycosis. The extent and timing of surgical debridement necessary to optimize outcomes of mucormycosis have not been defined. Limited data from a retrospective study support the use of intraoperative frozen sections to delineate the margins of infected tissues, with sparing of tissues lacking evidence of infection. A multidisciplinary team, including an internist, an infectious disease specialist, and surgical specialists relevant to the sites of infection, is typically required for the management of mucormycosis.

ANTIFUNGAL THERAPY

Primary antifungal therapy for mucormycosis should be based on a polyene antibiotic (**Table 205-1**), except perhaps for mild localized infection (e.g., isolated suprafascial cutaneous infection) in immunocompetent patients, which has been eradicated surgically. Amphotericin B (AmB) deoxycholate remains the only licensed antifungal agent for the treatment of mucormycosis. However, lipid formulations of AmB are significantly less nephrotoxic, can be administered at higher doses, and may be more efficacious than AmB deoxycholate for this purpose.

Table 205-1 First-Line Antifungal Options for the Treatment of Mucormycosis[a]

Drug	Recommended Dosage	Advantages and Supporting Studies	Disadvantages
Primary Antifungal Therapy			
AmB deoxycholate	• 1.0–1.5 mg/kg qd	• >5 decades of clinical experience • Inexpensive • Only licensed agent for treatment of	• Highly toxic • Poor CNS penetration

		mucormycosis	
LAmB	• 5–10 mg/kg qd	• Less nephrotoxic than AmB deoxycholate • Better CNS penetration than AmB deoxycholate or ABLC • Better outcomes than with AmB deoxycholate in murine models and a retrospective clinical review	• Expensive
ABLC	• 5–7.5 mg/kg qd	• Less nephrotoxic than AmB deoxycholate • Murine and retrospective clinical data suggest benefit of combination therapy with echinocandins.	• Expensive • Possibly less efficacious than LAmB for CNS infection
Primary Combination Therapy[b]			
Caspofungin plus Lipid polyene	• 70-mg IV loading dose, then 50 mg/d for ≥2 weeks • 50 mg/m^2 IV in children	• Favorable toxicity profile • Synergistic in murine disseminated mucormycosis • Retrospective clinical data suggest superior outcomes for rhino-orbital-cerebral mucormycosis.	• Very limited clinical data on combination therapy
Micafungin or anidulafungin plus Lipid polyene	• 100 mg/d for ≥2 weeks • Micafungin: 4 mg/kg qd in children • Micafungin: 10 mg/kg qd in low-birth-weight infants • Anidulafungin: 1.5 mg/kg qd in children	• Favorable toxicity profile • Synergistic with LAmB in murine model of disseminated mucormycosis	• No clinical data
Deferasirox *plus* Lipid polyene	• 20 mg/kg PO qd for 2–4 weeks	• Highly fungicidal against Mucorales in vitro • Synergistic with LAmB in murine model of disseminated mucormycosis	• Only available for enteral administration • No clinical data; phase II clinical trial completed

[a]Primary therapy should generally include a polyene. Non-polyene-based regimens may be appropriate for patients who refuse polyene therapy or for relatively immunocompetent patients with mild disease (e.g., isolated suprafascial cutaneous infection) that can be surgically eradicated.

[b]Prospective randomized trials are necessary to confirm the suggested benefit (from animal and small retrospective human studies) of combination therapy for mucormycosis. Dose escalation of any echinocandin is not recommended because of a paradoxical loss of benefit of combination therapy at echinocandin doses of ≥3 mg/kg qd.

Abbreviations: ABLC, AmB lipid complex; AmB, amphotericin B; CNS, central nervous system; LAmB, liposomal AmB.

Source: Reprinted from Spellberg et al, 2009.

The optimal dosages for antifungal treatment of mucormycosis are not known. Starting dosages of 1 mg/kg per day for AmB deoxycholate and 5–7.5 mg/kg per day for liposomal AmB (LAmB) and amphotericin B lipid complex (ABLC) are commonly given to adults and children. Whether higher dosages provide any additional benefit is unclear. However, dose escalation of LAmB to 10 mg/kg per day for CNS mucormycosis may be considered in light of the limited penetration of polyenes into the brain. ABLC dose escalation above 7.5 mg/kg per day is not advisable given the lack of relevant data.

Echinocandin–lipid polyene combinations improved survival rates among mice with disseminated mucormycosis (including CNS disease) and were associated with significantly better outcomes than polyene monotherapy in a small retrospective study involving primarily diabetic patients with rhino-orbital-cerebral mucormycosis. Although combination therapy may be considered on the basis of these limited data sets, definitive clinical trials are needed to establish whether it offers any real advantage over monotherapy for mucormycosis.

In contrast to deferoxamine, the iron chelator deferasirox is fungicidal for clinical isolates of the

Mucorales. In mice with DKA and disseminated mucormycosis, combination deferasirox-LAmB therapy resulted in synergistic improvement of survival rates. Enrollment in a double-blind, randomized, placebo-controlled, phase II safety/exploratory efficacy study of adjunctive deferasirox therapy (20 mg/kg per day for 14 days) [the Deferasirox-AmBisome Therapy for Mucormycosis (DEFEAT Mucor) Study, NCT00419770] has recently been completed; this study is likely to elucidate the potential risks and benefits of iron chelation therapy for mucormycosis.

Posaconazole is the only FDA-approved azole with in vitro activity against the Mucorales. However, pharmacokinetic/pharmacodynamic data raise concerns about the reliability of achieving adequate in vivo levels of orally administered posaconazole. Furthermore, posaconazole has been found to be inferior in efficacy to AmB for the treatment of murine mucormycosis and is not superior to placebo for treatment of murine infection with *R. oryzae*. Moreover, posaconazole-polyene combination therapy is not superior to polyene monotherapy for murine mucormycosis.

The roles of recombinant cytokines and neutrophil transfusions in the primary treatment of mucormycosis are not clear. Limited data indicate that hyperbaric oxygen may be useful in centers with the appropriate technical expertise and facilities.

In general, antifungal therapy for mucormycosis should be continued until all of the following objectives are attained: (1) resolution of clinical signs and symptoms of infection; (2) resolution or stabilization of residual radiographic signs of disease on serial imaging; and (3) resolution of underlying immunosuppression. For patients with mucormycosis who are receiving immunosuppressive medications, secondary antifungal prophylaxis is typically continued for as long as the immunosuppressive regimen is administered.

FURTHER READINGS

Chamilos G et al: Delaying amphotericin B–based frontline therapy significantly increases mortality among patients with hematologic malignancy who have zygomycosis. Clin Infect Dis 47:503, 2008[PMID: 18611163]

Ibrahim AS et al: The iron chelator deferasirox protects mice from mucormycosis through iron starvation. J Clin Invest 117:2649, 2007[PMID: 17786247]

Kontoyiannis DP et al: Zygomycosis in a tertiary-care cancer center in the era of *Aspergillus*-active antifungal therapy: A case-control observational study of 27 recent cases. J Infect Dis 191:1350, 2005[PMID: 15776383]

Marr KA et al: Epidemiology and outcome of mould infections in hematopoietic stem cell transplant recipients. Clin Infect Dis 34:909, 2002[PMID: 11880955]

Reed C et al: Combination polyene-caspofungin treatment of rhino-orbital-cerebral mucormycosis. Clin Infect Dis 47:364, 2008[PMID: 18558882]

Roden MM et al: Epidemiology and outcome of zygomycosis: A review of 929 reported cases. Clin Infect Dis 41:634, 2005[PMID: 16080086]

Spellberg B et al: Recent advances in the management of mucormycosis: From bench to bedside. Clin Infect Dis 48:1743, 2009[PMID: 19435437]

Spellberg B, Ibrahim AS: Recent advances in the treatment of mucormycosis. Curr Infect Dis Rep 12:423, 2010[PMID: 21308550]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ENDEMIC MYCOSES (DIMORPHIC FUNGI)

Dimorphic fungi exist in discrete environmental niches as molds that produce conidia, which are their infectious form. In tissues and at temperatures of >35°C, the mold converts to the yeast form. Other endemic mycoses—histoplasmosis, coccidioidomycosis, and blastomycosis—are discussed in Chaps. 199, 200, and 201, respectively.

Sporotrichosis

ETIOLOGIC AGENT

Sporothrix schenckii is a thermally dimorphic fungus that is found worldwide in sphagnum moss, decaying vegetation, and soil.

EPIDEMIOLOGY AND PATHOGENESIS

Sporotrichosis most commonly infects persons who participate in outdoor activities such as landscaping, gardening, and tree farming. Infected animals, especially cats, can transmit *S. schenckii* to humans. Sporotrichosis is primarily a localized infection of skin and subcutaneous tissues that follows traumatic inoculation of conidia. Osteoarticular sporotrichosis is uncommon, occurring most often in middle-aged men who abuse alcohol, and pulmonary sporotrichosis occurs almost exclusively in persons with chronic obstructive pulmonary disease who have inhaled the organism from the environment. Dissemination occurs rarely, almost always in markedly immunocompromised patients, especially those with AIDS.

CLINICAL MANIFESTATIONS

Days or weeks after inoculation, a papule develops at the site and then usually ulcerates but is not very painful. Similar lesions develop sequentially along the lymphatic channels proximal to the original lesion. Some patients develop a fixed cutaneous lesion that can be verrucous or ulcerative and that remains localized without lymphatic extension. The differential diagnosis of lymphocutaneous sporotrichosis includes nocardiosis, tularemia, nontuberculous mycobacterial infection (especially that due to *Mycobacterium marinum*), and leishmaniasis. Osteoarticular sporotrichosis can present as chronic synovitis or septic arthritis. Pulmonary sporotrichosis must be differentiated from tuberculosis or other fungal pneumonias. Numerous ulcerated skin lesions, with or without spread to visceral organs [including the central nervous system (CNS)], are characteristic of disseminated sporotrichosis.

DIAGNOSIS

S. schenckii usually grows readily as a mold when material from a cutaneous lesion is incubated at room temperature. Histopathologic examination of biopsy material shows a mixed granulomatous and pyogenic reaction, and tiny oval or cigar-shaped yeasts are sometimes visualized with special stains. Serologic testing is not useful.

TREATMENT AND PROGNOSIS

Guidelines for the management of the various forms of sporotrichosis have been published by the Infectious Diseases Society of America (**Table 206-1**). Itraconazole is the drug of choice for lymphocutaneous sporotrichosis. Fluconazole is less effective; voriconazole and posaconazole have not been used for sporotrichosis. Saturated solution of potassium iodide (SSKI) is also effective for lymphocutaneous infection and costs much less than itraconazole. However, SSKI is poorly tolerated because of adverse reactions, including metallic taste, salivary gland swelling, rash, and fever. Terbinafine appears to be effective but has been used in few patients. Treatment for lymphocutaneous sporotrichosis is continued for 2–4 weeks after all lesions have resolved, usually for a total of 3–6 months. Pulmonary and osteoarticular forms of sporotrichosis are treated with itraconazole for at least 1 year. Severe pulmonary infection and disseminated sporotrichosis, including that involving the CNS, are treated initially with amphotericin B (AmB), which is followed by itraconazole after improvement has been noted. Lifelong suppressive therapy with itraconazole is required for AIDS patients. The success rate for treatment of lymphocutaneous sporotrichosis is 90–100%, but other forms of the disease respond poorly to antifungal therapy.

Table 206-1 Suggested Treatment for Endemic Mycoses

Disease	First-Line Therapy	Alternatives/Comments

Sporotrichosis		
Cutaneous, lymphocutaneous	Itraconazole, 200 mg/d until 2–4 weeks after lesions resolve	SSKI, increasing doses[a] Terbinafine, 500 mg bid
Pulmonary, osteoarticular	Itraconazole, 200 mg bid for 12 months	Lipid AmB[b] for severe pulmonary disease until stable; then itraconazole
Disseminated, central nervous system	Lipid AmB[b] for 4–6 weeks	Itraconazole, 200 mg bid after AmB for 12 months Itraconazole maintenance for AIDS patients: 200 mg/d until CD4+ T cell count is >200/μL for 12 months
Paracoccidioidomycosis		
Chronic (adult form)	Itraconazole, 100–200 mg/d for 6–12 months	TMP-SMX, 160/800 mg bid for 12–36 months
Acute (juvenile form)	AmB[c] until improvement	Itraconazole, 200 mg bid after AmB for 12 months
Penicilliosis		
Mild or moderate	Itraconazole, 200 mg bid for 12 weeks	Itraconazole maintenance for AIDS patients: 200 mg/d until CD4+ T cell count is >100/μL for 6 months
Severe	AmB[c] until improvement	Itraconazole, 200 mg bid after AmB for 12 weeks Itraconazole maintenance: as for mild or moderate disease

[a]The starting dosage is 5–10 drops tid in water or juice. The dosage is increased weekly by 10 drops per dose, as tolerated, up to 40–50 drops tid.

[b]The dosage of lipid AmB is 3–5 mg/kg daily; the higher dosage should be used when the central nervous system is involved.

[c]The dosage of AmB deoxycholate is 0.6–1.0 mg/kg daily.

Abbreviations: AmB, amphotericin B; SSKI, saturated solution of potassium iodide; TMP-SMX, trimethoprim-sulfamethoxazole.

Paracoccidioidomycosis

ETIOLOGIC AGENT

Paracoccidioides brasiliensis is a thermally dimorphic fungus that is endemic in humid areas of Central and South America, especially in Brazil.

EPIDEMIOLOGY AND PATHOGENESIS

A striking male-to-female ratio varies from 14:1 to as high as 70:1 (in rural Brazil). Most patients are middle-aged or elderly men from rural areas. Paracoccidioidomycosis develops after the inhalation of aerosolized conidia encountered in the environment. For most patients, disease rarely develops at the time of the initial infection but appears years later, presumably after reactivation of a latent infection.

CLINICAL MANIFESTATIONS

Two major syndromes are associated with paracoccidioidomycosis: the acute or juvenile form and the chronic or adult form. The acute form is uncommon, occurs mostly in persons <30 years old, and manifests as disseminated infection of the reticuloendothelial system. Immunocompromised individuals can also manifest this type of rapidly progressive disease. The chronic form of paracoccidioidomycosis accounts for ~90% of cases and predominantly affects older men. The primary manifestation is progressive pulmonary disease, primarily in the lower lobes, with fibrosis. Ulcerative and nodular mucocutaneous lesions in the nares and mouth—another common manifestation of chronic paracoccidioidomycosis—must be differentiated from leishmaniasis (Chap. 212) and squamous cell carcinoma (Chap. 87).

DIAGNOSIS

The diagnosis is established by growth of the organism in culture. A presumptive diagnosis can be made by detection of the distinctive thick-walled yeast, with multiple narrow-necked buds attached circumferentially, in purulent material or tissue biopsies.

TREATMENT AND PROGNOSIS

Itraconazole is the treatment of choice for paracoccidioidomycosis (Table 206-1). Ketoconazole is also effective but more toxic; voriconazole and posaconazole have been used with success in a few cases. Sulfonamides are also effective and are the least costly agents, but the response is slower and the relapse rate higher. Seriously ill patients should be treated with AmB initially. Patients with paracoccidioidomycosis have an excellent response to therapy, but pulmonary fibrosis is often progressive in those with chronic disease.

Penicilliosis

ETIOLOGIC AGENT

Penicillium marneffei is a thermally dimorphic fungus that is endemic in the soil in certain areas of Vietnam, Thailand, and several other southeastern Asian countries.

EPIDEMIOLOGY AND PATHOGENESIS

The epidemiology of penicilliosis is linked to bamboo rats, which are infected with the fungus but rarely manifest disease. The disease occurs most often among persons living in rural areas in which the rats are found, but there is no evidence for transmission of the infection directly from rats to humans. Infection is rare in immunocompetent hosts, and most cases are reported in persons who have advanced AIDS. Infection results from the inhalation of conidia from the environment. The organism converts to the yeast phase in the lungs and then spreads hematogenously to the reticuloendothelial system.

CLINICAL MANIFESTATIONS

The clinical manifestations of penicilliosis mimic those of disseminated histoplasmosis and include fever, fatigue, weight loss, dyspnea, diarrhea (in some cases), lymphadenopathy, hepatosplenomegaly, and skin lesions, which appear as papules that often umbilicate and resemble molluscum contagiosum (Chap. 183).

DIAGNOSIS

Penicilliosis is diagnosed by culture of *P. marneffei* from blood or from biopsy samples of skin, bone marrow, or lymph node. The organism usually grows within 1 week as a mold that produces a distinctive red pigment. Histopathologic examination of tissues and smears of blood or material from skin lesions shows oval or elliptical yeast-like organisms with central septation and can quickly establish a presumptive diagnosis.

TREATMENT AND PROGNOSIS

Patients who have severe disease should be treated initially with AmB until their condition improves; therapy can then be changed to itraconazole (Table 206-1). Patients who have mild symptoms can be treated from the start with itraconazole. For patients with AIDS, suppressive therapy with itraconazole is recommended until immune reconstitution (related to successful therapy for HIV infection with antiretroviral drugs) is evident. Disseminated penicilliosis is usually fatal if not treated. With treatment, the mortality rate is ~10%.

PHAEOHYPHOMYCOSES

In these common soil organisms (also called *dematiaceous* fungi), melanin causes the hyphae and/or conidia to be darkly pigmented (brown/black). The term *phaeohyphomycosis* is used to describe any infection with a pigmented mold. This definition encompasses two specific syndromes—eumycetoma and chromoblastomycosis—as well as all other types of infections caused by these organisms. It is important to note that eumycetomas can be caused by hyaline molds as well as brown-black molds and that only about half of all mycetomas are due to fungi. Actinomycetes cause the remainder (Chap. 162). Most of these fungi cause localized subcutaneous infections after direct inoculation, but disseminated infection and serious visceral focal infections also occur, especially in immunocompromised patients.

ETIOLOGIC AGENTS

A large number of pigmented molds can cause human infection. All are found in the soil or on plants, and some cause economically important plant diseases. The most common cause of eumycetoma is *Madurella* species, but hyaline molds such as *Scedosporium* species also cause this syndrome. *Fonsecaea* and *Cladophialophora* species are responsible for most cases of chromoblastomycosis. Disseminated infection and focal visceral infections are caused by a variety of dematiaceous fungi; *Alternaria*, *Exophiala*, *Curvularia*, and *Wangiella* species are among the more common molds reported to cause human infection.

EPIDEMIOLOGY AND PATHOGENESIS

Most infections, including all cases of eumycetoma and chromoblastomycosis, are acquired by inoculation through the skin. These two syndromes are seen almost entirely in tropical and subtropical areas and occur mostly in rural laborers who are frequently exposed to the organisms. Inhalation into the upper or lower respiratory tract leads to localized infection of sinuses and, in immunocompromised

patients, to pneumonia and sometimes hematogenous dissemination. Several organisms, specifically *Cladophialophora bantiana* and *Rhinocladiella mackensiei*, are neurotropic and likely to cause CNS infection.

CLINICAL MANIFESTATIONS

Eumycetoma is a chronic subcutaneous and cutaneous infection that usually occurs on the lower extremities and that is characterized by swelling, development of sinus tracts, and the appearance of grains that are actually colonies of fungi discharged from the sinus tract. As the infection progresses, adjacent fascia and bony structures become involved. The disease is indolent and disfiguring, progressing slowly over years. Complications include fractures of infected bone and bacterial superinfections.

Chromoblastomycosis is an indolent subcutaneous infection characterized by nodular, verrucous, or plaque-like painless lesions that occur predominantly on the lower extremities and grow slowly over months or years. There is hardly ever extension to adjacent structures, as is seen with eumycetoma. Long-term consequences include bacterial superinfection, chronic lymphedema, and (rarely) the development of squamous cell carcinoma.

Dematiaceous molds are the most common cause of allergic fungal sinusitis and a less common cause of invasive fungal sinusitis. Keratitis occurs with corneal inoculation. Many patients, including some who are immunocompromised, develop only localized infection manifested by cyst-like subcutaneous lesions at the site of inoculation. Other immunocompromised patients have pneumonia and disseminated infection, including CNS involvement, and are quite ill.

DIAGNOSIS

The specific diagnosis of infection with a pigmented mold is established by growth of the organism in culture. However, in eumycetoma, a tentative clinical diagnosis can be made when a patient presents with a lesion characterized by swelling, sinus tracts, and grains. Histopathologic examination and culture are necessary to ensure that the etiologic agent is a mold and not an actinomycete. In chromoblastomycosis, the diagnosis rests on the histologic demonstration of sclerotic bodies in the tissues; culture merely establishes which pigmented mold is causing the infection. Sclerotic bodies are dark brown, thick-walled, septate fungal forms that resemble large yeasts and define the syndrome. For disseminated infections and nonsubcutaneous focal infections, growth of the organism is essential to differentiate infection with a hyaline mold (e.g., *Aspergillus* or *Fusarium*) from that due to a pigmented mold. No serologic or non-culture-based methods are currently available to aid in the diagnosis of phaeohyphomycoses.

TREATMENT AND PROGNOSIS

Treatment of eumycetoma and chromoblastomycosis involves both surgical extirpation of the lesion and use of antifungal agents. Surgical removal of the lesions of both eumycetoma and chromoblastomycosis is most effective if performed before extensive spread has occurred. In chromoblastomycosis, cryosurgery and laser therapy have been used with variable success. The antifungal agents of choice are itraconazole, voriconazole, and posaconazole. The most experience has accrued with itraconazole; there is less experience with the newer azoles, which are active in vitro and have been reported to be effective in a few patients. Flucytosine and terbinafine also have been used to treat chromoblastomycosis.

Disseminated and focal visceral infections are treated with the appropriate antifungal agent; the choice of agent is based on the location and extent of the infection, in vitro testing, and clinical experience with the specific infecting organism. AmB is not effective against many of these organisms but has been used successfully against others. Again, the most experience has been obtained with itraconazole, which is effective for localized infections. Voriconazole and posaconazole are likely to play an increasing role in therapy for both localized and disseminated infections. Chromoblastomycosis and eumycetoma are chronic indolent infections that are difficult to cure without surgical excision, but these are not life-threatening diseases. Disseminated or focal visceral infections with pigmented molds in immunocompromised patients are associated with high mortality rates unless immunosuppression can be diminished and effective antifungal agents prescribed promptly.

OPPORTUNISTIC FUNGAL INFECTIONS

Many environmental fungi can cause infection in markedly immunocompromised hosts, but two genera of hyaline (nonpigmented) molds, *Fusarium* and *Scedosporium*, and one yeast-like genus, *Trichosporon*, have become particularly prominent pathogens among such patients. *Fusarium* and *Scedosporium* species overlap with invasive aspergillosis in their clinical manifestations and, when seen in tissues, appear similar to *Aspergillus*. In the immunocompetent host, these fungi cause localized infections of skin, skin structures, and subcutaneous tissues, but their role in disseminated infection will be emphasized in this section.

Fusariosis

ETIOLOGIC AGENT

Fusarium species, which are found worldwide in soil and on plants, have emerged as major opportunists in markedly immunocompromised patients.

EPIDEMIOLOGY AND PATHOGENESIS

Most human infections follow inhalation of conidia, but ingestion and direct inoculation can also lead to disease. An outbreak of severe *Fusarium* keratitis among soft contact lens wearers was traced back to a particular brand of contact lens solution and individual contact lens cases that had been contaminated. Disseminated infection is reported most often in patients who have a hematologicmalignancy, are neutropenic, have received a stem cell or solid organ transplant, or have a severe burn.

CLINICAL MANIFESTATIONS

In immunocompetent persons, *Fusarium* species cause localized infections of various organs. These organisms commonly cause fungal keratitis, which can extend into the anterior chamber of the eye, cause loss of vision, and require corneal transplantation. Onychomycosis due to *Fusarium* species, while basically an annoyance in immunocompetent patients, is a source of subsequent hematogenous dissemination and should be aggressively sought and treated in neutropenic patients. In profoundly immunocompromised patients, fusariosis is angioinvasive, and clinical manifestations mimic those of aspergillosis. Pulmonary infection is characterized by multiple nodular lesions. Sinus infection is likely to lead to invasion of adjacent structures. Disseminated fusariosis occurs primarily in neutropenic patients with hematologicmalignancies and in allogeneic stem cell transplant recipients, especially those with graft-versus-host disease. Disseminated fusariosis differs from disseminated aspergillosis in that skin lesions are extremely common with fusariosis; the lesions are nodular or necrotic, are usually painful, and appear over time in different locations.

DIAGNOSIS

The diagnostic approach usually includes both documentation of the growth of *Fusarium* species from involved tissue and demonstration of invasion by histopathologic or microbiologic techniques that show septate hyphae in tissues or aspirates. The organism is difficult to differentiate from *Aspergillus* species in tissues; thus, identification with culture is imperative. An extremely helpful diagnostic clue is growth in blood cultures, which are positive in as many as 50% of patients with disseminated fusariosis. No serologic or non-culture-based techniques are currently available to aid in diagnosis.

TREATMENT AND PROGNOSIS

Fusarium species are resistant to many antifungal agents. A lipid formulation of AmB (at least 5 mg/kg daily), voriconazole (200–400 mg twice daily), or posaconazole (400 mg twice daily) is recommended. Many physicians use both a lipid formulation of AmB and either voriconazole or posaconazole because susceptibility information is not available when therapy must be initiated. Serum drug levels should be monitored with either azole to ensure that absorption is adequate and with voriconazole to avoid toxicity. Mortality rates for disseminated fusariosis have been as high as 85%. With the antifungal therapy noted above, mortality rates have fallen to ~50%. However, if neutropenia persists, the mortality rate approaches 100%.

Scedosporiosis

ETIOLOGIC AGENT

The genus *Scedosporium* includes several pathogens. The major causes of human infections are *Scedosporium apiospermum*, which in its sexual state is termed *Pseudallescheria boydii*, and *S. prolificans*. The *S. apiospermum* complex encompasses several species but will be referred to here simply as *S. apiospermum*.

EPIDEMIOLOGY AND PATHOGENESIS

S. apiospermum is found worldwide in temperate climates in tidal flats, swamps, ponds, manure, and soil. This organism is known as a cause of pneumonia, disseminated infection, and brain abscess in near-drowning victims. *S. prolificans* is also found in soil but is more geographically restricted; most cases are reported from Spain and Australia. Infection occurs predominantly through inhalation of conidia, but direct inoculation through the skin or into the eye also can occur.

CLINICAL MANIFESTATIONS

Among immunocompetent persons, *Scedosporium* species are a prominent cause of eumycetoma. Keratitis as a result of accidental corneal inoculation is a sight-threatening infection. In patients who have hematologicmalignancies (especially acute leukemia with neutropenia), recipients of solid organ or stem cell transplants, and patients receiving glucocorticoids, *Scedosporium* species are angioinvasive, causing pneumonia and widespread dissemination with abscesses. Pulmonary infection mimics that caused by aspergillosis; nodules, cavities, and lobar infiltrates are common. Disseminated infection involves the skin, heart, brain, and many other organs. Skin lesions are not as common or as painful as those of fusariosis.

DIAGNOSIS

Diagnosis depends on the growth of *Scedosporium* species from involved tissue and the demonstration of invasion by histopathologic or microbiologic techniques that show septate hyphae in tissues or aspirates. Culture evidence is essential because *Scedosporium* species are difficult to differentiate from *Aspergillus* in tissues. Demonstration of tissue invasion is essential because these ubiquitous environmental molds can be mere contaminants or colonizers. *S. prolificans* can grow in blood cultures, but *S. apiospermum* usually does not. No serologic or non-culture-based diagnostic methods are available to aid in diagnosis.

TREATMENT AND PROGNOSIS

Scedosporium species are resistant to AmB, echinocandins, and some azoles. Voriconazole is the agent of choice for *S. apiospermum*, and posaconazole has also been used for this infection. *S. prolificans* is resistant in vitro to almost every available antifungal agent; the addition of agents such as terbinafine to a voriconazole regimen has been attempted because in vitro data suggest possible synergy against some strains of *S. prolificans*. Mortality rates have been as high as 65–75% for *S. apiospermum* and 85–100% for *S. prolificans*. Mortality rates for *S. apiospermum* infection have decreased to 40–50% with the use of voriconazole, but those for disseminated *S. prolificans* infection remain extremely high.

TRICHOSPORONOSIS

Etiologic Agent

The genus *Trichosporon* contains many species, some of which cause localized infection of hair and nails. The major pathogen responsible for invasive infection is *Trichosporonasahii*. *Trichosporon* species grow as yeast-like colonies in vitro; in vivo, however, hyphae, pseudohyphae, and arthroconidia can also be seen.

EPIDEMIOLOGY AND PATHOGENESIS

These yeasts are commonly found in soil, sewage, and water and in rare instances can colonize human skin and the human gastrointestinal tract. Most infections follow fungal inhalation or entry via central venous catheters. Systemic infection occurs almost exclusively in immunocompromised hosts, including those who have hematologicmalignancies, are neutropenic, have received a solid organ transplant, or are receiving glucocorticoids.

CLINICAL MANIFESTATIONS

Disseminated trichosporonosis resembles invasive candidiasis, and fungemia is often the initial manifestation of infection. Pneumonia, skin lesions, and sepsis syndrome are common. The skin lesions begin as papules or nodules surrounded by erythema and progress to central necrosis. A chronic form of infection mimics hepatosplenic candidiasis (chronic disseminated candidiasis).

DIAGNOSIS

The diagnosis of systemic *Trichosporon* infection is established by growth of the organism from involved tissues or from blood. Histopathologic examination of a skin lesion showing a mixture of yeast forms, arthroconidia, and hyphae can lead to an early presumptive diagnosis of trichosporonosis. The serum cryptococcal antigen latex agglutination test may be positive in patients with disseminated trichosporonosis because *T. asahii* and *Cryptococcus neoformans* share polysaccharide antigens.

TREATMENT AND PROGNOSIS

Rates of response to AmB have been disappointing, and many *Trichosporon* isolates are resistant in vitro. Voriconazole appears to be the antifungal agent of choice and is used at a dosage of 200–400 mg twice daily. The mortality rates for disseminated *Trichosporon* infection have been as high as 70% but should decrease with the use of newer azoles, such as voriconazole; however, patients who remain neutropenic are likely to succumb to this infection.

SUPERFICIAL CUTANEOUS INFECTIONS

Fungal infections of the skin and skin structures are caused by molds and yeasts that do not invade deeper tissues but rather cause disease merely by inhabiting the superficial layers of skin, hair follicles, and nails. These agents are the most common cause of fungal diseases of humans but only rarely cause serious infections.

Yeast Infections

ETIOLOGIC AGENTS

The lipophilic yeast *Malassezia* is actually dimorphic in that it lives on the skin in the yeast phase but transforms to the mold phase as it causes disease. Of the seven species, six are grouped together in some classification systems as the *M. furfur* complex; all six of these species require exogenous lipids for growth. The seventh species, *M. pachydermatis*, is also lipophilic but does not have an absolute requirement for lipids.

EPIDEMIOLOGY AND PATHOGENESIS

Malassezia species are part of the indigenous human flora found in the stratum corneum of the back, chest, scalp, and face—areas rich in sebaceous glands. Disease is more common in moist humid areas of the world. The organisms do not invade below the stratum corneum and generally elicit little if any inflammatory response.

CLINICAL MANIFESTATIONS

Malassezia species cause tinea versicolor (also called *pityriasis versicolor*), folliculitis, and seborrheic dermatitis. Tinea versicolor presents as flat round scaly patches of hypo- or hyperpigmented skin on the neck, chest, or upper arms. The lesions are usually asymptomatic but can be pruritic. They can be mistaken for vitiligo, but the latter is not scaly. Folliculitis occurs over the back and chest and mimics bacterial folliculitis. Seborrheic dermatitis manifests as erythematous pruritic scaly lesions in the eyebrows, moustache, nasolabial folds, and scalp. The scalp lesions are termed cradle cap in babies and dandruff in adults. Seborrheic dermatitis can be severe in patients with advanced AIDS. Fungemia and disseminated infection occur rarely with *Malassezia* species—almost always in premature neonates receiving parenteral lipid preparations through a central venous catheter.

DIAGNOSIS

Malassezia infections are diagnosed clinically in most cases. If scrapings are collected on a microscope slide on which a drop of potassium hydroxide has been placed, a mixture of budding yeasts and short septate hyphae are seen. In order to culture *M. furfur* from those patients in whom disseminated infection is suspected, sterile olive oil must be added to the medium.

TREATMENT AND PROGNOSIS

Topical creams and lotions, including selenium sulfide shampoo, ketoconazole shampoo or cream, terbinafine cream, and ciclopirox cream, are effective in treating *Malassezia* infections and are usually given for 2 weeks. Mild topical steroid creams are sometimes used to treat seborrheic dermatitis. For extensive disease, itraconazole (200 mg/d) or fluconazole (200 mg/d) can be used for 5–7 days. The rare cases of fungemia caused by *Malassezia* species are treated with AmB or fluconazole, prompt removal of the catheter, and discontinuance of parenteral lipid infusions. *Malassezia* skin infections are benign and self-limited, although recurrences are the rule. The outcome of systemic infection depends on the host's underlying conditions, but most infants do well.

Dermatophyte (Mold) Infections

ETIOLOGIC AGENTS

The molds that cause skin infections in humans include the genera *Trichophyton*, *Microsporum*, and *Epidermophyton*. These organisms, which are not components of the normal skin flora, can live within the keratinized structures of the skin—hence the term *dermatophytes*.

EPIDEMIOLOGY AND PATHOGENESIS

Dermatophytes occur worldwide, and infections with these organisms are extremely common. Some organisms cause disease only in humans and can be transmitted by person-to-person contact and by fomites, such as hairbrushes or wet floors that have been contaminated by infected individuals. Several species cause infections in cats and dogs and can readily be transmitted from these animals to humans. Finally, some dermatophytes are spread from contact with soil. The characteristic ring shape of cutaneous lesions is the result of the organisms' outward growth in a centrifugal pattern in the stratum corneum. Fungal invasion of the nails usually occurs through the lateral or superficial nail plates and then spreads throughout the nails; when hair shafts are invaded, the organisms can be found either within the shaft or surrounding it. Symptoms are caused by the inflammatory reaction elicited by fungal antigens and not by tissue invasion. Dermatophyte infections occur more commonly in male than in female patients, and progesterone has been shown to inhibit dermatophyte growth.

CLINICAL MANIFESTATIONS

Dermatophyte infection of the skin is often called *ringworm*. This term is confusing because worms are not involved. *Tinea*, the Latin word for *worm*, describes the serpentine nature of the skin lesions and is a less confusing designation that is used in conjunction with the name of the body part affected—e.g., tinea capitis (head), tinea pedis (feet), tinea corporis (body), tinea cruris (crotch), and tinea unguium (nails, although infection at this site is more often termed *onychomycosis*).

Tinea capitis occurs most commonly in children 3–7 years old. Children with tinea capitis usually present with well-demarcated scaly patches in which hair shafts are broken off right above the skin; alopecia can result. Tinea corporis is manifested by well-demarcated, annular, pruritic, scaly lesions that undergo central clearing. Usually one or several small lesions are present. In some cases, tinea corporis can involve much of the trunk or manifest as folliculitis with pustule formation. The rash should be differentiated from contact dermatitis, eczema, and psoriasis. Tinea cruris is seen almost exclusively in men. The perineal rash is erythematous and pustular, has a discrete scaly border, is without satellite lesions, and is usually pruritic. The rash must be differentiated from intertriginous candidiasis,

erythrasma, and psoriasis.

Tinea pedis also is more common among men than among women. It usually starts in the web spaces of the toes; peeling, maceration, and pruritus are followed by development of a scaly pruritic rash along the lateral and plantar surfaces of the feet. Hyperkeratosis of the soles of the feet often ensues. Tinea pedis has been implicated in lower-extremity cellulitis, as streptococci and staphylococci can gain entrance to the tissues through fissures between the toes. Onychomycosis affects toenails more often than fingernails and is most common among persons who have tinea pedis. The nail becomes thickened and discolored and may crumble; onycholysis almost always occurs. Onychomycosis is more common in older adults and in persons with vascular disease, diabetes mellitus, and trauma to the nails. Fungal infection must be differentiated from psoriasis, which can mimic onychomycosis but usually has associated skin lesions.

DIAGNOSIS

Many dermatophyte infections are diagnosed by their clinical appearance. If the diagnosis is in doubt, as is often the case in children with tinea capitis, scrapings should be taken from the edge of a lesion with a scalpel blade, transferred to a slide to which a drop of potassium hydroxide is added, and examined under a microscope for the presence of hyphae. Cultures are indicated if an outbreak is suspected or the patient does not respond to therapy. Culture of the nail is especially useful as an aid to decisions about both diagnosis and treatment.

TREATMENT AND PROGNOSIS

Dermatophyte infections usually respond to topical therapy. Lotions or sprays are easier than creams to apply to large or hairy areas. Particularly for tinea cruris, the affected area should be kept as dry as possible. When patients have extensive skin lesions, oral itraconazole or terbinafine can hasten resolution (**Table 206-2**). Terbinafine interacts with fewer drugs than itraconazole and is generally the first-line agent. Onychomycosis does not respond to topical therapy, although ciclopirox nail lacquer applied daily for a year is occasionally beneficial. Itraconazole and terbinafine both accumulate in the nail plate and can be used to treat onychomycosis (Table 206-2). These agents are more effective and better tolerated than griseofulvin and ketoconazole. The major decision to be made with regard to therapy is whether the extent of nail involvement justifies the use of systemic antifungal agents that have adverse effects, may interact with other drugs, and are costly. Treating for cosmetic reasons alone is discouraged. Relapses of tinea cruris and tinea pedis are common and should be treated early with topical creams to avoid development of more extensive disease. Relapses of onychomycosis follow treatment in 25–30% of cases.

Table 206-2 Suggested Treatment for Extensive Tinea Infections and Onychomycosis

Antifungal Agent	Suggested Dosage	Comments
Extensive tinea skin infection		
Terbinafine	250 mg/d for 1–2 weeks	Adverse reactions minimal with short treatment period
Itraconazole[a]	200 mg/d for 1–2 weeks	Adverse reactions minimal with short treatment period except for drug interactions
Onychomycosis		
Terbinafine	250 mg/d for 3 months	Slightly superior to itraconazole; monitor for hepatotoxicity
Itraconazole[a]	200 mg/d for 3 months or 200 mg bid for 1 week each month for 3 months	Drug interactions frequent; monitor for hepatotoxicity; rarely causes hypokalemia, hypertension, edema; use with caution in patients with congestive heart failure

[a]Itraconazole capsules require food and gastric acid for absorption, whereas itraconazole solution is taken on an empty stomach.

FURTHER READINGS

Ben-Ami R et al: Phaeohyphomycosis in a tertiary care cancer center. Clin Infect Dis 48:1033, 2009[PMID: 19267655]

Chang DC et al: Multistate outbreak of *Fusarium* keratitis associated with use of a contact lens solution. JAMA 296:953, 2006[PMID: 16926355]

Cortez KJ et al: Infections caused by *Scedosporium* species. Clin Microbiol Rev 21:157, 2008[PMID: 18202441]

de Berker D: Clinical practice. Fungal nail disease. N Engl J Med 360:2108, 2009

Garnica M et al: Difficult mycoses of the skin: Advances in the epidemiology and management of eumycetoma, phaeohyphomycosis, and chromoblastomycosis. Curr Opin Infect Dis 22:559, 2009[PMID: 19773651]

Kauffman CA et al: Clinical practice guidelines for the management of sporotrichosis: 2007 Update by the Infectious Diseases Society of America. Clin Infect Dis 45:1255, 2007[PMID: 17968818]

Nucci M et al: *Fusarium* infections in hematopoietic stem cell transplant recipients. Clin Infect Dis 38:1237, 2004[PMID: 15127334]

Restrepo A et al: Pulmonary paracoccidioidomycosis. Semin Respir Crit Care Med 29:182, 2008[PMID: 18366000]

Ruan SY et al: Invasive trichosporonosis caused by *Trichosporon asahii* and other unusual *Trichosporon* species at a medical center in Taiwan. Clin Infect Dis 49:1457, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PNEUMOCYSTIS INFECTIONS: INTRODUCTION

Definition and Description

Pneumocystis is an opportunistic fungal pulmonary pathogen that is an important cause of pneumonia in the immunocompromised host. Although organisms within the *Pneumocystis* genus are morphologically very similar, they are genetically diverse and host-specific. *P. jirovecii* infects humans, whereas *P. carinii*—the original species described in 1909—infects rats. For clarity, only the genus designation *Pneumocystis* will be used in this chapter.

Developmental stages of the organism include the trophic form, the cyst, and the precyst (an intermediate stage). The life cycle of *Pneumocystis* probably involves sexual and asexual reproduction, although definitive proof awaits the development of a reliable culture system. *Pneumocystis* contains several different antigen groups, the most prominent of which are the 95- to 140-kDa major surface glycoprotein (MSG) and kexin (KEX1).

Epidemiology

Serologic surveys have demonstrated that *Pneumocystis* has a worldwide distribution and that most healthy children have been exposed to the organism by 3–4 years of age. Airborne transmission of *Pneumocystis* has been documented in animal studies; person-to-person transmission has been suggested by hospital outbreaks of *Pneumocystis* pneumonia (PcP) and by molecular epidemiologic analysis of isolates. Data suggest that the cyst is the transmissible form.

Pathogenesis and Pathology

Studies over the past several years have shown that *Pneumocystis* commonly colonizes patients who are immunosuppressed or who have chronic obstructive pulmonary disease. This colonization elicits an inflammatory response and is associated with a decline in lung function.

The host factors that predispose to the development of PcP include defects in cellular and humoral immunity. The risk of PcP among HIV-infected patients rises markedly when circulating CD4+ T cell counts fall below 200/μL. Other persons at risk for PcP are patients receiving immunosuppressive agents (particularly glucocorticoids) for cancer and organ transplantation; those receiving biologic agents such as infliximab and etanercept for rheumatoid arthritis and inflammatory bowel disease; children with primary immunodeficiency diseases; and premature malnourished infants.

The principal host effector cells against *Pneumocystis* are alveolar macrophages, which ingest and kill the organism, releasing a variety of inflammatory mediators. Proliferating organisms remain extracellular within the alveolus, attaching tightly to type I cells. Alveolar damage results in increased alveolar-capillary permeability and surfactant abnormalities, including a fall in phospholipids and an increase in surfactant proteins A and D. The host inflammatory response to lung injury leads to increases in levels of interleukin 8 and in neutrophil counts in bronchoalveolar lavage (BAL) fluid. These changes correlate with disease severity.

On lung sections stained with hematoxylin and eosin, the alveoli are filled with a typical foamy, vacuolated exudate. Severe disease may include interstitial edema, fibrosis, and hyaline membrane formation. The host inflammatory changes usually consist of hypertrophy of alveolar type II cells, a typical reparative response, and a mild mononuclear cell interstitial infiltrate. Malnourished infants display an intense plasma cell infiltrate that gave the disease its early name: interstitial plasma cell pneumonia.

Clinical Features

Patients with PcP develop dyspnea, fever, and nonproductive cough. HIV-infected patients are usually ill for several weeks and may have relatively subtle manifestations. Symptoms in non-HIV-infected patients are of shorter duration and often begin after the glucocorticoid dose has been tapered. A high index of suspicion and a thorough history are key factors in early detection.

Physical findings include tachypnea, tachycardia, and cyanosis, but lung auscultation reveals few abnormalities. Reduced arterial oxygen pressure (Pa_{O2}), increased alveolar-arterial oxygen gradient ($PA_{O2} – Pa_{O2}$), and respiratory alkalosis are evident. Diffusion capacity is reduced, and heightened uptake with nonspecific nuclear imaging techniques (gallium scan) may be noted. Elevated serum concentrations of lactate dehydrogenase, reflecting lung parenchymal damage, and β-d-glucan, a

component of the fungal cell wall, have been reported; however, these elevations are not specific for PcP infection.

The classic findings on chest radiography consist of bilateral diffuse infiltrates beginning in the perihilar regions **(Fig. 207-1*A*)**, but various atypical manifestations (nodular densities, cavitary lesions) have also been reported. Pneumothorax occurs, and its management is often difficult. Early in the course of PcP, the chest radiograph may be normal, although high-resolution CT of the lung may reveal ground-glass opacities at this stage (Fig. 207-1*B*).

Figure 207-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

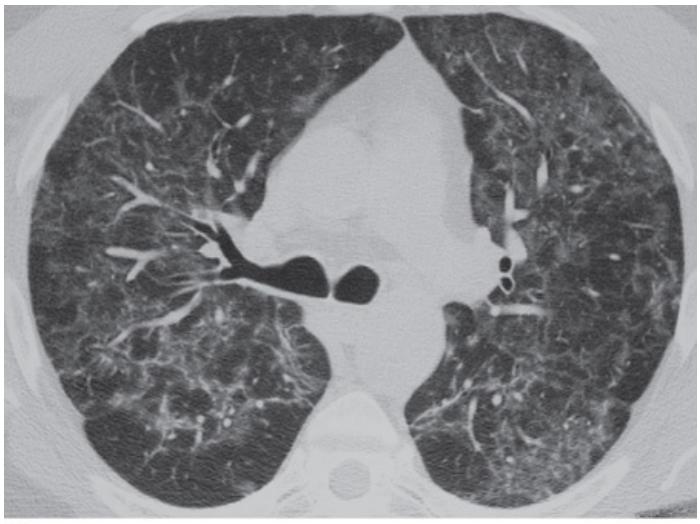

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Chest radiograph depicting diffuse infiltrates in an HIV-infected patient with PcP. **B.** High-resolution CT of the lung showing ground-glass opacification in an HIV-infected patient with PcP. (*Courtesy of Dr. Cristopher Meyer, with permission.*)

While *Pneumocystis* usually remains confined to the lungs, cases of disseminated infection have occurred in both HIV-infected and non-HIV-infected patients. Common sites of involvement include the lymph nodes, spleen, liver, and bone marrow.

Diagnosis

Because of the nonspecific nature of the clinical picture, the diagnosis must be based on specific identification of the organism. A definitive diagnosis is made by histopathologic staining. Traditional cell wall stains such as methenamine silver selectively stain the wall of *Pneumocystis* cysts, while

reagents such as Wright-Giemsa stain the nuclei of all developmental stages. Immunofluorescence with monoclonal antibodies is more sensitive and specific than histologic staining. DNA amplification by PCR may become part of routine diagnostics but may not distinguish colonization from infection.

The successful diagnosis of PcP depends on the collection of proper specimens. In general, the yield from different diagnostic procedures is higher for HIV-infected patients than for non-HIV-infected patients because of the higher organism burden in the former group. Sputum induction and oral washes have gained popularity as simple, noninvasive techniques; however, these procedures require trained and dedicated personnel. Fiberoptic bronchoscopy with BAL, which provides information about the organism burden, the host inflammatory response, and the presence of other opportunistic infections, continues to be the mainstay of *Pneumocystis* diagnosis. Transbronchial biopsy and open lung biopsy, the most invasive procedures, are used only when a diagnosis cannot be made by BAL.

Course and Prognosis

In the typical case of untreated PcP, progressive respiratory embarrassment leads to death. Therapy is most effective when instituted early, before there is extensive alveolar damage. If examination of induced sputum is nondiagnostic and BAL cannot be performed in a timely manner, empirical therapy for PcP is reasonable. However, this practice does not eliminate the need for a specific etiologic diagnosis. With improved management of HIV and its complications, mortality rates from PcP among HIV-infected patients are 0–15%. In contrast, rates of early death remain high among patients who require mechanical ventilation (60%) and among non-HIV-infected patients (40%).

Treatment: Pneumocystis Infections

Trimethoprim-sulfamethoxazole (TMP-SMX), which acts by inhibiting folic acid synthesis, is considered the drug of choice for all forms of PcP **(Table 207-1)**. Therapy is continued for 14 days in non-HIV-infected patients and for 21 days in persons infected with HIV. Since HIV-infected patients respond more slowly than non-HIV-infected patients, it is prudent to wait at least 7 days after the initiation of treatment before concluding that therapy has failed. TMP-SMX is well tolerated by non-HIV-infected patients, whereas more than half of HIV-infected patients experience serious adverse reactions.

Table 207-1 Treatment of Pneumocystosis

Drug(s), Dose, Route	Adverse Effects
First Choice[a] TMP-SMX (TMP: 5 mg/kg; SMX: 25 mg/kg[b]) q6–8 h PO or IV	Fever, rash, cytopenias, hepatitis, hyperkalemia, GI disturbances
Other Agents[a]	
TMP, 5 mg/kg q6–8h; plus dapsone, 100 mg qd PO	Hemolysis (G6PD deficiency), methemoglobinemia, fever, rash, GI disturbances
Atovaquone, 750 mg bid PO	Rash, fever, GI and hepatic disturbances
Clindamycin, 300–450 mg q6h PO or 600 mg q6–8h IV; plus primaquine, 15–30 mg qd PO	Hemolysis (G6PD deficiency), methemoglobinemia, rash, colitis, neutropenia
Pentamidine, 3–4 mg/kg qd IV	Hypotension, azotemia, cardiac arrhythmias, pancreatitis, dysglycemias, hypocalcemia, neutropenia, hepatitis
Trimetrexate, 45 mg/m^2 qd IV; plus leucovorin,[c] 20 mg/kg q6h PO or IV **Adjunctive Agent**	Cytopenias, peripheral neuropathy, hepatic disturbances
Prednisone, 40 mg bid x 5 d, 40 mg qd x 5 d, 20 mg qd x 11 d; PO or IV	Immunosuppression, peptic ulcer, hyperglycemia, mood changes, hypertension

[a]Therapy is administered for 14 days to non-HIV-infected patients and for 21 days to HIV-infected patients.

[b]Equivalent of 2 double-strength (DS) tablets. (One DS tablet contains 160 mg of TMP and 800 mg of SMX.)

[c]Leucovorin prevents bone marrow toxicity from trimetrexate.

Abbreviations: GI, gastrointestinal; G6PD, glucose-6-phosphate dehydrogenase; TMP-SMX, trimethoprim-sulfamethoxazole.

Several alternative regimens are available for the treatment of mild to moderate cases of PcP (a Pa_{O2}

of >70 mmHg or a $PA_{O2} – Pa_{O2}$ of <35 mmHg while breathing room air). TMP plus dapsone and clindamycin plus primaquine are about as effective as TMP-SMX. Dapsone and primaquine should not be administered to patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency. Atovaquone is less effective than TMP-SMX but is better tolerated. Since *Pneumocystis* lacks ergosterol, it is not susceptible to antifungal agents that inhibit ergosterol synthesis.

Alternative regimens that are recommended for the treatment of moderate to severe PcP (a Pa_{O2} of ≤ 70 mmHg or a $PA_{O2} – Pa_{O2}$ of ≥35 mmHg) are parenteral pentamidine, parenteral clindamycin plus primaquine, or trimetrexate plus leucovorin. Parenteral clindamycin plus primaquine may be more efficacious than pentamidine.

Molecular evidence of resistance to sulfonamides and to atovaquone has emerged among clinical *Pneumocystis* isolates. Although prior sulfonamide exposure is a risk factor, this resistance has also occurred in HIV-infected patients who have never received sulfonamides. The outcome of therapy appears to be linked more strongly to traditional measures—e.g., high Acute Physiology, Age, and Chronic Health Evaluation III (APACHE III) scores, need for positive-pressure ventilation, delayed intubation, and development of pneumothorax—than to the presence of molecular markers of sulfonamide resistance.

Early institution of antiretroviral therapy when HIV patients present with PcP has been associated with improved survival rates, but careful attention should be devoted to the possible development of the immune reconstitution inflammatory syndrome. HIV-infected patients frequently experience deterioration of respiratory function shortly after receiving anti-*Pneumocystis* drugs. The adjunctive administration of tapering doses of glucocorticoids to HIV-infected patients with moderate to severe PcP can prevent this problem and improve the rate of survival (Table 207-1). For maximal benefit, this adjunctive therapy should be started early in the course of the illness. The use of steroids as adjunctive therapy in HIV-infected patients with mild PcP or in non-HIV-infected patients remains to be evaluated.

Prevention

Prophylaxis is indicated for HIV-infected patients with CD4+ T cell counts of <200/μL or a history of oropharyngeal candidiasis and for both HIV-infected and non-HIV-infected patients who have recovered from PcP. Prophylaxis may be discontinued in HIV-infected patients once CD4+ T cell counts have risen to >200/μL and remained at that level for ≥3 months. Primary prophylaxis guidelines for immunocompromised hosts not infected with HIV are less clear.

TMP-SMX is the drug of choice for primary and secondary prophylaxis **(Table 207-2)**. This agent also provides protection against toxoplasmosis and some bacterial infections. Alternative regimens are available for individuals intolerant of TMP-SMX (Table 207-2). Although there are no specific recommendations for preventing the spread of *Pneumocystis* in health care facilities, it seems prudent to prevent direct contact between patients with PcP and other susceptible hosts.

Table 207-2 Prophylaxis of Pneumocystosis[a]

Drug(s), Dose, Route	Comments
First Choice	
TMP-SMX, 1 DS tablet or 1 SS tablet qd PO[b]	TMP-SMX can be safely reintroduced for treatment of some patients who have had mild to moderate side effects.
Other Agents	
Dapsone, 50 mg bid or 100 mg qd PO	—
Dapsone, 50 mg qd PO; plus pyrimethamine, 50 mg weekly PO; plus leucovorin, 25 mg weekly PO	Leucovorin prevents bone marrow toxicity from pyrimethamine.
Dapsone, 200 mg weekly PO; plus pyrimethamine, 75 mg weekly PO; plus leucovorin, 25 mg weekly PO	Leucovorin prevents bone marrow toxicity from pyrimethamine.
Pentamidine, 300 mg monthly via Respirgard II nebulizer	Adverse reactions include cough and bronchospasm.
Atovaquone, 1500 mg qd PO	—
TMP-SMX, 1 DS tablet three times weekly PO	TMP-SMX can be safely reintroduced for treatment of some patients who have had mild to moderate side effects.

[a]For a list of adverse effects, see Table 207-1.

[b]One DS tablet contains 160 mg of TMP and 800 mg of SMX.

Abbreviations: DS, double-strength; SS, single-strength; TMP-SMX, trimethoprim-sulfamethoxazole.

FURTHER READINGS

Cushion MT et al: Echinocandin treatment of *Pneumocystis* pneumonia in rodent models depletes cysts leaving trophic burdens that cannot transmit the infection. PloS One 5:e8524, 2010

Helweg-Larsen J et al: Clinical efficacy of first- and second-line treatments for HIV-associated *Pneumocystis jirovecii* pneumonia: A tri-centre cohort study. J Antimicrob Chemother 64:1282, 2009[PMID: 19858161]

Kaplan JE et al: Guidelines for prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: Recommendations from CDC, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. MMWR Recomm Rep 58(RR-4):1–207; quiz CE1–4, 2009 [whole issue]

Krajicek BJ et al: *Pneumocystis* pneumonia: Current concepts in pathogenesis, diagnosis, and treatment. Clin Chest Med 30:265, vi, 2009

Miller RF et al: Improved survival for HIV infected patients with severe *Pneumocystis jirovecii* pneumonias independent of highly active antiretroviral therapy. Thorax 61:716, 2006[PMID: 16601092]

Morris A et al: Epidemiology and clinical significance of Pneumocystis colonization. J Infect Dis 197:10, 2008[PMID: 18171279]

Tipirneni R et al: Healthcare worker occupation and immune response to *Pneumcystis jirovecii*. Emerg Infect Dis 15:1590, 2009[PMID: 19861050]

Walzer PD et al: Early predictors of mortality from *Pneumocystis jirovecii* pneumonia in HIV-infected patients: 1985–2006. Clin Infect Dis 46:634, 2008

Yazaki H et al: Outbreak of *Pneumocystis jiroveci* pneumonia in renal transplant recipients: *P. jiroveci* is contagious to the susceptible host. Transplantation 88:380, 2009[PMID: 19667941]

Zolopa A et al: Early antiretroviral therapy reduces AIDS progression/death in individuals with acute opportunistic infections; a multicenter randomized strategy trial. PloS One 4:e5575, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LABORATORY DIAGNOSIS OF PARASITIC INFECTIONS: INTRODUCTION

The cornerstone for the diagnosis of parasitic infections is a thorough history of the patient's illness. Epidemiologic aspects of the illness are especially important because the risks of acquiring many parasites are closely related to occupation, recreation, or travel to areas of high endemicity. Without a basic knowledge of the epidemiology and life cycles of the major parasites, it is difficult to approach the diagnosis of parasitic infections systematically. Accordingly, the medical classification of important human parasites in this chapter emphasizes their geographic distribution, their transmission, and the anatomic location and stages of their life cycle in humans. The text and tables are intended to serve as a guide to the correct diagnostic procedures for the major parasitic infections; in addition, the reader is referred to other chapters that contain more comprehensive information about each infection (Chaps. 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, and 220). Tables e25-1, e25-2, and e25-3 summarize the geographic distributions, the anatomic locations, and the methods employed for the diagnosis of flatworm, roundworm, and protozoal infections, respectively.

Table e25-1 Flatworm Infections

		Life-Cycle Hosts		Diagnosis			
Parasite	**Geographic Distribution**	**Intermediate (Transmission)**	**Definitive**	**Parasite Stage**	**Body Fluid or Tissue**	**Serologic Tests**	**Other**
Tapeworms (Cestodes)							
Intestinal tapeworms							
Taenia saginata (beef tapeworm)	Worldwide	Beef	Humans	Ova, segments	Feces	—	Motile segments
Hymenolepis nana (dwarf tapeworm)	Worldwide	Grain beetles	Humans, mice[a]	Ova	Feces	—	—
Diphyllobothrium latum (fish tapeworm)	Worldwide	Copepods –fish[b]	Humans, other mammals	Ova, segments	Feces	—	Megaloblastic anemia in 1%
T. solium[c] (pork tapeworm)	Worldwide	Swine	Humans	Ova, segments	Feces	WB	Especially Mexico, Central and South America, Africa
Somatic tapeworms							
Echinococcus granulosus (hydatid disease)	Sheep-raising and hunting areas	Sheep, camels, humans, others	Dogs	Hydatid	Lung, liver	WB, EIA	Chest radiography, CT, MRI
E. multilocularis (hydatid disease)	Subarctic areas	Rodents, humans	Foxes, dogs, cats	Hydatid	Liver	—	May resemble cholangiocellular carcinoma
T. solium[c] (pork tapeworm)	Worldwide	Swine, humans	Humans	Cysticercus	Muscles, CNS	WB	CT, MRI, radiography
Flukes (Trematodes)							
Intestinal flukes							
Fasciolopsis buski	China, India	Snails–water chestnuts	Humans	Ova	Feces	—	—
Heterophyes heterophyes	Far East, India	Snails–fish	Humans	Ova	Feces	—	—
Metagonimus yokogawai	Far East, Balkans, North Africa	Snails–fish	Humans	Ova	Feces	—	—
Liver flukes							
Clonorchis sinensis	China, Southeast Asia	Snails–fish	Humans	Ova	Feces, bile	—	Recurrent bacterial cholangitis
Fasciola hepatica	Sheep-raising areas	Snails –watercress	Humans, sheep	Ova	Feces,[d] bile	EIA	Cirrhosis, portal hypertension
Lung flukes							
Paragonimus spp.	Orient, Africa, the Americas	Snails –crabs/crayfish	Humans, other mammals	Adults, ova	Lung, sputum, feces	WB, EIA	Chest radiography, CT, MRI
Blood flukes							
Schistosoma mansoni	Africa, Central and South America, West Indies	Snails	Humans	Ova, adults	Feces	EIA, WB	Rectal snips, liver biopsy
S. haematobium	Africa	Snails	Humans	Ova, adults	Urine	WB	Liver, urine, or bladder biopsy
S. japonicum	Far East	Snails	Humans	Ova, adults	Feces	WB	Liver biopsy

[a]Larvae also can mature in intestinal villi of humans and mice.

[b]When there are two intermediate hosts, the first is separated from the second by a dash. Definitive hosts are infected by the second intermediate host.

[c]T. solium can cause either intestinal infections or cysticercosis. Its ova are identical to those of T. saginata; scolices and segments of the two species differ.

[d]Ova seldom reach the fecal stream during acute disease.

Note: CNS, central nervous system; EIA, enzyme immunoassay; WB, western blot. Serologic tests listed in Tables e25-1, e25-2, and e25-3 are available commercially or from the Centers for Disease Control and Prevention, Atlanta, GA.

Table e25-2 Roundworm Infections

		Life-Cycle Hosts			Diagnosis		
Parasite	**Geographic Distribution**	**Intermediate (Transmission)**	**Definitive**	**Parasite Stage**	**Body Fluid or Tissue**	**Serologic Tests**	**Other**
Intestinal Roundworms							
Enterobius vermicularis (pinworm)	Temperate and tropical zones	Fecal-oral	Humans	Ova	Perianal skin	—	"Cellophane tape" test
Trichuris trichiura (whipworm)	Temperate and tropical zones	Soil, fecal-oral	Humans	Ova	Feces	—	Rectal prolapse
Ascaris lumbricoides (roundworm of humans)	Temperate and tropical zones	Soil, fecal-oral	Humans	Ova	Feces	—	Sx of pulmonary migration
Ancylostoma duodenale (Old World hookworm)	Eurasia, Africa, Pacific	Soil to skin	Humans	Ova/larvae	Feces	—	Sx of pulmonary migration, anemia
Necator americanus (New World hookworm)	U.S., Africa, worldwide	Soil to skin	Humans	Ova/larvae	Feces	—	Sx of pulmonary migration, anemia
Strongyloides stercoralis (strongyloidiasis)	Moist tropics and subtropics	Soil to skin	Humans	Larvae	Feces, sputum, duodenal fluid	EIA	Dissemination in immunodeficiency
Capillaria philippinensis	Southeast Asia, Taiwan, Egypt	Raw fish	Birds	Ova, larvae, adults	Feces	—	Malabsorption/autoinfection, biopsy
Tissue Roundworms							
Trichinella spiralis (trichinellosis)	Worldwide	Swine/humans	Swine/humans	Larvae	Muscle	EIA	Muscle biopsy
Wuchereria bancrofti (filariasis)	Coastal areas in tropics and subtropics	Mosquitoes	Humans	Microfilariae	Blood, lymph nodes	EIA, RAPID	Nocturnal periodicity[a]
Brugia malayi (filariasis)	Asia, Indian subcontinent	Mosquitoes	Humans	Microfilariae	Blood	EIA, RAPID	Nocturnal
Loa loa (African eye worm)	West and Central Africa	Mango flies (*Chrysops*)	Humans	Microfilariae	Blood	RAPID	May be visible in eye, diurnal
Onchocerca volvulus (river blindness)	Africa, Mexico, Central and South America	Blackflies	Humans	Adults/larvae	Skin/eye	—	Examine nodules or skin snips
Dracunculus medinensis (guinea worm)	Africa	*Cyclops*	Humans	Adults/larvae	Skin	—	May be visible in lesion
Angiostrongylus cantonensis	Southeast Asia, Pacific, Caribbean	Snails/slugs, shrimp/fish	Rats	Larvae	CSF (rarely found)	—	Eosinophilic meningitis
Larva Migrans Syndromes							
Ancylostoma braziliense (creeping eruption)	Tropical and temperate zones	Soil to skin	Dogs/cats, humans	Larvae	Skin	—	Dog and cat hookworm
Toxocara canis and *cati* (visceral larva migrans)	Tropical and temperate zones	Soil, fecal-oral	Dogs/cats, humans	Larvae	Viscera, CNS, eye	EIA	Also caused by roundworms of other species

[a]Except for infection acquired in the South Pacific, blood should be drawn at midnight.

Note: CNS, central nervous system; CSF, cerebrospinal fluid; EIA, enzyme immunoassay; RAPID, rapid immunographic assay (available at the National Institutes of Health: 301-496-5398); Sx, signs/symptoms.

Table e25-3 Protozoal Infections

Parasite	Geographic Distribution	Life-Cycle Hosts		Diagnosis			
		Intermediate (Transmission)	Definitive	Parasite Stage	Body Fluid or Tissue	Serologic Tests	Other
Intestinal Protozoans							
Entamoeba histolytica (amebiasis)	Worldwide, especially tropics	Fecal-oral	Humans	Troph, cyst	Feces, liver	EIA, antigen detection	Ultrasound, liver CT, PCR
Giardia lamblia (giardiasis)	Worldwide	Fecal-oral	Humans	Troph, cyst	Feces	Antigen detection	String test, DFA, PCR
Isospora belli	Worldwide	Fecal-oral	Humans	Oocyst	Feces	—	Acid-fast[a]
Cryptosporidium	Worldwide	Fecal-oral	Humans, other animals	Oocyst	Feces	Antigen detection	Acid-fast,[a] DFA, biopsy, PCR
Cyclospora cayetanensis	Worldwide?	Fecal-oral	Humans, other animals?	Oocyst	Feces	—	Acid-fast,[a] modified safranin, autofluorescence, biopsy, PCR
Microsporidia (*Enterocytozoon bieneusi*, *Encephalitozoon* spp.) (microsporidiosis)	Worldwide?	?	Animals, humans	Spore	Feces	—	Modified trichrome, biopsy, PCR
Free-Living Amebas							
Naegleria	Worldwide	Warm water	Humans	Troph, cyst	CNS, nares	DFA	Biopsy, nasal swab, culture
Acanthamoeba	Worldwide	Soil, water	Humans	Troph, cyst	CNS, skin, cornea	DFA	Biopsy, scrapings, culture
Balamuthia	The Americas	Soil?	Humans, other animals	Troph, cyst	Brain	DFA	Biopsy
Blood and Tissue Protozoans							
Plasmodium spp. (malaria)	Subtropics and tropics	Mosquitoes	Humans	Asexual	Blood	Limited use	PCR
Babesia microti (babesiosis)	U.S., especially New England	Ticks	Rodents, humans	Asexual	Blood	IIF	Animal spp. in asplenia, PCR
Trypanosoma rhodesiense (African sleeping sickness)	Sub-Saharan East Africa	Tsetse flies	Humans, herbivores	Tryp	Blood, CSF	IIF[b]	Also chancre, lymph nodes
T. gambiense (African sleeping sickness)	Sub-Saharan West Africa	Tsetse flies	Humans, swine	Tryp	Blood, CSF	Card agglutination,[c] IIF[b]	Also chancre, lymph nodes
T. cruzi (Chagas' disease)	Mexico to South America	Reduviid bugs (triatomes)	Humans, dogs, wild animals	Amastigote, tryp	Multiple organs/blood	IIF, EIA	Reactivation in immunosuppression
Leishmania tropica, etc.	Widespread in tropics and subtropics	Sandflies (*Phlebotomus*)	Humans, dogs, rodents	Amastigote	Skin	IFA, EIA[d]	Biopsy, scrapings, culture
L. braziliensis (mucocutaneous)	Mexico to South America	Sandflies (*Lutzomyia*)	Humans, dogs, rodents	Amastigote	Skin, mucous membranes	IFA[b], EIA	Biopsy, scrapings, culture
L. donovani (kala-azar)	Widespread in tropics and subtropics	Sandflies (*Phlebotomus*)	Humans, dogs, wild animals	Amastigote	RE system	IFA[b], EIA	Biopsy, culture, PCR
Toxoplasma gondii (toxoplasmosis)	Worldwide	Humans, other mammals	Cats	Cyst, troph	CNS, eye, muscles, other	EIA, IIF	PCR

[a]Acid-fastness is best demonstrated by auramine fluorescence or modified acid-fast stain.

[b]Contact the CDC at 770-488-7760.

[c]Card agglutination is provided to endemic countries by the World Health Organization.

[d]Limited specificity; most sensitive for *L. donovani*.

Abbreviations: CNS, central nervous system; CSF, cerebrospinal fluid; DFA, direct fluorescent antibody; EIA, enzyme immunoassay; IFA, indirect fluorescent antibody; IIF, indirect immunofluorescence; PCR, polymerase chain reaction; RE, reticuloendothelial; troph, trophozoite; tryp, trypomastigote form.

In addition to selecting the correct diagnostic procedures, physicians must counsel their patients to ensure that specimens are collected properly and arrive at the laboratory promptly. For example, the diagnosis of bancroftian filariasis is unlikely to be confirmed by the laboratory unless blood is drawn near midnight, when the nocturnal microfilariae are active. Laboratory personnel and surgical pathologists should be notified in advance when a parasitic infection is suspected. Continuing interaction with the laboratory staff and the surgical pathologists increases the likelihood that parasites in body fluids or biopsy specimens will be examined carefully by the most capable individuals.

INTESTINAL PARASITES

Most helminths and protozoa exit the body in the fecal stream. The patient should be instructed to collect feces in a clean waxed or cardboard container and to record the time of collection on the container. Contamination with water (which could contain free-living protozoa) or with urine (which can damage trophozoites) should be avoided. Fecal samples should be collected before ingestion of barium or other contrast agents for radiologic procedures and before treatment with antidiarrheal agents and antacids, because these substances change the consistency of the feces and interfere with microscopic detection of parasites. Because of the cyclic shedding of most parasites in the feces, a minimum of three samples collected on alternate days should be examined. Examination of a single sample can be up to 50% less sensitive. When delays in transport to the laboratory are unavoidable, fecal samples should be kept in polyvinyl alcohol or another fixative to preserve protozoal trophozoites. New collection kits with instructions for the patient to transfer portions of the sample directly into fixative and bacterial carrier medium may enhance the recovery of trophozoites. Refrigeration will also preserve trophozoites for a few hours and protozoal cysts and helminthic ova for several days.

Analysis of fecal samples entails both macroscopic and microscopic examination. Watery or loose stools are more likely to contain protozoal trophozoites, but protozoal cysts and all stages of helminths may be found in formed feces. If adult worms or tapeworm segments are observed, they should be transported promptly to the laboratory or washed and preserved in fixative for later examination. The only tapeworm with motile segments is *Taenia saginata*, the beef tapeworm, which patients sometimes bring to the physician. Because the ova of *T. saginata* are morphologically indistinguishable from those of *Taenia solium* (the cause of cysticercosis), motility is an important distinguishing characteristic.

Microscopic examination of feces is not complete until direct wet mounts have been evaluated and concentration techniques as well as permanent stains have been applied. Before accepting a report of negativity for ova and parasites as final, the physician should insist that the laboratory undertake each of these procedures. Some intestinal parasites are more readily detected in material other than feces. For example, examination of duodenal contents is sometimes necessary to detect *Giardia lamblia*, *Cryptosporidium*, and *Strongyloides* larvae. Use of the "cellophane tape" technique to detect pinworm ova on the perianal skin sometimes also reveals ova of *T. saginata* deposited perianally when the motile segments disintegrate (Table e25-4).

Table e25-4 Alternative Procedures for Laboratory Diagnosis of Parasites Found in Feces[a]

Parasites and Fecal Stages	Alternative Diagnostic Procedures
Tapeworms (Cestodes)	
Taenia saginata ova and segments	Perianal "cellophane tape" test for ova
T. solium ova and segments	Serology; brain biopsy for neurocysticercosis
Flukes (Trematodes)	
Clonorchis (Opisthorchis) sinensis ova	Examination of bile for ova and adults in cholangitis
Fasciola hepatica ova	Examination of bile for ova and adults in cholangitis
Paragonimus ova	Serology; sputum; biopsy of lung or brain for larvae
Schistosoma ova	Serology for all; rectal snips (especially for *S. mansoni*), urine (*S. haematobium*), liver biopsy and liver ultrasound
Roundworms	
Enterobius vermicularis ova and adults	Perianal "cellophane tape" test for ova and adults
Trichuris trichiura ova	None
Ascaris lumbricoides ova and adults	Examination of sputum for larvae in lung disease
Hookworm ova and occasional larvae	Examination of sputum for larvae in lung disease
Strongyloides larvae	Duodenal aspirate or jejunal biopsy; serology; sputum or lung biopsy for filariform larvae in disseminated disease
Protozoans	
Entamoeba histolytica trophozoites and cysts	Serology; liver biopsy for trophozoites
Giardia lamblia trophozoites and cysts	Duodenal aspirate or jejunal biopsy[b]
Isospora belli oocysts	Duodenal aspirate or jejunal biopsy[b]
Cryptosporidium oocysts	Duodenal aspirate or jejunal biopsy[b]
Microsporidial spores	Duodenal aspirate or jejunal biopsy[b]

[a]Stains and concentration techniques are discussed in the text.

[b]Commercial string test is satisfactory; *Isospora* and *Cryptosporidium* are acid-fast.

Two routine solutions are used to make wet mounts for identification of the various life stages of helminths and protozoa: physiologic saline for trophozoites, cysts, ova, and larvae and dilute iodine solution for protozoal cysts and ova. Iodine solution must never be used to examine specimens for trophozoites because it kills the parasites and thus eliminates their characteristic motility.

The two most common concentration procedures for detecting small numbers of cysts and ova are formalin-ether sedimentation and zinc sulfate flotation. The formalin-ether technique is preferable because all parasites sediment but not all float. Slides permanently stained for trophozoites should be prepared before concentration. Additional slides stained for cysts and ova may be made from the concentrate.

In many instances, especially in the differentiation of *Entamoeba histolytica* from other amebas, identification of parasites from wet mounts or concentrates must be considered tentative. Permanently stained smears allow study of the cellular detail necessary for definitive identification. The iron-hematoxylin stain is excellent for critical work, but trichrome staining, which can be

completed in 1 h, is a satisfactory alternative that also reveals parasites in specimens preserved in polyvinyl alcohol fixative. Modified acid-fast staining and fluorescent auramine microscopy are useful adjuncts for detection and identification of several intestinal protozoa, including *Cryptosporidium* and *Cyclospora.* Microsporidia, which cause chronic diarrhea in HIV-infected patients, may be missed unless a special modified trichrome stain is requested (Table e25-3).

BLOOD AND TISSUE PARASITES

Invasion of tissue by protozoa and helminths renders the choice of diagnostic techniques more difficult. For example, physicians must understand that aspiration of an amebic liver abscess rarely reveals *E. histolytica* because the trophozoites are located primarily in the abscess wall. They must remember that the urine sediment offers the best opportunity to detect *Schistosoma haematobium* in the young Ethiopian immigrant or the American traveler who returns from Africa with hematuria. Tables e25-1, e25-2, and e25-3, which offer a quick guide to the geographic distribution and anatomic locations of the major tissue parasites, should help the physician to select the appropriate body fluid or biopsy site for microscopic examination. Tables e25-5 and e25-6 provide additional information about the identification of parasites in samples from specific anatomic locations. The laboratory procedures for detection of parasites in other body fluids are similar to those used in the examination of feces. The physician should insist on wet mounts, concentration techniques, and permanent stains for all body fluids. The trichrome or iron-hematoxylin stain is satisfactory for all tissue helminths in body fluids other than blood, but microfilarial worms and blood protozoa are more easily visualized with Giemsa or Wright's staining.

Table e25-5 Identification of Parasites in Blood and Other Body Fluids

Body Fluid, Parasite	Enrichment/Stain	Culture Technique
Blood		
Plasmodium spp.	Thick and thin smears/Giemsa or Wright's	Not useful for diagnosis
Leishmania spp.	Buffy coat/Giemsa	Media available from CDC
African trypanosomes[a]	Buffy coat, anion column/wet mount and Giemsa	Mouse or rat inoculation[b]
Trypanosoma cruzi[c]	As for African species	As above and xenodiagnosis
Toxoplasma gondii	Buffy coat/Giemsa	Fibroblast cell lines
Microfilariae[d]	Filtration/wet mount and Giemsa	None
Urine		
Schistosoma haematobium	Centrifugation/wet mount	None
Microfilariae (in chyluria)	As for blood	None
Spinal fluid		
African trypanosomes	Centrifugation, anion column/wet mount and Giemsa	As for blood
Naegleria fowleri	Centrifugation/wet mount and Giemsa or trichrome	Nonnutrient agar overlaid with *Escherichia coli*

[a] *Trypanosoma rhodesiense* and *T. gambiense*.

[b] Inject mice intraperitoneally with 0.2 mL of whole heparinized blood (0.5 mL for rats). After 5 days, check tail blood daily for trypanosomes as described above. Call the CDC (770-488-7775) for information on diagnosis and treatment.

[c] Detectable in blood by conventional techniques only during acute disease. Xenodiagnosis is successful in ~50% of patients with chronic Chagas' disease.

[d] Daytime (1000–1400 h) and nighttime (2200–0200 h) blood samples should be drawn to maximize the chance of detecting *Wuchereria* (nocturnal except for Pacific strains), *Brugia* (nocturnal), and *Loa loa* (diurnal).

Table e25-6 Minor Procedures for Diagnosis of Parasitic Infections

Parasite(s) and Stage	Procedure
Onchocerca volvulus and *Mansonella streptocerca* microfilariae	*Skin snips:* Lift skin with a needle and excise ~1 mg to a depth of 0.5 mm from several sites. Weigh each sample, place it in 0.5 mL of saline for 4 h, and examine wet mounts and Giemsa stains of the saline either directly or after filtration. Count microfilariae.[a]
Loa loa adults and *O. volvulus* adults and microfilariae	*Biopsies of subcutaneous nodules:* Stain routine histopathologic sections and impression smears with Giemsa.
Trichinella spiralis larvae (and perhaps *Taenia solium* cysticerci)	*Muscle biopsies:* Excise ~1.0 g of deltoid or gastrocnemius muscle and squash between two glass slides for direct microscopic examination.
Schistosoma ova of all species, but especially *S. mansoni*	*Rectal snips:* From four areas of mucosa, take 2-mg snips, tease onto a glass slide, and flatten with a second slide before examining directly at 10x. Preparations may be fixed in alcohol or stained.
Trypanosoma gambiense and *T. rhodesiense* trypomastigotes	*Aspirate of chancre or lymph node:* [b] Aspirate center with an 18-gauge needle, place a drop on a slide, and examine for motile forms. An otherwise insufficient volume of material may be stained with Giemsa.
Acanthamoeba spp. trophozoites or cysts	*Corneal scrapings:* Have an ophthalmologist obtain a sample for immediate Giemsa staining and culture on nutrient agar overlaid with *Escherichia coli*.
Cutaneous and mucocutaneous *Leishmania* spp.	*Swabs, aspirates, or punch biopsies of skin lesions:* Obtain a specimen from the margin of a lesion for Giemsa staining of impression smears; section and culture on special media from the CDC.

[a]Counts of >100/mg are associated with a significant risk of complications.

[b]Lymph node aspiration is contraindicated in some infections and should be used judiciously.

The most common parasites detected in Giemsa-stained blood smears are the plasmodia, microfilariae, and African trypanosomes (Table e25-5). Most patients with Chagas' disease present in the chronic phase, when *Trypanosoma cruzi* is no longer microscopically detectable in blood smears. Wet mounts are sometimes more sensitive than stained smears for the detection of microfilariae and African trypanosomes because these active parasites cause noticeable movement of the erythrocytes in the microscopic field. Filtration of blood through a polycarbonate filter (pore size, 3–5 μm) facilitates the detection of microfilariae. The intracellular amastigote forms of *Leishmania* spp. and *T. cruzi* can sometimes be visualized in stained smears of peripheral blood, but aspirates of the bone marrow, liver, and spleen are the best sources for microscopic detection and culture of *Leishmania* in kala-azar and of *T. cruzi* in chronic Chagas' disease. The diagnosis of malaria and the critical distinction among the various *Plasmodium* species are made by microscopic examination of stained thick and thin blood films (Chap. 210). Detection of subpatent infection and identification of *Plasmodium* species can be confirmed by PCR. Recently, *P. knowlesi,* a simian parasite, has been identified as the cause of an increasing number of infections in Malaysian Borneo and other areas of Southeast Asia. PCR or another molecular method is required to differentiate *P. knowlesi* from *P. malariae.*

Although most tissue parasites stain with the traditional hematoxylin and eosin, surgical biopsy specimens should also be stained with appropriate special stains. The surgical pathologist who is accustomed to applying silver stains for *Pneumocystis* to induced sputum and transbronchial biopsies may need to be reminded to examine wet mounts and iron-hematoxylin–stained preparations of pulmonary specimens for helminthic ova and *E. histolytica*. The clinician should also be able to advise the surgeon and pathologist about optimal techniques for the identification of parasites in specimens obtained by certain specialized minor procedures (Table e25-6). For example, the excision of skin snips for the diagnosis of onchocerciasis, the collection of rectal snips for the diagnosis of schistosomiasis, and punch biopsy of skin lesions for the identification and culture of cutaneous and mucocutaneous species of *Leishmania* are simple procedures, but the diagnosis can be missed if the specimens are improperly obtained or processed.

NONSPECIFIC TESTS

Eosinophilia (>500/μL) commonly accompanies infections with most of the tissue helminths; absolute numbers of eosinophils may be high in trichinellosis and the migratory phases of filariasis (Table e25-7). Intestinal helminths provoke eosinophilia only during pulmonary migration of the larval stages. Eosinophilia is not a manifestation of protozoal infections. Parasitic causes of eosinophilia in cerebrospinal fluid include nematodes (e.g., *Angiostrongylus*, *Gnathostoma*, *Toxocara*, and *Baylisascaris* species) as well as flatworms (e.g., *Taenia solium* and *Schistosoma* species).

Table e25-7 Parasites Frequently Associated With Eosinophilia[a]

Parasite	Comment
Tapeworms (Cestodes)	
Echinococcus granulosus	When hydatid cyst leaks
Taenia solium	During muscle encystation and in cerebrospinal fluid with neurocysticercosis
Flukes (Trematodes)	
Paragonimus spp.	Uniformly high in acute stage
Fasciola hepatica	May be high in acute stage
Clonorchis (Opisthorchis) sinensis	Variable
Schistosoma mansoni	50% of infected travelers
S. haematobium	25% of infected travelers
S. japonicum	Up to 6000/μL in acute infection
Roundworms	
Ascaris lumbricoides	During larval migration
Hookworm species	During larval migration
Strongyloides stercoralis	Profound during migration and early years of infection
Trichinella spiralis	Up to 7000/μL
Filarial species[b]	Varies but can reach 5000–8000/μL
Toxocara spp.	>3000/μL
Ancylostoma braziliense	With extensive cutaneous eruption
Gnathostoma spinigerum	In visceral larva migrans and eosinophilic meningitis
Angiostrongylus cantonensis	In eosinophilic meningitis
A. costaricensis	During larval migration in mesenteric vessels

[a]Virtually every helminth has been associated with eosinophilia. This table includes both common and uncommon parasites that frequently elicit eosinophilia during infection.

[b]*Wuchereria bancrofti, Brugia* spp., *Loa loa,* and *Onchocerca volvulus*.

Like the hypochromic microcytic anemia of heavy hookworm infections, other nonspecific laboratory abnormalities may suggest parasitic infection in patients with appropriate geographic and/or environmental exposures. Biochemical evidence of cirrhosis or an abnormal urine sediment in an African immigrant certainly raises the possibility of schistosomiasis, and anemia and thrombocytopenia in a febrile traveler or immigrant are among the hallmarks of malaria. CT and MRI also contribute to the diagnosis of infections with many tissue parasites and have become invaluable adjuncts in the diagnosis of neurocysticercosis and cerebral toxoplasmosis.

ANTIBODY AND ANTIGEN DETECTION

Useful antibody assays for many of the important tissue parasites are available; most of those listed in Table e25-8 can be obtained from the Centers for Disease Control and Prevention (CDC) in Atlanta. The results of serologic tests not listed in the tables should be interpreted with caution.

Table e25-8 Serologic and Molecular Tests for Parasitic Infections[a]

Parasite, Infection	Antibody	Antigen or DNA/RNA
Tapeworms		
Echinococcosis	WB, EIA	
Cysticercosis	WB	
Flukes		
Paragonimiasis	WB, EIA[b]	
Schistosomiasis	EIA, WB	
Fascioliasis	EIA[b]	
Roundworms		
Strongyloidiasis	EIA	
Trichinellosis	EIA	
Toxocariasis	EIA	
Filariasis	EIA[c]	RAPID[c]
Protozoans		
Amebiasis	EIA	EIA,[b] RAPID,[b] PCR
Giardiasis		EIA,[b] RAPID,[b] DFA, PCR
Cryptosporidiosis		EIA,[b] DFA, RAPID,[b] PCR
Malaria (all species)	IIF[d]	RAPID, PCR
Babesiosis	IIF	PCR
Chagas' disease	IIF, EIA	PCR
Leishmaniasis	IIF, EIA	PCR[b]
Toxoplasmosis	IIF, EIA (IgM)[e]	PCR[b]
Microsporidiosis		PCR
Cyclosporiasis		PCR
Acanthamebiasis		DFA, PCR
Naegleriasis		DFA, PCR
Balmuthiasis		DFA

[a]Unless otherwise noted, all tests are available at the CDC.

[b]Research or commercial laboratories only.

[c]Available at the NIH (301-496-5398) and commercially.

[d]Of limited use for management of acute disease.

[e]Determination of infection within the last 3 months may require additional tests by a research laboratory.

Note: DFA, direct fluorescent antibody; EIA, enzyme immunoassay; IIF, indirect immunofluorescence; PCR, polymerase chain reaction; RAPID, rapid immunographic assay; WB, western blot. Most antigen and antibody parasite detection kits are available commercially. Most PCRs listed are now available at the CDC and in commercial or research laboratories. Contact Dr. Alexandre da Silva at the CDC (770-488-4072).

The value of antibody assays is limited by several factors. For example, the preparation of thick and thin blood smears remains the procedure of choice for the diagnosis of malaria in individual patients because diagnostic titers to plasmodia develop slowly and do not differentiate species—a critical step in patient management. Filarial antigens cross-react with those from other nematodes; as in assays for antibody to most parasites, the presence of antibody in the filarial assay fails to distinguish between past and current infection. Despite these specific limitations, the restricted geographic distribution of many tropical parasites increases the diagnostic usefulness of both the presence and the absence of antibody in travelers from industrialized countries. In contrast, a large proportion of the world's population has been exposed to *Toxoplasma gondii*, and the presence of IgG antibody to *T. gondii* does not constitute proof of active disease.

Fewer antibody assays are available for the diagnosis of infection with intestinal parasites. *E. histolytica* is the major exception. Sensitive, specific serologic tests are invaluable in the diagnosis of amebiasis. Commercial kits for the detection of antigen by enzyme-linked immunosorbent assay or of whole organisms by fluorescent antibody assay are now available for several protozoan parasites. A rapid test (approved by the U.S. Food and Drug Administration) for the detection of *P. falciparum* in blood is less sensitive than thick smears read by an experienced microscopist, but its use is increasing in developing countries because of its simplicity (Table e25-8).

MOLECULAR TECHNIQUES

DNA hybridization with probes that are repeated many times in the genome of a specific parasite and amplification of a specific DNA fragment by the polymerase chain reaction (PCR) have now been established as useful techniques for the diagnosis of several protozoan infections (Table e25-8). Although PCR is very sensitive, it is an adjunct to conventional techniques for parasite detection and should be requested only when microscopic and immunodiagnostic procedures fail to establish the probable

diagnosis. For example, only multiple negative blood smears or the failure to identify the infecting species justifies PCR for the diagnosis or proper management of malaria. In addition to PCR of anticoagulated blood, the CDC (contact Dr. Alexandre da Silva, 770-488-4072, for details) and several commercial laboratories now perform PCRs for detection of certain specific parasites in stool samples, biopsy specimens, and bronchoalveolar lavage fluid (Table e25-8). Although PCRs are now used primarily for the detection of protozoans, active research efforts are likely to establish their feasibility for the detection of several helminths.

FURTHER READINGS

Cox-Singh J et al: *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. Clin Infect Dis 46:165, 2008[PMID: 18171245]

Freedman DO et al: Spectrum of disease and relation to place of exposure among ill returned travelers. N Engl J Med 354:119, 2006[PMID: 16407507]

Garcia LS et al: Algorithms for detection and identification of parasites, in *Manual of Clinical Microbiology* , 9th ed, vol 2, PR Murray et al (eds). Washington, DC, ASM Press, 2007, pp 2020–2039

Graff-Teixeira C et al: Update on eosinophilic meningoencephalitis and its clinical relevance. Clin Microbiol Rev 22:322, 2009

Murray CK et al: Update on rapid diagnostic testing for malaria. Clin Microbiol Rev 21:97, 2008[PMID: 18202438]

Rosenblatt JE: Laboratory diagnosis of infections due to blood and tissue parasites. Clin Infect Dis 49:1103, 2009[PMID: 19691431]

Schuster FL et al: Under the radar. *Balamuthia* amebic encephalitis. Clin Infect Dis 48:879, 2009[PMID: 19236272]

Seybolt LM et al: Diagnostic evaluation of newly arrived asymptomatic refugees with eosinophilia. Clin Infect Dis 42:363, 2006[PMID: 16392081]

Stark D et al: Clinical significance of enteric protozoa in the immunosuppressed human population. Clin Microbiol Rev 22:634, 2009[PMID: 19822892]

Thellier M, Breton J: Enterocytozoon bieneusi in humans and animals, focus on laboratory identification and molecular epidemiology. Parasite 15:349, 2008[PMID: 18814706]

Walker M et al: Parasitic central nervous system infections in immunocompromised hosts: Malaria, microsporidiosis, leishmaniasis, and African trypanosomiasis. Clin Infect Dis 42:115, 2006[PMID: 16323101]

Wilson M et al: Molecular immunological approaches to the diagnosis of parasitic infection, in *Manual of Molecular and Clinical Laboratory Immunology,* 7th ed, B Detrick et al (eds). Washington, DC, ASM Press, 2006, pp 557–568

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

AGENTS USED TO TREAT PARASITIC INFECTIONS: INTRODUCTION

Parasitic infections afflict more than half of the world's population and impose a substantial health burden, particularly in underdeveloped nations, where they are most prevalent. The reach of some parasitic diseases, including malaria, has expanded over the past few decades as a result of factors such as deforestation, population shifts, global warming, and other climatic events. Despite major efforts at vaccine development and vector control, chemotherapy remains the single most effective means of controlling parasitic infections. Efforts to combat the spread of some diseases are hindered by the development and spread of drug resistance, the limited introduction of new antiparasitic agents, and the proliferation of counterfeit medications. However, there are good reasons to be optimistic. The past 10 years have witnessed the launch of ambitious global initiatives aimed at controlling or eliminating threats such as AIDS, tuberculosis, and malaria. Recognition of the substantial burden imposed by the "neglected" tropical diseases has generated multinational partnerships to develop and deploy effective antiparasitic agents. Vaccines against several tropical diseases are being developed, and clinical trials have begun for vaccines against schistosomiasis, hookworm, and leishmaniasis.

This chapter deals exclusively with the agents used to treat infections due to parasites. Specific treatment recommendations for the parasitic diseases of humans are listed in subsequent chapters. **The pharmacology of the antiparasitic agents is discussed in great detail in Chap. e26.**

Table 208-1 presents a brief overview of each agent (including some drugs that are covered in other chapters), along with its major toxicities, spectrum of activity, and safety for use during pregnancy and lactation. Many of the agents are approved by the U.S. Food and Drug Administration but are considered investigational for the treatment of certain infections; these drugs are marked accordingly in the table. In addition, drugs available only through the Centers for Disease Control and Prevention (CDC) Drug Service (telephone: 404-639-3670 or 404-639-2888; *www.cdc.gov/laboratory/drugservice/*) or only through their manufacturers (whose contact information may be available from the CDC) are specified by footnotes in the table.

Table 208-1 Overview of Agents Used for the Treatment of Parasitic Infections

Drugs by Class	Parasitic Infection(s)	Adverse Effects	Major Drug-Drug Interactions	Pregnancy Class[a]	Breast Milk	
4-Aminoquinolines						
Amodiaquine	Malaria[b]	Agranulocytosis, hepatotoxicity	No information	Not assigned	No information	
Chloroquine	Malaria[b]	*Occasional:*pruritus, nausea, vomiting, headache, hair depigmentation, exfoliative dermatitis, reversible corneal opacity. *Rare:* irreversible retinal injury, nail discoloration, blood dyscrasias	Antacids and kaolin: reduced absorption of chloroquine Ampicillin: bioavailability reduced by chloroquine Cimetidine: increased serum levels of chloroquine Cyclosporine: serum levels increased by chloroquine	Not assigned[c]	Yes	
8-Aminoquinolines						
Primaquine	Malaria[b]		*Frequent:* hemolysis in patients with G6PD deficiency. *Occasional:* methemoglobinemia, GI disturbances. *Rare:* CNS symptoms	Quinacrine: potentiatedtoxicity of primaquine	Contraindicated	No information
Tafenoquine	Malaria[b]	**Frequent:** hemolysis in patients with G6PD deficiency, mild GI upset. *Occasional:* methemoglobinemia, headaches	No information	Not assigned	No information	
Aminoalcohols						
Halofantrine	Malaria[b]	*Frequent:* abdominal pain,diarrhea. *Occasional:* ECGdisturbances (dose-related prolongation of QTc and PR interval), nausea, pruritus. Contraindicated in persons who have cardiac disease or who have taken mefloquine in the preceding 3 weeks	Concomitant use of agents that prolong QTc intervalcontraindicated	C	No information	
Lumefantrine	Malaria[b]	*Occasional:* nausea, vomiting, diarrhea, abdominal pain, anorexia, headache, dizziness	No major interactions	Not assigned	No information	
Aminoglycosides						
Paromomycin	Amebiasis,[b] infection with *Dientamoeba fragilis,* giardiasis, cryptosporidiosis, leishmaniasis	*Frequent:* GI disturbances (oral dosing only). *Occasional:* nephrotoxicity, ototoxicity,vestibular toxicity (parenteraldosing only)	No major interactions	Not assigned[c]	No information	
Amphotericin B Amphotericin B deoxycholate	Leishmaniasis,[d] amebic meningoencephalitis	*Frequent:* fever, chills, hypokalemia, hypomagnesemia, nephrotoxicity. *Occasional:* vomiting, dyspnea, hypotension	Antineoplastic agents: renal toxicity, bronchospasm, hypotension Glucocorticoids, ACTH, digitalis: hypokalemia	B	No information	
Amphotec (InterMune)			Zidovudine: increased			

			myelo- and nephrotoxicity		
Amphotericin B lipid complex, ABLC (Abelcet)					
Amphotericin B, liposomal (AmBisome)					
Antimonials	Leishmaniasis				
Pentavalent antimony[e]		*Frequent:* arthralgias/myalgias, pancreatitis, ECG changes (QT prolongation, T wave flattening or inversion)	No major interactions	Not assigned	Yes
Meglumine antimoniate	*Frequent:* arthralgias/myalgias, pancreatitis, ECG changes (QT prolongation, T wave flattening or inversion)	Antiarrhythmics and tricyclic antidepressants: increased risk of cardiotoxicity	Not assigned	No information	
Artemisinin and derivatives	Malaria[f]	*Occasional:* neurotoxicity (ataxia, convulsions), nausea, vomiting, anorexia, contact dermatitis			
Arteether			No information	Not assigned	Yes[g]
Artemether			No clinically significant interactions	C	Yes[g]
Artesunate[e]			Mefloquine: levels decreased and clearance accelerated by artesunate	C	Yes[g]
Dihydroartemisinin	Mefloquine: increased absorption	Not assigned	Yes[g]		
Atovaquone	Malaria,[b] babesiosis	*Frequent:* nausea, vomiting. *Occasional:* abdominal pain, headache	Plasma levels decreased by rifampin, tetracycline;bioavailability decreased by metoclopramide	C	No information
Azoles	Leishmaniasis	Serious: hepatotoxicity. Rare: exfoliative skin disorders, anaphylaxis	Warfarin, oral hypoglycemics, phenytoin, cyclosporine, theophylline, digoxin, dofetilide, quinidine, carbamazepine, rifa –butin, busulfan, docetaxel, vinca alkaloids, pimozide, alprazolam, diazepam, midazolam, triazolam, verapamil, atorvastatin, ceriva–statin, lovastatin, simvastatin, tacrolimus, sirolimus, indinavir, ritonavir, saquinavir, alfentanil, buspirone, methylprednisolone, trimetrexate: plasma levels increased by azoles	C	Yes
Fluconazole			Carbamazepine, phenobarbital, phenytoin, isoniazid, rifabutin, rifampin, antacids, H2-receptor antagonists, proton pump inhibitors, nevirapine: decreased plasma levels of azoles		
Itraconazole			Clarithromycin, erythromycin, indinavir, ritonavir: increased plasma levels of azoles		
Ketoconazole					
Benzimidazoles					
Albendazole	Ascariasis, capillariasis, clonorchiasis, cutaneous larva migrans, cysticercosis,[b] echinococcosis,[b] enterobiasis, eosinophilic enterocolitis, gnathostomiasis, hookworm, lymphatic filariasis, microsporidiosis, strongyloidiasis, trichinellosis, trichostrongyliasis, trichuriasis, visceral larva migrans	*Occasional:* nausea, vomiting, abdominal pain, headache, reversible alopecia, elevated aminotransferases. *Rare:* leukopenia, rash	Dexamethasone, praziquantel: plasma level of albendazole sulfoxide increased by ~50%	C	Yes[g]
Mebendazole	Ascariasis,[b] capillariasis, eosinophilic enterocolitis, enterobiasis,[b] hookworm,[b] trichinellosis, trichostrongyliasis, trichuriasis,[b] visceral larva migrans	*Occasional:* diarrhea, abdominal pain, elevated aminotransferases. *Rare:* agranulocytosis, thrombocytopenia, alopecia	Cimetidine: inhibited mebendazole metabolism	C	No information
Thiabendazole	Strongyloidiasis,[b] cutaneous larva migrans,[b] visceral larva	*Frequent:* anorexia, nausea, vomiting, diarrhea, headache, dizziness, asparagus-	Theophylline: serum levels increased by thiabendazole	C	No information

	migrans[b]	dizziness, asparagus-like urine odor. *Occasional:* drowsiness, giddiness, crystalluria, elevated aminotransferases, psychosis. *Rare:* hepatitis, seizures, angioneurotic edema, Stevens-Johnson syndrome, tinnitus			
Triclabendazole	Fascioliasis, paragonimiasis	*Occasional:* abdominal cramps, diarrhea, biliary colic, transient headache	No information	Not assigned	Yes
Benznidazole	Chagas' disease	*Frequent:* rash, pruritus, nausea, leukopenia, paresthesias	No major interactions	Not assigned	No information
Bithionol[e]	Fascioliasis, paragonimiasis	Diarrhea, abdominal cramps (usually mild and transient)			
Clindamycin	Babesiosis, malaria, toxoplasmosis	*Occasional:* pseudomembranous colitis, abdominal pain, diarrhea, nausea/vomiting. *Rare:* pruritus, skin rashes	No major interactions	B	Yes[g]
Diloxanide furoate	Amebiasis	*Frequent:* flatulence. *Occasional:* nausea, vomiting, diarrhea. *Rare:* pruritus	None reported	Contraindicated	No information
Eflornithine[h] (difluoromethylornithine, DFMO)	Trypanosomiasis	*Frequent:* pancytopenia. *Occasional:* diarrhea, seizures. *Rare:* transient hearing loss	No major interactions	Contraindicated	No information
Emetine and dehydroemetine[e]	Amebiasis, fascioliasis	*Severe:* cardiotoxicity. *Frequent:* pain at injection site. *Occasional:* dizziness, headache, GI symptoms	None reported	X	No information
Folate antagonists					
Dihydrofolate reductase inhibitors					
Pyrimethamine	Malaria,[b] isosporiasis, toxoplasmosis[b]	*Occasional:* folate deficiency. *Rare:* rash, seizures, severe skin reactions (toxic epidermal necrolysis, erythema multiforme, Stevens-Johnson syndrome)	Sulfonamides, proguanil, zidovudine: increased risk of bone marrow suppression when used concomitantly	C	Yes
Proguanil and chlorproguanil	Malaria	*Occasional:* urticaria. *Rare:* hematuria, GI disturbances	No major interactions	C	Yes
Trimethoprim	Cyclosporiasis, isosporiasis	Hyperkalemia, GI upset, mild stomatitis	Methotrexate: reduced clearance Warfarin: effect prolonged Phenytoin: hepatic metabolism increased	C	Yes
Dihydropteroate synthetase inhibitors: sulfonamides	Malaria,[b] toxoplasmosis[b]	*Frequent:* GI disturbances, allergic skin reactions, crystalluria. *Rare:* severe skin reactions (toxic epidermal necrolysis, erythema multiforme, Stevens-Johnson syndrome), agranulocytosis, aplastic anemia, hypersensitivity of the respiratory tract, hepatitis, interstitial nephritis, hypoglycemia, aseptic meningitis	Thiazide diuretics: increased risk of thrombocytopenia in elderly patients Warfarin: effect prolonged by sulfonamides Methotrexate: levels increased by sulfonamides Phenytoin: metabolism impaired by sulfonamides	B	Yes
Sulfadiazine					
Sulfamethoxazole					
Sulfadoxine					
	Sulfonylureas: effect prolonged by sulfonamides				
Dihydropteroatesynthetase inhibitors: sulfones					
Dapsone	Leishmaniasis, malaria, toxoplasmosis	*Frequent:* rash, anorexia. *Occasional:* hemolysis, methemoglobinemia, neuropathy, allergic dermatitis, anorexia, nausea, vomiting, tachycardia, headache, insomnia, psychosis, hepatitis. Rare: agranulocytosis	Rifampin: lowered plasma levels of dapsone	C	Yes
Fumagillin	Microsporidiosis	*Rare:* neutropenia, thrombocytopenia	None reported	No information	No information
Furazolidone	Giardiasis	*Frequent:* nausea/vomiting, brown urine. *Occasional:* rectal itching, headache. *Rare:* hemolytic anemia, disulfiram-like reactions, MAO-inhibitor interactions	Risk of hypertensive crisis when administered for >5 days with MAO inhibitors	C	No information

Iodoquinol	Amebiasis,[b] balantidiasis, *D. fragilis* infection	*Occasional:* headache, rash, pruritus, thyrotoxicosis, nausea, vomiting, abdominal pain, diarrhea. *Rare:* optic neuritis, peripheral neuropathy, seizures,encephalopathy	No major interactions	C	No information
Ivermectin	Ascariasis, cutaneous larva migrans, gnathostomiasis, loiasis, lymphatic filariasis, onchocerciasis,[b] scabies, strongyloidiasis,[b] trichuriasis	*Occasional:* fever, pruritus, headache, myalgias. *Rare:* hypotension	No major interactions	C	Yes[g]
Levamisole	Ascariasis, hookworm	*Frequent:* GI disturbances,dizziness, headache. *Rare:* agranulocytosis, peripheralneuropathy	Alcohol: disulfiram-like effect Warfarin: prolonged prothrombin time	C	No information
Macrolides					
Azithromycin	Babesiosis	*Occasional:* nausea, vomiting, diarrhea, abdominal pain. *Rare:* angioedema, cholestatic jaundice	Cyclosporine and digoxin: levels increased by azithromycin Nelfinavir: increased levels of azithromycin	B	Yes
Spiramycin[h]	Toxoplasmosis	*Occasional:* GI disturbances,transient skin eruptions. *Rare:* thrombocytopenia, QT prolongation in an infant, cholestatic hepatitis	No major interactions	Not assigned[c]	Yes[g]
Mefloquine	Malaria[b]	*Frequent:* lightheadedness, nausea, headache. *Occasional:* confusion; nightmares; insomnia; visual disturbance; transient and clinically silent ECG abnormalities, including sinus bradycardia, sinus arrhythmia, first-degree AV block, prolongation of QTc interval, and abnormal T waves. *Rare:* psychosis, convulsions, hypotension	Administration of halofantrine <3 weeks after mefloquine use may produce fatal QTc prolongation. Mefloquine may lower plasma levels of anticonvulsants. Levels decreased and clearance accelerated by artesunate	C	Yes
Melarsoprol[e]	Trypanosomiasis	*Frequent:* myocardial injury, encephalopathy, peripheral neuropathy, hypertension. *Occasional:* G6PD-induced hemolysis, erythema nodosum leprosum. *Rare:* hypotension	No major interactions	Not assigned	No information
Metrifonate	Schistosomiasis	*Frequent:* abdominal pain, nausea, vomiting, diarrhea, headache, vertigo, bronchospasm. *Rare:* cholinergic symptoms	No major interactions	B	No
Miltefosine	Leishmaniasis	*Frequent:* mild and transient(1–2 days) GI disturbances within first 2 weeks of therapy (resolve after treatment completion); motion sickness. *Occasional:* reversible elevations of creatinine and aminotransferases	No major interactions	Not assigned	No information
Niclosamide	Intestinal cestodes[b]	*Occasional:* nausea, vomiting,dizziness, pruritus	No major interactions	B	No information
Nifurtimox[e]	Chagas' disease	*Frequent:* nausea, vomiting, abdominal pain, insomnia, paresthesias, weakness, tremors. *Rare:* seizures (all are reversible and dose-related)	No major interactions	Not assigned	No information
Nitazoxanide	Cryptosporidiosis,[b] giardiasis[b]	*Occasional:* abdominal pain, diarrhea. *Rare:* vomiting, headache	No major interactions	B	No information
Nitroimidazoles					
Metronidazole	Amebiasis,[b] balantidiasis, dracunculiasis, giardiasis, trichomoniasis,[b] D. fragilis infection	*Frequent:* nausea, headache, anorexia, metallic aftertaste. *Occasional:* vomiting, insomnia, vertigo, paresthesias, disulfiram-like effects. *Rare:* seizures,peripheral neuropathy	Warfarin: effect enhanced by metronidazole Disulfiram: psychotic reaction Phenobarbital, phenytoin: accelerate elimination of metronidazole Lithium: serum levels	B	Yes

			elevated by metronidazole Cimetidine: prolonged half-life of metronidazole		
Tinidazole	Amebiasis,[b] giardiasis, trichomoniasis	*Occasional:* nausea, vomiting, metallic taste	See metronidazole	C	Yes
Oxamniquine	Schistosomiasis	*Occasional:* dizziness, drowsiness, headache, orange urine, elevated aminotransferases. *Rare:* seizures	No major interactions	C	No information
Paromomycin	Amebiasis,[b] D. fragilis infection, giardiasis, cryptosporidiosis, leishmaniasis	*Frequent:* GI disturbances (oral dosing only). *Occasional:* nephrotoxicity, ototoxicity, vestibular toxicity (parenteral dosing only)	No major interactions	Oral: B Parenteral: not assigned[c]	No information
Pentamidine isethionate	Leishmaniasis, trypanosomiasis	*Frequent:* hypotension, hypoglycemia, pancreatitis, sterile abscesses at IM injection sites, GI disturbances, reversible renal failure. *Occasional:* hepatotoxicity, cardiotoxicity, delirium. *Rare:* anaphylaxis	No major interactions	C	No information
Piperazine and derivatives					
Piperazine	Ascariasis, enterobiasis	*Occasional:* nausea, vomiting, diarrhea, abdominal pain, headache. *Rare:* neurotoxicity, seizures	None reported	C	No information
Diethylcarbamazine[e]	Lymphatic filariasis, loiasis, tropical pulmonary eosinophilia	*Frequent:* dose-related nausea, vomiting. *Rare:* fever, chills, arthralgias, headaches	None reported	Not assigned[c]	No information
Praziquantel	Clonorchiasis,[b] cysticercosis, diphyllobothriasis, hymenolepiasis, taeniasis, opisthorchiasis, intestinal trematodes, paragonimiasis, schistosomiasis[b]	*Frequent:* abdominal pain, diarrhea, dizziness, headache, malaise. *Occasional:* fever, nausea. *Rare:* pruritus, singultus	No major interactions	B	Yes
Pyrantel pamoate	Ascariasis, eosinophilic enterocolitis, enterobiasis,[b] hookworm, trichostrongyliasis	*Occasional:* GI disturbances, headache, dizziness, elevated aminotransferases	No major interactions	C	No information
Quinacrine[h]	Giardiasis[b]	*Frequent:* headache, nausea, vomiting, bitter taste. *Occasional:* yellow-orange discoloration of skin, sclerae, urine; begins after 1 week of treatment and lasts up to 4 months after drug discontinuation. *Rare:* psychosis, exfoliative dermatitis, retinopathy, G6PD-induced hemolysis, exacerbation of psoriasis, disulfiram-like effects	Primaquine: toxicity potentiated by quinacrine	C	No information
Quinine and quinidine	Malaria, babesiosis	*Frequent:* cinchonism (tinnitus, high-tone deafness, headache, dysphoria, nausea, vomiting, abdominal pain, visual disturbances, postural hypotension), hyperinsulinemia resulting in life-threatening hypoglycemia. *Occasional:* deafness, hemolytic anemia, arrhythmias, hypotension due to rapid IV infusion	Carbonic-anhydrase inhibitors, thiazide diuretics: reduced renal elimination of quinidine Amiodarone, cimetidine: increased quinidine levels Nifedipine: decreased quinidine levels; quinidine slows metabolism of nifedipine Phenobarbital, phenytoin, rifampin: accelerated hepatic elimination of quinidine Verapamil: reduced hepatic clearance of quinidine Diltiazem: decreased clearance of quinidine	X	Yes[g]
Quinolones					
Ciprofloxacin	Cyclosporiasis, isosporiasis	*Occasional:* nausea, diarrhea, vomiting, abdominal pain/discomfort, headache, restlessness,	Probenecid: increased serum levels of ciprofloxacin Theophylline, warfarin: serum levels increased	C	Yes

		rash. *Rare:* myalgias/arthralgias, tendon rupture, CNS symptoms (nervousness, agitation, insomnia, anxiety, nightmares or paranoia); convulsions	serum levels increased by ciprofloxacin		
Suramin[e]	Trypanosomiasis	*Frequent:* immediate: fever, urticaria, nausea, vomiting, hypotension; delayed (up to 24 h): exfoliative dermatitis, stomatitis, paresthesias, photophobia, renal dysfunction. *Occasional:* nephrotoxicity, adrenal toxicity, optic atrophy, anaphylaxis	No major interactions	Not assigned	No information
Tetracyclines	Balantidiasis, *D. fragilis* infection, malaria; lymphatic filariasis (doxycycline)	*Frequent:* GI disturbances. *Occasional:* photosensitivity dermatitis. *Rare:* exfoliative dermatitis, esophagitis, hepatotoxicity	Warfarin: effect prolonged by tetracyclines	D	Yes

[a]Based on U.S. Food and Drug Administration pregnancy categories of A–D, X.

[b]Approved by the FDA for this indication.

[c]Use in pregnancy is recommended by international organizations outside the United States.

[d]Only AmBisome has been approved by the FDA for this indication.

[e]Available through the CDC.

[f]Only artemether (in combination with lumefantrine) and artesunate have been approved by the FDA for this indication.

[g]Not believed to be harmful.

[h]Available through the manufacturer. Abbreviations: ACTH, adrenocorticotropic hormone; AV, atrioventricular; CNS, central nervous system; ECG, electrocardiogram; G6PD, glucose 6-phosphate dehydrogenase; MAO, monoamine oxidase.

FURTHER READINGS

ABRAMOWICZ M (ed): Drugs for parasitic infections. Med Lett Drugs Ther 5:e1, 2007

HOTEZ PJ et al: Control of neglected tropical diseases. N Engl J Med 357:1018, 2007[PMID: 17804846]

KEISER J, UTZINGER J: Efficacy of current drugs against soil-transmitted helminth infections. JAMA 299:1937, 2008[PMID: 18430913]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PHARMACOLOGY OF AGENTS USED TO TREAT PARASITIC INFECTIONS: INTRODUCTION

This chapter deals exclusively with the pharmacologic properties of the agents used to treat infections due to parasites. Specific treatment recommendations for the parasitic diseases of humans are listed in the chapters on those diseases. Information on these agents' major toxicities, spectrum of activity, and safety for use during pregnancy and lactation is presented in Chap. 208. Many of the agents discussed herein are approved by the U.S. Food and Drug Administration (FDA) but are considered investigational for the treatment of certain infections (see Table 208-1). Drugs marked in the text with an asterisk (*) are available only through the Centers for Disease Control and Prevention (CDC) Drug Service (telephone: 404-639-3670 or 404-639-2888; www.cdc.gov/ncpdcid/dsr/). Drugs marked with a dagger (†) are available only through their manufacturers; contact information for these manufacturers may be available from the CDC.

ALBENDAZOLE

Like all benzimidazoles, albendazole acts by selectively binding to free β-tubulin in nematodes, inhibiting the polymerization of tubulin and the microtubule-dependent uptake of glucose. Irreversible damage occurs in gastrointestinal (GI) cells of the nematodes, resulting in starvation, death, and expulsion by the host. While highly injurious to nematodes, this fundamental disruption of cellular metabolism also offers treatment for a wide range of parasitic diseases.

Albendazole is poorly absorbed from the GI tract. Administration with a fatty meal increases its absorption by two- to sixfold. Poor absorption may be advantageous for the treatment of intestinal helminths, but successful treatment of tissue helminth infections (e.g., hydatid disease and neurocysticercosis) requires that a sufficient amount of active drug reach the site of infection. The metabolite albendazole sulfoxide is responsible for the drug's therapeutic effect outside the gut lumen. Albendazole sulfoxide crosses the blood-brain barrier, reaching a level significantly higher than that achieved in plasma. The high concentrations of albendazole sulfoxide attained in cerebrospinal fluid (CSF) probably explain the efficacy of albendazole in the treatment of neurocysticercosis.

Albendazole is extensively metabolized in the liver, but there are few data regarding the drug's use in patients with hepatic disease. Single-dose albendazole therapy in humans is largely without side effects (overall frequency, ≤1%). More prolonged courses (e.g., as administered for cystic and alveolar echinococcal disease) have been associated with liver function abnormalities and bone marrow toxicity. Thus, when prolonged use is anticipated, the drug should be administered in treatment cycles of 28 days interrupted by 14 days off therapy. Prolonged therapy with full-dose albendazole (800 mg/d) should be approached cautiously in patients also receiving drugs with known effects on the cytochrome P450 system.

AMODIAQUINE

Amodiaquine has been widely used in the treatment of malaria for >40 years. Like chloroquine (the other major 4-aminoquinoline), amodiaquine is now of limited use because of the spread of resistance. Amodiaquine interferes with hemozoin formation through complexation with heme. Although rapidly absorbed, amodiaquine behaves as a prodrug after oral administration, with the principal plasma metabolite monodesethylamodiaquine as the predominant antimalarial agent. Amodiaquine and its metabolites are all excreted in urine, but there are no recommendations concerning dosage adjustment in patients with impaired renal function. Severe adverse events can occur, albeit rarely (1 case in 2000 treatment courses), with amodiaquine administration. Agranulocytosis and hepatotoxicity can develop with repeated use; therefore, this drug should not be used for prophylaxis. Despite widespread resistance, amodiaquine has been shown to be effective in some areas when combined with other antimalarial drugs. Its use with artesunate in a fixed-dose combination with a soluble formulation creates an antimalarial specifically designed for children. Amodiaquine is not available in the United States.

AMPHOTERICIN B

See Table 208-1 and Chap. 198.

ANTIMONIALS*

Despite associated adverse reactions and the need for prolonged parenteral treatment, the pentavalent antimonial compounds (designated Sb^V) have remained the first-line therapy for all forms of leishmaniasis throughout the world, primarily because they are affordable, are effective, and have

survived the test of time. Although they have been used for almost 100 years, their mechanism of action against *Leishmania* species has only recently come to light. Pentavalent antimonials are active only after bioreduction to the trivalent Sb(III) form. This form inhibits trypanothione reductase, a critical enzyme involved in the oxidative stress management of *Leishmania* species. The fact that *Leishmania* species use trypanothione rather than glutathione (which is used by mammalian cells) may explain the parasite-specific activity of antimonials. The drugs are taken up by the reticuloendothelial system, and their activity against *Leishmania* species may be enhanced by this localization. Sodium stibogluconate is the only pentavalent antimonial available in the United States; meglumine antimoniate is principally used in francophone countries.

Resistance is a major problem in some areas. Although low-level unresponsiveness to Sb^V was identified in India in the 1970s, incremental increases in both the recommended daily dosage (to 20 mg/kg) and the duration of treatment (to 28 days) satisfactorily compensated for the growing resistance until around 1990. There has since been a steady erosion in the capacity of Sb^V to induce long-term cure in patients with kala-azar who live in eastern India. Foremost among the many factors that have probably contributed to this failure is the provision of suboptimal treatment for years, which led to the development of drug resistance among parasites. Co-infection with HIV impairs the treatment response.

Sodium stibogluconate is available in aqueous solution and is administered parenterally. Antimony appears to have two elimination phases. When administered IV, the mean half-life of the first phase is <2 h; the mean half-life of the terminal elimination phase is nearly 36 h. This slower phase may be due to conversion of pentavalent antimony to a trivalent form that is the likely cause of the side effects often seen with prolonged therapy.

ARTEMISININ DERIVATIVES*

Artesunate, artemether, arteether, and the parent compound artemisinin are sesquiterpene lactones derived from the wormwood plant *Artemisia annua*. These agents are at least tenfold more potent in vivo than other antimalarial drugs and presently show no cross-resistance with known antimalarial drugs; thus, they have become first-line agents for the treatment of severe falciparum malaria. The artemisinin compounds are rapidly effective against the asexual blood forms of *Plasmodium* species but are not active against intrahepatic forms. Artemisinin and its derivatives are highly lipid soluble and readily cross both host and parasite cell membranes. One factor that explains the drugs' highly selective toxicity against malaria is that parasitized erythrocytes concentrate artemisinin and its derivatives to concentrations one hundredfold higher than those in uninfected erythrocytes. The antimalarial effect of these agents results primarily from dihydroartemisinin, a compound to which artemether and artesunate are both converted. In the presence of heme or molecular iron, the endoperoxide moiety of dihydroartemisinin decomposes, generating free radicals and other metabolites that damage parasite proteins. The compounds are available for oral, rectal, IV, or IM administration, depending on the derivative. In the United States, IV artesunate is available for the treatment of severe, quinidine-unresponsive malaria through the CDC malaria hotline (770-488-7788, M-F, 0800-1630 EST; 770-488-7100 after hours). Artemisinin and its derivatives are cleared rapidly from the circulation. Their short half-lives limit their value for prophylaxis and monotherapy. These agents should be used only in combination with another, longer-acting agent (e.g., artesunate-mefloquine, dihydroartemisinin-piperaquine). A combined formulation of artemether and lumefantrine is now available for the treatment of acute uncomplicated falciparum malaria acquired in areas where *Plasmodium falciparum* is resistant to chloroquine and antifolates.

ATOVAQUONE

Atovaquone is a hydroxynaphthoquinone that exerts broad-spectrum antiprotozoal activity via selective inhibition of parasite mitochondrial electron transport. This agent exhibits potent activity against toxoplasmosis and babesiosis when used with pyrimethamine and azithromycin, respectively. Atovaquone possesses a novel mode of action against *Plasmodium* species, inhibiting the electron transport system at the level of the cytochrome bc1 complex. The drug is active against both the erythrocytic and the exoerythrocytic stages of *Plasmodium* species; however, because it does not eradicate hypnozoites from the liver, patients with *Plasmodium vivax* or *Plasmodium ovale* infections must be given radical prophylaxis.

Malarone is a fixed-dose combination of atovaquone and proguanil used for malaria prophylaxis as well as for the treatment of acute, uncomplicated *P. falciparum* malaria. Malarone has been shown to be effective in regions with multidrug-resistant *P. falciparum*. Resistance to atovaquone has yet to be reported.

The bioavailability of atovaquone varies considerably. Absorption after a single oral dose is slow, increases two- to threefold with a fatty meal, and is dose-limited above 750 mg. The elimination half-life is increased in patients with moderate hepatic impairment. Because of the potential for drug accumulation, the use of atovaquone is generally contraindicated in persons with a creatinine clearance rate <30 mL/min. No dosage adjustments are needed in patients with mild to moderate renal

impairment.

AZITHROMYCIN

See Table 208-1 and Chap. 133.

AZOLES

See Table 208-1 and Chap. 198.

BENZNIDAZOLE

This oral nitroimidazole derivative is used to treat Chagas' disease, with cure rates of 80–90% recorded in acute infections. Benznidazole is believed to exert its trypanocidal effects by generating oxygen radicals to which the parasites are more sensitive than mammalian cells because of a relative deficiency in antioxidant enzymes. Benznidazole also appears to alter the balance between pro- and anti-inflammatory mediators by downregulating the synthesis of nitrite, interleukin (IL) 6, and IL-10 in macrophages. Benznidazole is highly lipophilic and readily absorbed. The drug is extensively metabolized; only 5% of the dose is excreted unchanged in the urine. Benznidazole is well tolerated; adverse effects are rare and usually manifest as gastrointestinal upset or pruritic rash.

BITHIONOL*

Bithionol is a chlorinated bisphenol with activity against trematodes. *Fasciola hepatica* uses fumarate reduction coupled to rhodoquinone for anaerobic energy metabolism. Bithionol competitively inhibits electron transfer to fumarate by rhodoquinone; the result is impaired anaerobic energy metabolism and trematode death. Bithionol is parasite specific for two reasons: (1) fumarate reductase catalyzes the reverse of the reaction of mammalian succinic dehydrogenase in the Krebs cycle, and (2) the rhodoquinone respiratory chain link is unique to the parasite. In the mammalian respiratory chain, the quinone electron carrier is ubiquinone. Bithionol is readily absorbed from the GI tract. It is no longer produced, but limited supplies are available from the CDC.

CHLOROQUINE

This 4-aminoquinoline has marked, rapid schizonticidal and gametocidal activity against blood forms of *P. ovale* and *Plasmodium malariae* and against susceptible strains of *P. vivax* and *P. falciparum*. It is not active against intrahepatic forms (*P. vivax* and *P. ovale*). Parasitized erythrocytes accumulate chloroquine in significantly greater concentrations than do normal erythrocytes. Chloroquine, a weak base, concentrates in the food vacuoles of intraerythrocytic parasites because of a relative pH gradient between the extracellular space and the acidic food vacuole. Once it enters the acidic food vacuole, chloroquine is rapidly converted to a membrane-impermeable protonated form and is trapped. Continued accumulation of chloroquine in the parasite's acidic food vacuoles results in drug levels that are 600-fold higher at this site than in plasma. The high accumulation of chloroquine results in an increase in pH within the food vacuole to a level above that required for the acid proteases' optimal activity, inhibiting parasite heme polymerase; as a result, the parasite is effectively killed with its own metabolic waste. Compared with susceptible strains, chloroquine-resistant plasmodia transport chloroquine out of intraparasitic compartments more rapidly and maintain lower chloroquine concentrations in their acid vesicles. Hydroxychloroquine, a congener of chloroquine, is equivalent to chloroquine in its antimalarial efficacy but is preferred to chloroquine for the treatment of autoimmune disorders because it produces less ocular toxicity when used in high doses.

Chloroquine is well absorbed. However, because it exhibits extensive tissue binding, a loading dose is required to yield effective plasma concentrations. A therapeutic drug level in plasma is reached 2–3 h after oral administration (the preferred route). Chloroquine can be administered IV, but excessively rapid parenteral administration can result in seizures and death from cardiovascular collapse. The mean half-life of chloroquine is 4 days, but the rate of excretion decreases as plasma levels decline, making once-weekly administration possible for prophylaxis in areas with sensitive strains. About one-half of the parent drug is excreted in urine, but the dose should not be reduced for persons with acute malaria and renal insufficiency.

CIPROFLOXACIN

Table 208-1 and Chap. 133.

CLINDAMYCIN

See Table 208-1 and Chap. 133.

DAPSONE

See Table 208-1 and Chap. 168.

DEHYDROEMETINE

Emetine is an alkaloid derived from ipecac; dehydroemetine is synthetically derived from emetine and is considered less toxic. Both agents are active against *Entamoeba histolytica* and appear to work by blocking peptide elongation and thus inhibiting protein synthesis. Emetine is rapidly absorbed after parenteral administration, rapidly distributed throughout the body, and slowly excreted in the urine in

unchanged form. Both agents are contraindicated in patients with renal disease.

DIETHYLCARBAMAZINE*

A derivative of the antihelminthic agent piperazine with a long history of successful use, diethylcarbamazine (DEC) remains the treatment of choice for lymphatic filariasis and loiasis and has also been used for visceral larva migrans. While piperazine itself has no antifilarial activity, the piperazine ring of DEC is essential for the drug's activity. DEC's mechanism of action remains to be fully defined. Proposed mechanisms include immobilization due to inhibition of parasite cholinergic muscle receptors, disruption of microtubule formation, and alteration of helminthic surface membranes resulting in enhanced killing by the host's immune system. DEC enhances adherence properties of eosinophils. The development of resistance under drug pressure (i.e., a progressive decrease in efficacy when the drug is used widely in human populations) has not been observed, although DEC has variable effects when administered to persons with filariasis. Monthly administration provides effective prophylaxis against both bancroftian filariasis and loiasis.

DEC is well absorbed after oral administration, with peak plasma concentrations reached within 1–2 h. No parenteral form is available. The drug is eliminated largely by renal excretion, with <5% found in feces. If more than one dose is to be administered to an individual with renal dysfunction, the dose should be reduced commensurate with the reduction in creatinine clearance rate. Alkalinization of the urine prevents renal excretion and increases the half-life of DEC. Use in patients with onchocerciasis can precipitate a Mazzotti reaction, with pruritus, fever, and arthralgias. Like other piperazines, DEC is active against *Ascaris* species. Patients co-infected with this nematode may expel live worms after treatment.

DILOXANIDE FUROATE

Diloxanide furoate, a substituted acetanilide, is a luminally active agent used to eradicate the cysts of *E. histolytica*. After ingestion, diloxanide furoate is hydrolyzed by enzymes in the lumen or mucosa of the intestine, releasing furoic acid and the ester diloxanide; the latter acts directly as an amebicide.

Diloxanide furoate is given alone to asymptomatic cyst passers. For patients with active amebic infections, diloxanide is generally administered in combination with a 5-nitroimidazole such as metronidazole or tinidazole. Diloxanide furoate is rapidly absorbed after oral administration. When coadministered with a 5-nitroimidazole, only diloxanide appears in the systemic circulation; levels peak within 1 h and disappear within 6 h. About 90% of an oral dose is excreted in the urine within 48 h, chiefly as the glucuronide metabolite. Diloxanide furoate is contraindicated in pregnant and breast-feeding women and in children <2 years of age.

EFLORNITHINE†

Eflornithine (difluoromethylornithine, or DFMO) is a fluorinated analogue of the amino acidornithine. Although originally designed as an antineoplastic agent, eflornithine has proven effective against some trypanosomatids. At one point, the production of this effective agent ceased despite the increasing incidence of human African trypanosomiasis; however, production resumed after eflornithine was discovered to be an effective cosmetic depilatory agent.

Eflornithine has specific activity against all stages of infection with *Trypanosoma brucei gambiense*; however, it is inactive against *T. b. rhodesiense*. The drug acts as an irreversible suicide inhibitor of ornithine decarboxylase, the first enzyme in the biosynthesis of the polyamines putrescine and spermidine. Polyamines are essential for the synthesis of trypanothione, an enzyme required for the maintenance of intracellular thiols in the correct redox state and for the removal of reactive oxygen metabolites. However, polyamines are also essential for cell division in eukaryotes, and ornithine decarboxylase is similar in trypanosomes and mammals. The selective antiparasitic activity of eflornithine is partly explained by the structure of the trypanosomal enzyme, which lacks a 36-amino-acid C-terminal sequence found on mammalian ornithine decarboxylase. This difference results in a lower turnover of ornithine decarboxylase and a more rapid decrease of polyamines in trypanosomes than in the mammalian host. The diminished effectiveness of eflornithine against *T. b. rhodesiense* appears to be due to the parasite's ability to replace the inhibited enzyme more rapidly than *T. b. gambiense*.

Eflornithine is less toxic but more costly than conventional therapy. It can be administered IV or PO. The dose should be reduced in renal failure. Eflornithine readily crosses the blood-brain barrier; CSF levels are highest in persons with the most severe central nervous system (CNS) involvement.

FUMAGILLIN†

Fumagillin is a water-insoluble antibiotic that is derived from the fungus *Aspergillus fumigatus* and is active against microsporidia. This drug is effective when used topically to treat ocular infections due to *Encephalitozoon* species. When given systemically, fumagillin was effective but caused thrombocytopenia in all recipients in the second week of treatment; this side effect was readily reversed when administration of the drug was stopped. The mechanisms by which fumagillin inhibits microsporidial replication are poorly understood, although the drug may inhibit methionine aminopeptidase 2 by irreversibly blocking the active site.

FURAZOLIDONE

This nitrofuran derivative is an effective alternative agent for the treatment of giardiasis and also exhibits activity against *Isospora belli*. Since it is the only agent active against *Giardia* that is available in liquid form, it is often used to treat young children. Furazolidone undergoes reductive activation in *Giardia lamblia* trophozoites—an event that, unlike the reductive activation of metronidazole, involves an NADH oxidase. The killing effect correlates with the toxicity of reduced products, which damage important cellular components, including DNA. Although furazolidone had been thought to be largely unabsorbed when administered orally, the occurrence of systemic adverse reactions indicates that this is not the case. More than 65% of the drug dose can be recovered from the urine as colored metabolites. Omeprazole reduces the oral bioavailability of furazolidone.

Furazolidone is a monoamine oxidase (MAO) inhibitor; thus caution should be used in its concomitant administration with other drugs (especially indirectly acting sympathomimetic amines) and in the consumption of food and drink containing tyramine during treatment. However, hypertensive crises have not been reported in patients receiving furazolidone, and it has been suggested that—since furazolidone inhibits MAOs gradually over several days—the risks are small if treatment is limited to a 5-day course. Because hemolytic anemia can occur in patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency and glutathione instability, furazolidone treatment is contraindicated in mothers who are breast-feeding and in neonates.

HALOFANTRINE

This 9-phenanthrenemethanol is one of three classes of arylaminoalcohols first identified as potential antimalarial agents by the World War II Malaria Chemotherapy Program. Its activity is believed to be similar to that of chloroquine, although it is an oral alternative for the treatment of malaria due to chloroquine-resistant *P. falciparum*. Although the mechanism of action is poorly understood, halofantrine is thought to share mechanism(s) with the 4-aminoquinolines, forming a complex with ferriprotoporphyrin IX and interfering with the degradation of hemoglobin.

Halofantrine exhibits erratic bioavailability, but its absorption is significantly enhanced when it is taken with a fatty meal. The elimination half-life of halofantrine is 1–2 days; it is excreted mainly in feces. Halofantrine is metabolized into *N*-debutyl-halofantrine by the cytochrome P450 enzyme CYP3A4. Grapefruit juice should be avoided during treatment because it increases both halofantrine's bioavailability and halofantrine-induced QT interval prolongation by inhibiting CYP3A4 at the enterocyte level.

IODOQUINOL

Iodoquinol (diiodohydroxyquin), a hydroxyquinoline, is an effective luminal agent for the treatment of amebiasis, balantidiasis, and infection with *Dientamoeba fragilis*. Its mechanism of action is unknown. It is poorly absorbed. Because the drug contains 64% organically bound iodine, it should be used with caution in patients with thyroid disease. Iodine dermatitis occurs occasionally during iodoquinol treatment. Protein-bound serum iodine levels may be increased during treatment and can interfere with certain tests of thyroid function. These effects may persist for as long as 6 months after discontinuation of therapy. Iodoquinol is contraindicated in patients with liver disease. Most serious are the reactions related to prolonged high-dose therapy (optic neuritis, peripheral neuropathy), which should not occur if the recommended dosage regimens are followed.

IVERMECTIN

Ivermectin (22,23-dihydroavermectin) is a derivative of the macrocyclic lactone avermectin produced by the soil-dwelling actinomycete *Streptomyces avermitilis*. Ivermectin is active at low doses against a wide range of helminths and ectoparasites. It is the drug of choice for the treatment of onchocerciasis, strongyloidiasis, cutaneous larva migrans, and scabies. Ivermectin is highly active against microfilariae of the lymphatic filariases but has no macrofilaricidal activity. When ivermectin is used in combination with other agents such as DEC or albendazole for treatment of lymphatic filariasis, synergistic activity is seen. While active against the intestinal helminths *Ascaris lumbricoides* and *Enterobius vermicularis*, ivermectin is only variably effective in trichuriasis and is ineffective against hookworms. Widespread use of ivermectin for treatment of intestinal nematode infections in sheep and goats has led to the emergence of drug resistance in veterinary practice; this development may portend problems in human medical use.

Data suggest that ivermectin acts by opening the neuromuscular membrane-associated, glutamate-dependent chloride channels. The influx of chloride ions results in hyperpolarization and muscle paralysis—particularly of the nematode pharynx, with consequent blockage of the oral ingestion of nutrients. Because these chloride channels are present only in invertebrates, the paralysis is seen only in the parasite.

Ivermectin is available for administration to humans only as an oral formulation. The drug is highly protein bound; it is almost completely excreted in feces. Both food and beer increase the bioavailability of ivermectin significantly. Ivermectin is distributed widely throughout the body; animal studies indicate that it accumulates at the highest concentration in adipose tissue and liver, with little accumulation in

the brain. Few data exist to guide therapy in hosts with conditions that may influence drug pharmacokinetics.

Ivermectin is generally administered as a single dose of 150–200 μg/kg. In the absence of parasitic infection, the adverse effects of ivermectin in therapeutic doses are minimal. Adverse effects in patients with filarial infections include fever, myalgia, malaise, lightheadedness, and (occasionally) postural hypotension. The severity of such side effects is related to the intensity of parasite infection, with more symptoms in individuals with a heavy parasite burden. In onchocerciasis, skin edema, pruritus, and mild eye irritation may also occur. The adverse effects are generally self-limiting and only occasionally require symptom-based treatment with antipyretics or antihistamines. More severe complications of ivermectin therapy for onchocerciasis include encephalopathy in patients heavily infected with *Loa loa*.

LUMEFANTRINE

Lumefantrine (benflumetol), a fluorene arylaminoalcohol derivative synthesized in the 1970s by the Chinese Academy of Military Medical Sciences (Beijing), has marked blood schizonticidal activity against a wide range of plasmodia. This agent conforms structurally and in mode of action to other arylaminoalcohols (quinine, mefloquine, and halofantrine). Lumefantrine exerts its antimalarial effect as a consequence of its interaction with heme, a degradation product of hemoglobin metabolism. Although its antimalarial activity is slower than that of the artemisinin-based drugs, the recrudescence rate with the recommended lumefantrine regimen is lower. The pharmacokinetic properties of lumefantrine are reminiscent of those of halofantrine, with variable oral bioavailability, considerable augmentation of oral bioavailability by concomitant fat intake, and a terminal elimination half-life of ~4–5 days in patients with malaria.

Artemether and lumefantrine have synergistic activity, and clinical studies of several hundred patients in China show the combination to be safe and well tolerated. The combined formulation of artemether and lumefantrine has been developed for the treatment of falciparum malaria in areas where *P. falciparum* is resistant to chloroquine and antifolates. This combination has now been approved by the FDA.

MEBENDAZOLE

This benzimidazole is a broad-spectrum antiparasitic agent widely used to treat intestinal helminthiases. Its mechanism of action is similar to that of albendazole; however, it is a more potent inhibitor of parasite malic dehydrogenase and exhibits a more specific and selective effect against intestinal nematodes than the other benzimidazoles.

Mebendazole is available only in oral form but is poorly absorbed from the GI tract; only 5–10% of a standard dose is measurable in plasma. The proportion absorbed from the GI tract is extensively metabolized in the liver. Metabolites appear in the urine and bile; impaired liver or biliary function results in higher plasma mebendazole levels in treated patients. No dose reduction is warranted in patients with renal function impairment. Because mebendazole is poorly absorbed, its incidence of side effects is low. Transient abdominal pain and diarrhea sometimes occur, usually in persons with massive parasite burdens.

MEFLOQUINE

Mefloquine is the preferred drug for prophylaxis of chloroquine-resistant malaria; high doses can be used for treatment. Despite the development of drug-resistant strains of *P. falciparum* in parts of Africa and southeast Asia, mefloquine is an effective drug throughout most of the world. Cross-resistance of mefloquine with halofantrine and with quinine has been documented in limited areas. Like quinine and chloroquine, this quinoline is active only against the asexual erythrocytic stages of malarial parasites. Unlike quinine, however, mefloquine has a relatively poor affinity for DNA and, as a result, does not inhibit the synthesis of parasitic nucleic acids and proteins. Although both mefloquine and chloroquine inhibit hemozoin formation and heme degradation, mefloquine differs in that it forms a complex with heme that may be toxic to the parasite.

Mefloquine HCl is poorly water soluble and intensely irritating when given parenterally; thus it is available only in tablet form. Its absorption is adversely affected by vomiting and diarrhea but is significantly enhanced when the drug is administered with or after food. About 98% of the drug binds to protein. Mefloquine is excreted mainly in the bile and feces; therefore, no dose adjustment is needed in persons with renal insufficiency. The drug and its main metabolite are not appreciably removed by hemodialysis. No special chemoprophylactic dosage adjustments are indicated for the achievement of plasma concentrations in dialysis patients that are similar to those in healthy persons. Pharmacokinetic differences have been detected among various ethnic populations. In practice, however, these distinctions are of minor importance compared with host immune status and parasite sensitivity. In patients with impaired liver function, the elimination of mefloquine may be prolonged, leading to higher plasma levels.

Mefloquine should be used with caution by individuals participating in activities requiring alertness and fine-motor coordination. If the drug is to be administered for a prolonged period, periodic evaluations

are recommended, including liver function tests and ophthalmic examinations. Sleep abnormalities (insomnia, abnormal dreams) have occasionally been reported. Psychosis and seizures occur rarely; mefloquine should not be prescribed to patients with neuropsychiatric conditions, including depression, generalized anxiety disorder, psychosis, schizophrenia, and seizure disorder. If acute anxiety, depression, restlessness, or confusion develops during prophylaxis, these psychiatric symptoms may be considered prodromal to a more serious event, and the drug should be discontinued.

Concomitant use of quinine, quinidine, or drugs producing ß-adrenergic blockade may cause significant electrocardiographic abnormalities or cardiac arrest. Halofantrine must not be given simultaneously with or <3 weeks after mefloquine because a potentially fatal prolongation of the QTc interval on electrocardiography may occur. No data exist on mefloquine use after halofantrine use. Administration of mefloquine with quinine or chloroquine may increase the risk of convulsions. Mefloquine may lower plasma levels of anticonvulsants. Caution should be exercised with regard to concomitant antiretroviral therapy, since mefloquine has been shown to exert variable effects on ritonavir pharmacokinetics that are not explained by hepatic CYP3A4 activity or ritonavir protein binding. Vaccinations with attenuated live bacteria should be completed at least 3 days before the first dose of mefloquine.

Women of childbearing age who are traveling to areas where malaria is endemic should be warned against becoming pregnant and encouraged to practice contraception during malaria prophylaxis with mefloquine and for up to 3 months thereafter. However, in the case of unplanned pregnancy, use of mefloquine is not considered an indication for pregnancy termination.

MELARSOPROL*

Melarsoprol has been used since 1949 for the treatment of human African trypanosomiasis. This trivalent arsenical compound is indicated for the treatment of African trypanosomiasis with neurologic involvement and for the treatment of early disease that is resistant to suramin or pentamidine. Melarsoprol, like other drugs containing heavy metals, interacts with thiol groups of several different proteins; however, its antiparasitic effects appear to be more specific. Trypanothione reductase is a key enzyme involved in the oxidative stress management of both *Trypanosoma* and *Leishmania* species, helping to maintain an intracellular reducing environment by reduction of disulfide trypanothione to its dithiol derivative dihydrotrypanothione. Melarsoprol sequesters dihydrotrypanothione, depriving the parasite of its main sulfhydryl antioxidant, and inhibits trypanothione reductase, depriving the parasite of the essential enzyme system that is responsible for keeping trypanothione reduced. These effects are synergistic. The selectivity of arsenical action against trypanosomes is due at least in part to the greater melarsoprol affinity of reduced trypanothione than of other monothiols (e.g., cysteine) on which the mammalian host depends for maintenance of high thiol levels. Melarsoprol enters the parasite via an adenosine transporter; drug-resistant strains lack this transport system.

Melarsoprol is always administered IV. A small but therapeutically significant amount of the drug enters the CSF. The compound is excreted rapidly, with ~80% of the arsenic found in feces.

Melarsoprol is highly toxic. The most serious adverse reaction is reactive encephalopathy, which affects 6% of treated individuals and usually develops within 4 days of the start of therapy, with an average case-fatality rate of 50%. Glucocorticoids are administered with melarsoprol to prevent this development. Because melarsoprol is intensely irritating, care must be taken to avoid infiltration of the drug.

METRIFONATE

Metrifonate has selective activity against *Schistosoma haematobium*. This organophosphorous compound is a prodrug that is converted nonenzymatically to dichlorvos (2,2-dichlorovinyl dimethylphosphate, DDVP), a highly active chemical that irreversibly inhibits the acetylcholinesterase enzyme. Schistosomal cholinesterase is more susceptible to dichlorvos than is the corresponding human enzyme. The exact mechanism of action of metrifonate is uncertain, but the drug is believed to inhibit tegumental acetylcholine receptors that mediate glucose transport.

Metrifonate is administered in a series of three doses at 2-week intervals. After a single oral dose, metrifonate produces a 95% decrease in plasma cholinesterase activity within 6 h, with a fairly rapid return to normal. However, 2.5 months are required for erythrocyte cholinesterase levels to return to normal. Treated persons should not be exposed to neuromuscular blocking agents or organophosphate insecticides for at least 48 h after treatment.

METRONIDAZOLE AND OTHER NITROIMIDAZOLES

See Table 208-1 and Chap. 133.

MILTEFOSINE

In the early 1990s, miltefosine (hexadecylphosphocholine), originally developed as an antineoplastic agent, was discovered to have significant antiproliferative activity against *Leishmania* species, *Trypanosoma cruzi*, and *T. brucei* parasites in vitro and in experimental animal models. Miltefosine is the first oral drug that has proved to be highly effective and comparable to amphotericin B against

visceral leishmaniasis in India, where antimonial-resistant cases are prevalent. Miltefosine is also effective in previously untreated visceral infections. Cure rates in cutaneous leishmaniasis are comparable to those obtained with antimony.

The activity of miltefosine is attributed to interaction with cell signal transduction pathways and inhibition of phospholipid and sterol biosynthesis. Resistance to miltefosine has not been observed clinically. The drug is readily absorbed from the GI tract, is widely distributed, and accumulates in several tissues. The efficacy of a 28-day treatment course in Indian visceral leishmaniasis is equivalent to that of amphotericin B therapy; however, it appears that a shortened course of 21 days may be equally efficacious.

General recommendations for the use of miltefosine are limited by the exclusion of specific groups from the published clinical trials: persons <12 or >65 years of age, persons with the most advanced disease, breast-feeding women, HIV-infected patients, and individuals with significant renal or hepatic insufficiency.

NICLOSAMIDE

Niclosamide is active against a wide variety of adult tapeworms but not against tissue cestodes. It is also a molluscicide and is used in snail-control programs. The drug uncouples oxidative phosphorylation in parasite mitochondria, thereby blocking the uptake of glucose by the intestinal tapeworm and resulting in the parasite's death. Niclosamide rapidly causes spastic paralysis of intestinal cestodes in vitro. Its use is limited by its side effects, the necessarily long duration of therapy, the recommended use of purgatives, and—most important—limited availability (i.e., on a named-patient basis from the manufacturer).

Niclosamide is poorly absorbed. Tablets are given on an empty stomach in the morning after a liquid meal the night before, and this dose is followed by another 1 h later. For treatment of hymenolepiasis, the drug is administered for 7 days. A second course is often prescribed. The scolex and proximal segments of the tapeworms are killed on contact with niclosamide and may be digested in the gut. However, disintegration of the adult tapeworm results in the release of viable ova, which theoretically can result in autoinfection. Although fears of the development of cysticercosis in patients with *Taenia solium* infections have proved unfounded, it is still recommended that a brisk purgative be given 2 h after the first dose.

NIFURTIMOX*

This nitrofuran compound is an inexpensive and effective oral agent for the treatment of acute Chagas' disease. Trypanosomes lack catalase and have very low levels of peroxidase; as a result, they are very vulnerable to by-products of oxygen reduction. When nifurtimox is reduced in the trypanosome, a nitro anion radical is formed and undergoes autooxidation, resulting in the generation of the superoxide anion O_2^-, hydrogen peroxide (H_2O_2), hydroperoxyl radical (HO_2), and other highly reactive and cytotoxic molecules. Despite the abundance of catalases, peroxidases, and superoxide dismutases that neutralize these destructive radicals in mammalian cells, nifurtimox has a poor therapeutic index. Prolonged use is required, but the course may have to be interrupted because of drug toxicity, which develops in 40–70% of recipients. Nifurtimox is well absorbed and undergoes rapid and extensive biotransformation; <0.5% of the original drug is excreted in urine.

NITAZOXANIDE

Nitazoxanide is a 5-nitrothiazole compound used for the treatment of cryptosporidiosis and giardiasis; it is active against other intestinal protozoa as well. The drug is approved for use in children 1–11 years of age.

The antiprotozoal activity of nitazoxanide is believed to be due to interference with the pyruvate-ferredoxin oxidoreductase (PFOR) enzyme–dependent electron transfer reaction that is essential to anaerobic energy metabolism. Studies have shown that the PFOR enzyme from *G. lamblia* directly reduces nitazoxanide by transfer of electrons in the absence of ferredoxin. The DNA-derived PFOR protein sequence of *Cryptosporidium parvum* appears to be similar to that of *G. lamblia*. Interference with the PFOR enzyme–dependent electron transfer reaction may not be the only pathway by which nitazoxanide exerts antiprotozoal activity.

After oral administration, nitazoxanide is rapidly hydrolyzed to an active metabolite, tizoxanide (desacetyl-nitazoxanide). Tizoxanide then undergoes conjugation, primarily by glucuronidation. It is recommended that nitazoxanide be taken with food; however, no studies have been conducted to determine whether the pharmacokinetics of tizoxanide and tizoxanide glucuronide differ in fasted versus fed subjects. Tizoxanide is excreted in urine, bile, and feces, and tizoxanide glucuronide is excreted in urine and bile. The pharmacokinetics of nitazoxanide in patients with impaired hepatic and/or renal function have not been studied. Tizoxanide is highly bound to plasma protein (>99.9%). Therefore, caution should be used when administering this agent concurrently with other highly plasma protein–bound drugs with narrow therapeutic indices, as competition for binding sites may occur.

OXAMNIQUINE

This tetrahydroquinoline derivative is an effective alternative agent for the treatment of *Schistosoma mansoni*, although susceptibility to this drug exhibits regional variation. Oxamniquine exhibits anticholinergic properties, but its primary mode of action seems to rely on ATP-dependent enzymatic drug activation generating an intermediate that alkylates essential macromolecules, including DNA. In treated adult schistosomes, oxamniquine produces marked tegumental alterations that are similar to those seen with praziquantel but that develop less rapidly, becoming evident 4–8 days after treatment.

Oxamniquine is administered orally as a single dose and is well absorbed. Food retards absorption and reduces bioavailability. About 70% of an administered dose is excreted in urine as a mixture of pharmacologically inactive metabolites. Patients should be warned that their urine might have an intense orange-red color. Side effects are uncommon and usually mild, although hallucinations and seizures have been reported.

PAROMOMYCIN (AMINOSIDINE)

First isolated in 1956, this aminoglycoside is an effective oral agent for the treatment of infections due to intestinal protozoa. Parenteral paromomycin appears to be effective against visceral leishmaniasis in India.

Paromomycin inhibits protozoan protein synthesis by binding to the 30S ribosomal RNA in the aminoacyl-tRNA site, causing misreading of mRNA codons. Paromomycin is less active against *G. lamblia* than standard agents; however, like other aminoglycosides, paromomycin is poorly absorbed from the intestinal lumen, and the high levels of drug in the gut compensate for this relatively weak activity. If absorbed or administered systemically, paromomycin can cause ototoxicity and nephrotoxicity. However, systemic absorption is very limited, and toxicity should not be a concern in persons with normal kidneys. Topical formulations are not generally available.

PENTAMIDINE ISETHIONATE

This diamidine is an effective alternative agent for some forms of leishmaniasis and trypanosomiasis. It is available for parenteral and aerosolized administration. While its mechanism of action remains undefined, it is known to exert a wide range of effects, including interaction with trypanosomal kinetoplast DNA; interference with polyamine synthesis by a decrease in the activity of ornithine decarboxylase; and inhibition of RNA polymerase, ribosomal function, and the synthesis of nucleic acids and proteins.

Pentamidine isethionate is well absorbed, is highly tissue bound, and is excreted slowly over several weeks, with an elimination half-life of 12 days. No steady-state plasma concentration is attained in persons given daily injections; the result is extensive accumulation of pentamidine in tissues, primarily the liver, kidney, adrenal, and spleen. Pentamidine does not penetrate well into the CNS. Pulmonary concentrations of pentamidine are increased when the drug is delivered in aerosolized form.

PIPERAZINE

The antihelminthic activity of piperazine is confined to ascariasis and enterobiasis. Piperazine acts as an agonist at extrasynaptic γ-aminobutyric acid (GABA) receptors, causing an influx of chloride ions in the nematode somatic musculature. Although the initial result is hyperpolarization of the muscle fibers, the ultimate effect is flaccid paralysis leading to the expulsion of live worms. Patients should be warned, as this occurrence can be unsettling.

PRAZIQUANTEL

This heterocyclic pyrazinoisoquinoline derivative is highly active against a broad spectrum of trematodes and cestodes. It is the mainstay of treatment for schistosomiasis and is a critical part of community-based control programs.

All of the effects of praziquantel can be attributed either directly or indirectly to an alteration of intracellular calcium concentrations. Although the exact mechanism of action remains unclear, the major mechanism is disruption of the parasite tegument, causing tetanic contractures with loss of adherence to host tissues and, ultimately, disintegration or expulsion. Praziquantel induces changes in the antigenicity of the parasite by causing the exposure of concealed antigens. Praziquantel also produces alterations in schistosomal glucose metabolism, including decreases in glucose uptake, lactate release, glycogen content, and ATP levels.

Praziquantel exerts its parasitic effects directly and does not need to be metabolized to be effective. It is well absorbed but undergoes extensive first-pass hepatic clearance. Levels of the drug are increased when it is taken with food, particularly carbohydrates, or with cimetidine. Serum levels are reduced by glucocorticoids, chloroquine, carbamazepine, and phenytoin. Praziquantel is completely metabolized in humans, with 80% of the dose recovered as metabolites in urine within 4 days. It is not known to what extent praziquantel crosses the placenta, but retrospective studies suggest that it is safe in pregnancy.

Patients with schistosomiasis who have heavy parasite burdens may develop abdominal discomfort, nausea, headache, dizziness, and drowsiness. Symptoms begin 30 min after ingestion, may require spasmolytics for relief, and usually disappear spontaneously after a few hours.

PRIMAQUINE PHOSPHATE

Primaquine, an 8-aminoquinoline, has a broad spectrum of activity against all stages of plasmodial development in humans but has been used most effectively for eradication of the hepatic stage of these parasites. Despite its toxicity, it remains the drug of choice for radical cure of *P. vivax* infections. Primaquine must be metabolized by the host to be effective. It is, in fact, rapidly metabolized; only a small fraction of the dose of the parent drug is excreted unchanged. Although the parasiticidal activity of the three oxidative metabolites remains unclear, they are believed to affect both pyrimidine synthesis and the mitochondrial electron transport chain. The metabolites appear to have significantly less antimalarial activity than primaquine; however, their hemolytic activity is greater than that of the parent drug.

Primaquine causes marked hypotension after parenteral administration and therefore is given only by the oral route. It is rapidly and almost completely absorbed from the GI tract.

Patients should be tested for G6PD deficiency before they receive primaquine. The drug may induce the oxidation of hemoglobin into methemoglobin, irrespective of the G6PD status of the patient. Primaquine is otherwise well tolerated.

PROGUANIL (CHLOROGUANIDE)

Proguanil inhibits plasmodial dihydrofolate reductase and is used with atovaquone for oral treatment of uncomplicated malaria or with chloroquine for malaria prophylaxis in parts of Africa without widespread chloroquine-resistant *P. falciparum*.

Proguanil exerts its effect primarily by means of the metabolite cycloguanil, whose inhibition of dihydrofolate reductase in the parasite disrupts deoxythymidylate synthesis, thus interfering with a key pathway involved in the biosynthesis of pyrimidines required for nucleic acid replication. There are no clinical data indicating that folate supplementation diminishes drug efficacy; women of childbearing age for whom atovaquone/proguanil is prescribed should continue taking folate supplements to prevent neural tube birth defects.

Proguanil is extensively absorbed regardless of food intake. The drug is 75% protein bound. The main routes of elimination are hepatic biotransformation and renal excretion; 40–60% of the proguanil dose is excreted by the kidneys. Drug levels are increased and elimination is impaired in patients with hepatic insufficiency.

PYRANTEL PAMOATE

Pyrantel is a tetrahydropyrimidine formulated as pamoate. This safe, well-tolerated, inexpensive drug is used to treat a variety of intestinal nematode infections but is ineffective in trichuriasis. Pyrantel pamoate is usually effective in a single dose. Its target is the nicotinic acetylcholine receptor on the surface of nematode somatic muscle. Pyrantel depolarizes the neuromuscular junction of the nematode, resulting in its irreversible paralysis and allowing the natural expulsion of the worm.

Pyrantel pamoate is poorly absorbed from the intestine; >85% of the dose is passed unaltered in feces. The absorbed portion is metabolized and excreted in urine. Piperazine is antagonistic to pyrantel pamoate and should not be used concomitantly.

Pyrantel pamoate has minimal toxicity at the oral doses used to treat intestinal helminthic infection. It is not recommended for pregnant women or for children <12 months old.

PYRIMETHAMINE

When combined with short-acting sulfonamides, this diaminopyrimidine is effective in malaria, toxoplasmosis, and isosporiasis. Unlike mammalian cells, the parasites that cause these infections cannot utilize preformed pyrimidines obtained through salvage pathways but rather rely completely on de novo pyrimidine synthesis, for which folate derivatives are essential cofactors. The efficacy of pyrimethamine is increasingly limited by the development of resistant strains of *P. falciparum* and *P. vivax*. Single amino acid substitutions to parasite dihydrofolate reductase confer resistance to pyrimethamine by decreasing the enzyme's binding affinity for the drug.

Pyrimethamine is well absorbed; the drug is 87% bound to human plasma proteins. In healthy volunteers, drug concentrations remain at therapeutic levels for up to 2 weeks; drug levels are lower in patients with malaria.

At the usual dosage, pyrimethamine alone causes little toxicity except for occasional skin rashes and, more rarely, blood dyscrasias. Bone marrow suppression sometimes occurs at the higher doses used for toxoplasmosis; at these doses, the drug should be administered with folinic acid.

QUINACRINE*

Quinacrine is the only drug approved by the FDA for the treatment of giardiasis. Although its production was discontinued in 1992, quinacrine can be obtained from alternative sources through the CDC Drug Service. The antiprotozoal mechanism of quinacrine has not been fully elucidated. The drug inhibits NADH oxidase—the same enzyme that activates furazolidone. The differing relative quinacrine uptake rate between human cells and *G. lamblia* may explain the selective toxicity of the

drug. Resistance correlates with decreased drug uptake.

Quinacrine is rapidly absorbed from the intestinal tract and is widely distributed in body tissues. Alcohol is best avoided due to a disulfiram-like effect.

QUININE AND QUINIDINE

When combined with another agent, the cinchona alkaloid quinine is effective for the oral treatment of both uncomplicated, chloroquine-resistant malaria and babesiosis. Quinine acts rapidly against the asexual blood stages of all forms of human malaria. For severe malaria, only quinidine (the dextroisomer of quinine) is available in the United States. Quinine concentrates in the acidic food vacuoles of *Plasmodium* species. The drug inhibits the nonenzymatic polymerization of the highly reactive, toxic heme molecule into the nontoxic polymer pigment hemozoin.

Quinine is readily absorbed when given orally. In patients with malaria, the elimination half-life of quinine increases according to the severity of the infection. However, toxicity is avoided by an increase in the concentration of plasma glycoproteins. The cinchona alkaloids are extensively metabolized, particularly by CYP3A4; only 20% of the dose is excreted unchanged in urine. The drug's metabolites are also excreted in urine and may be responsible for toxicity in patients with renal failure. Renal excretion of quinine is decreased when cimetidine is taken and increased when the urine is acidic. The drug readily crosses the placenta.

Quinidine is both more potent as an antimalarial and more toxic than quinine. Its use requires cardiac monitoring. Dose reduction is necessary in persons with severe renal impairment.

SPIRAMYCIN†

This macrolide is used to treat acute toxoplasmosis in pregnancy and congenital toxoplasmosis. While the mechanism of action is similar to that of other macrolides, the efficacy of spiramycin in toxoplasmosis appears to stem from its rapid and extensive intracellular penetration, which results in macrophage drug concentrations 10–20 times greater than serum concentrations.

Spiramycin is rapidly and widely distributed throughout the body and reaches concentrations in the placenta up to five times those in serum. This agent is excreted mainly in bile. Indeed, in humans, the urinary excretion of active compounds represents only 20% of the administered dose.

Serious reactions to spiramycin are rare. Of the available macrolides, spiramycin appears to have the lowest risk of drug interactions. Complications of treatment are rare but, in neonates, can include life-threatening ventricular arrhythmias that disappear with drug discontinuation.

SULFONAMIDES

See Table 208-1 and Chap. 133.

SURAMIN*

This derivative of urea is the drug of choice for the early stage of African trypanosomiasis. The drug is polyanionic and acts by forming stable complexes with proteins, thus inhibiting multiple enzymes essential to parasite energy metabolism. Suramin appears to inhibit all trypanosome glycolytic enzymes more effectively than it inhibits the corresponding host enzymes.

Suramin is parenterally administered. It binds to plasma proteins and persists at low levels for several weeks after infusion. Its metabolism is negligible. This drug does not penetrate the CNS.

TAFENOQUINE

Tafenoquine is an 8-aminoquinoline with causal prophylactic activity. Its prolonged half-life (2–3 weeks) allows longer dosing intervals when the drug is used for prophylaxis. Tafenoquine has been well tolerated in clinical trials. When tafenoquine is taken with food, its absorption is increased by 50% and the most commonly reported adverse event—mild GI upset—is diminished. Like primaquine, tafenoquine is a potent oxidizing agent, causing hemolysis in patients with G6PD deficiency as well as methemoglobinemia.

TETRACYCLINES

See Table 208-1 and Chap. 133.

THIABENDAZOLE

Discovered in 1961, thiabendazole remains one of the most potent of the numerous benzimidazole derivatives. However, its use has declined significantly because of a higher frequency of adverse effects than is seen with other, equally effective agents.

Thiabendazole is active against most intestinal nematodes that infect humans. Although the exact mechanism of its antihelminthic activity has not been fully elucidated, it is likely to be similar to that of other benzimidazole drugs: namely, inhibition of polymerization of parasite β-tubulin. The drug also inhibits the helminth-specific enzyme fumarate reductase. In animals, thiabendazole has anti-inflammatory, antipyretic, and analgesic effects, which may explain its usefulness in dracunculiasis and trichinosis. Thiabendazole also suppresses egg and/or larval production by some nematodes and may inhibit the subsequent development of eggs or larvae passed in feces. Despite the emergence and

global spread of thiabendazole-resistant trichostrongyliasis among sheep, there have been no reports of drug resistance in humans.

Thiabendazole is available in tablet form and as an oral suspension. The drug is rapidly absorbed from the GI tract but can also be absorbed through the skin. Thiabendazole should be taken after meals. This agent is extensively metabolized in the liver before ultimately being excreted; most of the dose is excreted within the first 24 h. The usual dose of thiabendazole is determined by the patient's weight, but some treatment regimens are parasite specific. No particular adjustments are recommended in patients with renal or hepatic failure; only cautious use is advised.

Coadministration of thiabendazole in patients taking theophylline can result in an increase in theophylline levels by >50%. Therefore, serum levels of theophylline should be monitored closely in this situation.

TINIDAZOLE

This nitroimidazole is effective for the treatment of amebiasis, giardiasis, and trichomoniasis. Like metronidazole, tinidazole must undergo reductive activation by the parasite's metabolic system before it can act on protozoal targets. Tinidazole inhibits the synthesis of new DNA in the parasite and causes degradation of existing DNA. The reduced free-radical derivatives alkylate DNA, with consequent cytotoxic damage to the parasite. This damage appears to be produced by short-lived reduction intermediates, resulting in helix destabilization and strain breakage of DNA. The mechanism of action and side effects of tinidazole are similar to those of metronidazole, but adverse events appear to be less frequent and severe with tinidazole. In addition, the significantly longer half-life of tinidazole (>12 h) offers potential cure with a single dose.

TRICLABENDAZOLE

While most benzimidazoles have broad-spectrum antihelminthic activity, they exhibit minimal or no activity against *F. hepatica*. In contrast, the antihelminthic activity of triclabendazole is highly specific for *Fasciola* and *Paragonimus* species, with little activity against nematodes, cestodes, and other trematodes. Triclabendazole is effective against all stages of *Fasciola* species. The active sulfoxide metabolite of triclabendazole binds to fluke tubulin by assuming a unique nonplanar configuration and disrupts microtubule-based processes. Resistance to triclabendazole in veterinary use has been reported in Australia and Europe; however, no resistance has been documented in humans.

Triclabendazole is rapidly absorbed after oral ingestion; administration with food enhances its absorption and shortens the elimination half-life of the active metabolite. Both the sulfoxide and the sulfone metabolites are highly protein bound (>99%). Treatment with triclabendazole is typically given in one or two doses. No clinical data are available regarding dose adjustment in renal or hepatic insufficiency; however, given the short course of therapy and extensive hepatic metabolism of triclabendazole, dose adjustment is unlikely to be necessary. No information exists on drug interactions.

TRIMETHOPRIM-SULFAMETHOXAZOLE

See Table 208-1 and Chap. 133.

FURTHER READINGS

Hoerauf A: New strategies to combat filariasis. Expert Rev Anti Infect Ther 4:211, 2006[PMID: 16597203]

Keiser J, Utzinger J: The drugs we have and the drugs we need against major helminth infections. Adv Parasitol 73:197, 2010[PMID: 20627144]

Moore TA: Ivermectin, in *Kucer's Use of Antibiotics* , 6th ed, M L Grayson et al (eds). London, Oxford University Press, 2010, pp 2254–2262

———, McCarthy JS: Benzimidazoles, in *Antimicrobial Therapy and Vaccines* , 2nd ed, VL Yu et al (eds). Pittsburgh, ESun Technologies, 2005, pp 1021–1036

Schlagenhauf P et al: The position of mefloquine as a 21st century malaria chemoprophylaxis. Malar J 9:357, 2010[PMID: 21143906]

World Health Organization: Guidelines for the treatment of malaria, 2d ed. Geneva, WHO, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

AMEBIASIS

Definition

Amebiasis is infection with the parasitic intestinal protozoan *Entamoeba histolytica* (the "tissue-lysing ameba"). Most infections are probably asymptomatic, but *E. histolytica* can cause disease ranging from dysentery to extraintestinal infections, including liver abscesses.

Life Cycle and Transmission

E. histolytica exists in two stages: a hardy multinucleate cyst form (**Fig. 209-1**) and the motile trophozoite stage (**Fig. 209-2**). Infection (of which humans are the natural hosts) is acquired by ingestion of cysts contained in fecally contaminated food or water or, more rarely, through oral-anal sexual contact. Cysts survive stomach acidity and excyst within the small intestine to form the 20- to 50-μm trophozoite stage. Trophozoites can live within the large-bowel lumen without causing disease or can invade the intestinal mucosa, causing amebic colitis. In some cases, *E. histolytica* trophozoites invade through the mucosa and into the bloodstream, traveling through the portal circulation to reach the liver and causing amebic liver abscesses. Motile trophozoites may be excreted into the stool—a diagnostically important event—but are rapidly killed upon exposure to air or stomach acid and therefore are not infectious. Trophozoite cysts within the large bowel are excreted in the stool, continuing the life cycle.

Figure 209-1

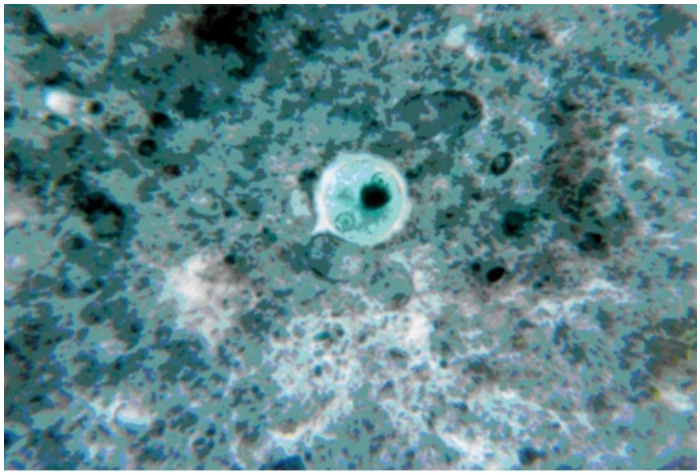

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***Entamoeba* cyst.** Three of the four nuclei are clearly visible. *(Courtesy of Dr. George Healy, Centers for Disease Control and Prevention.)*

Figure 209-2

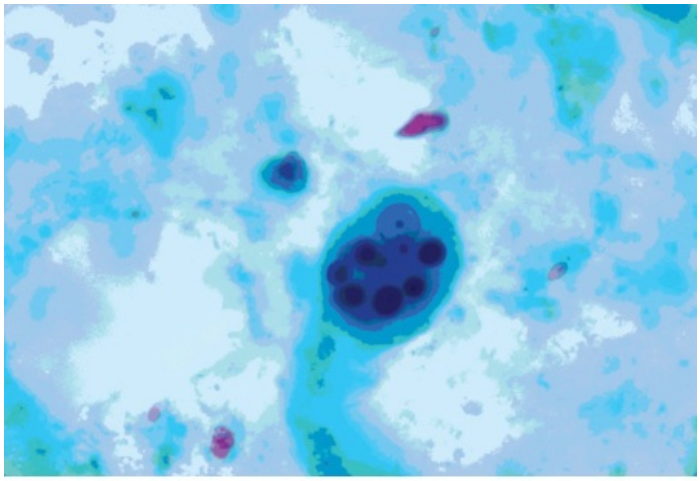

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***E. histolytica* trophozoite with ingested red blood cells.** Note the single nucleus with central nucleolus. *(Courtesy of the Centers for Disease Control and Prevention.)*

Epidemiology

Molecular diagnostics continue to clarify what was once a confusing picture of the true incidence and prevalence of *E. histolytica* infection and disease. It was a staple of most textbooks that 10% of the world's population was infected with *E. histolytica*. We now know that most asymptomatic individuals harboring amebic trophozoites or cysts in their stools are infected with a noninvasive species: *Entamoeba dispar* or *Entamoeba moshkovskii*. *E. dispar* appears not to cause disease, even in the most profoundly immunosuppressed individuals; furthermore, at this time, there is little evidence to suggest that *E. moshkovskii* causes disease, although epidemiologic studies of this species are in their infancy. In contrast, *E. histolytica* infection can cause disease, although not all patients develop symptoms. It remains unclear how frequently people infected with*E. histolytica* do develop symptoms; in one study in a highly endemic area, only 10% of infected patients developed symptoms over a 1-year observation period. A remarkable feature of amebiasis is its more common occurrence in men than in women, although the prevalence of infection with *E. histolytica* does not appear to differ between the sexes. This pattern is particularly pronounced for amebic liver abscess, whose prevalence is ~7 times higher among men than among women. The explanation for this difference remains unknown, but less efficient complement-mediated killing of amebic trophozoites by serum from men than by serum from women has been reported.

E. histolytica infections are most common in areas of the world where poor sanitation and crowding compromise the barriers to contamination of food and drinking water with human feces. Endemic areas include parts of Mexico, India, and nations in the tropical regions of Africa, South and Central America, and Asia. *E. histolytica* was present in ~2.1% of individuals presenting with diarrhea in a large series from Bangladesh and in 1.4% of the asymptomatic control group. In 2007, amebiasis was listed as the sixth most common cause of disease in Mexico, with an incidence of ~544 cases per 100,000 population. In the United States and other developed countries, disease is unusual and is found almost exclusively in travelers or immigrants from endemic areas. Rarely, outbreaks take place in institutionalized populations, and infections have been documented with increased frequency among men who have sex with men; however, most of the latter cases have been asymptomatic and probably represent *E. dispar* infections.

Pathogenesis and Pathology

E. histolytica trophozoites possess a potent repertoire of adhesins, proteinases, pore-forming proteins, and other effector molecules that enable them to lyse cells and tissue, induce both cellular necrosis and apoptosis, and resist both innate and adaptive immune defenses. Disease begins when *E. histolytica* trophozoites adhere to colonic mucosal epithelial cells. Disruption of the colonic mucin barrier is seen in pathologic sections from the diseased colon, but it is not clear whether this disruption is caused by the parasite, facilitating its adherence to mucosal cells, or occurs as a consequence of the adhesion event, with subsequent mucosal damage. Adherence is mediated primarily by a family of surface lectin molecules capable of binding to galactose and N-acetylgalactosamine residues. *E. histolytica* can lyse host cells upon contact through a family of amphipathic peptides called *amoebapores* that form barrel-stave pores in target cell membranes. Both cellular necrosis and apoptosis can occur after *E. histolytica* comes into contact with host cells, and which outcome predominates may relate to inherent characteristics of the target cell or the tissue environment. One consistent and unequivocal finding is the important role played by amebic cysteine proteinases in the

disease process. *E. histolytica* possesses a large family of cysteine proteinases that are capable of lysing the extracellular matrix between host cells (thus detaching cells and facilitating invasion) and cleaving host defense molecules (including complement components and antibodies). Studies in animal models, including chimeric mice with human intestinal xenografts, have shown that inhibition of *E. histolytica* cysteine proteinase activity, via either direct gene targeting or chemical inhibitors, significantly reduces disease. The ultimate effect of all these amebic virulence factors on the human colon is the production of small ulcers that have heaped borders and contain focal areas of epithelial cell loss, a modest inflammatory response, and mucosal hemorrhage. The intervening mucosa is usually normal, but diffuse hyperemia is sometimes seen. *E. histolytica* trophozoites can then invade laterally through the submucosal layer, creating the classic flask-shaped ulcers that appear on pathologic examination as narrow-necked lesions, broadening in the submucosal region, with *E. histolytica* trophozoites at the margins between dead and live tissues (**Fig. 209-3**). Ulcers tend to stop at the muscularis layer, and full-thickness lesions and colonic perforation are unusual. Amebomas, a rare complication of intestinal disease, are granulomatous mass lesions protruding into the bowel lumen, with a thickened edematous and hemorrhagic bowel wall that can cause obstructive symptoms.

Figure 209-3

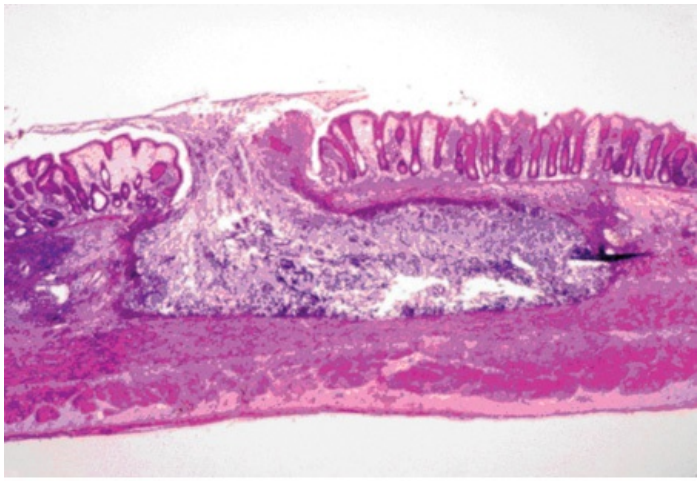

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

E. histolytica **flask-shaped intestinal ulceration from a kitten.***(Courtesy of Dr. Mae Melvin, Centers for Disease Control and Prevention.)*

In some individuals with *E. histolytica* colonic infection, trophozoites invade the portal venous system and reach the liver, where they cause amebic liver abscesses. *E. histolytica* trophozoites must resist lysis by serum complement to survive in the bloodstream. Amebic liver abscesses have a characteristic appearance on pathologic examination: the roughly circular abscesses contain a large necrotic center resembling anchovy paste that is surrounded by a narrow ring of a few inflammatory cells, fibrosis, and occasionally a few amebic trophozoites. The adjacent liver parenchyma is usually completely normal. Results in experimental rodent models of amebic liver abscess suggest that initial lesions may have more inflammatory cells and that lysis of neutrophils by *E. histolytica* trophozoites may contribute to tissue damage. In murine models of disease, apoptosis is a prominent component of hepatocyte death and the blockade of caspase activity can significantly reduce liver abscess formation, but whether any of these factors is applicable to human disease is unclear.

The role of innate and adaptive immunity in preventing*E. histolytica* infection or controlling disease needs further clarification. Studies of children in a highly endemic area have suggested that prior *E. histolytica* intestinal infection may stimulate mucosal IgA antibodies to amebic antigens, thereby reducing the likelihood of subsequent infections; this protection is relatively short lived. In contrast, among individuals in an area of Vietnam with a high prevalence of amebic liver abscess, a prior episode of disease did not reduce the risk of a second case, despite the presence of serum antibodies. Studies of animal models suggest that cell-mediated immunity may play a role in host defense, and glucocorticoid use has been associated with worse outcomes in patients with amebic colitis. However, individuals with HIV/AIDS do not appear to be at increased risk for infection with *E. histolytica*, and there is no evidence that they develop more severe disease than do immunocompetent hosts.

Clinical Syndromes

INTESTINAL AMEBIASIS

Most patients harboring *Entamoeba* species are asymptomatic, but individuals with *E. histolytica* infection can develop disease. Symptoms of amebic colitis generally appear 2–6 weeks after ingestion of the cyst form of the parasite. Diarrhea (classically heme-positive) and lower abdominal pain are the most common symptoms. Malaise and weight loss may be noted as disease progresses. Severe dysentery, with 10–12 small-volume, blood- and mucus-containing stools daily, may develop, but only ~40% of patients are febrile. Fulminant amebic colitis, with even more profuse diarrhea, severe abdominal pain (including peritoneal signs), fever, and pronounced leukocytosis are rare, disproportionately affecting young children, pregnant women, individuals being treated with glucocorticoids, and possibly individuals with diabetes or alcoholism. Paralytic ileus and colonic mucosal sloughing may be seen; intestinal perforation occurs in >75% of patients with this fulminant form of disease. Mortality rates from fulminant amebic colitis exceed 40% in some series. Recognized complications of amebic colitis also include toxic megacolon (documented in ~0.5% of patients with colitis), with severe bowel dilation and intramural air, and the aforementioned ameboma, which presents as an abdominal mass that may be confused with colon cancer.

AMEBIC LIVER ABSCESS

Just a century ago, amebic liver abscess—the most common extraintestinal manifestation of amebiasis—was almost always fatal; however, with current rapid diagnostic methods and effective medical treatment, mortality rates are now 1–3%. Disease begins when *E. histolytica* trophozoites penetrate through the colonic mucosa, travel through the portal circulation, and reach the liver. Most individuals with amebic liver abscess do not have concurrent signs or symptoms of colitis, and most do not have *E. histolytica* trophozoites in their stools. The exceptions are individuals with fulminant amebic colitis, in which concurrent amebic liver abscess is not uncommon. Disease can arise from months to years after travel to or residence in an endemic area; therefore, a careful travel history is key in making the diagnosis. The classic presentations of amebic liver abscess are right-upper-quadrant pain, fever, and hepatic tenderness. The pace of disease is usually acute, with symptoms lasting <10 days. However, a more chronic presentation, with weight loss and anorexia as prominent accompanying features, does occur. Jaundice is unusual, but dullness and rales at the right lung base (secondary to pleural effusion) are common. The most common laboratory findings are leukocytosis (without eosinophilia), an elevated alkaline phosphatase level, mild anemia, and an elevated erythrocyte sedimentation rate.

OTHER EXTRAINTESTINAL COMPLICATIONS OF AMEBIASIS

Right-sided pleural effusions and atelectasis are common in cases of amebic liver abscess and generally require no treatment. However, the abscess ruptures through the diaphragm in ~10% of patients, causing pleuropulmonary amebiasis. Suggestive symptoms are sudden-onset cough, pleuritic chest pain, and shortness of breath. In some patients, pleuropulmonary amebiasis is the presenting manifestation of amebic liver abscess and may be confused with bacterial pneumonia and empyema. A dramatic complication is the development of a hepatobronchial fistula, in which patients can cough up the contents of the liver abscess—copious amounts of brown sputum that may contain *E. histolytica* trophozoites. In ~1–3% of cases, the amebic liver abscess ruptures into the peritoneum, and peritoneal signs and shock develop. Even rarer is rupture of an amebic liver abscess into the pericardium; the signs and symptoms are those commonly seen with pericarditis (chest pain, pericardial rub, dyspnea, tachypnea, or cardiac tamponade), and nearly 30% of cases end in death. Cerebral abscesses complicate <0.1% of cases of amebic liver abscess and are associated with the sudden onset of headache, vomiting, seizures, and mental status changes and a high mortality rate. Cutaneous amebiasis (which usually involves the anal and perianal regions), genital disease (including rectovaginal fistulas), and urinary tract lesions are rare but reported complications of amebiasis.

Diagnostic Tests

The diagnosis of amebic colitis has traditionally been based on the demonstration of *E. histolytica* trophozoites or cysts in the stool or colonic mucosa of patients with diarrhea. However, the inability of microscopy to differentiate between *E. histolytica* and other *Entamoeba* species, such as *E. dispar* and *E. moshkovskii*, limits its effectiveness as a sole diagnostic method. Examination of three stool samples improves sensitivity for the detection of *Entamoeba* species, and it has been argued that the presence of amebic trophozoites containing red blood cells in a diarrheal stool is highly suggestive of *E. histolytica* infection. However, because trophozoites containing red blood cells are not found in most patients with *E. histolytica* infection, the applicability of this finding is limited.

Despite these inherent limitations, microscopy, often combined with serologic testing, remains the standard diagnostic approach in many hospitals and clinics worldwide. Culture of stools for *E. histolytica* trophozoites serves as a research tool but is generally not available for clinical use. PCR assay for DNA in stool samples is currently the most sensitive and specific method for identifying *E. histolytica* infection and has become a valuable epidemiologic and research tool; probes can be configured to detect *E. dispar* and *E. moshkovskii* as well. While significant advances are being made in reducing the costs of PCR-based diagnostics, this method still is not feasible for clinical diagnosis in most endemic areas. Commercially available tests that use enzyme-linked immunosorbent assays (ELISAs) or immunochromatographic techniques to detect *Entamoeba* antigens are less expensive and

more easily performed and are being used with increasing frequency. Greater sensitivity than microscopy and the ability to detect *E. histolytica* specifically are claimed by some of the leading kits, representing significant advantages over microscopy. Unfortunately, not all clinical studies have supported these claims, concerns have been raised about the specificity of the tests in nonendemic areas, and the ELISAs are less sensitive and specific than are PCR-based diagnostics. At this point, antigen detection–based ELISAs that can specifically identify *E. histolytica* in stool probably represent the best choice in endemic areas; however, the results of any of these diagnostic tests need to be interpreted in light of clinical presentation, and a second confirmatory test (e.g., microscopy and/or amebic serology) may be prudent. In instances in which amebiasis is suspected on clinical grounds in a patient with acute colitis but initial stool samples are negative, colonoscopy with examination of brushings or mucosal biopsies for *E. histolytica* trophozoites may be helpful in making the diagnosis or in identifying other diseases, such as inflammatory bowel disease or pseudomembranous colitis.

The diagnosis of amebic liver abscess is based on the detection (generally by ultrasound or CT; **Fig. 209-4**) of one or more space-occupying lesions in the liver and a positive serologic test for antibodies to *E. histolytica* antigens. As has been noted, amebiasis can present months or years after travel to or residence in an endemic area, and so a careful travel history is mandatory when anyone presents with a liver abscess. Amebic liver abscesses are classically described as single, large, and located in the right lobe of the liver, but sensitive imaging techniques have shown that multiple abscesses are more common than previously suspected. When a patient has a space-occupying lesion of the liver, a positive amebic serology is highly sensitive (>94%) and highly specific (>95%) for the diagnosis of amebic liver abscess. False-negative serologic tests have been reported when serum samples were obtained very early in the course of abscess (within 7–10 days of onset), but repeat tests are almost always positive.

Figure 209-4

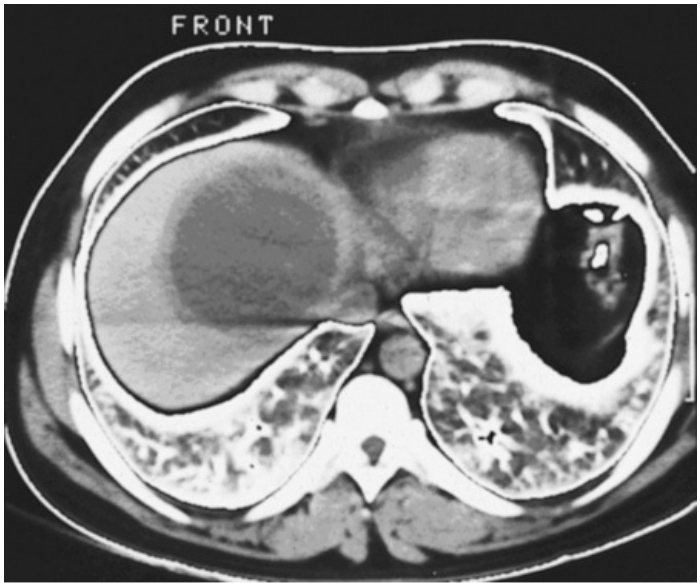

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Large amebic abscess in the right lobe of the liver visualized by CT. *(Courtesy of Dr. M. M. Reeder, International Registry of Tropical Imaging.)*

Differential Diagnosis

The differential diagnosis of amebic colitis includes bacterial dysentery (e.g., *Shigella* and *Campylobacter* infections), schistosomiasis, *Balantidium coli* infection, pseudomembranous colitis, inflammatory bowel disease, and ischemic colitis. Stool cultures for bacterial pathogens, microscopic examination of stools, and amebic serology help differentiate amebic colitis from these other entities. Amebomas may be confused with colonic carcinoma; several case reports describe instances in which amebomas and associated liver abscesses were initially considered to be colon cancer with liver metastases. Amebic liver abscess must be distinguished from pyogenic liver abscess, echinococcal cysts, and primary or metastatic liver tumors. It is difficult to differentiate pyogenic from amebic liver abscesses on purely clinical grounds, but amebic serology is usually the key test in excluding or diagnosing amebic liver abscess. Abscesses that rupture into the pleural space may be accompanied by cough, sputum production, and dyspnea and may initially be diagnosed as bronchopneumonia.

Treatment: Amebiasis

The nitroimidazole compounds tinidazole and metronidazole are the drugs of choice for the treatment of amebic colitis and amebic liver abscess (**Table 209-1**). To date, *E. histolytica* has not demonstrated resistance to any of the commonly used agents—a situation that greatly simplifies treatment. Tinidazole appears to be better tolerated and slightly more effective than metronidazole for amebic colitis and amebic liver abscess. Metronidazole is available as a parenteral formulation for patients who cannot take oral medications. Whenever possible, fulminant amebic colitis is managed conservatively, even in the presence of perforation, with the addition of antibiotics to treat gut bacteria and percutaneous catheter drainage of fluid collections if needed.

Table 209-1 Recommended Therapeutic Dosages of Antiamebic Drugs

Drug	Dosage	Duration, Days
Amebic Colitis or Amebic Liver Abscess		
Tinidazole	2 g/d PO with food	3
Metronidazole	750 mg tid PO or IV	5–10
***Entamoeba histolytica* Luminal Infection**		
Paromomycin	30 mg/kg qd PO in 3 divided doses	5–10
Iodoquinol	650 mg PO tid	20

Remarkably, given the large size of amebic liver abscesses, treatment with tinidazole or metronidazole in the same doses used for amebic colitis is almost always successful. More than 90% of patients respond with a decrease in abdominal pain and fever within 72 h of the initiation of therapy. Drainage of amebic liver abscesses is rarely needed; in one large series, neither time to becoming afebrile nor length of hospitalization was significantly different for patients who underwent percutaneous radiography-guided aspiration of the abscess accompanied by medical therapy than for those who received medical therapy alone. Aspiration should be reserved for individuals in whom pyogenic abscess or a bacterial superinfection is suspected but whose diagnosis is uncertain, for patients failing to respond to tinidazole or metronidazole (i.e., those who have persistent fever or abdominal pain after 4 days of treatment), for individuals with large liver abscesses in the left lobe (because of the risk of rupture into the pericardium), and for patients whose large abscesses and accelerated clinical course raise concerns about imminent rupture. In contrast, aspiration and/or percutaneous catheter drainage improves outcomes in patients with pleuropulmonary amebiasis and empyema (where amebic liver abscesses have ruptured into the pleural space), and percutaneous catheter or surgical drainage is absolutely indicated for cases of amebic pericarditis. Rupture of an amebic liver abscess into the peritoneum is generally managed conservatively, with medical therapy and percutaneous catheter drainage of fluid collections as needed.

Neither metronidazole nor tinidazole reaches high levels in the gut lumen; therefore, patients with amebic colitis or amebic liver abscess should also receive treatment with a luminal agent (paromomycin or iodoquinol) to ensure eradication of the infection (Table 209-1). Paromomycin is the preferred agent. Asymptomatic individuals with documented *E. histolytica* infection should be treated because of the risks of developing amebic colitis or amebic liver abscess in the future and of transmitting the infection to others. Paromomycin or iodoquinol in the doses listed in the table should be used in these cases.

Nitazoxanide, a broad-spectrum antiparasitic drug, is efficacious against *E. histolytica* trophozoites in both tissue and gut lumen and may become an important addition to the therapeutic repertoire. However, clinical experience with nitazoxanide for the treatment of *E. histolytica* infection remains limited at this point.

Prevention

Avoidance of the ingestion of food and water contaminated with human feces is the only way to prevent *E. histolytica* infection. Travelers to endemic areas should exercise the same measures used to reduce the risk of travelers' diarrhea (Chap. 117). Treatment of asymptomatic persons who pass *E. histolytica* cysts in the stool may help reduce opportunities for disease transmission. There is no evidence for any effective prophylaxis, and no vaccine is available.

INFECTION WITH FREE-LIVING AMEBAS

In contrast to the trophozoites of the parasitic *E. histolytica*, which can survive only in humans and some other primate hosts, free-living amebas of the genera *Naegleria*, *Acanthamoeba*, and *Balamuthia* live in brackish or freshwater habitats around the world (including lakes, tap water, swimming pools, and air conditioning and heating ducts) and are accidental and opportunistic agents of disease.

Naegleria Infections

Naegleria (the "brain-eating ameba") is the causative agent of primary amebic meningoencephalitis (PAM). Nearly always fatal but quite rare, cases of PAM have been reported from 15 countries and from all continents except Antarctica; 35 cases were reported in the United States between 1998 and 2009. *Naegleria* prefers warm freshwater, and most cases occur in otherwise healthy children, who usually have swum in lakes or swimming pools during the previous 2 weeks. *Naegleria* enters the central nervous system via water inhaled or splashed into the nose, with trophozoites disrupting the olfactory mucosa, invading through the cribriform plate, and ascending via the olfactory nerves into the brain. The earliest manifestations are anosmia (usually perceived as alterations in taste), headache, fever, photophobia, nausea, and vomiting. Cranial nerve palsies, especially of the third, fourth, and sixth nerves, are documented and rapid progression of disease, with seizures, coma, and death within 7–10 days of the onset of symptoms, are common. Pathologic examination reveals hemorrhagic necrosis of brain tissue (often most prominent in the olfactory bulbs), evidence of increased intracranial pressure, scant purulent material that may contain a few amebas, and marked leptomeningitis (**Fig. 209-5**).

Figure 209-5

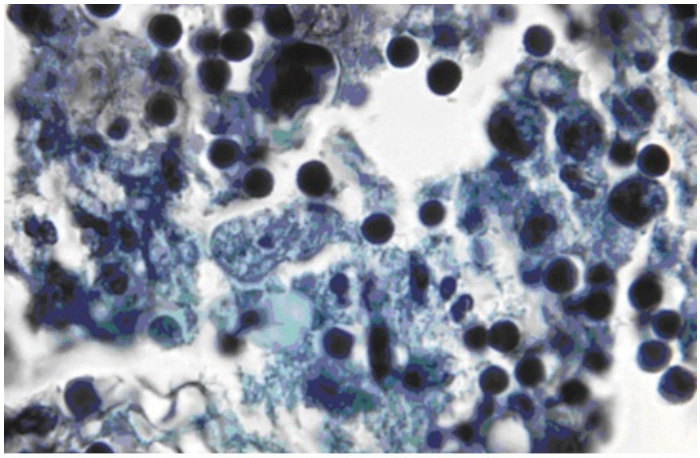

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***Naegleria* in a section of human brain tissue** from a patient with primary amebic meningoencephalitis. *(Courtesy of Dr. George Healy, Centers for Disease Control and Prevention .)*

The diagnosis of PAM is based on the finding of motile *Naegleria* trophozoites in wet mounts of freshly obtained cerebrospinal fluid (CSF). Laboratory findings in the CSF resemble those in bacterial meningitis, with high opening pressures, low glucose levels, high protein concentrations, and elevated polymorphonuclear cell–predominant white blood cell counts. PAM should be suspected in any patient who has an appropriate history and purulent meningoencephalitis with negative gram stains, negative antigen detection and PCR tests for other pathogens, and negative bacterial cultures. Unfortunately, the prognosis for PAM is dismal. The few survivors who have been reported were treated with high-dose amphotericin B and rifampin in combination.

Acanthamoeba Infections

Acanthamoeba species are free-living amebas that cause two major clinical syndromes: granulomatous amebic encephalitis and keratitis. Granulomatous amebic encephalitis occurs in debilitated, chronically ill, and immunosuppressed individuals who may be undergoing chemotherapy, receiving glucocorticoids, or suffering from lymphoproliferative diseases, systemic lupus erythematosus, or AIDS. It is believed that *Acanthamoeba* reaches the central nervous system through the bloodstream, traveling from a primary site of infection in the nares, skin, sinuses, or lungs. The pace of infection is indolent compared with that of PAM. Granulomatous amebic encephalitis tends to present as a space-occupying lesion in the brain. Common symptoms include altered mental status, stiff neck, and headache along with focal findings including hemiparesis, ataxia, and cranial nerve palsies. Seizures and coma often precede death. Pathologic findings in the brain include cerebral edema and multiple areas of necrosis and hemorrhage. Amebic trophozoites and cysts are scattered throughout the tissue and are often located near blood vessels (**Fig. 209-6**). Multinucleated giant cells forming granulomas give the syndrome its name but are seen less often in highly immunocompromised patients. The diagnosis is usually made by detection of *Acanthamoeba* trophozoites or cysts in biopsy specimens; a fluorescein-labeled antiserum is available from the Centers for Disease Control and Prevention (CDC) to help identify *Acanthamoeba* in microscopic sections. *Acanthamoeba* trophozoites and cysts are occasionally seen in CSF, but samples from most patients with granulomatous amebic encephalitis show mild lymphocyte-predominant pleocytosis, slightly elevated protein levels, and

normal or slightly depressed glucose concentrations without the presence of amebas. CT findings vary, with hypodense lesions that resemble infarcts in some patients and multiple enhancing lesions that resemble toxoplasmosis in others. Unfortunately, there are no therapies with proven efficacy against this disease, and almost all cases have ended in death. There have been case reports of survivors treated with multidrug combinations that included pentamidine, sulfadiazine, flucytosine, rifampin, and fluconazole.

Figure 209-6

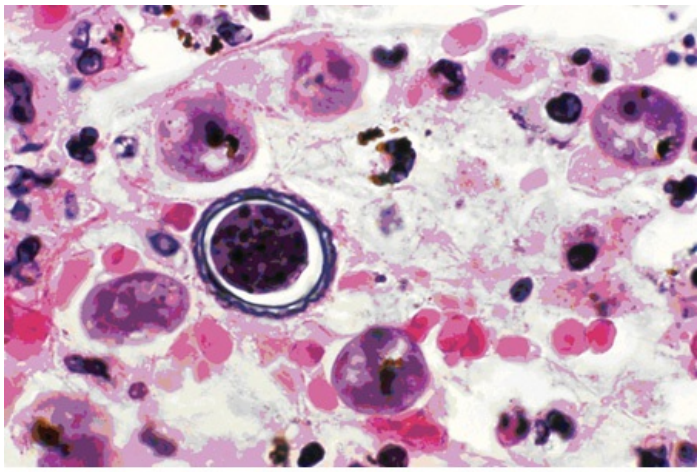

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***Acanthamoeba* cyst in brain tissue** from a patient with granulomatous amebic encephalitis. *(Courtesy of Dr. George Healy, Centers for Disease Control and Prevention.)*

Acanthamoeba keratitis is associated with corneal injuries complicated by exposure to water or soil and with the wearing of contact lenses. In contact lens–associated infection, extended wear, breaches in hygiene and disinfection procedures, swimming with contact lenses in place, and the use of homemade saline solutions contaminated with *Acanthamoeba* are important risk factors. The incidence of *Acanthamoeba* keratitis varies from 1.65–2.01 cases per million contact lens users in the United States to 17.53–19.5 cases per million users in the United Kingdom. Unilateralphotophobia, excessive tearing, redness, and foreign-body sensation are the earliest signs and symptoms; disease is bilateral in some contact lens users. *Acanthamoeba* keratitis can progress rapidly; abscesses, hypopyon, scleritis, and corneal perforation with vision loss can develop within weeks. The disease may be diagnosed by identification of the polygonal cyst form in corneal scrapings or biopsy material, by culture of biopsy samples or contact lenses on *Escherichia coli*–seeded agar plates, by confocal microscopy, and by PCR. The differential diagnosis includes bacterial, fungal, mycobacterial, and viral (particularly herpetic) causes. Current therapy involves topical administration of a cationic antiseptic agent such as a biguanide or chlorhexidine, with or without a diamidine agent. The persistence of the cyst form of *Acanthamoeba* complicates treatment, and long durations of therapy (6 months to 1 year) are required. In severe cases, particularly when vision is threatened or already diminished, penetrating keratoplasty may be indicated.

Balamuthia Infections

Balamuthiamandrillaris is a free-living ameba that causes meningoencephalitis in both immunosuppressed and immunocompetent hosts, particularly children and the elderly. The disease presents similarly to granulomatous amebic encephalitis caused by *Acanthamoeba*, and essentially all of the points made above with regard to the latter organism—in terms of clinical presentation, pathologic findings, and lack of proven therapies—apply to *Balamuthia* infections as well. Most cases are identified post mortem; the few cases identified before death have been found during histologic examination of brain biopsy specimens. A specific antiserum is available from the CDC to aid in identifying *B. mandrillaris* in clinical specimens.

FURTHER READINGS

Dart JKG et al: *Acanthamoeba* keratitis: Diagnosis and treatment update 2009. Am J Ophthalmol 148:487, 2009[PMID: 19660733]

Fotedar R et al: Laboratory diagnostic techniques for *Entamoeba* species. Clin Microbiol Rev 20:511, 2007[PMID: 17630338]

Haque R et al: Prospective case-control study of the association between common enteric protozoal parasites and diarrhea in Bangladesh. J Infect Dis 48:1191, 2009[PMID: 19323634]

Snow M et al: Differences in complement-mediated killing of *Entamoeba histolytica* between men and women—an explanation for the increased susceptibility of men to invasive amebiasis? Am J Trop Med Hyg 78:922, 2008[PMID: 18541770]

Stanley SL: Amoebiasis. Lancet 361:1025, 2003[PMID: 12660071]

Visvesvara GS et al: Pathogenic and opportunistic free-livingamoebae: *Acanthamoeba* spp., *Balamuthia mandrillaris, Naegleria fowleri*, and *Sappinia diploidea*. FEMS Immunol Med Microbiol 50:1, 2007[PMID: 17428307]

Ximénez C et al: Reassessment of the epidemiology of amebiasis: State of the art. Infect Genet Evol 9:1023, 2009[PMID: 21413565]

Zhang Z et al: *Entamoeba histolytica* cysteine proteinases with interleukin-1 beta converting enzyme (ICE) activity cause intestinal inflammation and tissue damage in amoebiasis. Mol Microbiol 37:542, 2000[PMID: 10931347]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

Humanity has but three great enemies: Fever, famine and war; of these by far the greatest, by far the most terrible, is fever.

William Osler

MALARIA: INTRODUCTION

Malaria is a protozoan disease transmitted by the bite of infected *Anopheles* mosquitoes. The most important of the parasitic diseases of humans, it is transmitted in 108 countries containing 3 billion people and causes nearly 1 million deaths each year. Malaria has been eliminated from the United States, Canada, Europe, and Russia; in the late twentieth and early twenty-first centuries, however, its prevalence rose in many parts of the tropics. Despite enormous control efforts, increases in the drug resistance of the parasite, the insecticide resistance of its vectors, and human travel and migration have contributed to this resurgence. Occasional local transmission after importation of malaria has occurred in several southern and eastern areas of the United States and in Europe, indicating the continual danger to nonmalarious countries. Although there are many promising new control and research initiatives, malaria remains today, as it has been for centuries, a heavy burden on tropical communities, a threat to nonendemic countries, and a danger to travelers.

ETIOLOGY AND PATHOGENESIS

Five species of the genus *Plasmodium* cause nearly all malarial infections in humans. These are *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae,* and—in Southeast Asia—the monkey malaria parasite *P. knowlesi*, which can be reliably identified only by molecular methods (Table 210-1). Almost all deaths are caused by falciparum malaria. Human infection begins when a female anopheline mosquito inoculates plasmodial *sporozoites* from its salivary gland during a blood meal (Fig. 210-1). These microscopic motile forms of the malarial parasite are carried rapidly via the bloodstream to the liver, where they invade hepatic parenchymal cells and begin a period of asexual reproduction. By this amplification process (known as *intrahepatic* or *preerythrocytic schizogony* or *merogony*), a single sporozoite eventually may produce from 10,000 to >30,000 daughter merozoites. The swollen infected liver cell eventually bursts, discharging motile *merozoites* into the bloodstream. These merozoites then invade the red blood cells (RBCs) and multiply six- to twentyfold every 48–72 h. When the parasites reach densities of ~50/μL of blood (~100 million parasites in the blood of an adult), the symptomatic stage of the infection begins. In *P. vivax* and *P. ovale* infections, a proportion of the intrahepatic forms do not divide immediately but remain dormant for a period ranging from 3 weeks to a year or longer before reproduction begins. These dormant forms, or *hypnozoites*, are the cause of the relapses that characterize infection with these two species.

Table 210-1 Characteristics of *Plasmodium* Species Infecting Humans

	Finding for Indicated Species[a]			
Characteristic	***P. falciparum***	***P. vivax***	***P. ovale***	***P. malariae***
Duration of intrahepatic phase (days)	5.5	8	9	15
Number of merozoites released per infected hepatocyte	30,000	10,000	15,000	15,000
Duration of erythrocytic cycle (hours)	48	48	50	72
Red cell preference	Younger cells (but can invade cells of all ages)	Reticulocytes and cells up to 2 weeks old	Reticulocytes	Older cells
Morphology	Usually only ring forms[b]; banana-shaped gametocytes	Irregularly shaped large rings and trophozoites; enlarged erythrocytes; Schüffner's dots	Infected erythrocytes, enlarged and oval with tufted ends; Schüffner's dots	Band or rectangular forms of trophozoites common
Pigment color	Black	Yellow-brown	Dark brown	Brown-black
Ability to cause relapses	No	Yes	Yes	No

[a] In Southeast Asia, the monkey malaria parasite *P. knowlesi* also causes disease in humans.

[b] Parasitemias of >2% are suggestive of *P. falciparum* infection.

Figure 210-1

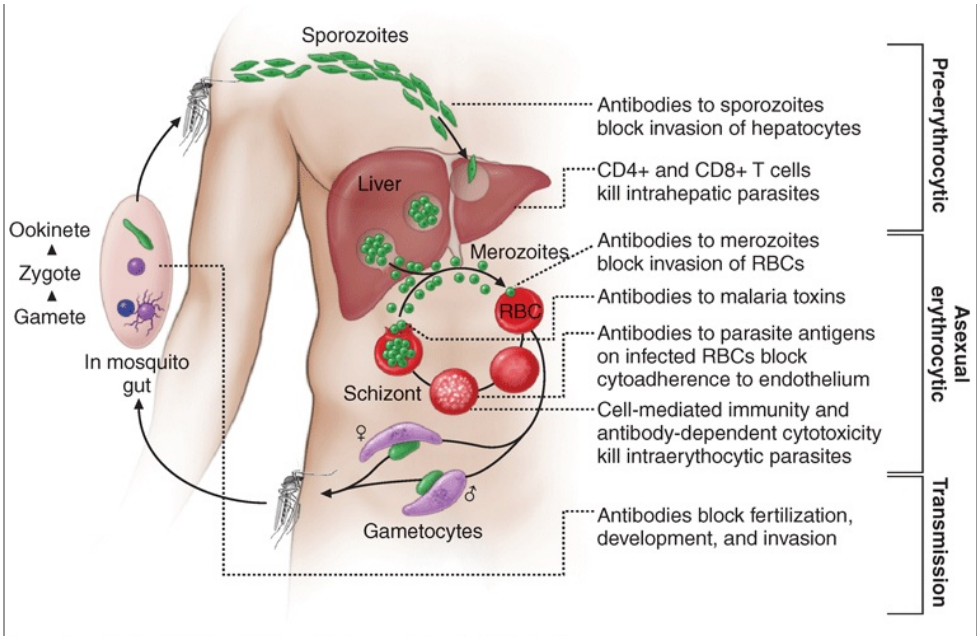

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The malaria transmission cycle from mosquito to human. RBC, red blood cell.

After entry into the bloodstream, merozoites rapidly invade erythrocytes and become *trophozoites*. Attachment is mediated via a specific erythrocyte surface receptor. In the case of *P. vivax*, this receptor is related to the Duffy blood-group antigen Fy^a or Fy^b. Most West Africans and people with origins in that region carry the Duffy-negative FyFy phenotype and are therefore resistant to *P. vivax* malaria. During the early stage of intraerythrocytic development, the small "ring forms" of the different parasitic species appear similar under light microscopy. As the trophozoites enlarge, species-specific characteristics become evident, pigment becomes visible, and the parasite assumes an irregular or ameboid shape. By the end of the 48-h intraerythrocytic life cycle (24 h for *P. knowlesi*, 72 h for *P. malariae*), the parasite has consumed two-thirds of the RBC's hemoglobin and has grown to occupy most of the cell. It is now called a *schizont*. Multiple nuclear divisions have taken place (*schizogony* or *merogony*), and the RBC then ruptures to release 6–30 daughter merozoites, each potentially capable of invading a new RBC and repeating the cycle. The disease in human beings is caused by the direct effects of RBC invasion and destruction by the asexual parasite and the host's reaction. After a series of asexual cycles (*P. falciparum*) or immediately after release from the liver (*P. vivax*, *P. ovale*, *P. malariae*, *P. knowlesi*), some of the parasites develop into morphologically distinct, longer-lived sexual forms (*gametocytes*) that can transmit malaria.

After being ingested in the blood meal of a biting female anopheline mosquito, the male and female gametocytes form a zygote in the insect's midgut. This zygote matures into an ookinete, which penetrates and encysts in the mosquito's gut wall. The resulting oocyst expands by asexual division until it bursts to liberate myriad motile sporozoites, which then migrate in the hemolymph to the salivary gland of the mosquito to await inoculation into another human at the next feeding.

EPIDEMIOLOGY

Malaria occurs throughout most of the tropical regions of the world (Fig. 210-2). *P. falciparum* predominates in Africa, New Guinea, and Hispaniola (i.e., the Dominican Republic and Haiti); *P. vivax* is more common in Central America. The prevalence of these two species is approximately equal in South America, the Indian subcontinent, eastern Asia, and Oceania. *P. malariae* is found in most endemic areas, especially throughout sub-Saharan Africa, but is much less common. *P. ovale* is relatively unusual outside of Africa and, where it is found, comprises <1% of isolates. Patients infected with *P. knowlesi* have been identified on the island of Borneo and to a lesser extent elsewhere in Southeast Asia where the main hosts, long-tailed and pig-tailed macaques, are found.

Figure 210-2

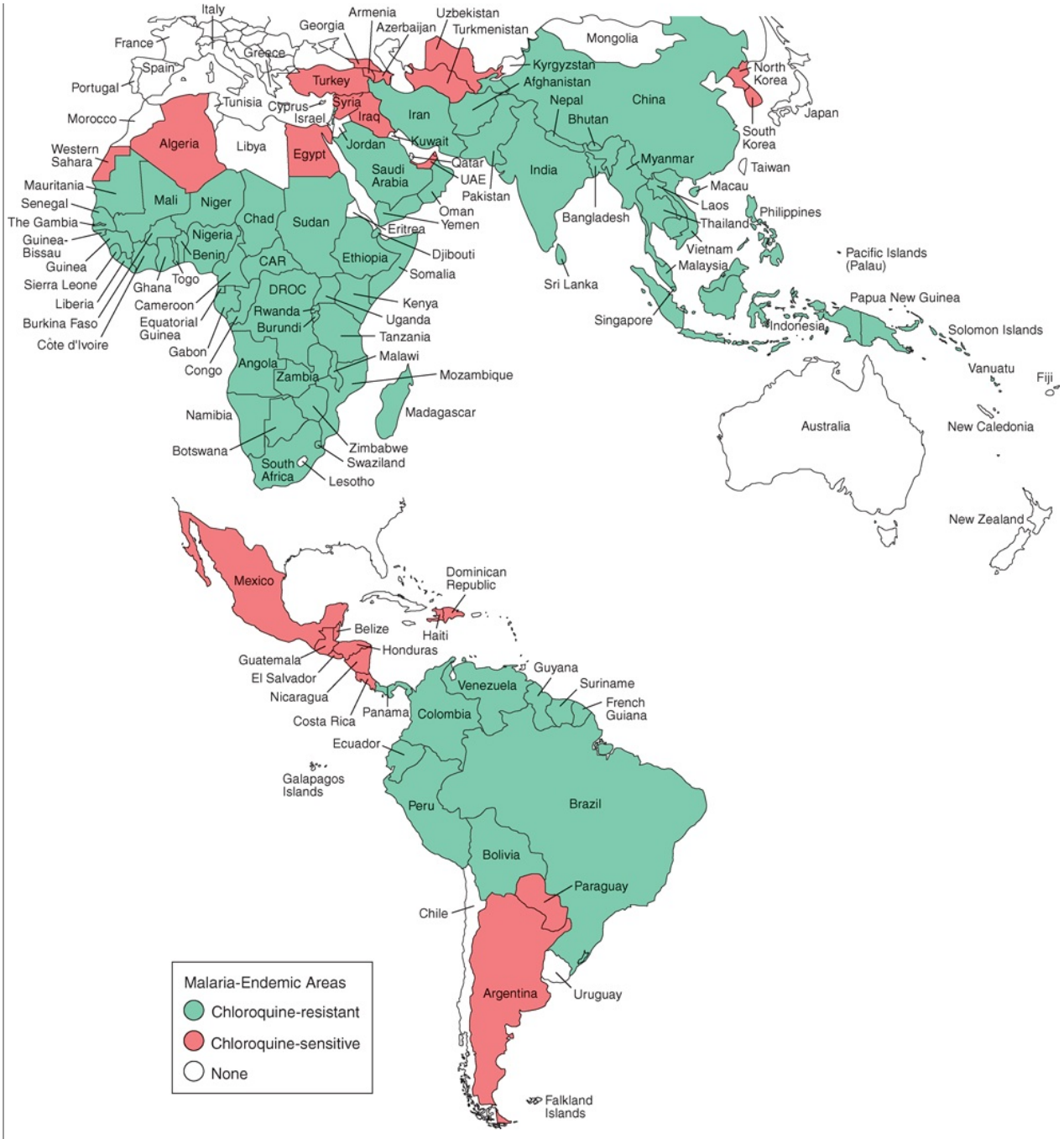

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Malaria-endemic countries in the Americas (***bottom***) and in Africa, the Middle East, Asia, and the South Pacific (***top***), 2007. CAR, Central African Republic; DROC, Democratic Republic of the Congo; UAE, United Arab Emirates.

The epidemiology of malaria is complex and may vary considerably even within relatively small geographic areas. Endemicity traditionally has been defined in terms of parasitemia rates or palpable-spleen rates in children 2–9 years of age as hypoendemic (<10%), mesoendemic (11–50%), hyperendemic (51–75%), and holoendemic (>75%); however, it is uncommon to use these indices for planning control programs because statistically valid national surveys are not done routinely in most endemic areas. In holo- and hyperendemic areas (e.g., certain regions of tropical Africa or coastal New Guinea) where there is intense *P. falciparum* transmission, people may sustain more than one infectious mosquito bite per day and are infected repeatedly throughout their lives. In such settings, rates of morbidity and mortality due to malaria are considerable during early childhood. Immunity against disease is hard won in these areas, and the burden of disease in young children is high; by adulthood, however, most malarial infections are asymptomatic. Constant, frequent, year-round infection is termed *stable transmission.* In areas where transmission is low, erratic, or focal, full protective immunity is not acquired, and symptomatic disease may occur at all ages. This situation usually exists in hypoendemic areas and is termed *unstable transmission*. Even in stable transmission areas, there is often an increased incidence of symptomatic malaria coinciding with increased mosquito breeding and transmission during the rainy season. Malaria can behave like an epidemic disease in some areas, particularly those with unstable malaria, such as northern India (the state of Rajasthan), Sri Lanka, Iraq, Turkey, the horn of Africa, Rwanda, Burundi, southern Africa, Madagascar, and central Asia. An epidemic can develop when there are changes in environmental, economic, or social conditions, such as heavy rains following drought or migrations (usually of refugees or workers) from a nonmalarious region to an area of high transmission; a breakdown in malaria control and prevention services can intensify epidemic conditions. This situation usually results in considerable mortality among all age groups.

The principal determinants of the epidemiology of malaria are the number (density), the human-biting habits, and the

longevity of the anopheline mosquito vectors. Not all of the >400 anophelines can transmit malaria, and the ~40 species that do so vary considerably in their efficiency as malaria vectors. More specifically, the transmission of malaria is directly proportional to the density of the vector, the square of the number of human bites per day per mosquito, and the tenth power of the probability of the mosquito's surviving for 1 day. Mosquito longevity is particularly important because the portion of the parasite's life cycle that takes place within the mosquito—from gametocyte ingestion to subsequent inoculation (*sporogony*)—lasts 8–30 days, depending on ambient temperature; thus, to transmit malaria, the mosquito must survive for >7 days. Sporogony is not completed at cooler temperatures—i.e., <16°C for *P. vivax* and <21°C for *P. falciparum*; thus transmission does not occur below these temperatures, although malaria outbreaks and transmission have occurred in the highlands (>1500 m) of east Africa, which were previously free of vectors. The most effective mosquito vectors of malaria are those, such as *Anopheles gambiae* in Africa, which are long-lived, occur in high densities in tropical climates, breed readily, and bite humans in preference to other animals. The entomologic inoculation rate (EIR; the number of sporozoite-positive mosquito bites per person per year) is the most common measure of malaria transmission and varies from <1 in some parts of Latin America and Southeast Asia to >300 in parts of tropical Africa.

ERYTHROCYTE CHANGES IN MALARIA

After invading an erythrocyte, the growing malarial parasite progressively consumes and degrades intracellular proteins, principally hemoglobin. The potentially toxic heme is detoxified by lipid-mediated crystallization to biologically inert hemozoin (malaria pigment). The parasite also alters the RBC membrane by changing its transport properties, exposing cryptic surface antigens, and inserting new parasite-derived proteins. The RBC becomes more irregular in shape, more antigenic, and less deformable.

In *P. falciparum* infections, membrane protuberances appear on the erythrocyte's surface 12–15 h after the cell's invasion. These "knobs" extrude a high-molecular-weight, antigenically variant, strain-specific erythrocyte membrane adhesive protein (PfEMP1) that mediates attachment to receptors on venular and capillary endothelium—an event termed *cytoadherence.* Several vascular receptors have been identified, of which intercellular adhesion molecule 1 (ICAM-1) is probably the most important in the brain, chondroitin sulfate B in the placenta, and CD36 in most other organs. Thus, the infected erythrocytes stick inside and eventually block capillaries and venules. At the same stage, these *P. falciparum*–infected RBCs may also adhere to uninfected RBCs (to form rosettes) and to other parasitized erythrocytes (agglutination). The processes of cytoadherence, rosetting, and agglutination are central to the pathogenesis of falciparum malaria. They result in the sequestration of RBCs containing mature forms of the parasite in vital organs (particularly the brain), where they interfere with microcirculatory flow and metabolism. Sequestered parasites continue to develop out of reach of the principal host defense mechanism: splenic processing and filtration. As a consequence, only the younger ring forms of the asexual parasites are seen circulating in the peripheral blood in falciparum malaria, and the level of peripheral parasitemia underestimates the true number of parasites within the body. Severe malaria is also associated with reduced deformability of the uninfected erythrocytes, which compromises their passage through the partially obstructed capillaries and venules and shortens RBC survival.

In the other three ("benign") human malarias, sequestration does not occur, and all stages of the parasite's development are evident on peripheral-blood smears. Whereas *P. vivax*, *P. ovale*, and *P. malariae* show a marked predilection for either young RBCs (*P. vivax*, *P. ovale*) or old cells (*P. malariae*) and produce a level of parasitemia that is seldom >2%, *P. falciparum* can invade erythrocytes of all ages and may be associated with very high levels of parasitemia.

HOST RESPONSE

Initially, the host responds to plasmodial infection by activating nonspecific defense mechanisms. Splenic immunologic and filtrative clearance functions are augmented in malaria, and the removal of both parasitized and uninfected erythrocytes is accelerated. The parasitized cells escaping splenic removal are destroyed when the schizont ruptures. The material released induces the activation of macrophages and the release of proinflammatory mononuclear cell–derived cytokines, which cause fever and exert other pathologic effects. Temperatures of ≥40°C damage mature parasites; in untreated infections, the effect of such temperatures is to further synchronize the parasitic cycle, with eventual production of the regular fever spikes and rigors that originally served to characterize the different malarias. These regular fever patterns (tertian, every 2 days; quartan, every 3 days) are seldom seen today in patients who receive prompt and effective antimalarial treatment.

The geographic distributions of sickle cell disease, hemoglobins C and E, hereditary ovalocytosis, the thalassemias, and glucose-6-phosphate dehydrogenase (G6PD) deficiency closely resemble that of falciparum malaria before the introduction of control measures. This similarity suggests that these genetic disorders confer protection against death from falciparum malaria. For example, HbA/S heterozygotes (sickle cell trait) have a sixfold reduction in the risk of dying from severe falciparum malaria. This decrease in risk appears to be related to impaired parasite growth at low oxygen tensions and reduced parasitized red cell cytoadherence. Parasite multiplication in HbA/E heterozygotes is reduced at high parasite densities. In Melanesia, children with α-thalassemia appear to have more frequent malaria (both vivax and falciparum) in the early years of life, and this pattern of infection appears to protect them against severe disease. In Melanesian ovalocytosis, rigid erythrocytes resist merozoite invasion, and the intraerythrocytic milieu is hostile.

Nonspecific host defense mechanisms stop the infection's expansion, and the subsequent strain-specific immune response then controls the infection. Eventually, exposure to sufficient strains confers protection from high-level parasitemia and disease but not from infection. As a result of this state of infection without illness (*premunition*), asymptomatic parasitemia is common among adults and older children living in regions with stable and intense transmission (i.e., holo- or hyperendemic areas). Immunity is mainly specific for both the species and the strain of infecting malarial parasite. Both humoral immunity and cellular immunity are necessary for protection, but the mechanisms of each are incompletely understood (Fig. 210-1). Immune individuals have a polyclonal increase in serum levels of IgM, IgG, and IgA, although much of this antibody is unrelated to protection. Antibodies to a variety of parasitic antigens presumably act in concert to limit in vivo replication of the parasite. In the case of falciparum malaria, the most important of these antigens is the surface adhesin—the variant protein PfEMP1 mentioned above. Passively transferred IgG from immune adults has been shown to reduce levels of parasitemia in children; although parasitemia in very young infants can occur, passive transfer of maternal antibody contributes to the relative (but not complete) protection of infants from severe malaria in the first months of life. This complex immunity to disease declines when a person lives outside an endemic area for several months or longer.

Several factors retard the development of cellular immunity to malaria. These factors include the absence of major histocompatibility antigens on the surface of infected RBCs, which precludes direct T cell recognition; malaria antigen–specific immune unresponsiveness; and the enormous strain diversity of malarial parasites, along with the ability of the parasites to express variant immunodominant antigens on the erythrocyte surface that change during the period of infection. Parasites may persist in the blood for months (or, in the case of *P. malariae*, for many years) if treatment is not given. The complexity of the immune response in malaria, the sophistication of the parasites' evasion mechanisms, and the lack of a good in vitro correlate with clinical immunity have all slowed progress toward an effective vaccine.

CLINICAL FEATURES

Malaria is a very common cause of fever in tropical countries. The first symptoms of malaria are nonspecific; the lack of a sense of well-being, headache, fatigue, abdominal discomfort, and muscle aches followed by fever are all similar to the symptoms of a minor viral illness. In some instances, a prominence of headache, chest pain, abdominal pain, arthralgia, myalgia, or diarrhea may suggest another diagnosis. Although headache may be severe in malaria, there is not neck stiffness or photophobia as occurs in meningitis. While myalgia may be prominent, it is not usually as severe as in dengue fever, and the muscles are not tender as in leptospirosis or typhus. Nausea, vomiting, and orthostatic hypotension are common. The classic malarial paroxysms, in which fever spikes, chills, and rigors occur at regular intervals, are relatively unusual and suggest infection with *P. vivax* or *P. ovale*. The fever is irregular at first (that of falciparum malaria may never become regular); the temperature of nonimmune individuals and children often rises above 40°C in conjunction with tachycardia and sometimes delirium. Although childhood febrile convulsions may occur with any of the malarias, generalized seizures are specifically associated with falciparum malaria and may herald the development of encephalopathy (cerebral malaria). Many clinical abnormalities have been described in acute malaria, but most patients with uncomplicated infections have few abnormal physical findings other than fever, malaise, mild anemia, and (in some cases) a palpable spleen. Anemia is common among young children living in areas with stable transmission, particularly where resistance has compromised the efficacy of antimalarial drugs. In nonimmune individuals with acute malaria, the spleen takes several days to become palpable, but splenic enlargement is found in a high proportion of otherwise healthy individuals in malaria-endemic areas and reflects repeated infections. Slight enlargement of the liver is also common, particularly among young children. Mild jaundice is common among adults; it may develop in patients with otherwise uncomplicated malaria and usually resolves over 1–3 weeks. Malaria is not associated with a rash like those seen in meningococcal septicemia, typhus, enteric fever, viral exanthems, and drug reactions. Petechial hemorrhages in the skin or mucous membranes—features of viral hemorrhagic fevers and leptospirosis—develop only rarely in severe falciparum malaria.

Severe Falciparum Malaria

Appropriately and promptly treated, uncomplicated falciparum malaria (i.e., the patient can swallow medicines and food) carries a mortality rate of ~0.1%. However, once vital-organ dysfunction occurs or the total proportion of erythrocytes infected increases to >2% (a level corresponding to $>10^{12}$ parasites in an adult), mortality risk rises steeply. The major manifestations of severe falciparum malaria are shown in Table 210-2, and features indicating a poor prognosis are listed in Table 210-3.

Table 210-2 Manifestations of Severe Falciparum Malaria

Signs	Manifestations
Major	
Unarousable coma/cerebral malaria	Failure to localize or respond appropriately to noxious stimuli; coma persisting for >30 min after generalized convulsion
Acidemia/acidosis	Arterial pH of <7.25 or plasma bicarbonate level of <15 mmol/L; venous lactate level of >5 mmol/L; manifests as labored deep breathing, often termed "respiratory distress"
Severe normochromic, normocytic anemia	Hematocrit of <15% or hemoglobin level of <50 g/L (<5 g/dL) with parasitemia level of >100,000/μL
Renal failure	Urine output (24 h) of <400 mL in adults or <12 mL/kg in children; no improvement with rehydration; serum creatinine level of >265 μmol/L (>3 mg/dL)
Pulmonary edema/adult respiratory distress syndrome	Noncardiogenic pulmonary edema, often aggravated by overhydration
Hypoglycemia	Plasma glucose level of <2.2 mmol/L (<40 mg/dL)
Hypotension/shock	Systolic blood pressure of <50 mmHg in children 1–5 years or <80 mmHg in adults; core/skin temperature difference of >10°C; capillary refill >2 s
Bleeding/disseminated intravascular coagulation	Significant bleeding and hemorrhage from the gums, nose, and gastrointestinal tract and/or evidence of disseminated intravascular coagulation
Convulsions	More than two generalized seizures in 24 h; signs of continued seizure activity sometimes subtle (e.g., tonic-clonic eye movements without limb or face movement)
Hemoglobinuria[a]	Macroscopic black, brown, or red urine; not associated with effects of oxidant drugs and red blood cell enzyme defects (such as G6PD deficiency)
Other	
Impaired consciousness/arousable	Unable to sit or stand without support
Extreme weakness	Prostration; inability to sit unaided[b]
Hyperparasitemia	Parasitemia level of >5% in nonimmune patients (>20% in any patient)
Jaundice	Serum bilirubin level of >50 mmol/L (>3 mg/dL) if combined with other evidence of vital-organ dysfunction

[a] Hemoglobinuria may occur in uncomplicated malaria.

[b] In a child who is normally able to sit.

Abbreviation: G6PD, glucose-6-phosphate dehydrogenase.

Table 210-3 Features Indicating a Poor Prognosis in Severe Falciparum Malaria

Clinical
Marked agitation
Hyperventilation (respiratory distress)
Hypothermia (<36.5°C)
Bleeding
Deep coma

- Repeated convulsions
- Anuria
- Shock

Laboratory

- Biochemistry
  - Hypoglycemia (<2.2 mmol/L)
  - Hyperlactatemia (>5 mmol/L)
  - Acidosis (arterial pH <7.3, serum HCO_3 <15 mmol/L)
  - Elevated serum creatinine (>265 μmol/L)
  - Elevated total bilirubin (>50 μmol/L)
  - Elevated liver enzymes (AST/ALT 3 times upper limit of normal)
  - Elevated muscle enzymes (CPK ↑, myoglobin ↑)
  - Elevated urate (>600 μmol/L)
- Hematology
  - Leukocytosis (>12,000/μL)
  - Severe anemia (PCV <15%)
  - Coagulopathy
    - Decreased platelet count (<50,000/μL)
    - Prolonged prothrombin time (>3 s)
    - Prolonged partial thromboplastin time
    - Decreased fibrinogen (<200 mg/dL)
- Parasitology
  - Hyperparasitemia
    - Increased mortality at >100,000/μL
    - High mortality at >500,000/μL
    - >20% of parasites identified as pigment-containing trophozoites and schizonts
    - >5% of neutrophils with visible pigment

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; CPK, creatine phosphokinase; PCV, packed cell volume.

CEREBRAL MALARIA

Coma is a characteristic and ominous feature of falciparum malaria and, despite treatment, is associated with death rates of ~20% among adults and 15% among children. Any obtundation, delirium, or abnormal behavior should be taken very seriously. The onset may be gradual or sudden following a convulsion.

Cerebral malaria manifests as diffuse symmetric encephalopathy; focal neurologic signs are unusual. Although some passive resistance to head flexion may be detected, signs of meningeal irritation are absent. The eyes may be divergent and a pout reflex is common, but other primitive reflexes are usually absent. The corneal reflexes are preserved, except in deep coma. Muscle tone may be either increased or decreased. The tendon reflexes are variable, and the plantar reflexes may be flexor or extensor; the abdominal and cremasteric reflexes are absent. Flexor or extensor posturing may be seen. On routine funduscopy, ~15% of patients have retinal hemorrhages; with pupillary dilatation and indirect ophthalmoscopy, this figure increases to 30–40%. Other funduscopic abnormalities (Fig. 210-3) include discrete spots of retinal opacification (30–60%), papilledema (8% among children, rare among adults), cotton wool spots (<5%), and decolorization of a retinal vessel or segment of vessel (occasional cases). Convulsions, usually generalized and often repeated, occur in ~10% of adults and up to 50% of children with cerebral malaria. More covert seizure activity is also common, particularly among children, and may manifest as repetitive tonic-clonic eye movements or even hypersalivation. Whereas adults rarely (i.e., in <3% of cases) suffer neurologic sequelae, ~5% of children surviving cerebral malaria—especially those with hypoglycemia, severe anemia, repeated seizures, and deep coma—have residual neurologic deficit when they regain consciousness; hemiplegia, cerebral palsy, cortical blindness, deafness, and impaired cognition and learning (all of varying duration) have been reported. The majority of these deficits improve markedly or resolve completely within 6 months. Approximately 10% of children surviving cerebral malaria have a persistent language deficit. The incidence of epilepsy is increased and the life expectancy decreased among these children.

Figure 210-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The eye in cerebral malaria: perimacular whitening and pale-centered retinal hemorrhages. *(Courtesy of N. Beare, T. Taylor, S. Harding, S. Lewallen, and M. Molyneux; with permission.)*

HYPOGLYCEMIA

Hypoglycemia, an important and common complication of severe malaria, is associated with a poor prognosis and is particularly problematic in children and pregnant women. Hypoglycemia in malaria results from a failure of hepatic gluconeogenesis and an increase in the consumption of glucose by both host and, to a much lesser extent, the malaria parasites. To compound the situation, quinine (and quinidine), which is still widely used for the treatment of both severe and uncomplicated falciparum malaria, is a powerful stimulant of pancreatic insulin secretion. Hyperinsulinemic hypoglycemia is especially troublesome in pregnant women receiving quinine treatment. In severe disease, the clinical diagnosis of hypoglycemia is difficult: the usual physical signs (sweating, gooseflesh, tachycardia) are absent, and the neurologic impairment caused by hypoglycemia cannot be distinguished from that caused by malaria.

ACIDOSIS

Acidosis, an important cause of death from severe malaria, results from accumulation of organic acids. Hyperlactatemia commonly coexists with hypoglycemia. In adults, coexisting renal impairment often compounds the acidosis; in children, ketoacidosis may also contribute. Other still-unidentified organic acids are major contributors to acidosis. Acidotic breathing, sometimes called respiratory distress, is a sign of poor prognosis. It is often followed by circulatory failure refractory to volume expansion or inotropic drug treatment and ultimately by respiratory arrest. The plasma concentrations of bicarbonate or lactate are the best biochemical prognosticators in severe malaria. Lactic acidosis is caused by the combination of anaerobic glycolysis in tissues where sequestered parasites interfere with microcirculatory flow, hypovolemia, lactate production by the parasites, and a failure of hepatic and renal lactate clearance. The prognosis of severe acidosis is poor.

NONCARDIOGENIC PULMONARY EDEMA

Adults with severe falciparum malaria may develop noncardiogenic pulmonary edema even after several days of antimalarial therapy. The pathogenesis of this variant of the adult respiratory distress syndrome is unclear. The mortality rate is >80%. This condition can be aggravated by overly vigorous administration of IV fluid. Noncardiogenic pulmonary edema can also develop in otherwise uncomplicated vivax malaria, where recovery is usual.

RENAL IMPAIRMENT

Renal impairment is common among adults with severe falciparum malaria but rare among children. The pathogenesis of renal failure is unclear but may be related to erythrocyte sequestration and agglutination interfering with renal microcirculatory flow and metabolism. Clinically and pathologically, this syndrome manifests as acute tubular necrosis. Renal cortical necrosis never develops. Acute renal failure may occur simultaneously with other vital-organ dysfunction (in which case the mortality risk is high) or may progress as other disease manifestations resolve. In survivors, urine flow resumes in a median of 4 days, and serum creatinine levels return to normal in a mean of 17 days (Chap. 279). Early dialysis or hemofiltration considerably enhances the likelihood of a patient's survival, particularly in acute hypercatabolic renal failure.

HEMATOLOGIC ABNORMALITIES

Anemia results from accelerated RBC removal by the spleen, obligatory RBC destruction at parasite schizogony, and ineffective erythropoiesis. In severe malaria, both infected and uninfected RBCs show reduced deformability, which correlates with prognosis and development of anemia. Splenic clearance of all RBCs is increased. In nonimmune individuals and in areas with unstable transmission, anemia can develop rapidly and transfusion is often required. As a consequence of repeated malarial infections, children in many areas of Africa may develop severe anemia resulting from both shortened survival of uninfected RBCs and marked dyserythropoiesis. Anemia is a common consequence of antimalarial drug resistance, which results in repeated or continued infection.

Slight coagulation abnormalities are common in falciparum malaria, and mild thrombocytopenia is usual. Of patients with severe malaria, <5% have significant bleeding with evidence of disseminated intravascular coagulation. Hematemesis from stress ulceration or acute gastric erosions may also occur rarely.

LIVER DYSFUNCTION

Mild hemolytic jaundice is common in malaria. Severe jaundice is associated with *P. falciparum* infections; is more common among adults than among children; and results from hemolysis, hepatocyte injury, and cholestasis. When accompanied by other vital-organ dysfunction (often renal impairment), liver dysfunction carries a poor prognosis. Hepatic dysfunction contributes to hypoglycemia, lactic acidosis, and impaired drug metabolism. Occasional patients with falciparum malaria may

develop deep jaundice (with hemolytic, hepatic, and cholestatic components) without evidence of other vital-organ dysfunction.

OTHER COMPLICATIONS

HIV/AIDS predisposes to more severe malaria in nonimmune individuals. Malaria anemia is worsened by concurrent infections with intestinal helminths, hookworm in particular. Septicemia may complicate severe malaria, particularly in children. Differentiating severe malaria from sepsis with incidental parasitemia in childhood is very difficult. In endemic areas, *Salmonella* bacteremia has been associated specifically with *P. falciparum* infections. Chest infections and catheter-induced urinary tract infections are common among patients who are unconscious for >3 days. Aspiration pneumonia may follow generalized convulsions. The frequency of complications of severe falciparum malaria is summarized in Table 210-4.

Table 210-4 Relative Incidence of Severe Complications of Falciparum Malaria

Complication	Nonpregnant Adults	Pregnant Women	Children
Anemia	+	++	+++
Convulsions	+	+	+++
Hypoglycemia	+	+++	+++
Jaundice	+++	+++	+
Renal failure	+++	+++	–
Pulmonary edema	++	+++	+

Note: –, rare; +, infrequent; ++, frequent; +++, very frequent.

Malaria in Pregnancy

In areas of high malaria transmission, falciparum malaria in primi- and secundigravid women is associated with low birth weight (average reduction, ~170 g) and consequently increased infant and childhood mortality rates. In general, infected mothers in areas of stable transmission remain asymptomatic despite intense accumulation of parasitized erythrocytes in the placental microcirculation. Maternal HIV infection predisposes pregnant women to malaria, predisposes their newborns to congenital malarial infection, and exacerbates the reduction in birth weight associated with malaria.

In areas with unstable transmission of malaria, pregnant women are prone to severe infections and are particularly vulnerable to high-level parasitemia with anemia, hypoglycemia, and acute pulmonary edema. Fetal distress, premature labor, and stillbirth or low birth weight are common results. Fetal death is usual in severe malaria. Congenital malaria occurs in <5% of newborns whose mothers are infected; its frequency and the level of parasitemia are related directly to the parasite density in maternal blood and in the placenta. *P. vivax* malaria in pregnancy is also associated with a reduction in birth weight (average, 110 g), but, in contrast to the situation in falciparum malaria, this effect is more pronounced in multigravid than in primigravid women. About 150,000 women die in childbirth yearly, with most deaths occurring in low-income countries; maternal death from hemorrhage at childbirth is correlated with malaria-induced anemia.

Malaria in Children

Most of the nearly 1 million persons who die of falciparum malaria each year are young African children. Convulsions, coma, hypoglycemia, metabolic acidosis, and severe anemia are relatively common among children with severe malaria, whereas deep jaundice, acute renal failure, and acute pulmonary edema are unusual. Severely anemic children may present with labored deep breathing, which in the past has been attributed incorrectly to "anemic congestive cardiac failure" but in fact is usually caused by metabolic acidosis, often compounded by hypovolemia. In general, children tolerate antimalarial drugs well and respond rapidly to treatment.

Transfusion Malaria

Malaria can be transmitted by blood transfusion, needle-stick injury, sharing of needles by infected injection drug users, or organ transplantation. The incubation period in these settings is often short because there is no preerythrocytic stage of development. The clinical features and management of these cases are the same as for naturally acquired infections. Radical chemotherapy with primaquine is unnecessary for transfusion-transmitted *P. vivax* and *P. ovale* infections.

CHRONIC COMPLICATIONS OF MALARIA

Tropical Splenomegaly (Hyperreactive Malarial Splenomegaly)

Chronic or repeated malarial infections produce hypergammaglobulinemia; normochromic, normocytic anemia; and, in certain situations, splenomegaly. Some residents of malaria-endemic areas in tropical Africa and Asia exhibit an abnormal immunologic response to repeated infections that is characterized by massive splenomegaly, hepatomegaly, marked elevations in serum titers of IgM and malarial antibody, hepatic sinusoidal lymphocytosis, and (in Africa) peripheral B cell lymphocytosis. This syndrome has been associated with the production of cytotoxic IgM antibodies to CD8+ T lymphocytes, antibodies to CD5+ T lymphocytes, and an increase in the ratio of CD4+ T cells to CD8+ T cells. These events may lead to uninhibited B cell production of IgM and the formation of cryoglobulins (IgM aggregates and immune complexes). This immunologic process stimulates reticuloendothelial hyperplasia and clearance activity and eventually produces splenomegaly. Patients with hyperreactive malarial splenomegaly (HMS) present with an abdominal mass or a dragging sensation in the abdomen and occasional sharp abdominal pains suggesting perisplenitis. Anemia and some degree of pancytopenia are usually evident, and in some cases malarial parasites cannot be found in peripheral-blood smears. Vulnerability to respiratory and skin infections is increased; many patients die of overwhelming sepsis. Persons with HMS who are living in endemic areas should receive antimalarial chemoprophylaxis; the results are usually good. In nonendemic areas, antimalarial treatment is advised. In some cases refractory to therapy, clonal lymphoproliferation may develop and then evolve into a malignant lymphoproliferative disorder.

Quartan Malarial Nephropathy

Chronic or repeated infections with *P. malariae* (and possibly with other malarial species) may cause soluble immune-complex injury to the renal glomeruli, resulting in the nephrotic syndrome. Other unidentified factors must contribute to this process since only a very small proportion of infected patients develop renal disease. The histologic appearance is that of focal or segmental glomerulonephritis with splitting of the capillary basement membrane. Subendothelial dense deposits are seen on electron microscopy, and immunofluorescence reveals deposits of complement and immunoglobulins; in samples of renal tissue from children, *P. malariae* antigens are often visible. A coarse-granular pattern of basement membrane immunofluorescent deposits (predominantly IgG3) with selective proteinuria carries a better prognosis than a fine-granular,

predominantly IgG2 pattern with nonselective proteinuria. Quartan nephropathy usually responds poorly to treatment with either antimalarial agents or glucocorticoids and cytotoxic drugs.

Burkitt's Lymphoma and Epstein-Barr Virus Infection

It is possible that malaria-related immune dysregulation provokes infection with lymphoma viruses. Burkitt's lymphoma is strongly associated with Epstein-Barr virus. The prevalence of this childhood tumor is high in malarious areas of Africa.

DIAGNOSIS

Demonstration of the Parasite

The diagnosis of malaria rests on the demonstration of asexual forms of the parasite in stained peripheral-blood smears. After a negative blood smear, repeat smears should be made if there is a high degree of suspicion. Of the Romanowsky stains, Giemsa at pH 7.2 is preferred; Field's, Wright's, or Leishman's stain can also be used. Both thin (Figs. 210-4 and 210-5; see also Figs. e27-3 and e27-4) and thick (Figs. 210-6, 210-7, 210-8, and 210-9) blood smears should be examined. The thin blood smear should be rapidly air-dried, fixed in anhydrous methanol, and stained; the RBCs in the tail of the film should then be examined under oil immersion (x1000 magnification). The level of parasitemia is expressed as the number of parasitized erythrocytes per 1000 RBCs. The thick blood film should be of uneven thickness. The smear should be dried thoroughly and stained without fixing. As many layers of erythrocytes overlie one another and are lysed during the staining procedure, the thick film has the advantage of concentrating the parasites (by 40- to 100-fold compared with a thin blood film) and thus increasing diagnostic sensitivity. Both parasites and white blood cells (WBCs) are counted, and the number of parasites per unit volume is calculated from the total leukocyte count. Alternatively, a WBC count of 8000/μL is assumed. This figure is converted to the number of parasitized erythrocytes per microliter. A minimum of 200 WBCs should be counted under oil immersion. Interpretation of blood smear films requires some experience because artifacts are common. Before a thick smear is judged to be negative, 100–200 fields should be examined under oil immersion. In high-transmission areas, the presence of up to 10,000 parasites/μL of blood may be tolerated without symptoms or signs in partially immune individuals. Thus in these areas the detection of malaria parasites is sensitive but has low specificity in identifying malaria as the cause of illness. Low-density parasitemia is common in other conditions causing fever.

Figure 210-4

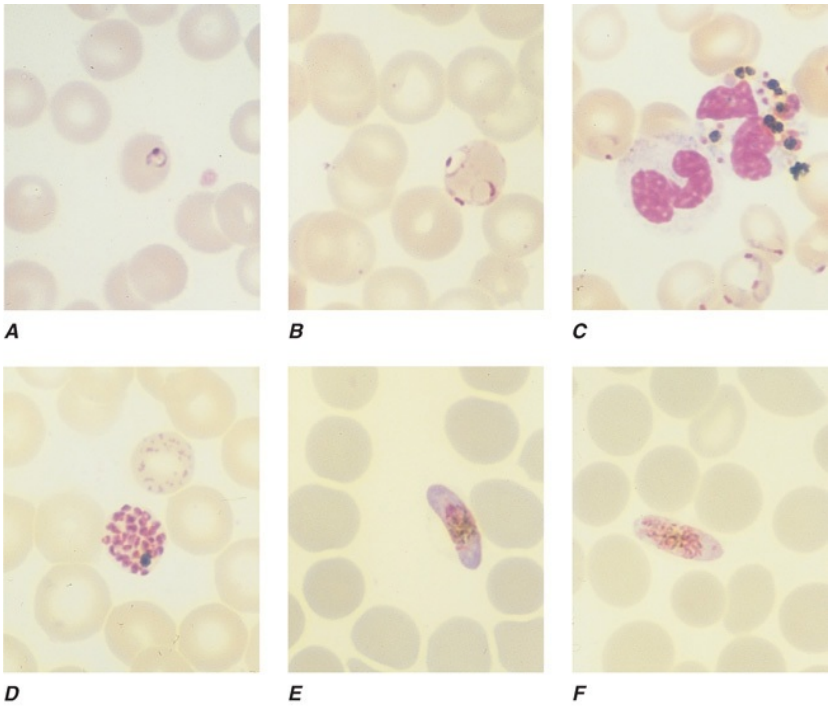

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thin blood films of *Plasmodium falciparum*. *A*. Young trophozoites. ***B.*** Old trophozoites. ***C.*** Pigment in polymorphonuclear cells and trophozoites. ***D.*** Mature schizonts. ***E.*** Female gametocytes. ***F.*** Male gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure 210-5

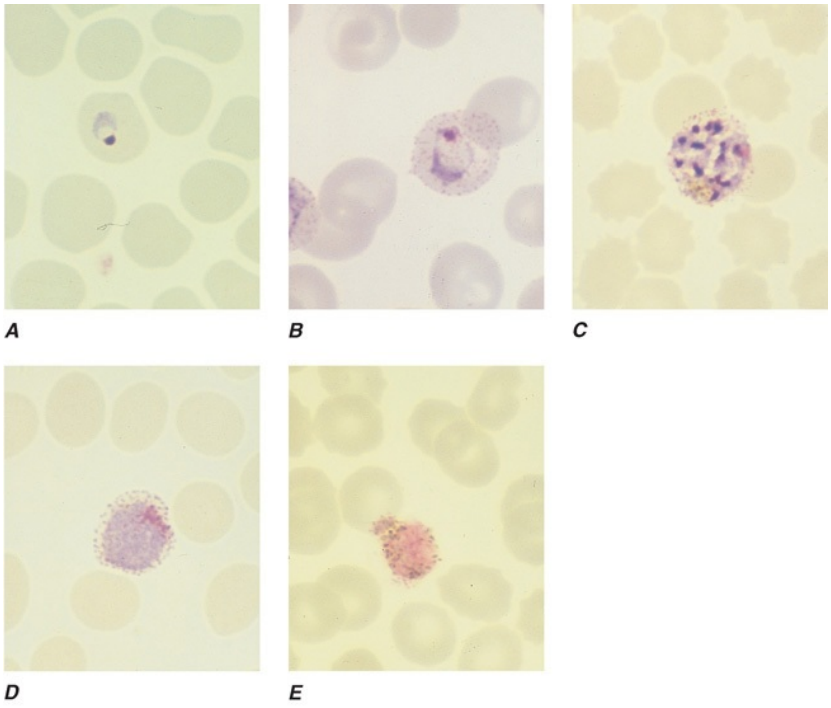

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thin blood films of *Plasmodium vivax*. *A*. Young trophozoites. ***B.*** Old trophozoites. ***C.*** Mature schizonts. ***D.*** Female gametocytes. ***E.*** Male gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure 210-6

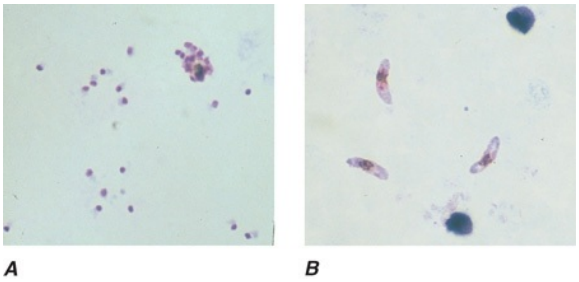

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium falciparum*. *A.* Trophozoites. ***B.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure 210-7

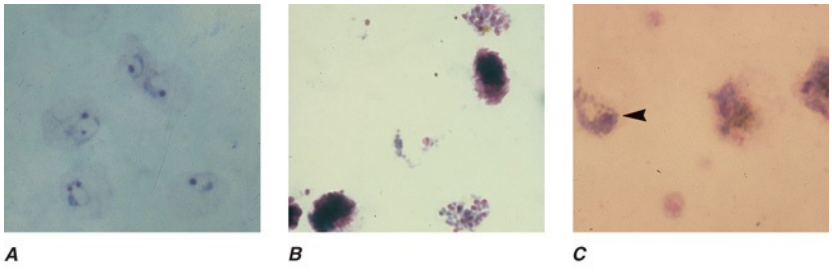

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium vivax*. *A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure 210-8

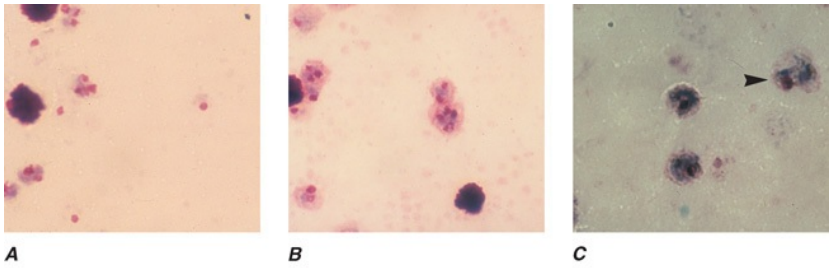

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium ovale*. *A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure 210-9

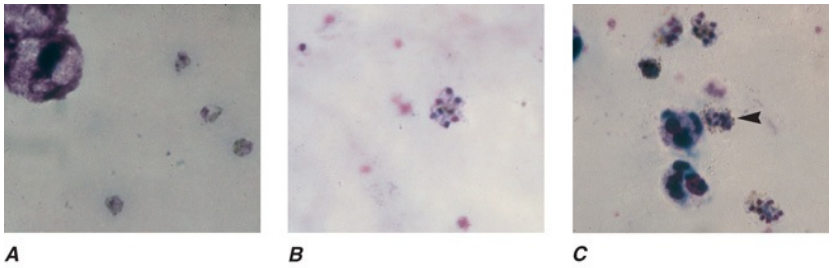

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium malariae*. *A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Rapid, simple, sensitive, and specific antibody-based diagnostic stick or card tests that detect *P. falciparum*–specific, histidine-rich protein 2 (PfHRP2) or lactate dehydrogenase antigens in finger-prick blood samples are now being used widely in control programs (Table 210-5). Some of these rapid diagnostic tests (RDTs) carry a second antibody, which allows falciparum malaria to be distinguished from the less dangerous malarias. PfHRP2-based tests may remain positive for several weeks after acute infection. This feature is a disadvantage in high-transmission areas where infections are frequent but is of value in the diagnosis of severe malaria in patients who have taken antimalarial drugs and cleared peripheral parasitemia (but in whom the PfHRP2 test remains strongly positive). RDTs are replacing microscopy in many areas because of their simplicity and speed, but they are relatively expensive and do not quantify parasitemia.

Table 210-5 Methods for the Diagnosis of Malaria[a]

Method	Procedure	Advantages	Disadvantages
Thick blood film[b]	Blood should be uneven in thickness but sufficiently thin to read the hands of a watch through part of the spot. Stain dried, unfixed blood spot with Giemsa, Field's, or another Romanowsky stain. Count number of asexual parasites per 200 WBCs (or per 500 at low densities). Count gametocytes separately.[c]	Sensitive (0.001% parasitemia); species specific; inexpensive	Requires experience (artifacts may be misinterpreted as low-level parasitemia); underestimates true count
Thin blood film[d]	Stain fixed smear with Giemsa, Field's, or another Romanowsky stain. Count number of RBCs containing asexual parasites per 1000 RBCs. In severe malaria, assess stage of parasite development and count neutrophils containing malaria pigment.[e] Count gametocytes separately.[c]	Rapid; species specific; inexpensive; in severe malaria, provides prognostic information[e]	Insensitive (<0.05% parasitemia); uneven distribution of *P. vivax*, as enlarged infected red cells concentrate at leading edge
PfHRP2 dipstick or card test	A drop of blood is placed on the stick or card, which is then immersed in washing solutions. Monoclonal antibody captures the parasite antigen and reads out as a colored band.	Robust and relatively inexpensive; rapid; sensitivity similar to or slightly lower than that of thick films (~0.001% parasitemia)	Detects only *Plasmodium falciparum*; remains positive for weeks after infection[f]; does not quantitate *P. falciparum* parasitemia
Plasmodium LDH dipstick or card test	A drop of blood is placed on the stick or card, which is then immersed in washing solutions. Monoclonal antibodies capture the parasite antigens and read out as colored bands. One band is genus specific (all malarias), and the other is specific for *P. falciparum.*	Rapid; sensitivity similar to or slightly lower than that of thick films for *P. falciparum* (~0.001% parasitemia)	Slightly more difficult preparation than PfHRP2 tests; may miss low-level parasitemia with *P. vivax, P. ovale*, and *P. malariae* and does not speciate these organisms; does not quantitate *P. falciparum* parasitemia
Microtube concentration methods with acridine	Blood is collected in a specialized tube containing acridine orange, anticoagulant, and a float. After centrifugation, which concentrates the parasitized cells around the float, fluorescence microscopy is	Sensitivity similar or superior to that of thick films (~0.001% parasitemia); ideal	Does not speciate or quantitate; requires fluorescence microscopy

orange staining	performed.	for processing large numbers of samples rapidly	

[a]Malaria cannot be diagnosed clinically with accuracy, but treatment should be started on clinical grounds if laboratory confirmation is likely to be delayed. In areas of the world where malaria is endemic and transmission is high, low-level asymptomatic parasitemia is common in otherwise healthy people. Thus malaria may not be the cause of a fever, although in this context the presence of >10,000 parasites/μL (~0.2% parasitemia) does indicate that malaria is the cause. Antibody and polymerase chain reaction tests have no role in the diagnosis of malaria except that PCR is increasingly used for genotyping and speciation in mixed infections.

[b]Asexual parasites/200 WBCs x 40 = parasite count/μL (assumes a WBC count of 8000/μL). See Figs. 210-6, 210-7, 210-8, and 210-9.

[c]Gametocytemia may persist for days or weeks after clearance of asexual parasites. Gametocytemia without asexual parasitemia does not indicate active infection.

[d]Parasitized RBCs (%) x hematocrit x 1256 = parasite count/μL. See Figs. 210-4 and 210-5.

[e]The presence of >100,000 parasites/μL (~2% parasitemia) is associated with an increased risk of severe malaria, but some patients have severe malaria with lower counts. At any level of parasitemia, the finding that >50% of parasites are tiny rings (cytoplasm width less than half of nucleus width) carries a relatively good prognosis. The presence of visible pigment in >20% of parasites or of phagocytosed pigment in >5% of polymorphonuclear leukocytes (indicating massive recent schizogony) carries a worse prognosis.

[f]Persistence of PfHRP2 is a disadvantage in high-transmission settings, where many asymptomatic people have positive tests, but can be used to diagnostic advantage in low-transmission settings when a sick patient has received previous unknown treatment (which, in endemic areas, often consists of antimalarial drugs). A positive PfHRP2 test indicates that the illness is falciparum malaria, even if the blood smear is negative.

Abbreviations: LDH, lactate dehydrogenase; PfHRP2, *P. falciparum* histidine-rich protein 2; RBCs, red blood cells; WBCs, white blood cells.

The relationship between parasitemia and prognosis is complex; in general, patients with >10^5 parasites/μL are at increased risk of dying, but nonimmune patients may die with much lower counts, and partially immune persons may tolerate parasitemia levels many times higher with only minor symptoms. In severe malaria, a poor prognosis is indicated by a predominance of more mature *P. falciparum* parasites (i.e., >20% of parasites with visible pigment) in the peripheral-blood film or by the presence of phagocytosed malarial pigment in >5% of neutrophils. In *P. falciparum* infections, gametocytemia peaks 1 week after the peak of asexual parasites. Because the mature gametocytes of *P. falciparum* are not affected by most antimalarial drugs, their persistence does not constitute evidence of drug resistance. Phagocytosed malarial pigment is sometimes seen inside peripheral-blood monocytes or polymorphonuclear leukocytes and may provide a clue to recent infection if malaria parasites are not detectable. After the clearance of the parasites, this intraphagocytic malarial pigment is often evident for several days in the peripheral blood or for longer in bone marrow aspirates or smears of fluid expressed after intradermal puncture. Staining of parasites with the fluorescent dye acridine orange allows more rapid diagnosis of malaria (but not speciation of the infection) in patients with low-level parasitemia.

Laboratory Findings

Normochromic, normocytic anemia is usual. The leukocyte count is generally normal, although it may be raised in very severe infections. There is slight monocytosis, lymphopenia, and eosinopenia, with reactive lymphocytosis and eosinophilia in the weeks after the acute infection. The erythrocyte sedimentation rate, plasma viscosity, and levels of C-reactive protein and other acute-phase proteins are high. The platelet count is usually reduced to ~10^5/μL. Severe infections may be accompanied by prolonged prothrombin and partial thromboplastin times and by more severe thrombocytopenia. Levels of antithrombin III are reduced even in mild infection. In uncomplicated malaria, plasma concentrations of electrolytes, blood urea nitrogen (BUN), and creatinine are usually normal. Findings in severe malaria may include metabolic acidosis, with low plasma concentrations of glucose, sodium, bicarbonate, calcium, phosphate, and albumin together with elevations in lactate, BUN, creatinine, urate, muscle and liver enzymes, and conjugated and unconjugated bilirubin. Hypergammaglobulinemia is usual in immune and semi-immune subjects. Urinalysis generally gives normal results. In adults and children with cerebral malaria, the mean opening pressure at lumbar puncture is ~160 mm of cerebrospinal fluid (CSF); usually the CSF is normal or has a slightly elevated total protein level [<1.0 g/L (<100 mg/dL)] and cell count (<20/μL).

Treatment: Malaria

(Table 210-6) When a patient in or from a malarious area presents with fever, thick and thin blood smears should be prepared and *examined immediately* to confirm the diagnosis and identify the species of infecting parasite (Figs. 210-4 through 210-9). Repeat blood smears should be performed at least every 12–24 h for 2 days if the first smears are negative and malaria is strongly suspected. Alternatively, a rapid antigen detection card or stick test should be performed. Patients with severe malaria or those unable to take oral drugs should receive parenteral antimalarial therapy. If there is any doubt about the resistance status of the infecting organism, it should be considered resistant. Antimalarial drug susceptibility testing can be performed but is rarely available, has poor predictive value in an individual case, and yields results too slowly to influence the choice of treatment. Several drugs are available for oral treatment. The choice of drug depends on the likely sensitivity of the infecting parasites. Despite increasing evidence of chloroquine resistance in *P. vivax* (from parts of Indonesia, Oceania, eastern and southern Asia, and Central and South America), chloroquine remains the treatment of choice for the non-falciparum malarias (*P. vivax*, *P. ovale*, *P. malariae*, *P. knowlesi*) except in Indonesia and Papua New Guinea, where high levels of resistance in *P. vivax* are prevalent.

Table 210-6 Regimens for the Treatment of Malaria

Type of Disease or Treatment	Regimen(s)
Uncomplicated Malaria	
Known chloroquine-sensitive strains of *Plasmodium vivax, P. malariae, P. ovale, P. knowlesi, P. falciparum*[a]	Chloroquine (10 mg of base/kg stat followed by 5 mg/kg at 12, 24, and 36 h or by 10 mg/kg at 24 h and 5 mg/kg at 48 h) *or* Amodiaquine (10–12 mg of base/kg qd for 3 days)

Radical treatment for *P. vivax or P. ovale* infection	In addition to chloroquine or amodiaquine as detailed above, primaquine (0.5 mg of base/kg qd) should be given for 14 days to prevent relapse. In mild G6PD deficiency, 0.75 mg of base/kg should be given once weekly for 6–8 weeks. Primaquine should not be given in severe G6PD deficiency.
Sensitive *P. falciparum* malaria[b]	Artesunate[c] (4 mg/kg qd for 3 days) plus sulfadoxine (25 mg/kg)/pyrimethamine (1.25 mg/kg) as a single dose *or* Artesunate[c] (4 mg/kg qd for 3 days) plus amodiaquine (10 mg of base/kg qd for 3 days)[d]
Multidrug-resistant *P. falciparum* malaria	Either artemether-lumefantrine[c] (1.5/9 mg/kg bid for 3 days with food) *or* Artesunate[c] (4 mg/kg qd for 3 days) *plus* Mefloquine (25 mg of base/kg—either 8 mg/kg qd for 3 days or 15 mg/kg on day 2 and then 10 mg/kg on day 3)[d]
Second-line treatment/treatment of imported malaria	Either artesunate[c] (2 mg/kg qd for 7 days) or quinine (10 mg of salt/kg tid for 7 days) *plus 1 of the following 3:* 1. Tetracycline[e] (4 mg/kg qid for 7 days) 2. Doxycycline[e] (3 mg/kg qd for 7 days) 3. Clindamycin (10 mg/kg bid for 7 days) *or* Atovaquone-proguanil (20/8 mg/kg qd for 3 days with food)
Severe Falciparum Malaria[f]	
	Artesunate[c] (2.4 mg/kg stat IV followed by 2.4 mg/kg at 12 and 24 h and then daily if necessary)[g] *or, if unavailable, one of the following:* Artemether[c] (3.2 mg/kg stat IM followed by 1.6 mg/kg qd) *or* Quinine dihydrochloride (20 mg of salt/kg[h] infused over 4 h, followed by 10 mg of salt/kg infused over 2–8 h q8h[i]) *or* Quinidine (10 mg of base/kg[h] infused over 1–2 h, followed by 1.2 mg of base/kg per hour[i] with electrocardiographic monitoring)

[a] Very few areas now have chloroquine-sensitive *P. falciparum* malaria (Fig. 210-2).

[b] In areas where the partner drug to artesunate is known to be effective.

[c] Artemisinin derivatives are not readily available in some temperate countries.

[d] Fixed-dose coformulated combinations are available. The World Health Organization now recommends artemisinin combination regimens as first-line therapy for falciparum malaria in all tropical countries and advocates use of fixed-dose combinations.

[e] Tetracycline and doxycycline should not be given to pregnant women or to children <8 years of age.

[f] Oral treatment should be substituted as soon as the patient recovers sufficiently to take fluids by mouth.

[g] Artesunate is the drug of choice when available. The data from large studies in Southeast Asia showed a 35% lower mortality rate than with quinine, and very large studies in Africa showed a 22.5% reduction in mortality rate compared with quinine.

[h] A loading dose should not be given if therapeutic doses of quinine or quinidine have definitely been administered in the previous 24 h. Some authorities recommend a lower dose of quinidine.

[i] Infusions can be given in 0.9% saline and 5–10% dextrose in water. Infusion rates for quinine and quinidine should be carefully controlled.

Abbreviation: G6PD, glucose-6-phosphate dehydrogenase.

The treatment of falciparum malaria has changed radically in recent years. In all endemic areas, the World Health Organization (WHO) now recommends artemisinin-based combinations as first-line treatment for uncomplicated falciparum malaria. These rapidly and reliably effective drugs are sometimes unavailable in temperate countries, where treatment recommendations are limited by the registered available drugs. Fake or substandard antimalarials are commonly sold in many Asian and African countries. Thus, careful attention is required at the time of purchase and later, especially when the patient fails to respond as expected. Characteristics of antimalarial drugs are shown in Table 210-7.

Table 210-7 Properties of Antimalarial Drugs

Drug(s)	Pharmacokinetic Properties	Antimalarial Activity	Minor Toxicity	Major Toxicity
Quinine,	Good oral and IM absorption	Acts mainly on	*Common:* "Cinchonism":	*Common:* Hypoglycemia

quinidine	(quinine); *Cl* and V_d reduced, but plasma protein binding (principally to ∞1 acid glycoprotein) increased (90%) in malaria; quinine $t_1/_2$: 16 h in malaria, 11 h in healthy persons; quinidine $t_1/_2$: 13 h in malaria, 8 h in healthy persons	trophozoite blood stage; kills gametocytes of *P. vivax*, *P. ovale*, and *P. malariae* (but not *P. falciparum*); no action on liver stages	tinnitus, high-tone hearing loss, nausea, vomiting, dysphoria, postural hypotension; ECG QT_C interval prolongation (quinine usually by <10% but quinidine by up to 25%) *Rare:* Diarrhea, visual disturbance, rashes *Note:* Very bitter taste	*Rare:* Hypotension, blindness, deafness, cardiac arrhythmias, thrombocytopenia, hemolysis, hemolytic-uremic syndrome, vasculitis, cholestatic hepatitis, neuromuscular paralysis *Note:* Quinidine more cardiotoxic
Chloroquine	Good oral absorption, very rapid IM and SC absorption; complex pharmacokinetics; enormous *Cl* and V_d (unaffected by malaria); blood concentration profile determined by distribution processes in malaria; $t_1/_2$: 1–2 months	As for quinine but acts slightly earlier in asexual cycle	*Common:* Nausea, dysphoria, pruritus in dark-skinned patients, postural hypotension *Rare:* Accommodation difficulties, keratopathy, rash *Note:* Bitter taste, well tolerated	*Acute:* Hypotensive shock (parenteral), cardiac arrhythmias, neuropsychiatric reactions *Chronic:* Retinopathy (cumulative dose, >100 g), skeletal and cardiac myopathy
Piperaquine	Adequate oral absorption, enhanced by fats; similar pharmacokinetics to chloroquine; $t_1/_2$: 21–28 days	As for chloroquine, but retains activity against multidrug-resistant *P. falciparum*	Epigastric pain, diarrhea, slight ECG QT_C prolongation	None identified
Amodiaquine	Good oral absorption; largely converted to active metabolite desethylamodiaquine	As for chloroquine	Nausea (tastes better than chloroquine)	Agranulocytosis; hepatitis, mainly with prophylactic use; should not be used with efavirenz
Primaquine	Complete oral absorption; active metabolite not known; $t_1/_2$: 7 h	Radical cure; eradicates hepatic forms of *P. vivax* and *P. ovale;* kills all stages of gametocyte development of *P. falciparum*	Nausea, vomiting, diarrhea, abdominal pain, hemolysis, methemoglobinemia	Massive hemolysis in subjects with severe G6PD deficiency
Mefloquine	Adequate oral absorption; no parenteral preparation; $t_1/_2$: 14–20 days (shorter in malaria)	As for quinine	Nausea, giddiness, dysphoria, fuzzy thinking, sleeplessness, nightmares, sense of dissociation	Neuropsychiatric reactions, convulsions, encephalopathy
Halofantrine[b]	Highly variable absorption related to fat intake; $t_1/_2$: 1–3 days (active desbutyl metabolite $t_1/_2$?: 3–7 days)	As for quinine	Diarrhea	Cardiac conduction disturbances; atrioventricular block; ECG QT_C interval prolongation; potentially lethal ventricular tachyarrhythmias
Lumefantrine	Highly variable absorption related to fat intake; $t_1/_2$: 3–4 days	As for quinine	None identified	None identified
Artemisinin and derivatives (artemether, artesunate)	Good oral absorption, slow and variable absorption of IM artemether; artesunate and artemether biotransformed to active metabolite dihydroartemisinin; all drugs eliminated very rapidly; $t_1/_2$: <1 h	Broader stage specificity and more rapid than other drugs; no action on liver stages; kills all but fully mature gametocytes of *P. falciparum*	Reduction in reticulocyte count (but not anemia); neutropenia at high doses	Anaphylaxis, urticaria, fever
Pyrimethamine	Good oral absorption, variable IM absorption; $t_1/_2$: 4 days	For blood stages, acts mainly on mature forms; causal prophylactic	Well tolerated	Megaloblastic anemia, pancytopenia, pulmonary infiltration
Proguanil (chloroguanide)	Good oral absorption; biotransformed to active metabolite cycloguanil; $t_1/_2$: 16 h; biotransformation reduced by oral contraceptive use and in pregnancy	Causal prophylactic; not used alone for treatment	Well tolerated; mouth ulcers and rare alopecia	Megaloblastic anemia in renal failure
Atovaquone	Highly variable absorption related to fat intake; $t_{1/2}$: 30–70 h	Acts mainly on trophozoite blood stage	None identified	None identified
Tetracycline, doxycycline[a]	Excellent absorption; $t_{1/2}$: 8 h for tetracycline, 18 h for doxycycline	Weak antimalarial activity; should not be used alone for treatment	Gastrointestinal intolerance, deposition in growing bones and teeth, photosensitivity	Renal failure in patients with impaired renal function tetracycline)

		moniliasis, benign intracranial hypertension

[a]Tetracycline and doxycycline should not be given to pregnant women or to children <8 years of age.

[b]Halofantrine should not be used by patients with long ECG QT[c] intervals or known conduction disturbances or by those taking drugs that may affect ventricular repolarization, e.g., quinidine, quinine, mefloquine, chloroquine, neuroleptics, antiarrhythmics, tricyclic antidepressants, terfenidine, or astemizole.

Abbreviations: *Cl*, systemic clearance; ECG, electrocardiogram; G6PD, glucose-6-phosphate dehydrogenase; V_d, total apparent volume of distribution.

Severe Malaria

In large studies conducted in Asia, parenteral artesunate, a water-soluble artemisinin derivative, has been shown to reduce mortality rates in severe falciparum malaria among adults by 35% from rates obtained with quinine. Recently, the largest trial ever in severe malaria showed that parenteral artesunate reduced the mortality rate among African children by 22.5% compared with that obtained with quinine. Artesunate has, therefore, become the drug of choice for all patients with severe malaria everywhere. Artesunate is given by IV injection but can also be given by IM injection. Artemether and the closely related drug artemotil (arteether) are oil-based formulations given by IM injection; they are erratically absorbed and do not confer the same survival benefit as artesunate. A rectal formulation of artesunate has been developed as a community-based pre-referral treatment for patients in the rural tropics who cannot take oral medications. Pre-referral administration of rectal artesunate has been shown to decrease mortality risk among severely ill children in communities without access to immediate parenteral treatment. Although the artemisinin compounds are safer than quinine and considerably safer than quinidine, only one formulation is available in the United States. IV artesunate has been approved by the U.S. Food and Drug Administration for emergency use against severe malaria through the Centers for Disease Control and Prevention (CDC) Drug Service (see end of chapter for contact information). The antiarrhythmic quinidine gluconate is as effective as quinine and, as it was more readily available, replaced quinine for the treatment of malaria in the United States. The administration of quinidine must be closely monitored if dysrhythmias and hypotension are to be avoided. A total plasma level >8 μg/mL, a QT_C interval >0.6 s, or QRS widening beyond 25% of baseline is an indication for slowing infusion rates. If arrhythmia or saline-unresponsive hypotension develops, treatment with this drug should be discontinued. Quinine is safer than quinidine; cardiovascular monitoring is not required except when the recipient has cardiac disease.

Severe falciparum malaria constitutes a medical emergency requiring intensive nursing care and careful management. The patient should be weighed and, if comatose, placed on his or her side or prone. Frequent evaluation of the patient's condition is essential. Ancillary drugs such as high-dose glucocorticoids, urea, heparin, dextran, desferrioxamine, antibody to tumor necrosis factor α, and high-dose phenobarbital (20 mg/kg) have proved either ineffective or harmful in clinical trials and should not be used. In acute renal failure or severe metabolic acidosis, hemofiltration or hemodialysis should be started as early as possible.

Parenteral antimalarial treatment should be started as soon as possible. Artesunate, given by either IV or IM injection, is the agent of choice; it is simple to administer, safe, and rapidly effective. If artesunate is unavailable and artemether, quinine, or quinidine is used, an initial loading dose must be given so that therapeutic concentrations are reached as soon as possible. Both quinine and quinidine will cause dangerous hypotension if injected rapidly; when given IV, they must be administered carefully by rate-controlled infusion only. If this approach is not possible, quinine may be given by deep IM injections into the anterior thigh. The optimal therapeutic range for quinine and quinidine in severe malaria is not known with certainty, but total plasma concentrations of 8–15 mg/L for quinine and 3.5–8.0 mg/L for quinidine are effective and do not cause serious toxicity. The systemic clearance and apparent volume of distribution of these alkaloids are markedly reduced and plasma protein binding is increased in severe malaria, so that the blood concentrations attained with a given dose are higher. If the patient remains seriously ill or in acute renal failure for >2 days, maintenance doses of quinine or quinidine should be reduced by 30–50% to prevent toxic accumulation of the drug. The initial doses should never be reduced. If one of the artemisinin derivatives is given, dose reductions are unnecessary, even in renal failure. Exchange transfusion may be considered for severely ill patients, although the precise indications for this procedure have not been agreed upon. It has been recommended that—if safe and feasible—exchange should be considered for patients with severe malaria, but there is no clear evidence that this measure is beneficial, particularly if artesunate is used. The role of prophylactic anticonvulsants in children is also uncertain. If respiratory support is not available, then a full loading dose of phenobarbital (20 mg/kg) to prevent convulsions should not be given as it may cause respiratory arrest.

When the patient is unconscious, the blood glucose level should be measured every 4–6 h. All patients should receive a continuous infusion of dextrose, and blood concentrations ideally should be maintained above 4 mmol/L. Hypoglycemia (<2.2 mmol/L or 40 mg/dL) should be treated immediately with bolus glucose. The parasite count and hematocrit level should be measured every 6–12 h. Anemia develops rapidly; if the hematocrit falls to <20%, then whole blood (preferably fresh) or packed cells should be transfused slowly, with careful attention to circulatory status. Renal function should be checked daily. Children presenting with severe anemia and acidotic breathing are often hypovolemic; in this situation, resuscitation with crystalloids or blood is indicated. Accurate assessment is vital. Management of fluid balance is difficult in severe malaria, particularly in adults, because of the thin dividing line between overhydration (leading to pulmonary edema) and underhydration (contributing to renal impairment). As soon as the patient can take fluids, oral therapy should be substituted for parenteral treatment.

Uncomplicated Malaria

Infections due to *P. vivax*, *P. knowlesi*, *P. malariae*, and *P. ovale* should be treated with oral chloroquine (total dose, 25 mg of base/kg). In much of the tropics, drug-resistant *P. falciparum* has been increasing in distribution, frequency, and intensity. It is now accepted that, to prevent resistance, falciparum malaria should be treated with drug combinations and not with single drugs in endemic areas; the same rationale has been applied successfully to the treatment of tuberculosis, HIV/AIDS, and cancers. This combination strategy is based on simultaneous use of two or more drugs with different modes of action. Artemisinin combination treatment (ACT) regimens are now recommended as first-line treatment for falciparum malaria throughout the malaria-affected world. The artemisinin component is usually an artemisinin derivative (artesunate, artemether, or dihydroartemisinin) given for 3 days, and the partner drug is usually a slower-acting antimalarial to which *P. falciparum* is sensitive. Five ACT regimens are currently recommended by the WHO. In areas with multidrug-resistant falciparum malaria (parts of Asia and South America, including those with mefloquine-resistant parasites; Fig. 210-10), artemether-lumefantrine, artesunate-mefloquine, or dihydroartemisinin-piperaquine should be used; these regimens provide cure rates of >90%. In areas with sensitive parasites, the aforementioned combinations, artesunate-sulfadoxine-pyrimethamine, or artesunate-amodiaquine may also be used. Atovaquone-proguanil is also highly effective everywhere, although it is seldom used in endemic areas because of its high cost. Of great concern is the emergence of artemisinin-resistant *P. falciparum* in western

Cambodia and adjacent Thailand. Infections with these parasites are cleared slowly from the blood, with clearance times typically exceeding 3 days.

The 3-day ACT regimens are all well tolerated, although mefloquine is associated with increased rates of vomiting and dizziness. As second-line treatments for recrudescence following first-line therapy, a different ACT regimen may be given; another alternative is a 7-day course of either artesunate or quinine plus tetracycline, doxycycline, or clindamycin. Tetracycline and doxycycline cannot be given to pregnant women or to children <8 years of age. Oral quinine is extremely bitter and regularly produces cinchonism comprising tinnitus, high-tone deafness, nausea, vomiting, and dysphoria. Adherence is poor with the required 7-day regimens of quinine.

Patients should be monitored for vomiting for 1 h after the administration of any oral antimalarial drug. If there is vomiting, the dose should be repeated. Symptom-based treatment, with tepid sponging and acetaminophen administration, lowers fever and thereby reduces the patient's propensity to vomit these drugs. Minor central nervous system reactions (nausea, dizziness, sleep disturbances) are common. The incidence of serious adverse neuropsychiatric reactions to mefloquine treatment is ~1 in 1000 in Asia but may be as high as 1 in 200 among Africans and Caucasians. All the antimalarial quinolines (chloroquine, mefloquine, and quinine) exacerbate the orthostatic hypotension associated with malaria, and all are tolerated better by children than by adults. Pregnant women, young children, patients unable to tolerate oral therapy, and nonimmune individuals (e.g., travelers) with suspected malaria should be evaluated carefully and hospitalization considered. If there is any doubt as to the identity of the infecting malarial species, treatment for falciparum malaria should be given. A negative blood smear makes malaria unlikely but does not rule it out completely; thick blood films should be checked again 1 and 2 days later to exclude the diagnosis. Nonimmune patients receiving treatment for malaria should have daily parasite counts performed until the thick films are negative. If the level of parasitemia does not fall below 25% of the admission value in 48 h or if parasitemia has not cleared by 7 days (and adherence is assured), drug resistance is likely and the regimen should be changed.

To eradicate persistent liver stages and prevent relapse (radical treatment), primaquine (0.5 mg of base/kg, adult dose) should be given daily for 14 days to patients with *P. vivax* or *P. ovale* infections after laboratory tests for G6PD deficiency have proved negative. If the patient has a mild variant of G6PD deficiency, primaquine can be given in a dose of 0.75 mg of base/kg (45 mg maximum) once weekly for 6 weeks. Pregnant women with vivax or ovale malaria should not be given primaquine but should receive suppressive prophylaxis with chloroquine (5 mg of base/kg per week) until delivery, after which radical treatment can be given.

Complications

ACUTE RENAL FAILURE

If the level of BUN or creatinine rises despite adequate rehydration, fluid administration should be restricted to prevent volume overload. As in other forms of hypercatabolic acute renal failure, renal replacement therapy is best performed early (Chap. 279). Hemofiltration and hemodialysis are more effective than peritoneal dialysis and are associated with lower mortality. Some patients with renal impairment pass small volumes of urine sufficient to allow control of fluid balance; these cases can be managed conservatively if other indications for dialysis do not arise. Renal function usually improves within days, but full recovery may take weeks.

ACUTE PULMONARY EDEMA (ACUTE RESPIRATORY DISTRESS SYNDROME)

Patients should be positioned with the head of the bed at a 45° elevation and given oxygen and IV diuretics. Pulmonary artery occlusion pressures may be normal, indicating increased pulmonary capillary permeability. Positive-pressure ventilation should be started early if the immediate measures fail (Chap. 235).

HYPOGLYCEMIA

An initial slow injection of 50% dextrose (0.5 g/kg) should be followed by an infusion of 10% dextrose (0.10 g/kg per hour). The blood glucose level should be checked regularly thereafter as recurrent hypoglycemia is common, particularly among patients receiving quinine or quinidine. In severely ill patients, hypoglycemia commonly occurs together with metabolic (lactic) acidosis and carries a poor prognosis.

OTHER COMPLICATIONS

Patients who develop spontaneous bleeding should be given fresh blood and IV vitamin K. Convulsions should be treated with IV or rectal benzodiazepines and, if necessary, respiratory support. Aspiration pneumonia should be suspected in any unconscious patient with convulsions, particularly with persistent hyperventilation; IV antimicrobial agents and oxygen should be administered, and pulmonary toilet should be undertaken. Hypoglycemia or gram-negative septicemia should be suspected when the condition of any patient suddenly deteriorates for no obvious reason during antimalarial treatment. In malaria-endemic areas where a high proportion of children are parasitemic, it is usually impossible to distinguish severe malaria from bacterial sepsis with confidence. It is increasingly accepted that these children should be treated with both antimalarials and broad-spectrum antibiotics from the outset. Because nontyphoidal *Salmonella* infections are particularly common, empirical antibiotics should be selected to cover these organisms. Antibiotics should be considered for severely ill patients of any age who are not responding to antimalarial treatment.

PREVENTION

In recent years, considerable progress has been made in malaria prevention, control, and research. New drugs have been discovered and developed, and one vaccine candidate has reached the stage of advanced field testing. Highly effective drugs, insecticide-treated nets, and insecticides for spraying dwellings are being purchased for endemic countries by the Global Fund to Fight AIDS, Tuberculosis, and Malaria; the President's Malaria Initiative; UNICEF; and other organizations. Malaria research and control are being strongly supported by the National Institute of Allergy and Infectious Diseases, the Wellcome Trust, the Bill & Melinda Gates Foundation, the WHO, the Multilateral Initiative on Malaria, the Roll Back Malaria Partnership, and the Global Health Council, among others. Still, the eradication of malaria is not feasible in the immediate future because of the widespread distribution of *Anopheles* breeding sites; the great number of infected persons; the continued use of ineffective antimalarial drugs; and inadequacies in human and material resources, infrastructure, and control programs. The call for and commitment to ultimate eradication of malaria by the Gates Foundation in 2007—seconded by Margaret Chan, Director General of the WHO—added great impetus to all malaria initiatives, especially those aimed at discovery and implementation of new interventions and other critical goals. Malaria may be contained by judicious use of insecticides to kill the mosquito vector, rapid diagnosis, appropriate patient management, and—where effective and feasible—administration of intermittent preventive treatment or chemoprophylaxis to high-risk groups such as pregnant women, young children, and travelers from nonendemic regions. Malaria researchers are intensifying their efforts to gain a better understanding of parasite-human-mosquito interactions and to develop more effective control and prevention interventions. Despite the enormous investment in efforts to develop a malaria vaccine and the 30–60% efficacy of a recombinant protein sporozoite-targeted adjuvanted vaccine in limited field trials, no safe, effective, long-lasting vaccine is likely to be available for

general use in the near future (Chap. 122). While there is great promise for one or more malaria vaccines on the more distant horizon, prevention and control measures continue to rely on antivector and drug-use strategies.

Personal Protection Against Malaria

Simple measures to reduce the frequency of infected-mosquito bites in malarious areas are very important. These measures include the avoidance of exposure to mosquitoes at their peak feeding times (usually dusk to dawn) as well as the use of insect repellents containing 10–35% DEET (or, if DEET is unacceptable, 7% picaridin), suitable clothing, and insecticide-impregnated bed nets or other materials. Widespread use of bed nets treated with residual pyrethroids reduces the incidence of malaria in areas where vectors bite indoors at night and has been shown to reduce mortality rates in western and eastern Africa.

Chemoprophylaxis

(Table 210-8; *wwwnc.cdc.gov/travel/yellowbook/2010/chapter-2/malaria.aspx*) Recommendations for prophylaxis depend on knowledge of local patterns of *Plasmodium* species drug sensitivity and the likelihood of acquiring malarial infection. When there is uncertainty, drugs effective against resistant *P. falciparum* should be used [atovaquone-proguanil (Malarone), doxycycline, or mefloquine]. Chemoprophylaxis is never entirely reliable, and malaria should always be considered in the differential diagnosis of fever in patients who have traveled to endemic areas, even if they are taking prophylactic antimalarial drugs.

Table 210-8 Drugs Used in the Prophylaxis of Malaria

Drug	Usage	Adult Dose	Pediatric Dose	Comments
Atovaquone/proguanil (Malarone)	Prophylaxis in areas with chloroquine- or mefloquine-resistant *Plasmodium falciparum*	1 adult tablet PO[a]	5–8 kg: 1/2 pediatric tablet[b] daily ≥8–10 kg: 3/4 pediatric tablet daily ≥10–20 kg: 1 pediatric tablet daily ≥20–30 kg: 2 pediatric tablets daily ≥30–40 kg: 3 pediatric tablets daily ≥40 kg: 1 adult tablet daily	Begin 1–2 days before travel to malarious areas. Take daily at the same time each day while in the malarious areas and for 7 days after leaving such areas. Atovaquone-proguanil is contraindicated in persons with severe renal impairment (creatinine clearance rate <30 mL/min). In the absence of data, it is not recommended for children weighing <5 kg, pregnant women, or women breast-feeding infants weighing <5 kg. Atovaquone/proguanil should be taken with food or a milky drink.
Chloroquine phosphate (Aralen and generic)	Prophylaxis only in areas with chloroquine-sensitive *P. falciparum*[c] or *P. vivax* only	300 mg of base (500 mg of salt) PO once weekly	5 mg/kg of base (8.3 mg of salt/kg) PO once weekly, up to maximum adult dose of 300 mg of base	Begin 1–2 weeks before travel to malarious areas. Take weekly on the same day of the week while in the malarious areas and for 4 weeks after leaving such areas. Chloroquine phosphate may exacerbate psoriasis.
Doxycycline (many brand names and generic)	Prophylaxis in areas with chloroquine- or mefloquine-resistant *P. falciparum*[c]	100 mg PO qd	≥8 years of age: 2 mg/kg, up to adult dose	Begin 1–2 days before travel to malarious areas. Take daily at the same time each day while in the malarious areas and for 4 weeks after leaving such areas. Doxycycline is contraindicated in children <8 years of age and in pregnant women.
Hydroxychloroquine sulfate (Plaquenil)	An alternative to chloroquine for primary prophylaxis only in areas with chloroquine-sensitive *P. falciparum*[c] or *P. vivax* only	310 mg of base (400 mg of salt) PO once weekly	5 mg of base/kg (6.5 mg of salt/kg) PO once weekly, up to maximum adult dose of 310 mg of base	Begin 1–2 weeks before travel to malarious areas. Take weekly on the same day of the week while in the malarious areas and for 4 weeks after leaving such areas. Hydroxychloroquine may exacerbate psoriasis.
Mefloquine (Lariam and generic)	Prophylaxis in areas with chloroquine-resistant *P. falciparum*	228 mg of base (250 mg of salt) PO once weekly	≤9 kg: 4.6 mg of base/kg (5 mg of salt/kg) PO once	Begin 1–2 weeks before travel to malarious areas. Take weekly on the same day of the week while in the malarious areas and for 4 weeks after leaving such areas. Mefloquine is contraindicated in persons allergic to this drug or related compounds (e.g., quinine and quinidine) and in persons with active or recent depression, generalized anxiety disorder,

			weekly 10–19 kg: 1/4 tablet once weekly 20–30 kg: 1/2 tablet once weekly 31–45 kg: 3/4 tablet once weekly ≥46 kg: 1 tablet once weekly	psychosis, schizophrenia, other major psychiatric disorders, or seizures. Use with caution in persons with psychiatric disturbances or a history of depression. Mefloquine is not recommended for persons with cardiac conduction abnormalities.
Primaquine	For prevention of malaria in areas with mainly *P. vivax*	30 mg of base (52.6 mg of salt) PO qd	0.5 mg of base/kg (0.8 mg of salt/kg) PO qd, up to adult dose; should be taken with food	Begin 1–2 days before travel to malarious areas. Take daily at the same time each day while in the malarious areas and for 7 days after leaving such areas. Primaquine is contraindicated in persons with G6PD deficiency. It is also contraindicated during pregnancy and in lactation unless the infant being breast-fed has a documented normal G6PD level.
Primaquine	Used for presumptive antirelapse therapy (terminal prophylaxis) to decrease risk of relapses of *P. vivax and P. ovale*	30 mg of base (52.6 mg of salt) PO qd for 14 days after departure from the malarious area	0.5 mg of base/kg (0.8 mg of salt/kg), up to adult dose, PO qd for 14 days after departure from the malarious area	This therapy is indicated for persons who have had prolonged exposure to *P. vivax* and/or *P. ovale*. It is contraindicated in persons with G6PD deficiency as well as during pregnancy and in lactation unless the infant being breast-fed has a documented normal G6PD level.

[a]An adult tablet contains 250 mg of atovaquone and 100 mg of proguanil hydrochloride.

[b]A pediatric tablet contains 62.5 mg of atovaquone and 25 mg of proguanil hydrochloride.

[c]Very few areas now have chloroquine-sensitive malaria (Fig. 210-2).

Source:CDC: *http://wwwn.cdc.gov/travel/contentMalariaDrugsHC.aspx*.

Pregnant women traveling to malarious areas should be warned about the potential risks. All pregnant women at risk in endemic areas should be encouraged to attend regular antenatal clinics. Mefloquine is the only drug advised for pregnant women traveling to areas with drug-resistant malaria; this drug is generally considered safe in the second and third trimesters of pregnancy, and the data on first-trimester exposure, although limited, are reassuring. Chloroquine and proguanil are regarded as safe. The safety of other prophylactic antimalarial agents in pregnancy has not been established. Antimalarial prophylaxis has been shown to reduce mortality rates among children between the ages of 3 months and 4 years in malaria-endemic areas; however, it is not a logistically or economically feasible option in many countries. The alternative—to give intermittent treatment doses [intermittent preventive treatment (IPT)]—shows promise for more widespread use in infants, young children, and pregnant women. Children born to nonimmune mothers in endemic areas (usually expatriates moving to malaria-endemic areas) should receive prophylaxis from birth.

Travelers should start taking antimalarial drugs 2 days to 2 weeks before departure so that any untoward reactions can be detected and so that therapeutic antimalarial blood concentrations will be present when needed (Table 210-8). Antimalarial prophylaxis should continue for 4 weeks after the traveler has left the endemic area, except if atovaquone-proguanil or primaquine has been taken; these drugs have significant activities against the liver stage of the infection (causal prophylaxis) and can be discontinued 1 week after departure from the endemic area. If suspected malaria develops while a traveler is abroad, obtaining a reliable diagnosis and antimalarial treatment locally is a top priority. Presumptive self-treatment for malaria with atovaquone-proguanil (for 3 consecutive days) or another drug can be considered under special circumstances; medical advice on self-treatment should be sought before departure for malarious areas and as soon as possible after illness begins. Every effort should be made to confirm the diagnosis by parasitologic studies.

Atovaquone-proguanil (Malarone; 3.75/1.5 mg/kg or 250/100 mg, daily adult dose) is a fixed-combination, once-daily prophylactic agent that is very well tolerated by adults and children, with fewer adverse gastrointestinal effects than chloroquine-proguanil and fewer adverse central nervous system effects than mefloquine. It is proguanil itself, rather than the antifolate metabolite cycloguanil, that acts synergistically with atovaquone. This combination is effective against all types of malaria, including multidrug-resistant falciparum malaria. Atovaquone-proguanil is best taken with food or a milky drink to optimize absorption. There are insufficient data on the safety of this regimen in pregnancy.

Mefloquine (250 mg of salt weekly, adult dose) has been widely used for malarial prophylaxis because it is usually effective against multidrug-resistant falciparum malaria and is reasonably well tolerated. The drug has been associated with rare episodes of psychosis and seizures at prophylactic doses; these reactions are more frequent at the higher doses used for treatment. More common side effects with prophylactic doses of mefloquine include mild nausea, dizziness, fuzzy thinking, disturbed sleep patterns, vivid dreams, and malaise. The drug is contraindicated for use by travelers with known hypersensitivity to mefloquine or related compounds (e.g., quinine, quinidine) and by persons with active or recent depression, anxiety disorder, psychosis, schizophrenia, another major psychiatric disorder, or seizures; mefloquine is not recommended for persons with cardiac conduction abnormalities. The role of mefloquine prophylaxis during pregnancy remains uncertain; in studies in Africa, mefloquine prophylaxis was found to be effective and safe during pregnancy. However,

in one study from Thailand, treatment of malaria with mefloquine was associated with an increased risk of stillbirth.

Daily administration of doxycycline (100 mg daily, adult dose) is an effective alternative to atovaquone-proguanil or mefloquine. Doxycycline is generally well tolerated but may cause vulvovaginal thrush, diarrhea, and photosensitivity and cannot be used by children <8 years old or by pregnant women.

Chloroquine can no longer be relied upon to prevent *P. falciparum* infections in most areas but is used to prevent and treat malaria due to the other human *Plasmodium* species and for *P. falciparum* malaria in Central American countries west and north of the Panama Canal, Caribbean countries, and some countries in the Middle East. Chloroquine-resistant *P. vivax* has been reported from parts of eastern Asia, Oceania, and Central and South America. This drug is generally well tolerated, although some patients cannot take it because of malaise, headache, visual symptoms (due to reversible keratopathy), gastrointestinal intolerance, or pruritus. Chloroquine is considered safe in pregnancy. With chronic administration for >5 years, a characteristic dose-related retinopathy may develop, but this condition is rare at the doses used for antimalarial prophylaxis. Idiosyncratic or allergic reactions are also rare. Skeletal and/or cardiac myopathy is a potential problem with protracted prophylactic use; they are more likely to occur at the high doses used in the treatment of rheumatoid arthritis. Neuropsychiatric reactions and skin rashes are unusual. When used continuously, amodiaquine, a related aminoquinoline, is associated with a high risk of agranulocytosis (~1 person in 2000) and hepatotoxicity (~1 person in 16,000); thus this agent should not be used for prophylaxis.

Primaquine (daily adult dose, 0.5 mg of base/kg or 30 mg taken with food), an 8-aminoquinoline compound, has proved safe and effective in the prevention of drug-resistant falciparum and vivax malaria in adults. This drug can be considered for persons who are traveling to areas with or without drug-resistant *P. falciparum* and who are intolerant to other recommended drugs. Abdominal pain and oxidant hemolysis—the principal adverse effects—are not common as long as the drug is taken with food and is not given to G6PD-deficient persons, in whom it can cause hemolysis that is sometimes fatal. Travelers must be tested for G6PD deficiency and be shown to have a level in the normal range before receiving primaquine. Primaquine should not be given to pregnant women or neonates. The 8-aminoquinolines (primaquine, tafenoquine) given in a single dose with ACT are being considered for widespread use in treatment regimens in malaria elimination programs because of their gametocytocidal effect on *P. falciparum*.

In the past, the dihydrofolate reductase inhibitors pyrimethamine and proguanil (chloroguanide) were administered widely, but the rapid selection of resistance in both *P. falciparum* and *P. vivax* has limited their use. Whereas antimalarial quinolines such as chloroquine (a 4-aminoquinoline) act on the erythrocyte stage of parasitic development, the dihydrofolate reductase inhibitors also inhibit preerythrocytic growth in the liver (causal prophylaxis) and development in the mosquito (sporontocidal activity). Proguanil is safe and well tolerated, although mouth ulceration occurs in ~8% of persons using this drug; it is considered safe for antimalarial prophylaxis in pregnancy. The prophylactic use of the combination of pyrimethamine and sulfadoxine is not recommended because of an unacceptable incidence of severe toxicity, principally exfoliative dermatitis and other skin rashes, agranulocytosis, hepatitis, and pulmonary eosinophilia (incidence, 1:7000; fatal reactions, 1:18,000). The combination of pyrimethamine with dapsone (0.2/1.5 mg/kg weekly; 12.5/100 mg, adult dose) has been used in some countries. Dapsone may cause methemoglobinemia and allergic reactions and (at higher doses) may pose a significant risk of agranulocytosis. Proguanil and the pyrimethamine-dapsone combination are not available in the United States.

Because of the increasing spread and intensity of antimalarial drug resistance (Figs. 210-2 and 210-10), the CDC recommends that travelers and their providers consider their destination, type of travel, and current medications and health risks when choosing antimalarial chemoprophylaxis. Consultation for the evaluation of prophylaxis failures or treatment of malaria can be obtained from state and local health departments and the CDC Malaria Hotline (770-488-7788) or the CDC Emergency Operations Center (770-488-7100).

Figure 210-10

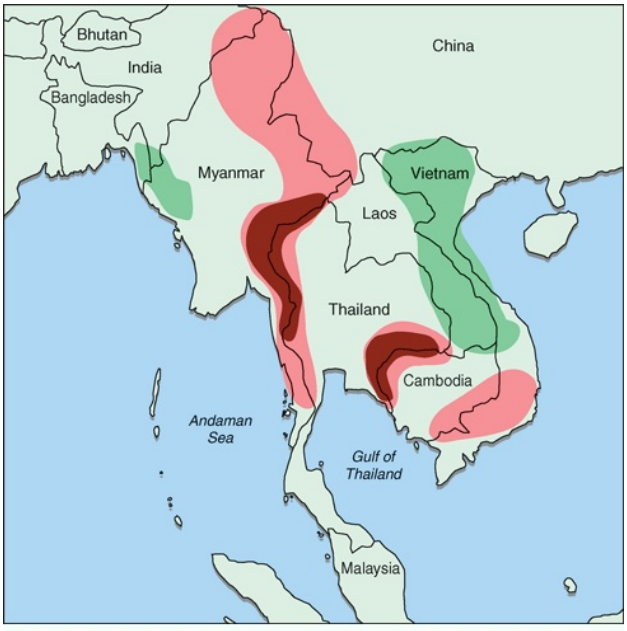

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mefloquine resistance in *Plasmodium falciparum* in Southeast Asia: high-level mefloquine resistance (*brown*), low-level mefloquine resistance (*red*), and mefloquine sensitivity (failure rate, <20%; *green*). There is insufficient information for other areas.

FURTHER READINGS

Baird JK et al: Prevention and treatment of vivax malaria. Curr Infect Dis Rep 9:39, 2007[PMID: 17254503]

Breman JG: Eradicating malaria. Sci Prog 92:1, 2009[PMID: 19544698]

Centers for Disease Control and Prevention: Treatment of malaria (guidelines for clinicians). Atlanta, Department of Health and Human Services, 2010; available online at *wwwnc.cdc.gov/travel/yellowbook/2010/chapter-2/malaria.aspx*

Dondorp A et al: Artesunate versus quinine in the treatment of severe falciparum malaria in African children (AQUAMAT): An open-label randomized trial. Lancet 376:1647, 2010 (erratum Lancet 377:126, 2011)

Gomes MS et al: Pre-referral rectal artesunate to prevent death and disability in severe malaria: A placebo-controlled trial. Lancet 373:557, 2009[PMID: 19059639]

White NJ: How antimalarial drug resistance affects post-treatment prophylaxis. Mal J 7:9, 2008[PMID: 18186948]

———: The assessment of antimalarial drug efficacy. Trends Parasitol 18:865, 2002

World Health Organization: *Guidelines for the Treatment of Malaria* , 2nd ed. Geneva, World Health Organization, 2010 *(who.int/malaria/publications/atoz/9789241547925/en/)*

———: Assessment and monitoring of antimalarial drug efficacy for the treatment of uncomplicated falciparum malaria. WHO/HTM/RBM/2003.50; available online at *www.who.int/malaria/publications/atoz/whohtmrbm200350/en/*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BABESIOSIS: INTRODUCTION

Babesiosis is a tick-borne infectious disease caused by parasites of the genus *Babesia*. These protozoans are obligate parasites of red blood cells (RBCs). Wild and domestic animals are the natural reservoirs of *Babesia*. Transmission to humans is incidental and was recognized only half a century ago. The vast majority of cases occur in the United States, where babesiosis has the status of an emerging infectious disease. Sporadic cases are reported in Europe and the rest of the world (see "Global Considerations," below).

Epidemiology

GEOGRAPHIC DISTRIBUTION

Babesia microti, a parasite of small rodents, is the etiologic agent of babesiosis in the northeastern United States. Highly endemic areas include Nantucket Island, Martha's Vineyard, Block Island, Shelter Island, eastern Long Island, and Fire Island. On the mainland, babesiosis is endemic in southeastern Massachusetts, coastal Rhode Island and Connecticut, central New Jersey, Wisconsin, and Minnesota. On the West Coast, the etiologic agent is *Babesia duncani*, a species closely related to those found in wildlife. The index case of infection with this species was reported from Washington State. Several cases were identified in northern California, and one case may have been contracted in central Oregon. Three cases of babesiosis caused by *Babesia divergens*–like organisms—from Washington State, Missouri, and Kentucky, respectively—have been reported.

PREVALENCE

The number of cases of *B. microti* illness has increased steadily over the last decade. In 2009, more than 700 cases were reported to the public health departments in endemic states. The prevalence of babesiosis caused by *B. microti* is underestimated because young healthy individuals typically experience a mild and self-limiting illness and may not seek medical attention. Accordingly, seroprevalence is much higher than the prevalence of clinical babesiosis.

Modes of Transmission

The nymphal stage of the deer tick *Ixodes scapularis* is the primary vector for transmission of *B. microti*. Transmission occurs from May through September. The incubation period is 1–6 weeks long, with three-fourths of cases presenting in June and July. The vectors for transmission of *B. duncani* and *B. divergens*–like organisms remain unknown.

Babesiosis is occasionally acquired by transfusion of blood products, primarily packed RBCs; more than 70 such cases have been caused by *B. microti*, and two cases caused by *B. duncani* have been transmitted by this route. The incubation period lasts 1–9 weeks. Most cases occur in endemic areas during fall and winter. Some cases are diagnosed in nonendemic areas to which blood products have been imported from endemic areas. Of the 11 transfusion-related babesiosis deaths reported to the U.S. Food and Drug Administration since 1998, 10 have occurred since 2005.

Three cases of congenital babesiosis have been attributed to *B. microti*. Other cases of neonatal babesiosis have been acquired by transfusion or tick bite.

Clinical Manifestations

MILD *B. MICROTI* ILLNESS

Patients experience a gradual onset of malaise, fatigue, and weakness. Fever exceeds 38°C, can reach 40.6°C, and is accompanied by one or several of the following: chills, sweats, headache, myalgia, anorexia, dry cough, arthralgia, and nausea. Less common symptoms include neck stiffness, sore throat, shortness of breath, abdominal pain, and weight loss. On physical examination, fever is the salient feature. Development of *erythema chronicum migrans* is suggestive of intercurrent Lyme disease. Ecchymoses and petechiae have been reported. Mild splenomegaly and hepatomegaly are occasionally noted. Lymphadenopathy is absent. Jaundice, slight pharyngeal erythema, retinal infarcts, and retinopathy with splinter hemorrhages are rare.

SEVERE *B. MICROTI* ILLNESS

Severe babesiosis is associated with parasitemia levels of >4% and requires hospitalization. Risk factors include an age of>50 years, male gender, asplenia, HIV/AIDS, malignancy, and immunosuppression. Compared with patients hospitalized for other febrile illnesses, patients with

severe babesiosis are more likely to report malaise, myalgia or arthralgia, and shortness of breath. Complications develop in ~40% of hospitalized patients. Risk factors for complications are severe anemia (hemoglobin level ≤10 g/dL) and high-level parasitemia (>10%). Acute respiratory distress syndrome is the most common complication. Other complications include disseminated intravascular coagulation, congestive heart failure, and renal failure. Splenic infarcts and rupture have been reported. Strong predictors of poor outcome—defined as hospitalization for >2 weeks, stay in an intensive care unit for >2 days, or death—are male gender, alkaline phosphatase levels of >125 U/L, and white blood cell (WBC) counts of >5 x 10^9/L. The fatality rate is 5% among all hospitalized patients but is much higher (20%) among immunocompromised patients.

Other Babesial Infections

Cases of *B. duncani* infection range in severity from asymptomatic to fatal. Clinical manifestations are those reported for *B. microti*. All three reported patients infected with *B. divergens*–like organisms required hospitalization; one died.

Diagnosis

A diagnosis of babesiosis should be considered for any patient who (1) presents with flu-like symptoms and has recently resided in or traveled to an endemic area or received a blood transfusion or (2) presents with symptoms of or has been diagnosed with Lyme disease or human granulocytotropic anaplasmosis.

Babesiosis is diagnosed by microscopic examination of Giemsa-stained thin blood smears, on which *Babesia* species appear as round or pear-shaped organisms. The ring form is most common and lacks the central brownish deposit (hemozoin) typical of *Plasmodium falciparum* trophozoites (Chap. e27). Other distinguishing features are the absence of schizonts and gametocytes and the occasional presence of tetrads ("Maltese crosses"), which are pathognomonic of infection with *B. microti* or *B. duncani* but are also noted in human RBCs infected with *B. divergens*–like organisms. When parasitized RBCs are rare (particularly at the onset of symptoms), identification of the parasite may require multiple blood smears over several days. If babesiosis is suspected but the parasite cannot be identified by microscopy, amplification of babesial 18S rRNA by polymerase chain reaction (PCR) is recommended. Serology is useful to confirm the diagnosis. An indirect immunofluorescent antibody test for *B. microti* is available through the Centers for Disease Control and Prevention. IgM titers of ≥1:64 and IgG titers of ≥1:1024 signify active or recent infection. Titers typically decline over 6–12 months. Titers of <1:64 suggest complete clearance. Titers that remain positive (≥1:64) suggest persistent low-level parasitemia. Antibodies to *B. microti* do not react with *B. duncani* or *B. divergens*–like organisms.

Parasitemia levels typically range from 1% to 20% in immunocompetent hosts but can reach 85% in asplenic patients. Low hematocrit, low hemoglobin, low haptoglobin, and elevated lactate dehydrogenase levels are consistent with hemolytic anemia. Reticulocyte counts are elevated; thrombocytopenia is common. WBC counts are normal or slightly depressed. Liver function tests (alkaline phosphatase, aspartate and alanine aminotransferases, bilirubin) yield elevated values. Urinalysis may detect hemoglobinuria, excess urobilinogen, and proteinuria. Elevated concentrations of blood urea nitrogen and serum creatinine indicate renal compromise.

Treatment: Babesiosis

(See **Table 211-1**) Whether *B. microti* infection should be treated depends on the clinical context. Asymptomatic infections need not be treated unless *Babesia* organisms are detected on blood smear or by PCR for >3 months. Symptomatic infections need not be treated if *Babesia* is not detected, despite positive serology. When *Babesia* is detected in symptomatic patients, treatment should be initiated.

Table 211-1 Treatment of Human Babesiosis

Organism	Severity	Adults	Children
B. microti	Mild[a]	Atovaquone (750 mg q12h PO) *plus* Azithromycin (500–1000 mg/d PO on day 1, 250 mg/d PO thereafter)	Atovaquone (20 mg/kg q12h PO; maximum, 750 mg/dose) *plus* Azithromycin [10 mg/kg qd PO on day 1 (maximum, 500 mg/dose), 5 mg/kg qd PO thereafter (maximum, 250 mg/dose)]
	Severe[a]	Clindamycin (300–600 mg q6h IV or 600 mg q8h PO) *plus* Quinine (650 mg q6–8h PO) *plus*	Clindamycin (7–10 mg/kg q6–8h IV or 7–10 mg/kg q6–8h PO; maximum, 600 mg/dose) *plus* Quinine (8 mg/kg q8h PO; maximum, 650 mg/dose)

		Consider RBC exchange transfusion	*plus* Consider RBC exchange transfusion
B. divergens[b]		Immediate complete RBC exchange transfusion *plus* Clindamycin (7–10 mg/kg q6–8h IV; maximum, 600 mg/dose) *plus* Quinine (8 mg/kg q8h PO; maximum, 650 mg/dose)	Immediate complete RBC exchange transfusion *plus* Clindamycin (600 mg q6–8h IV) *plus* Quinine (650 mg q8h PO)

[a]Treatment for 7–10 days. In asplenic individuals and in immunocompromised patients, therapy should last for at least 6 weeks, including 2 weeks after parasites are no longer detected on blood smear.

[b]Treatment for 7–10 days, but duration may vary.

Abbreviation: RBC, red blood cell.

MILD *B. MICROTI* ILLNESS

The recommended regimen for treatment of mild illness due to *B. microti* is oral atovaquone plus azithromycin for 7–10 days. Clindamycin plus quinine is the second choice. The two regimens are equally effective, but atovaquone plus azithromycin is better tolerated. Symptoms should begin to abate within 48 h of the initiation of therapy and should resolve within 3 months. For immunocompromised patients, high-dose azithromycin (600–1000 mg/d) is recommended. Relevant antimicrobial therapy should be initiated when patients are also diagnosed with Lyme disease (Chap. 173) or human granulocytotropic anaplasmosis (Chap. 174)—conditions that should be sought whenever *B. microti* is diagnosed.

SEVERE *B. MICROTI* ILLNESS

The recommended regimen for treatment of severe illness caused by *B. microti* is IV clindamycin plus oral quinine for 7–10 days. Oral quinine may be replaced with IV quinidine. Partial or complete RBC exchange transfusion is advised in cases of high-level parasitemia (>10%); severe anemia (hemoglobin, ≤10 g/dL); or pulmonary, hepatic, or renal compromise. Parasitemia and hematocrit should be monitored every day or every other day until symptoms recede and the parasitemia level is <5%.

A single course of standard antimicrobial therapy is often insufficient to eradicate symptoms and parasitemia in patients immunocompromised by splenectomy, HIV/AIDS, malignancy, or immunosuppressive therapy, including rituximab (anti-CD20) therapy for B cell lymphomas. In these patients, cure is more likely when therapy is administered for at least 6 weeks, including 2 weeks after parasites are no longer observed on blood smear. In immunocompromised patients with relapsing babesiosis, a second course of atovaquone plus azithromycin may fail to result in cure, despite >28 days of uninterrupted therapy. In the setting of severe disease, consideration should always be given to the possibility of intercurrent Lyme disease or human granulocytotropic anaplasmosis and empirical therapy for these infections considered until they are excluded.

When clindamycin plus quinine fails to clear *B. microti* or when quinine must be discontinued, alternative regimens have been successful according to anecdotal reports. These empirical regimens have consisted of two- or three-drug combinations such as azithromycin plus quinine and clindamycin plus azithromycin added to doxycycline. In one immunocompromised patient, B. microti was cleared after administration of a five-drug regimen that consisted of clindamycin, quinine, azithromycin, atovaquone, and the combination of atovaquone-proguanil.

OTHER BABESIAL INFECTIONS

The regimen for *B. duncani* infections typically consists of IV clindamycin (600 mg tid/qid or 1200 mg bid) plus oral quinine (600–650 mg tid) for 7–10 days. One relapse was cured with IV clindamycin (1200 mg tid) for10 days. A pediatric case was successfully treated with IV clindamycin (40 mg/kg per day) plus oral quinine (25 mg/kg per day) for 15 days. The regimen for *B. divergens*–like infections typically consists of IV clindamycin (600 mg tid/qid, 900 mg tid, or 1200 mg bid) plus oral quinine or quinidine (650 mg tid).

Prevention

No vaccine is available for human use. There is no role for antibiotic prophylaxis. Individuals who reside in or travel to endemic areas, especially those at risk for severe babesiosis, should wear clothing that covers the lower part of the body, apply tick repellents (such as DEET) to clothing, and

limit outdoor activities from May through September. A thorough skin examination should be conducted after outdoor activities and ticks removed using tweezers. Individuals with a history of symptomatic babesiosis or with positive antibody titers are indefinitely deferred from donating blood.

Global Considerations

EUROPE

About 30 cases of babesiosis have been attributed to *B. divergens*, a pathogen of cattle. Most cases have occurred in France, Ireland, and Great Britain. Sporadic cases have been reported from Croatia (index case), Spain, Portugal, and Sweden. Three cases—in Italy, Austria, and Germany, respectively—were caused by *Babesia* EU1, a parasite of roe deer. Both parasites are transmitted by the sheep tick *Ixodes ricinus*. The single reported case of *B. microti* infection was probably acquired by transfusion of a contaminated platelet concentrate.

Asplenia is a major risk factor for symptomatic *B. divergens* infection. The incubation period is 1–3 weeks long. The onset of hemoglobinuria and jaundice is sudden. Other symptoms include persistent fever (>41°C), shaking chills, drenching sweats, headache, myalgia, and lumbar and abdominal pain. Mild hepatomegaly may be noted. If the infection is not rapidly treated, the patient's condition deteriorates, with pulmonary edema and renal failure. In the past, most patients with severe cases died. Since exchange transfusion has been combined with chemotherapy, fatal cases have been rare. The overall fatality rate remains ~40%. All patients infected with *Babesia* EU1 had been splenectomized, but their illness ranged from mild to severe; none died.

Because *B. divergens* infections are fulminant, the parasite is readily identified on blood smears. The parasitemia level may reach 80%. Serology is of no use because symptoms appear before antibody levels rise. In *Babesia* EU1 infection, the parasitemia level ranges from 1% to 30%. No serologic test specific for *Babesia* EU1 is available, but sera from patients infected with *Babesia* EU1 react with *B. divergens* antigen.

Babesiosis caused by *B. divergens* is a medical emergency. The recommended treatment is immediate complete blood exchange transfusion and therapy with IV clindamycin plus oral quinine or IV quinidine. In some cases, cure has been obtained with exchange transfusion and clindamycin monotherapy. Although uninfected RBCs are introduced by exchange transfusion, anemia may persist for >1 month. If so, additional transfusion is needed. One mild case was cured with IV pentamidine plus oral trimethoprim-sulfamethoxazole. *Babesia* EU1 infections have been treated with IV or oral clindamycin (600 mg tid) alone or in combination with oral quinine (650 mg tid). One patient who became intolerant to quinine was treated with oral atovaquone (750 mg bid) plus oral azithromycin (500 mg/d).

REST OF THE WORLD

A case of *B. divergens*–like infection has been identified on the Canary Islands, and isolated cases have been reported from South Africa, Mozambique, Egypt, and India. Two cases of *B. microti*–like infection have been reported in Taiwan and another in Japan. A patient in South Korea was infected with a *Babesia* species (KO1) related to species found in sheep. Asymptomatic infections have been identified in Mexico (*B. bigemina*, *B. canis*) and Colombia (*B. bigemina*, *B. bovis*). A case of *B. microti* infection diagnosed in Poland most likely was imported from Brazil.

FURTHER READINGS

Gubernot DM et al: Transfusion-transmitted babesiosis in the United States: Summary of a workshop. Transfusion 49:2759, 2009[PMID: 19821952]

Krause PJ et al: Persistent and relapsing babesiosis in immunocompromised patients. Clin Infect Dis 46:370, 2008[PMID: 18181735]

Tonnetti L et al: Transfusion-transmitted *Babesia microti* identified through hemovigilance. Transfusion 49:2557, 2009[PMID: 19624607]

Vannier E et al: Human babesiosis. Infect Dis Clin North Am 22:469, 2008[PMID: 18755385]

Wormser GP et al: Emergence of resistance to azithromycin–atovaquone in immunocompromised patients with *Babesia microti* infection. Clin Infect Dis 50:381, 2010[PMID: 20047477]

——— et al: The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: Clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 43:1089, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF BLOOD SMEARS OF MALARIA AND BABESIOSIS: INTRODUCTION

Five species of blood protozoan parasites cause human malaria: the potentially lethal and often drug-resistant Plasmodium falciparum; the relapsing parasites P. vivax and P. ovale; P. malariae, which can persist at low densities for years; and *P. knowlesi*, a monkey parasite that causes occasional infections in humans in tropical forests in Southeast Asia. *P. knowlesi* resembles *P. falciparum* and *P. malariae* microscopically but is identified definitively by molecular methods (see Table e27-1, footnote a).

Table e27-1 Morphologic Characteristics of Human Malaria Parasites[a]

	P. falciparum	***P. vivax***	***P. ovale***	***P. malariae***
Asexual parasites	Usually only fine blue ring forms (some resembling stereo headsets) are seen. Parasitemia level may exceed 2%.	Irregular, large, fairly thick rings become highly pleomorphic as the parasite grows. Parasitemia level is low.	Regular, dense ring enlarges to compact, blue, mature tropho-zoite (rectangular or band-form). Parasitemia level is low.	Dense, thick rings mature to dense, round trophozoites. Parasitemia level is low.
Schizonts	Rare in peripheral blood; 8–32 merozoites, dark brown-black pigment	Common; 12–18 merozoites, orange-brown pigment	8–14 merozoites, brown or black pigment	8–10 merozoites, dark brown or black pigment
Gametocytes	Banana-shaped; male: light blue; female: darker blue; a few scattered blue-black pigment granules in cytoplasm	Round or oval; male: round, pale blue; female: oval, dark blue; triangular nucleus, a few orange pigment granules	Large, round, dense, and blue (like P. malariae), but prominent James's dots; brown pigment	Large, oval; male: pale blue; female: dense blue; large black pigment granules
RBC changes	RBCs are normal in size. As the parasite matures, the RBC cytoplasm becomes pale, the cells become crenated, and a few small red dots may appear over the cytoplasm (Maurer's clefts).	RBCs are enlarged. Pale red Schüffner's dots increase in number as the parasite matures.	RBCs become oval with tufted ends. Red James's dots are prominent.	RBCs are normal in size and shape. No red dots are seen.

[a] The early trophozoites of *P. knowlesi* resemble those of P. *falciparum*. The late and mature trophozoites, schizonts, and gametocytes of P. knowlesi appear very similar to those of P. malariae; the differences are that *P. knowlesi* trophozoites may have double chromatin dots and two or three parasites per RBC and that *P. knowlesi* mature schizonts have 16 merozoites rather than the 8–10 found with *P. malariae.*

Abbreviation: RBC, red blood cell.

The malaria parasites are readily seen under the microscope (x1000 magnification) in thick and thin blood smears stained with supravital dyes (e.g., Giemsa's, Field's, Wright's, Leishman's). The morphologic characteristics of the parasites are summarized in Table e27-1. In the thick film, lysis of red blood cells by water leaves the stained white cells and parasites, allowing detection of densities as low as 50 parasites/μL. This degree of sensitivity is up to 100 times greater than that of the thin film, in which the red cells are fixed and the malaria parasites are seen inside the cells. The thin film is better for speciation and provides useful prognostic information in severe falciparum malaria. Several findings are associated with increased mortality risk: high parasite counts, more mature parasites (>20% containing visible malaria pigment), and phagocytosed malaria pigment in >5% of neutrophils.

Babesia microti appears as a small ring form resembling P. falciparum. Unlike Plasmodium, Babesia does not cause the production of pigment in parasites, nor are schizonts or gametocytes formed.

Figure e27-1

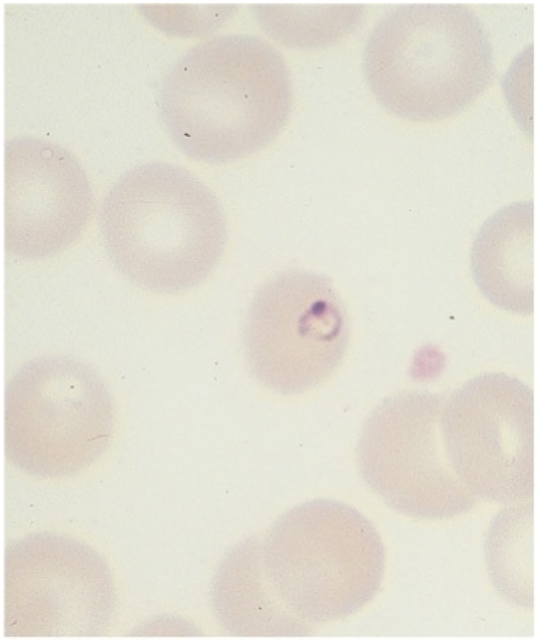

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

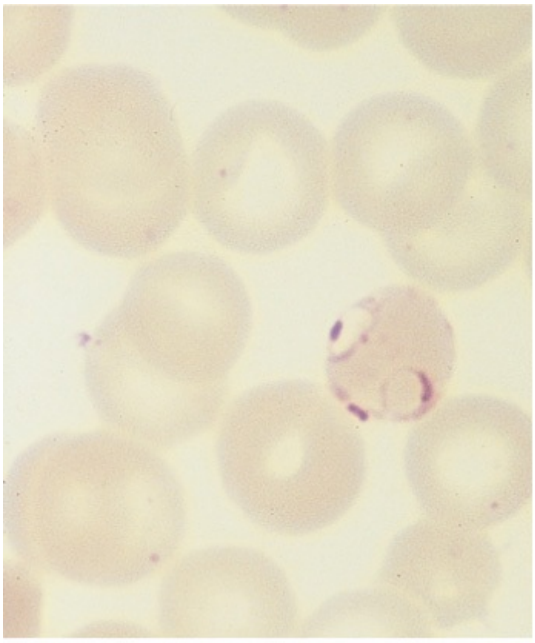

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

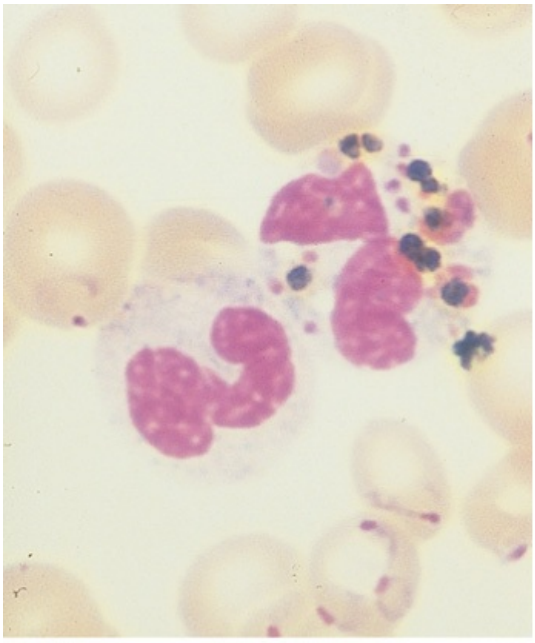

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

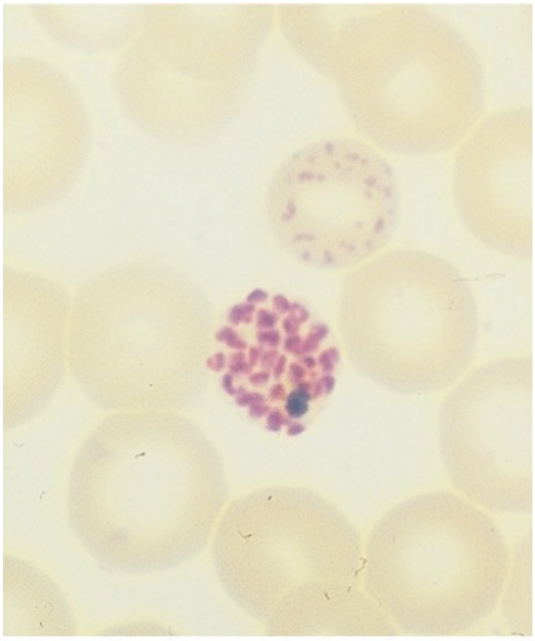

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

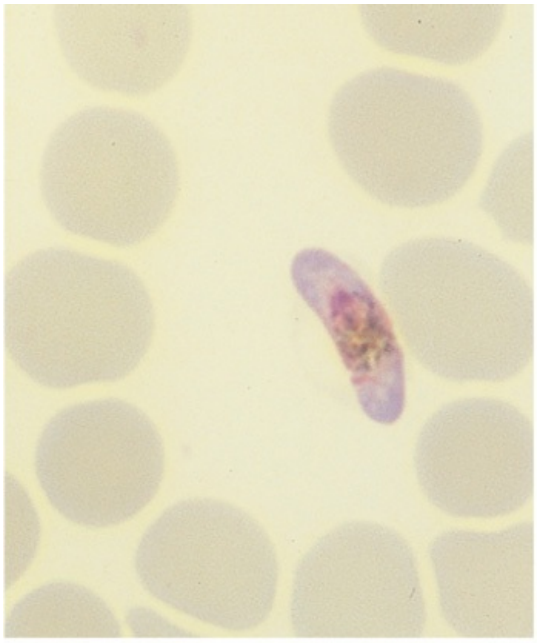

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

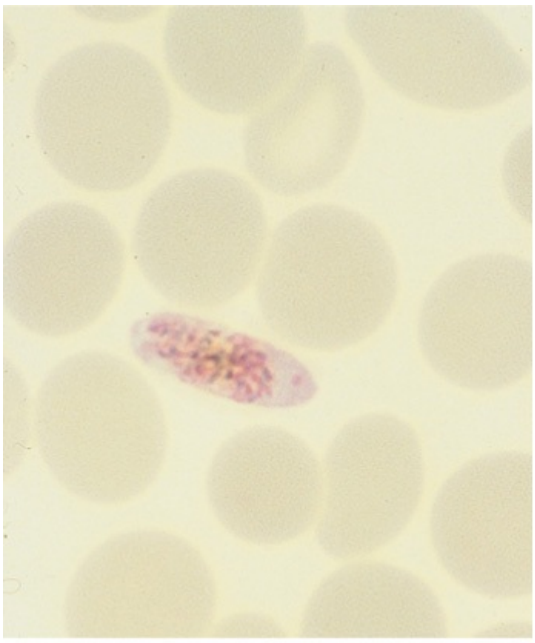

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thin blood films of *Plasmodium falciparum*. *A.* Young trophozoites. ***B.*** Old trophozoites. ***C.*** Pigment in polymorphonuclear cells and trophozoites. ***D.*** Mature schizonts. ***E.*** Female gametocytes. **F.** Male gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure e27-2

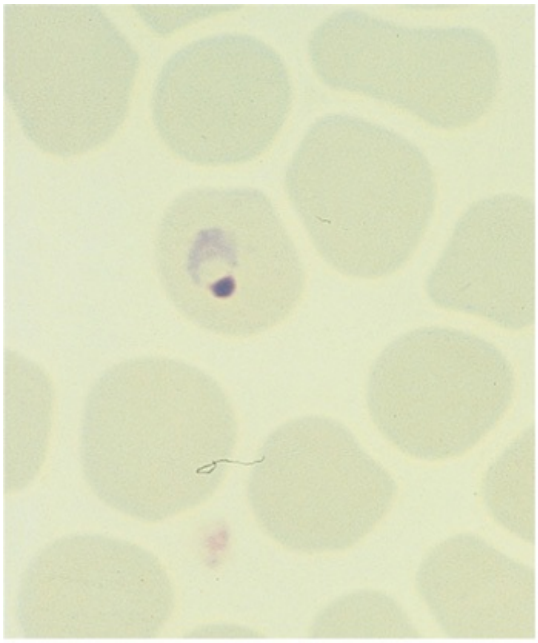

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

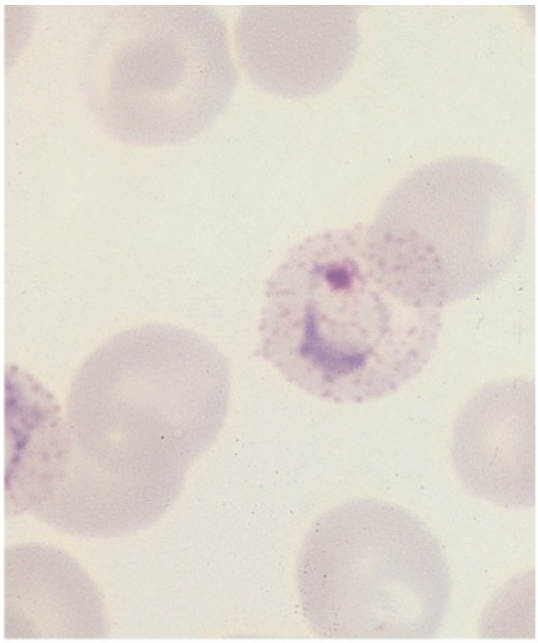

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

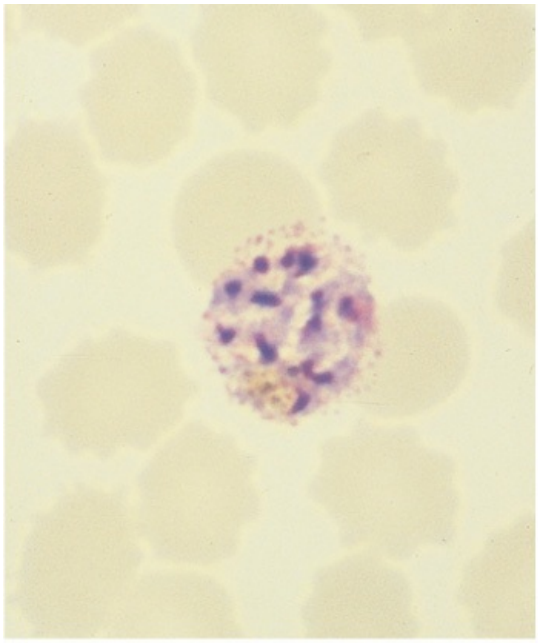

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

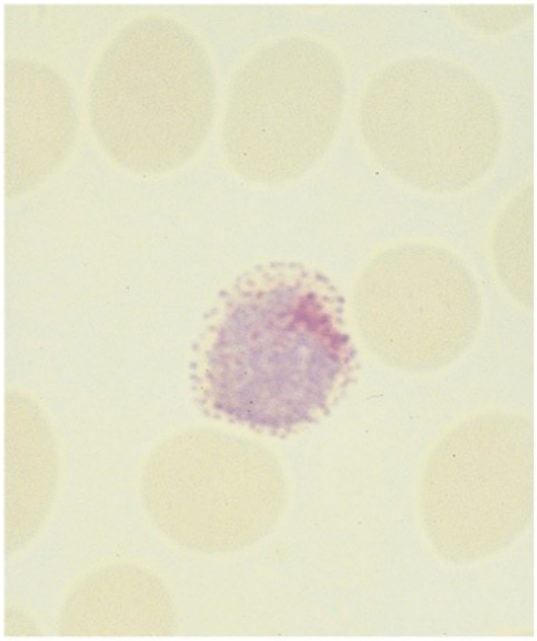

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

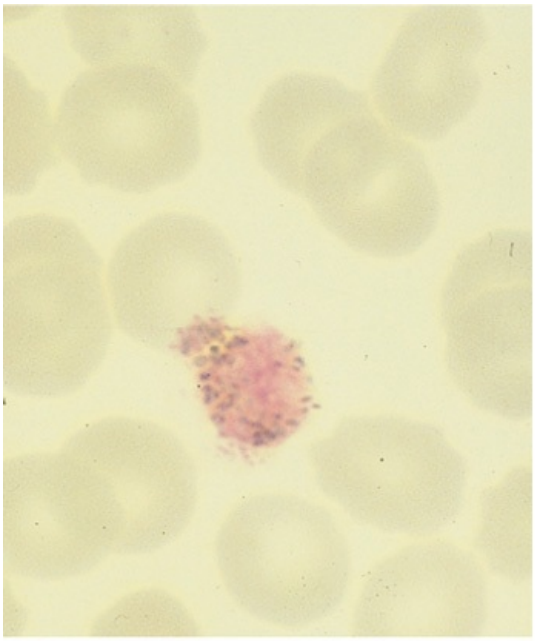

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thin blood films of *Plasmodium vivax*. *A.* Young trophozoites. ***B.*** Old trophozoites. ***C.*** Mature schizonts. ***D.*** Female gametocytes. ***E.*** Male gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure e27-3

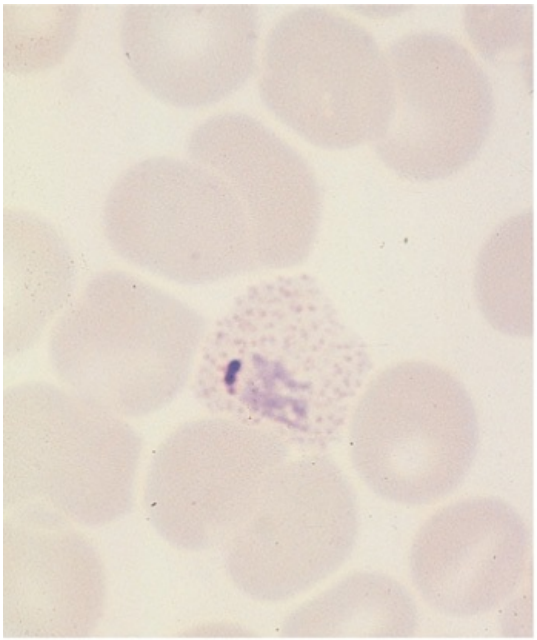

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

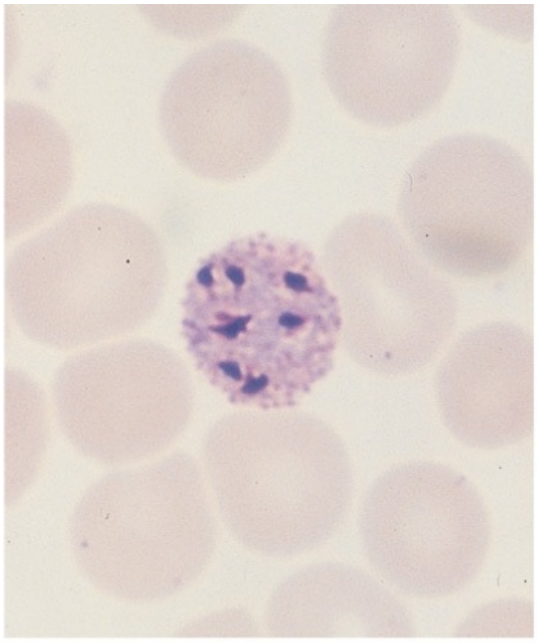

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

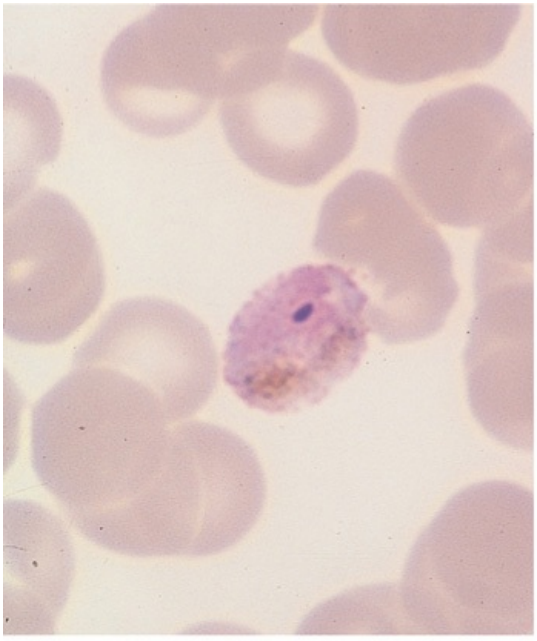

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

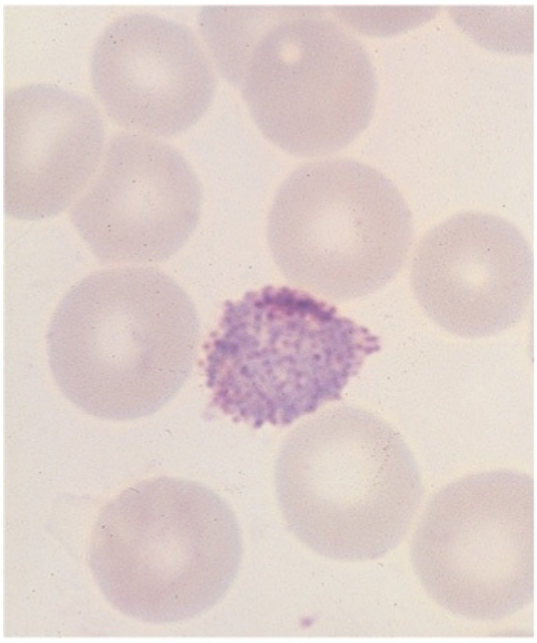

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thin blood films of *Plasmodium ovale*. *A.* Old trophozoites. ***B.*** Mature schizonts. ***C.*** Male gametocytes. ***D.*** Female gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure e27-4

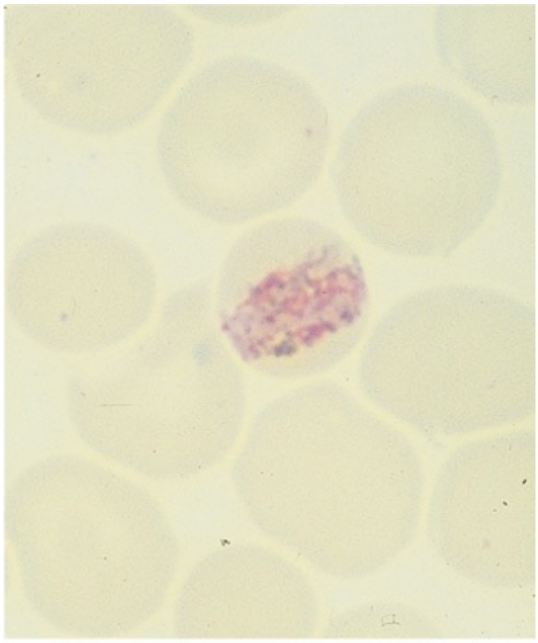

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

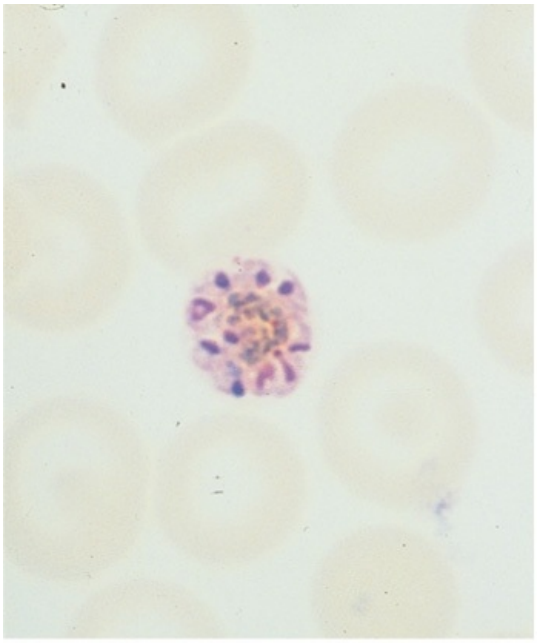

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

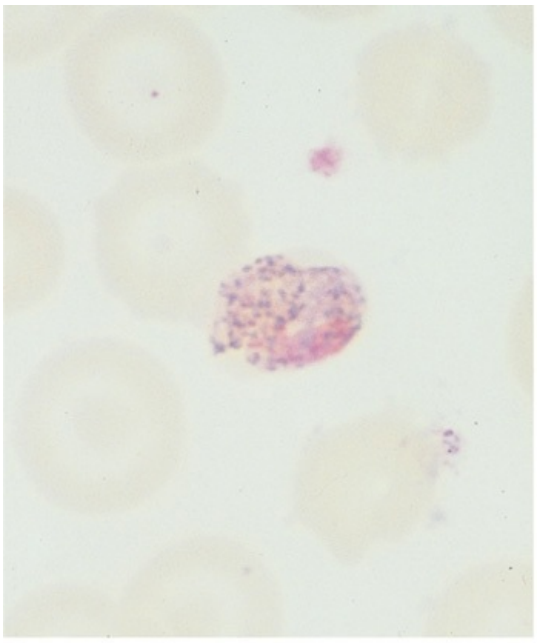

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

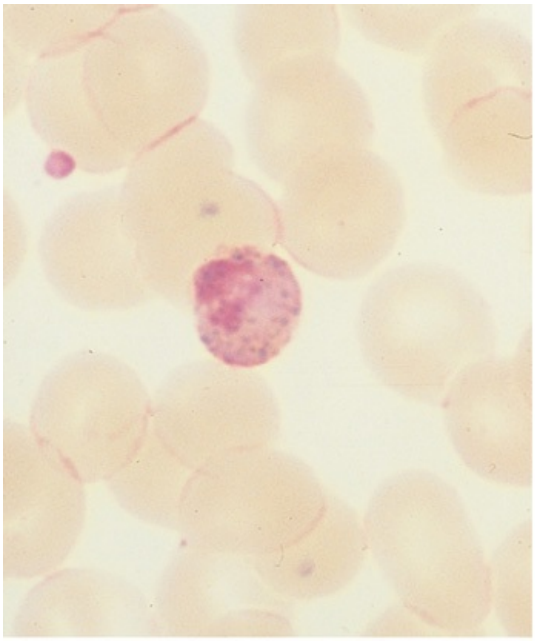

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thin blood films of *Plasmodium malariae*. *A.* Old trophozoites. ***B.*** Mature schizonts. ***C.*** Male gametocytes. ***D.*** Female gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure e27-5

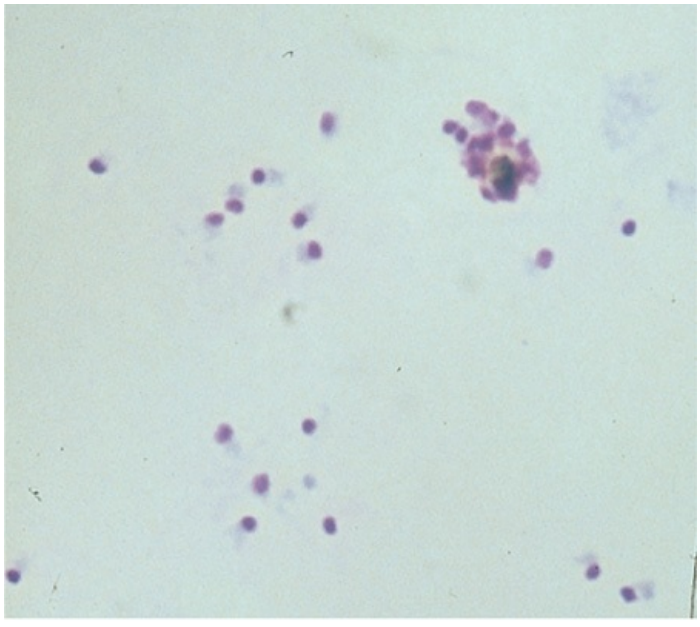

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

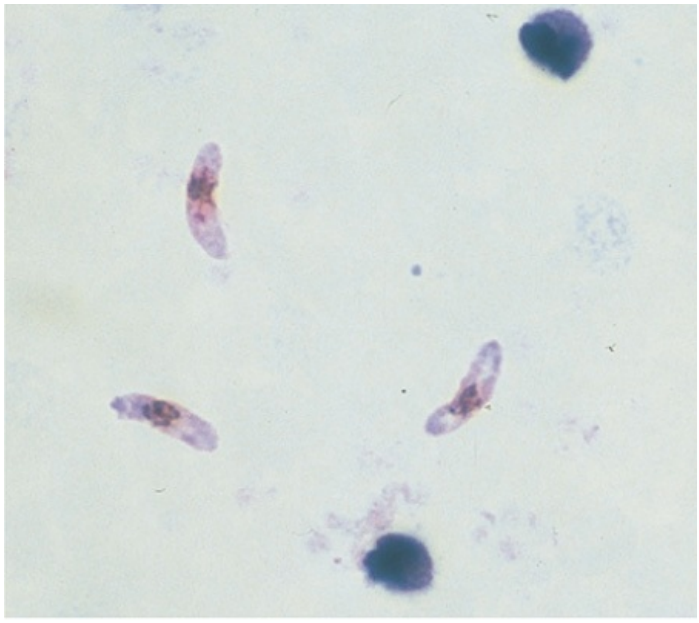

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium falciparum*. *A.* Trophozoites. ***B.*** Gametocytes. (Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)

Figure e27-6

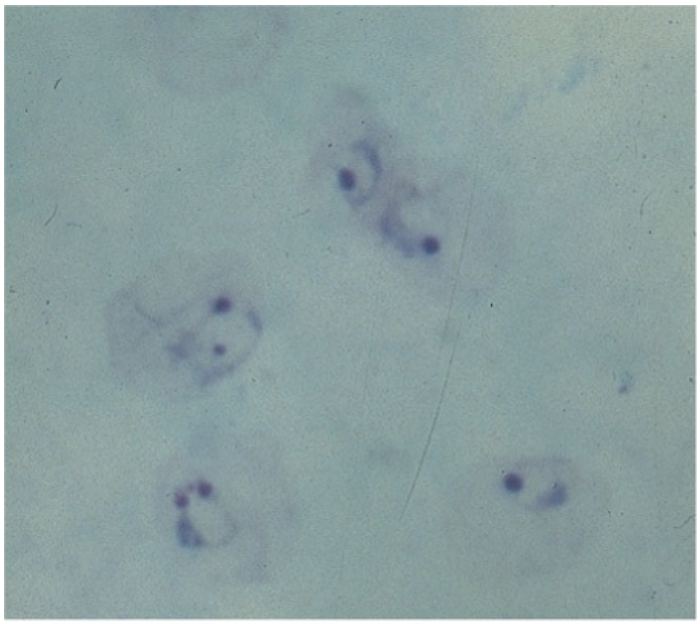

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

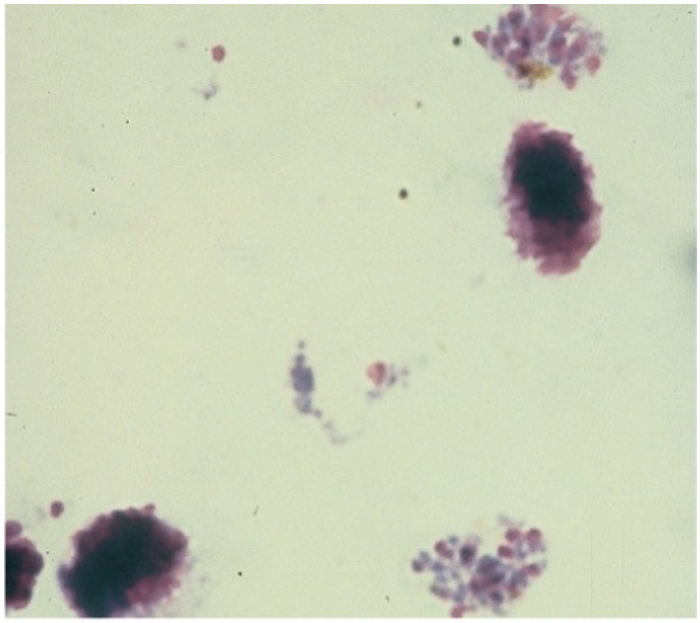

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

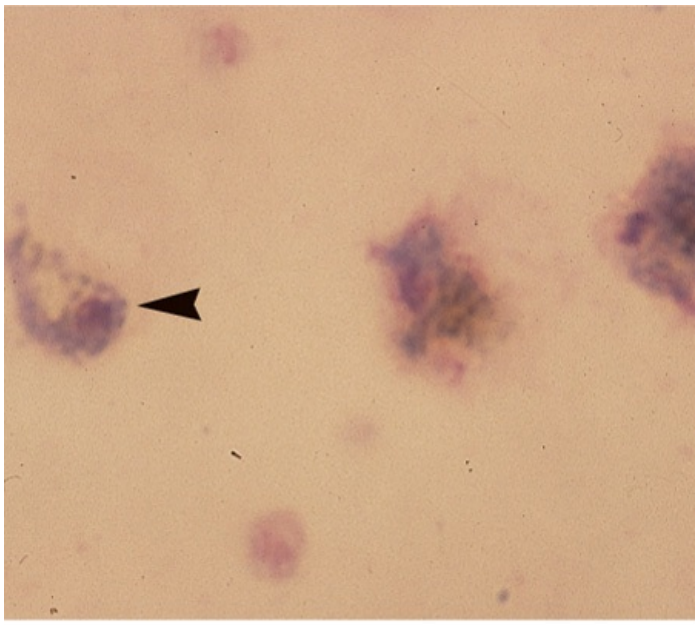

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium vivax*. *A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure e27-7

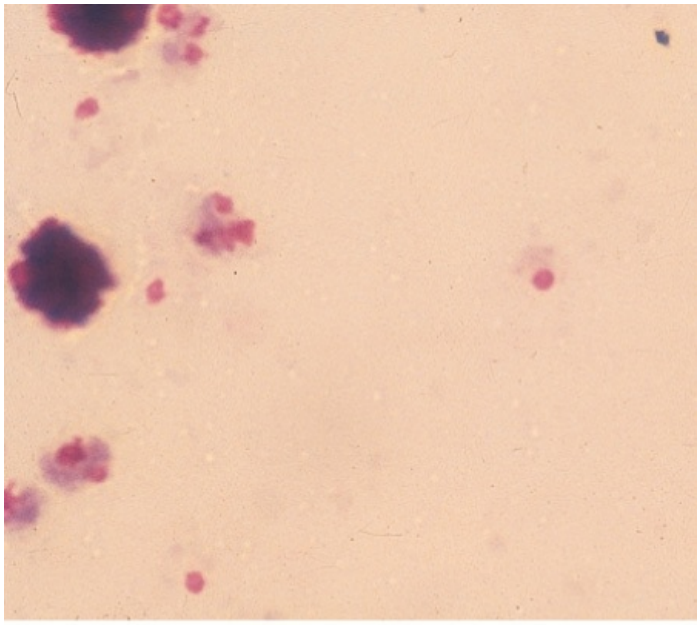

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

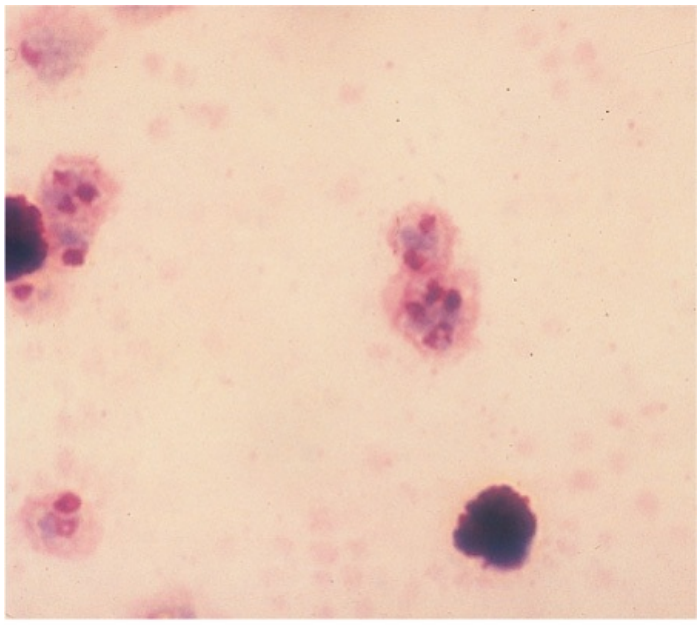

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

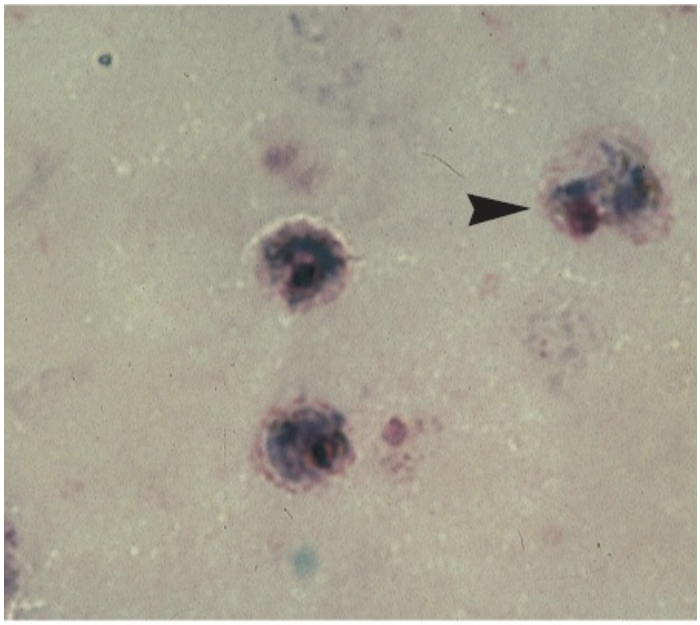

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium ovale*. *A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure e27-8

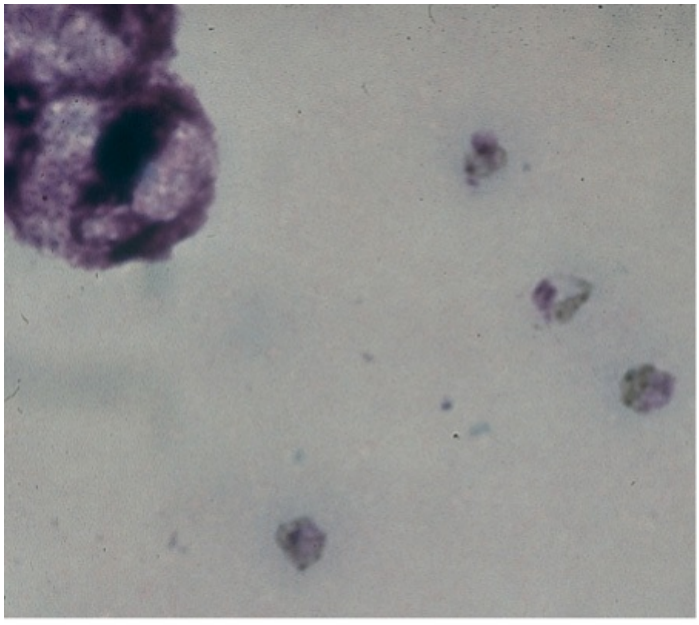

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

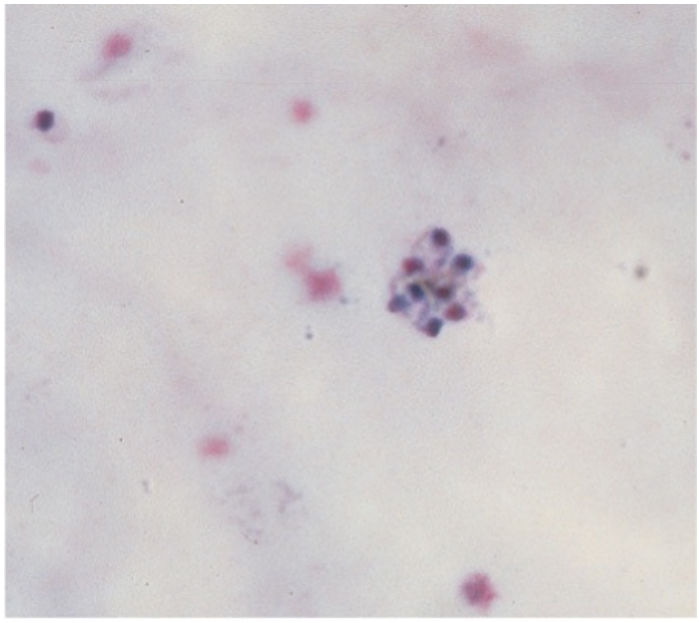

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

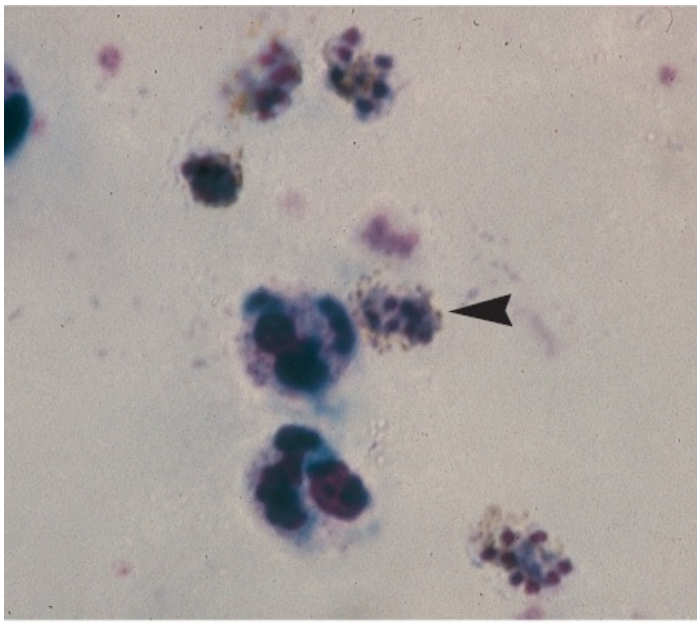

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thick blood films of *Plasmodium malariae*.*A.* Trophozoites. ***B.*** Schizonts. ***C.*** Gametocytes. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

Figure e27-9

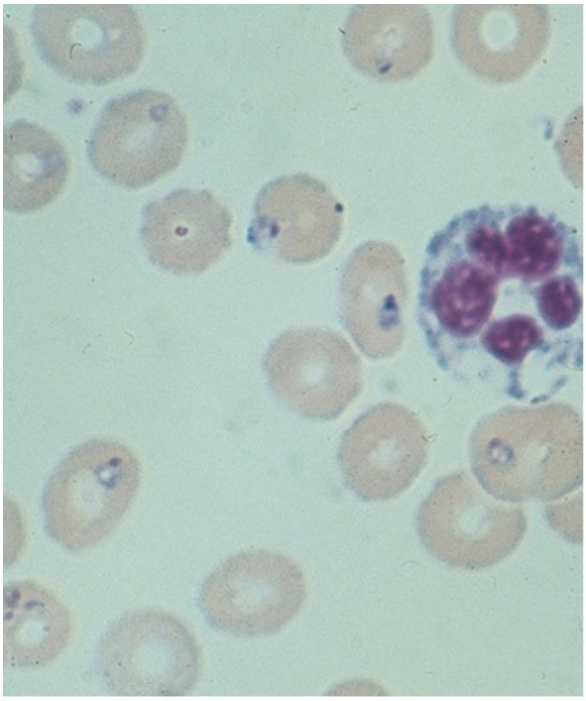

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thin blood film showing trophozoites of *Babesia*. *(Reproduced from Bench Aids for the Diagnosis of Malaria Infections, 2nd ed, with the permission of the World Health Organization.)*

FURTHER READINGS

Warhurst C, Williams JE: Laboratory procedures for diagnosis of malaria, in Abdalla SH, Pasvol G (series eds): *Malaria*: A *Hematological Perspective*. G Pasvol, SL Hoffman (eds): *Tropical Medicine: Science and Practice,* vol 4. London, Imperial College Press, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LEISHMANIASIS: INTRODUCTION

Definition

Encompassing a complex group of disorders, leishmaniasis is caused by unicellular eukaryotic obligatory intracellular protozoa of the genus *Leishmania* and primarily affects the host's reticuloendothelial system. *Leishmania* species produce widely varying clinical syndromes ranging from self-healing cutaneous ulcers to fatal visceral disease. These syndromes fall into three broad categories: visceral leishmaniasis (VL), cutaneous leishmaniasis (CL), and mucosal leishmaniasis (ML).

Etiology and Life Cycle

Leishmaniasis is caused by ~20 species of the genus *Leishmania* in the order Kinetoplastida and the family Trypanosomatidae (**Table 212-1**). Several clinically important species are of the subspecies *Viannia*. The organisms are transmitted by phlebotomine sandflies of the genus *Phlebotomus* in the "Old World" (Asia, Africa, and Europe) and the genus Lutzomyia in the "New World" (the Americas). Transmission may be anthroponotic (i.e., the vector transmits the infection from infected humans to healthy humans) or zoonotic (i.e., the vector transmits the infection from an animal reservoir to humans). Human-to-human transmission via shared infected needles has been documented in IV drug users in the Mediterranean region. In utero transmission to the fetus occurs rarely.

Table 212-1 Geographic Distribution and Characteristic Epidemiology of Leishmaniases

Organism, Endemic Region	Clinical Syndrome	Species	Vector	Reservoir	Transmission	Setting
***L. donovani* Complex**						
South Asia	VL, PKDL	*L. donovani*	Phlebotomus argentipes	Humans	Anthroponotic	Rural, domestic
Sudan, Somalia, Ethiopia, Kenya, Uganda	VL, PKDL	*L. donovani*	P. orientalis, P. martini	Humans, rodents in Sudan, canines	Anthroponotic, occasionally zoonotic	Majority peridomestic, occasionally sylvatic
Mediterranean basin, Middle East, Central Asia, China	VL, CL	*L. infantum*	P. perniciosus, P. ariasi	Dogs, foxes, jackals	Zoonotic	Domestic, peridomestic
Middle East, Saudi Arabia, Yemen	VL	*L. donovani*	P. perniciosus, P. ariasi	Dogs, foxes, jackals	Zoonotic	Domestic, peridomestic
Central and South America	VL, CL	*L. infantum*	Lutzomyia longipalpis	Foxes, dogs, opossums	Zoonotic	Domestic, peridomestic, periurban
Azerbaijan, Armenia, Georgia, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan	VL	*L. infantum*	P. turanicus	Humans, dogs, foxes	Anthroponotic, zoonotic	Domestic
L. tropica						
Western India to Turkey, parts of North and East Africa	CL, leishmaniasis recidivans	*L. tropica*	*P. sergenti*	Humans	Anthroponotic	Urban domestic, peridomestic
L. major						
Western and Central Asia, North and sub-Saharan Africa	CL	*L. major*	*P. papatasi, P. duboscqi*	Nile rats, rodents	Zoonotic	Sylvatic, peridomestic
Kazakhstan, Turkmenistan, and Uzbekistan	CL	*L. major*	*P. papatasi, P. duboscqi*	Gerbils	Zoonotic	Rural
L. aethiopica						
Ethiopia, Uganda, Kenya	CL, DCL	*L. aethiopica*	*P. longipes, P. pedifer*	Hyraxes	Zoonotic	Sylvatic, peridomestic
Subspecies *Viannia*						
Peru, Ecuador	CL, ML	*L. (V.) peruviana*	Lutzomyia verrucarum, *L. peruensis*	Wild rodents	Zoonotic	Andean Valleys
Guyana, Surinam, French Guyana, Ecuador. Brazil.	CL, ML	*L. (V.) guyanensis*	*L. umbratilis*	Sloths, arboreal anteaters.	Zoonotic	Tropical forests

Colombia, Bolivia				opossums		
Central America, Ecuador, Colombia	CL, ML	*L. (V.) panamensis*	*L. trapidoi*	Sloths	Zoonotic	Tropical forest and deforested areas
South and Central America	CL, ML	*L. (V.) braziliensis*	*Lutzomyia* spp., *L. umbratilis, Psychodopygus wellcomei*	Forest rodents, peridomestic animals	Zoonotic	Tropical forest and deforested areas
***L. mexicana* Complex**						
Central America and northern parts of South America	CL, ML, DCL	*L. amazonensis*	*L. flaviscutellata*	Forest rodents	Zoonotic	Tropical forest and deforested areas
	CL, ML, DCL	*L. mexicana*	*L. olmeca*	Variety of forest rodents and marsupials	Zoonotic	
	CL, DCL	*L. pifanoi*	*L. olmeca*	Variety of forest rodents and marsupials	Zoonotic	Tropical forest and deforested areas

Abbreviations: CL, cutaneous leishmaniasis; DCL, diffuse cutaneous leishmaniasis; ML, mucosal leishmaniasis; PKDL, post–kala-azar dermal leishmaniasis; VL, visceral leishmaniasis.

Leishmania organisms occur in two forms: extracellular, flagellate promastigotes (length, 10–20 μm) in the sandfly vector and intracellular, nonflagellate amastigotes (length, 2–4 μm; **Fig. 212-1**) in vertebrate hosts, including humans. Promastigotes are introduced through the proboscis of the female sandfly into the skin of the vertebrate host. Neutrophils predominate among the host cells that first encounter and take up promastigotes at the site of parasite delivery. The infected neutrophils may undergo apoptosis and release viable parasites that are taken up by macrophages, or the apoptotic cells may themselves be taken up by macrophages and dendritic cells. The parasites multiply as amastigotes inside macrophages, causing cell rupture with subsequent invasion of other macrophages. While feeding on infected hosts, sandflies pick up amastigotes, which transform into the flagellate form in the flies' posterior midgut and multiply by binary fission; the promastigotes then migrate to the anterior midgut and can infect a new host when flies take another blood meal.

Figure 212-1

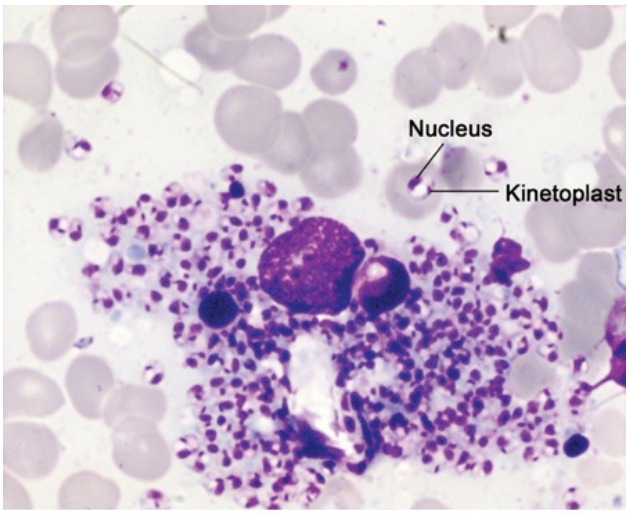

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A macrophage with numerous intracellular amastigotes (2–4 μm) in a Giemsa-stained splenic smear from a patient with visceral leishmaniasis. Each amastigote contains a nucleus and a characteristic kinetoplast consisting of multiple copies of mitochondrial DNA. A few extracellular parasites are also visible.

Epidemiology

Leishmaniasis occurs in 98 countries—most of them developing—in tropical and temperate regions (**Fig. 212-2**). Two million cases occur annually, of which 1–1.5 million are CL (and its variations) and 500,000 are VL. More than 350 million people are at risk, with an overall prevalence of 12 million. Although the distribution of *Leishmania* is limited by the distribution of sandfly vectors, human leishmaniasis is on the increase worldwide.

Figure 212-2

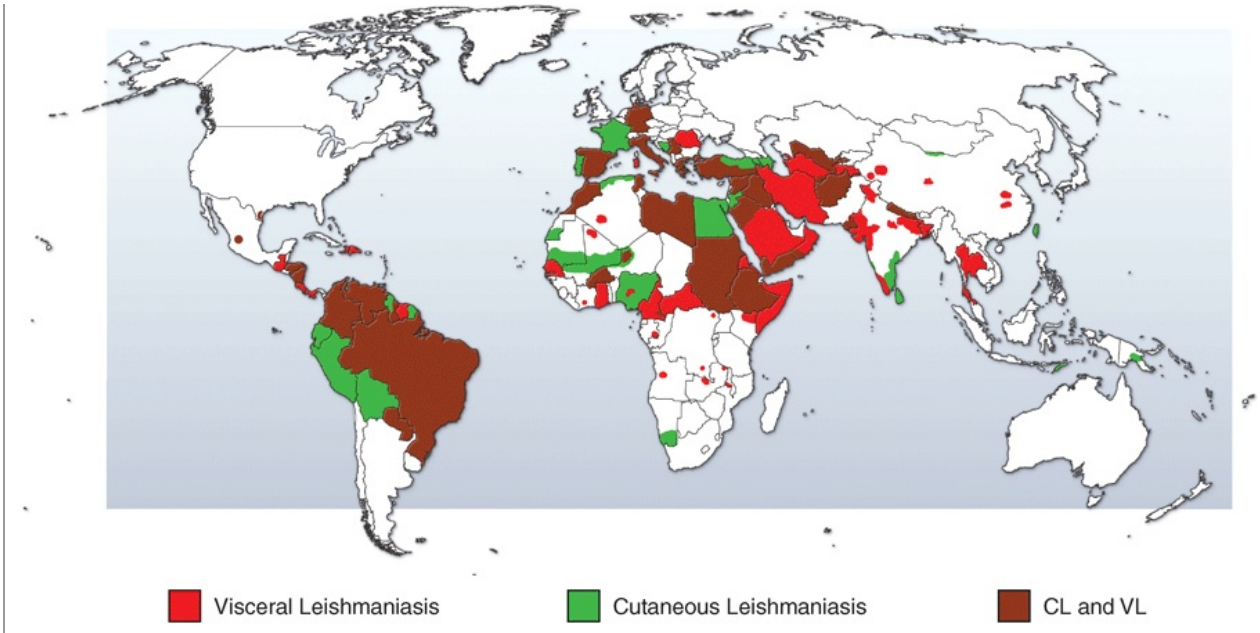

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Worldwide distribution of human leishmaniasis.

Visceral Leishmaniasis

VL (also known as *kala-azar*, a Hindi term meaning "black fever") is caused by the *L. donovani* complex, which includes *L. donovani* and *L. infantum* (the latter designated *L. chagasi* in the New World); these species are responsible for anthroponotic and zoonotic transmission, respectively. India and neighboring Nepal, Bangladesh, Sudan, and Brazil are the four largest foci of VL and account for 90% of the world's VL burden, with India the worst affected. Zoonotic VL is reported from all countries in the Middle East, Pakistan, and other countries from western Asia to China. Endemic foci also exist in the independent states of the former Soviet Union, mainly Georgia and Azerbaijan. In the Horn of Africa, Sudan, Ethiopia, Kenya, Uganda, and Somalia report VL. In Sudan, large outbreaks are thought to be anthroponotic, although zoonotic transmission also occurs. VL is rare in West and sub-Saharan Africa.

Mediterranean VL, long an established endemic disease due to *L. infantum*, has a large canine reservoir and was seen primarily in infants before the advent of HIV. In Mediterranean Europe, 70% of adult VL cases are associated with HIV co-infection. The combination is deadly because of the impact of the two infections together on the immune system. IV drug users are at particular risk. Other forms of immunosuppression (e.g., that associated with organ transplantation) also predispose to VL. In the Americas, disease caused by *L. infantum* is endemic from Mexico to Argentina, but 90% of cases in the New World are reported from northeastern Brazil.

IMMUNOPATHOGENESIS

The majority of individuals infected by *L. donovani* or *L. infantum* mount a successful immune response and control the infection, never developing symptomatic disease. Forty-eight hours after intradermal injection of killed promastigotes, these individuals exhibit delayed-type hypersensitivity (DTH) to leishmanial antigens in the leishmanin skin test (also called the Montenegro skin test). Results in mouse models indicate that the development of acquired resistance to leishmanial infection is controlled by the production of interleukin (IL) 12 by antigen-presenting cells and the subsequent secretion of interferon (IFN) γ, tumor necrosis factor (TNF) α, and other proinflammatory cytokines by the T helper 1 (T_H1) subset of T lymphocytes. The immune response in patients developing active VL is complex; in addition to increased production of multiple proinflammatory cytokines and chemokines, patients with active disease have markedly elevated levels of IL-10 in serum as well as enhanced IL-10 mRNA expression in lesional tissues. The main disease-promoting activity of IL-10 in VL may be to condition host macrophages for enhanced survival and growth of the parasite. IL-10 can render macrophages unresponsive to activation signals and inhibit killing of amastigotes by downregulating the production of TNF-α and nitric oxide. Multiple antigen-presentation functions of dendritic cells and macrophages are also suppressed by IL-10. Patients with such suppression do not have positive leishmanin skin tests, nor do their peripheral-blood mononuclear cells respond to leishmanial antigens in vitro. Organs of the reticuloendothelial system are predominantly affected, with remarkable enlargement of the spleen, the liver, and lymph nodes in some regions. The tonsils and intestinal submucosa are also heavily infiltrated with parasites. Bone marrow dysfunction results in pancytopenia.

CLINICAL FEATURES

On the Indian subcontinent and in the Horn of Africa, persons of all ages are affected by VL. In endemic areas of the Americas and the Mediterranean basin, immunocompetent infants and small children and immunodeficient adults are affected especially often. The most common presentation of VL is an abrupt onset of moderate- to high-grade fever associated with rigor and chills. Fever may continue for several weeks with decreasing intensity, and the patient may become afebrile for a short period before experiencing another bout of fever. The spleen may be palpable by the second week of illness and, depending on the duration of illness, may become hugely enlarged (**Fig. 212-3**). Hepatomegaly (usually moderate in degree) soon follows. Lymphadenopathy is common in most endemic regions of the world except the Indian subcontinent, where it is rare. Patients lose weight and feel weak,

and the skin gradually develops dark discoloration due to hyperpigmentation that is most easily seen in brown-skinned individuals. In advanced illness, hypoalbuminemia may manifest as pedal edema and ascites. Anemia appears early and may become severe enough to cause congestive heart failure. Epistaxis, retinal hemorrhages, and gastrointestinal bleeding are associated with thrombocytopenia. Secondary infections such as measles, pneumonia, tuberculosis, bacillary or amebic dysentery, and gastroenteritis are common. Herpes zoster, chickenpox, boils in the skin, and scabies may also occur. Untreated, the disease is fatal in most patients, including 100% of those with HIV co-infection.

Figure 212-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A patient with visceral leishmaniasis has a hugely enlarged spleen visible through the surface of the abdomen. Splenomegaly is the most important feature of visceral leishmaniasis.

Leukopenia and anemia occur early and are followed by thrombocytopenia. There is a marked polyclonal increase in serum immunoglobulins. Serum levels of hepatic aminotransferases are raised in a significant proportion of patients, and serum bilirubin levels are elevated occasionally. Renal dysfunction is uncommon.

LABORATORY DIAGNOSIS

Demonstration of amastigotes in smears of tissue aspirates is the gold standard for the diagnosis of VL (Fig. 212-1). The sensitivity of splenic smears is >95%, whereas smears of bone marrow (60–85%) and lymph node aspirates (50%) are less sensitive. Culture of tissue aspirates increases sensitivity. Splenic aspiration is invasive and may be dangerous in untrained hands. Several serologic techniques are currently used to detect antibodies to *Leishmania*. An enzyme-linked immunosorbent assay (ELISA) and the indirect immunofluorescent antibody test (IFAT) are used in sophisticated laboratories. In the field, however, a rapid immunochromatographic test based on the detection of antibodies to a recombinant antigen (rK39) consisting of 39 amino acids conserved in the kinesin region of *L. infantum* is used worldwide. The test requires only a drop of fingerprick blood or serum, and the result can be read within 15 minutes. Except in East Africa (where both its sensitivity and its specificity are lower), the sensitivity of the rK39 rapid diagnostic test in immunocompetent individuals is ~98% and its specificity is 90%. Qualitative detection of leishmanial nucleic acid by polymerase chain reaction (PCR) and quantitative detection by real-time PCR are confined to specialized laboratories and have yet to be used for routine diagnosis of VL in endemic areas. PCR can distinguish among the major species of *Leishmania* infecting humans.

DIFFERENTIAL DIAGNOSIS

VL is easily mistaken for malaria. Other febrile illnesses that may mimic VL include typhoid fever, tuberculosis, brucellosis, schistosomiasis, and histoplasmosis. Splenomegaly due to portal hypertension, chronic myeloid leukemia, tropical splenomegaly syndrome, and (in Africa) schistosomiasis may also be confused with VL. Fever with neutropenia or pancytopenia in patients from an endemic region strongly suggests a diagnosis of VL; hypergammaglobulinemia in patients with long-standing illness strengthens the diagnosis. In nonendemic countries, a careful travel history is essential when any patient presents with fever.

TREATMENT: VISCERAL LEISHMANIASIS

General Considerations

Severe anemia should be corrected by blood transfusion, and other comorbid conditions should be managed promptly. Treatment of VL is complex, as the optimal drug, dosage, and duration vary with the endemic region. In spite of completing recommended treatment, some patients experience relapse (most often within 6 months), and prolonged follow-up is recommended. A pentavalent antimonial is the drug of choice in most endemic regions of the world, but there is widespread resistance to antimony in the Indian state of Bihar, where either amphotericin B (AmB) deoxycholate or miltefosine is preferred. Dose requirements for AmB are lower in India than in the Americas,

Africa, or the Mediterranean region. In Mediterranean countries, where cost is seldom an issue, liposomal AmB is the drug of choice. In immunocompetent patients, relapses are uncommon with AmB in its deoxycholate and lipid formulations. Antileishmanial therapy has recently evolved as new drugs and delivery systems have become available and resistance to antimonial compounds has emerged.

Except for AmB (deoxycholate and lipid formulations), antileishmanial drugs are available in the United States only from the Centers for Disease Control and Prevention.

Pentavalent Antimonial Compounds

Two pentavalent antimonial (Sb^V) preparations are available: sodium stibogluconate (100 mg of Sb^V/mL) and meglumine antimonate (85 mg of Sb^V/mL). The daily dose is 20 mg/kg by rapid IV infusion or IM injection, and therapy continues for 28–30 days. Cure rates exceed 90% in Africa, the Americas, and most of the Old World but are <50% in Bihar, India, as a result of resistance. Adverse reactions to Sb^V treatment are common and include arthralgia, myalgia, and elevated serum levels of aminotransferases. Electrocardiographic changes are common. Concave ST segment elevation is not significant, but prolongation of QT_C to >0.5 s may herald ventricular arrhythmia and sudden death. Chemical pancreatitis is common but usually does not require discontinuation of treatment; severe clinical pancreatitis occurs in immunosuppressed patients.

Amphotericin B

AmB is currently used as a first-line drug in Bihar. In others parts of the world, it is used when initial antimonial treatment fails. Conventional AmB deoxycholate is administered in doses of 0.75–1.0 mg/kg on alternate days for a total of 15 infusions. Fever with chills is an almost universal adverse reaction to AmB infusions. Nausea and vomiting are also common, as is thrombophlebitis in the infused veins. Acute toxicities can be minimized by administration of antihistamines like chlorpheniramine and antipyretic agents like acetaminophen before each infusion. AmB can cause renal dysfunction and hypokalemia and in rare instances elicits hypersensitivity reactions, bone marrow suppression, and myocarditis, all of which can be fatal.

The several lipid formulations of AmB developed to replace the deoxycholate formulation are preferentially taken up by reticuloendothelial tissues. Because very little free drug is available to cause toxicity, a large amount of drug can be delivered over a short period. Liposomal AmB has been used extensively to treat VL in all parts of the world. With a terminal half-life of ~150 h, liposomal AmB can be detected in the liver and spleen of animals for several weeks after a single dose. This is the only drug approved by the U.S. Food and Drug Administration (FDA) for the treatment of VL; the regimen is 3 mg/kg daily on days 1–5, 14, and 21 (total dose, 21 mg/kg). However, the total dose requirement for different regions of the world varies widely. In Asia, it is 10–15 mg/kg; in Africa, ~18 mg/kg; and in Mediterranean/American regions, not less than 20 mg/kg. The daily dose is flexible (1–10 mg/kg). In a study in India, a single dose of 10 mg/kg cured infection in 96% of patients. Adverse effects of liposomal AmB are usually mild and include infusion reactions, backache, and occasional reversible nephrotoxicity.

Paromomycin

Paromomycin (aminosidine) is an aminocyclitol-aminoglycoside antibiotic with antileishmanial activity. Its mechanism of action against *Leishmania* has yet to be established. Paromomycin is approved in India for the treatment of VL at an IM dose of 11 mg of base/kg daily for 21 days; this regimen produces a cure rate of 95%. However, the optimal dose has not been established in other endemic regions. Paromomycin is a relatively safe drug, but some patients develop hepatotoxicity, reversible ototoxicity, and (in rare instances) nephrotoxicity and tetany.

Miltefosine

Miltefosine, an alkylphosphocholine, is the first oral compound approved for the treatment of leishmaniasis. This drug has a long half-life (150–200 h); its mechanism of action is not clearly understood. The recommended therapeutic regimens for patients on the Indian subcontinent are a daily dose of 50 mg for 28 days for patients weighing <25 kg, a twice-daily dose of 50 mg for 28 days for patients weighing ≥25 kg, and 2.5 mg/kg for 28 days for children 2–11 years of age. These regimens result in a cure rate of 94% in India. Doses in other regions remain to be established. Because of its long half-life, miltefosine is prone to induce resistance in *Leishmania*. Its adverse effects include mild to moderate vomiting and diarrhea in 40% and 20% of patients, respectively; these reactions usually clear spontaneously after a few days. Rare cases of severe allergic dermatitis, hepatotoxicity, and nephrotoxicity have been reported. Because miltefosine is expensive and is associated with significant adverse events, it is best administered as directly observed therapy to ensure completion of treatment and to minimize the risk of resistance induction. Because miltefosine is teratogenic in rats, its use is contraindicated during pregnancy and (unless contraceptive measures are strictly adhered to for at least 3 months after treatment) in women of childbearing age.

Multidrug Therapy

Multidrug therapy for leishmaniasis is likely to be preferred in the future. Its potential advantages in VL include (1) better compliance and lower costs associated with shorter treatment courses and decreased hospitalization, (2) less toxicity due to lower drug doses and/or shorter duration of treatment, and (3) a reduced likelihood that resistance to either agent will develop. Trials of multidrug therapy are under way in Asia and Africa.

Prognosis of Treated VL Patients

Recovery from VL is quick. Within a week of the start of treatment, defervescence, regression of splenomegaly, weight gain, and recovery of hematologicparameters are evident. With effective treatment, no parasites are recovered from tissue aspirates at the posttreatment evaluation. Continued clinical improvement over 6–12 months is suggestive of cure. A small percentage of patients (with the exact figure depending on the regimen used) relapse but respond well to treatment with AmB deoxycholate or lipid formulations.

VL in the Immunocompromised Host

HIV/VL co-infection has been reported from 35 countries. VL behaves as an opportunistic infection in HIV-1-

infected patients where both infections are endemic. HIV infection can increase the risk of developing VL severalfold in endemic areas. Co-infected patients usually show the classic signs of VL, but they may present with atypical features due to loss of immunity and involvement of unusual anatomic locations, with, for example, infiltration of the skin, oral mucosa, gastrointestinal tract, lungs, and other organs. Serodiagnostic tests are commonly negative. Parasites can be recovered from unusual sites such as bronchoalveolar lavage fluid and buffy coat. Liposomal AmB is the drug of choice for HIV/VL co-infection—both for primary treatment and for treatment of relapses. A total dose of 40 mg/kg, administered as 4 mg/kg on days 1–5, 10, 17, 24, 31, and 38, is considered optimal and is approved by the FDA, but most patients relapse within 1 year. Pentavalent antimonials and AmB deoxycholate can also be used where liposomal AmB is not accessible. Reconstitution of patients' immunity by antiretroviral therapy has led to a dramatic decline in the incidence of co-infection in the Mediterranean basin. In contrast, HIV/VL co-infection is on the rise in African and Asian countries. Ethiopia is worst affected: up to 30% of VL patients are also infected with HIV. Since restoration of the CD4+ T cell count to >200/μL does decrease the frequency of relapse, antiretroviral therapy (in addition to antileishmanial therapy) is a cornerstone for the management of HIV/VL co-infection. Secondary prophylaxis with liposomal AmB has been shown to delay relapses, but no regimen has been established as optimal.

Post–Kala-Azar Dermal Leishmaniasis

On the Indian subcontinent and in Sudan and other East African countries, 2–50% of patients develop skin lesions concurrent with or after the cure of VL. Most common are hypopigmented macules, papules, and/or nodules or diffuse infiltration of the skin and sometimes of the oral mucosa. The African and Indian diseases differ in several respects; important features of post–kala-azar dermal leishmaniasis (PKDL) in these two regions are listed in **Table 212-2**, and disease in an Indian patient is depicted in **Fig. 212-4**.

Table 212-2 Clinical, Epidemiologic, and Therapeutic Features of Post–Kala-Azar Dermal Leishmaniasis: East Africa and the Indian Subcontinent

Feature	East Africa	Indian Subcontinent
Most affected country	Sudan	Bangladesh
Incidence among patients with VL	~50%	~2%
Interval between VL and PKDL	During VL to 6 months	6 months to 3 years
Age distribution	Mainly children	Any age
History of prior VL	Yes	Not necessarily
Rashes of PKDL in presence of active VL	Yes	No
Treatment with sodium stibogluconate	2–3 months	2–4 months
Natural course	Spontaneous cure in majority of patients	Spontaneous cure not reported

Abbreviations: PKDL, post–kala-azar dermal leishmaniasis; VL, visceral leishmaniasis.

Figure 212-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Post–kala-azar dermal leishmaniasis in an Indian patient. Note nodules of varying size involving the entire face. The face is erythematous, and the surface of some of the large nodules is discolored.

In PKDL, parasites are scanty in hypopigmented macules but may be seen and cultured more easily from nodular

lesions. Cellular infiltrates are heavier in nodules than in macules. Lymphocytes are the dominant cells; next most common are histiocytes and plasma cells. In about half of cases, epithelioid cells—scattered individually or forming compact granulomas—are seen. The diagnosis is based on history and clinical findings, but rK39 and other serologic tests are positive in most cases. Indian PKDL is treated with pentavalent antimonials for 60–120 days. This prolonged course frequently leads to noncompliance. The alternative—several courses of AmB spread over several months—is expensive and unacceptable for most patients. In East Africa, a majority of patients experience spontaneous healing. In those with persistent lesions, the response to 60 days of treatment with a pentavalent antimonial is good.

Cutaneous Leishmaniasis

CL can be broadly divided into Old World and New World forms. Old World CL caused by *L. tropica* is anthroponotic and is confined to urban or suburban areas throughout its range. Zoonotic CL is most commonly due to *L. major*, which naturally parasitizes several species of desert rodents that act as reservoirs over wide areas of the Middle East, Africa, and central and southern Asia. Local outbreaks of human disease are common. Major outbreaks currently affect Afghanistan and Pakistan in association with refugees and population movement. CL is increasingly seen in tourists and military personnel on mission in CL-endemic regions of countries like Afghanistan and Iraq and as a co-infection in HIV-infected patients. *L. aethiopica* is restricted to the highlands of Ethiopia, Kenya, and Uganda, where it is a natural parasite of hyraxes. New World CL is mainly zoonotic and is most often caused by *L. mexicana*, *L. (V.) panamensis*, and *L. amazonensis*. A wide range of forest animals act as reservoirs, and human infections with these species are predominantly rural. As a result of extensive urbanization and deforestation, *L. (V.) braziliensis* has adapted to peridomestic and urban animals, and CL due to this organism is increasingly becoming an urban disease. In the United States, a few cases of CL have been acquired indigenously in Texas.

IMMUNOPATHOGENESIS

As in VL, the proinflammatory (T_H1) response in CL may result in either asymptomatic or subclinical infection. However, in some individuals, the immune response causes ulcerative skin lesions, the majority of which will heal spontaneously, leaving a scar. Healing is usually followed by immunity to reinfection with that species of parasite.

CLINICAL FEATURES

A few days or weeks after the bite of a sandfly, a papule develops and grows into a nodule that ulcerates over some weeks or months. The base of the ulcer, which is usually painless, consists of necrotic tissue and crusted serum, but secondary bacterial infection sometimes occurs. The margins of the ulcer are raised and indurated. Lesions may be single or multiple and vary in size from 0.5 to >3 cm (**Fig. 212-5**). Lymphatic spread and lymph gland involvement may be palpable and may precede the appearance of the skin lesion. There may be satellite lesions, especially in *L. major* and *L. tropica* infections. The lesions usually heal spontaneously after 2–15 months. Lesions due to *L. major* and *L. mexicana* tend to heal rapidly, while those due to *L. tropica* and parasites of subspecies *Viannia* heal more slowly. In CL caused by *L. tropica*, new lesions—usually scaly, erythematous papules and nodules—develop in the center or periphery of a healed sore, a condition known as *leishmaniasis recidivans*. Lesions of *L. mexicana* and *L. (V.) peruviana* closely resemble those seen in the Old World; however, lesions on the pinna of the ear are common, chronic, and destructive in the former infections. *L. mexicana* is responsible for chiclero's ulcer, the so-called self-healing sore of Mexico. CL lesions on exposed body parts (e.g., the face and hands), permanent scar formation, and social stigmatization may cause anxiety and depression and may affect the quality of life of CL patients.

Figure 212-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cutaneous leishmaniasis in a Bolivian child. There are multiple ulcers resulting from several sandfly bites. The edges of the ulcers are raised. *(Courtesy of P. Desjeux.)*

DIFFERENTIAL DIAGNOSIS

A typical history (an insect bite followed by the events leading to ulceration) in a resident of or a traveler to an endemic focus strongly suggests CL. Cutaneous tuberculosis, fungal infections, leprosy, sarcoidosis, and malignant ulcers are sometimes mistaken for CL.

LABORATORY DIAGNOSIS

Demonstration of amastigotes in material obtained from a lesion remains the diagnostic gold standard. Microscopic examination of slit skin smears, aspirates, or biopsies of the lesion is used for detection of parasites. Culture of smear or biopsy material may yield *Leishmania*. PCR is more sensitive than microscopy and culture and allows identification of *Leishmania* to the species level. This information is important in decisions about therapy since responses to treatment can vary with the species. Isoenzyme profiling is used to determine species for research purposes.

TREATMENT: CUTANEOUS LEISHMANIASIS

Although lesions heal spontaneously in the majority of cases, their spread or persistence indicates that treatment may be needed. One or a few small lesions due to "self-healing species" can be treated with topical agents. Systemic treatment is required for lesions over the face, hands, or joints; multiple lesions; large ulcers; lymphatic spread; New World CL with the potential for development of ML; and CL in HIV co-infected patients.

A pentavalent antimonial is the first-line drug for all forms of CL and is used in a dose of 20 mg/kg for 20 days, as for VL. The exceptions to this rule are CL caused by *L. (V.) guyanensis*, for which pentamidine isethionate is the drug of choice (two injections of 4 mg of salt/kg separated by a 48-h interval), and CL due to *L. aethiopica*, which responds to paromomycin (16 mg/kg daily) but not to antimonials. Relapses usually respond to a second course of treatment. In Peru, topical imiquimod (5–7.5%) plus parenteral antimonials have been shown to cure CL more rapidly than antimonials alone. Azoles and triazoles have been used with mixed responses in both Old and New World CL but have not been adequately assessed for this indication in clinical trials. In *L. major* infection, oral fluconazole (200 mg/d for 6 weeks) resulted in a higher rate of cure than placebo (79% vs. 34%) and also cured infection faster. Adverse effects include gastrointestinal symptoms and hepatotoxicity. Ketoconazole (600 mg/d for 28 days) is 76–90% effective in CL due to *L. (V.) panamensis* and *L. mexicana* in Panama and Guatemala. Miltefosine has been used in CL in doses of 2.5 mg/kg for 28 days. This agent is effective against *L. major* infections. In Colombia, where CL is due to *L. (V.) panamensis*, miltefosine was also effective, with a cure rate of 91%. For *L. (V.) braziliensis* infections, however, the results with miltefosine are less consistent. Other drugs, such as dapsone, allopurinol, rifampin, azithromycin, and pentoxifylline, have been used either alone or in combinations, but most of the relevant studies have had design limitations that preclude meaningful conclusions.

Small lesions (≲3 cm in diameter) may conveniently be treated weekly until cure with an intralesional injection of a pentavalent antimonial at a dose adequate to blanch the lesion (0.2–2.0 mL). An ointment containing 15% paromomycin sulfate plus 12% methylbenzonium chloride cures 70% of lesions due to *L. major* in 20 days and may be suitable for lesions caused by other species. Heat therapy with an FDA-approved radiofrequency generator and cryotherapy with liquid nitrogen have also been used successfully.

DIFFUSE CUTANEOUS LEISHMANIASIS (DCL)

DCL is a rare form of leishmaniasis caused by *L. amazonensis* and *L. mexicana* in South and Central America and by *L. aethiopica* in Ethiopia and Kenya. DCL is characterized by the lack of a cell-mediated immune response to the parasite, the uncontrolled multiplication of which thus continues unabated. The DTH response is negative, and lymphocytes do not respond to leishmanial antigens in vitro. DCL patients have a polarized immune response with high levels of immunosuppressive cytokines, including IL-10, transforming growth factor (TGF) β, and IL-4, and low concentrations of IFN-γ. Profound immunosuppression leads to widespread cutaneous disease. Lesions may initially be confined to the face or a limb but spread over months or years to other areas of the skin. They may be symmetrically or asymmetrically distributed and include papules, nodules, plaques, and areas of diffuse infiltration. These lesions do not ulcerate. The overlying skin is usually erythematous in pale-skinned patients. The lesions are teeming with parasites, which are therefore easy to recover. DCL does not heal spontaneously and is difficult to treat. If relapse and drug resistance are to be prevented, treatment should be continued for some time after lesions have healed and parasites can no longer be isolated. In the New World, repeated 20-day courses of pentavalent antimonials are given, with an intervening drug-free period of 10 days. Miltefosine has been used for several months with a good initial response. Combinations should be tried. In Ethiopia, a combination of paromomycin (14 mg/kg per day) and sodium stibogluconate (10 mg/kg per day) is effective.

Mucosal Leishmaniasis

The subgenus *Viannia* is widespread from the Amazon basin to Paraguay and Costa Rica and is responsible for deep sores and for ML (Table 212-1). In *L. (V.) braziliensis* infections, cutaneous lesions may be simultaneously accompanied by mucosal spread of the disease or followed by spread years later. ML is caused typically by *L. (V.) braziliensis* and rarely by *L. amazonensis*, *L. (V.) guyanensis*, and *L. (V.) panamensis*. Young men with chronic lesions of CL are at particular risk. Overall, ~3% of infected persons develop ML. Not every patient with ML has a history of prior CL. ML is almost entirely confined to the Americas. In rare cases, ML may also be caused by Old World species like *L. major*, *L. infantum*, or *L. donovani*.

IMMUNOPATHOGENESIS AND CLINICAL FEATURES

The immune response is polarized toward a T_H1 response, with marked increases of IFN-γ and TNF-α and varying levels of T_H2 cytokines (IL-10 and TGF-β). Patients have a stronger DTH response with ML than with CL, and their peripheral-blood mononuclear cells respond strongly to leishmanial antigens. The parasite spreads via the lymphatics or the bloodstream to mucosal tissues of the upper respiratory tract. Intense inflammation leads to destruction, and severe disability ensues. Lesions in or around the nose or mouth (espundia; **Fig. 212-6**) are the typical presentation of ML. Patients usually provide a history of self-healed CL preceding ML by 1–5 years. Typically, ML presents as nasal stuffiness and bleeding followed by destruction of nasal cartilage, perforation of the nasal septum, and collapse of the nasal bridge. Subsequent involvement of the pharynx and larynx leads to difficulty in swallowing and phonation. The lips, cheeks, and soft palate may also be affected. Secondary bacterial infection is common, and aspiration pneumonia may be fatal. Despite the high degree of T_H1 immunity and the strong DTH response, ML does not heal spontaneously.

Figure 212-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mucosal leishmaniasis in a Brazilian patient. There is extensive inflammation around the nose and mouth, destruction of the nasal mucosa, ulceration of the upper lip and nose, and destruction of the nasal septum. *(Courtesy of R. Dietz.)*

LABORATORY DIAGNOSIS

Tissue biopsy is essential for identification of parasites, but the rate of detection is poor unless PCR techniques are used. The strongly positive DTH response fails to distinguish between past and present infection.

TREATMENT: MUCOSAL LEISHMANIASIS

The regimen of choice is a pentavalent antimonial agent administered at a dose of 20 mg of Sb^{V}/kg for 30 days. Patients with ML require long-term follow-up with repeated oropharyngeal and nasal examination. With failure of therapy or relapse, patients may receive another course of an antimonial but then become unresponsive, presumably because of resistance in the parasite. In this situation, AmB should be used. An AmB deoxycholate dose totaling 25–45 mg/kg is appropriate. There are no controlled trials of liposomal AmB, but administration of 2–3 mg/kg for 20 days is considered adequate. Miltefosine (2.5 mg/kg for 28 days) cured 71% of ML patients in Bolivia. The more extensive the disease, the worse the prognosis; thus prompt, effective treatment and regular follow-up are essential.

Prevention of Leishmaniasis

No vaccine is available for any form of leishmaniasis. Inoculation with live *L. major* ("leishmanization") is practiced in Iran. Anthroponotic leishmaniasis is controlled by case finding, treatment, and vector control with insecticide-impregnated bed nets and curtains and residual insecticide spraying. Control of zoonotic leishmaniasis is more difficult. Use of insecticide-impregnated collars for dogs, treatment of infected domestic dogs, and culling of street dogs are measures that have been used with uncertain efficacy to prevent transmission of *L. infantum*. Personal prophylaxis with bed nets and repellants may reduce the risk of CL infections in the New World.

FURTHER READINGS

Alvar J et al: The relationship between leishmaniasis and AIDS: The second 10 years. Clin Microbiol Rev 21:334, 2008[PMID: 18400800]

Amato VS et al: Mucosal leishmaniasis. Current scenario and prospects for treatment. Acta Trop 105:1, 2008[PMID: 17884002]

Chappuis F et al: Visceral leishmaniasis: What are the needs for diagnosis, treatment and control? Nat Rev Microbiol 11:873, 2007

den Boer ML et al: Developments in the treatment of visceral leishmaniasis.

Expert Opin Emerg Drugs 14:395, 2009

Magill AJ: *Leishmania* species: Visceral (kala-azar), cutaneous and mucosal leishmaniasis, in *Principles and Practice of Infectious Diseases,* 7th ed, GL Mandell et al (eds). Philadelphia, Churchill Livingstone Elsevier, 2010, pp 3463–3480

Reithinger R et al: Cutaneous leishmaniasis. Lancet Infect Dis 7:581, 2007[PMID: 17714672]

Sundar S et al: Single-dose liposomal amphotericin B for visceral leishmaniasis in India. N Engl J Med 362:504, 2010[PMID: 20147716]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHAGAS' DISEASE AND TRYPANOSOMIASIS: INTRODUCTION

Although the genus *Trypanosoma* contains many species of protozoans, only *T. cruzi*, *T. brucei gambiense*, and *T. brucei rhodesiense* cause disease in humans. *T. cruzi* is the etiologic agent of Chagas' disease in the Americas; *T. b. gambiense* and *T. b. rhodesiense* cause African trypanosomiasis.

CHAGAS' DISEASE

Definition

Chagas' disease, or American trypanosomiasis, is a zoonosis caused by the protozoan parasite *T. cruzi*. Acute Chagas' disease is usually a mild febrile illness that results from initial infection with the organism. After spontaneous resolution of the acute illness, most infected persons remain for life in the indeterminate phase of chronic Chagas' disease, which is characterized by subpatent parasitemia, easily detectable antibodies to *T. cruzi*, and an absence of associated signs and symptoms. In 10–30% of chronically infected patients, cardiac and/or gastrointestinal lesions develop that can result in serious morbidity and even death.

Life Cycle and Transmission

T. cruzi is transmitted among its mammalian hosts by hematophagous triatomine insects, often called reduviid bugs. The insects become infected by sucking blood from animals or humans who have circulating parasites. Ingested organisms multiply in the gut of the triatomines, and infective forms are discharged with the feces at the time of subsequent blood meals. Transmission to a second vertebrate host occurs when breaks in the skin, mucous membranes, or conjunctivae become contaminated with bug feces that contain infective parasites. *T. cruzi* can also be transmitted by the transfusion of blood donated by infected persons, by organ transplantation, from mother to unborn child, by ingestion of contaminated food or drink, and in laboratory accidents.

Pathology

Initial infection at the site of parasite entry is characterized by local histologic changes that include the presence of parasites within leukocytes and cells of subcutaneous tissues and the development of interstitial edema, lymphocytic infiltration, and reactive hyperplasia of adjacent lymph nodes. After dissemination of the organisms through the lymphatics and the bloodstream, primarily muscles (including the myocardium) **(Fig. 213-1)** and ganglion cells may become heavily parasitized. The characteristic pseudocysts present in sections of infected tissues are intracellular aggregates of multiplying parasites.

Figure 213-1

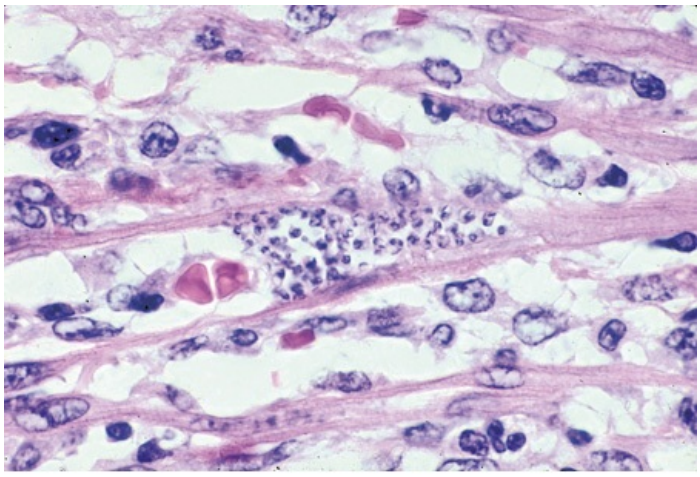

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***Trypanosoma cruzi* in the heart muscle of a child who died of acute Chagas' myocarditis.** An infected myocyte containing several dozen *T. cruzi* amastigotes is in the center of the field (hematoxylin and eosin, 900x).

In individuals with chronic *T. cruzi* infections who develop related clinical manifestations, the heart is the organ most commonly affected. Changes include thinning of the ventricular walls, biventricular enlargement, apical aneurysms, and mural thrombi. Widespread lymphocytic infiltration, diffuse interstitial fibrosis, and atrophy of myocardial cells are often apparent, but parasites are difficult to find in myocardial tissue by conventional histologic methods. Conduction-system abnormalities often affect the right branch and the left anterior branch of the bundle of His. In chronic Chagas' disease of the gastrointestinal tract (megadisease), the esophagus and colon may exhibit varying degrees of dilatation. On microscopic examination, focal inflammatory lesions with lymphocytic infiltration are seen, and the number of neurons in the myenteric plexus may be markedly reduced. Accumulating evidence implicates the persistence of parasites and the accompanying chronic inflammation—rather than autoimmune mechanisms—as the basis for the pathology in patients with chronic *T. cruzi* infection.

Epidemiology

T. cruzi is found only in the Americas. Wild and domestic mammals harboring *T. cruzi* and infected triatomines are found in spotty distributions from the southern United States to southern Argentina. Humans become involved in the cycle of transmission when infected vectors take up residence in the primitive wood, adobe, and stone houses common in much of Latin America. Thus human *T. cruzi* infection is a health problem primarily among the poor in rural areas of Mexico and Central and South America. Most new *T. cruzi* infections in rural settings occur in children, but the incidence is unknown because most cases go undiagnosed. Historically, transfusion-associated transmission of *T. cruzi* was a serious public health problem in many endemic countries. However, with some notable exceptions, transmission by this route has been essentially eliminated as effective programs for the screening of donated blood have been implemented. Several dozen patients with HIV and chronic *T. cruzi* infections who underwent acute recrudescence of the latter have been described. These patients generally presented with *T. cruzi* brain abscesses, a manifestation of the illness that does not occur in immunocompetent persons. Currently, it is estimated that 8 million people are chronically infected with *T. cruzi* and that 14,000 deaths due to the illness occur each year. The resulting morbidity and mortality make Chagas' disease the most important parasitic disease burden in Latin America.

In recent years, the rate of *T. cruzi* transmission has decreased markedly in several endemic countries as a result of successful programs involving vector control, blood-bank screening, and education of at-risk populations. A major program, which began in 1991 in the "southern cone" nations of South America (Uruguay, Paraguay, Bolivia, Brazil, Chile, and Argentina), has provided the framework for much of this progress. Uruguay and Chile were certified free of transmission by the main domiciliary vector species (*Triatoma infestans*) in the late 1990s, and Brazil was declared transmission-free in 2006. Transmission has been reduced markedly in Argentina as well. Similar control programs have been initiated in the countries of northern South America and in the Central American nations.

Acute Chagas' disease is rare in the United States. Five cases of autochthonous transmission and five instances of transmission by blood transfusion have been reported. Moreover, *T. cruzi* was transmitted to five recipients of organs from three *T. cruzi*–infected donors. Two of these recipients became infected through cardiac transplants. Acute Chagas' disease has not been reported in tourists returning to the United States from Latin America, although three such instances have been reported in Europe. In contrast, the prevalence of chronic *T. cruzi* infections in the United States has increased considerably in recent years. An estimated 23 million immigrants from Chagas'-endemic countries currently live in the United States, ~17 million of whom are Mexicans. The total number of *T. cruzi*–infected persons living in the United States is estimated to be 300,000. Screening of the U.S. blood supply for *T. cruzi* infection began in January 2007. The overall prevalence of *T. cruzi* infection among donors is about 1 in 29,000, and to date more than 1200 infected donors have been identified and deferred permanently (see "Diagnosis," below).

Clinical Course

The first signs of acute Chagas' disease develop at least 1 week after invasion by the parasites. When the organisms enter through a break in the skin, an indurated area of erythema and swelling (the chagoma), accompanied by local lymphadenopathy, may appear. *Romaña's sign*—the classic finding in acute Chagas' disease, which consists of unilateral painless edema of the palpebrae and periocular tissues—can result when the conjunctiva is the portal of entry. These initial local signs may be followed by malaise, fever, anorexia, and edema of the face and lower extremities. Generalized lymphadenopathy and hepatosplenomegaly may develop. Severe myocarditis develops rarely; most deaths in acute Chagas' disease are due to heart failure. Neurologic signs are not common, but meningoencephalitis occurs occasionally, especially in children <2 years old. Usually within 4–8 weeks, acute signs and symptoms resolve spontaneously in virtually all patients, who then enter the asymptomatic or indeterminate phase of chronic *T. cruzi* infection.

Symptomatic chronic Chagas' disease becomes apparent years or even decades after the initial infection. The heart is commonly involved, and symptoms are caused by rhythm disturbances, segmental or dilated cardiomyopathy, and thromboembolism. Right bundle-branch block is a common electrocardiographic abnormality, but other types of intraventricular and atrioventricular blocks, premature ventricular contractions, and tachy- and bradyarrhythmias occur frequently. Cardiomyopathy often results in biventricular heart failure with a predominance of right-sided failure at advanced stages. Embolization of mural thrombi to the brain or other areas may take place. Sudden death is the main cause of death in Chagas' heart disease. Patients with megaesophagus suffer from dysphagia, odynophagia, chest pain, and regurgitation. Aspiration can occur (especially during sleep) in patients with severe esophageal dysfunction, and repeated episodes of aspiration pneumonitis are common. Weight loss, cachexia, and pulmonary infection can result in death. Patients with megacolon are plagued by abdominal pain and chronic constipation, which predisposes to fecaloma formation. Advanced megacolon can cause obstruction, volvulus, septicemia, and death.

Diagnosis

The diagnosis of acute Chagas' disease requires the detection of parasites. Microscopic examination of fresh anticoagulated blood or the buffy coat is the simplest way to see the motile organisms. Parasites also can be seen in Giemsa-stained thin and thick blood smears. Microhematocrit tubes containing acridine orange as a stain can be used for the same purpose. When used by experienced personnel, all of these methods yield positive results in a high proportion of cases of acute Chagas' disease. Serologic testing plays no role in diagnosing acute Chagas' disease.

Chronic Chagas' disease is diagnosed by the detection of specific IgG antibodies that bind to *T. cruzi* antigens. Demonstration of the parasite is not of primary importance. In Latin America, ~30 assays are commercially available, including several based on recombinant antigens. Although these tests usually show good sensitivity and reasonable specificity, false-positive reactions may occur—typically with samples from patients who have other infectious and parasitic diseases or autoimmune disorders. In addition, confirmatory testing has presented a persistent challenge. For these reasons, the World Health Organization recommends that specimens be tested in at least two assays and that well-characterized positive and negative comparison samples be included in each run. The radioimmune precipitation assay (Chagas RIPA) is a highly sensitive and specific confirmatory method for detecting antibodies to *T. cruzi* (approved under the Clinical Laboratory Improvement Amendment and available in the authors' laboratory). In December 2006, the U.S. Food and Drug Administration (FDA) approved a test to screen blood and organ donors for *T. cruzi* infection (Ortho *T. cruzi* ELISA Test System, Ortho-Clinical Diagnostics, Raritan, NJ). Since January 2007, the vast majority of U.S. blood donors have been screened with the Ortho test, and positive units have undergone confirmatory testing in the Chagas RIPA. A second test for donor screening was approved by the FDA in April 2010 (Abbott PRISM® Chagas Assay, Abbott Laboratories, Abbott Park, IL). The use of PCR assays to detect *T. cruzi* DNA in chronically infected persons has been studied extensively. The sensitivity of this approach has not been shown to be reliably greater than that of serology, and no PCR assays are commercially available.

Treatment: Chagas' Disease

Therapy for Chagas' disease is still unsatisfactory. For many years now, only two drugs—nifurtimox and benznidazole—have been available for this purpose. Unfortunately, both drugs lack efficacy and may cause bothersome side effects.

In acute Chagas' disease, nifurtimox markedly reduces the duration of symptoms and parasitemia and decreases the mortality rate. Nevertheless, limited studies have shown that only ~70% of acute infections are cured parasitologically by a full course of treatment. Common adverse effects of nifurtimox include anorexia, nausea, vomiting, weight loss, and abdominal pain. Neurologic reactions to the drug may include restlessness, disorientation, insomnia, twitching, paresthesia, polyneuritis, and seizures. These symptoms usually disappear when the dosage is reduced or treatment is discontinued. The recommended daily dosage is 8–10 mg/kg for adults, 12.5–15 mg/kg for adolescents, and 15–20 mg/kg for children 1–10 years of age. The drug should be given orally in four divided doses each day, and therapy should be continued for 90–120 days. Nifurtimox is available from the Drug Service of the Centers for Disease Control and Prevention (CDC) in Atlanta (telephone number, 404-639-3670).

The efficacy of benznidazole is similar or even superior to that of nifurtimox. A cure rate of 90% among congenitally infected infants treated before their first birthday has been reported. Adverse effects include rash, peripheral neuropathy, and rarely granulocytopenia. The recommended oral dosage is 5 mg/kg per day for 60 days for adults and 5–10 mg/kg per day for 60 days for children, with administration of two or three divided doses. Benznidazole is generally considered the drug of choice in Latin America.

The question of whether adults in the indeterminate or chronic symptomatic phase of Chagas' disease should be treated with nifurtimox or benznidazole has been debated for years. The fact that parasitologic cure rates in chronically infected persons are notably inferior to those in patients with

acute or recent chronic infection is central to this controversy. No convincing evidence from randomized controlled trials indicates that nifurtimox or benznidazole treatment of adults in the indeterminate or chronic symptomatic phase reduces either the appearance and progression of symptoms or mortality rates. On the basis of results of some observational studies, a panel of experts convened by the CDC in 2006 recommended that adults <50 years old with presumably long-standing indeterminate *T. cruzi* infections—or even with mild to moderate disease—be offered treatment. A large randomized clinical trial (the BENEFIT multicenter trial) designed to assess the parasitologic and clinical efficacy of benznidazole in adults (18–75 years old) with chronic Chagas' heart disease (without advanced lesions) is being performed in Brazil, Argentina, Colombia, and Bolivia, but results will not be available until 2012 at the earliest. In contrast, randomized studies have shown that treatment of children is useful, and the current consensus of Latin American authorities is that all *T. cruzi*–infected persons up to 18 years old and all adults known to have become infected recently should be given benznidazole or nifurtimox.

The usefulness of antifungal azoles for the treatment of Chagas' disease has been studied in laboratory animals and to a lesser extent in humans. To date, none of these drugs has exhibited a level of anti–*T. cruzi* activity that would justify its use in humans. Several newer drugs in this class have shown promise in animal studies and are likely to undergo human trials in the near future.

Patients who develop cardiac and/or gastrointestinal disease in association with *T. cruzi* infection should be referred to appropriate subspecialists for further evaluation and treatment. Cardiac transplantation is an option for patients with end-stage chagasic cardiomyopathy; more than 150 such transplantations have been done in Brazil and the United States. The survival rate among Chagas' disease cardiac transplant recipients seems to be higher than that among persons receiving cardiac transplants for other reasons. This better outcome may be due to the fact that lesions are limited to the heart in most patients with symptomatic chronic Chagas' disease.

Prevention

Since drug therapy has limitations and vaccines are not available, the control of *T. cruzi* transmission in endemic countries depends on the reduction of domiciliary vector populations by spraying of insecticides, improvements in housing, and education of at-risk persons. As noted above, these measures, coupled with serologic screening of blood donors, have markedly reduced transmission of the parasite in many endemic countries. Tourists would be wise to avoid sleeping in dilapidated houses in rural areas of endemic countries. Mosquito nets and insect repellent can provide additional protection.

In view of the possibly serious consequences of chronic *T. cruzi* infection, it would be prudent for all immigrants from endemic regions who are living in the United States to be tested for evidence of infection. Identification of persons harboring the parasite would permit periodic electrocardiographic monitoring, which can be important because pacemakers benefit some patients who develop ominous rhythm disturbances. The possibility of congenital transmission is yet another justification for screening. *T. cruzi* is classified as a Risk Group 2 agent in the United States and a Risk Group 3 agent in some European countries. Laboratory staff should work with the parasite or infected vectors at containment levels consistent with the risk group designation in their areas.

SLEEPING SICKNESS

Definition

Sleeping sickness, or human African trypanosomiasis (HAT), is caused by flagellated protozoan parasites that belong to the *T. brucei* complex and are transmitted to humans by tsetse flies. In untreated patients, the trypanosomes first cause a febrile illness that is followed months or years later by progressive neurologic impairment and death.

The Parasites and Their Transmission

The East African (*rhodesiense*) and the West African (*gambiense*) forms of sleeping sickness are caused, respectively, by two trypanosome subspecies: *T. b. rhodesiense* and *T. b. gambiense*. These subspecies are morphologically indistinguishable but cause illnesses that are epidemiologically and clinically distinct **(Table 213-1)**. The parasites are transmitted by blood-sucking tsetse flies of the genus *Glossina*. The insects acquire the infection when they ingest blood from infected mammalian hosts. After many cycles of multiplication in the midgut of the vector, the parasites migrate to the salivary glands. Their transmission takes place when they are inoculated into a mammalian host during a subsequent blood meal. The injected trypanosomes multiply in the blood **(Fig. 213-2)** and other extracellular spaces and evade immune destruction for long periods by undergoing antigenic variation, a process driven by gene switching in which the antigenic structure of the organisms' surface coat of glycoproteins changes periodically.

Table 213-1 Comparison of West African and East African Trypanosomiases

Point of Comparison	West African (*Gambiense*)	East African (*Rhodesiense*)
Organism	*T. b. gambiense*	*T. b. rhodesiense*
Vectors	Tsetse flies (palpalis group)	Tsetse flies (morsitans group)
Primary reservoir	Humans	Antelope and cattle
Human illness	Chronic (late CNS disease)	Acute (early CNS disease)
Duration of illness	Months to years	<9 months
Lymphadenopathy	Prominent	Minimal
Parasitemia	Low	High
Diagnosis by rodent inoculation	No	Yes
Epidemiology	Rural populations	Workers in wild areas, rural populations, tourists in game parks

Abbreviation: CNS, central nervous system.

Source: Reprinted with permission from LV Kirchhoff in GL Mandell et al (eds): *Principles and Practice of Infectious Diseases,* 7th ed. Philadelphia, Elsevier Churchill Livingstone, 2010.

Figure 213-2

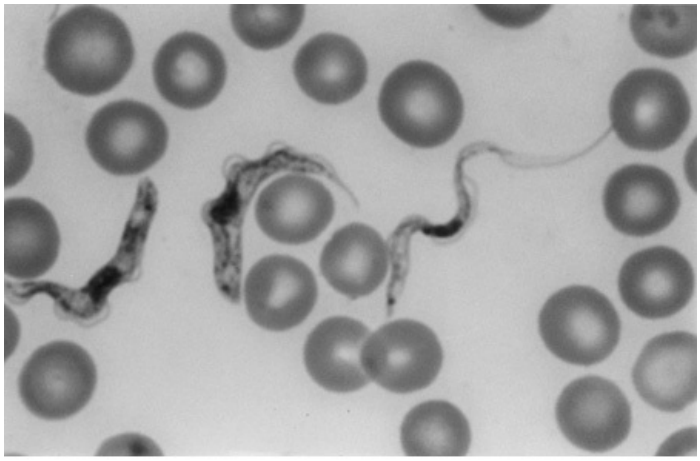

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Trypanosoma brucei rhodesiense parasites in rat blood. The slender parasite is thought to be the form that multiplies in mammalian hosts, while the stumpy forms are nondividing and are capable of infecting insect vectors (Giemsa, 1200x). *(Courtesy of Dr. G. A. Cook, Madison, WI; with permission.)*

Pathogenesis and Pathology

A self-limited inflammatory lesion (trypanosomal chancre) may appear a week or so after the bite of an infected tsetse fly. A systemic febrile illness then evolves as the parasites are disseminated through the lymphatics and bloodstream. Systemic HAT without central nervous system (CNS) involvement is generally referred to as *stage I disease*. In this stage, widespread lymphadenopathy and splenomegaly reflect marked lymphocytic and histiocytic proliferation and invasion of morular cells, which are plasmacytes that may be involved in the production of IgM. Endarteritis, with perivascular infiltration of both parasites and lymphocytes, may develop in lymph nodes and the spleen. Myocarditis develops frequently in patients with stage I disease and is especially common in *T. b. rhodesiense* infections.

Hematologic manifestations that accompany stage I HAT include moderate leukocytosis, thrombocytopenia, and anemia. High levels of immunoglobulins, consisting primarily of polyclonal IgM, are a constant feature, and heterophile antibodies, antibodies to DNA, and rheumatoid factor are often detected. High levels of antigen-antibody complexes may play a role in the tissue damage and increased vascular permeability that facilitate dissemination of the parasites.

Stage II disease involves invasion of the CNS. The presence of trypanosomes in perivascular areas is accompanied by intense infiltration of mononuclear cells. Abnormalities in cerebrospinal fluid (CSF) include increased pressure, elevated total protein concentration, and pleocytosis. In addition,

trypanosomes are frequently found in CSF.

Epidemiology

The trypanosomes that cause sleeping sickness are found only in sub-Saharan Africa. After its near-eradication in the mid-1960s, sleeping sickness underwent a resurgence in the 1990s, primarily in Uganda, Sudan, the Central African Republic, the Democratic Republic of the Congo, and Angola. Although a subsequent increase in control activities reduced the incidence in many endemic areas, the World Health Organization estimated that there were 50,000–70,000 new cases in 2004, the vast majority of which were caused by *T. b. gambiense*. Approximately 50 million persons are at risk of acquiring HAT.

Humans are the only reservoir of *T. b. gambiense*, which occurs in widely distributed foci in tropical rain forests of Central and West Africa. *Gambiense* trypanosomiasis is primarily a problem in rural populations; tourists rarely become infected. Trypanotolerant antelope species in savanna and woodland areas of Central and East Africa are the principal reservoir of *T. b. rhodesiense*. Cattle can also be infected with this and other trypanosome species but generally succumb to the infection. Since risk results from contact with tsetse flies that feed on wild animals, humans acquire *T. b. rhodesiense* infection only incidentally, usually while visiting or working in areas where infected game and vectors are present. Roughly one or two imported cases of HAT acquired in East African parks are reported to the CDC each year.

Clinical Course

A painful trypanosomal chancre appears in some patients at the site of inoculation of the parasite. Hematogenous and lymphatic dissemination (stage I disease) is marked by the onset of fever. Typically, bouts of high temperatures lasting several days are separated by afebrile periods. Lymphadenopathy is prominent in *T. b. gambiense* trypanosomiasis. The nodes are discrete, movable, rubbery, and nontender. Cervical nodes are often visible, and enlargement of the nodes of the posterior cervical triangle, or *Winterbottom's sign*, is a classic finding. Pruritus and maculopapular rashes are common. Inconstant findings include malaise, headache, arthralgias, weight loss, edema, hepatosplenomegaly, and tachycardia. The differential diagnosis of stage I HAT includes many diseases that are common in the tropics and are associated with fevers. HIV infection, malaria, and typhoid fever are common in populations at risk for HAT and need to be considered.

CNS invasion (stage II disease) is characterized by the insidious development of protean neurologic manifestations that are accompanied by progressive abnormalities in the CSF. A picture of progressive indifference and daytime somnolence develops (hence the designation "sleeping sickness"), sometimes alternating with restlessness and insomnia at night. A listless gaze accompanies a loss of spontaneity, and speech may become halting and indistinct. Extrapyramidal signs may include choreiform movements, tremors, and fasciculations. Ataxia is frequent, and the patient may appear to have Parkinson's disease, with a shuffling gait, hypertonia, and tremors. In the final phase, progressive neurologic impairment ends in coma and death.

The most striking difference between the *gambiense* and *rhodesiense* forms of HAT is that the latter illness tends to follow a more acute course. Typically, in tourists with *T. b. rhodesiense* disease, systemic signs of infection, such as fever, malaise, and headache, appear before the end of the trip or shortly after the return home. Persistent tachycardia unrelated to fever is common early in the course of *T. b. rhodesiense* trypanosomiasis, and death may result from arrhythmias and congestive heart failure before CNS disease develops. In general, untreated *T. b. rhodesiense* trypanosomiasis leads to death in a matter of weeks to months, often without a clear distinction between the hemolymphatic and CNS stages. In contrast, *T. b. gambiense* disease can smolder for many months or even for years.

Diagnosis

A definitive diagnosis of HAT requires detection of the parasite. If a chancre is present, fluid should be expressed and examined directly by light microscopy for the highly motile trypanosomes. The fluid also should be fixed and stained with Giemsa. Material obtained by needle aspiration of lymph nodes early in the illness should be examined similarly. Examination of wet preparations and Giemsa-stained thin and thick films of serial blood samples is also useful. If parasites are not seen initially in blood, efforts should be made to concentrate the organisms, which can be done in microhematocrit tubes containing acridine orange. Alternatively, the buffy coat from 10–15 mL of anticoagulated blood can be examined directly under a microscope. The likelihood of finding parasites in blood is higher in stage I than in stage II disease and in patients infected with *T. b. rhodesiense* rather than *T. b. gambiense*. Trypanosomes may also be seen in material aspirated from the bone marrow; the aspirate can be inoculated into liquid culture medium, as can blood, buffy coat, lymph node aspirates, and CSF. Finally, *T. b. rhodesiense* infection can be detected by inoculation of these specimens into mice or rats, which —when positive—results in patent parasitemias in a week or two. Although this method is highly sensitive for the detection of *T. b. rhodesiense*, it does not detect *T. b. gambiense* because of host specificity.

It is essential to examine CSF from all patients in whom HAT is suspected. Abnormalities in the CSF

that may be associated with stage II disease include an increase in the CSF mononuclear cell count as well as increases in opening pressure and in levels of total protein and IgM. Trypanosomes may be seen in the sediment of centrifuged CSF. Any CSF abnormality in a patient in whom trypanosomes have been found at other sites must be viewed as pathognomonic for CNS involvement and thus must prompt specific treatment for CNS disease. In patients with CSF pleocytosis in whom parasites are not found, tuberculous meningitis and HIV-associated CNS infections such as cryptococcosis should be considered in the differential diagnosis.

A number of serologic assays, such as the card agglutination test for trypanosomes (CATT) for *T. b. gambiense*, are available to aid in the diagnosis of HAT, but their variable sensitivity and specificity mandate that decisions about treatment be based on demonstration of the parasite. These tests are of value for epidemiologic surveys. PCR assays for detecting African trypanosomes in humans have been developed, but none is commercially available.

Treatment: Sleeping Sickness

The drugs used for treatment of HAT are suramin, pentamidine, eflornithine, and the organic arsenical melarsoprol. In the United States, these drugs can be obtained from the CDC. Therapy for HAT must be individualized on the basis of the infecting subspecies, the presence or absence of CNS disease, adverse reactions, and occasionally drug resistance. The choices of drugs for the treatment of HAT are summarized in **Table 213-2**.

Table 213-2 Treatment of Human African Trypanosomiases[a]

	Clinical Stage	
Causative Organism	**I (Normal CSF)**	**II (Abnormal CSF)**
T. brucei gambiense (West African)	Pentamidine Alternative: Suramin	Eflornithine Alternative: Melarsoprol
T. brucei rhodesiense(East African)	Suramin	Melarsoprol

[a]For doses and duration, see text.

Abbreviation: CSF, cerebrospinal fluid.

Suramin is highly effective against stage I *rhodesiense* HAT. However, it can cause serious adverse effects and must be administered under the close supervision of a physician. A 100- to 200-mg IV test dose should be given to detect hypersensitivity. The dosage for adults is 20 mg/kg on days 1, 5, 12, 18, and 26. The drug is given by slow IV infusion of a freshly prepared 10% aqueous solution. Approximately 1 patient in 20,000 has an immediate, severe, and potentially fatal reaction to the drug, developing nausea, vomiting, shock, and seizures. Less severe reactions include fever, photophobia, pruritus, arthralgias, and skin eruptions. Renal damage is the most common important adverse effect of suramin. Transient proteinuria often appears during treatment. A urinalysis should be done before each dose, and treatment should be discontinued if proteinuria increases or if casts and red cells appear in the sediment. Suramin should not be given to patients with renal insufficiency.

Pentamidine is the first-line drug for treatment of stage I *gambiense* HAT. The dose for both adults and children is 4 mg/kg per day, given IM or IV for 7–10 days. Frequent, immediate adverse reactions include nausea, vomiting, tachycardia, and hypotension. These reactions are usually transient and do not warrant cessation of therapy. Other adverse reactions include nephrotoxicity, abnormal liver function tests, neutropenia, rashes, hypoglycemia, and sterile abscesses. Suramin is an alternative agent for stage I *T. b. gambiense* disease.

Eflornithine is highly effective for treatment of both stages of *gambiense* sleeping sickness. In the trials on which the FDA based its approval, this agent cured >90% of 600 patients with stage II disease. The recommended treatment schedule is 400 mg/kg per day, given IV in four divided doses, for 2 weeks. Adverse reactions include diarrhea, anemia, thrombocytopenia, seizures, and hearing loss. The high dosage and duration of therapy required are disadvantages that make widespread use of eflornithine difficult. A randomized trial comparing the standard eflornithine regimen (400 mg/kg per day infused over 6 h for 14 days) with nifurtimox-eflornithine combination therapy (oral nifurtimox, 15 mg/kg per day for 10 days; plus eflornithine, 400 mg/kg per day infused over 12 h for 7 days) in adults with stage II *gambiense* HAT showed improved efficacy and reduced adverse effects with combination therapy, making it suitable for first-line use.

The arsenical melarsoprol is the drug of choice for the treatment of *rhodesiense* HAT with CNS involvement and is an alternative agent for stage II *gambiense* disease. For *rhodesiense* disease, the drug should be given to adults in three courses of 3 days each. The dosage is 2.0–3.6 mg/kg per day, given IV in three divided doses for 3 days and followed 1 week later by 3.6 mg/kg per day, also in three divided doses and for 3 days. The latter course is repeated 7 days later. In debilitated patients,

suramin is administered for 2–4 days before therapy with melarsoprol is initiated; an 18-mg initial dose of the latter drug, followed by progressive increases to the standard dose, has been recommended. For children, a total of 18–25 mg/kg should be given over 1 month. An IV starting dose of 0.36 mg/kg should be increased gradually to a maximum of 3.6 mg/kg at 1- to 5-day intervals, for a total of 9 or 10 doses. The regimen for *gambiense* disease is 2.2 mg/kg per day, given IV for 10 days.

Melarsoprol is highly toxic and should be administered with great care. To reduce the likelihood of drug-induced encephalopathy, all patients receiving melarsoprol should be given prednisolone at a dose of 1 mg/kg (up to 40 mg) per day, beginning 1–2 days before the first dose of melarsoprol and continuing through the last dose. Without prednisolone prophylaxis, the incidence of reactive encephalopathy has been reported to be as high as 18% in some series. Clinical manifestations of reactive encephalopathy include high fever, headache, tremor, impaired speech, seizures, and even coma and death. Treatment with melarsoprol should be discontinued at the first sign of encephalopathy but may be restarted cautiously at lower doses a few days after signs have resolved. Extravasation of the drug results in intense local reactions. Vomiting, abdominal pain, nephrotoxicity, and myocardial damage can occur.

Prevention

HAT poses complex public-health and epizootic problems in Africa. Considerable progress has been made in some areas through control programs that focus on eradication of vectors and drug treatment of infected humans. People can reduce their risk of acquiring trypanosomiasis by avoiding areas known to harbor infected insects, by wearing protective clothing, and by using insect repellent. Chemoprophylaxis is not recommended, and no vaccine is available to prevent transmission of the parasites.

FURTHER READINGS

Bacal F et al: Transplantation for Chagas' disease: An overview of immunosuppression and reactivation in the last two decades. Clin Transplant 24:E29, 2010

Bern C et al: Evaluation and treatment of Chagas disease in the United States: A systematic review. JAMA 298:2171, 2007[PMID: 18000201]

———, Montgomery SP: An estimate of the burden of Chagas disease in the United States. Clin Infect Dis 49:e52, 2009

Brun R et al: Human African trypanosomiasis. Lancet 375:148, 2010[PMID: 19833383]

Chang CD et al: Evaluation of a prototype *Trypanosoma cruzi* antibody assay with recombinant antigens on a fully automated chemiluminescence analyzer for blood donor screening. Transfusion 46:1737, 2006[PMID: 17002630]

Kirchhoff LV et al: Transfusion-associated Chagas' disease (American trypanosomiasis) in Mexico: Implications for transfusion medicine in the United States. Transfusion 46:298, 2006[PMID: 16441610]

Priotto G et al: Nifurtimox-eflornithine combination therapy for second-stage African *Trypanosoma brucei gambiense* trypanosomiasis: A multicentre, randomised, phase III, non-inferiority trial. Lancet 374:56, 2009[PMID: 19559476]

Rassi A Jr et al: Development and validation of a risk score for predicting death in Chagas' heart disease. N Engl J Med 355:799, 2006[PMID: 16928995]

——— et al: Chagas disease. Lancet 375:1388, 2010

Sartori AM et al: Manifestations of Chagas disease (American trypanosomiasis) in patients with HIV/AIDS. Ann Trop Med Hyg 101:31, 2007[PMID: 17244408]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TOXOPLASMA INFECTIONS: INTRODUCTION

Definition

Toxoplasmosis is caused by infection with the obligate intracellular parasite *Toxoplasma gondii*. Acute infection acquired after birth may be asymptomatic but is thought to result in the lifelong chronic persistence of cysts in the host's tissues. In both acute and chronic toxoplasmosis, the parasite is responsible for clinically evident disease, including lymphadenopathy, encephalitis, myocarditis, and pneumonitis. Congenital toxoplasmosis is an infection of newborns that results from the transplacental passage of parasites from an infected mother to the fetus. These infants may be asymptomatic at birth, but most later manifest a wide range of signs and symptoms, including chorioretinitis, strabismus, epilepsy, and psychomotor retardation. In immunocompetent individuals, toxoplasmosis can also present as acute disease (typically chorioretinitis) associated with food- or waterborne sources.

Etiology

T. gondii is an intracellular coccidian that infects both birds and mammals. There are two distinct stages in the life cycle of *T. gondii* **(Fig. 214-1)**. In the *nonfeline* stage, tissue cysts that contain bradyzoites or sporulated oocysts are ingested by an intermediate host (e.g., a human, mouse, sheep, pig, or bird). The cyst is rapidly digested by the acidic-pH gastric secretions. Bradyzoites or sporozoites are released, enter the small-intestinal epithelium, and transform into rapidly dividing tachyzoites. The tachyzoites can infect and replicate in all mammalian cells except red blood cells. Once attached to the host cell, the parasite penetrates the cell and forms a parasitophorous vacuole within which it divides. Parasite replication continues until the number of parasites within the cell approaches a critical mass and the cell ruptures, releasing parasites that infect adjoining cells. As a result of this process, an infected organ soon shows evidence of cytopathology. Most tachyzoites are eliminated by the host's humoral and cell-mediated immune responses. Tissue cysts containing many bradyzoites develop 7–10 days after systemic tachyzoite infection. These tissue cysts occur in various host organs but persist principally within the central nervous system (CNS) and muscle. The development of this chronic stage completes the nonfeline portion of the life cycle. Active infection in the immunocompromised host is most likely to be due to the spontaneous release of encysted parasites that undergo rapid transformation into tachyzoites within the CNS.

Figure 214-1

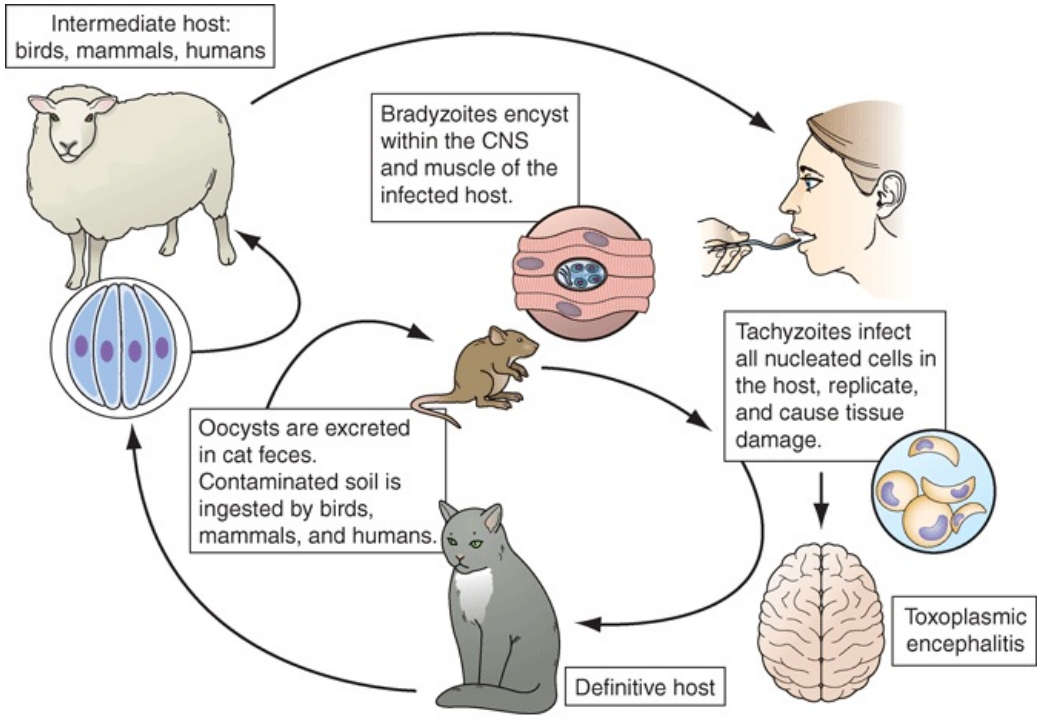

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Life cycle of *Toxoplasma gondii*. The cat is the definitive host in which the sexual phase of the cycle is completed. Oocysts shed in cat feces can infect a wide range of animals, including birds, rodents, grazing domestic animals, and humans. The bradyzoites found in the muscle of food animals may infect humans who eat insufficiently cooked meat products, particularly lamb and pork. Although human disease can take many forms, congenital infection and encephalitis from reactivation of latent infection in the brains of immunosuppressed persons are the most important manifestations. CNS, central nervous system. *(Courtesy of Dominique Buzoni-Gatel, Institut Pasteur, Paris; with permission.)*

The principal (*feline*) stage in the life cycle takes place in the cat (the definitive host) and its prey. The parasite's sexual phase is defined by the formation of oocysts within the feline host. This enteroepithelial cycle begins with the ingestion of the bradyzoite tissue cysts and culminates (after several intermediate stages) in the production of gametes. Gamete fusion produces a zygote, which envelops itself in a rigid wall and is secreted in the feces as an unsporulated oocyst. After 2–3 days of exposure to air at ambient temperature, the noninfectious oocyst sporulates to produce eight sporozoite progeny. The sporulated oocyst can be ingested by an intermediate host, such as a person emptying a cat's litter box or a pig rummaging in a barnyard. It is in the intermediate host that *T. gondii* completes its life cycle. Sporulated oocysts, which are environmentally hardy and very infectious, are thought to be sources of waterborne outbreaks such as those reported in Victoria (British Columbia, Canada) and in South America.

Epidemiology

T. gondii infects a wide range of mammals and birds. Its seroprevalence depends on the locale and the age of the population. Generally, hot arid climatic conditions are associated with a low prevalence of infection. In the United States and most European countries, the seroprevalence increases with age and exposure. For example, in the United States, 5–30% of individuals 10–19 years old and 10–67% of those >50 years old have serologic evidence of exposure. In Central America, France, Turkey, and Brazil, the seroprevalence is higher. Because of increased awareness of food-borne infections, the prevalence of seropositivity has decreased worldwide. There may be as many as 2100 cases of toxoplasmic encephalitis (TE) each year in the United States.

Transmission

ORAL TRANSMISSION

The principal source of human *Toxoplasma* infection remains uncertain, but infection is thought to occur by the oral route. Transmission can be attributable to ingestion of either sporulated oocysts from contaminated soil, food, or water or bradyzoites from undercooked meat. During acute feline infection, a cat may excrete as many as 100 million parasites per day. These very stable sporozoite-containing oocysts are highly infectious and may remain viable for many years in soil or water. Humans infected during a well-documented outbreak of oocyst-transmitted infection develop stage-specific antibodies to the oocyst/sporozoite.

Children and adults also can acquire infection from tissue cysts containing bradyzoites. The ingestion of a single cyst is all that is required for human infection. Undercooking or insufficient freezing of meat is an important source of infection in the developed world. In the United States, lamb products and pork products may show evidence of cysts that contain bradyzoites, but the overall prevalence of *T. gondii* has been gradually decreasing. The incidence in beef is much lower—perhaps as low as 1%. Direct ingestion of bradyzoite cysts in these various meat products leads to acute infection.

TRANSMISSION VIA BLOOD OR ORGANS

In addition to being transmitted orally, *T. gondii* can be transmitted directly from a seropositive donor to a seronegative recipient in a transplanted heart, heart-lung, kidney, liver, or pancreas. Viable parasites can be cultured from refrigerated anticoagulated blood, which may be a source of infection in individuals receiving blood transfusions. *T. gondii* reactivation has been reported in bone marrow, hematopoietic stem cell, and liver transplant recipients as well as in individuals with AIDS. Although antibody titers generally are not useful in monitoring *T. gondii* infection, individuals with higher antibody titers reportedly may be at relatively high risk for reactivation after hematopoietic stem cell transplantation; thus routine polymerase chain reaction (PCR) screening of blood from these patients may be in order. Finally, laboratory personnel can be infected after contact with contaminated needles or glassware or with infected tissue.

TRANSPLACENTAL TRANSMISSION

On average, about one-third of all women who acquire infection with *T. gondii* during pregnancy transmit the parasite to the fetus; the remainder give birth to normal, uninfected babies. Of the various factors that influence fetal outcome, gestational age at the time of infection is the most critical (see below). Few data support a role for recrudescent maternal infection as the source of congenital disease, although rare cases of transmission by immunocompromised women (e.g., those infected with HIV or those receiving high-dose glucocorticoids) have been reported. Thus, women who are seropositive before pregnancy usually are protected against acute infection and do not give birth to congenitally infected neonates.

The following general guidelines can be used to evaluate congenital infection. There is essentially no risk if the mother becomes infected ≥6 months before conception. If infection is acquired <6 months before conception, the likelihood of transplacental infection increases as the interval between infection and conception decreases. In pregnancy, if the mother becomes infected during the first trimester, the incidence of transplacental infection is lowest (~15%), but the disease in the neonate is most severe. If maternal infection occurs during the third trimester, the incidence of transplacental infection is greatest (65%), but the infant is usually asymptomatic at birth. Infected infants who are normal at birth may have a higher incidence of learning disabilities and chronic neurologic sequelae than uninfected children. Only a small proportion (20%) of women infected with *T. gondii* develop clinical signs of infection. Often the diagnosis is first appreciated when routine postconception serologic tests show evidence of specific antibody.

Pathogenesis

Upon the host's ingestion of either tissue cysts containing bradyzoites or oocysts containing sporozoites, the parasites are released from the cysts by a digestive process. Bradyzoites are resistant to the effect of pepsin and invade the host's gastrointestinal tract. Within enterocytes (or other gut-associated cells), the parasites undergo morphologic transformation, giving rise to invasive tachyzoites. These tachyzoites induce a parasite-specific secretory IgA response. From the gastrointestinal tract, parasites are disseminated to a variety of organs, particularly lymphatic tissue, skeletal muscle, myocardium, retina, placenta, and the CNS. At these sites, the parasite infects host cells, replicates, and invades the adjoining cells. In this fashion, the hallmarks of the infection develop: cell death and focal necrosis surrounded by an acute inflammatory response.

In the immunocompetent host, both the humoral and the cellular immune responses control infection; parasite virulence and tissue tropism may be strain specific. Tachyzoites are sequestered by a variety of immune mechanisms, including induction of parasiticidal antibody, activation of macrophages with radical intermediates, production of interferon γ (IFN-γ), and stimulation of CD8+ cytotoxic T lymphocytes. These antigen-specific lymphocytes are capable of killing both extracellular parasites and target cells infected with parasites. As tachyzoites are cleared from the acutely infected host, tissue cysts containing bradyzoites begin to appear, usually within the CNS and the retina. Studies indicate that *Toxoplasma* secretes signaling molecules into infected host cells and that these molecules modulate host gene expression, host metabolism, and host immune response.

In the immunocompromised or fetal host, the immune factors necessary to control the spread of tachyzoite infection are lacking. This altered immune state allows the persistence of tachyzoites and gives rise to progressive focal destruction that results in organ failure (i.e., necrotizing encephalitis, pneumonia, and myocarditis).

Persistence of infection with cysts containing bradyzoites is common in the immunocompetent host. This lifelong infection usually remains subclinical. Although bradyzoites are in a slow metabolic phase,

cysts do degenerate and rupture within the CNS. This degenerative process, with the development of new bradyzoite-containing cysts, is the most probable source of recrudescent infection in immunocompromised individuals and the most likely stimulus for the persistence of antibody titers in the immunocompetent host. Although the concept is controversial, the persistence of toxoplasmosis has been hypothesized to be a contributing factor to a variety of neuropsychiatric conditions, including schizophrenia and bipolar disease. In rodents, infection clearly has significant effects on behavior, increasing predation.

Pathology

Cell death and focal necrosis due to replicating tachyzoites induce an intense mononuclear inflammatory response in any tissue or cell type infected. Tachyzoites rarely can be visualized by routine histopathologic staining of these inflammatory lesions. However, immunofluorescent staining with parasitic antigen–specific antibodies can reveal either the organism itself or evidence of antigen. In contrast to this inflammatory process caused by tachyzoites, bradyzoite-containing cysts cause inflammation only at the early stages of development, and even this inflammation may be a response to the presence of tachyzoite antigens. Once the cysts reach maturity, the inflammatory process can no longer be detected, and the cysts remain immunologically quiescent within the brain matrix until they rupture.

LYMPH NODES

During acute infection, lymph node biopsy demonstrates characteristic findings, including follicular hyperplasia and irregular clusters of tissue macrophages with eosinophilic cytoplasm. Granulomas rarely are evident in these specimens. Although tachyzoites are not usually visible, they can be sought either by subinoculation of infected tissue into mice, with resultant disease, or by PCR. PCR amplification of DNA fragments of *Toxoplasma* genes is effective and sensitive in establishing lymph node infection by tachyzoites.

EYES

In the eye, infiltrates of monocytes, lymphocytes, and plasma cells may produce uni- or multifocal lesions. Granulomatous lesions and chorioretinitis can be observed in the posterior chamber after acute necrotizing retinitis. Other ocular complications include iridocyclitis, cataracts, and glaucoma.

CENTRAL NERVOUS SYSTEM

During CNS involvement, both focal and diffuse meningoencephalitis can be documented, with evidence of necrosis and microglial nodules. Necrotizing encephalitis in patients without AIDS is characterized by small diffuse lesions with perivascular cuffing in contiguous areas. In the AIDS population, polymorphonuclear leukocytes may be present in addition to monocytes, lymphocytes, and plasma cells. Cysts containing bradyzoites frequently are found contiguous with the necrotic tissue border. As stated previously, it is estimated that there are as many as 2100 cases of TE in the United States each year.

LUNGS AND HEART

Among patients with AIDS who die of toxoplasmosis, 40–70% have involvement of the lungs and heart. Interstitial pneumonitis can develop in neonates and immunocompromised patients. Thickened and edematous alveolar septa infiltrated with mononuclear and plasma cells are apparent. This inflammation may extend to the endothelial walls. Tachyzoites and bradyzoite-containing cysts have been observed within the alveolar membrane. Superimposed bronchopneumonia can be caused by other microbial agents. Cysts and aggregates of parasites in cardiac muscle tissue are evident in patients with AIDS who die of toxoplasmosis. Focal necrosis surrounded by inflammatory cells is associated with hyaline necrosis and disrupted myocardial cells. Pericarditis is associated with toxoplasmosis in some patients.

GASTROINTESTINAL TRACT

Rare cases of human gastrointestinal tract infection with *T. gondii* have presented as ulcerations in the mucosa. Acute infection in certain strains of inbred mice (C57BL/6) results in lethal ileitis within 7–9 days. This inflammatory bowel disease has been recognized in several other mammalian species, including pigs and nonhuman primates. Although the association between human inflammatory bowel disease and either acute or recurrent *Toxoplasma* infection has not been established, studies have demonstrated recognition of the infection by human intestinal epithelial cells, as evidenced by mitogen-activated protein kinase phosphorylation, nuclear factor κB translocation, and interleukin 8 (IL-8) secretion.

OTHER SITES

Pathologic changes during disseminated infection are similar to those described for the lymph nodes, eyes, and CNS. In patients with AIDS, the skeletal muscle, pancreas, stomach, and kidneys can be involved, with necrosis, invasion by inflammatory cells, and (rarely) tachyzoites detectable by routine staining. Large necrotic lesions may cause direct tissue destruction. In addition, secondary effects from acute infection of these various organs, including pancreatitis, myositis, and glomerulonephritis, have

been reported.

Host Immune Response

Acute *Toxoplasma* infection evokes a cascade of protective immune responses in the immunocompetent host. *Toxoplasma* enters the host at the gut mucosal level and evokes a mucosal immune response that includes the production of antigen-specific secretory IgA. Titers of serum IgA antibody directed at p30 (SAG-1) are a useful marker for congenital and acute toxoplasmosis. Milk-whey IgA from acutely infected mothers contains a high titer of antibody to *T. gondii* and can block infection of enterocytes in vitro. In mice, IgA intestinal secretions directed at the parasite are abundant and are associated with the induction of mucosal T cells.

Within the host, *T. gondii* rapidly induces detectable levels of both IgM and IgG serum antibodies. Monoclonal gammopathy of the IgG class can occur in congenitally infected infants. IgM levels may be increased in newborns with congenital infection. The polyclonal IgG antibodies evoked by infection are parasiticidal in vitro in the presence of serum complement and are the basis for the Sabin-Feldman dye test. However, cell-mediated immunity is the major protective response evoked by the parasite during host infection. Macrophages are activated after phagocytosis of antibody-opsonized parasites. This activation can lead to death of the parasite by either an oxygen-dependent or an oxygen-independent process. If the parasite is not phagocytosed and enters the macrophage by active penetration, it continues to replicate, and this replication may represent the mechanism for transport and dissemination to distant organs. *Toxoplasma* stimulates a robust IL-12 response by human dendritic cells. The requirement for costimulation via CD40/154 has been established. The CD4+ and CD8+ T cell responses are antigen-specific and further stimulate the production of a variety of important lymphokines that expand the T cell and natural killer cell repertoire. *T. gondii* is a potent inducer of a T_H1 phenotype, with IL-12 and IFN-γ playing an essential role in the control of the parasites' growth in the host. Regulation of the inflammatory response is at least partially under the control of a T_H2 response that includes the production of IL-4 and IL-10 in seropositive individuals. Both asymptomatic patients and those with active infection may have a depressed CD4+-to-CD8+ ratio. This shift may be correlated with a disease syndrome but is not necessarily correlated with disease outcome. Human T cell clones of both the CD4+ and the CD8+ phenotypes are cytolytic against parasite-infected macrophages. These T cell clones produce cytokines that are "microbistatic." IL-18, IL-7, and IL-15 upregulate the production of IFN-γ and may be important during acute and chronic infection. The effect of IFN-γ may be paradoxical, with stimulation of a host downregulatory response as well.

Although in patients with AIDS *T. gondii* infection is believed to be recrudescent, determination of antibody titers generally is not helpful in establishing reactivation. Because of the severe depletion in CD4+ T cells, quite frequently there is no observed increase in antibody titer during exacerbation of infection. T cells from AIDS patients with reactivation of toxoplasmosis fail to secrete both IFN-γ and IL-2. This alteration in the production of these critical immune cytokines contributes to the persistence of infection. *Toxoplasma* infection frequently develops late in the course of AIDS, when the loss of T cell –dependent protective mechanisms, particularly CD8+ T cells, becomes most pronounced.

Clinical Manifestations

In persons whose immune systems are intact, acute toxoplasmosis is usually asymptomatic and self-limited. This condition can go unrecognized in 80–90% of adults and children with acquired infection. The asymptomatic nature of this infection makes diagnosis difficult in mothers infected during pregnancy. In contrast, the wide range of clinical manifestations in congenitally infected children includes severe neurologic complications such as hydrocephalus, microcephaly, mental retardation, and chorioretinitis. If prenatal infection is severe, multiorgan failure and subsequent intrauterine fetal death can occur. In children and adults, chronic infection can persist throughout life, with little consequence to the immunocompetent host.

TOXOPLASMOSIS IN IMMUNOCOMPETENT PATIENTS

The most common manifestation of acute toxoplasmosis is cervical lymphadenopathy. The nodes may be single or multiple, are usually nontender, are discrete, and vary in firmness. Lymphadenopathy also may be found in suboccipital, supraclavicular, inguinal, and mediastinal areas. Generalized lymphadenopathy occurs in 20–30% of symptomatic patients. Between 20 and 40% of patients with lymphadenopathy also have headache, malaise, fatigue, and fever [usually with a temperature of <40°C (<104°F)]. A smaller proportion of symptomatic individuals have myalgia, sore throat, abdominal pain, maculopapular rash, meningoencephalitis, and confusion. Rare complications associated with infection in the normal immune host include pneumonia, myocarditis, encephalopathy, pericarditis, and polymyositis. Signs and symptoms associated with acute infection usually resolve within several weeks, although the lymphadenopathy may persist for some months. In one epidemic, toxoplasmosis was diagnosed correctly in only 3 of the 25 patients who consulted physicians. If toxoplasmosis is considered in the differential diagnosis, routine laboratory and serologic screening should precede node biopsy. It is now appreciated that genotypes of *T. gondii* prevalent in South America may be more virulent than those typically seen in North America or Europe. These genotypes may be associated with acute or recurrent ocular disease in immunocompetent individuals. Thus a detailed history is

critical for establishing a diagnosis.

The results of routine laboratory studies are usually unremarkable except for minimal lymphocytosis, an elevated erythrocyte sedimentation rate, and a nominal increase in serum aminotransferase levels. Evaluation of cerebrospinal fluid (CSF) in cases with evidence of encephalopathy or meningoencephalitis shows an elevation of intracranial pressure, mononuclear pleocytosis (10–50 cells/mL), a slight increase in protein concentration, and (occasionally) an increase in the gamma globulin level. PCR amplification of the *Toxoplasma* DNA target sequence in CSF may be beneficial. The CSF of chronically infected individuals is normal.

INFECTION OF IMMUNOCOMPROMISED PATIENTS

Patients with AIDS and those receiving immunosuppressive therapy for lymphoproliferative disorders are at greatest risk for developing acute toxoplasmosis. This predilection may be due either to reactivation of latent infection or to acquisition of parasites from exogenous sources such as blood or transplanted organs. In individuals with AIDS, >95% of cases of TE are believed to be due to recrudescent infection. In most of these cases, encephalitis develops when the CD4+ T cell count falls below 100/μL. In immunocompromised hosts, the disease may be rapidly fatal if untreated. Thus, accurate diagnosis and initiation of appropriate therapy are necessary to prevent fulminant infection.

Toxoplasmosis is a principal opportunistic infection of the CNS in persons with AIDS. Although geographic origin may be related to frequency of infection, it has no correlation with the severity of disease in immunocompromised hosts. Individuals with AIDS who are seropositive for *T. gondii* are at high risk for encephalitis. Before the advent of current combination antiretroviral treatment (ART), about one-third of the 15–40% of adult AIDS patients in the United States who were latently infected with *T. gondii* developed TE. TE may still be a presenting infection in individuals who are unaware of their positive HIV status.

The signs and symptoms of acute toxoplasmosis in immunocompromised patients principally involve the CNS **(Fig. 214-2)**. More than 50% of patients with clinical manifestations have intracerebral involvement. Clinical findings at presentation range from nonfocal to focal dysfunction. CNS findings include encephalopathy, meningoencephalitis, and mass lesions. Patients may present with altered mental status (75%), fever (10–72%), seizures (33%), headaches (56%), and focal neurologic findings (60%), including motor deficits, cranial nerve palsies, movement disorders, dysmetria, visual-field loss, and aphasia. Patients who present with evidence of diffuse cortical dysfunction develop evidence of focal neurologic disease as infection progresses. This altered condition is due not only to the necrotizing encephalitis caused by direct invasion by the parasite but also to secondary effects, including vasculitis, edema, and hemorrhage. The onset of infection can range from an insidious process over several weeks to an acute confusional state with fulminant focal deficits, including hemiparesis, hemiplegia, visual-field defects, localized headache, and focal seizures.

Figure 214-2

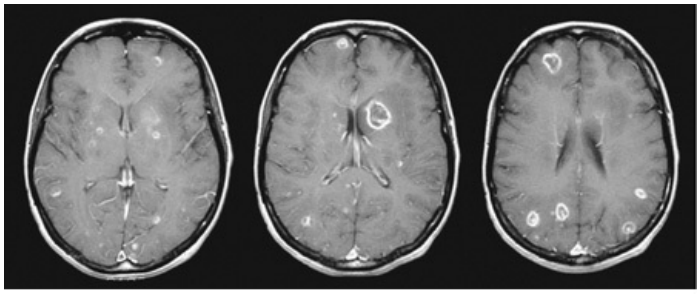

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Toxoplasmic encephalitis in a 36-year-old patient with AIDS. The multiple lesions are demonstrated by magnetic resonance scanning (T1 weighted with gadolinium enhancement). *(Courtesy of Clifford Eskey, Dartmouth Hitchcock Medical Center, Hanover, NH; with permission.)*

Although lesions can occur anywhere in the CNS, the areas most often involved appear to be the brainstem, basal ganglia, pituitary gland, and corticomedullary junction. Brainstem involvement gives rise to a variety of neurologic dysfunctions, including cranial nerve palsy, dysmetria, and ataxia. With basal ganglionic infection, patients may develop hydrocephalus, choreiform movements, and choreoathetosis. Because *Toxoplasma* usually causes encephalitis, meningeal involvement is uncommon, and thus CSF findings may be unremarkable or may include a modest increase in cell count and in protein—but not glucose—concentration.

Cerebral toxoplasmosis must be differentiated from other opportunistic infections or tumors in the CNS of AIDS patients. The differential diagnosis includes herpes simplex encephalitis, cryptococcal meningitis, progressive multifocal leukoencephalopathy, and primary CNS lymphoma. Involvement of

the pituitary gland can give rise to panhypopituitarism and hyponatremia from inappropriate secretion of vasopressin (antidiuretic hormone). HIV-associated neurocognitive disorder (HAND) may present as cognitive impairment, attention loss, and altered memory. Brain biopsy in patients who have been treated for TE but who continue to exhibit neurologic dysfunction often fails to identify organisms.

Autopsies of *Toxoplasma*-infected patients have demonstrated the involvement of multiple organs, including the lungs, gastrointestinal tract, pancreas, skin, eyes, heart, and liver. *Toxoplasma* pneumonia can be confused with *Pneumocystis* pneumonia (PcP). Respiratory involvement usually presents as dyspnea, fever, and a nonproductive cough and may rapidly progress to acute respiratory failure with hemoptysis, metabolic acidosis, hypotension, and (occasionally) disseminated intravascular coagulation. Histopathologic studies demonstrate necrosis and a mixed cellular infiltrate. The presence of organisms is a helpful diagnostic indicator, but organisms can also be found in healthy tissue. Infection of the heart is usually asymptomatic but can be associated with cardiac tamponade or biventricular failure. Infections of the gastrointestinal tract and the liver have been documented.

CONGENITAL TOXOPLASMOSIS

Between 400 and 4000 infants born each year in the United States are affected by congenital toxoplasmosis. Acute infection in mothers acquiring *T. gondii* during pregnancy is usually asymptomatic; most such women are diagnosed via prenatal serologic screening. Infection of the placenta leads to hematogenous infection of the fetus. As gestation proceeds, the proportion of fetuses that become infected increases but the clinical severity of the infection declines. Although infected children may initially be asymptomatic, the persistence of *T. gondii* can result in reactivation and clinical disease—most frequently chorioretinitis—decades later. Factors associated with relatively severe disabilities include delays in diagnosis and in initiation of therapy, neonatalhypoxia and hypoglycemia, profound visual impairment (see "Ocular Infection," below), uncorrected hydrocephalus, and increased intracranial pressure. If treated appropriately, upwards of 70% of children have normal developmental, neurologic, and ophthalmologic findings at follow-up evaluations. Treatment for 1 year with pyrimethamine, a sulfonamide, and folinic acid is tolerated with minimal toxicity (see "Treatment," below).

OCULAR INFECTION

Infection with *T. gondii* is estimated to cause 35% of all cases of chorioretinitis in the United States and Europe. Most ocular involvement is believed to be due to congenital infection, but acquired infection can be associated with outbreaks of ocular disease even in immunocompetent individuals (as seen in Victoria, British Columbia, and in South America). A variety of ocular manifestations are documented, including blurred vision, scotoma, photophobia, and eye pain. Macular involvement occurs, with loss of central vision, and nystagmus is secondary to poor fixation. Involvement of the extraocular muscles may lead to disorders of convergence and to strabismus. Ophthalmologic examination should be undertaken in newborns with suspected congenital infection. As the inflammation resolves, vision improves, but episodic flare-ups of chorioretinitis, which progressively destroy retinal tissue and lead to glaucoma, are common. The ophthalmologic examination reveals yellow-white, cotton-like patches with indistinct margins of hyperemia. As the lesions age, white plaques with distinct borders and black spots within the retinal pigment become more apparent. Lesions usually are located near the posterior pole of the retina; they may be single but are more commonly multiple. Congenital lesions may be unilateral or bilateral and show evidence of massive chorioretinal degeneration with extensive fibrosis. Surrounding these areas of involvement are a normal retina and vasculature. In patients with AIDS, retinal lesions are often large, with diffuse retinal necrosis, and include both free tachyzoites and cysts containing bradyzoites. Toxoplasmic chorioretinitis may be a prodrome to the development of encephalitis.

Diagnosis

TISSUE AND BODY FLUIDS

The differential diagnosis of acute toxoplasmosis can be made by appropriate culture, serologic testing, and PCR **(Table 214-1)**. Although difficult and available only at specialized laboratories, the isolation of *T. gondii* from blood or other body fluids can be accomplished after subinoculation of the sample into the peritoneal cavity of mice. If no parasites are found in the mouse's peritoneal fluid 6–10 days after inoculation, its anti-*Toxoplasma* serum titer can be evaluated 4–6 weeks after inoculation. Isolation of *T. gondii* from the patient's body fluids reflects acute infection, whereas isolation from biopsied tissue is an indication only of the presence of tissue cysts and should not be misinterpreted as evidence of acute toxoplasmosis. Persistent parasitemia in patients with latent, asymptomatic infection is rare. Histologic examination of lymph nodes may suggest the characteristic changes described above. Demonstration of tachyzoites in lymph nodes establishes the diagnosis of acute toxoplasmosis. Like subinoculation into mice, histologic demonstration of cysts containing bradyzoites confirms prior infection with *T. gondii* but is nondiagnostic for acute infection.

Table 214-1 Differential Laboratory Diagnosis of Toxoplasmosis

Clinical Setting	Alternative Diagnosis	Distinguishing Characteristics	
Mononucleosis syndrome	Epstein-Barr virus	Serology	
	Cytomegalovirus	Serology/PCR or culture	
	HIV	Serology/viral load	
	Bartonella (cat-scratch disease)	Biopsy (PCR or culture)/serology	
	Lymphoma	Biopsy	
Congenital infection	Cytomegalovirus	Viral culture/PCR	
	Herpes simplex virus	Viral culture/PCR	
	Rubella virus	Viral culture/serology	
	Syphilis	Serology	
Chorioretinitis in immunocompetent individual	Listeriosis	Bacterial culture	
	Tuberculosis	Bacterial culture	
	Syphilis	Serology	
	Histoplasmosis	Serology/culture	
Chorioretinitis in AIDS patient	Cytomegalovirus	Viral culture/PCR	
	Syphilis	Serology	
	Herpes simplex virus	Viral culture/PCR	
	Varicella-zoster virus	Viral culture/PCR	
	Fungal infection	Culture	
CNS lesions in AIDS patient	Lymphoma or metastatic tumor	Tissue biopsy	
		Brain abscess	Bacterial culture
		Progressive multifocal leukoencephalopathy	PCR for JC Virus
		Fungal infection	Biopsy and culture
	Mycobacterial infection	Biopsy and culture	

Source: Adapted from JD Schwartzman: Toxoplasmosis, *in Principles and Practice of Clinical Parasitology*. Hoboken, Wiley, 2001.

SEROLOGY

The procedures mentioned above have great diagnostic value but are limited by difficulties encountered either in the growth of parasites in vivo or in the identification of tachyzoites by histochemical methods. Serologic testing has become the routine method of diagnosis.

Diagnosis of acute infection with *T. gondii* can be established by detection of the simultaneous presence of IgG and IgM antibodies to *Toxoplasma* in serum. The presence of circulating IgA favors the diagnosis of an acute infection. The Sabin-Feldman dye test, the indirect fluorescent antibody test, and the enzyme-linked immunosorbent assay (ELISA) all satisfactorily measure circulating IgG antibody to *Toxoplasma*. Positive IgG titers (>1:10) can be detected as early as 2–3 weeks after infection. These titers usually peak at 6–8 weeks and decline slowly to a new baseline level that persists for life. Antibody avidity increases with time and can be useful in difficult cases during pregnancy for establishing when infection may have occurred. The serum IgM titer should be measured in concert with the IgG titer to better establish the time of infection; either the double-sandwich IgM-ELISA or the IgM-immunosorbent assay (IgM-ISAGA) should be used. Both assays are specific and sensitive, with fewer false-positive results than other commercial tests. The double-sandwich IgA-ELISA is more sensitive than the IgM-ELISA for detecting congenital infection in the fetus and newborn. Although a negative IgM result with a positive IgG titer indicates distant infection, IgM can persist for >1 year and should not necessarily be considered a reflection of acute disease. If acute toxoplasmosis is suspected, a more extensive panel of serologic tests can be performed at the *Toxoplasma* reference laboratory at the Palo Alto Medical Foundation (http://www.pamf.org/serology/clinicianguide.html).

MOLECULAR DIAGNOSTICS

Molecular approaches can directly detect *T. gondii* in biologic samples independent of the serologic

response. Results obtained with PCR have suggested high sensitivity, specificity, and clinical utility in the diagnosis of TE in resource-poor settings. Real-time PCR is a promising technique that can provide quantitative results. Isolates can be genotyped and polymorphic sequences can be obtained, with consequent identification of the precise strain. Molecular epidemiologic studies with polymorphic markers have been useful in correlating clinical signs and symptoms of disease with different *T. gondii* genotypes.

THE IMMUNOCOMPETENT ADULT OR CHILD

For the patient who presents with lymphadenopathy only, a positive IgM titer is an indication of acute infection—and an indication for therapy, if clinically warranted (see "Treatment," below). The serum IgM titer should be determined again in 3 weeks. An elevation in the IgG titer without an increase in the IgM titer suggests that infection is present but is not acute. If there is a borderline increase in either IgG or IgM, the titers should be reassessed in 3–4 weeks.

THE IMMUNOCOMPROMISED HOST

A presumptive clinical diagnosis of TE in patients with AIDS is based on clinical presentation, history of exposure (as evidenced by positive serology), and radiologic evaluation. To detect latent infection with *T. gondii*, HIV-infected persons should be tested for IgG antibody to *Toxoplasma* soon after HIV infection is diagnosed. When these criteria are used, the predictive value is as high as 80%. More than 97% of patients with AIDS and toxoplasmosis have IgG antibody to *T. gondii* in serum. IgM serum antibody usually is not detectable. Although IgG titers do not correlate with active infection, serologic evidence of infection virtually always precedes the development of TE. It is therefore important to determine the *Toxoplasma* antibody status of all patients infected with HIV. Antibody titers may range from negative to 1:1024 in patients with AIDS and TE. Fewer than 3% of patients have no demonstrable antibody to *Toxoplasma* at diagnosis of TE.

Patients with TE have focal or multifocal abnormalities demonstrable by CT or MRI. Neuroradiologic evaluation should include double-dose contrast CT of the head. By this test, single and frequently multiple contrast-enhancing lesions (<2 cm) may be identified. MRI usually demonstrates multiple lesions located in both hemispheres, with the basal ganglia and corticomedullary junction most commonly involved; MRI provides a more sensitive evaluation of the efficacy of therapy than does CT (Fig. 214-2). These findings are not pathognomonic of *Toxoplasma* infection, since 40% of CNS lymphomas are multifocal and 50% are ring-enhancing. For both MRI and CT scans, the rate of false-negative results is ~10%. The finding of a single lesion on an MRI scan increases the likelihood of primary CNS lymphoma (in which solitary lesions are four times more likely than in TE) and strengthens the argument for the performance of a brain biopsy. A therapeutic trial of anti-*Toxoplasma* medications is frequently used to assess the diagnosis. Treatment of presumptive TE with pyrimethamine plus sulfadiazine or clindamycin results in quantifiable clinical improvement in >50% of patients by day 3. By day 7, >90% of treated patients show evidence of improvement. In contrast, if patients fail to respond or have lymphoma, clinical signs and symptoms worsen by day 7. Patients in this category require brain biopsy with or without a change in therapy. This procedure can now be performed by a stereotactic CT-guided method that reduces the potential for complications. Brain biopsy for *T. gondii* identifies organisms in 50–75% of cases. PCR amplification of CSF may also confirm toxoplasmosis or suggest alternative diagnoses, such as progressive multifocal leukoencephalopathy (JC virus positive) or primary CNS lymphoma (Epstein-Barr virus positive).

Both positron emission tomography (PET) and single-photon emission CT (SPECT) have been touted as means of detecting or ruling out *Toxoplasma* infection when a CNS lesion is suspected. However, CT and MRI are currently the standard diagnostic imaging tests for TE. As in other conditions, the radiologic response may lag behind the clinical response. Resolution of lesions may take from 3 weeks to 6 months. Some patients show clinical improvement despite worsening radiographic findings.

CONGENITAL INFECTION

The issue of concern when a pregnant woman has evidence of recent *T. gondii* infection is whether the fetus is infected. PCR analysis of the amniotic fluid for the B1 gene of *T. gondii* has replaced fetal blood sampling. Serologic diagnosis is based on the persistence of IgG antibody or a positive IgM titer after the first week of life (a time frame that excludes placental leak). The IgG determination should be repeated every 2 months. An increase in IgM beyond the first week of life is indicative of acute infection. However, up to 25% of infected newborns may be seronegative and have normal routine physical examinations. Thus assessment of the eye and the brain, with ophthalmologic testing, CSF evaluation, and radiologic studies, is important in establishing the diagnosis.

OCULAR TOXOPLASMOSIS

The serum antibody titer may not correlate with the presence of active lesions in the fundus, particularly in cases of congenital toxoplasmosis. In general, a positive IgG titer (measured in undiluted serum if necessary) in conjunction with typical lesions establishes the diagnosis. Antibody production in ocular fluids, expressed in terms of the Goldmann-Witmer coefficient, can also be used for diagnosis of ocular disease. Confirmation of local specific antibody production in the eye indicates that the site of

inflammatory activity is localized to this organ. However, two-thirds of patients without evidence of specific antibody production at initial clinical presentation later develop a detectable titer. If lesions are atypical and the titer is in the low-positive range, the diagnosis is presumptive. The parasitic antigen–specific polyclonal IgG assay as well as parasitic antigen–specific PCR may facilitate the diagnosis. Accordingly, the clinical diagnosis of ocular toxoplasmosis can be supported in 60–90% of cases by laboratory tests, depending on the time of anterior chamber puncture and the panel of antibody analyses used. In the remaining cases, the possibility of a falsely negative laboratory diagnosis or of an incorrect clinical diagnosis cannot be clarified further.

Treatment: Toxoplasmosis

CONGENITAL INFECTION

Congenitally infected neonates are treated with daily oral pyrimethamine (1 mg/kg) and sulfadiazine (100 mg/kg) with folinic acid for 1 year. Depending on the signs and symptoms, prednisone (1 mg/kg per day) may be used for congenital infection. Some U.S. states and some countries routinely screen pregnant women (France, Austria) and/or newborns (Denmark, Massachusetts). Management and treatment regimens vary with the country and the treatment center. Most experts use spiramycin to treat pregnant women who have acute toxoplasmosis early in pregnancy and use pyrimethamine/sulfadiazine/folinic acid to treat women who seroconvert after 18 weeks of pregnancy or in cases of documented fetal infection. This treatment is somewhat controversial: clinical studies, which have included few untreated women, have not proven the efficacy of such therapy in preventing congenital toxoplasmosis. However, studies do suggest that treatment during pregnancy decreases the severity of infection. Many women who are infected in the first trimester elect termination of pregnancy. Those who do not terminate pregnancy are offered prenatal antibiotic therapy to reduce the frequency and severity of *Toxoplasma* infection in the infant.

INFECTION IN IMMUNOCOMPETENT PATIENTS

Immunologically competent adults and older children who have only lymphadenopathy do not require specific therapy unless they have persistent, severe symptoms. Patients with ocular toxoplasmosis are usually treated for 1 month with pyrimethamine plus either sulfadiazine or clindamycin and sometimes with prednisone. Treatment should be supervised by an ophthalmologist familiar with *Toxoplasma* disease. Ocular disease can be self-limited without treatment, but therapy is typically considered for lesions that are severe or close to the fovea or optic disc.

INFECTION IN IMMUNOCOMPROMISED PATIENTS

Primary Prophylaxis

Patients with AIDS should be treated for acute toxoplasmosis; in immunocompromised patients, toxoplasmosis is rapidly fatal if untreated. Before the introduction of ART, the median survival time was >1 year for patients who could tolerate treatment for TE. Despite their toxicity, the drugs used to treat TE were required for survival prior to ART. The incidence of TE has declined as the survival of patients with HIV infection has increased through the use of ART.

In Africa, many patients are diagnosed with HIV infection only after developing opportunistic infections such as TE. Hence, the optimal management of these opportunistic infections is important if the benefits of subsequent ART are to be realized. AIDS patients who are seropositive for *T. gondii* and who have a CD4+ T lymphocyte count of <100/μL should receive prophylaxis against TE.

Of the currently available agents, trimethoprim-sulfamethoxazole (TMP-SMX) appears to be an effective alternative for treatment of TE in resource-poor settings where the preferred combination of pyrimethamine plus sulfadiazine is not available. The daily dose of TMP-SMX recommended as the preferred regimen for PcP prophylaxis (one double-strength tablet) is effective against TE. If patients cannot tolerate TMP-SMX, the recommended alternative is dapsone-pyrimethamine, which is also effective against PcP. Atovaquone with or without pyrimethamine also can be considered. Prophylactic monotherapy with dapsone, pyrimethamine, azithromycin, clarithromycin, or aerosolized pentamidine is probably insufficient. AIDS patients who are seronegative for *Toxoplasma* and are not receiving prophylaxis for PcP should be retested for IgG antibody to *Toxoplasma* if their CD4+ T cell count drops to <100/μL. If seroconversion has taken place, then the patient should be given prophylaxis as described above.

Discontinuing Primary Prophylaxis

Current studies indicate that prophylaxis against TE can be discontinued in patients who have responded to ART and whose CD4+ T lymphocyte count has been >200/μL for 3 months. Although patients with CD4+ T lymphocyte counts of <100/μL are at greatest risk for developing TE, the risk that this condition will develop when the count has increased to 100–200/μL has not been established. Thus, prophylaxis should be discontinued when the count has increased to >200/μL. Discontinuation of therapy reduces the pill burden; the potential for drug toxicity, drug interaction, or selection of drug-resistant pathogens; and cost. Prophylaxis should be recommenced if the CD4+ T lymphocyte count again decreases to <100–200/μL.

Individuals who have completed initial therapy for TE should receive treatment indefinitely unless

immune reconstitution, with a CD4+ T cell count of >200/μL, occurs as a consequence of ART. Combination therapy with pyrimethamine plus sulfadiazine plus leucovorin is effective for this purpose. An alternative to sulfadiazine in this regimen is clindamycin.

Discontinuing Secondary Prophylaxis (Long-Term Maintenance Therapy)

Patients receiving secondary prophylaxis for TE are at low risk for recurrence when they have completed initial therapy for TE, remain asymptomatic, and have a CD4+ T lymphocyte count of >200/μL for at least 6 months after ART. This recommendation is based on recent observations in a large cohort (381 patients) and is consistent with more extensive data indicating the safety of discontinuing secondary prophylaxis for other opportunistic infections during advanced HIV disease. Discontinuation of long-term maintenance therapy among these patients appears reasonable. A repeat MRI brain scan is recommended. Secondary prophylaxis should be reintroduced if the CD4+ T lymphocyte count decreases to <200/μL.

Prevention

All HIV-infected persons, including those who lack IgG antibody to *Toxoplasma*, should be counseled regarding sources of *Toxoplasma* infection. The chances of primary infection with *Toxoplasma* can be reduced by not eating undercooked meat and by avoiding oocyst-contaminated material (i.e., a cat's litter box). Specifically, lamb, beef, and pork should be cooked to an internal temperature of 165°–170°F; from a more practical perspective, meat cooked until it is no longer pink inside usually satisfies this requirement. Hands should be washed thoroughly after work in the garden, and all fruits and vegetables should be washed. Ingestion of raw shellfish is a risk factor for toxoplasmosis, given that the filter-feeding mechanism of clams and mussels concentrates oocysts.

If the patient owns a cat, the litter box should be cleaned or changed daily, preferably by an HIV-negative, nonpregnant person; alternatively, patients should wash their hands thoroughly after changing the litter box. Litter boxes should be changed daily if possible, as freshly excreted oocysts will not have sporulated and will not be infectious. Patients should be encouraged to keep their cats inside and not to adopt or handle stray cats. Cats should be fed only canned or dried commercial food or well-cooked table food, not raw or undercooked meats. Patients need not be advised to part with their cats or to have their cats tested for toxoplasmosis. Blood intended for transfusion into *Toxoplasma*-seronegative immunocompromised individuals should be screened for antibody to *T. gondii*. Although such serologic screening is not routinely performed, seronegative women should be screened for evidence of infection several times during pregnancy if they are exposed to environmental conditions that put them at risk for infection with *T. gondii*. HIV-positive individuals should adhere closely to these preventive measures.

FURTHER READINGS

Dedicoat M: Management of toxoplasmic encephalitis in HIV-infected adults (with an emphasis on resource-poor settings). Cochrane Database Syst Rev 3:CD005420, 2006

Jones JL et al: *Toxoplasma gondii* infection in the United States, 1999–2004, decline from the prior decade. Am J Trop Med Hyg 77:405, 2007[PMID: 17827351]

——— et al: Risk factors for *Toxoplasma gondii* infection in the United States. Clin Infect Dis 49:878, 2009

Kaplan JE et al: Guidelines for prevention and treatment of opportunistic infections in HIV-infected adults and adolescents. Recommendations from CDC, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. MMWR Recomm Rep 58:1–207, 2009[PMID: 19834454]

Lehmann T et al: Globalization and the population structure of *Toxoplasma gondii*. Proc Natl Acad Sci USA 103:11423, 2006[PMID: 16849431]

Mcleod R et al: Outcome of treatment for congenital toxoplasmosis, 1981–2004: The National Collaborative Chicago-Based, Congenital Toxoplasmosis Study. Clin Infect Dis 42:1383, 2006[PMID: 16619149]

Meers S et al: Myeloablative conditioning predisposes patients for *Toxoplasma gondii* reactivation after allogeneic stem cell transplantation. Clin Infect Dis 50:1127, 2010[PMID: 20218875]

Montoya JG, Remington JS: Management of *Toxoplasma gondii* infection during pregnancy. Clin Infect Dis 47:554, 2008[PMID: 18624630]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PROTOZOAL INFECTIONS

Giardiasis

Giardia intestinalis (also known as *G. lamblia* or *G. duodenalis*) is a cosmopolitan protozoal parasite that inhabits the small intestines of humans and other mammals. Giardiasis is one of the most common parasitic diseases in both developed and developing countries worldwide, causing both endemic and epidemic intestinal disease and diarrhea.

LIFE CYCLE AND EPIDEMIOLOGY

(Fig. 215-1) Infection follows the ingestion of environmentally hardy cysts, which excyst in the small intestine, releasing flagellated trophozoites **(Fig. 215-2)** that multiply by binary fission. *Giardia* remains a pathogen of the proximal small bowel and does not disseminate hematogenously. Trophozoites remain free in the lumen or attach to the mucosal epithelium by means of a ventral sucking disk. As a trophozoite encounters altered conditions, it forms a morphologically distinct cyst, which is the stage of the parasite usually found in the feces. Trophozoites may be present and even predominate in loose or watery stools, but it is the resistant cyst that survives outside the body and is responsible for transmission. Cysts do not tolerate heating, desiccation, or continued exposure to feces but do remain viable for months in cold fresh water. The number of cysts excreted varies widely but can approach 10^7 per gram of stool.

Figure 215-1

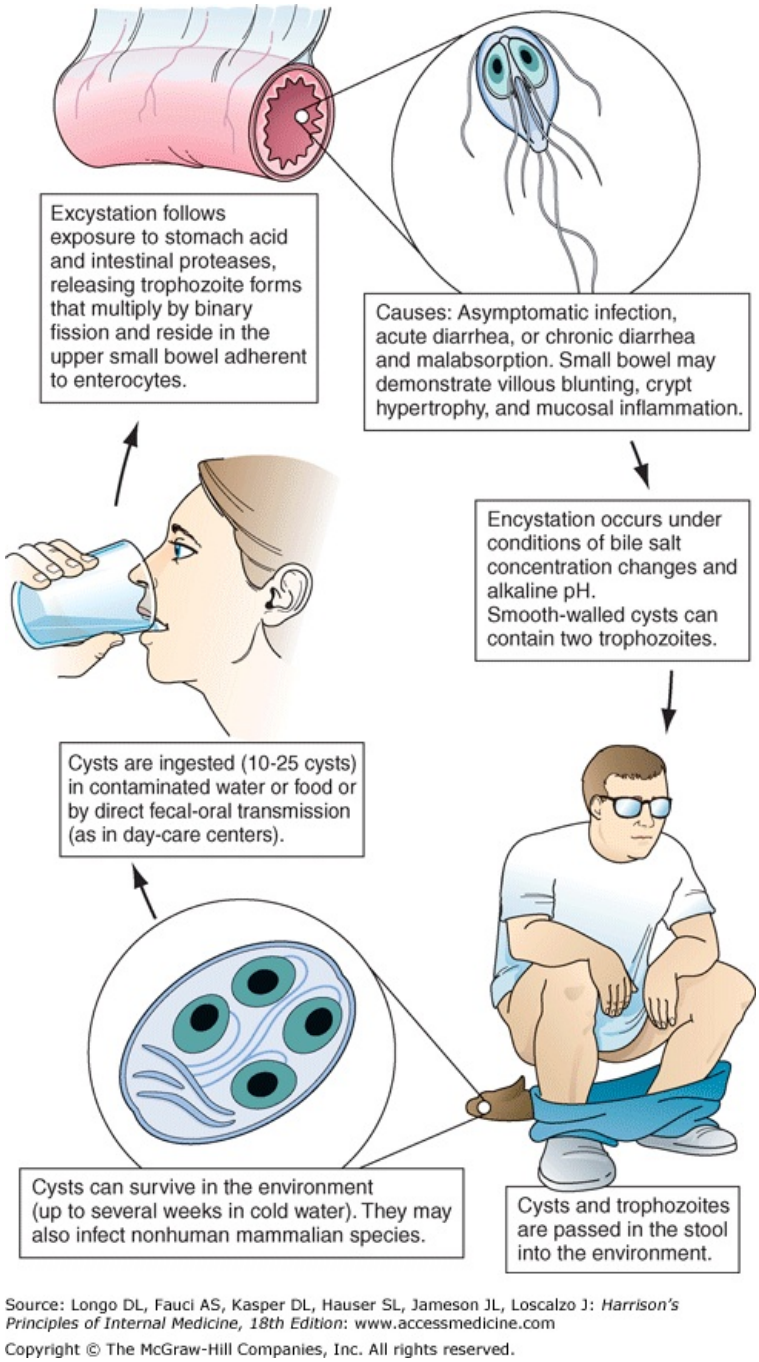

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Life cycle of *Giardia*. *(Reprinted from RL Guerrant et al: Tropical Infectious Disease: Principles, Pathogens and Practice, 2nd ed, 2006, p 987, with permission from Elsevier Science.)*

Ingestion of as few as 10 cysts is sufficient to cause infection in humans. Because cysts are infectious when excreted, person-to-person transmission occurs where fecal hygiene is poor. Giardiasis (symptomatic or asymptomatic) is especially prevalent in day-care centers; person-to-person spread also takes place in other institutional settings with poor fecal hygiene and during anal-oral contact. If food is contaminated with *Giardia* cysts after cooking or preparation, food-borne transmission can occur. Waterborne transmission accounts for episodic infections (e.g., in campers and travelers) and for major epidemics in metropolitan areas. Surface water, ranging from mountain streams to large municipal reservoirs, can become contaminated with fecally derived *Giardia* cysts; outmoded water systems are subject to cross-contamination from leaking sewer lines. The efficacy of water as a means of transmission is enhanced by the small infectious inoculum of *Giardia*, the prolonged survival of cysts in cold water, and the resistance of cysts to killing by routine chlorination methods that are adequate for controlling bacteria. Viable cysts can be eradicated from water by either boiling or filtration. In the United States, *Giardia* (like *Cryptosporidium*; see below) is a common cause of waterborne epidemics of gastroenteritis. *Giardia* is common in developing countries, and infections may be acquired by travelers.

Giardia parasites genotypically similar to those in humans are found in many mammals, including beavers from reservoirs implicated in epidemics. The importance of dogs and cats as sources of

infection for humans is unclear.

Giardiasis, like cryptosporidiosis, creates a significant economic burden because of the costs incurred in the installation of water filtration systems required to prevent waterborne epidemics, in the management of epidemics that involve large communities, and in the evaluation and treatment of endemic infections.

PATHOPHYSIOLOGY

The reasons that some, but not all, infected patients develop clinical manifestations and the mechanisms by which *Giardia* causes alterations in small-bowel function are largely unknown. Although trophozoites adhere to the epithelium, they do not cause invasive or locally destructive alterations. The lactose intolerance and, in a minority of infected adults and children, significant malabsorption that develop are clinical signs of the loss of brush-border enzyme activities. In most infections, the morphology of the bowel is unaltered; however, in a few cases (usually in chronically infected, symptomatic patients), the histopathologic findings (including flattened villi) and the clinical manifestations resemble those of tropical sprue and gluten-sensitive enteropathy. The pathogenesis of diarrhea in giardiasis is not known.

The natural history of *Giardia* infection varies markedly. Infections may be aborted, transient, recurrent, or chronic. Parasite as well as host factors may be important in determining the course of infection and disease. Both cellular and humoral responses develop in human infections, but their precise roles in the control of infection and/or disease are unknown. Because patients with hypogammaglobulinemia suffer from prolonged, severe infections that are poorly responsive to treatment, humoral immune responses appear to be important. The greater susceptibility of the young than of the old and of newly exposed persons than of chronically exposed populations suggests that at least partial protective immunity may develop. *Giardia* isolates vary genotypically, biochemically, and biologically, and variations among isolates may contribute to different courses of infection.

CLINICAL MANIFESTATIONS

Disease manifestations of giardiasis range from asymptomatic carriage to fulminant diarrhea and malabsorption. Most infected persons are asymptomatic, but in epidemics the proportion of symptomatic cases may be higher. Symptoms may develop suddenly or gradually. In persons with acute giardiasis, symptoms develop after an incubation period that lasts at least 5–6 days and usually 1–3 weeks. Prominent early symptoms include diarrhea, abdominal pain, bloating, belching, flatus, nausea, and vomiting. Although diarrhea is common, upper intestinal manifestations such as nausea, vomiting, bloating, and abdominal pain may predominate. The duration of acute giardiasis is usually >1 week, although diarrhea often subsides. Individuals with chronic giardiasis may present with or without having experienced an antecedent acute symptomatic episode. Diarrhea is not necessarily prominent, but increased flatus, loose stools, sulfurous belching, and (in some instances) weight loss occur. Symptoms may be continual or episodic and can persist for years. Some persons who have relatively mild symptoms for long periods recognize the extent of their discomfort only in retrospect. Fever, the presence of blood and/or mucus in the stools, and other signs and symptoms of colitis are uncommon and suggest a different diagnosis or a concomitant illness. Symptoms tend to be intermittent yet recurring and gradually debilitating, in contrast with the acute disabling symptoms associated with many enteric bacterial infections. Because of the less severe illness and the propensity for chronic infections, patients may seek medical advice late in the course of the illness; however, disease can be severe, resulting in malabsorption, weight loss, growth retardation, and dehydration. A number of extraintestinal manifestations have been described, such as urticaria, anterior uveitis, and arthritis; whether these are caused by giardiasis or concomitant processes is unclear.

Giardiasis can be severe in patients with hypogammaglobulinemia and can complicate other preexisting intestinal diseases, such as that occurring in cystic fibrosis. In patients with AIDS, *Giardia* can cause enteric illness that is refractory to treatment.

DIAGNOSIS

(Table 215-1) Giardiasis is diagnosed by detection of parasite antigens in the feces or by identification of cysts in the feces or of trophozoites in the feces or small intestines. Cysts are oval, measure 8–12 μm x 7–10 μm, and characteristically contain four nuclei. Trophozoites are pear-shaped, dorsally convex, flattened parasites with two nuclei and four pairs of flagella (Fig. 215-2). The diagnosis is sometimes difficult to establish. Direct examination of fresh or properly preserved stools as well as concentration methods should be used. Because cyst excretion is variable and may be undetectable at times, repeated examination of stool, sampling of duodenal fluid, and biopsy of the small intestine may be required to detect the parasite. Tests for parasitic antigens in stool are at least as sensitive and specific as good microscopic examinations and are easier to perform. All of these methods occasionally yield false-negative results.

Table 215-1 Diagnosis of Intestinal Protozoal Infections

Parasite	Stool	Fecal Acid-Fast	Stool Antigen	Other

	O+P[a]	Stain	Immunoassays	
Giardia	+		+	
Cryptosporidium	–	+	+	
Isospora	–	+		
Cyclospora	–	+		
Microsporidia	–			Special fecal stains, tissue biopsies

[a]O+P, ova and parasites.

Figure 215-2

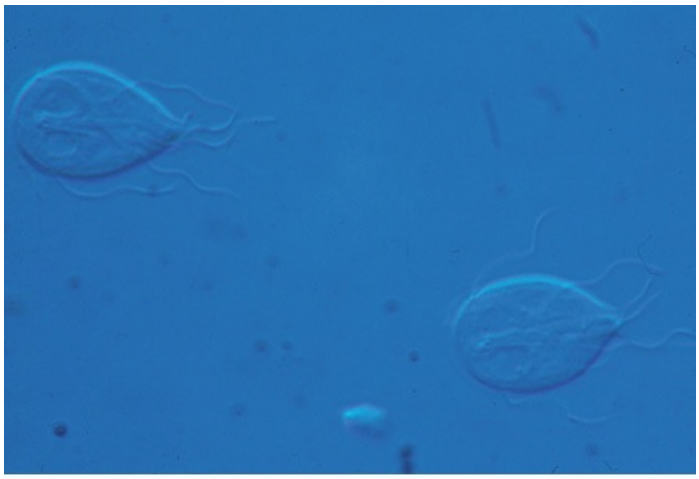

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flagellated, binucleate ***Giardia*** **trophozoites.**

TREATMENT: GIARDIASIS

Cure rates with metronidazole (250 mg thrice daily for 5 days) are usually >90%. Tinidazole (2 g once by mouth) is reportedly more effective than metronidazole. Nitazoxanide (500 mg twice daily for 3 days) is an alternative agent for treatment of giardiasis. Paromomycin, an oral aminoglycoside that is not well absorbed, can be given to symptomatic pregnant patients, although information is limited on how effectively this agent eradicates infection.

Almost all patients respond to therapy and are cured, although some with chronic giardiasis experience delayed resolution of symptoms after eradication of *Giardia*. For many of the latter patients, residual symptoms probably reflect delayed regeneration of intestinal brush-border enzymes. Continued infection should be documented by stool examinations before treatment is repeated. Patients who remain infected after repeated treatments should be evaluated for reinfection through family members, close personal contacts, and environmental sources as well as for hypogammaglobulinemia. In cases refractory to multiple treatment courses, prolonged therapy with metronidazole (750 mg thrice daily for 21 days) has been successful.

PREVENTION

Although giardiasis is extremely infectious, disease can be prevented by consumption of noncontaminated food and water and by personal hygiene when caring for infected children. Boiling or filtering potentially contaminated water prevents infection.

Cryptosporidiosis

The coccidian parasite *Cryptosporidium* causes diarrheal disease that is self-limited in immunocompetent human hosts but can be severe in persons with AIDS or other forms of immunodeficiency. Two species of *Cryptosporidium*, *C. hominis* and *C. parvum*, cause most human infections.

LIFE CYCLE AND EPIDEMIOLOGY

Cryptosporidium species are widely distributed in the world. Cryptosporidiosis is acquired by the consumption of oocysts (50% infectious dose: ~132 oocysts in nonimmune individuals), which excyst to liberate sporozoites that in turn enter and infect intestinal epithelial cells. The parasite's further development involves both asexual and sexual cycles, which produce forms capable of infecting other

epithelial cells and of generating oocysts that are passed in the feces. *Cryptosporidium* species infect a number of animals, and *C. parvum* can spread from infected animals to humans. Since oocysts are immediately infectious when passed in feces, person-to-person transmission takes place in day-care centers and among household contacts and medical providers. Waterborne transmission (especially that of *C. hominis*) accounts for infections in travelers and for common-source epidemics. Oocysts are quite hardy and resist killing by routine chlorination. Both drinking water and recreational water (e.g., pools, waterslides) have been increasingly recognized as sources of infection.

PATHOPHYSIOLOGY

Although intestinal epithelial cells harbor cryptosporidia in an intracellular vacuole, the means by which secretory diarrhea is elicited remain uncertain. No characteristic pathologic changes are found by biopsy. The distribution of infection can be spotty within the principal site of infection, the small bowel. Cryptosporidia are found in the pharynx, stomach, and large bowel of some patients and at times in the respiratory tract. Especially in patients with AIDS, involvement of the biliary tract can cause papillary stenosis, sclerosing cholangitis, or cholecystitis.

CLINICAL MANIFESTATIONS

Asymptomatic infections can occur in both immunocompetent and immunocompromised hosts. In immunocompetent persons, symptoms develop after an incubation period of ~1 week and consist principally of watery nonbloody diarrhea, sometimes in conjunction with abdominal pain, nausea, anorexia, fever, and/or weight loss. In these hosts, the illness usually subsides after 1–2 weeks. In contrast, in immunocompromised hosts (especially those with AIDS and CD4+ T cell counts <100/μL), diarrhea can be chronic, persistent, and remarkably profuse, causing clinically significant fluid and electrolyte depletion. Stool volumes may range from 1 to 25 L/d. Weight loss, wasting, and abdominal pain may be severe. Biliary tract involvement can manifest as midepigastric or right-upper-quadrant pain.

DIAGNOSIS

(Table 215-1) Evaluation starts with fecal examination for small oocysts, which are smaller (4–5 μm in diameter) than the fecal stages of most other parasites. Because conventional stool examination for ova and parasites does not detect *Cryptosporidium*, specific testing must be requested. Detection is enhanced by evaluation of stools (obtained on multiple days) by several techniques, including modified acid-fast and direct immunofluorescent stains and enzyme immunoassays. Cryptosporidia can also be identified by light and electron microscopy at the apical surfaces of intestinal epithelium from biopsy specimens of the small bowel and, less frequently, the large bowel.

TREATMENT: CRYPTOSPORIDIOSIS

Nitazoxanide is approved by the U.S. Food and Drug Administration for the treatment of cryptosporidiosis and is available in tablet form for adults (500 mg twice daily for 3 days) and as an elixir for children. To date, however, this agent has not been effective for the treatment of HIV-infected patients, in whom improved immune status due to antiretroviral therapy can lead to amelioration of cryptosporidiosis. Otherwise, treatment includes supportive care with replacement of fluids and electrolytes and administration of antidiarrheal agents. Biliary tract obstruction may require papillotomy or T-tube placement. Prevention requires minimizing exposure to infectious oocysts in human or animal feces. Use of submicron water filters may minimize acquisition of infection from drinking water.

Isosporiasis

The coccidian parasite *Isospora belli* causes human intestinal disease. Infection is acquired by the consumption of oocysts, after which the parasite invades intestinal epithelial cells and undergoes both sexual and asexual cycles of development. Oocysts excreted in stool are not immediately infectious but must undergo further maturation.

Although *I. belli* infects many animals, little is known about the epidemiology or prevalence of this parasite in humans. It appears to be most common in tropical and subtropical countries. Acute infections can begin abruptly with fever, abdominal pain, and watery nonbloody diarrhea and can last for weeks or months. In patients who have AIDS or are immunocompromised for other reasons, infections often are not self-limited but rather resemble cryptosporidiosis, with chronic, profuse watery diarrhea. Eosinophilia, which is not found in other enteric protozoan infections, may be detectable. The diagnosis (Table 215-1) is usually made by detection of the large (~25-μm) oocysts in stool by modified acid-fast staining. Oocyst excretion may be low-level and intermittent; if repeated stool examinations are unrevealing, sampling of duodenal contents by aspiration or small-bowel biopsy (often with electron-microscopic examination) may be necessary.

TREATMENT: ISOSPORIASIS

Trimethoprim-sulfamethoxazole (TMP-SMX, 160/800 mg four times daily for 10 days; and for HIV-infected patients, then three times daily for 3 weeks) is effective. For patients intolerant of sulfonamides, pyrimethamine (50–75 mg/d) can be used. Relapses can occur in persons with AIDS and necessitate maintenance therapy with TMP-SMX (160/800 mg three times per week).

Cyclosporiasis

Cyclospora cayetanensis, a cause of diarrheal illness, is globally distributed: illness due to *C. cayetanensis* has been reported in the United States, Asia, Africa, Latin America, and Europe. The epidemiology of this parasite has not yet been fully defined, but waterborne transmission and food-borne transmission by basil and imported raspberries have been recognized. The full spectrum of illness attributable to *Cyclospora* has not been delineated. Some patients may harbor the infection without symptoms, but many have diarrhea, flulike symptoms, and flatulence and belching. The illness can be self-limited, can wax and wane, or in many cases can involve prolonged diarrhea, anorexia, and upper gastrointestinal symptoms, with sustained fatigue and weight loss in some instances. Diarrheal illness may persist for >1 month. *Cyclospora* can cause enteric illness in patients infected with HIV.

The parasite is detectable in epithelial cells of small-bowel biopsy samples and elicits secretory diarrhea by unknown means. The absence of fecal blood and leukocytes indicates that disease due to *Cyclospora* is not caused by destruction of the small-bowel mucosa. The diagnosis (Table 215-1) can be made by detection of spherical 8- to 10-μm oocysts in the stool, although routine stool ova andparasite (O+P) examinations are not sufficient. Specific fecal examinations must be requested to detect the oocysts, which are variably acid-fast and are fluorescent when viewed with ultraviolet light microscopy. Cyclosporiasis should be considered in the differential diagnosis of prolonged diarrhea, with or without a history of travel by the patient to other countries.

TREATMENT: CYCLOSPORIASIS

Cyclosporiasis is treated with TMP-SMX (160/800 mg twice daily for 7 days). HIV-infected patients may experience relapses after such treatment and thus may require longer-term suppressive maintenance therapy.

Microsporidiosis

Microsporidia are obligate intracellular spore-forming protozoa that infect many animals and cause disease in humans, especially as opportunistic pathogens in AIDS. Microsporidia are members of a distinct phylum, Microspora, which contains dozens of genera and hundreds of species. The various microsporidia are differentiated by their developmental life cycles, ultrastructural features, and molecular taxonomy based on ribosomal RNA. The complex life cycles of the organisms result in the production of infectious spores **(Fig. 215-3)**. Currently, eight genera of microsporidia —*Encephalitozoon*, *Pleistophora*, *Nosema*, *Vittaforma*, *Trachipleistophora*, *Brachiola*, *Microsporidium*, and *Enterocytozoon*—are recognized as causes of human disease. Although some microsporidia are probably prevalent causes of self-limited or asymptomatic infections in immunocompetent patients, little is known about how microsporidiosis is acquired.

Figure 215-3

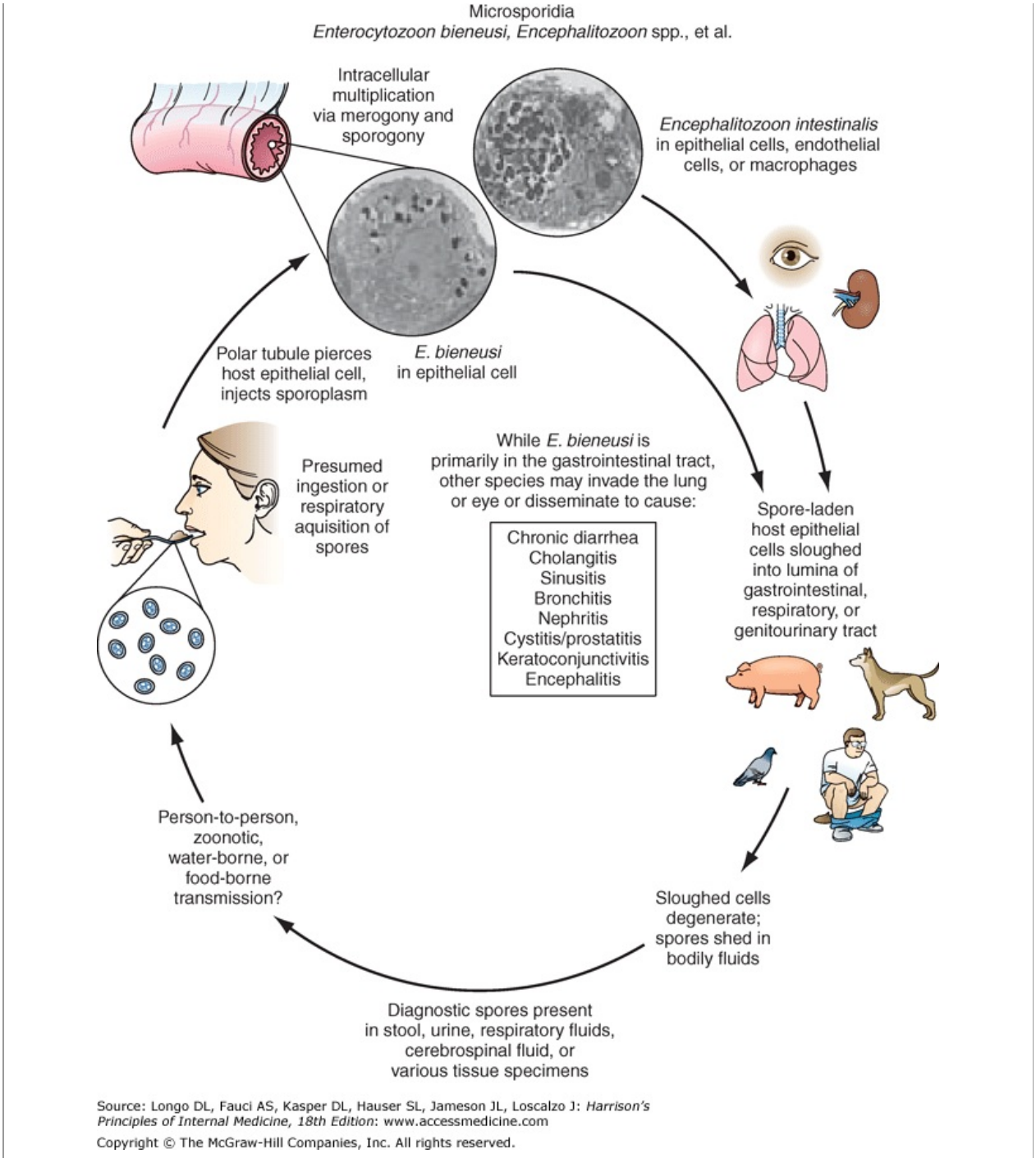

Life cycle of microsporidia. *(Reprinted from RL Guerrant et al: Tropical Infectious Disease: Principles, Pathogens and Practice, 2nd ed, 2006, p 1128, with permission from Elsevier Science.)*

Microsporidiosis is most common among patients with AIDS, less common among patients with other types of immunocompromise, and rare among immunocompetent hosts. In patients with AIDS, intestinal infections with *Enterocytozoon bieneusi* and *Encephalitozoon* (formerly *Septata*) *intestinalis* are recognized to contribute to chronic diarrhea and wasting; these infections are found in 10–40% of patients with chronic diarrhea. Both organisms have been found in the biliary tracts of patients with cholecystitis. *E. intestinalis* may also disseminate to cause fever, diarrhea, sinusitis, cholangitis, and bronchiolitis. In patients with AIDS, *Encephalitozoon hellem* has caused superficial keratoconjunctivitis as well as sinusitis, respiratory tract disease, and disseminated infection. Myositis due to *Pleistophora* has been documented. *Nosema*, *Vittaforma*, and *Microsporidium* have caused stromal keratitis associated with trauma in immunocompetent patients.

Microsporidia are small gram-positive organisms with mature spores measuring 0.5–2 μm x 1–4 μm. Diagnosis of microsporidial infections in tissue often requires electron microscopy, although intracellular spores can be visualized by light microscopy with hematoxylin and eosin, Giemsa, or tissue Gram's stain. For the diagnosis of intestinal microsporidiosis, modified trichrome or chromotrope 2R-based staining and Uvitex 2B or calcofluor fluorescent staining reveal spores in smears of feces or duodenal aspirates. Definitive therapies for microsporidial infections remain to be established. For

superficial keratoconjunctivitis due to *E. hellem*, topical therapy with fumagillin suspension has shown promise (Chap. 208). For enteric infections with *E. bieneusi* and *E. intestinalis* in HIV-infected patients, therapy with albendazole may be efficacious (Chap. 208).

Other Intestinal Protozoa

BALANTIDIASIS

Balantidium coli is a large ciliated protozoal parasite that can produce a spectrum of large-intestinal disease analogous to amebiasis. The parasite is widely distributed in the world. Since it infects pigs, cases in humans are more common where pigs are raised. Infective cysts can be transmitted from person to person and through water, but many cases are due to the ingestion of cysts derived from porcine feces in association with slaughtering, with use of pig feces for fertilizer, or with contamination of water supplies by pig feces.

Ingested cysts liberate trophozoites, which reside and replicate in the large bowel. Many patients remain asymptomatic, but some have persisting intermittent diarrhea, and a few develop more fulminant dysentery. In symptomatic individuals, the pathology in the bowel—both gross and microscopic—is similar to that seen in amebiasis, with varying degrees of mucosal invasion, focal necrosis, and ulceration. Balantidiasis, unlike amebiasis, does not spread hematogenously to other organs. The diagnosis is made by detection of the trophozoite stage in stool or sampled colonic tissue. Tetracycline (500 mg four times daily for 10 days) is an effective therapeutic agent.

BLASTOCYSTIS HOMINIS INFECTION

B. hominis, while believed by some to be a protozoan capable of causing intestinal disease, remains an organism of uncertain pathogenicity. Some patients who pass *B. hominis* in their stools are asymptomatic, whereas others have diarrhea and associated intestinal symptoms. Diligent evaluation reveals other potential bacterial, viral, or protozoal causes of diarrhea in some but not all patients with symptoms. Because the pathogenicity of *B. hominis* is uncertain and because therapy for *Blastocystis* infection is neither specific nor uniformly effective, patients with prominent intestinal symptoms should be fully evaluated for other infectious causes of diarrhea. If diarrheal symptoms associated with *Blastocystis* are prominent, either metronidazole (750 mg thrice daily for 10 days) or TMP-SMX (160 mg/800 mg twice daily for 7 days) can be used.

DIENTAMOEBA FRAGILIS INFECTION

D. fragilis is unique among intestinal protozoa in that it has a trophozoite stage but not a cyst stage. How trophozoites survive to transmit infection is not known. When symptoms develop in patients with *D. fragilis* infection, they are generally mild and include intermittent diarrhea, abdominal pain, and anorexia. The diagnosis is made by the detection of trophozoites in stool; the lability of these forms accounts for the greater yield when fecal samples are preserved immediately after collection. Since fecal excretion rates vary, examination of several samples obtained on alternate days increases the rate of detection. Iodoquinol (650 mg three times daily for 20 days), paromomycin (25–35 mg/kg per day in three doses for 7 days), metronidazole (500–750 mg three times daily for 10 days), or tetracycline (500 mg four times daily for 10 days) is appropriate for treatment.

TRICHOMONIASIS

Various species of trichomonads can be found in the mouth (in association with periodontitis) and occasionally in the gastrointestinal tract. *Trichomonas vaginalis*—one of the most prevalent protozoal parasites in the United States—is a pathogen of the genitourinary tract and a major cause of symptomatic vaginitis.

Life Cycle and Epidemiology

T. vaginalis is a pear-shaped, actively motile organism that measures about 10 x 7 μm, replicates by binary fission, and inhabits the lower genital tract of females and the urethra and prostate of males. In the United States, it accounts for ~3 million infections per year in women. While the organism can survive for a few hours in moist environments and could be acquired by direct contact, person-to-person venereal transmission accounts for virtually all cases of trichomoniasis. Its prevalence is greatest among persons with multiple sexual partners and among those with other sexually transmitted diseases (Chap. 130).

Clinical Manifestations

Many men infected with *T. vaginalis* are asymptomatic, although some develop urethritis and a few have epididymitis or prostatitis. In contrast, infection in women, which has an incubation period of 5–28 days, is usually symptomatic and manifests with malodorous vaginal discharge (often yellow), vulvar erythema and itching, dysuria or urinary frequency (in 30–50% of patients), and dyspareunia. These manifestations, however, do not clearly distinguish trichomoniasis from other types of infectious vaginitis.

Diagnosis

Detection of motile trichomonads by microscopic examination of wet mounts of vaginal or prostatic secretions has been the conventional means of diagnosis. Although this approach provides an immediate diagnosis, its sensitivity for the detection of *T. vaginalis* is only ~50–60% in routine evaluations of vaginal secretions. Direct immunofluorescent antibody staining is more sensitive (70–90%) than wet-mount examinations. *T. vaginalis* can be recovered from the urethra of both males and females and is detectable in males after prostatic massage. Culture of the parasite is the most sensitive means of detection; however, facilities for culture are not generally available, and detection of the organism takes 3–7 days.

Treatment: Trichomoniasis

Metronidazole, given either as a single 2-g dose or in 500-mg doses twice daily for 7 days, is usually effective. Tinidazole (a single 2-g dose) is also effective. All sexual partners must be treated concurrently to prevent reinfection, especially from asymptomatic males. In males with persistent symptomatic urethritis after therapy for nongonococcal urethritis, metronidazole therapy should be considered for possible trichomoniasis. Alternatives to metronidazole for treatment during pregnancy are not readily available, although use of 100-mg clotrimazole vaginal suppositories nightly for 2 weeks may cure some infections in pregnant women. Reinfection often accounts for apparent treatment failures, but strains of *T. vaginalis* exhibiting high-level resistance to metronidazole have been encountered. Treatment of these resistant infections with higher oral doses, parenteral doses, or concurrent oral and vaginal doses of metronidazole or with tinidazole has been successful.

FURTHER READINGS

Articles

Abubakar I et al: Treatment of cryptosporidiosis in immunocompromised individuals: Systematic review and meta-analysis. Br J Clin Pharmacol 63:387, 2007[PMID: 17335543]

Chalmers RM, Davies AP: Minireview: Clinical cryptosporidiosis. Exp Parasitol 124:138, 2010[PMID: 19545516]

Daly ER et al: Outbreak of giardiasis associated with a community drinking-water source. Epidemiol Infect 138:491, 2010[PMID: 19751538]

Gupta S et al: Chronic diarrhoea in HIV patients: Prevalence of coccidian parasites. Indian J Med Microbiol 26:172, 2008[PMID: 18445958]

Huppert JS: Trichomoniasis in teens: An update. Curr Opin Obstet Gynecol 21:371, 2009[PMID: 19491679]

Lanternier F et al: Microsporidiosis in solid organ transplant recipients: Two *Enterocytozoon bieneusi* cases and review. Transpl Infect Dis 11:83, 2009[PMID: 18803616]

Ortega YR, Sanchez R: Update on *Cyclospora cayetanensis*, a food-borne and waterborne parasite. Clin Microbiol Rev 23:218, 2010[PMID: 20065331]

Robertson LJ et al: Giardiasis—Why do the symptoms sometimes never stop? Trends Parasitol 26:75, 2010[PMID: 20186998]

Schuster FL, Ramirez-Avila L: Current world status of *Balantidium coli*. Clin Microbiol Rev 21:626, 2008[PMID: 18854484]

Tan KS: New insights on classification, identification, and clinical relevance of *Blastocystis* spp. Clin Microbiol Rev 21:639, 2008[PMID: 18854485]

Vandenberg O et al: Clinical and microbiological features of dientamoebiasis in patients suspected of suffering from a parasitic gastrointestinal illness: A comparison of *Dientamoeba fragilis* and *Giardia lamblia* infections. Int J Infect Dis 10:255, 2006[PMID: 16469517]

Weitzel T et al: Epidemiological and clinical features of travel-associated cryptosporidiosis. Clin Microbiol Infect 12:921, 2006[PMID: 16882300]

Yoder JS, Beach MJ: *Cryptosporidium* surveillance and risk factors in the United States. Exp Parasitol 124:31, 2010[PMID: 19786022]

———,———: Giardiasis surveillance—United States, 2003–2005. MMWR Surveill Summ 56:11, 2007

Book

Weiss LM, Schwartz DA: Microsporidiosis, in *Tropical Infectious Diseases: Principles, Pathogens and Practice*, 2nd ed, RL Guerrant et al (eds). Elsevier, Philadelphia, 2006, pp 1126–1140

Web Site

CDC Division of Parasitic Diseases. *http://www.cdc.gov/ ncidod/dpd/default.htm*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TRICHINELLOSIS AND OTHER TISSUE NEMATODE INFECTIONS: INTRODUCTION

Nematodes are elongated, symmetric roundworms. Parasitic nematodes of medical significance may be broadly classified as either predominantly intestinal or tissue nematodes. This chapter covers the tissue nematodes that cause trichinellosis, visceral and ocular larva migrans, cutaneous larva migrans, cerebral angiostrongyliasis, and gnathostomiasis. All of these zoonotic infections result from incidental exposure to infectious nematodes. The clinical symptoms of these infections are due largely to invasive larval stages that (except in the case of *Trichinella*) do not reach maturity in humans.

TRICHINELLOSIS

Trichinellosis develops after the ingestion of meat containing cysts of *Trichinella* (e.g., pork or other meat from a carnivore). Although most infections are mild and asymptomatic, heavy infections can cause severe enteritis, periorbital edema, myositis, and (infrequently) death.

LIFE CYCLE AND EPIDEMIOLOGY

Eight species of *Trichinella* are recognized as causes of infection in humans. Two species are distributed worldwide: *T. spiralis*, which is found in a great variety of carnivorous and omnivorous animals, and *T. pseudospiralis*, which is found in mammals and birds. *T. nativa* is present in Arctic regions and infects bears; *T. nelsoni* is found in equatorial eastern Africa, where it is common among felid predators and scavengers such as hyenas and bush pigs; and *T. britovi* is found in Europe, western Africa, and western Asia among carnivores but not among domestic swine. *T. murrelli* is present in North American game animals.

After human consumption of trichinous meat, encysted larvae are liberated by digestive acid and proteases **(Fig. 216-1)**. The larvae invade the small-bowel mucosa and mature into adult worms. After ~1 week, female worms release newborn larvae that migrate via the circulation to striated muscle. The larvae of all species except *T. pseudospiralis*, *T. papuae*, and *T. zimbabwensis* then encyst by inducing a radical transformation in the muscle cell architecture. Although host immune responses may help to expel intestinal adult worms, they have little effect on muscle-dwelling larvae.

Figure 216-1

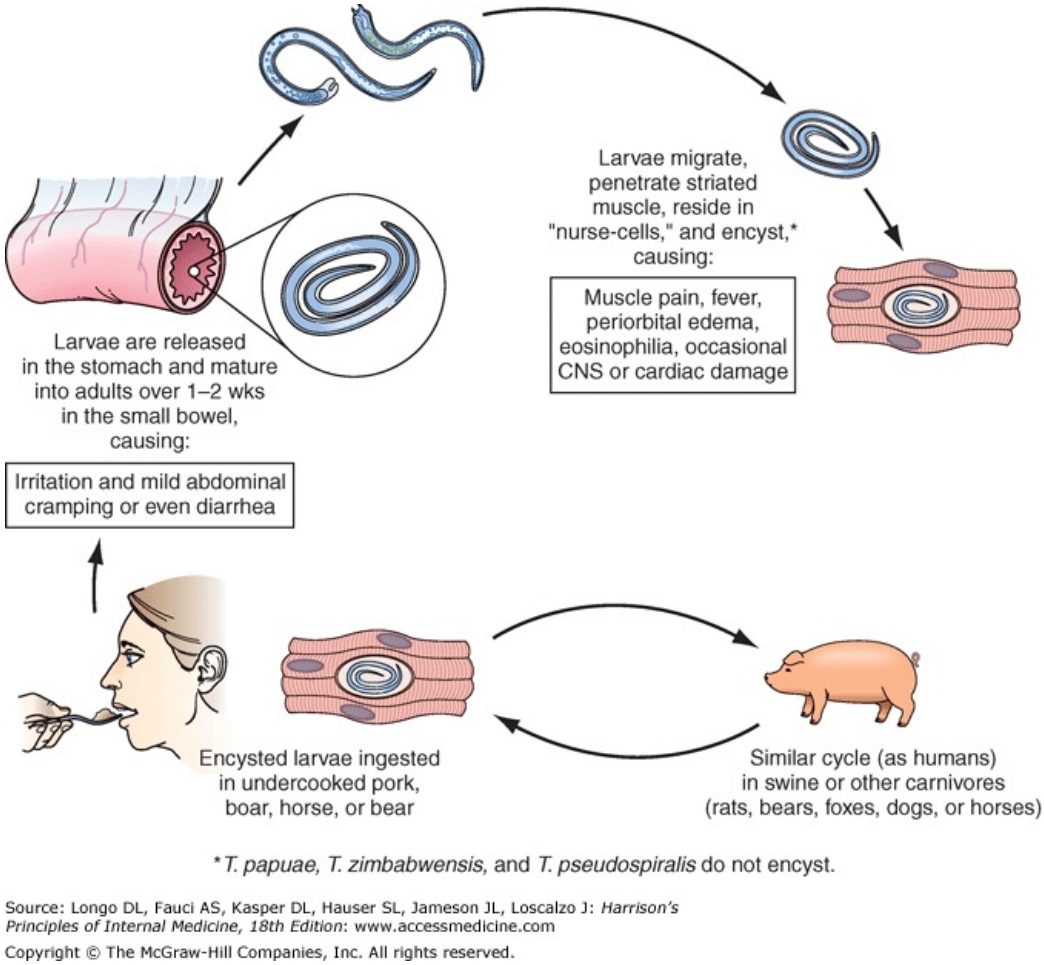

Life cycle of *Trichinella spiralis* (cosmopolitan); *nelsoni* (equatorial Africa); *britovi* (Europe, western Africa, western Asia); *nativa* (Arctic); *murrelli* (North America); *papuae* (Papua New Guinea); *zimbabwensis* (Tanzania); and *pseudospiralis* (cosmopolitan). CNS, central nervous system. *[Reprinted from Guerrant RL et al (eds): Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed, p 1218. © 2006, with permission from Elsevier Science.]*

Human trichinellosis is often caused by the ingestion of infected pork products and thus can occur in almost any location where the meat of domestic or wild swine is eaten. Human trichinellosis also may be acquired from the meat of other animals, including dogs (in parts of Asia and Africa), horses (in Italy and France), and bears and walruses (in northern regions). Although cattle (being herbivores) are not natural hosts of *Trichinella*, beef has been implicated in outbreaks when contaminated or adulterated with trichinous pork. Laws that prohibit the feeding of uncooked garbage to pigs have greatly reduced the transmission of trichinellosis in the United States. About 12 cases of trichinellosis are reported annually in this country, but most mild cases probably remain undiagnosed. Recent U.S. and Canadian outbreaks have been attributable to consumption of wild game (especially bear meat) and, less frequently, of pork.

PATHOGENESIS AND CLINICAL FEATURES

Clinical symptoms of trichinellosis arise from the successive phases of parasite enteric invasion, larval migration, and muscle encystment (Fig. 216-1). Most light infections (those with <10 larvae per gram of muscle) are asymptomatic, whereas heavy infections (which can involve >50 larvae per gram of muscle) can be life-threatening. Invasion of the gut by large numbers of parasites occasionally provokes diarrhea during the first week after infection. Abdominal pain, constipation, nausea, or vomiting also may be prominent.

Symptoms due to larval migration and muscle invasion begin to appear in the second week after infection. The migrating *Trichinella* larvae provoke a marked local and systemic hypersensitivity reaction, with fever and hypereosinophilia. Periorbital and facial edema is common, as are hemorrhages in the subconjunctivae, retina, and nail beds ("splinter" hemorrhages). A maculopapular rash, headache, cough, dyspnea, or dysphagia sometimes develops. Myocarditis with tachyarrhythmias or heart failure—and, less commonly, encephalitis or pneumonitis—may develop and accounts for most deaths of patients with trichinellosis.

Upon onset of larval encystment in muscle 2–3 weeks after infection, symptoms of myositis with myalgias, muscle edema, and weakness develop, usually overlapping with the inflammatory reactions to migrating larvae. The most commonly involved muscle groups include the extraocular muscles; the

biceps; and the muscles of the jaw, neck, lower back, and diaphragm. Peaking ~3 weeks after infection, symptoms subside only gradually during a prolonged convalescence. Uncommon infections with *T. pseudospiralis*, whose larvae do not encapsulate in muscles, elicit prolonged polymyositis-like illness.

LABORATORY FINDINGS AND DIAGNOSIS

Blood eosinophilia develops in >90% of patients with symptomatic trichinellosis and may peak at a level of >50% 2–4 weeks after infection. Serum levels of muscle enzymes, including creatine phosphokinase, are elevated in most symptomatic patients. Patients should be questioned thoroughly about their consumption of pork or wild animal meat and about illness in other individuals who ate the same meat. A presumptive clinical diagnosis can be based on fevers, eosinophilia, periorbital edema, and myalgias after a suspect meal. A rise in the titer of parasite-specific antibody, which usually does not occur until after the third week of infection, confirms the diagnosis. Alternatively, a definitive diagnosis requires surgical biopsy of at least 1 g of involved muscle; the yields are highest near tendon insertions. The fresh muscle tissue should be compressed between glass slides and examined microscopically **(Fig. 216-2)**, because larvae may be missed by examination of routine histopathologic sections alone.

Figure 216-2

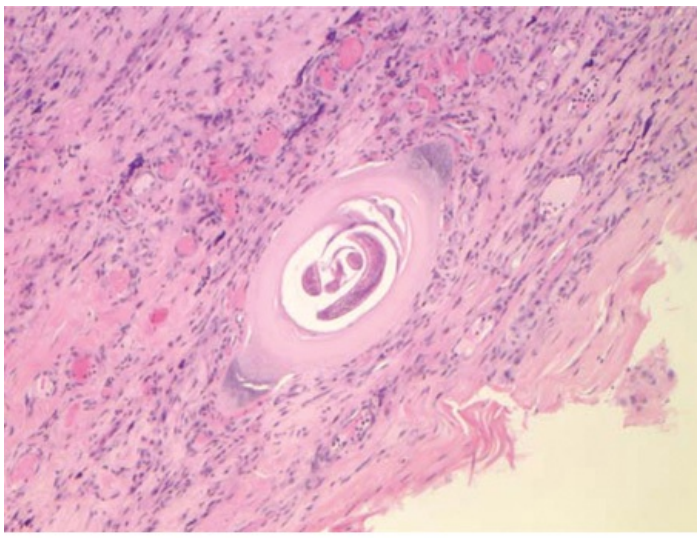

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***Trichinella* larva** encysted in a characteristic hyalinized capsule in striated muscle tissue. *(Photo/Wadsworth Center, New York State Department of Health. Reprinted from MMWR 53:606, 2004; public domain.)*

Treatment: Trichinellosis

Most lightly infected patients recover uneventfully with bed rest, antipyretics, and analgesics. Glucocorticoids like prednisone **(Table 216-1)** are beneficial for severe myositis and myocarditis. Mebendazole and albendazole are active against enteric stages of the parasite, but their efficacy against encysted larvae has not been conclusively demonstrated.

Table 216-1 Therapy for Tissue Nematode Infections

Infection	Severity	Treatment
Trichinellosis	Mild	Supportive
	Moderate	Albendazole (400 mg bid x 8–14 days) or Mebendazole (200–400 mg tid x 3 days, then 400 mg tid x 8–14 days)
	Severe	Add glucocorticoids (e.g., prednisone, 1 mg/kg qd x 5 days)
Visceral larva migrans	Mild to moderate	Supportive
	Severe	Glucocorticoids (as above)
Cutaneous larva migrans	Ocular	Not fully defined; albendazole (800 mg bid for adults, 400 mg bid for children) with glucocorticoids x 5–20 days has been effective Ivermectin (single dose, 200 μg/kg) or

migrans		Albendazole (200 mg bid x 3 days)
Angiostrongyliasis	Mild to moderate	Supportive
Gnathostomiasis	Severe	Glucocorticoids (as above) Ivermectin (200 mg/kg per day x 2 days) or Albendazole (400 mg bid x 21 days)

PREVENTION

Larvae may be killed by cooking pork until it is no longer pink or by freezing it at –15°C for 3 weeks. However, Arctic *T. nativa* larvae in walrus or bear meat are relatively resistant and may remain viable despite freezing.

VISCERAL AND OCULAR LARVA MIGRANS

Visceral larva migrans is a syndrome caused by nematodes that are normally parasitic for nonhuman host species. In humans, these nematode larvae do not develop into adult worms but instead migrate through host tissues and elicit eosinophilic inflammation. The most common form of visceral larva migrans is toxocariasis due to larvae of the canine ascarid *Toxocara canis*; the syndrome is due less commonly to the feline ascarid *T. cati* and even less commonly to the pig ascarid *Ascaris suum*. Rare cases with eosinophilic meningoencephalitis have been caused by the raccoon ascarid *Baylisascaris procyonis*.

LIFE CYCLE AND EPIDEMIOLOGY

The canine roundworm *T. canis* is distributed among dogs worldwide. Ingestion of infective eggs by dogs is followed by liberation of *Toxocara* larvae, which penetrate the gut wall and migrate intravascularly into canine tissues, where most remain in a developmentally arrested state. During pregnancy, some larvae resume migration in bitches and infect puppies prenatally (through transplacental transmission) or after birth (through suckling). Thus, in lactating bitches and puppies, larvae return to the intestinal tract and develop into adult worms, which produce eggs that are released in the feces. Eggs must undergo embryonation over several weeks to become infectious. Humans acquire toxocariasis mainly by eating soil contaminated by puppy feces that contains infective *T. canis* eggs. Visceral larva migrans is most common among children who habitually eat dirt.

PATHOGENESIS AND CLINICAL FEATURES

Clinical disease most commonly afflicts preschool children. After humans ingest *Toxocara* eggs, the larvae hatch and penetrate the intestinal mucosa, from which they are carried by the circulation to a wide variety of organs and tissues. The larvae invade the liver, lungs, central nervous system (CNS), and other sites, provoking intense local eosinophilic granulomatous responses. The degree of clinical illness depends on larval number and tissue distribution, reinfection, and host immune responses. Most light infections are asymptomatic and may be manifest only by blood eosinophilia. Characteristic symptoms of visceral larva migrans include fever, malaise, anorexia and weight loss, cough, wheezing, and rashes. Hepatosplenomegaly is common. These features are often accompanied by extraordinary peripheral eosinophilia, which may approach 90%. Uncommonly, seizures or behavioral disorders develop. Rare deaths are due to severe neurologic, pneumonic, or myocardial involvement.

The ocular form of the larva migrans syndrome occurs when *Toxocara* larvae invade the eye. An eosinophilic granulomatous mass, most commonly in the posterior pole of the retina, develops around the entrapped larva. The retinal lesion can mimic retinoblastoma in appearance, and mistaken diagnosis of the latter condition can lead to unnecessary enucleation. The spectrum of eye involvement also includes endophthalmitis, uveitis, and chorioretinitis. Unilateral visual disturbances, strabismus, and eye pain are the most common presenting symptoms. In contrast to visceral larva migrans, ocular toxocariasis usually develops in older children or young adults with no history of pica; these patients seldom have eosinophilia or visceral manifestations.

DIAGNOSIS

In addition to eosinophilia, leukocytosis and hypergammaglobulinemia may be evident. Transient pulmonary infiltrates are apparent on chest x-rays of about one-half of patients with symptoms of pneumonitis. The clinical diagnosis can be confirmed by an enzyme-linked immunosorbent assay for toxocaral antibodies. Stool examination for parasite eggs, while important in the evaluation of unexplained eosinophilia, is worthless for toxocariasis, since the larvae do not develop into egg-producing adults in humans.

Treatment: Visceral and Ocular Larva Migrans

The vast majority of *Toxocara* infections are self-limited and resolve without specific therapy. In patients with severe myocardial, CNS, or pulmonary involvement, glucocorticoids may be employed to reduce inflammatory complications. Available antihelminthic drugs, including mebendazole and

albendazole, have not been shown conclusively to alter the course of larva migrans. Control measures include prohibiting dog excreta in public parks and playgrounds, deworming dogs, and preventing pica in children. Treatment of ocular disease is not fully defined, but the administration of albendazole in conjunction with glucocorticoids has been effective (Table 216-1).

CUTANEOUS LARVA MIGRANS

Cutaneous larva migrans ("creeping eruption") is a serpiginous skin eruption caused by burrowing larvae of animal hookworms, usually the dog and cat hookworm *Ancylostoma braziliense*. The larvae hatch from eggs passed in dog and cat feces and mature in the soil. Humans become infected after skin contact with soil in areas frequented by dogs and cats, such as areas underneath house porches. Cutaneous larva migrans is prevalent among children and travelers in regions with warm humid climates, including the southeastern United States.

After larvae penetrate the skin, erythematous lesions form along the tortuous tracks of their migration through the dermal-epidermal junction; the larvae advance several centimeters in a day. The intensely pruritic lesions may occur anywhere on the body and can be numerous if the patient has lain on the ground. Vesicles and bullae may form later. The animal hookworm larvae do not mature in humans and, without treatment, will die after an interval ranging from weeks to a couple of months, with resolution of skin lesions. The diagnosis is made on clinical grounds. Skin biopsies only rarely detect diagnostic larvae. Symptoms can be alleviated by ivermectin or albendazole (Table 216-1).

ANGIOSTRONGYLIASIS

Angiostrongylus cantonensis, the rat lungworm, is the most common cause of human eosinophilic meningitis **(Fig. 216-3)**.

Figure 216-3

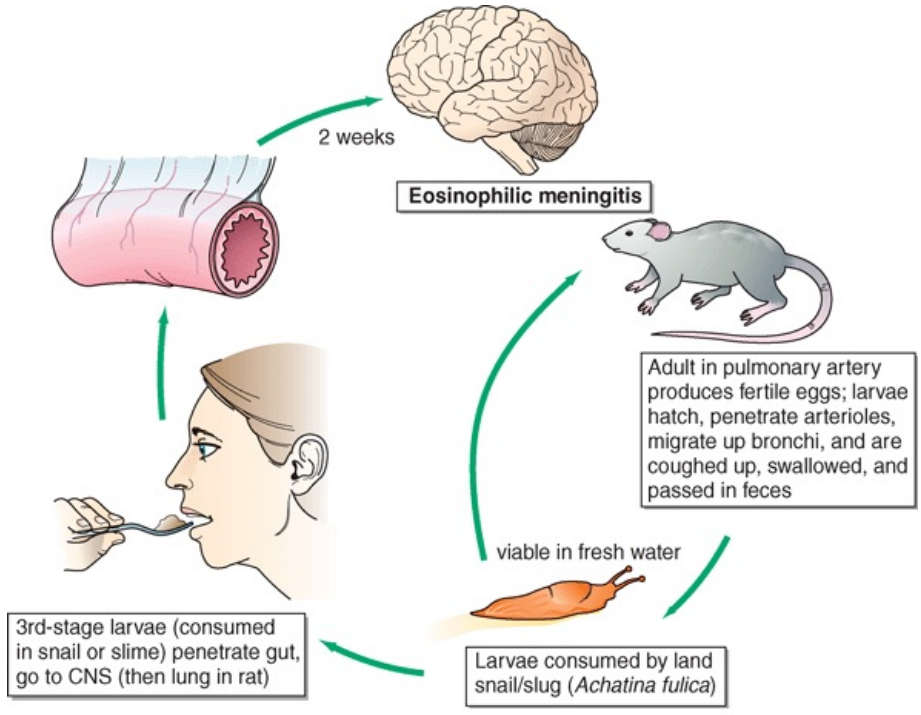

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Life cycle of *Angiostrongylus cantonensis* (rat lung worm), found in Southeast Asia, Pacific Islands, Cuba, Australia, Japan, China, Mauritius, and U.S. ports. CNS, central nervous system. *[Reprinted from Guerrant RL et al (eds): Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed, p 1225. © 2006, with permission from Elsevier Science.]*

LIFE CYCLE AND EPIDEMIOLOGY

This infection occurs principally in Southeast Asia and the Pacific Basin but has spread to other areas of the world. *A. cantonensis* larvae produced by adult worms in the rat lung migrate to the gastrointestinal tract and are expelled with the feces. They develop into infective larvae in land snails and slugs. Humans acquire the infection by ingesting raw infected mollusks; vegetables contaminated by mollusk slime; or crabs, freshwater shrimp, and certain marine fish that have themselves eaten infected mollusks. The larvae then migrate to the brain.

PATHOGENESIS AND CLINICAL FEATURES

The parasites eventually die in the CNS, but not before initiating pathologic consequences that, in heavy infections, can result in permanent neurologic sequelae or death. Migrating larvae cause marked local eosinophilic inflammation and hemorrhage, with subsequent necrosis and granuloma formation around dying worms. Clinical symptoms develop 2–35 days after the ingestion of larvae. Patients usually present with an insidious or abrupt excruciating frontal, occipital, or bitemporal headache. Neck stiffness, nausea and vomiting, and paresthesias are also common. Fever, cranial and extraocular nerve palsies, seizures, paralysis, and lethargy are uncommon.

LABORATORY FINDINGS

Examination of cerebrospinal fluid (CSF) is mandatory in suspected cases and usually reveals an elevated opening pressure, a white blood cell count of 150–2000/μL, and an eosinophilic pleocytosis of >20%. The protein concentration is usually elevated and the glucose level normal. The larvae of *A. cantonensis* are only rarely seen in CSF. Peripheral-blood eosinophilia may be mild. The diagnosis is generally based on the clinical presentation of eosinophilic meningitis together with a compatible epidemiologic history.

Treatment: Angiostrongyliasis

Specific chemotherapy is not of benefit in angiostrongyliasis; larvicidal agents may exacerbate inflammatory brain lesions. Management consists of supportive measures, including the administration of analgesics, sedatives, and—in severe cases—glucocorticoids (Table 216-1). Repeated lumbar punctures with removal of CSF can relieve symptoms. In most patients, cerebral angiostrongyliasis has a self-limited course, and recovery is complete. The infection may be prevented by adequately cooking snails, crabs, and prawns and inspecting vegetables for mollusk infestation. Other parasitic or fungal causes of eosinophilic meningitis in endemic areas may include gnathostomiasis (see below), paragonimiasis (Chap. 219), schistosomiasis (Chap. 219), neurocysticercosis (Chap. 220), and coccidioidomycosis (Chap. 200).

GNATHOSTOMIASIS

Infection of human tissues with larvae of *Gnathostoma spinigerum* can cause eosinophilic meningoencephalitis, migratory cutaneous swellings, or invasive masses of the eye and visceral organs.

LIFE CYCLE AND EPIDEMIOLOGY

Human gnathostomiasis occurs in many countries and is notably endemic in Southeast Asia and parts of China and Japan. In nature, the mature adult worms parasitize the gastrointestinal tract of dogs and cats. First-stage larvae hatch from eggs passed into water and are ingested by *Cyclops* species (water fleas). Infective third-stage larvae develop in the flesh of many animal species (including fish, frogs, eels, snakes, chickens, and ducks) that have eaten either infected *Cyclops* or another infected second intermediate host. Humans typically acquire the infection by eating raw or undercooked fish or poultry. Raw fish dishes, such as *som fak* in Thailand and *sashimi* in Japan, account for many cases of human gnathostomiasis. Some cases in Thailand result from the local practice of applying frog or snake flesh as a poultice.

PATHOGENESIS AND CLINICAL FEATURES

Clinical symptoms are due to the aberrant migration of a single larva into cutaneous, visceral, neural, or ocular tissues. After invasion, larval migration may cause local inflammation, with pain, cough, or hematuria accompanied by fever and eosinophilia. Painful, itchy, migratory swellings may develop in the skin, particularly in the distal extremities or periorbital area. Cutaneous swellings usually last ~1 week but often recur intermittently over many years. Larval invasion of the eye can provoke a sight-threatening inflammatory response. Invasion of the CNS results in eosinophilic meningitis with myeloencephalitis, a serious complication due to ascending larval migration along a large nerve track. Patients characteristically present with agonizing radicular pain and paresthesias in the trunk or a limb, which are followed shortly by paraplegia. Cerebral involvement, with focal hemorrhages and tissue destruction, is often fatal.

DIAGNOSIS AND TREATMENT

Cutaneous migratory swellings with marked peripheral eosinophilia, supported by an appropriate geographic and dietary history, generally constitute an adequate basis for a clinical diagnosis of gnathostomiasis. However, patients may present with ocular or cerebrospinal involvement without antecedent cutaneous swellings. In the latter case, eosinophilic pleocytosis is demonstrable (usually along with hemorrhagic or xanthochromic CSF), but worms are almost never recovered from CSF. Surgical removal of the parasite from subcutaneous or ocular tissue, though rarely feasible, is both diagnostic and therapeutic. Albendazole or ivermectin may be helpful (Table 216-1). At present, cerebrospinal involvement is managed with supportive measures and generally with a course of glucocorticoids. Gnathostomiasis can be prevented by adequate cooking of fish and poultry in endemic areas.

FURTHER READINGS

Articles

Bouchard O et al: Cutaneous larva migrans in travelers: A prospective study, with assessment of therapy with ivermectin. Clin Infect Dis 31:493, 2000

Bowman DD et al: Hookworms of dogs and cats as agents of cutaneous larva migrans. Trends Parasitol 26:162, 2010[PMID: 20189454]

Herman JS, Chiodini PL: Gnathostomiasis, another emerging imported disease. Clin Microbiol Rev 22:484, 2009[PMID: 19597010]

Kennedy ED et al: Trichinellosis surveillance—United States, 2002–2007. MMWR Surveill Summ 58:1, 2009[PMID: 19959986]

Lederman ER et al: Dermatologic conditions of the ill returned traveler: An analysis from the GeoSentinel Surveillance Network. Int J Infect Dis 12:593, 2008[PMID: 18343180]

Ramirez-Avila L et al: Eosinophilic meningitis due to *Angiostrongylus* and *Gnathostoma* species. Clin Infect Dis 48:322, 2009[PMID: 19123863]

Rubinsky-Elefant G et al: Human toxocariasis: Diagnosis, worldwide seroprevalences and clinical expression of the systemic and ocular forms. Ann Trop Med Parasitol 104:3, 2010[PMID: 20149289]

Sakai S et al: Pulmonary lesions associated with visceral larva migrans due to *Ascaris suum* or *Toxocara canis*: Imaging of six cases. AJR Am J Roentgenol 186:1697, 2006[PMID: 16714661]

Shimoni Z et al: The use of prednisone in the treatment of trichinellosis. Isr Med Assoc J 9:537, 2007[PMID: 17710786]

Wang QP et al: Human angiostrongyliasis. Lancet Infect Dis 8:621, 2008[PMID: 18922484]

Won ky et al: National seroprevalence and risk factors for zoonotic *Toxocara* spp. infection. Am J Trop Med Hyg 79:552, 2008[PMID: 18840743]

Web Sites

CDC Division of Parasitic Diseases. *http://www.cdc.gov/ncidod/dpd/default.htm*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTESTINAL NEMATODE INFECTIONS: INTRODUCTION

More than a billion persons worldwide are infected with one or more species of intestinal nematodes. **Table 217-1** summarizes biologic and clinical features of infections due to the major intestinal parasitic nematodes. These parasites are most common in regions with poor fecal sanitation, particularly in resource-poor countries in the tropics and subtropics, but they have also been seen with increasing frequency among immigrants and refugees to resource-rich countries. Although nematode infections are not usually fatal, they contribute to malnutrition and diminished work capacity. It is interesting that these helminth infections may protect some individuals from allergic disease. Humans may on occasion be infected with nematode parasites that ordinarily infect animals; these zoonotic infections produce diseases such as trichostrongyliasis, anisakiasis, capillariasis, and abdominal angiostrongyliasis.

Table 217-1 Major Human Intestinal Parasitic Nematodes

	Parasitic Nematode				
Feature	***Ascaris lumbricoides* (Roundworm)**	***Necator americanus*, *Ancylostoma duodenale* (Hookworm)**	***Strongyloides stercoralis***	***Trichuris trichiura* (Whipworm)**	***Enterobius vermicularis* (Pinworm)**
Global prevalence in humans (millions)	807	576	100	604	209
Endemic areas	Worldwide	Hot, humid regions	Hot, humid regions	Worldwide	Worldwide
Infective stage	Egg	Filariform larva	Filariform larva	Egg	Egg
Route of infection	Oral	Percutaneous	Percutaneous or autoinfection	Oral	Oral
Gastrointestinal location of worms	Jejunal lumen	Jejunal mucosa	Small-bowel mucosa	Cecum, colonic mucosa	Cecum, appendix
Adult worm size	15–40 cm	7–12 mm	2 mm	30–50 mm	8–13 mm (female)
Pulmonary passage of larvae	Yes	Yes	Yes	No	No
Incubation period[a] (days)	60–75	40–100	17–28	70–90	35–45
Longevity	1 y	*N. americanus*: 2–5 y *A. duodenale*: 6–8 y	Decades (owing to autoinfection)	5 y	2 months
Fecundity (eggs/day/worm)	240,000	*N. americanus*: 4000–10,000 *A. duodenale*: 10,000–25,000	5000–10,000	3000–7000	2000
Principal symptoms	Rarely gastrointestinal or biliary obstruction	Iron-deficiency anemia in heavy infection	Gastrointestinal symptoms; malabsorption or sepsis in hyperinfection	Gastrointestinal symptoms, anemia	Perianal pruritus
Diagnostic stage	Eggs in stool	Eggs in fresh stool, larvae in old stool	Larvae in stool or duodenal aspirate; sputum in	Eggs in stool	Eggs from perianal skin on cellulose

			hyperinfection		acetate tape
Treatment	Mebendazole Albendazole Pyrantel pamoate Ivermectin Nitazoxanide	Mebendazole Pyrantel pamoate Albendazole	1. Ivermectin 2. Albendazole	Mebendazole Albendazole Ivermectin	Mebendazole Pyrantel pamoate Albendazole

[a]Time from infection to egg production by mature female worm.

Intestinal nematodes are roundworms; they range in length from 1 mm to many centimeters when mature (Table 217-1). Their life cycles are complex and highly varied; some species, including *Strongyloides stercoralis* and *Enterobius vermicularis*, can be transmitted directly from person to person, while others, such as *Ascaris lumbricoides*, *Necator americanus*, and *Ancylostoma duodenale*, require a soil phase for development. Because most helminth parasites do not self-replicate, the acquisition of a heavy burden of adult worms requires repeated exposure to the parasite in its infectious stage, whether larval or egg. Hence, clinical disease, as opposed to asymptomatic infection, generally develops only with prolonged residence in an endemic area and is typically related to infection intensity. In persons with marginal nutrition, intestinal helminth infections may impair growth and development. Eosinophilia and elevated serum IgE levels are features of many helminth infections and, when unexplained, should always prompt a search for intestinal helminths. Significant protective immunity to intestinal nematodes appears not to develop in humans, although mechanisms of parasite immune evasion and host immune responses to these infections have not been elucidated in detail.

ASCARIASIS

A. lumbricoides is the largest intestinal nematode parasite of humans, reaching up to 40 cm in length. Most infected individuals have low worm burdens and are asymptomatic. Clinical disease arises from larval migration in the lungs or effects of the adult worms in the intestines.

LIFE CYCLE

Adult worms live in the lumen of the small intestine. Mature female *Ascaris* worms are extraordinarily fecund, each producing up to 240,000 eggs a day, which pass with the feces. Ascarid eggs, which are remarkably resistant to environmental stresses, become infective after several weeks of maturation in the soil and can remain infective for years. After infective eggs are swallowed, larvae hatched in the intestine invade the mucosa, migrate through the circulation to the lungs, break into the alveoli, ascend the bronchial tree, and return—through swallowing—to the small intestine, where they develop into adult worms. Between 2 and 3 months elapse between initial infection and egg production. Adult worms live for 1–2 years.

EPIDEMIOLOGY

Ascaris is widely distributed in tropical and subtropical regions as well as in other humid areas, including the rural southeastern United States. Transmission typically occurs through fecally contaminated soil and is due either to a lack of sanitary facilities or to the use of human feces as fertilizer. With their propensity for hand-to-mouth fecal carriage, younger children are most affected. Infection outside endemic areas, though uncommon, can occur when eggs on transported vegetables are ingested.

CLINICAL FEATURES

During the lung phase of larval migration, ~9–12 days after egg ingestion, patients may develop an irritating nonproductive cough and burning substernal discomfort that is aggravated by coughing or deep inspiration. Dyspnea and blood-tinged sputum are less common. Fever is usually reported. Eosinophilia develops during this symptomatic phase and subsides slowly over weeks. Chest x-rays may reveal evidence of eosinophilic pneumonitis (Löffler's syndrome), with rounded infiltrates a few millimeters to several centimeters in size. These infiltrates may be transient and intermittent, clearing after several weeks. Where there is seasonal transmission of the parasite, seasonal pneumonitis with eosinophilia may develop in previously infected and sensitized hosts.

In established infections, adult worms in the small intestine usually cause no symptoms. In heavy infections, particularly in children, a large bolus of entangled worms can cause pain and small-bowel obstruction, sometimes complicated by perforation, intussusception, or volvulus. Single worms may cause disease when they migrate into aberrant sites. A large worm can enter and occlude the biliary tree, causing biliary colic, cholecystitis, cholangitis, pancreatitis, or (rarely) intrahepatic abscesses. Migration of an adult worm up the esophagus can provoke coughing and oral expulsion of the worm. In highly endemic areas, intestinal and biliary ascariasis can rival acute appendicitis and gallstones as causes of surgical acute abdomen.

LABORATORY FINDINGS

Most cases of ascariasis can be diagnosed by microscopic detection of characteristic *Ascaris* eggs (65 by 45 μm) in fecal samples. Occasionally, patients present after passing an adult worm —identifiable by its large size and smooth cream-colored surface—in the stool or through the mouth or nose. During the early transpulmonary migratory phase, when eosinophilic pneumonitis occurs, larvae can be found in sputum or gastric aspirates before diagnostic eggs appear in the stool. The eosinophilia that is prominent during this early stage usually decreases to minimal levels in established infection. Adult worms may be visualized, occasionally serendipitously, on contrast studies of the gastrointestinal tract. A plain abdominal film may reveal masses of worms in gas-filled loops of bowel in patients with intestinal obstruction. Pancreaticobiliary worms can be detected by ultrasound and endoscopic retrograde cholangiopancreatography; the latter method also has been used to extract biliary *Ascaris* worms.

Treatment: Ascariasis

Ascariasis should always be treated to prevent potentially serious complications. Albendazole (400 mg once), mebendazole (100 g twice daily for 3 days or 500 mg once), or ivermectin (150–200 mg/kg once) is effective. These medications are contraindicated in pregnancy, however. Pyrantel pamoate (11 mg/kg once; maximum, 1 g) is safe in pregnancy. Nitazoxanide (7.5 mg/kg once; maximum, 500 mg) has also been used in ascariasis. Mild diarrhea and abdominal pain are uncommon side effects of these agents. Partial intestinal obstruction should be managed with nasogastric suction, IV fluid administration, and instillation of piperazine through the nasogastric tube, but complete obstruction and its severe complications require immediate surgical intervention.

HOOKWORM

Two hookworm species (*A. duodenale* and *N. americanus*) are responsible for human infections. Most infected individuals are asymptomatic. Hookworm disease develops from a combination of factors—a heavy worm burden, a prolonged duration of infection, and an inadequate iron intake—and results in iron-deficiency anemia and, on occasion, hypoproteinemia.

LIFE CYCLE

Adult hookworms, which are ~1 cm long, use buccal teeth (*Ancylostoma*) or cutting plates (*Necator*) to attach to the small-bowel mucosa and suck blood (0.2 mL/d per *Ancylostoma* adult) and interstitial fluid. The adult hookworms produce thousands of eggs daily. The eggs are deposited with feces in soil, where rhabditiform larvae hatch and develop over a 1-week period into infectious filariform larvae. Infective larvae penetrate the skin and reach the lungs by way of the bloodstream. There they invade alveoli and ascend the airways before being swallowed and reaching the small intestine. The prepatent period from skin invasion to appearance of eggs in the feces is ~6–8 weeks, but it may be longer with *A. duodenale*. Larvae of *A. duodenale*, if swallowed, can survive and develop directly in the intestinal mucosa. Adult hookworms may survive over a decade but usually live ~6–8 years for *A. duodenale* and 2–5 years for *N. americanus*.

EPIDEMIOLOGY

A. duodenale is prevalent in southern Europe, North Africa, and northern Asia, and *N. americanus* is the predominant species in the Western Hemisphere and equatorial Africa. The two species overlap in many tropical regions, particularly Southeast Asia. In most areas, older children have the highest incidence and greatest intensity of hookworm infection. In rural areas where fields are fertilized with human feces, older working adults also may be heavily infected.

CLINICAL FEATURES

Most hookworm infections are asymptomatic. Infective larvae may provoke pruritic maculopapular dermatitis ("ground itch") at the site of skin penetration as well as serpiginous tracks of subcutaneous migration (similar to those of cutaneous larva migrans; Chap. 216) in previously sensitized hosts. Larvae migrating through the lungs occasionally cause mild transient pneumonitis, but this condition develops less frequently in hookworm infection than in ascariasis. In the early intestinal phase, infected persons may develop epigastric pain (often with postprandial accentuation), inflammatory diarrhea, or other abdominal symptoms accompanied by eosinophilia. The major consequence of chronic hookworm infection is iron deficiency. Symptoms are minimal if iron intake is adequate, but marginally nourished individuals develop symptoms of progressive iron-deficiency anemia and hypoproteinemia, including weakness and shortness of breath.

LABORATORY FINDINGS

The diagnosis is established by the finding of characteristic 40- by 60-μm oval hookworm eggs in the feces. Stool-concentration procedures may be required to detect light infections. Eggs of the two species are indistinguishable by light microscopy. In a stool sample that is not fresh, the eggs may have hatched to release rhabditiform larvae, which need to be differentiated from those of *S. stercoralis*. Hypochromic microcytic anemia, occasionally with eosinophilia or hypoalbuminemia, is characteristic of hookworm disease.

Treatment: Hookworm Infection

Hookworm infection can be eradicated with several safe and highly effective antihelminthic drugs, including albendazole (400 mg once), mebendazole (500 mg once), and pyrantel pamoate (11 mg/kg for 3 days). Mild iron-deficiency anemia can often be treated with oral iron alone. Severe hookworm disease with protein loss and malabsorption necessitates nutritional support and oral iron replacement along with deworming. There is some concern that the benzimidazoles (mebendazole and albendazole) are becoming less effective against human hookworms than in the past.

ANCYLOSTOMA CANINUM AND ANCYLOSTOMA BRAZILIENSE

A. caninum, the canine hookworm, has been identified as a cause of human eosinophilic enteritis, especially in northeastern Australia. In this zoonotic infection, adult hookworms attach to the small intestine (where they may be visualized by endoscopy) and elicit abdominal pain and intense local eosinophilia. Treatment with mebendazole (100 mg twice daily for 3 days) or albendazole (400 mg once) or endoscopic removal is effective. Both of these animal hookworm species can cause cutaneous larva migrans ("creeping eruption"; Chap. 216).

STRONGYLOIDIASIS

S. stercoralis is distinguished by its ability—unique among helminths (except for Capillaria; see below)—to replicate in the human host. This capacity permits ongoing cycles of autoinfection as infective larvae are internally produced. Strongyloidiasis can thus persist for decades without further exposure of the host to exogenous infective larvae. In immunocompromised hosts, large numbers of invasive *Strongyloides* larvae can disseminate widely and can be fatal.

LIFE CYCLE

In addition to a parasitic cycle of development, *Strongyloides* can undergo a free-living cycle of development in the soil **(Fig. 217-1)**. This adaptability facilitates the parasite's survival in the absence of mammalian hosts. Rhabditiform larvae passed in feces can transform into infectious filariform larvae either directly or after a free-living phase of development. Humans acquire strongyloidiasis when filariform larvae in fecally contaminated soil penetrate the skin or mucous membranes. The larvae then travel through the bloodstream to the lungs, where they break into the alveolar spaces, ascend the bronchial tree, are swallowed, and thereby reach the small intestine. There the larvae mature into adult worms that penetrate the mucosa of the proximal small bowel. The minute (2-mm-long) parasitic adult female worms reproduce by parthenogenesis; adult males do not exist. Eggs hatch in the intestinal mucosa, releasing rhabditiform larvae that migrate to the lumen and pass with the feces into soil. Alternatively, rhabditiform larvae in the bowel can develop directly into filariform larvae that penetrate the colonic wall or perianal skin and enter the circulation to repeat the migration that establishes ongoing internal reinfection. This autoinfection cycle allows strongyloidiasis to persist for decades.

Figure 217-1

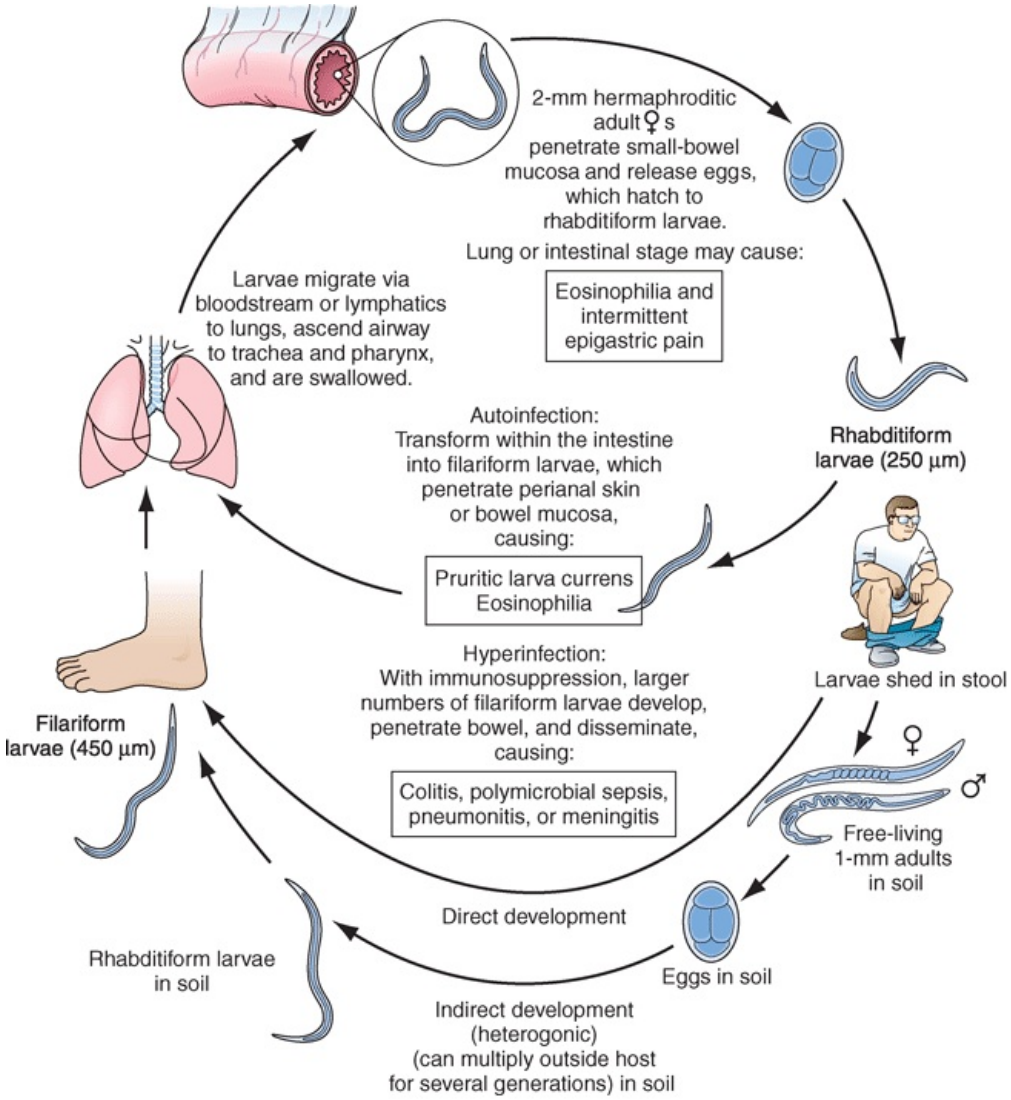

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Life cycle of *Strongyloides stercoralis.* *(Adapted from Guerrant RL et al (eds): Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed, p 1276. © 2006, with permission from Elsevier Science.)*

EPIDEMIOLOGY

S. stercoralis is spottily distributed in tropical areas and other hot, humid regions and is particularly common in Southeast Asia, sub-Saharan Africa, and Brazil. In the United States, the parasite is endemic in parts of the Southeast and is found in immigrants, refugees, travelers, and military personnel who have lived in endemic areas.

CLINICAL FEATURES

In uncomplicated strongyloidiasis, many patients are asymptomatic or have mild cutaneous and/or abdominal symptoms. Recurrent urticaria, often involving the buttocks and wrists, is the most common cutaneous manifestation. Migrating larvae can elicit a pathognomonic serpiginous eruption, *larva currens* ("running larva"). This pruritic, raised, erythematous lesion advances as rapidly as 10 cm/h along the course of larval migration. Adult parasites burrow into the duodenojejunal mucosa and can cause abdominal (usually midepigastric) pain, which resembles peptic ulcer pain except that it is aggravated by food ingestion. Nausea, diarrhea, gastrointestinal bleeding, mild chronic colitis, and weight loss can occur. Small-bowel obstruction may develop with early, heavy infection. Pulmonary symptoms are rare in uncomplicated strongyloidiasis. Eosinophilia is common, with levels fluctuating over time.

The ongoing autoinfection cycle of strongyloidiasis is normally constrained by unknown factors of the host's immune system. Abrogation of host immunity, especially with glucocorticoid therapy and much less commonly with other immunosuppressive medications, leads to hyperinfection, with the generation of large numbers of filariform larvae. Colitis, enteritis, or malabsorption may develop. In disseminated strongyloidiasis, larvae may invade not only gastrointestinal tissues and the lungs but also the central

nervous system, peritoneum, liver, and kidneys. Moreover, bacteremia may develop because of the passage of enteric flora through disrupted mucosal barriers. Gram-negative sepsis, pneumonia, or meningitis may complicate or dominate the clinical course. Eosinophilia is often absent in severely infected patients. Disseminated strongyloidiasis, particularly in patients with unsuspected infection who are given glucocorticoids, can be fatal. Strongyloidiasis is a frequent complication of infection with human T cell lymphotropic virus type I, but disseminated strongyloidiasis is not common among patients infected with HIV-1.

DIAGNOSIS

In uncomplicated strongyloidiasis, the finding of rhabditiform larvae in feces is diagnostic. Rhabditiform larvae are ~250 μm long, with a short buccal cavity that distinguishes them from hookworm larvae. In uncomplicated infections, few larvae are passed and single stool examinations detect only about one-third of cases. Serial examinations and the use of the agar plate detection method improve the sensitivity of stool diagnosis. In uncomplicated strongyloidiasis (but not in hyperinfection), stool examinations may be repeatedly negative. *Strongyloides* larvae may also be found by sampling of the duodenojejunal contents by aspiration or biopsy. An enzyme-linked immunosorbent assay for serum antibodies to antigens of *Strongyloides* is a sensitive method of diagnosing uncomplicated infections. Such serologic testing should be performed for patients whose geographic histories indicate potential exposure, especially those who exhibit eosinophilia and/or are candidates for glucocorticoid treatment of other conditions. In disseminated strongyloidiasis, filariform larvae should be sought in stool as well as in samples obtained from sites of potential larval migration, including sputum, bronchoalveolar lavage fluid, or surgical drainage fluid.

Treatment: Strongyloidiasis

Even in the asymptomatic state, strongyloidiasis must be treated because of the potential for subsequent fatal hyperinfection. Ivermectin (200 mg/kg daily for 2 days) is more effective than albendazole (400 mg daily for 3 days). For disseminated strongyloidiasis, treatment with ivermectin should be extended for at least 5–7 days or until the parasites are eradicated.

TRICHURIASIS

Most infections with *Trichuris trichiura* are asymptomatic, but heavy infections may cause gastrointestinal symptoms. Like the other soil-transmitted helminths, whipworm is distributed globally in the tropics and subtropics and is most common among poor children from resource-poor regions of the world.

LIFE CYCLE

Adult *Trichuris* worms reside in the colon and cecum, the anterior portions threaded into the superficial mucosa. Thousands of eggs laid daily by adult female worms pass with the feces and mature in the soil. After ingestion, infective eggs hatch in the duodenum, releasing larvae that mature before migrating to the large bowel. The entire cycle takes ~3 months, and adult worms may live for several years.

CLINICAL FEATURES

Tissue reactions to *Trichuris* are mild. Most infected individuals have no symptoms or eosinophilia. Heavy infections may result in abdominal pain, anorexia, and bloody or mucoid diarrhea resembling inflammatory bowel disease. Rectal prolapse can result from massive infections in children, who often suffer from malnourishment and other diarrheal illnesses. Moderately heavy *Trichuris* burdens also contribute to growth retardation.

DIAGNOSIS AND TREATMENT

The characteristic 50- by 20-μm lemon-shaped *Trichuris* eggs are readily detected on stool examination. Adult worms, which are 3–5 cm long, are occasionally seen on proctoscopy. Mebendazole (500 mg once) or albendazole (400 mg daily for 3 doses) is safe and moderately effective for treatment, with cure rates of 70–90%. Ivermectin (200 mg/kg daily for 3 doses) is also safe but is not quite as efficacious as the benzimidazoles.

ENTEROBIASIS (PINWORM)

E. vermicularis is more common in temperate countries than in the tropics. In the United States, ~40 million persons are infected with pinworms, with a disproportionate number of cases among children.

LIFE CYCLE AND EPIDEMIOLOGY

Enterobius adult worms are ~1 cm long and dwell in the cecum. Gravid female worms migrate nocturnally into the perianal region and release up to 10,000 immature eggs each. The eggs become infective within hours and are transmitted by hand-to-mouth passage. From ingested eggs, larvae hatch and mature into adults. This life cycle takes ~1 month, and adult worms survive for ~2 months. Self-infection results from perianal scratching and transport of infective eggs on the hands or under the nails to the mouth. Because of the ease of person-to-person spread, pinworm infections are common

among family members.

CLINICAL FEATURES

Most pinworm infections are asymptomatic. Perianal pruritus is the cardinal symptom. The itching, which is often worse at night as a result of the nocturnal migration of the female worms, may lead to excoriation and bacterial superinfection. Heavy infections have been claimed to cause abdominal pain and weight loss. On rare occasions, pinworms invade the female genital tract, causing vulvovaginitis and pelvic or peritoneal granulomas. Eosinophilia is uncommon.

DIAGNOSIS

Since pinworm eggs are not released in feces, the diagnosis cannot be made by conventional fecal ova and parasite tests. Instead, eggs are detected by the application of clear cellulose acetate tape to the perianal region in the morning. After the tape is transferred to a slide, microscopic examination will detect pinworm eggs, which are oval, measure 55 by 25 μm, and are flattened along one side.

Treatment: Enterobiasis

Infected children and adults should be treated with mebendazole (100 mg once), albendazole (400 mg once), or pyrantel pamoate (11 mg/kg once; maximum, 1 g), with the same treatment repeated after 2 weeks. Treatment of household members is advocated to eliminate asymptomatic reservoirs of potential reinfection.

TRICHOSTRONGYLIASIS

Trichostrongylus species, which are normally parasites of herbivorous animals, occasionally infect humans, particularly in Asia and Africa. Humans acquire the infection by accidentally ingesting *Trichostrongylus* larvae on contaminated leafy vegetables. The larvae do not migrate in humans but mature directly into adult worms in the small bowel. These worms ingest far less blood than hookworms; most infected persons are asymptomatic, but heavy infections may give rise to mild anemia and eosinophilia. *Trichostrongylus* eggs in stool examinations resemble those of hookworms but are larger (85 by 115 μm). Treatment consists of mebendazole or albendazole (Chap. 208).

ANISAKIASIS

Anisakiasis is a gastrointestinal infection caused by the accidental ingestion in uncooked saltwater fish of nematode larvae belonging to the family Anisakidae. The incidence of anisakiasis in the United States has increased as a result of the growing popularity of raw fish dishes. Most cases occur in Japan, the Netherlands, and Chile, where raw fish—sashimi, pickled green herring, and ceviche, respectively—are national culinary staples. Anisakid nematodes parasitize large sea mammals such as whales, dolphins, and seals. As part of a complex parasitic life cycle involving marine food chains, infectious larvae migrate to the musculature of a variety of fish. Both *Anisakis simplex* and *Pseudoterranova decipiens* have been implicated in human anisakiasis, but an identical gastric syndrome may be caused by the red larvae of eustrongylid parasites of fish-eating birds.

When humans consume infected raw fish, live larvae may be coughed up within 48 h. Alternatively, larvae may immediately penetrate the mucosa of the stomach. Within hours, violent upper abdominal pain accompanied by nausea and occasionally vomiting ensues, mimicking an acute abdomen. The diagnosis can be established by direct visualization on upper endoscopy, outlining of the worm by contrast radiographic studies, or histopathologic examination of extracted tissue. Extraction of the burrowing larvae during endoscopy is curative. In addition, larvae may pass to the small bowel, where they penetrate the mucosa and provoke a vigorous eosinophilic granulomatous response. Symptoms may appear 1–2 weeks after the infective meal, with intermittent abdominal pain, diarrhea, nausea, and fever resembling the manifestations of Crohn's disease. The diagnosis may be suggested by barium studies and confirmed by curative surgical resection of a granuloma in which the worm is embedded. Anisakid eggs are not found in the stool, since the larvae do not mature in humans. Serologic tests have been developed but are not widely available.

Anisakid larvae in saltwater fish are killed by cooking to 60°C, freezing at –20°C for 3 days, or commercial blast freezing, but not usually by salting, marinating, or cold smoking. No medical treatment is available; surgical or endoscopic removal should be undertaken.

CAPILLARIASIS

Intestinal capillariasis is caused by ingestion of raw fish infected with *Capillaria philippinensis*. Subsequent autoinfection can lead to a severe wasting syndrome. The disease occurs in the Philippines and Thailand and, on occasion, elsewhere in Asia. The natural cycle of *C. philippinensis* involves fish from fresh and brackish water. When humans eat infected raw fish, the larvae mature in the intestine into adult worms, which produce invasive larvae that cause intestinal inflammation and villus loss. Capillariasis has an insidious onset with nonspecific abdominal pain and watery diarrhea. If untreated, progressive autoinfection can lead to protein-losing enteropathy, severe malabsorption, and ultimately death from cachexia, cardiac failure, or superinfection. The diagnosis is established by

identification of the characteristic peanut-shaped (20- by 40-μm) eggs on stool examination. Severely ill patients require hospitalization and supportive therapy in addition to prolonged antihelminthic treatment with albendazole (200 mg twice daily for 10 days; Chap. 208).

ABDOMINAL ANGIOSTRONGYLIASIS

Abdominal angiostrongyliasis is found in Latin America and Africa. The zoonotic parasite *Angiostrongylus costaricensis* causes eosinophilic ileocolitis after the ingestion of contaminated vegetation. *A. costaricensis* normally parasitizes the cotton rat and other rodents, with slugs and snails serving as intermediate hosts. Humans become infected by accidentally ingesting infective larvae in mollusk slime deposited on fruits and vegetables; children are at highest risk. The larvae penetrate the gut wall and migrate to the mesenteric artery, where they develop into adult worms. Eggs deposited in the gut wall provoke an intense eosinophilic granulomatous reaction, and adult worms may cause mesenteric arteritis, thrombosis, or frank bowel infarction. Symptoms may mimic those of appendicitis, including abdominal pain and tenderness, fever, vomiting, and a palpable mass in the right iliac fossa. Leukocytosis and eosinophilia are prominent. CT with contrast medium typically shows inflamed bowel, often with concomitant obstruction, but a definitive diagnosis is usually made surgically with partial bowel resection. Pathologic study reveals a thickened bowel wall with eosinophilic granulomas surrounding the *Angiostrongylus* eggs. In nonsurgical cases, the diagnosis rests solely on clinical grounds because larvae and eggs cannot be detected in the stool. Medical therapy for abdominal angiostrongyliasis is of uncertain efficacy. Careful observation and surgical resection for severe symptoms are the mainstays of treatment.

FURTHER READINGS

Audicana MT, Kennedy MW: *Anisakis* simplex: From obscure infectious worm to inducer of immune hypersensitivity. Clin Microbiol Rev 21:360, 2008[PMID: 18400801]

Bethony J et al: Soil-transmitted helminth infections: Ascariasis, trichuriasis, and hookworm. Lancet 367:1521, 2006[PMID: 16679166]

Fox LM, Saravolatz LD: Nitazoxanide: A new thiazolide antiparasitic agent. Clin Infect Dis 40:1173, 2005[PMID: 15791519]

Harhay MO et al: Epidemiology and control of human gastrointestinal parasites in children. Expert Rev Anti Infect Ther 8:219, 2010[PMID: 20109051]

Hotez PJ et al: Hookworm infection. N Engl J Med 351:799, 2004[PMID: 15317893]

Keiser J, Utzinger J: Efficacy of current drugs against soil-transmitted helminth infections: Systematic review and meta-analysis. JAMA 299:1937, 2008[PMID: 18430913]

Lu LH et al: Human intestinal capillariasis (*Capillaria philippinensis*) in Taiwan. Am J Trop Med Hyg 74:810, 2006[PMID: 16687685]

Ramanathan R, Nutman T: *Strongyloides stercoralis* infection in the immunocompromised host. Curr Infect Dis Rep 10:105, 2008[PMID: 18462583]

Roxby AC et al: Strongyloidiasis in transplant patients. Clin Infect Dis 49:1411, 2009[PMID: 19807271]

Saichua P et al: Human intestinal capillariasis in Thailand. World J Gastroenterol 14:506, 2008[PMID: 18203280]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

FILARIAL AND RELATED INFECTIONS: INTRODUCTION

Filarial worms are nematodes that dwell in the subcutaneous tissues and the lymphatics. Eight filarial species infect humans **(Table 218-1)**; of these, four—*Wuchereria bancrofti*, *Brugia malayi*, *Onchocerca volvulus*, and *Loa loa*—are responsible for most serious filarial infections. Filarial parasites, which infect an estimated 170 million persons worldwide, are transmitted by specific species of mosquitoes or other arthropods and have a complex life cycle, including infective larval stages carried by insects and adult worms that reside in either lymphatic or subcutaneous tissues of humans. The offspring of adults are microfilariae, which, depending on their species, are 200–250 μm long and 5–7 μm wide, may or may not be enveloped in a loose sheath, and either circulate in the blood or migrate through the skin (Table 218-1). To complete the life cycle, microfilariae are ingested by the arthropod vector and develop over 1–2 weeks into new infective larvae. Adult worms live for many years, whereas microfilariae survive for 3–36 months. The *Rickettsia*-like endosymbiont *Wolbachia* has been found intracellularly in all stages of *Brugia*, *Wuchereria*, *Mansonella*, and *Onchocerca* and has become a target for antifilarial chemotherapy.

Table 218-1 Characteristics of the Filariae

Organism	Periodicity	Distribution	Vector	Location of Adult	Microfilarial Location	Sheath
Wuchereria bancrofti	Nocturnal	Cosmopolitan areas worldwide, including South America, Africa, southern Asia, Papua New Guinea, China, Indonesia	*Culex, Anopheles* (mosquitoes)	Lymphatic tissue	Blood	+
	Subperiodic	Eastern Pacific	*Aedes* (mosquitoes)	Lymphatic tissue	Blood	+
Brugia malayi	Nocturnal	Southeast Asia, Indonesia, India	*Mansonia*, *Anopheles* (mosquitoes)	Lymphatic tissue	Blood	+
	Subperiodic	Indonesia, Southeast Asia	*Coquillettidia*, *Mansonia* (mosquitoes)	Lymphatic tissue	Blood	+
B. timori	Nocturnal	Indonesia	*Anopheles* (mosquitoes)	Lymphatic tissue	Blood	+
Loa loa	Diurnal	West and Central Africa	*Chrysops* (deerflies)	Subcutaneous tissue	Blood	+
Onchocerca volvulus	None	South and Central America, Africa	*Simulium* (blackflies)	Subcutaneous tissue	Skin, eye	–
Mansonella ozzardi	None	South and Central America	*Culicoides* (midges)	Undetermined site	Blood	–
		Caribbean	*Simulium* (blackflies)			
M. perstans	None	South and Central America, Africa	*Culicoides* (midges)	Body cavities, mesentery, perirenal tissue	Blood	–
M. streptocerca	None	West and Central Africa	*Culicoides* (midges)	Subcutaneous tissue	Skin	–

Usually, infection is established only with repeated, prolonged exposures to infective larvae. Since the clinical manifestations of filarial diseases develop relatively slowly, these infections should be considered to induce chronic diseases with possible long-term debilitating effects. In terms of the nature, severity, and timing of clinical manifestations, patients with filarial infections who are native to endemic areas and have lifelong exposure may differ significantly from those who are travelers or who

have recently moved to these areas. Characteristically, filarial disease is more acute and intense in newly exposed individuals than in natives of endemic areas.

LYMPHATIC FILARIASIS

Lymphatic filariasis is caused by *W. bancrofti*, *B. malayi*, or *B. timori*. The threadlike adult parasites reside in lymphatic channels or lymph nodes, where they may remain viable for more than two decades.

Epidemiology

W. bancrofti, the most widely distributed filarial parasite of humans, affects an estimated 110 million people and is found throughout the tropics and subtropics, including Asia and the Pacific Islands, Africa, areas of South America, and the Caribbean basin. Humans are the only definitive host for the parasite. Generally, the subperiodic form is found only in the Pacific Islands; elsewhere, *W. bancrofti* is nocturnally periodic. (Nocturnally periodic forms of microfilariae are scarce in peripheral blood by day and increase at night, whereas subperiodic forms are present in peripheral blood at all times and reach maximal levels in the afternoon.) Natural vectors for *W. bancrofti* are *Culex fatigans* mosquitoes in urban settings and anopheline or aedean mosquitoes in rural areas.

Brugian filariasis due to *B. malayi* occurs primarily in eastern India, Indonesia, Malaysia, and the Philippines. *B. malayi* also has two forms distinguished by the periodicity of microfilaremia. The more common nocturnal form is transmitted in areas of coastal rice fields, while the subperiodic form is found in forests. *B. malayi* naturally infects cats as well as humans. The distribution of *B. timori* is limited to the islands of southeastern Indonesia.

Pathology

The principal pathologic changes result from inflammatory damage to the lymphatics, which is typically caused by adult worms and not by microfilariae. Adult worms live in afferent lymphatics or sinuses of lymph nodes and cause lymphatic dilatation and thickening of the vessel walls. The infiltration of plasma cells, eosinophils, and macrophages in and around the infected vessels, along with endothelial and connective tissue proliferation, leads to tortuosity of the lymphatics and damaged or incompetent lymph valves. Lymphedema and chronic stasis changes with hard or brawny edema develop in the overlying skin. These consequences of filarial infection are due both to the direct effects of the worms and to the host's inflammatory response to the parasite. Inflammatory responses are believed to cause the granulomatous and proliferative processes that precede total lymphatic obstruction. It is thought that the lymphatic vessel remains patent as long as the worm remains viable and that the death of the worm leads to enhanced granulomatous reaction and fibrosis. Lymphatic obstruction results, and, despite collateralization of the lymphatics, lymphatic function is compromised.

Clinical Features

The most common presentations of the lymphatic filariases are asymptomatic (or subclinical) microfilaremia, hydrocele **(Fig. 218-1)**, acute adenolymphangitis (ADL), and chronic lymphatic disease. In areas where *W. bancrofti* or *B. malayi* is endemic, the overwhelming majority of infected individuals have few overt clinical manifestations of filarial infection despite large numbers of circulating microfilariae in the peripheral blood. Although they may be clinically asymptomatic, virtually all persons with *W. bancrofti* or *B. malayi* microfilaremia have some degree of subclinical disease that includes microscopic hematuria and/or proteinuria, dilated (and tortuous) lymphatics (visualized by imaging), and—in men with *W. bancrofti* infection—scrotal lymphangiectasia (detectable by ultrasound). In spite of these findings, the majority of individuals appear to remain clinically asymptomatic for years; in relatively few does the infection progress to either acute or chronic disease.

Figure 218-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hydrocele associated with *Wuchereria bancrofti* infection.

ADL is characterized by high fever, lymphatic inflammation (lymphangitis and lymphadenitis), and transient local edema. The lymphangitis is retrograde, extending peripherally from the lymph node draining the area where the adult parasites reside. Regional lymph nodes are often enlarged, and the entire lymphatic channel can become indurated and inflamed. Concomitant local thrombophlebitis can occur as well. In brugian filariasis, a single local abscess may form along the involved lymphatic tract and subsequently rupture to the surface. The lymphadenitis and lymphangitis can involve both the upper and lower extremities in both bancroftian and brugian filariasis, but involvement of the genital lymphatics occurs almost exclusively with *W. bancrofti* infection. This genital involvement can be manifested by funiculitis, epididymitis, and scrotal pain and tenderness. In endemic areas, another type of acute disease—dermatolymphangioadenitis (DLA)—is recognized as a syndrome that includes high fever, chills, myalgias, and headache. Edematous inflammatory plaques clearly demarcated from normal skin are seen. Vesicles, ulcers, and hyperpigmentation may also be noted. There is often a history of trauma, burns, radiation, insect bites, punctiform lesions, or chemical injury. Entry lesions, especially in the interdigital area, are common. DLA is often diagnosed as cellulitis.

If lymphatic damage progresses, transient lymphedema can develop into lymphatic obstruction and the permanent changes associated with elephantiasis **(Fig. 218-2)**. Brawny edema follows early pitting edema, and thickening of the subcutaneous tissues and hyperkeratosis occur. Fissuring of the skin develops, as do hyperplastic changes. Superinfection of these poorly vascularized tissues becomes a problem. In bancroftian filariasis, in which genital involvement is common, hydroceles may develop (Fig. 218-1); in advanced stages, this condition may evolve into scrotal lymphedema and scrotal elephantiasis. Furthermore, if there is obstruction of the retroperitoneal lymphatics, increased renal lymphatic pressure leads to rupture of the renal lymphatics and the development of chyluria, which is usually intermittent and most prominent in the morning.

Figure 218-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Elephantiasis of the lower extremity associated with *Wuchereria bancrofti* infection.

The clinical manifestations of filarial infections in travelers or transmigrants who have recently entered an endemic region are distinctive. Given a sufficient number of bites by infected vectors, usually over a 3- to 6-month period, recently exposed patients can develop acute lymphatic or scrotal inflammation with or without urticaria and localized angioedema. Lymphadenitis of epitrochlear, axillary, femoral, or inguinal lymph nodes is often followed by retrogradely evolving lymphangitis. Acute attacks are short-lived and are not usually accompanied by fever. With prolonged exposure to infected mosquitoes, these attacks, if untreated, become more severe and lead to permanent lymphatic inflammation and obstruction.

Diagnosis

A definitive diagnosis can be made only by detection of the parasites and hence can be difficult. Adult worms localized in lymphatic vessels or nodes are largely inaccessible. Microfilariae can be found in blood, in hydrocele fluid, or (occasionally) in other body fluids. Such fluids can be examined microscopically, either directly or—for greater sensitivity—after concentration of the parasites by the passage of fluid through a polycarbonate cylindrical-pore filter (pore size, 3 μm) or by the centrifugation of fluid fixed in 2% formalin (Knott's concentration technique). The timing of blood collection is critical and should be based on the periodicity of the microfilariae in the endemic region involved. Many infected individuals do not have microfilaremia, and definitive diagnosis in such cases can be difficult. Assays for circulating antigens of *W. bancrofti* permit the diagnosis of microfilaremic and cryptic (amicrofilaremic) infection. Two tests are commercially available: an enzyme-linked immunosorbent assay (ELISA) and a rapid-format immunochromatographic card test. Both assays have sensitivities of 93–100% and specificities approaching 100%. There are currently no tests for circulating antigens in brugian filariasis.

Polymerase chain reaction (PCR)-based assays for DNA of *W. bancrofti* and *B. malayi* in blood have been developed. A number of studies indicate that the sensitivity of this diagnostic method is equivalent to or greater than that of parasitologic methods.

In cases of suspected lymphatic filariasis, examination of the scrotum, the lymph nodes, or (in female patients) the breast by means of high-frequency ultrasound in conjunction with Doppler techniques may result in the identification of motile adult worms within dilated lymphatics. Worms may be visualized in the lymphatics of the spermatic cord in up to 80% of men infected with *W. bancrofti*. Live adult worms have a distinctive pattern of movement within the lymphatic vessels (termed the *filaria dance sign*). Radionuclide lymphoscintigraphic imaging of the limbs reliably demonstrates widespread lymphatic abnormalities in both subclinical microfilaremic persons and those with clinical manifestations of lymphatic pathology. While of potential utility in the delineation of anatomic changes associated with infection, lymphoscintigraphy is unlikely to assume primacy in the diagnostic evaluation of individuals with suspected infection; it is principally a research tool, although it has been used more widely for assessment of lymphedema of any cause. Eosinophilia and elevated serum concentrations of IgE and antifilarial antibody support the diagnosis of lymphatic filariasis. There is, however, extensive cross-

reactivity between filarial antigens and antigens of other helminths, including the common intestinal roundworms; thus, interpretations of serologic findings can be difficult. In addition, residents of endemic areas can become sensitized to filarial antigens (and thus be serologically positive) through exposure to infected mosquitoes without having patent filarial infections.

The ADL associated with lymphatic filariasis must be distinguished from thrombophlebitis, infection, and trauma. Retrograde evolution is a characteristic feature that helps distinguish filarial lymphangitis from ascending bacterial lymphangitis. Chronic filarial lymphedema must also be distinguished from the lymphedema of malignancy, postoperative scarring, trauma, chronic edematous states, and congenital lymphatic system abnormalities.

Treatment: Lymphatic Filariasis

With newer definitions of clinical syndromes in lymphatic filariasis and new tools to assess clinical status (e.g., ultrasound, lymphoscintigraphy, circulating filarial antigen assays, PCR), approaches to treatment based on infection status can be considered.

Diethylcarbamazine [(DEC), 6 mg/kg daily for 12 days], which has both macro- and microfilaricidal properties, remains the drug of choice for the treatment of active lymphatic filariasis (defined by microfilaremia, antigen positivity, or adult worms on ultrasound), although albendazole (400 mg twice daily for 21 days) has also demonstrated macrofilaricidal efficacy. A 4- to 6-week course of doxycycline (targeting the intracellular *Wolbachia*) also has significant macrofilaricidal activity, as has DEC/albendazole used daily for 7 days. The addition of DEC to a 3-week course of doxycycline has recently been shown to be efficacious in lymphatic filariasis.

Regimens that combine single doses of albendazole (400 mg) with either DEC (6 mg/kg) or ivermectin (200 μg/kg) all have a sustained microfilaricidal effect and are the mainstay of programs for the eradication of lymphatic filariasis in Africa (albendazole/ivermectin) and elsewhere (albendazole/DEC) (see "Prevention and Control," below).

As has already been mentioned, a growing body of evidence indicates that, although they may be asymptomatic, virtually all persons with *W. bancrofti* or *B. malayi* microfilaremia have some degree of subclinical disease (hematuria, proteinuria, abnormalities on lymphoscintigraphy). Thus, early treatment of asymptomatic persons is recommended to prevent further lymphatic damage. For ADL, supportive treatment (including the administration of antipyretics and analgesics) is recommended, as is antibiotic therapy if secondary bacterial infection is likely. Similarly, because lymphatic disease is associated with the presence of adult worms, treatment with DEC is recommended for microfilaria-negative carriers of adult worms.

In persons with chronic manifestations of lymphatic filariasis, treatment regimens that emphasize hygiene, prevention of secondary bacterial infections, and physiotherapy have gained wide acceptance for morbidity control. These regimens are similar to those recommended for lymphedema of most nonfilarial causes and known by a variety of names, including *complex decongestive physiotherapy* and *complex lymphedema therapy*. Hydroceles (Fig. 218-1) can be managed surgically. With chronic manifestations of lymphatic filariasis, drug treatment should be reserved for individuals with evidence of active infection.

Side effects of DEC treatment include fever, chills, arthralgias, headaches, nausea, and vomiting. Both the development and the severity of these reactions are directly related to the number of microfilariae circulating in the bloodstream. The adverse reactions may represent either an acute hypersensitivity reaction to the antigens being released by dead and dying parasites or an inflammatory reaction induced by the intracellular *Wolbachia* endosymbionts freed from their intracellular niche.

Ivermectin has a side effect profile similar to that of DEC when used in lymphatic filariasis. In patients infected with *L. loa*, who have high levels of *Loa* microfilaremia, DEC—like ivermectin (see "Loiasis," below)—can elicit severe encephalopathic complications. When used in single-dose regimens for the treatment of lymphatic filariasis, albendazole is associated with relatively few side effects.

Prevention and Control

To protect themselves against filarial infection, individuals must avoid contact with infected mosquitoes by using personal protective measures including bed nets, particularly those impregnated with insecticides such as permethrin. Community-based intervention is the current approach to elimination of lymphatic filariasis as a public health problem. The underlying tenet of this approach is that mass annual distribution of antimicrofilarial chemotherapy—albendazole with either DEC (for all areas except those where onchocerciasis is coendemic; see section on onchocerciasis treatment, below) or ivermectin—will profoundly suppress microfilaremia. If the suppression is sustained, then transmission can be interrupted.

Created by the World Health Organization in 1997, the Global Programme to Eliminate Lymphatic Filariasis is based on mass administration of single annual doses of DEC plus albendazole in non-African regions and of albendazole plus ivermectin in Africa. Available information at the end of 2008 indicated that >695 million persons in 51 countries had thus far participated. Not only has lymphatic filariasis been eliminated in some defined areas, but collateral benefits—avoidance of disability and

treatment of intestinal helminths and other conditions (e.g., scabies and louse infestation)—have also been noted. The strategy of the global program is being refined, and attempts are being made to integrate this effort with other mass-treatment strategies (e.g., deworming programs, malaria control, and trachoma control).

TROPICAL PULMONARY EOSINOPHILIA

Tropical pulmonary eosinophilia (TPE) is a distinct syndrome that develops in some individuals infected with the lymphatic-dwelling filarial species. This syndrome affects males and females in a ratio of 4:1, often during the third decade of life. The majority of cases have been reported from India, Pakistan, Sri Lanka, Brazil, Guyana, and Southeast Asia.

Clinical Features

The main features include a history of residence in filarial-endemic regions, paroxysmal cough and wheezing (usually nocturnal and probably related to the nocturnal periodicity of microfilariae), weight loss, low-grade fever, lymphadenopathy, and pronounced blood eosinophilia (>3000 eosinophils/μL). Chest x-rays or CT scans may be normal but generally show increased bronchovascular markings. Diffuse miliary lesions or mottled opacities may be present in the middle and lower lung fields. Tests of pulmonary function show restrictive abnormalities in most cases and obstructive defects in half. Characteristically, total serum IgE levels (10,000–100,000 ng/mL) and antifilarial antibody titers are markedly elevated.

Pathology

In TPE, microfilariae and parasite antigens are rapidly cleared from the bloodstream by the lungs. The clinical symptoms result from allergic and inflammatory reactions elicited by the cleared parasites. In some patients, trapping of microfilariae in other reticuloendothelial organs can cause hepatomegaly, splenomegaly, or lymphadenopathy. A prominent, eosinophil-enriched, intraalveolar infiltrate is often reported, and with it comes the release of cytotoxic proinflammatory eosinophil granule proteins that may mediate some of the pathology seen in TPE. In the absence of successful treatment, interstitial fibrosis can lead to progressive pulmonary damage.

Differential Diagnosis

TPE must be distinguished from asthma, Löffler's syndrome, allergic bronchopulmonary aspergillosis, allergic granulomatosis with angiitis (Churg-Strauss syndrome), the systemic vasculitides [most notably periarteritis nodosa and granulomatosis with polyangiitis (Wegener's)], chronic eosinophilic pneumonia, and the idiopathichypereosinophilic syndrome.

Treatment: Tropical Pulmonary Eosinophilia

DEC is used at a daily dosage of 4–6 mg/kg for 14 days. Symptoms usually resolve within 3–7 days after the initiation of therapy. Relapse, which occurs in ~12–25% of cases (sometimes after an interval of years), requires re-treatment.

ONCHOCERCIASIS

Epidemiology

Onchocerciasis ("river blindness") is caused by the filarial nematode *O. volvulus*, which infects an estimated 37 million individuals in 35 countries worldwide. The majority of individuals infected with *O. volvulus* live in the equatorial region of Africa extending from the Atlantic coast to the Red Sea. In the Americas, isolated foci have been identified in Mexico, Guatemala, Colombia, Ecuador, Venezuela, and Brazil. The infection is also found in Yemen.

Etiology

Infection in humans begins with the deposition of infective larvae on the skin by the bite of an infected blackfly. The larvae develop into adults, which are typically found in subcutaneous nodules. About 7 months to 3 years after infection, the gravid female releases microfilariae that migrate out of the nodule and throughout the tissues, concentrating in the dermis. Infection is transmitted to other persons when a female fly ingests microfilariae from the host's skin and these microfilariae then develop into infective larvae. Adult *O. volvulus* females and males are ~40–60 cm and ~3–6 cm in length, respectively. The life span of adults can be as long as 18 years, with an average of ~9 years. Because the blackfly vector breeds along free-flowing rivers and streams (particularly in rapids) and generally restricts its flight to an area within several kilometers of these breeding sites, both biting and disease transmission are most intense in these locations.

Pathology

Onchocerciasis primarily affects the skin, eyes, and lymph nodes. In contrast to the pathology in lymphatic filariasis, the damage in onchocerciasis is elicited by microfilariae and not by adult parasites. In the skin, there are mild but chronic inflammatory changes that can result in loss of elastic fibers, atrophy, and fibrosis. The subcutaneous nodules, or onchocercomata, consist primarily of fibrous

tissues surrounding the adult worm, often with a peripheral ring of inflammatory cells (characterized as lymphatic in origin) surrounded by an endothelial layer. In the eye, neovascularization and corneal scarring lead to corneal opacities and blindness. Inflammation in the anterior and posterior chambers frequently results in anterior uveitis, chorioretinitis, and optic atrophy. Although punctate opacities are due to an inflammatory reaction surrounding dead or dying microfilariae, the pathogenesis of most manifestations of onchocerciasis is still unclear.

Clinical Features

SKIN

Pruritus and rash are the most common manifestations of onchocerciasis. The pruritus can be incapacitating; the rash is typically a papular eruption **(Fig. 218-3)** that is generalized rather than localized to a particular region of the body. Long-term infection results in exaggerated and premature wrinkling of the skin, loss of elastic fibers, and epidermal atrophy that can lead to loose, redundant skin and hypo- or hyperpigmentation. Localized eczematoid dermatitis can cause hyperkeratosis, scaling, and pigmentary changes. In an immunologically hyperreactive form of onchodermatitis (commonly termed *sowdah* or localized onchodermatitis), the affected skin darkens as a consequence of the profound inflammation that occurs as microfilariae in the skin are cleared.

Figure 218-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Papular eruption as a consequence of onchocerciasis.

ONCHOCERCOMATA

These subcutaneous nodules, which can be palpable and/or visible, contain the adult worm. In African patients, they are common over the coccyx and sacrum, the trochanter of the femur, the lateral anterior crest, and other bony prominences; in patients from South and Central America, nodules tend to develop preferentially in the upper part of the body, particularly on the head, neck, and shoulders. Nodules vary in size and characteristically are firm and not tender. It has been estimated that, for every palpable nodule, there are four deeper nonpalpable ones.

OCULAR TISSUE

Visual impairment is the most serious complication of onchocerciasis and usually affects only those persons with moderate or heavy infections. Lesions may develop in all parts of the eye. The most common early finding is conjunctivitis with photophobia. Punctate keratitis—acute inflammatory reactions surrounding dying microfilariae and manifested as "snowflake" opacities—is common among younger patients and resolves without apparent complications.

Sclerosing keratitis occurs in 1–5% of infected persons and is the leading cause of onchocercal blindness in Africa. Anterior uveitis and iridocyclitis develop in ~5% of infected persons in Africa. In Latin America, complications of the anterior uveal tract (pupillary deformity) may cause secondary glaucoma. Characteristic chorioretinal lesions develop as a result of atrophy and hyperpigmentation of

the retinal pigment epithelium. Constriction of the visual fields and overt optic atrophy may occur.

LYMPH NODES

Mild to moderate lymphadenopathy is common, particularly in the inguinal and femoral areas, where the enlarged nodes may hang down in response to gravity ("hanging groin"), sometimes predisposing to inguinal and femoral hernias.

SYSTEMIC MANIFESTATIONS

Some heavily infected individuals develop cachexia with loss of adipose tissue and muscle mass. Among adults who become blind, there is a three- to fourfold increase in the mortality rate.

Diagnosis

Definitive diagnosis depends on the detection of an adult worm in an excised nodule or, more commonly, of microfilariae in a skin snip. Skin snips are obtained with a corneal-scleral punch, which collects a blood-free skin biopsy sample extending to just below the epidermis, or by lifting of the skin with the tip of a needle and excision of a small (1- to 3-mm) piece with a sterile scalpel blade. The biopsy tissue is incubated in tissue culture medium or in saline on a glass slide or flat-bottomed microtiter plate. After incubation for 2–4 h (or occasionally overnight in light infections), microfilariae emergent from the skin can be seen by low-power microscopy.

Eosinophilia and elevated serum IgE levels are common but, because they occur in many parasitic infections, are not diagnostic in themselves. Assays to detect specific antibodies to *Onchocerca* and PCR to detect onchocercal DNA in skin snips are used in specialized laboratories and are highly sensitive and specific.

Treatment

The main goals of therapy are to prevent the development of irreversible lesions and to alleviate symptoms. Surgical excision is recommended when nodules are located on the head (because of the proximity of microfilaria-producing adult worms to the eye), but chemotherapy is the mainstay of management. Ivermectin, a semisynthetic macrocyclic lactone active against microfilariae, is the first-line agent for the treatment of onchocerciasis. It is given orally in a single dose of 150 μg/kg, either yearly or semiannually. Recently, more frequent ivermectin administration (every 3 months) has been suggested to ameliorate pruritus and skin disease. Moreover, quadrennial administration of ivermectin has some macrofilaricidal activity.

After treatment, most individuals have few or no reactions. Pruritus, cutaneousedema, and/or maculopapular rash occurs in ~1–10% of treated individuals. In areas of Africa coendemic for *O. volvulus* and *L. loa*, however, ivermectin is contraindicated (as it is for pregnant or breast-feeding women) because of severe posttreatment encephalopathy seen in patients, especially children, who are heavily microfilaremic for *L. loa* (>2000–5000 microfilariae/mL). Although ivermectin treatment results in a marked drop in microfilarial density, its effect can be short-lived (<3 months in some cases). Thus, it is occasionally necessary to give ivermectin more frequently for persistent symptoms.

A 6-week course of doxycycline is macrofilaristatic, rendering female adult worms sterile for long periods. Because this agent targets the *Wolbachia* endosymbiont of the filarial parasite, new approaches for definitive treatment (i.e., cure) may become available.

Prevention

Vector control has been beneficial in highly endemic areas in which breeding sites are vulnerable to insecticide spraying, but most areas endemic for onchocerciasis are not suited to this type of control. Community-based administration of ivermectin every 6–12 months is being used to interrupt transmission in endemic areas. This measure, in conjunction with vector control, has already helped reduce the prevalence of disease in endemic foci in Africa and Latin America. No drug has proved useful for prophylaxis of *O. volvulus* infection.

LOIASIS

Etiology and Epidemiology

Loiasis is caused by *L. loa* (the African eye worm), which is present in the rain forests of West and Central Africa. Adult parasites (females, 50–70 mm long and 0.5 mm wide; males, 25–35 mm long and 0.25 mm wide) live in subcutaneous tissues. Microfilariae circulate in the blood with a diurnal periodicity that peaks between 12:00 noon and 2:00 p.m.

Clinical Features

Manifestations of loiasis in natives of endemic areas may differ from those in temporary residents or visitors. Among the indigenous population, loiasis is often an asymptomatic infection with microfilaremia. Infection may be recognized only after subconjunctival migration of an adult worm **(Fig. 218-4)** or may be manifested by episodic Calabar swellings—evanescent localized areas of angioedema and erythema developing on the extremities and less frequently at other sites.

Nephropathy, encephalopathy, and cardiomyopathy can occur but are rare. In patients who are not residents of endemic areas, allergic symptoms predominate, episodes of Calabar swelling tend to be more frequent and debilitating, microfilaremia is less common, and eosinophilia and increased levels of antifilarial antibodies are characteristic.

Figure 218-4

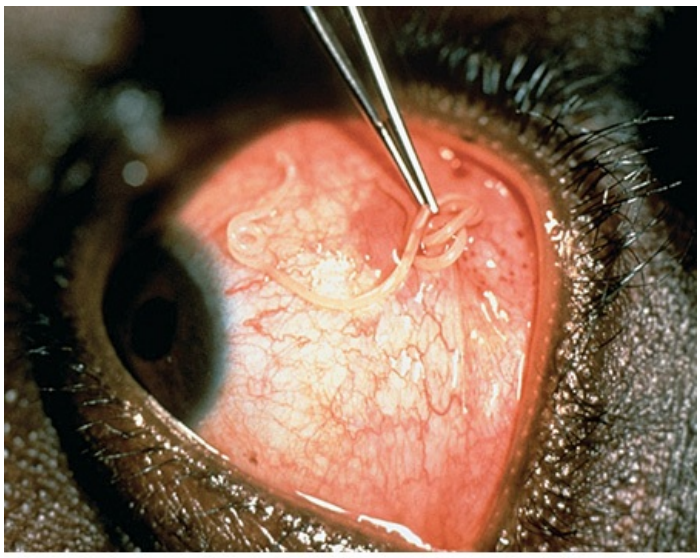

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adult *Loa loa* being surgically removed after its subconjunctival migration.

Pathology

The pathogenesis of the manifestations of loiasis is poorly understood. Calabar swellings are thought to result from a hypersensitivity reaction to adult worm antigens.

Diagnosis

Definitive diagnosis of loiasis requires the detection of microfilariae in the peripheral blood or the isolation of the adult worm from the eye (Fig. 218-4) or from a subcutaneous biopsy specimen from a site of swelling developing after treatment. PCR-based assays for the detection of *L. loa* DNA in blood are available in specialized laboratories and are highly sensitive and specific, as are some newer recombinant antigen–based serologic techniques. In practice, the diagnosis must often be based on a characteristic history and clinical presentation, blood eosinophilia, and elevated levels of antifilarial antibodies, particularly in travelers to an endemic region, who are usually amicrofilaremic. Other clinical findings in the travelers include hypergammaglobulinemia, elevated levels of serum IgE, and elevated leukocyte and eosinophil counts.

Treatment: Loiasis

DEC (8–10 mg/kg per day for 21 days) is effective against both the adult and the microfilarial forms of *L. loa*, but multiple courses are frequently necessary before loiasis resolves completely. In cases of heavy microfilaremia, allergic or other inflammatory reactions can take place during treatment, including central nervous system involvement with coma and encephalitis. Heavy infections can be treated initially with apheresis to remove the microfilariae and with glucocorticoids (40–60 mg of prednisone per day) followed by doses of DEC (0.5 mg/kg per day). If antifilarial treatment has no adverse effects, the prednisone dose can be rapidly tapered and the dose of DEC gradually increased to 8–10 mg/kg per day.

Albendazole or ivermectin is effective in reducing microfilarial loads, although neither is approved for this purpose by the U.S. Food and Drug Administration. Moreover, ivermectin is contraindicated in patients with >5000 microfilariae/mL because this drug has been associated with >1200 deaths in heavily infected patients with loiasis in West and Central Africa. DEC (300 mg weekly) is an effective prophylactic regimen for loiasis.

STREPTOCERCIASIS

Mansonella streptocerca, found mainly in the tropical forest belt of Africa from Ghana to the Democratic Republic of the Congo, is transmitted by biting midges. The major clinical manifestations involve the skin and include pruritus, papular rashes, and pigmentation changes. Many infected individuals have inguinal adenopathy, although most are asymptomatic. The diagnosis is made by

detection of the characteristic microfilariae in skin snips. Ivermectin at a single dose of 150 μg/kg leads to sustained suppression of microfilariae in the skin and is probably the treatment of choice for streptocerciasis.

MANSONELLA PERSTANS INFECTION

M. perstans, distributed across the center of Africa and in northeastern South America, is transmitted by midges. Adult worms reside in serous cavities—pericardial, pleural, and peritoneal—as well as in the mesentery and the perirenal and retroperitoneal tissues. Microfilariae circulate in the blood without periodicity. The clinical and pathologic features of the infection are poorly defined. Most patients appear to be asymptomatic, but manifestations may include transient angioedema and pruritus of the arms, face, or other parts of the body (analogous to the Calabar swellings of loiasis); fever; headache; arthralgias; and right-upper-quadrant pain. Occasionally, pericarditis and hepatitis occur. The diagnosis is based on the demonstration of microfilariae in blood or serosal effusions. Perstans filariasis is often associated with peripheral-blood eosinophilia and antifilarial antibody elevations.

With the identification of a *Wolbachia* endosymbiont in *M. perstans*, doxycycline (200 mg twice a day) for 6 weeks has been established as the first effective treatment for this infection.

MANSONELLA OZZARDI INFECTION

The distribution of *M. ozzardi* is restricted to Central and South America and certain Caribbean islands. Adult worms are rarely recovered from humans. Microfilariae circulate in the blood without periodicity. Although this organism has often been considered nonpathogenic, headache, articular pain, fever, pulmonary symptoms, adenopathy, hepatomegaly, pruritus, and eosinophilia have been ascribed to *M. ozzardi* infection. The diagnosis is made by detection of microfilariae in peripheral blood. Ivermectin (a single dose of 6 mg) is effective in treating this infection.

DRACUNCULIASIS (GUINEA WORM INFECTION)

Etiology and Epidemiology

The incidence of dracunculiasis, caused by *Dracunculus medinensis*, has declined dramatically because of global eradication efforts. Current estimates suggest that there are slightly more than 3000 cases worldwide; the infection is endemic only in Ethiopia, Ghana, Mali, Niger, Nigeria, and Sudan. Asia has now been deemed dracunculiasis-free.

Humans acquire *D. medinensis* when they ingest water containing infective larvae derived from *Cyclops*, a crustacean that is the intermediate host. Larvae penetrate the stomach or intestinal wall, mate, and mature. The adult male probably dies; the female worm develops over a year and migrates to subcutaneous tissues, usually in the lower extremity. As the thin female worm, ranging in length from 30 cm to 1 m, approaches the skin, a blister forms that, over days, breaks down and forms an ulcer. When the blister opens, large numbers of motile, rhabditiform larvae can be released into stagnant water; ingestion by *Cyclops* completes the life cycle.

Clinical Features

Few or no clinical manifestations of dracunculiasis are evident until just before the blister forms, when there is an onset of fever and generalized allergic symptoms, including periorbital edema, wheezing, and urticaria. The emergence of the worm is associated with local pain and swelling. When the blister ruptures (usually as a result of immersion in water) and the adult worm releases larva-rich fluid, symptoms are relieved. The shallow ulcer surrounding the emerging adult worm heals over weeks to months. Such ulcers, however, can become secondarily infected, the result being cellulitis, local inflammation, abscess formation, or (uncommonly) tetanus. Occasionally, the adult worm does not emerge but becomes encapsulated and calcified.

Diagnosis

The diagnosis is based on the findings developing with the emergence of the adult worm, as described above.

Treatment

Gradual extraction of the worm by winding of a few centimeters on a stick each day remains the common and effective practice. Worms may be excised surgically. No drug is effective in treating dracunculiasis.

Prevention

Prevention, which remains the only real control measure, depends on the provision of safe drinking water.

ZOONOTIC FILARIAL INFECTIONS

Dirofilariae that affect primarily dogs, cats, and raccoons occasionally infect humans incidentally, as do

Brugia and *Onchocerca* parasites that affect small mammals. Because humans are an abnormal host, the parasites never develop fully. Pulmonary dirofilarial infection caused by the canine heartworm *Dirofilaria immitis* generally presents in humans as a solitary pulmonary nodule. Chest pain, hemoptysis, and cough are uncommon. Infections with *D. repens* (from dogs) or *D. tenuis* (from raccoons) can cause local subcutaneous nodules in humans. Zoonotic *Brugia* infection can produce isolated lymph node enlargement, whereas zoonotic *Onchocerca* can cause subconjunctival masses. Eosinophilia levels and antifilarial antibody titers are not commonly elevated. Excisional biopsy is both diagnostic and curative. These infections usually do not respond to chemotherapy.

FURTHER READINGS

Coulibaly YI et al: A randomized trial of doxycycline for *Mansonella perstans* infection. N Engl J Med 361:1448, 2009[PMID: 19812401]

Dreyer G et al: Pathogenesis of lymphatic disease in bancroftian filariasis: A clinical perspective. Parasitol Today 16:544, 2000[PMID: 11121854]

Gardon J et al: Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for *Loa loa* infection. Lancet 350:18, 1997[PMID: 9217715]

Gonzalez RJ et al: Successful interruption of transmission of *Onchocerca volvulus* in the Escuintla-Guatemala focus, Guatemala. PLoS Negl Trop Dis 3:e404, 2009

Hoerauf A: Filariasis: New drugs and new opportunities for lymphatic filariasis and onchocerciasis. Curr Opin Infect Dis 21:673, 2008[PMID: 18978537]

——— et al: Depletion of *Wolbachia* endobacteria in *Onchocerca volvulus* by doxycycline and microfilaridermia after ivermectin treatment. Lancet 357:1415, 2001

Hooper PJ et al: The Global Programme to Eliminate Lymphatic Filariasis: Health impact during its first 8 years (2000–2007). Ann Trop Med Parasitol 103(Suppl 1):S17, 2009

Ottesen EA: Lymphatic filariasis: Treatment, control and elimination. Adv Parasitol 61:395, 2006[PMID: 16735170]

Stingl P: Onchocerciasis: Developments in diagnosis, treatment and control. Int J Dermatol 48:393, 2009[PMID: 19335425]

Taylor MJ et al: Lymphatic filariasis and onchocerciasis. Lancet 376:1175, 2010[PMID: 20739055]

——— et al: Macrofilaricidal activity after doxycycline treatment of *Wuchereria bancrofti*: A double-blind, randomised placebo-controlled trial. Lancet 365:2116, 2005

Vijayan VK: Tropical pulmonary eosinophilia: Pathogenesis, diagnosis and management. Curr Opin Pulm Med 13:428, 2007[PMID: 17940489]

Weil GJ, Ramzy RM: Diagnostic tools for filariasis elimination programs. Trends Parasitol 23:78, 2007[PMID: 17174604]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SCHISTOSOMIASIS AND OTHER TREMATODE INFECTIONS: INTRODUCTION

Trematodes, or flatworms, are a group of morphologically and biologically heterogeneous organisms that belong to the phylum Platyhelminthes. Human infection with trematodes occurs in many geographic areas and can cause considerable morbidity and mortality. The dependence on one drug—praziquantel—for treatment of most infections caused by helminths, including trematodes, raises the specter of developing resistance in these worms; several instances of reduced drug efficacy have already been reported.

ETIOLOGIC AGENTS AND THEIR LIFE CYCLES

For clinical purposes, significant trematode infections of humans may be divided according to tissues invaded by adult flukes: blood, biliary tree, intestines, and lungs **(Table 219-1)**. Trematodes share some common morphologic features, including macroscopic size (from one to several centimeters); dorsoventral, flattened, bilaterally symmetric bodies (adult worms); and the prominence of two suckers. Except for schistosomes, all human parasitic trematodes are hermaphroditic. Their life cycles involve a definitive host (mammalian/human), in which adult worms initiate sexual reproduction, and an intermediate host (snails), in which asexual multiplication of larvae occurs. More than one intermediate host may be necessary for some species of trematodes. Human infection is initiated either by direct penetration of intact skin or by ingestion. Upon maturation within humans, adult flukes initiate sexual reproduction and egg production. Helminth ova leave the definitive host in excreta or sputum and, upon reaching suitable environmental conditions, they hatch, releasing free-living miracidia that seek specific snail intermediate hosts. After asexual reproduction, cercariae are released from infected snails. In certain species, these organisms infect humans; in others, they find a second intermediate host to allow encystment into metacercariae—the infective stage.

Table 219-1 Major Human Trematode Infections

Trematode	Transmission	Endemic Area(s)
Blood Flukes		
Schistosoma mansoni	Skin penetration by cercariae released from snails	Africa, South America, Middle East
S. japonicum	Skin penetration by cercariae released from snails	China, Philippines, Indonesia
S. intercalatum	Skin penetration by cercariae released from snails	West Africa
S. mekongi	Skin penetration by cercariae released from snails	Southeast Asia
S. haematobium	Skin penetration by cercariae released from snails	Africa, Middle East
Biliary (Hepatic) Flukes		
Clonorchis sinensis	Ingestion of metacercariae in freshwater fish	Far East
Opisthorchis viverrini	Ingestion of metacercariae in freshwater fish	Far East, Thailand
O. felineus	Ingestion of metacercariae in freshwater fish	Far East, Europe
Fasciola hepatica	Ingestion of metacercariae on aquatic plants or in water	Worldwide
F. gigantica	Ingestion of metacercariae on aquatic plants or in water	Sporadic, Africa
Intestinal Flukes		
Fasciolopsis buski	Ingestion of metacercariae on aquatic plants	Southeast Asia
Heterophyes heterophyes	Ingestion of metacercariae in freshwater or brackish-water fish	Far East, North Africa
Lung Flukes		

Paragonimus westermani	Ingestion of metacercariae in crayfish or crabs	Global except North America and Europe

The host-parasite relationship in trematode infections is a product of certain biologic features of these organisms: they are multicellular, undergo several developmental changes within the host, and usually result in chronic infections. In general, the distribution of worm infections in human populations is *overdispersed*; i.e., it follows a negative binomial mathematical relationship in which most infected individuals harbor low worm burdens while a small percentage are heavily infected. It is the heavily infected minority who are particularly prone to disease sequelae and who constitute an epidemiologically significant reservoir of infection in endemic areas. Equally important is an appreciation that worms do not multiply within the definitive host and that they have a relatively long life span, ranging from a few months to a few years. Morbidity and death due to trematode infections reflect a multifactorial process that results from the tipping of a delicate balance between intensity of infection and host reactions, which initiate and modulate immunologic and pathologic outcome. Furthermore, the genetics of the parasite and of the human host contribute to the outcome of infection and disease. Infections with trematodes that migrate through or reside in host tissues are associated with a moderate to high degree of peripheral-blood eosinophilia; this association is of significance in protective and immunopathologic sequelae and is a useful clinical indicator of infection.

Approach to the Patient: Trematode Infection

The approach to individuals with suspected trematode infection begins with a question: Where have you been? Details of geographic history, exposure to freshwater bodies, and indulgence in local eating habits without ensuring safety of food and drink are all essential elements eliciting the history of present illness. The workup plan must include a detailed physical examination and tests appropriate for suspected infection. Diagnosis is based either on detection of the relevant stage of the parasite in excreta, sputum, or (rarely) tissue samples or on sensitive and specific serologic tests. Consultation with physicians familiar with these infections or with the U.S. Centers for Disease Control and Prevention (CDC) is helpful in guiding diagnosis and selecting therapy.

BLOOD FLUKES: SCHISTOSOMIASIS

Human schistosomiasis is caused by five species of the parasitic trematode genus *Schistosoma*: the intestinal species *S. mansoni*, *S. japonicum*, *S. mekongi*, and *S. intercalatum* and the urinary species *S. haematobium*. Infection may cause considerable morbidity in the intestines, liver, and urinary tract, and a proportion of affected individuals die. Other schistosomes (e.g., avian species) may invade human skin but then die in subcutaneous tissue, producing only self-limiting cutaneous manifestations.

Etiology

Human infection is initiated by penetration of intact skin with infective cercariae. These organisms, which are released from infected snails in freshwater bodies, measure ~2 mm in length and possess an anterior and a ventral sucker that attach to the skin and facilitate penetration. Once in subcutaneous tissue, cercariae transform into schistosomula, with morphologic, membrane, and immunologic changes. The cercarial outer membrane changes from a trilaminar to a heptalaminar structure that is then maintained throughout the organism's life span in humans. This transformation is thought to be the schistosome's main adaptive mechanism for survival in humans. Schistosomula begin their migration within 2–4 days via venous or lymphatic vessels, reaching the lungs and finally the liver parenchyma. Sexually mature worms descend into the venous system at specific anatomic locations: intestinal veins (*S. mansoni*, *S. japonicum*, *S. mekongi*, and *S. intercalatum*) and vesical veins (*S. haematobium*). After mating, adult gravid females travel against venous blood flow to small tributaries, where they deposit their ova intravascularly. Schistosome ova **(Fig. 219-1)** have specific morphologic features that vary with the species. Aided by enzymatic secretions through minipores in eggshells, ova move through the venous wall, traversing host tissues to reach the lumen of the intestinal or urinary tract, and are voided with stools or urine. Approximately 50% of ova are retained in host tissues locally (intestines or urinary tract) or are carried by venous blood flow to the liver and other organs. Schistosome ova that reach freshwater bodies hatch, releasing free-living miracidia that seek the snail intermediate host and undergo several cycles of asexual multiplication. Finally, infective cercariae are shed from snails.

Figure 219-1

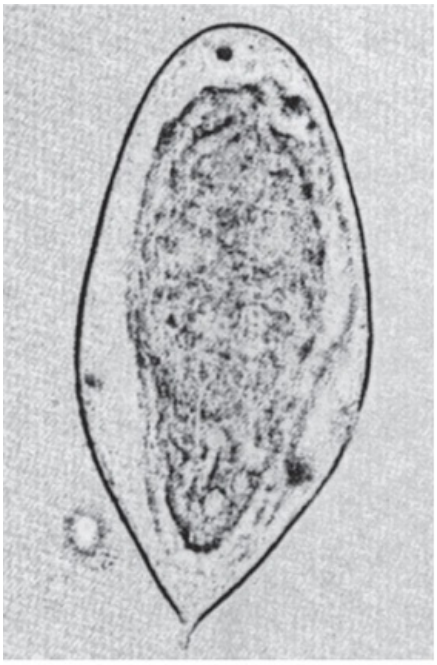

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

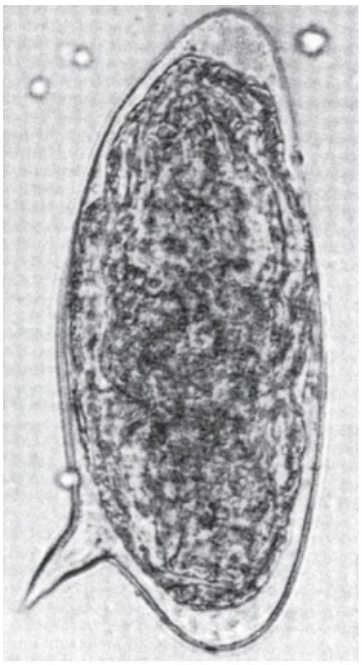

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

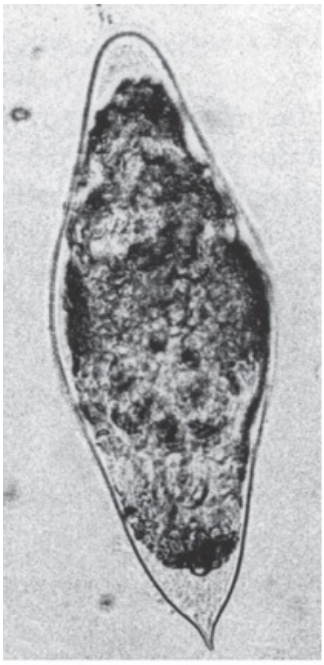

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Morphology of schistosome eggs, the diagnostic stage of the parasite's life cycle. ***A.*** *S. haematobium* egg (in a urine sample) is large (~140 mm long), with a terminal spine. ***B.*** *S. mansoni* egg (in a fecal sample) is large (~150 mm long), with a thin shell and lateral spine. ***C.*** *S. japonicum* egg (fecal) is smaller than that of *S. mansoni* (~90 mm long), with a small spine or hooklike structure. ***D.*** *S. mekongi* egg (fecal) is similar to that of *S. japonicum* but smaller (~65 mm long). *E. S. intercalatum* egg (fecal) is larger than that of *S. haematobium* (~190 mm long), with a longer, sharply pointed spine. (*From LR Ash, TC Orihel: Atlas of Human Parasitology, 3rd ed. Chicago, ASCP Press, 1990; with permission.*)

Adult schistosomes are ~1–2 cm long. Males are slightly shorter than females, with flattened bodies and anteriorly curved edges forming the gynecophoral canal, in which mature adult females are usually held. Females are longer, slender, and rounded in cross-section. The precise nature of biochemical and reproductive exchanges between the two sexes is unknown, as are the regulatory mechanisms for pairing. Adult schistosomes parasitize specific sites in the host venous system. What guides adult intestinal schistosomes to branches of the superior or inferior mesenteric veins or adult *S. haematobium* worms to the vesical plexus is unknown. In addition, adult worms inhibit the coagulation cascade and evade the effector arms of the host immune responses by still-undetermined mechanisms. The genome of schistosomes is relatively large (~270 Mb) and is arrayed on seven pairs of autosomes and one pair of sex chromosomes. Sequencing of the *S. japonicum* and *S. mansoni* genomes has provided the first insight into the worms' genomic and proteomic features, offering an opportunity to discover new drug targets and to understand the molecular basis of pathogenesis.

Epidemiology

The global distribution of schistosome infection in human populations **(Fig. 219-2)** is dependent on both parasite and host factors. Information on prevalence and global distribution is inexact. The five *Schistosoma* species are estimated to infect 200–300 million individuals in South America, the Caribbean, Africa, the Middle East, and Southeast Asia. The total population living under conditions favoring transmission approximates double or triple that number—a fact reflecting the public health significance of schistosomiasis.

Figure 219-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

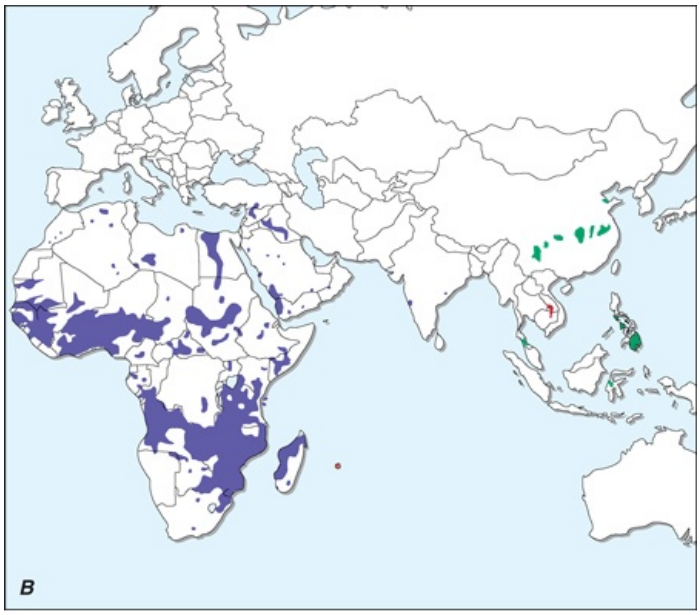

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Global distribution of schistosomiasis. ***A.*** *S. mansoni* infection (*dark blue*) is endemic in Africa, the Middle East, South America, and a few Caribbean countries. *S. intercalatum* infection (*green*) is endemic in sporadic foci in West and Central Africa. ***B.*** *S. haematobium* infection (*purple*) is endemic in Africa and the Middle East. The major endemic countries for *S. japonicum* infection (*green*) are China, the Philippines, and Indonesia. *S. mekongi* infection (*red*) is endemic in sporadic foci in Southeast Asia.

In endemic areas, the rate of yearly onset of new infection, or incidence, is generally low. Prevalence, on the other hand, starts to be appreciable by the age of 3–4 years and builds to a maximum that varies by endemic region (up to 100%) in the 15- to 20-year age group. Prevalence then stabilizes or decreases slightly in older age groups (>40 years). Intensity of infection (as measured by fecal or urinary egg counts, which correlate with adult worm burdens in most circumstances) follows the increase in prevalence up to the age of 15–20 years and then declines markedly in older age groups. This decline may reflect acquisition of resistance or may be due to changes in water contact patterns, since older people have less exposure. Infection with schistosomes in human populations has a peculiar pattern. Most infected individuals harbor low worm burdens, and only a small proportion suffer from high-intensity infection. This pattern may be due to differences in worm infectivity or to a spectrum of genetic susceptibilities in human populations.

Disease due to schistosome infection is the outcome of parasitologic, host, and associated viral infections or nutritional and environmental factors. Most disease syndromes relate to the presence of one or more of the parasite stages in humans. Disease manifestations in the populations of endemic areas correlate, in general, with intensity and duration of infection as well as with age and genetic susceptibility of the host. Overall, disease manifestations are clinically relevant in only a small proportion of persons infected with any of the intestinal schistosomes. In contrast, urinary schistosomiasis manifests clinically in most infected individuals. Estimates of total morbidity due to chronic schistosomiasis indicate a significantly greater burden than was previously appreciated.

Patients with both HIV infection and schistosomiasis excrete far fewer eggs in their stools than those infected with *S. mansoni* alone; the mechanism underlying this difference is unknown. Treatment with praziquantel may result in reduced HIV replication and increased CD4+ T lymphocyte counts.

Pathogenesis and Immunity

Cercarial invasion is associated with dermatitis arising from dermal and subdermal inflammatory responses, both humoral and cell-mediated. As the parasites approach sexual maturity in the liver of infected individuals and as oviposition commences, acute schistosomiasis or Katayama fever (a serum sickness–like illness; see "Clinical Features," below) may occur. The associated antigen excess results in formation of soluble immune complexes, which may be deposited in several tissues, initiating multiple pathologic events. In chronic schistosomiasis, most disease manifestations are due to eggs retained in host tissues. The granulomatous response around these ova is cell-mediated and is regulated both positively and negatively by a cascade of cytokine, cellular, and humoral responses. Granuloma formation begins with recruitment of a host of inflammatory cells in response to antigens secreted by the living organism within the ova. Cells recruited initially include phagocytes, antigen-specific T cells, and eosinophils. Fibroblasts, giant cells, and B lymphocytes predominate later. These lesions reach a size many times that of parasite eggs, thus inducing organomegaly and obstruction. Immunomodulation or downregulation of host responses to schistosome eggs plays a significant role in limiting the extent of the granulomatous lesions—and consequently disease—in chronically infected experimental animals or humans. The underlying mechanisms involve another cascade of regulatory cytokines and idiotypic antibodies. Subsequent to the granulomatous response, fibrosis sets in, resulting in more permanent disease sequelae. Because schistosomiasis is also a chronic infection, the accumulation of antigen-antibody complexes results in deposits in renal glomeruli and may cause significant kidney disease.

The better-studied pathologic sequelae in schistosomiasis are those observed in liver disease. Ova that are carried by portal blood embolize to the liver. Because of their size (~150 x 60 μm in the case of *S. mansoni*), they lodge at presinusoidal sites, where granulomas are formed. These granulomas contribute to the hepatomegaly observed in infected individuals (Fig. 219-3). Schistosomal liver enlargement is also associated with certain class I and class II human leukocyte antigen (HLA)haplotypes and markers; its genetic basis appears to be multigenic. Presinusoidal portal blockage causes several hemodynamic changes, including portal hypertension and associated development of portosystemic collaterals at the esophagogastric junction and other sites. Esophageal varices are most likely to break and cause repeated episodes of hematemesis. Because changes in hepatic portal blood flow occur slowly, compensatory arterialization of the blood flow through the liver is established. While this compensatory mechanism may be associated with certain metabolic side effects, retention of hepatocyte perfusion permits maintenance of normal liver function for several years.

Figure 219-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic hepatosplenomegaly caused by schistosomiasis mansoni. Liver and spleen enlargement, ascites, and wasting are characteristically seen in patients with chronic *S. mansoni* infection.

The second most significant pathologic change in the liver relates to fibrosis. It is characteristically periportal (Symmers' clay pipe–stem fibrosis) but may be diffuse. Fibrosis, when diffuse, may be seen in areas of egg deposition and granuloma formation but is also seen in distant locations such as portal tracts. Schistosomiasis results in pure fibrotic lesions in the liver; cirrhosis occurs when other nutritional factors or infectious agents (e.g., hepatitis B or C virus) are involved. Deposition of fibrotic tissue in the extracellular matrix results from the interaction of T lymphocytes with cells of the fibroblast series; several cytokines, such as interleukin (IL) 2, IL-4, IL-1, and transforming growth factor β (TGF-β), are known to stimulate fibrogenesis. The process may be dependent on the genetic constitution of the host. Furthermore, regulatory cytokines that can suppress fibrogenesis, such as interferon γ (IFN-γ) or IL-12, may play a role in modulating the response.

While the above description focuses on granuloma formation and fibrosis of the liver, similar processes occur in urinary schistosomiasis. Granuloma formation at the lower end of the ureters obstructs urinary flow, with subsequent development of hydroureter and hydronephrosis. Similar lesions in the urinary bladder cause the protrusion of papillomatous structures into its cavity; these may ulcerate and/or bleed. The chronic stage of infection is associated with scarring and deposition of calcium in the bladder wall.

Studies on immunity to schistosomiasis, whether innate or adaptive, have expanded our knowledge of the components of these responses and target antigens. The critical question, however, is whether humans acquire immunity to schistosomes. Epidemiologic data suggest the onset of acquired immunity during the course of infection in young adults. Curative treatment of infected populations in endemic areas is followed by differentiation in the pattern of reinfection. Some (susceptible) individuals acquire reinfection rapidly, whereas other (resistant) individuals are reinfected slowly. This difference may be explained by differences in transmission, immunologic response, or genetic susceptibility. The mechanism of acquired immunity involves antibodies, complement, and several effector cells, particularly eosinophils. Furthermore, the intensity of schistosome infection has been correlated with a region in chromosome 5. In several studies, a few protective schistosome antigens have been identified as vaccine candidates, but none has been evaluated in human populations to date.

Clinical Features

In general, disease manifestations of schistosomiasis occur in three stages, which vary not only by species but also by intensity of infection and other host factors, such as age and genetics. During the phase of cercarial invasion, a form of dermatitis may be observed. This so-called swimmers' itch occurs most often with *S. mansoni* and *S. japonicum* infections, manifesting 2 or 3 days after invasion as an itchy maculopapular rash on the affected areas of the skin. The condition is particularly severe when humans are exposed to avian schistosomes. This form of cercarial dermatitis is also seen around freshwater lakes in the northern United States, particularly in the spring. Cercarial dermatitis is a self-limiting clinical entity. During worm maturation and at the beginning of oviposition (i.e., 4–8 weeks after skin invasion), acute schistosomiasis or Katayama fever—a serum sickness–like syndrome with fever, generalized lymphadenopathy, and hepatosplenomegaly—may develop. Individuals with acute schistosomiasis show a high degree of peripheral-blood eosinophilia. Parasite-specific antibodies may be detected before schistosome eggs are identified in excreta.

Acute schistosomiasis has become an important clinical entity worldwide because of increased travel to endemic areas. Travelers are exposed to parasites while swimming or wading in freshwater bodies and upon their return present with acute manifestations. The course of acute schistosomiasis is generally benign, but deaths are occasionally reported in association with heavy exposure to schistosomes.

The main clinical manifestations of chronic schistosomiasis are species-dependent. Intestinal species (*S. mansoni*, *S. japonicum*, *S. mekongi*, and *S. intercalatum*) cause intestinal and hepatosplenic disease as well as several manifestations associated with portal hypertension. During the intestinal phase, which may begin a few months after infection and may last for years, symptomatic patients characteristically have colicky abdominal pain, bloody diarrhea, and anemia. Patients may also report fatigue and an inability to perform daily routine functions and may show evidence of growth retardation. Schistosomiasis morbidity is generally underappreciated. The severity of intestinal schistosomiasis is often related to the intensity of the worm burden. The disease runs a chronic course and may result in colonic polyposis, which has been reported from some endemic areas, such as Egypt.

The hepatosplenic phase of disease manifests early (during the first year of infection, particularly in children) with liver enlargement due to parasite-induced granulomatous lesions. Hepatomegaly is seen in ~15–20% of infected individuals; it correlates roughly with intensity of infection, occurs more often in children, and may be related to specific HLAhaplotypes. In subsequent phases of infection, presinusoidal blockage of blood flow leads to portal hypertension and splenomegaly **(Fig. 219-3)**. Moreover, portal hypertension may lead to varices at the lower end of the esophagus and at other

sites. Patients with schistosomal liver disease may have right-upper-quadrant "dragging" pain during the hepatomegaly phase, and this pain may move to the left upper quadrant as splenomegaly progresses. Bleeding from esophageal varices may, however, be the first clinical manifestation of this phase. Patients may experience repeated bleeding but seem to tolerate its impact, since an adequate total hepatic blood flow permits normal liver function for a considerable duration. In late-stage disease, typical fibrotic changes occur along with liver function deterioration and the onset of ascites, hypoalbuminemia, and defects in coagulation. Intercurrent viral infections of the liver (especially hepatitis B and C) or nutritional deficiencies may well accelerate or exacerbate the deterioration of hepatic function.

The extent and severity of intestinal and hepatic disease in schistosomiasis mansoni and japonica have been well described. While it was originally thought that *S. japonicum* might induce more severe disease manifestations because the adult worms can produce 10 times more eggs than *S. mansoni*, subsequent field studies have not supported this claim. Clinical observations of individuals infected with *S. mekongi* or *S. intercalatum* have been less detailed, partly because of the limited geographic distribution of these organisms.

The clinical manifestations of *S. haematobium* infection occur relatively early and involve a high percentage of infected individuals. Up to 80% of children infected with *S. haematobium* have dysuria, frequency, and hematuria, which may be terminal. Urine examination reveals blood and albumin as well as an unusually high frequency of bacterial urinary tract infection and urinary sediment cellular metaplasia. These manifestations correlate with intensity of infection, the presence of urinary bladder granulomas, and subsequent ulceration. Along with local effects of granuloma formation in the urinary bladder, obstruction of the lower end of the ureters results in hydroureter and hydronephrosis, which may be seen in 25–50% of infected children. As infection progresses, bladder granulomas undergo fibrosis, which results in typical sandy patches visible on cystoscopy. In many endemic areas, an association between squamous cell carcinoma of the bladder and *S. haematobium* infection has been observed. Such malignancy is detected in a younger age group than is transitional cell carcinoma. In fact, *S. haematobium* has now been classified as a human carcinogen.

Significant disease may occur in other organs during chronic schistosomiasis. Lung and central nervous system (CNS) disease have been documented; other sites, such as the skin and the genital organs, are far less frequently affected. In pulmonary schistosomiasis, embolized eggs lodge in small arterioles, producing acute necrotizing arteriolitis and granuloma formation. During *S. mansoni* and *S. japonicum* infection, schistosome eggs reach the lungs after the development of portosystemic collateral circulation; in *S. haematobium* infection, ova may reach the lungs directly via connections between the vesical and systemic circulation. Subsequent fibrous tissue deposition leads to endarteritis obliterans, pulmonary hypertension, and cor pulmonale. The most common symptoms are cough, fever, and dyspnea. Cor pulmonale may be diagnosed radiologically on the basis of prominence of the right side of the heart and dilation of the pulmonary artery. Frank evidence of right-sided heart failure may be seen in late cases.

Although less common than pulmonary manifestations, CNS schistosomiasis is important, characteristically occurring in association with *S. japonicum* infection. Migratory worms deposit eggs in the brain and induce a granulomatous response. The frequency of this manifestation among infected individuals in some endemic areas (e.g., the Philippines) is calculated at 2–4%. Jacksonian epilepsy due to *S. japonicum* infection is the second most common cause of epilepsy in these areas. *S. mansoni* and *S. haematobium* infections have been associated with transverse myelitis. This syndrome is thought to be due to eggs traveling to the venous plexus around the spinal cord. In schistosomiasis mansoni, transverse myelitis is usually seen in the chronic stage after the development of portal hypertension and portosystemic shunts, which allow ova to travel to the spinal cord veins. This proposed sequence of events has been challenged because of a few reports of transverse myelitis occurring early in the course of *S. mansoni* infection. More information is needed to confirm these observations. During schistosomiasis haematobia, ova may travel through communication between vesical and systemic veins, resulting in spinal cord disease that may be detected at any stage of infection. Pathologic study of lesions in schistosomal transverse myelitis may reveal eggs along with necrotic or granulomatous lesions. Patients usually present with acute or rapidly progressing lower-leg weakness accompanied by sphincter dysfunction.

Diagnosis

Physicians in areas not endemic for schistosomiasis face considerable diagnostic challenges. In the most common clinical presentation, a traveler returns with symptoms and signs of acute syndromes of schistosomiasis—namely, cercarial dermatitis or Katayama fever. Central to a correct diagnosis is a thorough inquiry into the patient's history of travel and exposure to freshwater bodies—whether slow- or fast-running—in a nonendemic area. Differential diagnosis of fever in returned travelers includes a spectrum of infections whose etiologies are viral (e.g., Dengue fever), bacterial (e.g., enteric fever, leptospirosis), rickettsial, or protozoal (e.g., malaria). In cases of Katayama fever, prompt diagnosis is essential and is based on clinical presentation, high-level peripheral-blood eosinophilia, and a positive serologic assay for schistosomal antibodies. Two tests are available at the CDC: the Falcon assay

screening test/enzyme-linked immunosorbent assay (FAST-ELISA) and the confirmatory enzyme-linked immunoelectrotransfer blot (EITB). Both tests are highly sensitive and ~96% specific. In some instances, examination of stool or urine for ova may yield positive results.

Individuals with established infection are diagnosed by a combination of geographic history, characteristic clinical presentation, and presence of schistosome ova in excreta. The diagnosis may also be established with the serologic assays mentioned above or with those that detect circulating schistosome antigens. These assays can be applied either to blood or to other body fluids (e.g., cerebrospinal fluid). For suspected schistosome infection, stool examination by the Kato thick smear or any other concentration method generally identifies all but the most lightly infected individuals. For *S. haematobium*, urine may be examined by microscopy of sediment or by filtration of a known volume through Nuclepore filters. Kato thick smear and Nuclepore filtration provide quantitative data on the intensity of infection, which is of value in assessing the degree of tissue damage and in monitoring the effect of chemotherapy. Schistosome infection may also be diagnosed by examination of tissue samples, typically rectal biopsies; other biopsyprocedures (e.g., liver biopsy) are not needed, except in rare circumstances.

The differential diagnosis of schistosomal hepatomegaly must include viral hepatitis of all etiologies, miliary tuberculosis, malaria, visceral leishmaniasis, ethanol abuse, and causes of hepatic and portal vein obstruction. The differential diagnosis of hematuria in *S. haematobium* infection includes bacterial cystitis, tuberculosis, urinary stones, and malignancy.

Treatment: Schistosomiasis

Treatment of schistosomiasis depends on the stage of infection and the clinical presentation. Other than topical dermatologic applications for relief of itching, no specific treatment is indicated for cercarial dermatitis caused by avian schistosomes. Therapy for acute schistosomiasis or Katayama fever needs to be adjusted appropriately for each case. While antischistosomal chemotherapy may be used, it does not have a significant impact on maturing worms. In severe acute schistosomiasis, management in an acute-care setting is necessary, with supportive measures and consideration of glucocorticoid treatment. Once the acute critical phase is over, specific chemotherapy is indicated for parasite elimination. For all individuals with established infection, treatment to eradicate the parasite should be administered. The drug of choice is praziquantel, which—depending on the infecting species **(Table 219-2)**—is administered PO as a total of 40 or 60 mg/kg in two or three doses over a single day. Praziquantel treatment results in parasitologic cure in ~85% of cases and reduces egg counts by >90%. Few side effects have been encountered, and those that do develop usually do not interfere with completion of treatment. Dependence on a single chemotherapeutic agent has raised the possibility of development of resistance in schistosomes; to date, such resistance does not seem to be clinically significant. The effect of antischistosomal treatment on disease manifestations varies by stage. Early hepatomegaly and bladder lesions are known to resolve after chemotherapy, but the late established manifestations, such as fibrosis, do not recede. Additional management modalities are needed for individuals with other manifestations, such as hepatocellular failure or recurrent hematemesis. The use of these interventions is guided by general medical and surgical principles.

Table 219-2 Drug Therapy for Human Trematode Infections

Infection	Drug of Choice	Adult Dose and Duration
Blood Flukes		
S. mansoni, S. intercalatum, S. haematobium	Praziquantel	20 mg/kg, 2 doses in 1 day
S. japonicum, S. mekongi	Praziquantel	20 mg/kg, 3 doses in 1 day
Biliary (Hepatic) Flukes		
C. sinensis, O. viverrini, O. felineus	Praziquantel	25 mg/kg, 3 doses in 1 day
F. hepatica, F. gigantica	Triclabendazole	10 mg/kg once
Intestinal Flukes		
F. buski, H. heterophyes	Praziquantel	25 mg/kg, 3 doses in 1 day
Lung Flukes		
P. westermani	Praziquantel	25 mg/kg, 3 doses per day for 2 days

Prevention and Control

Transmission of schistosomiasis is dependent on human behavior. Since the geographic distribution of infections in endemic regions of the world is not clearly demarcated, it is prudent for travelers to endemic areas to avoid contact with all freshwater bodies, irrespective of the speed of water flow or unsubstantiated claims of safety. Some topical agents, when applied to skin, may inhibit cercarial penetration, but none is currently available. If exposure occurs, a follow-up visit with a health care

provider is strongly recommended. Prevention of infection in inhabitants of endemic areas is a significant challenge. Residents of these regions use freshwater bodies for sanitary, domestic, recreational, and agricultural purposes. Several control measures have been used, including application of molluscicides, provision of sanitary water and sewage disposal, chemotherapy, and health education. Current recommendations to countries endemic for schistosomiasis emphasize the use of multiple approaches. With the advent of an oral, safe, and effective antischistosomal agent, chemotherapy has been most successful in reducing the intensity of infection and reversing disease. The duration of this positive impact depends on the transmission dynamics of the parasite in any specific endemic region. The ultimate goal of research on prevention and control is the development of a vaccine. Although there are a few promising leads, this goal is probably not within reach during the next decade or so.

LIVER (BILIARY) FLUKES

Several species of biliary fluke infecting humans are particularly common in Southeast Asia and Russia. Other species are transmitted in Europe, Africa, and the Americas. On the basis of their migratory pathway in humans, these infections may be divided into the *Clonorchis* and *Fasciola* groups (Table 219-1).

Clonorchiasis and Opisthorchiasis

Infection with *Clonorchis sinensis*, the Chinese or oriental fluke, is endemic among fish-eating mammals in Southeast Asia. Humans are an incidental host; the prevalence of human infection is highest in China, Vietnam, and Korea. Infection with *Opisthorchis viverrini* and *O. felineus* is zoonotic in cats and dogs. Transmission to humans occurs occasionally, particularly in Thailand (*O. viverrini*) and in Southeast Asia and eastern Europe (*O. felineus*). Data on the exact geographic distribution of these infectious agents in human populations are rudimentary.

Infection with any of these three species is established by ingestion of raw or inadequately cooked freshwater fish harboring metacercariae. These organisms excyst in the duodenum, releasing larvae that travel through the ampulla of Vater and mature into adult worms in bile canaliculi. Mature flukes are flat and elongated, measuring 1–2 cm in length. The hermaphroditic worms reproduce by releasing small operculated eggs, which pass with bile into the intestines and are voided with stools. The life cycle is completed in the environment in specific freshwater snails (the first intermediate host) and encystment of metacercariae in freshwater fish.

Except for late sequelae, the exact clinical syndromes caused by clonorchiasis and opisthorchiasis are not well defined. Since most infected individuals harbor a low worm burden, many are asymptomatic. Moderate to heavy infection may be associated with vague right-upper-quadrant pain. In contrast, chronic or repeated infection is associated with manifestations such as cholangitis, cholangiohepatitis, and biliary obstruction. Cholangiocarcinoma is epidemiologically related to *C. sinensis* infection in China and to *O. viverrini* infection in northeastern Thailand. This association has resulted in classification of these infectious agents as human carcinogens.

Fascioliasis

Infections with *Fasciola hepatica* and *F. gigantica* are worldwide zoonoses that are particularly endemic in sheep-raising countries. Human cases have been reported in South America, Europe, Africa, Australia, and the Far East. Recent estimates indicate a worldwide prevalence of 17 million cases. High endemicity has been reported in certain areas of Peru and Bolivia. In most endemic areas the predominant species is *F. hepatica*, but in Asia and Africa a varying degree of overlap with *F. gigantica* has been observed.

Humans acquire fascioliasis by ingestion of metacercariae attached to certain aquatic plants, such as watercress. Infection may also be acquired by consumption of contaminated water or ingestion of food items washed with such water. Acquisition of human infection through consumption of freshly prepared raw liver containing immature flukes has been reported. Infection is initiated when metacercariae excyst, penetrate the gut wall, and travel through the peritoneal cavity to invade the liver capsule. Adult worms finally reach bile ducts, where they produce large operculated eggs, which are voided in bile through the gastrointestinal tract to the outside environment. The flukes' life cycle is completed in specific snails (the first intermediate host) and encystment on aquatic plants.

Clinical features of fascioliasis relate to the stage and intensity of infection. Acute disease develops during parasite migration (1–2 weeks after infection) and includes fever, right-upper-quadrant pain, hepatomegaly, and eosinophilia. CT of the liver may show migratory tracks. Symptoms and signs usually subside as the parasites reach their final habitat. In individuals with chronic infection, bile duct obstruction and biliary cirrhosis are infrequently demonstrated. No relation to hepatic malignancy has been ascribed to fascioliasis.

Diagnosis

Diagnosis of infection with any of the biliary flukes depends on a high degree of suspicion, elicitation of an appropriate geographic history, and stool examination for characteristically shaped parasite ova.

Additional evidence may be obtained by documenting peripheral-blood eosinophilia or imaging the liver. Serologic testing is helpful, particularly in lightly infected individuals.

Treatment: Biliary Flukes

Drug therapy (praziquantel or triclabendazole) is summarized in Table 219-2. Patients with anatomic lesions in the biliary tract or malignancy are managed according to general medical guidelines.

INTESTINAL FLUKES

Two species of intestinal flukes cause human infection in defined geographic areas worldwide (Table 219-1). The large *Fasciolopsis buski* (adults measure 2 x 7 cm) is endemic in Southeast Asia, while the smaller *Heterophyes heterophyes* is found in the Nile Delta of Egypt and in the Far East. Infection is initiated by ingestion of metacercariae attached to aquatic plants (*F. buski*) or encysted in freshwater or brackish-water fish (*H. heterophyes*). Flukes mature in human intestines, and eggs are passed with stools. Most individuals infected with intestinal flukes are asymptomatic. In heavy *F. buski* infection, diarrhea, abdominal pain, and malabsorption may be encountered. Heavy infection with *H. heterophyes* may be associated with abdominal pain and mucous diarrhea. The diagnosis is established by detection of characteristically shaped ova in stool samples. The drug of choice for treatment is praziquantel (Table 219-2).

LUNG FLUKES

Infection with the lung fluke *Paragonimus westermani* (Table 219-1) and related species (e.g., *P. africanus*) is endemic in many parts of the world, excluding North America and Europe. Endemicity is particularly noticeable in West Africa, Central and South America, and Asia. In nature, the reservoir hosts of *P. westermani* are wild and domestic felines. In Africa, *P. africanus* has been found in other species, such as dogs. Adult lung flukes, which are 7–12 mm in length, are found encapsulated in the lungs of infected persons. In rare circumstances, flukes are found encysted in the CNS (cerebral paragonimiasis) or the abdominal cavity. Humans acquire lung fluke infection by ingesting infective metacercariae encysted in the muscles and viscera of crayfish and freshwater crabs. In endemic areas, these crustaceans are consumed either raw or pickled. Once the organisms reach the duodenum, they excyst, penetrate the gut wall, and travel through the peritoneal cavity, diaphragm, and pleural space to reach the lungs. Mature flukes are found in the bronchioles surrounded by cystic lesions. Parasite eggs are either expectorated with sputum or swallowed and passed to the outside environment with feces. The life cycle is completed in snails and freshwater crustacea.

When maturing flukes lodge in lung tissues, they cause hemorrhage and necrosis, resulting in cyst formation. The adjacent lung parenchyma shows evidence of inflammatory infiltration, predominantly by eosinophils. Cysts usually measure 1–2 cm in diameter and may contain one or two worms each. With the onset of oviposition, cysts usually rupture in adjacent bronchioles—an event allowing ova to exit the human host. Older cysts develop thickened walls, which may undergo calcification. During the active phase of paragonimiasis, lung tissues surrounding parasite cysts may show evidence of pneumonia, bronchitis, bronchiectasis, and fibrosis.

Pulmonary paragonimiasis is particularly symptomatic in persons with moderate to heavy infection. Productive cough with brownish sputum or frank hemoptysis associated with peripheral-blood eosinophilia is usually the presenting feature. Chest examination may reveal signs of pleurisy. In chronic cases, bronchitis or bronchiectasis may predominate, but these conditions rarely proceed to lung abscess. Imaging of the lungs demonstrates characteristic features, including patchy densities, cavities, pleural effusion, and ring shadows. Cerebral paragonimiasis presents as either space-occupying lesions or epilepsy.

Diagnosis

Pulmonary paragonimiasis is diagnosed by detection of parasite ova in sputum and/or stools. Serology is of considerable help in egg-negative cases and in cerebral paragonimiasis.

Treatment: Lung Flukes

The drug of choice for treatment is praziquantel (Table 219-2). Other medical or surgical management may be needed for pulmonary or cerebral lesions.

CONTROL AND PREVENTION OF TISSUE FLUKES

For residents of nonendemic areas who are visiting an endemic region, the only effective preventive measure is to avoid ingestion of local plants, fish, or crustaceans; if their ingestion is necessary, these items should be washed or cooked thoroughly. Instruction on water and food preparation and consumption should be included in physicians' advice to travelers (Chap. 123). Interruption of transmission among residents of endemic areas depends on avoiding ingestion of infective stages and disposing of feces and sputum appropriately to prevent hatching of eggs in the environment. These two approaches rely greatly on socioeconomic development and health education. In countries where economic progress has resulted in financial and social improvements, transmission has decreased.

The third approach to control in endemic communities entails selective use of chemotherapy for individuals posing the highest risk of transmission (i.e., those with heavy infections). The availability of praziquantel—a broad-spectrum, safe, and effective anthelmintic agent—provides a means for reducing the reservoirs of infection in human populations. However, the existence of most of these helminthic infections as zoonoses in several animal species complicates control efforts.

FURTHER READINGS

Curtale F et al: Comprehensive primary health care, a viable strategy for the elimination of schistosomiasis. Trans R Soc Trop Med Hyg 104:70, 2010[PMID: 19775714]

King CH: Lifting the burden of schistosomiasis—defining elements of infection-associated disease and the benefits of antiparasite treatment. J Infect Dis 196:653, 2007[PMID: 17674304]

Lim JH et al: Parasitic diseases of the biliary tract. AJR Am J Roentgenol 188:1596, 2007[PMID: 17515382]

Lun ZR et al: Clonorchiasis: A key foodborne zoonosis in China. Lancet Infect Dis 5:31, 2005[PMID: 15620559]

Robinson MW, Dalton JP: Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiases. Philos Trans R Soc Lond B Biol Sci 364:2763, 2009[PMID: 19687044]

Schistosoma japonicum genome sequencing and functional analysis consortium: The *Schistosoma japonicum* genome reveals features of host-parasite interplay. Nature 460:345, 2009

Stauffer WM et al: Biliary liver flukes (opisthorchiasis and clonorchiasis) in immigrants in the United States: Often subtle and diagnosed years after arrival. J Travel Med 11:157, 2004[PMID: 15710057]

Walson JL et al: Prevalence and correlates of helminth co-infection in Kenyan HIV-1 infected adults. PLoS Negl Trop Dis 4:e644, 2010

Wilson S et al: Health implications of chronic hepatosplenomegaly in Kenyan school-aged children chronically exposed to malarial infections and *Schistosoma mansoni*. Trans R Soc Trop Med Hyg 104:110, 2010[PMID: 19818465]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CESTODE INFECTIONS: INTRODUCTION

Cestodes, or tapeworms, are segmented worms. The adults reside in the gastrointestinal tract, but the larvae can be found in almost any organ. Human tapeworm infections can be divided into two major clinical groups. In one group, humans are the definitive hosts, with the adult tapeworms living in the gastrointestinal tract (*Taenia saginata*, *Diphyllobothrium*, *Hymenolepis*, and *Dipylidium caninum*). In the other, humans are intermediate hosts, with larval-stage parasites present in the tissues; diseases in this category include echinococcosis, sparganosis, and coenurosis. Humans may be either the definitive or the intermediate hosts for *Taenia solium*. Both stages of *Hymenolepis nana* are found simultaneously in the human intestines.

The ribbon-shaped tapeworm attaches to the intestinal mucosa by means of sucking cups or hooks located on the scolex. Behind the scolex is a short, narrow neck from which proglottids (segments) form. As each proglottid matures, it is displaced further back from the neck by the formation of new, less mature segments. The progressively elongating chain of attached proglottids, called the *strobila*, constitutes the bulk of the tapeworm. The length varies among species. In some, the tapeworm may consist of more than 1000 proglottids and may be several meters long. The mature proglottids are hermaphroditic and produce eggs, which are subsequently released. Since eggs of the different *Taenia* species are morphologically identical, differences in the morphology of the scolex or proglottids provide the basis for diagnostic identification to the species level.

Most human tapeworms require at least one intermediate host for complete larval development. After ingestion of the eggs or proglottids by an intermediate host, the larval oncospheres are activated, escape the egg, and penetrate the intestinal mucosa. The oncosphere migrates to tissues and develops into an encysted form known as a *cysticercus* (single scolex), a *coenurus* (multiple scolices), or a *hydatid* (cyst with daughter cysts, each containing several protoscolices). Ingestion by the definitive host of tissues containing a cyst enables a scolex to develop into a tapeworm.

Taeniasis Saginata and Taeniasis Asiatica

The beef tapeworm *T. saginata* occurs in all countries where raw or undercooked beef is eaten. It is most prevalent in sub-Saharan African and Middle Eastern countries. *T. asiatica* is closely related to *T. saginata* and is found in Asia with pigs as intermediate hosts. The clinical manifestations and morphology of these two species are very similar and are therefore discussed together.

ETIOLOGY AND PATHOGENESIS

Humans are the only definitive host for the adult stage of *T. saginata* and *T. asiatica*. The tapeworms, which can reach 8 m in length with 1000–2000 proglottids, inhabit the upper jejunum. The scolex of *T. saginata* has four prominent suckers, whereas *T. asiatica* has an unarmed rostellum. Each gravid segment has 15–30 uterine branches (in contrast to 8–12 for *T. solium*). The eggs are indistinguishable from those of *T. solium*; they measure 30–40 μm, contain the oncosphere, and have a thick brown striated shell. Eggs deposited on vegetation can live for months or years until they are ingested by cattle or other herbivores (*T. saginata*) or pigs *(T. asiatica*). The embryo released after ingestion invades the intestinal wall and is carried to striated muscle or viscera, where it transforms into the cysticercus. When ingested in raw or undercooked meat, this form can infect humans. After the cysticercus is ingested, it takes ~2 months for the mature adult worm to develop.

CLINICAL MANIFESTATIONS

Patients become aware of the infection most commonly by noting passage of proglottids in their feces. The proglottids are often motile, and patients may experience perianal discomfort when proglottids are discharged. Mild abdominal pain or discomfort, nausea, change in appetite, weakness, and weight loss can occur.

DIAGNOSIS

The diagnosis is made by the detection of eggs or proglottids in the stool. Eggs may also be present in the perianal area; thus, if proglottids or eggs are not found in the stool, the perianal region should be examined with use of a cellophane-tape swab (as in pinworm infection; Chap. 217). Distinguishing *T. saginata* or *T. asiatica* from *T. solium* requires examination of mature proglottids. All three species can be distinguished by examining the scolex. Available serologic tests are not helpful diagnostically. Eosinophilia and elevated levels of serum IgE may be detected.

TREATMENT: TAENIASIS SAGINATA AND TAENIASIS ASIATICA

A single dose of praziquantel (10 mg/kg) is highly effective.

PREVENTION

The major method of preventing infection is the adequate cooking of beef or pork viscera; exposure to temperatures as low as 56°C for 5 min will destroy cysticerci. Refrigeration or salting for long periods or freezing at –10°C for 9 days also kills cysticerci in beef. General preventive measures include inspection of beef and proper disposal of human feces.

Taeniasis Solium and Cysticercosis

The pork tapeworm *T. solium* can cause two distinct forms of infection in humans: adult tapeworms in the intestine or larval forms in the tissues (cysticercosis). Humans are the only definitive hosts for *T. solium*; pigs are the usual intermediate hosts, although other animals may harbor the larval forms.

T. solium exists worldwide but is most prevalent in Latin America, sub-Saharan Africa, China, India, and Southeast Asia. Cysticercosis occurs in industrialized nations largely as a result of the immigration of infected persons from endemic areas.

ETIOLOGY AND PATHOGENESIS

The adult tapeworm generally resides in the upper jejunum. The scolex attaches by both sucking disks and two rows of hooklets. Often only one adult worm is present, but that worm may live for years. The tapeworm, usually ~3 m in length, may have as many as 1000 proglottids, each of which produces up to 50,000 eggs. Groups of 3–5 proglottids are generally released and excreted into the feces, and the eggs in these proglottids are infective for both humans and animals. The eggs may survive in the environment for several months. After ingestion of eggs by the pig intermediate host, the larvae are activated, escape the egg, penetrate the intestinal wall, and are carried to many tissues; they are most frequently identified in striated muscle of the neck, tongue, and trunk. Within 60 –90 days, the encysted larval stage develops. These cysticerci can survive for months to years. By ingesting undercooked pork containing cysticerci, humans acquire infections that lead to intestinal tapeworms. Infections that cause human cysticercosis follow the ingestion of *T. solium* eggs, usually from close contact with a tapeworm carrier. Autoinfection may occur if an individual with an egg-producing tapeworm ingests eggs derived from his or her own feces.

CLINICAL MANIFESTATIONS

Intestinal infections with *T. solium* may be asymptomatic. Fecal passage of proglottids may be noted by patients. Other symptoms are

infrequent.

In cysticercosis, the clinical manifestations are variable. Cysticerci can be found anywhere in the body but are most commonly detected in the brain, cerebrospinal fluid (CSF), skeletal muscle, subcutaneous tissue, or eye. The clinical presentation of cysticercosis depends on the number and location of cysticerci as well as the extent of associated inflammatory responses or scarring. Neurologic manifestations are the most common **(Fig. 220-1)**. Seizures are associated with inflammation surrounding cysticerci in the brain parenchyma. These seizures may be generalized, focal, or Jacksonian. Hydrocephalus results from CSF flow obstruction by cysticerci and accompanying inflammation or by CSF outflow obstruction from arachnoiditis. Signs of increased intracranial pressure, including headache, nausea, vomiting, changes in vision, dizziness, ataxia, or confusion, are often evident. Patients with hydrocephalus may develop papilledema or display altered mental status. When cysticerci develop at the base of the brain or in the subarachnoid space, they may cause chronic meningitis or arachnoiditis, communicating hydrocephalus, or strokes.

Figure 220-1

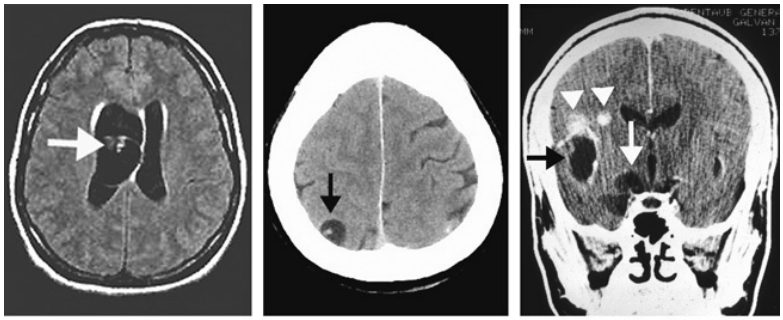

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurocysticercosis is caused by *Taenia solium*. Neurologic infection can be classified on the basis of the location and viability of the parasites. When the parasites are in the ventricles, they often cause obstructive hydrocephalus. ***Left:*** MRI showing a cysticercus in the lateral ventricle, with resultant hydrocephalus. The arrow points to the scolex within the cystic parasite. ***Center:*** CT showing a parenchymal cysticercus, with enhancement of the cyst wall and an internal scolex (*arrow*). ***Right:*** Multiple cysticerci, including calcified lesions from prior infection (*arrowheads*), viable cysticerci in the basilar cisterns (*white arrow*), and a large degenerating cysticercus in the Sylvian fissure (*black arrow*). *(Modified with permission from JC Bandres et al: Clin Infect Dis 15:799, 1992. © The University of Chicago Press.)*

DIAGNOSIS

The diagnosis of intestinal *T. solium* infection is made by the detection of eggs or proglottids, as described for *T. saginata*. More sensitive methods, including antigen-capture ELISA, PCR, and serology for tapeworm stage-specific antigens, are currently available only as research techniques. In cysticercosis, diagnosis can be difficult. A consensus conference has delineated absolute, major, minor, and epidemiologic criteria for diagnosis **(Table 220-1)**. Diagnostic certainty is possible only with definite demonstration of the parasite (absolute criteria). This task can be accomplished by histologic observation of the parasite in excised tissue, by funduscopic visualization of the parasite in the eye (in the anterior chamber, vitreous, or subretinal spaces), or by neuroimaging studies demonstrating cystic lesions containing a characteristic scolex. With improving resolution of neuroimaging studies, the scolex can now be identified in many cases. In other instances, a clinical diagnosis is made on the basis of a combination of clinical presentation, radiographic studies, serologic tests, and exposure history.

Table 220-1 Diagnostic Criteria for Human Cysticercosis[a]

1. Absolute criteria
a. Demonstration of cysticerci by histologic or microscopic examination of biopsy material
b. Visualization of the parasite in the eye by funduscopy
c. Neuroradiologic demonstration of cystic lesions containing a characteristic scolex
2. Major criteria
a. Neuroradiologic lesions suggestive of neurocysticercosis
b. Demonstration of antibodies to cysticerci in serum by enzyme-linked immunoelectrotransfer blot
c. Resolution of intracranial cystic lesions spontaneously or after therapy with albendazole or praziquantel alone
3. Minor criteria
a. Lesions compatible with neurocysticercosis detected by neuroimaging studies
b. Clinical manifestations suggestive of neurocysticercosis
c. Demonstration of antibodies to cysticerci or cysticercal antigen in cerebrospinal fluid by ELISA
d. Evidence of cysticercosis outside the central nervous system (e.g., cigar-shaped soft-tissue calcifications)
4. Epidemiologic criteria
a. Residence in a cysticercosis-endemic area
b. Frequent travel to a cysticercosis-endemic area
c. Household contact with an individual infected with *Taenia solium*

[a]Diagnosis is confirmed by either one absolute criterion or a combination of two major criteria, one minor criterion, and one epidemiologic criterion. A probable diagnosis is supported by the fulfillment of (1) one major criterion plus two minor criteria; (2) one major criterion plus one minor criterion and one epidemiologic criterion; or (3) three minor criteria plus one epidemiologic criterion.

Abbreviation: ELISA, enzyme-linked immunosorbent assay.

Source: Modified from Del Brutto et al., 2001.

Neuroimaging findings suggestive of neurocysticercosis constitute the primary major diagnostic criterion. These findings include cystic lesions with or without enhancement (e.g., ring enhancement), one or more nodular calcifications (which may also have associated enhancement), or focal enhancing lesions. Cysticerci in the brain parenchyma are usually 5–20 mm in diameter and rounded. Cystic lesions in the subarachnoid space or fissures may enlarge up to 6 cm in diameter and may be lobulated. For cysticerci within the subarachnoid space or ventricles, the walls may be very thin and the cyst fluid is often isodense with CSF. Thus, obstructive hydrocephalus or enhancement of the basilar meninges may be the only finding on CT in extraparenchymal neurocysticercosis.

Cysticerci in the ventricles or subarachnoid space are usually visible to an experienced neuroradiologist on MRI or on CT with intraventricular contrast injection. CT is more sensitive than MRI in identifying calcified lesions, whereas MRI is better for identifying cystic lesions, scolices, and enhancement.

The second major diagnostic criterion is detection of specific antibodies to cysticerci. While most tests employing unfractionated antigen have high rates of false-positive and false-negative results, this problem can be overcome by using the more specific immunoblot assay. An immunoblot assay using lentil-lectin purified glycoproteins has >99% specificity and is highly sensitive. However, patients with single intracranial lesions or with calcifications may be seronegative. With this assay, serum samples provide greater diagnostic sensitivity than CSF. All of the diagnostic antigens have been cloned, and assays using recombinant antigens are being developed. Antigen detection assays using monoclonal antibodies to detect parasite antigen in the blood or CSF may also facilitate diagnosis and patient follow-up. However, these assays are not widely available at present.

Studies have demonstrated that clinical criteria can aid in diagnosis in selected cases. In patients from endemic areas who had single enhancing lesions presenting with seizures, a normal physical examination, and no evidence of systemic disease (e.g., no fever, adenopathy, or chest radiographic abnormalities), the constellation of rounded CT lesions 5–20 mm in diameter with no midline shift was almost always caused by neurocysticercosis. Finally, spontaneous resolution or resolution after therapy with albendazole alone is consistent with neurocysticercosis.

Minor diagnostic criteria include neuroimaging findings consistent with but less characteristic of cysticercosis, clinical manifestations suggestive of neurocysticercosis (e.g., seizures, hydrocephalus, or altered mental status), evidence of cysticercosis outside the central nervous system (CNS) (e.g., cigar-shaped soft-tissue calcifications), or detection of antibody in CSF by ELISA. Epidemiologic criteria include exposure to a tapeworm carrier or household member infected with *T. solium*, current or prior residence in an endemic area, and frequent travel to an endemic area.

The diagnosis is confirmed in patients with either one absolute criterion or a combination of two major criteria, one minor criterion, and one epidemiologic criterion (Table 220-1). A probable diagnosis is supported by the fulfillment of (1) one major criterion plus two minor criteria; (2) one major criterion plus one minor criterion and one epidemiologic criterion; or (3) three minor criteria plus one epidemiologic criterion. While the CSF is usually abnormal in neurocysticercosis, CSF abnormalities are not pathognomonic. Patients may have CSF pleocytosis with a predominance of lymphocytes, neutrophils, or eosinophils. The protein level in CSF may be elevated; the glucose concentration is usually normal but may be depressed.

TREATMENT: TAENIASIS SOLIUM AND CYSTICERCOSIS

Intestinal *T. solium* infection is treated with a single dose of praziquantel (10 mg/kg). However, praziquantel occasionally evokes an inflammatory response in the CNS if concomitant cryptic cysticercosis is present. Niclosamide (2 g) is also effective but is not widely available.

The initial management of neurocysticercosis should focus on symptom-based treatment of seizures or hydrocephalus. Seizures can usually be controlled with antiepileptic treatment. If parenchymal lesions resolve without development of calcifications and patients remain free of seizures, antiepileptic therapy can usually be discontinued after 1–2 years. Placebo-controlled trials are beginning to clarify the clinical advantage of antiparasitic drugs for parenchymal neurocysticercosis. Trends toward faster resolution of neuroradiologic abnormalities have been observed in most studies. The clinical benefits are less dramatic and consist mainly of shortening the period during which recurrent seizures occur and decreasing the number of patients who have many recurrent seizures. For the treatment of patients with brain parenchymal cysticerci, most authorities favor antiparasitic drugs, including albendazole (15 mg/kg per day for 8–28 days) or praziquantel (50–100 mg/kg daily in three divided doses for 15–30 days). Longer courses are often needed in patients with multiple subarachnoid cysticerci. Both agents may exacerbate the inflammatory response around the dying parasite, thereby exacerbating seizures or hydrocephalus as well. Thus, patients receiving these drugs should be carefully monitored, and high-dose glucocorticoids should be used during treatment. Since glucocorticoids induce first-pass metabolism of praziquantel and may decrease its antiparasitic effect, cimetidine should be co-administered to inhibit praziquantel metabolism. Pilot studies suggest that the two drugs in combination may be more effective than the individual agents.

For patients with hydrocephalus, the emergent reduction of intracranial pressure is the mainstay of therapy. In the case of obstructive hydrocephalus, the preferred approach is removal of the cysticercus via endoscopic surgery. However, this intervention is not always possible. An alternative approach is initially to perform a diverting procedure, such as ventriculoperitoneal shunting. Historically, shunts have usually failed, but low failure rates have been attained with administration of antiparasitic drugs and glucocorticoids. Open craniotomy to remove cysticerci is now required only infrequently. For patients with subarachnoid cysts or giant cysticerci, glucocorticoids are needed to reduce arachnoiditis and accompanying vasculitis. Most authorities recommend prolonged courses of antiparasitic drugs and shunting when hydrocephalus is present. Methotrexate can be used as a steroid-sparing agent in patients requiring prolonged therapy. In patients with diffuse cerebral edema and elevated intracranial pressure due to multiple inflamed lesions, glucocorticoids are the mainstay of therapy, and antiparasitic drugs should be avoided. For ocular and spinal medullary lesions, drug-induced inflammation may cause irreversible damage. Most patients should be managed surgically, although case reports have described cures with medical therapy.

PREVENTION

Measures for the prevention of intestinal *T. solium* infection consist of the application to pork of precautions similar to those described above for beef with regard to *T. saginata* infection. The prevention of cysticercosis involves minimizing the opportunities for ingestion of fecally derived eggs by means of good personal hygiene, effective fecal disposal, and treatment and prevention of human intestinal infections. Mass chemotherapy has been administered to human and porcine populations in efforts at disease eradication. Finally, vaccines to prevent porcine cysticercosis have shown promise in studies and are under development.

Echinococcosis

Echinococcosis is an infection caused in humans by the larval stage of the *Echinococcus granulosus* complex, *E. multilocularis*, or *E. vogeli*. *E. granulosus* complex parasites, which produce unilocular cystic lesions and are prevalent in areas where livestock is raised in association with dogs, cause cystic hydatid disease. Molecular evidence suggests that *E. granulosus* strains may actually belong to more than one species; specifically, strains from sheep, cattle, pigs, horses, and camels probably represent separate species. These parasites are found on all continents, with areas of high prevalence in China, central Asia, the Middle East, the Mediterranean region, eastern Africa, and parts of South America. *E. multilocularis*, which causes multilocular alveolar lesions that are locally invasive, is found in Alpine, sub-Arctic, or Arctic regions, including Canada, the United States, and central and northern Europe; China; and central Asia. *E. vogeli* causes polycystic hydatid disease and is found only in Central and South America.

Like other cestodes, echinococcal species have both intermediate and definitive hosts. The definitive hosts are canines that pass eggs in their feces. After the ingestion of eggs, cysts develop in the intermediate hosts—sheep, cattle, humans, goats, camels, and horses for the *E. granulosus* complex and mice and other rodents for *E. multilocularis*. When a dog (*E. granulosus*) or fox (*E. multilocularis*) ingests infected meat containing cysts, the life cycle is completed.

ETIOLOGY

The small (5-mm-long) adult *E. granulosus* complex worms, which live for 5–20 months in the jejunum of dogs, have only three proglottids: one immature, one mature, and one gravid. The gravid segment splits to release eggs that are morphologically similar to *Taenia* eggs and are extremely hardy. After humans ingest the eggs, embryos escape from the eggs, penetrate the intestinal mucosa, enter the portal circulation, and are carried to various organs, most commonly the liver and lungs. Larvae develop into fluid-filled unilocular hydatid cysts that consist of an external membrane and an inner germinal layer. Daughter cysts develop from the inner aspect of the germinal layer, as do germinating cystic structures called *brood capsules*. New larvae, called *protoscolices*, develop in

large numbers within the brood capsule. The cysts expand slowly over a period of years.

The life cycle of *E. multilocularis* is similar except that wild canines, such as foxes, serve as the definitive hosts and small rodents serve as the intermediate hosts. The larval form of *E. multilocularis*, however, is quite different in that it remains in the proliferative phase, the parasite is always multilocular, and vesicles without brood capsule or protoscolices progressively invade the host tissue by peripheral extension of processes from the germinal layer.

CLINICAL MANIFESTATIONS

Slowly enlarging echinococcal cysts generally remain asymptomatic until their expanding size or their space-occupying effect in an involved organ elicits symptoms. The liver and the lungs are the most common sites of these cysts. The liver is involved in about two-thirds of *E. granulosus* infections and in nearly all *E. multilocularis* infections. Since a period of years elapses before cysts enlarge sufficiently to cause symptoms, they may be discovered incidentally on a routine x-ray or ultrasound study.

Patients with hepatic echinococcosis who are symptomatic most often present with abdominal pain or a palpable mass in the right upper quadrant. Compression of a bile duct or leakage of cyst fluid into the biliary tree may mimic recurrent cholelithiasis, and biliary obstruction can result in jaundice. Rupture of or episodic leakage from a hydatid cyst may produce fever, pruritus, urticaria, eosinophilia, or anaphylaxis. Pulmonary hydatid cysts may rupture into the bronchial tree or peritoneal cavity and produce cough, salty phlegm, dyspnea, chest pain, or hemoptysis. Rupture of hydatid cysts, which can occur spontaneously or at surgery, may lead to multifocal dissemination of protoscolices, which can form additional cysts. Other presentations are due to the involvement of bone (invasion of the medullary cavity with slow bone erosion producing pathologic fractures), the CNS (space-occupying lesions), the heart (conduction defects, pericarditis), and the pelvis (pelvic mass).

The larval forms of *E. multilocularis* characteristically present as a slowly growing hepatic tumor, with progressive destruction of the liver and extension into vital structures. Patients commonly report upper-quadrant and epigastric pain. Liver enlargement and obstructive jaundice may be apparent. The lesions may infiltrate adjoining organs (e.g., diaphragm, kidneys, or lungs) or may metastasize to the spleen, lungs, or brain.

DIAGNOSIS

Radiographic and related imaging studies are important in detecting and evaluating echinococcal cysts. Plain x-rays will define pulmonary cysts of *E. granulosus*—usually as rounded masses of uniform density—but may miss cysts in other organs unless there is cyst wall calcification (as occurs in the liver). MRI, CT, and ultrasound reveal well-defined cysts with thick or thin walls. When older cysts contain a layer of hydatid sand that is rich in accumulated protoscolices, these imaging methods may detect this fluid layer of different density. However, the most pathognomonic finding, if demonstrable, is that of daughter cysts within the larger cyst. This finding, like eggshell or mural calcification on CT, is indicative of *E. granulosus* infection and helps to distinguish the cyst from carcinomas, bacterial or amebic liver abscesses, or hemangiomas. In contrast, ultrasound or CT of alveolar hydatid cysts reveals indistinct solid masses with central necrosis and plaquelike calcifications.

A specific diagnosis of *E. granulosus* infection can be made by the examination of aspirated fluids for protoscolices or hooklets, but diagnostic aspiration is not usually recommended because of the risk of fluid leakage resulting in either dissemination of infection or anaphylactic reactions. Serodiagnostic assays can be useful, although a negative test does not exclude the diagnosis of echinococcosis. Cysts in the liver elicit positive antibody responses in ~90% of cases, whereas up to 50% of individuals with cysts in the lungs are seronegative. Detection of antibody to specific echinococcal antigens by immunoblotting has the highest degree of specificity.

TREATMENT: ECHINOCOCCOSIS

Therapy for cystic echinococcosis is based on considerations of the size, location, and manifestations of cysts and the overall health of the patient. Surgery has traditionally been the principal definitive method of treatment. Currently, ultrasound staging is recommended for *E. granulosus* infections **(Fig. 220-2)**. Small C1, CE1, and CE3 lesions may respond to chemotherapy with albendazole. For CE1 lesions and uncomplicated CE3 lesions, PAIR (*p*ercutaneous *a*spiration, *i*nfusion of scolicidal agents, and *r*easpiration) is now recommended instead of surgery. PAIR is contraindicated for superficially located cysts (because of the risk of rupture), for cysts with multiple thick internal septal divisions (honeycombingpattern), and for cysts communicating with the biliary tree. For prophylaxis of secondary peritoneal echinococcosis due to inadvertent spillage of fluid during PAIR, the administration of albendazole (15 mg/kg daily in two divided doses) should be initiated at least 4 days before the procedure and continued for at least 4 weeks afterward. Ultrasound- or CT-guided aspiration allows confirmation of the diagnosis by demonstration of protoscolices in the aspirate. After aspiration, contrast material should be injected to detect occult communications with the biliary tract. Alternatively, the fluid should be checked for bile staining by dipstick. If no bile is found and no communication visualized, the contrast material is reaspirated, with subsequent infusion of scolicidal agents (usually 95% ethanol; alternatively, hypertonic saline). This approach, when implemented by a skilled practitioner, yields rates of cure and relapse equivalent to those following surgery, with less perioperative morbidity and shorter hospitalization. In experienced hands, some CE2 lesions can be treated by aspiration with a trochar. Daughter cysts within the primary cyst may need to be punctured separately, and catheter drainage may be required.

Figure 220-2

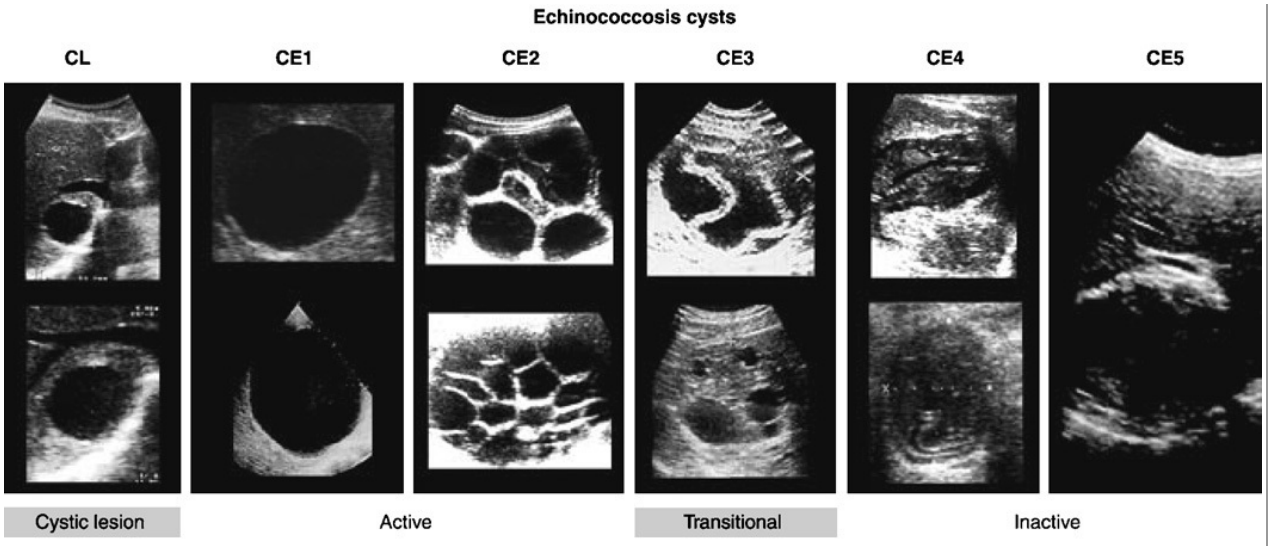

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of cystic hydatid disease caused by *Echinococcus granulosus* should be based on viability of the parasite, which can be estimated from radiographic appearance. The ultrasound appearance includes lesions classified as active, transitional, and inactive. *Active* cysts include types CL (with a cystic lesion and no visible cyst wall), CE1 [with a visible cyst wall and internal echoes (snowflake sign)], and CE2 (with a visible cyst wall and internal septation). *Transitional cysts* (CE3) may have detached laminar membranes or may be partially collapsed. *Inactive cysts* include types CE4 (a nonhomogeneous mass) and CE5 (a cyst with a thick calcified wall). *[Adapted from RL Guerrant et al (eds): Tropical Infectious Diseases: Principles, Pathogens and Practice, 2nd ed, p 1312. © 2005, with permission from Elsevier Science.]*

Surgery remains the treatment of choice for complicated *E. granulosus* cysts (e.g., those communicating with the biliary tract) or for areas where PAIR is not possible. For *E. granulosus*, the preferred surgical approach is pericystectomy, in which the entire cyst and the surrounding fibrous tissue are removed. The risks posed by leakage of fluid during surgery or PAIR include anaphylaxis and dissemination of infectious protoscolices. The latter complication has been minimized by careful attention to the prevention of spillage of the cyst and by soaking of the drapes with hypertonic saline. Infusion of scolicidal agents is no longer recommended because of problems with hypernatremia, intoxication, or sclerosing cholangitis. Albendazole, which is active against *Echinococcus*, should be administered adjunctively, beginning several days before resection and continuing for several weeks for *E. granulosus*. Praziquantel (50 mg/kg daily for 2 weeks) may hasten the death of the protoscolices. Medical therapy with albendazole alone for 12 weeks to 6 months results in cure in ~30% of cases and in improvement in another 50%. In many instances of treatment failure, *E. granulosus* infections are subsequently treated successfully with PAIR or additional courses of medical therapy. Response to treatment is best assessed by serial imaging studies, with attention to cyst size and consistency. Some cysts may not demonstrate complete radiologic resolution even though no viable protoscolices are present. Some of these cysts with partial radiologic resolution (e.g., CE4 or CE5) can be managed with observation only.

Surgical resection remains the treatment of choice for *E. multilocularis* infection. Complete removal of the parasite continues to offer the best chance for cure. Ongoing therapy with albendazole for at least 2 years after presumptively curative surgery is recommended. Positron emission tomography (PET) scanning can be used to follow disease activity. Most cases are diagnosed at a stage at which complete resection is not possible; in these cases, albendazole treatment should be continued indefinitely, with careful monitoring. In some cases, liver transplantation has been used because of the size of the necessary liver resection. However, continuous immunosuppression favors the proliferation of *E. multilocularis* larvae and reinfection of the transplant. Thus, indefinite treatment with albendazole is required.

PREVENTION

In endemic areas, echinococcosis can be prevented by administering praziquantel to infected dogs, by denying dogs access to infected animals, or by vaccinating sheep. Limitation of the number of stray dogs is helpful in reducing the prevalence of infection among humans.

Hymenolepiasis Nana

Infection with *H. nana*, the dwarf tapeworm, is the most common of all the cestode infections. *H. nana* is endemic in both temperate and tropical regions of the world. Infection is spread by fecal/oral contamination and is common among institutionalized children.

ETIOLOGY AND PATHOGENESIS

H. nana is the only cestode of humans that does not require an intermediate host. Both the larval and adult phases of the life cycle take place in the human. The adult—the smallest tapeworm parasitizing humans—is ~2 cm long and dwells in the proximal ileum. Proglottids, which are quite small and are rarely seen in the stool, release spherical eggs 30–44 μm in diameter, each of which contains an oncosphere with six hooklets. The eggs are immediately infective and are unable to survive for >10 days in the external environment. When the egg is ingested by a new host, the oncosphere is freed and penetrates the intestinal villi, becoming a cysticercoid larva. Larvae migrate back into the intestinal lumen, attach to the mucosa, and mature into adult worms over 10–12 days. Eggs may also hatch before passing into the stool, causing internal autoinfection with increasing numbers of intestinal worms. Although the life span of adult *H. nana* worms is only ~4–10 weeks, the autoinfection cycle perpetuates the infection.

CLINICAL MANIFESTATIONS

H. nana infection, even with many intestinal worms, is usually asymptomatic. When infection is intense, anorexia, abdominal pain, and diarrhea develop.

DIAGNOSIS

Infection is diagnosed by the finding of eggs in the stool.

TREATMENT: HYMENOLEPIASIS NANA

Praziquantel (25 mg/kg once) is the treatment of choice, since it acts against both the adult worms and the cysticercoids in the intestinal villi. Nitazoxanide (500 mg bid for 3 days) may be used as an alternative.

PREVENTION

Good personal hygiene and improved sanitation can eradicate the disease. Epidemics have been controlled by mass chemotherapy coupled with improved hygiene.

Hymenolepiasis Diminuta

Hymenolepis diminuta, a cestode of rodents, occasionally infects small children, who ingest the larvae in uncooked cereal foods contaminated by fleas and other insects in which larvae develop. Infection is usually asymptomatic and is diagnosed by the detection of eggs in the stool. Treatment with praziquantel results in cure in most cases.

Diphyllobothriasis

Diphyllobothrium latum and other *Diphyllobothrium* species are found in the lakes, rivers, and deltas of the Northern Hemisphere, Central Africa, and South America.

ETIOLOGY AND PATHOGENESIS

The adult worm—the longest tapeworm (up to 25 m)—attaches to the ileal and occasionally to the jejunal mucosa by its suckers, which are located on its elongated scolex. The adult worm has 3000–4000 proglottids, which release ~1 million eggs daily into the feces. If an egg reaches water, it hatches and releases a free-swimming embryo that can be eaten by small freshwater crustaceans (*Cyclops* or *Diaptomus* species). After an infected crustacean containing a developed procercoid is swallowed by a fish, the larva migrates into the fish's flesh and grows into a plerocercoid, or sparganum larva. Humans acquire the infection by ingesting infected raw or smoked fish. Within 3–5 weeks, the tapeworm matures into an adult in the human intestine.

CLINICAL MANIFESTATIONS

Most *D. latum* infections are asymptomatic, although manifestations may include transient abdominal discomfort, diarrhea, vomiting, weakness, and weight loss. Occasionally, infection can cause acute abdominal pain and intestinal obstruction; in rare cases, cholangitis or cholecystitis may be produced by migrating proglottids. Because the tapeworm absorbs large quantities of vitamin B12 and interferes with ileal B_{12} absorption, vitamin B_{12} deficiency can develop, but this effect has been noted only in Scandinavia, where up to 2% of infected patients, especially the elderly, have megaloblastic anemia resembling pernicious anemia and may exhibit neurologic sequelae of B_{12} deficiency.

DIAGNOSIS

The diagnosis is made readily by the detection of the characteristic eggs in the stool. The eggs possess a single shell with an operculum at one end and a knob at the other. Mild to moderate eosinophilia may be detected.

TREATMENT: DIPHYLLOBOTHRIASIS

Praziquantel (5–10 mg/kg once) is highly effective. Parenteral vitamin B12 should be given if B_{12} deficiency is manifest.

PREVENTION

Infection can be prevented by heating fish to 54°C for 5 min or by freezing it at –18°C for 24 h. Placing fish in brine with a high salt concentration for long periods kills the eggs.

Dipylidiasis

Dipylidium caninum, a common tapeworm of dogs and cats, may accidentally infect humans. Dogs, cats, and occasionally humans become infected by ingesting fleas harboring cysticercoids. Children are more likely to become infected than adults. Most infections are asymptomatic, but abdominal pain, diarrhea, anal pruritus, urticaria, eosinophilia, or passage of segments in the stool may occur. The diagnosis is made by the detection of proglottids or ova in the stool. As in *D. latum* infection, therapy consists of praziquantel. Prevention requires anthelmintic treatment and flea control for pet dogs or cats.

Sparganosis

Humans can be infected by the sparganum, or plerocercoid larva, of a diphyllobothrid tapeworm of the genus *Spirometra*. Infection can be acquired by the consumption of water containing infected *Cyclops*; by the ingestion of infected snakes, birds, or mammals; or by the application of infected flesh as poultices. The worm migrates slowly in tissues, and infection commonly presents as a subcutaneous swelling. Periorbital tissues can be involved, and ocular sparganosis may destroy the eye. Surgical excision is used to treat localized sparganosis.

Coenurosis

This rare infection of humans by the larval stage (coenurus) of the dog tapeworm *Taenia multiceps* or *T. serialis* results in a space-occupying cystic lesion. As in cysticercosis, involvement of the CNS and subcutaneous tissue is most common. Both definitive diagnosis and treatment require surgical excision of the lesion. Chemotherapeutic agents generally are not effective.

FURTHER READINGS

Articles

Brunetti E et al: Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. Acta Trop 114:1, 2010[PMID: 19931502]

Craig P, Ito A: Intestinal cestodes. Curr Opin Infect Dis 20:524, 2007[PMID: 17762788]

Del Brutto OH et al: Proposed diagnostic criteria for neurocysticercosis. Neurology 57:177, 2001

——— et al: Meta-analysis: Cysticidal drugs for neurocysticercosis: Albendazole and praziquantel. Ann Intern Med 145:43, 2006

Garcia HH et al: A trial of antiparasitic treatment to reduce the rate of seizures due to cerebral cysticercosis. N Engl J Med 350:249, 2004[PMID: 14724304]

Nash TE et al: Treatment of neurocysticercosis: Current status and future research needs. Neurology 67:1120, 2006[PMID: 17030744]

Rangel-Castilla L et al: Contemporary neurosurgical approaches to neurocysticercosis. Am J Trop Med Hyg 80:373, 2009[PMID: 19270284]

Scholz T et al: Update on the human broad tapeworm (genus *Diphyllobothrium*), including clinical relevance. Clin Microbiol Rev 22:146, 2009[PMID: 19136438]

Serpa JA et al: Neurocysticercosis in Houston, Texas: An update. Medicine (Baltimore) 90:81, 2011[PMID: 21200189]

World health organization informal working group on echinococcosis: International classification of ultrasound images in cystic echinococcosis for application in clinical and field epidemiological settings. Acta Tropica 85:253, 2003

Web Sites

CDC Division of Parasitic Diseases. *www.cdc.gov/ncidod/dpd/default.htm*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MICROBIAL BIOTERRORISM: INTRODUCTION

Descriptions of the use of microbial pathogens as potential weapons of war or terrorism date from ancient times. Among the most frequently cited of such episodes are the poisoning of water supplies in the sixth century b.c. with the fungus *Claviceps purpurea* (rye ergot) by the Assyrians, the hurling of the dead bodies of plague victims over the walls of the city of Kaffa by the Tartar army in 1346, and the efforts by the British to spread smallpox to the Native American population loyal to the French via contaminated blankets in 1767. Although the use of chemical weapons in wartime took place in the not-too-distant past (Chap. 222), the tragic events of September 11, 2001, followed closely by the mailing of letters containing anthrax spores to media and congressional offices through the U.S. Postal Service, dramatically changed the mindset of the American public regarding both our vulnerability to microbial bioterrorist attacks and the seriousness and intent of the federal government to protect its citizens against future attacks. Modern science has revealed methods of deliberately spreading or enhancing disease in ways not appreciated by our ancestors. The combination of basic research, good medical practice, and constant vigilance will be needed to defend against such attacks.

Although the potential impact of a bioterrorist attack could be enormous, leading to thousands of deaths and high morbidity rates, acts of bioterrorism would be expected to produce their greatest impact through the fear and terror they generate. In contrast to biowarfare, where the primary goal is destruction of the enemy through mass casualties, an important goal of bioterrorism is to destroy the morale of a society through fear and uncertainty. While the actual biologic impact of a single act may be small, the degree of disruption created by the realization that such an attack is possible may be enormous. This was readily apparent with the impact on the U.S. Postal Service and the functional interruption of the activities of the legislative branch of the United States government following the anthrax attacks noted above. Thus, the key to the defense against these attacks is a highly functioning system of public health surveillance and education so that attacks can be quickly recognized and effectively contained. This is complemented by the availability of appropriate countermeasures in the form of diagnostics, therapeutics, and vaccines, both in response to and in anticipation of bioterrorist attacks.

The Working Group for Civilian Biodefense has put together a list of key features that characterize the elements of biologic agents that make them particularly effective as weapons **(Table 221-1)**. Included among these are the ease of spread and transmission of the agent as well as the presence of an adequate database to allow newcomers to the field to quickly apply the good science of others to bad intentions of their own. Agents of bioterrorism may be used in their naturally occurring forms or they can be deliberately modified to provide maximal impact. Among the approaches to maximizing the deleterious effects of biologic agents are the genetic modification of microbes for the purposes of antimicrobial resistance or evasion by the immune system, creation of fine-particle aerosols, chemical treatment to stabilize and prolong infectivity, and alteration of host range through changes in surface proteins. Certain of these approaches fall under the category of *weaponization*, which is a term generally used to describe the processing of microbes or toxins in a manner that would ensure a devastating effect of a release. For example, weaponization of anthrax by the Soviets comprised the production of vast amounts of spores in a form that maintained aerosolization for prolonged periods of time; the spores were of appropriate size to reach the lower respiratory tract easily and could be delivered in a massive release, such as via widely dispersed bomblets.

Table 221-1 Key Features of Biologic Agents Used as Bioweapons

1. High morbidity and mortality rates
2. Potential for person-to-person spread
3. Low infective dose and highly infectious by aerosol
4. Lack of rapid diagnostic capability
5. Lack of universally available effective vaccine
6. Potential to cause anxiety
7. Availability of pathogen and feasibility of production
8. Environmental stability
9. Database of prior research and development
10. Potential to be "weaponized"

Source: From L Borio et al: JAMA 287:2391, 2002; with permission.

The U.S. Centers for Disease Control and Prevention (CDC) classifies potential biologic threats into three categories, A, B, and C **(Table 221-2)**. Category A agents are the highest-priority pathogens. They pose the greatest risk to national security because they (1) can be easily disseminated or transmitted from person to person, (2) result in high mortality rates and have the potential for major public health impact, (3) might cause public panic and social disruption, and (4) require special action for public health preparedness. Category B agents are the second highest priority pathogens and include those that are moderately easy to disseminate, result in moderate morbidity rates and low mortality rates, and require specifically enhanced diagnostic capacity. Category C agents are the third highest priority. These include certain emerging pathogens to which the general population lacks immunity that could be engineered for mass dissemination in the future because of availability, ease of production, to ease of dissemination, and that have a major public health impact and the potential for high morbidity and mortality rates. It should be pointed out, however, that these A, B, and C designations are empirical, and, depending on evolving circumstances such as intelligence-based threat assessments, the priority rating of any given microbe or toxin could change. The CDC classification system also largely reflects the severity of illness produced by a given agent, rather than its accessibility to potential terrorists.

Table 221-2 CDC Category A, B, and C Agents

Category A
Anthrax (*Bacillus anthracis*)
Botulism (*Clostridium botulinum* toxin)
Plague (*Yersinia pestis*)

Smallpox (*Variola major*)
Tularemia (*Francisella tularensis*)
Viral hemorrhagic fevers
Arenaviruses: Lassa, New World (Machupo, Junin, Guanarito, and Sabia) Bunyaviridae: Crimean-Congo, Rift Valley
Filoviridae: Ebola, Marburg
Category B
Brucellosis (*Brucella* spp.)
Epsilon toxin of *Clostridium perfringens*
Food safety threats (e.g., *Salmonella* spp., *Escherichia coli* 0157:H7, *Shigella*)
Glanders (*Burkholderia mallei*)
Melioidosis (*B. pseudomallei*)
Psittacosis (*Chlamydophila psittaci*)
Q fever (*Coxiella burnetii*)
Ricin toxin from *Ricinus communis* (castor beans)
Staphylococcal enterotoxin B
Typhus fever (*Rickettsia prowazekii*)
Viral encephalitis [alphaviruses (e.g., Venezuelan, eastern, and western equine encephalitis)]
Water safety threats (e.g., *Vibrio cholerae, Cryptosporidium parvum*)
Category C
Emerging infectious diseases threats such as Nipah, hantavirus, SARS coronavirus, and pandemic influenza.

Abbreviation: SARS, severe acute respiratory syndrome.

Source: Centers for Disease Control and Prevention and the National Institute of Allergy and Infectious Diseases.

CATEGORY A AGENTS

Anthrax

See also Chap. 138.

BACILLUS ANTHRACIS AS A BIOWEAPON

Anthrax may be the prototypic disease of bioterrorism. Although rarely, if ever, spread from person to person, the illness embodies the other major features of a disease introduced through terrorism, as outlined in Table 221-1. U.S. and British government scientists studied anthrax as a potential biologic weapon beginning approximately at the time of World War II (WWII). Offensive bioweapons activity including bioweapons research on microbes and toxins in the United States ceased in 1969 as a result of two executive orders by President Richard M. Nixon. Although the 1972 Biological and Toxin Weapons Convention Treaty outlawed research of this type worldwide, the Soviet Union produced and stored tons of anthrax spores for potential use as a bioweapon until at least the late 1980s. At present, there is suspicion that research on anthrax as an agent of bioterrorism is ongoing by several nations and extremist groups. One example of this is the release of anthrax spores by the Aum Shinrikyo cult in Tokyo in 1993. Fortunately, there were no casualties associated with this episode because of the inadvertent use of a nonpathogenic strain of anthrax by the terrorists.

The potential impact of anthrax spores as a bioweapon was clearly demonstrated in 1979 following the accidental release of spores into the atmosphere from a Soviet Union bioweapons facility in Sverdlosk, Russia. While actual figures are not known, at least 77 cases of anthrax were diagnosed with certainty, of which 66 were fatal. These victims were exposed in an area within 4 km downwind of the facility, and deaths due to anthrax were also noted in livestock up to 50 km further downwind. Based on recorded wind patterns, the interval between the time of exposure and development of clinical illness ranged from 2 to 43 days. The majority of cases were within the first 2 weeks. Death typically occurred within 1–4 days following the onset of symptoms. It is likely that the widespread use of postexposure penicillin prophylaxis limited the total number of cases. The extended period of time between exposure and disease in some individuals supports the data from nonhuman primate studies, suggesting that the anthrax spores can lie dormant in the respiratory tract for at least 4–6 weeks without evoking an immune response. This extended period of microbiologic latency following exposure poses a significant challenge for management of victims in the postexposure period.

In September 2001, the American public was exposed to anthrax spores as a bioweapon delivered through the U.S. Postal Service by an employee of the United States Army Research Institute for Infectious Diseases (USAMRIID) who had access to such materials and who committed suicide prior to being indicted for this crime. The CDC identified 22 confirmed or suspected cases of anthrax as a consequence of this attack. These included 11 patients with inhalational anthrax, of whom 5 died, and 11 patients with cutaneous anthrax (7 confirmed), all of whom survived **(Fig. 221-1)**. Cases occurred in individuals who opened contaminated letters as well as in postal workers involved in the processing of mail. A minimum of five letters mailed from Trenton, NJ, served as the vehicles for these attacks. One of these letters was reported to contain 2 g of material, equivalent to 100 billion to 1 trillion weapon-grade spores. Since studies performed in the 1950s using monkeys exposed to aerosolized anthrax suggested that ~10,000 spores were required to produce lethal disease in 50% of animals exposed to this dose (the LD_{50}), the contents of one letter had the theoretical potential, under optimal conditions, of causing illness or death in up to 50 million individuals when one considers an LD_{50} of 10,000 spores. The strain used in this attack was the Ames strain. Although it was noted to have an inducible β-lactamase and to constitutively express a cephalosporinase, it was susceptible to all antibiotics standard for *B. anthracis*.

Figure 221-1

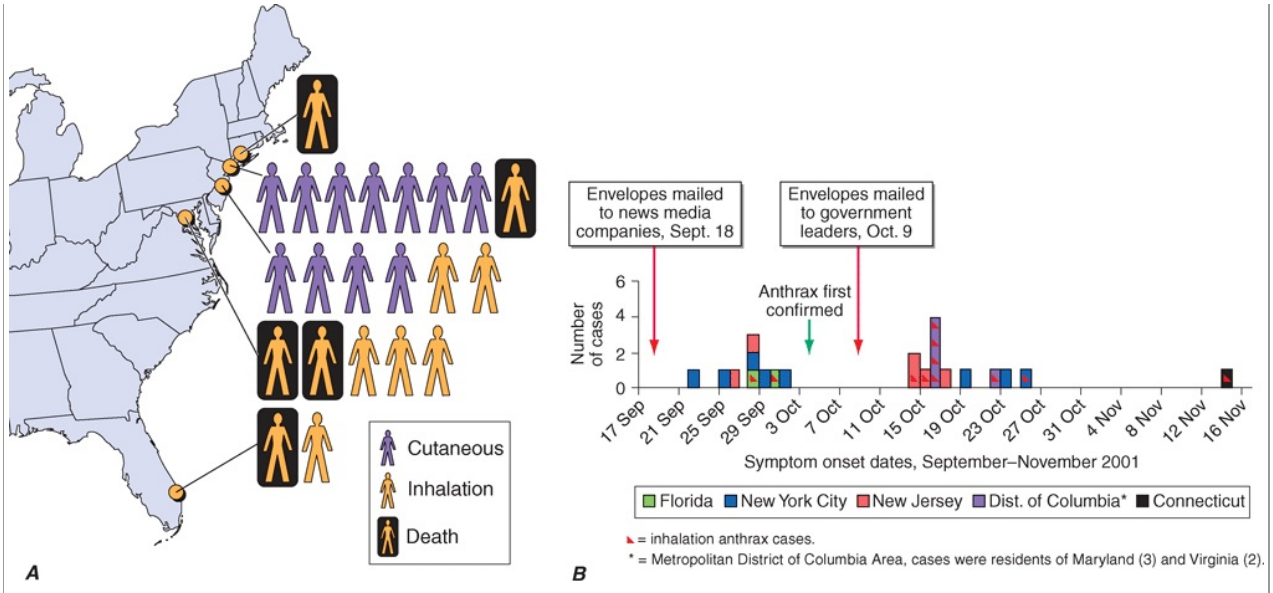

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Confirmed anthrax cases associated with bioterrorism: United States, 2001. *A.* Geographic location, clinical manifestation, and outcome of the 11 cases of confirmed inhalational and 11 cases of confirmed cutaneous anthrax. ***B.*** Epidemic curve for 22 cases of anthrax. (*From DB Jernigan et al: Emerg Infect Dis 8:1019, 2002; with permission* .)

MICROBIOLOGY AND CLINICAL FEATURES

Anthrax is caused by *B. anthracis*, a gram-positive, nonmotile, spore-forming rod that is found in soil and predominantly causes disease in herbivores such as cattle, goats, and sheep. Anthrax spores can remain viable for decades. The remarkable stability of these spores makes them an ideal bioweapon, and their destruction in decontamination activities can be a challenge. Naturally occurring human infection is generally the result of contact with anthrax-infected animals or animal products such as goat hair in textile mills or animal skins used in making drums. While an LD_{50} of 10,000 spores is a generally accepted number, it has also been suggested that as few as one to three spores may be adequate to cause disease in some settings. Advanced technology is likely to be necessary to generate spores of the optimal size (1–5 μm) to travel to the alveolar spaces as a bioweapon.

The three major clinical forms of anthrax are gastrointestinal, cutaneous, and inhalational. *Gastrointestinal anthrax* typically results from the ingestion of contaminated meat; the condition is rarely seen and is unlikely to be the result of a bioterrorism event. The lesion of *cutaneous anthrax* typically begins as a papule following the introduction of spores through an opening in the skin. This papule then evolves to a painless vesicle followed by the development of a coal-black, necrotic eschar **(Fig. 221-2)**. It is the Greek word for coal (*anthrax*) that gives the organism and the disease its name. Cutaneous anthrax was ~20% fatal prior to the availability of antibiotics. *Inhalational anthrax* is the form most likely to be responsible for death in the setting of a bioterrorist attack. It occurs following the inhalation of spores that become deposited in the alveolar spaces. These spores are phagocytosed by macrophages and transported to the mediastinal and peribronchial lymph nodes where they germinate, leading to active bacterial growth and elaboration of the bacterial products edema toxin and lethal toxin. Subsequent hematogenous spread of bacteria is accompanied by cardiovascular collapse and death. The earliest symptoms are typically a viral-like prodrome with fever, malaise, and abdominal and/or chest symptoms that progress over the course of a few days to a moribund state. A characteristic finding is mediastinal widening and pleural effusions on chest x-ray **(Fig. 221-3)**. While initially thought to be 100% fatal, the experiences at Sverdlosk in 1979 and in the United States in 2001 (see below) indicate that with prompt initiation of antibiotic therapy, survival is possible. The characteristics of the 11 cases of inhalational anthrax diagnosed in the United States in 2001 following exposure to contaminated letters postmarked September 18 or October 9, 2001, followed the classic pattern established for this illness, with patients presenting with a rapidly progressive course characterized by fever, fatigue or malaise, nausea or vomiting, cough, and shortness of breath. At presentation, the total white blood cell counts were ~ 10,000 cells/μL; transaminases tended to be elevated, and all 11 had abnormal findings on chest x-ray and CT. Radiologic findings included infiltrates, mediastinal widening, and hemorrhagic pleural effusions. For cases in which the dates of exposure were known, symptoms appeared within 4–6 days. Death occurred within 7 days of diagnosis in the five fatal cases (overall mortality rate 55%). Rapid diagnosis and prompt initiation of antibiotic therapy were key to survival.

Figure 221-2

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical manifestations of a pediatric case of cutaneous anthrax associated with the bioterrorism attack of 2001. The lesion progresses from vesicular on day 5 **(*A*)** to necrotic with the classic black eschar on day 12 **(*B*)** to a healed scar 2 months later **(*C*)**. *(Photographs provided by Dr. Mary Wu Chang and (A) reprinted with permission of the New England Journal of*

Medicine.)

Figure 221-3

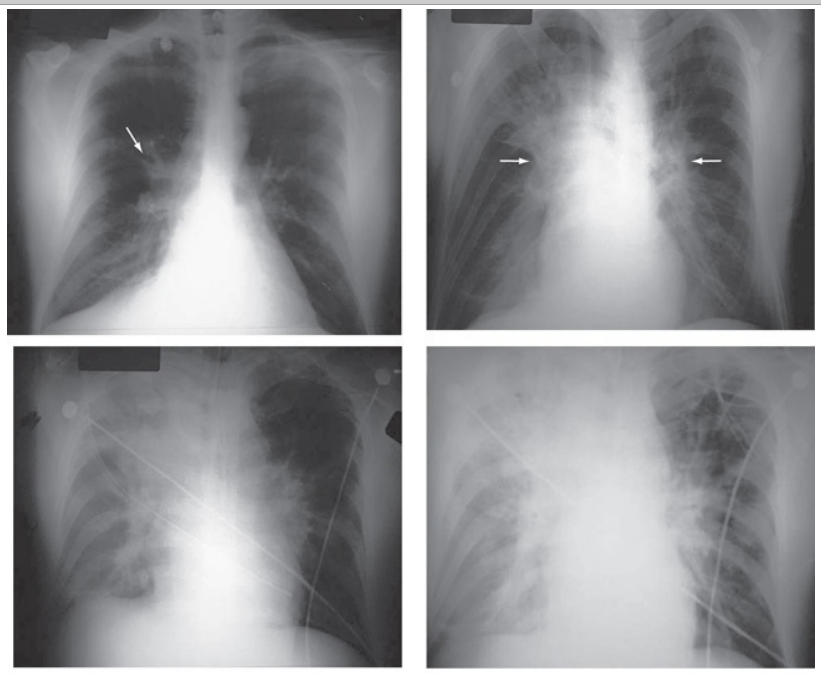

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Progression of chest x-ray findings in a patient with inhalational anthrax. Findings evolved from subtle hilar prominence and right perihilar infiltrate to a progressively widened mediastinum, marked perihilar infiltrates, peribronchial cuffing, and air bronchograms. (*From L Borio et al: JAMA 286:2554, 2001; with permission*.)

TREATMENT: ANTHRAX

Anthrax can be successfully treated if the disease is promptly recognized and appropriate therapy is initiated early. While penicillin, ciprofloxacin, and doxycycline are the currently licensed antibiotics for this indication, clindamycin and rifampin also have in vitro activity against the organism and have been used as part of treatment regimens. Until sensitivity results are known, suspected cases are best managed with a combination of broadly active agents **(Table 221-3)**. Patients with inhalational anthrax are not contagious and do not require special isolation procedures.

Table 221-3 Clinical Syndromes, Prevention, and Treatment Strategies for Diseases Caused by Category A Agents

Agent	Clinical Syndrome	Incubation Period	Diagnosis	Treatment	Prophylaxis
Bacillus anthracis (anthrax)	Cutaneous lesion: Papule to eschar	1–12 days	Culture, Gram stain, PCR, Wright stain of peripheral smear	Postexposure: Ciprofloxacin, 500 mg, PO bid x 60 d *or* Doxycycline, 100 mg PO bid x 60 d or Amoxicillin, 500 mg PO q8h, likely to be effective if strain penicillin sensitive *Active disease:* Ciprofloxacin, 400 mg IV q12h *or* Doxycycline, 100 mg IV q12h *plus* Clindamycin, 900 mg IV q8h and/or rifampin, 300 mg IV q12h; switch to PO when stable x 60 d total *Antitoxin strategies:* Neutralizing monoclonal and polyclonal antibodies are under study	Anthrax vaccine adsorbed Recombinant protective antigen vaccines are under study
	Inhalational disease: Fever, malaise, chest and abdominal discomfort Pleural effusion, widened mediastinum on chest x-ray	1–60 days			
Yersinia pestis (pneumonic plague)	Fever, cough, dyspnea, hemoptysis Infiltrates and consolidation on chest x-ray	1–6 days	Culture, Gram stain, direct fluorescent antibody, PCR	Gentamicin, 2.0 mg/kg IV loading then 1.7 mg/kg q8h IV *or* Streptomycin, 1.0 g q12h IM or IV Alternatives include	Doxycycline, 100 mg PO bid (ciprofloxacin may also be active) Formalin-fixed vaccine (FDA licensed; not

				doxycycline, 100 mg bid PO or IV; chloramphenicol, 500 mg qid PO or IV	licensed, not available)
Variola major (smallpox)	Fever, malaise, headache, backache, emesis Maculopapular to vesicular to pustular skin lesions	7–17 days	Culture, PCR, electron microscopy	Supportive measures; consideration for cidofovir, antivaccinia immunoglobulin	Vaccinia immunization
Francisella tularensis (tularemia)	Fever, chills, malaise, myalgia, chest discomfort, dyspnea, headache, skin rash, pharyngitis, conjunctivitis Hilar adenopathy on chest x-ray	1–14 days	Gram stain, culture, immunohistochemistry, PCR	Streptomycin, 1 g IM bid *or* Gentamicin, 5 mg/kg per day div q8h IV for 14 days *or* Doxycycline, 100 mg IV bid *or* Chloramphenicol, 15 mg/kg up to 1 gm IV qid *or* Ciprofloxacin, 400 mg IV bid	Doxycycline, 100 mg PO bid x 14 days *or* Ciprofloxacin, 500 mg PO bid x 14 days
Viral hemorrhagic fevers	Fever, myalgia, rash, encephalitis, prostration	2–21 days	RT-PCR, serologic testing for antigen or antibody Viral isolation by CDC or U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID)	Supportive measures Ribavirin 30 mg/kg up to 2 g x 1, followed by 16 mg/kg IV up to 1 g q6h for 4 days, followed by 8 mg/kg IV up to 0.5 g q8h x 6 days	No known chemoprophylaxis Consideration for ribavirin in high-risk situations
Botulinum toxin (*Clostridium botulinum*)	Dry mouth, blurred vision, ptosis, weakness, dysarthria, dysphagia, dizziness, respiratory failure, progressive paralysis, dilated pupils	12–72 h	Mouse bioassay, toxin immunoassay	Supportive measures including ventilation, HBAT equine antitoxin from the CDC Emergency Operations Center, 770-488-7100	Administration of antitoxin

Abbreviations: CDC, U.S. Centers for Disease Control and Prevention; FDA, U.S. Food and Drug Administration; HBAT, heptavalent botulinum antitoxin; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase PCR.

VACCINATION AND PREVENTION

The first successful vaccine for anthrax was developed for animals by Louis Pasteur in 1881. At present, the single vaccine licensed for human use is a product produced from the cell-free culture supernatant of an attenuated, nonencapsulated strain of *B. anthracis* (Stern strain), referred to as *anthrax vaccine adsorbed* (AVA). Clinical trials for safety in humans and efficacy in animals are currently under way to evaluate the role of recombinant protective antigen (one of the major components, along with lethal factor and edema factor, of *B. anthracis* toxins) as an alternative to AVA. In a postexposure setting in non-human primates, a 2-week course of AVA + ciprofloxacin was found to be superior to ciprofloxacin alone in preventing the development of clinical disease and death. While the current recommendation for postexposure prophylaxis is 60 days of antibiotics, it would seem prudent to include immunization with anthrax vaccine if available. Given the potential for *B. anthracis* to be engineered to express penicillin resistance, the empirical regimen of choice in this setting is either ciprofloxacin or doxycycline.

Plague

See also Chap. 159.

***YERSINIA PESTIS* AS A BIOWEAPON**

Although it lacks the environmental stability of anthrax, the highly contagious nature and high mortality rate of plague make it a close to ideal agent of bioterrorism, particularly if delivered in a weaponized form. Occupying a unique place in history, plague has been alleged to have been used as a biologic weapon for centuries. The catapulting of plague-infected corpses into besieged fortresses is a practice that was first noted in 1346 during the assault of the city of Kaffa by the Tartars. Although unlikely to have resulted in disease transmission, some believe that this event may have played a role in the start of the Black Death pandemic of the fourteenth and fifteenth centuries in Europe. Given that plague was already moving across Asia toward Europe at this time, it is unclear whether such an allegation is accurate. During WWII, the infamous Unit 731 of the Japanese army was reported to have repeatedly dropped plague-infested fleas over parts of China, including Manchuria. These drops were associated with subsequent outbreaks of plague in the targeted areas. Following WWII, the United States and the Soviet Union conducted programs of research on how to create aerosolized *Y. pestis* that could be used as a bioweapon to cause primary pneumonic plague. As mentioned above, plague was thought to be an excellent bioweapon due to the fact that in addition to causing infection in those inhaling the aerosol, significant numbers of secondary cases of primary pneumonic plague would likely occur due to the contagious nature of the disease and person-to-person transmission via respiratory aerosol. Secondary reports of research conducted during that time suggest that organisms remain viable for up to 1 h and can be dispersed for distances up to 10 km. While the offensive bioweapons program in the United States was terminated prior to production of sufficient quantities of plague organisms for use as a weapon, it is believed that Soviet scientists did manufacture quantities sufficient for such a purpose. It has also been reported that more than 10 Soviet Institutes and >1000 scientists were working with plague as a biologic weapon. Of concern is the fact that in 1995 a microbiologist in Ohio was arrested for having obtained *Y. pestis* in the mail from the American Type Culture Collection, using a credit card and a false letterhead. In the wake of this incident, the U.S. Congress passed a law in 1997 requiring that anyone intending to send or receive any of 42 different agents that could potentially be used as bioweapons first register with the CDC.

MICROBIOLOGY AND CLINICAL FEATURES

Plague is caused by *Y. pestis*, a nonmotile, gram-negative bacillus that exhibits bipolar, or "safety pin," staining with Wright, Giemsa, or Wayson stains. It has had a major impact on the course of history, thus adding to the element of fear evoked by its mention. The earliest reported plague epidemic was in 224 b.c. in China. The most infamous pandemic began in Europe in the fourteenth century, during which time one-third to one-half of the entire population of Europe was killed. During a plague outbreak in India in 1994, even though the number of confirmed cases was relatively small, it is estimated that 500,000 individuals fled their homes in fear of this disease.

The clinical syndromes of plague generally reflect the mode of infection. *Bubonic plague* is the consequence of an insect bite;

primary *pneumonic plague* arises through the inhalation of bacteria. Most of the plague seen in the world today is bubonic plague and is the result of a bite by a plague-infected flea. In part as a consequence of past pandemics, plague infection of rodents exists widely in nature, including in the southwestern United States, and each year thousands of cases of plague occur worldwide through contact with infected animals or fleas. Following inoculation of regurgitated bacteria into the skin by a flea bite, organisms travel through the lymphatics to regional lymph nodes, where they are phagocytized but not destroyed. Inside the cell, they multiply rapidly leading to inflammation, painful lymphadenopathy with necrosis, fever, bacteremia, septicemia, and death. The characteristic enlarged, inflamed lymph nodes, or *buboes*, give this form of plague its name. In some instances, patients may develop bacteremia without lymphadenopathy following infection, a condition referred to as *primary septicemic plague*. Extensive ecchymoses may develop due to disseminated intravascular coagulation, and gangrene of the digits and/or nose may develop in patients with advanced septicemic plague. It is thought that this appearance of some patients gave rise to the term *Black Death* in reference to the plague epidemic of the fourteenth and fifteenth centuries. Some patients may develop pneumonia (secondary pneumonic plague) as a complication of bubonic or septicemic plague. These patients may then transmit the agent to others via the respiratory route, causing cases of primary pneumonic plague. Primary pneumonic plague is the manifestation most likely to occur as the result of a bioterrorist attack, with an aerosol of bacteria spread over a wide area or a particular environment that is densely populated. In this setting, patients would be expected to develop fever, cough with hemoptysis, dyspnea, and gastrointestinal symptoms 1–6 days following exposure. Clinical features of pneumonia would be accompanied by pulmonary infiltrates and consolidation on chest x-ray. In the absence of antibiotics, the mortality rate of this disease is on the order of 85%, and death usually occurs within 2–6 days.

TREATMENT: PLAGUE

Streptomycin, tetracycline, and doxycycline are licensed by the U.S. Food and Drug Administration (FDA) for the treatment of plague. Multiple additional antibiotics licensed for other infections are commonly used and are likely effective. Among these are aminoglycosides such as gentamicin, cephalosporins, trimethoprim/sulfamethoxazole, chloramphenicol, and ciprofloxacin (Table 221-3). A multidrug-resistant strain of *Y. pestis* was identified in 1995 from a patient with bubonic plague in Madagascar. While this organism was resistant to streptomycin, ampicillin, chloramphenicol, sulfonamides, and tetracycline, it retained its susceptibility to other aminoglycosides and cephalosporins. Given the subsequent identification of a similar organism in 1997 coupled with the fact that this resistance is plasmid-mediated, it seems likely that genetically modifying *Y. pestis* to a multidrug-resistant form is possible. Unlike patients with inhalational anthrax (see above), patients with pulmonary plague should be cared for under conditions of strict respiratory isolation comparable to that used for multidrug-resistant tuberculosis.

VACCINATION AND PREVENTION

A formalin-fixed, whole-organism vaccine was licensed by the FDA for the prevention of plague. That vaccine is no longer being manufactured, but its potential value as a current countermeasure against bioterrorism would likely have been modest at best as it was ineffective against animal models of primary pneumonic plague. Efforts are under way to develop a second generation of vaccines that will protect against aerosol challenge. Among the candidates being tested are recombinant forms of the F1 and V antigens of *Y. pestis*. It is likely that doxycycline or ciprofloxacin would provide coverage in a chemoprophylaxis setting. Unlike the case with anthrax, in which one has to be concerned about the persistence of ungerminated spores in the respiratory tract, the duration of prophylaxis against plague need only extend to 7 days following exposure.

Smallpox

See also Chap. 183.

VARIOLA VIRUS AS A BIOWEAPON

Given that most of the world's population was once vaccinated against smallpox, variola virus would not have been considered a good candidate as a bioweapon 30 years ago. However, with the cessation of immunization programs in the United States in 1972 and throughout the world in 1980 due to the successful global eradication of smallpox, close to 50% of the U.S. population is fully susceptible to smallpox today. Given its infectious nature and the 10–30% mortality rate in unimmunized individuals, the deliberate spread of this virus could have a devastating effect on our society and unleash a previously conquered deadly disease. It is estimated that an initial infection of 50–100 persons in a first generation of cases could expand by a factor of 10–20 with each succeeding generation in the absence of any effective containment measures. While the likely implementation of an effective public health response makes this scenario unlikely, it does illustrate the potential damage and disruption that can result from a smallpox outbreak.

In 1980, the World Health Organization (WHO) recommended that all immunization programs be terminated; that representative samples of variola virus be transferred to two locations, one at the CDC in Atlanta, GA, in the United States and the other at the Institute of Virus Preparations in the Soviet Union; and that all other stocks of smallpox be destroyed. Several years later, it was recommended that these two authorized collections be destroyed. However, these latter recommendations were placed on hold in the wake of increased concerns on the use of variola virus as a biologic weapon and thus the need to maintain an active program of defensive research. Many of these concerns were based upon allegations made by former Soviet officials that extensive programs had been in place in that country for the production and weaponization of large quantities of smallpox virus. The dismantling of these programs with the fall of the Soviet Union and the subsequent weakening of security measures led to fears that stocks of *V. major* may have made their way to other countries or terrorist organizations. In addition, accounts that efforts had been taken to produce recombinant strains of *Variola* that would be more virulent and more contagious than the wild-type virus have led to an increase in the need to be vigilant for the reemergence of this often fatal infectious disease.

MICROBIOLOGY AND CLINICAL FEATURES

Smallpox is caused by one of two variants of variola virus, *V. major* and *V. minor*. Variola is a double-strand DNA virus and member of the Orthopoxvirus genus of the Poxviridae family. Infections with *V. minor* are generally less severe than those of *V. major*, with milder constitutional symptoms and lower mortality rates; thus *V. major* is the only one considered to be a viable bioweapon. Infection with *V. major* typically occurs following contact with an infected person. Patients are infectious from the time that a maculopapular rash appears on the skin and oropharynx through the resolution and scabbing of the pustular lesions. Infection occurs principally during close contact, through the inhalation of saliva droplets containing virus from the oropharyngeal exanthem. Aerosolized material from contaminated clothing or linen can also spread infection. Several days after exposure, a primary viremia is believed to occur that results in dissemination of virus to lymphoid tissues. A secondary viremia occurs ~4 days later that leads to localization of infection in the dermis. Approximately 12–14 days following the initial exposure, the patient develops high fever, malaise, vomiting, headache, backache, and a maculopapular rash that begins on the face and extremities and spreads to the trunk (centripetal) with lesions in the same developmental stage in any given location. This is in contrast to the rash of varicella (chickenpox) that begins on the trunk and face and spreads to the extremities (centrifugal) with lesions at all stages of development. The lesions are initially maculopapular and evolve to vesicles that eventually become pustules and then scabs. The oral mucosa also develops maculopapular lesions that evolve to ulcers. The lesions appear over a period of 1–2 days and evolve at the same rate. Although virus can be isolated from the scabs on the skin, the conventional thinking is that once the scabs have formed the patient is no longer contagious. Smallpox is associated with 10–30% mortality rates, with patients typically dying of severe systemic illness during the second week of symptoms. Historically, ~5–10% of naturally occurring smallpox cases take either of two highly virulent atypical forms, classified as *hemorrhagic* and *malignant*. These are

difficult to diagnose because of their atypical presentations. The hemorrhagic form is uniformly fatal and begins with the relatively abrupt onset of a severely prostrating illness characterized by high fevers and severe headache and back and abdominal pain. This form of the illness resembles a severe systemic inflammatory syndrome, in which patients have a high viremia but die without developing the characteristic rash. Cutaneous erythema develops accompanied by petechiae and hemorrhages into the skin and mucous membranes. Death usually occurs within 5–6 days. The malignant, or "flat," form of smallpox is frequently fatal and has an onset similar to the hemorrhagic form, but with confluent skin lesions developing more slowly and never progressing to the pustular stage.

TREATMENT: SMALLPOX

Given the infectious nature of smallpox and the extreme vulnerability of contemporary society, patients who are suspected cases should be handled with strict isolation procedures. While laboratory confirmation of a suspected case by culture, PCR and electron microscopy is essential, it is equally important that appropriate precautions be employed when obtaining samples for culture and laboratory testing. All health care and laboratory workers caring for patients should have been recently immunized with vaccinia, and all samples should be transported in doubly sealed containers. Patients should be cared for in negative-pressure rooms with strict isolation precautions.

There is no licensed specific therapy for smallpox, and historic treatments have focused solely on supportive care. While several antiviral agents, including cidofovir, that are licensed for other diseases have in vitro activity against *V. major*, they have never been tested in the setting of human disease. For this reason, it is difficult to predict whether or not they would be effective in cases of smallpox and, if effective, whether or not they would be of value in patients with advanced disease. Research programs studying the efficacy of new antiviral compounds (ST-246 and others) against *V. major* are currently under way.

VACCINATION AND PREVENTION

In 1796, Edward Jenner demonstrated that deliberate infection with cowpox virus could prevent illness on subsequent exposure to smallpox. Today, smallpox is a preventable disease following immunization with vaccinia. The current dilemma facing our society regarding assessment of the risk and benefit of smallpox vaccination is that the degree of risk that someone will deliberately and effectively release smallpox into our society is unknown. As a prudent first step in preparedness for a smallpox attack, virtually all members of the U.S. armed services have received primary or booster immunizations with vaccinia. In addition, tens of thousands of civilian health care workers who comprise smallpox-response teams at the state and local public health level have been vaccinated.

Initial fears regarding the immunization of a segment of the American population with vaccinia when there are more individuals receiving immunosuppressive drugs and other immunocompromised patients than ever before were dispelled by the data generated from the military and civilian immunization campaigns of 2002–2004. Adverse event rates for the first 450,000 immunizations were similar to and, in certain categories of adverse events, even lower than those from prior historic data, in which most severe sequelae of vaccination occurred in young infants **(Table 221-4)**. In addition, 11 patients with early-stage HIV infection were inadvertently immunized without problem. One significant concern during that immunization campaign, however, was the description of a syndrome of myopericarditis, which had not been appreciated during prior immunization campaigns with vaccinia. In an effort to provide a safer vaccine to protect against smallpox, ACAM 2000, a cloned virus propagated in tissue culture, was developed and became the first second-generation smallpox vaccine to be licensed. This vaccine is now used by the U.S. military and is part of the U.S. government stockpile. Research continues on attenuated forms of vaccinia such as modified vaccinia Ankara (MVA). Vaccinia immune globulin is available to treat those who experience a severe reaction to immunization with vaccinia.

Table 221-4 Complications from 438,134 Administrations of Vaccinia during the United States Department of Defense (DOD) Smallpox Immunization Campaign Initiated in December 2002

Complication	Number of Cases	DOD Rate per Million Vaccinees (95% Confidence Interval)	Historic Rate Per Million Vaccinees
Mild or temporary:			
Generalized vaccinia, mild	35	67 (52, 85)	45–212[a]
Inadvertent inoculation, self	62	119 (98, 142)	606[a]
Vaccinia transfer to contact	28	53 (40, 69)	8–27[a]
Moderate or serious:			
Encephalitis	1	2.2 (0.6, 7.2)	2.6–8.7[a]
Acute myopericarditis	69	131 (110, 155)	100[b]
Eczema vaccinatum	0	0 (0, 3.7)	2–35[a]
Progressive vaccinia	0	0 (0, 3.7)	1–7[a]
Death	1[c]	1.9 (0.2, 5.6)	1–2[a]

[a] Based on adolescent and adult smallpox vaccinations from 1968 studies, both primary and revaccinations.

[b] Based on case series in Finnish military recruits given the Finnish strain of smallpox vaccine.

[c] Potentially attributable to vaccination; after lupus-like illness.

Source: From JD Grabenstein and W Winkenwerder: *http://www.smallpox.mil/event/SPSafetySum.asp.*

Tularemia

See also Chap. 158.

***FRANCISELLA TULARENSIS* AS A BIOWEAPON**

Tularemia has been studied as an agent of bioterrorism since the mid-twentieth century. It has been speculated by some that the outbreak of tularemia among German and Soviet soldiers during fighting on the Eastern Front during WWII was the consequence of a deliberate release. Unit 731 of the Japanese Army studied the use of tularemia as a bioweapon during WWII. Large preparations were made for mass production of *F. tularensis* by the United States, but no stockpiling of any agent took place. Stocks of *F. tularensis* were reportedly generated by the Soviet Union in the mid-1950s. It has also been suggested that the Soviet program extended into the era of molecular biology and that some strains were engineered to be resistant to common

antibiotics. *F. tularensis* is an extremely infectious organism, and human infections have occurred from merely examining an uncovered petri dish streaked with colonies. Given these facts, it is reasonable to conclude that this organism might be utilized as a bioweapon through either an aerosol or contamination of food or drinking water.

MICROBIOLOGY AND CLINICAL FEATURES

While similar in many ways to anthrax and plague, tularemia, also referred to as rabbit fever or deer fly fever, is neither as lethal nor as fulminant as either of these other two category A bacterial infections. It is, however, extremely infectious, and as few as 10 organisms can lead to establishment of infection. Despite this fact, it is not spread from person to person. Tularemia is caused by *F. tularensis*, a small, nonmotile, gram-negative coccobacillus. Although it is not a spore-forming organism, it is a hardy bacterium that can survive for weeks in the environment. Infection typically comes from insect bites or contact with organisms in the environment. Infections have occurred in laboratory workers studying the agent. Large waterborne outbreaks have been recorded. It is most likely that the outbreak among German and Russian soldiers and Russian civilians noted above during WWII represented a large waterborne tularemia outbreak in a *Tularensis*-enzootic area devastated by warfare.

Humans can become infected through a variety of environmental sources. Infection is most common in rural areas where a variety of small mammals may serve as reservoirs. Human infections in the summer are often the result of insect bites from ticks, flies, or mosquitoes that have bitten infected animals. In colder months, infections are most likely the result of direct contact with infected mammals and are most common in hunters. In these settings, infection typically presents as a systemic illness with an area of inflammation and necrosis at the site of tissue entry. Drinking of contaminated water may lead to an oropharyngeal form of tularemia characterized by pharyngitis with cervical and/or retropharyngeal lymphadenopathy (Chap. 158). The most likely mode of dissemination of tularemia as a biologic weapon would be as an aerosol, as has occurred in a number of natural outbreaks in rural areas, including Martha's Vineyard in the United States. Approximately 1–14 days following exposure by this route, one would expect to see inflammation of the airways with pharyngitis, pleuritis, and bronchopneumonia. Typical symptoms would include the abrupt onset of fever, fatigue, chills, headache, and malaise (Table 221-3). Some patients might experience conjunctivitis with ulceration, pharyngitis, and/or cutaneous exanthems. A pulse-temperature dissociation might be present. Approximately 50% of patients would show a pulmonary infiltrate on chest x-ray. Hilar adenopathy might also be present, and a small percentage of patients could have adenopathy without infiltrates. The highly variable presentation makes acute recognition of aerosol-disseminated tularemia very difficult. The diagnosis would likely be made by immunohistochemistry or culture of infected tissues or blood. Untreated, mortality rates range from 5 to 15% for cutaneous routes of infection and from 30 to 60% for infection by inhalation. Since the advent of antibiotic therapy, these rates have dropped to <2%.

TREATMENT: TULAREMIA

Both streptomycin and doxycycline are licensed for treatment of tularemia. Other agents likely to be effective include gentamicin, chloramphenicol, and ciprofloxacin (Table 221-3). Given the potential for genetic modification of this organism to yield antibiotic-resistant strains, broad-spectrum coverage should be the rule until sensitivities have been determined. As mentioned above, special isolation procedures are not required.

VACCINATION AND PREVENTION

There are no vaccines currently licensed for the prevention of tularemia. While a live, attenuated strain of the organism has been used in the past with some reported success, there are inadequate data to support its widespread use at this time. Development of a vaccine for this agent is an important part of the current biodefense research agenda. In the absence of an effective vaccine, postexposure chemoprophylaxis with either doxycycline or ciprofloxacin appears to be a reasonable approach (Table 221-3).

Viral Hemorrhagic Fevers

See also Chaps. 196 and 197.

HEMORRHAGIC FEVER VIRUSES AS BIOWEAPONS

Several of the hemorrhagic fever viruses have been reported to have been weaponized by the Soviet Union and the United States. Nonhuman primate studies indicate that infection can be established with very few virions and that infectious aerosol preparations can be produced. Under the guise of wanting to aid victims of an Ebola outbreak, members of the Aum Shinrikyo cult in Japan were reported to have traveled to central Africa in 1992 in an attempt to obtain Ebola virus for use in a bioterrorist attack. Thus, while there has been no evidence that these agents have ever been used in a biologic attack, there is clear interest in their potential for this purpose.

MICROBIOLOGY AND CLINICAL FEATURES

The viral hemorrhagic fevers are a group of illnesses caused by any one of a number of similar viruses (Table 221-2). These viruses are all enveloped, single-strand RNA viruses that are thought to depend upon a host reservoir for long-term survival. While rodents or insects have been indentified as the hosts for some of these viruses, for others the hosts are unknown. These viruses tend to be geographically restricted according to the migration patterns of their hosts. Great apes are not a natural reservoir for Ebola virus, but large numbers of these animals in sub-Saharan Africa have died from Ebola infection over the past decade. Humans can become infected with hemorrhagic fever viruses if they come into contact with an infected host or other infected animals. Person-to-person transmission, largely through direct contact with virus-containing body fluids, has been documented for Ebola, Marburg, and Lassa viruses and rarely for the New World arenaviruses. While there is no clear evidence of respiratory spread among humans, these viruses have been shown in animal models to be highly infectious by the aerosol route. This, coupled with mortality rates as high as 90%, makes them excellent candidate agents of bioterrorism.

The clinical features of the viral hemorrhagic fevers vary depending upon the particular agent (Table 221-3). Initial signs and symptoms typically include fever, myalgia, prostration, and disseminated intravascular coagulation with thrombocytopenia and capillary hemorrhage. These findings are consistent with a cytokine-mediated systemic inflammatory syndrome. A variety of different maculopapular or erythematous rashes may be seen. Leukopenia, temperature-pulse dissociation, renal failure, and seizures may also be part of the clinical presentation.

Outbreaks of most of these diseases are sporadic and unpredictable. As a consequence, most studies of pathogenesis have been performed using laboratory animals. The diagnosis should be suspected in anyone with temperature >38.3°C for <3 weeks who also exhibits at least two of the following: hemorrhagic or purpuric rash, epistaxis, hematemesis, hemoptysis, or hematochezia in the absence of any other identifiable cause. In this setting, samples of blood should be sent after consultation to the CDC or the USAMRIID for serologic testing for antigen and antibody as well as reverse transcriptasepolymerase chain reaction (RT-PCR) testing for hemorrhagic fever viruses. All samples should be handled with double-bagging. Given how little is known regarding the human-to-human transmission of these viruses, appropriate isolation measures would include full barrier precautions with negative-pressure rooms and use of powered air-purifying respirators (PAPRs). Unprotected skin contact with cadavers has been implicated in the transmission of certain hemorrhagic fever viruses such as Ebola, so it is recommended that autopsies of suspected cases be performed using the strictest measures for protection and that burial or cremation be performed promptly without embalming.

TREATMENT: VIRAL HEMORRHAGIC FEVERS

There are no approved and effective antiviral therapies for this class of viruses (Table 221-3). While there are anecdotal reports of the efficacy of ribavirin, interferon-α, or hyperimmune immunoglobulin, definitive data are lacking. The best data for ribavirin are in arenavirus (Lassa and New World) infections. In some in vitro systems, specific immunoglobulin has been reported to enhance infectivity, and thus these potential treatments must be approached with caution.

VACCINATION AND PREVENTION

There are no licensed and effective vaccines for these agents. Studies are currently under way examining the potential role of DNA, recombinant viruses, and attenuated viruses as vaccines for several of these infections. Among the most promising at present are vaccines for Argentine, Ebola, Rift Valley, and Kayasanur Forest viruses.

Botulism Toxin (*Clostridium Botulinum*)

See also Chap. 141.

BOTULINUM TOXIN AS A BIOWEAPON

In a bioterrorist attack, botulinum toxin would likely be dispersed as an aerosol or as contamination of a food supply. While contamination of a water supply is possible, it is likely that any toxin would be rapidly inactivated by the chlorine used to purify drinking water. Similarly, toxin can be inactivated by heating any food to >85°C for >5 min. Without external facilitation, the environmental decay rate is estimated at 1% per minute, and thus the time interval between weapon release and ingestion or inhalation needs to be rather short. The Japanese biologic warfare group, Unit 731, is reported to have conducted experiments on botulism poisoning in prisoners in the 1930s. The United States and the Soviet Union both acknowledged producing botulinum toxin, and there is some evidence that the Soviet Union attempted to create recombinant bacteria containing the gene for botulinum toxin. In records submitted to the United Nations, Iraq admitted to having produced 19,000 L of concentrated toxin'enough toxin to kill the entire population of the world three times over. By many accounts, botulinum toxin was the primary focus of the pre-1991 Iraqi bioweapons program. In addition to these examples of state-supported research into the use of botulinum toxin as a bioweapon, the Aum Shinrikyo cult unsuccessfully attempted on at least three occasions to disperse botulinum toxin into the civilian population of Tokyo.

MICROBIOLOGY AND CLINICAL FEATURES

Unique among the category A agents for not being a live microorganism, botulinum toxin is one of the most potent toxins ever described and is thought by some to be the most poisonous substance in existence. It is estimated that 1 g of botulinum toxin would be sufficient to kill 1 million individuals if adequately dispersed. Botulinum toxin is produced by the gram-positive, spore-forming anaerobe *C. botulinum* (Chap. 141). Its natural habitat is soil. There are seven antigenically distinct forms of botulinum toxin, designated A–G. The majority of naturally occurring human cases are of types A, B, and E. Antitoxin directed toward one of these will have little to no activity against the others. The toxin is a 150-kDa zinc-containing protease that prevents the intracellular fusion of acetylcholine vesicles with the motor neuron membrane, thus preventing the release of acetylcholine. In the absence of acetylcholine-dependent triggering of muscle fibers, a flaccid paralysis develops. Although botulism does not spread from person to person, the ease of production of the toxin coupled with its high morbidity and 60–100% mortality make it a close to ideal bioweapon.

Botulism can result from the growth of *C. botulinum* infection in a wound or the intestine, the ingestion of contaminated food, or the inhalation of aerosolized toxin. The latter two forms are the most likely modes of transmission for bioterrorism. Once toxin is absorbed into the bloodstream, it binds to the neuronal cell membrane, enters the cell, and cleaves one of the proteins required for the intracellular binding of the synaptic vesicle to the cell membrane, thus preventing release of the neurotransmitter to the membrane of the adjacent muscle cell. Patients initially develop multiple cranial nerve palsies that are followed by a descending flaccid paralysis. The extent of the neuromuscular compromise is dependent upon the level of toxemia. The majority of patients experience diplopia, dysphagia, dysarthria, dry mouth, ptosis, dilated pupils, fatigue, and extremity weakness. There are minimal true central nervous system effects, and patients rarely show significant alterations in mental status. Severe cases can involve complete muscular collapse, loss of the gag reflex, and respiratory failure, requiring weeks or months of ventilator support. Recovery requires the regeneration of new motor neuron synapses with the muscle cell, a process that can take weeks to months. In the absence of secondary infections, which may be common during the protracted recovery phase of this illness, patients remain afebrile. The diagnosis is suspected on clinical grounds and confirmed by a mouse bioassay or toxin immunoassay.

TREATMENT: BOTULISM

Treatment for botulism is mainly supportive and may require intubation, mechanical ventilation, and parenteral nutrition (Table 221-3). If diagnosed early enough, administration of equine antitoxin may reduce the extent of nerve injury and decrease the severity of disease. At present, a heptavalent botulinum antitoxin (HBAT) is available through the CDC as an investigational agent for treatment of naturally occurring noninfant botulism. HBAT contains horse serum—derived antibody fragments to all seven known botulinum toxins (A—G). It is composed of <2% intact immunoglobulin and ≥90% Fab and $F(ab')_2$ immunoglobulin fragments. A single dose of antitoxin is usually adequate to neutralize any circulating toxin. Repeat dosing may be needed in a setting of continued toxin exposure. Given that this product is derived from horse serum, one needs to be vigilant for hypersensitivity reactions, including serum sickness and anaphylaxis following its administration. Once the damage to the nerve axon has been done, however, there is little possible in the way of specific therapy. At this point, vigilance for secondary complications such as infections during the protracted recovery phase is of the utmost importance. Due to their ability to worsen neuromuscular blockade, aminoglycosides and clindamycin should be avoided in the treatment of these infections.

VACCINATION AND PREVENTION

A botulinum toxoid preparation has been used as a vaccine for laboratory workers at high risk of exposure and in certain military situations; however, it is not currently available in quantities that could be used for the general population. At present, early recognition of the clinical syndrome and use of appropriate equine antitoxin is the mainstay of prevention of full-blown disease in exposed individuals. The development of human monoclonal antibodies as a replacement for equine antitoxin antibodies is an area of active research interest.

CATEGORY B AND C AGENTS

The category B agents include those that are easy or moderately easy to disseminate and result in moderate morbidity and low mortality rates. A listing of the current category B agents is provided in Table 221-2. As can be seen, it includes a wide array of microorganisms and products of microorganisms. Several of these agents have been used in bioterrorist attacks, although never with the impact of the Category A agents described above. Among the more notorious of these was the contamination of salad bars in Oregon in 1984 with *Salmonella typhimurium* by the religious cult Rajneeshee. In this outbreak, which many consider to be the first bioterrorist attack against U.S. citizens, >750 individuals were poisoned and 40 were hospitalized in an effort to influence a local election. The intentional nature of this outbreak went unrecognized for more than a decade.

Category C agents are the third highest priority agents in the biodefense agenda. These agents include emerging pathogens to which little or no immunity exists in the general population, such as the severe acute respiratory syndrome (SARS) coronavirus or pandemic-potential strains of influenza that could be obtained from nature and deliberately disseminated. These agents are

characterized as being relatively easy to produce and disseminate, having high morbidity and mortality rates and having a significant public health impact. There is no running list of category C agents at the present time.

PREVENTION AND PREPAREDNESS

As noted above, a large and diverse array of agents has the potential to be used in a bioterrorist attack. In contrast to the military situation with biowarfare, where the primary objective is to inflict mass casualties on a healthy and prepared militia, the objectives of bioterrorism are to harm civilians as well as to create fear and disruption among the civilian population. While the military needs only to prepare their troops to deal with the limited number of agents that pose a legitimate threat of biowarfare, the public health system needs to prepare the entire civilian population to deal with the multitude of agents and settings that could be utilized in a bioterrorism attack. This includes anticipating issues specific to the very young and the very old, the pregnant patient, and the immunocompromised individual. The challenges in this regard are enormous and immediate. While military preparedness emphasizes vaccines toward a limited number of agents, civilian preparedness needs to rely upon rapid diagnosis and treatment of a wide array of conditions.

The medical profession must maintain a high index of suspicion that unusual clinical presentations or the clustering of cases of a rare disease may not be a chance occurrence but rather the first sign of a bioterrorist event. This is particularly true when such diseases occur in traditionally healthy populations, when surprisingly large numbers of rare conditions occur, and when diseases commonly seen in rural settings appear in urban populations. Given the importance of rapid diagnosis and early treatment for many of these conditions, it is essential that the medical care team report any suspected cases of bioterrorism immediately to local and state health authorities and/or to the CDC (888-246-2675). Enhancements have been made to the public health surveillance network to facilitate the rapid sharing of information among public health agencies.

At present a series of efforts are in place to ensure the biomedical security of the civilian population of the United States. The Public Health Service is moving toward a more highly trained, fully deployable force. The Strategic National Stockpile (SNS) maintained by the CDC provides rapid access to quantities of pharmaceuticals, antidotes, vaccines, and other medical supplies that may be of value in the event of biologic or chemical terrorism. The SNS has two basic components. The first of these consists of "push packages" that can be deployed anywhere in the United States within 12 h. These push packages are a preassembled set of supplies, pharmaceuticals, and medical equipment ready for immediate delivery to the field. They provide treatment for a variety of conditions given the fact that an actual threat may not have been precisely identified at the time of stockpile deployment. The contents of the push packs are constantly updated to ensure that they reflect current needs as determined by national security threat assessments; they include antibiotics for treatment of anthrax, plague, and tularemia as well as a cache of vaccine to deal with a smallpox threat. The second component of the SNS comprises inventories managed by specific vendors and consists of the provision of additional pharmaceuticals, supplies, and/or products tailored to the specific attack.

The number of FDA-approved and -licensed drugs and vaccines for category A and B agents is currently limited and not reflective of the pharmacy of today. In an effort to speed the licensure of additional drugs and vaccines for these diseases, the FDA has a rule for the licensure of such countermeasures against agents of bioterrorism when adequate and well-controlled clinical efficacy studies cannot be ethically conducted in humans. This is commonly referred to as the "Animal Rule." Thus, for indications in which field trials of prophylaxis or therapy for a naturally occurring disease are not feasible, the FDA will rely on evidence solely from laboratory animal studies. For this rule to apply, it must be shown that (1) there are reasonably well-understood pathophysiologic mechanisms for the condition and its treatment; (2) the effect of the intervention is independently substantiated in at least two animal species, including species expected to react with a response predictive for humans; (3) the animal study endpoint is clearly related to the desired benefit in humans; and (4) the data in animals allow selection of an effective dose in humans.

Finally, the Biomedical Advanced Research and Development Authority (BARDA) was established within the U.S. Department of Health and Human Services to provide an integrated, systematic approach to the development and purchase of the necessary vaccines, drugs, therapies, and diagnostic tools for public health medical emergencies. BARDA manages Project BioShield, an initiative established to facilitate biodefense research within the federal government, create a stable source of funding for the purchase of countermeasures against agents of bioterrorism, and create a category of "emergency use authorization" to allow the FDA to approve the use of unlicensed countermeasures during times of extraordinary unmet needs, as might be present in the context of a bioterrorist attack.

While the prospect of a deliberate attack on civilians with disease-producing agents may seem to be an act of incomprehensible evil, history shows us that it is something that has been done in the past and will likely be done again in the future. It is the responsibility of health care providers to be aware of this possibility, to be able to recognize early signs of a potential bioterrorist attack and alert the public health system, and to respond quickly to provide care to the individual patient. Among the web sites with current information on microbial bioterrorism are *www.bt.cdc.gov*, *www.niaid.nih.gov*, and *www.cidrap.umn.edu*.

FURTHER READINGS

Alibek K, Handelman S: *Biohazard: The Chilling True Story of the Largest Covert Biological Weapons in the World, Told from the Inside by the Man Who Ran It*. New York, Random House, 1999

Croddy E (with C Perey-Armendariz and J Hart): *Chemical and Biological Warfare: A Comprehensive Survey for the Concerned Citizen*. New York, Copernicus Books, 2001

Henderson DA et al (eds): *Bioterrorism: Guidelines for Medical and Public Health Management* . JAMA and Archives Journals, AMA Press, 2002

Jernigan JA et al: Bioterrorism-related inhalational anthrax: The first 10 cases reported in the United States. Emerg Infect Dis 7:933, 2001[PMID: 11747719]

Kennedy RB et al: Smallpox vaccines for biodefense. Vaccine 27: D73, 2009

Nalca A, Zumbrun EE: ACAM2000™: The new smallpox vaccine for United States Strategic National Stockpile. Drug Des Devel Ther 4:71, 2010[PMID: 20531961]

Rusnak JM et al: Risk of occupationally acquired illnesses from biological threat agents in unvaccinated laboratory workers. Biosecur Bioterror 2:281, 2004[PMID: 15650438]

Wilkening DA: Sverdlovsk revisited: Modeling human inhalation anthrax. Proc Natl Acad Sci USA 103:20, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHEMICAL TERRORISM: INTRODUCTION

The use of chemical warfare agents (CWAs) in modern warfare dates back to World War I (WWI). Most recently, sulfur mustard and nerve agents were used by Iraq against the Iranian military and Kurdish civilians. Since the Japanese sarin attacks in 1994–1995 and the terrorist strikes of September 11, 2001, the all too real possibility of chemical or biological terrorism against civilian populations anywhere in the world has attracted increased attention.

Military planners consider the WWI blistering agent sulfur mustard and the organophosphorus nerve agents as the most likely agents to be used on the battlefield. In a civilian or terrorist scenario, the choice widens considerably. For example, many of the chemical warfare agents of WWI, including chlorine, phosgene, and cyanide, are used today in large amounts in industry. They are produced in chemical plants, are stockpiled in large tanks, and travel up and down highways and railways in large tanker cars. The rupture of any of these agents by accident or purposely could cause many injuries and deaths. Countless hazardous materials (HAZMATs) that are not used on the battlefield can be used as terrorist weapons. Some of them, including insecticides and ammonia, could wreak as much damage and injury as the weaponized chemical agents.

In a recent example, insurgents in Iraq used chlorine gas released from tankers after explosions as a crude form of chemical weaponry. Using this gas, they killed 12 people and intoxicated more than 140 others in three attacks in February 2007. Table 222-1 describes the physical appearance and initial physiologic effects, and Table 222-2 provides guidelines for immediate treatment of chlorine intoxication. Nonetheless, the focus of this chapter is on the blister and nerve chemical warfare agents, which have been employed in battle and against civilians and have demonstrated a significant public health impact.

Table 222-1 Recognizing and Diagnosing Health Effects of Chemical Terrorism

Agent	Agent Name	Unique Characteristics	Initial Effects
Nerve	Cyclohexyl sarin (GF)[a] Sarin (GB) Soman (GD) Tabun (GA) VX VR	Miosis (pinpoint pupils) Copious secretions Muscle twitching/fasciculations	Miosis (pinpoint pupils) Blurred/dim vision Headache Nausea, vomiting, diarrhea Copious secretions/sweating Muscle twitching/fasciculations Breathing difficulty Seizures
Asphyxiant/blood	Arsine Cyanogen chloride Hydrogen cyanide	Possible cherry-red skin Possible cyanosis Possible frostbite[b]	Confusion Nausea Patients may gasp for air, similar to asphyxiation but more abrupt onset Seizures before death
Choking/pulmonary-damaging	Chlorine Hydrogen chloride Nitrogen oxides Phosgene	Chlorine is a greenish-yellow gas with pungent odor Phosgene gas smells like newly mown hay or grass Possible frostbite[a]	Eye and skin irritation Airway irritation Dyspnea, cough Sore throat Chest tightness

Blistering/vesicant	Mustard/sulfur mustard (HD, H) Mustard gas (H) Nitrogen mustard (HN-1, HN-2, HN-3) Lewisite (L) Phosgene oxime (CX)	Mustard (HD) has an odor like burning garlic or horseradish Lewisite (L) has an odor like penetrating geranium Phosgene oxime (CX) has a pepperish or pungent odor	Severe irritation Redness and blisters of the skin Tearing, conjunctivitis, corneal damage Mild respiratory distress to marked airway damage May cause death
Incapacitating/behavior-altering	Agent 15/BZ	May appear as mass drug intoxication with erratic behaviors, shared realistic and distinct hallucinations, disrobing, and confusion Hyperthermia Mydriasis (dilated pupils)	Dry mouth and skin Initial tachycardia Altered consciousness, delusions, denial of illness, belligerence Hyperthermia Ataxia (lack of coordination) Hallucinations Mydriasis (dilated pupils)

[a] Letters in parentheses indicate NATO codes for designated agents.

[b] Frostbite may occur from skin contact with liquid arsine, cyanogen chloride, or phosgene.

Source: State of New York, Department of Health, as modified by the Chemical Casualty Care Division, U.S. Army Medical Research Institute of Chemical Defense.

Table 222-2 Decontamination and Treatment of Chemical Terrorism

Agent	Decontamination	First Aid	Other Patient Considerations
Nerve	Remove clothing immediately Gently wash skin with soap and water Do not abrade skin For eyes, flush with plenty of water or normal saline	Atropine before other measures Pralidoxime (2-PAM) chloride	Onset of symptoms from dermal contact with liquid forms may be delayed Repeated antidote administration may be necessary
Asphyxiant/blood	Remove clothing immediately if no frostbite[a] Gently wash skin with soap and water Do not abrade skin For eyes, flush with plenty of water or normal saline	Rapid treatment with oxygen For cyanide, use antidotes (sodium nitrite and then sodium thiosulfate)	Arsine and cyanogen chloride may cause delayed pulmonary edema
Choking/pulmonary-damaging	Remove clothing immediately if no	Fresh air, forced rest Semi-upright position	May cause delayed pulmonary edema, even after

	frostbite[a] Gently wash skin with soap and water Do not abrade skin For eyes, flush with plenty of water or normal saline	Semiupright position If signs of respiratory distress are present, oxygen with or without positive airway pressure may be needed Other supportive therapy, as needed	a symptom-free period that varies in duration with the amount inhaled
Blistering/vesicant	Immediate decontamination is essential to minimize damage Remove clothing immediately Gently wash skin with soap and water Do not abrade skin For eyes, flush with plenty of water or normal saline	Immediately decontaminate skin Flush eyes with water or normal saline for 10–15 min If breathing difficulty, give oxygen Supportive care	Mustard has an asymptomatic latent period There is no antidote or treatment for mustard Lewisite has immediate burning pain, blisters later Specific antidote British anti-lewisite (BAL) may decrease systemic effects of lewisite Phosgene oxime causes immediate pain Possible pulmonary edema
Incapacitating/behavior-altering	Remove clothing immediately Gently wash skin with water or soap and water Do not abrade skin	Remove heavy clothing Evaluate mental status Use restraints as needed Monitor core temperature carefully Supportive care	Hyperthermia and self-injury are largest risks Hard to detect because it is an odorless and nonirritating substance Possible serious arrhythmias Specific antidote (physostigmine) may be available

[a] For frostbite areas, DO NOT remove any adhering clothing. Wash area with plenty of warm water to release clothing.

Source: State of New York, Department of Health.

Many mistakenly believe that chemical attacks will always be so severe that little can be done except to bury the dead. History proves the opposite. Even in WWI, when IV fluids, endotracheal tubes, and antibiotics were unavailable, the mortality rate among U.S. forces on the battlefield from chemical warfare agents, chiefly sulfur mustard and the pulmonary intoxicants, was only 1.9%. That was far lower than the 7% mortality rate from conventional wounds. In the 1995 Tokyo subway sarin incident, among the 5500 patients who sought medical attention at hospitals, 80% of whom were not actually symptomatic, only 12 died. Recent events should produce not a fatalistic attitude but a realistic wish to understand the pathophysiology of the syndromes these agents cause, with a view to treating expeditiously all patients who present for care and an expectation of saving the vast majority. As we prepare to defend our civilian population from the effects of chemical terrorism, we also must consider the fact that terrorism itself can produce sequelae such as physiologic or neurologic effects that may resemble the effects of nonlethal exposures to CWAs. These effects are due to a general fear of chemicals, fear of decontamination, fear of protective ensemble, or other phobic reactions.

Many writers have pointed out the increased difficulty in differentiating between stress reactions and nerve agent—induced organic brain syndromes. Knowledge of the behavioral effects of CWAs and their medical countermeasures is imperative to ensure that military and civilian medical and mental health organizations can deal with possible incidents involving weapons of mass destruction.

For the reader's benefit, the chemical warfare agents, their North Atlantic Treaty Organization (NATO) codes, and their initial effects are listed in **Table 222-1**.

Vesicants

SULFUR MUSTARD

Sulfur mustard has been a military threat since it first appeared on the battlefield in Belgium during

WWI. In modern times it remains a threat on the battlefield as well as a potential terrorist threat for bioterrorism because of simplicity of manufacture and extreme effectiveness. Sulfur mustard accounted for 70% of the 1.3 million chemical casualties in WWI. Occasional cases occur in the United States in people exposed to WWI and WWII-era munitions.

Mechanism

Sulfur mustard constitutes both a vapor and a liquid threat to all exposed epithelial surfaces. The effects are delayed, appearing hours after exposure. The organs most commonly affected are the skin (with erythema and vesicles), eyes (ranging from mild conjunctivitis to severe eye damage), and airways (ranging from mild upper airway irritation to severe bronchiolar damage). After exposure to large quantities of mustard, precursor cells of the bone marrow are damaged, leading to pancytopenia and secondary infection. The gastrointestinal mucosa may be damaged, and there are sometimes central nervous system (CNS) signs of unknown mechanism. No specific antidotes exist; management is entirely supportive. Immediate decontamination of the liquid is the only way to reduce damage. Complete decontamination in 2 minutes stops clinical injury; decontamination at 5 minutes will reduce skin injury by ~50%. **Table 222-2** lists approaches to decontamination of mustard and other CWAs.

Mustard dissolves slowly in aqueous media such as sweat, but once dissolved, it rapidly forms extremely reactive cyclic ethylene sulfonium ions, which react with cell proteins, cell membranes, and especially DNA in rapidly dividing cells. The ability of mustard to react with and alkylate DNA gives rise to the effects by which it has been characterized as "radiomimetic," similar to radiation injury. Mustard has many biologic actions, but its actual mechanism of action is largely unknown. Much of the biologic damage from mustard results from DNA alkylation and cross-linking in rapidly dividing cells: corneal epithelium, basal keratinocytes, bronchial mucosal epithelium, gastrointestinal mucosal epithelium, and bone marrow precursor cells. This may lead to cellular death and inflammatory reactions. In the skin, proteolytic digestion of anchoring filaments at the epidermal-dermal junction may be the major mechanism of action resulting in blister formation. Mustard also has mild cholinergic activity, which may be responsible for effects such as early gastrointestinal and CNS symptoms.

Mustard reacts with tissue within minutes of entering the body. Its circulating half-life in unaltered form is extremely brief.

Clinical Features

Topical effects of mustard occur in the skin, airways, and eyes, with the eyes being most sensitive, followed by the airways. Absorbed mustard may produce effects in the bone marrow, gastrointestinal tract, and CNS. Direct injury to the gastrointestinal tract also may occur after ingestion of the compound through contamination of water or food.

Erythema is the mildest and earliest form of mustard skin injury. It resembles sunburn and is associated with pruritus, burning, or stinging pain. Erythema begins to appear within 2 hours to 2 days after vapor exposure. Time of onset depends on severity of exposure, ambient temperature and humidity, and type of skin. The most sensitive sites are the warm moist locations and thin delicate skin, such as the perineum, external genitalia, axillae, antecubital fossae, and neck.

Within the erythematous areas, small vesicles can develop, which may later coalesce to form bullae **(Fig. 222-1)**. The typical bulla is large, dome-shaped, flaccid, thin-walled, translucent, and surrounded by erythema. The blister fluid, a transudate, is clear to straw colored and becomes yellow, tending to coagulate. The fluid does not contain mustard and is not itself a vesicant. Lesions from high-dose liquid exposure may develop a central zone of coagulation necrosis with blister formation at the periphery. These lesions take longer to heal and are more prone to secondary infection than are the uncomplicated lesions seen at lower exposure levels. Severe lesions may require skin grafting.

Figure 222-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Large bulla formation from mustard burn in a patient. Although the blisters on this patient involved only 7% of the body surface area, this patient still required hospitalization in a burn intensive care unit.

The primary airway lesion is necrosis of the mucosa with possible damage to underlying smooth muscle. The damage begins in the upper airways and descends to the lower airways in a dose-dependent manner. Usually the terminal airways and alveoli are affected only as a terminal event. Pulmonary edema is not usually present unless the damage is very severe, and then it becomes hemorrhagic.

The earliest effects from mustard and perhaps the only effects from a low concentration involve the nose, sinuses, and pharynx. There may be irritation or burning of the nares, epistaxis, sinus pain, and pharyngeal pain. As the concentration increases, laryngitis, voice changes, and nonproductive cough develop. Damage to the trachea and upper bronchi leads to a productive cough. Lower airway involvement causes dyspnea, severe cough, and increasing quantities of sputum. Terminally, there may be necrosis of the smaller airways with hemorrhagic edema into surrounding alveoli. Hemorrhagic pulmonary edema is rare.

Necrosis of airway mucosa causes "pseudomembrane" formation. These membranes may cause obstruction of the bronchi. During WWI, high-dose mustard exposure caused acute death via this mechanism in a small minority of cases **(Fig. 222-2)**.

Figure 222-2

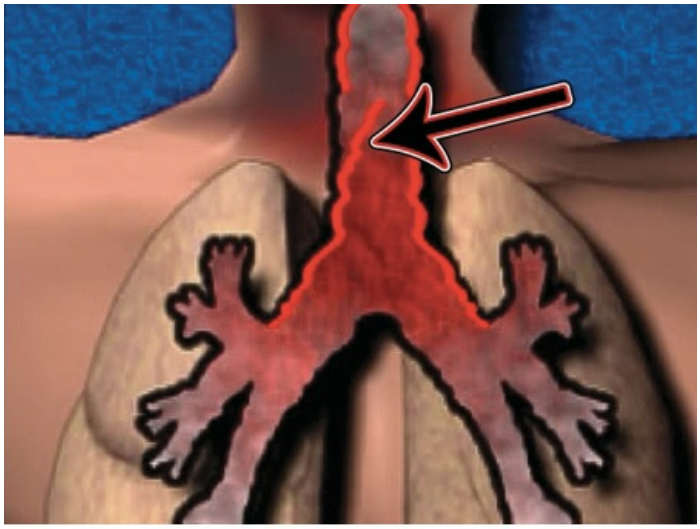

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of pseudomembrane formation as is seen in high-dose sulfur mustard vapor inhalation exposure. In World War I, severe inhalation exposure often caused death via obstruction of large airways.

The eyes are the organs most sensitive to mustard vapor injury. The latent period is shorter for eye injury than for skin injury and is also exposure concentration—dependent. After low-dose vapor exposure, irritation evidenced by reddening of the eyes may be the only effect. As the dose increases, the injury includes progressively more severe conjunctivitis, photophobia, blepharospasm, pain, and corneal damage **(Fig. 222-3)**.

Figure 222-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

World War I photograph of troops exposed to sulfur mustard vapor. The vast majority of these troops survived with no long-term damage to the eyes; however, they were effectively rendered blind for days to weeks.

About 90% of eye injuries related to mustard heal in 2 weeks to 2 months without sequelae. Scarring between the iris and the lens may follow severe effects; this scarring may restrict pupillary movements and predispose victims to glaucoma. The most severe damage is caused by liquid mustard. After extensive eye exposure, severe corneal damage with possible perforation of the cornea and loss of the eye can occur. In some individuals, a latent chronic keratitis sometimes associated with corneal ulcerations has been described as early as 8 months and as late as 20 years after initial exposure.

The mucosa of the gastrointestinal tract is susceptible to mustard damage from either systemic absorption or ingestion of the agent. Mustard exposure in small amounts will cause nausea and vomiting lasting up to 24 hours. The mechanism of the nausea and vomiting is not understood, but mustard does have a cholinergic-like effect. The CNS effects of mustard also remain poorly defined. Large exposures can cause seizures in animals. Reports from WWI and from the Iran-Iraq war described people exposed to small amounts of mustard acting sluggish, apathetic, and lethargic. These reports suggest that minor psychological problems could linger for a year or longer.

The cause of death in the majority of mustard poisoning cases is sepsis and respiratory failure. Mechanical obstruction via pseudomembrane formation and agent-induced laryngospasm is important in the first 24 hours, but only in cases of severe exposure. From the third through the fifth day after exposure, one can expect a secondary pneumonia due to bacterial invasion of denuded necrotic mucosa. The third wave of death is caused by agent-induced bone marrow suppression, which peaks 7–21 days after exposure and causes death via sepsis.

Treatment: Sulfur Mustard

A patient severely ill from mustard poisoning requires the general supportive care provided for any severely ill patient as well as the specific care given to a burn patient. Liberal use of systemic analgesics, maintenance of fluid and electrolyte balance, nutrition, appropriate antibiotics, and other supportive measures are necessary (Table 222-2).

The management of a patient exposed to mustard may range from simple, as in the provision of symptomatic care for a sunburn-like erythema, to complex, as in providing total management for a severely ill patient with burns, immunosuppression, and multisystem involvement. Before raw denuded areas of skin develop, especially with less severe exposures, topical cortisone creams or lotions may be of benefit. Some very basic research data point to the early use of anti-inflammatory preparations. Small blisters (<1–2 cm) should be left intact. Because larger bullae eventually will break, they should be unroofed carefully. Denuded areas should be irrigated three to four times daily with saline, other sterile solutions, or soapy water and then liberally covered with the topical antibiotic of choice, such as silver sulfadiazine or mafenide acetate, to a thickness of 1–2 mm. Some physicians advocate sterile needle drainage of large blisters, collapsing the blister roof to form a sterile dressing. Mustard blister

fluid does not contain sulfur mustard, only sterile tissue fluid. Health care staff should not fear possible contamination. If an antibiotic cream is not available, sterile petrolatum will be useful. Modified Dakins solution (sodium hypochlorite 0.5%) was used both in WWI and in Iranian casualties (1984–1987) for field-expedient irrigation and antisepsis. Large areas of vesication require hospitalization, IV therapy, and whirlpool bath irrigation.

Systemic analgesics should be used liberally, particularly before manipulation of the patient. Monitoring of fluids and electrolytes is important in any sick patient, but it must be recognized that fluid loss is not of the magnitude seen with thermal burns. Overly rigorous hydration seems to have precipitated pulmonary edema in a few Iranian casualties sent to European hospitals.

Conjunctival irritation from a low vapor exposure will respond to any of a number of available ophthalmic solutions after the eyes are irrigated thoroughly. A topical antibiotic applied several times a day will reduce the incidence and severity of infection. Animal laboratory data have shown remarkable results with commercially available topical antibiotic/glucocorticoid ophthalmologic ointments applied early. An ophthalmologist should be consulted. Topical glucocorticoids are not of proven value, but their use during the first few hours or days might significantly reduce inflammation and subsequent damage. Further use should be relegated to an ophthalmologist.

Vaseline or a similar substance should be applied regularly to the edges of the lids to prevent them from sticking together. Topical analgesics may be useful initially if blepharospasm is too severe to permit an adequate examination; however, topical analgesics have limited value.

A productive cough and dyspnea accompanied by fever and leukocytosis occurring within 12–24 hours are indicative of a chemical pneumonitis. The clinician must resist the urge to use prophylactic antibiotics for this process. Infection often occurs on the third to fifth day and is signaled by an increased fever, pulmonary infiltrate, and an increase in sputum production with a change in color. Appropriate antibiotic therapy should await confirmation by Gram stain and, later, positive culture and sensitivity.

Intubation may be necessary if laryngeal spasm or edema makes it difficult or becomes life threatening. Intubation permits better ventilation and facilitates suction of the necrotic and inflammatory debris. Early use of positive end-expiratory pressure (PEEP) or continuous positive airway pressure (CPAP) may be beneficial. Pseudomembrane formation may require fiberoptic bronchoscopy for suctioning of the necrotic debris.

Bronchodilators are of benefit for bronchospasm. If additional relief of bronchospasm is needed, glucocorticoids should be used. There is little evidence that the routine use of glucocorticoids is beneficial, except for additional relief of bronchospasm.

Leukopenia begins around day 3 with major systemic absorption. Marrow suppression peaks at 7–14 days. In the Iran-Iraq war, a white blood cell count of ≥200/μL usually resulted in death of the patient. Sterilization of the gut by nonabsorbable antibiotics should be considered to reduce the possibility of sepsis from enteric organisms. Cellular replacement (bone marrow transplants or transfusions) may be successful. Granulocyte colony-stimulating factor (G-CSF) produced a 50% reduction in the time for the bone marrow to recover in nonhuman primates exposed to sulfur mustard. Medication for nausea and vomiting may be necessary for gastrointestinal side effects. Lymphopenia precedes general leukopenia by a day or more, and may be a useful clinical tip-off to impending leukopenia.

Excellent assessments of the contributions of DNA alkylation, inflammation, activation of proteolytic enzymes, or lipid peroxidation to the mustard injury have been developed in the last 15–20 years. Some examples include (1) the demonstration of up to 75% reduction of inflammation and tissue damage in the mouse ear swelling test by vanilloid compounds and (2) the demonstration of 50–60% protection by *N*-acetylcysteine in the generation of free radicals within guinea pig lung exposed to mustard. In many cases, the demonstration of protection is dependent on the availability of sufficient amounts of drug with adequate half-lives. Strategies to enhance bioavailability include attachment of polyethylene glycol to the antioxidant drug/enzyme and delivery of the drug/enzyme in a liposome (or both).

Nerve Agents

The organophosphorus nerve agents are the deadliest of the CWAs. They work by inhibition of tissue synaptic acetylcholinesterase, creating an acute cholinergic crisis. Death ensues because of respiratory depression and can occur within seconds to minutes.

The nerve agents tabun and sarin were first used on the battlefield by Iraq against Iran during the first Persian Gulf War (1984–1987). Estimates of casualties from these agents range from 20,000 –100,000. In 1994 and 1995, the Japanese cult Aum Shinrikyo used sarin in two terrorist attacks in Matsumoto and Tokyo. Two US soldiers were exposed to sarin while rendering safe an improvised explosive device in Iraq in 2004.

The "classic" nerve agents include tabun (GA), sarin (GB), soman (GD), cyclosarin (GF), and VX. VR, similar to VX, was manufactured in the former Soviet Union (Table 222-1). The two-letter codes were established by a NATO international convention and convey no clinical implications. All the nerve

agents are organophosphorus compounds, which are liquid at standard temperature and pressure. The "G" agents evaporate at about the rate of water, except for GF, which is oily and thus probably will have evaporated within 24 hours after deposition on the ground. Their high volatility thus makes a spill of any amount a serious vapor hazard. In the Tokyo subway attack in which sarin was used, 100% of the symptomatic patients inhaled sarin vapor that spilled out on the floor of the subway cars. VX, an oily liquid, is the exception. Its low vapor pressure makes it much less of a vapor hazard but potentially a greater environmental hazard because it persists in the environment far longer.

Mechanism

Acetylcholinesterase inhibition accounts for the major life-threatening effects of nerve agent poisoning. Reversal of this inhibition by antidotal therapy is effective, proving that this is the primary toxic action of these poisons. At cholinergic synapses, acetylcholinesterase, bound to the postsynaptic membrane, functions as a turn-off switch to regulate cholinergic transmission. Inhibition of acetylcholinesterases causes the released neurotransmitter, acetylcholine, to accumulate abnormally. End organ overstimulation, which is recognized by clinicians as a cholinergic crisis, ensues **(Fig. 222-4)**.

Figure 222-4

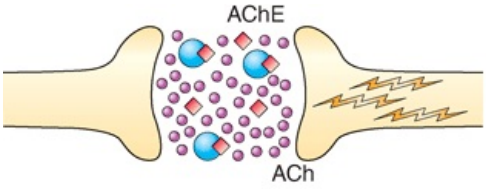

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of the pathophysiology of nerve agent exposure. Nerve agent (◆) binds to the active site of acetylcholinesterase (AChE), which is shown as floating free in space but is in reality a postsynaptic membrane-bound enzyme. As a result, acetylcholine (●), which normally is released from presynaptic membrane but normally is degraded, accumulates, and this leads (↯) to organ overstimulation and cholinergic crisis.

Clinical Features

Clinical effects of nerve agent exposure are identical for vapor and liquid exposure routes if the dose is sufficiently large. The speed and order of symptom onset will differ (Table 222-2).

Exposure of a patient to nerve agent vapor, overwhelmingly the more likely route of exposure in both battlefield and terrorist scenarios, will cause cholinergic symptoms in the order in which the toxin encounters cholinergic synapses. The most exposed synapses on the integument of the human are in the pupillary muscles. Nerve agent vapor easily crosses the cornea, interacts with these synapses, and produces miosis, described by Tokyo subway victims as "the world going black." Rarely, this also can cause eye pain and nausea. Exocrine glands in the nose, mouth, and pharynx next become exposed to the vapor, and cholinergic overload here causes increased secretions, rhinorrhea, excess salivation, and drooling. Next, toxin interacts with exocrine glands in the upper airway, causing bronchorrhea, and with bronchial smooth muscle, causing bronchospasm. This combination of events can cause hypoxia.

Once the victim has inhaled, vapor can passively cross the alveolar-capillary membrane, enter the bloodstream, and incidentally and asymptomatically inhibit circulating cholinesterases, particularly free butyrylcholinesterase and erythrocyte acetylcholinesterase, both of which can be assayed. Unfortunately, the assay may not be easily interpreted without a baseline, since cholines-terase levels vary enormously between subjects and over time in an individual, healthy patient.

Usually the first organ system to become symptomatic from bloodborne nerve agent exposure is the gastrointestinal tract, where cholinergic overload causes abdominal cramping and pain, nausea, vomiting, and diarrhea. After the gastrointestinal tract is involved, nerve agents will affect the heart, distant exocrine glands, muscles, and brain. Because there are cholinergic synapses on both the vagal (parasympathetic) and sympathetic sides of the autonomic input to the heart, one cannot predict how heart rate and blood pressure will change once intoxication has occurred. Remote exocrine activity will include oversecretion in the salivary, nasal, respiratory, and sweat glands—the patient will be "wet all over." Bloodborne nerve agents will overstimulate neuromuscular junctions in skeletal muscles, causing fasciculations followed by frank twitching. If the process goes on long enough, eventually ATP in muscles will be depleted and flaccid paralysis will ensue.

In the brain, since the cholinergic system is so widely distributed, bloodborne nerve agents will, in sufficient doses, cause rapid loss of consciousness, seizures, and central apnea leading to death within minutes. If respiration is supported, status epilepticus that does not respond to usual anticonvulsants may ensue (Chap. 369). If status epilepticus persists, neuronal death and permanent

brain dysfunction may occur. Even in mild nerve agent intoxication, patients may recover but may experience weeks of irritability, sleep disturbance, and nonspecific neurobehavioral symptoms.

The time from exposure to development of the full-blown cholinergic crisis from nerve agent vapor inhalation can be minutes or even seconds, yet there is no depot effect. Since nerve agents have a short circulating half-life, if the patient is supported and, ideally, treated with antidotes, improvement should be rapid without subsequent deterioration.

Liquid exposure to nerve agents differs in speed and order of symptom onset. A nerve agent on intact skin will partially evaporate and partially begin to travel through the skin, causing localized sweating and then localized fasciculations when it encounters neuromuscular junctions. Once in muscle, it will cross into the circulation and cause gastrointestinal discomfort, respiratory distress, heart rate changes, generalized fasciculations and twitching, loss of consciousness, seizures, and central apnea. The time course will be much longer than with vapor inhalation; even a large, lethal droplet can take up to 30 minutes to have an effect, and a small, sublethal dose could continue to take effect over 18 hours. Clinical worsening that occurs hours after treatment has started is far more likely with liquid than with vapor exposure. Additionally, miosis, which is practically unavoidable with vapor exposure, is not always present with liquid exposure and may be the last symptom to present in this situation. This is due to the relative insulation of the pupillary muscle from the systemic circulation.

Unless removed by specific therapy (oximes), binding of a nerve agent to cholinesterase is essentially irreversible. Erythrocyte acetylcholinesterase activity recovers at about 1% per day. Plasma butyrylcholinesterase recovers more quickly and is a better guide to recovery of tissue enzyme activity.

TREATMENT: NERVE AGENTS

Acute nerve agent poisoning is treated by decontamination, respiratory support, and three antidotes: an anticholinergic, an oxime, and an anticonvulsant (Tables 222-3 and **222-4**). In acute cases, all these forms of therapy may be given simultaneously.

Table 222-3 Antidote Recommendations After Exposure to Cyanide

Patient	Mild (Conscious)	Severe (Unconscious)	Other Treatment
Child	If patient is conscious and has no other signs or symptoms, antidotes may not be necessary.	Sodium nitrite[a]: 0.12–0.33 mL/kg, not to exceed 10 mL of 3% solution[b] slow IV over no less than 5 min, or slower if hypotension develops *and* Sodium thiosulfate: 1.65 mL/kg of 25% solution IV over 10–20 min	Nasal oxygen supplementation For sodium nitrite–induced orthostatic hypotension, normal saline infusion and supine position are recommended. If still apneic after antidote administration, consider sodium bicarbonate for severe acidosis.
Adult	If patient is conscious and has no other signs or symptoms, antidotes may not be necessary.	Sodium nitrite[a]: 10–20 mL of 3% solution[b] slow IV over no less than 5 min, or slower if hypotension develops *and* Sodium thiosulfate: 50 mL of 25% solution IV over 10–20 min Alternate: hydroxocobalamin 5g in reconstituted solution IV over 15 minutes	For amyl nitrite, inhaled ampules titrated to need or until other forms of IV therapy are initiated

[a] If sodium nitrite is unavailable, administer amyl nitrite by inhalation from crushable ampoules.

[b] Available in Pasadena Cyanide Antidote Kit, formerly Lilly Cyanide Kit.

Note: Victims whose clothing or skin is contaminated with hydrogen cyanide liquid or solution can secondarily contaminate response personnel by direct contact or through off-gassing vapors. Avoid dermal contact with cyanide-contaminated victims or with gastric contents of victims who may have ingested cyanide-containing materials. Victims exposed only to hydrogen cyanide gas do not pose contamination risks to rescuers. *If the patient is a victim of recent smoke inhalation (may have high carboxyhemoglobin levels), administer only sodium thiosulfate.*

Source: State of New York, Department of Health.

Table 222-4 Antidote Recommendations After Exposure to Nerve Agents

		Antidotes	
Patient Age	**Mild/Moderate Effects[a]**	**Severe Effects[b]**	**Other Treatment**
Infants (0 –2 years)	Atropine: 0.05 mg/kg IM, *or* 0.02 mg/kg IV; *and* 2-PAM chloride: 15 mg/kg IM *or* IV slowly	Atropine: 0.1 mg/kg IM, *or* 0.02 mg/kg IV; *and* 2-PAM chloride: 25 mg/kg IM, *or* 15 mg/kg IV slowly	Assisted ventilation after antidotes for severe exposure.
Child (2 –10 years)	Atropine: 1 mg IM, *or* 0.02 mg/kg IV; *and* 2-PAM chloride[c]: 15 mg/kg IM *or* IV slowly	Atropine: 2 mg IM, *or* 0.02 mg/kg IV; *and* 2-PAM chloride[c]: 25 mg/kg IM, *or* 15 mg/kg IV slowly	Repeat atropine (2 mg IM, *or* 1 mg IM for infants) at 5- to 10-min intervals until secretions have diminished and breathing is comfortable *or* airway resistance has returned to near normal.
Adolescent (>10 years)	Atropine: 2 mg IM, *or* 0.02 mg/kg IV; *and* 2-PAM chloride[c]: 15 mg/kg IM *or* IV slowly	Atropine: 4 mg IM, *or* 0.02 mg/kg IV; *and* 2-PAM chloride[c]: 25 mg/kg IM, *or* 15 mg/kg IV slowly	
Adult	Atropine: 2 to 4 mg IM *or* IV; *and* 2-PAM chloride: 600 mg IM, *or* 15 mg/kg IV slowly	Atropine: 6 mg IM; *and* 2-PAM chloride: 1800 mg IM, *or* 15 mg/kg IV slowly	Phentolamine for 2-PAM-induced hypertension: (5 mg IV for adults; 1 mg IV for children). Diazepam for convulsions: (0.2 to 0.5 mg IV for infants <5 years: 1 mg IV for children >5 years; 5 mg IV for adults).
Elderly, frail	Atropine: 1 mg IM; *and* 2-PAM chloride: 10 mg/kg IM, *or* 5 to 10 mg/kg IV slowly	Atropine: 2 to 4 mg IM; and 2-PAM chloride: 25 mg/kg IM, *or* 5 to 10 mg/kg IV slowly	

[a] Mild/moderate effects include localized sweating, muscle fasciculations, nausea, vomiting, weakness, dyspnea.

[b] Severe effects include unconsciousness, convulsions, apnea, flaccid paralysis.

[c] If calculated dose exceeds the adult IM dose, adjust accordingly.

Note: 2-PAM chloride is pralidoxime chloride or protopam chloride.

Source: State of New York, Department of Health.

Decontamination

Decontamination of a vapor is formally not necessary, but in the Tokyo subway attack, sarin vapor trapped in patients' clothing caused miosis in 10% of emergency personnel. Removal of clothing would have obviated most of this problem. Decontamination of liquid is accomplished in the military by using Reactive Skin Decontamination Lotion (RSDL), which absorbs liquid off skin. Civilian agencies now stockpile this product approved by the U.S. Food and Drug Administration (FDA). At hospitals, soap and copious amounts of water should suffice. RSDL is only approved for spot decontamination of small areas; whole-body decontamination requires water with or without soap. Physical removal of the agent is superior to all known decontamination solutions and lotions. In any event, decontamination must be accomplished before the patient enters the hospital facility to avoid contaminating the facility and its staff. In patients with contaminated wounds, one should extract potentially contaminated clothing and other foreign material that may serve as a depot for the liquid agent.

Respiratory Support

Death from nerve agent poisoning is almost always from respiratory causes. Ventilation will be complicated by increased resistance and secretions. Atropine should be given before ventilation or as ventilation begins, since it will make ventilation far easier.

Antidotal Therapy

Atropine

In theory, any anticholinergic could be used to treat nerve agent poisoning, but worldwide the choice is invariably atropine due to its wide temperature stability and rapid effectiveness either IM or IV and because inadvertent administration of this drug usually causes little CNS dysfunction (Table 222-4). Atropine rapidly reverses cholinergic overload at muscarinic synapses but has little effect at nicotinic synapses. Practically, this implies that atropine can quickly treat the life-threatening respiratory effects of nerve agents but probably will not help with neuromuscular and possibly sympathetic effects. In the field, military personnel are given MARK I kits **(Fig. 222-5*A*)** containing 2 mg atropine in autoinjector

form for IM use. Civilian agencies are now stockpiling this FDA-approved product as well. The MARK 1 kit is being replaced for both civilian and military use by a combined autoinjector containing both 2.1 mg atropine and oxime (2-PAM Cl), licensed by the FDA under the trade name Duodote(TM). One can only give full autoinjector doses and not divide them. The field-loading dose is 2, 4, or 6 mg, with retreatment every 5–10 minutes until the patient's breathing and secretions improve. The Iranians used larger doses initially during the Iran-Iraq war, in which oximes were in short supply. When the patient reaches a level of medical care at which drugs can be given IV, this is the preferred route; in small children this may be the initial form of atropine therapy. However, pediatric autoinjectors of 0.5 mg and 1 mg are now manufactured. There is no upper limit to atropine therapy in a patient either IM or IV; however, a total average adult dose for a severely afflicted patient would usually be 20–30 mg.

Figure 222-5

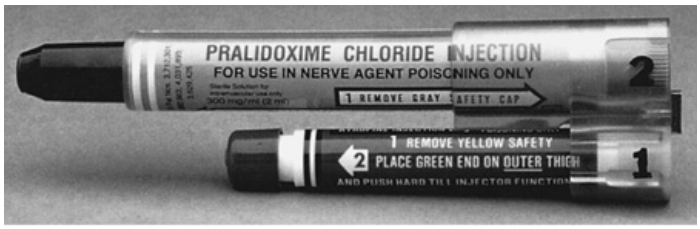

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

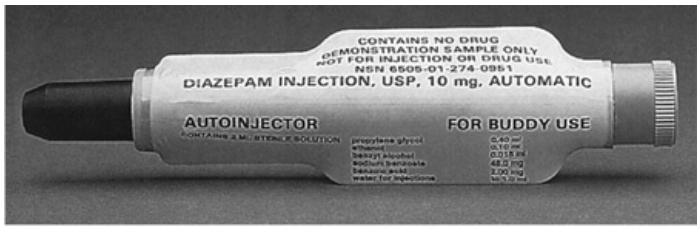

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Antidotes to nerve agents. ***A.*** MARK I autoinjector set containing one 600-mg dose of 2-pralidoxime chloride and one 2-mg dose of atropine. Such sets are carried by all U.S. military forces in a potentially chemical battlefield and are being stockpiled by civilian first responders. ***B.*** Diazepam 10-mg autoinjector. These are carried by all U.S. military forces in a potential chemical battlefield and are being stockpiled by civilian first responders.

In a mildly afflicted patient with miosis and no other systemic symptoms, atropine or homatropine eyedrops may suffice for therapy. This will produce ~24 hours of mydriasis. Frank miosis or imperfect accommodation may persist for weeks or even months after all other signs and symptoms have resolved.

Oxime Therapy

Oximes are nucleophiles that reactivate the cholinesterase whose active site has been occupied and bound to nerve agent (Table 222-4). Therapy with oximes therefore restores normal enzyme function. Oxime therapy is limited by a second side reaction, called "aging," in which a side chain on nerve agents falls off the complex at a characteristic rate. "Aged" complexes are negatively charged, and oximes cannot reactivate negatively charged complexes. The practical effect of this differs from one nerve agent to another since each ages at a characteristic rate. VX, for practical purposes, never ages, sarin ages in 3–5 hours, and tabun ages over a longer period. All these are so much longer than the patient's expected life span after acute nerve agent toxicity that they may be ignored. Soman, in contrast, ages in 2 minutes. Thus, after only a few minutes after exposure, oximes are useless in treating soman poisoning. The oxime used varies by country; the United States has approved and fielded 2-pralidoxime chloride (2-PAM Cl). MARK 1 kits (Fig. 222-5A) and Duodotes(TM) both contain autoinjectors 600 mg 2-PAM Cl. Initial field loading doses are 600, 1200, and 1800 mg. Since blood pressure elevation may occur after administration of 45 mg/kg in adults, field use of 2-PAM Cl is restricted to 1800 mg, IM, per hour. During the time when more oxime cannot be given, atropine alone is recommended. In the hospital setting, 2.5–25 mg/kg IV 2-PAM Cl has been found to reactivate 50% of inhibited cholinesterase. The usual recommendation is 1000 mg through slow IV drip over 20–30 minutes, with ≥2500 mg over a period of 1–1.5 hours. Active research is ongoing to field a more effective and broader-spectrum oxime than 2-PAM Cl.

Anticonvulsants

Nerve agent—induced seizures do not respond to the usual anticonvulsants used for status epilepticus, including phenytoin, phenobarbital, carbamazepine, valproic acid, and lamotrigine (Chap. 369). The only anticonvulsants that have been shown to stop this form of status are the benzodiazepines. Diazepam is the only benzodiazepine approved for seizures in humans, although other FDA-approved benzodiazepines work well against nerve agent—induced seizures in animal models. Diazepam therefore is manufactured in 10-mg injectors for IM use and given to U.S. forces for this purpose **(Fig. 222-5*B*)**. Civilian agencies are stockpiling this field product (convulsive antidote for nerve agent, CANA), which generally is not used in hospital practice. Extrapolation from animal studies indicates that adults will probably require 30–40 mg diazepam, IM, to stop nerve agent—induced status epilepticus. In the hospital or in a small child unable to receive the autoinjector, IV diazepam may be used at similar doses. The clinician may confuse seizures with the neuromuscular signs of nerve agent poisoning. In the hospital, early electroencephalography is advised to distinguish among nonconvulsive status epilepticus, actual seizures, and postictal paralysis. Recent animal studies have shown that the most effective benzodiazepine in this situation is midazolam, which is not FDA-approved for seizures.

Peripheral neuropathy and the so-called intermediate syndrome, which are prominent long-term effects of insecticide poisoning, are not described in nerve agent survivors.

Recent research has explored approaches leading to a transient "immunity," or drugs that would provide protection against lethal nerve agents yet be devoid of side effects. A novel approach is to use enzymes to scavenge these highly toxic nerve agents before they attack their intended targets. The accumulated work has shown that if a scavenger is present at the time of nerve agent exposure, rapid reduction of toxicant levels is observed. This reduction is so rapid and profound that the need to administer a host of pharmacologically active drugs as antidotes is, according to laboratory studies, eliminated.

Cyanide

Cyanide (CN^-) has become an agent of particular interest in terrorist scenarios because of its applicability to indoor targets. Attacks have occurred in recent years using this agent, for example, against the water supply of the U.S. Embassy in Italy. The 1993 World Trade Center bombing in New York may have been intended as a cyanide release as well.

Hydrogen cyanide and cyanogen chloride, the major forms of cyanide, are either true gases or liquids very close to their boiling points at standard room temperature. Hydrogen cyanide gas is lighter than air and does not remain concentrated outdoors for long; this makes it a poor military weapon but an effective weapon in an indoor space such as a train station or sports arena. Cyanide is also water-soluble and poses a threat to the food and water supply from either accidents or malign intent. It is well absorbed from the GI tract, through skin, or via inhalation, the preferred terrorist route. Cyanide smells like bitter almonds, but 50% of persons lack the ability to smell it.

Unique among chemical warfare agents, cyanide is a normal constituent of the environment and actually is a required cofactor in many compounds important in metabolism, including vitamin B12. Cyanide is present in many plants; this explains why smokers, for instance, have three times the usual cyanide level chronically. In consequence, humans have evolved a detoxification mechanism for cyanide. Cyanide poisoning results if a large challenge of CN^- overwhelms this mechanism, while treatment of cyanide poisoning exploits it.

MECHANISM

Cyanide directly poisons the last step in the mitochondrial electron transport chain, cytochrome a3, which results in a shutdown of cellular energy production. Tissues are poisoned in direct proportion to their metabolic rate, with the carotid baroreceptors and the brain, the most metabolically active tissues in the body, affected fastest and most severely. This results from cyanide's high affinity for certain metals, notably Co and Fe+++. Cytochrome a3 contains Fe+++, to which CN^- binds. Cyanide-poisoned tissues cannot extract oxygen from the blood; even though pulmonary oxygen exchange and cardiac function are preserved, cells die of hypoxia—of histotoxic rather than cardiopulmonary cause.

CLINICAL FEATURES

Approximately 15 seconds after inhalation of a high concentration of cyanide, there is a transient hyperpnea, followed within 15–30 s by the onset of convulsions and electrical status epilepticus. Respiratory activity stops 2–3 min later, and cardiac activity ceases several minutes after that.

Exposure, especially via inhalation, to a large challenge of CN^- can cause death in as little as 8 min. Smaller challenges will cause symptoms spread over a longer period; very low doses may produce no effects at all because of the body's ability to detoxify small amounts. Cyanogen chloride additionally produces mucous membrane irritation. Many but not all patients will display a cherry red appearance because their venous blood remains oxygenated.

DIFFERENTIAL DIAGNOSIS

In a mass casualty caused by a chemical agent, the primary differential diagnosis of cyanide poisoning will be nerve agent poisoning. Cyanide-poisoned patients lack the prominent cholinergic signs such as miosis and increased secretions seen in nerve agent poisoning. The cherry red appearance often seen in cyanide poisoning is never seen in nerve agent poisoning. Cyanosis, confusingly, is not a prominent early sign in cyanide poisoning.

TREATMENT: CYANIDE

Treatment of cyanide poisoning may require simply evacuating the patient from the source of contamination. Decontamination of a true gas, other than clothing removal to avoid gas trapped in clothing air cells, is probably not a major concern.

Oxygen, via mask, nasal cannula, or endotracheal tube, has been shown to benefit patients, although the benefit is not explained by the known mechanism of action of CN^-.

Antidotal Therapy

Cyanide antidotes exploit the body's innate detoxification mechanism, the hepatic enzyme rhodanese. They also exploit cyanide's affinity for certain metal ions. The antidotes are summarized in Table 222-3.

The classic two-step cyanide antidote kit includes two IV solutions: sodium nitrite and sodium thiosulfate. It may also include amyl nitrite perles for inhalation.

Nitrites are methemoglobin formers. Giving a patient nitrite converts a fraction of the body's hemoglobin into methemoglobin by converting heme iron from Fe++ to Fe+++. CN^- has a greater affinity for methemoglobin Fe+++ than for cytochrome a3. As a result, administration of nitrite creates a "sink" of cyanmethemoglobin; creation of methemoglobin pulls CN^- off mitochondrial cytochrome a3, allowing cellular respiration to resume. Recent work suggests that nitrites may also work via a second mechanism, involving the neurotransmitter nitrous oxide, which may explain why cyanide-poisoned patients improve after nitrite administration faster than is explained by the known rate of methemoglobin formation.

Nitrite administration may save the patient acutely but creates an unstable pool of cyanmethemoglobin, the elimination of which requires a sulfur donor: sodium thiosulfate. Sodium thiosulfate donates sulfur to the reaction catalyzed by rhodanese; this converts cyanide to thiocyanate, a compound the body eliminates harmlessly in urine. In fire victims'whose oxygen-carrying capacity is already reduced, and in whom the administration of nitrite may form so much methemoglobin as to make the blood unable to carry oxygen at all—sodium thiosulfate alone may be administered.

Hydroxocobalamin, or vitamin B12a, has recently been approved for use as an alternate cyanide antidote. It must be reconstituted at the scene, unlike sodium nitrite or sodium thiosulfate. It lacks the propensity for hypotension of nitrites, but many of the cases in which it has been beneficial have also required the use of sodium thiosulfate. It causes an orange discoloration of the skin of no functional significance.

All of these antidotes require the placement of an IV line. Amyl nitrite is currently the only nonintravenous cyanide antidote available, although it has never been formally approved by the U.S. Food and Drug Administration. Cyanide antidote kits contain amyl nitrite perles that can be crushed and inhaled by a patient who is still breathing. Amyl nitrite can also be given through a respirator.

None of the cyanide antidotes is specifically approved for pediatric use.

PROGNOSIS

Cyanide casualties tend to recover very quickly compared with casualties of other chemical agents. Many industrial cases have returned to work within the same shift. If a patient suffers a large challenge and expires, it is usually within minutes of exposure.

FURTHER READINGS

Hurst CG, Smith WJ: Health effects of exposure to vesicant agents, in *Chemical Warfare Agents: Chemistry, Pharmacology, Toxicology and Therapeutics,* JA Romano et al (eds). Boca Raton, FL, CRC Press, 2008

Karalliedde L et al: Possible immediate and long-term health effects following exposure to chemical warfare agents. *Public Health* 114:238–248, 2000[PMID: 10962584]

Lenz DE et al: Nerve agent bioscavengers: Progress in the development of a new mode of protection against organophosphorus poisoning, in *Chemical Warfare Agents: Chemistry, Pharmacology, Toxicology and Therapeutics,* JA Romano et al (eds). Boca Raton, FL, CRC Press, 2008

McDonough JH et al: Anticonvulsant treatment of nerve agent seizures: Anticholinergics versus diazepam in soman-intoxicated guinea pigs. Epilepsy Res 38:1–14, 2000[PMID: 10604601]

Newmark J: Nerve agents, in *Clinical Neurotoxicology: Syndromes, Substances, Environments*, MR Dobbs (ed). Philadelphia, Saunders-Elsevier, 2009, pp 646–659

Okumura T et al: Report on 640 victims of the Tokyo subway sarin attack. Ann Emerg Med 28:129–135, 1996[PMID: 8759575]

Russell D, simpson J: Emergency planning and preparedness for the deliberate release of toxic industrial chemicals. Clin Toxicol 48:171–176,2010[PMID: 20334545]

Sidell FR: Clinical considerations in nerve agent intoxication, in *Chemical Warfare Agents*, SM Somani (ed). San Diego, Academic Press, 1992, pp 156–194

Tuorinsky SD (ed.): *Medical Aspects of Chemical Warfare* . Walter Reed Army Medical Center, Washington, DC, Borden Institute, 2008. [Available on the website: http://ccc.apgea.army.mil]

U.S. Army Medical Research Institute of Chemical Defense, Chemical Casualty Care Division: *Medical Management of Chemical Casualties Handbook*, 4th ed. Aberdeen Proving Ground, MD, 2007 [Available on the website: http://ccc.apgea.army.mil]

Willems JL: Clinical management of mustard gas casualties. *Ann Med Milit Belg* 3(Suppl.):1–61, 1989

Wurzburg H. Treatment of cyanide poisoning in an industrial setting. Vet Human Toxicol 38:44–47, 1996[PMID: 8825751]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

RADIATION TERRORISM: INTRODUCTION

Terror attacks using nuclear or radiation-related devices are an unequivocal threat in the twenty-first century and are capable of having unique medical and psychological effects. This chapter will focus on the most probable scenarios of possible attacks and the medical principles of handling such threats.

There are two major categories of potential terrorist incidents with widespread radiologic consequences. The first is the use of radiologic dispersal devices. This could cause a purposeful dissemination of radioactive material without nuclear detonation by using conventional explosives with radionuclides, attacking fixed nuclear facilities, or attacking nuclear-powered surface vessels or submarines. Malfunctioning nuclear weapons that are detonated with no nuclear yield (nuclear "duds") and installation of radionuclides in food or water are also possible means of generating a terror attack. The second and less probable scenario is the actual use of nuclear weapons. Each scenario has its own medical aspects, including "conventional" blast or thermal injury, introduction to a radiation field, and exposure to either external or internal contamination from a radioactive explosion.

Types of Radioisotopic Radiation

Isotopes of atoms with uneven numbers of protons and/or neutrons are typically unstable; such isotopes discharge particles or energy to matter, a process that is defined as *radiation*. The main radiation types are alpha, beta, gamma, and neutrons.

Alpha (α) radiation consists of heavy, positively charged particles that contain two protons and two neutrons. Alpha particles usually are emitted from isotopes with an atomic number of ≥82, such as uranium and plutonium. Due to their large size, alpha particles have limited penetrating power. Fine obstacles such as cloth and human skin usually can stop them from penetrating into the body, and they represent a small risk with external exposure due to their limited penetration. If they somehow are internalized, alpha particles can cause significant cellular damage in their immediate proximity.

Beta (β) radiation consists of electrons, which are small, light, negatively charged particles (about 1/2000 the mass of a neutron or proton). They can travel only a short finite distance in tissue, depending on their energy. Exposure to beta particles is common in many radiation accidents. Radioactive iodine released in nuclear plant accidents is the best known member of this group. Plastic layers and clothing can stop most beta particles, and their penetration is measured to be a few millimeters. A large quantum of energy to the basal stratum of the skin can cause a burn that is similar to a thermal burn and is treated as such.

Gamma (γ) rays and *x-rays* (both photons) are similar. Gamma rays are uncharged electromagnetic radiation discharged from a nucleus as a wave or photons of energy. X-rays are the product of abrupt mechanical deceleration of electrons striking a heavy target such as tungsten. Gamma rays and x-rays have similar properties, (i.e., no charge and no mass, just energy). Both travel easily through matter, sometimes called *penetrating radiation*, and are the principal type of radiation that causes total-body exposure. If the energy of gamma rays and x-rays is the same, their biologic effects will be the same.

Neutron (π) particles are heavy and uncharged, often emitted during nuclear detonation. They possess a wide energy range; their ability to penetrate tissues is variable, depending on their energy. They are less likely to be present in most scenarios of radiation bioterrorism.

The ionization resulting from protons, electrons, and gamma rays is either a direct or an indirect (i.e., mediated through water) effect of particles or photons on DNA. Ionization of DNA resulting from neutrons is secondary to the neutrons knocking electrons out of their atomic orbit and the formation of free radicals, which can also damage DNA directly.

The commonly used units of radiation are the rad and the gray (Gy). The rad (radiation absorbed dose) is energy deposited within living matter and is equal to 100 ergs/g of tissue.

The traditional rad has been replaced by the Syst`me Internationale (SI) unit of the gray; 100 rad = 1 Gy.

Types of Exposure

Whole-body exposure represents deposition of radiation energy over the entire body. Alpha and beta particles have limited penetration and do not cause significant noncutaneous injury unless emission results from an internalized source. Whole-body exposure from gamma rays, x-rays, or neutrons, which can penetrate through the body (depending on their energy), can result in damage to multiple tissues and organs. The tissue damage is proportional to the radiation exposure of the specific organ or tissue.

External contamination is a result of fallout of radioactive particles that land on the body surface, clothing, skin, and hair. This is the dominant element to consider in the mass casualty situation resulting from a radioactive terrorist strike. The common contaminants primarily emit alpha and beta radiation. Alpha particles do not penetrate beyond the skin and thus have minimal systemic effects. Beta emitters can cause significant cutaneous burns and scarring. Gamma emitters not only may cause local damage but also can cause whole-body radiation exposures and injury. The medical treatment is primarily decontamination of the body, including

wounds and burns, to prevent the contamination from becoming internalized. Removing the contaminated clothing reduces the contamination significantly and is a first step in the decontamination process. Generally, patients will not constitute a significant radiation hazard to health care providers, and lifesaving treatment should not be delayed for fear of secondary contamination of the medical team. Any damage to health care personnel will depend directly on the duration of exposure and will be inversely proportional to the square of the distance from any radioactive source. Gowns that can be easily removed are essential to protect health care personnel.

Internal contamination will occur when radioactive material is inhaled or ingested or is able to enter the body through open wounds or burns or via skin absorption. In principle, any externally contaminated casualty should be evaluated for internal contamination. Some isotopes may have toxic effects on specific target organs due to their chemical properties, in addition to radiologic injury. The respiratory system is the main portal of entrance for internal contamination, and the lung is the organ at greatest risk. Aerosol particles <5 μm can reach the alveoli, whereas larger particles will remain in proximal airways. The tiny particles can be absorbed by the lymphatic system or the bloodstream. Bronchial lavage is often a helpful treatment in this situation. Radioactive material entering the gastrointestinal (GI) tract will be absorbed according to its chemical structure and solubility. The insoluble radionuclides may affect the lower GI tract. Intact skin is normally a good barrier to most radionuclides. Penetration through the skin usually takes place when wounds or burns have altered the skin barriers. Therefore, any skin erosion should be cleaned and decontaminated promptly.

Absorbed radioactive materials will travel throughout the body. Liver, kidney, adipose tissue, thyroid, and bone and bone marrow tend to bind and retain the radioactive material more than other tissues do. The medical treatment includes preventing absorption, reducing incorporation, and enhancing elimination (see below).

Localized exposure means close contact between a highly radioactive source and a part of the body, causing discrete damage to the skin and deeper tissues, similar to a thermal burn. Later signs include epilation, erythema, moist desquamation, ulceration, blistering, and necrosis in proportion to exposure. Alopecia, transient or permanent, is dose related and starts at cutaneous doses >3 Gy. Overt tissue damage can take weeks and even months to develop; the healing process can also be very slow, lasting for months. Long-term cutaneous changes, including keratosis, fibrosis, and telangiectasias, may appear years after the exposure. Treatment is based on analgesia and infection prophylaxis. Nevertheless, severe burns often require grafting or even amputation. Long-term radiation effects are characterized by cell loss and cell death.

Radiologic Dispersal Events

Radiologic dispersal incidents are generally of two types, resulting from (1) small, usually localized sources or (2) wide dispersals over large areas. The radioactive materials can take the form of solid state, aerosol, gas, or liquid. They can be put into food or water, released from vehicles, or be spread by explosion. The principal route of exposure is usually direct contact between the victim's skin and the radioactive particles, although internal contamination could occur if the material were inhaled or ingested. The radiation field is also a potential source of whole-body exposure. The psychosocial effects that accompany such an event are significant and are beyond the scope of this chapter. A list of radioactive materials, including information on their major properties and medical treatment, is given in **Table 223-1**.

Table 223-1 Internal Contaminant Radionuclides: Properties and Treatment

Isotope Name	Symbol	Common Usage	Radiation Type $t_{1/2}$ Radiologic $t_{1/2}$ Biologic, days	Exposure Type	Mode of Contamination	Focal Accumulation in Body	Treatment
Manganese	Mn-56	Reactors, research laboratories	β, γ 2.6 h 5.7	External, internal	N/A	Liver	N/A
Cobalt	Co-60	Medical radiotherapy devices, commercial food irradiators	β, γ 5.26 y 9.5	External, internal	Lungs	Liver	Gastric lavage, purgatives; penicillamine in severe cases
Strontium	Sr-90	Fission product of uranium	β 28 y 18,000	Internal	Moderate GI tract	Bones—similar to calcium	Strontium, calcium, ammonium chloride
Molybdenum	Mo-99	Hospitals—scans	β, γ 66.7 h 3	External, internal	N/A	Kidneys	N/A
Technetium	Tc-99m	Hospitals—scans	β, γ 6.049 h 1	External, internal	IV administration	Kidneys, total body	Potassium perchlorate to reduce thyroid dose
Cesium	Cs-137	Medical radiotherapy devices	β, γ 30 y 70	External, internal	Lungs, GI tract, wounds, follows potassium	Renal excretion	Ion-exchange resins, Prussian blue
Gadolinium	Gd-153	Hospitals	β, γ 242 d 1000	External, internal	N/A	N/A	N/A

Iridium	Ir-192	Commercial radiography	β, γ 74 d 50	External, internal	N/A	Spleen	N/A
Radium	Ra-226	Instrument illumination, industrial applications, old medical equipment, former Soviet Union military equipment	α, β, γ 1602 *y* 16,400	External, internal	GI tract	Bones	$MgSO_4$ lavage, ammonium chloride, calcium alginates
Tritium	H-3	Luminescent gun sights, muzzle-velocity detectors, nuclear weapons	β 12.5 y 12	Internal	Inhalation, GI tract, wounds	Total body	Dilution with controlled water intake, diuretics
Iodine-131	131_I	Reactor accidents, thyroid ablators	β, γ 8.1 d 138	Internal	Inhalation, GI tract, wounds	Thyroid	Potassium/sodium iodide, propylthiouracil, methimazole
Uranium	U-235	Depleted uranium, natural uranium, fuel rods, weapons-grade material	α, (α, β, γ) 7.1×10^8 *y* 15	Internal	GI tract	Kidneys, bones	$NaHCO_3$, chelation with EDTA
Plutonium	Pu-239	Produced from uranium in reactors, nuclear weapons	α 2.2×10^4 *y* 73,000	Internal	Limited lung absorption, high retention	Lungs, bones, bone marrow, liver, gonads	Chelating with DTPA or EDTA
Americium	Am-241	Smoke detectors, nuclear weapon detonation fallout	α 458 *y* 73,000	Internal	Inhalation, skin wounds	Lungs, liver, bones, bone marrow	Chelating with DTPA or EDTA
Polonium	Po-210	Calibration source	α 138.4 *d* 60	Internal	Inhalation, wounds	Spleen, kidneys	Lavage, dimercaprol
Thorium	Th-232	Calibration source	α 1.41×10^{10} *y* 73,000	Internal	N/A	N/A	N/A
Phosphorus	P-32	Research laboratories, medical facilities	β 14.3 d 1155	Internal	Inhalation, GI tract, wounds	Bones, bone marrow, rapidly replicating cells	Lavage, aluminum hydroxide, phosphate

Abbreviations: DTPA, diethylenetriamine pentaacetic acid; EDTA, ethylenediamine tetraacetic acid; GI, gastrointestinal; h, hours; N/A, not available; y, years.

In a localized event, the amount and spread of the radioactive materials are usually limited and can be treated like a spill of hazardous material. Protective clothing prevents or minimizes the contamination of emergency responders.

The use of explosives coupled with a large amount of radioactive materials can result in wide dispersion of radiation, which is of far greater concern. Other potential sources of radiation are nuclear reactors, spent nuclear fuel, and transport vehicles. Less probable but still possible is the use of a large source of penetrating radiation without explosion. It is expected that most exposures would be low, and the principal health and psychosocial effects would be similar to those in the former scenario but on a larger scale.

Whenever an explosion is involved, conventional lifesaving treatment should be given first priority. Only then should decontamination and specific treatment be given for the radiation exposure.

Silent exposure represents a scenario in which a powerful radiologic source, often also called a radiologic exposure device, could be hidden in a crowded place and spread radioactive materials without any awareness or announcement. It might take a long time to recognize the event and the source of exposure. One of the major clues to this situation is the appearance of unusual clinical manifestations in many individuals; such manifestations are often nonspecific and include symptoms of acute radiation sickness (see below) such as

headache, fatigue, malaise, and opportunistic infections. GI phenomena such as diarrhea, nausea, vomiting, and anorexia may occur. Dermatologic symptoms (burns, ulceration, and epilation) and hematopoietic manifestations such as bleeding tendency, thrombocytopenia, purpura, lymphopenia, and neutropenia are also possible and are dose related. Careful epidemiologic studies may be necessary to identify the source of such exposure.

Nuclear Weapons

The most likely scenario of nuclear terror would be the detonation of a single low-yield device. The estimated yield of such a device is anywhere between 0.01 and 10 kilotons of TNT, although the probability more likely would be toward the lower yield. Coping with such an event is certainly possible. The effects of such an explosion are a combination of several components: ground shock, air blast, thermal radiation, initial nuclear radiation, crater formation, and radioactive fallout.

The nuclear detonation, like a conventional explosion, will produce a shock wave that can further damage structures and cause many casualties. In addition, the detonation can produce an extremely hot fireball that can ignite materials and cause severe burns. The detonation also releases an intense pulse of ionizing radiation, mainly gamma rays and neutrons. The radiation produced in the first minute is termed *initial radiation*, whereas the ongoing radiation due to fallout is termed *residual radiation*. Both types of radiation can cause acute radiation sickness. The $LD_{50/30}$ (i.e., a dose that causes a 50% mortality rate at 30 days) is ~4 Gy for whole-body exposure without medical support; with medical support, the $LD_{50/30}$ ranges between 8 and 10 Gy. Winds can carry fallout and contaminate large areas.

On top of its effects, a massive blast forms a crater in the soil and usually produces a ground shock that compounds the damage and the number of casualties. Inhalation of large amounts of radioactive dust causes pneumonitis that can lead to pulmonary fibrosis. Use of a mask covering the mouth and nose can be very helpful. The intense flash of infrared and visible light can cause either temporary or permanent blindness. Cataracts can develop months to years later among those who survive.

Acute Radiation Syndrome

Radiation interactions with atoms can result in ionization and the formation of free radicals that damage tissue by disrupting chemical bonds and molecular structures in the cell, including DNA. Radiation damage can lead to cell death; the cells that recover may be mutated and at higher risk for subsequent cancer. Cell sensitivity increases as the replication rate increases and cell differentiation decreases. Bone marrow and mucosal surfaces of the GI tract, which have vast mitotic activity, are significantly more sensitive to radiation than are slowly dividing tissues such as bones and muscles. After exposure of either all or most of the human body to ionizing radiation, acute radiation syndrome (ARS) can develop. The clinical manifestations of ARS reflect the dose and type of radiation as well as the part of the body exposed.

ARS manifests as three major groups of signs and symptoms: hematopoietic, GI, and neurovascular. There are four major stages in ARS: prodrome, latent phase, illness, and recovery or death. The higher the radiation doses, the shorter and more severe each stage. The prodrome appears within minutes to 4 days after exposure, lasts from a few hours to a few days, and can include nausea, vomiting, anorexia, and diarrhea. At the end of the prodrome, ARS progresses to the latent phase. Minimal or no symptoms are present during the latent phase, which commonly lasts up to 2.5 weeks but can last up to 6 weeks. The duration depends on the radiation dose, the health of the patient, and the coexisting illness or injury. After the latent phase, the exposed person manifests illness that may eventuate in recovery or lead to death.

With exposure to doses <1 Gy, ARS is generally mild. At this dose symptoms can be minimal or nonexistent even if the entire body is exposed to penetrating radiation. The clinical picture will mainly be transient depression of bone marrow (lymphopenia) that lasts up to 2 to 3 weeks and then improves.

ARS is significantly more acute and severe with exposure to very high doses: >30 Gy. At this dose the prodrome appears in minutes and is followed by 5 to 6 hours of latency before a cardiovascular collapse occurs secondary to irreversible damage to the microcirculation.

The type and dose of radiation and the part of the body exposed will determine not only the timing of the different stages of ARS but also the dominant clinical picture. At low radiation doses of 0.7–4 Gy, hematopoietic depression due to bone marrow suppression takes place and constitutes the main illness. The patient may develop infections and bleeding secondary to low leukocyte and platelet counts, respectively. The bone marrow eventually will recover in almost all patients if they are supported with transfusions and fluids; antibiotics are often needed in addition. With exposure to 6–8 Gy, the clinical picture is significantly more complicated. At these doses, the bone marrow will not always recover and death may ensue. A GI syndrome may accompany the hematopoietic manifestations and further worsen the patient's condition. Compromise of the absorptive layer of the gut alters absorption of fluids, electrolytes, and nutrients. GI injury can lead to vomiting, diarrhea, GI bleeding, sepsis, and electrolyte and fluid imbalance in a patient whose blood counts are compromised for a period of weeks, often leading to death. Whole-body exposure to doses >9–10 Gy is almost always fatal. Crucial elements of the bone marrow simply will not recover. In addition to the GI syndrome associated with very large exposures, patients may develop a neurovascular syndrome; the latter dominates with whole-body doses >20 Gy. Vascular collapse, seizures, confusion, and death usually occur within days. In this variant the prodrome and latent phase both shorten to a few hours.

TREATMENT: ACUTE RADIATION SICKNESS

The treatment of ARS is focused on maintaining homeostasis, giving damaged organs a chance to recover. Aggressive support is given to every damaged system. Treatment for the hematopoietic system includes mainly therapy for neutropenia and infection, transfusion and blood products such as leukoreduced irradiated blood as needed, and hematopoietic growth factors. The value of bone marrow transplantation in this situation is questionable. None of the transplants that were performed among the victims of the nuclear reactor accident in Chernobyl proved successful. Bone marrow transplantation could be considered for casualties with whole-body

exposure to 6–10 Gy when the hematopoietic syndrome is dominant and the bone marrow is less likely to recover with time. Another major component of the treatment of ARS is partial or total parenteral nutrition to bypass the damaged GI system. For blast and thermal injuries, standard therapy for trauma is given. Psychological support is essential in many cases.

Medical Management of Radiation Bioterrorism

Victims of radiation bioterrorism can suffer from conventional thermal or blast injuries, exposure to radiation, and contamination by radioactive materials. Many will have combinations of the above, which can be synergistic and cause higher morbidity and mortality rates than is the case when they occur alone. The number of casualties will be a major factor in determining the response of the medical system to an act of radiation bioterrorism. If only a few persons are affected, no significant changes and adaptation of the system are needed to treat the victims. However, if a terror attack results in a large number (dozens or more) of casualties, an organized disaster plan at the local and state levels must be invoked to deal with the crisis properly. Useful U.S. planning documents that include many universal planning concepts can be found at *http://www.remm.nlm.gov/remm_Preplanning.htm*. Medical personnel should have a prior assignment and training and be prepared to function in a scenario with which they are familiar. Stockpiles of specific equipment and medications have to be preplanned (see the Centers for Disease Control and Prevention Web site at *http://www.bt.cdc.gov/stockpile/*). One of the goals of terrorists is to overwhelm medical facilities and minimize the salvage of casualties.

Initial management consists of *primary triage and transportation* of the wounded to medical facilities for treatment. The rationale behind the triage is to sort patients into classes according to the severity of injury for the purpose of expediting clinical care and maximizing the use of the available clinical services and facilities. Triage requires determination of the level of emergency care needed. The higher the number and range of casualties are, the more complex and difficult triage becomes. The mildly wounded and victims of contamination only can be sent to evacuation, registration with disaster response teams, and decontamination and treatment centers. **Figure 223-1** illustrates evacuation in a multicasualties radiologic event. In this way, the hospitals can avoid being directly overwhelmed, and those who are severely wounded can receive better treatment. Emergency treatment will be administered initially according to the presence of conventional injuries such as wounds, trauma, and thermal or chemical burns. Individuals with such injuries should be stabilized, if possible, and immediately transported to a medical facility. Removing the victim's clothes and wrapping him or her in clean blankets or nylon sheets reduces both the exposure of the patient and the contamination risk to the staff. However, the possibility of contamination needs to be determined. Less severely injured victims should receive a preliminary decontamination before or during evacuation to a hospital.

Figure 223-1

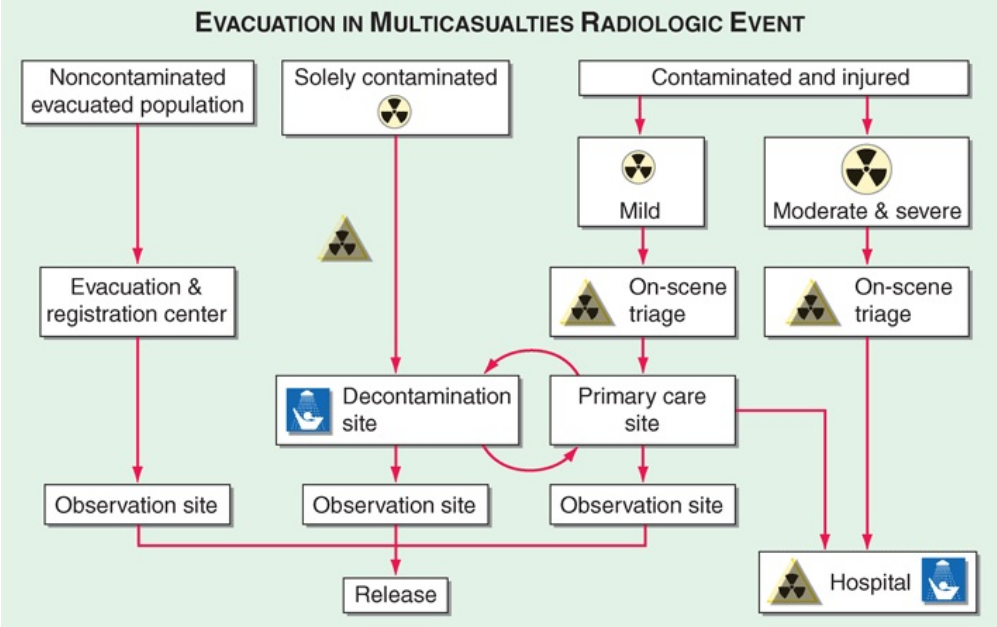

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for evacuation in a multicasualties radiologic event.

One must remember that radionuclide contamination of the skin is commonly not an acute life-threatening situation for the patient or the personnel who care for the patient. Only powerful gamma emitters are likely to cause real damage from contamination. It is important to emphasize that exposure to a radiation field alone does not necessarily create any contamination. The exposed person, if not contaminated, is not radioactive and does not directly emit any radiation.

To protect the staff, protective gear (gowns, gloves, masks, and caps) should be used. Protective masks with filters and chemically protective overgarments provide excellent protection from contamination. Waterproof shoe covers are also important. Remaining in the contaminated area and dealing with lifesaving procedures should

take place according to the "ALARA" principle: as low as reasonably achievable. It is better to send many people for short exposure times than to send a few people for longer periods of time to do the same job.

Decontamination of victims should take place in the field before their arrival at medical facilities, but radiologic decontamination should never interfere with medical care. Removal of outer clothing and shoes usually will reduce a patient's contamination by 80–90%. Contaminated clothes should be carefully removed by rolling them over themselves, placing them in marked plastic bags, and removing them to a predefined area for contaminated clothes and equipment. A radiation detector should then be used to check for the presence of any residual radiologic contamination on the patient's body. To prevent internalization of the radioactive materials, one should cover open wounds before decontamination. Showering or washing of the entire skin and hair is very important. The skin is dried and reassessed for residual contamination until no radiation is found. Contamination-removing chemical agents are more than sufficient to remove radiologic contamination.

Wound decontamination should be as conservative as possible. The main goal is to prevent both extensive local damage and internal contamination through lacerated skin. The bandages should be removed, and the wounds flushed. The wound should then be dried and assessed for radiation. This procedure can be repeated again and again until contamination is undetectable. Excision of contaminated wounds should be attempted only when surgically necessary. Radioactive shrapnel that can penetrate through the skin should be removed.

In the hospital, staff can wear normal hospital barrier clothing, including two pairs of gloves, a gown, shoe covers, a head cover, and a face mask. Eye protection is recommended. Decontamination of medical personnel is obligatory after emergency treatment and decontamination of the patient. All protective clothing should be placed after use in a designated container for contaminated clothing.

Radiation intensity decays rapidly with the square of the distance from the source, and increasing the distance from the source and decreasing the time spent near it are basic principles of radiation safety. Shielding with lead can be used as protection from small radioactive gamma sources. Geiger counters can detect gamma and beta radiation. Pocket chamber (pencil) dosimeters, film badges, and thermoluminescent dosimeters can measure accumulated exposure to gamma radiation. All these detectors are in common use in medical facilities and should be used to help define the level of contamination. Alpha radiation is harder to detect due to its poor penetration. An alpha scintillation counter, which is capable of detecting alpha radiation, is not commonly used in medical facilities.

Guidelines for Hospital Management

Figure 223-2 shows a model for hospital arrangement for triage. Persons contaminated either externally or internally should be identified, externally decontaminated, and, if needed, treated immediately and specifically for internal contamination. In all other cases, the need for treatment of radiation injuries does not constitute a medical emergency. Early actions, such as blood sampling both for assessing the degree of severity of the exposure and for blood type and cross-matching for possible transfusion, need to be taken promptly if ARS is evident or if exposure is suspected.

Figure 223-2

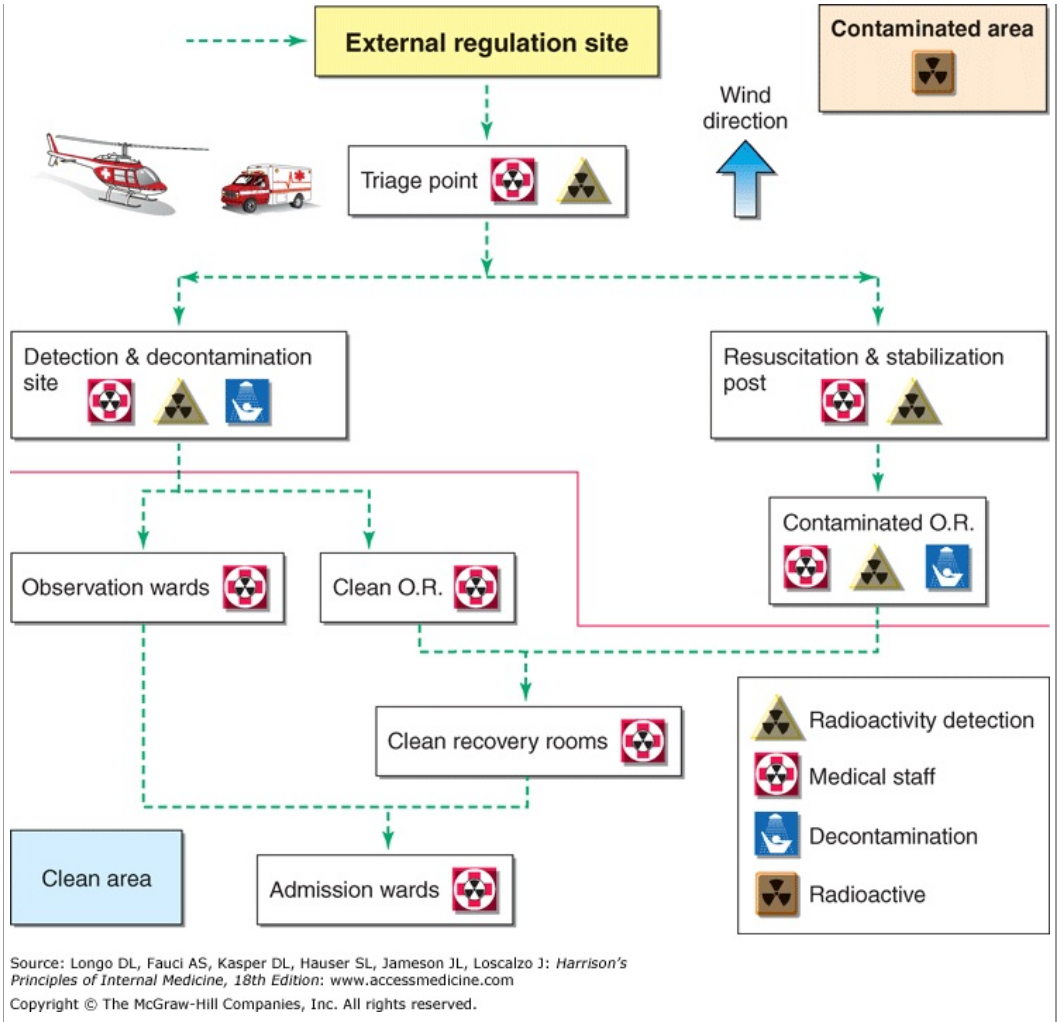

Flow chart of hospital triage. O.R., operating room.

In the hospital entrance, a distinct decontamination area should be set up promptly. Separation between clean and contaminated areas is essential. Medical personnel in this area should wear protective gear as noted above. They also should be rotated in their assignments every 1 to 2 hours to ensure minimal exposure to radiation. If patients are critically wounded and require either surgery or resuscitation, they need to pass directly to "contaminated" operating rooms or resuscitation sites for lifesaving procedures. Once such patients are stable, they should be decontaminated. It is important to obtain details concerning the exposure, look for prodromal signs of radiation sickness, and do a physical examination. One of the best ways to estimate exposure clinically is to measure the time of prodromal appearance. The earlier the prodromal signs and symptoms appear, the higher the dose is of radiation exposure. A few laboratory tests need to be taken routinely, such as complete blood count and urinalysis. If internal contamination is suspected, specific treatment should be given, as outlined below.

TREATMENT: RADIONUCLIDE CONTAMINATION

Treatment for internal radionuclide contamination, decorporation, should be started as soon as possible after suspected or known exposure. The approximate upper limit of radionuclide contamination that can reasonably be ignored from a radiation safety point of view is not well defined. These are judgments that will depend on the circumstances of the event and the resources available. One method to determine a level of internal contamination that will trigger decorporation would be the upper limit of the yearly radionuclide contamination permitted for radiation workers [allowable levels of intake (ALIs)]. Radiation workers are permitted 50 times more radiation dose each year than are members of the general public. A new concept called the Clinical Decision Guide is contained in the National Council on Radiation Protection and Measurements (NCRP) Report 161 and is expected to replace the use of ALI.

The goal is to leave the smallest amount of radionuclides possible in the body. Treatment is given to reduce absorption and enhance elimination and excretion. Some of the decorporation agents are not U.S. Food and Drug Administration (FDA)-approved for these indications, and there is very little clinical data to support the efficacy of their use.

Clearance of the GI tract may be achieved by stomach lavage, with emetics (such as apomorphine, 5 to 10 mg, or ipecac, 1- to 2-g capsules or 15 mL in syrup), or by using purgatives, laxatives, ion exchangers, and aluminum antacids. Prussian blue, 1 g tid for a minimum of 3 weeks, is an ion exchanger used to treat cesium 137 internal contamination. Aluminum antacids (such as aluminum phosphate gel) may reduce strontium uptake in the gut if given immediately after exposure. Aluminum hydroxide is less effective.

Prevention or reversal of radionuclide interaction with tissues can be done by blocking, diluting, mobilizing, and chelating agents. *Blocking agents* prevent the entrance of radioactive materials. A good example is potassium

iodide (KI), which blocks the uptake of radioactive iodine (^{131}I) by the thyroid. KI is most effective if taken within the first hour after exposure and is still effective 6 hours after exposure. The effectiveness subsequently declines until 24 hours after exposure; however, it is recommended that KI be taken up to 48 hours after exposure. The KI dose is based on age, predicted thyroid exposure, and pregnancy and lactation status. Adults between the ages of 18–40 should receive 130 mg/d for 7–14 days if exposed to ≥10 cGy of radioactive iodine. Other thyroid-blocking agents include prophylthiouracil, 100 mg tid for 8 days, and methimazole, 10 mg tid for 2 days followed by 5 mg tid for 6 days, but they are somewhat less effective.

Diluting agents decrease the absorption of the radionuclide; for example, water may be used as a diluting agent in the treatment for tritium (^{3}H) contamination. The recommended amount is 3–4 L/d for at least 3 weeks.

Mobilizing agents are most effective when given immediately; however, they may be effective for up to 2 weeks after exposure. These agents include antithyroid drugs, parathyroid extract, glucocorticoids, ammonium chloride, diuretics, expectorants, and inhalants. All of them should induce the release of radionuclides from tissues.

Chelating agents can bind many radioactive materials, after which the complexes are excreted from the human body. In this regard, diethylenetriaminepentaacetic acid (DTPA) as either Ca-DTPA or Zn-DTPA is superior to ethylenediamine tetraacetic acid (EDTA); DTPA was approved by the FDA to treat internal contamination with plutonium, americium, and curium, but it also chelates berkelium, californium, and any material with an atomic number >92. Ca-DTPA is more effective than Zn-DTPA during the first 24 hours after internal contamination, and both drugs are equally effective after the initial 24 hours. If both drugs are available, Ca-DTPA should be given as the first dose. If additional treatment is needed, treatment should be switched to Zn-DTPA. The dose is 1 g Ca-DTPA or Zn-DTPA, dissolved in 250 mL of normal saline or 5% glucose, given intravenously over 1 hour daily. The duration of chelation treatment depends on the amount of internal contamination and the individual response to treatment. DTPA also can be administrated by nebulized inhalation; 1 g is given in 1:1 dilution with water or saline over 15–20 minutes. Nebulized Zn-DTPA is recommended if the internal contamination is only by inhalation. The IV route is recommended and should be used if the route of internal contamination is not known or if multiple routes of internal contamination are likely. Treating uranium contamination with DTPA is contraindicated due to its synergistic damage to the kidneys.

Lung lavage can reduce radiation-induced pneumonitis and is indicated only when a large amount of radionuclide enters the lungs and has the potential for acute radiation injury. The procedure requires anesthesia. **Table 223-2** summarizes the common treatment regimens for internal radionuclide contamination.

Table 223-2 Common Drugs[a] for Treatment of Internal Contamination

Medication	Administered for Radionuclides	Route of Administration	Dosage	Duration	Mechanism of Action
KI	^{131}I	PO	130 mg/d for adults >40 with thyroid exposure >500 cGy; 130 mg/d for adults 18–40 with thyroid exposure >10 cGy; 130 mg/d for pregnant or lactating women with thyroid exposure >5 cGy; 65 mg/d for children and adolescents 3–18 with thyroid exposure >5 cGy; 32.5 mg/d for infants 1 mo to 3 y with thyroid exposure >5 cGy; 16 mg/d for neonates from birth to 1 mo with thyroid exposure >5 cGy	7–14 d	Blocking agent
Zn-DTPA	Plutonium, *trans*-plutonium, yttrium, americium, curium	IV	1 g in 250 mL NS or 5% glucose, given in 1–2 h, *or* bolus over 3–4 min	Up to 5 d	Chelating agent
		Inhalation	1 g in 1:1 dilution with water or NS over 15–20 min		
		IM	1 g; not recommended because of pain		
Ca-DTPA	Plutonium, *trans*-plutonium, yttrium, americium, curium	IV	1 g in 250 mL NS or 5% glucose, given in 1–2 h, *or* bolus over 3–4 min	Up to 5 d	Chelating agent
		Inhalation	1 g in 1:1 dilution with water or NS over 15–20 min		
		IM	1 g; not recommended because of pain		
Bicarbonate	Uranium	IV	2 ampoules sodium bicarbonate (44.3 meq each, 7.5%) in 1000 mL NS, 125 mL/L, *or* 1 ampoule of sodium bicarbonate (44.3 meq,	Usually IV for the first 24 h, PO for additional 2 d;	Increased excretion via the kidneys

			7.5%) in 500 mL NS, 500 mL/h	continuation of treatment for >3 d is rare and can be done according to titration of uranium amounts in the body	
		PO	2 tablets every 4 h until urine pH = 7–8, *or* 4 g (8 tablets) 3 tid		
Prussian blue	Cesium-137	PO	1 g tid with 100–200 mL water, up to 10 g/d	=3 wk titrated by urine and fecal bioassay and whole-body counting	Ion exchanger
Water	Tritium (H-3)	PO	>3–4 L per d	3 wk	Excretion of water
Aluminum phosphate gel	Strontium	PO	100 mL immediately after exposure	Once	Decreased gut absorption
Aluminum hydroxide		PO	60–100 mL	Once	Decreased gut absorption

[a]Excluding KI, these drugs have not been approved for this purpose by the U.S. Food and Drug Administration at the time of publication.

Abbreviations: NS, normal saline.

Medical Assay of the Radiation-Exposed Patient

One of the major difficulties in treating victims exposed to radiation is determination of the amount of exposure. Clinical assessment of the patient is the best approach. Biodosimetry, when available, can lead to better assessment of the level of exposure. The clinical assessment is based primarily on the timing and severity of the prodrome of ARS. Appearance of an early prodrome indicates high exposure to radiation. Victims who arrive at the hospital complaining of severe weakness, nausea, vomiting, diarrhea, or seizures probably will not survive despite supportive measures. Decontamination and the use of radiation-detection equipment are both very important. Few tests can be performed to estimate the radiation exposure and the contamination. Biodosimetry Assessment Tool (BAT) is a tool to aid treatment decisions during radiation exposure incidents; it was developed by the U.S. Armed Forces Radiobiology Research Institute (*http://www.afrri.usuhs.mil*). Baseline laboratory tests should include a complete blood count with differential and platelet count, renal evaluation, and determination of electrolytes. Urine and stool samples should be obtained if internal contamination is suspected. Nasal swabs should be taken from each nostril for determination of inhalation of radionuclides. The nasal swabs are useful if taken within the first 1–2 hours after the exposure. After exhalation, each swab is labeled and sealed in a plastic bag and sent for analysis to appropriate laboratories. Patients exposed to 0.7–4 Gy will develop pancytopenia from as early as 10 days to as long as 8 weeks after exposure. Lymphocytes show the most rapid decline, whereas other leukocytes and platelets decline less rapidly. Erythrocytes are the least vulnerable blood elements.

Absolute lymphocyte counts should be taken every 4–6 hours for 5–6 days; they are the most valuable early indicator because they are recognized to be a sensitive marker for radiation damage and correlate with both the exposure and the prognosis. A 50% drop in absolute lymphocyte count within the first 24 hours indicates a significant injury. HLA typing is necessary whenever there is suspicion of irreversible bone marrow damage. Lymphocyte chromosomal analysis can detect radiation exposure as low as 0.03–0.06 Gy, and 15 mL of blood should be drawn as early as possible in a heparinized collection tube and kept cool. Radiation-induced chromosomal aberrations in peripheral blood lymphocytes include dicentric chromosomes and ring forms that last for a few weeks. Calibration of a dose-response curve makes it possible to assess the radiation dose. Dicentric quantification requires multiple days to perform and is available only in select centers.

Another method for estimating exposure is the in vitro cytokinesis-block micronucleus assay. Micronuclei can be the result of small acentric chromosome fragments that arise during exposure to radiation. The technique to score the micronuclei in peripheral blood lymphocytes has been standardized in the last few years. It can be a useful tool in small-scale exposure but is not feasible in a mass casualty setting. An algorithm for the treatment of radiation casualties is shown in **Fig. 223-3**.

Figure 223-3

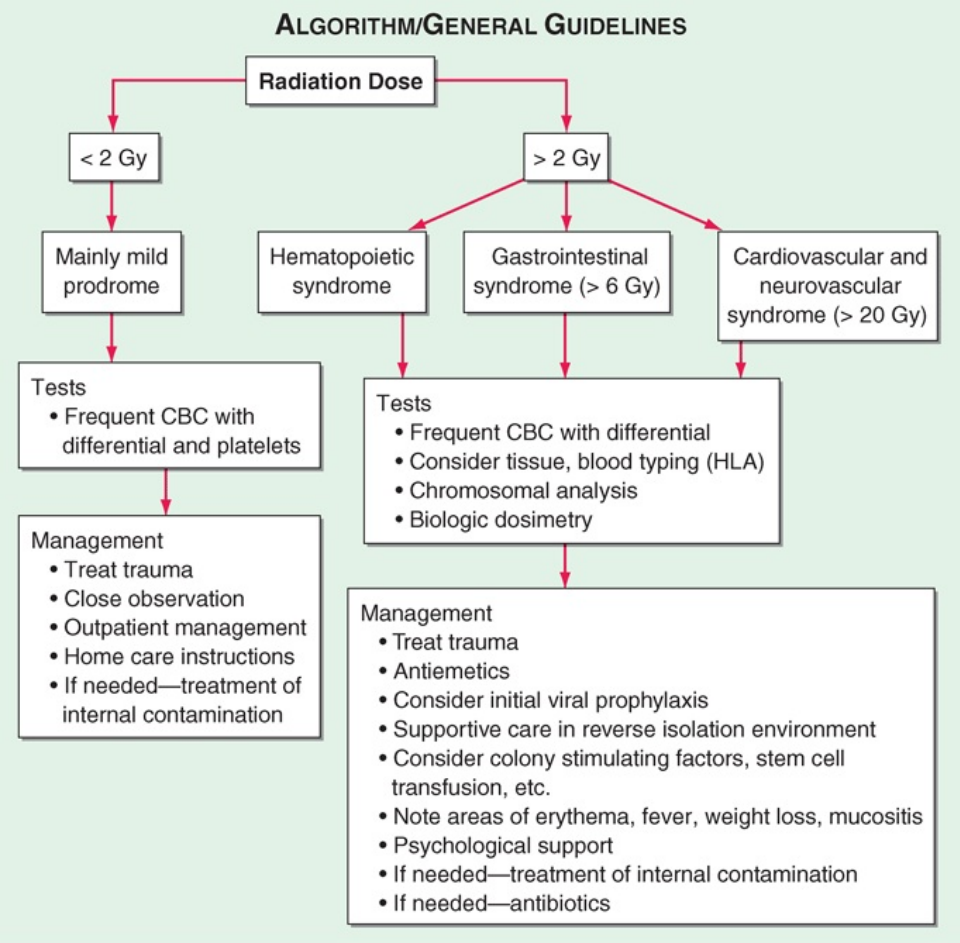

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

General guidelines for treatment of radiation casualties. CBC, complete blood count.

Follow-Up

It is desirable to continue follow-up in some circumstances. In general, only persons who are exposed to <8–10 Gy whole-body irradiation have a chance to survive in the long term, and they are at risk of developing cataracts, sterility, and lung, kidney, and bone marrow problems. Based on their age, their gender, and the amount and type of exposure, they should be followed for many years. A major public health issue is the risk of secondary malignancy in individuals and populations that were exposed to low doses of radiation. Leukemia and breast, brain, thyroid, and lung cancer are the most common, but the exposed population is at increased risk for many other cancers as well. Appropriate follow-up protocols should be developed, based on the type of exposure and the exposed population. In cases of internal contamination, the long-term follow-up should be focused on the organ at risk. Such is the case with uranium contamination, with its nephrotoxic properties.

FURTHER READINGS

Dainiak N: Hematologic consequences of exposure to ionizing radiation. Exp Hematol 30:513, 2002[PMID: 12063018]

Harper FT et al: Realistic radiological dispersal device hazard boundaries and ramifications for early consequence management decisions. Health Phys 93:1, 2007[PMID: 17563488]

Homeland Security Council: Planning Guidance for Response to a Nuclear Detonation. *http://www.remm.nlm.gov/planning-guidance.pdf 1/2009*

Lawrence Livermore National Laboratory: Key Response Planning Factors for the Aftermath of Nuclear Terrorism. *http://www.remm.nlm.gov/IND_ResponsePlanning_LLNL-TR-410067.pdf*

Mettler FA JR, Voelz GL: Major radiation exposure—what to expect and how to respond. N Engl J Med 346:1554, 2002[PMID: 12015396]

Moulder JE: Report on an interagency workshop on the radiobiology of nuclear terrorism: Molecular and Cellular Biology of Moderate Dose (1–10 Sv) Radiation and Potential Mechanisms of Radiation Protection (Bethesda, MD, December 17–18, 2001). Radiat Res 158:118, 2002[PMID: 12071812]

U.S. Department of Health and Human Services: Radiation event medical management: Guidance on diagnosis

& treatment for health care providers. http://www.remm.nlm.gov/remm_Preplanning.htm

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE BLOOD VESSEL

Vascular Ultrastructure

Blood vessels participate in homeostasis on a moment-to-moment basis and contribute to the pathophysiology of diseases of virtually every organ system. Hence, an understanding of the fundamentals of vascular biology furnishes a foundation for understanding the normal function of all organ systems and many diseases. The smallest blood vessels—capillaries—consist of a monolayer of endothelial cells apposed to a basement membrane, adjacent to occasional smooth-muscle-like cells known as *pericytes*(**Fig. 224-1*A***). Unlike larger vessels, pericytes do not invest the entire microvessel to form a continuous sheath. Veins and arteries typically have a trilaminar structure **(Fig. 224-1*B*–*E*)**. The *intima* consists of a monolayer of endothelial cells continuous with those of the capillaries. The middle layer, or *tunica media*, consists of layers of smooth-muscle cells; in veins, the media can contain just a few layers of smooth-muscle cells **(Fig. 224-1*B*).** The outer layer, the *adventitia*, consists of looser extracellular matrix with occasional fibroblasts, mast cells, and nerve terminals. Larger arteries have their own vasculature, the *vasa vasorum*, which nourishes the outer aspects of the tunica media. The adventitia of many veins surpasses the intima in thickness.

Figure 224-1

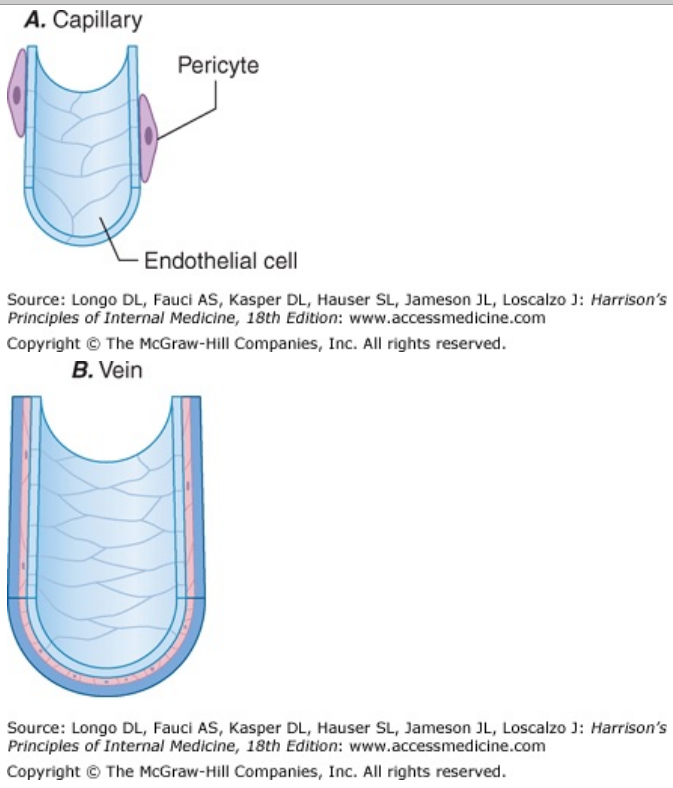

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

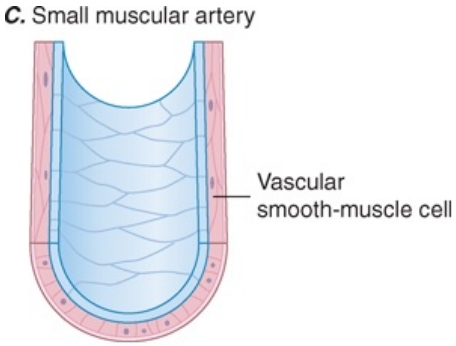

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

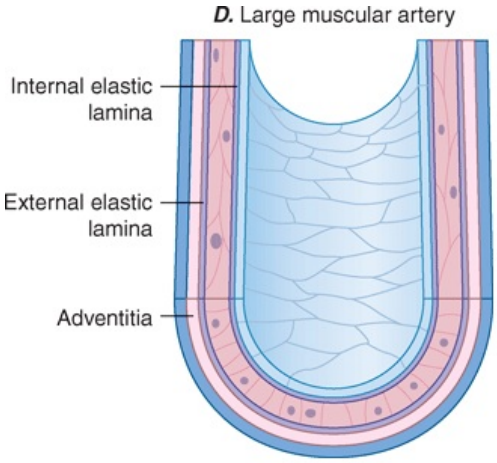

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

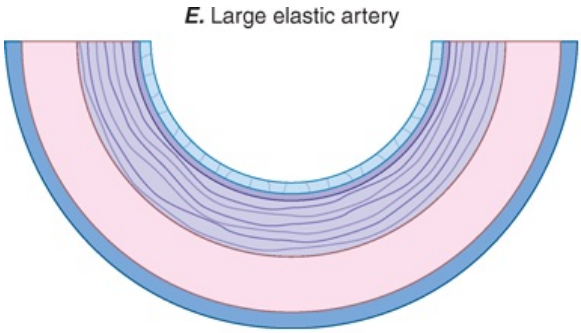

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematics of the structures of various types of blood vessels. ***A.*** Capillaries consist of an endothelial tube in contact with a discontinuous population of pericytes. ***B.*** Veins typically have thin medias and thicker adventitias. ***C.*** A small muscular artery features a prominent tunica media. ***D.*** Larger muscular arteries have a prominent media with smooth-muscle cells embedded in a complex extracellular matrix. ***E.*** Larger elastic arteries have cylindrical layers of elastic tissue alternating with concentric rings of smooth-muscle cells.

The tone of muscular arterioles regulates blood pressure and flow through various arterial beds. These smaller arteries have a relatively thick tunica media in relation to the adventitia **(Fig. 224-1*C*).** Medium-size muscular arteries similarly contain a prominent tunica media **(Fig. 224-1*D*)**; atherosclerosis commonly affects this type of muscular artery. The larger elastic arteries have a much more structured tunica media consisting of concentric bands of smooth-muscle cells, interspersed with strata of elastin-rich extracellular matrix sandwiched between layers of smooth-muscle cells **(Fig. 224-1*E*)**. Larger arteries have a clearly demarcated internal elastic lamina that forms the barrier between the intima and the media. An external elastic lamina demarcates the media of arteries from the surrounding adventitia.

Origin of Vascular Cells

The intima in human arteries often contains occasional resident smooth-muscle cells beneath the monolayer of vascular endothelial cells. The embryonic origin of smooth-muscle cells in various types of artery differs. Some upper-body arterial smooth-muscle cells derive from the neural crest, whereas lower-body arteries generally recruit smooth-muscle cells from neighboring mesodermal structures during

development. Derivatives of the proepicardial organ, which gives rise to the epicardial layer of the heart, contribute to the vascular smooth-muscle cells of the coronary arteries. Recent evidence suggests that bone marrow may give rise to both vascular endothelial cells and smooth-muscle cells, particularly under conditions of injury repair or vascular lesion formation. Indeed, the ability of bone marrow to repair an injured endothelial monolayer may contribute to maintenance of vascular health, whereas failure to do so may lead to arterial disease. The precise sources of endothelial and mesenchymal progenitor cells or their stem cell precursors remain the subject of active investigation **(Chaps. 65, 66, and 67)**.

Vascular Cell Biology

ENDOTHELIAL CELL

The key cell of the vascular intima, the endothelial cell, has manifold functions in health and disease. Most obviously, the endothelium forms the interface between tissues and the blood compartment. It therefore must regulate the entry of molecules and cells into tissues in a selective manner. The ability of endothelial cells to serve as a selectively permeable barrier fails in many vascular disorders, including atherosclerosis and hypertension. This dysregulation of permeability also occurs in pulmonary edema and other situations of "capillary leak."

The endothelium also participates in the local regulation of blood flow and vascular caliber. Endogenous substances produced by endothelial cells such as prostacyclin, endothelium-derived hyperpolarizing factor, nitric oxide (NO), and hydrogen peroxide (H_2O_2) provide tonic vasodilatory stimuli under physiologic conditions in vivo **(Table 224-1)**. Impaired production or excess catabolism of NO impairs this endothelium-dependent vasodilator function and may contribute to excessive vasoconstriction in various pathologic situations. By contrast, endothelial cells also produce potent vasoconstrictor substances such as endothelin in a regulated fashion. Excessive production of reactive oxygen species, such as superoxide anion (O_2^-), by endothelial or smooth-muscle cells under pathologic conditions (e.g., excessive exposure to angiotensin II) can promote local oxidative stress and inactivate NO.

Table 224–1. Endothelial Functions in Health and Disease

Homeostatic Phenotype	Dysfunctional Phenotype
Vasodilation	Impaired dilation, vasoconstriction
Antithrombotic, profibrinolytic	Prothrombotic, antifibrinolytic
Anti-inflammatory	Proinflammatory
Antiproliferative	Proproliferative
Antioxidant	Prooxidant
Permselectivity	Impaired barrier function

The endothelial monolayer contributes critically to inflammatory processes involved in normal host defenses and pathologic states. The normal endothelium resists prolonged contact with blood leukocytes; however, when activated by bacterial products such as endotoxin or proinflammatory cytokines released during infection or injury, endothelial cells express an array of leukocyte adhesion molecules that bind various classes of leukocytes. The endothelial cells appear to recruit selectively different classes of leukocytes in different pathologic conditions. The gamut of adhesion molecules and chemokines generated during acute bacterial infection tends to recruit granulocytes. In chronic inflammatory diseases such as tuberculosis and atherosclerosis, endothelial cells express adhesion molecules that favor the recruitment of mononuclear leukocytes that characteristically accumulate in these conditions.

The endothelium also dynamically regulates thrombosis and hemostasis. Nitric oxide, in addition to its vasodilatory properties, can limit platelet activation and aggregation. Like NO, prostacyclin produced by endothelial cells under normal conditions not only provides a vasodilatory stimulus but also antagonizes platelet activation and aggregation. Thrombomodulin expressed on the surface of endothelial cells binds thrombin at low concentrations and inhibits coagulation through activation of the protein C pathway, inactivating clotting factors Va and VIIIa and thus combating thrombus formation. The surface of endothelial cells contains heparan sulfate glycosaminoglycans that furnish an endogenous antithrombotic coating to the vasculature. Endothelial cells also participate actively in fibrinolysis and its regulation. They express receptors for plasminogen and plasminogen activators and produce tissue-type plasminogen activator. Through local generation of plasmin, the normal endothelial monolayer can promote the lysis of nascent thrombi.

When activated by inflammatory cytokines, bacterial endotoxin, or angiotensin II, for example, endothelial cells can produce substantial quantities of the major inhibitor of fibrinolysis, plasminogen activator inhibitor 1 (PAI-1). Thus, in pathologic circumstances, the endothelial cell may promote local thrombus accumulation rather than combat it. Inflammatory stimuli also induce the expression of the potent procoagulant tissue factor, a contributor to disseminated intravascular coagulation in sepsis.

Endothelial cells also participate in the pathophysiology of a number of immune-mediated diseases. Lysis of endothelial cells mediated by complement provides an example of immunologically mediated tissue

injury. The presentation of foreign histocompatibility complex antigens by endothelial cells in solid-organ allografts can trigger immunologic rejection. In addition, immune-mediated endothelial injury may contribute in some patients with thrombotic thrombocytopenic purpura and patients with hemolytic-uremic syndrome. Thus, in addition to contributing to innate immune responses, endothelial cells participate actively in both humoral and cellular limbs of the immune response.

Endothelial cells regulate growth of the subjacent smooth-muscle cells as well. Heparan sulfate glycosaminoglycans elaborated by endothelial cells can hold smooth-muscle proliferation in check. In contrast, when exposed to various injurious stimuli, endothelial cells can elaborate growth factors and chemoattractants, such as platelet-derived growth factor, that can promote the migration and proliferation of vascular smooth-muscle cells. Dysregulated elaboration of these growth-stimulatory molecules may promote smooth-muscle accumulation in atherosclerotic lesions.

CLINICAL ASSESSMENT OF ENDOTHELIAL FUNCTION

Various invasive and noninvasive approaches can be used to evaluate endothelial vasodilator function in humans. Either pharmacologic agonists or increased flow stimulates the endothelium to release acutely molecular effectors that alter underlying smooth-muscle cell tone. Invasively, infusion of the cholinergic agonists acetylcholine and methacholine stimulates the release of NO from normal endothelial cells. Changes in coronary diameter can be quantitatively measured in response to an intracoronary infusion of these short-lived, rapidly acting agents. Noninvasive assessment of endothelial function in the forearm circulation typically involves occlusion of brachial artery blood flow with a blood pressure cuff, which elicits reactive hyperemia after release; the resulting flow increase normally causes endothelium-dependent vasodilation, which is measured as the change in brachial artery blood flow and diameter by ultrasound **(Fig. 224-2)**. This approach depends on shear stress–dependent changes in endothelial release of NO after restoration of blood flow, as well as the effect of adenosine released (transiently) from ischemic tissue in the forearm.

Figure 224-2

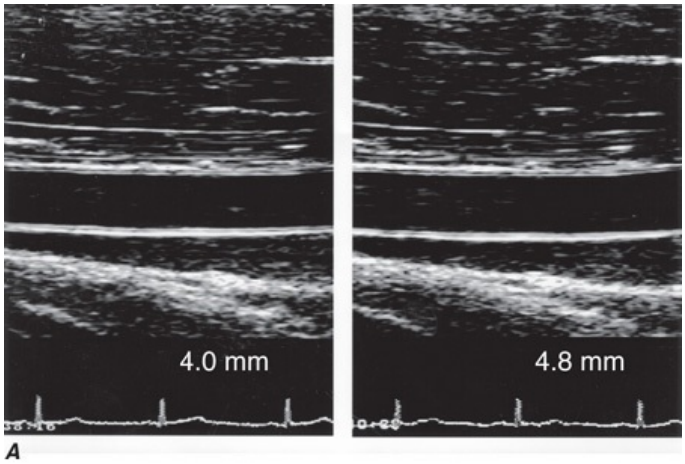

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

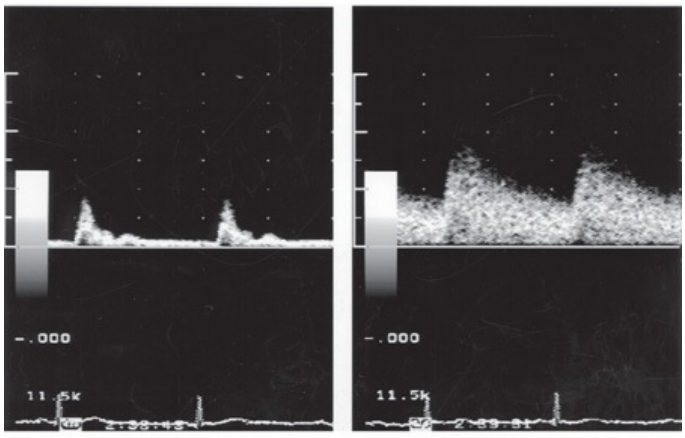

B Baseline Hyperemia

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

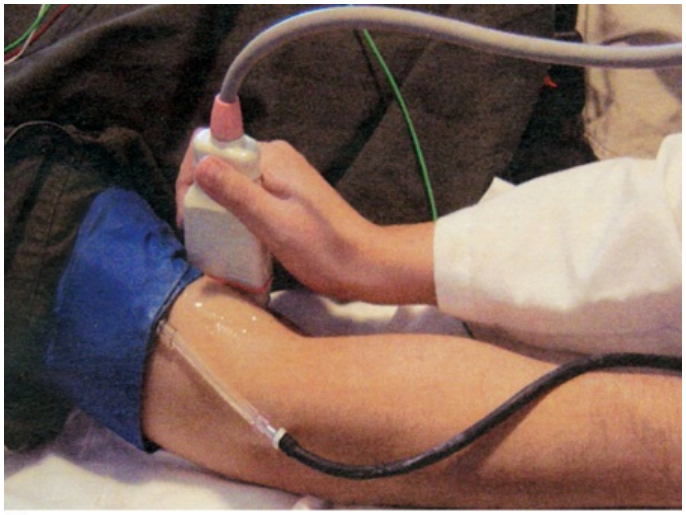

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Assessment of endothelial function in vivo using blood pressure cuff-occlusion and release. Upon deflation of the cuff, changes in diameter (***A***) and blood flow (***B***) of the brachial artery are monitored with an ultrasound probe (***C***). *(Reproduced with permission of J. Vita, MD.)*

Typically, these invasive and noninvasive approaches detect inducible vasodilatory changes in vessel diameter of ~10%. In individuals with atherosclerosis or its risk factors (especially hypertension, hypercholesterolemia, diabetes mellitus, and smoking), such studies can detect endothelial dysfunction as defined by a smaller change in diameter and, in the extreme case, a so-called paradoxical vasoconstrictor response owing to the direct effect of cholinergic agonists on vascular smooth-muscle cell tone.

VASCULAR SMOOTH-MUSCLE CELL

The vascular smooth-muscle cell, the major cell type of the media layer of blood vessels, also contributes actively to vascular pathobiology. Contraction and relaxation of smooth-muscle cells at the level of the muscular arteries controls blood pressure, and, hence, regional blood flow and the afterload experienced by the left ventricle (see below). The vasomotor tone of veins, which is governed by smooth-muscle cell tone, regulates the capacitance of the venous tree and influences the preload experienced by both ventricles. Smooth-muscle cells in the adult vessel seldom replicate. This homeostatic quiescence of smooth-muscle cells changes in conditions of arterial injury or inflammatory activation. Proliferation and migration of arterial smooth-muscle cells, which is associated with a change in phenotype characterized by lower content of contractile proteins and greater production of extracellular matrix macromolecules, can contribute to the development of arterial stenoses in atherosclerosis, arteriolar remodeling that can sustain and propagate hypertension, and the hyperplastic response of arteries injured by angioplasty or stent deployment. In the pulmonary circulation, smooth-muscle migration and proliferation contribute decisively to the pulmonary vascular disease that gradually occurs in response to sustained high-flow states such as left-to-right shunts. Such pulmonary vascular disease provides a major obstacle to the management of many patients with adult congenital heart disease. Elucidation of the signaling pathways that regulate the reversible transition of the vascular smooth-muscle cell phenotype remains an active focus of investigation. Among other mediators, microRNAs have emerged as powerful regulators of this transition, offering new targets for intervention.

The activated, phenotypically modulated smooth-muscle cells secrete the bulk of vascular extracellular matrix. Excessive production of collagen and glycosaminoglycans contributes to the remodeling and altered biology and biomechanics of arteries affected by hypertension or atherosclerosis. In larger elastic arteries, the elastin synthesized by smooth-muscle cells serves to maintain not only normal arterial structure but also hemodynamic function. The ability of the larger arteries, such as the aorta, to store the kinetic energy of systole promotes tissue perfusion during diastole. Arterial stiffness associated with aging or disease, as manifested by a widening pulse pressure, increases left ventricular afterload and portends a poor outcome.

Like endothelial cells, vascular smooth-muscle cells do not merely respond to vasomotor or inflammatory stimuli elaborated by other cell types but can themselves serve as a source of such stimuli. For example, when exposed to bacterial endotoxin or other proinflammatory stimuli, smooth-muscle cells can elaborate cytokines and other inflammatory mediators. Like endothelial cells, upon inflammatory activation, arterial smooth-muscle cells can produce prothrombotic mediators such as tissue factor, the antifibrinolytic protein PAI-1, and other molecules that modulate thrombosis and fibrinolysis. Smooth-muscle cells also elaborate autocrine growth factors that can amplify hyperplastic responses to arterial injury.

VASCULAR SMOOTH-MUSCLE CELL FUNCTION

Vascular smooth-muscle cells govern vessel tone. Those cells contract when stimulated by a rise in intracellular calcium concentration by calcium influx through the plasma membrane and by calcium release from intracellular stores **(Fig. 224-3)**. In vascular smooth-muscle cells, voltage-dependent L-type calcium channels open with membrane depolarization, which is regulated by energy-dependent ion pumps such as the Na^+,K^+-ATPase pump and ion channels such as the Ca^{2+}-sensitive K^+ channel. Local changes in intracellular calcium concentration, termed *calcium sparks*, result from the influx of calcium through the voltage-dependent calcium channel and are caused by the coordinated activation of a cluster of ryanodine-sensitive calcium release channels in the sarcoplasmic reticulum (see below). Calcium sparks directly augment intracellular calcium concentration and indirectly increase intracellular calcium concentration by activating chloride channels. In addition, calcium sparks reduce smooth-muscle contractility by activating large-conductance calcium-sensitive K^+ channels, hyperpolarizing the cell membrane and thereby limiting further voltage-dependent increases in intracellular calcium.

Figure 224-3

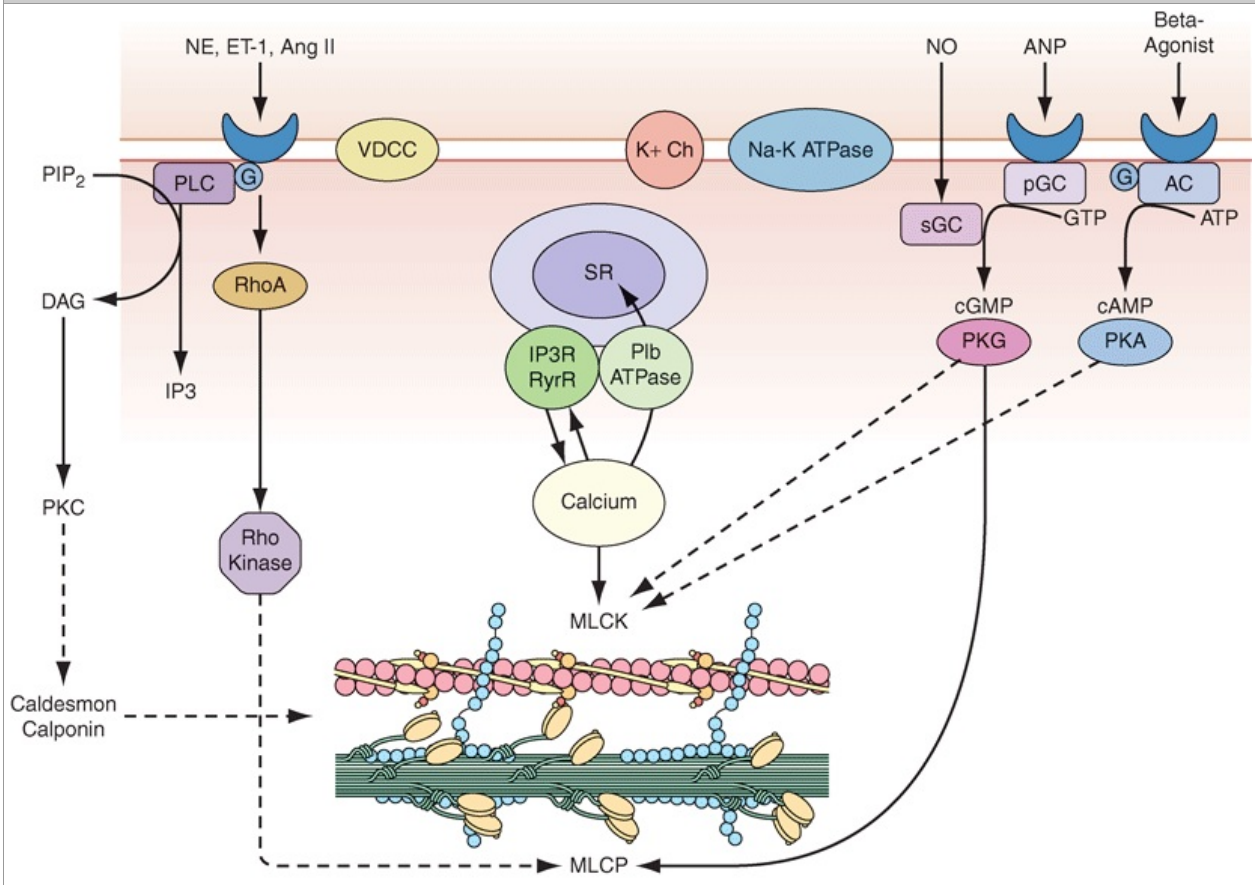

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Regulation of vascular smooth-muscle cell calcium concentration and actomyosin ATPase-dependent contraction. AC, adenylyl cyclase; Ang II, angiotensin II; ANP, antrial natriuretic peptide; DAG, diacylglycerol; ET-1, endothelin-1; G, G-protein; IP_3, inositol 1,4,5-trisphosphate; MLCK, myosin light chain kinase; MLCP, myosin light chain phosphatase; NE, norepinephrine; NO, nitric oxide; pGC, particular guanylyl cyclase; PIP_2, phosphatidylinositol 4,5-bisphosphate; PKA, protein kinase A; PKC, protein kinase C; PKG, protein kinase G; PLC, phospholipase C; sGC, soluble guanylyl cyclase; SR, sarcoplasmic reticulum; VDCC, voltage-dependent calcium channel. *(Modified from B Berk, in Vascular Medicine, 3rd ed, p 23. Philadelphia, Saunders, Elsevier, 2006; with permission.)*

Biochemical agonists also increase intracellular calcium concentration, in this case by receptor-dependent activation of phospholipase C with hydrolysis of phosphatidylinositol 4,5-bisphosphate, resulting in generation of diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3). These membrane lipid derivatives in turn activate protein kinase C and increase intracellular calcium concentration. In addition, IP_3 binds to specific receptors on the sarcoplasmic reticulum membrane to increase calcium efflux from this calcium storage pool into the cytoplasm.

Vascular smooth-muscle cell contraction is controlled principally by the phosphorylation of myosin light chain, which in the steady state depends on the balance between the actions of myosin light chain kinase and myosin light chain phosphatase. Calcium activates myosin light chain kinase through the formation of a calcium-calmodulin complex. Phosphorylation of myosin light chain by this kinase

augments myosin ATPase activity and enhances contraction. Myosin light chain phosphatase dephosphorylates myosin light chain, reducing myosin ATPase activity and contractile force. Phosphorylation of the myosin-binding subunit (thr695) of myosin light chain phosphatase by Rho kinase inhibits phosphatase activity and induces calcium sensitization of the contractile apparatus. Rho kinase is itself activated by the small GTPase RhoA, which is stimulated by guanosine exchange factors and inhibited by GTPase-activating proteins.

Both cyclic AMP and cyclic GMP relax vascular smooth-muscle cells through complex mechanisms. β agonists, acting through their G-protein-coupled receptors activate adenylyl cyclase to convert ATP to cyclic AMP; NO and atrial natriuretic peptide acting directly and via a G-protein-coupled receptor, respectively, activate guanylyl cyclase to convert GTP to cyclic GMP. These agents in turn activate protein kinase A and protein kinase G, respectively, which inactivate myosin light chain kinase and decrease vascular smooth-muscle cell tone. In addition, protein kinase G can interact directly with the myosin-binding substrate subunit of myosin light chain phosphatase, increasing phosphatase activity and decreasing vascular tone. Finally, several mechanisms drive NO-dependent, protein kinase G–mediated reductions in vascular smooth-muscle cell calcium concentration, including phosphorylation-dependent inactivation of RhoA; decreased IP_3 formation; phosphorylation of the IP_3 receptor–associated cyclic GMP kinase substrate, with subsequent inhibition of IP_3 receptor function; phosphorylation of phospholamban, which increases calcium ATPase activity and sequestration of calcium in the sarcoplasmic reticulum; and protein kinase G–dependent stimulation of plasma membrane calcium ATPase activity, perhaps by activation of the Na^+,K^+-ATPase pump or hyperpolarization of the cell membrane by activation of calcium-dependent K^+ channels.

CONTROL OF VASCULAR SMOOTH-MUSCLE CELL TONE

The tone of vascular smooth-muscle cells is governed by the autonomic nervous system and by the endothelium in tightly regulated control networks. Autonomic neurons enter the blood vessel medial layer from the adventitia and modulate vascular smooth-muscle cell tone in response to baroreceptors and chemoreceptors within the aortic arch and carotid bodies and in response to thermoreceptors in the skin. These regulatory components include rapidly acting reflex arcs modulated by central inputs that respond to sensory inputs (olfactory, visual, auditory, and tactile) as well as emotional stimuli. Three classes of nerves mediate autonomic regulation of vascular tone: *sympathetic*, whose principal neurotransmitters are epinephrine and norepinephrine; *parasympathetic*, whose principal neurotransmitter is acetylcholine; and *nonadrenergic/noncholinergic*, which include two subgroups—nitrergic, whose principal neurotransmitter is NO, and peptidergic, whose principal neurotransmitters are substance P, vasoactive intestinal peptide, calcitoningene-related peptide, and ATP.

Each of these neurotransmitters acts through specific receptors on the vascular smooth-muscle cell to modulate intracellular calcium and, consequently, contractile tone. Norepinephrine activates α receptors, and epinephrine activates α and β receptors (adrenergic receptors); in most blood vessels, norepinephrine activates postjunctional α_1 receptors in large arteries and α_2 receptors in small arteries and arterioles, leading to vasoconstriction. Most blood vessels express β_2-adrenergic receptors on their vascular smooth-muscle cells and respond to β agonists by cyclic AMP–dependent relaxation. Acetylcholine released from parasympathetic neurons binds to muscarinic receptors (of which there are five subtypes, $M_{1–5}$) on vascular smooth-muscle cells to yield vasorelaxation. In addition, NO stimulates presynaptic neurons to release acetylcholine, which can stimulate the release of NO from the endothelium. Nitrergic neurons release NO produced by neuronal NO synthase, which causes vascular smooth-muscle cell relaxation via the cyclic GMP–dependent and –independent mechanisms described above. The peptidergic neurotransmitters all potently vasodilate, acting either directly or through endothelium-dependent NO release to decrease vascular smooth-muscle cell tone. **For the detailed molecular physiology of the autonomic nervous system, see Chap. 375.**

The endothelium modulates vascular smooth-muscle tone by the direct release of several effectors, including NO, prostacyclin, hydrogen sulfide, and endothelium-derived hyperpolarizing factor, all of which cause vasorelaxation, and endothelin, which causes vasoconstriction. The release of these endothelial effectors of vascular smooth-muscle cell tone is stimulated by mechanical (shear stress, cyclic strain, etc.) and biochemical mediators (purinergic agonists, muscarinic agonists, peptidergic agonists), with the biochemical mediators acting through endothelial receptors specific to each class. In addition to these local paracrine modulators of vascular smooth-muscle cell tone, circulating mediators can affect tone, including norepinephrine and epinephrine, vasopressin, angiotensin II, bradykinin, and the natriuretic peptides (ANP, BNP, CNP, and DNP), as discussed above.

Vascular Regeneration

Growth of new blood vessels can occur in response to conditions such as chronic hypoxemia and tissue ischemia. Growth factors, including vascular endothelial growth factor (VEGF) and forms of fibroblast growth factor (FGF), activate a signaling cascade that stimulates endothelial proliferation and tube formation, defined as *angiogenesis*. The development of collateral vascular networks in the ischemic myocardium reflects this process and can result from selective activation of endothelial progenitor cells, which may reside in the blood vessel wall or home to the ischemic tissue subtended by an occluded or severely stenotic vessel from the bone marrow. True arteriogenesis, or the development of a new blood

vessel that includes all three cell layers, normally does not occur in the cardiovascular system of adult mammals. The molecular mechanisms and progenitor cells that can recapitulate blood vessel development de novo are under rapidly advancing study (Chaps. 65, 66, and 67).

Vascular Pharmacogenomics

The last decade has witnessed considerable progress in efforts to define the genetic differences underlying individual variations in vascular pharmacologic responses. Many investigators have focused on receptors and enzymes associated with neurohumoral modulation of vascular function as well as hepatic enzymes that metabolize drugs that affect vascular tone. The genetic polymorphisms thus far associated with differences in vascular response often (but not invariably) relate to functional differences in the activity or expression of the receptor or enzyme of interest. Some of these polymorphisms appear to have different allele frequencies in specific ethnic groups. A summary of recently identified polymorphisms defining these vascular pharmacogenomic differences is provided in **Table 224-2. For more detailed discussion, see Chap. 5.**

Table 224–2. Genetic Polymorphisms in Vascular Function and Disease Risk

Gene	Polymorphic Allele	Clinical Implications
α-adrenergic receptors		
α1A	Arg492Cys	None
α2B	Glu9/G1712	Increased CHD events
α2C	A2cDcl3232-325	Ethnic differences in risk of hypertension or heart failure
Angiotensin-converting enzyme (ACE)	Insertion/deletion polymorphism in intron 16	D allele or DD genotype-increased response to ACE inhibitors; inconsistent data for increased risk of atherosclerotic heart disease, and hypertension
Ang II type I receptor	1166A→C Ala-Cys	Increased response to Ang II and increased risk of pregnancy-associated hypertension
β-Adrenergic receptors		
β1		
	Ser49Gly	Increased HR and DCM risk
	Arg389Gly	Increased heart failure in blacks
β2		
	Arg16Gly	Familial hypertension, increased obesity risk
	Glu27Gln	Hypertension in white type II diabetics
	Thr164Ile	Decreased agonist affinity and worse HF outcome
B2-Bradykinin receptor	Cys58Thr, Cys412Gly, Thr21Met	Increased risk of hypertension in some ethnic groups
Endothelial nitric oxide synthase (eNOS)	Nucleotide repeats in introns 4 and 13, Glu298Asp	Increased MI and venous thrombosis
	Thr785Cys	Early coronary artery disease

Abbreviations: CHD, coronary heart disease; HR, heart rate; DCM, dilated cardiomyopathy; HF, heart failure; MI, myocardial infarction.

Source: Derived from B Schaefer et al: Heart Dis 5:129, 2003.

CELLULAR BASIS OF CARDIAC CONTRACTION

Cardiac Ultrastructure

About three-fourths of the ventricular mass is composed of cardiomyocytes, normally 60–140 μm in length and 17–25 μm in diameter **(Fig. 224-4*A*)**. Each cell contains multiple, rodlike cross-banded strands (myofibrils) that run the length of the cell and are composed of serially repeating structures, the sarcomeres. The cytoplasm between the myofibrils contains other cell constituents, including the single centrally located nucleus, numerous mitochondria, and the intracellular membrane system, the sarcoplasmic reticulum.

Figure 224-4

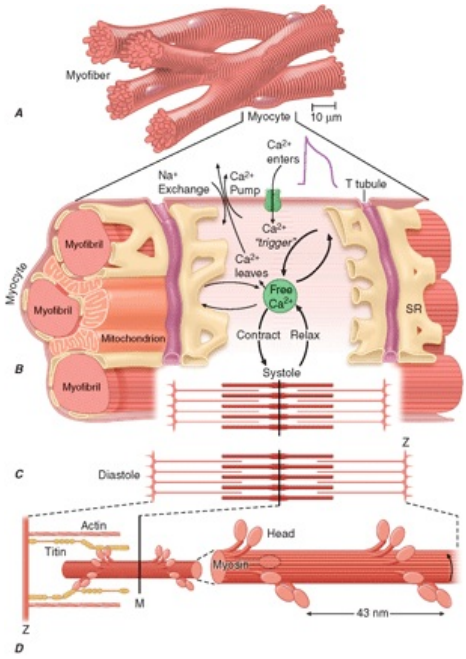

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A shows the branching myocytes making up the cardiac myofibers. ***B*** illustrates the critical role played by the changing [Ca^{2+}] in the myocardial cytosol. Ca^{2+} ions are schematically shown as entering through the calcium channel that opens in response to the wave of depolarization that travels along the sarcolemma. These Ca^{2+} ions "trigger" the release of more calcium from the sarcoplasmic reticulum (SR) and thereby initiate a contraction-relaxation cycle. Eventually the small quantity of Ca^{2+} that has entered the cell leaves predominantly through an Na^{+}/Ca^{2+} exchanger, with a lesser role for the sarcolemmal Ca^{2+} pump. The varying actin-myosin overlap is shown for (***B***) systole, when [Ca^{2+}] is maximal, and (***C***) diastole, when [Ca^{2+}] is minimal. ***D.*** The myosin heads, attached to the thick filaments, interact with the thin actin filaments. *(From LH Opie, Heart Physiology, reprinted with permission. Copyright LH Opie, 2004.)*

The *sarcomere*, the structural and functional unit of contraction, lies between adjacent Z lines, which are dark repeating bands that are apparent on transmission electron microscopy. The distance between Z lines varies with the degree of contraction or stretch of the muscle and ranges between 1.6 and 2.2 μm. Within the confines of the sarcomere are alternating light and dark bands, giving the myocardial fibers their striated appearance under the light microscope. At the center of the sarcomere is a dark band of constant length (1.5 μm), the A band, which is flanked by two lighter bands, the I bands, which are of variable length. The sarcomere of heart muscle, like that of skeletal muscle, consists of two sets of interdigitating myofilaments. Thicker filaments, composed principally of the protein myosin, traverse the A band; they are about 10 nm (100 Å) in diameter, with tapered ends. Thinner filaments, composed primarily of actin, course from the Z lines through the I band into the A band; they are approximately 5 nm (50 Å) in diameter and 1.0 μm in length. Thus, thick and thin filaments overlap only within the (dark) A band, whereas the (light) I band contains only thin filaments. On electron-microscopic examination, bridges may be seen to extend between the thick and thin filaments within the A band; these are myosin heads (see below) bound to actin filaments.

The Contractile Process

The sliding filament model for muscle contraction rests on the fundamental observation that both the thick and the thin filaments are constant in overall length during both contraction and relaxation. With activation, the actin filaments are propelled farther into the A band. In the process, the A band remains constant in length, whereas the I band shortens and the Z lines move toward one another.

The *myosin* molecule is a complex, asymmetric fibrous protein with a molecular mass of about 500,000 Da; it has a rodlike portion that is about 150 nm (1500 Å) in length with a globular portion (head) at its end. These globular portions of myosin form the bridges between the myosin and actin molecules and are the site of ATPase activity. In forming the thick myofilament, which is composed of ~300 longitudinally stacked myosin molecules, the rodlike segments of the myosin molecules are laid down in an orderly, polarized manner, leaving the globular portions projecting outward so that they can interact with actin to generate force and shortening **(Fig. 224-4*B*)**.

Actin has a molecular mass of about 47,000 Da. The thin filament consists of a double helix of two chains of actin molecules wound about each other on a larger molecule, tropomyosin. A group of

regulatory proteins—troponins C, I, and T—are spaced at regular intervals on this filament **(Fig. 224-5)**. In contrast to myosin, actin lacks intrinsic enzymatic activity but does combine reversibly with myosin in the presence of ATP and Ca^{2+}. The calcium ion activates the myosin ATPase, which in turn breaks down ATP, the energy source for contraction (Fig. 224-5). The activity of myosin ATPase determines the rate of forming and breaking of the actomyosin cross-bridges and ultimately the velocity of muscle contraction. In relaxed muscle, tropomyosin inhibits this interaction. *Titin***(Fig. 224-4*D*)** is a large, flexible, myofibrillar protein that connects myosin to the Z line; its stretching contributes to the elasticity of the heart. Dystrophin is a long cytoskeletal protein that has an amino-terminal actin-binding domain and a carboxy-terminal domain that binds to the dystroglycan complex at adherens junctions on the cell membrane, thus tethering the sarcomere to the cell membrane at regions tightly coupled to adjacent contracting myocytes. Mutations in components of the dystrophin complex lead to muscular dystrophy and associated cardiomyopathy.

Figure 224-5

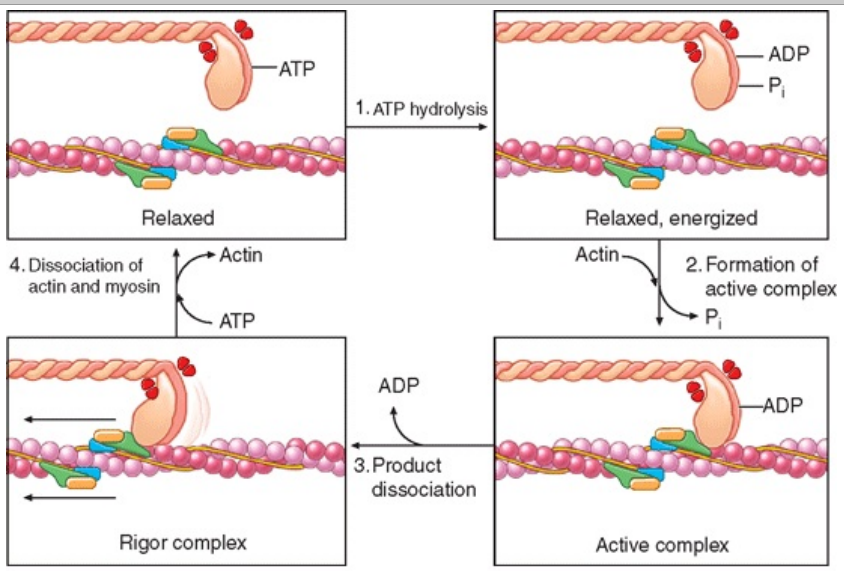

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Four steps in cardiac muscle contraction and relaxation. In relaxed muscle (***upper left***), ATP bound to the myosin cross-bridge dissociates the thick and thin filaments. ***Step 1:*** Hydrolysis of myosin-bound ATP by the ATPase site on the myosin head transfers the chemical energy of the nucleotide to the activated cross-bridge **(*upper right*)**. When cytosolic Ca^{2+} concentration is low, as in relaxed muscle, the reaction cannot proceed because tropomyosin and the troponin complex on the thin filament do not allow the active sites on actin to interact with the cross-bridges. Therefore, even though the cross-bridges are energized, they cannot interact with actin. ***Step 2:*** When Ca^{2+} binding to troponin C has exposed active sites on the thin filament, actin interacts with the myosin cross-bridges to form an active complex **(*lower right*)** in which the energy derived from ATP is retained in the actin-bound cross-bridge, whose orientation has not yet shifted. ***Step 3:*** The muscle contracts when ADP dissociates from the cross-bridge. This step leads to the formation of the low-energy rigor complex **(*lower left*)** in which the chemical energy derived from ATP hydrolysis has been expended to perform mechanical work (the "rowing" motion of the cross-bridge). ***Step 4:*** The muscle returns to its resting state, and the cycle ends when a new molecule of ATP binds to the rigor complex and dissociates the cross-bridge from the thin filament. This cycle continues until calcium is dissociated from troponin C in the thin filament, which causes the contractile proteins to return to the resting state with the cross-bridge in the energized state. ATP, adenosine triphosphate; ATPase, adenosine triphosphatase; ADP, adenosine diphosphate. *[From AM Katz: Heart failure: Cardiac function and dysfunction, in Atlas of Heart Diseases, 3d ed, WS Colucci (ed). Philadelphia, Current Medicine, 2002. Reprinted with permission.]*

During activation of the cardiac myocyte, Ca^{2+} becomes attached to one of three components of the heterotrimer troponin C, which results in a conformational change in the regulatory protein tropomyosin; the latter, in turn, exposes the actin cross-bridge interaction sites (Fig. 224-5). Repetitive interaction between myosin heads and actin filaments is termed *cross-bridge cycling*, which results in sliding of the actin along the myosin filaments, ultimately causing muscle shortening and/or the development of tension. The splitting of ATP then dissociates the myosin cross-bridge from actin. In the presence of ATP (Fig. 224-5), linkages between actin and myosin filaments are made and broken cyclically as long as sufficient Ca^{2+} is present; these linkages cease when $[Ca^{2+}]$ falls below a critical level, and the troponin-tropomyosin complex once more prevents interactions between the myosin cross-bridges and actin filaments **(Fig. 224-6)**.

Figure 224-6

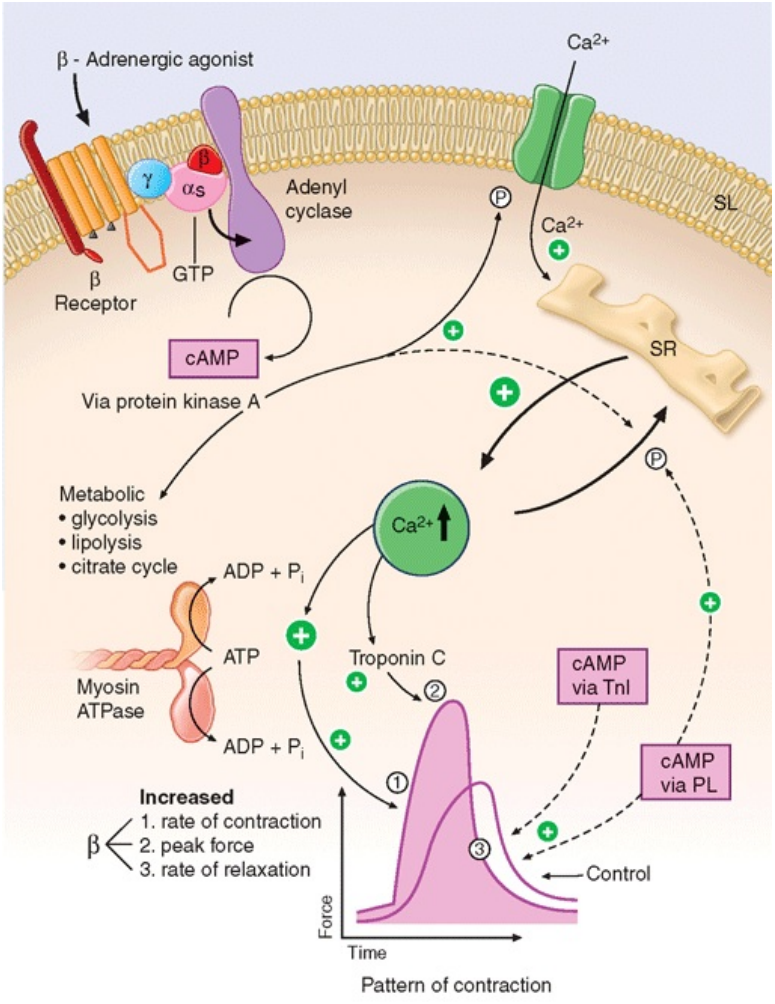

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Signal systems involved in positive inotropic and lusitropic (enhanced relaxation) effects of β-adrenergic stimulation. When the β-adrenergic agonist interacts with the β receptor, a series of G protein–mediated changes leads to activation of adenylyl cyclase and the formation of cyclic adenosine monophosphate (cAMP). The latter acts via protein kinase A to stimulate metabolism ***(left)*** and phosphorylate the Ca^{2+} channel protein ***(right)***. The result is an enhanced opening probability of the Ca^{2+} channel, thereby increasing the inward movement of Ca^{2+} ions through the sarcolemma (SL) of the T tubule. These Ca^{2+} ions release more calcium from the sarcoplasmic reticulum (SR) to increase cytosolic Ca^{2+} and activate troponin C. Ca^{2+} ions also increase the rate of breakdown of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) and inorganic phosphate (Pi). Enhanced myosin ATPase activity explains the increased rate of contraction, with increased activation of troponin C explaining increased peak force development. An increased rate of relaxation is explained by the fact that cAMP also activates the protein phospholamban, situated on the membrane of the SR, that controls the rate of uptake of calcium into the SR. The latter effect explains enhanced relaxation (lusitropic effect). P, phosphorylation; PL, phospholamban; TnI, troponin I. *(Modified from LH Opie, Heart Physiology, reprinted with permission. Copyright LH Opie, 2004.)*

Intracytoplasmic Ca^{2+} is a principal determinant of the inotropic state of the heart. Most agents that stimulate myocardial contractility (positive inotropic stimuli), including the digitalis glycosides and β-adrenergic agonists, increase the $[Ca^{2+}]$ in the vicinity of the myofilaments, which in turn triggers cross-bridge cycling. Increased impulse traffic in the cardiac adrenergic nerves stimulates myocardial contractility as a consequence of the release of norepinephrine from cardiac adrenergic nerve endings. Norepinephrine activates myocardial β receptors and, through the G_S-stimulated guanine nucleotide-binding protein, activates the enzyme adenylyl cyclase, which leads to the formation of the intracellular second messenger cyclic AMP from ATP (Fig. 224-6). Cyclic AMP in turn activates protein kinase A (PKA), which phosphorylates the Ca^{2+} channel in the myocardial sarcolemma, thereby enhancing the influx of Ca^{2+} into the myocyte. Other functions of PKA are discussed below.

The *sarcoplasmic reticulum* (SR) **(Fig. 224-7)**, a complex network of anastomosing intracellular channels, invests the myofibrils. Its longitudinally disposed tubules closely invest the surfaces of individual sarcomeres but have no direct continuity with the outside of the cell. However, closely related to the SR, both structurally and functionally, are the transverse tubules, or T system, formed by tubelike invaginations of the sarcolemma that extend into the myocardial fiber along the Z lines, i.e., the ends of the sarcomeres.

Figure 224-7

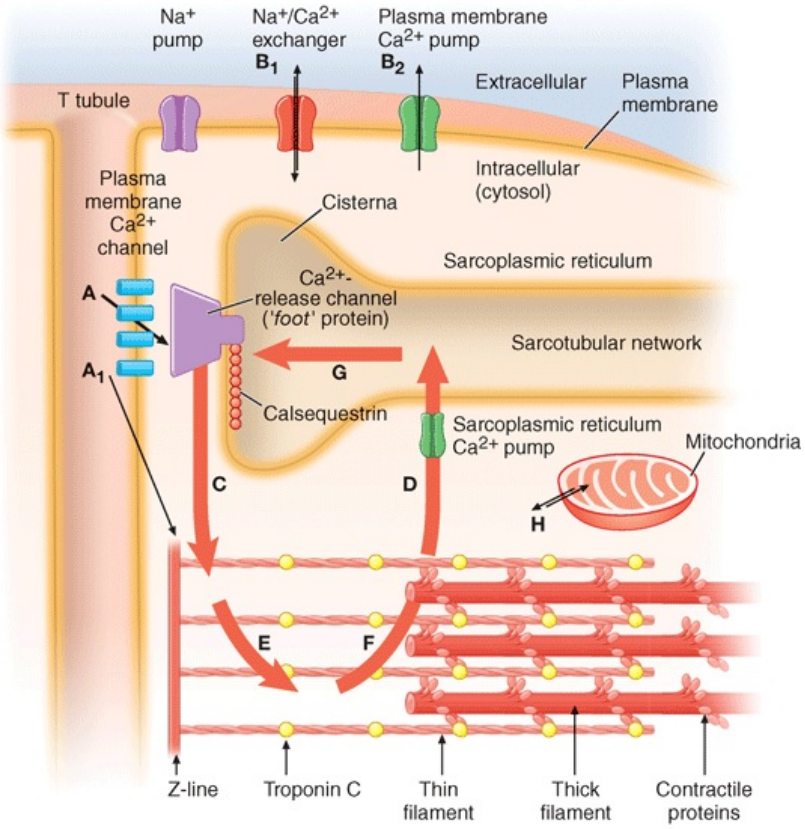

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The Ca^{2+} fluxes and key structures involved in cardiac excitation-contraction coupling. The arrows denote the direction of Ca^{2+} fluxes. The thickness of each arrow indicates the magnitude of the calcium flux. Two Ca^{2+} cycles regulate excitation-contraction coupling and relaxation. The larger cycle is entirely intracellular and involves Ca^{2+} fluxes into and out of the sarcoplasmic reticulum, as well as Ca^{2+} binding to and release from troponin C. The smaller extracellular Ca^{2+} cycle occurs when this cation moves into and out of the cell. The action potential opens plasma membrane Ca^{2+} channels to allow passive entry of Ca^{2+} into the cell from the extracellular fluid **(*arrow A*)**. Only a small portion of the Ca^{2+} that enters the cell directly activates the contractile proteins **(*arrow A*$_1$)**. The extracellular cycle is completed when Ca^{2+} is actively transported back out to the extracellular fluid by way of two plasma membrane fluxes mediated by the sodium-calcium exchanger **(*arrow B*$_1$)** and the plasma membrane calcium pump **(*arrow B*$_2$)**. In the intracellular Ca^{2+} cycle, passive Ca^{2+} release occurs through channels in the cisternae **(*arrow C*)** and initiates contraction; active Ca^{2+} uptake by the Ca^{2+} pump of the sarcotubular network **(*arrow D*)** relaxes the heart. Diffusion of Ca^{2+} within the sarcoplasmic reticulum **(*arrow G*)** returns this activator cation to the cisternae, where it is stored in a complex with calsequestrin and other calcium-binding proteins. Ca^{2+} released from the sarcoplasmic reticulum initiates systole when it binds to troponin C **(*arrow E*)**. Lowering of cytosolic [Ca^{2+}] by the sarcoplasmic reticulum (SR) causes this ion to dissociate from troponin **(*arrow F*)** and relaxes the heart. Ca^{2+} also may move between mitochondria and cytoplasm (H). *(Adapted from Katz, with permission.)*

Cardiac Activation

In the inactive state, the cardiac cell is electrically polarized; i.e., the interior has a negative charge relative to the outside of the cell, with a transmembrane potential of –80 to –100 mV (Chap. 231). The

sarcolemma, which in the resting state is largely impermeable to Na^+, has a Na^+- and K^+-stimulating pump energized by ATP that extrudes Na^+ from the cell; this pump plays a critical role in establishing the resting potential. Thus, intracellular [K^+] is relatively high and [Na^+] is far lower; conversely, extracellular [Na^+] is high and [K^+] is low. At the same time, in the resting state, extracellular [Ca^{2+}] greatly exceeds free intracellular [Ca^{2+}].

The action potential has four phases (Fig. 231-1*B*). During the plateau of the action potential (phase 2), there is a slow inward current through L-type Ca^{2+} channels in the sarcolemma (Fig. 224-7). The depolarizing current not only extends across the surface of the cell but penetrates deeply into the cell by way of the ramifying T tubular system. The absolute quantity of Ca^{2+} that crosses the sarcolemma and the T system is relatively small and by itself appears to be insufficient to bring about full activation of the contractile apparatus. However, this Ca^{2+} current triggers the release of much larger quantities of Ca^{2+} from the SR, a process termed *Ca^{2+}-induced Ca^{2+} release*. The latter is a major determinant of intracytoplasmic [Ca^{2+}] and therefore of myocardial contractility.

Ca^{2+} is released from the SR through a Ca^{2+} release channel, a cardiac isoform of the ryanodine receptor (RyR2), which controls intracytoplasmic [Ca^{2+}] and, as in vascular smooth-muscle cells, leads to the local changes in intracellular [Ca^{2+}] called calcium sparks. A number of regulatory proteins, including *calstabin 2*, inhibit RyR2 and thereby the release of Ca^{2+} from the SR. PKA dissociates calstabin from the RyR2, enhancing Ca^{2+} release and thereby myocardial contractility. Excessive plasma catecholamine levels and cardiac sympathetic neuronal release of norepinephrine cause hyperphosphorylation of PKA, leading to calstabin 2–depleted RyR2. The latter depletes SR Ca^{2+} stores and thereby impairs cardiac contraction, leading to heart failure, and also triggers ventricular arrhythmias.

The Ca^{2+} released from the SR then diffuses toward the myofibrils, where, as already described, it combines with troponin C (Fig. 224-6). By repressing this inhibitor of contraction, Ca^{2+} activates the myofilaments to shorten. During repolarization, the activity of the Ca^{2+} pump in the SR, the SR Ca^{2+} ATPase ($SERCA_2A$), reaccumulates Ca^{2+} against a concentration gradient, and the Ca^{2+} is stored in the SR by its attachment to a protein, *calsequestrin*. This reaccumulation of Ca^{2+} is an energy (ATP)-requiring process that lowers the cytoplasmic [Ca^{2+}] to a level that inhibits the actomyosin interaction responsible for contraction, and in this manner leads to myocardial relaxation. Also, there is an exchange of Ca^{2+} for Na^+ at the sarcolemma (Fig. 224-7), reducing the cytoplasmic [Ca^{2+}]. Cyclic AMP–dependent PKA phosphorylates the SR protein *phospholamban*; the latter, in turn, permits activation of the Ca^{2+} pump, thereby increasing the uptake of Ca^{2+} by the SR, accelerating the rate of relaxation and providing larger quantities of Ca^{2+} in the SR for release by subsequent depolarization, thereby stimulating contraction.

Thus, the combination of the cell membrane, transverse tubules, and SR, with their ability to transmit the action potential and release and then reaccumulate Ca^{2+}, plays a fundamental role in the rhythmic contraction and relaxation of heart muscle. Genetic or pharmacologic alterations of any component, whatever its etiology, can disturb these functions.

CONTROL OF CARDIAC PERFORMANCE AND OUTPUT

The extent of shortening of heart muscle and, therefore, the stroke volume of the ventricle in the intact heart depend on three major influences: (1) the length of the muscle at the onset of contraction, i.e., the preload; (2) the tension that the muscle is called on to develop during contraction, i.e., the afterload; and (3) the contractility of the muscle, i.e., the extent and velocity of shortening at any given preload and afterload. The major determinants of preload, afterload, and contractility are shown in **Table 224-3**.

Table 224–3. Determinants of Stroke Volume

I. Ventricular Preload
A. Blood volume
B. Distribution of blood volume
1. Body position
2. Intrathoracic pressure
3. Intrapericardial pressure
4. Venous tone

5. Pumping action of skeletal muscles
C. Atrial contraction
II. Ventricular Afterload
A. Systemic vascular resistance
B. Elasticity of arterial tree
C. Arterial blood volume
D. Ventricular wall tension
1. Ventricular radius
2. Ventricular wall thickness
III. Myocardial Contractility[a]
A. Intramyocardial $[Ca^{2+}]$ ↑↓
B. Cardiac adrenergic nerve activity ↑↓[b]
C. Circulating catecholamines ↑↓[b]
D. Cardiac rate ↑↓[b]
E. Exogenous inotropic agents↑
F. Myocardial ischemia ↓
G. Myocardial cell death (necrosis, apoptosis, autophagy) ↓
H. Alterations of sarcomeric and cytoskeletal proteins ↓
1. Genetic
2. Hemodynamic overload
I. Myocardial fibrosis ↓
J. Chronic overexpression of neurohormones ↓
K. Ventricular remodeling ↓
L. Chronic and/or excessive myocardial hypertrophy ↓

[a]Arrows indicate directional effects of determinants of contractility.

[b]Contractility rises initially but later becomes depressed.

THE ROLE OF MUSCLE LENGTH (PRELOAD)

The preload determines the length of the sarcomeres at the onset of contraction. The length of the sarcomeres associated with the most forceful contraction is ~2.2 μm. This length provides the optimum configuration for the interaction between the two sets of myofilaments. The length of the sarcomere also regulates the extent of activation of the contractile system, i.e., its sensitivity to Ca^{2+}. According to this concept, termed *length-dependent activation*, myofilament sensitivity to Ca^{2+} is also maximal at the optimal sarcomere length. The relation between the initial length of the muscle fibers and the developed force has prime importance for the function of heart muscle. This relationship forms the basis of Starling's law of the heart, which states that within limits, the force of ventricular contraction depends on the end-diastolic length of the cardiac muscle; in the intact heart, the latter relates closely to the ventricular end-diastolic volume.

CARDIAC PERFORMANCE

The ventricular end-diastolic or "filling" pressure sometimes is used as a surrogate for the end-diastolic volume. In isolated heart and heart-lung preparations, the stroke volume varies directly with the end-diastolic fiber length (preload) and inversely with the arterial resistance (afterload), and as the heart fails—i.e., as its contractility declines—it delivers a progressively smaller stroke volume from a normal or even elevated end-diastolic volume. The relation between the ventricular end-diastolic pressure and the stroke work of the ventricle (the ventricular function curve) provides a useful definition of the level of contractility of the heart in the intact organism. An increase in contractility is accompanied by a shift of the ventricular function curve upward and to the left (greater stroke work at any level of ventricular end-diastolic pressure, or lower end-diastolic volume at any level of stroke work), whereas a shift downward and to the right characterizes depression of contractility **(Fig. 224-8)**.

Figure 224-8

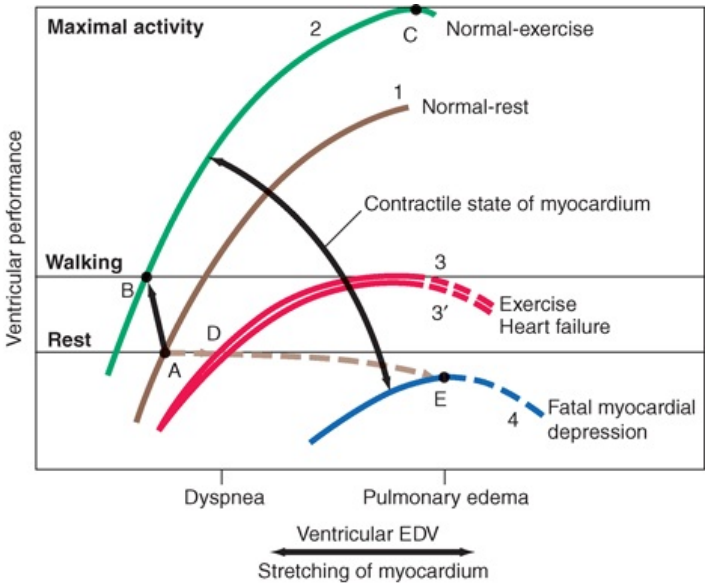

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The interrelations among influences on ventricular end-diastolic volume (EDV) through stretching of the myocardium and the contractile state of the myocardium. Levels of ventricular EDV associated with filling pressures that result in dyspnea and pulmonary edema are shown on the abscissa. Levels of ventricular performance required when the subject is at rest, while walking, and during maximal activity are designated on the ordinate. The broken lines are the descending limbs of the ventricular-performance curves, which are rarely seen during life but show the level of ventricular performance if end-diastolic volume could be elevated to very high levels. For further explanation, see text. *[Modified from WS Colucci and E Braunwald: Pathophysiology of heart failure, in Braunwald's Heart Disease, 7th ed, DP Zipes et al (eds). Philadelphia: Elsevier, 2005, pp 509–538.]*

VENTRICULAR AFTERLOAD

In the intact heart, as in isolated cardiac muscle, the extent and velocity of shortening of ventricular muscle fibers at any level of preload and of myocardial contractility relate inversely to the afterload, i.e., the load that opposes shortening. In the intact heart, the afterload may be defined as the tension developed in the ventricular wall during ejection. Afterload is determined by the aortic pressure as well as by the volume and thickness of the ventricular cavity. Laplace's law states that the tension of the myocardial fiber is the product of the intracavitary ventricular pressure and ventricular radius divided by wall thickness. Therefore, at any particular level of aortic pressure, the afterload on a dilated left ventricle exceeds that on a normal-sized ventricle. Conversely, at the same aortic pressure and ventricular diastolic volume, the afterload on a hypertrophied ventricle is lower than of a normal chamber. The aortic pressure in turn depends on the peripheral vascular resistance, the physical characteristics of the arterial tree, and the volume of blood it contains at the onset of ejection.

Ventricular afterload critically regulates cardiovascular performance **(Fig. 224-9)**. As already noted, elevations in both preload and contractility increase myocardial fiber shortening, whereas increases in afterload reduce it. The extent of myocardial fiber shortening and left ventricular size determine stroke volume. An increase in arterial pressure induced by vasoconstriction, for example, augments afterload, which opposes myocardial fiber shortening, reducing stroke volume.

Figure 224-9

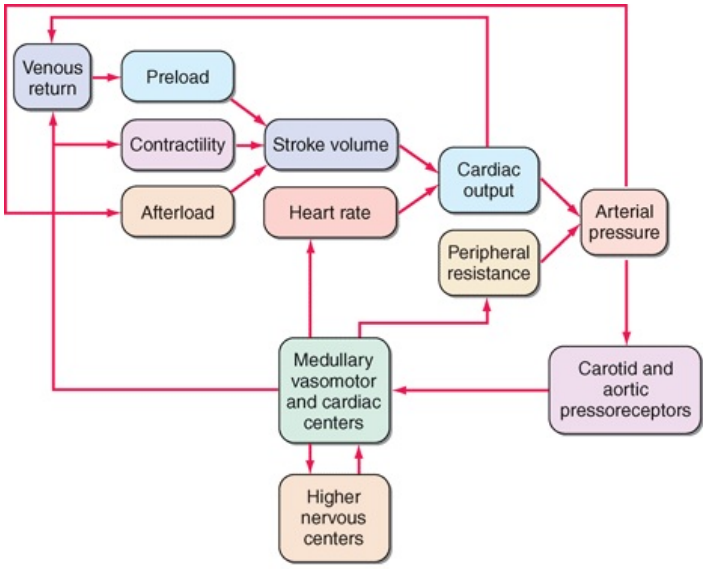

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Interactions in the intact circulation of preload, contractility, and afterload in producing stroke volume. Stroke volume combined with heart rate determines cardiac output, which, when combined with peripheral vascular resistance, determines arterial pressure for tissue perfusion. The characteristics of the arterial system also contribute to afterload, an increase which reduces stroke volume. The interaction of these components with carotid and aortic arch baroreceptors provides a feedback mechanism to higher medullary and vasomotor cardiac centers and to higher levels in the central nervous system to effect a modulating influence on heart rate, peripheral vascular resistance, venous return, and contractility. *[From MR Starling; Physiology of myocardial contraction, in Atlas of Heart Failure: Cardiac Function and Dysfunction, 3rd ed, WS Colucci and E Braunwald (eds). Philadelphia: Current Medicine, 2002, pp 19–35.]*

When myocardial contractility becomes impaired and the ventricle dilates, afterload rises (Laplace's law) and limits cardiac output. Increased afterload also may result from neural and humoral stimuli that occur in response to a fall in cardiac output. This increased afterload may reduce cardiac output further, thereby increasing ventricular volume and initiating a vicious circle, especially in patients with ischemic heart disease and limited myocardial O_2 supply. Treatment with vasodilators has the opposite effect; when afterload is reduced, cardiac output rises (Chap. 234).

Under normal circumstances, the various influences acting on cardiac performance enumerated above interact in a complex fashion to maintain cardiac output at a level appropriate to the requirements of the metabolizing tissues (Fig. 224-9); interference with a single mechanism may not influence the cardiac output. For example, a moderate reduction of blood volume or the loss of the atrial contribution to ventricular contraction ordinarily can be sustained without a reduction in the cardiac output at rest. Under these circumstances, other factors, such as increases in the frequency of adrenergic nerve impulses to the heart, heart rate, and venous tone, will serve as compensatory mechanisms and sustain cardiac output in a normal individual.

EXERCISE

The integrated response to exercise illustrates the interactions among the three determinants of stroke volume: preload, afterload, and contractility (Fig. 224-8). Hyperventilation, the pumping action of the exercising muscles, and venoconstriction during exercise all augment venous return and hence ventricular filling and preload (Table 224-3). Simultaneously, the increase in the adrenergic nerve impulse traffic to the myocardium, the increased concentration of circulating catecholamines, and the tachycardia that occur during exercise combine to augment the contractility of the myocardium (Fig. 224-8, curves 1 and 2) and together elevate stroke volume and stroke work, without a change in or even a reduction of end-diastolic pressure and volume (Fig. 224-8, points A and B). Vasodilation occurs in the exercising muscles, thus tending to limit the increase in arterial pressure that otherwise would occur as cardiac output rises to levels as high as five times greater than basal levels during maximal exercise. This vasodilation ultimately allows the achievement of a greatly elevated cardiac output during exercise at an arterial pressure only moderately higher than in the resting state.

ASSESSMENT OF CARDIAC FUNCTION

Several techniques can define impaired cardiac function in clinical practice. The cardiac output and stroke volume may be depressed in the presence of heart failure, but not uncommonly, these variables are within normal limits in this condition. A somewhat more sensitive index of cardiac function is the ejection fraction, i.e., the ratio of stroke volume to end-diastolic volume (normal value = 67 ± 8%), which is frequently depressed in systolic heart failure even when the stroke volume itself is normal.

Alternatively, abnormally elevated ventricular end-diastolic volume (normal value = 75 ± 20 mL/m^2) or end-systolic volume (normal value = 25 ± 7 mL/m^2) signifies impairment of left ventricular systolic function.

Noninvasive techniques, particularly echocardiography as well as radionuclide scintigraphy and cardiac magnetic resonance imaging (MRI) (Chap. 229), have great value in the clinical assessment of myocardial function. They provide measurements of end-diastolic and end-systolic volumes, ejection fraction, and systolic shortening rate, and they allow assessment of ventricular filling (see below) as well as regional contraction and relaxation. The latter measurements are particularly important in ischemic heart disease, as myocardial infarction causes regional myocardial damage.

A limitation of measurements of cardiac output, ejection fraction, and ventricular volumes in assessing cardiac function is that ventricular loading conditions strongly influence these variables. Thus, a depressed ejection fraction and lowered cardiac output may be observed in patients with normal ventricular function but reduced preload, as occurs in hypovolemia, or with increased afterload, as occurs in acutely elevated arterial pressure.

The end-systolic left ventricular pressure-volume relationship is a particularly useful index of ventricular performance since it does not depend on preload and afterload **(Fig. 224-10)**. At any level of myocardial contractility, left ventricular end-systolic volume varies inversely with end-systolic pressure; as contractility declines, end-systolic volume (at any level of end-systolic pressure) rises.

Figure 224-10

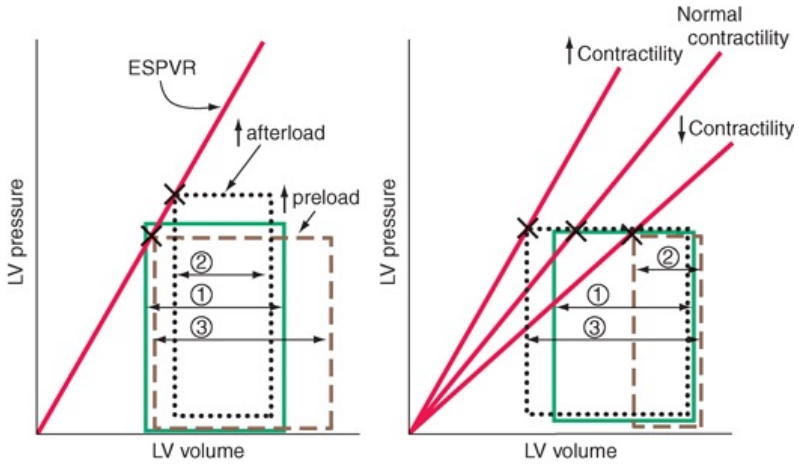

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The responses of the left ventricle to increased afterload, increased preload, and increased and reduced contractility are shown in the pressure-volume plane. ***Left.*** Effects of increases in preload and afterload on the pressure-volume loop. Since there has been no change in contractility, ESPVR (the end-systolic pressure-volume relation) is unchanged. With an increase in afterload, stroke volume falls (1 → 2); with an increase in preload, stroke volume rises (1 → 3). ***Right.*** With increased myocardial contractility and constant LV end-diastolic volume, the ESPVR moves to the left of the normal line (lower end-systolic volume at any end-systolic pressure) and stroke volume rises (1 → 3). With reduced myocardial contractility, the ESPVR moves to the right; end-systolic volume is increased, and stroke volume falls (1 → 2).

Diastolic Function

Ventricular filling is influenced by the extent and speed of myocardial relaxation, which in turn depends on the rate of uptake of Ca^{2+} by the SR; the latter may be enhanced by adrenergic activation and reduced by ischemia, which reduces the ATP available for pumping Ca^{2+} into the SR (see above). The stiffness of the ventricular wall also may impede filling. Ventricular stiffness increases with hypertrophy and conditions that infiltrate the ventricle, such as amyloid, or is caused by an extrinsic constraint (e.g., pericardial compression) **(Fig. 224-11)**.

Figure 224-11

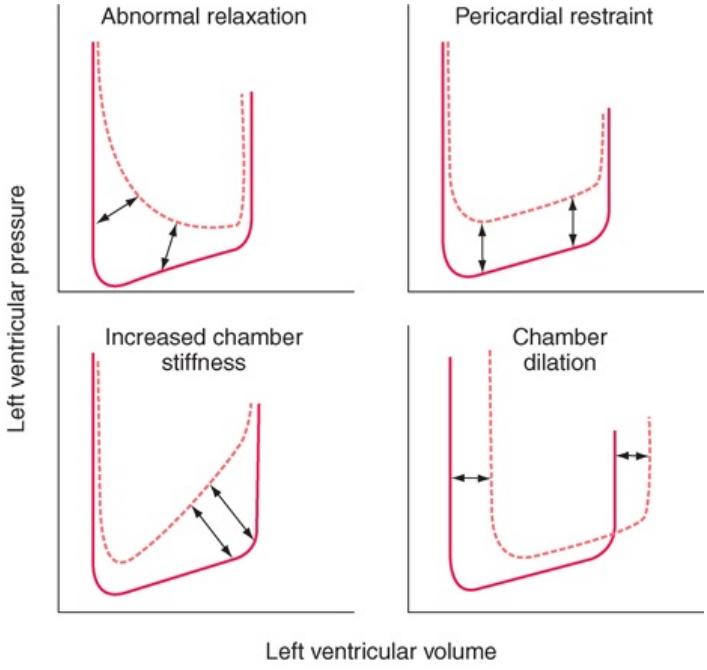

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanisms that cause diastolic dysfunction reflected in the pressure-volume relation. The bottom half of the pressure-volume loop is depicted. Solid lines represent normal subjects; broken lines represent patients with diastolic dysfunction. *(From JD Carroll et al: The differential effects of positive inotropic and vasodilator therapy on diastolic properties in patients with congestive cardiomyopathy. Circulation 74:815, 1986; with permission.)*

Ventricular filling can be assessed by continuously measuring the velocity of flow across the mitral valve, using Doppler ultrasound. Normally, the velocity of inflow is more rapid in early diastole than during atrial systole; with mild to moderately impaired relaxation, the rate of early diastolic filling declines, whereas the rate of presystolic filling rises. With further impairment of filling, the pattern is "pseudo-normalized," and early ventricular filling becomes more rapid as left atrial pressure upstream to the stiff left ventricle rises.

Cardiac Metabolism

The heart requires a continuous supply of energy (in the form of ATP) not only to perform its mechanical pumping functions, but also to regulate intracellular and transsarcolemmal ionic movements and concentration gradients. Among its pumping functions, the development of tension, the frequency of contraction, and the level of myocardial contractility are the principal determinants of the heart's substantial energy needs, making its O_2 requirements approximately 15% of that of the entire organism.

Most ATP production depends on the oxidation of substrate [glucose and free fatty acids (FFAs)]. Myocardial FFAs are derived from circulating FFAs, which result principally from lipolysis in adipose tissue, whereas the myocyte's glucose derives from plasma as well as from the cell's breakdown of its glycogen stores (glycogenolysis). These two principal sources of acetyl coenzyme A in cardiac muscle vary reciprocally. Glucose is broken down in the cytoplasm into a three-carbon product, pyruvate, which passes into the mitochondria, where it is metabolized to the two-carbon fragment, acetyl-Co-A, and undergoes oxidation. FFAs are converted to acyl-CoA in the cytoplasm and acetyl-CoA in the mitochondria. Acetyl-CoA enters the citric acid (Krebs) cycle to produce ATP by oxidative phosphorylation within the mitochondria; ATP then enters the cytoplasm from the mitochondrial compartment. Intracellular ADP, resulting from the breakdown of ATP, enhances mitochondrial ATP production.

In the fasted, resting state, circulating FFA concentrations and their myocardial uptake are high, and they furnish most of the heart's acetyl-CoA (~70%). In the fed state, with elevations of blood glucose and insulin, glucose oxidation increases and FFA oxidation subsides. Increased cardiac work, the administration of inotropic agents, hypoxia, and mild ischemia all enhance myocardial glucose uptake, glucose production resulting from glycogenolysis, and glucose metabolism to pyruvate (glycolysis). By contrast, β-adrenergic stimulation, as occurs during stress, raises the circulating levels and metabolism of FFAs in favor of glucose. Severe ischemia inhibits the cytoplasmic enzyme pyruvate dehydrogenase, and despite both glycogen and glucose breakdown, glucose is metabolized only to lactic acid (anaerobic glycolysis), which does not enter the citric acid cycle. Anaerobic glycolysis produces much less ATP than does aerobic glucose metabolism, in which glucose is metabolized to pyruvate and subsequently oxidized to CO_2. High concentrations of circulating FFAs, which can occur when adrenergic stimulation is superimposed on severe ischemia, reduce oxidative phosphorylation and also cause ATP wastage; the myocardial content of ATP declines and impairs myocardial contraction. In addition, products of FFA breakdown can exert toxic effects on cardiac cell membranes and may be arrhythmogenic.

Myocardial energy is stored as creatine phosphate (CP), which is in equilibrium with ATP, the immediate source of energy. In states of reduced energy availability, the CP stores decline first. Cardiac hypertrophy, fibrosis, tachycardia, increased wall tension resulting from ventricular dilation, and increased intracytoplasmic [Ca^{2+}] all contribute to increased myocardial energy needs. When coupled with reduced coronary flow reserve, as occurs with obstruction of coronary arteries or abnormalities of the coronary microcirculation, an imbalance in myocardial ATP production relative to demand may occur, and the resulting ischemia can worsen or cause heart failure.

DEVELOPMENTAL BIOLOGY OF THE CARDIOVASCULAR SYSTEM

The heart is the first organ to form during embryogenesis (**Fig. 224-12**) and must accomplish the simultaneous challenges of circulating blood, nutrients, and oxygen to the other forming organs while continuing to grow and undergo complex morphogenetic changes. Early progenitors of the heart arise within very early crescent-shaped fields of lateral splanchnic mesoderm under the influence of multiple signals, including those derived from neural ectoderm long before neural tube closure. Early cardiac precursors express regulatory transcription factors that play reiterated roles in cardiac development, such as *NKX2-5* and *GATA4;* these mutations are responsible for some forms of inherited congenital heart disease. Early cardiac precursors form two bilateral heart tubes, each composed of a single cell layer of endocardium surrounded by a single layer of myocardial precursors. Subsequently, a single midline heart tube is formed by the medial migration and midline fusion of these bilateral structures. The caudal, inflow region of the heart tube adopts a more rostral final position and represents the atrial anlagen, whereas the rostral, outflow portion of the tube forms the truncus arteriosus, which divides to produce the aorta and the proximal pulmonary artery. Between these extremes lie the structural precursors of the ventricles.

Figure 224-12

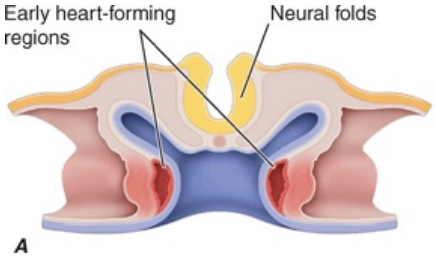

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

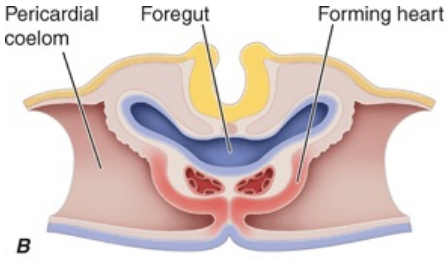

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

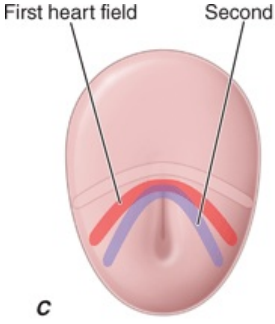

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

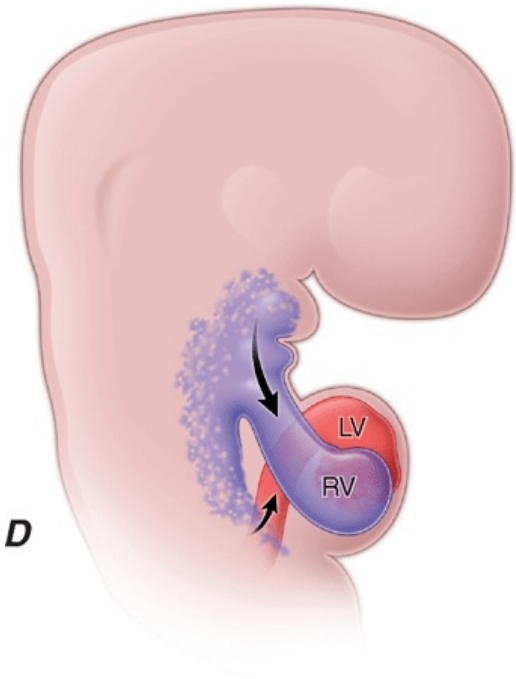

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

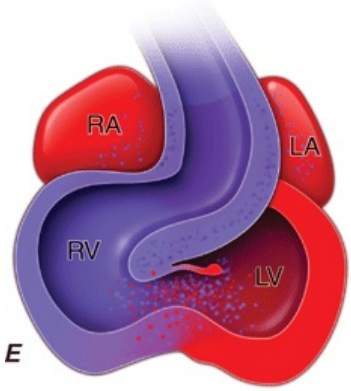

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

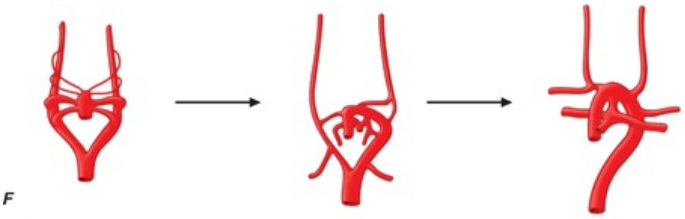

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Schematic depiction of a transverse section through an early embryo depicts the bilateral regions where early heart tubes form. ***B.*** The bilateral heart tubes subsequently migrate to the midline and fuse to form the linear heart tube. ***C.*** At the early cardiac crescent stage of embryonic development, cardiac precursors include a primary heart field fated to form the linear heart tube and a second heart field fated to add myocardium to the inflow and outflow poles of the heart. ***D.*** Second heart field cells populate the pharyngeal region before subsequently migrating to the maturing heart. ***E.*** Large portions of the right ventricle and outflow tract and some cells within the atria derive from the second heart field. ***F.*** The aortic arch arteries form as symmetric sets of vessels that then remodel under the influence of the neural crest to form the asymmetric mature vasculature.

The linear heart tube undergoes an asymmetric looping process (the first gross evidence of left-right asymmetry in the developing embryo), which positions the portion of the heart tube destined to become the left ventricle to the left of the more rostral precursors of the right ventricle and outflow tract. Looping is coordinated with chamber specification and ballooning of various regions of the heart tube to produce the presumptive atria and ventricles.

Relatively recent work has demonstrated that significant portions of the right ventricle are formed by cells that are added to the developing heart after looping has occurred. These cells, which are derived from what is called the second heart field, derive from progenitors in the ventral pharynx and express markers that allow for their identification, including islet-1. Different embryologic origins of cells within the right and left ventricles may help explain why some forms of congenital and adult heart diseases affect these regions of the heart to varying degrees.

After looping and chamber formation, a series of septation events divide the left and right sides of the heart, separate the atria from the ventricles, and form the aorta and pulmonary artery from the truncus arteriosus. Cardiac valves form between the atria and the ventricles and between the ventricles and the outflow vessels. Early in development, the single layer of myocardial cells secretes an extracellular matrix rich in hyaluronic acid. This extracellular matrix, termed "cardiac jelly," accumulates within the endocardial cushions, precursors of the cardiac valves. Signals from overlying myocardial cells, including members of the transforming growth factor β family, trigger migration, invasion, and phenotypic changes of underlying endocardial cells, which undergo an epithelial-mesenchymal transformation and invade the cardiac jelly to cellularize the endocardial cushions. Mesenchymal components proliferate and remodel to form the mature valve leaflets.

The great vessels form as a series of bilaterally symmetric aortic arch arteries that undergo asymmetric remodeling events to form the mature vasculature. The immigration of neural crest cells that arise in the dorsal neural tube orchestrates this process. These cells are required for aortic arch remodeling and septation of the truncus arteriosus. They develop into smooth-muscle cells within the tunica media of the aortic arch, the ductus arteriosus, and the carotid arteries. Smooth-muscle cells within the descending aorta arise from a different embryologic source, the lateral plate mesoderm. Neural crest cells are sensitive to both vitamin A and folic acid, and congenital heart disease involving abnormal remodeling of the aortic arch arteries has been associated with maternal deficiencies of these vitamins. Genetic syndromes associated with aortic arch defects can be associated with other abnormalities of neural crest craniofacial derivatives, including the palate.

Coronary artery formation requires yet another cell population that initiates extrinsic to the embryonic heart fields. Epicardial cells arise in the proepicardial organ, a derivative of the septum transversum, which also contributes to the fibrous portion of the diaphragm and to the liver. Proepicardial cells contribute to the smooth-muscle cells of the coronary arteries and are required for their proper patterning. Other cell types within the heart, including fibroblasts and potentially some myocardial cells, also can arise from the proepicardium.

The cardiac conduction system, which functions both to generate and to propagate electrical impulses, develops primarily from multipotential cardiac precursors. The conduction system is composed of slow (proximal) components, such as the sinoatrial (SA) and atrioventricular (AV) nodes, as well as fast (distal) components, including the His bundle, bundle branches, and Purkinje fibers. The AV node primarily serves to delay the electrical impulse between atria and ventricles (manifesting decremental conduction), whereas the distal conduction system rapidly delivers the impulse throughout the ventricles. Significant recent attention has been focused on the embryologic origins of various components of the specialized conduction network. Precursors within the sinus venosus give rise to the SA node, whereas those within the AV canal mature into heterogeneous cell types that compose the AV node. Myocardial cells transdifferentiate into Purkinje fibers to form the distal conduction system. Fast and slow conducting cell types within the nodes and bundles are characterized by expression of distinct gap junction proteins, including connexins, and ion channels that characterize unique cell fates and electrical properties of the tissues. Developmental defects in conduction system morphogenesis and lineage determination can lead to various electrophysiologic disorders, including congenital heart block and preexcitation syndromes such as Wolff-Parkinson-White syndrome (Chap. 233).

Studies of cardiac stem and progenitor cells suggest that progressive lineage restriction results in the gradual and stepwise determination of mature cell fates within the heart, with early precursors capable of adopting endothelial, smooth-muscle, or cardiac phenotypes, and subsequent further specialization into atrial, ventricular, and specialized conduction cell types.

Regenerating Cardiac Tissue

Until very recently, adult mammalian myocardial cells were viewed as fully differentiated and without regenerative potential. Evidence currently supports the existence of limited endogenous regenerative potential of mature cardiac myocytes, resident cardiac progenitors, and/or bone marrow–derived stem cells. Considerable current effort is being devoted to evaluating the utility of cells from these sources to enhance the regenerative potential of the heart. The success of such approaches would offer the exciting possibility of reconstructing an infarcted or failing ventricle (Chaps. 65 and 67).

FURTHER READINGS

Colucci WS, Braunwald E (eds): *Atlas of Heart Failure: Cardiac Function and Dysfunction* , 4th ed. Philadelphia, Current Medicine, 2004

Deanfield JE et al: Endothelial function and dysfunction: Testing and clinical relevance. Circulation 115:1285, 2007[PMID: 17353456]

Katz am: *Physiology of the Heart* , 4th ed. Philadelphia, Lippincott, Williams & Wilkins, 2005

Kirby ML: *Cardiac Development.* New York, Oxford University Press, 2007

Libby P et al: The vascular endothelium and atherosclerosis, in *The Handbook of Experimental Pharmacology*, S Moncada and EA Higgs (eds). Berlin-Heidelberg, Springer-Verlag, 2006

Mahoney WM, Schwartz SM: Defining smooth muscle cells and smooth muscle cell injury. J Clin Invest 15:221, 2005

Opie LH: *Heart Physiology: From Cell to Circulation*, 4th ed. Philadelphia, Lippincott, Williams & Wilkins, 2004

———: Mechanisms of cardiac contraction and relaxation, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Elsevier, 2008

Wehrens XH et al: Intracellular calcium release and cardiac disease. Annu Rev Physiol 67:69, 2005[PMID: 15709953]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EPIDEMIOLOGY OF CARDIOVASCULAR DISEASE: INTRODUCTION

Cardiovascular disease (CVD) is now the most common cause of death worldwide. Before 1900, infectious diseases and malnutrition were the most common causes and CVD was responsible for less than 10% of all deaths. Today, CVD accounts for approximately 30% of deaths worldwide, including nearly 40% in high-income countries and about 28% in low- and middle-income countries.

THE EPIDEMIOLOGIC TRANSITION

The global rise in CVD is the result of an unprecedented transformation in the causes of morbidity and mortality during the twentieth and twenty-first centuries. Known as the epidemiologic transition, this shift is driven by industrialization, urbanization, and associated lifestyle changes and is taking place in every part of the world among all races, ethnic groups, and cultures. The transition is divided into four basic stages: pestilence and famine, receding pandemics, degenerative and human-made diseases, and delayed degenerative diseases. A fifth stage, characterized by an epidemic of inactivity and obesity, may be emerging in some countries **(Table 225-1**).

Table 225–1. Five Stages of the Epidemiologic Transition

Stage	Description	Deaths Related to CVD, %	Predominant CVD Type
Pestilence and famine	Predominance of malnutrition and infectious diseases as causes of death; high rates of infant and child mortality; low mean life expectancy	<10	Rheumatic heart disease, cardiomopathies caused by infection and malnutrition
Receding pandemics	Improvements in nutrition and public health lead to decrease in rates of deaths related to malnutrition and infection; precipitous decline in infant and child mortality rates	10–35	Rheumatic valvular disease, hypertension, CHD, and stroke (predominantly hemorrhagic)
Degenerative and human-made diseases	Increased fat and caloric intake and decrease in physical activity lead to emergence of hypertension and atherosclerosis; with increase in life expectancy, mortality from chronic, noncommunicable diseases exceeds mortality from malnutrition and infectious disease	35–65	CHD and stroke (ischemic and hemorrhagic)
Delayed degenerative diseases	CVD and cancer are the major causes of morbidity and mortality; better treatment and prevention efforts help avoid deaths among those with disease and delay primary events; age-adjusted CVD morality rate declines; CVD affecting older and older individuals	40–50	CHD, stroke, and congestive heart failure
Inactivity and obesity	Overweight and obesity increase at alarming rate; diabetes and hypertension increase; decline in smoking rates levels off; a minority of the population meets physical activity recommendations	Possible reversal of age-adjusted declines in mortality	CHD, stroke, and congestive heart failure, peripheral vascular disease

Abbreviations: CHD, coronary heart disease; CVD, cardiovascular disease. ***Source:*** Adapted from AR Omran: Milbank Mem Fund Q 49:509, 1971; and SJ Olshansky, AB Ault: Milbank Q 64:355, 1986.

Malnutrition, infectious diseases, and high infant and child mortality rates that are offset by high fertility mark the *age of pestilence and famine* . Tuberculosis, dysentery, cholera, and influenza are often fatal, resulting in a mean life expectancy of about 30 years. Cardiovascular disease, which accounts for less than 10% of deaths, takes the form of rheumatic heart disease and cardiomyopathies due to infection and malnutrition. Approximately 10% of the world's population remains in the age of pestilence and

famine.

Per capita income and life expectancy increase during the *age of receding pandemics* as the emergence of public health systems, cleaner water supplies, and improved nutrition combine to drive down deaths from infectious disease and malnutrition. Infant and childhood mortality rates also decline, but deaths due to CVD increase to between 10% and 35% of all deaths. Rheumatic valvular disease, hypertension, coronary heart disease, and stroke are the predominant forms of CVD. Almost 40% of the world's population is currently in this stage.

The *age of degenerative and human-made diseases* is distinguished by mortality from noncommunicable diseases—primarily CVD—surpassing mortality from malnutrition and infectious diseases. Caloric intake, particularly from animal fat, increases. Coronary heart disease and stroke are prevalent, and 35%–65% of all deaths can be traced to CVD. Typically, the rate of CHD deaths exceeds that of stroke by a ratio of 2:1 to 3:1. During this period, average life expectancy surpasses 50 years. Roughly 35% of the world's population falls into this category.

In the *age of delayed degenerative diseases* , CVD and cancer remain the major causes of morbidity and mortality, with CVD accounting for 40% of all deaths. However, age-adjusted CVD mortality declines, aided by preventive strategies such as smoking cessation programs and effective blood pressure control, acute hospital management, and technological advances such as the availability of bypass surgery. Coronary heart disease (CHD), stroke, and congestive heart failure are the primary forms of CVD. About 15% of the world's population is now in the age of delayed degenerative diseases or is exiting this age and moving into the fifth stage of the epidemiologic transition.

In the industrialized world, physical activity continues to decline while total caloric intake increases. The resulting epidemic of overweight and obesity may signal the start of the *age of inactivity and obesity*. Rates of Type 2 diabetes mellitus, hypertension, and lipid abnormalities are on the rise, trends that are particularly evident in children. If these risk factor trends continue, age-adjusted CVD mortality rates could increase in the coming years.

The Epidemiologic Transition in the United States

The United States, like other high-income countries, has proceeded through four stages of the epidemiologic transition. Recent trends, however, suggest that the rates of decline of some chronic and degenerative diseases have slowed. Because of the large amount of available data, the United States serves as a useful reference point for comparisons.

THE AGE OF PESTILENCE AND FAMINE (BEFORE 1900)

The American colonies were born into pestilence and famine, with half the Pilgrims who arrived in 1620 dying of infection and malnutrition by the following spring. At the end of the 1800s, the U.S. economy was still largely agrarian, with more than 60% of the population living in rural settings. By 1900, average life expectancy had increased to about 50 years. However, tuberculosis, pneumonia, and other infectious diseases still accounted for more deaths than any other cause. CVD accounted for less than 10% of all deaths.

THE AGE OF RECEDING PANDEMICS (1900–1930)

By 1900, a public health infrastructure was in place: Forty states had health departments, many larger towns had major public works efforts to improve the water supply and sewage systems, municipal use of chlorine to disinfect water was widespread, pasteurization and other improvements in food handling were introduced, and the educational quality of health care personnel improved. Those changes led to dramatic declines in infectious disease mortality rates. However, the continued shift from a rural, agriculture-based economy to an urban, industrial economy had a number of consequences on risk behaviors and factors for CVD. Owing to a lack of refrigerated transport from farms to urban centers, consumption of fresh fruits and vegetables declined and consumption of meat and grains increased, resulting in diets that were higher in animal fat and processed carbohydrates. In addition, the availability of factory-rolled cigarettes made tobacco more accessible and affordable for the mass population. Age-adjusted CVD mortality rates rose from 300 per 100,000 people in 1900 to approximately 390 per 100,000 during this period, driven by rapidly rising CHD rates.

THE AGE OF DEGENERATIVE AND HUMAN-MADE DISEASES (1930–1965)

During this period, deaths from infectious diseases fell to fewer than 50 per 100,000 per year and life expectancy increased to almost 70 years. At the same time, the country became increasingly urbanized and industrialized, precipitating a number of important lifestyle changes. By 1955, 55% of adult men were smoking, and fat consumption accounted for approximately 40% of total calories. Lower activity levels, high-fat diets, and increased smoking pushed CVD death rates to peak levels.

THE AGE OF DELAYED DEGENERATIVE DISEASES (1965–2000)

Substantial declines in age-adjusted CVD mortality rates began in the mid-1960s. In the 1970s and 1980s, age-adjusted CHD mortality rates fell approximately 2% per year and stroke rates fell 3% per year. A main characteristic of this phase is the steadily rising age at which a first CVD event occurs. Two significant advances have been credited with the decline in CVD mortality rates: new therapeutic

approaches and the implementation of prevention measures. Treatments once considered advanced, such as angioplasty, bypass surgery, and implantation of defibrillators, are now considered the standard of care. Treatments for hypertension and elevated cholesterol along with the widespread use of aspirin have also contributed significantly to reducing deaths from CVD. In addition, Americans have been exposed to public health campaigns promoting lifestyle modifications effective at reducing the prevalence of smoking, hypertension, and dyslipidemia.

IS THE UNITED STATES ENTERING THE FIFTH AGE?

The decline in the age-adjusted CVD death rate of 3% per year through the 1970s and 1980s tapered off in the 1990s to 2%. However, CVD death rates declined by 3–5% per year during the first decade of the new millennium. In 2000, the age-adjusted CVD death rate was 341 per 100,000. By 2006, it had fallen to 263 per 100,000. Competing trends appear to be in play. On the one hand, the well-recognized increase in the prevalence of diabetes and obesity, a slowing in the rate of decline of smoking, and a leveling off in the rate of detection and treatment for hypertension are in the negative column. On the other hand, cholesterol levels continue to decline in the face of increased statin use.

Current Worldwide Variations

An epidemiologic transition much like that which occurred in the United States is occurring throughout the world, but unique regional features have modified aspects of the transition in various parts of the world. In terms of economic development, the world can be divided into two broad categories, (1) high-income countries and (2) low- and middle-income countries, which can be further subdivided into six distinct economic/geographic regions. Currently, 85% of the world's population lives in low- and middle-income countries, and it is those countries which are driving the rates of change in the global burden of CVD (**Fig. 225-1**). Three million CVD deaths occurred in high-income countries in 2001, compared with 13 million in the rest of the world.

Figure 225-1

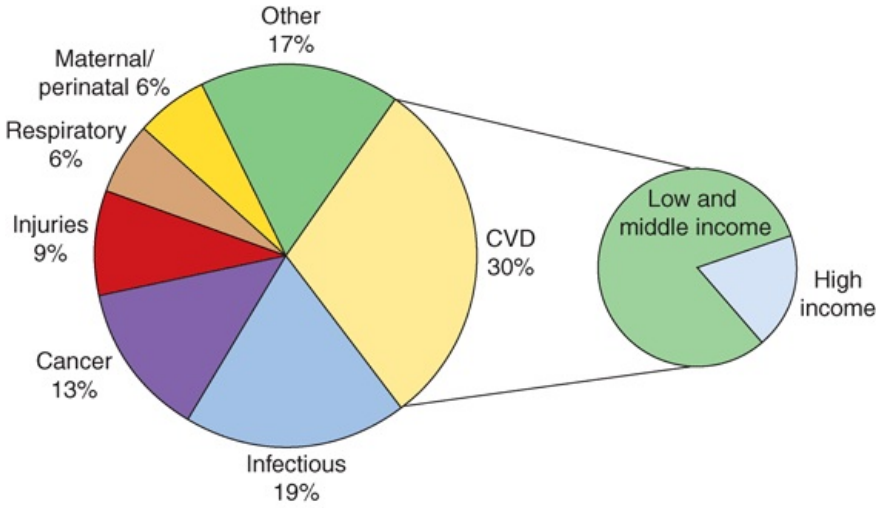

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CVD data compared with other causes of death. CVD: cardiovascular disease. *(Based on data from CD Mathers et al: Deaths and Disease Burden by Cause: Global Burden of Disease Estimates for 2001 byWorld Bank Country Groups. Disease Control Priorities Working Paper 18. April 2004, revised January 2005.)*

HIGH-INCOME COUNTRIES

Approximately 940 million people live in high-income countries, where CHD is the dominant form of CVD, with rates that tend to be twofold to fivefold higher than stroke rates. The rates of CVD in Canada, New Zealand, Australia, and Western Europe tend to be similar to those in the United States; however, among the countries of Western Europe, the absolute rates vary threefold with a clear north/south gradient. The highest CVD death rates are in the northern countries, such as Finland, Ireland, and Scotland, and the lowest rates are in the Mediterranean countries of France, Spain, and Italy. Japan is unique among the high-income countries: stroke rates increased dramatically, but CHD rates did not rise as sharply over the last century. This difference may stem in part from genetic factors, but it is more likely that the fish- and plant-based low-fat diet and resulting low cholesterol levels have played a larger role. Importantly, Japanese dietary habits are undergoing substantial changes, reflected in an increase in cholesterol levels.

LOW- AND MIDDLE-INCOME COUNTRIES

The World Bank groups the low- and middle-income countries (gross national income per capita less than US $9200) into six geographic regions: East Asia and the Pacific, (Eastern) Europe and Central Asia, Latin America and the Caribbean, Middle East and North Africa, South Asia, and Sub-Saharan Africa. Although communicable diseases continue to be a major cause of death, CVD has emerged as a significant health concern in low- and middle-income countries. In most, an urban/rural gradient has emerged for CHD, stroke, and hypertension, with higher rates in urban centers.

Although CVD rates are rapidly rising, there are vast differences among the regions and countries and even within individual countries (**Fig. 225-2**). Many factors contribute to this heterogeneity. First, the regions are in various stages of the epidemiologic transition. Second, vast differences in lifestyle and behavioral risk factors exist. Third, racial and ethnic differences may lead to altered susceptibilities to various forms of CVD. In addition, it should be noted that for most countries in these regions, accurate countrywide data on cause-specific mortality are not complete, as death certificate completion is not routine and most of those countries do not have a centralized registry for deaths.

Figure 225-2

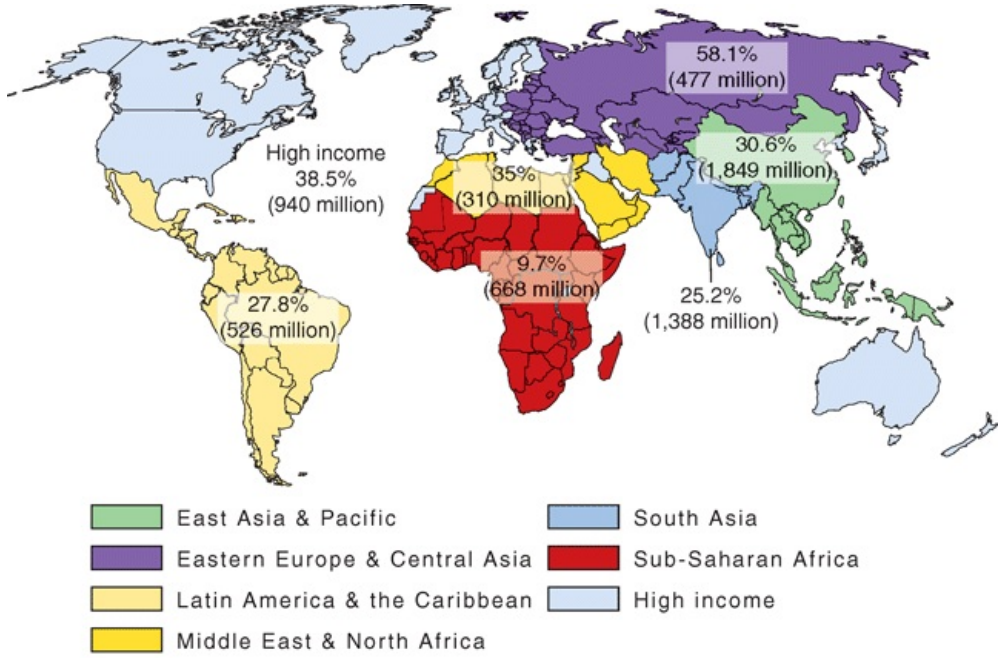

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CVD death as a percentage of total deaths and total population in seven economic regions of the world defined by the World Bank *(Based on data from CD Mathers et al: Deaths and Disease Burden by Cause: Global Burden of Disease Estimates for 2001 by World Bank Country Groups. Disease Control Priorities Working Paper 18. April 2004, revised January 2005.)*

The *East Asia and Pacific* region, home to nearly 2 billion people, appears to be straddling the second and third phases of the epidemiologic transition, with China, Indonesia, and Sri Lanka's large combined population driving most of the trends. Overall, CVD is a major cause of death in China, but as in Japan, stroke (particularly hemorrhagic) causes more deaths than does CHD, in a ratio of about 3:1. However, age-adjusted CHD mortality increased 40% from 1984 to 1999, suggesting further epidemiologic transition. China also appears to have a geographic gradient like that of Western Europe, with higher CVD rates in northern China than in southern China by a factor of 6. Other countries, such as Vietnam and Cambodia, are just emerging from the pestilence and famine era.

The *Eastern Europe and Central Asia* region is firmly at the peak of the third phase, with the highest death rates (58%) due to CVD in the world, nearly double the rate of high-income countries. More troubling is that nearly 35% of deaths from CHD occur among working-age adults, which is three times the rate in the United States. In Russia, increased CVD rates have contributed to falling life expectancy, particularly for men, whose life expectancy dropped from 71.6 in 1986 to 59 years today. In Poland, by contrast, the age-adjusted mortality rate decreased by approximately 30% for men during the 1990s and slightly more among women. Slovenia, Hungary, the Czech Republic, and Slovakia have had similar declines.

In general, *Latin America* appears to be in the third phase of the epidemiologic transition, although as in other low- and middle-income regions, there is vast regional heterogeneity, with some areas in the second phase of the transition and some in the fourth. Today, approximately 28% of all deaths in this region are attributable to CVD, with CHD rates (35%) higher than stroke rates (29%). As in Eastern

Europe, some countries—Mexico, Costa Rica, and Venezuela—continued an overall increase in age-adjusted CHD mortality of 3%–10% between 1970 and 2002, whereas in others—Argentina, Brazil, Chile, and Columbia—rates appear to have declined by as much as 2% per year over the same period. The *Middle East and North Africa* region appears to be entering the third phase of the epidemiologic transition, with increasing life expectancy overall and CVD death rates just below those of developed nations. CHD is responsible for 17% of all deaths, and stroke for 7%. The traditional high-fiber diet, low in fat and cholesterol, has changed rapidly. Over the last few decades, daily fat consumption has increased in most of these countries, ranging from a 13.6% increase in Sudan to a 143.3% increase in Saudi Arabia. Over 75% of Egyptians are overweight or obese, and the rate is 67% in Iraq and Jordan. Nearly 60% of Syrians of physical activity and Iraqis report that they are physically inactive (less than 10 min per day).

Most people in *South Asia* live in rural India, a country that is experiencing an alarming increase in heart disease. CVD accounted for 32% of all deaths in 2000, and an estimated 2 million deaths were expected to occur due to CHD by 2010, representing a 30% increase over the preceding decade. The transition appears to be in the Western style, with CHD as the dominant form of CVD. In 1960, CHD represented 4% of all CVD deaths in India, whereas in 1990 the proportion was >50%. This is somewhat unexpected because stroke tends to be a more dominant factor early in the epidemiologic transition. This finding may reflect inaccuracies in cause-specific mortality estimates or possibly an underlying genetic component. It has been suggested that Indians have exaggerated insulin insensitivity in response to the Western lifestyle pattern that may differentially increase rates of CHD over stroke. The South Asia region has the highest overall prevalence of diabetes in the low-income regions, with rates as high as 14% in urban centers. In certain rural areas, the prevalence of CVD and its risk factors is approaching urban rates. Nonetheless, rheumatic heart disease continues to be a major cause of morbidity and mortality.

For the most part, *sub-Saharan Africa* remains in the first phase of the epidemiologic transition, with CVD rates half those in developed nations. Life expectancy has decreased by an average of 5 years since the early 1990s largely because of HIV/AIDS and other chronic diseases, according to the World Bank; life expectancies are the lowest in the world. Still, CVD accounts for 46% of noncommunicable deaths and is the leading cause of death among adults >age 35. As more HIV/AIDS patients receive antiretroviral treatment, managing CVD risk factors such as dyslipidemia in this population requires more attention. However, hypertension continues to be the major public health concern and has resulted in stroke being the dominant form of CVD. Rheumatic heart disease is still an important cause of CVD mortality and morbidity.

Global Trends in Cardiovascular Disease

In 1990, CVD accounted for 28% of the world's 50.4 million deaths and 9.7% of the 1.4 billion lost disability-adjusted life years (DALYs), and by 2001, CVD was responsible for 29% of all deaths and 14% of the 1.5 billion lost DALYs. By 2030, when the population is expected to reach 8.2 billion, 33% of all deaths will be the result of CVD (**Table 225-2**). Of these, 14.9% of deaths in men and 13.1% of deaths in women will be due to CHD. Stroke will be responsible for 10.4% of all male deaths and 11.8% of all female deaths.

Table 225–2. Estimated Morbidity Related to Heart Disease: 2010-2030

Deaths	By 2010	By 2030
CVD deaths: annual number of all deaths	18.1 million	24.2 million
CVD deaths: percentage of all deaths	30.8%	32.5%
CHD deaths: percentage of all male deaths	13.1%	14.9%
CHD deaths: percentage of all female deaths	13.6%	13.1%
Stroke deaths: percentage of all male deaths	9.2%	10.4%
Stroke deaths: percentage of all female deaths	11.5%	11.8%

Abbreviations: CVD, cardiovascular disease; CHD, coronary heart disease.

Source: Adapted from J Mackay, G Mensah: *Atlas of Heart Disease and Stroke* . Geneva, World Health Organization, 2004.

In the *high-income countries*, population growth will be fueled by emigration from the low- and middle-income countries, but the populations of high-income countries will shrink as a proportion of the world's population. The modest decline in CVD death rates that began in the high-income countries in the latter third of the twentieth century will continue, but the rate of decline appears to be slowing. However, these countries are expected to see an increase in the prevalence of CVD, as well as the absolute number of deaths as the population ages.

Significant proportions of the population living in *low- and middle-income countries* have entered the third phase of the epidemiologic transition, and some are entering the fourth stage. Changing

demographics play a significant role in future predictions for CVD throughout the world. For example, between 1990 and 2001, the population of Eastern Europe and Central Asia grew by 1 million people per year, whereas South Asia added 25 million people each year.

CVD rates will also have an economic impact. Even assuming no increase in CVD risk factors, most countries, but especially India and South Africa, will see a large number of people between 35 and 64 die of CVD over the next 30 years as well as an increasing level of morbidity among middle-aged people related to heart disease and stroke. In China, it is estimated that there will be 9 million deaths from CVD in 2030—up from 2.4 million in 2002—with half occurring in individuals between 35 and 64 years old.

Regional Trends in Risk Factors

As indicated earlier, the global variation in CVD rates is related to temporal and regional variations in known risk behaviors and factors. Ecological analyses of major CVD risk factors and mortality demonstrate high correlations between expected and observed mortality rates for the three main risk factors—smoking, serum cholesterol, and hypertension—and suggest that many of the regional variations are based on differences in conventional risk factors.

Behavioral Risk Factors

TOBACCO

Every year, more than 5.5 trillion cigarettes are produced, enough to provide every person on the planet with 1,000 cigarettes. Worldwide, 1.3 billion people smoked in 2003, a number that is projected to increase to 1.6 billion by 2030. Tobacco currently causes about 5 million deaths—9% of all deaths—annually. Approximately 1.6 million are CVD-related. If current smoking patterns continue, by 2030 the global burden of disease attributable to tobacco will reach 10 million deaths annually. A unique feature of the low- and middle-income countries is easy access to smoking during the early stages of the epidemiologic transition due to the availability of relatively inexpensive tobacco products. In South Asia, the prominence of locally produced forms of tobacco other than manufactured cigarettes makes control of consumption more challenging.

DIET

Total caloric intake per capita increases as countries develop. With regard to cardiovascular disease, a key element of dietary change is an increase in intake of saturated animal fats and hydrogenated vegetable fats, which contain atherogenic trans-fatty acids, along with a decrease in intake of plant-based foods and an increase in simple carbohydrates. Fat contributes less than 20% of calories in rural China and India, less than 30% in Japan, and well above 30% in the United States. Caloric contributions from fat appear to be falling in the high-income countries. In the United States, between 1971 and 2000, the percentage of calories derived from saturated fat decreased from 13% to 11%.

PHYSICAL INACTIVITY

The increased mechanization that accompanies the economic transition leads to a shift from physically demanding agriculture-based work to largely sedentary industry- and office-based work. In the United States, approximately one-quarter of the population does not participate in any leisure-time physical activity and only 22% report engaging in sustained physical activity for at least 30 minutes on 5 or more days per week (the current recommendation). In contrast, in countries such as China, physical activity is still integral to everyday life. Approximately 90% of the urban population walks or rides a bicycle to work, shopping, or school daily.

Metabolic Risk Factors

LIPID LEVELS

Worldwide, high cholesterol levels are estimated to cause 56% of ischemic heart disease and 18% of strokes, amounting to 4.4 million deaths annually. As countries move through the epidemiologic transition, mean population plasma cholesterol levels tend to rise. Social and individual changes that accompany urbanization clearly play a role because plasma cholesterol levels tend to be higher among urban residents than among rural residents. This shift is driven largely by greater consumption of dietary fats—primarily from animal products and processed vegetable oils—and decreased physical activity. In the high-income countries, in general, mean population cholesterol levels are falling, whereas wide variation is seen in the low- and middle-income countries.

HYPERTENSION

Elevated blood pressure is an early indicator of the epidemiologic transition. Worldwide, approximately 62% of strokes and 49% of cases of ischemic heart disease are attributable to suboptimal (>115 mmHg systolic) blood pressure, which is believed to account for more than 7 million deaths annually. Remarkably, nearly half of this burden occurs among those with systolic blood pressure <140 mmHg, even as this level is used at the arbitrary threshold for defining hypertension in many national guidelines. Rising mean population blood pressure is apparent as populations industrialize and move from rural to urban settings. Among urban-dwelling men and women in India, for example, the prevalence of hypertension is 25.5% and 29.0%, respectively, whereas it is 14.0% and 10.8%,

respectively, in rural communities. One major concern in low- and middle-income countries is the high rate of undetected, and therefore untreated, hypertension. This may explain, at least in part, the higher stroke rates in these countries in relation to CHD rates during the early stages of the transition. The high rates of hypertension, especially undiagnosed hypertension, throughout Asia probably contribute to the high prevalence of hemorrhagic stroke in the region.

OBESITY

Although clearly associated with increased risk of CHD, much of the risk posed by obesity may be mediated by other CVD risk factors, including hypertension, diabetes mellitus, and lipid profile imbalances. In the mid-1980s, the World Health Organization's MONICA Project sampled 48 populations for cardiovascular risk factors. In all but one male population (China) and in most of the female populations, between 50% and 75% of adults age 35–64 years were overweight or obese. In addition, the prevalence of extreme obesity (BMI >40 kg/m^2) more than tripled, increasing from 1.3% to 4.9%. In many of the low- and middle-income countries, obesity appears to coexist with undernutrition and malnutrition. Obesity is increasing throughout the world, particularly in developing countries, where the trajectories are steeper than those experienced in the developed countries. According to the latest World Health Organization (WHO) data, this is equivalent to about 1.3 billion overweight adults in the world. A survey undertaken in 1998 found that as many as 58% of African women living in South Africa might have been overweight or obese.

DIABETES MELLITUS

As a consequence of, or in addition to, increasing body mass index and decreasing levels of physical activity, worldwide rates of diabetes—predominantly Type 2 diabetes—are on the rise. In 2003, 194 million adults, or 5% of the world's population, had diabetes. By 2025, this number is predicted to increase 72 percent to 333 million. By 2025, the number of people with Type 2 diabetes is projected to double in three of the six low- and middle-income regions: Middle East and North Africa, South Asia, and Sub-Saharan Africa. There appear to be clear genetic susceptibilities to diabetes mellitus in various racial and ethnic groups. For example, migration studies suggest that South Asians and Indians tend to be at higher risk than are people of European ancestry.

Summary

Although CVD rates are declining in the high-income countries, they are increasing in virtually every other region of the world. The consequences of this preventable epidemic will be substantial on many levels: individual mortality and morbidity rates, family suffering, and staggering economic costs.

Three complementary strategies can be used to lessen the impact. First, the overall burden of CVD risk factors can be lowered through populationwide public health measures such as national campaigns against cigarette smoking, unhealthy diets, and physical inactivity. Second, it is important to identify higher-risk subgroups of the population that stand to benefit the most from specific, low-cost prevention interventions, including screening for and treatment of hypertension and elevated cholesterol. Simple, low-cost interventions, such as the "polypill," a regimen of aspirin, a statin, and an anithypertensive agent, also need to be explored. Third, resources should be allocated to acute as well as secondary prevention interventions. For countries with limited resources, a critical first step in developing a comprehensive plan is better assessment of cause-specific mortality and morbidity, as well as the prevalence of the major preventable risk factors.

In the meantime, the high-income countries must continue to bear the burden of research and development aimed at prevention and treatment, being mindful of the economic limitations of many countries. The concept of the epidemiologic transition provides insight into methods to alter the course of the CVD epidemic. The efficient transfer of low-cost preventive and therapeutic strategies could alter the natural course of this epidemic and thereby reduce the excess global burden of preventable CVD.

FURTHER READINGS

Gaziano T, Gaziano JM: Global burden of cardiovascular disease, in *Heart Disease: A Textbook of Cardiovascular Medicine*, 9th ed, E Braunwald (ed). Philadelphia, Elsevier Saunders, 2009

Jamison DT et al (eds): *Disease Control Priorities in Developing Countries*, 2nd ed. Washington, DC, Oxford University Press, 2006

Lawes CM et al: Global burden of blood-pressure-related disease, 2001. Lancet 371:1513, 2008[PMID: 18456100]

Lopez AD et al (eds): *Global Burden of Disease and Risk Factors.* Washington, DC, Oxford University Press, 2006

Shafey O et al: *The Tobacco Atlas,* 3rd ed. Atlanta: American Cancer Society, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE MAGNITUDE OF THE PROBLEM

Cardiovascular diseases comprise the most prevalent serious disorders in industrialized nations and are a rapidly growing problem in developing nations (Chap. 225). Age-adjusted death rates for coronary heart disease have declined by two-thirds in the last 4 decades in the United States, reflecting the identification and reduction of risk factors as well as improved treatments and interventions for the management of coronary artery disease, arrhythmias, and heart failure. Nonetheless, cardiovascular diseases remain the most common causes of death, responsible for 35% of all deaths, almost 1 million deaths each year. Approximately one-fourth of these deaths are sudden. In addition, cardiovascular diseases are highly prevalent, diagnosed in 80 million adults, or ~35% of the adult population. The growing prevalence of obesity (Chap. 77), type 2 diabetes mellitus (Chap. 344), and metabolic syndrome (Chap. 242), which are important risk factors for atherosclerosis, now threatens to reverse the progress that has been made in the age-adjusted reduction in the mortality rate of coronary heart disease.

For many years cardiovascular disease was considered to be more common in men than in women. In fact, the percentage of all deaths secondary to cardiovascular disease is higher among women (43%) than among men (37%) (Chap. 6). In addition, although the absolute number of deaths secondary to cardiovascular disease has declined over the past decades in men, this number has actually risen in women. Inflammation, obesity, type 2 diabetes mellitus, and the metabolic syndrome appear to play more prominent roles in the development of coronary atherosclerosis in women than in men. Coronary artery disease (CAD) is more frequently associated with dysfunction of the coronary microcirculation in women than in men. Exercise electrocardiography has a lower diagnostic accuracy in the prediction of epicardial obstruction in women than in men.

CARDIAC SYMPTOMS

The symptoms caused by heart disease result most commonly from myocardial ischemia, disturbance of the contraction and/or relaxation of the myocardium, obstruction to blood flow, or an abnormal cardiac rhythm or rate. Ischemia, which is caused by an imbalance between the heart's oxygen supply and demand, is manifest most frequently as chest discomfort (Chap. 12), whereas reduction of the pumping ability of the heart commonly leads to fatigue and elevated intravascular pressure upstream of the failing ventricle. The latter results in abnormal fluid accumulation, with peripheral edema (Chap. 36) or pulmonary congestion and dyspnea (Chap. 33). Obstruction to blood flow, as occurs in valvular stenosis, can cause symptoms resembling those of myocardial failure (Chap. 234). Cardiac arrhythmias often develop suddenly, and the resulting symptoms and signs—palpitations (Chap. 37), dyspnea, hypotension, and syncope (Chap. 20)—generally occur abruptly and may disappear as rapidly as they develop.

Although dyspnea, chest discomfort, edema, and syncope are cardinal manifestations of cardiac disease, they occur in other conditions as well. Thus, dyspnea is observed in disorders as diverse as pulmonary disease, marked obesity, and anxiety (Chap. 33). Similarly, chest discomfort may result from a variety of noncardiac and cardiac causes other than myocardial ischemia (Chap. 12). Edema, an important finding in untreated or inadequately treated heart failure, also may occur with primary renal disease and in hepatic cirrhosis (Chap. 36). Syncope occurs not only with serious cardiac arrhythmias but in a number of neurologic conditions as well (Chap. 20). Whether heart disease is responsible for these symptoms frequently can be determined by carrying out a careful clinical examination (Chap. 227), supplemented by noninvasive testing using electrocardiography at rest and during exercise (Chap. 228), echocardiography, roentgenography, and other forms of myocardial imaging (Chap. 229).

Myocardial or coronary function that may be adequate at rest may be insufficient during exertion. Thus, dyspnea and/or chest discomfort that appear during activity are characteristic of patients with heart disease, whereas the opposite pattern, i.e., the appearance of these symptoms at rest and their remission during exertion, is rarely observed in such patients. It is important, therefore, to question the patient carefully about the relation of symptoms to exertion.

Many patients with cardiovascular disease may be asymptomatic both at rest and during exertion but may present with an abnormal physical finding such as a heart murmur, elevated arterial pressure, or an abnormality of the electrocardiogram (ECG) or the cardiac silhouette on the chest roentgenogram or other imaging test. It is important to assess the global risk of CAD in asymptomatic individuals, using a combination of clinical assessment and measurement of cholesterol and its fractions, as well as

other biomarkers, such as C-reactive protein, in some patients (Chap. 241). Since the first clinical manifestation of CAD may be catastrophic—sudden cardiac death, acute myocardial infarction, or stroke in previous asymptomatic persons—it is mandatory to identify those at high risk of such events and institute further testing and preventive measures.

Diagnosis

As outlined by the New York Heart Association (NYHA), the elements of a complete cardiac diagnosis include the systematic consideration of the following:

1. *The underlying etiology*. Is the disease congenital, hypertensive, ischemic, or inflammatory in origin?
2. *The anatomical abnormalities*. Which chambers are involved? Are they hypertrophied, dilated, or both? Which valves are affected? Are they regurgitant and/or stenotic? Is there pericardial involvement? Has there been a myocardial infarction?
3. *The physiological disturbances*. Is an arrhythmia present? Is there evidence of congestive heart failure or myocardial ischemia?
4. *Functional disability*. How strenuous is the physical activity required to elicit symptoms? The classification provided by the NYHA has been found to be useful in describing functional disability **(Table 226-1)**.

Table 226–1. New York Heart Association Functional Classification

Class I	Class III
No limitation of physical activity	Marked limitation of physical activity
No symptoms with ordinary exertion	Less than ordinary activity causes symptoms
Class II	Asymptomatic at rest
Slight limitation of physical activity	Class IV
Ordinary activity causes symptoms	Inability to carry out any physical activity without discomfort Symptoms at rest

Source: Modified from The Criteria Committee of the New York Heart Association.

One example may serve to illustrate the importance of establishing a complete diagnosis. In a patient who presents with exertional chest discomfort, the identification of myocardial ischemia as the etiology is of great clinical importance. However, the simple recognition of ischemia is insufficient to formulate a therapeutic strategy or prognosis until the underlying anatomical abnormalities responsible for the myocardial ischemia, e.g., coronary atherosclerosis or aortic stenosis, are identified and a judgment is made about whether other physiologic disturbances that cause an imbalance between myocardial oxygen supply and demand, such as severe anemia, thyrotoxicosis, or supraventricular tachycardia, play contributory roles. Finally, the severity of the disability should govern the extent and tempo of the workup and strongly influence the therapeutic strategy that is selected.

The establishment of a correct and complete cardiac diagnosis usually commences with the history and physical examination (Chap. 227). Indeed, the clinical examination remains the basis for the diagnosis of a wide variety of disorders. The clinical examination may then be supplemented by five types of laboratory tests: (1) ECG (Chap. 228), (2) noninvasive imaging examinations (chest roentgenogram, echocardiogram, radionuclide imaging, computed tomographic imaging, and magnetic resonance imaging (Chap. 229), (3) blood tests to assess risk [e.g., lipid determinations, C-reactive protein (Chap. 241)] or cardiac function [e.g., brain natriuretic peptide (BNP) (Chap. 234)], (4) occasionally specialized invasive examinations [i.e., cardiac catheterization and coronary arteriography (Chap. 230)], and (5) genetic tests to identify monogenic cardiac diseases [e.g., hypertrophic cardiomyopathy (Chap. 238), Marfan syndrome (Chap. 363), and abnormalities of cardiac ion channels that lead to prolongation of the QT interval and an increase in the risk of sudden death (Chap. 233)]. These tests are becoming more widely available.

Family History

In eliciting the history of a patient with known or suspected cardiovascular disease, particular attention should be directed to the family history. Familial clustering is common in many forms of heart disease. Mendelian transmission of single-gene defects may occur, as in hypertrophic cardiomyopathy (Chap. 238), Marfan syndrome (Chap. 363), and sudden death associated with a prolonged QT syndrome (Chap. 233). Premature coronary disease and essential hypertension, type 2 diabetes mellitus, and hyperlipidemia (the most important risk factors for coronary artery disease) are usually polygenic disorders. Although familial transmission may be less obvious than in the monogenic disorders, it is helpful in assessing risk and prognosis in polygenic disorders. Familial clustering of cardiovascular diseases not only may occur on a genetic basis but also may be related to familial dietary or behavior

patterns such as excessive ingestion of salt or calories and cigarette smoking.

Assessment of Functional Impairment

When an attempt is made to determine the severity of functional impairment in a patient with heart disease, it is helpful to ascertain the level of activity and the rate at which it is performed before symptoms develop. Thus, it is not sufficient to state that the patient complains of dyspnea. The breathlessness that occurs after running up two long flights of stairs denotes far less functional impairment than do similar symptoms that occur after taking a few steps on level ground. Also, the degree of customary physical activity at work and during recreation should be considered. The development of two-flight dyspnea in a well-conditioned marathon runner may be far more significant than the development of one-flight dyspnea in a previously sedentary person. The history should include a detailed consideration of the patient's therapeutic regimen. For example, the persistence or development of edema, breathlessness, and other manifestations of heart failure in a patient who is receiving optimal doses of diuretics and other therapies for heart failure (Chap. 234) is far graver than are similar manifestations in the absence of treatment. Similarly, the presence of angina pectoris despite treatment with optimal doses of multiple antianginal drugs (Chap. 243) is more serious than it is in a patient on no therapy. In an effort to determine the progression of symptoms, and thus the severity of the underlying illness, it may be useful to ascertain what, if any, specific tasks the patient could have carried out 6 months or 1 year earlier that he or she cannot carry out at present.

Electrocardiogram

(See also Chap. 228) Although an ECG usually should be recorded in patients with known or suspected heart disease, with the exception of the identification of arrhythmias, conduction abnormalities, ventricular hypertrophy, and acute myocardial infarction, it generally does not establish a specific diagnosis. The range of normal electrocardiographic findings is wide, and the tracing can be affected significantly by many noncardiac factors, such as age, body habitus, and serum electrolyte concentrations. In general, electrocardiographic changes should be interpreted in the context of other abnormal cardiovascular findings.

Assessment of the Patient with a Heart Murmur

(Fig. 226-1) The cause of a heart murmur can often be readily elucidated from a systematic evaluation of its major attributes: timing, duration, intensity, quality, frequency, configuration, location, and radiation when considered in the light of the history, general physical examination, and other features of the cardiac examination, as described in Chap. 227.

Figure 226-1

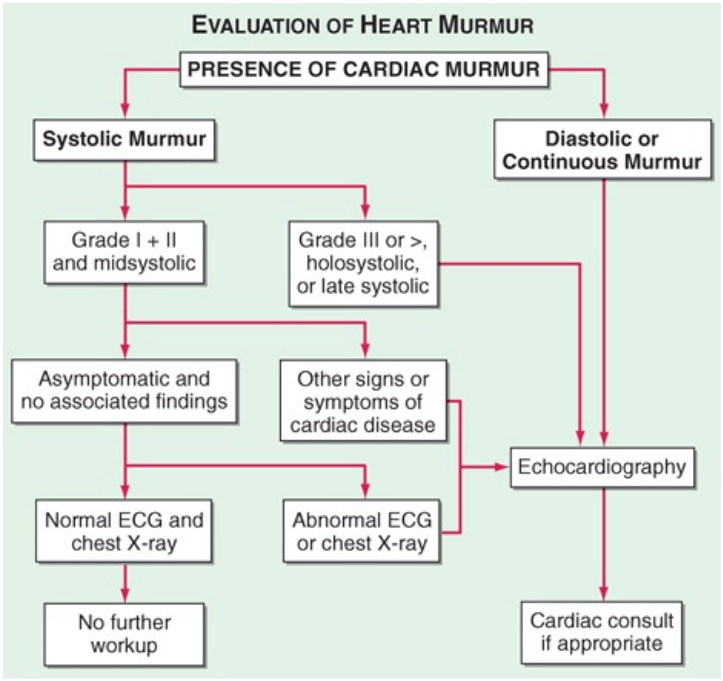

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An alternative "echocardiography first" approach to the evaluation of a heart murmur that also uses the results of the electrocardiogram (ECG) and chest x-ray in asymptomatic patients with soft midsystolic murmurs and no other physical findings. This algorithm is useful for patients over age 40 years in whom the prevalence of coronary artery disease and aortic stenosis increases as the

cause of systolic murmur. *[From RA O'Rourke, in Primary Cardiology, 2d ed, E Braunwald, L Goldman (eds). Philadelphia, Saunders, 2003.]*

The majority of heart murmurs are midsystolic and soft (grades I–II/VI). When such a murmur occurs in an asymptomatic child or young adult *without* other evidence of heart disease on clinical examination, it is usually benign and echocardiography generally is not required. By contrast, two-dimensional and Doppler echocardiography (Chap. 229) are indicated in patients with loud systolic murmurs (grades ≥ III/VI), especially those that are holosystolic or late systolic, and in most patients with diastolic or continuous murmurs.

Natural History

Cardiovascular disorders often present acutely, as in a previously asymptomatic person who develops an acute myocardial infarction (Chap. 245), or a previously asymptomatic patient with hypertrophic cardiomyopathy (Chap. 238), or with a prolonged QT interval (Chap. 233) whose first clinical manifestation is syncope or even sudden death. However, the alert physician may recognize the patient at risk for these complications long before they occur and often can take measures to prevent their occurrence. For example, a patient with acute myocardial infarction will often have had risk factors for atherosclerosis for many years. Had these risk factors been recognized, their elimination or reduction might have delayed or even prevented the infarction. Similarly, a patient with hypertrophic cardiomyopathy may have had a heart murmur for years and a family history of this disorder. These findings could have led to an echocardiographic examination, recognition of the condition, and appropriate therapy long before the occurrence of a serious acute manifestation.

Patients with valvular heart disease or idiopathic dilated cardiomyopathy, by contrast, may have a prolonged course of gradually increasing dyspnea and other manifestations of chronic heart failure that is punctuated by episodes of acute deterioration only late in the course of the disease. Understanding the natural history of various cardiac disorders is essential for applying appropriate diagnostic and therapeutic measures to each stage of the condition, as well as for providing the patient and family with the likely prognosis.

PITFALLS IN CARDIOVASCULAR MEDICINE

Increasing subspecialization in internal medicine and the perfection of advanced diagnostic techniques in cardiology can lead to several undesirable consequences. Examples include the following:

1. Failure by the *noncardiologist* to recognize important cardiac manifestations of systemic illnesses. For example, the presence of mitral stenosis, patent foramen ovale, and/or transient atrial arrhythmia should be considered in a patient with stroke, or the presence of pulmonary hypertension and cor pulmonale should be considered in a patient with scleroderma or Raynaud's syndrome. A cardiovascular examination should be carried out to identify and estimate the severity of the cardiovascular involvement that accompanies many noncardiac disorders.
2. Failure by the *cardiologist* to recognize underlying systemic disorders in patients with heart disease. For example, hyperthyroidism should be considered in an elderly patient with atrial fibrillation and unexplained heart failure, and Lyme disease should be considered in a patient with an unexplained fluctuating atrioventricular block. A cardiovascular abnormality may provide the clue critical to the recognition of some systemic disorders. For instance, an unexplained pericardial effusion may provide an early clue to the diagnosis of tuberculosis or a neoplasm.
3. Overreliance on and overutilization of laboratory tests, particularly invasive techniques, for the evaluation of the cardiovascular system. Cardiac catheterization and coronary arteriography (Chap. 230) provide precise diagnostic information that may be crucial in developing a therapeutic plan in patients with known or suspected CAD. Although a great deal of attention has been directed to these examinations, it is important to recognize that they serve to *supplement*, not *supplant*, a careful examination carried out with clinical and noninvasive techniques. A coronary arteriogram should not be performed in lieu of a careful history in patients with chest pain suspected of having ischemic heart disease. Although coronary arteriography may establish whether the coronary arteries are obstructed and to what extent, the results of the procedure by themselves often do not provide a definitive answer to the question of whether a patient's complaint of chest discomfort is attributable to coronary atherosclerosis and whether or not revascularization is indicated.

Despite the value of invasive tests in certain circumstances, they entail some small risk to the patient, involve discomfort and substantial cost, and place a strain on medical facilities. Therefore, they should be carried out only if the results can be expected to modify the patient's management.

Disease Prevention and Management

The prevention of heart disease, especially of CAD, is one of the most important tasks of primary care health givers as well as cardiologists. Prevention begins with risk assessment, followed by attention to lifestyle, such as achieving optimal weight, physical activity, smoking cessation, and then aggressive treatment of all abnormal risk factors, such as hypertension, hyperlipidemia, and diabetes mellitus (Chap. 344).

After a complete diagnosis has been established in patients with known heart disease, a number of management options are usually available. Several examples may be used to demonstrate some of the principles of cardiovascular therapeutics:

1. In the absence of evidence of heart disease, the patient should be clearly informed of this assessment and *not* be asked to return at intervals for repeated examinations. If there is no evidence of disease, such continued attention may lead to the patient's developing inappropriate concern about the possibility of heart disease.
2. If there is no evidence of cardiovascular disease but the patient has one or more risk factors for the development of ischemic heart disease (Chap. 243), a plan for their reduction should be developed and the patient should be retested at intervals to assess compliance and efficacy in risk reduction.
3. Asymptomatic or mildly symptomatic patients with valvular heart disease that is anatomically severe should be evaluated periodically, every 6 to 12 months, by clinical and noninvasive examinations. Early signs of deterioration of ventricular function may signify the need for surgical treatment before the development of disabling symptoms, irreversible myocardial damage, and excessive risk of surgical treatment (Chap. 237).
4. In patients with CAD (Chap. 243), available practice guidelines should be considered in the decision on the form of treatment (medical, percutaneous coronary intervention, or surgical revascularization). Mechanical revascularization may be employed too frequently in the United States and too infrequently in Eastern Europe and developing nations. The mere presence of angina pectoris and/or the demonstration of critical coronary arterial narrowing at angiography should not reflexively evoke a decision to treat the patient by revascularization. Instead, these interventions should be limited to patients with CAD whose angina has not responded adequately to medical treatment or in whom revascularization has been shown to improve the natural history (e.g., acute coronary syndrome or multivessel CAD with left ventricular dysfunction).

ACKNOWLEDGMENT

Dr. Eugene Braunwald authored this chapter for the previous edition Some of the material from the 17th edition has been carried forward.

FURTHER READINGS

Abrams J: *Synopsis of Cardiac Physical Diagnosis* , 2d ed. Oxford, Butterworth Heinemann, 2001

American Heart Association: *Heart Disease and Stroke Statistics 2009 Update*. Dallas, TX, American Heart Association, www.americanheart.org

The Criteria Committee of the New York Heart Association: *Nomenclature and Criteria for Diagnosis* , 9th ed. Boston, Little, Brown, 1994

Morrow DA et al: Chronic coronary artery disease, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

National Heart, Lung and Blood Institute: *FY 2005 Fact Book* . Bethesda, MD, National Heart, Lung and Blood Institute, 2006

The Nhlbi-Sponsored Women's Ischemia Syndrome Evaluation (WISE): Challenging existing paradigms in ischemic heart disease. J Am Coll Cardiol 47(Suppl S):1, 2006

Vanden Belt J: The history, in *Classic Teachings in Clinical Cardiology: A Tribute to W. Proctor Harvey*, M Chizner (ed). Cedar Grove, NJ, Laennec, 1996, pp 41–54

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PHYSICAL EXAMINATION OF THE CARDIOVASCULAR SYSTEM: INTRODUCTION

The approach to a patient with known or suspected cardiovascular disease begins with the time-honored traditions of a directed history and a targeted physical examination. The scope of these activities depends on the clinical context at the time of presentation, ranging from an elective ambulatory follow-up visit to a more focused emergency department encounter. There has been a gradual decline in physical examination skills over the last two decades at every level, from student to faculty specialist, a development of great concern to both clinicians and medical educators. Classic cardiac findings are recognized by only a minority of internal medicine and family practice residents. Despite popular perceptions, clinical performance does not improve predictably as a function of experience; instead, the acquisition of new examination skills may become more difficult for a busy individual practitioner. Less time is now devoted to mentored cardiovascular examinations during the training of students and residents. One widely recognized outcome of these trends is the progressive overutilization of noninvasive imaging studies to establish the presence and severity of cardiovascular disease even when the examination findings imply a low pretest probability of significant pathology. Educational techniques to improve bedside skills include repetition, patient-centered teaching conferences, and visual display feedback of auscultatory events with Doppler echocardiographic imaging.

The evidence base that links the findings from the history and physical examination to the presence, severity, and prognosis of cardiovascular disease has been established most rigorously for coronary artery disease, heart failure, and valvular heart disease. For example, observations regarding heart rate, blood pressure, signs of pulmonary congestion, and the presence of mitral regurgitation (MR) contribute importantly to bedside risk assessment in patients with acute coronary syndromes. Observations from the physical examination in this setting can inform clinical decision making before the results of cardiac biomarkers testing are known. The prognosis of patients with systolic heart failure can be predicted on the basis of the jugular venous pressure (JVP) and the presence or absence of a third heart sound (S_3). Accurate characterization of cardiac murmurs provides important insight into the natural history of many valvular and congenital heart lesions. Finally, the important role played by the physical examination in enhancing the clinician-patient relationship cannot be overestimated.

The General Physical Examination

Any examination begins with an assessment of the general appearance of the patient, with notation of age, posture, demeanor, and overall health status. Is the patient in pain or resting quietly, dyspneic or diaphoretic? Does the patient choose to avoid certain body positions to reduce or eliminate pain, as might be the case with suspected acute pericarditis? Are there clues indicating that dyspnea may have a pulmonary cause, such as a barrel chest deformity with an increased anterior-posterior diameter, tachypnea, and pursed-lip breathing? Skin pallor, cyanosis, and jaundice can be appreciated readily and provide additional clues. A chronically ill-appearing emaciated patient may suggest the presence of long-standing heart failure or another systemic disorder, such as a malignancy. Various genetic syndromes, often with cardiovascular involvement, can also be recognized easily, such as trisomy 21, Marfan syndrome, and Holt-Oram syndrome. Height and weight should be measured routinely, and both body mass index and body surface area should be calculated. Knowledge of the waist circumference and the waist-to-hip ratio can be used to predict long-term cardiovascular risk. Mental status, level of alertness, and mood should be assessed continuously during the interview and examination.

SKIN

Central cyanosis occurs with significant right-to-left shunting at the level of the heart or lungs, allowing deoxygenated blood to reach the systemic circulation. Peripheral cyanosis or acrocyanosis, in contrast, is usually related to reduced extremity blood flow due to small vessel constriction, as seen in patients with severe heart failure, shock, or peripheral vascular disease; it can be aggravated by the use of β-adrenergic blockers with unopposed α-mediated constriction. Differential cyanosis refers to isolated cyanosis affecting the lower but not the upper extremities in a patient with a large patent ductus arteriosus (PDA) and secondary pulmonary hypertension with right-to-left to shunting at the great vessel level. Hereditary telangiectasias on the lips, tongue, and mucous membranes, as part of the Osler-Weber-Rendu syndrome (hereditary hemorrhagic telangiectasia), resemble spider nevi and can be a source of right-to-left shunting when also present in the lung. Malar telangiectasias also are

seen in patients with advanced mitral stenosis and scleroderma. An unusually tan or bronze discoloration of the skin may suggest hemochromatosis as the cause of the associated systolic heart failure. Jaundice, which may be visible first in the sclerae, has a broad differential diagnosis but in the appropriate setting can be consistent with advanced right heart failure and congestive hepatomegaly or late-term "cardiac cirrhosis." Cutaneous ecchymoses are seen frequently among patients taking vitamin K antagonists or antiplatelet agents such as aspirin and thienopyridines. Various lipid disorders sometimes are associated with subcutaneous xanthomas, particularly along the tendon sheaths or over the extensor surfaces of the extremities. Severe hypertriglyceridemia can be associated with eruptive xanthomatosis and lipemia retinalis. Palmar crease xanthomas are specific for type III hyperlipoproteinemia. Pseudoxanthoma elasticum, a disease associated with premature atherosclerosis, is manifested by a leathery, cobblestoned appearance of the skin in the axilla and neck creases and by angioid streaks on funduscopic examination. Extensive lentiginoses have been described in a variety of development delay–cardiovascular syndromes, including Carney syndrome, which includes multiple atrial myxomas. Cutaneous manifestations of sarcoidosis such as lupus pernio and erythema nodosum may suggest this disease as a cause of an associated dilated cardiomyopathy, especially with heart block, intraventricular conduction delay, or ventricular tachycardia.

HEAD AND NECK

Dentition and oral hygiene should be assessed in every patient both as a source of potential infection and as an index of general health. A high-arched palate is a feature of Marfan syndrome and other connective tissue disease syndromes. Bifid uvula has been described in patients with Loeys-Dietz syndrome, and orange tonsils are characteristic of Tangier disease. The ocular manifestations of hyperthyroidism have been well described. Many patients with congenital heart disease have associated hypertelorism, low-set ears, or micrognathia. Blue sclerae are a feature of osteogenesis imperfecta. An arcus senilis pattern lacks specificity as an index of coronary heart disease risk. The funduscopic examination is an often underutilized method by which to assess the microvasculature, especially among patients with established atherosclerosis, hypertension, or diabetes mellitus. A mydriatic agent may be necessary for optimal visualization. A funduscopic examination should be performed routinely in the assessment of patients with suspected endocarditis and those with a history of acute visual change. Branch retinal artery occlusion or visualization of a Hollenhorst plaque can narrow the differential diagnosis rapidly in the appropriate setting. Relapsing polychondritis may manifest as an inflamed pinna or, in its later stages, as a saddle-nose deformity because of destruction of nasal cartilage; Wegener's granulomatosis can also lead to a saddle-nose deformity.

CHEST

Midline sternotomy, left posterolateral thoracotomy, or infraclavicular scars at the site of pacemaker/defibrillator generator implantation should not be overlooked and may provide the first clue regarding an underlying cardiovascular disorder in patients unable to provide a relevant history. A prominent venous collateral pattern may suggest subclavian or vena caval obstruction. If the head and neck appear dusky and slightly cyanotic and the venous pressure is grossly elevated without visible pulsations, a diagnosis of superior vena cava syndrome should be entertained. Thoracic cage abnormalities have been well described among patients with connective tissue disease syndromes. They include pectus carinatum ("pigeon chest") and pectus excavatum ("funnel chest"). Obstructive lung disease is suggested by a barrel chest deformity, especially with tachypnea, pursed-lip breathing, and use of accessory muscles. The characteristically severe kyphosis and compensatory lumbar, pelvic, and knee flexion of ankylosing spondylitis should prompt careful auscultation for a murmur of aortic regurgitation (AR). Straight back syndrome refers to the loss of the normal kyphosis of the thoracic spine and has been described in patients with mitral valve prolapse (MVP) and its variants. In some patients with cyanotic congenital heart disease, the chest wall appears to be asymmetric, with anterior displacement of the left hemithorax. The respiratory rate and pattern should be noted during spontaneous breathing, with additional attention to depth, audible wheezing, and stridor. Lung examination can reveal adventitious sounds indicative of pulmonary edema, pneumonia, or pleuritis.

ABDOMEN

In some patients with advanced obstructive lung disease, the point of maximal cardiac impulse may be in the epigastrium. The liver is frequently enlarged and tender in patients with chronic heart failure. Systolic pulsations over the liver signify severe tricuspid regurgitation (TR). Splenomegaly may be a feature of infective endocarditis, particularly when symptoms have persisted for weeks or months. Ascites is a nonspecific finding but may be present with advanced chronic right heart failure, constrictive pericarditis, hepatic cirrhosis, or an intraperitonealmalignancy. The finding of an elevated JVP implies a cardiovascular etiology. In nonobese patients, the aorta typically is palpated between the epigastrium and the umbilicus. The sensitivity of palpation for the detection of an abdominal aortic aneurysm (pulsatile and expansile mass) decreases as a function of body size. Because palpation alone is not sufficiently accurate to establish this diagnosis, a screening ultrasound examination is advised. The presence of an arterial bruit over the abdomen suggests high-grade atherosclerotic disease, though precise localization is difficult.

EXTREMITIES

The temperature and color of the extremities, the presence of clubbing, arachnodactyly, and pertinent nail findings can be surmised quickly during the examination. Clubbing implies the presence of central right-to-left shunting, although it has also been described in patients with endocarditis. Its appearance can range from cyanosis and softening of the root of the nail bed, to the classic loss of the normal angle between the base of the nail and the skin, to the skeletal and periosteal bony changes of hypertrophic osteoarthropathy, which is seen rarely in patients with advanced lung or liver disease. Patients with the Holt-Oram syndrome have an unopposable, "fingerized" thumb, whereas patients with Marfan syndrome may have arachnodactyly and a positive "wrist" (overlapping of the thumb and fifth finger around the wrist) or "thumb" (protrusion of the thumb beyond the ulnar aspect of the hand when the fingers are clenched over the thumb in a fist) sign. The Janeway lesions of endocarditis are nontender, slightly raised hemorrhages on the palms and soles, whereas Osler's nodes are tender, raised nodules on the pads of the fingers or toes. Splinter hemorrhages are classically identified as linear petechiae in the midposition of the nail bed and should be distinguished from the more common traumatic petechiae, which are seen closer to the distal edge.

Lower extremity or presacral edema in the setting of an elevated JVP defines volume overload and may be a feature of chronic heart failure or constrictive pericarditis. Lower extremity edema in the absence of jugular venous hypertension may be due to lymphatic or venous obstruction or, more commonly, to venous insufficiency, as further suggested by the appearance of varicosities, venous ulcers (typically medial in location), and brownish cutaneous discoloration from hemosiderin deposition (eburnation). Pitting edema can also be seen in patients who use dihydropyridine calcium channel blockers. A Homan's sign (posterior calf pain on active dorsiflexion of the foot against resistance) is neither specific nor sensitive for deep venous thrombosis. Muscular atrophy or the absence of hair along an extremity is consistent with severe arterial insufficiency or a primary neuromuscular disorder.

Cardiovascular Examination

JUGULAR VENOUS PRESSURE AND WAVE FORM

Jugular venous pressure is the single most important bedside measurement from which to estimate the volume status. The internal jugular vein is preferred because the external jugular vein is valved and not directly in line with the superior vena cava and right atrium. Nevertheless, the external jugular vein has been used to discriminate between high and low central venous pressure (CVP) when tested among medical students, residents, and attending physicians. Precise estimation of the central venous or right atrial pressure from bedside assessment of the jugular venous waveform has proved difficult. Venous pressure traditionally has been measured as the vertical distance between the top of the jugular venous pulsation and the sternal inflection point (angle of Louis). A distance >4.5 cm at 30° elevation is considered abnormal. However, the actual distance between the mid-right atrium and the angle of Louis varies considerably as a function of both body size and the patient angle at which the assessment is made (30°, 45°, or 60°). The use of the sternal angle as a reference point leads to systematic underestimation of CVP, and this method should be used less for semiquantification than to distinguish a normal from an abnormally elevated CVP. The use of the clavicle may provide an easier reference for standardization. Venous pulsations above this level in the sitting position are clearly abnormal, as the distance between the clavicle and the right atrium is at least 10 cm. The patient should always be placed in the sitting position, with the legs dangling below the bedside, when an elevated pressure is suspected in the semisupine position. It should also be noted that bedside estimates of CVP are made in centimeters of water but must be converted to millimeters of mercury to provide correlation with accepted hemodynamic norms (1.36 cmH_2O = 1.0 mmHg).

The venous waveform sometimes can be difficult to distinguish from the carotid pulse, especially during casual inspection. Nevertheless, the venous waveform has several characteristic features, and its individual components can be appreciated in most patients **(Fig. 227-1)**. In patients in sinus rhythm, the venous waveform is typically biphasic, whereas the carotid upstroke is monophasic.

Figure 227-1

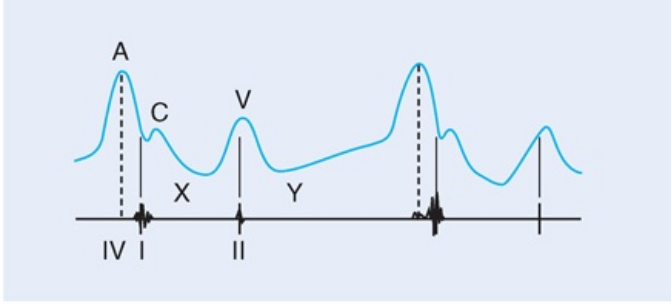

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

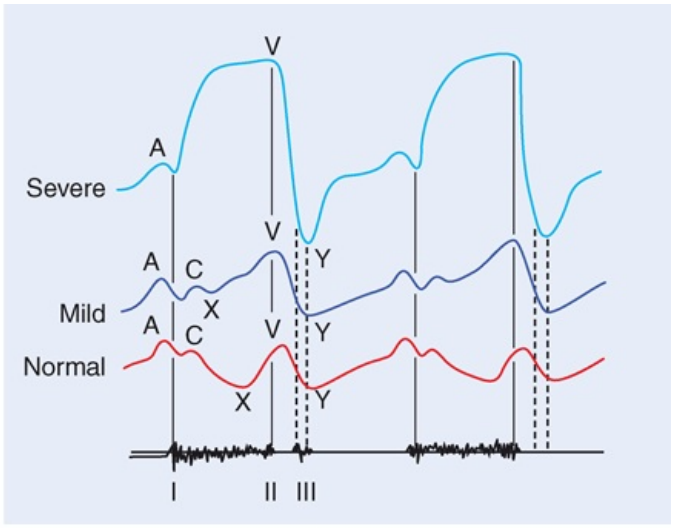

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

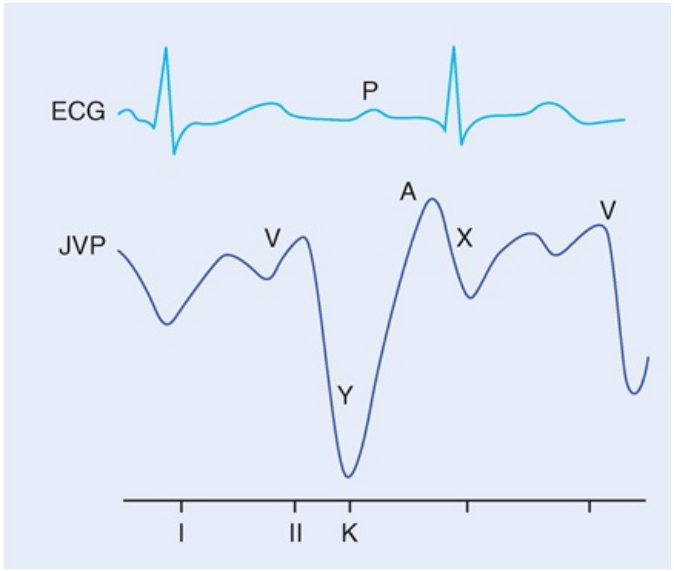

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A*. Jugular venous pulse wave tracing** (*top*) with heart sounds (*bottom*). The A wave represents right atrial presystolic contraction and occurs just after the electrocardiographic P wave and just before the first heart sound (I). In this example, the A wave is accentuated and larger than normal due to decreased right ventricular compliance, as also suggested by the right-sided S_4 (IV). The C wave may reflect the carotid pulsation in the neck and/or an early systolic increase in right atrial pressure as the right ventricle pushes the closed tricuspid valve into the right atrium. The *x* descent follows the A wave just as atrial pressure continues to fall. The V wave represents atrial filling during ventricular systole and peaks at the second heart sound (II). The *y* descent corresponds to the fall in right atrial pressure after tricuspid valve opening. ***B. Jugular venous wave forms in mild (*middle*) and severe (*top*) tricuspid regurgitation, compared with normal, with phonocardiographic representation of the corresponding heart sounds below. With increasing degrees of tricuspid regurgitation, the waveform becomes "ventricularized." ***C.*** ECG (*top*), jugular venous waveform (*middle*), and heart sounds (*bottom*) in pericardial constriction. Note the prominent and rapid *y* descent, corresponding in timing to the pericardial knock (K). *(From J Abrams: Synopsis of Cardiac Physical Diagnosis, 2nd ed. Boston, Butterworth Heinemann, 2001, pp 25–35.)*

The venous waveform is divided into several distinct peaks. The *a* wave reflects right atrial presystolic contraction and occurs just after the electrocardiographic P wave, preceding the first heart sound (S_1). A prominent *a* wave is seen in patients with reduced right ventricular compliance; a cannon *a* wave occurs with atrioventricular (AV) dissociation and right atrial contraction against a closed tricuspid valve. In a patient with a wide complex tachycardia, the appreciation of cannon *a* waves in the jugular venous waveform identifies the rhythm as ventricular in origin. The *a* wave is not present with atrial fibrillation. The *x* descent defines the fall in right atrial pressure after inscription of the *a* wave. The *c* wave interrupts this *x* descent and is followed by a further descent. The *v* wave represents atrial filling (atrial diastole) and occurs during ventricular systole. The height of the *v* wave is determined by right

atrial compliance as well as the volume of blood returning to the right atrium either antegrade from the cavae or retrograde through an incompetent tricuspid valve. In patients with TR, the *v* wave is accentuated and the subsequent fall in pressure (*y* descent) is rapid. With progressive degrees of TR, the *v* wave merges with the *c* wave, and the right atrial and jugular vein waveforms become "ventricularized." The *y* descent, which follows the peak of the *v* wave, can become prolonged or blunted with obstruction to right ventricular inflow, as may occur with tricuspid stenosis (TS) or pericardial tamponade. Normally, the venous pressure should fall by at least 3 mmHg with inspiration. Kussmaul's sign is defined by either a rise or a lack of fall of the JVP with inspiration and is classically associated with constrictive pericarditis, although it has been reported in patients with restrictive cardiomyopathy, massive pulmonary embolism, right ventricular infarction, and advanced left ventricular systolic heart failure.

Venous hypertension sometimes can be elicited by performance of the abdominojugular reflex or with passive leg elevation. When these signs are positive, a volume-overloaded state with limited compliance of an overly distended or constricted venous system is present. The abdominojugular reflex is elicited with firm and consistent pressure over the upper portion of the abdomen, preferably over the right upper quadrant, for at least 10 s. A positive response is defined by a sustained rise of more than 3 cm in JVP for at least 15 s after release of the hand. Patients must be coached to refrain from breath holding or a Valsalva-like maneuver during the procedure. The abdominojugular reflex is useful in predicting a pulmonary artery wedge pressure in excess of 15 mmHg in patients with heart failure.

Although the JVP estimates right ventricular filling pressure, it has a predictable relationship with the pulmonary artery wedge pressure. In a large study of patients with advanced heart failure, the presence of a right atrial pressure >10 mm Hg (as predicted on bedside examination) had a positive value of 88% for the prediction of a pulmonary artery wedge pressure of >22 mmHg. In addition, an elevated JVP has prognostic significance in patients with both symptomatic heart failure and asymptomatic left ventricular systolic dysfunction. The presence of an elevated JVP is associated with a higher risk of subsequent hospitalization for heart failure, death from heart failure, or both.

ASSESSMENT OF BLOOD PRESSURE

Measurement of blood pressure usually is delegated to a medical assistant but should be repeated by the clinician. Accurate measurement depends on body position, arm size, time of measurement, place of measurement, device, device size, technique, and examiner. In general, physician-recorded blood pressures are higher than nurse-recorded pressures. Blood pressure is best measured in the seated position with the arm at the level of the heart, using an appropriately sized cuff, after 5–10 min of relaxation. When it is measured in the supine position, the arm should be raised to bring it to the level of the mid-right atrium. The length and width of the blood pressure cuff bladder should be 80% and 40% of the arm's circumference, respectively. A common source of error in practice is to use an inappropriately small cuff, resulting in marked overestimation of true blood pressure, or an inappropriately large cuff, resulting in underestimation of true blood pressure. The cuff should be inflated to 30 mmHg above the expected systolic pressure and the pressure released at a rate of 2–3 mmHg/s. Systolic and diastolic pressures are defined by the first and fifth Korotkoff sounds, respectively. Very low (even 0 mmHg) diastolic blood pressures may be recorded in patients with chronic, severe AR or a large arteriovenous fistula because of enhanced diastolic "run-off." In these instances, both the phase IV and phase V Korotkoff sounds should be recorded. Blood pressure is best assessed at the brachial artery level, though it can be measured at the radial, popliteal, or pedal pulse level. In general, systolic pressure increases and diastolic pressure decreases when measured in more distal arteries. Blood pressure should be measured in both arms, and the difference should be less than 10 mmHg. A blood pressure differential that exceeds this threshold may be associated with atherosclerotic or inflammatory subclavian artery disease, supravalvular aortic stenosis, aortic coarctation, or aortic dissection. Systolic leg pressures are usually as much as 20 mmHg higher than systolic arm pressures. Greater leg–arm pressure differences are seen in patients with chronic severe AR as well as patients with extensive and calcified lower extremity peripheral arterial disease. The ankle-brachial index (lower pressure in the dorsalis pedis or posterior tibial artery divided by the higher of the two brachial artery pressures) is a powerful predictor of long-term cardiovascular mortality.

The blood pressure measured in an office or hospital setting may not accurately reflect the pressure in other venues. "White coat hypertension" is defined by at least three separate clinic-based measurements >140/90 mmHg and at least two non-clinic-based measurements <140/90 mmHg in the absence of any evidence of target organ damage. Individuals with white coat hypertension may not benefit from drug therapy, although they may be more likely to develop sustained hypertension over time. Masked hypertension should be suspected when normal or even low blood pressures are recorded in patients with advanced atherosclerotic disease, especially when evidence of target organ damage is present or bruits are audible.

Orthostatic hypotension is defined by a fall in systolic pressure >20 mmHg or in diastolic pressure >10 mmHg in response to assumption of the upright posture from a supine position within 3 min. There may also be a lack of a compensatory tachycardia, an abnormal response that suggests autonomic

insufficiency, as may be seen in patients with diabetes or Parkinson's disease. Orthostatic hypotension is a common cause of postural lightheadedness/syncope and should be assessed routinely in patients for whom this diagnosis might pertain. It can be exacerbated by advanced age, dehydration, certain medications, food, deconditioning, and ambient temperature.

ARTERIAL PULSE

The carotid artery pulse occurs just after the ascending aortic pulse. The aortic pulse is best appreciated in the epigastrium, just above the level of the umbilicus. Peripheral arterial pulses that should be assessed routinely include the subclavian, brachial, radial, ulnar, femoral, popliteal, dorsalis pedis, and posterior tibial. In patients in whom the diagnosis of either temporal arteritis or polymyalgia rheumatica is suspected, the temporal arteries also should be examined. Although one of the two pedal pulses may not be palpable in up to 10% of normal subjects, the pair should be symmetric. The integrity of the arcuate system of the hand is assessed by Allen's test, which is performed routinely before instrumentation of the radial artery. The pulses should be examined for their symmetry, volume, timing, contour, amplitude, and duration. If necessary, simultaneous auscultation of the heart can help identify a delay in the arrival of an arterial pulse. Simultaneous palpation of the radial and femoral pulses may reveal a femoral delay in a patient with hypertension and suspected aortic coarctation. The carotid upstrokes should never be examined simultaneously or before listening for a bruit. Light pressure should always be used to avoid precipitation of carotid hypersensitivity syndrome and syncope in a susceptible elderly individual. The arterial pulse usually becomes more rapid and spiking as a function of its distance from the heart, a phenomenon that reflects the muscular status of the more peripheral arteries and the summation of the incident and reflected waves. In general, the character and contour of the arterial pulse depend on the stroke volume, ejection velocity, vascular compliance, and systemic vascular resistance. The pulse examination can be misleading in patients with reduced cardiac output and in those with stiffened arteries from aging, chronic hypertension, or peripheral arterial disease.

The character of the pulse is best appreciated at the carotid level **(Fig. 227-2)**. A weak and delayed pulse (*pulsus parvus et tardus*) defines severe aortic stenosis (AS). Some patients with AS may also have a slow, notched, or interrupted upstroke (anacrotic pulse) with a thrill or shudder. With chronic severe AR, by contrast, the carotid upstroke has a sharp rise and rapid fall-off (Corrigan's or water-hammer pulse). Some patients with advanced AR may have a bifid or bisferiens pulse, in which two systolic peaks can be appreciated. A bifid pulse is also described in patients with hypertrophic obstructive cardiomyopathy (HOCM), with inscription of percussion and tidal waves. A bifid pulse is easily appreciated in patients on intra-aortic balloon counterpulsation (IABP), in whom the second pulse is diastolic in timing.

Figure 227-2

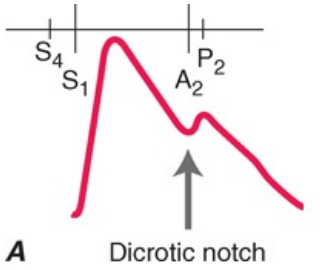

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

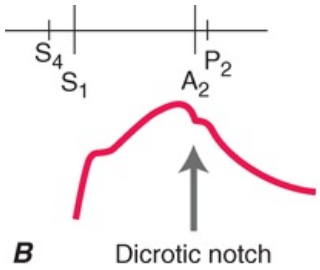

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

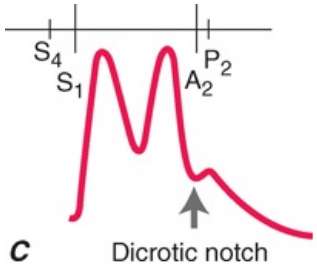

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

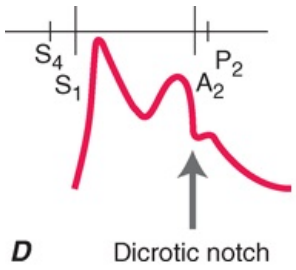

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

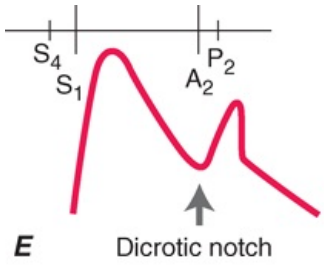

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagrams of the configurational changes in carotid pulse and their differential diagnoses. Heart sounds are also illustrated. ***A***. Normal. S_4, fourth heart sound; S_1, first heart sound; A_2 aortic component of second heart sound; P_2 pulmonic component of second heart sound. ***B***. Aortic stenosis. Anacrotic pulse with slow upstroke to a reduced peak. ***C.*** Bisferiens pulse with two peaks in systole. This pulse is rarely appreciated in patients with severe aortic regurgitation. ***D***. Bisferiens pulse in hypertrophic obstructive cardiomyopathy. There is a rapid upstroke to the first peak (percussion wave) and a slower rise to the second peak (tidal wave). ***E***. Dicrotic pulse with peaks in systole and diastole. This waveform may be seen in patients with sepsis or during intra-aortic balloon counterpulsation with inflation just after the dicrotic notch. *(From K Chatterjee, W Parmley [eds]: Cardiology: An Illustrated Text/Reference. Philadelphia, JB Lippincott, 1991).*

Pulsus paradoxus refers to a fall in systolic pressure >10 mmHg with inspiration that is seen in patients with pericardial tamponade but also is described in those with massive pulmonary embolism, hemorrhagic shock, severe obstructive lung disease, and tension pneumothorax. Pulsus paradoxus is measured by noting the difference between the systolic pressure at which the Korotkoff sounds are first heard (during expiration) and the systolic pressure at which the Korotkoff sounds are heard with each heartbeat, independent of the respiratory phase. Between these two pressures, the Korotkoff sounds are heard only intermittently and during expiration. The cuff pressure must be decreased slowly to appreciate the finding. It can be difficult to measure pulsus paradoxus in patients with tachycardia, atrial fibrillation, or tachypnea. A pulsus paradoxus may be palpable at the brachial artery or femoral artery level when the pressure difference exceeds 15 mmHg. This inspiratory fall in systolic pressure is an exaggerated consequence of interventricular dependence.

Pulsus alternans, in contrast, is defined by beat-to-beat variability of pulse amplitude. It is present only when every other phase I Korotkoff sound is audible as the cuff pressure is lowered slowly, typically in a patient with a regular heart rhythm and independent of the respiratory cycle. Pulsus alternans is seen in patients with severe left ventricular systolic heart failure and is thought to be due to cyclic changes in intracellular calcium and action potential duration. Interestingly, when pulsus alternans is associated with electrocardiographic T-wave alternans, the risk for an arrhythmic event appears to be increased.

Ascending aortic aneurysms can rarely be appreciated as a pulsatile mass in the right parasternal.

Appreciation of a prominent abdominal aortic pulse should prompt noninvasive imaging for better characterization. Femoral and/or popliteal artery aneurysms should be sought in patients with abdominal aortic aneurysm disease.

The level of a claudication-producing arterial obstruction can often be identified on physical examination **(Fig. 227-3)**. For example, in a patient with calf claudication, a decrease in pulse amplitude between the common femoral and popliteal arteries will localize the obstruction to the level of the superficial femoral artery, although inflow obstruction above the level of the common femoral artery may coexist. Auscultation for carotid, subclavian, abdominal aortic, and femoral artery bruits should be routine. However, the correlation between the presence of a bruit and the degree of vascular obstruction is poor. A cervical bruit is a weak indicator of the degree of carotid artery stenosis; the absence of a bruit does not exclude the presence of significant luminal obstruction. If a bruit extends into diastole or if a thrill is present, the obstruction is generally severe. Other causes of arterial bruits include an arteriovenous fistula with enhanced flow.

Figure 227-3.

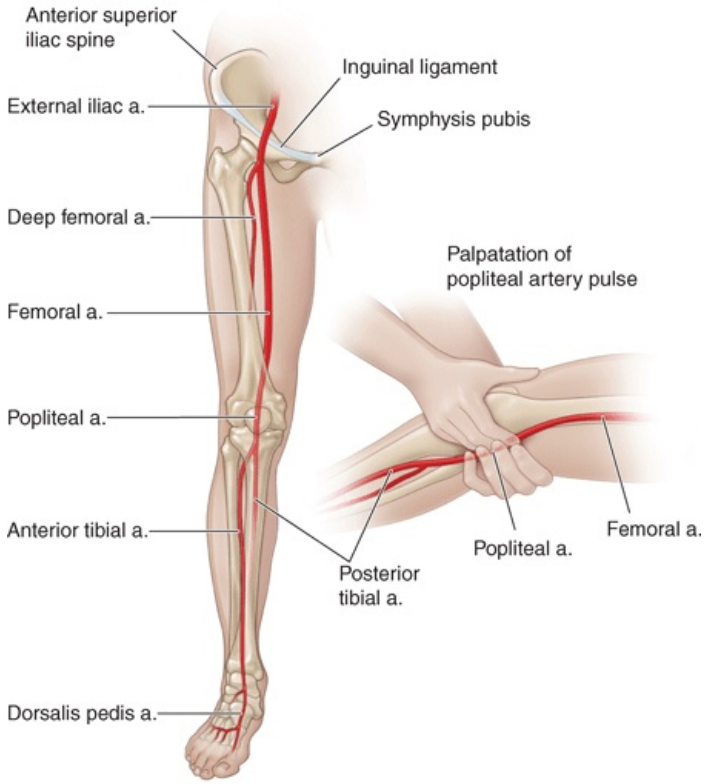

A **Major arteries of the lower limb**

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

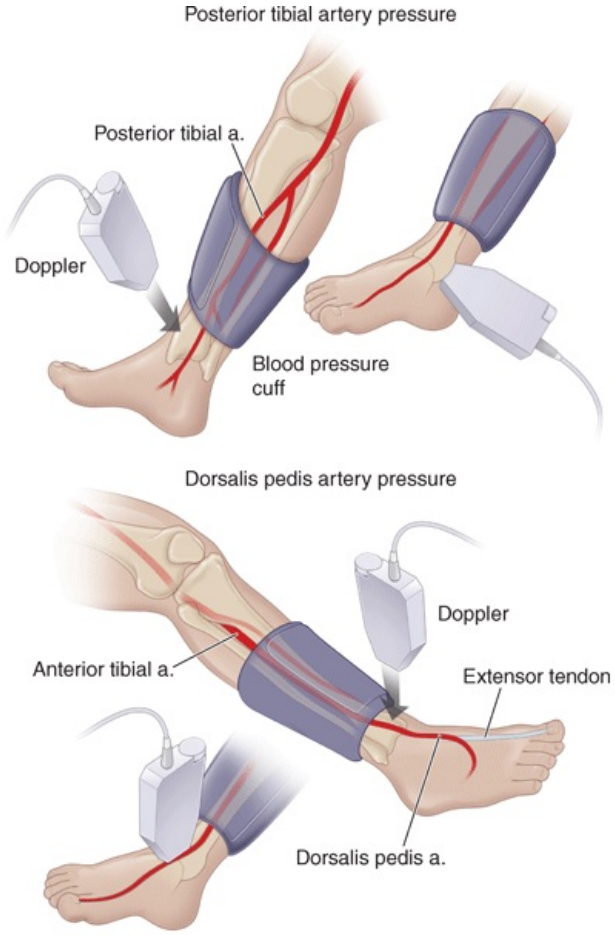

A. Anatomy of the major arteries of the leg. ***B.*** Measurement of the ankle systolic pressure. *(From NA Khan et al: JAMA 295:536–546, 2006.)*

The likelihood of significant lower extremity peripheral arterial disease increases with typical symptoms of claudication, cool skin, abnormalities on pulse examination, or the presence of a vascular bruit. Abnormal pulse oximetry (a >2% difference between finger and toe oxygen saturation) can be used to detect lower extremity peripheral arterial disease and is comparable in its performance characteristics to the ankle-brachial index.

INSPECTION AND PALPATION OF THE HEART

The left ventricular apex beat may be visible in the midclavicular line at the fifth intercostal space in thin-chested adults. Visible pulsations anywhere other than this expected location are abnormal. The left anterior chest wall may heave in patients with an enlarged or hyperdynamic left or right ventricle. As noted previously, a visible right upper parasternal pulsation may be suggestive of ascending aortic aneurysm disease. In thin, tall patients and patients with advanced obstructive lung disease and flattened diaphragms, the cardiac impulse may be visible in the epigastrium and should be distinguished from a pulsatile liver edge.

Palpation of the heart begins with the patient in the supine position at 30° and can be enhanced by placing the patient in the left lateral decubitus position. The normal left ventricular impulse is less than 2 cm in diameter and moves quickly away from the fingers; it is better appreciated at end expiration, with the heart closer to the anterior chest wall. Characteristics such as size, amplitude, and rate of force development should be noted.

Enlargement of the left ventricular cavity is manifested by a leftward and downward displacement of an enlarged apex beat. A sustained apex beat is a sign of pressure overload, such as that which may be present in patients with AS or chronic hypertension. A palpable presystolic impulse corresponds to the fourth heart sound (S_4) and is indicative of reduced left ventricular compliance and the forceful contribution of atrial contraction to ventricular filling. A palpable third sound (S_3), which is indicative of a rapid early filling wave in patients with heart failure, may be present even when the gallop itself is not audible. A large left ventricular aneurysm may sometimes be palpable as an ectopic impulse, discrete from the apex beat. Hypertrophic obstructive cardiomyopathy may very rarely cause a triple cadence

beat at the apex with contributions from a palpable S_4 and the two components of the bisferiens systolic pulse.

Right ventricular pressure or volume overload may create a sternal lift. Signs of either TR (*cv* waves in the jugular venous pulse) and/or pulmonary arterial hypertension (a loud single or palpable P_2) would be confirmatory. The right ventricle can enlarge to the extent that left-sided events cannot be appreciated. A zone of retraction between the right and left ventricular impulses sometimes can be appreciated in patients with right ventricle pressure or volume overload when they are placed in the left lateral decubitus position. Systolic and diastolic thrills signify turbulent and high-velocity blood flow. Their locations help identify the origin of heart murmurs.

Cardiac Auscultation

HEART SOUNDS

Ventricular systole is defined by the interval between the first (S_1) and second (S_2) heart sounds **(Fig. 227-4)**. The first heart sound (S_1) includes mitral and tricuspid valve closure. Normal splitting can be appreciated in young patients and those with right bundle branch block, in whom tricuspid valve closure is relatively delayed. The intensity of S_1 is determined by the distance over which the anterior leaflet of the mitral valve must travel to return to its annular plane, leaflet mobility, left ventricular contractility, and the PR interval. S_1 is classically loud in the early phases of rheumatic mitral stenosis (MS) and in patients with hyperkinetic circulatory states or short PR intervals. S_1 becomes softer in the later stages of MS when the leaflets are rigid and calcified, after exposure to ß-adrenergic receptor blockers, with long PR intervals, and with left ventricular contractile dysfunction. The intensity of any heart sound, however, can be reduced by any process that increases the distance between the stethoscope and the responsible cardiac event, including mechanical ventilation, obstructive lung disease, obesity, pneumothorax, and a pericardial effusion.

Figure 227-4

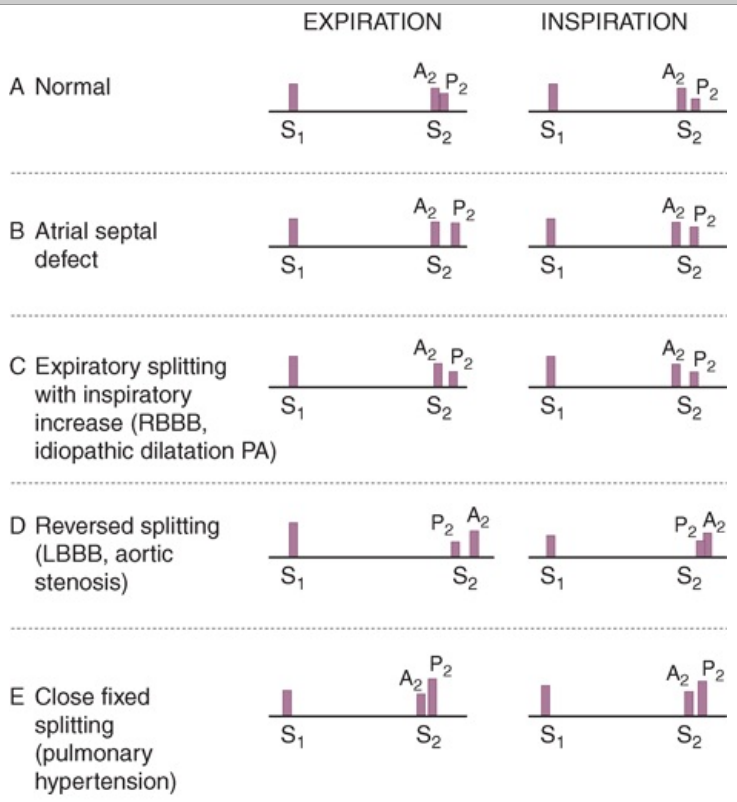

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Heart sounds. ***A***. Normal. S_1, first heart sound; S_2, second heart sound; A_2, aortic component of the second heart sound; P_2, pulmonic component of the second heart sound. ***B.*** Atrial septal defect with fixed splitting of S_2. ***C***. Physiologic but wide splitting of S_2 with right bundle branch block. ***D***. Reversed or paradoxical splitting of S_2 with left bundle branch block. ***E***. Narrow splitting of S_2 with pulmonary hypertension. *(From NO Fowler: Diagnosis of Heart Disease. New York, Springer-Verlag, 1991, p 31.)*

Aortic and pulmonic valve closure constitutes the second heart sound (S_2). With normal or physiologic

splitting, the A_2–P_2 interval increases with inspiration and narrows during expiration. This physiologic interval will widen with right bundle branch block because of the further delay in pulmonic valve closure and in patients with severe MR because of the premature closure of the aortic valve. An unusually narrowly split or even a singular S_2 is a feature of pulmonary arterial hypertension. Fixed splitting of S_2, in which the A_2–P_2 interval is wide and does not change during the respiratory cycle, occurs in patients with a secundum atrial septal defect. Reversed or paradoxical splitting refers to a pathologic delay in aortic valve closure, such as that which occurs in patients with left bundle branch block, right ventricular apical pacing, severe AS, HOCM, and acute myocardial ischemia. With reversed or paradoxical splitting, the individual components of S_2 are audible at end expiration, and their interval narrows with inspiration, the opposite of what would be expected under normal physiologic conditions. P_2 is considered loud when its intensity exceeds that of A_2 at the base, when it can be palpated in the area of the proximal pulmonary artery (second left interspace), or when both components of S_2 can be appreciated at the lower left sternal border or apex. The intensity of A_2 and P_2 decreases with aortic and pulmonic stenosis, respectively. In these conditions, a single S_2 may result.

SYSTOLIC SOUNDS

An ejection sound is a high-pitched early systolic sound that corresponds in timing to the upstroke of the carotid pulse. It usually is associated with congenital bicuspid aortic or pulmonic valve disease; however, ejection sounds are also sometimes audible in patients with isolated aortic or pulmonary root dilation and normal semilunar valves. The ejection sound that accompanies bicuspid aortic valve disease becomes softer and then inaudible as the valve calcifies and becomes more rigid. The ejection sound that accompanies pulmonic stenosis (PS) moves closer to the first heart sound as the severity of the stenosis increases. In addition, the pulmonic ejection sound is the only right-sided acoustic event that decreases in intensity with inspiration. Ejection sounds are often heard more easily at the lower left sternal border than they are at the base. Nonejection sounds (clicks), which occur after the onset of the carotid upstroke, are related to mitral valve prolapse and may be single or multiple. The nonejection click may introduce a murmur. This click-murmur complex will move away from the first heart sound with maneuvers that increase ventricular preload, such as squatting. On standing, the click and murmur move closer to S_1.

DIASTOLIC SOUNDS

The high pitched opening snap (OS) of MS occurs after a very short interval after the second heart sound. The A_2–OS interval is inversely proportional to the height of the left atrial–left ventricular diastolic pressure gradient. The intensity of both S_1 and the OS of MS decreases with progressive calcification and rigidity of the anterior mitral leaflets. The pericardial knock (PK) is also high-pitched and occurs slightly later than the opening snap, corresponding in timing to the abrupt cessation of ventricular expansion after tricuspid valve opening and to an exaggerated *y* descent seen in the jugular venous waveform in patients with constrictive pericarditis. A tumor plop is a lower-pitched sound that rarely can be heard in patients with atrial myxoma. It may be appreciated only in certain positions and arises from the diastolic prolapse of the tumor across the mitral valve.

The third heart sound (S_3) occurs during the rapid filling phase of ventricular diastole. It can be a normal finding in children, adolescents, and young adults; however, in older patients it signifies heart failure. A left-sided S_3 is a low-pitched sound best heard over the left ventricular (LV) apex. A right-sided S_3 is usually better heard over the lower left sternal border and becomes louder with inspiration. A left-sided S_3 in patients with chronic heart failure is predictive of cardiovascular morbidity and mortality. Interestingly, an S_3 is equally prevalent among heart failure patients with and without LV systolic dysfunction.

The fourth heart sound (S_4) occurs during the atrial filling phase of ventricular diastole and indicates left ventricular presystolic expansion. An S_4 is more common among patients who derive significant benefit from the atrial contribution to ventricular filling, such as those with chronic left ventricular hypertrophy or active myocardial ischemia. An S_4 is not present with atrial fibrillation.

CARDIAC MURMURS

Heart murmurs result from audible vibrations that are caused by increased turbulence and are defined by their timing within the cardiac cycle. Not all murmurs are indicative of structural heart disease, and the accurate identification of a benign or functional systolic murmur often can obviate the need for additional testing in healthy subjects. The duration, frequency, configuration, and intensity of a heart murmur are dictated by the magnitude, variability, and duration of the responsible pressure difference between two cardiac chambers, the two ventricles, or the ventricles and their respective great arteries. The intensity of a heart murmur is graded on a scale of 1 to 6; a thrill is present with murmurs of grade 4 or greater intensity. Other attributes of the murmur that aid in its accurate identification include its location, radiation, and response to bedside maneuvers. Although clinicians can detect and correctly

identify heart murmurs with only fair reliability, a careful and complete bedside examination usually can identify individuals with valvular heart disease for whom transthoracic echocardiography and clinical follow-up are indicated and exclude subjects for whom no further evaluation is necessary.

Systolic murmurs can be early, mid-, late, or holosystolic in timing **(Fig. 227-5)**. Acute severe MR results in a decrescendo early systolic murmur the characteristics of which are related to the progressive attenuation of the left ventricular to left atrial pressure gradient during systole because of the steep and rapid rise in left atrial pressure in this clinical context. Severe MR associated with posterior leaflet prolapse or flail radiates anteriorly and to the base, where it can be confused with the murmur of aortic stenosis. MR that is due to anterior leaflet involvement radiates posteriorly and to the axilla. With acute TR in patients with normal pulmonary arterty (PA) pressures, an early systolic murmur that may increase in intensity with inspiration may be heard at the left lower sternal border, with regurgitant *cv* waves visible in the jugular venous pulse.

Figure 227-5

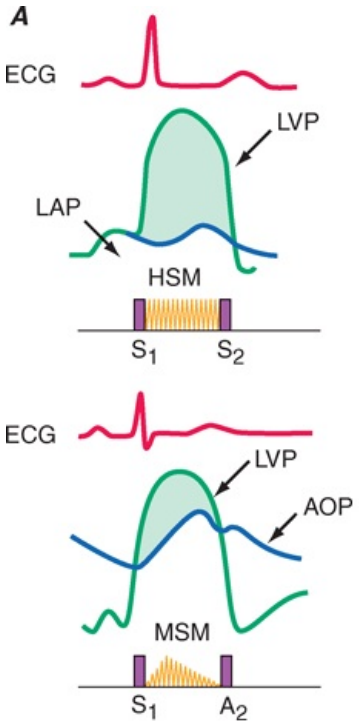

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

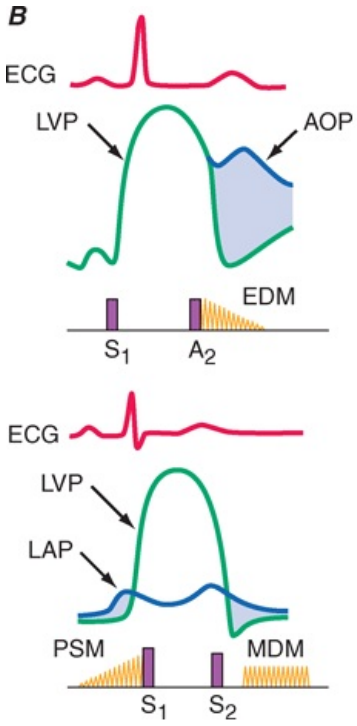

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. *Top*. Graphic representation of the systolic pressure difference (green shaded area) between left ventricle and left atrium with phonocardiographic recording of a holosystolic murmur (HSM) indicative of mitral regurgitation. ECG, electrocardiogram; LVP, left ventricular pressure; LAP, left atrial pressure; S_1, first heart sound; S_2 second heart sound. *Bottom*. Graphic representation of the systolic pressure gradient (green shaded area) between left ventricle and aorta in patient with aortic stenosis. A midsystolic murmur (MSM) with a crescendo-decrescendo configuration is recorded. AOP, aortic pressure. ***B.*** *Top*. Graphic representation of the diastolic pressure difference between the aorta and left ventricle (blue shaded area) in a patient with aortic regurgitation, resulting in a decrescendo, early diastolic murmur (EDM) beginning with A_2. *Bottom*. Graphic representation of the diastolic left atrial –left ventricular gradient (blue areas) in a patient with mitral stenosis with a mid-diastolic murmur (MDM) and late presystolic murmurs (PSM).

A midsystolic murmur begins after S_1 and ends before S_2; it is typically crescendo-decrescendo in configuration. Aortic stenosis is the most common cause of a midsystolic murmur in an adult. It is often difficult to estimate the severity of the valve lesion on the basis of the physical examination findings, especially in older hypertensive patients with stiffened carotid arteries or patients with low cardiac output in whom the intensity of the systolic heart murmur is misleadingly soft. Examination findings consistent with severe AS would include parvus et tardus carotid upstrokes, a late-peaking grade 3 or greater midsystolic murmur, a soft A_2, a sustained LV apical impulse, and an S_4. It is sometimes difficult to distinguish aortic sclerosis from more advanced degrees of valve stenosis. The former is defined by focal thickening and calcification of the aortic valve leaflets that is not severe enough to result in obstruction. These valve changes are associated with a Doppler jet velocity across the aortic valve of 2.5 m/s or less. Patients with aortic sclerosis can have grade 2 or 3 midsystolic murmurs identical in their acoustic characteristics to the murmurs heard in patients with more advanced degrees of AS. Other causes of a midsystolic heart murmur include pulmonic valve stenosis (with or without an ejection sound), HOCM, increased pulmonary blood flow in patients with a large atrial septal defect and left-to-right shunting, and several states associated with accelerated blood flow in the absence of structural heart disease, such as fever, thyrotoxicosis, pregnancy, anemia, and normal adolescence.

The murmur of hypertrophic obstructive cardiomyopathy has features of both obstruction to left ventricular outflow and MR, as would be expected from knowledge of the pathophysiology of this condition. The systolic murmur of HOCM usually can be distinguished from other causes on the basis of its response to bedside maneuvers, including Valsalva, passive leg raising, and standing/squatting. In general, maneuvers that decrease left ventricular preload (or increase left ventricular contractility) will cause the murmur to intensify, whereas maneuvers that increase left ventricular preload or afterload will cause a decrease in the intensity of the murmur. Accordingly, the systolic murmur of HOCM becomes louder during the strain phase of the Valsalva maneuver and after standing quickly from a squatting position. The murmur becomes softer with passive leg raising and when squatting. The murmur of AS is typically loudest in the second right interspace with radiation into the carotids, whereas the murmur of HOCM is best heard between the lower left sternal border and the apex. The

murmur of PS is best heard in the second left interspace. The midsystolic murmur associated with enhanced pulmonic blood flow in the setting of a large atrial septal defect (ASD) is usually loudest at the mid-left sternal border.

A late systolic murmur, heard best at the apex, indicates MVP. As previously noted, the murmur may or may not be introduced by a nonejection click. Differential radiation of the murmur, as previously described, may help identify the specific leaflet involved by the myxomatous process. The click-murmur complex behaves in a manner directionally similar to that demonstrated by the murmur of HOCM during the Valsalva and stand/squat maneuvers **(Fig. 227-6)**. The murmur of MVP can be identified by the accompanying nonejection click.

Figure 227-6.

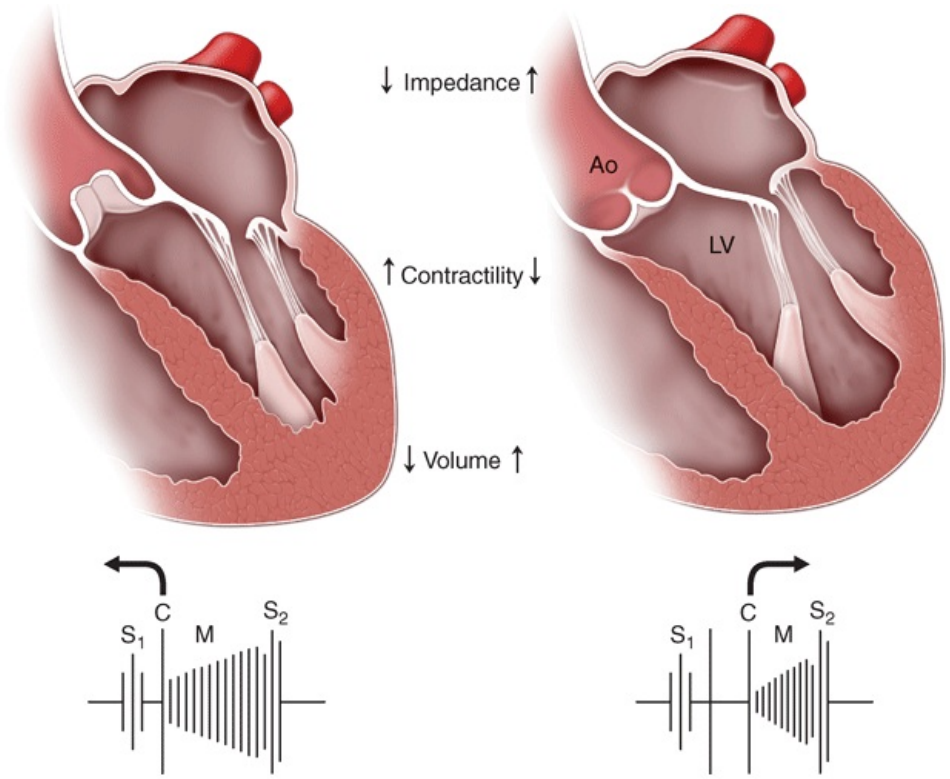

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Behavior of the click (C) and murmur (M) of mitral valve prolapse with changes in loading (volume, impedance) and contractility. S_1, first heart sound; S_2, second heart sound. With standing (left side of figure), volume and impedance decrease, as a result of which the click and murmur move closer to S_1. With squatting (right), the click and murmur move away from S_1 owing to the increases in left ventricular volume and impedance (afterload). *(Adapted from RA O'Rourke, MH Crawford: Curr Prob Cardiol 1:9, 1976.)*

Holosystolic murmurs are plateau in configuration and reflect a continuous and wide pressure gradient between the left ventricle and left atrium with chronic MR, the left ventricle and right ventricle with a ventricular septal defect (VSD), and the right ventricle and right atrium with TR. In contrast to acute MR, in chronic MR the left atrium is enlarged and its compliance is normal or increased to the extent that there is little if any further increase in left atrial pressure from any increase in regurgitant volume. The murmur of MR is best heard over the cardiac apex. The intensity of the murmur increases with maneuvers that increase left ventricular afterload, such as sustained hand grip. The murmur of a VSD (without significant pulmonary hypertension) is holosystolic and loudest at the mid-left sternal border, where a thrill is usually present. The murmur of TR is loudest at the lower left sternal border, increases in intensity with inspiration (Carvallo's sign), and is accompanied by visible *cv* waves in the jugular venous wave form and, on occasion, by pulsatile hepatomegaly.

DIASTOLIC MURMURS

In contrast to some systolic murmurs, diastolic heart murmurs always signify structural heart disease **(**Fig. 227-5**)**. The murmur associated with acute, severe AR is relatively soft and of short duration because of the rapid rise in left ventricular diastolic pressure and the progressive diminution of the aortic-left ventricular diastolic pressure gradient. In contrast, the murmur of chronic severe AR is classically heard as a decrescendo, blowing diastolic murmur along the left sternal border in patients with primary valve pathology and sometimes along the right sternal border in patients with primary

aortic root pathology. With chronic AR, the pulse pressure is wide and the arterial pulses are bounding in character. These signs of significant diastolic run-off are absent in the acute phase. The murmur of pulmonic regurgitation (PR) is also heard along the left sternal border. It is most commonly due to pulmonary hypertension and enlargement of the annulus of the pulmonic valve. S_2 is single and loud and may be palpable. There is a right ventricular/parasternal lift that is indicative of chronic right ventricular pressure overload. A less impressive murmur of PR is present after repair of tetralogy of Fallot or pulmonic valve atresia. In this postoperative setting, the murmur is softer and lower-pitched and the severity of the accompanying pulmonic regurgitation can be underestimated significantly.

Mitral stenosis is the classic cause of a mid- to late diastolic murmur, which is best heard over the apex in the left lateral decubitus position and is low-pitched or rumbling and is introduced by an OS in the early stages of the rheumatic disease process. Presystolic accentuation refers to an increase in the intensity of the murmur just before the first heart sound and occurs in patients with sinus rhythm. It is absent in patients with atrial fibrillation. The auscultatory findings in patients with rheumatic tricuspid stenosis typically are obscured by left-sided events, though they are similar in nature to those described in patients with MS. "Functional" mitral or tricuspid stenosis refers to the generation of mid-diastolic murmurs that are created by increased and accelerated transvalvular diastolic flow, even in the absence of valvular obstruction, in the setting of severe MR, severe TR, or a large ASD with left-to-right shunting. The Austin Flint murmur of chronic severe AR is a low-pitched mid- to late apical diastolic murmur that sometimes can be confused with MS. The Austin Flint murmur typically decreases in intensity after exposure to vasodilators, whereas the murmur of MS may be accompanied by an opening snap and also may increase in intensity after vasodilators because of the associated increase in cardiac output. Unusual causes of a mid-diastolic murmur include atrial myxoma, complete heart block, and acute rheumatic mitral valvulitis.

CONTINUOUS MURMUR

A continuous murmur is predicated on a pressure gradient that persists between two cardiac chambers or blood vessels across systole and diastole. The murmurs typically begin in systole, envelop the second heart sound (S_2), and continue through some portion of diastole. They can often be difficult to distinguish from individual systolic and diastolic murmurs in patients with mixed valvular heart disease. The classic example of a continuous murmur is that associated with a patent ductus arteriosus, which usually is heard in the second or third interspace at a slight distance from the sternal border. Other causes of a continuous murmur include a ruptured sinus of Valsalva aneurysm with creation of an aortic–right atrial or right ventricular fistula, a coronary or great vessel arteriovenous fistula, and an arteriovenous fistula constructed to provide dialysis access. There are two types of benign continuous murmurs. The cervical venous hum is heard in children or adolescents in the supraclavicular fossa. It can be obliterated with firm pressure applied to the diaphragm of the stethoscope, especially when the subject turns his or her head toward the examiner. The mammary souffle of pregnancy relates to enhanced arterial blood flow through engorged breasts. The diastolic component of the murmur can be obliterated with firm pressure over the stethoscope.

DYNAMIC AUSCULTATION

Diagnostic accuracy can be enhanced by the performance of simple bedside maneuvers to identify heart murmurs and characterize their significance **(Table 227-1)**. Except for the pulmonic ejection sound, right-sided events increase in intensity with inspiration and decrease with expiration; left-sided events behave oppositely (100% sensitivity, 88% specificity). As previously noted, the intensity of the murmurs associated with MR, VSD, and AR will increase in response to maneuvers that increase LV afterload, such as hand grip and vasopressors. The intensity of these murmurs will decrease after exposure to vasodilating agents. Squatting is associated with an abrupt increase in LV preload and afterload, whereas rapid standing results in a sudden decrease in preload. In patients with MVP, the click and murmur move away from the first heart sound with squatting because of the delay in onset of leaflet prolapse at higher ventricular volumes. With rapid standing, however, the click and murmur move closer to the first heart sound as prolapse occurs earlier in systole at a smaller chamber dimension. The murmur of HOCM behaves similarly, becoming softer and shorter with squatting (95% sensitivity, 85% specificity) and longer and louder on rapid standing (95% sensitivity, 84% specificity). A change in the intensity of a systolic murmur in the first beat after a premature beat or in the beat after a long cycle length in patients with atrial fibrillation suggests valvular AS rather than MR, particularly in an older patient in whom the murmur of the AS may be well transmitted to the apex (Gallavardin effect). Of note, however, the systolic murmur of HOCM also increases in intensity in the beat after a premature beat. This increase in intensity of any LV outflow murmur in the beat after a premature beat relates to the combined effects of enhanced LV filling (from the longer diastolic period) and postextrasystolic potentiation of LV contractile function. In either instance, forward flow will accelerate, causing an increase in the gradient across the LV outflow tract (dynamic or fixed) and a louder systolic murmur. In contrast, the intensity of the murmur of MR does not change in a postpremature beat, as there is relatively little change in the nearly constant LV to left atrial pressure gradient or further alteration in mitral valve flow. Bedside exercise can sometimes be performed to increase cardiac output and, secondarily, the intensity of both systolic and diastolic heart murmurs. Most left-sided heart murmurs decrease in intensity and duration during the strain phase of the Valsalva maneuver. The

murmurs associated with MVP and HOCM are the two notable exceptions. The Valsalva maneuver also can be used to assess the integrity of the heart and vasculature in the setting of advanced heart failure.

Table 227–1. Effects of Physiologic and Pharmacologic Interventions on the Intensity of Heart Murmurs and Sounds

Respiration Right-sided murmurs and sounds generally increase with inspiration, except for the PES. Left-sided murmurs and sounds are usually louder during expiration.
Valsalva maneuver Most murmurs decrease in length and intensity. Two exceptions are the systolic murmur of HOCM, which usually becomes much louder, and that of MVP, which becomes longer and often louder. After release of the Valsalva maneuver, right-sided murmurs tend to return to control intensity earlier than do left-sided murmurs.
After VPB or AF Murmurs originating at normal or stenotic semilunar valves increase in the cardiac cycle after a VPB or in the cycle after a long cycle length in AF. By contrast, systolic murmurs due to AV valve regurgitation do not change, diminish (papillary muscle dysfunction), or become shorter (MVP).
Positional changes With *standing*, most murmurs diminish, with two exceptions being the murmur of HOCM, which becomes louder, and that of MVP, which lengthens and often is intensified. With *squatting*, most murmurs become louder, but those of HOCM and MVP usually soften and may disappear. Passive leg raising usually produces the same results.
Exercise Murmurs due to blood flow across normal or obstructed valves (e.g,. PS, MS) become louder with both isotonic and submaximal isometric (hand grip) exercise. Murmurs of MR, VSD, and AR also increase with hand grip exercise. However, the murmur of HOCM often decreases with nearly maximum hand grip exercise. Left-sided S_4 and S_3 sounds are often accentuated by exercise, particularly when due to ischemic heart disease.

Abbreviations: TR, tricuspid regurgitation; TS, tricuspid stenosis; PR, pulmonic regurgitation; HOCM, hypertrophic obstructive cardiomyopathy; MVP, mitral valve prolapse; PS, pulmonic stenosis; MS, mitral stenosis; MR, mitral regurgitation; PES, pulmonic ejection sound; VSD, ventricular septal defect; AR, aortic regurgitation; VPB, ventricular premature beat; AF, atrial fibrillation.

PROSTHETIC HEART VALVES

The first clue that prosthetic valve dysfunction may contribute to recurrent symptoms is frequently a change in the quality of the heart sounds or the appearance of a new murmur. The heart sounds with a bioprosthetic valve resemble those generated by native valves. A mitral bioprosthesis usually is associated with a grade 2 or 3 midsystolic murmur along the left sternal border (created by turbulence across the valve struts as they project into the LV outflow tract) as well as by a soft mid-diastolic murmur that occurs with normal LV filling. This diastolic murmur often can be heard only in the left lateral decubitus position and after exercise. A high pitched or holosystolic apical murmur is indicative of paravalvular leak or bioprosthetic regurgitation, for which additional imagining is indicated. Clinical deterioration can occur rapidly after the first expression of bioprosthetic failure. A tissue valve in the aortic position is always associated with a grade 2 to 3 midsystolic murmur at the base or just below the suprasternal notch. A diastolic murmur of AR is abnormal in any circumstances. Mechanical valve dysfunction may first be suggested by a decrease in the intensity of either the opening or the closing sound. A high-pitched apical systolic murmur in patients with a mechanical mitral prosthesis and a diastolic decrescendo murmur in patients with a mechanical aortic prosthesis indicate paravalvular regurgitation. Patients with prosthetic valve thrombosis may present clinically with signs of shock, muffled heart sounds, and soft murmurs.

PERICARDIAL DISEASE

A pericardial friction rub is nearly 100% specific for the diagnosis of acute pericarditis, though the sensitivity of this finding is not nearly as high, as the rub may come and go over the course of an acute illness or be very difficult to elicit. The rub is heard as a leathery or scratchy three-component or two-component sound, though it may be monophasic. Classically, the three components are ventricular systole, rapid early diastolic filling, and late presystolic filling after atrial contraction in patients in sinus rhythm. It is necessary to listen to the heart in several positions. Additional clues may be present from the history and 12-lead electrocardiogram. The rub typically disappears as the volume of any pericardial effusion increases. Pericardial tamponade can be diagnosed with a sensitivity of 98%, a specificity of 83%, and a positive likelihood ratio of 5.9 (95% confidence intervals 2.4 to 14) by a pulsus paradoxus that exceeds 12 mmHg in a patient with a large pericardial effusion.

The findings on physical examination are integrated with the symptoms previously elicited with a careful history to construct an appropriate differential diagnosis and proceed with indicated imaging and laboratory assessment. The physical examination is an irreplaceable component of the diagnostic algorithm and in selected patients can inform prognosis. Educational efforts to improve clinician competence eventually may result in cost saving, particularly if the indications for imaging can be

influenced by the examination findings.

FURTHER READINGS

Drazner MH et al; Prognostic importance of elevated jugular venous pressure and a third heart sound in patients with heart failure. N Engl J Med 345:574, 2001[PMID: 11529211]

Fang JC, O'gara PT: The history and physical examination: An evidence-based approach, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

Markel H: The stethoscope and the art of listening. N Engl J Med 354:551, 2006[PMID: 16467541]

O'rourke RA et al: The history, physical examination, and cardiac auscultation, in *Hurst's The Heart*, 12th ed, V Fuster et al (eds). New York, McGraw-Hill, 2008

Roy CL et al: Does this patient with a pericardial effusion have cardiac tamponade? JAMA 297:1810, 2007[PMID: 17456823]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ELECTROCARDIOGRAPHY: INTRODUCTION

An electrocardiogram (ECG or EKG) is a graphic recording of electric potentials generated by the heart. The signals are detected by means of metal electrodes attached to the extremities and chest wall and then are amplified and recorded by the electrocardiograph. ECG *leads* actually display the instantaneous *differences* in potential between the electrodes.

The clinical utility of the ECG derives from its immediate availability as a noninvasive, inexpensive, and highly versatile test. In addition to its use in detecting arrhythmias, conduction disturbances, and myocardial ischemia, electrocardiography may reveal other findings related to life-threatening metabolic disturbances (e.g., hyperkalemia) or increased susceptibility to sudden cardiac death (e.g., QT prolongation syndromes).

Electrophysiology

(See also Chaps.232 and 233) Depolarization of the heart is the initiating event for cardiac contraction. The electric currents that spread through the heart are produced by three components: cardiac pacemaker cells, specialized conduction tissue, and the heart muscle itself. The ECG, however, records only the depolarization (stimulation) and repolarization (recovery) potentials generated by the atrial and ventricular myocardium.

The depolarization stimulus for the normal heartbeat originates in the *sinoatrial* (SA) *node* **(Fig. 228-1)**, or *sinus node*, a collection of *pacemaker cells*. These cells fire spontaneously; that is, they exhibit *automaticity*. The first phase of cardiac electrical activation is the spread of the depolarization wave through the right and left atria, followed by atrial contraction. Next, the impulse stimulates pacemaker and specialized conduction tissues in the atrioventricular (AV) nodal and His-bundle areas; together, these two regions constitute the AV junction. The bundle of His bifurcates into two main branches, the right and left bundles, which rapidly transmit depolarization wavefronts to the right and left ventricular myocardium by way of Purkinje fibers. The main left bundle bifurcates into two primary subdivisions: a left anterior fascicle and a left posterior fascicle. The depolarization wavefronts then spread through the ventricular wall, from endocardium to epicardium, triggering ventricular contraction.

Figure 228-1

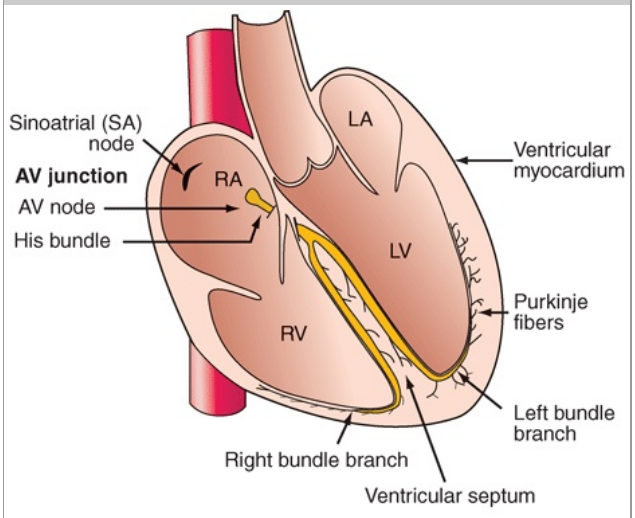

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic of the cardiac conduction system.

Since the cardiac depolarization and repolarization waves have direction and magnitude, they can be represented by vectors. Vector analysis illustrates a central concept of electrocardiography: The ECG records the complex spatial and temporal summation of electrical potentials from multiple myocardial fibers conducted to the surface of the body. This principle accounts for inherent limitations in both ECG *sensitivity* (activity from certain cardiac regions may be canceled out or may be too weak to be recorded) and *specificity* (the same vectorial sum can result from either a selective gain or a loss of forces in opposite directions).

ECG Waveforms and Intervals

The ECG waveforms are labeled alphabetically, beginning with the P wave, which represents atrial depolarization **(Fig. 228-2)**. The QRS complex represents ventricular depolarization, and the ST-T-U complex (ST segment, T wave, and U wave) represents ventricular repolarization. The J point is the junction between the end of the QRS complex and the beginning of the ST segment. Atrial repolarization is usually too low in amplitude to be detected, but it may become apparent in conditions such as acute pericarditis and atrial infarction.

Figure 228-2

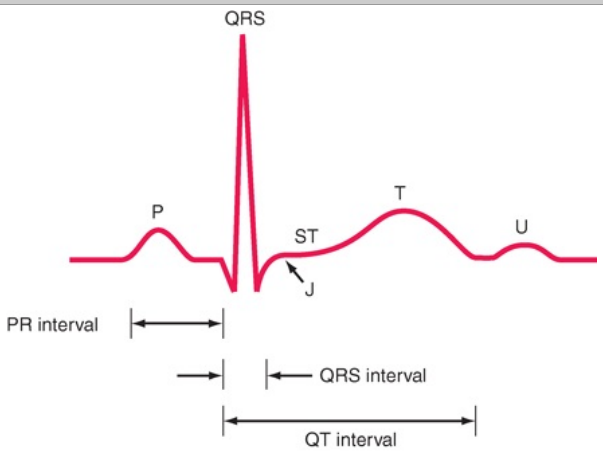

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Basic ECG waveforms and intervals. Not shown is the R-R interval, the time between consecutive QRS complexes.

The QRS-T waveforms of the surface ECG correspond in a general way with the different phases of simultaneously obtained ventricular *action potentials*, the intracellular recordings from single myocardial fibers (Chap. 232). The rapid upstroke (phase 0) of the action potential corresponds to the onset of QRS. The plateau (phase 2) corresponds to the isoelectric ST segment, and active repolarization (phase 3) corresponds to the inscription of the T wave. Factors that decrease the slope of phase 0 by impairing the influx of Na^+ (e.g., hyperkalemia and drugs such as flecainide) tend to increase QRS duration. Conditions that prolong phase 2 (amiodarone, hypocalcemia) increase the QT interval. In contrast, shortening of ventricular repolarization (phase 2), such as by digitalis administration or hypercalcemia, abbreviates the ST segment.

The electrocardiogram ordinarily is recorded on special graph paper that is divided into 1-mm^2 gridlike boxes. Since the ECG paper speed is generally 25 mm/s, the smallest (1 mm) horizontal divisions correspond to 0.04 (40 ms), with heavier lines at intervals of 0.20 s (200 ms). Vertically, the ECG graph measures the amplitude of a specific wave or deflection (1 mV = 10 mm with standard calibration; the voltage criteria for hypertrophy mentioned below are given in millimeters). There are four major ECG intervals: R-R, PR, QRS, and QT (Fig. 228-2). The heart rate (beats per minute) can be computed readily from the interbeat (R-R) interval by dividing the number of large (0.20 s) time units between consecutive R waves into 300 or the number of small (0.04 s) units into 1500. The PR interval measures the time (normally 120–200 ms) between atrial and ventricular depolarization, which includes the physiologic delay imposed by stimulation of cells in the AV junction area. The QRS interval (normally 100–110 ms or less) reflects the duration of ventricular depolarization. The QT interval includes both ventricular depolarization and repolarization times and varies inversely with the heart rate. A rate-related ("corrected") QT interval, QT_C, can be calculated as $QT/\sqrt{R\text{-}R}$ and normally is ≤0.44 s. (Some references give QT_C upper normal limits as 0.43 s in men and 0.45 s in women. Also, a number of different formulas have been proposed, without consensus, for calculating the QT_C.)

The QRS complex is subdivided into specific deflections or waves. If the initial QRS deflection in a particular lead is negative, it is termed a *Q wave*; the first positive deflection is termed an *R wave*. A negative deflection after an R wave is an *S wave*. Subsequent positive or negative waves are labeled R' and S', respectively. Lowercase letters (qrs) are used for waves of relatively small amplitude. An entirely negative QRS complex is termed a *QS wave*.

ECG Leads

The 12 conventional ECG leads record the difference in potential between electrodes placed on the surface of the body. These leads are divided into two groups: six limb (extremity) leads and six chest (precordial) leads. The limb leads record potentials transmitted onto the *frontal plane* **(Fig. 228-3*A*)**, and the chest leads record potentials transmitted onto the *horizontal plane* (Fig. 228-3*B*).

Figure 228-3

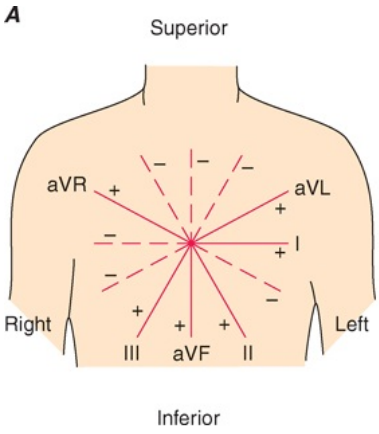

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

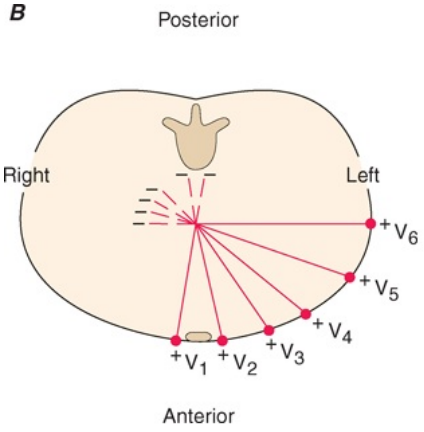

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The six frontal plane (*A*) and six horizontal plane (*B*) leads provide a three-dimensional representation of cardiac electrical activity.

The spatial orientation and polarity of the six frontal plane leads is represented on the hexaxial diagram **(Fig. 228-4)**. The six chest leads **(Fig. 228-5)** are unipolar recordings obtained by electrodes in the following positions: lead V_1, fourth intercostal space, just to the right of the sternum; lead V_2, fourth intercostal space, just to the left of the sternum; lead V_3, midway between V_2 and V_4; lead V_4, midclavicular line, fifth intercostal space; lead V_5, anterior axillary line, same level as V_4; and lead V_6, midaxillary line, same level as V_4 and V_5.

Figure 228-4

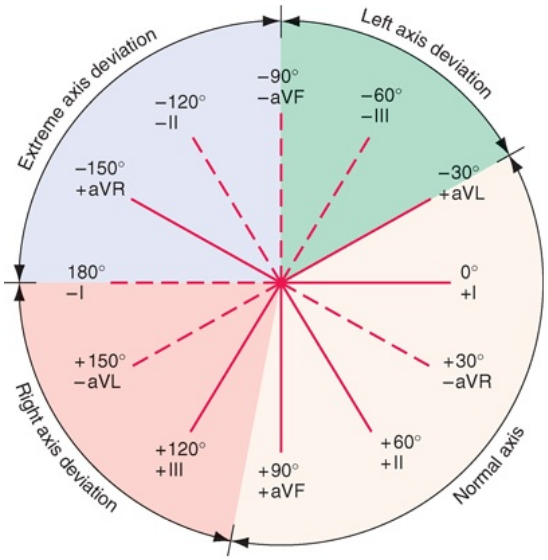

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The frontal plane (limb or extremity) leads are represented on a hexaxial diagram. Each ECG lead has a specific spatial orientation and polarity. The positive pole of each lead axis (*solid line*) and the negative pole (*hatched line*) are designated by their angular position relative to the positive pole of lead I (0°). The mean electrical axis of the QRS complex is measured with respect to this display.

Figure 228-5

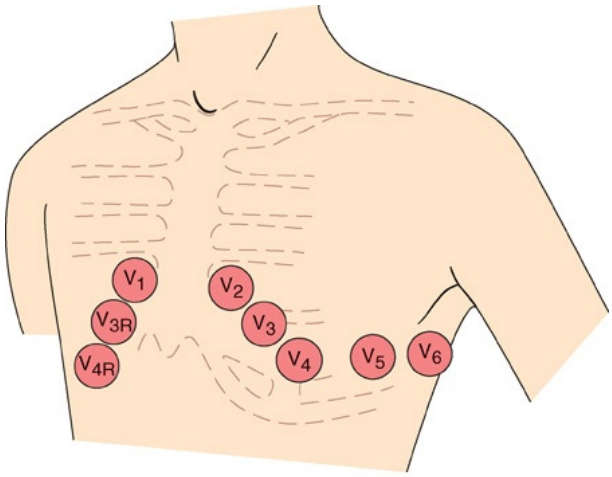

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The horizontal plane (chest or precordial) leads are obtained with electrodes in the locations shown.

Together, the frontal and horizontal plane electrodes provide a three-dimensional representation of cardiac electrical activity. Each lead can be likened to a different video camera angle "looking" at the same events—atrial and ventricular depolarization and repolarization—from different spatial orientations. The conventional 12-lead ECG can be supplemented with additional leads in special circumstances. For example, right precordial leads V_3R, V_4R, etc., are useful in detecting evidence of acute right ventricular ischemia. Bedside monitors and ambulatory ECG (Holter) recordings usually employ only one or two modified leads. **Intracardiac electrocardiography and electrophysiologic testing are discussed in Chaps. 232 and 233.**

The ECG leads are configured so that a positive (upright) deflection is recorded in a lead if a wave of depolarization spreads toward the positive pole of that lead, and a negative deflection is recorded if the wave spreads toward the negative pole. If the mean orientation of the depolarization vector is at right angles to a particular lead axis, a biphasic (equally positive and negative) deflection will be recorded.

GENESIS OF THE NORMAL ECG

P Wave

The normal atrial depolarization vector is oriented downward and toward the subject's left, reflecting the spread of depolarization from the sinus node to the right and then the left atrial myocardium. Since this vector points toward the positive pole of lead II and toward the negative pole of lead aVR, the normal P wave will be positive in lead II and negative in lead aVR. By contrast, activation of the atria from an ectopic pacemaker in the lower part of either atrium or in the AV junction region may produce retrograde P waves (negative in lead II, positive in lead aVR). The normal P wave in lead V_1 may be biphasic with a positive component reflecting right atrial depolarization, followed by a small (<1 mm^2) negative component reflecting left atrial depolarization.

QRS Complex

Normal ventricular depolarization proceeds as a rapid, continuous spread of activation wave fronts. This complex process can be divided into two major sequential phases, and each phase can be represented by a mean vector **(Fig. 228-6)**. The first phase is depolarization of the interventricular septum from the left to the right and anteriorly (vector 1). The second results from the simultaneous depolarization of the right and left ventricles; it normally is dominated by the more massive left ventricle, so that vector 2 points leftward and posteriorly. Therefore, a right precordial lead (V_1) will record this biphasic depolarization process with a small positive deflection (septal r wave) followed by a larger negative deflection (S wave). A left precordial lead, e.g., V_6, will record the same sequence with a small negative deflection (septal q wave) followed by a relatively tall positive deflection (R wave). Intermediate leads show a relative increase in R-wave amplitude (normal R-wave progression) and a decrease in S-wave amplitude progressing across the chest from right to left. The precordial lead where the R and S waves are of approximately equal amplitude is referred to as the *transition zone* (usually V_3 or V_4) **(Fig. 228-7)**.

Figure 228-6

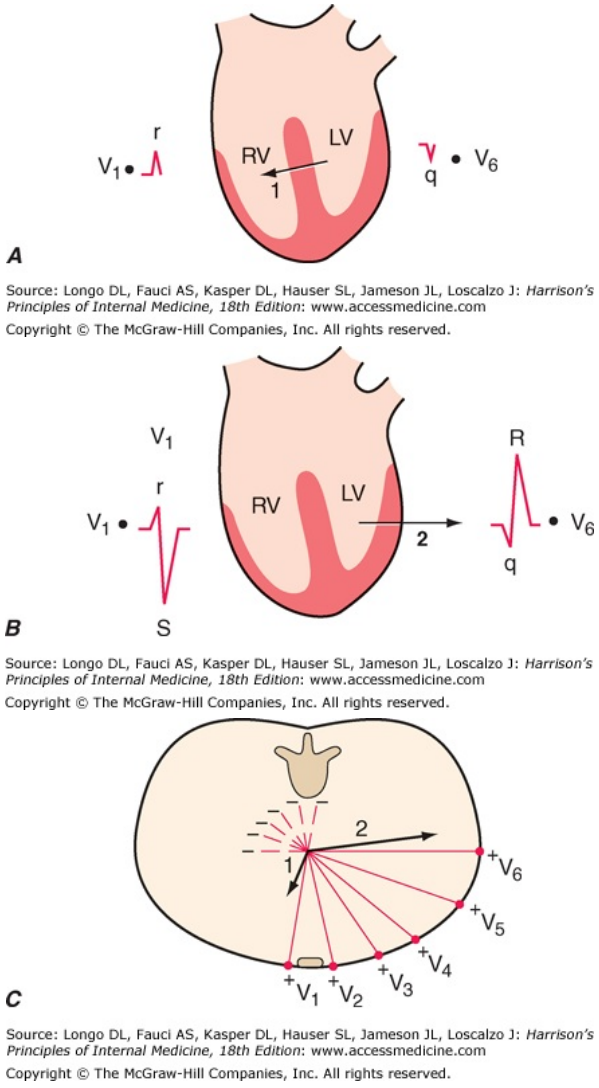

Ventricular depolarization can be divided into two major phases, each represented by a vector. ***A.*** The first phase (*arrow 1*) denotes depolarization of the ventricular septum, beginning on the left side and spreading to the right. This process is represented by a small "septal" r wave in lead V_1 and a small septal q wave in lead V_6. ***B.*** Simultaneous depolarization of the left and right ventricles (LV and RV) constitutes the second phase. Vector 2 is oriented to the left and posteriorly, reflecting the electrical predominance of the LV. ***C.*** Vectors (*arrows*) representing these two phases are shown in reference to the horizontal plane leads. *(After Goldberger.)*

Figure 228-7

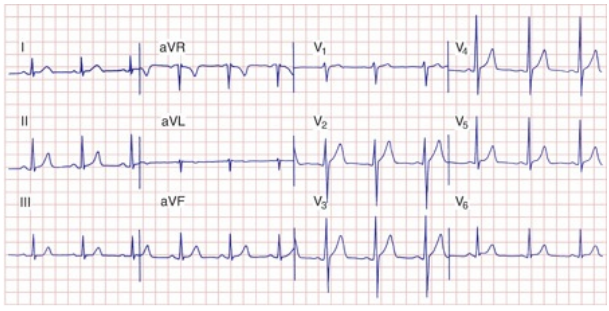

Normal electrocardiogram from a healthy subject. Sinus rhythm is present with a heart rate of 75 beats per minute. PR interval is 0.16 s; QRS interval (duration) is 0.08 s; QT interval is 0.36 s; QT_C is 0.40 s; the mean QRS axis is about +70°. The precordial leads show normal R-wave progression with the transition zone (R wave = S wave) in lead V_3.

The QRS pattern in the extremity leads may vary considerably from one normal subject to another depending on the *electrical axis* of the QRS, which describes the mean orientation of the QRS vector with reference to the six frontal plane leads. Normally, the QRS axis ranges from –30° to +100° (Fig. 228-4). An axis more negative than –30° is referred to as *left axis deviation*, and an axis more positive than +100° is referred to as *right axis deviation*. Left axis deviation may occur as a normal variant but is more commonly associated with left ventricular hypertrophy, a block in

the anterior fascicle of the left bundle system (left anterior fascicular block or hemiblock), or inferior myocardial infarction. Right axis deviation also may occur as a normal variant (particularly in children and young adults), as a spurious finding due to reversal of the left and right arm electrodes, or in conditions such as right ventricular overload (acute or chronic), infarction of the lateral wall of the left ventricle, dextrocardia, left pneumothorax, and left posterior fascicular block.

T Wave and U Wave

Normally, the mean T-wave vector is oriented roughly concordant with the mean QRS vector (within about 45° in the frontal plane). Since depolarization and repolarization are electrically opposite processes, this normal QRS–T-wave vector concordance indicates that repolarization normally must proceed in the reverse direction from depolarization (i.e., from ventricular epicardium to endocardium). The normal U wave is a small, rounded deflection (≤1 mm) that follows the T wave and usually has the same polarity as the T wave. An abnormal increase in U-wave amplitude is most commonly due to drugs (e.g., dofetilide, amiodarone, sotalol, quinidine, procainamide, disopyramide) or to hypokalemia. Very prominent U waves are a marker of increased susceptibility to the *torsades de pointes* type of ventricular tachycardia (Chap. 233). Inversion of the U wave in the precordial leads is abnormal and may be a subtle sign of ischemia.

MAJOR ECG ABNORMALITIES

Cardiac Enlargement and Hypertrophy

Right atrial overload (acute or chronic) may lead to an increase in P-wave amplitude (≥2.5 mm) **(Fig. 228-8)**. Left atrial overload typically produces a biphasic P wave in V_1 with a broad negative component or a broad (≥120 ms), often notched P wave in one or more limb leads (Fig. 228-8). This pattern may also occur with left atrial conduction delays in the absence of actual atrial enlargement, leading to the more general designation of *left atrial abnormality*.

Figure 228-8

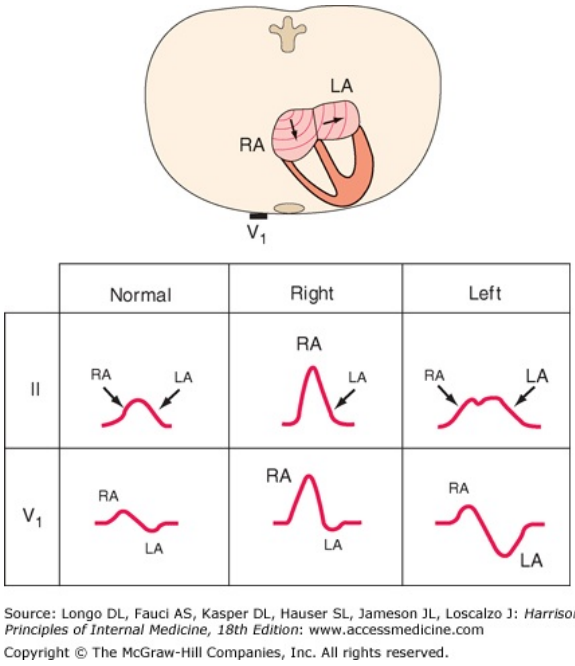

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Right atrial (RA) overload may cause tall, peaked P waves in the limb or precordial leads. Left atrial (LA) abnormality may cause broad, often notched P waves in the limb leads and a biphasic P wave in lead V_1 with a prominent negative component representing delayed depolarization of the LA. (*After MK Park, WG Guntheroth: How to Read Pediatric ECGs, 4th ed. St. Louis, Mosby/Elsevier, 2006.*)

Right ventricular hypertrophy due to a pressure load (as from pulmonic valve stenosis or pulmonary artery hypertension) is characterized by a relatively tall R wave in lead V_1 (R ≥ S wave), usually with right axis deviation **(Fig. 228-9)**; alternatively, there may be a qR pattern in V_1 or V_3R. ST depression and T-wave inversion in the right-to-midprecordial leads are also often present. This pattern, formerly called right ventricular "strain," is attributed to repolarization abnormalities in acutely or chronically overloaded muscle. Prominent S waves may occur in the left lateral precordial leads. Right ventricular hypertrophy due to ostium secundum–type atrial septal defects, with the accompanying right ventricular volume overload, is commonly associated with an incomplete or complete right bundle branch block pattern with a rightward QRS axis.

Figure 228-9

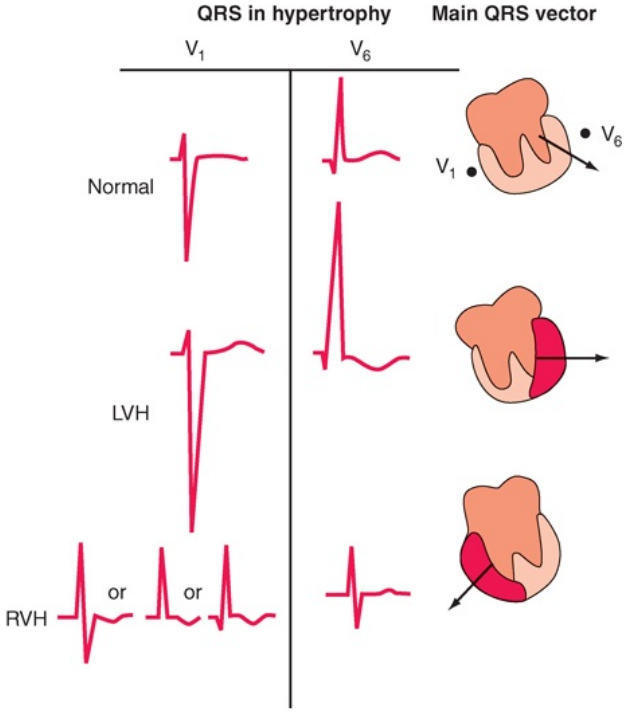

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left ventricular hypertrophy (LVH) increases the amplitude of electrical forces directed to the left and posteriorly. In addition, repolarization abnormalities may cause ST-segment depression and T-wave inversion in leads with a prominent R wave. Right ventricular hypertrophy (RVH) may shift the QRS vector to the right; this effect usually is associated with an R, RS, or qR complex in lead V_1. T-wave inversions may be present in right precordial leads.

Acute cor pulmonale due to pulmonary embolism (Chap. 262), for example, may be associated with a normal ECG or a variety of abnormalities. Sinus tachycardia is the most common arrhythmia, although other tachyarrhythmias, such as atrial fibrillation or flutter, may occur. The QRS axis may shift to the right, sometimes in concert with the so-called $S_1Q_3T_3$ pattern (prominence of the S wave in lead I and the Q wave in lead III, with T-wave inversion in lead III). Acute right ventricular dilation also may be associated with slow R-wave progression and ST-T abnormalities in V_1 to V_4 simulating acute anterior infarction. A right ventricular conduction disturbance may appear.

Chronic cor pulmonale due to obstructive lung disease (Chap. 234) usually does not produce the classic ECG patterns of right ventricular hypertrophy noted above. Instead of tall right precordial R waves, chronic lung disease more typically is associated with small R waves in right-to-midprecordial leads (slow R-wave progression) due in part to downward displacement of the diaphragm and the heart. Low-voltage complexes are commonly present, owing to hyperaeration of the lungs.

A number of different voltage criteria for *left ventricular hypertrophy* (Fig. 228-9) have been proposed on the basis of the presence of tall left precordial R waves and deep right precordial S waves [e.g., SV_1 + (RV_5 or RV_6) >35 mm]. Repolarization abnormalities (ST depression with T-wave inversions, formerly called the left ventricular "strain" pattern) also may appear in leads with prominent R waves. However, prominent precordial voltages may occur as a normal variant, especially in athletic or young individuals. Left ventricular hypertrophy may increase limb lead voltage with or without increased precordial voltage (e.g., RaVL + SV_3 >20 mm in women and >28 mm in men). The presence of left atrial abnormality increases the likelihood of underlying left ventricular hypertrophy in cases with borderline voltage criteria. Left ventricular hypertrophy often progresses to incomplete or complete left bundle branch block. The sensitivity of conventional voltage criteria for left ventricular hypertrophy is decreased in obese persons and smokers. ECG evidence for left ventricular hypertrophy is a major noninvasive marker of increased risk of cardiovascular morbidity and mortality rates, including sudden cardiac death. However, because of false-positive and false-negative diagnoses, the ECG is of limited utility in diagnosing atrial or ventricular enlargement. More definitive information is provided by echocardiography (Chap. 229).

Bundle Branch Blocks

Intrinsic impairment of conduction in either the right or the left bundle system (intraventricular conduction disturbances) leads to prolongation of the QRS interval. With complete bundle branch blocks, the QRS interval is ≥120 ms in duration; with incomplete blocks, the QRS interval is between 100 and 120 ms. The QRS vector usually is oriented in the direction of the myocardial region where depolarization is delayed **(Fig. 228-10)**. Thus, with right bundle branch block, the terminal QRS vector is oriented to the right and anteriorly (rSR' in V_1 and qRS in V_6, typically). Left bundle branch block alters both early and later phases of ventricular depolarization. The major QRS vector is directed to the left and posteriorly. In addition, the normal early left-to-right pattern of septal activation is disrupted such that septal depolarization proceeds from right to left as well. As a result, left bundle branch block generates wide, predominantly negative (QS) complexes in lead V_1 and entirely positive (R) complexes in lead V_6. A pattern identical to that of left bundle branch block, preceded by a sharp spike, is seen in most cases of electronic right ventricular pacing because of the relative delay in left ventricular activation.

Figure 228-10

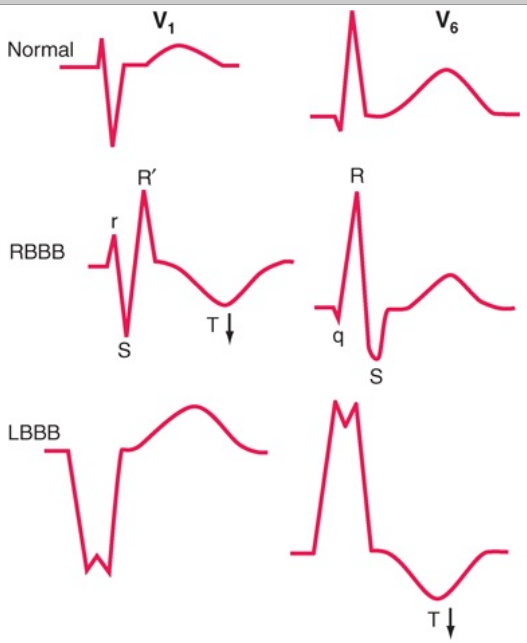

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Comparison of typical QRS-T patterns in right bundle branch block (RBBB) and left bundle branch block (LBBB) with the normal pattern in leads V_1 and V_6. Note the secondary T-wave inversions (*arrows*) in leads with an rSR' complex with RBBB and in leads with a wide R wave with LBBB.

Bundle branch block may occur in a variety of conditions. In subjects without structural heart disease, right bundle branch block is seen more commonly than left bundle branch block. Right bundle branch block also occurs with heart disease, both congenital (e.g., atrial septal defect) and acquired (e.g., valvular, ischemic). Left bundle branch block is often a marker of one of four underlying conditions associated with increased risk of cardiovascular morbidity and mortality rates: coronary heart disease (frequently with impaired left ventricular function), hypertensive heart disease, aortic valve disease, and cardiomyopathy. Bundle branch blocks may be chronic or intermittent. A bundle branch block may be rate-related; for example, it often occurs when the heart rate exceeds some critical value.

Bundle branch blocks and depolarization abnormalities secondary to artificial pacemakers not only affect ventricular depolarization (QRS) but also are characteristically associated with *secondary repolarization* (ST-T) abnormalities. With bundle branch blocks, the T wave is typically opposite in polarity to the last deflection of the QRS (Fig. 228-10). This discordance of the QRS–T-wave vectors is caused by the altered sequence of repolarization that occurs secondary to altered depolarization. In contrast, *primary repolarization* abnormalities are independent of QRS changes and are related instead to actual alterations in the electrical properties of the myocardial fibers themselves (e.g., in the resting membrane potential or action potential duration), not just to changes in the sequence of repolarization. Ischemia, electrolyte imbalance, and drugs such as digitalis all cause such primary ST–T-wave changes. Primary and secondary T-wave changes may coexist. For example, T-wave inversions in the right precordial leads with left bundle branch block or in the left precordial leads with right bundle branch block may be important markers of underlying ischemia or other abnormalities. A distinctive abnormality simulating right bundle branch block with ST-segment elevations in the right chest leads is seen with the Brugada pattern (Chap. 233).

Partial blocks (fascicular or "hemiblocks") in the left bundle system (left anterior or posterior fascicular blocks) generally do not prolong the QRS duration substantially but instead are associated with shifts in the frontal plane QRS axis (leftward or rightward, respectively). More complex combinations of fascicular and bundle branch blocks may occur that involve the left and right bundle system. Examples of *bifascicular block* include right bundle branch block and left posterior fascicular block, right bundle branch block with left anterior fascicular block, and complete left bundle branch block. Chronic bifascicular block in an asymptomatic individual is associated with a relatively low risk of progression to high-degree AV heart block. In contrast, new bifascicular block with acute anterior myocardial infarction carries a much greater risk of complete heart block. Alternation of right and left bundle branch block is a sign of *trifascicular disease*. However, the presence of a prolonged PR interval and bifascicular block does not necessarily indicate trifascicular involvement, since this combination may arise with AV node disease and bifascicular block. Intraventricular conduction delays also can be caused by extrinsic (toxic) factors that slow ventricular conduction, particularly hyperkalemia or drugs (e.g., class 1 antiarrhythmic agents, tricyclic antidepressants, phenothiazines).

Prolongation of QRS duration does not necessarily indicate a conduction delay but may be due to *preexcitation* of the ventricles via a bypass tract, as in Wolff-Parkinson-White (WPW) patterns (Chap. 233) and related variants. The diagnostic triad of WPW consists of a wide QRS complex associated with a relatively short PR interval and slurring of the initial part of the QRS (delta wave), with the latter effect being due to aberrant activation of ventricular myocardium. The presence of a bypass tract predisposes to reentrant supraventricular tachyarrhythmias.

Myocardial Ischemia and Infarction

(See also Chap. 245) The ECG is a cornerstone in the diagnosis of acute and chronic ischemic heart disease. The findings depend on several key factors: the nature of the process [reversible (i.e., ischemia) versus irreversible (i.e., infarction)], the duration (acute versus chronic), the extent (transmural versus subendocardial), and localization (anterior versus inferoposterior), as well as the presence of other underlying abnormalities (ventricular hypertrophy, conduction defects).

Ischemia exerts complex time-dependent effects on the electrical properties of myocardial cells. Severe, acute ischemia lowers the resting membrane potential and shortens the duration of the action potential. Such changes cause a voltage gradient between normal and ischemic zones. As a consequence, current flows between those regions. These currents of injury are represented on the surface ECG by deviation of the ST segment **(Fig. 228-11)**. When the acute ischemia is *transmural*, the ST vector usually is shifted in the direction of the outer (epicardial) layers, producing ST elevations and sometimes, in the earliest stages of ischemia, tall, positive so-called hyperacute T waves over the ischemic zone. With ischemia confined primarily to the *subendocardium*, the ST vector typically shifts toward the subendocardium and ventricular cavity, so that overlying (e.g., anterior precordial) leads show ST-segment depression (with ST elevation in lead aVR). Multiple factors affect the amplitude of acute ischemic ST deviations. Profound ST elevation or depression in multiple leads usually indicates very severe ischemia. From a clinical viewpoint, the division of acute myocardial infarction into ST-segment elevation and non-ST elevation types is useful since the efficacy of acute reperfusion therapy is limited to the former group.

Figure 228-11

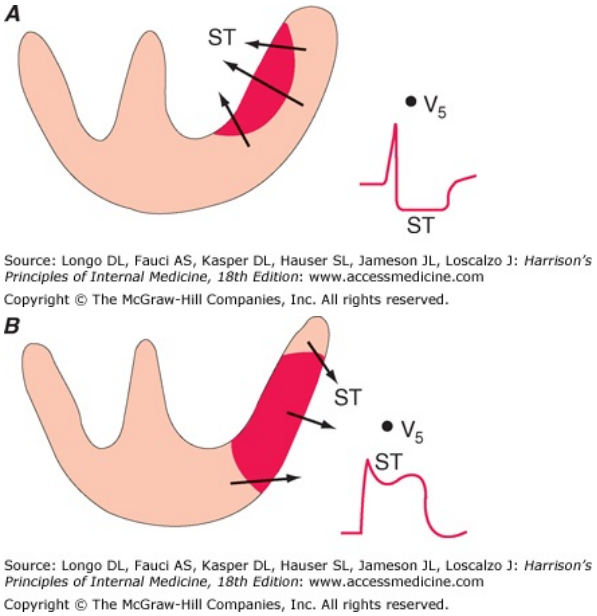

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute ischemia causes a current of injury. With predominant subendocardial ischemia (***A***), the resultant ST vector will be directed toward the inner layer of the affected ventricle and the ventricular cavity. Overlying leads therefore will record ST depression. With ischemia involving the outer ventricular layer (***B***) (transmural or epicardial injury), the ST vector will be directed outward. Overlying leads will record ST elevation.

The ECG leads are usually more helpful in localizing regions of ST elevation than non-ST elevation ischemia. For example, acute transmural anterior (including apical and lateral) wall ischemia is reflected by ST elevations or increased T-wave positivity in one or more of the precordial leads (V_1–V_6) and leads I and aVL. Inferior wall ischemia produces changes in leads II, III, and aVF. "Posterior" wall ischemia (usually associated with lateral or inferior involvement) may be indirectly recognized by *reciprocal* ST depressions in leads V_1 to V_3 (thus constituting an ST elevation "equivalent" acute coronary syndrome). Right ventricular ischemia usually produces ST elevations in right-sided chest leads (Fig. 228-5). When ischemic ST elevations occur as the earliest sign of acute infarction, they typically are followed within a period ranging from hours to days by evolving T-wave inversions and often by Q waves occurring in the same lead distribution. Reversible transmural ischemia, for example, due to coronary vasospasm (Prinzmetal's variant angina and probably the Tako-Tsubo "stress" cardiomyopathy syndrome), may cause transient ST-segment elevations without development of Q waves, as may very early reperfusion in acute coronary syndromes. Depending on the severity and duration of ischemia, the ST elevations may resolve completely in minutes or be followed by T-wave inversions that persist for hours or even days. Patients with ischemic chest pain who present with deep T-wave inversions in multiple precordial leads (e.g., V_1–V_4) with or without cardiac enzyme elevations typically have severe obstruction in the left anterior descending coronary artery system **(Fig. 228-12)**. In contrast, patients whose baseline ECG already shows abnormal T-wave inversions may develop T-wave normalization (pseudonormalization) during episodes of acute transmural ischemia.

Figure 228-12

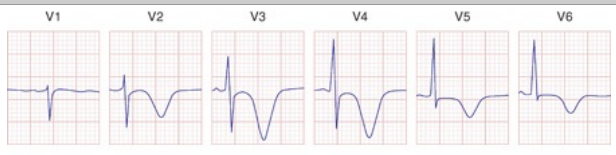

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe anterior wall ischemia (with or without infarction) may cause prominent T-wave inversions in the precordial leads. This pattern (sometimes referred to as Wellens T waves) is usually associated with a high-grade stenosis of the left anterior descending coronary artery.

With infarction, depolarization (QRS) changes often accompany repolarization (ST-T) abnormalities. Necrosis of sufficient myocardial tissue may lead to decreased R-wave amplitude or abnormal Q waves (even in the absence of transmurality) in the anterior or inferior leads **(Fig. 228-13)**. Previously, abnormal Q waves were considered markers of transmuralmyocardial infarction, whereas subendocardial infarcts were thought not to produce Q waves. However, careful ECG-pathology correlative studies have indicated that transmural infarcts may occur without Q waves and that subendocardial (nontransmural) infarcts sometimes may be associated with Q waves. Therefore, infarcts are more appropriately classified as "Q-wave" or "non-Q-wave." The major acute ECG changes in syndromes of ischemic heart disease are summarized schematically in **Fig. 228-14**. Loss of depolarization forces due to posterior or lateral infarction may cause reciprocal increases in R-wave amplitude in leads V_1 and V_2 without diagnostic Q waves in any of the conventional leads. Atrial infarction may be associated with PR-segment deviations due to an atrial current of injury, changes in P-wave morphology, or atrial arrhythmias.

Figure 228-13

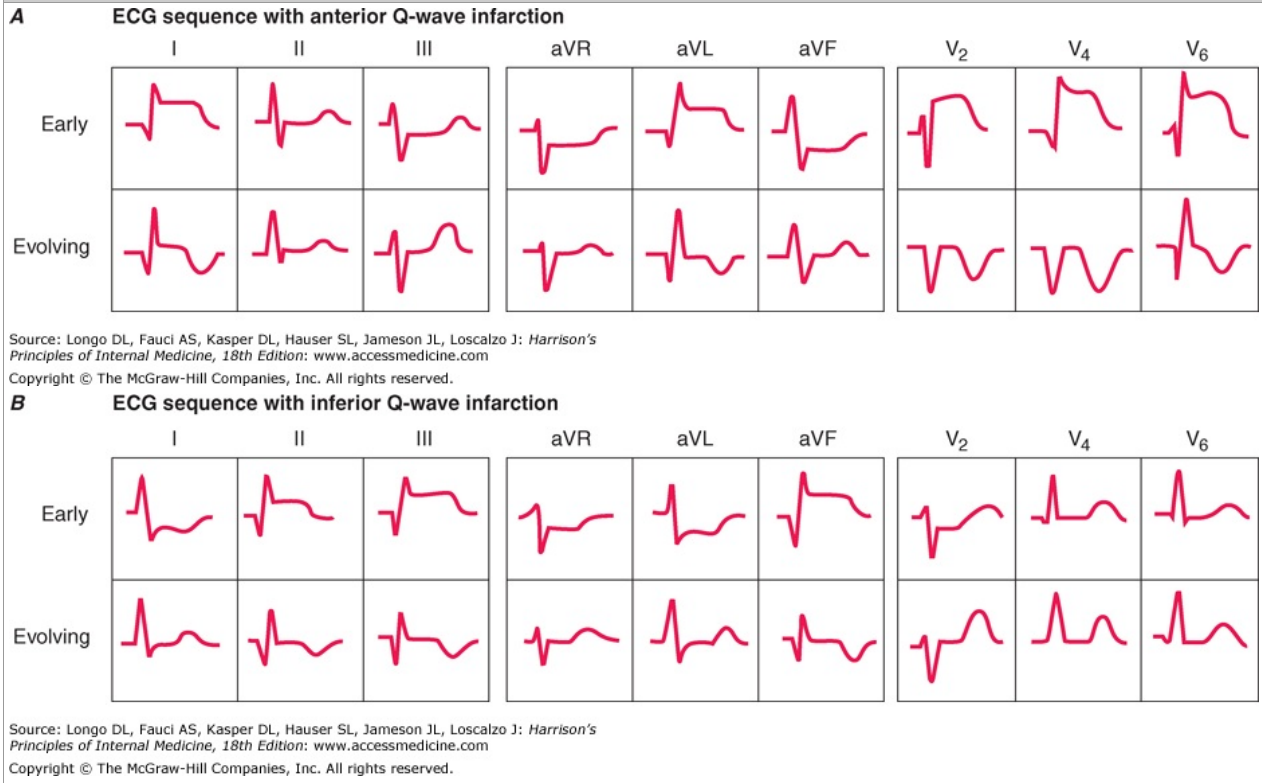

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sequence of depolarization and repolarization changes with (***A***) acute anterior and (***B***) acute inferior wall Q-wave infarctions. With anterior infarcts, ST elevation in leads I and aVL and the precordial leads may be accompanied by reciprocal ST depressions in leads II, III, and aVF. Conversely, acute inferior (or posterolateral) infarcts may be associated with reciprocal ST depressions in leads V_1 to V_3. *(After Goldberger.)*

Figure 228-14

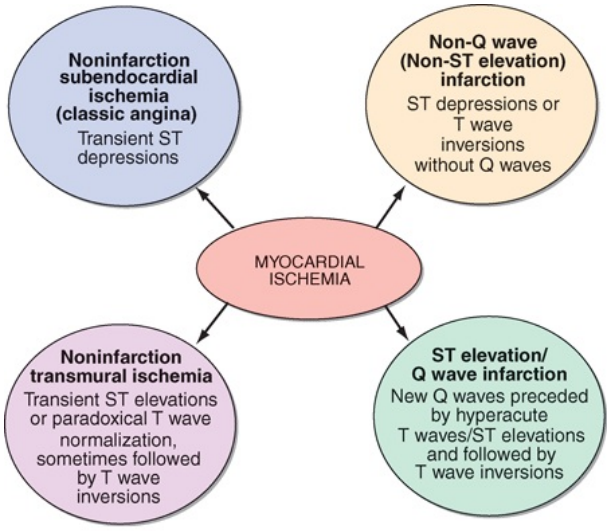

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Variability of ECG patterns with acute myocardial ischemia. The ECG also may be normal or nonspecifically abnormal. Furthermore, these categorizations are not mutually exclusive. *(After Goldberger, 2006.)*

In the weeks and months after infarction, these ECG changes may persist or begin to resolve. Complete normalization of the ECG after Q-wave infarction is uncommon but may occur, particularly with smaller infarcts. In contrast, ST-segment elevations that persist for several weeks or more after a Q-wave infarct usually correlate with a severe underlying wall motion disorder (akinetic or dyskinetic zone), although not necessarily a frank ventricular aneurysm. ECG changes due to ischemia may occur spontaneously or may be provoked by various exercise protocols (stress electrocardiography; Chap. 243).

The ECG has important limitations in both sensitivity and specificity in the diagnosis of ischemic heart disease. Although a single normal ECG does not exclude ischemia or even acute infarction, a normal ECG *throughout* the course of an acute infarct is distinctly uncommon. Prolonged chest pain without diagnostic ECG changes therefore should always prompt a careful search for other noncoronary causes of chest pain (Chap. 12). Furthermore, the diagnostic changes of acute or evolving ischemia are often masked by the presence of left bundle branch block, electronic ventricular pacemaker patterns, and Wolff-Parkinson-White preexcitation. However, clinicians continue to overdiagnose ischemia or infarction based on the presence of ST-segment elevations or depressions; T-wave inversions; tall, positive T waves; or Q waves *not* related to ischemic heart disease (pseudoinfarct patterns). For example, ST-segment elevations simulating ischemia may occur with acute pericarditis or myocarditis, as a normal variant (including the typical "early repolarization" pattern), or in a variety of other conditions **(Table 228-1)**. Similarly, tall, positive T waves do not invariably represent hyperacute ischemic changes but may also be caused by normal variants, hyperkalemia, cerebrovascular injury, and left ventricular volume overload due to mitral or aortic regurgitation, among other causes.

Table 228–1. Differential Diagnosis of ST-Segment Elevations

Ischemia/myocardial infarction
Noninfarction, transmural ischemia (Prinzmetal's angina, and probably Tako-Tsubo syndrome, which may also exactly simulate classical acute infarction)
Acute myocardial infarction
Postmyocardial infarction (ventricular aneurysm pattern)
Acute pericarditis
Normal variants (including "early repolarization" patterns)
Left ventricular hypertrophy/left bundle branch block[a]
Other (rarer)
Acute pulmonary embolism[a]
Brugada patterns (right bundle branch block–like pattern with ST elevations in right precordial leads)[a]
Class 1C antiarrhythmic drugs[a]
DC cardioversion
Hypercalcemia[a]
Hyperkalemia[a]
Hypothermia [J (Osborn) waves]
Nonischemic myocardial injury

Myocarditis
Tumor invading left ventricle
Trauma to ventricles

[a]Usually localized to V_1–V_2 or V_3.

Source: Modified from Goldberger.

ST-segment elevations and tall, positive T waves are common findings in leads V_1 and V_2 in left bundle branch block or left ventricular hypertrophy in the absence of ischemia. The differential diagnosis of Q waves includes physiologic or positional variants, ventricular hypertrophy, acute or chronic noncoronary myocardial injury, hypertrophic cardiomyopathy, and ventricular conduction disorders. Digoxin, ventricular hypertrophy, hypokalemia, and a variety of other factors may cause ST-segment depression mimicking subendocardial ischemia. Prominent T-wave inversion may occur with ventricular hypertrophy, cardiomyopathies, myocarditis, and cerebrovascular injury (particularly intracranial bleeds), among many other conditions.

Metabolic Factors and Drug Effects

A variety of metabolic and pharmacologic agents alter the ECG and, in particular, cause changes in repolarization (ST-T-U) and sometimes QRS prolongation. Certain life-threatening electrolyte disturbances may be diagnosed initially and monitored from the ECG. *Hyperkalemia* produces a sequence of changes **(Fig. 228-15)**, usually beginning with narrowing and peaking (tenting) of the T waves. Further elevation of extracellular K^+ leads to AV conduction disturbances, diminution in P-wave amplitude, and widening of the QRS interval. Severe hyperkalemia eventually causes cardiac arrest with a slow sinusoidal type of mechanism ("sine-wave" pattern) followed by asystole. *Hypokalemia***(Fig. 228-16)** prolongs ventricular repolarization, often with prominent U waves. Prolongation of the QT interval is also seen with drugs that increase the duration of the ventricular action potential: class 1A antiarrhythmic agents and related drugs (e.g., quinidine, disopyramide, procainamide, tricyclic antidepressants, phenothiazines) and class III agents [e.g., amiodarone (Fig. 228-16), dofetilide, dronedarone, sotalol, ibutilide]. Marked QT prolongation, sometimes with deep, wide T-wave inversions, may occur with intracranial bleeds, particularly subarachnoid hemorrhage ("CVA T-wave" pattern) (Fig. 228-16). Systemic *hypothermia* also prolongs repolarization, usually with a distinctive convex elevation of the J point (Osborn wave). *Hypocalcemia* typically prolongs the QT interval (ST portion), whereas *hypercalcemia* shortens it **(Fig. 228-17)**. Digitalis glycosides also shorten the QT interval, often with a characteristic "scooping" of the ST–T-wave complex (*digitalis effect*).

Figure 228-15

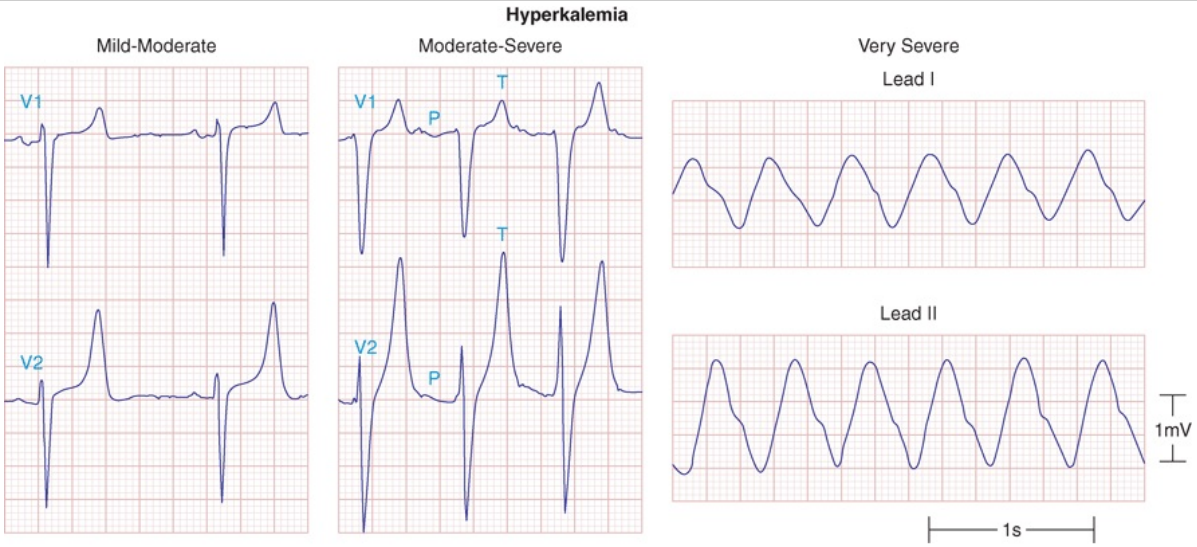

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The earliest ECG change with hyperkalemia is usually peaking ("tenting") of the T waves. With further increases in the serum potassium concentration, the QRS complexes widen, the P waves decrease in amplitude and may disappear, and finally a sine-wave pattern leads to asystole unless emergency therapy is given. *(After Goldberger.)*

Figure 228-16

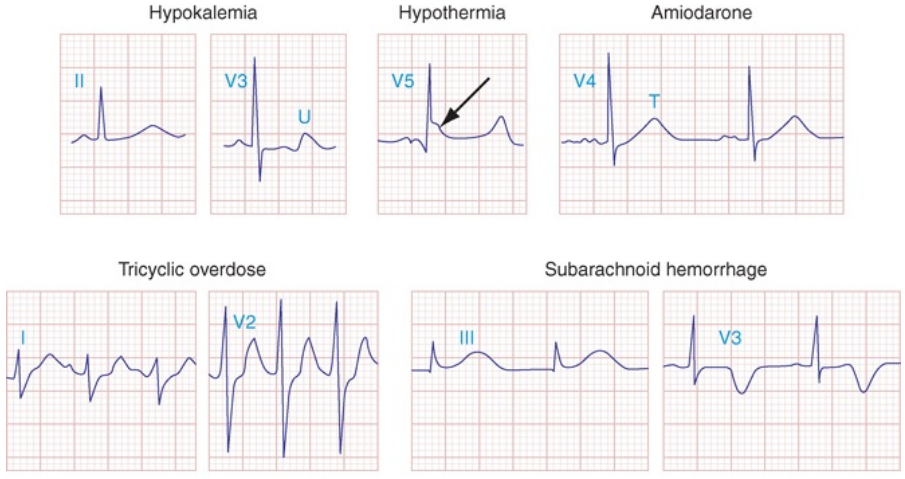

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A variety of metabolic derangements, drug effects, and other factors may prolong ventricular repolarization with QT prolongation or prominent U waves. Prominent repolarization prolongation, particularly if due to hypokalemia, inherited "channelopathies," or certain pharmacologic agents, indicates increased susceptibility to *torsades des pointes*–type ventricular tachycardia (Chap. 233). Marked systemic hypothermia is associated with a distinctive convex "hump" at the J point (Osborn wave, *arrow*) due to altered ventricular action potential characteristics. Note QRS and QT prolongation along with sinus tachycardia in the case of tricyclic antidepressant overdose.

Figure 228-17

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prolongation of the Q-T interval (ST-segment portion) is typical of hypocalcemia. Hypercalcemia may cause abbreviation of the ST segment and shortening of the QT interval.

Many other factors are associated with ECG changes, particularly alterations in ventricular repolarization. T-wave flattening, minimal T-wave inversions, or slight ST-segment depression ("nonspecific ST–T-wave changes") may occur with a variety of electrolyte and acid-base disturbances, a variety of infectious processes, central nervous system disorders, endocrine abnormalities, many drugs, ischemia, hypoxia, and virtually any type of cardiopulmonary abnormality. Although subtle ST–T-wave changes may be markers of ischemia, transient nonspecific repolarization changes may also occur after a meal or with postural (orthostatic) change, hyperventilation, or exercise in healthy individuals.

Electrical Alternans

Electrical alternans—a beat-to-beat alternation in one or more components of the ECG signal—is a common type of nonlinear cardiovascular response to a variety of hemodynamic and electrophysiologic perturbations. Total electrical alternans (P-QRS-T) with sinus tachycardia is a relatively specific sign of pericardial effusion, usually with cardiac tamponade. The mechanism relates to a periodic swinging motion of the heart in the effusion at a frequency exactly one-half the heart rate. Repolarization (ST-T or U wave) alternans is a sign of electrical instability and may precede ventricular tachyarrhythmias.

Clinical Interpretation of the ECG

Accurate analysis of ECGs requires thoroughness and care. The patient's age, gender, and clinical status should always be taken into account. Many mistakes in ECG interpretation are errors of omission. Therefore, a systematic approach is essential. The following 14 points should be analyzed carefully in every ECG: (1) standardization (calibration) and technical features (including lead placement and artifacts), (2) rhythm, (3) heart rate, (4) PR interval/AV conduction, (5) QRS interval, (6) QT/QT_C interval, (7) mean QRS electrical axis, (8) P waves, (9) QRS voltages, (10) precordial R-wave progression, (11) abnormal Q waves, (12) ST segments, (13) T waves, (14) U waves.

Only after analyzing all these points should the interpretation be formulated. Where appropriate, important clinical correlates or inferences should be mentioned. For example, sinus tachycardia with QRS and QT-(U) prolongation, especially in the context of changes in mental status, suggests tricyclic antidepressant overdose (Fig. 228-16). The triad of peaked T waves (hyperkalemia), a long QT due to ST-segment lengthening (hypocalcemia), and left ventricular hypertrophy (systemic hypertension) suggests chronic renal failure. Comparison with any previous ECGs is invaluable. **The diagnosis and management of specific cardiac arrhythmias and conduction disturbances are discussed in Chaps. 232 and 233**.

Computerized Electrocardiography

Computerized ECG systems are widely used for immediate retrieval of thousands of ECG records. Computer interpretation of ECGs still has major limitations. Incomplete or inaccurate readings are most likely with arrhythmias and complex abnormalities. Therefore, computerized interpretation (including measurements of basic ECG intervals) should not be accepted without careful clinician review.

FURTHER READINGS

Goldberger AL: *Clinical Electrocardiography: A Simplified Approach* , 8th ed. St. Louis, Mosby/Elsevier, in press

Kligfield P et al: Recommendations for the standardization and interpretation of the electrocardiogram: Part I. The electrocardiogram and its standardization. J Am Coll Cardiol 49:1109, 2007[PMID: 17349896]

Mirvis DM, Goldberger AL: Electrocardiography, in *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 9th ed, RW Bonow et al (eds). Philadelphia, Saunders, 2010

Surawicz B, Knilans TK: *Chou's Electrocardiography in Clinical Practice*, 6th ed. Philadelphia, Saunders, 2008

Wagner G et al: Recommendations for the standardization and interpretation of the electrocardiogram: Part VI. Acute myocardial ischemia. J Am Coll Cardiol 53:1003, 2009[PMID: 19281933]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF ELECTROCARDIOGRAPHY: INTRODUCTION

The electrocardiograms (ECGs) in this Atlas supplement those illustrated in Chap. 228. The interpretations emphasize findings of specific teaching value.

All of the figures are from ECG Wave-Maven, Copyright 2003, Beth Israel Deaconess Medical Center, *http://ecg.bidmc.harvard.edu*.

The abbreviations used in this chapter are as follows:

AF—atrial fibrillation

HCM—hypertrophic cardiomyopathy

LVH—left ventricular hypertrophy

MI—myocardial infarction

NSR—normal sinus rhythm

RBBB—right bundle branch block

RV—right ventricular

RVH—right ventricular hypertrophy

Myocardial Ischemia and Infarction

Figure e28-1

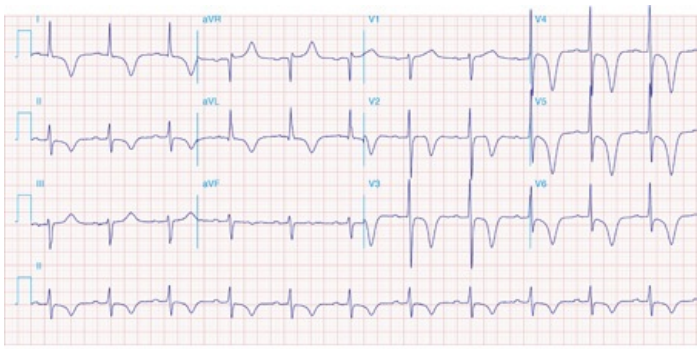

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anterior wall ischemia (deep T-wave inversions and ST-segment depressions in I, aVL, V_3–V_6) in a patient with **LVH** (increased voltage in V_2–V_5).

Figure e28-2

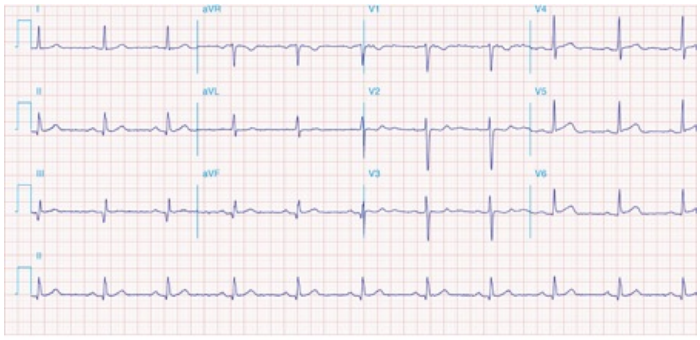

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute anterolateral wall ischemia with ST elevations in V_4–V_6. Probable prior inferior MI with Q waves in leads II, III, and aVF.

Figure e28-3

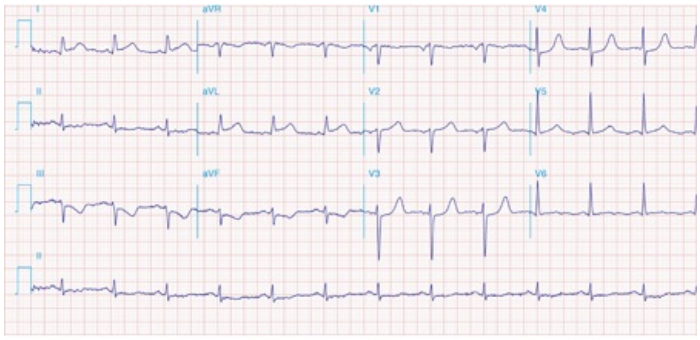

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute lateral ischemia with ST elevations in I and aVL with probable reciprocal ST depressions inferiorly (II, III, and aVF). Ischemic ST depressions also in V_3 and V_4. **Left atrial abnormality**.

Figure e28-4

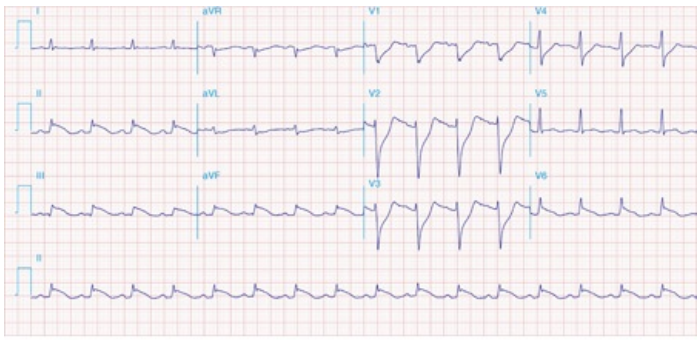

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus tachycardia. Marked ischemic ST-segment elevations in inferior limb leads (II, III, aVF) and laterally (V_6) suggestive of **acute inferolateral MI,** and prominent ST-segment depressions with upright T waves in V_1–V_4 are consistent with associated **acute posterior MI**.

Figure e28-5

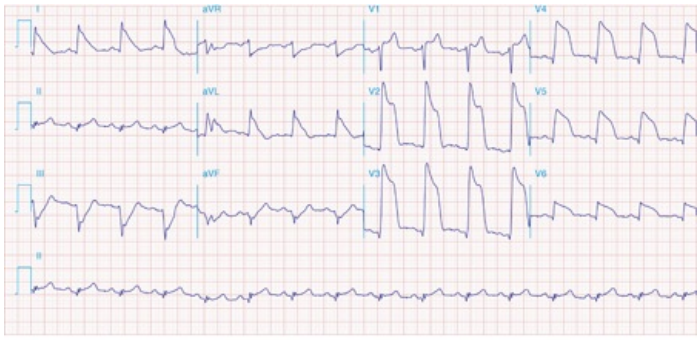

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute, extensive anterior MI with marked ST elevations in I, aVL, V_1–V_6 and small pathologic Q waves in V_3–V_6. Marked reciprocal ST-segment depressions in III and aVF.

Figure e28-6

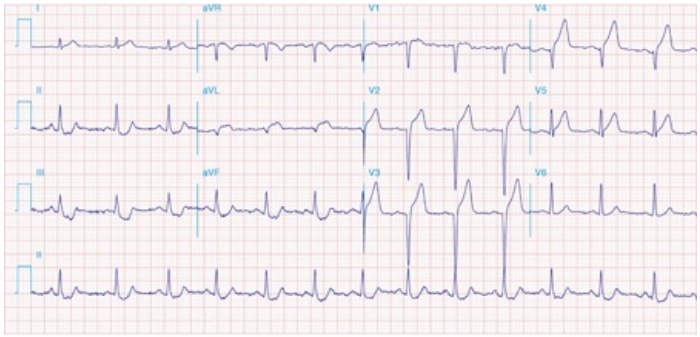

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute anterior wall MI with ST elevations and Q waves in V_1–V_4 and aVL and reciprocal inferior ST depressions.

Figure e28-7

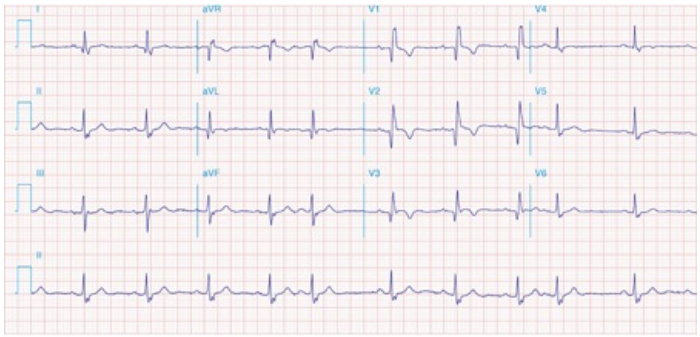

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR with premature atrial complexes. **RBBB;** pathologic Q waves and ST elevation due to **acute anterior/septal MI** in V_1–V_3.

Figure e28-8

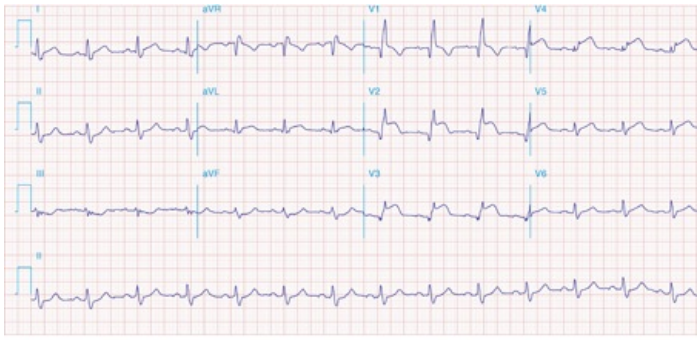

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute anteroseptal MI (Q waves and ST elevations in V_1–V_4) with **RBBB** (note terminal R waves in V_1).

Figure e28-9

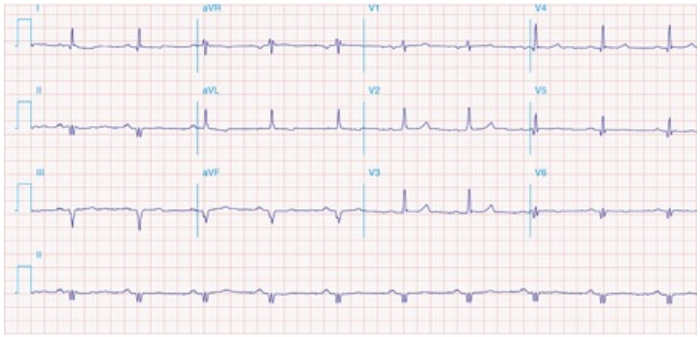

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Extensive prior MI involving inferior-posterior-lateral wall (Q waves in leads II, III, aVF, tall R waves in V_1, V_2, and Q waves in V_5, V_6). T-wave abnormalities in leads I and aVL, V_5, and V_6.

Figure e28-10

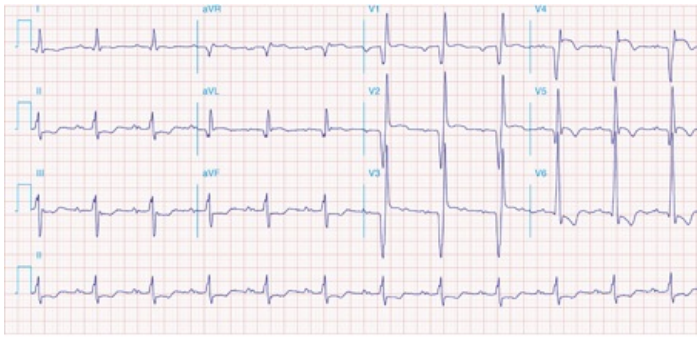

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR with PR prolongation ("1st degree AV block"), left atrial abnormality, LVH, and RBBB. Pathologic Q waves in V_1–V_5 and aVL with ST elevations (a chronic finding in this patient). Findings compatible with **prior anterolateral MI and LV aneurysm**.

Figure e28-11

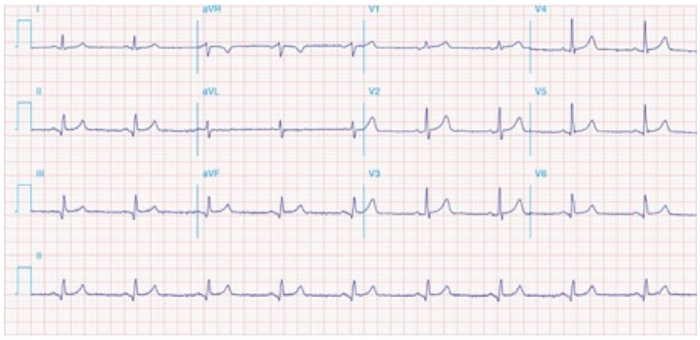

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prior inferior-posterior MI. Wide (0.04 s) Q waves in the inferior leads (II, III, aVF); broad R wave in V_1 (a Q wave "equivalent" here). Absence of right-axis deviation and the presence of upright T waves in V_1–V_2 are also against RVH.

Figure e28-12

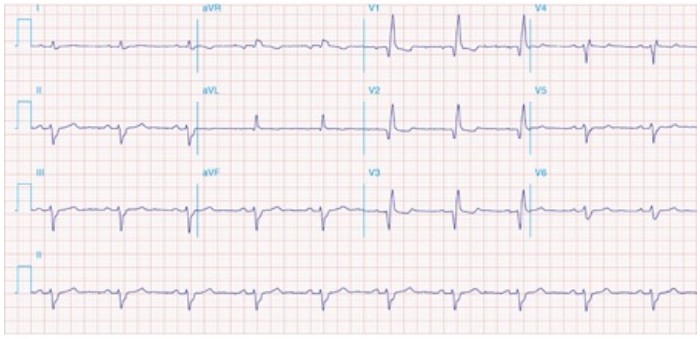

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR with RBBB (broad terminal R wave in V_1) and left anterior fascicular block (hemiblock) and pathologic anterior Q waves in V_1–V_3. Patient had **severe multivessel coronary artery disease,** with echocardiogram showing septal dyskinesis and apical akinesis.

Pericarditis

Figure e28-13

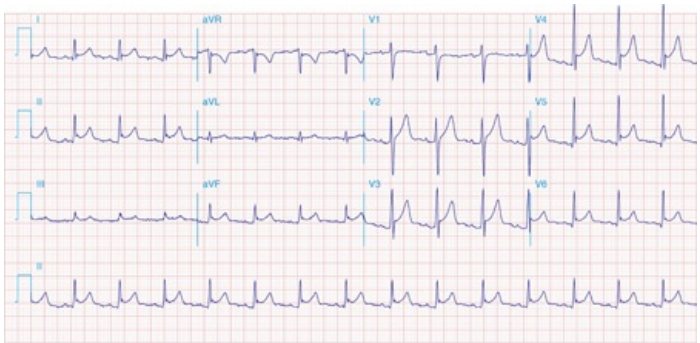

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute pericarditis with diffuse ST elevations in I, II, III, aVF, V_3–V_6, without T-wave inversions. Also note concomitant PR-segment elevation in aVR and PR depression in the inferolateral leads.

Figure e28-14

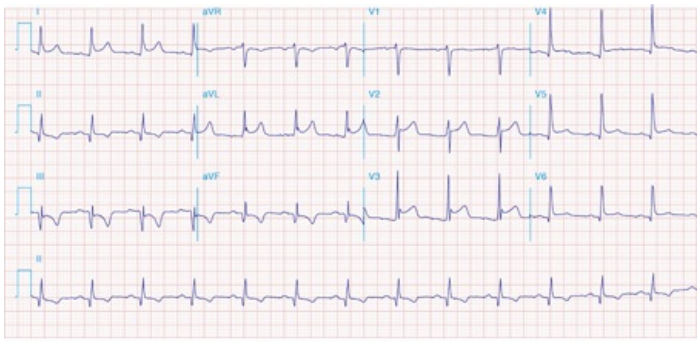

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus rhythm; diffuse ST elevations (I, II, aVL, aVF, V_2–V_6) with associated PR deviations (elevated PR in aVR; depressed in V_4–V_6); borderline low voltage. Q-wave and T-wave inversions in II, III, and aVF. Diagnosis: **acute pericarditis with inferior Q-wave MI**.

Valvular Heart Disease and Hypertrophic Cardiomyopathy

Figure e28-15

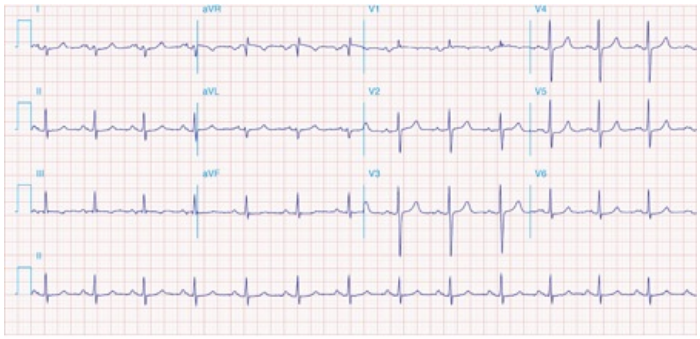

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR, prominent left atrial abnormality (see I, II, V_1), right-axis deviation and **RVH** (tall, relatively narrow R wave in V_1) in a patient with **mitral stenosis**.

Figure e28-16

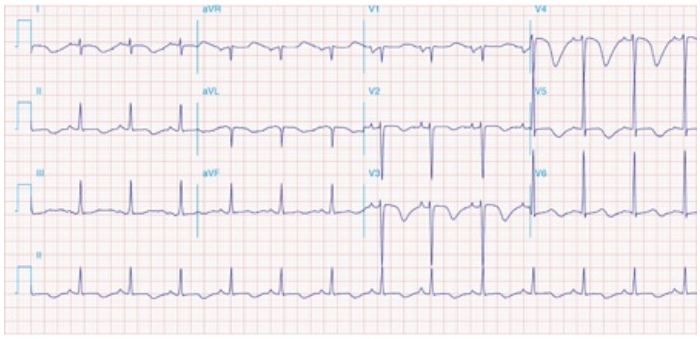

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR, left atrial abnormality, and LVH by voltage criteria with borderline right-axis deviation in a patient with **mixed mitral stenosis** (left atrial abnormality and right-axis deviation) and **mitral regurgitation** (LVH). Prominent precordial T-wave inversions and QT prolongation also present.

Figure e28-17

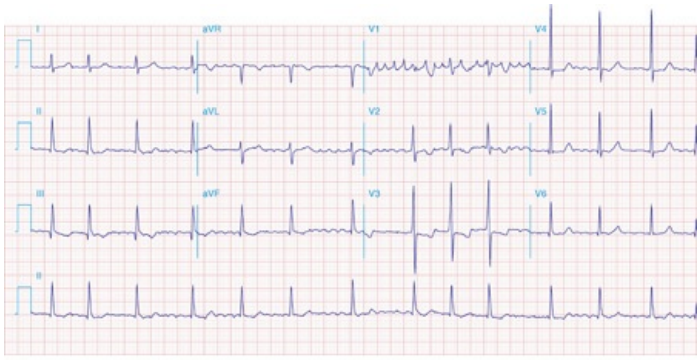

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coarse AF, tall R in V_2 with vertical QRS axis (positive R in aVF) indicating RVH. Tall R in V_4 may be due to concomitant LVH. Patient had **severe mitral stenosis with moderate mitral regurgitation**.

Figure e28-18

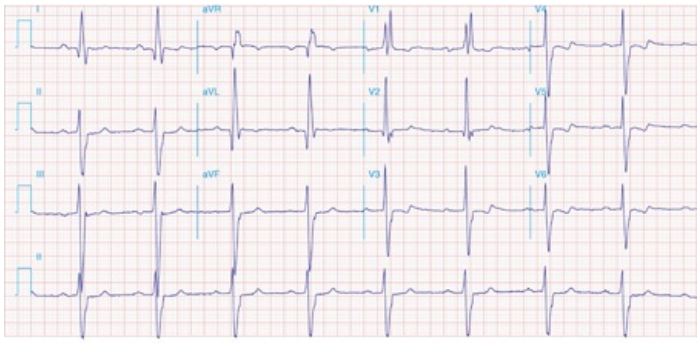

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR; first-degree A-V "block" (P-R prolongation); LVH (tall R in aVL); RBBB (wide multiphasic R wave in V1) and left anterior fascicular block in a patient with **HCM.** Deep Q waves in I and aVL are consistent with **septal hypertrophy**.

Figure e28-19

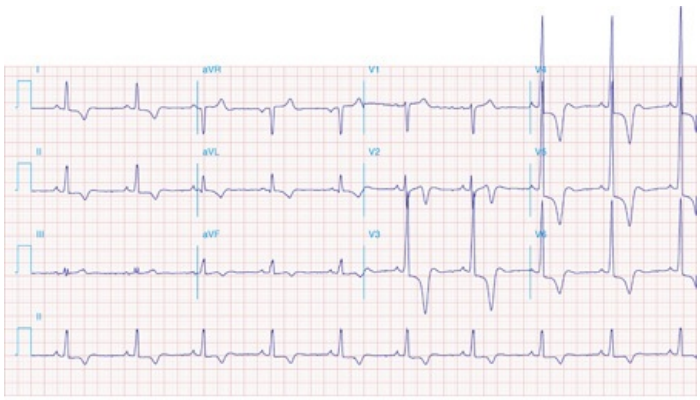

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

LVH with deep T-wave inversions in limb leads and precordial leads. Striking T-wave inversions in mid-precordial leads suggest **apical HCM** (Yamaguchi's syndrome).

Pulmonary Embolism and Chronic Pulmonary Hypertension

Figure e28-20

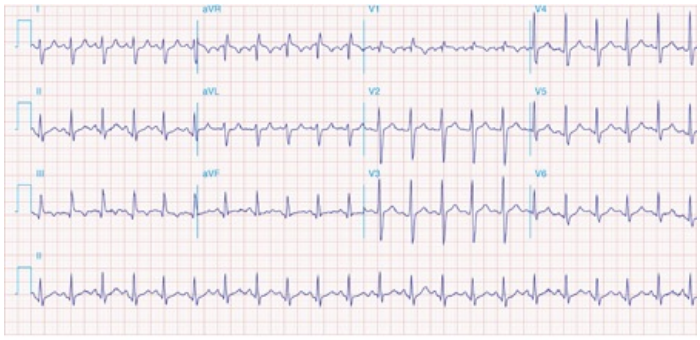

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus tachycardia with S1Q3T3 pattern (T-wave inversion in III), incomplete RBBB, and right precordial T-wave inversions consistent with acute RV overload in a patient with **pulmonary emboli**.

Figure e28-21

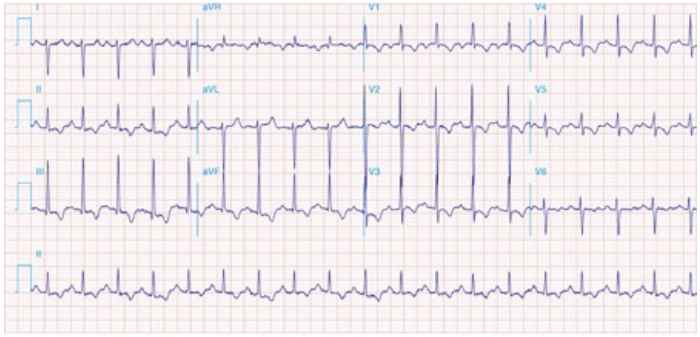

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus tachycardia, right-axis deviation, RVH with tall R in V_1 and deep S in V_6 and inverted T waves in II, III, aVF, and V_1–V_5 in a patient with **atrial septal defect and severe pulmonary hypertension**.

Figure e28-22

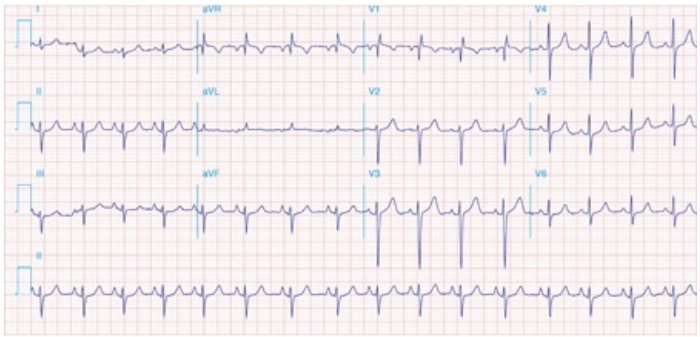

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Signs of right atrial/RV overload in a patient with **chronic obstructive lung disease:** (1) peaked P waves in II; (2) QR in V_1 with narrow QRS; (3) delayed precordial transition, with terminal S waves in V_5/V_6; (4) superior axis deviation with an S_1-S_2-S_3 pattern.

Figure e28-23

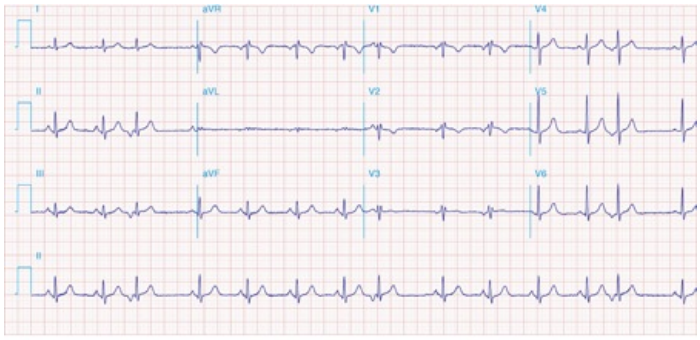

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

(1) Low voltage; (2) incomplete RBBB (rsr' in V_1–V_3); (3) borderline peaked P waves in lead II with vertical P-wave axis (probable right atrial overload); (4) slow R-wave progression in V_1–V_3; (5) prominent S waves in V_6; and (6) atrial premature beats. This combination is seen typically in **severe chronic obstructive lung disease**.

Electrolyte Disorders

Figure e28-24

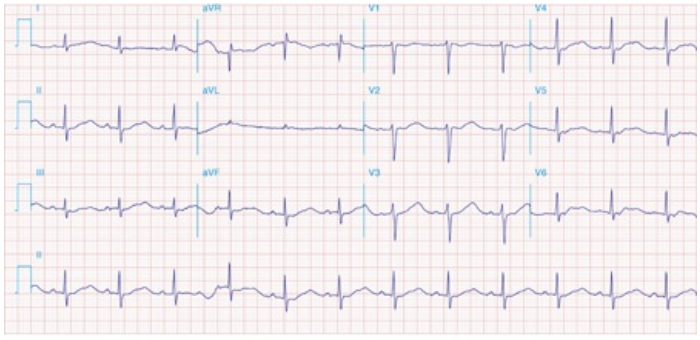

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prominent U waves (II, III, and V_4–V_6) with ventricular repolarization prolongation in a patient with **severe hypokalemia**.

Figure e28-25

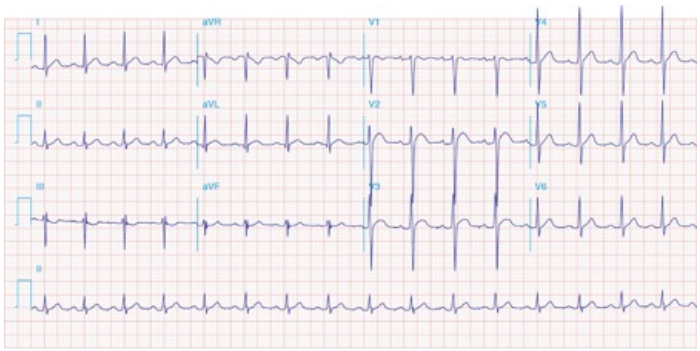

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Abbreviated ST segment such that the T wave looks like it takes off directly from QRS in some leads (I, V_4, aVL, and V_5) in a patient with severe **hypercalcemia**. Note also high takeoff of ST segment in V_2/V_3 simulating acute ischemia.

Figure e28-026

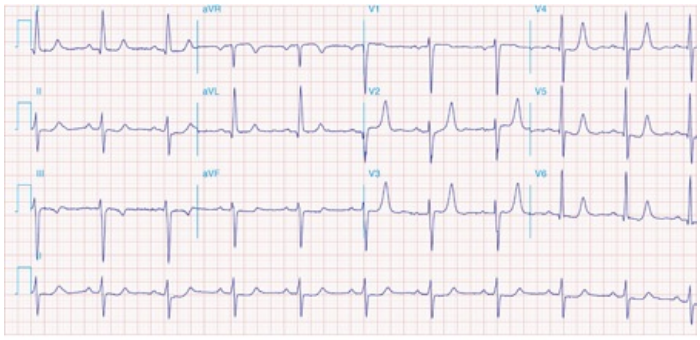

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR with LVH, left atrial abnormality, and tall peaked T waves in the precordial leads with inferolateral ST depressions (II, III, aVF, and V_6); left anterior fascicular block and borderline prolonged QT interval in a patient with **renal failure, hypertension, and hyperkalemia;** prolonged QT is secondary to **associated hypocalcemia**.

Miscellaneous

Figure e28-27

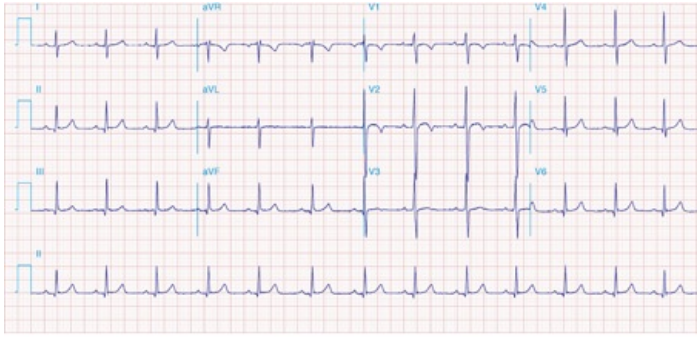

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal ECG in an 11-year-old male. T-wave inversions in V_1–V_2. Vertical QRS axis (+90°) and early precordial transition between V_2 and V_3 are **normal findings in children**.

Figure e28-28

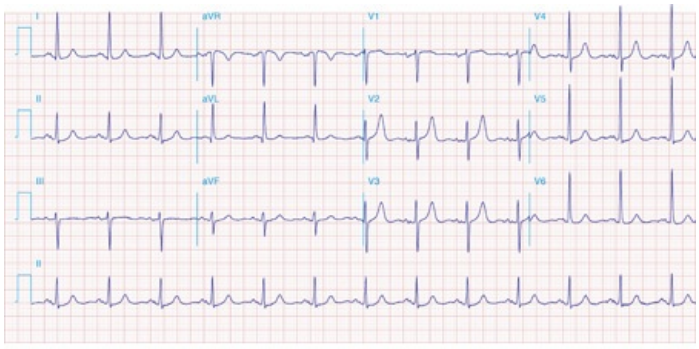

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left atrial abnormality and **LVH** in a patient with **long-standing hypertension**.

Figure e28-29

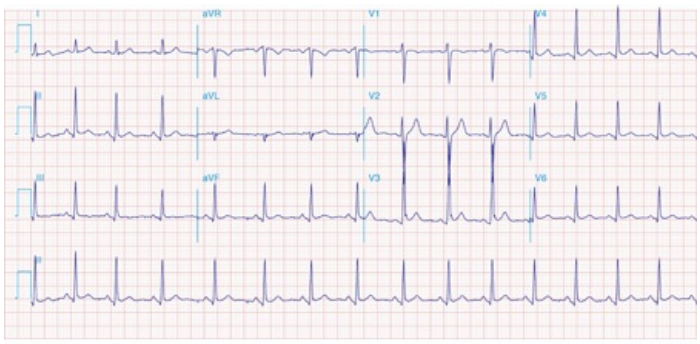

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal variant ST-segment elevations in a healthy 21-year-old male (commonly referred to as *benignearly repolarization pattern*). ST elevations exhibit upward concavity and are most apparent in V_3 and V_4, and less than 1 mm in the limb leads. Precordial QRS voltages are prominent, but within normal limits for a young adult. No evidence of left atrial abnormality or ST depression/T wave inversions to go along with **LVH**.

Figure e28-30

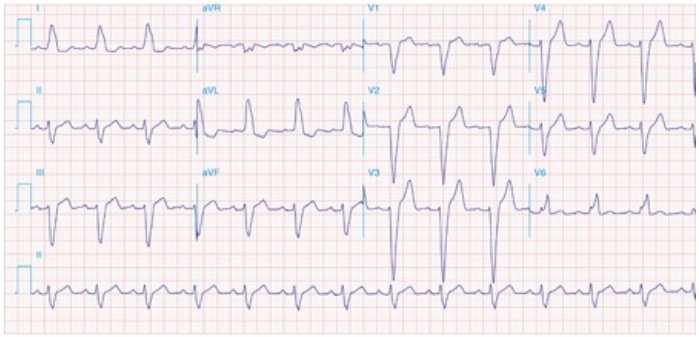

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR with first-degree AV "block" (PR interval = 0.24 s) and complete **left bundle branch block**.

Figure e28-31

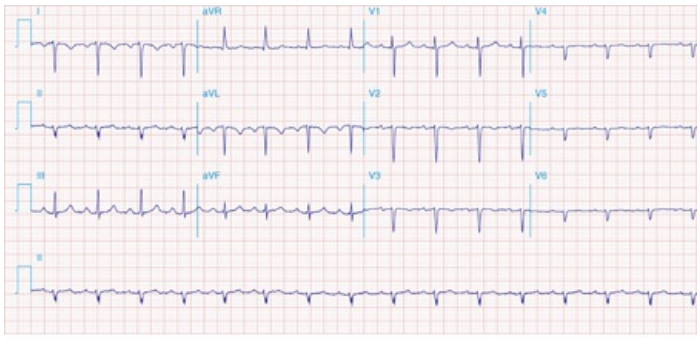

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dextrocardia with: (1) inverted P waves in I and aVL; (2) negative QRS complex and T wave in I; and (3) progressively decreasing voltage across the precordium.

Figure e28-32

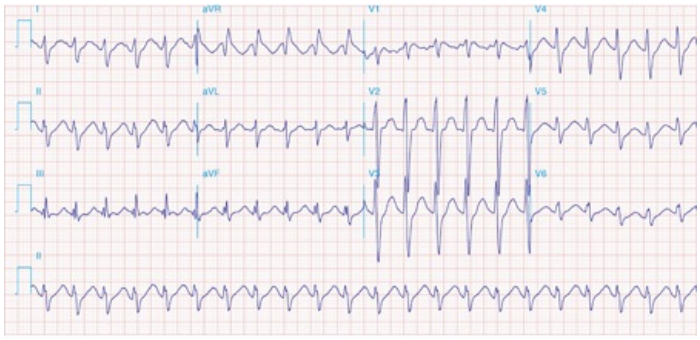

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus tachycardia; intraventricular conduction delay (IVCD) with a rightward QRS axis. QT interval is prolonged for the rate. The triad of sinus tachycardia, a wide QRS complex, and a long QT in appropriate clinical context suggests **tricyclic antidepressant overdose**. Terminal S wave (rS) in I, and terminal R wave (qR) in aVR are also noted as part of this IVCD variant.

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NONINVASIVE CARDIAC IMAGING: ECHOCARDIOGRAPHY, NUCLEAR CARDIOLOGY, AND MRI/CT IMAGING: INTRODUCTION

Cardiovascular imaging plays an essential role in the practice of cardiology. Two-dimensional (2D) echocardiography is able to visualize the heart directly in real time using ultrasound, providing instantaneous assessment of the myocardium, cardiac chambers, valves, pericardium, and great vessels. Doppler echocardiography measures the velocity of moving red blood cells and has become a noninvasive alternative to cardiac catheterization for assessment of hemodynamics. Transesophageal echocardiography (TEE) provides a unique window for high-resolution imaging of posterior structures of the heart, particularly the left atrium, mitral valve, and aorta. Nuclear cardiology uses radioactive tracers to provide assessment of myocardial perfusion and metabolism, along with ventricular function, and is applied primarily to the evaluation of patients with ischemic heart disease. Cardiac MRI and CT can delineate cardiac structure and function with high resolution. They are particularly useful in the examination of cardiac masses, the pericardium, the great vessels, and ventricular function and perfusion. Gadolinium enhancement during cardiac MRI adds information on myocardial perfusion. Detection of coronary calcification by CT as well as direct visualization of coronary arteries by CT angiography (CTA) may be useful in selected patients with suspected coronary artery disease (CAD). This chapter provides an overview of the basic concepts of these cardiac imaging modalities as well as the clinical indications for each procedure.

ECHOCARDIOGRAPHY

Two-Dimensional Echocardiography

BASIC PRINCIPLES

2D echocardiography uses the principle of ultrasound reflection off cardiac structures to produce images of the heart **(Table 229-1)**. For a transthoracic echocardiogram (TTE), the imaging is performed with a handheld transducer placed directly on the chest wall. In selected patients, a TEE may be performed, in which an ultrasound transducer is mounted on the tip of an endoscope placed in the esophagus and directed toward the cardiac structures.

Table 229–1. Clinical Uses of Echocardiography

Two-Dimensional Echocardiography
Cardiac chambers
Chamber size
Left ventricular hypertrophy
Regional wall motion abnormalities
Valve
Morphology and motion
Pericardium
Effusion
Tamponade
Masses
Great vessels
Stress Echocardiography
Two-dimensional
Myocardial ischemia Viable myocardium
Doppler
Valve disease
Doppler Echocardiography
Valve stenosis

Gradient
Valve area
Valve regurgitation
Semiquantitation
Intracardiac pressures
Volumetric flow
Diastolic filling
Intracardiac shunts
Transesophageal Echocardiography Inadequate transthoracic images
Aortic disease Infective endocarditis
Source of embolism
Valve prosthesis
Intraoperative

Current echocardiographic machines are portable and can be wheeled directly to the patient's bedside. Thus, a major advantage of echocardiography over other imaging modalities is the ability to obtain instantaneous images of the cardiac structures for immediate interpretation. Thus, echocardiography has become an ideal imaging modality for cardiac emergencies. A limitation of TTE is the inability to obtain high-quality images in all patients, especially those with a thick chest wall or severe lung disease, as ultrasound waves are poorly transmitted through lung parenchyma. Technology such as harmonic imaging and IV contrast agents (which traverse the pulmonary circulation) can be used to enhance endocardial borders in patients with poor acoustic windows.

CHAMBER SIZE AND FUNCTION

2D echocardiography is an ideal imaging modality for assessing left ventricular (LV) size and function **(Fig. 229-1)**. A qualitative assessment of the ventricular cavity and systolic function can be made directly from the 2D image by experienced observers. 2D echocardiography is useful in the diagnosis of LV hypertrophy and is the imaging modality of choice for the diagnosis of hypertrophic cardiomyopathy. Other chamber sizes are assessed by visual analysis, including the left atrium and right-sided chambers.

Figure 229-1

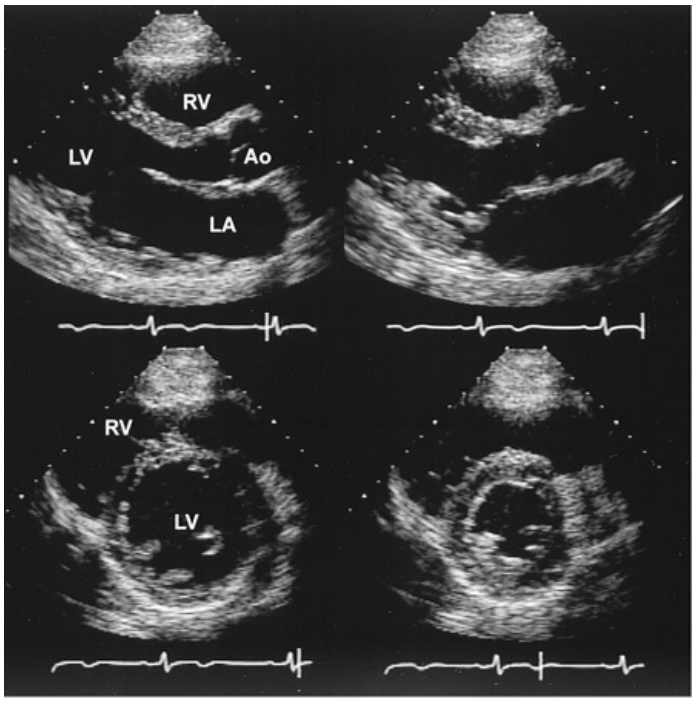

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Two-dimensional echocardiographic still-frame images from a normal patient with a normal heart. Upper: Parasternal long-axis view during systole and diastole (*left*) and systole (*right*). During systole, there is thickening of the myocardium and reduction in the size of the left ventricle (LV). The

valve leaflets are thin and open widely. Lower: Parasternal short-axis view during diastole (left) and systole (right) demonstrating a decrease in the left ventricular cavity size during systole as well as an increase in wall thickening. LA, left atrium; RV, right ventricle; Ao, aorta.

VALVE ABNORMALITIES

2D echocardiography is the "gold standard" for imaging valve morphology and motion. Leaflet thickness and mobility, valve calcification, and the appearance of subvalvular and supravalvular structures can be assessed. Valve stenosis is reliably diagnosed by the thickening and decreased mobility of the valve. 2D echocardiography is also the gold standard for the diagnosis of mitral stenosis, which produces typical tethering and diastolic doming, and the severity of the stenosis can be ascertained from a direct planimetry measurement of the mitral valve orifice. The presence and often the etiology of stenosis of the semilunar valves can be made by 2D echocardiography **(Fig. 229-2)**, but evaluation of the severity of the stenosis requires Doppler echocardiography (see below). The diagnosis of valvular regurgitation must be made by Doppler echocardiography, but 2D echocardiography is valuable for determining the etiology of the regurgitation, as well as its effects on ventricular dimensions, shape, and function.

Figure 229-2

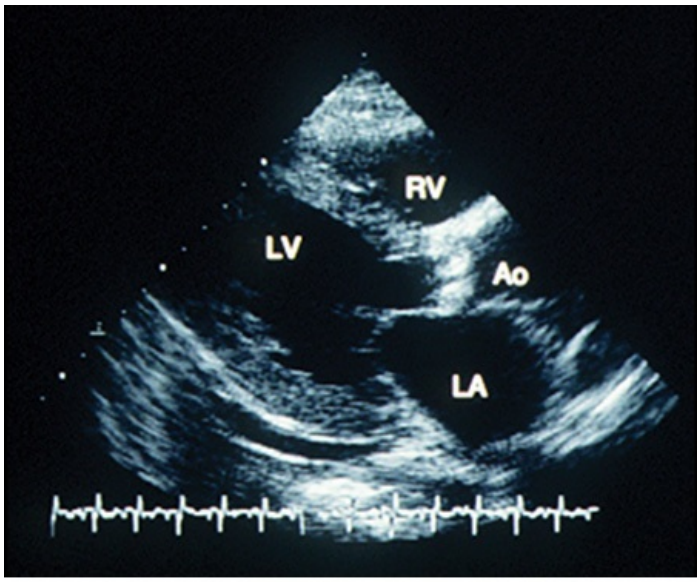

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Two-dimensional echocardiographic still-frame images from a patient with aortic stenosis. Parasternal long-axis view shows a heavily calcified aortic valve. RV, right ventricle; LV, left ventricle; AO, aorta; LA, left atrium.

PERICARDIAL DISEASE

2D echocardiography is the imaging modality of choice for the detection of pericardial effusion, which is easily visualized as a black echolucent ovoid structure surrounding the heart **(Fig. 229-3)**. In the hemodynamically unstable patient with pericardial tamponade, typical echo findings include a dilated inferior vena cava, right atrial collapse, and then right ventricular collapse. Echocardiographically guided pericardiocentesis has now become a standard of care.

Figure 229-3

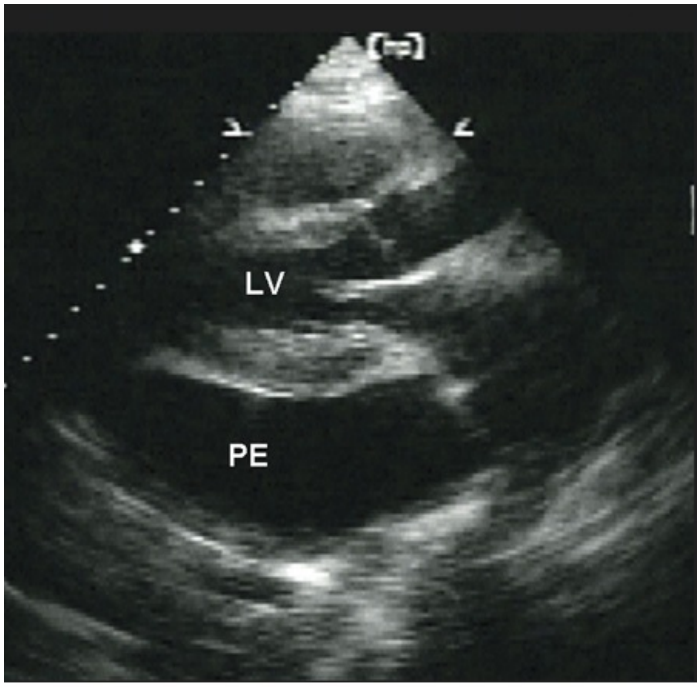

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Two-dimensional echocardiographic still-frame image of a patient with a pericardial effusion. Pericardial effusion (PE) is shown as black echo-free space surrounding the heart. LV, left ventricle.

INTRACARDIAC MASSES

Intracardiac masses can be visualized on 2D echocardiography, provided that image quality is adequate. Solid masses appear as echo-dense structures, which can be located inside the cardiac chambers or infiltrating into the myocardium or pericardium. LV thrombus appears as an echo-dense structure, usually in the apical region associated with regional wall motion abnormalities. The appearance and mobility of the thrombus are predictive of embolic events. Vegetations appear as mobile linear echo densities attached to valve leaflets. Atrial myxoma can be diagnosed by the appearance of a well-circumscribed mobile mass with attachments to the atrial septum **(Fig. 229-4)**. The high-resolution images provided by TEE may be required for further delineation of myocardial masses, especially those <1 cm in diameter.

Figure 229-4

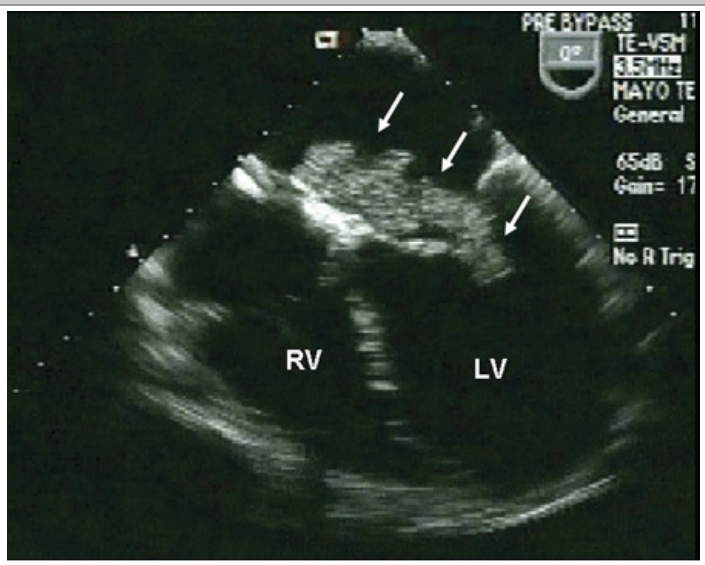

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transesophageal still-frame echocardiographic images of a patient with a left atrial myxoma. There is a large echo-dense mass in the left atrium, attached to the atrial septum. The mass moves across the mitral valve in diastole. LV, left ventricle; RV, right ventricle.

AORTIC DISEASE

2D echocardiography can provide extremely useful information on diseases of the aorta. The proximal ascending aorta, the arch, and the distal descending aorta can usually be visualized via the transthoracic approach. The definitive diagnosis of a suspected aortic dissection usually requires a TEE, which can rapidly provide high-resolution images of the proximal ascending and descending thoracic aorta **(Fig. 229-5)**.

Figure 229-5

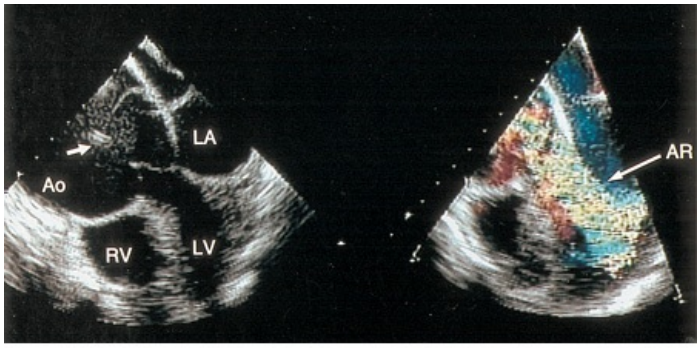

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transesophageal still-frame echocardiographic view of a patient with a dilated aorta, aortic dissection, and severe aortic regurgitation. The arrow points to the intimal flap that is seen in the dilated ascending aorta. *Left:* The long-axis apex-down view of the black-and-white two-dimensional image in diastole. *Right:* Color-flow imaging that demonstrates a large mosaic jet of aortic regurgitation. Ao, aorta; RV, right ventricle; AR, aortic regurgitation.

Doppler Echocardiography

BASIC PRINCIPLES

Doppler echocardiography uses ultrasound reflecting off moving red blood cells to measure the velocity of blood flow across valves, within cardiac chambers, and through the great vessels. Normal and abnormal blood flow patterns can be assessed noninvasively. Color-flow Doppler imaging displays the blood velocities in real time superimposed upon a 2D echocardiographic image. The different colors indicate the direction of blood flow (red toward and blue away from the transducer), with green superimposed when there is turbulent flow. Pulsed-wave Doppler measures the blood flow velocity in a specific location on the 2D echocardiographic image. Continuous-wave Doppler echocardiography can measure high velocities of blood flow directed along the line of the Doppler beam, such as occur in the presence of valve stenosis, valve regurgitation, or intracardiac shunts. These high velocities can be used to determine intracardiac pressure gradients by a modified Bernoulli equation:

$$\text{Pressure change} = 4 \text{ times } (\text{velocity})^2$$

Tissue Doppler echocardiography measures the velocity of myocardial motion. Myocardial velocities can be used to determine myocardial strain rate, which is a quantitative measure of regional myocardial contraction and relaxation.

VALVE GRADIENTS

In the presence of valvular stenosis, there is an increase in the velocity of blood flow across the stenotic valve. A continuous-wave Doppler can be used to determine the pressure gradient across the valve **(Fig. 229-6)**. A valve area can also be calculated from the Doppler velocities.

Figure 229-6

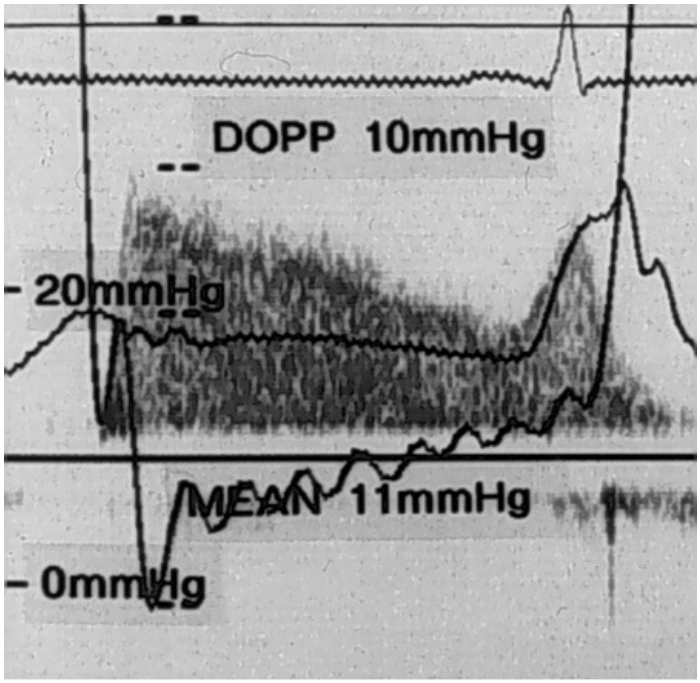

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Continuous-wave Doppler of mitral valve velocities in a patient with mitral stenosis. The mean gradient calculated from Doppler (DOPP) of 10 mmHg is similar to the mean gradient of 11 mmHg from simultaneous cardiac catheterization in this patient.

VALVULAR REGURGITATION

Valvular regurgitation is diagnosed by Doppler echocardiography when there is abnormal retrograde flow across the valve. Color-flow imaging is the Doppler method used most frequently to detect valve regurgitation by visualization of a high-velocity turbulent jet in the chamber proximal to the regurgitant valve **(Fig. 229-7)**. The size and extent of the color-flow jet into the receiving cardiac chamber provide a semiquantitative estimate of the severity of regurgitation.

Figure 229-7

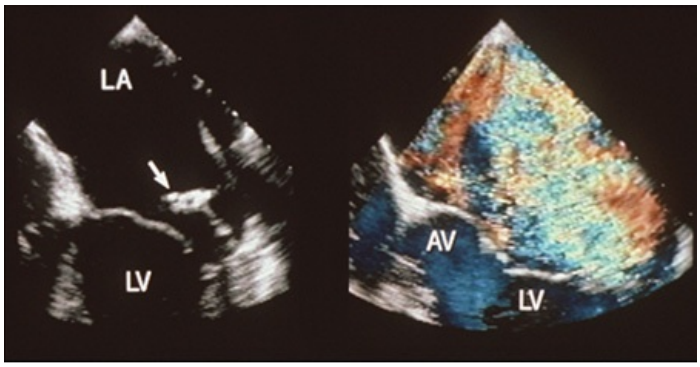

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left: Transesophageal echocardiographic view of a patient with severe mitral regurgitation due to a flail posterior leaflet. The arrow points to the portion of the posterior leaflet that is unsupported and moves into the left atrium during systole. *Right:* Color-flow imaging demonstrating a large mosaic jet of mitral regurgitation during systole. LA, left atrium; LV, left ventricle; AV, aortic valve.

INTRACARDIAC PRESSURES

These can be calculated from the peak continuous-wave Doppler signal of a regurgitant lesion, which reflects the pressure gradient between two cardiac chambers. This approach is commonly applied to a tricuspid regurgitant jet, from which the systolic pressure gradient between the right atrium and right ventricle can be calculated, yielding an accurate measurement of pulmonary artery systolic pressure **(Fig. 229-8)**.

Figure 229-8

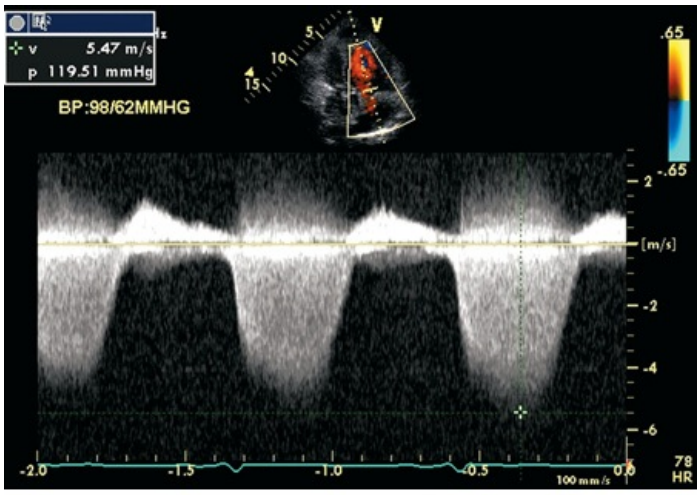

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Continuous-wave Doppler of tricuspid regurgitation in a patient with pulmonary hypertension. There is an increase in the velocity to 5.4 m/s. Using the modified Bernoulli equation, the peak pressure gradient between right ventricle and right atrium during systole is 116 mmHg. Assuming a right atrial pressure of 10 mmHg, the right ventricular systolic pressure is 126 mmHg. In the absence of right ventricular outflow tract obstruction, this indicates there is severe pulmonary hypertension with a pulmonary artery systolic pressure of 126 mmHg.

CARDIAC OUTPUT

Volume flow rates (or stroke volume and cardiac output) can be reliably measured noninvasively by Doppler echocardiography. Flow is calculated as the product of the cross-sectional area of the vessel or chamber through which blood moves and the velocity of blood flow as assessed by Doppler.

DIASTOLIC FILLING

Doppler echocardiography allows noninvasive evaluation of ventricular diastolic filling. The transmitral velocity curves reflect the relative pressure gradients between the left atrium and ventricle throughout diastole and are influenced by the rate of ventricular relaxation, the driving force across the valve, and the compliance of the ventricle. In the early phase of diastolic dysfunction there is primarily an impairment of LV relaxation, with reduced early transmitral flow and a compensatory increase in flow during atrial contraction **(Fig. 229-9)**. As disease progresses and ventricular compliance declines, left atrial pressure rises, resulting in a higher early transmitral velocity and shortening of the deceleration of flow in early diastole. Analysis of Doppler tissue velocities of annular motion and myocardial strain provides further information concerning the diastolic properties of the heart.

Figure 229-9

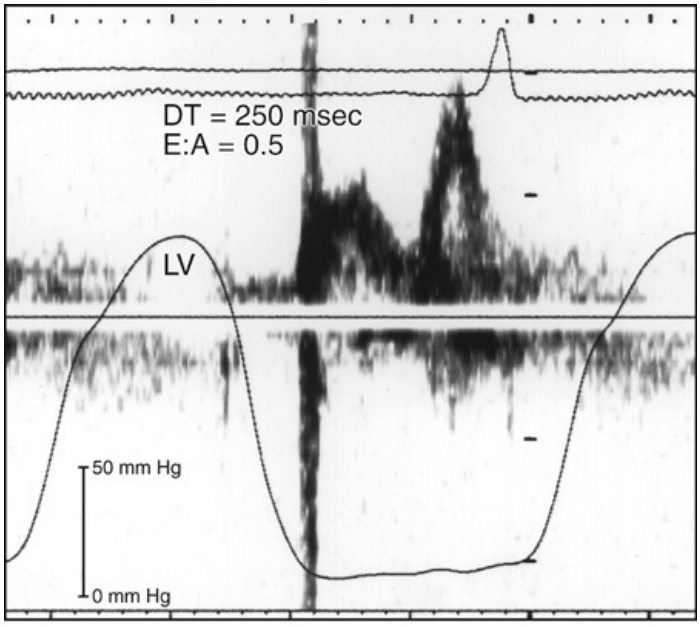

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

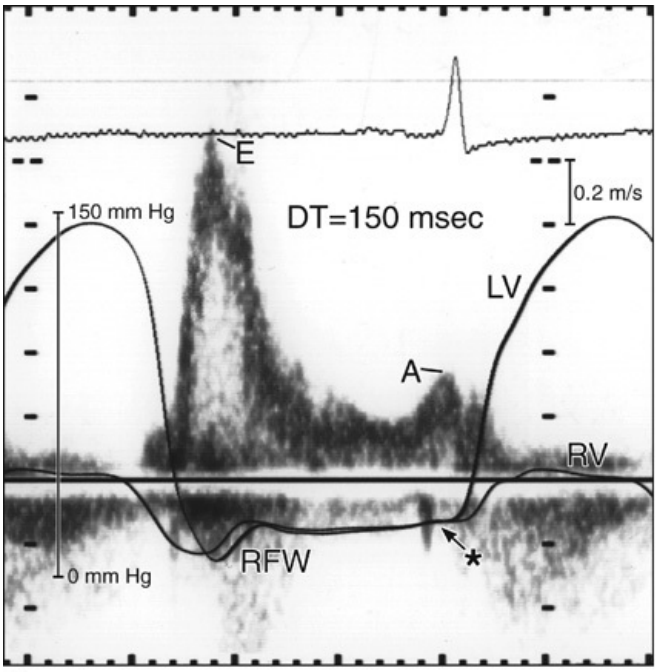

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

High-fidelity left ventricular (LV) pressure curves superimposed on a mitral inflow velocity curve obtained by Doppler echocardiography. The ratio of early and late diastolic flows is termed the E:A ratio. The deceleration time (DT) measures the rate of decline of early velocity and reflects the effective operative compliance of the left ventricle. *Left:* In early stages of diastolic dysfunction, there is an abnormality of relaxation. There is a decrease in the early diastolic filling and an increase with filling at atrial contraction, resulting in a low E:A ratio of 0.5, with deceleration time (DT) of 250 ms. In this instance, the LV diastolic pressure is low at 6 mmHg. *Right:* As diastolic dysfunction progresses, there is a restriction to filling, in which there is a high early diastolic velocity and low velocity at atrial contraction resulting in a high E:A ratio of 3.0, with DT of 150 ms. In this instance, the LV diastolic pressure is markedly elevated to 34 mmHg.

CONGENITAL HEART DISEASE

2D and Doppler echocardiography have been useful in the evaluation of patients with congenital heart disease. Congenital stenotic or regurgitant valve lesions can be assessed. The detection of intracardiac shunts is possible by 2D and Doppler echocardiography. Patency of surgical shunts and conduits can also be evaluated.

Stress Echocardiography

2D and Doppler echocardiography are usually performed with the patient in the resting state. Further information can be obtained by reimaging during either exercise or pharmacologic stress. The primary indications for stress echocardiography are to confirm the suspicion of ischemic heart disease and determine the extent of ischemia.

A decrease in systolic contraction of an ischemic area (segment) of myocardium, termed a regional wall motion abnormality, occurs before symptoms or electrocardiographic changes **(Fig. 229-10)**. New regional wall motion abnormalities, a decline in ejection fraction, and an increase in end-systolic volume with stress are all indicators of myocardial ischemia. Exercise stress testing is usually done with exercise protocols using either upright treadmill or bicycle exercise. In patients who are not able to exercise, pharmacologic testing can be performed by infusion of dobutamine to increase myocardial oxygen demand. Dobutamine echocardiography has also been used to assess myocardial viability in patients with poor systolic function and concomitant CAD; when used for this purpose, dobutamine is administered at lower doses than standard pharmacologic stress doses.

Figure 229-10

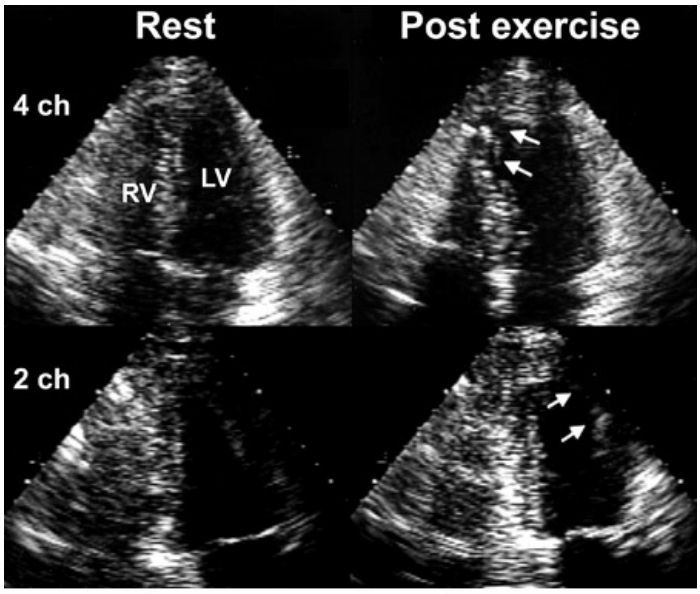

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Systolic still-frame two-dimensional echocardiographic images of a patient undergoing a stress echocardiogram. During rest (*left*), there is contraction of all segments of the myocardium. During exercise (*right*), there are regional wall motion abnormalities in the anterior and anteroapical segments (*arrows*). 4ch = four-chamber view, 2ch = two-chamber views, LV = left ventricle, RV = right ventricle.

Doppler echocardiography can be used at rest and during exercise in patients with valvular heart disease to determine the hemodynamic response of valve gradients and pulmonary pressures **(Fig. 229-11)**. In patients with low-output, low-gradient aortic stenosis, the response of the gradient to dobutamine stimulation is of diagnostic and therapeutic value.

Figure 229-11

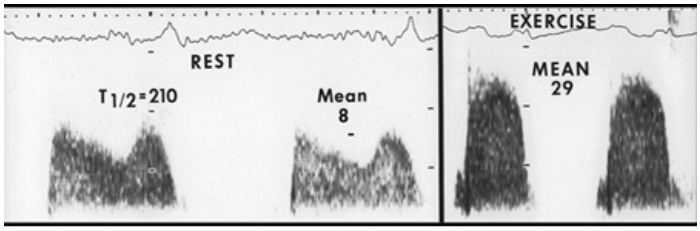

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Continuous-wave Doppler echocardiogram across the mitral valve of a patient with mitral stenosis. In

the resting state (*left*), there is a mean gradient of 8 mmHg. During exercise (*right*), the mean gradient rises to 29 mmHg, indicating a hemodynamically significant mitral stenosis.

Transesophageal Echocardiography

When limited information is obtained from a TTE due to poor imaging windows, TEE can be useful. Diseases of the aorta, such as aortic dissection, can be readily diagnosed by TEE. Defining the source of embolism is a common indication for TEE, as abnormalities such as atrial thrombi, patent foramen ovale, and aortic plaques can be detected. Other masses, particularly those in the atria, can be visualized. The presence of vegetations for the diagnosis of infective endocarditis and its complications can be assessed by TEE. This technique has been used before cardioversion in patients with atrial fibrillation to rule out a thrombus in the left atrium or left atrial appendage.

NUCLEAR CARDIOLOGY

Basic Principles of Nuclear Cardiology

Nuclear (or radionuclide) imaging requires intravenous administration of radiopharmaceuticals (isotopes or tracers). Once injected, the isotope traces physiologic processes and undergoes uptake in specific organs. During this process, radiation is emitted in the form of photons, generally gamma rays, generated during radioactive decay when the nucleus of an isotope changes from one energy level to a lower one. A special camera detects these photons and creates images via a computer interface. The two most commonly used technologies in clinical nuclear cardiology are single- photon emission computed tomography (SPECT) and positron emission tomography (PET). These technologies differ in instrumentation, acquisition, resolution, and nuclides used.

Clinical Applications

ASSESSMENT OF MYOCARDIAL PERFUSION AND CORONARY ARTERY DISEASE

Nuclear myocardial perfusion imaging (MPI) using SPECT and more recently PET has an established role in the evaluation and management of patients with known or suspected coronary artery disease (CAD). Both SPECT and PET MRI require the injection of isotopes at rest and during stress to produce images of regional myocardial uptake proportional to regional blood flow. Normally, myocardial blood flow can be increased up to fivefold above the resting state to meet the increased myocardial oxygen demand during stress. In the presence of a fixed coronary stenosis, the inability to increase myocardial perfusion in the territory supplied by the stenosis creates a flow differential and inhomogeneous myocardial tracer uptake. In patients unable to exercise, pharmacologic agents are used to increase blood flow and create similar inhomogeneities.

The most commonly used SPECT perfusion tracers are thallium-201 (^{201}Tl) and technetium-99m (^{99m}Tc) labeled isonitriles. ^{99m}Tc isonitriles have higher photon energies and shorter physical half-lives than ^{201}Tl, permitting injection of higher doses with less radiation exposure while concurrently producing higher-quality images. The FDA-approved PET tracers are rubidium-82 (^{82}Rb) and ^{13}N ammonia ($^{13}NH_3$) for high-dose administration and shorter imaging protocols.

Both SPECT and PET myocardial perfusion images are commonly interpreted by visual analysis, which may be supplemented with quantitative software. Normal myocardial perfusion images demonstrate uniform tracer uptake throughout the LV myocardium **(Fig. 229-12)**. In contrast, regions with reduced myocardial blood flow demonstrate varying degrees of reduced tracer uptake **(Fig. 229-13)**, which can be graded on a semiquantitative scale. Reduced tracer uptake in a myocardial region on both resting and stress images is called a fixed defect and is consistent with infarction. Reduced tracer uptake on the stress image with relatively preserved or improved uptake on the rest image is called a reversible defect and indicates ischemia. PET has the ability to quantify myocardial blood flow and flow reserve in absolute terms.

Figure 229-12

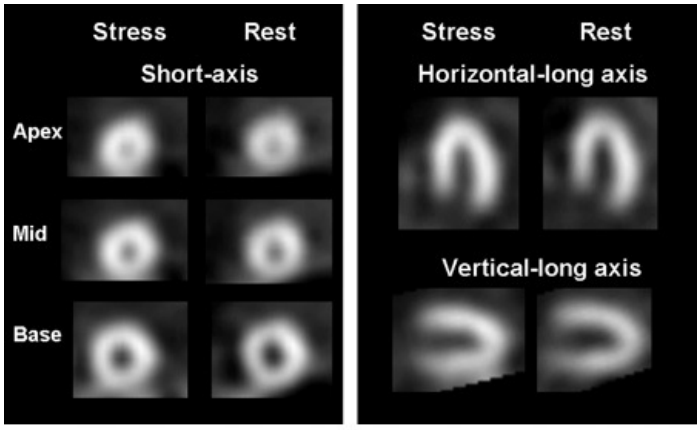

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Exercise technetium-99m sestamibi images in a 65-year-old man with atypical angina. Images are shown in three standard views; stress (*left*) and rest (*right*) in each panel. There is uniform tracer uptake throughout the left ventricular myocardium at rest and peak stress in all three views.

Figure 229-13

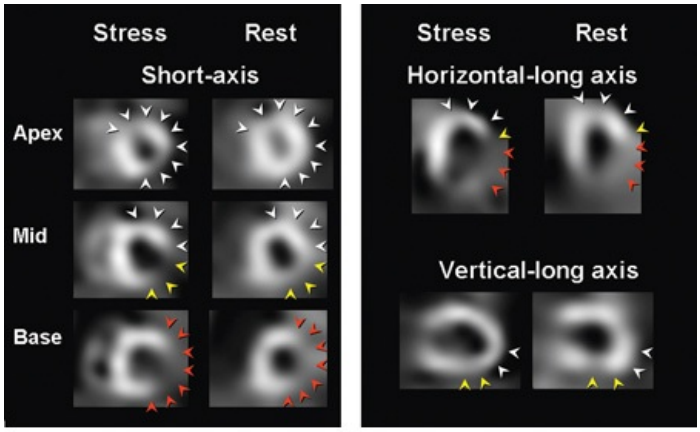

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Exercise technetium-99m sestamibi and rest thallium-201 images in a 72-year-old woman with typical angina. Images are shown in three standard views, with stress (*left*) and rest (*right*) in each panel. Stress images demonstrate reduced tracer uptake in the apical, mid-anterior, mid-lateral, and mid-inferior regions (*white arrowheads*) with normal or near-normal tracer uptake in the corresponding regions on the rest images (white arrowheads), signifying a reversible defect consistent with ischemia. The lack of complete normalization (or reversibility) of tracer uptake on the rest images at the mid-inferior and mid-lateral region represents associated infarction in that area (*yellow arrowheads*). On both stress and rest images, the basal inferior and basal lateral regions exhibit severely reduced tracer uptake, signifying a fixed defect consistent with infarction (*red arrowheads*). Subsequent invasive coronary angiography demonstrated severe stenosis of the mid left anterior descending coronary artery and occlusion of the left circumflex coronary artery with collaterals.

For the diagnosis of angiographically significant CAD, SPECT using ^{201}T1 and ^{99m}Tc isonitriles and either exercise or pharmacologic stress has an average sensitivity of 87% and specificity of 73%. In comparison, PET MPI has higher accuracy (average sensitivity 90%; specificity of 89%). The robust methods for attenuation correction with PET improve the specificity, particularly in obese populations and women, while the superior resolution and higher extraction fraction of PET tracers increase the sensitivity **(Fig. 229-14)**. PET has not been as widely used as SPECT due to decreased availability and less local experience, but PET scanners are becoming more widely available (**Table 229-2**).

Figure 229-14

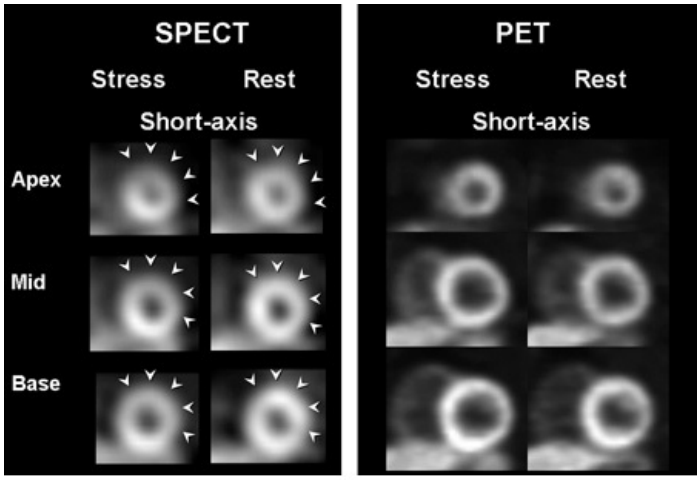

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

SPECT and PET images in a 67-year-old woman with atypical angina. Images are shown in short-axis views, with stress (*left*) and rest (*right*) in each panel. Shifting breast position between the rest and stress SPECT acquisitions produced an apparent reversible apical, anterior, and anterolateral attenuation artifact (*arrowheads*) resembling ischemia. With PET and its built-in attenuation correction in the same patient, the defect was not present. SPECT, single-photon emission computed tomography; PET, positron emission tomography.

Table 229–2. Relative Advantages and Disadvantages of SPECT and PET

SPECT
Thallium-201
Lower radiopharmaceutical cost
Measurement of increased pulmonary uptake
Less hepatobiliary and bowel uptake
Detection of resting ischemia (hibernating myocardium)
Longer physical half-life of tracers (limiting dose administration)
Lower energy level
Technetium-99m isonitriles
Better image quality
Ventricular function assessment (gated SPECT)
Shorter imaging time
Shorter imaging protocols (patient/scheduling convenience)
Acute imaging in myocardial infarction and unstable angina
Superior quantification
PET
Robust attenuation correction
Short physical half-life of tracers
Best image quality (particularly in obese patients and women)
Shorter imaging time (particularly for rubidium-82)
Very short imaging protocols (particularly for rubidium-82)
More complex imaging protocols (particularly viability assessment)
Detection of viability
Absolute quantification
High diagnostic accuracy
Limited prognostic studies
N-13 ammonia requires on-site cyclotron
Rubidium-82 generally requires costly commercial generator
Lower radiation exposure, particularly for N-13 ammonia
High Risk Perfusion Imaging Features
Severe resting or exercise LV systolic dysfunction (EF<35%)

Stress-induced large perfusion defect (especially if anterior)
Stress-induced multiple perfusion defects of moderate size
Large, fixed perfusion defect with LV dilatation
Transient (poststress) LV dilatation Increased lung uptake (thallium)

Abbreviatons: EF, ejection fraction; LV, left ventricle; PET, positron emission tomography; SPECT, single photon emission computed tomography.

Both SPECT and PET MPI have powerful prognostic value. In patients with normal SPECT MPI results, the annual rate of cardiac death or myocardial infarction is generally very low (<0.7%). Annual death/event rates increase with the extent and severity of imaging abnormalities and are generally about 3% in those with mild to moderate abnormalities and about 7% in those with severe abnormalities; rates are higher in specific populations such as diabetics and those with high-risk exercise treadmill results. High-risk SPECT MPI findings include severe resting or poststress LV systolic dysfunction, large or multiple stress-induced defects, or a large fixed defect with LV dilation or increased ^{201}Tl lung uptake. The incremental prognostic value of SPECT MPI has been established in many clinical settings, including populations with known CAD, prior myocardial infarction and/or revascularization, and acute chest pain in the emergency department.

ASSESSMENT OF MYOCARDIAL METABOLISM AND VIABILITY

PET has traditionally been regarded as the gold standard technique for the assessment of myocardial viability. The positron-emitting tracer F-18 fluorodeoxyglucose (FDG) assesses myocardial glucose metabolism and is an indicator of myocardial viability. Because uptake is heterogeneous in normal myocardium in the fasting state, oral glucose loading or a combination of insulin and glucose infusions is used to enhance myocardial uptake. With reduced myocardial blood flow and ischemia, substrate utilization switches from fatty acids and lactate toward glucose, leading to enhanced myocardial FDG uptake. This pattern of enhanced FDG uptake in regions of decreased perfusion (termed flow/metabolism "mismatch") identifies areas of ischemic or hibernating myocardium that are likely to improve in function after revascularization **(Fig. 229-15)**. This mismatch has a sensitivity and specificity of 92% and 63%, respectively, for regional contractile recovery after revascularization. The SPECT radiopharmaceuticals, ^{201}Tl and ^{99m}Tc isonitriles, require an intact (viable) cell membrane for uptake and also provide an assessment of myocardial viability in addition to perfusion. However, PET identifies ischemic or hibernating myocardium in 10–20% of regions otherwise classified as fibrotic (infarcted) by SPECT perfusion tracers. Patients with ischemic heart failure, who have viable myocardium identified by PET or SPECT and undergo revascularization, have a better survival than those who do not have viable myocardium or do not undergo revascularization.

Figure 229-15

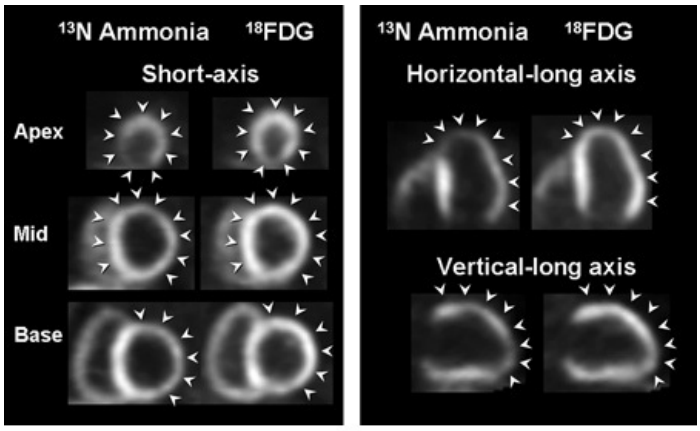

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

PET viability study in a 63-year-old woman with heart failure, severe LV systolic dysfunction, and severe coronary artery disease. Images are shown in three standard views, with perfusion (*left*) and glucose metabolism (*right*) in each panel. The N-13 ammonia images show a very large apical, septal, anterior, and lateral perfusion defect (arrowheads), but F-18 fluorodeoxyglucose (18FDG) images demonstrate relatively preserved glucose uptake in the corresponding segments (*arrowheads*). This PET perfusion-metabolism mismatch is consistent with hibernating myocardium. The patient underwent coronary artery bypass grafting surgery with improvement in left ventricular systolic function (ejection fraction increased from 26% pre- to 45% postoperatively). All regions identified as viable recovered contractile function after revascularization.

ASSESSMENT OF VENTRICULAR FUNCTION

In addition to perfusion and metabolic information, LV systolic function and volumes are now routinely obtained with gated SPECT and PET acquisitions, as long as the heart rate is relatively constant. An automated technique determines the endocardial borders of the LV cavity, and a geometric model is used to calculate the LVEF and volumes with a high level of reproducibility. Regional wall motion can also be assessed by visual examination. The combined variables of perfusion and function are more effective in risk stratification than either alone.

Another established but less widely available nuclear technique for assessing LV function and volumes is equilibrium radionuclide angiography RNA, also known as multiple-gated blood pool acquisition (MUGA). This technique involves imaging of ^{99m}Tc-labeled albumin or red cells that are uniformly distributed throughout the blood volume. LV volumes throughout the cardiac cycle are calculated from a time-activity curve generated using regions of interest.

Innovations in hybrid imaging technology, especially PET/CT and SPECT/CT, are occurring rapidly and contribute to their emerging role in the combined assessments of anatomy and physiology in patients with suspected or known CAD. The diagnostic literature is evolving for these hybrid technologies but radiation exposure is a concern and large-scale clinical trials are still needed to validate their clinical applications, determine their prognostic value, and address their cost-effectiveness and appropriateness.

MRI AND CT IMAGING

Magnetic Resonance Imaging

BASIC PRINCIPLES

MRI is a technique based on the magnetic properties of hydrogen nuclei. In the presence of a large magnetic field, nuclear spin transitions from the ground state to excited states can be induced by an electric field, and as the nuclei relax and return to their ground state, they release energy in the form of electromagnetic radiation that is detected and processed into an image. Although the large vascular vessels can be visualized on MRI without contrast agents, gadolinium is frequently employed as a contrast agent to produce magnetic resonance angiograms (MRAs). Contrast agents also provide enhanced soft tissue contrast as well as the opportunity to obtain rapid angiographic images during the first pass of contrast through the vascular system.

Cardiac MRI is challenging because of the rapid motion of the heart and coronary arteries. However, both static and cine images can usually now be obtained using electrocardiographic triggering, often within short breath-holds of 10–15 s. Cine images can be acquired in any plane with excellent blood-myocardial contrast. These images can be used to quantify accurately ejection fraction, end-systolic and end-diastolic volumes, and cardiac mass with high accuracy, reliability, and reproducibility, and without the need for ionizing radiation.

CLINICAL UTILITY

The multiplanar capabilities of MRI, coupled with excellent contrast and spatial resolution, provide superb images of the myocardium and great vessels. MRI is of great value in defining anatomic relationships in patients with complex congenital heart disease (**Fig. 229-16**) and cardiomyopathies **(Fig. 229-17)**. Cardiac masses can be characterized and distinguished from thrombus **(Fig. 229-18)**. In addition to defining their relationship to normal anatomic structures, MRI can determine whether a mediastinal or pulmonary mass has invaded the pericardium or heart. The entire pericardium can be visualized in multiple planes, and MRI has proved useful in characterizing pericardial effusions, pericardial thickening, and inflammation. Specialized pulse sequences can measure the velocity of blood in each pixel of the image, so that flow across valves and within blood vessels may be determined with accuracy, thereby aiding in the evaluation of valvular disease and intracardiac shunts.

Figure 229-16

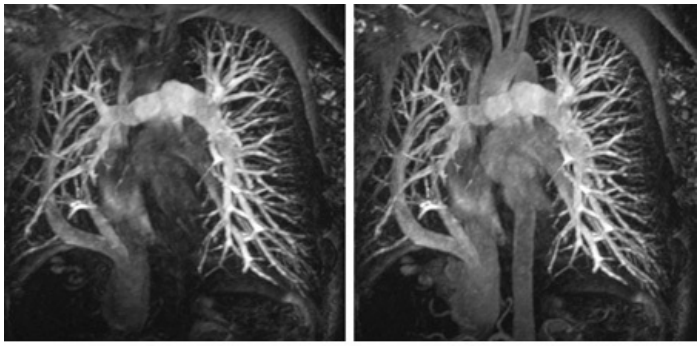

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRA scan of a patient with partial anomalous pulmonary venous drainage of the right lung into the inferior vena cava (scimitar syndrome). MRA is able to define the abnormal anatomic relationships of cardiac structures and great vessels in patients with congenital heart disease.

Figure 229-17

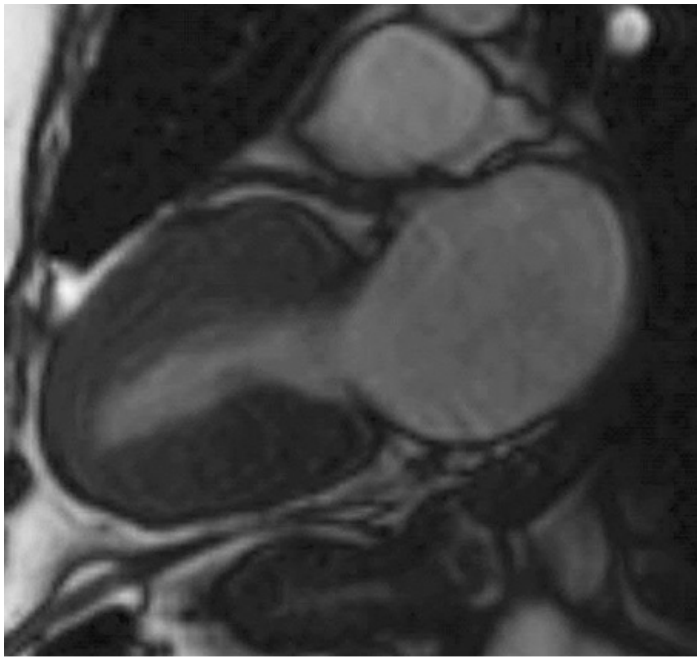

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI scan of a patient with hypertrophic cardiomyopathy, showing the severe increase in left ventricular wall thickness. Cardiac MRI is an ideal imaging modality for diagnosing cardiomyopathies.

Figure 229-18

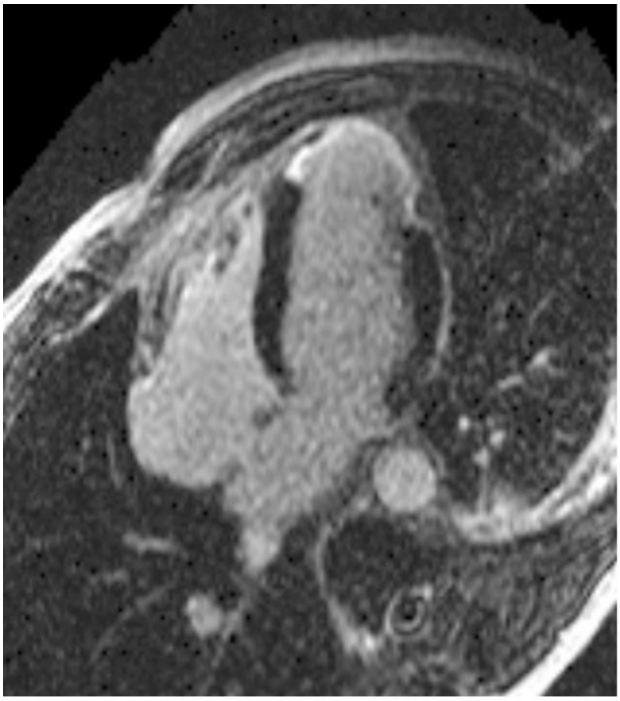

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI scan with delayed gadolinium enhancement in a patient with a large anteroapical infarction. The gadolinium (white area) accumulated in the extracellular space in the presence of cell death from myocardial infarction.

MRA is a standard technique for imaging the aorta and large vessels of the chest and abdomen, with results essentially identical to conventional angiography. MRA of the coronary arteries is a much more difficult challenge, both because of the small size of these vessels and because of their rapid and complex motion during the cardiac cycle; thus, coronary MRA is not yet a reliable clinical technique.

MRI is now an accepted technology for the evaluation of patients with suspected or known coronary disease. Ventricular function and wall motion can be assessed at rest and during infusion of inotropic agents. Assessment of myocardial perfusion can be performed by injecting a bolus of gadolinium contrast and then continuously scanning the heart as the gadolinium passes through the cardiac chambers and into the myocardium. Relative perfusion deficits are reflected as regions of low signal intensity within the myocardium. Pharmacologic stress (typically achieved with vasodilators) can be applied during perfusion imaging to detect physiologically significant coronary artery lesions. Myocardial perfusion imaging with cardiac MRI is more sensitive than SPECT imaging for detecting subendocardial ischemia due to its enhanced spatial resolution.

Myocardial viability and infarction may be determined by imaging the heart 10–20 min after gadolinium injection, known as delayed enhancement magnetic resonance imaging. In normal myocardium, gadolinium cannot penetrate the membranes of the densely packed myocytes. Abnormal myocardial tissue accumulates excess gadolinium following intravenous injection, as ruptured myocyte membranes allow gadolinium to passively diffuse into intracellular space. In chronic infarction, the tissue concentration of gadolinium is increased due to an expansion of the intracellular space from collagenous scar (Fig. 229-18). Thus, delayed enhancement is indicative of nonviable or infarcted myocardium, the subendocardial versus transmural extent of which is accurately assessed by the high spatial resolution of MRI. The presence and pattern of gadolinium enhancement not only is useful for determining viability but also has prognostic value in the patient with an ischemic cardiomyopathy. "Myocardium at risk" following myocardial infarction can be assessed by examining the amount of myocardial edema, using T_2-weighted sequences **(Fig. 229-19)**.

Figure 229-19

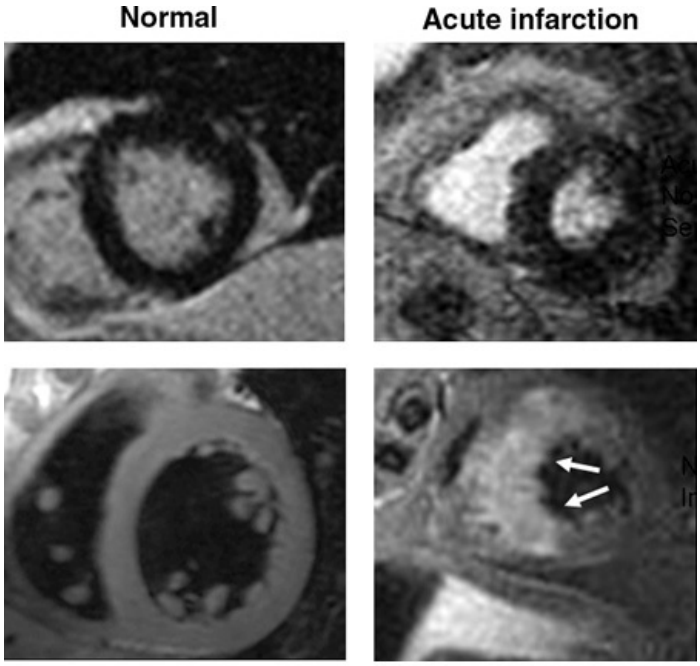

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left: Normal delayed enhancement and "edema"-sensitive images. Top (*left*): Delayed enhancement image illustrating normal black myocardium without infarction/fibrosis. Bottom (*left*): A triple inversion recovery sequence that is T2-weighted demonstrating normal homogenousappearing gray myocardium. *Right:* A patient postmyocardial infarction and early revascularization without evidence of an infarction and an edematous myocardium in the septum. Top (*right*): Delayed enhanced with normal black myocardium without infarction or fibrosis. Bottom (*right*): A triple inversion recovery sequence illustrating edema in the septum without infarction. This is the area of "salvaged" myocardium.

LIMITATIONS OF MRI

Relative contraindications to MRI include the presence of pacemakers, internal defibrillators, or cerebral aneurysm clips. A small percentage of patients are claustrophobic and unable to tolerate the examination within the relatively confined quarters of the magnet bore. Examination of clinically unstable patients and those undergoing stress testing is problematic, since close hemodynamic and electrocardiographic monitoring is difficult. Image quality in patients with significant arrhythmias is often limited. Patients with renal disease receiving gadolinium contrast may be at risk of developing nephrogenic systemic fibrosis, characterized by increased tissue deposition of collagen in the skin and development of fibrosis in skin and other organs.

Computed Tomographic Imaging

BASIC PRINCIPLES

CT is a fast, simple, noninvasive technique that provides images of the myocardium and great vessels with excellent spatial resolution and good soft tissue contrast. The development of electron-beam CT and multidetector-row CT have led to improved temporal resolution and routine imaging of the beating heart. Motion-free high-spatial-resolution images are now possible with multidetector CT technology (≥ 64 channel) that allows imaging of the coronary arteries.

CLINICAL APPLICATIONS

Cardiac CT has important clinical applications. Pericardial calcification is easily detected by CT **(Fig. 229-20)**. CT is useful in characterizing cardiac masses, particularly those containing fat or calcium. The ability to detect small amounts of fat with high spatial resolution makes CT an attractive technique for imaging patients with suspected arrhythmogenic right ventricular dysplasia. Cine images can be used to evaluate wall motion and to determine ejection fraction, end-diastolic and end-systolic volumes, and cardiac mass.

Figure 229-20

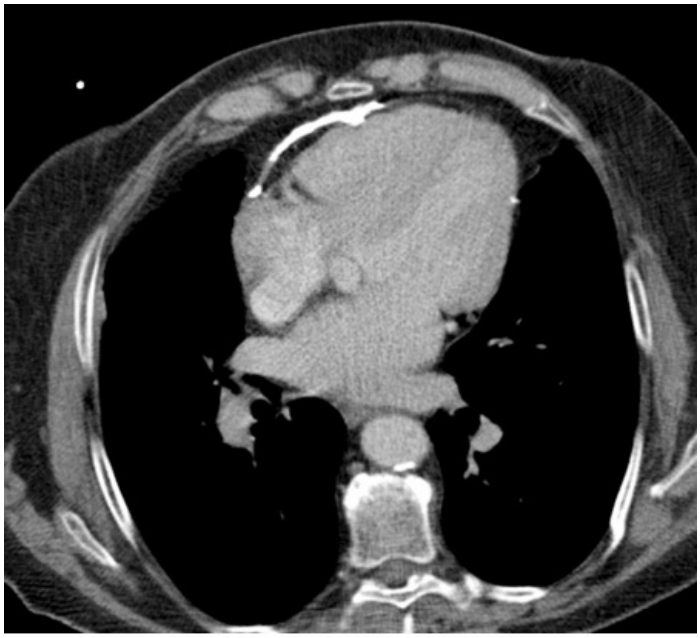

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan showing pericardial calcification, seen as a white linear density anterior to the myocardium.

CT angiography (CTA) has demonstrated accuracy similar to MRA in imaging the aorta and great vessels, and CTA is the examination of choice in the evaluation of patients with suspected pulmonary embolus. CTA is an excellent imaging modality for the diagnosis of aortic dissection or penetrating ulcers. Complete visualization of the entire aorta and its branches is possible by CTA using a single contrast medial bolus injection.

CORONARY CALCIFICATION

Calcium in the coronary arteries occurs in atherosclerosis and is absent in the normal coronary artery **(Fig. 229-21)**. CT is very sensitive for the detection of coronary artery calcification, and the absence of coronary calcification excludes significant epicardial coronary disease. The quantity of coronary calcification (coronary calcium score) is related to the severity of CAD and prognosis. However, the utility of CT calcification score in clinical practice in the asymptomatic patient is limited to those with a moderate risk of coronary heart disease in whom the result will change management.

Figure 229-21

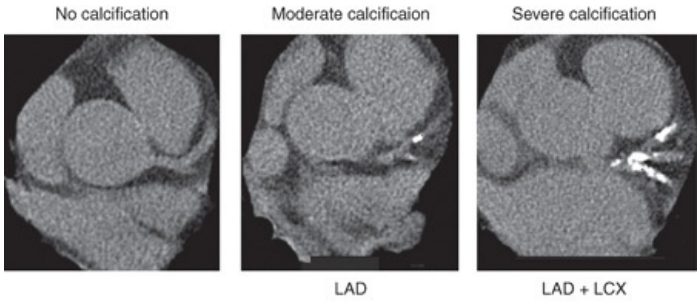

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scans of three patients showing the ability to detect coronary calcification. *Left:* Normal coronary arteries without calcification. *Middle:* Calcification in the left anterior artery (LAD). *Right:* Severe calcification in the LAD and circumflex (CX) arteries.

CONTRAST-ENHANCED CT ANGIOGRAPHY

With the high temporal and spatial resolution of multislice spiral CT, accurate assessment of luminal narrowing in the major branches of the coronary arteries is possible in selected patients. Studies at experienced centers have shown a sensitivity and specificity of >90% for detecting coronary artery lesions as compared to cardiac catheterization. The highest accuracy has been noted in the left main and the proximal portions of the left-sided coronary arteries with decreased sensitivities in the more distal segments and in the more rapidly moving right coronary artery **(Fig. 229-22)**.

Figure 229-22

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three-dimensional volume rendered image of a contrastenhanced CT angiogram demonstrating a normal left main coronary artery arising from the aorta and its two branches, the left anterior descending artery (*left*) and the circumflex artery (*right*).

The concept of "noninvasive coronary angiography" has generated great interest in CTA. However, as with any imaging modalities, CTA has technical limitations requiring proper patient selection and preparation. The integration of CTA into clinical practice requires knowledge of pretest diagnostic and prognostic data and the incremental information that will alter management. The well-accepted indication for coronary CTA is in the evaluation of suspected coronary artery anomalies for which CTA not only confirms the diagnosis but also shows the course of the arteries related to the great vessels **(Fig. 229-23)**. For patients with chest pain syndromes, CTA is best used to rule out significant coronary disease, given its high specificity. Thus, it is the patient with an intermediate pretest probability of CAD who cannot exercise or has uninterpretable or equivocal results on prior testing who would be best suited for CTA. The benefit of CTA in other groups of patients is still unclear.

Figure 229-23

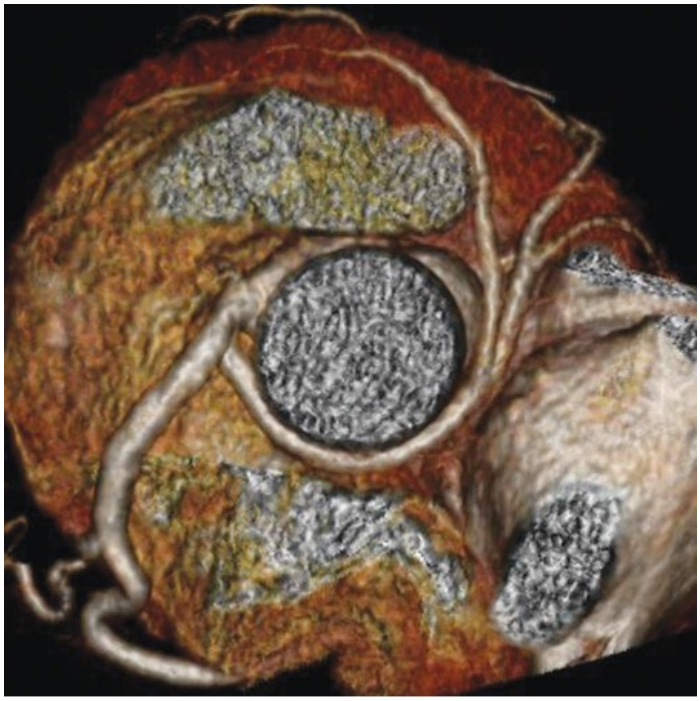

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three-dimensional volume rendered image of a contrastenhanced CT angiogram illustrating an anomalous left coronary artery arising from the right coronary artery and traveling posterior to the aorta.

LIMITATIONS OF CT

Limitations of CT include its dependence on ionizing radiation (in contrast to MRI) and the need for iodinated contrast. Techniques to lower radiation doses continue to evolve, as the radiation doses for coronary CTA generally exceed those delivered during standard diagnostic cardiac catheterization. Fast or irregular heart rhythms and body motion limit the accuracy of CTA. Heavy calcification and artifacts from stents preclude accurate assessment of the severity of a stenosis.

SELECTION OF IMAGING TESTS

(Table 229-3)

Table 229–3. Selection of Imaging Tests

	Echo	Nuclear	CT[a]	MRI[b]
LV size/function	Initial modality of choice Low cost, portable Provides ancillary structural and hemodynamic information	Available from gated SPECT or PET imaging.	Best resolution Highest cost	Best resolution Highest cost
Valve disease	Initial modality of choice Valve motion Doppler hemodynamics			Visualize valve motion Delineate abnormal flow
Pericardial disease Aortic disease	Pericardial effusion Doppler hemodynamics TEE rapid diagnosis[c] Acute dissection		Pericardial thickening Image entire aorta Acute aneurysm Aortic dissection	Pericardial thickening Image entire aorta Aortic aneurysm Chronic dissection

Cardiac masses	TTE—large intracardiac masses TEE—smaller intracardiac masses[c]		Extracardiac masses Myocardial masses	Extracardiac masses Myocardiac masses

[a]Contrast required.

[b]Relative contraindication: pacemakers, metallic objects, claustrophobic.

[c]When not seen on TTE.

Abbreviations: Echo, echocardiography; PET, positron emission tomography; SPECT, single-photon emission computed tomography; TEE, transesophageal echocardiography; TTE, transthoracic echocardiogram.

Basic Premise

The choice of the optimal imaging modality for a particular patient should be based upon the major problem being addressed, other concomitant clinical questions, as well as the local expertise and equipment available in an institution. The clinical urgency and costs of each test also need to be considered. To ensure the effective use of cardiovascular imaging tools, Appropriateness Criteria have been developed by the national societies to examine the incremental clinical benefit of imaging modalities.

Common Clinical Questions

LEFT VENTRICULAR SIZE AND FUNCTION

2D echocardiography is the primary imaging modality obtained for assessment of LV cavity size, systolic function, and wall thickness. Echocardiography can also provide concomitant information on valve function, pulmonary artery pressures, and diastolic filling, which are valuable in the patient presenting with possible heart failure. The disadvantage is poor endocardial resolution in some patients and the lack of reproducible quantitative measurements.

Equilibrium radionuclide angiography can provide an accurate quantitative measurement of LV volumes and function but is not widely available and cannot be used in patients with irregular rhythms. Gated SPECT and PET measure LV systolic function and volumes as a part of myocardial perfusion and/or viability imaging but also require relatively regular rhythm. Both MRI and CT scanning provide the highest quality resolution of the endocardial border and, thus, are the most accurate of all modalities. However, they are of higher cost, lack portability, and do not provide concomitant hemodynamic information as echocardiography does.

VALVULAR HEART DISEASE

2D and Doppler echocardiography provide both anatomic and hemodynamic information regarding valve disease, and are the first test of choice. MRI can also visualize valve motion and determine abnormal flow velocities across valves, but there is less validation of quantitative hemodynamic measurements in comparison to echocardiography.

PERICARDIAL DISEASE

Echocardiography is the first imaging modality of choice in patients with suspected pericardial effusion and tamponade owing to its rapid image display and portability. For patients with suspected constrictive pericarditis, either MRI or CT scanning is the imaging modality that best delineates pericardial thickness. Hemodynamic analysis of the enhancement of ventricular interaction that occurs in pericardial constriction can be assessed by Doppler echocardiography.

AORTIC DISEASE

Both CT scanning and MRI are the imaging modalities of choice for the evaluation of the stable patient with suspected aortic aneurysm or aortic dissection. In the acutely ill patient with suspected aortic dissection, either TEE or CT scanning is a reliable imaging modality.

CARDIAC MASSES

2D TTE is the first test to rule out an intracardiac mass; masses >1.0 cm in diameter are usually well visualized. Intracardiac masses of smaller size may be visualized by TEE. CT scanning and MRI are optimal for evaluating masses extrinsic to the heart or involving the myocardium.

Choosing the Appropriate Imaging Test for the Evaluation of Known or Suspected CAD

The choice of an initial test should be based on the evaluation of the patient's resting electrocardiogram, the ability to perform exercise, the clinical features, the patient's body habitus, and the available local expertise and technology **(Fig. 229-24)**. For the standard assessment of CAD, the exercise electrocardiographic test should be the initial consideration in patients with an interpretable electrocardiogram who are able to exercise. If there are resting electrocardiographic abnormalities, or if the patient has had prior coronary revascularization, an imaging modality (either nuclear imaging or

echocardiography) should be used for initial evaluation. Imaging tests can add prognostic information to a standard exercise electrocardiographic test and, thus, are especially useful when the initial results fall into an intermediate risk category. Pharmacologic stress testing with imaging should be used in patients who are unable to exercise. The utility of CT coronary angiography is evolving.

Figure 229-24

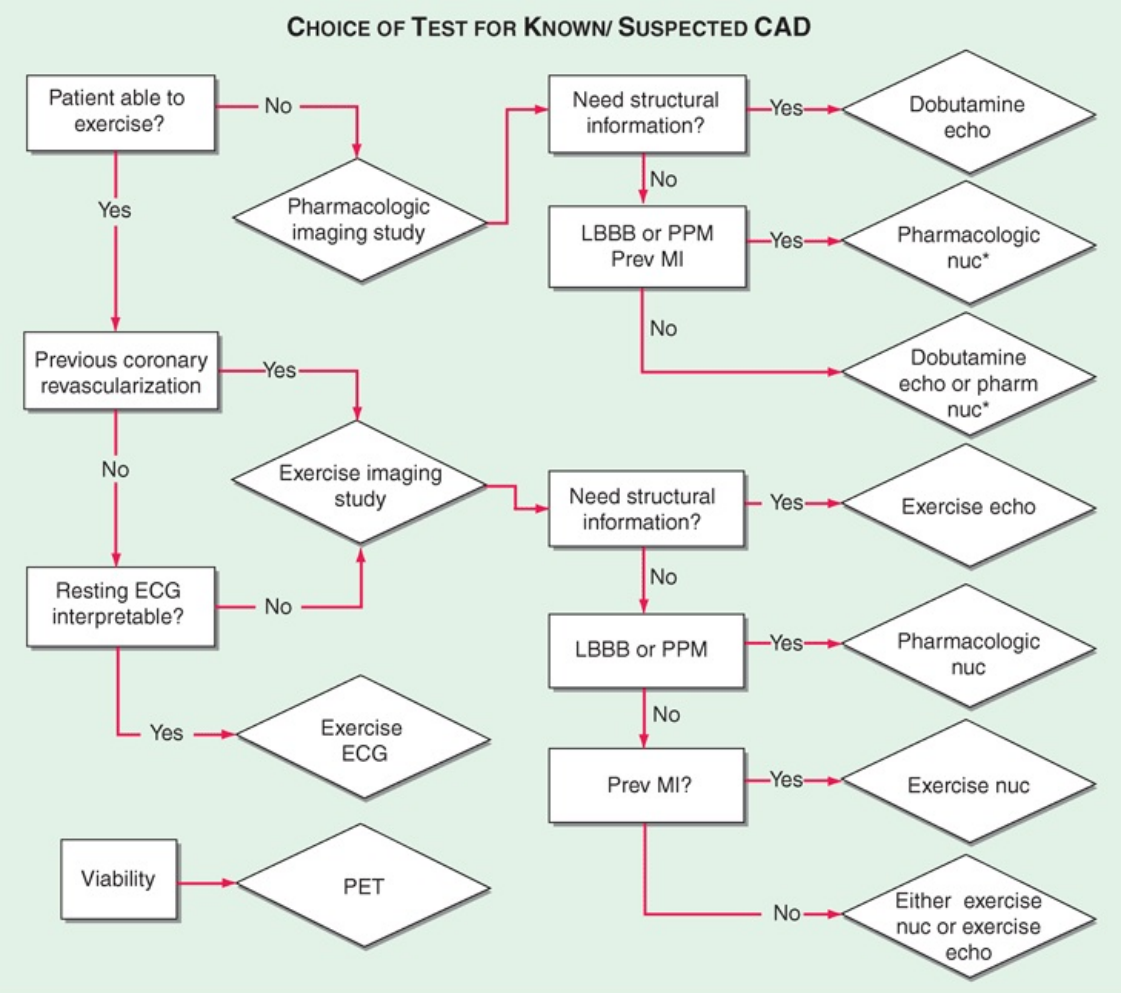

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flow diagram showing selection of initial stress test in a patient with chest pain. Patients who are able to exercise, without previous revascularization, and with an interpretable resting ECG can be tested with an exercise ECG. The appropriate imaging study for other patients depends on multiple factors (see text). LBBB, left bundle branch block; Prev MI-Reg ischemia, previous MI with a need to detect regional ischemia; Nuc, SPECT nuclear imaging study; Pharm, pharmacologic. *Consider PET if morbidly obese or female with large/dense breasts.

While the patient is often best evaluated with the imaging modality for which most experience and expertise are available, there are additional considerations and certain situations where one imaging modality has an advantage over another. Echocardiography provides structural information. Therefore, if there is a question of concomitant valve disease, pericardial disease, or aortic disease, stress echocardiography should be considered. In patients with previous infarction and/or LV systolic dysfunction on the basis of CAD, nuclear imaging, particularly PET, or MRI, is the preferred modality as it also establishes viability. In general, nuclear imaging is more sensitive and less specific than echocardiography for the detection of myocardial ischemia and viability.

FURTHER READINGS

Berman DS et al: ACCF/ACR/AHA/ASNC/NASCI/SAIP/SCAI/SCCT 2009 Expert Consensus Document on Coronary CT Angiography. Circ (In press)

Greenland P et al: ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: A report of the American College of Cardiology Foundation Clinical Expert

Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography) developed in collaboration with the Society of Atherosclerosis Imaging and Prevention and the Society of Cardiovascular Computed Tomography. JACC 49:378, 2007[PMID: 17239724]

Hendel RC et al: ACCF/ACR/SCCT/SCMR/ASNC/NASCI/SCAI/SIR 2006 Appropriateness criteria for cardiac computed tomography and cardiac magnetic resonance imaging: A report of the American College of Cardiology Foundation Quality Strategic Directions Committee Appropriateness Criterial Working Group, American College of Radiology, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, American Society of Nuclear Cardiology, North American Society for Cardiac Imaging, Society for Cardiovascular Angiography and Interventions, and Society of Interventional Radiology. JACC 48:1475, 2006[PMID: 17010819]

——— et al: ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 Appropriate use criteria for cardiac radionuclide imaging: A report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the American Society of Nuclear Cardiology, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the Society of Cardiovascular Computed Tomography, the Society for Cardiovascular Magnetic Resonance, and the American College of Emergency Physicians. JACC 53:2201, 2009

Lester SJ et al: Unlocking the mysteries of diastolic function: Deciphering the Rosetta stone 10 years later. JACC 51:679, 2008[PMID: 18279730]

Miller JM et al: Diagnostic performance of coronary angiography by 64-row CT. N Engl J Med 359:2324, 2008[PMID: 19038879]

Rispler S et al: Integrated single-photon emission computed tomography and computed tomography coronary angiography for the assessment of hemodynamically significant coronary artery lesion. JACC 49:1059, 2007[PMID: 17349885]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ECHOCARDIOGRAPHIC IMAGES

Video e29-1A

Real-time two-dimensional echocardiographic images of a patient with a normal heart. ***A.*** Parasternal long-axis view. There is symmetric contraction of the ventricles, evidenced by a decrease in cavity size and increase in wall thickness during systole. Echocardiographic imaging is performed in multiple acoustic windows with different transducer rotations so that the entire heart and great vessels can be displayed in various planes. Most information from a study is obtained from visual analysis of the two-dimensional images, although objective measurements of cardiac dimensions can be made.

Video e29-1B

Real-time two-dimensional echocardiographic images of a patient with a normal heart. ***B.*** Parasternal short-axis view. There is symmetric contraction of the ventricles, evidenced by a decrease in cavity size and increase in wall thickness during systole. Echocardiographic imaging is performed in multiple acoustic windows with different transducer rotations so that the entire heart and great vessels can be displayed in various planes. Most information from a study is obtained from visual analysis of the two-dimensional images, although objective measurements of cardiac dimensions can be made.

Video e29-2A

Real-time two-dimensional echocardiographic images of a patient with a severe decrease in left ventricular systolic function. The estimated ejection fraction is 20%. ***A.*** Parasternal long-axis view.

Video e29-2B

Real-time two-dimensional echocardiographic images of a patient with a severe decrease in left ventricular systolic function. The estimated ejection fraction is 20%. ***B.*** Parasternal short-axis view.

Video e29-3A

Real-time two-dimensional echocardiographic images of a patient with hypertrophic cardiomyopathy. There is a marked increase in left ventricular wall thickness with hyperdynamic systolic function. ***A.*** Parasternal long-axis view.

Video e29-3B

Real-time two-dimensional echocardiographic images of a patient with hypertrophic cardiomyopathy. There is a marked increase in left ventricular wall thickness with hyperdynamic systolic function. ***B.*** Parasternal short-axis view.

Video e29-4

Real-time two-dimensional parasternal long-axis images from a patient with aortic stenosis. There is normal left ventricular cavity size with normal systolic function. The aortic valve is thickened and calcified, with restricted opening.

Video e29-5A

Real-time two-dimensional echocardiographic images of a patient with mitral stenosis. There is diastolic doming and restricted leaflet opening secondary to fusion of the commissures. ***A.*** Parasternal long-axis view.

Video e29-5B

Real-time two-dimensional echocardiographic images of a patient with mitral stenosis. There is diastolic doming and restricted leaflet opening secondary to fusion of the commissures. ***B.*** Parasternal short-axis view.

Video e29-6A

Real-time two-dimensional echocardiographic images from the parasternal long-axis view of a patient with mitral valve prolapse. During systole, both anterior and posterior leaflet of the mitral valve prolapse into the left atrium. ***A.*** Gray-scale images demonstrate a leaflet morphology and motion. Abnormalities of the valve apparatus such as annular dilatation, prolapse, flail leaflets, vegetation, and rheumatic involvement can be diagnosed by two-dimensional echocardiography. The left ventricular response to volume overload can be assessed by two-dimensional echocardiography.

Video e29-6B

Real-time two-dimensional echocardiographic images from the parasternal long-axis view of a patient with mitral valve prolapse. During systole, both anterior and posterior leaflet of the mitral valve prolapse into the left atrium. ***B.*** Color flow imaging demonstrating late systolic blue-colored jet of mitral regurgitation. Abnormalities of the valve apparatus such as annular dilatation, prolapse, flail leaflets, vegetation, and rheumatic involvement can be diagnosed by two-dimensional echocardiography. The left ventricular response to volume overload can be assessed by two-dimensional echocardiography.

Video e29-7A

Real-time two-dimensional images with color flow Doppler imaging of a patient with mitral regurgitation due to ruptured chordae tendineae. ***A.*** Gray-scale image showing a thickened redundant posterior leaflet of the mitral valve with loss of coaptation during systole.

Video e29-7B

Real-time two-dimensional images with color flow Doppler imaging of a patient with mitral regurgitation due to ruptured chordae tendineae. ***B***. Color flow imaging showing severe mitral regurgitation as high velocity turbulence (mosaic pattern) extending into the left atrium during systole.

Video e29-8

Real-time transesophageal echocardiographic images of a patient with severe mitral regurgitation due to a flail posterior leaflet. The posterior mitral valve leaflet is completely unsupported and moves into the left atrium during systole. Transesophageal echocardiography provides high-resolution images of posterior structure such as the left atrium, mitral valve, and aorta.

Video e29-9

Real-time two-dimensional echocardiographic images of a patient with a vegetation on the mitral valve. There is a mobile echo density attached directly to the mitral valve apparatus that intermittently appears in the left atrium.

Video e29-10

Real-time transesophageal echocardiographic images of a patient with a left atrial myxoma. There is a large echo-dense mass in the left atrium that is attached to the atrial septum. The mass moves across the mitral valve during diastole. Although an echocardiographic image cannot provide pathologic confirmation of the etiology of a mass, the diagnosis of atrial myxoma can be suspected from the appearance, mobility, and attachment to the atrial septum.

Video e29-11

Real-time two-dimensional echocardiographic images from the parasternal long-axis view of a patient with a large aneurysm of the ascending aorta.

Video e29-12

Real-time two-dimensional echocardiographic images of a patient with pericardial effusion. The effusion is shown as a black echo-free space surrounding the heart.

Video e29-13

Real-time two-dimensional echocardiographic images from a subcostal view showing a large secundum atrial septal defect. There is a "drop out" in the region of the mid atrial septum. The right ventricle is enlarged from right ventricular volume overload.

Video e29-14A

Real-time two-dimensional echocardiographic images showing a close-up view of the atrial septum in a patient with the question of an atrial septal defect. ***A.*** Gray-scale image showing a questionable "drop out" in the atrial septum.

Video e29-14B

Real-time two-dimensional echocardiographic images showing a close-up view of the atrial septum in a patient with the question of an atrial septal defect. ***B.*** Color flow imaging confirms left to right flow across the atrial septum.

Video e29-15A

Real-time two-dimensional stress echocardiogram in a normal subject. The studies at rest are shown on the left and the studies during peak exercise are shown on the right. ***A.*** Parasternal long-axis (*top*) and short-axis (*bottom*) views. At rest, there is contraction of all segments of the myocardium. During exercise, there is an increase in contractility and in the thickening of all segments of the myocardium with a decrease in end-systolic volume.

Video e29-15B

Real-time two-dimensional stress echocardiogram in a normal subject. The studies at rest are shown on the left and the studies during peak exercise are shown on the right.*B.* Apical four-chamber (*top*) and two-chamber (*bottom*) views. At rest, there is contraction of all segments of the myocardium. During exercise, there is an increase in contractility and in the thickening of all segments of the myocardium with a decrease in end-systolic volume.

Video e29-16A

Real-time two-dimensional stress echocardiogram of a patient with coronary artery disease. The studies at rest are shown on the left and studies during peak exercise are shown on the right. ***A.*** Parasternal long-axis (*top*) and short-axis (*bottom*) views. The images during peak exercise show regional wall motion abnormalities in the anteroseptal distribution indicative of myocardial ischemia. This was subsequently found to be associated with a high-grade lesion on the left anterior descending artery.

Video e29-16B

Real-time two-dimensional stress echocardiogram of a patient with coronary artery disease. The studies at rest are shown on the left and studies during peak exercise are shown on the right. ***B.*** Apical four-chamber (*top*) and two-chamber (*bottom*) views. The images during peak exercise show regional wall motion abnormalities in the anteroseptal distribution indicative of myocardial ischemia. This was subsequently found to be associated with a high-grade lesion on the left anterior descending artery.

NUCLEAR IMAGES

Figure e29-1

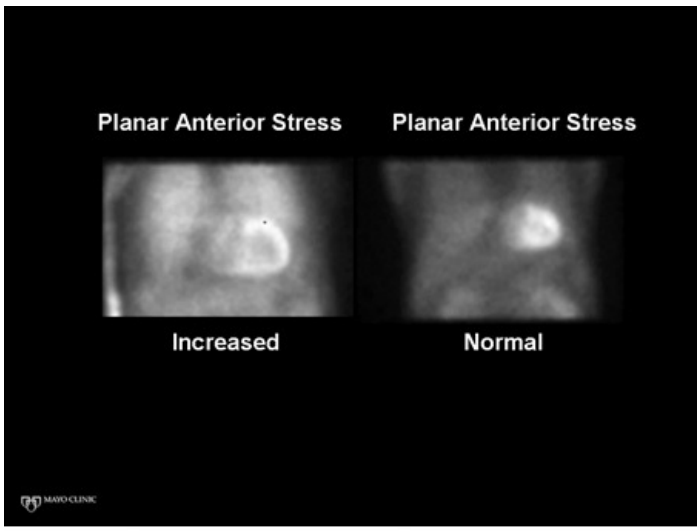

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anterior planar thallium images following stress, showing increased lung uptake on the left (count intensity in lung >50% of that in myocardium) and normal lung uptake on the right (count intensity in lung <50% of that in myocardium).

Increased lung uptake of thallium may be seen immediately after stress. It reflects increased pulmonary capillary wedge pressure and occurs in the presence of severe coronary artery disease and/or left ventricular systolic dysfunction. It provides important adverse prognostic information that is

incremental to other clinical, stress, and coronary angiographic variables.

Figure e29-2

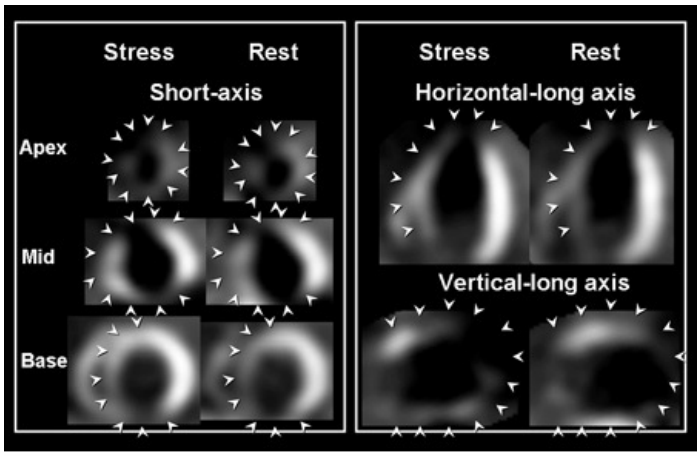

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Exercise SPECT ^{99m}Tc sestamibi scan in a 64-year-old male patient with a previous infarct. The stress images *(left)* show a large defect involving the apex, anterior, septal, and inferior walls (arrowheads) with little change from the rest images (right), signifying a fixed defect consistent with infarction. There is also severe left ventricular enlargement and severely reduced left ventricular systolic function by gated images (next image).The relative advantage of ^{201}Tl and ^{99m}Tc are detailed in Table e20-1a. The better image quality and assessment of ventricular function permitted by ^{99m}Tc compounds have contributed to their more common use for stress imaging, although both ^{201}Tl- and ^{99m}Tc-labeled compounds provide clinically useful myocardial perfusion images in the majority of patients. A "dual-isotope" protocol is employed in some centers. This uses ^{201}Tl for the initial rest image and a ^{99m}Tc-labeled compound for the subsequent stress image, primarily for patient and scheduling convenience. SPECT, single photon emission computed tomography; ^{99m}Tc, technetium 99m.

Figure e29-3

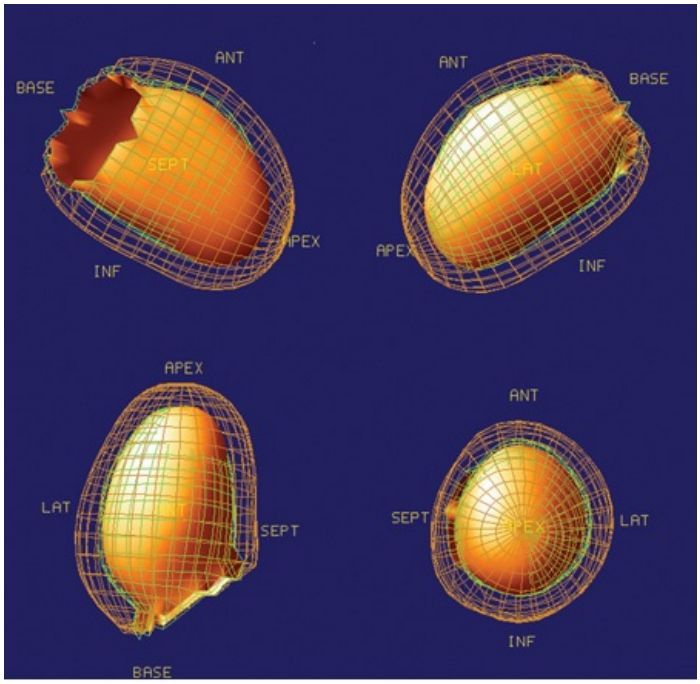

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mesh cinegraphic display of images obtained from electrocardiographic gating of a SPECT ^{99m}Tc sestamibi scan in a 64-year-old male patient with a previous infarct (same patient as in Fig. e29-2). Gated images are generally acquired about 30–45 min following stress. Electrocardiographic gating allows calculation of left ventricular volumes and global systolic function, as well as visual assessment of regional wall motion. In this patient, the calculated left ventricular ejection fraction was 13%. There was severe global hypokinesis.

Figure e29-4

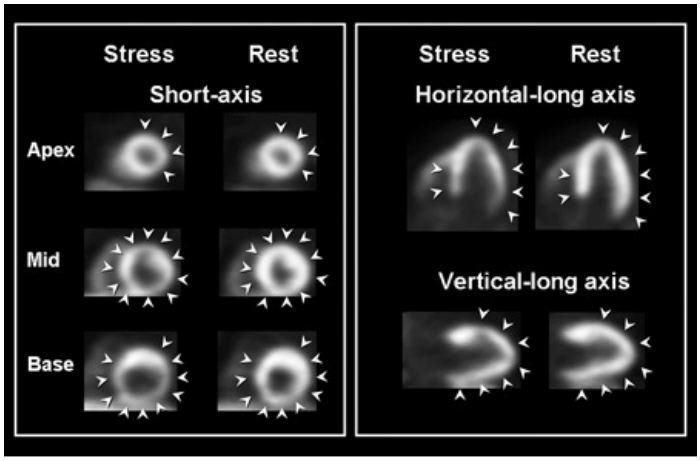

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adenosine PET N-13 ammonia in a 55-year-old obese male patient with typical angina. The stress images *(left)* show a large defect involving the apex, anterior, septal, inferior, and lateral walls *(arrowhead)* with normal or near-normal tracer uptake in the corresponding regions on the rest images *(arrowheads)*, signifying a reversible defect consistent with ischemia. The patient was found to have severe multivessel coronary artery disease on subsequent invasive angiography.

The robust methods for attenuation correction with PET improve the specificity, particularly in obese populations and women, while the superior resolution and higher extraction fraction of PET tracers increase the sensitivity of detection of coronary artery disease. The excellent image quality of PET has contributed to their emerging use for stress imaging. Additional advantages include assessment of left ventricular volumes and systolic function, shorter imaging protocols, and lower radiation exposure, particularly with N-13 ammonia.

PET, positron emission tomography. SPECT, single photon emission computed tomography.

MRI/CT IMAGES

Video e29-17

MRI scan in real time of a patient with a large left ventricular apical aneurysm. The long axis-view demonstrates a thin dyskinetic apical aneurysm with a preserved systolic function of the basal anterior and basal inferior wall. MRI scanning allows excellent visualization of endocardial border.

Figure e29-5

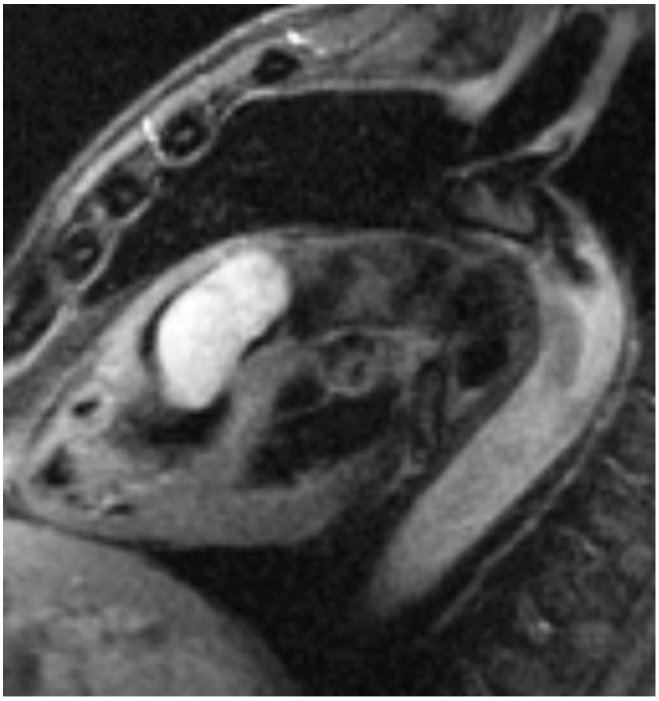

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MR image of a patient with a right ventricular myxoma, which is shown as a bright oblong structure in the right ventricular outflow tract.

Video e29-18

Cine MRI scan of a patient with a dilated ascending aorta (annulo-aortic ectasia). There is a central jet of aortic regurgitation entering the left ventricular outflow tract.

Figure e29-6

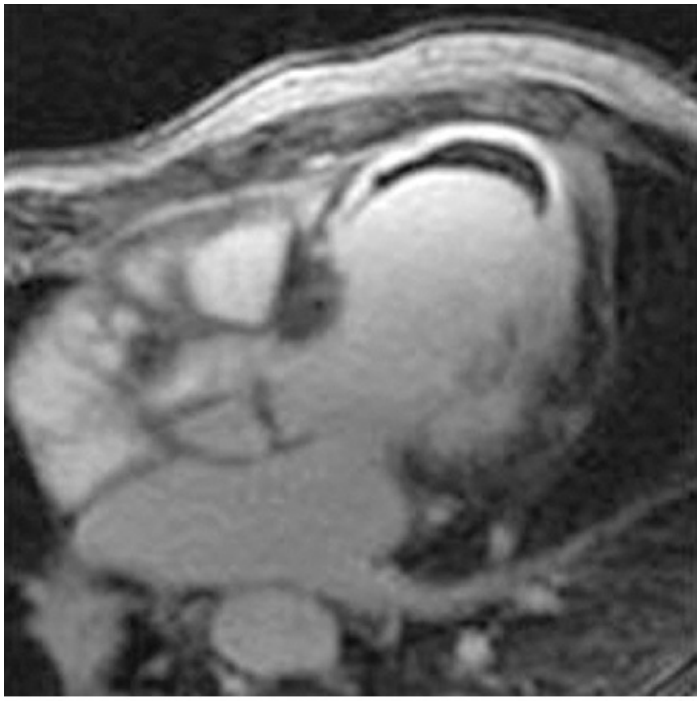

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MR image with contrast enhancement of a patient with a large apical aneurysm and thrombus. Imaging the heart 10–20 min after gadolinium injection demonstrates enhancement of the infarcted tissue (visible as dense white image). The infarcted tissue retains contrast by virtual of its large extracellular volume. The left ventricular thrombus adherent to the infarcted myocardium is shown as a dark laminated area adjacent to the white myocardium.

Figure e29-7

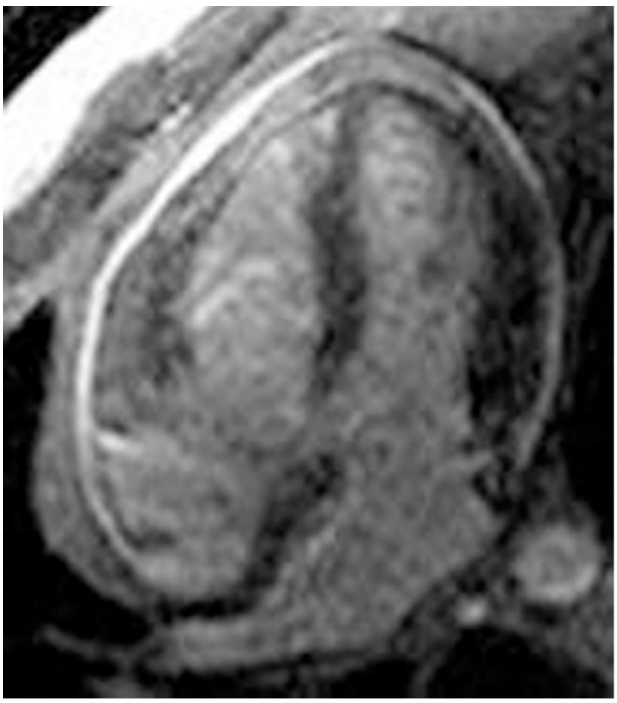

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MR images with contrast enhancement in a patient with acute pericarditis. In the presence of pericardial inflammation, the gadolinium enhancement occurs, seen as a white layer in the pericardium.

Figure e29-8

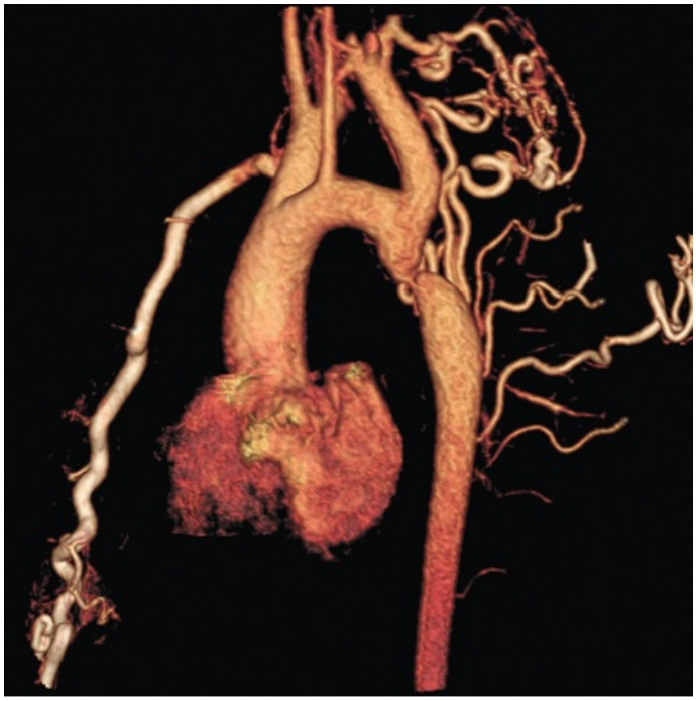

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three-dimensional reconstruction of a CT angiogram, showing a severe coarctation of the descending aorta. The large collateral vessels are the result of the severe stenosis of the distal thoracic aorta.

Figure e29-9

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three-dimensional reconstruction of a CT angiogram of the pulmonary veins, demonstrating an anomalous pulmonary venous drainage into the inferior vena cava. ***A.***: frontal view. ***B.***: posterior view.

Figure e29-10

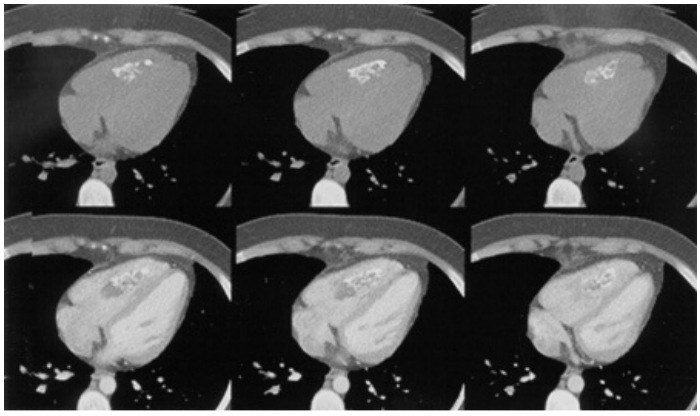

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cardiac CT images demonstrating a calcified mass in the right ventricle, which at pathologic examination was a chronic thrombus. Calcification is seen as a bright signal in both the noncontrast (*upper*) and contrast-enhanced (*lower*) images.

Figure e29-11

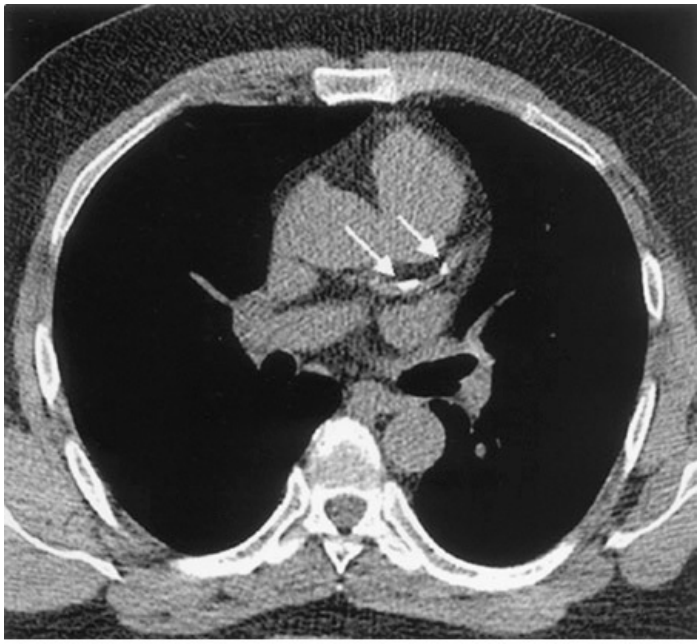

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Noncontrast image from an electron beam CT revealing two small foci of calcification in the left anterior descending artery (*arrows*).

Figure e29-12

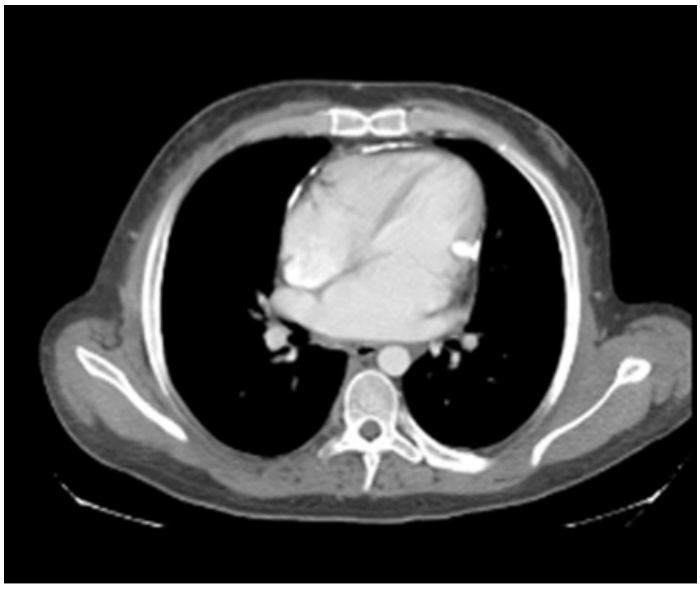

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT image from a patient with calcific constrictive pericarditis. Calcification is seen as a bright signal in the anterior pericardium as well as calcification extending into the lateral wall of the left ventricle.

Figure e29-13

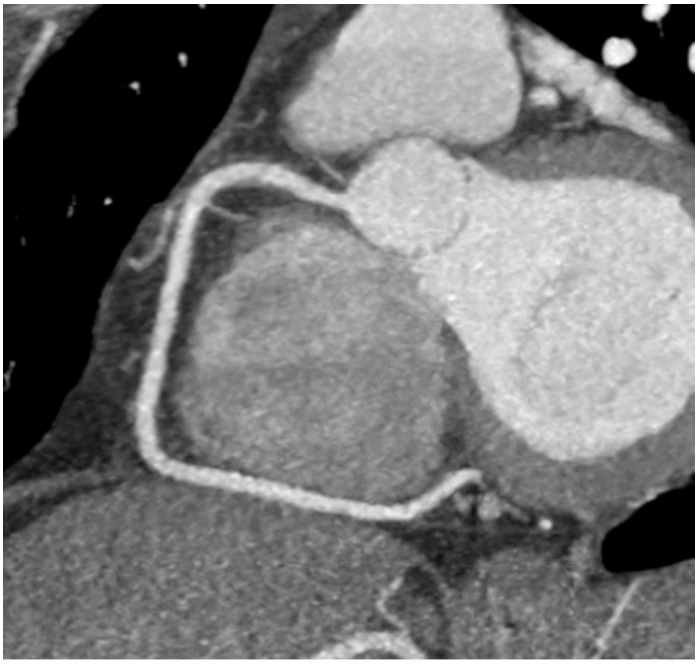

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A reconstructed CT coronary angiogram showing a normal right coronary artery.

Video e29-19

CT coronary angiogram showing a normal right coronary artery. The movie highlights multiple thin slices through the right coronary artery.

Figure e29-14

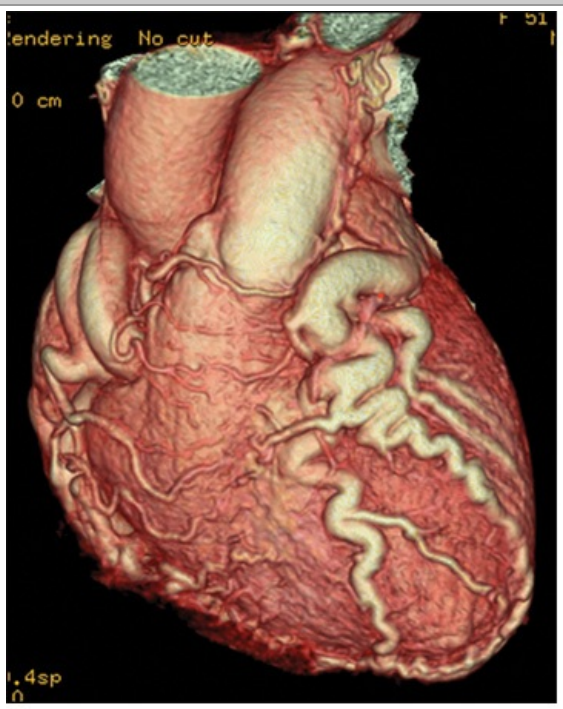

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three-dimensional reconstruction of a CT angiogram showing a large fistula of the left anterior descending artery.

Figure e29-15

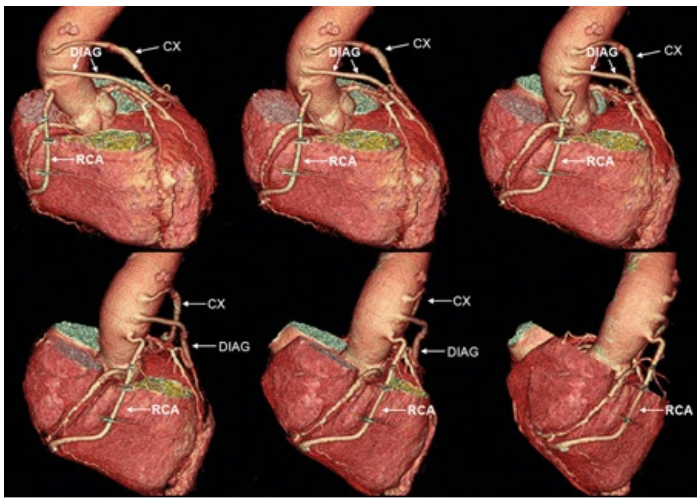

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three-dimensional reconstruction of a CT angiogram demonstrating three saphenous vein coronary artery bypass grafts in different views. In the upper left-hand panel is an anterior posterior view of the heart and grafts. The heart is sequentially rotated clockwise in the panels going from left to right to illustrate the ability of CT angiography to visualize the saphenous vein grafts. RCA: saphenous vein graft to the right coronary artery; CX: saphenous vein graft to the circumflex artery; DIAG: saphenous vein graft to the diagonal artery.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIAGNOSTIC CARDIAC CATHETERIZATION AND CORONARY ANGIOGRAPHY: INTRODUCTION

Diagnostic cardiac catheterization and coronary angiography are considered the gold standard in the assessment of the anatomy and physiology of the heart and its associated vasculature. In 1929, Forssmann demonstrated the feasibility of cardiac catheterization in humans when he passed a urological catheter from a vein in his arm to his right atrium and documented the catheter's position in the heart by x-ray. In the 1940s, Cournand and Richards applied this technique to patients with cardiovascular disease to evaluate cardiac function.These three physicians were awarded the Nobel Prize in 1956. In 1958, Sones inadvertently performed the first selective coronary angiography when a catheter in the left ventricle slipped back across the aortic valve, engaged the right coronary artery, and power-injected 40 mL of contrast down the vessel. The resulting angiogram provided superb anatomic detail of the artery, and the patient suffered no adverse effects. Sones went on to develop selective coronary catheters, which were modified further by Judkins, who developed preformed catheters and allowed coronary artery angiography to gain widespread use as a diagnostic tool. In the United States, cardiac catheterization is the second most common operative procedure, with nearly three million procedures performed annually.

CARDIAC CATHETERIZATION

Indications, Risks, and Preprocedure Management

Cardiac catheterization and coronary angiography are indicated to evaluate the extent and severity of cardiac disease in symptomatic patients and to determine if medical, surgical, or catheter-based interventions are warranted **(Table 230-1)**. They are also used to exclude severe disease in symptomatic patients with equivocal findings on noninvasive studies and in patients with chest-pain syndromes of unclear etiology for whom a definitive diagnosis is necessary for management. Cardiac catheterization is not mandatory prior to cardiac surgery in some younger patients who have congenital or valvular heart disease that is well defined by noninvasive imaging and who do not have symptoms or risk factors that suggest concomitant coronary artery disease.

Table 230–1. Indications for Cardiac Catheterization and Coronary Angiography

Coronary Artery Disease
Asymptomatic or Symptomatic
High risk for adverse outcome based on noninvasive testing
Sudden cardiac death
Sustained (>30 sc) monomorphic ventricular tachycardia
Nonsustained (<30 sc) polymorphic ventricular tachycardia
Symptomatic
Canadian Cardiology Society class III or IV angina on medical therapy
Unstable angina—high or intermediate risk
Chest-pain syndrome of unclear etiology and equivocal findings on noninvasive tests
Acute Myocardial Infarction
Reperfusion with primary percutaneous coronary intervention
Persistent or recurrent ischemia
Severe pulmonary edema
Cardiogenic shock or hemodynamic instability
Mechanical complications—mitral regurgitation, ventricular septal defect
Valvular Heart Disease
Suspected valve disease in symptomatic patients—dyspnea, angina, heart failure, syncope
Infective endocarditis with coronary embolization
Asymptomatic patients with aortic regurgitation and cardiac enlargement or ↓ ejection fraction
Prevalve surgery in older patients with coronary artery diseaserisk factors
Congestive Heart Failure

New onset with angina or suspected undiagnosed coronary artery disease
Congenital Heart Disease
Prior to surgical correction, when symptoms or noninvasive testing suggests coronary disease
Suspicion for congenital coronary anomalies
Forms of congenital heart disease associated with coronary anomalies
Pericardial Disease
Symptomatic patients with suspected cardiac tamponade or constrictive pericarditis
Cardiac Transplantation
Preoperative and postsurgical evaluation
Other Conditions
Hypertrophic cardiomyopathy with angina
Diseases of the aorta when knowledge of coronary artery involvement is necessary for management

Source: Adapted from American College of Cardiology/American Heart Association Ad Hoc Task Force on Practice Guidelines: ACC/AHA guidelines for coronary angiography. Circulation 1999;99:2345–2357.

The risks associated with elective cardiac catheterization are relatively low, with a reported risk of 0.05% for myocardial infarction, 0.07% for stroke, and 0.08–0.14% for death. These risks increase substantially if the catheterization is performed emergently, during acute myocardial infarction or in hemodynamically unstable patients. Additional risks of the procedure include tachy- or brady-arrhythmias that require countershock or pharmacologic therapy, acute renal failure leading to transient or permanent dialysis, vascular complications that necessitate surgical repair, and significant access-site bleeding. Of these risks, vascular access-site bleeding is the most common complication, occurring in 1.5–2.0% of patients, with major bleeding events associated with a worse short- and long-term outcome.

In patients who understand and accept the risks associated with cardiac catheterization, there are no absolute contraindications when the procedure is performed in anticipation of a life-saving intervention. Relative contraindications do, however, exist; these include decompensated congestive heart failure; acute renal failure; severe chronic renal insufficiency, unless dialysis is planned; bacteremia; acute stroke; active gastrointestinal bleeding; severe, uncorrected electrolyte abnormalities; a history of an anaphylactic/anaphylactoid reaction to iodinated contrast agents; and a history of allergy/bronchospasm to aspirin in patients for whom progression to a percutaneous coronary intervention is likely.

Contrast allergy and contrast-induced renal failure merit further consideration, because these adverse events may occur in otherwise healthy individuals and prophylactic measures exist to reduce risk. Allergic reactions to contrast agents occur in <5% of cases with severe anaphylactoid (clinically indistinguishable from anaphylaxis, but not mediated by an IgE mechanism) reactions occurring in 0.1%–0.2% of patients. Mild reactions manifest as nausea, vomiting, and urticaria, while severe anaphylactoid reactions lead to hypotensive shock, pulmonary edema, and cardiorespiratory arrest. Patients with a history of significant contrast allergy should be premedicated with corticosteroids and antihistamines (H_1- and H_2-blockers) and studies performed with nonionic, low-osmolar contrast agents that have a lower reported rate of allergic reactions.

Contrast-induced nephropathy, defined as an increase in creatinine >0.5 mg/dL or 25% above baseline that occurs 48–72 hours after contrast administration, occurs in ~2–7% of patients with rates of 20–30% reported in high-risk patients, including those with diabetes mellitus, congestive heart failure, chronic kidney disease, anemia, and older age. Dialysis is required in 0.3–0.7% of patients and is associated with a 5-fold increase in in-hospital mortality. For all patients, adequate intravascular volume expansion with intravenous 0.9% saline (1.0–1.5 mL/kg per hour) for 3–12 hours before and continued 6–24 hours after the procedure limits the risk of contrast-induced nephropathy. In patients with chronic kidney disease, additional pretreatment with *N*-acetylcysteine (Mucomist, 600 mg bid orally before and two days after catheterization) also decreases risk. Diabetic patients treated with metformin should stop the drug 48 hours prior to the procedure to limit the associated risk of lactic acidosis. Other strategies to decrease risk include the administration of sodium bicarbonate, although there is conflicting data regarding its efficacy; use of low- or iso-osmolar contrast agents; and limiting the volume of contrast to <100 mL per procedure.

Cardiac catheterization is performed after the patient has fasted for six hours and has received IV conscious sedation to remain awake but sedated during the procedure. All patients with suspected coronary artery disease are pretreated with 325 mg aspirin. In patients in whom the procedure is likely to progress to a percutaneous coronary intervention, a clopidogrel 600-mg loading dose followed by 75 mg daily should be started. Warfarin is held starting 48 hours prior to the catheterization to allow the international normalized ratio (INR) to fall to <2.0 and limit access-site bleeding complications. Cardiac catheterization is a sterile procedure, so antibiotic prophylaxis is not required.

Technique

Cardiac catheterization and coronary angiography provide a detailed hemodynamic and anatomic assessment of the heart and coronary arteries. The selection of procedures is dependent upon the patient's symptoms and clinical condition, with some direction provided by noninvasive studies.

VASCULAR ACCESS

Cardiac catheterization procedures are performed using a percutaneous technique to enter the femoral artery and vein as the preferred access sites for left and right heart catheterization, respectively. A flexible sheath is inserted into the vessel over a guidewire, allowing diagnostic catheters to be introduced into the vessel and advanced toward the heart using fluoroscopic guidance. The brachial or radial artery may also be used as an arterial access site in patients with peripheral arterial disease that involves the abdominal aorta, iliac, or femoral vessels; severe iliac-artery tortuosity; morbid obesity; or preference for early postprocedure ambulation. Use of radial-artery access is gaining popularity owing to a lower rate of access-site bleeding complications. A normal Allen's test confirming dual blood supply to the hand from the radial and ulnar arteries is a prerequisite to access this site. The internal jugular vein serves as an alternate access site to the right heart when the patient has an inferior vena cava filter in place or requires prolonged hemodynamic monitoring.

RIGHT HEART CATHETERIZATION

This procedure measures pressures in the right heart. Right heart catheterization is no longer a routine part of diagnostic cardiac catheterization, but it is reasonable in patients with unexplained dyspnea, valvular heart disease, pericardial disease, right and/or left ventricular dysfunction, congenital heart disease, and suspected intracardiac shunts. Right heart catheterization uses a balloon-tipped flotation catheter that is inserted into the femoral or jugular vein. Using fluoroscopic guidance, the catheter is advanced sequentially to the right atrium, right ventricle, pulmonary artery, and pulmonary wedge position (as a surrogate for left atrial pressure); in each cardiac chamber, pressure is measured and blood samples are obtained for oxygen-saturation analysis to screen for intracardiac shunts.

LEFT HEART CATHETERIZATION

This procedure measures pressures in the left heart as a determinant of left ventricular performance. With the aid of fluoroscopy, a catheter is guided to the ascending aorta and across the aortic valve into the left ventricle to provide a direct measure of left ventricular pressure. In patients with a tilting-disc prosthetic aortic valve, crossing the valve with a catheter is contraindicated and the left heart may be accessed from the right atrium using a needle-tipped catheter to puncture the atrial septum at the fossa ovalis. Once the catheter crosses from the right to the left atrium, it can be advanced across the mitral valve to the left ventricle. This technique is also used for mitral valvuloplasty. Heparin is given for prolonged procedures to limit the risk of stroke from embolism of clots that may form on the catheter.

Hemodynamics

A comprehensive hemodynamic assessment involves obtaining pressure measurements in the right and left heart and peripheral arterial system and determining the cardiac output **(Table 230-2)**. The shape and magnitude of the pressure waveforms provide important diagnostic information; an example of normal pressure tracings is shown in **Fig. 230-1**. In the absence of valvular heart disease, the atria and ventricles are "one chamber" during diastole when the tricuspid and mitral valves are open while in systole, when the pulmonary and aortic valves are open, the ventricles and their respective outflow tracts are considered "one chamber." These concepts form the basis by which hemodynamic measurements are used to assess valvular stenosis. When aortic stenosis is present, there is a systolic pressure gradient between the left ventricle and the aorta; when mitral stenosis is present, there is a diastolic pressure gradient between the pulmonary capillary wedge (left atrial) pressure and the left ventricle **(Fig. 230-2)**. Hemodynamic measurements also discriminate between aortic stenosis and hypertrophic obstructive cardiomyopathy where the asymmetrically hypertrophied septum creates a dynamic intraventricular pressure gradient during ventricular systole. The magnitude of this obstruction is measured using an end-hole catheter positioned at the left ventricular apex that is pulled back while recording pressure; once the catheter has passed the septal obstruction and is positioned in the apex of the left ventricle, a gradient can be measured between the left ventricular apex and the aorta. Hypertrophic obstructive cardiomyopathy is confirmed by the Brockenbrough-Braunwald sign: following a premature ventricular contraction, there is an increase in the left ventricular–aorta pressure gradient with a simultaneous decrease in the aortic pulse pressure. These findings are absent in aortic stenosis.

Table 230–2. Normal Values for Hemodynamic Measurements

Pressures (mmHg)	
Right atrium	
Mean	0–5
a wave	1–7

v wave	1–7
Right ventricle	
Peak systolic/end diastolic	17–32/1–7
Pulmonary artery	
Peak systolic/end diastolic	17–32/1–7
Mean	9–19
Pulmonary capillary wedge (mean)	4–12
Left atrium Mean	 4–12
a wave	4–15
v wave	4–15
Left ventricle	
Peak systolic/end diastolic	90–140/5–12
Aorta	
Peak systolic/end diastolic	90–140/60–90
Mean **[Resistances (dyn-s)/cm^5]**	70–105
Systemic vascular resistance	900–1400
Pulmonary vascular resistance	40–120
Oxygen Consumption Index [(L-min)/m^2]	115–140
Arteriovenous oxygen difference (vol %) Cardiac index [(L-min)/m^2]	3.5–4.8 2.8–4.2

Figure 230-1

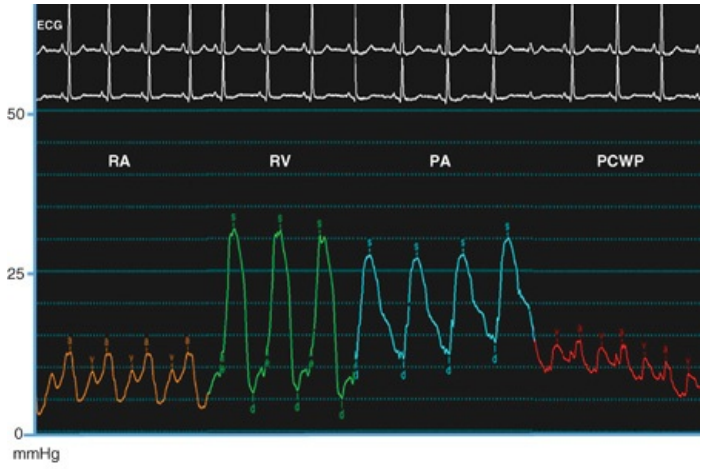

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal hemodynamic waveforms recorded during right heart catheterization. Atrial pressure tracings have a characteristic "a" wave that reflects atrial contraction and a "*v*" wave that reflects pressure changes in the atrium during ventricular systole. Ventricular pressure tracings have a low pressure diastolic filling period and a sharp rise in pressure that occurs during ventricular systole. ra, right atrium; rv, right ventricle; pa, pulmonary artery; pcwp, pulmonary capillary wedge pessure; s, systole; d, diastole.

Figure 230-2

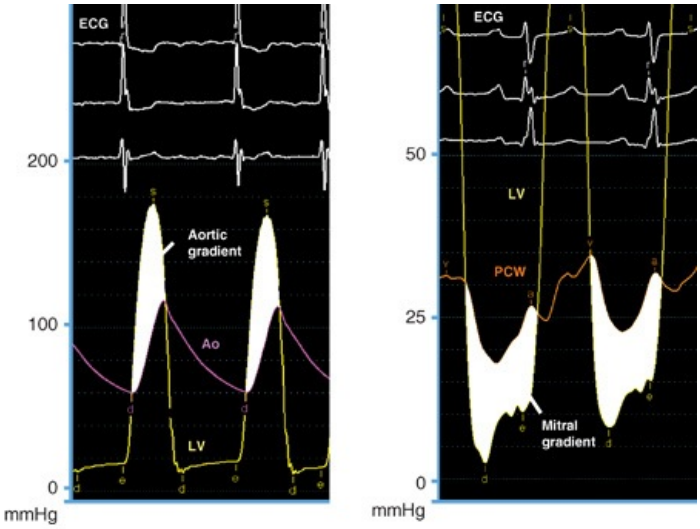

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe aortic and mitral stenosis. Simultaneous recording of left ventricular (LV) and aortic (Ao) pressure tracings demonstrate a 62-mmHg mean systolic gradient (shaded area) that corresponds to an aortic valve area of 0.6 cm^2 (left). Simultaneous recording of LV and pulmonary capillary wedge (PCW) pressure tracings reveal a 14-mmHg mean diastolic gradient (shaded area) that is consistent with critical mitral stenosis (mitral valve area = 0.5 cm^2). s, systole; d, diastole; e, end diastole.

Regurgitant valvular lesions increase volume (and pressure) in the "receiving" cardiac chamber. In severe mitral and tricuspid regurgitation, the increase in blood flow to the atria takes place during ventricular systole, leading to an increase in the *v* wave (two times greater than the mean pressure). Severe aortic regurgitation leads to a decrease in aortic diastolic pressure with a concomitant rise in left ventricular end-diastolic pressure, resulting in equalization of pressures between the two chambers at enddiastole.

Hemodynamic measurements are also used to differentiate between cardiac tamponade, constrictive pericarditis, and restrictive cardiomyopathy. In cardiac tamponade, right atrial pressure is increased with a decreased "y" descent, indicative of impaired right atrial emptying in diastole, and there is diastolic equalization of pressures in all cardiac chambers. In constrictive pericarditis, right atrial pressure is elevated with a prominent "y" descent, indicating rapid filling of the right ventricle during early diastole. A diastolic dip and plateau or "square root sign," in the ventricular waveforms due to an abrupt halt in ventricular filling during diastole; right ventricular and pulmonary artery pressures are elevated; and discordant pressure changes in the right and left ventricles with inspiration (right ventricular systolic pressure increases while left ventricular systolic pressure decreases) are observed. The latter hemodynamic phenomenon is the most specific for constriction. Restrictive cardiomyopathy may be distinguished from constrictive pericarditis by a marked increase in right ventricular and pulmonary artery systolic pressures (usually >60 mmHg), a separation of the left and right ventricular diastolic pressures by >5 mmHg (at baseline or with acute volume loading), and concordant changes in left and right ventricular diastolic filling pressures with inspiration (both increase).

CARDIAC OUTPUT

Cardiac output is measured by the Fick method or the thermodilution technique or calculated from left ventricular angiography. Typically, the Fick method and thermodilution technique are both performed during cardiac catheterization, although the Fick method is considered more reliable in the presence of tricuspid regurgitation and in low-output states. The Fick method uses oxygen as the indicator substance and is based on the principle that the amount of a substance taken up or released by an organ (oxygen consumption) is equal to the product of its blood flow (cardiac output) and the difference in the concentration of the substance in the arterial and venous circulation (arterial-venous oxygen difference). Thus, the formula for calculating the Fick cardiac output is:

Cardiac output (L/min) = [oxygen consumption (mL/min)]/[arterial-venous oxygen difference (mL/L)]

Oxygen consumption is estimated as 125 mL oxygen/minute x body surface area, and the arterial-venous oxygen difference is determined by first calculating the oxygen carrying capacity of blood [hemoglobin (g/100 mL) x 1.36 (mL oxygen/g hemoglobin) x 10] and multiplying this product by the fractional oxygen saturation. The thermodilution method measures a substance that is injected into and adequately mixes with blood. In contemporary practice, thermodilution cardiac outputs are measured using temperature as the indicator. Measurements are made with a thermistor-tipped catheter that detects temperature deviations in the pulmonary artery after the injection of 10 mL of room-temperature normal saline into the right atrium. Cardiac output may also be calculated from the left

ventriculogram by first determining left ventricular volumes in end-diastole and end-systole using the area-length method. Cardiac output is equal to the heart rate x stroke volume, which is the difference between the end-diastolic volume and the end-systolic volume.

VASCULAR RESISTANCE

Resistance across the systemic and pulmonary circulations is calculated by extrapolating from Ohm's law of electrical resistance and is equal to the mean pressure gradient divided by the mean flow (cardiac output). Therefore, systemic vascular resistance is [(mean aortic pressure – mean right atrial pressure)/cardiac output] multiplied by 80 to convert the resistance from Wood units to dyn-s-cm^{-5}. Similarly, the pulmonary vascular resistance is [(mean pulmonary artery – mean pulmonary capillary wedge pressure)/cardiac output] x 80. Pulmonary vascular resistance is lowered by oxygen, nitroprusside, calcium channel blockers, prostacyclin infusions, and inhaled nitric oxide; these therapies may be administered during catheterization to determine if increased pulmonary vascular resistance is fixed or reversible.

VALVE AREA

Hemodynamic data may also be used to calculate the valve area using the Gorlin formula that equates the area to the flow across the valve divided by the pressure gradient between the cardiac chambers surrounding the valve. The formula for the assessment of valve area is: Area = [cardiac output (cm^3/min)/(systolic ejection period or diastolic filling period)(heart rate)]/44.3 C x square root of the pressure gradient, where C = 1 for aortic valve and 0.85 for the mitral valve. A valve area of <1.0 cm^2 and a mean gradient of greater than 40 mmHg indicate severe aortic stenosis, while a valve area of <1.5 cm^2 and a mean gradient >5–10 mmHg is consistent with moderate-to-severe mitral stenosis; in symptomatic patients with a mitral valve area >1.5 cm^2, a mean gradient >15 mmHg, pulmonary artery pressure >60 mmHg, or a pulmonary artery wedge pressure >25 mmHg after exercise is also considered significant and may warrant intervention. The modified Hakki formula has also been used to estimate aortic valve area. This formula calculates the valve area as the cardiac output (L/min) divided by the square root of the pressure gradient. Aortic valve area calculations based on the Gorlin formula are flow-dependent and, therefore, for patients with low cardiac outputs, it is imperative to determine if a decreased valve area actually reflects a fixed stenosis or is overestimated by a low cardiac output and stroke volume that is insufficient to open the valve leaflets fully. In these instances, cautious hemodynamic manipulation using dobutamine to increase the cardiac output and recalculation of the aortic valve area may be necessary.

INTRACARDIAC SHUNTS

In patients with congenital heart disease, detection, localization, and quantification of the intracardiac shunt should be evaluated. A shunt should be suspected when there is unexplained arterial desaturation or increased oxygen saturation of venous blood. A "step up" or increase in oxygen content indicates the presence of a left-to-right shunt while a "step down" indicates a right-to-left shunt. The shunt is localized by detecting a difference in oxygen saturation levels of 5–7% between adjacent cardiac chambers. The severity of the shunt is determined by the ratio of pulmonary blood flow (Q_p) to the systemic blood flow (Q_S), or Q_p/Q_S = [(systemic arterial oxygen content – mixed venous oxygen content)/pulmonary vein oxygen content – pulmonary artery oxygen content]. For an atrial septal defect, a shunt ratio of 1.5 is considered significant and factored with other clinical variables to determine the need for intervention. When a congenital ventricular septal defect is present, a shunt ratio of ≥2.0 with evidence of left ventricular volume overload is a class I indication for surgical correction.

Ventriculography and Aortography

Ventriculography to assess left ventricular function may be performed during cardiac catheterization. A pigtail catheter is advanced retrograde across the aortic valve into the left ventricle and 30–45 mL of contrast is power-injected to visualize the left ventricular chamber during the cardiac cycle. The ventriculogram is usually performed in the right anterior oblique projection to examine wall motion and mitral valve function. Normal wall motion is observed as symmetric contraction of all segments; hypokinetic segments have decreased contraction, akinetic segments do not contract, and dyskinetic segments appear to bulge paradoxically during systole **(Fig. 230-3)**. Ventriculography may also reveal a left ventricular aneurysm, pseudoaneurysm, or diverticulum and can be used to assess mitral valve prolapse and the severity of mitral regurgitation. The degree of mitral regurgitation is estimated by comparing the density of contrast opacification of the left atrium with that of the left ventricle. Minimal contrast reflux into the left atrium is considered 1+ mitral regurgitation while contrast density in the left atrium that is greater than that in the left ventricle with reflux of contrast into the pulmonary veins within three beats defines 4+ mitral regurgitation.

Figure 230-3

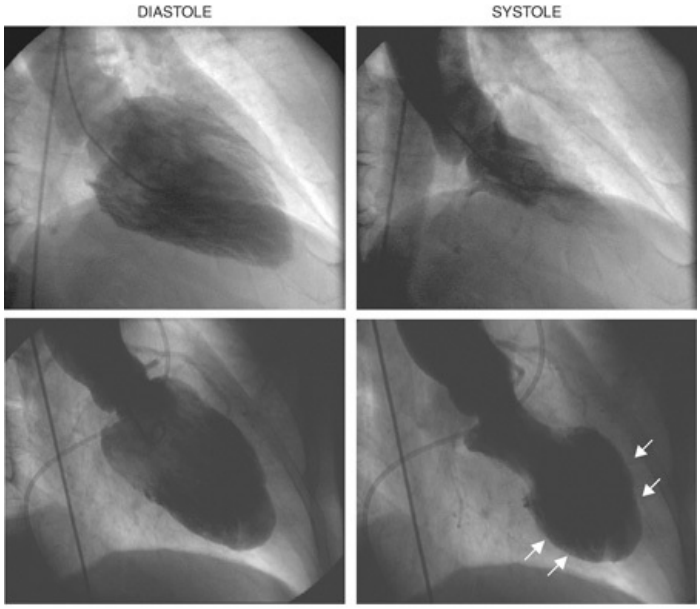

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left ventriculogram at end diastole (left) and end systole (right). In patients with normal left ventricular function, the ventriculogram reveals symmetric contraction of all walls (top). Patients with coronary artery disease may have wall motion abnormalities on ventriculography as seen in this 60-year-old male following a large anterior myocardial infarction. In systole, the anterior, apical, and inferior walls are akinetic (white arrows) (bottom).

Aortography in the cardiac catheterization laboratory visualizes abnormalities of the ascending aorta, including aneurysmal dilation and involvement of the great vessels, as well as dissection with compression of the true lumen by an intimal flap that separates the true and false lumina. Aortography can also be used to identify patent saphenous vein grafts that elude selective cannulation, identify shunts that involve the aorta such as a patent ductus arteriosus, and provide a qualitative assessment of aortic regurgitation using a 1+–4+ scale similar to that used for mitral regurgitation.

CORONARY ANGIOGRAPHY

Selective coronary angiography is almost always performed during cardiac catheterization and is used to define the coronary anatomy and determine the extent of epicardial coronary artery and coronary artery bypass graft disease. Specially shaped coronary catheters are used to engage the left and right coronary ostia. Hand injection of radiopaque contrast agents create a coronary "luminogram" that is recorded on a radiographic images (cine angiography). Because the coronary arteries are three-dimensional objects that are in motion with the cardiac cycle, angiograms of the vessels using several different orthogonal projections are taken to best visualize the vessels without overlap or foreshortening.

The normal coronary anatomy is highly variable between individuals, but, in general, there are two coronary ostia and three major coronary vessels—the left anterior descending, the left circumflex, and the right coronary arteries with the left anterior descending and left circumflex arteries arising from the left main coronary artery **(Fig. 230-4)**. When the right coronary artery is the origin of the A-V nodal branch, the posterior descending artery, and the posterior lateral vessels, the circulation is defined as right dominant; this is found in ~85% of individuals. When these branches arise from the left circumflex artery as occurs in ~5% of individuals, the circulation is defined as left dominant. The remaining ~10% of patients have a codominant circulation with vessels arising from both the right and left coronary circulation. In some patients, a ramus intermedius branch arises directly from the left main coronary artery; this finding is a normal variant. Coronary artery anomalies occur in 1–2% of patients, with separate ostia for the left anterior descending and left circumflex arteries being the most common (0.41%).

Figure 230-4

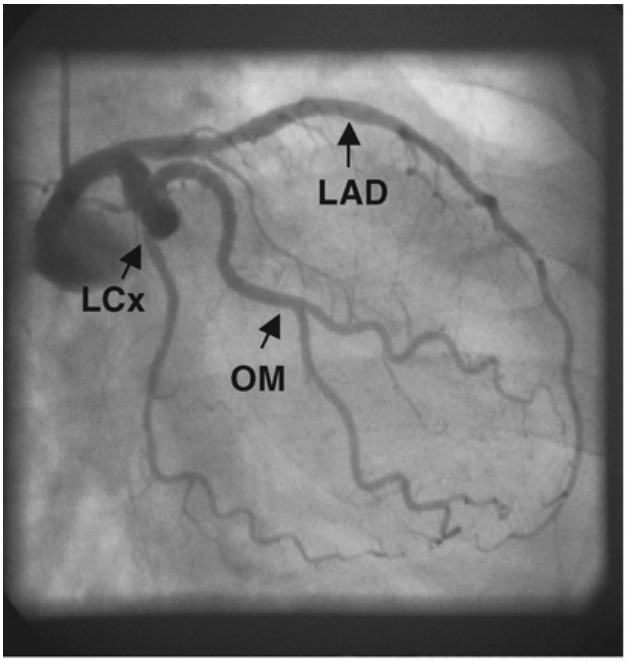

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

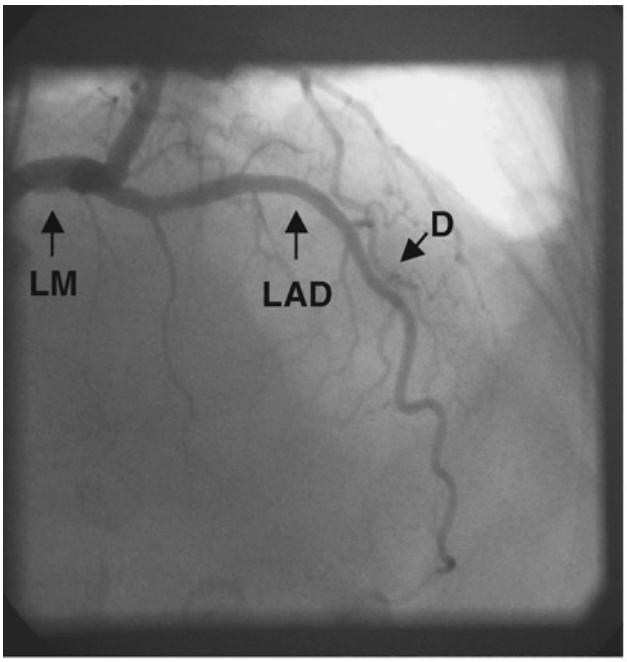

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

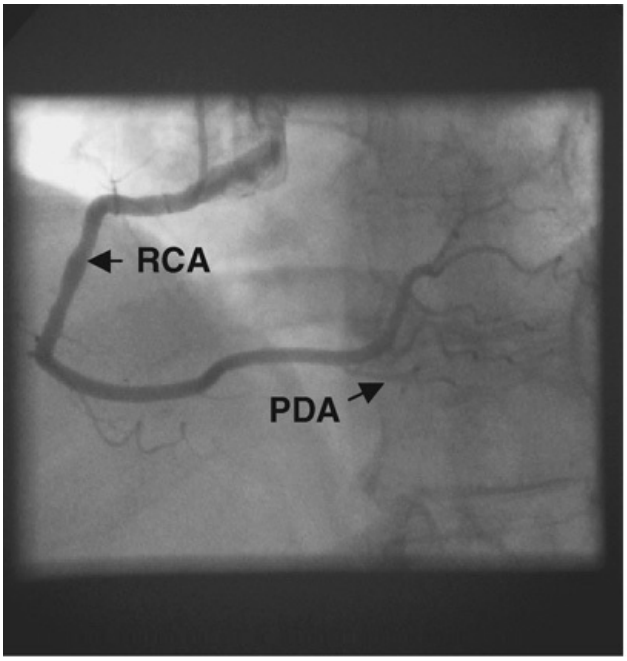

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal coronary artery anatomy. ***A.*** Coronary angiogram showing the left circumflex (LCx) artery and its obtuse marginal (OM) branches. The left anterior descending artery (LAD) is also seen but may be foreshortened in this view. ***B***. The LAD and its diagonal (D) branches are best seen in cranial views. In this angiogram, the left main (LM) coronary artery is also seen. ***C.*** The right coronary artery gives off the posterior descending artery (PDA) so this is a right dominant circulation.

Coronary angiography visualizes coronary artery stenoses as luminal narrowings on the cine angiogram. The degree of narrowing is referred to as the percent stenosis and is determined visually by comparing the most severely diseased segment with a proximal or distal "normal segment" mg; a stenosis >50% is considered significant **(Fig. 230-5)**. Online quantitative coronary angiography can provide a more accurate assessment of the percent stenosis and lessen the tendency to overestimate lesion severity visually. The presence of a myocardial bridge, which most commonly involves the left anterior descending artery, may be mistaken for a significant stenosis; this occurs when a portion of the vessel dips below the epicardial surface into the myocardium and is subject to compressive forces during ventricular systole. The key to differentiating a myocardial bridge from a fixed stenosis is that the "stenosed" part of the vessel returns to normal during diastole. Coronary calcification is also seen during angiography prior to the injection of contrast agents. Collateral blood vessels may be seen traversing from one vessel to the distal vasculature of a severely stenosed or totally occluded vessel. Thrombolysis in myocardial infarction (TIMI) flow grade, a measure of the relative duration of time that it takes for contrast to opacify the coronary artery fully, may provide an additional clue to the degree of lesion severity, and the presence of TIMI grade 1 or 2 flow suggests that a significant coronary artery stenosis is present.

Figure 230-5

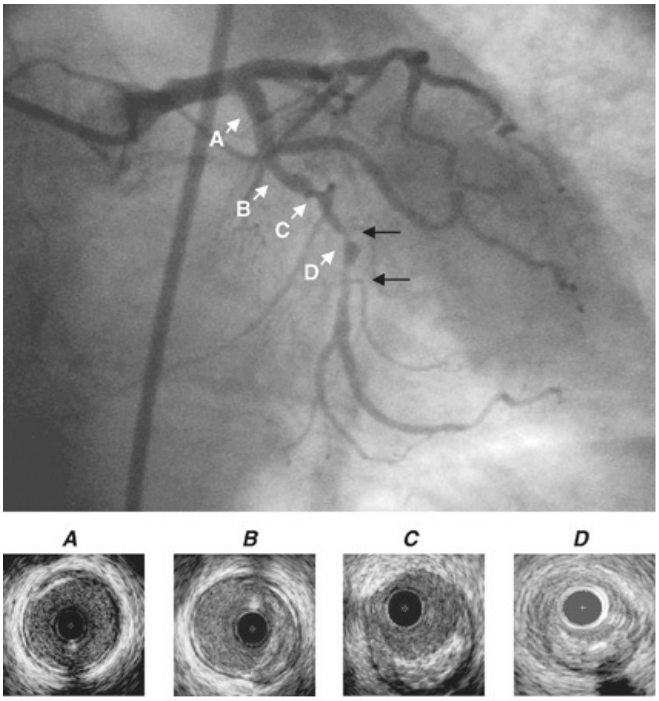

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronary stenoses on cine angiogram and intravascular ultrasound. Significant stenoses in the coronary artery are seen as narrowings (black arrows) of the vessel. Intravascular ultrasound shows a normal segment of artery (A), areas with eccentric plaque (B, C), and near total obliteration of the lumen at the site of the significant stenosis (D). Note that the intravascular ultrasound catheter is present in the images as a black circle.

Intravascular Ultrasound, Fractional Flow Reserve, and Coronary Flow Reserve

During coronary angiography, intermediate stenoses (40–70%), indeterminate findings, or anatomic findings that are incongruous with the patient's symptoms may require further interrogation. In these cases, intravascular ultrasound provides a more accurate anatomic assessment of the coronary artery and the degree of coronary atherosclerosis **(Fig. 230-5)**. Intravascular ultrasound is performed using a small flexible catheter with a 40-mHz transducer at its tip that is advanced into the coronary artery over a guidewire. Data from intravascular ultrasound studies may be used to image atherosclerotic plaque precisely, determine luminal cross-sectional area, and measure vessel size; it is also used during or following percutaneous coronary intervention to assess the stenosis and determine the adequacy of stent placement. Measurement of the fractional flow reserve provides a functional assessment of the stenosis. The fractional flow reserve is the ratio of the pressure in the coronary artery distal to the stenosis divided by the pressure in the artery proximal to the stenosis at maximal vasodilation. Fractional flow reserve is measured using a coronary pressure–sensor guidewire at rest and at maximal hyperemia following the injection of adenosine. A fractional flow reserve of <0.75 indicates a hemodynamically significant stenosis that would benefit from intervention. Measurement of coronary flow reserve is another technique to assess the functional severity of a stenosis, although this technique is used with less frequency than fractional flow reserve. The coronary flow reserve is the maximal coronary blood flow increase above resting conditions during maximal vasodilation and is a measure of both epicardial coronary artery and microvascular function. Coronary flow reserve is determined using a Doppler flow guidewire before and after the administration of adenosine to induce hyperemia. A coronary flow reserve <2:1 after maximal hyperemia is considered abnormal.

Postprocedure Care

Once the procedure is completed, vascular access sheaths are removed. If the femoral approach is used, direct manual compression or vascular closure devices that immediately close the arteriotomy site with a staple/clip, collagen plug, or suture are used to achieve hemostasis. These devices decrease the length of bed rest (from 6 hours to 2–4 hours) and improve patient satisfaction but have not been shown definitively to be superior to manual compression with respect to access-site complications. When cardiac catheterization is performed as an elective outpatient procedure, the patient completes postprocedure bed rest in a monitored setting and is discharged home with instructions to liberalize fluids because contrast agents promote an osmotic diuresis, to avoid strenuous activity, and to observe the vascularaccess site for signs of complications. Overnight hospitalization may be required for high-risk patients with significant comorbidities, patients with complications occurring during the catheterization, or in patients who have undergone a percutaneous coronary intervention. Hypotension early after the procedure may be due to inadequate fluid

replacement or retroperitoneal bleeding from the access site.

FURTHER READINGS

Baim DS (ed): *Grossman's Cardiac Catheterization, Angiography, and Intervention*, 7th ed. Baltimore, Lippincott Williams & Wilkins, 2006

Kern MJ (ed): *Hemodynamic Rounds*: *Interpretation of Cardiac Pathophysiology From Pressure Waveform Analysis*, 3rd ed. Hoboken, Wiley-Blackwell, 2009

Nicholls SJ et al: Intravascular ultrasound-derived measures of cornoary atherosclerotic plaque burden and clinical outcome. JACC 55:2399, 2010[PMID: 20488313]

Ryan TJ: The coronary angiogram and its seminal contributions to cardiovascular medicine over five decades. Circulation 106:752, 2002[PMID: 12163439]

Tobis J: Assessment of intermediate severity coronary lesions in the catheterization laboratory. JACC 49:839-48, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HISTORY AND INTRODUCTION

The field of cardiac electrophysiology was ushered in with the development of the electrocardiogram (ECG) by Einthoven at the turn of the twentieth century. The recording of cellular membrane currents revealed that the body surface ECG is the timed sum of the cellular action potentials in the atria and ventricles. In the late 1960s, the development of intracavitary recording, in particular the ability to record His bundle electrograms with programmed stimulation of the heart, marked the beginning of contemporary clinical electrophysiology. Adoption of radio frequency technology to ablate cardiac tissue in the early 1990s heralded the birth of interventional cardiac electrophysiology.

The clinical problem of sudden death caused by ventricular arrhythmias, most commonly in the setting of coronary obstruction, was recognized as early as the late nineteenth century. The problem was vexing and led to the development of pharmacologic and nonpharmacologic therapies, including transthoracic defibrillators, cardiac massage, and, most recently, implantable defibrillators. Over time the limitations of antiarrhythmic drug therapy have been highlighted repeatedly in clinical trials, and now ablation and devices are first-line therapy for a number of cardiac arrhythmias.

In the last two decades, the genetic basis of a number of heritable arrhythmias has been elucidated, revealing important insights into the mechanisms not only of these rare arrhythmias but also of similar rhythm disturbances observed in more common forms of heart disease.

Descriptive Physiology

The normal cardiac impulse is generated by pacemaker cells in the sinoatrial node situated at the junction of the right atrium and the superior vena cava (see Fig. 228-1). This impulse is transmitted slowly through nodal tissue to the anatomically complex atria, where it is conducted more rapidly to the atrioventricular node (AVN), inscribing the P wave of the ECG (see Fig. 228-2). There is a perceptible delay in conduction through the anatomically and functionally heterogeneous AVN. The time needed for activation of the atria and the AVN delay is represented as the PR interval of the ECG. The AVN is the only electrical connection between the atria and the ventricles in the normal heart. The electrical impulse emerges from the AVN and is transmitted to the His-Purkinje system, specifically the common bundle of His, then the left and right bundle branches, and then to the Purkinje network, facilitating activation of ventricular muscle. In normal circumstances, the ventricles are activated rapidly in a well-defined fashion that is determined by the course of the Purkinje network, and this inscribes the QRS complex (see Fig. 228-2). Recovery of electrical excitability occurs more slowly and is governed by the time of activation and duration of regional action potentials. The relative brevity of epicardial action potentials in the ventricle results in repolarization that occurs first on the epicardial surface and then proceeds to the endocardium, which inscribes a T wave normally of the same polarity as the QRS complex. The duration of activation and recovery is determined by the action potential duration represented on the body surface ECG by the QT interval (see Fig. 228-2).

Cardiac myocytes exhibit a characteristically long action potential (200–400 ms) compared with neurons and skeletal muscle cells (1–5 ms). The action potential profile is sculpted by the orchestrated activity of multiple distinctive time- and voltage-dependent ionic currents **(Fig. 231-1*A*)**. The currents in turn are carried by transmembrane proteins that passively conduct ions down their electrochemical gradients through selective pores (ion channels), actively transport ions against their electrochemical gradient (pumps, transporters), or electrogenically exchange ionic species (exchangers).

Figure 231-1

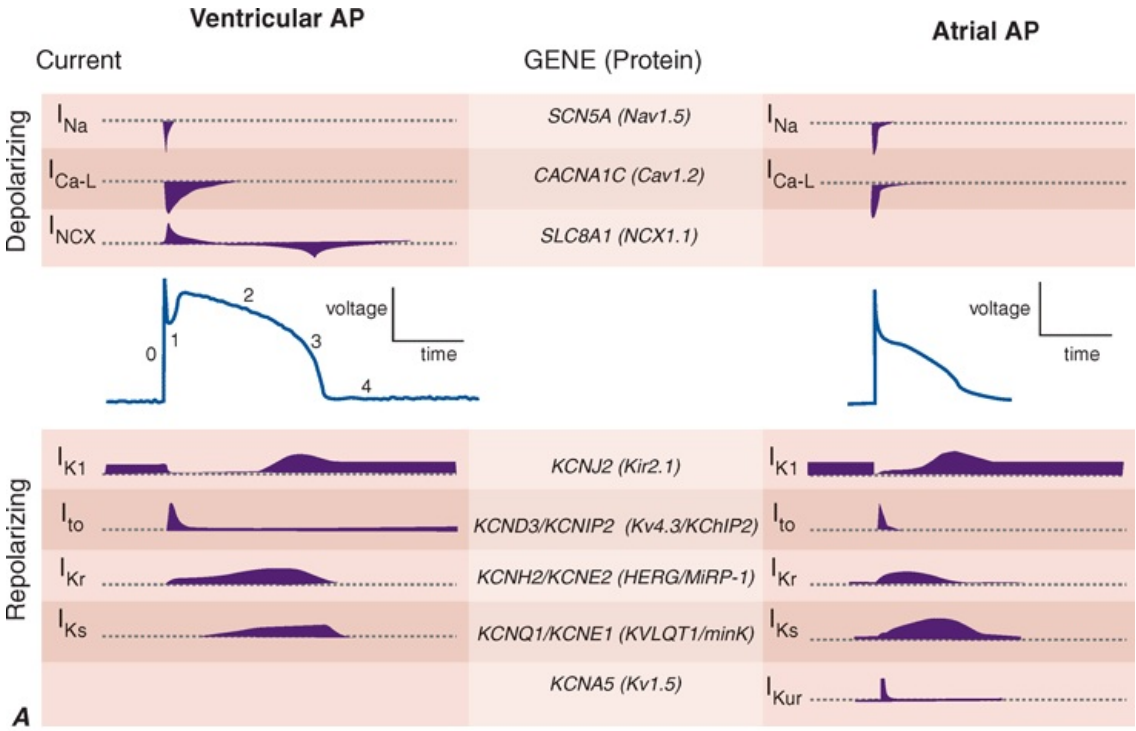

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

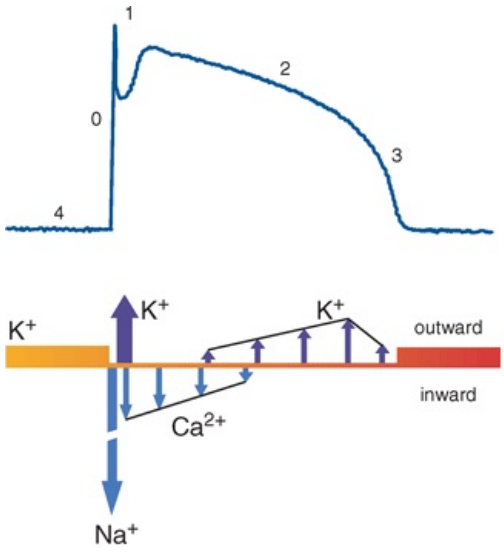

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Cellular atrial and ventricular action potentials. Phases 0–4 are the rapid upstroke, early repolarization, plateau, late repolarization, and diastole, respectively. The ionic currents and their respective genes are shown above and below the action potentials. The currents that underlie the action potentials vary in atrial and ventricular myocytes. ***B.*** A ventricular action potential with a schematic of the ionic currents flowing during the phases of the action potential. Potassium current (I_{K1}) is the principal current during phase 4 and determines the resting membrane potential of the myocyte. Sodium current generates the upstroke of the action potential (phase 0); activation of I_{to} with inactivation of the Na current inscribes early repolarization (phase 1). The plateau (phase 2) is generated by a balance of repolarizing potassium currents and depolarizing calcium current. Inactivation of the calcium current with persistent activation of potassium currents (predominantly I_{Kr} and I_{Ks}) causes phase 3 repolarization.

Action potentials in the heart are regionally distinct. The regional variability in cardiac action potentials is a result of differences in the number and types of ion channel proteins expressed by different cell

types in the heart. Further, unique sets of ionic currents are active in pacemaking and muscle cells, and the relative contributions of these currents may vary in the same cell type in different regions of the heart (Fig. 231-1*A*).

Ion channels are complex, multisubunit transmembrane glycoproteins that open and close in response to a number of biologic stimuli, including a change in membrane voltage, ligand binding (directly to the channel or to a G protein–coupled receptor), and mechanical deformation **(Fig. 231-2)**. Other ion motive exchangers and transporters contribute importantly to cellular excitability in the heart. Ion pumps establish and maintain the ionic gradients across the cell membrane that serve as the driving force for current flow through ion channels. Transporters or exchangers that do not move ions in an electrically neutral manner (e.g., the sodium-calcium exchanger transports three Na^+ for one Ca^{2+}) are termed *electrogenic* and contribute directly to the action potential profile.

Figure 231-2

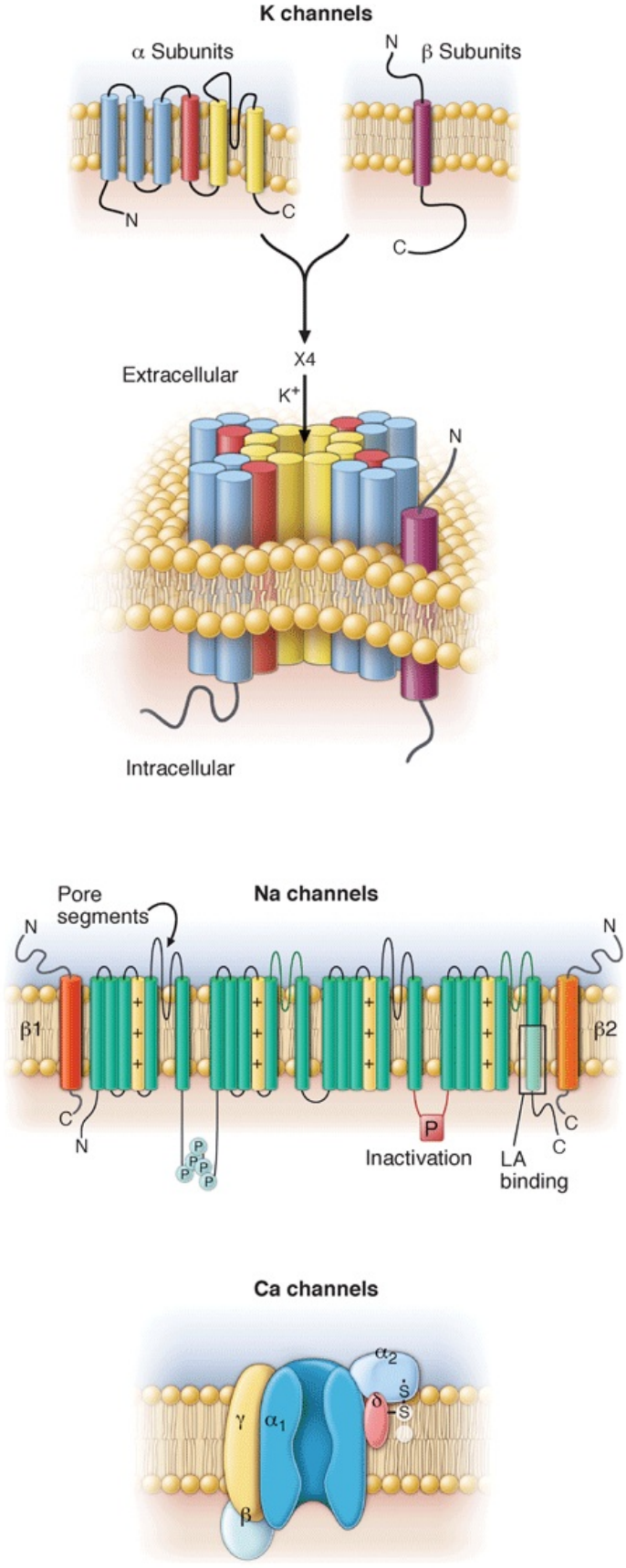

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Topology and subunit composition of the voltage-dependent ion channels. Potassium channels are formed by the tetramerization of α or pore-forming subunits and one or more β subunits; only single β subunits are shown for clarity. Sodium and calcium channels are composed of α subunits with four homologous domains and one or more ancillary subunits. In all channel types the loop of protein between the fifth and sixth membrane-spanning repeat in each subunit or domain forms the ion-selective pore. In the case of the sodium channel, the channel is a target for phosphorylation, the

ion selective pore. In the case of the sodium channel, this channel is a target for phosphorylation, the linker between the third and fourth homologous domain is critical to inactivation, and the sixth membrane-spanning repeat in the fourth domain is important in local anesthetic antiarrhythmic drug binding.

The most abundant superfamily of ion channels expressed in the heart is voltage gated. Several structural themes are common to all voltage-dependent ion channels. First, the architecture is modular, consisting either of four homologous subunits (e.g., K channels) or of four internally homologous domains (e.g., Na and Ca channels). Second, the proteins fold around a central pore lined by amino acids that exhibit exquisite conservation within a given channel family of like selectivity (e.g., jellyfish, eel, fruit fly, and human Na channels have very similar P segments). Third, the general strategy for activation gating (opening and closing in response to changes in membrane voltage) is highly conserved: the fourth transmembrane segment (S4), studded with positively charged residues, lies within the membrane field and moves in response to depolarization, opening the channel. Fourth, most ion channel complexes include not only the pore-forming proteins (α subunits) but also auxiliary subunits (e.g., β subunits) that modify channel function (Fig. 231-2).

Na and Ca channels are the primary carriers of depolarizing current in both the atria and the ventricles; inactivation of these currents and activation of repolarizing K currents hyperpolarize the heart cells, reestablishing the negative resting membrane potential **(Fig. 231-1*B*)**. The *plateau phase* is a time when little current is flowing, and relatively minor changes in depolarizing or repolarizing currents can have profound effects on the shape and duration of the action profile. Mutations in subunits of these channel proteins produce arrhythmogenic alterations in the action potentials that cause long and short QT syndrome, idiopathic ventricular fibrillation, familial atrial fibrillation, and some forms of conduction system disease.

Mechanisms of Cardiac Arrhythmias

Cardiac arrhythmias result from abnormalities of electrical impulse generation, conduction, or both. Bradyarrhythmias typically arise from disturbances in impulse formation at the level of the sinoatrial node or from disturbances in impulse propagation at any level, including exit block from the sinus node, conduction block in the AV node, and impaired conduction in the His-Purkinje system. Tachyarrhythmias can be classified according to mechanism, including enhanced automaticity (spontaneous depolarization of atrial, junctional, or ventricular pacemakers), reentry (circus propagation of a depolarizing wavefront), or triggered arrhythmias (initiated by afterdepolarizations) occurring during or immediately after cardiac repolarization, during phase 3 or 4 of the action potential. A variety of mapping and pacing maneuvers typically performed during invasive electrophysiologic testing can often determine the underlying mechanism of a tachyarrhythmia (**Table 231-1**).

Table 231–1. Arrhythmia Mechanisms

Electrophysiologic Property	Molecular Components	Mechanism	Prototypic Arrhythmias
Cellular			
Impulse Initiation			
Automaticity	I_f, $I_{Ca\text{-}L}$, $I_{Ca\text{-}T}$, I_K, I_{K1}	Suppression/acceleration of phase 4	Sinus bradycardia, sinus tachycardia
Triggered automaticity	Calcium overload, I_{TI}	DADs	Digitalis toxicity, reperfusion VT
	$I_{Ca\text{-}L}$, I_K, I_{Na}	EADs	Torsades des pointes, congenital and acquired
Excitation	I_{Na}	Suppression of phase 0	Ischemic VF
	$I_{K\text{-}ATP}$	AP shortening, inexcitability	
	$I_{Ca\text{-}L}$	Suppression	AV block
Repolarization	I_{Na}, $I_{Ca\text{-}L}$, I_K, I_{K1}, Ca^{2+} homeostasis	AP prolongation, EADs, DADs	Polymorphic VT (HF, LVH)
	$I_{Ca\text{-}L}$, K channels, Ca^{2+} homeostasis	AP shortening	Atrial fibrillation
Multicellular			
Cellular Coupling	Connexins (Cx43), I_{Na},	Decreased coupling	Ischemic VT/VF

	$I_{K\text{-}ATP}$		
Tissue Structure	Extracellular matrix, collagen	Excitable gap and functional reentry	Monomorphic VT, atrial fibrillation

Abbreviations: AP, action potential; AV, atrioventricular; DADs, delayed afterdepolarizations; EADs, early afterdepolarizations; HF, heart failure; LVH, left ventricular hypertrophy; VF, ventricular fibrillation; VT, ventricular tachyarrhythmia.

ALTERATIONS IN IMPULSE INITIATION: AUTOMATICITY

Spontaneous (phase 4) diastolic depolarization underlies the property of automaticity (pacemaking) characteristic of cells in the sinoatrial (SA) and atrioventricular (AV) nodes, His-Purkinje system, coronary sinus, and pulmonary veins. Phase 4 depolarization results from the concerted action of a number of ionic currents, including K^+ currents, Ca^{2+} currents, electrogenic Na, K-ATPase, the Na-Ca exchanger, and the so-called funny, or pacemaker, current (I_f); however, the relative importance of these currents remains controversial.

The rate of phase 4 depolarization and, therefore, the firing rates of pacemaker cells are dynamically regulated. Prominent among the factors that modulate phase 4 is autonomic nervous system tone. The negative chronotropic effect of activation of the parasympathetic nervous system is a result of the release of acetylcholine that binds to muscarinic receptors, releasing G protein $\beta\gamma$ subunits that activate a potassium current (I_{KACh}) in nodal and atrial cells. The resulting increase in K^+ conductance opposes membrane depolarization, slowing the rate of rise of phase 4 of the action potential. Conversely, augmentation of sympathetic nervous system tone increases myocardial catecholamine concentrations, which activate both α– and β–adrenergic receptors. The effect of β_1-adrenergic stimulation predominates in pacemaking cells, augmenting both L-type Ca current ($I_{Ca\text{-}L}$) and I_f, thus increasing the slope of phase 4. Enhanced sympathetic nervous system activity can dramatically increase the rate of firing of SA nodal cells, producing sinus tachycardia with rates >200 beats/min. By contrast, the increased rate of firing of Purkinje cells is more limited, rarely producing ventricular tachyarrhythmias >120 beats/min.

Normal automaticity may be affected by a number other factors associated with heart disease. Hypokalemia and ischemia may reduce the activity of Na, K-ATPase, thereby reducing the background repolarizing current and enhancing phase 4 diastolic depolarization. The end result would be an increase in the spontaneous firing rate of pacemaking cells. Modest increases in extracellular potassium may render the maximum diastolic potential more positive, thereby also increasing the firing rate of pacemaking cells. A more significant increase in $[K^+]_o$, however, renders the heart inexcitable by depolarizing the membrane potential.

Normal or enhanced automaticity of subsidiary latent pacemakers produces escape rhythms in the setting of failure of more dominant pacemakers. Suppression of a pacemaker cell by a faster rhythm leads to an increased intracellular Na^+ load ($[Na^+]_i$), and extrusion of Na^+ from the cell by Na, K-ATPase produces an increased background repolarizing current that slows phase 4 diastolic depolarization. At slower rates, $[Na^+]_i$ is decreased, as is the activity of the Na, K-ATPase, resulting in progressively more rapid diastolic depolarization and warm-up of the tachycardia rate. Overdrive suppression and warm-up are characteristic of, but may not be observed in, all automatic tachycardias. Abnormal conduction into tissue with enhanced automaticity (*entrance block*) may blunt or eliminate the phenomena of overdrive suppression and warm-up of automatic tissue.

Abnormal automaticity may underlie atrial tachycardia, accelerated idioventricular rhythms, and ventricular tachycardia, particularly that associated with ischemia and reperfusion. It has also been suggested that injury currents at the borders of ischemic myocardium may depolarize adjacent nonischemic tissue, predisposing to automatic ventricular tachycardia.

AFTERDEPOLARIZATIONS AND TRIGGERED AUTOMATICITY

Triggered automaticity or activity refers to impulse initiation that is dependent on afterdepolarizations **(Fig. 231-3)**. Afterdepolarizations are membrane voltage oscillations that occur during (early afterdepolarizations, EADs) or after (delayed afterdepolarizations, DADs) an action potential.

Figure 231-3

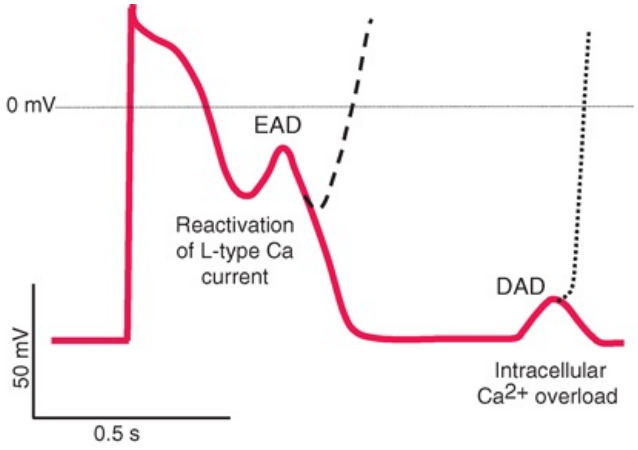

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic action potentials with early after depolarizations (EADs) and delayed afterdepolarizations (DADs). Afterdepolarizations are spontaneous depolarizations in cardiac myocytes. EADs occur before the end of the action potential (phases 2 and 3), interrupting repolarization. DADs occur during phase 4 of the action potential after completion of repolarization. The cellular mechanisms of EADs and DADs differ (see text).

The cellular feature common to the induction of DADs is the presence of an increased Ca^{2+} load in the cytosol and sarcoplasmic reticulum. Digitalis glycoside toxicity, catecholamines, and ischemia all can enhance Ca^{2+} loading sufficiently to produce DADs. Accumulation of lysophospholipids in ischemic myocardium with consequent Na^{+} and Ca^{2+} overload has been suggested as a mechanism for DADs and triggered automaticity. Cells from damaged areas or cells that survive a myocardial infarction may display spontaneous release of calcium from the sarcoplasmic reticulum, and this may generate "waves" of intracellular calcium elevation and arrhythmias.

EADs occur during the action potential and interrupt the orderly repolarization of the myocyte. Traditionally, EADs have been thought to arise from action potential prolongation and reactivation of depolarizing currents, but more recent experimental evidence suggests a previously unappreciated interrelationship between intracellular calcium loading and EADs. Cytosolic calcium may increase when action potentials are prolonged. This, in turn, appears to enhance L-type Ca current, further prolonging action potential duration as well as providing the inward current driving EADs. Intracellular calcium loading by action potential prolongation may also enhance the likelihood of DADs. The interrelationship among intracellular $[Ca^{2+}]$, EADs, and DADs may be one explanation for the susceptibility of hearts that are calcium loaded (e.g., in ischemia or congestive heart failure) to develop arrhythmias, particularly on exposure to action potential–prolonging drugs.

EAD-triggered arrhythmias exhibit rate dependence. In general, the amplitude of an EAD is augmented at slow rates when action potentials are longer. Indeed, a fundamental condition that underlies the development of EADs is action potential and QT prolongation. Hypokalemia, hypomagnesemia, bradycardia, and, most commonly, drugs can predispose to the generation of EADs, invariably in the context of prolonging the action potential. Antiarrhythmics with class IA and III action (see below) produce action potential and QT prolongation intended to be therapeutic but frequently causing arrhythmias. Noncardiac drugs such as phenothiazines, nonsedating antihistamines, and some antibiotics can also prolong the action potential duration and predispose to EAD-mediated triggered arrhythmias. Decreased $[K^{+}]_O$ paradoxically may decrease membrane potassium currents (particularly the delayed rectifier current, I_{Kr}) in the ventricular myocyte, explaining why hypokalemia causes action potential prolongation and EADs. In fact, potassium infusions in patients with the congenital long QT syndrome (LQTS) and in those with drug-induced acquired QT prolongation shorten the QT interval.

EAD-mediated triggered activity probably underlies initiation of the characteristic polymorphic ventricular tachycardia, torsades des pointes, seen in patients with congenital and acquired forms of LQTS. Structural heart disease such as cardiac hypertrophy and failure may also delay ventricular repolarization (so-called electrical remodeling) and predispose to arrhythmias related to abnormalities of repolarization. The abnormalities of repolarization in hypertrophy and heart failure are often magnified by concomitant drug therapy or electrolyte disturbances.

ABNORMAL IMPULSE CONDUCTION: REENTRY

The most common arrhythmia mechanism is reentry. Fundamentally, reentry is defined as circulation of an activation wave around an inexcitable obstacle. Thus, the requirements for reentry are two electrophysiologically dissimilar pathways for impulse propagation around an inexcitable region such

that unidirectional block occurs in one of the pathways and a region of excitable tissue exists at the head of the propagating wavefront **(Fig. 231-4)**. Structural and electrophysiologic properties of the heart may contribute to the development of the inexcitable obstacle and that of unidirectional block. The complex geometry of muscle bundles in the heart and spatial heterogeneity of cellular coupling or other active membrane properties (i.e., ionic currents) appear to be critical.

Figure 231-4

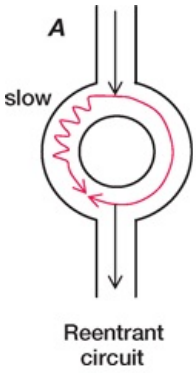

Reentrant circuit

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

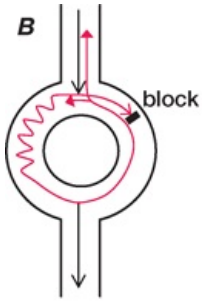

Initiation of reentry

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sustained reentry

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of reentry. *A.* The circuit contains two limbs, one with slow conduction. ***B.*** A premature impulse blocks in the fast pathway and conducts over the slow pathway, allowing the fast pathway to recover so that the activation wave can reenter the fast pathway from the retrograde direction. ***C.*** During sustained reentry utilizing such a circuit, a gap (excitable gap) exists between the activating head of the wave and the recovering tail. ***D.*** One mechanism of termination of reentry occurs when the conduction and recovery characteristics of the circuit change and the activating head of the wave collides with the tail, extinguishing the tachycardia.

A key feature in classifying reentrant arrhythmias, particularly for therapy, is the presence and size of an excitable gap. An excitable gap exists when the tachycardia circuit is longer than the tachycardia wavelength (λ = conduction velocity x refractory period, representing the size of the circuit that can sustain reentry), allowing appropriately timed stimuli to reset propagation in the circuit. Reentrant arrhythmias may exist in the heart in the absence of an excitable gap and with a tachycardia wavelength nearly the same size as the path length. In this case, the wavefront propagates through partially refractory tissue with no anatomic obstacle and no fully excitable gap; this is referred to as *leading circle reentry*, a form of functional reentry (reentry that depends on functional properties of the tissue). Unlike excitable gap reentry, there is no fixed anatomic circuit in leading circle reentry, and it may therefore not be possible to disrupt the tachycardia with pacing or destruction of a part of the circuit. Furthermore, the circuit in leading circle reentry tends to be less stable than that in excitable gap reentrant arrhythmias, with large variations in cycle length and a predilection to termination.

Anatomically determined, excitable gap reentry can explain several clinically important tachycardias, such as AV reentry, atrial flutter, bundle branch reentry ventricular tachycardia, and ventricular tachycardia in scarred myocardium. There is strong evidence to suggest that other, less organized arrhythmias, such as atrial and ventricular fibrillation, are associated with more complex activation of the heart and are due to functional reentry.

Structural heart disease is associated with changes in conduction and refractoriness that increase the risk of reentrant arrhythmias. Chronically ischemic myocardium exhibits a downregulation of the gap junction channel protein (connexin 43) that carries intercellular ionic current. The border zones of infarcted and failing ventricular myocardium exhibit not only functional alterations of ionic currents but also remodeling of tissue and altered distribution of gap junctions. The changes in gap junction channel expression and distribution, in combination with macroscopic tissue alterations, support a role for slowed conduction in reentrant arrhythmias that complicate chronic coronary artery disease (CAD). Aged human atrial myocardium exhibits altered conduction, manifest as highly fractionated atrial electrograms, producing an ideal substrate for the reentry that may underlie the very common development of atrial fibrillation in the elderly.

Approach to the Patient: Cardiac Arrhythmia

The evaluation of patients with suspected cardiac arrhythmias is highly individualized; however, two key features—the history and ECG—are pivotal in directing the diagnostic workup and therapy. Patients with cardiac arrhythmias exhibit a wide spectrum of clinical presentations that range from asymptomatic ECG abnormalities to survival from cardiac arrest. In general, the more severe the presenting symptoms are, the more aggressive the evaluation and treatment are. Loss of consciousness that is believed to be of cardiac origin typically mandates an exhaustive search for the etiology and often requires invasive, device-based therapy. The presence of structural heart disease and prior myocardial infarction dictates a change in the approach to the management of syncope or ventricular arrhythmias. The presence of a family history of serious ventricular arrhythmias or premature sudden death will influence the evaluation of presumed heritable arrhythmias.

The physical examination is focused on determining whether there is cardiopulmonary disease that is associated with specific cardiac arrhythmias. The absence of significant cardiopulmonary disease often, but not always, suggests benignity of the rhythm disturbance. In contrast, palpitations, syncope, or near syncope in the setting of significant heart or lung disease have more ominous implications. In addition, the physical examination may reveal the presence of a persistent arrhythmia such as atrial

fibrillation.

The judicious use of noninvasive diagnostic tests is an important element in the evaluation of patients with arrhythmias, and there is no test more important than the ECG, particularly if recorded at the time of symptoms. Uncommon but diagnostically important signatures of electrophysiologic disturbances may be unearthed on the resting ECG, such as delta waves in Wolff-Parkinson-White (WPW) syndrome, prolongation or shortening of the QT interval, right precordial ST-segment abnormalities in Brugada syndrome, and epsilon waves in arrhythmogenic right ventricular dysplasia. Variants of body surface ECG recording can provide important information about arrhythmia substrates and triggers. Holter monitoring and event recording, either continuous or intermittent, record the body surface ECG over longer periods, enhancing the possibility of observing the cardiac rhythm during symptoms. Implantable long-term monitors and commercial ambulatory ECG monitoring services permit prolonged telemetric monitoring both for diagnosis and to assess the efficacy of therapy.

Long-term recordings permit the assessment of the time-varying behavior of the heart rhythm. Heart rate variability (HRV) and QT interval variability (QTV) provide noninvasive methods to assess autonomic nervous system influence on the heart. A decrease in HRV has been associated with increased sympathetic nervous system tone and increased mortality rates in patients after myocardial infarction. Signal-averaged electrocardiography (SAECG) uses signal-averaging techniques to amplify small potentials in the body surface ECG that are associated with slow conduction in the myocardium. The presence of these small potentials, referred to as *late potentials* because of their timing with respect to the QRS complex, and prolongation of the filtered (or averaged) QRS duration are indicative of slowed conduction in the ventricle and have been associated with an increased risk of ventricular arrhythmias after myocardial infarction. Exercise electrocardiography is important in determining the presence of myocardial demand ischemia; more recently, analysis of the morphology of the QT interval with exercise has been used to assess the risk of serious ventricular arrhythmias. Microscopic alterations in the T wave (T wave alternans, TWA) at low heart rates may identify patients at risk for ventricular arrhythmias. Cardiac imaging plays an important role in the detection and characterization of myocardial structural abnormalities that may render the heart more susceptible to arrhythmia. Ventricular tachyarrhythmias, for instance, occur more frequently in patients with ventricular systolic dysfunction and chamber dilation, in hypertrophic cardiomyopathy, and in the setting of infiltrative diseases such as sarcoidosis. Supraventricular arrhythmias may be associated with particular congenital conditions, including AV reentry in the setting of Ebstein's anomaly. Echocardiography is a frequently employed imaging technique to screen for disorders of cardiac structure and function. Increasingly, magnetic resonance (MR) imaging of the myocardium is being used to screen for scar burden, fibrofatty infiltration of the myocardium as seen in arrhythmogenic right ventricular cardiomyopathy, and other structural changes that affect arrhythmia susceptibility.

Head-up tilt (HUT) testing is useful in the evaluation of some patients with syncope. The physiologic response to HUT is incompletely understood; however, redistribution of blood volume and increased ventricular contractility occur consistently. Exaggerated activation of a central reflex in response to HUT produces a stereotypic response of an initial increase in heart rate, then a drop in blood pressure followed by a reduction in heart rate characteristic of neurally mediated hypotension. Other responses to HUT may be observed in patients with orthostatic hypotension and autonomic insufficiency. HUT is used most often in patients with recurrent syncope, although it may be useful in patients with single syncopal episodes with associated injury, particularly in the absence of structural heart disease. In patients with structural heart disease, HUT may be indicated in those with syncope, in whom other causes (e.g., asystole, ventricular tachyarrhythmias) have been excluded. HUT has been suggested as a useful tool in the diagnosis of and therapy for recurrent idiopathic vertigo, chronic fatigue syndrome, recurrent transient ischemic attacks, and repeated falls of unknown etiology in the elderly. Importantly, HUT is relatively contraindicated in the presence of severe CAD with proximal coronary stenoses, known severe cerebrovascular disease, severe mitral stenosis, and obstruction to left ventricular outflow (e.g., aortic stenosis). The method of HUT is variable, but the angle of tilt and the duration of upright posture are central to the diagnostic utility of the test. Pharmacologic provocation of orthostatic stress with isoproterenol, nitrates, adenosine, and edrophonium has been used to shorten the test and enhance specificity.

Electrophysiologic testing is central to the understanding and treatment of many cardiac arrhythmias. Indeed, most frequently, electrophysiologic testing is interventional, providing both diagnosis and therapy. The components of the electrophysiologic test are baseline measurements of conduction under resting and stressed (rate or pharmacologic) conditions and maneuvers, both pacing and pharmacologic, to induce arrhythmias. A number of sophisticated electrical mapping and catheter-guidance techniques have been developed to facilitate catheter-based therapeutics in the electrophysiology laboratory.

Treatment: Cardiac Arrhythmias

ANTIARRHYTHMIC DRUG THERAPY

The interaction of antiarrhythmic drugs with cardiac tissues and the resulting electrophysiologic changes are complex. An incomplete understanding of the effects of these drugs has produced serious

missteps that have had adverse effects on patient outcomes and the development of newer pharmacologic agents. Currently, antiarrhythmic drugs have been relegated to an ancillary role in the treatment of most cardiac arrhythmias.

There are several explanations for the complexity of antiarrhythmic drug action: the structural similarity of target ion channels; regional differences in the levels of expression of channels and transporters, which change with disease; time and voltage dependence of drug action; and the effect of these drugs on targets other than ion channels. Because of the limitations of any scheme to classify antiarrhythmic agents, a shorthand that is useful in describing the major mechanisms of action is of some utility. Such a classification scheme was proposed in 1970 by Vaughan-Williams and later modified by Singh and Harrison. The classes of antiarrhythmic action are class I, local anesthetic effect due to blockade of Na^+ current; class II, interference with the action of catecholamines at the β-adrenergic receptor; class III, delay of repolarization due to inhibition of K^+ current or activation of depolarizing current; class IV, interference with calcium conductance **(Table 231-2)**. The limitations of the Vaughan-Williams classification scheme include multiple actions of most drugs, overwhelming consideration of antagonism as a mechanism of action, and the fact that several agents have none of the four classes of action in the scheme.

Table 231–2. Antiarrhythmic Drug Actions

	Class Actions				
Drug	**I**	**II**	**III**	**IV**	**Miscellaneous Action**
Quinidine	++		++		α-Adrenergic blockade
Procainamide	++		++		Ganglionic blockade
Flecainide	+++		+		
Propafenone	++	+			
Sotalol		++	+++		
Dofetilide			+++		
Amiodarone	++	++	+++	+	α-Adrenergic blockade
Ibutilide			+++		Na+ channel activator

CATHETER ABLATION

The use of catheter ablation is based on the principle that there is a critical anatomic region of impulse generation or propagation that is required for the initiation and maintenance of cardiac arrhythmias. Destruction of such a critical region results in the elimination of the arrhythmia. The use of radio frequency (RF) energy in clinical medicine is nearly a century old. The first catheter ablation using a DC energy source was performed in the early 1980s by Scheinman and colleagues. By the early 1990s, RF had been adapted for use in catheter-based ablation in the heart **(Fig. 231-5)**.

Figure 231-5

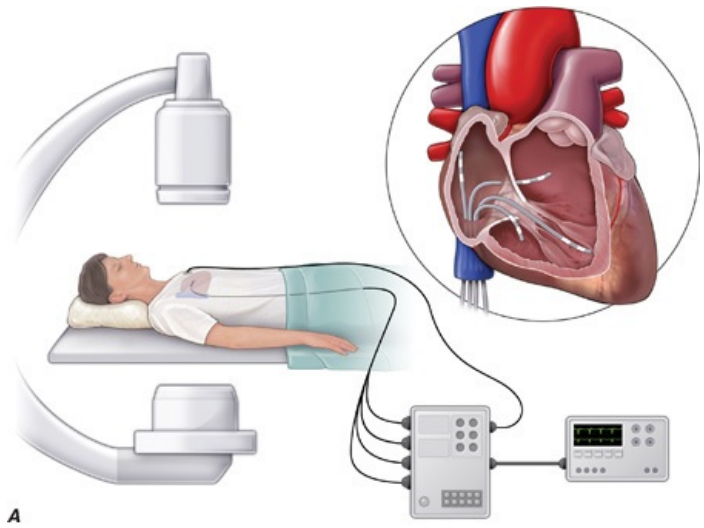

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

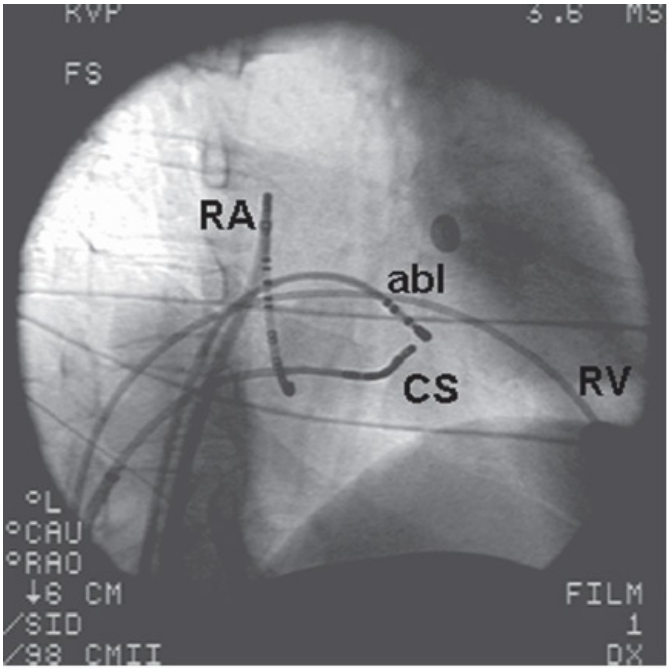

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Catheter ablation of cardiac arrhythmias. ***A.*** A schematic of the catheter system and generator in a patient undergoing radiofrequency catheter ablation (RFCA); the circuit involves the catheter in the heart and a dispersive patch placed on the body surface (usually the back). The inset shows a diagram of the heart with a catheter located at the AV valve ring for ablation of an accessory pathway. ***B.*** A right anterior oblique fluoroscopic image of the catheter position for ablation of a left-sided accessory pathway. A catheter is placed in the atrial side of the mitral valve ring (abl) via a transseptal puncture. Other catheters are placed in the coronary sinus, in the right atrium (RA), and in the right ventricular (RV) apex to record local electrical activation. ***C.*** Body surface ECG recordings (I, II, V_1) and endocardial electrograms (HRA, high right atrium; HISp, proximal His bundle electrogram; CS 7,8, recordings from poles 7 and 8 of a decapolar catheter placed in the coronary sinus) during RFCA of a left-sided accessory pathway in a patient with Wolff-Parkinson-White syndrome. The QRS narrows at the fourth complex; the arrow shows the His bundle electrogram, which becomes apparent with elimination of ventricular preexcitation over the accessory pathway.

The RF frequency band (300–30,000 kHz) is used to generate energy for several biomedical applications, including coagulation and cauterization of tissues. Energy of this frequency will not stimulate skeletal muscle or the heart and heats tissue by a resistive mechanism, with the intensity of heating and tissue destruction being proportional to the delivered power. Alternative, less frequently used energy sources for catheter ablation of cardiac arrhythmias include microwaves (915 MHz or 2450 MHz), lasers, ultrasound, and freezing (cryoablation). Of these alternative ablation techniques, cryoablation is being used clinically with the most frequency, especially ablation in the region of the AV node. At temperatures just below 32°C, membrane ion transport is disrupted, producing depolarization of cells, decreased action potential amplitude and duration, and slowed conduction velocity (resulting in local conduction block)—all of which are reversible if the tissue is rewarmed in a timely fashion. Tissue cooling can be used for mapping and ablation. Cryomapping can be used to confirm the location of a desired ablation target, such as an accessory pathway in WPW syndrome, or can be

used to determine the safety of ablation around the AV node by monitoring AV conduction during cooling. Another advantage of cryoablation is that once the catheter tip cools below freezing, it adheres to the tissue, increasing catheter stability independent of the rhythm or pacing.

DEVICE THERAPY

Bradyarrhythmias due either to primary sinus node dysfunction or to atrioventricular conduction defects are readily treated through implantation of a permanent pacemaker. Clinical indications for pacemaker implantation often depend on the presence either of symptomatic bradycardia or of an unreliable endogenous escape rhythm and are more fully reviewed in Chap. 232.

Ventricular tachyarrhythmias, particularly those occurring in the context of progressive structural heart diseases such as ischemic cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy, may recur despite therapy with antiarrhythmic drugs or catheter ablation. In appropriate candidates, implantation of an internal cardioverter-defibrillator (ICD) may reduce mortality rates from sudden cardiac death. In a subset of patients with congestive heart failure (CHF) and ventricular mechanical dyssynchrony, ICD or pacemaker platforms can be used to provide cardiac resynchronization therapy, typically through implantation of a left ventricular pacing lead. In patients with dyssynchronous CHF, such therapy has been shown to improve both morbidity and mortality rates.

FURTHER READINGS

Akar FG, Tomaselli GF: Genetic basis of cardiac arrhythmias, in *Hurst's The Heart*, 12th ed, V Fuster et al (eds). New York, McGraw-Hill, 2007

Hille B: *Ion Channels of Excitable Membranes*, 3rd ed. Sunderland, MA, Sinauer Associates, 2001

Josephson ME: *Clinical Cardiac Electrophysiology Techniques and Interpretations*, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2008

Saksena S, Camm AJ (eds): *Electrophysiological Disorders of the Heart* . Philadelphia, Elsevier Churchill Livingstone, 2005

Zipes DP, Jalife J (eds): *Cardiac Electrophysiology: From Cell to Bedside*, 5th ed. Philadelphia, Elsevier, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE BRADYARRHYTHMIAS: INTRODUCTION

Electrical activation of the heart normally originates in the sinoatrial (SA) node, the predominant pacemaker. Other subsidiary pacemakers in the atrioventricular (AV) node, specialized conducting system, and muscle may initiate electrical activation if the SA node is dysfunctional or suppressed. Typically, subsidiary pacemakers discharge at a slower rate and, in the absence of an appropriate increase in stroke volume, may result in tissue hypoperfusion.

Spontaneous activation and contraction of the heart are a consequence of the specialized pacemaking tissue in these anatomic locales. As described in Chap. 231, action potentials in the heart are regionally heterogeneous. The action potentials in cells isolated from nodal tissue are distinct from those recorded from atrial and ventricular myocytes **(Fig. 232-1)**. The complement of ionic currents present in nodal cells results in a less negative resting membrane potential compared with atrial or ventricular myocytes. Electrical diastole in nodal cells is characterized by slow diastolic depolarization (phase 4), which generates an action potential as the membrane voltage reaches threshold. The action potential upstrokes (phase 0) are slow compared with atrial or ventricular myocytes, being mediated by calcium rather than sodium current. Cells with properties of SA and AV nodal tissue are electrically connected to the remainder of the myocardium by cells with an electrophysiologic phenotype between that of nodal cells and that of atrial or ventricular myocytes. Cells in the SA node exhibit the most rapid phase 4 depolarization and thus are the dominant pacemakers in a normal heart.

Figure 232-1

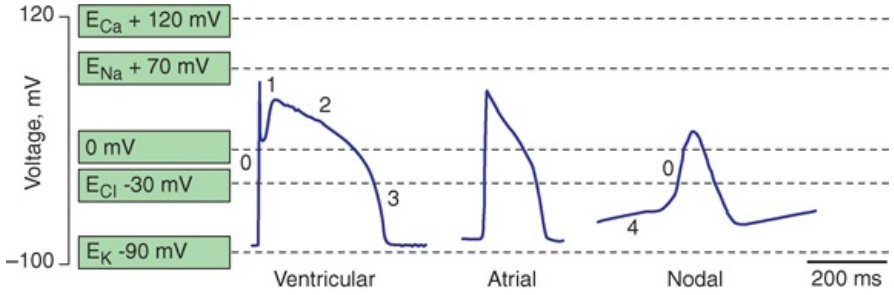

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Action potential profiles recorded in cells isolated from sinoatrial or atrioventricular nodal tissue compared with those of cells from atrial or ventricular myocardium. Nodal cell action potentials exhibit more depolarized resting membrane potentials, slower phase 0 upstrokes, and phase 4 diastolic depolarization.

Bradycardia results from a failure of either impulse initiation or impulse conduction. Failure of impulse initiation may be caused by depressed automaticity resulting from a slowing or failure of phase 4 diastolic depolarization **(Fig. 232-2)**, which may result from disease or exposure to drugs. Prominently, the autonomic nervous system modulates the rate of phase 4 diastolic depolarization and thus the firing rate of both primary (SA node) and subsidiary pacemakers. Failure of conduction of an impulse from nodal tissue to atrial or ventricular myocardium may produce bradycardia as a result of exit block. Conditions that alter the activation and connectivity of cells (e.g., fibrosis) in the heart may result in failure of impulse conduction.

Figure 232-2

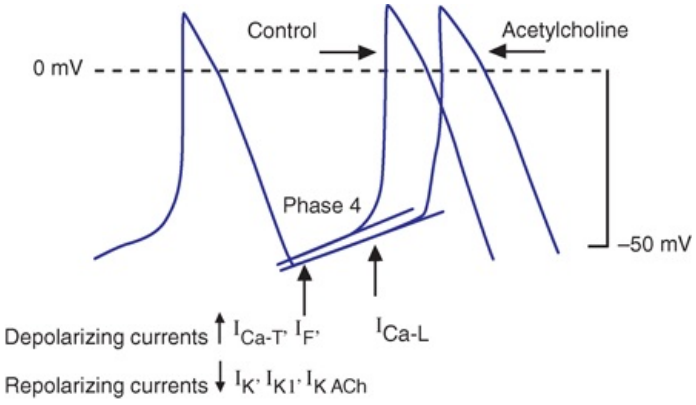

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematics of nodal action potentials and the currents that contribute to phase 4 depolarization. Relative increases in depolarizing L- ($I_{Ca\text{-}L}$) and T- ($I_{Ca\text{-}T}$) type calcium and pacemaker currents (I_f) along with a reduction in repolarizing inward rectifier (I_{K1}) and delayed rectifier (I_K) potassium currents result in depolarization. Activation of ACh-gated (I_{KACh}) potassium current and beta blockade slow the rate of phase 4 and decrease the pacing rate. *(Modified from J Jalife et al: Basic Cardiac Electrophysiology for the Clinician, Blackwell Publishing, 1999.*

SA node dysfunction and AV conduction block are the most common causes of pathologic bradycardia. SA node dysfunction may be difficult to distinguish from physiologic sinus bradycardia, particularly in the young. SA node dysfunction increases in frequency between the fifth and sixth decades of life and should be considered in patients with fatigue, exercise intolerance, or syncope and sinus bradycardia. Transient AV block is common in the young and probably is a result of the high vagal tone found in up to 10% of young adults. Acquired and persistent failure of AV conduction is decidedly rare in healthy adult populations, with an estimated incidence of ~200/million population per year.

Permanent pacemaking is the only reliable therapy for symptomatic bradycardia in the absence of extrinsic and reversible etiologies such as increased vagal tone, hypoxia, hypothermia, and drugs **(Table 232-1)**. Approximately 50% of the 150,000 permanent pacemakers implanted in the United States and 20–30% of the 150,000 of those in Europe were implanted for SA node disease.

Table 232–1. Etiologies of SA Node Dysfunction

Extrinsic	Intrinsic
Autonomic	Sick-sinus syndrome (SSS)
Carotid sinus hypersensitivity	Coronary artery disease (chronic and acute MI)
Vasovagal (cardioinhibitory) stimulation	Inflammatory
Drugs	Pericarditis
Beta blockers	Myocarditis (including viral)
Calcium channel blockers	Rheumatic heart disease
Digoxin	Collagen vascular diseases
Antiarrhythmics (class I and III)	Lyme disease
Adenosine	Senile amyloidosis
Clonidine (other sympatholytics)	Congenital heart disease
Lithium carbonate	TGA/Mustard and Fontan repairs
Cimetidine	Iatrogenic
Amitriptyline	Radiation therapy
Phenothiazines	Postsurgical
Narcotics (methadone)	Chest trauma
Pentamidine	Familial
Hypothyroidism	AD SSS, OMIM #163800 (15q24-25)
Sleep apnea Hypoxia	AR SSS, OMIM #608567 (3p21)
Endotracheal suctioning (vagal maneuvers)	SA node disease with myopia, OMIM 182190
Hypothermia	Kearns-Sayre syndrome, OMIM #530000
Increased intracranial pressure	Myotonic dystrophy
	Type 1, OMIM #160900 (19q13.2-13.3)

	Type 2, OMIM #602668 (3q13.3-q24)
	Friedreich's ataxia, OMIM #229300 9q13, 9p23-p11)

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; MI, myocardial infarction; OMIM, Online Mendelian Inheritance in Man (database); TGA, transposition of the great arteries;.

SA Node Disease

STRUCTURE AND PHYSIOLOGY OF THE SA NODE

The SA node is composed of a cluster of small fusiform cells in the sulcus terminalis on the epicardial surface of the heart at the right atrial–superior vena caval junction, where they envelop the SA nodal artery. The SA node is structurally heterogeneous, but the central prototypic nodal cells have fewer distinct myofibrils than does the surrounding atrial myocardium, no intercalated disks visible on light microscopy, a poorly developed sarcoplasmic reticulum, and no T-tubules. Cells in the peripheral regions of the SA node are transitional in both structure and function. The SA nodal artery arises from the right coronary artery in 55–60% and the left circumflex artery in 40–45% of persons. The SA node is richly innervated by sympathetic and parasympathetic nerves and ganglia.

Irregular and slow propagation of impulses from the SA node can be explained by the electrophysiology of nodal cells and the structure of the SA node itself. The action potentials of SA nodal cells are characterized by a relatively depolarized membrane potential (Fig. 232-1) of –40 to –60 mV, slow phase 0 upstroke, and relatively rapid phase 4 diastolic depolarization compared with the action potentials recorded in cardiac muscle cells. The relative absence of inward rectifier potassium current (I_{K1}) accounts for the depolarized membrane potential; the slow upstroke of phase 0 results from the absence of available fast sodium current (I_{Na}) and is mediated by L-type calcium current (I_{Ca-L}); and phase 4 depolarization is a result of the aggregate activity of a number of ionic currents. Prominently, both L- and T-type (I_{Ca-T}) calcium currents, the pacemaker current (so-called funny current, or I_f) formed by the tetramerization of hyperpolarization-activated cyclic nucleotide-gated channels, and the electrogenic sodium-calcium exchanger provide depolarizing current that is antagonized by delayed rectifier (I_{Kr}) and acetylcholine-gated ($IKAC_h$) potassium currents. I_{Ca-L}, I_{Ca-T}, and I_f are modulated by β-adrenergic stimulation and $IKAC_h$ by vagal stimulation, explaining the exquisite sensitivity of diastolic depolarization to autonomic nervous system activity. The slow conduction within the SA node is explained by the absence of I_{Na} and poor electrical coupling of cells in the node, resulting from sizable amounts of interstitial tissue and a low abundance of gap junctions. The poor coupling allows for graded electrophysiologic properties within the node, with the peripheral transitional cells being silenced by electrotonic coupling to atrial myocardium.

ETIOLOGY OF SA NODAL DISEASE

SA nodal dysfunction has been classified as intrinsic or extrinsic. The distinction is important because extrinsic dysfunction is often reversible and generally should be corrected before pacemaker therapy is considered (Table 232-1). The most common causes of extrinsic SA node dysfunction are drugs and autonomic nervous system influences that suppress automaticity and/or compromise conduction. Other extrinsic causes include hypothyroidism, sleep apnea, and conditions likely to occur in critically ill patients such as hypothermia, hypoxia, increased intracranial pressure (Cushing's response), and endotracheal suctioning via activation of the vagus nerve.

Intrinsic sinus node dysfunction is degenerative and often is characterized pathologically by fibrous replacement of the SA node or its connections to the atrium. Acute and chronic coronary artery disease (CAD) may be associated with SA node dysfunction, although in the setting of acute myocardial infarction (MI; typically inferior), the vabnormalities are transient. Inflammatory processes may alter SA node function, ultimately producing replacement fibrosis. Pericarditis, myocarditis, and rheumatic heart disease have been associated with SA nodal disease with sinus bradycardia, sinus arrest, and exit block. Carditis associated with systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and mixed connective tissue disorders (MCTDs) may also affect SA node structure and function. Senile amyloidosis is an infiltrative disorder in patients typically in the ninth decade of life; deposition of amyloid protein in the atrial myocardium can impair SA node function. Some SA node disease is iatrogenic and results from direct injury to the SA node during cardiothoracic surgery.

Rare heritable forms of sinus node disease have been described, and several have been characterized genetically. Autosomal dominant sinus node dysfunction in conjunction with supraventricular tachycardia [(i.e., tachycardia-bradycardia variant of sick-sinus syndrome (SSS)] has been linked to mutations in the pacemaker current (I_f) subunit gene *HCN4* on chromosome 15. An autosomal recessive form of SSS with the prominent feature of atrial inexcitability and absence of P waves on the electrocardiogram (ECG) is caused by mutations in the cardiac sodium channel gene, *SCN5A*, on chromosome 3. SSS associated with myopia has been described but not genetically characterized. There are several neuromuscular diseases, including Kearns-Sayre syndrome (ophthalmoplegia, pigmentary degeneration of the retina, and cardiomyopathy) and myotonic

dystrophy, that have a predilection for the conducting system and SA node.

SSS in both the young and the elderly is associated with an increase in fibrous tissue in the SA node. The onset of SSS may be hastened by coexisting disease, such as CAD, diabetes mellitus, hypertension, and valvular diseases and cardiomyopathies.

CLINICAL FEATURES OF SA NODE DISEASE

SA node dysfunction may be completely asymptomatic and manifest as an ECG anomaly such as sinus bradycardia; sinus arrest and exit block; or alternating supraventricular tachycardia, usually atrial fibrillation, and bradycardia. Symptoms associated with SA node dysfunction, in particular tachycardia-bradycardia syndrome, may be related to both slow and fast heart rates. For example, tachycardia may be associated with palpitations, angina pectoris, and heart failure, and bradycardia may be associated with hypotension, syncope, presyncope, fatigue, and weakness. In the setting of SSS, overdrive suppression of the SA node may result in prolonged pauses and syncope upon termination of the tachycardia. In many cases, symptoms associated with SA node dysfunction result from concomitant cardiovascular disease. A significant minority of patients with SSS develop signs and symptoms of heart failure that may be related to slow or fast heart rates.

One-third to one-half of patients with SA node dysfunction develop supraventricular tachycardia, usually atrial fibrillation or atrial flutter. The incidence of persistent atrial fibrillation in patients with SA node dysfunction increases with advanced age, hypertension, diabetes mellitus, left ventricular dilation, valvular heart disease, and ventricular pacing. Remarkably, some symptomatic patients may experience an improvement in symptoms with the development of atrial fibrillation, presumably from an increase in their average heart rate. Patients with the tachycardia-bradycardia variant of SSS, similar to patients with atrial fibrillation, are at risk for thromboembolism, and *those at greatest risk*, including patients ≥65 years and patients with a prior history of stroke, valvular heart disease, left ventricular dysfunction, or atrial enlargement, should be treated with anticoagulants. Up to one-quarter of patients with SA node disease will have concurrent AV conduction disease, although only a minority will require specific therapy for high-grade AV block.

The natural history of SA node dysfunction is one of varying intensity of symptoms even in patients who present with syncope. Symptoms related to SA node dysfunction may be significant, but overall mortality usually is not compromised in the absence of other significant comorbid conditions. These features of the natural history need to be taken into account in considering therapy for these patients.

ELECTROCARDIOGRAPHY OF SA NODE DISEASE

The electrocardiographic manifestations of SA node dysfunction include sinus bradycardia, sinus pauses, sinus arrest, sinus exit block, tachycardia (in SSS), and chronotropic incompetence. It is often difficult to distinguish pathologic from physiologic sinus bradycardia. By definition, sinus bradycardia is a rhythm driven by the SA node with a rate of <60 beats/min; sinus bradycardia is very common and typically benign. Resting heart rates <60 beats/min are very common in young healthy individuals and physically conditioned subjects. A sinus rate of <40 beats/min in the awake state in the absence of physical conditioning generally is considered abnormal. Sinus pauses and sinus arrest result from failure of the SA node to discharge, producing a pause without P waves visible on the ECG **(Fig. 232-3)**. Sinus pauses of up to 3 s are common in awake athletes, and pauses of this duration or longer may be observed in asymptomatic elderly subjects. Intermittent failure of conduction from the SA node produces sinus exit block. The severity of sinus exit block may vary in a manner similar to that of AV block (see below). Prolongation of conduction from the sinus node will not be apparent on the ECG; second-degree SA block will produce intermittent conduction from the SA node and a regularly irregular atrial rhythm.

Figure 232-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus slowing and pauses on the ECG. The ECG is recorded during sleep in a young patient without heart disease. The heart rate before the pause is slow, and the PR interval is prolonged, consistent with an increase in vagal tone. The P waves have a morphology consistent with sinus rhythm. The recording is from a two-lead telemetry system in which the tracing labeled II mimics frontal lead II and V represents Modified Central Lead 1, which mimics lead V1 of the standard 12-lead ECG.

Type I second-degree SA block results from progressive prolongation of SA node conduction with intermittent failure of the impulses originating in the sinus node to conduct to the surrounding atrial tissue. Second-degree SA block appears on the ECG as an intermittent absence of P waves **(Fig. 232-4)**. In type II second-degree SA block, there is no change in SA node conduction before the pause. Complete or third-degree SA block results in no P waves on the ECG. Tachycardia-bradycardia syndrome is manifest as alternating sinus bradycardia and atrial tachyarrhythmias. Although atrial tachycardia, atrial flutter, and atrial fibrillation may be observed, the latter is the most common tachycardia. Chronotropic incompetence is the inability to increase the heart rate in response to exercise or other stress appropriately and is defined in greater detail below.

Figure 232-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mobitz type I SA nodal exit block. A theoretical SA node electrogram (SAN EG) is shown. Note that there is grouped beating producing a regularly irregular heart rhythm. The SA node EG rate is constant with progressive delay in exit from the node and activation of the atria, inscribing the P wave. This produces subtly decreasing P-P intervals before the pause, and the pause is less than twice the cycle length of the last sinus interval.

DIAGNOSTIC TESTING

SA node dysfunction is most commonly a clinical or electrocardiographic diagnosis. Sinus bradycardia or pauses on the resting ECG are rarely sufficient to diagnose SA node disease, and longer-term recording and symptom correlation generally are required. Symptoms in the absence of sinus bradyarrhythmias may be sufficient to exclude a diagnosis of SA node dysfunction.

Electrocardiographic recording plays a central role in the diagnosis and management of SA node dysfunction. Despite the limitations of the resting ECG, longer-term recording employing Holter or event monitors may permit correlation of symptoms with the cardiac rhythm. Many contemporary event monitors may be automatically triggered to record the ECG when certain programmed heart rate criteria are met. Implantable ECG monitors permit long-term recording (12–18 months) in particularly challenging patients.

Failure to increase the heart rate with exercise is referred to as *chronotropic incompetence*. This is alternatively defined as failure to reach 85% of predicted maximal heart rate at peak exercise or failure to achieve a heart rate >100 beats/min with exercise or a maximal heart rate with exercise less than two standard deviations below that of an age-matched control population. Exercise testing may be useful in discriminating chronotropic incompetence from resting bradycardia and may aid in the identification of the mechanism of exercise intolerance.

Autonomic nervous system testing is useful in diagnosing carotid sinus hypersensitivity; pauses >3 s are consistent with the diagnosis but may be present in asymptomatic elderly subjects. Determining the intrinsic heart rate (IHR) may distinguish SA node dysfunction from slow heart rates that result from high vagal tone. The normal IHR after administration of 0.2 mg/kg propranolol and 0.04 mg/kg atropine is 117.2 – (0.53 x age) in beats/min; a low IHR is indicative of SA disease.

Electrophysiologic testing may play a role in the assessment of patients with presumed SA node dysfunction and in the evaluation of syncope, particularly in the setting of structural heart disease. In this circumstance, electrophysiologic testing is used to rule out more malignant etiologies of syncope, such as ventricular tachyarrhythmias and AV conduction block. There are several ways to assess SA node function invasively. They include the sinus node recovery time (SNRT), defined as the longest pause after cessation of overdrive pacing of the right atrium near the SA node (normal: <1500 ms or, corrected for sinus cycle length, <550 ms), and the sinoatrial conduction time (SACT), defined as one-half the difference between the intrinsic sinus cycle length and a noncompensatory pause after a premature atrial stimulus (normal <125 ms). The combination of an abnormal SNRT, an abnormal SACT, and a low IHR is a sensitive and specific indicator of intrinsic SA node disease.

TREATMENT: SINOATRIAL NODE DYSFUNCTION

Since SA node dysfunction is not associated with increased mortality rates, the aim of therapy is alleviation of symptoms. Exclusion of extrinsic causes of SA node dysfunction and correlation of the cardiac rhythm with symptoms is an essential part of patient management. Pacemaker implantation is the primary therapeutic intervention in patients with symptomatic SA node dysfunction. Pharmacologic considerations are important in the evaluation and management of patients with SA nodal disease. A number of drugs modulate SA node function and are extrinsic causes of dysfunction (Table 232-1).

Beta blockers and calcium channel blockers increase SNRT in patients with SA node dysfunction, and antiarrhythmic drugs with class I and III action may promote SA node exit block. In general, such agents should be discontinued before decisions regarding the need for permanent pacing in patients with SA node disease are made. Chronic pharmacologic therapy for sinus bradyarrhythmias is limited. Some pharmacologic agents may improve SA node function; digitalis, for example, has been shown to shorten SNRT in patients with SA node dysfunction. Isoproterenol or atropine administered IV may increase the sinus rate acutely. Theophylline has been used both acutely and chronically to increase heart rate but has liabilities when used in patients with tachycardia-bradycardia syndrome, increasing the frequency of supraventricular tachyarrhythmias, and in patients with structural heart disease, increasing the risk of potentially serious ventricular arrhythmias. At the current time, there is only a single randomized study of therapy for SA node dysfunction. In patients with resting heart rates <50 and >30 beats/min on a Holter monitor, patients who received dual-chamber pacemakers experienced significantly fewer syncopal episodes and had symptomatic improvement compared with patients randomized to theophylline or no treatment.

In certain circumstances, sinus bradycardia requires no specific treatment or only temporary rate support. Sinus bradycardia is common in patients with acute inferior or posterior MI and can be exacerbated by vagal activation induced by pain or the use of drugs such as morphine. Ischemia of the SA nodal artery probably occurs in acute coronary syndromes more typically with involvement with the right coronary artery, and even with infarction, the effect on SA node function most often is transient.

Sinus bradycardia is a prominent feature of carotid sinus hypersensitivity and neurally mediated hypotension associated with vasovagal syncope that responds to pacemaker therapy. Carotid hypersensitivity with recurrent syncope or presyncope associated with a predominant cardioinhibitory component responds to pacemaker implantation. Several randomized trials have investigated the efficacy of permanent pacing in patients with drug-refractory vasovagal syncope, with mixed results. Although initial trials suggested that patients undergoing pacemaker implantation have fewer recurrences and a longer time to recurrence of symptoms, at least one follow-up study did not confirm these results.

The details of pacing modes and indications for pacing in SA node dysfunction are discussed below.

Atrioventricular Conduction Disease

STRUCTURE AND PHYSIOLOGY OF THE AV NODE

The AV conduction axis is structurally complex, involving the atria and ventricles as well as the AV node. Unlike the SA node, the AV node is a subendocardial structure originating in the transitional zone, which is composed of aggregates of cells in the posterior-inferior right atrium. Superior, medial, and posterior transitional atrionodal bundles converge on the compact AV node. The compact AV node (~1 x 3 x 5 mm) is situated at the apex of the triangle of Koch, which is defined by the coronary sinus ostium posteriorly, the septal tricuspid valve annulus anteriorly, and the tendon of Todaro superiorly. The compact AV node continues as the penetrating AV bundle where it immediately traverses the central fibrous body and is in close proximity to the aortic, mitral, and tricuspid valve annuli; thus, it is subject to injury in the setting of valvular heart disease or its surgical treatment. The penetrating AV bundle continues through the annulus fibrosis and emerges along the ventricular septum adjacent to the membranous septum as the bundle of His. The right bundle branch (RBB) emerges from the distal AV bundle in a band that traverses the right ventricle (moderator band). In contrast, the left bundle branch (LBB) is a broad subendocardial sheet of tissue on the septal left ventricle. The Purkinje fiber network emerges from the RBB and LBB and extensively ramifies on the endocardial surfaces of the right and left ventricles, respectively.

The blood supply to the penetrating AV bundle is from the AV nodal artery and first septal perforator of the left anterior descending coronary artery. The bundle branches also have a dual blood supply from the septal perforators of the left anterior descending coronary artery and branches of the posterior descending coronary artery. The AV node is highly innervated with postganglionic sympathetic and parasympathetic nerves. The bundle of His and distal conducting system are minimally influenced by autonomic tone.

The cells that constitute the AV node complex are heterogeneous with a range of action potential profiles. In the transitional zones, the cells have an electrical phenotype between those of atrial myocytes and cells of the compact node (Fig. 232-1). Atrionodal transitional connections may exhibit *decremental conduction*, defined as slowing of conduction with increasingly rapid rates of stimulation. Fast and slow AV nodal pathways have been described, but it is controversial whether these two types of pathway are anatomically distinct or represent functional heterogeneities in different regions of the AV nodal complex. Myocytes that constitute the compact node are depolarized (resting membrane potential ~–60 mV) and exhibit action potentials with low amplitudes, slow upstrokes of phase 0 (<10 V/s), and phase 4 diastolic depolarization; high-input resistance; and relative insensitivity to external [K+]. The action potential phenotype is explained by the complement of ionic currents expressed. AV nodal cells lack I_{K1} and I_{Na}; $I_{Ca\text{-}L}$ is responsible for phase 0; and phase 4 depolarization reflects the composite activity of the depolarizing currents I_f, $I_{Ca\text{-}L}$, $I_{Ca\text{-}T}$, and INCX and the repolarizing currents

I_{Kr} and $IKAC_h$. Electrical coupling between cells in the AV node is tenuous due to the relatively sparse expression of gap junction channels (predominantly connexin-40) and increased extracellular volume.

The His bundle and the bundle branches are insulated from ventricular myocardium. The most rapid conduction in the heart is observed in these tissues. The action potentials exhibit very rapid upstrokes (phase 0), prolonged plateaus (phase 2), and modest automaticity (phase 4 depolarization). Gap junctions, composed largely of connexin-40, are abundant, but bundles are poorly connected transversely to ventricular myocardium.

ETIOLOGY OF AV CONDUCTION DISEASE

Conduction block from the atrium to the ventricle can occur for a variety of reasons in a number of clinical situations, and AV conduction block may be classified in a number of ways. The etiologies may be functional or structural, in part analogous to extrinsic and intrinsic causes of SA nodal dysfunction. The block may be classified by its severity from first to third degree or complete AV block or by the location of block within the AV conduction system. **Table 232-2** summarizes the etiologies of AV conduction block. Those which are functional (autonomic, metabolic/endocrine, and drug-related) tend to be reversible. Most other etiologies produce structural changes, typically fibrosis, in segments of the AV conduction axis that are generally permanent. Heightened vagal tone during sleep or in well-conditioned individuals can be associated with all grades of AV block. Carotid sinus hypersensitivity, vasovagal syncope, and cough and micturition syncope may be associated with SA node slowing and AV conduction block. Transient metabolic and endocrinologic disturbances as well as a number of pharmacologic agents also may produce reversible AV conduction block.

Table 232–2. Etiologies of Atrioventricular Block

Autonomic	
Carotid sinus hypersensitivity	Vasovagal
Metabolic/Endocrine	
Hyperkalemia	Hypermagnesemia
Hypothyroidism	Adrenal insufficiency
Drug-Related	
Beta blockers	Adenosine
Calcium channel blockers	Antiarrhythmics (class I and III)
Digitalis	Lithium
Infectious	
Endocarditis	Tuberculosis
Lyme disease	Diphtheria
Chagas' disease	Toxoplasmosis
Syphilis	
Heritable/Congenital	
Congenital heart disease	Facioscapulohumeral MD, OMIM #158900 (4q35)
Maternal SLE	
Kearns-Sayre syndrome, OMIM #530000	Emery-Dreifuss MD, OMIM #310300 (Xq28)
Myotonic dystrophy Type 1, OMIM #160900 (19q13.2-13.3) Type 2, OMIM #602668 (3q13.3-q24)	Progressive familial heart block, OMIM #113900 (19q13.2-q13.3, 3p21)
Inflammatory	
SLE	MCTD
Rheumatoid arthritis	Scleroderma
Infiltrative	
Amyloidosis	Hemochromatosis
Sarcoidosis	
Neoplastic/Traumatic	
Lymphoma	Radiation
Mesothelioma	Catheter ablation

Melanoma	
Degenerative	
Lev disease **Coronary Artery Disease**	Lenègre disease
Acute MI	

Abbreviations: MCTD, mixed connective tissue disease; MI, myocardial infarction; OMIM, Online Mendelian Inheritance in Man (database); SLE, systemic lupus erythematosus.

Several infectious diseases have a predilection for the conducting system. Lyme disease may involve the heart in up to 50% of cases; 10% of patients with Lyme carditis develop AV conduction block, which is generally reversible but may require temporary pacing support. Chagas' disease, which is common in Latin America, and syphilis may produce more persistent AV conduction disturbances. Some autoimmune and infiltrative diseases may produce AV conduction block, including SLE, RA, MCTD, scleroderma, amyloidosis (primary and secondary), sarcoidosis, and hemochromatosis; rare malignancies also may impair AV conduction.

Idiopathic progressive fibrosis of the conduction system is one of the more common and degenerative causes of AV conduction block. Aging is associated with degenerative changes in the summit of the ventricular septum, central fibrous body, and aortic and mitral annuli and has been described as "sclerosis of the left cardiac skeleton." The process typically begins in the fourth decade of life and may be accelerated by atherosclerosis, hypertension, and diabetes mellitus. Accelerated forms of progressive familial heart block have been identified in families with mutations in the cardiac sodium channel gene (*SCN5A*) and other loci that have been mapped to chromosomes 1 and 19.

AV conduction block has been associated with heritable neuromuscular diseases, including the nucleotide repeat disease myotonic dystrophy, the mitochondrial myopathy Kearns-Sayre syndrome (Chap. 387), and several of the monogenic muscular dystrophies. Congenital AV block may be observed in complex congenital cardiac anomalies (Chap. 236), such as transposition of the great arteries, ostium primum atrial septal defects (ASDs), ventricular septal defects (VSDs), endocardial cushion defects, and some single-ventricle defects. Congenital AV block in the setting of a structurally normal heart has been seen in children born to mothers with SLE. Iatrogenic AV block may occur during mitral or aortic valve surgery, rarely in the setting of thoracic radiation, and as a consequence of catheter ablation. AV block is a decidedly rare complication of the surgical repair of VSDs or ASDs but may complicate repairs of transposition of the great arteries.

CAD may produce transient or persistent AV block. In the setting of coronary spasm, ischemia, particularly in the right coronary artery distribution, may produce transient AV block. In acute MI, AV block transiently develops in 10–25% of patients; most commonly this is first- or second-degree AV block, but complete heart block (CHB) may also occur. Second-degree and higher-grade AV block tends to occur more often in inferior than in anterior acute MI; however, the level of block in inferior MI tends to be in the AV node with more stable, narrow escape rhythms. In contrast, acute anterior MI is associated with block in the distal AV nodal complex, His bundle, or bundle branches and results in wide complex, unstable escape rhythms and a worse prognosis with high mortality rates.

ELECTROCARDIOGRAPHY AND ELECTROPHYSIOLOGY OF AV CONDUCTION BLOCK

Atrioventricular conduction block typically is diagnosed electrocardiographically, which characterizes the severity of the conduction disturbance and allows one to draw inferences about the location of the block. AV conduction block manifests as slow conduction in its mildest forms and failure to conduct, either intermittent or persistently, in more severe varieties. First-degree AV block (PR interval >200 ms) is a slowing of conduction through the AV junction **(Fig. 232-5)**. The site of delay is typically in the AV node but may be in the atria, bundle of His, or His-Purkinje system. A wide QRS is suggestive of delay in the distal conduction system, whereas a narrow QRS suggests delay in the AV node proper or, less commonly, in the bundle of His. In second-degree AV block there is an intermittent failure of electrical impulse conduction from atrium to ventricle. Second-degree AV block is subclassified as Mobitz type I (Wenckebach) or Mobitz type II. The periodic failure of conduction in Mobitz type I block is characterized by a progressively lengthening PR interval, shortening of the RR interval, and a pause that is less than two times the immediately preceding RR interval on the ECG. The ECG complex after the pause exhibits a shorter PR interval than that immediately preceding the pause **(Fig. 232-6)**. This ECG pattern most often arises because of decremental conduction of electrical impulses in the AV node.

Figure 232-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

First-degree AV block with slowing of conduction in the AV node as indicated by the prolonged atrial-to-His bundle electrogram (AH) interval, in this case 157 ms. The His bundle-to-earliest ventricular activation on the surface ECG (HV) interval is normal. The normal HV interval suggests normal conduction below the AV node to the ventricle. I and V1 are surface ECG leads, HIS is the recording of the endocavitary electrogram at the His bundle position. A, H, and V are labels for the atrial, His bundle, and right ventricular electrograms, respectively.

Figure 232-6

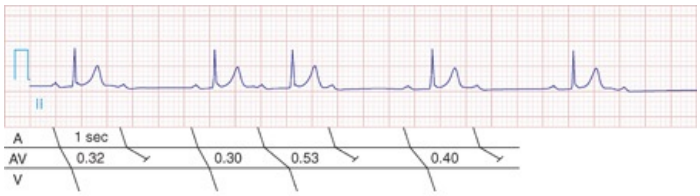

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mobitz type I second-degree AV block. The PR interval prolongs before the pause, as shown in the ladder diagram. The ECG pattern results from slowing of conduction in the AV node.

It is important to distinguish type I from type II second-degree AV nodal block because the latter has more serious prognostic implications. Type II second-degree AV block is characterized by intermittent failure of conduction of the P wave without changes in the preceding PR or RR intervals. When AV block is 2:1, it may be difficult to distinguish type I from type II block. Type II second-degree AV block typically occurs in the distal or infra-His conduction system, is often associated with intraventricular conduction delays (e.g., bundle branch block), and is more likely to proceed to higher grades of AV block than is type I second-degree AV block. Second-degree AV block (particularly type II) may be associated with a series of nonconducted P waves, referred to as *paroxysmal AV block* **(Fig. 232-7)**, and implies significant conduction system disease and is an indication for permanent pacing. Complete failure of conduction from atrium to ventricle is referred to as complete or third-degree AV block. AV block that is intermediate between second degree and third degree is referred to as high-grade AV block and, as with CHB, implies advanced AV conduction system disease. In both cases, the block is most often distal to the AV node, and the duration of the QRS complex can be helpful in determining the level of the block. In the absence of a preexisting bundle branch block, a wide QRS escape rhythm **(Fig. 232-8*B*)** implies a block in the distal His or bundle branches; in contrast, a narrow QRS rhythm implies a block in the AV node or proximal His and an escape rhythm originating in the AV junction (Fig. 232-8*A*). Narrow QRS escape rhythms are typically faster and more stable than wide QRS escape rhythms and originate more proximally in the AV conduction system.

Figure 232-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Paroxysmal AV block. Multiple nonconducted P waves after a period of sinus bradycardia with a normal PR interval. This implies significant conduction system disease, requiring permanent pacemaker implantation.

Figure 232-8

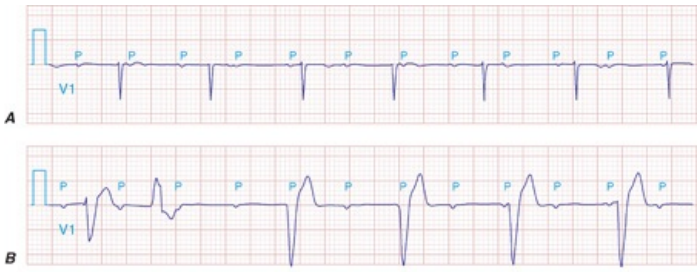

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

High-grade AV block. ***A***. Multiple nonconducted P waves with a regular narrow complex QRS escape rhythm probably emanating from the AV junction. ***B.*** A wide complex QRS escape and a single premature ventricular contraction. In both cases there is no consistent temporal relationship between the P waves and QRS complexes.

DIAGNOSTIC TESTING

Diagnostic testing in the evaluation of AV block is aimed at determining the level of conduction block, particularly in asymptomatic patients, since the prognosis and therapy depend on whether the block is in or below the AV node. Vagal maneuvers, carotid sinus massage, exercise, and administration of drugs such as atropine and isoproterenol may be diagnostically informative. Owing to the differences in innervation of the AV node and infranodal conduction system, vagal stimulation and carotid sinus massage slow conduction in the AV node but have less of an effect on infranodal tissue and may even improve conduction due to a reduced rate of activation of distal tissues. Conversely, atropine, isoproterenol, and exercise improve conduction through the AV node and impair infranodal conduction. In patients with congenital CHB and a narrow QRS complex, exercise typically increases heart rate; by contrast, those with acquired CHB, particularly with wide QRS, do not respond to exercise with an increase in heart rate.

Additional diagnostic evaluation, including electrophysiologic testing, may be indicated in patients with syncope and suspected high-grade AV block. This is particularly relevant if noninvasive testing does not reveal the cause of syncope or if the patient has structural heart disease with ventricular tachyarrhythmias as a cause of symptoms. Electrophysiologic testing provides more precise information regarding the location of AV conduction block and permits studies of AV conduction under conditions of pharmacologic stress and exercise. Recording of the His bundle electrogram by a catheter positioned at the superior margin of the tricuspid valve annulus provides information about conduction at all levels of the AV conduction axis. A properly recorded His bundle electrogram reveals local atrial activity, the His electrogram, and local ventricular activation; when it is monitored simultaneously with recorded body surface electrocardiographic traces, intraatrial, AV nodal, and infranodal conduction times can be assessed (Fig. 232-5). The time from the most rapid deflection of the atrial electrogram in the His bundle recording to the His electrogram (*AH interval*) represents conduction through the AV node and is normally <130 ms. The time from the His electrogram to the earliest onset of the QRS on the surface ECG (*HV interval*) represents the conduction time through the His-Purkinje system and is normally ≲55 ms.

Rate stress produced by pacing can unveil abnormal AV conduction. Mobitz I second-degree AV block at short atrial paced cycle lengths is a normal response. However, when it occurs at atrial cycle lengths >500 ms (<120 beats/min) in the absence of high vagal tone, it is abnormal. Typically, type I second-degree AV block is associated with prolongation of the AH interval, representing conduction slowing and block in the AV node. AH prolongation occasionally is due to the effect of drugs (beta blockers, calcium channel blockers, digitalis) or increased vagal tone. Atropine can be used to reverse high vagal tone; however, if AH prolongation and AV block at long pacing cycle lengths persist, intrinsic AV node disease is likely. Type II second-degree block is typically infranodal, often in the His-Purkinje system. Block below the node with prolongation of the HV interval or a His bundle electrogram with no ventricular activation **(Fig. 232-9)** is abnormal unless it is elicited at fast pacing rates or short coupling intervals with extra stimulation. It is often difficult to determine the type of second-degree AV block when 2:1 conduction is present; however, the finding of a His bundle electrogram after every atrial electrogram indicates that block is occurring in the distal conduction system.

Figure 232-9

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

High-grade AV block below the His. The AH interval is normal and is not changing before the block. Atrial and His bundle electrograms are recorded consistent with block below the distal AV junction. I, II, III, and V1 are surface ECG leads. HISp, HISd, and RVA are the proximal HIS, distal HIS, and right ventricular apical electrical recordings. A, H, and V represent the atrial, His, and ventricular electrograms on the His bundle recording. *(Tracing courtesy of Dr. Joseph Marine; with permission.)*

Intracardiac recording at electrophysiologic study that reveals prolongation of conduction through the His-Purkinje system (i.e., long HV interval) is associated with an increased risk of progression to higher grades of block and is generally an indication for pacing. In the setting of bundle branch block, the HV interval may reveal the condition of the unblocked bundle and the prognosis for developing more advanced AV conduction block. Prolongation of the HV interval in patients with asymptomatic bundle branch block is associated with an increased risk of developing higher-grade AV block. The risk increases with greater prolongation of the HV interval such that in patients with an HV interval >100 ms, the annual incidence of complete AV block approaches 10%, indicating a need for pacing. In patients with acquired CHB, even if intermittent, there is little role for electrophysiologic testing, and pacemaker implantation is almost always indicated.

TREATMENT: MANAGEMENT OF AV CONDUCTION BLOCK

Temporary or permanent artificial pacing is the most reliable treatment for patients with symptomatic AV conduction system disease. However, exclusion of reversible causes of AV block and the need for temporary heart rate support based on the hemodynamic condition of the patient are essential considerations in each patient. Correction of electrolyte derangements and ischemia, inhibition of excessive vagal tone, and withholding of drugs with AV nodal blocking properties may increase the heart rate. Adjunctive pharmacologic treatment with atropine or isoproterenol may be useful if the block is in the AV node. Since most pharmacologic treatment may take some time to initiate and become effective, temporary pacing may be necessary. The most expeditious technique is the use of transcutaneous pacing, where pacing patches are placed anteriorly over the cardiac apex (cathode)

and posteriorly between the spine and the scapula or above the right nipple (anode). Acutely, transcutaneous pacing is highly effective, but its duration is limited by patient discomfort and longer-term failure to capture the ventricle owing to changes in lead impedance. If a patient requires more than a few minutes of pacemaker support, transvenous temporary pacing should be instituted. Temporary pacing leads can be placed from the jugular or subclavian venous system and advanced to the right ventricle, permitting stable temporary pacing for many days, if necessary. In most circumstances, in the absence of prompt resolution, conduction block distal to the AV node requires permanent pacemaking.

Permanent Pacemakers

NOMENCLATURE AND COMPLICATIONS

The main therapeutic intervention in SA node dysfunction and AV conduction block is permanent pacing. Since the first implementation of permanent pacing in the 1950s, many advances in technology have resulted in miniaturization, increased longevity of pulse generators, improvement in leads, and increased functionality. To better understand pacemaker therapy for bradycardias, it is important to be familiar with the fundamentals of pacemaking. Pacemaker modes and function are named using a five-letter code. The first letter indicates the chamber(s) that is paced (O, none; A, atrium; V, ventricle; D, dual; S, single), the second is the chamber(s) in which sensing occurs (O, none; A, atrium; V, ventricle; D, dual; S, single), the third is the response to a sensed event (O, none; I, inhibition; T, triggered; D, inhibition + triggered), the fourth refers to the programmability or rate response (R, rate responsive), and the fifth refers to the existence of antitachycardia functions if present (O, none; P, antitachycardia pacing; S, shock; D, pace + shock). Almost all modern pacemakers are multiprogrammable and have the capability for rate responsiveness using one of several rate sensors: activity or motion, minute ventilation, or QT interval. The most commonly programmed modes of implanted single- and dual-chamber pacemakers are VVIR and DDDR, respectively, although multiple modes can be programmed in modern pacemakers.

Although pacemakers are highly reliable, they are subject to a number of complications related to implantation and electronic function. In adults, permanent pacemakers are most commonly implanted with access to the heart by way of the subclavian–superior vena cava venous system. Rare, but possible, acute complications of transvenous pacemaker implantation include infection, hematoma, pneumothorax, cardiac perforation, diaphragmatic/phrenic nerve stimulation, and lead dislodgment. Limitations of chronic pacemaker therapy include infection, erosion, lead failure, and abnormalities resulting from inappropriate programming or interaction with the patient's native electrical cardiac function. Rotation of the pacemaker pulse generator in its subcutaneous pocket, either intentionally or inadvertently, often referred to as "twiddler's syndrome," can wrap the leads around the generator and produce dislodgment with failure to sense or pace the heart. The small size and light weight of contemporary pacemakers make this a rare complication.

Complications stemming from chronic cardiac pacing also result from disturbances in atrioventricular synchrony and/or left ventricular mechanical synchrony. Pacing modes that interrupt or fail to restore atrioventricular synchrony may lead to a constellation of signs and symptoms, collectively referred to as pacemaker syndrome, that include neck pulsation, fatigue, palpitations, cough, confusion, exertional dyspnea, dizziness, syncope, elevation in jugular venous pressure, canon A waves, and stigmata of congestive heart failure, including edema, rales, and a third heart sound. Right ventricular apical pacing can induce dyssynchronous activation of the left ventricle, leading to compromised left ventricular (LV) systolic function, mitral valve regurgitation, and the previously mentioned stigmata of congestive heart failure. Maintenance of AV synchrony can minimize the sequelae of pacemaker syndrome. Selection of pacing modes that minimize unnecessary ventricular pacing or implantation of a device capable of right and left ventricular pacing (biventricular pacing) can help minimize the deleterious consequences of pacing-induced mechanical dyssynchrony at the ventricular level.

PACEMAKER THERAPY IN SA NODE DYSFUNCTION

Pacing in SA nodal disease is indicated to alleviate symptoms of bradycardia. Consensus guidelines published by the American Heart Association (AHA)/American College of Cardiology/Heart Rhythm Society (ACC/HRS) outline the indications for the use of pacemakers and categorize them by class based on levels of evidence. Class I conditions are those for which there is evidence or consensus of opinion that therapy is useful and effective. In class II conditions there is conflicting evidence or a divergence of opinion about the efficacy of a procedure or treatment; in class IIa conditions the weight of evidence or opinion favors treatment, and in class IIb conditions efficacy is less well established by the evidence or opinion of experts. In class III conditions, the evidence or weight of opinion indicates that the therapy is not efficacious or useful and may be harmful.

Class I indications for pacing in SA node dysfunction include documented symptomatic bradycardia, sinus node dysfunction–associated long-term drug therapy for which there is no alternative, and symptomatic chronotropic incompetence. Class IIa indications include those outlined previously in which sinus node dysfunction is suspected but not documented and for syncope of unexplained origin in the presence of major abnormalities of SA node dysfunction. Mildly symptomatic individuals with heart rates consistently <40 beats/min constitute a class IIb indication for pacing. Pacing is not

indicated in patients with SA node dysfunction who do not have symptoms and in those in whom bradycardia is associated with the use of nonessential drugs **(Table 232-3)**.

Table 232–3. Summary of Guidelines for Pacemaker Implantation in SA Node Dysfunction

Class I
1. SA node dysfunction with symptomatic bradycardia or sinus pauses
2. Symptomatic SA node dysfunction as a result of essential long-term drug therapy with no acceptable alternatives
3. Symptomatic chronotropic incompetence
4. Atrial fibrillation with bradycardia and pauses >5s
Class IIa
1. SA node dysfunction with heart rates <40 beats/min without a clear and consistent relationship between bradycardia and symptoms
2. SA node dysfunction with heart rates <40 beats/min on an essential long-term drug therapy with no acceptable alternatives, without a clear and consistent relationship between bradycardia and symptoms
3. Syncope of unknown origin when major abnormalities of SA node dysfunction are discovered or provoked by electrophysiologic testing
Class IIb
1. Mildly symptomatic patients with waking chronic heart rates <40 beats/min
Class III
1. SA node dysfunction in asymptomatic patients, even those with heart rates <40 beats/min
2. SA node dysfunction in which symptoms suggestive of bradycardia are not associated with a slow heart rate
3. SA node dysfunction with symptomatic bradycardia due to nonessential drug therapy

Source: Modified from Epstein et al, J. Am. Coll. Cardiol. 51:e1, 2008 and Gregoratos et al, J. Am. Coll. Cardiol. 40:703, 2002.

There is some controversy about the mode of pacing that should be employed in SA node disease. A number of randomized, single-blind trials of pacing mode have been performed. There are no trials that demonstrate an improvement in mortality rate with AV synchronous pacing compared with single-chamber pacing in SA node disease. In some of these studies, the incidence of atrial fibrillation and thromboembolic events was reduced with AV synchronous pacing. In trials of patients with dual-chamber pacemakers designed to compare single-chamber with dual-chamber pacing by crossover design, the need for AV synchronous pacing due to pacemaker syndrome was common. Pacing modes that preserve AV synchrony appear to be associated with a reduction in the incidence of atrial fibrillation and improved quality of life. Because of the low but finite incidence of AV conduction disease, patients with SA node dysfunction usually undergo dual-chamber pacemaker implantation.

PACEMAKER THERAPY IN CAROTID SINUS HYPERSENSITIVITY AND VASOVAGAL SYNCOPE

Carotid sinus hypersensitivity, if accompanied by a significant cardioinhibitory component, responds well to pacing. In this circumstance, pacing is required only intermittently and single-chamber ventricular pacing is often sufficient. The mechanism of vasovagal syncope is incompletely understood but appears to involve activation of cardiac mechanoreceptors with consequent activation of neural centers that mediate vagal activation and withdrawal of sympathetic nervous system tone. Several randomized clinical trials have been performed in patients with drug-refractory vasovagal syncope, with some studies suggesting reduction in the frequency and the time to recurrent syncope in patients who were paced compared with those who were not. A recent follow-up study to one of those initial trials, however, found less convincing results, casting some doubt on the utility of pacing for vagally mediated syncope.

PACEMAKERS IN AV CONDUCTION DISEASE

There are no randomized trials that evaluate the efficacy of pacing in patients with AV block, as there are no reliable therapeutic alternatives for AV block and untreated high-grade AV block is potentially lethal. The consensus guidelines for pacing in acquired AV conduction block in adults provide a general outline for situations in which pacing is indicated **(Table 232-4)**. Pacemaker implantation should be performed in any patient with symptomatic bradycardia and irreversible second- or third-degree AV block, regardless of the cause or level of block in the conducting system. Symptoms may include those directly related to bradycardia and low cardiac output or to worsening heart failure, angina, or intolerance to an essential medication. Pacing in patients with asymptomatic AV block should be individualized; situations in which pacing should be considered are patients with acquired CHB, particularly in the setting of cardiac enlargement; left ventricular dysfunction; and waking heart rates ≤40 beats/min. Patients who have asymptomatic second-degree AV block of either type should be

considered for pacing if the block is demonstrated to be intra- or infra-His or is associated with a wide QRS complex. Pacing may be indicated in asymptomatic patients in special circumstances, in patients with profound first-degree AV block and left ventricular dysfunction in whom a shorter AV interval produces hemodynamic improvement, and in the setting of milder forms of AV conduction delay (first-degree AV block, intraventricular conduction delay) in patients with neuromuscular diseases that have a predilection for the conduction system, such as myotonic dystrophy and other muscular dystrophies, and Kearns-Sayre syndrome.

Table 232–4. Guideline Summary for Pacemaker Implantation in Acquired AV Block

Class I
1. Third-degree or high-grade AV block at any anatomic level associated with:
a. Symptomatic bradycardia
b. Essential drug therapy that produces symptomatic bradycardia
c. Periods of asystole >3 s or any escape rate <40 beats/min while awake
d. Postoperative AV block not expected to resolve
e. Catheter ablation of the AV junction
f. Neuromuscular diseases such as myotonic dystrophy, Kearns-Sayre syndrome, Erb dystrophy, and peroneal muscular atrophy, regardless of the presence of symptoms
2. Second-degree AV block with symptomatic bradycardia
3. Type II second-degree AV block with a wide QRS complex with or without symptoms
4. Exercise-induced second- or third-degree AV block in the absence of ischemia
5. Atrial fibrillation with bradycardia and pauses >5s
Class IIa
1. Asymptomatic third-degree AV block regardless of level
2. Asymptomatic type II second-degree AV block with a narrow QRS complex
3. Asymptomatic type II second-degree AV block with block within or below the His at electrophysiologic study
4. First- or second-degree AV block with symptoms similar to pacemaker syndrome
Class IIb
1. Asymptomatic third-degree AV block regardless of level
2. Asymptomatic type II second-degree AV block with a narrow
QRS complex
3. Asymptomatic type II second-degree AV block with block within or below the His at electrophysiologic study
4. First- or second-degree AV block with symptoms similar to pacemaker syndrome
Class IIb
1. Marked first-degree AV block (PR interval >300 ms) in patients with LV dysfunction in whom shortening the AV delay would improve hemodynamics
2. Neuromuscular diseases such as myotonic dystrophy, Kearns-Sayre syndrome, Erb dystrophy, and peroneal muscular atrophy with any degree of AV block regardless of the presence of symptoms
Class III
1. Asymptomatic first-degree AV block
2. Asymptomatic type I second-degree AV block at the AV node level
3. AV block that is expected to resolve or is unlikely to recur (Lyme disease, drug toxicity)

Source: Modified from Epstein et al, J. Am. Coll. Cardiol. 51:e1, 2008 and Gregoratos et al, J. Am. Coll. Cardiol. 40:703, 2002.

PACEMAKER THERAPY IN MYOCARDIAL INFARCTION

Atrioventricular block in acute MI is often transient, particularly in inferior infarction. The circumstances in which pacing is indicated in acute MI are persistent second- or third-degree AV block, particularly if symptomatic, and transient second- or third-degree AV block associated with bundle branch block **(Table 232-5)**. Pacing is generally not indicated in the setting of transient AV block in the absence of intraventricular conduction delays or in the presence of fascicular block or first-degree AV block that develops in the setting of preexisting bundle branch block. Fascicular blocks that develop in acute MI in the absence of other forms of AV block also do not require pacing (Table 232-5 and **Table 232-6**).

Table 232–5. Guideline Summary for Pacemaker Implantation in AV Conduction Block in Acute Myocardial Infarction (AMI)

Class I
1. Persistent second-degree AV block in the His-Purkinje system with bilateral bundle branch block or third-degree block within or below the His after AMI
2. Transient advanced (second- or third-degree) infranodal AV block and associated bundle branch block. If the site of block is uncertain, an electrophysiologic study may be necessary
3. Persistent and symptomatic second- or third-degree AV block
Class IIb
1. Persistent second- or third-degree AV block at the AV node level
Class III
1. Transient AV block in the absence of intraventricular conduction defects
2. Transient AV block in the presence of isolated left anterior fascicular block
3. Acquired left anterior fascicular block in the absence of AV block
4. Persistent first-degree AV block in the presence of bundle branch block that is old or age-indeterminate

Source: Modified from Epstein et al, J. Am. Coll. Cardiol. 51:e1, 2008 and Gregoratos et al, J. Am. Coll. Cardiol. 40:703, 2002.

Table 232–6. Indications for Pacemaker Implantation in Chronic Bifascicular and Trifascicular Block

Class I
1. Intermittent third-degree AV block
2. Type II second-degree AV block
3. Alternating bundle branch block
Class IIa
1. Syncope not demonstrated to be due to AV block when other likely causes (e.g., ventricular tachycardia) have been excluded
2. Incidental finding at electrophysiologic study of a markedly prolonged HV interval (>100 ms) in asymptomatic patients
3. Incidental finding at electrophysiologic study of pacing-induced infra-His block that is not physiologic
Class IIb
1. Neuromuscular diseases such as myotonic dystrophy, Kearns-Sayre syndrome, Erb dystrophy, and peroneal muscular atrophy with any degree of fascicular block regardless of the presence of symptoms, because there may be unpredictable progression of AV conduction disease
Class III
1. Fascicular block without AV block or symptoms
2. Fascicular block with first-degree AV block without symptoms

Source: Modified from Epstein et al, J. Am. Coll. Cardiol. 51:e1, 2008 and Gregoratos et al, J. Am. Coll. Cardiol. 40:703, 2002.

PACEMAKER THERAPY IN BIFASCICULAR AND TRIFASCICULAR BLOCK

Distal forms of AV conduction block may require pacemaker implantation in certain clinical settings. Patients with bifascicular or trifascicular block and symptoms, particularly syncope that is not attributable to other causes, should undergo pacemaker implantation. Pacemaking is indicated in asymptomatic patients with bifascicular or trifascicular block who experience intermittent third-degree, type II second-degree AV block or alternating bundle branch block. In patients with fascicular block who are undergoing electrophysiologic study, a markedly prolonged HV interval or block below the His at long cycle lengths also may constitute an indication for permanent pacing. Patients with fascicular block and the neuromuscular diseases previously described should also undergo pacemaker implantation (Table 232-6).

SELECTION OF PACING MODE

In general, a pacing mode that maintains AV synchrony reduces complications of pacing such as pacemaker syndrome and pacemaker-mediated tachycardia. This is particularly true in younger patients; the importance of dual-chamber pacing in the elderly, however, is not well established. Several studies have failed to demonstrate a difference in mortality rate in older patients with AV block treated with a single- (VVI) compared with a dual- (DDD) chamber pacing mode. In some of the studies that randomized pacing mode, the risk of chronic atrial fibrillation and stroke risk decreased with

physiologic pacing. In patients with sinus rhythm and AV block, the very modest increase in risk with dual-chamber pacemaker implantation appears to be justified to avoid the possible complications of single-chamber pacing.

FURTHER READINGS

Bharati S et al: Sinus node dysfunction, in *Electrophysiological Disorders of the Heart* , S Saksena, AJ Camm (eds). Philadelphia, Elsevier Churchill Livingstone, 2005

Epstein AE et al: ACC/AHA/HRS 2008 Guidelines for Device-Based Therapy of Cardiac Rhythm Abnormalities: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (writing Committee to Revise the ACC/AHA/NASPE 2002 Guideline Update for Implantation of Cardiac Pacemakers and Antiarrhythmia Devices) developed in collaboration with the American Association for Thoracic Surgery and Society of Thoracic Surgeons. J Am Coll Cardiol 51:el, 2008

Goldschlager N et al: Atrioventricular block, in *Electrophysiological Disorders of the Heart* , S Saksena, AJ Camm (eds). Philadelphia, Elsevier Churchill Livingstone, 2005

———: ACC/AHA/NASPE 2008 Guidelines for device-based therapy of cardiac rhythm abnormalities. J Am Coll Cardiol 51:21, 2008

Gregoratos G et al: ACC/AHA/NASPE 2002 Guideline Update for Implantation of Cardiac Pacemakers and Antiarrhythmia Devices–summary article: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/NASPE Committee to Update the 1998 Pacemaker Guidelines). J Am Coll Cardiol 40:1703, 2002[PMID: 12427427]

Josephson ME: *Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2008

Vijayaraman P, Ellenbogen KA: Bradyarrhythmias and pacemakers, in *Hurst's The Heart,* 12th ed, V Fuster et al (eds). New York, McGraw-Hill, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE TACHYARRHYTHMIAS: INTRODUCTION

The term *tachyarrhythmias* typically refers to nonsustained and sustained forms of tachycardia originating from myocardial foci or reentrant circuits. The standard definition of tachycardia is a rhythm that produces a ventricular rate >100 beats per minute. This definition has some limitations in that atrial rates can exceed 100 beats per minute despite a slow ventricular rate. Furthermore, ventricular rates may exceed the baseline sinus rate and be <100 beats per minute but still represent an important "tachycardia" response, as is observed with accelerated ventricular rhythms. Premature complexes (depolarizations) are considered under the category of tachyarrhythmias because they may cause arrhythmia-related symptoms and/or serve as triggering events for more sustained forms of tachycardia.

Symptoms Due to Tachyarrhythmias

Tachyarrhythmias classically produce symptoms of palpitations or racing of the pulse. With premature beats, skipping of the pulse or a pause may be experienced, and patients may even sense slowing of the heart rate or dizziness. A more dramatic irregularity of the pulse is experienced with chaotic rapid rhythms or tachyarrhythmias that originate in the atrium and conduct variably to the ventricles. With rapid tachyarrhythmias, hemodynamic compromise can occur, as can dizziness or syncope due to a decrease in cardiac output or breathlessness due to a marked increase in cardiac filling pressures. Occasionally, chest discomfort may be experienced that mimics symptoms of myocardial ischemia. The underlying cardiac condition typically dictates the severity of symptoms at any specific heart rate. Even patients with normal systolic left ventricular (LV) function may experience severe symptoms if diastolic compliance due to hypertrophy or valvular obstruction is present and a tachycardia develops. Hemodynamic collapse with the development of ventricular fibrillation (VF) can lead to sudden cardiac death (SCD) (Chap. 273).

Diagnostic Tests in Evaluating Tachyarrhythmias

In patients who present with nonlife-threatening symptoms such as palpitations or dizziness, electrocardiographic (ECG) confirmation of an arrhythmia with the development of recurrent symptoms is essential. A 24-hour Holter monitor should be considered only for patients with daily symptoms. For intermittent symptoms that are of prolonged duration, a patient-activated event monitor can be used to obtain the ECG information without the need for continuous ECG lead attachment and recordings. A patient-activated monitor with a continuously recorded memory loop ("loop recorder") can be used to document short-lived episodes and the onset of the arrhythmia. This is the preferred monitoring technique for symptomatic patients with less frequent arrhythmia events, but it requires continuous ECG recording. A monitor that automatically triggers to record a fast rhythm can be used to detect asymptomatic arrhythmias. Patients with infrequent, severe symptoms that cannot be identified by intermittent ECG monitoring may receive an implanted loop ECG monitor that provides more extended periods of monitoring and automatic arrhythmia detection **(Fig. 233-1)**.

Figure 233-1

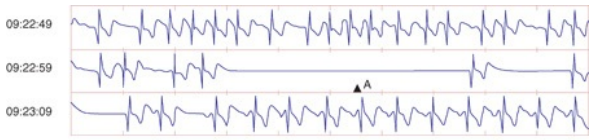

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spontaneous termination of atrial fibrillation at the time of a syncopal episode identified from implantable loop ECG recording.

In patients who present with more severe symptoms, such as syncope, outpatient monitoring may be insufficient. In patients with structural heart disease and syncope in whom there is suspicion of ventricular tachycardia (VT), hospitalization and diagnostic electrophysiologic testing are warranted, with strong consideration of an implantable cardioverter/defibrillator (ICD) device. The 12-lead ECG recorded in sinus rhythm should be assessed carefully in patients without structural heart disease for evidence of ST-segment elevation in leads V_1 and V_2 consistent with the Brugada syndrome, QT interval changes consistent with long or short QT syndromes, or a short PR interval and delta wave consistent with Wolff-Parkinson-White (WPW) syndrome. These ECG patterns identify a possible arrhythmogenic substrate that may cause intermittent life-threatening symptoms and warrant further evaluation and therapy. The individual syndromes are discussed in detail later in this chapter.

Monitoring for asymptomatic tachyarrhythmias is indicated in several specific situations. In patients with a suspected tachycardia-induced cardiomyopathy marked by chamber dilation and depression in systolic function, the demonstration of arrhythmia control is essential. Monitoring for asymptomatic ventricular premature complexes (VPCs) and nonsustained VT can be helpful in stratifying the risk of SCD in patients with depressed LV function after myocardial infarction (MI). Finally, in patients with asymptomatic atrial fibrillation (AF), anticoagulation treatment strategies depend on an accurate assessment of the presence of this arrhythmia. The duration of monitoring for asymptomatic arrhythmias may have to be extended to optimize detection capabilities.

A 12-lead ECG recording during the tachycardia can be an important diagnostic tool in identifying the mechanism and origin of a tachycardia to a degree not afforded by one- or two-lead ECG recordings. A 12-lead ECG of the tachyarrhythmia should be recorded and incorporated as a permanent part of the medical record whenever possible. For patients whose arrhythmias are provoked by exercise, an exercise test may provide an opportunity to obtain 12-lead ECG recordings of the arrhythmia and may obviate the need for more extended periods of monitoring.

Many paroxysmal supraventricular tachyarrhythmias are not associated with a significant risk of structural heart disease,

and an evaluation for the presence of ischemic heart disease and cardiac function is required infrequently unless dictated by the severity or characteristics of the symptoms. However, in patients with focal or macroreentrant atrial tachycardias (ATs), atrial flutter (AFL), or AF, an evaluation of cardiac chamber size and function and of valve function is warranted. In patients with VT, an echocardiographic assessment of LV and right ventricular (RV) size and function should be the norm. Ventricular tachycardia that occurs in the setting of depressed LV function should raise the suspicion of advanced coronary artery disease (CAD). Ventricular tachycardia in the setting of isolated RV dilation should raise concern about the diagnosis of arrhythmogenic RV cardiomyopathy. Polymorphic VT in the absence of QT prolongation should always raise concern for a potentially unstable ischemic process that may need to be corrected to effect VT control.

Mechanisms of Tachyarrhythmias

Tachycardias are due to abnormalities of impulse formation and/or abnormalities of impulse propagation **(Fig. 233-2)**.

Figure 233-2

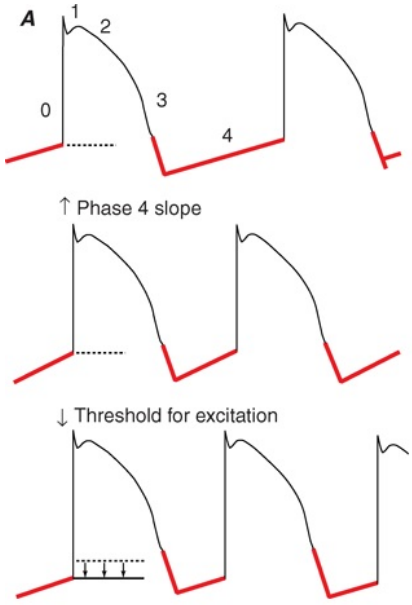

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

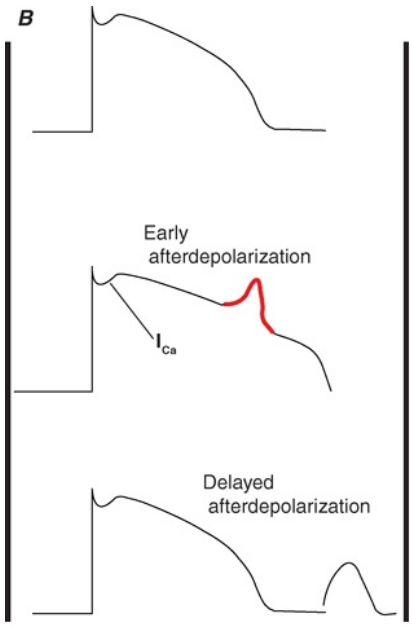

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

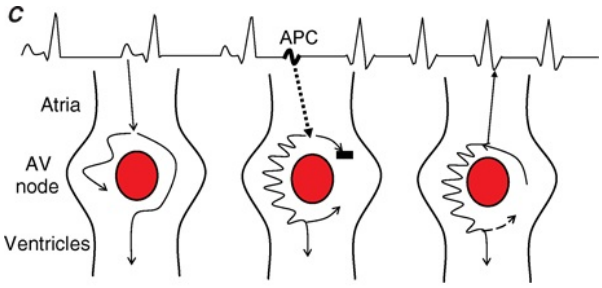

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of the different mechanisms for arrhythmias. *A.* Abnormal automaticity due to an increased slope of phase 4 of the action potential or a decrease in the threshold for phase 0. ***B.*** Triggered activity due to early afterdepolarizations (EADs) during phase 3 of the action potential due to alteration of plateau currents or delayed afterdepolarizations (DADs) during phase 4 of the action potential due to intracellular calcium accumulation. ***C.*** Reentry with basic requirements of two pathways that have heterogeneous electrophysiologic properties which allows conduction to block in one pathway and propagate slowly in the other, allowing for sufficient delay so that the blocked site has time for recovery to allow for reentry or circus movement tachycardia. Shown is a typical schema for reentry in the AV node. AV, atrioventricular; APC, atrial premature complex.

ABNORMALITIES IN IMPULSE FORMATION

An increase in automaticity normally causes an increase in sinus rate and sinus tachycardia (Fig. 233-2*A*). Abnormal automaticity is due to an increase in the slope of phase 4 depolarization or a reduced threshold for action potential depolarization in myocardium other than the sinus node. Abnormal automaticity is thought to be responsible for most atrial premature complexes (APCs) and VPCs and some ATs. Pacing does not provoke automatic rhythms. Less commonly, abnormal impulse formation is due to triggered activity. Triggered activity is related to cellular afterdepolarizations that occur at the end of the action potential, during phase 3, and are referred to as *early afterdepolarizations;* when they occur after the action potential, during phase 4, they are referred to as *late afterdepolarizations*. Afterdepolarizations are attributable to an increase in intracellular calcium accumulation. If sufficient afterdepolarization amplitude is achieved, repeated myocardial depolarization and a tachycardic response can occur. Early afterdepolarizations may be responsible for the VPCs that trigger the polymorphic ventricular arrhythmia known as *torsades des pointes* (TDP) (p. 31). Late afterdepolarizations are thought to be responsible for atrial, junctional, and fascicular tachyarrhythmias caused by digoxin toxicity and also appear to be the basis for catecholamine-sensitive VT originating in the outflow tract. In contrast to automatic tachycardias, those due to triggered activity (Fig. 233-2*B*) frequently can be provoked with pacing maneuvers.

ABNORMALITIES IN IMPULSE PROPAGATION

Reentry is due to inhomogeneities in myocardial conduction and/or recovery properties. The presence of a unidirectional block with slow conduction to allow for retrograde recovery of the blocked myocardium allows the formation of a circuit that, if perpetuated, can sustain a tachycardia (Fig. 233-2*C*). These inhomogeneities are somewhat inherent but are minimized in normal myocardial activation/recovery. The inhomogeneities can be exaggerated by the presence of extra pathways, as occurs with the WPW syndrome; generalized genetically determined myocardial ion channel abnormalities, as occur with long QT syndrome (LQTS); or the interruption of normal myocardial patterns of activation due to the development of fibrosis.

Reentry appears to be the basis for most abnormal sustained supraventricular tachycardias (SVTs) and VTs. In general, reentry can be anatomically driven (fixed) based on the presence of "extra" pathways, natural anatomic barriers of conduction such as the crista terminalis, the vertical crest on the interior wall of the right atrium that separates the nontrabeculated posterior right atrium from the rest of the trabeculated right atrium located lateral to the structure, and/or extensive fibrosis created by underlying myocardial disease. This form of reentry seems to be more stable and results in a tachycardia that has a uniform (often monomorphic), repetitive appearance. Other forms of reentry appear to be more functional and are more dependent on dynamic changes in electrophysiologic properties of the myocardium. These tachycardias tend to be more unstable and may result in tachycardias that have a polymorphic appearance. Two classic examples of reentry that is primarily functional are VF due to acute myocardial ischemia and polymorphic VT in patients with a genetically determined ion channel abnormality such as the Brugada syndrome, LQTS, or catecholaminergic polymorphic VT (pp. 35-36).

SUPRAVENTRICULAR TACHYARRHYTHMIAS

Atrial Premature Complexes

Atrial premature complexes are the most common arrhythmia identified during extended ECG monitoring. The incidence of APCs frequently increases with age and with the presence of structural heart diseases. Atrial premature complexes typically are asymptomatic, although some patients experience palpitations or an irregularity of the pulse.

ECG DIAGNOSIS OF APCS

The ECG diagnosis of APCs is based on the identification of a P wave that occurs before the anticipated sinus beat **(Fig. 233-3*A*** and ***B*)**. The source of the APC appears to parallel the typical sites of origin for ATs. The P-wave contour typically differs from that noted during sinus rhythm, although APCs from the right atrial appendage, superior vena cava (SVC), and superior aspect of the crista terminalis in the region of the sinus node may mimic the sinus P-wave morphology. In response to an APC, the PR interval lengthens, although APCs that originate near the atrioventricular (AV) nodal region may actually have a shorter PR interval because the atrial conduction time to the junction is shortened. A very early APC may not conduct to the ventricle and can create a pulse irregularity that may be perceived as a pause or "dropped beat." If the APC conducts rapidly through the AV node, a partially recovered His-Purkinje system will be encountered and a QRS pattern consistent with a right or left bundle branch block may occur. This wide QRS pattern and the failure to recognize the preceding P wave may result in misdiagnosis of VPCs. APCs characteristically reset the sinus node. The resulting sum of the pre- and post-APC RR interval is less than two sinus PP intervals.

Figure 233-3

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

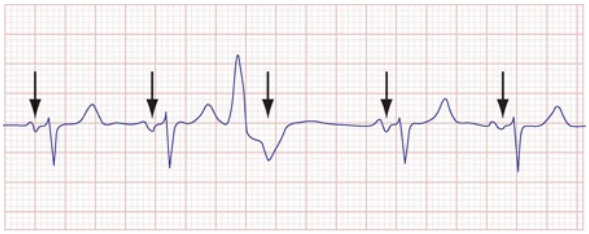

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial and ventricular premature complexes (APCs, VPCs). The APC resets the sinus node, and no compensatory pause is present ***A.*** even when conducted aberrantly in the ventricles with a bundle branch block-type QRS pattern ***B.*** VPCs tend not to reset the sinus activity (*arrows*) and will demonstrate a full compensatory pause ***C.***

TREATMENT: ATRIAL PREMATURE COMPLEXES

Atrial premature complexes generally do not require intervention. For extremely symptomatic patients who do not respond to explanation and reassurance, an attempt can be made to suppress the APCs with pharmacologic agents. The repetitive focus can even be targeted for catheter ablation. Beta blockers may be tried. Of note, these agents may uncommonly exacerbate symptoms if AV block occurs with the APC and irregularity of the pulse consequently becomes more profound. The use of class IC antiarrhythmic agents may eliminate the APCs but should be avoided if structural heart disease is present.

Junctional Premature Complexes

Junctional premature complexes are extremely uncommon. The complexes originate from the AV node and His bundle region and may produce retrograde atrial activation with the P wave distorting the initial or terminal portions of the QRS complex, producing pseudo Q or S waves in leads II, III, and aVF. Extrasystoles originating in the bundle of His that do not conduct to the ventricle and also block the atria can produce unexplained surface ECG PR prolongation that does not follow a typical Wenckebach periodicity (i.e., gradual PR prolongation culminating in atrial activity that fails to conduct to the ventricles). Intracardiac recordings frequently can identify a His depolarization, thus identifying the origin of the complex to the AV junction. Symptomatic patients typically may be treated with beta blockers or, if there is no structural heart disease, class IC antiarrhythmic agents.

Sinus Tachycardia

Physiologic sinus tachycardia represents a normal or appropriate response to physiologic stress, such as that which occurs with exercise, anxiety, or fever. Pathologic conditions such as thyrotoxicosis, anemia, and hypotension also may produce sinus tachycardia. It is important to distinguish sinus tachycardia from other SVTs. Sinus tachycardia will produce a P-wave contour consistent with its origin from the sinus node located in the superior-lateral and posterior aspect of the right atrium. The P wave is upright in leads II, III, and aVF and negative in lead aVR. The P-wave morphology in lead V_1 characteristically has a biphasic, positive/negative contour. Onset of sinus tachycardia is gradual, and in response to carotid sinus pressure there may be some modest and transient slowing but no abrupt termination. Importantly, the diagnosis should not be based on the PR interval or the presence of a P wave before every QRS complex. The PR interval and the presence of 1:1 AV conduction properties are determined by AV nodal and His-Purkinje conduction; therefore, the PR interval can be dramatically prolonged while sinus tachycardia remains the atrial mechanism.

TREATMENT: PHYSIOLOGIC SINUS TACHYCARDIA

Treatment of physiologic sinus tachycardia is directed at the underlying condition causing the tachycardia response. Uncommonly, beta blockers are used to minimize the tachycardia response if it is determined to be potentially harmful, as may occur in a patient with ischemic heart disease and rate-related anginal symptoms.

Inappropriate sinus tachycardia represents an uncommon but important medical condition in which the heart rate increases either spontaneously or out of proportion to the degree of physiologic stress/exercise. Dizziness and even frank syncope often accompany the sinus tachycardia and symptoms of palpitations. The syndrome can be quite disabling. Associated symptoms of chest pain, headaches, and gastrointestinal upset are common. In many patients, the syndrome occurs after a

viral illness and may resolve spontaneously over the course of 3–12 months, suggesting a postviral dysautonomia.

Excluding the diagnosis of an automatic AT that originates in the region of the sinus node can be difficult and may require invasive electrophysiologic evaluation. Frequently, patients are misdiagnosed as having an anxiety disorder with physiologic sinus tachycardia.

TREATMENT: INAPPROPRIATE SINUS TACHYCARDIA

For symptomatic patients, maintaining an increased state of hydration, salt loading, and careful titration of beta blockers to the maximum tolerated dose, administered in divided doses, frequently minimize symptoms. For severely symptomatic patients who are intolerant of or unresponsive to beta blockers, catheter ablation directed at modifying the sinus node may be effective. Because of the high recurrence rate after ablation and the frequent need for atrial pacing therapy, this intervention remains second-line treatment.

Atrial Fibrillation

(Fig. 233-4) Atrial fibrillation is the most common sustained arrhythmia. It is marked by disorganized, rapid, and irregular atrial activation. The ventricular response to the rapid atrial activation is also irregular. In an untreated patient, the ventricular rate also tends to be rapid and is entirely dependent on the conduction properties of the AV junction. Although typically the rate will vary between 120 and 160 beats per minute, in some patients it can be >200 beats per minute. In other patients, because of heightened vagal tone or intrinsic AV nodal conduction properties, the ventricular response is <100 beats per minute and occasionally even profoundly slow. The mechanism for AF initiation and maintenance, although still debated, appears to be a complex interaction between drivers responsible for the initiation and the complex anatomic atrial substrate that promotes the maintenance of multiple wavelets of (micro)reentry. The drivers appear to originate predominantly from the atrialized musculature that enters the pulmonary veins and represent either focal abnormal automaticity or triggered firing that is somewhat modulated by autonomic influences. Sustained forms of microreentry as drivers also have been documented around the orifice of pulmonary veins; nonpulmonary vein drivers also have been demonstrated. The role these drivers play in maintaining the tachycardias may be significant and may explain the success of pulmonary vein isolation procedures in eliminating more chronic or persistent forms of AF.

Figure 233-4

Supraventricular tachycardias with irregular ventricular rates. Atrial fibrillation ***A.*** atrial flutter ***B.*** atrial tachycardia ***C.*** and multifocal atrial tachycardia (MAT; ***D.***) are shown. The characteristics of the atrial activity with respect to the morphology and rate provide the clues to the diagnosis. The variable ventricular response to the atrial flutter and the atrial tachycardia suggest a Wenckebach-type periodicity.

Although AF is common in the adult population, it is extremely unusual in children unless structural heart disease is present or there is another arrhythmia that precipitates the AF, such as paroxysmal SVT in patients with WPW syndrome. The incidence of AF increases with age such that >5% of the adult population over 70 will experience the arrhythmia. As many patients are asymptomatic with AF, it is anticipated that the overall incidence, particularly that noted in the elderly, may be more than double previously reported rates. Occasionally, AF appears to have a well-defined etiology, such as acute hyperthyroidism, an acute vagotonic episode, or acute alcohol intoxication. Acute AF is particularly common during the acute or early recovery phase of major vascular, abdominal, and thoracic surgery, in which case autonomic fluxes and/or direct mechanical irritation potentiate the arrhythmia. AF also may be triggered by other supraventricular tachycardias (p. 29), such as AV nodal reentrant tachycardia (AVNRT), and elimination of these arrhythmias may prevent AF recurrence.

AF has clinical importance related to (1) the loss of atrial contractility, (2) the inappropriate fast ventricular response, and (3) the loss of atrial appendage contractility and emptying leading to the risk of clot formation and subsequent thromboembolic events.

Symptoms from AF vary dramatically. Many patients are asymptomatic and have no apparent hemodynamic consequences from the development of AF. Other patients experience only minor palpitations or sense irregularity of the pulse. Many patients, however, experience severe palpitations. The hemodynamic effect in patients can be quite dramatic, depending on the need for normal atrial contractility and the ventricular response. Hypotension, pulmonary congestion, and anginal symptoms may be severe in some patients. In patients with the LV diastolic dysfunction that occurs with hypertension, hypertrophic cardiomyopathy, or obstructive aortic valvular disease, symptoms may be even more dramatic, especially if the ventricular rate does not permit adequate ventricular filling. Exercise intolerance and easy fatigability are the hallmarks of poor rate control with exertion. Occasionally, the only manifestation of AF is severe dizziness or syncope associated with the pause that occurs upon termination of AF before sinus rhythm resumes (Fig. 233-1).

The ECG in AF is characterized by the lack of organized atrial activity and the irregularly irregular ventricular response. Occasionally, one needs to record from multiple ECG leads simultaneously to identify the chaotic continuous atrial activation. Lead V_1 frequently shows the appearance of organized atrial activity that mimics AFL. This occurs because the crista terminalis serves as an effective anatomic barrier to electrical conduction, and the activation of the lateral right atrium may be represented by a more uniform activation wavefront that originates over the roof of the right atrium. ECG assessment of the PP interval (<200 ms) and the chaotic P-wave morphology in the remaining ECG leads will confirm the presence of AF.

Evaluation of a patient with AF should include a search for a reversible cause of the arrhythmia, such as hyperthyroidism or anemia. An echocardiogram should be performed to determine whether there is structural heart disease. Persistent or labile hypertension should be identified and treated.

TREATMENT: ATRIAL FIBRILLATION

Treatment for AF must take into account the clinical situation in which the arrhythmia is encountered, the chronicity of the AF, the status of the patient's level of anticoagulation, risk factors for stroke, the patient's symptoms, the hemodynamic impact of the AF, and the ventricular rate.

Acute Rate Control

In the absence of hemodynamic compromise that might warrant emergent cardioversion to terminate the AF, the initial goals of therapy are to (1) establish control of the ventricular rate and (2) address anticoagulation status and begin IV heparin treatment if the duration of AF is >12 h and risk factors for stroke with AF are present **(Table 233-1)**. Ventricular rate control for acute AF is best established with beta blockers and/or the calcium channel blocking agents verapamil and diltiazem. The route of administration and dose will be dictated by the ventricular rate and clinical status. Digoxin may add to the rate-controlling benefit of the other agents but is uncommonly used as a stand-alone agent, especially in acute AF.

Table 233–1. Risk Factors for Stroke in Atrial Fibrillation

History of stroke or transient ischemic attack	Age >75 years
Mitral stenosis	Congestive heart failure
Hypertension	Left ventricular dysfunction
Diabetes mellitus	Marked left atrial enlargement (>5.0 cm)
	Spontaneous echo contrast

Anticoagulation is of particular importance in patients who have known risk factors for stroke associated with AF. Factors associated with the highest risk of stroke include a history of stroke, transient ischemic attack (TIA) or systemic embolism, and the presence of rheumatic mitral stenosis. Other identified risk factors include age >65 years, history of congestive heart failure (CHF), diabetes mellitus, hypertension, LV dysfunction, and evidence of marked left atrial enlargement (>5.0 cm). Chronic anticoagulation with warfarin targeted to achieve an international normalized ratio (INR) between 2.0 and 3.0 is recommended in patients with persistent or frequent and long-lived paroxysmal AF and risk factors. If patients have not been adequately anticoagulated and the AF is more than 24–48 h in duration, a transesophageal echocardiogram (TEE) can be performed to exclude the presence of a left atrial thrombus that might dislodge with the attempted restoration of sinus rhythm with either nonpharmacologic or pharmacologic therapy. Anticoagulation must be instituted coincident with the TEE and maintained for at least 1 month after restoration of sinus rhythm if the duration of AF has been prolonged or is unknown. Heparin is maintained routinely until the INR is 1.8 with the administration of warfarin after the TEE. For patients who do not warrant early cardioversion of AF, anticoagulation should be maintained for at least 3 weeks with the INR confirmed to be >1.8 on at least two separate occasions before attempts at cardioversion.

Termination of AF acutely may be warranted on the basis of clinical parameters and/or hemodynamic status. Confirmation of appropriate anticoagulation status as described above must be documented unless symptoms and clinical status warrant emergent intervention. Direct current transthoracic cardioversion during short-acting anesthesia is a reliable way to terminate AF. Conversion rates using a 200-J biphasic shock delivered synchronously with the QRS complex typically are >90%. Pharmacologic therapy to terminate AF is less reliable. Oral and/or IV administration of amiodarone or procainamide has only modest success. The acute IV administration of ibutilide appears to be somewhat more effective and may be used in selected patients to facilitate termination with direct current (DC) cardioversion **(Tables 233-2** and **233-3)**.

Table 233–2. Commonly Used Antiarrhythmic Agents—Intravenous Dose Range/Primary Indication

Drug	Loading	Maintenance	Primary Indication	Class*
Adenosine	6–18 mg (rapid bolus)	N/A	Terminate reentrant SVT involving AV node	—
Amiodarone	15 mg/min for 10 min, 1 mg/min for 6 h	0.5–1 mg/min	AF, AFL, SVT, VT/VF	III
Digoxin	0.25 mg q2h until 1 mg total	0.125–0.25 mg/d	AF/AFL rate control	—
Diltiazem	0.25 mg/kg over 3–5 min (max 20 mg)	5–15 mg/h	SVT, AF/AFL rate control	IV
Esmolol	500 μg/kg over 1 min	50 μg/kg per min	AF/AFL rate control	II
Ibutilide	1 mg over 10 min if over 60 kg	N/A	Terminate AF/AFL	III

Lidocaine	1–3 mg/kg at 20–50 mg/min	1–4 mg/min	VT	IB
Metoprolol	5 mg over 3–5 min x 3 doses	1.25–5 mg q6h	SVT, AF rate control; exercise-induced VT; long QT	II
Procainamide	15 mg/kg over 60 min	1–4 mg/min	Convert/prevent AF/VT	IA
Quinidine	6–10 mg/kg at 0.3–0.5 mg/kg per min	N/A	Convert/prevent AF/VT	IA
Verapamil	5–10 mg over 3–5 min	2.5–10 mg/h	SVT, AF rate control	IV

* Classification of antiarrhythmic drugs: Class I—agents that primarily block inward sodium current; class IA agents also prolong action potential duration; class II—antisympathetic agents; class III—agents that primarily prolong action potential duration; class IV—calcium channel-blocking agents.

Abbreviations: AF, atrial fibrillation; AFL, atrial flutter; AV, atrioventricular; SVT, supraventricular tachycardia; VF, ventricular fibrillation; VT, ventricular tachycardia.

Table 233–3. Commonly Used Antiarrhythmic Agents: Chronic Oral Dosing/Primary Indications

Drug	Dosing Oral, mg, Maintenance	Half-Life, h	Primary Route(s) of Metabolism/Elimination	Most Common Indication	Class[a]
Acebutolol	200–400 q12h	6–7	Renal/hepatic	AF rate control/SVT	II
				Long QT/RVOT VT	
Amiodarone	100–400 qd	40–55 d	Hepatic	AF/VT prevention	III[b]
Atenolol	25–100 per d	6–9	Renal	AF rate control/SVT	II
				Long QT/RVOT VT	
Digoxin	0.125–0.5 qd	38–48	Renal	AF rate control	—
Diltiazem	30–60 q6h	3–4.5	Hepatic	AF rate control/SVT	IV
Disopyramide	100–300 q6–8h	4–10	Renal 50%/hepatic	AF/SVT prevention	Ia
Dofetilide	0.125–0.5 q12h	10	Renal	AF prevention	III
Dronedarone	400 q12 hr	13-19	Hepatic	AF prevention	IIIb
Flecainide	50–200 q12h	7–22	Hepatic 75%/renal	AF/SVT/VT prevention	Ic
Metoprolol	25–100 q6h	3–8	Hepatic	AF rate control/SVT	II
				Long QT/RVOT VT	
Mexiletine	150–300 q8–12h	10–14	Hepatic	VT prevention	Ib
Moricizine	100–400 q8h	3–13	Hepatic 60%/renal	AF prevention	Ic
Nadolol	40–240 per d	10–24	Renal	Same as metoprolol	II
Procainamide	250–500 q3–6h	3–5	Hepatic/renal	AF/SVT/VT prevention	Ia
Propafenone	150–300 q8h	2–8	Hepatic	AF/SVT/VT prevention	Ic
Quinidine	300–600 q6h	6–8	Hepatic 75%/renal	AF/SVT/VT prevention	Ia
Sotalol	80–160 q12h	12	Renal	AF/VT prevention	III
Verapamil	80–120 q6–8h	4.5–12	Hepatic/renal	AF rate control/RVOT VT	IV
				Idiopathic LV VT	

[a]Classification of antiarrhythmic drugs: Class I—agents that primarily block inward sodium current; class II—antisympathetic agents; class III—agents that primarily prolong action potential duration; class IV—calcium channel-blocking agents.

[b]Amiodarone and dronedarone both are grouped in class III, but both also have class I, II, and IV properties.

Abbreviations: AF, atrial fibrillation; LV, left ventricular; RVOT, right ventricular outflow tract; SVT, supraventricular tachycardia; VT, ventricular tachycardia.

Pharmacologic therapy to maintain sinus rhythm can be instituted once sinus rhythm has been established or in anticipation of cardioversion to attempt to maintain sinus rhythm (Table 233-3). A single episode of AF may not warrant any intervention or only a short course of beta blocker therapy. To prevent recurrent AF unresponsive to beta blockade, a trial of antiarrhythmic therapy may be warranted, particularly if the AF is associated with rapid rates and/or significant symptoms. The selection of antiarrhythmic agents should be dictated primarily by the presence or absence of CAD, depressed LV function not attributable to a reversible tachycardia-induced cardiomyopathy, and/or severe hypertension with evidence of marked LV hypertrophy. The presence of any significant structural heart disease typically narrows treatment to the use of sotalol, amiodarone, dofetilide, or dronedarone. Severely depressed LV function with heart failure symptoms precludes the use of dronedarone and may limit sotalol therapy. Owing to the risk of QT prolongation and polymorphic VT, sotalol and dofetilide have to be initiated in the hospital in most cases.

In patients without evidence of structural heart disease or hypertensive heart disease without evidence of severe hypertrophy, the use of the class IC antiarrhythmic agents flecainide or propafenone appears to be well tolerated and does not have significant proarrhythmia risk. It is important to recognize that no drug is uniformly effective, and arrhythmia recurrence should be anticipated in over one-half of the patients during long-term follow-up regardless of the type and number of agents tried. It is also important to recognize that although the maintenance of sinus rhythm has been associated with improved long-term survival, the survival outcome of patients randomized to the pharmacologic maintenance of sinus rhythm was not superior to that of patients treated with rate control and anticoagulation in the AFFIRM and RACE trials. The AFFIRM and RACE trials compared outcome with respect to survival and thromboembolic

events in patients with AF and risk factors for stroke using the two treatment strategies. It is believed that the poor outcome related to pharmacologic therapy used to maintain sinus rhythm was primarily due to the common inefficacy of such drug therapy and an increased incidence of asymptomatic AF. Many of the drugs used for rhythm control, including sotalol, amiodarone, propafenone, dronedarone, and flecainide, enhance slowing of AV nodal conduction. The absence of symptoms frequently leads to stopping anticoagulant therapy, and asymptomatic AF without anticoagulation increases stroke risk. Any consideration for stopping anticoagulation therefore must be accompanied by a prolonged period of ECG monitoring to document asymptomatic AF. It is also recommended that patients participate in monitoring by learning to take their pulse on a twice-daily basis and reliably identify its regularity if discontinuing anticoagulant therapy is contemplated seriously.

It is clear that to reduce the risk of drug-induced complications in treating AF, a thorough understanding of the drug planned to be used is critical—its dosing, metabolism, and common side effects and important drug-drug interactions. This information has been summarized in Tables 233-2, 233-3, **233-4**, and **233-5** and serves as a starting point for a more complete review. In using antiarrhythmic agents that slow atrial conduction, strong consideration should be given to adding a beta blocker or a calcium channel blocker (verapamil or diltiazem) to the treatment regimen. This should help avoid a rapid ventricular response if AF is converted to "slow" AFL with the drug therapy **(Fig. 233-5)**.

Table 233–4. Common Nonarrhythmic Toxicity of Most Frequently Used Antiarrhythmic Agents

Drug	Common Nonarrhythmic Toxicity
Amiodarone	Tremor, peripheral neuropathy, pulmonary inflammation, hypo- and hyperthyroidism, photosensitivity
Adenosine	Cough, flushing
Digoxin	Anorexia, nausea, vomiting, visual changes
Disopyramide	Anticholinergic effects, decreased myocardial contractility
Dofetilide	Nausea
Dronedarone	Gastointestinal intolerance, exacerbation of heart failure
Flecainide	Dizziness, nausea, headache, decreased myocardial contractility
Ibutilide	Nausea
Lidocaine	Dizziness, confusion, delirium, seizures, coma
Mexiletine	Ataxia, tremor, gait disturbances, rash, nausea
Moricizine	Mood changes, tremor, loss of mental clarity, nausea
Procainamide	Lupus erythematosus-like syndrome (more common in slow acetylators), anorexia, nausea, neutropenia
Propafenone	Taste disturbance, dyspepsia, nausea, vomiting
Quinidine	Diarrhea, nausea, vomiting, cinchonism, thrombocytopenia
Sotalol	Hypotension, bronchospasm

Table 233–5. Proarrhythmic Manifestations of Most Frequently Used Antiarrhythmic Agents

Drug	Common Proarrhythmic Toxicity
Amiodarone	Sinus bradycardia, AV block, increase in defibrillation threshold
	Rare: long QT and torsades des pointes, 1:1 ventricular conduction with atrial flutter
Adenosine	All arrhythmias potentiated by profound pauses, atrial fibrillation
Digoxin	High-grade AV block, fascicular tachycardia, accelerated junctional rhythm, atrial tachycardia
Disopyramide	Long QT and torsades des pointes, 1:1 ventricular response to atrial flutter; increased risk of some ventricular tachycardias in patients with structural heart disease
Dofetilide	Long QT and torsades des pointes
Dronedarone	Bradyarrhythmias and AV block, long QT and torsades des pointes
Flecainide	1:1 Ventricular response to atrial flutter; increased risk of some ventricular tachycardias in patients with structural heart disease; sinus bradycardia
Ibutilide	Long QT and torsades des pointes
Procainamide	Long QT and torsades des pointes, 1:1 ventricular response to atrial flutter; increased risk of some ventricular tachycardias in patients with structural heart disease
Propafenone	1:1 Ventricular response to atrial flutter; increased risk of some ventricular tachycardias in patients with structural heart disease; sinus bradycardia
Quinidine	Long QT and torsades des pointes, 1:1 ventricular response to atrial flutter; increased risk of some ventricular tachycardias in patients with structural heart disease
Sotalol	Long QT and torsades des pointes, sinus bradycardia

Abbreviation: AV, atrioventricular.

Figure 233-5

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial fibrillation *A.* transitions to "slow" atrial flutter during antiarrhythmic drug therapy. ***B.*** A rapid ventricular response with 1:1 atrioventricular conduction occurred with exercise, leading to ***C.*** symptoms of dizziness.

Chronic Rate Control

This is an option in patients who are asymptomatic or symptomatic due to the resulting tachycardia. Rate control is frequently difficult to achieve in patients who have paroxysmal AF. In patients with more persistent forms of AF, rate control with beta blockers, the calcium channel blockers diltiazem and verapamil, and/or digoxin frequently can be achieved. Using the drugs in combination may avoid some of the common side effects seen with high-dose monotherapy. An effort should be made to document the adequacy of rate control to reduce the risk of a tachycardia-induced cardiomyopathy. Heart rates >80 beats/min at rest or 100 beats/min with very modest physical activity are indications that rate control may be inadequate in persistent AF. Extended periods of ECG monitoring and assessment of heart rate with exercise should be considered.

In patients with symptoms resulting from inadequate rate control with pharmacologic therapy or worsening LV function due to the persistent tachycardia, ablative therapy to attempt to eliminate atrial fibrillation, or an AV junction ablation can be performed. The AV junction ablation must be coupled with the implantation of an activity sensor pacemaker to maintain a physiologic range of heart rates. Recent evidence that RV pacing can occasionally modestly depress LV function should be taken into consideration in identifying which patients are appropriate candidates for the "ablate and pace" treatment strategy. Occasionally, biventricular pacing may be used to minimize the degree of dyssynchronization that can occur with RV apical pacing alone. Rate control treatment options must be coupled with chronic anticoagulation therapy in all cases. Trials evaluating the elimination of embolic risk by elimination or isolation of the left atrial appendage or by endovascular insertion of a left atrial appendage-occluding device may provide other treatment options that can eliminate the need for chronic anticoagulation.

Catheter and Surgical Ablative Therapy to Prevent Recurrent AF

Although the optimum ablation strategy has not been defined, most ablation strategies incorporate techniques that isolate the atrial muscle sleeves entering the pulmonary veins; these muscle sleeves have been identified as the source of the majority of triggers responsible for the initiation of AF. Ablation therapy is currently considered an alternative to additional pharmacologic therapy trials in patients with recurrent symptomatic AF or AF associated with poor rate control who have failed an initial attempt at rhythm control with pharmacologic management Ablative therapy appears superior to additional pharmacologic treatment aimed at rhythm control in this setting. Elimination of AF in 50–80% of patients with a catheter-based ablation procedure should be anticipated, depending on the chronicity of the AF, with additional patients becoming responsive to previously ineffective medications.

Catheter ablative therapy also holds promise in patients with more persistent forms of AF and even those with severe atrial dilation. Its confirmed efficacy suggests an important alternative to His bundle ablation and pacemaker insertion in many patients. Serious risks related to the left atrial ablation procedure, albeit low (overall 2–4%), include pulmonary vein stenosis, atrioesophageal fistula, systemic embolic events, perforation/tamponade, and phrenic nerve injury.

Surgical ablation of AF is typically performed at the time of other cardiac valve or coronary artery surgery and, less commonly, as a stand-alone procedure. The surgical Cox-Maze procedure is designed to interrupt all macroreentrant circuits that might potentially develop in the atria, thereby precluding the ability of the atria to fibrillate. In an attempt to simplify the operation, the multiple incisions of the traditional Cox-Maze procedure have been replaced with linear lines of ablation and pulmonary vein isolation using a variety of energy sources.

Severity of AF symptoms and difficulties in rate and/or rhythm control with pharmacologic therapy frequently dictate the optimum AF treatment strategy. Similar to the approach with pharmacologic rhythm control, a cautious approach to eliminating anticoagulant therapy is recommended after catheter or surgical ablation. Careful ECG monitoring for asymptomatic AF, particularly in patients with multiple risk factors for stroke, should be considered until guidelines are firmly established. If the left atrial appendage has been removed surgically, the threshold for stopping anticoagulation may be lowered. Antiarrhythmic therapy typically can be discontinued after catheter or surgical ablation of AF. However, in selected patients, satisfactory AF control may require maintenance of previously ineffective drug therapy after the ablation intervention.

Atrial Flutter and Macroreentrant Atrial Tachycardias

Macroreentrant arrhythmias involving the atrial myocardium are referred to collectively as AFL. The terms *AFL* and *macroreentrant AT* frequently are used interchangeably, with both denoting a nonfocal source of an atrial arrhythmia. The typical or most common AFL circuit rotates in a clockwise or counterclockwise direction in the right atrium around the tricuspid valve annulus. The posterior boundary of the right AFL circuit is defined by the crista terminalis, the eustachian ridge, and the inferior and superior vena cavae. Counterclockwise right AFL represents ~80% of all AFL with superiorly directed activation of the interatrial septum, which produces the saw-toothed appearance of the P waves in ECG leads II, III, and aVF. Clockwise rotation of the same right atrial circuit produces predominantly positive P waves in leads II, III, and aVF (Fig. 233-4). Macroreentrant left AFL also may develop, albeit much less commonly. This type of arrhythmia may be

the sequela of surgical or catheter-based ablation procedures that create large anatomic barriers or promote slowing of conduction in the left atrium, especially around the mitral valve annulus or partially disconnected pulmonary veins. Atypical AFL or macroreentrant AT can also develop around incisions created during surgery for valvular or congenital heart disease or in and/or around large areas of atrial fibrosis.

Classic or typical right AFL has an atrial rate of 260–300 beats per minute with a ventricular response that tends to be 2:1, or typically 130–150 beats per minute. In the setting of severe atrial conduction disease and or antiarrhythmic drug therapy, the atrial rate can slow to <200 beats per minute. In this setting, a 1:1 rapid ventricular response may occur, particularly with exertion, and produce adverse hemodynamic effects (Fig. 233-5). Atypical AFL or macroreentrant AT related to prior surgical incisions and atrial fibrosis demonstrates less predictability in terms of the atrial rate and is more likely to demonstrate slower rates that overlap with those identified with focal atrial tachycardias (p. 27).

Because lead V_1 is frequently monitored in a hospitalized patient, coarse AF may be misdiagnosed as AFL. This occurs because in both typical right AFL and coarse AF the crista terminalis in the right atrium may serve as an effective anatomic barrier. The free wall of the right atrium, whose electrical depolarization is best reflected on the body surface by lead V_1, may demonstrate a uniform wavefront of atrial activation in both conditions. The timing of atrial activation is much more rapid in AF and always demonstrates variable atrial intervals with some intervals between defined P waves <200 ms **(Fig. 233-6)**. A review of the other ECG leads demonstrates the disorganized atrial depolarization that is characteristic of AF. Frequently, an individual patient may alternate between AF and AFL or, less commonly, may manifest AF in one atrium and AFL in the other, making the distinction more difficult.

Figure 233-6

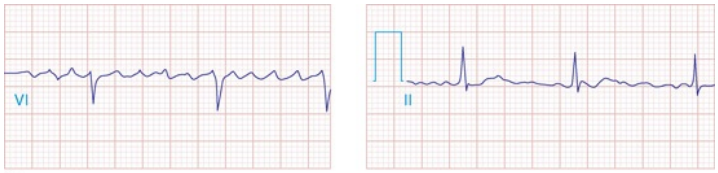

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

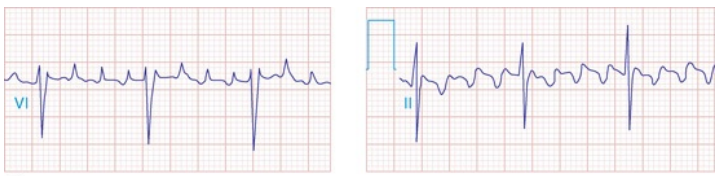

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial flutter/atrial fibrillation. Coarse atrial fibrillation ***A.*** contrasted with organized atrial flutter ***B.***

TREATMENT: ATRIAL FLUTTER

Because of the anticipated rapid regular ventricular rate associated with AFL and the failure to respond to pharmacologic therapy directed at slowing the ventricular rate, patients frequently are treated with DC cardioversion. The organized atrial flutter activity frequently can be terminated with low-energy external cardioversion of 50–100 J. The risk of thromboembolic events associated with typical AFL is high, and anticoagulation must be managed similarly to what was described for patients with AF (p. 23).

Asymptomatic patients with AFL may develop heart failure symptoms with tachycardia-induced severe LV dysfunction. In all patients, an effort should be made to control the ventricular rate pharmacologically or restore sinus rhythm. Rate control with calcium antagonists (diltiazem or verapamil), beta blockers, and/or digoxin may be difficult. Even higher-grade AV slowing, such as a 4:1 AV response, may be only transient and is easily overcome with activity or emotional stress. Owing to the typically faster ventricular rate, AFL tends to be poorly tolerated in comparison to AF.

In selected patients with high anesthetic risk, an attempt at pharmacologic cardioversion with procainamide, amiodarone, or ibutilide is appropriate. Antiarrhythmic drug therapy may also enhance the efficacy of DC cardioversion and the maintenance of sinus rhythm after cardioversion. Recurrence rates of AFL with pharmacologic attempts at rhythm control exceed 80% by 1 year.

Patients who manifest recurrent AFL appear to be effectively treated with catheter ablative therapy. For typical right AFL, an isthmus ablation line from the tricuspid annulus to the opening of the inferior vena cava can permanently eliminate flutter, with an anticipated success rate of >90% in most experienced centers. In patients with macroreentrant atrial tachycardia or AFL involving prior surgical incisions or catheter ablation or in areas of atrial fibrosis, detailed mapping of the arrhythmia circuit is required to design the best ablation strategy to interrupt the circuit. In selected patient with AF and typical right AFL, pharmacologic therapy may help prevent the AF but not the AFL. In this type of patient, hybrid therapy with antiarrhythmic agents coupled with a right atrial isthmus ablation may produce AF and AFL control.

Multifocal Atrial Tachycardia

Multifocal AT (MAT) is the signature tachycardia of patients with significant pulmonary disease. The atrial rhythm is characterized by at least three distinct P-wave morphologies and often at least three different PR intervals, and the associated atrial and ventricular rates are typically between 100 and 150 beats per minute. The presence of an isoelectric baseline distinguishes this arrhythmia from AF (Fig. 233-4). The absence of any intervening sinus rhythm distinguishes MAT from normal sinus rhythm with frequent multifocal APCs, although this distinction may be moot as these processes define an electrophysiologic continuum.

TREATMENT: MULTIFOCAL ATRIAL TACHYCARDIA

Therapy for MAT should be directed at improving the underlying medical condition, which is typically, although not

invariably, chronic obstructive or restrictive lung disease. Treatment with the calcium channel blocker verapamil also may provide some benefit. The judicious use of flecainide or propafenone may also decrease atrial arrhythmias. Patients should be screened for the presence of significant ventricular dysfunction or CAD before these agents are started. Low-dose amiodarone therapy may also control the arrhythmia and minimize the risk of pulmonary toxicity noted with the drug.

Focal Atrial Tachycardias

The two general mechanisms for focal ATs can be distinguished by observations made at AT initiation and in response to adenosine. *Automatic ATs* start with a "warm-up" period over the first 3–10 complexes and, similarly, slow in rate before termination. They may respond to adenosine not only with evidence of AV block but also with gradual slowing of the atrial rhythm and termination. The initiation of automatic ATs frequently can be provoked by isoproterenol infusion. The first P wave of the tachycardia has the same morphology as the remaining waves. Some of the ATs may be triggered or provoked by burst atrial pacing but are not reliably initiated by programmed atrial stimulation.

In contrast, evidence supporting a focal reentrant AT includes the initiation of the tachycardia with programmed atrial stimulation or spontaneous premature beats. The P wave initiating the tachycardia will characteristically have a different morphology than the P wave during the sustained AT. In response to adenosine, reentrant ATs will demonstrate AV block but typically do not slow and/or terminate. Most focal ATs in the absence of structural heart disease originate from specific anatomic locations. These anatomic locations appear to be associated with anatomic ridges, such as the crista terminalis, the valve annuli, and the limbus of the fossa ovalis. ATs also appear to originate from the muscular sleeves associated with the cardiac thoracic veins, i.e., the SVC, the coronary sinus, and the pulmonary veins. As was indicated, repetitive firing of these foci also appears to serve as the triggering mechanism for AF in most patients.

It is important to distinguish focal ATs from reentrant tachycardias that incorporate the AV node in the circuit **(Fig. 233-7)**. The primary distinction is related to the persistence of the AT in the presence of AV block that occurs spontaneously or is created by carotid sinus massage or the administration of adenosine (Fig. 233-4). Atrial activity drives the ventricles in AT and all changes in the PP interval accompanied by correlative changes in the RR intervals; in addition, the V–A relationship changes when the atrial rate changes. The P wave in AT is characteristically distinct from the sinus P-wave morphology, and unless there is significant AV nodal conduction delay, the PR interval is shorter than the measured RP interval when there is a 1:1 relationship between atria and ventricles (Fig. 233-7).

Figure 233-7

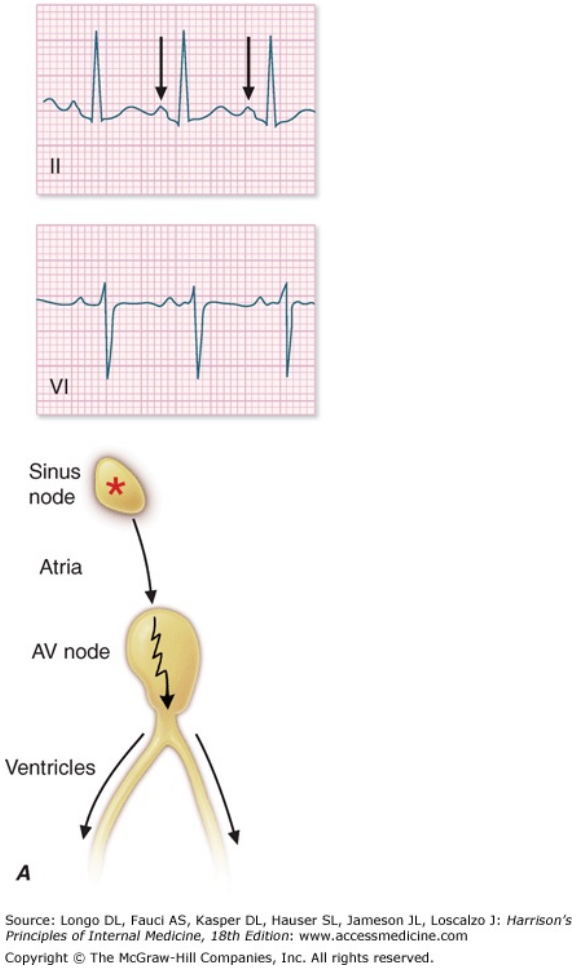

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

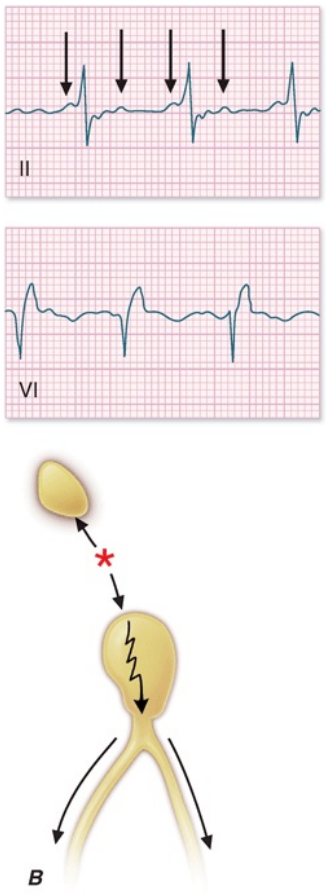

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

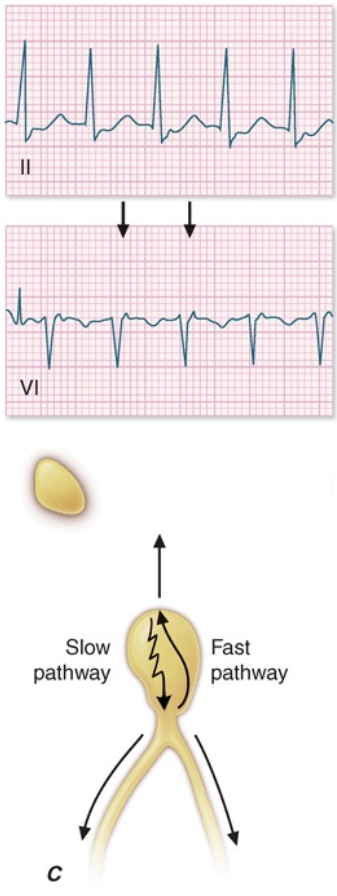

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

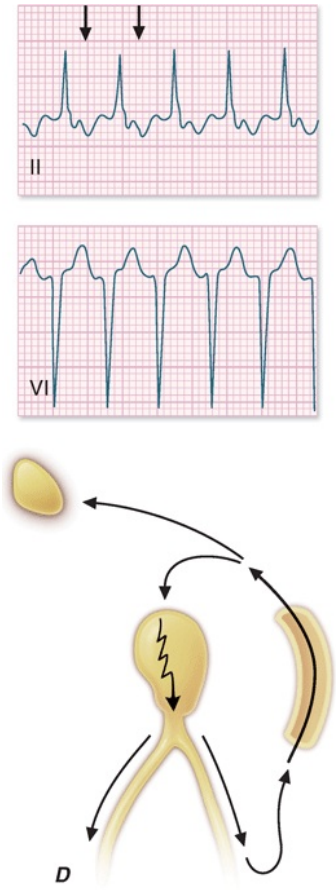

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pattern of atrial and ventricular activation and characteristic relationship of P-wave and QRS complex as recorded in leads II and V_1 during regular supraventricular tachycardias. ***A.*** Sinus tachycardia. ***B.*** Atrial tachycardia from top of the atria. ***C.*** Atrioventricular nodal reentry. ***D.*** Accessory pathway–mediated orthodromic supraventricular tachycardia.

The P wave for ATs depends on the anatomic site of origin. In addition to attempting to create AV block to establish the diagnosis of AT, analysis of the P-wave morphology on the 12-lead ECG may help exclude AV nodal reentry, AV bypass tract–mediated reentrant tachycardias, and physiologic or inappropriate sinus tachycardia (Fig. 233-7).

The ECG distinction between focal automatic or microreentrant and macroreentrant AT or atypical AFL is not always possible. Although sustained focal ATs tend to be slower, the atrial rates frequently overlap. Focal ATs, which are more common in the absence of structural heart disease, tend to demonstrate an isolectric baseline between P waves, whereas macroreentrant ATs represent atrial activation that is continuous and an isoelectric baseline between P waves frequently is absent. In patients with a history of prior atrial surgery, one must suspect a macroreentrant mechanism. These distinctions are less important with respect to acute management but have importance related to ablation strategies and anticipated outcome (pp. 28-29).

TREATMENT: ATRIAL TACHYCARDIA

Pharmacologic treatment of AT generally is approached in a similar fashion to that of AF and AFL. AV nodal blocking agents are administered in the setting of rapid ventricular rates. Acute IV administration of procainamide or amiodarone may terminate the tachycardia. Tachycardias that do not respond to pharmacologic therapy may be terminated with electrical cardioversion. Typically, anticoagulation before treatment is not needed unless there is evidence of severe atrial dilatation, >5 cm left atrial diameter with a high risk of AF, and/or a history of coincident paroxysmal AF. Most focal ATs are readily amenable to catheter ablative therapy. In patients who fail to respond to medical therapy or who are reluctant to take chronic drug therapy, this option should be considered, with an anticipated 90% cure rate. A parahisian location for the AT and/or a focus that is located in the left atrium may modestly increase the risk related to the procedure, and for this reason, every effort should be made to determine the likely origin of the AT based on an analysis of the P-wave morphology on 12-lead ECG before the procedure.

AV Nodal Tachycardias

AV NODAL REENTRANT TACHYCARDIA

Atrioventricular nodal reentrant tachycardia is the most common paroxysmal regular SVT. It is more commonly observed in women than in men and is typically manifest in the second to fourth decades of life. In general, because AVNRT tends to occur in the absence of structural heart disease, it is usually well tolerated. Neck pulsations are usually felt because of the simultaneous atrial and ventricular contraction, and a "frog sign" can be identified on physical examination during the arrhythmia. In the presence of hypertension or other forms of structural heart disease that limit ventricular filing, hypotension or syncope may occur.

Atrioventricular nodal reentrant tachycardia develops because of the presence of two electrophysiologically distinct pathways for conduction in the complex syncytium of muscle fibers that make up the AV node. The fast pathway in the more superior part of the node has a longer refractory period, whereas the pathway lower in the AV node region conducts more slowly but has a shorter refractory period. As a result of the inhomogeneities of conduction and refractoriness, a reentrant

circuit can develop in response to premature stimulation. Although conduction occurs over both pathways during sinus rhythm, only the conduction over the fast pathway is manifest, and as a result, the PR interval is normal. APCs occurring at a critical coupling interval are blocked in the fast pathway because of the longer refractory period and are conducted slowly over the slow pathway. When sufficient conduction slowing occurs, the blocked fast pathway can recover excitability and atrial activation can occur over the fast pathway to complete the circuit. Repetitive activation down the slow and up the fast pathway results in typical AV nodal reentrant tachycardia (Fig. 233-7).

ECG Findings in AVNRT

The APC initiating AVNRT is characteristically followed by a long PR interval consistent with conduction via the slow pathway. AVNRT is manifest typically as a narrow QRS complex tachycardia at rates that range from 120 to 250 beats/min. The QRS-P wave pattern associated with typical AVNRT is quite characteristic, with simultaneous activation of the atria and ventricles from the reentrant AV nodal circuit. The P wave frequently is buried inside the QRS complex and either will not be visible or will distort the initial or terminal portion of the QRS complex (Fig. 233-7). Because atrial activation originates in the region of the AV node, a negative deflection will be generated by retrograde atrial depolarization when recording ECG leads II, III, or AVF.

Occasionally, AVNRT occurs with activation in the reverse direction, conducting down the fast pathway and returning up the slow pathway. This form of AVNRT occurs much less commonly and produces a prolonged RP interval during the tachycardia with a negative P wave in leads II, III, and aVF. This atypical form of AVNRT is more easily precipitated by ventricular stimulation.

TREATMENT: ATRIOVENTRICULAR NODAL REENTRANT TACHYCARDIA

Acute Treatment

Treatment is directed at altering conduction within the AV node. Vagal stimulation, such as that which occurs with the Valsalva maneuver or carotid sinus massage, can slow conduction in the AV node sufficiently to terminate AVNRT. In patients in whom physical maneuvers do not terminate the tachyarrhythmia, the administration of adenosine, 6–12 mg IV, frequently does so. Intravenous beta blockade or calcium channel therapy should be considered as second-line treatment. If hemodynamic compromise is present, R-wave synchronous DC cardioversion using 100–200 J can terminate the tachyarrhythmia.

Prevention

Prevention may be achieved with drugs that slow conduction in the antegrade slow pathway, such as digitalis, beta blockers, and calcium channel blockers. In patients who have a history of exercise-precipitated AVNRT, the use of beta blockers frequently eliminates symptoms. In patients who do not respond to drug therapy directed at the antegrade slow pathway, treatment with class IA or IC agents directed at altering conduction of the fast pathway may be considered.

Catheter ablation, directed at elimination or modification of slow pathway conduction, is very effective in permanently eliminating AVNRT. Patients with recurrent AVNRT that produces significant symptoms or heart rates >200 beats/min and patients reluctant to take chronic drug therapy should be considered for ablative therapy. Catheter ablation can cure AV nodal reentry in >95% of patients with a single procedure. The risk of AV block requiring a permanent pacemaker is ~1% with the ablation procedure.

AV JUNCTIONAL TACHYCARDIAS

These can also occur in the setting of enhanced normal automaticity, abnormal automaticity, or triggered activity. These tachycardias may or may not be associated with retrograde conduction to the atria, and the P waves may appear dissociated or produce intermittent conduction and early activation of the junction. These arrhythmias may occur as a manifestation of increased adrenergic tone or drug effect in patients with sinus node dysfunction or after surgical or catheter ablation. The arrhythmia may also be a manifestation of digoxin toxicity. The most common manifestation of digoxin intoxication is the sudden regularization of the response to AF. A junctional tachycardia due to digoxin toxicity typically does not manifest retrograde conduction. Sinus activity may appear dissociated or result in intermittent capture beats with a long PR interval. If the rate is >50 beats per minute and <100 beats per minute, the term *accelerated junctional rhythm* applies. Occasionally, automatic rhythms are mimicked by AVNRT that fails to conduct to the atrium. The triggering events associated with the onset of the tachycardia may provide a clue to the appropriate diagnosis. Initiation of the tachycardia without an atrial premature beat with a gradual acceleration in rate suggests an automatic focus.

Treatment: Atrioventricular Junctional Tachycardias

Treatment of automatic/triggered junctional tachycardias is directed at decreasing adrenergic stimulation and reversing digoxin toxicity, if present. Digoxin therapy can be withheld if toxicity is suspected, and the administration of digoxin-specific antibody fragments can rapidly reverse digoxin toxicity if the tachycardia is producing significant symptoms and rapid termination is indicated. Junctional tachycardia due to abnormal automaticity can be treated pharmacologically with beta blockers. A trial of class IA or IC drugs may also be attempted. For incessant automatic junctional tachycardia, focal catheter ablation can be performed but is associated with an increased risk of AV block.

Tachycardias Associated with Accessory AV Pathways

Tachycardias that involve accessory pathways (APs) between atria and ventricles commonly manifest a normal QRS complex with a short or long RP interval. They must be considered in the differential diagnosis of other narrow-complex tachycardias. Importantly, most tachycardias associated with APs involve a large macroreentrant circuit that includes the ventricles (Fig. 233-7). Thus, identifying these arrhythmias as "supraventricular" is actually a misnomer, and they deserve separate consideration.

Accessory pathways are typically capable of conducting rapidly in both an antegrade and a retrograde direction. In the absence of an AP, the sinus impulse normally activates the ventricles via the AV node and His-Purkinje system, resulting in a PR interval of 120–200 ms. When an antegradely conducting AP is present, the sinus impulse bypasses the AV node and can activate the ventricles rapidly, resulting in ventricular preexcitation. The resulting PR interval is shorter than anticipated. In addition, because the initial ventricular activation is due to muscle-to-muscle conduction, as opposed to rapid spread of activation via the His-Purkinje system, the initial portion of the QRS complex is slurred, creating the characteristic "delta wave." The remaining portion of the QRS complex in sinus rhythm is created by a fusion of the ventricular activation wavefront originating from the Purkinje network and the continued spread of activation from the site of insertion of the AP **(Fig. 233-8)**. Evidence of ventricular preexcitation includes evidence in sinus rhythm of a short PR interval and a delta wave.

Figure 233-8

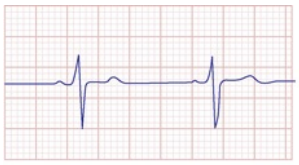

V_2

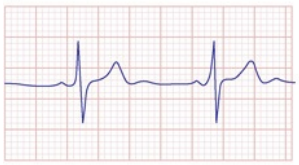

V_3

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

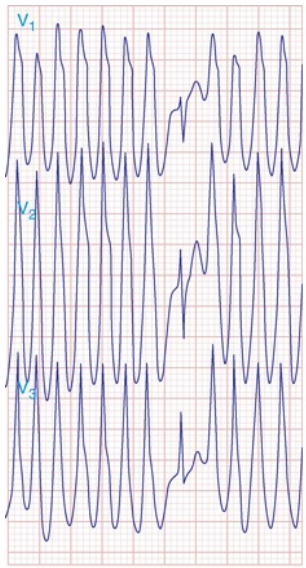

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Sinus rhythm tracing of leads V_1–V_3 showing evidence of Wolff-Parkinson-White syndrome with short PR interval and delta wave. ***B.*** During atrial fibrillation, rapid conduction to the ventricles is observed producing a wide QRS complex tachycardia with marked irregularity of the ventricular response and morphology of the QRS complex.

The most common AP connects the left atrium to the left ventricle, followed by posterior septal, right free wall, and anterior septal APs. APs typically insert from the atrium into the adjacent ventricular myocardium. However, occasionally pathways, particularly those originating from the right atrium, can have a ventricular insertion at a site distant from the AV groove in the fascicles. These pathways conduct more slowly and are referred to as *atriofascicular accessory pathways* . Atriofascicular APs are unique in their tendency to demonstrate decremental antegrade conduction.

Other accessory pathway connections from the AV node to the fascicles may exist. These pathways are referred to as *Mahaim fibers* and typically manifest a normal PR interval with a delta wave.

Patients with manifest preexcitation and WPW syndrome are typically subject to both macroreentrant tachycardias and a rapid response to AF (Fig. 233-8). The most common macroreentrant tachycardia associated with the WPW syndrome is referred to as *orthodromic AV reentry* . Ventricular activation occurs via the AV node and the His-Purkinje system. Conduction then returns or reenters the atria via retrograde conduction over the AP. The reentrant circuit develops because of the inhomogeneity in conduction and refractoriness in the AP and the normal AV node.

Characteristically, the AP has more rapid conduction but a longer refractory period than that of the AV node. Typical APs do not show evidence of antegrade decremental conduction. An APC can block in the AP and conduct sufficiently slowly or with decrement via the AV node to allow for retrograde recovery of activation of the AP and, in turn, of the atria (Fig. 233-7). This retrograde activation of the atria via the AP is referred to as an *echo beat*. If the pattern repeats itself, a tachycardia develops. Uncommonly, the reentrant circuit can be reversed so that the impulse reaches the ventricle via the AP and conducts retrogradely through to the atria via the His-Purkinje system and the AV node; this is referred to as *antidromic AV reentry* and/or *preexcitation macroreentry*, with the entire activation of the ventricle originating from the site of insertion of

the AP. Although it is uncommon, it is important to recognize antidromic SVT. The ECG pattern during the tachycardia mimics VT originating from the site of ventricular insertion of the AP. The presence of manifest preexcitation in sinus rhythm provides a valuable clue to the diagnosis.

The second most common and potentially more serious arrhythmia associated with the WPW syndrome is rapidly conducting AF. Nearly 50% of patients with evidence of APs are predisposed to episodes of AF. In patients who have rapid antegrade conduction from the atria to the ventricles over the AP, the AP can conduct rapidly in response to AF, resulting in a faster ventricular rate than would occur normally via the AV node. The rapid ventricular rates can result in hemodynamic compromise and even precipitate VF. The QRS pattern during AF in patients with manifest preexcitation can appear quite bizarre and change on a beat-to-beat basis due to the variability in the degree of fusion from activation over the AV node (Fig. 233-8).

CONCEALED APS

In ~50% of patients with APs, there is no antegrade conduction over the AP; however, retrograde conduction is preserved. As a result, the AP is not manifest in sinus rhythm and is manifest only during the sustained tachycardia. The presence of a concealed AP is suggested by the timing and pattern of atrial activation during the tachycardia: the P wave typically follows ventricular activation with a short RP wave interval (Fig. 233-7). Because many APs connect the left ventricle to the left atrium, the pattern of atrial activation during the tachycardia frequently produces negative P waves in leads I and aVL. The tachycardia circuit and therefore its ECG manifestation during orthodromic tachycardia are identical both in patients with overt preexcitation in sinus rhythm and in those with concealed APs. Patients with concealed APs, although prone to episodes of AF, are not at risk for developing a rapid ventricular response to the AF.

Occasionally, APs conduct extremely slowly in a retrograde fashion, resulting in longer retrograde conduction and the development of a long RP interval during the tachycardia (*long RP tachycardia*). Because of the presence of this dramatically slowed conduction, additional conduction slowing created by premature atrial complexes is not required for tachycardia to ensue. These patients are more prone to frequent episodes of tachycardia and can present with "incessant" tachycardias and tachycardia-induced LV cardiomyopathy. The correct diagnosis of a long RP tachycardia may be suggested by the pattern of initiation and the P-wave morphology. Frequently, however, an electrophysiologic evaluation is required to establish the diagnosis.

TREATMENT: ACCESSORY PATHWAY-MEDIATED TACHYCARDIAS

Acute treatment of AP-mediated macroreentrant orthodromic tachycardias is similar to that for AV nodal reentry and is directed at altering conduction in the AV node. Vagal stimulation with the Valsalva maneuver and carotid sinus pressure may create sufficient AV nodal slowing to terminate the AVRT. Intravenous administration of adenosine, 6–12 mg, is first-line pharmacologic therapy; IV, the calcium channel blockers verapamil and diltiazem or beta blockers may also be effective. In patients who manifest preexcitation and AF, therapy should be aimed at preventing a rapid ventricular response. In life-threatening situations, DC cardioversion should be used to terminate the AF. In nonlife-threatening situations, procainamide at a dose of 15 mg/kg administered IV over 20–30 min will slow the ventricular response and may organize and terminate AF. Ibutilide can also be used to facilitate termination of AF. During AF there may be rapid conduction over the AV node as well as the AP. Caution should be used in attempting to slow AV nodal conduction with the use of digoxin or verapamil; when administered IV, these drugs may actually result in an acute increase in rate over the AP, placing the patient at risk for development of VF. Digoxin appears to shorten the refractory period of the AP directly and thus increases the ventricular rate. Verapamil appears to shorten the refractory period indirectly by causing vasodilation and a reflex increase in sympathetic tone.

Chronic oral administration of beta blockers and/or verapamil or diltiazem may be used to prevent recurrent supraventricular reentrant tachycardias associated with APs. In patients with evidence of AF and a rapid ventricular response and in those with recurrences of SVT on AV nodal blocking drugs, strong consideration should be given to the administration of a class IA or IC antiarrhythmic drug such as quinidine, flecainide, or propafenone because these drugs slow conduction and increase refractoriness in the AP.

Patients with a history of recurrent symptomatic SVT episodes, incessant SVT, and heart rates >200 beats/min with SVT should be given strong consideration for undergoing catheter ablation. Patients who have demonstrated rapid antegrade conduction over their AP or the potential for rapid conduction should also be considered for catheter ablation. Catheter ablation therapy has been demonstrated to be successful in >95% of patients with documented WPW syndrome and appears effective regardless of age. The risk of catheter ablative therapy is low and is dictated primarily by the location of the AP. Ablation of parahisian APs is associated with a risk of heart block, and ablation in the left atrium is associated with a small but definite risk of thromboembolic phenomenon. These risks must be weighed against the potential serious complications associated with hemodynamic compromise, the risk of VF, and the severity of the patient's symptoms with AP-mediated tachycardias.

Patients who demonstrate evidence of ventricular preexcitation in the absence of any prior arrhythmia history merit special consideration. The first arrhythmia manifestation can be a rapid SVT or, albeit of low risk (<1%), a life-threatening rapid response to AF. Patients who demonstrate intermittent preexcitation during ECG monitoring or an abrupt loss of AP conduction during exercise testing are at low risk of a life-threatening rapid response to AF. All other patients should be advised of their risks and therapeutic options in advance of a documented arrhythmia event.

VENTRICULAR TACHYARRHYTHMIAS

Ventricular Premature Complexes

The origin of premature beats in the ventricle at sites remote from the Purkinje network produces slow ventricular activation and a wide QRS complex that is typically >140 ms in duration. Ventricular premature complexes are common and increase with age and the presence of structural heart disease. VPCs can occur with a certain degree of periodicity that has become incorporated into the lexicon of electrocardiography. Ventricular premature complexes may occur in patterns of *bigeminy*, in which every sinus beat is followed by a VPC, or *trigeminy*, in which two sinus beats are followed by a VPC. VPCs may have different morphologies and are thus referred to as *multiformed*. Two successive VPCs are termed *pairs* or *couplets*. Three or more consecutive VPCs are termed *VT* when the rate is >100 beats per minute. If the repetitive VPCs terminate spontaneously and are more than three beats in duration, the arrhythmia is referred to as *nonsustained VT*.

APCs with aberrant ventricular conduction may also create a wide and early QRS complex. The premature P wave can occasionally be difficult to discern when it falls on the preceding T wave, and other clues must be used to make the diagnosis. The QRS pattern for a VPC does not appear to follow a typical right or left bundle branch block pattern as the QRS morphology is associated with aberrant atrial conduction and can be quite bizarre. On occasion, VPCs can arise from the Purkinje network of the ventricles, in which case the QRS pattern mimics aberration. The 12-lead ECG recording of the

VPC may be required to identify subtle morphologic clues regarding the QRS complex to confirm its ventricular origin. Most commonly, VPCs are associated with a "fully compensatory pause" [i.e., the duration between the last QRS before the PVC and the next QRS complex is equal to twice the sinus rate (Fig. 233-3)]. The VPC typically does not conduct to the atrium. If the VPC does conduct to the atrium, it may not be sufficiently early to reset the sinus node. As a result, sinus activity will occur and the antegrade wavefront from the sinus node may encounter some delay in the AV node or His-Purkinje system from the blocked VPC wavefront, or it may collide with the retrograde atrial wavefront. Sinus activity will continue undisturbed, resulting in a delay to the next QRS complex (Fig. 233-3). Occasionally the VPC can occur early enough and conduct retrograde to the atrium to reset the sinus node; the pause that results will be less than compensatory. VPCs that fail to influence the oncoming sinus impulse are termed *interpolated VPCs*. A ventricular focus that fires repetitively at a fixed interval may produce variably coupled VPCs, depending on the sinus rate. This type of focus is referred to as a *parasystolicfocus* because its firing does not appear to be modulated by sinus activity and the conducted QRS complex. The ventricular ectopy will occur at a characteristic fixed integer or multiple of these intervals. The variability in coupling relative to the underlying QRS complex and a fixed interval between complexes of ventricular origin provide the diagnostic information necessary to identify a parasystolic focus.

TREATMENT: VENTRICULAR PREMATURE COMPLEXES

The threshold for treatment of VPCs is high, and the treatment is directed primarily at eliminating severe symptoms associated with palpitations. VPCs of sufficient frequency can cause a reversible cardiomyopathy. Depressed LV function in the setting of ventricular bigeminy and/or frequent nonsustained VT should raise the possibility of a cardiomyopathy that is reversible with control of the ventricular arrhythmia. In the absence of structural heart disease, VPCs do not appear to have prognostic significance. In patients with structural heart disease, frequent VPCs and runs of nonsustained VT have prognostic significance and may portend an increased risk of SCD. However, no study has documented that elimination of VPCs with antiarrhythmic drug therapy reduces the risk of arrhythmic death in patients with severe structural heart disease. In fact, drug therapies that slow myocardial conduction and/or enhance dispersion of refractoriness can actually increase the risk of life-threatening arrhythmias (drug-induced QT prolongation and TDP) despite being effective at eliminating VPCs.

Accelerated Idioventricular Rhythm (AIVR)

AIVR refers to a ventricular rhythm that is characterized by three or more complexes at a rate >40 beats per minute and <120 beats per minute. The arrhythmia mechanism causing AIVR is thought to be due to abnormal automaticity. By definition, there is an overlap between AIVR and "slow" VT; both rhythms can manifest rates between 90 and 120 beats per minute. Because AIVR tends to be a benign rhythm with different therapeutic implications, it is worthwhile to attempt to distinguish it from "slow" VT. AIVR has a characteristic gradual onset and offset and more variability in cycle length. It is typically a brief, self-limiting arrhythmia. AIVR can be seen in the absence of any structural heart disease, but it is frequently present in the setting of acute myocardial infarction (MI), cocaineintoxication, acute myocarditis, digoxin intoxication, and postoperative cardiac surgery. Sustained forms of AIVR can exist, particularly in the setting of acute MI and postoperatively. In the setting of sustained AIVR, hemodynamic compromise can occur because of the loss of AV synchrony. Patients with RV infarction associated with proximal right coronary artery occlusion are most susceptible to associated bradyarrhythmias and the hemodynamic consequences of AIVR. In these patients, acceleration of the atrial rate either by the cautious administration of atropine or by atrial pacing may be an important treatment consideration.

Ventricular Tachycardia

VT originates below the bundle of His at a rate >100 beats per minute; most VT patients have rates >120 beats per minute. Sustained VT at rates <120 beats per minute and even <100 beats per minute can be observed, particularly in association with the administration of antiarrhythmic agents that can slow the rate. Because of the overlap in rates with AIVR, the arrhythmia ECG characteristics and the clinical circumstance sometimes can be used to distinguish the two forms of tachycardia. Slow sustained VT is less likely to show a marked warm up in rate and the marked cycle-length oscillations seen with AIVR, and it is more likely to occur in the setting of chronic infarction or cardiomyopathy and less likely with acute infarction or myocarditis. Obviously, significant overlap may exist. Typically, slow VT will be initiated with programmed stimulation and is found to represent a large macroreentrant circuit in chronically diseased myocardium capable of supporting markedly slow conduction.

The QRS complex during VT may be uniform (monomorphic) or may vary from beat to beat (polymorphic). Polymorphic VT in patients who demonstrate a long QT interval during their baseline rhythm typically is referred to as *torsades des pointes*. The polymorphic VT associated with QT prolongation dramatically oscillates around the baseline on most of the monitored ECG leads, mimicking the "turning of the points" stitching pattern **(Fig. 233-9)**.

Figure 233-9

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus rhythm with long QT interval and the polymorphic ventricular arrhythmia torsades des pointes. Dramatic T wave alternans is present in sinus rhythm.

Monomorphic VT suggests a stable tachycardia focus in the absence of structural heart disease or a fixed anatomic abnormality that can create the substrate for a stable reentrant VT circuit when structural disease is present. Monomorphic VT tends to be a reproducible and recurrent phenomenon and may be initiated with pacing and programmed ventricular

stimulation. In contrast, polymorphic VT suggests a more dynamic and/or unstable process and, by its very nature, is less reproducible. Polymorphic VT may be produced by acute ischemia, myocarditis, or dynamic changes in the QT interval and enhanced dispersion of ventricular refractoriness. Polymorphic VTs are not reliably initiated with pacing or programmed stimulation.

A time duration of 30 seconds frequently is used to distinguish sustained from nonsustained VT. Hemodynamically unstable VT that requires termination before 30 seconds or VT that is terminated by therapy from an implantable defibrillator is also typically classified as sustained. Ventricular flutter appears as a sine wave on the ECG and has a rate of >250 beats per minute. A rapid rate coupled with the sine wave nature of the arrhythmia makes it impossible to identify a discrete QRS morphology. When antiarrhythmic drugs are being administered, a sine wave appearance of the QRS complex can be observed, even at rates as low as 200 beats per minute. VF is characterized by completely disorganized ventricular activation on the surface ECG. Polymorphic ventricular arrhythmias, ventricular flutter, and VF always produce hemodynamic collapse if allowed to continue. The hemodynamic stability of a unimorphic VT depends on the presence and severity of the underlying structural heart disease, the location of the site of origin of the arrhythmia, and the heart rate.

It is important to distinguish monomorphic VT from SVT with aberrant ventricular conduction due to right or left bundle branch block.

Importantly, the sinus or baseline 12-lead ECG tracing can provide important clues that help establish the correct diagnosis of a wide complex tachycardia. The presence of an aberrant QRS pattern that matches exactly that of the wide complex rhythm strongly supports the diagnosis of SVT. A right or left bundle branch block QRS pattern that does not match the QRS and/or that is wider in duration than the QRS during the wide complex tachycardia supports the diagnosis of VT. Most patients with VT have structural heart disease and show evidence of a prior Q wave MI during sinus rhythm. Important exceptions to this rule are discussed (pp. 35-36). Finally, the presence of a preexcited QRS pattern on the 12-lead ECG in sinus rhythm suggests that the wide complex rhythm represents an atrial arrhythmia, such as AFL or a focal AT, with rapid conduction over an AP or antidromic macroreentrant tachycardia (Fig. 233-8). If the arrhythmia is irregular with changing QRS complexes, the diagnosis of AF with ventricular preexcitation should be considered.

With the exception of some idiopathic outflow tract tachycardias, most VTs do not respond to vagal stimulation provoked by carotid sinus massage, the Valsalva maneuver, or adenosine administration. The IV administration of verapamil and/or adenosine is not recommended as a diagnostic test. Verapamil has been associated with hemodynamic collapse when administered to patients with structural heart disease and VT.

Patients with VT frequently demonstrate AV disassociation. Findings on physical examination of intermittent cannon *a* waves and variability of the first heart sound are consistent with AV dissociation. The presence of AV dissociation is characteristically marked by the presence of sinus capture or fusion beats. The presence of 1:1 ventriculoatrial conduction does not preclude a diagnosis of VT.

Additional characteristics of the 12-lead ECG during the tachycardia that suggest VT include (1) the presence of a QRS duration >140 ms in the absence of drug therapy, (2) a superior and rightward QRS frontal plane axis, (3) a bizarre QRS complex that does not mimic the characteristic QRS pattern associated with left or right bundle branch block, and (4) slurring of the initial portion of the QRS **(Fig. 233-10)**. **Table 233-6** provides a useful summary of ECG criteria that have evolved based on the described characteristics of VT.

Figure 233-10

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ventricular tachycardia. ECG showing AV dissociation (arrows mark P waves), wide QRS >200 ms, superior frontal plane axis, slurring of the initial portion of the QRS, and large S wave in V_6—all clues to the diagnosis of ventricular tachycardia.

Table 233–6. ECG Clues Supporting the Diagnosis of Ventricular Tachycardia

AV dissociation (atrial capture, fusion beats)
QRS duration >140 ms for RBBB type V_1 morphology; V_1>160 ms for LBBB type V_1 morphology
Frontal plane axis –90° to 180°
Delayed activation during initial phase of the QRS complex
LBBB pattern—R wave in V_1, V_2>40 ms
RBBB pattern—onset of R wave to nadir of S >100 ms
Bizarre QRS pattern that does not mimic typical RBBB or LBBB QRS complex
Concordance of QRS complex in all precordial leads
RS or dominant S in V_6 for RBBB VT

Q wave in V_6 with LBBB QRS pattern
Monophasic R or biphasic qR or R/S in V_1 with RBBB pattern

Abbreviations: AV, atrioventricular; RBBB/LBBB, right/left bundle branch block.

TREATMENT: VENTRICULAR TACHYCARDIA/FIBRILLATION

Sustained polymorphic VT, ventricular flutter, and VF all lead to immediate hemodynamic collapse. Emergency asynchronous defibrillation is therefore required, with at least 200-J monophasic or 100-J biphasic shock. The shock should be delivered asynchronously to avoid delays related to sensing of the QRS complex. If the arrhythmia persists, repeated shocks with the maximum energy output of the defibrillator are essential to optimize the chance of successful resuscitation. Intravenous lidocaine and/or amiodarone should be administered but should not delay repeated attempts at defibrillation.

For any monomorphic wide complex rhythm that results in hemodynamic compromise, a prompt R-wave synchronous shock is required. Conscious sedation should be provided if the hemodynamic status permits. For patients with a well-tolerated wide complex tachycardia, the appropriate diagnosis should be established on the basis of strict ECG criteria (Table 233-6). Pharmacologic treatment to terminate monomorphic VT is not typically successful (<30%). Intravenous procainamide, lidocaine, or amiodarone can be utilized. If the arrhythmia persists, synchronous R-wave cardioversion after the administration of conscious sedation is appropriate. Selected patients with focal outflow tract tachycardias (p. 35) who demonstrate triggered or automatic VT may respond to IV beta blocker administration. Idiopathic LV septal VT (see p. 35) appears to respond uniquely to IV verapamil administration.

VT in patients with structural heart disease is now almost always treated with the implantation of an ICD to manage anticipated VT recurrence. The ICD can provide rapid pacing and shock therapy to treat most VTs effectively **(Fig. 233-11)**.

Figure 233-11

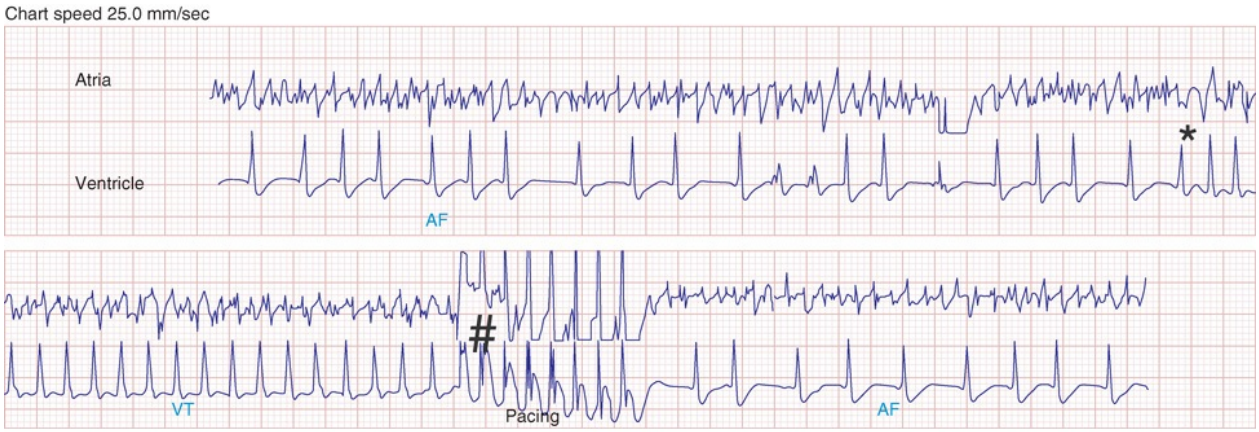

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ventricular tachycardia (VT) (*) during atrial fibrillation stopped by pacing (#) from an implantable cardioverter defibrillator (ICD) from recording stored by ICD. The atrial electrogram shows characteristic fibrillatory waves through the tracing. The ventricular electrogram shows an irregularly irregular response consistent with atrial fibrillation at the beginning of the tracing. The ventricular electrogram suddenly changes in morphology (*) and becomes regular, consistent with the diagnosis of VT. Pacing transiently accelerates the rate and interrupts the rapid VT. The patient was unaware of the life-threatening event.

Prevention of VT remains important, and >50% of patients with a history of VT and an ICD may need to be treated with adjunctive antiarrhythmic drug therapy to prevent VT recurrences or to manage atrial arrhythmias. Because of the presence of an ICD, there is more flexibility with respect to the selection of antiarrhythmic drug therapy. The use of sotalol or amiodarone represents first-line therapy for patients with a history of structural heart disease and life-threatening monomorphic or polymorphic VT not due to long QT syndrome. Importantly, sotalol has been associated with a decrease in the defibrillation threshold, which reflects the amount of energy necessary to terminate VF. Amiodarone may be better tolerated in patients with a more marginal hemodynamic status and systolic blood pressure. The risk of end-organ toxicity from amiodarone must be weighed against the ease of use and general efficacy. Antiarrhythmic drug therapy with agents such as quinidine, procainamide, and propafenone, which might not normally be used in patients with structural heart disease because of the risk of proarrhythmia, may be considered in patients with an ICD and recurrent VT.

Catheter ablative therapy for VT in patients without structural heart disease results in cure rates >90%. In patients with structural heart disease, catheter ablation that includes a strategy for eliminating unmappable/rapid VT and one that incorporates endocardial as well as epicardial mapping and ablation should be employed. In most patients, catheter ablation can reduce or eliminate the requirement for toxic drug therapy and should be considered in any patient with recurrent VT. The utilization of ablative therapy to reduce the incidence of ICD shocks for VT in patients who receive the ICD as part of primary prevention for VT is being actively investigated.

MANAGEMENT OF VT STORM

Repeated VT episodes requiring external cardioversion/defibrillation or repeated appropriate ICD shock therapy are referred to as *VT storm*. Although a definition of more than two episodes in 24 h is used, most patients with VT storm will experience many more episodes. In the extreme form of VT storm, the tachycardia becomes incessant and the baseline rhythm cannot be restored for any extended period. In patients with recurrent polymorphic VT in the absence of the long QT interval, one should have a high suspicion of active ischemic disease or fulminant myocarditis. Intravenous lidocaine or amiodarone administration should be coupled with prompt assessment of the status of the coronary anatomy. Endomyocardial biopsy, if indicated by clinical circumstances, may be used to confirm the diagnosis of myocarditis,

although the diagnostic yield is low. In patients who demonstrate QT prolongation and recurrent pause-dependent polymorphic VT (TDP), removal of an offending QT-prolonging drug, correction of potassium or magnesium deficiencies, and emergency pacing to prevent pauses should be considered. Intravenous beta blockade therapy should be considered for polymorphic VT storm. A targeted treatment strategy should be employed if the diagnosis of the polymorphic VT syndrome can be established. For example, quinidine or isoproterenol can be used in the treatment of Brugada syndrome (p. 40). Intraaortic balloon counterpulsation or acute coronary angioplasty may be needed to stop recurrent polymorphic VT precipitated by acute ischemia. In selected patients with a repeating VPC trigger for their polymorphic VT, the VPC can be targeted for ablation to prevent recurrent VT.

In patients with recurrent monomorphic VT, acute IV administration of lidocaine, procainamide, or amiodarone can prevent recurrences. The use of such therapy is empirical, and a clinical response is not certain. Procainamide and amiodarone are more likely to slow the tachycardia and make it hemodynamically tolerated. Unfortunately, antiarrhythmic drugs, especially those that slow conduction (e.g., amiodarone, procainamide), can also facilitate recurrent VT or even result in incessant VT. VT catheter ablation can eliminate frequent recurrent or incessant VT and frequent ICD shocks. Such therapy should be deployed earlier in the course of arrhythmia events to prevent adverse consequences of recurrent VT episodes and adverse effects from antiarrhythmic drugs.

Unique VT Syndromes

Although most ventricular arrhythmias occur in the setting of CAD with prior MI, a significant number of patients develop VT in other settings. A brief discussion of each unique VT syndrome is warranted. Information that illustrates a unique pathogenesis and enhances the ability to make the correct diagnosis and institute appropriate therapy will be highlighted.

IDIOPATHIC OUTFLOW TRACT VT

VT in the absence of structural heart disease is referred to as *idiopathic VT*. There are two major varieties of these VTs. Outflow tachycardias originate in the RV and LV outflow tract regions. Approximately 80% of outflow tract VTs originate in the RV and ~20% in the LV outflow tract regions. Outflow tract VTs appear to originate from anatomic sites that form an arc that begins just above the tricuspid valve and extends along the roof of the outflow tract region to include the free wall and septal aspect of the right ventricle just beneath the pulmonic valve, the aortic valve region, and then the anterior/superior margin of the mitral valve annulus. These arrhythmias appear more commonly in women. Importantly, these ventricular arrhythmias are *rarely* associated with SCD unless manifest by very short coupled premature complexes that trigger VF. Patients manifest symptoms of palpitations with exercise, stress, and caffeine ingestion. In women, the arrhythmia is more commonly associated with hormonal triggers and can frequently be timed to the premenstrual period, gestation, and menopause. Uncommonly, the VPCs and VTs can be of sufficient frequency and duration to cause a tachycardia-induced cardiomyopathy.

The pathogenesis of outflow tract VT remains unknown, and there is no definite anatomic abnormality identified with these VTs. Vagal maneuvers, adenosine, and beta blockers tend to terminate the VTs, whereas catecholamine infusion, exercise, and stress tend to potentiate the outflow tract VTs. Based on these observations, the mechanism of the arrhythmia is most likely calcium-dependent triggered activity. Preliminary data suggest that at least in some patients, a somatic mutation of the inhibitory G protein ($G\alpha_{i2}$) may serve as the genetic basis for the VT. In contrast to VT in patients with CAD, outflow tract VTs are uncommonly initiated with programmed stimulation but are able to be initiated by rapid-burst atrial or ventricular pacing, particularly when coupled with the infusion of isoproterenol.

Outflow tract VT typically produces large monophasic R waves in the inferior frontal plane leads II, III, and aVF and typically occurs as nonsustained bursts of VT and/or frequent premature beats. Cycle length oscillations during the tachycardia are common. Since most VT originates in the RV outflow tract, the VT typically has a left bundle branch block (LBBB) pattern in lead V_1 (negative QRS vector) **(Fig. 233-12)**. Outflow tract VTs, originating in the left ventricle, particularly those associated with an origin from the mitral valve annulus, have a right bundle branch block (RBBB) pattern in lead V_1 (positive QRS vector).

Figure 233-12

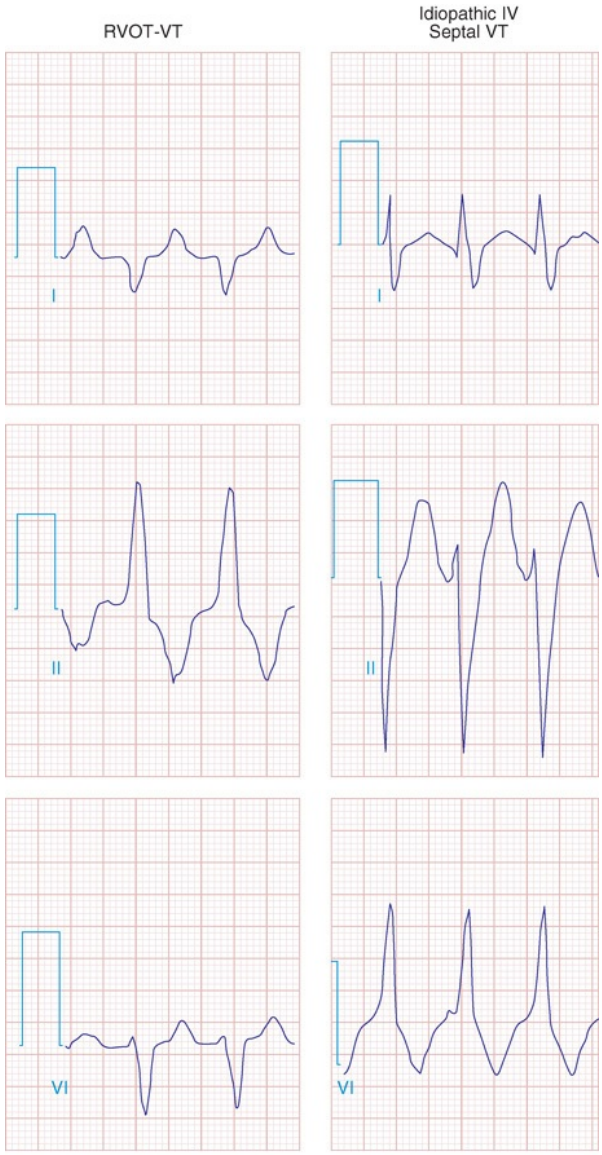

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Common idiopathic ventricular tachycardia (VT) ECG patterns. Right ventricular outflow tract (RVOT) VT with typical left bundle QRS pattern in V_1 and inferiorly directed frontal plane axis, and left ventricular septal VT from the inferior septum with a narrow QRS RBBB pattern in V_1 and superior and leftward front plane QRS axis.

Treatment: Idiopathic Outflow Tract Ventricular Tachycardia

Acute medical therapy for idiopathic outflow tract VT is rarely required because the VT is hemodynamically tolerated and is typically nonsustained. Intravenous beta blockers frequently terminate the tachycardia. Chronic therapy with beta or calcium channel blockers frequently prevents recurrent episodes of the tachycardia. The arrhythmia also appears to respond to treatment with class IA or IC agents or with sotalol. Catheter ablative therapy has been utilized successfully to eliminate the tachycardia with success rates >90%. Because of the absence of structural heart disease and the focal nature of these arrhythmias, the 12-lead ECG pattern during VT can help localize the site of origin of the arrhythmia and help facilitate catheter ablation. Efficacy of therapy is assessed with treadmill testing and/or ECG monitoring, and electrophysiologic study is performed only when the diagnosis is in question or to perform catheter ablation.

IDIOPATHIC LV SEPTAL/FASCICULAR VT

The second most common idiopathic VT is linked anatomically to the Purkinje system in the left ventricle. The arrhythmia mechanism appears to be macroreentry involving calcium-dependent slow response fibers that are part of the Purkinje network, although automatic tachycardias have also been observed. A 12-lead ECG morphology of the VT shows a narrow RBBB pattern and a superior leftward axis or an inferior rightward axis, depending on whether the VT originates from the posterior or anterior fascicles (Fig. 233-12). Idiopathic LV septal VT is unique in its suppression with verapamil. Beta blockers also have been used with some success as primary or effective adjunctive therapy. Catheter ablation is very effective therapy for VT resistant to drug therapy or in patients reluctant to take daily therapy, with anticipated successful elimination of VT in >90% of patients.

VT ASSOCIATED WITH LV DILATED CARDIOMYOPATHY

Monomorphic and polymorphic VTs may occur in patients with nonischemic dilated cardiomyopathy (Chap. 238). Although the myopathic process may be diffuse, there appears to be a predilection for the development of fibrosis around the mitral and aortic valvular regions. Most uniform sustained VT can be mapped to these regions of fibrosis. Drug therapy is usually ineffective in preventing VT, and empirical trials of sotalol or amiodarone are usually initiated only for recurrent VT episodes after ICD implantation. VT associated with nonischemic dilated cardiomyopathy appears to be less amenable to catheter

ablative therapy from the endocardium; frequently, the VT originates from epicardial areas of fibrosis and catheter access to the epicardium can be gained via a percutaneous pericardial puncture to improve the outcome of ablation techniques. In patients with a history of depressed myocardial dysfunction due to a nonischemic cardiomyopathy with an LV ejection fraction <30%, data now support the implantation of a prophylactic ICD device to reduce the risk of SCD from the first VT/VF episode effectively.

BUNDLE BRANCH REENTRANT VT

Monomorphic VT in patients with idiopathic nonischemic cardiomyopathy or valvular cardiomyopathy is frequently due to a large macroreentrant circuit involving the various elements of the His-Purkinje network. The arrhythmia usually occurs in the presence of underlying disease of the His-Purkinje system. In sinus rhythm, an incomplete left bundle block is typically present and the time it takes to traverse the His-Purkinje network is delayed; this slow conduction serves as the substrate for reentry. Characteristically, the VT circuit rotates in an antegrade direction down the right bundle and retrograde up the left posterior or anterior fascicles and left bundle branch. As a result, bundle branch reentrant VT typically has a QRS morphology with a left bundle branch block type of pattern and a leftward superior axis **(Fig. 233-13)**. The circuit for bundle branch (LBB) reentrant VT can occasionally rotate in the opposite direction, antegrade through the left bundle and retrograde through the right bundle, in which case a RBBB pattern during VT will be manifest.

Figure 233-13

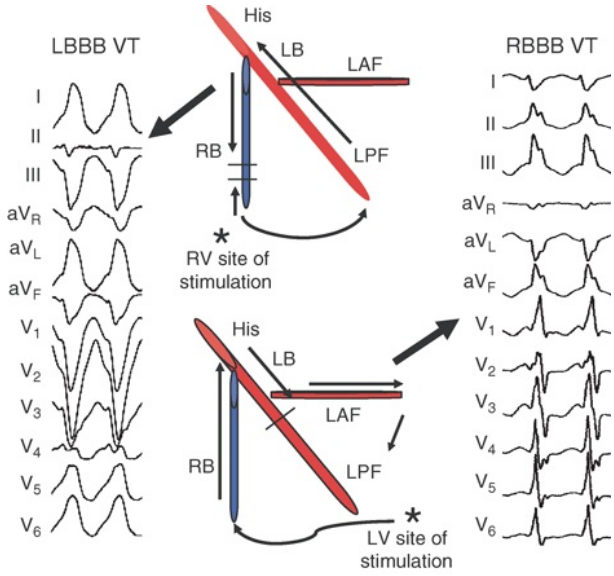

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bundle branch reentrant ventricular tachycardia (VT) showing typical QRS morphologies when VT is initiated with stimulation from the right ventricle [left bundle branch block (LBBB) VT pattern] or left ventricle [right bundle branch block (RBBB) VT pattern] and schema for circuit involving the His-Purkinje network.

It is important to recognize bundle branch reentrant VT because it is readily amenable to ablative therapy that targets a component of the His-Purkinje system, typically the right bundle, to block the VT circuit. Less commonly, bundle branch reentry may occur in the absence of structural heart disease or in the setting of CAD. The use of adjunctive ICD therapy is dictated by the ability to eliminate the VT successfully and the severity of the LV dysfunction.

VT ASSOCIATED WITH HYPERTROPHIC CARDIOMYOPATHY

(See also Chap. 238) VT and VF have also been associated with hypertrophic cardiomyopathy. In patients with hypertrophic cardiomyopathy and a history of sustained VT/VF, unexplained syncope, a strong family history of SCD, LV septal thickness >30 mm, or nonsustained spontaneous VT, the risk of SCD is high and ICD implantation is usually indicated. Amiodarone, sotalol, and beta blockers have been used to control recurrent VT. Experience with ablative therapy is limited because of the infrequency with which the VT is tolerated hemodynamically. Ablation procedures that target the substrate for VT/VF and ablate areas of low voltage consistent with fibrosis, which frequently are located in apical aneurysms, appear to have promise in this setting. The WPW syndrome has been observed in patients with hypertrophic cardiomyopathy associated with *PRKAG2* mutations.

VT ASSOCIATED WITH OTHER INFILTRATIVE CARDIOMYOPATHIES AND NEUROMUSCULAR DISORDERS

An increased arrhythmia risk has been identified when cardiac involvement occurs in a variety of infiltrative diseases and neuromuscular disorders **(Table 233-7)**. Many patients manifest AV conduction disturbances and may require permanent pacemaker insertion. The decision to implant an ICD device should follow current established guidelines for patients with nonischemic cardiomyopathy, which include an LV ejection fraction < 35% or a history of unexplained syncope with significant LV dysfunction. A recent report identified AF, PR interval >240 ms, QRS 120 ms, or heart block and type 1 myotonic dystrophy as predicting a risk of sudden death. Additional study will be required to determine if patients with lesser degrees of LV dysfunction or other more diffuse myopathic disease processes also have identifiable risk and warrant primary ICD implantation. A potential proarrhythmic risk of antiarrhythmic drug therapy should be acknowledged, and drug therapy should be reserved for symptomatic arrhythmias and limited to amiodarone or sotalol if an ICD is not present.

Table 233–7. Infiltrative/Inflammatory and Neuromuscular Disorders Associated with an Increased Ventricular Arrhythmia Risk

Sarcoidosis[a]	Emery-Dreyfuss muscular dystrophy[a]

Chagas' disease[a]	Limb-girdle muscular dystrophy[a]
Amyloidosis[a]	Duchenne muscular dystrophy
Fabry disease	Becker muscular dystrophy
Hemochromatosis	Kearn-Sayre syndrome[a]
Myotonic muscular dystrophy[a]	Friedreich's ataxia

[a] High frequency of ventricular arrhythmias noted.

ARRHYTHMOGENIC RV CARDIOMYOPATHY/DYSPLASIA (ARVCM/D)

(See also Chap. 238) ARVCM/D due to a genetically determined dysplastic process or after a suspected viral myocarditis is also associated with VT/VF. The sporadic nonfamilial/nondysplastic form of RV cardiomyopathy appears to be more common; however, this may vary with ethnicity. In patients predisposed to VT, there appears to be a predominance of perivalvular fibrosis involving mostly the free wall of the right ventricle in proximity to the tricuspid and pulmonic valves. The surface ECG leads that reflect RV activation, including V_1–V_3, may show terminal notching of the QRS complex and inverted T waves in sinus rhythm. When the terminal notching is distinct and appears separated from the QRS complex, it is referred to as an *epsilon wave* **(Fig. 233-14)**. Epsilon waves are consistent with markedly delayed ventricular activation in the region of the RV free wall near the base of the tricuspid and pulmonic valves in areas of extensive fibrosis.

Figure 233-14

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

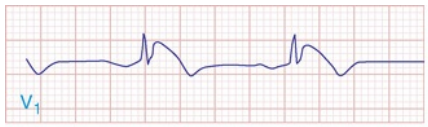

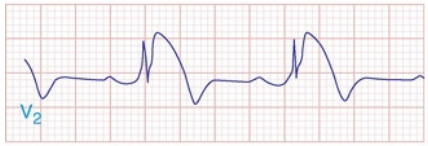

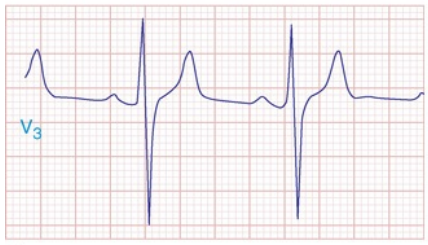

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Leads V_1 to V_3 in sinus rhythm from a normal subject ***A.***, from a patient with arrhythmogenic right ventricular cardiomyopathy showing epsilon waves (*arrow*) and T-wave inversion ***B.***, and from a patient with Brugada syndrome with ST-segment elevation in V_1 and V_2 ***C.***

In patients with ARVCM/D, echocardiography demonstrates RV enlargement with RV wall motion abnormalities and RV apical aneurysm formation. MRI may show fatty replacement of the ventricle, thinning of the RV free wall with increased fibrosis, and associated wall motion abnormalities. Because of the presence of extensive amounts of fat normally covering the epicardium in the region of the RV, caution must be used to avoid overinterpreting the MRI in trying to determine the appropriate diagnosis. Patients tend to have multiple VT morphologies. The VT will typically have a LBBB type QRS pattern in V_1 and tend to have poor R-wave progression in V_1 through V_6, consistent with an RV free-wall origin. Areas of low electrogram voltage that are identified during RV catheter endocardial sinus rhythm voltage mapping may be helpful in confirming the diagnosis. Importantly, endocardial biopsy may not identify the presence of fatty replacement or fibrosis unless directed to the basal RV free wall. The familial forms of this syndrome have been linked to a number of desmosomal protein mutations. A distinct genetic form of this syndrome, Naxos disease, consists of arrhythmogenic RV dysplasia coupled with palmar-plantar keratosis and woolly hair and is associated with a high risk of SCD in adolescents and young adults.

Treatment: Arrhythmogenic Right Ventricular Cardiomyopathy/Dysplasia

The threshold for ICD implantation in patients with an established diagnosis of ARVCM/D is low. An ICD typically is implanted in patients deemed to have a persistent VT risk, those who have had spontaneous or inducible rapid VTs, and those who show concomitant LV cardiomyopathy. Treatment options for recurrent VT in patients with ARVCM/D include the use of the antiarrhythmic agent sotalol. Beta blockers serve as useful adjunctive therapy when coupled with other antiarrhythmic agents. Catheter ablative therapy directed at mappable sustained ventricular arrhythmias is also highly successful in controlling recurrent VT. In selected patients with multiple VT morphologies and unstable VT, linear ablation lesions directed at endocardial scars and, if required, targeting late potentials in epicardial scars, defined by catheter-based bipolar voltage mapping, provide significant amelioration of the recurrent VT episodes.

VT AFTER OPERATIVE TETRALOGY OF FALLOT REPAIR

VT may also occur after surgical repair of tetralogy of Fallot. Patients typically develop VT many years after the surgery. VT tends to occur in patients with evidence of RV systolic dysfunction. The VT mechanism and location are typically a macroreentrant circuit around the right ventriculotomy scar to the valve annuli. Catheter ablation creating linear lesions that extend from either the pulmonic or the tricuspid annulus to the ventriculotomy scar is typically effective in preventing arrhythmia recurrences. An ICD is usually implanted in patients who manifest rapid VT, have persistent inducible VT after ablation, or have concomitant LV dysfunction.

FASCICULAR TACHYCARDIA CAUSED BY DIGOXIN TOXICITY

Digoxin toxicity can produce increased ventricular ectopy and, when coupled with bradyarrhythmias caused by digoxin toxicity, may predispose to sustained polymorphic ventricular arrhythmias and VF. The signature VT associated with digoxin toxicity is bidirectional VT **(Fig. 233-15)**. This unique VT is due to triggered activity associated with calcium overload resulting from the inhibition of Na^+,K^+-ATPase by digoxin. Bidirectional VT originates from the left anterior and posterior fascicles, creating a relatively narrow QRS right bundle branch (RBB) configuration with a beat-to-beat alternating right and left frontal plane QRS axis. This VT seldom is observed in the absence of digoxin toxicity. Treatment for bidirectional VT or other hemodynamically significant arrhythmias due to digoxin excess includes correction of electrolyte disorders and IV infusion of digoxin-specific Fab fragments. The antibody fragments will, over the course of 1 hour, bind digoxin and eliminate toxic effects. In the setting of normal renal function, the bound complex is secreted.

Figure 233-15

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Digoxin toxic bidirectional fascicular tachycardia.

Genetically Determined Abnormalities that Predispose to Polymorphic Ventricular Arrhythmias

Ion channel defects that affect cardiac depolarization and repolarization may predispose to life-threatening polymorphic VT and SCD. These defects frequently produce unique ECG characteristics during sinus rhythm that facilitate the diagnosis.

LONG QT SYNDROME

The congenital form of LQTS consists of defects in cardiac ion channels that are responsible for cardiac repolarization. Defects that enhance sodium or calcium inward currents or inhibit outward potassium currents during the plateau phase of the action potential lengthen action potential duration and, hence, the QT interval. Of the eight genetic mutations identified to date, five affect the α or β subunits of the three different potassium channels involved with repolarization **(Table 233-8)**. Since many patients with QT prolongation do not have one of the defined mutations, it is anticipated that other genetic abnormalities affecting repolarization channel function will be identified.

Table 233–8. Inherited Arrhythmia Disorders: "Channelopathies" with High Risk of Ventricular Arrhythmias

Disorder	Gene	Protein/Channel Affected
LQT1	*KCNQ1*	I_{Ks} channel α subunit
LQT2	*KCNH2 (HERG)*	I_{Kr} channel α subunit
LQT3	*SCN5A*	I_{NA} channel α subunit
LQT4	*ANK2*	Ankyrin-B
LQT5	*KCNE1*	I_{Ks} channel β subunit
LQT6	*KCNE2*	I_{Kr} channel β subunit

LQT7	*KCNJ2*	I_{K1} channel α subunit
LQT8	*CACNA1C*	I_{Ca} channel α subunit
Jervell LN1	*KCNQ1*	I_{Ks} channel ß subunit
Jervell LN2	*KCNE1*	I_{Kr} channel ß subunit
Brugada syndrome	*SCN5A*	I_{Na} channel
Catecholaminergic VT	*Ry R2*	Ryanodine receptor, calsequestrin receptor
SQTS1	*KCNH2 (HERG)*	I_{Kr} channel α subunit
SQTS2	*KCNQ1(KvLQT1)*	I_{Ks} channel α subunit
SQTS3	*KCNJ2*	I_{K1} channel

Abbreviations: LQT, long QT (interval); SQT, short QT (interval).

The triggers for the ventricular arrhythmias are thought to be due to early afterdepolarizations potentiated by intracellular calcium accumulation from a prolonged action potential plateau. Heterogeneity of myocardial repolarization indexed by a longer QT interval predisposes to polymorphic ventricular arrhythmias in response to the triggers (Fig. 233-9).

In most patients with LQTS, the QT interval corrected for heart rate using Bazett's formula is >460 ms in men and >480 ms in women with LQTS. Marked lengthening of the QT interval to >500 ms is clearly associated with a greater arrhythmia risk in patients with LQTS. Many affected individuals may have QT intervals that intermittently measure within a normal range or fail to shorten appropriately with exercise. Some individuals manifest the syndrome only when exposed to a drug, such as sotalol, that alters channel function.

The genotype associated with LQTS appears to influence prognosis, and identification of the genotype appears to help optimize clinical management. The first three genotypic designations of the mutations identified, LQT1, LQT2, and LQT3, appear to account for >99% of patients with clinically relevant LQTSs. Surface ECG characteristics may be helpful in distinguishing the three most common genotypes, with genetic testing being definitive.

LQT1 represents the most common genotypic abnormality. Patients with LQT1 fail to shorten or actually prolong their QT interval with exercise. The T wave in patients with LQT1 tends to be broad and constitutes the majority of the prolonged QT interval. The most common trigger for potentiating cardiac arrhythmias in patients with LQT1 is exercise, followed by emotional stress.

More than 80% of male patients have their first cardiac event by age 20, so competitive exercise should be restricted and swimming avoided for these patients. Patients tend to respond to beta blocker therapy. Patients with two LQT1 alleles have the Jervell and Lange-Nielsen syndrome, with more dramatic QT prolongation and deafness and a worse arrhythmia prognosis.

LQT2 is the second most common genotypic abnormality. The T wave tends to be notched and bifid. In LQT2 patients, the most common precipitant is emotional stress, followed by sleep or auditory stimulation. Despite the occurrence during sleep, patients typically respond to beta blocker therapy.

LQT3 is due to a mutation in the gene that encodes the cardiac sodium channel on chromosome 3. Prolongation of the action potential duration occurs because of failure to inactivate this channel. LQT3 patients have either late-onset peaked biphasic T waves or asymmetric peaked T waves. The arrhythmia events tend to be more life-threatening, and thus the prognosis for LQT3 is the poorest of all the LQTs. Male patients appear to have the worst prognosis among patients with LQT3. Most events in LQT3 patients occur during sleep, suggesting that they are at higher risk during periods of slow heart rates. Beta blockers are not recommended, and exercise is not restricted in LQT3.

Treatment: Long QT Syndrome

The institution of ICD therapy should be strongly considered in any patient with LQTS who has demonstrated any life-threatening arrhythmia. Patients with syncope with a confirmed diagnosis based on unequivocal ECG criteria or positive genetic testing should also be given the same strong consideration. Primary prevention with prophylactic ICD implantation should be considered in male patients with LQT3 and in all patients with marked QT prolongation (>500 ms), particularly when coupled with an immediate family history of SCD. Future epidemiologic investigation may provide firmer guidelines to sort patients further on the basis of risks such as age, gender, arrhythmia history, and genetic characteristics. In all patients with documented or suspected LQTS, drugs that prolong the QT interval must be avoided. For an updated list of drugs, see *www.qtdrugs.org*.

ACQUIRED LQTS

Patients with a genetic predisposition related to what appear to be sporadic mutations and/or single nucleotide polymorphisms can develop marked QT prolongation in response to drugs that alter repolarization currents. The QT prolongation and associated polymorphic ventricular tachycardia (TDP) are seen more frequently in women and may be a manifestation of subclinical LQTS. Drug-induced long QT and TDP frequently are potentiated by the development of hypokalemia and bradycardia. The offending drugs typically block the potassium I_{Kr} channel (Table 233-5). Since most drug effects are dose dependent, important drug-drug interactions that alter metabolism and/or alterations in elimination kinetics because of hepatic or renal dysfunction frequently contribute to the arrhythmias.

Treatment: Acquired Long QT Syndrome

Acute therapy for acquired LQTS is directed at eliminating the offending drug therapy, reversing metabolic abnormalities by the infusion of magnesium and/or potassium, and preventing pause-dependent arrhythmias by temporary pacing or the cautious infusion of isoproterenol. Class IB antiarrhythmic agents (e.g., lidocaine) that do not cause QT prolongation may also be used, though they are frequently ineffective. Supportive therapy to allay anxiety and prevent pain with required DC shock therapy for sustained arrhythmias and efforts to facilitate drug elimination are important.

SHORT QT SYNDROME

A gain in function of repolarization currents can result in a shortening of atrial and ventricular refractoriness and marked QT shortening on the surface ECG (Table 233-8). The T wave tends to be tall and peaked. A QT interval <320 ms is required to establish the diagnosis of this uncommon syndrome. Mutations in the *HERG*, *KvLQT1*, and *KCNJ2* genes have been identified. Patients with the syndrome are predisposed to both AF and VF. ICD implantation is recommended. Double counting of QRS and T waves may lead to inappropriate ICD shocks. Drug therapy with quinidine has been used to lengthen the QT interval and reduce the amplitude of the T wave. This therapy is being evaluated to determine long-term efficacy in preventing arrhythmias in this syndrome.

BRUGADA SYNDROME

The major clinical features of Brugada syndrome include manifest, transient, or concealed ST segment elevation in V_1 to V_3 that typically can be provoked with the sodium channel-blocking drugs ajmaline, flecainide, and procainamide and a risk of polymorphic ventricular arrhythmias. It appears that a diminished inward sodium current in the region of the RV outflow tract epicardium is responsible for Brugada syndrome (Table 233-8). A loss of the action potential dome in the RV epicardium due to unopposed I_{TO} potassium outward current results in dramatic shortening of the action potential. The large potential difference between the normal endocardium and rapidly depolarized RV outflow epicardium gives rise to ST-segment elevation in V_1–V_3 in sinus rhythm and predisposes to local ventricular reentry (Fig. 233-14). The majority of genetic abnormalities responsible for the syndrome have not been described; however, in ~20% of patients, mutations of *SCN5A* genes have been identified. Although identified in both genders and all races with an autosomal dominant inheritance pattern, the arrhythmia syndrome is most common in young male patients (~75%) and is thought to be responsible for the sudden and unexpected nocturnal death syndrome (SUDS) described in Southeast Asian men. The ventricular arrhythmia characteristically occurs with rest or during sleep. Fever and other sodium channel-blocking drugs have also precipitated ventricular arrhythmias.

The presence of spontaneous coved-type ST elevation in the right precordial leads and a history of syncope or aborted sudden cardiac death are predictors of an adverse outcome. Because of the overlap in *SCN5A* mutations, the association of Brugada syndrome with phenotypic LQT3 and conduction disturbances has been noted.

Treatment: Brugada Syndrome

A drug challenge with procainamide may be important to establish the diagnosis and the probable cause of unexplained syncope when the surface ECG is equivocal (saddleback ST elevation pattern). Ajmaline and intravenous flecainide, which are not available in the United States, may have higher sensitivities for identifying the syndrome. Successful acute management of recurrent VT has been reported with isoproterenol or quinidine administration, although experience has been limited. Patients who do not benefit from beta blockers and chronic suppression with quinidine, which may lengthen epicardial action potential duration by blocking I_{TO} current, may be considered for ICD implantation. ICD treatment to manage recurrences and prevent sudden death is recommended for all patients who have had documented arrhythmia episodes and patients with syncope and positive spontaneous or provoked coved-type ECG ST-segment changes in V_1–V_3. Family members should undergo ECG screening for the presence of the abnormality. The role of programmed cardiac stimulation and the use of ICD therapy in asymptomatic patients with the Brugada-type ECG pattern remain somewhat controversial, as is provocative drug infusion and programmed stimulation in patients with borderline abnormalities and no arrhythmia symptoms. Longer-term follow-up in a larger group of these relatively low-risk patients may be required before definitive recommendations can be provided. Counseling on controversies that exist, the potential risk of fever, and inadvertent administration of tricyclic antidepressants should be considered. Genetic testing may be helpful in confirming the presence of the genetic abnormality in family members of patients who manifest the arrhythmia syndrome.

CATECHOLAMINERGIC POLYMORPHIC VT

A mutation of the myocardial ryanodine release channel, which effectively creates a "leak" in calcium from the sarcoplasmic reticulum, has been identified in patients with catecholaminergic VT (Table 233-8). The accumulation of intracellular calcium potentiates delayed afterdepolarizations and triggered activity. Patients can manifest bidirectional VT, nonsustained polymorphic VT, or recurrent VF. Both an autosomal dominant familial form and sporadic forms of the disease have been described. More recently, an autosomal recessive variant associated with a mutation in the sarcoplasmic reticulum calcium-buffering protein calsequestrin has also been identified. The arrhythmias are precipitated by exercise and emotional stress **(Fig. 233-16)**. Exercise restriction is warranted. Treatment with beta blockers and ICD implantation has been recommended. Prevention of inappropriate or easily triggered ICD shocks by proper ICD programming is essential to prevent VT storm from endogenous catecholamine release.

Figure 233-16

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Catecholaminergic polymorphic ventricular tachycardia noted during an exercise stress test.

Special Consideration: Approach to Tachyarrhythmias in Athletes

The first manifestation of a tachyarrhythmia, whether benign or malignant, may occur during athletic activity. Fortunately, successful cardiopulmonary resuscitation of life-threatening ventricular arrhythmias has increased with the use of automatic external defibrillators at major sporting events and schools. Rarely, VF may be precipitated by blunt precordial blows without structural injury to the heart or chest wall (commotio cordis).

The approach to the athlete should begin with an assessment of the severity and significance of symptoms. Syncope with exertion should be assumed to be caused by a potentially lethal arrhythmia. A thorough cardiac evaluation and restricted participation in competitive sports are in order until a less life-threatening diagnosis can be established. ECG recording at the time of the symptomatic events usually can establish the diagnosis, although it may be difficult to obtain.

In patients with syncope and no ECG-documented arrhythmia, a systematic attempt to define a cardiac structural or primarily electrical abnormality with a routine ECG and a transthoracic echocardiogram is in order. Common structural abnormalities associated with fatal or life-threatening ventricular arrhythmias include hypertrophic cardiomyopathy, arrhythmogenic cardimyopathy, and acute myocarditis. Coronary anomalies should be suspected if the arrhythmia symptoms are preceded in onset by chest pain. The 12-lead ECG should be screened for the presence of preexcitation, QT prolongation, a Brugada-type ECG pattern or epsilon waves, and T-wave inversions consistent with a nonischemic RV or LV cardiomyopathy or myocarditis. Additional ECG monitoring may be required. A stress test may be appropriate to provoke arrhythmias, especially if there are recurrent symptoms. It is critical to achieve the level of exercise that precipitated the arrhythmia, which for some athletes may be a challenge for the exercise lab.

Management of an athlete with cardiac arrhythmias may be a challenge, with a tendency to discourage participation in competitive sports and institute treatment whenever there is a perception of increased risk. Guidelines for restricting athletic activity have been promulgated on the basis of expert consensus and evidence-based data and can facilitate management once a diagnosis has been established (**Table 233-9**). Treatment should be based on standards established for each arrhythmia syndrome. Curative catheter ablative therapy should be applied if indicated. ICD therapy, if required, is incompatible with contact sports because of the potential for blunt trauma and consequent damage to the device. Although ICDs are effective, their psychosocial impact, the potential for inappropriate shocks for sinus tachycardia, and lead-related complications must be recognized.

Table 233–9. Recommendations for Competitive Athletes with Selected Cardiovascular Abnormalities

Clinical Entity	Clinical Criteria	Sports Permitted
Gene carriers for arrhythmia syndromes without phenotype VT		All sports
Long QT syndrome	>0.47 s in men, >0.48 s in women	Low-intensity competitive sports
Brugada syndrome		Low-intensity competitive sports
Catecholaminergic polymorphic VT		No competitive sports
Asymptomatic Wolff-Parkinson-White syndrome	Electrophysiological study not mandatory	All sports except restriction in dangerous environments
Premature ventricular complexes		All competitive sports when no increase in PVCs or symptoms occur with exercise
Nonsustained ventricular tachycardia	No structural heart disease	All competitive sports
Nonsustained ventricular tachycardia	Structural heart disease	Only low-intensity competitive sports

Source: Adapted from ACC Bethesda Conference #36 from Pelliccia et al: J Am Coll Cardiol 52:1990–1996, 2008.

FURTHER READINGS

Antzelevitch C et al: Brugada syndrome: Report of the second consensus conference. Heart Rhythm 2:429, 2005[PMID: 15898165]

Cappato R et al: Updated worldwide survey on the methods, efficacy, and safety of catheter ablation for human atrial fibrillation. Circ Arrhythm Electrophysiol 3:32, 2010[PMID: 19995881]

Delacretaz E: Clinical practice: Supraventricular tachycardia. N Engl J Med 354:1039, 2006[PMID: 16525141]

Epstein AE et al: ACC/AHA/HRS 2008 Guidelines for Device-Based Therapy of Cardiac Rhythm Abnormalities: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the ACC/AHA/NASPE 2002 Guideline Update for Implantation of Cardiac Pacemaker and Antiarrhythmia Devices): Developed in collaboration with the American Association for Thoracic Surgery and Society of Thoracic Surgeons. Circulation 117:e350, 2008

Fuster V et al: ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation. Circulation 114:e257, 2006

Groh WJ et al. Electrocardiographic abnormalities and sudden death in myotonic dystrophy type 1. N Engl J Med 358(25):2688, 2008[PMID: 18565861]

Josephson ME: *Clinical Cardiac Electrophysiology*: *Techniques and Interpretations*, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2008

Lehnart SE et al: Inherited arrhythmias: Consensus report about the diagnosis, phenotyping, molecular mechanisms, and therapeutic approaches for primary cardiomyopathies of gene mutations affecting ion channel function. Circulation 116:2325, 2007[PMID: 17998470]

Marchlinski FE et al: Ventricular tachycardia/ventricular fibrillation ablation in the setting of ischemic heart disease. J Cardiovasc Electrophysiol 16:59, 2005

Maron BJ, Zipes DP: Introduction: Eligibility recommendations for competitive athletes with cardiovascular abnormalities—general considerations. J Am Coll Cardiol 45:1318, 2005[PMID: 15837280]

Miller JM, Zipes DP: Therapy for cardiac arrhythmias, in *Braunwald's Heart Disease*: *A Textbook of Cardiovascular Medicine*, 7th ed. DP Zipes et al (eds). Philadelphia, Saunders, 2005, pp 712–756

Page RL: Medical management of atrial fibrillation: Future directions. Heart Rhythm 4:580, 2007

Priori SG et al: Risk stratification in the long QT syndrome. N Engl J Med 348:1866, 2003[PMID: 12736279]

Roden DM: Drug-induced prolongation of the QT interval. N Engl J Med 350:1013, 2004[PMID: 14999113]

Taylor FC et al: Systematic review of long term anticoagulation or antiplatelet treatment in patients with non-rheumatic atrial fibrillation. BMJ 322:321, 2001[PMID: 11159653]

Wyse DG et al: Affirm: A comparison of rate control and rhythm control in patients with atrial fibrillation. N Engl J Med 347:1825, 2002[PMID: 12466506]

Zipes DP et al: ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. Circulation 114:e385, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF CARDIAC ARRHYTHMIAS: INTRODUCTION

The electrocardiograms in this Atlas supplement those illustrated in Chaps. 232 and 233. The interpretations emphasize findings of specific teaching value.

All of the figures are adapted from cases in ECG Wave-Maven, Copyright 2003, Beth Israel Deaconess Medical Center,*http://ecg.bidmc.harvard.edu*.

The abbreviations used in this chapter are as follows:

AF—atrial fibrillation

AV—atrioventricular

AVRT—atrioventricular reentrant tachycardia

LBBB—left bundle branch block

LV—left ventricular

LVH—left ventricular hypertrophy

MI—myocardial infarction

NSR—normal sinus rhythm

RBBB—right bundle branch block

VT—ventricular tachycardia

WPW—Wolff-Parkinson-White

Figure e30-1

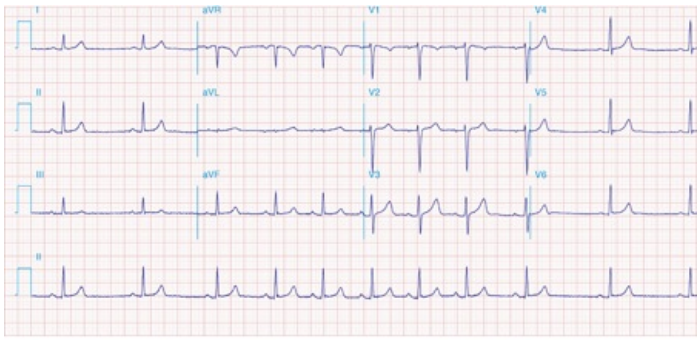

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Respiratory sinus arrhythmia, a physiologic finding in a healthy young adult. The rate of the sinus pacemaker is relatively slow at the beginning of the strip during expiration, then accelerates during inspiration and slows again with expiration. Changes are due to cardiac vagal tone modulation with breathing.

Figure e30-2

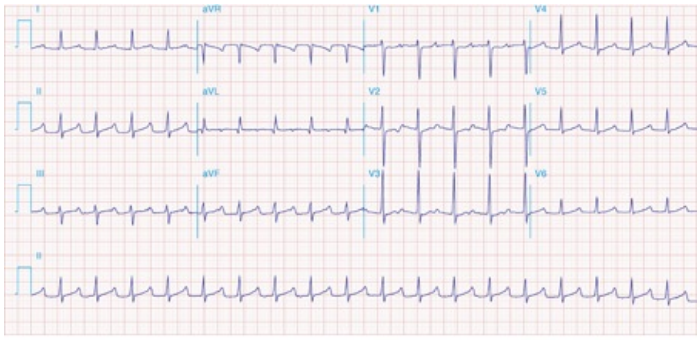

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus tachycardia (110/min) with first-degree AV "block" (conduction delay) with PR interval = 0.28 s. The P wave is visible after the ST-T wave in V_1–V_3 and superimposed on the T wave in other leads. Atrial (non-sinus) tachycardias may produce a similar pattern, but the rate is usually faster.

Figure e30-3

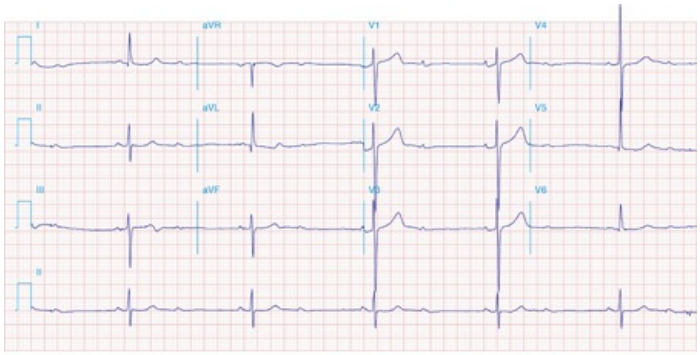

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus rhythm (P wave rate about 60/min) with **2:1 AV (second-degree) block** causing **marked bradycardia** (ventricular rate of about 30/min). **LVH** is also present.

Figure e30-4

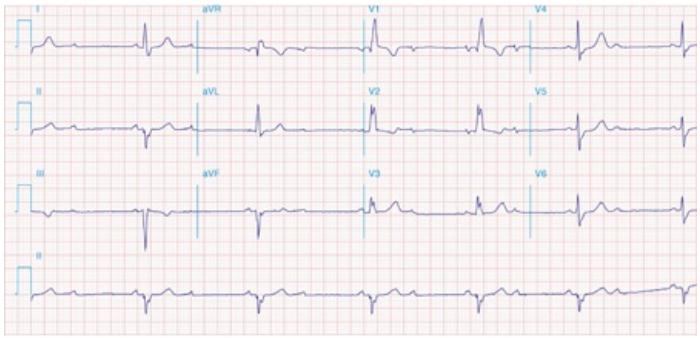

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus rhythm (P wave rate about 60/min) **with 2:1 (second-degree) AV block** yielding a ventricular (pulse) rate of about 30/min.**Left atrialabnormality**. **RBBB** with **left anterior fascicular block.** Possible **inferior MI.**

Figure e30-5

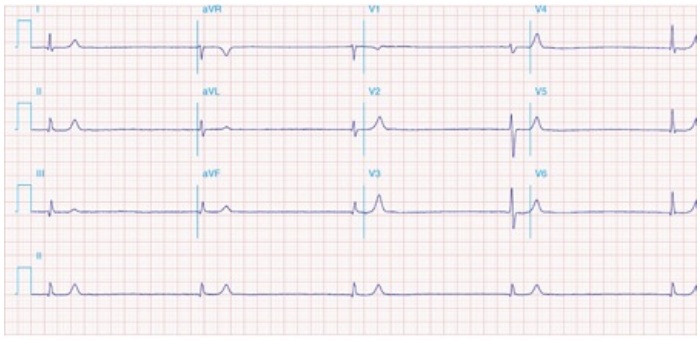

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Marked junctional bradycardia (25 beats/min). Rate is regular with a flat baseline between narrow QRS complexes, without evident P waves. Patient was on atenolol, with **possible underlying sick sinus syndrome**.

Figure e30-6

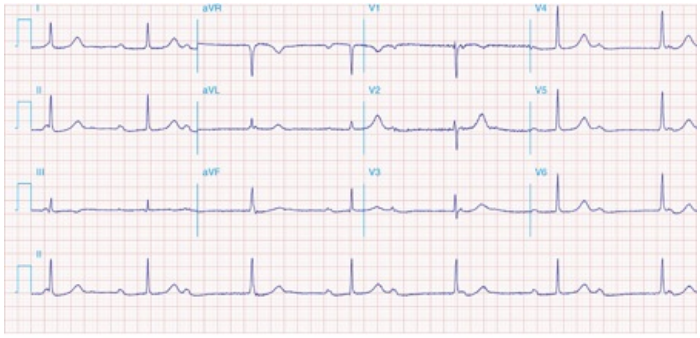

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus rhythm at a rate of 64/min (P wave rate) with **third-degree (complete) AV block** yielding an effective heart (pulse) rate of 40/min. The slow, narrow QRS complexes indicate an A-V junctional escape pacemaker. **Left atrial abnormality.**

Figure e30-7

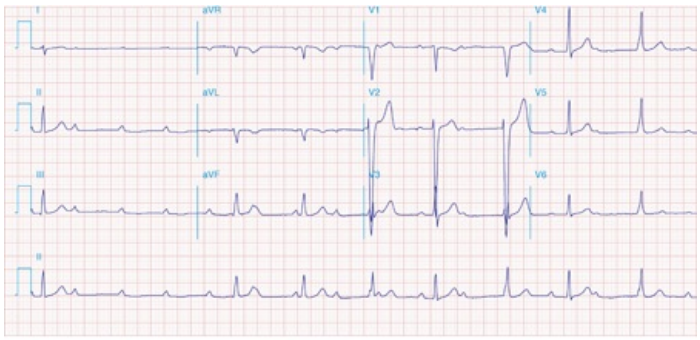

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sinus rhythm at a rate of 90/min with **advanced second-degree AV block** and **possible transient complete heart block** with **Lyme carditis.**

Figure e30-8

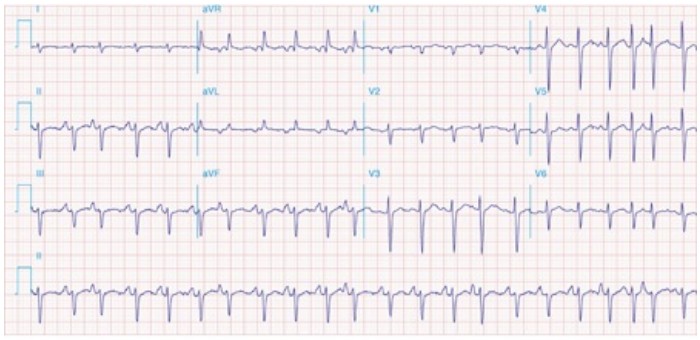

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multifocal atrial tachycardia with varying P-wave morphologies and P-P intervals; right atrial overload with peaked P waves in II, III, and aVF (with vertical P wave axis); superior QRS axis; slow R-wave progression with delayed transition in precordial leads in patient with **severe chronic obstructive lung disease.**

Figure e30-9

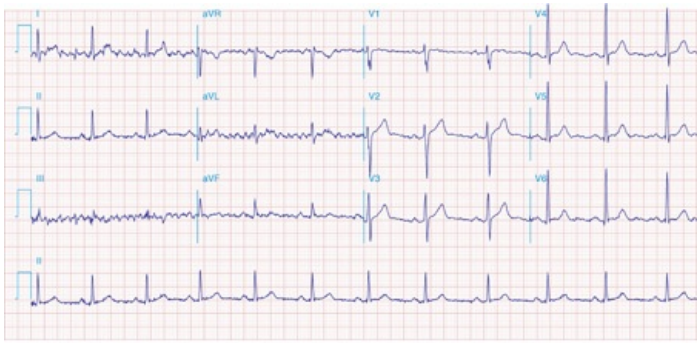

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

NSR in a patient with **Parkinson's disease.** Tremor artifact, best seen in limb leads. This **tremor artifact** may sometimes be confused with atrial flutter/fibrillation. Borderline voltage criteria for LVH are present.

Figure e30-10

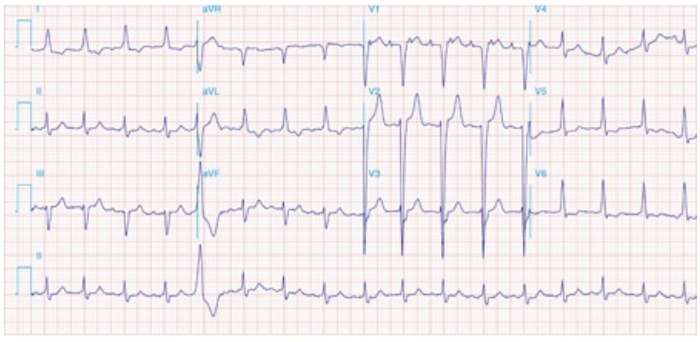

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial tachycardia with atrial rate of about 200/min (note lead V_1), **2:1 AV block (conduction),** and one premature ventricular complex. Also present: LVH with intraventricular conduction delay and slow precordial R-wave progression (cannot rule out prior anterior MI).

Figure e30-11

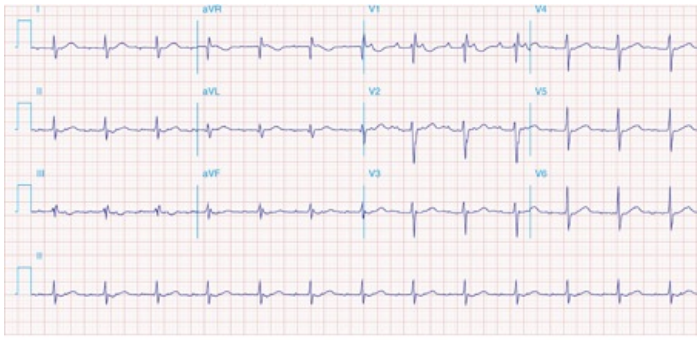

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial tachycardia with 2:1 block. P-wave rate is about 150/min, with ventricular (QRS) rate of about 75/min. The nonconducted ("extra") P waves just after the QRS complex are best seen in lead V_1. Also, note incomplete RBBB and borderline QT prolongation.

Figure e30-12

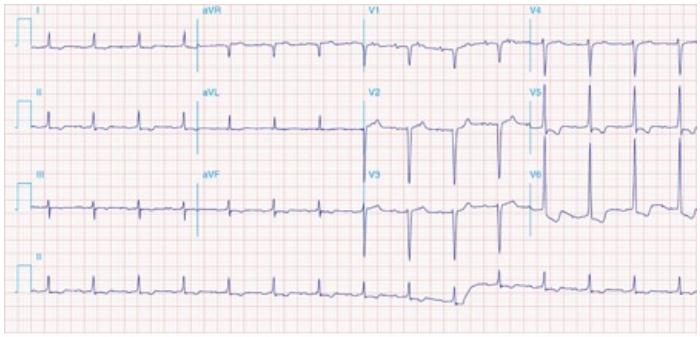

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial tachycardia [180/min with 2:1 AV block (see lead V_1)]. **LVH** by precordial voltage and nonspecific ST-T changes. Slow R-wave progression (V_1–V_4) raises consideration of prior anterior MI.

Figure e30-13

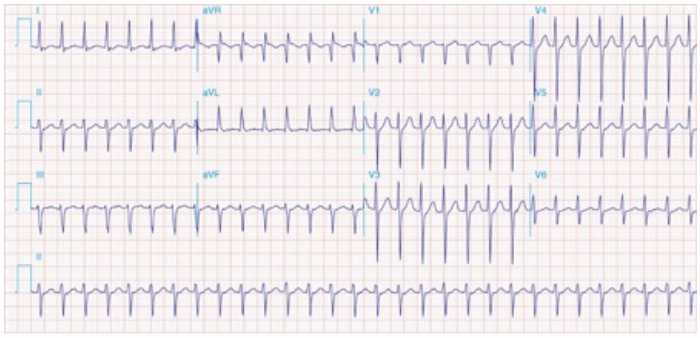

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

AV nodal reentrant tachycardia (AVNRT) at a rate of 150/min. Note subtle "pseudo" R waves in lead aVR due to retrograde atrial activation, which occurs nearly simultaneous with ventricles in AVNRT. Left-axis deviation consistent with left anterior fascicular block (hemiblock) is also present.

Figure e30-14

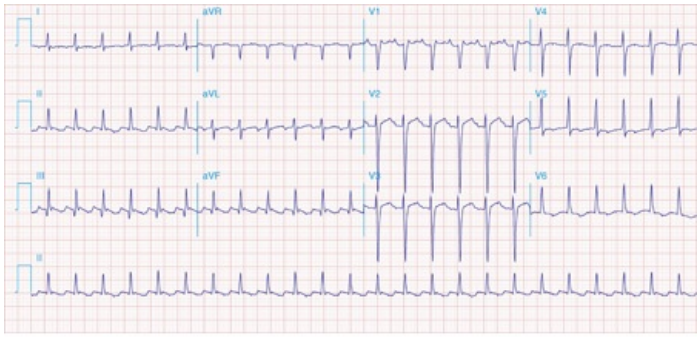

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial flutter with 2:1 AV conduction. Note atrial flutter waves, partly hidden in the early ST segment, seen, for example, in leads II and V_1.

Figure e30-15

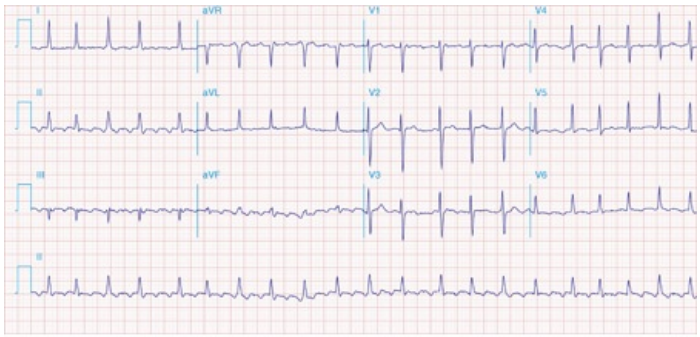

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atrial flutter with atrial rate 300/min and variable (predominant 2:1 and 3:1) AV conduction. Typical flutter waves best seen in lead II.

Figure e30-16

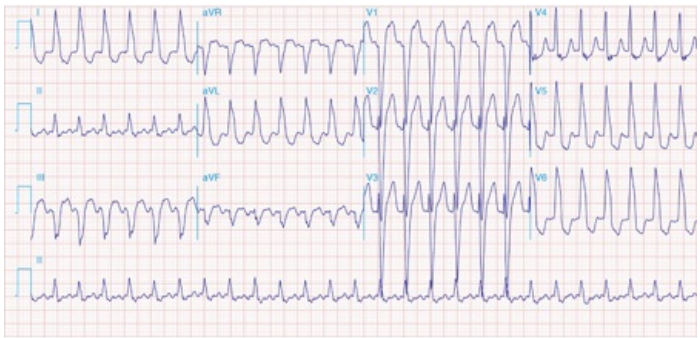

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Wide complex tachycardia.Atrial flutter with 2:1 AV conduction (block) and **LBBB**, not to be mistaken for VT. Typical atrial flutter activity is clearly present in lead II at a cycle rate of about 320/min, with an effective ventricular rate of about 160/min.

Figure e30-17

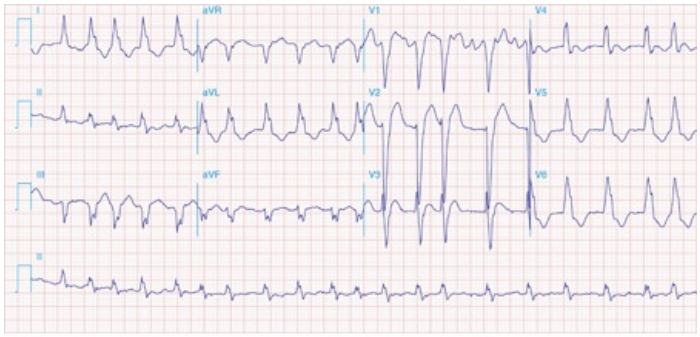

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

AF with LBBB. The ventricular rhythm is erratically irregular. Coarse fibrillatory waves are best seen in lead V_1, with a typical LBBB pattern.

Figure e30-18

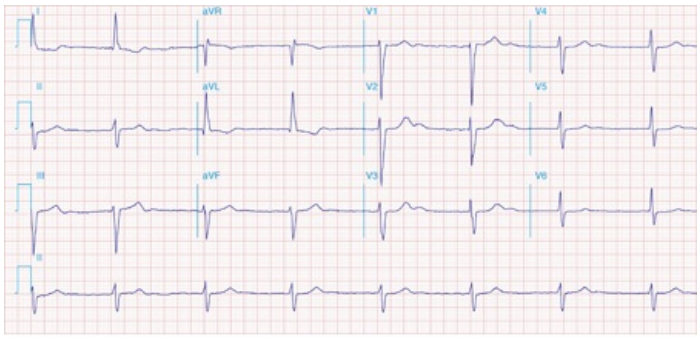

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

AF with complete heart block and a junctional escape mechanism causing a slow regular ventricular response (45/min). The QRS complexes show an intraventricular conduction delay with left-axis deviation and LVH. Q-T (U) prolongation is also present.

Figure e30-19

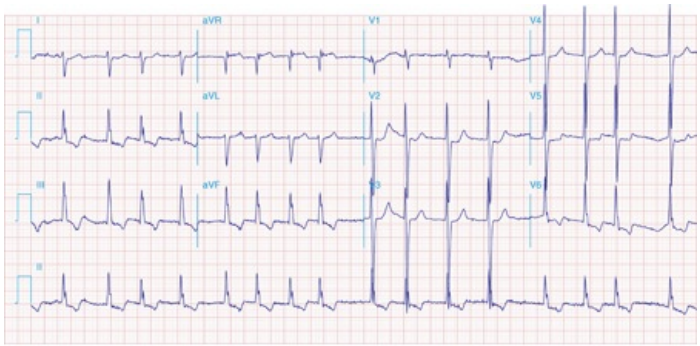

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

AF with right-axis deviation and LVH. Tracing suggests biventricular hypertrophy in a patient with **mitral stenosis and aortic valve disease.**

Figure e30-20

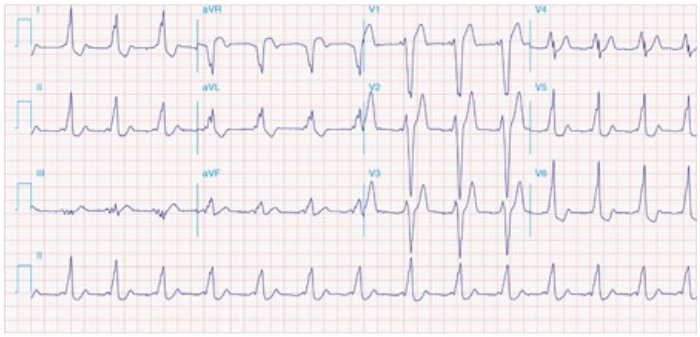

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

WPW pre-excitation pattern, with triad of short PR, wide QRS, and delta waves. Polarity of the delta waves (slightly positive in leads V_1 and V_2 and most positive in lead II and lateral chest leads) is consistent with a right-sided bypass tract.

Figure e30-21

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

AF in patient with the WPW syndrome and antegrade conduction down the bypass tract leading to a wide complex tachycardia. Rhythm is "irregularly irregular" and rate is extremely rapid (about 230/min). Not all beats are pre-excited.

Figure e30-22

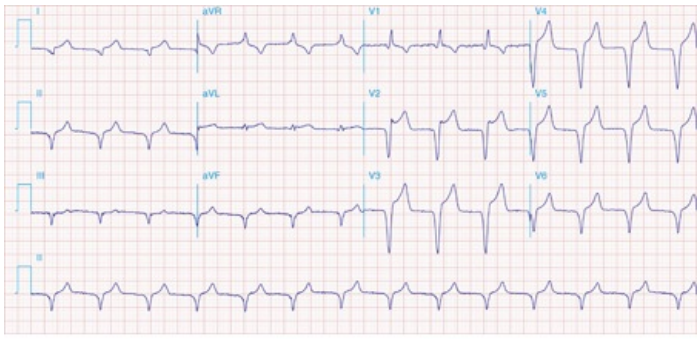

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Accelerated idioventricular rhythm (AIVR) originating from the LV and accounting for RBBB morphology. ST elevations in the precordial leads from **underlying acute MI.**

Figure e30-23

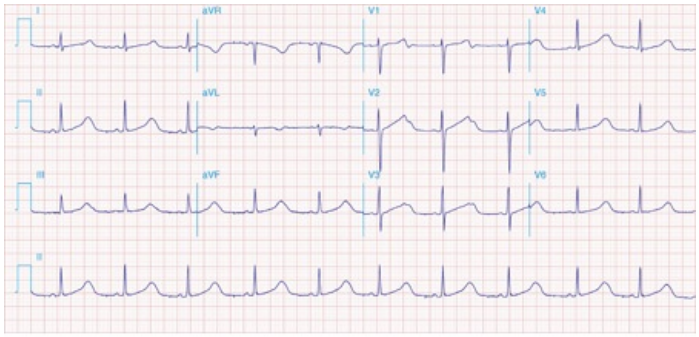

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prolonged (0.60 s) QT interval in a patient with **hereditary long-QT syndrome.**

Figure e30-24

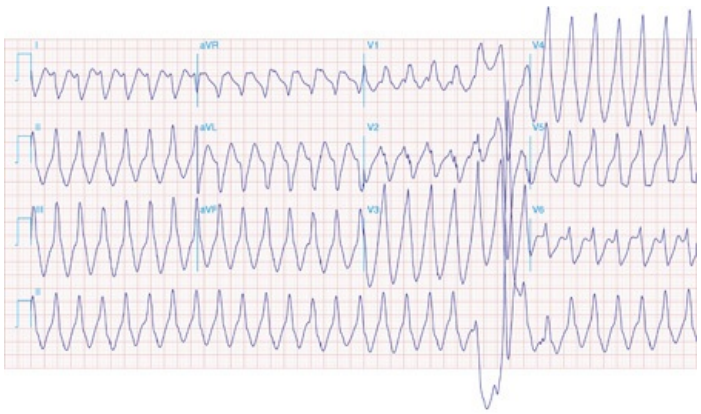

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Monomorphic VT at rate of 170/min. The RBBB morphology in V_1 and the R:S ratio <1 in V_6 are both suggestive of VT. The morphology of the VT is suggestive of origin from the left side of the heart, near the base (RBBB with inferior/rightward axis). Baseline artifact is present in leads V_1–V_3.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEART FAILURE

Definition

Heart failure (HF) is a clinical syndrome that occurs in patients who, because of an inherited or acquired abnormality of cardiac structure and/or function, develop a constellation of clinical symptoms (dyspnea and fatigue) and signs (edema and rales) that lead to frequent hospitalizations, a poor quality of life, and a shortened life expectancy.

Epidemiology

HF is a burgeoning problem worldwide, with more than 20 million people affected. The overall prevalence of HF in the adult population in developed countries is 2%. HF prevalence follows an exponential pattern, rising with age, and affects 6–10% of people over age 65. Although the relative incidence of HF is lower in women than in men, women constitute at least one-half the cases of HF because of their longer life expectancy. In North America and Europe, the lifetime risk of developing HF is approximately one in five for a 40-year-old. The overall prevalence of HF is thought to be increasing, in part because current therapies for cardiac disorders, such as myocardial infarction (MI), valvular heart disease, and arrhythmias, are allowing patients to survive longer. Very little is known about the prevalence or risk of developing HF in emerging nations because of the lack of population-based studies in those countries. Although HF once was thought to arise primarily in the setting of a depressed left ventricular (LV) ejection fraction (EF), epidemiologic studies have shown that approximately one-half of patients who develop HF have a normal or preserved EF (EF ≥40–50%). Accordingly, HF patients are now broadly categorized into one of two groups: (1) HF with a depressed EF (commonly referred to as *systolic failure*) or (2) HF with a preserved EF (commonly referred to as *diastolic failure*).

Etiology

As shown in **Table 234-1**, any condition that leads to an alteration in LV structure or function can predispose a patient to developing HF. Although the etiology of HF in patients with a preserved EF differs from that of patients with depressed EF, there is considerable overlap between the etiologies of these two conditions. In industrialized countries, coronary artery disease (CAD) has become the predominant cause in men and women and is responsible for 60–75% of cases of HF. Hypertension contributes to the development of HF in 75% of patients, including most patients with CAD. Both CAD and hypertension interact to augment the risk of HF, as does diabetes mellitus.

Table 234-1 Etiologies of Heart Failure

Depressed Ejection Fraction (<40%)	
Coronary artery disease	Nonischemic dilated cardiomyopathy
Myocardial infarction[a]	Familial/genetic disorders
Myocardial ischemia[a]	Infiltrative disorders[a]
Chronic pressure overload	Toxic/drug-induced damage
Hypertension[a]	Metabolic disorder[a]
Obstructive valvular disease[a]	Viral
Chronic volume overload	Chagas' disease
Regurgitant valvular disease	Disorders of rate and rhythm
Intracardiac (left-to-right) shunting	Chronic bradyarrhythmias
Extracardiac shunting	Chronic tachyarrhythmias
Preserved Ejection Fraction (>40–50%)	
Pathologic hypertrophy	Restrictive cardiomyopathy

Primary (hypertrophic cardiomyopathies)	Infiltrative disorders (amyloidosis, sarcoidosis)
Secondary (hypertension)	Storage diseases (hemochromatosis)
Aging	Fibrosis
	Endomyocardial disorders
Pulmonary Heart Disease	
Cor pulmonale	
Pulmonary vascular disorders	
High-Output States	
Metabolic disorders	Excessive blood-flow requirements
Thyrotoxicosis	Systemic arteriovenous shunting
Nutritional disorders (beriberi)	Chronic anemia

Note: [a] Indicates conditions that can also lead to heart failure with a preserved ejection fraction.

In 20–30% of the cases of HF with a depressed EF, the exact etiologic basis is not known. These patients are referred to as having nonischemic, dilated, or idiopathic cardiomyopathy if the cause is unknown (Chap. 238). Prior viral infection or toxin exposure (e.g., alcoholic or chemotherapeutic) also may lead to a dilated cardiomyopathy. Moreover, it is becoming increasingly clear that a large number of cases of dilated cardiomyopathy are secondary to specific genetic defects, most notably those in the cytoskeleton. Most forms of familial dilated cardiomyopathy are inherited in an autosomal dominant fashion. Mutations of genes that encode cytoskeletal proteins (desmin, cardiac myosin, vinculin) and nuclear membrane proteins (laminin) have been identified thus far. Dilated cardiomyopathy also is associated with Duchenne's, Becker's, and limb-girdle muscular dystrophies. Conditions that lead to a high cardiac output (e.g., arteriovenous fistula, anemia) are seldom responsible for the development of HF in a normal heart; however, in the presence of underlying structural heart disease, these conditions can lead to overt HF.

Global Considerations

Rheumatic heart disease remains a major cause of HF in Africa and Asia, especially in the young. Hypertension is an important cause of HF in the African and African-American populations. Chagas' disease is still a major cause of HF in South America. Not surprisingly, anemia is a frequent concomitant factor in HF in many developing nations. As developing nations undergo socioeconomic development, the epidemiology of HF is becoming similar to that of Western Europe and North America, with CAD emerging as the single most common cause of HF. Although the contribution of diabetes mellitus to HF is not well understood, diabetes accelerates atherosclerosis and often is associated with hypertension.

Prognosis

Despite many recent advances in the evaluation and management of HF, the development of symptomatic HF still carries a poor prognosis. Community-based studies indicate that 30–40% of patients die within 1 year of diagnosis and 60–70% die within 5 years, mainly from worsening HF or as a sudden event (probably because of a ventricular arrhythmia). Although it is difficult to predict prognosis in an individual, patients with symptoms at rest [New York Heart Association (NYHA) class IV] have a 30–70% annual mortality rate, whereas patients with symptoms with moderate activity (NYHA class II) have an annual mortality rate of 5–10%. Thus, functional status is an important predictor of patient outcome (**Table 234-2**).

Table 234-2 New York Heart Association Classification

Functional Capacity	Objective Assessment
Class I	Patients with cardiac disease but without resulting limitation of physical activity. Ordinary physical activity does not cause undue fatigue, palpitations, dyspnea, or anginal pain.
Class II	Patients with cardiac disease resulting in slight limitation of physical activity. They are comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea, or anginal pain.
Class III	Patients with cardiac disease resulting in marked limitation of physical activity. They are comfortable at rest. Less than ordinary activity causes fatigue, palpitation, dyspnea, or anginal pain.
Class IV	Patients with cardiac disease resulting in inability to carry on any physical activity without discomfort. Symptoms of heart failure or the anginal syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.

Source: Adapted from New York Heart Association, Inc., *Diseases of the Heart and Blood Vessels:*

Nomenclature and Criteria for Diagnosis, 6th ed. Boston, Little Brown, 1964, p. 114.

Pathogenesis

Figure 234-1 provides a general conceptual framework for considering the development and progression of HF with a depressed EF. As shown, HF may be viewed as a progressive disorder that is initiated after an *index event* either damages the heart muscle, with a resultant loss of functioning cardiac myocytes, or, alternatively, disrupts the ability of the myocardium to generate force, thereby preventing the heart from contracting normally. This index event may have an abrupt onset, as in the case of a myocardial infarction (MI); it may have a gradual or insidious onset, as in the case of hemodynamic pressure or volume overloading; or it may be hereditary, as in the case of many of the genetic cardiomyopathies. Regardless of the nature of the inciting event, the feature that is common to each of these index events is that they all in some manner produce a decline in the pumping capacity of the heart. In most instances, patients remain asymptomatic or minimally symptomatic after the initial decline in pumping capacity of the heart or develop symptoms only after the dysfunction has been present for some time.

Figure 234-1

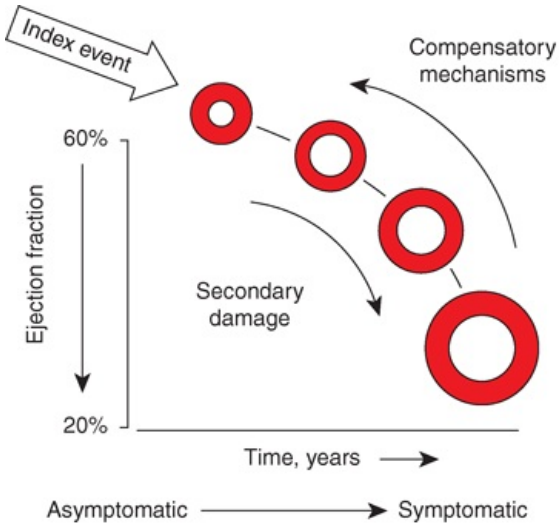

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis of heart failure with a depressed ejection fraction. Heart failure begins after an index event produces an initial decline in the heart's pumping capacity. After this initial decline in pumping capacity, a variety of compensatory mechanisms are activated, including the adrenergic nervous system, the renin-angiotensin-aldosterone system, and the cytokine system. In the short term, these systems are able to restore cardiovascular function to a normal homeostatic range with the result that the patient remains asymptomatic. However, with time the sustained activation of these systems can lead to secondary end-organ damage within the ventricle, with worsening left ventricular remodeling and subsequent cardiac decompensation. *(From D Mann: Circulation 100:999, 1999.)*

Although the precise reasons why patients with LV dysfunction may remain asymptomatic is not certain, one potential explanation is that a number of compensatory mechanisms become activated in the presence of cardiac injury and/or LV dysfunction allowing patients to sustain and modulate LV function for a period of months to years. The list of compensatory mechanisms that have been described thus far include (1) activation of the renin-angiotensin-aldosterone (RAA) and adrenergic nervous systems, which are responsible for maintaining cardiac output through increased retention of salt and water **(Fig. 234-2)**, and (2) increased myocardial contractility. In addition, there is activation of a family of countervailing vasodilatory molecules, including the atrial and brain natriuretic peptides (ANP and BNP), prostaglandins (PGE2 and PGI2), and nitric oxide (NO), that offsets the excessive peripheral vascular vasoconstriction. Genetic background, sex, age, or environment may influence these compensatory mechanisms, which are able to modulate LV function within a physiologic/homeostatic range so that the functional capacity of the patient is preserved or is depressed only minimally. Thus, patients may remain asymptomatic or minimally symptomatic for a period of years; however, at some point patients become overtly symptomatic, with a resultant striking increase in morbidity and mortality rates. Although the exact mechanisms that are responsible for this transition are not known, as will be discussed below, the transition to symptomatic HF is accompanied by increasing activation of neurohormonal, adrenergic, and cytokine systems that lead to a series of adaptive changes within the myocardium collectively referred to as *LV remodeling* .

Figure 234-2

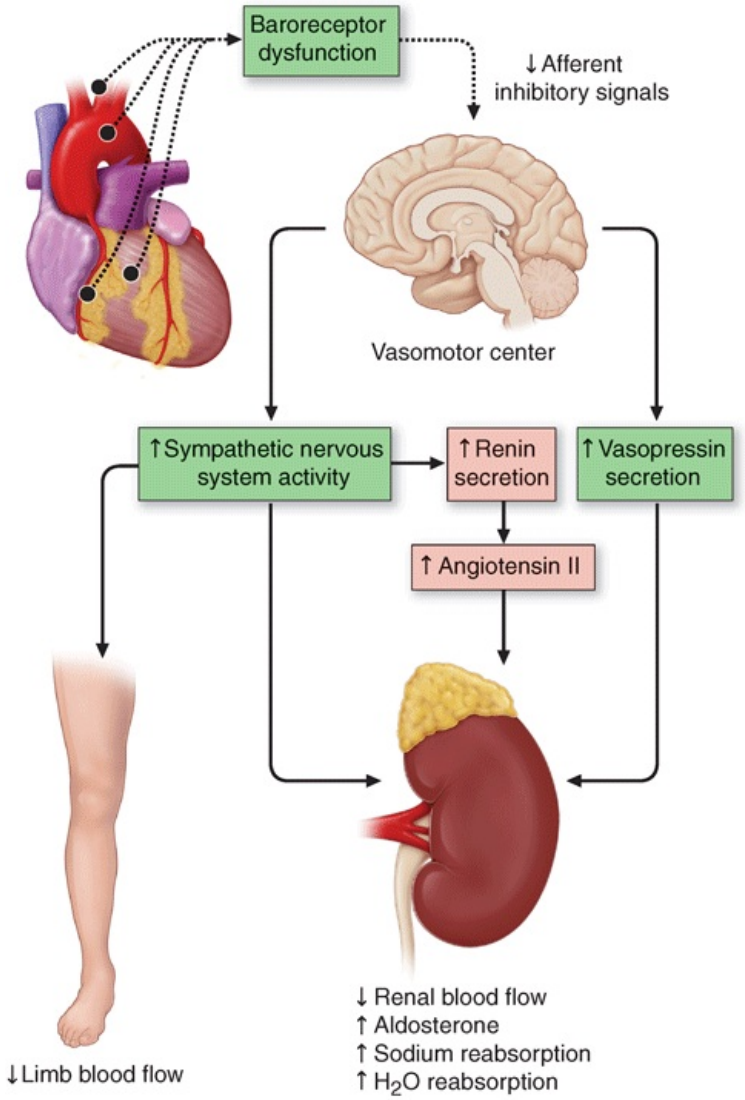

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Activation of neurohormonal systems in heart failure. The decreased cardiac output in HF patients results in an "unloading" of high-pressure baroceptors (circles) in the left ventricle, carotid sinus, and aortic arch. This unloading of the peripheral baroreceptors leads to a loss of inhibitory parasympathetic tone to the central nervous system (CNS), with a resultant generalized increase in efferent sympathetic tone, and non-osmotic release of arginine vasopressin (AVP) from the pituitary. AVP [or antidiuretic hormone (ADH)] is a powerful vasoconstrictor that increases the permeability of the renal collecting ducts, leading to the reabsorption of free water. These afferent signals to the CNS also activate efferent sympathetic nervous system pathways that innervate the heart, kidney, peripheral vasculature, and skeletal muscles.

Sympathetic stimulation of the kidney leads to the release of renin, with a resultant increase in the circulating levels of angiotensin II and aldosterone. The activation of the renin-angiotensin-aldosterone system promotes salt and water retention and leads to vasoconstriction of the peripheral vasculature, myocyte hypertrophy, myocyte cell death, and myocardial fibrosis. Although these neurohormonal mechanisms facilitate short-term adaptation by maintaining blood pressure, and hence perfusion to vital organs, the same neurohormonal mechanisms are believed to contribute to end-organ changes in the heart and the circulation and to the excessive salt and water retention in advanced HF. *[Modified from A Nohria et al: Neurohormonal, renal and vascular adjustments, in Atlas of Heart Failure: Cardiac Function and Dysfunction, 4th ed, WS Colucci (ed). Philadelphia, Current Medicine Group 2002, p. 104.]*

In contrast to our understanding of the pathogenesis of HF with a depressed EF, our understanding of the mechanisms that contribute to the development of HF with a preserved EF is still evolving. That is, although diastolic dysfunction (see below) was thought to be the only mechanism responsible for the development of HF with a preserved EF, community-based studies suggest that additional extracardiac mechanisms may be important, such as increased vascular stiffness and impaired renal function.

Basic Mechanisms of Heart Failure

SYSTOLIC DYSFUNCTION

LV remodeling develops in response to a series of complex events that occur at the cellular and molecular levels **(Table 234-3).** These changes include (1) myocyte hypertrophy, (2) alterations in the contractile properties of the myocyte, (3) progressive loss of myocytes through necrosis, apoptosis, and autophagic cell death, (4) ß-adrenergic desensitization, (5) abnormal myocardial energetics and metabolism, and (6) reorganization of the extracellular matrix with dissolution of the organized structural collagen weave surrounding myocytes and subsequent replacement by an interstitial collagen matrix that does not provide structural support to the myocytes. The biologic stimuli for these profound changes include mechanical stretch of the myocyte, circulating neurohormones (e.g., norepinephrine, angiotensin II), inflammatory cytokines [e.g., tumor necrosis factor (TNF)], other peptides and growth factors (e.g., endothelin), and reactive oxygen species (e.g., superoxide). The sustained overexpression of these biologically active molecules is believed to contribute to the progression of HF by virtue of the deleterious effects they exert on the heart and the circulation. Indeed, this insight forms the clinical rationale for using pharmacologic agents that antagonize these systems [e.g., angiotensin-converting enzyme (ACE) inhibitors and beta blockers] in treating patients with HF.

Table 234-3 Overview of Left Ventricular Remodeling

Alterations in Myocyte Biology
Excitation-contraction coupling
Myosin heavy chain (fetal) gene expression
ß-adrenergic desensitization
Hypertrophy
Myocytolysis
Cytoskeletal proteins
Myocardial Changes
Myocyte loss
Necrosis
Apoptosis
Autophagy
Alterations in extracellular matrix
Matrix degradation
Myocardial fibrosis
Alterations in Left Ventricular Chamber Geometry
Left ventricular (LV) dilation
Increased LV sphericity
LV wall thinning
Mitral valve incompetence

Source: Adapted from D Mann: Pathophysiology of heart failure, in *Braunwald's Heart Disease*, 8th ed, PL Libby et al (eds). Philadelphia, Elsevier, 2008, p. 550.

In order to understand how the changes that occur in the failing cardiac myocyte contribute to depressed LV systolic function in HF, it is instructive first to review the biology of the cardiac muscle cell (Chap. 224). Sustained neurohormonal activation and mechanical overload result in transcriptional and posttranscriptional changes in the genes and proteins that regulate excitation-contraction coupling and cross-bridge interaction (see Figs. 224-6 and 224-7). The changes that regulate excitation-contraction include decreased function of sarcoplasmic reticulum Ca^{2+} adenosine triphosphatase (SERCA2A), resulting in decreased calcium uptake into the sarcoplamsic reticulum (SR), and hyperphosphorylation of the ryanodine receptor, leading to calcium leakage from the SR. The changes that occur in the cross-bridges include decreased expression of α-myosin heavy chain and increased expression of ß-myosin heavy chain, myocytolysis, and disruption of the cytoskeletal links between the sarcomeres and the extracellular matrix. Collectively, these changes impair the ability of the myocyte to contract and therefore contribute to the depressed LV systolic function observed in patients with HF.

DIASTOLIC DYSFUNCTION

Myocardial relaxation is an adenosine triphosphate (ATP)-dependent process that is regulated by uptake of cytoplasmic calcium into the SR by SERCA2A and extrusion of calcium by sarcolemmal pumps (see Fig. 224-7). Accordingly, reductions in ATP concentration, as occurs in ischemia, may interfere with these processes and lead to slowed myocardial relaxation. Alternatively, if LV filling is

delayed because LV compliance is reduced (e.g., from hypertrophy or fibrosis), LV filling pressures will similarly remain elevated at end diastole (see Fig. 224-11). An increase in heart rate disproportionately shortens the time for diastolic filling, which may lead to elevated LV filling pressures, particularly in noncompliant ventricles. Elevated LV end-diastolic filling pressures result in increases in pulmonary capillary pressures, which can contribute to the dyspnea experienced by patients with diastolic dysfunction. In addition to impaired myocardial relaxation, increased myocardial stiffness secondary to cardiac hypertrophy and increased myocardial collagen content may contribute to diastolic failure. Importantly, diastolic dysfunction can occur alone or in combination with systolic dysfunction in patients with HF.

LEFT VENTRICULAR REMODELING

Ventricular remodeling refers to the changes in LV mass, volume, and shape and the composition of the heart that occur after cardiac injury and/or abnormal hemodynamic loading conditions. LV remodeling may contribute independently to the progression of HF by virtue of the mechanical burdens that are engendered by the changes in the geometry of the remodeled LV. In addition to the increase in LV end-diastolic volume, LV wall thinning occurs as the left ventricle begins to dilate. The increase in wall thinning, along with the increase in afterload created by LV dilation, leads to a functional *afterload mismatch* that may contribute further to a decrease in stroke volume. Moreover, the high end-diastolic wall stress might be expected to lead to (1) hypoperfusion of the subendocardium, with resultant worsening of LV function, (2) increased oxidative stress, with the resultant activation of families of genes that are sensitive to free radical generation (e.g., TNF and interleukin 1β), and (3) sustained expression of stretch-activated genes (angiotensin II, endothelin, and TNF) and/or stretch activation of hypertrophic signaling pathways. Increasing LV dilation also results in tethering of the papillary muscles with resulting incompetence of the mitral valve apparatus and functional mitral regurgitation, which in turn leads to further hemodynamic overloading of the ventricle. Taken together, the mechanical burdens that are engendered by LV remodeling contribute to the progression of HF.

Clinical Manifestations

SYMPTOMS

The cardinal symptoms of HF are fatigue and shortness of breath. Although fatigue traditionally has been ascribed to the low cardiac output in HF, it is likely that skeletal-muscle abnormalities and other noncardiac comorbidities (e.g., anemia) also contribute to this symptom. In the early stages of HF, dyspnea is observed only during exertion; however, as the disease progresses, dyspnea occurs with less strenuous activity, and it ultimately may occur even at rest. The origin of dyspnea in HF is probably multifactorial (Chap. 33). The most important mechanism is pulmonary congestion with accumulation of interstitial or intra-alveolar fluid, which activates juxtacapillary J receptors, which in turn stimulate the rapid, shallow breathing characteristic of cardiac dyspnea. Other factors that contribute to dyspnea on exertion include reductions in pulmonary compliance, increased airway resistance, respiratory muscle and/or diaphragm fatigue, and anemia. Dyspnea may become less frequent with the onset of right ventricular (RV) failure and tricuspid regurgitation.

Orthopnea

Orthopnea, which is defined as dyspnea occurring in the recumbent position, is usually a later manifestation of HF than is exertional dyspnea. It results from redistribution of fluid from the splanchnic circulation and lower extremities into the central circulation during recumbency, with a resultant increase in pulmonary capillary pressure. Nocturnal cough is a common manifestation of this process and a frequently overlooked symptom of HF. Orthopnea generally is relieved by sitting upright or sleeping with additional pillows. Although orthopnea is a relatively specific symptom of HF, it may occur in patients with abdominal obesity or ascites and patients with pulmonary disease whose lung mechanics favor an upright posture.

Paroxysmal Nocturnal Dyspnea (PND)

This term refers to acute episodes of severe shortness of breath and coughing that generally occur at night and awaken the patient from sleep, usually 1–3 hours after the patient retires. PND may be manifest by coughing or wheezing, possibly because of increased pressure in the bronchial arteries leading to airway compression, along with interstitial pulmonary edema that leads to increased airway resistance. Whereas orthopnea may be relieved by sitting upright at the side of the bed with the legs in a dependent position, patients with PND often have persistent coughing and wheezing even after they have assumed the upright position. *Cardiac asthma* is closely related to PND, is characterized by wheezing secondary to bronchospasm, and must be differentiated from primary asthma and pulmonary causes of wheezing.

Cheyne-Stokes Respiration

Also referred to as periodic respiration or cyclic respiration, Cheyne-Stokes respiration is present in 40% of patients with advanced HF and usually is associated with low cardiac output. Cheyne-Stokes respiration is caused by a diminished sensitivity of the respiratory center to arterial P_{CO2}. There is an apneic phase, during which arterial P_{O2} falls and arterial P_{CO2} rises. These changes in the arterial

blood gas content stimulate the depressed respiratory center, resulting in hyperventilation and hypocapnia, followed by recurrence of apnea. Cheyne-Stokes respirations may be perceived by the patient or the patient's family as severe dyspnea or as a transient cessation of breathing.

Acute Pulmonary Edema

See Chap. 272

OTHER SYMPTOMS

Patients with HF also may present with gastrointestinal symptoms. Anorexia, nausea, and early satiety associated with abdominal pain and fullness are common complaints and may be related to edema of the bowel wall and/or a congested liver. Congestion of the liver and stretching of its capsule may lead to right-upper-quadrant pain. Cerebral symptoms such as confusion, disorientation, and sleep and mood disturbances may be observed in patients with severe HF, particularly elderly patients with cerebral arteriosclerosis and reduced cerebral perfusion. Nocturia is common in HF and may contribute to insomnia.

Physical Examination

A careful physical examination is always warranted in the evaluation of patients with HF. The purpose of the examination is to help determine the cause of HF as well as to assess the severity of the syndrome. Obtaining additional information about the hemodynamic profile and the response to therapy and determining the prognosis are important additional goals of the physical examination.

GENERAL APPEARANCE AND VITAL SIGNS

In mild or moderately severe HF, the patient appears to be in no distress at rest except for feeling uncomfortable when lying flat for more than a few minutes. In more severe HF, the patient must sit upright, may have labored breathing, and may not be able to finish a sentence because of shortness of breath. Systolic blood pressure may be normal or high in early HF, but it generally is reduced in advanced HF because of severe LV dysfunction. The pulse pressure may be diminished, reflecting a reduction in stroke volume. Sinus tachycardia is a nonspecific sign caused by increased adrenergic activity. Peripheral vasoconstriction leading to cool peripheral extremities and cyanosis of the lips and nail beds is also caused by excessive adrenergic activity.

JUGULAR VEINS

(See also Chap. 227) Examination of the jugular veins provides an estimation of right atrial pressure. The jugular venous pressure is best appreciated with the patient lying recumbent, with the head tilted at 45°. The jugular venous pressure should be quantified in centimeters of water (normal ≤8 cm) by estimating the height of the venous column of blood above the sternal angle in centimeters and then adding 5 cm. In the early stages of HF, the venous pressure may be normal at rest but may become abnormally elevated with sustained (~1 min) pressure on the abdomen (positive abdominojugular reflux). Giant *v* waves indicate the presence of tricuspid regurgitation.

PULMONARY EXAMINATION

Pulmonary crackles (rales or crepitations) result from the transudation of fluid from the intravascular space into the alveoli. In patients with pulmonary edema, rales may be heard widely over both lung fields and may be accompanied by expiratory wheezing (cardiac asthma). When present in patients without concomitant lung disease, rales are specific for HF. Importantly, rales are frequently absent in patients with chronic HF, even when LV filling pressures are elevated, because of increased lymphatic drainage of alveolar fluid. Pleural effusions result from the elevation of pleural capillary pressure and the resulting transudation of fluid into the pleural cavities. Since the pleural veins drain into both the systemic and the pulmonary veins, pleural effusions occur most commonly with biventricular failure. Although pleural effusions are often bilateral in HF, when they are unilateral, they occur more frequently in the right pleural space.

CARDIAC EXAMINATION

Examination of the heart, although essential, frequently does not provide useful information about the severity of HF. If cardiomegaly is present, the point of maximal impulse (PMI) usually is displaced below the fifth intercostal space and/or lateral to the midclavicular line, and the impulse is palpable over two interspaces. Severe LV hypertrophy leads to a sustained PMI. In some patients, a third heart sound (S_3) is audible and palpable at the apex. Patients with enlarged or hypertrophied right ventricles may have a sustained and prolonged left parasternal impulse extending throughout systole. An S_3 (or *protodiastolic gallop*) is most commonly present in patients with volume overload who have tachycardia and tachypnea, and it often signifies severe hemodynamic compromise. A fourth heart sound (S_4) is not a specific indicator of HF but is usually present in patients with diastolic dysfunction. The murmurs of mitral and tricuspid regurgitation are frequently present in patients with advanced HF.

ABDOMEN AND EXTREMITIES

Hepatomegaly is an important sign in patients with HF. When it is present, the enlarged liver is frequently tender and may pulsate during systole if tricuspid regurgitation is present. Ascites, a late

sign, occurs as a consequence of increased pressure in the hepatic veins and the veins draining the peritoneum. Jaundice, also a late finding in HF, results from impairment of hepatic function secondary to hepatic congestion and hepatocellular hypoxemia and is associated with elevations of both direct and indirect bilirubin.

Peripheral edema is a cardinal manifestation of HF, but it is nonspecific and usually is absent in patients who have been treated adequately with diuretics. Peripheral edema is usually symmetric and dependent in HF and occurs predominantly in the ankles and the pretibial region in ambulatory patients. In bedridden patients, edema may be found in the sacral area (*presacral edema*) and the scrotum. Long-standing edema may be associated with indurated and pigmented skin.

CARDIAC CACHEXIA

With severe chronic HF, there may be marked weight loss and cachexia. Although the mechanism of cachexia is not entirely understood, it is probably multifactorial and includes elevation of the resting metabolic rate; anorexia, nausea, and vomiting due to congestive hepatomegaly and abdominal fullness; elevation of circulating concentrations of cytokines such as TNF; and impairment of intestinal absorption due to congestion of the intestinal veins. When present, cachexia augurs a poor overall prognosis.

Diagnosis

The diagnosis of HF is relatively straightforward when the patient presents with classic signs and symptoms of HF; however, the signs and symptoms of HF are neither specific nor sensitive. Accordingly, the key to making the diagnosis is to have a high index of suspicion, particularly for high-risk patients. When these patients present with signs or symptoms of HF, additional laboratory testing should be performed.

ROUTINE LABORATORY TESTING

Patients with new-onset HF and those with chronic HF and acute decompensation should have a complete blood count, a panel of electrolytes, blood urea nitrogen, serum creatinine, hepatic enzymes, and a urinalysis. Selected patients should have assessment for diabetes mellitus (fasting serum glucose or oral glucose tolerance test), dyslipidemia (fasting lipid panel), and thyroid abnormalities (thyroid-stimulating hormone level).

ELECTROCARDIOGRAM (ECG)

A routine 12-lead ECG is recommended. The major importance of the ECG is to assess cardiac rhythm and determine the presence of LV hypertrophy or a prior MI (presence or absence of Q waves) as well as to determine QRS width to ascertain whether the patient may benefit from resynchronization therapy (see below). A normal ECG virtually excludes LV systolic dysfunction.

CHEST X-RAY

A chest x-ray provides useful information about cardiac size and shape, as well as the state of the pulmonary vasculature, and may identify noncardiac causes of the patient's symptoms. Although patients with acute HF have evidence of pulmonary hypertension, interstitial edema, and/or pulmonary edema, the majority of patients with chronic HF do not. The absence of these findings in patients with chronic HF reflects the increased capacity of the lymphatics to remove interstitial and/or pulmonary fluid.

ASSESSMENT OF LV FUNCTION

Noninvasive cardiac imaging (Chap. 229) is essential for the diagnosis, evaluation, and management of HF. The most useful test is the two-dimensional (2-D) echocardiogram/Doppler, which can provide a semiquantitative assessment of LV size and function as well as the presence or absence of valvular and/or regional wall motion abnormalities (indicative of a prior MI). The presence of left atrial dilation and LV hypertrophy, together with abnormalities of LV diastolic filling provided by pulse-wave and tissue Doppler, is useful for the assessment of HF with a preserved EF. The 2-D echocardiogram/Doppler is also invaluable in assessing RV size and pulmonary pressures, which are critical in the evaluation and management of cor pulmonale (see below). Magnetic resonance imaging (MRI) also provides a comprehensive analysis of cardiac anatomy and function and is now the gold standard for assessing LV mass and volumes. MRI also is emerging as a useful and accurate imaging modality for evaluating patients with HF, both in terms of assessing LV structure and for determining the cause of HF (e.g., amyloidosis, ischemic cardiomyopathy, hemochromatosis).

The most useful index of LV function is the EF (stroke volume divided by end-diastolic volume). Because the EF is easy to measure by noninvasive testing and easy to conceptualize, it has gained wide acceptance among clinicians. Unfortunately, the EF has a number of limitations as a true measure of contractility, since it is influenced by alterations in afterload and/or preload. Nonetheless, with the exceptions indicated above, when the EF is normal (≥50%), systolic function is usually adequate, and when the EF is significantly depressed (<30–40%), contractility is usually depressed.

BIOMARKERS

Circulating levels of natriuretic peptides are useful adjunctive tools in the diagnosis of patients with HF.

Both B-type natriuretic peptide (BNP) and N-terminal pro-BNP, which are released from the failing heart, are relatively sensitive markers for the presence of HF with depressed EF; they also are elevated in HF patients with a preserved EF, albeit to a lesser degree. However, it is important to recognize that natriuretic peptide levels increase with age and renal impairment, are more elevated in women, and can be elevated in right HF from any cause. Levels can be falsely low in obese patients and may normalize in some patients after appropriate treatment. At present, serial measurements of BNP are not recommended as a guide to HF therapy. Other biomarkers, such as troponin T and I, C-reactive protein, TNF receptors, and uric acid, may be elevated in HF and provide important prognostic information. Serial measurements of one or more biomarkers ultimately may help guide therapy in HF, but they are not currently recommended for this purpose.

EXERCISE TESTING

Treadmill or bicycle exercise testing is not routinely advocated for patients with HF, but either is useful for assessing the need for cardiac transplantation in patients with advanced HF (Chap. 235). A peak oxygen uptake (V_{O2}) <14 mL/kg per min is associated with a relatively poor prognosis. Patients with a V_{O2}<14 mL/kg per min have been shown, in general, to have better survival when transplanted than when treated medically.

Differential Diagnosis

HF resembles but should be distinguished from (1) conditions in which there is circulatory congestion secondary to abnormal salt and water retention but in which there is no disturbance of cardiac structure or function (e.g., renal failure) and (2) noncardiac causes of pulmonary edema (e.g., acute respiratory distress syndrome). In most patients who present with classic signs and symptoms of HF, the diagnosis is relatively straightforward. However, even experienced clinicians have difficulty differentiating the dyspnea that arises from cardiac and pulmonary causes (Chap. 33). In this regard, noninvasive cardiac imaging, biomarkers, pulmonary function testing, and chest x-ray may be useful. A very low BNP or N-terminal pro-BNP may be helpful in excluding a cardiac cause of dyspnea in this setting. Ankle edema may arise secondary to varicose veins, obesity, renal disease, or gravitational effects. When HF develops in patients with a preserved EF, it may be difficult to determine the relative contribution of HF to the dyspnea that occurs in chronic lung disease and/or obesity.

Treatment: Heart Failure

HF should be viewed as a continuum that is composed of four interrelated stages. *Stage A* includes patients who are at high risk for developing HF but do not have structural heart disease or symptoms of HF (e.g., patients with diabetes mellitus or hypertension). *Stage B* includes patients who have structural heart disease but do not have symptoms of HF (e.g., patients with a previous MI and asymptomatic LV dysfunction). *Stage C* includes patients who have structural heart disease and have developed symptoms of HF (e.g., patients with a previous MI with dyspnea and fatigue). *Stage D* includes patients with refractory HF requiring special interventions (e.g., patients with refractory HF who are awaiting cardiac transplantation). In this continuum, every effort should be made to prevent HF not only by treating the preventable causes of HF (e.g., hypertension) but also by treating the patient in stages B and C with drugs that prevent disease progression (e.g., ACE inhibitors and beta blockers) and by symptomatic management of patients in stage D.

DEFINING AN APPROPRIATE THERAPEUTIC STRATEGY FOR CHRONIC HF

Once patients have developed structural heart disease, their therapy depends on their NYHA functional classification **(**Table 234-2**)**. Although this classification system is notoriously subjective and has large interobserver variability, it has withstood the test of time and continues to be widely applied to patients with HF. For patients who have developed LV systolic dysfunction but remain asymptomatic (class I), the goal should be to slow disease progression by blocking neurohormonal systems that lead to cardiac remodeling (see below). For patients who have developed symptoms (class II–IV), the primary goal should be to alleviate fluid retention, lessen disability, and reduce the risk of further disease progression and death. These goals generally require a strategy that combines diuretics (to control salt and water retention) with neurohormonal interventions (to minimize cardiac remodeling).

MANAGEMENT OF HF WITH DEPRESSED EJECTION FRACTION (<40%)

General Measures

Clinicians should aim to screen for and treat comorbidities such as hypertension, CAD, diabetes mellitus, anemia, and sleep-disordered breathing, as these conditions tend to exacerbate HF. HF patients should be advised to stop smoking and to limit alcohol consumption to two standard drinks per day in men or one per day in women. Patients suspected of having an alcohol-induced cardiomyopathy should be urged to abstain from alcohol consumption indefinitely. Extremes of temperature and heavy physical exertion should be avoided. Certain drugs are known to make HF worse and should be avoided **(Table 234-4)**. For example, nonsteroidal anti-inflammatory drugs, including cyclooxygenase 2 inhibitors, are not recommended in patients with chronic HF because the risk of renal failure and fluid retention is markedly increased in the presence of reduced renal function or ACE inhibitor therapy. Patients should receive immunization with influenza and pneumococcal

vaccines to prevent respiratory infections. It is equally important to educate the patient and family about HF, the importance of proper diet, and the importance of compliance with the medical regimen. Supervision of outpatient care by a specially trained nurse or physician assistant and/or in specialized HF clinics has been found to be helpful, particularly in patients with advanced disease.

Table 234-4 Factors That May Precipitate Acute Decompensation in Patients with Chronic Heart Failure

Dietary indiscretion
Myocardial ischemia/infarction
Arrhythmias (tachycardia or bradycardia)
Discontinuation of HF therapy
Infection
Anemia
Initiation of medications that worsen HF
Calcium antagonists (verapamil, diltiazem)
Beta blockers
Nonsteroidal anti-inflammatory drugs
Antiarrhythmic agents [all class I agents, sotalol (class III)]
Anti-TNF antibodies
Alcohol consumption
Pregnancy
Worsening hypertension
Acute valvular insufficiency

Abbreviations: HF, heart failure; TNF, tumor necrosis factor.

Activity

Although heavy physical labor is not recommended in HF, routine modest exercise has been shown to be beneficial in patients with NYHA class I–III HF. For euvolemic patients, regular isotonic exercise such as walking or riding a stationary-bicycle ergometer, as tolerated, should be encouraged. Exercise training results in reduced HF symptoms, increased exercise capacity, and improved quality of life.

Diet

Dietary restriction of sodium (2–3 g daily) is recommended in all patients with HF and preserved or depressed EF. Further restriction (<2 g daily) may be considered in moderate to severe HF. Fluid restriction is generally unnecessary unless the patient develops hyponatremia (<130 meq/L), which may develop because of activation of the renin-angiotensin system, excessive secretion of antidiuretic hormone, or loss of salt in excess of water from diuretic use. Fluid restriction (<2 L/day) should be considered in hyponatremic patients or those whose fluid retention is difficult to control despite high doses of diuretics and sodium restriction. Vasopressin antagonists also may be useful in severe hyponatremia. Caloric supplementation is recommended for patients with advanced HF and unintentional weight loss or muscle wasting (cardiac cachexia); however, anabolic steroids are not recommended for these patients because of the potential problems with volume retention. The use of dietary supplements ("nutriceuticals") should be avoided in the management of symptomatic HF because of the lack of proven benefit and the potential for significant (adverse) interactions with proven HF therapies.

Diuretics

Many of the clinical manifestations of moderate to severe HF result from excessive salt and water retention that leads to volume expansion and congestive symptoms. Diuretics **(Table 234-5)** are the only pharmacologic agents that can adequately control fluid retention in advanced HF, and they should be used to restore and maintain normal volume status in patients with congestive symptoms (dyspnea, orthopnea, edema) or signs of elevated filling pressures (rales, jugular venous distention, peripheral edema). Furosemide, torsemide, and bumetanide act at the loop of Henle (*loop diuretics*) by reversibly inhibiting the reabsorption of Na^+, K^+, and Cl^- in the thick ascending limb of Henle's loop; thiazides and metolazone reduce the reabsorption of Na^+ and Cl^- in the first half of the distal convoluted tubule; and potassium-sparing diuretics such as spironolactone act at the level of the collecting duct.

Table 234-5 Drugs for the Treatment of Chronic Heart Failure (EF <40%)

	Initiating Dose	Maximal Dose

Diuretics		
Furosemide	20–40 mg qd or bid	400 mg/d[a]
Torsemide	10–20 mg qd bid	200 mg/d[a]
Bumetanide	0.5–1 mg qd or bid	10 mg/d[a]
Hydrochlorthiazide	25 mg qd	100 mg/d[a]
Metolazone	2.5–5 mg qd or bid	20 mg/d[a]
Angiotensin-Converting Enzyme Inhibitors		
Captopril	l6.25 mg tid	50 mg tid
Enalapril	2.5 mg bid	10 mg bid
Lisinopril	2.5–5 mg qd	20–35 mg qd
Ramipril	1.25–2.5 mg bid	2.5–5 mg bid
Trandolapri	0.5 mg qd	4 mg qd
Angiotensin Receptor Blockers		
Valsartan	40 mg bid	160 mg bid
Candesartan	4 mg qd	32 mg qd
Irbesartan	75 mg qd	300 mg qd[b]
Losartan	12.5 mg qd	50 mg qd
β **Receptor Blockers**		
Carvedilol	3.125 mg bid	25–50 mg bid
Bisoprolol	1.25 mg qd	10 mg qd
Metoprolol succinate CR	12.5–25 mg qd	Target dose 200 mg qd
Additional Therapies		
Spironolactone	12.5–25 mg qd	25–50 mg qd
Eplerenone	25 mg qd	50 mg qd
Combination of hydralazine/isosorbide dinitrate	10–25 mg/10 mg tid	75 mg/40 mg tid
Fixed dose of hydralazine/isosorbide dinitrate	37.5 mg/20 mg (one tablet) tid	75 mg/40 mg (two tablets) tid
Digoxin	0.125 mg qd	≤0.375 mg/d[b]

Notes: [a] Dose must be titrated to reduce the patient's congestive symptoms.

[b] Target dose not established.

Although all diuretics increase sodium excretion and urinary volume, they differ in their potency and pharmacologic properties. Whereas loop diuretics increase the fractional excretion of sodium by 20–25%, thiazide diuretics increase it by only 5–10% and tend to lose their effectiveness in patients with moderate or severe renal insufficiency (creatinine >2.5 mg/dL). Hence, loop diuretics generally are required to restore normal volume status in patients with HF. Diuretics should be initiated in low doses (Table 234-5) and then carefully titrated upward to relieve signs and symptoms of fluid overload in an attempt to obtain the patient's "dry weight." This typically requires multiple dose adjustments over many days and occasionally weeks in patients with severe fluid overload. Intravenous administration of diuretics may be necessary to relieve congestion acutely and can be done safely in the outpatient setting. Once the congestion has been relieved, treatment with diuretics should be continued to prevent the recurrence of salt and water retention.

Refractoriness to diuretic therapy may represent patient nonadherence, a direct effect of chronic diuretic use on the kidney, or progression of underlying HF. The addition of thiazides or metolazone, once or twice daily, to loop diuretics may be considered in patients with persistent fluid retention despite high-dose loop diuretic therapy. Metolazone is generally more potent and much longer-acting than the thiazides in this setting as well as in patients with chronic renal insufficiency. However, chronic daily use, especially of metolazone, should be avoided if possible because of the potential for electrolyte shifts and volume depletion. Ultrafiltration and dialysis may be used in cases of refractory

fluid retention that are unresponsive to high doses of diuretics and have been shown to be helpful in the short term.

Adverse Effects

Diuretics have the potential to produce electrolyte and volume depletion as well as worsening azotemia. In addition, they may lead to worsening neurohormonal activation and disease progression. One of the most important adverse consequences of diuresis is alterations in potassium homeostasis (hypokalemia or hyperkalemia), which increase the risk of life-threatening arrhythmias. In general, both loop- and thiazide-type diuretics lead to hypokalemia, whereas spironolactone, eplerenone, and triamterene lead to hyperkalemia.

PREVENTING DISEASE PROGRESSION (TABLE 234-5)

Drugs that interfere with excessive activation of the RAA system and the adrenergic nervous system can relieve the symptoms of HF with a depressed EF by stabilizing and/or reversing cardiac remodeling. In this regard, ACE inhibitors and beta blockers have emerged as the cornerstones of modern therapy for HF with a depressed EF.

ACE Inhibitors

There is overwhelming evidence that ACE inhibitors should be used in symptomatic and asymptomatic patients **(Figs. 234-3** and **234-4)** with a depressed EF (<40%). ACE inhibitors interfere with the renin-angiotensin system by inhibiting the enzyme that is responsible for the conversion of angiotensin I to angiotensin II. However, because ACE inhibitors also inhibit kininase II, they may lead to the upregulation of bradykinin, which may further enhance the beneficial effects of angiotensin suppression. ACE inhibitors stabilize LV remodeling, improve symptoms, reduce hospitalization, and prolong life. Because fluid retention can attenuate the effects of ACE inhibitors, it is preferable to optimize the dose of diuretic before starting the ACE inhibitor. However, it may be necessary to reduce the dose of diuretic during the initiation of ACE inhibition to prevent symptomatic hypotension. ACE inhibitors should be initiated in low doses, followed by gradual increments if the lower doses have been well tolerated. The doses of ACE inhibitors should be increased until they are similar to those which have been shown to be effective in clinical trials (Table 234-5). Higher doses are more effective than lower doses in preventing hospitalization.

Figure 234-3

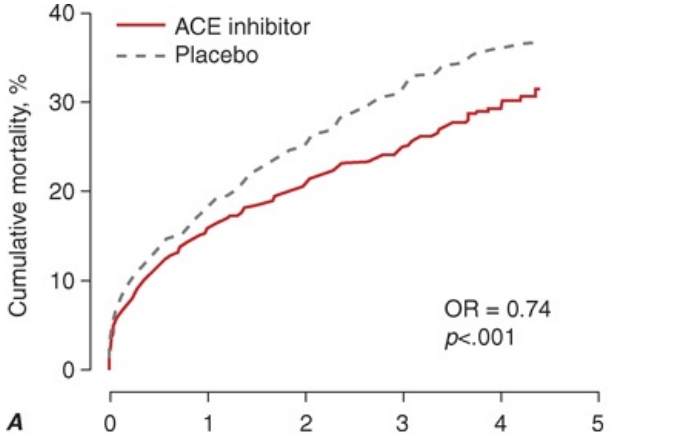

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

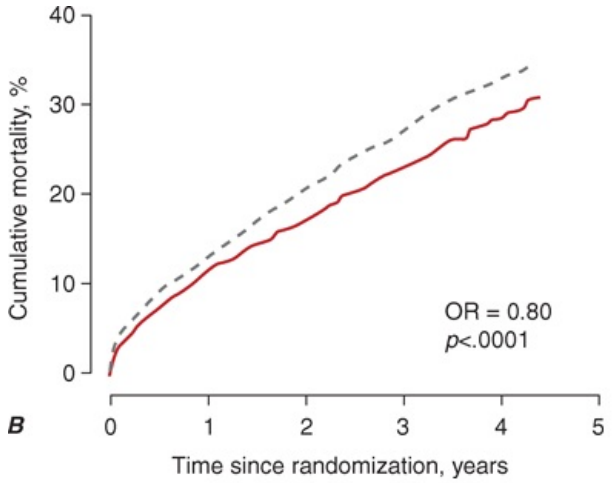

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Meta-analysis of angiotensin-converting enzyme (ACE) inhibitors in heart failure patients with a depressed ejection fraction.

A. The Kaplan-Meier curves for mortality for 5966 HF patients with a depressed EF treated with an ACE inhibitor after acute myocardial infarction (three trials).

B. The Kaplan-Meier curves for mortality for 12,763 HF patients with a depressed EF treated with an ACE inhibitor in five clinical trials, including postinfarction trials. The benefits of ACE inhibitors were observed early and persisted long-term. *(Modified from MD Flather et al: Lancet 355:1575, 2000.)*

Figure 234-4

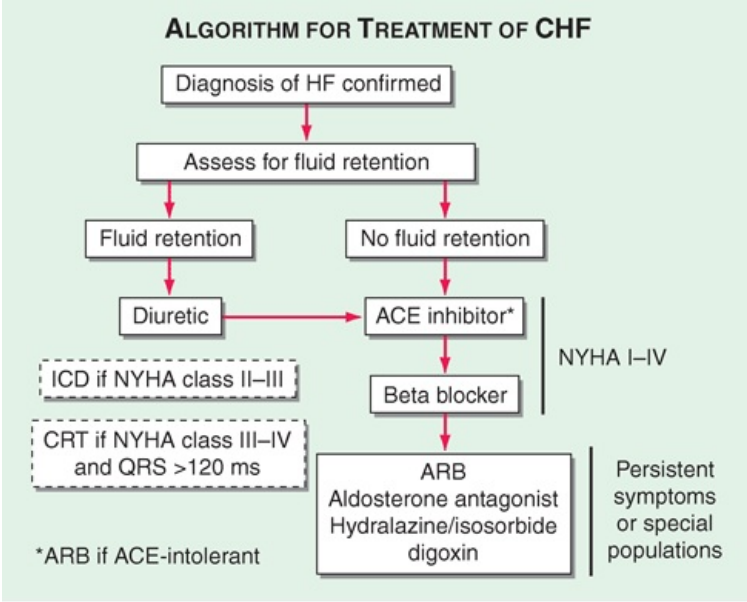

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment algorithm for chronic heart failure patients with a depressed ejection fraction. After the clinical diagnosis of HF is made, it is important to treat the patient's fluid retention before starting an ACE inhibitor (or an ARB if the patient is ACE-intolerant). Beta blockers should be started after the fluid retention has been treated and/or the ACE inhibitor has been uptitrated. If the patient remains symptomatic, an ARB, an aldosterone antagonist, or digoxin can be added as "triple therapy." The fixed-dose combination of hydralazine/isosorbide dinitrate should be added to an ACE inhibitor and a beta blocker in African-American patients with NYHA class II–IV HF. Device therapy should be considered in addition to pharmacologic therapy in appropriate patients. HF, heart failure; ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; NYHA, New York Heart Association; CRT, cardiac resynchronization therapy; ICD, implantable cardiac defibrillator.

Adverse Effects

The majority of adverse effects are related to suppression of the renin-angiotensin system. The decreases in blood pressure and mild azotemia that may occur during the initiation of therapy generally

are well tolerated and do not require a decrease in the dose of the ACE inhibitor. However, if hypotension is accompanied by dizziness or if the renal dysfunction becomes severe, it may be necessary to reduce the dose of the inhibitor. Potassium retention may also become problematic if the patient is receiving potassium supplements or a potassium-sparing diuretic. Potassium retention that is not responsive to these measures may require a reduction in the dose of ACE inhibitor.

The side effects of ACE inhibitors related to kinin potentiation include a nonproductive cough (10–15% of patients) and angioedema (1% of patients). In patients who cannot tolerate ACE inhibitors because of cough or angioedema, angiotensin receptor blockers (ARBs) are the recommended first line of therapy (see below). Patients intolerant of ACE inhibitors because of hyperkalemia or renal insufficiency are likely to experience the same side effects with ARBs. In these cases, the combination of hydralazine and an oral nitrate should be considered (Table 234-5).

Angiotensin Receptor Blockers

These drugs are well tolerated in patients who are intolerant of ACE inhibitors because of cough, skin rash, and angioedema. ARBs should be used in symptomatic and asymptomatic patients with an EF <40% who are ACE-intolerant for reasons other than hyperkalemia or renal insufficiency (Table 234-5). Although ACE inhibitors and ARBs inhibit the renin-angiotensin system, they do so by different mechanisms. Whereas ACE inhibitors block the enzyme responsible for converting angiotensin I to angiotensin II, ARBs block the effects of angiotensin II on the angiotensin type 1 receptor. Some clinical trials have demonstrated a therapeutic benefit from the addition of an ARB to an ACE inhibitor in patients with chronic HF. When given in concert with beta blockers, ARBs reverse the process of LV remodeling, improve patient symptoms, prevent hospitalization, and prolong life.

Adverse Effects

Both ACE inhibitors and ARBs have similar effects on blood pressure, renal function, and potassium. Therefore, the problems of symptomatic hypotension, azotemia, and hyperkalemia are similar for both of these agents.

β-Adrenergic Receptor Blockers

Beta-blocker therapy represents a major advance in the treatment of patients with a depressed EF **(Fig. 234-5)**. These drugs interfere with the harmful effects of sustained activation of the adrenergic nervous system by competitively antagonizing one or more adrenergic receptors (α_1, β_1, and β_2). Although there are a number of potential benefits to blocking all three receptors, most of the deleterious effects of adrenergic activation are mediated by the β_1 receptor. When given in concert with ACE inhibitors, beta blockers reverse the process of LV remodeling, improve patient symptoms, prevent hospitalization, and prolong life. Therefore, beta blockers are indicated for patients with symptomatic or asymptomatic HF and a depressed EF <40%.

Figure 234-5

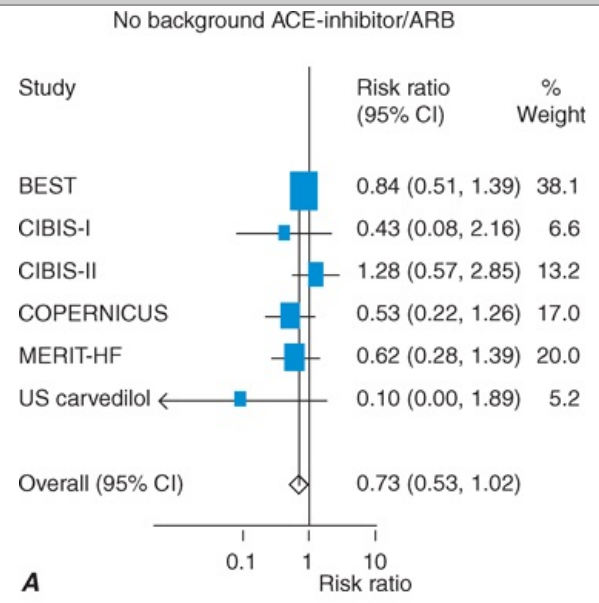

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

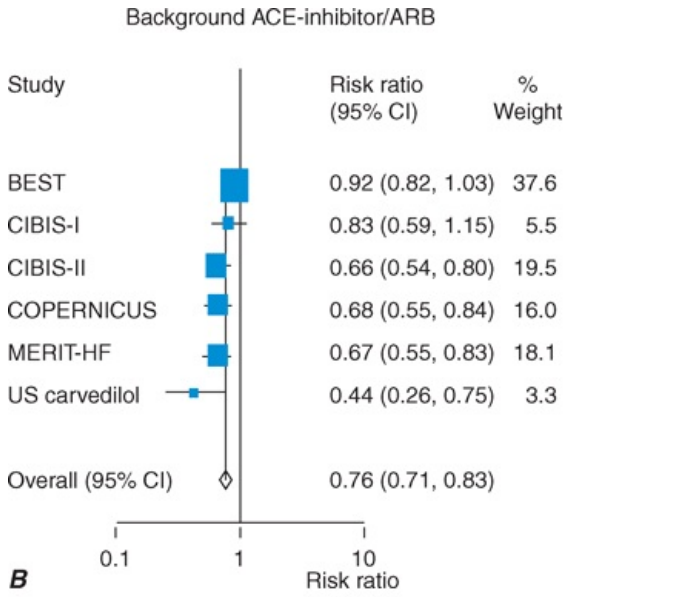

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Meta-analysis of beta blockers on mortality rates in HF patients with a depressed EF. Effect of beta blockers vs. placebo in patients who were not (***A***) or who were (***B***) receiving an angiotensin-converting enzyme (ACE) inhibitor or an angiotensin receptor blocker (ARB) at baseline in six clinical trials. There was a similar impact of beta-blocker therapy on the endpoints of all-cause mortality as well as death and heart failure hospitalization in both the presence and the absence of ACE inhibitor or ARB at baseline. BEST, Beta-Blocker Evaluation of Survival Trial (bucindolol); CIBIS, Cardiac Insufficiency Bisoprolol Study (bisoprolol); COPERNICUS, Carvedilol prOsPEctive RaNdomIzed Cumulative Survival (carvedilol); MERIT-HF, Metoprolol CR/XL Randomized Intervention Trial in Heart Failure (metoprolol CR/XL). *(Modified from H Krum et al: Eur Heart J 26:2154, 2005.)*

Analogous to the use of ACE inhibitors, beta blockers should be initiated in low doses (Table 234-5), followed by gradual increments in the dose if lower doses have been well tolerated. The dose of beta blocker should be increased until the doses used are similar to those which have been reported to be effective in clinical trials (Table 234-5). However, unlike ACE inhibitors, which may be titrated upward relatively rapidly, the titration of beta blockers should proceed no more rapidly than at 2-week intervals, because the initiation and/or increased dosing of these agents may lead to worsening fluid retention consequent to the withdrawal of adrenergic support to the heart and the circulation. Thus, it is important to optimize the dose of diuretic before starting therapy with beta blockers. If worsening fluid retention does occur, it is likely to do so within 3–5 days of the initiation of therapy, and it will be manifest as an increase in body weight and/or symptoms of worsening HF. The increased fluid retention usually can be managed by increasing the dose of diuretics. In some patients the dose of the beta blocker may have to be reduced.

Contrary to early reports, the aggregate results of clinical trials suggest that beta-blocker therapy is well tolerated by the great majority (≥85%) of HF patients, including patients with comorbid conditions such as diabetes mellitus, chronic obstructive lung disease, and peripheral vascular disease. Nonetheless, there is a subset of patients (10–15%) who remain intolerant to beta blockers because of worsening fluid retention or symptomatic hypotension or bradycardia.

Adverse Effects

The adverse effects of beta-blocker use generally are related to the predictable complications that arise from interfering with the adrenergic nervous system. These reactions generally occur within several days of the initiation of therapy and generally are responsive to adjustments of concomitant medications, as described above. Therapy with beta blockers can lead to bradycardia and/or exacerbate heart block. Accordingly, the dose of beta blocker should be reduced if the heart rate decreases to <50 beats/min and/or second- or third-degree heart block or symptomatic hypotension develops. Beta blockers are not recommended for patients who have asthma with active bronchospasm. Beta blockers that also block the α_1 receptor can lead to vasodilatory side effects.

Aldosterone Antagonists

Although classified as potassium-sparing diuretics, drugs that block the effects of aldosterone (spironolactone or eplerenone) have beneficial effects that are independent of the effects of these agents on sodium balance. Although ACE inhibition may transiently decrease aldosterone secretion, with chronic therapy there is a rapid return of aldosterone to levels similar to those before ACE inhibition. Accordingly, the administration of an aldosterone antagonist is recommended for patients with NYHA class IV or class III (previously class IV) HF who have a depressed EF (<35%) and are receiving standard therapy, including diuretics, ACE inhibitors, and beta blockers. The dose of

aldosterone antagonist should be increased until the doses used are similar to those which have been shown to be effective in clinical trials (Table 234-5).

Adverse Effects

The major problem with the use of aldosterone antagonists is the development of life-threatening hyperkalemia, which is more prone to occur in patients who are receiving potassium supplements or who have underlying renal insufficiency. Aldosterone antagonists are not recommended when the serum creatinine is >2.5 mg/dL (or creatinine clearance is <30 mL/min) or when the serum potassium is >5 mmol/L. Painful gynecomastia may develop in 10–15% of patients who use spironolactone, in which case eplerenone may be substituted.

SPECIAL POPULATIONS

The combination of hydralazine and isosorbide dinitrate (Table 234-5) is recommended as part of standard therapy in addition to beta blockers and ACE inhibitors for African Americans with NYHA class II–IV HF. Although the exact mechanism for the effects of this combination is not known, it is believed to be secondary to the beneficial effects of NO on the peripheral circulation.

MANAGEMENT OF PATIENTS WHO REMAIN SYMPTOMATIC

Additional pharmacologic therapy should be considered in patients who have persistent symptoms or progressive worsening despite optimized therapy with an ACE inhibitor and a beta blocker. Agents that may be considered as part of additional therapy include an ARB, spironolactone, the combination of hydralazine and isosorbide dinitrate, and digitalis. The optimal choice of additional drug therapy to improve the outcome further has not been firmly established. Thus, the choice of a specific agent will be influenced by clinical considerations, including renal function, serum potassium concentration, blood pressure, and race. The triple combination of an ACE inhibitor, an ARB, and an aldosterone antagonist should not be used because of the high risk of hyperkalemia.

Digoxin is recommended for patients with symptomatic LV systolic dysfunction who have concomitant atrial fibrillation, and it should be considered for patients who have signs or symptoms of HF while receiving standard therapy, including ACE inhibitors and beta blockers. Therapy with digoxin is commonly initiated and maintained at a dose of 0.125–0.25 mg daily. For the great majority of patients, the dose should be 0.125 mg daily, and the serum digoxin level should be <1 ng/mL, especially in elderly patients, patients with impaired renal function, and patients with a low lean body mass. Higher doses (and serum concentrations) appear to be less beneficial. There is no indication for using loading doses of digoxin to initiate therapy in patients with HF.

ANTICOAGULATION AND ANTIPLATELET THERAPY

Patients with HF have an increased risk for arterial or venous thromboembolic events. In clinical HF trials, the rate of stroke ranges from 1.3 to 2.4% per year. Depressed LV function is believed to promote relative stasis of blood in dilated cardiac chambers with increased risk of thrombus formation. Treatment with warfarin [goal international normalized ratio (INR) 2–3] is recommended for patients with HF and chronic or paroxysmal atrial fibrillation or with a history of systemic or pulmonary emboli, including stroke or transient ischemic attack. Patients with symptomatic or asymptomatic ischemic cardiomyopathy and documented recent large anterior MI or recent MI with documented LV thrombus should be treated with warfarin (goal INR 2–3) for the initial 3 months after the MI unless there are contraindications to its use.

Aspirin is recommended in HF patients with ischemic heart disease for the prevention of MI and death. However, lower doses of aspirin (75 or 81 mg) may be preferable because of the concern of worsening of HF at higher doses.

MANAGEMENT OF CARDIAC ARRHYTHMIAS

(See also Chap. 233) Atrial fibrillation occurs in 15–30% of patients with HF and is a common cause of cardiac decompensation. Most antiarrhythmic agents, with the exception of amiodarone and dofetilide, have negative inotropic effects and are proarrhythmic. Amiodarone is a class III antiarrhythmic that has few or no negative inotropic and/or proarrhythmic effects and is effective against most supraventricular arrhythmias. Amiodarone is the preferred drug for restoring and maintaining sinus rhythm, and it may improve the success of electrical cardioversion in patients with HF. Amiodarone increases the level of phenytoin and digoxin and prolongs the INR in patients taking warfarin. Therefore, it is often necessary to reduce the dose of these drugs by as much as 50% when initiating therapy with amiodarone. The risk of adverse events such as hyperthyroidism, hypothyroidism, pulmonary fibrosis, and hepatitis is relatively low, particularly when lower doses of amiodarone are used (100–200 mg/d).

Implantable cardiac defibrillators (ICDs; see below) are highly effective in treating recurrences of sustained ventricular tachycardia and/or ventricular fibrillation in HF patients with recurrent arrhythmias and/or cardiac syncope, and they may be used as stand-alone therapy or in combination with amiodarone and/or a beta blocker (Chap. 233). There is no role for treating ventricular arrhythmias with an antiarrhythmic agent without an ICD.

DEVICE THERAPY

Cardiac Resynchronization

Approximately one-third of patients with a depressed EF and symptomatic HF (NYHA class III–IV) manifest a QRS duration >120 ms. This ECG finding of abnormal inter- or intraventricular conduction has been used to identify patients with dyssynchronous ventricular contraction. The mechanical consequences of ventricular dyssynchrony include suboptimal ventricular filling, a reduction in LV contractility, prolonged duration (and therefore greater severity) of mitral regurgitation, and paradoxical septal wall motion. *Biventricular pacing* , also termed *cardiac resynchronization therapy* (CRT), stimulates both ventricles nearly simultaneously, thereby improving the coordination of ventricular contraction and reducing the severity of mitral regurgitation. When CRT is added to optimal medical therapy in patients in sinus rhythm, there is a significant decrease in patient mortality rates and hospitalization and a reversal of LV remodeling, as well as improved quality of life and exercise capacity. Accordingly, CRT is recommended for patients in sinus rhythm with an EF <35% and a QRS >120 ms and those who remain symptomatic (NYHA III–IV) despite optimal medical therapy. The benefits of CRT in patients with atrial fibrillation have not been clearly established.

Implantable Cardiac Defibrillators

(See also Chap. 233) The prophylactic implantation of ICDs in patients with mild to moderate HF (NYHA class II–III) has been shown to reduce the incidence of sudden cardiac death in patients with ischemic or nonischemic cardiomyopathy. Accordingly, implantation of an ICD should be considered for patients in NYHA class II–III HF with a depressed EF of <35% who are already on optimal background therapy, including an ACE inhibitor (or ARB), a beta blocker, and an aldosterone antagonist. An ICD may also be combined with a biventricular pacemaker in patients with NYHA class III–IV HF.

MANAGEMENT OF HF WITH A PRESERVED EJECTION FRACTION (>40–50%)

Despite the wealth of information with respect to the evaluation and management of HF with a depressed EF, there are no proven and/or approved pharmacologic or device therapies for the management of patients with HF and a preserved EF. Therefore, it is recommended that initial treatment efforts should be focused, wherever possible, on the underlying disease process (e.g., myocardial ischemia, hypertension) associated with HF with preserved EF. Precipitating factors such as tachycardia and atrial fibrillation should be treated as quickly as possible through rate control and restoration of sinus rhythm when appropriate. Dyspnea may be treated by reducing total blood volume (dietary sodium restriction and diuretics), decreasing central blood volume (nitrates), or blunting neurohormonal activation with ACE inhibitors, ARBs, and/or beta blockers. Treatment with diuretics and nitrates should be initiated at low doses to avoid hypotension and fatigue.

ACUTE DECOMPENSATED HF

Defining an Appropriate Therapeutic Strategy

The therapeutic goals for the management of acute decompensated HF (ADHF) therapy are to (1) stabilize the hemodynamic derangements that provoked the symptoms responsible for the hospitalization, (2) identify and treat the reversible factors that precipitated decompensation, and (3) reestablish an effective outpatient medical regimen that will prevent disease progression and relapse. In most instances this will require hospitalization, often in an intensive care unit (ICU) setting. Every effort should be made to identify the precipitating causes, such as infection, arrhythmias, dietary indiscretion, pulmonary embolism, infective endocarditis, occult myocardial ischemia/infarction, and environmental and/or emotional stress (Table 234-4), since removal of these precipitating events is critical to the success of treatment.

The two primary hemodynamic determinants of ADHF are elevated LV filling pressures and a depressed cardiac output. Frequently the depressed cardiac output is accompanied by an increase in systemic vascular resistance (SVR) as a result of excessive neurohormonal activation. Because these hemodynamic derangements may occur singly or together, patients with acute HF generally present with one of four basic hemodynamic profiles **(Fig. 234-6)**: normal LV filling pressure with normal perfusion (Profile A), elevated LV filling pressure with normal perfusion (Profile B), elevated LV filling pressures with decreased perfusion (Profile C), and normal or low LV filling pressure with decreased tissue perfusion (Profile L).

Figure 234-6

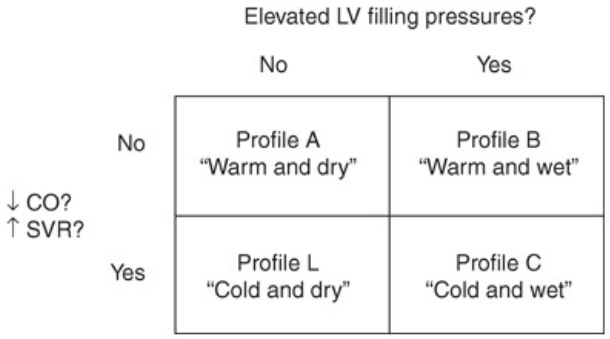

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hemodynamic profiles in patients with acute heart failure. Most patients can be categorized into one of the four hemodynamic profiles by performing a brief bedside examination that includes examination of the neck veins, lungs, and peripheral extremities. More definitive hemodynamic information may be obtained by performing invasive hemodynamic monitoring, particularly if the patient is gravely ill or if the clinical presentation is unclear. This hemodynamic classification provides a useful guide for selecting the initial optimal therapies for the management of acute HF. LV, left ventricular; CO, cardiac output; SVR, systemic vascular resistance. *(Modified from Grady et al: Circulation 102:2443, 2000.)*

Accordingly, the therapeutic approach to treating patients with acute HF should be tailored to reflect the patient's hemodynamic presentation. The goal should be, whenever possible, to restore the patient to a normal hemodynamic profile (Profile A). In many instances the patient's hemodynamic presentation can be approximated from the clinical examination. For example, patients with elevated LV filling pressures may have signs of fluid retention (rales, elevated neck veins, peripheral edema) and are referred to as being "wet," whereas patients with a depressed cardiac output and an elevated SVR generally have poor tissue perfusion manifested by cool distal extremities and are referred to as being "cold." Nonetheless, it should be emphasized that patients with chronic heart failure may not have rales or evidence of peripheral edema at the time of the initial presentation with acute decompensation, and this may lead to the underrecognition of elevated filling pressures. In these patients, it may be appropriate to perform invasive hemodynamic monitoring.

Patients who are not congested and have normal tissue perfusion are referred to as being "dry" and "warm," respectively. When acute HF patients present to the hospital with Profile A, their symptoms are often due to conditions other than HF (e.g., pulmonary or hepatic disease or transient myocardial ischemia). More commonly, however, acute HF patients present with congestive symptoms ["warm and wet" (Profile B)], in which case treatment of the elevated filling pressures with diuretics and vasodilators is warranted to reduce LV filling pressures. Profile B includes most patients with acute pulmonary edema. **The treatment of this life-threatening condition is described in Chap. 272.**

Patients also may present with congestion and a significantly elevated SVR and reduction of cardiac output ["cold and wet" (Profile C)]. In these patients, cardiac output can be increased and LV filling pressures reduced by using intravenous vasodilators. Intravenous inotropic agents with vasodilating action [dobutamine, low-dose dopamine, milrinone **(Table 234-6)**] augment cardiac output by stimulating myocardial contractility as well as by functionally unloading the heart.

Table 234-6 Drugs for the Treatment for Acute Heart Failure

	Initiating Dose	Maximal Dose
Vasodilators		
Nitroglycerin	20 μg/min	40–400 μg/min
Nitroprusside	10 μg/min	30–350 μg/min
Nesiritide	Bolus 2 μg/kg	0.01–0.03 μg/kg per min[a]
Inotropes		
Dobutamine	1–2 μg/kg per min	2–10 μg/kg per min[b]
Milrinone	Bolus 50 μg/kg	0.1–0.75 μg/kg per min[b]
Dopamine	1–2 μg/kg per min	2–4 μg/kg per min[b]
Levosimendan	Bolus 12 μg/kg	0.1–0.2 μg/kg per min[c]

Vasoconstrictors		
Dopamine for hypotension	5 μg/kg per min	5–15 μg/kg per min
Epinephrine	0.5 μg/kg per min	50 μg/kg per min
Phenylephrine	0.3 μg/kg per min	3 μg/kg per min
Vasopressin	0.05 units/min	0.1–0.4 units/min

Notes: [a] Usually <4 μg/kg/min.

[b] Inotropes will also have vasodilatory properties.

[c] Approved outside the United States for management of acute heart failure.

Patients who present with Profile L ("cold and dry") should be carefully evaluated by right-heart catheterization for the presence of an occult elevation of LV filling pressures. If LV filling pressures are low [pulmonary capillary wedge pressure (PCWP) <12 mmHg], a cautious trial of fluid repletion may be considered. The goals of further therapy depend on the clinical situation. Therapy to reach the aforementioned goals may not be possible in some patients, particularly if they have disproportionate RV dysfunction or if they develop cardiorenal syndrome, in which renal function deteriorates during aggressive diuresis. Worsening renal dysfunction occurs in approximately 25% of patients hospitalized with HF and is associated with prolonged hospital stays and higher mortality rates after discharge.

Pharmacologic Management of Acute HF

(Table 234-6)

Vasodilators

After diuretics, intravenous vasodilators are the most useful medications for the management of acute HF. By stimulating guanylyl cyclase within smooth-muscle cells, nitroglycerin, nitroprusside, and nesiritide exert dilating effects on arterial resistance and venous capacitance vessels, which results in a lowering of LV filling pressure, a reduction in mitral regurgitation, and improved forward cardiac output without increasing heart rate or causing arrhythmias. Hypotension is the most common side effect of all vasodilating agents.

Intravenous *nitroglycerin* generally is begun at 20 μg/min and is increased in 20-μg increments until patient symptoms are improved or PCWP is decreased to 16 mmHg without reducing systolic blood pressure below 80 mmHg. The most common side effect of IV or oral nitrates is headache, which, if mild, can be treated with analgesics and often resolves during continued therapy. *Nitroprusside* generally is initiated at 10 μg/min and increased by 10–20 μg every 10–20 min as tolerated, with the same hemodynamic goals as described above. The rapidity of onset and offset, with a half-life of approximately 2 min, facilitates early establishment of an individual patient's optimal level of vasodilation in the ICU. The major limitation of nitroprusside is side effects from cyanide toxicity, which manifests predominantly as gastrointestinal and central nervous system manifestations and is most likely to occur in patients receiving >250 μg/min for over 48 h.

Nesiritide, the newest vasodilator, is a recombinant form of brain-type natriuretic peptide, which is an endogenous peptide secreted primarily from the LV in response to an increase in wall stress. Nesiritide is given as a bolus (2 μg/kg) followed by a fixed-dose infusion (0.01–0.03 μg/kg per min). Nesiritide effectively lowers LV filling pressures and improves symptoms during the treatment of acute HF. Headache is less common with nesiritide than with nitroglycerin. Although termed a *natriuretic peptide*, nesiritide has not been associated with major diuresis when used alone in clinical trials. It does, however, appear to potentiate the effect of concomitant diuretics such that the total required diuretic dose may be slightly lower. There have, however, been recent concerns about the adverse effects of neseritide on renal function in acute decompensated HF which may be related to the initial bolus.

Inotropic Agents

Positive inotropic agents produce direct hemodynamic benefits by stimulating cardiac contractility as well as by producing peripheral vasodilation. Collectively, these hemodynamic effects result in an improvement in cardiac output and a fall in LV filling pressures.

Dobutamine, which is the most commonly used inotropic agent for the treatment of acute HF, exerts its effects by stimulating β_1 and β_2 receptors, with little effect on α_1 receptors. Dobutamine is given as a continuous infusion at an initial infusion rate of 1–2 μg/kg per min. Higher doses (>5 μg/kg per min) are frequently necessary for severe hypoperfusion; however, there is little added benefit to increasing the dose above 10 μg/kg per min. Patients maintained on chronic infusions for >72 h generally develop tachyphylaxis and require increasing doses.

Milrinone is a phosphodiesterase III inhibitor that leads to increased cyclic AMP by inhibiting its breakdown. Milrinone may act synergistically with β-adrenergic agonists to achieve a greater increase in cardiac output than is achieved with either agent alone, and it may also be more effective than dobutamine in increasing cardiac output in the presence of beta blockers. Milrinone may be

administered as a bolus dose of 50 μg/kg per min, followed by a continuous infusion rate of 0.1–0.75 μg/kg per min. If the patient has a low blood pressure, many clinicians will omit the bolus dose. Because milrinone is a more effective vasodilator than dobutamine, it produces a greater reduction in LV filling pressures, albeit with a greater risk of hypotension.

Although short-term use of inotropes provides hemodynamic benefits, these agents are more prone to cause tachyarrhythmias and ischemic events than vasodilators are. Therefore, inotropes are most appropriately used in clinical settings in which vasodilators and diuretics are not helpful, such as in patients with poor systemic perfusion and/or cardiogenic shock, patients requiring short-term hemodynamic support after an MI or surgery, and patients awaiting cardiac transplantation, or as palliative care in patients with advanced HF. If patients require sustained use of intravenous inotropes, strong consideration should be given to the use of an ICD to safeguard against the proarrhythmic effects of these agents.

Vasoconstrictors

Vasoconstrictors are used to support systemic blood pressure in patients with HF. Of the three agents that are commonly used (Table 234-6), dopamine is generally the first choice for therapy in situations in which modest inotropy and pressor support are required. Dopamine is an endogenous catecholamine that stimulates β_1 and α_1 receptors and dopaminergic receptors (DA_1 and DA_2) in the heart and circulation. The effects of dopamine are dose-dependent. Low doses of dopamine (<2 μg/kg per min) stimulate the DA_1 and DA_2 receptors and cause vasodilation of the splanchnic and renal vasculature. Moderate doses (2–4 μg/kg per min) stimulate the β_1 receptors and cause an increase in cardiac output with little or no change in heart rate or SVR. At higher doses (≥5 μg/kg per min) the effects of dopamine on the α_1 receptors overwhelm the dopaminergic receptors, and vasoconstriction ensues, leading to an increase in SVR, LV filling pressures, and heart rate. Significant additional inotropic and blood pressure support can be provided by epinephrine, phenylephrine, and vasopressin (Table 234-6); however, prolonged use of these agents can lead to renal and hepatic failure and can cause gangrene of the limbs. Therefore, these agents should not be administered except in true emergency situations.

Vasopressin Antagonists

Vasopressin levels are often elevated in patients with HF and LV dysfunction and may contribute to the hyponatremia that develops in HF patients. Vasopressin antagonists reduce body weight and edema and normalize serum sodium in patients with hyponatremia but have not been associated with improved patient outcomes in clinical trials. Tolvaptan (oral) and conivaptan (IV) are currently approved for the treatment of hyponatremia but are not approved for the treatment of HF.

Mechanical and Surgical Interventions

If pharmacologic interventions fail to stabilize a patient with refractory HF, mechanical and surgical interventions may provide effective circulatory support. These interventions include intraaortic balloon counter pulsation, percutaneous and surgically implanted LV assist devices, and cardiac transplantation (Chap. 235).

Planning for Hospital Discharge

Patient education should take place during the entire hospitalization, with a specific focus on salt and fluid status and obtaining daily weights, in addition to medication schedules. Although the majority of patients hospitalized with HF can be stabilized and returned to a good level of function on an oral regimen designed to maintain stability, 30–50% of patients discharged with a diagnosis of HF are rehospitalized within 3–6 months. Although there are multiple reasons for rehospitalization, failure to meet criteria for discharge is perhaps the most common. Criteria for discharge should include at least 24 h of stable fluid status, blood pressure, and renal function on the oral regimen planned for home. Before discharge, patients should be free of dyspnea or symptomatic hypotension while at rest, washing, and walking on the ward.

COR PULMONALE

Definition

Cor pulmonale, often referred to as *pulmonary heart disease*, is defined as dilation and hypertrophy of the right ventricle in response to diseases of the pulmonary vasculature and/or lung parenchyma. Historically, this definition has excluded congenital heart disease and those diseases in which the right heart fails secondary to dysfunction of the left side of the heart.

Etiology and Epidemiology

Cor pulmonale develops in response to acute or chronic changes in the pulmonary vasculature and/or the lung parenchyma that are sufficient to cause pulmonary hypertension. The true prevalence of cor pulmonale is difficult to ascertain for two reasons. First, not all patients with chronic lung disease will develop cor pulmonale, and second, our ability to diagnose pulmonary hypertension and cor pulmonale by routine physical examination and laboratory testing is relatively insensitive. However, advances in

2-D echo/Doppler imaging and biomarkers (BNP) make it easier to screen for and detect cor pulmonale.

Once patients with chronic pulmonary or pulmonary vascular disease develop cor pulmonale, the prognosis worsens. Although chronic obstructive pulmonary disease (COPD) and chronic bronchitis are responsible for approximately 50% of the cases of cor pulmonale in North America (Chap. 260), any disease that affects the pulmonary vasculature (Chap. 250) or parenchyma can lead to cor pulmonale. **Table 234-7** provides a list of common diseases that may lead to cor pulmonale. In contrast to COPD, the elevation in pulmonary artery pressure appears to be substantially higher in the interstitial lung diseases (Chap. 261), in which there is an inverse correlation between pulmonary artery pressure and the diffusion capacity for carbon monoxide, as well as patient survival. When cor pulmonale occurs in conjunction with obstructive sleep apnea, typically COPD or a hypoventilation syndrome [e.g., obesity hypoventilation syndrome (OHS)] is present concurrently (Chap. 265).

Table 234-7 Etiology of Chronic Cor Pulmonale

Diseases Leading to Hypoxemic Vasoconstriction
Chronic bronchitis Chronic obstructive pulmonary disease
Cystic fibrosis
Chronic hypoventilation
Obesity
Neuromuscular disease
Chest wall dysfunction
Living at high altitudes
Diseases that Cause Occlusion of the Pulmonary Vascular Bed
Thromboembolic disease, acute or chronic
Pulmonary arterial hypertension
Pulmonary veno-occlusive disease
Diseases that Lead to Parenchymal Disease
Chronic bronchitis
Chronic obstructive pulmonary disease
Bronchiectasis
Cystic fibrosis Pneumoconiosis
Sarcoidosis
Interstitial lung disease

Pathophysiology and Basic Mechanisms

Although many conditions can lead to cor pulmonale, the common pathophysiologic mechanism in each case is pulmonary hypertension that is sufficient to lead to RV dilation, with or without the development of concomitant RV hypertrophy. The systemic consequences of cor pulmonale relate to alterations in cardiac output as well as salt and water homeostasis. Anatomically, the RV is a thin-walled, compliant chamber that is better suited to handle volume overload than pressure overload. Thus, the sustained pressure overload imposed by pulmonary hypertension and increased pulmonary vascular resistance eventually causes the RV to fail.

The response of the RV to pulmonary hypertension depends on the acuteness and severity of the pressure overload. Acute cor pulmonale occurs after a sudden and severe stimulus (e.g., massive pulmonary embolus), with RV dilatation and failure but no RV hypertrophy (Chap. 262). Chronic cor pulmonale, however, is associated with a more slowly evolving and progressive pulmonary hypertension that leads to initial modest RV hypertrophy and subsequent RV dilation.

Decompensation of chronic cor pulmonale can be aggravated by intermittent events that induce pulmonary vasoconstriction and RV afterload, such as hypoxemia and especially hypercarbia-induced respiratory acidosis (e.g., OHS), as well as sustained events, including COPD exacerbations, acute pulmonary emboli, and positive-pressure (mechanical) ventilation. RV failure also can be precipitated by alterations in RV volume that occur in various settings, including increased salt and fluid retention, atrial arrhythmias, polycythemia, sepsis, and a large left-to-right (extracardiac) shunt. **The most common mechanisms that lead to pulmonary hypertension, including vasoconstriction, activation of the clotting cascade, and obliteration of pulmonary arterial vessels, are discussed in Chap. 250.**

Clinical Manifestations

SYMPTOMS

The symptoms of chronic cor pulmonale generally are related to the underlying pulmonary disorder. Dyspnea, the most common symptom, is usually the result of the increased work of breathing secondary to changes in elastic recoil of the lung (fibrosing lung diseases), altered respiratory mechanics (e.g., overinflation with COPD), or inefficient ventilation (e.g., primary pulmonary vascular disease).Orthopnea and paroxysmal nocturnal dyspnea are rarely symptoms of isolated right HF and usually point toward concurrent left heart dysfunction. Rarely, these symptoms reflect increased work of breathing in the supine position resulting from compromised diaphragmatic excursion. Tussive or effort-related syncope may occur because of the inability of the RV to deliver blood adequately to the left side of the heart. Abdominal pain and ascites that occur with cor pulmonale are similar to the right-heart failure that ensues in chronic HF. Lower-extremity edema may occur secondary to neurohormonal activation, elevated RV filling pressures, or increased levels of carbon dioxide and hypoxemia, which can lead to peripheral vasodilation and edema formation. **The symptoms of acute cor pulmonale with pulmonary embolus are reviewed in Chap. 262.**

SIGNS

Many of the signs encountered in cor pulmonale are also present in HF patients with a depressed EF, including tachypnea, elevated jugular venous pressures, hepatomegaly, and lower-extremity edema. Patients may have prominent *v* waves in the jugular venous pulse as a result of tricuspid regurgitation. Other cardiovascular signs include an RV heave palpable along the left sternal border or in the epigastrium. The increase in intensity of the holosystolic murmur of tricuspid regurgitation with inspiration ("Carvallo's sign") may be lost eventually as RV failure worsens. Cyanosis is a late finding in cor pulmonale and is secondary to a low cardiac output with systemic vasoconstriction and ventilation-perfusion mismatches in the lung.

Diagnosis

The most common cause of right-heart failure is not pulmonary parenchymal or vascular disease but left heart failure. Therefore, it is important to evaluate the patient for LV systolic and diastolic dysfunction. The ECG in severe pulmonary hypertension shows P pulmonale, right axis deviation, and RV hypertrophy. Radiographic examination of the chest may show enlargement of the main pulmonary artery, the hilar vessels, and the descending right pulmonary artery. Spirometry and lung volumes can identify obstructive and/or restrictive defects indicative of parenchymal lung diseases; arterial blood gases can demonstrate hypoxemia and/or hypercapnia. Spiral computed tomography (CT) scans of the chest are useful in diagnosing acute thromboembolic disease; however, ventilation-perfusion lung scanning remains best suited for diagnosing *chronic thromboembolic disease* (Chap. 262). A high-resolution CT scan of the chest can identify interstitial lung disease.

Two-dimensional echocardiography is useful for measuring RV thickness and chamber dimensions as well as the anatomy of the pulmonary and tricuspid valves. Location of the RV behind the sternum and its crescent shape challenge assessment of RV function by echocardiography, especially when parenchymal lung disease is present. Calculated measures of RV function [e.g., tricuspid annular plane systolic excursion (TAPSE) or the Tei Index] supplement more subjective assessments of RV function. The interventricular septum may move paradoxically during systole in the presence of pulmonary hypertension. As noted, Doppler echocardiography can be used to assess pulmonary artery pressures. MRI is also useful for assessing RV structure and function, particularly in patients who are difficult to image with 2-D echocardiography because of severe lung disease. Right-heart catheterization is useful for confirming the diagnosis of pulmonary hypertension and for excluding elevated left-heart pressures (measured as the PCWP) as a cause for right-heart failure. BNP and N-terminal BNP levels are elevated in patients with cor pulmonale secondary to RV stretch and may be dramatically elevated in acute pulmonary embolism.

Treatment: Cor Pulmonale

The primary treatment goal of cor pulmonale is to target the underlying pulmonary disease, since this will decrease pulmonary vascular resistance and lessen RV afterload. Most pulmonary diseases that lead to chronic cor pulmonale are advanced and therefore are less amenable to treatment. General principles of treatment include decreasing work of breathing by using noninvasive mechanical ventilation and bronchodilation, as well as treating any underlying infection (Chaps. 260 and 261). Adequate oxygenation (oxygen saturation ≥90–92%) and correcting respiratory acidosis are vital for decreasing pulmonary vascular resistance and reducing demands on the RV. Patients should be transfused if they are anemic, and phlebotomy may be considered in extreme cases of polycythemia.

Diuretics are effective in RV failure, and indications are similar to those for chronic HF. One caveat of chronic diuretic use is to avoid inducing contraction alkalosis and worsening hypercapnia. Digoxin is of uncertain benefit in the treatment of cor pulmonale and may lead to arrhythmias in the setting of tissue hypoxemia and acidosis. Therefore, if digoxin is administered, it should be given at low doses and monitored carefully.

Pulmonary vasodilators can effectively improve symptoms through modest reduction of pulmonary pressures and RV afterload when isolated pulmonary arterial hypertension is present. Vasodilators are unproven in cases of pulmonary hypertension and cor pulmonale due to parenchymal lung diseases or hypoventilation syndromes. **The treatment of the acute cor pulmonale that occurs with pulmonary embolus is described in Chap. 262. The treatment of pulmonary hypertension is discussed in Chap. 250.**

FURTHER READINGS

Ashrafian H et al: Metabolic mechanisms in heart failure. Circulation 116:434, 2007[PMID: 17646594]

Bardy GH et al: Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. N Engl J Med 352:225, 2005[PMID: 15659722]

Chapman HA: Disorders of lung matrix remodeling. J Clin Invest 113:148, 2004[PMID: 14722604]

Cleland JG et al: The effect of cardiac resynchronization on morbidity and mortality in heart failure. N Engl J Med 352:1539, 2005[PMID: 15753115]

Friedrich EB, Bohm M: Management of end stage heart failure. Heart 93:626, 2007[PMID: 17435073]

Hunt SA et al: 2009 focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the International Society for Heart and Lung Transplantation. Circulation 119:e391, 2009

Kessler R et al: "Natural history" of pulmonary hypertension in a series of 131 patients with chronic obstructive lung disease. Am J Respir Crit Care Med 164:219, 2001[PMID: 11463591]

Mann DL, Bristow MR: Mechanisms and models in heart failure: The biomechanical model and beyond. Circulation 111:2837, 2005[PMID: 15927992]

Mosterd A, Hoes AW: Clinical epidemiology of heart failure. Heart 93:1137, 2007[PMID: 17699180]

Pengo V et al: Incidence of chronic thromboembolic pulmonary hypertension after pulmonary embolism. N Engl J Med 350:2257, 2004[PMID: 15163775]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CARDIAC TRANSPLANTATION AND PROLONGED ASSISTED CIRCULATION: INTRODUCTION

Advanced or end-stage heart failure is an increasingly frequent sequela, as progressively more effective palliation for the earlier stages of heart disease and prevention of sudden death associated with heart disease become more widely recognized and employed (Chap. 234). When patients with end-stage or refractory heart failure are identified, the physician is faced with the decision of advising compassionate end-of-life care or choosing to recommend extraordinary life-extending measures. For the occasional patient who is relatively young and without serious comorbidities, the latter may represent a reasonable option. Current therapeutic options are limited to cardiac transplantation (with the option of mechanical cardiac assistance as a "bridge" to transplantation) or (at least in theory) the option of permanent mechanical assistance of the circulation. In the future, it is possible that genetic modulation of ventricular function or cell-based cardiac repair will be options for such patients. Currently, both approaches are considered to be experimental.

CARDIAC TRANSPLANTATION

Surgical techniques for orthotopic transplantation of the heart were devised in the 1960s and taken into the clinical arena in 1967. The procedure did not gain widespread clinical acceptance until the introduction of "modern" and more effective immunosuppression in the early 1980s. By the 1990s, the demand for transplantable hearts met, and then exceeded, the available donor supply and leveled off at about 4,000 heart transplants annually worldwide, according to data from the Registry of the International Society for Heart and Lung Transplantation (ISHLT). Subsequently, heart transplant activity in the United States has remained stable at ~2,200/year, but worldwide activity reported to this registry has decreased somewhat. This apparent decline in numbers may be a result of the fact that reporting is legally mandated in the United States, but not elsewhere, and several countries have started their own databases.

Surgical Technique

Donor and recipient hearts are excised in virtually identical operations with incisions made across the atria and atrial septum at the midatrial level (leaving the posterior walls of the atria in place) and across the great vessels just above the semilunar valves. The donor heart is generally "harvested" in an anatomically identical manner by a separate surgical team and transported from the donor hospital in a bag of iced saline solution and then is reanastomosed into the waiting recipient in the orthotopic or normal anatomic position. The only change in surgical technique since this method was first described has been a movement in recent years to move the right atrial anastamosis back to the level of the superior and inferior vena cavae to better preserve right atrial geometry and prevent atrial arrhythmias. Both methods of implantation leave the recipient with a surgically denervated heart that does not respond to any direct sympathetic or parasympathetic stimuli but does respond to circulating catecholamines. The physiologic responses of the denervated heart to the demands of exercise are atypical but quite adequate to carry on normal physical activity.

Donor Allocation System

In the United States the allocation of donor organs is accomplished under the supervision of the United Network for Organ Sharing (UNOS), a private organization under contract to the federal government. The United States is divided geographically into eleven regions for donor heart allocation. Allocation of donor hearts within a region is decided according to a system of priority that takes into account (1) the severity of illness, (2) geographic distance from the donor, and (3) patient time on the waiting list. A physiologic limit of ~3 h of "ischemic" (out-of-body) time for hearts precludes a national sharing of hearts. This allocation system design is reissued annually and is responsive to input from a variety of constituencies, including both donor families and transplant professionals.

At the current time, highest priority according to severity of illness is assigned to patients requiring hospitalization at the transplant center for IV inotropic support with a pulmonary artery catheter in place for hemodynamic monitoring or to patients requiring mechanical circulatory support [i.e., intra-aortic balloon pump (IABP), right or left ventricular assist device (RVAD, LVAD), extracorporeal membrane oxygenation (ECMO), or mechanical ventilation]. Second highest priority is given to patients requiring ongoing inotropic support, but without a pulmonary artery catheter in place. All other patients have priority according to their time accrued on the waiting list, and matching is achieved only according to ABO blood group compatibility and gross body size compatibility, although some patients who are "pre-sensitized" and have preexisting anti-HLA antibodies (commonly multiparous women or patients previously multiply transfused) undergo prospective cross-matching with the donor. While HLA matching of donor and recipient would be ideal, the relatively small numbers of patients, as well as the time constraints involved, make such matching impractical.

Indications/Contraindications

Heart failure is an increasingly common cause of death, particularly in the elderly. Most patients who reach what has recently been categorized as stage D, or refractory end-stage heart failure, are appropriately treated with compassionate end-of-life care. A subset of such patients who are younger and without significant comorbidities can be considered as candidates for heart transplantation. Exact criteria vary in different centers but generally take into consideration the patient's physiologic age and the existence of comorbidities such as peripheral or cerebrovascular disease, obesity, diabetes, cancer, or chronic infection.

Results

A registry organized by the ISHLT has tracked worldwide and U.S. survival rates after heart transplantation since 1982. The most recent update reveals 83% and 76% survival 1 and 3 years posttransplant, or a posttransplant "half-life" of 10.00 years **(Fig. 235-1)**. The quality of life in these patients is generally excellent, with well over 90% of patients in the registry returning to normal and unrestricted function following transplantation.

Figure 235-1

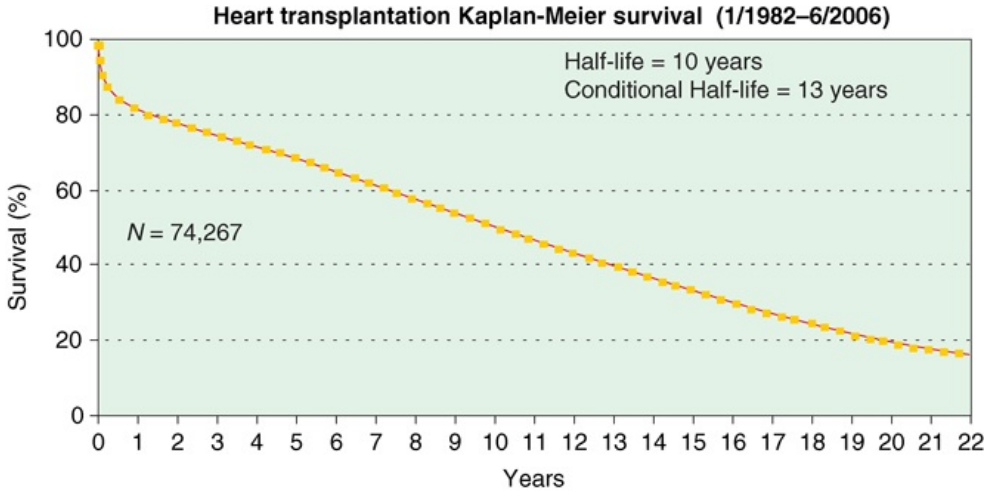

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Survival was calculated using the Kaplan-Meier method, which incorporates information from all transplants for whom any follow-up has been provided. Because many patients are still alive and some patients have been lost to follow-up, the survival rates are estimates rather than exact rates because the time of death is not known for all patients. Therefore, 95% confidence limits are provided. (*From J Heart Lung Transplant 2008; 27:937–983.)*

Immunosuppression

Medical regimens employed to provide suppression of the normal immune response to a solid organ allograft vary from center to center and are in a constant state of evolution, as more effective agents with improved side-effect profiles and less toxicity are introduced. All currently used regimens are nonspecific, providing general hyporeactivity to foreign antigens rather than donor-specific hyporeactivity, and also providing the attendant, and unwanted, susceptibility to infections and malignancy. Most cardiac transplant programs currently use a three-drug regimen including a calcineurin inhibitor (cyclosporine or tacrolimus), an inhibitor of T cell proliferation or differentiation (azathioprine, mycophenolate mofetil, or sirolimus), and at least a short initial course of glucocorticoids. Many programs also include an initial "induction" course of polyclonal or monoclonal anti-T cell antibodies in the perioperative period to decrease the frequency or severity of early posttransplant rejection. Most recently introduced have been monoclonal antibodies (daclizumab and basiliximab) that block the interleukin 2 receptor and may provide prevention of allograft rejection without additional global immunosuppression.

Diagnosis of cardiac allograft rejection is usually made with the use of endomyocardial biopsy, either done on a surveillance basis or in response to clinical deterioration. Biopsy surveillance is performed on a regular basis in most programs for the first year postoperatively and for the first five years in many programs. Therapy consists of augmentation of immunosuppression, the intensity and duration of which is dictated by the severity of the rejection.

Late Posttransplant Management Issues

Increasing numbers of heart transplant patients are surviving for years following transplantation and constitute a population of patients with a number of long-term management issues.

ALLOGRAFT CORONARY ARTERY DISEASE

Despite usually having young donor hearts, cardiac allograft recipients are prone to develop coronary artery disease (CAD). This CAD is generally a diffuse, concentric, and longitudinal process that is quite

different from "ordinary" atherosclerotic CAD, which is more focal and often eccentric. The underlying etiology is most likely primarily immunologic injury of the vascular endothelium, but a variety of risk factors influence its existence and progression and include nonimmunologic factors such as dyslipidemia, diabetes mellitus, and cytomegalovirus (CMV) infection. It is hoped that newer and improved immunosuppressive modalities will reduce the incidence and impact of these devastating complications, which currently account for the majority of late posttransplant deaths. Thus far, the immunosuppressive agents mycophenolate mofetil and the mammalian target of rapamycin (mTOR) inhibitors sirolimus and everolimus have been shown to be associated with short-term lesser incidence and extent of coronary intimal thickening; in anecdotal reports, institution of sirolimus was associated with some reversal of the disease. The use of statins has also been shown to be associated with a reduced incidence of this vasculopathy, and these drugs are now almost universally used in transplant recipients unless contraindicated. Palliation of the disease with percutaneous interventions is probably safe and effective in the short term, although the disease often advances relentlessly. Because of the denervated status of the organ, patients rarely experience angina pectoris, even with advanced stages of the disease.

Retransplantation is the only definitive form of therapy for advanced allograft CAD, but the scarcity of donor hearts makes the decision to pursue retransplantation a difficult one in an individual patient, as well as a difficult ethical issue.

MALIGNANCY

The occurrence of an increased incidence of malignancy is a well-recognized sequela of any program of chronic immunosuppression, and organ transplantation is no exception. Lymphoproliferative disorders are among the most frequent posttransplant complications and, in most cases, seem to be driven by the Epstein-Barr virus. Effective therapy includes reduction of immunosuppression (a clear "double-edged sword" in the setting of a life-sustaining organ), antiviral agents, and traditional chemo- and radiotherapy. Most recently, specific antilymphocyte (CD20) therapy has shown great promise. Cutaneousmalignancies (both basal cell and squamous cell carcinomas) also occur with increased frequency in transplant recipients and can pursue very aggressive courses. The role of decreasing immunosuppression for treatment of these cancers is far less clear.

INFECTIONS

The use of currently available nonspecific immunosuppressive modalities to prevent allograft rejection naturally results in an increased susceptibility to infectious complications in transplant recipients. Although their incidence has decreased since the introduction of cyclosporine, infections with unusual and opportunistic organisms remain the major cause of death during the first postoperative year and remain a threat to the chronically immunosuppressed patient throughout life. Effective therapy depends on careful surveillance for early signs and symptoms of opportunistic infection and an extremely aggressive approach to obtaining a specific diagnosis as well as expertise in recognizing the more common clinical presentations of CMV, *Aspergillus*, and other opportunistic infectious agents.

PROLONGED ASSISTED CIRCULATION

The modern era of mechanical circulatory support can be traced back to 1953, when cardiopulmonary bypass was first used in a clinical setting and ushered in the possibility of brief periods of circulatory support to permit open-heart surgery. Subsequently, a variety of extracorporeal pumps to provide circulatory support for brief periods of time have been developed. The use of a mechanical device to support the circulation for more than a few hours initially developed slowly, with the implant of a total artificial heart in 1969 in Texas by Cooley. This patient survived for 60 hours until a donor organ became available, at which point he was transplanted. Unfortunately, the patient died of pulmonary complications after transplantation. The entire field of mechanical replacement of the heart took a decade-long hiatus until the 1980s, when total artificial hearts were reintroduced with much publicity; however, they failed to produce the hoped-for treatment of end-stage heart disease. Starting in the 1970s, parallel to the development of the total artificial heart, there was intense research in the development of ventricular assist devices, which provide mechanical assistance to (rather than replacement of) the failing ventricle (currently, newer versions of the total artificial heart are in preliminary clinical trials).

Although conceived of initially as alternatives to biologic replacement of the heart, LVADs were introduced as, and are still employed primarily as, temporary "bridges" to heart transplantation in candidates who begin to fail medical therapy before a donor heart becomes available. Several devices are approved by the U.S. Food and Drug Administration (FDA) and are currently in widespread use. Those that are implantable within the body are compatible with hospital discharge and offer the patient a chance for life at home while waiting for a donor heart. However successful such "bridging" is for the individual patient, it does nothing to alleviate the scarcity of donor hearts; the ultimate goal in the field remains that of providing a reasonable alternative to biologic replacement of the heart—one that is widely and easily available and cost-effective.

Current Indications and Applications of Ventricular Assist Devices

Currently, there are two major indications for long-term ventricular assistance. First, patients with chronic end-stage heart failure are eligible for mechanical support if they are at risk of imminent death from cardiogenic shock. Second, if patients have a left ventricular ejection fraction < 25%, peak VO2 < 14 mL/kg/min or are dependent on inotropic therapy or support with intra-aortic balloon counterpulsation, they may be eligible for mechanical support. If they are eligible for heart transplantation, the mechanical circulatory assistance is termed "bridge to transplantation." By contrast, if the patient has a contraindication

to heart transplantation, the device therapy is deemed to be "destination" left ventricular assistance therapy.

Available Devices

In the United States, there are currently four FDA-approved devices that are used as bridges to transplantation in adults. Of these four devices, one is also approved for use as destination therapy or long-term mechanical support of the heart. There are a number of other devices that are approved only for short-term support for post-cardiac surgery shock or for patients with cardiogenic shock secondary to acute myocardial infarction or fulminant myocarditis; these will not be considered here. None of the long-term devices as yet are totally implantable and, because of this need for transcutaneous connections, all share a common problem with infectious complications. Likewise, all share some tendency to thromboembolic complications as well as the expected possibility of mechanical device failure common to any machine.

The CardioWest total artificial heart (TAH) (Syncardia, Tucson, AZ) is a pneumatic, biventricular, orthotopically implanted total artificial heart with an externalized driveline connecting it to its console. It consists of two spherical polyurethane chambers with polyurethane diaphragms. Inflow and outflow conduits are constructed of Dacron and contain Medtronic-Hall (Medtronic, Inc., Minneapolis, MN) valves. It is currently the only FDA-approved device for use as a bridge to transplantation in patients who have severe biventricular failure.

The Thoratec LVAD (Thoratec Corp., Pleasanton, CA) is an extracorporeal pump that takes blood from a large cannula placed in the left ventricular apex and propels it forward through an outflow cannula inserted into the ascending aorta. The pump itself sits in the paracorporeal position on the abdomen and is attached to a device console cart with wheels, allowing for limited ambulation. The extracorporeal nature of this pump allows it to be used in small adults for whom intracorporeal pumps would be too large.

The Novacor LVAD (WorldHeart, Inc., Oakland, CA) also takes blood from the left ventricular apex through a cannula and propels it into the ascending aorta through a second cannula. With this device, the pump itself is placed in a surgically created pocket in the peritoneal fascia in the abdomen. A driveline that connects to the power source is tunneled subcutaneously and usually exits in the right upper quadrant of the abdomen.

The HeartMate XVE LVAD (Thoratec Corp., Pleasanton, CA) is an intracorporeal left ventricular assist device that has an externalized driveline. The pump sits in the anterior abdominal wall with cannulae that traverse across the diaphragm. There is a drainage cannula in the left ventricular apex, and the blood is expelled from the pump into the ascending aorta via a synthetic vascular graft. This device may be used as a bridge to transplantation and patients may be discharged from the hospital with this device to await transplantation. The HeartMate XVE LVAD is now one of two FDA-approved devices for destination therapy.

The HeartMate II LVAS (Thoratec Corp., Pleasanton, CA) similarly uses a drainage cannula in the left ventricular apex to drain blood into a small chamber, where the blood is driven by an electrically powered motor that spins a rotor, accelerating blood outflow into the ascending aorta **(Figure 235-2)**. This device is currently the only FDA-approved axial-flow pump that can be used both as a bridge to transplantation and as destination therapy. There are several other axial-flow pumps currently being investigated. These devices have fewer moving parts than previous devices and provide non-pulsatile blood flow. All current axial-flow pumps continue to require transcutaneous connections to power the electric motor. Newer, third-generation devices, which also provide non-pulsatile flow, work through a different mechanism than the axial-flow pumps and are currently being investigated. These devices are even smaller than the currently available axial-flow pumps, and their mechanism of action involves less trauma to blood cells, which may result in better durability and decreased long-term complications.

Figure 235-2

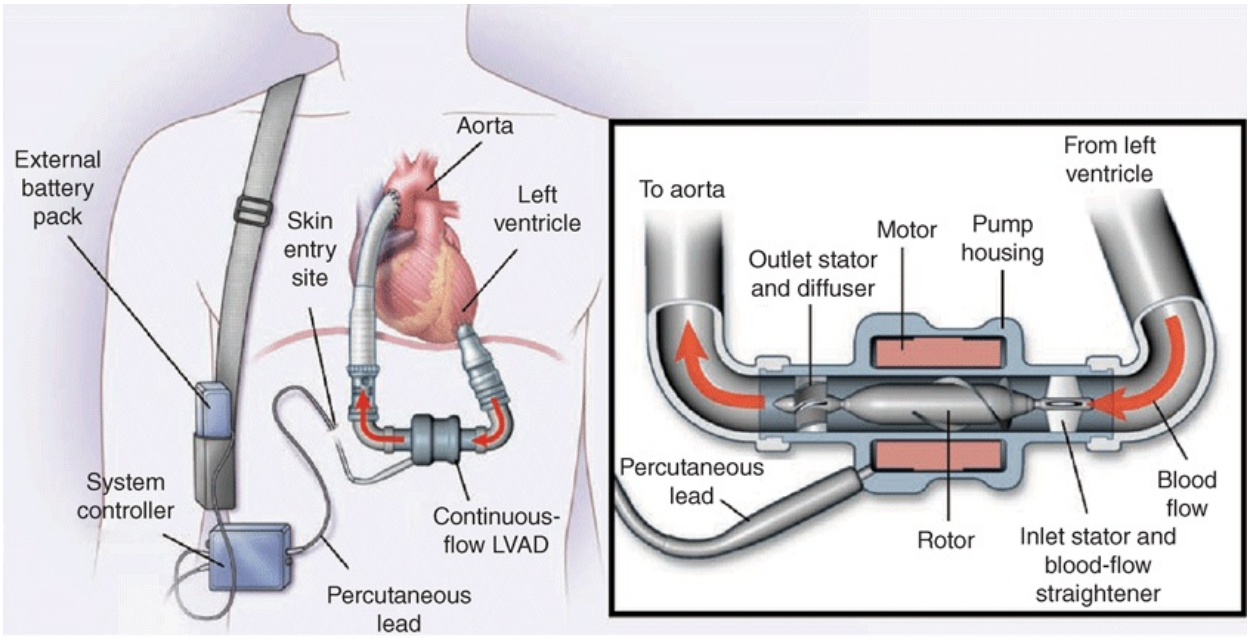

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of HeartMate II left ventricular assist device. *(Reprinted with permission from Thoratec Corp., Pleasanton, CA.)*

Results

The use of these devices in the United States is limited mainly to patients with post-cardiac surgery shock and to those who are bridged to transplantation. The results of bridging to transplantation with the available devices are quite good, with nearly 75% of younger patients receiving a transplant by 1 year and having excellent posttransplant survival rates.

Publication of the REMATCH (Randomized Evaluation of Mechanical Assistance in the Treatment of Heart Failure) trial in 2001 documented a somewhat improved survival rate in nontransplant candidates with end-stage heart disease randomized to a HeartMate XVE LVAD (albeit with a high rate of complications, especially neurologic ones) as opposed to continued medical therapy; this led to renewed interest in also using the devices as non- biologic permanent replacement of heart function, as well as to FDA approval of one device for this indication. This outcome, in turn, led the ISHLT to initiate a Mechanical Circulatory Support database in 2002, which collects voluntary data from 60 international centers and contained data from 655 patients in its most recent publication. Only 12% of these patients had the device placed with the intention of permanent, or "destination," use, with survival rates of only 65% at 6 months and 34% at 1 year.

Several studies have evaluated the benefit of LVAD therapy as a bridge to transplantation, with the most recent data taken from a series of 133 patients who underwent implantation of a HeartMate II device. In this group of patients, 80% achieved the principal outcome (defined as survival to transplantation, recovery of heart function, or ongoing device support) at 180 days. With increased experience and improved outcomes using LVADs as a bridge to transplantation, the ability to maintain end-organ function and limit the progression of pulmonary hypertension, or even decrease pulmonary vascular resistance, makes mechanical unloading an attractive option when compared with continued inotropic support. The early bridge-to-transplantation experience demonstrated reduced posttransplantation survival when compared with medical management; however, more recent experience has shown equivalent outcomes following transplantation. This result is likely secondary to a trend toward earlier device implantation, prior to the onset of irreversible end-organ damage.

FURTHER READINGS

Garnier JL et al: Treatment of post-transplant lymphomas with anti-B-cell monoclonal antibodies. Recent Results Cancer Res 159:113, 2002[PMID: 11785835]

Gray Na Jr, Selzman CH: Current status of the total artificial heart. Am Heart J 152:4, 2006[PMID: 16824826]

Hunt SA et al: ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing committee to update the 2001 guidelines for the evaluation and management of heart failure). Available at http://www.acc.org/clinical/guidelines/failure//index.pdf

Kirklin JK et al: INTERMACS Database for Durable Devices for Circulatory Support: First Annual Report. J

Heart Lung Transplant 27:1065, 2008[PMID: 18926395]

Mancini D et al: Use of rapamycin slows progression of cardiac transplantation vasculopathy. Circulation 108:48, 2003[PMID: 12742978]

Taylor DO et al: Registry of the International Society for Heart and Lung Transplantation: Twenty-sixth official adult heart transplant report—2005. J Heart Lung Transplant 228: 1007, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CONGENITAL HEART DISEASE IN THE ADULT: INTRODUCTION

A little over a hundred years ago, Sir William Osler, in his classic textbook *The Principles and Practice of Medicine* (New York, Appleton & Co, 1892, pp 659–663), devoted only five pages to "Congenital Affections of the Heart," with the first sentence declaring, that "[t]hese [disorders] have only limited clinical interest, as in a large proportion of cases the anomaly is not compatible with life, and in others nothing can be done to remedy the defect or even to relieve symptoms." Fortunately, in the intervening century, considerable progress has been made in understanding the basis for these disorders and their effective treatment.

The most common birth defects are cardiovascular in origin. These malformations are due to complex multifactorial genetic and environmental causes, but recognized chromosomal aberrations and mutations of single genes account for <10% of all cardiac malformations. Congenital heart disease (CHD) complicates ~1% of all live births in the general population—about 40,000 births/year—but occurs more frequently in the offspring (about 4–5%) of women with CHD. Owing to the remarkable surgical advances over the last 60 years, >90% of afflicted neonates and children now reach adulthood; women with CHD may now frequently successfully bear children after competent repairs. As such, the population with CHD is steadily increasing. Women with aortic disease (e.g., aortic coarctation or Marfan's syndrome) risk aortic dissection. Patients with cyanotic heart disease, pulmonary hypertension, or Marfan's syndrome with a dilated aortic root generally should not become pregnant; those with correctable lesions should be counseled about the risks of pregnancy with an uncorrected malformation versus repair and later pregnancy.

More than one million adults with operated or unoperated CHD live in the United States today and, thus, outnumber the 800,000 children with CHD. Because true surgical cures are rare, and all repairs —be they palliative or corrective—may leave residua, sequelae, or complications, most require some degree of lifetime expert surveillance. The anatomic and physiologic changes in the heart and circulation due to any specific CHD lesion are not static but, rather, progress from prenatal life to adulthood. Malformations that are benign or escape detection in childhood may become clinically significant in the adult. For example, a functionally normal congenitally bicuspid aortic valve may thicken and calcify with time, resulting in significant aortic stenosis; a well-tolerated left-to-right shunt of an atrial septal defect (ASD) may result in cardiac decompensation or pulmonary hypertension only after the fourth to fifth decade.

CARDIAC DEVELOPMENT

(Also See Chapter 224)

CHD is generally the result of aberrant embryonic development of a normal structure or failure of such a structure to progress beyond an early stage of embryonic or fetal development. This brief section serves to introduce the reader to normal development so that defects may be better understood; by necessity, it is not exhaustive. Cardiogenesis is a finely tuned process with transcriptional control of a complex group of regulatory proteins that activate or inhibit their gene targets in a location- and time-dependent manner. At about three weeks of embryonic development, two cardiac cords form and become canalized; at that point, the primordial cardiac tube develops from two sources (cardiac crescent or the first heart field, pharyngeal mesoderm or the second heart field); by 21 days, these fuse into a single cardiac tube beginning at the cranial end. The cardiac tube then elongates and develops discrete constrictions with the following segments from caudal to cranial location: *sinus venosus* receives the umbilical, vitelline, and common cardinal veins: *atrium*, *ventricle*, *bulbus cordis*, *truncus arteriosus*, *aortic sac*, and the *aortic arches*. The cardiac tube is fixed at the sinus venosus and arterial ends.

Subsequently, in the next few weeks, differential growth of cells causes the tube to elongate and loop as an "S" with the bulboventricular portion moving rightward and the atrium and sinus venosus moving posterior to the ventricle. The primitive atrium and ventricle communicate via the atrioventricular canal from which the endocardial cushion develops into two parts (ventrally and dorsally). The cushions fuse and divide the atrioventricular canal into two atrioventricular inlets and also migrate to help form the ventricular septum. The primitive atrium is divided first by a *septum primum* membrane, which grows down from the superior wall to the cushions; as this fusion occurs, the mid-portion resorbs in the center forming the *ostium secundum*. Rightward of the *septum primum*, a second *septum secundum* membrane grows down from the ventral-cranial wall toward—but not reaching—the cushions, and

covering most, but not all, of the *ostium secundum*, resulting in a flap of the *foramen ovale*. The primitive ventricle is partitioned by a finely tuned set of events. The interventricular septum grows up toward the cushions, and the cushions form an upper inlet septum; between the two portions is a hole called the interventricular foramen. The left and right ventricles begin to develop side by side, and the atria and their respective inlet valves align over their ventricles. Finally, these two parts of the septum fuse with the bulboventricular ridges, which, once having septated the truncus arteriosus, extend into the ventricle. The bulbocordis divides into a subaortic portion as the muscular conus resorbs, while the subpulmonary section has elongation of its muscular conus. Spiral division of the common truncus arteriosus rotates and aligns the pulmonary artery and aortic portions over their respective outflow tracts, the aortic valve moving posterior over the left ventricle (LV) outflow tract and the pulmonary valve moving anterior over the right ventricle (RV) outflow tract, with a wraparound relationship of the two great arteries.

Early on, the venous systems are bilateral and symmetric and enter two horns of the sinus venosus. Ultimately, except for the coronary sinus, most of the left-sided portions and the left sinus–venosus horn regress and the systemic venous system empties into the right horn via the inferior and superior vena cavae. The pulmonary venous system, initially connecting to the systemic venous system, develops as buds from the developing lungs fuse together in the *pulmonary venous* confluence at which point the connection to the systemic system regresses. Simultaneously, a projection from the back wall of the left atrium (the *common pulmonary vein*) grows posteriorly to merge with the confluence, which then becomes a part of the posterior left atrial wall.

The truncus arteriosus and aortic sac initially develop six paired symmetric arches, which curve posteriorly and become the paired dorsal aortae. The detailed description of the selective regression of some of the arches is not presented in this chapter. In brief summary, this process results in the development of arch 3 as the internal carotid arteries, left arch 4 as the aortic arch and right subclavian artery, and part of arch 6 as the patent ductus arteriosus. The two dorsal thoracic aortae fuse in the abdomen with persistence of the left dorsal aorta.

SPECIFIC CARDIAC DEFECTS

Tables 236-1, 236-2, and 236-3 list CHD malformations as simple, intermediate, or complex. Simple defects generally are single lesions with a shunt or a valvular malformation. Intermediate defects may have two or more simple defects. Complex defects generally have components of an intermediate defect plus more complex cardiac and vascular anatomy, often with cyanosis, and frequently with transposition complexes. The goal of these tables is to suggest when cardiology consultation or advanced CHD specialty care is needed. Patients with complex CHD (which includes most "named" surgeries that usually involve complex CHD) should virtually always be managed in conjunction with an experienced specialty adult CHD center. Patients with intermediate lesions should have an initial consultation and subsequent occasional intermittent follow-up with a cardiologist. Patients with simple lesions often may be managed by a well-informed internist or general cardiologist, although consultation with a specifically trained adult congenital cardiologist is occasionally advisable.

Table 236-1 Simple Adult Congenital Heart Disease

Native disease
Uncomplicated congenital aortic valve disease
Mild congenital mitral valve disease (e.g., except parachute valve, cleft leaflet)
Uncomplicated small atrial septal defect
Uncomplicated small ventricular septal defect
Mild pulmonic stenosis
Repaired conditions
Previously ligated or occluded ductus arteriosus
Repaired secundum or sinus venosus atrial septal defect without residua
Repaired ventricular septal defect without residua

Table 236-2 Intermediate Complexity Congenital Heart Disease

Ostium primum or sinus venosus atrial septal defect
Anomalous pulmonary venous drainage, partial or total
Atrioventricular canal defects (partial or complete)
Ventricular septal defect, complicated (e.g., absent or abnormal valves or with associated obstructive lesions, aortic regurgitation)

Coarctation of the aorta
Pulmonic valve stenosis (moderate to severe)
Infundibular right ventricular outflow obstruction of significance
Pulmonary valve regurgitation (moderate to severe)
Patent ductus arteriosus (non-closed)—moderate to large
Sinus of Valsalva fistula/aneurysm
Subvalvular or supravalvular aortic stenosis

Table 236-3 Complex Adult Congenital Heart Disease
Cyanotic congenital heart diseases (all forms)
Eisenmenger's syndrome
Ebstein's anomaly
Tetralogy of Fallot or pulmonary atresia (all forms)
Transposition of the great arteries
Single ventricle; tricuspid or mitral atresia
Double-outlet ventricle
Truncus arteriosus
Fontan or Rastelli procedures

Atrial Septal Defect

ASD is a common cardiac anomaly that may be first encountered in the adult and occurs more frequently in females. *Sinus venosus* ASD occurs high in the atrial septum near the entry of the superior vena cava into the right atrium and is associated frequently with anomalous pulmonary venous connection from the right lung to the superior vena cava or right atrium. *Ostium primum* ASDs lie adjacent to the atrioventricular valves, either of which may be deformed and regurgitant. Ostium primum ASDs are common in Down syndrome; the more complex atrioventricular septal defects with a common atrioventricular valve and a posterior defect of the basal portion of the interventricular septum are more typical of this chromosomal defect. The most common *ostium secundum* ASD involves the fossa ovalis and is midseptal in location; this should not be confused with a *patent foramen ovale*. Anatomic obliteration of the foramen ovale ordinarily follows its functional closure soon after birth, but residual "probe patency" is a common normal variant; ASD denotes a true deficiency of the atrial septum and implies functional and anatomic patency. The magnitude of the left-to-right shunt depends on the ASD size, ventricular diastolic properties, and the relative impedance in the pulmonary and systemic circulations. The left-to-right shunt causes diastolic overloading of the right ventricle and increased pulmonary blood flow. Patients with ASD are usually asymptomatic in early life, although there may be some physical underdevelopment and an increased tendency for respiratory infections; cardiorespiratory symptoms occur in many older patients. Beyond the fourth decade, a significant number of patients develop atrial arrhythmias, pulmonary arterial hypertension, bidirectional and then right-to-left shunting of blood, and right heart failure. Patients exposed to the chronic environmental hypoxemia of high altitude tend to develop pulmonary hypertension at younger ages. In older patients, left-to-right shunting across the ASD increases as progressive systemic hypertension and/or coronary artery disease (CAD) result in reduced compliance of the left ventricle.

PHYSICAL EXAMINATION

Examination usually reveals a prominent RV impulse and palpable pulmonary artery pulsation. The first heart sound is normal or split, with accentuation of the tricuspid valve closure sound. Increased flow across the pulmonic valve is responsible for a midsystolic pulmonary outflow murmur. The second heart sound is widely split and is relatively fixed in relation to respiration. A mid-diastolic rumbling murmur, loudest at the fourth intercostal space and along the left sternal border, reflects increased flow across the tricuspid valve. In ostium primum ASD, an apical holosystolic murmur indicates associated mitral or tricuspid regurgitation or a ventricular septal defect (VSD).

These findings are altered when increased pulmonary vascular resistance causes diminution of the left-to-right shunt. Both the pulmonary outflow and tricuspid inflow murmurs decrease in intensity, the pulmonic component of the second heart sound and a systolic ejection sound are accentuated, the two components of the second heart sound may fuse, and a diastolic murmur of pulmonic regurgitation appears. Cyanosis and clubbing accompany the development of a right-to-left shunt (see "Ventricular Septal Defect" below). In adults with an ASD and atrial fibrillation, the physical findings may be confused with mitral stenosis with pulmonary hypertension because the tricuspid diastolic flow murmur and widely split second heart sound may be mistakenly thought to represent the diastolic murmur of mitral stenosis and the mitral "opening snap," respectively.

ELECTROCARDIOGRAM

In ostium secundum ASD, electrocardiogram (ECG) usually shows right-axis deviation and an rSr' pattern in the right precordial leads representing enlargement of the RV outflow tract. An ectopic atrial pacemaker or first-degree heart block may occur with the sinus venous ASD. In ostium primum ASD, the RV conduction defect is accompanied by left superior axis deviation and counterclockwise rotation of the frontal plane QRS loop. Varying degrees of RV and right atrial (RA) enlargement or hypertrophy may occur with each type of defect, depending on the height of the pulmonary artery pressure. *Chest x-ray* shows an enlarged right atrium and right ventricle, and pulmonary artery and its branches; increased pulmonary vascular markings of left-to-right shunt vascularity will diminish if pulmonary vascular disease develops.

ECHOCARDIOGRAM

Echocardiography reveals pulmonary arterial and RV and RA dilatation with abnormal (paradoxical) ventricular septal motion in the presence of a significant right heart volume overload. The ASD may be visualized directly by two-dimensional imaging, color-flow imaging, or echocontrast. In most institutions, two-dimensional echocardiography and Doppler examination have supplanted cardiac catheterization. Transesophageal echocardiography is indicated if the transthoracic echocardiogram is ambiguous, which is often the case with sinus venosus defects, or during catheter device closure **(Fig. 236-1)**. Cardiac catheterization is performed if inconsistencies exist in the clinical data, if significant pulmonary hypertension or associated malformations are suspected, or if CAD is a possibility.

Figure 236-1

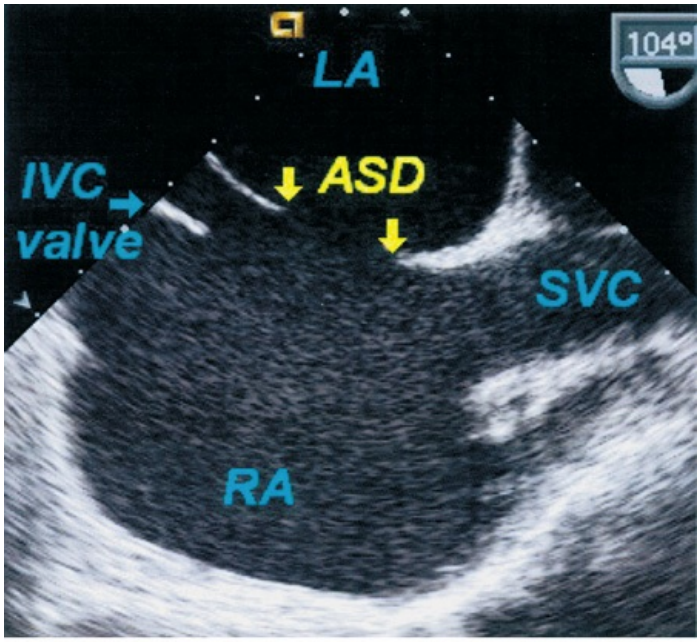

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

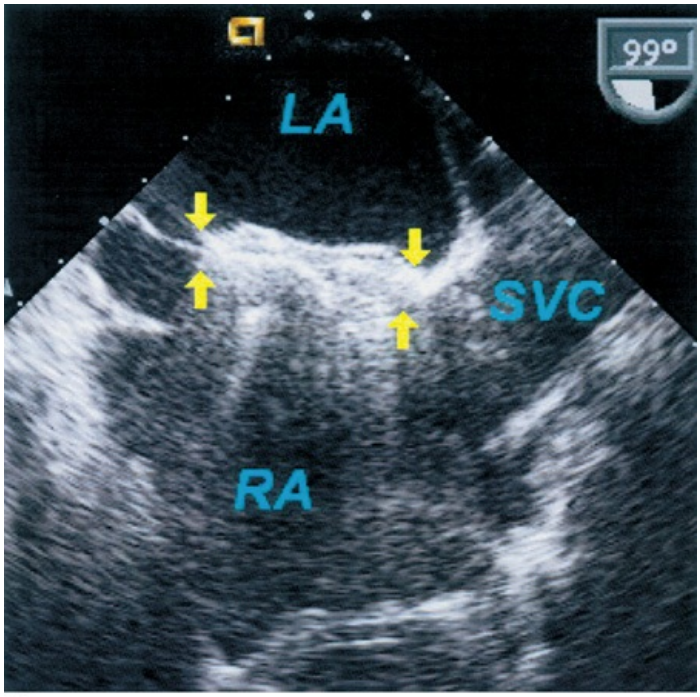

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Secundum atrial septal defect. Transesophageal echocardiogram of secundum ASD and device closure. ***A.*** The atrial septal defect (ASD) between the left atrium (LA) and right atrium (RA) is shown. ***B.*** A percutaneous catheter–delivered device has occluded the defect. IVC, inferior vena cava; SVC, superior vena cava.

TREATMENT: ATRIAL SEPTAL DEFECT

Operative repair, usually with a patch of pericardium or of prosthetic material or percutaneous transcatheter device closure, if the ASD is of an appropriate size and shape, should be advised for all patients with uncomplicated secundum ASD with significant left-to-right shunting, i.e., pulmonary-to-systemic flow ratios ≥2:1. Excellent results may be anticipated, at low risk, even in patients >40 years, in the absence of severe pulmonary hypertension. In ostium primum ASD, cleft mitral valves may require repair in addition to patch closure of the ASD. Closure should not be carried out in patients with small defects and trivial left-to-right shunts or in those with severe pulmonary vascular disease without a significant left-to-right shunt.

Patients with sinus venosus or ostium secundum ASDs rarely die before the fifth decade. During the fifth and sixth decades, the incidence of progressive symptoms, often leading to severe disability, increases substantially. Medical management should include prompt treatment of respiratory-tract infections; anti-arrhythmic medications for atrial fibrillation or supraventricular tachycardia; and the usual measures for hypertension, coronary disease, or heart failure (Chap. 234), if these complications occur. The risk of infective endocarditis is quite low unless the defect is complicated by valvular regurgitation or has recently been repaired with a patch or device (Chap. 124).

VENTRICULAR SEPTAL DEFECT

A VSD is one of the most common of all cardiac birth defects, either as isolated defects or as a component of a combination of anomalies. The VSD is usually single and situated in the membranous or midmuscular portion of the septum. The functional disturbance depends on its size and on the status of the pulmonary vascular bed. Only small- or moderate-size VSDs are seen initially in adulthood, as most patients with an isolated large VSD come to medical or surgical attention early in life.

A wide spectrum exists in the natural history of VSD, ranging from spontaneous closure to congestive cardiac failure and death in infancy. Included within this spectrum are the possible development of pulmonary vascular obstruction, RV outflow tract obstruction, aortic regurgitation, or infective endocarditis. Spontaneous closure is more common in patients born with a small VSD, which occurs in early childhood in most. The pulmonary vascular bed is often a principal determinant of the clinical manifestations and course of a given VSD and feasibility of surgical repair. Increased pulmonary arterial pressure results from increased pulmonary blood flow and/or resistance, the latter usually the result of obstructive, obliterative structural changes within the pulmonary vascular bed. It is important to quantitate and compare pulmonary-to-systemic flows and resistances in patients with severe pulmonary hypertension. The term *Eisenmenger's syndrome* is applied to patients with a large

communication between the two circulations at the aortopulmonary, ventricular, or atrial levels and bidirectional or predominantly right-to-left shunts because of high resistance and obstructive pulmonary hypertension.

Patients with large VSDs and pulmonary hypertension are at greatest risk for developing pulmonary vascular obstruction. Large VSDs should be corrected surgically early in life when pulmonary vascular disease is still reversible or not yet developed. In patients with Eisenmenger syndrome, symptoms in adult life consist of exertional dyspnea, chest pain, syncope, and hemoptysis. The right-to-left shunt leads to cyanosis, clubbing, and erythrocytosis (see below). The degree to which pulmonary vascular resistance is elevated before operation is a critical factor determining prognosis. If the pulmonary vascular resistance is one-third or less of the systemic value, progression of pulmonary vascular disease after operation is unusual; however, if a moderate to severe increase in pulmonary vascular resistance exists preoperatively, either no change or a progression of pulmonary vascular disease is common postoperatively. Pregnancy is contraindicated in Eisenmenger syndrome. The mother's health is most at risk if she has a cardiovascular lesion associated with pulmonary vascular disease and pulmonary hypertension (e.g., Eisenmenger physiology or mitral stenosis) or LV outflow tract obstruction (e.g., aortic stenosis), but she is also at risk of death with any malformation that may cause heart failure or a hemodynamically important arrhythmia. The fetus is most at risk with maternal cyanosis, heart failure, or pulmonary hypertension.

RV outflow tract obstruction develops in ~5–10% of patients who present in infancy with a moderate to large left-to-right shunt. With time, as subvalvular RV outflow tract obstruction progresses, the findings in these patients whose VSD remains sizable begin to resemble more closely those of the cyanotic tetralogy of Fallot. In ~5% of patients, aortic valve regurgitation results from insufficient cusp tissue or prolapse of the cusp through the interventricular defect; the aortic regurgitation then complicates and dominates the clinical course. Two-dimensional *echocardiography* with spectral and color Doppler examination defines the number and location of defects in the ventricular septum and associated anomalies and the hemodynamic physiology of the defect(s). Hemodynamic and angiographic study may be occasionally required to assess the status of the pulmonary vascular bed and clarify details of the altered anatomy.

Treatment: Ventricular Septal Defect

Surgery is not recommended for patients with normal pulmonary arterial pressures with small shunts (pulmonary-to-systemic flow ratios of <1.5 to 2:1). Operative correction or transcatheter closure is indicated when there is a moderate to large left-to-right shunt with a pulmonary-to-systemic flow ratio >1.5:1 or 2:1, in the absence of prohibitively high levels of pulmonary vascular resistance.

In the Eisenmenger VSD patient, pulmonary arterial vasodilators and both single-lung transplantation with intracardiac defect repair and total heart-lung transplantation show promise for improvement in symptoms (Chaps. 235 and 266). Chronic hypoxemia in cyanotic CHD results in secondary *erythrocytosis* due to increased erythropoietin production (Chap. 35). The term *polycythemia* is a misnomer; white cell counts are normal and platelet counts are normal to decreased. Compensated erythrocytosis with iron-replete equilibrium hematocrits rarely result in symptoms of hyperviscosity at hematocrits < 65% and occasionally not even with hematocrits ≥ 70%. For this reason, therapeutic phlebotomy is rarely required in compensated erythrocytosis. In contrast, patients with decompensated erythrocytosis fail to establish equilibrium with unstable, rising hematocrits and recurrent hyperviscosity symptoms. Therapeutic phlebotomy, a two-edged sword, allows temporary relief of symptoms but limits oxygen delivery, begets instability of the hematocrit, and compounds the problem by iron depletion. Iron-deficiency symptoms are usually indistinguishable from those of hyperviscosity; progressive symptoms after recurrent phlebotomy are usually due to iron depletion with hypochromic microcytosis. Iron depletion results in a larger number of smaller (microcytic) hypochromic red cells that are less capable of carrying oxygen and less deformable in the microcirculation; with more of them relative to plasma volume, viscosity is greater than for an equivalent hematocrit with fewer, larger, iron-replete, deformable cells. As such, iron-depleted erythrocytosis results in increasing symptoms due to decreased oxygen delivery to the tissues.

Hemostasis is abnormal in cyanotic CHD, due, in part, to the increased blood volume and engorged capillaries, abnormalities in platelet function, and sensitivity to aspirin or nonsteroidal anti-inflammatory agents, as well as abnormalities of the extrinsic and intrinsic coagulation system. Oral contraceptives are often contraindicated for cyanotic women because of the enhanced risk of vascular thrombosis. Adults with cyanotic CHD do not appear to be at increased risk for stroke unless there are excessive injudicious phlebotomies, inappropriate use of aspirin or anticoagulants, or the presence of atrial arrhythmias or infective endocarditis. Symptoms of hyperviscosity can be produced in any cyanotic patient with erythrocytosis if dehydration reduces plasma volume. Phlebotomy for symptoms of hyperviscosity not due to dehydration or iron deficiency is a simple outpatient removal of 500 mL of blood over 45 min with isovolumetric replacement with isotonic saline. Acute phlebotomy without volume replacement is contraindicated. Iron repletion in decompensated iron-depleted erythrocytosis reduces iron-deficiency symptoms, but must be done gradually to avoid an excessive rise in hematocrit and resulting hyperviscosity.

PATENT DUCTUS ARTERIOSUS

The ductus arteriosus is a vessel leading from the bifurcation of the pulmonary artery to the aorta just distal to the left subclavian artery. Normally, the vascular channel is open in the fetus but closes immediately after birth. The flow across the ductus is determined by the pressure and resistance relationships between the systemic and pulmonary circulations and by the cross-sectional area and length of the ductus. In most adults with this anomaly, pulmonary pressures are normal and a gradient and shunt from aorta to pulmonary artery persist throughout the cardiac cycle, resulting in a characteristic thrill and a continuous "machinery" murmur with late systolic accentuation at the upper left sternal edge. In adults who were born with a large left-to-right shunt through the ductus arteriosus, pulmonary vascular obstruction (Eisenmenger syndrome) with pulmonary hypertension, right-to-left shunting, and cyanosis have usually developed. Severe pulmonary vascular disease results in reversal of flow through the ductus; unoxygenated blood is shunted to the descending aorta; and the toes—but not the fingers—become cyanotic and clubbed, a finding termed *differential cyanosis*. The leading causes of death in adults with patent ductus are cardiac failure and infective endocarditis; occasionally severe pulmonary vascular obstruction may cause aneurysmal dilatation, calcification, and rupture of the ductus.

Treatment: Patent Ductus Arteriosus

In the absence of severe pulmonary vascular disease and predominant left-to-right shunting of blood, the patent ductus should be surgically ligated or divided. Transcatheter closure using coils, buttons, plugs, and umbrellas has become commonplace for appropriately shaped defects. Thoracoscopic surgical approaches are considered experimental. Operation should be deferred for several months in patients treated successfully for infective endocarditis because the ductus may remain somewhat edematous and friable.

AORTIC ROOT–TO-RIGHT-HEART SHUNTS

The three most common causes of aortic root–to-right-heart shunts are congenital aneurysm of an aortic sinus of Valsalva with fistula, coronary arteriovenous fistula, and anomalous origin of the left coronary artery from the pulmonary trunk. *Aneurysm of an aortic sinus of Valsalva* consists of a separation or lack of fusion between the media of the aorta and the annulus of the aortic valve. Rupture usually occurs in the third or fourth decade of life; most often, the aorticocardiac fistula is between the right coronary cusp and the right ventricle; but occasionally, when the noncoronary cusp is involved, the fistula drains into the right atrium. Abrupt rupture causes chest pain, bounding pulses, a continuous murmur accentuated in diastole, and volume overload of the heart. Diagnosis is confirmed by two-dimensional and Doppler echocardiographic studies; cardiac catheterization quantitates the left-to-right shunt, and thoracic aortography visualizes the fistula. Medical management is directed at cardiac failure, arrhythmias, or endocarditis. At operation, the aneurysm is closed and amputated, and the aortic wall is reunited with the heart, either by direct suture or with a patch or prosthesis.

Coronary arteriovenous fistula, an unusual anomaly, consists of a communication between a coronary artery and another cardiac chamber, usually the coronary sinus, right atrium, or right ventricle. The shunt is usually of small magnitude and myocardial blood flow is not usually compromised; if the shunt is large, there may be a coronary "steal" syndrome with myocardial ischemia and possible angina or ventricular arrhythmias. Potential complications include infective endocarditis; thrombus formation with occlusion or distal embolization with myocardial infarction; rupture of an aneurysmal fistula; and, rarely, pulmonary hypertension and congestive failure. A loud, superficial, continuous murmur at the lower or midsternal border usually prompts a further evaluation of asymptomatic patients. Doppler echocardiography demonstrates the site of drainage; if the site of origin is proximal, it may be detectable by two-dimensional echocardiography. Angiography (classic catheterization, CT, or magnetic resonance angiography) permits identification of the size and anatomic features of the fistulous tract, which may be closed by suture or transcatheter obliteration.

The third anomaly causing a shunt from the aortic root to the right heart is *anomalous origin of the left coronary artery from the pulmonary artery*. Myocardial infarction and fibrosis commonly lead to death within the first year, although up to 20% of patients survive to adolescence and beyond without surgical correction. The diagnosis is supported by the ECG findings of an anterolateral myocardial infarction and left ventricular hypertrophy (LVH). Operative management of adults consists of coronary artery bypass with an internal mammary artery graft or saphenous vein–coronary artery graft.

CONGENITAL AORTIC STENOSIS

Malformations that cause obstruction to LV outflow include congenital valvular aortic stenosis, discrete subaortic stenosis, or supravalvular aortic stenosis. Bicuspid aortic valves are more common in males than in females. The congenital bicuspid aortic valve, which may initially be functionally normal, is one of the most common congenital malformations of the heart and may go undetected in early life. Because bicuspid valves may develop stenosis or regurgitation with time or be the site of infective endocarditis, the lesion may be difficult to distinguish in older adults from acquired rheumatic or degenerative calcific aortic valve disease. The dynamics of blood flow associated with a congenitally deformed, rigid aortic valve commonly lead to thickening of the cusps and, in later life, to calcification.

Hemodynamically significant obstruction causes concentric hypertrophy of the LV wall. The ascending aorta is often dilated, misnamed "poststenotic" dilatation; this is due to histologic abnormalities of the aortic media similar to those in Marfan's syndrome and may result in aortic dissection. Diagnosis is made by echocardiography, which reveals the morphology of the aortic valve and aortic root and quantitates severity of stenosis or regurgitation. **The clinical manifestations and hemodynamic abnormalities are discussed in Chap. 237.**

Treatment: Valvular Aortic Stenosis

Medical management includes prophylaxis against infective endocarditis and, in patients with diminished cardiac reserve, the administration of digoxin and diuretics and sodium restriction while awaiting operation. A dilated aortic root may require beta blockers. Aortic valve replacement is indicated in adults with critical obstruction, i.e., with an aortic valve area <0.45 cm^2/m^2, with symptoms secondary to LV dysfunction or myocardial ischemia, or with hemodynamic evidence of LV dysfunction. In asymptomatic children or adolescents or young adults with critical aortic stenosis without valvular calcification or these features, aortic balloon valvuloplasty is often useful (Chap. 246). If surgery is contraindicated in older patients because of a complicating medical problem such as malignancy or renal or hepatic failure, balloon valvuloplasty may provide short-term improvement. This procedure may serve as a bridge to aortic valve replacement in patients with severe heart failure.

Subaortic Stenosis

The *discrete* form of subaortic stenosis consists of a membranous diaphragm or fibromuscular ring encircling the LV outflow tract just beneath the base of the aortic valve. The jet impact from the subaortic stenotic jet on the underside of the aortic valve often begets progressive aortic valve fibrosis and valvular regurgitation. Echocardiography demonstrates the anatomy of the subaortic obstruction; Doppler studies show turbulence proximal to the aortic valve and can quantitate the pressure gradient and severity of aortic regurgitation. Treatment consists of complete excision of the membrane or fibromuscular ring.

Supravalvular Aortic Stenosis

This is a localized or diffuse narrowing of the ascending aorta originating just above the level of the coronary arteries at the superior margin of the sinuses of Valsalva. In contrast to other forms of aortic stenosis, the coronary arteries are subjected to elevated systolic pressures from the left ventricle, are often dilated and tortuous, and are susceptible to premature atherosclerosis. In most patients, a genetic defect for the anomaly is located in the same chromosomal region as elastin on chromosome 7. Supravalvular aortic stenosis is the most commonly associated cardiac defect in Williams-Beuren syndrome, typically comprising the following: "elfin" facies, low nasal bridge, cheerful demeanor, mental retardation with retained language skills and love of music, supravalvular aortic stenosis, and transient hypercalcemia.

COARCTATION OF THE AORTA

Narrowing or constriction of the lumen of the aorta may occur anywhere along its length but is most common distal to the origin of the left subclavian artery near the insertion of the ligamentum arteriosum. Coarctation occurs in ~7% of patients with congenital heart disease, is more common in males than females, and is particularly frequent in patients with gonadal dysgenesis (e.g., Turner syndrome). Clinical manifestations depend on the site and extent of obstruction and the presence of associated cardiac anomalies; most commonly a bicuspid aortic valve. Circle of Willis aneurysms may occur in up to 10%, and pose a high risk of sudden rupture and death.

Most children and young adults with isolated, discrete coarctation are asymptomatic. Headache, epistaxis, cold extremities, and claudication with exercise may occur, and attention is usually directed to the cardiovascular system when a heart murmur or hypertension in the upper extremities and absence, marked diminution, or delayed pulsations in the femoral arteries are detected on physical examination. Enlarged and pulsatile collateral vessels may be palpated in the intercostal spaces anteriorly, in the axillae, or posteriorly in the interscapular area. The upper extremities and thorax may be more developed than the lower extremities. A midsystolic murmur over the left interscapular space may become continuous if the lumen is narrowed sufficiently to result in a high- velocity jet across the lesion throughout the cardiac cycle. Additional systolic and continuous murmurs over the lateral thoracic wall may reflect increased flow through dilated and tortuous collateral vessels. The ECG usually reveals LV hypertrophy. Chest x-ray may show a dilated left subclavian artery high on the left mediastinal border and a dilated ascending aorta. Indentation of the aorta at the site of coarctation and pre- and poststenotic dilatation (the "3" sign) along the left paramediastinal shadow are essentially pathognomonic. Notching of the third to ninth ribs, an important radiographic sign, is due to inferior rib erosion by dilated collateral vessels **(Figs. 236-2** and **236-3)**. Two-dimensional echocardiography from suprasternal windows identifies the site of coarctation; Doppler quantitates the pressure gradient. Transesophageal echocardiography and MRI or three-dimensional CT allow visualization of the length and severity of the obstruction and associated collateral arteries (Figs. 236-2 and 236-3). In adults, cardiac catheterization is indicated primarily to evaluate the coronary arteries or to perform catheter-based intervention (angioplasty and stent of the coarctation).

Figure 236-2

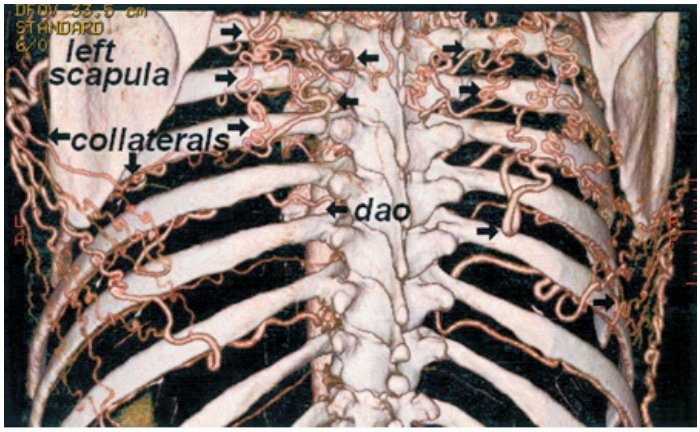

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Aortic coarctation. The extensive collaterals (*left*) underneath the ribs and in the periscapular region are shown on a posterior view of a three-dimensional CT angiogram, which are responsible for rib notching on chest x-ray. dao, descending aorta.

Figure 236-3

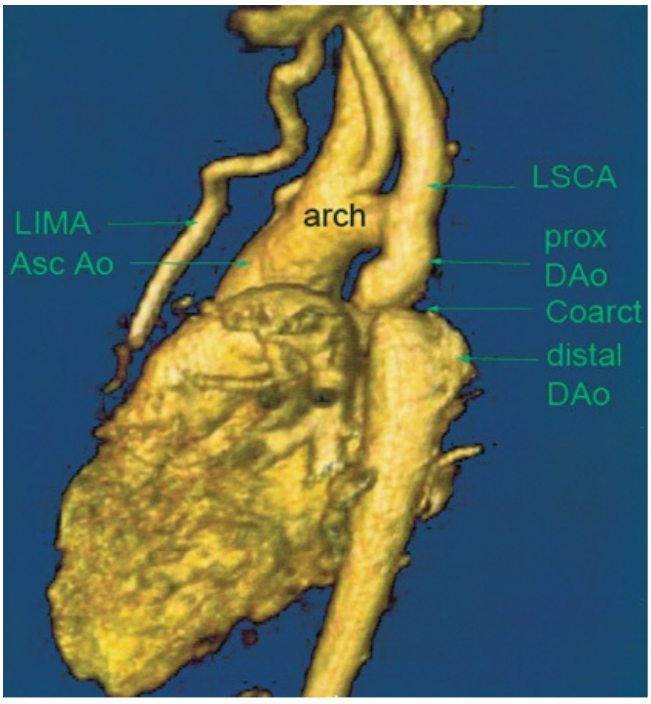

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Aortic coarctation. The coarctation (Coarct) of the aorta is shown in the typical "adult" location in the descending aorta (DAo) just distal to the dilated left subclavian artery (LSCA) in this three-dimensional reconstruction of an MR angiogram. There is a post-coarct aneurysm that is in part due to intrinsic aortic medial tissue weakness. The left internal mammary artery (LIMA) is dilated. Asc Ao, ascending aorta; prox, proximal.

The chief hazards of proximal aortic severe hypertension include cerebral aneurysms and hemorrhage, aortic dissection and rupture, premature coronary arteriosclerosis, and LV failure; infective endarteritis may occur on the coarctation site or endocarditis may settle on an associated bicuspid aortic valve, which is estimated to be present in up to 75% of patients.

Treatment: Coarctation of the Aorta

Treatment is surgical or involves percutaneous catheter balloon dilatation with stent placement; the

details of selection of therapy are beyond this review. Late postoperative systemic hypertension in the absence of residual coarctation is related partly to the duration of preoperative hypertension. Follow-up of rest and exercise blood pressures is important; many have systolic hypertension only during exercise, in part due to a diffuse vasculopathy. All operated or stented coarctation patients deserve a high-quality MRI or CT procedure in follow-up.

PULMONARY STENOSIS WITH INTACT VENTRICULAR SEPTUM

Obstruction to RV outflow may be localized to the supravalvular, valvular, or subvalvular levels or occur at a combination of these sites. Multiple sites of narrowing of the peripheral pulmonary arteries are a feature of *rubella embryopathy* and may occur with both the familial and sporadic forms of supravalvular aortic stenosis. Valvular pulmonic stenosis (PS) is the most common form of isolated RV obstruction.

The severity of the obstructing lesion, rather than the site of narrowing, is the most important determinant of the clinical course. In the presence of a normal cardiac output, a peak systolic pressure gradient <30 mmHg indicates mild PS and >50 mmHg indicates severe PS; processes between these limits are considered to indicate moderate stenosis. Patients with mild PS are generally asymptomatic and demonstrate little or no progression in the severity of obstruction with age. In patients with more significant stenosis, the severity may increase with time. Symptoms vary with the degree of obstruction. Fatigue, dyspnea, RV failure, and syncope may limit the activity of older patients, in whom moderate or severe obstruction may prevent an augmentation of cardiac output with exercise. In patients with severe obstruction, the systolic pressure in the RV may exceed that in the LV, because the ventricular septum is intact. RV ejection is prolonged with moderate or severe stenosis, and the sound of pulmonary valve closure is delayed and soft. RV hypertrophy reduces the compliance of that chamber, and a forceful RA contraction is necessary to augment RV filling. A fourth heart sound; prominent *a* waves in the jugular venous pulse; and, occasionally, presystolic pulsations of the liver reflect vigorous atrial contraction. The clinical diagnosis is supported by a left parasternal lift and harsh systolic crescendo- decrescendo murmur and thrill at the upper left sternal border, typically preceded by a systolic ejection sound if the obstruction is due to a mobile nondysplastic pulmonary valve. The holosystolic murmur of tricuspid regurgitation may accompany severe PS, especially in the presence of congestive heart failure. Cyanosis usually reflects right-to-left shunting through a patent foramen ovale or ASD. In patients with supravalvular or peripheral pulmonary arterial stenosis, the murmur is systolic or continuous and is best heard over the area of narrowing, with radiation to the peripheral lung fields.

In mild cases, the ECG is normal, whereas moderate and severe stenoses are associated with RV hypertrophy. The chest x-ray with mild or moderate PS shows a heart of normal size with normal lung vascularity. In pulmonary valvular stenosis, dilatation of the main and left pulmonary arteries occurs in part due to the direction of the PS jet and in part due to intrinsic tissue weakness. With severe obstruction, RV hypertrophy is generally evident. The pulmonary vascularity may be reduced with severe stenosis, RV failure, and/or a right-to-left shunt at the atrial level. Two-dimensional echocardiography visualizes pulmonary-valve morphology; the outflow tract pressure gradient is quantitated by Doppler echocardiography.

Treatment: Pulmonary Stenosis

The cardiac catheter technique of balloon valvuloplasty (Chap. 230) is usually effective. Direct surgical relief of moderate and severe obstruction may be accomplished at a low risk. Multiple stenoses of the peripheral pulmonary arteries are usually inoperable, but narrowing of a proximal branch or at the bifurcation of the main pulmonary trunk may be surgically corrected or undergo balloon dilatation and stenting.

TETRALOGY OF FALLOT

The four components of the tetralogy of Fallot are malaligned VSD, obstruction to RV outflow, aortic override of the VSD, and RV hypertrophy due to the RV's response to aortic pressure via the large VSD **(Fig. 236-4)**.

Figure 236-4

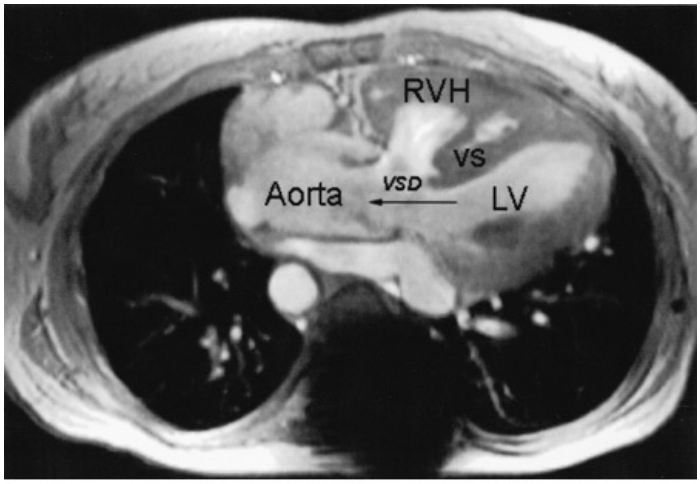

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tetralogy of Fallot. Magnetic resonance angiogram. A mid-systolic frame showing the malaligned ventricular septal defect (VSD) with the aorta overriding the ventricular septal defect. LV, left ventricle; RVH, RV hypertrophy; VS, ventricular septum.

The severity of RV outflow obstruction determines the clinical presentation. The severity of hypoplasia of the RV outflow tract varies from mild to complete (pulmonary atresia). Pulmonary valve stenosis and supravalvular and peripheral pulmonary arterial obstruction may coexist; rarely, there is unilateral absence of a pulmonary artery (usually the left). A right-sided aortic arch and descending thoracic aorta occur in ~25%.

The relationship between the resistance of blood flow from the ventricles into the aorta and into the pulmonary artery plays a major role in determining the hemodynamic and clinical picture. When the RV outflow obstruction is severe, pulmonary blood flow is reduced markedly, and a large volume of desaturated systemic venous blood shunts right-to-left across the VSD. Severe cyanosis and erythrocytosis occur, and symptoms of systemic hypoxemia are prominent. In many infants and children, the obstruction is mild but progressive.

The ECG shows RV hypertrophy. Chest x-ray shows a normal-sized, boot-shaped heart (*coeur en sabot*) with a prominent right ventricle and a concavity in the region of the pulmonary conus. Pulmonary vascular markings are typically diminished, and the aortic arch and knob may be on the right side. Two-dimensional echocardiography demonstrates the malaligned VSD with the overriding aorta and the site and severity of PS, which may be subpulmonic (fixed or dynamic), at the pulmonary valve or in the main or branch pulmonary arteries. Classic contrast angiography may provide details regarding the RV outflow tract, pulmonary valve and annulus, and caliber of the main branches of the pulmonary artery as well of possible associated aortopulmonary collaterals. Coronary arteriography identifies the anatomy and course of the coronary arteries. In experienced centers, these issues are often well demonstrated in adults by MRI (Fig. 236-4) or CT angiography with three-dimensional reconstruction.

Treatment: Tetralogy of Fallot

For a variety of reasons, only a few adults with tetralogy of Fallot have not had some form of previous surgical intervention. Reoperation in adults is most commonly for severe pulmonary regurgitation. Long-term concerns about ventricular function persist. Ventricular and atrial arrhythmias may require medical treatment or electrophysiologic study and ablation. Interventional catheterization may be needed in selected patients (i.e., angioplasty and stenting of branch pulmonary stenosis). The aortic root has a medial tissue defect; it is commonly enlarged and associated with aortic regurgitation. Endocarditis remains a risk despite surgical repair.

COMPLETE TRANSPOSITION OF THE GREAT ARTERIES

This condition is commonly called *dextro-* or *d-transposition of the great arteries*. The aorta arises rightward anteriorly from the right ventricle, and the pulmonary artery emerges leftward and posteriorly from the LV, which results in two separate parallel circulations; some communication between them must exist after birth to sustain life. Most patients have an interatrial communication, two-thirds have a patent ductus arteriosus, and about one-third have an associated VSD. Transposition is more common in males and accounts for ~10% of cyanotic heart disease. The course is determined by the degree of tissue hypoxemia, the ability of each ventricle to sustain an increased workload in the presence of reduced coronary arterial oxygenation, the nature of the associated cardiovascular anomalies, and the status of the pulmonary vascular bed. By the third decade of life, ~30% of patients will have developed decreased RV function and progressive tricuspid regurgitation, which may lead to congestive heart failure. Pulmonary vascular obstruction develops by one to two years of age in patients with an

associated large VSD or large patent ductus arteriosus in the absence of obstruction to LV outflow.

Treatment: Transposition of the Great Arteries

The balloon or blade catheter or surgical creation or enlargement of an interatrial communication in the neonate is the simplest procedure for providing increased intracardiac mixing of systemic and pulmonary venous blood. Systemic pulmonary–artery anastomosis may be indicated in the patient with severe obstruction to LV outflow and diminished pulmonary blood flow. Intracardiac repair may be accomplished by rearranging the venous returns (intra-atrial switch, i.e., Mustard or Senning operation) so that the systemic venous blood is directed to the mitral valve and, thence, to the left ventricle and pulmonary artery, while the pulmonary venous blood is diverted through the tricuspid valve and right ventricle to the aorta. The late survival after these repairs is good, but arrhythmias (e.g., atrial flutter) or conduction defects (e.g., sick sinus syndrome) occur in ~50% of such patients by 30 years after the intra-atrial switch surgery. Progressive dysfunction of the systemic subaortic right ventricle, tricuspid regurgitation, ventricular arrhythmias, or cardiac arrest and late sudden death are worrisome features. Preferably, this malformation is corrected in infancy by transposing both coronary arteries to the posterior artery and transecting, contraposing, and anastomosing the aorta and pulmonary arteries (arterial-switch operation). For those patients with a VSD in whom it is necessary to bypass a severely obstructed LV outflow tract, corrective operation employs an intracardiac ventricular baffle and extracardiac prosthetic conduit to replace the pulmonary artery (Rastelli procedure).

SINGLE VENTRICLE

This is a family of complex lesions with both atrioventricular valves or a common atrioventricular valve opening to a single ventricular chamber. Associated anomalies include abnormal great artery positional relationships, pulmonic valvular or subvalvular stenosis, and subaortic stenosis. Survival to adulthood depends on a relatively normal pulmonary blood flow, yet normal pulmonary resistance and good ventricular function. Modifications of the Fontan approach are generally applied to carefully selected patients with creation of a pathway(s) from the systemic veins to the pulmonary arteries.

TRICUSPID ATRESIA

This malformation is characterized by atresia of the tricuspid valve; an interatrial communication; and, frequently, hypoplasia of the right ventricle and pulmonary artery. The clinical picture is usually dominated by severe cyanosis due to obligatory admixture of systemic and pulmonary venous blood in the left ventricle. The ECG characteristically shows RA enlargement, left-axis deviation, and LV hypertrophy.

Atrial septostomy and palliative operations to increase pulmonary blood flow, often by anastomosis of a systemic artery or vein to a pulmonary artery, may allow survival to the second or third decade. A Fontan atriopulmonary or total cavopulmonary connection may then allow functional correction in those patients with normal or low pulmonary arterial resistance pressure and good LV function.

EBSTEIN'S ANOMALY

Characterized by a downward displacement of the tricuspid valve into the right ventricle, due to anomalous attachment of the tricuspid leaflets, the Ebstein tricuspid valve tissue is dysplastic and results in tricuspid regurgitation. The abnormally situated tricuspid orifice produces an "atrialized" portion of the RV lying between the atrioventricular ring and the origin of the valve, which is continuous with the RA chamber. Often, the RV is hypoplastic. Although the clinical manifestations are variable, some patients come to initial attention because of either (1) progressive cyanosis from right-to-left atrial shunting, (2) symptoms due to tricuspid regurgitation and RV dysfunction, or (3) paroxysmal atrial tachyarrhythmias with or without atrioventricular bypass tracts [Wolff-Parkinson-White (WPW) syndrome]. Diagnostic findings by two-dimensional echocardiography include the abnormal positional relation between the tricuspid and mitral valves with abnormally increased apical displacement of the septal tricuspid leaflet. Tricuspid regurgitation is quantitated by Doppler examination. Surgical approaches include prosthetic replacement of the tricuspid valve when the leaflets are tethered or repair of the native valve.

CONGENITALLY CORRECTED TRANSPOSITION

The two fundamental anatomic abnormalities in this malformation are transposition of the ascending aorta and pulmonary trunk and inversion of the ventricles. This arrangement results in desaturated systemic venous blood passing from the right atrium through the mitral valve to the LV and into the pulmonary trunk, whereas oxygenated pulmonary venous blood flows from the left atrium through the tricuspid valve to the RV and into the aorta. Thus, the circulation is corrected functionally. The clinical presentation, course, and prognosis of patients with congenitally corrected transposition vary depending on the nature and severity of any complicating intracardiac anomalies and of development of dysfunction of the systemic subaortic RV. Progressive RV dysfunction and tricuspid regurgitation may also develop in one-third of patients by age 30; Ebstein-type anomalies of the left-side tricuspid atrioventricular valve are common. VSD or PS due to obstruction to outflow from the right-sided subpulmonary (anatomic left) ventricle may coexist. Complete heart block occurs at a rate of 2–10% per decade. The diagnosis of the malformation and associated lesions can be established by comprehensive two-dimensional echocardiography and Doppler examination.

MALPOSITIONS OF THE HEART

Positional anomalies refer to conditions in which the cardiac apex is in the right side of the chest (*dextrocardia*), or at the midline (*mesocardia*), or in which there is a normal location of the heart in the left side of the chest but abnormal position of the viscera (*isolated levocardia*). Knowledge of the position of the abdominal organs and of the branching pattern of the main stem bronchi is important in categorizing these malpositions. When dextrocardia occurs without situs inversus, when the visceral situs is indeterminate, or if isolated levocardia is present, associated, often complex, multiple cardiac anomalies are usually present. In contrast, mirror-image dextrocardia is usually observed with complete situs inversus, which occurs most frequently in individuals whose hearts are otherwise normal.

Surgically Modified Congenital Heart Disease

Owing to the enormous strides in cardiovascular surgical techniques that have occurred in the past 60 years, a large number of long-term survivors of corrective operations in infancy and childhood have reached adulthood. These patients are often challenging because of the diversity of anatomic, hemodynamic, and electrophysiologic residua and sequelae of cardiac operations.

The proper care of the survivor of operation for CHD requires that the clinician understand the details of the malformation before operation; pay meticulous attention to the details of the operative procedure; and recognize the postoperative residua (conditions left totally or partially uncorrected), the sequelae (conditions caused by surgery), and the complications that may have resulted from the operation. Except for ligation of an uncomplicated patent ductus arteriosus, almost every other surgical repair leaves behind or causes some abnormality of the heart and circulation that may range from trivial to serious. Intraoperative transesophageal echocardiography assists in detecting unsuspected lesions, in monitoring the repair, and in verifying a satisfactory result or directing further repair. Thus, even with results that are considered clinically to be good to excellent, continued long-term postoperative follow-up is advisable.

Cardiac operations importantly involving the atria, such as closure of ASD, repair of total or partial anomalous pulmonary venous return, or venous switch corrections of complete transposition of the great arteries (the Mustard or Senning operations), may be followed years later by sinus node or atrioventricular node dysfunction or by atrial arrhythmias (especially atrial flutter). Intraventricular surgery may also result in electrophysiologic consequences, including complete heart block necessitating pacemaker insertion to avoid sudden death. Valvular problems may arise late after initial cardiac operation. An example is the progressive stenosis of an initially nonobstructive bicuspid aortic valve in the patient who underwent aortic coarctation repair. Such aortic valves may also be the site of infective endocarditis. After repair of the ostium primum ASD, the cleft mitral valve may become progressively regurgitant. Tricuspid regurgitation may also be progressive in the postoperative patient with tetralogy of Fallot if RV outflow tract obstruction was not relieved adequately at initial surgery. In many patients with surgically modified CHD, inadequate relief of an obstructive lesion, or a residual regurgitant lesion, or a residual shunt will cause or hasten the onset of clinical signs and symptoms of myocardial dysfunction. Despite a good hemodynamic repair, many patients with a subaortic RV develop RV decompensation and signs of left heart failure. In many patients, particularly those who were cyanotic for many years before operation, a preexisting compromise in ventricular performance is due to the original underlying malformation.

A final category of postoperative problems involves the use of prosthetic valves, patches, or conduits in the operative repair. The special risks include infective endocarditis, thrombus formation, and premature degeneration and calcification of the prosthetic materials. There are many patients in whom extracardiac conduits are required to correct the circulation functionally and often to carry blood to the lungs from the right atrium or right ventricle. These conduits may develop intraluminal obstruction, and, if they include a prosthetic valve, it may show progressive calcification and thickening. Many such patients face reintervention (interventional cardiac catheterization or surgical reoperation) one or more times in their lives. Such care should be directed to centers specializing in adults with complex congenital cardiovascular malformations. The effect of pregnancy in postoperative patients depends on the outcome of the repair, including the presence and severity of residua, sequelae, or complications. Contraception is an important topic with such patients. Tubal ligation should be considered in those in whom pregnancy is strictly contraindicated.

ENDOCARDITIS PROPHYLAXIS

Two major predisposing causes of infective endocarditis are a susceptible cardiovascular substrate and a source of bacteremia. The clinical and bacteriologic profile of infective endocarditis in patients with CHD has changed with the advent of intracardiac surgery and of prosthetic devices. Prophylaxis includes both antimicrobial and hygienic measures. Meticulous dental and skin care are required. Routine antimicrobial prophylaxis is recommended for bacteremic dental procedures or instrumentation through an infected site in most patients with operated CHD, particularly if foreign material, such as a prosthetic valve, conduit, surgically constructed shunt, etc., is in place. In the case of patches, in the absence of a high-pressure patch leak, prophylaxis is usually recommended for six months until there is endothelialization. Individuals with unrepaired cyanotic heart disease are also generally

recommended to receive prophylaxis (Chap. 124).

FURTHER READINGS

Aboulhosn J, Child JS: *Congenital Heart Disease in Adults*, in Hurst's The Heart 12th ed, V Fuster et al (eds), New York, McGraw Medical, 2008, pp 1922–1948

———: *Congenital Heart Disease in Adults*, in Hurst's The Heart Manual of Cardiology 12th ed, V Fuster et al (eds), New York, McGraw Medical, 2009, pp 546–556

Perloff JK, et al: *Congenital Heart Disease in Adults* 3rd ed., Philadelphia, Saunders Elsevier, 2009, pp 1–488

Warnes CA, et al: ACC/AHA 2008 Guidelines for the Management of Adults with Congenital Heart Disease. Circulation 118:e714 2008

Webb G et al: Congenital heart disease, in *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 7th ed., D. Zipes, et al. (eds) Philadelphia, Elsevier Saunders, 2005, pp 1489–1552

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VALVULAR HEART DISEASE: INTRODUCTION

The role of the physical examination in the evaluation of patients with valvular heart disease is also considered in Chaps. e13 and 227; of electrocardiography (ECG) in Chap. 228; of echocardiography and other noninvasive imaging techniques in Chap. 229; and of cardiac catheterization and angiography in Chap. 230.

MITRAL STENOSIS

Etiology and Pathology

Rheumatic fever is the leading cause of mitral stenosis (MS) **(Table 237-1)**. Other less common etiologies of obstruction to left atrial outflow include congenital mitral valve stenosis, cor triatriatum, mitral annular calcification with extension onto the leaflets, systemic lupus erythematosus, rheumatoid arthritis, left atrial myxoma, and infective endocarditis with large vegetations. Pure or predominant MS occurs in approximately 40% of all patients with rheumatic heart disease and a history of rheumatic fever (Chap. 322). In other patients with rheumatic heart disease, lesser degrees of MS may accompany mitral regurgitation (MR) and aortic valve disease. With reductions in the incidence of acute rheumatic fever, particularly in temperate climates and developed countries, the incidence of MS has declined considerably over the past few decades. However, it remains a major problem in developing nations, especially in tropical and semitropical climates (see p. 1949).

Table 237-1 Major Causes of Valvular Heart Diseases

Valve Lesion	Etiologies
Mitral stenosis	Rheumatic fever Congenital Severe mitral annular calcification SLE, RA
Mitral regurgitation	Acute Endocarditis Papillary muscle rupture (post-MI) Trauma Chordal rupture/leaflet flail (MVP, IE) Chronic Myxomatous (MVP) Rheumatic fever Endocarditis (healed) Mitral annular calcification Congenital (cleft, AV canal) HOCM with SAM Ischemic (LV remodeling) Dilated cardiomyopathy Radiation
Aortic stenosis	Congenital (bicuspid, unicuspid) Degenerative calcific Rheumatic fever Radiation
Aortic regurgitation	Valvular Congenital (bicuspid) Endocarditis Rheumatic fever Myxomatous (prolapse) Traumatic Syphilis Ankylosing spondylitis Root disease Aortic dissection Cystic medial degeneration Marfan's syndrome Bicuspid aortic valve Nonsyndromic familial aneurysm Aortitis

	Hypertension
Tricuspid stenosis	Rheumatic Congenital
Tricuspid regurgitation	Primary Rheumatic Endocarditis Myxomatous (TVP) Carcinoid Radiation Congenital (Ebstein's) Trauma Papillary muscle injury (post-MI) Secondary RV and tricuspid annular dilatation Multiple causes of RV enlargement (e.g., long-standing pulmonary HTN) Chronic RV apical pacing
Pulmonic stenosis	Congenital Carcinoid
Pulmonic regurgitation	Valve disease Congenital Postvalvotomy Endocarditis Annular enlargement Pulmonary hypertension Idiopathic dilation Marfan's syndrome

Abbreviations: AV, atrioventricular; HOCM, hypertrophic obstructive cardiomyopathy; HTN, hypertension; IE, infective endocarditis; LV, left ventricular; MI, myocardial infarction; MVP, mitral valve prolapse; RA, rheumatoid arthritis; RV, right ventricular; SAM, systolic anterior motion of the anterior mitral valve leaflet; SLE, systemic lupus erythematosus; TVP, tricuspid valve prolapse.

In rheumatic MS, the valve leaflets are diffusely thickened by fibrous tissue and/or calcific deposits. The mitral commissures fuse, the chordae tendineae fuse and shorten, the valvular cusps become rigid, and these changes, in turn, lead to narrowing at the apex of the funnel-shaped ("fish-mouth") valve. Although the initial insult to the mitral valve is rheumatic, the later changes may be a nonspecific process resulting from trauma to the valve caused by altered flow patterns due to the initial deformity. Calcification of the stenotic mitral valve immobilizes the leaflets and narrows the orifice further. Thrombus formation and arterial embolization may arise from the calcific valve itself, but in patients with atrial fibrillation (AF), thrombi arise more frequently from the dilated left atrium (LA), particularly from within the left atrial appendage.

Pathophysiology

In normal adults, the area of the mitral valve orifice is 4–6 cm^2. In the presence of significant obstruction, i.e., when the orifice area is reduced to < ~2 cm^2, blood can flow from the LA to the left ventricle (LV) only if propelled by an abnormally elevated left atrioventricular pressure gradient, the hemodynamic hallmark of MS. When the mitral valve opening is reduced to <1 cm^2, often referred to as "severe" MS, a LA pressure of ~25 mmHg is required to maintain a normal cardiac output (CO). The elevated pulmonary venous and pulmonary arterial (PA) wedge pressures reduce pulmonary compliance, contributing to exertional dyspnea. The first bouts of dyspnea are usually precipitated by clinical events that increase the rate of blood flow across the mitral orifice, resulting in further elevation of the LA pressure (see below).

To assess the severity of obstruction hemodynamically, both the transvalvular pressure gradient and the flow rate must be measured (Chap. 230). The latter depends not only on the CO but on the heart rate, as well. An increase in heart rate shortens diastole proportionately more than systole and diminishes the time available for flow across the mitral valve. Therefore, at any given level of CO, tachycardia, including that associated with rapid AF, augments the transvalvular pressure gradient and elevates further the LA pressure. Similar considerations apply to the pathophysiology of tricuspid stenosis.

The LV diastolic pressure and ejection fraction (EF) are normal in isolated MS. In MS and sinus rhythm, the elevated LA and PA wedge pressures exhibit a prominent atrial contraction pattern (*a* wave) and a gradual pressure decline after the *v* wave and mitral valve opening (*y* descent). In severe MS and whenever pulmonary vascular resistance is significantly increased, the pulmonary arterial pressure (PAP) is elevated at rest and rises further during exercise, often causing secondary elevations of right ventricular (RV) end-diastolic pressure and volume.

CARDIAC OUTPUT

In patients with moderate MS (mitral valve orifice 1 cm^2–1.5 cm^2), the CO is normal or almost so at rest, but rises subnormally during exertion. In patients with severe MS (valve area <1 cm^2), particularly those in whom pulmonary vascular resistance is markedly elevated, the CO is subnormal at rest and may fail to rise or may even decline during activity.

PULMONARY HYPERTENSION

The clinical and hemodynamic features of MS are influenced importantly by the level of the PAP. Pulmonary hypertension results from: (1) passive backward transmission of the elevated LA pressure; (2) pulmonary arteriolar constriction (the so-called "second stenosis"), which presumably is triggered by LA and pulmonary venous hypertension (reactive pulmonary hypertension); (3) interstitial edema in the walls of the small pulmonary vessels; and (4) at end stage, organic obliterative changes in the pulmonary vascular bed. Severe pulmonary hypertension results in RV enlargement, secondary tricuspid regurgitation (TR) and

pulmonic regurgitation (PR), as well as right-sided heart failure.

Symptoms

In temperate climates, the latent period between the initial attack of rheumatic carditis (in the increasingly rare circumstances in which a history of one can be elicited) and the development of symptoms due to MS is generally about two decades; most patients begin to experience disability in the fourth decade of life. Studies carried out before the development of mitral valvotomy revealed that once a patient with MS became seriously symptomatic, the disease progressed continuously to death within 2–5 years.

In patients whose mitral orifices are large enough to accommodate a normal blood flow with only mild elevations of LA pressure, marked elevations of this pressure leading to dyspnea and cough may be precipitated by sudden changes in the heart rate, volume status, or CO, as, for example, with severe exertion, excitement, fever, severe anemia, paroxysmal AF and other tachycardias, sexual intercourse, pregnancy, and thyrotoxicosis. As MS progresses, lesser degrees of stress precipitate dyspnea, the patient becomes limited in daily activities, and orthopnea and paroxysmal nocturnal dyspnea develop. The development of permanent AF often marks a turning point in the patient's course and is generally associated with acceleration of the rate at which symptoms progress.

Hemoptysis (Chap. 34) results from rupture of pulmonary-bronchial venous connections secondary to pulmonary venous hypertension. It occurs most frequently in patients who have elevated LA pressures without markedly elevated pulmonary vascular resistances and is rarely fatal. *Recurrent pulmonary emboli* (Chap. 262), sometimes with infarction, are an important cause of morbidity and mortality rates late in the course of MS. *Pulmonary infections,* i.e., bronchitis, bronchopneumonia, and lobar pneumonia, commonly complicate untreated MS, especially during the winter months.

PULMONARY CHANGES

In addition to the aforementioned changes in the pulmonary vascular bed, fibrous thickening of the walls of the alveoli and pulmonary capillaries occurs commonly in MS. The vital capacity, total lung capacity, maximal breathing capacity, and oxygen uptake per unit of ventilation are reduced (Chap. 252). Pulmonary compliance falls further as pulmonary capillary pressure rises during exercise.

THROMBI AND EMBOLI

Thrombi may form in the left atria, particularly within the enlarged atrial appendages of patients with MS. Systemic embolization, the incidence of which is 10–20%, occurs more frequently in patients with AF, in patients >65 years of age, and in those with a reduced CO. However, systemic embolization may be the presenting feature in otherwise asymptomatic patients with only mild MS.

Physical Findings

(See also Chaps. e13 and 227)

INSPECTION AND PALPATION

In patients with severe MS, there may be a malar flush with pinched and blue facies. In patients with sinus rhythm and severe pulmonary hypertension or associated tricuspid stenosis (TS), the jugular venous pulse reveals prominent *a* waves due to vigorous right atrial systole. The systemic arterial pressure is usually normal or slightly low. An RV tap along the left sternal border signifies an enlarged RV. A diastolic thrill may rarely be present at the cardiac apex, with the patient in the left lateral recumbent position.

AUSCULTATION

The first heart sound (S_1) is usually accentuated and slightly delayed. The pulmonic component of the second heart sound (P_2) also is often accentuated, and the two components of the second heart sound (S_2) are closely split. The opening snap (OS) of the mitral valve is most readily audible in expiration at, or just medial to, the cardiac apex. This sound generally follows the sound of aortic valve closure (A_2) by 0.05–0.12 s. The time interval between A_2 and OS varies inversely with the severity of the MS. The OS is followed by a low-pitched, rumbling, diastolic murmur, heard best at the apex with the patient in the left lateral recumbent position (see Fig. 227-5); it is accentuated by mild exercise (e.g., a few rapid sit-ups) carried out just before auscultation. In general, the duration of this murmur correlates with the severity of the stenosis in patients with preserved CO. In patients with sinus rhythm, the murmur often reappears or becomes louder during atrial systole (presystolic accentuation). Soft, grade I or II/VI systolic murmurs are commonly heard at the apex or along the left sternal border in patients with pure MS and do not necessarily signify the presence of MR. Hepatomegaly, ankle edema, ascites, and pleural effusion, particularly in the right pleural cavity, may occur in patients with MS and RV failure.

ASSOCIATED LESIONS

With severe pulmonary hypertension, a pansystolic murmur produced by functional TR may be audible along the left sternal border. This murmur is usually louder during inspiration and diminishes during forced expiration (Carvallo's sign). When the CO is markedly reduced in MS, the typical auscultatory findings, including the diastolic rumbling murmur, may not be detectable (silent MS), but they may reappear as compensation is restored. The *Graham Steell murmur* of PR, a high-pitched, diastolic, decrescendo blowing murmur along the left sternal border, results from dilation of the pulmonary valve ring and occurs in patients with mitral valve disease and severe pulmonary hypertension. This murmur may be indistinguishable from the more common murmur produced by aortic regurgitation (AR), although it may increase in intensity with inspiration and is accompanied by a loud and often palpable P2.

Laboratory Examination

ECG

In MS and sinus rhythm, the P wave usually suggests LA enlargement (see Fig. 228-8). It may become tall and peaked in lead II and upright in lead V_1 when severe pulmonary hypertension or TS complicates MS and right atrial (RA) enlargement occurs. The QRS complex is usually normal. However, with severe pulmonary hypertension, right-axis deviation and RV hypertrophy are often present.

ECHOCARDIOGRAM

(See also Chap. 229) Transthoracic echocardiography (TTE) with color flow and spectral Doppler imaging provides critical information, including measurements of mitral inflow velocity during early (E wave) and late (A wave in patients in sinus rhythm) diastolic filling, estimates of the transvalvular peak and mean gradients and of the mitral orifice area, the presence and severity of any associated MR, the extent of leaflet calcification and restriction, the degree of distortion of the subvalvular apparatus, and the anatomic suitability for percutaneous mitral balloon valvotomy [percutaneous mitral balloon valvuloplasty (PMBV); see below]. In addition, TTE provides an assessment of LV and RV function, chamber sizes, an estimation of the pulmonary artery pressure (PAP) based on the tricuspid regurgitant jet velocity, and an indication of the presence and severity of any associated valvular lesions. Transesophageal echocardiography (TEE) provides superior images and should be employed when TTE is

inadequate for guiding management decisions. TEE is especially indicated to exclude the presence of left atrial thrombus prior to PMBV.

CHEST X-RAY

The earliest changes are straightening of the upper left border of the cardiac silhouette, prominence of the main pulmonary arteries, dilation of the upper lobe pulmonary veins, and posterior displacement of the esophagus by an enlarged LA. Kerley B lines are fine, dense, opaque, horizontal lines that are most prominent in the lower and mid-lung fields and that result from distention of interlobular septae and lymphatics with edema when the resting mean LA pressure exceeds approximately 20 mmHg.

Differential Diagnosis

Like MS, significant MR may also be associated with a prominent diastolic murmur at the apex due to increased antegrade transmitral flow, but in patients with isolated MR, this diastolic murmur commences slightly later than in patients with MS, and there is often clear-cut evidence of LV enlargement. An opening snap and increased $P_2$are absent, and S_1 is soft or absent. An apical pansystolic murmur of at least grade III/VI intensity as well as an S_3 suggest significant MR. Similarly, the apical mid-diastolic murmur associated with severe AR (*Austin Flint murmur*) may be mistaken for MS but can be differentiated from it because it is not intensified in presystole and becomes softer with administration of amyl nitrite. TS, which occurs rarely in the absence of MS, may mask many of the clinical features of MS or be clinically silent; when present, the diastolic murmur of TS increases with inspiration.

Atrial septal defect (Chap. 236) may be mistaken for MS; in both conditions there is often clinical, ECG, and chest x-ray evidence of RV enlargement and accentuation of pulmonary vascularity. However, the absence of LA enlargement and of Kerley B lines and the demonstration of fixed splitting of S_2 with a grade 2 or 3 mid-systolic murmur at the mid to upper left sternal border all favor atrial septal defect over MS. Atrial septal defects with large left-to-right shunts may result in functional TS because of the enhanced diastolic flow.

Left atrial myxoma (Chap. 240) may obstruct LA emptying, causing dyspnea, a diastolic murmur, and hemodynamic changes resembling those of MS. However, patients with an LA myxoma often have features suggestive of a systemic disease, such as weight loss, fever, anemia, systemic emboli, and elevated serum IgG and interleukin 6 (IL-6) concentrations. The auscultatory findings may change markedly with body position. The diagnosis can be established by the demonstration of a characteristic echo-producing mass in the LA with TTE.

Cardiac Catheterization

Left and right heart catheterization is useful when there is a discrepancy between the clinical and TTE findings that cannot be resolved with either TEE or cardiac magnetic resonance (CMR) imaging. The growing experience with CMR for the assessment of patients with valvular heart disease may decrease the need for invasive catheterization. Catheterization is helpful in assessing associated lesions, such as aortic stenosis (AS) and AR. Catheterization and coronary angiography are not usually necessary to aid in decision-making about surgery in patients younger than 65 years of age, with typical findings of severe mitral obstruction on physical examination and TTE. In men older than 40 years of age, women older than 45 years of age, and younger patients with coronary risk factors, especially those with positive noninvasive stress tests for myocardial ischemia, coronary angiography is advisable preoperatively to identify patients with critical coronary obstructions that should be bypassed at the time of operation. Computed tomographic coronary angiography (CTCA) (Chap. 229) is now often used to screen preoperatively for the presence of coronary artery disease (CAD) in patients with valvular heart disease and low pretest likelihood of CAD. Catheterization and left ventriculography are indicated in most patients who have undergone PMBV or previous mitral valve surgery, and who have redeveloped limiting symptoms, especially if questions regarding the severity of the valve lesion(s) remain after echocardiography.

Treatment: Mitral Stenosis

(See Fig. 237-1) Penicillin prophylaxis of group A β-hemolytic streptococcal infections (Chap. 322) for secondary prevention of rheumatic fever is important for at-risk patients with rheumatic MS **(Table 237-2)**. Recommendations for infective endocarditis prophylaxis have recently changed. In symptomatic patients, some improvement usually occurs with restriction of sodium intake and small doses of oral diuretics. Beta blockers, nondihydropyridine calcium channel blockers (e.g., verapamil or diltiazem), and digitalis glycosides are useful in slowing the ventricular rate of patients with AF. Warfarin to an international normalized ratio (INR) of 2–3 should be administered indefinitely to patients with MS, who have AF or a history of thromboembolism. The routine use of warfarin in patients in sinus rhythm with LA enlargement (maximal dimension >5.5 cm) with or without spontaneous echo contrast is more controversial.

Figure 237-1

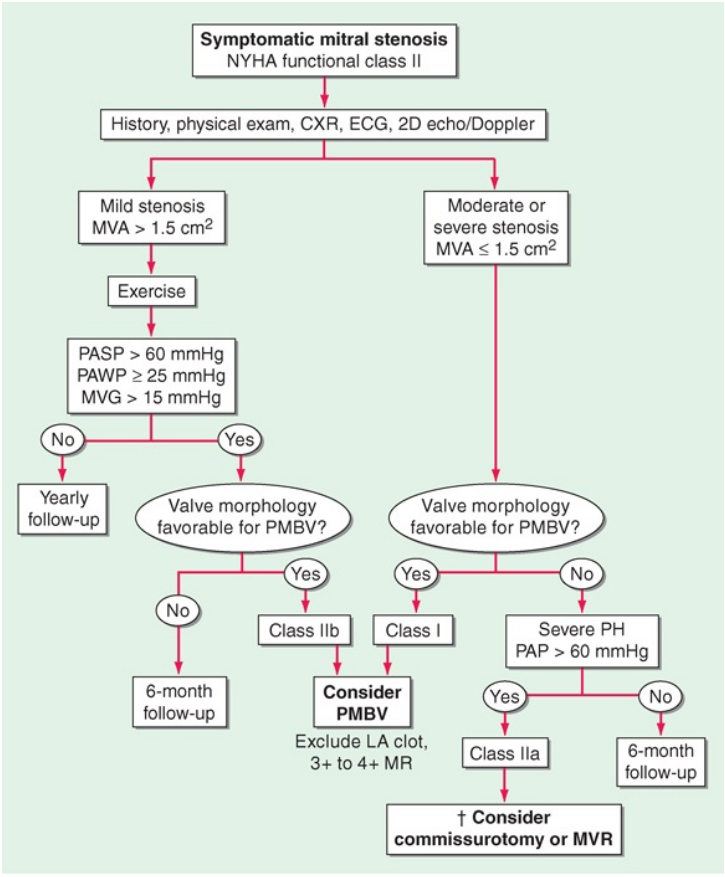

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management Strategy for Patients with Mitral Stenosis (MS) and Mild Symptoms. †There is controversy as to whether patients with severe MS (MVA <1 cm^2) and severe pulmonary hypertension (PH) (PASP >60 mmHg) should undergo percutaneous mitral balloon valvotomy (PMBV) or mitral valve replacement (MVR) to prevent right ventricular failure. CXR, chest x-ray; ECG, electrocardiogram; echo, echocardiography; LA, left atrial; MR, mitral regurgitation; MVA, mitral valve area; MVG, mean mitral valve pressure gradient; NYHA, New York Heart Association; PASP, pulmonary artery systolic pressure; PAWP, pulmonary artery wedge pressure; 2D, 2-dimensional. *(From RO Bonow et al: J Am Coll Cardiol 48:e1, 2006; with permission.)*

Table 237-2 Medical Therapy of Valvular Heart Disease

Lesion	Symptom Control	Natural History
Mitral stenosis	Beta blockers, nondihydropyridine calcium channel blockers, or digoxin for rate control of AF; cardioversion for new-onset AF and HF; diuretics for HF	Warfarin for AF or thromboembolism; PCN for RF prophylaxis
Mitral regurgitation	Diuretics for HF	Warfarin for AF or thromboembolism
	Vasodilators for acute MR	Vasodilators for HTN
Aortic stenosis	Diuretics for HF	No proven therapy
Aortic regurgitation	Diuretics and vasodilators for HF	Vasodilators for HTN

Note: Antibiotic prophylaxis is recommended according to current American Heart Association guidelines. For patients with these forms of valvular heart disease, prophylaxis is indicated for a prior history of endocarditis. HF is an indication for surgical or percutaneous treatment, and the recommendations here pertain to short-term therapy prior to definitive correction of the valve lesion. For patients whose comorbidities prohibit surgery, the medical therapies listed can be continued according to available guidelines for the management of HF. See text.

Abbreviations: AF, atrial fibrillation; HF, heart failure; HTN, systemic hypertension; PCN, penicillin; RF, rheumatic fever.

Source: Adapted from NA Boon, P Bloomfield: Heart 87:395, 2002, with permission.

If AF is of relatively recent onset in a patient whose MS is not severe enough to warrant PMBV or surgical commissurotomy, reversion to sinus rhythm pharmacologically or by means of electrical countershock is indicated. Usually, cardioversion should be undertaken after the patient has had at least 3 consecutive weeks of anticoagulant treatment to a therapeutic INR. If cardioversion is indicated more urgently, then intravenous heparin should be provided and TEE performed to exclude the presence of left atrial thrombus before the procedure. Conversion to sinus rhythm is rarely successful or sustained in patients with severe MS, particularly those in whom the LA is especially enlarged or in whom AF has been present for more than 1 year.

MITRAL VALVOTOMY

Unless there is a contraindication, mitral valvotomy is indicated in symptomatic [New York Heart Association (NYHA) Functional Class II–IV] patients with isolated MS, whose effective orifice (valve area) is < ~1 cm^2/m^2 body surface area, or <1.5 cm^2 in normal-sized adults. Mitral valvotomy can be carried out by two techniques: PMBV and surgical valvotomy. In PMBV (Figs. 237-2 and 237-3), a catheter is directed into the LA after transseptal puncture, and a single balloon is directed across the valve and

inflated in the valvular orifice. Ideal patients have relatively pliable leaflets with little or no commissural calcium. In addition, the subvalvular structures should not be significantly scarred or thickened, and there should be no left atrial thrombus. The short- and long-term results of this procedure in appropriate patients are similar to those of surgical valvotomy, but with less morbidity and a lower periprocedural mortality rate. Event-free survival in younger (<45 years) patients with pliable valves is excellent, with rates as high as 80–90% over 3–7 years. Therefore, PMBV has become the procedure of choice for such patients when it can be performed by a skilled operator in a high-volume center.

Figure 237-2

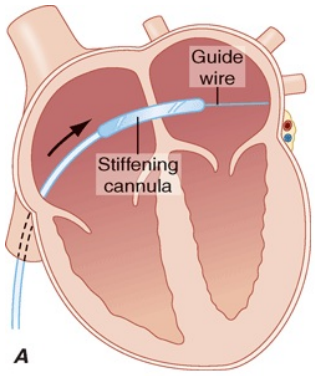

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

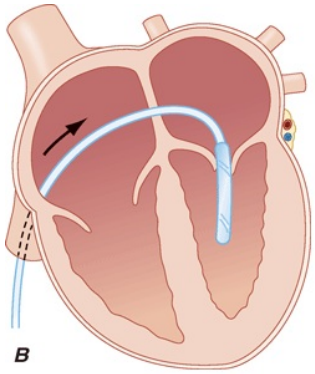

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

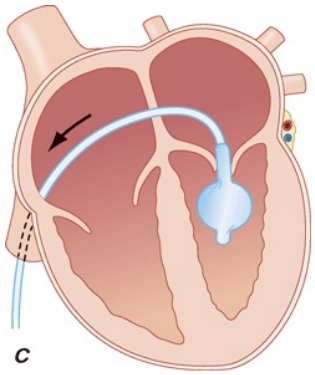

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

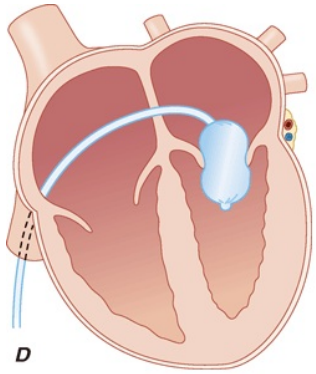

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Inoue Balloon Technique for Percutaneous Mitral Balloon Valvotomy.*A.* After transseptal puncture, the deflated balloon catheter is advanced across the interatrial septum, then across the mitral valve and into the left ventricle. ***B.-D.*** The balloon is inflated stepwise within the mitral orifice.

Figure 237-3

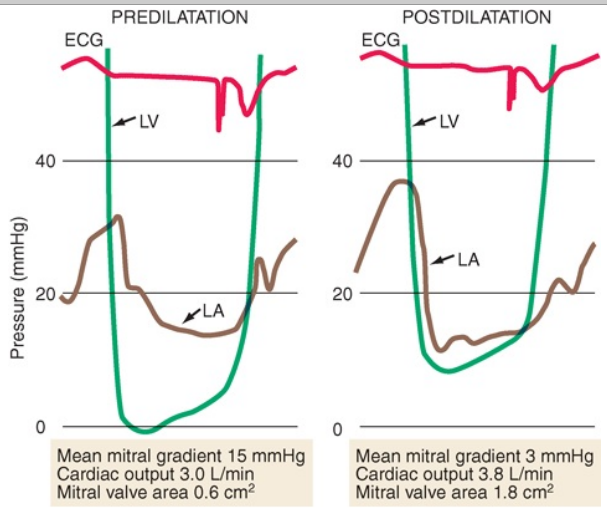

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Simultaneous Left Atrial (LA) and Left Ventricular (LV) Pressure before and after percutaneous mitral balloon valvuloplasty (PMBV) in a patient with severe mitral stenosis. *(Courtesy of Raymond G. McKay, MD; with permission.)*

Transthoracic echocardiography is helpful in identifying patients for the percutaneous procedure, and TEE is performed routinely to exclude left atrial thrombus at the time of the scheduled procedure. An "echo score" has been developed to help guide decision-making. The score accounts for the degree of leaflet thickening, calcification, and mobility, and for the extent of subvalvular thickening. A lower score predicts a higher likelihood of successful PMBV.

In patients in whom PMBV is not possible or unsuccessful, or in many patients with restenosis, an "open" valvotomy using cardiopulmonary bypass is necessary. In addition to opening the valve commissures, it is important to loosen any subvalvular fusion of papillary muscles and chordae tendineae and to remove large deposits of calcium, thereby improving valvular function, as well as to remove atrial thrombi. The perioperative mortality rate is ~2%.

Successful valvotomy is defined by a 50% reduction in the mean mitral valve gradient and a doubling of the mitral valve area. Successful valvotomy, whether balloon or surgical, usually results in striking symptomatic and hemodynamic improvement and prolongs survival. However, there is no evidence that the procedure improves the prognosis of patients with slight or no functional impairment. Therefore, unless recurrent systemic embolization or severe pulmonary hypertension has occurred (PA systolic pressures >50 mmHg at rest or >60 mmHg with exercise), valvotomy is *not* recommended for patients who are entirely asymptomatic and/or who have mild stenosis (mitral valve area >1.5 cm^2). When there is little symptomatic improvement after valvotomy, it is likely that the procedure was ineffective, that it induced MR, or that associated valvular or myocardial disease was present. About half of all patients undergoing surgical mitral valvotomy require reoperation by 10 years. In the pregnant patient with MS, valvotomy should be carried out if pulmonary congestion occurs despite intensive medical treatment. PMBV is the preferred strategy in this setting and is performed with TEE and no or minimal x-ray exposure.

Mitral valve replacement (MVR) is necessary in patients with MS and significant associated MR, those in whom the valve has been severely distorted by previous transcatheter or operative manipulation, or those in whom the surgeon does not find it possible to improve valve function significantly with valvotomy. MVR is now routinely performed with preservation of the chordal attachments to optimize LV functional recovery. Perioperative mortality rates with MVR vary with age, LV function, the presence of CAD, and associated comorbidities. They average 5% overall but are lower in young patients and may be twice as high in patients >65 years of age with comorbidity rates (Table 237-3). Since there are also long-term complications of valve replacement (p. 1949), patients in whom preoperative evaluation suggests the possibility that MVR may be required should be operated on only if they have severe MS—i.e., an orifice area ≤1 cm^2—and are in NYHA Class III, i.e., symptomatic with ordinary activity despite optimal medical therapy. The overall 10-year survival of surgical survivors is ~70%. Long-term prognosis is worse in patients >65 years of age and those with marked disability and marked depression of the CO preoperatively.

Pulmonary hypertension and RV dysfunction are additional risk factors for poor outcome.

Table 237-3 Mortality Rates After Valve Surgery*

Operation	Number	Unadjusted Operative Mortality (%)
AVR (isolated)	20,168	3.2
MVR (isolated)	4,616	5.0
AVR + CAB	16,678	5.0
MVR + CAB	2,479	8.8
AVR + MVR	1,239	9.0
MVP	5,617	1.8
MVP + CAB	4,932	4.8
TV surgery	6,235	9.2
PV surgery	480	6.0

*Data are for calendar year 2008, in which 912 sites reported a total of 276,308 procedures. Data are available from the Society of Thoracic Surgeons at http://www.sts.org/documents/pdf/ndb/2ndHarvestExecutiveSummary_2009.pdf.

Abbreviations: AVR, aortic valve replacement; CAB, coronary artery bypass; MVR, mitral valve replacement; MVP, mitral valve repair; TV surgery, tricuspid valve repair and replacement; PV surgery, pulmonic valve repair and replacement.

MITRAL REGURGITATION

Etiology

MR may result from an abnormality or disease process that affects any one or more of the five functional components of the mitral valve apparatus (leaflets, annulus, chordae tendineae, papillary muscles, and subjacent myocardium) (Table 237-1). Acute MR can occur in the setting of acute myocardial infarction (MI) with papillary muscle rupture (Chap. 245), following blunt chest wall trauma, or during the course of infective endocarditis. With acute MI, the posteromedial papillary muscle is involved much more frequently than the anterolateral papillary muscle because of its singular blood supply. Transient, acute MR can occur during periods of active ischemia and bouts of angina pectoris. Rupture of chordae tendineae can result in "acute-on-chronic MR" in patients with myxomatous degeneration of the valve apparatus.

Chronic MR can result from rheumatic disease, mitral valve prolapse (MVP), extensive mitral annular calcification, congenital valve defects, hypertrophic obstructive cardiomyopathy (HOCM), and dilated cardiomyopathy (Chap. 238). The rheumatic process produces rigidity, deformity, and retraction of the valve cusps and commissural fusion, as well as shortening, contraction, and fusion of the chordae tendineae. The MR associated with both MVP and HOCM is usually dynamic in nature. MR in HOCM occurs as a consequence of anterior papillary muscle displacement and systolic anterior motion of the anterior mitral valve leaflet into the narrowed LV outflow tract. Annular calcification is especially prevalent among patients with advanced renal disease and is commonly observed in women >65 years of age with hypertension and diabetes. MR may occur as a congenital anomaly (Chap. 236), most commonly as a defect of the endocardial cushions (atrioventricular cushion defects). A cleft anterior mitral valve leaflet accompanies primum atrial septal defect. Chronic MR is frequently secondary to ischemia and may occur as a consequence of ventricular remodeling, papillary muscle displacement, and leaflet tethering, or with fibrosis of a papillary muscle, in patients with healed myocardial infarction(s) and ischemic cardiomyopathy. Similar mechanisms of annular dilation and ventricular remodeling contribute to the MR that occurs among patients with nonischemic forms of dilated cardiomyopathy once the left ventricular end-diastolic dimension reaches 6 cm.

Irrespective of cause, chronic severe MR is often progressive, since enlargement of the LA places tension on the posterior mitral leaflet, pulling it away from the mitral orifice and thereby aggravating the valvular dysfunction. Similarly, LV dilation increases the regurgitation, which, in turn, enlarges the LA and LV further, causing chordal rupture and resulting in a vicious circle; hence the aphorism, "mitral regurgitation begets mitral regurgitation."

Pathophysiology

The resistance to LV emptying (LV afterload) is reduced in patients with MR. As a consequence, the LV is decompressed into the LA during ejection, and with the reduction in LV size during systole, there is a rapid decline in LV tension. The initial compensation to MR is more complete LV emptying. However, LV volume increases progressively with time as the severity of the regurgitation increases and as LV contractile function deteriorates. This increase in LV volume is often accompanied by a reduced forward CO, although LV compliance is often increased and, thus, LV diastolic pressure does not increase until late in the course. The regurgitant volume varies directly with the LV systolic pressure and the size of the regurgitant orifice; as mentioned above, the latter, in turn, is influenced by the extent of LV and mitral annular dilation. Since ejection fraction (EF) rises in severe MR in the presence of normal LV function, even a modest reduction in this parameter (<60%) reflects significant dysfunction.

During early diastole, as the distended LA empties, there is a particularly rapid *y* descent in the absence of accompanying MS. A brief, early diastolic LA-LV pressure gradient [often generating a rapid filling sound (S_3) and mid-diastolic murmur masquerading as MS] may occur in patients with pure MR as a result of the very rapid flow of blood across a normal-sized mitral orifice.

Semi-quantitative estimates of left ventricular ejection fraction (LVEF), CO, PA systolic pressure, regurgitant volume, regurgitant fraction (RF), and the effective regurgitant orifice area can be obtained during a careful Doppler echocardiographic examination. These measurements can also be obtained with CMR. Left and right heart catheterization with contrast ventriculography is used less frequently. Severe, nonischemic MR is defined by a regurgitant volume ≥60 mL/beat, regurgitant fraction (RF) ≥50%, and effective regurgitant orifice area ≥0.40 cm^2. Severe ischemic MR is usually associated with an effective regurgitant orifice area of >0.3 cm^2.

LA COMPLIANCE

In acute severe MR, the regurgitant volume is delivered into a normal-sized LA having normal or reduced compliance. As a result, LA pressures rise markedly for any increase in LA volume. The *v* wave in the LA pressure pulse is usually prominent, LA and pulmonary venous pressures are markedly elevated, and pulmonary edema is common. Because of the rapid rise in LA pressures during ventricular systole, the murmur of acute MR is early in timing and decrescendo in configuration ending well before S_2, as a reflection of the progressive diminution in the LV-LA pressure gradient. LV systolic function in acute MR may be normal, hyperdynamic, or reduced, depending on the clinical context.

Patients with chronic severe MR, on the other hand, develop marked LA enlargement and *increased* LA compliance with little if any increase in LA and pulmonary venous pressures for any increase in LA volume. The LA *v* wave is relatively less prominent.

The murmur of chronic MR is classically holosystolic in timing and plateau in configuration, as a reflection of the near-constant LV-LA pressure gradient. These patients usually complain of severe fatigue and exhaustion secondary to a low forward CO, while symptoms resulting from pulmonary congestion are less prominent initially; AF is almost invariably present once the LA dilates significantly.

Symptoms

Patients with chronic mild-to-moderate isolated MR are usually asymptomatic. This form of LV volume overload is well tolerated. Fatigue, exertional dyspnea, and orthopnea are the most prominent complaints in patients with chronic severe MR. Palpitations are common and may signify the onset of AF. Right-sided heart failure, with painful hepatic congestion, ankle edema, distended neck veins, ascites, and secondary TR, occurs in patients with MR who have associated pulmonary vascular disease and marked pulmonary hypertension. Conversely, acute pulmonary edema is common in patients with acute severe MR.

Physical Findings

In patients with chronic severe MR, the arterial pressure is usually normal, although the carotid arterial pulse may show a sharp upstroke owing to the reduced forward cardiac output. A systolic thrill is often palpable at the cardiac apex, the LV is hyperdynamic with a brisk systolic impulse and a palpable rapid-filling wave (S_3), and the apex beat is often displaced laterally.

In patients with acute severe MR, the arterial pressure may be reduced with a narrow pulse pressure, the jugular venous pressure and wave forms may be normal or increased and exaggerated, the apical impulse is not displaced, and signs of pulmonary congestion are prominent.

AUSCULTATION

S_1 is generally absent, soft, or buried in the holosystolic murmur of chronic MR. In patients with severe MR, the aortic valve may close prematurely, resulting in wide but physiologic splitting of S_2. A low-pitched S_3 occurring 0.12–0.17 s after the aortic valve closure sound, i.e., at the completion of the rapid-filling phase of the LV, is believed to be caused by the sudden tensing of the papillary muscles, chordae tendineae, and valve leaflets. It may be followed by a short, rumbling, mid-diastolic murmur, even in the absence of structural MS. A fourth heart sound is often audible in patients with *acute* severe MR who are in sinus rhythm. A presystolic murmur is not ordinarily heard with isolated MR.

A systolic murmur of at least grade III/VI intensity is the most characteristic auscultatory finding in chronic severe MR. It is usually holosystolic (see Fig. 227-5A), but as previously noted it is decrescendo and ceases in mid- to late systole in patients with acute severe MR. The systolic murmur of chronic MR is usually most prominent at the apex and radiates to the axilla. However, in patients with ruptured chordae tendineae or primary involvement of the posterior mitral leaflet with prolapse or flail, the regurgitant jet is eccentric, directed anteriorly, and strikes the LA wall adjacent to the aortic root. In this situation, the systolic murmur is transmitted to the base of the heart and, therefore, may be confused with the murmur of AS. In patients with ruptured chordae tendineae, the systolic murmur may have a cooing or "sea gull" quality, while a flail leaflet may produce a murmur with a musical quality. The systolic murmur of chronic MR not due to MVP is intensified by isometric exercise (handgrip) but is reduced during the strain phase of the Valsalva maneuver because of the associated decrease in LV preload.

Laboratory Examination

ECG

In patients with sinus rhythm, there is evidence of LA enlargement, but RA enlargement also may be present when pulmonary hypertension is severe. Chronic severe MR is generally associated with AF. In many patients, there is no clear-cut ECG evidence of enlargement of either ventricle. In others, the signs of eccentric LV hypertrophy are present.

ECHOCARDIOGRAM

TTE is indicated to assess the mechanism of the MR and its hemodynamic severity. LV function can be assessed from LV end-diastolic and end-systolic volumes and EF. Observations can be made regarding leaflet structure and function, chordal integrity, LA and LV size, annular calcification, and regional and global LV systolic function. Doppler imaging should demonstrate the width or area of the color flow MR jet within the LA, the intensity of the continuous wave Doppler signal, the pulmonary venous flow contour, the early peak mitral inflow velocity, and the quantitative measures of regurgitant volume, RF, and effective regurgitant orifice area. In addition, the PA pressures can be estimated from the TR jet velocity. TTE is also indicated to follow the course of patients with chronic MR and to provide rapid assessment for any clinical change. The echocardiogram in patients with MVP is described in the next section. TEE provides greater detail than TTE (see Fig. 229-5).

CHEST X-RAY

The LA and LV are the dominant chambers in chronic MR. Late in the course of the disease, the LA may be massively enlarged and forms the right border of the cardiac silhouette. Pulmonary venous congestion, interstitial edema, and Kerley B lines are sometimes noted. Marked calcification of the mitral leaflets occurs commonly in patients with longstanding, combined rheumatic MR and MS. Calcification of the mitral annulus may be visualized, particularly on the lateral view of the chest. Patients with acute severe MR may have asymmetric pulmonary edema if the regurgitant jet is directed predominantly to the orifice of an upper lobe pulmonary vein.

Treatment: Mitral Regurgitation

MEDICAL TREATMENT

(See Fig. 237-4 and Table 237-2) The management of chronic severe MR depends to some degree on its cause. Warfarin should be provided once AF intervenes with a target INR of 2–3. Cardioversion should be considered depending on the clinical context and left atrial size. In contrast to the acute setting, there are no large, long-term prospective studies to substantiate the use of vasodilators for the treatment of chronic, isolated severe MR with preserved LV systolic function *in the absence of systemic hypertension.* The severity of MR in the setting of an ischemic or nonischemic dilated cardiomyopathy may diminish with aggressive, evidence-based treatment of heart failure, including the use of diuretics, beta blockers, angiotensin-converting enzyme (ACE) inhibitors, digitalis, and biventricular pacing (cardiac resynchronization therapy [CRT]). Asymptomatic patients with severe MR in sinus rhythm with normal LV size and systolic function should avoid isometric forms of exercise.

Figure 237-4

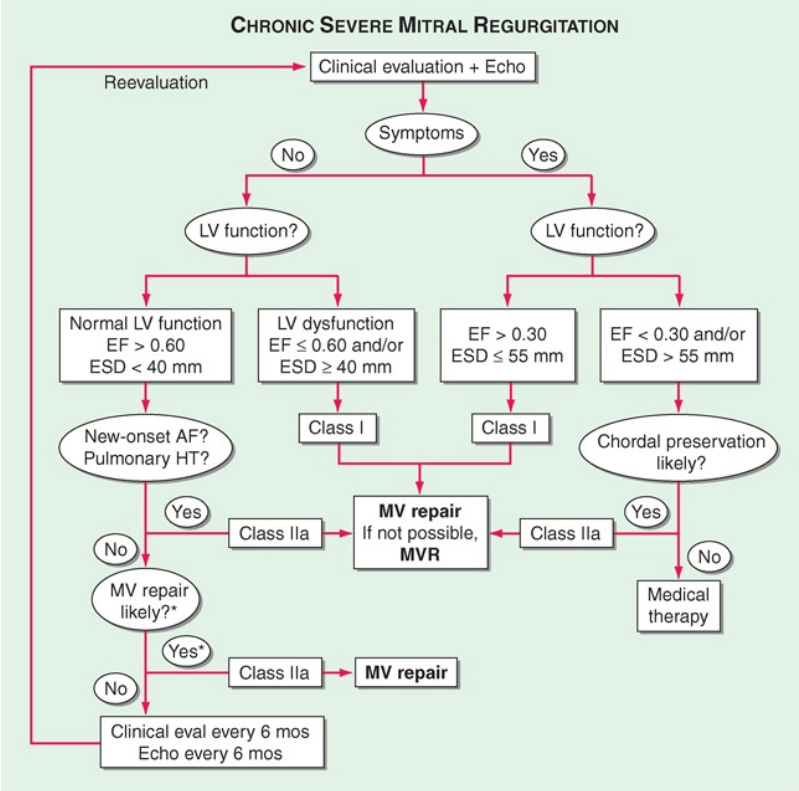

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management Strategy for Patients with Chronic Severe Nonischemic Mitral Regurgitation. *Mitral valve (MV) repair may be performed in asymptomatic patients with normal left ventricular (LV) function if performed by an experienced surgical team and if the likelihood of successful MV repair is >90%. AF, atrial fibrillation; Echo, echocardiography; EF, ejection fraction; ESD, end-systolic dimension; eval, evaluation; HT, hypertension; MVR, mitral valve replacement. *(From RO Bonow et al: J Am Coll Cardiol 48:e1, 2006; with permission.)*

Patients with acute severe MR require urgent stabilization and preparation for surgery. Diuretics, intravenous vasodilators (particularly sodium nitroprusside), and even intraaortic balloon counterpulsation may be needed for patients with post-MI papillary muscle rupture or other forms of acute severe MR.

SURGICAL TREATMENT

In the selection of patients with chronic, nonischemic, severe MR for surgical treatment, the often slowly progressive nature of the condition must be balanced against the immediate and long-term risks associated with operation. These risks are significantly lower for primary valve repair than for valve replacement (Table 237-3). Repair usually consists of valve reconstruction using a variety of valvuloplasty techniques and insertion of an annuloplasty ring. Repair spares the patient the long-term adverse consequences of valve replacement, i.e., thromboembolic and hemorrhagic complications in the case of mechanical prostheses and late valve failure necessitating repeat valve replacement in the case of bioprostheses (p. 1949). In addition, by preserving the integrity of the papillary muscles, subvalvular apparatus, and chordae tendineae, mitral repair and valvuloplasty maintain LV function to a relatively greater degree.

Surgery for chronic, nonischemic, severe MR is indicated once symptoms occur, especially if valve repair is feasible (Fig. 237-4). Other indications for early consideration of mitral valve repair include recent-onset AF and pulmonary hypertension, defined as a PA pressure ≥50 mmHg at rest or ≥60 mmHg with exercise. Surgical treatment of chronic, nonischemic severe MR is indicated for asymptomatic patients when LV dysfunction is progressive, with LVEF falling below 60% and/or end-systolic dimension increasing beyond 40 mm. These aggressive recommendations for surgery are predicated on the outstanding results achieved with mitral valve repair, particularly when applied to patients with myxomatous disease, such as that associated with prolapse or flail leaflet. Indeed, primary valvuloplasty repair of patients younger than 75 years with normal LV systolic function and no CAD can now be performed by experienced surgeons with <1% perioperative mortality risk. Repair is feasible in up to 95% of patients with myxomatous disease. Long-term durability is excellent; the incidence of reoperative surgery for failed primary repair is ~1% per year for 10 years after surgery. For patients with AF, left or bi-atrial Maze surgery or radiofrequency isolation of the pulmonary veins is often performed to reduce the risk of recurrent, postoperative AF.

The surgical management of patients with ischemic MR is more complicated and almost always involves simultaneous coronary artery revascularization. Although current surgical practice includes annuloplasty repair with an undersized ring for patients with moderate or greater degrees of MR at the time of coronary artery bypass surgery, the efficacy of this approach has not been established in prospective, randomized trials. There is also uncertainty as to whether or not valve repair or replacement is the preferred strategy, given the higher incidence of residual or recurrent MR after repair in this context compared with outcomes in patients with organic (myxomatous) disease. In patients with significantly impaired LV function (EF <30%), the risk of surgery increases, the recovery of LV performance is incomplete, and the long-term survival is reduced. However, conservative management has little to offer these patients, so operative treatment may be indicated, and the clinical and hemodynamic improvement that follows surgical treatment of patients with advanced disease is occasionally dramatic, especially when severe CAD is present and bypass grafting can be performed. The routine performance of valve repair in patients with significant MR in the setting of severe, dilated cardiomyopathy has not been shown to improve long-term survival. Patients with acute severe MR can often be stabilized temporarily with appropriate medical therapy, but surgical correction will be necessary, emergently in the case of papillary muscle rupture and within days to weeks in most other settings.

When surgical treatment is contemplated, left and right heart catheterization and left ventriculography *may* be helpful in confirming the presence of severe MR in patients in whom there is a discrepancy between the clinical and TTE findings that cannot be resolved with TEE or CMR. Coronary angiography identifies patients who require concomitant coronary

revascularization.

PERCUTANEOUS MITRAL VALVE REPAIR

A transcatheter approach to the treatment of either organic or functional MR may be feasible in selected patients with appropriate anatomy, although the proper role of current techniques remains under active investigation. One approach involves the deployment of a clip delivered via transeptal puncture that grasps the leading edges of the mitral leaflets in their mid-portion (anterior scallop 2–posterior scallop 2 or A2-P2, Fig. 237-5). The length and width of the gap between these leading edges have dictated patient eligibility in the trials reported to date. Preliminary results with this technically demanding technique have been favorable. A second approach involves the deployment of a device within the coronary sinus that can be adjusted to reduce its circumference, thus secondarily decreasing the circumference of the mitral annulus and the effective orifice area of the valve, much like a surgically implanted ring. Variations in the anatomic relationship of the coronary sinus to the mitral annulus and circumflex coronary artery have limited the applicability of this technique. Attempts to reduce the septal-lateral dimension of a dilated annulus using adjustable cords placed across the LV in a subvalvular location have also been investigated.

Figure 237-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clip Used to Grasp the Free Edges of the Anterior and Posterior Leaflets in Their Mid-Sections during Percutaneous Repair of Selected Patients with Mitral Regurgitation. *(Courtesy of Abbott Vascular. © 2010 Abbott Laboratories. All rights reserved.)*

MITRAL VALVE PROLAPSE

MVP, also variously termed the *systolic click-murmur syndrome, Barlow's syndrome, floppy-valve syndrome,* and *billowing mitral leaflet syndrome,* is a relatively common but highly variable clinical syndrome resulting from diverse pathologic mechanisms of the mitral valve apparatus. Among these are excessive or redundant mitral leaflet tissue, which is commonly associated with myxomatous degeneration and greatly increased concentrations of certain glycosaminoglycans.

In most patients with MVP, the cause is unknown, but in some it appears to be a genetically determined collagen disorder. A reduction in the production of type III collagen has been incriminated, and electron microscopy has revealed fragmentation of collagen fibrils.

MVP is a frequent finding in patients with heritable disorders of connective tissue, including Marfan's syndrome (Chap. 363), osteogenesis imperfecta, and Ehlers-Danlos syndrome. MVP may be associated with thoracic skeletal deformities similar to but not as severe as those in Marfan's syndrome, such as a high-arched palate and alterations of the chest and thoracic spine, including the so-called straight back syndrome.

In most patients with MVP, myxomatous degeneration is confined to the mitral valve, although the tricuspid and aortic valves may also be affected. The posterior mitral leaflet is usually more affected than the anterior, and the mitral valve annulus is often dilated. In many patients, elongated, redundant, or ruptured chordae tendineae cause or contribute to the regurgitation.

MVP also may occur rarely as a sequel to acute rheumatic fever, in ischemic heart disease, and in various cardiomyopathies, as well as in 20% of patients with ostium secundum atrial septal defect.

MVP may lead to excessive stress on the papillary muscles, which, in turn, leads to dysfunction and ischemia of the papillary muscles and the subjacent ventricular myocardium. Rupture of chordae tendineae and progressive annular dilation and calcification contribute to valvular regurgitation, which then places more stress on the diseased mitral valve apparatus, thereby creating a vicious circle. The ECG changes (see below) and ventricular arrhythmias appear to result from regional ventricular dysfunction related to the increased stress placed on the papillary muscles.

Clinical Features

MVP is more common in women and occurs most frequently between the ages of 15 and 30 years; the clinical course is most often benign. MVP may also be observed in older (>50 years) patients, often men, in whom MR is often more severe and requires surgical treatment. There is an increased familial incidence for some patients, suggesting an autosomal dominant form of inheritance with incomplete penetrance. MVP varies in its clinical expression, ranging from only a systolic click and murmur with mild prolapse of the posterior leaflet to severe MR due to chordal rupture and leaflet flail. The degree of myxomatous change of the leaflets can also vary widely. In many patients, the condition progresses over years or decades; in others it worsens rapidly as a result of chordal rupture or endocarditis.

Most patients are asymptomatic and remain so for their entire lives. However, in North America, MVP is now the most common cause of isolated severe MR requiring surgical treatment. Arrhythmias, most commonly ventricular premature contractions and paroxysmal supraventricular and ventricular tachycardia, as well as AF, have been reported and may cause palpitations, light-headedness, and syncope. Sudden death is a very rare complication and occurs most often in patients with severe MR and depressed LV systolic function. There may be an excess risk of sudden death among patients with a flail leaflet. Many patients have chest pain that is difficult to evaluate; it is often substernal, prolonged, and not related to exertion, but may rarely resemble angina pectoris. Transient cerebral ischemic attacks secondary to emboli from the mitral valve due to endothelial disruption have been reported. Infective endocarditis may occur in patients with MR and/or leaflet thickening.

AUSCULTATION

The most important finding is the mid- or late (nonejection) systolic click, which occurs 0.14 s or more after S_1 and is thought to be generated by the sudden tensing of slack, elongated chordae tendineae or by the prolapsing mitral leaflet when it reaches its maximum excursion. Systolic clicks may be multiple and may be followed by a high-pitched, late systolic crescendo-decrescendo murmur, which occasionally is "whooping" or "honking" and is heard best at the apex. The click and murmur occur earlier with standing, during the strain phase of the Valsalva maneuver, and with any intervention that decreases LV volume, exaggerating the propensity of mitral leaflet prolapse. Conversely, squatting and isometric exercises, which increase LV volume, diminish MVP; the click-murmur complex is delayed, moves away from S_1, and may even disappear. Some patients have a mid-systolic click without a murmur; others have a murmur without a click. Still others have both sounds at different times.

Laboratory Examination

The ECG most commonly is normal but may show biphasic or inverted T waves in leads II, III, and aVF, and occasionally supraventricular or ventricular premature beats. TTE is particularly effective in identifying the abnormal position and prolapse of the mitral valve leaflets. A useful echocardiographic definition of MVP is systolic displacement (in the parasternal long axis view) of the mitral valve leaflets by at least 2 mm into the LA superior to the plane of the mitral annulus. Color flow and continuous wave Doppler imaging is helpful to evaluate the associated MR and provide semiquantitative estimates of severity. The jet lesion of MR due to MVP is most often eccentric, and assessment of RF and effective regurgitant orifice area can be difficult. TEE is indicated when more accurate information is required and is performed routinely for intraoperative guidance for valve repair. Invasive left ventriculography is rarely necessary but can also show prolapse of the posterior and sometimes of both mitral valve leaflets.

Treatment: Mitral Valve Prolapse

Infective endocarditis prophylaxis is indicated only for patients with a prior history of endocarditis. Beta blockers sometimes relieve chest pain and control palpitations. If the patient is symptomatic from severe MR, mitral valve repair (or rarely, replacement) is indicated (Fig. 237-4). Antiplatelet agents, such as aspirin, should be given to patients with transient ischemic attacks, and if these are not effective, anticoagulants, such as warfarin, should be considered. Warfarin is also indicated once AF intervenes.

AORTIC STENOSIS

AS occurs in about one-fourth of all patients with chronic valvular heart disease; approximately 80% of adult patients with symptomatic valvular AS are male.

Etiology and Pathogenesis

(Table 237-1)

AS in adults is due to degenerative calcification of the aortic cusps and occurs most commonly on a substrate of congenital disease (bicuspid aortic valve [BAV]), chronic (tri-)leaflet deterioration, or previous rheumatic inflammation. A recent pathologic study of specimens removed at the time of aortic valve replacement for AS showed that 53% were bicuspid and 4% unicuspid. Contrary to previous teachings, the process of aortic valve deterioration and calcification is not a passive one, but rather one that shares many features with vascular atherosclerosis, including endothelial dysfunction, lipid accumulation, inflammatory cell activation, cytokine release, and upregulation of several signaling pathways (Fig. 237-6). Eventually, valvular myofibroblasts differentiate phenotypically into osteoblasts and actively produce bone matrix proteins that allow for the deposition of calcium hydroxyapatite crystals. Genetic polymorphisms involving the vitamin D receptor, the estrogen receptor in postmenopausal women, interleukin 10, and apolipoprotein E4 have been linked to the development of calcific AS, and a strong familial clustering of cases has been reported from western France. Several traditional atherosclerotic risk factors have also been associated with the development and progression of calcific AS, including low-density lipoprotein (LDL)-cholesterol, lipoprotein a [Lp(a)], diabetes mellitus, smoking, chronic kidney disease, and the metabolic syndrome. The presence of aortic valve sclerosis (focal thickening and calcification of the leaflets not severe enough to cause obstruction) is associated with an excess risk of cardiovascular death and MI among persons older than age 65. Approximately 30% of persons older than 65 years exhibit aortic valve sclerosis, whereas 2% exhibit frank stenosis.

Figure 237-6

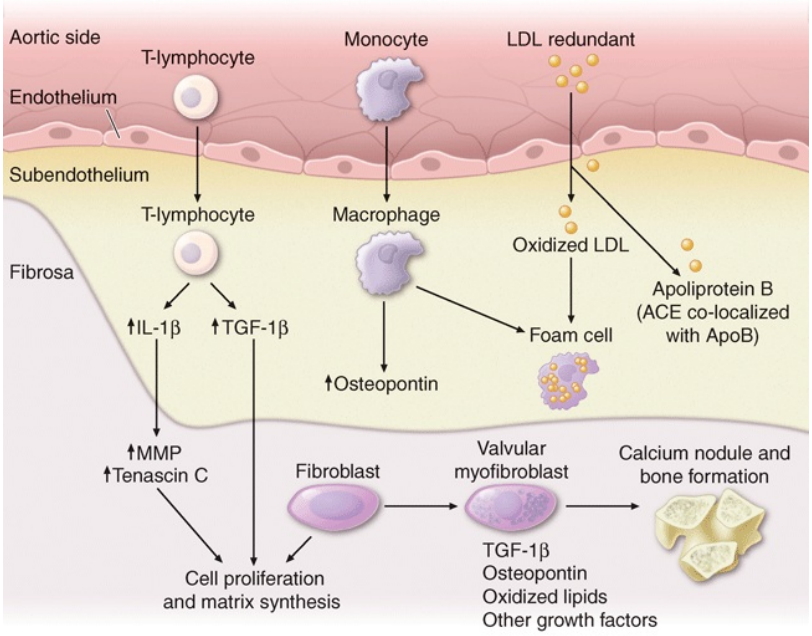

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis of Calcific Aortic Stenosis. Inflammatory cells infiltrate across the endothelial barrier and release cytokines

Pathogenesis of Calcific Aortic Stenosis: Inflammatory cells infiltrate across the endothelial barrier and release cytokines that act on fibroblasts to promote cellular proliferation and matrix remodeling. LDL is oxidatively modified and taken up by macrophage scavengers to become foam cells. Angiotensin-converting enzyme colocalizes with Apo-B. A subset of myofibroblasts differentiate into an osteoblast phenotype capable of promoting bone formation. ACE, angiotensin-converting enzyme; ApoB, apolipoprotein B; LDL, low-density lipoprotein; IL, interleukin; MMP, matrix metalloproteinase; TGF, transforming growth factor. *(From RV Freeman, CM Otto: Circulation 111:3316, 2005; with permission.)*

Rheumatic disease of the aortic leaflets produces commissural fusion, sometimes resulting in a bicuspid-appearing valve. This condition, in turn, makes the leaflets more susceptible to trauma and ultimately leads to fibrosis, calcification, and further narrowing. By the time the obstruction to LV outflow causes serious clinical disability, the valve is usually a rigid calcified mass, and careful examination may make it difficult or even impossible to determine the etiology of the underlying process. Rheumatic AS is almost always associated with involvement of the mitral valve and with AR. Mediastinal radiation can also result in late scarring, fibrosis, and calcification of the leaflets with AS.

Bicuspid Aortic Valve Disease

A bicuspid aortic valve (BAV) is the most common congenital heart valve defect and occurs in 0.5–1.4% of the population with a 2–4:1 male to female predominance. The inheritance pattern appears to be autosomal dominant with incomplete penetrance, although some have questioned an X-linked component as suggested by the prevalence of BAV disease among patients with Turner's syndrome. The prevalence of BAV disease among first-degree relatives of an affected individual is approximately 10%. A single gene defect to explain the majority of cases has not been identified, although a mutation in the *NOTCH1* gene has been described in some families. Abnormalities in endothelial nitric oxide synthase and NKX2.5 have been implicated, as well. Aortic coarctation or medial degeneration with ascending aortic aneurysm formation occurs commonly among patients with BAV disease. Patients with BAV disease have larger aortas than patients with comparable tricuspid aortic valve disease. The aortopathy develops independent of the hemodynamic severity of the valve lesion and is a risk factor for dissection. A BAV can be a component of more complex congenital heart disease with or without other left heart obstructing lesions.

Other Forms of Obstruction to Left Ventricular Outflow

In addition to valvular AS, three other lesions may be responsible for obstruction to LV outflow: *hypertrophic obstructive cardiomyopathy* (Chap. 238), *discrete fibromuscular/membranous subaortic stenosis*, and *supravalvular AS* (Chap. 236). The causes of left ventricular outflow obstruction can be differentiated on the basis of the cardiac examination and Doppler echocardiographic findings.

Pathophysiology

The obstruction to LV outflow produces a systolic pressure gradient between the LV and aorta. When severe obstruction is suddenly produced experimentally, the LV responds by dilation and reduction of stroke volume. However, in some patients, the obstruction may be present at birth and/or increase gradually over the course of many years, and LV contractile performance is maintained by the presence of concentric LV hypertrophy. Initially, this serves as an adaptive mechanism because it reduces toward normal the systolic stress developed by the myocardium, as predicted by the Laplace relation ($S = Pr/h$, where S = systolic wall stress, P = pressure, r = radius, and h = wall thickness). A large transaortic valve pressure gradient may exist for many years without a reduction in CO or LV dilation; ultimately, however, excessive hypertrophy becomes maladaptive, LV systolic function declines, abnormalities of diastolic function progress, and irreversible myocardial fibrosis develops.

A mean systolic pressure gradient >40 mmHg with a normal CO or an effective aortic orifice area <~1 cm^2 (or ~<0.6 cm^2/m^2 body surface area in a normal-sized adult)—i.e., less than approximately one-third of the normal orifice area—is generally considered to represent severe obstruction to LV outflow. The elevated LV end-diastolic pressure observed in many patients with severe AS and preserved EF signifies the presence of diminished compliance of the hypertrophied LV. Although the CO at rest is within normal limits in most patients with severe AS, it usually fails to rise normally during exercise. Loss of an appropriately timed, vigorous atrial contraction, as occurs in AF or atrioventricular dissociation, may cause rapid progression of symptoms. Late in the course, contractile function deteriorates because of afterload excess, the CO and LV–aortic pressure gradient decline, and the mean LA, PA, and RV pressures rise. LV performance can be further compromised by superimposed CAD.

The hypertrophied LV causes an increase in myocardial oxygen requirements. In addition, even in the absence of obstructive CAD, coronary blood flow is impaired to the extent that ischemia can be precipitated under conditions of excess demand. Capillary density is reduced relative to wall thickness, compressive forces are increased, and the elevated LV end-diastolic pressure reduces the coronary driving pressure. The subendocardium is especially vulnerable to ischemia by this mechanism.

Symptoms

AS is rarely of clinical importance until the valve orifice has narrowed to approximately 1 cm^2. Even severe AS may exist for many years without producing any symptoms because of the ability of the hypertrophied LV to generate the elevated intraventricular pressures required to maintain a normal stroke volume. Once symptoms occur, valve replacement is indicated.

Most patients with pure or predominant AS have gradually increasing obstruction over years, but do not become symptomatic until the sixth to eighth decades. Adult patients with BAV disease, however, develop significant valve dysfunction and symptoms one to two decades sooner. Exertional dyspnea, angina pectoris, and syncope are the three cardinal symptoms. Often, there is a history of insidious progression of fatigue and dyspnea associated with gradual curtailment of activities. *Dyspnea* results primarily from elevation of the pulmonary capillary pressure caused by elevations of LV diastolic pressures secondary to reduced left ventricular compliance and impaired relaxation. *Angina pectoris* usually develops somewhat later and reflects an imbalance between the augmented myocardial oxygen requirements and reduced oxygen availability. CAD may or may not be present, although its coexistence is common among AS patients older than age 65. *Exertional syncope* may result from a decline in arterial pressure caused by vasodilation in the exercising muscles and inadequate vasoconstriction in nonexercising muscles in the face of a fixed CO, or from a sudden fall in CO produced by an arrhythmia.

Because the CO at rest is usually well maintained until late in the course, marked fatigability, weakness, peripheral cyanosis, cachexia, and other clinical manifestations of a low CO are usually not prominent until this stage is reached. Orthopnea, paroxysmal nocturnal dyspnea, and pulmonary edema, i.e., symptoms of LV failure, also occur only in the advanced stages of the disease. Severe pulmonary hypertension leading to RV failure and systemic venous hypertension, hepatomegaly, AF, and TR are usually late findings in patients with isolated severe AS.

When AS and MS coexist, the reduction in CO induced by MS lowers the pressure gradient across the aortic valve and, thereby, masks many of the clinical findings produced by AS.

Physical Findings

The rhythm is generally regular until late in the course; at other times, AF should suggest the possibility of associated mitral valve disease. The systemic arterial pressure is usually within normal limits. In the late stages, however, when stroke volume declines, the systolic pressure may fall and the pulse pressure narrow. The peripheral arterial pulse rises slowly to a delayed peak (*pulsus parvus et tardus*. A thrill or anacrotic "shudder" may be palpable over the carotid arteries, more commonly the left. In the elderly, the stiffening of the arterial wall may mask this important physical sign. In many patients, the *a* wave in the jugular

venous pulse is accentuated. This results from the diminished distensibility of the RV cavity caused by the bulging, hypertrophied interventricular septum.

The LV impulse is usually displaced laterally. A double apical impulse (with a palpable S_4) may be recognized, particularly with the patient in the left lateral recumbent position. A systolic thrill may be present at the base of the heart to the right of the sternum when leaning forward or in the suprasternal notch.

AUSCULTATION

An early systolic ejection sound is frequently audible in children, adolescents, and young adults with congenital BAV disease. This sound usually disappears when the valve becomes calcified and rigid. As AS increases in severity, LV systole may become prolonged so that the aortic valve closure sound no longer precedes the pulmonic valve closure sound, and the two components may become synchronous, or aortic valve closure may even follow pulmonic valve closure, causing paradoxical splitting of S_2 (Chap. 227). The sound of aortic valve closure can be heard most frequently in patients with AS who have pliable valves, and calcification diminishes the intensity of this sound. Frequently, an S_4 is audible at the apex and reflects the presence of LV hypertrophy and an elevated LV end-diastolic pressure; an S_3 generally occurs late in the course, when the LV dilates and its systolic function becomes severely compromised.

The murmur of AS is characteristically an ejection (mid) systolic murmur that commences shortly after the S_1, increases in intensity to reach a peak toward the middle of ejection, and ends just before aortic valve closure. It is characteristically low-pitched, rough and rasping in character, and loudest at the base of the heart, most commonly in the second right intercostal space. It is transmitted upward along the carotid arteries. Occasionally it is transmitted downward and to the apex, where it may be confused with the systolic murmur of MR (Gallavardin effect). In almost all patients with severe obstruction and preserved CO, the murmur is at least grade III/VI. In patients with mild degrees of obstruction or in those with severe stenosis with heart failure and low CO in whom the stroke volume and, therefore, the transvalvular flow rate are reduced, the murmur may be relatively soft and brief.

Laboratory Examination

ECG

In most patients with severe AS there is LV hypertrophy. In advanced cases, ST-segment depression and T-wave inversion (LV "strain") in standard leads I and aVL and in the left precordial leads are evident. However, there is no close correlation between the ECG and the hemodynamic severity of obstruction, and the absence of ECG signs of LV hypertrophy does not exclude severe obstruction. Many patients with AS have systemic hypertension, which can also contribute to the development of hypertrophy.

ECHOCARDIOGRAM

The key findings on TTE are thickening, calcification, and reduced systolic opening of the valve leaflets and LV hypertrophy. Eccentric closure of the aortic valve cusps is characteristic of congenitally bicuspid valves. TEE imaging usually displays the obstructed orifice extremely well, but it is not routinely required for accurate characterization. The valve gradient and aortic valve area can be estimated by Doppler measurement of the transaortic velocity. Severe AS is defined by a valve area <1 cm^2, whereas moderate AS is defined by a valve area of 1–1.5 cm^2 and mild AS by a valve area of 1.5–2 cm^2. Aortic valve sclerosis, conversely, is accompanied by a jet velocity of less than 2.5 meters/s (peak gradient <25 mmHg). LV dilation and reduced systolic shortening reflect impairment of LV function. There is increasing experience with the use of longitudinal strain and strain rate to characterize earlier changes in LV systolic function, well before a decline in EF can be appreciated. Doppler indices of impaired diastolic function are frequently seen.

Echocardiography is useful for identifying coexisting valvular abnormalities; for differentiating valvular AS from other forms of LV outflow obstruction; and for measurement of the aortic root and proximal ascending aortic dimension. These aortic measurements are particularly important for patients with BAV disease. Dobutamine stress echocardiography is useful for the evaluation of patients with AS and severe LV systolic dysfunction (EF <0.35), in whom the severity of the AS can often be difficult to judge.

CHEST X-RAY

The chest x-ray may show no or little overall cardiac enlargement for many years. Hypertrophy without dilation may produce some rounding of the cardiac apex in the frontal projection and slight backward displacement in the lateral view. A dilated proximal ascending aorta may be seen along the upper right heart border in the frontal view. Aortic valve calcification may be discernible in the lateral view, but is usually readily apparent on fluoroscopic examination or by echocardiography; the absence of valvular calcification in an adult suggests that severe valvular AS is *not* present. In later stages of the disease, as the LV dilates there is increasing roentgenographic evidence of LV enlargement, pulmonary congestion, and enlargement of the LA, PA, and right heart chambers.

CATHETERIZATION

Right and left heart catheterization for invasive assessment of AS is now performed infrequently but can be useful when there is a discrepancy between the clinical and Doppler echocardiographic findings. Appropriate concerns have been raised that attempts to cross the aortic valve for measurement of left ventricular pressures are associated with a risk of cerebral embolization. Catheterization is also useful in three distinct categories of patients: (1) *patients with multivalvular disease,* in whom the role played by each valvular deformity should be defined to aid in the planning of operative treatment; (2) *young, asymptomatic patients with noncalcific congenital AS,* to define the severity of obstruction to LV outflow, since operation or percutaneous aortic balloon valvoplasty (PABV) may be indicated in these patients if severe AS is present, even in the absence of symptoms; balloon valvotomy may follow left heart catheterization in the same sitting; and (3) *patients in whom it is suspected that the obstruction to LV outflow may not be at the level of the aortic valve* but rather at the sub- or supravalvular level.

Coronary angiography is indicated to detect or exclude CAD in appropriate patients with severe AS who are being considered for surgery. The incidence of significant CAD for which bypass grafting is indicated at the time of aortic valve replacement (AVR) exceeds 50% among adult patients.

Natural History

Death in patients with severe AS occurs most commonly in the seventh and eighth decades. Based on data obtained at postmortem examination in patients before surgical treatment became widely available, the average time to death after the onset of various symptoms was as follows: angina pectoris, 3 years; syncope, 3 years; dyspnea, 2 years; congestive heart failure, 1.5–2 years. Moreover, in >80% of patients who died with AS, symptoms had existed for <4 years. Among adults dying with valvular AS, sudden death, which presumably resulted from an arrhythmia, occurred in 10–20%; however, most sudden deaths occurred in patients who had previously been symptomatic. Sudden death as the first manifestation of severe AS is very uncommon (<1% per year) in asymptomatic adult patients. Calcific AS is a progressive disease, with an annual reduction in valve area averaging

0.1 cm^2 and annual increases in the peak jet velocity and mean valve gradient averaging 0.3 meters/s and 7 mmHg, respectively (Table 237-2, Fig. 237-7).

Figure 237-7

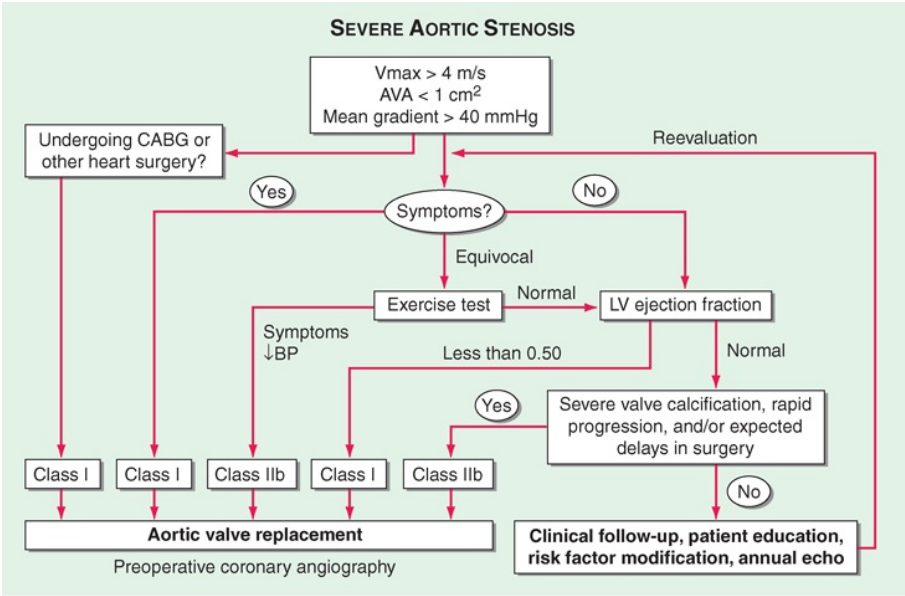

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management Strategy for Patients with Severe Aortic Stenosis. Preoperative coronary angiography should be performed routinely as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical findings and echocardiography. AVA, aortic valve area; BP, blood pressure; CABG, coronary artery bypass graft surgery; echo, echocardiography; LV, left ventricle; Vmax, maximal velocity across aortic valve by Doppler echocardiography. *(From Bonow et al. Modified from CM Otto: J Am Coll Cardiol 47:2141, 2006.)*

Treatment: Aortic Stenosis

MEDICAL TREATMENT

In patients with severe AS (valve area <1 cm^2), strenuous physical activity and competitive sports should be avoided, even in the asymptomatic stage. Care must be taken to avoid dehydration and hypovolemia to protect against a significant reduction in CO. Medications used for the treatment of hypertension or CAD, including beta blockers and ACE inhibitors, are generally safe for asymptomatic patients with preserved left ventricular systolic function. Nitroglycerin is helpful in relieving angina pectoris in patients with CAD. Retrospective studies have shown that patients with degenerative calcific AS who receive HMG-CoA reductase inhibitors ("statins") exhibit slower progression of leaflet calcification and aortic valve area reduction than those who do not. However, randomized prospective studies with either high-dose atorvastatin or combination simvastatin/ezetimibe have failed to show a measurable effect on valve-related outcomes. The use of statin medications should continue to be driven by considerations regarding primary and secondary prevention of CAD. ACE inhibitors have not been studied prospectively for AS-related outcomes. The need for endocarditis prophylaxis is restricted to AS patients with a prior history of endocarditis.

SURGICAL TREATMENT

Asymptomatic patients with calcific AS and severe obstruction should be followed carefully for the development of symptoms and by serial echocardiograms for evidence of deteriorating LV function. Operation is indicated in patients with severe AS (valve area <1 cm^2 or 0.6 cm^2/m^2 body surface area) who are symptomatic, those who exhibit LV dysfunction (EF <50%), as well as those with BAV disease and an aneurysmal or expanding aortic root (maximal dimension >4.5 cm or annual increase in size >0.5 cm/year), even if they are asymptomatic. Patients with asymptomatic moderate or severe AS who are referred for CABG surgery should also have AVR. In patients without heart failure, the operative risk of AVR is approximately 3% (Table 237-3) but increases as a function of age and the need for concomitant coronary revascularization with bypass grafting. The indications for AVR in the asymptomatic patient have been the subject of intense debate over the past 5 years, as surgical outcomes in selected patients have continued to improve. Relative indications for which surgery can be considered include an abnormal response to treadmill exercise; rapid progression of AS, especially when urgent access to medical care might be compromised; very severe AS defined by a valve area <0.6 cm^2; and severe LV hypertrophy suggested by a wall thickness of >15 mm. Exercise testing can be safely performed in the asymptomatic patient, as many as one-third of whom will show signs of functional impairment.

Operation should be carried out within 3–4 months of symptom onset and certainly well before frank LV failure develops; at this late stage, the aortic valve pressure gradient declines in parallel with the CO and stroke volume (low gradient, low output AS). In such patients, the perioperative mortality risk is high (15–20%), and evidence of myocardial disease may persist even when the operation is technically successful. Long-term postoperative survival correlates with preoperative LV function. Nonetheless, in view of the even worse prognosis of such patients when they are treated medically, there is usually little choice but to advise surgical treatment, especially in patients in whom contractile reserve can be demonstrated by dobutamine echocardiography (defined by a ≥20% in stroke volume after dobutamine challenge). In patients in whom severe AS and CAD coexist, relief of the AS and revascularization may sometimes result in striking clinical and hemodynamic improvement (Table 237-3).

Because many patients with calcific AS are elderly, particular attention must be directed to the adequacy of hepatic, renal, and pulmonary function before AVR is recommended. Age alone is not a contraindication to AVR for AS. The mortality rate depends to a substantial extent on the patient's preoperative clinical and hemodynamic state. The 10-year survival rate of patients with AVR is approximately 60%. Approximately 30% of bioprosthetic valves evidence primary valve failure in 10 years, requiring re-replacement, and an approximately equal percentage of patients with mechanical prostheses develop significant hemorrhagic complications as a consequence of treatment with anticoagulants (p. 1949). Homograft AVR is usually reserved for patients with aortic valve endocarditis.

The Ross procedure involves replacement of the diseased aortic valve with the autologous pulmonic valve and implantation of a homograft in the native pulmonic position. Its use has declined considerably in the United States because of the technical complexity of the procedure and the incidence of late postoperative aortic root dilation and autograft failure with AR. There is also a low incidence of homograft stenosis.

PERCUTANEOUS BALLOON AORTIC VALVULOPLASTY

This procedure is preferable to operation in many children and young adults with congenital, noncalcific AS (Chap. 236). It is not commonly used in adults with severe calcific AS because of a very high restenosis rate (80% within 1 year) and the risk of procedural complications, but on occasion it has been used successfully as a "bridge to operation" in patients with severe LV dysfunction and shock who are too ill to tolerate surgery.

PERCUTANEOUS AORTIC VALVE REPLACEMENT

Transcatheter aortic valve implantation (TAVI) for treatment of AS has been performed in more than 20,000 high-risk (Society for Thoracic Surgery mortality risk >10%) adult patients worldwide using one of two available systems, a balloon-expandable valve and a self-expanding valve, both of which incorporate a pericardial prosthesis (Fig. 237-8). The aorto-iliofemoral anatomy must allow for passage of large-bore catheters; a direct, trans-LV apical approach under surgical guidance is an alternative means of delivering the valve. There are also several reports of the successful use of axillary or subclavian artery access with construction of a vascular surgical sleeve. Aortic balloon valvuloplasty is performed as a first step to create an orifice sufficient for the prosthesis. Procedural success rates now exceed 90%, and intermediate-term prosthetic function is excellent. A mild degree of paravalvular AR is common with TAVI; postprocedural heart block is observed more frequently with the self-expanding valve. Preliminary results with TAVI have been very favorable (Fig. 237-9), and it is anticipated that this technology, now clinically available in Canada and Europe, will gain regulatory approval in the United States for the treatment of patients with severe AS considered too high risk for surgical AVR. For elderly patients with symptomatic, severe AS who are deemed too high risk for surgery, TAVI has been shown in a randomized prospective trial to prolong life and improve functionality. The use of these devices for the treatment of prosthetic valve failure not due to paravalvular regurgitation ("valve-in-valve"), as an alternative to reoperative valve replacement, is also under active study.

Figure 237-8.

A. Balloon-expandable and ***B*** and self-expanding valves for percutaneous aortic valve replacement. B, inflated balloon; N, nose cone; V, valve. *(Part A, courtesy of Edwards Lifesciences, Irvine, CA; with permission. RetroFlex 3 is a trademark of Edwards Lifesciences Corporation. Part B, © Medtronic, Inc. 2010. Medtronic CoreValve Transcatheter Aortic Valve. CoreValve is a registered trademark of Medtronic, Inc.)*

Figure 237-9

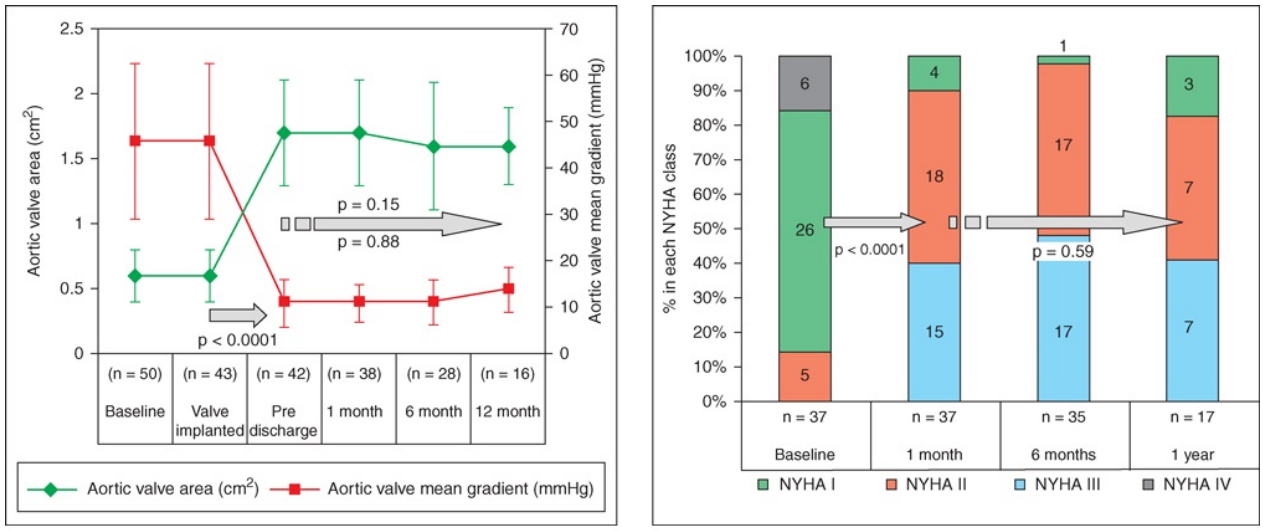

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Twelve-Month Outcomes Following Percutaneous Aortic Valve Replacement. *(Adapted from JG Webb et al: Circulation 116:755, 2007; with permission.)*

AORTIC REGURGITATION

Etiology

(Table 237-1) AR may be caused by primary valve disease or by primary aortic root disease.

PRIMARY VALVE DISEASE

Rheumatic disease results in thickening, deformity, and shortening of the individual aortic valve cusps, changes that prevent their proper opening during systole and closure during diastole. A rheumatic origin is much less common in patients with isolated AR who do not have associated rheumatic mitral valve disease. Patients with congenital BAV disease may develop predominant AR, and approximately 20% of patients will require aortic valve surgery between 10 and 40 years of age. Congenital fenestrations of the aortic valve occasionally produce mild AR. Membranous subaortic stenosis often leads to thickening and scarring of the aortic valve leaflets with secondary AR. Prolapse of an aortic cusp, resulting in progressive chronic AR, occurs in approximately 15% of patients with ventricular septal defect (Chap. 236) but may also occur as an isolated phenomenon or as a consequence of myxomatous degeneration sometimes associated with mitral (p. 1937) and/or tricuspid valve involvement.

AR may result from infective endocarditis, which can develop on a valve previously affected by rheumatic disease, a congenitally deformed valve, or on a normal aortic valve, and may lead to perforation or erosion of one or more leaflets. The aortic valve leaflets may become scarred and retracted during the course of syphilis or ankylosing spondylitis and contribute further to the

AR that derives primarily from the associated root disease. Although traumatic rupture or avulsion of the aortic valve is an uncommon cause of acute AR, it does represent the most frequent serious lesion in patients surviving nonpenetrating cardiac injuries. The coexistence of hemodynamically significant AS with AR usually excludes all the rarer forms of AR because it occurs almost exclusively in patients with rheumatic or congenital AR. In patients with AR due to primary valvular disease, dilation of the aortic annulus may occur secondarily and lead to worsening regurgitation.

PRIMARY AORTIC ROOT DISEASE

AR may also be due entirely to marked aortic dilation, i.e., aortic root disease, without primary involvement of the valve leaflets; widening of the aortic annulus and separation of the aortic leaflets are responsible for the AR (Chap. 248). Cystic medial degeneration of the ascending aorta, which may or may not be associated with other manifestations of Marfan's syndrome; idiopathic dilation of the aorta; annuloaortic ectasia; osteogenesis imperfecta; and severe hypertension may all widen the aortic annulus and lead to progressive AR. Occasionally AR is caused by retrograde dissection of the aorta involving the aortic annulus. Syphilis and ankylosing spondylitis, both of which may affect aortic valves, may also be associated with cellular infiltration and scarring of the media of the thoracic aorta, leading to aortic dilation, aneurysm formation, and severe regurgitation. In syphilis of the aorta (Chap. 169), now a very rare condition, the involvement of the intima may narrow the coronary ostia, which in turn may be responsible for myocardial ischemia.

Pathophysiology

The total stroke volume ejected by the LV (i.e., the sum of the effective forward stroke volume and the volume of blood that regurgitates back into the LV) is increased in patients with AR. In patients with severe AR, the volume of regurgitant flow may equal the effective forward stroke volume. In contrast to MR, in which a portion of the LV stroke volume is delivered into the low-pressure LA, in AR the entire LV stroke volume is ejected into a high-pressure zone, the aorta. An increase in the LV end-diastolic volume (increased preload) constitutes the major hemodynamic compensation for AR. The dilation and eccentric hypertrophy of the LV allow this chamber to eject a larger stroke volume without requiring any increase in the relative shortening of each myofibril. Therefore, severe AR may occur with a normal effective forward stroke volume and a normal left ventricular EF [total (forward plus regurgitant) stroke volume/end-diastolic volume], together with an elevated LV end-diastolic pressure and volume. However, through the operation of Laplace's law, LV dilation increases the LV systolic tension required to develop any given level of systolic pressure. Chronic AR is, thus, a state in which LV preload and afterload are both increased. Ultimately, these adaptive measures fail. As LV function deteriorates, the end-diastolic volume rises further and the forward stroke volume and EF decline. Deterioration of LV function often precedes the development of symptoms. Considerable thickening of the LV wall also occurs with chronic AR, and at autopsy the hearts of these patients may be among the largest encountered, sometimes weighing >1000 g.

The reverse pressure gradient from aorta to LV, which drives the AR flow, falls progressively during diastole, accounting for the decrescendo nature of the diastolic murmur. Equilibration between aortic and LV pressures may occur toward the end of diastole in patients with chronic severe AR, particularly when the heart rate is slow. In patients with acute severe AR, the LV is unprepared for the regurgitant volume load. LV compliance is normal or reduced, and LV diastolic pressures rise rapidly, occasionally to levels >40 mmHg. The LV pressure may exceed the LA pressure toward the end of diastole, and this reversed pressure gradient closes the mitral valve prematurely.

In patients with chronic severe AR, the effective forward CO usually is normal or only slightly reduced at rest, but often it fails to rise normally during exertion. An early sign of LV dysfunction is a reduction in the EF. In advanced stages there may be considerable elevation of the LA, PA wedge, PA, and RV pressures and lowering of the forward CO at rest.

Myocardial ischemia may occur in patients with AR because myocardial oxygen requirements are elevated by LV dilation, hypertrophy, and elevated LV systolic tension, and coronary blood flow may be compromised. A large fraction of coronary blood flow occurs during diastole, when arterial pressure is low, thereby reducing coronary perfusion or driving pressure. This combination of increased oxygen demand and reduced supply may cause myocardial ischemia, particularly of the subendocardium, even in the absence of concomitant CAD.

History

Approximately three-fourths of patients with pure or predominant valvular AR are men; women predominate among patients with primary valvular AR who have associated rheumatic mitral valve disease. A history compatible with infective endocarditis may sometimes be elicited from patients with rheumatic or congenital involvement of the aortic valve, and the infection often precipitates or seriously aggravates preexisting symptoms.

In patients with *acute severe AR,* as may occur in infective endocarditis, aortic dissection, or trauma, the LV cannot dilate sufficiently to maintain stroke volume, and LV diastolic pressure rises rapidly with associated marked elevations of LA and PA wedge pressures. Pulmonary edema and/or cardiogenic shock may develop rapidly.

Chronic severe AR may have a long latent period, and patients may remain relatively asymptomatic for as long as 10–15 years. However, uncomfortable awareness of the heartbeat, especially on lying down, may be an early complaint. Sinus tachycardia, during exertion or with emotion, or premature ventricular contractions may produce particularly uncomfortable palpitations as well as head pounding. These complaints may persist for many years before the development of exertional dyspnea, usually the first symptom of diminished cardiac reserve. The dyspnea is followed by orthopnea, paroxysmal nocturnal dyspnea, and excessive diaphoresis. Anginal chest pain even in the absence of CAD may occur in patients with severe AR, even in younger patients. Anginal pain may develop at rest as well as during exertion. Nocturnal angina may be a particularly troublesome symptom, and it may be accompanied by marked diaphoresis. The anginal episodes can be prolonged and often do not respond satisfactorily to sublingual nitroglycerin. Systemic fluid accumulation, including congestive hepatomegaly and ankle edema, may develop late in the course of the disease.

Physical Findings

In chronic severe AR, the jarring of the entire body and the bobbing motion of the head with each systole can be appreciated, and the abrupt distention and collapse of the larger arteries are easily visible. The examination should be directed toward the detection of conditions predisposing to AR, such as bicuspid valve, endocarditis, Marfan's syndrome, and ankylosing spondylitis.

ARTERIAL PULSE

A rapidly rising "water-hammer" pulse, which collapses suddenly as arterial pressure falls rapidly during late systole and diastole (Corrigan's pulse), and capillary pulsations, an alternate flushing and paling of the skin at the root of the nail while pressure is applied to the tip of the nail (Quincke's pulse), are characteristic of chronic severe AR. A booming "pistol-shot" sound can be heard over the femoral arteries (Traube's sign), and a to-and-fro murmur (Duroziez's sign) is audible if the femoral artery is lightly compressed with a stethoscope.

The arterial pulse pressure is widened as a result of both systolic hypertension and a lowering of the diastolic pressure. The measurement of arterial diastolic pressure with a sphygmomanometer may be complicated by the fact that systolic sounds are frequently heard with the cuff completely deflated. However, the level of cuff pressure at the time of muffling of the Korotkoff sounds (phase IV) generally corresponds fairly closely to the true intraarterial diastolic pressure. As the disease progresses and

the LV end-diastolic pressure rises, the arterial diastolic pressure may actually rise as well, because the aortic diastolic pressure cannot fall below the LV end-diastolic pressure. For the same reason, acute severe AR may also be accompanied by only a slight widening of the pulse pressure. Such patients are invariably tachycardic as the heart rate increases in an attempt to preserve the CO.

PALPATION

In patients with chronic severe AR, the LV impulse is heaving and displaced laterally and inferiorly. The systolic expansion and diastolic retraction of the apex are prominent. A diastolic thrill may be palpable along the left sternal border in thin-chested individuals, and a prominent systolic thrill may be palpable in the suprasternal notch and transmitted upward along the carotid arteries. This systolic thrill and the accompanying murmur do not necessarily signify the coexistence of AS. In many patients with pure AR or with combined AS and AR, the carotid arterial pulse is bisferiens, i.e., with two systolic waves separated by a trough (see Fig. 227-2*D*).

AUSCULTATION

In patients with severe AR, the aortic valve closure sound (A_2) is usually absent. A systolic ejection sound is audible in patients with BAV disease, and occasionally an S_4 also may be heard. The murmur of chronic AR is typically a high-pitched, blowing, decrescendo diastolic murmur, heard best in the third intercostal space along the left sternal border (see Fig. 227-5*B*). In patients with mild AR, this murmur is brief, but as the severity increases, it generally becomes louder and longer, indeed holodiastolic. When the murmur is soft, it can be heard best with the diaphragm of the stethoscope and with the patient sitting up, leaning forward, and with the breath held in forced expiration. In patients in whom the AR is caused by primary valvular disease, the diastolic murmur is usually louder along the left than the right sternal border. However, when the murmur is heard best along the right sternal border, it suggests that the AR is caused by aneurysmal dilation of the aortic root. "Cooing" or musical diastolic murmurs suggest eversion of an aortic cusp vibrating in the regurgitant stream.

A mid-systolic ejection murmur is frequently audible in isolated AR. It is generally heard best at the base of the heart and is transmitted along the carotid arteries. This murmur may be quite loud without signifying aortic obstruction. A third murmur sometimes heard in patients with severe AR is the *Austin Flint murmur,* a soft, low-pitched, rumbling mid-to-late diastolic murmur. It is probably produced by the diastolic displacement of the anterior leaflet of the mitral valve by the AR stream and is not associated with hemodynamically significant mitral obstruction. The auscultatory features of AR are intensified by strenuous and sustained handgrip, which augments systemic vascular resistance.

In acute severe AR, the elevation of LV end-diastolic pressure may lead to early closure of the mitral valve, a soft S_1, a pulse pressure that is not particularly wide, and a soft, short, early diastolic murmur of AR.

Laboratory Examination

ECG

In patients with chronic severe AR, the ECG signs of LV hypertrophy become manifest (Chap. 228). In addition, these patients frequently exhibit ST-segment depression and T-wave inversion in leads I, aVL, V_5, and V_6 ("LV strain"). Left-axis deviation and/or QRS prolongation denote diffuse myocardial disease, generally associated with patchy fibrosis, and usually signify a poor prognosis.

ECHOCARDIOGRAM

LV size is increased in chronic AR and systolic function is normal or even supernormal until myocardial contractility declines, as signaled by a decrease in ejection or increase in the end-systolic dimension. A rapid, high-frequency diastolic fluttering of the anterior mitral leaflet produced by the impact of the regurgitant jet is a characteristic finding. The echocardiogram is also useful in determining the cause of AR, by detecting dilation of the aortic annulus and root, aortic dissection (see Fig. 229-5), or primary leaflet pathology. With severe AR, the central jet width assessed by color flow Doppler imaging exceeds 65% of the left ventricular outflow tract, the regurgitant volume is ≥60 mL/beat, the regurgitant fraction is ≥50%, and there is diastolic flow reversal in the proximal descending thoracic aorta. The continuous wave Doppler profile shows a rapid deceleration time in patients with acute severe AR, due to the rapid increase in LV diastolic pressure. Surveillance transthoracic echocardiography forms the cornerstone of longitudinal follow-up and allows for the early detection of changes in LV size and/or function. For patients in whom echocardiography is limited by poor acoustical windows or inadequate semiquantitative assessment, gated cardiac MR imaging can be performed. This modality also allows for accurate assessment of aortic size and contour.

CHEST X-RAY

In chronic severe AR, the apex is displaced downward and to the left in the frontal projection. In the left anterior oblique and lateral projections, the LV is displaced posteriorly and encroaches on the spine. When AR is caused by primary disease of the aortic root, aneurysmal dilation of the aorta may be noted, and the aorta may fill the retrosternal space in the lateral view. Echocardiography, cardiac MR, and CT angiography are more sensitive than the chest x-ray for the detection of aortic root enlargement.

CARDIAC CATHETERIZATION AND ANGIOGRAPHY

When needed, right and left heart catheterization with contrast aortography can provide confirmation of the magnitude of regurgitation and the status of LV function. Coronary angiography is performed routinely in appropriate patients prior to surgery.

Treatment: Aortic Regurgitation

ACUTE AORTIC REGURGITATION

(See Fig. 237-10) Patients with acute severe AR may respond to intravenous diuretics and vasodilators (such as sodium nitroprusside), but stabilization is usually short-lived and operation is indicated urgently. Intraaortic balloon counterpulsation is contraindicated. Beta blockers are also best avoided so as not to reduce the CO further or slow the heart rate. Surgery is the treatment of choice and is usually necessary within 24 hours of diagnosis.

Figure 237-10

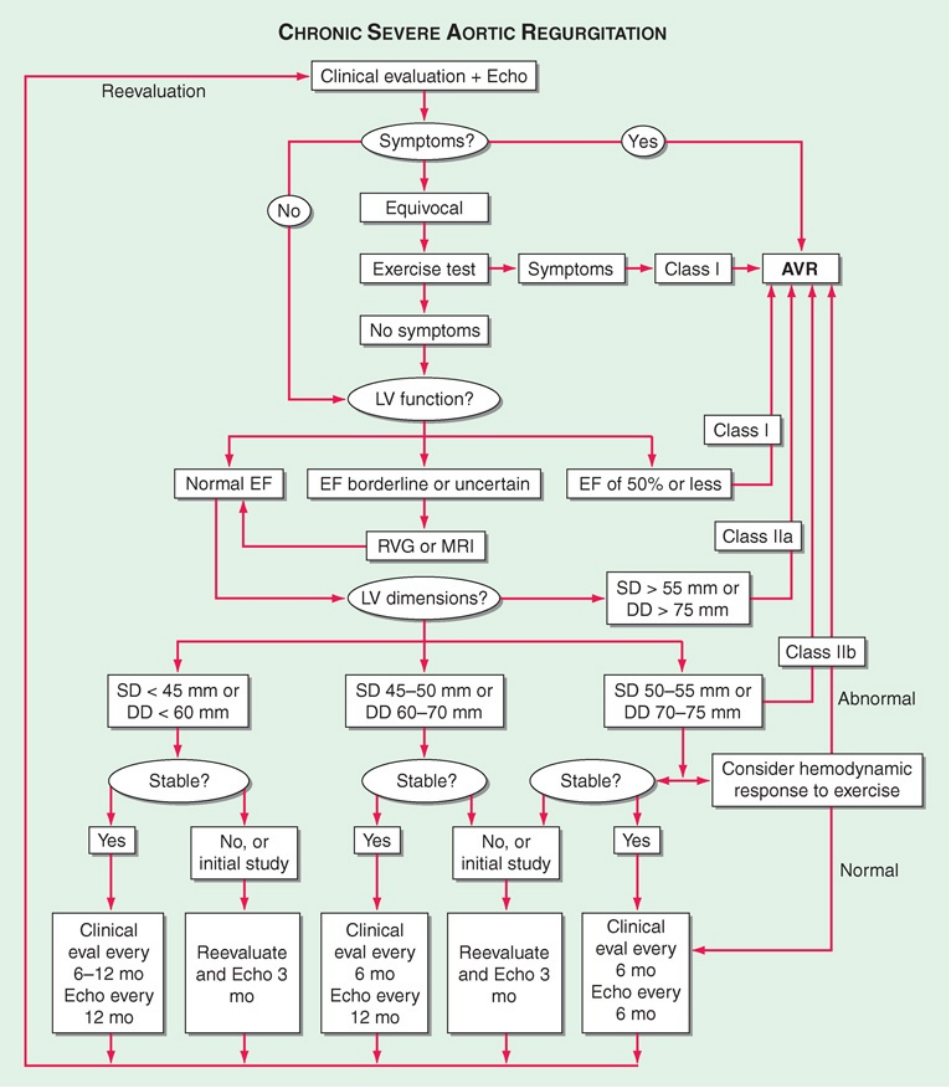

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management Strategy for Patients with Chronic Severe Aortic Regurgitation. Preoperative coronary angiography should be performed routinely, as determined by age, symptoms, and coronary risk factors. Cardiac catheterization and angiography may also be helpful when there is a discrepancy between clinical and echocardiographic findings. "Stable" refers to stable echocardiographic measurements. In some centers, serial follow-up may be performed with radionuclide ventriculography (RVG) or magnetic resonance imaging (MRI) rather than echocardiography (echo) to assess left ventricular (LV) volume and systolic function. AVR, aortic valve replacement; DD, end-diastolic dimension; EF, ejection fraction; eval, evaluation; SD, end-systolic dimension. *(Modified from Bonow et al.)*

CHRONIC AORTIC REGURGITATION

Early symptoms of dyspnea and effort intolerance respond to treatment with diuretics; vasodilators (ACE inhibitors, dihydropyridine calcium channel blockers, or hydralazine) may be useful as well. Surgery can then be performed in a more controlled setting. The use of vasodilators to extend the compensated phase of chronic severe AR before the onset of symptoms or the development of LV dysfunction is more controversial. Expert consensus is strong regarding the need to control systolic blood pressure (goal <140 mmHg) in patients with chronic AR, and vasodilators are an excellent first choice as antihypertensive agents. It is often difficult to achieve adequate control because of the increased stroke volume that accompanies severe AR. Cardiac arrhythmias and systemic infections are poorly tolerated in patients with severe AR and must be treated promptly and vigorously. Although nitroglycerin and long-acting nitrates are not as helpful in relieving anginal pain as they are in patients with ischemic heart disease, they are worth a trial. Patients with syphilitic aortitis should receive a full course of penicillin therapy (Chap. 169). Beta blockers and the angiotensin receptor blocker, losartan, may be useful to retard the rate of aortic root enlargement in young patients with Marfan's syndrome and aortic root dilation. Early reports of the efficacy of losartan in patients with Marfan's syndrome have led to its use in other populations of patients, including those with BAV disease and aortopathy. The use of beta blockers in patients with valvular AR was previously felt to be relatively contraindicated due to concerns that the resulting slowing of the heart rate would allow more time for diastolic regurgitation. More recent observational reports, however, suggest that beta blockers may provide functional benefit in patients with chronic AR. Beta blockers can sometimes provide incremental blood pressure lowering in patients with chronic AR and hypertension. Patients with severe AR, particularly those with an associated aortopathy, should avoid isometric exercises.

SURGICAL TREATMENT

In deciding on the advisability and proper timing of surgical treatment, two points should be kept in mind: (1) patients with chronic severe AR usually do not become symptomatic until *after* the development of myocardial dysfunction; and (2) when delayed too long (defined as >1 year from onset of symptoms or LV dysfunction), surgical treatment often does not restore normal LV function. Therefore, in patients with chronic severe AR, careful clinical follow-up and noninvasive testing with echocardiography at approximately 6-month intervals are necessary if operation is to be undertaken at the optimal time, i.e., *after* the onset of LV dysfunction but *prior* to the development of severe symptoms. Operation can be deferred as long as the patient both remains asymptomatic and retains normal LV function without severe chamber dilation (end-diastolic dimension >75

mm).

AVR is indicated for the treatment of severe AR in symptomatic patients irrespective of LV function. In general, the operation should be carried out in asymptomatic patients with severe AR and progressive LV dysfunction defined by an LVEF <50%, an LV end-systolic dimension >55 mm or end-systolic volume >55 mL/m^2, or an LV diastolic dimension >75 mm. Smaller dimensions may be appropriate thresholds in individuals of smaller stature. Patients with severe AR without indications for operation should be followed by clinical and echocardiographic examination every 3–12 months.

Surgical options for management of aortic valve and root disease have expanded considerably over the past decade. AVR with a suitable mechanical or tissue prosthesis is generally necessary in patients with rheumatic AR and in many patients with other forms of regurgitation. Rarely, when a leaflet has been perforated during infective endocarditis or torn from its attachments to the aortic annulus by thoracic trauma, primary surgical repair may be possible. When AR is due to aneurysmal dilation of the root, or proximal ascending aorta, rather than to primary valve involvement, it may be possible to reduce or eliminate the regurgitation by narrowing the annulus or by excising a portion of the aortic root without replacing the valve. Elective, valve-sparing aortic root reconstruction generally involves reimplantation of the valve in a contoured graft with reattachment of the coronary artery buttons into the side of the graft and is best undertaken in specialized surgical centers (Fig. 237-11). Resuspension of the native aortic valve leaflets is possible in approximately 50% of patients with acute AR in the setting of type A aortic dissection. In other conditions, however, regurgitation can be effectively eliminated only by replacing the aortic valve, the dilated or aneurysmal ascending aorta responsible for the regurgitation, and implanting a composite valve-graft conduit. This formidable procedure entails a higher risk than isolated AVR.

Figure 237-11

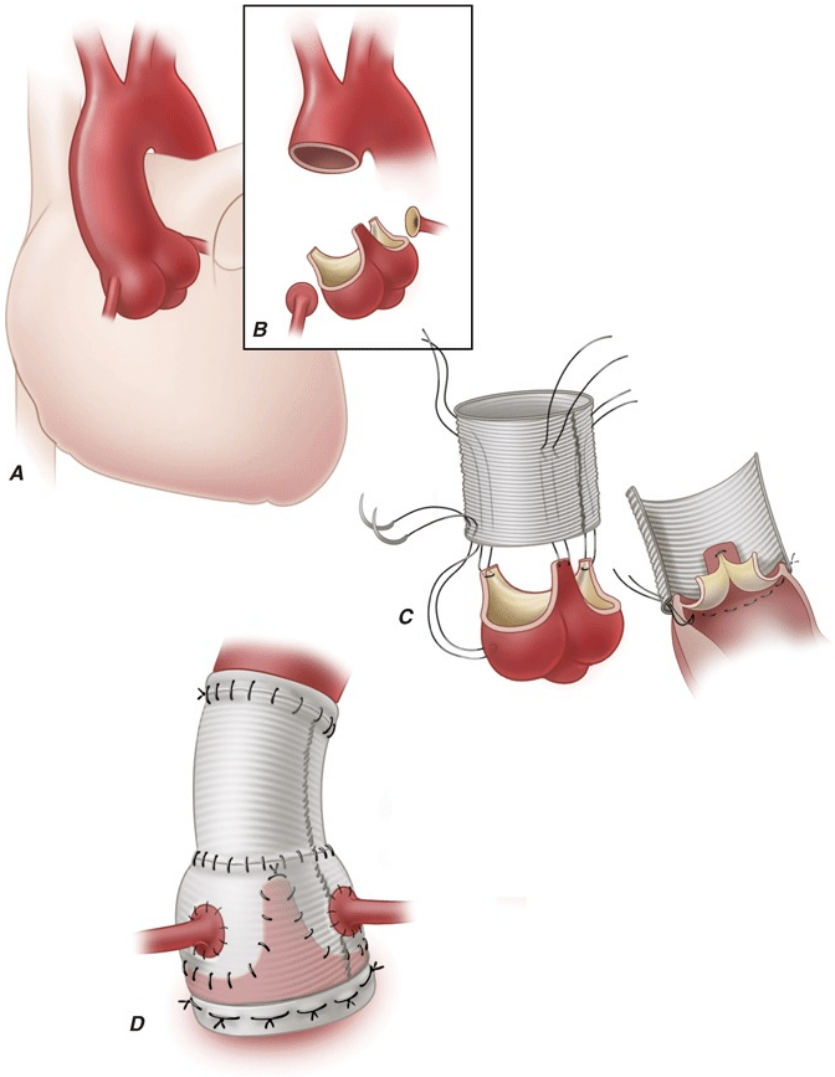

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Valve-Sparing Aortic Root Reconstruction (David Procedure). *(From P Steltzer et al (eds): Valvular Heart Disease: A Companion to Braunwald's Heart Disease, 3rd ed, Fig 12-27, p. 200.)*

As in patients with other valvular abnormalities, both the operative risk and the late mortality rate are largely dependent on the stage of the disease and myocardial function at the time of operation. The overall operative mortality rate for isolated AVR is approximately 3% (Table 237-3). However, patients with marked cardiac enlargement and prolonged LV dysfunction experience an operative mortality rate of approximately 10% and a late mortality rate of approximately 5% per year due to LV failure despite a technically satisfactory operation. Nonetheless, because of the very poor prognosis with medical management, even patients with LV systolic failure should be considered for operation.

Patients with acute severe AR require prompt surgical treatment, which may be lifesaving.

TRICUSPID STENOSIS

TS, which is much less prevalent than MS in North America and Western Europe, is generally rheumatic in origin and more common in women than men. It does not occur as an isolated lesion and is usually associated with MS. Hemodynamically significant TS occurs in 5–10% of patients with severe MS; rheumatic TS is commonly associated with some degree of TR. Nonrheumatic causes of TS are rare.

Pathophysiology

A diastolic pressure gradient between the RA and RV defines TS. It is augmented when the transvalvular blood flow increases during inspiration and declines during expiration. A mean diastolic pressure gradient of 4 mmHg is usually sufficient to elevate the mean RA pressure to levels that result in systemic venous congestion. Unless sodium intake has been restricted and diuretics administered, this venous congestion is associated with hepatomegaly, ascites, and edema, sometimes severe. In patients with sinus rhythm, the RA *a* wave may be extremely tall and may even approach the level of the RV systolic pressure. The *y* descent is prolonged. The CO at rest is usually depressed, and it fails to rise during exercise. The low CO is responsible for the normal or only slightly elevated LA, PA, and RV systolic pressures despite the presence of MS. Thus, the presence of TS can mask the hemodynamic and clinical features of any associated MS.

Symptoms

Because the development of MS generally precedes that of TS, many patients initially have symptoms of pulmonary congestion and fatigue. Characteristically, patients with severe TS complain of relatively little dyspnea for the degree of hepatomegaly, ascites, and edema that they have. However, fatigue secondary to a low CO and discomfort due to refractory edema, ascites, and marked hepatomegaly are common in patients with TS and/or TR. In some patients, TS may be suspected for the first time when symptoms of right-sided failure persist after an adequate mitral valvotomy.

Physical Findings

Because TS usually occurs in the presence of other obvious valvular disease, the diagnosis may be missed unless it is considered. Severe TS is associated with marked hepatic congestion, often resulting in cirrhosis, jaundice, serious malnutrition, anasarca, and ascites. Congestive hepatomegaly and, in cases of severe tricuspid valve disease, splenomegaly are present. The jugular veins are distended, and in patients with sinus rhythm there may be giant *a* waves. The *v* waves are less conspicuous, and because tricuspid obstruction impedes RA emptying during diastole, there is a slow *y* descent. In patients with sinus rhythm there may be prominent presystolic pulsations of the enlarged liver as well.

On auscultation, an OS of the tricuspid valve may rarely be heard approximately 0.06 s after pulmonic valve closure. The diastolic murmur of TS has many of the qualities of the diastolic murmur of MS, and because TS almost always occurs in the presence of MS, it may be missed. However, the tricuspid murmur is generally heard best along the left lower sternal border and over the xiphoid process, and is most prominent during presystole in patients with sinus rhythm. The murmur of TS is augmented during inspiration, and it is reduced during expiration and particularly during the strain phase of the Valsalva maneuver, when tricuspid transvalvular flow is reduced.

Laboratory Examination

The ECG features of RA enlargement (see Fig. 228-8) include tall, peaked P waves in lead II, as well as prominent, upright P waves in lead V_1. The *absence* of ECG evidence of right ventricular hypertrophy (RVH) in a patient with right-sided heart failure who is believed to have MS should suggest associated tricuspid valve disease. The chest x-ray in patients with combined TS and MS shows particular prominence of the RA and superior vena cava without much enlargement of the PA and with less evidence of pulmonary vascular congestion than occurs in patients with isolated MS. On echocardiographic examination, the tricuspid valve is usually thickened and domes in diastole; the transvalvular gradient can be estimated routinely by continuous wave Doppler echocardiography. TTE provides additional information regarding mitral valve structure and function, LV and RV size and function, and PA pressure.

Treatment: Tricuspid Stenosis

Patients with TS generally exhibit marked systemic venous congestion; intensive salt restriction, bed rest, and diuretic therapy are required during the preoperative period. Such a preparatory period may diminish hepatic congestion and thereby improve hepatic function sufficiently so that the risks of operation, particularly bleeding, are diminished. Surgical relief of the TS should be carried out, preferably at the time of surgical mitral valvotomy or MVR, in patients with moderate or severe TS who have mean diastolic pressure gradients exceeding ~4 mmHg and tricuspid orifice areas <1.5–2 cm^2. TS is almost always accompanied by significant TR. Operative repair may permit substantial improvement of tricuspid valve function. If repair cannot be accomplished, the tricuspid valve may have to be replaced with a prosthesis, preferably a large bioprosthetic valve. Mechanical valves in the tricuspid position are more prone to thromboembolic complications than in other positions.

TRICUSPID REGURGITATION

Most commonly, TR is secondary to marked dilation of the tricuspid annulus from RV enlargement due to PA hypertension. Functional TR may complicate RV enlargement of any cause, including an inferior MI that involves the RV. It is commonly seen in the late stages of heart failure due to rheumatic or congenital heart disease with severe PA hypertension (pulmonary artery systolic pressure >55 mmHg), as well as in ischemic and idiopathic dilated cardiomyopathies. It is reversible in part if PA hypertension can be relieved. Rheumatic fever may produce organic (primary) TR, often associated with TS. Infarction of RV papillary muscles, tricuspid valve prolapse, carcinoid heart disease, endomyocardial fibrosis, radiation, infective endocarditis, and trauma all may produce TR. Less commonly, TR results from congenitally deformed tricuspid valves, and it occurs with defects of the atrioventricular canal, as well as with Ebstein's malformation of the tricuspid valve (Chap. 236). TR also develops eventually in patients with chronic RV apical pacing.

As is the case for TS, the clinical features of TR result primarily from systemic venous congestion and reduction of CO. With the onset of TR in patients with PA hypertension, symptoms of pulmonary congestion diminish, but the clinical manifestations of right-sided heart failure become intensified. The neck veins are distended with prominent *v* waves and rapid *y* descents, marked hepatomegaly, ascites, pleural effusions, edema, systolic pulsations of the liver, and a positive hepatojugular reflex. A prominent RV pulsation along the left parasternal region and a blowing holosystolic murmur along the lower left sternal margin, which may be intensified during inspiration and reduced during expiration or the strain of the Valsalva maneuver, are characteristic findings; AF is usually present.

The ECG may show changes characteristic of the lesion responsible for the enlargement of the RV that leads to TR, e.g., an inferior Q-wave MI or RVH. Echocardiography may be helpful by demonstrating RV dilation and prolapsing, flail, scarred, or displaced tricuspid leaflets; the diagnosis and assessment of TR can be made by color flow Doppler imaging (see Fig. 229-8). Severe TR is accompanied by hepatic vein systolic flow reversal. Continuous wave Doppler of the TR velocity profile is useful in estimating PA systolic pressure. Roentgenographic examination usually reveals enlargement of both the RA and RV.

In patients with severe TR, the CO is usually markedly reduced, and the RA pressure pulse may exhibit no *x* descent during early systole but a prominent *c-v* wave with a rapid *y* descent. The mean RA and the RV end-diastolic pressures are often elevated.

Treatment: Tricuspid Regurgitation

Isolated TR, in the absence of PA hypertension, such as that occurring as a consequence of infective endocarditis or trauma, is usually well tolerated and does not require operation. Indeed, even total excision of an infected tricuspid valve may be well tolerated for several years if the PA pressure and resistance are normal. Treatment of the underlying cause of left heart failure usually reduces the severity of functional TR, by reducing the size of the tricuspid annulus. In patients with mitral valve disease and TR secondary to PA hypertension and RV enlargement, effective surgical correction of the mitral valve abnormality results in lowering of the PA pressures and a gradual reduction or disappearance of the TR without direct treatment of the tricuspid valve. However, recovery may be more rapid in patients with severe functional TR if, at the time of mitral valve surgery and especially when there is enlargement of the tricuspid valve annulus, tricuspid valve annuloplasty (generally with the insertion of a ring) or in the rare instance of severe organic tricuspid valve disease, tricuspid valve replacement, is performed (Table 237-3). Tricuspid valve annuloplasty or replacement may rarely be required for severe, primary TR.

PULMONIC VALVE DISEASE

The pulmonic valve is affected by rheumatic fever far less frequently than are the other valves, and it is uncommonly a nidus for infective endocarditis. The most common *acquired* abnormality affecting the pulmonic valve is regurgitation secondary to dilation of the pulmonic valve ring as a consequence of severe PA hypertension. This dilation produces the *Graham Steell murmur,* a high-pitched, decrescendo, diastolic blowing murmur along the left sternal border, which is difficult to differentiate from the far more common murmur produced by AR. Pulmonic regurgitation is usually of little hemodynamic significance; indeed, surgical removal or destruction of the pulmonic valve by infective endocarditis does not produce heart failure unless significant PA hypertension is also present.

The *carcinoid syndrome* may cause pulmonic stenosis and/or regurgitation. Pulmonic regurgitation occurs universally among patients who have undergone childhood repair of tetralogy of Fallot with reconstruction of the RV outflow tract. Congenital pulmonic stenosis is discussed in Chap. 236.

Percutaneous pulmonic valve replacement has been successfully performed in many patients with severe PR after childhood repair of tetralogy of Fallot or pulmonic valve stenosis or atresia. This procedure was introduced clinically prior to percutaneous aortic valve replacement.

MULTIPLE AND MIXED VALVULAR HEART DISEASE

Many acquired and congenital cardiac lesions may result in stenosis and/or regurgitation of one or more heart valves. For example, rheumatic fever may present with mitral (MS, MR, MS and MR), aortic (AS, AR, AS and AR), and/or tricuspid (TS, TR, TS and TR) valve involvement. The common association of secondary TR with significant mitral valve disease has been discussed previously. Aortic valve infective endocarditis may secondarily involve the mitral apparatus. Mediastinal radiation may result in aortic, mitral, and even tricuspid valve disease, most often with mixed stenosis and regurgitation. Ergotamines, and the previously used combination of fenfluramine and phentermine, may result in mixed valve lesions, as are also seen in patients with carcinoid heart disease. Patients with Marfan's syndrome may have both AR from root dilation and MR due to MVP.

The clinical assessment of patients with multiple or mixed valve disease can be challenging, With combined mitral and aortic valve disease, the hemodynamic derangements associated with the mitral lesion(s) may mask the full expression of the aortic valve disease. For example, severe mitral stenosis may decrease LV preload to the extent that the severity of AS or AR can be underappreciated. Alternatively, the development of AF during the course of MS can lead to sudden worsening in a patient whose aortic valve disease was not previously felt to be significant. In patients with mixed MS/MR or AS/AR, the regurgitant lesion usually predominates, and can be followed primarily as the trigger for surgical or transcatheter intervention as appropriate. Nevertheless, there are frequent exceptions to the rules, and careful assessment is indicated. The LV in a patient with hypertrophy in the setting of both AS and AR may not be able to dilate in response to the volume load imposed by the regurgitant lesion and, thus, show signs or lead to the development of symptoms of compromise at a relatively earlier point in the natural history. Serial echocardiography can aid in decision-making.

Similar considerations pertain to the evaluation of patients with heart valve disease and other cardiac and systemic disorders that can contribute to impaired exercise tolerance, such as hypertension, coronary artery disease, obesity, sleep apnea, and skeletal muscle deconditioning.

VALVE REPLACEMENT AND REPAIR

The results of valve replacement surgery are dependent on (1) the patient's myocardial function and general medical condition at the time of operation; (2) the technical abilities of the operative team and the quality of the postoperative care; and (3) the durability, hemodynamic characteristics, and thrombogenicity of the prosthesis. Increased perioperative mortality rate is associated with advanced age and comorbidity rate (e.g., pulmonary or renal disease, the need for nonvalvular cardiovascular surgery, diabetes mellitus) as well as with greater levels of preoperative functional disability and PA hypertension. Late complications of valve replacement include thromboemboli, bleeding due to anticoagulants, mechanical valve thrombosis, pannus ingrowth, paravalvular leak, hemolysis, structural deterioration, infective endocarditis, and prosthesis-patient mismatch (functional prosthetic valve stenosis that occurs when the prosthesis is too small relative to the patient's anatomy).

The choice between a tissue and mechanical prosthesis is essentially a trade-off between the risk of structural valve deterioration and the possible need for reoperation with a tissue valve versus the obligate need for lifelong anticoagulation and its attendant risks with a mechanical valve. The incidence of structural valve deterioration varies inversely with patient age and may be accelerated further by pregnancy or end-stage renal disease. Failure of a tissue prosthesis results in the need for reoperation in up to 30% of patients by 10 years and in 50% by 15 years. Rates of structural valve deterioration are higher for mitral than for aortic bioprostheses. This phenomenon may be due in part to the greater closing pressure to which a mitral prosthesis is exposed.

Traditionally, a mechanical prosthesis was considered preferable for a patient younger than age 65 who could take anticoagulants reliably. Bioprostheses were recommended for older patients (>65 years) who did not otherwise have an indication for anticoagulation (e.g., AF). Recent surveys of cardiac surgery in the United States, as reflected in the Society of Thoracic Surgeons database, show a clear and progressive trend favoring the implantation of bioprosthetic valves in younger (<65 years) patients. Reasons for this development include improved durability of newer generation bioprostheses, decreased risk of death or major complication at time of reoperation, the hazards of long-term anticoagulation, and patient preference to avoid anticoagulation for lifestyle reasons. Patient preference must be factored into any decision regarding the type of valve replacement. A mechanical prosthesis is reasonable for aortic or mitral valve replacement in patients younger than 65 years without a contraindication to anticoagulation. A bioprosthesis is equally reasonable for aortic or mitral valve replacement in patients younger than 65 years who elect this strategy for lifestyle reasons with full knowledge of the likely need for reoperation over time.

Bioprostheses remain the preferred valve choice for patients older than 65 years, in both the aortic and mitral position. Bioprosthetic valves are also indicated for women who expect to become pregnant, as well as for others who refuse to take anticoagulation or for whom anticoagulation may be contraindicated. Types of bioprostheses include xenografts (e.g., porcine

aortic valves; cryopreserved, mounted bovine pericardial valves), homograft (allograft) aortic valves obtained from cadavers, and pulmonary autografts transplanted into the aortic position (Ross procedure). Homograft replacement may be preferred for the management of complicated aortic valve infective endocarditis.

In patients without contraindications to anticoagulants, particularly those younger than 65 years, a mechanical prosthesis is reasonable. Many surgeons now select the St. Jude prosthesis, a double-disk tilting prosthesis, for replacement of both aortic and mitral valves because of favorable hemodynamic characteristics and possible association with lower thrombogenicity. A tissue valve is preferred for tricuspid replacement.

The decision to proceed with valve replacement should be finalized only after an experienced surgeon has agreed that valve repair is either not appropriate or not feasible. As noted previously, valve repair techniques have improved considerably over the past 10–15 years, for both mitral and aortic lesions. Primary repair is often associated with a lower risk of postoperative LV functional impairment, particularly for patients with MR, and avoids the long-term risks of a prosthesis.

Antibiotic prophylaxis prior to dental procedures that involve manipulation of gingival tissue or the periapical region of teeth or perforation of the oral mucosa is indicated for patients after valve replacement or ring annuloplasty. Prosthetic valve and ventricular function should be assessed with transthoracic echocardiography 3 months after surgery and a baseline complete blood count, reticulocyte count, and serum lactate dehydrogenase (LDH) test obtained to serve as a baseline set of values should the question of hemolysis arise in the future. The intensity of anticoagulation should follow recommended guidelines.

GLOBAL BURDEN OF VALVULAR HEART DISEASE

Primary valvular heart disease ranks well below coronary heart disease, stroke, hypertension, obesity, and diabetes as major threats to the public health. Nevertheless, it is the source of significant morbidity and mortality rates. Rheumatic fever (Chap. 322) is the dominant cause of valvular heart disease in developing countries. Its prevalence has been estimated to range from as low as 1 per 100,000 school-age children in Costa Rica to as high as 150 per 100,000 in China. Rheumatic heart disease accounts for 12–65% of hospital admissions related to cardiovascular disease and 2–10% of hospital discharges in some developing countries. Prevalence and mortality rates vary among communities even within the same country as a function of crowding and the availability of medical resources and population-wide programs for detection and treatment of group A streptococcal pharyngitis. In economically deprived areas, tropical and subtropical climates (particularly on the Indian subcontinent), Central America, and the Middle East, rheumatic valvular disease progresses more rapidly than in more-developed nations and frequently causes serious symptoms in patients younger than 20 years of age. This accelerated natural history may be due to repeated infections with more virulent strains of rheumatogenic streptococci. Approximately 16 million people live with rheumatic heart disease worldwide. As of 2000, worldwide death rates for rheumatic heart disease approximated 5.5 per 100,000 population (n = 332,000), with the highest rates reported from Southeast Asia (7.6 per 100,000).

Although there have been recent reports of isolated outbreaks of streptococcal infection in North America, valve disease in developed countries is now dominated by degenerative or inflammatory processes that lead to valve thickening, calcification, and dysfunction. The prevalence of valvular heart disease increases with age. Important left-sided valve disease may affect as many as 12–13% of adults older than the age of 75. In the United States, there were 1.5 million hospital discharges with any diagnosis of valvular heart disease in 2005, and 94,000 of these were related to surgical procedures for heart valve disease (mostly involving the aortic and mitral valves).

The incidence of infective endocarditis (Chap. 124) has increased with the aging of the population, the more widespread prevalence of vascular grafts and intracardiac devices, the emergence of more virulent multidrug-resistant microorganisms, and the growing epidemic of diabetes. Infective endocarditis has become a more frequent cause of acute valvular regurgitation. Bicuspid aortic valve disease affects as many as 0.5–1.4% of the population, and an increasing number of childhood survivors of congenital heart disease present later in life with valvular dysfunction. The global burden of valvular heart disease is expected to progress.

As is true for other health conditions, disparities in access to and quality of care for patients with valvular heart disease have been well documented. Management decisions and outcome differences based on age, gender, and race require educational efforts across all levels of providers.

FURTHER READINGS

Bonow RO et al: ACC/AHA 2006 Guidelines for the management of patients with valvular heart disease: Executive summary. Circulation 114:450, 2006

Carabello BA, Paulus WJ: Aortic stenosis. Lancet 373:956, 2009[PMID: 19232707]

Coats L, Bonhoeffer P: New percutaneous treatments for valve disease. Heart 93:639, 2007[PMID: 17435075]

Enriquez-Sarano M, Tajik AJ: Aortic regurgitation. N Engl J Med 351:1539, 2004[PMID: 15470217]

Feldman T, Leon MB: Prospects for percutaneous valve therapies. Circulation 116:2866, 2007[PMID: 18071090]

Freeman RV, Otto CM. Spectrum of calcific aortic valve disease: Pathogenesis, disease progression, and treatment strategies. Circulation 111:3316, 2005[PMID: 15967862]

Grube E et al: Progress and current status of percutaneous aortic valve replacement: Results of three device generations of the CoreValve Revalving system. Circ Cardiovasc Interven 1:167, 2008[PMID: 20031675]

Nkomo VT et al: Burden of valvular heart diseases: A population-based study. Lancet 368:1005, 2006[PMID: 16980116]

Otto CM, Bonow RO (eds): Valvular Heart Disease. A Companion to Braunwald's Heart Disease, 3rd ed. Philadelphia, Saunders Elsevier, 2009

———: Valvular heart disease, in Braunwald's Heart Disease, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

Pellikka PA et al: Outcome of 622 adults with asymptomatic, hemodynamically significant aortic stenosis during prolonged follow-up. Circulation 111:3290, 2005[PMID: 15956131]

Rosenhek R et al: Outcome of watchful waiting in asymptomatic severe mitral regurgitation. Circulation 113:2238, 2006[PMID: 16651470]

Society of Thoracic Surgeons: STS Adult Cardiovascular National Surgery Database. Period Ending 03/31/2009 Executive Summary. http://www.sts.org/documents/pdf/ndb/2ndHarvestExecutiveSummary_2009.pdf

———: STS national database risk calculator. http://www.sts.org/sections/stsnationaldatabase/riskcalculator/

Tadros TM et al: Ascending aortic dilatation associated with bicuspid aortic valve: pathophysiology, molecular biology, and clinical implications. Circulation 119:880, 2009[PMID: 19221231]

Webb JG et al: Percutaneous transarterial aortic valve replacement in selected high risk patients with aortic stenosis. Circulation 116:755, 2007[PMID: 17646579]

Wilson W et al: Prevention of infective endocarditis. Guidelines from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee, Council on Cardiovascular Disease in the Young, and the Council on Clinical Cardiology, Council on Cardiovascular Surgery and Anesthesia and the Quality of Care and Outcomes Research Interdisciplinary Working Group. April 2007, http://circ.ahajournals.org

World Health Organization: Rheumatic fever and rheumatic heart disease. World Health Organ Tech Rep Ser 923:1, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION AND CLASSIFICATION

Cardiomyopathy is disease of the heart muscle. It is estimated that cardiomyopathy accounts for 5–10% of the 5–6 million patients already diagnosed with heart failure in the United States. This term is intended to exclude cardiac dysfunction that results from other structural heart disease, such as coronary artery disease, primary valve disease, or severe hypertension; however, in general usage the phrase *ischemic cardiomyopathy* is sometimes applied to describe diffuse dysfunction occurring in the presence of multivessel coronary artery disease, and *nonischemic cardiomyopathy* to describe cardiomyopathy from other causes. As of 2006, cardiomyopathies are defined as "a heterogeneous group of diseases of the myocardium associated with mechanical and/or electrical dysfunction that usually (but not invariably) exhibit inappropriate ventricular hypertrophy or dilatation and are due to a variety of causes that frequently are genetic."

The traditional classification of cardiomyopathies into a triad of dilated, restrictive, and hypertrophic was based initially on autopsy specimens and later on echocardiographic findings. Dilated and hypertrophic cardiomyopathies can be distinguished on the basis of left ventricular wall thickness and cavity dimension; however, restrictive cardiomyopathy can have variably increased wall thickness and chamber dimensions that range from reduced to slightly increased, with prominent atrial enlargement. Restrictive cardiomyopathy is now defined more on the basis of abnormal diastolic function, which is also present but initially less prominent in dilated and hypertrophic cardiomyopathy. Restrictive cardiomyopathy can overlap in presentation, gross morphology, and etiology with both hypertrophic and dilated cardiomyopathies (**Table 238-1**).

Table 238-1 Presentation with Symptomatic Cardiomyopathy

	Dilated	Restrictive	Hypertrophic
Ejection fraction (normal ≥55%)	Usually <30% when symptoms severe	25-50%	>60%
Left ventricular diastolic dimension (normal <55 mm)	≥60 mm	>60 mm (may be decreased)	Often decreased
Left ventricular wall thickness	Decreased	Normal or increased	Markedly increased
Atrial size	Increased	Increased; may be massive	Increased; related to abnormal
Valvular regurgitation	Related to annular dilation; mitral appears earlier, during decompensation; tricuspid regurgitation in late stages	Related to endocardial involvement; frequent mitral and tricuspid regurgitation, rarely severe	Related to valve-septum interaction; mitral regurgitation
Common first symptoms	Exertional intolerance	Exertional intolerance, fluid retention early	Exertional intolerance; may have chest pain
Congestive symptoms*	Left before right, except right prominent in young adults	Right often dominates	Left-sided congestion may develop late
Arrhythmia	Ventricular tachyarrhythmia; conduction block in Chagas' disease, and some families. Atrial fibrillation.	Ventricular uncommon except in sarcoidosis conduction block in sarcoidosis and amyloidosis. Atrial fibrillation.	Ventricular tachyarrhythmias; atrial fibrillation

*Left-sided symptoms of pulmonary congestion; dyspnea on exertion, orthopnea, paroxysmal nocturnal dyspnea. Right-sided symptoms of systemic versus congestion: discomfort on bending, hepatic and abdominal distention, peripheral edema.

Expanding information renders this classification triad based on phenotype increasingly inadequate to define disease or therapy. Identification of more genetic determinants of cardiomyopathy has suggested a four-way classification scheme of etiology as primary (affecting primarily the heart) and secondary to other systemic disease. The primary causes are then divided into genetic, mixed genetic and acquired, and acquired; however, in current practice the genetic information is often unavailable at the time of initial presentation, particularly in the absence of extracardiac manifestations. Many mutated genes can be associated with the same general phenotype, and one defective gene may manifest as multiple phenotypes. In addition, the bases of evidence for most therapies are still driven by clinical phenotypes. Although the proposed genetic classification does not yet guide many current clinical strategies, it will become increasingly relevant as classification of disease moves beyond individual organ pathology to more integrated systems approaches.

GENERAL PRESENTATION

For all cardiomyopathies, the early symptoms often relate to exertional intolerance with breathlessness or fatigue, usually from inadequate cardiac reserve during exercise. These symptoms may initially go unnoticed or be attributed to other causes, commonly pulmonary. As fluid retention leads to elevation of resting filling pressures, shortness of breath may occur during routine daily activity, such as dressing, and may manifest as dyspnea or cough in the supine position. Although often considered the hallmark of congestion, peripheral edema may not appear despite severe fluid retention, particularly in younger patients. The nonspecific term *congestive heart failure* describes only the resulting syndrome of fluid retention, which is common to the three types of cardiomyopathy and also to other cardiac diseases associated with elevated filling pressures. Despite the different structural basis, all three types of cardiomyopathy can be associated with atrioventricular valve regurgitation, typical and atypical chest pain, atrial and ventricular tachyarrhythmias, and embolic events (Table 238-1). Initial evaluation begins with a detailed clinical history and examination, looking for clues to cardiac, extracardiac, and familial disease (**Table 238-2**). The initial evaluation, prognosis, and therapy are generally defined by the severity of cardiac and clinical dysfunction, with some distinctive features according to etiology.

Table 238-2 Initial Evaluation of Cardiomyopathy

Clinical Evaluation
Thorough history and physical examination to identify cardiac and noncardiac disorders[a]
Detailed family history of heart failure, cardiomyopathy, skeletal myopathy, conduction disorders and tachyarrhythmias, sudden

death
History of alcohol, illicit drugs, chemotherapy or radiation therapy[a] Assessment of ability to perform routine and desired activities[a]
Assessment of volume status, orthostatic blood pressure, body mass index[a]
Laboratory Evaluation
Electrocardiogram[a]
Chest radiograph[a] Two-dimensional and Doppler echocardiogram[a]
Chemistry: Serum sodium, potassium,[a] calcium,[a] magnesium[a] Fasting glucose (glycohemoglobin in DM) Creatinine,[a] blood urea nitrogen[a] Albumin,[a] total protein,[a] liver function tests[a] Lipid profile Thyroid-stimulating hormone[a] Serum iron, transferrin saturation Urinalysis Creatine kinase Hematology: Hemoglobin/hematocrit[a] White blood cell count with differential,[a] including eosinophils Erythrocyte sedimentation rate
Initial Evaluation Only in Patients Selected for Possible Specific Diagnosis
Titers for infection in presence of clinical suspicion: Acute viral (coxsackie virus, echovirus, influenza virus) Human immunodeficiency virus, Chagas' disease, Lyme disease, toxoplasmosis
Catheterization with coronary angiography in patients with angina who are candidates for intervention[a]
Serologies for active rheumatologic disease
Endomyocardial biopsy including sample for electron microscopy when suspecting specific diagnosis with therapeutic implications
Screening for sleep-disordered breathing

[a] Level I Recommendations from ACC/AHA Practice Guidelines for Chronic Heart Failure in the adult.

Source: From Hunt et al.

GENETIC ETIOLOGIES OF CARDIOMYOPATHY

The estimated prevalence of a genetic etiology for cardiomyopathy continues to increase with increasing awareness of the importance of the family history and the availability of genetic testing. Well-recognized in hypertrophic cardiomyopathy, heritability is present in at least 30% of dilated cardiomyopathy without other clear etiology. Careful family history should elicit not only known cardiomyopathy and heart failure, but also family members who have had sudden death, often incorrectly attributed to "a massive heart attack," who have had atrial fibrillation or pacemaker implantation by middle age, or who have muscular dystrophy. The family history should be reviewed at subsequent intervals particularly regarding siblings and cousins, who may tend to manifest disease at similar ages.

Most familial cardiomyopathies are inherited in an autosomal dominant pattern, with occasional autosomal recessive and X-linked inheritance (**Table 238-3**). The penetrance and phenotype of a given mutation varies with other genetic, epigenetic, and environmental determinants. Some mutations are associated with primary conduction system disease as well as dilated cardiomyopathy (CDDC). With rare exceptions, such as the replacements for defective metabolic enzymes, current therapy is based on the phenotype rather than the genetic defect. However, knowledge of the genetic defect may influence prognosis and in some cases provide indication for implantable defibrillators.

Table 238-3 Inherited Genetic Defects Associated With Cardiomyopathy

	Gene Product	Inheritance	Cardiac Phenotype	Isolated Cardiac Phenotype	Extracardiac Manifestations
Sarcomere	*MYH7* (β myosin heavy chain)	AD	HCM, DCM, LVNC	yes	Skeletal myopathy
	MYBPC3 (myosin binding protein C)	AD	HCM, (DCM)	yes	
	TNNT2 (cardiac troponin T)	AD	HCM, DCM,	yes	

			LVNC		
	TNNI3 (cardiac troponin I)	AD, AR	HCM, DCM, RCM	yes	
	TTN (Titin)	AD	HCM, DCM	yes	
	TPM1 (α-tropomyosin)	AD	HCM, DCM	yes	
	TNNC1 (slow troponin C)	AD	DCM	yes	
	ACTC (α-actin)	AD	HCM, DCM, (LVNC)	yes	
	MYL2 (myosin regulatory light chain)	AD	HCM	yes	Skeletal myopathy
	MYL3 (myosin essential light chain)	AD	HCM	yes	
	MYH6 (α-myosin heavy chain)	AD	HCM, (DCM)	yes	
Z-disk and Cytoskeleton	*DES* (Desmin)	AD	DCM	yes	Skeletal myopathy
	LDB3 (Cypher-ZASP)	AD	DCM, LVNC	yes	Skeletal myopathy
	MYOZ2 (Myozenin)	AD	HCM	yes	
	TCAP (Telethonin)	AD	DCM, HCM	yes	
	ANKRD1 (CARP)	AD	HCM, (DCM)	yes	
	CSRP3 (MLP)	AD	DCM, (HCM)	yes	
	OBSCN (Obscurin)	AD	HCM	yes	
	ACTN2 (α-actinin-2)	AD	DCM	yes	
	CRYAB (αB-crystallin)	AD	DCM	yes	
Nuclear Membrane	*LMNA* (Lamin A/C)	AD	CDDC	yes	Skeletal myopathy
	EMD (Emerin)	X-linked	CDDC	no	Skeletal myopathy, contractures
	TMPO (Thymopoietin)	AD	DCM	yes	
Excitation-Contraction Coupling	*PLN* (Phospholamban)	AD	DCM	yes	
	SCN5A (NAV 1.5)	AD	CDDC	yes	
	RYR2 (cardiac ryanodine receptor)	AD	ARVC	yes	
Cellular Metabolism	*PRKAG2* (γ-subunit of AMP Kinase)	AD	HCM+	yes	
	LAMP2 (lysosomal associated membrane protein)	X-linked	HCM+	no*	Danon's disease: skeletal myopathy, cognitive impairment
	TAZ (Tafazzin)	X-linked	DCM, LVNC	no	Barth's syndrome: skeletal myopathy, cognitive impairment, neutropenia
	FXN (Frataxin)	AR	HCM	no	Friedreich's ataxia: ataxia, diabetes mellitus type 2
	ABCC9 (sulfonylurea receptor 2)	AD	DCM	yes	
	TMEM43 (transmembrane protein 43)	AD	ARVC	yes	
	GLA (α-galactosidase-A) (other systemic metabolic defects, see Table 238-4)	X-linked	HCM	yes	Fabry's disease: renal failure, angiokeratomas and painful neuropathy
Mitochondria	mitochondrial DNA	Maternal transmission	DCM, HCMc	no	MELAS, MERRF, Kearns-Sayre syndrome, ocular myopathy
Sarcolemmal Membrane	*DMD* (Dystrophin)	X-linked	DCM	no*	Duchenne's and Becker's muscular dystrophy
	DMPK (dystrophica myotonica protein kinase)	AD	DCM	no	Myotonic dystrophy type 1
	SGCD (δ-sarcoglycan)	AD	DCM	yes	
	VCL (Metavinculin)	AD	DCM	yes	
Desmosome	*DSP* (Desmoplakin)	AD, AR	ARVC	yes	Carvajal syndrome (AR)
	DSG2 (Desmoglein 2)	AD	ARVC	yes	
	DSC2 (Desmocollin 2)	AD	ARVC	yes	
	PKP2 (Plakophilin 2)	AD	ARVC	yes	
	JUP (Plakoglobin)	AD, AR	ARVC	yes	Naxos syndrome (AR)
Other	*EYA4* (Eyes Absent 4)	AD	DCM	no	Sensorineural deafness
	RBM20 (RNA binding motif 20)	AD	DCM	yes	
	PSEN1 (Presenilin-1,2)	AD	DCM	yes	Dementia

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ARVC, arrhythmogenic right ventricular cardiomyopathy; CDDC, conduction disease with dilated cardiomyopathy; DCM, dilated cardiomyopathy; HCM+, HCM with preexcitation; HCMc, HCM with conduction disease; LVNC, left ventricular noncompaction; MELAS, (mitochondrial) myopathy, encephalopathy, lactic acidosis, and strokelike episodes syndrome; MERRF, myoclonic epilepsy with ragged red fibers; RCM, restrictive cardiomyopathy.

Note: *indicates that isolated cardiac phenotype can occur in women with the x-linked defects.

Source: From Neal Lakdawala, MD, Cardiovascular Genetics, Brigham and Women's Hospital.

Defects in sarcomeric proteins of myosin, actin, and troponin are the best characterized. While the majority of these are associated

with hypertrophic cardiomyopathy, an increasing number of sarcomeric mutations have now been implicated in dilated cardiomyopathy, and some have also been associated with left ventricular noncompaction. Thus far, few mutations have been identified in excitation-contraction coupling proteins, perhaps because they are too crucial for survival to allow variation.

Many of the proteins encoded by abnormal structural genes span more than one functional area of the myocyte (**Fig. 238-1**). Proteins contributing to the Z-disk organize and stabilize the sarcomeres. Multiple other proteins are involved in connecting and maintaining the cytoskeleton of the myocyte. For example, desmin forms intermediate filaments that connect the nuclear and plasma membranes, Z-lines, and the intercalated disks between muscle cells. Desmin mutations impair the transmission of force and signaling for both cardiac and skeletal muscle, and are, thus, associated with a peripheral myopathy as well as a dilated cardiomyopathy. Most of the identified genetic defects in the Z-disk and cytoskeleton are associated with dilated cardiomyopathy.

Figure 238-1

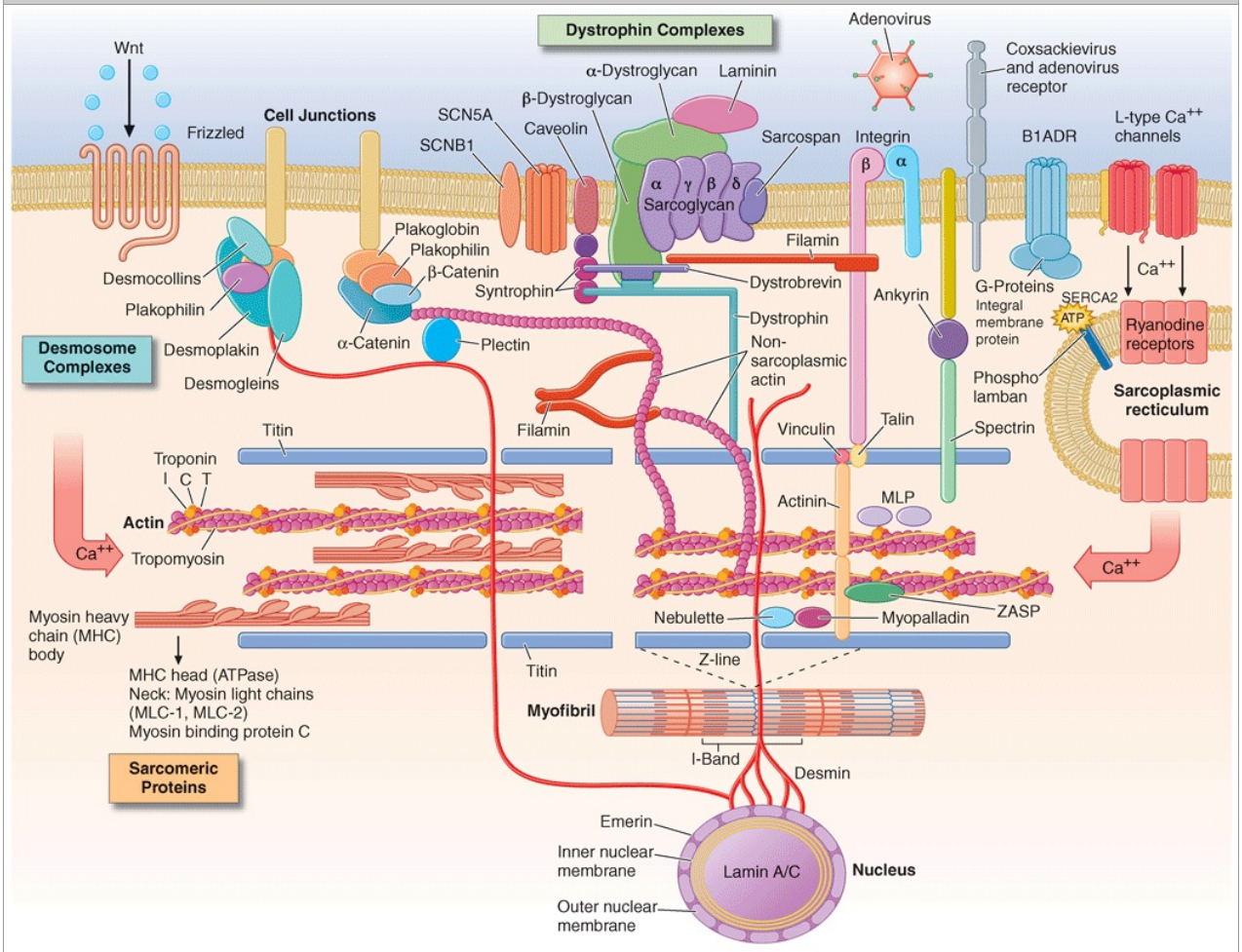

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Drawing of myocyte indicating multiple sites of abnormal gene products associated with cardiomyopathy. Major functional groups include the sarcomeric proteins (actin, myosin, tropomyosin, and the associated regulatory proteins), the dystrophin complex stabilizing and connecting the cell membrane to intracellular structures, the desmosome complexes associated with cell-cell connections and stability, and multiple cytoskeletal proteins that integrate and stabilize the myocyte. ATP, adenosine triphosphate. (*Figure adapted from Jeffrey A. Towbin, MD, University of Cincinnati, with permission.*)

Proteins in the sarcolemmal membrane are associated with dilated cardiomyopathy. The best known is the X-linked dystrophin, abnormalities of which cause Duchenne's and Becker's muscle dystrophy. (Interestingly, abnormal dystrophin can be acquired when the Coxsackie virus cleaves dystrophin during viral myocarditis.) This protein provides a network that supports the sarcolemma and also connects to the sarcomere. The progressive functional defect in both cardiac and skeletal muscle reflects vulnerability to mechanical stress. Dystrophin is associated at the membrane with a complex of other proteins, such as metavinculin, abnormalities of which cause dilated cardiomyopathy with autosomal dominant inheritance. Defects in the sarcolemmal channel proteins (*channelopathies*) are generally associated with primary arrhythmias, but mutations in SCN5A, distinct from those which cause the Brugada or long-QT syndromes, have been implicated in dilated cardiomyopathy.

Nuclear membrane protein defects in the myocyte can also cause skeletal myopathy in either autosomal dominant (lamin proteins) or X-linked (emerin) patterns. These are associated with a high prevalence of atrial arrhythmias and conduction system disease which, in some family members, occur without detectable cardiomyopathy.

Intercalated disks between cardiac myocytes allow mechanical and electrical coupling between cells and also connect to desmin filaments within the cell. Mutations in proteins of the desmosomal complex compromise attachment of the myocytes, which can become disconnected and die, to be replaced by fat and fibrous tissue. These areas are highly arrhythmogenic and may go on to aneurysm formation. Although it is more noticeable in the thinner right ventricle, this condition often affects both ventricles. As desmosomes are also important for elasticity of hair and skin, some defective desmosomal proteins are associated with striking "woolly hair" and thickened skin on the palm and soles.

Owing to the conservation of signaling pathways in multiple systems, we may expect to discover more extracardiac manifestations of genetic abnormalities initially considered to manifest exclusively in the heart. In contrast, the monogenic disorders of metabolism that affect the heart are already clearly recognized to affect multiple organ systems (**Table 238-4**). The most important currently are those defective enzymes for which specific enzyme replacement therapy can now ameliorate the course of disease, as with alpha-

galactosidase-A (Fabry's disease). Abnormalities of mitochondrial DNA (maternally transmitted) impair energy production with multiple clinical manifestations, including impaired cognitive function and skeletal myopathy. The phenotypic expression is highly variable depending on the distribution of the maternal mitochondria during embryonic development. Heritable systemic diseases, such as familial amyloidosis and hemochromatosis, can affect the heart without abnormal expression of specific cardiac genes.

Table 238-4 Examples of Inherited Defects in Metabolic Pathways Associated With Cardiomyopathy, Usually Restrictive or Pseudohypertrophic Phenotype

Glycogen Storage Diseases
II—Pompe's (alpha 1,4 glucosidase)
III—Forbes: de-branching enzyme (amylo 1,6 glucosidase)
Glucose Metabolism (Defective PRKAG2[a])
Fatty acid metabolism
Carnitine transport defect
Medium chain Acyl-CoA dehydrogenase
Long chain Acyl-CoA dehydrogenase
Sphingolipidoses
Fabry's disease (alpha galactosidase A)
Gaucher disease (beta-glucocerebroside)
Disorders of lysosomal function
Danon's disease—(lysosome-associated membrane protein, LAMP2)
Miscellaneous
Hemochromatosis—Fe metabolism
Familial amyloidosis—abnormal transthyretin
Barth syndrome—tafazzin defect affecting cardiolipin
Friedreich's ataxia—frataxin

[a]Gamma-2 regulatory subunit of the AMP-activated protein kinase important for glucose metabolism.

For any patient with suspected or proven genetic disease, family members should be considered and evaluated in a longitudinal fashion. Screening includes an echocardiogram and electrocardiogram (ECG). The indications and implications for confirmatory specific genetic testing vary depending upon the specific mutation. The profound questions raised by families about diseases shared and passed down merit serious and sensitive discussion, ideally provided by a trained genetic counselor.

DILATED CARDIOMYOPATHY

An enlarged left ventricle with decreased systolic function as measured by left ventricular ejection fraction characterizes dilated cardiomyopathy (**Figs. 238-2, 238-3, 238-4**). *Systolic failure* is more marked than the frequently accompanying diastolic dysfunction, although the latter may be functionally severe in the setting of marked volume overload. The syndrome of dilated cardiomyopathy has multiple etiologies (**Table 238-5**). Up to one-third of cases may be familial, as discussed below. Acquired cardiomyopathy is often attributed to a brief primary injury such as infection or toxin exposure. Some myocytes may die during the initial injury, while others survive only to have later programmed cell death, (apoptosis). As the surviving myocytes hypertrophy to accommodate the increased burden of wall stress, local and circulating factors stimulate deleterious responses that contribute to progression of disease, even in the absence of further primary injury. Dynamic remodeling of the interstitial scaffolding affects diastolic function and the amount of ventricular dilation. Mitral regurgitation commonly develops as the valvular apparatus is distorted by ventricular dilation and sometimes by focal injury to underlying myocardium and is usually substantial by the time heart failure is severe. Many cases that present "acutely" have progressed silently through these stages over months to years.

Figure 238-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dilated cardiomyopathy. This gross specimen of a heart removed at the time of transplantation shows massive left ventricular dilation and moderate right ventricular dilation. Although the left ventricular wall in particular appears thinned, there is significant hypertrophy of this heart, which weighs more than 800 gm (upper limit of normal = 360 g). A defibrillator lead is seen traversing the tricuspid valve into the right ventricular apex. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Figure 238-3

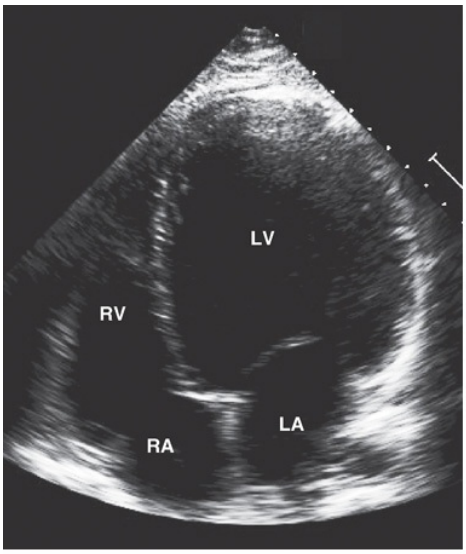

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dilated cardiomyopathy. This echocardiogram of a young man with dilated cardiomyopathy shows massive global dilation and thinning of the walls of the left ventricle (LV). The left atrium (LA) is also enlarged compared to normal. Note that the echocardiographic and pathologic images are vertically opposite, such that the LV is by convention on the top right in the echocardiographic image and bottom right in the pathologic images. (*Image courtesy of Justina Wu, MD, Brigham and Women's Hospital, Boston.*)

Figure 238-4

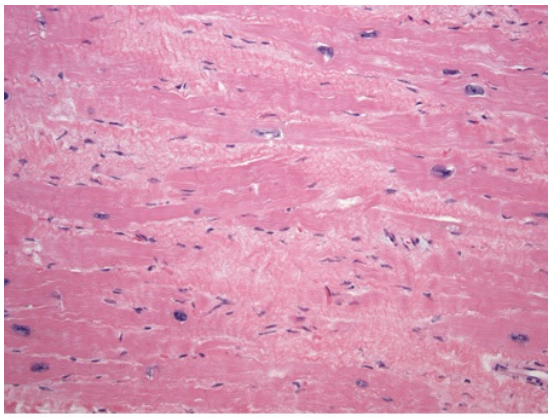

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dilated cardiomyopathy. Microscopic specimen of a dilated cardiomyopathy showing the nonspecific changes of interstitial fibrosis and myocyte hypertrophy characterized by increased myocyte size and enlarged, irregular nuclei. Hematoxylin and eosin stained section, 100 x original magnification. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Table 238–5 Major Causes of Dilated Cardiomyopathy (With Common Examples)

Inflammatory Myocarditis
Infective
Viral (Coxsackie, adenovirus, HIV, hepatitis C)
Parasitic (*T. cruzi*—Chagas' disease, toxoplasmosis)
Bacterial (diphtheria)
Spirochetal (*Borellia burgdorferi*—Lyme disease)
Rickettsial—(Q fever)
Fungal (with systemic infection)
Noninfective

Granulomatous inflammatory disease Sarcoidosis Giant cell myocarditis Hypersensitivity myocarditis Polymyositis, dermatomyositis Collagen vascular disease Peripartum cardiomyopathy Transplant rejection
Toxic
Alcohol Catecholamines: amphetamines, cocaine Chemotherapeutic agents: (anthracyclines, trastuzumab) Interferon Other therapeutic agents (hydroxychloroquine, chloroquine) Drugs of misuse (emetine, anabolic steroids) Heavy metals: lead, mercury Occupational exposure: hydrocarbons, arsenicals
Metabolic[a]
Nutritional deficiencies: thiamine, selenium, carnitine
Electrolyte deficiencies: calcium, phosphate, magnesium
Endocrinopathy: Thyroid disease Pheochromocytoma Diabetes
Obesity
Hemochromatosis
Inherited Metabolic Pathway Defects (See Table 238-4)
Familial[a] (See Table 238-3)
Skeletal and cardiac myopathy
Dystrophin-related dystrophy (Duchenne's, Becker's)
Mitochondrial myopathies (e.g., Kearns-Sayre syndrome)
Arrhythmogenic ventricular dysplasia
Hemochromatosis
Associated with other systemic diseases
Susceptibility to immune-mediated myocarditis
Overlap with Restrictive Cardiomyopathy
"Minimally dilated cardiomyopathy"
Hemochromatosis
Amyloidosis
Hypertrophic cardiomyopathy ("burned-out")
"Idiopathic"[a]
Miscellaneous (Shared Elements of Above Etiologies)
Arrhythmogenic right ventricular dysplasia (may also affect left ventricle)[a] Left ventricular noncompaction[a]
Peripartum cardiomyopathy

[a]Some specific cases can be linked now to specific genetic mutation in a familial cardiomyopathy; others with similar phenotypes that appear to be acquired or idiopathic may represent genetic factors not yet identified.

Regardless of the nature and degree of direct cell injury, the resulting functional impairment often includes some contribution from secondary responses that may be reversible. The potential reversibility of cardiomyopathy in the absence of ongoing injury remains a subject of active controversy. Almost half of all patients with truly recent onset cardiomyopathy demonstrate substantial spontaneous recovery. Some patients have dramatic improvement to near-normal ejection fractions during pharmacologic therapy, particularly notable with the β-adrenergic antagonists coupled with renin-angiotensin system inhibition. Interest in the potential for recovery of cardiomyopathy in the absence of coronary artery disease has been further stimulated by occasional "recovery" of left ventricular function in young patients after a year or more of mechanical circulatory support. The diagnosis and therapy for dilated cardiomyopathy is generally dictated by the stage of heart failure (Chap. 234), with specific aspects discussed with the relevant etiology below.

Infective Myocarditis

Myocarditis is an inflammatory process, most commonly attributed to infectious organisms that can invade the myocardium directly, produce cardiotoxins, and trigger chronic inflammatory responses. Infective myocarditis has been reported with almost all types of infectious agents, but is most commonly associated with viral infections, the protozoan *Trypanosoma cruzi* in South America, and endomyocardial fibrosis in equatorial Africa.

Viral myocarditis in murine models begins with acute infection. After viruses enter the circulation through the respiratory or gastrointestinal tract, they can infect other organs possessing specific receptors, such as the coxsackie-adenovirus receptor on the

heart. Viral invasion and replication can lead directly to myocardial injury and lysis. Viral proteases have multiple actions, of which one is to degrade the protein, dystrophin, in the myocyte membrane complex that is genetically abnormal in some muscular dystrophies. Viral antigens activate immune responses that help to contain the initial infection but may persist into later phases. Components include nonspecific cytokines, specific antibodies, and cytotoxic T-lymphocytes, which in some cases recognize myocyte proteins. There is varying evidence for a latent phase of ongoing infection with persistence of the viral genome and some viral proteins. The relative contributions of viral persistence and deleterious host immune responses to progressive dysfunction have not been clearly delineated in human disease (**Fig. 238-5**). The late stages are dominated by nonspecific secondary changes in gene expression, and by local and systemic neurohormonal responses, as seen for other etiologies of heart failure.

Figure 238-5

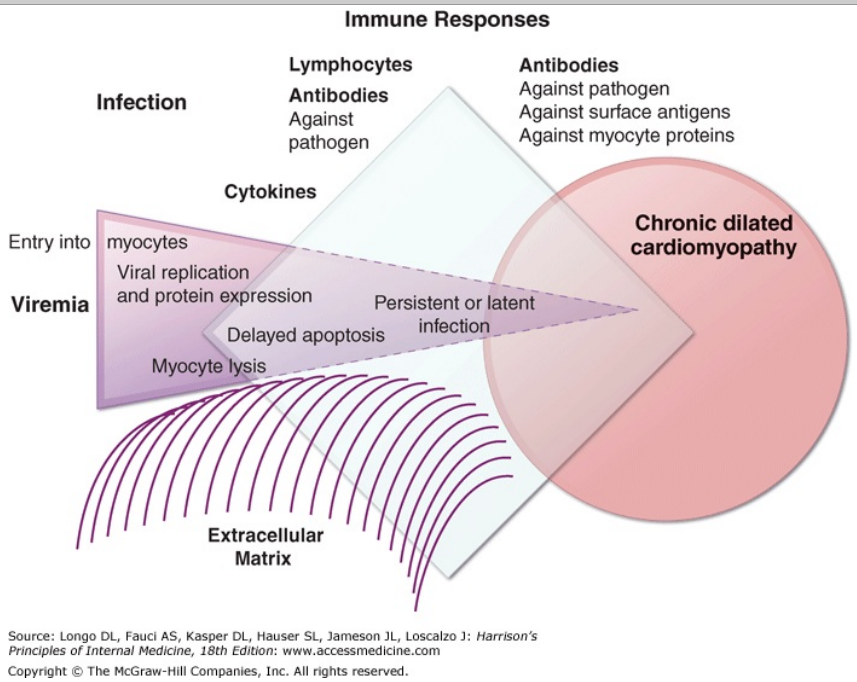

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram demonstrating the possible progression from infection through direct, secondary, and autoimmune responses to dilated cardiomyopathy. Most of the supporting evidence for this sequence is derived from animal models. It is not known to what degree persistent infection and/or ongoing immune responses contribute to ongoing myocardial injury in the chronic phase.

Although viral myocarditis is generally considered to be an acquired cardiomyopathy, families have been reported whose clinical disease appeared after a syndrome consistent with viral myocarditis. One possible explanation for this apparent mixed etiology is that some genetic variants of myocardial cell surface receptors bind more avidly to certain viruses, particularly coxsackie virus and adenovirus.

The typical clinical picture of myocarditis is a young adult with progressive dyspnea and weakness over a few days to weeks after a recent viral syndrome with fevers and often myalgias indicative of skeletal muscle inflammation. Some patients present with atypical or anginal-type chest pain, or with pleuritic, positional chest pain due to pericarditis with some degree of underlying myocarditis. Patients in whom ventricular tachyarrhythmias dominate the presentation may have viral myocarditis but should be evaluated for sarcoidosis or giant cell myocarditis. Patients presenting with pulmonary or systemic embolic events from intracardiac thrombi generally already have chronic, severe cardiac dysfunction.

A small number of patients present with *acute fulminant myocarditis,* with rapid progression from a severe febrile respiratory syndrome to cardiogenic shock from which multiple organ system failure, including coagulopathy, may develop. Such patients have often been discharged from the emergency department with antibiotic therapy only to return in extremis. Prompt triage is vital to provide aggressive support with high-level intravenous inotropic therapy and on occasion, mechanical circulatory support; importantly, more than half of patients with this acute presentation can survive with marked improvement within the first few weeks, often returning to near-normal systolic function.

Many patients presenting with heart failure after a viral illness actually have a long-standing cardiomyopathy that was acutely exacerbated but not caused by the new viral illness. Heart failure from any cause often worsens transiently during infection, presumably due to the myocardial depressant effects of circulating cytokines. Marked left ventricular dilation and the presence of severely elevated left-ventricular filling pressures without frank pulmonary edema suggest chronic, slowly progressive disease, which is often further supported by a history of gradual changes in exercise tolerance before the viral syndrome.

For the usual subacute presentation, the diagnosis of cardiomyopathy is confirmed by echocardiography, and further evaluation is directed to ascertain whether myocarditis is present. Troponin is often mildly elevated, and creatine kinase may be released from the cardiac injury or skeletal muscle involvement. In some cases, cardiac catheterization is performed to rule out acute ischemia. Magnetic resonance imaging is increasingly used for the diagnosis of myocarditis, which is supported by evidence of increased tissue edema and gadolinium enhancement, particularly in the mid-wall distinct from the usual coronary artery territories. Endomyocardial biopsy criteria for myocarditis require lymphocytic infiltration with evidence of myocyte necrosis (**Fig. 238-6**), but are met in only about 10–20% of classic presentations. Most biopsies in fulminant myocarditis show only marked tissue edema without a cellular infiltrate, and it is likely that many less acute cases may similarly be characterized by tissue edema and cytokine depression of myocardial function, possibly including some antibody-mediated endothelial injury, without marked cellular infiltrates. Acute and convalescent viral titers are usually sent but are more likely to be important from the public health standpoint than for the individual.

Figure 238-6

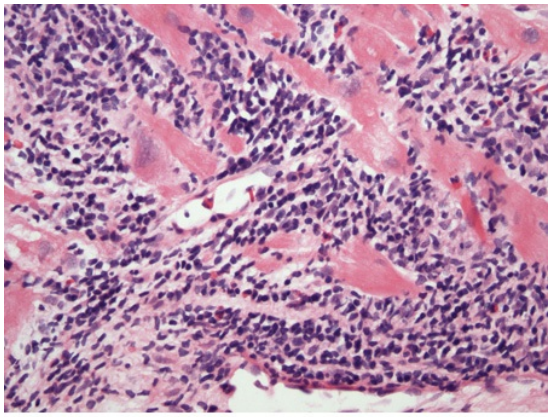

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute myocarditis. Microscopic image of an endomyocardial biopsy showing massive infiltration with mononuclear cells and occasional eosinophils associated with clear myocyte damage. The myocyte nuclei are enlarged and reactive. Such extensive involvement of the myocardium would lead to extensive replacement fibrosis even if the inflammatory response could be suppressed. Hematoxylin and eosin stained section, 200x original magnification. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Viral myocarditis treatment is initially directed toward stabilizing the hemodynamic status and then toward adjusting neurohormonal antagonists for the treatment of heart failure as tolerated. Presentation with fulminant disease requires rapid evaluation and therapy as discussed above. For patients with subacute presentation, randomized trials have shown no benefit of immunosuppression with glucocorticoid combinations or intravenous immunoglobulin, even when the biopsy is positive for lymphocytic infiltrates; yet, immunosuppression is often used even in the absence of evidence of benefit, in part due to perceived analogy to acute cardiac transplant rejection. Animal models have shown that viral replication and myocardial injury can be worsened by immunosuppression during the early phase of infection; however, patients with persistent inflammatory myocarditis and a progressive downhill course over weeks may be treated empirically with glucocorticoids in an attempt to avoid the need for cardiac transplantation.

The true prognosis of viral myocarditis is not known, as most unrecognized cases probably resolve spontaneously, while others progress to cardiomyopathy without other obvious cause. However, among patients who have truly recent onset cardiomyopathy of less than 3–6 months' duration without other apparent etiology, almost half will have major improvement in left ventricular ejection fraction during the subsequent 6–12 months. Those patients in whom left ventricular ejection fraction and dimensions return to normal are usually considered to have residual subclinical cardiomyopathy. Neurohormonal antagonist therapy is usually continued indefinitely as tolerated, with dose adjustments to avoid side effects.

SPECIFIC VIRUSES

In humans, viruses are often suspected but rarely confirmed as the direct cause of myocarditis. Often implicated is the picornavirus family of RNA viruses, with the enteroviruses *Coxsackie, echovirus,* and *poliovirus. Influenza,* another RNA virus, is implicated in myocarditis with varying frequency from year to year as the epitopes change. Of the DNA viruses, *adenovirus,variola* (smallpox) and *vaccinia* (smallpox vaccine), and the *herpesviruses (Varicella zoster, Cytomegalovirus, and Epstein-Barr virus)* are well-recognized as causes of myocarditis. From genetic analyses of biopsy tissue, parvovirus B19, coxsackie, adenovirus, and Epstein-Barr virus are the agents most often implicated. The role of *parvovirus B19* as a cause of myocarditis or cardiomyopathy is difficult to determine, as almost half of individuals show evidence of prior infection with this small DNA virus that causes "fifth disease" in children.

Human immunodeficiency virus (HIV) has been associated with echocardiographic abnormalities in 10–40% patients with clinical disease. Cardiomyopathy in HIV may result from cardiac involvement with other associated viruses, such as cytomegalovirus and hepatitis C. Antiviral drugs to treat chronic HIV can cause cardiomyopathy, both directly as cardiotoxins and through drug hypersensitivity. The clinical picture may be complicated by pericardial effusions and pulmonary hypertension. There is a high frequency of lymphocytic myocarditis found at autopsy, and viral particles have been demonstrated in the myocardium in some cases, consistent with direct causation.

Hepatitis C has been repeatedly implicated in cardiomyopathy, particularly in Germany and Asia. Cardiac function may improve after interferon therapy. As this cytokine itself often depresses cardiac function transiently, careful coordination of administration and ongoing clinical evaluation are critical. Involvement of the heart with hepatitis B is uncommon but can be seen when associated with systemic vasculitis (polyarteritis nodosa).

Other viral infections in which cardiac involvement is specifically implicated, beyond the depression of cardiac function during any systemic cytokine activation, include *mumps, respiratory syncytial virus,* the *arboviruses (dengue fever and yellow fever),* and *arenaviruses (Lassa fever)*.

PARASITIC MYOCARDITIS

Chagas' disease is the third most common parasitic infection in the world and the most common cause of cardiomyopathy. The protozoan *Trypanosoma cruzi* (*T. cruzi*) is usually transmitted by the bite of the reduviid bug, endemic in the rural areas of South and Central America. Transmission can also occur through blood transfusion, organ donation, from mother to fetus, and occasionally orally. While programs to eradicate the insect vector have decreased the prevalence from about 16 million to less than 10 million in South America, cases are increasingly recognized in Western developed countries. Approximately 100,000 affected individuals are currently living in the United States, most of whom contracted the disease in endemic areas.

The acute phase of Chagas' disease with parasitemia is usually unrecognized, but in fewer than 5% of cases presents clinically within a few weeks of infection, with nonspecific symptoms or occasionally with acute myocarditis and meningoencephalitis. In the absence of antiparasitic therapy, the silent stage progresses slowly over 10–30 years in almost half of patients to manifest in the cardiac and gastrointestinal systems in the chronic stages. Survival is less than 30% at 5 years after the onset of overt clinical heart failure.

Multiple pathogenetic mechanisms are implicated. The parasite itself can cause myocyte lysis and primary neuronal damage, and specific immune responses may recognize the parasites or related antigens and lead to chronic immune activation in the absence of detectable parasites. Molecular techniques have revealed persistent parasite DNA fragments in infected individuals. Further evidence for persistent infection is the eruption of parasitic skin lesions during immunosuppression after cardiac transplantation. As in postviral myocarditis, the relative roles of persistent infection and of secondary autoimmune injury have not been resolved (Fig. 238-5). An additional factor in progression of Chagas' disease is the autonomic dysfunction and microvascular damage that may contribute to cardiac and gastrointestinal disease.

Features typical of Chagas' disease are conduction system abnormalities, particularly sinus node and atrioventricular (AV) node dysfunction and right bundle branch block. Atrial fibrillation and ventricular tachyarrhythmias also occur. Small ventricular aneurysms are common, particularly at the apex. The dilated ventricles are particularly thrombogenic, giving rise to pulmonary and systemic emboli. The serologic enzyme-linked immunosorbent assay (ELISA) for the IgM has largely replaced the previous complement fixation test for diagnosis.

Treatment of the advanced stages focused on the clinical manifestations of the disease, with heart failure regimens, pacemaker-defibrillators, and anticoagulation; however, increasing emphasis is placed on antiparasitic therapy even in chronic disease. The most common effective antiparasitic therapies are benznidazole and nifurtimox, both associated with multiple severe reactions, including dermatitis, gastrointestinal distress, and neuropathy. Patients without major extracardiac disease have occasionally undergone transplantation, after which they require lifelong therapy to suppress reactivation of infection.

African trypanosomiasis infection results from the tsetse fly bite and can occur in travelers exposed during trips to Africa. The West African form is caused by *Trypanosoma brucei gambiense* and progresses silently over years. The East African form caused by *T. brucei rhodesiense* can progress rapidly through perivascular infiltration to myocarditis and heart failure, with frequent arrhythmias. The diagnosis is made by identification of trypanosomes in blood, lymph nodes, or other affected sites. Development of optimal drug regimens remains limited, and depends on the type and the stage (hemolymphatic or neurologic).

Toxoplasmosis is contracted through undercooked infected beef or pork, transmission from feline feces, organ transplantation, transfusion, or maternal-fetal transmission. Immunocompromised hosts are at greatest risk for reactivation of latent infection from cysts. The cysts have been found in up to 40% of autopsies of patients dying from HIV infection. Toxoplasmosis may present with encephalitis or chorioretinitis, and in the heart can cause myocarditis, pericardial effusion, constrictive pericarditis, and heart failure. The diagnosis may be suspected in an immunocompromised patient with myocarditis and serologic evidence of toxoplasmosis. Fortuitous sampling may reveal the cysts in the myocardium. Combination therapy can include pyrimethamine and sulfadiazine or clindamycin.

Trichinellosis is caused by *Trichinella spiralis* larva ingested with undercooked meat. Larva migrating into skeletal muscles cause myalgias, weakness, and fever. Periorbital and facial edema, and conjunctival and retinal hemorrhage may also be seen. Although the larva may occasionally invade the myocardium, clinical heart failure is rare, and when observed, attributed to the eosinophilic inflammatory response. The diagnosis is made from the specific serum antibody and is further supported by the presence of eosinophilia. Treatment includes antihelminthic drugs and glucocorticoids if inflammation is severe.

Cardiac involvement with *Echinococcus* is rare, but cysts can form and rupture in the myocardium and pericardium.

BACTERIAL INFECTIONS

Most bacterial infections can involve the heart occasionally through direct invasion and abscess formation, but do so rarely. More commonly, contractility is depressed globally in severe infection and sepsis through systemic inflammatory responses. *Diphtheria* specifically affects the heart in almost one-half of cases and is the most common cause of death in patients with this infection. Once a disease of children, the prevalence of vaccines has shifted the incidence of this disease to countries where immunization is not routine and to older populations who have lost their immunity. The bacillus releases a toxin that impairs protein synthesis and may particularly affect the conduction system. The specific antitoxin should be administered as soon as possible, with higher priority than antibiotic therapy. Other systemic bacterial infections that can involve the heart include *brucellosis, chlamydophila, legionella, meningococcus, mycoplasma, psittacosis, and salmonellosis,* for which treatment is directed at the systemic infection.

Clostridial infections cause myocardial damage from the released toxin. Gas bubbles can be detected in the myocardium, and occasionally abscesses can form in the myocardium and pericardium. *Streptococcal infection* with β-hemolytic streptococci is most commonly associated with acute rheumatic fever, and is characterized by inflammation and fibrosis of cardiac valves and systemic connective tissue, but can also lead to a myocarditis with focal or diffuse infiltrates of mononuclear cells.

Tuberculosis can involve the myocardium directly as well as through tuberculous pericarditis, but rarely does so when the disease is treated with antibiotics. *Whipple's disease* is caused by *Tropheryma whippleii.* The usual manifestations are in the gastrointestinal tract, but pericarditis, coronary arteritis, valvular lesions, and occasionally clinical heart failure may also occur. Multidrug antituberculous regimens are effective, but the disease tends to relapse even with appropriate treatment.

OTHER INFECTIONS

Spirochetal myocarditis has been diagnosed from myocardial biopsies containing *Borrelia burgdorferi* that causes *Lyme disease.* Lyme carditis most often presents with arthritis and conduction system disease that resolves within 1–2 weeks of antibiotic treatment, only rarely causing clinical heart failure.

Fungal myocarditis can occur due to hematogenous or direct spread of infection from other sites, as has been described for aspergillosis, actinomycosis, blastomycosis, candidiasis, coccidioidomycosis, cryptococcosis, histoplasmosis, and mucormycosis. However, cardiac infection is rarely the dominant clinical feature of these infections.

The **rickettsial infections,** *Q fever, Rocky Mountain spotted fever, and scrub typhus,* are frequently accompanied by ECG changes, but most clinical manifestations relate to systemic vascular involvement.

Noninfective Myocarditis

Myocardial inflammation can occur without apparent preceding infection. The paradigm of noninfectious inflammation without infection is cardiac transplant rejection, from which we have learned that myocardial depression can develop and reverse quickly, that noncellular mediators such as antibodies and cytokines play a major role in addition to lymphocytes, and that myocardial antigens are exposed by prior physical injury and viral infection.

The most commonly diagnosed noninfectious inflammation is granulomatous myocarditis, including both sarcoidosis and giant cell myocarditis. Sarcoidosis, as discussed in Chap. 329, is a multisystem disease most commonly affecting the lungs, presenting in young adults with higher prevalence in African-American males. Patients with pulmonary sarcoid are at high risk for cardiac involvement, but cardiac sarcoidosis may also occur without clinical lung disease in middle-aged whites of both genders. Regional clustering of the disease supports the suspicion that the granulomatous reaction is triggered by an infectious or environmental allergen not yet identified.

The sites and density of cardiac granulomata, the time course, and the degree of extracardiac involvement are remarkably variable. Patients may present with rapid onset heart failure and ventricular tachyarrhythmias, conduction block, chest pain syndromes, or minor cardiac findings in the setting of ocular involvement, an infiltrative skin rash, or a nonspecific febrile illness. They may also present less acutely after months to years of fluctuating cardiac symptoms. When ventricular tachycardia or conduction block dominate the initial presentation of heart failure without coronary artery disease, suspicion should be high for these granulomatous myocarditides.

Depending on the time course, the ventricles may appear restrictive or dilated, at times with right ventricular predominance. Small ventricular aneurysms are common. Computed tomography of the chest often reveals pulmonary lymphadenopathy even in the absence of clinical lung disease. Metabolic imaging [positron emission tomography (PET)] of the whole chest can highlight active sarcoid lesions that are avid for glucose. Magnetic resonance imaging (MRI) of the heart can identify areas likely to be inflammatory. To rule out chronic granulomatous infections, the diagnosis usually requires pathologic confirmation. Biopsy of enlarged mediastinal nodes may provide the highest yield. The scattered granulomata of sarcoidosis can be missed on cardiac biopsy (**Fig. 238-7**).

Figure 238-7

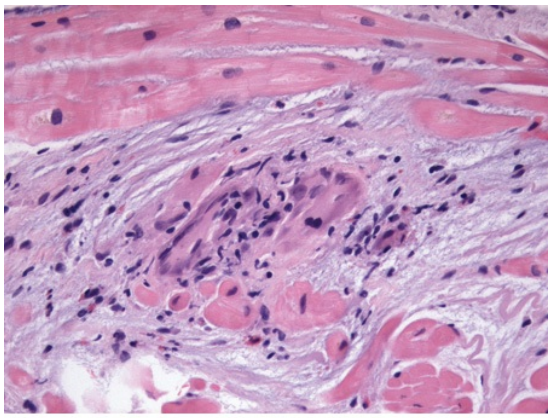

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sarcoidosis. Microscopic image of an endomyocardial biopsy showing a noncaseating granuloma and associated interstitial fibrosis typical of sarcoidosis. No microorganisms were present on special stains, and no foreign material was identified. Hematoxylin and eosin stained section, 200x original magnification. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Immunosuppressive treatment for sarcoidosis is initiated with high-dose glucocorticoids, which are often more effective for arrhythmias than for the heart failure. Pacemakers and implantable defibrillators are generally indicated to prevent life-threatening heart block or ventricular tachycardia, respectively. Because the inflammation often resolves into extensive fibrosis that impairs cardiac function and provides pathways for reentrant arrhythmias, the prognosis is best when the granulomata are not extensive.

Giant cell myocarditis is less common than sarcoidosis, but accounts for 10–20% of biopsy-positive cases of myocarditis. Giant cell myocarditis typically presents with rapidly progressive heart failure and tachyarrhythmias. Diffuse granulomatous lesions are surrounded by extensive inflammatory infiltrate unlikely to be missed on endomyocardial biopsy. Associated conditions are thymomas, thyroiditis, pernicious anemia, other autoimmune diseases, and occasionally recent infections. Glucocorticoid therapy is less effective than for sarcoidosis, and is sometimes combined with other immunosuppressive agents. The course is generally of rapid deterioration requiring urgent transplantation. Although the severity of presentation and myocardial histology are more fulminant than sarcoidosis, the occasional finding of giant cell myocarditis after sarcoidosis suggests that they may in some cases represent different stages of a similar disease.

Hypersensitivity myocarditis is usually an unexpected diagnosis, made when the biopsy reveals infiltration with lymphocytes and mononuclear cells with a high proportion of eosinophils. (Sometimes called eosinophilic myocarditis, this should not be confused with the hypereosinophilic syndrome in which very high circulating, often clonal populations of eosinophils cause endomyocardial fibrosis.) Most commonly the reaction is attributed to antibiotics, particularly those taken chronically, but thiazides, anticonvulsants, indomethacin, and methyldopa have also been implicated. High-dose glucocorticoids can be curative.

Myocarditis can be associated with systemic inflammatory diseases such as polymyositis and *dermatomyositis.* While sometimes considered as an explanation for cardiac findings in patients with other inflammatory disease, such as systemic lupus erythematosus, the more common causes are pericarditis, vasculitis, pulmonary hypertension, or accelerated coronary artery disease.

Peripartum cardiomyopathy develops during the last trimester or within the first 6 months after pregnancy, with a frequency between 1:3,000 and 1:15,000 deliveries. The mechanisms remain controversial, but inflammation has been implicated. Risk factors are increased maternal age, increased parity, twin pregnancy, malnutrition, use of tocolytic therapy for premature labor, and preeclampsia or toxemia of pregnancy. As the increased circulatory demand of pregnancy can aggravate other cardiac disease that was clinically unrecognized, it is crucial to the diagnosis that there be no evidence for preexisting cardiac disorder.

Heart failure early after delivery was previously common in Nigeria, when the custom for new mothers included salt ingestion while reclining on a warm bed, which likely impaired mobilization of the excess circulating volume after delivery. In the Western world, lymphocytic myocarditis has often been found on myocardial biopsy. This inflammation has been hypothesized to reflect increased susceptibility to viral myocarditis or an autoimmune myocarditis due to cross-reactivity of anti-uterine antibodies against cardiac muscle. Another mechanism involving a prolactin cleavage fragment has been proposed based on an animal model.

Toxic Cardiomyopathy

Cardiotoxicity has been reported with multiple environmental and pharmacologic agents. Often these associations are seen only with very high levels of exposure or acute overdoses, respectively, in which acute electrocardiographic and hemodynamic abnormalities may reflect both direct drug effect and systemic toxicity.

Alcohol is the most common toxin implicated in chronic dilated cardiomyopathy. Excess consumption may contribute to more than 10% of cases of heart failure, including exacerbation of cases with other primary etiologies such as valvular disease or previous infarction. Toxicity is attributed both to alcohol and to its primary metabolite acetaldehyde. Polymorphisms of the genes encoding alcohol dehydrogenase and the angiotensin-converting enzyme increase the likelihood of alcoholic cardiomyopathy. Superimposed vitamin deficiencies and toxic alcohol additives are rarely implicated. The alcohol consumption necessary to produce cardiomyopathy in an otherwise normal heart has been estimated to be six drinks (about 4 ounces of pure ethanol) daily for 5–10 years, but frequent binge drinking may also be sufficient. Many patients with alcoholic cardiomyopathy are fully functional without apparent stigmata of alcoholism.

Diastolic dysfunction, mild ventricular dilation, and subclinical depression of contractility can be seen before the development of clinical heart failure. Atrial fibrillation occurs commonly. The cardiac impairment in severe alcoholic cardiomyopathy is the sum of both permanent damage and a substantial component that is reversible after cessation of alcohol consumption. Medical therapy includes neurohormonal antagonists and diuretics as needed for fluid management. Withdrawal should be supervised to avoid exacerbations of heart failure or arrhythmias, and ongoing support arranged. Even with severe disease, marked improvement can occur within 3–6 months of abstinence. Implantable defibrillators are generally deferred until an adequate period of abstinence, after which they may not be necessary if the ejection fraction has improved. With continued consumption, the prognosis is grim.

Cocaine, amphetamines, and related catecholaminergic stimulants can produce chronic cardiomyopathy as well as acute ischemia and tachyarrhythmias. Pathology reveals tiny microinfarcts consistent with small vessel ischemia. Similar findings can be seen with pheochromocytoma.

Chemotherapy agents are the most common drugs implicated in cardiomyopathy. Judicious use of these drugs requires balancing the risks of the malignancy and the risks of cardiotoxicity, as many cancers have a chronic course with prognosis no worse than heart failure.

Anthracyclines cause characteristic histologic changes of vacuolar degeneration and myofibrillar loss. Generation of reactive oxygen species involving heme compounds is currently the favored explanation for myocyte injury and fibrosis. Disruption of the large titin protein may contribute to loss of sarcomere organization. There are three different presentations of anthracycline-induced cardiomyopathy. Acute heart failure developing during administration of a single dose can be severe, but may clinically resolve in a few weeks. Early onset doxorubicin cardiotoxicity develops in about 3% of patients during or shortly after a chronic course, relating closely to total dose. It may be rapidly progressive, but may also improve to restore reasonable ventricular function. The chronic presentation differs according to whether therapy was given before or after puberty. Patients who received doxorubicin while still growing may have inadequate development of the heart to support cardiac function into the early twenties. Late after adult exposure, patients may develop the gradual onset of symptoms or an acute onset precipitated by a reversible insult, such as influenza or atrial fibrillation. Doxorubicin cardiotoxicity leads to a relatively nondilated ventricle, perhaps due to the accompanying fibrosis. Thus, the stroke volume may be severely reduced with an ejection fraction of 30–40%, which would be well tolerated in a patient with a more dilated ventricle typical of other cardiomyopathies. Therapy is that for heart failure, with careful suppression of "inappropriate" sinus tachycardia, and attention to postural hypotension that can occur in these patients. Once thought to have an inexorable downward course, some patients with doxorubicin cardiotoxicity improve under careful management to near-normal clinical function for many years.

Trastuzumab is a monoclonal antibody that interferes with cell surface receptors crucial for some tumor growth and for cardiac adaptation. The incidence of cardiotoxicity is lower than for anthracyclines but enhanced by coadministration with them. Although considered to be more often reversible, trastuzumab cardiotoxicity does not always resolve, and some patients progress to clinical heart failure and death. As with anthracycline cardiotoxicity, therapy is as usual for heart failure, but it is not clear whether or not the spontaneous rate of improvement is enhanced by neurohormonal antagonists.

Cardiotoxicity with *cyclophosphamide and ifosfamide* generally occurs acutely and with very high doses. 5-Fluorouracil, cisplatin, and some other alkylating agents can cause recurrent coronary spasm that occasionally leads to depressed contractility. Many small molecule *tyrosine kinase inhibitors* are under development for different malignancies. Although these agents are "targeted" at specific tumor receptors or pathways, the biologic conservation of signaling pathways can cause inhibitors to have "off-target" effects that include the heart and vasculature. Acute administration of *interferon-α* can cause hypotension and arrhythmias. Clinical heart failure occurring during repeated chronic administration usually resolves after discontinuation.

Other therapeutic drugs that can cause cardiotoxicity during chronic use include hydroxychloroquine, chloroquine, emetine, and antiretroviral therapies.

Toxic exposures are most commonly implicated in arrhythmias or respiratory injury acutely during accidents. Chronic exposures that can cause cardiotoxicity include hydrocarbons, fluorocarbons, arsenicals, lead, and mercury.

Metabolic Causes of Dilated Cardiomyopathy

Endocrine disorders affect multiple organ systems, including the heart. *Hyperthyroidism and hypothyroidism* do not often cause clinical heart failure in an otherwise normal heart, but commonly exacerbate heart failure. The most common, current reason for thyroid abnormalities in the heart failure population is the use of amiodarone, a drug with substantial iodine content. Clinical signs of thyroid disease may be masked, so tests of thyroid function are part of the routine evaluation of cardiomyopathy. Hypothyroidism should be treated with very slow escalation of doses to avoid exacerbating tachyarrhythmias and heart failure. Hyperthyroidism should always be considered with new onset atrial fibrillation or ventricular tachycardia, or atrial fibrillation in which the rapid ventricular response is difficult to control. Hyperthyroidism and heart failure are a dangerous combination that merits very close supervision, often hospitalization, during titration of antithyroid medications, which may lead to precipitous worsening of heart failure. *Pheochromocytoma* is rare, but should be considered when a patient has heart failure and very labile blood pressure and heart rate, sometimes with episodic palpitations (Chap. 343). Most patients with pheochromocytoma have postural hypotension. In addition to α-adrenergic receptor antagonists, definitive therapy requires surgical extirpation. Very high renin states, such as those caused by renal artery stenosis, can lead to modest depression in ejection fraction with little or no ventricular dilation and markedly labile symptoms with flash pulmonary edema, related to sudden shifts in vascular tone and intravascular volume.

Controversies remain regarding whether *diabetes* and *obesity* are sufficient to cause cardiomyopathy. Most heart failure in diabetes results from epicardial coronary disease, with further increase in coronary artery risk due to accompanying hypertension and renal dysfunction. Cardiomyopathy may result in part from insulin resistance and increased advanced-glycosylation end products, which impair both systolic and diastolic function. However, much of the dysfunction can be attributed to scattered focal ischemia resulting from distal coronary artery tapering and limited microvascular perfusion even without proximal focal stenoses. Diabetes is a typical factor, along with hypertension, advanced age, and female gender, in heart failure with "preserved" ejection fraction.

The existence of a cardiomyopathy due to *obesity* is generally accepted. In addition to cardiac involvement from associated diabetes, hypertension, and vascular inflammation of the metabolic syndrome, obesity alone is associated with impaired excretion of excess volume load, which, over time, can lead to increased wall stress and secondary adaptive neurohumoral responses. The rapid clearance of natriuretic peptides by adipose tissue may contribute to fluid retention. In the absence of another obvious cause of cardiomyopathy in an obese patient with systolic dysfunction without marked ventricular dilation, effective weight reduction is often associated with major improvement in ejection fraction and clinical function.

Nutritional deficiencies can occasionally cause dilated cardiomyopathy but are not commonly implicated in developed Western countries. *Beri-beri heart disease* due to thiamine deficiency can result from poor nutrition in undernourished populations, and in patients deriving most of their calories from alcohol, and has been reported in teenagers subsisting only on highly processed foods. This disease is initially a vasodilated state with very high output heart failure that can later progress to a low output state; thiamine repletion can lead to prompt recovery of cardiovascular function. Abnormalities in *carnitine* metabolism can cause dilated or restrictive cardiomyopathies, usually in children. Deficiency of trace elements such as *selenium* can cause cardiomyopathy (Keshan's disease).

Calcium is essential for excitation-contraction coupling, serving as an inotrope when administered. Chronic deficiencies of calcium, such as can occur with hypoparathyroidism (particularly postsurgical) or intestinal dysfunction (from diarrheal syndromes and following extensive resection), can cause severe chronic heart failure that responds over days or weeks to vigorous calcium repletion. *Phosphate* is a component of high-energy compounds needed for efficient energy transfer and multiple signaling pathways. Hypophosphatemia can develop during starvation and early refeeding following a prolonged fast, and occasionally during hyperalimentation. *Magnesium* is a cofactor for thiamine-dependent reactions and for the sodium-potassiumadenosine triphosphatase (ATPase), but hypomagnesemia rarely becomes sufficiently profound to cause clinical cardiomyopathy.

Hemochromatosis is variably classified as a metabolic or storage disease. It is included among the causes of restrictive cardiomyopathy, but the clinical presentation is often that of a dilated cardiomyopathy. The autosomal recessive form is related to the HFE gene. With up to 10% of the population heterozygous for one mutation, the clinical prevalence might be as high as 1 in 500. The lower rates observed highlight the limited penetrance of the disease, suggesting the role of additional genetic and environmental factors for clinical expression. The clinical syndrome includes cirrhosis, diabetes, and hypogonadism (Chap. 357). Hemochromatosis can also be acquired from iron overload due to hemolytic anemia and transfusions. Excess iron is deposited in the perinuclear compartment of cardiomyocytes, with resulting disruption of intracellular architecture and mitochondrial function.

Diagnosis is easily made from measurement of serum iron and transferrin saturation, with a threshold of >60% for men, and >45 –50% for women. Magnetic resonance imaging can help to quantitate iron stores in the liver and heart, and endomyocardial biopsy tissue can be stained for iron (**Fig. 238-8**). If diagnosed early, hemochromatosis can often be managed by repeated phlebotomy to remove iron. For more severe iron overload, iron chelation therapy with desferrioxamine (deferoxamine) or deferasirox can help to improve cardiac function if myocyte loss and replacement fibrosis are not too severe. Inborn disorders of metabolism occasionally present with dilated cardiomyopathy, although are most often associated with restrictive cardiomyopathy (Table 238-4).

Figure 238-8

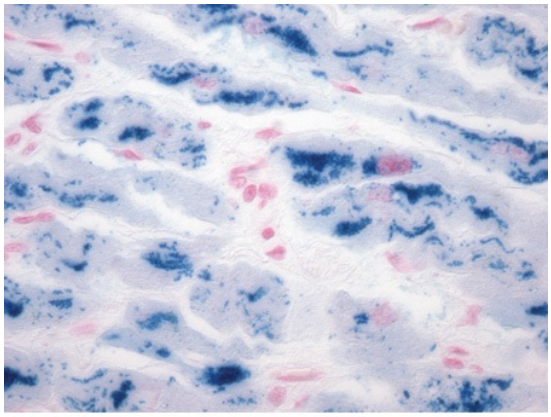

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hemochromatosis. Microscopic image of an endomyocardial biopsy showing extensive iron deposition within the cardiac myocytes with the Prussian blue stain (400x original magnification). (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Familial Dilated Cardiomyopathy

The recognized frequency of familial involvement in dilated cardiomyopathy has now increased to an estimated 30% (Table 238-3). The most recognizable familial syndromes are the *muscular dystrophies.* Both Duchenne's and the milder Becker's dystrophy result from abnormalities in the X-linked dystrophin gene of the sarcolemmal membrane. Skeletal myopathy is present in multiple other genetic cardiomyopathies (Table 238-3), some of which are associated with creatine kinase elevations. Mitochondrial myopathies are associated with varying degrees of skeletal involvement, biopsies of which show the characteristic "ragged red fiber" appearance. Some patients with mitochondrial myopathy have characteristic drooping eyelids. The energy deficits associated with mitochondrial abnormalities lead to multiple systemic syndromes. Other familial metabolic defects more often present as restrictive disease, but can sometimes be identified on electron microscopy of endomyocardial biopsies.

Families with a history of atrial arrhythmias, conduction system disease, and cardiomyopathy may have abnormalities of the nuclear membrane lamin proteins. While all dilated cardiomyopathies carry a risk of sudden death, a family history of cardiomyopathy with sudden death raises suspicion for a particularly arrhythmogenic mutation; affected family members may be considered for implantable defibrillators even before meeting the reduced ejection fraction threshold for primary prevention of sudden death.

A prominent family history of sudden death or ventricular tachycardia before clinical cardiomyopathy suggests genetic defects in the desmosomal proteins causing *arrhythmogenic ventricular dysplasia* (**Fig. 238-9**). Originally described as affecting the right ventricle [arrhythmogenic right ventricular dysplasia (ARVD)], this disorder can affect either or both ventricles. Patients often present first with ventricular tachycardia. Genetic defects in proteins of the desmosomal complex disrupt myocyte junctions and adhesions, leading to replacement of myocardium by deposits of fat. Thin ventricular walls may be recognized on echocardiography but are better visualized on MRI. The same protein also affects hair and skin, leading in some cases to a distinct syndrome of "woolly hair," and thickened palms and soles. Implantable defibrillators are usually indicated to prevent sudden death. There is variable progression to right, left, or biventricular failure.

Figure 238-9

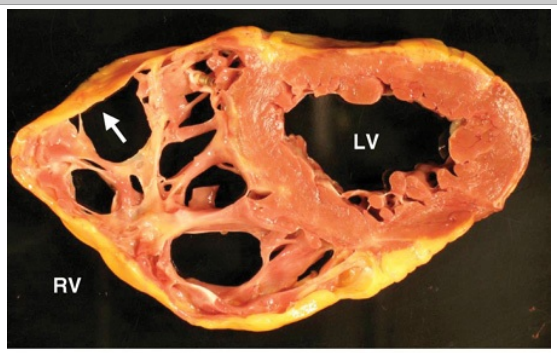

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arrhythmogenic right ventricular dysplasia. ***A.*** Cross-sectional slice of a pathology specimen removed at transplantation, showing severe dysplasia of the right ventricle (RV) with extensive fatty replacement of right ventricular myocardium. The remarkably thin right ventricular free wall is revealed by transillumination ***B.*** (*Images courtesy of Gayle Winters, MD, and Richard Mitchell, MD, PhD, Division of Pathology, Brigham and Women's Hospital, Boston.*)

Left ventricular noncompaction is a condition of unknown prevalence that is increasingly revealed by better imaging techniques, first by two-dimensional echocardiography and more recently by magnetic resonance imaging. The diagnostic criteria includes the presence of multiple trabeculations in the left ventricle distal to the papillary muscles, creating a "spongy" appearance of the apex; it has been associated with multiple genetic variants in the sarcomeric and other proteins such as tafazzin. The condition may be diagnosed incidentally or in patients carrying previous diagnoses of dilated, restrictive, or hypertrophic cardiomyopathy. The three cardinal clinical features are ventricular arrhythmias, embolic events, and heart failure. Treatment generally includes anticoagulation and consideration for an implantable defibrillator.

Some families inherit a susceptibility to viral-induced myocarditis. This propensity may relate to abnormalities in cell surface receptors, such as the coxsackie-adenovirus receptor, that bind viral proteins. Some may have partial homology with viral proteins such that an autoimmune response is triggered against the myocardium.

The therapy of familial dilated cardiomyopathy is dictated primarily by the stage of clinical disease and the risk for sudden death. In some cases, the familial etiology facilitates prognostic decisions, particularly regarding the likelihood of recovery after a new diagnosis, which is unlikely for familial disease and frequent if the disease is acquired. The rate of progression of disease is to some extent heritable, although marked variation can be seen; however, there have been cases of remarkable clinical remission after acute presentation, likely after a reversible insult, such as infective myocarditis.

Genetic testing is less robust for dilated cardiomyopathy, for which our current understanding is similar to that for hypertrophic cardiomyopathy a decade ago. Newer molecular techniques, animal models, and databanks of cardiomyopathy patients are all contributing to the rapid expansion of the data presented in Table 238-3. However, serendipitous identification of inherited cardiomyopathy, its systemic signature, and clinical course remain crucial to continue to advance the field, one family and one gene at a time.

Tako-Tsubo Cardiomyopathy

The apical ballooning syndrome, or stress-induced cardiomyopathy, occurs typically in older women after sudden intense emotional or physical stress. The ventricle shows global ventricular dilation with basal contraction, forming the shape of the narrow-necked jar (*tako-tsubo*) used in Japan to trap octopi. Originally described there, it is increasingly recognized in other countries and may go unrecognized during intensive care unit (ICU) admission for noncardiac conditions. Presentations include pulmonary edema, hypotension, and chest pain with ECG changes mimicking an acute infarction. The left ventricular dysfunction extends beyond a specific coronary artery distribution and generally resolves within days to weeks, but may recur in up to 10% of patients. Animal models and ventricular biopsies suggest that this acute cardiomyopathy may result from intense sympathetic activation with heterogeneity of myocardial autonomic innervation, diffuse microvascular spasm, and/or direct catecholamine toxicity. Coronary angiography may be required to rule out acute coronary occlusion. No therapies have been proven beneficial, but reasonable strategies include nitrates for pulmonary edema, intraaortic balloon pump if needed for low output, combined alpha- and beta blockers rather than selective beta blockade if hemodynamically stable, and magnesium for arrhythmias related to QT prolongation. Anticoagulation is generally withheld due to the occasional occurrence of ventricular rupture.

Idiopathic Dilated Cardiomyopathy

Idiopathic dilated cardiomyopathy is a diagnosis of exclusion, when all other known factors have been excluded. Approximately two-thirds of dilated cardiomyopathies are still labeled as idiopathic; however, a substantial proportion of these may reflect unrecognized genetic disease. Continued reconsideration of etiology often reveals specific causes later in a patient's course.

OVERLAP BETWEEN CARDIOMYOPATHIES

The limitations of our phenotypic classification are revealed through the multiple overlaps between the etiologies and presentations of the three types. Cardiomyopathy with reduced systolic function but without severe dilation can represent early dilated cardiomyopathy, "minimally dilated cardiomyopathy," or restrictive diseases without marked increases in ventricular wall thickness. For example, sarcoidosis and hemochromatosis can present as dilated or restrictive disease. Early stages of amyloidosis sometimes appear as dilated cardiomyopathy, but can also be mistaken for hypertrophic cardiomyopathy. Progression of hypertrophic cardiomyopathy into a "burned-out" phase occurs occasionally, with decreased contractility and modest ventricular dilation. Overlaps are particularly common with the inherited metabolic disorders, which can present as any of the three major phenotypes (Fig. 238-4).

RESTRICTIVE CARDIOMYOPATHY

The least common of the triad of cardiomyopathies is restrictive cardiomyopathy, which is dominated by abnormal diastolic function, often with mildly decreased contractility and ejection fraction (usually >30–50%). Both atria are enlarged, sometimes massively.

Modest left ventricular dilation can be present, usually with an end-diastolic dimension <6 cm. End-diastolic pressures are elevated in both ventricles, with preservation of cardiac output until late in the disease. Subtle exercise intolerance is usually the first symptom but is often not recognized until after clinical presentation with congestive symptoms. The restrictive diseases often present with relatively more right-sided symptoms, such as edema, abdominal discomfort, and ascites, although filling pressures are elevated in both ventricles. The cardiac impulse is less displaced than in dilated cardiomyopathy and less dynamic than in hypertrophic cardiomyopathy. A fourth heart sound is more common than a third heart sound in sinus rhythm, but atrial fibrillation is common. Jugular venous pressures often show rapid Y descents, and may increase during inspiration (positive Kussmaul's sign). Most restrictive cardiomyopathies are due to infiltration of abnormal substances between myocytes, storage of abnormal metabolic products within myocytes, or fibrotic injury (**Table 238-6**).

Table 238-6 Causes of Restrictive Cardiomyopathies

Infiltrative (Between Myocytes)
Amyloidosis
Primary (light chain amyloid)
Familial (abnormal transthyretin)[a]
Senile (normal transthyretin or atrial peptides)
Inherited metabolic defects[a] (see Table 238-4)
Storage (Within Myocytes)
Hemochromatosis (iron)[a]
Inherited metabolic defectsa (see Table 238-4)
Fabry's disease
Glycogen storage disease (II, III)
Fibrotic
Radiation
Scleroderma
Endomyocardial
Possibly related fibrotic diseases
Tropical endomyocardial fibrosis
Hypereosinophilic syndrome (Löffler's endocarditis)
Carcinoid syndrome
Radiation
Drugs: e.g., serotonin, ergotamine
Overlap with Other Cardiomyopathies
Hypertrophic cardiomyopathy/"pseudohypertrophic"[a]
"Minimally dilated" cardiomyopathy
Early stage dilated cardiomyopathy
Partial recovery from dilated cardiomyopathy
Sarcoidosis
Idiopathic[a]

[a]Can be familial.

Infiltrative Disease

Amyloidosis is the major cause of restrictive cardiomyopathy (**Figs. 238-10, 238-11, 238-12**), most often due to "primary amyloidosis" (Chap. 112) caused by abnormal production of immunoglobulin light chains. Familial amyloidosis results from an autosomal dominant mutation in transthyretin, a carrier protein for thyroxine and retinol, that is more common in African Americans than whites. Amyloidosis secondary to other chronic diseases rarely involves the heart. Senile amyloidosis with deposition of normal transthyretin or atrial natriuretic peptide usually has an indolent course and is very common beyond the seventh decade.

Figure 238-10

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Restrictive cardiomyopathy—amyloidosis. Gross specimen of a heart with amyloidosis. The heart is firm and rubbery with a waxy cut surface. The atria are markedly dilated and the left atrial endocardium, normally smooth, has yellow-brown amyloid deposits that give texture to the surface. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Figure 238-11

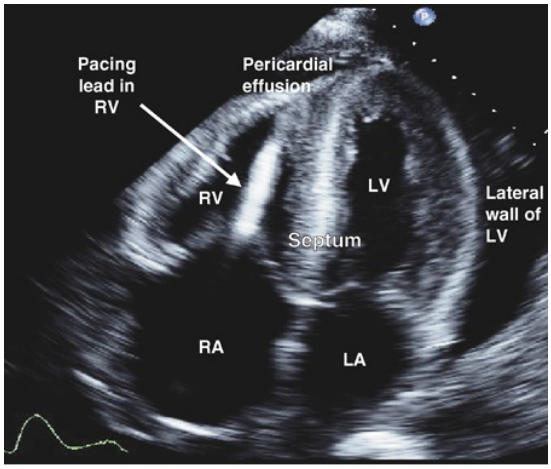

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Restrictive cardiomyopathy—amyloidosis. Echocardiogram showing thickened walls of both ventricles without major chamber dilation. The atria are markedly dilated, consistent with chronically elevated ventricular filling pressures. In this example, there is a characteristic hyper-refractile "glittering" of the myocardium typical of amyloid infiltration, which is often absent (especially with more recent echocardiographic systems of better resolution). The mitral and tricuspid valves are thickened. A pacing lead is visible in the right ventricle and a pericardial effusion is evident. Note that the echocardiographic and pathologic images are vertically opposite, such that the LV is by convention on the top right in the echocardiographic image and bottom right in the pathologic images. (*Image courtesy of Justina Wu, MD, Brigham and Women's Hospital, Boston.*)

Figure 238-12

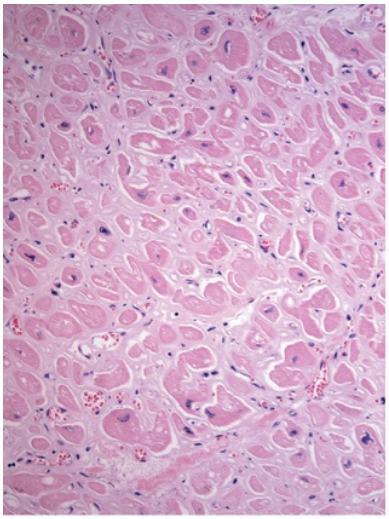

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

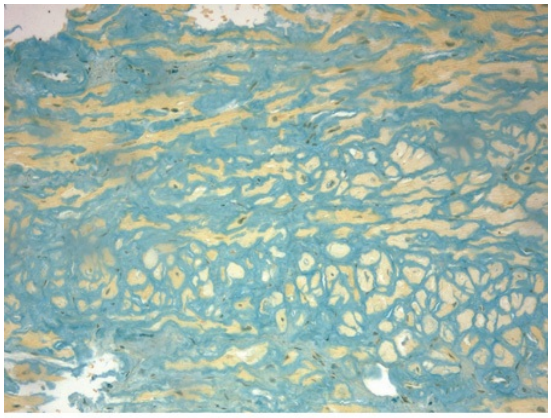

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Amyloidosis—microscopic images of amyloid involving the myocardium. The left panel (hematoxylin-eosin stain) shows glassy, grey-pink amorphous material infiltrating between cardiomyocytes, which stain a darker pink. The right panel shows a sulfated blue stain that highlights the amyloid green and stains the cardiac myocytes yellow. (The Congo red stain can also be used to highlight amyloid; under polarized light, amyloid will have an apple-green birefringence when stained with Congo red). Images at 100x original magnification. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Amyloid fibrils infiltrate the myocardium, especially around the conduction system and coronary vessels. Typical clinical features are conduction block, autonomic neuropathy, renal involvement, and occasionally thickened skin lesions. Cardiac amyloid is suspected from thickened ventricular walls in conjunction with an electrocardiogram that shows low voltage. A characteristic refractile brightness in the septum on echocardiography is suggestive, but neither sensitive nor specific. Both atria are dilated, often dramatically so. The diagnosis of primary or familial can be made from biopsies of an abdominal fat pad or the rectum, but cardiac amyloidosis is most reliably identified from the myocardium (Fig. 238-12). Therapy is largely symptomatic, using diuretics as needed to treat fluid retention, which often requires high doses. Digoxin bound to the amyloid fibrils can reach toxic levels, and should therefore be used only in very low doses, if at all. There is no evidence regarding use of neurohormonal antagonists in amyloid heart disease, where the possible theoretical benefit has to be balanced against their potential side effects in light of frequent autonomic neuropathy and dependence on heart rate reserve. The risk of intracardiac thrombi may warrant chronic anticoagulation. Once heart failure develops, the median survival is 6–12 months in primary amyloidosis. Multiple myeloma is treated with chemotherapy (prednisone, melphalan, bortezomib), the extent of which is usually limited by the potential of worsening cardiac dysfunction. Colchicine can be of some benefit in inflammation-associated (AA) amyloid. Transthyretin-associated cardiac amyloid requires heart and liver transplantation, while senile cardiac amyloid is treated with conventional heart failure regimens. Immunoglobulin-associated amyloid has occasionally been treated with sequential heart transplantation and delayed bone marrow transplant, with frequent recurrence of amyloid in the transplanted heart.

Disorders of Metabolic Pathways

Multiple genetic disorders of metabolic pathways can cause myocardial disease, due to infiltration of abnormal products or cells containing them between the myocytes, and storage disease, due to their accumulation within cells (**Tables 238-4, 238-6**). The restrictive phenotype is most common but mildly dilated cardiomyopathy may occur. Hypertrophic cardiomyopathy may be mimicked by the myocardium thickened with these abnormal products causing "pseudohypertrophy." Most of these diseases are diagnosed during childhood.

Fabry's disease results from a deficiency of the lysosomal enzyme alpha-galactosidase A caused by one of more than 160 mutations. This disorder of glycosphingolipid metabolism is an X-linked recessive disorder that may also cause clinical disease in female carriers. Glycolipid accumulation may be limited to the cardiac tissues or may also involve the skin and kidney. Electron microscopy of endomyocardial biopsy tissue shows diagnostic vesicles containing concentric lamellar figures (**Fig. 238-13**). Diagnosis is crucial because enzyme replacement can reduce abnormal deposits and improve cardiac and clinical function. Enzyme replacement can also improve the course of Gaucher's disease, in which cerebroside-rich cells accumulate in multiple organs due to a deficiency of beta-glucosidase. Cerebroside-rich cells infiltrate the heart, which can also lead to a hemorrhagic pericardial effusion

and valvular disease.

Figure 238-13

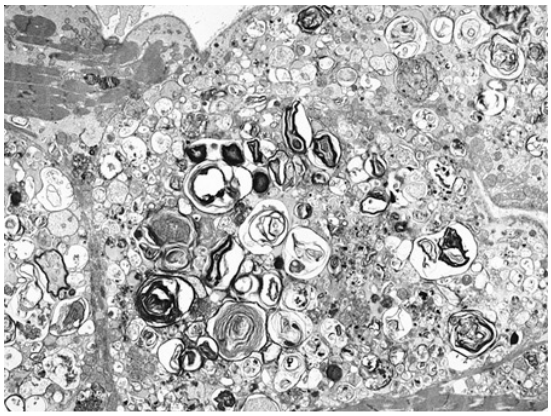

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fabry's disease. Transmission electron micrograph of a right ventricular endomyocardial biopsy specimen at high magnification showing the characteristic concentric lamellar inclusions of glycosphingolipids accumulating as a result of deficiency of the lysosomal enzyme alpha-galactosidase A. Image taken at 15,000x original magnification. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Glycogen storage diseases lead to accumulation of lysosomal storage products and intracellular glycogen accumulation, particularly with *glycogen storage disease type III,* due to a defective debranching enzyme. There are more than 10 types of *mucopolysaccharidoses,* in which autosomal dominant or X-linked deficiencies of lysosomal enzymes lead to the accumulation of glycosaminoglycans in the skeleton, nervous system, and heart. With characteristic facies, short stature, and frequent cognitive impairment, most individuals are diagnosed early in childhood and die before adulthood.

Carnitine is an essential cofactor in long-chain fatty acid metabolism. Multiple defects have been described that lead to carnitine deficiency, causing intracellular lipid inclusions and restrictive or dilated cardiomyopathy, often presenting in children. Fatty acid oxidation requires many metabolic steps with specific enzymes that can be deficient, with complex interactions with carnitine. Depending on the defect, cardiac and skeletal myopathy can be ameliorated with replacement of fatty acid intermediates and carnitine.

Two monogenic metabolic cardiomyopathies have recently been described as causes of increased ventricular wall thickness without an increase of muscle subunits or an increase in contractility. Mutations in the gamma-2 regulatory subunit of the adenosine monophosphate (AMP)-activated protein kinase important for glucose metabolism (PRKAG2) have been associated with a high prevalence of conduction abnormalities, such as AV block and ventricular preexcitation (Wolff-Parkinson-White syndrome). Several defects have been reported in an X-linked lysosome-associated membrane protein (LAMP2). This defect can be maternally transmitted or sporadic and has occasionally been isolated to the heart, although it often leads to a syndrome of skeletal myopathy, mental retardation, and hepatic dysfunction referred to as *Danon's disease.* Left ventricular hypertrophy appears early, often in childhood, and can progress rapidly to end-stage heart failure with low ejection fraction. Electron microscopy of these metabolic disorders shows that the myocytes are enlarged by multiple intracellular vacuoles of metabolic by-products.

Fibrotic Restrictive Cardiomyopathy

Progressive fibrosis can cause restrictive myocardial disease without dilation. Thoracic radiation, common for breast and lung cancer or mediastinal lymphoma, can produce early or late restrictive cardiomyopathy. Patients with *radiation cardiomyopathy* may present with a possible diagnosis of constrictive pericarditis, as the two conditions often coexist. Careful hemodynamic evaluation and, often, endomyocardial biopsy should be performed if considering pericardial stripping surgery, which is unlikely to be successful in the presence of underlying restrictive cardiomyopathy.

Scleroderma causes small vessel spasm and ischemia that can lead to a small, stiff heart with reduced ejection fraction without dilation. Doxorubicin causes direct myocyte injury usually leading to dilated cardiomyopathy, but the limited degree of dilation may result from increased fibrosis, which restricts remodeling.

Endomyocardial Disease

The physiologic picture of elevated filling pressures with atrial enlargement and preserved ventricular contractility with normal or reduced ventricular volumes can also result from extensive fibrosis of the endocardium, without transmural myocardial disease. For patients who have not lived in the equatorial regions, this picture is rare, and when seen is usually associated with a history of chronic hypereosinophilic syndrome (*Löffler's endocarditis*), which is more common in men than women. In this disease, persistent hyper-eosinophilia of >1500 eos/mm^3 for at least 6 months can cause an acute phase of eosinophilic injury in the endocardium, with systemic illness and injury to other organs. There is usually no obvious cause, but the hypereosinophilia can occasionally be explained by allergic, parasitic, or malignant disease. It is postulated to be followed by a period in which cardiac inflammation is replaced by evidence of fibrosis with superimposed thrombosis. In severe disease, the dense fibrotic layer can obliterate the ventricular apices and extend to thicken and tether the atrioventricular valve leaflets. The clinical disease may present with heart failure, embolic events, and atrial arrhythmias. While plausible, the sequence of transition has not been clearly demonstrated.

In tropical countries, up to one-quarter of heart failure may be due to *endomyocardial fibrosis,* affecting either or both ventricles. This condition shares with the previous condition the partial obliteration of the ventricular apex with fibrosis extending into the valvular inflow tract and leaflets; however, it is not clear that the etiologies are the same for all cases. Pericardial effusions frequently accompany endomyocardial fibrosis but are not common in Löffler's endocarditis. For endomyocardial fibrosis, there is no gender difference, but a higher prevalence in African American populations. While tropical endomyocardial fibrosis could represent the end-stage of previous hypereosinophilic disease triggered by endemic parasites, neither prior parasitic infection nor hypereosinophilia is usually documented. Geographic nutritional deficiencies have also been proposed as an etiology.

Medical treatment focuses on glucocorticoids and chemotherapy to suppress hypereosinophilia when present. Fluid retention may become increasingly resistant to diuretic therapy. Anticoagulation is recommended. Atrial fibrillation is associated with worse symptoms and prognosis, but may be difficult to suppress. Surgical resection of the apices and replacement of the fibrotic valves can improve symptoms, but surgical morbidity and mortality and later recurrence rates are high.

The serotonin secreted by *carcinoid* tumors can produce fibrous plaques in the endocardium and right-sided cardiac valves,

occasionally affecting left-sided valves, as well. Valvular lesions may be stenotic or regurgitant. Systemic symptoms include flushing and diarrhea. Liver disease from hepatic metastases may play a role by limiting hepatic function and thereby allowing more serotonin to reach the venous circulation.

HYPERTROPHIC CARDIOMYOPATHY

Hypertrophic cardiomyopathy is characterized by marked left ventricular hypertrophy in the absence of other causes, such as hypertension or valve disease (**Figs. 238-14 and 238-15**). The systolic function as measured by ejection fraction is often supranormal, at times with virtual obliteration of the left ventricular cavity during systole. The hypertrophy may be asymmetric, involving the septum more than the free wall of the ventricle. Approximately one-third of symptomatic patients demonstrate a resting intraventricular gradient that impedes outflow during systole and is exacerbated by increased contractility. This was previously termed *hypertrophic obstructive cardiomyopathy* (HOCM), as distinguished from *nonobstructive hypertrophic cardiomyopathy*. Other terms that have been used include *asymmetric septal hypertrophy* (ASH) and *idiopathic hypertrophic subaortic stenosis* (IHSS). However, the accepted terminology is now hypertrophic cardiomyopathy with or without an obstructive gradient. Classically, the microscopic picture shows marked disarray of individual fibers in a characteristic whorled pattern and disarray, also at the level of the larger bundles, interspersed with fibrosis (**Fig. 238-16**).

Figure 238-14

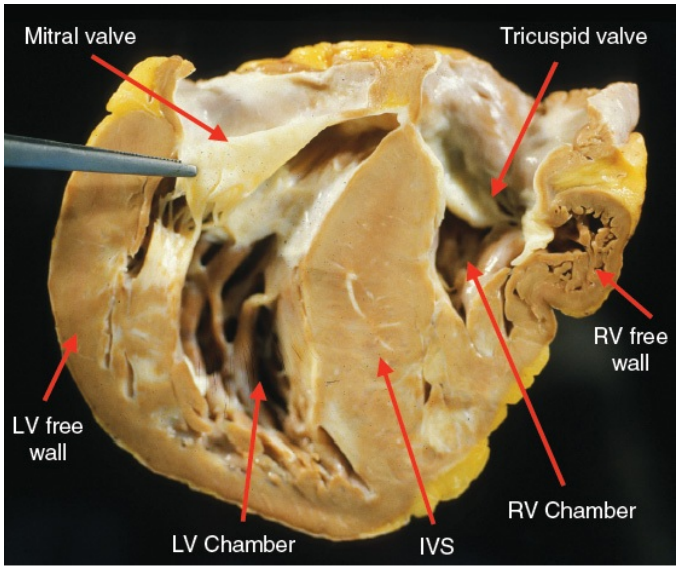

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypertrophic cardiomyopathy. Gross specimen of a heart with hypertrophic cardiomyopathy removed at the time of transplantation, showing asymmetric septal hypertrophy (septum much thicker than left ventricular free wall) with the septum bulging into the left ventricular outflow tract causing obstruction. The forceps are retracting the anterior leaflet of the mitral valve, demonstrating the characteristic plaque of systolic anterior motion, manifest as endocardial fibrosis on the interventricular septum in a mirror-image pattern to the valve leaflet. There is patchy replacement fibrosis, and small thick walled arterioles can be appreciated grossly, especially in the interventricular septum. IVS, interventricular septum; LV, left ventricle; RV, right ventricle. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

Figure 238-15

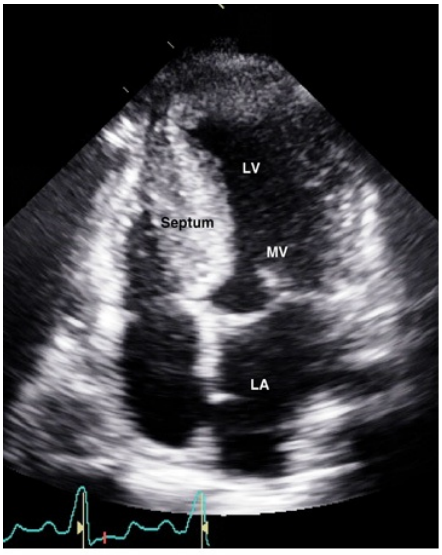

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypertrophic cardiomyopathy. This echocardiogram of hypertrophic cardiomyopathy shows asymmetric hypertrophy of the septum compared to the lateral wall of the left ventricle (LV). The mitral valve is moving anteriorly toward the hypertrophied septum

in systole. The left atrium (LA) is enlarged. Note that the echocardiographic and pathologic images are vertically opposite, such that the LV is by convention on the top right in the echocardiographic image and bottom right in the pathologic images. (*Image courtesy of Justina Wu, MD, Brigham and Women's Hospital, Boston.*)

Figure 238-16

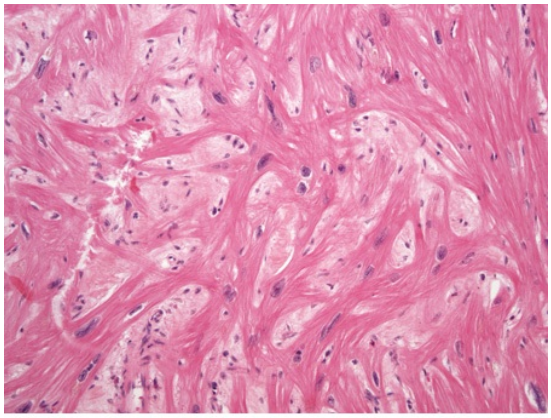

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypertrophic cardiomyopathy. Microscopic image of hypertrophic cardiomyopathy showing the characteristic disordered myocyte architecture with swirling and branching rather than the usual parallel arrangement of myocyte fibers. Myocyte nuclei vary markedly in size and interstitial fibrosis is present. (*Image courtesy of Robert Padera, MD, PhD, Department of Pathology, Brigham and Women's Hospital, Boston.*)

The prevalence of hypertrophic cardiomyopathy is 1:500 adults. Approximately one-half of these cases occur in a recognizable autosomal dominant pattern, and spontaneous mutations also arise. This is the best characterized genetic cardiomyopathy, for which more than 400 individual mutations have been identified in 11 sarcomeric genes. More than 80% of the mutations are in the beta-myosin heavy chain, the cardiac myosin-binding protein C, or cardiac troponin T. Some families may demonstrate a higher incidence of early progression to end-stage heart failure or death, suggesting that their mutations are more "malignant." However, the heterogeneity of phenotypic expression within and between families confirms the influence of modifying factors from other genes and the environment.

Hypertrophic cardiomyopathy is characterized hemodynamically by diastolic dysfunction, originally attributed to the hypertrophy, fibrosis, and intraventricular gradient when present. However, studies of asymptomatic family members indicate that diastolic dysfunction is a more fundamental abnormality that can precede evidence of hypertrophy. Resting ejection fraction and cardiac output are usually normal, but peak cardiac output during exercise may be reduced due to inadequate ventricular filling at high heart rates.

Diagnosis

Hypertrophic cardiomyopathy usually presents between the ages of 20 and 40 years. Dyspnea on exertion is the most common presenting symptom, reflecting elevated intracardiac filling pressures. Chest pain with either an atypical or typical exertional pattern occurs in more than half of symptomatic patients and is attributed to myocardial ischemia from high demand and abnormal intramural coronary arteries in the hypertrophied myocardium. Palpitations may result from atrial fibrillation or ventricular arrhythmias. Much less common are episodes of presyncope or syncope, often related to heavy exertion. Of grave concern is the possibility that the first manifestation of disease may be sudden death from ventricular tachycardia or fibrillation. Hypertrophic cardiomyopathy is the most common lesion found at autopsy of young athletes dying suddenly.

The physical examination typically reveals a harsh murmur heard best at the left lower sternal border, arising from both the outflow tract turbulence during ventricular ejection and the commonly associated mitral regurgitation. The gradient and the murmur may be enhanced by maneuvers that decrease ventricular volume, such as the Valsalva maneuver, or standing after squatting. They may be decreased by increasing ventricular volume or vascular resistance, such as with squatting or handgrip. A fourth heart sound is commonly heard due to decreased ventricular compliance. In patients with a significant outflow tract gradient, palpation of the carotid pulse may reveal a bifid systolic impulse, from early and delayed ejection. Patients with chronic, severe elevations in filling pressures may show signs of systemic fluid retention.

The electrocardiogram usually shows left ventricular hypertrophy, often with prominent septal Q waves that can be misdiagnosed as indicative of infarction. The diagnosis of hypertrophic cardiomyopathy is confirmed by echocardiography demonstrating left ventricular hypertrophy, which may or may not be more marked in the septum (Fig. 238-15). Intraventricular gradients to outflow can be identified by Doppler echocardiography at rest or during provocative maneuvers, such as the Valsalva maneuver. Systolic anterior motion (SAM) of the mitral valve is a classic finding on the echocardiogram. Mitral regurgitation may become severe. Cardiac catheterization can be performed to quantify the gradient, which characteristically increases after a premature ventricular contraction.

Apical hypertrophic cardiomyopathy is a variant that is uncommon in the United States; however, this variant accounts for about one-fourth of patients with hypertrophic cardiomyopathy in Japan. The electrocardiogram shows deep T-wave inversions in the precordial leads, and the echocardiogram shows a characteristic spade-like appearance with apical obliteration. It has been associated with a specific genetic defect in cardiac actin (Glu 101 Lys), but may occur with other sarcomere mutations.

The differential diagnosis of hypertrophic cardiomyopathy is limited in most patients once other cardiovascular causes for secondary hypertrophy are excluded. However, other diseases that result in thickened myocardium can appear indistinguishable on echocardiography, and are considered "pseudohypertrophic," particularly the inherited metabolic diseases (Table 238-4). The differential diagnosis between hypertrophic and restrictive cardiomyopathy may be particularly difficult when considering a diagnosis of "burned-out" hypertrophic cardiomyopathy in which systolic function has decreased. Overlap with infiltrative and restrictive myocardial diseases should be considered in the evaluation of increased left ventricular wall thickness on echocardiography, particularly when clinical features are atypical for classic hypertrophic cardiomyopathy. The metabolic defects in PRKAG2, alpha-galactosidase (Fabry's disease), and LAMP2 mutations (**Tables 238-3 and 238-4**) should routinely be considered during evaluation of apparent hypertrophic cardiomyopathy. With late onset without a family history of hypertrophic cardiomyopathy, amyloidosis should be carefully considered.

Treatment: Hypertrophic Cardiomyopathy

Therapy of hypertrophic cardiomyopathy is directed to symptom management and the prevention of sudden death (**Fig. 238-17**); it is not known whether treatment will decrease disease progression in asymptomatic family members. Exertional dyspnea and chest pain are treated by medication to reduce heart rate and ventricular contractility with hopes of improving diastolic filling patterns. Beta-adrenergic blocking drugs and verapamil are most commonly used as initial therapy. These agents both act to decrease heart rate and increase the length of time for diastolic filling, as well as to decrease the inotropic state. If there is fluid retention, diuretic therapy will usually be necessary, but requires careful titration to avoid hypovolemia, particularly in the presence of a resting or inducible obstruction to ventricular outflow. When symptoms persist and an outflow gradient is present, addition of disopyramide is sometimes effective. Amiodarone can also improve symptoms, but is usually initiated for control of arrhythmias rather than symptoms. Anticoagulation is recommended to prevent embolic events for patients who have had atrial fibrillation.

Figure 238-17

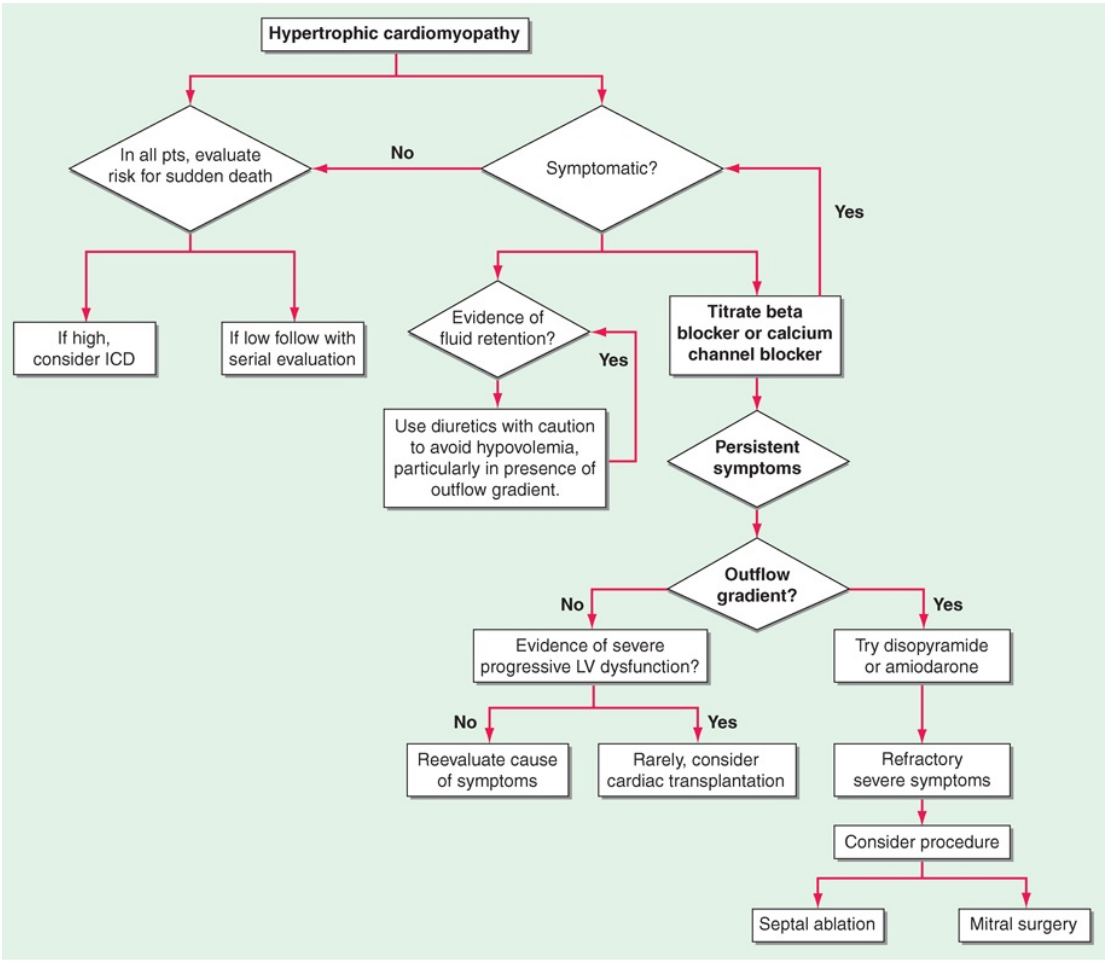

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment algorithm for hypertrophic cardiomyopathy depending on the presence and severity of symptoms, and the presence of an intraventricular gradient with obstruction to outflow. Note that all patients with hypertrophic cardiomyopathy should be evaluated for risk of sudden death, whether or not they require treatment for symptoms. ICD, implantable cardioverter-defibrillator; LV, left ventricular.

Symptoms that limit routine daily life despite adjustment of medical therapies develop in fewer than 5–10% of patients, generally those with substantial obstruction to ventricular outflow. Further therapies are directed to reduce this obstruction by changing ventricular mechanics. Most of the reported symptom improvement with dual chamber pacing is now attributed to placebo effect. Cardiac surgery can be directed to reduce the size of the upper septum that contributes to the obstruction (myomectomy), which generally also dominates the contribution of the anterior mitral leaflet to outflow tract obstruction. Reduction of the septum has also been accomplished by a catheter-based procedure in which ethanol is injected into a septal artery to cause a controlled septal infarction. The purpose of these interventions is to improve symptoms; none has been shown to prolong survival.

Cardiac transplantation is considered in fewer than 5% of patients with hypertrophic cardiomyopathy. It is rarely necessary to control symptoms in patients with preserved contractility and is more often considered for patients who progress to "burned-out" cardiomyopathy with decreased ejection fraction.

THERAPY TO PREVENT SUDDEN DEATH IN HYPERTROPHIC CARDIOMYOPATHY

Originally reported as 3–4% in referral populations, the annual incidence of sudden death in less selected populations of hypertrophic cardiomyopathy is approximately 1%. This outcome is attributed primarily to ventricular tachyarrhythmias, for which a myocardium with abnormal disorganized myocytes with patchy fibrosis is at exceptional risk, possibly further compromised by impaired coronary perfusion and aggravated by sudden increases in wall stress. Sudden death may precede the diagnosis of the disease, as in the death of young athletes. The highest risk is in patients with previous sustained ventricular tachyarrhythmias, a family history of sudden death, or, in cases of common mutations, a genetic mutation associated with sudden death (although the association of specific genotypes with the risk of sudden death remains controversial). Septal wall thickness >30 mm, recurrent syncope, exercise-induced hypotension, and nonsustained ventricular tachycardia are also risk factors. Areas of ventricular fibrosis detected on MRI may further identify susceptibility to life-threatening arrhythmias. Implantable cardioverter-defibrillators should be considered for those at highest risk (**Table 238-7**). Although low-risk patients can engage in regular physical activity with casual intent, it has been recommended that all patients with hypertrophic cardiomyopathy avoid intense training and competition.

Table 238-7 Risk Factors for Sudden Death in Hypertrophic Cardiomyopathy

Major Risk Factor		Screening Technique
History of cardiac arrest or spontaneous sustained ventricular tachycardia		History
Syncope	Usually with or after exertion	History
Family history of sudden cardiac death	Or possibly with a documented gene mutation associated with high risk	Family history
Spontaneous nonsustained ventricular tachycardia	>3 beats at rate >120	Exercise or 24–48 hour ambulatory recording
LV thickness >30 mm	Present in about 10% of patients, but many sudden deaths occur with wall thickness <30 mm	Echocardiography
Abnormal blood pressure response to exercise	Systolic blood pressure fall or failure to increase at peak exercise	Maximal upright exercise testing

Abbreviation: LV, left ventricle.

FURTHER READINGS

Baughman KL, Wynne J: Myocarditis, in *Braunwald's Heart Disease,* DP Zipes et al (eds). Philadelphia, Elsevier, 2005

Jessup M et al: 2009 focused update: ACCF/AHA Guidelines for the Diagnosis and Management of Heart Failure in Adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: Developed in collaboration with the International Society for Heart and Lung Transplantation. Circulation 119:1977, 2009[PMID: 19324967]

Maron BJ et al: American College of Cardiology/European Society of Cardiology Clinical Expert Consensus Document on Hypertrophic Cardiomyopathy. J Am Coll Cardiol 42:1688, 2003

———: Contemporary definitions and classifications of the cardiomyopathies. Circulation 113:1807, 2006

Rassi A Jr et al: Challenges and opportunities for primary, secondary, and tertiary prevention of Chagas' disease. Heart 95:524, 2009[PMID: 19131444]

Towbin JA, Vatta M: Genetics and genomics of dilated cardiomyopathy, in *Cardiovascular Genetics and Genomics,* American Heart Association Clinical Series, D Rodin (ed). Oxford, Wiley-Blackwell, 2009,

Yeh ET, Bickford CL: Cardiovascular complications of cancer therapy: Incidence, pathogenesis, diagnosis, and management. J Am Coll Cardiol 53:2231, 2009[PMID: 19520246]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PERICARDIAL DISEASE: INTRODUCTION

Normal Functions of the Pericardium

The normal pericardium is a double-layered sac; the visceral pericardium is a serous membrane that is separated by a small quantity (15–50 mL) of fluid, an ultrafiltrate of plasma, from the fibrous parietal pericardium. The normal pericardium, by exerting a restraining force, prevents sudden dilation of the cardiac chambers, especially the right atrium and ventricle, during exercise and with hypervolemia. It also restricts the anatomic position of the heart, minimizes friction between the heart and surrounding structures, prevents displacement of the heart and kinking of the great vessels, and probably retards the spread of infections from the lungs and pleural cavities to the heart. Nevertheless, total absence of the pericardium, either congenital or after surgery, does not produce obvious clinical disease. In partial left pericardial defects, the main pulmonary artery and left atrium may bulge through the defect; very rarely, herniation and subsequent strangulation of the left atrium may cause sudden death.

ACUTE PERICARDITIS

Acute pericarditis, by far the most common pathologic process involving the pericardium, may be classified both clinically and etiologically **(Table 239-1)**. There are four principal diagnostic features:

1. *Chest pain* is an important but not invariable symptom in various forms of acute pericarditis (Chap. 12); it is usually present in the acute infectious types and in many of the forms presumed to be related to hypersensitivity or autoimmunity. Pain is often absent in slowly developing tuberculous, postirradiation, neoplastic, and uremic pericarditis. The pain of acute pericarditis is often severe, retrosternal and left precordial, and referred to the neck, arms, or left shoulder. Often the pain is pleuritic, consequent to accompanying pleural inflammation (i.e., sharp and aggravated by inspiration and coughing), but sometimes it is a steady, constricting pain that radiates into either arm or both arms and resembles that of myocardial ischemia; therefore, confusion with acute myocardial infarction (AMI) is common. Characteristically, however, pericardial pain may be relieved by sitting up and leaning forward and is intensified by lying supine. The differentiation of AMI from acute pericarditis becomes perplexing when, with acute pericarditis, serum biomarkers of myocardial damage such as creatine kinase and troponin rise, presumably because of concomitant involvement of the epicardium in the inflammatory process (an epi-myocarditis) with resulting myocyte necrosis. However, these elevations, if they occur, are quite modest given the extensive electrocardiographic ST-segment elevation in pericarditis. This dissociation is useful in differentiating between these conditions.
2. A *pericardial friction rub* is audible in about 85% of these patients, may have up to three components per cardiac cycle, is high-pitched, and is described as rasping, scratching, or grating (Chap. 227); it can be elicited sometimes when the diaphragm of the stethoscope is applied firmly to the chest wall at the left lower sternal border. It is heard most frequently at end expiration with the patient upright and leaning forward. The rub is often inconstant, and the loud to-and-fro leathery sound may disappear within a few hours, possibly to reappear on the next day. A pericardial rub is heard throughout the respiratory cycle, whereas a pleural rub disappears when respiration is suspended.
3. The *electrocardiogram* (ECG) in acute pericarditis without massive effusion usually displays changes secondary to acute subepicardial inflammation (**Fig. 239-1**). It typically evolves through four stages. In stage 1, there is widespread elevation of the ST segments, often with upward concavity, involving two or three standard limb leads and V_2 to V_6, with reciprocal depressions only in aVR and sometimes V_1, as well as depression of the PR segment below the TP segment reflecting atrial involvement. Usually there are no significant changes in QRS complexes. In stage 2, after several days, the ST segments return to normal, and only then, or even later, do the T waves become inverted (stage 3). Ultimately, weeks or months after the onset of acute pericarditis, the ECG returns to normal in stage 4. In contrast, in AMI, ST elevations are convex, and reciprocal depression is usually more prominent; QRS changes occur, particularly the development of Q waves, as well as notching and loss of R-wave amplitude, and T-wave inversions are usually seen within hours *before* the ST segments have become isoelectric. Sequential ECGs are useful in distinguishing acute pericarditis from AMI. In the latter, elevated ST segments return to normal within hours (Chaps. 244 and 245).

   Early repolarization is a normal variant and may also be associated with widespread ST-segment elevation, most prominent in left precordial leads. However, in this condition the T waves are usually tall and the ST/T ratio is <0.25; importantly, this ratio is higher in acute pericarditis.

4. *Pericardial effusion* is usually associated with pain and/or the ECG changes mentioned above, as well as with an enlargement of the cardiac silhouette. Pericardial effusion is especially important clinically when it develops within a relatively short time as it may lead to cardiac tamponade (see below). Differentiation from cardiac enlargement may be difficult on physical examination, but heart sounds may be fainter with pericardial effusion. The friction rub may disappear, and the apex impulse may vanish, but sometimes it remains palpable, albeit medial to the left border of cardiac dullness. The base of the left lung may be compressed by pericardial fluid, producing *Ewart's sign*, a patch of dullness and increased fremitus (and egophony) beneath the angle of the left scapula. The chest roentgenogram may show a "water bottle" configuration of the cardiac silhouette **(Fig. 239-2)** but may be normal.

Table 239-1 Classification of Pericarditis

Clinical Classification
I. Acute pericarditis (<6 weeks)
A. Fibrinous
B. Effusive (serous or sanguineous)
II. Subacute pericarditis (6 weeks to 6 months)
A. Effusive-constrictive
B. Constrictive
III. Chronic pericarditis (>6 months)
A. Constrictive
B. Effusive
C. Adhesive (nonconstrictive)
Etiologic Classification
I. Infectious pericarditis
A. Viral (coxsackievirus A and B, echovirus, mumps, adenovirus, hepatitis, HIV)
B. Pyogenic (pneumococcus, streptococcus, staphylococcus, *Neisseria, Legionella*)
C. Tuberculous
D. Fungal (histoplasmosis, coccidioidomycosis, *Candida*, blastomycosis)
E. Other infections (syphilitic, protozoal, parasitic)
II. Noninfectious pericarditis
A. Acute myocardial infarction
B. Uremia
C. Neoplasia
1. Primary tumors (benign or malignant, mesothelioma)
2. Tumors metastatic to pericardium (lung and breast cancer, lymphoma, leukemia)
D. Myxedema
E. Cholesterol
F. Chylopericardium
G. Trauma
1. Penetrating chest wall
2. Nonpenetrating
H. Aortic dissection (with leakage into pericardial sac)
I. Postirradiation
J. Familial Mediterranean fever
K. Familial pericarditis
1. Mulibrey nanism*
L. Acute idiopathic
M. Whipple's disease
N. Sarcoidosis
III. Pericarditis presumably related to hypersensitivity or autoimmunity
A. Rheumatic fever

B. Collagen vascular disease (systemic lupus erythematosus, rheumatoid arthritis, ankylosing spondylitis, scleroderma, acute rheumatic fever, granulomatosis with polyangiitis (Wegener's)

C. Drug-induced (e.g., procainamide, hydralazine, phenytoin, isoniazide, minoxidil, anticoagulants, methysergide)

D. Post-cardiac injury

1. Postmyocardial infarction (Dressler's syndrome)
2. Postpericardiotomy
3. Posttraumatic

*An autosomal recessive syndrome characterized by growth failure, muscle hypotonia, hepatomegaly, ocular changes, enlarged cerebral ventricles, mental retardation, ventricular hypertrophy, and chronic constrictive pericarditis.

Figure 239-1

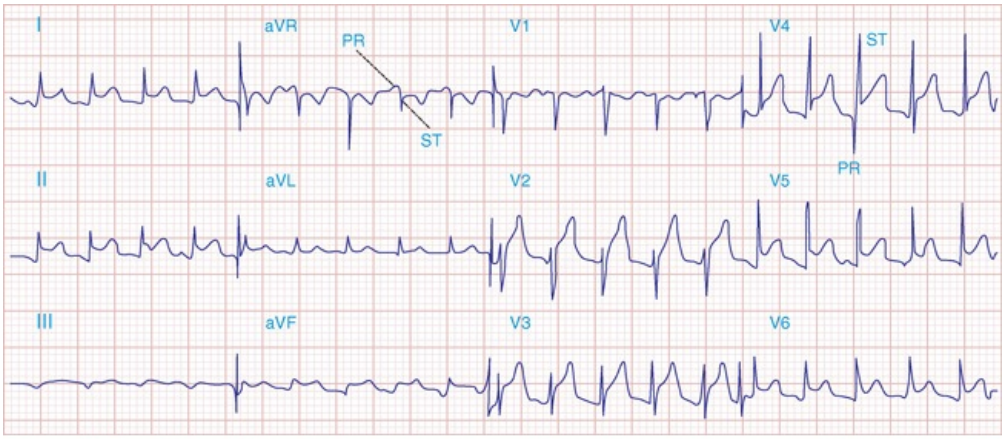

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute pericarditis often produces diffuse ST-segment elevations (in this case in leads I, II, aVF, and V_2 to V_6) due to a ventricular current of injury. Note also the characteristic PR-segment deviation (opposite in polarity to the ST segment) due to a concomitant atrial injury current.

Figure 239-2

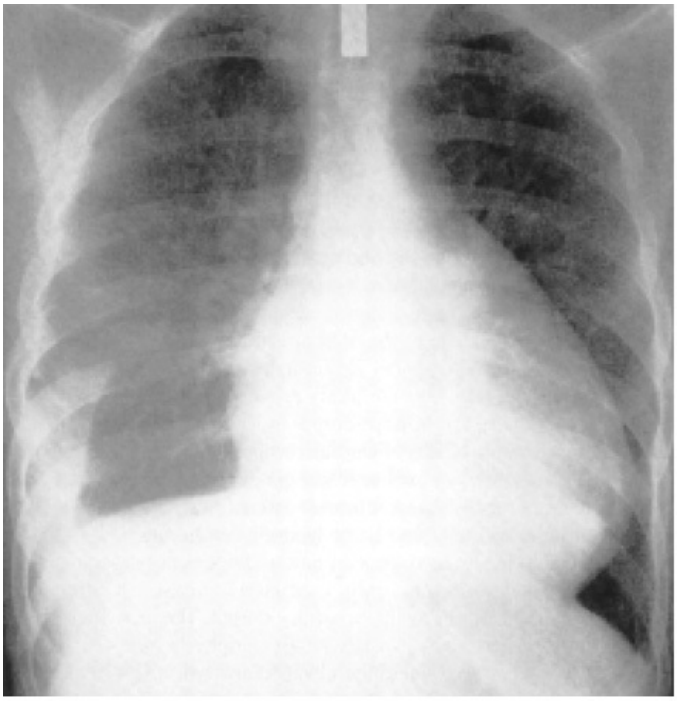

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiogram from a patient with a pericardial effusion showing typical "water bottle" heart. There is also a right pleural effusion. *[From SS Kabbani, M LeWinter, in MH Crawford et al (eds): Cardiology. London, Mosby, 2001.]*

DIAGNOSIS

Echocardiography (Chap. 229) is the most widely used imaging technique since it is sensitive, specific, simple, and noninvasive; may be performed at the bedside; and can identify accompanying cardiac tamponade (see below) **(Fig. 239-3)**. The presence of pericardial fluid is recorded by two-dimensional transthoracic echocardiography as a relatively echo-free space between the posterior pericardium and left ventricular epicardium in patients with small effusions and as a space between the anterior right ventricle and the parietal pericardium just beneath the anterior chest wall in those with larger effusions. In the latter, the heart may swing freely within the pericardial sac. When severe, the extent of this motion alternates and may be associated with electrical alternans. Echocardiography allows localization and estimation of the quantity of pericardial fluid.

Figure 239-3

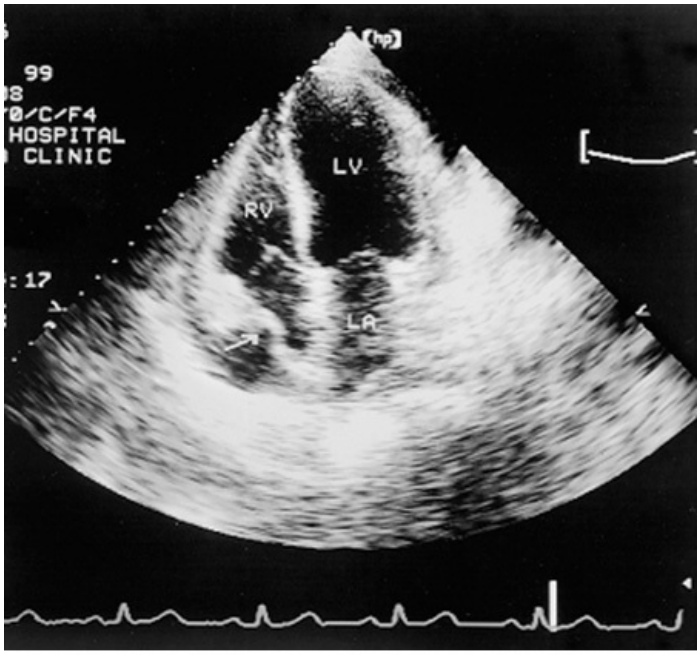

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Apical four-chamber echocardiogram recorded in a patient with a moderate pericardial effusion and evidence of hemodynamic compromise. The frame is recorded in early ventricular systole, immediately after atrial contraction. Note that the right atrial wall is indented inward and its curvature is frankly reversed (*arrow*), implying elevated intrapericardial pressure above right atrial pressure. LA, left atrium; LV, left ventricle; RV, right ventricle. *[From WF Armstrong: Echocardiography, in DP Zipes et al (eds): Braunwald's Heart Disease, 7th ed. Philadelphia, Elsevier, 2005.]*

The diagnosis of pericardial fluid or thickening may be confirmed by computed tomography (CT) or magnetic resonance imaging (MRI) **(Fig. 239-4)**. These techniques may be superior to echocardiography in detecting loculated pericardial effusions, pericardial thickening, and the presence of pericardial masses.

Figure 239-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic pericardial effusion in a 54-year-old female patient with Hodgkin's disease seen in

contrast-enhanced 64-slice CT. The arrows point at the pericardial effusion (LV, left ventricle; RV, right ventricle; RA, right atrium). Due to the timing of the scan relative to contrast injection, only the blood in the left ventricle is contrast-enhanced, hence, the low attenuation in the right-sided chambers. *[From Achenbach S, Daniel WG: Computed Tomography of the Heart, in P Libby et al (eds): Braunwald's Heart Disease, 8th ed. Philadelphia, Elsevier, 2008.]*

Cardiac Tamponade

The accumulation of fluid in the pericardial space in a quantity sufficient to cause serious obstruction to the inflow of blood to the ventricles results in cardiac tamponade. This complication may be fatal if it is not recognized and treated promptly. The three most common causes of tamponade are neoplastic disease, idiopathic pericarditis, and renal failure. Tamponade may also result from bleeding into the pericardial space after cardiac operations, trauma, and treatment of patients with acute pericarditis with anticoagulants.

The three principal features of tamponade (*Beck's triad*) are hypotension, soft or absent heart sounds, and jugular venous distention with a prominent *x* descent but an absent *y* descent. There are both limitation of ventricular filling and reduction of cardiac output. The quantity of fluid necessary to produce this critical state may be as small as 200 mL when the fluid develops rapidly or >2000 mL in slowly developing effusions when the pericardium has had the opportunity to stretch and adapt to an increasing volume. Tamponade may also develop more slowly, and in these circumstances the clinical manifestations may resemble those of heart failure, including dyspnea, orthopnea, and hepatic engorgement. A high index of suspicion for cardiac tamponade is required since in many instances no obvious cause for pericardial disease is apparent, and it should be considered in any patient with otherwise unexplained enlargement of the cardiac silhouette, hypotension, and elevation of jugular venous pressure. There may be reduction in amplitude of the QRS complexes, and *electrical alternans* of the P, QRS, or T waves should raise the suspicion of cardiac tamponade.

Table 239-2 lists the features that distinguish acute cardiac tamponade from constrictive pericarditis.

Table 239-2 Features That Distinguish Cardiac Tamponade from Constrictive Pericarditis and Similar Clinical Disorders

Characteristic	Tamponade	Constrictive Pericarditis	Restrictive Cardiomyopathy	RVMI
Clinical				
Pulsus paradoxus	Common	Usually absent	Rare	Rare
Jugular veins				
Prominent *y* descent	Absent	Usually present	Rare	Rare
Prominent *x* descent	Present	Usually present	Present	Rare
Kussmaul's sign	Absent	Present	Present	Present
Third heart sound	Absent	Absent	Rare	May be present
Pericardial knock	Absent	Often present	Absent	Absent
Electrocardiogram				
Low ECG voltage	May be present	May be present	May be present	Absent
Electrical alternans	May be present	Absent	Absent	Absent
Echocardiography				
Thickened pericardium	Absent	Present	Absent	Absent
Pericardial calcification	Absent	Often present	Absent	Absent
Pericardial effusion	Present	Absent	Absent	Absent
RV size	Usually small	Usually normal	Usually normal	Enlarged
Myocardial thickness	Normal	Normal	Usually increased	Normal
Right atrial collapse and RVDC	Present	Absent	Absent	Absent
Increased early filling, ↑ mitral flow velocity	Absent	Present	Present	May be present
Exaggerated respiratory variation in flow velocity	Present	Present	Absent	Absent
CT/MRI				
Thickened/calcific pericardium	Absent	Present	Absent	Absent
Cardiac catheterization				

Equalization of diastolic pressures	Usually present	Usually present	Usually absent	Absent or present
Cardiac biopsy helpful?	No	No	Sometimes	No

Abbreviations: ECG, electrocardiograph; RV, right ventricle; RVDC, right ventricular diastolic collapse; RVMI, right ventricular myocardial infarction.

Source: From GM Brockington et al: Cardiol Clin 8:645, 1990, with permission.

PARADOXICAL PULSE

This important clue to the presence of cardiac tamponade consists of a greater than normal (10 mmHg) inspiratory decline in systolic arterial pressure. When severe, it may be detected by palpating weakness or disappearance of the arterial pulse during inspiration, but usually sphygmomanometric measurement of systolic pressure during slow respiration is required.

Since both ventricles share a tight incompressible covering, i.e., the pericardial sac, the inspiratory enlargement of the right ventricle in cardiac tamponade compresses and reduces left ventricular volume; leftward bulging of the interventricular septum further reduces the left ventricular cavity as the right ventricle enlarges during inspiration. Thus, in cardiac tamponade the normal inspiratory augmentation of right ventricular volume causes an exaggerated reciprocal reduction in left ventricular volume. Also, respiratory distress increases the fluctuations in intrathoracic pressure, which exaggerates the mechanism just described. Right ventricular infarction (Chap. 245) may resemble cardiac tamponade with hypotension, elevated jugular venous pressure, an absent *y* descent in the jugular venous pulse, and, occasionally, pulsus paradoxus. The differences between these two conditions are shown in Table 239-2.

Paradoxical pulse occurs not only in cardiac tamponade but also in approximately one-third of patients with constrictive pericarditis (see below). This physical finding is not pathognomonic of pericardial disease because it may be observed in some cases of hypovolemic shock, acute and chronic obstructive airway disease, and pulmonary embolus.

Low-pressure tamponade refers to mild tamponade in which the intrapericardial pressure is increased from its slightly subatmospheric levels to +5 to +10 mmHg; in some instances, hypovolemia coexists. As a consequence, the central venous pressure is normal or only slightly elevated, whereas arterial pressure is unaffected and there is no paradoxical pulse. These patients are asymptomatic or complain of mild weakness and dyspnea. The diagnosis is aided by echocardiography, and both hemodynamic and clinical manifestations improve after pericardiocentesis.

DIAGNOSIS

Since immediate treatment of cardiac tamponade may be lifesaving, prompt measures to establish the diagnosis by echocardiography should be undertaken (Fig. 239-3). When pericardial effusion causes tamponade, Doppler ultrasound shows that tricuspid and pulmonic valve flow velocities increase markedly during inspiration, whereas pulmonic vein, mitral, and aortic flow velocities diminish. Often the right ventricular cavity is reduced in diameter, and there is late diastolic inward motion (collapse) of the right ventricular free wall and the right atrium. Transesophageal echocardiography may be necessary to diagnose a loculated or hemorrhagic effusion responsible for cardiac tamponade.

TREATMENT: CARDIAC TAMPONADE

Patients with acute pericarditis should be observed frequently for the development of an effusion; if a large effusion is present, the patient should be hospitalized and pericardiocentesis carried out or the patient should be watched closely for signs of tamponade. Arterial and venous pressures and heart rate should be monitored or followed carefully, and serial echocardiograms obtained.

Pericardiocentesis

If manifestations of tamponade appear, echocardiographically or fluoroscopically guided pericardiocentesis using an apical, parasternal, or, most commonly, subxiphoid approach must be carried out at once as reduction of the elevated intrapericardial pressure may be lifesaving. Intravenous saline may be administered as the patient is being readied for the procedure, but the pericardiocentesis must not be delayed. If possible, intrapericardial pressure should be measured before fluid is withdrawn, and the pericardial cavity should be drained as completely as possible. A small, multiholed catheter advanced over the needle inserted into the pericardial cavity may be left in place to allow draining of the pericardial space if fluid reaccumulates. Surgical drainage through a limited (subxiphoid) thoracotomy may be required in recurrent tamponade, when it is necessary to remove loculated effusions, and/or when it is necessary to obtain tissue for diagnosis.

Pericardial fluid obtained from an effusion often has the physical characteristics of an exudate. Bloody fluid is most commonly due to neoplasm in the United States and tuberculosis in developing nations but may also be found in the effusion of acute rheumatic fever, post-cardiac injury, and post-myocardial infarction, as well as in the pericarditis associated with renal failure or dialysis. Transudative pericardial effusions may occur in heart failure.

The pericardial fluid should be analyzed for red and white blood cells, and cytologic studies for cancer, microscopic studies, and cultures should be obtained. The presence of DNA of *Mycobacterium tuberculosis* determined by polymerase chain reaction or an elevated adenosine deaminase activity (>30 U/L) strongly supports the diagnosis of tuberculous pericarditis (Chap. 165).

Viral or Idiopathic Form of Acute Pericarditis

In many instances, acute pericarditis occurs in association with illnesses of known or presumed viral origin and probably is caused by the same agent. Commonly, there is an antecedent infection of the respiratory tract, and viral isolation and serologic studies are negative. In some cases, coxsackievirus A or B or the virus of influenza, echovirus, mumps, herpes simplex, chickenpox, adenovirus, cytomegalovirus, Epstein-Barr, or HIV has been isolated from pericardial fluid and/or appropriate elevations in viral antibody titers have been noted. Pericardial effusion is a common cardiac manifestation of HIV; it is usually secondary to infection (often mycobacterial) or neoplasm, most frequently lymphoma. Most frequently, a viral causation cannot be established; the term *idiopathic acute pericarditis* is then appropriate. Viral or idiopathic acute pericarditis occurs at all ages but is more common in young adults and is often associated with pleural effusions and pneumonitis. The almost simultaneous development of fever and precordial pain, often 10 to 12 days after a presumed viral illness, constitutes an important feature in the differentiation of acute pericarditis from AMI, in which chest pain precedes fever. The constitutional symptoms are usually mild to moderate, and a pericardial friction rub is often audible. The disease ordinarily runs its course in a few days to 4 weeks. The ST-segment alterations in the ECG usually disappear after 1 or more weeks, but the abnormal T waves may persist for several years and be a source of confusion in persons without a clear history of pericarditis.

Pleuritis and pneumonitis frequently accompany pericarditis. Accumulation of some pericardial fluid is common, and both tamponade and constrictive pericarditis are possible complications. Recurrent (relapsing) pericarditis occurs in about one-fourth of patients with acute idiopathic pericarditis. In a smaller number, there are multiple recurrences.

TREATMENT: IDIOPATHIC ACUTE PERICARDITIS

In acute idiopathic pericarditis there is no specific therapy, but bed rest and anti-inflammatory treatment with aspirin (2–4 g/d) may be given. If this is ineffective, one of the nonsteroidal anti-inflammatory drugs (NSAIDs), such as ibuprofen (400–600 mg tid), indomethacin (25–50 mg tid), or colchicine (0.6 mg bid), is often effective. Glucocorticoids (e.g., prednisone, 40–80 mg daily) usually suppress the clinical manifestations of the acute illness and may be useful in patients in whom purulent bacterial pericarditis has been excluded and in patients with pericarditis secondary to connective tissue disorders and renal failure (see below). Anticoagulants should be avoided since their use could cause bleeding into the pericardial cavity and tamponade.

After the patient has been asymptomatic and afebrile for about a week, the dose of the NSAID may be tapered gradually. Colchicine may prevent recurrences, but when recurrences are multiple, frequent, and disabling; continued beyond 2 years; and are not controlled by glucocorticoids, pericardiectomy may be necessary to terminate the illness.

POSTCARDIAC INJURY SYNDROME

Acute pericarditis may appear in a variety of circumstances that have one common feature: previous injury to the myocardium with blood in the pericardial cavity. The syndrome may develop after a cardiac operation (postpericardiotomy syndrome), after blunt or penetrating cardiac trauma (Chap. 240), or after perforation of the heart with a catheter. Rarely, it follows AMI.

The clinical picture mimics acute viral or idiopathic pericarditis. The principal symptom is the pain of acute pericarditis, which usually develops 1 to 4 weeks after the cardiac injury (1 to 3 days after AMI) but sometimes appears only after an interval of months. Recurrences are common and may occur up to 2 years or more after the injury. Pericarditis, fever with temperature up to 39°C (102.2°F), pleuritis, and pneumonitis are the outstanding features, and the bout of illness usually subsides in 1 or 2 weeks. The pericarditis may be of the fibrinous variety, or it may be a pericardial effusion, which is often serosanguineous but rarely causes tamponade. Leukocytosis, an increased sedimentation rate, and ECG changes typical of acute pericarditis may also occur.

This syndrome is probably the result of a hypersensitivity reaction to antigen that originates from injured myocardial tissue and/or pericardium. Circulating myocardial antisarcolemmal and antifibrillar autoantibodies occur frequently, but their precise role in the development of this syndrome has not been defined. Viral infection may also play an etiologic role, since antiviral antibodies are often elevated in patients who develop this syndrome after cardiac surgery.

Often no treatment is necessary aside from aspirin and analgesics. When the illness is followed by a series of disabling recurrences, therapy with an NSAID, colchicine, or a glucocorticoid is usually effective.

Differential Diagnosis

Since there is no specific test for *acute idiopathic pericarditis*, the diagnosis is one of exclusion. Consequently, all other disorders that may be associated with acute fibrinous pericarditis must be considered. A common diagnostic error is mistaking acute viral or idiopathic pericarditis for AMI and vice versa. When acute fibrinous pericarditis is associated with AMI (Chap. 245), it is characterized by fever, pain, and a friction rub in the first 4 days after the development of the infarct. ECG abnormalities (such as the appearance of Q waves, brief ST-segment elevations with reciprocal changes, and earlier T-wave changes in AMI) and the extent of the elevations of myocardial enzymes are helpful in differentiating pericarditis from AMI.

Pericarditis secondary to postcardiac injury is differentiated from acute idiopathic pericarditis chiefly by timing. If it occurs within a few days or weeks of an AMI, a chest blow, a cardiac perforation, or a cardiac operation, it may be justified to conclude that the two are probably related.

It is important to distinguish *pericarditis due to collagen vascular disease* from acute idiopathic pericarditis. Most important in the differential diagnosis is the pericarditis due to systemic lupus erythematosus (SLE; Chap. 319) or drug-induced (procainamide or hydralazine) lupus. When pericarditis occurs in the absence of any obvious underlying disorder, the diagnosis of SLE may be suggested by a rise in the titer of antinuclear antibodies. Acute pericarditis is an occasional complication of *rheumatoid arthritis*, *scleroderma*, and *polyarteritis nodosa*, and other evidence of these diseases is usually obvious. Asymptomatic pericardial effusion is also common in these disorders. The pericarditis of *acute rheumatic fever* is generally associated with evidence of severe pancarditis and with cardiac murmurs (Chap. 322).

Pyogenic (purulent) pericarditis is usually secondary to cardiothoracic operations, by extension of infection from the lungs or pleural cavities, from rupture of the esophagus into the pericardial sac, or from rupture of a ring abscess in a patient with infective endocarditis, or it can occur if septicemia complicates aseptic pericarditis. It is usually accompanied by fever, chills, septicemia, and evidence of infection elsewhere and generally has a poor prognosis. The diagnosis is made by examination of the pericardial fluid. Acute pericarditis may also complicate the viral, pyogenic, mycobacterial, and fungal infections that occur with HIV infection.

Pericarditis of renal failure (Chap. 280) occurs in up to one-third of patients with chronic uremia *(uremic pericarditis)*, is also seen in patients undergoing chronic dialysis with normal levels of blood urea and creatinine, and is termed *dialysis-associated pericarditis*. These two forms of pericarditis may be fibrinous and are generally associated with an effusion that may be sanguineous. A pericardial friction rub is common, but pain is usually absent or mild. Treatment with an NSAID and intensification of dialysis are usually adequate. Occasionally, tamponade occurs and pericardiocentesis is required. When the pericarditis of renal failure is recurrent or persistent, a pericardial window should be created or pericardiectomy may be necessary.

Pericarditis due to *neoplastic diseases* results from extension or invasion of metastatic tumors (most commonly carcinoma of the lung and breast, malignant melanoma, lymphoma, and leukemia) to the pericardium; pain, atrial arrhythmias, and tamponade are complications that occur occasionally. Diagnosis is made by pericardial fluid cytology or pericardial biopsy. *Mediastinal irradiation* for neoplasm may cause acute pericarditis and/or chronic constrictive pericarditis. Unusual causes of acute pericarditis include syphilis, fungal infection (histoplasmosis, blastomycosis, aspergillosis, and candidiasis), and parasitic infestation (amebiasis, toxoplasmosis, echinococcosis, trichinosis).

Chronic Pericardial Effusions

Chronic pericardial effusions are sometimes encountered in patients without an antecedent history of acute pericarditis. They may cause few symptoms per se, and their presence may be detected by finding an enlarged cardiac silhouette on chest roentgenogram. Tuberculosis is a common cause (Chap. 165).

OTHER CAUSES

Myxedema may be responsible for chronic pericardial effusion that is sometimes massive but rarely, if ever, causes cardiac tamponade. The cardiac silhouette is markedly enlarged, and an echocardiogram distinguishes cardiomegaly from pericardial effusion. The diagnosis of myxedema can be confirmed by tests for thyroid function (Chap. 341). Myxedematous pericardial effusion responds to thyroid hormone replacement.

Neoplasms, systemic lupus erythematosus (SLE), rheumatoid arthritis, mycotic infections, radiation therapy to the chest, pyogenic infections, and chylopericardium may also cause chronic pericardial effusion and should be considered and specifically sought in such patients.

Aspiration and analysis of the pericardial fluid are often helpful in diagnosis. Pericardial fluid should be analyzed as described on p. 1974. Grossly sanguineous pericardial fluid results most commonly from a neoplasm, tuberculosis, renal failure, or slow leakage from an aortic aneurysm. Pericardiocentesis may resolve large effusions, but pericardiectomy may be required with recurrence. Intrapericardial instillation of sclerosing agents or antineoplastic agents may be used to prevent reaccumulation of fluid.

CHRONIC CONSTRICTIVE PERICARDITIS

This disorder results when the healing of an acute fibrinous or serofibrinous pericarditis or the resorption of a chronic pericardial effusion is followed by obliteration of the pericardial cavity with the formation of granulation tissue. The latter gradually contracts and forms a firm scar, which may be calcified, encasing the heart and interfering with filling of the ventricles. In developing nations where the condition is prevalent, a high percentage of cases are of tuberculous origin, but this is now an uncommon cause in North America. Chronic constrictive pericarditis may follow acute or relapsing viral or idiopathic pericarditis, trauma with organized blood clot, cardiac surgery of any type, mediastinal irradiation, purulent infection, histoplasmosis, neoplastic disease (especially breast cancer, lung cancer, and lymphoma), rheumatoid arthritis, SLE, and chronic renal failure with uremia treated by chronic dialysis. In many patients the cause of the pericardial disease is undetermined, and in them an asymptomatic or forgotten bout of viral pericarditis, acute or idiopathic, may have been the inciting event.

The basic physiologic abnormality in patients with chronic constrictive pericarditis is the inability of the ventricles to fill because of the limitations imposed by the rigid, thickened pericardium. In constrictive pericarditis, ventricular filling is unimpeded during early diastole but is reduced abruptly when the elastic limit of the pericardium is reached, whereas in cardiac tamponade, ventricular filling is impeded throughout diastole. In both conditions, ventricular end-diastolic and stroke volumes are reduced and the end-diastolic pressures in both ventricles and the mean pressures in the atria, pulmonary veins, and systemic veins are all elevated to similar levels (i.e., within 5 mmHg of one another). Despite these hemodynamic changes, myocardial function may be normal or only slightly impaired in chronic constrictive pericarditis. However, the fibrotic process may extend into the myocardium and cause myocardial scarring and atrophy, and venous congestion may then be due to the combined effects of the pericardial and myocardial lesions.

In constrictive pericarditis, the right and left atrial pressure pulses display an M-shaped contour, with prominent *x* and *y* descents. The *y* descent, which is absent or diminished in cardiac tamponade, is the most prominent deflection in constrictive pericarditis; it reflects rapid early filling of the ventricles. The *y* descent is interrupted by a rapid rise in atrial pressure during early diastole, when ventricular filling is impeded by the constricting pericardium. These characteristic changes are transmitted to the jugular veins, where they may be recognized by inspection. In constrictive pericarditis, the ventricular pressure pulses in both ventricles exhibit characteristic "square root" signs during diastole. These hemodynamic changes, although characteristic, are not pathognomonic of constrictive pericarditis and may also be observed in cardiomyopathies characterized by restriction of ventricular filling (Chap. 238) (Table 239-2).

Clinical and Laboratory Findings

Weakness, fatigue, weight gain, increased abdominal girth, abdominal discomfort, a protuberant abdomen, and edema are common. The patient often appears chronically ill, and in advanced cases there are anasarca, skeletal muscle wasting, and cachexia. Exertional dyspnea is common, and orthopnea may occur, although it is usually not severe. Acute left ventricular failure (acute pulmonary edema) is very uncommon. The cervical veins are distended and may remain so even after intensive diuretic treatment, and venous pressure may fail to decline during inspiration (*Kussmaul's sign*). The latter is common in chronic pericarditis but may also occur in tricuspid stenosis, right ventricular infarction, and restrictive cardiomyopathy.

The pulse pressure is normal or reduced. In about one-third of cases, a paradoxical pulse (p. 0000) can be detected. Congestive hepatomegaly is pronounced and may impair hepatic function and cause jaundice; ascites is common and is usually more prominent than dependent edema. The apical pulse is reduced and may retract in systole (*Broadbent's sign*). The heart sounds may be distant; an early third heart sound (i.e., a pericardial knock, occurring at the cardiac apex 0.09–0.12 s after aortic valve closure) is often conspicuous; it occurs with the abrupt cessation of ventricular filling. A systolic murmur of tricuspid regurgitation may be present.

The *ECG* frequently displays low voltage of the QRS complexes and diffuse flattening or inversion of the T waves. Atrial fibrillation is present in about one-third of patients. The *chest roentgenogram* shows a normal or slightly enlarged heart; pericardial calcification is most common in tuberculous pericarditis. Pericardial calcification may, however, occur in the absence of constriction.

Inasmuch as the usual physical signs of cardiac disease (murmurs, cardiac enlargement) may be inconspicuous or absent in chronic constrictive pericarditis, hepatic enlargement and dysfunction associated with jaundice and intractable ascites may lead to a mistaken diagnosis of hepatic cirrhosis. This error can be avoided if the neck veins are inspected carefully in patients with ascites and hepatomegaly. Given a clinical picture resembling hepatic cirrhosis, but with the added feature of distended neck veins, a careful search for thickening of the pericardium by imaging (see Fig. 229-6) should be carried out and may disclose this curable or remediable form of heart disease.

The transthoracic *echocardiogram* typically shows pericardial thickening, dilation of the inferior vena

cava and hepatic veins, and a sharp halt in ventricular filling in early diastole, with normal ventricular systolic function and flattening of the left ventricular posterior wall. Atrial enlargement may be seen, especially in patients with long-standing constrictive physiology. There is a distinctive pattern of transvalvular flow velocity on Doppler flow-velocity echocardiography. During inspiration there is an exaggerated reduction in blood flow velocity in the pulmonary veins and across the mitral valve and a leftward shift of the ventricular septum; the opposite occurs during expiration. Diastolic flow velocity in the vena cavae into the right atrium and across the tricuspid valve increases in an exaggerated manner during inspiration and declines during expiration **(Fig. 239-5)**. However, echocardiography cannot definitively exclude the diagnosis of constrictive pericarditis. MRI and CT scanning **(Fig. 239-6)** are more accurate than echocardiography in establishing or excluding the presence of a thickened pericardium. Pericardial thickening and even pericardial calcification, however, are not synonymous with constrictive pericarditis since they may occur without seriously impairing ventricular filling.

Figure 239-5

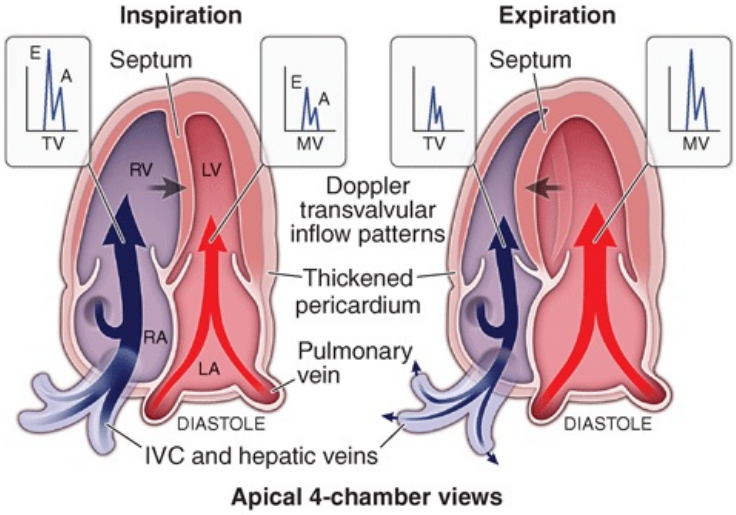

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Constrictive pericarditis. Doppler schema of respirophasic changes in mitral and tricuspid inflow. Reciprocal patterns of ventricular filling are assessed on pulsed Doppler examination of mitral valve (MV) and tricuspid valve (TV) inflow.

Figure 239-6

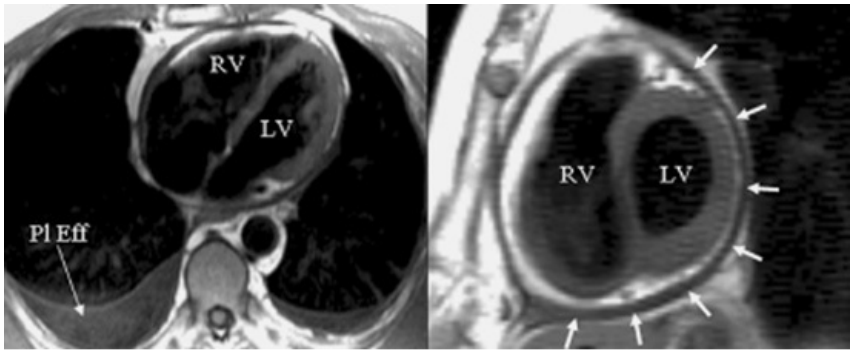

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cardiovascular magnetic resonance in a patient with constrictive pericarditis. On the right is a basal short-axis view of the ventricles showing a thickened pericardium encasing the heart (*arrows*). On the left is a transaxial view, again showing the thickened pericardium, particularly over the right heart, but also a pleural effusion (Pl Eff). LV, left ventricle; RV, right ventricle. *[From D Pennell: Cardiovascular Magnetic Resonance, in P Libby et al (eds): Braunwald's Heart Disease, 8th ed. Philadelphia, Elsevier, 2005.]*

Differential Diagnosis

Like chronic constrictive pericarditis, cor pulmonale (Chap. 234) may be associated with severe

systemic venous hypertension but little pulmonary congestion; the heart is usually not enlarged, and a paradoxical pulse may be present. However, in cor pulmonale, advanced parenchymal pulmonary disease is usually obvious and venous pressure *falls* during inspiration (i.e., Kussmaul's sign is negative). *Tricuspid stenosis* (Chap. 237) may also simulate chronic constrictive pericarditis; congestive hepatomegaly, splenomegaly, ascites, and venous distention may be equally prominent. However, in tricuspid stenosis, a characteristic murmur as well as the murmur of accompanying mitral stenosis is usually present.

Because constrictive pericarditis can be corrected surgically, it is important to distinguish chronic constrictive pericarditis from restrictive cardiomyopathy (Chap. 238), which has a similar physiologic abnormality (i.e., restriction of ventricular filling). In many patients with restrictive cardiomyopathy the ventricular wall is thickened as shown on echocardiographic examination (Table 239-2). The features favoring the diagnosis of restrictive cardiomyopathy over chronic constrictive pericarditis include a well-defined apex beat, cardiac enlargement, and pronounced orthopnea with attacks of acute left ventricular failure, left ventricular hypertrophy, gallop sounds (in place of a pericardial knock), bundle branch block, and, in some cases, abnormal Q waves on the ECG. The typical echocardiographic features of constrictive pericarditis (see above) are useful in the differential diagnosis in chronic constrictive pericarditis (Fig. 239-5). CT imaging (usually with contrast) and MRI are key in distinguishing between restrictive cardiomyopathy and chronic constrictive pericarditis. In the former, the ventricular walls are hypertrophied, whereas in the latter, the pericardium is thickened and sometimes calcified. When a patient has progressive, disabling, and unresponsive congestive heart failure and displays any of the features of constrictive heart disease, Doppler echocardiography to record respiratory effects on transvalvular flow and an MRI or CT scan should be obtained to detect or exclude constrictive pericarditis, since the latter is usually curable.

Treatment: Constrictive Pericarditis

Pericardial resection is the only definitive treatment of constrictive pericarditis and should be as complete as possible. Dietary sodium restriction and diuretics are useful during preoperative preparation. Coronary arteriography should be carried out preoperatively in patients older than 50 years of age to exclude unsuspected coronary artery disease. The benefits derived from cardiac decortication are usually progressive over a period of months. The risk of this operation depends on the extent of penetration of the myocardium by the fibrotic and calcific process, the severity of myocardial atrophy, the extent of secondary impairment of hepatic and/or renal function, and the patient's general condition. Operative mortality is in the range of 5 to 10%; the patients with the most severe disease are at highest risk. Therefore, surgical treatment should, if possible, be carried out relatively early in the course.

SUBACUTE EFFUSIVE-CONSTRICTIVE PERICARDITIS

This form of pericardial disease is characterized by the combination of a tense effusion in the pericardial space and constriction of the heart by thickened pericardium. It shares a number of features both with chronic pericardial effusion producing cardiac compression and with pericardial constriction. It may be caused by tuberculosis (see below), multiple attacks of acute idiopathic pericarditis, radiation, traumatic pericarditis, renal failure, scleroderma, and neoplasms. The heart is generally enlarged, and a paradoxical pulse and a prominent *x* descent (without a prominent *y* descent) are present in the atrial and jugular venous pressure pulses. After pericardiocentesis, the physiologic findings may change from those of cardiac tamponade to those of pericardial constriction. Furthermore, the intrapericardial pressure and the central venous pressure may decline, but not to normal. The diagnosis can be established by pericardiocentesis followed by pericardial biopsy. Wide excision of both the visceral and parietal pericardium is usually effective therapy.

TUBERCULOUS PERICARDIAL DISEASE

This chronic infection is a common cause of chronic pericardial effusion, although less so in the United States than in Africa, Asia, the Middle East, and other parts of the developing world where active tuberculosis is endemic (Chap. 165). The clinical picture is that of a chronic, systemic illness in a patient with pericardial effusion. It is important to consider this diagnosis in a patient with known tuberculosis, with HIV, and with fever, chest pain, weight loss, and enlargement of the cardiac silhouette of undetermined origin. If the etiology of chronic pericardial effusion remains obscure despite detailed analysis of the pericardial fluid (see above), a pericardial biopsy, preferably by a limited thoracotomy, should be performed. If definitive evidence is still lacking but the specimen shows granulomas with caseation, antituberculous chemotherapy (Chap. 165) is indicated.

If the biopsy specimen shows a thickened pericardium, pericardiectomy should be carried out to prevent the development of constriction. Tubercular cardiac constriction should be treated surgically while the patient is receiving antituberculous chemotherapy.

OTHER DISORDERS OF THE PERICARDIUM

Pericardial cysts appear as rounded or lobulated deformities of the cardiac silhouette, most commonly at the right cardiophrenic angle. They do not cause symptoms, and their major clinical significance lies

in the possibility of confusion with a tumor, ventricular aneurysm, or massive cardiomegaly. *Tumors* involving the pericardium are most commonly secondary to malignantneoplasms originating in or invading the mediastinum, including carcinoma of the bronchus and breast, lymphoma, and melanoma. The most common *primary* malignant tumor is the mesothelioma. The usual clinical picture of malignant pericardial tumor is an insidiously developing, often bloody pericardial effusion. Surgical exploration is required to establish a definitive diagnosis and to carry out definitive or, more commonly, palliative treatment.

FURTHER READINGS

Imazio M et al: Diagnosis and management of pericardial diseases. Nat Rev Cardiol 6:743, 2009[PMID: 19859068]

Khandaker MH et al: Pericardial disease: Diagnosis and managemnt. Mayo Clin Proc 85:572,2010[PMID: 20511488]

LeWinter M: Pericardial diseases, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

Maisch B et al: Guidelines on the diagnosis and management of pericardial diseases executive summary: The Task Force on the Diagnosis and Management of Pericardial Diseases of the European Society of Cardiology. Eur Heart J 25:587, 2004[PMID: 15120056]

Mayosi BM: Contemporary trends in the epidemiology and management of cardiomyopathy and pericarditis in sub-Saharan Africa. Heart 93:1176, 2007[PMID: 17890693]

——— et al: Tuberculous pericarditis. Circulation 112:3608, 2005

O'Leary SM et al: Imaging the pericardium: Appearances on ECG-gated 64-detector row cardiac computed tomography. Br J Radiol 83:194,2010[PMID: 21660345]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TUMORS OF THE HEART

Primary Tumors

Primary tumors of the heart are rare. Approximately three-quarters are histologically benign, and the majority of these tumors are myxomas. Malignant tumors, almost all of which are sarcomas, account for 25% of primary cardiac tumors **(Table 240-1)**. All cardiac tumors, regardless of pathologic type, have the potential to cause life-threatening complications. Many tumors are now surgically curable; thus, early diagnosis is imperative.

Table 240-1 Relative Incidence of Primary Tumors of the Heart

Type	Number	Percent
Benign	199	58.0
Myxoma	114	33.2
Rhabdomyoma	20	5.8
Fibroma	20	5.8
Hemangioma	17	5.0
Atrioventricular nodal	10	2.9
Granular cell	4	1.2
Lipoma	2	0.6
Paraganglioma	2	0.6
Myocytic hamartoma	2	0.6
Histiocytoid cardiomyopathy	2	0.6
Inflammatory psuedotumor	2	0.6
Other benign tumors	4	1.2
Malignant	144	42.0
Sarcoma	137	39.9
Lymphoma	7	2.1

Source: Modified from A Burke, R Virmani: *Atlas of Tumor Pathology: Tumors of the Heart and Great Vessels*. Washington, DC, Armed Forces Institute of Pathology 1996, p. 231; with permission.

CLINICAL PRESENTATION

Cardiac tumors may present with a wide array of cardiac and noncardiac manifestations. These manifestations depend in large part on the location and size of the tumor and are often nonspecific features of more common forms of heart disease, such as chest pain, syncope, heart failure, murmurs, arrhythmias, conduction disturbances, and pericardial effusion with or without tamponade. Additionally, embolic phenomena and constitutional symptoms may occur.

MYXOMA

Myxomas are the most common type of primary cardiac tumor in all age groups, accounting for one-third to one-half of all cases at postmortem and about three-quarters of the tumors treated surgically. They occur at all ages, most commonly in the third through sixth decades, with a female predilection. Approximately 90% of myxomas are sporadic; the remainder are familial with autosomal dominant transmission. The familial variety often occurs as part of a syndrome complex (Carney complex) that includes (1) myxomas (cardiac, skin, and/or breast), (2) lentigines and/or pigmented nevi, and (3) endocrine overactivity (primary nodular adrenal cortical disease with or without Cushing's syndrome, testicular tumors, and/or pituitary adenomas with gigantism or acromegaly). Certain constellations of findings have been referred to as the *NAME* syndrome (nevi, atrial myxoma, myxoid neurofibroma, and ephelides) or the *LAMB* syndrome (lentigines, atrial myxoma, and blue nevi), although these syndromes probably represent subsets of the Carney complex. The genetic basis of this complex has not been elucidated completely; however, patients frequently have inactivating mutations in the tumor-suppressor gene *PRKAR1A*, which encodes the protein kinase A type I-α regulatory subunit.

Pathologically, myxomas are gelatinous structures that consist of myxoma cells embedded in a stroma rich in glycosaminoglycans. Most are solitary, are located in the atria (particularly the left atrium, where they usually arise from the interatrial septum in the vicinity of the fossa ovalis), and are often pedunculated on a fibrovascular stalk. In contrast to sporadic tumors, familial or syndromic tumors tend to occur in younger individuals, are often multiple, may be ventricular in location, and are more likely to recur after initial resection.

Myxomas commonly present with obstructive signs and symptoms. The most common clinical presentation mimics that of mitral valve disease: either stenosis owing to tumor prolapse into the mitral orifice or regurgitation resulting from tumor-induced valvular trauma. Ventricular myxomas may cause outflow obstruction similar to that caused by subaortic or subpulmonic stenosis. The symptoms and signs of myxoma may be sudden in onset or positional in nature, owing to the effects of gravity on tumor position. A characteristic low-pitched sound, a "tumor plop," may be appreciated on auscultation during early or mid-diastole and is thought to result from the impact of the tumor against the mitral valve or ventricular wall. Myxomas also may present with peripheral or pulmonary emboli or with constitutional signs and symptoms, including fever, weight loss, cachexia, malaise, arthralgias, rash, digital clubbing, Raynaud's phenomenon, hypergammaglobulinemia, anemia, polycythemia, leukocytosis, elevated erythrocyte sedimentation rate, thrombocytopenia, and thrombocytosis. These factors account for the frequent misdiagnosis of patients with myxomas as having endocarditis, collagen vascular disease, or a paraneoplastic syndrome.

Two-dimensional transthoracic or omniplane transesophageal echocardiography is useful in the diagnosis of cardiac myxoma and allows assessment of tumor size and determination of the site of tumor attachment, both of which are important considerations in the planning of surgical excision **(Fig. 240-1)**. CT and MRI may provide important information regarding size, shape, composition, and surface characteristics of the tumor **(Fig. 240-2)**.

Figure 240-1

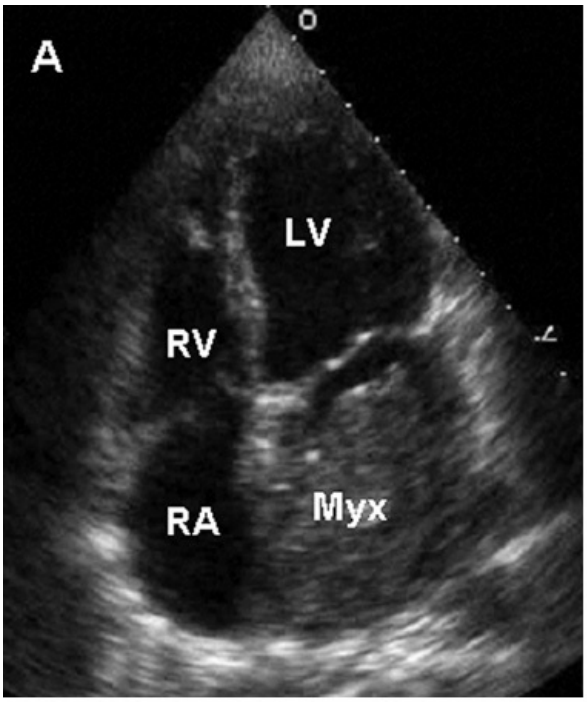

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

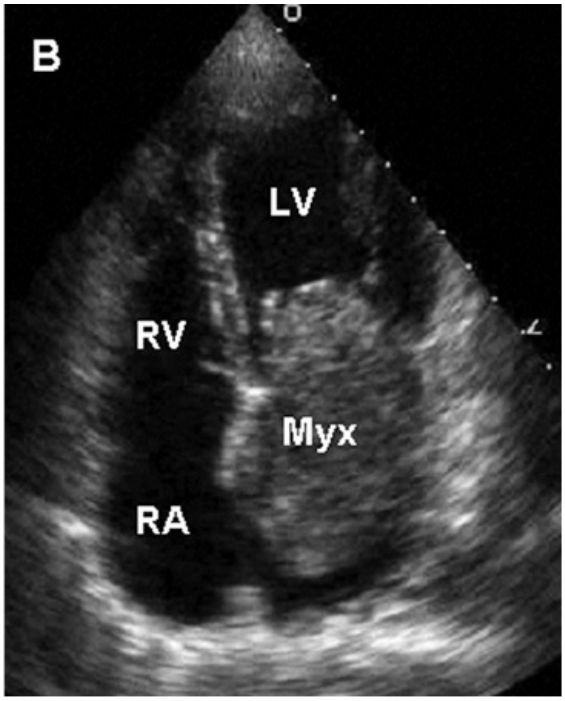

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transthoracic echocardiogram demonstrating a large atrial myxoma. The myxoma (Myx) fills the entire left atrium in systole (*panel A*) and prolapses across the mitral valve and into the left ventricle (LV) during diastole (*panel B*). RA, right atrium; RV, right ventricle. (*Courtesy of Dr. Michael Tsang; with permission.*)

Figure 240-2

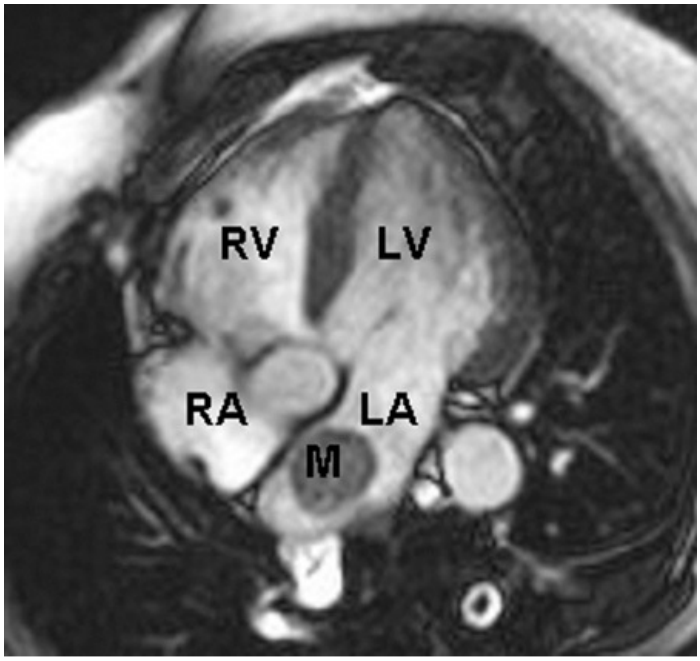

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cardiac MRI demonstrating a rounded mass (M) within the left atrium (LA). Pathologic evaluation at the time of surgery revealed it to be an atrial myxoma. LV, left ventricle; RA, right atrium; RV, right ventricle.

Although cardiac catheterization and angiography were previously performed routinely before tumor resection, they no longer are considered mandatory when adequate noninvasive information is available and other cardiac disorders (e.g., coronary artery disease) are not considered likely.

Additionally, catheterization of the chamber from which the tumor arises carries the risk of tumor embolization. Because myxomas may be familial, echocardiographic screening of first-degree relatives is appropriate, particularly if the patient is young and has multiple tumors or evidence of myxoma syndrome.

TREATMENT MYXOMA

Surgical excision utilizing cardiopulmonary bypass is indicated regardless of tumor size and is generally curative. Myxomas recur in 12–22% of familial cases but in only 1–2% of sporadic cases. Tumor recurrence most likely is due to multifocal lesions in the former and inadequate resection in the latter.

OTHER BENIGN TUMORS

Cardiac *lipomas*, although relatively common, are usually incidental findings at postmortem examination; however, they may grow as large as 15 cm and may present with symptoms owing to mechanical interference with cardiac function, arrhythmias, or conduction disturbances or as an abnormality of the cardiac silhouette on chest x-ray. *Papillary fibroelastomas* are the most common tumors of the cardiac valves. Although usually clinically silent, they can cause valve dysfunction and may embolize distally, resulting in transient ischemic attacks, stroke, or myocardial infarction. Therefore, these tumors should be resected even when asymptomatic. *Rhabdomyomas* and *fibromas* are the most common cardiac tumors in infants and children and usually occur in the ventricles, where they may produce mechanical obstruction to blood flow, thereby mimicking valvular stenosis, congestive heart failure (CHF), restrictive or hypertrophic cardiomyopathy, or pericardial constriction. Rhabdomyomas are probably hamartomatous growths, are multiple in 90% of cases, and are strongly associated with tuberous sclerosis. These tumors have a tendency to regress completely or partially; only tumors that cause obstruction require surgical resection. Fibromas are usually single, are often calcified, tend to grow and cause obstructive symptoms, and should be resected. *Hemangiomas* and *mesotheliomas* are generally small tumors, most often intramyocardial in location, and may cause atrioventricular (AV) conduction disturbances and even sudden death as a result of their propensity to develop in the region of the AV node. Other benign tumors arising from the heart include *teratoma*, *chemodectoma*, *neurilemoma*, *granular cell myoblastoma* , and *bronchogenic cysts*.

SARCOMA

Almost all primary cardiac malignancies are sarcomas, which may be of several histologic types. In general, these tumors are characterized by rapid progression that culminates in the patient's death within weeks to months from the time of presentation as a result of hemodynamic compromise, local invasion, or distant metastases. Sarcomas commonly involve the right side of the heart, are characterized by rapid growth, frequently invade the pericardial space, and may obstruct the cardiac chambers or venae cavae. Sarcomas also may occur on the left side of the heart and may be mistaken for myxomas.

Treatment: Sarcoma

At the time of presentation these tumors have often spread too extensively to allow for surgical excision. Although there are scattered reports of palliation with surgery, radiotherapy, and/or chemotherapy, the response of cardiac sarcomas to these therapies is generally poor. The one exception appears to be cardiac lymphosarcomas, which may respond to a combination of chemo- and radiotherapy.

Tumors Metastatic to the Heart

Tumors metastatic to the heart are much more common than primary tumors, and their incidence is likely to increase as the life expectancy of patients with various forms of malignantneoplasms is extended by more effective therapy. Although cardiac metastases may occur with any tumor type, the relative incidence is especially high in malignant melanoma and, to a somewhat lesser extent, leukemia and lymphoma. In absolute terms, the most common primary originating sites of cardiac metastases are carcinoma of the breast and lung, reflecting the high incidence of those cancers. Cardiac metastases almost always occur in the setting of widespread primary disease, and most often there is either primary or metastatic disease elsewhere in the thoracic cavity. Nevertheless, cardiac metastasis occasionally may be the initial presentation of an extrathoracic tumor.

Cardiac metastases may occur via hematogenous or lymphangitic spread or by direct tumor invasion. They generally manifest as small, firm nodules; diffuse infiltration also may occur, especially with sarcomas or hematologicneoplasms. The pericardium is most often involved, followed by myocardial involvement of any chamber and, rarely, by involvement of the endocardium or cardiac valves.

Cardiac metastases are clinically apparent only ~10% of the time, are usually not the cause of the patient's presentation, and rarely are the cause of death. The vast majority occur in the setting of a previously recognized malignantneoplasm. When symptomatic, cardiac metastases may result in a variety of clinical features, including dyspnea, acute pericarditis, cardiac tamponade, ectopic tachyarrhythmias, heart block, and CHF. As with primary cardiac tumors, the clinical presentation

reflects more the location and size of the tumor than its histologic type. Many of these signs and symptoms may also result from myocarditis, pericarditis, or cardiomyopathy induced by radiotherapy or chemotherapy.

Electrocardiographic (ECG) findings are nonspecific. On chest x-ray, the cardiac silhouette is most often normal but may be enlarged or exhibit a bizarre contour. Echocardiography is useful for identifying pericardial effusions and visualizing larger metastases, although CT and radionuclide imaging with gallium or thallium may define the tumor burden more clearly. Cardiac MRI offers superb image quality and plays a central role in the diagnostic evaluation of cardiac metastases and cardiac tumors in general. Pericardiocentesis may allow for a specific cytologic diagnosis in patients with malignant pericardial effusions. Angiography is rarely necessary but may delineate discrete lesions.

TREATMENT: TUMORS METASTATIC TO THE HEART

Most patients with cardiac metastases have advanced malignant disease; thus, therapy is generally palliative and consists of treatment of the primary tumor. Symptomatic malignant pericardial effusions should be drained by pericardiocentesis. Concomitant instillation of a sclerosing agent (e.g., tetracycline) may delay or prevent reaccumulation of the effusion, and creation of a pericardial window allows drainage of the effusion to the pleural space.

TRAUMATIC CARDIAC INJURY

Traumatic cardiac injury may be caused by either penetrating or nonpenetrating trauma. *Penetrating injuries* most often result from gunshot or knife wounds, and the site of entry is usually obvious. *Nonpenetrating injuries* most often occur during motor vehicle accidents, either from a rapid deceleration injury or from impact of the chest against the steering wheel, and may be associated with significant cardiac injury even in the absence of external signs of thoracic trauma.

Myocardial contusions are the most common form of nonpenetrating cardiac injury and may initially be overlooked in trauma patients as the clinical focus is directed toward other, more obvious injuries. Myocardial necrosis may occur as a direct result of the blunt injury or as a result of traumatic coronary laceration or thrombosis. The contused myocardium is pathologically similar to infarcted myocardium and may be associated with atrial or ventricular arrhythmias; conduction disturbances, including bundle branch block; or ECG abnormalities resembling those of infarction or pericarditis. Thus, it is important to consider contusion as a cause of otherwise unexplained ECG changes in a trauma patient. Serum creatine kinase, myocardial bound (CK-MB) isoenzyme levels are increased in ~20% of patients who experience blunt chest trauma but may be falsely elevated in the presence of massive skeletal muscle injury. Cardiac troponin levels are more specific for identifying cardiac injury in this setting. Echocardiography is useful in detecting structural and functional sequelae of contusion, including wall motion abnormalities, pericardial effusion, valvular dysfunction, and ventricular rupture.

Rupture of the cardiac valves or their supporting structures, most commonly of the tricuspid or mitral valve, leads to acute valvular incompetence. This complication is usually heralded by the development of a loud murmur, may be associated with rapidly progressive heart failure, and can be diagnosed by either transthoracic or transesophageal echocardiography.

The most serious consequence of nonpenetrating cardiac injury is myocardial rupture, which may result in hemopericardium and tamponade (free wall rupture) or intracardiac shunting (ventricular septal rupture). Although it generally is fatal, up to 40% of patients with cardiac rupture have been reported to survive long enough to reach a specialized trauma center. Hemopericardium also may result from traumatic rupture of a pericardial vessel or a coronary artery. Additionally, a pericardial effusion may develop weeks or even months after blunt chest trauma as a manifestation of the post-cardiac injury syndrome, which resembles the post-pericardiotomy syndrome (Chap. 239).

Blunt, nonpenetrating, often innocent-appearing injuries to the chest may trigger ventricular fibrillation even in absence of overt signs of injury. This syndrome, referred to as *commotio cordis*, occurs most often in adolescents during sporting events (e.g., baseball, hockey, football, and lacrosse) and probably results from an impact to the chest wall overlying the heart during the susceptible phase of repolarization just before the peak of the T wave. Survival depends on prompt defibrillation.

Rupture of the aorta, usually just above the aortic valve or at the site of the ligamentum arteriosum, is a common consequence of nonpenetrating chest trauma and is the most common vascular deceleration injury. The clinical presentation is similar to that of aortic dissection (Chap. 248); the arterial pressure and pulse amplitude may be increased in the upper extremities and decreased in the lower extremities, and chest x-ray may reveal mediastinal widening. Occasionally, aortic rupture is contained by the aortic adventitia, resulting in a false, or *pseudo*-, aneurysm that may be discovered months or years after the initial injury.

Sudden emotional or physical trauma may precipitate a transient catecholamine-mediated cardiomyopathy referred to as *Tako-Tsubo syndrome* or the *apical ballooning syndrome (Chap. 238).*

Penetrating injuries of the heart produced by knife or bullet wounds usually result in rapid clinical deterioration and frequently in death as a result of hemopericardium/pericardial tamponade or massive

hemorrhage. Nonetheless, up to half of such patients may survive long enough to reach a specialized trauma center if immediate resuscitation is performed. Prognosis in these patients relates to the mechanism of injury, their clinical condition at presentation, and the specific cardiac chamber(s) involved. Iatrogenic cardiac or coronary arterial perforation may complicate placement of central venous or intracardiac catheters, pacemaker leads, or intracoronary stents and is associated with a better prognosis than are other forms of penetrating cardiac trauma.

Traumatic rupture of a great vessel from penetrating injury is usually associated with hemothorax and, less often, hemopericardium. Local hematoma formation may compress major vessels and produce ischemic symptoms, and AV fistulas may develop, occasionally resulting in high-output CHF.

Occasionally, patients who survive penetrating cardiac injuries may subsequently present with a new cardiac murmur or CHF as a result of mitral regurgitation or an intracardiac shunt (i.e., ventricular or atrial septal defect, aortopulmonary fistula, or coronary AV fistula) that was undetected at the time of the initial injury or developed subsequently. Therefore, trauma patients should be examined carefully several weeks after the injury. If a mechanical complication is suspected, it can be confirmed by echocardiography or cardiac catheterization.

Treatment: Traumatic Cardiac Injury

The treatment of an uncomplicated myocardial contusion is similar to the medical therapy for a myocardial infarction, except that anticoagulation is contraindicated, and should include monitoring for the development of arrhythmias and mechanical complications such as cardiac rupture (Chap. 245). Acute myocardial failure resulting from traumatic valve rupture usually requires urgent operative correction. Immediate thoracotomy should be carried out for most cases of penetrating injury or if there is evidence of cardiac tamponade and/or shock regardless of the type of trauma. Pericardiocentesis may be lifesaving in patients with tamponade but is usually only a temporizing measure while awaiting definitive surgical therapy. Pericardial hemorrhage often leads to constriction (Chap. 239), which must be treated by surgical decortication.

FURTHER READINGS

Burke A et al: Cardiac tumors: An update. Heart 94:117, 2008[PMID: 18083956]

Kalra MK, Abbara S: Imaging cardiac tumors. Cancer Treat Res 143:177, 2008[PMID: 18619218]

Mattox KL: Traumatic heart disease, in *Braunwald's Heart Disease*, 7th ed, DP Zipes et al (eds). Philadelphia, Saunders, 2005

Pretre R, Chilcott M: Blunt trauma to the heart and great vessels. N Engl J Med 336:626, 1997[PMID: 9032049]

Reyman K: Cardiac myxomas. N Engl J Med 333:1610, 1995

Rhee PM et al: Penetrating cardiac injuries: A population-based study. J Trauma 45:366, 1998[PMID: 9715197]

Sabatine M, Schoen F: Primary tumors of the heart, in *Braunwald's Heart Disease*, 7th ed, DP Zipes et al (eds). Philadelphia, Saunders, 2005

Simmers TA et al: Traumatic papillary muscle rupture. Ann Thorac Surg 72:257, 2001[PMID: 11465192]

Sybrandy KC et al: Diagnosing cardiac contusion: Old wisdom and new insights. Heart 89:485, 2003[PMID: 12695446]

Tyburski JG et al: Factors affecting prognosis with penetrating wounds of the heart. J Trauma 48:587, 2000[PMID: 10780588]

Vaughan CJ et al: Tumors and the heart: Molecular genetic advances. Curr Opin Cardiol 16:195, 2001[PMID: 11357016]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CARDIAC MANIFESTATIONS OF SYSTEMIC DISEASE: INTRODUCTION

The common systemic disorders that have associated cardiac manifestations are summarized in **Table e31-1**.

Table e31-1 Common Systemic Disorders and Their Associated Cardiac Manifestations

Systemic Disorder	Common Cardiac Manifestations	Chapter
Diabetes mellitus	CAD, atypical angina, CMP, systolic or diastolic CHF	344
Protein-calorie malnutrition	Dilated CMP, CHF	75
Thiamine deficiency	High-output failure, dilated CMP	74
Hyperhomocysteinemia	Premature atherosclerosis	74
Obesity	CMP, systolic or diastolic CHF	77
Hyperthyroidism	Palpitations, SVT, atrial fibrillation, hypertension	341
Hypothyroidism	Hypotension, bradycardia, dilated CMP, CHF, pericardial effusion	341
Malignantcarcinoid	Tricuspid and pulmonary valve disease, right heart failure	350
Pheochromocytoma	Hypertension, palpitations, CHF	343
Acromegaly	Systolic or diastolic heart failure	339
Rheumatoid arthritis	Pericarditis, pericardial effusions, coronary arteritis, myocarditis, valvulitis	321
Seronegativearthropathies	Aortitis, aortic and mitral insufficiency, conduction abnormalities	325
Systemic lupus erythematosus	Pericarditis, Libman-Sacks endocarditis, myocarditis, arterial and venous thrombosis	319
HIV	Myocarditis, dilated CMP, pericardial effusion	189
Amyloidosis	CHF, restrictive CMP, valvular regurgitation, pericardial effusion	112
Sarcoidosis	CHF, dilated or restrictive CMP, ventricular arrhythmias, heart block	329
Hemochromatosis	CHF, arrhythmias, heart block	357
Marfan syndrome	Aortic aneurysm and dissection, aortic insufficiency, mitral valve prolapse	363
Ehlers-Danlos syndrome	Aortic and coronary aneurysms, mitral and tricuspid valve prolapse	363

Abbreviations: CAD, coronary artery disease; CHF, congestive heart failure; CMP, cardiomyopathy; SVT, supraventricular tachycardia.

Diabetes Mellitus

(See also Chap. 344) Diabetes mellitus, both insulin- and non-insulin-dependent, is an independent risk factor for coronary artery disease (CAD; Chap. 241) and accounts for 14–50% of new cases of cardiovascular disease. Furthermore, CAD is the most common cause of death in adults with diabetes mellitus. In the diabetic population the incidence of CAD relates to the duration of diabetes and the level of glycemic control, and its pathogenesis involves endothelial dysfunction, increased lipoprotein peroxidation, increased inflammation, a prothrombotic state, and associated metabolic abnormalities.

Diabetic patients are more likely to have a myocardial infarction, have a greater burden of CAD, have larger infarct size, and have more postinfarct complications, including heart failure, shock, and death, than are nondiabetics. Importantly, diabetic patients are more likely to have atypical ischemic symptoms; nausea, dyspnea, pulmonary edema, arrhythmias, heart block, or syncope may be their anginal equivalent. Additionally, "silent ischemia," resulting from autonomic nervous system dysfunction, is more common in diabetic patients, accounting for up to 90% of their ischemic episodes. Thus, one must have a low threshold for suspecting CAD in diabetic patients. The treatment of diabetic patients with CAD must include aggressive risk factor management (Chap. 344).

Pharmacologic therapy and revascularization are similar in diabetic patients and nondiabetics except that diabetic patients have higher morbidity and mortality rates associated with revascularization, have an increased risk of restenosis after percutaneous coronary intervention (PCI), and probably have improved survival when treated with surgical bypass compared with PCI for multivessel CAD.

Patients with diabetes mellitus also may have abnormal left ventricular systolic and diastolic function, reflecting concomitant epicardial CAD and/or hypertension, coronary microvascular disease, endothelial dysfunction, ventricular hypertrophy, and autonomic dysfunction. A restrictive cardiomyopathy may be present with abnormal myocardial relaxation and elevated ventricular filling pressures. Histologically, interstitial fibrosis is seen, and intramural arteries may demonstrate intimal thickening, hyaline deposition, and inflammatory changes. Diabetic patients have an increased risk of developing clinical heart failure, which probably contributes to their excessive cardiovascular morbidity and mortality rates. There is some evidence that insulin therapy may ameliorate diabetes-related myocardial dysfunction.

Malnutrition and Vitamin Deficiency

MALNUTRITION

(See also Chap. 75) In patients whose intake of protein, calories, or both is severely deficient, the heart may become thin, pale, and hypokinetic with myofibrillar atrophy and interstitial edema. The systolic pressure and cardiac output fall, and the pulse pressure narrows. Generalized edema is common and relates to a variety of factors, including reduced serum oncotic pressure and myocardial dysfunction. Such profound states of protein and calorie malnutrition, termed *kwashiorkor* and *marasmus,* respectively, are most common in underdeveloped countries. However, significant nutritional heart disease also may occur in developed nations, particularly in patients with chronic diseases such as AIDS, patients with anorexia nervosa, and patients with severe cardiac failure in whom gastrointestinal hypoperfusion and venous congestion may lead to anorexia and malabsorption. Open-heart surgery poses increased risk in malnourished patients, and those patients may benefit from preoperative hyperalimentation.

THIAMINE DEFICIENCY (BERIBERI)

(See also Chap. 74) Generalized malnutrition often is accompanied by thiamine deficiency; however, this hypovitaminosis also may occur in the presence of an adequate protein and caloric intake, particularly in the Far East, where polished rice deficient in thiamine may be a major dietary component. In Western nations where the use of thiamine-enriched flour is widespread, clinical thiamine deficiency is limited primarily to alcoholics, food faddists, and patients receiving chemotherapy. Nonetheless, when thiamine stores are measured using the thiamine-pyrophosphate effect (TPPE), thiamine deficiency has been found in 20–90% of patients with chronic heart failure. This deficiency appears to result from both reduced dietary intake and a diuretic-induced increase in the urinary excretion of thiamine. The acute administration of thiamine to these patients increases the left ventricular ejection fraction and the excretion of salt and water.

Clinically, patients with thiamine deficiency usually have evidence of generalized malnutrition, peripheral neuropathy, glossitis, and anemia. The classic associated cardiovascular syndrome is characterized by high-output heart failure, tachycardia, and often elevated biventricular filling pressures. The major cause of the high-output state is vasomotor depression leading to reduced systemic vascular resistance, the precise mechanism of which is not understood. The cardiac examination reveals a wide pulse pressure, tachycardia, a third heart sound, and, frequently, an apical systolic murmur. The electrocardiogram (ECG) may reveal decreased voltage, a prolonged QT interval, and T-wave abnormalities. The chest x-ray generally reveals cardiomegaly and signs of congestive heart failure (CHF). The response to thiamine is often dramatic, with an increase in systemic vascular resistance, a decrease in cardiac output, clearing of pulmonary congestion, and a reduction in heart size often occurring in 12–48 h. Although the response to inotropes and diuretics may be poor before thiamine therapy, these agents may be important *after* thiamine is given, since the left ventricle may not be able to handle the increased work load presented by the return of vascular tone.

VITAMIN B_6, B_{12}, AND FOLATE DEFICIENCY

(See also Chap. 74) Vitamin B6, B_{12}, and folate are cofactors in the metabolism of homocysteine. Their deficiency probably contributes to the majority of cases of hyperhomocysteinemia, a disorder associated with increased atherosclerotic risk. Supplementation of these vitamins has reduced the incidence of hyperhomocysteinemia in the United States; however, the clinical cardiovascular benefit of normalizing elevated homocysteine levels has not been proved.

Obesity

(See also Chap. 77) Severe obesity, especially abdominal obesity, is associated with an increase in cardiovascular morbidity and mortality rates. Although obesity itself is not considered a disease, it is associated with an increased prevalence of hypertension, glucose intolerance, and atherosclerotic CAD. In addition, obese patients have a distinct cardiovascular abnormality characterized by increased

total and central blood volumes, increased cardiac output, and elevated left ventricular filling pressure. The elevated cardiac output appears to be required to support the metabolic demands of the excess adipose tissue. Left ventricular filling pressure is often at the upper limits of normal at rest and rises excessively with exercise. In part as a result of chronic volume overload, eccentric cardiac hypertrophy with cardiac dilation and ventricular dysfunction may develop. In addition, altered levels of adipokines secreted by adipose tissue may contribute to adverse myocardial remodeling via direct effects on cardiac myocytes and other cells. Pathologically, there is left and, in some cases, right ventricular hypertrophy and generalized cardiac dilation. Pulmonary congestion, peripheral edema, and exercise intolerance may all ensue; however, the recognition of these findings may be difficult in massively obese patients.

Weight reduction is the most effective therapy and results in reduction in blood volume and the return of cardiac output toward normal. However, rapid weight reduction may be dangerous, as cardiac arrhythmias and sudden death owing to electrolyte imbalance have been described. Treatment with angiotensin-converting enzyme inhibitors, sodium restriction, and diuretics may be useful to control heart failure symptoms. This form of heart disease should be distinguished from the Pickwickian syndrome (Chap. 264), which may share several of the cardiovascular features of heart disease secondary to severe obesity but, in addition, frequently has components of central apnea, hypoxemia, pulmonary hypertension, and cor pulmonale.

Thyroid Disease

(See also Chap. 341) Thyroid hormone exerts a major influence on the cardiovascular system by a number of direct and indirect mechanisms, and, not surprisingly, cardiovascular effects are prominent in both hypo- and hyperthyroidism. Thyroid hormone causes increases in total-body metabolism and oxygen consumption that indirectly increase the cardiac workload. In addition, thyroid hormone exerts direct inotropic, chronotropic, and dromotropic effects that are similar to those seen with adrenergic stimulation (e.g., tachycardia, increased cardiac output); they are mediated at least partly by both transcriptional and nontranscriptional effects of thyroid hormone on myosin, calcium-activated ATPase, Na^+-K^+-ATPase, and myocardial ß-adrenergic receptors.

HYPERTHYROIDISM

Common cardiovascular manifestations of hyperthyroidism include palpitations, systolic hypertension, and fatigue. Sinus tachycardia is present in ~40% of hyperthyroid patients, and atrial fibrillation in ~15%. Physical examination may reveal a hyperdynamic precordium, a widened pulse pressure, increases in the intensity of the first heart sound and the pulmonic component of the second heart sound, and a third heart sound. An increased incidence of mitral valve prolapse has been described in hyperthyroid patients, in which case a midsystolic murmur may be heard at the left sternal border with or without a midsystolic click. A systolic pleuropericardial friction rub (*Means-Lerman scratch*) may be heard at the left second intercostal space during expiration and is thought to result from the hyperdynamic cardiac motion.

Elderly patients with hyperthyroidism may present with only cardiovascular manifestations of thyrotoxicosis such as sinus tachycardia, atrial fibrillation, and hypertension, all of which may be resistant to therapy until the hyperthyroidism is controlled. Angina pectoris and CHF are unusual with hyperthyroidism unless there is coexistent heart disease; in such cases, symptoms often resolve with treatment of the hyperthyroidism.

HYPOTHYROIDISM

Cardiac manifestations of hypothyroidism include a reduction in cardiac output, stroke volume, heart rate, blood pressure, and pulse pressure. Pericardial effusions are present in about one-third of patients, rarely progress to tamponade, and probably result from increased capillary permeability. Other clinical signs include cardiomegaly, bradycardia, weak arterial pulses, distant heart sounds, and pleural effusions. Although the signs and symptoms of myxedema may mimic those of CHF, in the absence of other cardiac disease, myocardial failure is uncommon. The ECG generally reveals sinus bradycardia and low voltage and may show prolongation of the QT interval, decreased P-wave voltage, prolonged AV conduction time, intraventricular conduction disturbances, and nonspecific ST-T-wave abnormalities. Chest x-ray may show cardiomegaly, often with a "water bottle" configuration; pleural effusions; and, in some cases, evidence of CHF. Pathologically, the heart is pale and dilated and often demonstrates myofibrillar swelling, loss of striations, and interstitial fibrosis.

Patients with hypothyroidism frequently have elevations of cholesterol and triglycerides, resulting in premature atherosclerotic CAD. Before treatment with thyroid hormone, patients with hypothyroidism frequently do not have angina pectoris, presumably because of the low metabolic demands caused by their condition. However, angina and myocardial infarction may be precipitated during initiation of thyroid hormone replacement, especially in elderly patients with underlying heart disease. Therefore, replacement should be done with care, starting with low doses that are increased gradually.

Malignant Carcinoid

(See also Chap. 350) Carcinoid tumors most often originate in the small bowel and elaborate a variety

of vasoactive amines (e.g., serotonin), kinins, indoles, and prostaglandins that are believed to be responsible for the diarrhea, flushing, and labile blood pressure that characterize the carcinoid syndrome. Some 50% of patients with carcinoid syndrome have cardiac involvement, usually manifesting as abnormalities of the right-sided cardiac structures. These patients invariably have hepatic metastases that allow vasoactive substances to circumvent hepatic metabolism. Left-sided cardiac involvement is rare and indicates either pulmonary carcinoid or an intracardiac shunt. Pathologically, carcinoid lesions are fibrous plaques that consist of smooth-muscle cells embedded in a stroma of glycosaminoglycans and collagen. They occur on the cardiac valves, where they cause valvular dysfunction, as well as on the endothelium of the cardiac chambers and great vessels.

Carcinoid heart disease most often presents as tricuspid regurgitation, pulmonic stenosis, or both. In some cases a high cardiac output state may occur, presumably as a result of a decrease in systemic vascular resistance resulting from vasoactive substances released by the tumor. Treatment with somatostatin analogues (e.g., octreotide) or interferon α improves symptoms and survival in patients with carcinoid heart disease but does not appear to improve valvular abnormalities. In some severely symptomatic patients, valve replacement is indicated. Coronary artery spasm, presumably due to a circulating vasoactive substance, may occur in patients with carcinoid syndrome.

Pheochromocytoma

(See also Chap. 343) In addition to causing labile or sustained hypertension, the high circulating levels of catecholamines resulting from a pheochromocytoma may cause direct myocardial injury. Focal myocardial necrosis and inflammatory cell infiltration are present in ~50% of patients who die with pheochromocytoma and may contribute to clinically significant left ventricular failure and pulmonary edema. In addition, associated hypertension results in left ventricular hypertrophy. Left ventricular dysfunction and CHF may resolve after removal of the tumor.

Acromegaly

(See also Chap. 339) Exposure of the heart to excessive growth hormone may cause CHF as a result of high cardiac output, diastolic dysfunction owing to ventricular hypertrophy (with increased left ventricular chamber size or wall thickness), or global systolic dysfunction. Hypertension occurs in up to one-third of patients with acromegaly and is characterized by suppression of the renin-angiotensin-aldosterone axis and increases in total-body sodium and plasma volume. Some form of cardiac disease occurs in about one-third of patients with acromegaly and is associated with a doubling of the risk of cardiac death.

Rheumatoid Arthritis and the Collagenvascular Diseases

RHEUMATOID ARTHRITIS

(See also Chap. 321) Rheumatoid arthritis may be associated with inflammatory changes in any or all cardiac structures, although pericarditis is the most common clinical entity. Pericardial effusions are found on echocardiography in 10–50% of patients with rheumatoid arthritis, particularly those with subcutaneous nodules. Nonetheless, only a small fraction of these patients have symptomatic pericarditis, and when present, it usually follows a benign course, only occasionally progressing to cardiac tamponade or constrictive pericarditis. The pericardial fluid is generally exudative, with decreased concentrations of complement and glucose and elevated cholesterol. Coronary arteritis with intimal inflammation and edema is present in ~20% of cases but only rarely results in angina pectoris or myocardial infarction. Inflammation and granuloma formation may affect the cardiac valves, most often the mitral and aortic valves, and may cause clinically significant regurgitation owing to valve deformity. Myocarditis is uncommon and rarely results in cardiac dysfunction.

Treatment is directed at the underlying rheumatoid arthritis and may include glucocorticoids. Urgent pericardiocentesis should be performed in patients with tamponade, but pericardiectomy usually is required in cases of pericardial constriction.

SERONEGATIVE ARTHROPATHIES

(See also Chap. 325) The seronegativearthropathies, including ankylosing spondylitis, reactive arthritis, psoriatic arthritis, and the arthritides associated with ulcerative colitis and regional enteritis, are all strongly associated with the HLA-B27 histocompatibility antigen and may be accompanied by a pancarditis and proximal aortitis. The aortic inflammation usually is limited to the aortic root but may extend to involve the aortic valve, mitral valve, and ventricular myocardium, resulting in aortic and mitral regurgitation, conduction abnormalities, and ventricular dysfunction. One-tenth of these patients have significant aortic insufficiency, and one-third have conduction disturbances; both are more common in patients with peripheral joint involvement and long-standing disease. Treatment with aortic valve replacement and permanent pacemaker implantation may be required. Occasionally, aortic regurgitation precedes the onset of arthritis, and therefore, the diagnosis of a seronegative arthritis should be considered in young males with isolated aortic regurgitation.

SYSTEMIC LUPUS ERYTHEMATOSUS (SLE)

(See also Chap. 319) A significant percentage of patients with SLE have cardiac involvement. Pericarditis is common, occurring in about two-thirds of patients, and generally follows a benign

course, although rarely tamponade or constriction may result. The characteristic endocardial lesions of SLE are verrucous valvular abnormalities known as *Libman-Sacks endocarditis*. They most often are located on the left-sided cardiac valves, particularly on the ventricular surface of the posterior mitral leaflet, and are made up almost entirely of fibrin. These lesions may embolize or become infected but rarely cause hemodynamically important valvular regurgitation. Myocarditis generally parallels the activity of the disease and, although common histologically, seldom results in clinical heart failure unless associated with hypertension. Although arteritis of epicardial coronary arteries may occur, it rarely results in myocardial ischemia. There is, however, an increased incidence of coronary atherosclerosis that probably is related more to associated risk factors and glucocorticoid use than to SLE itself. Patients with the antiphospholipid antibody syndrome may have a higher incidence of cardiovascular abnormalities, including valvular regurgitation, venous and arterial thrombosis, premature stroke, myocardial infarction, pulmonary hypertension, and cardiomyopathy.

FURTHER READINGS

Bhattacharyya S et al: Carcinoid heart disease. Circulation 116:2860, 2007[PMID: 18071089]

Gerdes AM, Iervasi G: Thyroid replacement therapy and heart failure. Circulation 122:385, 2010[PMID: 20660814]

Klein I, Danzi S: Thyroid disease and the heart. Circulation 116:1725, 2007[PMID: 17923583]

Lavie CJ et al: Obesity and cardiovascular disease: Risk factor, paradox, and impact of weight loss. J Am Coll Cardiol 26:1925, 2009

Mottillo S et al: The metabolic syndrome and cardiovascular risk: A systematic review and meta-analysis. J Am Coll Cardiol 56:1113, 2010[PMID: 20863953]

Retnakaran R, Zinman B: Type 1 diabetes, hyperglycaemia, and the heart. Lancet 24:1790, 2008

Rodondi N et al: Subclinical hypothyroidism and the risk of coronary heart disease and mortality. JAMA 304:1365, 2010[PMID: 20858880]

Roldan CA: Valvular and coronary heart disease in systemic inflammatory disease: Systemic disorders in heart disease. Heart 94:1089, 2008[PMID: 18625800]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PATHOGENESIS

Atherosclerosis remains the major cause of death and premature disability in developed societies. Moreover, current predictions estimate that by the year 2020 cardiovascular diseases, notably atherosclerosis, will become the leading global cause of total disease burden. Although many generalized or systemic risk factors predispose to its development, atherosclerosis affects various regions of the circulation preferentially and has distinct clinical manifestations that depend on the particular circulatory bed affected. Atherosclerosis of the coronary arteries commonly causes myocardial infarction (MI) (Chap. 245) and angina pectoris (Chap. 243). Atherosclerosis of the arteries supplying the central nervous system frequently provokes strokes and transient cerebral ischemia (Chap. 370). In the peripheral circulation, atherosclerosis causes intermittent claudication and gangrene and can jeopardize limb viability. Involvement of the splanchnic circulation can cause mesenteric ischemia. Atherosclerosis can affect the kidneys either directly (e.g., renal artery stenosis) or as a common site of atheroembolic disease (Chap. 248).

Even within a particular arterial bed, stenoses due to atherosclerosis tend to occur focally, typically in certain predisposed regions. In the coronary circulation, for example, the proximal left anterior descending coronary artery exhibits a particular predilection for developing atherosclerotic disease. Similarly, atherosclerosis preferentially affects the proximal portions of the renal arteries and, in the extracranial circulation to the brain, the carotid bifurcation. Indeed, atherosclerotic lesions often form at branching points of arteries which are regions of disturbed blood flow. Not all manifestations of atherosclerosis result from stenotic, occlusive disease. Ectasia and the development of aneurysmal disease, for example, frequently occur in the aorta (Chap. 248). In addition to focal, flow-limiting stenoses, nonocclusive intimal atherosclerosis also occurs diffusely in affected arteries, as shown by intravascular ultrasound and postmortem studies.

Atherogenesis in humans typically occurs over a period of many years, usually many decades. Growth of atherosclerotic plaques probably does not occur in a smooth, linear fashion but discontinuously, with periods of relative quiescence punctuated by periods of rapid evolution. After a generally prolonged "silent" period, atherosclerosis may become clinically manifest. The clinical expressions of atherosclerosis may be *chronic*, as in the development of stable, effort-induced angina pectoris or predictable and reproducible intermittent claudication. Alternatively, a dramatic *acute* clinical event such as MI, stroke, or sudden cardiac death may first herald the presence of atherosclerosis. Other individuals may never experience clinical manifestations of arterial disease despite the presence of widespread atherosclerosis demonstrated postmortem.

Initiation of Atherosclerosis

An integrated view of experimental results in animals and studies of human atherosclerosis suggests that the "fatty streak" represents the initial lesion of atherosclerosis. These early lesions most often seem to arise from focal increases in the content of lipoproteins within regions of the intima. This accumulation of lipoprotein particles may not result simply from increased permeability, or "leakiness," of the overlying endothelium **(Fig. 241-1)**. Rather, the lipoproteins may collect in the intima of arteries because they bind to constituents of the extracellular matrix, increasing the residence time of the lipid-rich particles within the arterial wall. Lipoproteins that accumulate in the extracellular space of the intima of arteries often associate with glycosaminoglycans of the arterial extracellular matrix, an interaction that may slow the egress of these lipid-rich particles from the intima. Lipoprotein particles in the extracellular space of the intima, particularly those retained by binding to matrix macromolecules, may undergo oxidative modifications. Considerable evidence supports a pathogenic role for products of oxidized lipoproteins in atherogenesis. Lipoproteins sequestered from plasma antioxidants in the extracellular space of the intima become particularly susceptible to oxidative modification, giving rise to hydroperoxides, lysophospholipids, oxysterols, and aldehydic breakdown products of fatty acids and phospholipids. Modifications of the apoprotein moieties may include breaks in the peptide backbone as well as derivatization of certain amino acid residues. Local production of hypochlorous acid by myeloperoxidase associated with inflammatory cells within the plaque yields chlorinated species such as chlorotyrosyl moieties. High-density lipoprotein (HDL) particles modified by HOCl-mediated chlorination function poorly as cholesterol acceptors, a finding that links oxidative stress with impaired reverse cholesterol transport, which is one likely mechanism of the antiatherogenic action of HDL (see below). Considerable evidence supports the presence of such oxidation products in atherosclerotic lesions. A particular member of the phospholipase family, lipoprotein-associated phospholipase A_2

(LpPL A_2), can generate proinflammatory lipids, including lysophosphatidyl choline-bearing oxidized lipid moieties from oxidized phospholipids found in oxidized low-density lipoproteins (LDLs). An inhibitor of this enzyme is in clinical development.

Figure 241-1

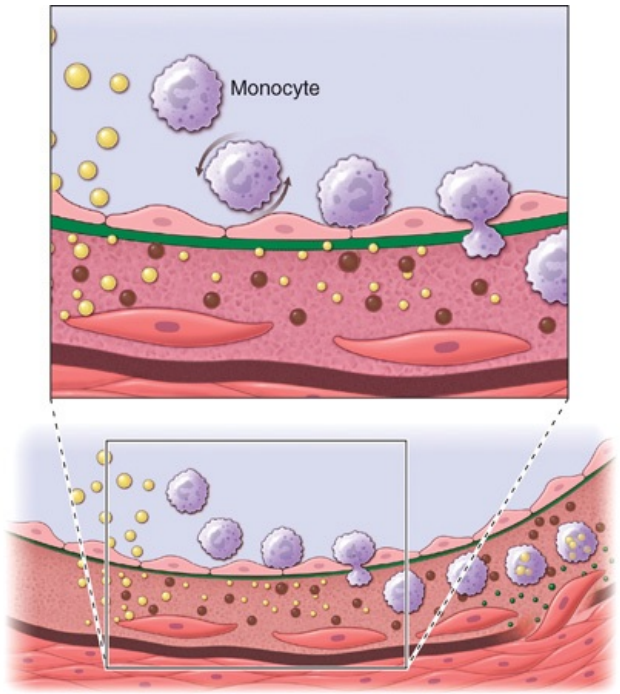

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cross-sectional view of an artery depicting steps in development of an atheroma, from left to right. The *upper panel* shows a detail of the boxed area below. The endothelial monolayer overlying the intima contacts blood. Hypercholesterolemia promotes accumulation of LDL particles (*light spheres*) in the intima. The lipoprotein particles often associate with constituents of the extracellular matrix, notably proteoglycans. Sequestration within the intima separates lipoproteins from some plasma antioxidants and favors oxidative modification. Such modified lipoprotein particles (*darker spheres*) may trigger a local inflammatory response that signals subsequent steps in lesion formation. The augmented expression of various adhesion molecules for leukocytes recruits monocytes to the site of a nascent arterial lesion.

Once adherent, some white blood cells migrate into the intima. The directed migration of leukocytes probably depends on chemoattractant factors, including modified lipoprotein particles themselves and chemoattractant cytokines (depicted by the smaller spheres), such as the chemokine macrophage chemoattractant protein-1 produced by vascular wall cells in response to modified lipoproteins. Leukocytes in the evolving fatty streak can divide and exhibit augmented expression of receptors for modified lipoproteins (scavenger receptors). These mononuclear phagocytes ingest lipids and become foam cells, represented by a cytoplasm filled with lipid droplets. As the fatty streak evolves into a more complicated atherosclerotic lesion, smooth muscle cells migrate from the media (*bottom of lower panel hairline*) through the internal elastic membrane (*solid wavy line*) and accumulate within the expanding intima, where they lay down extracellular matrix that forms the bulk of the advanced lesion (*bottom panel*, *right side*).

LEUKOCYTE RECRUITMENT

Accumulation of leukocytes characterizes the formation of early atherosclerotic lesions (Fig. 241-1). Thus, from its very inception, atherogenesis involves elements of inflammation, a process that now provides a unifying theme in the pathogenesis of this disease. The inflammatory cell types typically found in the evolving atheroma include monocyte-derived macrophages and lymphocytes. A number of adhesion molecules or receptors for leukocytes expressed on the surface of the arterial endothelial cell probably participate in the recruitment of leukocytes to the nascent atheroma. Constituents of oxidatively modified low-density lipoprotein can augment the expression of leukocyte adhesion molecules. This example illustrates how the accumulation of lipoproteins in the arterial intima may link mechanistically with leukocyte recruitment, a key event in lesion formation.

Laminar shear forces such as those encountered in most regions of normal arteries also can suppress the expression of leukocyte adhesion molecules. Sites of predilection for atherosclerotic lesions (e.g., branch points) often have disturbed flow. Ordered, pulsatile laminar shear of normal blood flow augments the production of nitric oxide by endothelial cells. This molecule, in addition to its vasodilator properties, can act at the low levels constitutively produced by arterial endothelium as a local anti-

inflammatory autacoid, e.g., limiting local adhesion molecule expression. Exposure of endothelial cells to laminar shear stress increases the transcription of Krüppel-like factor 2 (KLF2) and reduces the expression of a thioredoxin-interacting protein (Txnip) that inhibits the activity of the endogenous antioxidant thioredoxin. KLF2 augments the activity of endothelial nitric oxide synthase, and reduced Txnip levels boost the function of thioredoxin. Laminar shear stress also stimulates endothelial cells to produce superoxide dismutase, an antioxidant enzyme. These examples indicate how hemodynamic forces may influence the cellular events that underlie atherosclerotic lesion initiation and potentially explain the favored localization of atherosclerotic lesions at sites that experience disturbance to laminar shear stress.

Once captured on the surface of the arterial endothelial cell by adhesion receptors, the monocytes and lymphocytes penetrate the endothelial layer and take up residence in the intima. In addition to products of modified lipoproteins, cytokines (protein mediators of inflammation) can regulate the expression of adhesion molecules involved in leukocyte recruitment. For example, interleukin 1 (IL-1) or tumor necrosis factor α (TNF-α) induce or augment the expression of leukocyte adhesion molecules on endothelial cells. Because products of lipoprotein oxidation can induce cytokine release from vascular wall cells, this pathway may provide an additional link between arterial accumulation of lipoproteins and leukocyte recruitment. Chemoattractant cytokines such as monocyte chemoattractant protein 1 appear to direct the migration of leukocytes into the arterial wall.

FOAM-CELL FORMATION

Once resident within the intima, the mononuclear phagocytes mature into macrophages and become lipid-laden foam cells, a conversion that requires the uptake of lipoprotein particles by receptor-mediated endocytosis. One might suppose that the well-recognized "classic" receptor for LDL mediates this lipid uptake; however, humans or animals lacking effective LDL receptors due to genetic alterations (e.g., familial hypercholesterolemia) have abundant arterial lesions and extraarterial xanthomata rich in macrophage-derived foam cells. In addition, the exogenous cholesterol suppresses expression of the LDL receptor; thus, the level of this cell-surface receptor for LDL decreases under conditions of cholesterol excess. Candidates for alternative receptors that can mediate lipid loading of foam cells include a growing number of macrophage "scavenger" receptors, which preferentially endocytose modified lipoproteins, and other receptors for oxidized LDL or very low-density lipoprotein (VLDL). Monocyte attachment to the endothelium, migration into the intima, and maturation to form lipid-laden macrophages thus represent key steps in the formation of the fatty streak, the precursor of fully formed atherosclerotic plaques.

Atheroma Evolution and Complications

Although the fatty streak commonly precedes the development of a more advanced atherosclerotic plaque, not all fatty streaks progress to form complex atheromata. By ingesting lipids from the extracellular space, the mononuclear phagocytes bearing such scavenger receptors may remove lipoproteins from the developing lesion. Some lipid-laden macrophages may leave the artery wall, exporting lipid in the process. Lipid accumulation, and hence the propensity to form an atheroma, ensues if the amount of lipid entering the artery wall exceeds that removed by mononuclear phagocytes or other pathways.

Export by phagocytes may constitute one response to local lipid overload in the evolving lesion. Another mechanism, reverse cholesterol transport mediated by high-density lipoproteins, probably provides an independent pathway for lipid removal from atheroma. This transfer of cholesterol from the cell to the HDL particle involves specialized cell-surface molecules such as the ATP binding cassette (ABC) transporters. *ABCA1*, the gene mutated in Tangier disease, a condition characterized by very low HDL levels, transfers cholesterol from cells to nascent HDL particles and ABCG1 to mature HDL particles. "Reverse cholesterol transport" mediated by these ABC transporters allows HDL loaded with cholesterol to deliver it to hepatocytes by binding to scavenger receptor B 1 or other receptors. The liver cell can metabolize the sterol to bile acids that can be excreted. This export pathway from macrophage foam cells to peripheral cells such as hepatocytes explains part of the antiatherogenic action of HDLs. (Anti-inflammatory and antioxidant properties also may contribute to the atheroprotective effects of HDLs.) Thus, macrophages may play a vital role in the dynamic economy of lipid accumulation in the arterial wall during atherogenesis.

Some lipid-laden foam cells within the expanding intimal lesion perish. Some foam cells may die as a result of programmed cell death, or *apoptosis*. This death of mononuclear phagocytes results in the formation of the lipid-rich center, often called the *necrotic core*, in established atherosclerotic plaques. Macrophages loaded with modified lipoproteins may elaborate cytokines and growth factors that can further signal some of the cellular events in lesion complication. Whereas accumulation of lipid-laden macrophages characterizes the fatty streak, buildup of fibrous tissue formed by extracellular matrix typifies the more advanced atherosclerotic lesion. The smooth muscle cell synthesizes the bulk of the extracellular matrix of the complex atherosclerotic lesion. A number of growth factors or cytokines elaborated by mononuclear phagocytes can stimulate smooth muscle cell proliferation and production of extracellular matrix. Cytokines found in the plaque, including IL-1 and TNF-α, can induce local production of growth factors, including forms of platelet-derived growth factor (PDGF), fibroblast

growth factors, and others, which may contribute to plaque evolution and complication. Other cytokines, notably interferon ɣ (IFN-ɣ) derived from activated T cells within lesions, can limit the synthesis of interstitial forms of collagen by smooth muscle cells. These examples illustrate how atherogenesis involves a complex mix of mediators that in the balance determines the characteristics of particular lesions.

The arrival of smooth muscle cells and their elaboration of extracellular matrix probably provide a critical transition, yielding a fibrofatty lesion in place of a simple accumulation of macrophage-derived foam cells. For example, PDGF elaborated by activated platelets, macrophages, and endothelial cells can stimulate the migration of smooth muscle cells normally resident in the tunica media into the intima. Such growth factors and cytokines produced locally can stimulate the proliferation of resident smooth muscle cells in the intima as well as those that have migrated from the media. Transforming growth factor ß (TGF-ß), among other mediators, potently stimulates interstitial collagen production by smooth muscle cells. These mediators may arise not only from neighboring vascular cells or leukocytes (a "paracrine" pathway), but also, in some instances, may arise from the same cell that responds to the factor (an "autocrine" pathway). Together, these alterations in smooth muscle cells, signaled by these mediators acting at short distances, can hasten transformation of the fatty streak into a more fibrous smooth muscle cell and extracellular matrix-rich lesion.

In addition to locally produced mediators, products of blood coagulation and thrombosis likely contribute to atheroma evolution and complication. This involvement justifies the use of the term *atherothrombosis* to convey the inextricable links between atherosclerosis and thrombosis. Fatty streak formation begins beneath a morphologically intact endothelium. In advanced fatty streaks, however, microscopic breaches in endothelial integrity may occur. Microthrombi rich in platelets can form at such sites of limited endothelial denudation, owing to exposure of the thrombogenic extracellular matrix of the underlying basement membrane. Activated platelets release numerous factors that can promote the fibrotic response, including PDGF and TGF-ß. Thrombin not only generates fibrin during coagulation, but also stimulates protease-activated receptors that can signal smooth muscle migration, proliferation, and extracellular matrix production. Many arterial mural microthrombi resolve without clinical manifestation by a process of local fibrinolysis, resorption, and endothelial repair, yet can lead to lesion progression by stimulating these profibrotic functions of smooth muscle cells **(Fig. 241-2*D*)**.

Figure 241-2

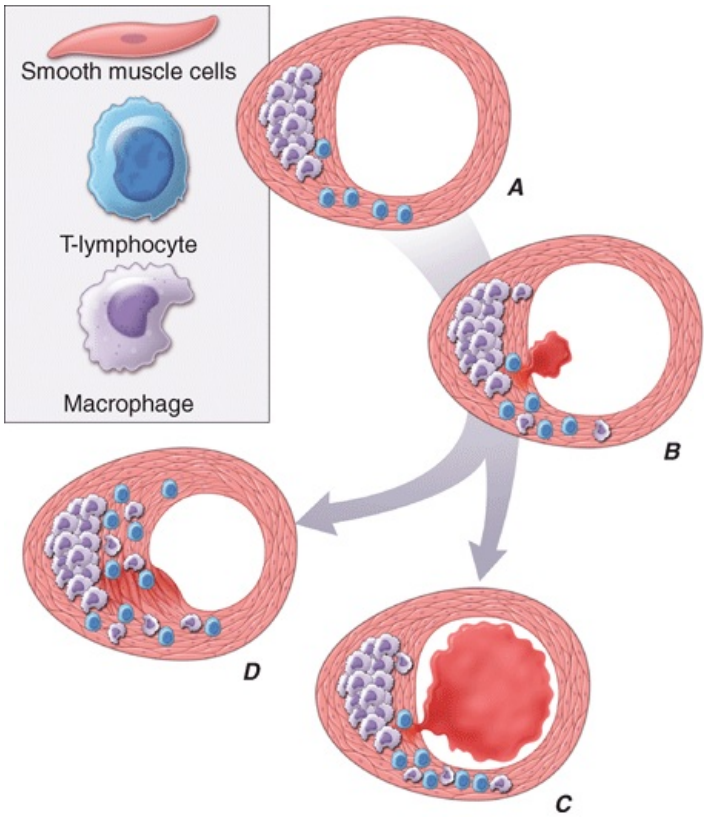

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Plaque rupture, thrombosis, and healing. *A.* Arterial remodeling during atherogenesis. During the initial part of the life history of an atheroma, growth is often outward, preserving the caliber of the

initial part of the life history of an atheroma, growth is often outward, preserving the caliber of the lumen. This phenomenon of "compensatory enlargement" accounts in part for the tendency of coronary arteriography to underestimate the degree of atherosclerosis. ***B.*** Rupture of the plaque's fibrous cap causes thrombosis. Physical disruption of the atherosclerotic plaque commonly causes arterial thrombosis by allowing blood coagulant factors to contact thrombogenic collagen found in the arterial extracellular matrix and tissue factor produced by macrophage-derived foam cells in the lipid core of lesions. In this manner, sites of plaque rupture form the nidus for thrombi. The normal artery wall has several fibrinolytic or antithrombotic mechanisms that tend to resist thrombosis and lyse clots that begin to form in situ. Such antithrombotic or thrombolytic molecules include thrombomodulin, tissue- and urokinase-type plasminogen activators, heparan sulfate proteoglycans, prostacyclin, and nitric oxide. ***C.*** When the clot overwhelms the endogenous fibrinolytic mechanisms, it may propagate and lead to arterial occlusion. The consequences of this occlusion depend on the degree of existing collateral vessels. In a patient with chronic multivessel occlusive coronary artery disease (CAD), collateral channels have often formed. In such circumstances, even a total arterial occlusion may not lead to myocardial infarction (MI), or it may produce an unexpectedly modest or a non-ST-segment elevation infarct because of collateral flow. In a patient with less advanced disease and without substantial stenotic lesions to provide a stimulus for collateral vessel formation, sudden plaque rupture and arterial occlusion commonly produces an ST-segment elevation infarction. These are the types of patients who may present with MI or sudden death as a first manifestation of coronary atherosclerosis. In some cases, the thrombus may lyse or organize into a mural thrombus without occluding the vessel. Such instances may be clinically silent. ***D.*** The subsequent thrombin-induced fibrosis and healing causes a fibroproliferative response that can lead to a more fibrous lesion that can produce an eccentric plaque that causes a hemodynamically significant stenosis. In this way, a nonocclusive mural thrombus, even if clinically silent or causing unstable angina rather than infarction, can provoke a healing response that can promote lesion fibrosis and luminal encroachment. Such a sequence of events may convert a "vulnerable" atheroma with a thin fibrous cap that is prone to rupture into a more "stable" fibrous plaque with a reinforced cap. Angioplasty of unstable coronary lesions may "stabilize" the lesions by a similar mechanism, producing a wound followed by healing.

MICROVESSELS

As atherosclerotic lesions advance, abundant plexuses of microvessels develop in connection with the artery's vasa vasorum. Newly developing microvascular networks may contribute to lesion complications in several ways. These blood vessels provide an abundant surface area for leukocyte trafficking and may serve as the portal for entry and exit of white blood cells from the established atheroma. Microvessels in the plaques may also furnish foci for intraplaque hemorrhage. Like the neovessels in the diabetic retina, microvessels in the atheroma may be friable and prone to rupture and can produce focal hemorrhage. Such a vascular leak can provoke thrombosis in situ, yielding local thrombin generation, which in turn can activate smooth muscle and endothelial cells through ligation of protease-activated receptors. Atherosclerotic plaques often contain fibrin and hemosiderin, an indication that episodes of intraplaque hemorrhage contribute to plaque complications.

Calcification

As they advance, atherosclerotic plaques also accumulate *calcium*. Proteins usually found in bone also localize in atherosclerotic lesions (e.g., osteocalcin, osteopontin, and bone morphogenetic proteins). Mineralization of the atherosclerotic plaque recapitulates many aspects of bone formation, including the regulatory participation of transcription factors such as Runx2.

PLAQUE EVOLUTION

Although atherosclerosis research has focused much attention on proliferation of smooth muscle cells, as in the case of macrophages, smooth muscle cells also can undergo apoptosis in the atherosclerotic plaque. Indeed, complex atheromata often have a mostly fibrous character and lack the cellularity of less advanced lesions. This relative paucity of smooth muscle cells in advanced atheromata may result from the predominance of cytostatic mediators such as TGF-β and IFN-γ (which can inhibit smooth muscle cell proliferation), and also from smooth muscle cell apoptosis. Some of the same proinflammatory cytokines that activate atherogenic functions of vascular wall cells can also sensitize these cells to undergo apoptosis.

Thus, during the evolution of the atherosclerotic plaque, a complex balance between entry and egress of lipoproteins and leukocytes, cell proliferation and cell death, extracellular matrix production, and remodeling, as well as calcification and neovascularization, contribute to lesion formation. Multiple and often competing signals regulate these various cellular events. Many mediators related to atherogenic risk factors, including those derived from lipoproteins, cigarette smoking, and angiotensin II, provoke the production of proinflammatory cytokines and alter the behavior of the intrinsic vascular wall cells and infiltrating leukocytes that underlie the complex pathogenesis of these lesions. Thus, advances in vascular biology have led to increased understanding of the mechanisms that link risk factors to the pathogenesis of atherosclerosis and its complications.

Clinical Syndromes of Atherosclerosis

Atherosclerotic lesions occur ubiquitously in Western societies. Most atheromata produce no symptoms, and many never cause clinical manifestations. Numerous patients with diffuse atherosclerosis may succumb to unrelated illnesses without ever having experienced a clinically significant manifestation of atherosclerosis. What accounts for this variability in the clinical expression

of atherosclerotic disease?

Arterial remodeling during atheroma formation (Fig. 241-2*A*) represents a frequently overlooked but clinically important feature of lesion evolution. During the initial phases of atheroma development, the plaque usually grows outward, in an abluminal direction. Vessels affected by atherogenesis tend to increase in diameter, a phenomenon known as *compensatory enlargement*, a type of vascular remodeling. The growing atheroma does not encroach on the arterial lumen until the burden of atherosclerotic plaque exceeds ~40% of the area encompassed by the internal elastic lamina. Thus, during much of its life history, an atheroma will not cause stenosis that can limit tissue perfusion.

Flow-limiting stenoses commonly form later in the history of the plaque. Many such plaques cause stable syndromes such as demand-induced angina pectoris or intermittent claudication in the extremities. In the coronary circulation and other circulations, even total vascular occlusion by an atheroma does not invariably lead to infarction. The hypoxic stimulus of repeated bouts of ischemia characteristically induces formation of collateral vessels in the myocardium, mitigating the consequences of an acute occlusion of an epicardial coronary artery. By contrast, many lesions that cause acute or unstable atherosclerotic syndromes, particularly in the coronary circulation, may arise from atherosclerotic plaques that do not produce a flow-limiting stenosis. Such lesions may produce only minimal luminal irregularities on traditional angiograms and often do not meet the traditional criteria for "significance" by arteriography. Thrombi arising from such nonocclusive stenoses may explain the frequency of MI as an initial manifestation of coronary artery disease (CAD) (in at least one-third of cases) in patients who report no prior history of angina pectoris, a syndrome usually caused by flow-limiting stenoses.

PLAQUE INSTABILITY AND RUPTURE

Postmortem studies afford considerable insight into the microanatomic substrate underlying the "instability" of plaques that do not cause critical stenoses. A superficial erosion of the endothelium or a frank plaque rupture or fissure usually produces the thrombus that causes episodes of unstable angina pectoris or the occlusive and relatively persistent thrombus that causes acute MI (Fig. 241-2*B*). In the case of carotid atheromata, a deeper ulceration that provides a nidus for the formation of platelet thrombi may cause transient cerebral ischemic attacks.

Rupture of the plaque's fibrous cap (Fig. 241-2*C*) permits contact between coagulation factors in the blood and highly thrombogenic tissue factor expressed by macrophage foam cells in the plaque's lipid-rich core. If the ensuing thrombus is nonocclusive or transient, the episode of plaque disruption may not cause symptoms or may result in episodic ischemic symptoms such as rest angina. Occlusive thrombi that endure often cause acute MI, particularly in the absence of a well-developed collateral circulation that supplies the affected territory. Repetitive episodes of plaque disruption and healing provide one likely mechanism of transition of the fatty streak to a more complex fibrous lesion (Fig. 241-2*D*). The healing process in arteries, as in skin wounds, involves the laying down of new extracellular matrix and fibrosis.

Not all atheromata exhibit the same propensity to rupture. Pathologic studies of culprit lesions that have caused acute MI reveal several characteristic features. Plaques that have caused fatal thromboses tend to have thin fibrous caps, relatively large lipid cores, and a high content of macrophages. Morphometric studies of such culprit lesions show that at sites of plaque rupture, macrophages and T lymphocytes predominate and contain relatively few smooth muscle cells. The cells that concentrate at sites of plaque rupture bear markers of inflammatory activation. In addition, patients with active atherosclerosis and acute coronary syndromes display signs of disseminated inflammation. For example, atherosclerotic plaques and even microvascular endothelial cells at sites remote from the "culprit" lesion of an acute coronary syndrome can exhibit markers of inflammatory activation.

Inflammatory mediators regulate processes that govern the integrity of the plaque's fibrous cap and, hence, its propensity to rupture. For example, the T cell-derived cytokine IFN-γ, which is found in atherosclerotic plaques, can inhibit growth and collagen synthesis of smooth muscle cells, as noted above. Cytokines derived from activated macrophages and lesional T cells can boost production of proteolytic enzymes that can degrade the extracellular matrix of the plaque's fibrous cap. Thus, inflammatory mediators can impair the collagen synthesis required for maintenance and repair of the fibrous cap and trigger degradation of extracellular matrix macromolecules, processes that weaken the plaque's fibrous cap and enhance its susceptibility to rupture (so-called vulnerable plaques). In contrast to plaques with these features of vulnerability, those with a dense extracellular matrix and relatively thick fibrous cap without substantial tissue factor–rich lipid cores seem generally resistant to rupture and unlikely to provoke thrombosis.

Features of the biology of the atheromatous plaque, in addition to its degree of luminal encroachment, influence the clinical manifestations of this disease. This enhanced understanding of plaque biology provides insight into the diverse ways in which atherosclerosis can present clinically and the reasons why the disease may remain silent or stable for prolonged periods, punctuated by acute complications at certain times. Increased understanding of atherogenesis provides new insight into the mechanisms

linking it to the risk factors discussed below, indicates the ways in which current therapies may improve outcomes, and suggests new targets for future intervention.

PREVENTION AND TREATMENT

The Concept of Atherosclerotic Risk Factors

The systematic study of risk factors for atherosclerosis emerged from a coalescence of experimental results, as well as from cross-sectional and ultimately longitudinal studies in humans. The prospective, community-based Framingham Heart Study provided rigorous support for the concept that hypercholesterolemia, hypertension, and other factors correlate with cardiovascular risk. Similar observational studies performed worldwide bolstered the concept of "risk factors" for cardiovascular disease.

From a practical viewpoint, the cardiovascular risk factors that have emerged from such studies fall into two categories: those modifiable by lifestyle and/or pharmacotherapy, and those that are immutable, such as age and sex. The weight of evidence supporting various risk factors differs. For example, hypercholesterolemia and hypertension certainly predict coronary risk, but the magnitude of the contributions of other so-called nontraditional risk factors, such as levels of homocysteine, levels of lipoprotein (a) [Lp(a)], and infection, remains controversial. Moreover, some biomarkers that predict cardiovascular risk may not participate in the causal pathway for the disease or its complications. For example, recent genetic studies suggest that C-reactive protein (CRP) does not itself mediate atherogenesis, despite its ability to predict risk. **Table 241-1** lists the risk factors recognized by the current National Cholesterol Education Project Adult Treatment Panel III (ATP III). The sections below will consider some of these risk factors and approaches to their modification.

Table 241–1. Major Risk Factors (Exclusive of LDL Cholesterol) that Modify LDL Goals

Cigarette smoking
Hypertension (BP ≥140/90 mmHg or on antihypertensive medication)
Low HDL cholesterol* [<1.0 mmol/L (<40 mg/dL)]
Diabetes mellitus
Family history of premature CHD
CHD in male first-degree relative <55 years
CHD in female first-degree relative <65 years
Age (men ≥45 years; women ≥55 years)
Lifestyle risk factors
Obesity (BMI ≥30 kg/m^2)
Physical inactivity
Atherogenic diet
Emerging risk factors
Lipoprotein(a) Homocysteine
Prothrombotic factors
Proinflammatory factors
Impaired fasting glucose
Subclinical atherosclerosis

*HDL cholesterol ≥1.6 mmol/L (≥60 mg/dL) counts as a "negative" risk factor; its presence removes one risk factor from the total count.

Abbreviations: BMI, body mass index; BP, blood pressure; CHD, coronary heart disease; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

Source: Modified from Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), Executive Summary. (Bethesda, MD: National Heart, Lung and Blood Institute, National Institutes of Health, 2001. NIH Publication No. 01-3670.)

LIPID DISORDERS

Abnormalities in plasma lipoproteins and derangements in lipid metabolism rank among the most firmly established and best understood risk factors for atherosclerosis. Chapter 356 describes the lipoprotein

classes and provides a detailed discussion of lipoprotein metabolism. Current ATP III guidelines recommend lipid screening in all adults >20 years of age. The screen should include a fasting lipid profile (total cholesterol, triglycerides, LDL cholesterol, and HDL cholesterol) repeated every 5 years.

ATP III guidelines strive to match the intensity of treatment to an individual's risk. A quantitative estimate of risk places individuals in one of three treatment strata **(Table 241-2)**. The first step in applying these guidelines involves counting an individual's risk factors (Table 241-1). Individuals with fewer than two risk factors fall into the lowest treatment intensity stratum [LDL goal <4.1 mmol/L (<160 mg/dL)]. In those with two or more risk factors, the next step involves a simple calculation that estimates the 10-year risk of developing coronary heart disease (CHD) (Table 241-2); see http://www.nhlbi.nih.gov/guidelines/cholesterol/ for the algorithm and a downloadable risk calculator. Those with a 10-year risk ≤20% fall into the intermediate stratum [LDL goal <3.4 mmol/L (<130 mg/dL)]. Those with a calculated 10-year CHD risk of >20%, any evidence of established atherosclerosis, or diabetes (now considered a CHD risk-equivalent) fall into the most intensive treatment group [LDL goal <2.6 mmol/L (<100 mg/dL)]. Members of the ATP III panel recently suggested <1.8 mmol/L (<70 mg/dL) as a goal for very high-risk patients and an optional goal for high-risk patients based on recent clinical trial data (Table 241-2). Beyond the Framingham algorithm, there are multiple risk calculators for various countries or regions. Risk calculators that incorporate family history of premature (CAD) and a marker of inflammation (CRP) have been validated for U.S. women and men.

Table 241–2. LDL Cholesterol Goals and Cutpoints for Therapeutic Lifestyle Changes (TLC) and Drug Therapy in Different Risk Categories

	LDL Level, mmol/L (mg/dL)		
Risk Category	**Goal**	**Initiate TLC**	**Consider Drug Therapy**
Very high	<1.8 (<70)	≥1.8 (≥70)	≥1.8 (≥70)
ACS, or CHD w/DM, or multiple CRFs			
High	<2.6 (<100)	≥2.6 (≥100)	≥2.6 (≥100) [<2.6 (<100): consider drug Rx]
If LDL <2.6 (<100)	[optional goal: <1.8 (<70)] <1.8 (<70)		
Moderately high	<2.6 (<100)	≥3.4 (≥130)	≥3.4 (≥130) [2.6–3.3 (100–129): consider drug Rx]
2+ risk factors (10-year risk, 10–20%)			
Moderate	<3.4 (<130)	≥3.4 (≥130)	≥4.1 (≥160)
2+ risk factors (risk <10%)			
Lower	<4.1 (<160)	≥4.1 (≥160)	≥4.9 (≥190)
0–1 risk factor			

Abbreviations: ACS, acute coronary syndrome; CHD, coronary heart disease; CRFs, coronary risk factors; DM, diabetes mellitus; LDL, low-density lipoprotein.

Source: Adapted from S Grundy et al: Circulation 110:227, 2004.

The first maneuver to achieve the LDL goal involves therapeutic lifestyle changes (TLC), including specific diet and exercise recommendations established by the guidelines. According to ATP III criteria, those with LDL levels exceeding goal for their risk group by >0.8 mmol/L (>30 mg/dL) merit consideration for drug therapy. In patients with triglycerides >2.6 mmol/L (>200 mg/dL), ATP III guidelines specify a secondary goal for therapy: "non-HDL cholesterol" (simply, the HDL cholesterol level subtracted from the total cholesterol). Cutpoints for the therapeutic decision for non-HDL cholesterol are 0.8 mmol/L (30 mg/dL) more than those for LDL.

An extensive and growing body of rigorous evidence now supports the effectiveness of aggressive management of LDL. Addition of drug therapy to dietary and other nonpharmacologic measures reduces cardiovascular risk in patients with established coronary atherosclerosis and also in individuals who have not previously experienced CHD events **(Fig. 241-3)**. As guidelines often lag the emerging clinical trial evidence base, the practitioner may elect to exercise clinical judgment in making therapeutic decisions in individual patients.

Figure 241-3

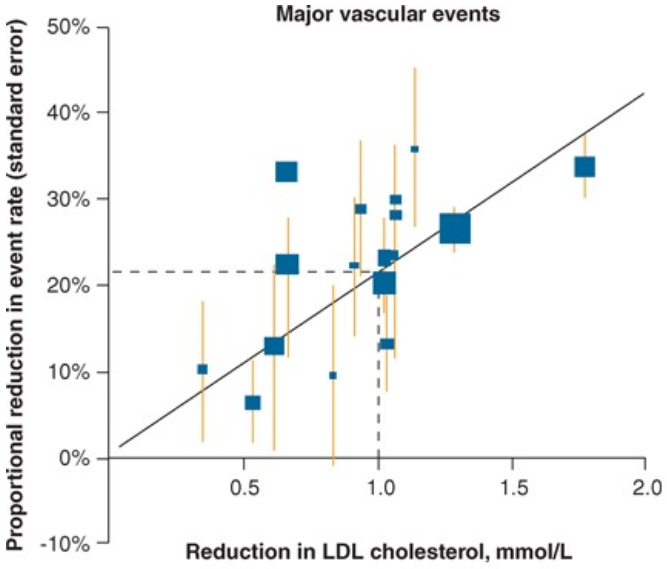

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lipid lowering reduces coronary events, as reflected on this graph showing the reduction in major cardiovascular events as a function of low-density lipoprotein level in a compendium of clinical trials with statins. *(Adapted from CTT Collaborators, Lancet 366:1267, 2005.)* The Management of Elevated Cholesterol in the Primary Prevention Group of Adult Japanese (MEGA), Treating to New Targets (TNT), and Incremental Decrease in Endpoints through Aggressive Lipid Lowering (IDEAL) studies have been added.

LDL-lowering therapies do not appear to exert their beneficial effect on cardiovascular events by causing a marked "regression" of stenoses. Angiographically monitored studies of lipid lowering have shown at best a modest reduction in coronary artery stenoses over the duration of study, despite abundant evidence of event reduction. These results suggest that the beneficial mechanism of lipid lowering does not require a substantial reduction in the fixed stenoses. Rather, the benefit may derive from "stabilization" of atherosclerotic lesions without decreased stenosis. Such stabilization of atherosclerotic lesions and the attendant decrease in coronary events may result from the egress of lipids or from favorably influencing aspects of the biology of atherogenesis discussed above. In addition, as sizable lesions may protrude abluminally rather than into the lumen due to complementary enlargement, shrinkage of such plaques may not be apparent on angiograms. The consistent benefit of LDL lowering by 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors (statins) observed in many risk groups may depend not only on their salutary effects on the lipid profile but also on direct modulation of plaque biology independent of lipid lowering.

A new class of LDL-lowering medications reduces cholesterol absorption from the proximal small bowel by targeting an enterocyte cholesterol transporter denoted Niemann-Pick C1-like 1 protein (NPC1L1). The NPC1L1 inhibitor ezetimibe provides a useful adjunct to current therapies to achieve LDL goals; however, no clinical trial evidence has yet demonstrated that ezetimibe improves CHD outcomes.

As the mechanism by which elevated LDL levels promote atherogenesis probably involves oxidative modification, several trials have tested the possibility that antioxidant vitamin therapy might reduce CHD events. Rigorous and well-controlled clinical trials have failed to demonstrate that antioxidant vitamin therapy improves CHD outcomes. Therefore, the current evidence base does *not* support the use of antioxidant vitamins for this indication.

The clinical use of effective pharmacologic strategies for lowering LDL has reduced cardiovascular events markedly, but even their optimal utilization in clinical trials prevents only a minority of these endpoints. Hence, other aspects of the lipid profile have become tempting targets for addressing the residual burden of cardiovascular disease that persists despite aggressive LDL lowering. Indeed, in the "poststatin" era, patients with LDL levels at or below target not infrequently present with acute coronary syndromes. Low levels of HDL present a growing problem in patients with CAD as the prevalence of metabolic syndrome and diabetes increases. Blood HDL levels vary inversely with those of triglycerides, and the independent role of triglycerides as a cardiovascular risk factor remains unsettled. For these reasons, approaches to raising HDL have emerged as a prominent next hurdle in the management of dyslipidemia. Weight loss and physical activity can raise HDL. Nicotinic acid, particularly in combination with statins, can robustly raise HDL. Some clinical trial data support the effectiveness of nicotinic acid in cardiovascular risk reduction. However, flushing and pruritus remain a challenge to patient acceptance, even with improved dosage forms of nicotinic acid. A combination of

nicotinic acid with an inhibitor of prostaglandin D receptor, a mediator of flushing, may limit this unwanted effect of nicotinic acid and is currently in clinical trials, but it has not received regulatory approval.

Agonists of nuclear receptors provide another potential avenue for raising HDL levels. Yet patients treated with peroxisome proliferator–activated receptors alpha and gamma (PPAR-α and -γ) agonists have not consistently shown improved cardiovascular outcomes, and at least some PPAR-agonists have been associated with worsened cardiovascular outcomes. Other agents in clinical development raise HDL levels by inhibiting cholesteryl ester transfer protein (CETP). The first of these agents to undergo large-scale clinical evaluation showed increased adverse events, leading to cessation of its development. Clinical studies currently underway will assess the effectiveness of other CETP inhibitors that lack some of the adverse off-target actions encountered with the first agent.

HYPERTENSION

(See also Chap. 247) A wealth of epidemiologic data support a relationship between hypertension and atherosclerotic risk, and extensive clinical trial evidence has established that pharmacologic treatment of hypertension can reduce the risk of stroke, heart failure, and CHD events.

DIABETES MELLITUS, INSULIN RESISTANCE, AND THE METABOLIC SYNDROME

(See also Chap. 344) Most patients with diabetes mellitus die of atherosclerosis and its complications. Aging and rampant obesity underlie a current epidemic of type 2 diabetes mellitus. The abnormal lipoprotein profile associated with insulin resistance, known as *diabetic dyslipidemia*, accounts for part of the elevated cardiovascular risk in patients with type 2 diabetes. Although diabetic individuals often have LDL cholesterol levels near the average, the LDL particles tend to be smaller and denser and, therefore, more atherogenic. Other features of diabetic dyslipidemia include low HDL and elevated triglyceride levels. Hypertension also frequently accompanies obesity, insulin resistance, and dyslipidemia. Indeed, the ATP III guidelines now recognize this cluster of risk factors and provide criteria for diagnosis of the "metabolic syndrome" **(Table 241-3)**. Despite legitimate concerns about whether clustered components confer more risk than an individual component, the metabolic syndrome concept may offer clinical utility.

Table 241–3. Clinical Identification of the Metabolic Syndrome—Any Three Risk Factors

Risk Factor	Defining Level
Abdominal obesity[a]	
Men (waist circumference)[b]	>102 cm (>40 in.)
Women	>88 cm (>35 in.)
Triglycerides	>1.7 mmol/L (>150 mg/dL)
HDL cho-lesterol	
Men	<1 mmol/L (<40 mg/dL)
Women	<1.3 mmol/L (<50 mg/dL)
Blood pressure	≥130/≥85 mmHg
Fasting glucose	>6.1 mmol/L (>110 mg/dL)

[a]Overweight and obesity are associated with insulin resistance and the metabolic syndrome. However, the presence of abdominal obesity is more highly correlated with the metabolic risk factors than is an elevated body mass index (BMI). Therefore, the simple measure of waist circumference is recommended to identify the BMI component of the metabolic syndrome.

[b]Some male patients can develop multiple metabolic risk factors when the waist circumference is only marginally increased [e.g., 94–102 cm (37–39 in.)]. Such patients may have a strong genetic contribution to insulin resistance. They should benefit from lifestyle changes, similarly to men with categorical increases in waist circumference.

Therapeutic objectives for intervention in these patients include addressing the underlying causes, including obesity and low physical activity, by initiating TLC. The ATP III guidelines provide an explicit step-by-step plan for implementing TLC, and treatment of the component risk factors should accompany TLC. Establishing that strict glycemic control reduces the risk of macrovascular complications of diabetes has proved much more elusive than the established beneficial effects on microvascular complications such as retinopathy and renal disease. Indeed, "tight" glycemic control may increase adverse events in patients with type 2 diabetes, lending even greater importance to aggressive control of other aspects of risk in this patient population. In this regard, multiple clinical trials, including the Collaborative Atorvastatin Diabetes Study (CARDS) that addressed specifically the diabetic population, have demonstrated unequivocal benefit of HMG-CoA reductase inhibitor therapy in diabetic patients over all ranges of LDL cholesterol levels (but not those with end-stage renal

disease). In view of the consistent benefit of statin treatment for diabetic populations and the thus far equivocal results with PPAR agonists, the current stance of the American Diabetic Association that statins be considered for persons with diabetes older than age 40 who have a total cholesterol level ≥ 135 appears amply justified. Among the oral hypoglycemic agents, metformin possesses the best evidence base for cardiovascular event reduction.

Diabetic populations appear to derive particular benefit from antihypertensive strategies that block the action of angiotensin II. Thus, the antihypertensive regimen for patients with the metabolic syndrome should include angiotensin converting-enzyme inhibitors or angiotensin receptor blockers when possible. Most of these individuals will require more than one antihypertensive agent to achieve the recently updated American Diabetes Association blood pressure goal of 130/80 mmHg.

MALE SEX/POSTMENOPAUSAL STATE

Decades of observational studies have verified excess coronary risk in men compared with premenopausal women. After menopause, however, coronary risk accelerates in women. At least part of the apparent protection against CHD in premenopausal women derives from their relatively higher HDL levels compared with those of men. After menopause, HDL values fall in concert with increased coronary risk. Estrogen therapy lowers LDL cholesterol and raises HDL cholesterol, changes that should decrease coronary risk.

Multiple observational and experimental studies have suggested that estrogen therapy reduces coronary risk. However, a spate of clinical trials has failed to demonstrate a net benefit of estrogen with or without progestins on CHD outcomes. In the Heart and Estrogen/Progestin Replacement Study (HERS), postmenopausal female survivors of acute MI were randomized to an estrogen/progestin combination or to placebo. This study showed no overall reduction in recurrent coronary events in the active treatment arm. Indeed, early in the 5-year course of this trial, there was a trend toward an actual increase in vascular events in the treated women. Extended follow-up of this cohort did not disclose an accrual of benefit in the treatment group. The Women's Health Initiative (WHI) study arm, using a similar estrogen plus progesterone regimen, was halted due to a small but significant hazard of cardiovascular events, stroke, and breast cancer. The estrogen without progestin arm of WHI (conducted in women without a uterus) was stopped early due to an increase in strokes, and failed to afford protection from MI or CHD death during observation over 7 years. The excess cardiovascular events in these trials may result from an increase in thromboembolism (Chap. 348). Physicians should work with women to provide information and help weigh the small but evident CHD risk of estrogen ± progestin versus the benefits for postmenopausal symptoms and osteoporosis, taking personal preferences into account. Post hoc analyses of observational studies suggest that estrogen therapy in women younger than or closer to menopause than the women enrolled in WHI might confer cardiovascular benefit. Thus, the timing in relation to menopause or the age at which estrogen therapy begins may influence its risk/benefit balance.

The lack of efficacy of estrogen therapy in cardiovascular risk reduction highlights the need for redoubled attention to known modifiable risk factors in women. The recent JUPITER trials randomized over 6000 women over age 65 without known cardiovascular disease with LDL <130 mg/dL and high-sensitivity (hs) CRP >2 mg/L to a statin or placebo. The statin-treated women had a striking reduction in cardiovascular events, as did the men. This trial, which included more women than any prior statin study, provides strong evidence supporting the efficacy of statins in women who meet those entry criteria.

DYSREGULATED COAGULATION OR FIBRINOLYSIS

Thrombosis ultimately causes the gravest complications of atherosclerosis. The propensity to form thrombi and/or lyse clots once they form clearly influences the manifestations of atherosclerosis. Thrombosis provoked by atheroma rupture and subsequent healing may promote plaque growth. Certain individual characteristics can influence thrombosis or fibrinolysis and have received attention as potential coronary risk factors. For example, fibrinogen levels correlate with coronary risk and provide information about coronary risk independent of the lipoprotein profile.

The stability of an arterial thrombus depends on the balance between fibrinolytic factors such as plasmin, and inhibitors of the fibrinolytic system such as plasminogen activator inhibitor 1 (PAI-1). Individuals with diabetes mellitus or the metabolic syndrome have elevated levels of PAI-1 in plasma, and this probably contributes to the increased risk of thrombotic events. Lp(a) (Chap. 356) may modulate fibrinolysis, and individuals with elevated Lp(a) levels have increased CHD risk.

Aspirin reduces CHD events in several contexts. Chapter 243 discusses aspirin therapy in stable ischemic heart disease, Chap. 244 reviews recommendations for aspirin treatment in acute coronary syndromes, and Chap. 370 describes aspirin's role in preventing recurrent ischemic stroke. In primary prevention, pooled trial data show that low-dose aspirin treatment (81 mg/d to 325 mg on alternate days) can reduce the risk of a first MI in men. Although the recent Women's Health Study (WHS) showed that aspirin (100 mg on alternate days) reduced strokes by 17%, it did not prevent MI in women. Current American Heart Association (AHA) guidelines recommend the use of low-dose aspirin (75–160 mg/d) for women with high cardiovascular risk (≥20% 10-year risk), for men with a ≥10% 10-

year risk of CHD, and for all aspirin-tolerant patients with established cardiovascular disease who lack contraindications.

HOMOCYSTEINE

A large body of literature suggests a relationship between hyperhomocysteinemia and coronary events. Several mutations in the enzymes involved in homocysteine accumulation correlate with thrombosis and, in some studies, with coronary risk. Prospective studies have not shown a robust utility of hyperhomocysteinemia in CHD risk stratification. Clinical trials have not shown that intervention to lower homocysteine levels reduces CHD events. Fortification of the U.S. diet with folic acid to reduce neural tube defects has lowered homocysteine levels in the population at large. Measurement of homocysteine levels should be reserved for individuals with atherosclerosis at a young age or out of proportion to established risk factors. Physicians who advise consumption of supplements containing folic acid should consider that this treatment may mask pernicious anemia.

INFLAMMATION

An accumulation of clinical evidence shows that markers of inflammation correlate with coronary risk. For example, plasma levels of CRP, as measured by a high-sensitivity assay (hsCRP), prospectively predict the risk of MI. CRP levels also correlate with the outcome in patients with acute coronary syndromes. In contrast to several other novel risk factors, CRP adds predictive information to that derived from established risk factors, such as those included in the Framingham score **(Fig. 241-4)**. Recent Mendelian randomization studies do not support a causal role for CRP in cardiovascular disease. Thus, CRP serves as a validated biomarker of risk but probably not as a direct contributor to pathogenesis.

Figure 241-4

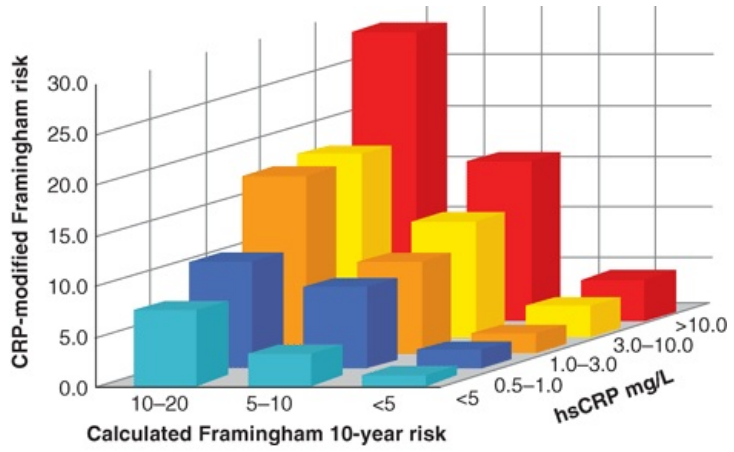

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C-reactive protein (CRP) level adds to the predictive value of the Framingham score. hsCRP, high-sensitivity measurement of CRP. *(Adapted from PM Ridker et al: Circulation 109:2818, 2004.)*

Elevations in acute-phase reactants such as fibrinogen and CRP reflect the overall inflammatory burden, not just vascular foci of inflammation. Visceral adipose tissue releases proinflammatory cytokines that drive CRP production and may represent a major extravascular stimulus to elevation of inflammatory markers in obese and overweight individuals. Indeed, CRP levels rise with body mass index (BMI), and weight reduction lowers CRP levels. Infectious agents might also furnish inflammatory stimuli related to cardiovascular risk. To date, randomized clinical trials have not supported the use of antibiotics to reduce CHD risk.

Intriguing evidence suggests that lipid-lowering therapy reduces coronary events in part by muting the inflammatory aspects of the pathogenesis of atherosclerosis. For example, in the JUPITER trial, a prespecified analysis showed that those who achieved lower levels of both LDL and CRP had better clinical outcomes than did those who only reached the lower level of either the inflammatory marker or the atherogenic lipoprotein **(Fig. 241-5)**. Similar analyses of studies of statin treatment in patients after acute coronary syndromes showed the same pattern. The anti-inflammatory effect of statins appears independent of LDL lowering, as these two variables correlated very poorly in individual subjects in multiple clinical trials.

Figure 241-5

Group	N	Rate
Placebo	7832	1.11
LDL ≥ 70 mg/dL, hsCRP ≥ 2 mg/L	1384	1.11
LDL < 70 mg/dL, hsCRP ≥ 2 mg/L	2921	0.62
LDL ≥ 70 mg/dL, hsCRP < 2 mg/L	726	0.54
LDL < 70 mg/dL, hsCRP < 2 mg/L	2685	0.38

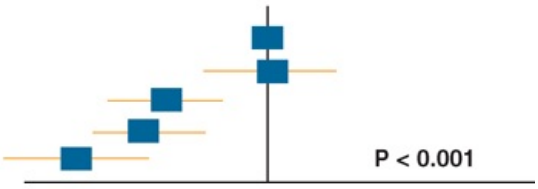

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evidence from the JUPITER study that both LDL-lowering and anti-inflammatory actions contribute to the benefit of statin therapy in primary prevention. See text for explanation. hsCRP, high-sensitivity measurement of C-reactive protein (CRP). *(Adapted from PM Ridker et al: Lancet 373:1175, 2009.)*

LIFESTYLE MODIFICATION

The prevention of atherosclerosis presents a long-term challenge to all health care professionals and for public health policy. Both individual practitioners and organizations providing health care should strive to help patients optimize their risk factor profiles long before atherosclerotic disease becomes manifest. The current accumulation of cardiovascular risk in youth and in certain minority populations presents a particularly vexing concern from a public health perspective.

The care plan for all patients seen by internists should include measures to assess and minimize cardiovascular risk. Physicians must counsel patients about the health risks of tobacco use and provide guidance and resources regarding smoking cessation. Similarly, physicians should advise all patients about prudent dietary and physical activity habits for maintaining ideal body weight. Both National Institutes of Health (NIH) and AHA statements recommend at least 30 minutes of moderate-intensity physical activity per day. Obesity, particularly the male pattern of centripetal or visceral fat accumulation, can contribute to the elements of the metabolic syndrome (Table 241-3). Physicians should encourage their patients to take personal responsibility for behavior related to modifiable risk factors for the development of premature atherosclerotic disease. Conscientious counseling and patient education may forestall the need for pharmacologic measures intended to reduce coronary risk.

ISSUES IN RISK ASSESSMENT

A growing panel of markers of coronary risk presents a perplexing array to the practitioner. Markers measured in peripheral blood include size fractions of LDL particles and concentrations of homocysteine, Lp(a), fibrinogen, CRP, PAI-1, myeloperoxidase, and lipoprotein-associated phospholipase A_2, among many others. In general, such specialized tests add little to the information available from a careful history and physical examination combined with measurement of a plasma lipoprotein profile and fasting blood glucose. The high-sensitivity CRP measurement may well prove an exception in view of its robustness in risk prediction, ease of reproducible and standardized measurement, relative stability in individuals over time, and, most important, ability to add to the risk information disclosed by standard measurements such as the components of the Framingham risk score (Fig. 241-4). The addition of information regarding a family history of premature atherosclerosis in parents (a simply obtained indicator of genetic susceptibility), together with the marker of inflammation hsCRP, permits correct reclassification of risk in individuals—especially those whose Framingham scores place them at intermediate risk. Current advisories, however, recommend the use of the hsCRP test only in individuals in this CHD event risk group (10–20%, 10-year risk).

Available data do not support the use of imaging studies to screen for subclinical disease (e.g., measurement of carotid-intima/media thickness, coronary artery calcification, and use of computed tomographic coronary angiograms). Inappropriate use of such imaging modalities may promote excessive alarm in asymptomatic individuals and prompt invasive diagnostic and therapeutic procedures of unproven value. Widespread application of such modalities for screening should await proof that clinical benefit derives from their application.

Progress in human genetics holds considerable promise for risk prediction and for individualization of cardiovascular therapy. Many reports have identified single-nucleotide polymorphisms (SNPs) in candidate genes as predictors of cardiovascular risk. To date, the validation of such genetic markers of risk and drug responsiveness in multiple populations has often proved disappointing. The advent of technology that permits relatively rapid and inexpensive genome-wide screens, in contrast to most SNP studies, has led to identification of sites of genetic variation that do reproducibly indicate heightened cardiovascular risk (e.g., chromosome 9p21). The results of genetic studies should identify new potential therapeutic targets (e.g., the enzyme mutated in autosomal dominant hypercholesterolemia, abbreviated *PCSK9*) and may lead to genetic tests that help refine cardiovascular risk assessment in the future.

The Challenge of Implementation: Changing Physician and Patient Behavior

Despite declining age-adjusted rates of coronary death, cardiovascular mortality worldwide is rising due to the aging of the population, and the subsiding of communicable diseases and increased

prevalence of risk factors in developing countries. Enormous challenges remain regarding translation of the current evidence base into practice. Physicians must learn how to help individuals adopt a healthy lifestyle in a culturally appropriate manner and to deploy their increasingly powerful pharmacologic tools most economically and effectively. The obstacles to implementation of current evidence-based prevention and treatment of atherosclerosis involve economics, education, physician awareness, and patient adherence to recommended regimens. Future goals in the treatment of atherosclerosis should include more widespread implementation of the current evidence-based guidelines regarding risk factor management and, when appropriate, drug therapy.

FURTHER READINGS

Alberti KG et al: Harmonizing the metabolic syndrome: A joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 120:1640, 2009[PMID: 19805654]

D'Agostino RB et al: General cardiovascular risk profile for use in primary care: The Framingham Heart Study. Circulation 117:743, 2008[PMID: 18212285]

Delahoy PJ et al: The relationship between reduction in low-density lipoprotein cholesterol by statins and reduction in risk of cardiovascular outcomes: An updated meta-analysis. Clin Ther 31:236, 2009[PMID: 19302897]

Hamsten A, Eriksson P: Identifying the susceptibility genes for coronary artery disease: From hyperbole through doubt to cautious optimism. J Intern Med 263:538, 2008[PMID: 18410597]

Libby P et al: Inflammation in atherosclerosis: From pathophysiology to practice. J Am Coll Cardiol 54:2129, 2009[PMID: 19942084]

Mosca L et al: Evidence-based guidelines for cardiovascular disease prevention in women: 2007 update. Circulation 115:1481, 2007[PMID: 17309915]

Ridker PM et al: C-reactive protein and parental history improve global cardiovascular risk prediction: The Reynolds Risk Score for men. Circulation 118:2243, 2008[PMID: 18997194]

Shao B, Heinecke JW: HDL, lipid peroxidation, and atherosclerosis. J Lipid Res 50:599, 2009[PMID: 19141435]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF ATHEROSCLEROSIS: INTRODUCTION

Much knowledge about the biology of human atherosclerosis and the risk factors for the disease has expanded considerably. The application of vascular biology to human atherosclerosis has revealed many new insights into the mechanisms that promote clinical events. The series of animated video presentations presented here illustrates some of the evolving information about risk factors for atherosclerosis and the pathophysiology of clinical events.

The importance of blood pressure as a risk factor for atherosclerosis and cardiovascular events has long been recognized. More recent clinical information has highlighted the importance of pulse pressure—the difference between the systolic pressure and minimum diastolic arterial pressure—as a prognostic indicator of cardiovascular risk. The video clip on pulse pressure explains the pathophysiology of this readily measured clinical variable.

Physicians possess a great deal of knowledge about the role of cholesterol in the prediction of atherosclerosis and its complications, but knowledge about the mechanism that links hypercholesterolemia to cardiovascular events has lagged the epidemiologic and observational findings. Low-density lipoprotein (LDL) provides an example of a well-understood cardiovascular risk factor. Several of the animations included in this series highlight the role of modified LDL as a trigger for inflammation and other aspects of the pathobiology of arterial plaques that lead to their aggravation and clinical events. Physicians have useful tools for modulating LDL, but other aspects of dyslipidemia are on the rise and provide a growing challenge to the practitioner. In particular, low levels of high-density lipoprotein (HDL) and elevated levels of triglycerides characterize the constellation of findings denoted by some as the "metabolic syndrome." In the wake of increasing obesity worldwide, these features of the lipoprotein profile require renewed focus. Several of the animations in this collection discuss the concept of the metabolic syndrome and the role of lipid profile components other than LDL in atherogenesis.

The traditional approach to atherosclerosis focused on arterial stenoses as a cause of ischemia and cardiovascular events. Physicians now have effective revascularization modalities for addressing flow-limiting stenoses, but atherosclerotic plaques that do not cause stenoses nonetheless may precipitate clinical events, such as unstable angina and acute myocardial infarction. Thus, it is necessary to add to the traditional focus on stenosis an enlarged appreciation of the pathobiology of atherosclerosis that underlies many acute coronary syndromes. The animation on the development and complication of atherosclerotic plaque explains some of these emerging concepts in plaque activation as they apply to the precipitation of thrombotic complications of atherosclerosis.

Video e32-1

Pulse pressure. Considerable evidence suggests that pulse pressure serves as an important risk factor for future cardiovascular events. This video clip explains the derivation of pulse pressure and some of the pathophysiology that determines this parameter. *(Animation by Animation MD; with permission.)*

Video e32-2

Plaque instability. Most coronary thromboses result from a physical disruption of the atherosclerotic plaque. This animation explains some of the current concepts of the pathophysiology of atherosclerotic plaque disruption and how it triggers arterial thrombosis.

Video e32-3

Lipoprotein menagerie. The lipid profile confers important information regarding cardiovascular risk and the effects of therapies; understanding lipoprotein metabolism provides insight into the pathophysiology of arterial disease. This animation presents the rudiments of lipoprotein metabolism that are important in clinical medicine.

Video e32-4

Formation and complication of atherosclerotic plaques. Physicians now understand the generation of atherosclerotic plaques as a dynamic process involving an interchange between cells of the artery wall, inflammatory cells recruited from blood, and risk factors such as lipoproteins. This animation

reviews current thinking about how risk factors alter the biology of the artery wall and can incite initiation and progression of atherosclerosis. It also discusses the importance of inflammation in these processes and portrays the role of inflammation in plaque disruption and thrombosis. Finally, this animation depicts the concept of stabilization of atherosclerotic plaques by interventions such as lipid lowering.

Video e32-5

Atherogenesis. This video clip highlights some of the current thinking about mechanisms of atherogenesis.

Video e32-6

Metabolic syndrome. A number of important cardiovascular risk factors tend to cluster in a pattern that has been described by some as the metabolic syndrome. Although controversy persists regarding whether cardiovascular risk due to these factors is additive or synergistic, their clinical importance is growing. This animation discusses some of the metabolic derangements that underlie the metabolic syndrome.

ACKNOWLEDGMENTS

From Peter Libby, MD: Changes and Challenges in Cardiovascular Protection: A Special CME Activity for Physicians. Created under an unrestricted educational grant from Merck & Co., Inc. Copyright © 2002, Cardinal Health; used with permission.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE METABOLIC SYNDROME: INTRODUCTION

The metabolic syndrome (syndrome X, insulin resistance syndrome) consists of a constellation of metabolic abnormalities that confer increased risk of cardiovascular disease (CVD) and diabetes mellitus (DM). The criteria for the metabolic syndrome have evolved since the original definition by the World Health Organization in 1998, reflecting growing clinical evidence and analysis by a variety of consensus conferences and professional organizations. The major features of the metabolic syndrome include central obesity, hypertriglyceridemia, low high-density lipoprotein (HDL) cholesterol, hyperglycemia, and hypertension **(Table 242-1)**.

Table 242–1. NCEP:ATPIII 2001 and IDF Criteria for the Metabolic Syndrome

NCEP:ATPIII 2001	**IDF Criteria for Central Adiposity[a]**		
Three or more of the following:	**Waist circumference**		
Central obesity: Waist circumference >102 cm (M), >88 cm (F)	**Men**	**Women**	**Ethnicity**
Hypertriglyceridemia: Triglycerides ≥150 mg/dL or specific medication	≥94 cm	≥80 cm	Europid, Sub-Saharan African, Eastern and Middle Eastern
Low HDL cholesterol: <40 mg/dL and <50 mg/dL, respectively, or specific medication	≥90 cm	≥80 cm	South Asian, Chinese, and ethnic South and Central American
Hypertension: Blood pressure ≥130 mm systolic or ≥85 mm diastolic or specific medication	≥85 cm	≥90 cm	Japanese
Fasting plasma glucose ≥100 mg/dL or specific medication or previously diagnosed Type 2 diabetes	**Two or more of the following:**		
	Fasting triglycerides >150 mg/dL or specific medication		
	HDL cholesterol <40 mg/dL and <50 mg/dL for men and women, respectively, or specific medication		
	Blood pressure >130 mm systolic or >85 mm diastolic or previous diagnosis or specific medication		
	Fasting plasma glucose ≥100 mg/dL or previously diagnosed Type 2 diabetes		

[a]In this analysis, the following thresholds for waist circumference were used: white men, ≥94 cm; African-American men, ≥94 cm; Mexican-American men, ≥90 cm; white women, ≥80 cm; African-American women, ≥80 cm; Mexican-American women, ≥80 cm. For participants whose designation was "other race—including multiracial," thresholds that were once based on Europid cut points (≥94 cm for men and ≥80 cm for women) and once based on South Asian cut points (≥90 cm for men and ≥80 cm for women) were used. For participants who were considered "other Hispanic," the IDF thresholds for ethnic South and Central Americans were used. ***Abbreviations:*** HDL, high-density lipoprotein; IDF, International Diabetes Foundation; NCEP:ATPIII, National Cholesterol Education Program, Adult Treatment Panel III.

Epidemiology

The prevalence of metabolic syndrome varies around the world, in part reflecting the age and ethnicity of the populations studied and the diagnostic criteria applied. In general, the prevalence of metabolic syndrome increases with age. The highest recorded prevalence worldwide is in Native Americans, with nearly 60% of women ages 45–49 and 45% of men ages 45–49 meeting National Cholesterol Education Program and Adult Treatment Panel III (NCEP:ATPIII) criteria. In the United States, metabolic syndrome is less common in African-American men and more common in Mexican-American women. Based on data from the National Health and Nutrition Examination Survey (NHANES) 1999–2000, the age-adjusted prevalence of the metabolic syndrome in United States adults who did not have diabetes is 28% for men and 30% for women. In France, a cohort 30 to 60 years old has shown a <10%

prevalence for each sex, although 17.5% are affected in the age range 60–64. Greater industrialization worldwide is associated with rising rates of obesity, which is anticipated to increase prevalence of the metabolic syndrome dramatically, especially as the population ages. Moreover, the rising prevalence and severity of obesity in children is initiating features of the metabolic syndrome in a younger population.

The frequency distribution of the five components of the syndrome for the U.S. population (NHANES III) is summarized in **Fig. 242-1**. Increases in waist circumference predominate in women, whereas fasting triglycerides >150 mg/dL and hypertension are more likely in men.

Figure 242-1

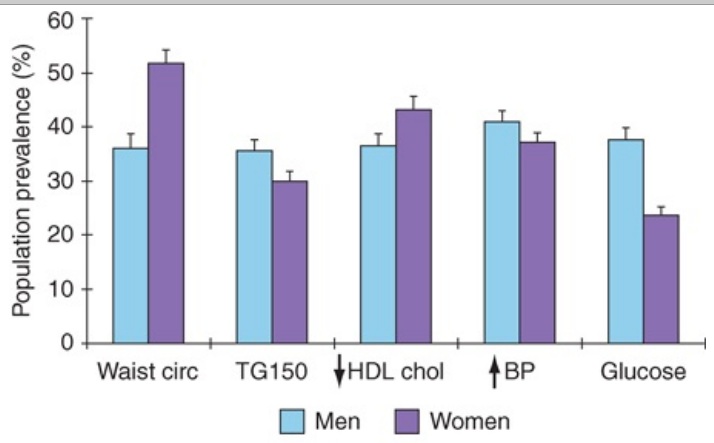

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prevalence of the metabolic syndrome components, from NHANES III. NHANES, National Health and Nutrition Examination Survey; TG, triglyceride; HDL, high-density lipoprotein; BP, blood pressure. The prevalence of elevated glucose includes individuals with known diabetes mellitus. *(Created from data in ES Ford et al: Diabetes Care 27:2444, 2004.)*

Risk Factors

OVERWEIGHT/OBESITY

Although the first description of the metabolic syndrome occurred in the early twentieth century, the worldwide overweight/obesity epidemic has been the driving force for more recent recognition of the syndrome. Central adiposity is a key feature of the syndrome, reflecting the fact that the syndrome's prevalence is driven by the strong relationship between waist circumference and increasing adiposity. However, despite the importance of obesity, patients who are normal weight may also be insulin-resistant and have the syndrome.

SEDENTARY LIFESTYLE

Physical inactivity is a predictor of CVD events and related mortality rate. Many components of the metabolic syndrome are associated with a sedentary lifestyle, including increased adipose tissue (predominantly central), reduced HDL cholesterol, and a trend toward increased triglycerides, high blood pressure, and increased glucose in the genetically susceptible. Compared with individuals who watched television or videos or used the computer <1 h daily, those who carried out those behaviors for >4 h daily had a twofold increased risk of the metabolic syndrome.

AGING

The metabolic syndrome affects 44% of the U.S. population older than age 50. A greater percentage of women over age 50 have the syndrome than men. The age dependency of the syndrome's prevalence is seen in most populations around the world.

DIABETES MELLITUS

DM is included in both the NCEP and International Diabetes Foundation (IDF) definitions of the metabolic syndrome. It is estimated that the great majority (~75%) of patients with Type 2 diabetes or impaired glucose tolerance (IGT) have the metabolic syndrome. The presence of the metabolic syndrome in these populations relates to a higher prevalence of CVD compared with patients with Type 2 diabetes or IGT without the syndrome.

CORONARY HEART DISEASE

The approximate prevalence of the metabolic syndrome in patients with coronary heart disease (CHD) is 50%, with a prevalence of 37% in patients with premature coronary artery disease (≤age 45), particularly in women. With appropriate cardiac rehabilitation and changes in lifestyle (e.g., nutrition,

physical activity, weight reduction, and, in some cases, pharmacologic agents), the prevalence of the syndrome can be reduced.

LIPODYSTROPHY

Lipodystrophic disorders in general are associated with the metabolic syndrome. Both genetic (e.g., Berardinelli-Seip congenital lipodystrophy, Dunnigan familial partial lipodystrophy) and acquired (e.g., HIV-related lipodystrophy in patients treated with highly active antiretroviral therapy) forms of lipodystrophy may give rise to severe insulin resistance and many of the components of the metabolic syndrome.

Etiology

INSULIN RESISTANCE

The most accepted and unifying hypothesis to describe the pathophysiology of the metabolic syndrome is insulin resistance, which is caused by an incompletely understood defect in insulin action (Chap. 344). The onset of insulin resistance is heralded by postprandial hyperinsulinemia, followed by fasting hyperinsulinemia and, ultimately, hyperglycemia.

An early major contributor to the development of insulin resistance is an overabundance of circulating fatty acids **(Fig. 242-2)**. Plasma albumin-bound free fatty acids (FFAs) are derived predominantly from adipose tissue triglyceride stores released by lipolytic enzymes lipase. Fatty acids are also derived from the lipolysis of triglyceride-rich lipoproteins in tissues by lipoprotein lipase (LPL). Insulin mediates both antilipolysis and the stimulation of LPL in adipose tissue. Of note, the inhibition of lipolysis in adipose tissue is the most sensitive pathway of insulin action. Thus, when insulin resistance develops, increased lipolysis produces more fatty acids, which further decrease the antilipolytic effect of insulin. Excessive fatty acids enhance substrate availability and create insulin resistance by modifying downstream signaling. Fatty acids impair insulin-mediated glucose uptake and accumulate as triglycerides in both skeletal and cardiac muscle, whereas increased glucose production and triglyceride accumulation are seen in liver.

Figure 242-2

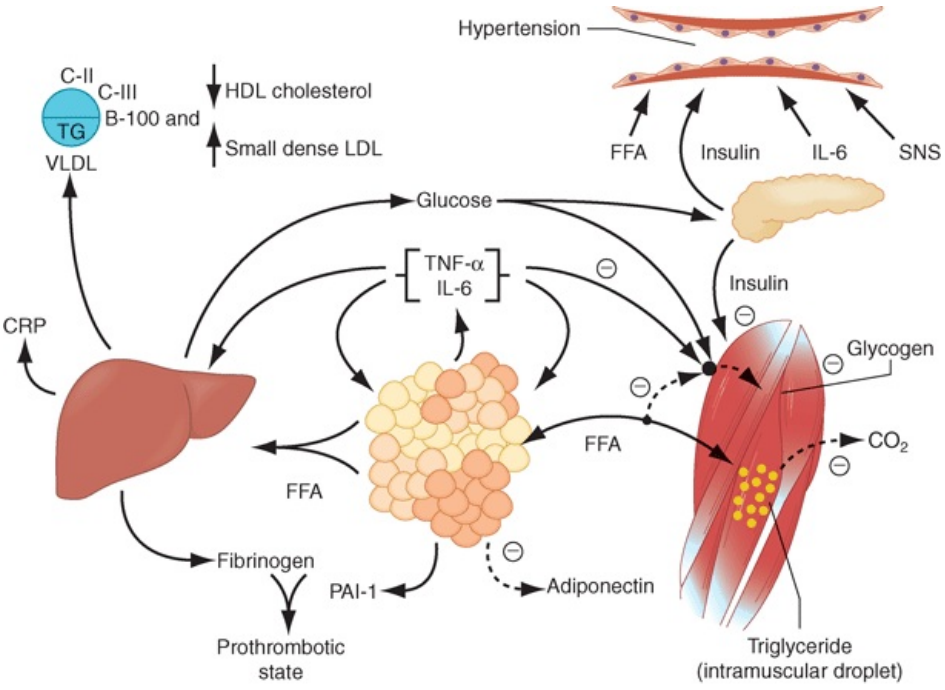

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathophysiology of the metabolic syndrome. Free fatty acids (FFAs) are released in abundance from an expanded adipose tissue mass. In the liver, FFAs result in an increased production of glucose and triglycerides and secretion of very low density lipoproteins (VLDLs). Associated lipid/lipoprotein abnormalities include reductions in high-density lipoprotein (HDL) cholesterol and an increased density of low-density lipoproteins (LDLs). FFAs also reduce insulin sensitivity in muscle by inhibiting insulin-mediated glucose uptake. Associated defects include a reduction in glucose partitioning to glycogen and increased lipid accumulation in triglyceride (TG). Increases in circulating glucose, and to some extent FFA, increase pancreatic insulin secretion, resulting in hyperinsulinemia. Hyperinsulinemia may result in enhanced sodium reabsorption and increased sympathetic nervous system (SNS) activity and contribute to the hypertension, as might increased levels of circulating FFAs. The proinflammatory state is superimposed and contributory to the insulin resistance produced by excessive FFAs. The enhanced secretion of interleukin 6 (IL-6) and tumor necrosis factor (TNF-α)

by excessive FFA. The enhanced secretion of interleukin 6 (IL-6) and tumor necrosis factor (TNF-α) produced by adipocytes and monocyte-derived macrophages results in more insulin resistance and lipolysis of adipose tissue triglyceride stores to circulating FFAs. IL-6 and other cytokines also enhance hepatic glucose production, VLDL production by the liver, and insulin resistance in muscle. Cytokines and FFAs also increase the hepatic production of fibrinogen and adipocyte production of plasminogen activator inhibitor 1 (PAI-1), resulting in a prothrombotic state. Higher levels of circulating cytokines also stimulate the hepatic production of C-reactive protein (CRP). Reduced production of the anti-inflammatory and insulin-sensitizing cytokine adiponectin is also associated with the metabolic syndrome. *(Reprinted from Eckel et al., with permission from Elsevier.)*

The oxidative stress hypothesis provides a unifying theory for aging and the predisposition to the metabolic syndrome. In studies carried out in insulin-resistant subjects with obesity or Type 2 diabetes, the offspring of patients with Type 2 diabetes, and the elderly, a defect has been identified in mitochondrial oxidative phosphorylation, leading to the accumulation of triglycerides and related lipid molecules in muscle. The accumulation of lipids in muscle is associated with insulin resistance.

INCREASED WAIST CIRCUMFERENCE

Waist circumference is an important component of the most recent and frequently applied diagnostic criteria for the metabolic syndrome. However, measuring waist circumference does not reliably distinguish increases in subcutaneous adipose tissue vs. visceral fat; this distinction requires CT or MRI. With increases in visceral adipose tissue, adipose tissue-derived FFAs are directed to the liver. In contrast, increases in abdominal subcutaneous fat release lipolysis products into the systemic circulation and avoid more direct effects on hepatic metabolism. Relative increases in visceral versus subcutaneous adipose tissue with increasing waist circumference in Asians and Asian Indians may explain the greater prevalence of the syndrome in those populations compared with African-American men in whom subcutaneous fat predominates. It is also possible that visceral fat is a marker for, but not the source of, excess postprandial FFAs in obesity.

DYSLIPIDEMIA

(See also Chap. 356) In general, FFA flux to the liver is associated with increased production of apoB-containing, triglyceride-rich very low density lipoproteins (VLDLs). The effect of insulin on this process is complex, but *hypertriglyceridemia* is an excellent marker of the insulin-resistant condition.

The other major lipoprotein disturbance in the metabolic syndrome is a *reduction in HDL cholesterol*. This reduction is a consequence of changes in HDL composition and metabolism. In the presence of hypertriglyceridemia, a decrease in the cholesterol content of HDL is a consequence of reduced cholesteryl ester content of the lipoprotein core in combination with cholesteryl ester transfer protein–mediated alterations in triglyceride, making the particle small and dense. This change in lipoprotein composition also results in increased clearance of HDL from the circulation. The relationships of these changes in HDL to insulin resistance are probably indirect, occurring in concert with the changes in triglyceride-rich lipoprotein metabolism.

In addition to HDL, low-density lipoproteins (LDLs) are modified in composition. With fasting serum triglycerides >2.0 mM (~180 mg/dL), there is almost always a predominance of small dense LDLs. Small dense LDLs are thought to be more atherogenic. They may be toxic to the endothelium, and they are able to transit through the endothelial basement membrane and adhere to glycosaminoglycans. They also have increased susceptibility to oxidation and are selectively bound to scavenger receptors on monocyte-derived macrophages. Subjects with increased small dense LDL particles and hypertriglyceridemia also have increased cholesterol content of both VLDL1 and VLDL2 subfractions. This relatively cholesterol-rich VLDL particle may contribute to the atherogenic risk in patients with metabolic syndrome.

GLUCOSE INTOLERANCE

(See also Chap. 344) The defects in insulin action lead to impaired suppression of glucose production by the liver and kidney and reduced glucose uptake and metabolism in insulin-sensitive tissues, i.e., muscle and adipose tissue. The relationship between impaired fasting glucose (IFG) or impaired glucose tolerance (IGT) and insulin resistance is well supported by human, nonhuman primate, and rodent studies. To compensate for defects in insulin action, insulin secretion and/or clearance must be modified to sustain euglycemia. Ultimately, this compensatory mechanism fails, usually because of defects in insulin secretion, resulting in progress from IFG and/or IGT to DM.

HYPERTENSION

The relationship between insulin resistance and hypertension is well established. Paradoxically, under normal physiologic conditions, insulin is a vasodilator with secondary effects on sodium reabsorption in the kidney. However, in the setting of insulin resistance, the vasodilatory effect of insulin is lost but the renal effect on sodium reabsorption is preserved. Sodium reabsorption is increased in whites with the metabolic syndrome but not in Africans or Asians. Insulin also increases the activity of the sympathetic nervous system, an effect that also may be preserved in the setting of the insulin resistance. Finally, insulin resistance is characterized by pathway-specific impairment in phosphatidylinositol-3-kinase signaling. In the endothelium, this may cause an imbalance between the production of nitric oxide and the secretion of endothelin 1, leading to decreased blood flow. Although these mechanisms are

provocative, when insulin action is assessed by levels of fasting insulin or by the Homeostasis Model Assessment (HOMA), insulin resistance contributes only modestly to the increased prevalence of hypertension in the metabolic syndrome.

PROINFLAMMATORY CYTOKINES

The increases in proinflammatory cytokines, including interleukin (IL)-1, IL-6, IL-18, resistin, tumor necrosis factor (TNF) α, and C-reactive protein (CRP), reflect overproduction by the expanded adipose tissue mass **(Fig. 242-2**). Adipose tissue-derived macrophages may be the primary source of proinflammatory cytokines locally and in the systemic circulation. It remains unclear, however, how much of the insulin resistance is caused by the paracrine vs. endocrine effects of these cytokines.

ADIPONECTIN

Adiponectin is an anti-inflammatory cytokine produced exclusively by adipocytes. Adiponectin enhances insulin sensitivity and inhibits many steps in the inflammatory process. In the liver, adiponectin inhibits the expression of gluconeogenic enzymes and the rate of glucose production. In muscle, adiponectin increases glucose transport and enhances fatty acid oxidation, partially due to activation of adenosine monophosphate (AMP) kinase. Adiponectin is reduced in the metabolic syndrome. The relative contribution of adiponectin deficiency versus overabundance of the proinflammatory cytokines is unclear.

Clinical Features

SYMPTOMS AND SIGNS

The metabolic syndrome is typically not associated with symptoms. On physical examination, waist circumference may be expanded and blood pressure elevated. The presence of one or either of these signs should alert the clinician to search for other biochemical abnormalities that may be associated with the metabolic syndrome. Less frequently, lipoatrophy or acanthosis nigricans is found on examination. Because these physical findings typically are associated with severe insulin resistance, other components of the metabolic syndrome should be expected.

ASSOCIATED DISEASES

Cardiovascular Disease

The relative risk for new-onset CVD in patients with the metabolic syndrome, in the absence of diabetes, averages between 1.5-fold and threefold. However, in an 8-year follow-up of middle-aged men and women in the Framingham Offspring Study (FOS), the population-attributable risk for patients with the metabolic syndrome to develop CVD was 34% in men and only 16% in women. In the same study, both the metabolic syndrome and diabetes predicted ischemic stroke, with greater risk for patients with the metabolic syndrome than for those with diabetes alone (19% vs. 7%), particularly in women (27% vs. 5%). Patients with metabolic syndrome are also at increased risk for peripheral vascular disease.

Type 2 Diabetes

Overall, the risk for Type 2 diabetes in patients with the metabolic syndrome is increased three- to fivefold. In the FOS's 8-year follow-up of middle-aged men and women, the population-attributable risk for developing Type 2 diabetes was 62% in men and 47% in women.

OTHER ASSOCIATED CONDITIONS

In addition to the features specifically associated with metabolic syndrome, insulin resistance is accompanied by other metabolic alterations. Those alterations include increases in apoB and apoC-III, uric acid, prothrombotic factors (fibrinogen, plasminogen activator inhibitor 1), serum viscosity, asymmetric dimethylarginine, homocysteine, white blood cell count, proinflammatory cytokines, CRP, microalbuminuria, nonalcoholic fatty liver disease (NAFLD) and/or nonalcoholic steatohepatitis (NASH), polycystic ovarian disease (PCOS), and obstructive sleep apnea (OSA).

Nonalcoholic Fatty Liver Disease

(See also Chap. 309) Fatty liver is relatively common. However, in NASH, both triglyceride accumulation and inflammation coexist. NASH is now present in 2–3% of the population in the United States and other Western countries. As the prevalence of overweight/obesity and the metabolic syndrome increases, NASH may become one of the more common causes of end-stage liver disease and hepatocellular carcinoma.

Hyperuricemia

(See also Chap. 359) Hyperuricemia reflects defects in insulin action on the renal tubular reabsorption of uric acid, whereas the increase in asymmetric dimethylarginine, an endogenous inhibitor of nitric oxide synthase, relates to endothelial dysfunction. Microalbuminuria also may be caused by altered endothelial pathophysiology in the insulin-resistant state.

Polycystic Ovary Syndrome

(See also Chap. 347) PCOS is highly associated with the metabolic syndrome, with a prevalence

between 40 and 50%. Women with PCOS are 2–4 times more likely to have the metabolic syndrome than are women without PCOS.

Obstructive Sleep Apnea

(See also Chap. 27) OSA is commonly associated with obesity, hypertension, increased circulating cytokines, IGT, and insulin resistance. With these associations, it is not surprising that the metabolic syndrome is frequently present. Moreover, when biomarkers of insulin resistance are compared between patients with OSA and weight-matched controls, insulin resistance is more severe in patients with OSA. Continuous positive airway pressure (CPAP) treatment in OSA patients improves insulin sensitivity.

Diagnosis

The diagnosis of the metabolic syndrome relies on satisfying the criteria listed in **Table 242-1** by using tools at the bedside and in the laboratory. The medical history should include evaluation of symptoms for OSA in all patients and PCOS in premenopausal women. Family history will help determine risk for CVD and DM. Blood pressure and waist circumference measurements provide information necessary for the diagnosis.

LABORATORY TESTS

Fasting lipids and glucose are needed to determine if the metabolic syndrome is present. The measurement of additional biomarkers associated with insulin resistance can be individualized. Such tests might include apoB, high-sensitivity CRP, fibrinogen, uric acid, urinary microalbumin, and liver function tests. A sleep study should be performed if symptoms of OSA are present. If PCOS is suspected on the basis of clinical features and anovulation, testosterone, luteinizing hormone, and follicle-stimulating hormone should be measured.

Treatment: The Metabolic Syndrome

LIFESTYLE

(See also Chap. 78) Obesity is the driving force behind the metabolic syndrome. Thus, weight reduction is the primary approach to the disorder. With weight reduction, the improvement in insulin sensitivity is often accompanied by favorable modifications in many components of the metabolic syndrome. In general, recommendations for weight loss include a combination of caloric restriction, increased physical activity, and behavior modification. For weight reduction, caloric restriction is the most important component, whereas increases in physical activity are important for maintenance of weight loss. Some, but not all, evidence suggests that the addition of exercise to caloric restriction may promote relatively greater weight loss from the visceral depot. The tendency for weight regain after successful weight reduction underscores the need for long-lasting behavioral changes.

Diet

Before prescribing a weight-loss diet, it is important to emphasize that it takes a long time for a patient to achieve an expanded fat mass; thus, the correction need not occur quickly. On the basis of ~3500 kcal = 1 lb of fat, ~500 kcal restriction daily equates to weight reduction of 1 lb per week. Diets restricted in carbohydrate typically provide a rapid initial weight loss. However, after 1 year, the amount of weight reduction is usually unchanged. Thus, adherence to the diet is more important than which diet is chosen. Moreover, there is concern about diets enriched in saturated fat, particularly for patients at risk for CVD. Therefore, a high-quality diet— i.e., enriched in fruits, vegetables, whole grains, lean poultry, and fish—should be encouraged to provide the maximum overall health benefit.

Physical Activity

Before a physical activity recommendation is provided to patients with the metabolic syndrome, it is important to ensure that the increased activity does not incur risk. Some high-risk patients should undergo formal cardiovascular evaluation before initiating an exercise program. For an inactive participant, gradual increases in physical activity should be encouraged to enhance adherence and avoid injury. Although increases in physical activity can lead to modest weight reduction, 60–90 min of daily activity is required to achieve this goal. Even if an overweight or obese adult is unable to achieve this level of activity, he or she will still derive a significant health benefit from at least 30 min of moderate-intensity daily activity. The caloric value of 30 min of a variety of activities can be found at http://www.americanheart.org/presenter.jhtml?identifier=3040364. Of note, a variety of routine activities, such as gardening, walking, and housecleaning, require moderate caloric expenditure. Thus, physical activity need not be defined solely in terms of formal exercise such as jogging, swimming, or tennis.

Obesity

(See also Chap. 78) In some patients with the metabolic syndrome, treatment options need to extend beyond lifestyle intervention. Weight-loss drugs come in two major classes: appetite suppressants and absorption inhibitors. Appetite suppressants approved by the U.S. Food and Drug Administration include phentermine (for short-term use only, 3 months) and sibutramine. Orlistat inhibits fat absorption by ~30% and is moderately effective compared to placebo (~5% weight loss). Orlistat has been shown

to reduce the incidence of Type 2 diabetes, an effect that was especially evident in patients with baseline IGT.

Bariatric surgery is an option for patients with the metabolic syndrome who have a body mass index (BMI) >40 kg/m^2 or >35 kg/m^2 with comorbidities. Gastric bypass results in a dramatic weight reduction and improvement in the features of metabolic syndrome. A survival benefit has also been realized.

LDL CHOLESTEROL

(See also Chap. 356) The rationale for the NCEP:ATPIII panel to develop criteria for the metabolic syndrome was to go beyond LDL cholesterol in identifying and reducing risk for CVD. The working assumption by the panel was that LDL cholesterol goals had already been achieved, and increasing evidence supports a linear reduction in CVD events with progressive lowering of LDL cholesterol. For patients with the metabolic syndrome and diabetes, LDL cholesterol should be reduced to <100 mg/dL and perhaps further in patients with a history of CVD events. For patients with the metabolic syndrome without diabetes, the Framingham risk score may predict a 10-year CVD risk that exceeds 20%. In these subjects, LDL cholesterol should also be reduced to <100 mg/dL. With a 10-year risk of <20%, however, the targeted LDL cholesterol goal is <130 mg/dL.

Diets restricted in saturated fats (<7% of calories), trans-fats (as few as possible), and cholesterol (<200 mg daily) should be applied aggressively. If LDL cholesterol remains above goal, pharmacologic intervention is needed. Statins (HMG-CoA reductase inhibitors), which produce a 20–60% lowering of LDL cholesterol, are generally the first choice for medication intervention. Of note, for each doubling of the statin dose, there is only ~6% additional lowering of LDL cholesterol. Side effects are rare and include an increase in hepatic transaminases and/or myopathy. The cholesterol absorption inhibitor ezetimibe is well tolerated and should be the second choice. Ezetimibe typically reduces LDL cholesterol by 15–20%. The bile acid sequestrants cholestyramine and colestipol are more effective than ezetimibe but must be used with caution in patients with the metabolic syndrome because they can increase triglycerides. In general, bile sequestrants should not be administered when fasting triglycerides are >200 mg/dL. Side effects include gastrointestinal symptoms (palatability, bloating, belching, constipation, anal irritation). Nicotinic acid has modest LDL cholesterol–lowering capabilities (<20%). Fibrates are best employed to lower LDL cholesterol when both LDL cholesterol and triglycerides are elevated. Fenofibrate may be more effective than gemfibrozil in this group.

TRIGLYCERIDES

The NCEP:ATPIII has focused on non-HDL cholesterol rather than triglycerides. However, a fasting triglyceride value of <150 mg/dL is recommended. In general, the response of fasting triglycerides relates to the amount of weight reduction achieved. A weight reduction of >10% is necessary to lower fasting triglycerides.

A fibrate (gemfibrozil or fenofibrate) is the drug of choice to lower fasting triglycerides and typically achieve a 35–50% reduction. Concomitant administration with drugs metabolized by the 3A4 cytochrome P450 system (including some statins) greatly increases the risk of myopathy. In these cases, fenofibrate may be preferable to gemfibrozil. In the Veterans Affairs HDL Intervention Trial (VA-HIT), gemfibrozil was administered to men with known CHD and levels of HDL cholesterol <40 mg/dL. A coronary disease event and mortality rate benefit was experienced predominantly in men with hyperinsulinemia and/or diabetes, many of whom retrospectively were identified as having the metabolic syndrome. Of note, the amount of triglyceride lowering in the VA-HIT did not predict benefit. Although levels of LDL cholesterol did not change, a decrease in LDL particle number correlated with benefit. Although several additional clinical trials have been performed, they have not shown clear evidence that fibrates reduce CVD risk as a consequence of triglyceride lowering.

Other drugs that lower triglycerides include statins, nicotinic acid, and high doses of omega-3 fatty acids. In choosing a statin for this purpose, the dose must be high for the "less potent" statins (lovastatin, pravastatin, fluvastatin) or intermediate for the "more potent" statins (simvastatin, atorvastatin, rosuvastatin). The effect of nicotinic acid on fasting triglycerides is dose-related and less than that of fibrates (~20–40%). In patients with the metabolic syndrome and diabetes, nicotinic acid may increase fasting glucose. Omega-3 fatty acid preparations that include high doses of docosahexaenoic acid and eicosapentaenoic acid (~3.0–4.5 g daily) lower fasting triglycerides by ~40%. No interactions with fibrates or statins occur, and the main side effect is eructation with a fishy taste. This can be partially blocked by ingestion of the nutraceutical after freezing. Clinical trials of nicotinic acid or high-dose omega-3 fatty acids in patients with the metabolic syndrome have not been reported.

HDL CHOLESTEROL

Beyond weight reduction, there are very few lipid-modifying compounds that increase HDL cholesterol. Statins, fibrates, and bile acid sequestrants have modest effects (5–10%), and there is no effect on HDL cholesterol with ezetimibe or omega-3 fatty acids. Nicotinic acid is the only currently available drug with predictable HDL cholesterol-raising properties. The response is dose-related and can increase HDL cholesterol ~30% above baseline. There is limited evidence at present that raising HDL

has a benefit on CVD events independent of lowering LDL cholesterol, particularly in patients with the metabolic syndrome.

BLOOD PRESSURE

(See also Chap. 247) The direct relationship between blood pressure and all-cause mortality rate has been well established, including patients with hypertension (>140/90) versus prehypertension (>120/80 but <140/90) versus individuals with normal blood pressure (<120/80). In patients with the metabolic syndrome without diabetes, the best choice for the first antihypertensive should usually be an angiotensin-converting enzyme (ACE) inhibitor or an angiotensin II receptor blocker, as these two classes of drugs appear to reduce the incidence of new-onset Type 2 diabetes. In all patients with hypertension, a sodium-restricted diet enriched in fruits and vegetables and low-fat dairy products should be advocated. Home monitoring of blood pressure may assist in maintaining good blood pressure control.

IMPAIRED FASTING GLUCOSE

(See also Chap. 344) In patients with the metabolic syndrome and Type 2 diabetes, aggressive glycemic control may favorably modify fasting triglycerides and/or HDL cholesterol. In patients with IFG without a diagnosis of diabetes, a lifestyle intervention that includes weight reduction, dietary fat restriction, and increased physical activity has been shown to reduce the incidence of Type 2 diabetes. Metformin has also been shown to reduce the incidence of diabetes, although the effect was less than that seen with lifestyle intervention.

INSULIN RESISTANCE

(See also Chap. 344) Several drug classes [biguanides, thiazolidinediones (TZDs)] increase insulin sensitivity. Because insulin resistance is the primary pathophysiologic mechanism for the metabolic syndrome, representative drugs in these classes reduce its prevalence. Both metformin and TZDs enhance insulin action in the liver and suppress endogenous glucose production. TZDs, but not metformin, also improve insulin-mediated glucose uptake in muscle and adipose tissue. Benefits of both drugs have also been seen in patients with NAFLD and PCOS, and the drugs have been shown to reduce markers of inflammation and small dense LDL.

FURTHER READINGS

Alberti KG et al: Harmonizing the metabolic syndrome: A joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 120:1640, 2009[PMID: 19805654]

Cornier MA et al: The metabolic syndrome. Endocr Rev 29:777, 2008[PMID: 18971485]

Eckel RH et al: The metabolic syndrome. Lancet 365:1415, 2005[PMID: 15836891]

Ford ES: Prevalence of the metabolic syndrome defined by the International Diabetes Federation among adults in the U.S. Diabetes Care 28:2745, 2005[PMID: 16249550]

Grundy SM et al: Diagnosis and management of the metabolic syndrome. An American Heart Association/National Heart, Lung, and Blood Institute scientific statement. Circulation 112:2735, 2005[PMID: 16157765]

———: Metabolic syndrome: Connecting and reconciling cardiovascular and diabetes worlds. J Am Coll Cardiol 47:1093, 2006

Stewart PM et al: Selective inhibitors of 11beta-hydroxysteroid dehydrogenase type 1 for patients with metabolic syndrome: Is the target liver, fat, or both? Diabetes 58:14, 2009[PMID: 19114723]

Valle-Gottlieb MG et al: Associations among metabolic syndrome, ischemia, inflammatory, oxidatives, and lipids biomarkers. J Clin Endocrinol Metab 95:586, 2010[PMID: 20016051]

Zimmet P et al: The metabolic syndrome: Progress towards one definition for an epidemic of our time. Nat Clin Pract Endocrinol Metab 4:239, 2008[PMID: 18427543]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ISCHEMIC HEART DISEASE: INTRODUCTION

Ischemic heart disease (IHD) is a condition in which there is an inadequate supply of blood and oxygen to a portion of the myocardium; it typically occurs when there is an imbalance between myocardial oxygen supply and demand. The most common cause of myocardial ischemia is atherosclerotic disease of an epicardial coronary artery (or arteries) sufficient to cause a regional reduction in myocardial blood flow and inadequate perfusion of the myocardium supplied by the involved coronary artery. Chapter 241 deals with the development and treatment of atherosclerosis. This chapter focuses on the chronic manifestations and treatment of ischemic heart disease. The following chapters address the acute phases of this disease.

Epidemiology

IHD causes more deaths and disability and incurs greater economic costs than any other illness in the developed world. IHD is the most common, serious, chronic, life-threatening illness in the United States, where 13 million persons have IHD, >6 million have angina pectoris, and >7 million have sustained a myocardial infarction. Genetic factors, a high-fat and energy-rich diet, smoking, and a sedentary lifestyle are associated with the emergence of IHD (Chap. 241). In the United States and Western Europe, it is growing among low-income groups, but primary prevention has delayed the disease to later in life in all socioeconomic groups. Despite these sobering statistics, it is worth noting that epidemiologic data show a decline in the rate of deaths due to IHD, about half of which is attributable to treatments and half to prevention by risk factor modification.

Obesity, insulin resistance, and type 2 diabetes mellitus are increasing and are powerful risk factors for IHD. With urbanization in countries with emerging economies and a growing middle class, elements of the energy-rich Western diet are being adopted. As a result, the prevalence of risk factors for and of IHD itself are both increasing rapidly in those regions such that a majority of the global burden of IHD occurs there. Population subgroups that appear to be particularly affected are men in South Asian countries, especially India and the Middle East. In light of the projection of large increases in IHD throughout the world, IHD is likely to become the most common cause of death worldwide by 2020.

Pathophysiology

Central to an understanding of the pathophysiology of myocardial ischemia is the concept of myocardial supply and demand. In normal conditions, for any given level of a demand for oxygen, the myocardium will control the supply of oxygen-rich blood to prevent underperfusion of myocytes and the subsequent development of ischemia and infarction. The major determinants of myocardial oxygen demand (MVO_2) are heart rate, myocardial contractility, and myocardial wall tension (stress). An adequate supply of oxygen to the myocardium requires a satisfactory level of oxygen-carrying capacity of the blood (determined by the inspired level of oxygen, pulmonary function, and hemoglobin concentration and function) and an adequate level of coronary blood flow. Blood flows through the coronary arteries in a phasic fashion, with the majority occurring during diastole. About 75% of the total coronary resistance to flow occurs across three sets of arteries: (1) large epicardial arteries (Resistance 1 = R_1), (2) prearteriolar vessels (R_2), and (3) arteriolar and intramyocardial capillary vessels (R_3). In the absence of significant flow-limiting atherosclerotic obstructions, R_1 is trivial; the major determinant of coronary resistance is found in R_2 and R_3.

The normal coronary circulation is dominated and controlled by the heart's requirements for oxygen. This need is met by the ability of the coronary vascular bed to vary its resistance (and, therefore, blood flow) considerably while the myocardium extracts a high and relatively fixed percentage of oxygen. Normally, intramyocardial resistance vessels demonstrate an immense capacity for dilation (R_2 and R_3 decrease). For example, the changing oxygen needs of the heart with exercise and emotional stress affect coronary vascular resistance and in this manner regulate the supply of oxygen and substrate to the myocardium (*metabolic regulation*). The coronary resistance vessels also adapt to physiologic alterations in blood pressure to maintain coronary blood flow at levels appropriate to myocardial needs (*autoregulation*).

By reducing the lumen of the coronary arteries, atherosclerosis limits appropriate increases in perfusion when the demand for flow is augmented, as occurs during exertion or excitement. When the luminal reduction is severe, myocardial perfusion in the basal state is reduced. Coronary blood flow also can be limited by spasm (see Prinzmetal's angina in Chap. 244), arterial thrombi, and, rarely, coronary emboli as well as by ostial narrowing due to aortitis. Congenital abnormalities such as the origin of the left anterior descending coronary artery from the pulmonary artery may cause myocardial ischemia and infarction in infancy, but this cause is very rare in adults.

Myocardial ischemia also can occur if myocardial oxygen demands are markedly increased and particularly when coronary blood flow may be limited, as occurs in severe left ventricular hypertrophy due to aortic stenosis. The latter can present with angina that is indistinguishable from that caused by coronary atherosclerosis largely owing to subendocardial ischemia (Chap. 237). A reduction in the oxygen-carrying capacity of the blood, as in extremely severe anemia or in the presence of carboxyhemoglobin, rarely causes myocardial ischemia by itself but may lower the threshold for ischemia in patients with moderate coronary obstruction.

Not infrequently, two or more causes of ischemia coexist in a patient, such as an increase in oxygen demand due to left ventricular hypertrophy secondary to hypertension and a reduction in oxygen supply secondary to coronary atherosclerosis and anemia. Abnormal constriction or failure of normal dilation of the coronary resistance vessels also can cause ischemia. When it causes angina, this condition is referred to as *microvascular angina*.

CORONARY ATHEROSCLEROSIS

Epicardial coronary arteries are the major site of atherosclerotic disease. The major risk factors for atherosclerosis [high levels of plasma low-density lipoprotein (LDL), low plasma high-density lipoprotein (HDL), cigarette smoking, hypertension, and diabetes mellitus (Chap. 241)] disturb the normal functions of the vascular endothelium. These functions include local control of vascular tone, maintenance of an antithrombotic surface, and control of inflammatory cell adhesion and

diapedesis. The loss of these defenses leads to inappropriate constriction, luminal thrombus formation, and abnormal interactions between blood cells, especially monocytes and platelets, and the activated vascular endothelium. Functional changes in the vascular milieu ultimately result in the subintimal collections of fat, smooth muscle cells, fibroblasts, and intercellular matrix that define the atherosclerotic plaque. This process develops at irregular rates in different segments of the epicardial coronary tree and leads eventually to segmental reductions in cross-sectional area, i.e., plaque formation.

There is also a predilection for atherosclerotic plaques to develop at sites of increased turbulence in coronary flow, such as at branch points in the epicardial arteries. When a stenosis reduces the diameter of an epicardial artery by 50%, there is a limitation of the ability to increase flow to meet increased myocardial demand. When the diameter is reduced by ~80%, blood flow at rest may be reduced, and further minor decreases in the stenotic orifice area can reduce coronary flow dramatically to cause myocardial ischemia at rest or with minimal stress.

Segmental atherosclerotic narrowing of epicardial coronary arteries is caused most commonly by the formation of a plaque, which is subject to rupture or erosion of the cap separating the plaque from the bloodstream. Upon exposure of the plaque contents to blood, two important and interrelated processes are set in motion: (1) platelets are activated and aggregate, and (2) the coagulation cascade is activated, leading to deposition of fibrin strands. A thrombus composed of platelet aggregates and fibrin strands traps red blood cells and can reduce coronary blood flow, leading to the clinical manifestations of myocardial ischemia.

The location of the obstruction influences the quantity of myocardium rendered ischemic and determines the severity of the clinical manifestations. Thus, critical obstructions in vessels, such as the left main coronary artery and the proximal left anterior descending coronary artery, are particularly hazardous. Chronic severe coronary narrowing and myocardial ischemia frequently are accompanied by the development of collateral vessels, especially when the narrowing develops gradually. When well developed, such vessels can by themselves provide sufficient blood flow to sustain the viability of the myocardium at rest but not during conditions of increased demand.

With progressive worsening of a stenosis in a proximal epicardial artery, the distal resistance vessels (when they function normally) dilate to reduce vascular resistance and maintain coronary blood flow. A pressure gradient develops across the proximal stenosis, and poststenotic pressure falls. When the resistance vessels are maximally dilated, myocardial blood flow becomes dependent on the pressure in the coronary artery distal to the obstruction. In these circumstances, ischemia, manifest clinically by angina or electrocardiographically by ST-segment deviation, can be precipitated by increases in myocardial oxygen demand caused by physical activity, emotional stress, and/or tachycardia. Changes in the caliber of the stenosed coronary artery due to physiologic vasomotion, loss of endothelial control of dilation (as occurs in atherosclerosis), pathologic spasm (Prinzmetal's angina), or small platelet-rich plugs also can upset the critical balance between oxygen supply and demand and thereby precipitate myocardial ischemia.

Effects of Ischemia

During episodes of inadequate perfusion caused by coronary atherosclerosis, myocardial tissue oxygen tension falls and may cause transient disturbances of the mechanical, biochemical, and electrical functions of the myocardium. Coronary atherosclerosis is a focal process that usually causes nonuniform ischemia. During ischemia, regional disturbances of ventricular contractility cause segmental hypokinesia, akinesia, or, in severe cases, bulging (dyskinesia), which can reduce myocardial pump function.

The abrupt development of severe ischemia, as occurs with total or subtotal coronary occlusion, is associated with almost instantaneous failure of normal muscle relaxation and then contraction. The relatively poor perfusion of the subendocardium causes more intense ischemia of this portion of the wall (compared with the subepicardial region). Ischemia of large portions of the ventricle causes transient left ventricular failure, and if the papillary muscle apparatus is involved, mitral regurgitation can occur. When ischemia is transient, it may be associated with angina pectoris; when it is prolonged, it can lead to myocardial necrosis and scarring with or without the clinical picture of acute myocardial infarction (Chap. 245).

A wide range of abnormalities in cell metabolism, function, and structure underlie these mechanical disturbances during ischemia. The normal myocardium metabolizes fatty acids and glucose to carbon dioxide and water. With severe oxygen deprivation, fatty acids cannot be oxidized, and glucose is converted to lactate; intracellular pH is reduced, as are the myocardial stores of high-energy phosphates, i.e., ATP and creatine phosphate. Impaired cell membrane function leads to the leakage of potassium and the uptake of sodium by myocytes as well as an increase in cytosolic calcium. The severity and duration of the imbalance between myocardial oxygen supply and demand determine whether the damage is reversible (≤20 min for total occlusion in the absence of collaterals) or permanent, with subsequent myocardial necrosis (>20 min).

Ischemia also causes characteristic changes in the electrocardiogram (ECG) such as repolarization abnormalities, as evidenced by inversion of T waves and, when more severe, displacement of ST segments (Chap. 228). Transient T-wave inversion probably reflects nontransmural, intramyocardial ischemia; transient ST-segment depression often reflects patchy subendocardial ischemia; and ST-segment elevation is thought to be caused by more severe transmural ischemia. Another important consequence of myocardial ischemia is electrical instability, which may lead to isolated ventricular premature beats or even ventricular tachycardia or ventricular fibrillation (Chap. 233). Most patients who die suddenly from IHD do so as a result of ischemia-induced ventricular tachyarrhythmias (Chap. 273).

Asymptomatic versus Symptomatic Ihd

Postmortem studies of accident victims and military casualties in Western countries have shown that coronary atherosclerosis often begins to develop before age 20 and is widespread even among adults who were asymptomatic during life. Exercise stress tests in asymptomatic persons may show evidence of silent myocardial ischemia, i.e., exercise-induced ECG changes not accompanied by angina pectoris; coronary angiographic studies of such persons may reveal coronary artery plaques and previously unrecognized obstructions (Chap. 230). Postmortem examination of patients with such obstructions without a history of clinical manifestations of myocardial ischemia often shows macroscopic scars secondary to myocardial infarction in regions supplied by diseased coronary arteries, with or without collateral circulation. According to population studies, ~25% of patients who survive acute myocardial infarction may not come to medical attention, and these patients have the same adverse prognosis as do those who present with the classic clinical picture of acute myocardial infarction (Chap. 245). Sudden death may be unheralded and is a common presenting manifestation of IHD (Chap. 273).

Patients with IHD also can present with cardiomegaly and heart failure secondary to ischemic damage of the left ventricular myocardium that may have caused no symptoms before the development of heart failure; this condition is referred to as *ischemic cardiomyopathy*. In contrast to the asymptomatic phase of IHD, the symptomatic phase is

characterized by chest discomfort due to either angina pectoris or acute myocardial infarction (Chap. 245). Having entered the symptomatic phase, the patient may exhibit a stable or progressive course, revert to the asymptomatic stage, or die suddenly.

STABLE ANGINA PECTORIS

This episodic clinical syndrome is due to transient myocardial ischemia. Various diseases that cause myocardial ischemia as well as the numerous forms of discomfort with which it may be confused are discussed in Chap. 12. Males constitute ~70% of all patients with angina pectoris and an even greater proportion of those less than 50 years of age. It is, however, important to note that angina pectoris in women is often atypical in presentation (see below).

History

The typical patient with angina is a man >50 years or a woman >60 years of age who complains of episodes of chest discomfort, usually described as heaviness, pressure, squeezing, smothering, or choking and only rarely as frank pain. When the patient is asked to localize the sensation, he or she typically places a hand over the sternum, sometimes with a clenched fist, to indicate a squeezing, central, substernal discomfort (Levine's sign). Angina is usually crescendo-decrescendo in nature, typically lasts 2 to 5 min, and can radiate to either shoulder and to both arms (especially the ulnar surfaces of the forearm and hand). It also can arise in or radiate to the back, interscapular region, root of the neck, jaw, teeth, and epigastrium. Angina is rarely localized below the umbilicus or above the mandible. A useful finding in assessing a patient with chest discomfort is the fact that myocardial ischemic discomfort does not radiate to the trapezius muscles; that radiation pattern is more typical of pericarditis.

Although episodes of angina typically are caused by exertion (e.g., exercise, hurrying, or sexual activity) or emotion (e.g., stress, anger, fright, or frustration) and are relieved by rest, they also may occur at rest [see "Unstable Angina Pectoris," (Chap. 244)] and while the patient is recumbent (angina decubitus). The patient may be awakened at night by typical chest discomfort and dyspnea. Nocturnal angina may be due to episodic tachycardia, diminished oxygenation as the respiratory pattern changes during sleep, or expansion of the intrathoracic blood volume that occurs with recumbency; the latter causes an increase in cardiac size (end-diastolic volume), wall tension, and myocardial oxygen demand that can lead to ischemia and transient left ventricular failure.

The threshold for the development of angina pectoris may vary by time of day and emotional state. Many patients report a fixed threshold for angina, which occurs predictably at a certain level of activity, such as climbing two flights of stairs at a normal pace. In these patients, coronary stenosis and myocardial oxygen supply are fixed, and ischemia is precipitated by an increase in myocardial oxygen demand; they are said to have stable exertional angina. In other patients, the threshold for angina may vary considerably within any particular day and from day to day. In such patients, variations in myocardial oxygen supply, most likely due to changes in coronary vasomotor tone, may play an important role in defining the pattern of angina. A patient may report symptoms upon minor exertion in the morning (a short walk or shaving) yet by midday be capable of much greater effort without symptoms. Angina may also be precipitated by unfamiliar tasks, a heavy meal, exposure to cold, or a combination of these factors.

Exertional angina typically is relieved in 1 to 5 min by slowing or ceasing activities and even more rapidly by rest and sublingual nitroglycerin (see below). Indeed, the diagnosis of angina should be suspect if it does not respond to the combination of these measures. The severity of angina can be conveniently summarized by the Canadian Cardiac Society functional classification (**Table 243-1**). Its impact on the patient's functional capacity can be described by using the New York Heart Association functional classification (Table 243-1).

Table 243–1. Cardiovascular Disease Classification Chart

Class	New York Heart Association Functional Classification	Canadian Cardiovascular Society Functional Classification
I	Patients have cardiac disease but *without* the resulting *limitations* of physical activity. Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea, or anginal pain.	Ordinary physical activity, such as walking and climbing stairs, *does not cause angina*. Angina present with strenuous or rapid or prolonged exertion at work or recreation.
II	Patients have cardiac disease resulting in *slight limitation* of physical activity. They are comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea, or anginal pain.	*Slight limitation* of ordinary activity. Walking or climbing stairs rapidly, walking uphill, walking or stair climbing after meals, in cold, or when under emotional stress or only during the few hours after awakening. Walking more than two blocks on the level and climbing more than one flight of stairs at a normal pace and in normal conditions.
III	Patients have cardiac disease resulting in *marked limitation* of physical activity. They are comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, dyspnea, or anginal pain.	*Marked limitation* of ordinary physical activity. Walking one to two blocks on the level and climbing more than one flight of stairs in normal conditions.
IV	Patients have cardiac disease resulting in *inability* to carry on any physical activity without discomfort. Symptoms of cardiac insufficiency or of the anginal syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.	*Inability* to carry on any physical activity without discomfort —anginal syndrome *may* be present at rest.

Source: Modified from Goldman L et al: Circulation 64:1227, 1981.

Sharp, fleeting chest pain or a prolonged, dull ache localized to the left submammary area is rarely due to myocardial ischemia. However, especially in women and diabetic patients, angina pectoris may be atypical in location and not strictly related to provoking factors. In addition, this symptom may exacerbate and remit over days, weeks, or months. Its occurrence can be seasonal, occurring more frequently in the winter in temperate climates. Anginal "equivalents" are symptoms of myocardial ischemia other than angina. They include dyspnea, nausea, fatigue, and faintness and are more common in the elderly and in diabetic patients.

Systematic questioning of a patient with suspected IHD is important to uncover the features of an unstable syndrome associated with increased risk, such as angina occurring with less exertion than in the past, occurring at rest, or awakening

the patient from sleep. Since coronary atherosclerosis often is accompanied by similar lesions in other arteries, a patient with angina should be questioned and examined for peripheral arterial disease [intermittent claudication (Chap. 249)], stroke, or transient ischemic attacks (Chap. 370). It is also important to uncover a family history of premature IHD (<55 years in first-degree male relatives and <65 in female relatives) and the presence of diabetes mellitus, hyperlipidemia, hypertension, cigarette smoking, and other risk factors for coronary atherosclerosis (Chap. 241).

The history of typical angina pectoris establishes the diagnosis of IHD until proven otherwise. In patients with atypical angina (Chap. 12), the coexistence of advanced age, male sex, the postmenopausal state, and risk factors for atherosclerosis increase the likelihood of hemodynamically significant coronary disease. A particularly challenging problem is the evaluation and management of patients with persistent ischemic-type chest discomfort but no flow-limiting obstructions in their epicardial coronary arteries. This situation arises more often in women than in men. Potential etiologies include microvascular coronary disease (detectable on coronary reactivity testing in response to vasoactive agents such as intracoronary adenosine, acetylcholine, and nitroglycerin) and abnormal cardiac nociception. Treatment of microvascular coronary disease should focus on efforts to improve endothelial function, including nitrates, beta blockers, calcium antagonists, statins, and angiotensin-converting enzyme (ACE) inhibitors. Abnormal cardiac nociception is more difficult to manage and may be ameliorated in some cases by imipramine.

Physical Examination

The physical examination is often normal in patients with stable angina when they are asymptomatic. However, because of the increased likelihood of ischemic heart disease in patients with diabetes and/or peripheral arterial disease, clinicians should search for evidence of atherosclerotic disease at other sites, such as an abdominal aortic aneurysm, carotid arterial bruits, and diminished arterial pulses in the lower extremities. The physical examination also should include a search for evidence of risk factors for atherosclerosis such as xanthelasmas and xanthomas (Chap. 241). Evidence for peripheral arterial disease should be sought by evaluating the pulse contour at multiple locations and comparing the blood pressure between the arms and between the arms and the legs (ankle-brachial index). Examination of the fundi may reveal an increased light reflex and arteriovenous nicking as evidence of hypertension. There also may be signs of anemia, thyroid disease, and nicotine stains on the fingertips from cigarette smoking.

Palpation may reveal cardiac enlargement and abnormal contraction of the cardiac impulse (left ventricular dyskinesia). Auscultation can uncover arterial bruits, a third and/or fourth heart sound, and, if acute ischemia or previous infarction has impaired papillary muscle function, an apical systolic murmur due to mitral regurgitation. These auscultatory signs are best appreciated with the patient in the left lateral decubitus position. Aortic stenosis, aortic regurgitation (Chap. 237), pulmonary hypertension (Chap. 250), and hypertrophic cardiomyopathy (Chap. 238) must be excluded, since these disorders may cause angina in the absence of coronary atherosclerosis. Examination during an anginal attack is useful, since ischemia can cause transient left ventricular failure with the appearance of a third and/or fourth heart sound, a dyskinetic cardiac apex, mitral regurgitation, and even pulmonary edema. Tenderness of the chest wall, localization of the discomfort with a single fingertip on the chest, or reproduction of the pain with palpation of the chest makes it unlikely that the pain is caused by myocardial ischemia. A protuberant abdomen may indicate that the patient has the metabolic syndrome and is at increased risk for atherosclerosis.

Laboratory Examination

Although the diagnosis of IHD can be made with a high degree of confidence from the history and physical examination, a number of simple laboratory tests can be helpful. The urine should be examined for evidence of diabetes mellitus and renal disease (including microalbuminuria) since these conditions accelerate atherosclerosis. Similarly, examination of the blood should include measurements of lipids (cholesterol—total, LDL, HDL—and triglycerides), glucose (hemoglobin A_{1C}), creatinine, hematocrit, and, if indicated based on the physical examination, thyroid function. A chest x-ray is important as it may show the consequences of IHD, i.e., cardiac enlargement, ventricular aneurysm, or signs of heart failure. These signs can support the diagnosis of IHD and are important in assessing the degree of cardiac damage. Evidence exists that an elevated level of high-sensitivity C-reactive protein (CRP) (specifically, between 0 and 3 mg/dL) is an independent risk factor for IHD and may be useful in therapeutic decision making about the initiation of hypolipidemic treatment. The major benefit of high-sensitivity CRP is in reclassifying the risk of IHD in patients in the "intermediate" risk category on the basis of traditional risk factors.

Electrocardiogram

A 12-lead ECG recorded at rest may be normal in patients with typical angina pectoris, but there may also be signs of an old myocardial infarction (Chap. 228). Although repolarization abnormalities, i.e., ST-segment and T-wave changes, as well as left ventricular hypertrophy and disturbances of cardiac rhythm or intraventricular conduction are suggestive of IHD, they are nonspecific, since they also can occur in pericardial, myocardial, and valvular heart disease or, in the case of the former, transiently with anxiety, changes in posture, drugs, or esophageal disease. The presence of left ventricular hypertrophy (LVH) is a significant indication of increased risk of adverse outcomes from ischemic heart disease. Of note, even though LVH and cardiac rhythm disturbances are nonspecific indicators of the development of IHD, they may be contributing factors to episodes of angina in patients in whom IHD has developed as a consequence of conventional risk factors. Dynamic ST-segment and T-wave changes that accompany episodes of angina pectoris and disappear thereafter are more specific.

Stress Testing

ELECTROCARDIOGRAPHIC

The most widely used test for both the diagnosis of IHD and the estimation of risk and prognosis involves recording the 12-lead ECG before, during, and after exercise, usually on a treadmill (**Fig. 243-1**). The test consists of a standardized incremental increase in external workload (**Table 243-2**) while symptoms, the ECG, and arm blood pressure are monitored. Exercise duration is usually symptom-limited, and the test is discontinued upon evidence of chest discomfort, severe shortness of breath, dizziness, severe fatigue, ST-segment depression >0.2 mV (2 mm), a fall in systolic blood pressure >10 mmHg, or the development of a ventricular tachyarrhythmia. This test is used to discover any limitation in exercise performance, detect typical ECG signs of myocardial ischemia, and establish their relationship to chest discomfort. The ischemic ST-segment response generally is defined as flat or downsloping depression of the ST segment >0.1 mV below baseline (i.e., the PR segment) and lasting longer than 0.08 s (Fig. 243-1). Upsloping or junctional ST-segment changes are not considered characteristic of ischemia and do not constitute a positive test. Although T-wave abnormalities, conduction disturbances, and ventricular arrhythmias that develop during exercise should be noted, they are also not diagnostic. Negative exercise tests in which the target heart rate (85% of maximal predicted heart rate for age and sex) is not achieved are considered nondiagnostic.

Table 243–2. Relation of Metabolic Equivalent Tasks (METs) to Stages in Various Testing Protocolsfunctional Class

Functional Class	Clinical Status	O_2 Cost mL/kg/min	METS	Treadmill Protocols

Functional Class	Clinical Status	O_2 Cost mL/kg/min	METS	BRUCE Modified 3 min Stage MPH	BRUCE Modified 3 min Stage %GR	BRUCE 3 min Stages MPH	BRUCE 3 min Stages %GR
				6.0	22	6.0	22
				5.5	20	5.2	20
				5.0	18	5.0	18
NORMAL AND I	HEALTHY, DEPENDENT ON AGE, ACTIVITY		56.0	16				
			52.5	15				
			49.0	14				
			45.5	13	4.2	16	4.2	16
			42.0	12				
			38.5	11	3.4	14	3.4	14
	SEDENTARY HEALTHY		35.0	10				
			31.5	9				
			28.0	8				
		LIMITED	24.5	7	2.5	12	2.5	12
II			21.0	6				
			17.5	5	1.7	10	1.7	10
III			14.0	4				
			10.5	3	1.7	5		
		SYMPTOMATIC	7.0	2	1.7	0		
IV			3.5	1				

Source: Modified from Fletcher GF et al: Circulation 104:1694, 2001.

Figure 243-1.

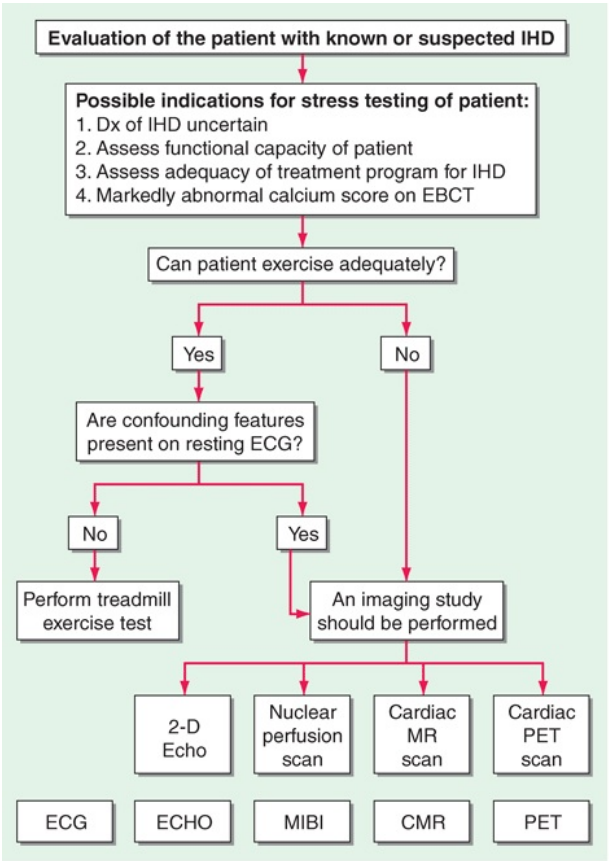

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

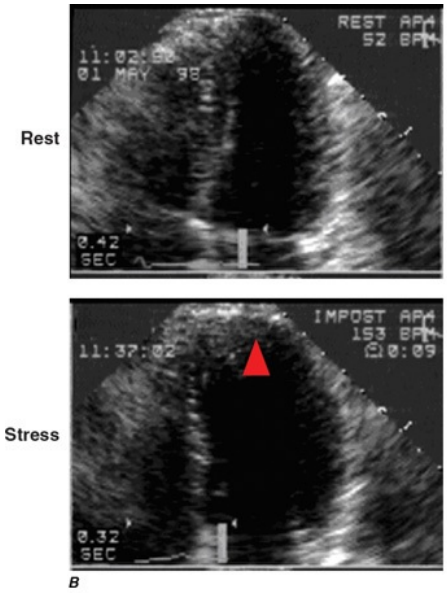

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

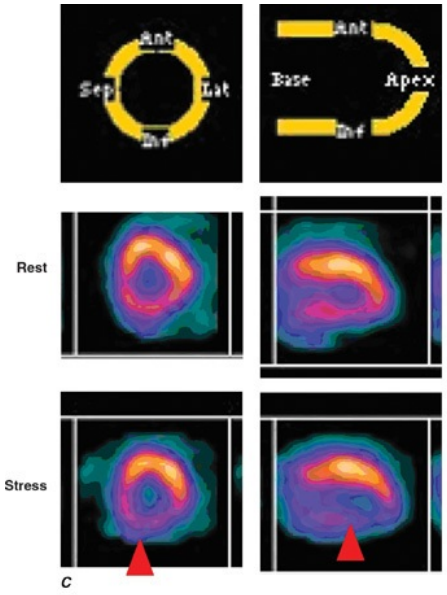

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

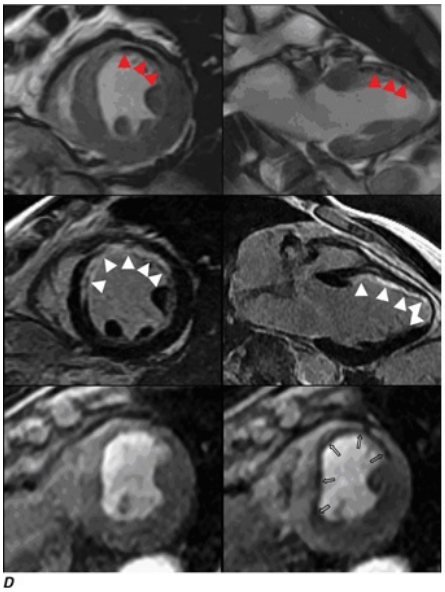

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

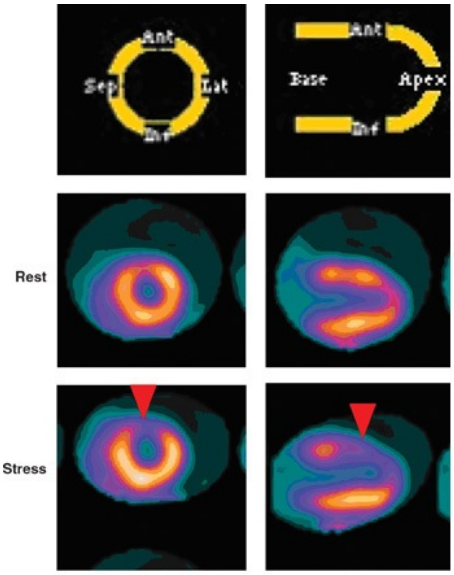

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evaluation of the patient with known or suspected ischemic heart disease. At the top of the figure is an algorithm for identifying patients who should be referred for stress testing and the decision pathway for determining whether a standard treadmill exercise with ECG monitoring alone is adequate. A specialized imaging study is necessary if the patient cannot exercise adequately (pharmacologic challenge is given) or if there are confounding features on the resting ECG (symptom-limited treadmill exercise may be used to stress the coronary circulation). At the bottom of the figure are examples of the data obtained with ECG monitoring and specialized imaging procedures. CMR, cardiac magnetic resonance; EBCT, electron beam computed tomography; ECG, electrocardiogram; ECHO, echocardiography; IHD, ischemic heart disease; MIBI, methoxyisobutyl isonitrite; MR, magnetic resonance; PET, positron emission tomography.

A. Lead V_4 at rest (*top*) and after 41/2 min of exercise (*bottom*). There is 3 mm (0.3 mV) of horizontal ST-segment depression, indicating a positive test for ischemia. *[Modified from BR Chaitman, in E Braunwald et al (eds): Heart Disease, 6th ed, Philadelphia, Saunders, 2001.]*

B. 45-year-old avid jogger who began experiencing classic substernal chest pressure underwent an exercise echo study. With exercise the patient's heart rate increased from 52 to 153 bpm. The left ventricular chamber dilated with exercise, and the septal and apical portions became akinetic to dyskinetic (*red arrow*). These findings are strongly suggestive of a significant flow-limiting stenosis in the proximal left anterior descending artery, which was confirmed at coronary angiography. *[Modified from SD Solomon, in E. Braunwald et al (eds): Primary Cardiology, 2nd ed, Philadelphia, Saunders, 2003.]*

C. Stress and rest myocardial perfusion SPECT images obtained with 99m-technetium sestamibi in a patient with chest pain and dyspnea on exertion. The images demonstrate a medium-size and severe stress perfusion defect involving the inferolateral and basal inferior walls, showing nearly complete reversibility, consistent with moderate ischemia in the right coronary artery territory (*red arrows*). *(Images provided by Dr. Marcello Di Carli, Nuclear Medicine Division, Brigham and Women's Hospital, Boston, MA.)*

D. A patient with a prior myocardial infarction presented with recurrent chest discomfort. On cardiac magnetic resonance (CMR) cine imaging, a large area of anterior akinesia was noted (marked by the arrows in the top left and right images, systolic frame only). This area of akinesia was matched by a larger extent of late gadolinium-DTPA enhancements consistent with a large transmural myocardial infarction (marked by arrows in the middle left and right images). Resting (*bottom left*) and adenosine vasodilating stress (*bottom right*) first-pass perfusion images revealed reversible perfusion abnormality that extended to the inferior septum. This patient was found to have an occluded proximal left anterior

abnormality that extended to the inferior septum. This patient was found to have an occluded proximal left anterior descending coronary artery with extensive collateral formation. This case illustrates the utility of different modalities in a CMR examination in characterizing ischemic and infarcted myocardium. DTPA, diethylenetriamine penta-acetic acid. (*Images provided by Dr. Raymond Kwong, Cardiovascular Division, Brigham and Women's Hospital, Boston, MA.)*

E. Stress and rest myocardial perfusion PET images obtained with rubidium-82 in a patient with chest pain on exertion. The images demonstrate a large and severe stress perfusion defect involving the mid and apical anterior, anterolateral, and anteroseptal walls and the LV apex, showing complete reversibility, consistent with extensive and severe ischemia in the mid-left anterior descending coronary artery territory (*red arrows*). *(Images provided by Dr. Marcello Di Carli, Nuclear Medicine Division, Brigham and Women's Hospital, Boston, MA.)*

In interpreting ECG stress tests, the probability that coronary artery disease (CAD) exists in the patient or population under study (i.e., pretest probability) should be considered. Overall, false-positive or false-negative results occur in one-third of cases. However, a positive result on exercise indicates that the likelihood of CAD is 98% in males who are >50 years with a history of typical angina pectoris and who develop chest discomfort during the test. The likelihood decreases if the patient has atypical or no chest pain by history and/or during the test.

The incidence of false-positive tests is significantly increased in patients with low probabilities of IHD, such as asymptomatic men < age 40 or premenopausal women with no risk factors for premature atherosclerosis. It is also increased in patients taking cardioactive drugs, such as digitalis and antiarrhythmic agents, and in those with intraventricular conduction disturbances, resting ST-segment and T-wave abnormalities, ventricular hypertrophy, or abnormal serum potassium levels. Obstructive disease limited to the circumflex coronary artery may result in a false-negative stress test since the lateral portion of the heart that this vessel supplies is not well represented on the surface 12-lead ECG. Since the overall sensitivity of exercise stress electrocardiography is only ~75%, a negative result does not exclude CAD, although it makes the likelihood of three-vessel or left main CAD extremely unlikely.

The physician should be present throughout the exercise test. It is important to measure total duration of exercise, the times to the onset of ischemic ST-segment change and chest discomfort, the external work performed (generally expressed as the stage of exercise), and the internal cardiac work performed, i.e., by the heart rate–blood pressure product. The depth of the ST-segment depression and the time needed for recovery of these ECG changes are also important. Because the risks of exercise testing are small but real—estimated at one fatality and two nonfatal complications per 10,000 tests—equipment for resuscitation should be available. Modified (heart rate–limited rather than symptom-limited) exercise tests can be performed safely in patients as early as 6 days after uncomplicated myocardial infarction (Table 243-2). Contraindications to exercise stress testing include rest angina within 48 h, unstable rhythm, severe aortic stenosis, acute myocarditis, uncontrolled heart failure, severe pulmonary hypertension, and active infective endocarditis.

The normal response to graded exercise includes progressive increases in heart rate and blood pressure. Failure of the blood pressure to increase or an actual decrease with signs of ischemia during the test is an important adverse prognostic sign, since it may reflect ischemia-induced global left ventricular dysfunction. The development of angina and/or severe (>0.2 mV) ST-segment depression at a low workload, i.e., before completion of stage II of the Bruce protocol, and/or ST-segment depression that persists >5 min after the termination of exercise increases the specificity of the test and suggests severe IHD and a high risk of future adverse events.

CARDIAC IMAGING

(See also Chap. 229) When the resting ECG is abnormal (e.g., preexcitation syndrome, >1 mm of resting ST-segment depression, left bundle branch block, paced ventricular rhythm), information gained from an exercise test can be enhanced by stress myocardial radionuclide perfusion imaging after the intravenous administration of thallium-201 or 99m-technetium sestamibi during exercise (or with pharmacologic) stress. Contemporary data also suggest positron emission tomography (PET) imaging (with exercise or pharmacologic stress) using N-13 ammonia or rubidium-82 nuclide as another technique for assessing perfusion. Images obtained immediately after cessation of exercise to detect regional ischemia are compared with those obtained at rest to confirm reversible ischemia and regions of persistently absent uptake that signify infarction.

A sizable fraction of patients who need noninvasive stress testing to identify myocardial ischemia and increased risk of coronary events cannot exercise because of peripheral vascular or musculoskeletal disease, exertional dyspnea, or deconditioning. In these circumstances, an intravenous pharmacologic challenge is used in place of exercise. For example, dipyridamole or adenosine can be given to create a coronary "steal" by temporarily increasing flow in nondiseased segments of the coronary vasculature at the expense of diseased segments. Alternatively, a graded incremental infusion of dobutamine may be administered to increase MVO_2. A variety of imaging options are available to accompany these pharmacologic stressors (Fig. 243-1). The development of a transient perfusion defect with a tracer such as thallium-201 or 99m-technetium sestamibi is used to detect myocardial ischemia.

Echocardiography is used to assess left ventricular function in patients with chronic stable angina and patients with a history of a prior myocardial infarction, pathologic Q waves, or clinical evidence of heart failure. Two-dimensional echocardiography can assess both global and regional wall motion abnormalities of the left ventricle that are transient when due to ischemia. Stress (exercise or dobutamine) echocardiography may cause the emergence of regions of akinesis or dyskinesis that are not present at rest. Stress echocardiography, like stress myocardial perfusion imaging, is more sensitive than exercise electrocardiography in the diagnosis of IHD. Cardiac magnetic resonance (CMR) stress testing is also evolving as an alternative to radionuclide, PET, or echocardiographic stress imaging. CMR stress testing performed with dobutamine infusion can be used to assess wall motion abnormalities accompanying ischemia, as well as myocardial perfusion. CMR can be used to provide more complete ventricular evaluation using multislice MR imaging (MRI) studies.

Atherosclerotic plaques become progressively calcified over time, and coronary calcification in general increases with age. For this reason, methods for detecting coronary calcium have been developed as a measure of the presence of coronary atherosclerosis. These methods involve computed tomography (CT) applications that achieve rapid acquisition of images [electron beam (EBCT), and multidetector (MDCT) detection]. Coronary calcium detected by these imaging techniques most commonly is quantified by using the Agatston score, which is based on the area and density of calcification. Although the diagnostic accuracy of this imaging method is high (sensitivity, 90–94%; specificity, 95–97%; negative predictive value, 93–99%), its prognostic utility has not been defined. Thus, its role in CT, EBCT, and MDCT scans for the detection and management of patients with IHD has not been clarified.

Coronary Arteriography

(See also Chap. 230) This diagnostic method outlines the lumina of the coronary arteries and can be used to detect or exclude serious coronary obstruction. However, coronary arteriography provides no information about the arterial wall, and severe atherosclerosis that does not encroach on the lumen may go undetected. Of note, atherosclerotic plaques characteristically are scattered throughout the coronary tree, tend to occur more frequently at branch points, and grow progressively in the intima and media of an epicardial coronary artery at first without encroaching on the lumen, causing an

outward bulging of the artery—a process referred to as remodeling (Chap. 241). Later in the course of the disease, further growth causes luminal narrowing.

INDICATIONS

Coronary arteriography is indicated in (1) patients with chronic stable angina pectoris who are severely symptomatic despite medical therapy and are being considered for revascularization, i.e., a percutaneous coronary intervention (PCI) or coronary artery bypass grafting (CABG), (2) patients with troublesome symptoms that present diagnostic difficulties in whom there is a need to confirm or rule out the diagnosis of IHD, (3) patients with known or possible angina pectoris who have survived cardiac arrest, (4) patients with angina or evidence of ischemia on noninvasive testing with clinical or laboratory evidence of ventricular dysfunction, and (5) patients judged to be at high risk of sustaining coronary events based on signs of severe ischemia on noninvasive testing, regardless of the presence or severity of symptoms (see below).

Examples of other indications for coronary arteriography include the following:

1. Patients with chest discomfort suggestive of angina pectoris but a negative or nondiagnostic stress test who require a definitive diagnosis for guiding medical management, alleviating psychological stress, career or family planning, or insurance purposes.
2. Patients who have been admitted repeatedly to the hospital for a suspected acute coronary syndrome (Chaps. 244 and 245) but in whom this diagnosis has not been established and in whom the presence or absence of CAD should be determined.
3. Patients with careers that involve the safety of others (e.g., pilots, firefighters, police) who have questionable symptoms or suspicious or positive noninvasive tests and in whom there are reasonable doubts about the state of the coronary arteries.
4. Patients with aortic stenosis or hypertrophic cardiomyopathy and angina in whom the chest pain could be due to IHD.
5. Male patients >45 years and females >55 years who are to undergo a cardiac operation such as valve replacement or repair and who may or may not have clinical evidence of myocardial ischemia.
6. Patients after myocardial infarction, especially those who are at high risk after myocardial infarction because of the recurrence of angina or the presence of heart failure, frequent ventricular premature contractions, or signs of ischemia on the stress test.
7. Patients with angina pectoris, regardless of severity, in whom noninvasive testing indicates a high risk of coronary events (poor exercise performance or severe ischemia).
8. Patients in whom coronary spasm or another nonatherosclerotic cause of myocardial ischemia (e.g., coronary artery anomaly, Kawasaki disease) is suspected.

Noninvasive alternatives to diagnostic coronary arteriography include CT angiography and cardiac MR angiography (Chap. 229). Although these new imaging techniques can provide information about obstructive lesions in the epicardial coronary arteries, their exact role in clinical practice has not been rigorously defined. Important aspects of their use that should be noted include the substantially higher radiation exposure with CT angiography compared to conventional diagnostic arteriography and the limitations on cardiac MR imposed by cardiac movement during the cardiac cycle, especially at high heart rates.

Prognosis

The principal prognostic indicators in patients known to have IHD are age, the functional state of the left ventricle, the location(s) and severity of coronary artery narrowing, and the severity or activity of myocardial ischemia. Angina pectoris of recent onset, unstable angina (Chap. 244), early postmyocardial infarction angina, angina that is unresponsive or poorly responsive to medical therapy, and angina accompanied by symptoms of congestive heart failure all indicate an increased risk for adverse coronary events. The same is true for the physical signs of heart failure, episodes of pulmonary edema, transient third heart sounds, and mitral regurgitation and for echocardiographic or radioisotopic (or roentgenographic) evidence of cardiac enlargement and reduced (<0.40) ejection fraction.

Most important, any of the following signs during noninvasive testing indicates a high risk for coronary events: inability to exercise for 6 min, i.e., stage II (Bruce protocol) of the exercise test; a strongly positive exercise test showing onset of myocardial ischemia at low workloads (≤0.1 mV ST-segment depression before completion of stage II, ≤0.2 mV ST depression at any stage, ST depression for >5 min after the cessation of exercise, a decline in systolic pressure >10 mmHg during exercise, the development of ventricular tachyarrhythmias during exercise); the development of large or multiple perfusion defects or increased lung uptake during stress radioisotope perfusion imaging; and a decrease in left ventricular ejection fraction during exercise on radionuclide ventriculography or during stress echocardiography. Conversely, patients who can complete stage III of the Bruce exercise protocol and have a normal stress perfusion scan or negative stress echocardiographic evaluation are at very low risk for future coronary events. The finding of frequent episodes of ST-segment deviation on ambulatory ECG monitoring (even in the absence of symptoms) is also an adverse prognostic finding.

On cardiac catheterization, elevations of left ventricular end-diastolic pressure and ventricular volume and reduced ejection fraction are the most important signs of left ventricular dysfunction and are associated with a poor prognosis. Patients with chest discomfort but normal left ventricular function and normal coronary arteries have an excellent prognosis. Obstructive lesions of the left main (>50% luminal diameter) or left anterior descending coronary artery proximal to the origin of the first septal artery are associated with a greater risk than are lesions of the right or left circumflex coronary artery because of the greater quantity of myocardium at risk. Atherosclerotic plaques in epicardial arteries with fissuring or filling defects indicate increased risk. These lesions go through phases of inflammatory cellular activity, degeneration, endothelial dysfunction, abnormal vasomotion, platelet aggregation, and fissuring or hemorrhage. These factors can temporarily worsen the stenosis and cause thrombosis and/or abnormal reactivity of the vessel wall, thus exacerbating the manifestations of ischemia. The recent onset of symptoms, the development of severe ischemia during stress testing (see above), and unstable angina pectoris (Chap. 244) all reflect episodes of rapid progression in coronary lesions.

With any degree of obstructive CAD, mortality is greatly increased when left ventricular function is impaired; conversely, at any level of left ventricular function, the prognosis is influenced importantly by the quantity of myocardium perfused by critically obstructed vessels. Therefore, it is essential to collect all the evidence substantiating past myocardial damage (evidence of myocardial infarction on ECG, echocardiography, radioisotope imaging, or left ventriculography), residual left ventricular function (ejection fraction and wall motion), and risk of future damage from coronary events (extent of coronary disease and severity of ischemia defined by noninvasive stress testing). The larger the quantity of established myocardial

necrosis is, the less the heart is able to withstand additional damage and the poorer the prognosis is. Risk estimation must include age, presenting symptoms, all risk factors, signs of arterial disease, existing cardiac damage, and signs of impending damage (i.e., ischemia).

The greater the number and severity of risk factors for coronary atherosclerosis [advanced age (>75 years), hypertension, dyslipidemia, diabetes, morbid obesity, accompanying peripheral and/or cerebrovascular disease, previous myocardial infarction], the worse the prognosis of an angina patient. Evidence exists that elevated levels of C-reactive protein in the plasma, extensive coronary calcification on electron beam CT (see above), and increased carotid intimal thickening on ultrasound examination also indicate an increased risk of coronary events.

Treatment: Stable Angina Pectoris

Once the diagnosis of ischemic heart disease has been made, each patient must be evaluated individually with respect to his or her level of understanding, expectations and goals, control of symptoms, and prevention of adverse clinical outcomes such as myocardial infarction and premature death. The degree of disability as well as the physical and emotional stress that precipitates angina must be recorded carefully to set treatment goals. The management plan should include the following components: (1) explanation of the problem and reassurance about the ability to formulate a treatment plan, (2) identification and treatment of aggravating conditions, (3) recommendations for adaptation of activity as needed, (4) treatment of risk factors that will decrease the occurrence of adverse coronary outcomes, (5) drug therapy for angina, and (6) consideration of revascularization.

EXPLANATION AND REASSURANCE

Patients with IHD need to understand their condition and realize that a long and productive life is possible even though they have angina pectoris or have experienced and recovered from an acute myocardial infarction. Offering results of clinical trials showing improved outcomes can be of great value in encouraging patients to resume or maintain activity and return to work. A planned program of rehabilitation can encourage patients to lose weight, improve exercise tolerance, and control risk factors with more confidence.

IDENTIFICATION AND TREATMENT OF AGGRAVATING CONDITIONS

A number of conditions may increase oxygen demand or decrease oxygen supply to the myocardium and may precipitate or exacerbate angina in patients with IHD. Left ventricular hypertrophy, aortic valve disease, and hypertrophic cardiomyopathy may cause or contribute to angina and should be excluded or treated. Obesity, hypertension, and hyperthyroidism should be treated aggressively to reduce the frequency and severity of anginal episodes. Decreased myocardial oxygen supply may be due to reduced oxygenation of the arterial blood (e.g., in pulmonary disease or, when carboxyhemoglobin is present, due to cigarette or cigar smoking) or decreased oxygen-carrying capacity (e.g., in anemia). Correction of these abnormalities, if present, may reduce or even eliminate angina pectoris.

ADAPTATION OF ACTIVITY

Myocardial ischemia is caused by a discrepancy between the demand of the heart muscle for oxygen and the ability of the coronary circulation to meet that demand. Most patients can be helped to understand this concept and utilize it in the rational programming of activity. Many tasks that ordinarily evoke angina may be accomplished without symptoms simply by reducing the speed at which they are performed. Patients must appreciate the diurnal variation in their tolerance of certain activities and should reduce their energy requirements in the morning, immediately after meals, and in cold or inclement weather. On occasion, it may be necessary to recommend a change in employment or residence to avoid physical stress.

Physical conditioning usually improves the exercise tolerance of patients with angina and has substantial psychological benefits. A regular program of isotonic exercise that is within the limits of the individual patient's threshold for the development of angina pectoris and that does not exceed 80% of the heart rate associated with ischemia on exercise testing should be strongly encouraged. Based on the results of an exercise test, the number of metabolic equivalent tasks (METs) performed at the onset of ischemia can be estimated (Table 243-2) and a practical exercise prescription can be formulated to permit daily activities that will fall below the ischemic threshold (**Table 243-3**).

Table 243–3. Energy Requirements for Some Common Activities

Less Than 3 METs	3–5 METs	5–7 METs	7–9 METs	More Than 9 METs
Self-care				
Washing/shaving	Cleaning windows	Easy digging in garden	Heavy shoveling	Carrying loads upstairs (objects more than 90 lb)
Dressing	Raking	Level hand lawn mowing	Carrying objects (60–90 lb)	Climbing stairs (quickly)
Light housekeeping	Power lawn mowing	Carrying objects (30–60 lb)		Shoveling heavy snow
Desk work	Bed making/stripping			
Driving auto	Carrying objects (15–30 lb)			
Occupational				
Sitting (clerical/ assembly)	Stocking shelves (light objects	Carpentry (exterior)	Digging ditches (pick and shovel)	Heavy labor
Desk work	Light welding/carpentry	Shoveling dirt		
Standing (store clerk)		Sawing wood		
Recreational				
Golf (cart)	Dancing (social)	Tennis (singles)	Canoeing	Squash
Knitting	Golf (walking)	Snow skiing (downhill)	Mountain climbing	Ski touring
	Sailing	Light		Vigorous basketball

	Tennis (doubles)	Light backpacking Basketball Stream fishing		
Physical conditioning				
Walking (2 mph)	Level walking (3–4 mph)	Level walking (4.5–5.0 mph)	Level jogging (5 mph)	Running more than 6 mph
Stationary bike	Level biking (6–8 mph)	Bicycling (9–10 mph)	Swimming, breast stroke	Bicycling (more than 13 mph)
Very light calisthenics	Light calisthenics	Swimming (crawl stroke)	Rowing machine Heavy calisthenics Bicycling (12 mph)	Rope jumping Walking uphill (5 mph)

Abbreviation: METs, metabolic equivalent tasks.

Source: Modified from WL Haskell: Rehabilitation of the coronary patient, in NK Wenger, HK Hellerstein (eds): *Design and Implementation of Cardiac Conditioning Program*. New York, Churchill Livingstone, 1978.

TREATMENT OF RISK FACTORS

A *family history* of premature IHD is an important indicator of increased risk and should trigger a search for treatable risk factors such as hyperlipidemia, hypertension, and diabetes mellitus. *Obesity* impairs the treatment of other risk factors and increases the risk of adverse coronary events. In addition, obesity often is accompanied by three other risk factors: diabetes mellitus, hypertension, and hyperlipidemia. The treatment of obesity and these accompanying risk factors is an important component of any management plan. A diet low in saturated and *trans*-unsaturated fatty acids and a reduced caloric intake to achieve optimal body weight are a cornerstone in the management of chronic IHD. It is especially important to emphasize weight loss and regular exercise in patients with the metabolic syndrome or overt diabetes mellitus.

Cigarette smoking accelerates coronary atherosclerosis in both sexes and at all ages and increases the risk of thrombosis, plaque instability, myocardial infarction, and death (Chap. 241). In addition, by increasing myocardial oxygen needs and reducing oxygen supply, it aggravates angina. Smoking cessation studies have demonstrated important benefits with a significant decline in the occurrence of these adverse outcomes. The physician's message must be clear and strong and supported by programs that achieve and monitor abstinence (Chap. 395). *Hypertension* (Chap. 247) is associated with an increased risk of adverse clinical events from coronary atherosclerosis as well as stroke. In addition, the left ventricular hypertrophy that results from sustained hypertension aggravates ischemia. There is evidence that long-term effective treatment of hypertension can decrease the occurrence of adverse coronary events.

Diabetes mellitus (Chap. 344) accelerates coronary and peripheral atherosclerosis and is frequently associated with dyslipidemias and increases in the risk of angina, myocardial infarction, and sudden coronary death. Aggressive control of the dyslipidemia (target LDL cholesterol <70 mg/dL) and hypertension (target BP 120/80) that are frequently found in diabetic patients is highly effective and therefore essential, as described below.

DYSLIPIDEMIA

The treatment of dyslipidemia is central in aiming for long-term relief from angina, reduced need for revascularization, and reduction in myocardial infarction and death. The control of lipids can be achieved by the combination of a diet low in saturated and *trans*-unsaturated fatty acids, exercise, and weight loss. Nearly always, HMG-CoA reductase inhibitors (statins) are required and can lower LDL cholesterol (25–50%), raise HDL cholesterol (5–9%), and lower triglycerides (5–30%). A powerful treatment effect of statins on atherosclerosis, IHD, and outcomes is seen regardless of the pretreatment LDL cholesterol level. Fibrates or niacin can be used to raise HDL cholesterol and lower triglycerides (Chaps. 241 and 356). Controlled trials with lipid-regulating regimens have shown equal proportional benefit for men, women, the elderly, diabetic patients, and even smokers.

Compliance with the health-promoting behaviors listed above is generally very poor, and a conscientious physician must not underestimate the major effort required to meet this challenge. Fewer than one-half of patients in the United States discharged from the hospital with proven coronary disease receive treatment for dyslipidemia. In light of the proof that treating dyslipidemia brings major benefits, physicians need to establish treatment pathways, monitor compliance, and follow up regularly.

RISK REDUCTION IN WOMEN WITH IHD

The incidence of clinical IHD in premenopausal women is very low; however, after menopause, the atherogenic risk factors increase (e.g., increased LDL, reduced HDL) and the rate of clinical coronary events accelerates to the levels observed in men. Women have not given up cigarette smoking as effectively as have men. Diabetes mellitus, which is more common in women, greatly increases the occurrence of clinical IHD and amplifies the deleterious effects of hypertension, hyperlipidemia, and smoking. Cardiac catheterization and coronary revascularization are underused in women and are performed at a later and more severe stage of the disease than in men. When cholesterol lowering, beta blockers after myocardial infarction, and coronary artery bypass grafting are applied in the appropriate patient groups, women receive the same benefits of improved outcome as do men.

DRUG THERAPY

The commonly used drugs for the treatment of angina pectoris are summarized in Tables 243-4, 243-5, and 243-6. Pharmacotherapy for IHD is designed to reduce the frequency of anginal episodes, myocardial infarction, and coronary death. There is a wealth of positive trial data to emphasize how important this medical management is when added to the health-promoting behaviors discussed above. To achieve maximum benefit from medical therapy for IHD, it is frequently necessary to combine agents from different classes and titrate the doses as guided by the individual profile of risk factors, symptoms, hemodynamic responses, and side effects.

Table 243–4. Nitroglycerin and Nitrates for Patients with Ischemic Heart Disease

Compound	Route	Dose	Duration of Effect
Nitroglycerin	Sublingual tablets	0.3–0.6 mg up to 1.5 mg	Approximately 10 min
	Spray	0.4 mg as needed	Similar to sublingual tablets

	Ointment	2% 6 x 6 in. 15 x15 cm 7.5–40 mg	Effect up to 7 h
	Transdermal	0.2–0.8 mg/h every 12 h	8–12 h during intermittent therapy
	Oral sustained release	2.5–13 mg	4–8 h
	Intravenous	5–200 mcg/min	Tolerance may be seen in 7–8 h
Isosorbide dinitrate	Sublingual	2.5–10 mg	Up to 60 min
	Oral	5–80 mg, 2–3 times daily	Up to 8 h
	Spray	1.25 mg daily	2–3 min
	Chewable	5 mg	2–2 1/2 h
	Oral slow release	40 mg 1–2 daily	Up to 8 h
	Intravenous	1.25–5.0 mg/h	Tolerance in 7–8 h
	Ointment	100 mg/24 h	Not effective
Isosorbide mononitrate	Oral	20 mg twice daily 60–240 mg once daily	12–24 h
Pentaerythritol tetranitrate	Sublingual	10 mg as needed	Not known

Source: Modified from RJ Gibbons et al.

Table 243–5. Properties of Beta Blockers in Clinical Use for Ischemic Heart Disease

Drugs	Selectivity	Partial Agonist Activity	Usual Dose for Angina
Acebutolol	β1	Yes	200–600 mg twice daily
Atenolol	β1	No	50–200 mg/d
Betaxolol	β1	No	10–20 mg/d
Bisoprolol	β1	No	10 mg/d
Esmolol (intravenous)[a]	β1	No	50–300 mcg/kg/min
Labetalol[b]	None	Yes	200–600 mg twice daily
Metoprolol	β1	No	50–200 mg twice daily
Nadolol	None	No	40–80 mg/day
Nebivolol	β1 (at low doses)	No	5–40 mg/day
Pindolol	None	Yes	2.5–7.5 mg 3 times daily
Propranolol	None	No	80–120 mg twice daily
Timolol	None	No	10 mg twice daily

Note: This list of beta blockers that may be used to treat patients with angina pectoris is arranged alphabetically. The agents for which there is the greatest clinical experience include atenolol, metoprolol, and propranolol. It is preferable to use a sustained-release formulation that may be taken once daily to improve the patient's compliance with the regimen.

[a]Esmolol is an ultra-short-acting beta blocker that is administered as a continuous intravenous infusion. Its rapid offset of action makes esmolol an attractive agent to use in patients with relative contraindications to beta blockade.

[b]Labetolol is a combined alpha and beta blocker.

Source: Modified from RJ Gibbons et al.

Table 243–6. Calcium Channel Blockers in Clinical Use for Ischemic Heart Disease

Drugs	Usual Dose	Duration of Action	Side Effects
Dihydropyridines			
Amlodipine	5–10 mg qd	Long	Headache, edema
Felodipine	5–10 mg qd	Long	Headache, edema
Isradipine	2.5–10 mg bid	Medium	Headache, fatigue
Nicardipine	20–40 mg tid	Short	Headache, dizziness, flushing, edema
Nifedipine	Immediate release:* 30–90 mg daily orally Slow release: 30–180 mg orally	Short	Hypotension, dizziness, flushing, nausea, constipation, edema
Nisoldipine	20–40 mg qd	Short	Similar to nifedipine
Nondihydropyridines			
Diltiazem	Immediate release: 30–80 mg 4 times daily Slow release: 120–320 mg qd	Short Long	Hypotension, dizziness, flushing, bradycardia, edema
Verapamil	Immediate release: 80–160 mg tid Slow release: 120–480 mg qd	Short Long	Hypotension, myocardial depression, heart failure, edema, bradycardia

Note: This list of calcium channel blockers that may be used to treat patients with angina pectoris is divided into two broad classes, dihydropyridines and nondihydropyridines, and arranged alphabetically within each class. Among the dihydropyridines, the greatest clinical experience has been obtained with amlodipine and nifedipine. After the initial period of dose titration with a short-acting formulation, it is preferable to switch to a sustained-release formulation that may be taken once daily to improve patient compliance with the regimen.

*May be associated with increased risk of mortality if administered during acute myocardial infarction.

Source: Modified from RJ Gibbons et al.

NITRATES

The organic nitrates are a valuable class of drugs in the management of angina pectoris (**Table 243-4**). Their major mechanisms of action include systemic venodilation with concomitant reduction in left ventricular end-diastolic volume and pressure, thereby reducing myocardial wall tension and oxygen requirements; dilation of epicardial coronary vessels; and increased blood flow in collateral vessels. When metabolized, organic nitrates release nitric oxide (NO) that binds to guanylyl cyclase in vascular smooth muscle cells, leading to an increase in cyclic guanosine monophosphate, which causes relaxation of vascular smooth muscle. Nitrates also exert antithrombotic activity by NO-dependent activation of platelet guanylyl cyclase, impairment of intraplatelet calcium flux, and platelet activation.

The absorption of these agents is most rapid and complete through the mucous membranes. For this reason, nitroglycerin is most commonly administered sublingually in tablets of 0.4 or 0.6 mg. Patients with angina should be instructed to take the medication both to relieve angina and also approximately 5 min before stress that is likely to induce an episode. The value of this prophylactic use of the drug cannot be overemphasized.

Nitrates improve exercise tolerance in patients with chronic angina and relieve ischemia in patients with unstable angina as well as patients with Prinzmetal's variant angina (Chap. 244). A diary of angina and nitroglycerin use may be valuable for detecting changes in the frequency, severity, or threshold for discomfort that may signify the development of unstable angina pectoris and/or herald an impending myocardial infarction.

Long-Acting Nitrates

None of the long-acting nitrates are as effective as sublingual nitroglycerin for the acute relief of angina. These organic nitrate preparations can be swallowed, chewed, or administered as a patch or paste by the transdermal route (Table 243-4). They can provide effective plasma levels for up to 24 h, but the therapeutic response is highly variable. Different preparations and/or administration during the daytime should be tried only to prevent discomfort while avoiding side effects such as headache and dizziness. Individual dose titration is important to prevent side effects. To minimize the effects of tolerance, the minimum effective dose should be used and a minimum of 8 h each day kept free of the drug to restore any useful response(s).

β-Adrenergic Blockers

These drugs represent an important component of the pharmacologic treatment of angina pectoris (**Table 243-5**). They reduce myocardial oxygen demand by inhibiting the increases in heart rate, arterial pressure, and myocardial contractility caused by adrenergic activation. Beta blockade reduces these variables most strikingly during exercise but causes only small reductions at rest. Long-acting beta-blocking drugs or sustained-release formulations offer the advantage of once-daily dosing (Table 243-5). The therapeutic aims include relief of angina and ischemia. These drugs also can reduce mortality and reinfarction rates in patients after myocardial infarction and are moderately effective antihypertensive agents.

Relative contraindications include asthma and reversible airway obstruction in patients with chronic lung disease, atrioventricular conduction disturbances, severe bradycardia, Raynaud's phenomenon, and a history of mental depression. Side effects include fatigue, reduced exercise tolerance, nightmares, impotence, cold extremities, intermittent claudication, bradycardia (sometimes severe), impaired atrioventricular conduction, left ventricular failure, bronchial asthma, worsening claudication, and intensification of the hypoglycemia produced by oral hypoglycemic agents and insulin. Reducing the dose or even discontinuation may be necessary if these side effects develop and persist. Since sudden discontinuation can intensify ischemia, the doses should be tapered over 2 weeks. Beta blockers with relative β_1-receptor specificity such as metoprolol and atenolol may be preferable in patients with mild bronchial obstruction and insulin-requiring diabetes mellitus.

Calcium Channel Blockers

Calcium channel blockers (**Table 243-6**) are coronary vasodilators that produce variable and dose-dependent reductions in myocardial oxygen demand, contractility, and arterial pressure. These combined pharmacologic effects are advantageous and make these agents as effective as beta blockers in the treatment of angina pectoris. They are indicated when beta blockers are contraindicated, poorly tolerated, or ineffective. Verapamil and diltiazem may produce symptomatic disturbances in cardiac conduction and bradyarrhythmias. They also exert negative inotropic actions and are more likely to aggravate left ventricular failure, particularly when used in patients with left ventricular dysfunction, especially if the patients are also receiving beta blockers. Although useful effects usually are achieved when calcium channel blockers are combined with beta blockers and nitrates, individual titration of the doses is essential with these combinations. Variant (Prinzmetal's) angina responds particularly well to calcium channel blockers (especially members of the dihydropyridine class), supplemented when necessary by nitrates (Chap. 244).

Verapamil ordinarily should not be combined with beta blockers because of the combined adverse effects on heart rate and contractility. Diltiazem can be combined with beta blockers in patients with normal ventricular function and no conduction disturbances. Amlodipine and beta blockers have complementary actions on coronary blood supply and myocardial oxygen demands. Whereas the former decreases blood pressure and dilates coronary arteries, the latter slows heart rate and decreases contractility. Amlodipine and the other second-generation dihydropyridine calcium antagonists (nicardipine, isradipine, long-acting nifedipine, and felodipine) are potent vasodilators and are useful in the simultaneous treatment of angina and hypertension. Short-acting dihydropyridines should be avoided because of the risk of precipitating infarction, particularly in the absence of concomitant beta-blocker therapy.

Choice between Beta Blockers and Calcium Channel Blockers for Initial Therapy

Since beta blockers have been shown to improve life expectancy after acute myocardial infarction (Chaps. 244 and 245) and calcium channel blockers have not, the former may also be preferable in patients with angina and a damaged left ventricle. However, calcium channel blockers are indicated in patients with the following: (1) inadequate responsiveness to the combination of beta blockers and nitrates; many of these patients do well with a combination of a beta blocker and a dihydropyridine calcium channel blocker, (2) adverse reactions to beta blockers such as depression, sexual disturbances,

and fatigue, (3) angina and a history of asthma or chronic obstructive pulmonary disease, (4) sick-sinus syndrome or significant atrioventricular conduction disturbances, (5) Prinzmetal's angina, or (6) symptomatic peripheral arterial disease.

Antiplatelet Drugs

Aspirin is an irreversible inhibitor of platelet cyclooxygenase and thereby interferes with platelet activation. Chronic administration of 75–325 mg orally per day has been shown to reduce coronary events in asymptomatic adult men over age 50, patients with chronic stable angina, and patients who have or have survived unstable angina and myocardial infarction. There is a dose-dependent increase in bleeding when aspirin is used chronically. It is preferable to use an enteric-coated formulation in the range of 81-162 mg/d. Administration of this drug should be considered in all patients with IHD in the absence of gastrointestinal bleeding, allergy, or dyspepsia. Clopidogrel (300–600 mg loading and 75 mg/d) is an oral agent that blocks P2Y12 ADP receptor–mediated platelet aggregation. It provides benefits similar to those of aspirin in patients with stable chronic IHD and may be substituted for aspirin if aspirin causes the side effects listed above. Clopidogrel combined with aspirin reduces death and coronary ischemic events in patients with an acute coronary syndrome (Chap. 244) and also reduces the risk of thrombus formation in patients undergoing implantation of a stent in a coronary artery (Chap. 246). Alternative antiplatelet agents that block the P2Y12 platelet receptor such as prasugrel have been shown to be more effective than clopidogrel for prevention of ischemic events after placement of a stent for an acute coronary syndrome but are associated with an increased risk of bleeding. Although combined treatment with clopidogrel and aspirin for at least a year is recommended in patients with an acute coronary syndrome treated with implantation of a drug-eluting stent, studies have not shown any benefit from the routine addition of clopidogrel to aspirin in patients with chronic stable IHD.

OTHER THERAPIES

The angiotensin-converting enzyme (ACE) inhibitors are widely used in the treatment of survivors of myocardial infarction, patients with hypertension or chronic IHD including angina pectoris, and those at high risk of vascular diseases such as diabetes. The benefits of ACE inhibitors are most evident in IHD patients at increased risk, especially if diabetes mellitus or LV dysfunction is present, and those who have not achieved adequate control of blood pressure and LDL cholesterol on beta blockers and statins. However, the routine administration of ACE inhibitors to IHD patients who have normal LV function and have achieved blood pressure and LDL goals on other therapies does not reduce the incidence of events and therefore is not cost-effective.

Despite treatment with nitrates, beta blockers, or calcium channel blockers, some patients with IHD continue to experience angina, and additional medical therapy is now available to alleviate their symptoms. Ranolazine, a piperazine derivative, may be useful for patients with chronic angina despite standard medical therapy. Its antianginal action is believed to occur via inhibition of the late inward sodium current (I_{Na}). The benefits of I_{Na} inhibition include limitation of the Na overload of ischemic myocytes and prevention of Ca^{2+} overload via the Na^{+}–Ca^{2+} exchanger. A dose of 500–1000 mg orally twice daily is usually well tolerated. Ranolazine is contraindicated in patients with hepatic impairment or with conditions or drugs associated with QT_C prolongation and when drugs that inhibit the CYP3A metabolic system (e.g., ketoconazole, diltiazem, verapamil, macrolide antibiotics, HIVprotease inhibitors, and large quantities of grapefruit juice) are being used.

Nonsteroidal anti-inflammatory drug (NSAID) use in patients with IHD may be associated with a small but finite increased risk of MI and mortality. For this reason, they generally should be avoided in IHD patients. If they are required for symptom relief, it is advisable to coadminister aspirin and strive to use the lowest NSAID dose required for the shortest period of time.

Another class of agents open ATP-sensitive potassium channels in myocytes, leading to a reduction of free intracellular calcium ions. The major drug in this class is nicorandil, which typically is administered orally in a dose of 20 mg twice daily for prevention of angina. (Nicorandil is not available for use in the United States but is used in several other countries.)

Angina and Heart Failure

Transient left ventricular failure with angina can be controlled by the use of nitrates. For patients with established congestive heart failure the increased left ventricular wall tension raises myocardial oxygen demand. Treatment of congestive heart failure with an angiotensin-converting enzyme inhibitor, a diuretic, and digoxin (Chap. 234) reduces heart size, wall tension, and myocardial oxygen demand, which helps control angina and ischemia. If the symptoms and signs of heart failure are controlled, an effort should be made to use beta blockers not only for angina but because trials in heart failure have shown significant improvement in survival. A trial of the intravenous ultra-short-acting beta blocker esmolol may be useful to establish the safety of beta blockade in selected patients. Nocturnal angina often can be relieved by the treatment of heart failure.

The combination of congestive heart failure and angina in patients with IHD usually indicates a poor prognosis and warrants serious consideration of cardiac catheterization and coronary revascularization.

CORONARY REVASCULARIZATION

Clinical trials have confirmed that with the initial diagnosis of stable IHD, it is first appropriate to initiate a thorough medical regimen as described above. Revascularization should be considered in the presence of unstable phases of the disease, intractable symptoms, severe ischemia or high-risk coronary anatomy, diabetes, and impaired LV function. *Revascularization should be employed in conjunction with but not replace the continuing need to modify risk factors and assess medical therapy.* An algorithm for integrating medical therapy and revascularization options in patients with IHD is shown in **Fig. 243-2**.

Figure 243-2.

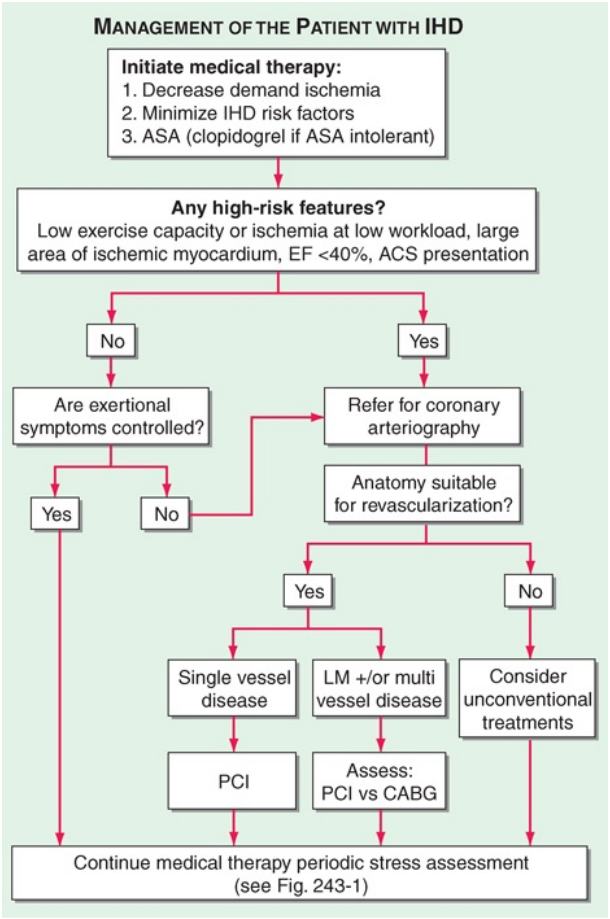

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for management of a patient with ischemic heart disease. All patients should receive the core elements of medical therapy as shown at the top of the algorithm. If high-risk features are present, as established by the clinical history, exercise test data, and imaging studies, the patient should be referred for coronary arteriography. Based on the number and location of the diseased vessels and their suitability for revascularization, the patient is treated with a percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) surgery or should be considered for unconventional treatments. See text for further discussion. IHD, ischemic heart disease; ASA, aspirin; EF, ejection fraction; ACS, acute coronary syndrome; LM, left main.

Percutaneous Coronary Intervention

(See also Chap. 246) Percutaneous coronary intervention (PCI) involving balloon dilatation usually accompanied by coronary stenting is widely used to achieve revascularization of the myocardium in patients with symptomatic IHD and suitable stenoses of epicardial coronary arteries. Whereas patients with stenosis of the left main coronary artery and those with three-vessel IHD (especially with diabetes and/or impaired left ventricular function) who require revascularization are best treated with CABG, PCI is widely employed in patients with symptoms and evidence of ischemia due to stenoses of one or two vessels and even in selected patients with three-vessel disease (and, perhaps, in some patients with left main disease) and may offer many advantages over surgery.

INDICATIONS AND PATIENT SELECTION

The most common clinical indication for PCI is symptom-limiting angina pectoris, despite medical therapy, accompanied by evidence of ischemia during a stress test. PCI is more effective than medical therapy for the relief of angina. PCI improves outcomes in patients with unstable angina or when used early in the course of myocardial infarction with and without cardiogenic shock. However, in patients with stable exertional angina, clinical trials have confirmed that PCI does not reduce the occurrence of death or myocardial infarction compared to optimum medical therapy. PCI can be used to treat stenoses in native coronary arteries as well as in bypass grafts in patients who have recurrent angina after CABG.

RISKS

When coronary stenoses are discrete and symmetric, two and even three vessels can be treated in sequence. However, case selection is essential to avoid a prohibitive risk of complications, which are usually due to dissection or thrombosis with vessel occlusion, uncontrolled ischemia, and ventricular failure (Chap. 246). Oral aspirin, a thienopyridine, and an antithrombin agent are given to reduce coronary thrombus formation. Left main coronary artery stenosis generally is regarded as a contraindication to PCI; such patients should be treated with CABG. In selected cases such as patients with prohibitive surgical risks, PCI of an unprotected left main can be considered, but such a procedure should be performed only by a highly skilled operator; importantly, there are regional differences in the use of this approach internationally.

EFFICACY

Primary success, i.e., adequate dilation (an increase in luminal diameter >20% to a residual diameter obstruction <50%) with relief of angina, is achieved in >95% of cases. Recurrent stenosis of the dilated vessels occurs in ~20% of cases within 6 months of PCI with bare metal stents, and angina will recur within 6 months in 10% of cases. Restenosis is more common in patients with diabetes mellitus, arteries with small caliber, incomplete dilation of the stenosis, long stents, occluded vessels, obstructed vein grafts, dilation of the left anterior descending coronary artery, and stenoses containing thrombi. In diseased vein grafts, procedural success has been improved by the use of capture devices or filters that

prevent embolization, ischemia, and infarction.

It is usual clinical practice to administer aspirin indefinitely and a thienopyridine for 1–3 months after the implantation of a bare metal stent. Although aspirin in combination with a thienopyridine may help prevent coronary thrombosis during and shortly after PCI with stenting, there is no evidence that these medications reduce the incidence of restenosis.

The use of drug-eluting stents that locally deliver antiproliferative drugs can reduce restenosis to less than 10%. Advances in PCI, especially the availability of drug-eluting stents, have vastly extended the use of this revascularization option in patients with IHD. Of note, however, the delayed endothelial healing in the region of a drug-eluting stent also extends the period during which the patient is at risk for subacute stent thrombosis. Current recommendations are to administer aspirin indefinitely and a thienopyridine daily for at least one year after implantation of a drug-eluting stent. When a situation arises in which temporary discontinuation of antiplatelet therapy is necessary, the clinical circumstances should be reviewed with the operator who performed the PCI and a coordinated plan should be established for minimizing the risk of late stent thrombus; central to this plan is the discontinuation of antiplatelet therapy for the shortest acceptable period. The risk of stent thrombosis is dependent on stent size and length, complexity of the lesions, age, diabetes, and technique. However, compliance with dual antiplatelet therapy and individual responsiveness to platelet inhibition are very important factors as well.

Successful PCI produces effective relief of angina in >95% of cases. More than one-half of patients with symptomatic IHD who require revascularization can be treated initially by PCI. Successful PCI is less invasive and expensive than CABG and permits savings in the *initial* cost of care. Successful PCI avoids the risk of stroke associated with CABG surgery, allows earlier return to work, and allows the resumption of an active life. However, the early health-related and economic benefit of PCI is reduced over time because of the greater need for follow-up and the increased need for repeat procedures. When directly compared in patients with diabetes or three-vessel or left main coronary artery disease, CABG was superior to PCI in preventing major adverse cardiac or cerebrovascular events over a 12-month follow-up.

Coronary Artery Bypass Grafting

Anastomosis of one or both of the internal mammary arteries or a radial artery to the coronary artery distal to the obstructive lesion is the preferred procedure. For additional obstructions that cannot be bypassed by an artery, a section of a vein (usually the saphenous) is used to form a connection between the aorta and the coronary artery distal to the obstructive lesion.

Although some indications for CABG are controversial, certain areas of agreement exist:

1. The operation is relatively safe, with mortality rates <1% in patients without serious comorbid disease and normal left ventricular function and when the procedure is performed by an experienced surgical team.
2. Intraoperative and postoperative mortality rates increase with the severity of ventricular dysfunction, comorbidities, age >80 years, and lack of surgical experience. The effectiveness and risk of CABG vary widely depending on case selection and the skill and experience of the surgical team.
3. Occlusion of *venous* grafts is observed in 10 to 20% of patients during the first postoperative year and in approximately 2% per year during 5- to 7-year follow-up and 4% per year thereafter. Long-term patency rates are considerably higher for internal mammary and radial artery implantations than for saphenous vein grafts. In patients with left anterior descending coronary artery obstruction, survival is better when coronary bypass involves the internal mammary artery rather than a saphenous vein. Graft patency and outcomes are improved by meticulous treatment of risk factors, particularly dyslipidemia.
4. Angina is abolished or greatly reduced in ~90% of patients after complete revascularization. Although this usually is associated with graft patency and restoration of blood flow, the pain may also have been alleviated as a result of infarction of the ischemic segment or a placebo effect. Within 3 years, angina recurs in about one-fourth of patients but is rarely severe.
5. Survival may be improved by operation in patients with stenosis of the left main coronary artery as well as in patients with three- or two-vessel disease with significant obstruction of the proximal left anterior descending coronary artery. The survival benefit is greater in patients with abnormal LV function (ejection fraction <50%). Survival *may* also be improved in the following patients: (a) with obstructive CAD who have survived sudden cardiac death or sustained ventricular tachycardia, (b) who have undergone previous CABG and have multiple saphenous vein graft stenoses, especially of a graft supplying the left anterior descending coronary artery, and (c) with recurrent stenosis after PCI and high-risk criteria on noninvasive testing.
6. Minimally invasive CABG through a small thoracotomy and/or off-pump surgery can reduce morbidity and shorten convalescence in suitable patients but does not appear to reduce significantly the risk of neurocognitive dysfunction postoperatively.
7. Among patients with type 2 diabetes mellitus and multivessel coronary disease, CABG surgery plus optimal medical therapy is superior to optimal medical therapy alone in preventing major cardiovascular events, a benefit mediated largely by a significant reduction in nonfatal MI. The benefits of CABG are especially evident in diabetic patients treated with an insulin-sensitizing strategy as opposed to an insulin-providing strategy.

Indications for CABG usually are based on the severity of symptoms, coronary anatomy, and ventricular function. The ideal candidate is male, is <80 years of age, has no other complicating disease, and has troublesome or disabling angina that is not adequately controlled by medical therapy or does not tolerate medical therapy. The patient wishes to lead a more active life and has severe stenoses of two or three epicardial coronary arteries with objective evidence of myocardial ischemia as a cause of the chest discomfort. Great symptomatic benefit can be anticipated in such patients. Congestive heart failure and/or left ventricular dysfunction, advanced age (>80 years), reoperation, urgent need for surgery, and the presence of diabetes mellitus are all associated with a higher perioperative mortality rate.

Left ventricular dysfunction can be due to noncontractile or hypocontractile segments that are viable but are chronically ischemic (hibernating myocardium). As a consequence of chronic reduction in myocardial blood flow, these segments downregulate their contractile function. They can be detected by using radionuclide scans of myocardial perfusion and metabolism, PET, cardiac MRI, or delayed scanning with thallium-201 or by improvement of regional functional impairment provoked by low-dose dobutamine. In such patients, revascularization improves myocardial blood flow, can return function, and can improve survival.

THE CHOICE BETWEEN PCI AND CABG

All the clinical characteristics of each individual patient must be used to decide on the method of revascularization (LV function, diabetes, lesion complexity, etc). A number of randomized clinical trials have compared PCI and CABG in patients with multivessel CAD who were suitable technically for both procedures. The redevelopment of angina requiring repeat

coronary angiography and repeat revascularization is higher with PCI. This is a result of restenosis in the stented segment (a problem largely solved with drug-eluting stents) and the development of new stenoses in unstented portions of the coronary vasculature. It has been argued that PCI with stenting focuses on culprit lesions whereas a bypass graft to the target vessel also provides a conduit around future culprit lesions proximal to the anastomosis of the graft to the native vessel (**Fig. 243-3**). By contrast, stroke rates are lower with PCI.

Figure 243-3.

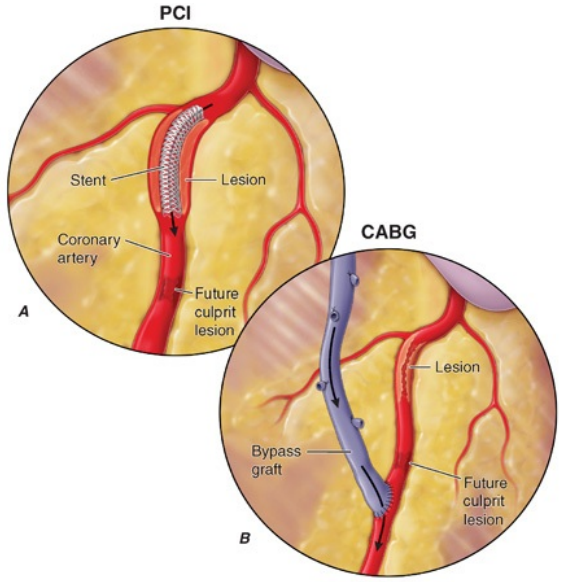

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Difference in the approach to the lesion with percutaneous coronary intervention (PCI) and coronary artery bypass grafting (CABG). PCI is targeted at the "culprit" lesion or lesions, whereas CABG is directed at the epicardial vessel, including the culprit lesion or lesions and future culprits, proximal to the insertion of the vein graft, a difference that may account for the superiority of CABG, at least in the intermediate term, in patients with multivessel disease. *(Reproduced from BJ Gersh et al: N Engl J Med 352:2235, 2005.)*

Comparison of mortality rates in patients treated with CABG versus PCI is a complex issue. There is an early increased risk of mortality with CABG, but mortality rates appear similar in the two revascularization strategies over the long term.

Based on available evidence, it is now recommended that patients with an unacceptable level of angina despite optimal medical management be considered for coronary revascularization. Patients with single- or two-vessel disease with normal LV function and anatomically suitable lesions ordinarily are advised to undergo PCI (Chap. 246). Patients with three-vessel disease (or two-vessel disease that includes the proximal left descending coronary artery) and impaired global LV function (LV ejection fraction < 50%) or diabetes mellitus and those with left main coronary artery disease or other lesions unsuitable for catheter-based procedures should be considered for CABG as the initial method of revascularization. In light of the complexity of the decision making, it is desirable to have a multidisciplinary team, including a cardiologist and a cardiac surgeon in conjunction with the patient's primary care physician, provide input in conjunction with ascertaining the patient's preferences before committing to a particular revascularization option.

Unconventional Treatments for Ihd

On occasion clinicians will encounter a patient who has persistent disabling angina despite maximally tolerated medical therapy and for whom revascularization is not an option (e.g., small diffusely diseased vessels not amenable to stent implantation or acceptable targets for bypass grafting). In such situations unconventional treatments should be considered.

Enhanced external counterpulsation utilizes pneumatic cuffs on the lower extremities to provide diastolic augmentation and systolic unloading of blood pressure to decrease cardiac work and oxygen consumption while enhancing coronary blood flow. Clinical trials have shown that regular application improves angina, exercise capacity, and regional myocardial perfusion. Experimental approaches such as gene and stem cell therapies are also under active study.

ASYMPTOMATIC (SILENT) ISCHEMIA

Obstructive CAD, acute myocardial infarction, and transient myocardial ischemia are frequently asymptomatic. During continuous ambulatory ECG monitoring, the majority of ambulatory patients with typical chronic stable angina are found to have objective evidence of myocardial ischemia (ST-segment depression) during episodes of chest discomfort while they are active outside the hospital. In addition, many of these patients also have more frequent episodes of asymptomatic ischemia. Frequent episodes of ischemia (symptomatic and asymptomatic) during daily life appear to be associated with an increased likelihood of adverse coronary events (death and myocardial infarction). In addition, patients with asymptomatic ischemia after a myocardial infarction are at greater risk for a second coronary event. The widespread use of exercise ECG during routine examinations has also identified some of these previously unrecognized patients with asymptomatic CAD. Longitudinal studies have demonstrated an increased incidence of coronary events in asymptomatic patients with positive exercise tests.

Treatment

ASYMPTOMATIC ISCHEMIA

The management of patients with asymptomatic ischemia must be individualized. When coronary disease has been confirmed, the aggressive treatment of hypertension and dyslipidemia is essential and will decrease the risk of infarction and death. In addition, the physician should consider the following: (1) the degree of positivity of the stress test,

particularly the stage of exercise at which ECG signs of ischemia appear; the magnitude and number of the ischemic zones of myocardium on imaging; and the change in LV ejection fraction that occurs on radionuclide ventriculography or echocardiography during ischemia and/or during exercise, (2) the ECG leads showing a positive response, with changes in the anterior precordial leads indicating a less favorable prognosis than changes in the inferior leads, and (3) the patient's age, occupation, and general medical condition.

Most would agree that an asymptomatic 45-year-old commercial airline pilot with significant (0.4-mV) ST-segment depression in leads V_1 to V_4 during mild exercise should undergo coronary arteriography, whereas an asymptomatic, sedentary 85-year-old retiree with 0.1-mV ST-segment depression in leads II and III during maximal activity need not. However, there is no consensus about the most appropriate approach in the large majority of patients for whom the situation is less extreme. Asymptomatic patients with silent ischemia, three-vessel CAD, and impaired LV function may be considered appropriate candidates for CABG.

The treatment of risk factors, particularly lipid lowering and blood pressure control as described above, and the use of aspirin, statins, and beta blockers after infarction have been shown to reduce events and improve outcomes in asymptomatic as well as symptomatic patients with ischemia and proven CAD. Although the incidence of asymptomatic ischemia can be reduced by treatment with beta blockers, calcium channel blockers, and long-acting nitrates, it is not clear whether this is necessary or desirable in patients who have not had a myocardial infarction.

FURTHER READINGS

Antman EM et al: Cyclooxygenase inhibition and cardiovascular risk. Circulation 112:759, 2005[PMID: 16061757]

Bhatt DL et al: Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. N Engl J Med 354:1706, 2006[PMID: 16531616]

Boden WE et al: Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 356:1503, 2007[PMID: 17387127]

Chaitman BR: Ranolazine for the treatment of chronic angina and potential use in other cardiovascular conditions. Circulation 113:2462, 2006[PMID: 16717165]

Cleeman JI et al: Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA 285:2486, 2001

DeFeyter PJ et al: Bypass surgery versus stenting for the treatment of multivessel disease in patients with unstable angina compared with stable angina. Circulation 105:2367, 2002

Frye RL et al: A randomized trial of therapies for type 2 diabetes and coronary artery disease. N Engl J Med 360:2503, 2009[PMID: 19502645]

Gibbons RJ et al: ACC/AHA 2002 guideline update for the management of patients with chronic stable angina: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1999 Guidelines for the Management of Patients with Chronic Stable Angina. J Am Coll Cardiol 41:159, 2002

Goldschmidt-Clermont PJ et al: Atherosclerosis 2005: Recent discoveries and novel hypotheses. Circulation 112:3348, 2005[PMID: 16301361]

Kim MC et al: Refractory angina pectoris: Mechanism and therapeutic options. J Am Coll Cardiol 39:923, 2002[PMID: 11897431]

Kushner FG et al: 2009 Focused Updates: ACC/AHA Guidelines for the Management of Patients With ST-Elevation Myocardial Infarction (updating the 2004 Guideline and 2007 Focused Update) and ACC/AHA/SCAI Guidelines on Percutaneous Coronary Intervention (updating the 2005 Guideline and 2007 Focused Update): A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 120:2271, 2009[PMID: 19923169]

Lee TH et al: Noninvasive tests in patients with stable coronary artery disease. N Engl J Med 344:1840, 2001[PMID: 11407346]

Marasco SF et al: No improvement in neurocognitive outcomes after off-pump versus on-pump coronary revascularisation: A meta-analysis. Eur J Cardiothorac Surg 33:961, 2008[PMID: 18424064]

McQueen MJ et al: Lipids, lipoproteins, and apolipoproteins as risk markers of myocardial infarction in 52 countries (the INTERHEART study): A case-control study. Lancet 372:224, 2008[PMID: 18640459]

Mega JL et al: Cytochrome p-450 polymorphisms and response to clopidogrel. N Engl J Med 360:354, 2009[PMID: 19106084]

Morrow D et al: Chronic coronary artery disease, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

Opie LH et al: Controversies in stable coronary artery disease. Lancet 367:69, 2006[PMID: 16399156]

Pretre R et al: Choice of revascularization strategy for patients with coronary artery disease. JAMA 285:992, 2001[PMID: 11209153]

Serruys PW et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. N Engl J Med 360:961, 2009[PMID: 19228612]

Van Den Brand MJBM et al: The effect of completeness of revascularization on event-free survival at one year in the ARTS trial. J Am Coll Cardiol 39:559, 2002

Weintraub WS et al: Predicting cardiovascular events with coronary calcium scoring. N Engl J Med 358:1394, 2008[PMID: 18367744]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

UNSTABLE ANGINA AND INON-ST-SEGMENT ELEVATION MYOCARDIAL INFARCTION:

INTRODUCTION

Patients with ischemic heart disease fall into two large groups: patients with chronic coronary artery disease (CAD) who most commonly present with stable angina (Chap. 243) and patients with acute coronary syndromes (ACSs). The latter group, in turn, is composed of patients with acute myocardial infarction (MI) with ST-segment elevation on their presenting electrocardiogram (ECG) (STEMI; Chap. 245) and those with unstable angina (UA) and non-ST-segment elevation MI (UA/NSTEMI; Fig. 245-1). Every year in the United States, approximately 1 million patients are admitted to hospitals with UA/NSTEMI as compared with ~300,000 patients with acute STEMI. The relative incidence of UA/NSTEMI compared to STEMI appears to be increasing. More than one-third of patients with UA/NSTEMI are women, while less than one-fourth of patients with STEMI are women.

Definition

The diagnosis of UA is based largely on the clinical presentation. *Stable* angina pectoris is characterized by chest or arm discomfort that may not be described as pain but is reproducibly associated with physical exertion or stress and is relieved within 5–10 minutes by rest and/or sublingual nitroglycerin (Chaps. 12 and 343). UA is defined as angina pectoris or equivalent ischemic discomfort with at least one of three features: (1) it occurs at rest (or with minimal exertion), usually lasting >10 minutes; (2) it is severe and of new onset (i.e., within the prior 4–6 weeks); and/or (3) it occurs with a crescendo pattern (i.e., distinctly more severe, prolonged, or frequent than previously). The diagnosis of NSTEMI is established if a patient with the clinical features of UA develops evidence of myocardial necrosis, as reflected in elevated cardiac biomarkers.

Pathophysiology

UA/NSTEMI is most commonly caused by a reduction in oxygen supply and/or by an increase in myocardial oxygen demand superimposed on a lesion that causes coronary arterial obstruction, usually an atherothrombotic coronary plaque. Four pathophysiologic processes that may contribute to the development of UA/NSTEMI have been identified: (1) plaque rupture or erosion with a superimposed nonocclusive thrombus, believed to be the most common cause; in such patients, NSTEMI may occur with downstream embolization of platelet aggregates and/or atherosclerotic debris; (2) dynamic obstruction [e.g., coronary spasm, as in Prinzmetal's variant angina (PVA) (p. 2020)]; (3) progressive mechanical obstruction [e.g., rapidly advancing coronary atherosclerosis or restenosis following percutaneous coronary intervention (PCI)]; and (4) UA secondary to increased myocardial oxygen demand and/or decreased supply (e.g., tachycardia, anemia). More than one of these processes may be involved.

Among patients with UA/NSTEMI studied at angiography, approximately 5% have stenosis of the left main coronary artery, 15% have three-vessel CAD, 30% have two-vessel disease, 40% have single-vessel disease, and 10% have no apparent critical epicardial coronary artery stenosis; some of the latter may have obstruction of the coronary microcirculation. The "culprit lesion" may show an eccentric stenosis with scalloped or overhanging edges and a narrow neck on angiography. Angioscopy has been reported to show "white" (platelet-rich) thrombi, as opposed to "red" (fibrin- and cell-rich) thrombi; the latter are more often seen in patients with acute STEMI. Patients with UA/NSTEMI frequently have multiple plaques at risk of disruption (vulnerable plaques).

Clinical Presentation

HISTORY AND PHYSICAL EXAMINATION

The clinical hallmark of UA/NSTEMI is chest pain, typically located in the substernal region or sometimes in the epigastrium, that radiates to the neck, left shoulder, and/or the left arm (Chap. 12). This discomfort is usually severe enough to be described as frank pain. Anginal "equivalents" such as dyspnea and epigastric discomfort may also occur, and these appear to be more frequent in women. The physical examination resembles that in patients with stable angina (Chap. 243) and may be unremarkable. If the patient has a large area of myocardial ischemia or a large NSTEMI, the physical findings can include diaphoresis; pale, cool skin; sinus tachycardia; a third and/or fourth heart sound; basilar rales; and, sometimes, hypotension, resembling the findings of large STEMI.

ELECTROCARDIOGRAM

In UA, ST-segment depression, transient ST-segment elevation, and/or T-wave inversion occur in 30 to 50% of patients. In patients with the clinical features of UA, the presence of new ST-segment deviation, even of only 0.05 mV, is an important predictor of adverse outcome. T-wave changes are sensitive for ischemia but less specific, unless they are new, deep T-wave inversions (≥0.3 mV).

CARDIAC BIOMARKERS

Patients with UA/NSTEMI who have elevated biomarkers of necrosis, such as CK-MB and troponin (a much more specific and sensitive marker of myocardial necrosis), are at increased risk for death or recurrent MI. Elevated levels of these markers distinguish patients with NSTEMI from those with UA. There is a direct relationship between the degree of troponin elevation and mortality. However, in patients *without* a clear clinical history of myocardial ischemia, minor troponin elevations have been reported and can be caused by congestive heart failure (CHF), myocarditis, or pulmonary embolism, or they may be false-positive readings. Thus, in patients with an *unclear* history, small troponin elevations may not be diagnostic of an ACS.

Diagnostic Evaluation

(See also Chap. 12) Approximately six million persons per year in the United States present to hospital emergency departments (EDs) with a complaint of chest pain or other symptoms suggestive of ACS. A diagnosis of an ACS is established in 20 to 25% of such patients. The first step in evaluating patients with possible UA/NSTEMI is to determine the *likelihood* that CAD is the cause of the presenting symptoms. The American College of Cardiology/American Heart Association (ACC/AHA) Guidelines include, among the factors associated with a high likelihood of ACS, a prior history typical of stable angina, a history of established CAD by angiography, prior MI, CHF, new ECG changes, or elevated cardiac biomarkers.

DIAGNOSTIC PATHWAYS

Four major diagnostic tools are used in the diagnosis of UA/NSTEMI in the ED: clinical history, the ECG, cardiac markers, and stress testing (coronary imaging is an emerging option). The goals are to: (1) recognize or exclude MI (using cardiac markers), (2) evaluate for rest ischemia (using serial or continuous ECGs), and (3) evaluate for significant CAD (using provocative stress testing). Patients with a low likelihood of ischemia are usually managed with an ED-based critical pathway (which, in some institutions, is carried out in a "chest-pain unit" **Fig. 244-1)**. Evaluation of such patients includes clinical monitoring for recurrent ischemic discomfort, serial ECGs, and cardiac markers, typically obtained at baseline and at 4–6 h and 12 h after presentation. If new elevations in cardiac markers or ECG changes are noted, the patient should be admitted to the hospital. If the patient remains pain free and the markers are negative, the patient may proceed to stress testing. CT angiography is used with increasing frequency to exclude obstructive CAD (Chap. 229).

Figure 244-1

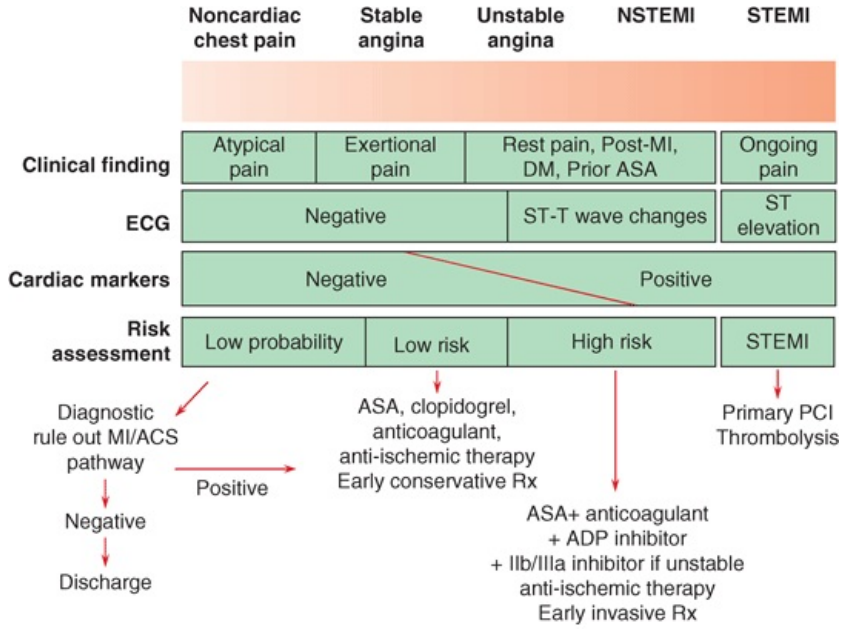

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for risk stratification and treatment of patients with suspected coronary artery disease. Using the clinical history of the type of pain and medical history, the ECG, and cardiac markers, one can identify patients who have a low likelihood of UA/NSTEMI, for whom a diagnostic "ruleout

myocardial infarction (MI) or acute coronary syndrome (ACS)" is warranted. If this is negative, the patient may be discharged, but if positive, the patient is admitted and treated for UA/NSTEMI. On the other end of the spectrum, patients with acute ongoing pain and ST-segment elevation are treated with percutaneous coronary intervention (PCI) or fibrinolysis (Chap. 245). For those with UA/NSTEMI, risk stratification is used to identify patients at medium to high risk, for whom an early invasive strategy is warranted. Antithrombotic therapy should include aspirin, an anticoagulant, an ADP antagonist (clopidogrel or prasugrel), with GP IIb/IIIa inhibition considered for use during PCI. For patients at low risk, treatment with aspirin, clopidogrel, an anticoagulant such as unfractionated or low molecular–weight heparin (LMWH) or fondaparinux and anti-ischemic therapy with beta blockers and nitrates, and a conservative strategy are indicated. ASA, aspirin; DM, diabetes mellitus; ECG, electrocardiogram; MI, myocardial infarction; Rx, treatment; STEMI, ST-segment elevation myocardial infarction. (*Adapted from CP Cannon, E Braunwald, in Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine, 9th ed, R Bonow et al (eds). Philadelphia, Saunders, 2011.*)

Risk Stratification and Prognosis

Patients with documented UA/NSTEMI exhibit a wide spectrum of early (30 days) risk of death, ranging from 1 to 10%, and of new or recurrent infarction of 3–5% or recurrent ACS (5-15%). Assessment of risk can be accomplished by clinical risk scoring systems such as that developed from the Thrombolysis in Myocardial Infarction (TIMI) Trials, which includes seven independent risk factors: age ≥ 65 years, three or more risk factors for CAD, documented CAD at catheterization, development of UA/NSTEMI while on aspirin, more than two episodes of angina within the preceding 24 h, ST deviation ≥0.5 mm, and an elevated cardiac marker **(Fig. 244-2)**. Other risk factors include diabetes mellitus, left ventricular dysfunction, renal dysfunction and elevated levels of brain natriuretic peptides and C-reactive protein. Multimarker strategies involving several biomarkers are now gaining favor, both to define more fully the pathophysiologic mechanisms underlying a given patient's presentation and to stratify the patient's risk further. Early risk assessment (especially using troponin, ST-segment changes, and/or a global risk-scoring system) is useful both in predicting the risk of recurrent cardiac events and in identifying those patients who would derive the greatest benefit from antithrombotic therapies more potent than unfractionated heparin, such as low molecular–weight heparin (LMWH) and glycoprotein IIb/IIIa inhibitors, and from an early invasive strategy. For example, in the TACTICS-TIMI 18 Trial, an early invasive strategy conferred a 40% reduction in recurrent cardiac events in patients with a positive troponin level, whereas no benefit was observed in those without detectable troponin.

Figure 244-2

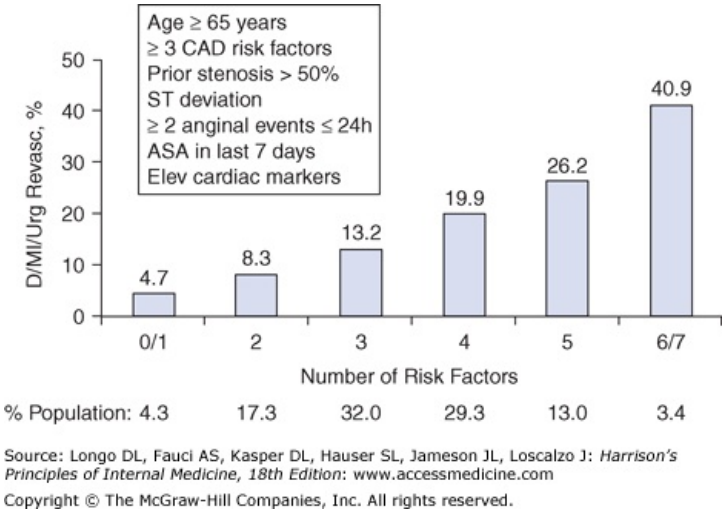

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The TIMI Risk Score for UA/NSTEMI, a simple but comprehensive clinical risk stratification score to identify increasing risk of death, myocardial infarction, or urgent revascularization to day 14. CAD, coronary artery disease; ASA, aspirin. *(Adapted from Antman et al.)*

Treatment: Unstable Angina and Non-ST-Segment Elevation Myocardial Infarction

MEDICAL TREATMENT

Patients with UA/NSTEMI should be placed at bed rest with continuous ECG monitoring for ST-segment deviation and cardiac arrhythmias. Ambulation is permitted if the patient shows no recurrence of ischemia (discomfort or ECG changes) and does not develop a biomarker of necrosis for 12–24 h. Medical therapy involves simultaneous anti-ischemic treatment and antithrombotic treatment.

ANTI-ISCHEMIC TREATMENT

(Table 244-1) To provide relief and prevention of recurrence of chest pain, initial treatment should include bed rest, nitrates, and beta blockers.

Table 244–1. Drugs Commonly Used in Intensive Medical Management of Patients with Unstable Angina and Non-ST Segment Elevation MI

Drug Category	Clinical Condition	When to Avoid[a]	Dosage
Nitrates	Administer sublingually, and, if symptoms persist, intravenously	Hypotension Patient receiving sildenafil or other PDE-5 inhibitor	Topical, oral, or buccal nitrates are acceptable alternatives for patients without ongoing or refractory symptoms 5–10 μg/min by continuous infusion titrated up to 75–100 μg/min until relief of symptoms or limiting side effects (headache or hypotension with a systolic blood pressure <90 mmHg or more than 30% below starting mean arterial pressure levels if significant hypertension is present)
Beta blockers[b]	Unstable angina	PR interval (ECG) >0.24 s 2° or 3° atrioventricular block Heart rate <60 beats/min Systolic pressure <90 mmHg Shock Left ventricular failure Severe reactive airway disease	Metoprolol 25–50 mg by mouth every 6 h If needed, and no heart failure, 5-mg increments by slow (over 1–2 min) IV administration
Calcium channel blockers	Patients whose symptoms are not relieved by adequate doses of nitrates and beta blockers, or in patients unable to tolerate adequate doses of one or both of these agents, or in patients with variant angina	Pulmonary edema Evidence of left ventricular dysfunction (for diltiazem or verapamil)	Dependent on specific agent
Morphine sulfate	Patients whose symptoms are not relieved after three serial sublingual nitroglycerin tablets or whose symptoms recur with adequate anti-ischemic therapy	Hypotension Respiratory depression Confusion Obtundation	2–5 mg IV dose May be repeated every 5–30 min as needed to relieve symptoms and maintain patient comfort

[a]Allergy or prior intolerance is a contraindication for all categories of drugs listed in this chart.

[b]Choice of the specific agent is not as important as ensuring that appropriate candidates receive this therapy. If there are concerns about patient intolerance owing to existing pulmonary disease, especially asthma, left ventricular dysfunction, risk of hypotension or severe bradycardia, initial selection should favor a short-acting agent, such as propranolol or metoprolol or the ultra-short-acting agent esmolol. Mild wheezing or a history of chronic obstructive pulmonary disease should prompt a trial of a short-acting agent at a reduced dose (e.g., 2.5 mg IV metoprolol, 12.5 mg oral metoprolol, or 25 μg/kg per min esmolol as initial doses) rather than complete avoidance of beta-blocker therapy.

Note: Some of the recommendations in this guide suggest the use of agents for purposes or in doses other than those specified by the U.S. Food and Drug Administration. Such recommendations are made after consideration of concerns regarding nonapproved indications. Where made, such recommendations are based on more recent clinical trials or expert consensus. IV, intravenous; ECG, electrocardiogram; 2°, second-degree; 3°, third-degree.

Source: Modified from E Braunwald et al: Circulation 90:613, 1994.

Nitrates

Nitrates should first be given sublingually or by buccal spray (0.3–0.6 mg) if the patient is experiencing ischemic pain. If pain persists after 3 doses given 5 min apart, intravenous nitroglycerin (5–10 μg/min using nonabsorbing tubing) is recommended. The rate of the infusion may be increased by 10 μg/min every 3–5 min until symptoms are relieved or systolic arterial pressure falls to <100 mmHg. Topical or oral nitrates (Chap. 243) can be used once the pain has resolved or they may replace intravenous nitroglycerin when the patient has been pain-free for 12–24 h. The only absolute contraindications to the use of nitrates are hypotension or the use of sildenafil or other drugs in that class within the previous 24–48 h.

Beta Adrenergic Blockers and Other Agents

Beta blockers are the other mainstay of anti-ischemic treatment. Oral beta blockade targeted to a heart rate of 50–60 beats/min is recommended as first-line treatment. A caution has been raised in the new ACC/AHA guidelines for use of intravenous beta blockade in patients with any evidence of acute heart failure, where they could increase the risk of cardiogenic shock. Heart rate–slowing calcium channel blockers, e.g., verapamil or diltiazem, are recommended for patients who have persistent or recurrent symptoms after treatment with full-dose nitrates and beta blockers and in patients with contraindications to beta blockade. Additional medical therapy includes angiotensin-converting enzyme (ACE) inhibition and HMG-CoA reductase inhibitors (statins) for long-term secondary prevention. Early administration of intensive statin therapy (e.g., atorvastatin 80 mg) prior to percutaneous coronary intervention (PCI) has been shown to reduce complications, suggesting that high-dose statin therapy should be started at the time of admission.

ANTITHROMBOTIC THERAPY

(Table 244-2) This is the other main component of treatment for UA/NSTEMI. Initial treatment should begin with the platelet cyclooxygenase inhibitor aspirin **(Fig. 244-3)**. The typical initial dose is 325 mg/d, with lower doses (75–162 mg/d) recommended for long-term therapy. The OASIS-7 trial randomized 25,087 ACS patients to receive high-dose (300–325 mg/d) vs. low-dose (75-100 mg/d) aspirin for 30 days and reported no differences in the risk of major bleeding or in efficacy over this period of time. "Aspirin resistance" has been noted in 5–10% of patients and more frequently in patients treated with lower doses of aspirin, but frequently has been related to noncompliance.

Table 244–2. Clinical Use of Antithrombotic Therapy

Oral Antiplatelet Therapy	
Aspirin	Initial dose of 162–325 mg nonenteric formulation followed by 75–162 mg/d of an enteric or a nonenteric formulation
Clopidogrel	Loading dose of 300-600 mg followed by 75 mg/d
Prasugrel	Pre-PCI: Loading dose 60 mg followed by 10 mg/d
Intravenous Antiplatelet Therapy	
Abciximab	0.25 mg/kg bolus followed by infusion of 0.125 μg/kg per min (maximum 10 μg/min) for 12 to 24 h
Eptifibatide	180 μg/kg bolus followed by infusion of 2.0 μg/kg per min for 72 to 96 h
Tirofiban	0.4 μg/kg per min for 30 min followed by infusion of 0.1 μg/kg per min for 48 to 96 h
Heparins*	
Unfractionated Heparin (UFH)	Bolus 60–70 U/kg (maximum 5000 U) IV followed by infusion of 12–15 U/kg per h (initial maximum 1000 U/h) titrated to a PTT 50–70 s
Enoxaparin	1 mg/kg SC every 12 h; the first dose may be preceded by a 30-mg IV bolus; renal adjustment to 1 mg/kg once daily if creatine Cl < 30 cc/min
Fondaparinux	2.5 mg SC qd
Bivalirudin	Initial bolus intravenous bolus of 0.1 mg/kg and an infusion of 0.25 mg/kg per hour. Before PCI, an additional intravenous bolus of 0.5 mg/kg was administered, and the infusion was increased to 1.75 mg/kg per hour.

*Other LMWH exist beyond those listed. ***Abbreviations:*** IV, intravenous; SC, subcutaneously.

Source: Modified from J Anderson et al: JACO 50:e1, 2007.

Figure 244-3

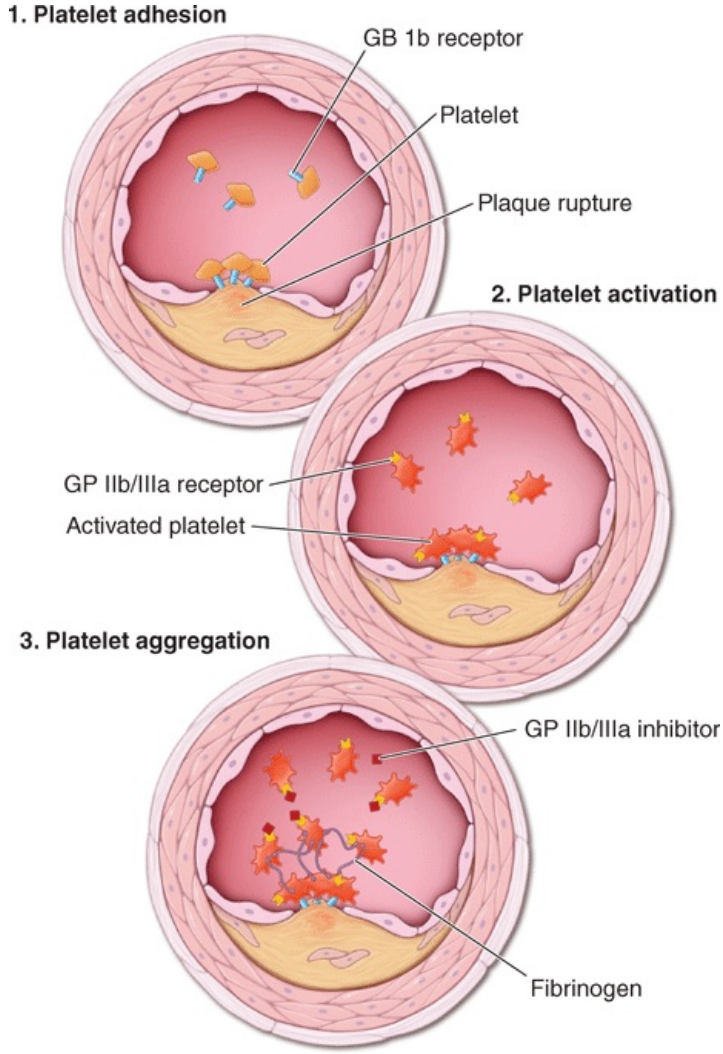

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Platelets initiate thrombosis at the site of a ruptured plaque with denuded endothelium: *platelet adhesion* occurs via (1) the GP 1b receptor in conjunction with von Willebrand factor. This is followed by *platelet activation* (2), which leads to a shape change of the platelet, degranulation of the alpha and dense granules, and expression of glycoprotein IIb/IIIa receptors on the platelet surface with activation of the receptor, such that it can bind fibrinogen. The final step is *platelet aggregation* (3), in which fibrinogen (or von Willebrand factor) binds to the activated GP IIb/IIIa receptors. Aspirin (ASA) and clopidogrel act to decrease platelet activation, whereas the GP IIb/IIIa inhibitors inhibit the final step of platelet aggregation. GP, glycoprotein. [*Modified from CP Cannon, E Braunwald, in Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine, 8th ed, R Bonow et al (eds). Philadelphia, Saunders, 2008.*]

The thienopyridine, clopidogrel, an inactive prodrug that is converted into an active metabolite, which blocks the platelet $P2Y_{12}$ component or the adenosine diphosphate receptor, in combination with aspirin, was shown in the CURE trial to confer a 20% relative reduction in cardiovascular death, MI, or stroke, compared with aspirin alone in both low- and high-risk patients, but to be associated with a moderate (absolute 1%) increase in major bleeding. Pretreatment with clopidogrel (a 300 or 600 mg loading dose, followed by 75 mg qd) is recommended prior to PCI. The OASIS-7 trial reported that one week of a higher dose of clopidogrel (600 mg loading dose and 150 mg/d for one week) did not result in an overall improvement in outcomes in ACS patients, but did so in patients receiving 325 mg of aspirin, especially those who underwent PCI.

Continued benefit of one year of treatment with the combination of clopidogrel and aspirin has been observed both in patients treated conservatively and in those who underwent PCI and should certainly continue for at least one year in patients with a drug-eluting stent. Up to one-third of patients have low response to clopidogrel, and a substantial proportion of these are related to a genetic variant of the cytochrome P450 system. A variant of the 2C19 gene leads to reduced conversion of clopidogrel into its active metabolite, which, in turn, causes lower platelet inhibition and a higher risk of cardiovascular events. Alternate agents, such as prasugrel, should be considered for ACS patients who are hyporesponsive to clopidogrel as identified by platelet and/or genetic testing, although such testing is

not yet widespread.

A recently approved thienopyridine, prasugrel, has been shown to achieve a more rapid onset, and higher level of platelet inhibition than clopidogrel. It has been used in ACS patients following angiography in whom PCI is planned at a dose of 60 mg load followed by 10 mg/d for up to 15 months. The TRITON-TIMI 38 trial showed that relative to clopidogrel, prasugrel reduced the risk of cardiovascular death, MI, or stroke significantly by 19%, albeit with an increase in major bleeding. Stent thrombosis was also reduced by 52%. This agent is contraindicated in patients with prior stroke or transient ischemic attack. Ticagrelor is a novel, *reversible* ADP inhibitor that has recently been reported to reduce the risk of cardiovascular death, MI, or stroke by 16% compared with clopidogrel in a broad population of ACS patients. This agent also reduced mortality and did not increase the risk of total bleeding; it is not yet FDA approved at the time of this writing.

Four options are available for anticoagulant therapy to be added to aspirin and clopidogrel. Unfractionated heparin (UFH) is the mainstay of therapy. The low-molecular-weight heparin (LMWH), enoxaparin, has been shown in several studies to be superior to UFH in reducing recurrent cardiac events, especially in conservatively managed patients. The indirect Factor Xa inhibitor, fondaparinux, is equivalent for early efficacy compared with enoxaparin but appears to have a lower risk of major bleeding. Bivalirudin, a direct thrombin inhibitor, is similar in efficacy to either UFH or LMWH among patients treated with a GP IIb/IIIa inhibitor, but use of bivalirudin alone causes less bleeding than the combination of heparin and a GP IIb/IIIa inhibitor in patients with UA/NSTEMI undergoing catheterization and/or PCI.

Prior to the advent of clopidogrel, many trials had shown the benefit of intravenous GP IIb/IIIa inhibitors. The benefit, however, has been small, i.e., only a 9% reduction in death or MI, with a significant increase in major bleeding. Two recent studies also failed to show a benefit for early initiation compared with use only for PCI. The use of these agents may be reserved for unstable patients with recurrent rest pain and ECG changes who undergo PCI.

Excessive bleeding is the most important adverse effect of all antithrombotic agents, including anticoagulants and antiplatelet agents. Therefore, attention must be directed to the doses of antithrombotic agents, accounting for weight, creatinine clearance, and a previous history of excessive bleeding, as a means of reducing the risk of bleeding.

INVASIVE VERSUS CONSERVATIVE STRATEGY

Multiple clinical trials have demonstrated the benefit of an early invasive strategy in high-risk patients, i.e., patients with multiple clinical risk factors, ST-segment deviation, and/or positive biomarkers **(Table 244-3)**. In this strategy, following treatment with anti-ischemic and antithrombotic agents, coronary arteriography is carried out within ~48 h of admission, followed by coronary revascularization (PCI or coronary artery bypass grafting), depending on the coronary anatomy.

Table 244–3. Class I Recommendations for Use of an Early Invasive Strategy*

Class I (Level of Evidence: A) Indications
Recurrent angina at rest/low-level activity despite Rx
Elevated TnT or TnI
New ST-segment depression
Rec. angina/ischemia with CHF symptoms, rales, MR
Positive stress test
EF < 0.40
Decreased BP
Sustained VT
PCI < 6 months, prior CABG
High-risk score

*Any one of the high-risk indicators.

Abbreviations: BP, blood pressure; CABG, coronary artery bypass grafting; CHF, congestive heart failure; EF, ejection fraction; MR, mitral regurgitation; PCI, percutaneous coronary intervention; Rec, recurrent; TnI, troponin I; TnT, troponin T; VT, ventricular tachycardia.

Source: J Anderson et al: JACO 50:e1, 2007.

In low-risk patients, the outcomes from an invasive strategy are similar to those obtained from a conservative strategy, which consists of anti-ischemic and antithrombotic therapy followed by "watchful waiting," and in which coronary arteriography is carried out only if rest pain or ST-segment changes recur or there is evidence of ischemia on a stress test.

Long-Term Management

The time of hospital discharge is a "teachable moment" for the patient with UA/NSTEMI, when the physician can review and optimize the medical regimen. Risk-factor modification is key, and the caregiver should discuss with the patient the importance of smoking cessation, achieving optimal weight, daily exercise following an appropriate diet, blood-pressure control, tight control of hyperglycemia (for diabetic patients), and lipid management, as recommended for patients with chronic stable angina (Chap. 243).

There is evidence of benefit with long-term therapy with five classes of drugs that are directed at different components of the atherothrombotic process. Beta blockers, statins (at a high dose, e.g., atorvastatin 80 mg/d), and ACE inhibitors or angiotensin receptor blockers are recommended for long-term plaque stabilization. Antiplatelet therapy, now recommended to be the combination of aspirin and clopidogrel (or prasugrel in post PCS patients) for one year, with aspirin continued thereafter, prevents or reduces the severity of any thrombosis that would occur if a plaque were to rupture.

Observational registries have shown that patients with UA/NSTEMI at high risk, including women and the elderly as well as racial minorities, are less likely to receive evidence-based pharmacologic and interventional therapies with resultant poorer clinical outcomes and quality of life.

Prinzmetal's Variant Angina

In 1959 Prinzmetal et al. described a syndrome of severe ischemic pain that occurs at rest but not usually with exertion and is associated with transient ST-segment elevation. This syndrome is due to focal spasm of an epicardial coronary artery, leading to severe myocardial ischemia. The cause of the spasm is not well defined, but it may be related to hypercontractility of vascular smooth muscle due to vasoconstrictor mitogens, leukotrienes, or serotonin.

CLINICAL AND ANGIOGRAPHIC MANIFESTATIONS

Patients with Prinzmetal's variant angina (PVA) are generally younger and have fewer coronary risk factors (with the exception of cigarette smoking) than patients with UA secondary to coronary atherosclerosis. Cardiac examination is usually unremarkable in the absence of ischemia. The clinical diagnosis of variant angina is made with the detection of transient ST-segment *elevation* with rest pain. Many patients also exhibit multiple episodes of asymptomatic ST-segment elevation (*silent ischemia*). Small elevations of troponin may occur in patients with prolonged attacks of variant angina.

Coronary angiography demonstrates transient coronary spasm as the diagnostic hallmark of PVA. Atherosclerotic plaques, which do not usually cause critical obstruction, in at least one proximal coronary artery occur in the majority of patients, and in them spasm usually occurs within 1 cm of the plaque. Focal spasm is most common in the right coronary artery, and it may occur at one or more sites in one artery or in multiple arteries simultaneously. Ergonovine, acetylcholine, other vasoconstrictor medications, and hyperventilation have been used to provoke focal coronary stenosis on angiography to establish the diagnosis. Hyperventilation has also been used to provoke rest angina, ST-segment elevation, and spasm on coronary arteriography.

TREATMENT: PRINZMETAL'S VARIANT ANGINA

Nitrates and calcium channel blockers are the main agents used to treat acute episodes and to abolish recurrent episodes of PVA. Aspirin may actually increase the severity of ischemic episodes, possibly as a result of the exquisite sensitivity of coronary tone to modest changes in the synthesis of prostacyclin. The response to beta blockers is variable. Coronary revascularization may be helpful in patients who also have discrete, proximal fixed obstructive lesions.

PROGNOSIS

Many patients with PVA pass through an acute, active phase, with frequent episodes of angina and cardiac events during the first 6 months after presentation. Long-term survival at 5 years is excellent (~ 90–95%). Patients with no or mild fixed coronary obstruction tend to experience a more benign course than do patients with associated severe obstructive lesions. Nonfatal MI occurs in up to 20% of patients by 5 years. Patients with PVA who develop serious arrhythmias during spontaneous episodes of pain are at a higher risk for sudden cardiac death. In most patients who survive an infarction or the initial 3- to 6-month period of frequent episodes, the condition stabilizes, and there is a tendency for symptoms and cardiac events to diminish over time.

FURTHER READINGS

Alexander KP et al: Excess dosing of antiplatelet and antithrombin agents in the treatment of non-ST-segment elevation acute coronary syndromes. JAMA 294:3108, 2005[PMID: 16380591]

Anderson JL et al: ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 116:e148, 2007

Antman EM et al: The TIMI risk score for unstable angina/non-ST elevation MI: A method for

prognostication and therapeutic decision-making. JAMA 284:835, 2000[PMID: 10938172]

Cannon CP, Braunwald E: Unstable angina, in *Braunwald's Heart Disease*, 9th ed, R Bonow et al (eds). Philadelphia, Saunders, 2011

Cannon CP et al: Intensive versus moderate lipid lowering with statins after acute coronary syndromes. N Engl J Med 350:1495, 2004[PMID: 15007110]

Giugliano RP et al: Early versus delayed, provisional eptifibatide in acute coronary syndromes. N Engl J Med 360:2176, 2009[PMID: 19332455]

O'Donoghue M et al: Early invasive vs. conservative treatment strategies in women and men with unstable angina and non-ST-segment elevation myocardial infarction: a meta-analysis. JAMA 300:71, 2008[PMID: 18594042]

Wallentin LT et al: Ticagrelor versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 361:1045, 2009[PMID: 19717846]

Wiviott SD et al: Prasugrel versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 357:2001, 2007[PMID: 17982182]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ST-SEGMENT ELEVATION MYOCARDIAL INFARCTION: INTRODUCTION

Acute myocardial infarction (AMI) is one of the most common diagnoses in hospitalized patients in industrialized countries. In the United States, approximately 650,000 patients experience a new AMI and 450,000 experience a recurrent AMI each year. The early (30-day) mortality rate from AMI is ~30%, with more than half of these deaths occurring before the stricken individual reaches the hospital. Although the mortality rate after admission for AMI has declined by ~30% over the past two decades, approximately 1 of every 25 patients who survives the initial hospitalization dies in the first year after AMI. Mortality is approximately fourfold higher in elderly patients (over age 75) as compared with younger patients.

When patients with prolonged ischemic discomfort at rest are first seen, the working clinical diagnosis is that they are suffering from an acute coronary syndrome **(Fig. 245-1)**. The 12-lead electrocardiogram (ECG) is a pivotal diagnostic and triage tool because it is at the center of the decision pathway for management; it permits distinction of those patients presenting with ST-segment elevation from those presenting without ST-segment elevation. Serum cardiac biomarkers are obtained to distinguish unstable angina (UA) from non-ST-segment MI (NSTEMI) and to assess the magnitude of an ST-segment elevation MI (STEMI). This chapter focuses on the evaluation and management of patients with STEMI, while Chapter 244 discusses UA/NSTEMI.

Figure 245-1

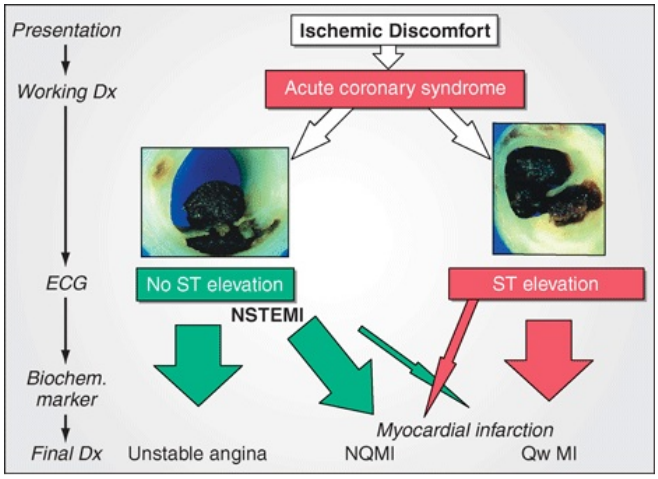

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute coronary syndromes. Following disruption of a vulnerable plaque, patients experience ischemic discomfort resulting from a reduction of flow through the affected epicardial coronary artery. The flow reduction may be caused by a completely occlusive thrombus (***right***) or subtotally occlusive thrombus (***left***). Patients with ischemic discomfort may present with or without ST-segment elevation. Of patients with ST-segment elevation, the majority (*wide red arrow*) ultimately develop a Q wave on the ECG (QwMI), while a minority (*thin red arrow*) do not develop Q wave and, in older literature, were said to have sustained a non-Q-wave MI (NQMI). Patients who present without ST-segment elevation are suffering from either unstable angina or a non-ST-segment elevation MI (NSTEMI) (*wide green arrows*), a distinction that is ultimately made on the presence or absence of a serum cardiac marker such as CKMB or a cardiac troponin detected in the blood. The majority of patients presenting with NSTEMI do not develop a Q wave on the ECG; a minority develop a QwMI (*thin green arrow*). (*Adapted from CW Hamm et al: Lancet 358:1533, 2001, and MJ Davies: Heart 83:361, 2000; with permission from the BMJ Publishing Group.*)

PATHOPHYSIOLOGY: ROLE OF ACUTE PLAQUE RUPTURE

STEMI usually occurs when coronary blood flow decreases abruptly after a thrombotic occlusion of a coronary artery previously affected by atherosclerosis. Slowly developing, high-grade coronary artery stenoses do not typically precipitate STEMI because of the development of a rich collateral network over time. Instead, STEMI occurs when a coronary artery thrombus develops rapidly at a site of vascular injury. This injury is produced or facilitated by factors such as cigarette smoking, hypertension, and lipid accumulation. In most cases, STEMI occurs when the surface of an atherosclerotic plaque becomes disrupted (exposing its contents to the blood) and conditions (local or systemic) favor thrombogenesis. A mural thrombus forms at the site of plaque disruption, and the involved coronary artery becomes occluded. Histologic studies indicate that the coronary plaques prone to disruption are those with a rich lipid core and a thin fibrous cap (Chap. 241). After an initial platelet monolayer forms at the site of the disrupted plaque, various agonists (collagen, ADP, epinephrine, serotonin) promote platelet activation. After agonist stimulation of platelets, thromboxane A_2 (a potent local vasoconstrictor) is

released, further platelet activation occurs, and potential resistance to fibrinolysis develops.

In addition to the generation of thromboxane A_2, activation of platelets by agonists promotes a conformational change in the glycoprotein IIb/IIIa receptor (Chap. 115). Once converted to its functional state, this receptor develops a high affinity for soluble adhesive proteins (i.e., integrins) such as fibrinogen. Since fibrinogen is a multivalent molecule, it can bind to two different platelets simultaneously, resulting in platelet cross-linking and aggregation.

The coagulation cascade is activated on exposure of tissue factor in damaged endothelial cells at the site of the disrupted plaque. Factors VII and X are activated, ultimately leading to the conversion of prothrombin to thrombin, which then converts fibrinogen to fibrin (Chap. 116). Fluid-phase and clot-bound thrombin participate in an autoamplification reaction leading to further activation of the coagulation cascade. The culprit coronary artery eventually becomes occluded by a thrombus containing platelet aggregates and fibrin strands.

In rare cases, STEMI may be due to coronary artery occlusion caused by coronary emboli, congenital abnormalities, coronary spasm, and a wide variety of systemic—particularly inflammatory—diseases. The amount of myocardial damage caused by coronary occlusion depends on (1) the territory supplied by the affected vessel, (2) whether or not the vessel becomes totally occluded, (3) the duration of coronary occlusion, (4) the quantity of blood supplied by collateral vessels to the affected tissue, (5) the demand for oxygen of the myocardium whose blood supply has been suddenly limited, (6) endogenous factors that can produce early spontaneous lysis of the occlusive thrombus, and (7) the adequacy of myocardial perfusion in the infarct zone when flow is restored in the occluded epicardial coronary artery.

Patients at increased risk for developing STEMI include those with multiple coronary risk factors (Chap. 241) and those with unstable angina (Chap. 244). Less common underlying medical conditions predisposing patients to STEMI include hypercoagulability, collagen vascular disease, cocaine abuse, and intracardiac thrombi or masses that can produce coronary emboli.

There have been major advances in the management of STEMI with recognition that the "chain of survival" involves a highly integrated system starting with prehospital care and extending to early hospital management so as to provide expeditious implementation of a reperfusion strategy.

CLINICAL PRESENTATION

In up to one-half of cases, a precipitating factor appears to be present before STEMI, such as vigorous physical exercise, emotional stress, or a medical or surgical illness. Although STEMI may commence at any time of the day or night, circadian variations have been reported such that clusters are seen in the morning within a few hours of awakening.

Pain is the most common presenting complaint in patients with STEMI. The pain is deep and visceral; adjectives commonly used to describe it are *heavy*, *squeezing*, and *crushing*, although, occasionally, it is described as stabbing or burning (Chap. 12). It is similar in character to the discomfort of angina pectoris (Chap. 243) but commonly occurs at rest, is usually more severe, and lasts longer. Typically, the pain involves the central portion of the chest and/or the epigastrium, and, on occasion, it radiates to the arms. Less common sites of radiation include the abdomen, back, lower jaw, and neck. The frequent location of the pain beneath the xiphoid and epigastrium and the patients' denial that they may be suffering a heart attack are chiefly responsible for the common mistaken impression of indigestion. The pain of STEMI may radiate as high as the occipital area but not below the umbilicus. It is often accompanied by weakness, sweating, nausea, vomiting, anxiety, and a sense of impending doom. The pain may commence when the patient is at rest, but when it begins during a period of exertion, it does not usually subside with cessation of activity, in contrast to angina pectoris.

The pain of STEMI can simulate pain from acute pericarditis (Chap. 239), pulmonary embolism (Chap. 262), acute aortic dissection (Chap. 248), costochondritis, and gastrointestinal disorders. These conditions should therefore be considered in the differential diagnosis. Radiation of discomfort to the trapezius is not seen in patients with STEMI and may be a useful distinguishing feature that suggests pericarditis is the correct diagnosis. However, *pain is not uniformly present in patients with STEMI* . The proportion of painless STEMIs is greater in patients with diabetes mellitus, and it increases with age. In the elderly, STEMI may present as sudden-onset breathlessness, which may progress to pulmonary edema. Other less common presentations, with or without pain, include sudden loss of consciousness, a confusional state, a sensation of profound weakness, the appearance of an arrhythmia, evidence of peripheral embolism, or merely an unexplained drop in arterial pressure.

Physical Findings

Most patients are anxious and restless, attempting unsuccessfully to relieve the pain by moving about in bed, altering their position, and stretching. Pallor associated with perspiration and coolness of the extremities occurs commonly. The combination of substernal chest pain persisting for >30 min and diaphoresis strongly suggests STEMI. Although many patients have a normal pulse rate and blood pressure within the first hour of STEMI, about one-fourth of patients with anterior infarction have manifestations of sympathetic nervous system hyperactivity (tachycardia and/or hypertension), and up to one-half with inferior infarction show evidence of parasympathetic hyperactivity (bradycardia and/or hypotension).

The precordium is usually quiet, and the apical impulse may be difficult to palpate. In patients with anterior wall infarction, an abnormal systolic pulsation caused by dyskinetic bulging of infarcted myocardium may develop in the periapical area within the first days of the illness and then may resolve. Other physical signs of ventricular dysfunction include fourth and third heart sounds, decreased intensity of the first heart sound, and paradoxical splitting of the second heart sound (Chap. 227). A transient midsystolic or late systolic apical systolic murmur due to dysfunction of the mitral valve apparatus may be present. A pericardial friction rub is heard in many patients with transmural STEMI at some time in the course of the disease, if the patients are examined frequently. The carotid pulse is often decreased in volume, reflecting reduced stroke volume. Temperature elevations up to 38°C

may be observed during the first week after STEMI. The arterial pressure is variable; in most patients with transmural infarction, systolic pressure declines by approximately 10–15 mmHg from the preinfarction state.

LABORATORY FINDINGS

Myocardial infarction (MI) progresses through the following temporal stages: (1) acute (first few hours–7 days), (2) healing (7–28 days), and (3) healed (≥29 days). When evaluating the results of diagnostic tests for STEMI, the temporal phase of the infarction must be considered. The laboratory tests of value in confirming the diagnosis may be divided into four groups: (1) ECG, (2) serum cardiac biomarkers, (3) cardiac imaging, and (4) nonspecific indices of tissue necrosis and inflammation.

Electrocardiogram

The electrocardiographic manifestations of STEMI are described in Chap. 228. During the initial stage, total occlusion of an epicardial coronary artery produces ST-segment elevation. Most patients initially presenting with ST-segment elevation ultimately evolve Q waves on the ECG. However, Q waves in the leads overlying the infarct zone may vary in magnitude and even appear only transiently, depending on the reperfusion status of the ischemic myocardium and restoration of transmembrane potentials over time. A small proportion of patients initially presenting with ST-segment elevation will not develop Q waves when the obstructing thrombus is not totally occlusive, obstruction is transient, or if a rich collateral network is present. Among patients presenting with ischemic discomfort but *without* ST-segment elevation, if a serum cardiac biomarker of necrosis (see below) is detected, the diagnosis of NSTEMI is ultimately made (Fig. 245-1). A minority of patients who present initially without ST-segment elevation may develop a Q-wave MI. Previously, it was believed that transmural MI is present if the ECG demonstrates Q waves or loss of R waves, and nontransmural MI may be present if the ECG shows only transient ST-segment and T-wave changes. However, electrocardiographic-pathologic correlations are far from perfect and terms such as *Q-wave MI*, *non-Q-wave MI*, *transmural MI*, and *nontransmural MI*, have been replaced by STEMI and NSTEMI (Fig. 245-1). Contemporary studies using MRI suggest that the development of a Q wave on the ECG is more dependent on the volume of infarcted tissue rather than the transmurality of infarction.

Serum Cardiac Biomarkers

Certain proteins, called serum cardiac biomarkers, are released from necrotic heart muscle after STEMI. The rate of liberation of specific proteins differs depending on their intracellular location, their molecular weight, and the local blood and lymphatic flow. Cardiac biomarkers become detectable in the peripheral blood once the capacity of the cardiac lymphatics to clear the interstitium of the infarct zone is exceeded and spillover into the venous circulation occurs. The temporal pattern of protein release is of diagnostic importance, but contemporary urgent reperfusion strategies necessitate making a decision (based largely on a combination of clinical and ECG findings) before the results of blood tests have returned from the laboratory. Rapid whole-blood bedside assays for serum cardiac markers are now available and may facilitate management decisions, particularly in patients with nondiagnostic ECGs.

Cardiac-specific troponin T (cTnT) and *cardiac-specific troponin I* (cTnI) have amino-acid sequences different from those of the skeletal muscle forms of these proteins. These differences permitted the development of quantitative assays for cTnT and cTnI with highly specific monoclonal antibodies. Since cTnT and cTnI are not normally detectable in the blood of healthy individuals but may increase after STEMI to levels >20 times higher than the upper reference limit (the highest value seen in 99% of a reference population not suffering from MI), the measurement of cTnT or cTnI is of considerable diagnostic usefulness, and they are now the preferred biochemical markers for MI **(Fig. 245-2)**. The cardiac troponins are particularly valuable when there is clinical suspicion of either skeletal muscle injury or a small MI that may be below the detection limit for creatine phosphokinase (CK) and its MB isoenzyme (CKMB) measurements, and they are, therefore, of particular value in distinguishing UA from NSTEMI. Levels of cTnI and cTnT may remain elevated for 7–10 days after STEMI.

Figure 245-2

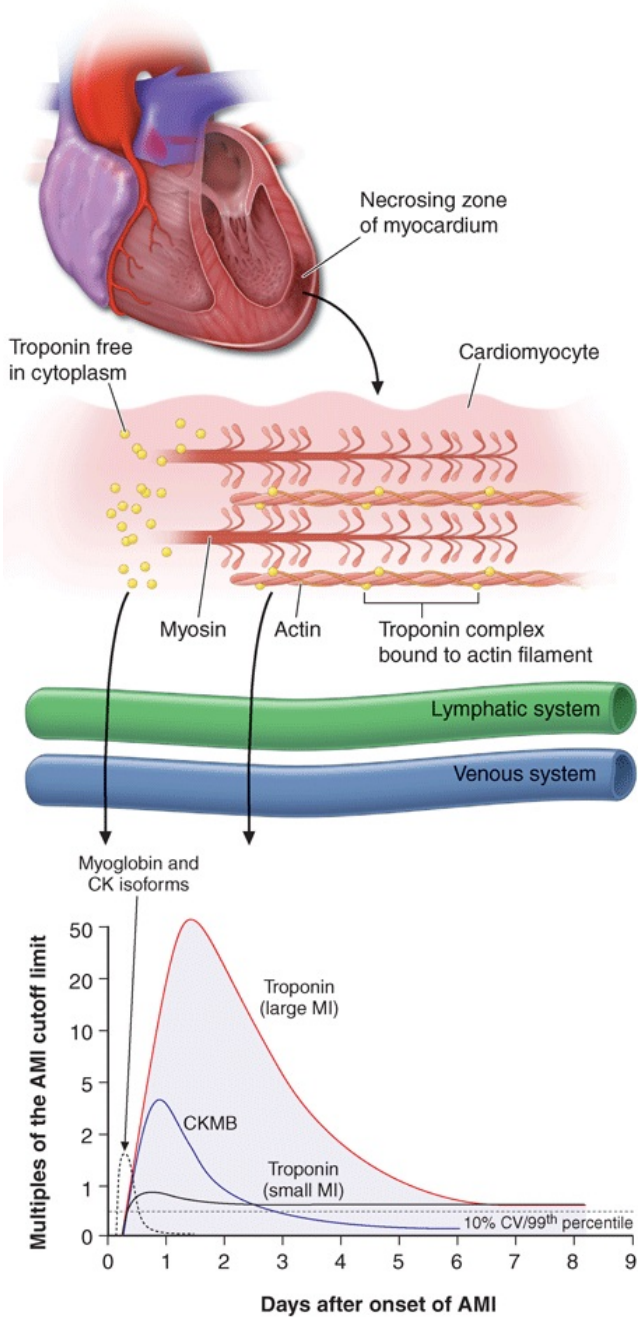

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The zone of necrosing myocardium is shown at the top of the figure, followed in the middle portion of the figure by a diagram of a cardiomyocyte that is in the process of releasing biomarkers. The biomarkers that are released into the interstitium are first cleared by lymphatics followed subsequently by spillover into the venous system. After disruption of the sarcolemmal membrane of the cardiomyocyte, the cytoplasmic pool of biomarkers is released first (left-most arrow in bottom portion of figure). Markers such as myoglobin and CK isoforms are rapidly released, and blood levels rise quickly above the cutoff limit; this is then followed by a more protracted release of biomarkers from the disintegrating myofilaments that may continue for several days. Cardiac troponin levels rise to about 20 to 50 times the upper reference limit (the 99th percentile of values in a reference control group) in patients who have a "classic" acute myocardial infarction (MI) and sustain sufficient myocardial necrosis to result in abnormally elevated levels of the MB fraction of creatine kinase (CKMB). Clinicians can now diagnose episodes of microinfarction by sensitive assays that detect cardiac troponin elevations above the upper reference limit, even though CKMB levels may still be in the normal reference range (not shown). CV = coefficient of variation. *(Modified from Antman EM: Decision making with cardiac troponin tests. N Engl J Med 346:2079, 2002 and Jaffe AS, Babiun L, Apple FS: Biomarkers in acute cardiac disease: The present and the future. J Am Coll Cardiol 48:1, 2006.)*

CK rises within 4–8 h and generally returns to normal by 48–72 h (Fig. 245-2). An important drawback of total CK measurement is its lack of specificity for STEMI, as CK may be elevated with skeletal muscle disease or trauma, including intramuscular injection. The MB isoenzyme of CK has the advantage over total CK that it is not present in significant concentrations in extracardiac tissue and, therefore, is considerably more specific. However, cardiac surgery, myocarditis, and electrical cardioversion often result in elevated serum levels of the MB isoenzyme. A ratio (relative index) of CKMB mass: CK activity ≥2.5 suggests but is not diagnostic of a myocardial rather than a skeletal muscle source for the CKMB elevation.

Many hospitals are using cTnT or cTnI rather than CKMB as the routine serum cardiac marker for diagnosis of STEMI, although any of these analytes remain clinically acceptable. It is *not* cost-effective to measure both a cardiac-specific troponin and CKMB at all time points in every patient.

While it has long been recognized that the total quantity of protein released correlates with the size of the infarct, the peak protein concentration correlates only weakly with infarct size. Recanalization of a coronary artery occlusion (either spontaneously or by mechanical or pharmacologic means) in the early hours of STEMI causes earlier peaking of biomarker measurements (Fig. 245-2) because of a rapid washout from the interstitium of the infarct zone, quickly overwhelming lymphatic clearance of the proteins.

The *nonspecific reaction* to myocardial injury is associated with polymorphonuclear leukocytosis, which appears within a few hours after the onset of pain and persists for 3–7 days; the white blood cell count often reaches levels of 12,000–15,000/μL. The erythrocyte sedimentation rate rises more slowly than the white blood cell count, peaking during the first week and sometimes remaining elevated for one or two weeks.

Cardiac Imaging

Abnormalities of wall motion on *two-dimensional echocardiography* (Chap. 229) are almost universally present. Although acute STEMI cannot be distinguished from an old myocardial scar or from acute severe ischemia by echocardiography, the ease and safety of the procedure make its use appealing as a screening tool in the Emergency Department setting. When the ECG is not diagnostic of STEMI, early detection of the presence or absence of wall motion abnormalities by echocardiography can aid in management decisions, such as whether the patient should receive reperfusion therapy [e.g., fibrinolysis or a percutaneous coronary intervention (PCI)]. Echocardiographic estimation of left ventricular (LV) function is useful prognostically; detection of reduced function serves as an indication for therapy with an inhibitor of the renin-angiotensin–aldosterone system. Echocardiography may also identify the presence of right ventricular (RV) infarction, ventricular aneurysm, pericardial effusion, and LV thrombus. In addition, Doppler echocardiography is useful in the detection and quantitation of a ventricular septal defect and mitral regurgitation, two serious complications of STEMI.

Several *radionuclide imaging techniques* (Chap. 229) are available for evaluating patients with suspected STEMI. However, these imaging modalities are used less often than echocardiography because they are more cumbersome and lack sensitivity and specificity in many clinical circumstances. Myocardial perfusion imaging with [^{201}Tl] or [^{99m}Tc]-sestamibi, which are distributed in proportion to myocardial blood flow and concentrated by viable myocardium (Chap. 243), reveal a defect ("cold spot") in most patients during the first few hours after development of a transmural infarct. Although perfusion scanning is extremely sensitive, it cannot distinguish acute infarcts from chronic scars and, thus, is not specific for the diagnosis of *acute* MI. Radionuclide ventriculography, carried out with [^{99m}Tc]-labeled red blood cells, frequently demonstrates wall motion disorders and reduction in the ventricular ejection fraction in patients with STEMI. While of value in assessing the hemodynamic consequences of infarction and in aiding in the diagnosis of RV infarction when the RV ejection fraction is depressed, this technique is nonspecific, as many cardiac abnormalities other than MI alter the radionuclide ventriculogram.

Myocardial infarction can be detected accurately with high-resolution cardiac MRI (Chap. 229) using a technique referred to as late enhancement. A standard imaging agent (gadolinium) is administered and images are obtained after a 10-min delay. Since little gadolinium enters normal myocardium, where there are tightly packed myocytes, but does percolate into the expanded intercellular region of the infarct zone, there is a bright signal in areas of infarction that appears in stark contrast to the dark areas of normal myocardium.

INITIAL MANAGEMENT

Prehospital Care

The prognosis in STEMI is largely related to the occurrence of two general classes of complications: (1) electrical complications (arrhythmias) and (2) mechanical complications ("pump failure"). Most out-of-hospital deaths from STEMI are due to the sudden development of ventricular fibrillation. The vast majority of deaths due to ventricular fibrillation occur within the first 24 h of the onset of symptoms, and of these, over half occur in the first hour. Therefore, the major elements of prehospital care of patients with suspected STEMI include (1) recognition of symptoms by the patient and prompt seeking of medical attention; (2) rapid deployment of an emergency medical team capable of performing resuscitative maneuvers, including defibrillation; (3) expeditious transportation of the patient to a hospital facility that is continuously staffed by physicians and nurses skilled in managing arrhythmias and providing advanced cardiac life support; and (4) expeditious implementation of reperfusion therapy **(Fig. 245-3)**. The greatest delay usually occurs not during transportation to the hospital but, rather, between the onset of pain and the patient's decision to call for help. This delay can best be reduced by health care professionals educating the public concerning the significance of chest discomfort and the importance of seeking early medical attention. Regular office visits with patients having a history of or who are at risk for ischemic heart disease are important "teachable moments" for clinicians to review the symptoms of STEMI and the appropriate action plan.

Figure 245-3

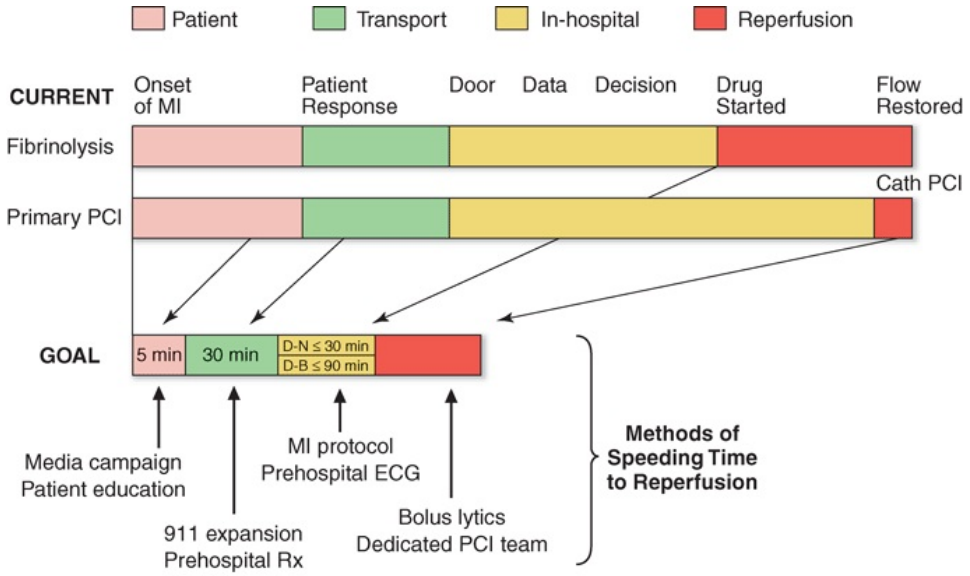

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Major components of time delay between onset of symptoms from STEMI and restoration of flow in the infarct-related artery. Plotted sequentially from left to right are the times for patients to recognize symptoms and seek medical attention, transportation to the hospital, in-hospital decision making, implementation of reperfusion strategy, and restoration of flow once the reperfusion strategy has been initiated. The time to initiate fibrinolytic therapy is the "door-to-needle" (D-N) time; this is followed by the period of time required for pharmacologic restoration of flow. More time is required to move the patient to the catheterization laboratory for a percutaneous coronary interventional (PCI) procedure, referred to as the "door-to-balloon" (D-B) time, but restoration of flow in the epicardial infarct–related artery occurs promptly after PCI. At the bottom is a variety of methods for speeding the time to reperfusion along with the goals for the time intervals for the various components of the time delay. (*Adapted from CP Cannon et al: J Thromb Thrombol 1:27, 1994* .)

Increasingly, monitoring and treatment are carried out by trained personnel in the ambulance, further shortening the time between the onset of the infarction and appropriate treatment. General guidelines for initiation of fibrinolysis in the prehospital setting include the ability to transmit 12-lead ECGs to confirm the diagnosis, the presence of paramedics in the ambulance, training of paramedics in the interpretation of ECGs and management of STEMI, and online medical command and control that can authorize the initiation of treatment in the field.

MANAGEMENT IN THE EMERGENCY DEPARTMENT

In the Emergency Department, the goals for the management of patients with suspected STEMI include control of cardiac discomfort, rapid identification of patients who are candidates for urgent reperfusion therapy, triage of lower-risk patients to the appropriate location in the hospital, and avoidance of inappropriate discharge of patients with STEMI. Many aspects of the treatment of STEMI are initiated in the Emergency Department and then continued during the in-hospital phase of management.

Aspirin is essential in the management of patients with suspected STEMI and is effective across the entire spectrum of acute coronary syndromes (Fig. 245-1). Rapid inhibition of cyclooxygenase-1 in platelets followed by a reduction of thromboxane A_2 levels is achieved by buccal absorption of a chewed 160–325-mg tablet in the Emergency Department. This measure should be followed by daily oral administration of aspirin in a dose of 75–162 mg.

In patients whose arterial O_2 saturation is normal, supplemental O_2 is of limited if any clinical benefit and therefore is not cost-effective. However, when hypoxemia is present, O_2 should be administered by nasal prongs or face mask (2–4 L/min) for the first 6–12 h after infarction; the patient should then be reassessed to determine if there is a continued need for such treatment.

Control of Discomfort

Sublingual *nitroglycerin* can be given safely to most patients with STEMI. Up to three doses of 0.4 mg should be administered at about 5-min intervals. In addition to diminishing or abolishing chest discomfort, nitroglycerin may be capable of both decreasing myocardial oxygen demand (by lowering preload) and increasing myocardial oxygen supply (by dilating infarct-related coronary vessels or collateral vessels). In patients whose initially favorable response to sublingual nitroglycerin is followed by the return of chest discomfort, particularly if accompanied by other evidence of ongoing ischemia such as further ST-segment or T-wave shifts, the use of intravenous nitroglycerin should be considered. Therapy with nitrates should be avoided in patients who present with low systolic arterial pressure (<90 mmHg) or in whom there is clinical suspicion of right ventricular infarction (inferior infarction on ECG, elevated jugular venous pressure, clear lungs, and hypotension). Nitrates should not be administered to patients who have taken the phosphodiesterase-5 inhibitor sildenafil for erectile dysfunction within the preceding 24 h, because it may potentiate the hypotensive effects of nitrates. An idiosyncratic reaction to nitrates, consisting of sudden marked hypotension, sometimes occurs but can usually be reversed promptly by the rapid administration of intravenous atropine.

Morphine is a very effective analgesic for the pain associated with STEMI. However, it may reduce sympathetically mediated arteriolar and venous constriction, and the resulting venous pooling may reduce cardiac

output and arterial pressure. These hemodynamic disturbances usually respond promptly to elevation of the legs, but in some patients volume expansion with intravenous saline is required. The patient may experience diaphoresis and nausea, but these events usually pass and are replaced by a feeling of well-being associated with the relief of pain. Morphine also has a vagotonic effect and may cause bradycardia or advanced degrees of heart block, particularly in patients with inferior infarction. These side effects usually respond to atropine (0.5 mg intravenously). Morphine is routinely administered by repetitive (every 5 min) intravenous injection of small doses (2–4 mg), rather than by the subcutaneous administration of a larger quantity, because absorption may be unpredictable by the latter route.

Intravenous *beta blockers* are also useful in the control of the pain of STEMI. These drugs control pain effectively in some patients, presumably by diminishing myocardial O_2 demand and hence ischemia. More important, there is evidence that intravenous beta blockers reduce the risks of reinfarction and ventricular fibrillation (see "Beta-Adrenoceptor Blockers" below). However, patient selection is important when considering beta blockers for STEMI. Oral beta-blocker therapy should be initiated in the first 24 h for patients who do not have any of the following: 1) signs of heart failure, 2) evidence of a low-output state, 3) increased risk for cardiogenic shock, or 4) other relative contraindications to beta blockade (PR interval greater than 0.24 seconds, second- or third-degree heart block, active asthma, or reactive airway disease). A commonly employed regimen is metoprolol, 5 mg every 2–5 min for a total of 3 doses, provided the patient has a heart rate >60 beats per minute (bpm), systolic pressure >100 mmHg, a PR interval <0.24 s, and rales that are no higher than 10 cm up from the diaphragm. Fifteen minutes after the last intravenous dose, an oral regimen is initiated of 50 mg every 6 h for 48 h, followed by 100 mg every 12 h.

Unlike beta blockers, calcium antagonists are of little value in the acute setting, and there is evidence that short-acting dihydropyridines may be associated with an increased mortality risk.

Management Strategies

The primary tool for screening patients and making triage decisions is the initial 12-lead ECG. When ST-segment elevation of at least 2 mm in 2 contiguous precordial leads and 1 mm in 2 adjacent limb leads is present, a patient should be considered a candidate for *reperfusion therapy* **(Fig. 245-4)**. The process of selecting patients for fibrinolysis versus primary PCI (angioplasty, or stenting; Chap. 246) is discussed below. In the absence of ST-segment elevation, fibrinolysis is not helpful, and evidence exists suggesting that it may be harmful.

Figure 245-4

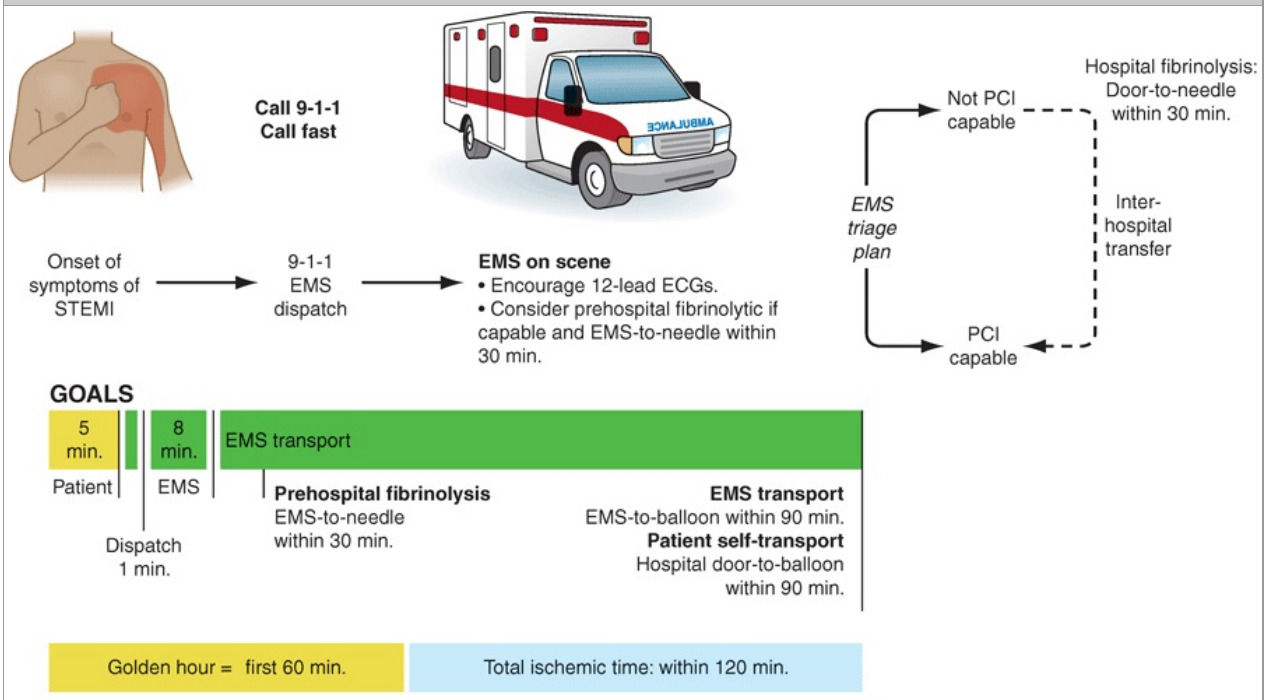

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Options for transportation of patients with STEMI and initial reperfusion treatment. Patient transported by EMS after calling 911: Reperfusion in patients with STEMI can be accomplished by the pharmacologic (fibrinolysis) or catheter-based (primary PCI) approaches. Implementation of these strategies varies based on the mode of transportation of the patient and capabilities at the receiving hospital. Transport time to the hospital is variable from case to case, but the goal is to keep total ischemic time within 120 min. There are three possible scenarios: (1) If EMS has fibrinolytic capability and the patient qualifies for therapy, prehospital fibrinolysis should be started within 30 min of EMS arrival on scene. (2) If EMS is not capable of administering prehospital fibrinolysis and the patient is transported to a non-PCI-capable hospital, the hospital door-to-needle time should be within 30 min for patients in whom fibrinolysis is indicated. (3) If EMS is not capable of administering prehospital fibrinolysis and the patient is transported to a PCI-capable hospital, the hospital door-to-balloon time should be within 90 min. *Interhospital transfer*: It is also appropriate to consider emergency interhospital transfer of the patient to a PCI-capable hospital for mechanical revascularization if: (1) there is a contraindication to fibrinolysis, (2) PCI can be initiated promptly (within 90 minutes after the patient presented to the initial receiving hospital or within 60 min compared to when fibrinolysis with a fibrin-specific agent could be initiated at the initial receiving

within 30 min compared to when fibrinolysis with a fibrin-specific agent could be initiated at the initial receiving hospital), (3) fibrinolysis is administered and is unsuccessful (i.e., "rescue PCI"). Secondary nonemergency interhospital transfer can be considered for recurrent ischemia. *Patient self-transport*: Patient self-transportation is discouraged. If the patient arrives at a non-PCI-capable hospital, the door-to-needle time should be within 30 min. If the patient arrives at a PCI-capable hospital, the door-to-balloon time should be within 90 min. The treatment options and time recommended after first hospital arrival are the same. *[Adapted with permission from Antman et al: ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1999 Guidelines for the Management of Patients with Acute Myocardial Infarction).* Circulation 110:e82, 2004.]

Limitation of Infarct Size

The quantity of myocardium that becomes necrotic as a consequence of a coronary artery occlusion is determined by factors other than just the site of occlusion. While the central zone of the infarct contains necrotic tissue that is irretrievably lost, the fate of the surrounding ischemic myocardium (ischemic penumbra) may be improved by timely restoration of coronary perfusion, reduction of myocardial O_2 demands, prevention of the accumulation of noxious metabolites, and blunting of the impact of mediators of reperfusion injury (e.g., calcium overload and oxygen-derived free radicals). Up to one-third of patients with STEMI may achieve *spontaneous* reperfusion of the infarct-related coronary artery within 24 h and experience improved healing of infarcted tissue. Reperfusion, either pharmacologically (by fibrinolysis) or by PCI, accelerates the opening of infarct-related arteries in those patients in whom spontaneous fibrinolysis ultimately would have occurred and also greatly increases the number of patients in whom restoration of flow in the infarct-related artery is accomplished. Timely restoration of flow in the epicardial infarct–related artery combined with improved perfusion of the downstream zone of infarcted myocardium results in a limitation of infarct size. Protection of the ischemic myocardium by the maintenance of an optimal balance between myocardial O_2 supply and demand through pain control, treatment of congestive heart failure (CHF), and minimization of tachycardia and hypertension extends the "window" of time for the salvage of myocardium by reperfusion strategies.

Glucocorticoids and nonsteroidal anti-inflammatory agents, with the exception of aspirin, should be avoided in patients with STEMI. They can impair infarct healing and increase the risk of myocardial rupture, and their use may result in a larger infarct scar. In addition, they can increase coronary vascular resistance, thereby potentially reducing flow to ischemic myocardium.

PRIMARY PERCUTANEOUS CORONARY INTERVENTION

(See also Chap. 246) PCI, usually angioplasty and/or stenting without preceding fibrinolysis, referred to as *primary PCI*, is effective in restoring perfusion in STEMI when carried out on an emergency basis in the first few hours of MI. It has the advantage of being applicable to patients who have contraindications to fibrinolytic therapy (see below) but otherwise are considered appropriate candidates for reperfusion. It appears to be more effective than fibrinolysis in opening occluded coronary arteries and, *when performed by experienced operators [≥75 PCI cases (not necessarily primary) per year] in dedicated medical centers (≥36 primary PCI cases per year)* , is associated with better short-term and long-term clinical outcomes. Compared with fibrinolysis, primary PCI is generally preferred when the diagnosis is in doubt, cardiogenic shock is present, bleeding risk is increased, or symptoms have been present for at least 2–3 h when the clot is more mature and less easily lysed by fibrinolytic drugs. However, PCI is expensive in terms of personnel and facilities, and its applicability is limited by its availability, around the clock, in only a minority of hospitals.

FIBRINOLYSIS

If no contraindications are present (see below), fibrinolytic therapy should ideally be initiated within 30 min of presentation (i.e., door-to-needle time ≤30 min). The principal goal of fibrinolysis is prompt restoration of full coronary arterial patency. The fibrinolytic agents tissue plasminogen activator (tPA), streptokinase, tenecteplase (TNK), and reteplase (rPA) have been approved by the U.S. Food and Drug Administration for intravenous use in patients with STEMI. These drugs all act by promoting the conversion of plasminogen to plasmin, which subsequently lyses fibrinthrombi. Although considerable emphasis was first placed on a distinction between more fibrin-specific agents, such as tPA, and non-fibrin-specific agents, such as streptokinase, it is now recognized that these differences are only relative, as some degree of systemic fibrinolysis occurs with the former agents. TNK and rPA are referred to as *bolus fibrinolytics* since their administration does not require a prolonged intravenous infusion.

When assessed angiographically, flow in the culprit coronary artery is described by a simple qualitative scale called the *thrombolysis in myocardial infarction (TIMI) grading system*: grade 0 indicates complete occlusion of the infarct-related artery; grade 1 indicates some penetration of the contrast material beyond the point of obstruction but without perfusion of the distal coronary bed; grade 2 indicates perfusion of the entire infarct vessel into the distal bed, but with flow that is delayed compared with that of a normal artery; and grade 3 indicates full perfusion of the infarct vessel with normal flow. The latter is the goal of reperfusion therapy, because full perfusion of the infarct-related coronary artery yields far better results in terms of limiting infarct size, maintenance of LV function, and reduction of both short- and long-term mortality rates. Additional methods of angiographic assessment of the efficacy of fibrinolysis include counting the number of frames on the cine film required for dye to flow from the origin of the infarct-related artery to a landmark in the distal vascular bed (*TIMI frame count*) and determining the rate of entry and exit of contrast dye from the microvasculature in the myocardial infarct zone (*TIMI myocardial perfusion grade*). These methods have an even tighter correlation with outcomes after STEMI than the more commonly employed TIMI flow grade.

Fibrinolytic therapy can reduce the relative risk of in-hospital death by up to 50% when administered within the first hour of the onset of symptoms of STEMI, and much of this benefit is maintained for at least 10 years. When appropriately used, fibrinolytic therapy appears to reduce infarct size, limit LV dysfunction, and reduce the incidence of serious complications such as septal rupture, cardiogenic shock, and malignant ventricular arrhythmias. Since myocardium can be salvaged only before it has been irreversibly injured, the timing of

reperfusion therapy, by fibrinolysis or a catheter-based approach, is of extreme importance in achieving maximum benefit. While the upper time limit depends on specific factors in individual patients, it is clear that every minute counts and that patients treated within 1–3 h of the onset of symptoms generally benefit most. Although reduction of the mortality rate is more modest, the therapy remains of benefit for many patients seen 3–6 h after the onset of infarction, and some benefit appears to be possible up to 12 h, especially if chest discomfort is still present and ST segments remain elevated. Compared with PCI for STEMI (primary PCI), fibrinolysis is generally the preferred reperfusion strategy for patients presenting in the first hour of symptoms, if there are logistical concerns about transportation of the patient to a suitable PCI center (experienced operator and team with a track record for a "door-to-balloon" time of <2 h), or there is an anticipated delay of at least 1 h between the time that fibrinolysis could be started versus implementation of PCI. Although patients <75 years achieve a greater relative reduction in the mortality rate with fibrinolytic therapy than do older patients, the higher *absolute* mortality rate (15–25%) in the latter results in similar absolute reductions in the mortality rates for both age groups.

tPA and the other relatively fibrin-specific plasminogen activators, rPA and TNK, are more effective than streptokinase at restoring full perfusion—i.e., TIMI grade 3 coronary flow—and have a small edge in improving survival as well. The current recommended regimen of tPA consists of a 15-mg bolus followed by 50 mg intravenously over the first 30 min, followed by 35 mg over the next 60 min. Streptokinase is administered as 1.5 million units (MU) intravenously over 1 h. rPA is administered in a double-bolus regimen consisting of a 10-MU bolus given over 2–3 min, followed by a second 10-MU bolus 30 min later. TNK is given as a single weight-based intravenous bolus of 0.53 mg/kg over 10 s. In addition to the fibrinolytic agents discussed above, pharmacologic reperfusion typically involves adjunctive antiplatelet and antithrombotic drugs, as discussed subsequently.

Alternative pharmacologic regimens for reperfusion combine an intravenous glycoprotein IIb/IIIa inhibitor with a reduced dose of a fibrinolytic agent. Compared with fibrinolytic agents that involve a prolonged infusion (e.g., tPA), such combination reperfusion regimens facilitate the rate and extent of fibrinolysis by inhibiting platelet aggregation, weakening the clot structure, and allowing penetration of the fibrinolytic agent deeper into the clot. However, combination reperfusion regimens have similar efficacy as compared with bolus fibrinolytics and are associated with an increased risk of bleeding, especially in patients >75 years. Therefore, combination reperfusion regimens are not recommended for routine use. Glycoprotein IIb/IIIa inhibitors, given alone or in combination with a reduced dose of a fibrinolytic agent as part of a preparatory regimen before planned immediate PCI (facilitated PCI), have not been shown to reduce infarct size or improve outcomes and, furthermore, are associated with increased bleeding. Facilitated PCI is, therefore, also not a strategy that is recommended for routine use.

INTEGRATED REPERFUSION STRATEGY

Evidence has emerged that suggests PCI plays an increasingly important role in the management of STEMI. Prior approaches that segregated the pharmacologic and catheter-based approaches to reperfusion have now been replaced with an integrated approach to triage and transfer of STEMI patients to receive PCI (Fig. 245-5).

Figure 245-5

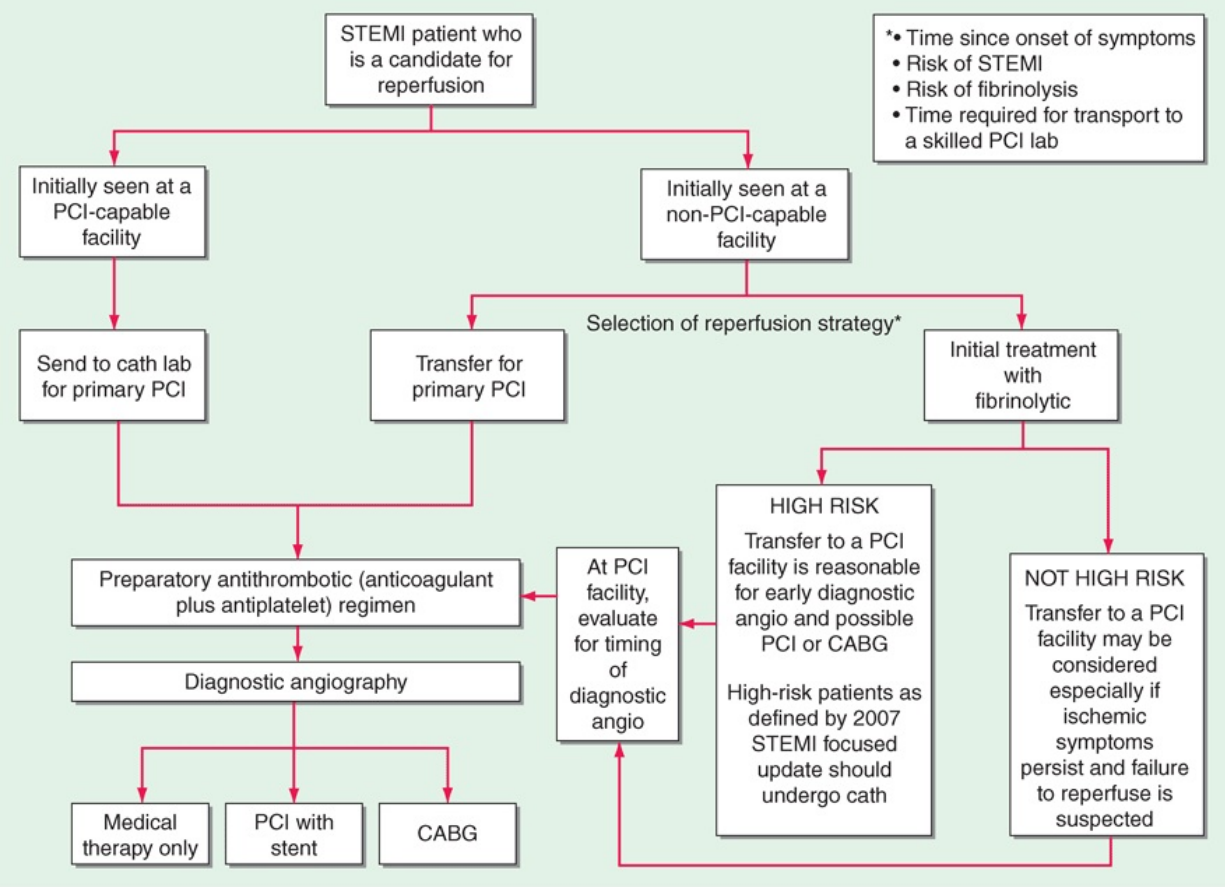

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Each community and each facility in that community should have an agreed-upon plan for how STEMI patients are to be treated that includes which hospitals should receive STEMI patients from EMS units capable of obtaining diagnostic ECGs, management at the initial receiving hospital, and written criteria and agreements for expeditious transfer of patients from non-PCI-capable facilities. Patients initially seen at a PCI-capable facility (left side of diagram) should be sent promptly to the cardiac catheterization laboratory with the intent to perform primary PCI. Patients initially seen at a non-PCI-capable facility (right side of diagram) should rapidly be assessed for the optimum reperfusion therapy (see box in top right corner for assessment criteria). This may include transfer for primary PCI or initial treatment with a fibrinolytic. Following administration of a fibrinolytic, management is dictated by the patient's overall risk for death/serious complications of STEMI, and whether or not they experience recurrent ischemic symptoms or left-ventricular failure (see the two boxes at the bottom right of diagram). *[Adapted from Kushner FG et al: 2009 focused update of the ACC/AHA Guidelines for the Management of Patients with ST-Elevation Myocardial Infarction (updating the 2004 guideline and 2007 focused update): a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 120:2271, 2009.]*

CONTRAINDICATIONS AND COMPLICATIONS

Clear contraindications to the use of fibrinolytic agents include a history of cerebrovascular hemorrhage at any time, a nonhemorrhagic stroke or other cerebrovascular event within the past year, marked hypertension (a reliably determined systolic arterial pressure >180 mmHg and/or a diastolic pressure >110 mmHg) at any time during the acute presentation, suspicion of aortic dissection, and active internal bleeding (excluding menses). While advanced age is associated with an increase in hemorrhagic complications, the benefit of fibrinolytic therapy in the elderly appears to justify its use if no other contraindications are present and the amount of myocardium in jeopardy appears to be substantial.

Relative contraindications to fibrinolytic therapy, which require assessment of the risk:benefit ratio, include current use of anticoagulants (international normalized ratio ≥2), a recent (<2 weeks) invasive or surgical procedure or prolonged (>10 min) cardiopulmonary resuscitation, known bleeding diathesis, pregnancy, a hemorrhagic ophthalmic condition (e.g., hemorrhagic diabetic retinopathy), active peptic ulcer disease, and a history of severe hypertension that is currently adequately controlled. Because of the risk of an allergic reaction, patients should not receive streptokinase if that agent had been received within the preceding five days to two years.

Allergic reactions to streptokinase occur in ~2% of patients who receive it. While a minor degree of hypotension occurs in 4–10% of patients given this agent, marked hypotension occurs, although rarely, in association with severe allergic reactions.

Hemorrhage is the most frequent and potentially the most serious complication. Because bleeding episodes that require transfusion are more common when patients require invasive procedures, unnecessary venous or arterial interventions should be avoided in patients receiving fibrinolytic agents. Hemorrhagic stroke is the most serious complication and occurs in ~0.5–0.9% of patients being treated with these agents. This rate increases with advancing age, with patients >70 years experiencing roughly twice the rate of intracranial hemorrhage as those

<65 years. Large-scale trials have suggested that the rate of intracranial hemorrhage with tPA or rPA is slightly higher than with streptokinase.

Cardiac catheterization and coronary angiography should be carried out after fibrinolytic therapy if there is evidence of either (1) failure of reperfusion (persistent chest pain and ST-segment elevation >90 min), in which case a *rescue PCI* should be considered; or (2) coronary artery reocclusion (re-elevation of ST segments and/or recurrent chest pain) or the development of recurrent ischemia (such as recurrent angina in the early hospital course or a positive exercise stress test before discharge), in which case an *urgent PCI* should be considered. The potential benefits of routine angiography and *elective* PCI even in asymptomatic patients following administration of fibrinolytic therapy are controversial, but such an approach may have merit given the numerous technological advances that have occurred in the catheterization laboratory and the increasing number of skilled interventionalists. Coronary artery bypass surgery should be reserved for patients whose coronary anatomy is unsuited to PCI but in whom revascularization appears to be advisable because of extensive jeopardized myocardium or recurrent ischemia.

HOSPITAL PHASE MANAGEMENT

Coronary Care Units

These units are routinely equipped with a system that permits continuous monitoring of the cardiac rhythm of each patient and hemodynamic monitoring in selected patients. Defibrillators, respirators, noninvasive transthoracic pacemakers, and facilities for introducing pacing catheters and flow-directed balloon-tipped catheters are also usually available. Equally important is the organization of a highly trained team of nurses who can recognize arrhythmias; adjust the dosage of antiarrhythmic, vasoactive, and anticoagulant drugs; and perform cardiac resuscitation, including electroshock, when necessary.

Patients should be admitted to a coronary care unit early in their illness when it is expected that they will derive benefit from the sophisticated and expensive care provided. The availability of electrocardiographic monitoring and trained personnel outside the coronary care unit has made it possible to admit lower-risk patients (e.g., those not hemodynamically compromised and without active arrhythmias) to "intermediate care units."

The duration of stay in the coronary care unit is dictated by the ongoing need for intensive care. If symptoms are controlled with oral therapy, patients may be transferred out of the coronary care unit. Also, patients who have a confirmed STEMI but who are considered to be at low risk [no prior infarction and no persistent chest discomfort, congestive heart failure (CHF), hypotension, or cardiac arrhythmias] may be safely transferred out of the coronary care unit within 24 h.

ACTIVITY

Factors that increase the work of the heart during the initial hours of infarction may increase the size of the infarct. Therefore, patients with STEMI should be kept at bed rest for the first 12 h. However, in the absence of complications, patients should be encouraged, under supervision, to resume an upright posture by dangling their feet over the side of the bed and sitting in a chair within the first 24 h. This practice is psychologically beneficial and usually results in a reduction in the pulmonary capillary wedge pressure. In the absence of hypotension and other complications, by the second or third day, patients typically are ambulating in their room with increasing duration and frequency, and they may shower or stand at the sink to bathe. By day 3 after infarction, patients should be increasing their ambulation progressively to a goal of 185 m (600 ft) at least 3 times a day.

DIET

Because of the risk of emesis and aspiration soon after STEMI, patients should receive either nothing or only clear liquids by mouth for the first 4–12 h. The typical coronary care unit diet should provide ≤30% of total calories as fat and have a cholesterol content of ≤300 mg/d. Complex carbohydrates should make up 50–55% of total calories. Portions should not be unusually large, and the menu should be enriched with foods that are high in potassium, magnesium, and fiber, but low in sodium. Diabetes mellitus and hypertriglyceridemia are managed by restriction of concentrated sweets in the diet.

BOWEL MANAGEMENT

Bed rest and the effect of the narcotics used for the relief of pain often lead to constipation. A bedside commode rather than a bedpan, a diet rich in bulk, and the routine use of a stool softener such as dioctyl sodium sulfosuccinate (200 mg/d) are recommended. If the patient remains constipated despite these measures, a laxative can be prescribed. Contrary to prior belief, it is safe to perform a gentle rectal examination on patients with STEMI.

SEDATION

Many patients require sedation during hospitalization to withstand the period of enforced inactivity with tranquillity. Diazepam (5 mg), oxazepam (15–30 mg), or lorazepam (0.5–2 mg), given 3–4 times daily, is usually effective. An additional dose of any of the above medications may be given at night to ensure adequate sleep. Attention to this problem is especially important during the first few days in the coronary care unit, where the atmosphere of 24-h vigilance may interfere with the patient's sleep. However, sedation is no substitute for reassuring, quiet surroundings. Many drugs used in the coronary care unit, such as atropine, H_2 blockers, and narcotics, can produce delirium, particularly in the elderly. This effect should not be confused with agitation, and it is wise to conduct a thorough review of the patient's medications before arbitrarily prescribing additional doses of anxiolytics.

PHARMACOTHERAPY

Antithrombotic Agents

The use of antiplatelet and anticoagulant therapy during the initial phase of STEMI is based on extensive

laboratory and clinical evidence that thrombosis plays an important role in the pathogenesis of this condition. The primary goal of treatment with antiplatelet and anticoagulant agents is to maintain patency of the infarct-related artery, in conjunction with reperfusion strategies. A secondary goal is to reduce the patient's tendency to thrombosis and, thus, the likelihood of mural thrombus formation or deep venous thrombosis, either of which could result in pulmonary embolization. The degree to which antiplatelet and anticoagulant therapy achieves these goals partly determines how effectively it reduces the risk of mortality from STEMI.

As noted previously (see "Management in the Emergency Department" above), aspirin is the standard antiplatelet agent for patients with STEMI. The most compelling evidence for the benefits of antiplatelet therapy (mainly with aspirin) in STEMI is found in the comprehensive overview by the Antiplatelet Trialists' Collaboration. Data from nearly 20,000 patients with MI enrolled in 15 randomized trials were pooled and revealed a relative reduction of 27% in the mortality rate, from 14.2% in control patients to 10.4% in patients receiving antiplatelet agents.

Inhibitors of the P2Y12 ADP receptor prevent activation and aggregation of platelets. The addition of the P2Y12 inhibitor clopidogrel to background treatment with aspirin to STEMI patients reduces the risk of clinical events (death, reinfarction, stroke) and, in patients receiving fibrinolytic therapy, has been shown to prevent reocclusion of a successfully reperfused infarct artery. New P2Y12 ADP receptor antagonists, such as prasugrel and ticagrelor, are more effective than clopidogrel in preventing ischemic complications in STEMI patients undergoing PCI, but are associated with an increased risk of bleeding. Glycoprotein IIb/IIIa receptor inhibitors appear useful for preventing thrombotic complications in patients with STEMI undergoing PCI.

The standard anticoagulant agent used in clinical practice is unfractionated heparin (UFH). The available data suggest that when UFH is added to a regimen of aspirin and a non-fibrin-specific thrombolytic agent such as streptokinase, additional mortality benefit occurs (about 5 lives saved per 1000 patients treated). It appears that the immediate administration of intravenous UFH, in addition to a regimen of aspirin and relatively fibrin-specific fibrinolytic agents (tPA, rPA, or TNK), helps to maintain patency of the infarct-related artery. This effect is achieved at the cost of a small increased risk of bleeding. The recommended dose of UFH is an initial bolus of 60 U/kg (maximum 4000 U) followed by an initial infusion of 12 U/kg per hour (maximum 1000 U/h). The activated partial thromboplastin time during maintenance therapy should be 1.5–2 times the control value.

Alternatives to UFH for anticoagulation of patients with STEMI are the low-molecular-weight heparin (LMWH) preparations, a synthetic version of the critical pentasaccharide sequence (fondaparinux), and the direct antithrombin bivalirudin. Advantages of LMWHs include high bioavailability permitting administration subcutaneously, reliable anticoagulation without monitoring, and greater antiXa:IIa activity. Enoxaparin has been shown to reduce significantly the composite endpoints of death/nonfatal reinfarction and death/nonfatal reinfarction/urgent revascularization compared with UFH in STEMI patients who receive fibrinolysis. Treatment with enoxaparin is associated with higher rates of serious bleeding, but net clinical benefit—a composite endpoint that combines efficacy and safety—still favors enoxaparin over UFH. Interpretation of the data on fondaparinux is difficult because of the complex nature of the pivotal clinical trial evaluating it in STEMI (OASIS-6). Fondaparinux appears superior to placebo in STEMI patients not receiving reperfusion therapy, but its relative efficacy and safety compared with UFH is less certain. Owing to the risk of catheter thrombosis, fondaparinux should not be used alone at the time of coronary angiography and PCI but should be combined with another anticoagulant with antithrombin activity such as UFH or bivalirudin. Contemporary trials of bivalirudin used an open-label design to evaluate its efficacy and safety compared with UFH plus a glycoprotein IIb/IIIa inhibitor. Bivalirudin was associated with a lower rate of bleeding, largely driven by reductions in vascular access site hematomas ≥5 cm or the administration of blood transfusions.

Patients with an anterior location of the infarction, severe LV dysfunction, heart failure, a history of embolism, two-dimensional echocardiographic evidence of mural thrombus, or atrial fibrillation are at increased risk of systemic or pulmonary thromboembolism. Such individuals should receive full therapeutic levels of anticoagulant therapy (LMWH or UFH) while hospitalized, followed by at least three months of warfarin therapy.

Beta-Adrenoceptor Blockers

The benefits of beta blockers in patients with STEMI can be divided into those that occur immediately when the drug is given acutely and those that accrue over the long term when the drug is given for secondary prevention after an infarction. Acute intravenous beta blockade improves the myocardial O_2 supply-demand relationship, decreases pain, reduces infarct size, and decreases the incidence of serious ventricular arrhythmias. In patients who undergo fibrinolysis soon after the onset of chest pain, no incremental reduction in mortality rate is seen with beta blockers, but recurrent ischemia and reinfarction are reduced.

Thus, beta-blocker therapy after STEMI is useful for most patients [including those treated with an angiotensin-converting enzyme (ACE) inhibitor] except those in whom it is specifically contraindicated (patients with heart failure or severely compromised LV function, heart block, orthostatic hypotension, or a history of asthma) and perhaps those whose excellent long-term prognosis (defined as an expected mortality rate of <1% per year, patients <55 years, no previous MI, with normal ventricular function, no complex ventricular ectopy, and no angina) markedly diminishes any potential benefit.

Inhibition of the Renin-Angiotensin-Aldosterone System

ACE inhibitors reduce the mortality rate after STEMI, and the mortality benefits are additive to those achieved with aspirin and beta blockers. The maximum benefit is seen in high-risk patients (those who are elderly or who have an anterior infarction, a prior infarction, and/or globally depressed LV function), but evidence suggests that a short-term benefit occurs when ACE inhibitors are prescribed unselectively to all hemodynamically stable patients with STEMI (i.e., those with a systolic pressure >100 mmHg). The mechanism involves a reduction in ventricular remodeling after infarction (see "Ventricular Dysfunction" below) with a subsequent reduction in the risk of CHF. The rate of recurrent infarction may also be lower in patients treated chronically with ACE inhibitors after infarction.

Before hospital discharge, LV function should be assessed with an imaging study. ACE inhibitors should be

continued indefinitely in patients who have clinically evident CHF, in patients in whom an imaging study shows a reduction in global LV function or a large regional wall motion abnormality, or in those who are hypertensive.

Angiotensin receptor blockers (ARBs) should be administered to STEMI patients who are intolerant of ACE inhibitors and who have either clinical or radiological signs of heart failure. Long-term aldosterone blockade should be prescribed for STEMI patients without significant renal dysfunction (creatinine ≥2.5 mg/dL in men and ≥ 2.0 mg/dL in women) or hyperkalemia (potassium ≥5.0 mEq/L) who are already receiving therapeutic doses of an ACE inhibitor, an LV ejection fraction ≤40 percent, and either symptomatic heart failure or diabetes mellitus. A multidrug regimen for inhibiting the renin-angiotensin-aldosterone system has been shown to reduce both heart failure–related and sudden cardiac death–related cardiovascular mortality after STEMI, but has not been as thoroughly explored as ACE inhibitors in STEMI patients.

Other Agents

Favorable effects on the ischemic process and ventricular remodeling (see below) previously led many physicians to routinely use *intravenous nitroglycerin* (5–10 μg/min initial dose and up to 200 μg/min as long as hemodynamic stability is maintained) for the first 24–48 h after the onset of infarction. However, the benefits of routine use of intravenous nitroglycerin are less in the contemporary era where beta-adrenoceptor blockers and ACE inhibitors are routinely prescribed for patients with STEMI.

Results of multiple trials of different calcium antagonists have failed to establish a role for these agents in the treatment of most patients with STEMI. Therefore, the routine use of calcium antagonists cannot be recommended. Strict control of blood glucose in diabetic patients with STEMI has been shown to reduce the mortality rate. Serum magnesium should be measured in all patients on admission, and any demonstrated deficits should be corrected to minimize the risk of arrhythmias.

COMPLICATIONS AND THEIR MANAGEMENT

Ventricular Dysfunction

After STEMI, the left ventricle undergoes a series of changes in shape, size, and thickness in both the infarcted and noninfarcted segments. This process is referred to as *ventricular remodeling* and generally precedes the development of clinically evident CHF in the months to years after infarction. Soon after STEMI, the left ventricle begins to dilate. Acutely, this results from expansion of the infarct, i.e., slippage of muscle bundles, disruption of normal myocardial cells, and tissue loss within the necrotic zone, resulting in disproportionate thinning and elongation of the infarct zone. Later, lengthening of the noninfarcted segments occurs as well. The overall chamber enlargement that occurs is related to the size and location of the infarct, with greater dilation following infarction of the anterior wall and apex of the left ventricle and causing more marked hemodynamic impairment, more frequent heart failure, and a poorer prognosis. Progressive dilation and its clinical consequences may be ameliorated by therapy with ACE inhibitors and other vasodilators (e.g., nitrates). In patients with an ejection fraction <40%, regardless of whether or not heart failure is present, ACE inhibitors or ARBs should be prescribed (see "Inhibition of the Renin-Angiotensin-Aldosterone System" above).

Hemodynamic Assessment

Pump failure is now the primary cause of in-hospital death from STEMI. The extent of infarction correlates well with the degree of pump failure and with mortality, both early (within 10 days of infarction) and later. The most common clinical signs are pulmonary rales and S_3 and S4 gallop sounds. Pulmonary congestion is also frequently seen on the chest roentgenogram. Elevated LV filling pressure and elevated pulmonary artery pressure are the characteristic hemodynamic findings, but these findings may result from a reduction of ventricular compliance (diastolic failure) and/or a reduction of stroke volume with secondary cardiac dilation (systolic failure) (Chap. 234).

A classification originally proposed by Killip divides patients into four groups: class I, no signs of pulmonary or venous congestion; class II, moderate heart failure as evidenced by rales at the lung bases, S_3 gallop, tachypnea, or signs of failure of the right side of the heart, including venous and hepatic congestion; class III, severe heart failure, pulmonary edema; and class IV, shock with systolic pressure <90 mmHg and evidence of peripheral vasoconstriction, peripheral cyanosis, mental confusion, and oliguria. When this classification was established in 1967, the expected hospital mortality rate of patients in these classes was as follows: class I, 0–5%; class II, 10–20%; class III, 35–45%; and class IV, 85–95%. With advances in management, the mortality rate in each class has fallen, perhaps by as much as one-third to one-half.

Hemodynamic evidence of abnormal global LV function appears when contraction is seriously impaired in 20–25% of the left ventricle. Infarction of ≥40% of the left ventricle usually results in cardiogenic shock (Chap. 272). Positioning of a balloon flotation (Swan-Ganz) catheter in the pulmonary artery permits monitoring of LV filling pressure; this technique is useful in patients who exhibit hypotension and/or clinical evidence of CHF. Cardiac output can also be determined with a pulmonary artery catheter. With the addition of intra-arterial pressure monitoring, systemic vascular resistance can be calculated as a guide to adjusting vasopressor and vasodilator therapy. Some patients with STEMI have markedly elevated LV filling pressures (>22 mmHg) and normal cardiac indices [2.6–3.6 L/(min/m^2)], while others have relatively low LV filling pressures (<15 mmHg) and reduced cardiac indices. The former patients usually benefit from diuresis, while the latter may respond to volume expansion.

Hypovolemia

This is an easily corrected condition that may contribute to the hypotension and vascular collapse associated with STEMI in some patients. It may be secondary to previous diuretic use, to reduced fluid intake during the early stages of the illness, and/or to vomiting associated with pain or medications. Consequently, hypovolemia should be identified and corrected in patients with STEMI and hypotension before more vigorous forms of therapy are begun. Central venous pressure reflects RV rather than LV filling pressure and is an inadequate guide for

adjustment of blood volume, because LV function is almost always affected much more adversely than RV function in patients with STEMI. The optimal LV filling or pulmonary artery wedge pressure may vary considerably among patients. Each patient's ideal level (generally ~20 mmHg) is reached by cautious fluid administration during careful monitoring of oxygenation and cardiac output. Eventually, the cardiac output level plateaus, and further increases in LV filling pressure only increase congestive symptoms and decrease systemic oxygenation without raising arterial pressure.

Treatment: Congestive Heart Failure

The management of CHF in association with STEMI is similar to that of acute heart failure secondary to other forms of heart disease (avoidance of hypoxemia, diuresis, afterload reduction, inotropic support) (Chap. 234), except that the benefits of digitalis administration to patients with STEMI are unimpressive. By contrast, diuretic agents are extremely effective, as they diminish pulmonary congestion in the presence of systolic and/or diastolic heart failure. LV filling pressure falls and orthopnea and dyspnea improve after the intravenous administration of furosemide or other loop diuretics. These drugs should be used with caution, however, as they can result in a massive diuresis with associated decreases in plasma volume, cardiac output, systemic blood pressure, and, hence, coronary perfusion. Nitrates in various forms may be used to decrease preload and congestive symptoms. Oral isosorbide dinitrate, topical nitroglycerin ointment, or intravenous nitroglycerin all have the advantage over a diuretic of lowering preload through venodilation without decreasing the total plasma volume. In addition, nitrates may improve ventricular compliance if ischemia is present, as ischemia causes an elevation of LV filling pressure. Vasodilators must be used with caution to prevent serious hypotension. As noted earlier, ACE inhibitors are an ideal class of drugs for management of ventricular dysfunction after STEMI, especially for the long term. (See "Inhibition of the Renin-Angiotensin-Aldosterone System" above.)

Cardiogenic Shock

Prompt reperfusion, efforts to reduce infarct size and treatment of ongoing ischemia and other complications of MI appear to have reduced the incidence of cardiogenic shock from 20% to about 7%. Only 10% of patients with this condition present with it on admission, while 90% develop it during hospitalization. Typically, patients who develop cardiogenic shock have severe multivessel coronary artery disease with evidence of "piecemeal" necrosis extending outward from the original infarct zone. **The evaluation and management of cardiogenic shock and severe power failure after STEMI are discussed in detail in Chap. 272.**

Right Ventricular Infarction

Approximately one-third of patients with inferior infarction demonstrate at least a minor degree of RV necrosis. An occasional patient with inferoposterior LV infarction also has extensive RV infarction, and rare patients present with infarction limited primarily to the RV. Clinically significant RV infarction causes signs of severe RV failure [jugular venous distention, Kussmaul's sign, hepatomegaly (Chap. 227)] with or without hypotension. ST-segment elevations of right-sided precordial ECG leads, particularly lead V_4R, are frequently present in the first 24 h in patients with RV infarction. Two-dimensional echocardiography is helpful in determining the degree of RV dysfunction. Catheterization of the right side of the heart often reveals a distinctive hemodynamic pattern resembling constrictive pericarditis (steep right atrial "y" descent and an early diastolic dip and plateau in RV waveforms) (Chap. 239). Therapy consists of volume expansion to maintain adequate RV preload and efforts to improve LV performance with attendant reduction in pulmonary capillary wedge and pulmonary arterial pressures.

Arrhythmias

(See also Chaps. 232 and 233) The incidence of arrhythmias after STEMI is higher in patients seen early after the onset of symptoms. The mechanisms responsible for infarction-related arrhythmias include autonomic nervous system imbalance, electrolyte disturbances, ischemia, and slowed conduction in zones of ischemic myocardium. An arrhythmia can usually be managed successfully if trained personnel and appropriate equipment are available when it develops. Since most deaths from arrhythmia occur during the first few hours after infarction, the effectiveness of treatment relates directly to the speed with which patients come under medical observation. The prompt management of arrhythmias constitutes a significant advance in the treatment of STEMI.

VENTRICULAR PREMATURE BEATS

Infrequent, sporadic ventricular premature depolarizations occur in almost all patients with STEMI and do not require therapy. Whereas in the past, frequent, multifocal, or early diastolic ventricular extrasystoles (so-called warning arrhythmias) were routinely treated with antiarrhythmic drugs to reduce the risk of development of ventricular tachycardia and ventricular fibrillation, pharmacologic therapy is now reserved for patients with sustained ventricular arrhythmias. Prophylactic antiarrhythmic therapy (either intravenous lidocaine early or oral agents later) is contraindicated for ventricular premature beats in the absence of clinically important ventricular tachyarrhythmias, as such therapy may actually increase the mortality rate. Beta-adrenoceptor blocking agents are effective in abolishing ventricular ectopic activity in patients with STEMI and in the prevention of ventricular fibrillation. As described above (see "Beta-Adrenoceptor Blockers"), they should be used routinely in patients without contraindications. In addition, hypokalemia and hypomagnesemia are risk factors for ventricular fibrillation in patients with STEMI; to reduce the risk, the serum potassium concentration should be adjusted to approximately 4.5 mmol/L and magnesium to about 2.0 mmol/L.

VENTRICULAR TACHYCARDIA AND FIBRILLATION

Within the first 24 h of STEMI, ventricular tachycardia and fibrillation can occur without prior warning arrhythmias. The occurrence of ventricular fibrillation can be reduced by prophylactic administration of intravenous lidocaine. However, prophylactic use of lidocaine has not been shown to reduce overall mortality from STEMI. In fact, in addition to causing possible noncardiac complications, lidocaine may predispose to an excess risk of bradycardia and asystole. For these reasons, and with earlier treatment of active ischemia, more frequent use of beta-blocking agents, and the nearly universal success of electrical cardioversion or defibrillation, routine prophylactic antiarrhythmic drug therapy *is no longer recommended* .

Sustained ventricular tachycardia that is well tolerated hemodynamically should be treated with an intravenous regimen of amiodarone (bolus of 150 mg over 10 min, followed by infusion of 1.0 mg/min for 6 h and then 0.5 mg/min) or procainamide (bolus of 15 mg/kg over 20–30 min; infusion of 1–4 mg/min); if it does not stop promptly, electroversion should be used (Chap. 233). An unsynchronized discharge of 200–300 J (monophasic wave form; approximately 50% of these energies with biphasic wave forms) is used immediately in patients with ventricular fibrillation or when ventricular tachycardia causes hemodynamic deterioration. Ventricular tachycardia or fibrillation that is refractory to electroshock may be more responsive after the patient is treated with epinephrine (1 mg intravenously or 10 mL of a 1:10,000 solution via the intracardiac route) or amiodarone (a 75–150-mg bolus).

Ventricular arrhythmias, including the unusual form of ventricular tachycardia known as torsades des pointes (Chap. 233), may occur in patients with STEMI as a consequence of other concurrent problems (such as hypoxia, hypokalemia, or other electrolyte disturbances) or of the toxic effects of an agent being administered to the patient (such as digoxin or quinidine). A search for such secondary causes should always be undertaken.

Although the in-hospital mortality rate is increased, the long-term survival is excellent in patients who survive to hospital discharge after *primary* ventricular fibrillation; i.e., ventricular fibrillation that is a primary response to acute ischemia that occurs during the first 48 h and is not associated with predisposing factors such as CHF, shock, bundle branch block, or ventricular aneurysm. This result is in sharp contrast to the poor prognosis for patients who develop ventricular fibrillation *secondary* to severe pump failure. For patients who develop ventricular tachycardia or ventricular fibrillation late in their hospital course (i.e., after the first 48 h), the mortality rate is increased both in-hospital and during long-term follow-up. Such patients should be considered for electrophysiologic study and implantation of a cardioverter/defibrillator (ICD) (Chap. 233). A more challenging issue is the prevention of sudden cardiac death from ventricular fibrillation late after STEMI in patients who have not exhibited sustained ventricular tachyarrhythmias during their index hospitalization. An algorithm for selection of patients who warrant prophylactic implantation of an ICD is shown in **Fig. 245-6**.

Figure 245-6

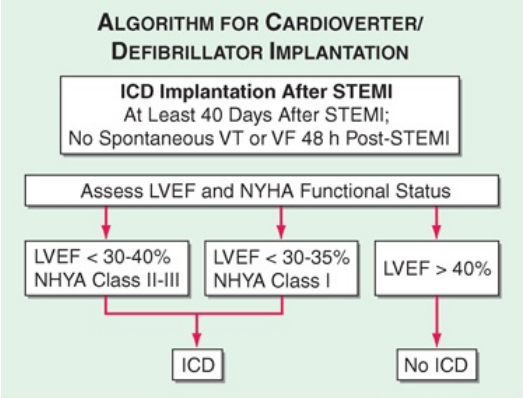

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for assessment of need for implantation of a cardioverter/defibrillator. The appropriate management is selected based upon measurement of left ventricular ejection fraction and assessment of the NYHA functional class. Patients with depressed left ventricular function at least 40 days post-STEMI are referred for insertion of an implantable cardioverter/defibrillator (ICD) if the LVEF is <30–40% and they are in NYHA class II–III or if the LVEF is <30–35% and they are in NYHA class I functional status. Patients with preserved left ventricular function (LVEF >40%) do not receive an ICD regardless of NYHA functional class. All patients are treated with medical therapy post-STEMI. *[Adapted from data contained in Zipes DP, et al: ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death; a report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Develop Guidelines for Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death). J Am Coll Cardiol 48:1064, 2006.]*

ACCELERATED IDIOVENTRICULAR RHYTHM

Accelerated idioventricular rhythm (AIVR, "slow ventricular tachycardia"), a ventricular rhythm with a rate of 60–100 bpm, often occurs transiently during fibrinolytic therapy at the time of reperfusion. For the most part, AIVR, whether it occurs in association with fibrinolytic therapy or spontaneously, is benign and does not presage the development of classic ventricular tachycardia. Most episodes of AIVR do not require treatment if the patient is monitored carefully, as degeneration into a more serious arrhythmia is rare.

SUPRAVENTRICULAR ARRHYTHMIAS

Sinus tachycardia is the most common supraventricular arrhythmia. If it occurs secondary to another cause (such as anemia, fever, heart failure, or a metabolic derangement), the primary problem should be treated first. However, if it appears to be due to sympathetic overstimulation (e.g., as part of a hyperdynamic state), then treatment with a beta blocker is indicated. Other common arrhythmias in this group are atrial flutter and atrial fibrillation, which are often secondary to LV failure. Digoxin is usually the treatment of choice for supraventricular arrhythmias if heart failure is present. If heart failure is absent, beta blockers, verapamil, or diltiazem are suitable alternatives for controlling the ventricular rate, as they may also help to control ischemia. If the abnormal rhythm persists for >2 h with a ventricular rate >120 bpm, or if tachycardia induces heart failure, shock, or ischemia (as manifested by recurrent pain or ECG changes), a synchronized electroshock (100–200 J monophasic wave form)

should be used.

Accelerated junctional rhythms have diverse causes but may occur in patients with inferoposterior infarction. Digitalis excess must be ruled out. In some patients with severely compromised LV function, the loss of appropriately timed atrial systole results in a marked reduction of cardiac output. Right atrial or coronary sinus pacing is indicated in such instances.

SINUS BRADYCARDIA

Treatment of sinus bradycardia is indicated if hemodynamic compromise results from the slow heart rate. Atropine is the most useful drug for increasing heart rate and should be given intravenously in doses of 0.5 mg initially. If the rate remains <50–60 bpm, additional doses of 0.2 mg, up to a total of 2.0 mg, may be given. Persistent bradycardia (<40 bpm) despite atropine may be treated with electrical pacing. Isoproterenol should be avoided.

ATRIOVENTRICULAR AND INTRAVENTRICULAR CONDUCTION DISTURBANCES

(See also Chap. 232) Both the in-hospital mortality rate and the post-discharge mortality rate of patients who have complete atrioventricular (AV) block in association with anterior infarction are markedly higher than those of patients who develop AV block with inferior infarction. This difference is related to the fact that heart block in inferior infarction is commonly a result of increased vagal tone and/or the release of adenosine and therefore is transient. In anterior wall infarction, however, heart block is usually related to ischemic malfunction of the conduction system, which is commonly associated with extensive myocardial necrosis.

Temporary electrical pacing provides an effective means of increasing the heart rate of patients with bradycardia due to AV block. However, acceleration of the heart rate may have only a limited impact on prognosis in patients with anterior wall infarction and complete heart block in whom the large size of the infarct is the major factor determining outcome. It should be carried out if it improves hemodynamics. Pacing does appear to be beneficial in patients with inferoposterior infarction who have complete heart block associated with heart failure, hypotension, marked bradycardia, or significant ventricular ectopic activity. A subgroup of these patients, those with RV infarction, often respond poorly to ventricular pacing because of the loss of the atrial contribution to ventricular filling. In such patients, dual-chamber AV sequential pacing may be required.

External noninvasive pacing electrodes should be positioned in a "demand" mode for patients with sinus bradycardia (rate <50 bpm) that is unresponsive to drug therapy, Mobitz II second-degree AV block, third-degree heart block, or bilateral bundle branch block (e.g., right bundle branch block plus left anterior fascicular block). Retrospective studies suggest that permanent pacing may reduce the long-term risk of sudden death due to bradyarrhythmias in the rare patient who develops combined persistent bifascicular and transient third-degree heart block during the acute phase of MI.

Other Complications

RECURRENT CHEST DISCOMFORT

Recurrent angina develops in ~25% of patients hospitalized for STEMI. This percentage is even higher in patients who undergo successful fibrinolysis. Because recurrent or persistent ischemia often heralds extension of the original infarct or reinfarction in a new myocardial zone and is associated with a near tripling of mortality after STEMI, patients with these symptoms should be referred for prompt coronary arteriography and mechanical revascularization. Repeat administration of a fibrinolytic agent is an alternative to early mechanical revascularization.

PERICARDITIS

(See also Chap. 239) Pericardial friction rubs and/or pericardial pain are frequently encountered in patients with STEMI involving the epicardium. This complication can usually be managed with aspirin (650 mg 4 times daily). It is important to diagnose the chest pain of pericarditis accurately, because failure to recognize it may lead to the erroneous diagnosis of recurrent ischemic pain and/or infarct extension, with resulting inappropriate use of anticoagulants, nitrates, beta blockers, or coronary arteriography. When it occurs, complaints of pain radiating to either trapezius muscle is helpful, because such a pattern of discomfort is typical of pericarditis but rarely occurs with ischemic discomfort. Anticoagulants potentially could cause tamponade in the presence of acute pericarditis (as manifested by either pain or persistent rub) and therefore should not be used unless there is a compelling indication.

THROMBOEMBOLISM

Clinically apparent thromboembolism complicates STEMI in ~10% of cases, but embolic lesions are found in 20% of patients in necropsy series, suggesting that thromboembolism is often clinically silent. Thromboembolism is considered to be an important contributing cause of death in 25% of patients with STEMI who die after admission to the hospital. Arterial emboli originate from LV mural thrombi, while most pulmonary emboli arise in the leg veins.

Thromboembolism typically occurs in association with large infarcts (especially anterior), CHF, and a LV thrombus detected by echocardiography. The incidence of arterial embolism from a clot originating in the ventricle at the site of an infarction is small but real. Two-dimensional echocardiography reveals LV thrombi in about one-third of patients with anterior wall infarction but in few patients with inferior or posterior infarction. Arterial embolism often presents as a major complication, such as hemiparesis when the cerebral circulation is involved or hypertension if the renal circulation is compromised. When a thrombus has been clearly demonstrated by echocardiographic or other techniques or when a large area of regional wall motion abnormality is seen even in the absence of a detectable mural thrombus, systemic anticoagulation should be undertaken (in the absence of contraindications), as the incidence of embolic complications appears to be markedly lowered by such therapy. The appropriate duration of therapy is unknown, but 3–6 months is probably prudent.

LEFT VENTRICULAR ANEURYSM

The term *ventricular aneurysm* is usually used to describe *dyskinesis* or local expansile paradoxical wall motion. Normally functioning myocardial fibers must shorten more if stroke volume and cardiac output are to be

maintained in patients with ventricular aneurysm; if they cannot, overall ventricular function is impaired. True aneurysms are composed of scar tissue and neither predispose to nor are associated with cardiac rupture.

The complications of LV aneurysm do not usually occur for weeks to months after STEMI; they include CHF, arterial embolism, and ventricular arrhythmias. Apical aneurysms are the most common and the most easily detected by clinical examination. The physical finding of greatest value is a double, diffuse, or displaced apical impulse. Ventricular aneurysms are readily detected by two-dimensional echocardiography, which may also reveal a mural thrombus in an aneurysm.

Rarely, myocardial rupture may be contained by a local area of pericardium, along with organizing thrombus and hematoma. Over time, this *pseudoaneurysm* enlarges, maintaining communication with the LV cavity through a narrow neck. Because a pseudoaneurysm often ruptures spontaneously, it should be surgically repaired if recognized.

POSTINFARCTION RISK STRATIFICATION AND MANAGEMENT

Many clinical and laboratory factors have been identified that are associated with an increase in cardiovascular risk after initial recovery from STEMI. Some of the most important factors include persistent ischemia (spontaneous or provoked), depressed LV ejection fraction (<40%), rales above the lung bases on physical examination or congestion on chest radiograph, and symptomatic ventricular arrhythmias. Other features associated with increased risk include a history of previous MI, age >75, diabetes mellitus, prolonged sinus tachycardia, hypotension, ST-segment changes at rest without angina ("silent ischemia"), an abnormal signal-averaged ECG, nonpatency of the infarct-related coronary artery (if angiography is undertaken), and persistent advanced heart block or a new intraventricular conduction abnormality on the ECG. Therapy must be individualized on the basis of the relative importance of the risk(s) present.

The goal of preventing reinfarction and death after recovery from STEMI has led to strategies to evaluate risk after infarction. In stable patients, submaximal exercise stress testing may be carried out before hospital discharge to detect residual ischemia and ventricular ectopy and to provide the patient with a guideline for exercise in the early recovery period. Alternatively, or in addition, a maximal (symptom-limited) exercise stress test may be carried out 4–6 weeks after infarction. Evaluation of LV function is usually warranted as well. Recognition of a depressed LV ejection fraction by echocardiography or radionuclide ventriculography identifies patients who should receive medications to inhibit the renin-angiotensin-aldosterone system. Patients in whom angina is induced at relatively low workloads, those who have a large reversible defect on perfusion imaging or a depressed ejection fraction, those with demonstrable ischemia, and those in whom exercise provokes symptomatic ventricular arrhythmias should be considered at high risk for recurrent MI or death from arrhythmia (Fig. 245-6). Cardiac catheterization with coronary angiography and/or invasive electrophysiologic evaluation is advised.

Exercise tests also aid in formulating an individualized exercise prescription, which can be much more vigorous in patients who tolerate exercise without any of the above-mentioned adverse signs. In addition, predischarge stress testing may provide an important psychological benefit, building the patient's confidence by demonstrating a reasonable exercise tolerance.

In many hospitals, a cardiac rehabilitation program with progressive exercise is initiated in the hospital and continued after discharge. Ideally, such programs should include an educational component that informs patients about their disease and its risk factors.

The usual duration of hospitalization for an uncomplicated STEMI is about 5 days. The remainder of the convalescent phase may be accomplished at home. During the first 1–2 weeks, the patient should be encouraged to increase activity by walking about the house and outdoors in good weather. Normal sexual activity may be resumed during this period. After 2 weeks, the physician must regulate the patient's activity on the basis of exercise tolerance. Most patients will be able to return to work within 2–4 weeks.

SECONDARY PREVENTION

Various secondary preventive measures are at least partly responsible for the improvement in the long-term mortality and morbidity rates after STEMI. Long-term treatment with an antiplatelet agent (usually aspirin) after STEMI is associated with a 25% reduction in the risk of recurrent infarction, stroke, or cardiovascular mortality (36 fewer events for every 1000 patients treated). An alternative antiplatelet agent that may be used for secondary prevention in patients intolerant of aspirin is clopidogrel (75 mg orally daily). ACE inhibitors or ARBs and, in appropriate patients, aldosterone antagonists should be used indefinitely by patients with clinically evident heart failure, a moderate decrease in global ejection fraction, or a large regional wall motion abnormality to prevent late ventricular remodeling and recurrent ischemic events.

The chronic routine use of oral beta-adrenoceptor blockers for at least two years after STEMI is supported by well-conducted, placebo-controlled trials.

Evidence suggests that warfarin lowers the risk of late mortality and the incidence of reinfarction after STEMI. Most physicians prescribe aspirin routinely for all patients without contraindications and add warfarin for patients at increased risk of embolism (see "Thromboembolism" above). Several studies suggest that in patients <75 years a low dose of aspirin (75–81 mg/d) in combination with warfarin administered to achieve an INR >2.0 is more effective than aspirin alone for preventing recurrent MI and embolic cerebrovascular accident. However, there is an increased risk of bleeding and a high rate of discontinuation of warfarin that has limited clinical acceptance of combination antithrombotic therapy. There is increased risk of bleeding when warfarin is added to dual antiplatelet therapy (aspirin and clopidogrel). However, patients who have had a stent implanted and have an indication for anticoagulation should receive dual antiplatelet therapies in combination with warfarin. Such patients should also receive a proton pump inhibitor to minimize the risk of gastrointestinal bleeding and should have regular monitoring of their hemoglobin levels and stool hematest while on combination antithrombotic therapy.

Finally, risk factors for *atherosclerosis* (Chap. 224) should be discussed with the patient, and, when possible, favorably modified.

FURTHER READINGS

Antman EM: ST-Elevation Myocardial Infarction: Management, in: P Libby, RO Bonow, DL Mann, DP Zipes (eds) *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine* , 8th ed, Philadelphia, Saunders Elsevier, 2008, pp 1233–1299

———: Time is muscle: Translation into practice. J Am Coll Cardiol 52:1216, 2008

Braunwald E, Antman EM: ST-Elevation Myocardial Infarction: Pathology, Pathophysiology, and Clinical Features in P Libby, RO Bonow, DL Mann, DP Zipes (eds) *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 8th ed, Philadelphia, Saunders Elsevier, 2008, pp 1207–1232

Jacobs AK et al: Development of systems of care for ST-elevation myocardial infarction patients: executive summary. Circulation 116:217, 2007[PMID: 17538045]

Kushner FG et al: 2009 Focused Updates: ACC/AHA Guidelines for the Management of Patients With ST-Elevation Myocardial Infarction (updating the 2004 Guideline and 2007 Focused Update) and ACC/AHA/SCAI Guidelines on Percutaneous Coronary Intervention (updating the 2005 Guideline and 2007 Focused Update): A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 120:2271, 2009[PMID: 19923169]

Lloyd-Jones D et al: Heart disease and stroke statistics—2009 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 119:e21, 2009

Mehran R et al: Bivalirudin in patients undergoing Primary angioplasty for acute myocardial infarction (HORIZONS-AMI): 1-year results of a randomised controlled trial. Lancet 374:1149, 2009[PMID: 19717185]

White HD, Chew DP: Acute myocardial infarction. Lancet 372:570, 2008[PMID: 18707987]

Wiviott SD et al: Prasugrel versus clopidogrel in patients with acute coronary syndromes. N Engl J Med 357:2001, 2007[PMID: 17982182]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PERCUTANEOUS CORONARY INTERVENTIONS AND OTHER INTERVENTIONAL PROCEDURES: INTRODUCTION

Percutaneous transluminal coronary angioplasty (PTCA) was first introduced by Andreas Gruentzig in 1977 as an alternative to coronary bypass surgery. The concept of percutaneous dilatation of the atherosclerotic peripheral vessels was initially demonstrated by Charles Dotter in 1964 in peripheral vessels where rigid catheters of graduated diameter were used to progressively enlarge the vessel lumen. The development of a small inelastic balloon catheter by Gruentzig allowed expansion of the technique into smaller peripheral and coronary vessels. Initial coronary experience was limited to the small percentage of patients who had single-vessel coronary disease and discrete proximal lesions due to the technical limitations of the equipment. Advances in technology and greater operator experience allowed the procedure to grow rapidly with expanded use in patients with more complex lesions and multivessel disease; by 1990 it was being performed in more than 300,000 patients annually. The addition of atherectomy devices that removed plaques aided in the growth of the procedure, but the introduction of coronary stents in 1994 was one of the major advances in the field. These devices reduced acute complications and reduced by half the significant problem of restenosis (or recurrence of the stenosis). Further reductions in restenosis were achieved by the introduction of drug-eluting stents in 2003. These stents have a polymer coating over the metal stent that is impregnated with antiproliferative agents that slowly release drugs directly into the plaque over a few months. Today, more than 1 million stents are placed in the United States per year and more than 4 million worldwide. Percutaneous coronary intervention (PCI) is the most common revascularization procedure in the United States and is performed nearly twice as often as coronary artery bypass surgery.

The field of interventional cardiology has matured to be recognized as a separate discipline in cardiology that requires specialized training. A dedicated 1-year interventional cardiology fellowship following a 3-year general cardiology fellowship and a separate board certification examination are now required to be certified in interventional cardiology. The discipline has also expanded to include interventions for structural heart disease including treatment of congenital heart disease, and valvular heart disease; it also includes interventions to treat peripheral vascular disease, including atherosclerotic and nonatherosclerotic lesions in the carotid, renal, aortic, and peripheral circulations.

TECHNIQUE

The initial procedure is performed in a similar manner as a diagnostic cardiac catheterization (Chap. 230). As is done with diagnostic catheterization, arterial access is obtained by percutaneous needle puncture into a peripheral artery. Most commonly, the arterial access site is the femoral artery, but radial artery access is gaining favor. To prevent thrombotic complications during the procedure, patients who are anticipated to need an angioplasty are given aspirin (325 mg) and clopidogrel (loading dose of 300–600 mg) before the procedure. During the procedure, anticoagulation is achieved by administration of unfractionated heparin, enoxaparin (a low-molecular-weight heparin), or bivalirudin (a direct thrombin inhibitor). In patients with ST-elevation myocardial infarction, high-risk acute coronary syndrome, or those with a large thrombus in the coronary artery, a glycoprotein IIb/IIIa inhibitor (abciximab, tirofiban, or eptifibatide) may also be given.

Following placement of an introducing sheath, preformed guiding catheters are used to cannulate selectively the origins of the coronary arteries. These catheters have larger internal diameters than diagnostic catheters in order to allow passage of the balloon catheter and wires. Through the guiding catheter, a flexible, steerable guidewire (diameter 0.4 mm) is negotiated down the coronary artery lumen using fluoroscopic guidance; it is then advanced through the stenosis and into the vessel beyond. This guidewire then serves as a "rail" over which angioplasty balloons, stents, or other therapeutic devices can be advanced to enlarge the narrowed segment of coronary artery. The artery is usually dilated with a balloon catheter and most often a stent is then placed with assessment of the final result by repeat angiography through the guiding catheter. The catheters and introducing sheath are removed and the artery manually held or closed using one of several arterial closure devices to achieve hemostasis. Because PCI is performed under local anesthesia and mild sedation, it requires only a short (1-day) hospitalization that decreases recovery time and hospital expense, as compared to coronary bypass surgery.

The inflated diameter of the angioplasty balloons range in size from 1.5 mm to 4.0 mm, and balloons are chosen to approximate the "normal" less diseased proximal or distal vessel without stenosis. The major advance introduced by Dr. Gruentzig was the use of inelastic balloons that do not overexpand the vessel beyond their predetermined size despite high pressures up to 10–20 atmospheres.

Angioplasty works by stretching the artery and compressing the plaque into the vessel wall, away from the lumen, enlarging the entire vessel (**Figs. 246-1** and 246-2). The procedure rarely results in embolization of atherosclerotic material. Owing to inelastic elements in the plaque, the stretching of the vessel by the balloon results in small localized dissections that can protrude into the lumen and be a nidus for acute thrombus formation. If the dissections are severe, then they can obstruct the lumen or induce a thrombotic occlusion of the artery (acute closure). Stents have largely prevented this complication by holding the dissection flaps up against the vessel wall (Fig. 246-1).

Figure 246-1

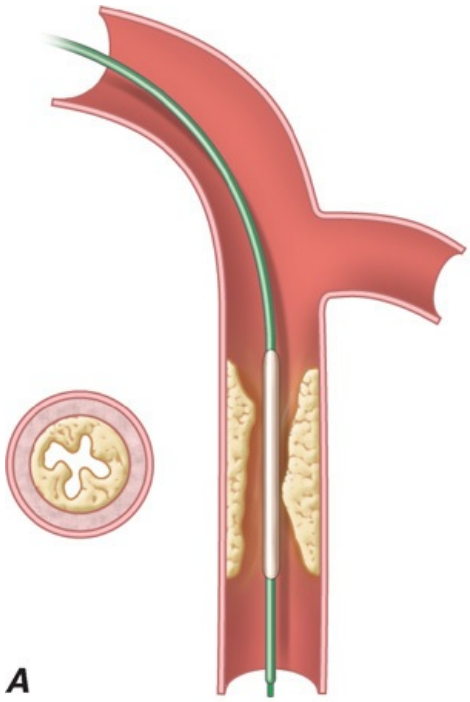

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

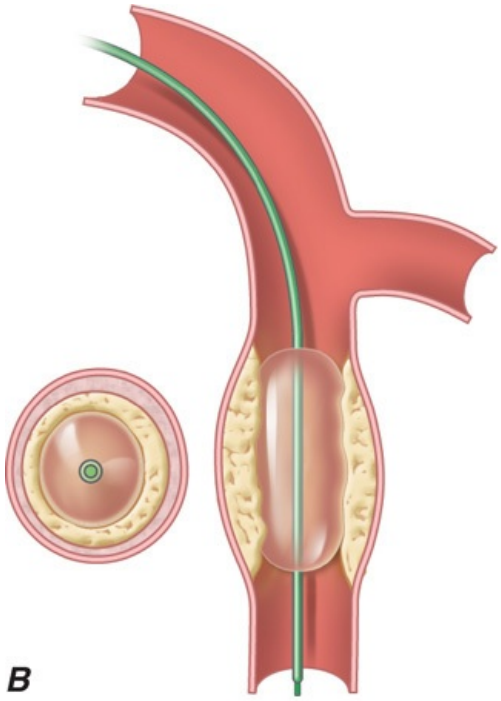

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

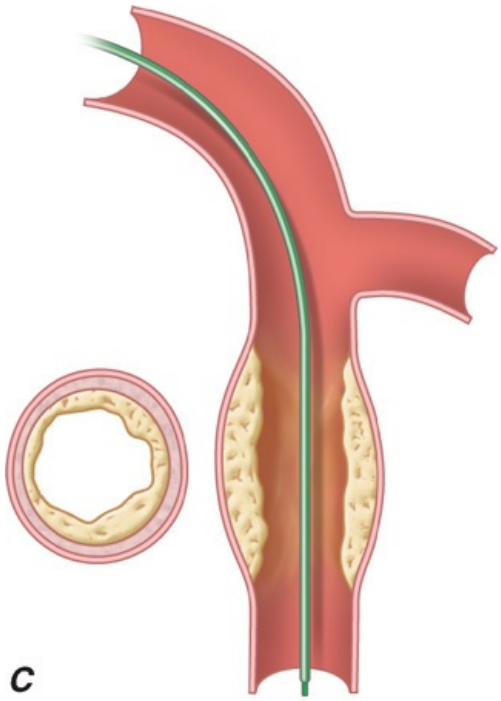

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

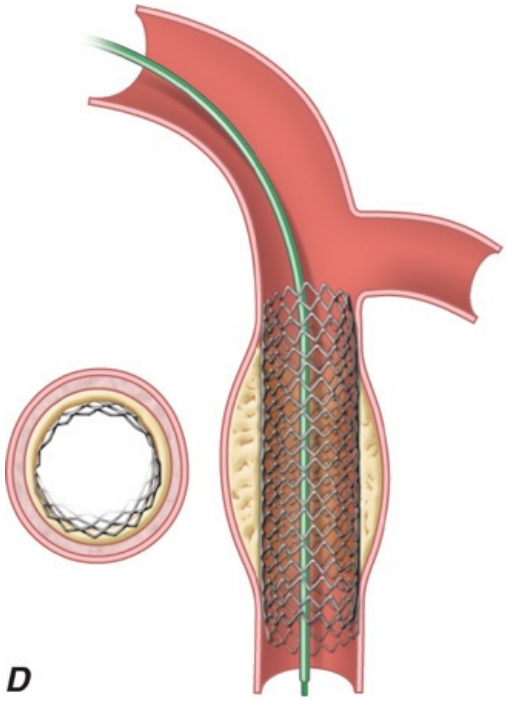

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of the primary mechanisms of balloon angioplasty and stenting. ***A.*** A balloon angioplasty catheter is positioned into the stenosis over a guidewire under fluoroscopic guidance. ***B***. The balloon is inflated temporarily occluding the vessel. ***C***. The lumen is enlarged primarily by stretching the vessel often resulting in small dissections in the neointima. ***D***. A stent mounted on a deflated balloon is placed into the lesion and pressed against the vessel wall with balloon inflation (not shown). The balloon is deflated and removed leaving the stent permanently against the wall acting as a scaffold to hold the dissections against the wall and prevent vessel recoil. *(Adapted from EJ Topol: Textbook of Cardiovascular Medicine, 2nd ed. Philadelphia, Lippincott Williams & Wilkins, 2002.)*

Figure 246-2

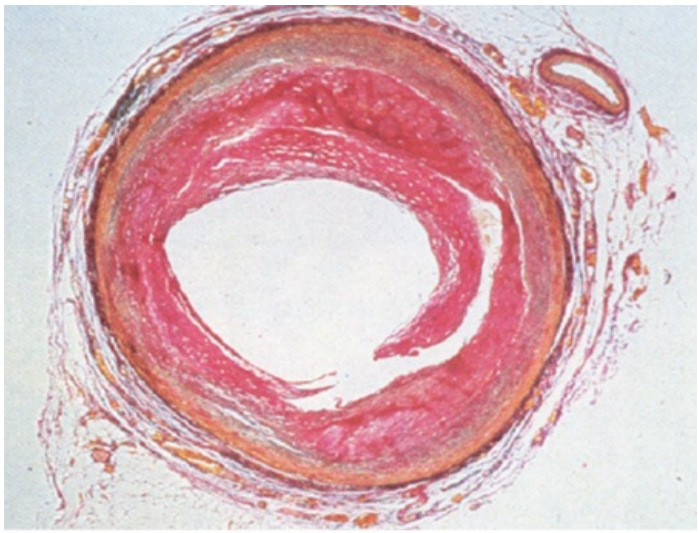

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

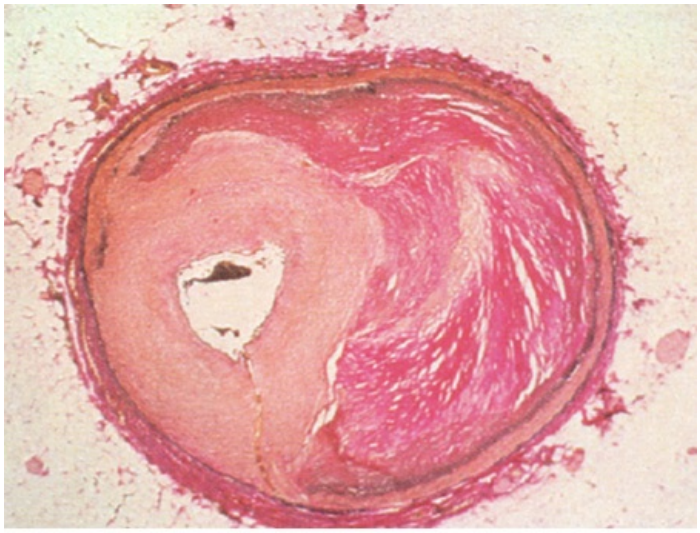

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathology of acute effects of balloon angioplasty with intimal dissection and vessel stretching (*panel A*) (*From M Ueda et al: Eur Heart J 12:937, 1991; with permission*) and an example of neointimal hyperplasia and restenosis showing renarrowing of the vessel (*panel B*). (*From CE Essed et al: Br Heart J 49:393, 1983; with permission.*)

Stents are currently used in more than 90% of coronary angioplasty procedures. Stents are wire meshes (usually made of stainless steel) that are compressed over a deflated angioplasty balloon. When the balloon is inflated, the stent is enlarged to approximate the "normal" vessel lumen. The balloon is then deflated and removed, leaving the stent behind to provide a permanent scaffold in the artery. Owing to the design of the struts, these devices are flexible, allowing their passage through diseased and tortuous coronary vessels. Stents are rigid enough to prevent elastic recoil of the vessel and have dramatically improved the success and safety of the procedure as a result.

Drug-eluting stents were first introduced in 2003. Using a metal stent, an antiproliferative agent is attached to the stent by use of a thin polymer coating. The antiproliferative drug elutes from the stent over a 1- to 3-month period after implantation. Drug-eluting stents have been shown to reduce clinical restenosis by 50% so that in uncomplicated lesions symptomatic restenosis occurs in 5–12% of patients. Not surprisingly, this led to the rapid acceptance of these devices; currently 50–90% of all stents implanted are drug-eluting. The first-generation devices were coated with either sirolimus or paclitaxel. *Sirolimus* is an immunosuppressive agent that arrests cell proliferation in the G_1 phase. *Paclitaxel* is an inhibitor of microtubules that can arrest cell division at the M phase in high concentrations, but can have cytostatic G_1, antimigratory, and antiinflammatory effects on smooth muscle cells at lower concentrations. Second-generation drug-eluting stents use newer agents such as everolimus, biolimus, and zotarolimus. These second-generation drug-eluting stents appear to be more effective with fewer complications than the first-generation devices. Preliminary data from long-term follow-up suggests that the second-generation drug-eluting stents have lower rates of stent thrombosis and myocardial infarction than the first-generation drug-eluting stents.

Other interventional devices include atherectomy devices, laser catheters, and thrombectomy catheters. These devices are designed to remove atherosclerotic plaque or thrombus and are used in conjunction with balloon dilatation and stent placement. Rotational atherectomy is the most commonly used adjunctive device for heavily calcified lesions and is modeled after a dentist's drill, with small round burrs of 1.5–2.5 mm at the tip of a flexible wire shaft. They are passed over the guidewire up to the stenosis and activated to rotate at 180,000 rpm in order to drill away atherosclerotic material. Because the atherosclerotic particles are <25 μm, they pass through the coronary microcirculation and rarely cause problems. The device is particularly useful in heavily calcified plaques that are resistant to balloon dilatation. Another available device is the directional atherectomy catheter. This catheter has a rigid housing at its tip that is open on one side, exposing a sliding rotating cutter. The catheter is placed in the stenosis, and a balloon on the noncutting side of the housing is inflated to push the housing up against the wall of the artery. When the cutter is rotated at 2500 rpm and advanced down the housing, it slices off atherosclerotic plaques into a distal collection chamber, allowing the plaque to be removed from the patient. Given the current advances in stents, neither rotational nor directional atherectomy is as frequently used today as in the past. Other devices include fiberoptic laser catheters that can vaporize atherosclerotic plaques. These are infrequently used today, as well. In acute myocardial infarction, specialized catheters without a balloon are used to aspirate thrombus in order to prevent embolization down the coronary vessel and to improve blood flow before angioplasty and stent

placement. Data suggest that manual catheter thrombus aspiration may even reduce mortality rate in primary PCI.

PCI of degenerated saphenous vein graft lesions has been associated with a significant incidence of distal embolization of atherosclerotic material, unlike PCI of native vessel disease. A number of distal protection devices have been shown to significantly reduce embolization and myocardial infarction in this setting. Most devices work by using a collapsible wire mesh at the end of a guidewire that is expanded in the distal vessel before angioplasty. If atherosclerotic debris is dislodged, the basket captures the material, and at the end of the PCI, the basket is pulled into a delivery catheter and the debris safely removed from the patient.

SUCCESS AND COMPLICATIONS

The advances in the technology have greatly improved the success and reduced the complications of the procedure. Currently, a successful procedure (angiographic success), defined as a reduction of the stenosis to less than a 20% diameter narrowing, occurs in 95–99% of patients. The success is dependent upon the coronary anatomy, with lower success rates in patients with tortuous, small, or calcified vessels or chronic total occlusions. Chronic total occlusions have the lowest success rates and their recanalization is usually not attempted unless the occlusion is recent (within 3 months) or there are favorable anatomic features. Improvements in equipment and technique have increased the success rates of recanalization of chronic total occlusions.

Serious complications are rare but include a mortality rate of 0.1–0.3% for elective cases, a large myocardial infarction occurs in less than 3%, and stroke in less than 0.1%. Patients who are elderly (>65 years), undergoing an emergent or urgent procedure, have chronic kidney disease, present with an ST-segment elevation myocardial infarction (STEMI), or are in shock have significantly higher risk. Scoring systems can help to estimate the risk of the procedure, although no perfect scoring system has yet been developed.

Myocardial infarction during PCI can occur for multiple reasons including an acute occluding thrombus, severe coronary dissection, embolization of thrombus or atherosclerotic material, or closure of a side branch vessel at the site of angioplasty. Most myocardial infarctions are small and only detected by a rise in the creatinine phosphokinase (CPK) or troponin level after the procedure. Only those with significant enzyme elevations (more than three times the upper limit of normal) are associated with a less favorable long-term outcome. Coronary stents have largely prevented coronary dissections due to the scaffolding effect of the stent. Metallic stents are also prone to thrombotic occlusion (1–3%), either acute (<24 h) or subacute (1–30 days), which can be ameliorated by greater attention to full initial stent deployment and the use of dual antiplatelet therapy [aspirin, plus a platelet P_2Y_{12}-receptor blocker (clopidogrel or prasugrel)]. Late (30 days–1 year) and very late stent thromboses (>1 year) occur very infrequently with stents but are slightly more common with drug-eluting stents, necessitating dual antiplatelet therapy with these stents for up to 1 year or longer. Premature discontinuation of dual antiplatelet therapy particularly in the first month after implantation is associated with a significantly increased risk for stent thrombosis (three- to ninefold greater). Stent thrombosis results in death in 10–20% and a myocardial infarction in 30–70% of patients. Elective surgery that requires discontinuation of antiplatelet therapy after drug-eluting stent implantation should be postponed until after 6 months and preferably after 1 year, if at all possible.

Restenosis, or renarrowing of the dilated coronary stenosis, is the most common complication of angioplasty and occurs in 20–50% of patients with balloon angioplasty alone, 10–30% of patients with bare metal stents, and in 5–15% of patients with drug-eluting stents. The fact that stent placement provides a larger acute luminal area than balloon angioplasty alone reduces the incidence of subsequent restenosis. Drug-eluting stents further reduce restenosis through a reduction in excessive neointimal growth over the stent. If restenosis does not occur, the long-term outcome is excellent (**Fig. 246-3**). Clinical restenosis is recognized by recurrence of angina or symptoms within 9 months of the procedure. Most commonly, patients with clinical restenosis present with worsening angina (60–70%), but patients can present with non-ST-elevation myocardial infarction (10%) or ST-elevation myocardial infarction (5%) as well. Clinical restenosis requires confirmation of a significant stenosis at the site of the prior PCI, with repeat PCI or coronary artery bypass grafting (CABG). This is termed *target lesion revascularization* (TLR) or *target vessel revascularization* (TVR). By angiography, the incidence of restenosis is significantly higher than clinical restenosis (TLR or TVR) because many patients have mild restenosis that does not result in a recurrence of symptoms. The management of clinical restenosis is usually to repeat the PCI with balloon dilatation and, placement of a bare metal or a drug-eluting stent. Rarely, intracoronary brachytherapy using beta radiation is used. Once a patient has had restenosis, the risk of a second restenosis is further increased. The risk factors for restenosis are diabetes, long lesions, small-diameter vessels, and suboptimal initial PCI result.

Figure 246-3

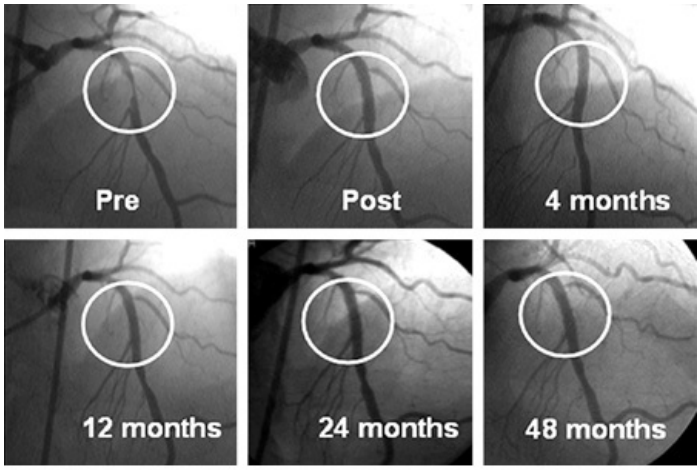

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Long-term results from one of the first patients to receive a sirolimus-eluting stent from early Sao Paulo experience. [*From: GW Stone, in D Baim (ed): Cardiac Catheterization, Angiography and Intervention, 7th ed, Philadelphia, Lippincott Williams & Wilkins, 2006; with permission.*]

The mechanism of restenosis is similar to that of wound healing, with inflammation and the migration and proliferation of smooth muscle cells that create a thick neointima (scar) that narrows the lumen at the site of dilatation (Fig. 246-2). The neointima is covered with endothelium, but it remains dysfunctional. The primary cause of restenosis in balloon angioplasty is adverse vessel remodeling with constriction of the vessel relative to the adjacent nondilated vessel. This change in remodeling can be appreciated by intravascular ultrasound but not by angiography since the latter only shows the lumen and not the entire vessel size. In addition to remodeling, excessive growth of the neointima further narrows the lumen. Stents prevent this unfavorable constrictive remodeling, and drug-eluting stents not only prevent this constriction but reduce the excessive neointimal growth as well. Common risk factors for atherosclerosis such as hyperlipidemia, hypertension, or cigarette smoking do not increase the risk of restenosis, although diabetes mellitus does.

Indications

The American College of Cardiology (ACC)/American Heart Association (AHA) guidelines extensively review the indications for PCI in patients with stable angina, unstable angina, non-ST-elevation, and ST-elevation myocardial infarction and should be referred to for a comprehensive discussion of the indications. Briefly, the two principal indications for coronary revascularization in patients with *chronic stable angina* (Chap. 243) are (1) to improve anginal symptoms in patients who remain symptomatic despite adequate medical therapy and (2) to reduce mortality rates in patients with severe coronary disease. In patients with stable angina, who are well controlled on medical therapy, older studies and the more current Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) and Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) trials have shown that revascularization does not lead to better outcomes and can be safely delayed until symptoms worsen or evidence of severe ischemia on noninvasive testing occurs. Randomized trials done in the 1960s and 1970s showed that CABG reduced mortality rates in patients with severe three-vessel or left main coronary disease when compared with medical therapy alone regardless of the degree of symptoms. Whether PCI also confers the same degree of protection is not known as trials of PCI versus medical therapy in patients with three-vessel disease have not been conducted, but randomized trials comparing CABG and PCI have shown equal rates of death and myocardial infarction (MI) rates over 5–10 years of follow-up. Consistently these studies have also shown that PCI, despite the use of stents, is associated with a 10–30% need for repeat PCI during the first year after the procedure due largely to restenosis, although drug-eluting stents have decreased this rate. This contrasts with a need for PCI or repeat CABG in bypass patients of 2–5%.

When revascularization is indicated, the choice of PCI or CABG depends upon a number of clinical and anatomic factors (**Fig. 246-4**). A subgroup analysis from the Bypass Angioplasty Revascularization Investigation (BARI) randomized trial showed that patients with treated diabetes mellitus and multivessel disease fared better with CABG; however, registry experiences suggest that PCI can be done in selected diabetic patients with less-severe multivessel disease with good long-term outcome. The Synergy between Percutaneous Coronary Intervention with Taxus and Cardiac Surgery (SYNTAX) trial compared PCI with the paclitaxel drug-eluting stent to CABG in 1800 patients with three-vessel coronary disease or left main disease. The study found no difference in death or myocardial infarction at 1 year, but repeat revascularization was significantly higher in the stent-treated group (13.5 vs. 5.9%), while stroke was higher in the surgical group (2.2 vs. 0.6%). The primary endpoint of death, MI,

stroke, or revascularization was significantly better with CABG due to the higher rate of revascularization in the drug-eluting stent group. Only 1 year of results are currently available and longer follow-up is needed to assess fully these two revascularization strategies in patients with severe coronary disease.

Figure 246-4

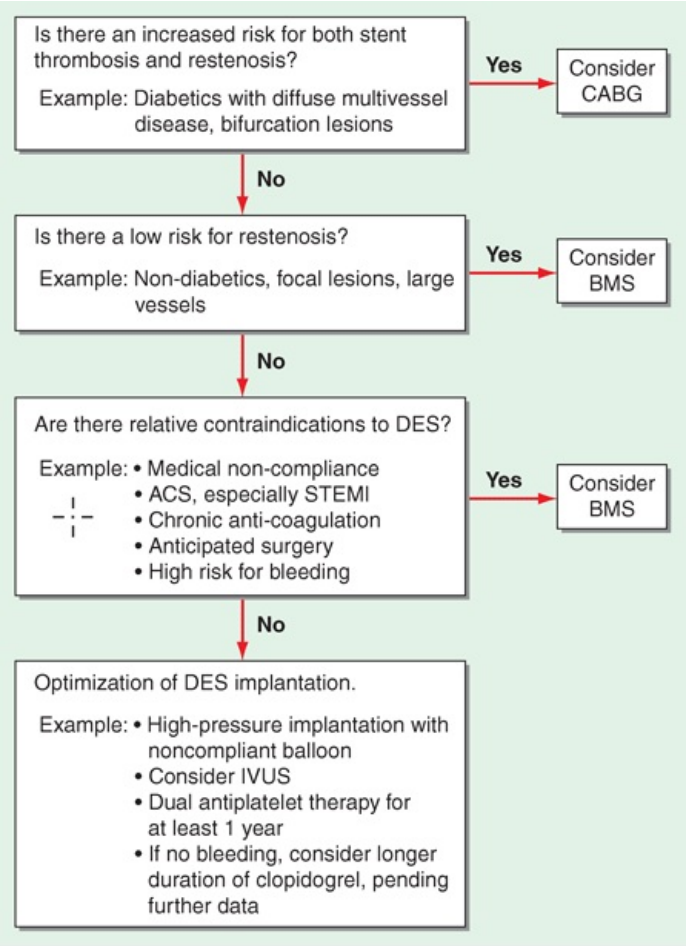

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

In patients requiring revascularization, several factors need to be considered in choosing between bare metal stents, drug-eluting stents, or coronary artery bypass surgery. ACS, acute coronary syndrome; BMS, bare metal stent; CABG, coronary artery bypass grafting; DES, drug-eluting stent; IVUS, intravascular ultrasound; STEMI, ST-segment elevation myocardial infarction. (*From AA Bavry, DL Bhatt: Circulation 116:696, 2007; with permission.*)

The choice of PCI versus CABG is also related to the anticipated procedural success and complications of PCI and the risks of CABG. For PCI, the characteristics of the coronary anatomy are critically important. The location of the lesion in the vessel (proximal or distal), the degree of tortuosity, and size of the vessel are considered. In addition, the lesion characteristics including the degree of the stenosis, the presence of calcium, lesion length, and presence of thrombus are assessed. The most common reason to decide not to do angioplasty is that the lesion felt to be responsible for the patient's symptoms is not treatable. This is most commonly due to the presence of a chronic total occlusion (>3 months in duration). In this setting the historical success rate has been low (30–70%) and complications are more common. A lesion classification to characterize the likelihood of success or failure of PCI has been developed by the ACC/AHA. Lesions with the highest success are called type A lesions (such as proximal noncalcified subtotal lesion) and those with the lowest success or highest complication rate are type C lesions (such as chronic total occlusions). Intermediate lesions are classified as type B1 or B2 depending on the number of unfavorable characteristics. Approximately 25–30% of patients will not be candidates for PCI due to unfavorable anatomy, whereas only 5% of CABG patients will not be candidates for surgery due to coronary anatomy. The primary reason for being considered inoperable is the presence of severe comorbidities such as advanced age, frailty, severe chronic obstructive pulmonary disease (COPD), or poor left ventricular function. Another consideration in choosing a revascularization strategy is the degree of revascularization. In patients with multivessel disease, bypass grafts can usually be placed in all vessels with significant stenosis,

while PCI may be able to treat only some of the lesions due to the presence of unfavorable anatomy. The decision to do PCI versus CABG will then depend upon the importance of complete revascularization in the patient. Given the multiple factors that need to be considered in choosing the best revascularization for an individual patient with multivessel disease, it is optimal to have a discussion between the cardiac surgeon and interventional cardiologist and the physicians caring for the patient to properly weigh the choices.

Patients with acute coronary syndrome are at excess risk of short- and long-term mortality. Randomized clinical trials have shown that PCI is superior to intensive medical therapy in reducing mortality rate and myocardial infarction, with the benefit largely confined to those patients who are high risk. This includes patients with refractory ischemia, recurrent angina, positive cardiac-specific enzymes, new ST-segment depression, low ejection fraction, severe arrhythmias, or a recent PCI or CABG. PCI is preferred over surgical therapy in most high-risk patients with acute coronary syndromes unless they have severe multivessel disease or the culprit lesion responsible for the unstable presentation cannot be adequately treated. In STEMI, thrombolysis or PCI (primary PCI) are effective methods to restore coronary blood flow and salvage myocardium within the first 12 hours after onset of chest pain. Because PCI is more effective than thrombolysis, it is preferred if readily available. PCI is also performed following thrombolysis to facilitate adequate reperfusion or as a rescue procedure in those who do not achieve reperfusion from thrombolysis or in those who develop cardiogenic shock.

Other Interventional Techniques

STRUCTURAL HEART DISEASE

Interventional treatment for structural heart disease (adult congenital heart disease and valvular heart disease) is a significant component of the field of interventional cardiology.

The most common adult congenital lesion to be treated with percutaneous techniques is closure of atrial septal defects (Chap. 236). The procedure is done as in a diagnostic right heart catheterization with the passage of a catheter up the femoral vein into the right atrium. With echo and fluoroscopic guidance the size and location of the defect can be accurately defined, and closure is accomplished using one of several approved devices. All devices use a left atrial and right atrial wire mesh or covered disk that are pulled together to capture the atrial septum around the defect and seal it off. The Amplatzer Septal Occluder device (AGA Medical, Minneapolis, Minnesota) is the most commonly used in the United States. The success rate in selected patients is 85–95%, and the device complications are rare and include device embolization, infection, or erosion. Closure of patent foramen ovale (PFO) is done in a similar way. PFO closure is an approved procedure in patients who have had recurrent paradoxical stroke despite adequate medical therapy including anticoagulation. The use in the treatment of migraine is under clinical investigation and is not an approved indication.

Similar devices can also be used to close patent ductus arteriosus and ventricular septal defects. Other congenital diseases that can be treated percutaneously include coarctation of the aorta, pulmonic stenosis, peripheral pulmonary stenosis, and other abnormal communications between the cardiac chambers or vessels.

The treatment of valvular heart disease is the most rapidly growing area in interventional cardiology. Until recently the only available techniques were balloon valvuoplasty for the treatment of aortic, mitral, or pulmonic stenosis (Chap. 237). Mitral valvuloplasty is the preferred treatment for symptomatic patients with rheumatic mitral stenosis who have favorable anatomy. The outcome in these patients is equal to that of surgical commissurotomy. The success is highly related to the echocardiographic appearance of the valve. The most favorable setting is commissural fusion without calcification or subchordal fusion and the absence of significant mitral regurgitation. Access is obtained from the femoral vein using a transseptal technique where a long metal catheter with a needle tip is advanced from the femoral vein through the right atrium and atrial septum at the level of the foramen ovale into the left atrium. A guidewire is advanced into the left ventricle, and a balloon-dilatation catheter is negotiated across the mitral valve and inflated to a predetermined size to enlarge the valve. The most commonly used dilatation catheter is the Inoue balloon. The technique splits the commissural fusion and commonly results in a doubling of the mitral valve area. The success of the procedure in favorable anatomy is 95% and severe complications are rare (1–2%). The most common complications are tamponade due to puncture into the pericardium and the creation of severe mitral regurgitation.

Similarly, severe aortic stenosis can be treated with balloon valvuloplasty. In this setting, the valvuloplasty balloon catheter is placed retrograde across the aortic valve from the femoral artery and briefly inflated to stretch open the valve. The success is much less favorable with an initial success rate of only 50% and a restenosis rate of 50% after 6 to 12 months. This poor success rate has limited its use to patients who are not surgical candidates or as a bridge to surgery in patients who are expected to improve sufficiently to become surgical candidates. In this setting, the mortality rate of the procedure is high (10%). Repeat aortic valvuloplasty as a treatment for aortic valve restenosis has been reported.

Percutaneous aortic valve replacement has been introduced to treat patients who are not suitable candidates for surgical aortic valve replacement. Currently, two valve models, the Edwards SAPIEN

valve (Edwards Lifescience, Irvine, California) and the CoreValve ReValving system (CoreValve Inc., Irvine, California), have been approved for use in Europe. In more than 4000 cases worldwide, follow-up shows no evidence of restenosis or prosthetic valve dysfunction in the midterm. Both are placed either retrograde from the femoral artery or can be placed via the left ventricular apex following surgical exposure. The CoreValve is self-expanding, while the Edwards valve is balloon expanded. Following balloon valvuloplasty the valve is positioned across the valve and deployed with post-deployment balloon inflation to ensure full contact with the aortic annulus. The success rate is 80–90% and the 30-day mortality rate is 10–15%, not unexpectedly as only high-risk patients are undergoing the procedure currently. Both valves are undergoing clinical testing in the United States.

Peripheral Arterial Interventions

The use of percutaneous interventions to treat symptomatic patients with arterial obstruction in the carotid, renal, aortic, and peripheral vessels is also part of the field of interventional cardiology. Randomized clinical trial data already support the use of carotid stenting in patients at high risk of complications from carotid endarterectomy (**Fig. 246-5**). Ongoing trials will determine whether carotid stenting should be used even more broadly. The success rate of peripheral interventional procedures has been improving, including for long segments of occlusive disease historically treated by peripheral bypass surgery (**Fig. 246-6**). Peripheral intervention is increasingly part of the training of an interventional cardiologist, and most programs now require an additional year of training after the interventional cardiology training year. The techniques and outcomes are described in detail in the chapter on peripheral vascular disease (Chap. 249).

Figure 246-5

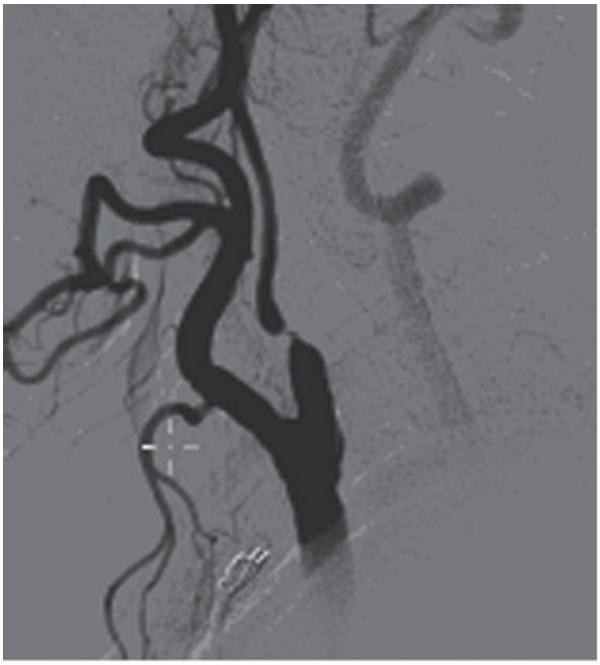

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An example of a high-risk patient who requires carotid revascularization, but who is not a candidate for carotid endarterectomy. Carotid artery stenting resulted in an excellent angiographic result. (*From M Belkin, DL Bhatt: Circulation 119:2302, 2009; with permission*.)

Figure 246-6

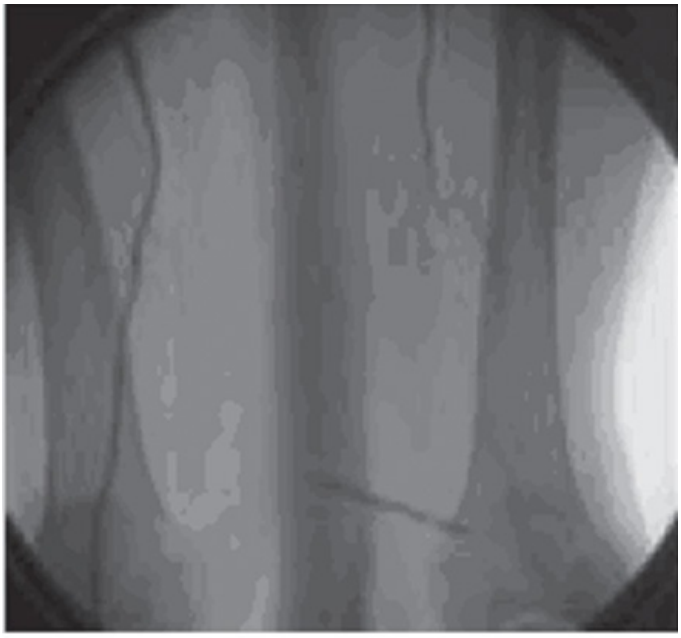

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

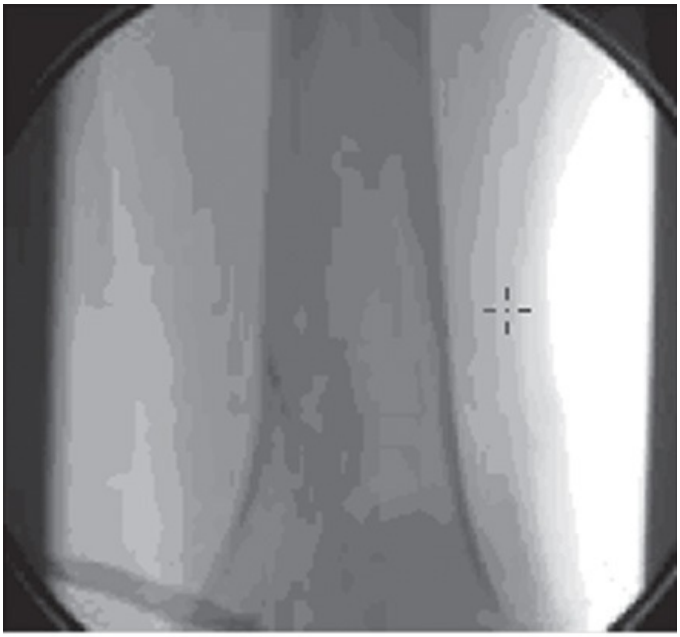

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

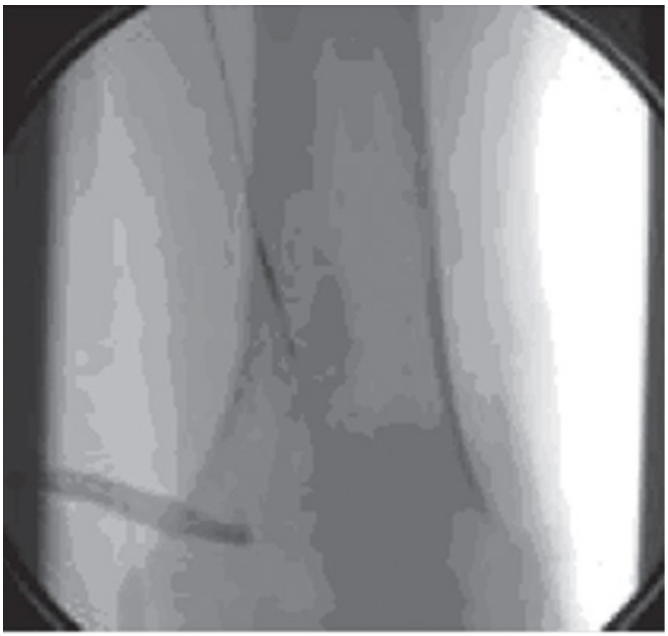

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

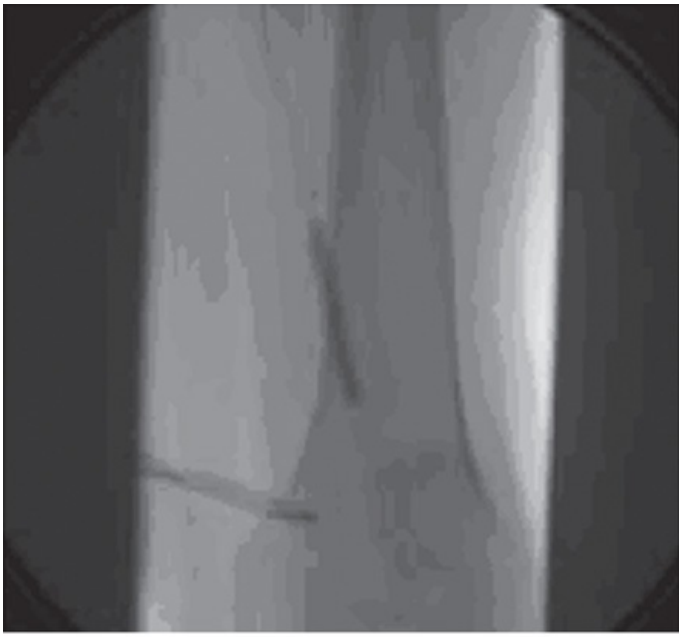

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

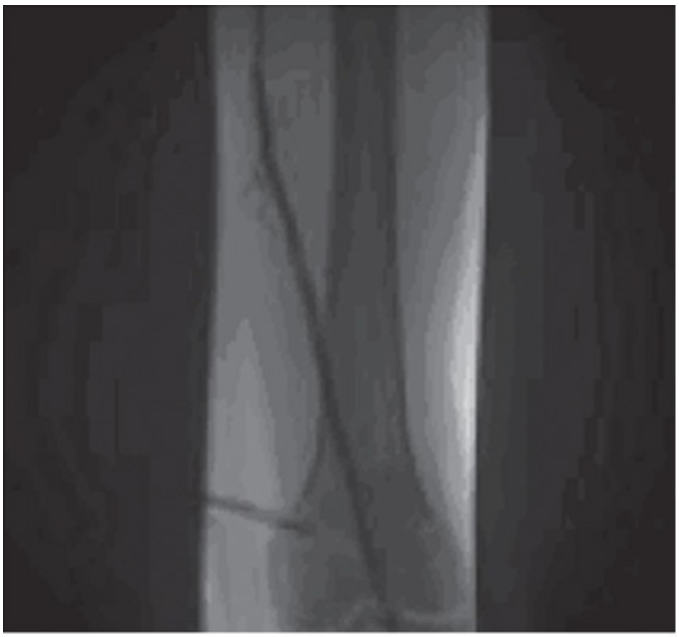

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Peripheral interventional procedures have become highly effective at treating anatomic lesions previously amenable only to bypass surgery.*A.* Complete occlusion of the left superficial femoral artery. ***B.*** Wire and catheter advanced into subintimal space. ***C.*** Intravascular ultrasound positioned in the subintimal space to guide retrograde wire placement through the occluded vessel. ***D.*** Balloon dilation of the occlusion. ***E.*** Stent placement with excellent angiographic result. (*From A Al Mahameed, DL Bhatt: Cleve Clin J Med 73:S45, 2006; with permission.*)

CIRCULATORY SUPPORT TECHNIQUES

The use of circulatory support techniques is occasionally needed in order to safely perform PCI on hemodynamically unstable patients. It also can be useful in helping to stabilize patients before surgical interventions. The most commonly used device is the percutaneous intraaortic balloon developed in the early 1960s. A 7-10 French 25- to 40-mL balloon catheter is placed retrograde from the femoral artery into the descending aorta between the aortic arch and the abdominal aortic bifurcation. It is connected to a helium gas inflation system that synchronizes the inflation to coincide with early

diastole with deflation by mid-diastole. As a result, it increases early diastolic pressure, lowers systolic pressure, and lowers late diastolic pressure through displacement of blood from the descending aorta (counterpulsation). This results in an increase in coronary blood flow and a decrease in afterload. It is contraindicated in patients with aortic regurgitation, aortic dissection, or severe peripheral vascular disease. The major complications are vascular and thrombotic. Intravenous heparin is given in order to reduce thrombotic complications.

Another useful tool is the Impella device (Abiomed, Danvers, Massachusetts). The catheter is placed percutaneously from the femoral artery into the left ventricle. The catheter has a small microaxial pump at its tip that can pump up to 2.5 liters per minute at a speed of 50,000 rpm from the left ventricle to the aorta. Other support devices include the hemopump and percutaneous cardiopulmonary bypass.

CONCLUSIONS

Interventional cardiology continues to expand its borders. Treatment for coronary artery disease, including complex anatomic subsets, continues to advance, encroaching on what has traditionally been treated by CABG. Technological advances such as drug-eluting stents, now already in their second generation, and manual aspiration devices are improving the results of PCI. In particular, the data for PCI preventing future ischemic events in unstable ischemic syndromes are substantial. For patients with stable coronary disease, PCI has an important role in symptom alleviation. Treatment of peripheral and cerebrovascular disease has also benefited from the application of percutaneous techniques. Structural heart disease is increasingly being treated with percutaneous options, with a high likelihood that interventional approaches will supplant open-heart surgery in a significant proportion of cases in years to come.

FURTHER READINGS

Baim DS: Percutaneous balloon angioplasty and general coronary intervention, in *Cardiac Catheterization, Angiography, and Intervention*, 7th ed, DS Baim (ed). Philadelphia, Lippincott Williams & Wilkins, 2006

BARI 2D Study Group: A randomized trial of therapies for type 2 diabetes and coronary artery disease. N Engl J Med 360:2503, 2009

Boden WE et al: Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med 356:1503, 2007[PMID: 17387127]

Hirsch AT et al: ACC/AHA 2005 practice guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric and abdominal aortic): Executive summary. Circulation 113:3463, 2006

King SB 3rd et al: 2007 Focused update of the ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines: 2007 writing group to review new evidence and update the ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention, writing on behalf of the 2005 writing committee. Circulation 117:261, 2008[PMID: 18079354]

Kushner FG et al: 2009 focused updates: ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction (updating the 2004 guideline and 2007 focused update) and ACC/AHA/SCAI guidelines on percutaneous coronary intervention (updating the 2005 guideline and 2007 focused update) a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 54:2205, 2009)

Serruys PW et al: Percutaneous coronary intervention versus coronary artery bypass grafting for severe coronary artery disease. N Engl J Med 360:961, 2009[PMID: 19228612]

Smith SC et al: ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention —summary article. A report of the American College of Cardiology/American Heart Association task force on practice guidelines (ACC/AHA/SCAI writing committee to update the 2001 guidelines for percutaneous coronary intervention). Circulation 113:156, 2006[PMID: 16391169]

White CJ: Catheter-based therapy for atherosclerotic renal artery stenosis. Circulation 113:1464, 2006[PMID: 16549651]

Yadav JS et al: Protected carotid-artery stenting versus endarterectomy in high-risk patients. N Eng J Med 351:1493, 2004[PMID: 15470212]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF PERCUTANEOUS REVASCULARIZATION: INTRODUCTION

Percutaneous coronary intervention (PCI) is the most widely employed coronary revascularization procedure worldwide (Chap. 246). It is now applied to patients with stable angina, acute coronary syndromes, including unstable angina and non-ST-segment elevation myocardial infarction (NSTEMI), and as a primary treatment strategy in patients with ST-segment elevation myocardial infarction (STEMI). PCI is also applicable to patients with either single- or multi-vessel disease.

In this chapter, the use of PCI will be illustrated in a variety of commonly encountered clinical and anatomic situations such as chronic total occlusion of a coronary artery, bifurcation disease, acute STEMI, saphenous vein graft disease, left main coronary artery disease, multivessel disease, and stent thrombosis. In addition, the use of interventional techniques to treat structural heart disease will be shown, including closure of an atrial septal defect (ASD) and percutaneous aortic valve implantation; the latter is approved in Europe but is under active investigation in clinical trials in the United States and not yet approved for use.

CASE 1: CHRONIC TOTAL OCCLUSION

(Videos e33-1 to e33-7)

- An 81-year-old man with angina, NYHA Class IV congestive heart failure and inferior-apical-posterior ischemia on an exercise technetium-99m scan.
- Diagnostic cardiac catheterization revealed a left dominant system with a totally occluded left circumflex (LCx) artery. The distal LCx filled via collaterals from the left anterior descending (LAD) artery, indicating chronicity of the total occlusion.

Video e33-1
Baseline left coronary angiogram shows an occluded LCx with left-to-left collaterals originating from LAD septal vessels.
Video e33-2
Attempts to cross the total occlusion in the LCx using a hydrophilic wire and an antegrade approach were not successful, with the wire tracking to the right of the trajectory.
Video e33-3
The LAD septal collateral is accessed with a guidewire and directed toward the distal LCx to cross the total occlusion retrograde.
Video e33-4
The total occlusion is crossed retrograde. The wire is snared in the guide, exteriorized, and used to provide antegrade access to the LCx.
Video e33-5
Antegrade flow in the LCx is restored after balloon inflation.
Video e33-6
Following stenting of the total occlusion, blood flow in the distal vessel is improved and a second significant stenosis is seen.
Video e33-7
Final result after LCx stenting.

SUMMARY

- Approximately 15–30% of all patients referred for cardiac catheterization will have a chronic

total occlusion (CTO) of a coronary artery.

- CTO often leads to a surgical referral for complete revascularization.
- Incomplete revascularization due to an untreated CTO is associated with an increased mortality rate (Hazard Ratio = 1.36, 95% CI = 1.12–1.66, $p < 0.05$).
- Successful PCI of a CTO leads to a 3.8–8.4% absolute reduction in mortality, symptom relief, and improved left ventricular function.
- Newer techniques, such as the retrograde approach to crossing total occlusions, are useful when the antegrade approach fails, or is not feasible, and there are well-developed collateral vessels.

(*Case contributed with permission by Dr. Frederick G. P. Welt)*

CASE 2: BIFURCATION STENTING

(Fig. e33-1; Videos e33-8 to e33-16)

- A 52-year-old man with an acute coronary syndrome and a troponin I = 0.18 (upper limit normal ≤ 0.04).
- Diagnostic cardiac catheterization showed single-vessel coronary artery disease with a significant stenosis in the mid-LAD and a bifurcation lesion involving a large diagonal branch.

Figure e33-1

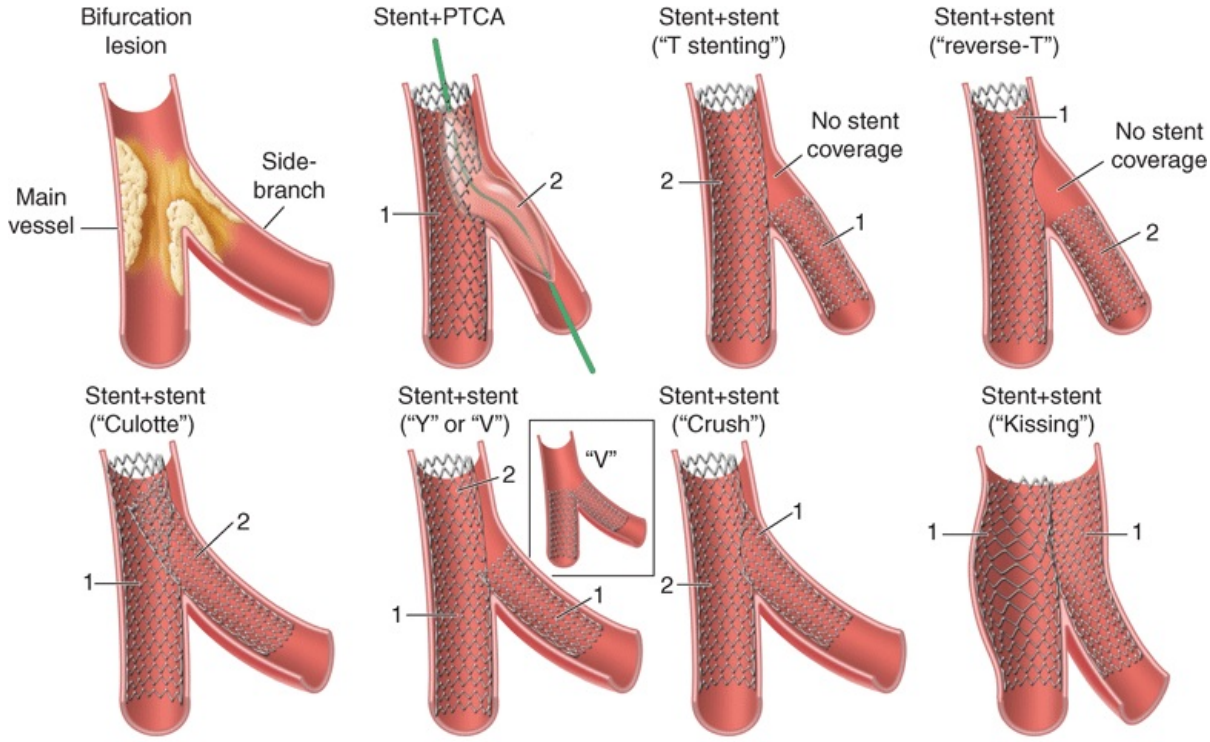

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of 1-stent and 2-stent techniques to treat bifurcation lesions. *(Reprinted with permission from SK Sharma and A Kini: Cardiol Clin 24:233, 2006.)*

Video e33-8

Baseline angiogram of the left coronary circulation shows the significant stenosis in the mid-LAD and the bifurcation lesion involving a large diagonal branch.

Video e33-9

Both vessels are accessed with guidewires and pretreated with balloon angioplasty.

Video e33-10

Result after balloon angioplasty.

Video e33-11

Stent being positioned in the LAD.

Video e33-12

LAD post-stent result.

Video e33-13

Stent deployed in diagonal branch through the stent struts in the LAD using the "culotte" technique.

Video e33-14

Diagonal branch post-stent result.

Video e33-15

Simultaneous inflation of two 2.5-mm "kissing" balloons.

Video e33-16

Final postbifurcation stenting result.

SUMMARY

- Approximately 15–20% of PCIs will involve the treatment of bifurcation lesions.
- Bifurcation lesions require consideration of PCI strategies that protect side-branch patency.
- There are both one-stent and two-stent techniques to treat bifurcation lesions; the selection of technique depends upon anatomic considerations, including plaque burden, angle of side-branch take-off, plaque shift during angioplasty, and side-branch distribution.
- Rates of target lesion revascularization and stent thrombosis are similar between one-stent and two-stent procedures.

CASE 3. INFERIOR MYOCARDIAL INFARCTION—THROMBUS AND MANUAL THROMBECTOMY

(Figs. e33-2 to e33-4; Videos e33-17 to e33-22)

- A 59-year-old man presented to the emergency room with 2 h of severe midsternal chest pressure.
- His systolic blood pressure was 100 mmHg and he was tachycardic in sinus rhythm with a heart rate of 90–100 bpm.
- His initial ECG showed inferior ST-segment elevations with lateral ST-segment depressions.
- He was referred emergently to the cardiac catheterization laboratory for primary PCI.

Figure e33-2

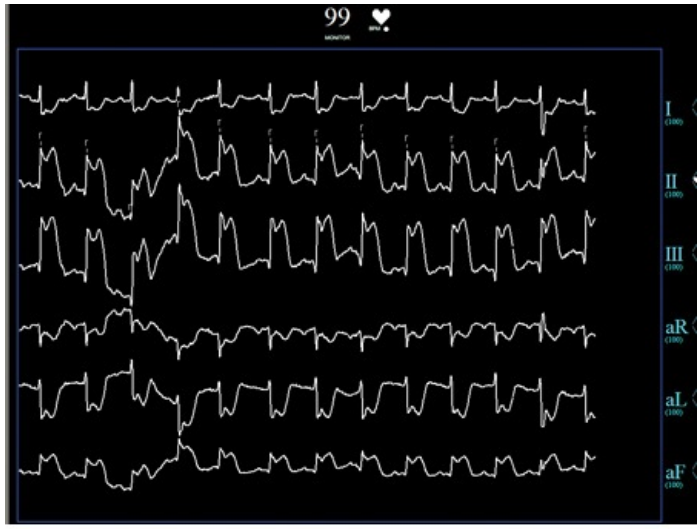

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pre-procedure ECG showing inferior ST-segment elevations and lateral ST-segment depressions.

Figure e33-3

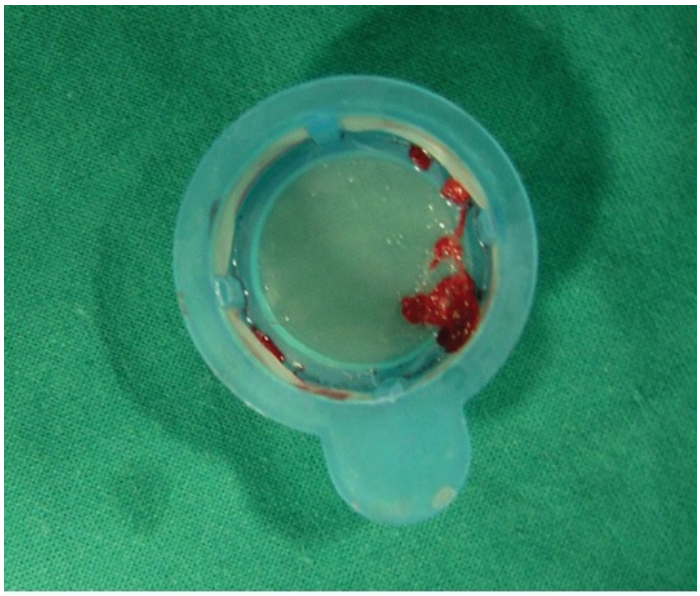

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Example of an organized red thrombus retrieved by manual thrombectomy. *(Reprinted with permission from C Trani et al: J Invasive Cardiol 19:E317, 2007.)*

Figure e33-4

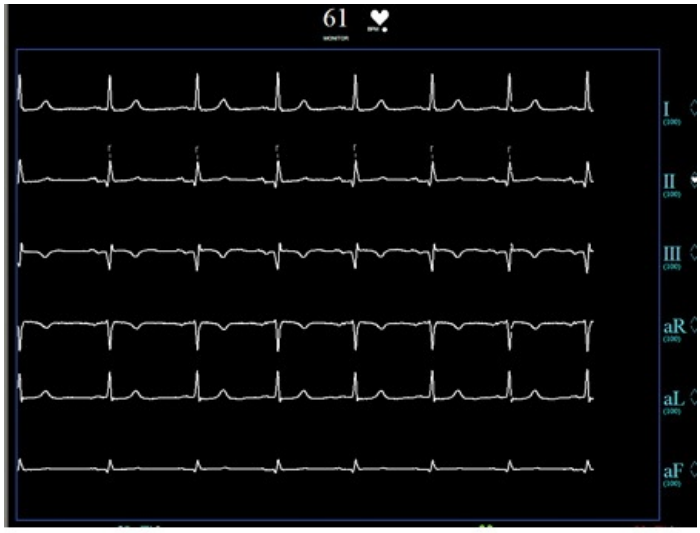

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Post-procedure ECG showing resolution of ST-segment elevations.

Video e33-17

The right coronary artery (RCA) is totally occluded with filling defects in the vessel after contrast injection, indicating thrombus is present in the vessel.

Video e33-18

An angioplasty wire is threaded through the thrombotic lesion, but this does not restore blood flow to the distal vessel.

Video e33-19

Result after manual thrombectomy and thrombus extraction. The "culprit" ruptured plaque and residual thrombus are now apparent in the vessel.

Video e33-20

After balloon angioplasty and stenting, thrombus is still present.

Video e33-21

After repeat manual thrombectomy and expansion of the stent, the thrombus is no longer present.

Video e33-22

Final result.

SUMMARY

- An acute STEMI occurs following plaque rupture that promotes thrombotic occlusion of a coronary artery.
- Despite successful revascularization of the epicardial coronary artery, microemboli liberated during balloon angioplasty and stenting may lead to persistent microvascular dysfunction. When present, microvascular dysfunction is associated with a larger infarct size, heart failure, malignant ventricular arrhythmias, and death.
- Manual thrombectomy is used to aspirate or remove thrombus in the vessel and limit distal embolization during angioplasty and stenting.
- Manual thrombectomy in primary PCI is associated with improved myocardial perfusion and a reduction in mortality.
- Adjunctive antiplatelet and antithrombin agents are important to aid in the resolution of intracoronary thrombus.

CASE 4. SAPHENOUS VEIN GRAFT INTERVENTION WITH DISTAL PROTECTION

(Fig. e33-5; Videos e33-23 to e33-26)

- A 62-year-old man with a history of chronic stable angina.
- A four-vessel coronary artery bypass grafting (CABG) surgery was performed 17 years earlier with a left internal mammary artery graft to the LAD, a right internal mammary artery graft to the RCA, a saphenous vein graft to the first obtuse marginal branch, and a saphenous vein graft to the first diagonal branch.
- The patient had a recent increase in angina with exertion and was found to have lateral ischemia on an exercise technetium 99m scan.
- Diagnostic cardiac catheterization revealed a significant stenosis in the body of the saphenous vein graft to the first obtuse marginal branch.

Figure e33-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Distal protection device showing captured atherosclerotic debris liberated by initial balloon dilation. *(Reprinted with permission from RA Aqel et al: J Invasive Cardiol 19:E104, 2007.)*

Video e33-23

Saphenous vein graft to a first obtuse marginal branch with an 80% eccentric stenosis in the midgraft.

Video e33-24

A distal protection device is deployed past the lesion.

Video e33-25

Angioplasty balloon inflation with the distal protection device in place.

Video e33-26

Final result after stent placement.

SUMMARY

- Saphenous vein grafts have a failure rate of up to 20% after 1 year and as high as 50% by 5 years.
- Graft failure (>1 month) results from intimal hyperplasia and atherosclerosis.
- Saphenous vein graft PCI is associated with distal embolization of atherosclerotic debris and microthrombi, leading to microvascular occlusion, reduced antegrade blood flow (the "no-reflow" phenomenon), and myocardial infarction.
- Embolic distal protection devices decrease the risk of distal embolization, as well as the incidence of no-reflow and myocardial infarction associated with saphenous vein graft interventions.

CASE 5: UNPROTECTED LEFT MAIN PCI IN A HIGH-RISK PATIENT

(Figs. e33-6 and e33-7; Videos e33-27 to e33-34)

- An 89-year-old woman presented with a NSTEMI associated with 5-mm ST-segment depression in the apical leads occurring 2 weeks after hospitalization for an NSTEMI that was treated conservatively.
- Chronic obstructive lung disease, elderly age, and the patient's refusal to consider cardiac surgery restricted the choice of therapeutic options to medical and/or percutaneous interventions.
- Diagnostic catheterization revealed a left dominant circulation with a heavily calcified 80% distal left main coronary artery stenosis extending into the LAD and into the proximal LCx

coronary arteries. A 70% proximal LAD lesion was also present.

- After consultation with the patient, family, and a cardiac surgeon, PCI was performed with intraaortic balloon pump support and a temporary pacemaker in the right ventricle.

Figure e33-6

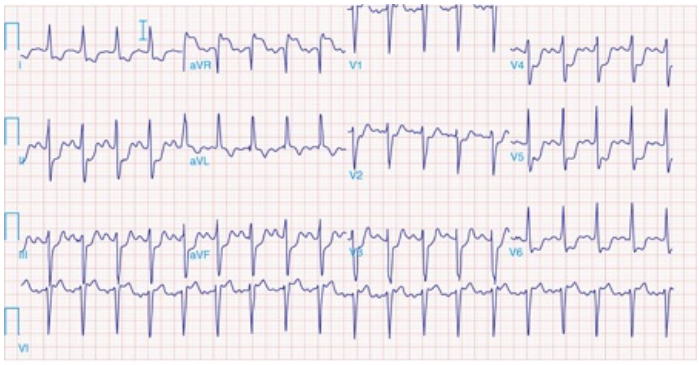

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

During chest pain, the ECG showed diffuse ST-segment depression of up to 5 mm in the inferior and lateral leads.

Figure e33-7

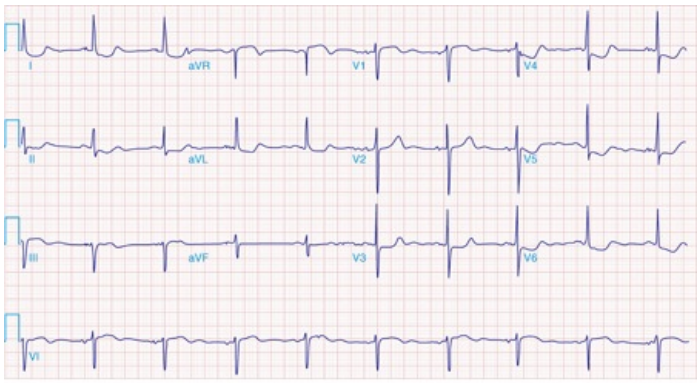

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Following resolution of the chest pain, the ST-segment depression is less marked.

Video e33-27

Baseline left coronary artery injection in right anterior oblique (RAO) cranial projection shows a high-grade calcified stenosis in the left main coronary artery and a significant stenosis in the proximal LAD.

Video e33-28

In the left anterior oblique (LAO) caudal view, the left main coronary artery lesion can be seen to extend into the ostia of both the LCx and the LAD.

Video e33-29

Guide wires were placed into both the LCx and LAD. After the left main coronary artery and LCx are dilated with balloon angioplasty, the proximal LAD is dilated and a long drug-eluting stent is placed to cover a lesion dissection that occurred with wiring of the vessel.

Video e33-30

The bifurcation lesion in the left main coronary artery extending into the LCx and LAD ostia is treated using a "culotte" technique. First, a drug-eluting stent is placed in the left main coronary artery and into the proximal LCx.

Video e33-31

Next, the LAD wire is removed and passed through the stent into the distal LAD. A second drug-eluting stent is deployed through the struts of the left main coronary artery/LCx stent.

Video e33-32

Following rewiring of the LCx, both stents are redilated simultaneously ("kissing" balloons).

Video e33-33

The final result in the LAO caudal view.

Video e33-34

The final result in the RAO cranial view showing patent left main, LCx, and LAD coronary arteries.

SUMMARY

- Left main coronary artery disease occurs in 5–10% of patients with symptomatic coronary artery disease.
- In patients with left main coronary artery disease, revascularization with CABG has been shown to decrease mortality significantly over 5–10 years of follow-up.
- PCI with drug-eluting stents in selected cases has been shown to have equal in-hospital and 1-year death and myocardial infarction rates as compared to CABG in the Synergy between PCI with Taxus and Cardiac Surgery (SYNTAX) trial. Long-term outcome differences between the two treatment strategies are not known.
- Indications for PCI of left main coronary artery lesions are high-risk surgical patients and patients with protected left main coronary artery disease (i.e., prior CABG with patent bypass grafts). Patients who are good candidates for bypass surgery may also undergo a stenting procedure, but discussion with the patient, the interventional cardiologist, and the cardiac surgeon should be undertaken to determine the best treatment option in an individual case.
- Outcomes are better for patients with an isolated lesion in the ostium and body of the left main coronary artery where a single stent can be placed compared to bifurcation lesions that involve the ostium of the LAD and LCx.

CASE 6: MULTIVESSEL PCI IN A DIABETIC PATIENT

(Videos e33-35 to e33-42)

- A 58-year-old man presented with an NSTEMI.
- The patient has hyperlipidemia and type 2 diabetes mellitus treated with oral hypoglycemic agents.
- Diagnostic catheterization revealed two-vessel disease with a total occlusion of the second obtuse marginal branch that was felt to be responsible for the patient's symptoms (culprit lesion).In addition, there was a high-grade stenosis in a large ramus intermedius branch and the RCA had a significant stenosis in the midsegment of the vessel.

Video e33-35

Baseline angiogram of the left coronary circulation in the RAO view shows the total occlusion of the second obtuse marginal branch with delayed retrograde filling via collateral vessels and a high-grade stenosis in the ramus intermedius.

Video e33-36

A guidewire is passed through the total occlusion and the lesion is pretreated with balloon angioplasty.

Video e33-37

Following placement of a drug-eluting stent in the lesion, the vessel is widely patent. A third obtuse marginal vessel, not previously seen, now fills faintly (TIMI 1 flow) with contrast but was not treated.

Video e33-38

The ramus intermedius lesion was crossed with a guidewire and pretreated with balloon angioplasty.

Video e33-39

A drug-eluting stent is placed across the ramus lesion and deployed. The final result shows no residual stenosis in either the ramus or second obtuse marginal vessels.

Video e33-40

Baseline angiogram of the RCA shows a high-grade lesion in the midsegment of the vessel.

Video e33-41

The lesion was pretreated with balloon dilation followed by stent deployment.

Video e33-42

The final result shows no residual stenosis in the mid-RCA.

SUMMARY

- Multivessel PCI is performed commonly and may be done in one setting or staged with two or more procedures.
- Acute and long-term studies of multivessel PCI have shown comparable rates of death and myocardial infarction when compared to CABG but a higher incidence of repeat revascularization as a result of restenosis associated with PCI.
- In the randomized Bypass Angioplasty Revascularization Investigation (BARI) trial, diabetic patients treated with PCI had a worse long-term mortality than diabetic patients treated with CABG. However, the BARI registry found that in selected diabetic patients with favorable anatomy, PCI can result in outcomes equal to that observed with CABG.

CASE 7: VERY LATE STENT THROMBOSIS OF A PROXIMAL LAD DRUG-ELUTING STENT

(Figs. e33-8 and e33-9; Videos e33-43 to e33-46)

- A 62-year-old man had a drug-eluting stent placed in a proximal LAD lesion to treat severe angina. He received dual antiplatelet therapy with aspirin and clopidogrel for 1 year and then discontinued clopidogrel per protocol.
- He remained asymptomatic until 15 months after the initial stent placement, when he presented with severe chest pain due to an acute anterior STEMI.
- He was taken to the catheterization laboratory within 70 min of presentation and his initial angiogram showed a total occlusion of the proximal LAD stent.

Figure e33-8

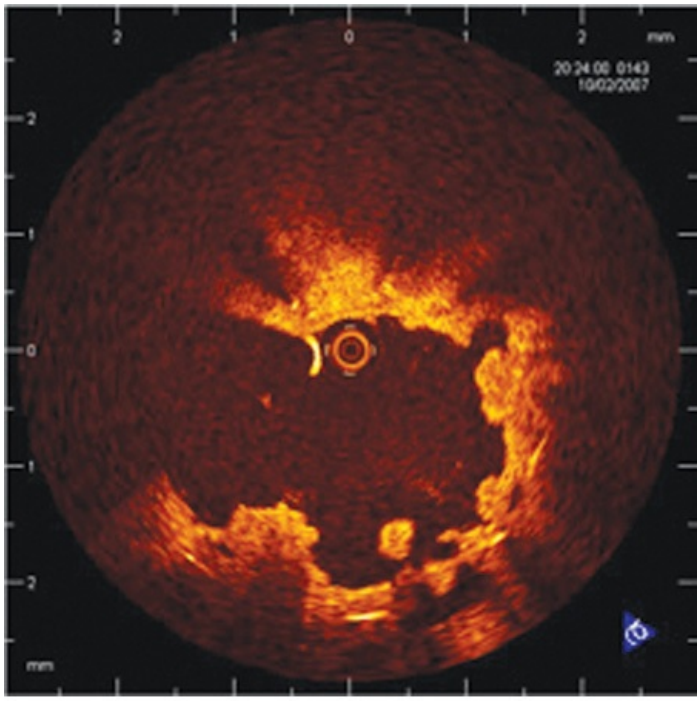

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Optical coherence tomography image following initial balloon dilation. Residual thrombus that is

adherent to the stent struts is seen. *(Reprinted with permission from AF Schinkel et al: JACC Cardiovasc Interv 1:449, 2008.)*

Figure e33-9

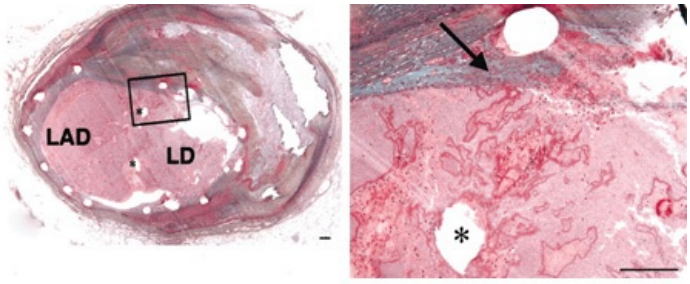

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathologic specimen of late stent thrombosis obtained at autopsy. Thrombus is seen filling the LAD vessel lumen and extending into a diagonal branch (LD). Stent struts occupied the space denoted by asterisk (*) (*left*). A magnified view of the vessel reveals thrombus around the stent strut and neointima formation (*arrow*) (*right*). *(Reprinted with permission from A Farb et al: Circulation 108:1701, 2003.)*

Video e33-43

Baseline angiogram showing a total occlusion of the proximal LAD within the drug-eluting stent and a significant stenosis at the origin of the LCx.

Video e33-44

The LAO view shows the LCx stenosis with a filling defect, indicating that thrombus is present in the vessel lumen.

Video e33-45

The LAD lesion was crossed with a guidewire, which resulted in slow filling of the mid-LAD (TIMI 2 flow), and revealed thrombus filling the stent.

Video e33-46

The final result after LAD and LCx stenting. The LAD lesion was pretreated with balloon angioplasty and a bare metal stent was deployed to cover the proximal lesion. The LCx ostial lesion was dilated with balloon angioplasty and a bare metal stent was placed using a "V stenting" technique.

SUMMARY

- Stent thrombosis is an infrequent (1–2%) but serious complication of stent placement. It occurs most commonly within the first month, but rarely, as late as 1 year (0.2–0.6%) with bare metal stents. Very late stent thrombosis (VLST), which occurs after 1 year, is very rare with bare metal stents but can occur with drug-eluting stents.
- Premature discontinuation of dual antiplatelet therapy is the most common cause of early and late stent thrombosis; however, the etiology of VLST is not clear.
- The majority of patients with stent thrombosis present with acute coronary syndromes or STEMI; this presentation is associated with a high mortality rate (10%).
- Treatment is immediate PCI with balloon angioplasty or restenting.

CASE 8: TRANSCATHETER AORTIC VALVE IMPLANTATION

(Figs. e33-10 to e33-14; Videos e33-47 to e33-50)

- A 75-year-old woman with symptomatic aortic stenosis and a valve area of 0.58 cm^2 by transthoracic echocardiogram.
- Chronic obstructive pulmonary disease (FEV1 = 0.54) and other comorbidities contributed to an unacceptably high cardiac surgical risk (calculated logistic Euroscore = 29.57%) for aortic valve replacement.
- She was referred for transcatheter aortic valve implantation (TAVI) as part of a clinical trial.

Figure e33-10

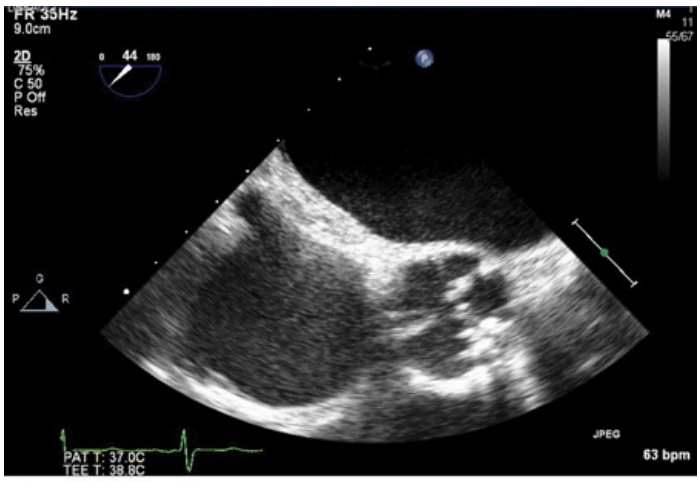

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

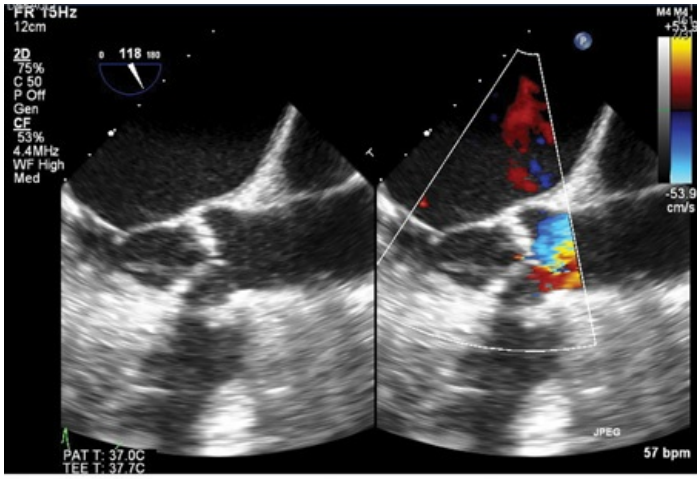

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transesophageal echocardiogram shows a calcified trileaflet aortic valve (*left*) with reduced leaflet excursion and a narrowed orifice in peak systole (*right*) .

Figure e33-11

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hemodynamically significant aortic (AO) stenosis. Simultaneous recording of AO and left ventricle (LV) pressures shows an 82 mmHg peak-to-peak gradient and a 63.3 mmHg mean gradient between the LV (154/9 mmHg) and AO (72/29 mmHg) pressures. This is consistent with an aortic valve area of 0.58 cm^2.

Figure e33-12

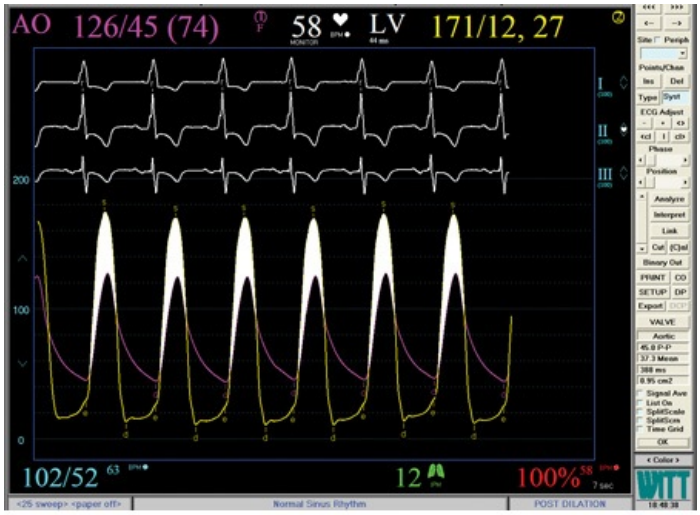

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

After balloon valvuloplasty, the LV–AO mean pressure gradient decreased to 37.3 mmHg, indicating that the aortic valve area increased to 0.95 cm^2.

Figure e33-13

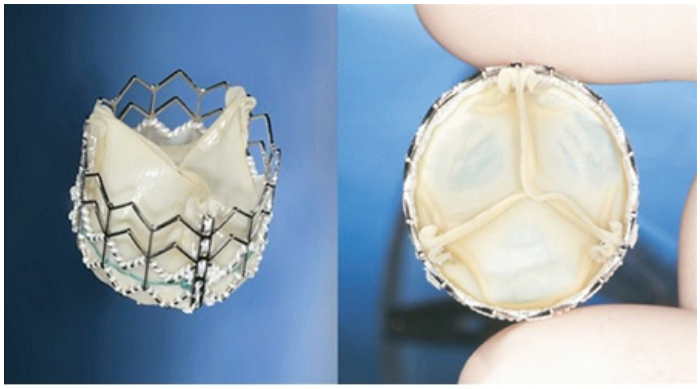

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The Edwards SAPIEN Transcatheter Heart Valve. *(Reprinted with permission from A Zajarias and A Cribier: JACC 53(20): 1829, 2009.)*

Figure e33-14

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Once the valve was deployed, the pressure gradient between the LV and AO decreased to 11.6 mmHg and the functional valve area is 1.34 cm^2.

Video e33-47

Aortogram shows patent coronary arteries and minimal aortic insufficiency.

Video e33-48

Balloon valvuloplasty is performed with rapid ventricular pacing at 180 bpm.

Video e33-49

A 26-mm Edwards-SAPIEN valve is positioned using fluoroscopic and transesophageal echo guidance and deployed.

Video e33-50

Aortogram after valve deployment shows a functional valve with mild aortic insufficiency and without impingement of the coronary ostia.

SUMMARY

- The prevalence of calcific aortic stenosis is 2–3% in individuals aged ≤75 years.
- Symptomatic aortic stenosis is associated with an average survival of 2–3 years and an

increased risk of sudden death; aortic valve replacement improves both symptoms and survival.

- In high-risk patients with severe aortic stenosis who are not surgical candidates, 1-year and 5-year survival rates are ~ 62% and 38%, respectively.
- Transcatheter aortic valve implantation (TAVI) is approved in Europe and being evaluated in the United States in clinical trials as an alternative to surgical aortic valve replacement in high-risk patients.

(Case contributed with permission by Dr. Andrew C. Eisenhauer)

CASE 9: ATRIAL SEPTAL DEFECT CLOSURE

(Figs. e33-15 to e33-19; Videos e33-51 to e33-53)

- A 48-year-old woman with increased shortness of breath, exercise intolerance, and an 18-mm secundum ASD.
- Echocardiogram showed a dilated right atrium (RA) and right ventricle (RV) with evidence of right ventricular volume overload.
- A shunt ratio (Qp/Qs) of 2.3:1 was determined at cardiac catheterization.
- Based on her symptoms, evidence of right-side chamber dilation, and a moderately sized atrial septal defect, the patient was referred for percutaneous closure of the ASD.

Figure e33-15

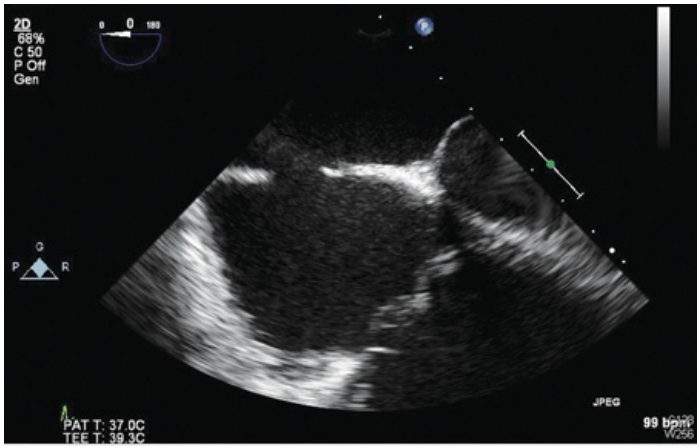

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

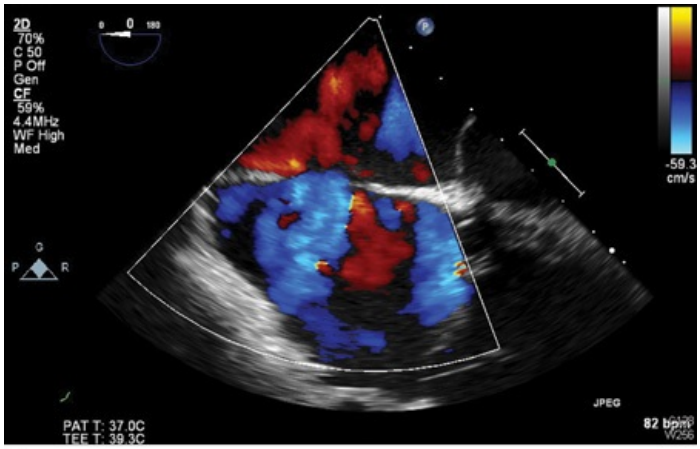

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transesophageal echocardiogram of a secundum ASD. The ASD is seen as "dropout" in the interatrial septum between the left atrium (LA) and RA (*left*). Doppler color flow imaging shows blue in the RA consistent with left-to-right flow (*right*).

Figure e33-16

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three-dimensional echocardiographic reconstruction of the secundum ASD. The ASD is round and has an acceptable margin of tissue to seat a septal occluder device.

Figure e33-17

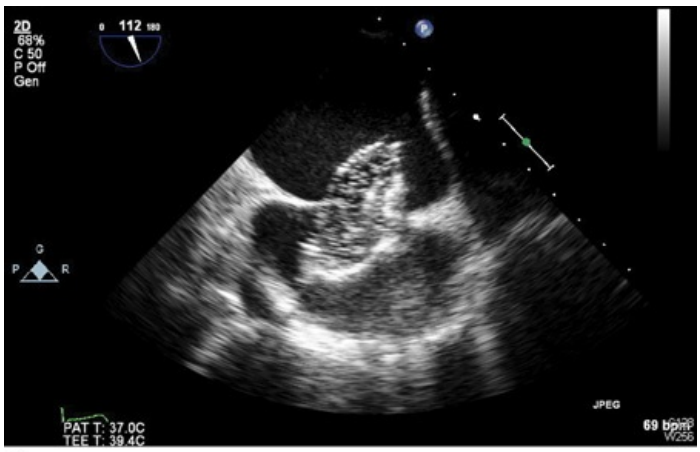

A

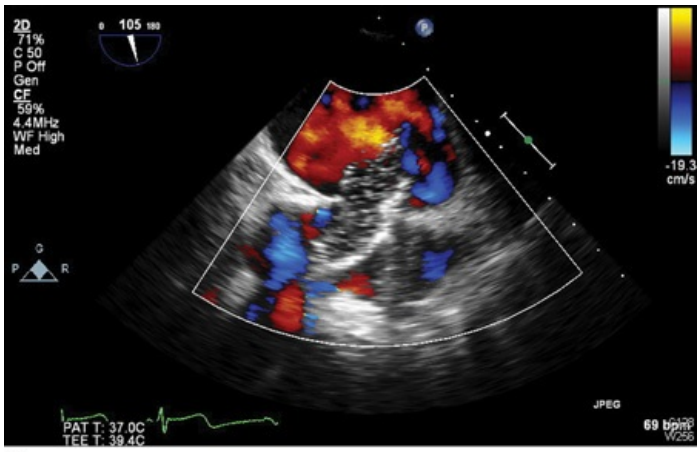

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transesophageal echocardiogram showing sizing balloon (*left*) and no flow (*right*) across the atrial

septal defect.

Figure e33-18

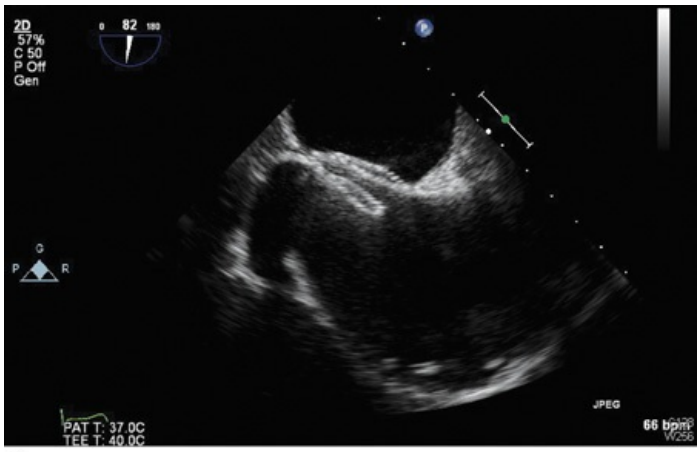

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

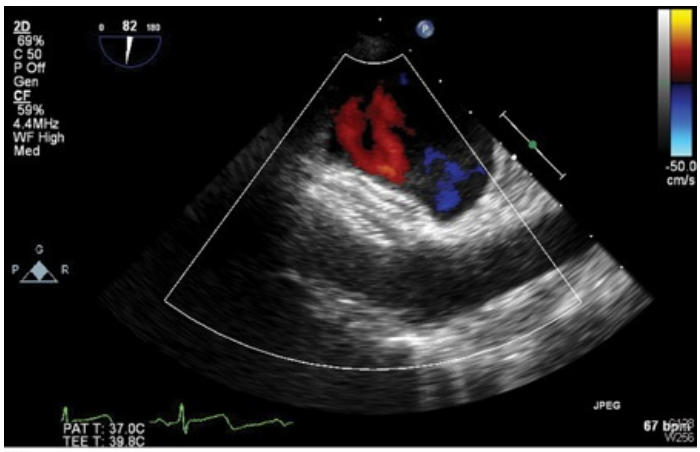

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Amplatzer septal occluder in place (*left*). There is no blood flow across the device (*right*).

Figure e33-19

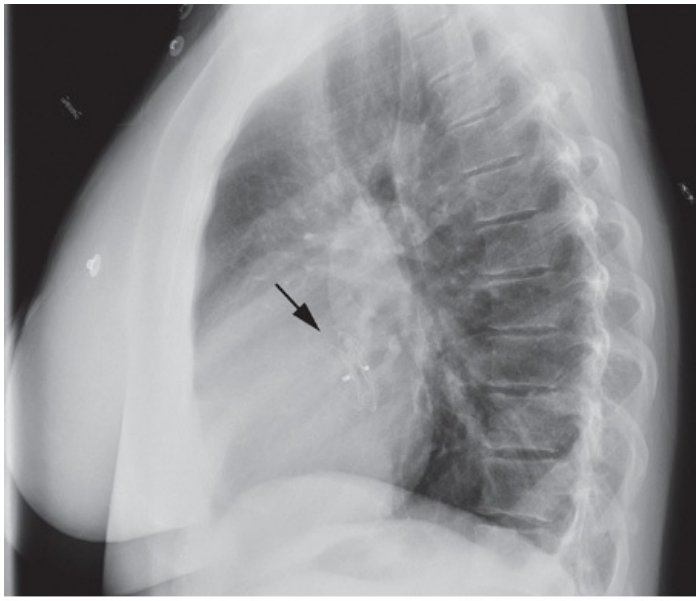

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postprocedure lateral chest x-ray showing the Amplatzer septal occluder in place.

Video e33-51

A sizing balloon is placed across the ASD.

Video e33-52

An Amplatzer septal occluder is being positioned across the ASD.

Video e33-53

The two discs of the device in place across the ASD.

SUMMARY

- Unrepaired ASDs lead to signs and symptoms of increased pulmonary blood flow and right heart failure; dyspnea, exercise intolerance, fatigue, palpitations and atrial arrhythmias, and pulmonary infections.
- Percutaneous closure of an ASD may be recommended for individuals with a secundum ASD and evidence of RA and RV enlargement with or without symptoms.
- Percutaneous closure is contraindicated in patients with irreversible pulmonary arterial hypertension and no left-to-right shunt. It is not recommended for closure of sinus venosus, coronary sinus, or primum ASDs.
- After the atrial septal occluder device is placed, patients are treated with antiplatelet agents and use antibiotic prophylaxis for certain procedures for 6 months. Follow-up echocardiograms to assess for device migration or erosion, residual shunting, thrombus or pericardial effusion are recommended at 1 day, 1 month, 6 months, 1 year, and periodically thereafter.

(Case contributed with permission by Dr. Andrew C. Eisenhauer)

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HYPERTENSIVE VASCULAR DISEASE: INTRODUCTION

Hypertension is one of the leading causes of the global burden of disease. Approximately 7.6 million deaths (13–15% of the total) and 92 million disability-adjusted life years worldwide were attributable to high blood pressure in 2001. Hypertension doubles the risk of cardiovascular diseases, including coronary heart disease (CHD), congestive heart failure (CHF), ischemic and hemorrhagic stroke, renal failure, and peripheral arterial disease. It often is associated with additional cardiovascular disease risk factors, and the risk of cardiovascular disease increases with the total burden of risk factors. Although antihypertensive therapy clearly reduces the risks of cardiovascular and renal disease, large segments of the hypertensive population are either untreated or inadequately treated.

EPIDEMIOLOGY

Blood pressure levels, the rate of age-related increases in blood pressure, and the prevalence of hypertension vary among countries and among subpopulations within a country. Hypertension is present in all populations except for a small number of individuals living in primitive, culturally isolated societies. In industrialized societies, blood pressure increases steadily during the first two decades of life. In children and adolescents, blood pressure is associated with growth and maturation. Blood pressure "tracks" over time in children and between adolescence and young adulthood. In the United States, average systolic blood pressure is higher for men than for women during early adulthood, although among older individuals the age-related rate of rise is steeper for women. Consequently, among individuals age 60 and older, systolic blood pressures of women are higher than those of men. Among adults, diastolic blood pressure also increases progressively with age until ~55 years, after which it tends to decrease. The consequence is a widening of pulse pressure (the difference between systolic and diastolic blood pressure) beyond age 60. The probability that a middle-aged or elderly individual will develop hypertension in his or her lifetime is 90%.

In the United States, based on results of the National Health and Nutrition Examination Survey (NHANES), approximately 30% (age-adjusted prevalence) of adults, or at least 65 million individuals, have hypertension (defined as any one of the following: systolic blood pressure ≥140 mmHg, diastolic blood pressure ≥90 mmHg, taking antihypertensive medications). Hypertension prevalence is 33.5% in non-Hispanic blacks, 28.9% in non-Hispanic whites, and 20.7% in Mexican Americans. The likelihood of hypertension increases with age, and among individuals age ≥60, the prevalence is 65.4%. Recent evidence suggests that the prevalence of hypertension in the United States may be increasing, possibly as a consequence of increasing obesity. The prevalence of hypertension and stroke mortality rates are higher in the southeastern United States than in other regions. In African Americans, hypertension appears earlier, is generally more severe, and results in higher rates of morbidity and mortality from stroke, left ventricular hypertrophy, CHF, and end-stage renal disease (ESRD) than in white Americans.

Both environmental and genetic factors may contribute to regional and racial variations in blood pressure and hypertensionprevalence. Studies of societies undergoing "acculturation" and studies of migrants from a less to a more urbanized setting indicate a profound environmental contribution to blood pressure. Obesity and weight gain are strong, independent risk factors for hypertension. It has been estimated that 60% of hypertensives are >20% overweight. Among populations, hypertension prevalence is related to dietary NaCl intake, and the age-related increase in blood pressure may be augmented by a high NaCl intake. Low dietary intakes of calcium and potassium also may contribute to the risk of hypertension. The urine sodium-to-potassium ratio is a stronger correlate of blood pressure than is either sodium or potassium alone. Alcohol consumption, psychosocial stress, and low levels of physical activity also may contribute to hypertension.

Adoption, twin, and family studies document a significant heritable component to blood pressure levels and hypertension. Family studies controlling for a common environment indicate that blood pressure heritabilities are in the range 15–35%. In twin studies, heritability estimates of blood pressure are ~60% for males and 30–40% for females. High blood pressure before age 55 occurs 3.8 times more frequently among persons with a positive family history of hypertension.

Genetic Considerations

Although specific genetic variants have been identified in rare Mendelian forms of hypertension (Table 247-5), these variants are not applicable to the vast majority (>98%) of patients with essential hypertension. For most individuals, it is likely that hypertension represents a polygenic disorder in which a combination of genes acts in concert with environmental exposures to make only a modest contribution to blood pressure. Further, different subsets of genes may lead to different phenotypes associated with hypertension, e.g., obesity, dyslipidemia, insulin resistance.

Several strategies are being utilized in the search for specific hypertension-related genes. Animal models (including selectively bred rats and congenic rat strains) provide a powerful approach for evaluating genetic loci and genes associated with hypertension. Comparative mapping strategies allow for the identification of syntenic genomic regions between the rat and human genomes that may be involved in blood pressure regulation. In association studies, different alleles (or combinations of alleles at different loci) of specific candidate genes or chromosomal regions are compared in hypertensive patients and normotensive control subjects. Current

evidence suggests that genes that encode components of the renin-angiotensin-aldosterone system, along with angiotensinogen and angiotensin-converting enzyme (ACE) polymorphisms, may be related to hypertension and to blood pressure sensitivity to dietary NaCl. The alpha-adducin gene is thought to be associated with increased renal tubular absorption of sodium, and variants of this gene may be associated with hypertension and salt sensitivity of blood pressure. Other genes possibly related to hypertension include genes encoding the AT_1 receptor, aldosterone synthase, and the β_2 adrenoreceptor. Genomewide association studies involve rapidly scanning markers across the entire genome to identify loci (not specific genes) associated with an observable trait (e.g., blood pressure) or a particular disease. This strategy has been facilitated by the availability of dense genotyping chips and the International HapMap. To date, the results of candidate gene studies often have not been replicated, and in contrast to several other polygenic disorders, genomewide association studies have had limited success in identifying genetic determinants of hypertension.

Preliminary evidence suggests that there may also be genetic determinants of target organ damage attributed to hypertension. Family studies indicate significant heritability of left ventricular mass, and there is considerable individual variation in the responses of the heart to hypertension. Family studies and variations in candidate genes associated with renal damage suggest that genetic factors also may contribute to hypertensive nephropathy. Specific genetic variants have been linked to CHD and stroke.

In the future, it is possible that DNA analysis will predict individual risk for hypertension and target organ damage and will identify responders to specific classes of antihypertensive agents. However, with the exception of the rare, monogenic hypertensive diseases, the genetic variants associated with hypertension remain to be confirmed, and the intermediate steps by which these variants affect blood pressure remain to be determined.

MECHANISMS OF HYPERTENSION

To provide a framework for understanding the pathogenesis of and treatment options for hypertensive disorders, it is useful to understand factors involved in the regulation of both normal and elevated arterial pressure. Cardiac output and peripheral resistance are the two determinants of arterial pressure **(Fig. 247-1)**. Cardiac output is determined by stroke volume and heart rate; stroke volume is related to myocardial contractility and to the size of the vascular compartment. Peripheral resistance is determined by functional and anatomic changes in small arteries (lumen diameter 100–400 μm) and arterioles.

Figure 247-1

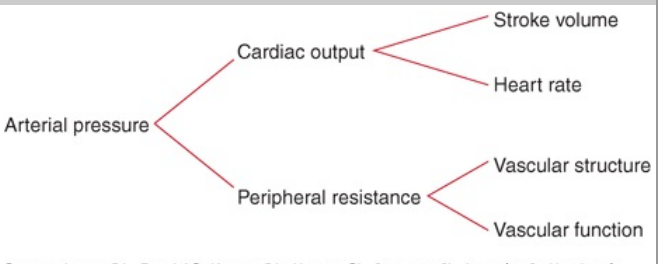

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Determinants of arterial pressure.

Intravascular Volume

Vascular volume is a primary determinant of arterial pressure over the long term. Sodium is predominantly an extracellular ion and is a primary determinant of the extracellular fluid volume. When NaCl intake exceeds the capacity of the kidney to excrete sodium, vascular volume initially expands and cardiac output increases. However, many vascular beds (including kidney and brain) have the capacity to autoregulate blood flow, and if constant blood flow is to be maintained in the face of increased arterial pressure, resistance within that bed must increase, since

$$\text{Blood flow} = \frac{\text{ressure across the vascular bed}}{\text{vascular resistance}}$$

The initial elevation of blood pressure in response to vascular volume expansion may be related to an increase of cardiac output; however, over time, peripheral resistance increases and cardiac output reverts toward normal. The effect of sodium on blood pressure is related to the provision of sodium with chloride; nonchloride salts of sodium have little or no effect on blood pressure. As arterial pressure increases in response to a high NaCl intake, urinary sodium excretion increases and sodium balance is maintained at the expense of an increase in arterial pressure. The mechanism for this "pressure-natriuresis" phenomenon may involve a subtle increase in the glomerular filtration rate, decreased absorbing capacity of the renal tubules, and possibly hormonal factors such as atrial natriuretic factor. In individuals with an impaired capacity to excrete sodium, greater increases in arterial pressure are required to achieve natriuresis and sodium balance.

NaCl-dependent hypertension may be a consequence of a decreased capacity of the kidney to excrete sodium, due either to intrinsic renal disease or to increased production of a salt-retaining hormone (mineralocorticoid) resulting in increased renal tubular reabsorption of sodium. Renal tubular sodium reabsorption also may be augmented by increased neural activity to the kidney. In each of these situations, a higher arterial pressure may be required to achieve sodium balance. Conversely, salt-wasting disorders are associated with low blood pressure levels. ESRD is an extreme example of volume-dependent hypertension. In ~80% of these patients, vascular volume and hypertension can be controlled with adequate dialysis; in the other 20%, the mechanism of hypertension is related to increased activity of the renin-angiotensin system and is likely to be responsive to

pharmacologic blockade of renin-angiotensin.

Autonomic Nervous System

The autonomic nervous system maintains cardiovascular homeostasis via pressure, volume, and chemoreceptor signals. Adrenergic reflexes modulate blood pressure over the short term, and adrenergic function, in concert with hormonal and volume-related factors, contributes to the long-term regulation of arterial pressure. The three endogenous catecholamines are norepinephrine, epinephrine, and dopamine. All three play important roles in tonic and phasic cardiovascular regulation.

The activities of the adrenergic receptors are mediated by guanosine nucleotide-binding regulatory proteins (G proteins) and by intracellular concentrations of downstream second messengers. In addition to receptor affinity and density, physiologic responsiveness to catecholamines may be altered by the efficiency of receptor-effector coupling at a site "distal" to receptor binding. The receptor sites are relatively specific both for the transmitter substance and for the response that occupancy of the receptor site elicits. Norepinephrine and epinephrine are agonists for all adrenergic receptor subtypes, although with varying affinities. Based on their physiology and pharmacology, adrenergic receptors have been divided into two principal types: α and β. These types have been differentiated further into α_1, α_2, β_1, and β_2 receptors. Recent molecular cloning studies have identified several additional subtypes. α Receptors are occupied and activated more avidly by norepinephrine than by epinephrine, and the reverse is true for β receptors. α_1 Receptors are located on postsynaptic cells in smooth muscle and elicit vasoconstriction. α_2 Receptors are localized on presynaptic membranes of postganglionic nerve terminals that synthesize norepinephrine. When activated by catecholamines, α_2 receptors act as negative feedback controllers, inhibiting further norepinephrine release. In the kidney, activation of α_1-adrenergic receptors increases renal tubular reabsorption of sodium. Different classes of antihypertensive agents either inhibit α_1 receptors or act as agonists of α_2 receptors and reduce systemic sympathetic outflow. Activation of myocardial β_1 receptors stimulates the rate and strength of cardiac contraction and consequently increases cardiac output. β_1 Receptor activation also stimulates renin release from the kidney. Another class of antihypertensive agents acts by inhibiting β_1 receptors. Activation of β_2 receptors by epinephrine relaxes vascular smooth muscle and results in vasodilation.

Circulating catecholamine concentrations may affect the number of adrenoreceptors in various tissues. Downregulation of receptors may be a consequence of sustained high levels of catecholamines and provides an explanation for decreasing responsiveness, or tachyphylaxis, to catecholamines. For example, orthostatic hypotension frequently is observed in patients with pheochromocytoma, possibly due to the lack of norepinephrine-induced vasoconstriction with assumption of the upright posture. Conversely, with chronic reduction of neurotransmitter substances, adrenoreceptors may increase in number or be upregulated, resulting in increased responsiveness to the neurotransmitter. Chronic administration of agents that block adrenergic receptors may result in upregulation, and withdrawal of those agents may produce a condition of temporary hypersensitivity to sympathetic stimuli. For example, clonidine is an antihypertensive agent that is a centrally acting α_2 agonist that inhibits sympathetic outflow. Rebound hypertension may occur with the abrupt cessation of clonidine therapy, probably as a consequence of upregulation of α_1 receptors.

Several reflexes modulate blood pressure on a minute-to-minute basis. One arterial baroreflex is mediated by stretch-sensitive sensory nerve endings in the carotid sinuses and the aortic arch. The rate of firing of these baroreceptors increases with arterial pressure, and the net effect is a decrease in sympathetic outflow, resulting in decreases in arterial pressure and heart rate. This is a primary mechanism for rapid buffering of acute fluctuations of arterial pressure that may occur during postural changes, behavioral or physiologic stress, and changes in blood volume. However, the activity of the baroreflex declines or adapts to sustained increases in arterial pressure such that the baroreceptors are reset to higher pressures. Patients with autonomic neuropathy and impaired baroreflex function may have extremely labile blood pressures with difficult-to-control episodic blood pressure spikes associated with tachycardia.

In both normal-weight and obese individuals, hypertension often is associated with increased sympathetic outflow. Based on recordings of postganglionic muscle nerve activity (detected by a microelectrode inserted in a peroneal nerve in the leg), sympathetic outflow tends to be higher in hypertensive than in normotensive individuals. Sympathetic outflow is increased in obesity-related hypertension and in hypertension associated with obstructive sleep apnea. Baroreceptor activation via electrical stimulation of carotid sinus afferent nerves has been shown to lower blood pressure in patients with "resistant" hypertension. Drugs that block the sympathetic nervous system are potent antihypertensive agents, indicating that the sympathetic nervous system plays a permissive, although not necessarily a causative, role in the maintenance of increased arterial pressure.

Pheochromocytoma is the most blatant example of hypertension related to increased catecholamine production, in this instance by a tumor. Blood pressure can be reduced by surgical excision of the tumor or by pharmacologic treatment with an α_1 receptor antagonist or with an inhibitor of tyrosine hydroxylase, the rate-limiting step in catecholamine biosynthesis.

Renin-Angiotensin-Aldosterone

The renin-angiotensin-aldosterone system contributes to the regulation of arterial pressure primarily via the vasoconstrictor properties of angiotensin II and the sodium-retaining properties of aldosterone. Renin is an aspartyl protease that is synthesized as an enzymatically inactive precursor, prorenin. Most renin in the circulation is synthesized in the renal afferent renal arteriole. Prorenin may be secreted directly into the circulation or may be activated within secretory cells and released as active renin. Although human plasma contains two to five times more prorenin than renin, there is no evidence that prorenin contributes to the physiologic activity of this system. There are three primary stimuli for renin secretion: (1) decreased NaCl transport in the distal portion of the thick ascending limb of the loop of Henle that abuts the corresponding

afferent arteriole (macula densa), (2) decreased pressure or stretch within the renal afferent arteriole (baroreceptor mechanism), and (3) sympathetic nervous system stimulation of renin-secreting cells via β_1 adrenoreceptors. Conversely, renin secretion is inhibited by increased NaCl transport in the thick ascending limb of the loop of Henle, by increased stretch within the renal afferent arteriole, and by β_1 receptor blockade. In addition, angiotensin II directly inhibits renin secretion due to angiotensin II type 1 receptors on juxtaglomerular cells, and renin secretion increases in response to pharmacologic blockade of either ACE or angiotensin II receptors.

Once released into the circulation, active renin cleaves a substrate, angiotensinogen, to form an inactive decapeptide, angiotensin I **(Fig. 247-2)**. A converting enzyme, located primarily but not exclusively in the pulmonary circulation, converts angiotensin I to the active octapeptide, angiotensin II, by releasing the C-terminal histidyl-leucine dipeptide. The same converting enzyme cleaves a number of other peptides, including and thereby inactivating the vasodilator bradykinin. Acting primarily through angiotensin II type 1 (AT_1) receptors on cell membranes, angiotensin II is a potent pressor substance, the primary tropic factor for the secretion of aldosterone by the adrenal zona glomerulosa, and a potent mitogen that stimulates vascular smooth muscle cell and myocyte growth. Independent of its hemodynamic effects, angiotensin II may play a role in the pathogenesis of atherosclerosis through a direct cellular action on the vessel wall. An angiotensin II type 2 (AT_2) receptor has been characterized. It is widely distributed in the kidney and has the opposite functional effects of the AT_1 receptor. The AT_2 receptor induces vasodilation, sodium excretion, and inhibition of cell growth and matrix formation. Experimental evidence suggests that the AT_2 receptor improves vascular remodeling by stimulating smooth muscle cell apoptosis and contributes to the regulation of glomerular filtration rate. AT_1 receptor blockade induces an increase in AT_2 receptor activity.

Figure 247-2

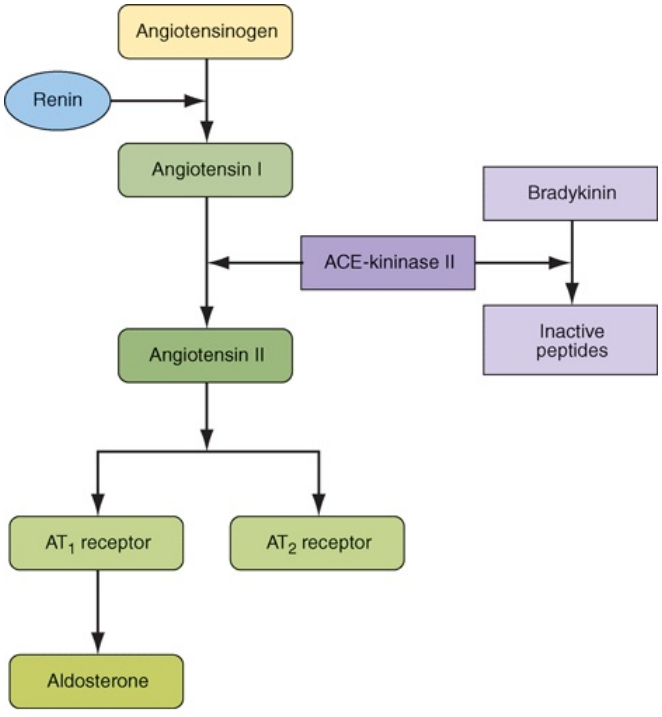

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Renin-angiotensin-aldosterone axis.

Renin-secreting tumors are clear examples of renin-dependent hypertension. In the kidney, these tumors include benign hemangiopericytomas of the juxtaglomerular apparatus and, infrequently, renal carcinomas, including Wilms' tumors. Renin-producing carcinomas also have been described in lung, liver, pancreas, colon, and adrenals. In these instances, in addition to excision and/or ablation of the tumor, treatment of hypertension includes pharmacologic therapies targeted to inhibit angiotensin II production or action. Renovascular hypertension is another renin-mediated form of hypertension. Obstruction of the renal artery leads to decreased renal perfusion pressure, thereby stimulating renin secretion. Over time, as a consequence of secondary renal damage, this form of hypertension may become less renin dependent.

Angiotensinogen, renin, and angiotensin II are also synthesized locally in many tissues, including the brain, pituitary, aorta, arteries, heart, adrenal glands, kidneys, adipocytes, leukocytes, ovaries, testes, uterus, spleen, and skin. Angiotensin II in tissues may be formed by the enzymatic activity of renin or by other proteases, e.g., tonin, chymase, and cathepsins. In addition to regulating local blood flow, tissue angiotensin II is a mitogen that stimulates growth and contributes to modeling and repair. Excess tissue angiotensin II may contribute to atherosclerosis, cardiac hypertrophy, and renal failure and consequently may be a target for pharmacologic therapy to prevent target organ damage.

Angiotensin II is the primary tropic factor regulating the synthesis and secretion of aldosterone by the zona glomerulosa of the adrenal cortex. Aldosterone synthesis is also dependent on potassium, and aldosterone secretion may be decreased in potassium-depleted individuals. Although acute elevations of adrenocorticotropic hormone (ACTH) levels also increase aldosterone secretion, ACTH is not an important tropic factor for the chronic regulation of aldosterone.

Aldosterone is a potent mineralocorticoid that increases sodium reabsorption by amiloride-sensitive epithelial sodium channels (ENaC) on the apical surface of the principal cells of the renal cortical collecting duct (Chap. 277). Electric neutrality is maintained by exchanging sodium for potassium and hydrogen ions. Consequently, increased aldosterone secretion may result in hypokalemia and alkalosis. Because potassium depletion may inhibit aldosterone synthesis, clinically, hypokalemia should be corrected before a patient is evaluated for hyperaldosteronism.

Mineralocorticoid receptors also are expressed in the colon, salivary glands, and sweat glands. Cortisol also binds to these receptors but normally functions as a less potent mineralocorticoid than aldosterone because cortisol is converted to cortisone by the enzyme 11 β-hydroxysteroid dehydrogenase type 2. Cortisone has no affinity for the mineralocorticoid receptor. Primary aldosteronism is a compelling example of mineralocorticoid-mediated hypertension. In this disorder, adrenal aldosterone synthesis and release are independent of renin-angiotensin, and renin release is suppressed by the resulting volume expansion.

Aldosterone also has effects on nonepithelial targets. Aldosterone and/or mineralocorticoid receptor activation induces structural and functional alterations in the heart, kidney, and blood vessels, leading to myocardial fibrosis, nephrosclerosis, and vascular inflammation and remodeling, perhaps as a consequence of oxidative stress. These effects are amplified by a high salt intake. In animal models, high circulating aldosterone levels stimulate cardiac fibrosis and left ventricular hypertrophy, and spironolactone (an aldosterone antagonist) prevents aldosterone-induced myocardial fibrosis. Pathologic patterns of left ventricular geometry also have been associated with elevations of plasma aldosterone concentration in patients with essential hypertension as well as in patients with primary aldosteronism. In patients with CHF, low-dose spironolactone reduces the risk of progressive heart failure and sudden death from cardiac causes by 30%. Owing to a renal hemodynamic effect, in patients with primary aldosteronism, high circulating levels of aldosterone also may cause glomerular hyperfiltration and albuminuria. These renal effects are reversible after removal of the effects of excess aldosterone by adrenalectomy or spironolactone.

Increased activity of the renin-angiotensin-aldosterone axis is not invariably associated with hypertension. In response to a low-NaCl diet or to volume contraction, arterial pressure and volume homeostasis may be maintained by increased activity of the renin-angiotensin-aldosterone axis. Secondary aldosteronism (i.e., increased aldosterone secondary to increased renin-angiotensin), but not hypertension, also is observed in edematous states such as CHF and liver disease.

Vascular Mechanisms

Vascular radius and compliance of resistance arteries are also important determinants of arterial pressure. Resistance to flow varies inversely with the fourth power of the radius, and consequently, small decreases in lumen size significantly increase resistance. In hypertensive patients, structural, mechanical, or functional changes may reduce the lumen diameter of small arteries and arterioles. Remodeling refers to geometric alterations in the vessel wall without a change in vessel volume. Hypertrophic (increased cell size, and increased deposition of intercellular matrix) or eutrophic vascular remodeling results in decreased lumen size and hence contributes to increased peripheral resistance. Apoptosis, low-grade inflammation, and vascular fibrosis also contribute to remodeling. Lumen diameter also is related to elasticity of the vessel. Vessels with a high degree of elasticity can accommodate an increase of volume with relatively little change in pressure, whereas in a semirigid vascular system, a small increment in volume induces a relatively large increment of pressure.

Hypertensive patients have stiffer arteries, and arteriosclerotic patients may have particularly high systolic blood pressures and wide pulse pressures as a consequence of decreased vascular compliance due to structural changes in the vascular wall. Recent evidence suggests that arterial stiffness has independent predictive value for cardiovascular events. Clinically, a number of devices are available to evaluate arterial stiffness or compliance, including ultrasound and magnetic resonance imaging (MRI).

Ion transport by vascular smooth muscle cells may contribute to hypertension-associated abnormalities of vascular tone and vascular growth, both of which are modulated by intracellular pH (pH_i). Three ion transport mechanisms participate in the regulation of pH_i: (1) Na^+-H^+ exchange, (2) Na^+-dependent HCO_3^--Cl^- exchange, and (3) cation-independent $HCO_3^-Cl^-$ exchange. Based on measurements in cell types that are more accessible than vascular smooth muscle (e.g., leukocytes, erythrocytes, platelets, skeletal muscle), activity of the Na^+-H^+ exchanger is increased in hypertension, and this may result in increased vascular tone by two mechanisms. First, increased sodium entry may lead to increased vascular tone by activating Na^+-Ca^{2+} exchange and thereby increasing intracellular calcium. Second, increased pH_i enhances calcium sensitivity of the contractile apparatus, leading to an increase in contractility for a given intracellular calcium concentration. Additionally, increased Na^+-H^+ exchange may stimulate growth of vascular smooth muscle cells by enhancing sensitivity to mitogens.

Vascular endothelial function also modulates vascular tone. The vascular endothelium synthesizes and releases a spectrum of vasoactive substances, including nitric oxide, a potent vasodilator. Endothelium-dependent vasodilation is impaired in hypertensive patients. This impairment often is assessed with high-resolution ultrasonography before and after the hyperemic phase of reperfusion that follows 5 minutes of forearm ischemia. Alternatively, endothelium-dependent vasodilation may be assessed in response to an intra-arterially infused endothelium-dependent vasodilator, e.g., acetylcholine. Endothelin is a vasoconstrictor peptide produced by the

endothelium, and that orally active endothelin antagonists may lower blood pressure in patients with resistant hypertension.

Currently, it is not known if the hypertension-related vascular abnormalities of ion transport and endothelial function are primary alterations or secondary consequences of elevated arterial pressure. Limited evidence suggests that vascular compliance and endothelium-dependent vasodilation may be improved by aerobic exercise, weight loss, and antihypertensive agents. It remains to be determined whether these interventions affect arterial structure and stiffness via a blood pressure–independent mechanism and whether different classes of antihypertensive agents preferentially affect vascular structure and function.

PATHOLOGIC CONSEQUENCES OF HYPERTENSION

Hypertension is an independent predisposing factor for heart failure, coronary artery disease, stroke, renal disease, and peripheral arterial disease (PAD).

Heart

Heart disease is the most common cause of death in hypertensive patients. Hypertensive heart disease is the result of structural and functional adaptations leading to left ventricular hypertrophy, CHF, abnormalities of blood flow due to atherosclerotic coronary artery disease and microvascular disease, and cardiac arrhythmias.

Both genetic and hemodynamic factors contribute to left ventricular hypertrophy. Clinically, left ventricular hypertrophy can be diagnosed by electrocardiography, although echocardiography provides a more sensitive measure of left ventricular wall thickness. Individuals with left ventricular hypertrophy are at increased risk for CHD, stroke, CHF, and sudden death. Aggressive control of hypertension can regress or reverse left ventricular hypertrophy and reduce the risk of cardiovascular disease. It is not clear whether different classes of antihypertensive agents have an added impact on reducing left ventricular mass, independent of their blood pressure–lowering effect.

CHF may be related to systolic dysfunction, diastolic dysfunction, or a combination of the two. Abnormalities of diastolic function that range from asymptomatic heart disease to overt heart failure are common in hypertensive patients. Patients with diastolic heart failure have a preserved ejection fraction, which is a measure of systolic function. Approximately one-third of patients with CHF have normal systolic function but abnormal diastolic function. Diastolic dysfunction is an early consequence of hypertension-related heart disease and is exacerbated by left ventricular hypertrophy and ischemia. Cardiac catheterization provides the most accurate assessment of diastolic function. Alternatively, diastolic function can be evaluated by several noninvasive methods, including echocardiography and radionuclide angiography.

Brain

Stroke is the second most frequent cause of death in the world; it accounts for 5 million deaths each year, with an additional 15 million persons having nonfatal strokes. Elevated blood pressure is the strongest risk factor for stroke. Approximately 85% of strokes are due to infarction, and the remainder are due to either intracerebral or subarachnoid hemorrhage. The incidence of stroke rises progressively with increasing blood pressure levels, particularly systolic blood pressure in individuals >65 years. Treatment of hypertension convincingly decreases the incidence of both ischemic and hemorrhagic strokes.

Hypertension also is associated with impaired cognition in an aging population, and longitudinal studies support an association between midlife hypertension and late-life cognitive decline. Hypertension-related cognitive impairment and dementia may be a consequence of a single infarct due to occlusion of a "strategic" larger vessel or multiple lacunar infarcts due to occlusive small vessel disease resulting in subcortical white matter ischemia. Several clinical trials suggest that antihypertensive therapy has a beneficial effect on cognitive function, although this remains an active area of investigation.

Cerebral blood flow remains unchanged over a wide range of arterial pressures (mean arterial pressure of 50–150 mmHg) through a process termed *autoregulation* of blood flow. In patients with the clinical syndrome of malignanthypertension, encephalopathy is related to failure of autoregulation of cerebral blood flow at the upper pressure limit, resulting in vasodilation and hyperperfusion. Signs and symptoms of hypertensive encephalopathy may include severe headache, nausea and vomiting (often of a projectile nature), focal neurologic signs, and alterations in mental status. Untreated, hypertensive encephalopathy may progress to stupor, coma, seizures, and death within hours. It is important to distinguish hypertensive encephalopathy from other neurologic syndromes that may be associated with hypertension, e.g., cerebral ischemia, hemorrhagic or thrombotic stroke, seizure disorder, mass lesions, pseudotumor cerebri, delirium tremens, meningitis, acute intermittent porphyria, traumatic or chemical injury to the brain, and uremic encephalopathy.

Kidney

The kidney is both a target and a cause of hypertension. Primary renal disease is the most common etiology of secondary hypertension. Mechanisms of kidney-related hypertension include a diminished capacity to excrete sodium, excessive renin secretion in relation to volume status, and sympathetic nervous system overactivity. Conversely, hypertension is a risk factor for renal injury and end-stage renal disease. The increased risk associated with high blood pressure is graded, continuous, and present throughout the distribution of blood pressure above optimal pressure. Renal risk appears to be more closely related to systolic than to diastolic blood pressure, and black men are at greater risk than white men for developing ESRD at every level of blood pressure. Proteinuria is a reliable marker of the severity of chronic kidney disease and is a predictor of its progression. Patients with high urine protein excretion (>3 g/24 h) have a more rapid rate of progression than do those with lower protein excretion rates.

Atherosclerotic, hypertension-related vascular lesions in the kidney primarily affect preglomerular arterioles, resulting in ischemic changes in the glomeruli and postglomerular structures. Glomerular injury also may be a consequence of direct damage to the glomerular capillaries due to glomerular hyperperfusion. Studies of

hypertension-related renal damage, primarily in experimental animals, suggest that loss of autoregulation of renal blood flow at the afferent arteriole results in transmission of elevated pressures to an unprotected glomerulus with ensuing hyperfiltration, hypertrophy, and eventual focal segmental glomerular sclerosis. With progressive renal injury there is a loss of autoregulation of renal blood flow and glomerular filtration rate, resulting in a lower blood pressure threshold for renal damage and a steeper slope between blood pressure and renal damage. The result may be a vicious cycle of renal damage and nephron loss leading to more severe hypertension, glomerular hyperfiltration, and further renal damage. Glomerular pathology progresses to glomerulosclerosis, and eventually the renal tubules may also become ischemic and gradually atrophic. The renal lesion associated with malignant hypertension consists of fibrinoid necrosis of the afferent arterioles, sometimes extending into the glomerulus, and may result in focal necrosis of the glomerular tuft.

Clinically, macroalbuminuria (a random urine albumin/creatinine ratio >300 mg/g) or microalbuminuria (a random urine albumin/creatinine ratio 30–300 mg/g) are early markers of renal injury. These are also risk factors for renal disease progression and cardiovascular disease.

Peripheral Arteries

In addition to contributing to the pathogenesis of hypertension, blood vessels may be a target organ for atherosclerotic disease secondary to long-standing elevated blood pressure. Hypertensive patients with arterial disease of the lower extremities are at increased risk for future cardiovascular disease. Although patients with stenotic lesions of the lower extremities may be asymptomatic, intermittent claudication is the classic symptom of PAD. This is characterized by aching pain in the calves or buttocks while walking that is relieved by rest. The ankle-brachial index is a useful approach for evaluating PAD and is defined as the ratio of noninvasively assessed ankle to brachial (arm) systolic blood pressure. An ankle-brachial index <0.90 is considered diagnostic of PAD and is associated with >50% stenosis in at least one major lower limb vessel. Several studies suggest that an ankle-brachial index <0.80 is associated with elevated blood pressure, particularly systolic blood pressure.

DEFINING HYPERTENSION

From an epidemiologic perspective, there is no obvious level of blood pressure that defines hypertension. In adults, there is a continuous, incremental risk of cardiovascular disease, stroke, and renal disease across levels of both systolic and diastolic blood pressure. The Multiple Risk Factor Intervention Trial (MRFIT), which included >350,000 male participants, demonstrated a continuous and graded influence of both systolic and diastolic blood pressure on CHD mortality, extending down to systolic blood pressures of 120 mmHg. Similarly, results of a meta-analysis involving almost 1 million participants indicate that ischemic heart disease mortality, stroke mortality, and mortality from other vascular causes are directly related to the height of the blood pressure, beginning at 115/75 mmHg, without evidence of a threshold. Cardiovascular disease risk doubles for every 20-mmHg increase in systolic and 10-mmHg increase in diastolic pressure. Among older individuals, systolic blood pressure and pulse pressure are more powerful predictors of cardiovascular disease than is diastolic blood pressure.

Clinically, hypertension may be defined as that level of blood pressure at which the institution of therapy reduces blood pressure–related morbidity and mortality. Current clinical criteria for defining hypertension generally are based on the average of two or more seated blood pressure readings during each of two or more outpatient visits. A recent classification recommends blood pressure criteria for defining normal blood pressure, prehypertension, hypertension (stages I and II), and isolated systolic hypertension, which is a common occurrence among the elderly **(Table 247-1)**. In children and adolescents, hypertension generally is defined as systolic and/or diastolic blood pressure consistently >95th percentile for age, sex, and height. Blood pressures between the 90th and 95th percentiles are considered prehypertensive and are an indication for lifestyle interventions.

Table 247–1. Blood Pressure Classification

Blood Pressure Classification	Systolic, mmHg	Diastolic, mmHg
Normal	<120	*and* <80
Prehypertension	120–139	*or* 80–89
Stage 1 hypertension	140–159	*or* 90–99
Stage 2 hypertension	≥160	*or* ≥100
Isolated systolic hypertension	≥140	*and* <90

Source: Adapted from Chobanian et al.

Home blood pressure and average 24-h ambulatory blood pressure measurements are generally lower than clinic blood pressures. Because ambulatory blood pressure recordings yield multiple readings throughout the day and night, they provide a more comprehensive assessment of the vascular burden of hypertension than do a limited number of office readings. Increasing evidence suggests that home blood pressures, including 24-h blood pressure recordings, more reliably predict target organ damage than do office blood pressures. Blood pressure tends to be higher in the early morning hours, soon after waking, than at other times of day. Myocardial infarction and stroke are more common in the early morning hours. Nighttime blood pressures are generally 10–20% lower than daytime blood pressures, and an attenuated nighttime blood pressure "dip" is associated with increased cardiovascular disease risk. Recommended criteria for a diagnosis of hypertension are average awake blood pressure ≥135/85 mmHg and asleep blood pressure ≥120/75 mmHg. These levels approximate a clinic blood pressure of 140/90 mmHg.

Approximately 15–20% of patients with stage 1 hypertension (as defined in Table 247-1) based on office blood pressures have average ambulatory readings <135/85 mmHg. This phenomenon, so-called white coat

hypertension, also may be associated with an increased risk of target organ damage (e.g., left ventricular hypertrophy, carotid atherosclerosis, overall cardiovascular morbidity), although to a lesser extent than in individuals with elevated office and ambulatory readings. Individuals with white coat hypertension are also at increased risk for developing sustained hypertension.

CLINICAL DISORDERS OF HYPERTENSION

Depending on methods of patient ascertainment, ~80–95% of hypertensive patients are diagnosed as having "essential" hypertension (also referred to as primary or idiopathic hypertension). In the remaining 5–20% of hypertensive patients, a specific underlying disorder causing the elevation of blood pressure can be identified **(Tables 247-2** and **247-3)**. In individuals with "secondary" hypertension, a specific mechanism for the blood pressure elevation is often more apparent.

Table 247–2. Systolic Hypertension with Wide Pulse Pressure
1. Decreased vascular compliance (arteriosclerosis)
2. Increased cardiac output.
a. Aortic regurgitation b. Thyrotoxicosis.
c. Hyperkinetic heart syndrome. d. Fever.
e. Arteriovenous fistula
f. Patent ductus arteriosus

Table 247–3. Secondary Causes of Systolic and Diastolic Hypertension	
Renal	Parenchymal diseases, renal cysts (including polycystic kidney disease), renal tumors (including renin-secreting tumors), obstructive uropathy
Renovascular	Arteriosclerotic, fibromuscular dysplasia
Adrenal	Primary aldosteronism, Cushing's syndrome, 17α-hydroxylase deficiency, 11β-hydroxylase deficiency, 11-hydroxysteroid dehydrogenase deficiency (licorice), pheochromocytoma
Aortic coarctation	
Obstructive sleep apnea	
Preeclampsia/eclampsia	
Neurogenic	Psychogenic, diencephalic syndrome, familial dysautonomia, polyneuritis (acute porphyria, lead poisoning), acute increased intracranial pressure, acute spinal cord section
Miscellaneous endocrine	Hypothyroidism, hyperthyroidism, hypercalcemia, acromegaly
Medications	High-dose estrogens, adrenal steroids, decongestants, appetite suppressants, cyclosporine, tricyclic antidepressants, monamine oxidase inhibitors, erythropoietin, nonsteroidal anti-inflammatory agents, cocaine
Mendelian forms of hypertension	See Table 247-4

Essential Hypertension

Essential hypertension tends to be familial and is likely to be the consequence of an interaction between environmental and genetic factors. The prevalence of essential hypertension increases with age, and individuals with relatively high blood pressures at younger ages are at increased risk for the subsequent development of hypertension. It is likely that essential hypertension represents a spectrum of disorders with different underlying pathophysiologies. In the majority of patients with established hypertension, peripheral resistance is increased and cardiac output is normal or decreased; however, in younger patients with mild or labile hypertension, cardiac output may be increased and peripheral resistance may be normal.

When plasma renin activity (PRA) is plotted against 24-h sodium excretion, ~10–15% of hypertensive patients have high PRA and 25% have low PRA. High-renin patients may have a vasoconstrictor form of hypertension, whereas low-renin patients may have volume-dependent hypertension. Inconsistent associations between plasma aldosterone and blood pressure have been described in patients with essential hypertension. The association between aldosterone and blood pressure is more striking in African Americans, and PRA tends to be low in hypertensive African Americans. This raises the possibility that subtle increases in aldosterone may contribute to hypertension in at least some groups of patients who do not have overt primary aldosteronism. Furthermore, spironolactone, an aldosterone antagonist, may be a particularly effective antihypertensive agent for some patients with essential hypertension, including some patients with "drug-resistant" hypertension.

Obesity and the Metabolic Syndrome

(See also Chap. 242) There is a well-documented association between obesity (body mass index >30 kg/m^2) and hypertension. Further, cross-sectional studies indicate a direct linear correlation between body weight (or

body mass index) and blood pressure. Centrally located body fat is a more important determinant of blood pressure elevation than is peripheral body fat. In longitudinal studies, a direct correlation exists between change in weight and change in blood pressure over time. Sixty percent of hypertensive adults are more than 20% overweight. It has been established that 60–70% of hypertension in adults may be directly attributable to adiposity.

Hypertension and dyslipidemia frequently occur together and in association with resistance to insulin-stimulated glucose uptake. This clustering of risk factors is often, but not invariably, associated with obesity, particularly abdominal obesity. Insulin resistance also is associated with an unfavorable imbalance in the endothelial production of mediators that regulate platelet aggregation, coagulation, fibrinolysis, and vessel tone. When these risk factors cluster, the risks for CHD, stroke, diabetes, and cardiovascular disease mortality are increased further.

Depending on the populations studied and the methodologies for defining insulin resistance, ~25–50% of nonobese, nondiabetic hypertensive persons are insulin resistant. The constellation of insulin resistance, abdominal obesity, hypertension, and dyslipidemia has been designated as the *metabolic syndrome*. As a group, first-degree relatives of patients with essential hypertension are also insulin resistant, and hyperinsulinemia (a surrogate marker of insulin resistance) may predict the eventual development of hypertension and cardiovascular disease. Although the metabolic syndrome may in part be heritable as a polygenic condition, the expression of the syndrome is modified by environmental factors, such as degree of physical activity and diet. Insulin sensitivity increases and blood pressure decreases in response to weight loss. The recognition that cardiovascular disease risk factors tend to cluster within individuals has important implications for the evaluation and treatment of hypertension. Evaluation of both hypertensive patients and individuals at risk for developing hypertension should include assessment of overall cardiovascular disease risk. Similarly, introduction of lifestyle modification strategies and drug therapies should address overall risk and not simply focus on hypertension.

Renal Parenchymal Diseases

Virtually all disorders of the kidney may cause hypertension (Table 247-3), and renal disease is the most common cause of secondary hypertension. Hypertension is present in >80% of patients with chronic renal failure. In general, hypertension is more severe in glomerular diseases than in interstitial diseases such as chronic pyelonephritis. Conversely, hypertension may cause nephrosclerosis, and in some instances it may be difficult to determine whether hypertension or renal disease was the initial disorder. Proteinuria >1000 mg/d and an active urine sediment are indicative of primary renal disease. In either instance, the goals are to control blood pressure and retard the rate of progression of renal dysfunction.

RENOVASCULAR HYPERTENSION

Hypertension due to an occlusive lesion of a renal artery, renovascular hypertension, is a potentially curable form of hypertension. In the initial stages, the mechanism of hypertension generally is related to activation of the renin-angiotensin system. However, renin activity and other components of the renin-angiotensin system may be elevated only transiently; over time, sodium retention and recruitment of other pressure mechanisms may contribute to elevated arterial pressure. Two groups of patients are at risk for this disorder: older arteriosclerotic patients who have a plaque obstructing the renal artery, frequently at its origin, and patients with fibromuscular dysplasia. Atherosclerosis accounts for the large majority of patients with renovascular hypertension. Although fibromuscular dysplasia may occur at any age, it has a strong predilection for young white women. The prevalence in females is eightfold that in males. There are several histologic variants of fibromuscular dysplasia, including medial fibroplasia, perimedial fibroplasia, medial hyperplasia, and intimal fibroplasia. Medial fibroplasia is the most common variant and accounts for approximately two-thirds of patients. The lesions of fibromuscular dysplasia are frequently bilateral and, in contrast to atherosclerotic renovascular disease, tend to affect more distal portions of the renal artery.

In addition to the age and sex of the patient, several clues from the history and physical examination suggest a diagnosis of renovascularhypertension. The diagnosis should be considered in patients with other evidence of atherosclerotic vascular disease. Although response to antihypertensive therapy does not exclude the diagnosis, severe or refractory hypertension, recent loss of hypertension control or recent onset of moderately severe hypertension, and unexplained deterioration of renal function or deterioration of renal function associated with an ACE inhibitor should raise the possibility of renovascular hypertension. Approximately 50% of patients with renovascular hypertension have an abdominal or flank bruit, and the bruit is more likely to be hemodynamically significant if it lateralizes or extends throughout systole into diastole.

If blood pressure is adequately controlled with a simple antihypertensive regimen and renal function remains stable, there may be little impetus to pursue an evaluation for renal artery stenosis, particularly in an older patient with atherosclerotic disease and comorbid conditions. Patients with long-standing hypertension, advanced renal insufficiency, or diabetes mellitus are less likely to benefit from renal vascular repair. The most effective medical therapies include an ACE inhibitor or an angiotensin II receptor blocker; however, these agents decrease glomerular filtration rate in a stenotic kidney owing to efferent renal arteriolar dilation. In the presence of bilateral renal artery stenosis or renal artery stenosis to a solitary kidney, progressive renal insufficiency may result from the use of these agents. Importantly, the renal insufficiency is generally reversible after discontinuation of the offending drug.

If renal artery stenosis is suspected and if the clinical condition warrants an intervention such as percutaneous transluminal renal angioplasty (PTRA), placement of a vascular endoprosthesis (stent), or surgical renal revascularization, imaging studies should be the next step in the evaluation. As a screening test, renal blood flow may be evaluated with a radionuclide [^{131}I]-orthoiodohippurate (OIH) scan or glomerular filtration rate may be evaluated with a [^{99}mTc]-diethylenetriamine pentaacetic acid (DTPA) scan before and after a single dose of captopril (or another ACE inhibitor). The following are consistent with a positive study: (1) decreased relative uptake by the involved kidney, which contributes <40% of total renal function, (2) delayed uptake on the affected

side, and (3) delayed washout on the affected side. In patients with normal, or nearly normal, renal function, a normal captopril renogram essentially excludes functionally significant renal artery stenosis; however, its usefulness is limited in patients with renal insufficiency (creatinine clearance <20 mL/min) or bilateral renal artery stenosis. Additional imaging studies are indicated if the scan is positive. Doppler ultrasound of the renal arteries produces reliable estimates of renal blood flow velocity and offers the opportunity to track a lesion over time. Positive studies usually are confirmed at angiography, whereas false-negative results occur frequently, particularly in obese patients. Gadolinium-contrast magnetic resonance angiography offers clear images of the proximal renal artery but may miss distal lesions. An advantage is the opportunity to image the renal arteries with an agent that is not nephrotoxic. Contrast arteriography remains the "gold standard" for evaluation and identification of renal artery lesions. Potential risks include nephrotoxicity, particularly in patients with diabetes mellitus or preexisting renal insufficiency.

Some degree of renal artery obstruction may be observed in almost 50% of patients with atherosclerotic disease, and there are several approaches for evaluating the functional significance of such a lesion to predict the effect of vascular repair on blood pressure control and renal function. Each approach has varying degrees of sensitivity and specificity, and no single test is sufficiently reliable to determine a causal relationship between a renal artery lesion and hypertension. Functionally significant lesions generally occlude more than 70% of the lumen of the affected renal artery. On angiography, the presence of collateral vessels to the ischemic kidney suggests a functionally significant lesion. A lateralizing renal vein renin ratio (ratio >1.5 of affected side/contralateral side) has a 90% predictive value for a lesion that would respond to vascular repair; however, the false-negative rate for blood pressure control is 50–60%. Measurement of the pressure gradient across a renal artery lesion does not reliably predict the response to vascular repair.

In the final analysis, a decision concerning vascular repair vs. medical therapy and the type of repair procedure should be individualized for each patient. Patients with fibromuscular disease have more favorable outcomes than do patients with atherosclerotic lesions, presumably owing to their younger age, shorter duration of hypertension, and less systemic disease. Because of its low risk-versus-benefit ratio and high success rate (improvement or cure of hypertension in 90% of patients and restenosis rate of 10%), PTRA is the initial treatment of choice for these patients. Surgical revascularization may be undertaken if PTRA is unsuccessful or if a branch lesion is present. In atherosclerotic patients, vascular repair should be considered if blood pressure cannot be controlled adequately despite optimal medical therapy or if renal function deteriorates. Surgery may be the preferred initial approach for younger atherosclerotic patients without comorbid conditions; however, for most atherosclerotic patients, depending on the location of the lesion, the initial approach may be PTRA and/or stenting. Surgical revascularization may be indicated if these approaches are unsuccessful, the vascular lesion is not amenable to PTRA or stenting, or concomitant aortic surgery is required, e.g., to repair an aneurysm. A National Institutes of Health–sponsored prospective, randomized clinical trial is in progress comparing medical therapy alone with medical therapy plus renal revascularization regarding Cardiovascular Outcomes for Renal Atherosclerotic Lesions (CORAL).

Primary Aldosteronism

Excess aldosterone production due to primary aldosteronism is a potentially curable form of hypertension. In patients with primary aldosteronism, increased aldosterone production is independent of the renin-angiotensin system, and the consequences are sodium retention, hypertension, hypokalemia, and low PRA. The reported prevalence of this disorder varies from <2% to ~15% of hypertensive individuals. In part, this variation is related to the intensity of screening and the criteria for establishing the diagnosis.

History and physical examination provide little information about the diagnosis. The age at the time of diagnosis is generally the third through fifth decade. Hypertension is usually mild to moderate but occasionally may be severe; primary aldosteronism should be considered in all patients with refractory hypertension. Hypertension in these patients may be associated with glucose intolerance. Most patients are asymptomatic, although, infrequently, polyuria, polydipsia, paresthesias, or muscle weakness may be present as a consequence of hypokalemic alkalosis. In a hypertensive patient with unprovoked hypokalemia (i.e., unrelated to diuretics, vomiting, or diarrhea), the prevalence of primary aldosteronism approaches 40–50%. In patients on diuretics, serum potassium <3.1 mmol/L (<3.1 meq/L) also raises the possibility of primary aldosteronism; however, serum potassium is an insensitive and nonspecific screening test. However, serum potassium is normal in ~25% of patients subsequently found to have an aldosterone-producing adenoma, and higher percentages of patients with other etiologies of primary aldosteronism are not hypokalemic. Additionally, hypokalemic hypertension may be a consequence of secondary aldosteronism, other mineralocorticoid- and glucocorticoid-induced hypertensive disorders, and pheochromocytoma.

The ratio of plasma aldosterone to plasma renin activity (PA/PRA) is a useful screening test. These measurements preferably are obtained in ambulatory patients in the morning. A ratio >30:1 in conjunction with a plasma aldosterone concentration >555 pmol/L (>20 ng/dL) reportedly has a sensitivity of 90% and a specificity of 91% for an aldosterone-producing adenoma. In a Mayo Clinic series, an aldosterone-producing adenoma subsequently was confirmed surgically in >90% of hypertensive patients with a PA/PRA ratio ≥20 and a plasma aldosterone concentration ≥415 pmol/L (≥15 ng/dL). There are, however, several caveats to interpreting the ratio. The cutoff for a "high" ratio is laboratory- and assay-dependent. Some antihypertensive agents may affect the ratio (e.g., aldosterone antagonists, angiotensin receptor antagonists, and ACE inhibitors may increase renin; aldosterone antagonists may increase aldosterone). Current recommendations are to withdraw aldosterone antagonists for at least 4 weeks before obtaining these measurements, with this caveat. The ratio has been reported to be useful as a screening test in measurements obtained with patients taking their usual antihypertensive medications. A high ratio in the absence of an elevated plasma aldosterone level is considerably less specific for primary aldosteronism since many patients with essential hypertension have low renin levels in this setting, particularly African Americans and elderly patients. In patients with renal insufficiency, the ratio may also be elevated because of decreased aldosterone clearance. In patients with an elevated PA/PRA ratio, the diagnosis of primary aldosteronism can be confirmed by demonstrating failure to suppress plasma aldosterone to

<277 pmol/L (<10 ng/dL) after IV infusion of 2 L of isotonic saline over 4 h; post-saline infusion plasma aldosterone values between 138 and 277 pmol/L (5–10 ng/dL) are not determinant. Alternative confirmatory tests include failure to suppress aldosterone (based on test specific criteria) in response to an oral NaCl load, fludrocortisone, or captopril.

Several adrenal abnormalities may culminate in the syndrome of primary aldosteronism, and appropriate therapy depends on the specific etiology. Some 60–70% of patients have an aldosterone-producing adrenal adenoma. The tumor is almost always unilateral, and most often measures <3 cm in diameter. Most of the remainder of these patients have bilateral adrenocortical hyperplasia (idiopathic hyperaldosteronism). Rarely, primary aldosteronism may be caused by an adrenal carcinoma or an ectopic malignancy, e.g., ovarian arrhenoblastoma. Most aldosterone-producing carcinomas, in contrast to adrenal adenomas and hyperplasia, produce excessive amounts of other adrenal steroids in addition to aldosterone. Functional differences in hormone secretion may assist in the differential diagnosis. Aldosterone biosynthesis is more responsive to adrenocorticotropic hormone (ACTH) in patients with adenoma and more responsive to angiotensin in patients with hyperplasia. Consequently, patients with adenoma tend to have higher plasma aldosterone in the early morning that decreases during the day, reflecting the diurnal rhythm of ACTH, whereas plasma aldosterone tends to increase with upright posture in patients with hyperplasia, reflecting the normal postural response of the renin-angiotensin- aldosterone axis. However, there is some overlap in the ability of these measurements to discriminate between adenoma and hyperplasia.

Adrenal computed tomography (CT) should be carried out in all patients diagnosed with primary aldosteronism. High-resolution CT may identify tumors as small as 0.3 cm and is positive for an adrenal tumor 90% of the time. If the CT is not diagnostic, an adenoma may be detected by adrenal scintigraphy with 6 ß-[I^{131}]iodomethyl-19-norcholesterol after dexamethasone suppression (0.5 mg every 6 h for 7 days); however, this technique has decreased sensitivity for adenomas <1.5 cm.

When carried out by an experienced radiologist, bilateral adrenal venous sampling for measurement of plasma aldosterone is the most accurate means of differentiating unilateral from bilateral forms of primary aldosteronism. The sensitivity and specificity of adrenal venous sampling (95% and 100%, respectively) for detecting unilateral aldosterone hypersecretion are superior to those of adrenal CT; success rates are 90–96%, and complication rates are <2.5%. One frequently used protocol involves sampling for aldosterone and cortisol levels in response to ACTH stimulation. An ipsilateral/ contralateral aldosterone ratio >4, with symmetric ACTH-stimulated cortisol levels, is indicative of unilateral aldosterone production.

Hypertension generally is responsive to surgery in patients with adenoma but not in patients with bilateral adrenal hyperplasia. Unilateral adrenalectomy, often done via a laparoscopic approach, is curative in 40–70% of patients with an adenoma. Surgery should be undertaken after blood pressure has been controlled and hypokalemia corrected. Transient hypoaldosteronism may occur up to 3 months postoperatively, resulting in hyperkalemia. Potassium should be monitored during this time, and hyperkalemia should be treated with potassium-wasting diuretics and with fludrocortisone, if needed. Patients with bilateral hyperplasia should be treated medically. The drug regimen for these patients, as well as for patients with an adenoma who are poor surgical candidates, should include an aldosterone antagonist and, if necessary, other potassium-sparing diuretics.

Glucocorticoid-remediable hyperaldosteronism is a rare, monogenic autosomal dominant disorder characterized by moderate to severe hypertension, often occurring at an early age. These patients may have a family history of hemorrhagic stroke at a young age. Hypokalemia is usually mild or absent. Normally, angiotensin II stimulates aldosterone production by the adrenal zona glomerulosa, whereas ACTH stimulates cortisol production in the zona fasciculata. Owing to a chimeric gene on chromosome 8, ACTH also regulates aldosterone secretion by the zona fasciculata in patients with glucocorticoid-remediable hyperaldosteronism. The consequence is overproduction in the zona fasciculata of both aldosterone and hybrid steroids (18-hydroxycortisol and 18-oxocortisol) due to oxidation of cortisol. The diagnosis may be established by urine excretion rates of these hybrid steroids that are 20 to 30 times normal or by direct genetic testing. Therapeutically, suppression of ACTH with low-dose glucocorticoids corrects the hyperaldosteronism, hypertension, and hypokalemia. Spironolactone is also a therapeutic option.

Cushing's Syndrome

(See also Chap. 342) Cushing's syndrome is related to excess cortisol production due either to excess ACTH secretion (from a pituitary tumor or an ectopic tumor) or to ACTH-independent adrenal production of cortisol. Hypertension occurs in 75–80% of patients with Cushing's syndrome. The mechanism of hypertension may be related to stimulation of mineralocorticoid receptors by cortisol and increased secretion of other adrenal steroids. If clinically suspected based on phenotypic characteristics, in patients not taking exogenous glucocorticoids, laboratory screening may be carried out with measurement of 24-h excretion rates of urine free cortisol or an overnight dexamethasone-suppression test. Recent evidence suggests that late night salivary cortisol is also a sensitive and convenient screening test. Further evaluation is required to confirm the diagnosis and identify the specific etiology of Cushing's syndrome. Appropriate therapy depends on the etiology.

Pheochromocytoma

(See also Chap. 343) Catecholamine-secreting tumors are located in the adrenal medulla (pheochromocytoma) or in extra-adrenal paraganglion tissue (paraganglioma) and account for hypertension in ~0.05% of patients. If unrecognized, pheochromocytoma may result in lethal cardiovascular consequences. Clinical manifestations, including hypertension, are primarily related to increased circulating catecholamines, although some of these tumors may secrete a number of other vasoactive substances. In a small percentage of patients, epinephrine is the predominant catecholamine secreted by the tumor, and these patients may present with hypotension rather than hypertension. The initial suspicion of the diagnosis is based on symptoms and/or the association of pheochromocytoma with other disorders (**Table 247-4**). Approximately 20% of pheochromocytomas are familial with autosomal dominant inheritance. Inherited pheochromocytomas may be associated with multiple endocrine

neoplasia (MEN) type 2A and type 2B, von Hippel-Lindau disease, and neurofibromatosis (Table 247-4). Each of these syndromes is related to specific, identifiable germ-line mutations. Additionally, mutations of succinate dehydrogenase genes are associated with paraganglioma syndromes, generally characterized by head and neck paragangliomas. Laboratory testing consists of measuring catecholamines in either urine or plasma. Genetic screening is available for evaluating patients and relatives suspected of harboring a pheochromocytoma associated with a familial syndrome. Surgical excision is the definitive treatment of pheochromocytoma and results in cure in ~90% of patients.

Table 247–4. Rare Mendelian Forms of Hypertension

Disease	Phenotype	Genetic Cause
Glucocorticoid-remediable hyperaldosteronism	Autosomal dominant Absent or mild hypokalemia	Chimeric 11β-hydroxylase/aldsoterone gene on chromosome 8
17α-hydroxylase deficiency	Autosomal recessive Males: pseudohermaphroditism Females: primary amenorrhea, absent secondary sexual characteristics	Random mutations of the CYP17 gene on chromosome 10
11β-hydroxylase deficiency	Autosomal recessive Masculinization	Mutations of the CYP11B1 gene on chromosome 8q21-q22
11β-hydroxysteroid dehydrogenase deficiency (apparent mineralocorticoid excess syndrome)	Autosomal recessive Hypokalemia, low renin, low aldosterone	Mutations in the 11β-hydroxysteroid dehydrogenase gene
Liddle's syndrome	Autosomal dominant Hypokalemia, low renin, low aldosterone	Mutation subunits of the epithelial sodium channel SCNN1B and SCNN1C genes
Pseudohypoaldosteronism type II (Gordon's syndrome)	Autosomal dominant Hyperkalemia, normal glomerular filtration rate	Linkage to chromosomes 1q31-q42 and 17p11-q21
Hypertension exacerbated in pregnancy	Autosomal dominant Severe hypertension in early pregnancy	Missense mutation with substitution of leucine for serine at codon 810 (MR_{L810})
Polycystic kidney disease	Autosomal dominant Large cystic kidneys, renal failure, liver cysts, cerebral aneurysms, valvular heart disease	Mutations in the PKD1 gene on chromosome 16 and PKD2 gene on chromosome 4
Pheochromocytoma	Autosomal dominant	
	(a) Multiple endocrine neoplasia, type 2A Medullary thyroid carcinoma, hyperparathyroidism	(a) Mutations in the RET protooncogene
	(b) Multiple endocrine neoplasia, type 2B Medullary thyroid carcinoma, mucosal neuromas, thickened corneal nerves, alimentary ganglioneuromatoses, marfanoid habitus	(b) Mutations in the RET protooncogene
	(c) von Hippel-Lindau disease Retinal angiomas, hemangioblastomas of the cerebellum and spinal cord, renal cell carcinoma	(c) Mutations in the VHL tumor-suppressor gene
	(d) Neurofibromatosis type 1 Multiple neurofibromas, café-au-lait spots	(d) Mutations in the NF1 tumor-suppressor gene

Miscellaneous Causes of Hypertension

Hypertension due to *obstructive sleep apnea* is being recognized with increasing frequency (Chap. 265). Independent of obesity, hypertension occurs in >50% of individuals with obstructive sleep apnea. The severity of hypertension correlates with the severity of sleep apnea. Approximately 70% of patients with obstructive sleep apnea are obese. Hypertension related to obstructive sleep apnea also should be considered in patients with drug-resistant hypertension and patients with a history of snoring. The diagnosis can be confirmed by polysomnography. In obese patients, weight loss may alleviate or cure sleep apnea and related hypertension. Continuous positive airway pressure (CPAP) administered during sleep is an effective therapy for obstructive sleep apnea. With CPAP, patients with apparently drug-resistant hypertension may be more responsive to antihypertensive agents.

Coarctation of the aorta is the most common congenital cardiovascular cause of hypertension (Chap. 236). The incidence is 1–8 per 1000 live births. It is usually sporadic but occurs in 35% of children with Turner syndrome. Even when the anatomic lesion is surgically corrected in infancy, up to 30% of patients develop subsequent hypertension and are at risk of accelerated coronary artery disease and cerebrovascular events. Patients with less severe lesions may not be diagnosed until young adulthood. The physical findings are diagnostic and include diminished and delayed femoral pulses and a systolic pressure gradient between the right arm and the

legs and, depending on the location of the coarctation, between the right and left arms. A blowing systolic murmur may be heard in the posterior left interscapular areas. The diagnosis may be confirmed by chest x-ray and transesophageal echocardiography. Therapeutic options include surgical repair and balloon angioplasty, with or without placement of an intravascular stent. Subsequently, many patients do not have a normal life expectancy but may have persistent hypertension, with death due to ischemic heart disease, cerebral hemorrhage, or aortic aneurysm.

Several additional endocrine disorders, including *thyroid diseases* and *acromegaly*, cause hypertension. Mild diastolic hypertension may be a consequence of hypothyroidism, whereas hyperthyroidism may result in systolic hypertension. *Hypercalcemia* of any etiology, the most common being primary hyperparathyroidism, may result in hypertension. Hypertension also may be related to a number of prescribed or over-the-counter *medications*.

MONOGENIC HYPERTENSION

A number of rare forms of monogenic hypertension have been identified (Table 247-4). These disorders may be recognized by their characteristic phenotypes, and in many instances the diagnosis may be confirmed by genetic analysis. Several inherited defects in adrenal steroid biosynthesis and metabolism result in mineralocorticoid-induced hypertension and hypokalemia. In patients with a 17α-hydroxylase deficiency, synthesis of sex hormones and cortisol is decreased **(Fig. 247-3)**. Consequently, these individuals do not mature sexually; males may present with pseudohermaphroditism and females with primary amenorrhea and absent secondary sexual characteristics. Because cortisol-induced negative feedback on pituitary ACTH production is diminished, ACTH-stimulated adrenal steroid synthesis proximal to the enzymatic block is increased. Hypertension and hypokalemia are consequences of increased synthesis of mineralocorticoids proximal to the enzymatic block, particularly desoxycorticosterone. Increased steroid production and, hence, hypertension may be treated with low-dose glucocorticoids. An 11β-hydroxylase deficiency results in a salt-retaining adrenogenital syndrome that occurs in 1 in 100,000 live births. This enzymatic defect results in decreased cortisol synthesis, increased synthesis of mineralocorticoids (e.g., desoxycorticosterone), and shunting of steroid biosynthesis into the androgen pathway. In the severe form, the syndrome may present early in life, including the newborn period, with virilization and ambiguous genitalia in females and penile enlargement in males, or in older children as precocious puberty and short stature. Acne, hirsutism, and menstrual irregularities may be the presenting features when the disorder is first recognized in adolescence or early adulthood. Hypertension is less common in the late-onset forms. Patients with an 11β-hydroxysteroid dehydrogenase deficiency have an impaired capacity to metabolize cortisol to its inactive metabolite, cortisone, and hypertension is related to activation of mineralocorticoid receptors by cortisol. This defect may be inherited or acquired, due to licorice-containing glycyrrhizin acid. The same substance is present in the paste of several brands of chewing tobacco. The defect in Liddle's syndrome (Chaps. 45 and 342) results from constitutive activation of amiloride-sensitive epithelial sodium channels on the distal renal tubule, resulting in excess sodium reabsorption; the syndrome is ameliorated by amiloride. Hypertension exacerbated in pregnancy (Chap. 7) is due to activation of the mineralocorticoid receptor by progesterone.

Figure 247-3

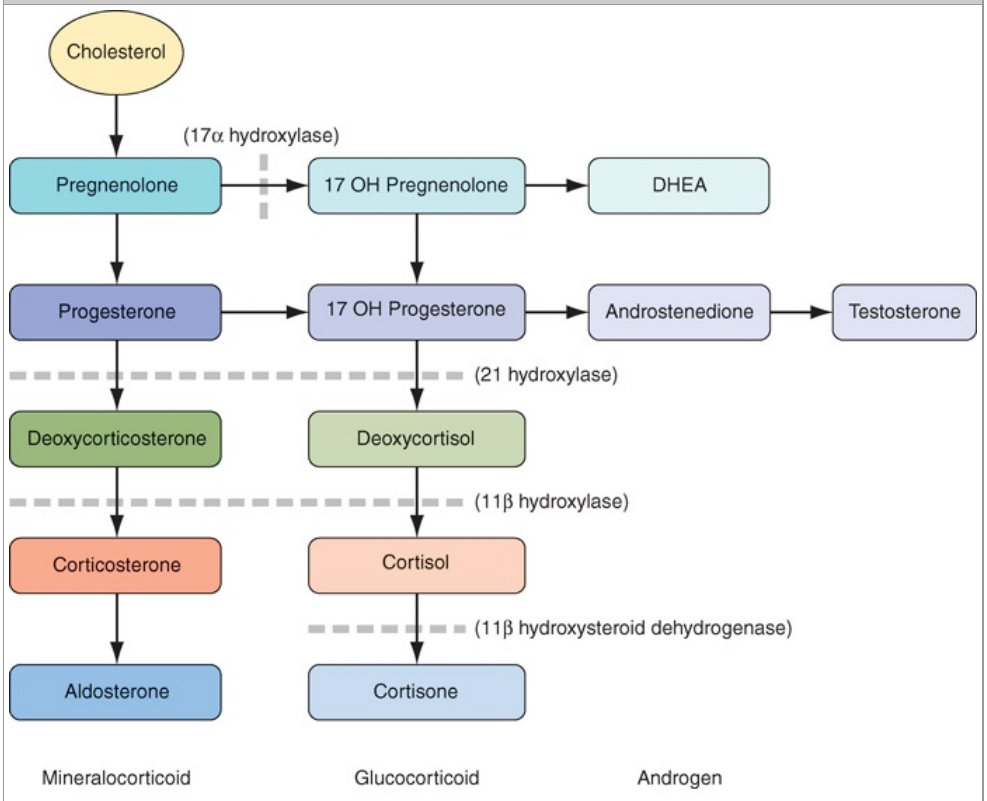

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adrenal enzymatic defects.

Approach to the Patient: Hypertension

HISTORY

The initial assessment of a hypertensive patient should include a complete history and physical examination to confirm a diagnosis of hypertension, screen for other cardiovascular disease risk factors, screen for secondary causes of hypertension, identify cardiovascular consequences of hypertension and other comorbidities, assess blood pressure–related lifestyles, and determine the potential for intervention.

Most patients with hypertension have no specific symptoms referable to their blood pressure elevation. Although popularly considered a symptom of elevated arterial pressure, headache generally occurs only in patients with severe hypertension. Characteristically, a "hypertensive headache" occurs in the morning and is localized to the occipital region. Other nonspecific symptoms that may be related to elevated blood pressure include dizziness, palpitations, easy fatigability, and impotence. When symptoms are present, they are generally related to hypertensive cardiovascular disease or to manifestations of secondary hypertension. **Table 247-5** lists salient features that should be addressed in obtaining a history from a hypertensive patient.

Table 247–5. Patient's Relevant History
Duration of hypertension
Previous therapies: responses and side effects
Family history of hypertension and cardiovascular disease Dietary and psychosocial history
Other risk factors: weight change, dyslipidemia, smoking, diabetes, physical inactivity
Evidence of secondary hypertension: history of renal disease; change in appearance; muscle weakness; spells of sweating, palpitations, tremor; erratic sleep, snoring, daytime somnolence; symptoms of hypo- or hyperthyroidism; use of agents that may increase blood pressure
Evidence of target organ damage: history of TIA, stroke, transient blindness; angina, myocardial infarction, congestive heart failure; sexual function Other comorbidities

Abbreviation: TIA, transient ischemic attack.

MEASUREMENT OF BLOOD PRESSURE

Reliable measurements of blood pressure depend on attention to the details of the technique and conditions of the measurement. Proper training of observers, positioning of the patient, and selection of cuff size are essential. Owing to recent regulations preventing the use of mercury because of concerns about its potential toxicity, most office measurements are made with aneroid sphygmomanometers or with oscillometric devices. These instruments should be calibrated periodically, and their accuracy confirmed. Before the blood pressure measurement is taken, the individual should be seated quietly in a chair (not the exam table) with feet on the floor for 5 min in a private, quiet setting with a comfortable room temperature. At least two measurements should be made. The center of the cuff should be at heart level, and the width of the bladder cuff should equal at least 40% of the arm circumference; the length of the cuff bladder should be enough to encircle at least 80% of the arm circumference. It is important to pay attention to cuff placement, stethoscope placement, and the rate of deflation of the cuff (2 mmHg/s). Systolic blood pressure is the first of at least two regular "tapping" Korotkoff sounds, and diastolic blood pressure is the point at which the last regular Korotkoff sound is heard. In current practice, a diagnosis of hypertension generally is based on seated, office measurements.

Currently available ambulatory monitors are fully automated, use the oscillometric technique, and typically are programmed to take readings every 15–30 min. Twenty-four-hour ambulatory blood pressure monitoring more reliably predicts cardiovascular disease risk than do office measurements. However, ambulatory monitoring is not used routinely in clinical practice and generally is reserved for patients in whom white coat hypertension is suspected. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) has also recommended ambulatory monitoring for treatment resistance, symptomatic hypotension, autonomic failure, and episodic hypertension.

PHYSICAL EXAMINATION

Body habitus, including weight and height, should be noted. At the initial examination, blood pressure should be measured in both arms and preferably in the supine, sitting, and standing positions to evaluate for postural hypotension. Even if the femoral pulse is normal to palpation, arterial pressure should be measured at least once in the lower extremity in patients in whom hypertension is discovered before age 30. Heart rate also should be recorded. Hypertensive individuals have an increased prevalence of atrial fibrillation. The neck should be palpated for an enlarged thyroid gland, and patients should be assessed for signs of hypo- and hyperthyroidism. Examination of blood vessels may provide clues about underlying vascular disease and should include funduscopic examination, auscultation for bruits over the carotid and femoral arteries, and palpation of femoral and pedal pulses. The retina is the only tissue in which arteries and arterioles can be examined directly. With increasing severity of hypertension and atherosclerotic disease, progressive funduscopic changes include increased arteriolar light reflex, arteriovenous crossing defects, hemorrhages and exudates, and, in patients with malignant hypertension, papilledema. Examination of the heart may reveal a loud second heart sound due to closure of the aortic valve and an S_4 gallop attributed to atrial contraction against a noncompliant left ventricle.

Left ventricular hypertrophy may be detected by an enlarged, sustained, and laterally displaced apical impulse. An abdominal bruit, particularly a bruit that lateralizes and extends throughout systole into diastole, raises the possibility of renovascular hypertension. Kidneys of patients with polycystic kidney disease may be palpable in the abdomen. The physical examination also should include evaluation for signs of CHF and a neurologic examination.

LABORATORY TESTING

Table 247-6 lists recommended laboratory tests in the initial evaluation of hypertensive patients. Repeat measurements of renal function, serum electrolytes, fasting glucose, and lipids may be obtained after the introduction of a new antihypertensive agent and then annually or more frequently if clinically indicated. More extensive laboratory testing is appropriate for patients with apparent drug-resistant hypertension or when the clinical evaluation suggests a secondary form of hypertension.

Table 247–6. Basic Laboratory Tests for Initial Evaluation

System	Test
Renal	Microscopic urinalysis, albumin excretion, serum BUN and/or creatinine
Endocrine	Serum sodium, potassium, calcium, ?TSH
Metabolic	Fasting blood glucose, total cholesterol, HDL and LDL (often computed) cholesterol, triglycerides
Other	Hematocrit, electrocardiogram

Abbreviations: BUN, blood urea nitrogen; HDL, LDL, high-/low-density lipoprotein; TSH, thyroid-stimulating hormone.

Treatment Hypertension

LIFESTYLE INTERVENTIONS

Implementation of lifestyles that favorably affect blood pressure has implications for both the prevention and the treatment of hypertension. Health-promoting lifestyle modifications are recommended for individuals with prehypertension and as an adjunct to drug therapy in hypertensive individuals. These interventions should address overall cardiovascular disease risk. Although the impact of lifestyle interventions on blood pressure is more pronounced in persons with hypertension, in short-term trials, weight loss and reduction of dietary NaCl have been shown to prevent the development of hypertension. In hypertensive individuals, even if these interventions do not produce a sufficient reduction in blood pressure to avoid drug therapy, the number of medications or doses required for blood pressure control may be reduced. Dietary modifications that effectively lower blood pressure are weight loss, reduced NaCl intake, increased potassium intake, moderation of alcohol consumption, and an overall healthy dietary pattern **(Table 247-7)**.

Table 247–7. Lifestyle Modifications to Manage Hypertension

Weight reduction	Attain and maintain BMI <25 kg/m^2
Dietary salt reduction	<6 g NaCl/d
Adapt DASH-type dietary plan	Diet rich in fruits, vegetables, and low-fat dairy products with reduced content of saturated and total fat
Moderation of alcohol consumption	For those who drink alcohol, consume ≤2 drinks/day in men and ≤1 drink/day in women
Physical activity	Regular aerobic activity, e.g., brisk walking for 30 min/d

Abbreviations: BMI, body mass index; DASH, Dietary Approaches to Stop Hypertension (trial).

Prevention and treatment of obesity are important for reducing blood pressure and cardiovascular disease risk. In short-term trials, even modest weight loss can lead to a reduction of blood pressure and an increase in insulin sensitivity. Average blood pressure reductions of 6.3/3.1 mmHg have been observed with a reduction in mean body weight of 9.2 kg. Regular physical activity facilitates weight loss, decreases blood pressure, and reduces the overall risk of cardiovascular disease. Blood pressure may be lowered by 30 min of moderately intense physical activity, such as brisk walking, 6–7 days a week, or by more intense, less frequent workouts.

There is individual variability in the sensitivity of blood pressure to NaCl, and this variability may have a genetic basis. Based on results of meta-analyses, lowering of blood pressure by limiting daily NaCl intake to 4.4–7.4 g (75–125 meq) results in blood pressure reductions of 3.7–4.9/0.9–2.9 mmHg in hypertensive individuals and lesser reductions in normotensive individuals. Dietary NaCl reduction also has been shown to reduce the long-term risk of cardiovascular events in adults with "prehypertension." Potassium and calcium supplementation have inconsistent, modest antihypertensive effects, and, independent of blood pressure, potassium supplementation may be associated with reduced stroke mortality. Alcohol use in persons consuming three or more drinks per day (a standard drink contains ~14 g ethanol) is associated with higher blood pressures, and a reduction of alcohol consumption is associated with a reduction of blood pressure. In patients with advanced renal disease, dietary protein restriction may have a modest effect in mitigating renal damage by reducing the intrarenal transmission of systemic arterial pressure.

The DASH (Dietary Approaches to Stop Hypertension) trial convincingly demonstrated that over an 8-week period a diet high in fruits, vegetables, and low-fat dairy products lowers blood pressure in individuals with high-normal blood pressures or mild hypertension. Reduction of daily NaCl intake to <6 g (100 meq) augmented the effect of this diet on blood pressure. Fruits and vegetables are enriched sources of potassium, magnesium, and fiber, and dairy products are an important source of calcium.

PHARMACOLOGIC THERAPY

Drug therapy is recommended for individuals with blood pressures ≥140/90 mmHg. The degree of benefit derived from antihypertensive agents is related to the magnitude of the blood pressure reduction. Lowering systolic blood

pressure by 10–12 mmHg and diastolic blood pressure by 5–6 mmHg confers relative risk reductions of 35–40% for stroke and 12–16% for CHD within 5 years of the initiation of treatment. Risk of heart failure is reduced by >50%. Hypertension control is the single most effective intervention for slowing the rate of progression of hypertension-related chronic kidney disease.

There is considerable variation in individual responses to different classes of antihypertensive agents, and the magnitude of response to any single agent may be limited by activation of counterregulatory mechanisms that oppose the hypotensive effect of the agent. Most available agents reduce systolic blood pressure by 7–13 mmHg and diastolic blood pressure by 4–8 mmHg when corrected for placebo effect. More often than not, combinations of agents, with complementary antihypertensive mechanisms, are required to achieve goal blood pressure reductions. Selection of antihypertensive agents and combinations of agents should be individualized, taking into account age, severity of hypertension, other cardiovascular disease risk factors, comorbid conditions, and practical considerations related to cost, side effects, and frequency of dosing **(Table 247-8)**.

Table 247–8. Examples of Oral Drugs Used in Treatment of Hypertension

Drug Class	Examples	Usual Total Daily Dose* (Dosing Frequency/Day)	Other Indications	Contraindications/Cautions
Diuretics				
Thiazides	Hydrochlorothiazide	6.25–50 mg (1–2)		Diabetes, dyslipidemia, hyperuricemia, gout, hypokalemia
	Chlorthalidone	25–50 mg (1)		
Loop diuretics	Furosemide	40–80 mg (2–3)	CHF due to systolic dys-function, renal failure	Diabetes, dyslipidemia, hyperuricemia, gout, hypokalemia
	Ethacrynic acid	50–100 mg (2–3)		
Aldosterone antagonists	Spironolactone	25–100 mg (1–2)	CHF due to systolic dysfunction, primary aldosteronism	Renal failure, hyperkalemia
	Eplerenone	50–100 mg (1–2)		
K^+ retaining	Amiloride	5–10 mg (1–2)		Renal failure, hyperkalemia
	Triamterene	50–100 mg (1–2)		
Beta blockers				
Cardioselective	Atenolol	25–100 mg (1)	Angina, CHF due to systolic dysfunction, post-MI, sinus tachycardia, ventricular tachyarrhythmias	Asthma, COPD, 2nd- or 3rd-degree heart block, sick-sinus syndrome
	Metoprolol	25–100 mg (1–2)		
Nonselective	Propranolol	40–160 mg (2)		
	Propranolol LA	60–180 (1)		
Combined alpha/beta	Labetalol	200–800 mg (2)	?Post-MI, CHF	
	Carvedilol	12.5–50 mg (2)		
Alpha antagonists				
Selective	Prazosin	2–20 mg (2–3)	Prostatism	
	Doxazosin	1–16 mg (1)		
	Terazosin	1–10 mg (1–2)		
Nonselective	Phenoxybenzamine	20–120 mg (2–3)	Pheochromocytoma	
Sympatholytics				
Central	Clonidine	0.1–0.6 mg (2)		
	Clonidine patch	0.1–0.3 mg (1/week)		
	Methyldopa	250–1000 mg (2)		
	Reserpine	0.05–0.25 mg (1)		
	Guanfacine	0.5–2 mg (1)		
ACE inhibitors	Captopril	25–200 mg (2)	Post-MI, coronary syndromes, CHF with low ejection fraction, nephropathy	Acute renal failure, bilateral renal artery stenosis, pregnancy, hyperkalemia
	Lisinopril	10–40 mg (1)		
	Ramipril	2.5–20 mg (1–2)		
Angiotensin II antagonists	Losartan	25–100 mg (1–2)	CHF with low ejection fraction, nephropathy, ACE inhibitor cough	Renal failure, bilateral renal artery stenosis, pregnancy, hyperkalemia
	Valsartan	80–320 mg (1)		
	Candesartan	2–32 mg (1–2)		
Renin inhibitors	Aliskiren	150–300 mg (1)	Diabetic nephropathy	Pregnancy
Calcium antagonists				
Dihydropyridines	Nifedipine (long-	30–60 mg (1)		

	acting)			
Nondihydropyridines	Verapamil (long-acting)	120–360 mg (1–2)	Post-MI, supraventricular tachycardias, angina	2nd- or 3rd-degree heart block
	Diltiazem (long-acting)	180-420 mg (1)		
Direct vasodilators	Hydralazine	25–100 mg (2)		Severe coronary artery disease
	Minoxidil	2.5–80 mg (1–2)		

*At the initiation of therapy, lower doses may be preferable for elderly patients and for select combinations of antihypertensive agents.

Abbreviations: ACE, angiotensin-converting enzyme; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; MI, myocardial infarction.

Diuretics

Low-dose thiazide diuretics often are used as first-line agents alone or in combination with other antihypertensive drugs. Thiazides inhibit the Na^+/Cl^- pump in the distal convoluted tubule and hence increase sodium excretion. In the long term, they also may act as vasodilators. Thiazides are safe, efficacious, inexpensive, and reduce clinical events. They provide additive blood pressure–lowering effects when combined with beta blockers, angiotensin-converting enzyme inhibitors (ACEIs), or angiotensin receptor blockers (ARBs). In contrast, addition of a diuretic to a calcium channel blocker is less effective. Usual doses of hydrochlorothiazide range from 6.25–50 mg/d. Owing to an increased incidence of metabolic side effects (hypokalemia, insulin resistance, increased cholesterol), higher doses generally are not recommended. Two potassium-sparing diuretics, amiloride and triamterene, act by inhibiting epithelial sodium channels in the distal nephron. These agents are weak antihypertensive agents but may be used in combination with a thiazide to protect against hypokalemia. The main pharmacologic target for loop diuretics is the Na^+-K^+-$2Cl^-$ cotransporter in the thick ascending limb of the loop of Henle. Loop diuretics generally are reserved for hypertensive patients with reduced glomerular filtration rates [reflected in serum creatinine >220 μmol/L (>2.5 mg/dL)], CHF, or sodium retention and edema for some other reason, such as treatment with a potent vasodilator, e.g., minoxidil.

Blockers of the Renin-Angiotensin System

ACEIs decrease the production of angiotensin II, increase bradykinin levels, and reduce sympathetic nervous system activity. ARBs provide selective blockade of AT_1 receptors, and the effect of angiotensin II on unblocked AT_2 receptors may augment their hypotensive effect. Both classes of agents are effective antihypertensive agents that may be used as monotherapy or in combination with diuretics, calcium antagonists, and alpha blocking agents. ACEIs and ARBs have been shown to improve insulin action and ameliorate the adverse effects of diuretics on glucose metabolism. Although the overall impact on the incidence of diabetes is modest, compared with amlodipine (a calcium antagonist), valsartan (an ARB) has been shown to reduce the risk of developing diabetes in high-risk hypertensive patients. ACEI/ARB combinations are less effective in lowering blood pressure than is the case when either class of these agents is used in combination with other classes of agents. In patients with vascular disease or a high risk of diabetes, combination ACEI/ARB therapy has been associated with more adverse events (e.g., cardiovascular death, myocardial infarction, stroke, and hospitalization for heart failure) without increases in benefit. However, in hypertensive patients with proteinuria, preliminary data suggest that reduction of proteinuria with ACEI/ARB combination treatment may be more effective than treatment with either agent alone.

Side effects of ACEIs and ARBs include functional renal insufficiency due to efferent renal arteriolar dilation in a kidney with a stenotic lesion of the renal artery. Additional predisposing conditions to renal insufficiency induced by these agents include dehydration, CHF, and use of nonsteroidal anti-inflammatory drugs. Dry cough occurs in ~15% of patients, and angioedema occurs in <1% of patients taking ACEIs. Angioedema occurs most commonly in individuals of Asian origin and more commonly in African Americans than in whites. Hyperkalemia due to hypoaldosteronism is an occasional side effect of both ACEIs and ARBs.

A new approach to blocking the renin-angiotensin system has been introduced into clinical practice for the treatment of hypertension: direct renin inhibitors. Blockade of the renin-angiotensin system is more complete with renin inhibitors than with ACEIs or ARBs. Aliskiren is the first of a class of oral, nonpeptide competitive inhibitors of the enzymatic activity of renin. Monotherapy with aliskiren seems to be as effective as an ACEI or ARB for lowering blood pressure, but not more effective. Further blood reductions may be achieved when aliskiren is used in combination with a thiazide diuretic, an ACEI, an ARB, or calcium antagonists. Currently, aliskiren is not considered a first-line antihypertensive agent.

Aldosterone Antagonists

Spironolactone is a nonselective aldosterone antagonist that may be used alone or in combination with a thiazide diuretic. It may be a particularly effective agent in patients with low-renin essential hypertension, resistant hypertension, and primary aldosteronism. In patients with CHF, low-dose spironolactone reduces mortality and hospitalizations for heart failure when given in addition to conventional therapy with ACEIs, digoxin, and loop diuretics. Because spironolactone binds to progesterone and androgen receptors, side effects may include gynecomastia, impotence, and menstrual abnormalities. These side effects are circumvented by a newer agent, eplerenone, which is a selective aldosterone antagonist. Eplerenone has recently been approved in the United States for the treatment of hypertension.

Beta Blockers

β-Adrenergic receptor blockers lower blood pressure by decreasing cardiac output, due to a reduction of heart

rate and contractility. Other proposed mechanisms by which beta blockers lower blood pressure include a central nervous system effect and inhibition of renin release. Beta blockers are particularly effective in hypertensive patients with tachycardia, and their hypotensive potency is enhanced by coadministration with a diuretic. In lower doses, some beta blockers selectively inhibit cardiac β_1 receptors and have less influence on β_2 receptors on bronchial and vascular smooth muscle cells; however, there seems to be no difference in the antihypertensive potencies of cardioselective and nonselective beta blockers. Certain beta blockers have intrinsic sympathomimetic activity, and it is uncertain whether this constitutes an overall advantage or disadvantage in cardiac therapy. Beta blockers without intrinsic sympathomimetic activity decrease the rate of sudden death, overall mortality, and recurrent myocardial infarction. In patients with CHF, beta blockers have been shown to reduce the risks of hospitalization and mortality. Carvedilol and labetalol block both β receptors and peripheral α-adrenergic receptors. The potential advantages of combined β- and α-adrenergic blockade in treating hypertension remain to be determined.

α-Adrenergic Blockers

Postsynaptic, selective α-adrenoreceptor antagonists lower blood pressure by decreasing peripheral vascular resistance. They are effective antihypertensive agents used either as monotherapy or in combination with other agents. However, in clinical trials of hypertensive patients, alpha blockade has not been shown to reduce cardiovascular morbidity and mortality or to provide as much protection against CHF as other classes of antihypertensive agents. These agents are also effective in treating lower urinary tract symptoms in men with prostatic hypertrophy. Nonselective α-adrenoreceptor antagonists bind to postsynaptic and presynaptic receptors and are used primarily for the management of patients with pheochromocytoma.

Sympatholytic Agents

Centrally acting α_2 sympathetic agonists decrease peripheral resistance by inhibiting sympathetic outflow. They may be particularly useful in patients with autonomic neuropathy who have wide variations in blood pressure due to baroreceptor denervation. Drawbacks include somnolence, dry mouth, and rebound hypertension on withdrawal. Peripheral sympatholytics decrease peripheral resistance and venous constriction by depleting nerve terminal norepinephrine. Although they are potentially effective antihypertensive agents, their usefulness is limited by orthostatic hypotension, sexual dysfunction, and numerous drug-drug interactions.

Calcium Channel Blockers

Calcium antagonists reduce vascular resistance through L-channel blockade, which reduces intracellular calcium and blunts vasoconstriction. This is a heterogeneous group of agents that includes drugs in the following three classes: phenylalkylamines (verapamil), benzothiazepines (diltiazem), and 1,4-dihydropyridines (nifedipine-like). Used alone and in combination with other agents (ACEIs, beta blockers, α_1-adrenergic blockers), calcium antagonists effectively lower blood pressure; however, it is unclear if adding a diuretic to a calcium blocker results in a further lowering of blood pressure. Side effects of flushing, headache, and edema with dihydropyridine use are related to their potencies as arteriolar dilators; edema is due to an increase in transcapillary pressure gradients, not to net salt and water retention.

Direct Vasodilators

Direct vasodilators decrease peripheral resistance and concomitantly activate mechanisms that defend arterial pressure, notably the sympathetic nervous system, the renin-angiotensin-aldosterone system, and sodium retention. Usually, they are not considered first-line agents but are most effective when added to a combination that includes a diuretic and a beta blocker. Hydralazine is a potent direct vasodilator that has antioxidant and nitric oxide–enhancing actions, and minoxidil is a particularly potent agent and is used most frequently in patients with renal insufficiency who are refractory to all other drugs. Hydralazine may induce a lupus-like syndrome, and side effects of minoxidil include hypertrichosis and pericardial effusion.

COMPARISONS OF ANTIHYPERTENSIVES

Based on pooling results from clinical trials, meta-analyses of the efficacy of different classes of antihypertensive agents suggest essentially equivalent blood pressure–lowering effects of the following six major classes of antihypertensive agents when used as monotherapy: thiazide diuretics, beta blockers, ACEIs, ARBs, calcium antagonists, and α_2 blockers. On average, standard doses of most antihypertensive agents reduce blood pressure by 8–10/4–7 mmHg; however, there may be subgroup differences in responsiveness. Younger patients may be more responsive to beta blockers and ACEIs, whereas patients over age 50 may be more responsive to diuretics and calcium antagonists. There is a limited relationship between plasma renin and blood pressure response. Patients with high-renin hypertension may be more responsive to ACEIs and ARBs than to other classes of agents, whereas patients with low-renin hypertension are more responsive to diuretics and calcium antagonists. Hypertensive African Americans tend to have low renin and may require higher doses of ACEIs and ARBs than whites for optimal blood pressure control, although this difference is abolished when these agents are combined with a diuretic. Beta blockers also appear to be less effective than thiazide diuretics in African Americans than in non-African Americans. Identification of genetic variants that influence blood pressure responsiveness would potentially provide a rational basis for the selection of a specific class of an antihypertensive agent in an individual patient. Early pharmacogenetic studies, utilizing either a candidate gene approach or genomewide scans, have shown associations of gene polymorphisms with blood pressure responsiveness to specific antihypertensive drugs. However, the reported effects have generally been too small to affect clinical decisions, and associated polymorphisms remain to be confirmed in subsequent studies. Currently, in practical terms, the presence of comorbidities often influences the selection of antihypertensive agents.

A recent meta-analysis of more than 30 randomized trials of blood pressure–lowering therapy indicates that for a given reduction in blood pressure, the major drug classes seem to produce similar overall net effects on total cardiovascular events. In both nondiabetic and diabetic hypertensive patients, most trials have failed to show

significant differences in cardiovascular outcomes with different drug regimens as long as equivalent decreases in blood pressure were achieved. For example, the Antihypertensive and Lipid-Lowering Treatment to prevent Heart Attack Trial (ALLHAT) demonstrated that the occurrence of coronary heart disease death and nonfatal myocardial infarction, as well as overall mortality, was virtually identical in hypertensive patients treated with either an ACEI (lisinopril), a diuretic (chlorthalidone), or a calcium antagonist (amlodipine).

However, in specific patient groups, ACEIs may have particular advantages, beyond that of blood pressure control, in reducing cardiovascular and renal outcomes. ACEIs and ARBs decrease intraglomerular pressure and proteinuria and may retard the rate of progression of renal insufficiency, not totally accounted for by their hypotensive effects, in both diabetic and nondiabetic renal diseases. Among African Americans with hypertension-related renal disease, ACEIs appear to be more effective than beta blockers or dihydropyridine calcium channel blockers in slowing, although not preventing, the decline of glomerular filtration rate. In experimental models of hypertension and diabetes, renal protection with aliskiren (a renin inhibitor) was comparable to that with ACEIs and ARBs. Independent of its blood pressure–lowering effect, aliskiren has renal protective effects in patients with hypertension, type 2 diabetes, and nephropathy. The renoprotective effect of these renin-angiotensin blockers, compared with other antihypertensive drugs, is less obvious at lower blood pressures. In most patients with hypertension and heart failure due to systolic and/or diastolic dysfunction, the use of diuretics, ACEIs or ARBs, and beta blockers is recommended to improve survival. Independent of blood pressure, in both hypertensive and normotensive individuals, ACEIs attenuate the development of left ventricular hypertrophy, improve symptomatology and risk of death from CHF, and reduce morbidity and mortality rates in post-myocardial infarction patients. Similar benefits in cardiovascular morbidity and mortality rates in patients with CHF have been observed with the use of ARBs. ACEIs provide better coronary protection than do calcium channel blockers, whereas calcium channel blockers provide more stroke protection than do either ACEIs or beta blockers. Results of a recent large, double-blind prospective clinical trial [Rationale and Design of the Avoiding Cardiovascular Events through Combination Therapy in Patients Living with Systolic Hypertension (ACCOMPLISH Trial)] indicated that combination treatment with an ACEI (benazepril) plus a calcium antagonist (amlodipine) was superior to treatment with the ACEI plus a diuretic (hydrochlorothiazide) in reducing the risk of cardiovascular events and death among high-risk patients with hypertension. However, the combination of an ACEI and a diuretic has recently been shown to produce major reductions in morbidity and mortality in the very elderly.

After a stroke, combination therapy with an ACEI and a diuretic, but not with an ARB, reduces the rate of recurrent stroke. Some of these apparent differences may reflect differences in trial design and/or patient groups.

BLOOD PRESSURE GOALS OF ANTIHYPERTENSIVE THERAPY

Based on clinical trial data, the maximum protection against combined cardiovascular endpoints is achieved with pressures <135–140 mmHg for systolic blood pressure and <80–85 mmHg for diastolic blood pressure; however, treatment has not reduced cardiovascular disease risk to the level in nonhypertensive individuals. More aggressive blood pressure targets for blood pressure control (e.g., office or clinic blood pressure < 130/80 mmHg) are generally recommended for patients with diabetes, coronary heart disease, chronic kidney disease, or additional cardiovascular disease risk factors. An even lower goal blood pressure (systolic blood pressure ~120 mmHg) may be desirable for patients with proteinuria (>1 g/d) since the decline of glomerular filtration rate in these patients is particularly blood pressure–dependent. In diabetic patients, effective blood pressure control reduces the risk of cardiovascular events and death as well as the risk for microvascular disease (nephropathy, retinopathy). Risk reduction is greater in diabetic than in nondiabetic individuals. Although the optimal target blood pressure in patients with heart failure has not been established, a reasonable goal is the lowest blood pressure that is not associated with evidence of hypoperfusion.

To achieve recommended blood pressure goals, the majority of individuals with hypertension will require treatment with more than one drug. Three or more drugs frequently are needed in patients with diabetes and renal insufficiency. For most agents, reduction of blood pressure at half-standard doses is only ~20% less than at standard doses. Appropriate combinations of agents at these lower doses may have additive or almost additive effects on blood pressure with a lower incidence of side effects.

Despite theoretical concerns about decreasing cerebral, coronary, and renal blood flow by overly aggressive antihypertensive therapy, clinical trials have found no evidence for a "J-curve" phenomenon; i.e., at blood pressure reductions achieved in clinical practice, there does *not* appear to be a lower threshold for increasing cardiovascular risk. A small nonprogressive increase in the serum creatinine concentration with blood pressure reduction may occur in patients with chronic renal insufficiency. This generally reflects a hemodynamic response, not structural renal injury, indicating that intraglomerular pressure has been reduced. Blood pressure control should not be allowed to deteriorate in order to prevent a modest rise in creatinine. Even among older patients with isolated systolic hypertension, further lowering of diastolic blood pressure does not result in harm. However, relatively little information is available concerning the risk-versus-benefit ratio of antihypertensive therapy in individuals >80 years, and in this population, gradual blood pressure reduction to less aggressive target levels of control may be appropriate.

The term *resistant hypertension* refers to patients with blood pressures persistently >140/90 mmHg despite taking three or more antihypertensive agents, including a diuretic, in a reasonable combination and at full doses. Resistant or difficult-to-control hypertension is more common in patients >60 years than in younger patients. Resistant hypertension may be related to "pseudoresistance" (high office blood pressures and lower home blood pressures), nonadherence to therapy, identifiable causes of hypertension (including obesity and excessive alcohol intake), and the use of any of a number of nonprescription and prescription drugs (Table 247-3). Rarely, in older patients, pseudohypertension may be related to the inability to measure blood pressure accurately in severely sclerotic arteries. This condition is suggested if the radial pulse remains palpable despite occlusion of the brachial artery by the cuff (Osler maneuver). The actual blood pressure can be determined by direct intra-arterial measurement. Evaluation of patients with resistant hypertension might include home blood pressure monitoring to determine if office blood pressures are representative of the usual blood pressure. A more extensive evaluation for a secondary form of hypertension should be undertaken if no other explanation for

hypertension resistance becomes apparent.

HYPERTENSIVE EMERGENCIES

Probably due to the widespread availability of antihypertensive therapy, in the United States there has been a decline in the numbers of patients presenting with "crisis levels" of blood pressure. Most patients who present with severe hypertension are chronically hypertensive, and in the absence of acute end organ damage, precipitous lowering of blood pressure may be associated with significant morbidity and should be avoided. The key to successful management of severe hypertension is to differentiate hypertensive crises from hypertensive urgencies. The degree of target organ damage, rather than the level of blood pressure alone, determines the rapidity with which blood pressure should be lowered. **Tables 247-9** and **247-10** list a number of hypertension-related emergencies and recommended therapies.

Table 247–9. Preferred Parenteral Drugs for Selected Hypertensive Emergencies

Hypertensive encephalopathy	Nitroprusside, nicardipine, labetalol
Malignanthypertension (when IV therapy is indicated)	Labetalol, nicardipine, nitroprusside, enalaprilat
Stroke	Nicardipine, labetalol, nitroprusside
Myocardial infarction/unstable angina	Nitroglycerin, nicardipine, labetalol, esmolol
Acute left ventricular failure	Nitroglycerin, enalaprilat, loop diuretics
Aortic dissection	Nitroprusside, esmolol, labetalol
Adrenergic crisis	Phentolamine, nitroprusside
Postoperative hypertension	Nitroglycerin, nitroprusside, labetalol, nicardipine
Preeclampsia/eclampsia of pregnancy	Hydralazine, labetalol, nicardipine

Source: Adapted from DG Vidt, in S Oparil, MA Weber (eds): *Hypertension*, 2nd ed. Philadelphia, Elsevier Saunders, 2005.

Table 247–10. Usual Intravenous Doses of Antihypertensive Agents Used in Hypertensive Emergencies.*

Antihypertensive Agent	Intravenous Dose
Nitroprusside	Initial 0.3 (μg/kg)/min; usual 2–4 (μg/kg)/min; maximum 10 (μg/kg)/min for 10 min
Nicardipine	Initial 5 mg/h; titrate by 2.5 mg/h at 5–15 min intervals; max 15 mg/h
Labetalol	2 mg/min up to 300 mg or 20 mg over 2 min, then 40–80 mg at 10-min intervals up to 300 mg total
Labetalol	2 mg/min up to 300 mg or 20 mg over 2 min, then 40–80 mg at 10-min intervals up to 300 mg total
Esmolol	Initial 80–500 μg/kg over 1 min, then 50–300 (μg/kg)/min
Phentolamine	5–15 mg bolus
Nitroglycerin	Initial 5 μg/min, then titrate by 5 μg/min at 3–5-min intervals; if no response is seen at 20 μg/min, incremental increases of 10–20 μg/min may be used
Hydralazine	10–50 mg at 30-min intervals

*Constant blood pressure monitoring is required. Start with the lowest dose. Subsequent doses and intervals of administration should be adjusted according to the blood pressure response and duration of action of the specific agent.

Malignanthypertension is a syndrome associated with an abrupt increase of blood pressure in a patient with underlying hypertension or related to the sudden onset of hypertension in a previously normotensive individual. The absolute level of blood pressure is not as important as its rate of rise. Pathologically, the syndrome is associated with diffuse necrotizing vasculitis, arteriolar thrombi, and fibrin deposition in arteriolar walls. Fibrinoid necrosis has been observed in arterioles of kidney, brain, retina, and other organs. Clinically, the syndrome is recognized by progressive retinopathy (arteriolar spasm, hemorrhages, exudates, and papilledema), deteriorating renal function with proteinuria, microangiopathic hemolytic anemia, and encephalopathy. In these patients, historic inquiry should include questions about the use of monamine oxidase inhibitors and recreational drugs (e.g., cocaine, amphetamines).

Although blood pressure should be lowered rapidly in patients with hypertensive encephalopathy, there are inherent risks of overly aggressive therapy. In hypertensive individuals, the upper and lower limits of autoregulation of cerebral blood flow are shifted to higher levels of arterial pressure, and rapid lowering of blood pressure to below the lower limit of autoregulation may precipitate cerebral ischemia or infarction as a consequence of decreased cerebral blood flow. Renal and coronary blood flows also may decrease with overly aggressive acute therapy. The initial goal of therapy is to reduce mean arterial blood pressure by no more than 25% within minutes to 2 h or to a blood pressure in the range of 160/100–110 mmHg. This may be accomplished with IV nitroprusside, a short-acting vasodilator with a rapid onset of action that allows for minute-to-minute

control of blood pressure. Parenteral labetalol and nicardipine are also effective agents for the treatment of hypertensive encephalopathy.

In patients with malignanthypertension without encephalopathy or another catastrophic event, it is preferable to reduce blood pressure over hours or longer rather than minutes. This goal may effectively be achieved initially with frequent dosing of short-acting oral agents such as captopril, clonidine, and labetalol.

Acute, transient blood pressure elevations that last days to weeks frequently occur after thrombotic and hemorrhagic strokes. Autoregulation of cerebral blood flow is impaired in ischemic cerebral tissue, and higher arterial pressures may be required to maintain cerebral blood flow. Although specific blood pressure targets have not been defined for patients with acute cerebrovascular events, aggressive reductions of blood pressure are to be avoided. With the increasing availability of improved methods for measuring cerebral blood flow (using CT technology), studies are in progress to evaluate the effects of different classes of antihypertensive agents on both blood pressure and cerebral blood flow after an acute stroke. Currently, in the absence of other indications for acute therapy, for patients with cerebral infarction who are not candidates for thrombolytic therapy, one recommended guideline is to institute antihypertensive therapy only for patients with a systolic blood pressure >220 mmHg or a diastolic blood pressure >130 mmHg. If thrombolytic therapy is to be used, the recommended goal blood pressure is <185 mmHg systolic pressure and <110 mmHg diastolic pressure. In patients with hemorrhagic stroke, suggested guidelines for initiating antihypertensive therapy are systolic >180 mmHg or diastolic pressure >130 mmHg. The management of hypertension after subarachnoid hemorrhage is controversial. Cautious reduction of blood pressure is indicated if mean arterial pressure is >130 mmHg.

In addition to pheochromocytoma, an adrenergic crisis due to catecholamine excess may be related to cocaine or amphetamine overdose, clonidine withdrawal, acute spinal cord injuries, and an interaction of tyramine-containing compounds with monamine oxidase inhibitors. These patients may be treated with phentolamine or nitroprusside.

Treatment of hypertension in patients with acute aortic dissection is discussed in Chap. 248, and treatment of hypertension in pregnancy is discussed in Chap. 7.

FURTHER READINGS

Accord Study Group: Effects of intensive blood pressure control in type 2 diabetes mellitus. N Engl J Med 362:1575, 2010

Adrogue JH, Madias NE: Sodium and potassium in the pathogenesis of hypertension. N Engl J Med 356:1966, 2007[PMID: 17494929]

Allhat Collaborative Research Group: Major outcomes in high-risk hypertensive patients randomized to angiotensin converting enzyme inhibitor or calcium channel blocker vs. diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). JAMA 288:2981, 2002

Appel LJ et al: Dietary approaches to prevent and treat hypertension: A scientific statement from the American Heart Association. Hypertension 47:296, 2006[PMID: 16434724]

Blood Pressure Lowering Treatment Trialists' Collaboration: Effects of different blood pressure-lowering regimens on major cardiovascular events in individuals with and without diabetes mellitus. Arch Intern Med 165:1410, 2005

Casas JP et al: Effect of inhibitors of the renin-angiotensin system and other antihypertensive drugs on renal outcomes: Systematic review and meta-analysis. Lancet 366:2026, 2005[PMID: 16338452]

Chobanian AV et al: The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. JAMA 289:2560, 2003[PMID: 12748199]

Law MR et al: Value of low dose combination treatment with blood pressure lowering drugs: Analysis of 354 randomised trials. Br Med J 326:1427, 2003[PMID: 12829555]

Lawes CMM et al: Global burden of blood pressure-related disease, 2001. Lancet 371:1513, 2008[PMID: 18456100]

Mancia G: Role of outcome trials in providing information on antihypertensive treatment: Importance and limitations. Am J Hypertens 19:1, 2006[PMID: 16461181]

Moser M, Setano JF: Resistant or difficult-to-control hypertension. N Engl J Med 355:385, 2006[PMID: 16870917]

Ong KL et al; Prevalence, awareness, treatment, and control of hypertension among United States adults 1999–2004. Hypertension 49:69, 2007[PMID: 17159087]

Pickering TG et al: Recommendations for blood pressure measurement in humans and experimental animals. Part 1: Blood pressure measurement in humans. A statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. Hypertension 45:142, 2005[PMID: 15611362]

———: Ambulatory blood-pressure monitoring. N Engl J Med 3554:2368, 2006

Textor SC: Current approaches to renovascular hypertension. Med Clin North Am 93:717, 2009[PMID: 19427501]

Weber MA, Materson BJ: Hypertension guidelines: A major reappraisal critically examines the available evidence. J Clin Hypertens (Greenwich) 12:229, 2010[PMID: 20433542]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISEASES OF THE AORTA: INTRODUCTION

The aorta is the conduit through which blood ejected from the left ventricle is delivered to the systemic arterial bed. In adults, its diameter is approximately 3 cm at the origin and in the ascending portion, 2.5 cm in the descending portion in the thorax, and 1.8–2 cm in the abdomen. The aortic wall consists of a thin intima composed of endothelium, subendothelial connective tissue, and an internal elastic lamina; a thick tunica media composed of smooth muscle cells and extracellular matrix; and an adventitia composed primarily of connective tissue enclosing the vasa vasorum and nervi vascularis. In addition to the conduit function of the aorta, its viscoelastic and compliant properties serve a buffering function. The aorta is distended during systole to allow a portion of the stroke volume and elastic energy to be stored, and it recoils during diastole so that blood continues to flow to the periphery. Because of its continuous exposure to high pulsatile pressure and shear stress, the aorta is particularly prone to injury and disease resulting from mechanical trauma. The aorta is also more prone to rupture than is any other vessel, especially with the development of aneurysmal dilation, since its wall tension, as governed by Laplace's law (i.e., proportional to the product of pressure and radius), will be increased.

CONGENITAL ANOMALIES OF THE AORTA

Congenital anomalies of the aorta usually involve the aortic arch and its branches. Symptoms such as dysphagia, stridor, and cough may occur if an anomaly causes a ring around or otherwise compresses the esophagus or trachea. Anomalies associated with symptoms include double aortic arch, origin of the right subclavian artery distal to the left subclavian artery, and right-sided aortic arch with an aberrant left subclavian artery. A Kommerell's diverticulum is an anatomic remnant of a right aortic arch. Most congenital anomalies of the aorta do not cause symptoms and are detected during catheter-based procedures. The diagnosis of suspected congenital anomalies of the aorta typically is confirmed by computed tomographic (CT) or magnetic resonance (MR) angiography. Surgery is used to treat symptomatic anomalies.

AORTIC ANEURYSM

An *aneurysm* is defined as a pathologic dilation of a segment of a blood vessel. A *true aneurysm* involves all three layers of the vessel wall and is distinguished from a *pseudoaneurysm*, in which the intimal and medial layers are disrupted and the dilated segment of the aorta is lined by adventitia only and, at times, by perivascular clot. Aneurysms also may be classified according to their gross appearance. A *fusiform aneurysm* affects the entire circumference of a segment of the vessel, resulting in a diffusely dilated artery. In contrast, a *saccular aneurysm* involves only a portion of the circumference, resulting in an outpouching of the vessel wall. Aortic aneurysms also are classified according to location, i.e., abdominal versus thoracic. Aneurysms of the descending thoracic aorta are usually contiguous with infradiaphragmatic aneurysms and are referred to as *thoracoabdominal aortic aneurysms*.

Etiology

Aortic aneurysms result from conditions that cause degradation or abnormal production of the structural components of the aortic wall: elastin and collagen. The causes of aortic aneurysms may be broadly categorized as degenerative diseases, inherited or developmental diseases, infections, vasculitis, and trauma (**Table 248-1**). Inflammation, proteolysis, and biomechanical wall stress contribute to the degenerative processes that characterize most aneurysms of the abdominal and descending thoracic aorta. These are mediated by B cell and T cell lymphocytes, macrophages, inflammatory cytokines, and matrix metalloproteinases that degrade elastin and collagen and alter the tensile strength and ability of the aorta to accommodate pulsatile stretch. The associated histopathology demonstrates destruction of elastin and collagen, decreased vascular smooth muscle, in-growth of new blood vessels, and inflammation. Factors associated with degenerative aortic aneurysms include aging, cigarette smoking, hypercholesterolemia, male sex, and a family history of aortic aneurysms.

Table 248–1. Diseases of the Aorta: Etiology and Associated Factors

Aortic aneurysm
Degenerative/atherosclerosis

Aging
Cigarette smoking
Male gender
Family history
Cystic medial necrosis
Marfan syndrome
Loeys-Dietz syndrome
Ehlers-Danlos syndrome type IV
Familial
Bicuspid aortic valve
Chronic aortic dissection
Infective (see below)
Trauma
Acute aortic syndromes (aortic dissection, acute intramural hematoma, penetrating atherosclerotic ulcer)
Atherosclerosis
Cystic medial necrosis (see above)
Hypertension
Vasculitis (see below)
Pregnancy
Trauma
Aortic occlusion
Atherosclerosis
Thromboembolism
Aortitis
Vasculitis
Takayasu's arteritis
Giant cell arteritis
Rheumatic
HLA-B27–associated spondyloarthropathies
Behçet's syndrome
Cogan's syndrome
Idiopathic aortitis
Infective
Syphilis
Tuberculosis
Mycotic (Salmonella, staphylococcal, streptococcal, fungal)

The most common pathologic condition associated with degenerative aortic aneurysms is *atherosclerosis*. Many patients with aortic aneurysms have coexisting risk factors for atherosclerosis (Chap. 241), as well as atherosclerosis in other blood vessels.

Cystic medial necrosis is the histopathologic term used to describe the degeneration of collagen and elastic fibers in the tunica media of the aorta as well as the loss of medial cells that are replaced by multiple clefts of mucoid material. Cystic medial necrosis characteristically affects the proximal aorta, results in circumferential weakness and dilation, and leads to the development of fusiform aneurysms involving the ascending aorta and the sinuses of Valsalva. This condition is particularly prevalent in patients with Marfan syndrome, Loeys-Dietz syndrome, Ehlers-Danlos syndrome type IV (Chap. 363), hypertension, congenital bicuspid aortic valves, and familial thoracic aortic aneurysm syndromes; sometimes it appears as an isolated condition in patients without any other apparent disease.

Familial clusterings of aortic aneurysms occur in 20% of patients, suggesting a hereditary basis for the disease. Mutations of the gene that encodes fibrillin-1 are present in patients with Marfan syndrome. Fibrillin-1 is an important component of extracellular micofibrils, which support the architecture of elastic fibers and other connective tissue. Deficiency of fibrillin-1 in the extracellular matrix leads to excessive signaling by transforming growth factor ß (TGF-ß). Loeys-Dietz syndrome is caused by mutations in the genes that encode TGF-ß receptors 1 (*TGFBR1*) and 2 (*TGFBR2*). Increased signaling by TGF-ß and mutations of *TGFBR1* and *TGFBR2* may cause thoracic aortic aneurysms. Mutations of type III procollagen have been implicated in Ehlers-Danlos type IV syndrome. Linkage

analysis has identified loci on chromosomes 5q13–14, 11q23.3–q24, and 3p24–25 in several families, although the specific alleles have not been described.

The infectious causes of aortic aneurysms include syphilis, tuberculosis, and other bacterial infections. *Syphilis* (Chap. 169) is a relatively uncommon cause of aortic aneurysm. Syphilitic periaortitis and mesoaortitis damage elastic fibers, resulting in thickening and weakening of the aortic wall. Approximately 90% of syphilitic aneurysms are located in the ascending aorta or aortic arch. *Tuberculous aneurysms* (Chap. 165) typically affect the thoracic aorta and result from direct extension of infection from hilar lymph nodes or contiguous abscesses as well as from bacterial seeding. Loss of aortic wall elasticity results from granulomatous destruction of the medial layer. A *mycotic aneurysm* is a rare condition that develops as a result of staphylococcal, streptococcal, *Salmonella*, or other bacterial or fungal infections of the aorta, usually at an atherosclerotic plaque. These aneurysms are usually saccular. Blood cultures are often positive and reveal the nature of the infective agent.

Vasculitides associated with aortic aneurysm include Takayasu's arteritis and giant cell arteritis, which may cause aneurysms of the aortic arch and descending thoracic aorta. Spondyloarthropathies such as ankylosing spondylitis, rheumatoid arthritis, psoriatic arthritis, relapsing polychondritis, and reactive arthritis (formerly known as Reiter's syndrome) are associated with dilation of the ascending aorta. Aortic aneurysms occur in patients with Behçet's syndrome (Chap. 327) and Cogan's syndrome. Aortic aneurysms also result from idiopathic aortitis. *Traumatic aneurysms* may occur after penetrating or nonpenetrating chest trauma and most commonly affect the descending thoracic aorta just beyond the site of insertion of the ligamentum arteriosum. Chronic aortic dissections are associated with weakening of the aortic wall that may lead to the development of aneurysmal dilatation.

Thoracic Aortic Aneurysms

The clinical manifestations and natural history of thoracic aortic aneurysms depend on their location. Cystic medial necrosis is the most common pathology associated with ascending aortic aneurysms, whereas atherosclerosis is the condition most frequently associated with aneurysms of the aortic arch and descending thoracic aorta. The average growth rate of thoracic aneurysms is 0.1–0.2 cm per year. Thoracic aortic aneurysms associated with Marfan syndrome or aortic dissection may expand at a greater rate. The risk of rupture is related to the size of the aneurysm and the presence of symptoms, ranging approximately from 2–3% per year for thoracic aortic aneurysms <4.0 cm in diameter to 7% per year for those >6 cm in diameter. Most thoracic aortic aneurysms are asymptomatic; however, compression or erosion of adjacent tissue by aneurysms may cause symptoms such as chest pain, shortness of breath, cough, hoarseness, and dysphagia. Aneurysmal dilation of the ascending aorta may cause congestive heart failure as a consequence of aortic regurgitation, and compression of the superior vena cava may produce congestion of the head, neck, and upper extremities.

A chest x-ray may be the first test that suggests the diagnosis of a thoracic aortic aneurysm **(Fig. 248-1)**. Findings include widening of the mediastinal shadow and displacement or compression of the trachea or left mainstem bronchus. Echocardiography, particularly transesophageal echocardiography, can be used to assess the proximal ascending aorta and descending thoracic aorta. Contrast-enhanced CT, magnetic resonance imaging (MRI), and conventional invasive aortography are sensitive and specific tests for assessment of aneurysms of the thoracic aorta and involvement of branch vessels **(Fig. 248-2)**. In asymptomatic patients whose aneurysms are too small to justify surgery, noninvasive testing with either contrast-enhanced CT or MRI should be performed at least every 6–12 months to monitor expansion.

Figure 248-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A chest x-ray of a patient with a thoracic aortic aneurysm.

Figure 248-2

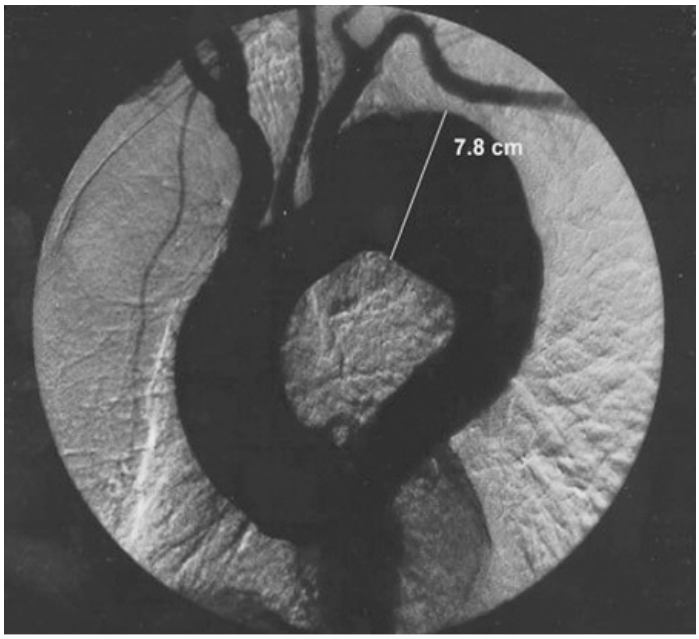

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An aortogram demonstrating a large fusiform aneurysm of the descending thoracic aorta.

TREATMENT: THORACIC AORTIC ANEURYSMS

β-Adrenergic blockers currently are recommended for patients with thoracic aortic aneurysms, particularly those with Marfan syndrome, who have evidence of aortic root dilatation to reduce the rate of further expansion. Additional medical therapy should be given as necessary to control hypertension. Recent preliminary studies indicate that angiotensin receptor antagonists and angiotensin-converting enzyme inhibitors will reduce the rate of aortic dilation in patients with Marfan syndrome by blocking TGF-β signaling; clinical outcome trials of this treatment approach are in progress. Operative repair with placement of a prosthetic graft is indicated in patients with symptomatic thoracic aortic aneurysms,

those in whom the ascending aortic diameter is >5.5–6 cm or the descending thoracic aortic diameter is >6.5-7 cm, and those with an aneurysm that has increased by >1 cm per year. In patients with Marfan syndrome or bicuspid aortic valve, ascending thoracic aortic aneurysms >5 cm should be considered for surgery. Endovascular repair is an alternative treatment for some patients with descending thoracic aortic aneurysms.

Abdominal Aortic Aneurysms

Abdominal aortic aneurysms occur more frequently in males than in females, and the incidence increases with age. Abdominal aortic aneurysms ≥4.0 cm may affect 1–2% of men older than 50 years. At least 90% of all abdominal aortic aneurysms >4.0 cm are related to atherosclerotic disease, and most of these aneurysms are below the level of the renal arteries. Prognosis is related to both the size of the aneurysm and the severity of coexisting coronary artery and cerebrovascular disease. The risk of rupture increases with the size of the aneurysm: the 5-year risk for aneurysms <5 cm is 1–2%, whereas it is 20–40% for aneurysms >5 cm in diameter. The formation of mural thrombi within aneurysms may predispose to peripheral embolization.

An abdominal aortic aneurysm commonly produces no symptoms. It usually is detected on routine examination as a palpable, pulsatile, expansile, and nontender mass, or it is an incidental finding observed on an abdominal x-ray or ultrasound study performed for other reasons. As abdominal aortic aneurysms expand, however, they may become painful. Some patients complain of strong pulsations in the abdomen; others experience pain in the chest, lower back, or scrotum. Aneurysmal pain is usually a harbinger of rupture and represents a medical emergency. More often, acute rupture occurs without any prior warning, and this complication is always life-threatening. Rarely, there is leakage of the aneurysm with severe pain and tenderness. Acute pain and hypotension occur with rupture of the aneurysm, which requires an emergency operation.

Abdominal radiography may demonstrate the calcified outline of the aneurysm; however, about 25% of aneurysms are not calcified and cannot be visualized by x-ray imaging. An abdominal ultrasound can delineate the transverse and longitudinal dimensions of an abdominal aortic aneurysm and may detect mural thrombus. Abdominal ultrasound is useful for serial documentation of aneurysm size and can be used to screen patients at risk for developing an aortic aneurysm. In one large study, ultrasound screening of men age 65–74 years was associated with a risk reduction in aneurysm-related death of 42%. For this reason, screening by ultrasonography is recommended for men age 65–75 years who have ever smoked. In addition, siblings or offspring of persons with abdominal aortic aneurysms, as well as individuals with thoracic aortic or peripheral arterial aneurysms, should be considered for screening for abdominal aortic aneurysms. CT with contrast and MRI are accurate noninvasive tests to determine the location and size of abdominal aortic aneurysms and to plan endovascular or open surgical repair **(Fig. 248-3)**. Contrast aortography may be used for the evaluation of patients with aneurysms, but the procedure carries a small risk of complications such as bleeding, allergic reactions, and atheroembolism. Since the presence of mural thrombi may reduce the luminal size, aortography may underestimate the diameter of an aneurysm.

Figure 248-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A computed tomographic angiogram (CTA) depicting a fusiform abdominal aortic aneurysm that has been treated with a bifurcated stent graft.

TREATMENT: ABDOMINAL AORTIC ANEURYSMS

Operative repair of the aneurysm with insertion of a prosthetic graft or endovascular placement of an aortic stent graft (Fig. 248-3) is indicated for abdominal aortic aneurysms of any size that are expanding rapidly or are associated with symptoms. For asymptomatic aneurysms, abdominal aortic aneurysm repair is indicated if the diameter is >5.5 cm. In randomized trials of patients with abdominal aortic aneurysms <5.5 cm, there was no difference in the long-term (5- to 8-year) mortality rate between those followed with ultrasound surveillance and those undergoing elective surgical repair. Thus, serial noninvasive follow-up of smaller aneurysms (<5 cm) is an alternative to immediate repair. The decision to perform an open surgical operation or endovascular repair is based in part on the vascular anatomy and comorbid conditions. Endovascular repair of abdominal aortic aneurysms has a lower short-term morbidity rate but a comparable long-term mortality rate with open surgical reconstruction. Long-term surveillance with CT or MR aortography is indicated after endovascular repair to detect leaks and possible aneurysm expansion.

In surgical candidates, careful preoperative cardiac and general medical evaluations (followed by appropriate therapy for complicating conditions) are essential. Preexisting coronary artery disease, congestive heart failure, pulmonary disease, diabetes mellitus, and advanced age add to the risk of surgery. ß-Adrenergic blockers decrease perioperative cardiovascular morbidity and mortality. With careful preoperative cardiac evaluation and postoperative care, the operative mortality rate approximates 1–2%. After acute rupture, the mortality rate of emergent operation is 45–50%. Endovascular repair with stent placement is an emerging approach but at the current time is associated with a mortality rate of approximately 40%.

ACUTE AORTIC SYNDROMES

The four major acute aortic syndromes are aortic rupture (discussed above), aortic dissection, intramural hematoma, and penetrating atherosclerotic ulcer. Aortic dissection is caused by a circumferential or, less frequently, transverse tear of the intima. It often occurs along the right lateral wall of the ascending aorta where the hydraulic shear stress is high. Another common site is the descending thoracic aorta just below the ligamentum arteriosum. The initiating event is either a primary intimal tear with secondary dissection into the media or a medial hemorrhage that dissects into and disrupts the intima. The pulsatile aortic flow then dissects along the elastic lamellar plates of the aorta and creates a false lumen. The dissection usually propagates distally down the descending aorta and into its major branches, but it may propagate proximally. Distal propagation may be limited by atherosclerotic plaque. In some cases, a secondary distal intimal disruption occurs, resulting in the reentry of blood from the false to the true lumen.

There are at least two important pathologic and radiologic variants of aortic dissection: intramural hematoma without an intimal flap and penetrating atherosclerotic ulcer. Acute intramural hematoma is

thought to result from rupture of the vasa vasorum with hemorrhage into the wall of the aorta. Most of these hematomas occur in the descending thoracic aorta. Acute intramural hematomas may progress to dissection and rupture. Penetrating atherosclerotic ulcers are caused by erosion of a plaque into the aortic media, are usually localized, and are not associated with extensive propagation. They are found primarily in the middle and distal portions of the descending thoracic aorta and are associated with extensive atherosclerotic disease. The ulcer can erode beyond the internal elastic lamina, leading to medial hematoma, and may progress to false aneurysm formation or rupture.

Several classification schemes have been developed for thoracic aortic dissections. DeBakey and colleagues initially classified aortic dissections as type I, in which an intimal tear occurs in the ascending aorta but involves the descending aorta as well; type II, in which the dissection is limited to the ascending aorta; and type III, in which the intimal tear is located in the descending aorta with distal propagation of the dissection **(Fig. 248-4)**. Another classification (Stanford) is that of type A, in which the dissection involves the ascending aorta (proximal dissection), and type B, in which it is limited to the descending aorta (distal dissection). From a management standpoint, classification of aortic dissections and intramural hematomas into type A or B is more practical and useful, since DeBakey types I and II are managed in a similar manner.

Figure 248-4

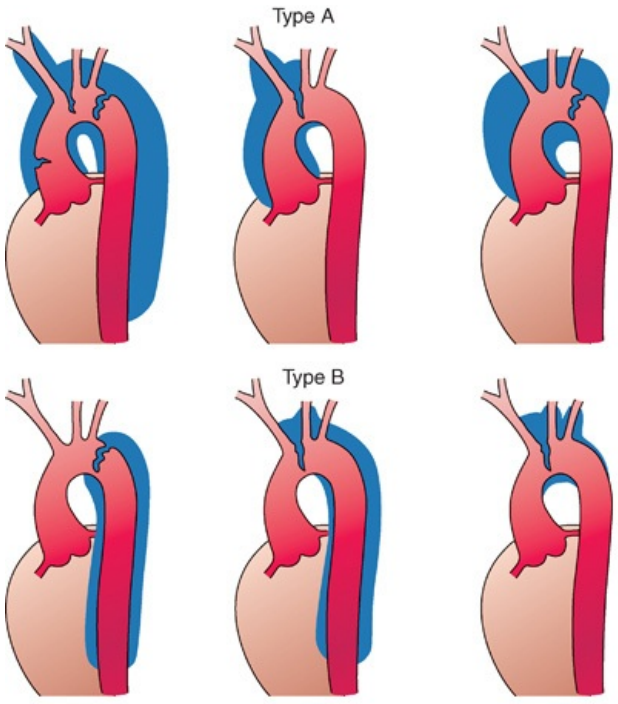

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Classification of aortic dissections. Stanford classification: Type A dissections (*top panels*) involve the ascending aorta independent of site of tear and distal extension; type B dissections (*bottom panels*) involve transverse and/or descending aorta without involvement of the ascending aorta. DeBakey classification: Type I dissection involves ascending to descending aorta (*top left*); type II dissection is limited to ascending or transverse aorta, without descending aorta (*top center + top right*); type III dissection involves descending aorta only (*bottom left*). *[From DC Miller, in RM Doroghazi, EE Slater (eds): Aortic Dissection. New York, McGraw-Hill, 1983, with permission.]*

The factors that predispose to aortic dissection include systemic hypertension, a coexisting condition in 70% of patients, and cystic medial necrosis. Aortic dissection is the major cause of morbidity and mortality in patients with Marfan syndrome (Chap. 363) and similarly may affect patients with Ehlers-Danlos syndrome. The incidence also is increased in patients with inflammatory aortitis (i.e., Takayasu's arteritis, giant cell arteritis), congenital aortic valve anomalies (e.g., bicuspid valve), coarctation of the aorta, and a history of aortic trauma. In addition, the risk of dissection is increased in otherwise normal women during the third trimester of pregnancy.

Clinical Manifestations

The peak incidence of aortic dissection is in the sixth and seventh decades. Men are more affected than women by a ratio of 2:1. The presentations of aortic dissection and its variants are the consequences of intimal tear, dissecting hematoma, occlusion of involved arteries, and compression of adjacent tissues. Acute aortic dissection presents with the sudden onset of pain (Chap. 12), which often is described as very severe and tearing and is associated with diaphoresis. The pain may be

localized to the front or back of the chest, often the interscapular region, and typically migrates with propagation of the dissection. Other symptoms include syncope, dyspnea, and weakness. Physical findings may include hypertension or hypotension, loss of pulses, aortic regurgitation, pulmonary edema, and neurologic findings due to carotid artery obstruction (hemiplegia, hemianesthesia) or spinal cord ischemia (paraplegia). Bowel ischemia, hematuria, and myocardial ischemia have all been observed. These clinical manifestations reflect complications resulting from the dissection occluding the major arteries. Furthermore, clinical manifestations may result from the compression of adjacent structures (e.g., superior cervical ganglia, superior vena cava, bronchus, esophagus) by the expanding dissection, causing aneurysmal dilation, and include Horner's syndrome, superior vena cava syndrome, hoarseness, dysphagia, and airway compromise. Hemopericardium and cardiac tamponade may complicate a type A lesion with retrograde dissection. Acute aortic regurgitation is an important and common (>50%) complication of proximal dissection. It is the outcome of either a circumferential tear that widens the aortic root or a disruption of the annulus by a dissecting hematoma that tears a leaflet(s) or displaces it, inferior to the line of closure. Signs of aortic regurgitation include bounding pulses, a wide pulse pressure, a diastolic murmur often radiating along the right sternal border, and evidence of congestive heart failure. The clinical manifestations depend on the severity of the regurgitation.

In dissections involving the ascending aorta, the chest x-ray often reveals a widened superior mediastinum. A pleural effusion (usually left-sided) also may be present. This effusion is typically serosanguineous and not indicative of rupture unless accompanied by hypotension and falling hematocrit. In dissections of the descending thoracic aorta, a widened mediastinum may be observed on chest x-ray. In addition, the descending aorta may appear to be wider than the ascending portion. An electrocardiogram that shows no evidence of myocardial ischemia is helpful in distinguishing aortic dissection from myocardial infarction. Rarely, the dissection involves the right or, less commonly, left coronary ostium and causes acute myocardial infarction.

The diagnosis of aortic dissection can be established by noninvasive techniques such as echocardiography, CT, and MRI. Aortography is used less commonly because of the accuracy of these noninvasive techniques. Transthoracic echocardiography can be performed simply and rapidly and has an overall sensitivity of 60–85% for aortic dissection. For diagnosing proximal ascending aortic dissections, its sensitivity exceeds 80%; it is less useful for detecting dissection of the arch and descending thoracic aorta. Transesophageal echocardiography requires greater skill and patient cooperation but is very accurate in identifying dissections of the ascending and descending thoracic aorta but not the arch, achieving 98% sensitivity and approximately 90% specificity. Echocardiography also provides important information regarding the presence and severity of aortic regurgitation and pericardial effusion. CT and MRI are both highly accurate in identifying the intimal flap and the extent of the dissection and involvement of major arteries; each has a sensitivity and specificity >90%. They are useful in recognizing intramural hemorrhage and penetrating ulcers. MRI also can detect blood flow, which may be useful in characterizing antegrade versus retrograde dissection. The relative utility of transesophageal echocardiography, CT, and MRI depends on the availability and expertise in individual institutions as well as on the hemodynamic stability of the patient, with CT and MRI obviously less suitable for unstable patients.

TREATMENT: AORTIC DISSECTION

Medical therapy should be initiated as soon as the diagnosis is considered. The patient should be admitted to an intensive care unit for hemodynamic monitoring. Unless hypotension is present, therapy should be aimed at reducing cardiac contractility and systemic arterial pressure, and thus shear stress. For acute dissection, unless contraindicated, β-adrenergic blockers should be administered parenterally, using intravenous propranolol, metoprolol, or the short-acting esmolol to achieve a heart rate of approximately 60 beats/min. This should be accompanied by sodium nitroprusside infusion to lower systolic blood pressure to ≤120 mmHg. Labetalol (Chap. 247), a drug with both β- and α-adrenergic blocking properties, also may be used as a parenteral agent in acute therapy for dissection.

The calcium channel antagonists verapamil and diltiazem may be used intravenously if nitroprusside or β-adrenergic blockers cannot be employed. The addition of a parenteral angiotensin- converting enzyme (ACE) inhibitor such as enalaprilat to a β-adrenergic blocker also may be considered. Isolated use of a direct vasodilator such as hydralazine is contraindicated because these agents can increase hydraulic shear and may propagate the dissection.

Emergent or urgent surgical correction is the preferred treatment for acute ascending aortic dissections and intramural hematomas (type A) and for complicated type B dissections, including those characterized by propagation, compromise of major aortic branches, impending rupture, or continued pain. Surgery involves excision of the intimal flap, obliteration of the false lumen, and placement of an interposition graft. A composite valve-graft conduit is used if the aortic valve is disrupted. The overall in-hospital mortality rate after surgical treatment of patients with aortic dissection is reported to be 15–25%. The major causes of perioperative mortality and morbidity include myocardial infarction, paraplegia, renal failure, tamponade, hemorrhage, and sepsis. Endoluminal stent grafts may be considered in selected patients. Other transcatheter techniques, such as fenestration of the intimal flaps and stenting of narrowed branch vessels to increase flow to compromised organs, are used in

selected patients. For uncomplicated and stable distal dissections and intramural hematomas (type B), medical therapy is the preferred treatment. The in-hospital mortality rate of medically treated patients with type B dissection is 10–20%. Long-term therapy for patients with aortic dissection and intramural hematomas (with or without surgery) consists of control of hypertension and reduction of cardiac contractility with the use of beta blockers plus other antihypertensive agents, such as ACE inhibitors or calcium antagonists. Patients with chronic type B dissection and intramural hematomas should be followed on an outpatient basis every 6–12 months with contrast-enhanced CT or MRI to detect propagation or expansion. Patients with Marfan syndrome are at high risk for postdissection complications. The long-term prognosis for patients with treated dissections is generally good with careful follow-up; the 10-year survival rate is approximately 60%.

Chronic Atherosclerotic Occlusive Disease

Atherosclerosis may affect the thoracic and abdominal aorta. Occlusive aortic disease caused by atherosclerosis usually is confined to the distal abdominal aorta below the renal arteries. Frequently the disease extends to the iliac arteries (Chap. 249). Claudication characteristically involves the buttocks, thighs, and calves and may be associated with impotence in males (Leriche syndrome). The severity of the symptoms depends on the adequacy of collaterals. With sufficient collateral blood flow, a complete occlusion of the abdominal aorta may occur without the development of ischemic symptoms. The physical findings include the absence of femoral and other distal pulses bilaterally and the detection of an audible bruit over the abdomen (usually at or below the umbilicus) and the common femoral arteries. Atrophic skin, loss of hair, and coolness of the lower extremities usually are observed. In advanced ischemia, rubor on dependency and pallor on elevation can be seen.

The diagnosis usually is established by physical examination and noninvasive testing, including leg pressure measurements, Doppler velocity analysis, pulse volume recordings, and duplex ultrasonography. The anatomy may be defined by MRI, CT, or conventional aortography, typically performed when one is considering revascularization. Catheter-based endovascular or operative treatment is indicated in patients with lifestyle-limiting or debilitating symptoms of claudication and patients with critical limb ischemia.

Acute Aortic Occlusion

Acute occlusion in the distal abdominal aorta constitutes a medical emergency because it threatens the viability of the lower extremities; it usually results from an occlusive (saddle) embolus that almost always originates from the heart. Rarely, acute occlusion may occur as the result of in situ thrombosis in a preexisting severely narrowed segment of the aorta.

The clinical picture is one of acute ischemia of the lower extremities. Severe rest pain, coolness, and pallor of the lower extremities and the absence of distal pulses bilaterally are the usual manifestations. Diagnosis should be established rapidly by MRI, CT, or aortography. Emergency thrombectomy or revascularization is indicated.

AORTITIS

Aortitis, a term referring to inflammatory disease of the aorta, may be caused by large vessel vasculitides such as Takayasu's arteritis and giant cell arteritis, rheumatic and HLA-B27–associated spondyloarthropathies, Behçet's syndrome, antineutrophil cytoplasmic antibodies (ANCA)-associated vasculitides, Cogan's syndrome, and infections such as syphilis, tuberculosis, and *Salmonella* or may be associated with retroperitoneal fibrosis. Aortitis may result in aneurysmal dilation and aortic regurgitation, occlusion of the aorta and its branch vessels, or acute aortic syndromes.

Takayasu's Arteritis

This inflammatory disease often affects the ascending aorta and aortic arch, causing obstruction of the aorta and its major arteries. Takayasu's arteritis is also termed *pulseless disease* because of the frequent occlusion of the large arteries originating from the aorta. It also may involve the descending thoracic and abdominal aorta and occlude large branches such as the renal arteries. Aortic aneurysms also may occur. The pathology is a panarteritis characterized by mononuclear cells and occasionally giant cells, with marked intimal hyperplasia, medial and adventitial thickening, and, in the chronic form, fibrotic occlusion. The disease is most prevalent in young females of Asian descent but does occur in women of other geographic and ethnic origins and also in young men. During the acute stage, fever, malaise, weight loss, and other systemic symptoms may be evident. Elevations of the erythrocyte sedimentation rate and C-reactive protein are common. The chronic stages of the disease, which is intermittently active, present with symptoms related to large artery occlusion, such as upper extremity claudication, cerebral ischemia, and syncope. The process is progressive, and there is no definitive therapy. Glucocorticoids and immunosuppressive agents have been reported to be effective in some patients during the acute phase. Surgical bypass or endovascular intervention of a critically stenotic artery may be necessary.

Giant Cell Arteritis

(See also Chap. 326) This vasculitis occurs in older individuals and affects women more often than

men. Primarily large and medium-size arteries are affected. The pathology is that of focal granulomatous lesions involving the entire arterial wall; it may be associated with polymyalgia rheumatica. Obstruction of medium-size arteries (e.g., temporal and ophthalmic arteries) and major branches of the aorta and the development of aortitis and aortic regurgitation are important complications of the disease. High-dose glucocorticoid therapy may be effective when given early.

Rheumatic Aortitis

Rheumatoid arthritis (Chap. 321), ankylosing spondylitis (Chap. 325), psoriatic arthritis (Chap. 325), reactive arthritis (formerly known as Reiter's syndrome) (Chap. 325), relapsing polychondritis, and inflammatory bowel disorders may all be associated with aortitis involving the ascending aorta. The inflammatory lesions usually involve the ascending aorta and may extend to the sinuses of Valsalva, the mitral valve leaflets, and adjacent myocardium. The clinical manifestations are aneurysm, aortic regurgitation, and involvement of the cardiac conduction system.

Idiopathic Aortitis

Idiopathic abdominal aortitis is characterized by adventitial and periaortic inflammation with thickening of the aortic wall. It is associated with abdominal aortic aneurysms and idiopathic retroperitoneal fibrosis. Affected individuals may present with vague constitutional symptoms, fever, and abdominal pain. Retroperitoneal fibrosis can cause ureteral obstruction and hydronephrosis. Glucocorticoids and immunosuppressive agents may reduce the inflammation.

Infective Aortitis

Infective aortitis may result from direct invasion of the aortic wall by bacterial pathogens such as *Staphylococcus*, *Streptococcus*, and *Salmonella* or by fungi. These bacteria cause aortitis by infecting the aorta at sites of atherosclerotic plaque. Bacterial proteases lead to degradation of collagen, and the ensuing destruction of the aortic wall leads to the formation of a saccular aneurysm referred to as a mycotic aneurysm. Mycotic aneurysms have a predilection for the suprarenal abdominal aorta. The pathologic characteristics of the aortic wall include acute and chronic inflammation, abscesses, hemorrhage, and necrosis. Mycotic aneurysms typically affect the elderly and occur in men three times more frequently than in women. Patients may present with fever, sepsis, and chest, back, or abdominal pain; there may have been a preceding diarrheal illness. Blood cultures are positive in the majority of patients. Both CT and MRI are useful to diagnose mycotic aneurysms. Treatment includes antibiotic therapy and surgical removal of the affected part of the aorta and revascularization of the lower extremities with grafts placed in uninfected tissue.

Syphilitic aortitis is a late manifestation of luetic infection (Chap. 169) that usually affects the proximal ascending aorta, particularly the aortic root, resulting in aortic dilation and aneurysm formation. Syphilitic aortitis occasionally may involve the aortic arch or the descending aorta. The aneurysms may be saccular or fusiform and are usually asymptomatic, but compression of and erosion into adjacent structures may result in symptoms; rupture also may occur.

The initial lesion is an obliterative endarteritis of the vasa vasorum, especially in the adventitia. This is an inflammatory response to the invasion of the adventitia by the spirochetes. Destruction of the aortic media occurs as the spirochetes spread into this layer, usually via the lymphatics accompanying the vasa vasorum. Destruction of collagen and elastic tissues leads to dilation of the aorta, scar formation, and calcification. These changes account for the characteristic radiographic appearance of linear calcification of the ascending aorta.

The disease typically presents as an incidental chest radiographic finding 15–30 years after initial infection. Symptoms may result from aortic regurgitation, narrowing of coronary ostia due to syphilitic aortitis, compression of adjacent structures (e.g., esophagus), or rupture. Diagnosis is established by a positive serologic test, i.e., rapid plasmin reagin (RPR) or fluorescent treponemal antibody. Treatment includes penicillin and surgical excision and repair.

FURTHER READINGS

Blankensteijn JD et al: Two-year outcomes after conventional or endovascular repair of abdominal aortic aneurysms. N Engl J Med 352:2398, 2005[PMID: 15944424]

Brook BS et al: Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. N Engl J Med 358:2787, 2008

Creager MA et al (eds): *Vascular Medicine: A Companion to Braunwald's Heart Disease* . Philadelphia, Saunders, 2006

Fleming C et al: Screening for abdominal aortic aneurysm: A best-evidence systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 142:203, 2005[PMID: 15684209]

Golledge J, Eagle KA: Acute aortic dissection. Lancet 372:55, 2008[PMID: 18603160]

Gornik HL, Creager MA: Diseases of the aorta, in *Textbook of Cardiovascular Medicine* , 3rd ed, E Topol (ed). Philadelphia, Lippincott, Williams & Wilkins, 2007

———: Aortitis. Circulation 117:3039, 2008

Hiratzka LF et al: 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with Thoracic Aortic Disease: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, American College of Radiology, American Stroke Association, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society of Interventional Radiology, Society of Thoracic Surgeons, and Society for Vascular Medicine. Circulation 121:e266, 2010

Hirsch AT et al: ACC/AHA 2005 Practice Guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic). Circulation 113:e463, 2006

Isselbacher EM: Thoracic and abdominal aortic aneurysms. Circulation 111:816, 2005[PMID: 15710776]

Lederle FA et al: Systematic review: Repair of unruptured abdominal aortic aneurysm. Ann Intern Med 146:735, 2007[PMID: 17502634]

Patel HJ, Deeb GM: Ascending and arch aorta: Pathology, natural history, and treatment. Circulation 118:188, 2008[PMID: 18606928]

Schermerhorn ML et al: Endovascular vs. open repair of abdominal aortic aneurysms in the Medicare population. N Engl J Med 358:464. 2008

Tsai TT et al: Acute aortic syndromes. Circulation 112:3802, 2005[PMID: 16344407]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ARTERIAL DISORDERS

Peripheral Artery Disease

Peripheral artery disease (PAD) is defined as a clinical disorder in which there is a stenosis or occlusion in the aorta or the arteries of the limbs. Atherosclerosis is the leading cause of PAD in patients >40 years old. Other causes include thrombosis, embolism, vasculitis, fibromuscular dysplasia, entrapment, cystic adventitial disease, and trauma. The highest prevalence of atherosclerotic PAD occurs in the sixth and seventh decades of life. As in patients with atherosclerosis of the coronary and cerebral vasculature, there is an increased risk of developing PAD in cigarette smokers and in persons with diabetes mellitus, hypercholesterolemia, hypertension, or hyperhomocysteinemia.

PATHOLOGY

(See also Chap. 241) Segmental lesions that cause stenosis or occlusion are usually localized to large and medium-size vessels. The pathology of the lesions includes atherosclerotic plaques with calcium deposition, thinning of the media, patchy destruction of muscle and elastic fibers, fragmentation of the internal elastic lamina, and thrombi composed of platelets and fibrin. The primary sites of involvement are the abdominal aorta and iliac arteries (30% of symptomatic patients), the femoral and popliteal arteries (80–90% of patients), and the more distal vessels, including the tibial and peroneal arteries (40–50% of patients). Atherosclerotic lesions occur preferentially at arterial branch points, which are sites of increased turbulence, altered shear stress, and intimal injury. Involvement of the distal vasculature is most common in elderly individuals and patients with diabetes mellitus.

CLINICAL EVALUATION

Fewer than 50% of patients with PAD are symptomatic, although many have a slow or impaired gait. The most common *symptom* is intermittent claudication, which is defined as a pain, ache, cramp, numbness, or a sense of fatigue in the muscles; it occurs during exercise and is relieved by rest. The site of claudication is distal to the location of the occlusive lesion. For example, buttock, hip, and thigh discomfort occurs in patients with aortoiliac disease, whereas calf claudication develops in patients with femoral-popliteal disease. Symptoms are far more common in the lower than in the upper extremities because of the higher incidence of obstructive lesions in the former region. In patients with severe arterial occlusive disease in whom resting blood flow cannot accommodate basal nutritional needs of the tissues, critical limb ischemia may develop. Patients complain of rest pain or a feeling of cold or numbness in the foot and toes. Frequently, these symptoms occur at night when the legs are horizontal and improve when the legs are in a dependent position. With severe ischemia, rest pain may be persistent.

Important *physical findings* of PAD include decreased or absent pulses distal to the obstruction, the presence of bruits over the narrowed artery, and muscle atrophy. With more severe disease, hair loss, thickened nails, smooth and shiny skin, reduced skin temperature, and pallor or cyanosis are common physical signs. In patients with critical limb ischemia, ulcers or gangrene may occur. Elevation of the legs and repeated flexing of the calf muscles produce pallor of the soles of the feet, whereas rubor, secondary to reactive hyperemia, may develop when the legs are dependent. The time required for rubor to develop or for the veins in the foot to fill when the patient's legs are transferred from an elevated to a dependent position is related to the severity of the ischemia and the presence of collateral vessels. Patients with severe ischemia may develop peripheral edema because they keep their legs in a dependent position much of the time. Ischemic neuropathy can result in numbness and hyporeflexia.

NONINVASIVE TESTING

The history and physical examination are often sufficient to establish the diagnosis of PAD. An objective assessment of the presence and severity of disease is obtained by noninvasive techniques. Arterial pressure can be recorded noninvasively in the legs by placement of sphygmomanometric cuffs at the ankles and the use of a Doppler device to auscultate or record blood flow from the dorsalis pedis and posterior tibial arteries. Normally, systolic blood pressure in the legs and arms is similar. Indeed, ankle pressure may be slightly higher than arm pressure due to pulse-wave amplification. In the presence of hemodynamically significant stenoses, the systolic blood pressure in the leg is decreased. Thus, the ratio of the ankle and brachial artery pressures (termed the *ankle:brachial index*, or ABI) is ≥1.0 in normal individuals and <1.0 in patients with PAD; a ratio of <0.5 is consistent

with severe ischemia.

Other noninvasive tests include segmental pressure measurements, segmental pulse volume recordings, duplex ultrasonography (which combines B-mode imaging and Doppler flow velocity waveform analysis examination), transcutaneous oximetry, and stress testing (usually using a treadmill). Placement of pneumatic cuffs enables assessment of systolic pressure along the legs. The presence of pressure gradients between sequential cuffs provides evidence of the presence and location of hemodynamically significant stenoses. In addition, the amplitude of the pulse volume contour becomes blunted in the presence of significant PAD. Duplex ultrasonography is used to image and detect stenotic lesions in native arteries and bypass grafts.

Treadmill testing allows the physician to assess functional limitations objectively. Decline of the ABI immediately after exercise provides further support for the diagnosis of PAD in patients with equivocal symptoms and findings on examination.

Magnetic resonance angiography (MRA), computed tomographic angiography (CTA), and conventional contrast angiography should not be used for routine diagnostic testing but are performed before potential revascularization **(Fig. 249-1)**. Each test is useful in defining the anatomy to assist planning for catheter-based and surgical revascularization procedures.

Figure 249-1

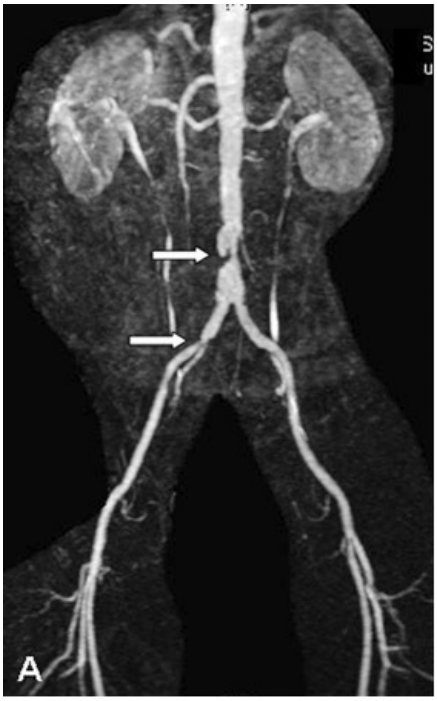

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

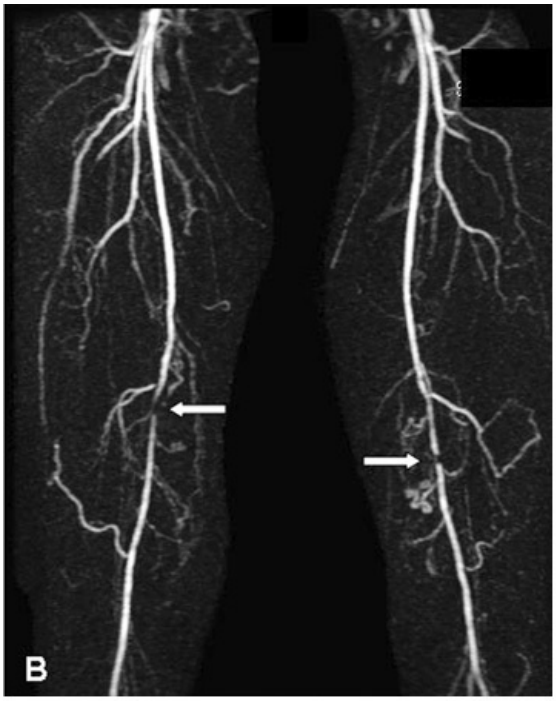

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Magnetic resonance angiography of a patient with intermittent claudication, showing stenoses of the distal abdominal aorta and right iliac common iliac artery (***A***) and stenoses of the right and left superficial femoral arteries (***B***). *(Courtesy of Dr. Edwin Gravereaux, with permission.)*

PROGNOSIS

The natural history of patients with PAD is influenced primarily by the extent of coexisting coronary artery and cerebrovascular disease. Approximately one-third to one-half of patients with symptomatic PAD have evidence of coronary artery disease (CAD) based on clinical presentation and electrocardiogram, and over one-half have significant CAD by coronary angiography. Patients with PAD have a 15–30% 5-year mortality rate and a two- to sixfold increased risk of death from coronary heart disease. Mortality rates are highest in those with the most severe PAD. Measurement of ABI is useful for detecting PAD and identifying persons at risk for future atherothrombotic events. The likelihood of symptomatic progression of PAD is lower than the chance of succumbing to CAD. Approximately 75–80% of nondiabetic patients who present with mild to moderate claudication remain symptomatically stable. Deterioration is likely to occur in the remainder, with approximately 1–2% of the group ultimately developing critical limb ischemia each year. Approximately 25–30% of patients with critical limb ischemia undergo amputation within 1 year. The prognosis is worse in patients who continue to smoke cigarettes or have diabetes mellitus.

TREATMENT: PERIPHERAL ARTERY DISEASE

Patients with PAD should receive therapies to reduce the risk of associated cardiovascular events, such as myocardial infarction and death, and to improve limb symptoms, prevent progression to critical limb ischemia, and preserve limb viability. Risk factor modification and antiplatelet therapy should be initiated to improve cardiovascular outcomes. The importance of discontinuing cigarette smoking cannot be overemphasized. The physician must assume a major role in this lifestyle modification. Counseling and adjunctive drug therapy with the nicotine patch, bupropion, or varenicline increase smoking cessation rates and reduce recidivism. It is important to control blood pressure in hypertensive patients. Angiotensin converting-enzyme inhibitors may reduce the risk of cardiovascular events in patients with symptomatic PAD. β-adrenergic blockers do not worsen claudication and may be used to treat hypertension, especially in patients with coexistent CAD. Treatment of hypercholesterolemia with statins is advocated to reduce the risk of myocardial infarction, stroke, and death. The National Cholesterol Education Program Adult Treatment Panel considers PAD a coronary heart disease equivalent and recommends treatment to reduce low-density lipoprotein (LDL) cholesterol to <100 mg/dL. Platelet inhibitors, including aspirin and clopidogrel, reduce the risk of adverse cardiovascular events in patients with atherosclerosis and are recommended for patients with PAD. Dual antiplatelet therapy with both aspirin and clopidogrel is not more effective than aspirin alone in reducing cardiovascular morbidity and mortality rates in patients with PAD. The anticoagulant warfarin is as effective as antiplatelet therapy in preventing adverse cardiovascular events but causes more major bleeding; therefore, it is not indicated to improve outcomes in patients with chronic PAD.

Therapies for intermittent claudication and critical limb ischemia include supportive measures,

medications, nonoperative interventions, and surgery. Supportive measures include meticulous care of the feet, which should be kept clean and protected against excessive drying with moisturizing creams. Well-fitting and protective shoes are advised to reduce trauma. Elastic support hose should be avoided, as it reduces blood flow to the skin. In patients with critical limb ischemia, shock blocks under the head of the bed together with a canopy over the feet may improve perfusion pressure and ameliorate some of the rest pain.

Patients with claudication should be encouraged to exercise regularly and at progressively more strenuous levels. Supervised exercise training programs for 30- to 45-min sessions, three to five times per week for at least 12 weeks, prolong walking distance. Patients also should be advised to walk until nearly maximum claudication discomfort occurs and then rest until the symptoms resolve before resuming ambulation. Pharmacologic treatment of PAD has not been as successful as the medical treatment of CAD (Chap. 243). In particular, vasodilators as a class have not proved to be beneficial. During exercise, peripheral vasodilation occurs distal to sites of significant arterial stenoses. As a result, perfusion pressure falls, often to levels lower than that generated in the interstitial tissue by the exercising muscle. Drugs such as α-adrenergic blocking agents, calcium channel antagonists, papaverine, and other vasodilators have not been shown to be effective in patients with PAD.

Cilostazol, a phosphodiesterase inhibitor with vasodilator and antiplatelet properties, increases claudication distance by 40–60% and improves measures of quality of life. The mechanism of action accounting for its beneficial effects is not known. Pentoxifylline, a substituted xanthine derivative, increases blood flow to the microcirculation and enhances tissue oxygenation. Although several placebo-controlled studies have found that pentoxifylline increases the duration of exercise in patients with claudication, its efficacy has not been confirmed in all clinical trials. Statins appeared promising for treatment of intermittent claudication in initial clinical trials, but more studies are needed to confirm their efficacy. There is no definitive medical therapy for critical limb ischemia, although several studies have suggested that long-term parenteral administration of vasodilator prostaglandins decreases pain and facilitates healing of ulcers. Clinical trials of angiogenic growth factors are proceeding. Intramuscular gene transfer of DNA encoding vascular endothelial growth factor, fibroblast growth factor, hepatocyte growth factor, or hypoxia-inducible factor 1α, as well as administration of endothelial progenitor cells, may promote collateral blood vessel growth in patients with critical limb ischemia. Some trial results have been negative, and others encouraging. The outcome of ongoing studies will further elucidate the potential role of therapeutic angiogenesis for PAD.

Revascularization

Revascularization procedures, including catheter-based and surgical interventions, are usually indicated for patients with disabling, progressive, or severe symptoms of intermittent claudication despite medical therapy and for those with critical limb ischemia. MRA, CTA, or conventional contrast angiography should be performed to assess vascular anatomy in patients who are being considered for revascularization. Nonoperative interventions include percutaneous transluminal angiography (PTA), stent placement, and atherectomy (Chap. 246). PTA and stenting of the iliac artery are associated with higher success rates than are PTA and stenting of the femoral and popliteal arteries. Approximately 90–95% of iliac PTAs are initially successful, and the 3-year patency rate is >75%. Patency rates may be higher if a stent is placed in the iliac artery. The initial success rates for femoral-popliteal PTA and stenting are approximately 80%, with 60% 3-year patency rates. Patency rates are influenced by the severity of pretreatment stenoses; the prognosis of occlusive lesions is worse than that of nonocclusive stenotic lesions. The role of drug-eluting stents in PAD is under investigation.

Several operative procedures are available for treating patients with aortoiliac and femoral-popliteal artery disease. The preferred operative procedure depends on the location and extent of the obstruction(s) and the general medical condition of the patient. Operative procedures for aortoiliac disease include aortobifemoral bypass, axillofemoral bypass, femoro-femoral bypass, and aortoiliac endarterectomy. The most frequently used procedure is the aortobifemoral bypass using knitted Dacron grafts. Immediate graft patency approaches 99%, and 5- and 10-year graft patency in survivors is >90% and 80%, respectively. Operative complications include myocardial infarction and stroke, infection of the graft, peripheral embolization, and sexual dysfunction from interruption of autonomic nerves in the pelvis. The operative mortality rate ranges from 1–3%, mostly due to ischemic heart disease.

Operative therapy for femoral-popliteal artery disease includes in situ and reverse autogenous saphenous vein bypass grafts, placement of polytetrafluoroethylene (PTFE) or other synthetic grafts, and thromboendarterectomy. The operative mortality rate ranges from 1–3%. The long-term patency rate depends on the type of graft used, the location of the distal anastomosis, and the patency of runoff vessels beyond the anastomosis. Patency rates of femoral-popliteal saphenous vein bypass grafts approach 90% at 1 year and 70–80% at 5 years. Five-year patency rates of infrapopliteal saphenous vein bypass grafts are 60–70%. In contrast, 5-year patency rates of infrapopliteal PTFE grafts are <30%. Lumbar sympathectomy alone or as an adjunct to aortofemoral reconstruction has fallen into disfavor.

Preoperative cardiac risk assessment may identify individuals who are especially likely to experience

an adverse cardiac event during the perioperative period. Patients with angina, prior myocardial infarction, ventricular ectopy, heart failure, or diabetes are among those at increased risk. Stress testing with treadmill exercise (if feasible), radionuclide myocardial perfusion imaging, or echocardiography permits further stratification of patient risk (Chap. 246). Patients with abnormal test results require close supervision and adjunctive management with anti-ischemic medications. β-adrenergic blockers and statins reduce the risk of postoperative cardiovascular complications. Coronary angiography and coronary artery revascularization compared with optimal medical therapy do not improve outcomes in most patients undergoing peripheral vascular surgery, but cardiac catheterization should be considered in patients with unstable angina and angina refractory to medical therapy as well as those suspected of having left main or three-vessel CAD.

Fibromuscular Dysplasia

Fibromuscular dysplasia is a hyperplastic disorder that affects medium-size and small arteries. It occurs predominantly in females and usually involves the renal and carotid arteries but can affect extremity vessels such as the iliac and subclavian arteries. The histologic classification includes intimal fibroplasia, medial dysplasia, and adventitial hyperplasia. Medial dysplasia is subdivided into medial fibroplasia, perimedial fibroplasia, and medial hyperplasia. Medial fibroplasia is the most common type and is characterized by alternating areas of thinned media and fibromuscular ridges. The internal elastic lamina usually is preserved. The iliac arteries are the limb arteries most likely to be affected by fibromuscular dysplasia. It is identified angiographically by a "string of beads" appearance caused by thickened fibromuscular ridges contiguous with thin, less-involved portions of the arterial wall. When limb vessels are involved, clinical manifestations are similar to those for atherosclerosis, including claudication and rest pain. PTA and surgical reconstruction have been beneficial in patients with debilitating symptoms or threatened limbs.

Thromboangiitis Obliterans

Thromboangiitis obliterans (Buerger's disease) is an inflammatory occlusive vascular disorder involving small and medium-size arteries and veins in the distal upper and lower extremities. Cerebral, visceral, and coronary vessels may be affected rarely. This disorder develops most frequently in men <40 years of age. The prevalence is higher in Asians and individuals of Eastern European descent. Although the cause of thromboangiitis obliterans is not known, there is a definite relationship to cigarette smoking in patients with this disorder.

In the initial stages of thromboangiitis obliterans, polymorphonuclear leukocytes infiltrate the walls of the small and medium-size arteries and veins. The internal elastic lamina is preserved, and a cellular, inflammatory thrombus develops in the vascular lumen. As the disease progresses, mononuclear cells, fibroblasts, and giant cells replace the neutrophils. Later stages are characterized by perivascular fibrosis, organized thrombus, and recanalization.

The clinical features of thromboangiitis obliterans often include a triad of claudication of the affected extremity, Raynaud's phenomenon, and migratory superficial vein thrombophlebitis. Claudication usually is confined to the calves and feet or the forearms and hands because this disorder primarily affects distal vessels. In the presence of severe digital ischemia, trophic nail changes, painful ulcerations, and gangrene may develop at the tips of the fingers or toes. The physical examination shows normal brachial and popliteal pulses but reduced or absent radial, ulnar, and/or tibial pulses. Arteriography is helpful in making the diagnosis. Smooth, tapering segmental lesions in the distal vessels are characteristic, as are collateral vessels at sites of vascular occlusion. Proximal atherosclerotic disease is usually absent. The diagnosis can be confirmed by excisional biopsy and pathologic examination of an involved vessel.

There is no specific treatment except abstention from tobacco. The prognosis is worse in individuals who continue to smoke, but results are discouraging even in those who stop smoking. Arterial bypass of the larger vessels may be used in selected instances, as well as local debridement, depending on the symptoms and severity of ischemia. Antibiotics may be useful; anticoagulants and glucocorticoids are not helpful. If these measures fail, amputation may be required.

Vasculitis

Other vasculitides may affect the arteries that supply the upper and lower extremities. **Takayasu's arteritis and giant cell (temporal) arteritis are discussed in Chap. 326.**

Acute Arterial Occlusion

Acute arterial occlusion results in the sudden cessation of blood flow to an extremity. The severity of ischemia and the viability of the extremity depend on the location and extent of the occlusion and the presence and subsequent development of collateral blood vessels. There are two principal causes of acute arterial occlusion: embolism and thrombus in situ.

The most common sources of arterial emboli are the heart, aorta, and large arteries. Cardiac disorders that cause thromboembolism include atrial fibrillation, both chronic and paroxysmal; acute myocardial infarction; ventricular aneurysm; cardiomyopathy; infectious and marantic endocarditis; thrombi associated with prosthetic heart valves; and atrial myxoma. Emboli to the distal vessels may also

originate from proximal sites of atherosclerosis and aneurysms of the aorta and large vessels. Less frequently, an arterial occlusion results paradoxically from a venous thrombus that has entered the systemic circulation via a patent foramen ovale or another septal defect. Arterial emboli tend to lodge at vessel bifurcations because the vessel caliber decreases at those sites; in the lower extremities, emboli lodge most frequently in the femoral artery, followed by the iliac artery, aorta, and popliteal and tibioperoneal arteries.

Acute arterial thrombosis in situ occurs most frequently in atherosclerotic vessels at the site of an atherosclerotic plaque or aneurysm and in arterial bypass grafts. Trauma to an artery may also result in the formation of an acute arterial thrombus. Arterial occlusion may complicate arterial punctures and placement of catheters; it also may result from arterial dissection if the intimal flap obstructs the artery. Less common causes include thoracic outlet compression syndrome, which causes subclavian artery occlusion, and entrapment of the popliteal artery by abnormal placement of the medial head of the gastrocnemius muscle. Polycythemia and hypercoagulable disorders (Chaps. 108 and 116) are also associated with acute arterial thrombosis.

CLINICAL FEATURES

The symptoms of an acute arterial occlusion depend on the location, duration, and severity of the obstruction. Often, severe pain, paresthesia, numbness, and coldness develop in the involved extremity within 1 hour. Paralysis may occur with severe and persistent ischemia. Physical findings include loss of pulses distal to the occlusion, cyanosis or pallor, mottling, decreased skin temperature, muscle stiffening, loss of sensation, weakness, and/or absent deep tendon reflexes. If acute arterial occlusion occurs in the presence of an adequate collateral circulation, as is often the case in acute graft occlusion, the symptoms and findings may be less impressive. In this situation, the patient complains about an abrupt decrease in the distance walked before claudication occurs or of modest pain and paresthesia. Pallor and coolness are evident, but sensory and motor functions generally are preserved. The diagnosis of acute arterial occlusion is usually apparent from the clinical presentation. In most circumstances, MRA, CTA, or catheter-based arteriography is used to confirm the diagnosis and demonstrate the location and extent of occlusion.

TREATMENT: ACUTE ARTERIAL OCCLUSION

Once the diagnosis is made, the patient should be anticoagulated with intravenous heparin to prevent propagation of the clot. In cases of severe ischemia of recent onset, particularly when limb viability is jeopardized, immediate intervention to ensure reperfusion is indicated. Endovascular or surgical thromboembolectomy or arterial bypass procedures are used to restore blood flow to the ischemic extremity promptly, particularly when a large proximal vessel is occluded.

Intraarterial thrombolytic therapy with recombinant tissue plasminogen activator, reteplase, or tenecteplase is often effective when acute arterial occlusion is caused by a thrombus in an atherosclerotic vessel or arterial bypass graft. Thrombolytic therapy may also be indicated when the patient's overall condition contraindicates surgical intervention or when smaller distal vessels are occluded, thus preventing surgical access. Meticulous observation for hemorrhagic complications is required during intraarterial thrombolytic therapy. Another endovascular approach to thrombus removal is percutaneous mechanical thrombectomy using devices that employ hydrodynamic forces or rotating baskets to fragment and remove the clot. These treatments may be used alone but usually are used in conjunction with pharmacologic thrombolysis. Amputation is performed when the limb is not viable, as characterized by loss of sensation, paralysis, and the absence of Doppler-detected blood flow in both arteries and veins.

If the limb is not in jeopardy, a more conservative approach that includes observation and administration of anticoagulants may be taken. Anticoagulation prevents recurrent embolism and reduces the likelihood of thrombus propagation; it can be initiated with intravenous heparin and followed by oral warfarin. Recommended doses are the same as those used for deep vein thrombosis (Chap. 262). Emboli resulting from infective endocarditis, the presence of prosthetic heart valves, or atrial myxoma often require surgical intervention to remove the cause.

Atheroembolism

Atheroembolism constitutes a subset of acute arterial occlusion. In this condition, multiple small deposits of fibrin, platelets, and cholesterol debris embolize from proximal atherosclerotic lesions or aneurysmal sites. Large protruding aortic atheromas are a source of emboli that may lead to stroke and renal insufficiency as well as limb ischemia. Atheroembolism may occur after intraarterial procedures. Since the emboli tend to lodge in the small vessels of the muscle and skin and may not occlude the large vessels, distal pulses usually remain palpable. Patients complain of acute pain and tenderness at the site of embolization. Digital vascular occlusion may result in ischemia and the "blue toe" syndrome; digital necrosis and gangrene may develop **(Fig. 249-2)**. Localized areas of tenderness, pallor, and livedo reticularis (see below) occur at sites of emboli. Skin or muscle biopsy may demonstrate cholesterol crystals.

Figure 249-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Atheroembolism causing cyanotic discoloration and impending necrosis of the toes ("blue toe" syndrome).

Ischemia resulting from atheroemboli is notoriously difficult to treat. Usually neither surgical revascularization procedures nor thrombolytic therapy is helpful because of the multiplicity, composition, and distal location of the emboli. Some evidence suggests that platelet inhibitors prevent atheroembolism. Surgical intervention to remove or bypass the atherosclerotic vessel or aneurysm that causes the recurrent atheroemboli may be necessary.

Thoracic Outlet Compression Syndrome

This is a symptom complex resulting from compression of the neurovascular bundle (artery, vein, or nerves) at the thoracic outlet as it courses through the neck and shoulder. Cervical ribs, abnormalities of the scalenus anticus muscle, proximity of the clavicle to the first rib, or abnormal insertion of the pectoralis minor muscle may compress the subclavian artery, subclavian vein, and brachial plexus as these structures pass from the thorax to the arm. Depending on the structures affected, thoracic outlet compression syndrome is divided into arterial, venous, and neurogenic forms. Patients with neurogenic thoracic outlet compression may develop shoulder and arm pain, weakness, and paresthesias. Patients with arterial compression may experience claudication, Raynaud's phenomenon, and even ischemic tissue loss and gangrene. Venous compression may cause thrombosis of the subclavian and axillary veins; this is often associated with effort and is referred to as *Paget-Schroetter syndrome*.

APPROACH TO THE PATIENT: THORACIC OUTLET COMPRESSION SYNDROME

Examination of a patient with thoracic outlet compression syndrome is often normal unless provocative maneuvers are performed. Occasionally, distal pulses are decreased or absent and digital cyanosis and ischemia may be evident. Tenderness may be present in the supraclavicular fossa. In patients with axillo-subclavian venous thrombosis, the affected extremity typically is swollen. Dilated collateral veins may be apparent around the shoulder and upper arm.

Several maneuvers that support the diagnosis of thoracic outlet compression syndrome may be used to precipitate symptoms, cause a subclavian artery bruit, and diminish arm pulses. These maneuvers include the abduction and external rotation test, in which the affected arm is abducted by 90° and the shoulder is externally rotated; the scalene maneuver (extension of the neck and rotation of the head to the side of the symptoms); the costoclavicular maneuver (posterior rotation of shoulders); and the hyperabduction maneuver (raising the arm 180°). A chest x-ray will indicate the presence of cervical ribs. Duplex ultrasonography, MRA, and contrast angiography can be performed during provocative maneuvers to demonstrate thoracic outlet compression of the subclavian artery. Duplex ultrasography, magnetic resonance venography, or contrast venography can be used to diagnose axillo-subclavian vein thrombosis. Neurophysiologic tests such as the electromyogram, nerve conduction studies, and somatosensory evoked potentials may be abnormal if the brachial plexus is involved, but the diagnosis of neurogenic thoracic outlet syndrome is not necessarily excluded if these tests are normal owing to their low sensitivity.

Most patients can be managed conservatively. They should be advised to avoid the positions that cause symptoms. Many patients benefit from shoulder girdle exercises. Surgical procedures such as removal of the first rib and resection of the scalenus anticus muscle are necessary occasionally for relief of symptoms or treatment of ischemia.

Popliteal Artery Entrapment

Popliteal artery entrapment typically affects young athletic men and women when the gastrocnemius or popliteus muscle compresses the popliteal artery and causes intermittent claudication. Thrombosis, embolism, or popliteal artery aneurysm may occur. The pulse examination may be normal unless provocative maneuvers such as ankle dorsiflexion and plantar flexion are performed. The diagnosis is confirmed by duplex ultrasound, CTA, MRA, or conventional angiography. Treatment involves surgical release of the popliteal artery or vascular reconstruction.

Popliteal Artery Aneurysm

Popliteal artery aneurysms are the most common peripheral artery aneurysms. Approximately 50% are bilateral. Patients with popliteal artery aneurysms often have aneurysms of other arteries, especially the aorta. The most common clinical presentation is limb ischemia secondary to thrombosis or embolism. Rupture occurs less frequently. Other complications include compression of the adjacent popliteal vein or peroneal nerve. Popliteal artery aneurysm can be detected by palpation and confirmed by duplex ultrasonography. Repair is indicated for symptomatic aneurysms or when the diameter exceeds 2–3 cm, owing to the risk of thrombosis, embolism, or rupture.

Arteriovenous Fistula

Abnormal communications between an artery and a vein, bypassing the capillary bed, may be congenital or acquired. Congenital arteriovenous fistulas are a result of persistent embryonic vessels that fail to differentiate into arteries and veins; they may be associated with birthmarks, can be located in almost any organ of the body, and frequently occur in the extremities. Acquired arteriovenous fistulas either are created to provide vascular access for hemodialysis or occur as a result of a penetrating injury such as a gunshot or knife wound or as complications of arterial catheterization or surgical dissection. An uncommon cause of arteriovenous fistula is rupture of an arterial aneurysm into a vein.

The clinical features depend on the location and size of the fistula. Frequently, a pulsatile mass is palpable, and a thrill and a bruit lasting throughout systole and diastole are present over the fistula. With long-standing fistulas, clinical manifestations of chronic venous insufficiency, including peripheral edema; large, tortuous varicose veins; and stasis pigmentation become apparent because of the high venous pressure. Evidence of ischemia may occur in the distal portion of the extremity. Skin temperature is higher over the arteriovenous fistula. Large arteriovenous fistulas may result in an increased cardiac output with consequent cardiomegaly and high-output heart failure (Chap. 234).

The diagnosis is often evident from the physical examination. Compression of a large arteriovenous fistula may cause reflex slowing of the heart rate (Nicoladoni-Branham sign). Duplex ultrasonography may detect an arteriovenous fistula, especially one that affects the femoral artery and vein at the site of catheter access. Computed tomographic and conventional angiography can confirm the diagnosis and are useful in demonstrating the site and size of the arteriovenous fistula.

Management of arteriovenous fistulas may involve surgery, radiotherapy, or embolization. Congenital arteriovenous fistulas are often difficult to treat because the communications may be numerous and extensive, and new communications frequently develop after ligation of the most obvious ones. Many of these lesions are best treated conservatively using elastic support hose to reduce the consequences of venous hypertension. Occasionally, embolization with autologous material, such as fat or muscle, or with hemostatic agents, such as gelatin sponges or silicon spheres, is used to obliterate the fistula. Acquired arteriovenous fistulas are usually amenable to surgical treatment that involves division or excision of the fistula. Occasionally, autogenous or synthetic grafting is necessary to reestablish continuity of the artery and vein.

Raynaud's Phenomenon

Raynaud's phenomenon is characterized by episodic digital ischemia, manifested clinically by the sequential development of digital blanching, cyanosis, and rubor of the fingers or toes after cold exposure and subsequent rewarming. Emotional stress may also precipitate Raynaud's phenomenon. The color changes are usually well demarcated and are confined to the fingers or toes. Typically, one or more digits will appear white when the patient is exposed to a cold environment or touches a cold object. The blanching, or pallor, represents the ischemic phase of the phenomenon and results from vasospasm of digital arteries. During the ischemic phase, capillaries and venules dilate, and cyanosis results from the deoxygenated blood that is present in these vessels. A sensation of cold or numbness or paresthesia of the digits often accompanies the phases of pallor and cyanosis.

With rewarming, the digital vasospasm resolves, and blood flow into the dilated arterioles and capillaries increases dramatically. This "reactive hyperemia" imparts a bright red color to the digits. In addition to rubor and warmth, patients often experience a throbbing, painful sensation during the hyperemic phase. Although the triphasic color response is typical of Raynaud's phenomenon, some patients may develop only pallor and cyanosis; others may experience only cyanosis.

Raynaud's phenomenon is broadly separated into two categories: the idiopathic variety, termed *Raynaud's disease*, and the secondary variety, which is associated with other disease states or known causes of vasospasm **(Table 249-1)**.

Table 249–1. Classification of Raynaud's Phenomenon

Primary or idiopathic Raynaud's phenomenon: Raynaud's disease
Secondary Raynaud's phenomenon
Collagen vascular diseases: scleroderma, systemic lupus erythematosus, rheumatoid arthritis, dermatomyositis, polymyositis
Arterial occlusive diseases: atherosclerosis of the extremities, thromboangiitis obliterans, acute arterial occlusion, thoracic outlet syndrome
Pulmonary hypertension
Neurologic disorders: intervertebral disk disease, syringomyelia, spinal cord tumors, stroke, poliomyelitis, carpal tunnel syndrome
Blood dyscrasias: cold agglutinins, cryoglobulinemia, cryofibrinogenemia, myeloproliferative disorders, Waldenström's macroglobulinemia
Trauma: vibration injury, hammer hand syndrome, electric shock, cold injury, typing, piano playing
Drugs: ergot derivatives, methysergide, β-adrenergic receptor blockers, bleomycin, vinblastine, cisplatin

RAYNAUD'S DISEASE

This appellation is applied when the secondary causes of Raynaud's phenomenon have been excluded. Over 50% of patients with Raynaud's phenomenon have Raynaud's disease. Women are affected about five times more often than men, and the age of presentation is usually between 20 and 40 years. The fingers are involved more frequently than the toes. Initial episodes may involve only one or two fingertips, but subsequent attacks may involve the entire finger and may include all the fingers. The toes are affected in 40% of patients. Although vasospasm of the toes usually occurs in patients with symptoms in the fingers, it may happen alone. Rarely, the earlobes, the tip of the nose, and the penis are involved. Raynaud's phenomenon occurs frequently in patients who also have migraine headaches or variant angina. These associations suggest that there may be a common predisposing cause for the vasospasm.

Results of physical examination are often entirely normal; the radial, ulnar, and pedal pulses are normal. The fingers and toes may be cool between attacks and may perspire excessively. Thickening and tightening of the digital subcutaneous tissue (*sclerodactyly*) develop in 10% of patients. Angiography of the digits for diagnostic purposes is not indicated.

In general, patients with Raynaud's disease have milder forms of Raynaud's phenomenon. Fewer than 1% of these patients lose a part of a digit. After the diagnosis is made, the disease improves spontaneously in approximately 15% of patients and progresses in about 30%.

SECONDARY CAUSES OF RAYNAUD'S PHENOMENON

Raynaud's phenomenon occurs in 80–90% of patients with systemic sclerosis (scleroderma) and is the presenting symptom in 30% (Chap. 323). It may be the only symptom of scleroderma for many years. Abnormalities of the digital vessels may contribute to the development of Raynaud's phenomenon in this disorder. Ischemic fingertip ulcers may develop and progress to gangrene and autoamputation. About 20% of patients with systemic lupus erythematosus (SLE) have Raynaud's phenomenon (Chap. 319). Occasionally, persistent digital ischemia develops and may result in ulcers or gangrene. In most severe cases, the small vessels are occluded by a proliferative endarteritis. Raynaud's phenomenon occurs in about 30% of patients with dermatomyositis or polymyositis (Chap. 388). It frequently develops in patients with rheumatoid arthritis and may be related to the intimal proliferation that occurs in the digital arteries.

Atherosclerosis of the extremities is a common cause of Raynaud's phenomenon in men >50 years. Thromboangiitis obliterans is an uncommon cause of Raynaud's phenomenon but should be considered in young men, particularly those who are cigarette smokers. The development of cold-induced pallor in these disorders may be confined to one or two digits of the involved extremity. Occasionally, Raynaud's phenomenon may follow acute occlusion of large and medium-size arteries by a thrombus or embolus. Embolization of atheroembolic debris may cause digital ischemia. The latter situation often involves one or two digits and should not be confused with Raynaud's phenomenon. In patients with thoracic outlet compression syndrome, Raynaud's phenomenon may result from diminished intravascular pressure, stimulation of sympathetic fibers in the brachial plexus, or a combination of both. Raynaud's phenomenon occurs in patients with primary pulmonary hypertension (Chap. 250); this is more than coincidental and may reflect a neurohumoral abnormality that affects both the pulmonary and digital circulations.

A variety of blood dyscrasias may be associated with Raynaud's phenomenon. Cold-induced precipitation of plasma proteins, hyperviscosity, and aggregation of red cells and platelets may occur in patients with cold agglutinins, cryoglobulinemia, or cryofibrinogenemia. Hyperviscosity syndromes that accompany myeloproliferative disorders and Waldenström macroglobulinemia should also be considered in the initial evaluation of patients with Raynaud's phenomenon.

Raynaud's phenomenon occurs often in patients whose vocations require the use of vibrating hand tools, such as chain saws or jackhammers. The frequency of Raynaud's phenomenon also seems to be increased in pianists and keyboard operators. Electric shock injury to the hands or frostbite may lead to the later development of Raynaud's phenomenon.

Several drugs have been causally implicated in Raynaud's phenomenon. They include ergot preparations, methysergide, β-adrenergic receptor antagonists, and the chemotherapeutic agents bleomycin, vinblastine, and cisplatin.

TREATMENT: RAYNAUD'S PHENOMENON

Most patients with Raynaud's phenomenon experience only mild and infrequent episodes. These patients need reassurance and should be instructed to dress warmly and avoid unnecessary cold exposure. In addition to gloves and mittens, patients should protect the trunk, head, and feet with warm clothing to prevent cold-induced reflex vasoconstriction. Tobacco use is contraindicated.

Drug treatment should be reserved for severe cases. Dihydropyridine calcium channel antagonists such as nifedipine, isradipine, felodipine, and amlodipine decrease the frequency and severity of Raynaud's phenomenon. Diltiazem may be considered but is less effective. The postsynaptic α_1-adrenergic antagonist prazosin has been used with favorable responses; doxazosin and terazosin may also be effective. Topical glyceryl trinitrate may be useful in some patients. Digital sympathectomy is helpful in some patients who are unresponsive to medical therapy.

Acrocyanosis

In this condition, there is arterial vasoconstriction and secondary dilation of the capillaries and venules with resulting persistent cyanosis of the hands and, less frequently, the feet. Cyanosis may be intensified by exposure to a cold environment. Acrocyanosis may be categorized as primary or secondary to an underlying condition. In primary acrocyanosis, women are affected much more frequently than men, and the age of onset is usually <30 years. Generally, patients are asymptomatic but seek medical attention because of the discoloration. The prognosis is favorable, and pain, ulcers, and gangrene do not occur. Examination reveals normal pulses, peripheral cyanosis, and moist palms. Trophic skin changes and ulcerations do *not* occur. The disorder can be distinguished from Raynaud's phenomenon because it is persistent and not episodic, the discoloration extends proximally from the digits, and blanching does not occur. Ischemia secondary to arterial occlusive disease can usually be excluded by the presence of normal pulses. Central cyanosis and decreased arterial oxygen saturation are not present. Patients should be reassured and advised to dress warmly and avoid cold exposure. Pharmacologic intervention is not indicated.

Secondary acrocyanosis may result from hypoxemia, connective tissue diseases, atheroembolism, antiphospholipid antibodies, cold agglutinins, or cryoglobulins and is associated with anorexia nervosa and orthostatic tachycardia syndrome. Treatment should be directed at the underlying disorder.

Livedo Reticularis

In this condition, localized areas of the extremities develop a mottled or rete (netlike) appearance of reddish to blue discoloration. The mottled appearance may be more prominent after cold exposure. There are primary and secondary forms of livedo reticularis. The primary, or idiopathic, form of this disorder may be benign or associated with ulcerations. The benign form occurs more frequently in women than in men, and the most common age of onset is the third decade. Patients with the benign form are usually asymptomatic and seek attention for cosmetic reasons. These patients should be reassured and advised to avoid cold environments. No drug treatment is indicated. Primary livedo reticularis with ulceration is also called *atrophie blanche en plaque*. The ulcers are painful and may take months to heal. Secondary livedo reticularis can occur with atheroembolism (see above), SLE and other vasculitides, anticardiolipin antibodies, hyperviscosity, cryoglobulinemia, and Sneddon's syndrome (ischemic stroke and livedo reticularis). Rarely, skin ulcerations develop.

Pernio (Chilblains)

Pernio is a vasculitic disorder associated with exposure to cold; acute forms have been described. Raised erythematous lesions develop on the lower part of the legs and feet in cold weather. They are associated with pruritus and a burning sensation, and they may blister and ulcerate. Pathologic examination demonstrates angiitis characterized by intimal proliferation and perivascular infiltration of mononuclear and polymorphonuclear leukocytes. Giant cells may be present in the subcutaneous tissue. Patients should avoid exposure to cold, and ulcers should be kept clean and protected with sterile dressings. Sympatholytic drugs and dihydropyridine calcium channel antagonists may be effective in some patients.

Erythromelalgia

This disorder is characterized by burning pain and erythema of the extremities. The feet are involved more frequently than the hands, and males are affected more frequently than females. Erythromelalgia may occur at any age but is most common in middle age. It may be primary (also termed erythermalgia) or secondary. The most common causes of secondary erythromelalgia are myeloproliferative disorders

such as polycythemia vera and essential thrombocytosis. Less common causes include drugs, such as calcium channel blockers, bromocriptine, and pergolide; neuropathies; connective tissue diseases such as SLE; and paraneoplastic syndromes. Patients complain of burning in the extremities that is precipitated by exposure to a warm environment and aggravated by a dependent position. The symptoms are relieved by exposing the affected area to cool air or water or by elevation. Erythromelalgia can be distinguished from ischemia secondary to peripheral arterial disorders and peripheral neuropathy because the peripheral pulses are present and the neurologic examination is normal. There is no specific treatment; aspirin may produce relief in patients with erythromelalgia secondary to myeloproliferative disease. Treatment of associated disorders in secondary erythromelalgia may be helpful.

Frostbite

In this condition, tissue damage results from severe environmental cold exposure or from direct contact with a very cold object. Tissue injury results from both freezing and vasoconstriction. Frostbite usually affects the distal aspects of the extremities or exposed parts of the face, such as the ears, nose, chin, and cheeks. Superficial frostbite involves the skin and subcutaneous tissue. Patients experience pain or paresthesia, and the skin appears white and waxy. After rewarming, there is cyanosis and erythema, wheal-and-flare formation, edema, and superficial blisters. Deep frostbite involves muscle, nerves, and deeper blood vessels. It may result in edema of the hand or foot, vesicles and bullae, tissue necrosis, and gangrene.

Initial treatment is rewarming, performed in an environment where reexposure to freezing conditions will not occur. Rewarming is accomplished by immersion of the affected part in a water bath at temperatures of 40°–44°C (104°–111°F). Massage, application of ice water, and extreme heat are contraindicated. The injured area should be cleansed with soap or antiseptic, and sterile dressings should be applied. Analgesics are often required during rewarming. Antibiotics are used if there is evidence of infection. The efficacy of sympathetic blocking drugs is not established. After recovery, the affected extremity may exhibit increased sensitivity to cold.

DISORDERS OF THE VEINS AND LYMPHATICS

Venous Disorders

Veins in the extremities can be broadly classified as either superficial or deep. In the lower extremity, the superficial venous system includes the greater and lesser saphenous veins and their tributaries. The deep veins of the leg accompany the major arteries. Perforating veins connect the superficial and deep systems at multiple locations. Bicuspid valves are present throughout the venous system to direct the flow of venous blood centrally.

VENOUS THROMBOSIS

The presence of thrombus within a superficial or deep vein, along with the accompanying inflammatory response in the vessel wall, is termed *venous thrombosis* or *thrombophlebitis*. Initially the thrombus is composed principally of platelets and fibrin. Red cells become interspersed with fibrin, and the thrombus tends to propagate in the direction of blood flow. The inflammatory response in the vessel wall may be minimal or characterized by granulocyte infiltration, loss of endothelium, and edema.

The factors that predispose to venous thrombosis were initially described by Virchow in 1856 and include stasis, vascular damage, and hypercoagulability. Accordingly, a variety of clinical situations are associated with increased risk of venous thrombosis **(Table 249-2)**. Venous thrombosis may occur in >50% of patients having orthopedic surgical procedures, particularly those involving the hip or knee, and in 10–40% of patients who undergo abdominal or thoracic operations. The prevalence of venous thrombosis is particularly high in patients with cancer of the pancreas, lungs, genitourinary tract, stomach, and breast. Approximately 10–20% of patients with idiopathicdeep vein thrombosis have or develop clinically overt cancer; there is no consensus on whether these individuals should be subjected to intensive diagnostic workup to search for occult malignancy.

Table 249–2. Conditions Associated with an Increased Risk for Development of Venous Thrombosis

Surgery
Orthopedic, thoracic, abdominal, and genitourinary procedures
Neoplasms
Pancreas, lung, ovary, testes, urinary tract, breast, stomach
Trauma
Fractures of spine, pelvis, femur, or tibia; spinal cord injuries
Immobilization
Acute myocardial infarction, congestive heart failure, stroke, postoperative convalescence

Pregnancy
Estrogen for replacement or contraception
Selective estrogen replacement modulators
Hypercoagulable states
Resistance to activated protein C; prothrombin 20210 A gene mutation deficiencies of antithrombin III, protein C, or protein S; antiphospholipid antibodies; myeloproliferative diseases; dysfibrinogenemia; disseminated intravascular coagulation
Venulitis Thromboangiitis obliterans, Behçet's disease, homocysteinuria
Previous deep vein thrombosis

The risk of thrombosis is increased after trauma such as fractures of the spine, pelvis, femur, and tibia. Immobilization, regardless of the underlying disease, is a major predisposing cause of venous thrombosis. This may account for the relatively high incidence in patients with acute myocardial infarction or congestive heart failure. The incidence of venous thrombosis is increased during pregnancy, particularly in the third trimester, and in the first month postpartum, as well as in individuals who use oral contraceptives, postmenopausal hormone replacement therapy, or selective estrogen receptor modulators. A variety of inherited and acquired disorders that produce systemic hypercoagulability, including resistance to activated protein C (factor V Leiden); prothrombin G20210A gene mutation; antithrombin III, protein C, and protein S deficiencies; antiphospholipid syndrome; hyperhomocysteinemia; SLE; myeloproliferative diseases; dysfibrinogenemia; heparin-induced thrombocytopenia; and disseminated intravascular coagulation, are associated with venous thrombosis. Venulitis occurring in thromboangiitis obliterans, Behçet's syndrome, and homocystinuria may also cause venous thrombosis.

DEEP VENOUS THROMBOSIS AND PULMONARY THROMBOEMBOLISM

See Chap. 262.

SUPERFICIAL VEIN THROMBOSIS

Thrombosis of the greater or lesser saphenous veins or their tributaries (i.e., superficial vein thrombosis) does not result in pulmonary embolism. It is associated with intravenous catheters and infusions, occurs in varicose veins, and may develop in association with deep venous thrombosis (DVT). Migrating superficial vein thrombosis is often a marker for a carcinoma and may also occur in patients with vasculitides, such as thromboangiitis obliterans. The clinical features of superficial vein thrombosis are easily distinguished from those of DVT. Patients complain of pain localized to the site of the thrombus. Examination reveals a reddened, warm, and tender cord extending along a superficial vein. The surrounding area may be red and edematous.

Treatment Superficial Vein Thrombosis

Treatment is primarily supportive. Initially, patients can be placed at bed rest with leg elevation and application of warm compresses. Nonsteroidal anti-inflammatory drugs may provide analgesia but may also obscure clinical evidence of thrombus propagation. If a thrombosis of the greater saphenous vein develops in the thigh and extends toward the saphenofemoral vein junction, it is reasonable to consider anticoagulant therapy to prevent extension of the thrombus into the deep system and a possible pulmonary embolism.

VARICOSE VEINS

Varicose veins are dilated, tortuous superficial veins that result from defective structure and function of the valves of the saphenous veins, intrinsic weakness of the vein wall, high intraluminal pressure, or, rarely, arteriovenous fistulas. Varicose veins can be categorized as primary or secondary. Primary varicose veins originate in the superficial system and occur two to three times as frequently in women as in men. Approximately one-half of these patients have a family history of varicose veins. Secondary varicose veins result from deep venous insufficiency and incompetent perforating veins or from deep venous occlusion that causes enlargement of superficial veins that are serving as collaterals.

Patients with venous varicosities are often concerned about the cosmetic appearance of their legs. Symptoms consist of a dull ache or pressure sensation in the legs after prolonged standing; this is relieved with leg elevation. The legs feel heavy, and mild ankle edema develops occasionally. Extensive venous varicosities may cause skin ulcerations near the ankle. Superficial venous thrombosis may be a recurring problem, and, rarely, a varicosity ruptures and bleeds. Visual inspection of the legs in the dependent position usually confirms the presence of varicose veins.

Varicose veins usually can be treated with conservative measures. Symptoms often decrease when the legs are elevated periodically, prolonged standing is avoided, and elastic support hose are worn. External compression stockings provide a counterbalance to the hydrostatic pressure in the veins. Ablative procedures, including sclerotherapy, endovenous radiofrequency or laser ablation, and surgery, may be considered to treat varicose veins in selected patients who have persistent symptoms,

have recurrent superficial vein thrombosis, and/or develop skin ulceration. Ablative therapy may also be indicated for cosmetic reasons. Small, symptomatic varicose veins can be treated with sclerotherapy, in which a sclerosing solution is injected into the involved varicose vein and a compression bandage is applied. Percutaneous, endovenous delivery of radiofrequency or laser energy can be used to treat incompetent great saphenous veins. Surgical therapy usually involves ligation and stripping of the great and small saphenous veins.

CHRONIC VENOUS INSUFFICIENCY

Chronic venous insufficiency may result from DVT and/or valvular incompetence. After DVT, the delicate valve leaflets become thickened and contracted so that they cannot prevent retrograde flow of blood; the vein becomes rigid and thick walled. Although most veins recanalize after an episode of thrombosis, the large proximal veins may remain occluded. Secondary incompetence develops in distal valves because high pressures distend the vein and separate the leaflets. Primary deep venous valvular dysfunction may also occur without previous thrombosis. Patients with venous insufficiency often complain of a dull ache in the leg that worsens with prolonged standing and resolves with leg elevation. Examination demonstrates increased leg circumference, edema, and superficial varicose veins. Erythema, dermatitis, and hyperpigmentation develop along the distal aspect of the leg, and skin ulceration may occur near the medial and lateral malleoli (**Fig. 249-3**). Cellulitis may be a recurring problem. The CEAP (clinical, etiologic, anatomic, pathophysiologic) classification schema incorporates the range of symptoms and signs of chronic venous insufficiency to characterize its severity (**Table 249-3**).

Figure 249-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Venous insufficiency with active venous ulcer near the medial malleolus. *(Courtesy of Dr. Steven Dean, with permission.)*

Table 249–3. CEAP (Clinical, Etiologic, Anatomic, Pathophysiologic) Classification

C0 No visible or palpable signs of venous disease
C1 Telangiectases, reticular veins
C2 Varicose veins
C3 Edema without skin changes
C4 Skin changes, including pigmentation, eczema, lipodermatosclerosis, and *atrophie blanche*
C5 Healed venous ulcer
C6 Active venous ulcer

Patients should be advised to avoid prolonged standing or sitting; frequent leg elevation is helpful. Graduated compression stockings should be worn during the day. These efforts should be intensified if

skin ulcers develop. Ulcers should be treated with applications of wet to dry dressings or occlusive hydrocolloid dressings. Commercially available compressive dressings that consist of paste with zinc oxide, calamine, glycerin, and gelatin may be applied and should be changed weekly until healing occurs. Recurrent ulceration and severe edema may be treated by surgical interruption of incompetent communicating veins. Subfascial endoscopic perforator surgery (SEPS) is a minimally invasive technique to interrupt incompetent communicating veins. Rarely, surgical valvuloplasty and bypass of venous occlusions are employed.

Lymphatic Disorders

Lymphatic capillaries are blind-ended tubes formed by a single layer of endothelial cells. The absent or widely fenestrated basement membrane of lymphatic capillaries allows access to interstitial proteins and particles. Lymphatic capillaries merge to form larger vessels that contain smooth muscle and are capable of vasomotion. Small- and medium-size lymphatic vessels empty into progressively larger channels, most of which drain into the thoracic duct. The lymphatic circulation is involved in the absorption of interstitial fluid and in the response to infection.

LYMPHEDEMA

Lymphedema may be categorized as primary or secondary **(Table 249-4)**. The prevalence of primary lymphedema is approximately 1 per 10,000 individuals. Primary lymphedema may be secondary to agenesis, hypoplasia, or obstruction of the lymphatic vessels. It may be associated with Turner's syndrome, Klinefelter's syndrome, Noonan's syndrome, yellow nail syndrome, intestinal lymphangiectasia syndrome, and lymphangiomyomatosis. Women are affected more frequently than are men. There are three clinical subtypes: congenital lymphedema, which appears shortly after birth; lymphedema praecox, which has its onset at the time of puberty; and lymphedema tarda, which usually begins after age 35. Familial forms of congenital lymphedema (Milroy's disease) and lymphedema praecox (Meige's disease) may be inherited in an autosomal dominant manner with variable penetrance; autosomal or sex-linked recessive forms are less common.

Table 249–4. Causes of Lymphedema

Primary	Secondary
Congenital (includes Milroy's disease) Lymphedema praecox [includes Nonne-Milroy-Meige (or chronic familial lymhedema of the limbs) disease] Lymphedema tarda	Recurrent lymphangitis Filariasis Tuberculosis Neoplasm Surgery Radiation therapy

Secondary lymphedema is an acquired condition that results from damage to or obstruction of previously normal lymphatic channels (Table 249-4). Recurrent episodes of bacterial lymphangitis, usually caused by streptococci, are a very common cause of lymphedema. The most common cause of secondary lymphedema worldwide is filariasis (Chap. 218). Tumors, such as prostate cancer and lymphoma, can also obstruct lymphatic vessels. Both surgery and radiation therapy for breast carcinoma may cause lymphedema of the upper extremity. Less common causes include tuberculosis, contact dermatitis, lymphogranuloma venereum, rheumatoid arthritis, pregnancy, and self-induced or factitious lymphedema after application of tourniquets.

Lymphedema is generally a painless condition, but patients may experience a chronic dull, heavy sensation in the leg, and most often they are concerned about the appearance of the leg. Lymphedema of the lower extremity, initially involving the foot, gradually progresses up the leg so that the entire limb becomes edematous. In the early stages, the edema is soft and pits easily with pressure. In the chronic stages, the limb has a woody texture, and the tissues become indurated and fibrotic. At this point the edema may no longer be pitting. The limb loses its normal contour, and the toes appear square. Lymphedema should be distinguished from other disorders that cause unilateral leg swelling, such as DVT and chronic venous insufficiency. In the latter condition, the edema is softer, and there is often evidence of a stasis dermatitis, hyperpigmentation, and superficial venous varicosities. Other causes of leg swelling that resemble lymphedema are pretibial myxedema and lipedema. Pretibial myxedema occurs in patients with hyperthyroidism, especially Graves' disease, and is caused by deposition of hyaluronic acid-rich protein in the dermis. Lipedema usually occurs in women and is caused by accumulation of adipose tissue in the leg from the thigh to the ankle with sparing of the feet. The evaluation of patients with lymphedema should include diagnostic studies to clarify the cause. Abdominal and pelvic ultrasound and CT can be used to detect obstructing lesions such as neoplasms. MRI may reveal edema in the epifascial compartment and identify lymph nodes and enlarged lymphatic channels. Lymphoscintigraphy and lymphangiography are rarely indicated, but

either can be used to confirm the diagnosis or differentiate primary from secondary lymphedema. Lymphoscintigraphy involves the injection of radioactively labeled technetium-containing colloid into the distal subcutaneous tissue of the affected extremity. In lymphangiography, contrast material is injected into a distal lymphatic vessel that has been isolated and cannulated. In primary lymphedema, lymphatic channels are absent, hypoplastic, or ectatic. In secondary lymphedema, lymphatic channels are usually dilated, and it may be possible to determine the level of obstruction.

Treatment: Lymphedema

Patients with lymphedema of the lower extremities must be instructed to take meticulous care of their feet to prevent recurrent lymphangitis. Skin hygiene is important, and emollients can be used to prevent drying. Prophylactic antibiotics are often helpful, and fungal infection should be treated aggressively. Patients should be encouraged to participate in physical activity; frequent leg elevation can reduce the amount of edema. Physical therapy, including massage to facilitate lymphatic drainage, may be helpful. Patients can be fitted with graduated compression hose to reduce the amount of lymphedema that develops with upright posture. Occasionally, intermittent pneumatic compression devices can be applied at home to facilitate reduction of the edema. Diuretics are contraindicated and may cause depletion of intravascular volume and metabolic abnormalities. Microsurgical lymphaticovenous anastomotic procedures have been performed to rechannel lymph flow from obstructed lymphatic vessels into the venous system.

FURTHER READINGS

Bergan J J et al: Chronic venous disease. N Engl J Med 355:488, 2006[PMID: 16885552]

Creager MA (ed): *Atlas of Vascular Disease, 3rd ed.* Philadelphia, Current Medicine Group, 2007

——— et al (eds): *Vascular Medicine*. Philadelphia, Saunders Elsevier, 2006

——— et al. Atherosclerotic Peripheral Vascular Disease Symposium II: Executive summary. Circulation 118:2811, 2008

Fowkes FG et al: Ankle brachial index combined with Framingham Risk Score to predict cardiovascular events and mortality: A meta-analysis. JAMA 300:197, 2008[PMID: 18612117]

Hirsch AT et al: ACC/AHA 2005 guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic). J Am Coll Cardiol 47:1239, 2006[PMID: 16545667]

Norgren L et al: Inter-Society Consensus for the Management of Peripheral Arterial Disease (TASC II). J Vasc Surg 45:S5, 2007

Raju S, Neglen P: Clinical practice: Chronic venous insufficiency and varicose veins. N Engl J Med 360:2319, 2009[PMID: 19474429]

Rockson SG: Current concepts and future directions in the diagnosis and management of lymphatic vascular disease. Vasc Med 15:223, 2010[PMID: 20483987]

White C: Clinical practice: Intermittent claudication. N Engl J Med 356:1241, 2007[PMID: 17377162]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PULMONARY HYPERTENSION: INTRODUCTION

Pulmonary hypertension, an abnormal elevation in pulmonary artery pressure, may be the result of left heart failure, pulmonary parenchymal or vascular disease, thromboembolism, or a combination of these factors. Whether the pulmonary hypertension arises from cardiac, pulmonary, or intrinsic vascular disease, it generally is a feature of advanced disease. Because the causes of pulmonary hypertension are so diverse, it is essential that the etiology underlying the pulmonary hypertension be clearly determined before beginning treatment.

Pathophysiology

The right ventricle responds to an increase in pulmonary vascular resistance by increasing right ventricular (RV) systolic pressure to preserve cardiac output. In some patients, chronic changes occur in the pulmonary circulation, resulting in progressive remodeling of the vasculature, which can sustain or promote pulmonary hypertension even if the initiating factor is removed.

The ability of the RV to adapt to increased vascular resistance is influenced by several factors, including age and the rapidity of the development of pulmonary hypertension. For example, a large acute pulmonary thromboembolism can result in RV failure and shock, whereas chronic thromboembolic disease of equal severity may result in only mild exercise intolerance. Coexisting hypoxemia can impair the ability of the ventricle to compensate. Studies support the concept that RV failure occurs in pulmonary hypertension when the RV myocardium becomes ischemic as a result of excessive demands and inadequate RV coronary blood flow. The onset of RV failure, often manifest by peripheral edema, is associated with a poor outcome.

Diagnosis

The most common symptom attributable to pulmonary hypertension is exertional dyspnea. Other common symptoms are fatigue, angina pectoris, syncope, near syncope, and peripheral edema.

The physical examination typically reveals increased jugular venous pressure, a reduced carotid pulse, and a palpable RV impulse. Most patients have an increased pulmonic component of the second heart sound, a right-sided fourth heart sound, and tricuspid regurgitation (Chap. 227). Peripheral cyanosis and/or edema tend to occur in later stages of the disease.

LABORATORY FINDINGS

(Fig. 250-1) The chest x-ray generally shows enlarged central pulmonary arteries. The lung fields may reveal other pathology. The electrocardiogram usually shows right axis deviation and RV hypertrophy. The echocardiogram commonly demonstrates RV and right atrial enlargement, a reduction in left ventricular (LV) cavity size, and a tricuspid regurgitant jet that can be used to estimate RV systolic pressure by Doppler. Pulmonary function tests are helpful in documenting underlying obstructive airways disease, whereas high-resolution chest computed tomography (CT) is preferred to diagnose restrictive lung disease. Hypoxemia and an abnormal diffusing capacity for carbon monoxide occur with pulmonary hypertension of many causes. A perfusion lung scan is almost always abnormal in patients with thromboembolic pulmonary hypertension (Chap. 262). However, diffuse defects of a nonsegmental nature often can be seen in long-standing pulmonary hypertension in the absence of thromboemboli. Laboratory tests should include antinuclear antibody and HIV testing. Because of the high frequency of thyroid abnormalities in patients with idiopathic pulmonary hypertension, it is recommended that the thyroid-stimulating hormone level be determined periodically.

Figure 250-1

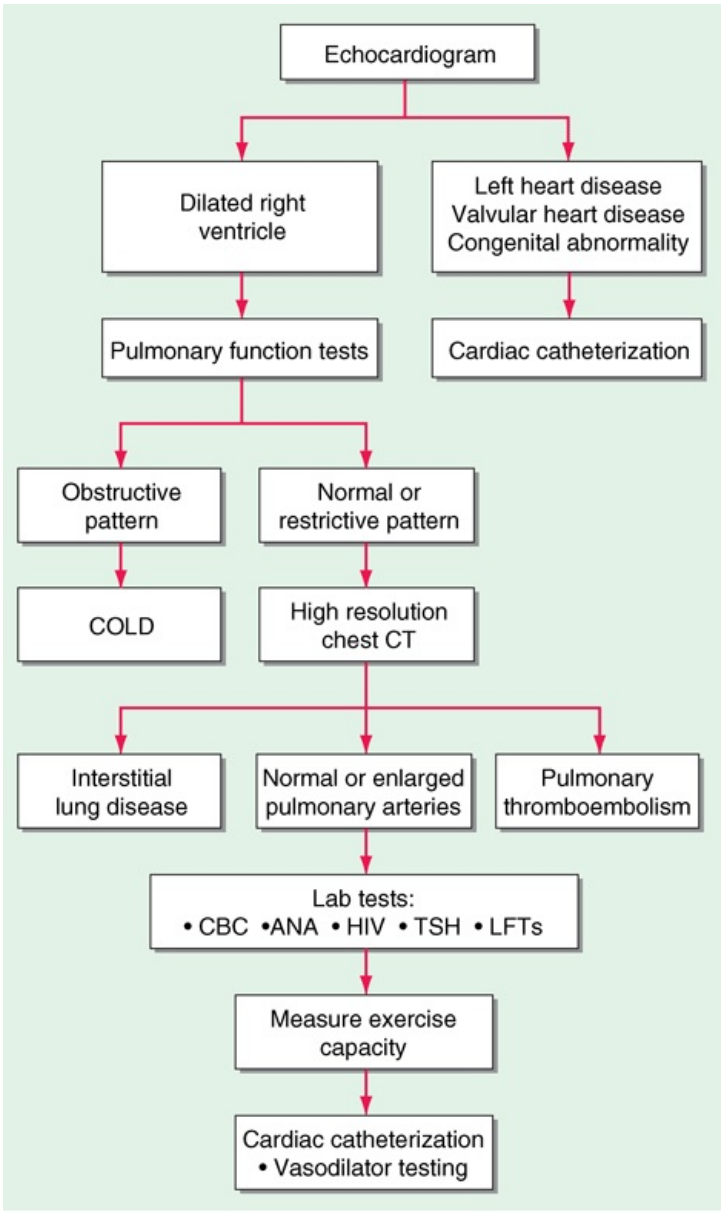

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An algorithm for the workup of a patient with unexplained pulmonary hypertension. All potential etiologies and associated conditions must be investigated in a patient with clinical findings consistent with pulmonary hypertension. COLD, chronic obstructive lung disease; CBC, complete blood count; ANA, antinuclear antibodies; HIV, human immunodeficiency virus; TSH, thyroid-stimulating hormone; LFTs, liver function tests.

CARDIAC CATHETERIZATION

Cardiac catheterization is mandatory for accurate measurement of pulmonary artery pressure, cardiac output, and LV filling pressure as well as documentation of an underlying cardiac shunt. Care should be taken to record pressures only at end expiration. It is recommended that patients with pulmonary arterial hypertension undergo drug testing with a short-acting pulmonary vasodilator to determine the extent of pulmonary vasodilator reactivity. Inhaled nitric oxide, intravenous adenosine, and intravenous epoprostenol have comparable effects in reducing pulmonary artery pressure acutely. Nitric oxide is administered via inhalation in 10–20 parts per million. Adenosine is given in doses of 50 μg/kg per min and increased every 2 min until side effects develop. Epoprostenol is given in doses of 2 ng/kg per min and increased every 30 min until side effects develop. Patients who respond usually can be treated with calcium channel blockers and have a more favorable prognosis.

PULMONARY ARTERIAL HYPERTENSION

Pulmonary arterial hypertension (PAH) refers to a variety of diseases that include idiopathic PAH, as noted in **Table 250-1**. Patients with PAH have a common histopathology characterized by medial hypertrophy, eccentric and concentric intimal fibrosis, recanalized thrombi appearing as fibrous webs, and plexiform lesions.

Table 250–1. A Clinical Classification of Pulmonary Hypertension

Category 1. Pulmonary arterial hypertension (PAH)

Key feature: elevation in pulmonary arterial pressure (PAP) with normal pulmonary capillary wedge pressure (pcwp)

Includes:

Idiopathic (IPAH)

- Sporadic
- Familial
- Exposure to drugs or toxins
- Persistent pulmonary hypertension of the newborn
- Pulmonary capillary hemangiomatosis (PCH)

Associated with other active conditions

- Collagen vascular disease
- Congenital systemic-to-pulmonary shunts
- Portal hypertension
- HIV infection

Category 2. Pulmonary venous hypertension

Key feature: elevation in PAP with elevation in pcwp

Includes:

- Left-sided atrial or ventricular heart disease
- Left-sided valvular heart disease
- Pulmonary venous obstruction
- Pulmonary venoocclusive disease (PVOD)

Category 3. Pulmonary hypertension associated with hypoxemic lung disease

Key feature: chronic hypoxia with mild elevation of PAP

Includes:

- Chronic obstructive lung disease
- Interstitial lung disease
- Sleep-disordered breathing
- Alveolar hypoventilation disorders
- Chronic exposure to high altitude
- Developmental abnormalities

Category 4. Pulmonary hypertension due to chronic thromboembolic disease

Key feature: elevation of PA pressure with documentation of pulmonary arterial obstruction for >3 months

Includes:

- Chronic pulmonary thromboembolism
- Nonthrombotic pulmonary embolism (tumor, foreign material)

Category 5. Miscellaneous

Key feature: elevation in PAP in association with a systemic disease where a causal relationship is not clearly understood.

Includes:

- Sarcoidosis
- Chronic anemias
- Histiocytosis X
- Lymphangiomatosis
- Schistosomiasis

Pathobiology

Vasoconstriction, vascular proliferation, thrombosis, and inflammation appear to underlie the development of PAH **(Fig. 250-2).** Abnormalities in multiple molecular pathways and genes that regulate the pulmonary vascular endothelial and smooth muscle cells have been identified. These abnormalities include decreased expression of the voltage-regulated potassium channel, mutations in the bone morphogenetic protein-2 receptor, increased tissue factor expression, overactivation of the serotonin transporter, transcription factor activation of hypoxia-inducible factor-1 alpha, and activation of nuclear factor of activated T cells. As a result, there appears to be loss of apoptosis of the smooth muscle cells that allows their proliferation and the emergence of apoptosis-resistant endothelial cells that can obliterate the vascular lumen. In addition, thrombin deposition in the pulmonary vasculature from a procoagulant state that develops as an independent abnormality or as a result of endothelial dysfunction may amplify the vascular proliferation.

Figure 250-2

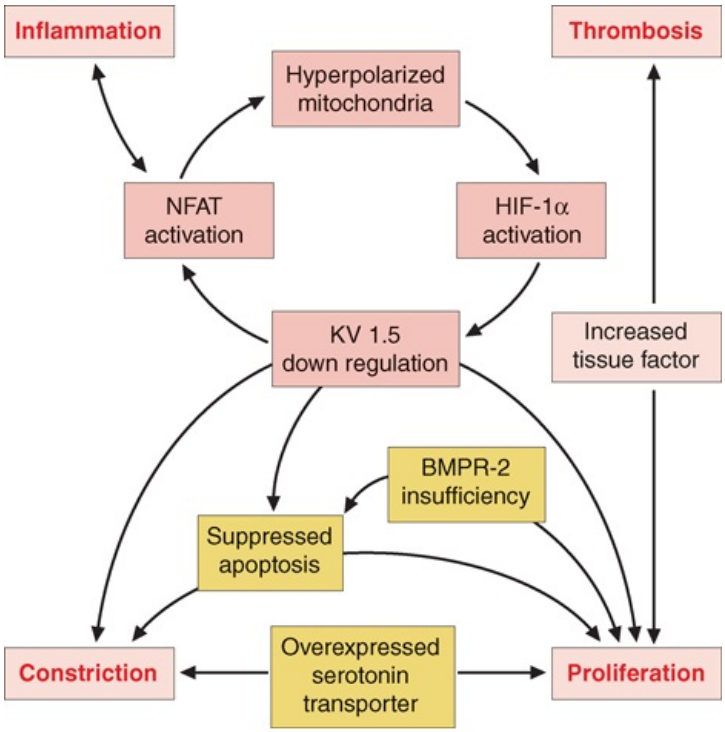

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple biologic pathways that can lead to pulmonary arterial hypertension. Some of the better-characterized ones are illustrated. Because of the redundancy in these pathways and the spectrum of abnormalities that may coexist, it is unlikely that a single agent will produce disease reversal. BMPR-2, bone morphogenetic protein receptor-2; HIF, hypoxia inducible factor; KV 1.5, voltage-regulated potassium channel 1.5; NFAT, nuclear factor of activated T cells.

IDIOPATHIC PULMONARY ARTERIAL HYPERTENSION

Idiopathic pulmonary arterial hypertension (IPAH), formerly referred to as primary pulmonary hypertension, is uncommon, with an estimated incidence of two cases per million. There is a female predominance, with most patients presenting in the fourth and fifth decades, although the age range is from infancy to >60 years.

Familial IPAH accounts for up to 20% of cases of IPAH and is characterized by autosomal dominant inheritance and incomplete penetrance. The clinical and pathologic features of familial and sporadic IPAH are identical. Heterozygous germ-line mutations involving the gene that code the type II bone morphogenetic protein receptor (BMPR II), a member of the transforming growth factor (TGF) ß superfamily, appear to account for most cases of familial IPAH. The TGF-ß superfamilies include multifunctional proteins that initiate diverse cellular responses by binding to and activating serine/threonine kinase receptors. The low gene penetrance indicates that other risk factors or abnormalities are necessary to manifest clinical disease. Germ-line mutations in the activin-like kinase gene and endoglin gene, which have been linked to hereditary hemorrhagic telangiectasia, coexist in some patients with familial IPAH.

Natural History

The natural history of IPAH is uncertain, but the disease typically is diagnosed late in its course. Before current therapies, a mean survival of 2–3 years from the time of diagnosis was reported.

Functional class remains a strong predictor of survival, with patients who are in New York Heart Association (NYHA) functional class IV having a mean survival of <6 months. The cause of death is usually RV failure, which is manifest by progressive hypoxemia, tachycardia, hypotension, and edema.

Treatment: Pulmonary Arterial Hypertension

Because the pulmonary artery pressure in PAH increases with exercise, patients should be cautioned against participating in activities that impose physical stress. Diuretic therapy relieves peripheral edema and may be useful in reducing RV volume overload. Pulse oximetry should be monitored, as O_2 supplementation helps alleviate dyspnea and RV ischemia in patients whose arterial O_2 saturation is reduced. Anticoagulant therapy is advocated for all patients with PAH based on studies demonstrating that warfarin increases survival of patients with PAH. The dose of warfarin generally is titrated to achieve an international normalized ratio (INR) of 2–3 times control.

Several treatments are approved for PAH; they are reviewed below without making a distinction among the different types. However, the efficacy and side effects of these drugs may not be the same in all types of PAH. Other than calcium channel blockers, none of the drugs produce a significant lowering of the pulmonary arterial pressure, and their long-term effects on survival are undefined. The principles for the selection and use of the approved drug treatments are reviewed in **Table 250-2**.

Table 250–2. Principles of Drug Treatment of Pulmonary Arterial Hypertension

• Establish a correct diagnosis:
Patients should undergo cardiac catheterization before initiating therapy.
• Obtain baseline assessments of the disease:
Tests should be obtained to monitor the patient's response to therapy to know whether the treatments are effective.
• Test vasoreactivity:
Patients should be tested at the time of diagnosis so that reactive patients are not missed.
• Reactive patients should be treated with calcium channel blockers:
Calcium blockers in high doses are the drugs of choice.
• Nonreactive patients should be offered other therapies:
No specific treatment has been established as first-line therapy.
• Periodic follow-up assessment of drug efficacy is essential:
Repeat assessments should be performed within 8 weeks of initiating a new drug, as patients who do not respond initially are not likely to respond with longer exposure.
Therapies can lose efficacy over time.
• Ineffective treatments should be replaced:
A different treatment should be substituted rather than added.
Patients who fail all treatments should be considered for lung transplantation.
• Benefits and risks of combination therapies are largely unknown:
Only the addition of sildenafil to epoprostenol has been shown to be efficacious.

CALCIUM CHANNEL BLOCKERS

Patients who respond to short-acting vasodilators at the time of cardiac catheterization (a fall in mean pulmonary arterial pressure ≥10 mmHg and a final mean pressure <40 mmHg) should be treated with calcium channel blockers. Typically, these patients require high doses (e.g., nifedipine, 240 mg/d, or amlodipine, 20 mg/d). Patients may have dramatic reductions in pulmonary artery pressure and pulmonary vascular resistance associated with improved symptoms, regression of RV hypertrophy, and improved survival now documented to exceed 20 years. However, <20% of patients respond to calcium channel blockers in the long term. These drugs are not effective in patients who are not vasoreactive. They also have not been approved for the treatment of PAH by the U.S. Food and Drug Administration.

ENDOTHELIN RECEPTOR ANTAGONISTS

The endothelin receptor antagonists *bosentan* and *ambrisentan* are approved treatments of PAH. In randomized clinical trials, both improved exercise tolerance as measured by an increase in 6-min walking distance. Therapy with bosentan is initiated at 62.5 mg bid for the first month and increased to 125 mg bid thereafter. Ambrisentan is initiated as 5 mg once daily and can be increased to 10 mg daily. Because of the high frequency of abnormal hepatic function tests associated with these drugs, primarily an increase in transaminases, it is recommended that liver function be monitored monthly throughout the duration of use. Bosentan is contraindicated in patients who are on cyclosporine or glyburide concurrently.

PHOSPHODIESTERASE-5 INHIBITORS

Sildenafil and *tadalafil*, phosphodiesterase-5 inhibitors, are approved for the treatment of PAH. Phosphodiesterase-5 is responsible for the hydrolysis of cyclic GMP in pulmonary vascular smooth muscle, the mediator through which nitric oxide lowers pulmonary artery pressure and inhibits pulmonary vascular growth. Clinical trials have shown that both drugs improve exercise tolerance in patients with PAH. The effective dose for sildenafil is 20–80 mg tid. The effective dose for tadalafil is 40 mg once daily. The most common side effect is headache. Neither drug should be given to patients who are taking nitrovasodilators.

PROSTACYCLINS

Iloprost, a prostacyclin analogue, is approved via inhalation for PAH. It has been shown to improve a composite measure of symptoms and exercise tolerance by 10%. Therapy can be given at either 2.5 or 5 μg per inhalation treatment via a dedicated nebulizer. The most common side effects are flushing and cough. Because of the very short half-life (<30 min) it is recommended that treatments be administered as often as every 2 h.

Epoprostenol is approved as a chronic IV treatment of PAH. Clinical trials have demonstrated an improvement in symptoms, exercise tolerance, and survival even if no acute hemodynamic response to drug challenge occurs. Drug administration requires placement of a permanent central venous catheter and infusion through an ambulatory infusion pump system. Side effects include flushing, jaw pain, and diarrhea, which are tolerated by most patients.

Treprostinil, an analogue of epoprostenol, is approved for PAH and may be given intravenously, subcutaneously, or via inhalation. Clinical trials have demonstrated an improvement in symptoms with exercise. Local pain at the infusion site with subcutaneous administration has caused most patients to switch to another therapy. Side effects are similar to those seen with epoprostenol.

The intravenous prostacyclins have the greatest efficacy as treatments for PAH and are often effective in patients who have failed all other treatments. Favorable properties include vasodilation, platelet inhibition, inhibition of vascular smooth muscle growth, and inotropic effects. It generally takes several months to titrate the dose of epoprostenol or treprostinil upward to achieve optimal clinical efficacy, which can be determined by symptoms, exercise testing, and catheterization. The optimal doses of these drugs have not been determined, but the typical doses of epoprostenol range from 25 to 40 ng/kg per min, and from 75 to 150 ng/kg per min for treprostinil. The major problem with intravenous therapy is infection related to the indwelling venous catheter, which requires close monitoring and diligence on the part of the patient. In addition, abrupt discontinuation of intravenous prostacyclins can lead to a rebound increase in pulmonary pressure.

It is recommended that every patient diagnosed with PAH be treated. Although no drug has been demonstrated to be superior as first-line therapy, many prefer to initiate treatment with an oral or inhaled form of therapy. Patients who fail to improve adequately within the first 2 months should be switched to a different therapy, as there is concern that delaying a more effective treatment may allow the disease to progress and become less responsive. The use of these drugs in combination has become popular, but the only randomized clinical trial demonstrating beneficial effects added sildenafil to patients treated with epoprostenol.

LUNG TRANSPLANTATION

(See also Chap. 266) Lung transplantation is considered for patients who, while on an intravenous prostacyclin, continue to manifest right heart failure. Acceptable results have been achieved with heart-lung, bilateral lung, and single-lung transplantation. The availability of donor organs often influences the choice of procedure.

CONDITIONS ASSOCIATED WITH PULMONARY HYPERTENSION

Collagen Vascular Disease

All the collagen vascular diseases may be associated with PAH. This complication occurs commonly with the CREST syndrome (calcinosis, Raynaud's phenomenon, esophageal involvement, sclerodactyly, and telangiectasia) and in scleroderma (Chap. 323) and less frequently in systemic lupus erythematosus (Chap. 319), Sjögren's syndrome (Chap. 324), dermatomyositis, polymyositis (Chap. 326), and rheumatoid arthritis (Chap. 321). Often these patients have coexistent interstitial pulmonary fibrosis even though it may not be apparent on chest x-ray, CT, or pulmonary function tests. Consequently, they tend to have hypoxemia as an important clinical feature, along with the other classic findings of pulmonary hypertension. A fall in diffusing capacity may precede the development of pulmonary hypertension. Treatment of these patients is identical to that of patients with IPAH (see above) but is less effective. The treatment of the pulmonary hypertension, however, does not affect the natural history of the underlying collagen vascular disease.

Congenital Systemic to Pulmonary Shunts

It is common for large post-tricuspid cardiac shunts (e.g., ventricular septal defect, patent ductus arteriosus) to produce severe PAH (Chap. 236). Although less common, this also may occur in

pretricuspid shunts (e.g., atrial septal defect, anomalous pulmonary venous drainage). In patients with uncorrected shunts, the clinical features include those associated with right-to-left shunting, such as hypoxemia and peripheral cyanosis, which worsen dramatically with exertion (Chap. 35). PAH may also occur years or even decades after surgical correction in the absence of right-to-left shunting. These patients present similarly to patients with IPAH but tend to have better long-term survival. The treatments are similar to those for IPAH.

Portal Hypertension

Portal hypertension is associated with PAH, but the mechanism is unknown. Patients with advanced cirrhosis can have the combined features of a high-output cardiac state in association with the features of pulmonary hypertension and RV failure. Thus, a normal cardiac output may actually reflect a marked impairment of RV function. The etiology of ascites and edema can be confusing in these patients because this condition can have both cardiac and hepatic causes. Overall, these patients have a worse prognosis than do patients with IPAH. Patients with mild pulmonary hypertension who have a favorable response to epoprostenol have undergone successful liver transplantation with improvement of the pulmonary vascular disease.

Anorexigens

A causal relationship has been established between exposure to several anorexigens, including aminorex and the fenfluramines, and the development of PAH. Often the pulmonary hypertension will develop years after the last exposure. Although the clinical features are identical to those of IPAH, the patients appear to be less responsive to medical treatments.

Pulmonary Capillary Hemangiomatosis

Pulmonary capillary hemangiomatosis is a very rare form of pulmonary hypertension. Histologically it is characterized by the presence of infiltrating thin-walled blood vessels throughout the pulmonary interstitium and walls of the pulmonary arteries and veins. The presenting symptoms are those of IPAH but often with hypoxemia or hemoptysis as a clinical feature. The diagnosis may be suggested by findings on chest CT. The clinical course is usually one of progressive deterioration leading to death. There is no established therapy.

PULMONARY VENOUS HYPERTENSION

Pulmonary hypertension occurs as a result of increased resistance to pulmonary venous drainage. It is associated with diastolic dysfunction of the left ventricle, diseases affecting the pericardium or mitral or aortic valves, and rare entities such as cor triatriatum, left atrial myxoma, extrinsic compression of the central pulmonary veins from fibrosing mediastinitis, and pulmonary venoocclusive disease. Pulmonary venous hypertension affects the pulmonary veins and venules, producing arterialization of the external elastic lamina, medial hypertrophy, and focal eccentric intimal fibrosis. Microcirculatory lesions include capillary congestion, focal alveolar edema, and dilation of the interstitial lymphatics. Although these lesions are potentially reversible, regression may take years after the underlying cause is removed. Pulmonary venous hypertension often triggers reactive vasoconstriction in the pulmonary arterial bed and results in proliferative changes of the intima and media that can produce severe elevations in pulmonary artery pressure. Clinically it may be confusing and appear as if two separate disease processes are occurring simultaneously. The distinction is important, however, as treatments that are effective in PAH may make patients with pulmonary venous hypertension worse.

Left Ventricular Diastolic Dysfunction

Pulmonary hypertension as a result of LV diastolic failure is common but often unrecognized (Chap. 234). It can occur with or without LV systolic failure. The most common risk factors are hypertensive heart disease; coronary artery disease; and impaired LV compliance related to age, diabetes, obesity, and hypoxemia. Symptoms of orthopnea and paroxysmal nocturnal dyspnea are prominent. Many patients improve considerably if LV end-diastolic pressure is lowered, but current treatments are unsatisfactory.

Mitral Valve Disease

Mitral stenosis and mitral regurgitation represent important causes of pulmonary hypertension (Chap. 237) from reactive pulmonary vasoconstriction resulting in marked elevations in pulmonary artery pressures. An echocardiogram usually shows abnormalities such as thickened mitral valve leaflets with reduced mobility or severe mitral regurgitation documented by Doppler echocardiography (Chap. 229). At cardiac catheterization, a pressure gradient between the pulmonary capillary wedge pressure and LV end-diastolic pressure is diagnostic of mitral stenosis.

In patients with mitral stenosis, corrective surgery of the mitral valve or mitral balloon valvuloplasty predictably results in a reduction in pulmonary artery pressure and pulmonary vascular resistance. Patients with mitral regurgitation, however, may not have as dramatic a response to surgery because of persistent elevations in LV end-diastolic pressure.

Pulmonary Venoocclusive Disease

Pulmonary venoocclusive disease is a rare and distinct pathologic entity found in <10% of patients who present with unexplained pulmonary hypertension. Histologically it is manifest by intimal proliferation and fibrosis of the intrapulmonary veins and venules, occasionally extending to the arteriolar bed. A CT scan may reveal septal thickening, diffuse or mosaic ground-glass opacities, multiple small nodules, or areas of alveolar consolidation. Advanced pulmonary venous obstruction explains the orthopnea that can mimic LV failure, pulmonary edema noted on chest x-ray, and the increase in pulmonary capillary wedge pressure at catheterization. Effective therapy for this condition has not been established.

PULMONARY HYPERTENSION ASSOCIATED WITH LUNG DISEASE AND HYPOXEMIA

The acute hypoxic response of the pulmonary arterial smooth muscle cells involves inhibition of the potassium current, membrane depolarization, and calcium entry through L-type calcium channels. Hypoxia, acting through the small G protein RhoA, stimulates Rho kinase, which inhibits myosin vs. heavy chain in light chain phosphatase, thereby increasing phosphorylation of the light chain and augmenting contraction. Chronic hypoxia results in muscularization of the arterioles with minimal effects on the intima. When it occurs as an isolated entity, the changes produced are potentially reversible.

Although chronic hypoxia is an established cause of pulmonary hypertension, it rarely leads to an increase in the systolic pulmonary artery pressure >50 mmHg. Polycythemia in response to the hypoxemia is a characteristic finding. Hypoxia also may occur in conjunction with other causes of pulmonary hypertension associated with more extensive vascular changes. Clinically, the hypoxia has an added adverse effect. Patients with chronic hypoxia who have a marked elevation in pulmonary pressure should be evaluated for the other causes of the pulmonary hypertension.

Chronic Obstructive Lung Disease

Chronic obstructive lung disease (COLD) is associated with mild pulmonary hypertension in the advanced stages (Chap. 260). The factors leading to an increase in pulmonary vascular resistance are numerous, but alveolar hypoxia is considered the predominant one. The presence of pulmonary hypertension in patients with COLD confers a worse outcome. The only effective therapy is supplemental oxygen. Clinical trials have documented that continuous oxygen therapy relieves the pulmonary vasoconstriction, reverses chronic ischemia throughout the systemic and pulmonary vascular beds, and improves survival. Long-term oxygen therapy is indicated if the resting arterial Po_2 remains <55 mmHg. Pulmonary vasodilators can worsen gas exchange and should not be used.

Interstitial Lung Disease

Pulmonary hypertension is common in interstitial lung disease that results from parenchymal and vascular remodeling (Chap. 261). Coexisting hypoxemia occurs frequently and contributes to morbidity. Interstitial lung disease often is associated with the collagen vascular diseases. Many patients have pulmonary fibrosis of unknown etiology. Patients are commonly older than 50 years and report an insidious onset of progressive dyspnea and cough for months to years. It is uncommon for the mean pulmonary artery pressure to exceed 40 mmHg. The pulmonary vasodilators approved for PAH have not been shown to be helpful.

Sleep-Disordered Breathing

The incidence of pulmonary hypertension in the setting of *obstructive sleep apnea*, a common condition (Chap. 265), is <20% and is generally mild. Some patients have severe pulmonary hypertension in conjunction with sleep apnea, which may be unrelated. It is recommended that the sleep apnea and the PAH be treated as coexisting problems.

Alveolar Hypoventilation

Pulmonary hypertension can occur in patients with chronic hypoventilation and hypoxia secondary to thoracovertebral deformities. Symptoms are slowly progressive and are related to hypoxemia (Chap. 264). In patients with advanced disease, intermittent positive-pressure breathing and supplemental oxygen have been used successfully.

Pulmonary hypertension secondary to hypoxemia has been reported in patients with neuromuscular disease as a result of generalized weakness of the respiratory muscles and in patients with diaphragmatic paralysis, generally from trauma to the phrenic nerve. Patients with nontraumatic bilateral diaphragmatic paralysis may go unrecognized until they present with either respiratory failure or pulmonary hypertension.

PULMONARY HYPERTENSION DUE TO THROMBOEMBOLIC DISEASE

Deep Venous Thrombosis and Pulmonary Embolism

See Chap. 262.

Chronic Thromboembolic Pulmonary Hypertension

Most patients treated for acute pulmonary thromboembolism with intravenous heparin and chronic oral warfarin do not develop chronic pulmonary hypertension. However, some patients have impaired fibrinolytic resolution of the thromboembolism, which leads to organization and incomplete recanalization and chronic obstruction of the pulmonary vascular bed. Because the initial pulmonary thromboembolism goes undetected or untreated, many patients are misdiagnosed as having IPAH. These patients may have underlying thrombophilic disorders, such as the lupus anticoagulant/anticardiolipin antibody syndrome, prothrombin gene mutation, or factor V Leiden (Chap. 117).

DIAGNOSIS

The physical examination is characteristic of pulmonary hypertension but may include bruits heard over areas of the lung, representing blood flow through vessels with partial occlusion. A perfusion lung scan or contrast-enhanced spiral CT scan should reveal multiple thromboemboli. High-resolution CT scanning is necessary to document the location and proximal extent of the thromboemboli and hence the potential for operability.

TREATMENT: CHRONIC THROMBOEMBOLIC PULMONARY HYPERTENSION

Pulmonary thromboendarterectomy is an established surgical treatment in patients whose thrombi are accessible to surgical removal. The operative mortality is <10% in experienced centers. Postoperative survivors can expect an improvement in functional class and exercise tolerance. Lifelong anticoagulation using warfarin is mandatory. Thrombolytic therapy is rarely helpful in patients with chronic thromboembolic pulmonary hypertension and may expose them to the increased risk of bleeding without potential benefit.

OTHER DISORDERS AFFECTING THE PULMONARY VASCULATURE

Sarcoidosis

Sarcoidosis can produce pulmonary hypertension as a result of fibrocystic lung involvement (Chap. 329) or direct cardiovascular involvement. Consequently, patients with sarcoidosis who present with progressive dyspnea and pulmonary hypertension require a thorough evaluation. There is a subset of patients with sarcoidosis and severe pulmonary hypertension who exhibit a favorable response to epoprostenol therapy.

Sickle Cell Disease

Cardiovascular system abnormalities are prominent in the clinical spectrum of sickle cell disease, including pulmonary hypertension. The etiology is multifactorial, including hemolysis, hypoxemia, thromboembolism, chronic high cardiac output, and chronic liver disease. The presence of pulmonary hypertension in patients with sickle cell disease is associated with higher mortality. Intensification of sickle cell disease–specific therapy appears to reduce the morbidity. Clinical trials assessing drugs to treat pulmonary hypertension are ongoing, but the efficacy of those drugs is unknown.

Schistosomiasis

Although extremely rare in North America, schistosomiasis is one of the most common causes of pulmonary hypertension worldwide (Chap. 219). The development of pulmonary hypertension occurs in the setting of hepatosplenic disease and portal hypertension. Studies suggest that inflammation from the infection triggers the pulmonary vascular changes that occur. The diagnosis is confirmed by finding the parasite ova in the urine or stools of patients with symptoms, which can be difficult. The efficacy of therapies directed toward pulmonary hypertension in these patients is unknown.

HIV Infection

The mechanism by which HIV infection produces pulmonary hypertension is unknown (Chap. 189). Although the incidence is estimated at 1 per 200 cases, the marked rise in the prevalence of HIV infection worldwide could have a significant impact on the frequency with which these entities are seen in combination. The evaluation and treatments are identical to those for IPAH. Treatment of the HIV infection does not appear to affect the severity or natural history of the underlying pulmonary hypertension.

FURTHER READINGS

Condliffe R: Connective tissue disease-associated pulmonary arterial hypertension in the modern treatment era. Am J Respir Crit Care Med 179:151, 2009[PMID: 18931333]

Diller GP, Gatzoulis MA: Pulmonary vascular disease in adults with congenital heart disease. Circulation 115:1039, 2007[PMID: 17325254]

Macchia A et al: Systematic review of trials using vasodilators in pulmonary arterial hypertension: Why a new approach is needed. Am Heart J 159:245, 2010[PMID: 20152223]

McLaughlin VV et al: ACCF/AHA Expert consensus document on pulmonary hypertension. J Am Coll

Cardiol 53;1573, 2009

Rabinovitch M: Molecular pathogenesis of pulmonary arterial hypertension. J Clin Invest 118:2372, 2008[PMID: 18596905]

Rich S, McLaughlin VV: Pulmonary hypertension, in *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine,* 8th ed, P Libby et al (eds). Philadelphia, Elsevier Saunders, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH DISEASE OF THE RESPIRATORY SYSTEM: INTRODUCTION

The majority of diseases of the respiratory system fall into one of three major categories: (1) obstructive lung diseases; (2) restrictive disorders; and (3) abnormalities of the vasculature. Obstructive lung diseases are most common and primarily include disorders of the airways such as asthma, chronic obstructive pulmonary disease (COPD), bronchiectasis, and bronchiolitis. Diseases resulting in restrictive pathophysiology include parenchymal lung diseases, abnormalities of the chest wall and pleura, as well as neuromuscular disease. Disorders of the pulmonary vasculature are not always recognized and include pulmonary embolism, pulmonary hypertension, and pulmonary venoocclusive disease. Although many specific diseases fall into these major categories, both infective and neoplastic processes can affect the respiratory system and may result in myriad pathologic findings, including obstruction, restriction, and pulmonary vascular disease (see Table 251-1).

Table 251–1. Categories of Respiratory Disease

Category	Examples
Obstructive lung disease	Asthma COPD
	Bronchiectasis
	Bronchiolitis
Restrictive pathophysiology—parenchymal disease	Idiopathic pulmonary fibrosis (IPF) Asbestosis
	Desquamative interstitial pneumonitis (DIP)
	Sarcoidosis
Restrictive pathophysiology—neuromuscular weakness	Amyotrophic lateral sclerosis (ALS) Guillain-Barré syndrome
Restrictive pathophysiology—chest wall/pleural disease	Kyphoscoliosis
	Ankylosing spondylitis
	Chronic pleural effusions
Pulmonary vascular disease	Pulmonary embolism
Malignancy	Pulmonary arterial hypertension (PAH) Bronchogenic carcinoma (non-small-cell and small cell)
	Metastatic disease
Infectious diseases	Pneumonia
	Bronchitis
	Tracheitis

Abbreviation: COPD, chronic obstructive pulmonary disease.

The majority of respiratory diseases present with abnormal gas exchange. Disorders can also be grouped into the categories of gas exchange abnormalities, including hypoxemic, hypercarbic, or

combined impairment. Importantly, many diseases of the lung do not manifest gas exchange abnormalities.

As with the evaluation of most patients, the approach to a patient with disease of the respiratory system begins with a thorough history. A focused physical examination is helpful in further categorizing the specific pathophysiology. Many patients will subsequently undergo pulmonary function testing, chest imaging, blood and sputum analysis, a variety of serologic or microbiologic studies, and diagnostic procedures, such as bronchoscopy. This step-wise approach is discussed in detail below.

HISTORY

Dyspnea and Cough

The cardinal symptoms of respiratory disease are dyspnea and cough (Chaps. 33 and 34). Dyspnea can result from many causes, some of which are not predominantly caused by lung pathology. The words a patient uses to describe breathlessness or shortness of breath can suggest certain etiologies of the dyspnea. Patients with obstructive lung disease often complain of "chest tightness" or "inability to get a deep breath," whereas patients with congestive heart failure more commonly report "air hunger" or a sense of suffocation.

The tempo of onset and duration of a patient's dyspnea are helpful in determining the etiology. Acute shortness of breath is usually associated with sudden physiological changes, such as laryngeal edema, bronchospasm, myocardial infarction, pulmonary embolism, or pneumothorax. Patients with underlying lung disease commonly have progressive shortness of breath or episodic dyspnea. Patients with COPD and idiopathic pulmonary fibrosis (IPF) have a gradual progression of dyspnea on exertion, punctuated by acute exacerbations of shortness of breath. In contrast, most asthmatics have normal breathing the majority of the time and have recurrent episodes of dyspnea usually associated with specific triggers, such as an upper respiratory tract infection or exposure to allergens.

Specific questioning should focus on factors that incite the dyspnea, as well as any intervention that helps resolve the patient's shortness of breath. Of the obstructive lung diseases, asthma is most likely to have specific triggers related to sudden onset of dyspnea, although this can also be true of COPD. Many patients with lung disease report dyspnea on exertion. It is useful to determine the degree of activity that results in shortness of breath as it gives the clinician a gauge of the patient's degree of disability. Many patients adapt their level of activity to accommodate progressive limitation. For this reason it is important, particularly in older patients, to delineate the activities in which they engage and how they have changed over time. Dyspnea on exertion is often an early symptom of underlying lung or heart disease and warrants a thorough evaluation.

Cough is the other common presenting symptom that generally indicates disease of the respiratory system. The clinician should inquire about the duration of the cough, whether or not it associated with sputum production, and any specific triggers that induce it. Acute cough productive of phlegm is often a symptom of infection of the respiratory system, including processes affecting the upper airway (e.g., sinusitis, tracheitis) as well as the lower airways (e.g., bronchitis, bronchiectasis) and lung parenchyma (e.g., pneumonia). Both the quantity and quality of the sputum, including whether it is blood-streaked or frankly bloody, should be determined. Hemoptysis warrants an evaluation as delineated in Chap. 34.

Chronic cough (defined as persisting for more than 8 weeks) is commonly associated with obstructive lung diseases, particularly asthma and chronic bronchitis, as well as "nonrespiratory" diseases, such as gastroesophageal reflux (GERD) and postnasal drip. Diffuse parenchymal lung diseases, including idiopathic pulmonary fibrosis, frequently present with a persistent, nonproductive cough. As with dyspnea, all causes of cough are not respiratory in origin, and assessment should consider a broad differential, including cardiac and gastrointestinal diseases as well as psychogenic causes.

Additional Symptoms

Patients with respiratory disease may complain of wheezing, which is suggestive of airways disease, particularly asthma. Hemoptysis, which must be distinguished from epistaxis or hematemesis, can be a symptom of a variety of lung diseases, including infections of the respiratory tract, bronchogenic carcinoma, and pulmonary embolism. Chest pain or discomfort is also often thought to be respiratory in origin. As the lung parenchyma is not innervated with pain fibers, pain in the chest from respiratory disorders usually results from either diseases of the parietal pleura (e.g., pneumothorax) or pulmonary vascular diseases (e.g., pulmonary hypertension). As many diseases of the lung can result in strain on the right side of the heart, patients may also present with symptoms of cor pulmonale, including abdominal bloating or distention, and pedal edema (Chap. 234).

Additional History

A thorough social history is an essential component of the evaluation of patients with respiratory disease. All patients should be asked about current or previous cigarette smoking as this exposure is associated with many diseases of the respiratory system, most notably COPD and bronchogenic lung cancer but also a variety of diffuse parenchymal lung diseases [e.g., desquamative interstitial pneumonitis (DIP) and pulmonary Langerhans cell histiocytosis]. For most disorders, the duration and

intensity of exposure to cigarette smoke increases the risk of disease. There is growing evidence that "second-hand smoke" is also a risk factor for respiratory tract pathology; for this reason, patients should be asked about parents, spouses, or housemates who smoke. It is becoming less common for patients to be exposed to cigarette smoke on the job, but for older patients, an occupational history should include the potential for heavy cigarette smoke exposure (e.g., flight attendants working prior to prohibition of smoking on airplanes).

Possible inhalational exposures should be explored, including those at the work place (e.g., asbestos, wood smoke) and those associated with leisure (e.g., pigeon excrement from pet birds, paint fumes) (Chap. 256). Travel predisposes to certain infections of the respiratory tract, most notably the risk of tuberculosis. Potential exposure to fungi found in specific geographic regions or climates (e.g., *Histoplasma capsulatum*) should be explored.

Associated symptoms of fever and chills should raise the suspicion of infective etiologies, both pulmonary and systemic. Some systemic diseases, commonly rheumatologic or autoimmune, present with respiratory tract manifestations. Review of systems should include evaluation for symptoms that suggest undiagnosed rheumatologic disease. These may include joint pain or swelling, rashes, dry eyes, dry mouth, or constitutional symptoms. Additionally, carcinomas from a variety of primary sources commonly metastasize to the lung and cause respiratory symptoms. Finally, therapy for other conditions, including both radiation and medications, can result in diseases of the chest.

Physical Examination

The clinician's suspicion for respiratory disease often begins with a patient's vital signs. The respiratory rate is often informative, whether elevated (tachypnea) or depressed (hypopnea). In addition, pulse oximetry should be measured as many patients with respiratory disease will have hypoxemia, either at rest or with exertion.

Simple observation of the patient is informative. Patients with respiratory disease may be in distress, often using accessory muscles of respiration to breathe. Severe kyphoscoliosis can result in restrictive pathophysiology. Inability to complete a sentence in conversation is generally a sign of severe impairment and should result in an expedited evaluation of the patient.

Auscultation

The majority of the manifestations of respiratory disease present with abnormalities of the chest examination. Wheezes suggest airway obstruction and are most commonly a manifestation of asthma. Peribronchial edema in the setting of congestive heart failure, often referred to as "cardiac asthma," can also result in diffuse wheezes as can any other process that causes narrowing of small airways. For this reason, clinicians must take care not to attribute all wheezing to asthma.

Rhonchi are a manifestation of obstruction of medium-sized airways, most often with secretions. In the acute setting, this may be a sign of viral or bacterial bronchitis. Chronic rhonchi suggest bronchiectasis or COPD. Bronchiectasis, or permanent dilation and irregularity of the bronchi, often causes what is referred to as a "musical chest" with a combination of rhonchi, pops, and squeaks. Stridor or a low-pitched, focal inspiratory wheeze usually heard over the neck, is a manifestation of upper airway obstruction and should result in an expedited evaluation of the patient as it can precede complete upper airway obstruction and respiratory failure.

Crackles, or rales, are commonly a sign of alveolar disease. A variety of processes that fill the alveoli with fluid result in crackles. Pneumonia, or infection of the lower respiratory tract and air spaces, may cause crackles. Pulmonary edema, of cardiogenic or noncardiogenic cause, is associated with crackles, generally more prominent at the bases. Interestingly, diseases that result in fibrosis of the interstitium (e.g., IPF) also result in crackles often sounding like Velcro being ripped apart. Although some clinicians make a distinction between "wet" and "dry" crackles, this has not been shown to be a reliable way to differentiate among etiologies of respiratory disease.

One way to help distinguish between crackles associated with alveolar fluid and those associated with interstitial fibrosis is to assess for egophony. Egophony is the auscultation of the sound "AH" instead of "EEE" when a patient phonates "EEE." This change in note is due to abnormal sound transmission through consolidated lung and will be present in pneumonia but not in IPF. Similarly, areas of alveolar filling have increased whispered pectoriloquy as well as transmission of larger airway sounds (i.e., bronchial breath sounds in a lung zone where vesicular breath sounds are expected).

The lack of breath sounds or diminished breath sounds can also help determine the etiology of respiratory disease. Patients with emphysema often have a quiet chest with diffusely decreased breath sounds. A pneumothorax or pleural effusion may present with an area of absent breath sounds, although this is not always the case.

Remainder of Chest Examination

In addition to auscultation, percussion of the chest helps distinguish among pathologic processes of the respiratory system. Diseases of the pleural space are often suggested by differences in percussion note. An area of dullness may suggest a pleural effusion, whereas hyperresonance, particularly at the

apex, can indicate air in the pleural space (i.e., pneumothorax).

Tactile fremitus will be increased in areas of lung consolidation, such as pneumonia, and decreased with pleural effusion. Decreased diaphragmatic excursion can suggest neuromuscular weakness manifesting as respiratory disease or hyperinflation associated with COPD.

Careful attention should also be paid to the cardiac examination with particular emphasis on signs of right heart failure as it is associated with chronic hypoxemic lung disease and pulmonary vascular disease. The clinician should feel for a right ventricular heave and listen for a prominent P2 component of the second heart sound, as well as a right-sided S4.

Other Systems

Pedal edema, if symmetric, may suggest cor pulmonale, and if asymmetric may be due to deep venous thrombosis and associated pulmonary embolism. Jugular venous distention may also be a sign of volume overload associated with right heart failure. Pulsus paradoxus is an ominous sign in a patient with obstructive lung disease as it is associated with significant negative intrathoracic (pleural) pressures required for ventilation, and impending respiratory failure.

As stated earlier, rheumatologic disease may manifest primarily as lung disease. Owing to this association, particular attention should be paid to joint and skin examination. Clubbing can be found in many lung diseases, including cystic fibrosis, IPF, and lung cancer, although it can also be associated with inflammatory bowel disease or as a congenital finding of no clinical importance. Patients with COPD do not usually have clubbing; thus, this sign should warrant an investigation for second process, most commonly an unrecognized bronchogenic carcinoma, in these patients. Cyanosis is seen in hypoxemic respiratory disorders that result in more than 5 g/dL deoxygenated hemoglobin.

DIAGNOSTIC EVALUATION

The sequence of studies is dictated by the clinician's differential diagnosis determined by the history and physical examination. Acute respiratory symptoms are often evaluated with multiple tests obtained at the same time in order to diagnose any life threatening diseases rapidly (e.g., pulmonary embolism or multilobar pneumonia). In contrast, chronic dyspnea and cough can be evaluated in a more protracted, step-wise fashion.

Pulmonary Function Testing

(See also Chap. 253) The initial pulmonary function test obtained is spirometry. This study is used to assess for obstructive pathophysiology as seen in asthma, COPD, and bronchiectasis. A diminished forced expiratory volume in 1 second (FEV_1)/forced vital capacity (FVC) (often defined as less than 70% of predicted value) is diagnostic of obstruction. History as well as further testing can help distinguish among different obstructive diseases. COPD is almost exclusively seen in cigarette smokers. Asthmatics often show an acute response to inhaled bronchodilators (e.g., albuterol). In addition to the measurements of FEV_1 and FVC, the clinician should examine the flow-volume loop. A plateau of the inspiratory or expiratory curves suggests large airway obstruction in extrathoracic and intrathoracic locations, respectively.

Normal spirometry or spirometry with symmetric decreases in FEV_1 and FVC warrants further testing, including lung volume measurement and the diffusion capacity of the lung for carbon monoxide (D_LCO). A total lung capacity (TLC) less than 80% of the predicted value for a patient's age, race, gender, and height defines restrictive pathophysiology. Restriction can result from parenchymal disease, neuromuscular weakness, or chest wall or pleural diseases. Restriction with impaired gas exchange, as indicated by a decreased D_LCO, suggests parenchymal lung disease. Additional testing, such as maximal expiratory pressure (MEP) and maximal inspiratory pressure (MIP), can help diagnose neuromuscular weakness. Normal spirometry, normal lung volumes, and a low D_LCO should prompt further evaluation for pulmonary vascular disease.

Arterial blood gas testing is often also helpful in assessing respiratory disease. Hypoxemia, while usually apparent with pulse oximetry, can be further evaluated with the measurement of arterial PO_2 and the calculation of an alveolar gas and arterial blood oxygen tension difference [$(A\text{-}a)DO_2$]. It should also be noted that at times, most often due to abnormal hemoglobins or non-oxygen hemoglobin-ligand complexes, pulse oximetry can be misleading (such as observed with carboxyhemoglobin). Diseases that cause ventilation-perfusion mismatch or shunt physiology will have an increased $(A\text{-}a)DO_2$ at rest. Arterial blood gas testing also allows for the measurement of arterial PCO_2. Most commonly, acute or chronic obstructive lung disease presents with hypercarbia; however, many diseases of the respiratory system can cause hypercarbia if the resulting increase in work of breathing is greater than that which allows a patient to sustain an adequate minute ventilation.

Chest Imaging

(See Chap. e34) Most patients with disease of the respiratory system will undergo imaging of the chest as part of initial evaluation. Clinicians should generally begin with a plain chest radiograph,

preferably posterior-anterior (PA) and lateral films. Several findings, including opacities of the parenchyma, blunting of the costophrenic angles, mass lesions, and volume loss, can be very helpful in determining an etiology. It should be noted that many diseases of the respiratory system, particularly those of the airways and pulmonary vasculature, are associated with a normal chest radiograph.

Subsequent computed tomography of the chest (CT scan) is often obtained. The CT scan allows better delineation of parenchymal processes, pleural disease, masses or nodules, and large airways. If administered with contrast, the pulmonary vasculature can be assessed with particular utility for determination of pulmonary emboli. Intravenous contrast also allows lymph nodes to be delineated in greater detail.

FURTHER STUDIES

Depending on the clinician's suspicion, a variety of other studies may be obtained. Concern for large airway lesions may warrant bronchoscopy. This procedure may also be used to sample the alveolar space with bronchoalveolar lavage (BAL) or to obtain nonsurgical lung biopsies. Blood testing may include assessment for hypercoagulable states in the setting of pulmonary vascular disease, serologic testing for infectious or rheumatologic disease, or assessment of inflammatory markers or leukocyte counts (e.g., eosinophils). Sputum evaluation for malignant cells or microorganisms may be appropriate. An echocardiogram to assess right- and left-sided heart function is often obtained. Finally, at times, a surgical lung biopsy is needed to diagnose certain diseases of the respiratory system. All of these studies will be guided by the preceding history, physical examination, pulmonary function testing, and chest imaging.

FURTHER READINGS

Irwin RS et al: Diagnosis and management of cough executive summary: ACCP evidence-based clinical practice guidelines. Chest 129:1S, 2006

Manning HL, Schwartzstein RM: Pathophysiology of dyspnea. N Engl J Med 333:1547, 1995[PMID: 7477171] [Full Text]

Schwartzstein RM, Parker MJ: *Respiratory Physiology: A Clinical Approach* . Philadelphia, Lippincott, 2006

Weinberger SE et al: *Principles of Pulmonary Medicine* , 5th ed. Philadelphia, Saunders, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

The primary function of the respiratory system is to oxygenate blood and eliminate carbon dioxide, which requires that blood come into virtual contact with fresh air to facilitate diffusion of respiratory gases between blood and gas. This process occurs in the lung alveoli, where blood flowing through alveolar wall capillaries is separated from alveolar gas by an extremely thin membrane of flattened endothelial and epithelial cells, across which respiratory gases diffuse and equilibrate. Blood flow through the lung is unidirectional via a continuous vascular path, along which venous blood absorbs oxygen from and loses CO_2 to inspired gas. The path for airflow, in contrast, reaches a dead end at the alveolar walls; as such, the alveolar space must be ventilated tidally, with inflow of fresh gas and outflow of alveolar gas alternating periodically at the respiratory rate (RR).

To achieve an enormous alveolar surface area (typically 70 m^2) for blood-gas diffusion within the modest volume of a thoracic cavity (typically 7 L), nature has distributed both blood flow and ventilation among millions of tiny alveoli through multigenerational branching of both pulmonary arteries and bronchial airways. As a consequence of variations in tube lengths and calibers along these pathways, and of the effects of gravity, tidal pressure fluctuations, and anatomic constraints from the chest wall, there is variation among alveoli in their relative ventilations and perfusions. Not surprisingly, for the lung to be most efficient in exchanging gas, the fresh gas ventilation of a given alveolus must be matched to its perfusion.

For the respiratory system to succeed in oxygenating blood and eliminating carbon dioxide, it must be able to ventilate the lung tidally to freshen alveolar gas; it must provide for perfusion of the individual alveolus in a manner proportional to its ventilation; and it must allow for adequate diffusion of respiratory gases between alveolar gas and capillary blood. Furthermore, it must be able to accommodate severalfold increases in the demand for oxygen uptake or CO_2 elimination imposed by metabolic needs or acid-base derangement. Given these multiple requirements for normal operation, it is not surprising that many diseases disturb respiratory function. Here, we consider in greater detail the physiologic determinants of lung ventilation and perfusion, and how their matching distributions and rapid gas diffusion allow for normal gas exchange. We also discuss how common diseases derange these normal functions, and thereby impair gas exchange—or at least raise the work of the respiratory muscles or heart to maintain adequate respiratory function.

VENTILATION

It is useful to think about the respiratory system as having three independently functioning components—the lung including its airways, the neuromuscular system, and the chest wall; the latter includes everything that is not lung or active neuromuscular system. As such, the mass of the respiratory muscles is part of the chest wall, while the force they generate is part of the neuromuscular system; the abdomen (especially an obese abdomen) and the heart (especially an enlarged heart) are, for these purposes, part of the chest wall. Each of these three components has mechanical properties that relate to its enclosed volume, or in the case of the neuromuscular system, the respiratory system volume at which it is operating, and to the rate of change of its volume (i.e., flow).

Volume-Related Mechanical Properties—Statics

Figure 252-1 shows the volume-related properties of each component of the respiratory system. Due both to surface tension at the air-liquid interface between alveolar wall lining fluid and alveolar gas and to elastic recoil of the lung tissue itself, the lung requires a positive transmural pressure difference between alveolar gas and its pleural surface to stay inflated; this difference is called the elastic recoil pressure of the lung, and it increases with lung volume. Importantly, the lung becomes rather stiff at high lung volumes, so that relatively small volume changes are accompanied by large changes in transpulmonary pressure; in contrast, the lung is compliant at lower lung volumes, including those at which tidal breathing normally occurs. Note that at zero inflation pressure, even normal lungs retain some air in the alveoli. This occurs because the small peripheral airways of the lung are tethered open by radially outward pull from inflated lung parenchyma attached to adventitia; as the lung deflates during exhalation, those small airways are pulled open progressively less, and eventually they close, trapping some gas in the alveoli. This effect can be exaggerated with age and especially with obstructive airways diseases, resulting in gas trapping at quite large lung volumes.

Figure 252-1

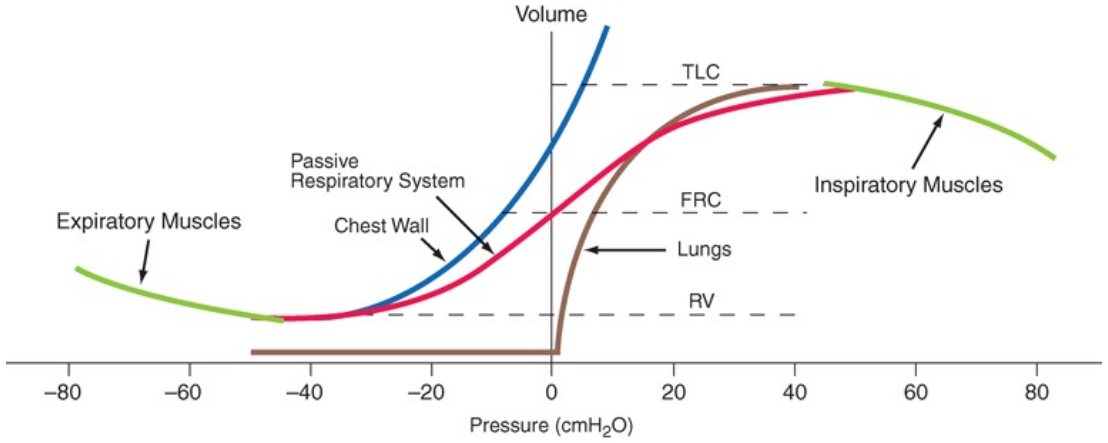

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pressure-volume curves of the isolated lung, isolated chest wall, combined respiratory system, inspiratory muscles, and expiratory muscles. FRC, functional residual capacity; RV, residual volume; TLC, total lung capacity.

The elastic behavior of the passive chest wall (i.e., in the absence of neuromuscular activation) differs markedly from that of the lung. Whereas the lung tends toward full deflation with no distending (transmural) pressure, the chest wall encloses a large volume when pleural pressure equals body surface (atmospheric) pressure. Furthermore, the chest wall is compliant at high enclosed volumes, readily expanding even further in response to increases in transmural pressure. The chest wall also remains compliant at small negative transmural pressures (i.e., when pleural pressure falls slightly below atmospheric pressure), but as the volume enclosed by the chest wall becomes quite small in response to large negative transmural pressures, the passive chest wall becomes stiff due to squeezing together of ribs and intercostal muscles, diaphragm stretch, displacement of abdominal contents, and straining of ligaments and bony articulations. Under normal circumstances, the lung and the passive chest wall enclose essentially the same volume, the only difference between these being the volumes of the pleural fluid and of the lung parenchyma (both quite small). As such, and because the lung and chest wall function in mechanical series, the pressure required to displace the passive respiratory system (lungs + chest wall) at any volume is simply the sum of the elastic recoil pressure of the lungs and the transmural pressure across the chest wall. When plotted against respiratory system volume, this relationship assumes a sigmoid shape, exhibiting stiffness at high lung volumes (imparted by the lung), stiffness at low lung volumes (imparted by the chest wall, or sometimes by airway closure), and compliance in the middle range of lung volumes. There is also a passive resting point of the respiratory system, attained when alveolar gas pressure equals body surface pressure (i.e., the transrespiratory system pressure is zero). At this volume [called *functional residual capacity* (FRC)], the outward recoil of the chest wall is balanced exactly by the inward recoil of the lung. As these recoils are transmitted through the pleural fluid, the latter is pulled both outward and inward simultaneously at FRC, and, thus, its pressure falls below atmospheric pressure (typically, –5 cmH_2O).

The normal passive respiratory system would equilibrate at FRC and remain there were it not for the actions of respiratory muscles. The inspiratory muscles act on the chest wall to generate the equivalent of positive pressure across the lungs and passive chest wall, while the expiratory muscles generate the equivalent of negative transrespiratory pressure. The maximal pressures these sets of muscles can generate varies with the lung volume at which they operate, due to length-tension relationships in striated muscle sarcomeres and to changes in mechanical advantage as the angles of insertion change with lung volume (Fig. 252-1). Nonetheless, under normal conditions the respiratory muscles are substantially "overpowered" for their roles, and generate more than adequate force to drive the respiratory system to its stiffness extremes, as determined by the lung [total lung capacity (TLC)] or chest wall or airway closure [residual volume (RV)]; importantly, the latter always prevents the adult lung from emptying completely under normal circumstances. The excursion between full and minimal lung inflation is called vital capacity (VC; **Fig. 252-2**), and is readily seen to be the difference between volumes at two unrelated stiffness extremes—one determined by the lung (TLC) and the other determined by the chest wall or airways (RV). Thus, although VC is easy to measure (see below), it tells one little about the intrinsic properties of the respiratory system. It is much more useful, as we shall see, for the clinician to know TLC and RV individually.

Figure 252-2

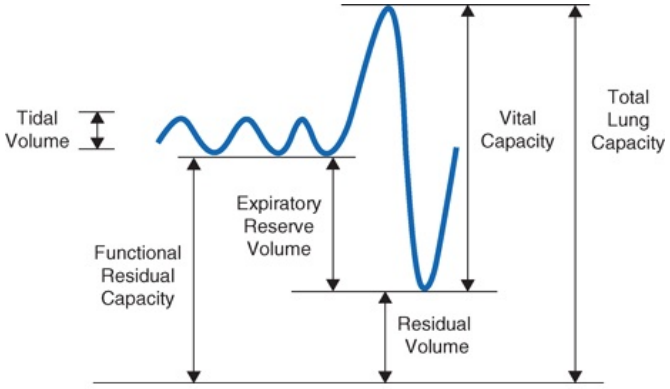

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spirogram demonstrating a slow vital capacity maneuver and various lung volumes.

Flow-Related Mechanical Properties—Dynamics

The passive chest wall and active neuromuscular system do exhibit mechanical behaviors related to the rate of change of volume, but these become quantitatively important only at markedly supraphysiologic breathing frequencies (e.g., during high-frequency mechanical ventilation), and, thus, we shall not address these here. In contrast, the dynamic airflow properties of the lung substantially determine the ability to ventilate and contribute importantly to the work of breathing, and are often deranged by disease. Understanding these properties is, therefore, well worthwhile.

As with flow of any fluid (gas or liquid) in any tube, maintenance of airflow within the pulmonary airways requires a pressure gradient that falls along the direction of flow, the magnitude of which is determined by the flow rate and the frictional resistance to flow. During quiet tidal breathing, the pressure gradients driving inspiratory or expiratory flow are small owing to the very low frictional resistance of normal pulmonary airways (normally <2 cmH_2O/L per second). However, during rapid exhalation another phenomenon reduces flow below that which would have been expected were frictional resistance the only impediment to flow. This phenomenon is called dynamic airflow limitation, and it occurs because the bronchial airways through which air is exhaled are collapsible rather than rigid (**Fig. 252-3**). An important anatomic feature of the pulmonary airways is its treelike branching structure. While the individual airways in each successive generation, from most proximal (trachea) to most distal (respiratory bronchioles), are smaller than those of the parent generation, their number increases exponentially such that the summed cross-sectional area of the airways becomes very large toward the lung periphery. Because flow (volume/time) is constant along the airway tree, the velocity of airflow (flow/summed cross-sectional area) is much greater in the central airways than in the peripheral airways. During exhalation, gas leaving the alveoli must therefore gain velocity as it proceeds toward the mouth. The energy required for this "convective" acceleration is drawn from the component of gas energy manifested as its local pressure, thereby reducing intraluminal gas pressure (the Bernoulli effect), reducing airway transmural pressure, reducing airway size (Fig. 252-3), and reducing flow. If one tries to exhale more forcefully, the local velocity increases further and reduces airway size further, resulting in no net increase in flow. Under these circumstances, flow has reached its maximum possible value, or its flow limit. Lungs normally exhibit such dynamic airflow limitation. The maximum value of flow is related to the gas density, airway cross-section and distensibility, the elastic recoil pressure of the lung, and the frictional pressure loss to the flow-limiting airway site. Under normal conditions, maximal expiratory flow falls with lung volume (**Fig. 252-4**), due primarily to the dependence of lung recoil pressure on lung volume (Fig. 252-1). In pulmonary fibrosis, lung recoil pressure is increased at any lung volume, and, thus, the maximum expiratory flow is relatively elevated when considered in relation to lung volume. Conversely, in emphysema, lung recoil pressure is reduced, which is a principal mechanism by which maximal expiratory flows fall. Diseases that narrow the airway lumen at any transmural pressure, such as asthma or chronic bronchitis, or which cause excessive airway collapsibility, like tracheomalacia, also reduce maximal expiratory flow.

Figure 252-3

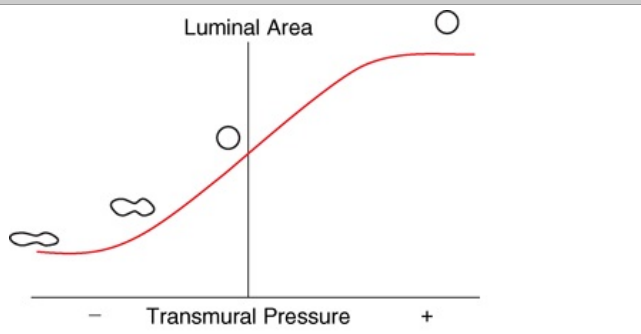

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Luminal area versus transmural pressure relationship. Transmural pressure represents the pressure difference across the airway wall from inside to outside.

Figure 252-4

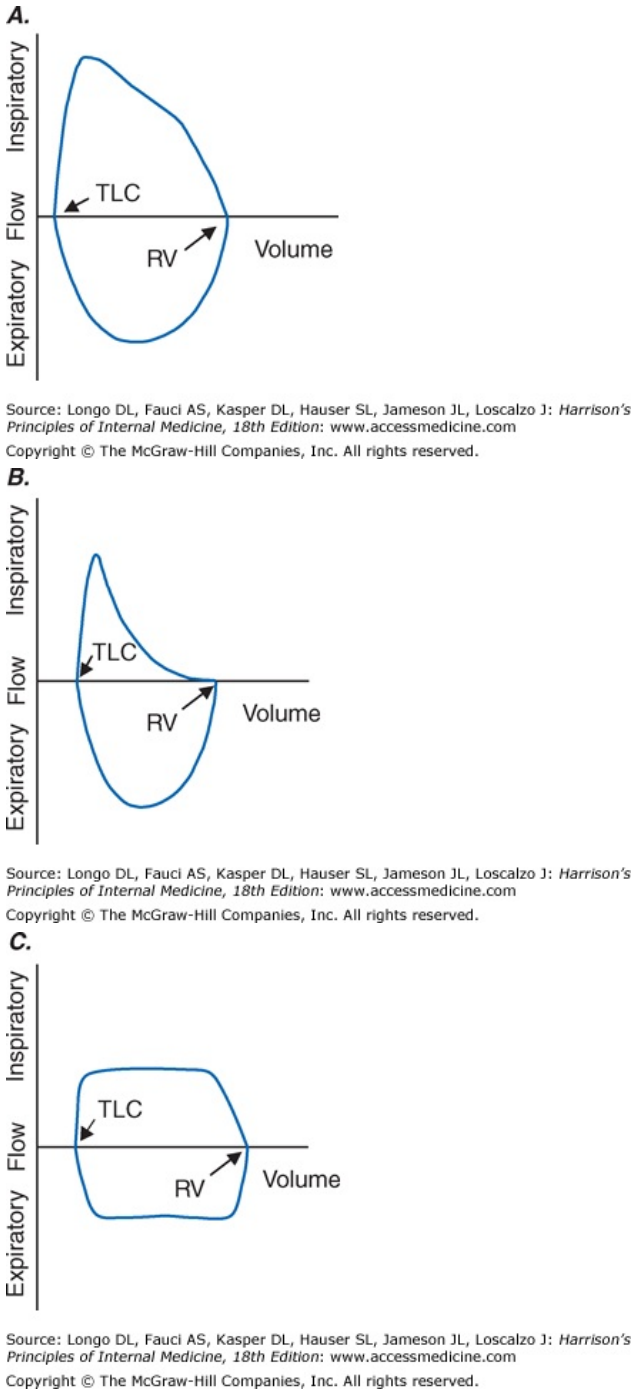

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flow-volume loops. ***A***. Normal. ***B***. Airflow obstruction. ***C***. Fixed central airway obstruction. RV, residual volume; TLC, total lung capacity.

The Bernoulli effect also acts during inspiration, but the more negative pleural pressures during inspiration lower the pressure outside of airways, thereby increasing transmural pressure and promoting airway expansion. Thus, inspiratory airflow limitation seldom occurs due to diffuse pulmonary airway disease. Conversely, extrathoracic airway narrowing (as due to a tracheal adenoma or post-tracheostomy stricture) can lead to inspiratory airflow limitation (Fig. 252-4).

The phenomenon of flow limitation and the importance of airway size and distensibility and of upstream pressure (lung elastic recoil pressure for forced exhalation) can easily be appreciated by sniffing through one's nose with low, medium, or substantial effort. If one keeps the nostrils relaxed, increasing from low to medium inspiratory effort raises inspiratory flow through the nose somewhat, but inhaling even harder will likely not increase inspiratory nasal airflow more but, rather, will just collapse the nares, a manifestation of dynamic airflow limitation. One can increase inspiratory nasal airflow, however, by flaring one's nostrils using the alae

nasi muscles. This increases nostril area (reducing velocity for a given flow through the nares) and stiffens the nostril walls (reducing their narrowing in response to negative transmural pressure). Springlike nasal strips sometimes used by football players have the same effect. In patients with obstructive sleep apnea (OSA), a narrowed and excessively compliant pharyngeal airway also collapses in response to negative transmural pressure generated by the Bernoulli effect and by inspiratory frictional pressure loss in the nose (which is why an upper respiratory infection often worsens OSA). Increasing the upstream driving pressure from which these phenomena lower intrapharyngeal gas pressure with positive nasal airway pressure keeps pharyngeal transmural pressure positive, preventing inspiratory airflow limitation. Inspiratory airflow limitation in the nose or in the pharynx of patients with OSA closely parallels expiratory flow limitation in the lung.

Work of Breathing

In health, the elastic (volume change–related) and dynamic (flow-related) loads that must be overcome to ventilate the lungs at rest are small, and the work required of the respiratory muscles is minimal. However, the work of breathing can increase considerably, due either to requirement for substantially increased ventilation, an abnormally increased mechanical load, or both. As discussed below, the rate of ventilation is primarily set by the need to eliminate carbon dioxide, and, thus, ventilation increases during exercise (sometimes more than 20-fold) and during metabolic acidosis as a compensatory response. Naturally, the work rate required to overcome the elasticity of the respiratory system increases with both the depth and frequency of tidal breaths, while the work required to overcome the dynamic load increases with total ventilation. A modest increase of ventilation is most efficiently achieved by increasing tidal volume but not respiratory rate, which is the normal ventilatory response to lower level exercise. At high levels of exercise, deep breathing persists, but respiratory rate also increases. The pattern chosen by the respiratory controller minimizes the work of breathing.

Work of breathing also increases when disease reduces the compliance of the respiratory system or increases the resistance to airflow. The former occurs commonly in diseases of the lung parenchyma (interstitial processes or fibrosis, alveolar filling diseases such as pulmonary edema or pneumonia, or substantial lung resection), and the latter occurs in obstructive airways diseases such as asthma, chronic bronchitis, emphysema, and cystic fibrosis. Furthermore, severe airflow obstruction can functionally reduce the compliance of the respiratory system by leading to dynamic hyperinflation. In this scenario, expiratory flows slowed by the obstructive airways disease may be insufficient to allow for complete exhalation during the expiratory phase of tidal breathing; as a result, the "functional residual capacity" from which the next breath is inhaled is greater than the static FRC. With repetition of incomplete exhalations of each tidal breath, the operating FRC becomes dynamically elevated, sometimes to a level that approaches TLC. At these high lung volumes, the respiratory system is much less compliant than at normal breathing volumes, and, thus, the elastic work of each tidal breath is also increased. The dynamic pulmonary hyperinflation that accompanies severe airflow obstruction causes patients to sense difficulty in breathing in—even though the pathophysiologic abnormality at root cause is expiratory airflow obstruction.

Adequacy of Ventilation

As noted above, the respiratory control system that sets the rate of ventilation responds to chemical signals, including arterial carbon dioxide and oxygen tensions and blood pH, and to volitional needs, such as the need to inhale deeply before playing a long phrase on the trumpet. Disturbances in ventilation are discussed in Chap. 264. Here, we focus on the relationship between ventilation of the lung and carbon dioxide elimination.

At the end of each tidal exhalation, the conducting airways are filled with alveolar gas that had not reached the mouth when expiratory flow stopped. During the ensuing inhalation, fresh gas immediately enters the airway tree at the mouth, but the gas first entering the alveoli at the start of inhalation is that same alveolar gas in the conducting airways that had just left the alveoli. As such, fresh gas does not enter the alveoli until the volume of the conducting airways has been inspired. This volume is called the anatomic dead space. Quiet breathing with tidal volumes smaller than the anatomic dead space introduces no fresh gas into the alveoli at all; only that part of the inspired tidal volume (V_T) that is greater than the dead space (V_D) introduces fresh gas into the alveoli. Importantly, the dead space can be further increased functionally if some of the inspired tidal volume is delivered to a part of the lung that receives no pulmonary blood flow, and, thus, cannot contribute to gas exchange, as can occur in the portion of the lung distal to a large pulmonary embolus. As such, exhaled minute ventilation ($\dot{V}_E = V_T \times RR$) includes a component of dead space ventilation ($\dot{V}_D = V_D \times RR$) and a component of fresh gas alveolar ventilation ($\dot{V}_A = [V_T - V_D] \times RR$). Carbon dioxide elimination from the alveoli is equal to $\dot{V}_A$ times the difference in CO_2 fraction between inspired air (essentially zero) and alveolar gas (typically ~5.6%, after correcting for humidification of inspired air, corresponding to 40 mmHg). In the steady state, the alveolar fraction of CO_2 is equal to the metabolic CO_2 production divided by the alveolar ventilation. Because, as discussed below, alveolar and arterial CO_2 tensions are equal, and because the respiratory controller normally strives to maintain arterial P_{CO2} (Pa_{CO2}) at ~40 mmHg, the adequacy of alveolar ventilation is reflected in Pa_{CO2}. If Pa_{CO2} falls much below 40 mmHg, alveolar hyperventilation is present, and if Pa_{CO2} exceeds 40 mmHg, then alveolar hypoventilation is present. Ventilatory failure is characterized by extreme alveolar hypoventilation.

As a consequence of oxygen uptake of alveolar gas into capillary blood, alveolar oxygen tension falls below that of inspired gas. The rate of oxygen uptake (determined by the body's metabolic oxygen consumption) is related to the average rate of metabolic carbon dioxide production and their ratio, called the "respiratory quotient" ($R = \dot{V}CO_2 / \dot{V}O_2$), depends largely on the fuel being metabolized. For a typical American diet, R is usually around 0.85, and more oxygen is absorbed than CO_2 is excreted. Together, these phenomena allow the estimation of alveolar oxygen tension, according to the following relationship, known as the alveolar gas equation:

$$Pa_{O2} = Fl_{O2} \times (P_{bar} - P_{H2O}) - Pa_{CO2}/R$$

The alveolar gas equation also highlights the influences of inspired oxygen fraction (Fl_{O2}), barometric pressure (P_{bar}), and vapor pressure of water (P_{H2O} = 47 mmHg at 37°C) in addition to alveolar ventilation (which sets PA_{CO2}) in determining P_AO_2. An implication of the alveolar gas equation is that severe arterial hypoxemia rarely occurs as a pure consequence of alveolar hypoventilation at sea level while breathing air. The potential for alveolar hypoventilation to induce severe hypoxemia with otherwise normal lungs increases as P_{bar} falls with increasing altitude.

GAS EXCHANGE

Diffusion

For oxygen to be delivered to the peripheral tissues, it must pass from alveolar gas into alveolar capillary blood by diffusing through alveolar membrane. The aggregate alveolar membrane is highly optimized for this process, with a very large surface area and minimal thickness. Diffusion through the alveolar membrane is so efficient in the human lung that in most circumstances its hemoglobin becomes fully oxygen saturated by the time a red blood cell has traveled just one-third the length of the alveolar capillary. As such, uptake of alveolar oxygen is ordinarily limited by the amount of blood transiting the alveolar capillaries rather than how rapidly oxygen can diffuse across the membrane; thus, oxygen uptake from the lung is said to be "perfusion limited." Carbon dioxide also equilibrates rapidly across the alveolar membrane. Thus, the oxygen and CO_2 tensions in capillary blood leaving a normal alveolus are essentially equal to those in alveolar gas. In only rare circumstances is oxygen uptake from normal lungs diffusion-limited, which can occur at high altitude and/or by high-performance athletes exerting maximum effort. Diffusion limitation can also occur in interstitial lung disease if substantially thickened alveolar walls remain perfused.

Ventilation-Perfusion Heterogeneity

As noted above, for gas exchange to be most efficient, the ventilation to each individual alveolus should be matched to the perfusion to its accompanying capillaries for each of millions of alveoli. Due to the differential effects of gravity on lung mechanics and blood flow throughout the lung, and due to differences of airway and vascular architecture among various respiratory paths, there is minor ventilation/perfusion heterogeneity even in the normal lung; however, $\dot{V}/\dot{Q}$ heterogeneity can be particularly marked in disease. Two extreme examples are (1) ventilation of unperfused lung distal to a pulmonary embolus, in which ventilation of the physiologic dead space is "wasted" in the sense that it does not contribute to gas exchange; and (2) perfusion of nonventilated lung, a condition known as a "shunt." The latter allows venous blood to pass through the lung unaltered; when mixed with fully oxygenated blood leaving other well-ventilated lung units, shunted venous blood disproportionately lowers the mixed arterial Pa_{O2}, due to the nonlinear oxygen content versus the PO_2 relationship of hemoglobin (**Fig. 252-5**). Furthermore, the resulting arterial hypoxemia is refractory to supplemental inspired oxygen. This is because raising inspired Fl_{O2} has no effect on alveolar gas tensions in nonventilated alveoli, and while raising inspired Fl_{O2} does increase PA_{O2} in ventilated alveoli, the oxygen content of blood exiting ventilated units increases only slightly as hemoglobin will already have been nearly fully saturated and the solubility of oxygen in plasma is quite small.

Figure 252-5

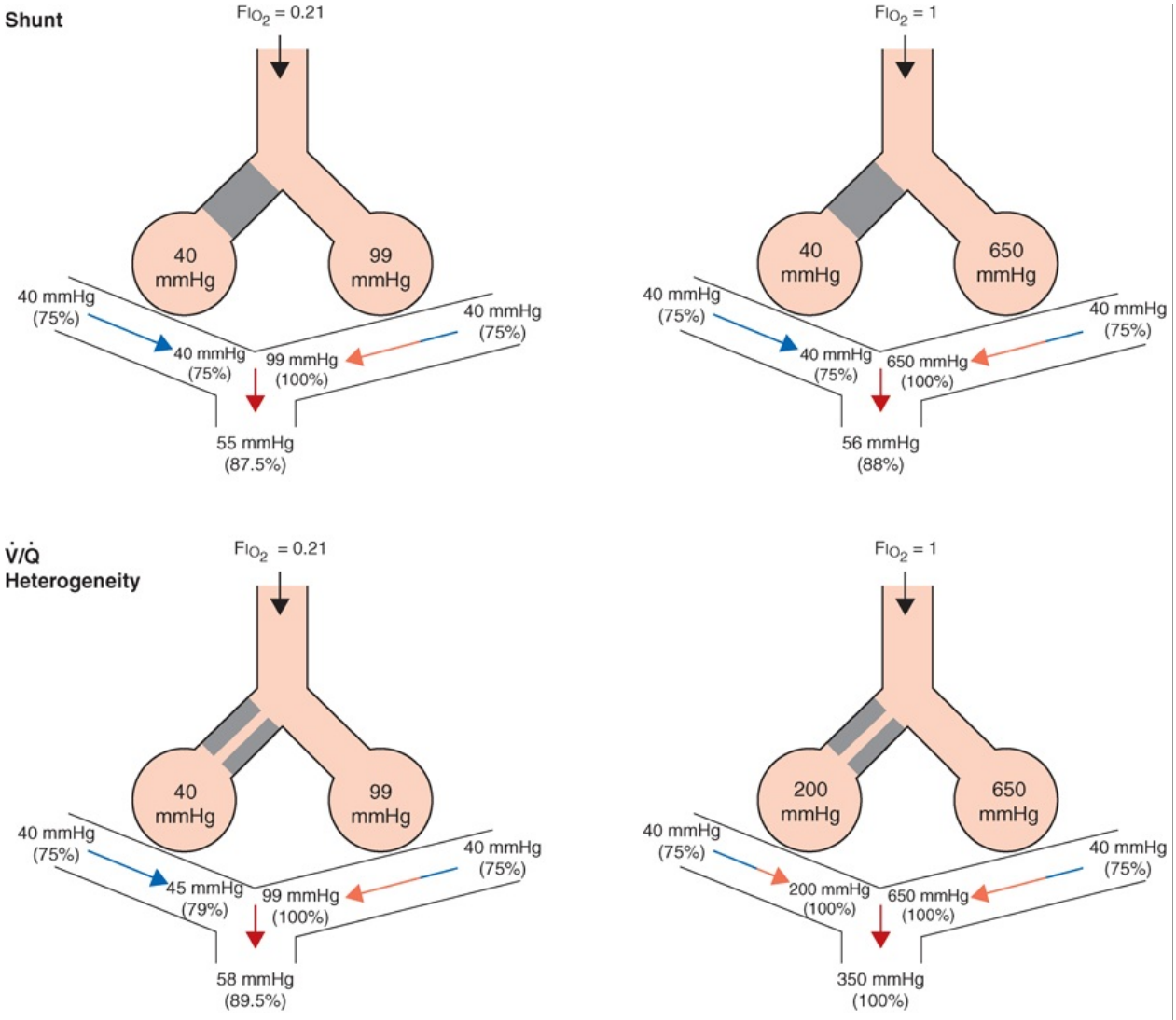

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Influence of air vs oxygen breathing on mixed arterial oxygenation in shunt and ventilation/perfusion heterogeneity. Partial pressure of oxygen (mmHg) and oxygen saturations are shown for mixed venous blood, end capillary blood for normal versus affected alveoli, and for mixed arterial blood.

More commonly occurring than the two extreme examples given above is a widening of the distribution of ventilation/ perfusion ratios; such V̇/ Q̇ heterogeneity is a common consequence of lung disease. In this circumstance, perfusion of relatively underventilated alveoli results in the incomplete oxygenation of exiting blood. When mixed with well-oxygenated blood leaving higher V̇/ Q̇ regions, this partially preoxygenated blood disproportionately lowers arterial Pa_{O2}, although to a lower extent than does a similar perfusion fraction of blood leaving regions of pure shunt. In addition, in contrast to shunt regions, inhalation of supplemental oxygen does raise the PA_{O2} even in relatively underventilated low V̇/ Q̇ regions, and so the arterial hypoxemia induced by V̇/ Q̇ heterogeneity is typically responsive to oxygen therapy (Fig. 252-5).

In sum, arterial hypoxemia can be caused by substantial reduction of inspired oxygen tension, by severe alveolar hypoventilation, or by perfusion of relatively underventilated (low V̇/ Q̇) or completely unventilated (shunt) lung regions, and, in unusual circumstances, by limitation of gas diffusion.

Approach to the Patient: Disturbances of Respiratory Function

There are many diseases that injure the respiratory system, but there are relatively few ways in which it responds to that injury. For this reason, the pattern of physiologic abnormalities may or may not provide sufficient information to discriminate among conditions. The following studies are commonly used to characterize a patient's respiratory function and often lead to a better understanding of the underlying disorder.

MEASUREMENT OF VENTILATORY FUNCTION

Lung Volumes

Figure 252-2 demonstrates a spirometry tracing in which the volume of air entering or exiting the lung is plotted over time. In a slow vital capacity maneuver, the subject inhales from FRC, fully inflating the lungs to TLC, and then the patient exhales slowly to RV; VC is the difference between TLC and RV, and represents the maximum excursion of the respiratory system. Spirometry discloses relative volume changes during these maneuvers, but cannot reveal the absolute volumes at which they occur. To determine absolute lung volumes, two approaches

are commonly used—inert gas dilution and body plethysmography. In the former, a known amount of a nonabsorbable inert gas (usually helium or neon) is inhaled in a single large breath or is rebreathed from a closed circuit; the inert gas is diluted by the gas resident in the lung at the time of inhalation, and its final concentration reveals the volume of resident gas contributing to the dilution. A drawback of this method is that regions of the lung that ventilate poorly (e.g., due to airflow obstruction) may not receive much inspired inert gas and so do not contribute to its dilution. As such, inert gas dilution often underestimates true lung volumes.

In the second approach, FRC is determined by measuring the compressibility of gas within the chest, which is proportional to the volume of gas being compressed. The patient sits in a body plethysmograph, a chamber usually made of transparent plastic to minimize claustrophobia, and at the end of a normal tidal breath (i.e., when lung volume is FRC) is instructed to pant against a closed shutter, thus, periodically compressing air within the lung slightly. Pressure fluctuations at the mouth and volume fluctuations within the body box (equal but opposite to those of the chest) are measured, and from these the thoracic gas volume is calculated using Boyle's law. Once FRC is obtained, TLC and RV are calculated by adding the inspiratory capacity or subtracting expiratory reserve volume, respectively, values determined during spirometry (Fig. 252-2). The most important determinants of healthy individuals' lung volumes are height, age, and gender, but there is considerable additional normal variation beyond that accounted for by these parameters. In addition, race influences lung volumes; on average TLC values are about 12% smaller in African Americans and 6% smaller in Asian Americans when compared with those of white Americans. In practice, a mean "normal" value is predicted by multivariate regression equations using height, age, and gender, and the patient's value is divided by the predicted value (often with "race correction" applied) to determine "percent predicted." For most measures of lung function, 85–115% of the predicted value can be normal, but in health the various lung volumes tend to scale together. For example, if one is "normal big" with TLC 110% of the predicted value, then all other lung volumes and spirometry values will also approximate 110% of their respective predicted values. This pattern is particularly helpful in evaluating airflow, as discussed below.

Air Flow

As noted above, spirometry plays a key role in lung volume determination. But even more often, spirometry is used to measure air flow, which reflects the dynamic properties of the lung. During a forced vital capacity maneuver, the patient inhales to TLC and then exhales rapidly and forcefully to RV; this ensures that flow limitation has been achieved, so that the precise effort made has little influence on actual flow. The total amount of air exhaled is the forced vital capacity (FVC) and the amount of air exhaled in the first second is the forced expiratory volume in one second (FEV_1); note that FEV_1 is a flow rate, as it reveals volume change per time. As with lung volumes, an individual's maximal expiratory flows should be compared to predicted values based on height, age, and gender. While the FEV_1/FVC ratio is typically reduced in airflow obstruction, airflow obstruction can also reduce FVC by raising RV. If this occurs, the FEV_1/FVC ratio may be "artifactually normal," erroneously suggesting that airflow obstruction is absent. To circumvent this problem, it is useful to compare FEV_1 as a fraction of its predicted value with TLC as a fraction of its predicted value. In health, these are usually similar. In contrast, even an FEV_1 value that is 95% of its predicted value may actually be relatively low if TLC is 110% of its respective prediction. In this case, airflow obstruction might be present, despite the "normal" value for FEV_1.

The relationships among volume, flow, and time during spirometry are best displayed in two plots—the spirogram (volume vs. time) and the flow-volume loop (flow vs. volume) (Fig. 252-4). In conditions that cause airflow obstruction, the site of obstruction can sometimes be correlated with the shape of the flow-volume loop. In diseases that cause lower airway obstruction such as asthma or emphysema, flows decrease more rapidly with declining lung volumes leading to a characteristic scooping of the flow-volume loop. In contrast, fixed upper airway obstruction typically leads to inspiratory and/or expiratory flow plateaus (Fig. 252-4).

Airways Resistance

The total resistance of the pulmonary and upper airways is measured in the same body plethysmography used to measure FRC. The patient is asked once again to pant, but this time against a closed and then opened shutter. Panting against the closed shutter reveals the thoracic gas volume as above. When the shutter is opened, flow is now directed to and from the body box, so that volume fluctuations in the box reveal the extent of thoracic gas compression, which in turn reveals the pressure fluctuations driving flow. Flow is measured simultaneously, allowing the calculation of lung resistance (as flow divided by pressure). In health, airways resistance is very small, <2 cmH_2O/L per second, and half of this resides within the upper airway. Of the lung's contribution, most of the resistance originates in the central airways. For this reason, airways resistance measurement tends to be insensitive to peripheral airflow obstruction.

Respiratory Muscle Strength

To measure respiratory muscle strength, the patient is instructed to exhale or inhale with maximum effort against a closed shutter while pressure is monitored at the mouth. Pressures greater than ±60 cmH_2O at FRC are considered adequate, making unlikely the possibility that respiratory muscle weakness accounts for any other ventilatory dysfunction that might be identified.

MEASUREMENT OF GAS EXCHANGE

Diffusing Capacity

This test uses a small (and safe) amount of carbon monoxide to measure gas exchange across the alveolar membrane during a 10-second breath hold. Carbon monoxide in exhaled breath is analyzed to determine the quantity of CO absorbed by crossing the alveolar membrane and combining with hemoglobin in red blood cells. This "single-breath diffusing capacity" [diffusion capacity of the lung for carbon monoxide (DL_{CO})] value

increases with the surface area available for diffusion and the amount of hemoglobin within the capillaries, and varies inversely with alveolar membrane thickness. Thus, DL_{CO} decreases in diseases that thicken or destroy alveolar membranes (e.g., pulmonary fibrosis, emphysema), curtail the pulmonary vasculature (e.g., pulmonary hypertension), or reduce alveolar capillary hemoglobin (e.g., anemia). Single-breath diffusing capacity may be elevated in acute congestive heart failure, asthma, polycythemia, and pulmonary hemorrhage.

Arterial Blood Gases

The effectiveness of gas exchange can be assessed by measuring the partial pressures of oxygen and carbon dioxide in a sample of blood obtained by arterial puncture. The oxygen content of blood (Ca_{O2}) depends upon arterial saturation ($\%O_2Sat$), which is set by Pa_{O2}, pH, and Pa_{CO2} according to the oxyhemoglobin dissociation curve; Ca_{O2} can also be measured by oximetry (see below):

$$Ca_{O2}\ (mL/dL) = 1.34\ (mL/dL/g) \times [hemoglobin](g) \times \%O_2Sat + 0.003\ (mL/dL/mmHg) \times Pa_{O2}\ (mmHg)$$

Pulse Oximetry

Continuous monitoring of arterial blood gases requires either repeated arterial punctures or an indwelling arterial catheter, and so may be difficult in many circumstances. Instead, the oxygen saturation fraction of hemoglobin can be measured continuously using pulse oximetry, a tool that measures the absorbance by hemoglobin of several wavelengths of light transmitted across a finger, toe, or ear by a noninvasive probe. However, since oxygen content varies relatively little with Pa_{O2} at saturations above 90%, it is difficult to know the precise Pa_{O2} using this device. In addition, as noted above, Pa_{CO2} is needed to fully assess the mechanism of hypoxemia, a value that is not revealed by pulse oximetry.

CLINICAL CORRELATIONS: TYPICAL EXAMPLES

This chapter has highlighted the physiologic processes underlying respiratory system function and the techniques used by clinicians to assess them. Figure 252-6 lists abnormalities in pulmonary function testing typically found in a number of common respiratory disorders and highlights the simultaneous occurrence of multiple physiologic abnormalities. Importantly, some of these respiratory disorders can coexist, which results in more complex superposition of these abnormalities.

Figure 252-6

	Restriction due to increased lung elastic recoil (pulmonary fibrosis)	Restriction due to chest wall abnormality (moderate obesity)	Restriction due to respiratory muscle weakness (myasthenia gravis)	Obstruction due to airway narrowing (acute asthma)	Obstruction due to decreased elastic recoil (severe emphysema)
TLC	60%	95%	75%	100%	130%
FRC	60%	65%	100%	104%	220%
RV	60%	100%	120%	120%	310%
FVC	60%	92%	60%	90%	60%
FEV_1	75%	92%	60%	35% pre-b.d. 75% post-b.d.	35% pre-b.d. 38% post-b.d.
R_{aw}	1.0	1.0	1.0	2.5	1.5
DL_{CO}	60%	95%	80%	120%	40%

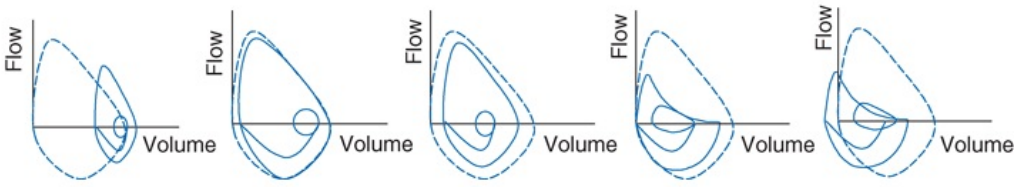

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Commonly seen abnormalities of pulmonary function (see text). Pulmonary function values are expressed as percent of normal predicted values, except for R_{aw}, which is expressed as $cmH_2O/L/s$ (normal <2 $cmH_2O/L/s$). The figures at the bottom of each column show typical configuration of flow-volume loops in each condition, including the flow-volume relationship during tidal breathing. b.d., bronchodilator; DLCO, diffusion capacity of lung for carbon monoxide; FEV_1, forced expiratory volume in one second; FRC, functional residual capacity; FVC, forced vital capacity; R_{aw}, airways resistance; RV, residual volume; TLC, total lung capacity.

Ventilatory Restriction Due to Increased Elastic Recoil—Example: Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis raises lung recoil at all lung volumes, thereby lowering TLC, FRC, and RV, as well as FVC. Maximal expiratory flows are also reduced compared with normal values, but are relatively elevated

when considered in relation to lung volumes. The latter occurs both because the increased lung recoil drives greater maximal flow at any lung volume and because airway diameters are relatively increased due to greater radially outward traction exerted on bronchi by the stiff lung parenchyma. Airway resistance is also normal, for the same reason. Pulmonary capillaries are destroyed by the fibrotic process resulting in marked reduction in diffusing capacity. Oxygenation is often severely reduced due to persistent perfusion of alveolar units that are relatively underventilated due to fibrosis of nearby (and mechanically linked) lung. The flow-volume loop looks like a miniature version of a normal loop but is shifted toward lower absolute lung volumes and displays maximum expiratory flows that are increased for any given volume when compared to the normal tracing.

Ventilatory Restriction Due to Chest Wall Abnormality—Example: Moderate Obesity

As the size of the average American continues to increase, this pattern may become the most commonly seen of pulmonary function abnormalities. In moderate obesity, the outward recoil of the chest wall is blunted due to the weight of chest wall fat and to the space occupied by intraabdominal fat. As such, preserved inward recoil of the lung now overbalances the reduced outward recoil of the chest wall, and FRC falls. Because respiratory muscle strength and lung recoil remain normal, TLC is typically unchanged (although TLC may fall in massive obesity) and RV is normal (but may be reduced in massive obesity). Mild hypoxemia may be present, due to perfusion of alveolar units that are poorly ventilated because of airway closure that occurs in dependent portions of the lung while breathing near the reduced FRC. Flows remain normal, as does DL_{CO}, unless obstructive sleep apnea (which often accompanies obesity) and associated chronic intermittent hypoxemia have induced pulmonary arterial hypertension, in which case DL_{CO} may be low.

Ventilatory Restriction Due to Reduced Muscle Strength—Example: Myasthenia Gravis

FRC remains normal, as both lung recoil and passive chest wall recoil are normal. However, TLC is low and RV is elevated, as respiratory muscle strength is insufficient to push the passive respiratory system fully toward either volume extreme. Caught between the low TLC and the elevated RV, FVC and FEV_1 are reduced as "innocent bystanders." As airway size and the lung vasculature are unaffected, both airways resistance (R_{aw}) and DL_{CO} are normal. Oxygenation is normal unless weakness becomes so severe that the patient has insufficient strength to reopen collapsed alveoli during sighs, with resulting atelectasis.

Airflow Obstruction Due to Decreased Airway Diameter—Example: Acute Asthma

During an episode of acute asthma, luminal narrowing due to smooth muscle constriction and inflammation and thickening within the small- and medium-sized bronchi raise frictional resistance and reduce airflow. Scooping of the flow-volume loop is caused by reduction of airflow, especially at lower lung volumes. Often, airflow obstruction can be reversed by inhalation of β_2-adrenergic agonists acutely or by treatment with inhaled steroids chronically. Total lung capacity (TLC) usually remains normal (although elevated TLC is sometimes seen in long-standing asthma), but FRC may be dynamically elevated. RV is often increased due to exaggerated airway closure at low lung volumes, and this elevation of RV reduces FVC. Because central airways are narrowed, airways resistance is usually elevated. Mild arterial hypoxemia is often present due to perfusion of relatively underventilated alveoli distal to obstructed airways (and is responsive to oxygen supplementation), but DL_{CO} is normal or mildly elevated.

Airflow Obstruction Due to Decreased Elastic Recoil—Example: Severe Emphysema

Loss of lung elastic recoil in severe emphysema results in pulmonary hyperinflation, of which elevated TLC is the hallmark. FRC is more severely elevated due both to loss of lung elastic recoil and to dynamic hyperinflation (the same phenomenon as autoPEEP, which is the unintended positive end-expiratory pressure). Residual volume is very severely elevated due to airway closure and because exhalation toward RV may take so long that RV cannot be reached before the patient must inhale again. Both FVC and FEV_1 are markedly decreased, the former due to the severe elevation of RV, and the latter because loss of lung elastic recoil reduces the pressure driving maximal expiratory flow and also reduces tethering open of small intrapulmonary airways. The flow-volume loop demonstrates marked scooping of the flow-volume loop, with an initial transient spike of flow attributable largely to expulsion of air from collapsing central airways at the onset of forced exhalation. Otherwise, the central airways remain relatively unaffected, so R_{aw} is normal in "pure" emphysema. Loss of alveolar surface and capillaries in the alveolar walls reduces DL_{CO}, but because poorly ventilated emphysematous acini are also poorly perfused (due to loss of their capillaries), arterial hypoxemia is usually not seen at rest until emphysema becomes very severe. However, during exercise, P_aO_2 may fall precipitously if extensive destruction of the pulmonary vasculature prevents a sufficient increase in cardiac output and mixed venous oxygen content falls substantially. Under these circumstances, any venous admixture through low $\dot{V}/\dot{Q}$ units has a particularly marked effect in lowering mixed arterial oxygen tension.

Acknowledgement

The authors wish to acknowledge the contribution of Dr. Steven E. Weinberger and Irene M. Rosen to this chapter in previous edits and the helpful contributions of Drs. Mary Strek and Jeff Jacobson.

FURTHER READINGS

Hyatt RE et al: *Interpretation of Pulmonary Function Testing,* 3rd ed. Philadelphia, Lippincott Williams & Wilkins, 2009

MacIntyre N: Standardization of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J 26:720, 2005[PMID: 16204605] [Full Text]

Miller MR et al: Standardization of spirometry. Eur Respir J 26:319, 2005[PMID: 16055882] [Full Text]

Wanger J: Standardization of measurements of lung volumes. Eur Respir J 26:511, 2005[PMID: 16135736] [Full Text]

Weinberger SE: *Principles of Pulmonary Medicine,* 4th ed. Philadelphia, Saunders, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIAGNOSTIC PROCEDURES IN RESPIRATORY DISEASE: INTRODUCTION

The diagnostic modalities available for assessing the patient with suspected or known respiratory system disease include imaging studies and techniques for acquiring biologic specimens, some of which involve direct visualization of part of the respiratory system. Methods to characterize the functional changes developing as a result of disease, including pulmonary function tests and measurements of gas exchange, are discussed in Chap. 252.

IMAGING STUDIES

Routine Radiography

Routine chest radiography, generally including both posteroanterior (PA) and lateral views, is an integral part of the diagnostic evaluation of diseases involving the pulmonary parenchyma, the pleura, and, to a lesser extent, the airways and the mediastinum (see Chaps. 251 and **e34**). Lateral decubitus views are often useful for determining whether pleural abnormalities represent freely flowing fluid, whereas apical lordotic views can often visualize disease at the lung apices better than the standard PA view. Portable equipment is often used for acutely ill patients who either cannot be transported to a radiology suite or cannot stand for PA and lateral views. Portable films are more difficult to interpret owing to several limitations: (1) the single antero posterior (AP) projection obtained; (2) variability in over- and underexposure of film; (3) a shorter focal spot-film distance leading to lack of edge sharpness, and loss of fine detail; and (4) magnification of the cardiac silhouette and other anterior structures by the AP projection. Common radiographic patterns and their clinical correlates are reviewed in Chap. e34.

Advances in computer technology and the availability of reusable radiation detectors have allowed the development of digital or computed radiography. The images obtained in this format can be subjected to significant postprocessing analysis to improve diagnostic information. In addition, the benefit of immediate availability of the images, the ability to store images electronically, and the facility of transfer within or between health care systems have led many hospital systems to convert to digital systems.

Computed Tomography

Computed tomography (CT) offers several advantages over routine chest radiography (**Figs. 253-1A, B and 253-2A, B**; see also Figs. 261-3, 261-4, and 268-4). First, the use of cross-sectional images allows distinction between densities that would be superimposed on plain radiographs. Second, CT is far better than routine radiographic studies at characterizing tissue density, distinguishing subtle density differences between adjacent structures, and providing accurate size assessment of lesions.

Figure 253-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest x-ray (A) and CT scan (B) from a patient with emphysema. The extent and distribution of emphysema are not well appreciated on plain film but clearly evident on CT scan obtained.

Figure 253-2

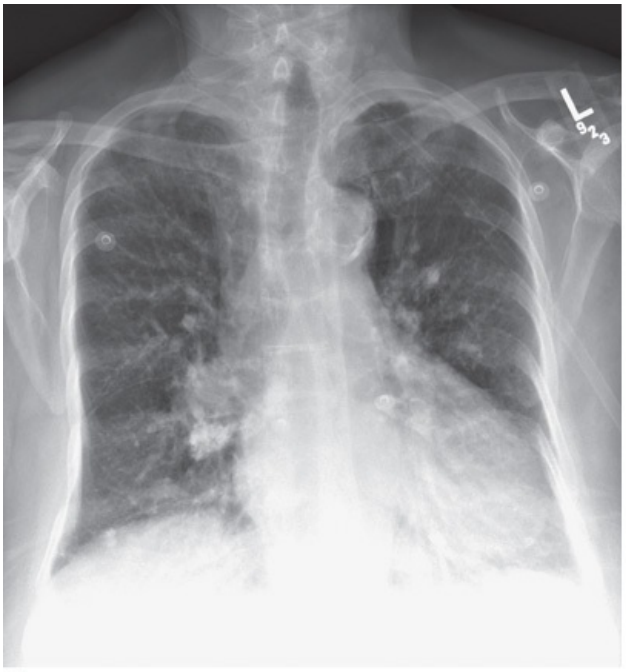

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest x-ray (A) and CT scan (B) demonstrating a right lower-lobe mass. The mass is not well appreciated on the plain film because of the hilar structures and known calcified adenopathy. CT is superior to plain radiography for the detection of abnormal mediastinal densities and the distinction of masses from adjacent vascular structures.

CT is particularly valuable in assessing hilar and mediastinal disease (which is often poorly characterized by plain radiography), in identifying and characterizing disease adjacent to the chest wall or spine (including pleural disease), and in identifying areas of fat density or calcification in pulmonary nodules (Fig. 253-2). Its utility in the assessment of mediastinal disease has made CT an important tool in the staging of lung cancer (Chap. 89), as an assessment of tumor involvement of mediastinal lymph nodes is critical to proper staging. With the additional use of contrast material, CT also makes it possible to distinguish vascular from nonvascular structures, which is particularly important in distinguishing lymph nodes and masses from vascular structures primarily in the mediastinum, and vascular disorders such as pulmonary embolism.

In high-resolution CT (HRCT), the thickness of individual cross-sectional images is ~1–2 mm, rather than the usual 7–10 mm in conventional CT. The visible detail on HRCT scans allows better recognition of subtle parenchymal and airway disease, thickened interlobular septa, ground-glass opacification, small nodules, and the abnormally thickened or dilated airways seen in bronchiectasis. Using HRCT, characteristic patterns are recognized for many interstitial lung diseases such as lymphangitic carcinoma, idiopathic pulmonary fibrosis, sarcoidosis, and eosinophilic granuloma. However, there is debate about, the settings in which the presence of a characteristic pattern on HRCT eliminates the need for obtaining lung tissue to make a diagnosis.

Helical CT Scanning

Recent advances in computer processing have allowed the development of helical CT scanning. Helical CT technology results in faster scans with improved contrast enhancement and thinner collimation. The image is obtained during a single breath-holding maneuver that allows less motion artifact. In addition, helical CT scanning allows the collection of continuous data over a larger volume of lung than is possible with conventional CT. Data from the imaging procedure can be reconstructed as images in planes other than the traditional cross-sectional (axial) view, including coronal, or sagittal planes (**Fig. 253-3**A). Finally, sophisticated volumetric "3D" representations of structures can be produced (Fig. 253-3B) including the ability to perform a *virtual bronchoscopy,* mimicking direct visualization through a bronchoscope (**Fig. 253-4**).

Figure 253-3

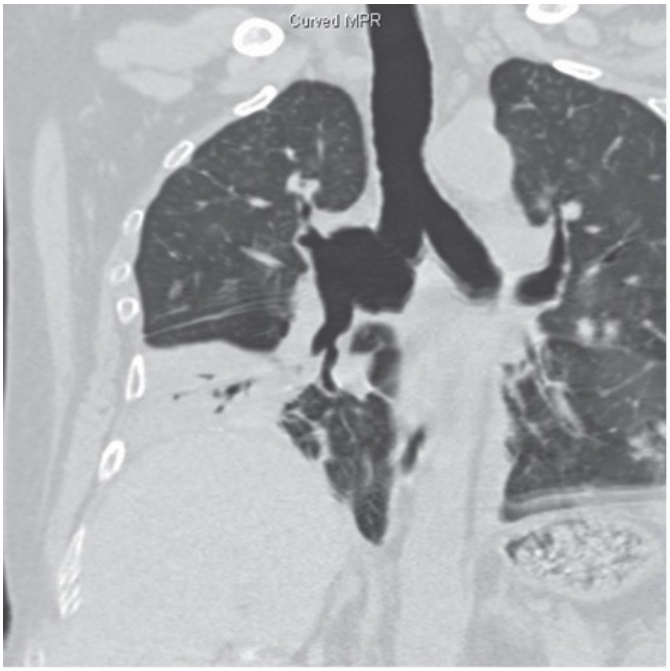

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

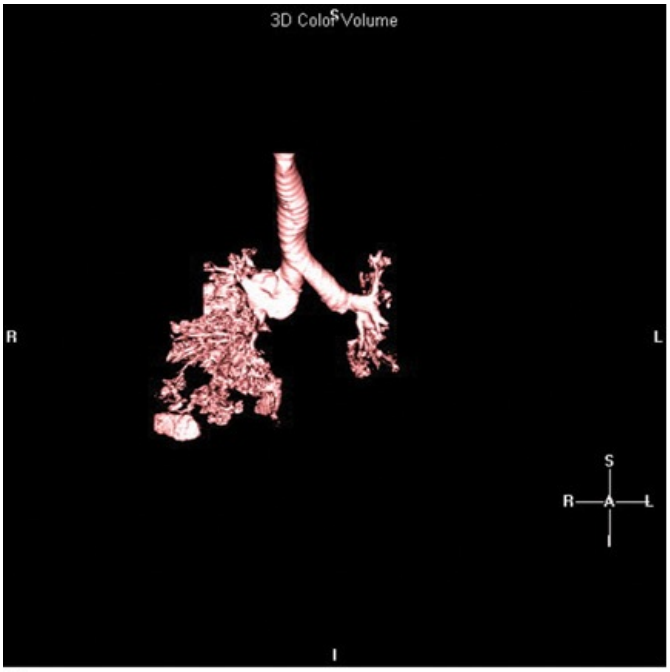

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spiral CT with reconstruction of images in planes other than axial view. Spiral CT in a lung transplant patient with a dehiscence and subsequent aneurysm of the anastomosis. CT images were reconstructed in the sagittal view (***A***) and using digital subtraction to view images of the airways only (***B***), which demonstrate the exact location and extent of the abnormality.

Figure 253-4

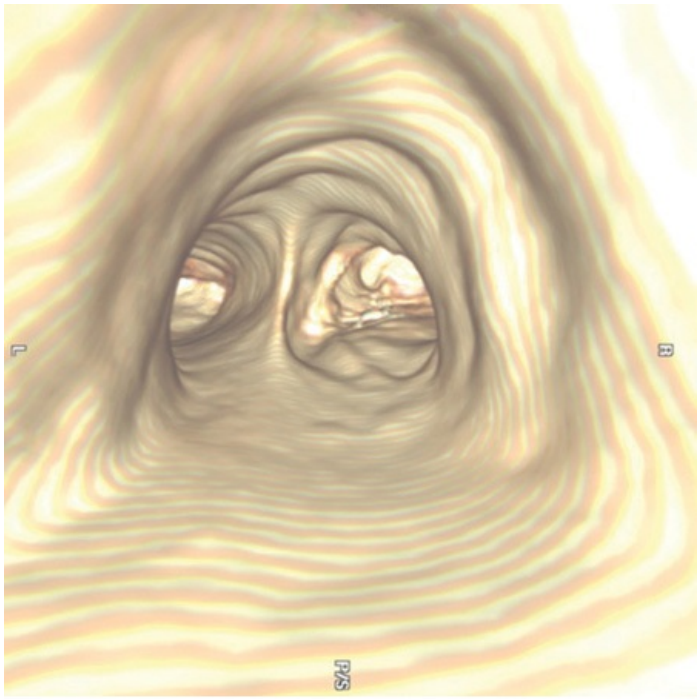

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRA image of the vasculature of a patient after lung transplant. The image demonstrates the detailed view of the vasculature that can be obtained using digital subtraction techniques. Images from a patient after lung transplant show the venous and arterial anastomosis on the right; a slight narrowing is seen at the site of the anastomosis, which is considered within normal limits and not suggestive of obstruction.

Multidetector CT (MDCT)

Refinements in detector technology have allowed production of scanners with additional detectors along the scanning axis (*z*-axis). These scanners, called *multidetector CT* (MDCT) scanners, can obtain multiple slices in a single rotation that are thinner and can be acquired in a shorter period of time. This results in enhanced resolution and increased image reconstruction ability. As the technology has progressed, higher numbers (2, 4, 6, 8, 10, 16, 32, 40 and currently up to 64) of detectors are used to produce clearer final images. The development of MDCT allows for even shorter breath holds, which are beneficial for all patients but especially children, the elderly, and the critically ill. However, it should be noted that despite the advantages of MDCT, there is an increase in radiation dose compared to single-detector CT to consider. With MDCT, the additional detectors along the *z*-axis result in improved use of the contrast bolus. In addition, the shorter breath holds secondary to faster scanning times and increased resolution have all led to improved imaging of the pulmonary vasculature and the ability to detect segmental and subsegmental emboli. In contrast to pulmonary angiography, CT pulmonary angiography (CTPA) also allows simultaneous detection of parenchymal abnormalities that may be contributing to a patient's clinical presentation. Secondary to these advantages and increasing availability, CTPA has rapidly become the test of choice for many clinicians in the evaluation of pulmonary embolism; it is considered equal to pulmonary angiography in terms of accuracy, and with less associated risks.

Magnetic Resonance Imaging

The role of magnetic resonance (MR) imaging in the evaluation of respiratory system disease is less well-defined than that of CT. Magnetic resonance provides poorer spatial resolution and less detail of the pulmonary parenchyma, and for these reasons is currently not considered a substitute for CT in imaging the thorax. However, the use of hyperpolarized gas in conjunction with MR has led to the investigational use of MR for imaging the lungs, particularly in obstructive lung disease. Of note, MR examinations are difficult to obtain among several subgroups of patients. Patients who cannot lie still or who cannot lie on their backs may have MR images that are of poor quality; some tests require patients to hold their breaths for 15 to 25 seconds at a time in order to get good MR images. MR is generally avoided in unstable and/or ventilated patients and those with severe trauma because of the hazards of the MR environment and the difficulties in monitoring patients within the MR room. The presence of metallic foreign bodies, pacemakers, and intracranial aneurysm clips also preclude use of MR.

An advantage of MR is the use of nonionizing electromagnetic radiation. Additionally, MR is well suited to distinguish vascular from nonvascular structures without the need for contrast. Blood vessels appear as hollow tubular structures because flowing blood does not produce a signal on MR imaging. Therefore, MR can be useful in demonstrating pulmonary emboli, defining aortic lesions such as aneurysms or dissection, or other vascular abnormalities (**Fig. 253-5**) if radiation and IV contrast medium cannot be used. Gadolinium can be used as an intravascular contrast agent for MR angiography (MRA); however, synchronization of data acquisition with the peak arterial bolus is one of the major challenges of MRA; the flow of contrast medium from the peripheral injection site to the vessel of interest is affected by a number of factors including heart rate, stroke volume, and the presence of proximal stenotic lesions.

Figure 253-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Virtual bronchoscopic image of the trachea. The view projected is one that would be obtained from the trachea looking down to the carina. The left and right main stem airways are seen bifurcating from the carina.

Nuclear Medicine Techniques

Nuclear imaging depends on the selective uptake of various compounds by organs of the body. In thoracic imaging, these compounds are concentrated by one of three mechanisms: blood pool or compartmentalization (e.g., within the heart), physiologic incorporation (e.g., bone or thyroid) and capillary blockage (e.g., lung scan). Radioactive isotopes can be administered by either the IV or inhaled routes or both. When injected intravenously, albumin macroaggregates labeled with [^{99m}Tc] become lodged in pulmonary capillaries; the distribution of the trapped radioisotope follows the distribution of blood flow. When inhaled, radiolabeled xenon gas can be used to demonstrate the distribution of ventilation. Using these techniques, ventilation-perfusion lung scanning was a commonly used technique for the evaluation of pulmonary embolism. Pulmonary thromboembolism produces one or more regions of ventilation-perfusion mismatch [i.e., regions in which there is a defect in perfusion that follows the distribution of a vessel and that is not accompanied by a corresponding defect in ventilation (Chap. 262)]. However, with advances in CT scanning, scintigraphic imaging has been largely replaced by CT angiography in patients with suspected pulmonary embolism.

Another common use of ventilation-perfusion scans is in patients with impaired lung function, who are being considered for lung resection. Because many patients with bronchogenic carcinoma have coexisting chronic obstructive pulmonary disease (COPD), the question arises as to whether or not a patient can tolerate lung resection. The distribution of the isotope(s) can be used to assess the regional distribution of blood flow and ventilation, allowing the physician to estimate the level of postoperative lung function.

Positron Emission Tomographic Scanning

Positron emission tomographic (PET) scanning is commonly used to identify malignant lesions in the lung, based on their increased uptake and metabolism of glucose. The technique involves injection of a radiolabeled glucose analogue, [^{18}F]-fluoro-2-deoxyglucose (FDG), which is taken up by metabolically active malignant cells. However, FDG is trapped within the cell following phosphorylation, and the unstable [^{18}F] decays by emission of positrons, which can be detected by a specialized PET camera or by a gamma camera that has been adapted for imaging of positron-emitting nuclides. This technique has been used in the evaluation of solitary pulmonary nodules and in staging lung cancer through the detection or exclusion of mediastinal lymph node involvement and identification of

extrathoracic disease. The limited anatomical definition of radionuclide imaging has been improved by the development of hybrid imaging that allows the superimposition of nuclear medicine and CT images, a technique known as functional–anatomical mapping. Today, most PET scans are performed using instruments with combined PET and CT scanners. The hybrid PET/CT scans provide images that help pinpoint the abnormal metabolic activity to anatomical structures seen on CT. The combined scans provide more accurate diagnoses than the two scans performed separately. FDG–PET can differentiate benign from malignant lesions as small as 1 cm. However, false-negative findings can occur in lesions with low metabolic activity such as carcinoid tumors and bronchioloalveolar cell carcinomas, or in lesions <1 cm in which the required threshold of metabolically active malignant cells is not present for PET diagnosis. False-positive results can be seen due to FDG uptake in inflammatory conditions such as pneumonia and granulomatous diseases.

Pulmonary Angiography

The pulmonary arterial system can be visualized by pulmonary angiography, in which radiopaque contrast medium is injected through a catheter placed in the pulmonary artery. When performed in cases of pulmonary embolism, pulmonary angiography demonstrates the consequences of an intravascular thrombus—either a defect in the lumen of a vessel (a filling defect) or an abrupt termination (cutoff) of the vessel. Other, less common indications for pulmonary angiography include visualization of a suspected pulmonary arteriovenous malformation and assessment of pulmonary arterial invasion by a neoplasm. The risks associated with modern arteriography are extremely small, generally of greatest concern in patients with severe pulmonary hypertension. With advances in CT scanning, MDCT angiography (MDCTA) is replacing conventional angiography for the diagnosis of pulmonary embolism.

Ultrasound

Diagnostic ultrasound (US) produces images using echoes or reflection of the ultrasound beam from interfaces between tissues with differing acoustic properties. US is nonionizing and safe to perform on pregnant patients and children. It is helpful in the detection and localization of pleural abnormalities, and a quick and effective way of guiding percutaneous needle biopsy of peripheral lung, pleural, or chest wall lesions. US is also helpful in identifying septations within loculated collections and can facilitate placement of a needle for sampling of pleural liquid (i.e., for thoracentesis), improving the yield and safety of the procedure. Bedside availability makes it valuable in the intensive care setting. Real-time imaging can be used to assess the movement of the diaphragm. Using the Doppler mode, patterns of blood flow in both large and small vessels can be visualized. Because US energy is rapidly dissipated in air, it is not useful for evaluation of the pulmonary parenchyma and cannot be used if there is any aerated lung between the US probe and the abnormality of interest.

Endobronchial US, in which the US probe is passed through a bronchoscope, is emerging as a valuable adjunct to bronchoscopy, allowing identification and localization of pathology adjacent to airway walls or within the mediastinum, discussed further below.

Virtual Bronchoscopy

The three-dimensional (3D) image of the thorax obtained by MDCT can be digitally stored, reanalyzed, and displayed as 3D reconstructions of the airways down to the sixth- to seventh-generation. Using these computed generated reconstructions, a "virtual" bronchoscopy can be performed (Fig. 253-5). Virtual bronchoscopy has been proposed as an adjunct to conventional bronchoscopy in several clinical situations: It can allow accurate assessment of the extent and length of an airway stenosis, including the airway distal to the narrowing; it can provide useful information about the relationship of the airway abnormality to adjacent mediastinal structures; and it allows preprocedure planning for therapeutic bronchoscopy to help ensure the appropriate equipment is available for the procedure. Virtual bronchoscopy can also be used to perform noninvasive follow-up of patients with treated airway lesions. Navigational systems using virtual bronchoscopy have been developed to allow pathfinding to guide the bronchoscopist to a peripheral region within the lung, allowing peripheral lung lesions to be sampled more efficiently. Finally, with the advent of endobronchial lung volume reduction surgery in the management of pulmonary emphysema, virtual bronchoscopy may be able to help target the area of peripheral lung for endobronchial valve procedures. The extent of emphysema in each segmental region together with other anatomic details may help in choosing the most appropriate subsegments. However, software packages for the generation of virtual bronchoscopic images are relatively early in development and their utilization and potential impact on patient care are still unknown. In addition to allowing virtual bronchoscopy, advances in computing capabilities and digital imaging allow the bronchoscopic images obtained through a real bronchoscopic examination to be stored as digital

images and reviewed after completion of the procedure.

MEDICAL TECHNIQUES FOR OBTAINING BIOLOGIC SPECIMENS

Collection of Sputum

Sputum can be collected either by spontaneous expectoration or after inhalation of an irritating aerosol such as hypertonic saline. The latter method, called *sputum induction,* is commonly used to obtain sputum for diagnostic studies, either because sputum is not spontaneously being produced or because of an expected higher yield of certain types of findings. Knowledge of the appearance and quality of the sputum specimen obtained is especially important when one is interested in Gram's method and culture. Because sputum consists mainly of secretions from the tracheobronchial tree rather than the upper airway, the finding of alveolar macrophages and other inflammatory cells is consistent with a lower respiratory tract origin of the sample, whereas the presence of squamous epithelial cells in a "sputum" sample indicates contamination by secretions from the upper airways.

In addition to processing for routine bacterial pathogens by Gram's method and culture, sputum can be processed for a variety of other pathogens, including staining and culture for mycobacteria or fungi, culture for viruses, and staining for *Pneumocystis jiroveci.* In the specific case of sputum obtained for evaluation of *P. jiroveci* pneumonia in a patient infected with HIV, for example, sputum should be collected by induction rather than spontaneous expectoration, and an immunofluorescent stain should be used to detect the organisms. Cytologic staining of sputum for malignant cells, using the traditional Papanicolaou method, allows noninvasive evaluation for suspected lung cancer. Traditional stains and cultures are now also being supplemented in some cases by immunologic techniques and by molecular biologic methods, including the use of polymerase chain reaction amplification and DNA probes.

Percutaneous Needle Aspiration (Transthoracic)

A needle can be inserted through the chest wall into a pulmonary lesion to aspirate material for analysis by cytologic or microbiologic techniques. Aspiration can be performed to obtain a diagnosis or to decompress and/or drain a fluid collection. The procedure is usually carried out under CT or ultrasound guidance to assist positioning of the needle and assure localization in the lesion. The low potential risk of this procedure (intrapulmonary bleeding or creation of a pneumothorax with collapse of the underlying lung) in experienced hands is usually acceptable owing to the information obtained. However, a limitation of the technique is sampling error due to the small size of the tissue sample. Thus, findings other than a specific cytologic or microbiologic diagnosis are of limited clinical value.

Thoracentesis

Sampling of pleural liquid by thoracentesis is commonly performed for diagnostic purposes or, in the case of a large effusion, for palliation of dyspnea. Diagnostic sampling, either by blind needle aspiration or after localization by US, allows the collection of liquid for microbiologic and cytologic studies. Analysis of the fluid obtained for its cellular composition and chemical constituents, including glucose, protein, and lactate dehydrogenase, allows the effusion to be classified as either exudative or transudative (Chap. 263).

Bronchoscopy

Bronchoscopy is the process of direct visualization of the tracheobronchial tree. Although bronchoscopy is now performed almost exclusively with flexible fiberoptic instruments, rigid bronchoscopy, generally performed in an operating room on a patient under general anesthesia, still has a role in selected circumstances, primarily because of a larger suction channel and the fact that the patient can be ventilated through the bronchoscope channel. These situations include the retrieval of a foreign body and the suctioning of a massive hemorrhage, for which the small suction channel of the bronchoscope may be insufficient.

Flexible Fiberoptic Bronchoscopy

This outpatient procedure is usually performed in an awake but sedated patient (conscious sedation). The bronchoscope is passed through either the mouth or the nose, between the vocal cords, and into the trachea. The ability to flex the scope makes it possible to visualize virtually all airways to the level of subsegmental bronchi. The bronchoscopist is able to identify endobronchial pathology, including tumors, granulomas, bronchitis, foreign bodies, and sites of bleeding. Samples from airway lesions can be taken by several methods, including washing, brushing, and biopsy. Washing involves instillation of sterile saline through a channel of the bronchoscope and onto the surface of a lesion. A portion of the liquid is collected by suctioning through the bronchoscope, and the recovered material can be analyzed for cells (cytology) or organisms (by standard stains and cultures). Brushing or biopsy of the surface of the lesion, using a small brush or biopsy forceps at the end of a long cable inserted through a channel of the bronchoscope, allows recovery of cellular material or tissue for analysis by standard cytologic and histopathologic methods.

The bronchoscope can be used to sample material not only from the regions that can be directly visualized (i.e., the airways) but also from the more distal pulmonary parenchyma. With the

bronchoscope wedged into a subsegmental airway, aliquots of sterile saline can be instilled through the scope, allowing sampling of cells and organisms even from alveolar spaces. This procedure, called *bronchoalveolar lavage,* has been particularly useful for the recovery of organisms such as *P. jiroveci* in patients with HIV infection.

Brushing and biopsy of the distal lung parenchyma can also be performed with the same instruments that are used for endobronchial sampling. These instruments can be passed through the scope into small airways, where they penetrate the airway wall, allowing biopsy of peribronchial alveolar tissue. This procedure, called *transbronchial biopsy,* is used when there is either relatively diffuse disease or a localized lesion of adequate size. With the aid of fluoroscopic imaging, the bronchoscopist is able to determine not only whether and when the instrument is in the area of abnormality, but also the proximity of the instrument to the pleural surface. If the forceps are too close to the pleural surface, there is a risk of violating the visceral pleura and creating a pneumothorax; the other potential complication of transbronchial biopsy is pulmonary hemorrhage. The incidence of these complications is less than several percent.

Transbronchial Needle Aspiration (TBNA)

Another procedure involves use of a hollow-bore needle passed through the bronchoscope for sampling of tissue adjacent to the trachea or a large bronchus. The needle is passed through the airway wall (transbronchial), and cellular material can be aspirated from mass lesions or enlarged lymph nodes, generally in a search for malignant cells. Other promising new techniques that are not yet widely available include fluorescence bronchoscopy (to detect early endobronchial malignancy) and endobronchial ultrasound (to better identify and localize peribronchial and mediastinal pathology). Mediastinoscopy has been considered the gold standard for mediastinal staging; however, TBNA allows sampling from the lungs and surrounding lymph nodes without the need for surgery or general anesthesia.

Endobronchial Ultrasound (EBUS)–Transbronchial Needle Aspiration (TBNA)

Further advances in needle aspiration techniques have been accomplished with the development of endobronchial ultrasound (EBUS). The technology uses an ultrasonic bronchoscope fitted with a probe that allows for needle aspiration of mediastinal and hilar lymph nodes guided by real-time US images. This procedure offers access to more difficult-to-reach areas and smaller lymph nodes in the staging of malignancies. EBUS–TBNA has the potential to access the same paratracheal and subcarinal lymph node stations as mediastinoscopy, but also extends out to the hilar lymph nodes (levels 10 and 11).The usefulness of EBUS for clinical indications other than lung cancer is unclear, although studies on sarcoidosis point to the effectiveness of endobronchial ultrasonography in diagnosing this disease.

Interventional Pulmonology (IP)

Interventional pulmonology was initially developed to focus on procedures to help palliate patients with advanced thoracic malignancies. However, the availability of advanced bronchoscopic and pleuroscopic techniques is enabling interventional pulmonologists to provide alternatives to surgery for patients with a wide variety of thoracic disorders and problems. IP can be defined as "the art and science of medicine as related to the performance of diagnostic and invasive therapeutic procedures that which require additional training and expertise beyond that which required in a standard pulmonary medicine training program."

A central role for an IP physician is the acquisition of tissue for diagnosing mass lesions within the thorax. Several techniques already discussed are part of the day-to-day procedural armamentarium used by an IP physician. TBNA to obtain cytologic, histologic, or microbiologic sampling of lesions within the airway wall, the lung parenchyma, and mediastinum. TBNA is frequently performed in combination with EBUS to improve diagnostic yield. Transthoracic needle aspiration and biopsy (TTNA/B) refers to the percutaneous sampling of lesions involving the chest wall, lung parenchyma, and mediastinum for cytologic, histopathologic, or microbiologic examinations.

Autofluorescence Bronchoscopy

Autofluorescence bronchoscopy (AFB) uses bronchoscopes with an additional light source that allows an experienced operator (interventional pulmonologist or surgeon) to distinguish between normal and abnormal tissue. This technique can be used as a screening tool in high-risk individuals to inspect the tracheobronchial tree in order to identify premalignant lesions (airway dysplasia) and carcinoma in situ.

MEDICAL THORACOSCOPY

Medical thoracoscopy (or pleuroscopy) focuses on the diagnosis of pleural-based problems. The procedure is performed with a conventional rigid or a semirigid pleuroscope (similar in design to a bronchoscope and enabling the operator to inspect the pleural surface, sample and/or drain pleural fluid, or perform targeted biopsies of the parietal pleura). Medical thoracoscopy can be performed in the endoscopy suite or operating room with the patient under conscious sedation and local anesthesia. In contrast, video-assisted thoracoscopic surgery (VATS) requires general anesthesia and is only

performed in the OR. A common diagnostic indication for medical thoracoscopy is the evaluation of a pleural effusion or biopsy of presumed parietal pleural carcinomatosis. It can also be used to place a chest tube under visual guidance, or perform chemical or talc pleurodesis, a therapeutic intervention to prevent a recurrent pleural effusion (usually malignant) or recurrent pneumothorax.

Therapeutic Bronchoscopy

The bronchoscope may provide the opportunity for treatment as well as diagnosis. A central role of the IP physican is the performance of therapeutic bronchoscopy. For example, an aspirated foreign body may be retrieved with an instrument passed through the bronchoscope (either flexible or rigid), and bleeding may be controlled with a balloon catheter similarly introduced. Newer interventional techniques performed through a bronchoscope include methods for achieving and maintaining patency of airways that are partially or completely occluded, especially by tumors. These techniques include laser therapy, cryotherapy, argon plasma coagulation, electrocautery, balloon bronchoplasty and dilation, and stent placement. Many IP physicians are also trained in performing percutaneous tracheotomy.

SURGICAL TECHNIQUES FOR OBTAINING BIOLOGIC SPECIMENS

Evaluation and diagnosis of disorders of the chest commonly involve collaboration between pulmonologists and thoracic surgeons. While procedures such as mediastinoscopy, VATS, and thoracotomy are performed by thoracic surgeons, there is overlap in many minimally invasive techniques that can be performed by a pulmonologist or a thoracic surgeon.

Mediastinoscopy and Mediastinotomy

Proper staging of lung cancer is of paramount concern when determining a treatment regimen. Although CT and PET scanning are useful for determining the size and nature of mediastinal lymph nodes as part of the staging of lung cancer, tissue biopsy and histopathologic examination are often critical for the diagnosis of mediastinal masses or enlarged mediastinal lymph nodes. The two major surgical procedures used to obtain specimens from masses or nodes in the mediastinum are mediastinoscopy (via a suprasternal approach) and mediastinotomy (via a parasternal approach). Both procedures are performed under general anesthesia by a qualified surgeon. In the case of suprasternal mediastinoscopy, a rigid mediastinoscope is inserted at the suprasternal notch and passed into the mediastinum along a pathway just anterior to the trachea. Tissue can be obtained with biopsy forceps passed through the scope, sampling masses or nodes that are in a paratracheal or pretracheal position (levels 2R, 2L, 3, 4R, 4L). Aortopulmonary lymph nodes (levels 5, 6) are not accessible by this route and thus are commonly sampled by parasternal mediastinotomy (the Chamberlain procedure). This approach involves a parasternal incision and dissection directly down to a mass or node that requires biopsy.

As an alternative to surgery, a bronchoscope can be used to perform TBNA (discussed above) to obtain tissue from the mediastinum, and, when combined with EBUS, can allow access to the same lymph node stations associated with mediastinoscopy, but also extend access out to the hilar lymph nodes (levels 10, 11). Finally, endoscopic ultrasound (EUS)–fine-needle aspiration (FNA) is a second procedure that complements EBUS–FNA in the staging of lung cancer. EUS–FNA is performed via the esophagus and is ideally suited for sampling lymph nodes in the posterior mediastinum (levels 7, 8, 9). Because US imaging cannot penetrate air filled spaces, the area directly anterior to the trachea cannot accurately be assessed and is a "blind spot" for EUS–FNA. However, EBUS–FNA can visualize the anterior lymph nodes and can complement EUS–FNA. The combination of EUS–FNA and EBUS–FNA is a technique that is becoming an alternative to surgery for staging the mediastinum in thoracic malignancies.

Video-Assisted Thoracic Surgery

Advances in video technology have allowed the development of thoracoscopy, or VATS, for the diagnosis and management of pleural as well as parenchymal lung disease. This procedure is performed in the operating room using single-lung ventilation with double-lumen endotracheal intubation and involves the passage of a rigid scope with a distal lens through a trocar inserted into the pleura. A high-quality image is shown on a monitor screen, allowing the operator to manipulate instruments passed into the pleural space through separate small intercostal incisions. With these instruments the operator can biopsy lesions of the pleura under direct visualization. In addition, this procedure is now used commonly to biopsy peripheral lung tissue or to remove peripheral nodules for both diagnostic and therapeutic purposes. This much less invasive procedure has largely supplanted the traditional "open lung biopsy" performed via thoracotomy. The decision to use a VATS technique versus performing an open thoracotomy is made by the thoracic surgeon and is based on whether a patient can tolerate the single-lung ventilation that is required to allow adequate visualization of the lung. With further advances in instrumentation and experience, VATS can be used to perform procedures previously requiring thoracotomy, including stapled lung biopsy, resection of pulmonary nodules, lobectomy, pneumonectomy, pericardial window, or other standard thoracic surgical procedures; but allows them to be performed in a minimally invasive manner.

Thoracotomy

Although frequently replaced by VATS, thoracotomy remains an option for the diagnostic sampling of lung tissue. It provides the largest amount of material, and it can be used to biopsy and/or excise lesions that are too deep or too close to vital structures for removal by VATS. The choice between VATS and thoracotomy needs to be made on a case-by-case basis.

Acknowledgment

We wish to acknowledge Dr. Scott Manaker and Dr. Steven Weinberger for their contributions to prior versions of this chapter.

FURTHER READINGS

Brown ma, Semelka RC: *MRI: Basic Principles and Applications,* 3rd ed. Hoboken, Wiley-Liss, 2003

Detterbeck FC. Evolution and science, progress and change. Thorax 62:654, 2007[PMID: 17687093] [Full Text]

———, et al: Seeking a home for a PET. Defining the appropriate place for positron emission tomography imaging in the diagnosis of pulmonary nodules or masses. Parts 1 and 2. Chest 125:2294, 2300, 2004

De Wever W et al: Multidetector CT-generated virtual bronchoscopy: An illustrated review of the potential clinical indications. Eur Respir J 23:776, 2004

———: Virtual bronchoscopy: Accuracy and usefulness—an overview. Semin Ultrasound CT MR 26:364, 2005

Kalra MK, Saini S: A practical approach to MDCT, in *MDCT: A Practical Approach,* S Saini et al (eds). Milan, Springer, 2006

Kavanagh JJ et al: Pulmonary embolism imaging with MDCT, in *MDCT: A Practical Approach* , S Saini et al (eds). Milan, Springer, 2006

Lee P, Colt HG: State of the art: Pleuroscopy. J Thorac Oncol 2:663, 2007[PMID: 17607126] [Full Text]

Medford ARL et al: Mediastinal staging procedures in lung cancer: EBUS, TBNA and mediastinoscopy. Curr Opin Pulmon Med 15:334, 2009[PMID: 19395972] [Full Text]

Moghissi K et al: Current indications and future perspective of fluorescence bronchoscopy: A review study. Photodiagn Photodyn Ther 5:238, 2008[PMID: 19356663] [Full Text]

Muller NL, Silva CIS: Normal chest radiograph, in *Imaging of the Chest* , vol. 1, NL Muller, CIS Silva (eds). Philadelphia, Saunders/Elsevier, 2008

Patel S, Kazerooni EA: Helical CT for the evaluation of acute pulmonary embolism. AJR Am J Roentgenol 185:135, 2005[PMID: 15972414] [Full Text]

Varela-Lema L et al: Effectiveness and safety of endobronchial ultrasound–transbronchial needle aspiration: A systematic review. Eur Respir J 33:1156, 2009[PMID: 19407050] [Full Text]

Wahidi MM et al: State of the art interventional pulmonology. Chest 131:261, 2007[PMID: 17218585] [Full Text]

Wallace MB et al. Minimally invasive endoscopic staging of suspected lung cancer. JAMA 299:540, 2008[PMID: 18252884] [Full Text]

Weinberger SE (ed): *Principles of Pulmonary Medicine,* 4th ed. Philadelphia, Saunders, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF CHEST IMAGING

This atlas of chest imaging is a collection of interesting chest radiographs and computed tomograms of the chest. The readings of the films are meant to be illustrative of specific, major findings. The associated text is not intended as a comprehensive assessment of the images.

EXAMPLES OF NORMAL IMAGING

Figure e34-1

Normal chest radiograph—review of anatomy.1. Trachea. 2. Carina. 3. Right atrium. 4. Right hemidiaphragm. 5. Aortic knob. 6. Left hilum. 7. Left ventricle. 8. Left hemidiaphragm (with stomach bubble). 9. Retrosternal clear space. 10. Right ventricle. 11. Left hemidiaphragm (with stomach bubble). 12. Left upper lobe bronchus.

Figure e34-2

Normal chest tomogram—note anatomy. 1. Superior vena cava. 2. Trachea. 3. Aortic arch. 4. Ascending aorta. 5. Right mainstem bronchus. 6. Descending aorta. 7. Left mainstem bronchus. 8. Main pulmonary artery. 9. Heart. 10. Esophagus. 11. Pericardium. 12. Descending aorta.

VOLUME LOSS

Figure e34-3

CT scan demonstrating left upper lobe collapse. The patient was found to have an endobronchial lesion (not visible on theCT scan) resulting in this finding. The superior vena cava (*black arrow*) is partially opacified by intravenous contrast.

Figure e34-4

CT scan revealing chronic left lower lobe collapse. Note dramatic volume loss with minimal aeration. There is subtle mediastinal shift to the left.

Figure e34-5

Left upper lobe scarring with hilar retraction with less prominent scarring in right upper lobe as well. Findings consistent with previous tuberculosis infection in an immigrant from Ecuador.

Figure e34-6

Apical scarring, traction bronchiectasis (*red arrow*), and decreased lung volume consistent with previous tuberculosis infection. Findings most significant in left lung.

Figure e34-7

Chest x-ray (CXR) demonstrating right upper lobe collapse (*yellow arrow*). Note the volume

loss as demonstrated by the elevated right hemidiaphragm as well as mediastinal shift to the right. Also apparent on the film are an endotracheal tube (*red arrow*) and a central venous catheter (*black arrow*).

Figure e34-8

Opacity in the right upper lobe. Note the volume loss as indicated by the elevation of the right hemidiaphragm, elevation of minor fissure (*yellow arrow*) and deviation of the trachea to the right (*blue arrow*).

Figure e34-9

CT scan of the same right upper lobe opacity. Note the air bronchograms and areas of consolidation.

LOSS OF PARENCHYMA

Figure e34-10

Emphysema with increased lucency, flattened diaphragms (*black arrows*), increased AP diameter, and increased retrosternal clear space (*red arrow*).

Figure e34-11

CT scan of diffuse, bilateral emphysema.

Figure e34-12

CT scan of bullous emphysema.

Figure e34-13

Lymphangioleiomyomatosis—note multiple thin-walled parenchymal cysts.

Figure e34-14

Two cavities on posteroanterior (PA) and lateral CXR. Cavities and air-fluid levels identified by *blue arrows*. The smaller cavity is in the right lower lobe (located below the major fissure, identified with the *yellow arrow*) and the larger cavity is located in the right middle lobe which is located between the minor (*blue arrow*) and major fissures. There is an associated opacity surrounding the cavity in the right lower lobe.

Figure e34-15

CT scan of parenchymal cavity.

Figure e34-16

Thick-walled cavitary lung lesions. The mass in the right lung has thick walls and advanced cavitation, while the smaller nodule on the left has early cavitary changes (*arrow*). This patient was diagnosed with Nocardia infection.

INTERSTITIAL PROCESSES

Figure e34-17

Mild congestive heart failure. Note the Kerley B lines (*black arrow*) and perivascular cuffing (*yellow arrow*) as well as the pulmonary vascular congestion (*red arrow*).

Figure e34-18

Pulmonary edema. Note indistinct vasculature, perihilar opacities, and peripheral interstitial reticular opacities. While this is an anteroposterior film making cardiac size more difficult to assess, the cardiac silhouette still appears enlarged.

Figure e34-19

CXR demonstrates reticular nodular opacities bilaterally with small lung volumes consistent with usual interstitial pneumonitis (UIP) on pathology. Clinically, UIP is used interchangeably with idiopathic pulmonary fibrosis (IPF).

Figure e34-20

CT scan of usual interstitial pneumonitis (UIP), also known as idiopathic pulmonary fibrosis (IPF). Classic findings include traction bronchiectasis (*black arrow*) and honeycombing (*red arrows*). Note subpleural, basilar predominance of the honeycombing.

Figure e34-21

(A) PA chest film—note presence of paratracheal (*blue arrow*), aortopulmonary window (*yellow arrow*) and hilar (*purple arrows*) lymphadenopathy.

(B) Lateral film—note hilar lymphadenopathy (*purple arrow*)

Figure e34-22

Sarcoid—CT scan of stage I demonstrating bulky hilar and mediastinal lymphadenopathy (*red arrows*).

Figure e34-23

Sarcoid—CXR of stage II.

(A) PA film with hilar lymphadenopathy (*green arrows*) and parenchymal changes.

(B) Lateral film with hilar adenopathy (*green arrow*) and parenchymal changes

Figure 34-24

Sarcoid—CT scan of stage II (calcified lymphadenopathy, parenchymal infiltrates).

Figure e34-25

Sarcoid—CT scan of stage II (nodular opacities tracking along bronchovascular bundles).

Figure e34-26

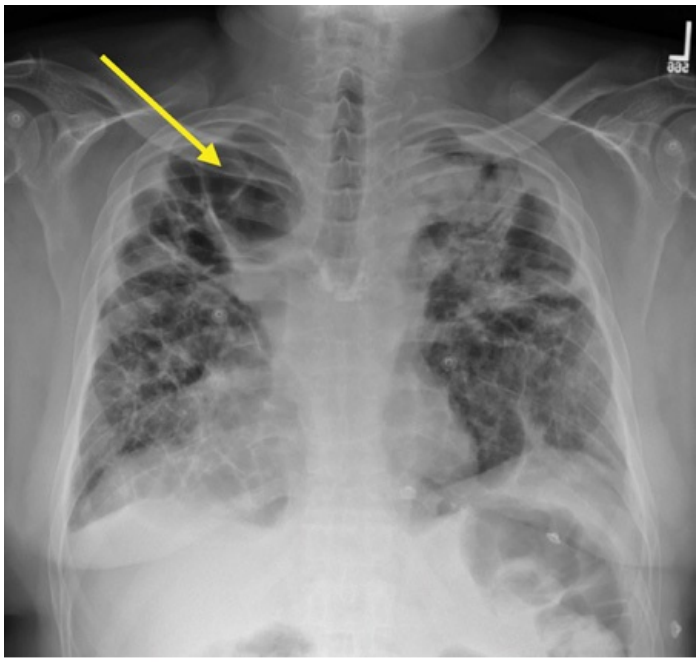

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sarcoid—stage IV with fibrotic lung disease and cavitary areas (*yellow arrow*).

ALVEOLAR PROCESSES

Figure e34-27

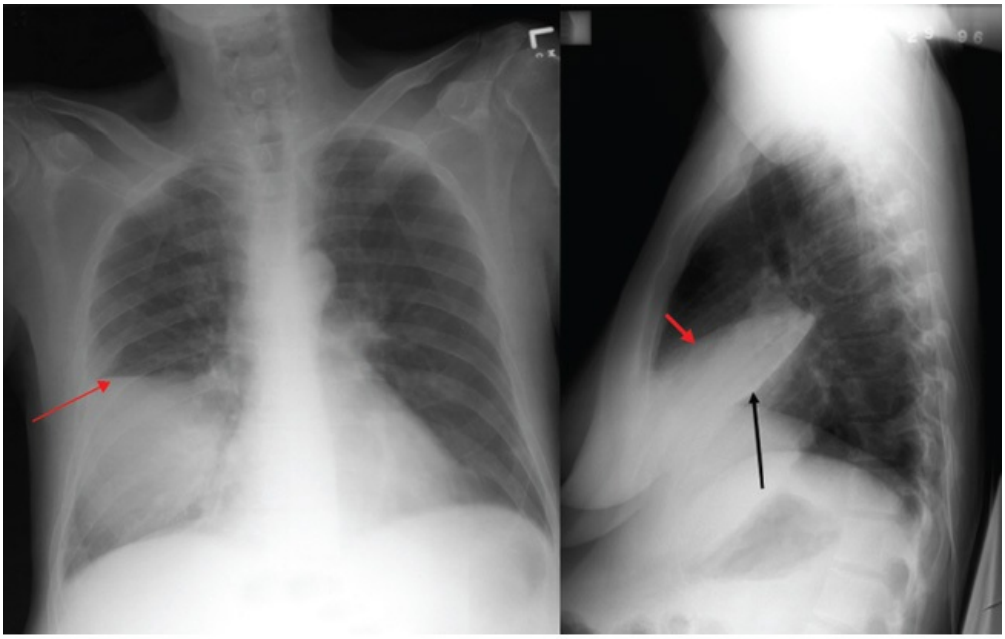

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Right middle lobe opacity illustrates major (*black arrow*) and minor fissures (*red arrows*) as well as the "silhouette sign" on the right heart border.

Figure e34-28

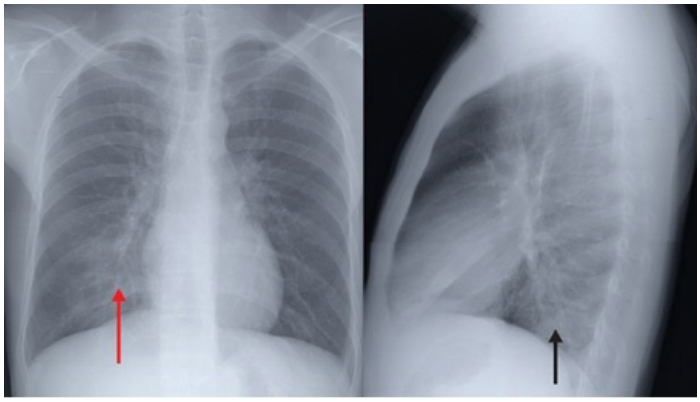

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Right lower lobe pneumonia—subtle opacity on PA film (*red arrow*), while the lateral film illustrates the "spine sign" (*black arrow*) where the lower spine does not become more lucent.

Figure e34-29

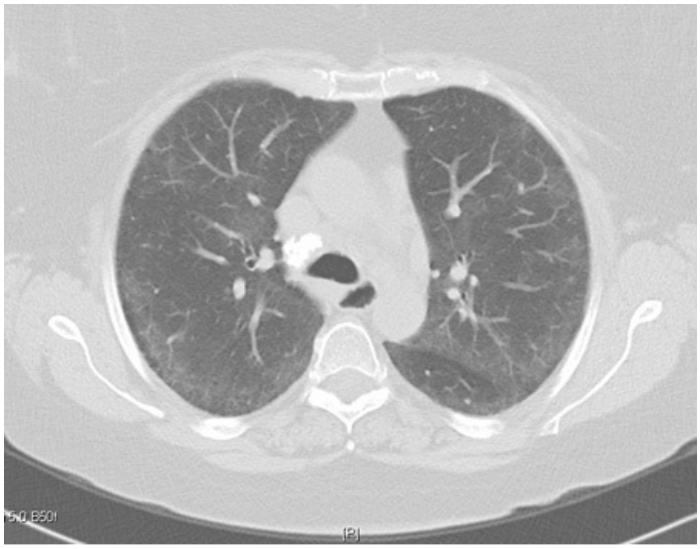

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of diffuse, bilateral "ground-glass" opacities. This finding is consistent with fluid density in the alveolar space.

Figure e34-30

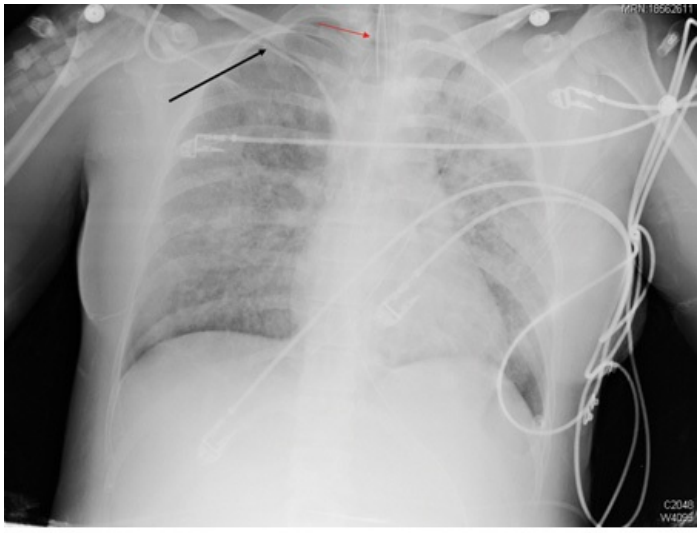

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CXR reveals diffuse, bilateral alveolar opacities without pleural effusions, consistent with acute respiratory distress syndrome (ARDS). Note that the patient has an endotracheal tube (*red arrow*) and a central venous catheter (*black arrow*).

Figure e34-31

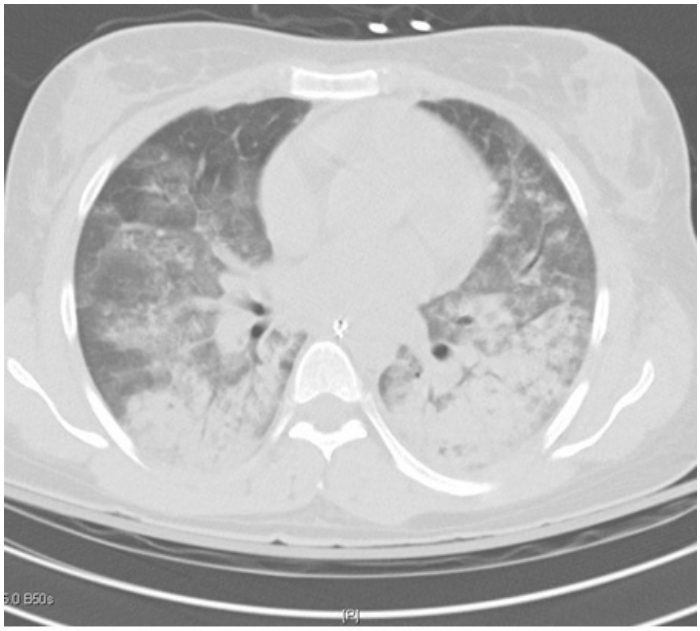

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of ARDS demonstrates "ground-glass" opacities with more consolidated areas in the dependent lung zones.

Figure e34-32

Three examples of air bronchograms (*red arrows*) on chest CT.

BRONCHIECTASIS

Figure e34-33

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cystic fibrosis with bronchiectasis, apical disease.

Figure e34-34

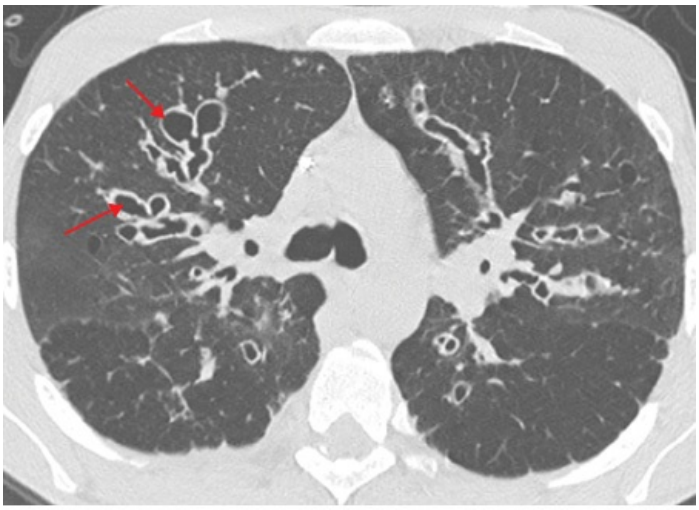

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of diffuse, cystic bronchiectasis (*red arrows*) in a patient with cystic fibrosis.

Figure e34-35

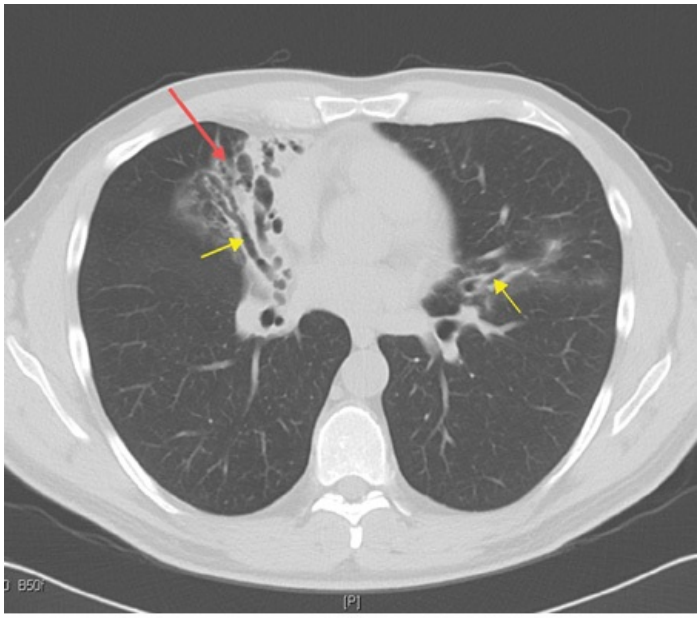

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of focal right middle lobe and lingular bronchiectasis (*yellow arrows*). Note that there is near total collapse of the right middle lobe (*red arrow*).

Figure e34-36

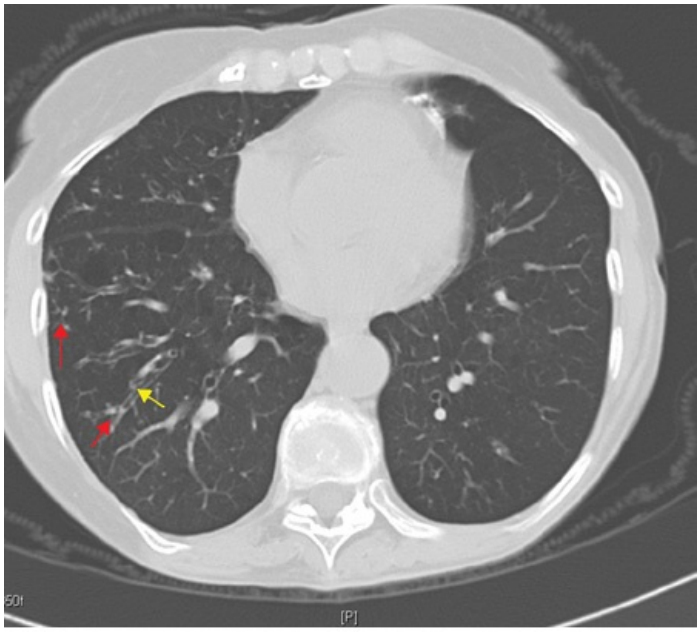

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

"Tree in bud" opacities (*red arrows*) and bronchiectasis (*yellow arrow*) consistent with atypical mycobacterial infection. "Tree in bud" refers to small nodules clustered around the centrilobular arteries as well as increased prominence of the centrilobular branching. These findings are consistent with bronchiolitis.

PLEURAL ABNORMALITIES

Figure e34-37

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Large right pneumothorax with near complete collapse of right lung. Pleural reflection highlighted with *red arrows*.

Figure e34-38

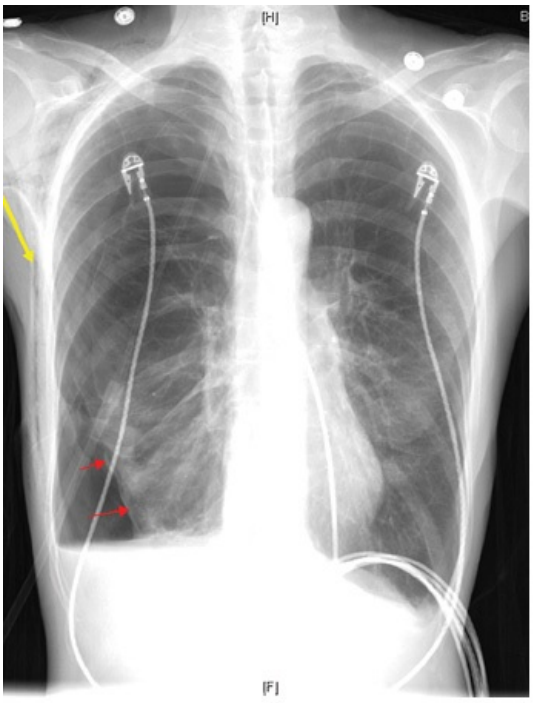

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Basilar pneumothorax with visible pleural reflection (*red arrows*). Also note, patient has subcutaneous emphysema (*yellow arrow*).

Figure e34-39

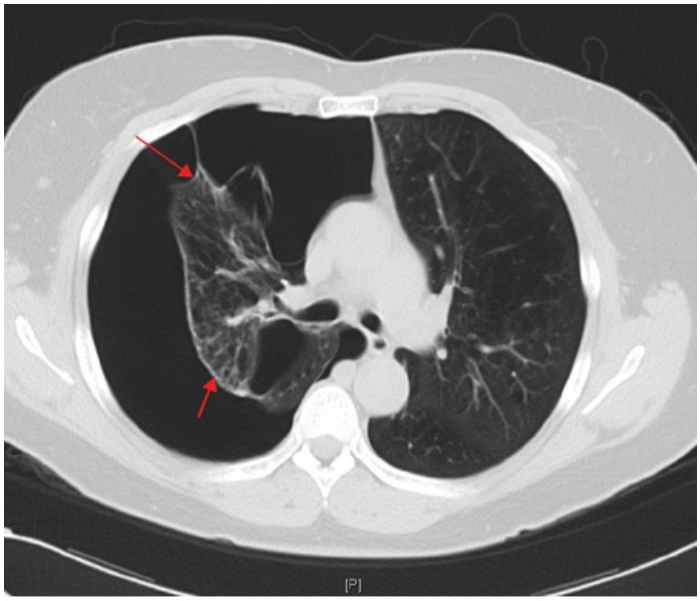

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of large right-sided pneumothorax. Note significant collapse of right lung with adhesion to anterior chest wall. Pleural reflection highlighted with *red arrows.* The patient has severe underlying emphysema.

Figure e34-40

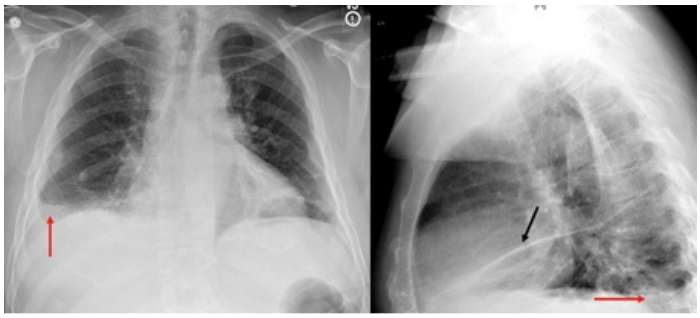

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Small right pleural effusion (*red arrows* highlight blunted right costophrenic angles) with associated pleural thickening. Note fluid in the major fissure (*black arrow*) visible on the lateral film as well as the meniscus of the right pleural effusion.

Figure e34-41

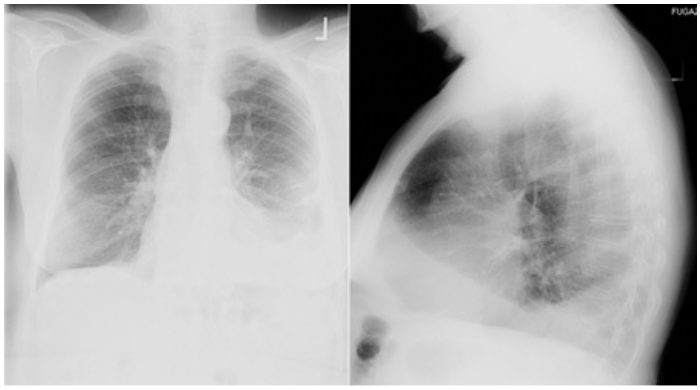

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left pleural effusion with clear meniscus seen on both PA and lateral chest radiographs.

NODULES AND MASSES

Figure e34-42

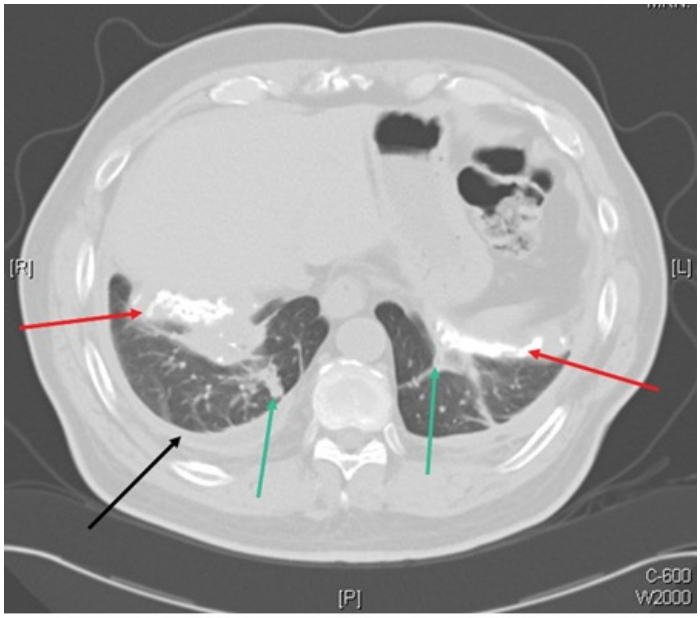

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Asbestosis. Note calcified pleural plaques (*red arrows*), pleural thickening (*black arrow*), and subpleural atelectasis (*green arrows*).

Figure e34-43

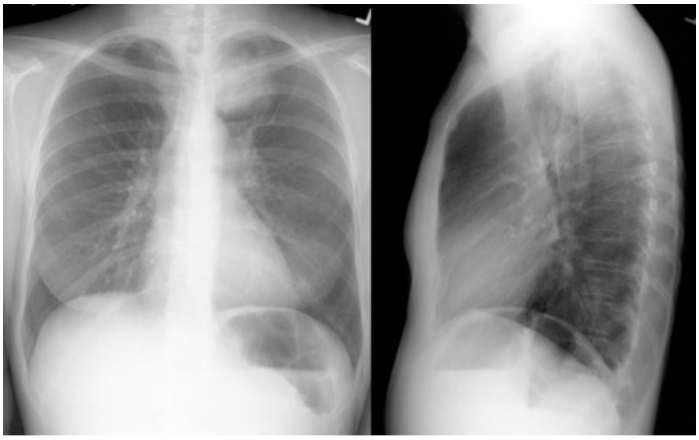

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left upper lobe mass, which biopsy revealed to be squamous cell carcinoma.

Figure e34-44

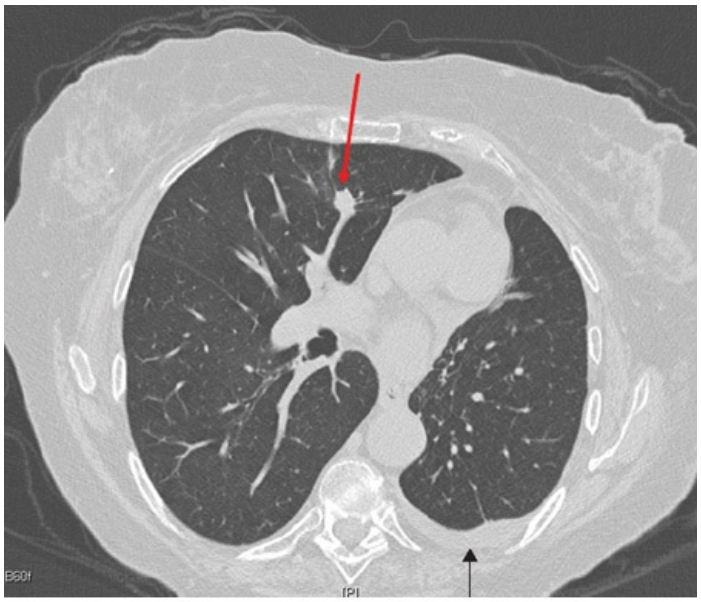

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Solitary pulmonary nodule on the right (*red arrow*) with a spiculated pattern concerning for lung cancer. Note also that the patient is status-post left upper lobectomy with resultant volume loss and associated effusion (*black arrow*).

Figure e34-45

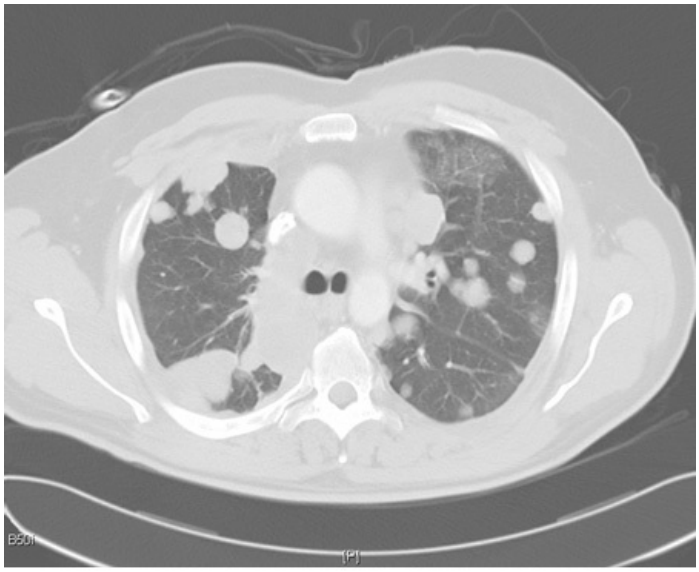

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Metastatic sarcoma. Note the multiple, well-circumscribed nodules of different size.

Figure e34-46

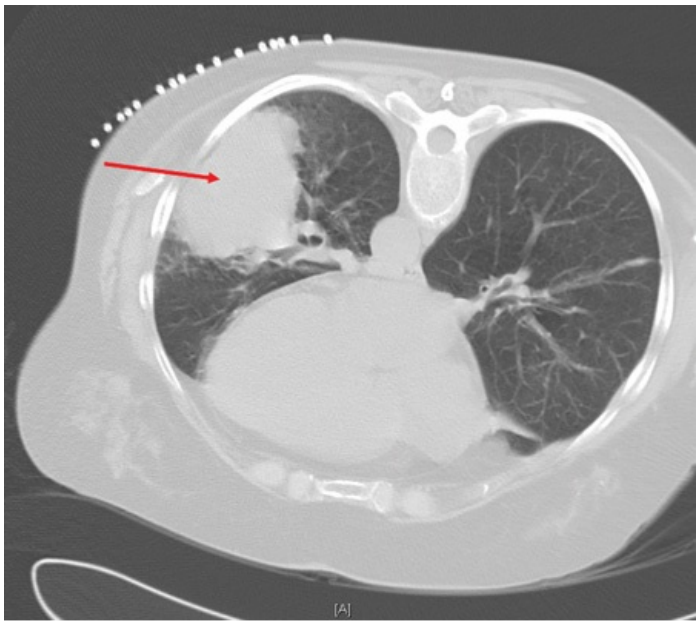

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left lower lobe lung mass (*red arrow*) abutting pleura. Biopsy demonstrated small cell lung cancer.

Figure e34-47

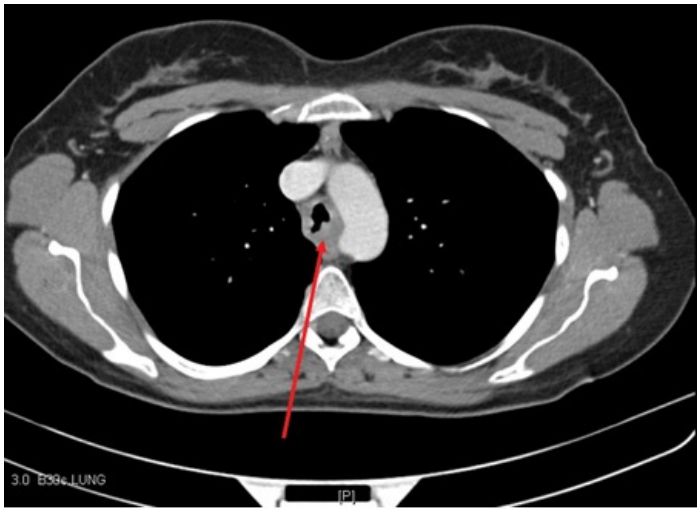

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of soft tissue mass encircling the trachea (*red arrow*) and invading tracheal lumen. Biopsy demonstrated adenoid cystic carcinoma (cylindroma).

PULMONARY VASCULAR ABNORMALITIES

Figure e34-48

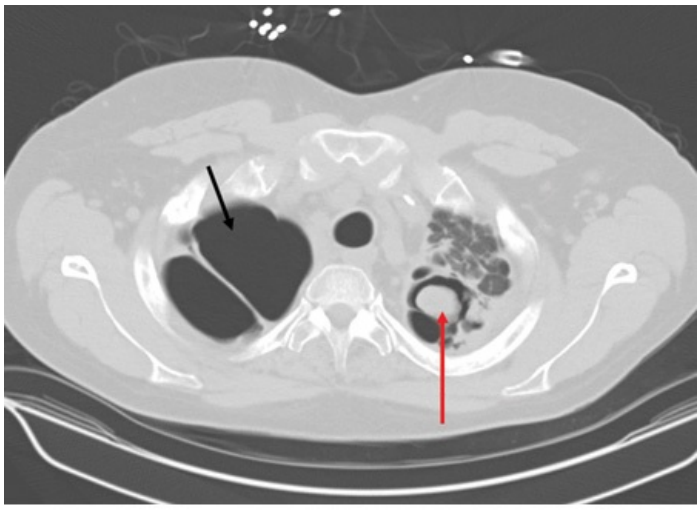

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mycetoma. Fungal ball (*red arrow*) growing in preexisting cavity on the left. Right upper lobe has a large bulla (*black arrow*).

Figure e34-49

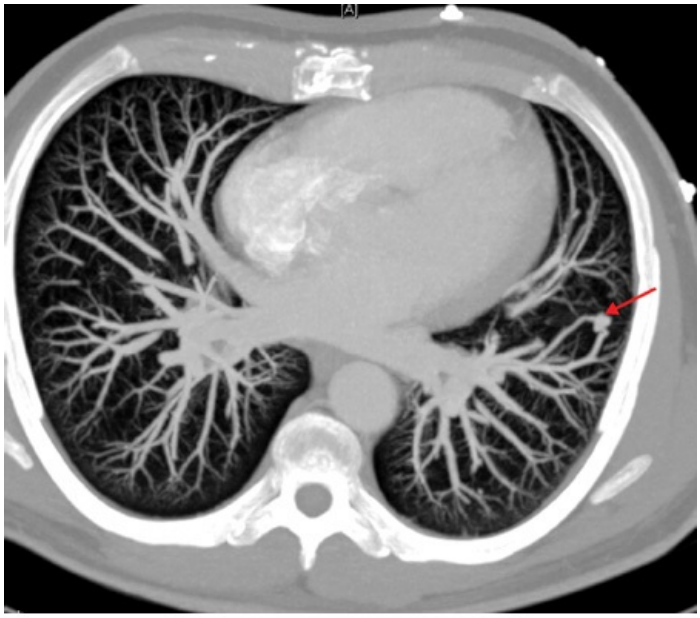

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pulmonary arteriovenous malformation (AVM) demonstrated on reformatted CT angiogram (*red arrow*).

Figure e34-50

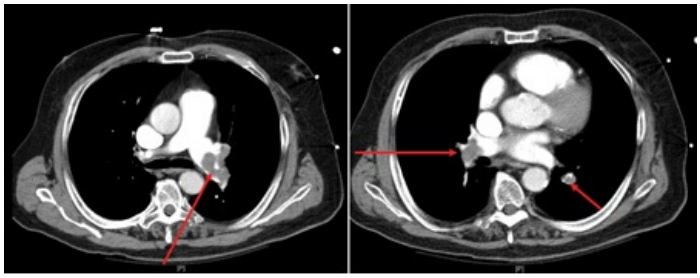

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Large bilateral pulmonary emboli (intravascular filling defects in contrast scan identified by *red arrows*).

Figure e34-51

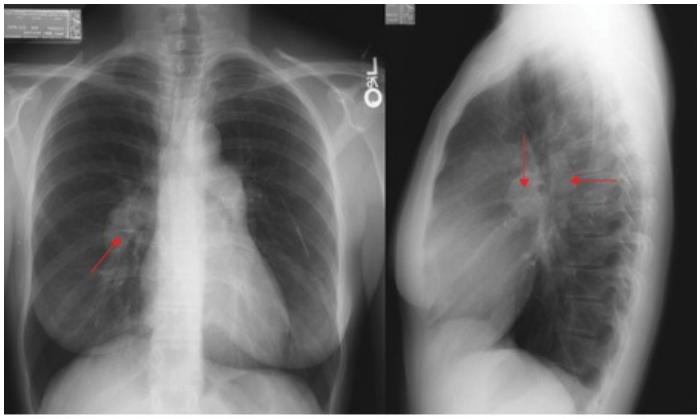

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CXR of a patient with severe pulmonary hypertension. Note the enlarged pulmonary arteries (*red arrows*) visible on both PA and lateral films.

Figure e34-52

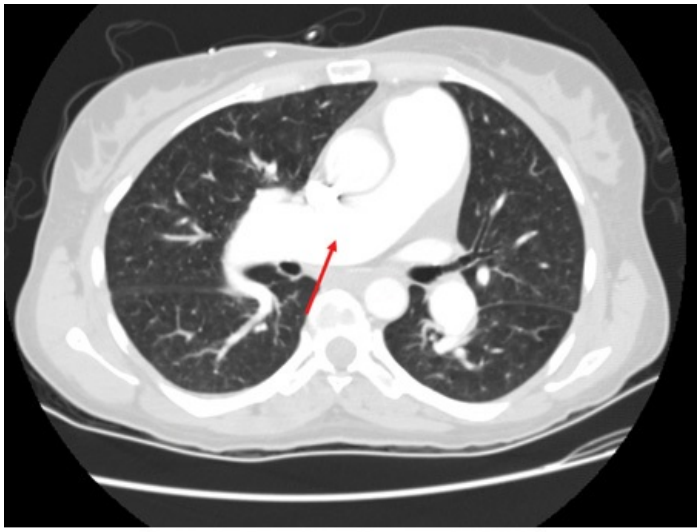

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of the same patient as in Fig. e34-51. Note the markedly enlarged pulmonary arteries (*red arrow*).

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ASTHMA: INTRODUCTION

Asthma is a syndrome characterized by airflow obstruction that varies markedly, both spontaneously and with treatment. Asthmatics harbor a special type of inflammation in the airways that makes them more responsive than nonasthmatics to a wide range of triggers, leading to excessive narrowing with consequent reduced airflow and symptomatic wheezing and dyspnea. Narrowing of the airways is usually reversible, but in some patients with chronic asthma there may be an element of irreversible airflow obstruction. The increasing global prevalence of asthma, the large burden it now imposes on patients, and the high health care costs have led to extensive research into its mechanisms and treatment.

PREVALENCE

Asthma is one of the most common chronic diseases globally and currently affects approximately 300 million people worldwide. The prevalence of asthma has risen in affluent countries over the last 30 years but now appears to have stabilized, with approximately 10–12% of adults and 15% of children affected by the disease. In developing countries where the prevalence of asthma had been much lower, there is a rising prevalence, which is associated with increased urbanization. The prevalence of atopy and other allergic diseases has also increased over the same time, suggesting that the reasons for the increase are likely to be systemic rather than confined to the lungs. This epidemiologic observation suggests that there is a maximum number of individuals in the community, who are likely to be affected by asthma, most likely by genetic predisposition. Most patients with asthma in affluent countries are atopic, with allergic sensitization to the house dust mite *Dermatophagoides pteronyssinus* and other environmental allergens.

Because asthma is both common and frequently complicated by the effects of smoking on the lungs, it is difficult to be certain about the natural history of the disease in adults. Asthma can present at any age, with a peak age of 3 years. In childhood, twice as many males as females are asthmatic, but by adulthood the sex ratio has equalized. The commonly held belief that children "grow out of their asthma" is justified to some extent. Long-term studies that have followed children until they reach the age of 40 years suggest that many with asthma become asymptomatic during adolescence but that asthma returns in some during adult life, particularly in those with persistent symptoms and severe asthma. Adults with asthma, including those with onset during adulthood, rarely become permanently asymptomatic. The severity of asthma does not vary significantly within a given patient; those with mild asthma rarely progress to more severe disease, whereas those with severe asthma usually have severe disease at the onset.

Deaths from asthma are uncommon, and in many affluent countries have been steadily declining over the last decade. A rise in asthma mortality seen in several countries during the 1960s was associated with increased use of short-acting β_2-adrenergic agonists (as rescue therapy), but there is now compelling evidence that the more widespread use of inhaled corticosteroids (ICS) in patients with persistent asthma is responsible for the decrease in mortality in recent years. Major risk factors for asthma deaths are poorly controlled disease with frequent use of bronchodilator inhalers, lack of corticosteroid therapy, and previous admissions to hospital with near-fatal asthma.

It has proved difficult to agree on a definition of asthma, but there is good agreement on the description of the clinical syndrome and disease pathology. Until the etiologic mechanisms of the disease are better understood, it will be difficult to provide an accurate definition.

ETIOLOGY

Asthma is a heterogeneous disease with interplay between genetic and environmental factors. Several risk factors have been implicated (**Table 254-1**).

Table 254-1 Risk Factors and Triggers Involved in Asthma

Endogenous Factors	Environmental Factors
Genetic predisposition	Indoor allergens
Atopy	Outdoor allergens
Airway hyperresponsiveness	Occupational sensitizers

Gender	Passive smoking
Ethnicity?	Respiratory infections
Obesity?	
Early viral infections?	
Triggers	
Allergens	
Upper respiratory tract viral infections	
Exercise and hyperventilation	
Cold air	
Sulfur dioxide and irritant gases	
Drugs (β-blockers, aspirin)	
Stress	
Irritants (household sprays, paint fumes)	

Atopy

Atopy is the major risk factor for asthma, and nonatopic individuals have a very low risk of developing asthma. Patients with asthma commonly suffer from other atopic diseases, particularly allergic rhinitis, which may be found in over 80% of asthmatic patients, and atopic dermatitis (eczema). Atopy may be found in 40–50% of the population in affluent countries, with only a proportion of atopic individuals becoming asthmatic. This observation suggests that some other environmental or genetic factor(s) predispose to the development of asthma in atopic individuals. The allergens that lead to sensitization are usually proteins that have protease activity, and the most common allergens are derived from house dust mites, cat and dog fur, cockroaches (in inner cities), grass and tree pollens, and rodents (in laboratory workers). Atopy is due to the genetically determined production of specific IgE antibody, with many patients showing a family history of allergic diseases.

Intrinsic Asthma

A minority of asthmatic patients (approximately 10%) have negative skin tests to common inhalant allergens and normal serum concentrations of IgE. These patients, with nonatopic or intrinsic asthma, usually show later onset of disease (adult-onset asthma), commonly have concomitant nasal polyps, and may be aspirin-sensitive. They usually have more severe, persistent asthma. Little is understood about mechanism, but the immunopathology in bronchial biopsies and sputum appears to be identical to that found in atopic asthma. There is recent evidence for increased local production of IgE in the airways, suggesting that there may be common IgE-mediated mechanisms; staphylococcal enterotoxins, which serve as "superantigens," have been implicated.

Infections

Although viral infections are common as triggers of asthma exacerbations, it is uncertain whether they play a role in etiology. There is some association between respiratory syncytial virus infection in infancy and the development of asthma, but the specific pathogenesis is difficult to elucidate, as this infection is very common in children. More recently, atypical bacteria such as *Mycoplasma* and *Chlamydophila,* have been implicated in the mechanism of severe asthma, but thus far, the evidence is not very convincing of a true association.

Genetic Considerations

The familial association of asthma and a high degree of concordance for asthma in identical twins indicate a genetic predisposition to the disease; however, whether or not the genes predisposing to asthma are similar or in addition to those predisposing to atopy is not yet clear. It now seems likely that different genes may also contribute to asthma specifically, and there is increasing evidence that the severity of asthma is also genetically determined. Genetic screens with classical linkage analysis and single-nucleotide polymorphisms of various candidate genes indicate that asthma is polygenic, with each gene identified having a small effect that is often not replicated in different populations. This observation suggests that the interaction of many genes is important, and these may differ in different populations. The most consistent findings have been associations with polymorphisms of genes on chromosome 5q, including the T helper 2 (T_H2) cells interleukin (IL)-4, IL-5, IL-9, and IL-13, which are associated with atopy. There is increasing evidence for a complex interaction between genetic polymorphisms and environmental factors that will require very large population studies to unravel. Novel genes that have been associated with asthma, including *ADAM-33, DPP-10*, and *GPRA*, have also been identified by positional cloning, but their function in disease pathogenesis is not yet clear. Recent genome-wide association studies have identified further novel genes, although, again, their functional role is not yet clear. Genetic polymorphisms may also be important in determining the response to asthma therapy. For example, the Arg-Gly-16 variant in the β_2-receptor has been associated with reduced response to β_2-agonists, and repeats of an Sp1 recognition sequence in the

promoter region of 5-lipoxygenase may affect the response to antileukotrienes. However, these effects are small and inconsistent and do not yet have any implications for asthma therapy.

Environmental Factors

It is likely that environmental factors in early life determine which atopic individuals become asthmatic. The increasing prevalence of asthma, particularly in developing countries, over the last few decades also indicates the importance of environmental mechanisms interacting with a genetic predisposition.

HYGIENE HYPOTHESIS

The observation that allergic sensitization and asthma were less common in children with older siblings first suggested that lower levels of infection may be a factor in affluent societies that increase the risks of asthma. This "hygiene hypothesis" proposes that lack of infections in early childhood preserves the T_H2 cell bias at birth, whereas exposure to infections and endotoxin results in a shift toward a predominant protective T_H1 immune response. Children brought up on farms who are exposed to a high level of endotoxin are less likely to develop allergic sensitization than children raised on dairy farms. Intestinal parasite infection may also be associated with a reduced risk of asthma. While there is considerable epidemiologic support for the hygiene hypothesis, it cannot account for the parallel increase in T_H1-driven diseases such as diabetes mellitus over the same period.

DIET

The role of dietary factors is controversial. Observational studies have shown that diets low in antioxidants such as vitamin C and vitamin A, magnesium, selenium, and omega-3 polyunsaturated fats (fish oil) or high in sodium and omega-6 polyunsaturates are associated with an increased risk of asthma. Vitamin D deficiency may also predispose to the development of asthma. However, interventional studies with supplementary diets have not supported an important role for these dietary factors. Obesity is also an independent risk factor for asthma, particularly in women, but the mechanisms are thus far unknown.

AIR POLLUTION

Air pollutants such as sulfur dioxide, ozone, and diesel particulates, may trigger asthma symptoms, but the role of different air pollutants in the etiology of the disease is much less certain. Most evidence argues against an important role for air pollution as asthma is no more prevalent in cities with a high ambient level of traffic pollution than in rural areas with low levels of pollution. Asthma had a much lower prevalence in East Germany compared to West Germany despite a much higher level of air pollution, but since reunification these differences have decreased as eastern Germany has become more affluent. Indoor air pollution may be more important with exposure to nitrogen oxides from cooking stoves and exposure to passive cigarette smoke. There is some evidence that maternal smoking is a risk factor for asthma, but it is difficult to dissociate this association from an increased risk of respiratory infections.

ALLERGENS

Inhaled allergens are common triggers of asthma symptoms and have also been implicated in allergic sensitization. Exposure to house dust mites in early childhood is a risk factor for allergic sensitization and asthma, but rigorous allergen avoidance has not shown any evidence for a reduced risk of developing asthma. The increase in house dust mites in centrally heated poorly ventilated homes with fitted carpets has been implicated in the increasing prevalence of asthma in affluent countries. Domestic pets, particularly cats, have also been associated with allergic sensitization, but early exposure to cats in the home may be protective through the induction of tolerance.

OCCUPATIONAL EXPOSURE

Occupational asthma is relatively common and may affect up to 10% of young adults. Over 200 sensitizing agents have been identified. Chemicals such as toluene diisocyanate and trimellitic anhydride, may lead to sensitization independent of atopy. Individuals may also be exposed to allergens in the workplace such as small animal allergens in laboratory workers and fungal amylase in wheat flour in bakers. Occupational asthma may be suspected when symptoms improve during weekends and holidays.

Other Factors

Several other factors have been implicated in the etiology of asthma, including lower maternal age, duration of breast-feeding, prematurity and low birthweight, and inactivity, but are unlikely to contribute to the recent global increase in asthma prevalence. There is also an association with acetaminophen (paracetamol) consumption in childhood, which remains unexplained.

PATHOGENESIS

Asthma is associated with a specific chronic inflammation of the mucosa of the lower airways. One of the main aims of treatment is to reduce this inflammation.

Pathology

The pathology of asthma has been revealed through examining the lungs at autopsy of patients who have died of asthma and from bronchial biopsies in patients with usually mild asthma. The airway mucosa is infiltrated with activated eosinophils and T lymphocytes, and there is activation of mucosal mast cells. The degree of inflammation is poorly related to disease severity and may be found in atopic patients without asthma symptoms. The inflammation is reduced by treatment with ICS. A characteristic finding is thickening of the basement membrane due to subepithelial collagen deposition. This feature is also found in patients with eosinophilic bronchitis presenting as cough who do not have asthma and is, therefore, likely to be a marker of eosinophilic inflammation in the airway as eosinophils release fibrogenic factors. The epithelium is often shed or friable, with reduced attachments to the airway wall and increased numbers of epithelial cells in the lumen. The airway wall itself may be thickened and edematous, particularly in fatal asthma. Another common finding in fatal asthma is occlusion of the airway lumen by a mucous plug, which is comprised of mucous glycoproteins secreted from goblet cells and plasma proteins from leaky bronchial vessels (**Fig. 254-1**). There is also vasodilation and increased numbers of blood vessels (angiogenesis). Direct observation by bronchoscopy indicates that the airways may be narrowed, erythematous, and edematous. The pathology of asthma is remarkably uniform in different types of asthma, including atopic, nonatopic, occupational, aspirin-sensitive, and pediatric asthma. These pathologic changes are found in all airways, but do not extend to the lung parenchyma; peripheral airway inflammation is found particularly in patients with severe asthma. The involvement of airways may be patchy and this is consistent with bronchographic findings of uneven narrowing of the airways.

Figure 254-1

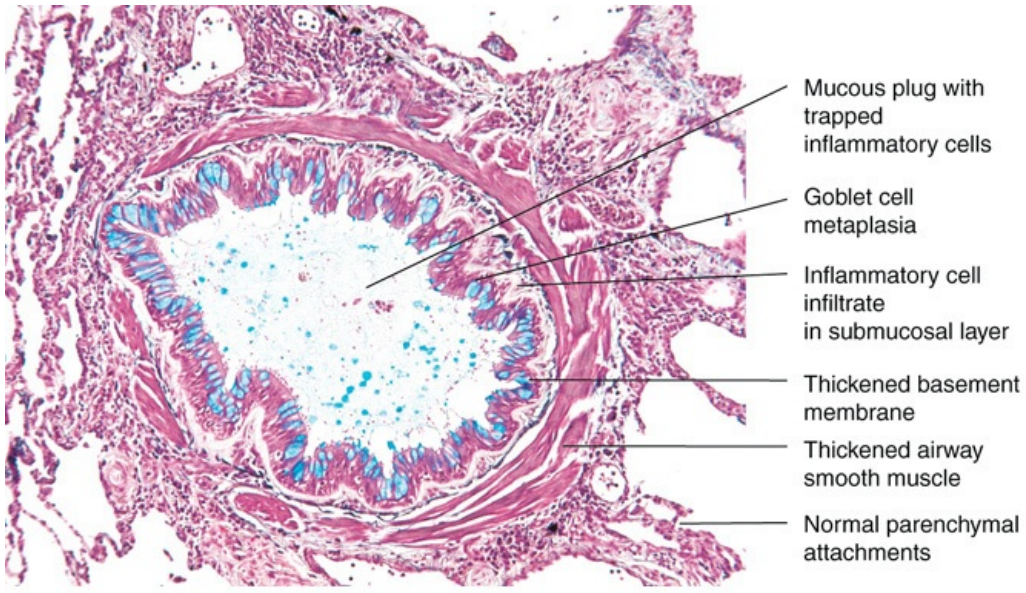

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histopathology of a small airway in fatal asthma. The lumen is occluded with a mucous plug, there is goblet cell metaplasia, and the airway wall is thickened, with an increase in basement membrane thickness and airway smooth muscle. *(Courtesy of Dr. J. Hogg, University of British Colombia.)*

Inflammation

There is inflammation in the respiratory mucosa from the trachea to terminal bronchioles, but with a predominance in the bronchi (cartilaginous airways). Considerable research has identified the major cellular components of inflammation, but it is still uncertain how inflammatory cells interact and how inflammation translates into the symptoms of asthma (**Fig. 254-2**). There is good evidence that the specific pattern of airway inflammation in asthma is associated with airway hyperresponsiveness (AHR), the physiologic abnormality of asthma, which is correlated with variable airflow obstruction. The pattern of inflammation in asthma is characteristic of allergic diseases, with similar inflammatory cells seen in the nasal mucosa in rhinitis. However, an indistinguishable pattern of inflammation is found in intrinsic asthma, and this may reflect local rather than systemic IgE production. Although most attention has focused on the acute inflammatory changes seen in asthma, this is a chronic condition, with inflammation persisting over many years in most patients. The mechanisms involved in persistence of inflammation in asthma are still poorly understood. Superimposed on this chronic inflammatory state are acute inflammatory episodes, which correspond to exacerbations of asthma. Many inflammatory cells are known to be involved in asthma with no key cell that is predominant (**Fig. 254-3**).

Figure 254-2

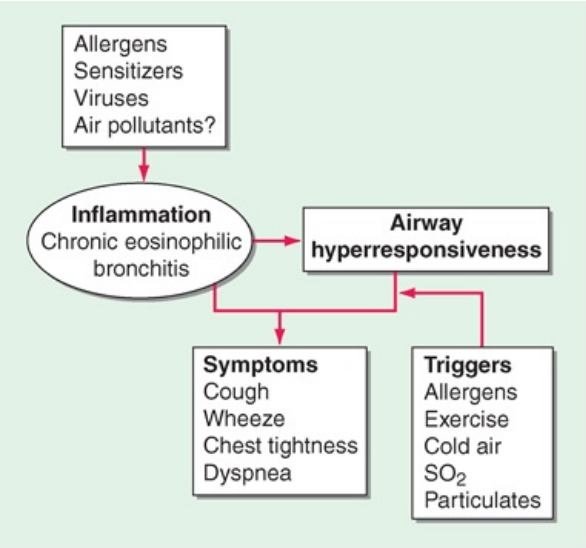

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Inflammation in the airways of asthmatic patients leads to airway hyperresponsiveness and symptoms. SO_2, sulfur dioxide.

Figure 254-3

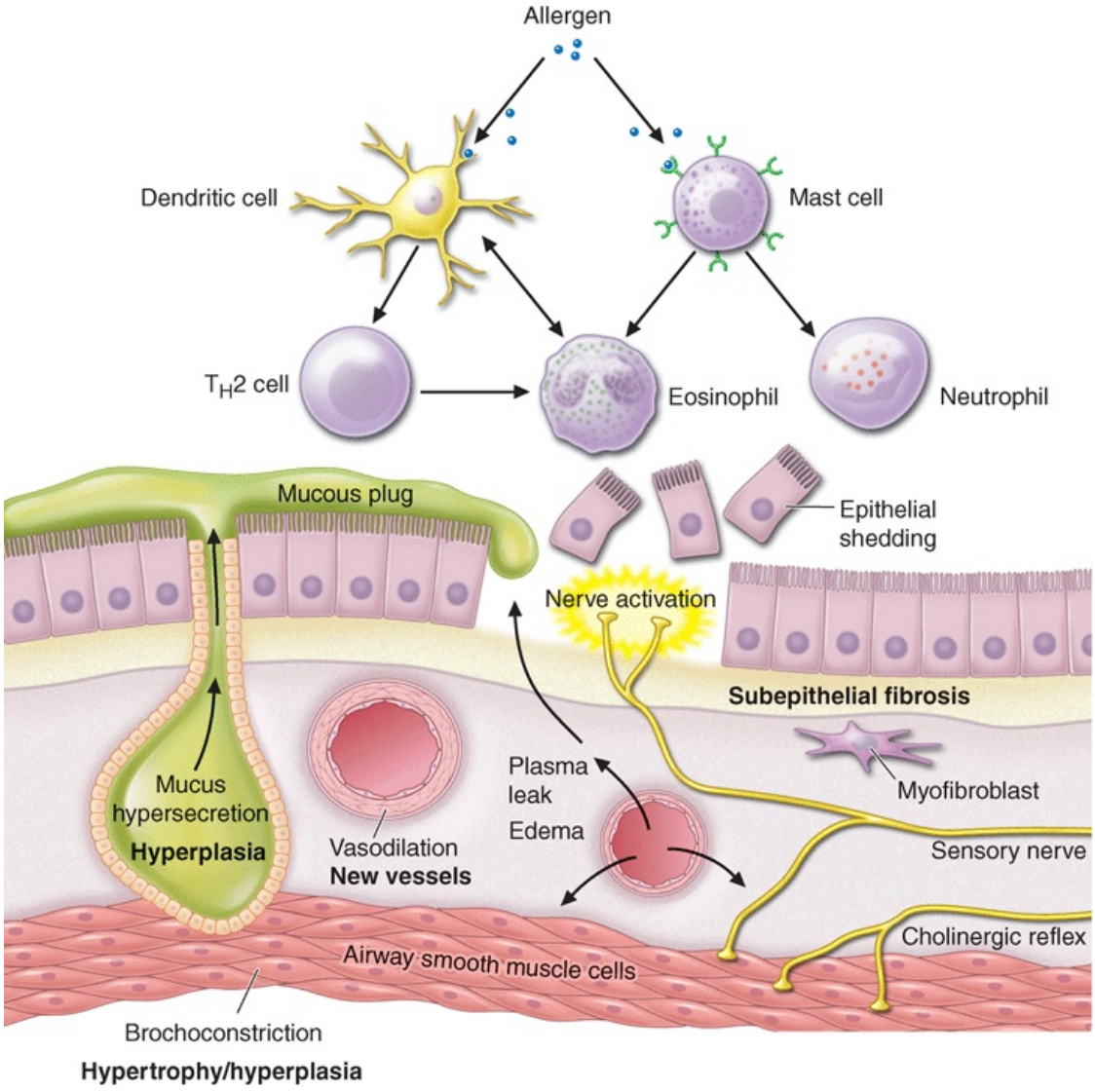

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The pathophysiology of asthma is complex with participation of several interacting inflammatory cells, which result in acute and chronic inflammatory effects on the airway.

MAST CELLS

Mast cells are important in initiating the acute bronchoconstrictor responses to allergens and several other indirectly acting stimuli such as exercise and hyperventilation (via osmolality or thermal changes), as well as fog. Activated mast cells are found at the airway surface in asthma patients and also in the airway smooth-muscle layer, whereas this is not seen in normal subjects or patients with eosinophilic bronchitis. Mast cells are activated by allergens through an IgE-dependent mechanism, and binding of specific IgE to mast cells renders them more sensitive to activation. The importance of IgE in the pathophysiology of asthma has been highlighted by clinical studies with humanized anti-IgE antibodies, which inhibit IgE-mediated effects, reduce asthma symptoms, and reduce exacerbations. There are, however, uncertainties about the role of mast cells in more chronic allergic inflammatory events. Mast cells release several bronchoconstrictor mediators, including histamine, prostaglandin D_2, and cysteinyl-leukotrienes, but also several cytokines, chemokines, growth factors, and neurotrophins.

MACROPHAGES AND DENDRITIC CELLS

Macrophages, which are derived from blood monocytes, may traffic into the airways in asthma and may be activated by allergens via low affinity IgE receptors ($Fc_{\varepsilon}RII$). Macrophages have the capacity to initiate a type of inflammatory response via the release of a certain pattern of cytokines, but these cells also release anti-inflammatory mediators (e.g., IL-10) and, thus, their roles in asthma are uncertain. Dendritic cells are specialized macrophage-like cells in the airway epithelium, which are the major antigen-presenting cells. Dendritic cells take up allergens, process them to peptides, and migrate to local lymph nodes where they present the allergenic peptides to uncommitted T-lymphocytes to program the production of allergen-specific T cells. Immature dendritic cells in the respiratory tract promote T_H2 cell differentiation and require cytokines such as IL-12 and tumor necrosis factor α (TNF-

α), to promote the normally preponderant T_H1 response. The cytokine thymic stromal lymphopoietin (TSLP) released from epithelial cells in asthmatic patients instructs dendritic cells to release chemokines that attract T_H2 cells into the airways.

EOSINOPHILS

Eosinophil infiltration is a characteristic feature of asthmatic airways. Allergen inhalation results in a marked increase in activated eosinophils in the airways at the time of the late reaction. Eosinophils are linked to the development of AHR through the release of basic proteins and oxygen-derived free radicals. Eosinophil recruitment involves adhesion of eosinophils to vascular endothelial cells in the airway circulation due to interaction between adhesion molecules, migration into the submucosa under the direction of chemokines, and their subsequent activation and prolonged survival. Blocking antibodies to IL-5 causes a profound and prolonged reduction in circulating and sputum eosinophils, but is not associated with reduced AHR or asthma symptoms, although in selected patients with steroid-resistant airway eosinophils, there is a reduction in exacerbations. Eosinophilic inflammation is also found in patients with chronic cough (eosinophilic bronchitis) who do not have AHR or clinical features of asthma. Increasing evidence suggests that eosinophils may be important in release of growth factors involved in airway remodeling, in exacerbations but not in AHR.

NEUTROPHILS

Increased numbers of activated neutrophils are found in sputum and airways of some patients with severe asthma and during exacerbations, although there is a proportion of patients even with mild or moderate asthma who have a predominance of neutrophils. The roles of neutrophils in asthma that are resistant to the anti-inflammatory effects of corticosteroids are currently unknown.

T LYMPHOCYTES

T lymphocytes play a very important role in coordinating the inflammatory response in asthma through the release of specific patterns of cytokines, resulting in the recruitment and survival of eosinophils and in the maintenance of a mast cell population in the airways. The naïve immune system and the immune system of asthmatics are skewed to express the T_H2 phenotype, whereas in normal airways T_H1 cells predominate. T_H2 cells, through the release of IL-5, are associated with eosinophilic inflammation and, through the release of IL-4 and IL-13, are associated with increased IgE formation. Recently, bronchial biopsies have demonstrated a preponderance of natural killer CD4+ T lymphocytes that express high levels of IL-4. Regulatory T cells play an important role in determining the expression of other T cells, and there is evidence for a reduction in a certain subset of regulatory T cells (CD4+CD25+) in asthma that is associated with increased T_H2 cells.

STRUCTURAL CELLS

Structural cells of the airways, including epithelial cells, fibroblasts, and airway smooth-muscle cells, are also important sources of inflammatory mediators such as cytokines and lipid mediators, in asthma. Indeed, because structural cells far outnumber inflammatory cells, they may become the major sources of mediators driving chronic inflammation in asthmatic airways. In addition, epithelial cells may have key roles in translating inhaled environmental signals into an airway inflammatory response, and are probably major target cells for ICS.

Inflammatory Mediators

Many different mediators have been implicated in asthma, and they may have a variety of effects on the airways that could account for the pathologic features of asthma (**Fig. 254-4**). Mediators such as histamine, prostaglandin D_2, and cysteinyl-leukotrienes contract airway smooth muscle, increase microvascular leakage, increase airway mucus secretion, and attract other inflammatory cells. Because each mediator has many effects, the role of individual mediators in the pathophysiology of asthma is not yet clear. Although the multiplicity of mediators makes it unlikely that preventing the synthesis or action of a single mediator will have a major impact in clinical asthma, recent clinical studies with antileukotrienes suggest that cysteinyl-leukotrienes have clinically important effects.

Figure 254-4

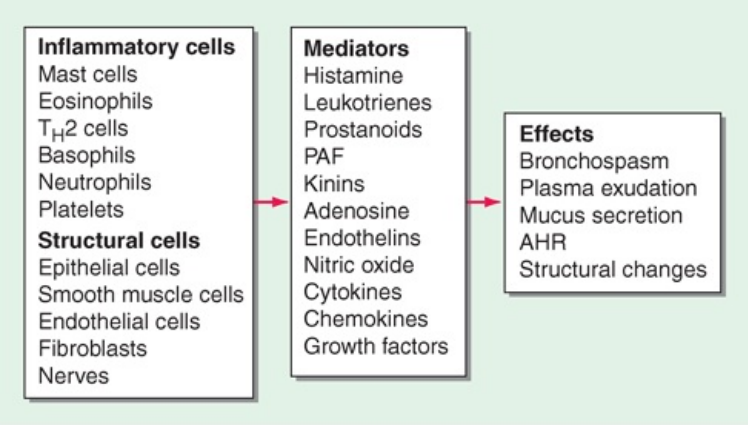

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Many cells and mediators are involved in asthma and lead to several effects on the airways.

CYTOKINES

Multiple cytokines regulate the chronic inflammation of asthma. The TH2 cytokines IL-4, IL-5, and IL-13 mediate allergic inflammation, whereas proinflammatory cytokines such as TNF-α and IL-1β, amplify the inflammatory response and play a role in more severe disease. Thymic stromal lymphopoietin is an upstream cytokine released from epithelial cells of asthmatics that orchestrates the release of chemokines that selectively attract T_H2 cells. Some cytokines such as IL-10 and IL-12 are anti-inflammatory and may be deficient in asthma.

CHEMOKINES

Chemokines are involved in attracting inflammatory cells from the bronchial circulation into the airways. Eotaxin (CCL11) is selectively attractant to eosinophils via CCR3 and is expressed by epithelial cells of asthmatics, whereas CCL17 (TARC) and CCL22 (MDC) from epithelial cells attract T_H2 cells via CCR4 (**Fig. 254-5**).

Figure 254-5

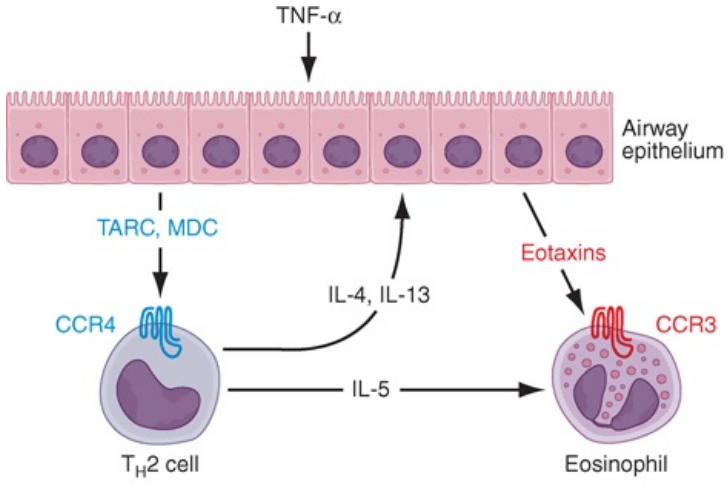

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chemokines in asthma. Tumor necrosis factor α (TNF-α) and other triggers of airway epithelial cells release thymus and activationregulated chemokine (TARC, CCL17) and macrophage-derived chemokine (MDC, CCL22) from epithelial cells that attract T_H2 cells via activation of their CCR4 receptors. These promote eosinophilic inflammation directly through the release of interleukin (IL)-5 and indirectly via the release of IL-4 and IL-13, which induce eotaxin (CCL11) formation in airway epithelial cells.

OXIDATIVE STRESS

There is increased oxidative stress in asthma as activated inflammatory cells such as macrophages and eosinophils that produce reactive oxygen species. Evidence for increased oxidative stress in asthma is provided by the increased concentrations of 8-isoprostane (a product of oxidized arachidonic acid) in exhaled breath condensates and increased ethane (a product of lipid peroxidation)

in the expired air of asthmatic patients. Increased oxidative stress is related to disease severity, may amplify the inflammatory response, and may reduce responsiveness to corticosteroids.

NITRIC OXIDE

Nitric oxide (NO) is produced by several cells in the airway by NO synthases, particularly airway epithelial cells and macrophages. The level of NO in the expired air of patients with asthma is higher than normal and is related to the eosinophilic inflammation. Increased NO may contribute to the bronchial vasodilation observed in asthma. Exhaled NO is increasingly used in the diagnosis and monitoring of asthmatic inflammation, although it is not yet used routinely in clinical practice.

TRANSCRIPTION FACTORS

Proinflammatory transcription factors such as nuclear factor-κB (NF-κB) and activator protein-1, are activated in asthmatic airways and orchestrate the expression of multiple inflammatory genes. More specific transcription factors that are involved include nuclear factor of activated T cells and GATA-3, which regulate the expression of T_H2 cytokines in T cells.

Effects of Inflammation

The chronic inflammatory response has several effects on the target cells of the airways, resulting in the characteristic pathophysiologic changes associated with asthma. Asthma may be regarded as a disease with continuous inflammation and repair proceeding simultaneously. Important advances continue to be made in our understanding of these changes, but, despite these new insights, the relationship between chronic inflammatory processes and asthma symptoms is often not clear.

AIRWAY EPITHELIUM

Airway epithelial shedding may be important in contributing to AHR and may explain how several mechanisms, such as ozone exposure, virus infections, chemical sensitizers, and allergen exposure, can lead to its development, as all of these stimuli may lead to epithelial disruption. Epithelial damage may contribute to AHR in a number of ways, including loss of its barrier function to allow penetration of allergens; loss of enzymes (such as neutral endopeptidase) that degrade certain peptide inflammatory mediators; loss of a relaxant factor (so called epithelial-derived relaxant factor); and exposure of sensory nerves, which may lead to reflex neural effects on the airway.

FIBROSIS

In all asthmatic patients, the basement membrane is apparently thickened due to subepithelial fibrosis with deposition of types III and V collagen below the true basement membrane and is associated with eosinophil infiltration, presumably through the release of profibrotic mediators such as transforming growth factor-β. Mechanical manipulations can alter the phenotype of airway epithelial cells in a profibrotic fashion. In more severe patients, there is also fibrosis within the airway wall, which may contribute to irreversible narrowing of the airways.

AIRWAY SMOOTH MUSCLE

There is still debate about the role of abnormalities in airway smooth muscle in asthmatic airways. In vitro airway smooth muscle from asthmatic patients usually shows no increased responsiveness to constrictors. Reduced responsiveness to β-agonists has also been reported in postmortem or surgically removed bronchi from asthmatics, although the number of β-receptors is not reduced, suggesting that β-receptors have been uncoupled. These abnormalities of airway smooth-muscle may be secondary to the chronic inflammatory process. Inflammatory mediators may modulate the ion channels that serve to regulate the resting membrane potential of airway smooth-muscle cells, thus altering the level of excitability of these cells. In asthmatic airways there is also a characteristic hypertrophy and hyperplasia of airway smooth muscle, which is presumably the result of stimulation of airway smooth-muscle cells by various growth factors such as platelet-derived growth factor (PDGF) or endothelin-1 released from inflammatory or epithelial cells.

VASCULAR RESPONSES

There is increased airway mucosal blood flow in asthma. The bronchial circulation may play an important role in regulating airway caliber, since an increase in the vascular volume may contribute to airway narrowing. Increased airway blood flow may be important in removing inflammatory mediators from the airway, and may play a role in the development of exercise-induced asthma. There is an increase in the number of blood vessels in asthmatic airways as a result of angiogenesis in response to growth factors, particularly vascular-endothelial growth factor. Microvascular leakage from postcapillary venules in response to inflammatory mediators is observed in asthma, resulting in airway edema and plasma exudation into the airway lumen.

MUCUS HYPERSECRETION

Increased mucus secretion contributes to the viscid mucous plugs that occlude asthmatic airways, particularly in fatal asthma. There is evidence for hyperplasia of submucosal glands that are confined to large airways and of increased numbers of epithelial goblet cells. IL-4 and IL-13 induce mucus hypersecretion in experimental models of asthma.

NEURAL EFFECTS

Various defects in autonomic neural control may contribute to AHR in asthma, but these are likely to be secondary to the disease, rather than primary defects. Cholinergic pathways, through the release of acetylcholine acting on muscarinic receptors, cause bronchoconstriction and may be activated reflexly in asthma. Inflammatory mediators may activate sensory nerves, resulting in reflex cholinergic bronchoconstriction or release of inflammatory neuropeptides. Inflammatory products may also sensitize sensory nerve endings in the airway epithelium such that the nerves become hyperalgesic. Neurotrophins, which may be released from various cell types in airways, including epithelial cells and mast cells, may cause proliferation and sensitization of airway sensory nerves. Airway nerves may also release neurotransmitters, such as substance P, which have inflammatory effects.

Airway Remodeling

Several changes in the structure of the airway are characteristically found in asthma, and these may lead to irreversible narrowing of the airways. Population studies have shown a greater decline in lung function over time than in normal subjects; however, most patients with asthma preserve normal or near-normal lung function throughout life if appropriately treated. This observation suggests that the accelerated decline in lung function occurs in a smaller proportion of asthmatics, and these are usually patients with more severe disease. There is some evidence that the early use of ICS may reduce the decline in lung function. The characteristic structural changes are increased airway smooth muscle, fibrosis, angiogenesis, and mucus hyperplasia.

ASTHMA TRIGGERS

Several stimuli trigger airway narrowing, wheezing, and dyspnea in asthmatic patients. While the previous view held that these should be avoided, it is now seen as evidence for poor control and an indicator of the need to increase controller (preventive) therapy.

Allergens

Inhaled allergens activate mast cells with bound IgE directly leading to the immediate release of bronchoconstrictor mediators, resulting in the early response that is reversed by bronchodilators. Often, experimental allergen challenge is followed by a late response when there is airway edema and an acute inflammatory response with increased eosinophils and neutrophils that are not very reversible with bronchodilators. The most common allergens to trigger asthma are *Dermatophagoides* species, and environmental exposure leads to low-grade chronic symptoms that are perennial. Other perennial allergens are derived from cats and other domestic pets, as well as cockroaches. Other allergens, including grass pollen, ragweed, tree pollen, and fungal spores, are seasonal. Pollens usually cause allergic rhinitis rather than asthma, but in thunderstorms the pollen grains are disrupted and the particles that may be released can trigger severe asthma exacerbations (thunderstorm asthma).

Virus Infections

Upper respiratory tract virus infections such as rhinovirus, respiratory syncytial virus, and coronavirus are the most common triggers of acute severe exacerbations and may invade epithelial cells of the lower as well as the upper airways. The mechanism whereby these viruses cause exacerbations is poorly understood, but there is an increase in airway inflammation with increased numbers of eosinophils and neutrophils. There is evidence for reduced production of type I interferons by epithelial cells from asthmatic patients, resulting in increased susceptibility to these viral infections and a greater inflammatory response.

Pharmacologic Agents

Several drugs may trigger asthma. Beta-adrenergic blockers commonly acutely worsen asthma, and their use may be fatal. The mechanisms are not clear but are likely mediated through increased cholinergic bronchoconstriction. All beta blockers need to be avoided and even selective β_2 blocker or topical application (e.g., timolol eye drops) may be dangerous. Angiotensin-converting enzyme inhibitors are theoretically detrimental as they inhibit breakdown of kinins, which are bronchoconstrictors; however, they rarely worsen asthma, and the characteristic cough is no more frequent in asthmatics than in nonasthmatics. Aspirin may worsen asthma in some patients (aspirin-sensitive asthma is discussed below under "Special Considerations").

Exercise

Exercise is a common trigger of asthma, particularly in children. The mechanism is linked to hyperventilation, which results in increased osmolality in airway lining fluid and triggers mast cell mediator release, resulting in bronchoconstriction. Exercise-induced asthma (EIA) typically begins after exercise has ended, and recovers spontaneously within about 30 minutes. EIA is worse in cold, dry climates than in hot, humid conditions. It is, therefore, more common in sports such as cross-country running in cold weather, overland skiing, and ice hockey than in swimming. It may be prevented by prior administration of β_2-agonists and antileukotrienes, but is best prevented by regular treatment with ICS, which reduce the population of surface mast cells required for this response.

Physical Factors

Cold air and hyperventilation may trigger asthma through the same mechanisms as exercise. Laughter may also be a trigger. Many patients report worsening of asthma in hot weather and when the weather changes. Some asthmatics become worse when exposed to strong smells or perfumes, but the mechanism of this response is uncertain.

Food

There is little evidence that allergic reactions to food lead to increased asthma symptoms, despite the belief of many patients that their symptoms are triggered by particular food constituents. Exclusion diets are usually unsuccessful at reducing the frequency of episodes. Some foods such as shellfish and nuts may induce anaphylactic reactions that may include wheezing. Patients with aspirin-induced asthma may benefit from a salicylate-free diet, but these are difficult to maintain. Certain food additives may trigger asthma. Metabisulfite, which is used as a food preservative, may trigger asthma through the release of sulfur dioxide gas in the stomach. Tartrazine, a yellow food-coloring agent, was believed to be a trigger for asthma, but there is little convincing evidence for this.

Air Pollution

Increased ambient levels of sulfur dioxide, ozone, and nitrogen oxides are associated with increased asthma symptoms.

Occupational Factors

Several substances found in the workplace may act as sensitizing agents, as discussed above, but may also act as triggers of asthma symptoms. Occupational asthma is characteristically associated with symptoms at work with relief on weekends and holidays. If removed from exposure within the first 6 months of symptoms, there is usually complete recovery. More persistent symptoms lead to irreversible airway changes, and, thus, early detection and avoidance are important.

Hormonal Factors

Some women show premenstrual worsening of asthma, which can occasionally be very severe. The mechanisms are not completely understood, but are related to a fall in progesterone and in severe cases may be improved by treatment with high doses of progesterone or gonadotropin-releasing factors. Thyrotoxicosis and hypothyroidism can both worsen asthma, although the mechanisms are uncertain.

Gastroesophageal Reflux

Gastroesophageal reflux is common in asthmatic patients as it is increased by bronchodilators. Although acid reflux might trigger reflex bronchoconstriction, it rarely causes asthma symptoms, and antireflux therapy fails to reduce asthma symptoms in most patients.

Stress

Many asthmatics report worsening of symptoms with stress. There is no doubt that psychological factors can induce bronchoconstriction through cholinergic reflex pathways. Paradoxically, very severe stress such as bereavement usually does not worsen, and may even improve, asthma symptoms.

PATHOPHYSIOLOGY

Limitation of airflow is due mainly to bronchoconstriction, but airway edema, vascular congestion, and luminal occlusion with exudate may also contribute. This results in a reduction in forced expiratory volume in 1 second (FEV_1), FEV_1/forced vital capacity (FVC) ratio, and peak expiratory flow (PEF), as well as an increase in airway resistance. Early closure of peripheral airway results in lung hyperinflation, (air trapping) and increased residual volume, particularly during acute exacerbations and in severe persistent asthma. In more severe asthma, reduced ventilation and increased pulmonary blood flow result in mismatching of ventilation and perfusion and in bronchial hyperemia. Ventilatory failure is very uncommon, even in patients with severe asthma, and arterial Pco_2 tends to be low due to increased ventilation.

Airway Hyperresponsiveness

AHR is the characteristic physiologic abnormality of asthma and describes the excessive bronchoconstrictor response to multiple inhaled triggers that would have no effect on normal airways. The increase in AHR is linked to the frequency of asthma symptoms, and, thus, an important aim of therapy is to reduce AHR. Increased bronchoconstrictor responsiveness is seen with *direct* broncho constrictors such as histamine and methacholine, which contract airway smooth muscle, but is characteristically also seen with many *indirect* stimuli, which release bronchoconstrictors from mast cells or activate sensory nerves. Most of the triggers for asthma symptoms appear to act indirectly, including allergens, exercise, hyperventilation, fog (via mast cell activation), irritant dusts, and sulfur dioxide (via a cholinergic reflex).

CLINICAL FEATURES AND DIAGNOSIS

The characteristic symptoms of asthma are wheezing, dyspnea, and coughing, which are variable, both spontaneously and with therapy. Symptoms may be worse at night and patients typically awake in the early morning hours. Patients may report difficulty in filling their lungs with air. There is increased mucus production in some patients, with typically tenacious mucus that is difficult to expectorate. There may be increased ventilation and use of accessory muscles of ventilation. Prodromal symptoms may precede an attack, with itching under the chin, discomfort between the scapulae, or inexplicable fear (impending doom).

Typical physical signs are inspiratory, and to a greater extent expiratory, rhonchi throughout the chest, and there may be hyperinflation. Some patients, particularly children, may present with a predominant nonproductive cough (cough-variant asthma). There may be no abnormal physical findings when asthma is under control.

Diagnosis

The diagnosis of asthma is usually apparent from the symptoms of variable and intermittent airways obstruction, but is usually confirmed by objective measurements of lung function.

LUNG FUNCTION TESTS

Simple spirometry confirms airflow limitation with a reduced FEV_1, FEV_1/FVC ratio, and PEF. Reversibility is demonstrated by a >12% and 200-mL increase in FEV_1 15 minutes after an inhaled short-acting β_2-agonist or in some patients by a 2 to 4 week trial of oral corticosteroids (OCS) (prednisone or prednisolone 30–40 mg daily). Measurements of PEF twice daily may confirm the diurnal variations in airflow obstruction. Flow-volume loops show reduced peak flow and reduced maximum expiratory flow. Further lung function tests are rarely necessary, but whole body plethysmography shows increased airway resistance and may show increased total lung capacity and residual volume. Gas diffusion is usually normal, but there may be a small increase in gas transfer in some patients.

AIRWAY RESPONSIVENESS

The increased AHR is normally measured by methacholine or histamine challenge with calculation of the provocative concentration that reduces FEV_1 by 20% (PC_{20}). This is rarely useful in clinical practice, but can be used in the differential diagnosis of chronic cough and when the diagnosis is in doubt in the setting of normal pulmonary function tests. Occasionally exercise testing is done to demonstrate the postexercise bronchoconstriction if there is a predominant history of EIA. Allergen challenge is rarely necessary and should only be undertaken by a specialist if specific occupational agents are to be identified.

HEMATOLOGIC TESTS

Blood tests are not usually helpful. Total serum IgE and specific IgE to inhaled allergens [radioallergosorbent test (RAST)] may be measured in some patients.

IMAGING

Chest roentgenography is usually normal but in more severe patients may show hyperinflated lungs. In exacerbations, there may be evidence of a pneumothorax. Lung shadowing usually indicates pneumonia or eosinophilic infiltrates in patients with bronchopulmonary aspergillosis. High-resolution CT may show areas of bronchiectasis in patients with severe asthma, and there may be thickening of the bronchial walls, but these changes are not diagnostic of asthma.

SKIN TESTS

Skin prick tests to common inhalant allergens are positive in allergic asthma and negative in intrinsic asthma, but are not helpful in diagnosis. Positive skin responses may be useful in persuading patients to undertake allergen avoidance measures.

EXHALED NITRIC OXIDE

Exhaled NO is now being used as a noninvasive test to measure eosinophilic airway inflammation. The typically elevated levels in asthma are reduced by ICS, so this may be a test of compliance with therapy. It may also be useful in demonstrating insufficient anti-inflammatory therapy.

Differential Diagnosis

It is usually not difficult to differentiate asthma from other conditions that cause wheezing and dyspnea. Upper airway obstruction by a tumor or laryngeal edema can mimic severe asthma, but patients typically present with stridor localized to large airways. The diagnosis is confirmed by a flow-volume loop that shows a reduction in inspiratory as well as expiratory flow, and bronchoscopy to demonstrate the site of upper airway narrowing. Persistent wheezing in a specific area of the chest may indicate endobronchial obstruction with a foreign body. Left ventricular failure may mimic the wheezing of asthma but basilar crackles are present in contrast to asthma.

Eosinophilic pneumonias and systemic vasculitis, including Churg-Strauss syndrome and polyarteritis

nodosa, may be associated with wheezing. Chronic obstructive pulmonary disease (COPD) is usually easy to differentiate from asthma as symptoms show less variability, never completely remit, and show much less (or no) reversibility to bronchodilators. Approximately 10% of COPD patients have features of asthma, with increased sputum eosinophils and a response to oral corticosteroids; these patients probably have both diseases concomitantly.

Treatment: Asthma

The treatment of asthma is straightforward and the majority of patients are now managed by internists and family doctors with effective and safe therapies. There are several aims of therapy (**Table 254-2**). Most emphasis has been placed on drug therapy, but several nonpharmacologic approaches have also been used. The main drugs for asthma can be divided into bronchodilators, which give rapid relief of symptoms mainly through relaxation of airway smooth muscle, and controllers, which inhibit the underlying inflammatory process.

Table 254-2 Aims of Asthma Therapy
• Minimal (ideally no) chronic symptoms, including nocturnal
• Minimal (infrequent) exacerbations
• No emergency visits
• Minimal (ideally no) use of a required β_2-agonist
• No limitations on activities, including exercise
• Peak expiratory flow circadian variation <20%
• (Near) normal PEF
• Minimal (or no) adverse effects from medicine

Abbreviation: PEF, peak expiratory flow.

BRONCHODILATOR THERAPIES

Bronchodilators act primarily on airway smooth muscle to reverse the bronchoconstriction of asthma. This gives rapid relief of symptoms but has little or no effect on the underlying inflammatory process. Thus, bronchodilators are not sufficient to control asthma in patients with persistent symptoms. There are three classes of bronchodilators in current use: β_2-adrenergic agonists, anticholinergics, and theophylline; of these, β_2-agonists are by far the most effective.

β_2-Agonists

β_2-Agonists activate β_2-adrenergic receptors, which are widely expressed in the airways. β_2-Receptors are coupled through a stimulatory G protein to adenylyl cyclase, resulting in increased intracellular cyclic adenosine monophosphate (AMP), which relaxes smooth muscle cells and inhibits certain inflammatory cells, particularly mast cells.

Mode of Action

The primary action of β_2-agonists is to relax airway smooth-muscle cells of all airways, where they act as functional antagonists, reversing and preventing contraction of airway smooth-muscle cells by all known bronchoconstrictors. This generalized action is likely to account for their great efficacy as bronchodilators in asthma. There are also additional nonbronchodilator effects that may be clinically useful, including inhibition of mast cell mediator release, reduction in plasma exudation, and inhibition of sensory nerve activation (**Table 254-3**). Inflammatory cells express small numbers of β_2-receptors, but these are rapidly downregulated with β_2-agonist activation so that, in contrast to corticosteroids, there are no effects on inflammatory cells in the airways and there is no reduction in AHR.

Table 254-3 Effects of β-Adrenergic Agonists on Airways
• Relaxation of airway smooth muscle (proximal and distal airways)
• Inhibition of mast cell mediator release
• Inhibition of plasma exudation and airway edema
• Increased mucociliary clearance
• Increased mucus secretion
• Decreased cough
• No effect on chronic inflammation

Clinical Use

ß2-Agonists are usually given by inhalation to reduce side effects. Short-acting ß2-agonists (SABAs) such as albuterol and terbutaline have a duration of action of 3–6 hours. They have a rapid onset of bronchodilation and are, therefore, used as needed for symptom relief. Increased use of SABAs indicates that asthma is not controlled. They are also useful in preventing EIA if taken prior to exercise. SABAs are used in high doses by nebulizer or via a metered-dose inhaler with a spacer. Long-acting ß2-agonists (LABAs) include salmeterol and formoterol, both of which have a duration of action over 12 hours and are given twice daily by inhalation. LABAs have replaced the regular use of SABAs, but LABAs should not be given in the absence of ICS therapy as they do not control the underlying inflammation. They do, however, improve asthma control and reduce exacerbations when added to ICS, which allows asthma to be controlled at lower doses of corticosteroids. This observation has led to the widespread use of fixed combination inhalers that contain a corticosteroid and a LABA, which have proved to be highly effective in the control of asthma.

Side Effects

Adverse effects are not usually a problem with ß2-agonists when given by inhalation. The most common side effects are muscle tremor and palpitations, which are seen more commonly in elderly patients. There is a small fall in plasma potassium due to increased uptake by skeletal muscle cells, but this effect does not usually cause any clinical problem.

Tolerance

Tolerance is a potential problem with any agonist given chronically, but while there is down-regulation of ß2-receptors, this does not reduce the bronchodilator response as there is a large receptor reserve in airway smooth-muscle cells. By contrast, mast cells become rapidly tolerant, but their tolerance may be prevented by concomitant administration of ICS.

Safety

The safety of ß2-agonists has been an important issue. There is an association between asthma mortality and the amount of SABA used, but careful analysis demonstrates that the increased use of rescue SABAs reflects poor asthma control, which is a risk factor for asthma death. The slight excess in mortality that has been associated with the use of LABAs is related to the lack of use of concomitant ICS, as the LABA therapy fails to suppress the underlying inflammation. This highlights the importance of always using an ICS when LABAs are given, which is most conveniently achieved by using a combination inhaler.

Anticholinergics

Muscarinic receptor antagonists such as ipratropium bromide, prevent cholinergic nerve-induced bronchoconstriction and mucus secretion. They are much less effective than ß2-agonists in asthma therapy as they inhibit only the cholinergic reflex component of bronchoconstriction, whereas ß2-agonists prevent all bronchoconstrictor mechanisms. Anticholinergics are, therefore, only used as an additional bronchodilator in patients with asthma that is not controlled by other inhaled medications. High doses may be given by nebulizer in treating acute severe asthma but should only be given following ß2-agonists, as they have a slower onset of bronchodilation.

Side effects are not usually a problem as there is little or no systemic absorption. The most common side effect is dry mouth; in elderly patients, urinary retention and glaucoma may also be observed.

Theophylline

Theophylline was widely prescribed as an oral bronchodilator several years ago, especially as it was inexpensive. It has now fallen out of favor as side effects are common and inhaled ß2-agonists are much more effective as bronchodilators. The bronchodilator effect is due to inhibition of phosphodiesterases in airway smooth-muscle cells, which increases cyclic AMP, but doses required for bronchodilation commonly cause side effects that are mediated mainly by phosphodiesterase inhibition. There is increasing evidence that theophylline at lower doses has anti-inflammatory effects, and these are likely to be mediated through different molecular mechanisms. There is evidence that theophylline activates the key nuclear enzyme histone deacetylase-2, which is a critical mechanism for switching off activated inflammatory genes.

Clinical Use

Oral theophylline is usually given as a slow-release preparation once or twice daily as this gives more stable plasma concentrations than normal theophylline tablets. It may be used as an additional bronchodilator in patients with severe asthma when plasma concentrations of 10–20 mg/L are required, although these concentrations are often associated with side effects. Low doses of theophylline, giving plasma concentrations of 5–10 mg/L, have additive effects to ICS and are particularly useful in patients with severe asthma. Indeed, withdrawal of theophylline from these patients may result in marked deterioration in asthma control. At low doses, the drug is well tolerated. IV aminophylline (a soluble salt of theophylline) was used for the treatment of severe asthma but has now been largely replaced by high doses of inhaled SABAs, which are more effective and have fewer

side effects. Aminophylline is occasionally used (via slow IV infusion) in patients with severe exacerbations that are refractory to SABAs.

Side Effects

Oral theophylline is well absorbed and is largely inactivated in the liver. Side effects are related to plasma concentrations; measurement of plasma theophylline may be useful in determining the correct dose. The most common side effects are nausea, vomiting, and headaches and are due to phosphodiesterase inhibition. Diuresis and palpitations may also occur, and at high concentrations cardiac arrhythmias, epileptic seizures, and death may occur due to adenosine A_1-receptor antagonism. Theophylline side effects are related to plasma concentration and are rarely observed at plasma concentrations below 10 mg/L. Theophylline is metabolized by CYP450 in the liver, and, thus, plasma concentrations may be elevated by drugs that block CYP450 such as erythromycin and allopurinol. Other drugs may also reduce clearance by other mechanisms leading to increased plasma concentrations (**Table 254-4**).

Table 254-4 Factors Affecting Clearance of Theophylline

Increased Clearance
• Enzyme induction (rifampicin, phenobarbitone, ethanol)
• Smoking (tobacco, marijuana) • High-protein, low-carbohydrate diet
• Barbecued meat
• Childhood
Decreased Clearance
• Enzyme inhibition (cimetidine, erythromycin, ciprofloxacin, allopurinol, zileuton, zafirlukast)
• Congestive heart failure
• Liver disease
• Pneumonia • Viral infection and vaccination
• High carbohydrate diet
• Old age

CONTROLLER THERAPIES

Inhaled Corticosteroids

ICS are by far the most effective controllers for asthma, and their early use has revolutionized asthma therapy.

Mode of Action

ICS are the most effective anti-inflammatory agents used in asthma therapy, reducing inflammatory cell numbers and their activation in the airways. ICS reduce eosinophils in the airways and sputum, and numbers of activated T lymphocytes and surface mast cells in the airway mucosa. These effects may account for the reduction in AHR that is seen with chronic ICS therapy.

The molecular mechanism of action of corticosteroids involves several effects on the inflammatory process. The major effect of corticosteroids is to switch off the transcription of multiple activated genes that encode inflammatory proteins such as cytokines, chemokines, adhesion molecules, and inflammatory enzymes. This effect involves several mechanisms, including inhibition of the transcription factors NF-κB and activator protein (AP)-1, but an important mechanism is recruitment of histone deacetylase-2 to the inflammatory gene complex, which reverses the histone acetylation associated with increased gene transcription. Corticosteroids also activate anti-inflammatory genes such as mitogen-activated protein (MAP) kinase phosphatase-1, and increase the expression of β_2-receptors. Most of the metabolic and endocrine side effects of corticosteroids are also mediated through transcriptional activation.

Clinical Use

ICS are by far the most effective controllers in the management of asthma and are beneficial in treating asthma of any severity and age. ICS are usually given twice daily, but some may be effective once daily in mildly symptomatic patients. ICS rapidly improve the symptoms of asthma, and lung function improves over several days. They are effective in preventing asthma symptoms, such as EIA and nocturnal exacerbations, but also prevent severe exacerbations. ICS reduce AHR, but maximal improvement may take several months of therapy. Early treatment with ICS appears to prevent irreversible changes in airway function that occur with chronic asthma. Withdrawal of ICS results in slow deterioration of asthma control, indicating that they suppress inflammation and symptoms, but do not cure the underlying condition. ICS are now given as first-line therapy for patients with persistent

asthma, but if they do not control symptoms at low doses, it is usual to add a LABA as the next step.

Side Effects

Local side effects include hoarseness (dysphonia) and oral candidiasis, which may be reduced with the use of a large-volume spacer device. There has been concern about systemic side effects from lung absorption, but many studies have demonstrated that ICS have minimal systemic effects (**Fig. 254-6**). At the highest recommended doses, there may be some suppression of plasma and urinary cortisol concentrations, but there is no convincing evidence that long-term treatment leads to impaired growth in children or to osteoporosis in adults. Indeed effective control of asthma with ICS reduces the number of courses of OCS that are needed and, thus, reduces systemic exposure to ICS.

Figure 254-6

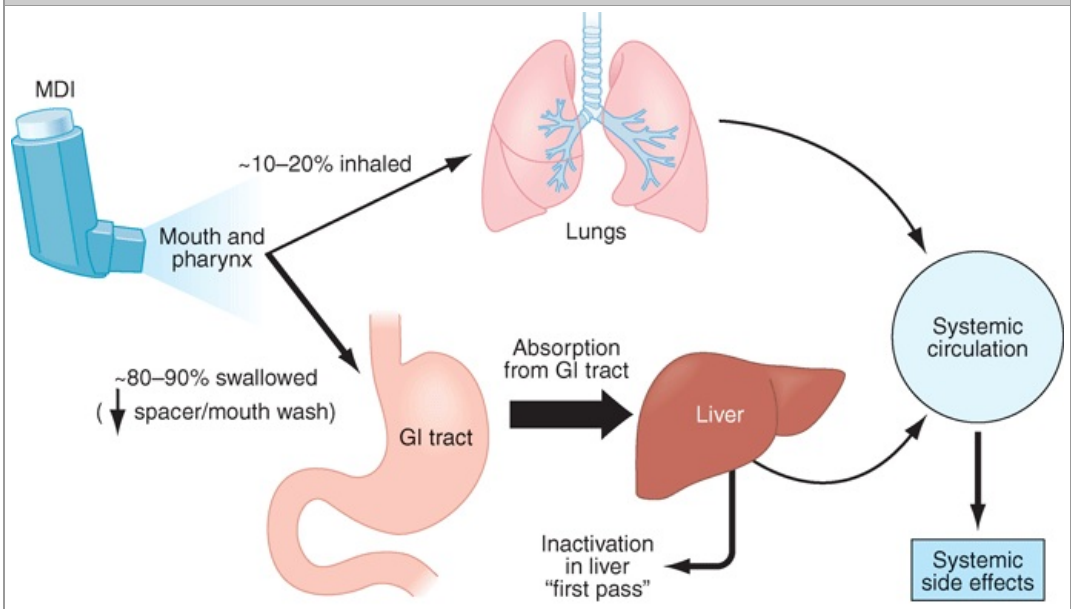

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pharmacokinetics of inhaled corticosteroids.

Systemic Corticosteroids

Corticosteroids are used intravenously (hydrocortisone or methylprednisolone) for the treatment of acute severe asthma, although several studies now show that OCS are as effective and easier to administer. A course of OCS (usually prednisone or prednisolone 30–45 mg once daily for 5–10 days) is used to treat acute exacerbations of asthma; no tapering of the dose is needed. Approximately 1% of asthma patients may require maintenance treatment with OCS; the lowest dose necessary to maintain control needs to be determined. Systemic side effects, including truncal obesity, bruising, osteoporosis, diabetes, hypertension, gastric ulceration, proximal myopathy, depression, and cataracts, may be a major problem, and steroid–sparing therapies may be considered if side effects are a significant problem. If patients require maintenance treatment with OCS, it is important to monitor bone density so that preventive treatment with bisphosphonates or estrogen in postmenopausal women may be initiated if bone density is low. Intramuscular triamcinolone acetonide is a depot preparation that is occasionally used in noncompliant patients, but proximal myopathy is a major problem with this therapy.

Antileukotrienes

Cysteinyl-leukotrienes are potent bronchoconstrictors, cause microvascular leakage, and increase eosinophilic inflammation through the activation of cys-LT_1-receptors. These inflammatory mediators are produced predominantly by mast cells and, to a lesser extent, eosinophils in asthma. Antileukotrienes such as montelukast and zafirlukast, block cys-LT_1-receptors and provide modest clinical benefit in asthma. They are less effective than ICS in controlling asthma and have less effect on airway inflammation, but are useful as an add-on therapy in some patients not controlled with low doses of ICS, although they are less effective than LABAs. They are given orally once or twice daily and are well tolerated. Some patients show a better response than others to antileukotrienes, but this has not been convincingly linked to any genomic differences in the leukotriene pathway.

Cromones

Cromolyn sodium and nedocromil sodium are asthma controller drugs that appear to inhibit mast cell and sensory nerve activation and are, therefore, effective in blocking trigger-induced asthma such as

EIA and allergen- and sulfur dioxide–induced symptoms. Cromones have relatively little benefit in the long-term control of asthma due to their short duration of action (at least four times daily by inhalation). They are very safe and were popular in the treatment of childhood asthma, although now low doses of ICS are preferred as they are more effective and have a proven safety profile.

Steroid-Sparing Therapies

Various immunomodulatory treatments have been used to reduce the requirement for OCS in patients with severe asthma, who have serious side effects with this therapy. Methotrexate, cyclosporin A, azathioprine, gold, and IV gamma globulin have all been used as steroid-sparing therapies, but none of these treatments has any long-term benefit and each is associated with a relatively high risk of side effects.

Anti-IgE

Omalizumab is a blocking antibody that neutralizes circulating IgE without binding to cell-bound IgE and, thus, inhibits IgE-mediated reactions. This treatment has been shown to reduce the number of exacerbations in patients with severe asthma and may improve asthma control. However, the treatment is very expensive and is only suitable for highly selected patients who are not controlled on maximal doses of inhaler therapy and have a circulating IgE within a specified range. Patients should be given a 3 to 4-month trial of therapy to show objective benefit. Omalizumab is usually given as a subcutaneous injection every 2–4 weeks and appears not to have significant side effects, although anaphylaxis is very occasionally seen.

Immunotherapy

Specific immunotherapy using injected extracts of pollens or house dust mites has not been very effective in controlling asthma and may cause anaphylaxis. Side effects may be reduced by sublingual dosing. It is not recommended in most asthma treatment guidelines because of lack of evidence of clinical efficacy.

Alternative Therapies

Nonpharmacologic treatments, including hypnosis, acupuncture, chiropraxis, breathing control, yoga, and speleotherapy, may be popular with some patients. However, placebo-controlled studies have shown that each of these treatments lacks efficacy and cannot be recommended. However, they are not detrimental and may be used as long as conventional pharmacologic therapy is continued.

Future Therapies

It has proved very difficult to discover novel pharmaceutical therapies, particularly as current therapy with corticosteroids and β_2-agonists is so effective in the majority of patients. There is, however, a need for the development of new therapies for patients with refractory asthma who have side effects with systemic corticosteroids. Antagonists of specific mediators have little or no benefit in asthma, apart from antileukotrienes, which have rather weak effects, presumably reflecting the fact that multiple mediators are involved. Blocking antibodies against IL-5 may reduce exacerbations in highly selected patients who have sputum eosinophils despite high doses of corticosteroids, whereas anti-TNF-α antibodies are not effective in severe asthma. Novel anti-inflammatory treatments that are in clinical development include inhibitors of phosphodiesterase-4, NF-κB and p38 MAP kinase. However, these drugs, which act on signal transduction pathways common to many cells, are likely to have troublesome side effects, necessitating their delivery by inhalation. Safer and more effective immunotherapy using T-cell peptide fragments of allergens or DNA vaccination are also being investigated. Bacterial products, such as CpG oligonucleotides that stimulate T_H1 immunity or regulatory T cells, are also currently under evaluation.

MANAGEMENT OF CHRONIC ASTHMA

There are several aims of chronic therapy in asthma (Table 254-2). It is important to establish the diagnosis objectively using spirometry or PEF measurements at home. Triggers that worsen asthma control, such as allergens or occupational agents, should be avoided, whereas triggers, such as exercise and fog, which result in transient symptoms, provide an indication that more controller therapy is needed.

Stepwise Therapy

For patients with mild, intermittent asthma, a short-acting β_2-agonist is all that is required (**Fig. 254-7**). However, use of a reliever medication more than three times a week indicates the need for regular controller therapy. The treatment of choice for all patients is an ICS given twice daily. It is usual to start with an intermediate dose [e.g., 200 (μg) bid of (beclomethasone dipropionate) BDP] or equivalent and to decrease the dose if symptoms are controlled after three months. If symptoms are not controlled, a LABA should be added, which is most conveniently given by switching to a combination inhaler. The dose of controller should be adjusted accordingly, as judged by the need for a rescue inhaler. Low doses of theophylline or an antileukotriene may also be considered as an add-on therapy, but these are less effective than LABAs. In patients with severe asthma, low-dose oral theophylline is also helpful, and when there is irreversible airway narrowing, the long-acting anticholinergic tiotropium

bromide may be tried. If asthma is not controlled despite the maximal recommended dose of inhaled therapy, it is important to check compliance and inhaler technique. In these patients, maintenance treatment with an OCS may be needed and the lowest dose that maintains control should be used. Occasionally omalizumab may be tried in steroid-dependent asthmatics who are not well controlled. Once asthma is controlled, it is important to slowly decrease therapy in order to find the optimal dose to control symptoms.

Figure 254-7

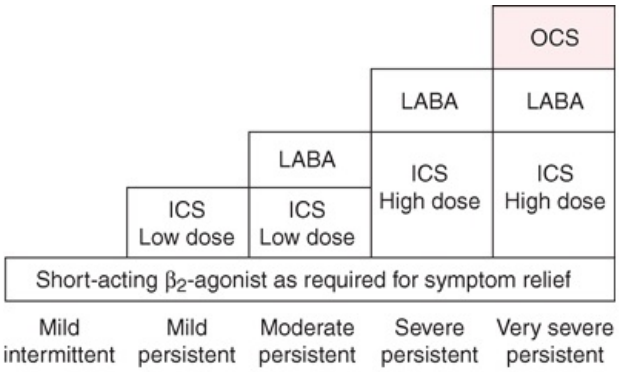

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stepwise approach to asthma therapy according to the severity of asthma and ability to control symptoms. ICS, inhaled corticosteroids; LABA, long-acting β_2-agonist; OCS, oral corticosteroid.

Education

Patients with asthma need to understand how to use their medications and the difference between reliever and controller therapies. Education may improve compliance, particularly with ICS. All patients should be taught how to use their inhalers correctly. In particular, they need to understand how to recognize worsening of asthma and how to step up therapy. Written action plans have been shown to reduce hospital admissions and morbidity rates in adults and children, and are recommended particularly in patients with unstable disease who have frequent exacerbations.

ACUTE SEVERE ASTHMA

Exacerbations of asthma are feared by patients and may be life threatening. One of the main aims of controller therapy is to prevent exacerbations; in this respect, ICS and combination inhalers are very effective.

Clinical Features

Patients are aware of increasing chest tightness, wheezing, and dyspnea that are often not or poorly relieved by their usual reliever inhaler. In severe exacerbations patients may be so breathless that they are unable to complete sentences and may become cyanotic. Examination usually shows increased ventilation, hyperinflation, and tachycardia. Pulsus paradoxus may be present, but this is rarely a useful clinical sign. There is a marked fall in spirometric values and PEF. Arterial blood gases on air show hypoxemia and P_{CO2} is usually low due to hyperventilation. A normal or rising P_{CO2} is an indication of impending respiratory failure and requires immediate monitoring and therapy. A chest roentgenogram is not usually informative, but may show pneumonia or pneumothorax.

Treatment: Acute Severe Asthma

A high concentration of oxygen should be given by face mask to achieve oxygen saturation of >90%. The mainstay of treatment are high doses of SABAs given either by nebulizer or via a metered-dose inhaler with a spacer. In severely ill patients with impending respiratory failure, IV β_2-agonists may be given. An inhaled anticholinergic may be added if there is not a satisfactory response to β_2-agonists alone, as there are additive effects. In patients who are refractory to inhaled therapies, a slow infusion of aminophylline may be effective, but it is important to monitor blood levels, especially if patients have already been treated with oral theophylline. Magnesium sulfate given intravenously or by nebulizer has also been shown to be effective when added to inhaled β_2-agonists, and is relatively well tolerated but is not routinely recommended. Prophylactic intubation may be indicated for impending respiratory failure, when the P_{CO2} is normal or rises. For patients with respiratory failure, it is necessary to intubate and institute ventilation. These patients may benefit from an anesthetic such as halothane if they have not responded to conventional bronchodilators. Sedatives should never be given as they may depress ventilation. Antibiotics should not be used routinely unless there are signs of pneumonia.

REFRACTORY ASTHMA

Although most patients with asthma are easily controlled with appropriate medication, a small proportion of patients (approximately 5% of asthmatics) are difficult to control despite maximal inhaled therapy. Some of these patients will require maintenance treatment with OCS. In managing these patients, it is important to investigate and correct any mechanisms that may be aggravating asthma. There are two major patterns of difficult asthma: some patients have persistent symptoms and poor lung function, despite appropriate therapy, whereas others may have normal or near normal lung function but intermittent, severe (sometimes life-threatening) exacerbations.

Mechanisms

The most common reason for poor control of asthma is noncompliance with medication, particularly ICS. Compliance with ICS may be low because patients do not feel any immediate clinical benefit or may be concerned about side effects. Compliance with ICS is difficult to monitor as there are no useful plasma measurements that can be made. Compliance may be improved by giving the ICS as a combination with a LABA that gives symptom relief. Compliance with OCS may be measured by suppression of plasma cortisol and the expected concentration of prednisone/prednisolone in the plasma. There are several factors that may make asthma more difficult to control, including exposure to high, ambient levels of allergens or unidentified occupational agents. Severe rhinosinusitis may make asthma more difficult to control; upper airway disease should be vigorously treated. Gastroesophageal reflux is common among asthmatics due to bronchodilator therapy, but there is little evidence that it is a significant factor in worsening asthma, and treatment of the reflux is not usually effective at improving asthma symptoms. Some patients may have chronic infection with *Mycoplasma pneumoniae* or *Chlamydophila pneumoniae* and benefit from treatment with a macrolide antibiotic. Drugs such as beta-adrenergic blockers, aspirin, and other cyclooxygenase (COX) inhibitors may worsen asthma. Some women develop severe premenstrual worsening of asthma, which is unresponsive to corticosteroids and requires treatment with progesterone or gonadotropin-releasing factors. Few systemic diseases make asthma more difficult to control, but hyper- and hypothyroidism may increase asthma symptoms and should be investigated if suspected.

Relatively little is known about the pathology of refractory asthma, as biopsy studies are more difficult in these patients. Some patients show the typical eosinophilic pattern of inflammation, whereas others have a predominantly neutrophilic pattern. There may be an increase in T_H1 cells and CD8 lymphocytes compared to mild asthma and increased expression of TNF-α. Structural changes in the airway, including fibrosis, angiogenesis, and airway smoothmuscle thickening, are more commonly seen in these patients.

Differential Diagnosis

Some patients who apparently have difficult-to-control asthma have vocal cord dysfunction, resulting in wheezing or stridor. This symptom is thought to be an attention-seeking hysterical conversion syndrome and may lead to escalating doses of asthma therapy with some patients taking high doses of oral corticosteroids. It may be recognized by the characteristic discrepancy between tests of forced expiration, such as FEV_1 and PEF, and relatively normal airway resistance. Direct inspection by laryngoscopy may confirm adduction of the vocal cords at the time of symptoms. This condition is usually difficult to manage, but it is important that patients be weaned off ocs and lcs. Speech therapy is sometimes beneficial. Some patients with COPD may be diagnosed as asthmatic and may show the characteristic poor response to corticosteroids and bronchodilators, but this situation is complicated by the fact that some patients with COPD also have concomitant asthma.

Corticosteroid-Resistant Asthma

A few patients with asthma show a poor response to corticosteroid therapy and may have various molecular abnormalities that impair the anti-inflammatory action of corticosteroids. Complete resistance to corticosteroids is extremely uncommon and affects less than 1 in 1000 patients. It is defined by a failure to respond to a high dose of oral prednisone/prednisolone (40 mg once daily over 2 weeks), ideally with a 2-week run-in with matched placebo. More common is reduced responsiveness to corticosteroids where control of asthma requires OCS (corticosteroid-dependent asthma). In all patients with poor responsiveness to corticosteroids, there is a reduction in the response of circulating monocytes and lymphocytes to the anti-inflammatory effects of corticosteroids in vitro and reduced skin blanching in response to topical corticosteroids. There are several mechanisms that have been described, including an excess of the transcription factor AP-1, an increase in the alternatively spliced form of the glucocorticoid receptor (GR)-β, an abnormal pattern of histone acetylation in response to corticosteroids, a defect in IL-10 production, and a reduction in histone deacetylase activity (as in COPD). These observations suggest that there are likely to be heterogeneous mechanisms for corticosteroid resistance; whether these mechanisms are genetically determined has yet to be decided.

Brittle Asthma

Some patients show chaotic variations in lung function despite taking appropriate therapy. Some show

a persistent pattern of variability and may require oral corticosteroids or, at times, continuous infusion of β_2-agonists (type I brittle asthma), whereas others have generally normal or near-normal lung function but precipitous, unpredictable falls in lung function that may result in death (type 2 brittle asthma). These latter patients are difficult to manage as they do not respond well to corticosteroids, and the worsening of asthma does not reverse well with inhaled bronchodilators. The most effective therapy is subcutaneous epinephrine, which suggests that the worsening is likely to be a localized airway anaphylactic reaction with edema. In some of these patients, there may be allergy to specific foods. These patients should be taught to self-administer epinephrine and should carry a medical warning accordingly.

Treatment: Refractory Asthma

Refractory asthma is difficult to control, by definition. It is important to check compliance and the correct use of inhalers and to identify and eliminate any underlying triggers. Low doses of theophylline may be helpful in some patients, and theophylline withdrawal has been found to worsen in many patients. Most of these patients will require maintenance treatment with oral corticosteroids, and the minimal dose that achieves satisfactory control should be determined by careful dose titration. Steroid-sparing therapies are rarely effective. In some patients with allergic asthma, omalizumab is effective, particularly when there are frequent exacerbations. Anti-TNF therapy is not effective in severe asthma and should not be used. A few patients may benefit from infusions of β_2-agonists. New therapies are needed for these patients, who currently consume a disproportionate amount of health care spending.

SPECIAL CONSIDERATIONS

Although asthma is usually straightforward to manage, there are some situations that may require additional investigation and different therapy.

Aspirin-Sensitive Asthma

A small proportion (1–5%) of asthmatics become worse with aspirin and other COX inhibitors, although this is much more commonly seen in severe cases and in those patients with frequent hospital admission. Aspirin-sensitive asthma is a well defined subtype of asthma that is usually preceded by perennial rhinitis and nasal polyps in nonatopic patients with a late onset of the disease. Aspirin, even in small doses, characteristically provokes rhinorrhea, conjunctival irritation, facial flushing, and wheezing. There is a genetic predisposition to increased production of cysteinyl-leukotrienes with functional polymorphism of *cys*-leukotriene C synthase. Asthma is triggered by COX inhibitors, but is persistent even in their absence. All nonselective COX inhibitors should be avoided, but selective COX2 inhibitors are safe to use when an anti-inflammatory analgesic is needed. Aspirin-sensitive asthma responds to usual therapy with ICS. Although antileukotrienes should be effective in these patients, they are no more effective than in allergic asthma. Occasionally, aspirin desensitization is necessary, but this should only be undertaken in specialized centers.

Asthma in the Elderly

Asthma may start at any age, including in elderly patients. The principles of management are the same as in other asthmatics, but side effects of therapy may be a problem, including muscle tremor with β_2-agonists and more systemic side effects with ICS. Comorbidities are more frequent in this age group, and interactions with drugs such as β_2-blockers, COX inhibitors, and agents that may affect theophylline metabolism need to be considered. COPD is more likely in elderly patients and may coexist with asthma. A trial of OCS may be very useful in documenting the steroid responsiveness of asthma.

Pregnancy

Approximately one-third of asthmatic patients who are pregnant improve during the course of a pregnancy, one-third deteriorate, and one-third are unchanged. It is important to maintain good control of asthma as poor control may have adverse effects on fetal development. Compliance may be a problem as there is often concern about the effects of antiasthma medications on fetal development. The drugs that have been used for many years in asthma therapy have now been shown to be safe and without teratogenic potential. These drugs include short-acting β_2-agonists, ICS, and theophylline; there is less safety information about newer classes of drugs such as LABAs, antileukotrienes, and anti-IgE. If an OCS is needed, it is better to use prednisone rather than prednisolone as it cannot be converted to the active prednisolone by the fetal liver, thus protecting the fetus from systemic effects of the corticosteroid. There is no contraindication to breast-feeding when patients are using these drugs.

Cigarette Smoking

Approximately 20% of asthmatics smoke, which may adversely affect asthma in several ways. Smoking asthmatics have more severe disease, more frequent hospital admissions, a faster decline in lung function, and a higher risk of death from asthma than nonsmoking asthmatics. There is evidence that smoking interferes with the anti-inflammatory actions of corticosteroids, necessitating higher doses for asthma control. Smoking cessation improves lung function and reduces the steroid resistance, and,

thus, vigorous smoking cessation strategies should be used. Some patients report a temporary worsening of asthma when they first stop smoking, which could be due to the loss of the bronchodilating effect of no in cigarette smoke.

Surgery

If asthma is well controlled, there is no contraindication to general anesthesia and intubation. Patients who are treated with OCS will have adrenal suppression and should be treated with an increased dose of OCS immediately prior to surgery. Patients with FEV_1 <80% of their normal levels should also be given a boost of OCS prior to surgery. High-maintenance doses of corticosteroids may be a contraindication to surgery because of increased risks of infection and delayed wound healing.

Bronchopulmonary Aspergillosis

Bronchopulmonary aspergillosis (BPA) is uncommon and results from an allergic pulmonary reaction to inhaled spores of *Aspergillus fumigatus* and, occasionally, other *Aspergillus* species. A skin prick test to *A. fumigatus* is always positive, whereas serum *Aspergillus* precipitins are low or undetectable. Characteristically, there are fleeting eosinophilic infiltrates in the lungs, particularly in the upper lobes. Airways become blocked with mucoid plugs rich in eosinophils, and patients may cough up brown plugs and have hemoptysis. BPA may result in bronchiectasis, particularly affecting central airways, if not suppressed by corticosteroids. Asthma is controlled in the usual way by ICS, but it is necessary to give a course of OCS if any sign of worsening or pulmonary shadowing is found. Treatment with the oral antifungal itraconazole is beneficial in preventing exacerbations.

FURTHER READINGS

Barnes PJ: How corticosteroids control inflammation. Br J Pharmacol 148:245, 2006.[PMID: 16604091] [Full Text]

_____: Cytokine networks in asthma and chronic obstructive pulmonary disease. J Clin Invest 118:3546, 2008

_____: Immunology of asthma and chronic obstructive pulmonary disease. Nat Immunol Rev 8:183, 2008

_____ et al: *Asthma and COPD*, 2nd ed. Amsterdam, Elsevier, 2009

Bateman ED et al: Global strategy for asthma management and prevention: GINA executive summary. Eur Respir J 31:143, 2008[PMID: 18166595] [Full Text]

Eder W et al: The asthma epidemic. N Engl J Med 355:2226, 2006[PMID: 17124020] [Full Text]

Fanta CH: Asthma. N Engl J Med 360:1002, 2009[PMID: 19264689] [Full Text]

Farooque SP, Lee TH: Aspirin-sensitive respiratory disease. Annu Rev Physiol 71:465, 2009[PMID: 19575683] [Full Text]

Hamid Q, Tulic M: Immunobiology of asthma. Annu Rev Physiol 71:489, 2009[PMID: 19575684] [Full Text]

_____: New therapies for asthma: Is there any progress? Trends Pharmacol Sci 31:355, 2010

Lazarus SC: Emergency treatment of asthma. N Engl J Med 363:755, 2010[PMID: 20818877] [Full Text]

Wenzel SE, Busse WW: Severe asthma: Lessons from the Severe Asthma Research Program. J Allergy Clin Immunol 119:14, 2007[PMID: 17208583] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HYPERSENSITIVITY PNEUMONITIS

First described in 1874, hypersensitivity pneumonitis (HP), or extrinsic allergic alveolitis, is an inflammatory disorder of the lung involving alveolar walls and terminal airways that is induced by repeated inhalation of a variety of organic agents in a susceptible host. The expression of HP depends on factors related to the host susceptibility and the inciting agent. The frequency of HP varies with the environmental exposure and the specific antigen involved, which often depends on season, geographic location, or presence of certain industries.

Etiology

Agents implicated as causes of HP are diverse and include those listed in **Table 255-1**. The common name of each disease often reflects the occupational or avocational risk associated with that disease. In the United States, the most common types of HP are farmer's lung, bird fancier's lung, and chemical worker's lung. In *farmer's lung*, inhalation of proteins, such as thermophilic bacteria and fungal spores that are present in moldy bedding and feed, are most commonly responsible for the development of HP. These antigens are probably also responsible for the etiology of *mushroom worker's disease* (moldy composted growth medium), bagassosis (moldy sugar cane), and water-related exposure (molds in air conditioners or humidifiers). *Hot tub lung* refers to a hypersensitivity reaction to *Mycobacterium avium complex*, which is present in hot tubs or whirlpools and is differentiated from actual infection. *Bird fancier's lung* (and the related disorders of duck fever, turkey handler's lung, and dove pillow's lung) is a response to inhalation of proteins from feathers and droppings. *Chemical worker's lung* is an example of how simple chemicals, such as isocyanates, may also cause immune-mediated diseases. Interestingly, cigarette smoking has been associated with decreased incidence of HP; however, smoking may lead to a more progressive or severe course of HP once the disease is present.

Table 255-1 Selected Examples of Hypersensitivity Pneumonitis (HP)

Disease	Antigen	Source of Antigen
Bagassosis	Thermophilic actinomycetes[a]	"Moldy" bagasse (sugar cane)
Bird fancier's, breeder's, or handler's lung[b]	Parakeet, pigeon, chicken, turkey proteins	Avian droppings or feathers
Cephalosporium HP	Contaminated basement (sewage)	*Cephalosporium*
Cheese washer's lung	*Penicillium casei*	Moldy cheese
Chemical worker's lung[b]	Isocyanates	Polyurethane foam, varnishes, lacquer
Coffee worker's lung	Coffee bean dust	Coffee beans
Compost lung	*Aspergillus*	Compost
Detergent worker's disease	*Bacillus subtilis enzymes (subtilisins)*	Detergent
Familial HP	*Bacillus subtilis*	Contaminated wood dust in walls
Farmer's lung[b]	Thermophilic actinomycetes[a]	"Moldy" hay, grain, silage
Fish food lung	Unknown	Fish food
Fish meal worker's lung	Fish meal dust	Fish meal

Furrier's lung	Animal fur dust	Animal pelts
Hot tub lung	*Cladosporium spp., Mycobacterium avium* complex	Mold on ceiling; contaminated water
Humidifier or air conditioner lung (ventilation pneumonitis)	*Aureobasidium pullulans, Candida albicans,* Thermophilic actinomycetes,[a] *mycobacterium* spp., other microorganisms	Contaminated water in humidification or forced-air air conditioning systems
Japanese summer-type HP	*Trichosporon cutaneum, T. asahii*, and *T. mucoides*	House dust, bird droppings
Laboratory worker's HP	Male rat urine	Laboratory rat
Lycoperdonosis	*Lycoperdon puffballs*	Puffball spores
Malt worker's lung	*Aspergillus fumigatus or A. clavatus*	Moldy barley
Maple bark disease	*Cryptostroma corticale*	Maple bark
Metalworking fluid lung	*Mycobacterium spp., Pseudomonas spp.*	Contaminated metalworking fluid
Miller's lung	*Sitophilus granarius (wheat weevil)*	Infested wheat flour
Miscellaneous medication	Amiodarone, bleomycin, efavirenz, gemcitabine, hydralazine, hydroxyurea, isoniazid, methotrexate, paclitaxel, penicillin, procarbazine, propranolol, riluzole, sirolimus, sulfasalazine	Medication
Mushroom worker's lung	Thermophilic actinomycetes,[a] *Hypsizygus marmoreus, Bunashimeji*, and other exotic mushrooms	Mushroom compost; mushrooms
Pituitary snuff taker's lung	Animal proteins	Heterologous pituitary snuff
Potato riddler's lung	Thermophilic actinomycetes,[a] *Aspergillus*	"Moldy" hay around potatoes
Sauna taker's lung	*Aureobasidium* spp., other	Contaminated sauna water
Sausage worker's lung	*Penicillium nalgiovense*	Dry sausage mold
Sequoiosis	*Aureobasidium, Graphium* spp.	Redwood sawdust
Streptomyces albus HP	*Streptomyces albus*	Contaminated fertilizer
Suberosis	*Penicillium glabrum and Chrysonilia sitophila*	Cork dust
Tap water lung	*Mycobacteria* spp.	Contaminated tap water
Thatched roof disease	*Saccharomonospora viridis*	Dried grasses and leaves
Tobacco worker's disease	*Aspergillus* spp.	Mold on tobacco
Winegrower's lung	*Botrytis cinerea*	Mold on grapes
Wood trimmer's disease	*Rhizopus spp., Mucor spp.*	Contaminated wood trimmings
Woodman's disease	*Penicillium spp.*	Oak and maple trees
Woodworker's lung	Wood dust, Alternaria	Oak, cedar, pine, and mahogany dusts

[a]Thermophilic actinomycetes species include *Micropolyspora faeni, Thermoactinomyces vulgaris, T. saccharri, T. viridis, and T. candidus.*

[b]Most common causes of hypersensitivity pneumonitis in the United States.

Pathogenesis

The finding that precipitating antibodies against extracts of moldy hay were demonstrable in most patients with farmer's lung led to the early conclusion that HP was an immune complex–mediated reaction. Subsequent investigations of HP in human beings and animal models provided evidence for the importance of cell-mediated hypersensitivity. The very early (acute) reaction is characterized by an

increase in polymorphonuclear leukocytes in the alveoli and small airways. This early lesion is followed by an influx of mononuclear cells into the lung and the formation of granulomas that appear to be the result of a classic delayed (T cell–mediated) hypersensitivity reaction to repeated inhalation of antigen and adjuvant-active materials. Studies in animal models suggest that the disease is a T_H1-mediated immune response to antigen, with interferon γ, interleukin (IL)-12, and possibly IL-18 contributing to disease expression. Most likely, multiple cytokines [including also IL-1β, transforming growth factor β (TGF-β), tumor necrosis factor α (TNF-α) and others] interact to promote HP; their source includes both alveolar macrophages and T lymphocytes in the lung. Data support a genetic predisposition to the development of HP; certain polymorphisms of the TNF-α promoter region and major histocompatibility complex reportedly confer an enhanced susceptibility to pigeon breeder's disease.

After inhalation of an antigenic particle, the attraction and accumulation of inflammatory cells in the lung may be due to one or more of the following mechanisms: induction of the adhesion molecules L-selectin and E-selectin, elaboration by dendritic cells of CC chemokine 1 (DC-CK-1/CCL18), or increased expression of CXCR3/CXCL10 by $CD4^+$ and $CD8^+$ lymphocytes. Increased levels of Fas protein and FasL in the lung (which would be expected to suppress inflammation by induction of T cell apoptosis) is counterbalanced by increased expression of the inducible antiapoptotic gene *Bcl-xL*, resulting in a lower overall level of pulmonary lymphocyte apoptosis in HP patients.

Bronchoalveolar lavage (BAL) in patients with HP consistently demonstrates an increase in T lymphocytes in lavage fluid (a finding that is also observed in patients with other granulomatous lung disorders). Patients with recent or continual exposure to antigen may have an increase in polymorphonuclear leukocytes in lavage fluid, which has been associated with lung fibrosis. A role for oxidant injury has been proposed in HP. Several markers of oxidative stress are reported to be increased during exacerbation of HP and are reduced by treatment with glucocorticoids.

Clinical Presentation

The clinical picture is that of an interstitial pneumonitis, which varies from patient to patient and seems related to the frequency and intensity of exposure to the causative antigen and, perhaps, other host factors. The presentation can be acute, subacute, or chronic. In the *acute* form, symptoms such as cough, fever, chills, malaise, and dyspnea may occur 6 to 8 h after exposure to the antigen and usually clear within a few days if there is no further exposure to antigen; it often closely resembles an influenza-like illness. The *subacute* form often appears insidiously over a period of weeks marked by cough and dyspnea and may progress to cyanosis and severe dyspnea, requiring hospitalization. In some patients, a subacute form of the disease may persist after an acute presentation of the disorder, especially if there is continued exposure to antigen. In most patients with the acute or subacute form of HP, the symptoms, signs, and other manifestations of HP disappear within days, weeks, or months if the causative agent is no longer inhaled. Transformation to a chronic form of the disease may occur, but the frequency of such progression is uncertain.

Continuous low-level antigen exposure or repeated episodes can also lead to chronic disease with more subtle symptoms, accounting for delayed or uncertain diagnosis over a long period of time. This may occur without a prior history of acute or subacute manifestations. The *chronic* form of HP may be clinically indistinguishable from pulmonary fibrosis in its later stages. Symptoms include cough, weight loss, malaise, and gradual increase in dyspnea. Physical examination may reveal inspiratory crackles and digital clubbing. Imaging shows interstitial fibrosis or emphysema. Progressive worsening may result in dependence on supplemental oxygen, pulmonary hypertension, or respiratory failure. Pulmonary fibrosis is the clinical manifestation of HP with the greatest predictive value for mortality. Fibrosis appears most prominent in hypersensitivity pneumonitis associated with birds, while emphysema is often more common in farmer's lung.

Diagnosis

All forms of the disease may be associated with elevations in erythrocyte sedimentation rate, C-reactive protein, rheumatoid factor, lactate dehydrogenase, or serum immunoglobulins. Following acute exposure to an antigen, neutrophilia and lymphopenia are frequently present. Eosinophilia is not a feature. Examination for *serum precipitins* against suspected antigens, such as those listed in Table 255-1, is an important part of the diagnostic workup and should be performed on any patient with interstitial lung disease, especially if a suggestive exposure history is elicited. The occurrence of precipitins indicates sufficient exposure to the causative agent for generation of an immunologic response and is one of the major diagnostic criteria; however, the diagnosis of HP is not established solely by the presence of precipitins, as they are found in sera of many individuals exposed to appropriate antigens who demonstrate no other evidence of HP. False-negative results may occur because of unreliable testing techniques or an inappropriate choice of antigens. Extraction of antigens from the suspected source may at times be helpful.

Chest x-ray shows no specific or distinctive changes in HP. It can be normal even in symptomatic patients. The acute or subacute phases may be associated with poorly defined, patchy, or diffuse infiltrates; with discrete, nodular infiltrates; or with air-space consolidation. In the chronic phase, the chest x-ray usually shows a diffuse reticulonodular infiltrate. Honeycombing may eventually develop as

the condition progresses. Apical sparing is common, suggesting that disease severity correlates with inhaled antigen load, but no particular distribution or pattern is classic for HP. Abnormalities rarely seen in HP include pleural effusion or thickening and significant hilar adenopathy.

High-resolution chest CT has become the procedure of choice for imaging of HP. Although pathognomonic features have not been identified, acute HP may appear with diffuse "ground-glass" infiltrates, a reticulonodular pattern, or confluent alveolar opacification. In subacute disease, centrilobularnodules and "ground-glass" changes predominate, and expiratory views may demonstrate air trapping or mosaic perfusion **(Fig. 255-1)**. This pattern is more common in individuals whose exposure to antigen continues rather than in those in whom removal from antigen exposure has occurred. In chronic HP, diffuse changes include patchy emphysema and interstitial fibrosis; subpleural linear opacities and honeycombing are also common. The findings are often similar (but not identical) to idiopathic pulmonary fibrosis.

Figure 255-1

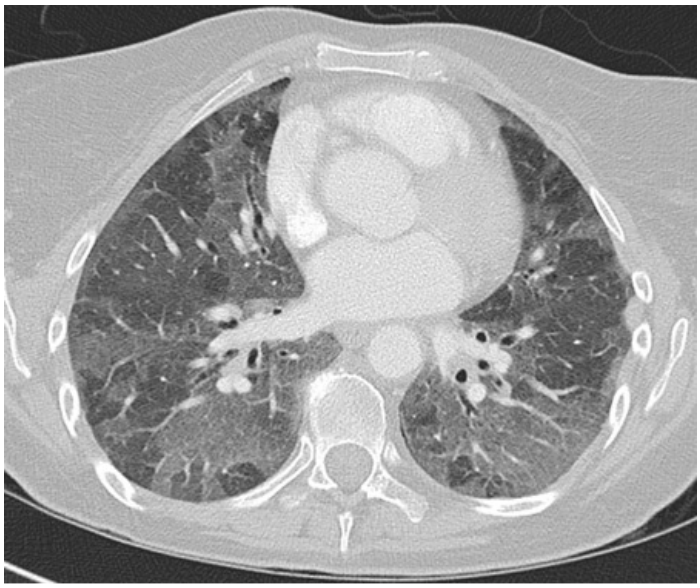

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest CT scan of a patient with subacute hypersensitivity pneumonitis in which scattered regions of ground-glass infiltrates in a mosaic pattern consistent with air trapping are seen bilaterally. This patient had *bird fancier's lung.(Courtesy of TJ Gross; with permission.)*

Pulmonary function studies in all forms of HP may show a restrictive or an obstructive pattern with loss of lung volumes, impaired diffusing capacity, and decreased compliance. Resting or exercise-induced hypoxemia may be seen. Bronchospasm and bronchial hyperreactivity are sometimes found in acute HP. With antigen avoidance, the pulmonary function abnormalities are usually reversible in acute and subacute disease.

BAL is used in some centers to aid in diagnostic evaluation. A marked lymphocytic alveolitis on BAL is almost universal, although not pathognomonic. Lymphocytes are typically activated and show a decreased helper/suppressor ratio, although this ratio can be variable depending on dose and duration of exposure. Alveolar neutrophilia is also prominent acutely, but tends to fade in the absence of recurrent exposure. Bronchoalveolar mastocytosis may correlate with disease activity.

Lung biopsy, obtained through flexible bronchoscopy, open-lung procedures, or thoracoscopy, may be diagnostic. Although the histopathology is distinctive, it may not be pathognomonic of HP **(Fig. 255-2)**. When the biopsy is taken during the active phase of disease, typical findings include an interstitial alveolar infiltrate consisting of plasma cells, lymphocytes, and occasional eosinophils and neutrophils, usually accompanied by loose, noncaseating peribronchial granulomas. Some degree of bronchiolitis is found in about one-half the cases. Rarely, bronchiolitis obliterans with organizing pneumonia (BOOP) (Chap. 261) may be present. In subacute disease, the triad of mononuclear bronchiolitis; interstitial infiltrates of lymphocytes and plasma cells; and single, nonnecrotizing, randomly scattered parenchymal granulomas without mural vascular involvement is consistent with HP. Interstitial fibrosis may be present, but most often is mild in earlier stages of the disease. Chronic HP has variable pathology and may resemble nonspecific interstitial pneumonia, organizing pneumonia, or usual interstitial pneumonia; granulomas may or may not be present. Centrilobular fibrosis, peribronchial inflammation with fibrosis, bridging fibrosis, and emphysema are common.

Figure 255-2

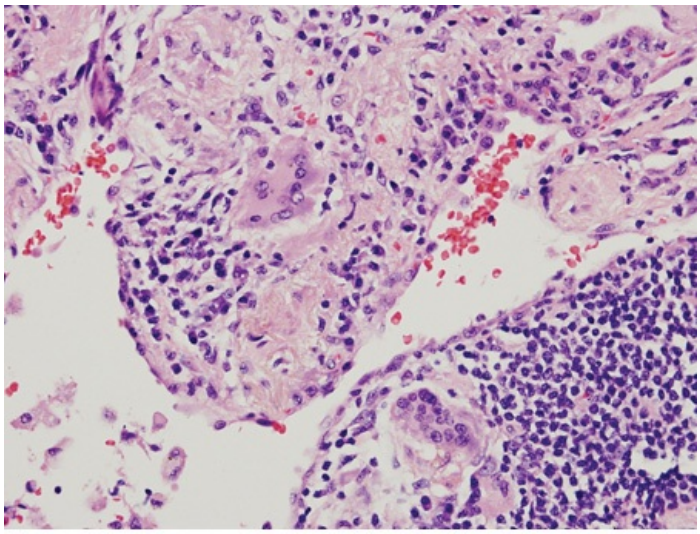

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Open-lung biopsy from a patient with subacute hypersensitivity pneumonitis demonstrating a loose, nonnecrotizing granuloma made up of histiocytes and multinucleated giant cells. Peribronchial inflammatory infiltrate made up of lymphocytes and plasma cells is also seen. *(Courtesy of TJ Gross; with permission.)*

A *prediction rule* for the clinical diagnosis of HP has been developed by the International HP Study Group. Six significant predictors of HP (exposure to a known antigen, positive predictive antibodies to the antigen, recurrent episodes of symptoms, inspiratory crackles, symptoms developing 4–8 h after exposure, and weight loss) were retrospectively developed then validated in a separate cohort. This diagnostic paradigm has a high predictive value in the diagnosis of HP, without the need for invasive testing. In cases where only a subset of the criteria is fulfilled, the diagnosis is less certain. It is clear, however, that the diagnosis of HP is established by (1) consistent symptoms, physical findings, pulmonary function tests, and radiographic tests; (2) a history of exposure to a recognized antigen; and (3) ideally, identification of an antibody to that antigen. Symptoms upon re-exposure to the suspected antigen also support the diagnosis. In some circumstances, BAL and/or lung biopsy may be needed. The most important tool in diagnosing HP continues to be a high index of suspicion.

Differential Diagnosis

Chronic HP may often be difficult to distinguish from a number of other interstitial lung disorders (Chap. 261). A negative history for use of relevant drugs and no evidence of a systemic disorder usually exclude the presence of drug-induced lung disease or a collagen vascular disorder. BAL often shows predominance of neutrophils in idiopathic pulmonary fibrosis and a predominance of CD4+ lymphocytes in sarcoidosis. Hilar/paratracheal lymphadenopathy or evidence of multisystem involvement also favors the diagnosis of sarcoidosis. In some patients, a lung biopsy may be required to differentiate chronic HP from other interstitial diseases. The lung disease associated with acute or subacute HP may clinically resemble other disorders that present with systemic symptoms and recurrent pulmonary infiltrates, including the allergic bronchopulmonary mycoses and other eosinophilic pneumonias. Eosinophilic pneumonia is often associated with asthma and is typified by peripheral eosinophilia; neither of these is a feature of HP. Allergic bronchopulmonary aspergillosis (ABPA) is the most common example of the allergic bronchopulmonary mycoses and is sometimes confused with HP because of the presence of precipitating antibodies to *Aspergillus fumigatus*. ABPA is associated with allergic (atopic) asthma. Acute HP may be confused with *organic dust toxic syndrome* (ODTS), a condition that is more common than HP. ODTS follows heavy exposure to organic dusts and is characterized by transient fever and muscle aches, with or without dyspnea and cough. Serum precipitins are absent, and the chest x-ray is usually normal. This distinction is important, as ODTS is a self-limited disorder without significant long-term sequelae, whereas continued antigen exposure in HP can result in permanent disability. Massive exposure to moldy silage may result in a syndrome termed *pulmonary mycotoxicosis*, with fever, chills, and cough and the presence of pulmonary infiltrates within a few hours of exposure. No previous sensitization is required, and precipitins are absent to *Aspergillus*, the suspected causative agent.

Treatment: Hypersensitivity Pneumonitis

Because effective treatment depends largely on avoiding the antigen, identification of the causative agent and its source is essential. This is usually possible if the physician takes a detailed

environmental and occupational history or, if necessary, visits the patient's environment. The simplest way to avoid the incriminated agent is to remove the patient from the environment or remove the source of the agent from the patient's environment. This recommendation cannot be taken lightly when it completely changes the lifestyle or livelihood of the patient. In many cases, the source of exposure (birds, humidifiers, molds, etc.) can be removed. Pollen masks, personal dust respirators, airstream helmets, and ventilated helmets with a supply of fresh air are increasingly efficient means of purifying inhaled air. If symptoms recur or physiologic abnormalities progress in spite of these measures, more effective measures to avoid antigen exposure must be pursued. The chronic form of HP typically results from low-grade or recurrent exposure over many months to years, and the lung disease may already be partially or completely irreversible. These patients are usually advised to avoid all possible contact with the offending agent.

Patients with the *acute,* recurrent form of HP usually recover without need for glucocorticoids. *Subacute* HP may be associated with severe symptoms and marked physiologic impairment and may continue to progress for several days despite hospitalization. Urgent establishment of the diagnosis and prompt institution of glucocorticoid treatment are indicated in such patients. Prednisone at a dosage of 1 mg/kg per d or its equivalent is continued for 7 to 14 days and then tapered over the ensuing 2 to 6 weeks at a rate that depends on the patient's clinical status. Patients with *chronic* HP may gradually recover without therapy following environmental control. In many patients, however, a trial of prednisone may be useful to obtain maximal reversibility of the lung disease. Following initial prednisone therapy (1 mg/kg per d for 2 to 4 weeks), the drug is tapered to the lowest dosage that will maintain the functional status of the patient. Many patients will not require or benefit from long-term therapy if there is no further exposure to antigen. Although a short course of corticosteroids has been shown to accelerate recovery from the acute stage, glucocorticoid therapy does not appear to have an effect on long-term prognosis of farmer's lung. Improvement of lung function may continue over a few months to years.

PULMONARY INFILTRATES WITH EOSINOPHILIA

Pulmonary infiltrates with eosinophilia (PIE, eosinophilic pneumonias) include distinct individual syndromes characterized by eosinophilic pulmonary infiltrates and, commonly, peripheral blood eosinophilia. Since Loeffler's initial description of a transient, benign syndrome of migratory pulmonary infiltrates and peripheral blood eosinophilia of unknown cause, this group of disorders has been enlarged to include several diseases of both known and unknown etiology **(Table 255-2)**. These diseases may be considered as immunologically mediated lung diseases, but are not to be confused with HP, in which eosinophilia is *not* a feature. In differentiating the etiologies of this heterogeneous group of lung disorders, an extensive history and full examination of all organ systems are essential.

Table 255-2 Pulmonary Infiltrates with Eosinophilia

Etiology known
Allergic bronchopulmonary mycoses
Parasitic infestations
Drug reactions
Eosinophilia-myalgia syndrome
Idiopathic
Loeffler's syndrome
Acute eosinophilic pneumonia
Chronic eosinophilic pneumonia
Allergic granulomatosis of Churg and Strauss
Hypereosinophilic syndrome

When an eosinophilic pneumonia is associated with bronchial asthma, it is important to determine if the patient has atopic asthma and has wheal-and-flare skin reactivity to *Aspergillus* or other relevant fungal antigens. If so, other criteria should be sought for the diagnosis of ABPA **(Table 255-3)** or other, rarer examples of allergic bronchopulmonary mycosis such as those caused by *Penicillium, Candida, Curvularia,* or *Helminthosporium* spp. *A. fumigatus* is the most common cause of ABPA. The chest roentgenogram in ABPA may show transient, recurrent infiltrates or may suggest the presence of proximal bronchiectasis. High-resolution chest CT is a sensitive, noninvasive technique for the recognition of proximal bronchiectasis. The bronchial asthma of ABPA likely involves an IgE-mediated hypersensitivity, whereas the bronchiectasis associated with this disorder is thought to result from a deposition of immune complexes in proximal airways. Adequate treatment usually requires the long-term use of systemic glucocorticoids. Another eosinophilic process associated with asthma is Churg-Strauss syndrome, or allergic angiitis granulomatosis, which presents with necrotizing eosinophilic vasculitis and eosinophilic infiltration of multiple organs, including the lung.

Table 255-3 Diagnostic Features of Allergic Bronchopulmonary Aspergillosis (ABPA)
Main diagnostic criteria
Bronchial asthma
Pulmonary infiltrates
Peripheral eosinophilia (>1000/μL)
Immediate wheal-and-flare response to *Aspergillus fumigatus*
Serum precipitins to *A. fumigatus*
Elevated serum IgE
Central bronchiectasis
Other diagnostic features
History of brownish plugs in sputum
Culture of *A. fumigatus* from sputum
Elevated IgE (and IgG) class antibodies specific for *A. fumigatus*

A travel history or evidence of recent immigration should prompt the consideration of parasite-associated disorders. *Tropical eosinophilia* is usually caused by filarial infection; however, eosinophilic pneumonias also occur with other parasites such as *Ascaris* spp., *Ancyclostoma* spp., *Toxocara* spp., and *Strongyloides stercoralis.* Tropical eosinophilia due to *Wuchereria bancrofti* or *W. malayi* occurs most commonly in southern Asia, Africa, and South America and is treated successfully with diethylcarbamazine. In the United States, *Strongyloides* is endemic to the Southeastern and Appalachian regions. Even in cases of known foreign travel, identification of the causative agent is not always possible, as exemplified by 18 cases (2 fatal) of acute eosinophilic pneumonia reported amongst U.S. military personnel deployed in Iraq.

In the United States, *drug-induced eosinophilic pneumonias* are the most common cause of eosinophilic pulmonary infiltrates. These are exemplified by acute reactions to nitrofurantoin, which may begin 2 h to 10 days after nitrofurantoin is started, with symptoms of dry cough, fever, chills, and dyspnea; an eosinophilic pleural effusion accompanying patchy or diffuse pulmonary infiltrates may also occur. Other drugs associated with eosinophilic pneumonias include sulfonamides, penicillin, chlorpropamide, thiazides, tricyclic antidepressants, hydralazine, gold salts, isoniazid, indomethacin, and others. One report has identified anti-TNF-α monoclonal antibody therapy as a cause of eosinophilic pneumonitis. Treatment consists of withdrawal of the incriminated drugs or toxins and the use of glucocorticoids, if necessary.

The group of primary (idiopathic) eosinophilic pneumonias consists of diseases of varying severity. *Loeffler's syndrome* was originally reported as a benign, acute eosinophilic pneumonia of unknown cause characterized by migrating pulmonary infiltrates and minimal clinical manifestations. In some patients, these clinical characteristics prove to be secondary to parasites or drugs. *Acute eosinophilic pneumonia* is an idiopathic acute febrile illness of <7 days' duration with severe hypoxemia, pulmonary infiltrates, pleural effusions, and no history of asthma. BAL fluid reveals greater than 25% eosinophils (normally less than 2% in nonsmokers); however, the peripheral eosinophilia tends to develop later in the course and may not be apparent on initial presentation. *Chronic eosinophilic pneumonia* presents with significant systemic symptoms including fever, chills, night sweats, cough, anorexia, and weight loss of several weeks' to months' duration. The chest x-ray classically shows peripheral infiltrates, and pulmonary function testing reveals obstruction. Peripheral blood and BAL eosinophilia is more pronounced than in the acute form. Some patients also have bronchial asthma of the intrinsic or nonallergic type. For both acute and chronic disease, dramatic clearing of symptoms and chest x-ray is often noted after initiation of glucocorticoid therapy In contrast to acute eosinophilic pneumonia, chronic eosinophilic pneumonia tends to recur and may require repeated treatment.

The *hypereosinophilic syndrome* is characterized by presence of >1500 eosinophils per microliter of peripheral blood for 6 months or longer; lack of evidence for parasitic, allergic, or other known causes of eosinophilia; and signs or symptoms of multisystem organ dysfunction. Consistent features are blood and bone marrow eosinophilia with tissue infiltration by relatively mature eosinophils. The heart may be involved with tricuspid valve abnormalities or endomyocardial fibrosis and a restrictive, biventricular cardiomyopathy (Chap. 238). Other organs affected typically include the lungs, liver, spleen, skin, and nervous system. Therapy of the disorder consists of glucocorticoids and/or hydroxyurea, plus therapy as needed for cardiac dysfunction, which is frequently responsible for much of the morbidity and mortality in this syndrome. Pulmonary eosinophilia has also been associated with T cell lymphoma and has been reported following *lung* and *bone marrow transplantation*.

GLOBAL PICTURE OF HYPERSENSITIVITY PNEUMONITIS AND PULMONARY INFILTRATES WITH EOSINOPHILIA

HP is more prevalent outside of the United States than within, and the range of antigen responses is somewhat different. Internationally, bird breeder's lung is the most common form of HP. Rather than being associated with avocational exposures, bird-raising practices, highlighted by the emerging threat of avian influenza, lead to substantial exposure to workers involved in poultry husbandry and processing. This increases antigen exposure enormously, in comparison to U.S. workers, and enhances the risk of HP. Importantly, it is the most common cause of pediatric HP and has been reported in individuals as young as 4 years, presenting as a chronic cough.

Farmer's lung, one of the earliest reported causes of HP, appears to be waning worldwide. This is likely in response to changing agricultural practices; increased use of masks and impermeable barriers in hay storage has reduced the exposure and proliferation of thermophilic bacteria, and thus HP.

The international manifestations of HP resemble those of the U.S. disease. Many industrialized nations have increasingly reported HP due to mycobacteria and pseudomonads in contaminated metalworking fluids. The prevalence of these environmental contaminants greatly depends on workplace hygiene practices. Some forms of HP are almost exclusively geographically limited, such as summer-type hypersensitivity pneumonitis in Japan from exposure to *Trichosporon cutaneum* associated with birds. Likewise, cork worker's pneumonitis (suberosis), caused by exposure to contaminated corks, is almost exclusively seen in Spain and southern Europe because of the regional cork industry. However, one of the causative antigens (*Chrysonilia sitophila*) is also reported to be an antigen in lung diseases associated with logging in Canada. In Spain, esparto, a member of the grass family, is used as a fiber for the weaving of mats, baskets, and ropes; it is also incorporated into traditional plaster construction. In both of its uses, it has been associated with HP (most likely due to contamination with *A. fumigatus*), again geographically limited because of the utility of the product, though not of the underlying fungal antigen. Exposure to exotic mushrooms is greater in Asia than in the United States, and this has recently been linked to cases of HP.

Pulmonary infiltrates with eosinophilia are also a greater international than U.S. health burden. In this case, parasitic infestation is far more common than drug-induced lung disease, but the manifestations are similar.

ACKNOWLEDGMENTS

We acknowledge the contribution of Dr. Joel N. Kline to this chapter in the previous edition.

FURTHER READINGS

Allen JN: Drug-induced eosinophilic lung disease. Clin Chest Med 25:77, 2004[PMID: 15062599] [Full Text]

Cooray JH, Ismail MM: Re-examination of the diagnostic criteria of tropical pulmonary eosinophilia. Respir Med 93:655, 1999[PMID: 10542980] [Full Text]

Cormier Y et al: High-resolution computed tomographic characteristics in acute farmer's lung and in its follow-up. Eur Resp J 16:56, 2000[PMID: 10933085] [Full Text]

Fink JN et al: Needs and opportunities for research in hypersensitivity pneumonitis. Am J Respir Crit Care Med 171:792, 2005[PMID: 15657460] [Full Text]

Lacasse Y et al: Clinical diagnosis of hypersensitivity pneumonitis. Am J Respir Crit Care Med 168:962, 2003

Miyazaki Y et al: Clinical predictors and histologic appearance of acute exacerbations in chronic hypersensitivity pneumonitis. Chest 134(6):1265, 2008[PMID: 18689595] [Full Text]

Selman M> et al: Hypersensitivity pneumonitis caused by fungi. Proc Am Thorac Soc 7:229, 2010[PMID: 20463253] [Full Text]

Shorr AF et al: Acute eosinophilic pneumonia among U.S. military personnel deployed in or near Iraq. JAMA 292:2997, 2004[PMID: 15613668] [Full Text]

Silva CI et al: Hypersensitivity pneumonitis: Spectrum of high-resolution CT and pathologic findings. Am J Roentgenol 188:334, 2007[PMID: 17242239] [Full Text]

Zacharisen MC et al: The long-term outcome in acute, subacute, and chronic forms of pigeon breeder's disease hypersensitivity pneumonitis. Ann Allergy Asthma Immunol 88:175, 2002[PMID: 11868922] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OCCUPATIONAL AND ENVIRONMENTAL LUNG DISEASE: INTRODUCTION

Occupational and environmental lung diseases are difficult to distinguish from those of nonenvironmental origin. Virtually all major categories of pulmonary disease can be caused by environmental agents, and environmentally related disease usually presents clinically in a manner indistinguishable from that of disease not caused by such agents. In addition, the etiology of many diseases may be multifactorial; occupational and environmental factors may interact with other factors (such as smoking and genetic risk). It is often only after a careful exposure history is taken that the underlying workplace or general environmental exposure is uncovered.

Why is knowledge of occupational or environmental etiology so important? Patient management and prognosis are affected significantly by such knowledge. For example, patients with occupational asthma or hypersensitivity pneumonitis often cannot be managed adequately without cessation of exposure to the offending agent. Establishment of cause may have significant legal and financial implications for a patient who no longer can work in his or her usual job. Other exposed people may be identified as having the disease or prevented from getting it. In addition, new associations between exposure and disease may be identified (e.g., nylon flock worker's lung disease and diacetyl-induced bronchiolitis obliterans).

Although the exact proportion of lung disease due to occupational and environmental factors is unknown, a large number of individuals are at risk. For example, 15–20% of the burden of adult asthma and chronic obstructive pulmonary disease (COPD) has been estimated to be due to occupational factors.

History and Physical Examination

The patient's history is of paramount importance in assessing any potential occupational or environmental exposure. Inquiry into specific work practices should include questions about the specific contaminants involved, the presence of visible dusts, chemical odors, the size and ventilation of workspaces, the use of respiratory protective equipment, and whether co-workers have similar complaints. The temporal association of exposure at work and symptoms may provide clues to occupation-related disease. In addition, the patient must be questioned about alternative sources of exposure to potentially toxic agents, including hobbies, home characteristics, exposure to secondhand smoke, and proximity to traffic or industrial facilities. Short-term and long-term exposures to potential toxic agents in the distant past also must be considered.

Workers in the United States have the right to know about potential hazards in their workplaces under federal Occupational Safety and Health Administration (OSHA) regulations. Employers must provide specific information about potential hazardous agents in products being used through Material Safety Data Sheets as well as training in personal protective equipment and environmental control procedures. Reminders posted in the workplace may warn workers about hazardous substances. However, the introduction of new processes and/or new chemical compounds may change exposure significantly, and often only the employee on the production line is aware of the change. For the physician caring for a patient with a suspected work-related illness, a visit to the work site can be very instructive. Alternatively, an affected worker can request an inspection by OSHA.

The physical examination of patients with environmentally related lung diseases may help determine the nature and severity of the pulmonary condition but usually does not contribute information that points to a specific etiology.

Pulmonary Function Tests and Chest Imaging

Exposures to inorganic and organic dusts can cause interstitial lung disease that presents with a restrictive pattern and a decreased diffusing capacity (Chap. 252). Similarly, exposures to a number of organic dusts or chemical agents may result in occupational asthma or COPD that is characterized by airway obstruction. Measurement of change in forced expiratory volume (FEV_1) before and after a working shift can be used to detect an acute bronchoconstrictive response. For example, an acute decrement of FEV_1 over the first work shift of the week is a characteristic feature of cotton textile workers with byssinosis (an obstructive airway disorder with features of both asthma and chronic bronchitis).

The chest radiograph is useful in detecting and monitoring the pulmonary response to mineral dusts, certain metals, and organic dusts capable of inducing hypersensitivity pneumonitis. The International

Labour Organisation (ILO) International Classification of Radiographs of Pneumoconioses classifies chest radiographs by the nature and size of opacities seen and the extent of involvement of the parenchyma. In general, small rounded opacities are seen in silicosis or coal worker's pneumoconiosis and small linear opacities are seen in asbestosis. The profusion of such opacities is rated by using a 12-point scheme. Although useful for epidemiologic studies and screening large numbers of workers, the ILO system can be problematic when applied to an individual worker's chest radiograph. With dusts causing rounded opacities, the degree of involvement on the chest radiograph may be extensive, whereas pulmonary function may be only minimally impaired. In contrast, in pneumoconiosis causing linear, irregular opacities like those seen in asbestosis, the radiograph may lead to underestimation of the severity of the impairment until relatively late in the disease. For patients with a history of asbestos exposure, conventional computed tomography (CT) is more sensitive for the detection of pleural thickening and high-resolution CT (HRCT) improves the detection of asbestosis.

Other procedures that may be of use in identifying the role of environmental exposures in causing lung disease include evaluation of heavy metal concentrations in urine (cadmium in battery plant workers), skin prick testing or specific IgE antibody titers for evidence of immediate hypersensitivity to agents capable of inducing occupational asthma (flour antigens in bakers), specific IgG precipitating antibody titers for agents capable of causing hypersensitivity pneumonitis (pigeon antigen in bird handlers), and assays for specific cell-mediated immune responses (beryllium lymphocyte proliferation testing in nuclear workers or tuberculin skin testing in health care workers). Sometimes a bronchoscopy to obtain transbronchial biopsies of lung tissue may be required for histologic diagnosis (chronic beryllium disease). Rarely, video-assisted thoracoscopic surgery to obtain a larger sample of lung tissue may be required to determine the specific diagnosis of environmentally induced lung disease (hypersensitivity pneumonitis or giant cell interstitial pneumonitis due to cobalt exposure).

Exposure Assessment

If reliable environmental sampling data are available, that information should be used in assessing a patient's exposure. Since many of the chronic diseases result from exposure over many years, current environmental measurements should be combined with work histories to arrive at estimates of past exposure.

In situations in which individual exposure to specific agents—either in a work setting or via ambient air pollutants—has been determined, the chemical and physical characteristics of those agents affect both the inhaled dose and the site of deposition in the respiratory tract. Water-soluble gases such as ammonia and sulfur dioxide are absorbed in the lining fluid of the upper and proximal airways and thus tend to produce irritative and bronchoconstrictive responses. In contrast, nitrogen dioxide and phosgene, which are less soluble, may penetrate to the bronchioles and alveoli in sufficient quantities to produce acute chemical pneumonitis that can be life-threatening.

Particle size of air contaminants must also be considered. Because of their settling velocities in air, particles >10–15 μm in diameter do not penetrate beyond the nose and throat. Particles <10 μm in size are deposited below the larynx. These particles are divided into three size fractions on the basis of their size characteristics and sources. Particles ~2.5–10 μm (coarse-mode fraction) contain crustal elements such as silica, aluminum, and iron. These particles mostly deposit relatively high in the tracheobronchial tree. Although the total mass of an ambient sample is dominated by these larger respirable particles, the number of particles, and therefore the surface area on which potential toxic agents can deposit and be carried to the lower airways, is dominated by particles <2.5 μm (fine-mode fraction). These fine particles are created primarily by the burning of fossil fuels or high-temperature industrial processes resulting in condensation products from gases, fumes, or vapors. The smallest particles, those <0.1 μm in size, represent the ultrafine fraction and make up the largest number of particles; they tend to remain in the airstream and deposit in the lung only on a random basis as they come into contact with the alveolar walls. If they do deposit, however, particles of this size range may penetrate into the circulation and be carried to extrapulmonary sites. New technologies create particles of this size ("nanoparticles") for use in many commercial applications. Besides the size characteristics of particles and the solubility of gases, the actual chemical composition, mechanical properties, and immunogenicity or infectivity of inhaled material determine in large part the nature of the diseases found among exposed persons.

OCCUPATIONAL EXPOSURES AND PULMONARY DISEASE

Table 256-1 provides broad categories of exposure in the workplace and diseases associated with chronic exposure in those industries.

Table 256-1 Categories of Occupational Exposure and Associated Respiratory Conditions

Occupational Exposures	Nature of Respiratory Responses	Comment
Inorganic Dusts		
Asbestos: mining, processing,	Fibrosis (asbestosis),	Virtually all new mining and

construction, ship repair	pleural disease, cancer, mesothelioma	construction with asbestos done in developing countries
Silica: mining, stone cutting, sandblasting, quarrying	Fibrosis (silicosis), progressive massive fibrosis (PMF), cancer, tuberculosis, chronic obstructive pulmonary disease (COPD)	Improved protection in United States, persistent risk in developing countries
Coal dust: mining	Fibrosis (coal worker's pneumoconiosis), PMF, COPD	Risk persists in certain areas of United States, increasing in countries where new mines open
Beryllium: processing alloys for high-tech industries	Acute pneumonitis (rare), chronic granulomatous disease, lung cancer (highly suspect)	Risk in high-tech industries persists
Other metals: aluminum, chromium, cobalt, nickel, titanium, tungsten carbide, or "hard metal" (contains cobalt)	Wide variety of conditions from acute pneumonitis to lung cancer and asthma	New diseases appear with new process development
Organic Dusts		
Cotton dust: milling, processing	Byssinosis (an asthma-like syndrome), chronic bronchitis, COPD	Increasing risk in developing countries with drop in United States as jobs shift overseas
Grain dust: elevator agents, dock workers, milling, bakers	Asthma, chronic bronchitis, COPD	Risk shifting more to migrant labor pool
Other agricultural dusts: fungal spores, vegetable products, insect fragments, animal dander, bird and rodent feces, endotoxins, microorganisms, pollens	Hypersensitivity pneumonitis (farmer's lung), asthma, chronic bronchitis	Important in migrant labor pool but also resulting from in-home exposures
Toxic chemicals: wide variety of industries, see Table 256-2	Asthma, chronic bronchitis, COPD, hypersensitivity pneumonitis, pneumoconiosis, and cancer	Reduced risk with recognized hazards; increasing risk for developing countries where controlled labor practices are less stringent
Other respiratory environmental agents: uranium and radon daughters, secondhand tobacco smoke, polycyclic hydrocarbons, biomass smoke, diesel exhaust, welding fumes, wood finishing	Occupational exposures estimated to contribute to up to 10% of all lung cancers; chronic bronchitis, COPD, and fibrosis	In-home exposures important; in developing countries biomass smoke is a major risk factor for COPD among women

Asbestos-Related Diseases

Asbestos is a generic term for several different mineral silicates, including chrysolite, amosite, anthophyllite, and crocidolite. In addition to workers involved in the production of asbestos products (mining, milling, and manufacturing), many workers in the shipbuilding and construction trades, including pipe fitters and boilermakers, were occupationally exposed because asbestos was widely used during the twentieth century for its thermal and electrical insulation properties. Asbestos also was used in the manufacture of fire-resistant textiles, in cement and floor tiles, and in friction materials such as brake and clutch linings.

Exposure to asbestos is not limited to persons who directly handle the material. Cases of asbestos-related diseases have been encountered in individuals with only bystander exposure, such as painters and electricians who worked alongside insulation workers in a shipyard. Community exposure resulted from the use of asbestos-containing mine and mill tailings as landfill, road surface, and playground material (e.g., Libby, MT, the site of a vermiculite mine in which the ore was contaminated with asbestos). Finally, exposure can occur from the disturbance of naturally occurring asbestos (e.g., from increasing residential development in the foothills of the Sierra Mountains in California).

Asbestos has largely been replaced in the developed world with synthetic mineral fibers such as fiberglass and refractory ceramic fibers, but it continues to be used increasingly in the developing world. Despite current OSHA regulations mandating adequate training for any worker potentially exposed to asbestos, exposure continues among inadequately trained and protected demolition workers. The major health effects from exposure to asbestos are pleural and pulmonary fibrosis, cancers of the respiratory tract, and pleural and peritoneal mesothelioma.

Asbestosis is a diffuse interstitial fibrosing disease of the lung that is directly related to the intensity and duration of exposure. The disease resembles other forms of diffuse interstitial fibrosis (Chap.

261). Usually, moderate to severe exposure has taken place for at least 10 years before the disease becomes manifest, and it may occur after exposure to any of the asbestiform fiber types. The mechanisms by which asbestos fibers induce lung fibrosis are not completely understood but are known to involve oxidative injury due to the generation of reactive oxygen species by the transition metals on the surface of the fibers as well as from cells engaged in phagocytosis.

The chest radiograph can be used to detect the pulmonary manifestations of asbestos exposure. Past exposure is specifically indicated by pleural plaques, which are characterized by either thickening or calcification along the parietal pleura, particularly along the lower lung fields, the diaphragm, and the cardiac border. Without additional manifestations, pleural plaques imply only exposure, not pulmonary impairment. Benign pleural effusions also may occur. The fluid is typically a serous or bloody exudate. The effusion may be slowly progressive or may resolve spontaneously. Irregular or linear opacities, evidence of asbestosis that usually are first noted in the lower lung fields and spreading into the middle and upper lung fields, occur as the disease progresses. An indistinct heart border or a "ground-glass" appearance in the lung fields is seen in some cases. In cases in which the x-ray changes are less obvious, HRCT may show distinct changes of subpleural curvilinear lines 5–10 mm in length that appear to be parallel to the pleural surface **(Fig. 256-1)**.

Figure 256-1

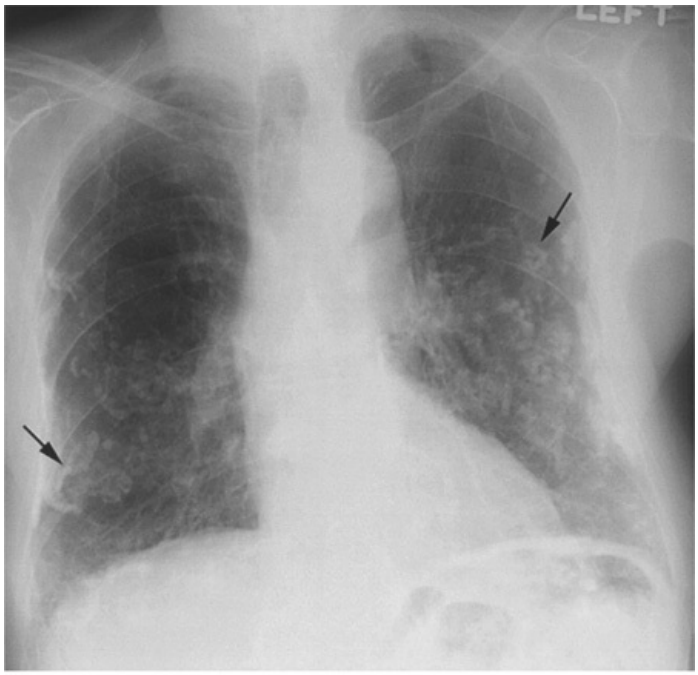

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

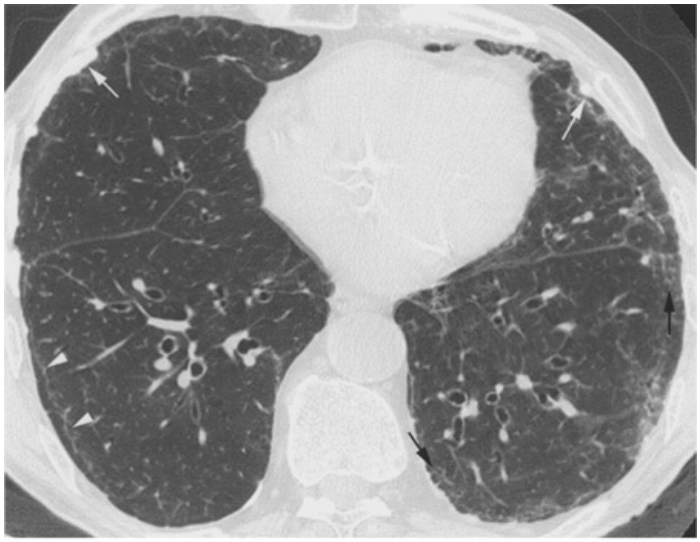

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Asbestosis: ***A.*** Frontal chest radiograph shows bilateral calcified pleural plaques consistent with asbestos-related pleural disease. Poorly defined linear and reticular abnormalities are seen in the lower lobes bilaterally. ***B.*** Axial high-resolution computed tomography of the thorax obtained through the lung bases shows bilateral, subpleural reticulation (*black arrows*), representing fibrotic lung disease due to asbestosis. Subpleural lines are also present (*arrowheads*), characteristic of, though not specific for, asbestosis. Calcified pleural plaques representing asbestos-related pleural disease (*white arrows*) are also evident.

Pulmonary function testing in asbestosis reveals a restrictive pattern with a decrease in both lung volumes and diffusing capacity. There may also be evidence of mild airflow obstruction (due to peribronchiolar fibrosis).

No specific therapy is available for the management of patients with asbestosis. The supportive care is the same as that given to any patient with diffuse interstitial fibrosis of any cause. In general, newly diagnosed cases will have resulted from exposures that occurred many years before.

Lung cancer (Chap. 89) is the most common cancer associated with asbestos exposure. The excess frequency of lung cancer (all histologic types) in asbestos workers is associated with a minimum latency of 15–19 years between first exposure and development of the disease. Persons with more exposure are at greater risk of disease. In addition, there is a significant interactive effect of smoking and asbestos exposure that results in greater risk than what would be expected from the additive effect of each factor.

Mesotheliomas (Chap. 263), both pleural and peritoneal, are also associated with asbestos exposure. In contrast to lung cancers, these tumors do not appear to be associated with smoking. Relatively short-term asbestos exposures of ≤1–2 years or less, occurring up to 40 years in the past, have been associated with the development of mesotheliomas (an observation that emphasizes the importance of obtaining a complete environmental exposure history). Although the risk of mesothelioma is much less than that of lung cancer among asbestos-exposed workers, over 2000 cases were reported in the United States per year at the start of the twenty-first century.

Although ~50% of mesotheliomas metastasize, the tumor generally is locally invasive, and death usually results from local extension. Most patients present with effusions that may obscure the underlying pleural tumor. In contrast to the findings in effusion due to other causes, because of the restriction placed on the chest wall, no shift of mediastinal structures toward the opposite side of the chest will be seen. The major diagnostic problem is differentiation from peripherally spreading pulmonary adenocarcinoma or adenocarcinoma that has metastasized to pleura from an extrathoracic primary site. Although cytologic examination of pleural fluid may suggest the diagnosis, biopsy of pleural tissue, generally with video-assisted thoracic surgery, and special immunohistochemical staining usually are required. There is no effective therapy.

Since epidemiologic studies have shown that >80% of mesotheliomas may be associated with asbestos exposure, documented mesothelioma in a patient with occupational or environmental exposure to asbestos may be compensable.

Silicosis

In spite of being one of the oldest known occupational pulmonary hazards, *free silica* (SiO_2), or crystalline quartz, is still a major cause of disease. The major occupational exposures include mining; stonecutting; employment in abrasive industries such as stone, clay, glass, and cement manufacturing;

foundry work; packing of silica flour; and quarrying, particularly of granite. Most often, pulmonary fibrosis due to silica exposure (silicosis) occurs in a dose-response fashion after many years of exposure.

Workers heavily exposed through sandblasting in confined spaces, tunneling through rock with a high quartz content (15–25%), or the manufacture of abrasive soaps may develop acute silicosis with as little as 10 months of exposure. The clinical and pathologic features of acute silicosis are similar to those of pulmonary alveolar proteinosis (Chap. 261). The chest radiograph may show profuse miliary infiltration or consolidation, and there is a characteristic HRCT pattern known as "crazy paving" **(Fig. 256-2)**. The disease may be quite severe and progressive despite the discontinuation of exposure. Whole-lung lavage may provide symptomatic relief and slow the progression.

Figure 256-2

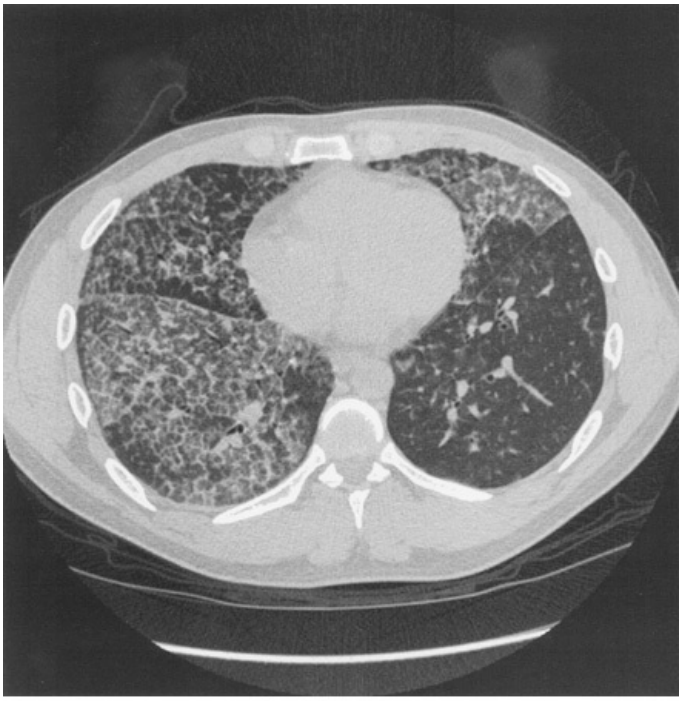

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute silicosis. This high-resolution computed tomography scan shows multiple small nodules consistent with silicosis but also diffuse ground-glass densities with thickened intralobular and interlobular septa producing polygonal shapes. This has been referred to as "crazy paving."

With long-term, less intense exposure, small rounded opacities in the upper lobes may appear on the chest radiograph after 15–20 years of exposure (*simple silicosis*). Calcification of hilar nodes may occur in as many as 20% of cases and produces a characteristic "eggshell" pattern. Silicotic nodules may be identified more readily by HRCT **(Fig. 256-3)**. The nodular fibrosis may be progressive in the absence of further exposure, with coalescence and formation of nonsegmental conglomerates of irregular masses >1 cm in diameter (*complicated silicosis*). These masses can become quite large, and when this occurs, the term *progressive massive fibrosis* (PMF) is applied. Significant functional impairment with both restrictive and obstructive components may be associated with this form of silicosis.

Figure 256-3

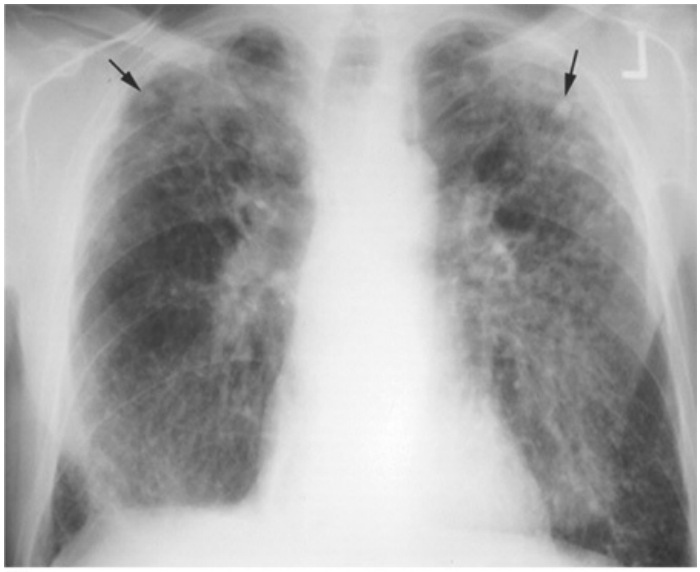

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

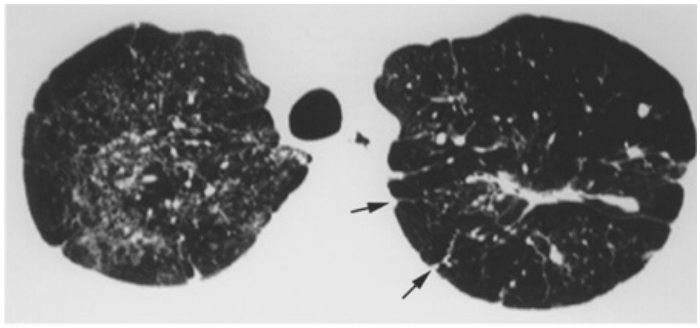

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic silicosis. *A.* Frontal chest radiograph in a patient with silicosis shows variably sized, poorly defined nodules (*arrows*) predominating in the upper lobes, ***B.*** Axial thoracic computed tomography image through the lung apices shows numerous small nodules, more pronounced in the right upper lobe. A number of the nodules are subpleural in location (*arrows*).

Because silica is cytotoxic to alveolar macrophages, patients with silicosis are at greater risk of acquiring lung infections that involve these cells as a primary defense (*Mycobacterium tuberculosis*, atypical mycobacteria and fungi). Because of the increased risk of active tuberculosis, the recommended treatment of latent tuberculosis in these patients is longer. Another potential clinical complication of silicosis is autoimmune connective tissue disorders such as rheumatoid arthritis and scleroderma. In addition, there are sufficient epidemiologic data that the International Agency for Research on Cancer lists silica as a probable lung carcinogen.

Other, less hazardous silicates include fuller's earth, kaolin, mica, diatomaceous earths, silica gel, soapstone, carbonate dusts, and cement dusts. The production of fibrosis in workers exposed to these agents is believed to be related either to the free silica content of these dusts or, for substances that contain no free silica, to the potentially large dust loads to which these workers may be exposed.

Other silicates, including *talc dusts*, may be contaminated with asbestos and/or free silica. Fibrosis and/or pleural or lung cancer have been associated with chronic exposure to commercial talc.

Coal Worker's Pneumoconiosis (CWP)

Occupational exposure to *coal dust* can lead to CWP, which has enormous social, economic, and medical significance in every nation in which coal mining is an important industry. Simple radiographically identified CWP is seen in ~10% of all coal miners and in as many as 50% of anthracite miners with more than 20 years' work on the coal face. The prevalence of disease is lower in workers in bituminous coal mines.

With prolonged exposure to coal dust (i.e., 15–20 years), small, rounded opacities similar to those of silicosis may develop. As in silicosis, the presence of these nodules (*simple CWP*) usually is not associated with pulmonary impairment. Much of the symptomatology associated with simple CWP

appears to be due to the effects of coal dust on the development of chronic bronchitis and COPD (Chap. 260). The effects of coal dust are additive to those of cigarette smoking.

Complicated CWP is manifested by the appearance on the chest radiograph of nodules ranging from 1 cm in diameter to the size of an entire lobe, generally confined to the upper half of the lungs. As in silicosis, this condition can progress to PMF that is accompanied by severe lung function deficits and associated with premature mortality. Despite improvements in technology to protect coal miners, cases of PMF still occur in the United States at a disturbing rate.

Caplan's syndrome (Chap. 321), first described in coal miners but subsequently found in patients with silicosis, includes seropositiverheumatoid arthritis with characteristic pneumoconiotic nodules. Silica has immunoadjuvant properties and is often present in anthracitic coal dust.

Chronic Beryllium Disease

Beryllium is a lightweight metal with tensile strength that has good electrical conductivity and is valuable in the control of nuclear reactions through its ability to quench neutrons. Although beryllium may produce an acute pneumonitis, it is far more commonly associated with a chronic granulomatous inflammatory disease that is similar to sarcoidosis (Chap. 329). Unless one inquires specifically about occupational exposures to beryllium in the manufacture of alloys, ceramics, or high-technology electronics in a patient with sarcoidosis, one may miss entirely the etiologic relationship to the occupational exposure. What distinguishes chronic beryllium disease (CBD) from sarcoidosis is evidence of a specific cell-mediated immune response (i.e., delayed hypersensitivity) to beryllium.

The test that usually provides this evidence is the beryllium lymphocyte proliferation test (BeLPT). The BeLPT compares the in vitro proliferation of lymphocytes from blood or bronchoalveolar lavage in the presence of beryllium salts with that of unstimulated cells. Proliferation is usually measured by lymphocyte uptake of radiolabeled thymidine.

Chest imaging findings are similar to those of sarcoidosis (nodules along septal lines) except that hilar adenopathy is somewhat less common. As with sarcoidosis, pulmonary function test results may show restrictive and/or obstructive ventilatory deficits and decreased diffusing capacity. With early disease, both chest imaging studies and pulmonary function tests may be normal. Fiberoptic bronchoscopy with transbronchial lung biopsy usually is required to make the diagnosis of CBD. In a beryllium-sensitized individual, the presence of noncaseating granulomas or monocytic infiltration in lung tissue establishes the diagnosis. Accumulation of beryllium-specific CD4+ T cells occurs in the granulomatous inflammation seen on lung biopsy. CBD is one of the best studied examples of gene-environment interaction. Susceptibility to CBD is highly associated with human leukocyte antigen DP (HLA-DP) alleles that have a glutamic acid in position 69 of the β chain.

Other metals, including aluminum and titanium dioxide, have been rarely associated with a sarcoid-like reaction in lung tissue. Exposure to dust containing tungsten carbide, also known as "hard metal," may produce giant cell interstitial pneumonitis. Cobalt is a constituent of tungsten carbide and is the likely etiologic agent of both the interstitial pneumonitis and the occupational asthma that may occur. The most common exposures to tungsten carbide occur in tool and dye, saw blade, and drill bit manufacture. Diamond polishing may also involve exposure to cobalt dust. The same Glu69 polymorphism of the HLA-DP β chain that confers increased risk of CBD also appears to increase the risk of cobalt-induced giant cell interstitial pneumonitis.

In patients with interstitial lung disease, one should always inquire about exposure to metal fumes and/or dusts. Especially when sarcoidosis appears to be the diagnosis, one should always consider possible CBD.

Other Inorganic Dusts

Most of the inorganic dusts discussed thus far are associated with the production of either dust macules or interstitial fibrotic changes in the lung. Other inorganic and organic dusts (see categories in Table 256-1), along with some of the dusts previously discussed, are associated with chronic mucus hypersecretion (chronic bronchitis), with or without reduction of expiratory flow rates. Cigarette smoking is the major cause of these conditions, and any effort to attribute some component of the disease to occupational and environmental exposures must take cigarette smoking into account. Most studies suggest an additive effect of dust exposure and smoking. The pattern of the irritant dust effect is similar to that of cigarette smoking, suggesting that small airway inflammation may be the initial site of pathologic response in those cases and continued exposure may lead to chronic bronchitis and COPD.

Organic Dusts

Some of the specific diseases associated with organic dusts are discussed in detail in the chapters on asthma (Chap. 254) and hypersensitivity pneumonitis (Chap. 255). Many of these diseases are named for the specific setting in which they are found, e.g., farmer's lung, malt worker's disease, and mushroom worker's disease. Often the temporal relation of symptoms to exposure furnishes the best evidence for the diagnosis. Three occupational exposures are singled out for discussion here because they affect the largest proportions of workers.

COTTON DUST (BYSSINOSIS)

Workers occupationally exposed to cotton dust (but also to flax, hemp, or jute dust) in the production of yarns for textiles and rope making are at risk for an asthma-like syndrome known as byssinosis. Exposure occurs throughout the manufacturing process but is most pronounced in the portions of the factory involved with the treatment of the cotton before spinning, i.e., blowing, mixing, and carding (straightening of fibers). The risk of byssinosis is associated with both cotton dust and endotoxin levels in the workplace environment.

Byssinosis is characterized clinically as occasional (early-stage) and then regular (late-stage) chest tightness toward the end of the first day of the workweek ("Monday chest tightness"). In epidemiologic studies, depending on the level of exposure via the carding room air, up to 80% of employees may show a significant drop in FEV_1 over the course of a Monday shift.

Initially the symptoms do not recur on subsequent days of the week. However, in 10–25% of workers, the disease may be progressive, with chest tightness recurring or persisting throughout the workweek. After >10 years of exposure, workers with recurrent symptoms are more likely to have an obstructive pattern on pulmonary function testing. The highest grades of impairment generally are seen in smokers.

Reduction of dust exposure is of primary importance to the management of byssinosis. Dust levels can be controlled by the use of exhaust hoods, general increases in ventilation, and wetting procedures, but respiratory protective equipment appears to be required during certain operations to prevent workers from being exposed to levels of cotton dust that exceed the current OSHA-permissible exposure level. Regular surveillance of pulmonary function in cotton dust–exposed workers using spirometry before and after the workshift is required by OSHA. All workers with persistent symptoms or significantly reduced levels of pulmonary function should be moved to areas of lower risk of exposure.

GRAIN DUST

Worldwide, many farmers and workers in grain storage facilities are exposed to grain dust. The presentation of obstructive airway disease in grain dust–exposed workers is virtually identical to the characteristic findings in cigarette smokers, i.e., persistent cough, mucus hypersecretion, wheeze and dyspnea on exertion, and reduced FEV_1 and FEV_1/FVC (forced vital capacity) ratio (Chap. 252).

Dust concentrations in grain elevators vary greatly but can be >10,000 $\mu g/m^3$; approximately one-third of the particles, by weight, are in the respirable range. The effect of grain dust exposure is additive to that of cigarette smoking, with ~50% of workers who smoke having symptoms. Among nonsmoking grain elevator operators, approximately one-quarter have mucus hypersecretion, about five times the number that would be expected in unexposed nonsmokers. Smoking grain dust–exposed workers are more likely to have obstructive ventilatory deficits on pulmonary function testing. As in byssinosis, endotoxin may play a role in grain dust–induced chronic bronchitis and COPD.

FARMER'S LUNG

This condition results from exposure to moldy hay containing spores of thermophilic actinomycetes that produce a hypersensitivity pneumonitis (Chap. 255). A patient with acute farmer's lung presents 4–8 h after exposure with fever, chills, malaise, cough, and dyspnea without wheezing. The history of exposure is obviously essential to distinguish this disease from influenza or pneumonia with similar symptoms. In the chronic form of the disease, the history of repeated attacks after similar exposure is important in differentiating this syndrome from other causes of patchy fibrosis (e.g., sarcoidosis).

A wide variety of other organic dusts are associated with the occurrence of hypersensitivity pneumonitis (Chap. 255). For patients who present with hypersensitivity pneumonitis, specific and careful inquiry about occupations, hobbies, and other home environmental exposures is necessary to uncover the source of the etiologic agent.

Toxic Chemicals

Exposure to toxic chemicals affecting the lung generally involves gases and vapors. A common accident is one in which the victim is trapped in a confined space where the chemicals have accumulated to toxic levels. In addition to the specific toxic effects of the chemical, the victim often sustains considerable anoxia, which can play a dominant role in determining whether the individual survives.

Table 256-2 lists a variety of toxic agents that can produce acute and sometimes life-threatening reactions in the lung. All these agents in sufficient concentrations have been demonstrated, at least in animal studies, to affect the lower airways and disrupt alveolar architecture, either acutely or as a result of chronic exposure. Some of these agents may be generated acutely in the environment (see below).

Table 256-2 Selected Common Toxic Chemical Agents that Affect the Lung

Agent(s)	Selected Exposures	Acute Effects from High or Accidental Exposure	Chronic Effects from Relatively Low Exposure
Acid anhydrides	Manufacture of resin esters, polyester resins, thermoactivated adhesives	Nasal irritation, cough	Asthma, chronic bronchitis, hypersensitivity pneumonitis
Acid fumes: H_2SO_4, HNO_3	Manufacture of fertilizers, chlorinated organic compounds, dyes, explosives, rubber products, metal etching, plastics	Mucous membrane irritation, followed by chemical pneumonitis 2–3 days later	Bronchitis and suggestion of mildly reduced pulmonary function in children with lifelong residential exposure to high levels
Acrolein and other aldehydes	By-product of burning plastics, woods, tobacco smoke	Mucous membrane irritant, decrease in lung function	Upper respiratory tract irritation
Ammonia	Refrigeration; petroleum refining; manufacture of fertilizers, explosives, plastics, and other chemicals	Same as for acid fumes, but bronchiectasis also has been reported	Upper respiratory tract irritation, chronic bronchitis
Cadmium fumes	Smelting, soldering, battery production	Mucous membrane irritant, acute respiratory distress syndrome (ARDS)	Chronic obstructive pulmonary disease (COPD)
Formaldehyde	Manufacture of resins, leathers, rubber, metals, and woods; laboratory workers, embalmers; emission from urethane foam insulation	Same as for acid fumes	Nasopharyngeal cancer
Halides and acid salts (Cl, Br, F)	Bleaching in pulp, paper, textile industry; manufacture of chemical compounds; synthetic rubber, plastics, disinfectant, rocket fuel, gasoline	Mucous membrane irritation, pulmonary edema; possible reduced FVC 1–2 yrs after exposure	Upper respiratory tract irritation, epistaxis, tracheobronchitis
Hydrogen sulfide	By-product of many industrial processes, oil, other petroleum processes and storage	Increase in respiratory rate followed by respiratory arrest, lactic acidosis, pulmonary edema, death	Conjunctival irritation, chronic bronchitis, recurrent pneumonitis
Isocyanates (TDI, HDI, MDI)	Production of polyurethane foams, plastics, adhesives, surface coatings	Mucous membrane irritation, dyspnea, cough, wheeze, pulmonary edema	Upper respiratory tract irritation, cough, asthma, hypersensitivity pneumonitis, reduced lung function
Nitrogen dioxide	Silage, metal etching, explosives, rocket fuels, welding, by-product of burning fossil fuels	Cough, dyspnea, pulmonary edema may be delayed 4–12 h; possible result from acute exposure: bronchiolitis obliterans in 2–6 wks	Emphysema in animals, ?chronic bronchitis, associated with reduced lung function in children with lifelong residential exposure
Ozone	Arc welding, flour bleaching, deodorizing, emissions from copying equipment, photochemical air pollutant	Mucous membrane irritant, pulmonary hemorrhage and edema, reduced pulmonary function transiently in children and adults, and increased hospitalization with exposure to summer haze	Excess cardiopulmonary mortality rates
Phosgene	Organic compound, metallurgy, volatilization of chlorine-containing compounds	Delayed onset of bronchiolitis and pulmonary edema	Chronic bronchitis
Sulfur dioxide	Manufacture of sulfuric acid, bleaches, coating of nonferrous metals, food processing, refrigerant, burning of fossil fuels.	Mucous membrane irritant, epistaxis, bronchospasm (especially in people with asthma)	Chronic bronchitis

	wood pulp industry	

Firefighters and fire victims are at risk of *smoke inhalation*, an important cause of acute cardiorespiratory failure. Smoke inhalation kills more fire victims than does thermal injury. Carbon monoxide poisoning with resulting significant hypoxemia can be life-threatening (Chap. e50). Synthetic materials (plastic, polyurethanes), when burned, may release a variety of other toxic agents (such as cyanide and hydrochloric acid), and this must be considered in evaluating smoke inhalation victims. Exposed victims may have some degree of lower respiratory tract inflammation and/or pulmonary edema.

Exposure to certain highly reactive, low-molecular-weight agents used in the manufacture of synthetic polymers, paints, and coatings (*diisocyanates* in polyurethanes, *aromatic amines* and *acid anhydrides* in epoxies) are associated with a high risk of occupational asthma. Although this occupational asthma manifests clinically as if sensitization has occurred, an IgE antibody–mediated mechanism is not necessarily involved. Hypersensitivity pneumonitis–like reactions also have been described in diisocyanate and acid anhydride–exposed workers.

Fluoropolymers such as Teflon, which at normal temperatures produce no reaction, become volatilized upon heating. The inhaled agents cause a characteristic syndrome of fever, chills, malaise, and occasionally mild wheezing, leading to the diagnosis of *polymer fume fever* . A similar self-limited, influenza-like syndrome— *metal fume fever* —results from acute exposure to fumes or smoke containing zinc oxide. The syndrome may begin several hours after work and resolves within 24 h, only to return on repeated exposure. Welding of galvanized steel is the most common exposure leading to metal fume fever.

Two other agents have been recently associated with potentially severe interstitial lung disease. Occupational exposure to nylon flock has been shown to induce a lymphocytic bronchiolitis, and workers exposed to diacetyl used to provide "butter" flavor in the manufacture of microwave popcorn and other foods have developed bronchiolitis obliterans (Chap. 261).

WORLD TRADE CENTER DISASTER

A consequence of the attack on the World Trade Center (WTC) on September 11, 2001, was relatively heavy exposure of a large number of firefighters and other rescue workers to the dust generated by the collapse of the buildings. Environmental monitoring and chemical characterization of WTC dust has revealed a wide variety of potentially toxic constituents, although much of the dust was pulverized cement. Possibly because of the high alkalinity of WTC dust, significant cough, wheeze, and phlegm production occurred among firefighters and cleanup crews. New cough and wheeze syndromes also occurred among local residents. Initial longitudinal follow-up of New York firefighters suggests that heavier exposure to WTC dust is associated with accelerated decline of lung function. Ongoing follow-up will provide data on whether massive exposure to this irritant dust has led to the development of chronic respiratory disease.

Occupational Respiratory Carcinogens

Exposures at work have been estimated to contribute to 10% of all lung cancer cases. In addition to asbestos, other agents either proven or suspected to be respiratory carcinogens include acrylonitrile, arsenic compounds, beryllium, bis(chloromethyl) ether, chromium (hexavalent), formaldehyde (nasal), isopropanol (nasal sinuses), mustard gas, nickel carbonyl (nickel smelting), polyaromatic hydrocarbons (coke oven emissions and diesel exhaust), secondhand tobacco smoke, silica (both mining and processing), talc (possible asbestos contamination in both mining and milling), vinyl chloride (sarcomas), wood (nasal cancer only), and uranium. Workers at risk of radiation-related lung cancer include not only those involved in mining or processing uranium but also those exposed in underground mining operations of other ores where radon daughters may be emitted from rock formations.

Assessment of Disability

Patients who have lung disease may not be able to continue to work in their usual jobs because of respiratory symptoms. *Disability* is the term used to describe the decreased ability to work due to the effects of a medical condition. Physicians are generally able to assess physiologic dysfunction, or *impairment*, but the rating of disability for compensation of loss of income also involves nonmedical factors such as the education and employability of the individual. The disability rating scheme differs with the compensation-granting agency. For example, the U.S. Social Security Administration requires that an individual be unable to do any work (i.e., *total* disability) before he or she will receive income replacement payments. Many state workers' compensation systems allow for payments for *partial* disability. In the Social Security scheme no determination of cause is done, whereas work-relatedness must be established in workers' compensation systems.

For respiratory impairment rating, resting pulmonary function tests (spirometry and diffusing capacity) are used as the initial assessment tool, with cardiopulmonary exercise testing (to assess maximal oxygen consumption) used if the results of the resting tests do not correlate with the patient's

symptoms. Methacholine challenge (to assess airway reactivity) can also be useful in patients with asthma who have normal spirometry when evaluated. Some compensation agencies (e.g., Social Security) have proscribed disability classification schemes based on pulmonary function test results. When no specific scheme is proscribed, the *Guidelines of the American Medical Association* should be used.

Evaluating relation to work exposure requires a detailed work history, as previously discussed in this chapter. Occasionally, as with some cases of suspected occupational asthma, challenge to the putative agent in the work environment with repeated pulmonary function measures may be required.

GENERAL ENVIRONMENTAL EXPOSURES

Outdoor Air Pollution

In 1971, the U.S. government established national air quality standards for several pollutants believed to be responsible for excess cardiorespiratory diseases. Primary standards regulated by the U.S. Environmental Protection Agency (EPA) designed to protect the public health with an adequate margin of safety exist for sulfur dioxide, particulates matter, nitrogen dioxide, ozone, lead, and carbon monoxide. Standards for each of these pollutants are updated regularly through an extensive review process conducted by the EPA. (For details on current standards, go to *http://www.epa.gov/air/criteria.html*.)

Pollutants are generated from both stationary sources (power plants and industrial complexes) and mobile sources (automobiles), and none of the regulated pollutants occurs in isolation. Furthermore, pollutants may be changed by chemical reactions after being emitted. For example, sulfur dioxide and particulate matter emissions from a coal-fired power plant may react in air to produce acid sulfates and aerosols, which can be transported long distances in the atmosphere. Oxidizing substances such as oxides of nitrogen and volatile organic compounds from automobile exhaust may react with sunlight to produce ozone. Although originally thought to be confined to Los Angeles, photochemically derived pollution ("smog") is now known to be a problem throughout the United States and in many other countries. Both acute and chronic effects of these exposures have been documented in large population studies.

The symptoms and diseases associated with air pollution are the same as conditions commonly associated with cigarette smoking. In addition, decreased growth of lung function and asthma have been associated with chronic exposure to only modestly elevated levels of traffic-related gases and respirable particles. Multiple population-based time-series studies within cities have demonstrated excess health care utilization for asthma and other cardiopulmonary conditions and mortality rates. Cohort studies comparing cities that have relatively high levels of particulate exposures with less polluted communities suggest excess morbidity and mortality rates from cardiopulmonary conditions in long-term residents of the former. The strong epidemiologic evidence that fine particulate matter is a risk factor for cardiovascular morbidity and mortality has prompted toxicologic investigations into the underlying mechanisms. The inhalation of fine particles from combustion sources probably generates oxidative stress followed by local injury and inflammation in the lungs that in turn lead to autonomic and systemic inflammatory responses that can induce endothelial dysfunction and/or injury. Recent research findings on the health effects of air pollutants have led to stricter U.S. ambient air quality standards for ozone, oxides of nitrogen, and particulate matter as well as greater emphasis on publicizing pollution alerts to encourage individuals with significant cardiopulmonary impairment to stay indoors during high-pollution episodes.

Indoor Exposures

Secondhand tobacco smoke (Chap. 395), radon gas, wood smoke, and other biologic agents generated indoors must be considered. Several studies have shown that the respirable particulate load in any household is directly proportional to the number of cigarette smokers living in that home. Increases in prevalence of respiratory illnesses, especially asthma, and reduced levels of pulmonary function measured with simple spirometry have been found in the children of smoking parents in a number of studies. Recent meta-analyses for lung cancer and cardiopulmonary diseases, combining data from multiple secondhand tobacco smoke epidemiologic studies, suggest an ~25% increase in relative risk for each condition, even after adjustment for major potential confounders.

Exposure to *radon gas* in homes is a risk factor for lung cancer. The main radon product (radon 222) is a gas that results from the decay series of uranium 238, with the immediate precursor being radium 226. The amount of radium in earth materials determines how much radon gas will be emitted. Outdoors, the concentrations are trivial. Indoors, levels are dependent on the sources, the ventilation rate of the space, and the size of the space into which the gas is emitted. Levels associated with excess lung cancer risk may be present in as many as 10% of the houses in the United States. When smokers reside in the home, the problem is potentially greater, since the molecular size of radon particles allows them to attach readily to smoke particles that are inhaled. Fortunately, technology is available for assessing and reducing the level of exposure.

Other indoor exposures of concern are bioaerosols that contain antigenic material (fungi, cockroaches,

dust mites, and pet danders) associated with an increased risk of atopy and asthma. Indoor chemical agents include strong cleaning agents (bleach, ammonia), formaldehyde, perfumes, pesticides, and oxides of nitrogen from gas appliances. Nonspecific responses associated with "tight-building syndrome," perhaps better termed "building-associated illness," in which no particular agent has been implicated, have included a wide variety of complaints, among them respiratory symptoms that are relieved only by avoiding exposure in the building in question. The degree to which "smells" and other sensory stimuli are involved in the triggering of potentially incapacitating psychological or physical responses has yet to be determined, and the long-term consequences of such environmental exposures are unknown.

Portal of Entry

The lung is a primary point of entry into the body for a number of toxic agents that affect other organ systems. For example, the lung is a route of entry for benzene (bone marrow), carbon disulfide (cardiovascular and nervous systems), cadmium (kidney), and metallic mercury (kidney, central nervous system). Thus, in any disease state of obscure origin, it is important to consider the possibility of inhaled environmental agents. Such consideration can sometimes furnish the clue needed to identify a specific external cause for a disorder that might otherwise be labeled "idiopathic."

GLOBAL CONSIDERATIONS

Indoor exposure to *biomass smoke* (wood, dung, crop residues, charcoal) is estimated to be responsible for ~3% of worldwide disability-adjusted life-years (DALYs) lost, due to acute lower respiratory infections in children and COPD and lung cancer in women. This burden of disease places indoor exposure to biomass smoke as the second leading environmental hazard for poor health, just behind unsafe water, sanitation, and hygiene, and is 3.5 times larger than the burden attributed to outdoor air pollution.

More than one-half of the world's population uses biomass fuel for cooking, heating, or baking. This occurs predominantly in the rural areas of developing countries. Because many families burn biomass fuels in open stoves, which are highly inefficient, and inside homes with poor ventilation, women and young children are exposed on a daily basis to high levels of smoke. In these homes, 24-h mean levels of fine particulate matter, a component of biomass smoke, have been reported to be 2–30 times higher than the National Ambient Air Quality Standards set by the U.S. EPA.

Epidemiologic studies have consistently shown associations between exposure to biomass smoke and both chronic bronchitis and COPD, with odds ratios ranging between 3 and 10 and increasing with longer exposures. In addition to the common occupational exposure to biomass smoke of women in developing countries, men from such countries may be occupationally exposed. Because of increased migration to the United States from developing countries, clinicians need to be aware of the chronic respiratory effects of exposure to biomass smoke, which can include interstitial lung disease **(Fig. 256-4)**. Evidence is beginning to emerge that improved stoves with chimneys can reduce biomass smoke–induced respiratory illness in both children and women.

Figure 256-4

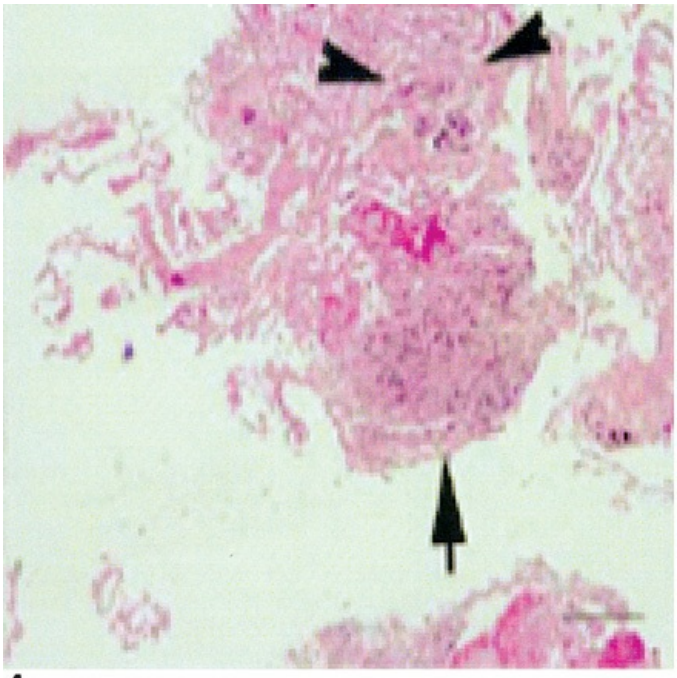

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

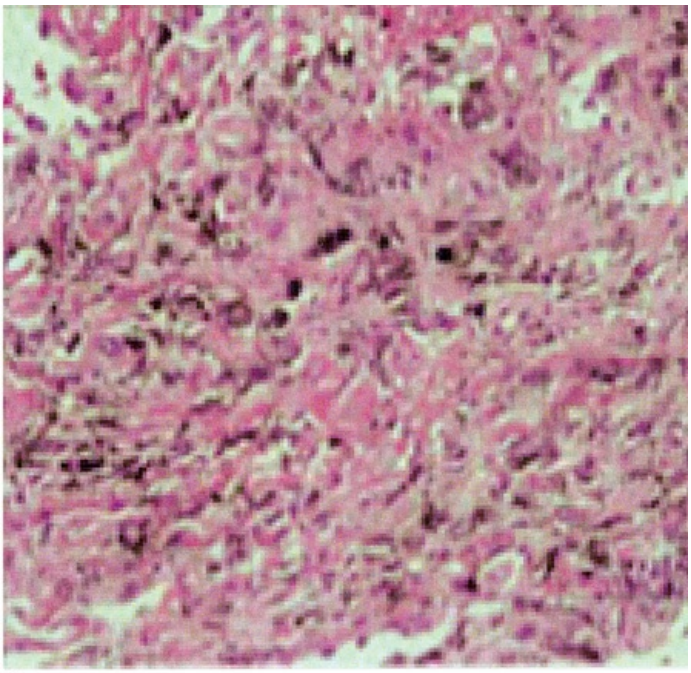

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histopathologic features of biomass smoke–induced interstitial lung disease. ***A.*** Anthracitic pigment is seen accumulating along alveolar septae (*arrowheads*) and within a pigmented dust macule (*single arrow*). ***B.*** A high-power photomicrograph contains a mixture of fibroblasts and carbon-laden macrophages.

FURTHER READINGS

Aldrich TK et al: Lung function in rescue workers at the World Trade Center after 7 years. N Engl J Med 362:1263, 2010[PMID: 20375403] [Full Text]

Balmes JR: When smoke gets in your lungs. Proc Am Thorac Soc 7:98, 2010[PMID: 20427578] [Full Text]

Chen TM et al: Outdoor air pollution: Overview and historical perspective. Am J Med Sci 333:230,

2007[PMID: 17435417] [Full Text]

Cummings KJ et al: A reconsideration of acute beryllium disease. Environ Health Perspect 117:1250, 2009[PMID: 19672405] [Full Text]

Currie GP et al: An overview of how asbestos exposure affects the lung. BMJ 339:b3209, 2009

Jerrett M et al: Long-term ozone exposure and mortality. N Engl J Med 360:1085, 2009[PMID: 19279340] [Full Text]

Rees D, Murray J: Silica, silicosis and tuberculosis. Int J Tuberc Lung Dis 11:474, 2007[PMID: 17439668] [Full Text]

Romieu I et al: Improved biomass stove intervention in rural Mexico: Impact on the respiratory health of women. Am J Respir Crit Care Med 180:649, 2009[PMID: 19556519] [Full Text]

Samuel G, Maier LA: Immunology of chronic beryllium disease. Curr Opin Allergy Clin Immunol 8:126, 2008[PMID: 18317020] [Full Text]

Suganuma N et al: Reliability of the proposed international classification of high-resolution computed tomography for occupational and environmental respiratory diseases. J Occup Health 51:210, 2009[PMID: 19372629] [Full Text]

Torén K, Blanc PD: Asthma caused by occupational exposures is common—a systematic analysis of estimates of the population-attributable fraction. BMC Pulm Med 9:7, 2009[PMID: 21698743] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION

Pneumonia is an infection of the pulmonary parenchyma. Despite being the cause of significant morbidity and mortality, pneumonia is often misdiagnosed, mistreated, and underestimated. In the past, pneumonia was typically classified as community-acquired (CAP), hospital-acquired (HAP), or ventilator-associated (VAP). Over the past two decades, however, some persons presenting as outpatients with onset of pneumonia have been found to be infected with the multidrug-resistant (MDR) pathogens previously associated with HAP. Factors responsible for this phenomenon include the development and widespread use of potent oral antibiotics, earlier transfer of patients out of acute-care hospitals to their homes or various lower-acuity facilities, increased use of outpatient IV antibiotic therapy, general aging of the population, and more extensive immunomodulatory therapies. The potential involvement of these MDR pathogens has led to a new category of pneumonia—termed *health care–associated pneumonia* (HCAP)—distinct from CAP. Conditions associated with HCAP and the likely pathogens are listed in Table 257-1.

Table 257-1 Clinical Conditions Associated with and Likely Pathogens in Health Care–Associated Pneumonia

	Pathogen			
Condition	**MRSA**	***Pseudomonas aeruginosa***	***Acinetobacter* spp.**	**MDR Enterobacteriaceae**
Hospitalization for ≥48 h	X	X	X	X
Hospitalization for ≥2 days in prior 3 months	X	X	X	X
Nursing home or extended-care-facility residence	X	X	X	X
Antibiotic therapy in preceding 3 months		X		X
Chronic dialysis	X			
Home infusion therapy	X			
Home wound care	X			
Family member with MDR infection	X			X

Abbreviations: MDR, multidrug-resistant; MRSA, methicillin-resistant *Staphylococcus aureus*.

Although the new classification system has been helpful in designing empirical antibiotic strategies, it is not without disadvantages. Not all MDR pathogens are associated with all risk factors (Table 257-1). Moreover, HCAP is a distillation of multiple risk factors, and each patient must be considered individually. For example, the risk of infection with MDR pathogens for a nursing home resident who has dementia but can independently dress, ambulate, and eat is quite different from the risk for a patient who is in a chronic vegetative state with a tracheostomy and a percutaneous feeding tube in place. In addition, risk factors for MDR infection do not preclude the development of pneumonia caused by the usual CAP pathogens.

This chapter deals with pneumonia in patients who are not considered to be immunocompromised. Pneumonia in severely immunocompromised patients, some of whom overlap with the groups of patients considered in this chapter, warrants separate discussion (see Chaps. 86, 132, and 189).

PATHOPHYSIOLOGY

Pneumonia results from the proliferation of microbial pathogens at the alveolar level and the host's response to those pathogens. Microorganisms gain access to the lower respiratory tract in several ways. The most common is by aspiration from the oropharynx. Small-volume aspiration occurs frequently during sleep (especially in the elderly) and in patients with decreased levels of consciousness. Many pathogens are inhaled as contaminated droplets. Rarely, pneumonia occurs via hematogenous spread (e.g., from tricuspid endocarditis) or by contiguous extension from an infected pleural or mediastinal space.

Mechanical factors are critically important in host defense. The hairs and turbinates of the nares capture larger inhaled particles before they reach the lower respiratory tract. The branching architecture of the tracheobronchial tree traps particles on the airway lining, where mucociliary clearance and local antibacterial factors either clear or kill the potential pathogen. The gag reflex and the cough mechanism offer critical protection from aspiration. In addition, the normal flora adhering to mucosal cells of the oropharynx, whose components are remarkably constant, prevents pathogenic bacteria from binding and thereby decreases the risk of pneumonia caused by these more virulent bacteria.

When these barriers are overcome or when the microorganisms are small enough to be inhaled to the alveolar level, resident alveolar macrophages are extremely efficient at clearing and killing pathogens. Macrophages are assisted by local proteins (e.g., surfactant proteins A and D) that have intrinsic opsonizing properties or antibacterial or antiviral activity. Once engulfed by the macrophage, the pathogens—even if they are not killed—are eliminated via either the mucociliary elevator or the lymphatics and no longer represent an infectious challenge. Only when the capacity of the alveolar macrophages to ingest or kill the microorganisms is exceeded does clinical pneumonia become manifest. In that situation, the alveolar macrophages initiate the inflammatory response to bolster lower respiratory tract defenses. The host inflammatory response, rather than the proliferation of microorganisms, triggers the clinical syndrome of pneumonia. The release of inflammatory mediators, such as interleukin (IL)-1 and tumor necrosis factor (TNF), results in fever. Chemokines, such as IL-8 and granulocyte colony-stimulating factor, stimulate the release of neutrophils and their attraction to the lung, producing both peripheral leukocytosis and increased purulent secretions. Inflammatory mediators released by macrophages and the newly recruited neutrophils create an alveolar capillary leak equivalent to that seen in the acute respiratory distress syndrome (ARDS), although in pneumonia this leak is localized (at least initially). Even erythrocytes can cross the alveolar-capillary membrane, with consequent hemoptysis. The capillary leak results in a radiographic infiltrate and rales detectable on auscultation, and hypoxemia results from alveolar filling. Moreover, some bacterial pathogens appear to interfere with the hypoxemic vasoconstriction that would normally occur with fluid-filled alveoli, and this interference can result in severe hypoxemia. Increased respiratory drive in the systemic inflammatory response syndrome (SIRS; Chap. 271) leads to respiratory alkalosis. Decreased compliance due to capillary leak, hypoxemia, increased respiratory drive, increased secretions, and occasionally infection-related bronchospasm all lead to dyspnea. If severe enough, the changes in lung mechanics secondary to reductions in lung volume and compliance and the intrapulmonary shunting of blood may cause the patient's death.

PATHOLOGY

Classic pneumonia evolves through a series of pathologic changes. The initial phase is one of *edema*, with the presence of a proteinaceous exudate—and often of bacteria—in the alveoli. This phase is rarely evident in clinical or autopsy specimens because it is so rapidly followed by a *red hepatization* phase. The presence of erythrocytes in the cellular intraalveolar exudate gives this second stage its name, but neutrophil influx is more important from the standpoint of host defense. Bacteria are occasionally seen in pathologic specimens collected during this phase. In the third phase, *gray hepatization*, no new erythrocytes are extravasating, and those already present have been lysed and degraded. The neutrophil is the predominant cell, fibrin deposition is abundant, and bacteria have disappeared. This phase corresponds with successful containment of the infection and improvement in gas exchange. In the final phase, *resolution*, the macrophage reappears as the dominant cell type in the alveolar space, and the debris of neutrophils, bacteria, and fibrin has been cleared, as has the inflammatory response.

This pattern has been described best for lobar pneumococcal pneumonia and may not apply to pneumonias of all etiologies, especially viral or *Pneumocystis* pneumonia. In VAP, respiratory bronchiolitis may precede the development of a radiologically apparent infiltrate. Because of the microaspiration mechanism, a bronchopneumonia pattern is most common in nosocomial pneumonias, whereas a lobar pattern is more common in bacterial CAP. Despite the radiographic appearance, viral and *Pneumocystis* pneumonias represent alveolar rather than interstitial processes.

COMMUNITY-ACQUIRED PNEUMONIA

Etiology

The extensive list of potential etiologic agents in CAP includes bacteria, fungi, viruses, and protozoa. Newly identified pathogens include hantaviruses, metapneumoviruses, the coronavirus responsible for severe acute respiratory syndrome (SARS), and community-acquired strains of methicillin-resistant *Staphylococcus aureus* (MRSA). Most cases of CAP, however, are caused by relatively few pathogens (Table 257-2). Although *Streptococcus pneumoniae* is most common, other organisms must also be considered in light of the patient's risk factors and severity of illness. In most cases, it is most useful to think of the potential causes as either "typical" bacterial pathogens or "atypical" organisms. The former category includes *S. pneumoniae*, *Haemophilus influenzae*, and (in selected patients) *S. aureus* and gram-negative bacilli such as *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*. The "atypical"

organisms include *Mycoplasma pneumoniae* and *Chlamydia pneumoniae* (in outpatients) and *Legionella* spp. (in inpatients) as well as respiratory viruses such as influenza viruses, adenoviruses, and respiratory syncytial viruses. Data suggest that a virus may be responsible for up to 18% of cases of CAP that require admission to the hospital. The atypical organisms cannot be cultured on standard media, nor can they be seen on Gram's stain. The frequency and importance of atypical pathogens have significant implications for therapy. These organisms are intrinsically resistant to all β-lactam agents and must be treated with a macrolide, a fluoroquinolone, or a tetracycline. In the ~10–15% of CAP cases that are polymicrobial, the etiology often includes a combination of typical and atypical pathogens.

Table 257-2 Microbial Causes of Community-Acquired Pneumonia, by Site of Care

	Hospitalized Patients	
Outpatients	**Non-ICU**	**ICU**
Streptococcus pneumoniae	*S. pneumoniae*	*S. pneumoniae*
Mycoplasma pneumoniae	*M. pneumoniae*	*Staphylococcus aureus*
Haemophilus influenzae	*Chlamydia pneumoniae*	*Legionella* spp.
C. pneumoniae	*H. influenzae*	Gram-negative bacilli
Respiratory viruses[a]	*Legionella* spp.	*H. influenzae*
	Respiratory viruses[a]	

[a]Influenza A and B viruses, adenoviruses, respiratory syncytial viruses, parainfluenza viruses.

Note: Pathogens are listed in descending order of frequency. ICU, intensive care unit.

Anaerobes play a significant role only when an episode of aspiration has occurred days to weeks before presentation for pneumonia. The combination of an unprotected airway (e.g., in patients with alcohol or drug overdose or a seizure disorder) and significant gingivitis constitutes the major risk factor. Anaerobic pneumonias are often complicated by abscess formation and significant empyemas or parapneumonic effusions.

S. aureus pneumonia is well known to complicate influenza infection. However, MRSA has been reported as the primary etiologic agent of CAP. While this entity is still relatively uncommon, clinicians must be aware of its potentially serious consequences such as necrotizing pneumonia. Two important developments have led to this problem: the spread of MRSA from the hospital setting to the community and the emergence of genetically distinct strains of MRSA in the community. The former circumstance is more likely to result in HCAP, whereas the novel community-acquired MRSA (CA-MRSA) strains have infected healthy individuals who have had no association with health care.

Unfortunately, despite a careful history and physical examination as well as routine radiographic studies, the causative pathogen in a case of CAP is difficult to predict with any degree of certainty; in more than one-half of cases, a specific etiology is never determined. Nevertheless, epidemiologic and risk factors may suggest the involvement of certain pathogens (Table 257-3).

Table 257-3 Epidemiologic Factors Suggesting Possible Causes of Community-Acquired Pneumonia

Factor	Possible Pathogen(s)
Alcoholism	*Streptococcus pneumoniae,* oral anaerobes, *Klebsiella pneumoniae, Acinetobacter* spp., *Mycobacterium tuberculosis*
COPD and/or smoking	*Haemophilus influenzae, Pseudomonas aeruginosa, Legionella* spp., *S. pneumoniae, Moraxella catarrhalis, Chlamydia pneumoniae*
Structural lung disease (e.g., bronchiectasis)	*P. aeruginosa, Burkholderia cepacia, Staphylococcus aureus*
Dementia, stroke, decreased level of consciousness	Oral anaerobes, gram-negative enteric bacteria
Lung abscess	CA-MRSA, oral anaerobes, endemic fungi, *M. tuberculosis*, atypical mycobacteria
Travel to Ohio or St. Lawrence river valleys	*Histoplasma capsulatum*
Travel to southwestern United States	Hantavirus, *Coccidioides* spp.
Travel to Southeast Asia	*Burkholderia pseudomallei,* avian influenza virus

Stay in hotel or on cruise ship in previous 2 weeks	*Legionella* spp.
Local influenza activity	Influenza virus, *S. pneumoniae*, *S. aureus*
Exposure to bats or birds	*H. capsulatum*
Exposure to birds	*Chlamydia psittaci*
Exposure to rabbits	*Francisella tularensis*
Exposure to sheep, goats, parturient cats	*Coxiella burnetii*

Abbreviations: CA-MRSA, community-acquired methicillin-resistant *Staphylococcus aureus;* COPD, chronic obstructive pulmonary disease.

Epidemiology

In the United States, ~80% of the 4 million CAP cases that occur annually are treated on an outpatient basis, and ~20% are treated in the hospital. CAP results in more than 600,000 hospitalizations, 64 million days of restricted activity, and 45,000 deaths annually. The overall yearly cost associated with CAP is estimated at $9–10 billion. The incidence rates are highest at the extremes of age. The overall annual rate in the United States is 12 cases per 1000 persons, but the figure increases to 12–18 per 1000 among children <4 years of age and to 20 per 1000 among persons >60 years of age.

The risk factors for CAP in general and for pneumococcal pneumonia in particular have implications for treatment regimens. Risk factors for CAP include alcoholism, asthma, immunosuppression, institutionalization, and an age of ≥70 years versus 60–69 years. Risk factors for pneumococcal pneumonia include dementia, seizure disorders, heart failure, cerebrovascular disease, alcoholism, tobacco smoking, chronic obstructive pulmonary disease (COPD), and HIV infection. CA-MRSA pneumonia is more likely in patients with skin colonization or infection with CA-MRSA. Enterobacteriaceae tend to infect patients who have recently been hospitalized and/or received antibiotic therapy or who have comorbidities such as alcoholism, heart failure, or renal failure. *P. aeruginosa* is a particular problem in patients with severe structural lung disease, such as bronchiectasis, cystic fibrosis, or severe COPD. Risk factors for *Legionella* infection include diabetes, hematologicmalignancy, cancer, severe renal disease, HIV infection, smoking, male gender, and a recent hotel stay or ship cruise. (Many of these risk factors would now reclassify as HCAP some cases that were previously designated CAP.)

Clinical Manifestations

CAP can vary from indolent to fulminant in presentation and from mild to fatal in severity. The various signs and symptoms that depend on the progression and severity of the infection include both constitutional findings and manifestations limited to the lung and associated structures. In light of the pathobiology of the disease, many of the findings are to be expected.

The patient is frequently febrile with tachycardia or may have a history of chills and/or sweats. Cough may be either nonproductive or productive of mucoid, purulent, or blood-tinged sputum. Depending on severity, the patient may be able to speak in full sentences or may be very short of breath. If the pleura is involved, the patient may experience pleuritic chest pain. Up to 20% of patients may have gastrointestinal symptoms such as nausea, vomiting, and/or diarrhea. Other symptoms may include fatigue, headache, myalgias, and arthralgias.

Findings on physical examination vary with the degree of pulmonary consolidation and the presence or absence of a significant pleural effusion. An increased respiratory rate and use of accessory muscles of respiration are common. Palpation may reveal increased or decreased tactile fremitus, and the percussion note can vary from dull to flat, reflecting underlying consolidated lung and pleural fluid, respectively. Crackles, bronchial breath sounds, and possibly a pleural friction rub may be heard on auscultation. The clinical presentation may not be so obvious in the elderly, who may initially display new-onset or worsening confusion and few other manifestations. Severely ill patients may have septic shock and evidence of organ failure.

Diagnosis

When confronted with possible CAP, the physician must ask two questions: Is this pneumonia, and, if so, what is the likely etiology? The former question is typically answered by clinical and radiographic methods, whereas the latter requires the aid of laboratory techniques.

CLINICAL DIAGNOSIS

The differential diagnosis includes both infectious and noninfectious entities such as acute bronchitis, acute exacerbations of chronic bronchitis, heart failure, pulmonary embolism, and radiation pneumonitis. The importance of a careful history cannot be overemphasized. For example, known cardiac disease may suggest worsening pulmonary edema, while underlying carcinoma may suggest lung injury secondary to irradiation. Epidemiologic clues, such as recent travel to areas with known endemic pathogens (e.g., the U.S. southwest), may alert the physician to specific possibilities (Table

257-3).

Unfortunately, the sensitivity and specificity of the findings on physical examination are less than ideal, averaging 58% and 67%, respectively. Therefore, chest radiography is often necessary to differentiate CAP from other conditions. Radiographic findings may include risk factors for increased severity (e.g., cavitation or multilobar involvement). Occasionally, radiographic results suggest an etiologic diagnosis. For example, pneumatoceles suggest infection with *S. aureus*, and an upper-lobe cavitating lesion suggests tuberculosis. CT is rarely necessary but may be of value in a patient with suspected postobstructive pneumonia caused by a tumor or foreign body. For outpatients, the clinical and radiologic assessments are usually all that is done before treatment for CAP is started since most laboratory results are not available soon enough to influence initial management significantly. In certain cases, the availability of rapid point-of-care outpatient diagnostic tests can be very important (e.g., rapid diagnosis of influenza virus infection can prompt specific anti-influenza drug treatment and secondary prevention).

ETIOLOGIC DIAGNOSIS

The etiology of pneumonia usually cannot be determined solely on the basis of clinical presentation; instead, the physician must rely upon the laboratory for support. Except for the 2% of CAP patients who are admitted to the intensive care unit (ICU), no data exist to show that treatment directed at a specific pathogen is statistically superior to empirical therapy. The benefit of establishing a microbial etiology can therefore be questioned, particularly in light of the cost of diagnostic testing. However, a number of reasons can be advanced for attempting an etiologic diagnosis. Identification of an unexpected pathogen allows narrowing of the initial empirical regimen that decreases antibiotic selection pressure, lessening the risk of resistance. Pathogens with important public safety implications such as *Mycobacterium tuberculosis* and influenza virus, may be found in some cases. Finally, without culture and susceptibility data, trends in resistance cannot be followed accurately, and appropriate empirical therapeutic regimens are harder to devise.

Gram's Stain and Culture of Sputum

The main purpose of the sputum Gram's stain is to ensure that a sample is suitable for culture. However, Gram's staining may also identify certain pathogens (e.g., *S. pneumoniae*, *S. aureus*, and gram-negative bacteria) by their characteristic appearance. To be adequate for culture, a sputum sample must have >25 neutrophils and <10 squamous epithelial cells per low-power field. The sensitivity and specificity of the sputum Gram's stain and culture are highly variable. Even in cases of proven bacteremic pneumococcal pneumonia, the yield of positive cultures from sputum samples is ≤ 50%.

Some patients, particularly elderly individuals, may not be able to produce an appropriate expectorated sputum sample. Others may already have started a course of antibiotics that can interfere with culture results at the time a sample is obtained. Inability to produce sputum can be a consequence of dehydration, and the correction of this condition may result in increased sputum production and a more obvious infiltrate on chest radiography. For patients admitted to the ICU and intubated, a deep-suction aspirate or bronchoalveolar lavage sample (obtained either via bronchoscopy or non-bronchoscopically) has a high yield on culture when sent to the microbiology laboratory as soon as possible. Since the etiologies in severe CAP are somewhat different from those in milder disease (Table 257-2), the greatest benefit of staining and culturing respiratory secretions is to alert the physician of unsuspected and/or resistant pathogens and to permit appropriate modification of therapy. Other stains and cultures (e.g., specific stains for *M. tuberculosis* or fungi) may be useful as well.

Blood Cultures

The yield from blood cultures, even when samples are collected before antibiotic therapy, is disappointingly low. Only ~5–14% of cultures of blood from patients hospitalized with CAP are positive, and the most frequently isolated pathogen is *S. pneumoniae*. Since recommended empirical regimens all provide pneumococcal coverage, a blood culture positive for this pathogen has little, if any, effect on clinical outcome. However, susceptibility data may allow narrowing of antibiotic therapy in appropriate cases. Because of the low yield and the lack of significant impact on outcome, blood cultures are no longer considered *de rigueur* for all hospitalized CAP patients. Certain high-risk patients—including those with neutropenia secondary to pneumonia, asplenia, or complement deficiencies; chronic liver disease; or severe CAP—should have blood cultured.

Antigen Tests

Two commercially available tests detect pneumococcal and certain *Legionella* antigens in urine. The test for *L. pneumophila* detects only serogroup 1, but this serogroup accounts for most community-acquired cases of Legionnaires' disease. The sensitivity and specificity of the *Legionella* urine antigen test are as high as 90% and 99%, respectively. The pneumococcal urine antigen test is also quite sensitive and specific (80% and >90%, respectively). Although false-positive results can be obtained with samples from pneumococcus-colonized children, the test is generally reliable. Both tests can detect antigen even after the initiation of appropriate antibiotic therapy. Other antigen tests include a rapid test for influenza virus and direct fluorescent antibody tests for influenza virus and respiratory

syncytial virus; the latter tests are only poorly sensitive.

Polymerase Chain Reaction

Polymerase chain reaction (PCR) tests, which amplify a microorganism's DNA or RNA, are available for a number of pathogens, including *L. pneumophila* and mycobacteria. In addition, a multiplex PCR can detect the nucleic acid of *Legionella* spp., *M. pneumoniae*, and *C. pneumoniae*. However, the use of these PCR assays is generally limited to research studies. In patients with pneumococcal pneumonia, an increased bacterial load documented by PCR is associated with an increased risk of septic shock, need for mechanical ventilation, and death. Such a test could conceivably help identify patients suitable for ICU admission.

Serology

A fourfold rise in specific IgM antibody titer between acute- and convalescent-phase serum samples is generally considered diagnostic of infection with the pathogen in question. In the past, serologic tests were used to help identify atypical pathogens as well as selected unusual organisms such as *Coxiella burnetii*. Recently, however, they have fallen out of favor because of the time required to obtain a final result for the convalescent-phase sample.

Treatment: Community-Acquired Pneumonia

SITE OF CARE

The cost of inpatient management exceeds that of outpatient treatment by a factor of 20, and hospitalization accounts for most CAP-related expenditures. Thus the decision to admit a patient with CAP to the hospital has considerable implications. Certain patients clearly can be managed at home, and others clearly require treatment in the hospital, but the choice is sometimes difficult. Tools that objectively assess the risk of adverse outcomes, including severe illness and death, can minimize unnecessary hospital admissions. There are currently two sets of criteria: the Pneumonia Severity Index (PSI), a prognostic model used to identify patients at low risk of dying; and the CURB-65 criteria, a severity-of-illness score.

To determine the PSI, points are given for 20 variables, including age, coexisting illness, and abnormal physical and laboratory findings. On the basis of the resulting score, patients are assigned to one of five classes with the following mortality rates: class 1, 0.1%; class 2, 0.6%; class 3, 2.8%; class 4, 8.2%; and class 5, 29.2%. Clinical trials demonstrate that routine use of the PSI results in lower admission rates for class 1 and class 2 patients. Patients in classes 4 and 5 should be admitted to the hospital, while those in class 3 should ideally be admitted to an observation unit until a further decision can be made.

The CURB-65 criteria include five variables: confusion (C); urea >7 mmol/L (U); respiratory rate ≥ 30/min (R); blood pressure, systolic ≤90 mmHg or diastolic ≤60 mmHg (B); and age ≥65 years (65). Patients with a score of 0, among whom the 30-day mortality rate is 1.5%, can be treated outside the hospital. With a score of 2, the 30-day mortality rate is 9.2%, and patients should be admitted to the hospital. Among patients with scores of ≥3, mortality rates are 22% overall; these patients may require admission to an ICU.

It is not clear which assessment tool is superior. The PSI is less practical in a busy emergency room setting because of the need to assess 20 variables. While the CURB-65 criteria are easily remembered, they have not been studied as extensively. Whichever system is used, these objective criteria must always be tempered by careful consideration of factors relevant to individual patients, including the ability to comply reliably with an oral antibiotic regimen and the resources available to the patient outside the hospital. In fact, neither the PSI nor CURB-65 is ideal for determining the need for ICU care. The severity criteria proposed by the Infectious Diseases Society of America (IDSA) and the American Thoracic Society (ATS) in their guidelines for the management of CAP are better suited to this purpose.

ANTIBIOTIC RESISTANCE

Antimicrobial resistance is a significant problem that threatens to diminish our therapeutic armamentarium. Misuse of antibiotics results in increased antibiotic selection pressure that can affect resistance locally or even globally by clonal dissemination. For CAP, the main resistance issues currently involve *S. pneumoniae* and CA-MRSA.

S. Pneumoniae

In general, pneumococcal resistance is acquired (1) by direct DNA incorporation and remodeling resulting from contact with closely related oral commensal bacteria, (2) by the process of natural transformation, or (3) by mutation of certain genes.

The cutoff for penicillin susceptibility in pneumonia has recently been raised from a minimal inhibitory concentration (MIC) of ≤0.6 μg/mL to an MIC of ≤2 μg/mL. Cutoffs for intermediate resistance have been raised to 4 μg/mL (from 0.1–1 μg/mL) and ≥8 μg/mL (from ≥2 μg/mL), respectively. These changes in susceptibility thresholds have resulted in a dramatic decrease in the proportion of pneumococcal isolates considered nonsusceptible. For meningitis, MIC thresholds remain at the former levels.

Fortunately, resistance to penicillin appeared to plateau even before the change in MIC thresholds. Pneumococcal resistance to β-lactam drugs is due solely to low-affinity penicillin-binding proteins. Risk factors for penicillin-resistant pneumococcal infection include recent antimicrobial therapy, an age of <2 years or >65 years, attendance at day-care centers, recent hospitalization, and HIV infection.

In contrast to penicillin resistance, resistance to macrolides is increasing through several mechanisms. *Target-site modification* is caused by ribosomal methylation in 23S rRNA encoded by the *ermB* gene, resulting in resistance to macrolides, lincosamides, and streptogramin B–type antibiotics. This MLSB phenotype is associated with high-level resistance, with typical MICs of ≥64 μg/mL. The *efflux mechanism* encoded by the *mef* gene (*M phenotype*) is usually associated with low-level resistance (MICs, 1–32 μg/mL). These two mechanisms account for ~45% and ~65%, respectively, of resistant pneumococcal isolates in the United States. High-level resistance to macrolides is more common in Europe, whereas lower-level resistance seems to predominate in North America. Although clinical failures with macrolides have been reported, many experts think that these drugs still have a role to play in the management of pneumococcal pneumonia in North America.

Pneumococcal resistance to fluoroquinolones (e.g., ciprofloxacin and levofloxacin) has been reported. Changes can occur in one or both target sites (topoisomerases II and IV); changes in these two sites usually result from mutations in the *gyrA* and *parC* genes, respectively. The increasing number of pneumococcal isolates that, although still testing susceptible to fluoroquinolones, already have a mutation in one target site is of concern. Such organisms may be more likely to undergo a second step mutation that will render them fully resistant to fluoroquinolones. In addition, an efflux pump may play a role in pneumococcal resistance to fluoroquinolones.

Isolates resistant to drugs from three or more antimicrobial classes with different mechanisms of action are considered MDR. The propensity for an association of pneumococcal resistance to penicillin with reduced susceptibility to other drugs such as macrolides, tetracyclines, and trimethoprim-sulfamethoxazole, is also of concern. In the United States, 58.9% of penicillin-resistant pneumococcal isolates from blood are also resistant to macrolides.

The most important risk factor for antibiotic-resistant pneumococcal infection is use of a specific antibiotic within the previous 3 months. Therefore, a patient's history of prior antibiotic treatment is a critical factor in avoiding the use of an inappropriate antibiotic.

CA-MRSA

CAP due to MRSA may be caused by infection with the classic hospital-acquired strains or with the more recently identified, genotypically and phenotypically distinct community-acquired strains. Most infections with the former strains have been acquired either directly or indirectly by contact with the health care environment and would now be classified as HCAP. In some hospitals, CA-MRSA strains are displacing the classic hospital-acquired strains—a trend suggesting that the newer strains may be more robust.

Methicillin resistance in *S. aureus* is determined by the *mecA* gene, which encodes for resistance to all β-lactam drugs. At least five *staphylococcal chromosomal cassette mec* (*SCCmec*) types have been described. The typical hospital-acquired strain usually has type II or III, whereas CA-MRSA has a type IV SCC*mec* element. CA-MRSA isolates tend to be less resistant than the older hospital-acquired strains and are often susceptible to trimethoprim-sulfamethoxazole, clindamycin, and tetracycline in addition to vancomycin and linezolid. However, CA-MRSA strains may also carry genes for superantigens, such as enterotoxins B and C and Panton-Valentine leukocidin, a membrane-tropic toxin that can create cytolytic pores in polymorphonuclear neutrophils, monocytes, and macrophages.

Gram-Negative Bacilli

A detailed discussion of resistance among gram-negative bacilli is beyond the scope of this chapter (see Chap. 149). Fluoroquinolone resistance among isolates of *Escherichia coli* from the community appears to be increasing. *Enterobacter* spp. are typically resistant to cephalosporins; the drugs of choice for use against these bacteria are usually fluoroquinolones or carbapenems. Similarly, when infections due to bacteria producing extended-spectrum β-lactamases are documented or suspected, a fluoroquinolone or a carbapenem should be used; these MDR strains are more likely to be involved in HCAP.

INITIAL ANTIBIOTIC MANAGEMENT

Since the physician rarely knows the etiology of CAP at the outset of treatment, initial therapy is usually empirical and is designed to cover the most likely pathogens (Table 257-4). In all cases, antibiotic treatment should be initiated as expeditiously as possible. The CAP treatment guidelines in the United States (summarized in Table 257-4) represent joint statements from the IDSA and the ATS; the Canadian guidelines come from the Canadian Infectious Disease Society and the Canadian Thoracic Society. In these guidelines, coverage is always provided for the pneumococcus and the atypical pathogens. In contrast, guidelines from some European countries do not always include atypical coverage based on local epidemiologic data. The U.S.–Canadian approach is supported by retrospective data from several studies of administrative databases including thousands of patients.

Atypical pathogen coverage provided by the addition of a macrolide to a cephalosporin or by the use of a fluoroquinolone alone has been consistently associated with a significant reduction in mortality rates compared with those for β-lactam coverage alone.

Table 257-4 Empirical Antibiotic Treatment of Community-Acquired Pneumonia

Outpatients
Previously healthy and no antibiotics in past 3 months
• A macrolide [clarithromycin (500 mg PO bid) or azithromycin (500 mg PO once, then 250 mg qd)] ***or***
• Doxycycline (100 mg PO bid)
Comorbidities or antibiotics in past 3 months: select an alternative from a different class
• A respiratory fluoroquinolone [moxifloxacin (400 mg PO qd), gemifloxacin (320 mg PO qd), levofloxacin (750 mg PO qd)] ***or***
• A β-lactam [preferred: high-dose amoxicillin (1 g tid) or amoxicillin/clavulanate (2 g bid); alternatives: ceftriaxone (1–2 g IV qd), cefpodoxime (200 mg PO bid), cefuroxime (500 mg PO bid)] ***plus*** a macrolide[a]
In regions with a high rate of "high-level" pneumococcal macrolide resistance,[b] consider alternatives listed above for patients with comorbidities.
Inpatients, Non-ICU
• A respiratory fluoroquinolone [moxifloxacin (400 mg PO or IV qd), gemifloxacin (320 mg PO qd), levofloxacin (750 mg PO or IV qd)]
• A β-lactam[c] [cefotaxime (1–2 g IV q8h), ceftriaxone (1–2 g IV qd), ampicillin (1–2 g IV q4–6h), ertapenem (1 g IV qd in selected patients)] ***plus*** a macrolide[d] [oral clarithromycin or azithromycin (as listed above for previously healthy patients) or IV azithromycin (1 g once, then 500 mg qd)]
Inpatients, ICU
• A β-lactam[e] [cefotaxime (1–2 g IV q8h), ceftriaxone (2 g IV qd), ampicillin-sulbactam (2 g IV q8h)] ***plus***
• Azithromycin or a fluoroquinolone (as listed above for inpatients, non-ICU)
Special Concerns
If *Pseudomonas* is a consideration
• An antipneumococcal, antipseudomonal β-lactam [piperacillin/tazobactam (4.5 g IV q6h), cefepime (1–2 g IV q12h), imipenem (500 mg IV q6h), meropenem (1 g IV q8h)] ***plus*** either ciprofloxacin (400 mg IV q12h) or levofloxacin (750 mg IV qd)
• The above β-lactams ***plus*** an aminoglycoside [amikacin (15 mg/kg qd) or tobramycin (1.7 mg/kg qd) and azithromycin]
• The above β-lactams[f] ***plus*** an aminoglycoside ***plus*** an antipneumococcal fluoroquinolone
If CA-MRSA is a consideration
• Add linezolid (600 mg IV q12h) or vancomycin (1 g IV q12h).

[a] Doxycycline (100 mg PO bid) is an alternative to the macrolide.

[b] MICs of >16 μg/mL in 25% of isolates.

[c] A respiratory fluoroquinolone should be used for penicillin-allergic patients.

[d] Doxycycline (100 mg IV q12h) is an alternative to the macrolide.

[e] For penicillin-allergic patients, use a respiratory fluoroquinolone and aztreonam (2 g IV q8h).

[f] For penicillin-allergic patients, substitute aztreonam.

Abbreviations: CA-MRSA, community-acquired methicillin-resistant *Staphylococcus aureus;* ICU, intensive care unit.

Therapy with a macrolide or a fluoroquinolone within the previous 3 months is associated with an increased likelihood of infection with a resistant strain of *S. pneumoniae*. For this reason, a fluoroquinolone-based regimen should be used for patients recently given a macrolide, and vice versa (Table 257-4).

Once the etiologic agent(s) and susceptibilities are known, therapy may be altered to target the specific pathogen(s). However, this decision is not always straightforward. If blood cultures yield *S. pneumoniae* sensitive to penicillin after 2 days of treatment with a macrolide plus a β-lactam or with a fluoroquinolone alone, should therapy be switched to penicillin alone? The concern here is that a β-lactam alone would not be effective in the potential 15% of cases with atypical co-infection. No

standard approach exists. In all cases, the individual patient and the various risk factors must be considered.

Management of bacteremic pneumococcal pneumonia is also controversial. Data from nonrandomized studies suggest that combination therapy (especially with a macrolide and a β-lactam) is associated with a lower mortality rate than monotherapy, particularly in severely ill patients. The exact reason is unknown, but possible explanations include an additive or synergistic antibacterial effect, antimicrobial tolerance, atypical co-infection, or the immunomodulatory effects of the macrolides.

For patients with CAP who are admitted to the ICU, the risk of infection with *P. aeruginosa* or CA-MRSA is increased, and coverage should be considered when a patient has risk factors or a Gram's stain suggestive of these pathogens (Table 257-4). If CA-MRSA infection is suspected, either linezolid or vancomycin should be added to the initial empirical regimen. There is concern about vancomycin's loss of potency against MRSA; in addition, vancomycin does not reach significant concentrations in epithelial lining fluid, whereas concentrations of linezolid at this site exceed the MIC for MRSA during the entire dosing interval.

Although hospitalized patients have traditionally received initial therapy by the IV route, some drugs —particularly the fluoroquinolones—are very well absorbed and can be given orally from the outset to select patients. For patients initially treated IV, a switch to oral treatment is appropriate as long as the patient can ingest and absorb the drugs, is hemodynamically stable, and is showing clinical improvement.

The duration of treatment for CAP has generated considerable interest. Patients were previously treated for 10–14 days, but studies with fluoroquinolones and telithromycin suggest that a 5-day course is sufficient for otherwise uncomplicated CAP. Even a single dose of ceftriaxone has been associated with a significant cure rate. A longer course is required for patients with bacteremia, metastatic infection, or infection with a virulent pathogen such as *P. aeruginosa* or CA-MRSA.

GENERAL CONSIDERATIONS

In addition to appropriate antimicrobial therapy, certain general considerations apply in dealing with CAP, HCAP, or HAP/VAP. Adequate hydration, oxygen therapy for hypoxemia, and assisted ventilation when necessary are critical to the success of therapy. Patients with severe CAP who remain hypotensive despite fluid resuscitation may have adrenal insufficiency and may respond to glucocorticoid treatment. Immunomodulatory therapy in the form of drotrecogin alfa (activated) should be considered for CAP patients with persistent septic shock and APACHE II scores of ≥25, particularly if the infection is caused by *S. pneumoniae*. The value of other forms of adjunctive therapy, including glucocorticoids, statins, and angiotensin-converting enzyme inhibitors, remains unproven in the management of CAP.

Failure to Improve

Patients who are slow to respond to therapy should be reevaluated at about day 3 (sooner if their condition is worsening rather than simply not improving), and a number of possible scenarios should be considered. A number of noninfectious conditions can mimic pneumonia, including pulmonary edema, pulmonary embolism, lung carcinoma, radiation and hypersensitivity pneumonitis, and connective tissue disease involving the lungs. If the patient has CAP and treatment is aimed at the correct pathogen, the lack of response may be explained in a number of ways. The pathogen may be resistant to the drug selected, or a sequestered focus (e.g., a lung abscess or empyema) may be blocking access of the antibiotic(s) to the pathogen. The patient may be getting either the wrong drug or the correct drug at the wrong dose or frequency of administration. It is also possible that CAP is the correct diagnosis but that an unsuspected pathogen (e.g., CA-MRSA, *M. tuberculosis*, or a fungus) is the cause. Nosocomial superinfections—both pulmonary and extrapulmonary—are possible explanations for failure to improve or worsening. In all cases of delayed response or deteriorating condition, the patient must be carefully reassessed and appropriate studies initiated. These studies may include such diverse procedures as CT and bronchoscopy.

Complications

As in other severe infections, common complications of severe CAP include respiratory failure, shock and multiorgan failure, coagulopathy, and exacerbation of comorbid illnesses. Three particularly noteworthy conditions are metastatic infection, lung abscess, and complicated pleural effusion. Metastatic infection (e.g., brain abscess or endocarditis), although unusual, deserves immediate attention by the physician, with a detailed workup and proper treatment. Lung abscess may occur in association with aspiration or with infection caused by a single CAP pathogen such as CA-MRSA, *P. aeruginosa*, or (rarely) *S. pneumoniae*. Aspiration pneumonia is typically a mixed polymicrobial infection involving both aerobes and anaerobes. In either scenario, drainage should be established, and antibiotics that cover the known or suspected pathogens should be administered. A significant pleural effusion should be tapped for both diagnostic and therapeutic purposes. If the fluid has a pH of <7, a glucose level of <2.2 mmol/L, and a lactate dehydrogenase concentration of >1000 U/L or if bacteria are seen or cultured, then the fluid should be drained; a chest tube is usually required.

Follow-Up

Fever and leukocytosis usually resolve within 2–4 days in otherwise healthy patients with CAP, but physical findings may persist longer. Chest radiographic abnormalities are slowest to resolve and may require 4–12 weeks to clear, with the speed of clearance depending on the patient's age and underlying lung disease. Patients may be discharged from the hospital once their clinical conditions are stable, with no active medical problems requiring hospital care. The site of residence after discharge (nursing home, home with family, home alone) is an important consideration, particularly for elderly patients. For a patient whose condition is improving and who (if hospitalized) has been discharged, a follow-up radiograph can be done ~4–6 weeks later. If relapse or recurrence is documented, particularly in the same lung segment, the possibility of an underlying neoplasm must be considered.

Prognosis

The prognosis of CAP depends on the patient's age, comorbidities, and site of treatment (inpatient or outpatient). Young patients without comorbidity do well and usually recover fully after ~2 weeks. Older patients and those with comorbid conditions can take several weeks longer to recover fully. The overall mortality rate for the outpatient group is <1%. For patients requiring hospitalization, the overall mortality rate is estimated at 10%, with ~50% of deaths directly attributable to pneumonia.

Prevention

The main preventive measure is vaccination. The recommendations of the Advisory Committee on Immunization Practices should be followed for influenza and pneumococcal vaccines. In the event of an influenza outbreak, unprotected patients at risk from complications should be vaccinated immediately and given chemoprophylaxis with either oseltamivir or zanamivir for 2 weeks—i.e., until vaccine-induced antibody levels are sufficiently high. Because of an increased risk of pneumococcal infection, even among patients without obstructive lung disease, smokers should be strongly encouraged to stop smoking.

An available 7-valent pneumococcal conjugate vaccine produces T cell–dependent antigens that result in long-term immunologic memory. Administration of this vaccine to children has led to an overall decrease in the prevalence of antimicrobial-resistant pneumococci and in the incidence of invasive pneumococcal disease among both children and adults. However, vaccination can be followed by the replacement of vaccine serotypes with nonvaccine serotypes (e.g., 19A and 35B).

HEALTH CARE–ASSOCIATED PNEUMONIA

HCAP represents a transition between classic CAP and typical HAP. The definition of HCAP is still in some degree of flux because of a lack of large-scale studies. Several of the studies that are available have been limited to patients with culture-positive pneumonia. In these studies, the incidence of MDR pathogens in HCAP was as high as or higher than in HAP/VAP. MRSA in particular was more common in HCAP than in traditional HAP/VAP. Conversely, prospective studies in nontertiary-care centers have found a low incidence of MDR pathogens in HCAP.

The patients at greatest risk for HCAP are not well defined. Patients from nursing homes are not always at elevated risk for infection with MDR pathogens. Careful evaluation of nursing home residents with pneumonia suggests that their risk of MDR infection is low if they have not recently received antibiotics and are independent in most activities of daily living. Conversely, nursing home patients are at increased risk of infection with influenza virus and other atypical pneumonia pathogens. Undue concern about MDR pathogens occasionally results in a failure to cover atypical pathogens in treating nursing home patients. In addition, patients receiving home infusion therapy or undergoing chronic dialysis are probably at particular risk for MRSA pneumonia but may not be at greater risk for infection with *Pseudomonas* or *Acinetobacter* than are other patients who develop CAP.

In general, the management of HCAP due to MDR pathogens is similar to that of MDR HAP/VAP. This topic will therefore be covered in subsequent sections on HAP and VAP. The prognosis of HCAP is intermediate between that of CAP and VAP and is closer to that of HAP.

Ventilator-Associated Pneumonia

Most research on VAP has focused on illness in the hospital setting. However, the information and principles based on this research can be applied to non-ICU HAP and HCAP as well. The greatest difference between VAP and HCAP/HAP is the return to dependence on expectorated sputum for a microbiologic diagnosis of VAP (as for that of CAP), which is further complicated by frequent colonization by pathogens in patients with HAP or HCAP.

ETIOLOGY

Potential etiologic agents of VAP include both MDR and non-MDR bacterial pathogens (Table 257-5). The non-MDR group is nearly identical to the pathogens found in severe CAP (Table 257-2); it is not surprising that such pathogens predominate if VAP develops in the first 5–7 days of the hospital stay. However, if patients have other risk factors for HCAP, MDR pathogens are a consideration, even early in the hospital course. The relative frequency of individual MDR pathogens can vary significantly from

hospital to hospital and even between different critical care units within the same institution. Most hospitals have problems with *P. aeruginosa* and MRSA, but other MDR pathogens are often institution-specific. Less commonly, fungal and viral pathogens cause VAP, usually affecting severely immunocompromised patients. Rarely, community-associated viruses cause mini-epidemics, usually when introduced by ill health care workers.

Table 257-5 Microbiologic Causes of Ventilator-Associated Pneumonia

Non-MDR Pathogens	MDR Pathogens
Streptococcus pneumoniae	*Pseudomonas aeruginosa*
Other *Streptococcus* spp.	MRSA
Haemophilus influenzae	*Acinetobacter* spp.
MSSA	Antibiotic-resistant Enterobacteriaceae
Antibiotic-sensitive Enterobacteriaceae	*Enterobacter* spp.
Escherichia coli	ESBL-positive strains
Klebsiella pneumoniae	*Klebsiella* spp.
Proteus spp.	*Legionella pneumophila*
Enterobacter spp.	Burkholderia cepacia
Serratia marcescens	*Aspergillus* spp.

Abbreviations: ESBL, extended-spectrum β-lactamase; MDR, multidrug-resistant; MRSA, methicillin-resistant *Staphylococcus aureus;* MSSA, methicillin-sensitive *S. aureus.*

EPIDEMIOLOGY

Pneumonia is a common complication among patients requiring mechanical ventilation. Prevalence estimates vary between 6 and 52 cases per 100 patients, depending on the population studied. On any given day in the ICU, an average of 10% of patients will have pneumonia—VAP in the overwhelming majority of cases. The frequency of diagnosis is not static but changes with the duration of mechanical ventilation, with the highest hazard ratio in the first 5 days and a plateau in additional cases (1% per day) after ~2 weeks. However, the cumulative rate among patients who remain ventilated for as long as 30 days is as high as 70%. These rates often do not reflect the recurrence of VAP in the same patient. Once a ventilated patient is transferred to a chronic-care facility or to home, the incidence of pneumonia drops significantly, especially in the absence of other risk factors for pneumonia. However, in chronic ventilator units, purulent tracheobronchitis becomes a significant issue, often interfering with efforts to wean patients off mechanical ventilation.

Three factors are critical in the pathogenesis of VAP: colonization of the oropharynx with pathogenic microorganisms, aspiration of these organisms from the oropharynx into the lower respiratory tract, and compromise of the normal host defense mechanisms. Most risk factors and their corresponding prevention strategies pertain to one of these three factors (Table 257-6).

Table 257-6 Pathogenic Mechanisms and Corresponding Prevention Strategies for Ventilator-Associated Pneumonia

Pathogenic Mechanism	Prevention Strategy
Oropharyngeal colonization with pathogenic bacteria	
Elimination of normal flora	Avoidance of prolonged antibiotic courses
Large-volume oropharyngeal aspiration around time of intubation	Short course of prophylactic antibiotics for comatose patients[a]
Gastroesophageal reflux	Postpyloric enteral feeding[b]; avoidance of high gastric residuals, prokinetic agents
Bacterial overgrowth of stomach	Prophylactic agents that raise gastric pH[b]; selective decontamination of digestive tract with nonabsorbable antibiotics[b]
Cross-infection from other colonized patients	Hand washing, especially with alcohol-based hand rub; intensive infection control education[a]; isolation; proper cleaning of reusable equipment
Large-volume aspiration Microaspiration around	Endotracheal intubation; avoidance of sedation; decompression of small-bowel obstruction

endotracheal tube	
Endotracheal intubation	Noninvasive ventilation[a]
Prolonged duration of ventilation	Daily awakening from sedation,[a] weaning protocols[a]
Abnormal swallowing function	Early percutaneous tracheostomy[a]
Secretions pooled above endotracheal tube	Head of bed elevated[a]; continuous aspiration of subglottic secretions with specialized endotracheal tube[a]; avoidance of reintubation; minimization of sedation and patient transport
Altered lower respiratory host defenses	Tight glycemic control[b]; lowering of hemoglobin transfusion threshold; specialized enteral feeding formula

[a]Strategies demonstrated to be effective in at least one randomized controlled trial.

[b]Strategies with negative randomized trials or conflicting results.

The most obvious risk factor is the endotracheal tube, which bypasses the normal mechanical factors preventing aspiration. While the presence of an endotracheal tube may prevent large-volume aspiration, microaspiration is actually exacerbated by secretions pooling above the cuff. The endotracheal tube and the concomitant need for suctioning can damage the tracheal mucosa, thereby facilitating tracheal colonization. In addition, pathogenic bacteria can form a glycocalyx biofilm on the tube's surface that protects them from both antibiotics and host defenses. The bacteria can also be dislodged during suctioning and can reinoculate the trachea, or tiny fragments of glycocalyx can embolize to distal airways, carrying bacteria with them.

In a high percentage of critically ill patients, the normal oropharyngeal flora is replaced by pathogenic microorganisms. The most important risk factors are antibiotic selection pressure, cross-infection from other infected/colonized patients or contaminated equipment, and malnutrition. Of these factors, antibiotic exposure poses the greatest risk by far. Pathogens such as *P. aeruginosa* almost never cause infection in patients without prior exposure to antibiotics. The recent emphasis on hand hygiene has lowered the cross-infection rate.

How the lower respiratory tract defenses become overwhelmed remains poorly understood. Almost all intubated patients experience microaspiration and are at least transiently colonized with pathogenic bacteria. However, only around one-third of colonized patients develop VAP. Colony counts increase to high levels, sometimes days before the development of clinical pneumonia; these increases suggest that the final step in VAP development, independent of aspiration and oropharyngeal colonization, is the overwhelming of host defenses. Severely ill patients with sepsis and trauma appear to enter a state of immunoparalysis several days after admission to the ICU—a time that corresponds to the greatest risk of developing VAP. The mechanism of this immunosuppression is not clear, although several factors have been suggested. Hyperglycemia affects neutrophil function, and trials suggest that keeping the blood sugar close to normal with exogenous insulin may have beneficial effects, including a decreased risk of infection. More frequent transfusions also adversely affect the immune response.

CLINICAL MANIFESTATIONS

The clinical manifestations are generally the same in VAP as in all other forms of pneumonia: fever, leukocytosis, increase in respiratory secretions, and pulmonary consolidation on physical examination, along with a new or changing radiographic infiltrate. The frequency of abnormal chest radiographs before the onset of pneumonia in intubated patients and the limitations of portable radiographic technique make interpretation of radiographs more difficult than in patients who are not intubated. Other clinical features may include tachypnea, tachycardia, worsening oxygenation, and increased minute ventilation.

DIAGNOSIS

No single set of criteria is reliably diagnostic of pneumonia in a ventilated patient. The inability to identify such patients compromises efforts to prevent and treat VAP and even calls into question estimates of the impact of VAP on mortality rates.

Application of clinical criteria consistently results in overdiagnosis of VAP, largely because of three common findings in at-risk patients: (1) tracheal colonization with pathogenic bacteria in patients with endotracheal tubes, (2) multiple alternative causes of radiographic infiltrates in mechanically ventilated patients, and (3) the high frequency of other sources of fever in critically ill patients. The differential diagnosis of VAP includes a number of entities such as atypical pulmonary edema, pulmonary contusion, alveolar hemorrhage, hypersensitivity pneumonitis, ARDS, and pulmonary embolism. Clinical findings in ventilated patients with fever and/or leukocytosis may have alternative causes, including antibiotic-associated diarrhea, sinusitis, urinary tract infection, pancreatitis, and drug fever.

Conditions mimicking pneumonia are often documented in patients in whom VAP has been ruled out by accurate diagnostic techniques. Most of these alternative diagnoses do not require antibiotic treatment; require antibiotics different from those used to treat VAP; or require some additional intervention, such as surgical drainage or catheter removal, for optimal management.

This diagnostic dilemma has led to debate and controversy. The major question is whether a quantitative-culture approach as a means of eliminating false-positive clinical diagnoses is superior to the clinical approach enhanced by principles learned from quantitative-culture studies. The most recent IDSA/ATS guidelines for HCAP suggest that either approach is clinically valid.

Quantitative-Culture Approach

The essence of the quantitative-culture approach is to discriminate between colonization and true infection by determining the bacterial burden. The more distal in the respiratory tree the diagnostic sampling, the more specific the results and therefore the lower the threshold of growth necessary to diagnose pneumonia and exclude colonization. For example, a quantitative endotracheal aspirate yields proximate samples, and the diagnostic threshold is 10^6 cfu/mL. The protected specimen brush method, in contrast, obtains distal samples and has a threshold of 10^3 cfu/mL. Conversely, sensitivity declines as more distal secretions are obtained, especially when they are collected blindly (i.e., by a technique other than bronchoscopy). Additional tests that may increase the diagnostic yield include Gram's stain, differential cell counts, staining for intracellular organisms, and detection of local protein levels elevated in response to infection.

Several studies have compared patient cohorts managed by the various quantitative-culture methods. While these studies documented issues of relative sensitivity and specificity, outcomes were not significantly different for the various groups of patients. The IDSA/ATS guidelines suggest that all these methods are appropriate and that the choice depends on availability and local expertise.

The Achilles heel of the quantitative approach is the effect of antibiotic therapy. With sensitive microorganisms, a single antibiotic dose can reduce colony counts below the diagnostic threshold. Recent changes in antibiotic therapy are the most significant. After 3 days, the operating characteristics of the tests are almost the same as if no antibiotic therapy has been given. Conversely, colony counts above the diagnostic threshold during antibiotic therapy suggest that the current antibiotics are ineffective. Even the normal host response may be sufficient to reduce quantitative-culture counts below the diagnostic threshold if sampling is delayed. In short, expertise in quantitative-culture techniques is critical, with a specimen obtained as soon as pneumonia is suspected and before antibiotic therapy is initiated or changed.

In a study comparing the quantitative with the clinical approach, use of bronchoscopic quantitative cultures resulted in significantly less antibiotic use at 14 days after study entry and lower rates of mortality and severity-adjusted mortality at 28 days. In addition, more alternative sites of infection were found in patients randomized to the quantitative-culture strategy. A critical aspect of this study was that antibiotic treatment was initiated only in patients whose gram-stained respiratory sample was positive or who displayed signs of hemodynamic instability. Fewer than one-half as many patients were treated for pneumonia in the bronchoscopy group, and only one-third as many microorganisms were cultured. Other studies that did not demonstrate a similar beneficial impact of quantitative culture on outcomes did not tightly link antibiotic treatment to the results of quantitative culture and other tests.

Clinical Approach

The lack of specificity of a clinical diagnosis of VAP has led to efforts to improve the diagnostic criteria. The Clinical Pulmonary Infection Score (CPIS) was developed by weighting of the various clinical criteria usually used for the diagnosis of VAP (Table 257-7). Use of the CPIS allows the selection of low-risk patients who may need only short-course antibiotic therapy or no treatment at all. Moreover, studies have demonstrated that the absence of bacteria in gram-stained endotracheal aspirates makes pneumonia an unlikely cause of fever or pulmonary infiltrates. These findings, coupled with a heightened awareness of the alternative diagnoses possible in patients with suspected VAP, can prevent inappropriate treatment for this disease. Furthermore, data show that the absence of an MDR pathogen in tracheal aspirate cultures eliminates the need for MDR coverage when empirical antibiotic therapy is narrowed. Since the most likely explanations for the mortality benefit of bronchoscopic quantitative cultures are decreased antibiotic selection pressure (which reduces the risk of subsequent infection with MDR pathogens) and identification of alternative sources of infection, a clinical diagnostic approach that incorporates such principles may result in similar outcomes.

Table 257-7 Clinical Pulmonary Infection Score (CPIS)

Criterion	Score
Fever (°C)	
≥38.5 but ≤38.9	1
>39 or <36	2

Leukocytosis	
<4000 or >11,000/μL	1
Bands >50%	1 (additional)
Oxygenation (mmHg)	
Pa_{O_2}/F_{IO_2} <250 and no ARDS	2
Chest radiograph	
Localized infiltrate	2
Patchy or diffuse infiltrate	1
Progression of infiltrate (no ARDS or CHF)	2
Tracheal aspirate	
Moderate or heavy growth	1
Same morphology on Gram's stain	1 (additional)
Maximal score[a]	12

[a]At the time of the original diagnosis, the progression of the infiltrate is not known and tracheal aspirate culture results are often unavailable; thus, the maximal score is initially 8–10.

Abbreviations: ARDS, acute respiratory distress syndrome; CHF, congestive heart failure.

Treatment: Ventilator-Associated Pneumonia

Many studies have demonstrated higher mortality rates with inappropriate than with appropriate empirical antibiotic therapy. The key to appropriate antibiotic management of VAP is an appreciation of the patterns of resistance of the most likely pathogens in any given patient.

ANTIBIOTIC RESISTANCE

If it were not for the risk of infection with MDR pathogens (Table 257-1), VAP could be treated with the same antibiotics used for severe CAP. However, antibiotic selection pressure leads to the frequent involvement of MDR pathogens by selecting either for drug-resistant isolates of common pathogens (MRSA and extended-spectrum β-lactamase–positive Enterobacteriaceae) or for intrinsically resistant pathogens (*P. aeruginosa* and *Acinetobacter* spp.). Frequent use of β-lactam drugs, especially cephalosporins, appears to be the major risk factor for infection with MRSA and extended spectrum β-lactamase–positive strains.

P. aeruginosa has demonstrated the ability to develop resistance to all routinely used antibiotics. Unfortunately, even if initially sensitive, *P. aeruginosa* isolates have also shown a propensity to develop resistance during treatment. Either derepression of resistance genes or selection of resistant clones within the large bacterial inoculum associated with most pneumonias may be the cause. *Acinetobacter* spp., *Stenotrophomonas maltophilia*, and *Burkholderia cepacia* are intrinsically resistant to many of the empirical antibiotic regimens employed (see below). VAP caused by these pathogens emerges during treatment of other infections, and resistance is always evident at initial diagnosis.

EMPIRICAL THERAPY

Recommended options for empirical therapy are listed in Table 257-8. Treatment should be started once diagnostic specimens have been obtained. The major factor in the selection of agents is the presence of risk factors for MDR pathogens. Choices among the various options listed depend on local patterns of resistance and the patient's prior antibiotic exposure.

Table 257-8 Empirical Antibiotic Treatment of Health Care–Associated Pneumonia

Patients without Risk Factors for MDR Pathogens
Ceftriaxone (2 g IV q24h) *or*
Moxifloxacin (400 mg IV q24h), ciprofloxacin (400 mg IV q8h), or levofloxacin (750 mg IV q24h) *or*
Ampicillin/sulbactam (3 g IV q6h) *or*
Ertapenem (1 g IV q24h)
Patients with Risk Factors for MDR Pathogens
1. A β-lactam: Ceftazidime (2 g IV q8h) or cefepime (2 g IV q8–12h) ***or*** Piperacillin/tazobactam (4.5 g IV q6h), imipenem (500 mg IV q6h or 1 g IV q8h), or meropenem (1 g IV q8h) ***plus***

2. A second agent active against gram-negative bacterial pathogens: Gentamicin or tobramycin (7 mg/kg IV q24h) or amikacin (20 mg/kg IV q24h) ***or*** Ciprofloxacin (400 mg IV q8h) or levofloxacin (750 mg IV q24h) ***plus***
3. An agent active against gram-positive bacterial pathogens: Linezolid (600 mg IV q12h) ***or*** Vancomycin (15 mg/kg, up to 1 g IV, q12h)

Abbreviation: MDR, multidrug-resistant.

The majority of patients *without* risk factors for MDR infection can be treated with a single agent. The major difference from CAP is the markedly lower incidence of atypical pathogens in VAP; the exception is *Legionella*, which can be a nosocomial pathogen, especially with breakdowns in the treatment of potable water in the hospital.

The standard recommendation for patients *with* risk factors for MDR infection is for three antibiotics: two directed at *P. aeruginosa* and one at MRSA. The choice of a β-lactam agent provides the greatest variability in coverage, yet the use of the broadest-spectrum agent—a carbapenem, even in an antibiotic combination—still represents inappropriate initial therapy in 10–15% of cases.

SPECIFIC TREATMENT

Once an etiologic diagnosis is made, broad-spectrum empirical therapy can be modified to address the known pathogen specifically. For patients with MDR risk factors, antibiotic regimens can be reduced to a single agent in more than one-half of cases and to a two-drug combination in more than one-quarter of cases. Only a minority of cases require a complete course with three drugs. A negative tracheal-aspirate culture or growth below the threshold for quantitative cultures, especially if the sample was obtained before any antibiotic change, strongly suggests that antibiotics should be discontinued. Identification of other confirmed or suspected sites of infection may require ongoing antibiotic therapy, but the spectrum of pathogens (and the corresponding antibiotic choices) may be different from those for VAP. If the CPIS decreases over the first 3 days, antibiotics should be stopped after 8 days. An 8-day course of therapy is just as effective as a 2-week course and is associated with less frequent emergence of antibiotic-resistant strains.

The major controversy regarding specific therapy for VAP concerns the need for ongoing combination treatment of *Pseudomonas* infection. No randomized controlled trials have demonstrated a benefit of combination therapy with a β-lactam and an aminoglycoside, nor have subgroup analyses in other trials found a survival benefit with such a regimen. The unacceptably high rates of clinical failure and death for VAP caused by *P. aeruginosa* despite combination therapy (see "Failure to Improve," below) indicate that better regimens are needed—including, perhaps, aerosolized antibiotics.

VAP caused by MRSA is associated with a 40% clinical failure rate when treated with standard-dose vancomycin. One proposed solution is the use of high-dose individualized treatment, although the risk of renal toxicity increases with this strategy. In addition, the MIC of vancomycin has been increasing, and a high percentage of clinical failures occur when the MIC is in the upper range of sensitivity (i.e., 1.5–2 μg/mL). Linezolid appears to be more efficacious than the standard dose of vancomycin and may be the preferred agent in patients with renal insufficiency and in those infected with high-MIC isolates of MRSA.

FAILURE TO IMPROVE

Treatment failure is not uncommon in VAP, especially in that caused by MDR pathogens. In addition to the 40% failure rate for MRSA infection treated with vancomycin, VAP due to *Pseudomonas* has a 50% failure rate, no matter what the regimen. The causes of clinical failure vary with the pathogen(s) and the antibiotic(s). Inappropriate therapy can usually be minimized by use of the recommended triple-drug regimen (Table 257-8). However, the emergence of β-lactam resistance during therapy is an important problem, especially in infection with *Pseudomonas* and *Enterobacter* spp. Recurrent VAP caused by the same pathogen is possible because the biofilm on endotracheal tubes allows reintroduction of the microorganism. However, studies of VAP caused by *Pseudomonas* show that approximately one-half of recurrent cases are caused by a new strain. Inadequate local levels of vancomycin are the likely cause of treatment failure in VAP due to MRSA.

Treatment failure is very difficult to diagnose. Pneumonia due to a new superinfection, the presence of extrapulmonary infection, and drug toxicity must be considered in the differential diagnosis of treatment failure. Serial CPIS appears to track the clinical response accurately, while repeat quantitative cultures may clarify the microbiologic response. A persistently elevated or rising CPIS value by day 3 of therapy is likely to indicate failure. The most sensitive component of the CPIS is improvement in oxygenation.

COMPLICATIONS

Apart from death, the major complication of VAP is prolongation of mechanical ventilation, with

corresponding increases in length of stay in the ICU and in the hospital. In most studies, an additional week of mechanical ventilation because of VAP is common. The additional expense of this complication often warrants costly and aggressive efforts at prevention.

In rare cases, some types of necrotizing pneumonia (e.g., that due to *P. aeruginosa*) result in significant pulmonary hemorrhage. More commonly, necrotizing infections result in the long-term complications of bronchiectasis and parenchymal scarring leading to recurrent pneumonias. The long-term complications of pneumonia are underappreciated. Pneumonia results in a catabolic state in a patient already nutritionally at risk. The muscle loss and general debilitation from an episode of VAP often require prolonged rehabilitation and, in the elderly, commonly result in an inability to return to independent function and the need for nursing home placement.

FOLLOW-UP

Clinical improvement, if it occurs, is usually evident within 48–72 h of the initiation of antimicrobial treatment. Because findings on chest radiography often worsen initially during treatment, they are less helpful than clinical criteria as an indicator of clinical response in severe pneumonia. Seriously ill patients with pneumonia often undergo follow-up chest radiography daily, at least until they are being weaned off mechanical ventilation. Once a patient has been extubated and is in stable condition, follow-up radiographs may not be necessary for a few weeks.

PROGNOSIS

VAP is associated with significant mortality. Crude mortality rates of 50–70% have been reported, but the real issue is attributable mortality. Many patients with VAP have underlying diseases that would result in death even if VAP did not occur. Attributable mortality exceeded 25% in one matched cohort study. Patients who develop VAP are at least twice as likely to die as those who do not. Some of the variability in VAP mortality rates is clearly related to the type of patient and ICU studied. VAP in trauma patients is not associated with attributable mortality, possibly because many of the patients were otherwise healthy before being injured. However, the causative pathogen also plays a major role. Generally, MDR pathogens are associated with significantly greater attributable mortality than non-MDR pathogens. Pneumonia caused by some pathogens (e.g., *S. maltophilia*) is simply a marker for a patient whose immune system is so compromised that death is almost inevitable.

PREVENTION

(Table 257-6) Because of the significance of the endotracheal tube as a risk factor for VAP, the most important preventive intervention is to avoid endotracheal intubation or at least to minimize its duration. Successful use of noninvasive ventilation via a nasal or full-face mask avoids many of the problems associated with endotracheal tubes. Strategies that minimize the duration of ventilation through daily holding of sedation and formal weaning protocols have also been highly effective in preventing VAP.

Unfortunately, a tradeoff in risks is sometimes required. Aggressive attempts to extubate early may result in reintubation(s) and increase aspiration, posing a risk of VAP. Heavy continuous sedation increases the risk, but self-extubation because of too little sedation is also a risk. The tradeoffs also apply to antibiotic therapy. Short-course antibiotic prophylaxis can decrease the risk of VAP in comatose patients requiring intubation, and data suggest that antibiotics decrease VAP rates in general. However, the major benefit appears to be a decrease in the incidence of early-onset VAP, which is usually caused by the less pathogenic non-MDR microorganisms. Conversely, prolonged courses of antibiotics consistently increase the risk of VAP caused by the more lethal MDR pathogens. Despite its virulence and associated mortality, VAP caused by *Pseudomonas* is rare among patients who have not recently received antibiotics.

Minimizing the amount of microaspiration around the endotracheal tube cuff is also a strategy for avoidance of VAP. Simply elevating the head of the bed (at least 30° above horizontal but preferably 45°) decreases VAP rates. Specially modified endotracheal tubes that allow removal of the secretions pooled above the cuff may also prevent VAP. The risk-to-benefit ratio of transporting the patient outside the ICU for diagnostic tests or procedures should be carefully considered, since VAP rates are increased among transported patients.

Emphasis on the avoidance of agents that raise gastric pH and on oropharyngeal decontamination has been diminished by the equivocal and conflicting results of more recent clinical trials. The role in the pathogenesis of VAP that is played by the overgrowth of bacterial components of the bowel flora in the stomach has also been downplayed. MRSA and the nonfermenters *P. aeruginosa* and *Acinetobacter* spp. are not normally part of the bowel flora but reside primarily in the nose and on the skin, respectively. Therefore, an emphasis on controlling overgrowth of the bowel flora may be relevant only in certain populations, such as liver transplant recipients and patients who have undergone other major intraabdominal procedures or who have bowel obstruction.

In outbreaks of VAP due to specific pathogens, the possibility of a breakdown in infection control measures (particularly contamination of reusable equipment) should be investigated. Even high rates of pathogens that are already common in a particular ICU may be a result of cross-infection. Education and reminders of the need for consistent hand washing and other infection control practices can

minimize this risk.

Hospital-Acquired Pneumonia

While significantly less well studied than VAP, HAP in nonintubated patients—both inside and outside the ICU—is similar to VAP. The main differences are in the higher frequency of non-MDR pathogens and the better underlying host immunity in nonintubated patients. The lower frequency of MDR pathogens allows monotherapy in a larger proportion of cases of HAP than of VAP.

The only pathogens that may be more common in the non-VAP population are anaerobes. The greater risk of macroaspiration by nonintubated patients and the lower oxygen tensions in the lower respiratory tract of these patients increase the likelihood of a role for anaerobes. While more common in patients with HAP, anaerobes are usually only contributors to polymicrobial pneumonias except in patients with large-volume aspiration or in the setting of bowel obstruction/ileus. As in the management of CAP, specific therapy targeting anaerobes probably is not indicated (unless gross aspiration is a concern) since many of the recommended antibiotics are active against anaerobes.

Diagnosis is even more difficult for HAP in the nonintubated patient than for VAP. Lower respiratory tract samples appropriate for culture are considerably more difficult to obtain from nonintubated patients. Many of the underlying diseases that predispose a patient to HAP are also associated with an inability to cough adequately. Since blood cultures are infrequently positive (<15% of cases), the majority of patients with HAP do not have culture data on which antibiotic modifications can be based. Therefore, de-escalation of therapy is less likely in patients with risk factors for MDR pathogens. Despite these difficulties, the better host defenses in non-ICU patients result in lower mortality rates than are documented for VAP. In addition, the risk of antibiotic failure is lower in HAP.

FURTHER READINGS

American Thoracic Society/Infectious Diseases Society of America: Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. Am J Respir Crit Care Med 171:388, 2005

Chastre J, Fagon JY: Ventilator-associated pneumonia. Am J Respir Crit Care Med 165:867, 2002[PMID: 11934711] [Full Text]

Fagon JY et al: Invasive and noninvasive strategies for management of suspected ventilator-associated pneumonia. A randomized trial. Ann Intern Med 132:621, 2000[PMID: 10766680] [Full Text]

Fine MJ et al: A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med 336:243, 1997[PMID: 8995086] [Full Text]

Lim WS et al: Defining community acquired pneumonia severity on presentation to hospital: An international derivation and validation study. Thorax 58:377, 2003[PMID: 12728155] [Full Text]

Mandell LA et al: Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia. Clin Infect Dis 44(Suppl 2): S27, 2007

Singh N et al: Short-course empiric antibiotic therapy for patients with pulmonary infiltrates in the intensive care unit. A proposed solution for indiscriminate antibiotic prescription. Am J Respir Crit Care Med 162:505, 2000[PMID: 10934078] [Full Text]

Vanderkooi OG et al: Predicting antimicrobial resistance in invasive pneumococcal infections. Clin Infect Dis 40:1288, 2005[PMID: 15825031] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BRONCHIECTASIS

Bronchiectasis refers to an irreversible airway dilation that involves the lung in either a focal or a diffuse manner and that classically has been categorized as cylindrical or tubular (the most common form), varicose, or cystic.

Etiology

Bronchiectasis can arise from infectious or noninfectious causes **(Table 258-1)**. Clues to the underlying etiology are often provided by the pattern of lung involvement. *Focal bronchiectasis* refers to bronchiectatic changes in a localized area of the lung and can be a consequence of obstruction of the airway—either extrinsic (e.g., due to compression by adjacent lymphadenopathy or parenchymal tumor mass) or intrinsic (e.g., due to an airway tumor or aspirated foreign body, a scarred/stenotic airway, or bronchial atresia from congenital underdevelopment of the airway). *Diffuse bronchiectasis* is characterized by widespread bronchiectatic changes throughout the lung and often arises from an underlying systemic or infectious disease process.

Table 258-1 Major Etiologies of Bronchiectasis and Proposed Workup

Pattern of Lung Involvement by Bronchiectasis	Etiology by Categories (with Specific Examples)	Workup
Focal	Obstruction (e.g., aspirated foreign body, tumor mass)	Chest imaging (chest x-ray and/or chest CT); bronchoscopy
Diffuse	Infection (e.g., bacterial, nontuberculous mycobacterial)	Gram's stain/culture; stains/cultures for acid-fast bacilli and fungi. If no pathogen is identified, consider bronchoscopy with bronchoalveolar lavage (BAL)
	Immunodeficiency (e.g., hypogammaglobulinemia, HIV infection, bronchiolitis obliterans after lung transplantation)	Complete blood count with differential; immunoglobulin measurement; HIV testing
	Genetic causes (e.g., cystic fibrosis, Kartagener's syndrome, α_1 antitrypsin deficiency)	Measurement of chloride levels in sweat (for cystic fibrosis), α_1 antitrypsin levels; nasal or respiratory tract brush/biopsy (for dyskinetic/immotile cilia syndrome); genetic testing
	Autoimmune or rheumatologic causes (e.g., rheumatoid arthritis, Sjögren's syndrome, inflammatory bowel disease); immune-mediated disease (e.g., allergic bronchopulmonary aspergillosis)	Clinical examination with careful joint exam, serologic testing (e.g., for rheumatoid factor). Consider workup for allergic bronchopulmonary aspergillosis, especially in patients with refractory asthma[a]
	Recurrent aspiration	Test of swallowing function and general neuromuscular strength
	Miscellaneous (e.g., yellow nail syndrome; traction bronchiectasis from postradiation fibrosis or idiopathic pulmonary fibrosis)	Guided by clinical condition
	Idiopathic	Exclusion of other causes

[a]Skin testing for *Aspergillus* reactivity; measurement of serum precipitins for *Aspergillus*, serum IgE levels, serum eosinophils, etc.

More pronounced involvement of the upper lung fields is most common in cystic fibrosis (CF) and is also observed in postradiation fibrosis, corresponding to the lung region encompassed by the radiation

port. Bronchiectasis with predominant involvement of the lower lung fields usually has its source in chronic recurrent aspiration (e.g., due to esophageal motility disorders like those in scleroderma), end-stage fibrotic lung disease (e.g., traction bronchiectasis from idiopathic pulmonary fibrosis), or recurrent immunodeficiency-associated infections (e.g., hypogammaglobulinemia). Bronchiectasis resulting from infection by nontuberculous mycobacteria [NTM; most commonly the *Mycobacterium avium-intracellulare* complex (MAC)] often preferentially affects the midlung fields. Congenital causes of bronchiectasis with predominant midlung field involvement include the dyskinetic/immotile cilia syndrome. Finally, predominant involvement of the central airways is reported in association with allergic bronchopulmonary aspergillosis (ABPA), in which an immune-mediated reaction to *Aspergillus* damages the bronchial wall. Congenital causes of central airway–predominant bronchiectasis resulting from cartilage deficiency include tracheobronchomegaly (Mounier-Kuhn syndrome) and Williams-Campbell syndrome.

In many cases, the etiology of bronchiectasis is not determined. In case series, as many as 25–50% of patients referred for bronchiectasis have idiopathic disease.

Epidemiology

The epidemiology of bronchiectasis varies greatly with the underlying etiology. For example, patients born with CF often develop significant clinical bronchiectasis in late adolescence or early adulthood, although atypical presentations of CF in adults in their thirties and forties are also possible. In contrast, bronchiectasis resulting from MAC infection classically affects nonsmoking women older than age 50 years. In general, the incidence of bronchiectasis increases with age. Bronchiectasis is more common among women than among men.

In areas where tuberculosis is prevalent, bronchiectasis more frequently occurs as a sequela of granulomatous infection. Focal bronchiectasis can arise from extrinsic compression of the airway by enlarged granulomatous lymph nodes and/or from development of intrinsic obstruction as a result of erosion of a calcified lymph node through the airway wall (e.g., broncholithiasis). Especially in reactivated tuberculosis, parenchymal destruction from infection can result in areas of more diffuse bronchiectasis. Apart from cases associated with tuberculosis, an increased incidence of non-CF bronchiectasis with an unclear underlying mechanism has been reported as a significant problem in developing nations. It has been suggested that the high incidence of malnutrition in certain areas may predispose to immune dysfunction and development of bronchiectasis.

Pathogenesis and Pathology

The most widely cited mechanism of infectious bronchiectasis is the "vicious cycle hypothesis," in which susceptibility to infection and poor mucociliary clearance result in microbial colonization of the bronchial tree. Some organisms, such as *Pseudomonas aeruginosa*, exhibit a particular propensity for colonizing damaged airways and evading host defense mechanisms. Impaired mucociliary clearance can result from inherited conditions such as CF or dyskinetic cilia syndrome, and it has been proposed that a single severe infection (e.g., pneumonia caused by *Bordetella pertussis* or *Mycoplasma pneumoniae*) can result in significant airway damage and poor secretion clearance. The presence of the microbes incites continued chronic inflammation, with consequent damage to the airway wall, continued impairment of secretion and microbial clearance, and ongoing propagation of the infectious/inflammatory cycle. Moreover, it has been proposed that mediators released directly from bacteria can interfere with mucociliary clearance.

Classic studies of the pathology of bronchiectasis from the 1950s demonstrated significant small-airway wall inflammation and larger-airway wall destruction as well as dilation, with loss of elastin, smooth muscle, and cartilage. It has been proposed that inflammatory cells in the small airways release proteases and other mediators, such as reactive oxygen species and proinflammatory cytokines, that damage the larger-airway walls. Furthermore, the ongoing inflammatory process in the smaller airways results in airflow obstruction. It is believed that antiproteases, such as α_1 antitrypsin, play an important role in neutralizing the damaging effects of neutrophil elastase and in enhancing bacterial killing. In addition to emphysema, bronchiectasis has been observed in patients with α_1 antitrypsin deficiency.

Proposed mechanisms for noninfectious bronchiectasis include immune-mediated reactions that damage the bronchial wall (e.g., those associated with systemic autoimmune conditions such as Sjögren's syndrome and rheumatoid arthritis). *Traction bronchiectasis* refers to dilated airways arising from parenchymal distortion as a result of lung fibrosis (e.g., postradiation fibrosis or idiopathic pulmonary fibrosis).

Clinical Manifestations

The most common clinical presentation is a persistent productive cough with ongoing production of thick, tenacious sputum. Physical findings often include crackles and wheezing on lung auscultation, and some patients with bronchiectasis exhibit clubbing of the digits. Mild to moderate airflow obstruction is often detected on pulmonary function tests, overlapping with that seen at presentation with other conditions, such as chronic obstructive pulmonary disease (COPD). Acute exacerbations of

bronchiectasis are usually characterized by changes in the nature of sputum production, with increased volume and purulence. However, typical signs and symptoms of lung infection, such as fever and new infiltrates, may not be present.

Diagnosis

The diagnosis is usually based on presentation with a persistent chronic cough and sputum production accompanied by consistent radiographic features. While chest radiographs lack sensitivity, the presence of "tram tracks" indicating dilated airways is consistent with bronchiectasis. Chest CT is more specific for bronchiectasis and is the imaging modality of choice for confirming the diagnosis. CT findings include airway dilation (detected as parallel "tram tracks" or as the "signet-ring sign"—a cross-sectional area of the airway with a diameter at least 1.5 times that of the adjacent vessel), lack of bronchial tapering (including the presence of tubular structures within 1 cm from the pleural surface), bronchial wall thickening in dilated airways, inspissated secretions (e.g., the "tree-in-bud" pattern), or cysts emanating from the bronchial wall (especially pronounced in cystic bronchiectasis; **Fig. 258-1**).

Figure 258-1

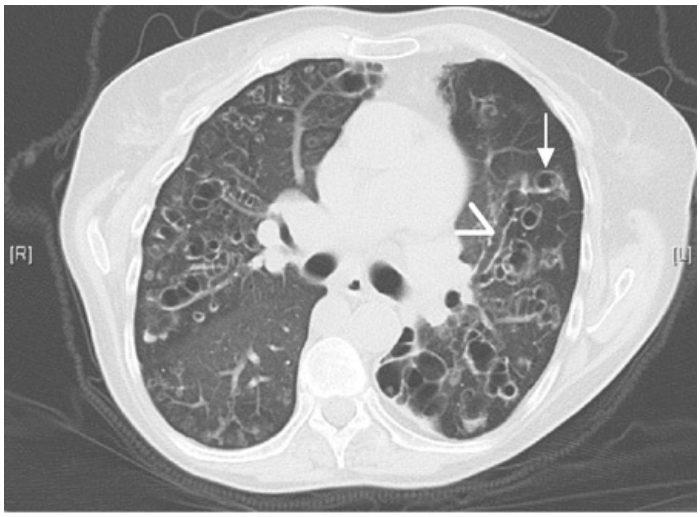

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Representative chest CT image of severe bronchiectasis. This patient's CT demonstrates many severely dilated airways, seen both longitudinally (arrowhead) and in cross-section (arrow).

Approach to the Patient: Bronchiectasis

The evaluation of a patient with bronchiectasis entails elicitation of a clinical history, chest imaging, and a workup to determine the underlying etiology. Evaluation of focal bronchiectasis almost always requires bronchoscopy to exclude airway obstruction by an underlying mass or foreign body. A workup for diffuse bronchiectasis includes analysis for the major etiologies (Table 258-1). Pulmonary function testing is an important component of a functional assessment of the patient.

Treatment: Bronchiectasis

Treatment of infectious bronchiectasis is directed at the control of active infection and improvements in secretion clearance and bronchial hygiene so as to decrease the microbial load within the airways and minimize the risk of repeated infections.

ANTIBIOTIC TREATMENT

Antibiotics targeting the causative or presumptive pathogen (with *Haemophilus influenzae* and *P. aeruginosa* isolated commonly) should be administered in acute exacerbations, usually for a minimum of 7–10 days. Decisions about treatment of NTM infection can be difficult, given that these organisms can be colonizers as well as pathogens and the prolonged treatment course often is not well tolerated. Consensus guidelines have advised that diagnostic criteria for true clinical infection with NTM should be considered in patients with symptoms and radiographic findings of lung disease who have at least two sputum samples positive on culture; at least one bronchoalveolar lavage (BAL) fluid sample positive on culture; a biopsy sample displaying histopathologic features of NTM infection (e.g., granuloma or a positive stain for acid-fast bacilli) along with one positive sputum culture; or a pleural fluid sample (or a sample from another sterile extrapulmonary site) positive on culture. MAC strains are the most common NTM pathogens, and the recommended regimen for HIV-negative patients includes a macrolide combined with rifampin and ethambutol. Consensus guidelines also recommend macrolide susceptibility testing for clinically significant MAC isolates.

BRONCHIAL HYGIENE

The numerous approaches employed to enhance secretion clearance in bronchiectasis include hydration and mucolytic administration, aerosolization of bronchodilators and hyperosmolar agents (e.g., hypertonic saline), and chest physiotherapy (e.g., postural drainage, traditional mechanical chest percussion via hand clapping to the chest, or use of devices such as an oscillatory positive expiratory pressure flutter valve or a high-frequency chest wall oscillation vest). The mucolytic dornase (DNase) is recommended routinely in CF-related bronchiectasis but not in non-CF bronchiectasis, given concerns about lack of efficacy and potential harm in the non-CF population.

ANTI-INFLAMMATORY THERAPY

It has been proposed that control of the inflammatory response may be of benefit in bronchiectasis, and relatively small-scale trials have yielded evidence of alleviated dyspnea, decreased need for inhaled β-agonists, and reduced sputum production with inhaled glucocorticoids. However, no significant differences in lung function or bronchiectasis exacerbation rates have been observed. Risks of immunosuppression and adrenal suppression must be carefully considered with use of anti-inflammatory therapy in infectious bronchiectasis. Nevertheless, administration of oral/systemic glucocorticoids may be important in treating bronchiectasis due to certain etiologies, such as ABPA, or noninfectious bronchiectasis due to underlying conditions, especially that in which an autoimmune condition is believed to be active (e.g., rheumatoid arthritis or Sjögren's syndrome). Patients with ABPA may also benefit from a prolonged course of treatment with the oral antifungal agent itraconazole.

REFRACTORY CASES

In select cases, surgery can be considered, with resection of a focal area of suppuration. In advanced cases, lung transplantation can be considered.

Complications

In more severe cases of infectious bronchiectasis, recurrent infections and repeated courses of antibiotics can lead to microbial resistance to antibiotics. In certain cases, combinations of antibiotics that have their own independent toxicity profiles may be necessary to treat resistant organisms.

Recurrent infections can result in injury to superficial mucosal vessels, with bleeding and, in severe cases, life-threatening hemoptysis. Management of massive hemoptysis usually requires intubation to stabilize the patient, identifying the source of bleeding, and protecting the nonbleeding lung. Control of bleeding often necessitates bronchial artery embolization and, in severe cases, surgery.

Prognosis

Outcomes of bronchiectasis vary widely with the underlying etiology and may also be influenced by the frequency of exacerbations and (in infectious cases) the specific pathogens involved. In one study, the decline of lung function in patients with non-CF bronchiectasis was similar to that in patients with COPD, with the forced expiratory volume in 1 s (FEV_1) declining by 50–55 mL per year as opposed to 20–30 mL per year for healthy controls.

Prevention

Reversal of an underlying immunodeficient state (e.g., by administration of gamma globulin for immunoglobulin-deficient patients) and vaccination of patients with chronic respiratory conditions (e.g., influenza and pneumococcal vaccines) can decrease the risk of recurrent infections. Patients who smoke should be counseled about smoking cessation.

After resolution of an acute infection in patients with recurrences (e.g., ≥3 episodes per year), the use of suppressive antibiotics to minimize the microbial load and reduce the frequency of exacerbations has been proposed, although there is less consensus with regard to this approach in non-CF-associated bronchiectasis than there is in patients with CF-related bronchiectasis. Possible suppressive treatments include (1) administration of an oral antibiotic (e.g., ciprofloxacin) daily for 1–2 weeks per month; (2) use of a rotating schedule of oral antibiotics (to minimize the risk of development of drug resistance); (3) administration of a macrolide antibiotic daily or three times per week (with mechanisms of possible benefit related to non-antimicrobial properties, such as anti-inflammatory effects and reduction of gram-negative bacillary biofilms); (4) inhalation of aerosolized antibiotics [e.g., tobramycin inhalation solution (TOBI)] by select patients on a rotating schedule (e.g., 30 days on, 30 days off) with the goal of decreasing the microbial load without encountering the side effects of systemic drug administration; and (5) intermittent administration of IV antibiotics (e.g., "clean-outs") for patients with more severe bronchiectasis and/or resistant pathogens.

In addition, ongoing, consistent attention to bronchial hygiene can promote secretion clearance and decrease the microbial load in the airways.

LUNG ABSCESS

The term *lung abscess* refers to a microbial infection of the lung that results in necrosis of the pulmonary parenchyma. *Necrotizing pneumonia* or *lung gangrene* refers to multiple small pulmonary abscesses in contiguous areas of the lung, usually resulting from a more virulent infection.

Classification

Lung abscesses are classified by clinical and pathologic features including the tempo of progression, the presence or absence of an associated underlying lesion, and the microbial pathogen responsible. Duration defines the infection as *acute* versus *chronic*, with the dividing line usually at 4–6 weeks. Abscesses occurring in the presence of underlying pulmonary lesions, including tumors or systemic conditions (e.g., HIV infection), are referred to as *secondary*; those that occur in the absence of underlying pulmonary lesions are considered *primary*. The term *nonspecific lung abscess* refers to cases in which no likely pathogen is recovered from expectorated sputum; most such cases are presumed to be due to anaerobic bacteria. *Putrid lung abscess* is a term applied to anaerobic bacterial lung abscesses, which are characterized by distinctive foul-smelling breath, sputum, or empyema fluid.

Etiology

The likely etiologic agent, appropriate diagnostic testing, and appropriate treatment are frequently indicated by the characteristics of the host and the disease process. A variety of microbial pathogens cause lung abscess **(Table 258-2)**. Most nonspecific lung abscesses are presumed to be due to anaerobic bacteria. Mycobacteria, especially *M. tuberculosis*, are a very important cause of pulmonary infections and abscess formation. Fungi and some parasites also cause lung abscess. An acute lung abscess developing in a young, previously healthy patient, especially in conjunction with influenza, is likely to involve *Staphylococcus aureus*; this pathogen generally is seen easily on sputum Gram's stain and culture, and presumptive treatment for methicillin-resistant *S. aureus* is urgent. In an immunocompromised host, suspected pathogens include enteric gram-negative bacilli—especially *Klebsiella pneumoniae* but also agents that are found almost exclusively in patients with defective cell-mediated immunity, such as *Nocardia asteroides* and *Rhodococcus equi*. Lung abscess acquired in other countries may involve *Burkholderia pseudomallei* or *Paragonimus westermani*.

Table 258-2 Microbial Pathogens Causing Cavitary Lung Infection

Aspiration-Prone Host
Anaerobic bacteria plus microaerophilic and/or anaerobic streptococci, *Gemella* spp.
Embolic (endovascular) lesions: usually *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Fusobacterium necrophorum*[a]
Endemic fungi: *Histoplasma*, *Blastomyces*, *Coccidioides* spp.
Mycobacteria: *M. tuberculosis*, *M. kansasii*, *M. avium*
Immunocompromised Host
M. tuberculosis, *Nocardia asteroides*, *Rhodococcus equi*, *Legionella* spp., *P. aeruginosa*, Enterobacteriaceae (especially *Klebsiella pneumoniae*), *Aspergillus* spp., *Cryptococcus* spp.
Previously Healthy Host
Bacteria: *S. aureus*,[b] *S. milleri*, *K. pneumoniae*, group A *Streptococcus*; *Gemella*, *Legionella*, and *Actinomyces* spp.
Parasites: *Entamoeba histolytica*, *Paragonimus westermani*, *Strongyloides stercoralis*

[a] Lemierre's disease.

[b] Often in a young patient with influenza.

Multiple pulmonary lesions that are not caused by microbes may resemble lung abscess. These include the lesions of pulmonary infarction, bronchiectasis, necrotizing carcinoma, pulmonary sequestration, vasculitides [e.g., periarteritis nodosa, granulomatosis with polyangiitis (Wegener's), Goodpasture syndrome], and cysts or bullae with fluid collections. In some cases, multiple lung abscesses result from septic emboli, most commonly in association with tricuspid valve endocarditis.

Clinical Features

The classic presentation of nonspecific lung abscess is an indolent infection that evolves over several days or weeks, usually in a host who has a predisposition to aspiration. A common feature is periodontal infection with pyorrhea or gingivitis. Anaerobes and aerobic or microaerophilic streptococci that colonize the upper airways are implicated in these lesions. The usual symptoms are fatigue, cough, sputum production, and fever. Chills are uncommon. Many patients have evidence of chronic disease, such as weight loss and anemia. Some patients have putrid-smelling sputum indicative of the presence of anaerobes; the foul odor is presumably due to the organisms' production of short-chain fatty acids, such as butyric or succinic acid. Some patients have pleurisy due to pleural involvement by contiguous spread or by a bronchopleural fistula. The pleurisy may be severe and may be the symptom that prompts medical evaluation. Sequential x-rays or CT scans show the evolution of this

lesion from pneumonitis to cavitation, a process that generally requires 7–14 days in experimental animals (Fig. 258-2).

Figure 258-2

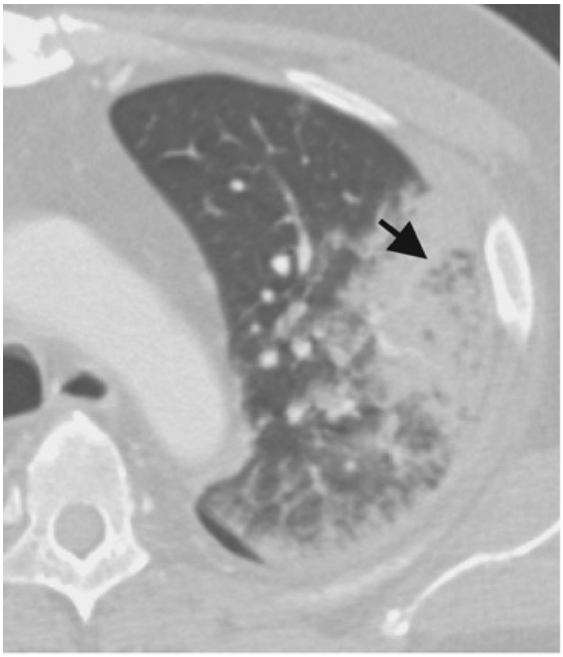

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Representative chest CT demonstrating development of lung abscesses. This patient was immunocompromised due to underlying lymphoma and developed severe *Pseudomonas aeruginosa* pneumonia, as represented by a left lung infiltrate with concern for central regions of necrosis (**panel A**, black arrow). Two weeks later, areas of cavitation with air fluid levels were visible in this region and were consistent with the development of lung abscesses (**panel B**, white arrow). (*Images provided by Dr. Ritu Gill, Division of Chest Radiology, Brigham and Women's Hospital, Boston.*)

Diagnosis

Lung abscess can usually be detected with standard imaging, including chest x-ray and CT (Fig. 258-

2). The latter is clearly preferred for precise definition of the lesion and its location and possibly for detection of underlying lesions. Lymphadenopathy is not associated with bacterial lung abscess; thus this finding suggests an alternative diagnosis.

Microbiologic studies include stains and cultures of expectorated sputum to detect aerobic bacterial pathogens. However, clinical correlations are very important because sputum cultures (especially those that do not satisfy standard cytologic criteria) are unreliable. In appropriate settings, it is important to consider cultures for fungi and mycobacteria. Anaerobic bacteria, the most common causes of primary lung abscess, are not detected in expectorated sputum cultures, and in any case the specimen is subject to anaerobic contamination as it traverses the upper airways. Alternative specimens that may be useful include pleural fluid obtained by thoracentesis in patients who have empyema and quantitative bronchoalveolar lavage (BAL) specimens if they are processed promptly and appropriately for anaerobic bacteria. Many reports describe the use of transtracheal aspiration to bypass the upper airways and obtain a specimen for meaningful anaerobic culture. This procedure, which was used extensively in the 1970s, has largely been abandoned out of concern about adverse consequences and because of a general decline in the pursuit of an etiologic agent in pulmonary infections. Another invasive method for bypassing contamination by the flora of the upper airways is transthoracic needle aspiration under CT guidance; the popularity of this procedure has increased in recent years. In most cases, the etiology of anaerobic lung abscess is clear: the host is prone to aspiration and has an abscess in a dependent pulmonary segment, with no other likely cause. As stated above, putrid breath, sputum, or empyema fluid indicates anaerobic infection.

Treatment: Lung Abscess

ANTIBIOTIC SELECTION

Treatment depends on the presumed or established etiology. Infections caused by anaerobic bacteria should usually be treated with clindamycin; the initial IV dosage of 600 mg four times daily can be changed to an oral dosage of 300 mg four times daily once the patient becomes afebrile and improves clinically. The duration of therapy is arbitrary, but many experts recommend continuation of oral treatment until imaging shows that chest lesions have cleared or have left a small, stable scar. A shorter course may be effective. An alternative to clindamycin is any β-lactam/β-lactamase inhibitor combination; parenteral treatment may be followed by orally administered amoxicillin/clavulanate. Carbapenems are also effective against anaerobic bacteria as well as streptococci, but the published data with these drugs in the treatment of anaerobic pulmonary infections are sparse. Penicillin was previously regarded as a preferred drug for these infections, but many oral anaerobes produce β-lactamases, and clindamycin proved superior to penicillin G in a randomized clinical trial. Metronidazole is highly active against virtually all anaerobes but not against aerobic microaerophilic streptococci, which play an important role in mixed infections. In therapeutic trials, metronidazole has done poorly unless combined with a β-lactam or another agent active against aerobic and microaerophilic streptococci.

Persistence of fever beyond 5–7 days or progression of the infiltrate suggests failure of therapy and a need to exclude factors such as obstruction, complicating empyema, and involvement of antibiotic-resistant bacteria. Many patients with uncomplicated lung abscesses and all those with atypical presentations or unresponsive abscesses should undergo bronchoscopy and/or CT to detect a possible associated anatomic lesion, such as a tumor, or a foreign body. Quantitative bacteriologic studies using a protected brush catheter or BAL are much less reliable when done after antibiotic therapy. Postural drainage was previously popular for patients with lung abscess, but aggressive attempts to implement this strategy may result in spillage to other pulmonary segments, leading to airway obstruction and clinical deterioration.

Lung abscess due to *S. aureus* is usually treated with vancomycin at a dosage that targets a trough serum level of 15–20 μg/mL. The main alternative is linezolid. Daptomycin should not be used for pulmonary infections. Lung abscesses caused by aerobic gram-negative bacteria need to be treated according to the results of antibiotic sensitivity tests. Most common among the pathogens involved are *K. pneumoniae* (especially the K1 strain in Taiwan) and *P. aeruginosa* in patients with severe chronic lung disease or compromised immune defenses. Pseudomonal lung abscesses usually require prolonged courses of parenteral antibiotics. Carbapenems or β-lactams are frequently combined with aminoglycosides; oral fluoroquinolones are often effective initially, but resistance is common with prolonged use. Aerosolized colistin and aminoglycosides are sometimes used to augment other therapy, but the efficacy of this approach is variable.

Surgery for lung abscesses was developed at the time penicillin became available in the late 1940s. The relative roles of penicillin and resectional surgery were hotly debated at that time, but by the late 1950s penicillin was favored. Initially the standard choice for most lung abscesses, penicillin was subsequently supplanted by the options summarized above. Recent large-scale reviews indicate that, in general, surgery is now reserved for ~10–12% of patients. The major indications for surgery are failure to respond to medical management, suspected neoplasm, and hemorrhage. Failure to respond to antibiotics is usually due to an obstructed bronchus and an extremely large abscess (>6 cm in diameter) or to infection involving relatively resistant bacteria, such as *P. aeruginosa*. The usual

procedure is lobectomy. An alternative intervention that is becoming popular is percutaneous drainage under CT guidance. Aspirate samples for assay of possible pathogens should be carefully collected.

RESPONSE TO THERAPY

Patients with lung abscess usually show clinical improvement, with decreased fever, within 3–5 days of initiation of antibiotic treatment. Defervescence can be expected within 5–10 days. Patients with fevers persisting for 7–14 days should undergo bronchoscopy or other diagnostic tests to better define anatomic changes and microbiologic findings. Cultures of expectorated sputum are not likely to be helpful at this juncture except for detecting pathogens such as mycobacteria and fungi. The response to therapy apparent on serial chest radiographs is delayed in comparison with the clinical course. In fact, infiltrates usually progress during the first 3 days of treatment in approximately one-half of patients. Pleural involvement is relatively common and may develop in dramatic fashion. The most common causes of failures of medical management include a failure to drain pleural collections, an inappropriate choice of antimicrobial therapy, an obstructed bronchus that prevents drainage, a "giant" abscess, a resistant pathogen, or refractory lesions due to immunocompromise.

FURTHER READINGS

Bronchiectasis

Barbato A et al: Primary ciliary dyskinesia: A consensus statement on diagnostic and treatment approaches in children. Eur Respir J 34:1264, 2009[PMID: 19948909] [Full Text]

Cantin L et al: Bronchiectasis. AJR Am J Roentgenol 193:W158, 2009

Griffith DE et al: An official ATS/IDSA statement: Diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. Am J Respir Crit Care Med 175:367, 2007[PMID: 17277290] [Full Text]

Kapur N et al: Inhaled steroids for bronchiectasis. Cochrane Database Syst Rev Jan 21(1):CD000996, 2009

Martinez-Garcia MA et al: Factors associated with lung function decline in adult patients with stable non-cystic fibrosis bronchiectasis. Chest 132:1565, 2007[PMID: 17998359] [Full Text]

Parr DG et al: Prevalence and impact of bronchiectasis in alpha1-antitrypsin deficiency. Am J Respir Crit Care Med 176:1215, 2007[PMID: 17872489] [Full Text]

Seitz AE et al: Trends and burdens of bronchiectasis-associated hospitalizations in the United States, 1993–2006. Chest 138:944, 2010[PMID: 20435655] [Full Text]

Shoemark A et al: Aetiology in adult patients with bronchiectasis. Respir Med 101:1163, 2007[PMID: 17223027] [Full Text]

Stevens DA et al: A randomized trial of itraconazole in allergic bronchopulmonary aspergillosis. N Engl J Med 342:756, 2000[PMID: 10717010] [Full Text]

Weycker D et al: Prevalence and economic burden of bronchiectasis (obstructive airways disease). Clin Pulm Med 12:205, 2005

Lung Abscess

Allewelt M et al: Ampicillin + sulbactam vs clindamycin +/– cephalosporin for the treatment of aspiration pneumonia and primary lung abscess. Clin Microbiol Infect 10:163, 2004[PMID: 14759242] [Full Text]

Bartlett JG: The role of anaerobic bacteria in lung abscess. Clin Infect Dis 40:923, 2005[PMID: 15824980] [Full Text]

———: Anaerobic bacterial infections of the lung and pleural space. Clin Infect Dis 16:S248, 1993

Fernández-Sabé N et al: Efficacy and safety of sequential amoxicillin-clavulanate in the treatment of anaerobic lung infections. Eur J Clin Microbiol Infect Dis 22:185, 2003[PMID: 21696853] [Full Text]

Francis JS et al: Severe community-onset pneumonia in healthy adults caused by methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes. Clin Infect Dis 40:100, 2005[PMID: 15614698] [Full Text]

Moreira Jda S et al: Lung abscess: Analysis of 252 consecutive cases diagnosed between 1968 and 2004. J Bras Pneumol 32:136, 2006[PMID: 17273583] [Full Text]

Shinzato T, Saito A: The *Streptococcus milleri* group as a cause of pulmonary infections. Clin Infect Dis 21(Suppl 3):S238, 1995

Takayanagi N et al: Etiology and outcome of community-acquired lung abscess. Respiration 80:98, 2010[PMID: 20389050] [Full Text]

Vardakas KZ et al: Comparison of community-acquired pneumonia due to methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* producing the Panton-Valentine leukocidin. Int J Tuberc Lung Dis 13:1476, 2009[PMID: 19919764] [Full Text]

Wang JL et al: Changing bacteriology of adult community-acquired lung abscess in Taiwan: *Klebsiella pneumoniae* vs. anaerobes. Clin Infect Dis 40:915, 2005[PMID: 16242281] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CYSTIC FIBROSIS: INTRODUCTION

Cystic fibrosis (CF) is a monogenic disorder that presents as a multisystem disease. The first signs and symptoms typically occur in childhood, but about 5% of patients in the United States are diagnosed as adults. Due to improvements in therapy, >46% of patients are now adults (≥18 years old) and 16.4% are past the age of 30. The median survival is >37.4 years for patients with CF; thus, CF is no longer only a pediatric disease, and internists must be prepared to recognize and treat its many complications. CF is characterized by chronic bacterial infection of the airways that leads to bronchiectasis and bronchiolectasis, exocrine pancreatic insufficiency and intestinal dysfunction, abnormal sweat gland function, and urogenital dysfunction.

PATHOGENESIS

Genetic Considerations

CF is an autosomal recessive disease resulting from mutations in the *CFTR* gene located on chromosome 7. The mutations in the *CFTR* gene fall into five major classes, as depicted in **Fig. 259-1**. Classes I–III mutations are considered "severe," as indexed by pancreatic insufficiency and high sweat NaCl values (see below). Class IV and V mutations can be "mild," i.e., associated with pancreatic sufficiency and intermediate/normal sweat NaCl values.

Figure 259-1

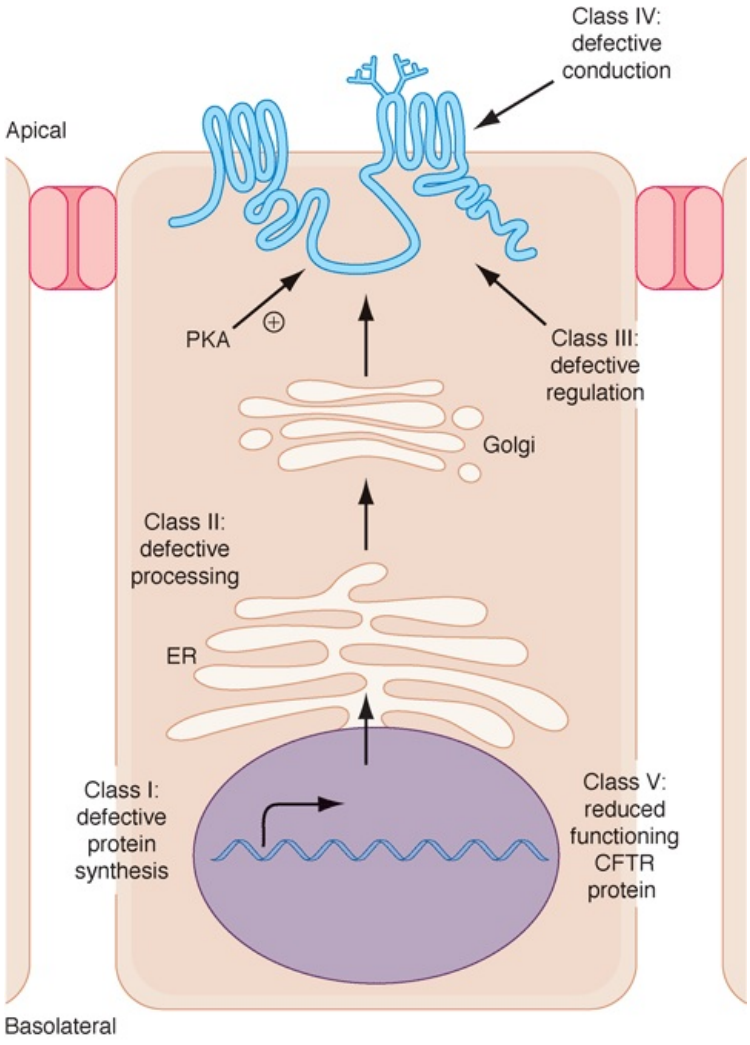

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schema describing classes of genetic mutations in CFTR gene and effects on CFTR protein/function. Note the ΔF_{508} mutation is a class II mutation and, like class I mutations, would be predicted to produce no mature CFTR protein in the apical membrane. CFTR, cystic fibrosis transmembrane conductance regulator.

The prevalence of CF varies with the ethnic origin of a population. CF is detected in approximately 1 in 3000 live births in the Caucasian population of North America and northern Europe, 1 in 17,000 live births of African Americans, and 1 in 90,000 live births of the Asian population of Hawaii. The most common mutation in the *CFTR* gene (~70% of CF chromosomes) is a 3-bp deletion (a class II mutation) that results in an absence of phenylalanine at amino acid position 508 (ΔF_{508}) of the CF gene protein product, known as cystic fibrosis transmembrane conductance regulator (CFTR). The large number (>1500) of relatively uncommon (<2% each) mutations identified in the *CFTR* gene makes genetic testing challenging.

CFTR Protein

The CFTR protein is a single polypeptide chain, containing 1480 amino acids, that functions both as a cyclic AMP–regulated Cl^- channel and a regulator of other ion channels. The fully processed form of CFTR localizes to the plasma membrane in normal epithelia. Biochemical studies indicate that the ΔF_{508} mutation leads to improper maturation and intracellular degradation of the mutant CFTR protein. Thus, absence of CFTR in the plasma membrane is central to the molecular pathophysiology of the ΔF_{508} mutation and other classes I–II mutations. Classes III–IV mutations produce CFTR proteins that are fully processed but are nonfunctional or only partially functional in the plasma membrane. Class V mutations include splicing mutations that produce small amounts of functional CFTR.

Epithelial Dysfunction

The epithelia affected by CF exhibit different functions in their native state, i.e., some are volume-absorbing (airways and distal intestinal epithelia), and some are salt- but not volume-absorbing (sweat duct), whereas others are volume-secretory (proximal intestine and pancreas). Given this diversity of

native activities, it is not surprising that CF produces organ-specific effects on electrolyte and water transport. However, the unifying concept is that all affected tissues express abnormal ion transport function.

Organ-Specific Pathophysiology

LUNG

The diagnostic biophysical hallmark of CF airway epithelia is the raised transepithelial electric potential difference (PD), which reflects both the rate of active ion transport and epithelial resistance to ion flow. CF airway epithelia exhibit abnormalities in both active Cl^- secretion and Na^+ absorption **(Fig. 259-2)**. The Cl^- secretory defect reflects the absence of cyclic AMP–dependent kinase and protein kinase C –regulated Cl^- transport mediated by CFTR itself. An important observation is that there is also a molecularly distinct Ca^{2+}-activated Cl^- channel (CaCC, TMEM16a) expressed in the apical membrane. This channel can substitute for CFTR with regard to Cl^- secretion and is a potential therapeutic target. The abnormal Na^+ transport reflects a second function of CFTR, its function as a tonic inhibitor of the epithelial Na^+ channel. The molecular mechanisms mediating this action of CFTR remain unknown.

Figure 259-2

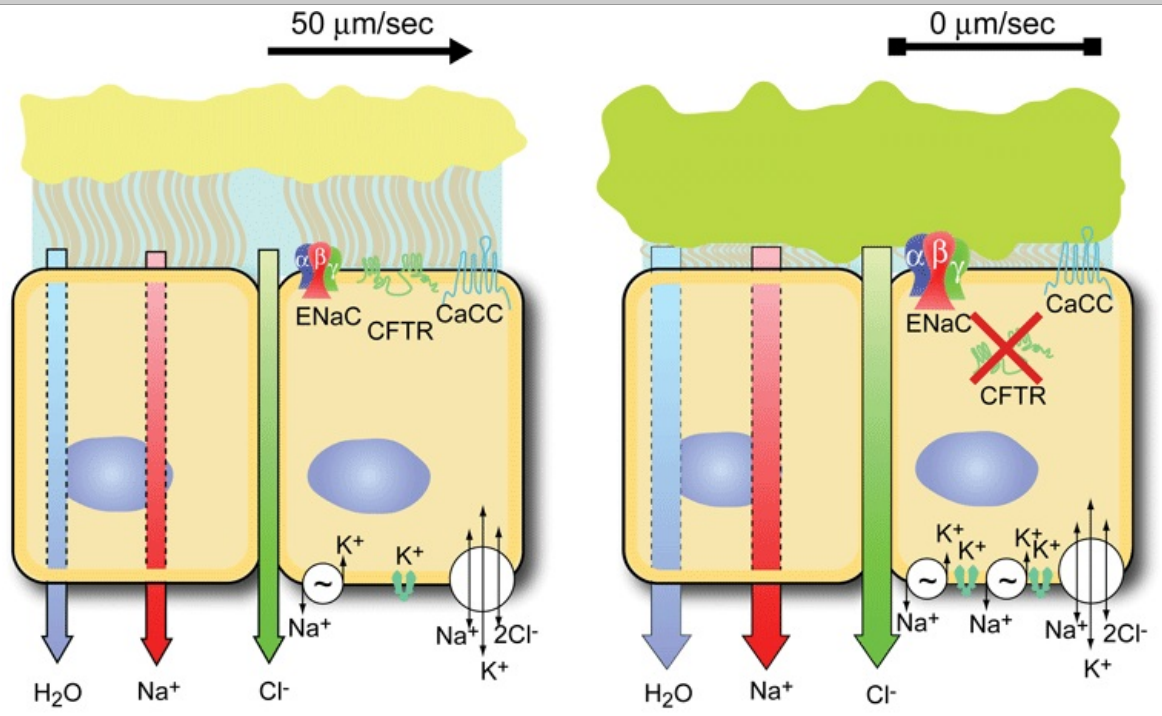

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Comparison of ion transport properties of normal (top) and CF (bottom) airway epithelia. The vectors describe routes and magnitudes of Na^+ and Cl^- transport that is accompanied by osmotically driven water flow. The normal basal pattern for ion transport is absorption of Na^+ from the lumen via an amiloride-sensitive epithelial Na^+ channel (ENaC) composed of α, β, and γ ENaC subunits. This process is accelerated in CF. The capacity to initiate cyclic AMP–mediated Cl^- secretion is diminished in CF airway epithelia due to abnormal maturation/dysfunction of the CFTR Cl^- channel. The accelerated Na^+ absorption in CF reflects the absence of CFTR inhibitory effects on Na^+ channels. A Ca^{2+}-activated Cl^- channel, likely a product of the TMEM16a gene, is expressed in normal and CF apical membranes and can be activated by extracellular ATP. Horizontal arrows depict velocity of mucociliary clearance (µm/sec).

Mucus clearance is the primary innate airways defense mechanism against infection by inhaled bacteria. Normal airways vary the rates of active Na^+ absorption and Cl^- secretion to adjust the volume of liquid (water), i.e., "hydration," on airway surfaces for efficient mucus clearance. The central hypothesis of CF airways pathophysiology is that the faulty regulation of Na^+ absorption and inability to secrete Cl^- via CFTR reduce the volume of liquid on airway surfaces; i.e., they are "dehydrated."

Dehydration of both the mucus and the periciliary liquid layers produces adhesion of mucus to the airway surface, which leads to a failure to clear mucus from the airways both by ciliary and cough-dependent mechanisms. The absence of a strict correspondence between gene-mutation class and severity of lung disease suggests important roles for modifier genes and gene-environmental interactions.

The infection that characterizes CF airways involves the mucus layer rather than epithelial or airway wall invasion. The predisposition of CF airways to chronic infection by *Staphylococcus aureus* and *Pseudomonas aeruginosa* is consistent with failure to clear mucus. Recently, it has been demonstrated that the O_2 tension is very low in CF mucus, and adaptations to hypoxemia are important determinants of the physiology of bacteria in the CF lung. Indeed, both mucus stasis and mucus hypoxemia contribute to (1) the propensity for *Pseudomonas* to grow in biofilm colonies within CF airway mucus and (2) the presence of strict anaerobes in CF lungs.

GASTROINTESTINAL TRACT

The gastrointestinal effects of CF are diverse. In the exocrine pancreas, the absence of the CFTR Cl^- channel in the apical membrane of pancreatic ductal epithelia limits the function of an apical membrane Cl^--HCO_3^- exchanger to secrete bicarbonate and Na^+ (by a passive process) into the duct. The failure to secrete Na^+ HCO_3^- and water leads to retention of enzymes in the pancreas and destruction of virtually all pancreatic tissue. Because of the lack of Cl^- and water secretion, the CF intestinal epithelium fails to flush secreted mucins and other macromolecules from intestinal crypts. The diminished CFTR-mediated liquid secretion may be exacerbated by excessive absorption of liquid, reflecting abnormalities of CFTR-mediated regulation of Na^+ absorption (both mediated by Na^+ channels and possibly other Na^+ transporters, e.g., Na^+-H^+ exchangers). Both dysfunctions lead to dehydrated intraluminal contents and intestinal obstruction. In the hepatobiliary system, defective hepatic ductal salt (Cl^-) and water secretion causes thickened biliary secretions, focal biliary cirrhosis, and bile-duct proliferation in approximately 25–30% of patients with CF. The inability of the CF gallbladder epithelium to secrete salt and water can lead to both chronic cholecystitis and cholelithiasis.

SWEAT GLAND

CF patients secrete nearly normal volumes of sweat in the sweat acinus, but are not able to absorb NaCl from the sweat duct due to the inability to absorb Cl^- via CFTR across ductal epithelial cells. This sweat gland dysfunction is typically measured by measuring Cl^- concentrations in sweat collected after iontophoresis of cholinergic agonists.

CLINICAL FEATURES

Most patients with CF present with signs and symptoms of the disease in childhood. Approximately 20% of patients present within the first 24 h of life with gastrointestinal obstruction, termed *meconium ileus*. Other common presentations within the first year or two of life include respiratory tract symptoms, most prominently cough and/or recurrent pulmonary infiltrates, and failure to thrive. A significant proportion of patients (~5%), however, are diagnosed after age 18.

Respiratory Tract

Upper respiratory tract disease is almost universal in patients with CF. Chronic sinusitis is common in childhood, and the incidence of nasal polyps, which often requires treatment with topical steroids and/or surgery, approaches 25%.

In the lower respiratory tract, the first symptom of CF is cough. With time, the cough becomes persistent and produces viscous, purulent, often greenish-colored sputum. There are protracted periods of clinical stability interrupted by "pulmonary exacerbations," often triggered by viral infections, and defined by increased cough, weight loss, low-grade fever, increased sputum volume, and decrements in pulmonary function. Over the course of years, the exacerbation frequency increases and the recovery of lost lung function becomes incomplete, leading to respiratory failure.

CF patients exhibit a characteristic sputum microbiology. *Haemophilus influenzae* and *S. aureus* are often the first organisms recovered from lung secretions in newly diagnosed CF patients. *P. aeruginosa*, often mucoid and antibiotic-resistant, is typically cultured from lower respiratory tract secretions thereafter. *Burkholderia cepacia* is also recovered from CF sputum and is pathogenic. Patient-to-patient spread of certain strains of these organisms mandates strict infection control in the hospital. Other gram-negative rods recovered from CF sputum include *Alcaligenes xylosoxidans*, *B. gladioli*; and, occasionally, *Proteus*, *Escherichia coli*, and *Klebsiella*. Up to 50% of CF patients have *Aspergillus fumigatus* in their sputum, and up to 10% of these patients exhibit the syndrome of allergic bronchopulmonary aspergillosis. *Mycobacterium tuberculosis* is rare in patients with CF. However, 10

–20% of adult patients with CF have sputum cultures positive for nontuberculous mycobacteria, and in some patients, these microorganisms are associated with disease.

The first lung-function abnormalities in CF children, increased ratios of residual volume to total lung capacity, suggest that small-airways disease is the first functional lung abnormality in CF. As disease progresses, both reversible and irreversible changes in forced vital capacity (FVC) and forced expiratory volume in 1 s (FEV_1) develop. The reversible component reflects the accumulation of intraluminal secretions and/or airway reactivity, which occurs in 40–60% of patients with CF. The irreversible component reflects chronic destruction of the airway wall and bronchiolitis.

The earliest chest x-ray change in CF lungs is hyperinflation, reflecting small-airways obstruction. Later, signs of luminal mucus impaction, bronchial cuffing, and finally bronchiectasis, e.g., ring shadows, are noted. For reasons that remain speculative, the right upper lobe displays the earliest and most severe changes.

CF pulmonary disease is associated with many intermittent complications. Pneumothorax is common (>10% of patients). The production of small amounts of blood in sputum is common in CF patients with advanced pulmonary disease. Massive hemoptysis is life-threatening. With advanced lung disease, clubbing of digits appears in virtually all patients with CF. As late events, respiratory failure and cor pulmonale are prominent features of CF.

Gastrointestinal Tract

Meconium ileus in infants presents with abdominal distention, failure to pass stool, and emesis. The abdominal flat plate can be diagnostic, with small intestinal air-fluid levels, a granular appearance representing meconium, and a small colon. In children and young adults, a syndrome termed *distal intestinal obstruction syndrome* (DIOS) occurs, which presents with right lower quadrant pain, loss of appetite, occasionally emesis, and often a palpable mass. DIOS can be confused with appendicitis, whose frequency is not increased in CF patients.

Exocrine pancreatic insufficiency occurs in >90% of patients with CF. Insufficient pancreatic enzyme secretion yields protein and fat malabsorption, with frequent, bulky, foul-smelling stools. Signs and symptoms of malabsorption of fat-soluble vitamins, including vitamins E and K, are also noted. Pancreatic beta cells are spared early, but function decreases with age. This effect, plus inflammation-induced insulin resistance, causes hyperglycemia and a requirement for insulin in >29% of older patients with CF (>35 years).

Genitourinary System

Late onset of puberty is common in both males and females with CF. More than 95% of male patients with CF are azoospermic, reflecting obliteration of the vas deferens due to defective liquid secretion. Some 20% of CF women are infertile due to effects of chronic lung disease on the menstrual cycle and thick, tenacious cervical mucus that blocks sperm migration. Most pregnancies produce viable infants, and CF women breast-feed infants normally.

DIAGNOSIS

The diagnosis of CF rests on the combination of clinical criteria and abnormal CFTR function as documented by sweat tests, nasal PD measurements, and *CFTR* mutation analysis. Elevated sweat Cl^- values are nearly pathognomonic for CF. The sweat concentration values for Cl^- (and Na^+) vary with age, but, typically, a Cl^- concentration of >70 meq/L in adults discriminates between CF and other lung diseases. DNA analysis of the most common mutations identify CF mutations in >90% of affected patients. The nasal PD measurement can document CFTR dysfunction if the sweat Cl^- test is normal or borderline and two CF mutations are not identified. DNA analysis is performed routinely in patients with CF, because pancreatic genotype-phenotype relationships have been identified and mutation class–specific treatments are being developed.

Between 1 and 2% of patients with the clinical syndrome of CF have normal sweat Cl^- values. In most of these patients, the nasal transepithelial PD is raised into the diagnostic range for CF, and sweat acini do not secrete in response to injected ß-adrenergic agonists. A single mutation of the *CFTR* gene, 3849 + 10 kb C→T, is associated with most CF patients with normal sweat Cl^- values.

Treatment: Cystic Fibrosis

The major objectives of therapy for CF are to promote clearance of secretions and control infection in the lung, provide adequate nutrition, and prevent intestinal obstruction. Ultimately, therapies that restore the processing of misfolded mutant *CFTR* or gene therapy may be the treatments of choice.

LUNG DISEASE

More than 95% of CF patients die of complications from lung infection. Theoretically, promoting clearance of adherent mucus should both treat and prevent progression of CF lung disease, whereas antibiotics principally reduce the bacterial burden in the CF lung.

The time-honored techniques for clearing pulmonary secretions are exercise, flutter valves, and chest percussion. Regular use of these maneuvers is effective in preserving lung function. Inhaled hypertonic saline (7%) has demonstrated efficacy in restoring mucus clearance and pulmonary function in short-term studies and in reducing acute exacerbations in a long-term (one-year) study. Hypertonic saline is safe but produces bronchoconstriction in some patients, which can be prevented with coadministered bronchodilators. Inhaled hypertonic saline is becoming a standard of care for all CF patients.

Pharmacologic agents for increasing mucus clearance are in use and in development. An important adjunct to secretion clearance can be recombinant human DNAse, which degrades DNA in CF sputum, increases airflow during short-term administration, and increases the time between pulmonary exacerbations. Most patients receive a therapeutic trial of DNAse for several months to test for efficacy. Clinical trials of experimental drugs aimed at restoring salt and water content of secretions are under way, but these drugs are not yet available for clinical use.

Antibiotics are used to treat lung infection, and their selection is guided by sputum culture results. However, because routine hospital microbiologic cultures are performed under conditions that do not mimic conditions in the CF lung, (e.g., hypoxemia) clinical efficacy often does not correlate with sensitivity testing. Because of increased total-body clearance and volume of distribution of antibiotics in CF patients, the required doses are higher for patients with CF.

Early intervention with antibiotics in infants with infection may eradicate *P. aeruginosa* for extended periods. In older patients with established infection, suppression of bacterial growth is the therapeutic goal. Azithromycin (250 mg/d or 500 mg three times weekly) is often used chronically, although it is unclear whether its actions are antimicrobial or anti-inflammatory. Inhaled aminoglycosides, (e.g., tobramycin 300 mg bid) are also used. "Mild exacerbations," as defined by increased cough and mucus production, are treated with additional oral antibiotics. Oral agents used to treat *Staphylococcus* include a semisynthetic penicillin or a cephalosporin. Oral ciprofloxacin may reduce pseudomonal bacterial counts and control symptoms, but its clinical utility is limited by emergence of resistant organisms. Accordingly, it is often used with an inhaled antibiotic, either tobramycin or colistin (75 mg bid). More severe exacerbations require intravenous antibiotics. Intravenous therapy is given both in hospital and outpatient settings. Usually, two drugs with different mechanisms of action (e.g., a cephalosporin and an aminoglycoside) are used to treat *P. aeruginosa* to minimize emergence of resistant organisms. Drug dosage should be monitored so that levels for gentamicin or tobramycin peak at ranges of ~10 μg/mL and exhibit troughs of <2 μg/mL. Antibiotics directed at *Staphylococcus* and/or *H. influenzae* are added, depending on culture results.

Inhaled β-adrenergic agonists can be useful to control airways constriction, but long-term benefit has not been shown. Oral glucocorticoids may reduce airway inflammation, but their long-term use is limited by adverse side effects; however, they may be useful for treating allergic bronchopulmonary aspergillosis.

The chronic damage to airway walls reflects in part the destructive activities of proteolytic enzymes generated, in part, by inflammatory cells. Specific antiprotease therapies are not available. However, a subset of adolescents with CF may benefit from long-term, high-dose nonsteroidal (ibuprofen) therapy.

Pulmonary complications often require acute interventions. Atelectasis requires treatment with inhaled hypertonic saline, chest physiotherapy, and antibiotics. Pneumothoraxes involving ≤10% of the lung can be observed, but chest tubes are required to expand a collapsed, diseased lung. Small-volume hemoptysis typically requires treatment of lung infection and assessment of coagulation and vitamin K status. For massive hemoptysis, bronchial artery embolization should be performed. The most ominous complications of CF are respiratory failure and cor pulmonale. The most effective conventional therapy for these conditions is vigorous medical management of the lung disease and O_2 supplementation.

Ultimately, the only effective treatment for respiratory failure in CF is lung transplantation (Chap. 266). The 2-year survival for lung transplantation exceeds 60%, and transplant-patient deaths result principally from obliterative bronchiolitis.

GASTROINTESTINAL DISEASE

Maintenance of adequate nutrition is critical for the health of the CF patient. Most (>90%) CF patients require pancreatic enzyme replacement. Capsules generally contain between 4000 and 20,000 units of lipase. The dose of enzymes (typically no more than 2500 units/kg per meal, to avoid risk of fibrosing colonopathy) is adjusted on the basis of weight, abdominal symptomatology, and stool character. Replacement of fat-soluble vitamins, particularly vitamins E and K, is usually required. Hyperglycemia most often becomes manifest in the adult and typically requires insulin treatment.

For treatment of acute distal intestinal obstruction, megalodiatrizoate or other hypertonic radiocontrast materials delivered by enema to the terminal ileum are used. For control of symptoms, adjustment of pancreatic enzymes and solutions containing osmotically active agents, (e.g., propyleneglycol) is recommended. Persistent symptoms may indicate a gastrointestinal malignancy, which is increased in incidence in patients with CF.

Cholestatic liver disease occurs in about 8% of CF patients. Treatment with urodeoxycholic acid is often initiated, but has not been shown to influence the course of hepatic disease. End-stage liver disease occurs in about 5% of CF patients and is treated by transplantation.

OTHER ORGAN COMPLICATIONS

Dehydration due to heat-induced salt loss occurs more readily in CF patients. CF patients also have an increased incidence of osteoarthropathy, renal stones, and osteoporosis, particularly following transplant.

PSYCHOSOCIAL FACTORS

CF imposes a tremendous burden on patients, and depression is common. Health insurance, career options, family planning, and life expectancy become major issues. Thus, assisting patients with the psychosocial adjustments required by CF is critical.

FURTHER READINGS

Boucher RC: Airway surface dehydration in cystic fibrosis: pathogenesis and therapy. Ann Rev Med 58:157, 2007[PMID: 17217330] [Full Text]

Boyle MP: Cystic fibrosis: Year in review. Curr Opin Pulm Med 16:583, 2010[PMID: 20829695] [Full Text]

Davies JC et al: Cystic fibrosis. Br Med J 335:1255, 2007[PMID: 18079549] [Full Text]

Davis PB: Cystic fibrosis since 1938. Am J Respir Crit Care Med 173:475, 2006[PMID: 16126935] [Full Text]

Elizur A et al: Airway inflammation in cystic fibrosis. Chest 133:489, 2008[PMID: 18252915] [Full Text]

Elkins MR et al: A controlled trial of long-term inhaled hypertonic saline in patients with cystic fibrosis. N Engl J Med 354:229, 2006[PMID: 16421364] [Full Text]

Ferkol T et al: Cystic fibrosis pulmonary exacerbations. J Pediatr 148:259, 2006[PMID: 16492439] [Full Text]

Flume PA, Stenbit A: Making the diagnosis of cystic fibrosis. Am J Med Sci 335:51, 2008[PMID: 18195584] [Full Text]

Guggino WB, Stanton BA: New insights into cystic fibrosis: molecular switches that regulate CFTR. Nat Rev Mol Cell Biol 7:426, 2006[PMID: 16723978] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHRONIC OBSTRUCTIVE PULMONARY DISEASE: INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is defined as a disease state characterized by airflow limitation that is not fully reversible (*http://www.goldcopd.com/*). COPD includes *emphysema*, an anatomically defined condition characterized by destruction and enlargement of the lung alveoli; *chronic bronchitis*, a clinically defined condition with chronic cough and phlegm; and *small airways disease*, a condition in which small bronchioles are narrowed. COPD is present only if chronic airflow obstruction occurs; chronic bronchitis *without* chronic airflow obstruction is *not* included within COPD.

COPD is the fourth leading cause of death and affects >10 million persons in the United States. COPD is also a disease of increasing public health importance around the world. Estimates suggest that COPD will rise from the sixth to the third most common cause of death worldwide by 2020.

RISK FACTORS

Cigarette Smoking

By 1964, the Advisory Committee to the Surgeon General of the United States had concluded that cigarette smoking was a major risk factor for mortality from chronic bronchitis and emphysema. Subsequent longitudinal studies have shown accelerated decline in the volume of air exhaled within the first second of the forced expiratory maneuver (FEV_1) in a dose-response relationship to the intensity of cigarette smoking, which is typically expressed as pack-years (average number of packs of cigarettes smoked per day multiplied by the total number of years of smoking). This dose-response relationship between reduced pulmonary function and cigarette smoking intensity accounts for the higher prevalence rates for COPD with increasing age. The historically higher rate of smoking among males is the likely explanation for the higher prevalence of COPD among males; however, the prevalence of COPD among females is increasing as the gender gap in smoking rates has diminished in the past 50 years.

Although the causal relationship between cigarette smoking and the development of COPD has been absolutely proved, there is considerable variability in the response to smoking. Although pack-years of cigarette smoking is the most highly significant predictor of FEV_1 (**Fig. 260-1**), only 15% of the variability in FEV_1 is explained by pack-years. This finding suggests that additional environmental and/or genetic factors contribute to the impact of smoking on the development of airflow obstruction.

Figure 260-1

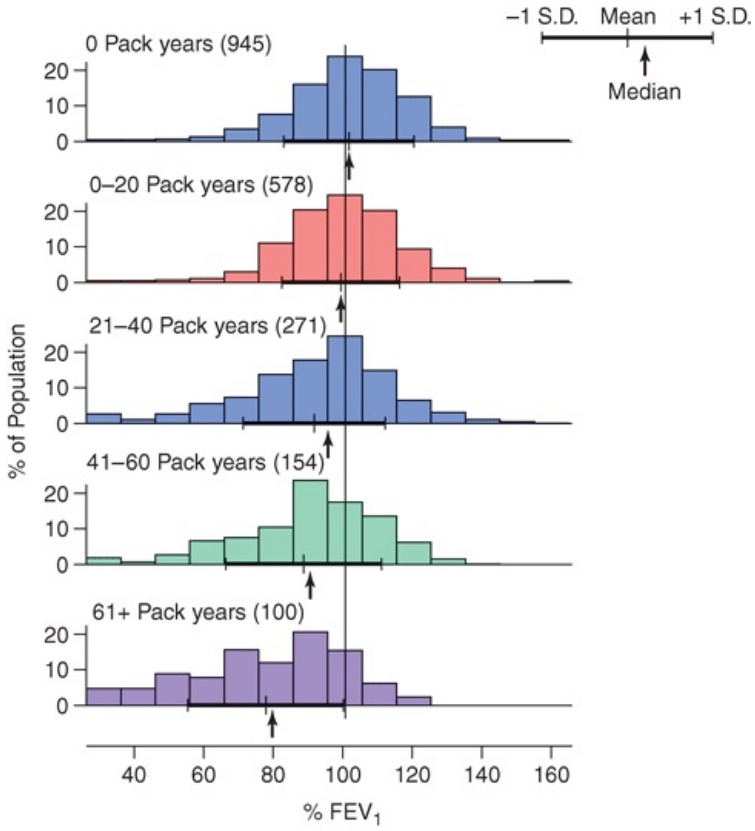

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Distributions of forced expiratory volume in 1 s (FEV_1) values in a general population sample, stratified by pack-years of smoking. Means, medians, and ±1 standard deviation of percent predicted FEV_1 are shown for each smoking group. Although a dose-response relationship between smoking intensity and FEV_1 was found, marked variability in pulmonary function was observed among subjects with similar smoking histories. (*From B Burrows et al: Am Rev Respir Dis 115:95, 1977; with permission.*)

Although cigar and pipe smoking may also be associated with the development of COPD, the evidence supporting such associations is less compelling, likely related to the lower dose of inhaled tobacco by-products during cigar and pipe smoking.

Airway Responsiveness and COPD

A tendency for increased bronchoconstriction in response to a variety of exogenous stimuli, including methacholine and histamine, is one of the defining features of asthma (Chap. 254). However, many patients with COPD also share this feature of airway hyperresponsiveness. The considerable overlap between persons with asthma and those with COPD in airway responsiveness, airflow obstruction, and pulmonary symptoms led to the formulation of the Dutch hypothesis. This suggests that asthma, chronic bronchitis, and emphysema are variations of the same basic disease, which is modulated by environmental and genetic factors to produce these pathologically distinct entities. The alternative British hypothesis contends that asthma and COPD are fundamentally different diseases: Asthma is viewed as largely an allergic phenomenon, while COPD results from smoking-related inflammation and damage. Determination of the validity of the Dutch hypothesis vs. the British hypothesis awaits identification of the genetic predisposing factors for asthma and/or COPD, as well as the interactions between these postulated genetic factors and environmental risk factors. Of note, several genes related to the proteinase-antiproteinase hypothesis have been implicated as genetic determinants for both COPD and asthma, including *ADAM33* and macrophage elastase (*MMP12*) as described below.

Longitudinal studies that compared airway responsiveness at the beginning of the study to subsequent decline in pulmonary function have demonstrated that increased airway responsiveness is clearly a significant predictor of subsequent decline in pulmonary function. Thus, airway hyperresponsiveness is a risk factor for COPD.

Respiratory Infections

The impact of adult respiratory infections on decline in pulmonary function is controversial, but significant long-term reductions in pulmonary function are not typically seen following an episode of bronchitis or pneumonia. The impact of the effects of childhood respiratory illnesses on the subsequent development of COPD has been difficult to assess due to a lack of adequate longitudinal

data. Thus, although respiratory infections are important causes of exacerbations of COPD, the association of both adult and childhood respiratory infections to the development and progression of COPD remains to be proven.

Occupational Exposures

Increased respiratory symptoms and airflow obstruction have been suggested to result from general exposure to dust and fumes at work. Several specific occupational exposures, including coal mining, gold mining, and cotton textile dust, have been suggested as risk factors for chronic airflow obstruction. Although nonsmokers in these occupations developed some reductions in FEV_1, the importance of dust exposure as a risk factor for COPD, independent of cigarette smoking, is not certain for most of these exposures. However, a recent study found that coal mine dust exposure was a significant risk factor for emphysema in both smokers and nonsmokers. In most cases, the magnitude of these occupational exposures on COPD risk is likely substantially less important than the effect of cigarette smoking.

Ambient Air Pollution

Some investigators have reported increased respiratory symptoms in those living in urban compared to rural areas, which may relate to increased pollution in the urban settings. However, the relationship of air pollution to chronic airflow obstruction remains unproved. Prolonged exposure to smoke produced by biomass combustion—a common mode of cooking in some countries—also appears to be a significant risk factor for COPD among women in those countries. However, in most populations, ambient air pollution is a much less important risk factor for COPD than cigarette smoking.

Passive, or Second-Hand, Smoking Exposure

Exposure of children to maternal smoking results in significantly reduced lung growth. In utero tobacco smoke exposure also contributes to significant reductions in postnatal pulmonary function. Although passive smoke exposure has been associated with reductions in pulmonary function, the importance of this risk factor in the development of the severe pulmonary function reductions in COPD remains uncertain.

Genetic Considerations

Although cigarette smoking is the major environmental risk factor for the development of COPD, the development of airflow obstruction in smokers is highly variable. Severe α_1 antitrypsin (α_1AT) deficiency is a proven genetic risk factor for COPD; there is increasing evidence that other genetic determinants also exist.

α_1 ANTITRYPSIN DEFICIENCY

Many variants of the protease inhibitor (PI or SERPINA1) locus that encodes α_1AT have been described. The common M allele is associated with normal α_1AT levels. The S allele, associated with slightly reduced α_1AT levels, and the Z allele, associated with markedly reduced α_1AT levels, also occur with frequencies >1% in most white populations. Rare individuals inherit null alleles, which lead to the absence of any α_1AT production through a heterogeneous collection of mutations. Individuals with two Z alleles or one Z and one null allele are referred to as Pi^Z, which is the most common form of severe α_1AT deficiency.

Although only 1–2% of COPD patients are found to have severe α_1AT deficiency as a contributing cause of COPD, these patients demonstrate that genetic factors can have a profound influence on the susceptibility for developing COPD. Pi^Z individuals often develop early-onset COPD, but the ascertainment bias in the published series of Pi^Z individuals—which have usually included many Pi^Z subjects who were tested for α_1AT deficiency because they had COPD—means that the fraction of Pi^Z individuals who will develop COPD and the age-of-onset distribution for the development of COPD in Pi^Z subjects remain unknown. Approximately 1 in 3000 individuals in the United States inherits severe α_1AT deficiency, but only a small minority of these individuals has been recognized. The clinical laboratory test used most frequently to screen for α_1AT deficiency is measurement of the immunologic level of α_1AT in serum (see "Laboratory Findings").

A significant percentage of the variability in pulmonary function among Pi^Z individuals is explained by cigarette smoking; cigarette smokers with severe α_1AT deficiency are more likely to develop COPD at early ages. However, the development of COPD in Pi^Z subjects, even among current or ex-smokers, is not absolute. Among Pi^Z nonsmokers, impressive variability has been noted in the development of airflow obstruction. Asthma and male gender also appear to increase the risk of COPD in Pi^Z subjects. Other genetic and/or environmental factors likely contribute to this variability.

Specific treatment in the form of α_1AT augmentation therapy is available for severe α_1AT deficiency as a weekly IV infusion (see "Treatment," below).

The risk of lung disease in heterozygous Pi^{MZ} individuals, who have intermediate serum levels of α_1AT (~60% of Pi^{MM} levels), is controversial. Although previous general population surveys have not typically shown increased rates of airflow obstruction in Pi^{MZ} compared to Pi^{MM} individuals, case-control studies that compared COPD patients to control subjects have usually found an excess of Pi^{MZ} genotypes in the COPD patient group. Several recent large population studies have suggested that Pi^{MZ} subjects are at slightly increased risk for the development of airflow obstruction, but it remains unclear if all Pi^{MZ} subjects are at slightly increased risk for COPD or if a subset of Pi^{MZ} subjects are at substantially increased risk for COPD due to other genetic or environmental factors.

OTHER GENETIC RISK FACTORS

Studies of pulmonary function measurements performed in general population samples have suggested that genetic factors other than PI type influence variation in pulmonary function. Familial aggregation of airflow obstruction within families of COPD patients has also been demonstrated.

Association studies have compared the distribution of variants in candidate genes hypothesized to be involved in the development of COPD in COPD patients and control subjects. However, the results have been quite inconsistent, often due to underpowered studies. However, a recent association study comprising 8300 patients and 7 separate cohorts found that a minor allele SNP of *MMP12* (rs2276109) associated with decreased MMP-12 expression has a positive effect on lung function in children with asthma and in adult smokers. Recent genome-wide association studies have identified several COPD loci, including a region near the hedgehog interacting protein (*HHIP*) gene on chromosome 4 and a cluster of genes on chromosome 15 (including components of the nicotinic acetylcholine receptor) that likely contain COPD susceptibility determinants, but the specific genetic determinants in those regions have yet to be definitively identified.

NATURAL HISTORY

The effects of cigarette smoking on pulmonary function appear to depend on the intensity of smoking exposure, the timing of smoking exposure during growth, and the baseline lung function of the individual; other environmental factors may have similar effects. Most individuals follow a steady trajectory of increasing pulmonary function with growth during childhood and adolescence, followed by a gradual decline with aging. Individuals appear to track in their quartile of pulmonary function based upon environmental and genetic factors that put them on different tracks. The risk of eventual mortality from COPD is closely associated with reduced levels of FEV_1. A graphic depiction of the natural history of COPD is shown as a function of the influences on tracking curves of FEV_1 in **Fig. 260-2**. Death or disability from COPD can result from a normal rate of decline after a reduced growth phase (curve *C*), an early initiation of pulmonary function decline after normal growth (curve *B*), or an accelerated decline after normal growth (curve *D*). The rate of decline in pulmonary function can be modified by changing environmental exposures (i.e., quitting smoking), with smoking cessation at an earlier age providing a more beneficial effect than smoking cessation after marked reductions in pulmonary function have already developed. Genetic factors likely contribute to the level of pulmonary function achieved during growth and to the rate of decline in response to smoking and potentially to other environmental factors as well.

Figure 260-2

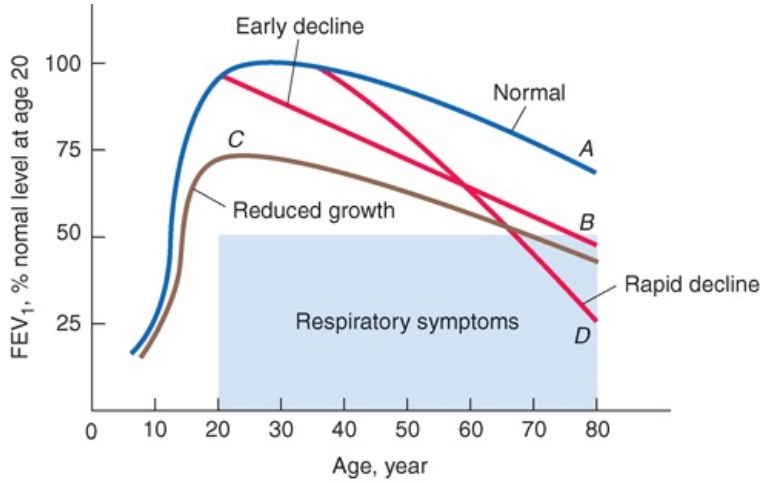

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypothetical tracking curves of FEV_1 for individuals throughout their life spans. The normal pattern of growth and decline with age is shown by curve *A*. Significantly reduced FEV_1 (<65% of predicted value at age 20) can develop from a normal rate of decline after a reduced pulmonary function growth phase (curve *C*), early initiation of pulmonary function decline after normal growth (curve *B*), or accelerated decline after normal growth (curve *D*). (*From B Rijcken: Doctoral dissertation, p 133, University of Groningen, 1991; with permission.*)

PATHOPHYSIOLOGY

Persistent reduction in forced expiratory flow rates is the most typical finding in COPD. Increases in the residual volume and the residual volume/total lung capacity ratio, nonuniform distribution of ventilation, and ventilation-perfusion mismatching also occur.

Airflow Obstruction

Airflow limitation, also known as airflow obstruction, is typically determined by spirometry, which involves forced expiratory maneuvers after the subject has inhaled to total lung capacity. Key parameters obtained from spirometry include FEV_1 and the total volume of air exhaled during the entire spirometric maneuver [forced vital capacity (FVC)]. Patients with airflow obstruction related to COPD have a chronically reduced ratio of FEV_1/FVC. In contrast to asthma, the reduced FEV_1 in COPD seldom shows large responses to inhaled bronchodilators, although improvements up to 15% are common. Asthma patients can also develop chronic (not fully reversible) airflow obstruction.

Airflow during forced exhalation is the result of the balance between the elastic recoil of the lungs promoting flow and the resistance of the airways limiting flow. In normal lungs, as well as in lungs affected by COPD, maximal expiratory flow diminishes as the lungs empty because the lung parenchyma provides progressively less elastic recoil and because the cross-sectional area of the airways falls, raising the resistance to airflow. The decrease in flow coincident with decreased lung volume is readily apparent on the expiratory limb of a flow-volume curve. In the early stages of COPD, the abnormality in airflow is only evident at lung volumes at or below the functional residual capacity (closer to residual volume), appearing as a scooped-out lower part of the descending limb of the flow-volume curve. In more advanced disease the entire curve has decreased expiratory flow compared to normal.

Hyperinflation

Lung volumes are also routinely assessed in pulmonary function testing. In COPD there is often "air trapping" (increased residual volume and increased ratio of residual volume to total lung capacity) and progressive hyperinflation (increased total lung capacity) late in the disease. Hyperinflation of the thorax during tidal breathing preserves maximum expiratory airflow, because as lung volume increases, elastic recoil pressure increases, and airways enlarge so that airway resistance decreases.

Despite compensating for airway obstruction, hyperinflation can push the diaphragm into a flattened position with a number of adverse effects. First, by decreasing the zone of apposition between the diaphragm and the abdominal wall, positive abdominal pressure during inspiration is not applied as effectively to the chest wall, hindering rib cage movement and impairing inspiration. Second, because the muscle fibers of the flattened diaphragm are shorter than those of a more normally curved diaphragm, they are less capable of generating inspiratory pressures than normal. Third, the flattened diaphragm (with increased radius of curvature, r) must generate greater tension (t) to develop the transpulmonary pressure (p) required to produce tidal breathing. This follows from Laplace's law, $p = 2t/r$. Also, because the thoracic cage is distended beyond its normal resting volume, during tidal

breathing the inspiratory muscles must do work to overcome the resistance of the thoracic cage to further inflation instead of gaining the normal assistance from the chest wall recoiling outward toward its resting volume.

Gas Exchange

Although there is considerable variability in the relationships between the FEV_1 and other physiologic abnormalities in COPD, certain generalizations may be made. The Pa_{O_2} usually remains near normal until the FEV_1 is decreased to ~50% of predicted, and even much lower FEV_1 values can be associated with a normal Pa_{O_2}, at least at rest. An elevation of arterial level of carbon dioxide (Pa_{CO_2}) is not expected until the FEV_1 is <25% of predicted and even then may not occur. Pulmonary hypertension severe enough to cause cor pulmonale and right ventricular failure due to COPD typically occurs in individuals who have marked decreases in FEV_1 (<25% of predicted) and chronic hypoxemia (Pa_{O_2} <55 mmHg); however, recent evidence suggests that some patients will develop significant pulmonary hypertension independent of COPD severity (Chap. 250).

Nonuniform ventilation and ventilation-perfusion mismatching are characteristic of COPD, reflecting the heterogeneous nature of the disease process within the airways and lung parenchyma. Physiologic studies are consistent with multiple parenchymal compartments having different rates of ventilation due to regional differences in compliance and airway resistance. Ventilation-perfusion mismatching accounts for essentially all of the reduction in Pa_{O_2} that occurs in COPD; shunting is minimal. This finding explains the effectiveness of modest elevations of inspired oxygen in treating hypoxemia due to COPD and therefore the need to consider problems other than COPD when hypoxemia is difficult to correct with modest levels of supplemental oxygen in the patient with COPD.

PATHOLOGY

Cigarette smoke exposure may affect the large airways, small airways (≲2 mm diameter), and alveoli. Changes in large airways cause cough and sputum, while changes in small airways and alveoli are responsible for physiologic alterations. Emphysema and small airway pathology are both present in most persons with COPD; however, they do not appear to be mechanistically related to each other, and their relative contributions to obstruction vary from one person to another.

Large Airway

Cigarette smoking often results in mucous gland enlargement and goblet cell hyperplasia leading to cough and mucus production that define chronic bronchitis, but these abnormalities are not related to airflow limitation. Goblet cells not only increase in number but in extent through the bronchial tree. Bronchi also undergo squamous metaplasia, predisposing to carcinogenesis and disrupting mucociliary clearance. Although not as prominent as in asthma, patients may have smooth-muscle hypertrophy and bronchial hyperreactivity leading to airflow limitation. Neutrophil influx has been associated with purulent sputum of upper respiratory tract infections. Independent of its proteolytic activity, neutrophil elastase is among the most potent secretagogues identified.

Small Airways

The major site of increased resistance in most individuals with COPD is in airways ≲2 mm diameter. Characteristic cellular changes include goblet cell metaplasia, with these mucus-secreting cells replacing surfactant-secreting Clara cells. Infiltration of mononuclear phagocytes is also prominent. Smooth-muscle hypertrophy may also be present. These abnormalities may cause luminal narrowing by fibrosis, excess mucus, edema, and cellular infiltration. Reduced surfactant may increase surface tension at the air-tissue interface, predisposing to airway narrowing or collapse. Respiratory bronchiolitis with mononuclear inflammatory cells collecting in distal airway tissues may cause proteolytic destruction of elastic fibers in the respiratory bronchioles and alveolar ducts where the fibers are concentrated as rings around alveolar entrances.

Because small airway patency is maintained by the surrounding lung parenchyma that provides radial traction on bronchioles at points of attachment to alveolar septa, loss of bronchiolar attachments as a result of extracellular matrix destruction may cause airway distortion and narrowing in COPD.

Lung Parenchyma

Emphysema is characterized by destruction of gas-exchanging air spaces, i.e., the respiratory bronchioles, alveolar ducts, and alveoli. Their walls become perforated and later obliterated with coalescence of small distinct air spaces into abnormal and much larger air spaces. Macrophages accumulate in respiratory bronchioles of essentially all young smokers. Bronchoalveolar lavage fluid from such individuals contains roughly five times as many macrophages as lavage from nonsmokers. In smokers' lavage fluid, macrophages comprise >95% of the total cell count, and neutrophils, nearly absent in nonsmokers' lavage, account for 1–2% of the cells. T lymphocytes, particularly CD8+ cells, are also increased in the alveolar space of smokers.

Emphysema is classified into distinct pathologic types, the most important being centriacinar and panacinar. *Centriacinar emphysema*, the type most frequently associated with cigarette smoking, is characterized by enlarged air spaces found (initially) in association with respiratory bronchioles. Centriacinar emphysema is usually most prominent in the upper lobes and superior segments of lower lobes and is often quite focal. *Panacinar emphysema* refers to abnormally large air spaces evenly distributed within and across acinar units. Panacinar emphysema is usually observed in patients with α_1AT deficiency, which has a predilection for the lower lobes. Distinctions between centriacinar and panacinar emphysema are interesting and may ultimately be shown to have different mechanisms of pathogenesis. However, garden-variety smoking-related emphysema is usually mixed, particularly in advanced cases, and these pathologic classifications are not helpful in the care of patients with COPD.

PATHOGENESIS

Airflow limitation, the major physiologic change in COPD, can result from both small airway obstruction and emphysema, as discussed above. Pathologic findings that can contribute to small airway obstruction are described above, but their relative importance is unknown. Fibrosis surrounding the small airways appears to be a significant contributor. Mechanisms leading to collagen accumulation around the airways in the face of increased collagenase activity remain an enigma. Although seemingly counterintuitive, there are several potential mechanisms whereby a proteinase can predispose to fibrosis, including proteolytic activation of transforming growth factor β (TGF-β). Largely due to greater similarity of animal air spaces than airways to humans, we know much more about mechanisms involved in emphysema than small airway obstruction.

The dominant paradigm of the pathogenesis of emphysema comprises four interrelated events (**Fig. 260-3**): (1) Chronic exposure to cigarette smoke may lead to inflammatory cell recruitment within the terminal air spaces of the lung. (2) These inflammatory cells release elastolytic proteinases that damage the extracellular matrix of the lung. (3) Structural cell death results from oxidant stress and loss of matrix-cell attachment. (4) Ineffective repair of elastin and other extracellular matrix components result in air space enlargement that defines pulmonary emphysema.

Figure 260-3

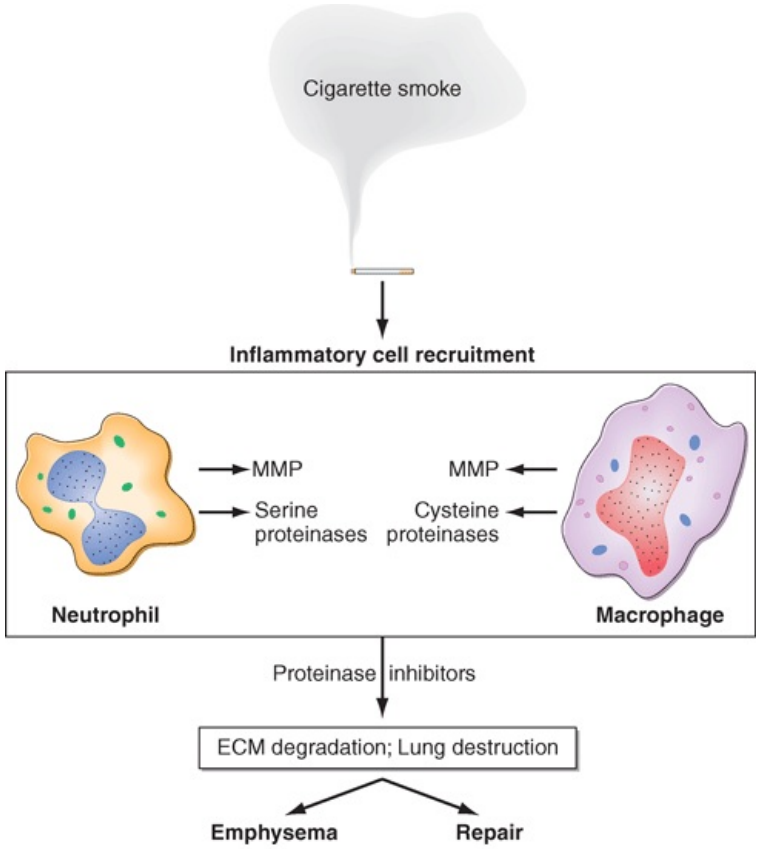

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis of emphysema. Upon long-term exposure to cigarette smoke, inflammatory cells are recruited to the lung; they release proteinases in excess of inhibitors, and if repair is abnormal, this leads to air space destruction and enlargement or emphysema. ECM, extracellular matrix; MMP,

leads to air space destruction and enlargement of emphysematic ECM; extracellular matrix; MMP, matrix metalloproteinase.

The Elastase:Antielastase Hypothesis

Elastin, the principal component of elastic fibers, is a highly stable component of the extracellular matrix that is critical to the integrity of the lung. The elastase:antielastase hypothesis proposed in the mid-1960s states that the balance of elastin-degrading enzymes and their inhibitors determines the susceptibility of the lung to destruction resulting in air space enlargement. This hypothesis was based on the clinical observation that patients with genetic deficiency in α_1AT, the inhibitor of the serine proteinase neutrophil elastase, were at increased risk of emphysema, and that instillation of elastases, including neutrophil elastase, to experimental animals results in emphysema. The elastase:antielastase hypothesis remains a prevailing mechanism for the development of emphysema. However, a complex network of immune and inflammatory cells and additional proteinases that contribute to emphysema have subsequently been identified.

INFLAMMATION AND EXTRACELLULAR MATRIX PROTEOLYSIS

Macrophages patrol the lower air space under normal conditions. Upon exposure to oxidants from cigarette smoke, macrophages become activated, producing proteinases and chemokines that attract other inflammatory cells. One mechanism of macrophage activation occurs via oxidant-induced inactivation of histone deacetylase-2, shifting the balance toward acetylated or loose chromatin, exposing nuclear factor κB sites and resulting in transcription of matrix metalloproteinases, proinflammatory cytokines such as interleukin 8 (IL-8), and tumor necrosis factor α (TNF-α); this leads to neutrophil recruitment. CD8+ T cells are also recruited in response to cigarette smoke and release interferon inducible protein-10 (IP-10, CXCL-7) that in turn leads to macrophage production of macrophage elastase [matrix metalloproteinase-12 (MMP-12)]. Matrix metalloproteinases and serine proteinases, most notably neutrophil elastase, work together by degrading the inhibitor of the other, leading to lung destruction. Proteolytic cleavage products of elastin also serve as a macrophage chemokine, fueling this destructive positive feedback loop.

Autoimmune mechanisms have recently been identified in COPD and may promote the progression of disease. Increased B cells and lymphoid follicles are present in patients, particularly those with advanced disease. Antibodies have been found against elastin fragments, as well; IgG autoantibodies with avidity for pulmonary epithelium and the potential to mediate cytotoxicity have been detected.

Concomitant cigarette smoke–induced loss of cilia in the airway epithelium and impaired macrophage phagocytosis predispose to bacterial infection with neutrophilia. In end-stage lung disease, long after smoking cessation there remains an exuberant inflammatory response, suggesting that mechanisms of cigarette smoke–induced inflammation that initiate the disease differ from mechanisms that sustain inflammation after smoking cessation.

CELL DEATH

Air space enlargement with loss of alveolar units obviously requires disappearance of both extracellular matrix and cells. Cell death can occur from increased oxidant stress both directly from cigarette smoke and from inflammation. Animal models have used endothelial and epithelial cell death as a means to generate transient air space enlargement. Uptake of apoptotic cells by macrophages results in production of growth factors and dampens inflammation, promoting lung repair. Cigarette smoke impairs macrophage uptake of apoptotic cells, limiting repair.

INEFFECTIVE REPAIR

The ability of the adult lung to repair damaged alveoli appears limited. It is unlikely that the process of septation that is responsible for alveologenesis during lung development can be reinitiated. The capacity of stem cells to repopulate the lung is under active investigation. It appears difficult for an adult human to completely restore an appropriate extracellular matrix, particularly functional elastic fibers.

CLINICAL PRESENTATION

History

The three most common symptoms in COPD are cough, sputum production, and exertional dyspnea. Many patients have such symptoms for months or years before seeking medical attention. Although the development of airflow obstruction is a gradual process, many patients date the onset of their disease to an acute illness or exacerbation. A careful history, however, usually reveals the presence of symptoms prior to the acute exacerbation. The development of exertional dyspnea, often described as increased effort to breathe, heaviness, air hunger, or gasping, can be insidious. It is best elicited by a careful history focused on typical physical activities and how the patient's ability to perform them has changed. Activities involving significant arm work, particularly at or above shoulder level, are particularly difficult for patients with COPD. Conversely, activities that allow the patient to brace the arms and use accessory muscles of respiration are better tolerated. Examples of such activities include pushing a shopping cart, walking on a treadmill, or pushing a wheelchair. As COPD advances, the principal feature is worsening dyspnea on exertion with increasing intrusion on the ability to

perform vocational or avocational activities. In the most advanced stages, patients are breathless doing simple activities of daily living.

Accompanying worsening airflow obstruction is an increased frequency of exacerbations (described below). Patients may also develop resting hypoxemia and require institution of supplemental oxygen.

Physical Findings

In the early stages of COPD, patients usually have an entirely normal physical examination. Current smokers may have signs of active smoking, including an odor of smoke or nicotine staining of fingernails. In patients with more severe disease, the physical examination is notable for a prolonged expiratory phase and may include expiratory wheezing. In addition, signs of hyperinflation include a barrel chest and enlarged lung volumes with poor diaphragmatic excursion as assessed by percussion. Patients with severe airflow obstruction may also exhibit use of accessory muscles of respiration, sitting in the characteristic "tripod" position to facilitate the actions of the sternocleidomastoid, scalene, and intercostal muscles. Patients may develop cyanosis, visible in the lips and nail beds.

Although traditional teaching is that patients with predominant emphysema, termed "pink puffers," are thin and noncyanotic at rest and have prominent use of accessory muscles, and patients with chronic bronchitis are more likely to be heavy and cyanotic ("blue bloaters"), current evidence demonstrates that most patients have elements of both bronchitis and emphysema and that the physical examination does not reliably differentiate the two entities.

Advanced disease may be accompanied by systemic wasting, with significant weight loss, bitemporal wasting, and diffuse loss of subcutaneous adipose tissue. This syndrome has been associated with both inadequate oral intake and elevated levels of inflammatory cytokines (TNF-α). Such wasting is an independent poor prognostic factor in COPD. Some patients with advanced disease have paradoxical inward movement of the rib cage with inspiration (Hoover's sign), the result of alteration of the vector of diaphragmatic contraction on the rib cage as a result of chronic hyperinflation.

Signs of overt right heart failure, termed *cor pulmonale*, are relatively infrequent since the advent of supplemental oxygen therapy.

Clubbing of the digits is not a sign of COPD, and its presence should alert the clinician to initiate an investigation for causes of clubbing. In this population, the development of lung cancer is the most likely explanation for newly developed clubbing.

Laboratory Findings

The hallmark of COPD is airflow obstruction (discussed above). Pulmonary function testing shows airflow obstruction with a reduction in FEV_1 and FEV_1/FVC (Chap. 252). With worsening disease severity, lung volumes may increase, resulting in an increase in total lung capacity, functional residual capacity, and residual volume. In patients with emphysema, the diffusing capacity may be reduced, reflecting the lung parenchymal destruction characteristic of the disease. The degree of airflow obstruction is an important prognostic factor in COPD and is the basis for the Global Initiative for Lung Disease (GOLD) redundant classification (**Table 260-1**). More recently it has been shown that a multifactorial index incorporating airflow obstruction, exercise performance, dyspnea, and body mass index is a better predictor of mortality rate than pulmonary function alone.

Table 260-1 Gold Criteria for COPD Severity

GOLD Stage	Severity	Symptoms	Spirometry
0	At Risk	Chronic cough, sputum production	Normal
I	Mild	With or without chronic cough or sputum production	FEV_1/FVC <0.7 and FEV_1 ≥80% predicted
IIA	Moderate	With or without chronic cough or sputum production	FEV_1/FVC <0.7 and 50% ≤FEV_1 <80% predicted
III	Severe	With or without chronic cough or sputum production	FEV_1/FVC <0.7 and 30% ≤FEV_1 <50% predicted
IV	Very Severe	With or without chronic cough or sputum production	FEV_1/FVC <0.7 and FEV_1 <30% predicted or FEV_1 <50% predicted with respiratory failure or signs of right heart failure

Abbreviation: GOLD, Global Initiative for Lung Disease.

Source: From RA Pauwels et al.

Arterial blood gases and oximetry may demonstrate resting or exertional hypoxemia. Arterial blood gases provide additional information about alveolar ventilation and acid-base status by measuring arterial P_{CO_2} and pH. The change in pH with P_{CO_2} is 0.08 units/10 mmHg acutely and 0.03 units/10 mmHg in the chronic state. Knowledge of the arterial pH therefore allows the classification of ventilatory failure, defined as P_{CO_2} >45 mmHg, into acute or chronic conditions. The arterial blood gas is an important component of the evaluation of patients presenting with symptoms of an exacerbation. An elevated hematocrit suggests the presence of chronic hypoxemia, as does the presence of signs of right ventricular hypertrophy.

Radiographic studies may assist in the classification of the type of COPD. Obvious bullae, paucity of parenchymal markings, or hyperlucency suggests the presence of emphysema. Increased lung volumes and flattening of the diaphragm suggest hyperinflation but do not provide information about chronicity of the changes. Computed tomography (CT) scan is the current definitive test for establishing the presence or absence of emphysema in living subjects (**Fig. 260-4**). From a practical perspective, the CT scan does little to influence therapy of COPD except in those individuals considering surgical therapy for their disease (described below).

Figure 260-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest CT scan of a patient with COPD who underwent a left single-lung transplant. Note the reduced parenchymal markings in the right lung (*left side of figure*) as compared to the left lung, representing emphysematous destruction of the lung, and mediastinal shift to the left, indicative of hyperinflation.

Recent guidelines have suggested testing for α_1AT deficiency in all subjects with COPD or asthma with chronic airflow obstruction. Measurement of the serum α_1AT level is a reasonable initial test. For subjects with low α_1AT levels, the definitive diagnosis of α_1AT deficiency requires protease inhibitor (PI) type determination. This is typically performed by isoelectric focusing of serum, which reflects the genotype at the PI locus for the common alleles and many of the rare PI alleles as well. Molecular genotyping of DNA can be performed for the common PI alleles (M, S, and Z).

Treatment: Chronic Obstructive Pulmonary Disease

STABLE PHASE COPD

Only three interventions—smoking cessation, oxygen therapy in chronically hypoxemic patients, and lung volume reduction surgery in selected patients with emphysema—have been demonstrated to influence the natural history of patients with COPD. There is currently suggestive, but not definitive, evidence that the use of inhaled glucocorticoids may alter mortality rate (but not lung function). All other current therapies are directed at improving symptoms and decreasing the frequency and severity

of exacerbations. The institution of these therapies should involve an assessment of symptoms, potential risks, costs, and benefits of therapy. This should be followed by an assessment of response to therapy, and a decision should be made whether or not to continue treatment.

PHARMACOTHERAPY

Smoking Cessation (See Also Chap. 395)

It has been shown that middle-aged smokers who were able to successfully stop smoking experienced a significant improvement in the rate of decline in pulmonary function, returning to annual changes similar to that of nonsmoking patients. Thus, all patients with COPD should be strongly urged to quit and educated about the benefits of quitting. An emerging body of evidence demonstrates that combining pharmacotherapy with traditional supportive approaches considerably enhances the chances of successful smoking cessation. There are three principal pharmacologic approaches to the problem: bupropion, originally developed as an antidepressant medication; nicotine replacement therapy available as gum, transdermal patch, inhaler, and nasal spray; and varenicline, a nicotinic acid receptor agonist/antagonist. Current recommendations from the U.S. Surgeon General are that all adult, nonpregnant smokers considering quitting be offered pharmacotherapy, in the absence of any contraindication to treatment.

Bronchodilators

In general, bronchodilators are used for symptomatic benefit in patients with COPD. The inhaled route is preferred for medication delivery as the incidence of side effects is lower than that seen with the use of parenteral medication delivery.

Anticholinergic Agents

Ipratropium bromide improves symptoms and produces acute improvement in FEV_1. Tiotropium, a long-acting anticholinergic, has been shown to improve symptoms and reduce exacerbations. Studies of both ipratropium and tiotropium have failed to demonstrate that either influences the rate of decline in FEV_1. In a large randomized clinical trial, there was a trend toward reduced mortality rate in the tiotropium-treated patients that approached, but did not reach, statistical significance. Side effects are minor, and a trial of inhaled anticholinergics is recommended in symptomatic patients with COPD.

Beta Agonists

These provide symptomatic benefit. The main side effects are tremor and tachycardia. Long-acting inhaled β agonists, such as salmeterol, have benefits comparable to ipratropium bromide. Their use is more convenient than short-acting agents. The addition of a β agonist to inhaled anticholinergic therapy has been demonstrated to provide incremental benefit. A recent report in asthma suggests that those patients, particularly African Americans, using a long-acting β agonist without concomitant inhaled corticosteroids have an increased risk of deaths from respiratory causes. The applicability of these data to patients with COPD is unclear.

Inhaled Glucocorticoids

Although a recent trial demonstrated an apparent benefit from the regular use of inhaled glucocorticoids on the rate of decline of lung function, a number of other well-designed randomized trials have not. Patients studied included those with mild to severe airflow obstruction and current and ex-smokers. Patients with significant acute response to inhaled β agonists were excluded from many of these trials, which may impact the generalizability of the findings. Their use has been associated with increased rates of oropharyngeal candidiasis and an increased rate of loss of bone density. Available data suggest that inhaled glucocorticoids reduce exacerbation frequency by ~25%. The impact of inhaled corticosteroids on mortality rates in COPD is controversial. A meta-analysis and several retrospective studies suggest a mortality benefit, but in a recently published randomized trial, differences in mortality rate approached, but did not reach, conventional criteria for statistical significance. A trial of inhaled glucocorticoids should be considered in patients with frequent exacerbations, defined as two or more per year, and in patients who demonstrate a significant amount of acute reversibility in response to inhaled bronchodilators.

Oral Glucocorticoids

The chronic use of oral glucocorticoids for treatment of COPD is not recommended because of an unfavorable benefit/risk ratio. The chronic use of oral glucocorticoids is associated with significant side effects, including osteoporosis, weight gain, cataracts, glucose intolerance, and increased risk of infection. A recent study demonstrated that patients tapered off chronic low-dose prednisone (~10 mg/d) did not experience any adverse effect on the frequency of exacerbations, health-related quality of life, or lung function. On average, patients lost ~4.5 kg (~10 lb) when steroids were withdrawn.

Theophylline

Theophylline produces modest improvements in expiratory flow rates and vital capacity and a slight improvement in arterial oxygen and carbon dioxide levels in patients with moderate to severe COPD. Nausea is a common side effect; tachycardia and tremor have also been reported. Monitoring of blood theophylline levels is typically required to minimize toxicity.

Oxygen

Supplemental O_2 is the only pharmacologic therapy demonstrated to unequivocally decrease mortality rates in patients with COPD. For patients with resting hypoxemia (resting O_2 saturation ≤88% or <90% with signs of pulmonary hypertension or right heart failure), the use of O_2 has been demonstrated to have a significant impact on mortality rate. Patients meeting these criteria should be on continual oxygen supplementation, as the mortality benefit is proportional to the number of hours/day oxygen is used. Various delivery systems are available, including portable systems that patients may carry to allow mobility outside the home.

Supplemental O_2 is commonly prescribed for patients with exertional hypoxemia or nocturnal hypoxemia. Although the rationale for supplemental O_2 in these settings is physiologically sound, the benefits of such therapy are not well substantiated.

Other Agents

N-acetyl cysteine has been used in patients with COPD for both its mucolytic and antioxidant properties. A prospective trial failed to find any benefit with respect to decline in lung function or prevention of exacerbations. Specific treatment in the form of IV α_1AT augmentation therapy is available for individuals with severe α_1AT deficiency. Despite sterilization procedures for these blood-derived products and the absence of reported cases of viral infection from therapy, some physicians recommend hepatitis B vaccination prior to starting augmentation therapy. Although biochemical efficacy of α_1AT augmentation therapy has been shown, a randomized controlled trial of α_1AT augmentation therapy has not definitively established the efficacy of augmentation therapy in reducing decline of pulmonary function. Eligibility for α_1AT augmentation therapy requires a serum α_1AT level <11 μM (approximately 50 mg/dL). Typically, Pi^Z individuals will qualify, although other rare types associated with severe deficiency (e.g., null-null) are also eligible. Since only a fraction of individuals with severe α_1AT deficiency will develop COPD, α_1AT augmentation therapy is not recommended for severely α_1AT-deficient persons with normal pulmonary function and a normal chest CT scan.

NONPHARMACOLOGIC THERAPIES

General Medical Care

Patients with COPD should receive the influenza vaccine annually. Polyvalent pneumococcal vaccine is also recommended, although proof of efficacy in this patient population is not definitive.

Pulmonary Rehabilitation

This refers to a treatment program that incorporates education and cardiovascular conditioning. In COPD, pulmonary rehabilitation has been demonstrated to improve health-related quality of life, dyspnea, and exercise capacity. It has also been shown to reduce rates of hospitalization over a 6- to 12-month period.

Lung Volume Reduction Surgery (LVRS)

Surgery to reduce the volume of lung in patients with emphysema was first introduced with minimal success in the 1950s and was reintroduced in the 1990s. Patients are excluded if they have significant pleural disease, a pulmonary artery systolic pressure >45 mmHg, extreme deconditioning, congestive heart failure, or other severe comorbid conditions. Patients with an FEV_1 <20% of predicted and either diffusely distributed emphysema on CT scan or diffusing capacity of lung for carbon monoxide (DL_{CO}) <20% of predicted have an increased mortality rate after the procedure and thus are not candidates for LVRS.

The National Emphysema Treatment trial demonstrated that LVRS offers both a mortality benefit and a symptomatic benefit in certain patients with emphysema. The anatomic distribution of emphysema and postrehabilitation exercise capacity are important prognostic characteristics. Patients with upper lobe–predominant emphysema and a low postrehabilitation exercise capacity are most likely to benefit from LVRS.

Lung Transplantation (See Also Chap. 266)

COPD is currently the second leading indication for lung transplantation (Fig. 260-4). Current recommendations are that candidates for lung transplantation should be <65 years; have severe disability despite maximal medical therapy; and be free of comorbid conditions such as liver, renal, or cardiac disease. In contrast to LVRS, the anatomic distribution of emphysema and the presence of pulmonary hypertension are not contraindications to lung transplantation. Unresolved issues concerning lung transplantation and COPD include whether single- or double-lung transplant is the preferred procedure.

EXACERBATIONS OF COPD

Exacerbations are a prominent feature of the natural history of COPD. Exacerbations are episodes of

increased dyspnea and cough and change in the amount and character of sputum. They may or may not be accompanied by other signs of illness, including fever, myalgias, and sore throat. Self-reported health-related quality of life correlates with frequency of exacerbations more closely than it does with the degree of airflow obstruction. Economic analyses have shown that >70% of COPD-related health care expenditures go to emergency department visits and hospital care; this translates to >$10 billion annually in the United States. The frequency of exacerbations increases as airflow obstruction increases; patients with moderate to severe airflow obstruction [GOLD stages III, IV (Table 260-1)] on average have one to three episodes per year. However, some individuals with very severe airflow obstruction do not have frequent exacerbations; the history of prior exacerbations is a strong predictor of future exacerbations.

The approach to the patient experiencing an exacerbation includes an assessment of the severity of the patient's illness, both acute and chronic components; an attempt to identify the precipitant of the exacerbation; and the institution of therapy.

Precipitating Causes and Strategies to Reduce Frequency of Exacerbations

A variety of stimuli may result in the final common pathway of airway inflammation and increased symptoms that are characteristic of COPD exacerbations. Bacterial infections play a role in many, but by no means all, episodes. Viral respiratory infections are present in approximately one-third of COPD exacerbations. In a significant minority of instances (20–35%), no specific precipitant can be identified.

Despite the frequent implication of bacterial infection, chronic suppressive or "rotating" antibiotics are not beneficial in patients with COPD. This is in contrast to their efficacy in patients with bronchiectasis due to cystic fibrosis, in whom suppressive antibiotics have been shown to reduce frequency of hospital admissions.

The role of pharmacotherapy in reducing exacerbation frequency is less well studied. Chronic oral glucocorticoids are not recommended for this purpose. Inhaled glucocorticoids reduce the frequency of exacerbations by 25–30% in most analyses. The use of inhaled glucocorticoids should be considered in patients with frequent exacerbations or those who have an asthmatic component, i.e., significant reversibility on pulmonary function testing or marked symptomatic improvement after inhaled bronchodilators. Similar magnitudes of reduction have been reported for anticholinergic and long-acting beta-agonist therapy. The influenza vaccine has been shown to reduce exacerbation rates in patients with COPD.

Patient Assessment

An attempt should be made to establish the severity of the exacerbation as well as the severity of preexisting COPD. The more severe either of these two components, the more likely that the patient will require hospital admission. The history should include quantification of the degree of dyspnea by asking about breathlessness during activities of daily living and typical activities for the patient. The patient should be asked about fever; change in character of sputum; any ill contacts; and associated symptoms such as nausea, vomiting, diarrhea, myalgias, and chills. Inquiring about the frequency and severity of prior exacerbations can provide important information.

The physical examination should incorporate an assessment of the degree of distress of the patient. Specific attention should be focused on tachycardia, tachypnea, use of accessory muscles, signs of perioral or peripheral cyanosis, the ability to speak in complete sentences, and the patient's mental status. The chest examination should establish the presence or absence of focal findings, degree of air movement, presence or absence of wheezing, asymmetry in the chest examination (suggesting large airway obstruction or pneumothorax mimicking an exacerbation), and the presence or absence of paradoxical motion of the abdominal wall.

Patients with severe underlying COPD, who are in moderate or severe distress or those with focal findings should have a chest x-ray. Approximately 25% of x-rays in this clinical situation will be abnormal, with the most frequent findings being pneumonia and congestive heart failure. Patients with advanced COPD, those with a history of hypercarbia, those with mental status changes (confusion, sleepiness), or those in significant distress should have an arterial blood-gas measurement. The presence of hypercarbia, defined as a P_{CO_2} >45 mmHg, has important implications for treatment (discussed below). In contrast to its utility in the management of exacerbations of asthma, measurement of pulmonary function has not been demonstrated to be helpful in the diagnosis or management of exacerbations of COPD.

There are no definitive guidelines concerning the need for inpatient treatment of exacerbations. Patients with respiratory acidosis and hypercarbia, significant hypoxemia, or severe underlying disease or those whose living situation is not conducive to careful observation and the delivery of prescribed treatment should be admitted to the hospital.

ACUTE EXACERBATIONS

Bronchodilators

Typically, patients are treated with an inhaled β agonist, often with the addition of an anticholinergic

agent. These may be administered separately or together, and the frequency of administration depends on the severity of the exacerbation. Patients are often treated initially with nebulized therapy, as such treatment is often easier to administer in older patients or to those in respiratory distress. It has been shown, however, that conversion to metered-dose inhalers is effective when accompanied by education and training of patients and staff. This approach has significant economic benefits and also allows an easier transition to outpatient care. The addition of methylxanthines (such as theophylline) to this regimen can be considered, although convincing proof of its efficacy is lacking. If added, serum levels should be monitored in an attempt to minimize toxicity.

Antibiotics

Patients with COPD are frequently colonized with potential respiratory pathogens, and it is often difficult to identify conclusively a specific species of bacteria responsible for a particular clinical event. Bacteria frequently implicated in COPD exacerbations include *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis*. In addition, *Mycoplasma pneumoniae* or *Chlamydia pneumoniae* are found in 5–10% of exacerbations. The choice of antibiotic should be based on local patterns of antibiotic susceptibility of the above pathogens as well as the patient's clinical condition. Most practitioners treat patients with moderate or severe exacerbations with antibiotics, even in the absence of data implicating a specific pathogen.

Glucocorticoids

Among patients admitted to the hospital, the use of glucocorticoids has been demonstrated to reduce the length of stay, hasten recovery, and reduce the chance of subsequent exacerbation or relapse for a period of up to 6 months. One study demonstrated that 2 weeks of glucocorticoid therapy produced benefit indistinguishable from 8 weeks of therapy. The GOLD guidelines recommend 30–40 mg of oral prednisolone or its equivalent for a period of 10–14 days. Hyperglycemia, particularly in patients with preexisting diagnosis of diabetes, is the most frequently reported acute complication of glucocorticoid treatment.

Oxygen

Supplemental O_2 should be supplied to keep arterial saturations ≥90%. Hypoxemic respiratory drive plays a small role in patients with COPD. Studies have demonstrated that in patients with both acute and chronic hypercarbia, the administration of supplemental O_2 does not reduce minute ventilation. It does, in some patients, result in modest increases in arterial P_{CO_2}, chiefly by altering ventilation-perfusion relationships within the lung. This should not deter practitioners from providing the oxygen needed to correct hypoxemia.

Mechanical Ventilatory Support

The initiation of noninvasive positive-pressure ventilation (NIPPV) in patients with respiratory failure, defined as Pa_{CO_2} >45 mmHg, results in a significant reduction in mortality rate, need for intubation, complications of therapy, and hospital length of stay. Contraindications to NIPPV include cardiovascular instability, impaired mental status or inability to cooperate, copious secretions or the inability to clear secretions, craniofacial abnormalities or trauma precluding effective fitting of mask, extreme obesity, or significant burns.

Invasive (conventional) mechanical ventilation via an endotracheal tube is indicated for patients with severe respiratory distress despite initial therapy, life-threatening hypoxemia, severe hypercarbia and/or acidosis, markedly impaired mental status, respiratory arrest, hemodynamic instability, or other complications. The goal of mechanical ventilation is to correct the aforementioned conditions. Factors to consider during mechanical ventilatory support include the need to provide sufficient expiratory time in patients with severe airflow obstruction and the presence of auto-PEEP (positive end-expiratory pressure), which can result in patients having to generate significant respiratory effort to trigger a breath during a demand mode of ventilation. The mortality rate of patients requiring mechanical ventilatory support is 17–30% for that particular hospitalization. For patients age >65 admitted to the intensive care unit for treatment, the mortality rate doubles over the next year to 60%, regardless of whether mechanical ventilation was required.

FURTHER READINGS

American Thoracic Society/European Respiratory Society Task Force: Standards for the diagnosis and management of patients with COPD [Internet]. Version 1.2. New York: American Thoracic Society; 2004 [updated September 8, 2005]. Available from *http://www-test.thoracic.org/copd/*

Fiore MC et al: *Treating Tobacco Use and Dependence*, Clinical Practice Guideline. Rockville, MD: U.S. Department of Health and Human Services. Public Health Service, June 2000

Hunninghake GM et al: MMP12, lung function, and COPD in high-risk populations. N Engl J Med 361:2599, 2009[PMID: 20018959] [Full Text]

Ito K et al: Decreased histone deacetylase activity in chronic obstructive pulmonary disease. N Engl J Med 352:1967, 2005[PMID: 15888697] [Full Text]

Mannino DM, Buist AS: Global burden of COPD: Risk factors, prevalence, and future trends. Lancet 370:765, 2007[PMID: 17765526] [Full Text]

Rabe KF et al: Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 176:532, 2007[PMID: 17507545] [Full Text]

———et al: Update in chronic obstructive pulmonary disease 2006. Am J Respir Crit Care Med 175:1222, 2007

Sin DD et al: Inhaled corticosteroids and mortality in chronic obstructive pulmonary disease. Thorax 60:992, 2005[PMID: 16227327] [Full Text]

Wise RA, Tashkin DP: Optimizing treatment of chronic obstructive pulmonary disease: An assessment of current therapies. Am J Med 120:S4, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTERSTITIAL LUNG DISEASES: INTRODUCTION

Patients with interstitial lung diseases (ILDs) come to medical attention mainly because of the onset of progressive exertional dyspnea or a persistent nonproductive cough. Hemoptysis, wheezing, and chest pain may be present. Often, the identification of interstitial opacities on chest x-ray focuses the diagnostic approach on one of the ILDs.

ILDs represent a large number of conditions that involve the parenchyma of the lung—the alveoli, the alveolar epithelium, the capillary endothelium, and the spaces between those structures—as well as the perivascular and lymphatic tissues. The disorders in this heterogeneous group are classified together because of similar clinical, roentgenographic, physiologic, or pathologic manifestations. These disorders often are associated with considerable rates of morbidity and mortality, and there is little consensus regarding the best management of most of them.

ILDs have been difficult to classify because >200 known individual diseases are characterized by diffuse parenchymal lung involvement, either as the primary condition or as a significant part of a multiorgan process, as may occur in the connective tissue diseases (CTDs). One useful approach to classification is to separate the ILDs into two groups based on the major underlying histopathology: (1) those associated with predominant inflammation and fibrosis and (2) those with a predominantly granulomatous reaction in interstitial or vascular areas **(Table 261-1)**. Each of these groups can be subdivided further according to whether the cause is known or unknown. For each ILD there may be an acute phase, and there is usually a chronic one as well. Rarely, some are recurrent, with intervals of subclinical disease.

Table 261-1 Major Categories of Alveolar and Interstitial Inflammatory Lung Disease

Lung Response: Alveolitis, Interstitial Inflammation, and Fibrosis	
Known Cause	
Asbestos	Residual of acute respiratory distress syndrome
Fumes, gases	Smoking-related
Drugs (antibiotics, amiodarone, gold) and chemotherapy drugs	Desquamative interstitial pneumonia
Radiation	Respiratory bronchiolitis–associated interstitial lung disease
Aspiration pneumonia	Langerhans cell granulomatosis (eosinophilic granuloma of the lung)
Unknown Cause	
Idiopathic interstitial pneumonias	Pulmonary alveolar proteinosis
Idiopathic pulmonary fibrosis (usual interstitial pneumonia)	Lymphocytic infiltrative disorders (lymphocytic interstitial pneumonitis associated with connective tissue disease)
Acute interstitial pneumonia (diffuse alveolar damage)	Eosinophilic pneumonias
Cryptogenic organizing pneumonia (bronchiolitis obliterans with organizing pneumonia)	Lymphangioleiomyomatosis
Nonspecific interstitial pneumonia	Amyloidosis
Connective tissue diseases	Inherited diseases
Systemic lupus erythematosus, rheumatoid arthritis, ankylosing spondylitis, systemic sclerosis, Sjögren's syndrome, polymyositis-dermatomyositis	Tuberous sclerosis, neurofibromatosis, Niemann-Pick disease, Gaucher's disease, Hermansky-Pudlak syndrome
Pulmonary hemorrhage syndromes	Gastrointestinal or liver diseases (Crohn's disease, primary biliary cirrhosis, chronic active hepatitis, ulcerative colitis)
Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, isolated pulmonary capillaritis	Graft-versus-host disease (bone marrow transplantation; solid organ transplantation)
Lung Response: Granulomatous	

Known Cause	
Hypersensitivity pneumonitis (organic dusts)	Inorganic dusts: beryllium, silica
Unknown Cause	
Sarcoidosis	Bronchocentric granulomatosis
Granulomatous vasculitides	Lymphomatoid granulomatosis
Granulomatosis with polyangiitis (Wegener's), allergic granulomatosis of Churg-Strauss	

Sarcoidosis (Chap. 329), idiopathic pulmonary fibrosis (IPF), and pulmonary fibrosis associated with CTDs (Chaps. 319–326) are the most common ILDs of unknown etiology. Among the ILDs of known cause, the largest group includes occupational and environmental exposures, especially the inhalation of inorganic dusts, organic dusts, and various fumes or gases (Chaps. 255 and 256) **(Table 261-2)**. A clinical diagnosis is possible for many forms of ILD, especially if an occupational and environmental history is pursued aggressively. High-resolution computed tomography (HRCT) scanning improves the diagnostic accuracy and may eliminate the need for tissue examination in many cases, especially in IPF. For other forms, tissue examination, usually obtained by thoracoscopic lung biopsy, is critical to confirmation of the diagnosis.

Table 261-2 Estimated Relative Frequency of the Interstitial Lung Diseases

Diagnosis	Relative Frequency, %
Idiopathic interstitial pneumonias	40
Idiopathic pulmonary fibrosis	55
Nonspecific interstitial pneumonia	25
Respiratory bronchiolitis—ILD and desquamative interstitial pneumonia	15
Cryptogenic organizing pneumonia	3
Acute interstitial pneumonia	<1
Occupational and environmental	26
Sarcoidosis	10
Connective tissue diseases	9
Drug and radiation	1
Pulmonary hemorrhage syndromes	<1
Other	13

Source: From DB Coultas, R Hubbard, in JP Lynch III (ed): *Lung Biology in Health and Disease.* New York, Marcel Dekker, 2004; S Garantziotis et al: J Clin Invest 114:319, 2004.

Pathogenesis

The ILDs are nonmalignant disorders and are not caused by identified infectious agents. The precise pathway(s) leading from injury to fibrosis is not known. Although there are multiple initiating agent(s) of injury, the immunopathogenic responses of lung tissue are limited, and the mechanisms of repair have common features **(Fig. 261-1)**.

Figure 261-1

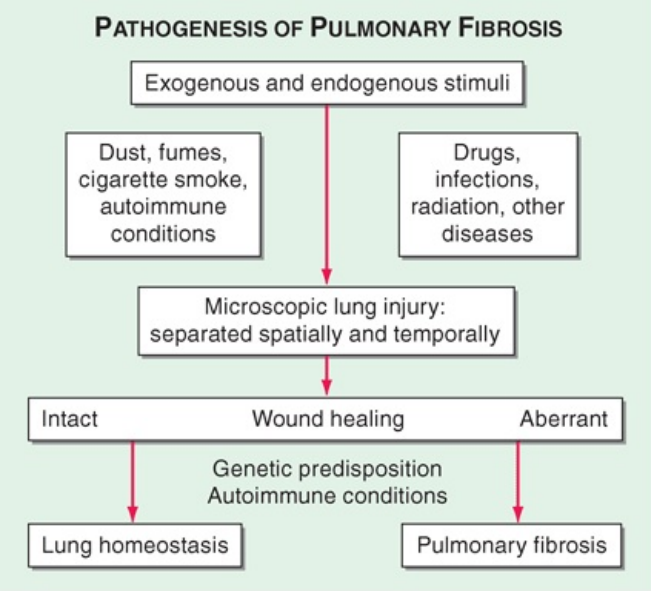

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Proposed mechanism for the pathogenesis of pulmonary fibrosis. The lung is naturally exposed to repetitive injury from a variety of exogenous and endogenous stimuli. Several local and systemic factors (e.g., fibroblasts, circulating fibrocytes, chemokines, growth factors, and clotting factors) contribute to tissue healing and functional recovery. Dysregulation of this intricate network through genetic predisposition, autoimmune conditions, or superimposed diseases can lead to aberrant wound healing, with the result of pulmonary fibrosis. Alternatively, excessive injury to the lung may overwhelm even intact reparative mechanisms and lead to pulmonary fibrosis. *(From S Garantziotis et al: J Clin Invest 114:319, 2004.)*

As mentioned above, the two major histopathologic patterns are a granulomatous pattern and a pattern in which inflammation and fibrosis predominate.

GRANULOMATOUS LUNG DISEASE

This process is characterized by an accumulation of T lymphocytes, macrophages, and epithelioid cells organized into discrete structures (granulomas) in the lung parenchyma. The granulomatous lesions can progress to fibrosis. Many patients with granulomatous lung disease remain free of severe impairment of lung function or, when symptomatic, improve after treatment. The main differential diagnosis is between sarcoidosis (Chap. 329) and hypersensitivity pneumonitis (Chap. 255).

INFLAMMATION AND FIBROSIS

The initial insult is an injury to the epithelial surface that causes inflammation in the air spaces and alveolar walls **(Fig. 261-2)**. If the disease becomes chronic, inflammation spreads to adjacent portions of the interstitium and vasculature and eventually causes interstitial fibrosis. Important histopathologic patterns found in the ILDs include usual interstitial pneumonia (UIP), nonspecific interstitial pneumonia, respiratory bronchiolitis/desquamative interstitial pneumonia, organizing pneumonia, diffuse alveolar damage (acute or organizing), and lymphocytic interstitial pneumonia. The development of irreversible scarring (fibrosis) of alveolar walls, airways, or vasculature is the most feared outcome in all of these conditions because it is often progressive and leads to significant derangement of ventilatory function and gas exchange.

Figure 261-2

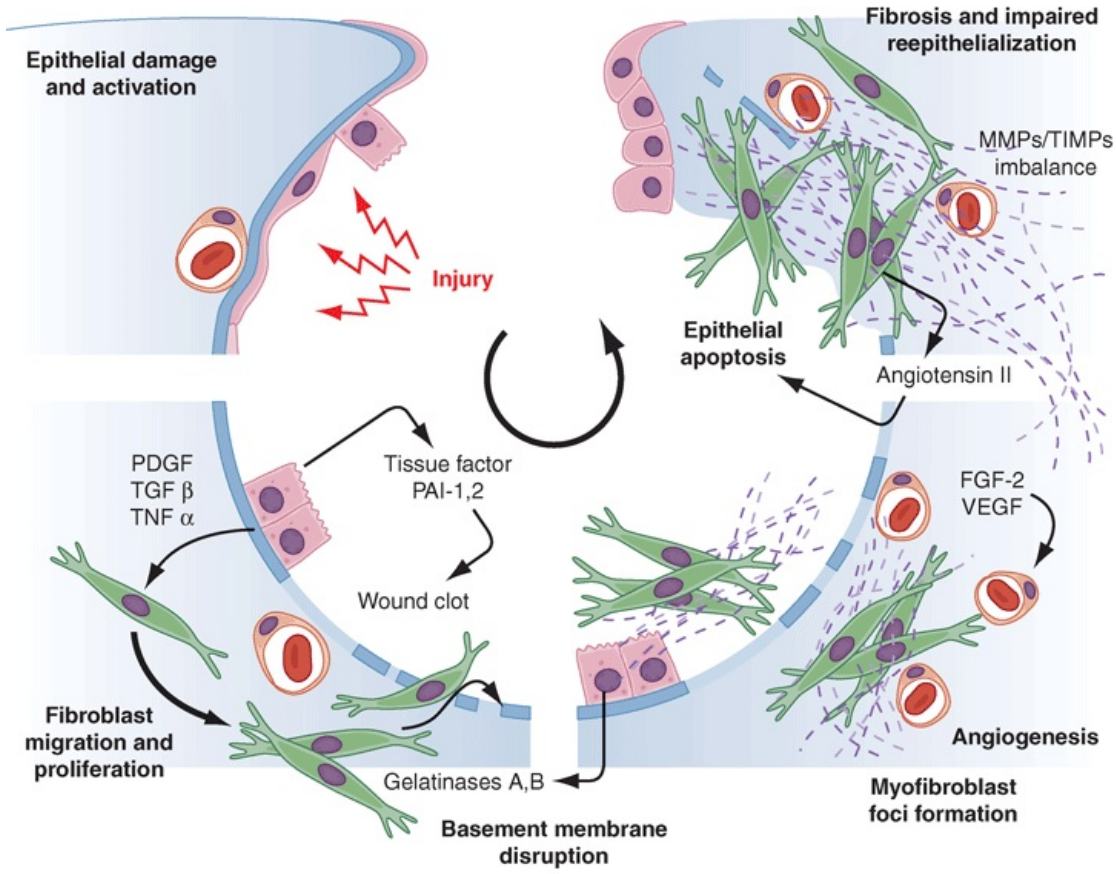

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cellular basis for the pathogenesis of interstitial lung disease. Multiple microinjuries damage and activate alveolar epithelial cells (*top left*), which in turn induce an antifibrinolytic environment in the alveolar spaces, enhancing wound clot formation. Alveolar epithelial cells secrete growth factors and induce migration and proliferation of fibroblasts and differentiation into myofibroblasts (*bottom left*). Subepithelial myofibroblasts and alveolar epithelial cells produce gelatinases that may increase basement membrane disruption and allow fibroblast–myofibroblast migration (*bottom right*). Angiogenic factors induce neovascularization. Both intraalveolar and interstitial myofibroblasts secrete extracellular matrix proteins, mainly collagens. An imbalance between interstitial collagenases and tissue inhibitors of metalloproteinases provokes the progressive deposit of extracellular matrix (*top right*). Signals responsible for myofibroblast apoptosis seem to be absent or delayed in usual interstitial pneumonia, increasing cell survival. Myofibroblasts produce angiotensinogen that, as angiotensin II, provokes alveolar epithelial cell death, further impairing reepithelialization. Abbreviations: FGF-2, fibroblast growth factor 2; MMPs, metalloproteinases; PAI-1, PAI-2, plasminogen activator inhibitor 1, 2; PDGF, platelet-derived growth factor; TGF-β, transforming growth factor β; TIMPs, tissue inhibitors of metalloproteinases; TNF-α, tumor necrosis factor α; VEGF, vascular endothelial growth factor. *(From M Selman et al: Ann Intern Med 134:136, 2001, with permission.)*

History

DURATION OF ILLNESS

Acute presentation (days to weeks), although unusual, occurs with allergy (drugs, fungi, helminths), acute interstitial pneumonia (AIP), eosinophilic pneumonia, and hypersensitivity pneumonitis. These conditions may be confused with atypical pneumonias because of diffuse alveolar opacities on chest x-ray. *Subacute presentation* (weeks to months) may occur in all ILDs but is seen especially in sarcoidosis, drug-induced ILDs, the alveolar hemorrhage syndromes, cryptogenic organizing pneumonia (COP), and the acute immunologic pneumonia that complicates systemic lupus erythematosus (SLE) or polymyositis. In most ILDs the symptoms and signs form a *chronic presentation* (months to years). Examples include IPF, sarcoidosis, pulmonary Langerhans cell histiocytosis (PLCH) (also known as Langerhans cell granulomatosis, eosinophilic granuloma, or histiocytosis X), pneumoconioses, and CTDs. *Episodic presentations* are unusual and include eosinophilic pneumonia, hypersensitivity pneumonitis, COP, vasculitides, pulmonary hemorrhage, and Churg-Strauss syndrome.

AGE

Most patients with sarcoidosis, ILD associated with CTD, lymphangioleiomyomatosis (LAM), PLCH, and inherited forms of ILD (familial IPF, Gaucher's disease, Hermansky-Pudlak syndrome) present

between the ages of 20 and 40 years. Most patients with IPF are older than 60 years.

GENDER

LAM and pulmonary involvement in tuberous sclerosis occur exclusively in premenopausal women. In addition, ILD in Hermansky-Pudlak syndrome and in the CTDs is more common in women; an exception is ILD in rheumatoid arthritis (RA), which is more common in men. IPF is more common in men. Because of occupational exposures, pneumoconioses also occur more frequently in men.

FAMILY HISTORY

Familial lung fibrosis has been associated with mutations in three genes: the surfactant protein Cgene, the surfactant protein A2 gene, and the ATP-binding cassette transporter A3 gene. Familial lung fibrosis is characterized by several patterns of interstitial pneumonia, including nonspecific interstitial pneumonia, desquamative interstitial pneumonia, and UIP. Older age, male sex, and a history of cigarette smoking have been identified as risk factors for familial lung fibrosis. Family associations (with an autosomal dominant pattern) have been identified in tuberous sclerosis and neurofibromatosis. Familial clustering has been identified increasingly in sarcoidosis. The genes responsible for several rare ILDs have been identified, i.e., alveolar microlithiasis, Gaucher's disease, Hermansky-Pudlak syndrome, and Niemann-Pick disease, along with the genes for surfactant homeostasis in pulmonary alveolar proteinosis and for control of cell growth and differentiation in LAM.

SMOKING HISTORY

Two-thirds to 75% of patients with IPF and familial lung fibrosis have a history of smoking. Patients with PLCH, respiratory bronchiolitis/desquamative interstitial pneumonia (DIP), Goodpasture's syndrome, respiratory bronchiolitis, and pulmonary alveolar proteinosis are usually current or former smokers.

OCCUPATIONAL AND ENVIRONMENTAL HISTORY

A strict chronologic listing of the patient's lifelong employment must be sought, including specific duties and known exposures. In hypersensitivity pneumonitis (see Fig. 255-1), respiratory symptoms, fever, chills, and an abnormal chest roentgenogram are often temporally related to a hobby (pigeon breeder's disease) or to the workplace (farmer's lung) (Chap. 255). Symptoms may diminish or disappear after the patient leaves the site of exposure for several days; similarly, symptoms may reappear when the patient returns to the exposure site.

OTHER IMPORTANT PAST HISTORY

Parasitic infections may cause pulmonary eosinophilia, and therefore a travel history should be taken in patients with known or suspected ILD. History of risk factors for HIV infection should be elicited because several processes may occur at the time of initial presentation or during the clinical course, e.g., HIV infection, organizing pneumonia, AIP, lymphocytic interstitial pneumonitis, and diffuse alveolar hemorrhage.

RESPIRATORY SYMPTOMS AND SIGNS

Dyspnea is a common and prominent complaint in patients with ILD, especially the idiopathic interstitial pneumonias, hypersensitivity pneumonitis, COP, sarcoidosis, eosinophilic pneumonias, and PLCH. Some patients, especially patients with sarcoidosis, silicosis, PLCH, hypersensitivity pneumonitis, lipoid pneumonia, or lymphangitis carcinomatosis, may have extensive parenchymal lung disease on chest x-ray without significant dyspnea, especially early in the course of the illness. Wheezing is an uncommon manifestation of ILD but has been described in patients with chronic eosinophilic pneumonia, Churg-Strauss syndrome, respiratory bronchiolitis, and sarcoidosis. Clinically significant chest pain is uncommon in most ILDs. However, substernal discomfort is common in sarcoidosis. Sudden worsening of dyspnea, especially if associated with acute chest pain, may indicate a spontaneous pneumothorax, which occurs in PLCH, tuberous sclerosis, LAM, and neurofibromatosis. Frank hemoptysis and blood-streaked sputum are rarely presenting manifestations of ILD but can be seen in the diffuse alveolar hemorrhage (DAH) syndromes, LAM, tuberous sclerosis, and the granulomatous vasculitides. Fatigue and weight loss are common in all ILDs.

Physical Examination

The findings are usually not specific. Most commonly, physical examination reveals tachypnea and bibasilar end-inspiratory dry crackles, which are common in most forms of ILD associated with inflammation but are less likely to be heard in the granulomatous lung diseases. Crackles may be present in the absence of radiographic abnormalities on the chest radiograph. Scattered late inspiratory high-pitched rhonchi—so-called inspiratory squeaks—are heard in patients with bronchiolitis. The cardiac examination is usually normal except in the middle or late stages of the disease, when findings of pulmonary hypertension and cor pulmonale may become evident (Chap. 250). Cyanosis and clubbing of the digits occur in some patients with advanced disease.

Laboratory

Antinuclear antibodies and anti-immunoglobulin antibodies (rheumatoid factors) are identified in some patients, even in the absence of a defined CTD. A raised lactate dehydrogenase (LDH) level is a

nonspecific finding common to ILDs. Elevation of the serum level of angiotensin-converting enzyme is common in sarcoidosis. Serum precipitins confirm exposure when hypersensitivity pneumonitis is suspected, although they are not diagnostic of the process. Antineutrophil cytoplasmic or anti-basement membrane antibodies are useful if vasculitis is suspected. The electrocardiogram is usually normal unless pulmonary hypertension is present; then it demonstrates right-axis deviation, right ventricular hypertrophy, or right atrial enlargement or hypertrophy. Echocardiography also reveals right ventricular dilation and/or hypertrophy in the presence of pulmonary hypertension.

Chest Imaging Studies

CHEST X-RAY

ILD may be first suspected on the basis of an abnormal chest radiograph, which most commonly reveals a bibasilar reticular pattern. A nodular or mixed pattern of alveolar filling and increased reticular markings also may be present. A subgroup of ILDs exhibit nodular opacities with a predilection for the upper lung zones [sarcoidosis, PLCH, chronic hypersensitivity pneumonitis, silicosis, berylliosis, RA (necrobiotic nodular form), ankylosing spondylitis]. The chest x-ray correlates poorly with the clinical or histopathologic stage of the disease. The radiographic finding of honeycombing correlates with pathologic findings of small cystic spaces and progressive fibrosis; when present, it portends a poor prognosis. In most cases, the chest radiograph is nonspecific and usually does not allow a specific diagnosis.

COMPUTED TOMOGRAPHY

High-resolution computed tomography is superior to the plain chest x-ray for early detection and confirmation of suspected ILD **(Fig. 261-3)**. In addition, HRCT allows better assessment of the extent and distribution of disease, and it is especially useful in the investigation of patients with a normal chest radiograph. Coexisting disease is often best recognized on HRCT scanning, e.g., mediastinal adenopathy, carcinoma, or emphysema. In the appropriate clinical setting HRCT may be sufficiently characteristic to preclude the need for lung biopsy in IPF, sarcoidosis, hypersensitivity pneumonitis, asbestosis, lymphangitic carcinoma, and PLCH. When a lung biopsy is required, HRCT scanning is useful for determining the most appropriate area from which biopsy samples should be taken.

Figure 261-3

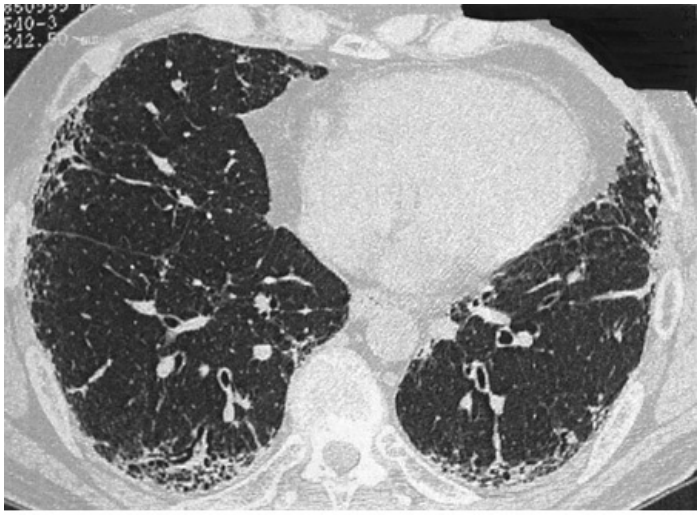

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Idiopathic pulmonary fibrosis. High-resolution CT image shows bibasal, peripheral predominant reticular abnormality with traction bronchiectasis and honeycombing. The lung biopsy showed the typical features of usual interstitial pneumonia.

Pulmonary Function Testing

SPIROMETRY AND LUNG VOLUMES

Measurement of lung function is important in assessing the extent of pulmonary involvement in patients with ILD. Most forms of ILD produce a restrictive defect with reduced total lung capacity (TLC), functional residual capacity, and residual volume (Chap. 252). Forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC) are reduced, but these changes are related to the decreased TLC. The FEV_1/FVC ratio is usually normal or increased. Lung volumes decrease as lung stiffness worsens with disease progression. A few disorders produce interstitial opacities on chest x-ray and obstructive airflow limitation on lung function testing (uncommon in sarcoidosis and hypersensitivity pneumonitis but common in tuberous sclerosis and LAM). Pulmonary function studies

have been proved to have prognostic value in patients with idiopathic interstitial pneumonias, particularly IPF and nonspecific interstitial pneumonia (NSIP).

DIFFUSING CAPACITY

A reduction in the diffusing capacity of the lung for carbon monoxide (DI_{CO}) is a common but nonspecific finding in most ILDs. This decrease is due in part to effacement of the alveolar capillary units but, more important, to mismatching of ventilation and perfusion (V./Q.). Lung regions with reduced compliance due to either fibrosis or cellular infiltration may be poorly ventilated but may still maintain adequate blood flow, and the ventilation-perfusion mismatch in these regions acts like true venous admixture. The severity of the reduction in DI_{CO} does not correlate with disease stage.

ARTERIAL BLOOD GAS

The resting arterial blood gas may be normal or reveal hypoxemia (secondary to a mismatching of ventilation to perfusion) and respiratory alkalosis. A normal arterial O_2 tension (or saturation by oximetry) at rest does not rule out significant hypoxemia during exercise or sleep. Carbon dioxide (CO_2) retention is rare and is usually a manifestation of end-stage disease.

Cardiopulmonary Exercise Testing

Because hypoxemia at rest is not always present and because severe exercise-induced hypoxemia may go undetected, it is useful to perform exercise testing with measurement of arterial blood gases to detect abnormalities of gas exchange. Arterial oxygen desaturation, a failure to decrease dead space appropriately with exercise [i.e., a high Vd/Vt (dead space/tidal volume) ratio (Chap. 252)], and an excessive increase in respiratory rate with a lower than expected recruitment of tidal volume provide useful information about physiologic abnormalities and extent of disease. Serial assessment of resting and exercise gas exchange is an excellent method for following disease activity and responsiveness to treatment, especially in patients with IPF. Increasingly, the 6-min walk test is used to obtain a global evaluation of submaximal exercise capacity in patients with ILD. The walk distance and level of oxygen desaturation tend to correlate with the patient's baseline lung function and mirror the patient's clinical course.

Fiberoptic Bronchoscopy and Bronchoalveolar Lavage (BAL)

In selected diseases (e.g., sarcoidosis, hypersensitivity pneumonitis, DAH syndrome, cancer, pulmonary alveolar proteinosis), cellular analysis of BAL fluid may be useful in narrowing the differential diagnostic possibilities among various types of ILD **(Table 261-3)**. The role of BAL in defining the stage of disease and assessment of disease progression or response to therapy remains poorly understood, and the usefulness of BAL in the clinical assessment and management remains to be established.

Table 261-3 Diagnostic Value of Bronchoalveolar Lavage in Interstitial Lung Disease

Condition	Bronchoalveolar Lavage Finding
Sarcoidosis	Lymphocytosis; CD4:CD8 ratio >3.5 most specific of diagnosis
Hypersensitivity pneumonitis	Marked lymphocytosis (>50%)
Organizing pneumonia	Foamy macrophages; mixed pattern of increased cells characteristic; decreased CD4:CD8 ratio
Eosinophilic lung disease	Eosinophils >25%
Diffuse alveolar bleeding	Hemosiderin-laden macrophages, red blood cells
Diffuse alveolar damage, drug toxicity	Atypical hyperplastic type II pneumocytes
Opportunistic infections	*Pneumocystis carinii*, fungi, cytomegalovirus-transformed cells
Lymphangitic carcinomatosis, alveolar cell carcinoma, pulmonary lymphoma	Malignant cells
Alveolar proteinosis	Milky effluent, foamy macrophages and lipoproteinaceous intraalveolar material (periodic acid–Schiff stain–positive)
Lipoid pneumonia	Fat globules in macrophages
Pulmonary Langerhans cell histiocytosis	Increased CD1+ Langerhans cells, electron microscopy demonstrating Birbeck granule in lavaged macrophage (expensive and difficult to perform)
Asbestos-related pulmonary disease	Dust particles, ferruginous bodies
Berylliosis	Positive lymphocyte transformation test to beryllium
Silicosis	Dust particles by polarized light microscopy

Lipoidosis	Accumulation of specific lipopigment in alveolar macrophages

Tissue and Cellular Examination

Lung biopsy is the most effective method for confirming the diagnosis and assessing disease activity. The findings may identify a more treatable process than originally suspected, particularly chronic hypersensitivity pneumonitis, COP, respiratory bronchiolitis–associated ILD, or sarcoidosis. Biopsy should be obtained before the initiation of treatment. A definitive diagnosis avoids confusion and anxiety later in the clinical course if the patient does not respond to therapy or experiences serious side effects from it.

Fiberoptic bronchoscopy with multiple transbronchial lung biopsies (four to eight biopsy samples) is often the initial procedure of choice, especially when sarcoidosis, lymphangitic carcinomatosis, eosinophilic pneumonia, Goodpasture's syndrome, or infection is suspected. If a specific diagnosis is not made by transbronchial biopsy, surgical lung biopsy by video-assisted thoracic surgery or open thoracotomy is indicated. Adequate-sized biopsies from multiple sites, usually from two lobes, should be obtained. Relative contraindications to lung biopsy include serious cardiovascular disease, honeycombing and other roentgenographic evidence of diffuse end-stage disease, severe pulmonary dysfunction, and other major operative risks, especially in the elderly.

Treatment: Interstitial Lung Disease

Although the course of ILD is variable, progression is common and often insidious. All treatable possibilities should be carefully considered. Since therapy does not reverse fibrosis, the major goals of treatment are permanent removal of the offending agent, when known, and early identification and aggressive suppression of the acute and chronic inflammatory process, thereby reducing further lung damage. Hypoxemia (PaO_2 <55 mmHg) at rest and/or with exercise should be managed with supplemental oxygen. Management of cor pulmonale may be required as the disease progresses (Chaps. 234 and 250). Pulmonary rehabilitation has been shown to improve the quality of life in patients with ILD.

DRUG THERAPY

Glucocorticoids are the mainstay of therapy for suppression of the alveolitis present in ILD, but the success rate is low. There have been no placebo-controlled trials of glucocorticoids in ILD, and so there is no direct evidence that steroids improve survival in many of the diseases for which they are commonly used. Glucocorticoid therapy is recommended for symptomatic ILD patients with eosinophilic pneumonias, COP, CTD, sarcoidosis, hypersensitivity pneumonitis, acute inorganic dust exposures, acute radiation pneumonitis, DAH, and drug-induced ILD. In organic dust disease, glucocorticoids are recommended for both the acute and chronic stages.

The optimal dose and proper length of therapy with glucocorticoids in the treatment of most ILDs are not known. A common starting dose is prednisone, 0.5–1 mg/kg in a once-daily oral dose (based on the patient's lean body weight). This dose is continued for 4–12 weeks, at which time the patient is reevaluated. If the patient is stable or improved, the dose is tapered to 0.25–0.5 mg/kg and is maintained at this level for an additional 4–12 weeks, depending on the course. Rapid tapering or a shortened course of glucocorticoid treatment can result in recurrence. If the patient's condition continues to decline on glucocorticoids, a second agent (see below) often is added and the prednisone dose is lowered to or maintained at 0.25 mg/kg per d.

Cyclophosphamide and azathioprine (1–2 mg/kg lean body weight per day), with or without glucocorticoids, have been tried with variable success in IPF, vasculitis, progressive systemic sclerosis, and other ILDs. An objective response usually requires at least 8–12 weeks to occur. In situations in which these drugs have failed or could not be tolerated, other agents, including methotrexate, colchicine, penicillamine, and cyclosporine, have been tried. However, their role in the treatment of ILDs remains to be determined.

Many cases of ILD are chronic and irreversible despite the therapy discussed above, and lung transplantation may then be considered (Chap. 266).

INDIVIDUAL FORMS OF ILD

Idiopathic Pulmonary Fibrosis

IPF is the most common form of idiopathic interstitial pneumonia. Separating IPF from other forms of lung fibrosis is an important step in the evaluation of all patients presenting with ILD. IPF has a distinctly poor response to therapy and a bad prognosis.

CLINICAL MANIFESTATIONS

Exertional dyspnea, a nonproductive cough, and inspiratory crackles with or without digital clubbing may be present on physical examination. HRCT lung scans typically show patchy, predominantly basilar, subpleural reticular opacities, often associated with traction bronchiectasis and honeycombing (Fig. 261-3). Atypical findings that should suggest an alternative diagnosis include extensive ground-

glass abnormality, nodular opacities, upper or midzone predominance, and prominent hilar or mediastinal lymphadenopathy. Pulmonary function tests often reveal a restrictive pattern, a reduced $DlCO$, and arterial hypoxemia that is exaggerated or elicited by exercise.

HISTOLOGIC FINDINGS

Confirmation of the presence of the UIP pattern on histologic examination is essential to confirm this diagnosis. Transbronchial biopsies are not helpful in making the diagnosis of UIP, and surgical biopsy usually is required. The histologic hallmark and chief diagnostic criterion of UIP is a heterogeneous appearance at low magnification with alternating areas of normal lung, interstitial inflammation, foci of proliferating fibroblasts, dense collagenfibrosis, and honeycomb changes. These histologic changes affect the peripheral, subpleural parenchyma most severely. The interstitial inflammation is usually patchy and consists of a lymphoplasmacytic infiltrate in the alveolar septa, associated with hyperplasia of type 2 pneumocytes. The fibrotic zones are composed mainly of dense collagen, although scattered foci of proliferating fibroblasts are a consistent finding. The extent of fibroblastic proliferation is predictive of disease progression. Areas of honeycomb change are composed of cystic fibrotic air spaces that frequently are lined by bronchiolar epithelium and filled with mucin. Smooth-muscle hyperplasia is commonly seen in areas of fibrosis and honeycomb change. A fibrotic pattern with some features similar to UIP may be found in the chronic stage of several specific disorders, such as pneumoconioses (e.g., asbestosis), radiation injury, certain drug-induced lung diseases (e.g., nitrofurantoin), chronic aspiration, sarcoidosis, chronic hypersensitivity pneumonitis, organized chronic eosinophilic pneumonia, and PLCH. Commonly, other histopathologic features are present in these situations, thus allowing separation of these lesions from the UIP-like pattern. Consequently, the term *usual interstitial pneumonia* is used for patients in whom the lesion is idiopathic and not associated with another condition.

Treatment: Management Issues in Patients with Ipf

Untreated patients with IPF show continued progression of their disease and have a high mortality rate. There is no effective therapy for IPF. Chronic microaspiration secondary to gastroesophageal reflux may play a role in the pathogenesis and natural history of IPF. Patients with IPF and coexisting emphysema [combined pulmonary fibrosis and emphysema (CPFE)] are more likely to require long-term oxygen therapy and develop pulmonary hypertension and may have a more dismal outcome than those without emphysema.

Patients with IPF may have acute deterioration secondary to infections, pulmonary embolism, or pneumothorax. Heart failure and ischemic heart disease are common problems in patients with IPF, accounting for nearly one-third of deaths. These patients also commonly experience an accelerated phase of rapid clinical decline that is associated with a poor prognosis (so-called acute exacerbations of IPF). These acute exacerbations are defined by worsening of dyspnea within a few days to 4 weeks; newly developing diffuse ground-glass abnormality and/or consolidation superimposed on a background reticular or honeycomb pattern consistent with the UIP pattern; worsening hypoxemia; and absence of infectious pneumonia, heart failure, and sepsis. The rate of these acute exacerbations ranges from 10–57%, apparently depending on the length of follow-up. During these episodes, the histopathologic pattern of diffuse alveolar damage is often found on the background of UIP. No therapy has been found to be effective in the management of acute exacerbations of IPF. Often mechanical ventilation is required, but it is usually not successful, with a hospital mortality rate of up to three-fourths of patients. In those who survive, a recurrence of acute exacerbation is common and usually results in death at those times.

Lung transplantation should be considered for patients who experience progressive deterioration despite optimal medical management and who meet the established criteria (Chap. 266).

Nonspecific Interstitial Pneumonia

This condition defines a subgroup of the idiopathic interstitial pneumonias that can be distinguished clinically and pathologically from UIP, DIP, AIP, and idiopathic BOOP. Importantly, many cases with this histopathologic pattern occur in the context of an underlying disorder, such as a connective tissue disease, drug-induced ILD, or chronic hypersensitivity pneumonitis.

Patients with idiopathic NSIP have clinical, serologic, radiographic, and pathologic characteristics highly suggestive of autoimmune disease and meet the criteria for undifferentiated connective tissue disease. Idiopathic NSIP is a subacute restrictive process with a presentation similar to that of IPF but usually at a younger age, most commonly in women who have never smoked. It is often associated with a febrile illness. HRCT shows bilateral, subpleural ground-glass opacities, often associated with lower lobe volume loss **(Fig. 261-4)**. Patchy areas of airspace consolidation and reticular abnormalities may be present, but honeycombing is unusual. The key histopathologic feature of NSIP is the uniformity of interstitial involvement across the biopsy section, and this may be predominantly cellular or fibrosing. There is less temporal and spatial heterogeneity than in UIP, and little or no honeycombing is found. The cellular variant is rare. Unlike patients with IPF (UIP), the majority of patients with NSIP have a good prognosis (5-year mortality rate estimated at <15%), with most showing improvement after treatment with glucocorticoids, often used in combination with azathioprine.

Figure 261-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nonspecific interstitial pneumonia. High-resolution CT through the lower lung shows volume loss with extensive ground-glass abnormality, reticular abnormality, and traction bronchiectasis. There is sparing on the lung immediately adjacent to the pleura. Histology showed a combination of inflammation and mild fibrosis.

Acute Interstitial Pneumonia (Hamman-Rich Syndrome)

AIP is a rare, fulminant form of lung injury characterized histologically by diffuse alveolar damage on lung biopsy. Most patients are older than 40 years. AIP is similar in presentation to the acute respiratory distress syndrome (ARDS) (Chap. 268) and probably corresponds to the subset of cases of idiopathic ARDS. The onset is usually abrupt in a previously healthy individual. A prodromal illness, usually lasting 7–14 days before presentation, is common. Fever, cough, and dyspnea are common manifestations at presentation. Diffuse, bilateral, air-space opacification is present on the chest radiograph. HRCT scans show bilateral, patchy, symmetric areas of ground-glass attenuation. Bilateral areas of air-space consolidation also may be present. A predominantly subpleural distribution may be seen. The diagnosis of AIP requires the presence of a clinical syndrome of idiopathic ARDS and pathologic confirmation of organizing diffuse alveolar damage. Therefore, lung biopsy is required to confirm the diagnosis. Most patients have moderate to severe hypoxemia and develop respiratory failure. Mechanical ventilation is often required. The mortality rate is high (>60%), with most patients dying within 6 months of presentation. Recurrences have been reported. However, those who recover often have substantial improvement in lung function. The main treatment is supportive. It is not clear that glucocorticoid therapy is effective.

Cryptogenic Organizing Pneumonia

COP is a clinicopathologic syndrome of unknown etiology. The onset is usually in the fifth and sixth decades. The presentation may be of a flulike illness with cough, fever, malaise, fatigue, and weight loss. Inspiratory crackles are frequently present on examination. Pulmonary function is usually impaired, with a restrictive defect and arterial hypoxemia being most common. The roentgenographic manifestations are distinctive, revealing bilateral, patchy, or diffuse alveolar opacities in the presence of normal lung volume. Recurrent and migratory pulmonary opacities are common. HRCT shows areas of air-space consolidation, ground-glass opacities, small nodular opacities, and bronchial wall thickening and dilation. These changes occur more frequently in the periphery of the lung and in the lower lung zone. Lung biopsy shows granulation tissue within small airways, alveolar ducts, and airspaces, with chronic inflammation in the surrounding alveoli. Glucocorticoid therapy induces clinical recovery in two-thirds of patients. A few patients have rapidly progressive courses with fatal outcomes despite glucocorticoids.

Foci of organizing pneumonia is a nonspecific reaction to lung injury found adjacent to other pathologic processes or as a component of other primary pulmonary disorders [e.g., cryptococcosis, granulomatosis with polyangiitis (Wegener's), lymphoma, hypersensitivity pneumonitis, and eosinophilic pneumonia]. Consequently, the clinician must carefully reevaluate any patient found to have this histopathologic lesion to rule out these possibilities.

Ild Associated with Cigarette Smoking

DESQUAMATIVE INTERSTITIAL PNEUMONIA

DIP is a rare but distinct clinical and pathologic entity found almost exclusively in cigarette smokers. The histologic hallmark is the extensive accumulation of macrophages in intraalveolar spaces with minimal interstitial fibrosis. The peak incidence is in the fourth and fifth decades. Most patients present with dyspnea and cough. Lung function testing shows a restrictive pattern with reduced DI_{CO} and arterial hypoxemia. The chest x-ray and HRCT scans usually show diffuse hazy opacities. Clinical recognition of DIP is important because the process is associated with a better prognosis (10-year survival rate is ~70%) in response to smoking cessation. There are no clear data showing that systemic glucocorticoids are effective in DIP.

RESPIRATORY BRONCHIOLITIS–ASSOCIATED ILD

Respiratory bronchiolitis–associated ILD (RB-ILD) is considered to be a subset of DIP and is characterized by the accumulation of macrophages in peribronchial alveoli. The clinical presentation is similar to that of DIP. Crackles are often heard on chest examination and occur throughout inspiration; sometimes they continue into expiration. The process is best seen on HRCT lung scanning, which shows bronchial wall thickening, centrilobularnodules, ground-glass opacity, and emphysema with air trapping. RB-ILD appears to resolve in most patients after smoking cessation alone.

PULMONARY LANGERHANS CELL HISTIOCYTOSIS

This is a rare, smoking-related, diffuse lung disease that primarily affects men between the ages of 20 and 40 years. The clinical presentation varies from an asymptomatic state to a rapidly progressive condition. The most common clinical manifestations at presentation are cough, dyspnea, chest pain, weight loss, and fever. Pneumothorax occurs in ~25% of patients. Hemoptysis and diabetes insipidus are rare manifestations. The radiographic features vary with the stage of the disease. The combination of ill-defined or stellate nodules (2–10 mm in diameter), reticular or nodular opacities, bizarre-shaped upper zone cysts, preservation of lung volume, and sparing of the costophrenic angles are characteristics of PLCH. HRCT that reveals a combination of nodules and thin-walled cysts is virtually diagnostic of PLCH. The most common pulmonary function abnormality is a markedly reduced DI_{CO}, although varying degrees of restrictive disease, airflow limitation, and diminished exercise capacity may occur. The characteristic histopathologic finding in PLCH is the presence of nodular sclerosing lesions that contain Langerhans cells accompanied by mixed cellular infiltrates. The nodular lesions are poorly defined and are distributed in a bronchiolocentric fashion with intervening normal lung parenchyma. As the disease advances, fibrosis progresses to involve adjacent lung tissue, leading to pericicatricial air space enlargement, which accounts for the concomitant cystic changes. Discontinuance of smoking is the key treatment, resulting in clinical improvement in one-third of patients. Most patients with PLCH experience persistent or progressive disease. Death due to respiratory failure occurs in ~10% of patients.

Ild Associated with Connective Tissue Disorders

Clinical findings suggestive of a CTD (musculoskeletal pain, weakness, fatigue, fever, joint pain or swelling, photosensitivity, Raynaud's phenomenon, pleuritis, dry eyes, dry mouth) should be sought in any patient with ILD. The CTDs may be difficult to rule out since the pulmonary manifestations occasionally precede the more typical systemic manifestations by months or years. The most common form of pulmonary involvement is the nonspecific interstitial pneumonia histopathologic pattern. However, determining the precise nature of lung involvement in most of the CTDs is difficult due to the high incidence of lung involvement caused by disease-associated complications of esophageal dysfunction (predisposing to aspiration and secondary infections), respiratory muscle weakness (atelectasis and secondary infections), complications of therapy (opportunistic infections), and associated malignancies.

PROGRESSIVE SYSTEMIC SCLEROSIS (PSS)

(See also Chap. 323) Clinical evidence of ILD is present in about one-half of patients with PSS, and pathologic evidence in three-quarters. Pulmonary function tests show a restrictive pattern and impaired diffusing capacity, often before any clinical or radiographic evidence of lung disease appears. Pulmonary vascular disease alone or in association with pulmonary fibrosis, pleuritis, or recurrent aspiration pneumonitis is strikingly resistant to current modes of therapy.

RHEUMATOID ARTHRITIS

(See also Chap. 321) ILD associated with RA is more common in men. Pulmonary manifestations of RA include pleurisy with or without effusion, ILD in up to 20% of cases, necrobiotic nodules (nonpneumoconiotic intrapulmonary rheumatoid nodules) with or without cavities, Caplan's syndrome (rheumatoid pneumoconiosis), pulmonary hypertension secondary to rheumatoid pulmonary vasculitis, organized pneumonia, and upper airway obstruction due to crico-arytenoid arthritis.

SYSTEMIC LUPUS ERYTHEMATOSUS

(See also Chap. 319) Lung disease is a common complication in SLE. Pleuritis with or without effusion is the most common pulmonary manifestation. Other lung manifestations include the following:

atelectasis, diaphragmatic dysfunction with loss of lung volumes, pulmonary vascular disease, pulmonary hemorrhage, uremic pulmonary edema, infectious pneumonia, and organized pneumonia. Acute lupus pneumonitis characterized by pulmonary capillaritis leading to alveolar hemorrhage is uncommon. Chronic, progressive ILD is uncommon. It is important to exclude pulmonary infection. Although pleuropulmonary involvement may not be evident clinically, pulmonary function testing, particularly DI_{CO}, reveals abnormalities in many patients with SLE.

POLYMYOSITIS AND DERMATOMYOSITIS (PM/DM)

(See also Chap. 388) ILD occurs in ~10% of patients with PM/DM. Diffuse reticular or nodular opacities with or without an alveolar component occur radiographically, with a predilection for the lung bases. ILD occurs more commonly in the subgroup of patients with an anti-Jo-1 antibody that is directed to histidyl tRNA synthetase. Weakness of respiratory muscles contributing to aspiration pneumonia may be present. A rapidly progressive illness characterized by diffuse alveolar damage may cause respiratory failure.

SJÖGREN'S SYNDROME

(See also Chap. 324) General dryness and lack of airway secretion cause the major problems of hoarseness, cough, and bronchitis. Lymphocytic interstitial pneumonitis, lymphoma, pseudolymphoma, bronchiolitis, and bronchiolitis obliterans are associated with this condition. Lung biopsy is frequently required to establish a precise pulmonary diagnosis. Glucocorticoids have been used in the management of ILD associated with Sjögren's syndrome with some degree of clinical success.

Drug-Induced Ild

Many classes of drugs have the potential to induce diffuse ILD, which is manifest most commonly as exertional dyspnea and nonproductive cough. A detailed history of the medications taken by the patient is needed to identify drug-induced disease, including over-the-counter medications, oily nose drops, and petroleum products (mineral oil). In most cases, the pathogenesis is unknown, although a combination of direct toxic effects of the drug (or its metabolite) and indirect inflammatory and immunologic events are likely. The onset of the illness may be abrupt and fulminant, or it may be insidious, extending over weeks to months. The drug may have been taken for several years before a reaction develops (e.g., amiodarone), or the lung disease may occur weeks to years after the drug has been discontinued (e.g., carmustine). The extent and severity of disease are usually dose-related. Treatment consists of discontinuation of any possible offending drug and supportive care.

Eosinophilic Pneumonia

(See Chap. 255)

Pulmonary Alveolar Proteinosis (Pap)

Although not strictly an ILD, PAP resembles and is therefore considered with these conditions. It has been proposed that a defect in macrophage function, more specifically an impaired ability to process surfactant, may play a role in the pathogenesis of PAP. This diffuse disease is characterized by the accumulation of an amorphous, periodic acid–Schiff-positive lipoproteinaceous material in the distal air spaces. There is little or no lung inflammation, and the underlying lung architecture is preserved. PAP is an autoimmune disease with a neutralizing antibody of immunoglobulin G isotype against granulocyte-macrophage colony-stimulating factor (GM-CSF). These findings suggest that neutralization of GM-CSF bioactivity by the antibody causes dysfunction of alveolar macrophages, which results in reduced surfactant clearance. There are three distinct classes of PAP: acquired (>90% of all cases), congenital, and secondary. *Congenital PAP* is transmitted in an autosomal recessive manner and is caused by homozygosity for a frameshift mutation (121ins2) in the *SP-B* gene, which leads to an unstable SP-B mRNA, reduced protein levels, and *secondary disturbances of SP-C processing. Secondary PAP* is rare among adults and is caused by lysinuric protein intolerance, acute silicosis and other inhalational syndromes, immunodeficiency disorders, and malignancies (almost exclusively of hematopoietic origin) and hematopoietic disorders.

The typical age of presentation is 30–50 years, and males predominate. The clinical presentation is usually insidious and is manifested by progressive exertional dyspnea, fatigue, weight loss, and low-grade fever. A nonproductive cough is common, but occasionally expectoration of "chunky" gelatinous material may occur. Polycythemia, hypergammaglobulinemia, and increased LDH levels are common. Markedly elevated serum levels of lung surfactant proteins A and D have been found in PAP. In the absence of any known secondary cause of PAP, an elevated serum anti-GM-CSF titer is highly sensitive and specific for the diagnosis of acquired PAP. BAL fluid levels of anti-GM-CSF antibodies correlate better with the severity of PAP than do serum titers. Radiographically, bilateral symmetric alveolar opacities located centrally in middle and lower lung zones result in a "bat-wing" distribution. HRCT shows a ground-glass opacification and thickened intralobular structures and interlobular septa. Whole-lung lavage(s) through a double-lumen endotracheal tube provides relief to many patients with dyspnea or progressive hypoxemia and also may provide long-term benefit.

Pulmonary Lymphangioleiomyomatosis

Pulmonary LAM is a rare condition that afflicts premenopausal women and should be suspected in young women with "emphysema," recurrent pneumothorax, or chylous pleural effusion. It is often misdiagnosed as asthma or chronic obstructive pulmonary disease. Pathologically, LAM is characterized by the proliferation of atypical pulmonary interstitial smooth muscle and cyst formation. The immature-appearing smooth-muscle cells react with monoclonal antibody HMB45, which recognizes a 100-kDa glycoprotein (gp100) originally found in human melanoma cells. Whites are affected much more commonly than are members of other racial groups. The disease accelerates during pregnancy and abates after oophorectomy. Common complaints at presentation are dyspnea, cough, and chest pain. Hemoptysis may be life threatening. Spontaneous pneumothorax occurs in 50% of patients; it may be bilateral and necessitate pleurodesis. Meningioma and renal angiomyolipomas (hamartomas), characteristic findings in the genetic disorder tuberous sclerosis, are also common in patients with LAM. Chylothorax, chyloperitoneum (chylous ascites), chyluria, and chylopericardium are other complications. Pulmonary function testing usually reveals an obstructive or mixed obstructive-restrictive pattern, and gas exchange is often abnormal. HRCT shows thin-walled cysts surrounded by normal lung without zonal predominance. Progression is common, with a median survival of 8–10 years from diagnosis. No therapy is of proven benefit in LAM. Progesterone (10 mg/d), luteinizing hormone–releasing hormone analogues, and sirolimus have been used. Oophorectomy is no longer recommended, and estrogen-containing drugs should be discontinued. Lung transplantation offers the only hope for cure despite reports of recurrent disease in the transplanted lung.

Syndromes of Ild with Diffuse Alveolar Hemorrhage

Injury to arterioles, venules, and the alveolar septal (alveolar wall or interstitial) capillaries can result in hemoptysis secondary to disruption of the alveolar-capillary basement membrane. This results in bleeding into the alveolar spaces, which characterizes DAH. Pulmonary capillaritis, characterized by a neutrophilic infiltration of the alveolar septae, may lead to necrosis of these structures, loss of capillary structural integrity, and the pouring of red blood cells into the alveolar space. Fibrinoid necrosis of the interstitium and red blood cells within the interstitial space are sometimes seen. Bland pulmonary hemorrhage (i.e., DAH without inflammation of the alveolar structures) also may occur.

The clinical onset is often abrupt, with cough, fever, and dyspnea. Severe respiratory distress requiring ventilatory support may be evident at initial presentation. Although hemoptysis is expected, it can be absent at the time of presentation in one-third of the cases. For patients without hemoptysis, new alveolar opacities, a falling hemoglobin level, and hemorrhagic BAL fluid point to the diagnosis. The chest radiograph is nonspecific and most commonly shows new patchy or diffuse alveolar opacities. Recurrent episodes of DAH may lead to pulmonary fibrosis, resulting in interstitial opacities on the chest radiograph. An elevated white blood cell count and falling hematocrit are common. Evidence for impaired renal function caused by focal segmental necrotizing glomerulonephritis, usually with crescent formation, also may be present.

Varying degrees of hypoxemia may occur and are often severe enough to require ventilatory support. DI_{CO} may be increased, resulting from the increased hemoglobin within the alveoli compartment.

Evaluation of either lung or renal tissue by immunofluorescent techniques indicates an absence of immune complexes (pauci-immune) in granulomatosis with polyangiitis (Wegener's), microscopic polyangiitis pauci-immune glomerulonephritis, and isolated pulmonary capillaritis. A granular pattern is found in the CTDs, particularly SLE, and a characteristic linear deposition is found in Goodpasture's syndrome. Granular deposition of IgA-containing immune complexes is present in Henoch-Schönlein purpura.

The mainstay of therapy for the DAH associated with systemic vasculitis, CTD, Goodpasture's syndrome, and isolated pulmonary capillaritis is IV methylprednisolone, 0.5–2 g daily in divided doses for up to 5 days, followed by a gradual tapering, and then maintenance on an oral preparation. Prompt initiation of therapy is important, particularly in the face of renal insufficiency, since early initiation of therapy has the best chance of preserving renal function. The decision to start other immunosuppressive therapy (cyclophosphamide or azathioprine) acutely depends on the severity of illness.

GOODPASTURE'S SYNDROME

Pulmonary hemorrhage and glomerulonephritis are features in most patients with this disease. Autoantibodies to renal glomerular and lung alveolar basement membranes are present. This syndrome can present and recur as DAH without an associated glomerulonephritis. In such cases, circulating anti-basement membrane antibody is often absent, and the only way to establish the diagnosis is by demonstrating linear immunofluorescence in lung tissue. The underlying histology may be bland hemorrhage or DAH associated with capillaritis. Plasmapheresis has been recommended as adjunctive treatment.

Inherited Disorders Associated with Ild

Pulmonary opacities and respiratory symptoms typical of ILD can develop in related family members and in several inherited diseases. These diseases include the phakomatoses, tuberous sclerosis and

neurofibromatosis (Chap. 379), and the lysosomal storage diseases, Niemann-Pick disease and Gaucher disease (Chap. 361). The Hermansky-Pudlak syndrome is an autosomal recessive disorder in which granulomatous colitis and ILD may occur. It is characterized by oculocutaneous albinism, bleeding diathesis secondary to platelet dysfunction, and the accumulation of a chromolipid, lipofuscin material in cells of the reticuloendothelial system. A fibrotic pattern is found on lung biopsy, but the alveolar macrophages may contain cytoplasmic ceroid-like inclusions.

Ild with a Granulomatous Response in Lung Tissue or Vascular Structures

Inhalation of organic dusts, which cause hypersensitivity pneumonitis, or of inorganic dust, such as silica, which elicits a granulomatous inflammatory reaction leading to ILD, produces diseases of known etiology (Table 261-1) that are discussed in Chaps. 255 and 256. Sarcoidosis (Chap. 329) is prominent among granulomatous diseases of unknown cause in which ILD is an important feature.

GRANULOMATOUS VASCULITIDES

(See also Chap. 326) The granulomatous vasculitides are characterized by pulmonary angiitis (i.e., inflammation and necrosis of blood vessels) with associated granuloma formation (i.e., infiltrates of lymphocytes, plasma cells, epithelioid cells, or histiocytes, with or without the presence of multinucleated giant cells, sometimes with tissue necrosis). The lungs are almost always involved, although any organ system may be affected. Granulomatosis with polyangiitis (Wegener's) and allergic angiitis and granulomatosis (Churg-Strauss syndrome) primarily affect the lung but are associated with a systemic vasculitis as well. The granulomatous vasculitides generally limited to the lung include necrotizing sarcoid granulomatosis and benign lymphocytic angiitis and granulomatosis. Granulomatous infection and pulmonary angiitis due to irritating embolic material (e.g., talc) are important known causes of pulmonary vasculitis.

Lymphocytic Infiltrative Disorders

This group of disorders features lymphocyte and plasma cell infiltration of the lung parenchyma. The disorders either are benign or can behave as low-grade lymphomas. Included are angioimmunoblastic lymphadenopathy with dysproteinemia, a rare lymphoproliferative disorder characterized by diffuse lymphadenopathy, fever, hepatosplenomegaly, and hemolytic anemia, with ILD in some cases.

LYMPHOCYTIC INTERSTITIAL PNEUMONITIS

This rare form of ILD occurs in adults, some of whom have an autoimmune disease or dysproteinemia. It has been reported in patients with Sjögren's syndrome and HIV infection.

LYMPHOMATOID GRANULOMATOSIS

This multisystem disorder of unknown etiology is an angiocentric malignant (T cell) lymphoma characterized by a polymorphic lymphoid infiltrate, an angiitis, and granulomatosis. Although it may affect virtually any organ, it is most frequently characterized by pulmonary, skin, and central nervous system involvement.

Bronchocentric Granulomatosis

Rather than a specific clinical entity, bronchocentric granulomatosis (BG) is a descriptive histologic term that is applied to an uncommon and nonspecific pathologic response to a variety of airway injuries. There is evidence that BG is caused by a hypersensitivity reaction to *Aspergillus* or other fungi in patients with asthma. About one-half of the patients described have had chronic asthma with severe wheezing and peripheral blood eosinophilia. In patients with asthma, BG probably represents one pathologic manifestation of allergic bronchopulmonary aspergillosis or another allergic mycosis. In patients without asthma, BG has been associated with RA and a variety of infections, including tuberculosis, echinococcosis, histoplasmosis, coccidioidomycosis, and nocardiosis. The chest roentgenogram reveals irregularly shaped nodular or mass lesions with ill-defined margins, which are usually unilateral and solitary, with upper lobe predominance. Glucocorticoids are the treatment of choice, often with an excellent outcome, although recurrences may occur as therapy is tapered or stopped.

Global Considerations

Limited epidemiologic data exist describing the prevalence or incidence of ILD in the general population. With a few exceptions, e.g., sarcoidosis and certain occupational and environmental exposures, there appear to be no significant differences in the prevalence or incidence of ILD among various populations. For sarcoidosis, there are important environmental, racial, and genetic differences (Chap. 329).

FURTHER READINGS

American Thoracic Society/European Respiratory Society: Idiopathic pulmonary fibrosis: Diagnosis and treatment: International consensus statement. Am J Respir Crit Care Med 161:646, 2000

———: International multidisciplinary consensus classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med 165:277, 2002

COLLARD HR et al: Acute exacerbations of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 176:636, 2007[PMID: 17585107] [Full Text]

EL-ZAMMAR OA, KATZENSTEIN AL: Pathological diagnosis of granulomatous lung disease: A review. Histopathology 50:289, 2007[PMID: 17257125] [Full Text]

KINDER BW et al: Idiopathic nonspecific interstitial pneumonia: Lung manifestation of undifferentiated connective tissue disease? Am J Respir Crit Care Med176:691, 2007[PMID: 17556720] [Full Text]

KING TE JR: Clinical advances in the diagnosis and therapy of the interstitial lung diseases. Am J Respir Crit Care Med 172:268, 2005[PMID: 15879420] [Full Text]

KLINGSBERG RC et al: Current clinical trials for the treatment of idiopathic pulmonary fibrosis. Respirology 15:19, 2010[PMID: 20199632] [Full Text]

KOTTMANN RM et al: Determinants of initiation and progression of idiopathic pulmonary fibrosis. Respirology 14:917, 2009[PMID: 19740254] [Full Text]

RAGHU G et al: An Official TS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. Am J Respir Crit Care Med 183:788, 2011[PMID: 21471066] [Full Text]

SELMAN M et al: Idiopathic pulmonary fibrosis: Prevailing and evolving hypotheses about its pathogenesis and implications for therapy. Ann Intern Med 134:136, 2001[PMID: 11177318] [Full Text]

SEYMOUR JF, PRESNEILL JJ: Pulmonary alveolar proteinosis: Progress in the first 44 years. Am J Respir Crit Care Med 166:215, 2002[PMID: 12119235] [Full Text]

TRAVIS WD et al: Idiopathic nonspecific interstitial pneumonia: Report of an American Thoracic Society project. Am J Respir Crit Care Med 177:1338, 2008[PMID: 18388353] [Full Text]

WELLS AU et al: Interstitial lung disease guideline. Thorax 63:v1, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEEP VENOUS THROMBOSIS AND PULMONARY THROMBOEMBOLISM: INTRODUCTION

Epidemiology

Venous thromboembolism (VTE), which encompasses deep venous thrombosis (DVT) and pulmonary embolism (PE), is one of the three major cardiovascular causes of death, along with myocardial infarction and stroke. VTE can cause death from PE or, among survivors, chronic thromboembolic pulmonary hypertension and postphlebitic syndrome. The U.S. Surgeon General has declared that PE is the most common preventable cause of death among hospitalized patients. Medicare has labeled PE and DVT occurring after total hip or knee replacement as unacceptable "never events" and no longer reimburses hospitals for the incremental expenses associated with treating this postoperative complication. New nonprofit organizations have begun educating health care professionals and the public on the medical consequences of VTE, along with risk factors and warning signs.

Between 100,000 and 300,000 VTE-related deaths occur annually in the United States. Mortality rates and length of hospital stay are decreasing as charges for hospital care increase. Approximately three of four symptomatic VTE events occur in the community, and the remainder are hospital acquired. Approximately 14 million (M) hospitalized patients are at moderate to high risk for VTE in the United States annually: 6 M major surgery patients and 8 M medical patients with comorbidities such as heart failure, cancer, and stroke. The prophylaxis paradigm has changed from voluntary to mandatory compliance with guidelines to prevent VTE among hospitalized patients. With an estimated 370,000 PE-related deaths annually in Europe, the projected direct cost for VTE-associated care exceeds 3 billion euros per year. In Japan, as the lifestyle becomes more westernized, the rate of VTE appears to be increasing.

The long-term effects of nonfatal VTE lower the quality of life. Chronic thromboembolic pulmonary hypertension is often disabling and causes breathlessness. A late effect of DVT is *postphlebitic syndrome*, which eventually occurs in more than one-half of DVT patients. Postphlebitic syndrome (also known as *postthrombotic syndrome* or *chronic venous insufficiency*) is a delayed complication of DVT that causes the venous valves of the leg to become incompetent and exude interstitial fluid. Patients complain of chronic ankle or calf swelling and leg aching, especially after prolonged standing. In its most severe form, postphlebitic syndrome causes skin ulceration, especially in the medial malleolus of the leg. There is no effective medical therapy for this condition.

PROTHROMBOTIC STATES

Thrombophilia contributes to the risk of venous thrombosis. The two most common autosomal dominant genetic mutations are factor V Leiden, which causes resistance to activated protein C (which inactivates clotting factors V and VIII), and the prothrombin gene mutation, which increases the plasma prothrombin concentration (Chaps. 58 and 117). Antithrombin, protein C, and protein S are naturally occurring coagulation inhibitors. Deficiencies of these inhibitors are associated with VTE but are rare. Hyperhomocysteinemia can increase the risk of VTE, but lowering the homocysteine level with folate, vitamin B6, or vitamin B12 does not reduce the incidence of VTE. Antiphospholipid antibody syndrome is the most common acquired cause of thrombophilia and is associated with venous or arterial thrombosis. Other common predisposing factors include cancer, systemic arterial hypertension, chronic obstructive pulmonary disease, long-haul air travel, air pollution, obesity, cigarette smoking, eating large amounts of red meat, oral contraceptives, pregnancy, postmenopausal hormone replacement, surgery, and trauma.

Pathophysiology

EMBOLIZATION

When venous thrombi are dislodged from their site of formation, they embolize to the pulmonary arterial circulation or, paradoxically, to the arterial circulation through a patent foramen ovale or atrial septal defect. About one-half of patients with pelvic vein thrombosis or proximal leg DVT develop PE, which is often asymptomatic. Isolated calf vein thrombi pose a much lower risk of PE but are the most common source of paradoxical embolism. These tiny thrombi can traverse a patent foramen ovale or atrial septal defect, unlike larger, more proximal leg thrombi. With increased use of chronic indwelling central venous catheters for hyperalimentation and chemotherapy, as well as more frequent insertion of permanent pacemakers and internal cardiac defibrillators, upper extremity venous thrombosis is becoming a more common problem. These thrombi rarely embolize and cause PE.

PHYSIOLOGY

The most common gas exchange abnormalities are hypoxemia (decreased arterial PO_2) and an increased alveolar-arterial O_2 tension gradient, which represents the inefficiency of O_2 transfer across the lungs. Anatomic dead space increases because breathed gas does not enter gas exchange units of the lung. Physiologic dead space increases because ventilation to gas exchange units exceeds venous blood flow through the pulmonary capillaries.

Other pathophysiologic abnormalities include the following:

1. *Increased pulmonary vascular resistance* due to vascular obstruction or platelet secretion of vasoconstricting neurohumoral agents such as serotonin. Release of vasoactive mediators can produce ventilation-perfusion mismatching at sites remote from the embolus, thereby accounting for a potential discordance between a small PE and a large alveolar-arterial O_2 gradient.
2. *Impaired gas exchange* due to increased alveolar dead space from vascular obstruction, hypoxemia from alveolar hypoventilation relative to perfusion in the nonobstructed lung, right-to-left shunting, and impaired carbon monoxide transfer due to loss of gas exchange surface.
3. *Alveolar hyperventilation* due to reflex stimulation of irritant receptors.
4. *Increased airway resistance* due to constriction of airways distal to the bronchi.
5. *Decreased pulmonary compliance* due to lung edema, lung hemorrhage, or loss of surfactant.

RIGHT-VENTRICULAR (RV) DYSFUNCTION

Progressive right heart failure is the usual cause of death from PE. As pulmonary vascular resistance increases, RV wall tension rises and causes further RV dilation and dysfunction. RV contraction continues even after the left ventricle (LV) starts relaxing at end-systole. Consequently, the interventricular septum bulges into and compresses an intrinsically normal left ventricle. Diastolic LV impairment develops, attributable to septal displacement, and results in reduced LV distensibility and impaired LV filling during diastole. Increased RV wall tension also compresses the right coronary artery, diminishes subendocardial perfusion, limits myocardial oxygen supply, and may precipitate myocardial ischemia and RV infarction. Underfilling of the LV may lead to a fall in left-ventricular cardiac output and systemic arterial pressure, thereby provoking myocardial ischemia due to compromised coronary artery perfusion. Eventually, circulatory collapse and death may ensue.

Diagnosis

CLINICAL EVALUATION

VTE mimics other illnesses, and PE is known as "the Great Masquerader," making diagnosis difficult. Occult PE is especially hard to detect when it occurs concomitantly with overt heart failure or pneumonia. In such circumstances, clinical improvement often fails to occur despite standard medical treatment of the concomitant illness. This scenario is a clinical clue to the possible coexistence of PE.

For patients who have DVT, the most common history is a cramp in the lower calf that persists for several days and becomes more uncomfortable as time progresses. For patients who have PE, the most common history is unexplained breathlessness.

In evaluating patients with possible VTE, the initial task is to decide on the clinical likelihood of the disorder. Patients with a low likelihood of DVT or a low-to-moderate likelihood of PE can undergo initial diagnostic evaluation with d-dimer testing alone (see "Blood tests") without obligatory imaging tests **(Fig. 262-1)**. If the d-dimer is abnormally elevated, imaging tests are the next step.

Figure 262-1

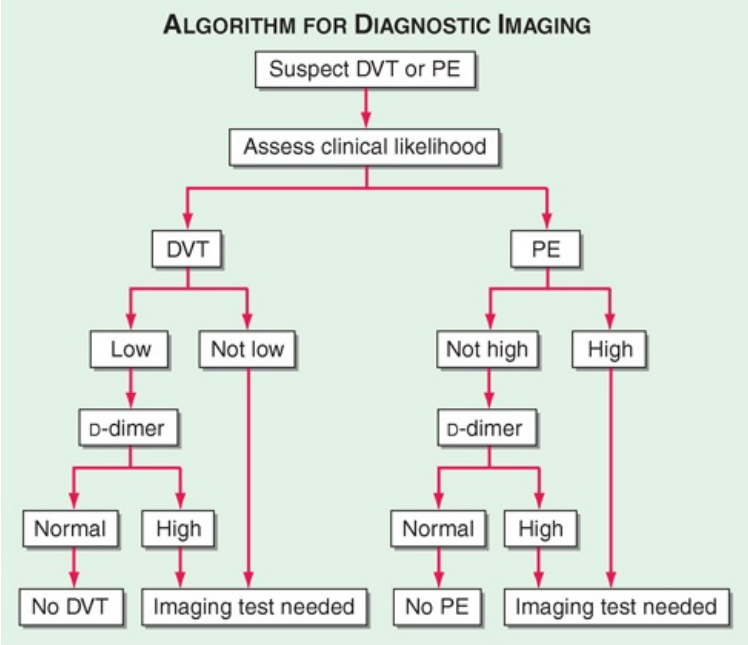

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

How to decide whether diagnostic imaging is needed. For assessment of clinical likelihood, see Table 262-1.

Point score methods are useful for estimating the clinical likelihood of DVT and PE **(Table 262-1).**

Table 262-1 Clinical Decision Rules

Low Clinical Likelihood of DVT if Point Score Is Zero or Less; Moderate-Likelihood Score Is 1 to 2; High-Likelihood Score Is 3 or Greater	
Clinical Variable	Score
Active cancer	1
Paralysis, paresis, or recent cast	1
Bedridden for >3 days; major surgery <12 weeks	1
Tenderness along distribution of deep veins	1
Entire leg swelling	1
Unilateral calf swelling >3 cm	1
Pitting edema	1
Collateral superficial nonvaricose veins	1
Alternative diagnosis at least as likely as DVT	–2
High Clinical Likelihood of PE if Point Score Exceeds 4	
Clinical Variable	Score
Signs and symptoms of DVT	3.0
Alternative diagnosis less likely than PE	3.0
Heart rate >100/min	1.5
Immobilization >3 days; surgery within 4 weeks	1.5
Prior PE or DVT	1.5
Hemoptysis	1.0
Cancer	1.0

CLINICAL SYNDROMES

The differential diagnosis is critical because not all leg pain is due to DVT and not all dyspnea is due to PE **(Table 262-2)**. Sudden, severe calf discomfort suggests a ruptured Baker's cyst. Fever and chills usually herald cellulitis rather than DVT, though DVT may be present concomitantly. Physical findings, if present at all, may consist only of mild palpation discomfort in the lower calf. Massive DVT is much easier to recognize. The patient presents with marked thigh swelling and tenderness during palpation of the common femoral vein. In extreme cases, patients are unable to walk or may require a cane, crutches, or a walker.

Table 262-2 Differential Diagnosis
DVT
Ruptured Baker's cyst
Cellulitis
Postphlebitic syndrome/venous insufficiency
PE
Pneumonia, asthma, chronic obstructive pulmonary disease
Congestive heart failure
Pericarditis
Pleurisy: "viral syndrome," costochondritis, musculoskeletal discomfort
Rib fracture, pneumothorax
Acute coronary syndrome
Anxiety

If the leg is diffusely edematous, DVT is unlikely. More probable is an acute exacerbation of venous insufficiency due to postphlebitic syndrome. Upper extremity venous thrombosis may present with asymmetry in the supraclavicular fossa or in the circumference of the upper arms. A prominent superficial venous pattern may be evident on the anterior chest wall.

Patients with *massive PE* present with systemic arterial hypotension and usually have anatomically widespread thromboembolism. Those with *moderate to large PE* have RV hypokinesis on echocardiography but normal systemic arterial pressure. Patients with *small to moderate PE* have both normal right heart function and normal systemic arterial pressure. They have an excellent prognosis with adequate anticoagulation.

The presence of *pulmonary infarction* usually indicates a small PE but one that is exquisitely painful because it lodges peripherally, near the innervation of pleural nerves. Pleuritic chest pain is more common with small, peripheral emboli. However, larger, more central PEs can occur concomitantly with peripheral pulmonary infarction.

Nonthrombotic PE may be easily overlooked. Possible etiologies include fat embolism after pelvic or long bone fracture, tumor embolism, bone marrow, and air embolism. Cement embolism and bony fragment embolism can occur after total hip or knee replacement. Intravenous drug users may inject themselves with a wide array of substances that can embolize such as hair, talc, and cotton. *Amniotic fluid embolism* occurs when fetal membranes leak or tear at the placental margin. Pulmonary edema in this syndrome probably is due to alveolar capillary leakage.

Dyspnea is the most common symptom of PE, and tachypnea is the most common sign. Dyspnea, syncope, hypotension, or cyanosis indicates a massive PE, whereas pleuritic pain, cough, or hemoptysis often suggests a small embolism situated distally near the pleura. On physical examination, young and previously healthy individuals may appear anxious but otherwise seem well, even with an anatomically large PE. They may have dyspnea only with moderate exertion. They often lack "classic" signs such as tachycardia, low-grade fever, neck vein distention, and an accentuated pulmonic component of the second heart sound. Sometimes paradoxical bradycardia occurs.

NONIMAGING DIAGNOSTIC MODALITIES

Nonimaging tests are best utilized in combination with clinical likelihood assessment of DVT or PE (Fig. 262-1).

Blood Tests

The quantitative *plasma d-dimer enzyme-linked immunosorbent assay (ELISA)* rises in the presence of DVT or PE because of the breakdown of fibrin by plasmin. Elevation of d-dimer indicates endogenous although often clinically ineffective thrombolysis. The sensitivity of the d-dimer is >80% for DVT (including isolated calf DVT) and >95% for PE. The d-dimer is less sensitive for DVT than for PE because the DVT thrombus size is smaller. The d-dimer is a useful "rule out" test. More than 95% of patients with a normal (<500 ng/mL) d-dimer do not have PE. The d-dimer assay is not specific. Levels increase in patients with myocardial infarction, pneumonia, sepsis, cancer, and the postoperative state and those in the second or third trimester of pregnancy. Therefore, d-dimer rarely has a useful role among hospitalized patients, because levels are frequently elevated due to systemic illness.

Contrary to classic teaching, *arterial blood gases* lack diagnostic utility for PE, even though both PO_2

and P_{co2} often decrease. Among patients suspected of having PE, neither the room air arterial PO_2 nor calculation of the alveolar-arterial O_2 gradient can reliably differentiate or triage patients who actually have PE at angiography.

Elevated Cardiac Biomarkers

Serum troponin and plasma heart-type fatty acid–binding protein levels increase because of RV microinfarction. Myocardial stretch results in elevation of brain natriuretic peptide or NT-pro-brain natriuretic peptide. Elevated cardiac biomarkers predict an increase in major complications and mortality from PE.

Electrocardiogram

The most frequently cited abnormality, in addition to sinus tachycardia, is the S1Q3T3 sign: an S wave in lead I, a Q wave in lead III, and an inverted T wave in lead III (Chap. 228). This finding is relatively specific but insensitive. Perhaps the most common abnormality is T-wave inversion in leads V_1 to V_4.

NONINVASIVE IMAGING MODALITIES

Venous Ultrasonography

Ultrasonography of the deep venous system **(Table 262-3)** relies on loss of vein compressibility as the primary criterion for DVT. When a normal vein is imaged in cross-section, it readily collapses with gentle manual pressure from the ultrasound transducer. This creates the illusion of a "wink." With acute DVT, the vein loses its compressibility because of passive distention by acute thrombus. The diagnosis of acute DVT is even more secure when thrombus is directly visualized. It appears homogeneous and has low echogenicity **(Fig. 262-2)**. The vein itself often appears mildly dilated, and collateral channels may be absent.

Table 262-3 Ultrasonography of the Deep Leg Veins

Criteria for Establishing the Diagnosis of Acute DVT
Lack of vein compressibility (principal criterion)
Vein does not "wink" when gently compressed in cross-section
Failure to appose walls of vein due to passive distention
Direct Visualization of Thrombus
Homogeneous
Low echogenicity
Abnormal Doppler Flow Dynamics
Normal response: calf compression augments Doppler flow signal and confirms vein patency proximal and distal to Doppler
Abnormal response: flow blunted rather than augmented with calf compression

Figure 262-2

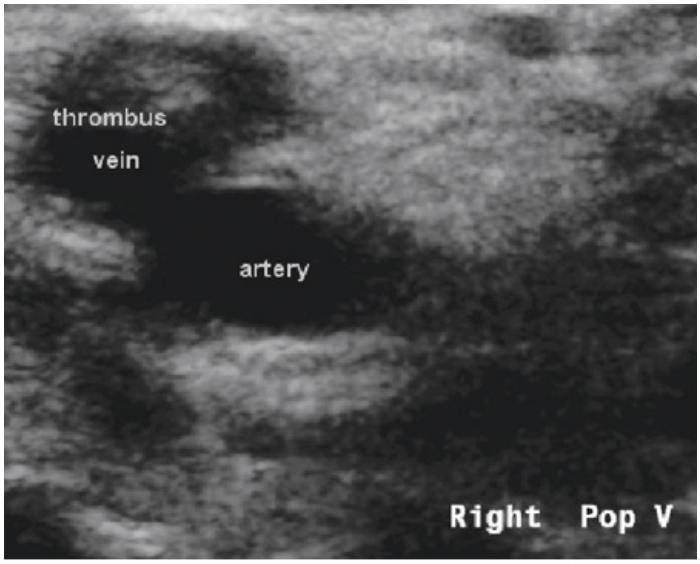

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute popliteal DVT on venous ultrasound in a 56-year-old man receiving chemotherapy for lung cancer.

Venous flow dynamics can be examined with Doppler imaging. Normally, manual calf compression causes augmentation of the Doppler flow pattern. Loss of normal respiratory variation is caused by an obstructing DVT or by any obstructive process within the pelvis. Because DVT and PE are so closely related and are both treated with anticoagulation (see "Treatment Deep Venous Thrombosis") confirmed DVT is usually an adequate surrogate for PE. In contrast, a normal venous ultrasound does not exclude PE. About one-half of patients with PE have no imaging evidence of DVT, probably because the clot already has embolized to the lung or is in the pelvic veins, where ultrasonography is usually inadequate. In patients without DVT, the ultrasound examination may identify other reasons for leg discomfort, such as a Baker's cyst (also known as a popliteal or synovial cyst) or a hematoma. For patients with a technically poor or nondiagnostic venous ultrasound, one should consider alternative imaging modalities for DVT, such as computed tomography (CT) and magnetic resonance imaging.

Chest Roentgenography

A normal or nearly normal chest x-ray often occurs in PE. Well-established abnormalities include focal oligemia (Westermark's sign), a peripheral wedged-shaped density above the diaphragm (Hampton's hump), and an enlarged right descending pulmonary artery (Palla's sign).

Chest CT

Computed tomography of the chest with intravenous contrast is the principal imaging test for the diagnosis of PE **(Fig. 262-3)**. Multidetector-row spiral CT acquires all chest images with ≤1 mm of resolution during a short breath hold. This generation of CT scanners can image small peripheral emboli. Sixth-order branches can be visualized with resolution superior to that of conventional invasive contrast pulmonary angiography. The CT scan also obtains excellent images of the RV and LV and can be used for risk stratification along with its use as a diagnostic tool. In patients with PE, RV enlargement on chest CT indicates an increased likelihood of death within the next 30 days compared with PE patients who have normal RV size on chest CT. When imaging is continued below the chest to the knee, pelvic and proximal leg DVT also can be diagnosed by CT scanning. In patients without PE, the lung parenchymal images may establish alternative diagnoses not apparent on chest x-ray that explain the presenting symptoms and signs such as pneumonia, emphysema, pulmonary fibrosis, pulmonary mass, and aortic pathology. Sometimes asymptomatic early-stage lung cancer is diagnosed incidentally.

Figure 262-3

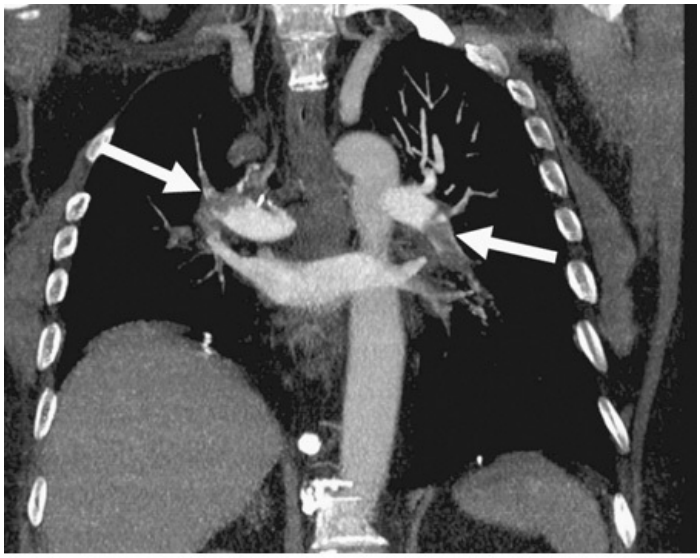

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Large bilateral proximal PE on a coronal chest CT image in a 54-year-old man with lung cancer and brain metastases. He had developed sudden onset of chest heaviness and shortness of breath while at home. There are filling defects in the main and segmental pulmonary arteries bilaterally (*white arrows*). Only the left upper lobe segmental artery is free of thrombus.

Lung Scanning

Lung scanning has become a second-line diagnostic test for PE, used mostly for patients who cannot tolerate intravenous contrast. Small particulate aggregates of albumin labeled with a gamma-emitting radionuclide are injected intravenously and are trapped in the pulmonary capillary bed. The perfusion scan defect indicates absent or decreased blood flow, possibly due to PE. Ventilation scans, obtained with a radiolabeled inhaled gas such as xenon or krypton, improve the specificity of the perfusion scan. Abnormal ventilation scans indicate abnormal nonventilated lung, thereby providing possible explanations for perfusion defects other than acute PE, such as asthma and chronic obstructive pulmonary disease. A high-probability scan for PE is defined as one that indicates two or more segmental perfusion defects in the presence of normal ventilation.

The diagnosis of PE is very unlikely in patients with normal and nearly normal scans but is about 90% certain in patients with high-probability scans. Unfortunately, most patients have nondiagnostic scans, and fewer than one-half of patients with angiographically confirmed PE have a high probability scan. As many as 40% of patients with high clinical suspicion for PE and "low-probability" scans do, in fact, have PE at angiography.

Magnetic Resonance (MR) (Contrast-Enhanced)

When ultrasound is equivocal, MR venography with gadolinium contrast is an excellent imaging modality to diagnose DVT. MR imaging should be considered for suspected VTE patients with renal insufficiency or contrast dye allergy. MR pulmonary angiography may detect large proximal PE but is not reliable for smaller segmental and subsegmental PE.

Echocardiography

Echocardiography is *not* a reliable diagnostic imaging tool for acute PE because most patients with PE have normal echocardiograms. However, echocardiography is a very useful diagnostic tool for detecting conditions that may mimic PE, such as acute myocardial infarction, pericardial tamponade, and aortic dissection.

Transthoracic echocardiography rarely images thrombus directly. The best-known indirect sign of PE on transthoracic echocardiography is McConnell's sign: hypokinesis of the RV free wall with normal motion of the RV apex.

One should consider transesophageal echocardiography when CT scanning facilities are not available or when a patient has renal failure or severe contrast allergy that precludes administration of contrast despite premedication with high-dose steroids. This imaging modality can identify saddle, right main, or left main PE.

INVASIVE DIAGNOSTIC MODALITIES

Pulmonary Angiography

Chest CT with contrast (see above) has virtually replaced invasive pulmonary angiography as a diagnostic test. Invasive catheter-based diagnostic testing is reserved for patients with technically

unsatisfactory chest CTs and those in whom an interventional procedure such as catheter-directed thrombolysis or embolectomy is planned. A definitive diagnosis of PE depends on visualization of an intraluminal filling defect in more than one projection. Secondary signs of PE include abrupt occlusion ("cut-off") of vessels, segmental oligemia or avascularity, a prolonged arterial phase with slow filling, and tortuous, tapering peripheral vessels.

Contrast Phlebography

Venous ultrasonography has virtually replaced contrast phlebography as the diagnostic test for suspected DVT.

INTEGRATED DIAGNOSTIC APPROACH

An integrated diagnostic approach (Fig. 262-1) streamlines the workup of suspected DVT and PE **(Fig. 262-4)**.

Figure 262-4

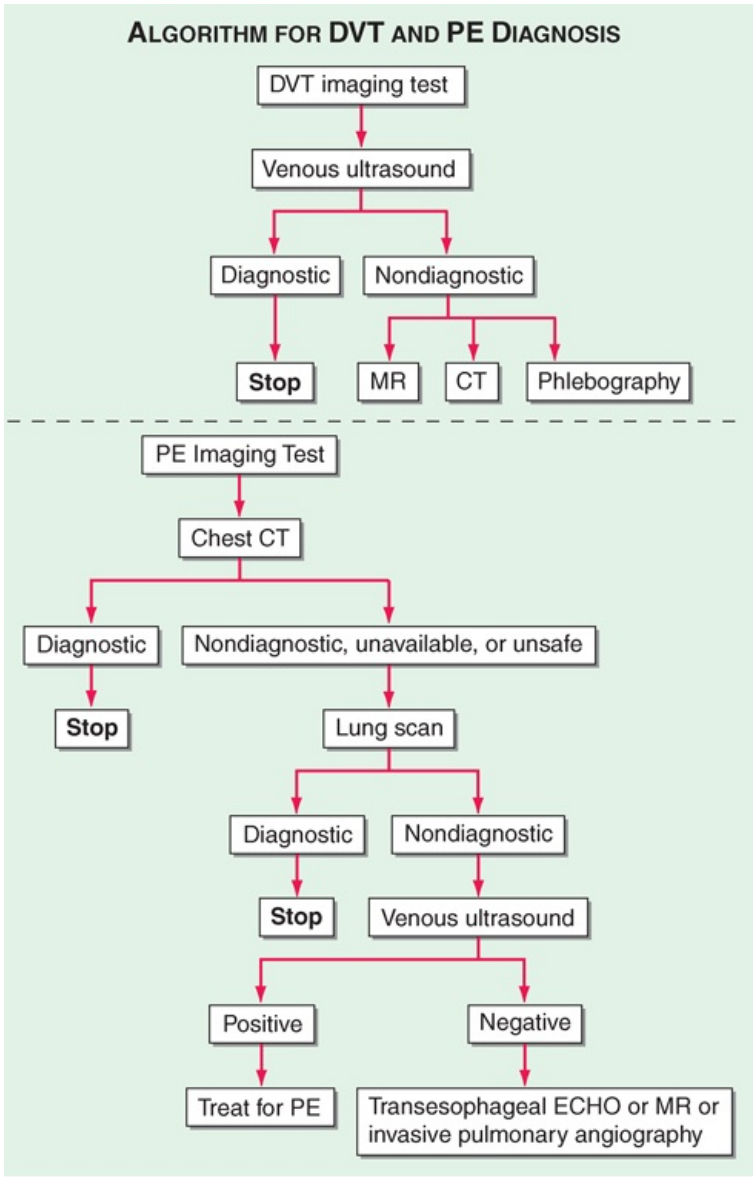

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Imaging tests to diagnose DVT and PE.

Treatment: Deep Venous Thrombosis

PRIMARY THERAPY VERSUS SECONDARY PREVENTION

Primary therapy consists of clot dissolution with thrombolysis or removal of PE by embolectomy. Anticoagulation with heparin and warfarin or placement of an inferior vena caval filter constitutes *secondary prevention* of recurrent PE rather than primary therapy.

RISK STRATIFICATION

Rapid and accurate risk stratification is critical in determining the optimal treatment strategy. The presence of hemodynamic instability, RV dysfunction, RV enlargement, or elevation of the troponin level due to RV microinfarction can identify high-risk patients. RV hypokinesis on echocardiography, RV enlargement on chest CT, and troponin elevation predict an increased mortality rate from PE.

Primary therapy should be reserved for patients at high risk of an adverse clinical outcome. When RV function remains normal in a hemodynamically stable patient, a good clinical outcome is highly likely with anticoagulation alone **(Fig. 262-5)**.

Figure 262-5

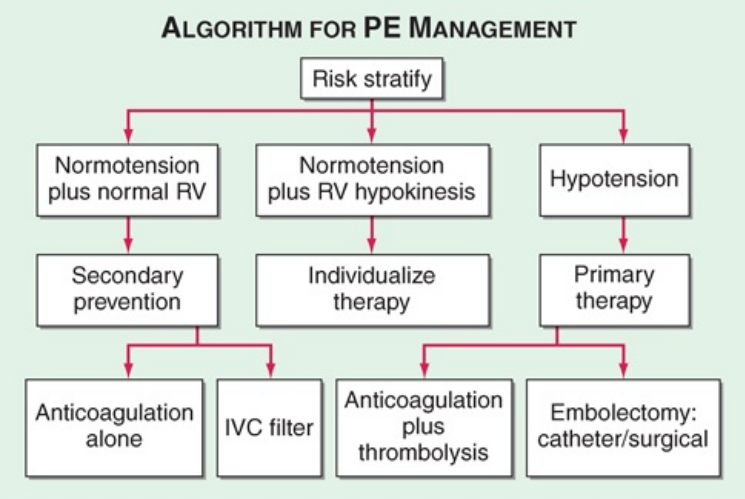

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute management of pulmonary thromboembolism. RV, right ventricular; IVC, inferior vena cava.

Treatment: Massive Pulmonary Embolism

ANTICOAGULATION

Anticoagulation is the foundation for successful treatment of DVT and PE **(Table 262-4)**. Immediately effective anticoagulation is initiated with a parenteral drug: unfractionated heparin (UFH), low-molecular-weight heparin (LMWH), or fondaparinux. One should use a direct thrombin inhibitor —argatroban, lepirudin, or bivalirudin—in patients with proven or suspected heparin-induced thrombocytopenia. Parenteral agents are continued as a transition or "bridge" to stable, long-term anticoagulation with a vitamin K antagonist (exclusively warfarin in the United States). Warfarin requires 5–7 days to achieve a therapeutic effect. During that period, one should overlap the parenteral and oral agents. After 5–7 days of anticoagulation, residual thrombus begins to endothelialize in the vein or pulmonary artery. However, anticoagulants do *not* directly dissolve thrombus that already exists.

Table 262-4 Anticoagulation of VTE

Immediate Parenteral Anticoagulation
Unfractionated heparin, bolus and continuous infusion, to achieve aPTT two to three times the upper limit of the laboratory normal, *or* Enoxaparin 1 mg/kg twice daily with normal renal function, *or* Dalteparin 200 U/kg once daily or 100 U/kg twice daily, with normal renal function, *or* Tinzaparin 175 U/kg once daily with normal renal function, *or* Fondaparinux weight-based once daily; adjust for impaired renal function
Warfarin Anticoagulation
Usual start dose is 5 mg Titrate to INR, target 2.0–3.0 Continue parenteral anticoagulation for a minimum of 5 days and until two sequential INR values, at least 1 day apart, achieve the target INR range.

Unfractionated Heparin

Unfractionated heparin anticoagulates by binding to and accelerating the activity of antithrombin, thus preventing additional thrombus formation and permitting endogenous fibrinolytic mechanisms to lyse clot that already has formed. UFH is dosed to achieve a target activated partial thromboplastin time (aPTT) that is 2–3 times the upper limit of the laboratory normal. This is usually equivalent to an aPTT of 60–80 s. For UFH, a typical intravenous bolus is 5000–10,000 units followed by a continuous infusion of 1000–1500 U/h. Nomograms based on a patient's weight may assist in adjusting the dose of heparin. The most popular nomogram utilizes an initial bolus of 80 U/kg, followed by an initial infusion rate of 18/kg per h.

The major advantage of UFH is its short half-life. This is especially useful if the patient may undergo an invasive procedure such as embolectomy. The major disadvantage of UFH is that achieving the target aPTT is empirical and may require repeated blood sampling and heparin dose adjustment every 4–6 hours. Furthermore, patients are at risk of developing heparin-induced thrombocytopenia.

Low-Molecular-Weight Heparins

These fragments of UFH exhibit less binding to plasma proteins and endothelial cells and consequently have greater bioavailability, a more predictable dose response, and a longer half-life than does UFH. No monitoring or dose adjustment is needed unless the patient is markedly obese or has chronic kidney disease.

There are two commonly used LMWH preparations in the United States: enoxaparin and dalteparin. *Enoxaparin* is approved as a bridge to warfarin for VTE. *Dalteparin* is also approved as monotherapy without warfarin for symptomatic VTE patients with cancer in a dose of 200 U/kg once daily for 30 days, followed by 150 U/kg once daily for months 2–6. These weight-adjusted LMWH doses must be reduced in patients with chronic kidney disease because the kidneys metabolize LMWH.

Fondaparinux

Fondaparinux, an anti-Xa pentasaccharide, is administered as a once-daily subcutaneous injection in a prefilled syringe to treat DVT and PE as a "bridge" to warfarin. No laboratory monitoring is required. Patients weighing <50 kg receive 5 mg, patients weighing 50–100 kg receive 7.5 mg, and patients weighing >100 kg receive 10 mg. Fondaparinux is synthesized in a laboratory and, unlike LMWH or UFH, is not derived from animal products. It does not cause heparin-induced thrombocytopenia. The dose must be adjusted downward for patients with renal dysfunction because the kidneys metabolize the drug.

Warfarin

This vitamin K antagonist prevents carboxylation activation of coagulation factors II, VII, IX, and X. The full effect of warfarin requires at least 5 days even if the prothrombin time, used for monitoring, becomes elevated more rapidly. If warfarin is initiated as monotherapy during an acute thrombotic illness, a paradoxical exacerbation of hypercoagulability can increase the likelihood of thrombosis rather than prevent it. Overlapping UFH, LMWH, or fondaparinux with warfarin for at least 5 days can counteract the early procoagulant effect of unopposed warfarin.

Warfarin Dosing

In an average-size adult, warfarin usually is initiated in a dose of 5 mg. Doses of 7.5 or 10 mg can be used in obese or large-framed young patients who are otherwise healthy. Patients who are malnourished or who have received prolonged courses of antibiotics are probably deficient in vitamin K and should receive smaller initial doses of warfarin, such as 2.5 mg. The prothrombin time is standardized by calculating the international normalized ratio (INR), which assesses the anticoagulant effect of warfarin (Chap. 58). The target INR is usually 2.5, with a range of 2.0–3.0.

The warfarin dose is titrated to achieve the target INR. Proper dosing is difficult because hundreds of drug-drug and drug-food interactions affect warfarin metabolism. Variables such as increasing age and comorbidities such as systemic illness reduce the required warfarin dose. Pharmacogenomics may provide more precise initial dosing of warfarin, especially for patients who require unusually large or small doses. *CYP2C9* variant alleles impair the hydroxylation of S-warfarin, thereby lowering the dose requirement. Variants in the gene encoding the vitamin K epoxide reductase complex 1 (*VKORC1*) can predict whether patients require low, moderate, or high warfarin doses. Nevertheless, more than half of warfarin dosing variability is caused by clinical factors such as age, sex, weight, concomitant drugs, and comorbid illnesses.

Nomograms have been developed (www.warfarindosing.org) to help clinicians initiate warfarin dosing based on clinical information and, if available, pharmacogenetic data. However, most practitioners utilize empirical dosing with an "educated guess." Centralized anticoagulation clinics have improved the efficacy and safety of warfarin dosing. Patients maintain a therapeutic INR more often if they self-monitor their INR with a home point-of-care fingerstick machine rather than obtaining a coagulation laboratory INR. The patient subgroup with the best results self-adjusts warfarin doses as well as self-tests INRs.

Novel Anticoagulants

Novel oral anticoagulants are administered in a fixed dose, establish effective anticoagulation within hours of administration, require no laboratory coagulation monitoring, and have few of the drug-drug or drug-food interactions that make warfarin so difficult to dose. Rivaroxaban, a factor Xa inhibitor, and dabigatran, a direct thrombin inhibitor, are approved in Canada and Europe for prevention of VTE after total hip and total knee replacement. In a large-scale trial of acute VTE treatment, dabigatran was as effective as warfarin and had less nonmajor bleeding. Because of these drugs' rapid onset of action and relatively short half-life compared with warfarin, "bridging" with a parenteral anticoagulant is not required.

Complications of Anticoagulants

The most serious adverse effect of anticoagulation is hemorrhage. For life-threatening or intracranial hemorrhage due to heparin or LMWH, protamine sulfate can be administered. There is no specific antidote for bleeding caused by fondaparinux or direct thrombin inhibitors.

Major bleeding from warfarin is best managed with prothrombin complex concentrate. With non-life threatening bleeding in a patient who can tolerate large volume, fresh-frozen plasma can be used. Recombinant human coagulation factor VIIa (rFVIIa), FDA-approved for bleeding in hemophiliacs, is an off-label option to manage catastrophic bleeding from warfarin. For minor bleeding or to manage an excessively high INR in the absence of bleeding, oral vitamin K may be administered.

Heparin-induced thrombocytopenia (HIT) and osteopenia are far less common with LMWH than with UFH. Thrombosis due to HIT should be managed with a direct thrombin inhibitor: argatroban for patients with renal insufficiency and lepirudin for patients with hepatic failure. In the setting of percutaneous coronary intervention, one should administer bivalirudin.

During pregnancy, warfarin should be avoided if possible because of warfarin embryopathy, which is most common with exposure during the sixth through twelfth week of gestation. However, women can take warfarin postpartum and breast-feed safely. Warfarin can also be administered safely during the second trimester.

Duration of Hospital Stay

Acute DVT patients with good family and social support, permanent residence, telephone service, and no hearing or language impairment often can be managed as outpatients. They, a family member, or a visiting nurse must administer a parenteral anticoagulant. Warfarin dosing can be titrated to the INR and adjusted on an outpatient basis.

Acute PE patients, who traditionally have required hospital stays of 5–7 days for intravenous heparin as a "bridge" to warfarin, can be considered for abbreviated hospitalization if they have a reliable support system at home and an excellent prognosis. Criteria include clinical stability, absence of chest pain or shortness of breath, normal RV size and function, and normal levels of cardiac biomarkers.

Duration of Anticoagulation

Patients with PE after surgery, trauma, or estrogen exposure (from oral contraceptives, pregnancy, or postmenopausal therapy) ordinarily have a low rate of recurrence after 3–6 months of anticoagulation. For DVT isolated to an upper extremity or calf that has been provoked by surgery, trauma, estrogen, or an indwelling central venous catheter or pacemaker, 3 months of anticoagulation suffices. For provoked proximal leg DVT or PE, 3 to 6 months of anticoagulation is sufficient. For patients with cancer and VTE, the consensus is to prescribe 3–6 months of LMWH as monotherapy without warfarin and to continue anticoagulation indefinitely unless the patient is rendered cancer-free. However, there is uncertainty whether subsequent anticoagulation should continue with LMWH or whether the patient should be placed on warfarin.

Among patients with idiopathic, unprovoked VTE, the recurrence rate is high after cessation of anticoagulation. VTE that occurs during long-haul air travel is considered unprovoked. It appears that unprovoked VTE is often a chronic illness, with latent periods between flares of recurrent episodes. American College of Chest Physicians (ACCP) guidelines recommend considering anticoagulation for an indefinite duration with a target INR between 2 and 3 for patients with idiopathic VTE. An alternative approach after the first 6 months of anticoagulation is to reduce the intensity of anticoagulation and to lower the target INR range to between 1.5 and 2.

Counterintuitively, the presence of genetic mutations such as heterozygous factor V Leiden and prothrombin gene mutation do not appear to increase the risk of recurrent VTE. However, patients with moderate or high levels of anticardiolipin antibodies probably warrant indefinite-duration anticoagulation even if the initial VTE was provoked by trauma or surgery.

INFERIOR VENA CAVAL (IVC) FILTERS

The two principal indications for insertion of an IVC filter are (1) active bleeding that precludes anticoagulation and (2) recurrent venous thrombosis despite intensive anticoagulation. Prevention of recurrent PE in patients with right heart failure who are not candidates for fibrinolysis and prophylaxis of extremely high-risk patients are "softer" indications for filter placement. The filter itself may fail by

permitting the passage of small- to medium-size clots. Large thrombi may embolize to the pulmonary arteries via collateral veins that develop. A more common complication is caval thrombosis with marked bilateral leg swelling.

Paradoxically, by providing a nidus for clot formation, filters double the DVT rate over the ensuing 2 years after placement. Retrievable filters can now be placed for patients with an anticipated temporary bleeding disorder or for patients at temporary high risk of PE, such as individuals undergoing bariatric surgery who have a prior history of perioperative PE. The filters can be retrieved up to several months after insertion unless thrombus forms and is trapped within the filter. The retrievable filter becomes permanent if it remains in place or if, for technical reasons such as rapid endothelialization, it cannot be removed.

MAINTAINING ADEQUATE CIRCULATION

For patients with massive PE and hypotension, one should administer 500 mL of normal saline. Additional fluid should be infused with extreme caution because excessive fluid administration exacerbates RV wall stress, causes more profound RV ischemia, and worsens LV compliance and filling by causing further interventricular septal shift toward the LV. Dopamine and dobutamine are first-line inotropic agents for treatment of PE-related shock. There should be a low threshold for initiating these pressors. Often, a "trial-and-error" approach works best; one should consider norepinephrine, vasopressin, or phenylephrine.

FIBRINOLYSIS

Successful fibrinolytic therapy rapidly reverses right heart failure and may result in a lower rate of death and recurrent PE by (1) dissolving much of the anatomically obstructing pulmonary arterial thrombus, (2) preventing the continued release of serotonin and other neurohumoral factors that exacerbate pulmonary hypertension, and (3) lysing much of the source of the thrombus in the pelvic or deep leg veins, thereby decreasing the likelihood of recurrent PE.

The preferred fibrinolytic regimen is 100 mg of recombinant tissue plasminogen activator (tPA) administered as a continuous peripheral intravenous infusion over 2 hours. Patients appear to respond to fibrinolysis for up to 14 days after the PE has occurred.

Contraindications to fibrinolysis include intracranial disease, recent surgery, and trauma. The overall major bleeding rate is about 10%, including a 1–3% risk of intracranial hemorrhage. Careful screening of patients for contraindications to fibrinolytic therapy (Chap. 245) is the best way to minimize bleeding risk.

The only FDA-approved indication for PE fibrinolysis is massive PE. For patients with preserved systolic blood pressure and submassive PE with moderate or severe RV dysfunction, ACCP guidelines for fibrinolysis recommend individual patient risk assessment of the thrombotic burden versus the bleeding risk.

PULMONARY EMBOLECTOMY

The risk of intracranial hemorrhage with fibrinolysis has prompted a renaissance of surgical embolectomy. More prompt referral before the onset of irreversible cardiogenic shock and multisystem organ failure and improved surgical technique have resulted in a high survival rate. A possible alternative to open surgical embolectomy is catheter embolectomy. New-generation catheters are under development.

PULMONARY THROMBOENDARTERECTOMY

Chronic thromboembolic pulmonary hypertension occurs in 2–4% of acute PE patients. Therefore, PE patients who have initial pulmonary hypertension (usually diagnosed with Doppler echocardiography) should be followed up at about 6 weeks with a repeat echocardiogram to determine whether pulmonary arterial pressure has normalized. Patients impaired by dyspnea due to chronic thromboembolic pulmonary hypertension should be considered for pulmonary thromboendarterectomy, which, if successful, can markedly reduce, and at times even cure, pulmonary hypertension (Chap. 250). The operation requires median sternotomy, cardiopulmonary bypass, deep hypothermia, and periods of hypothermic circulatory arrest. The mortality rate at experienced centers is approximately 5%.

EMOTIONAL SUPPORT

Patients with VTE may feel overwhelmed when they learn that they are susceptible to recurrent PE or DVT. They worry about the health of their families and the genetic implications of their illness. Those who are advised to discontinue warfarin after 3–6 months of therapy may feel especially vulnerable. At Brigham and Woman's Hospital a physican-nurse–facilitated PE support group has been maintained for patients and has met monthly for more than 15 years.

PREVENTION OF POSTPHLEBITIC SYNDROME

Daily use of below-knee 30- to 40-mmHg vascular compression stockings will halve the rate of developing postphlebitic syndrome. These stockings should be prescribed as soon as DVT is diagnosed and should be fitted carefully to maximize their benefit. When patients are in bed, the

stockings need not be worn.

Prevention of VTE

Prophylaxis **(Table 262-5)** is of paramount importance because VTE is difficult to detect and poses a profound medical and economic burden. Computerized reminder systems can increase the use of preventive measures and at Brigham and Women's Hospital have reduced the symptomatic VTE rate by more than 40%. Patients who have undergone total hip or knee replacement or cancer surgery will benefit from extended pharmacologic prophylaxis for a total of 4–5 weeks.

Table 262-5 Prevention of Venous Thromboembolism

Condition	Prophylaxis Strategy
High-risk general surgery	Mini-UFH *or* LMWH
Thoracic surgery	Mini-UFH + IPC
Cancer surgery, including gynecologic cancer surgery	LMWH, consider 1 month of prophylaxis
Total hip replacement, total knee replacement, hip fracture surgery	LMWH, fondaparinux (a pentasaccharide) 2.5 mg SC, once daily, *or* (except for total knee replacement) warfarin (target INR 2.5); rivaroxaban or dalteparin in countries where it is approved
Neurosurgery	IPC
Neurosurgery for brain tumor	Mini-UFH *or* LMWH, + IPC + predischarge venous ultrasonography
Benign gynecologic surgery	Mini-UFH
Medically ill patients	Mini-UFH *or* LMWH
Anticoagulation contraindicated	IPC
Long-haul air travel	Consider LMWH for very high-risk patients

Note: Mini-UFH, mini-dose unfractionated heparin, 5000 units subcutaneously twice (less effective) or three times daily (more effective); LMWH, low-molecular-weight heparin, typically in the United States enoxaparin, 40 mg once daily, or dalteparin, 2500 or 5000 units once daily; IPC, intermittent pneumatic compression devices.

FURTHER READINGS

AGENO W et al: Cardiovascular risk factors and venous thromboembolism: A meta-analysis. Circulation 117:93, 2008[PMID: 18086925] [Full Text]

DALEN JE: Should patients with venous thromboembolism be screened for thrombophilia? Am J Med 121:458, 2008[PMID: 18501222] [Full Text]

DENNIS M et al: Effectiveness of thigh-length graduated compression stockings to reduce the risk of deep vein thrombosis after stroke (CLOTS trial 1): A multicentre, randomised controlled trial. Lancet 373:1958, 2009[PMID: 19477503] [Full Text]

DENTALI F et al: Meta-analysis: Anticoagulant prophylaxis to prevent symptomatic venous thromboembolism in hospitalized medical patients. Ann Intern Med 146:278, 2007[PMID: 17310052] [Full Text]

GEERTS WH et al: Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th ed). Chest 133:381S, 2008

GLYNN RJ et al: A randomized trial of rosuvastatin in the prevention of venous thromboembolism. N Engl J Med 360:1851, 2009[PMID: 19329822] [Full Text]

KUCHER N, GOLDHABER SZ: Management of massive pulmonary embolism. Circulation 112:e28, 2005

PARK B et al: Recent trends in clinical outcomes and resource utilization for pulmonary embolism in the United States: Findings from the nationwide inpatient sample. Chest 136: 983, 2009[PMID: 19525357] [Full Text]

PIAZZA G, GOLDHABER SZ: The acutely decompensated right ventricle. Chest 128:1836, 2005[PMID: 16162794] [Full Text]

PRANDONI P et al: Residual thrombosis on ultrasonography to guide the duration of anticoagulation in patients with deep venous thrombosis: A randomized trial. Ann Intern Med 150:577, 2009[PMID: 19414836] [Full Text]

SPENCER FA et al: Venous thromboembolism in the outpatient setting. Arch Intern Med 167:1471, 2007[PMID: 17646600] [Full Text]

TODD JL, TAPSON VF: Thrombolytic therapy for acute pulmonary embolism: A critical appraisal. Chest 135:1321, 2009[PMID: 19420199] [Full Text]

TORBICKI A et al: Guidelines on the diagnosis and management of acute pulmonary embolism: The Task Force for the Diagnosis and Management of Acute Pulmonary Embolism of the European Society of Cardiology (ESC). Eur Heart J 29:2276, 2008[PMID: 18757870] [Full Text]

VAN BELLE A et al: Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. JAMA 295:172, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE PLEURA

Pleural Effusion

The pleural space lies between the lung and the chest wall and normally contains a very thin layer of fluid, which serves as a coupling system. A pleural effusion is present when there is an excess quantity of fluid in the pleural space.

ETIOLOGY

Pleural fluid accumulates when pleural fluid formation exceeds pleural fluid absorption. Normally, fluid enters the pleural space from the capillaries in the parietal pleura and is removed via the lymphatics in the parietal pleura. Fluid also can enter the pleural space from the interstitial spaces of the lung via the visceral pleura or from the peritoneal cavity via small holes in the diaphragm. The lymphatics have the capacity to absorb 20 times more fluid than is formed normally. Accordingly, a pleural effusion may develop when there is excess pleural fluid formation (from the interstitial spaces of the lung, the parietal pleura, or the peritoneal cavity) or when there is decreased fluid removal by the lymphatics.

DIAGNOSTIC APPROACH

When a patient is found to have a pleural effusion, an effort should be made to determine the cause **(Fig. 263-1)**. The first step is to determine whether the effusion is a transudate or an exudate. A *transudative pleural effusion* occurs when *systemic factors* that influence the formation and absorption of pleural fluid are altered. The leading causes of transudative pleural effusions in the United States are left-ventricular failure and cirrhosis. An *exudative pleural effusion* occurs when *local factors* that influence the formation and absorption of pleural fluid are altered. The leading causes of exudative pleural effusions are bacterial pneumonia, malignancy, viral infection, and pulmonary embolism. The primary reason for making this differentiation is that additional diagnostic procedures are indicated with exudative effusions to define the cause of the local disease.

Figure 263-1

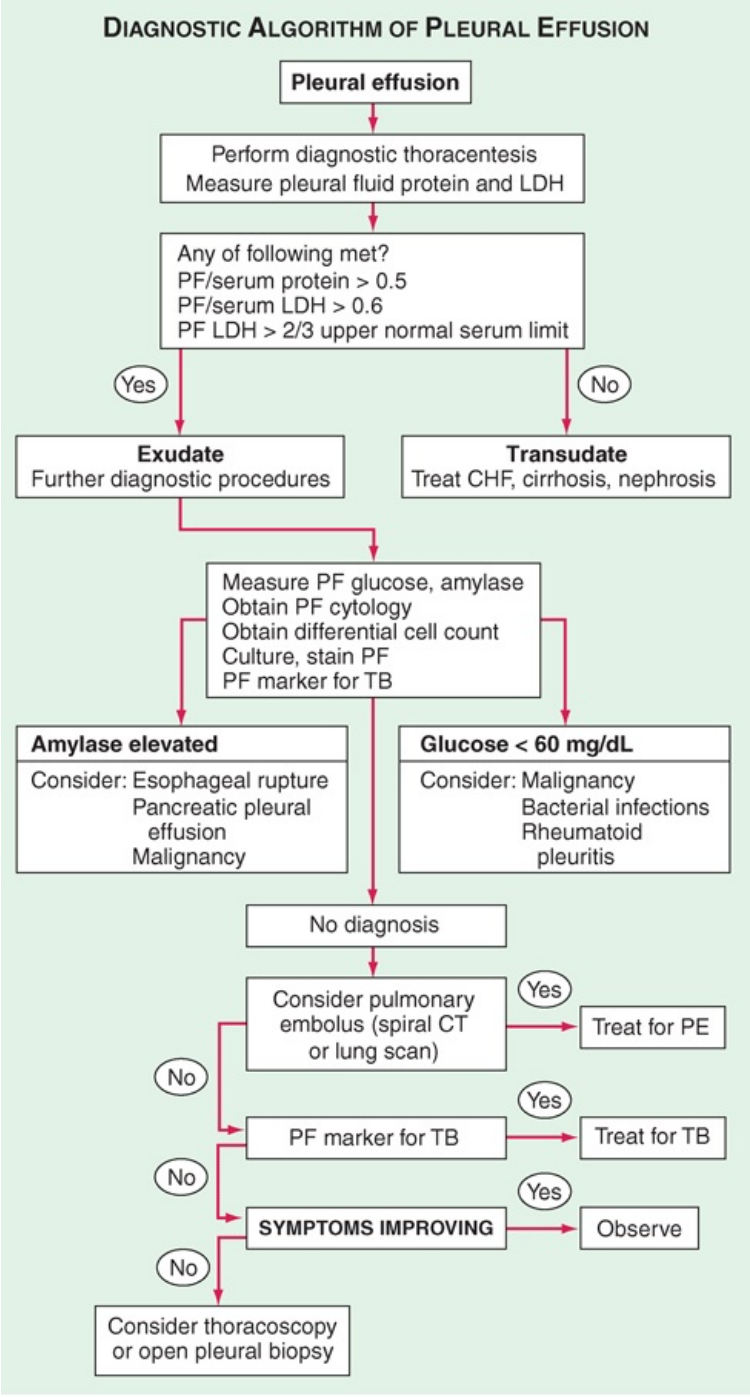

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the diagnosis of pleural effusions. CHF, congestive heart failure; CT, computed tomography; LDH, lactate dehydrogenase; PE, pulmonary embolism; TB, tuberculosis; PF, pleural fluid.

Transudative and exudative pleural effusions are distinguished by measuring the lactate dehydrogenase (LDH) and protein levels in the pleural fluid. Exudative pleural effusions meet at least one of the following criteria, whereas transudative pleural effusions meet none:

1. Pleural fluid protein/serum protein >0.5
2. Pleural fluid LDH/serum LDH >0.6
3. Pleural fluid LDH more than two-thirds normal upper limit for serum

These criteria misidentify ~25% of transudates as exudates. If one or more of the exudative criteria are met and the patient is clinically thought to have a condition producing a transudative effusion, the difference between the protein levels in the serum and the pleural fluid should be measured. If this gradient is >31 g/L (3.1 g/dL), the exudative categorization by these criteria can be ignored because

almost all such patients have a transudative pleural effusion.

If a patient has an exudative pleural effusion, the following tests on the pleural fluid should be obtained: description of the appearance of the fluid, glucose level, differential cell count, microbiologic studies, and cytology.

EFFUSION DUE TO HEART FAILURE

The most common cause of pleural effusion is left-ventricular failure. The effusion occurs because the increased amounts of fluid in the lung interstitial spaces exit in part across the visceral pleura; this overwhelms the capacity of the lymphatics in the parietal pleura to remove fluid. In patients with heart failure, a diagnostic thoracentesis should be performed if the effusions are not bilateral and comparable in size, if the patient is febrile, or if the patient has pleuritic chest pain to verify that the patient has a transudative effusion. Otherwise the patient's heart failure is treated. If the effusion persists despite therapy, a diagnostic thoracentesis should be performed. A pleural fluid N-terminal pro-brain natriuretic peptide (NT-proBNP) >1500 pg/mL is virtually diagnostic of an effusion secondary to congestive heart failure.

HEPATIC HYDROTHORAX

Pleural effusions occur in ~5% of patients with cirrhosis and ascites. The predominant mechanism is the direct movement of peritoneal fluid through small openings in the diaphragm into the pleural space. The effusion is usually right-sided and frequently is large enough to produce severe dyspnea.

PARAPNEUMONIC EFFUSION

Parapneumonic effusions are associated with bacterial pneumonia, lung abscess, or bronchiectasis and are probably the most common cause of exudative pleural effusion in the United States. *Empyema* refers to a grossly purulent effusion.

Patients with aerobic bacterial pneumonia and pleural effusion present with an acute febrile illness consisting of chest pain, sputum production, and leukocytosis. Patients with anaerobic infections present with a subacute illness with weight loss, a brisk leukocytosis, mild anemia, and a history of some factor that predisposes them to aspiration.

The possibility of a parapneumonic effusion should be considered whenever a patient with bacterial pneumonia is initially evaluated. The presence of free pleural fluid can be demonstrated with a lateral decubitus radiograph, computed tomography (CT) of the chest, or ultrasound. If the free fluid separates the lung from the chest wall by >10 mm, a therapeutic thoracentesis should be performed. Factors indicating the likely need for a procedure more invasive than a thoracentesis (in increasing order of importance) include the following:

1. Loculated pleural fluid
2. Pleural fluid pH <7.20
3. Pleural fluid glucose <3.3 mmol/L (<60 mg/dL)
4. Positive Gram stain or culture of the pleural fluid
5. Presence of gross pus in the pleural space

If the fluid recurs after the initial therapeutic thoracentesis and if any of these characteristics are present, a repeat thoracentesis should be performed. If the fluid cannot be completely removed with the therapeutic thoracentesis, consideration should be given to inserting a chest tube and instilling a fibrinolytic agent (e.g., tissue plasminogen activator, 10 mg) or performing a thoracoscopy with the breakdown of adhesions. Decortication should be considered when these measures are ineffective.

EFFUSION SECONDARY TO MALIGNANCY

Malignant pleural effusions secondary to metastatic disease are the second most common type of exudative pleural effusion. The three tumors that cause ~75% of all malignant pleural effusions are lung carcinoma, breast carcinoma, and lymphoma. Most patients complain of dyspnea, which is frequently out of proportion to the size of the effusion. The pleural fluid is an exudate, and its glucose level may be reduced if the tumor burden in the pleural space is high.

The diagnosis usually is made via cytology of the pleural fluid. If the initial cytologic examination is negative, thoracoscopy is the best next procedure if malignancy is strongly suspected. At the time of thoracoscopy, a procedure such as pleural abrasion should be performed to effect a pleurodesis. An alternative to thoracoscopy is CT- or ultrasound-guided needle biopsy of pleural thickening or nodules. Patients with a malignant pleural effusion are treated symptomatically for the most part, since the presence of the effusion indicates disseminated disease and most malignancies associated with pleural effusion are not curable with chemotherapy. The only symptom that can be attributed to the effusion itself is dyspnea. If the patient's lifestyle is compromised by dyspnea and if the dyspnea is relieved with a therapeutic thoracentesis, one of the following procedures should be considered: (1) insertion of a small indwelling catheter or (2) tube thoracostomy with the instillation of a sclerosing agent such as doxycycline, 500 mg.

MESOTHELIOMA

Malignant mesotheliomas are primary tumors that arise from the mesothelial cells that line the pleural cavities; most are related to asbestos exposure. Patients with mesothelioma present with chest pain and shortness of breath. The chest radiograph reveals a pleural effusion, generalized pleural thickening, and a shrunken hemithorax. Thoracoscopy or open pleural biopsy is usually necessary to establish the diagnosis. Chest pain should be treated with opiates, and shortness of breath with oxygen and/or opiates.

EFFUSION SECONDARY TO PULMONARY EMBOLIZATION

The diagnosis most commonly overlooked in the differential diagnosis of a patient with an undiagnosed pleural effusion is pulmonary embolism. Dyspnea is the most common symptom. The pleural fluid is almost always an exudate. The diagnosis is established by spiral CT scan or pulmonary arteriography (Chap. 262). Treatment of a patient with a pleural effusion secondary to pulmonary embolism is the same as it is for any patient with pulmonary emboli. If the pleural effusion increases in size after anticoagulation, the patient probably has recurrent emboli or another complication, such as a hemothorax or a pleural infection.

TUBERCULOUS PLEURITIS

(See also Chap. 165) In many parts of the world, the most common cause of an exudative pleural effusion is tuberculosis (TB), but tuberculous effusions are relatively uncommon in the United States. Tuberculous pleural effusions usually are associated with primary TB and are thought to be due primarily to a hypersensitivity reaction to tuberculous protein in the pleural space. Patients with tuberculous pleuritis present with fever, weight loss, dyspnea, and/or pleuritic chest pain. The pleural fluid is an exudate with predominantly small lymphocytes. The diagnosis is established by demonstrating high levels of TB markers in the pleural fluid (adenosine deaminase >40 IU/L or interferon γ >140 pg/mL). Alternatively, the diagnosis can be established by culture of the pleural fluid, needle biopsy of the pleura, or thoracoscopy. The recommended treatments of pleural and pulmonary TB are identical (Chap. 165).

EFFUSION SECONDARY TO VIRAL INFECTION

Viral infections are probably responsible for a sizable percentage of undiagnosed exudative pleural effusions. In many series, no diagnosis is established for ~20% of exudative effusions, and these effusions resolve spontaneously with no long-term residua. The importance of these effusions is that one should not be too aggressive in trying to establish a diagnosis for the undiagnosed effusion, particularly if the patient is improving clinically.

CHYLOTHORAX

A chylothorax occurs when the thoracic duct is disrupted and chyle accumulates in the pleural space. The most common cause of chylothorax is trauma (most frequently thoracic surgery), but it also may result from tumors in the mediastinum. Patients with chylothorax present with dyspnea, and a large pleural effusion is present on the chest radiograph. Thoracentesis reveals milky fluid, and biochemical analysis reveals a triglyceride level that exceeds 1.2 mmol/L (110 mg/dL). Patients with chylothorax and no obvious trauma should have a lymphangiogram and a mediastinal CT scan to assess the mediastinum for lymph nodes. The treatment of choice for most chylothoraxes is insertion of a chest tube plus the administration of octreotide. If these modalities fail, a pleuroperitoneal shunt should be placed unless the patient has chylous ascites. An alternative treatment is ligation of the thoracic duct. Patients with chylothoraxes should not undergo prolonged tube thoracostomy with chest tube drainage because this will lead to malnutrition and immunologic incompetence.

HEMOTHORAX

When a diagnostic thoracentesis reveals bloody pleural fluid, a hematocrit should be obtained on the pleural fluid. If the hematocrit is more than one-half of that in the peripheral blood, the patient is considered to have a hemothorax. Most hemothoraxes are the result of trauma; other causes include rupture of a blood vessel or tumor. Most patients with hemothorax should be treated with tube thoracostomy, which allows continuous quantification of bleeding. If the bleeding emanates from a laceration of the pleura, apposition of the two pleural surfaces is likely to stop the bleeding. If the pleural hemorrhage exceeds 200 mL/h, consideration should be given to thoracoscopy or thoracotomy.

MISCELLANEOUS CAUSES OF PLEURAL EFFUSION

There are many other causes of pleural effusion **(Table 263-1)**. Key features of some of these conditions are as follows: If the pleural fluid amylase level is elevated, the diagnosis of esophageal rupture or pancreatic disease is likely. If the patient is febrile, has predominantly polymorphonuclear cells in the pleural fluid, and has no pulmonary parenchymal abnormalities, an intraabdominal abscess should be considered.

Table 263-1 Differential Diagnoses of Pleural Effusions

Transudative Pleural Effusions	

1. Congestive heart failure	5. Peritoneal dialysis
2. Cirrhosis	6. Superior vena cava obstruction
3. Pulmonary embolization	7. Myxedema
4. Nephrotic syndrome	8. Urinothorax
Exudative Pleural Effusions	
1. Neoplastic diseases	6. Post-coronary artery bypass surgery
a. Metastatic disease	7. Asbestos exposure
b. Mesothelioma	8. Sarcoidosis
2. Infectious diseases	9. Uremia
a. Bacterial infections	10. Meigs' syndrome
b. Tuberculosis	11. Yellow nail syndrome
c. Fungal infections	12. Drug-induced pleural disease
d. Viral infections	a. Nitrofurantoin
e. Parasitic infections	b. Dantrolene
3. Pulmonary embolization	c. Methysergide
4. Gastrointestinal disease	d. Bromocriptine
a. Esophageal perforation	e. Procarbazine
b. Pancreatic disease	f. Amiodarone
c. Intraabdominal abscesses	g. Dasatinib
d. Diaphragmatic hernia	13. Trapped lung
e. After abdominal surgery	14. Radiation therapy
f. Endoscopic variceal sclerotherapy	15. Post-cardiac injury syndrome
g. After liver transplant	16. Hemothorax
5. Collagen vascular diseases	17. Iatrogenic injury
a. Rheumatoid pleuritis	18. Ovarian hyperstimulation syndrome
b. Systemic lupus erythematosus	19. Pericardial disease
c. Drug-induced lupus	20. Chylothorax
d. Immunoblastic lymphadenopathy	
e. Sjögren's syndrome	
f. Granulomatosis with polyangiitis (Wegener's)	
g. Churg-Strauss syndrome	

The diagnosis of an asbestos pleural effusion is one of exclusion. Benign ovarian tumors can produce ascites and a pleural effusion (Meigs' syndrome), as can the ovarian hyperstimulation syndrome. Several drugs can cause pleural effusion; the associated fluid is usually eosinophilic. Pleural effusions commonly occur after coronary artery bypass surgery. Effusions occurring within the first weeks are typically left-sided and bloody, with large numbers of eosinophils, and respond to one or two therapeutic thoracenteses. Effusions occurring after the first few weeks are typically left-sided and clear yellow, with predominantly small lymphocytes, and tend to recur. Other medical manipulations that induce pleural effusions include abdominal surgery; radiation therapy; liver, lung, or heart transplantation; and the intravascular insertion of central lines.

Pneumothorax

Pneumothorax is the presence of gas in the pleural space. A *spontaneous pneumothorax* is one that occurs without antecedent trauma to the thorax. A *primary spontaneous pneumothorax* occurs in the absence of underlying lung disease, whereas a *secondary pneumothorax* occurs in its presence. A *traumatic pneumothorax* results from penetrating or nonpenetrating chest injuries. A *tension pneumothorax* is a pneumothorax in which the pressure in the pleural space is positive throughout the respiratory cycle.

PRIMARY SPONTANEOUS PNEUMOTHORAX

Primary spontaneous pneumothoraxes are usually due to rupture of apical pleural blebs, small cystic spaces that lie within or immediately under the visceral pleura. Primary spontaneous pneumothoraxes occur almost exclusively in smokers; this suggests that these patients have subclinical lung disease. Approximately one-half of patients with an initial primary spontaneous pneumothorax will have a

recurrence. The initial recommended treatment for primary spontaneous pneumothorax is simple aspiration. If the lung does not expand with aspiration or if the patient has a recurrent pneumothorax, thoracoscopy with stapling of blebs and pleural abrasion is indicated. Thoracoscopy or thoracotomy with pleural abrasion is almost 100% successful in preventing recurrences.

SECONDARY PNEUMOTHORAX

Most secondary pneumothoraxes are due to chronic obstructive pulmonary disease, but pneumothoraxes have been reported with virtually every lung disease. Pneumothorax in patients with lung disease is more life-threatening than it is in normal individuals because of the lack of pulmonary reserve in these patients. Nearly all patients with secondary pneumothorax should be treated with tube thoracostomy. Most should also be treated with thoracoscopy or thoracotomy with the stapling of blebs and pleural abrasion. If the patient is not a good operative candidate or refuses surgery, pleurodesis should be attempted by the intrapleural injection of a sclerosing agent such as doxycycline.

TRAUMATIC PNEUMOTHORAX

Traumatic pneumothoraxes can result from both penetrating and nonpenetrating chest trauma. Traumatic pneumothoraxes should be treated with tube thoracostomy unless they are very small. If a hemopneumothorax is present, one chest tube should be placed in the superior part of the hemithorax to evacuate the air and another should be placed in the inferior part of the hemithorax to remove the blood. Iatrogenic pneumothorax is a type of traumatic pneumothorax that is becoming more common. The leading causes are transthoracic needle aspiration, thoracentesis, and the insertion of central intravenous catheters. Most can be managed with supplemental oxygen or aspiration, but if these measures are unsuccessful, a tube thoracostomy should be performed.

TENSION PNEUMOTHORAX

This condition usually occurs during mechanical ventilation or resuscitative efforts. The positive pleural pressure is life-threatening both because ventilation is severely compromised and because the positive pressure is transmitted to the mediastinum, resulting in decreased venous return to the heart and reduced cardiac output.

Difficulty in ventilation during resuscitation or high peak inspiratory pressures during mechanical ventilation strongly suggest the diagnosis. The diagnosis is made by physical examination showing an enlarged hemithorax with no breath sounds, hyperresonance to percussion, and shift of the mediastinum to the contralateral side. Tension pneumothorax must be treated as a medical emergency. If the tension in the pleural space is not relieved, the patient is likely to die from inadequate cardiac output or marked hypoxemia. A large-bore needle should be inserted into the pleural space through the second anterior intercostal space. If large amounts of gas escape from the needle after insertion, the diagnosis is confirmed. The needle should be left in place until a thoracostomy tube can be inserted.

DISORDERS OF THE MEDIASTINUM

The mediastinum is the region between the pleural sacs. It is separated into three compartments. The *anterior mediastinum* extends from the sternum anteriorly to the pericardium and brachiocephalic vessels posteriorly. It contains the thymus gland, the anterior mediastinal lymph nodes, and the internal mammary arteries and veins. The *middle mediastinum* lies between the anterior and posterior mediastina and contains the heart; the ascending and transverse arches of the aorta; the venae cavae; the brachiocephalic arteries and veins; the phrenic nerves; the trachea, the main bronchi, and their contiguous lymph nodes; and the pulmonary arteries and veins. The *posterior mediastinum* is bounded by the pericardium and trachea anteriorly and the vertebral column posteriorly. It contains the descending thoracic aorta, the esophagus, the thoracic duct, the azygos and hemiazygos veins, and the posterior group of mediastinal lymph nodes.

Mediastinal Masses

The first step in evaluating a mediastinal mass is to place it in one of the three mediastinal compartments, since each has different characteristic lesions. The most common lesions in the anterior mediastinum are thymomas, lymphomas, teratomatous neoplasms, and thyroid masses. The most common masses in the middle mediastinum are vascular masses, lymph node enlargement from metastases or granulomatous disease, and pleuropericardial and bronchogenic cysts. In the posterior mediastinum, neurogenic tumors, meningoceles, meningomyeloceles, gastroenteric cysts, and esophageal diverticula are commonly found.

CT scanning is the most valuable imaging technique for evaluating mediastinal masses and is the only imaging technique that should be done in most instances. Barium studies of the gastrointestinal tract are indicated in many patients with posterior mediastinal lesions, since hernias, diverticula, and achalasia are readily diagnosed in this manner. An ^{131}I scan can efficiently establish the diagnosis of intrathoracic goiter.

A definite diagnosis can be obtained with mediastinoscopy or anterior mediastinotomy in many patients with masses in the anterior or middle mediastinal compartments. A diagnosis can be established without thoracotomy via percutaneous fine-needle aspiration biopsy or endoscopic transesophageal or

endobronchial ultrasound-guided biopsy of mediastinal masses in most cases. Alternative ways to establish the diagnosis are video-assisted thoracoscopy, mediastinoscopy, and mediastinotomy. In many cases the diagnosis can be established and the mediastinal mass removed with video-assisted thoracoscopy.

Acute Mediastinitis

Most cases of acute mediastinitis either are due to esophageal perforation or occur after median sternotomy for cardiac surgery. Patients with esophageal rupture are acutely ill with chest pain and dyspnea due to the mediastinal infection. The esophageal rupture can occur spontaneously or as a complication of esophagoscopy or the insertion of a Blakemore tube. Appropriate treatment consists of exploration of the mediastinum with primary repair of the esophageal tear and drainage of the pleural space and the mediastinum.

The incidence of mediastinitis after median sternotomy is 0.4–5.0%. Patients most commonly present with wound drainage. Other presentations include sepsis and a widened mediastinum. The diagnosis usually is established with mediastinal needle aspiration. Treatment includes immediate drainage, debridement, and parenteral antibiotic therapy, but the mortality rate still exceeds 20%.

Chronic Mediastinitis

The spectrum of chronic mediastinitis ranges from granulomatous inflammation of the lymph nodes in the mediastinum to fibrosing mediastinitis. Most cases are due to histoplasmosis or TB, but sarcoidosis, silicosis, and other fungal diseases are at times causative. Patients with granulomatous mediastinitis are usually asymptomatic. Those with fibrosing mediastinitis usually have signs of compression of a mediastinal structure such as the superior vena cava or large airways, phrenic or recurrent laryngeal nerve paralysis, or obstruction of the pulmonary artery or proximal pulmonary veins. Other than antituberculous therapy for tuberculous mediastinitis, no medical or surgical therapy has been demonstrated to be effective for mediastinal fibrosis.

Pneumomediastinum

In this condition, there is gas in the interstices of the mediastinum. The three main causes are (1) alveolar rupture with dissection of air into the mediastinum, (2) perforation or rupture of the esophagus, trachea, or main bronchi, and (3) dissection of air from the neck or the abdomen into the mediastinum. Typically, there is severe substernal chest pain with or without radiation into the neck and arms. The physical examination usually reveals subcutaneous emphysema in the suprasternal notch and *Hamman's sign*, which is a crunching or clicking noise synchronous with the heartbeat and is best heard in the left lateral decubitus position. The diagnosis is confirmed with the chest radiograph. Usually no treatment is required, but the mediastinal air will be absorbed faster if the patient inspires high concentrations of oxygen. If mediastinal structures are compressed, the compression can be relieved with needle aspiration.

FURTHER READINGS

GILBERT S et al: Endobronchial ultrasound as a diagnostic tool in patients with mediastinal lymphadenopathy. Ann Thorac Surg 88:896, 2009[PMID: 19699917] [Full Text]

LIGHT RW: Pleural effusion. N Engl J Med 346:1971, 2002[PMID: 12075059] [Full Text]

———: *Pleural Diseases*, 5th ed. Philadelphia, Lippincott Williams & Wilkins, 2006

PORCEL JM et al: Biomarkers of heart failure in pleural fluid. Chest 163:671, 2009.

———: Pleural fluid tests to identify complicated parapneumonic effusions. Curr Opin Pulm Med 16:357, 2010

WARREN WH et al: Identification of clinical factors predicting Pleurx catheter removal in patients treated for malignant pleural effusion. Eur J Cardiothorac Surg 33:89, 2008[PMID: 17983758] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION AND PHYSIOLOGY

In health the arterial level of carbon dioxide (Pa_{CO2}) is maintained between 37 and 43 mmHg at sea level. All disorders of ventilation result in abnormal measurements of Pa_{CO2}. This chapter reviews chronic ventilatory disorders that are reflected in abnormal Pa_{CO2}.

The continuous production of CO_2 by cellular metabolism necessitates its efficient elimination by the respiratory system. The relationship between CO_2 production and Pa_{CO2} is described by the equation, $Pa_{CO2} = (k)(V._{CO2})/V._A$, where $V._{CO2}$ represents the carbon dioxide production, k is a constant and $V._A$ is fresh gas alveolar ventilation (Chap. 252). $V._A$ can be calculated as minute ventilation x(1-Vd/Vt), where the dead space fraction Vd/Vt represents the portion of a tidal breath that remains within the conducting airways at the conclusion of inspiration and does not, therefore, contribute to alveolar ventilation. As such, all disturbances of Pa_{CO2} must reflect altered CO_2 production, minute ventilation, or dead space fraction.

Diseases that alter $V._{CO2}$ are often acute (sepsis, burns, or pyrexia, for example), and their contribution to ventilatory abnormalities and/or respiratory failure is reviewed elsewhere. Chronic ventilatory disorders typically involve inappropriate levels of minute ventilation or increased dead space fraction. Characterization of these disorders requires a review of the normal respiratory cycle.

The spontaneous cycle of inspiration and expiration is automatically generated in the brainstem. Two groups of neurons located within the medulla are particularly important: the dorsal respiratory group (DRG) and the ventral respiratory column (VRC). These neurons have widespread projections, including the descending projections into the contralateral spinal cord, where they perform many functions. They initiate activity in the phrenic nerve/diaphragm, project to the upper airway muscle groups and spinal respiratory neurons, and innervate the intercostal and abdominal muscles that participate in normal respiration. The DRG acts as the initial integration site for many of the afferent nerves relaying information about the partial pressure of arterial oxygen (Pa_{O2}), Pa_{CO2}, pH, and blood pressure from the carotid and aortic chemoreceptors and baroreceptors to the central nervous system (CNS). In addition, the vagus nerve relays information from stretch receptors and juxtapulmonary-capillary receptors in the lung parenchyma and chest wall to the DRG. The respiratory rhythm is generated within the VRC, as well as the more rostrally located parafacial respiratory group (pFRG), which is particularly important for the generation of active expiration. One particularly important area within the VRC is the so-called pre-Bötzinger complex. This area is responsible for the generation of various forms of inspiratory activity, and lesioning of the pre-Bötzinger complex leads to the complete cessation of breathing. The neural output of these medullary respiratory networks can be voluntarily suppressed or augmented by input from higher brain centers and the autonomic nervous system. During normal sleep there is an attenuated response to hypercapnia and hypoxemia resulting in mild nocturnal hypoventilation that corrects upon awakening.

Once neural input has been delivered to the respiratory pump muscles, normal gas exchange requires an adequate amount of respiratory muscle strength to overcome the elastic and resistive loads of the respiratory system (**Fig. 264-1A**, Chap. 252). In health, the strength of the respiratory muscles readily accomplishes this, and normal respiration continues indefinitely. Reduction in respiratory drive or neuromuscular competence or substantial increase in respiratory load can diminish minute ventilation, resulting in hypercapnia (Fig. 264-1B). Alternatively, if normal respiratory muscle strength is coupled with excessive respiratory drive, then alveolar hyperventilation ensues and leads to hypocapnia (Fig. 264-1C).

Figure 264-1

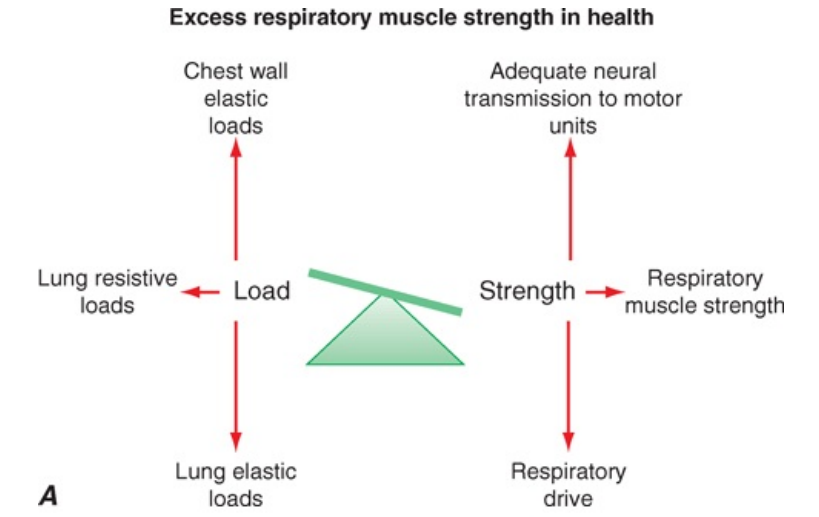

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

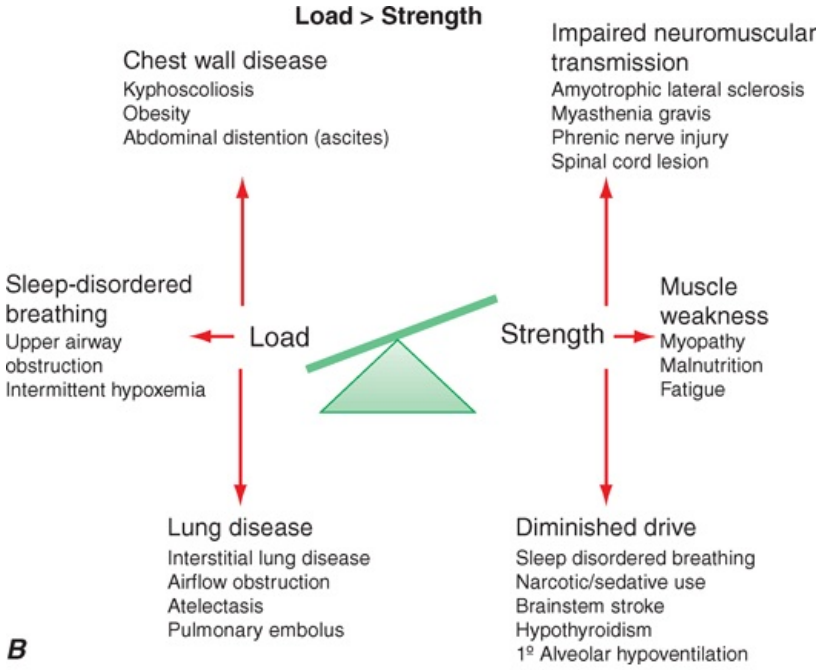

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

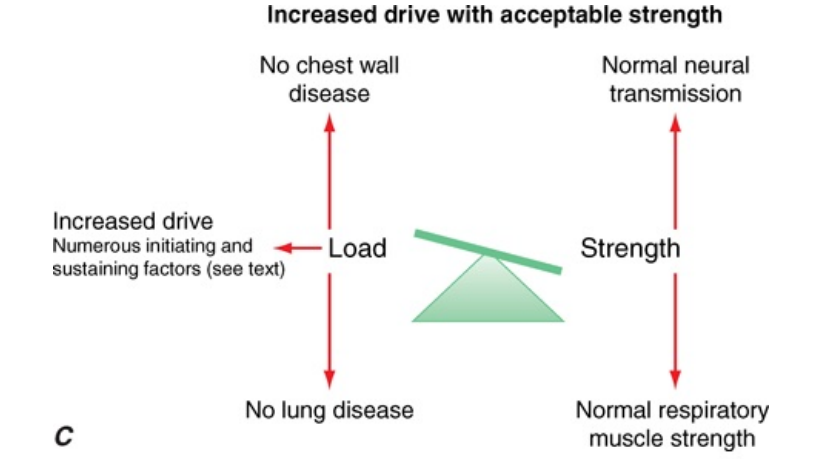

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examples of balance between respiratory system strength and load. *A*. Excess respiratory muscle strength in health. *B*. Load greater than strength. *C*. Increased drive with acceptable strength.

HYPOVENTILATION

Clinical Features

Diseases that reduce minute ventilation or increase dead space fall into four major categories: parenchymal lung and chest wall disease, sleep disordered breathing, neuromuscular disease, and respiratory drive disorders (Fig. 264-1B). The clinical manifestations of hypoventilation syndromes are nonspecific (Table 264-1) and vary depending on the severity of hypoventilation, the rate at which hypercapnia develops, the degree of compensation for respiratory acidosis, and the underlying disorder. Patients with parenchymal lung or chest wall disease typically present with shortness of breath and diminished exercise tolerance. Episodes of increased dyspnea and sputum production are hallmarks of obstructive lung diseases, such as chronic obstructive pulmonary disease (COPD), whereas progressive dyspnea and cough are common in interstitial lung diseases. Excessive daytime somnolence, poor quality sleep, and snoring are common among patients with sleep-disordered breathing. Sleep disturbance and orthopnea are also described in neuromuscular disorders. As neuromuscular weakness progresses, the respiratory muscles, including the diaphragm, are placed at a mechanical disadvantage in the supine position due to the upward movement of the abdominal contents. New-onset orthopnea is frequently a sign of reduced respiratory muscle force generation. More commonly, however, extremity weakness or bulbar symptoms develop prior to sleep disturbance in neuromuscular diseases, such as amyotrophic lateral sclerosis (ALS) or muscular dystrophy. Patients with respiratory drive disorders do not have symptoms distinguishable from other causes of chronic hypoventilation.

Table 264-1 Signs and Symptoms of Hypoventilation

Dyspnea during activities of daily living
Orthopnea in diseases affecting diaphragm function
Poor quality sleep Daytime hypersomnolence
Early morning headaches
Anxiety
Impaired cough in neuromuscular diseases

The clinical course of patients with chronic hypoventilation from neuromuscular or chest wall disease follows a characteristic sequence: An asymptomatic stage where daytime Pa_{O2} and Pa_{CO2} are normal, followed by nocturnal hypoventilation, initially during rapid eye movement (REM) sleep and later in non-REM sleep. Finally, if vital capacity drops further, daytime hypercapnia develops. Symptoms can develop at any point along this time course and often depend on the pace of respiratory muscle functional decline. Regardless of cause, the hallmark of all alveolar hypoventilation syndromes is an increase in alveolar P_{CO2} (PA_{CO2}) and, therefore, in Pa_{CO2}. The resulting respiratory acidosis eventually leads to a compensatory increase in plasma bicarbonate concentration. The increase in PA_{CO2} results in an obligatory decrease in PA_{O2}, often resulting in hypoxemia. If severe, the hypoxemia manifests clinically as cyanosis and can stimulate erythropoiesis, thereby inducing secondary erythrocytosis. The combination of chronic hypoxemia and hypercapnia may also induce pulmonary vasoconstriction, leading eventually to pulmonary hypertension, right-ventricular hypertrophy, and right heart failure.

Diagnosis

Elevated plasma bicarbonate in the absence of volume depletion is suggestive of hypoventilation. An arterial blood gas demonstrating elevated Pa_{CO2} with a normal pH confirms chronic alveolar hypoventilation. The subsequent evaluation to identify an etiology should initially focus on whether the patient has lung disease or chest wall abnormalities. Physical examination, imaging studies (chest x-ray and/or CT scan), and pulmonary function tests are sufficient to identify most lung/chest wall disorders leading to hypercapnia. If these evaluations are unrevealing, then the clinician should screen for obstructive sleep apnea (OSA), the most frequent sleep disorder leading to chronic hypoventilation. Several screening tools have been developed to identify patients at risk for OSA. The Berlin Questionnaire has been validated in a primary care setting and identifies patients likely to have OSA. The Epworth Sleepiness Scale (ESS) and the STOP-Bang questionnaire have not been validated in outpatient primary care settings but are quick and easy to use. The ESS measures daytime sleepiness, with a score of ≥10 identifying individuals who warrant additional investigation. The STOP-Bang survey has been used in preoperative clinics to identify patients at risk of having OSA. In this population, it has 93% sensitivity and 90% negative predictive value.

If the ventilatory apparatus (lung, airways, chest wall) is not responsible for chronic hypercapnia, then the focus should shift to respiratory drive and neuromuscular disorders. There is an attenuated

increase in minute ventilation in response to elevated CO_2 and/or low O_2 in respiratory drive disorders. These diseases are difficult to diagnose and should be suspected when patients with hypercapnia are found to have normal respiratory muscle strength, normal pulmonary function, and normal alveolar-arterial P_{O2} difference. Hypoventilation is more marked during sleep in patients with respiratory drive defects, and polysomnography often reveals central apneas, hypopneas, or hypoventilation. Brain imaging (CT scan or MRI) can sometimes identify structural abnormalities in the pons or medulla that result in hypoventilation. Chronic narcotic use or significant hypothyroidism can depress the central respiratory drive and lead to chronic hypercapnia, as well.

Respiratory muscle weakness has to be profound before lung volumes are compromised and hypercapnia develops. Typically, physical examination reveals decreased strength in major muscle groups prior to the development of hypercapnia. Measurement of maximum inspiratory and expiratory pressures or forced vital capacity (FVC) can be used to monitor for respiratory muscle involvement in diseases with progressive muscle weakness. These patients also have increased risk for sleep-disordered breathing, including hypopneas, central and obstructive apneas, and hypoxemia. Nighttime oximetry and polysomnography are helpful in better characterizing sleep disturbances in this patient population.

Treatment: Hypoventilation

Nocturnal noninvasive positive-pressure ventilation (NIPPV) has been used successfully in the treatment of hypoventilation and apneas, both central and obstructive, in patients with neuromuscular and chest wall disorders. Nocturnal NIPPV has been shown to improve daytime hypercapnia, prolong survival, and improve health-related quality of life when daytime hypercapnia is documented. ALS guidelines recommend nocturnal NIPPV if symptoms of hypoventilation exist AND one of the following criteria is present: $Paco_2 \geq 45$ mmHg; nocturnal oximetry demonstrates oxygen saturation ≤88% for 5 consecutive minutes; maximal inspiratory pressure <60 cm H_2O; and FVC <50% predicted. However, at present there is inconclusive evidence to support preemptive nocturnal NIPPV use in all patients with neuromuscular and chest wall disorders who demonstrate nocturnal, but not daytime, hypercapnia. Nevertheless, at some point the institution of full-time ventilatory support with either pressure or volume-preset modes is required in progressive neuromuscular disorders. There is less evidence to direct the timing of this decision, but ventilatory failure requiring mechanical ventilation and chest infections related to ineffective cough are frequent triggers for the institution of full-time ventilatory support.

Treatment of chronic hypoventilation from lung or neuromuscular diseases should be directed at the underlying disorder. Pharmacologic agents that stimulate respiration, such as medroxyprogesterone and acetazolamide, have been poorly studied in chronic hypoventilation and should not replace treatment of the underlying disease process. Regardless of the cause, excessive metabolic alkalosis should be corrected, as plasma bicarbonate levels elevated out of proportion to the degree of chronic respiratory acidosis can result in additional hypoventilation. When indicated, administration of supplemental oxygen is effective in attenuating hypoxemia, polycythemia, and pulmonary hypertension.

Phrenic nerve or diaphragm pacing is a potential therapy for patients with hypoventilation from high cervical spinal cord lesions or respiratory drive disorders. Prior to surgical implantation, patients should have nerve conduction studies to ensure normal bilateral phrenic nerve function. Small case series suggest that effective diaphragmatic pacing can improve quality of life in these patients.

HYPOVENTILATION SYNDROMES

Obesity Hypoventilation Syndrome

The diagnosis of obesity hypoventilation syndrome (OHS) requires: body mass index (BMI) ≥30 kg/m^2, sleep-disordered breathing and chronic daytime alveolar hypoventilation, defined as $Pa_{CO2} \geq 45$ mmHg, and Pa_{O2} < 70 mmHg in the absence of other known causes of hypercapnia. In almost 90% of cases, the sleep-disordered breathing is in the form of OSA. Several international studies in different populations confirm that the overall prevalence of obstructive sleep apnea syndrome, defined by an apnea hypopnea index ≥5 AND daytime sleepiness, is approximately 3–4% in middle-aged men and 2% in middle-aged women. Thus, the population at risk for the development of OHS continues to rise as the worldwide obesity epidemic persists. Although no population-based prevalence studies of OHS have been performed, some estimates suggest there may be as many as 500,000 individuals with OHS in the United States.

Several studies suggest that severe obesity (BMI >40 kg/m^2) and severe OSA apnea-hypopnea index [(AHI) >30 events per hour] are risk factors for the development of OHS. The pathogenesis of hypoventilation in these patients is multifactorial and incompletely understood. Defects in central respiratory drive have been demonstrated in OHS patients but often improve with treatment. This suggests central defects may not be the primary disturbance that leads to chronic hypercapnia. The treatment of OHS is similar to that for OSA: weight reduction and continuous positive airway pressure (CPAP) therapy during sleep. CPAP improves daytime hypercapnia and hypoxemia in the majority of

patients with OHS. There is not conclusive evidence to suggest that bilevel positive airway pressure (BiPAP) is superior to CPAP. Bilevel positive airway pressure should be reserved for patients not able to tolerate high levels of CPAP support or patients that remain hypoxemic despite resolution of obstructive respiratory events.

Central Hypoventilation Syndrome

This syndrome can present later in life or in the neonatal period where it is often called Ondine's curse or congenital central hypoventilation syndrome (CCHS). Abnormalities in the gene encoding PHOX2b, a transcription factor with a role in neuronal development, have been implicated in the pathogenesis of congenital central hypoventilation syndrome. Regardless of the age of onset, these patients have absent respiratory response to hypoxia or hypercapnia, mildly elevated Pa_{CO2} while awake, and markedly elevated Pa_{CO2} during non-REM sleep. Interestingly, these patients are able to augment their ventilation and "normalize" Pa_{CO2} during exercise. These patients typically require NIPPV or mechanical ventilation as therapy and should be considered for phrenic nerve or diaphragmatic pacing at centers with experience performing these procedures.

HYPERVENTILATION

Clinical Features

Hyperventilation is defined as ventilation in excess of metabolic requirements (CO_2 production) leading to a reduction in Pa_{CO2}. The physiology of patients with chronic hyperventilation is poorly understood, and there is no typical clinical presentation. Symptoms can include dyspnea, paresthesias, tetany, headache, dizziness, visual disturbances, and atypical chest pain. Because symptoms can be so diverse, patients with chronic hyperventilation present to a variety of health care providers, including internists, neurologists, psychologists, psychiatrists, and pulmonologists.

It is helpful to think of hyperventilation as having initiating and sustaining factors. Some investigators believe that an initial event leads to increased alveolar ventilation and a drop in Pa_{CO2} to ~20 mmHg. The ensuing onset of chest pain, breathlessness, paresthesia, or altered consciousness can be alarming. The resultant increase in minute volume to relieve these acute symptoms only serves to exacerbate symptoms that are often misattributed by the patient and health care workers to cardiopulmonary disorders. An unrevealing evaluation for causes of these symptoms often results in patients being anxious and fearful of additional attacks. It is important to note that **anxiety disorders and panic attacks are NOT synonymous with hyperventilation**. Anxiety can be both an initiating and sustaining factor in the pathogenesis of chronic hyperventilation, but these are not necessary for the development of chronic hypocapnia.

Diagnosis

Respiratory symptoms associated with acute hyperventilation can be the initial manifestation of systemic illnesses such as diabetic ketoacidosis. Causes of acute hyperventilation need to be excluded before a diagnosis of chronic hyperventilation is considered. Arterial blood gas sampling that demonstrates a compensated respiratory alkalosis with a near-normal pH, low Pa_{CO2} and low calculated bicarbonate are necessary to confirm chronic hyperventilation. Other causes of respiratory alkalosis, such as mild asthma, need to be diagnosed and treated before chronic hyperventilation can be considered. A high index of suspicion is required as increased minute ventilation can be difficult to detect on physical examination. Once chronic hyperventilation is established, a sustained 10% increase in alveolar ventilation is enough to perpetuate hypocapnia. This increase can be accomplished with subtle changes in the respiratory pattern, such as occasional sigh breaths or yawning two to three times per minute.

Treatment: Hyperventilation

There are few well-controlled treatment studies of chronic hyperventilation because of its diverse features and the lack of a universally accepted diagnostic process. Clinicians often spend considerable time identifying initiating factors, excluding alternative diagnoses and discussing the patient's concerns and fears. In some patients, reassurance and frank discussion about hyperventilation can be liberating. Identifying and eliminating habits that perpetuate hypocapnia, such as frequent yawning or sigh breathing, can be helpful. Some evidence suggests that breathing exercises and diaphragmatic retraining may be beneficial for some patients. The evidence for using medications to treat hyperventilation is scant. Beta-blockers may be helpful in patients with sympathetically mediated symptoms, such as palpitations and tremors.

ACKNOWLEDGMENT

We would like to acknowledge Eliot A Eliot A. Phillipson for earlier versions of this chapter and Jan-Marino Ramirez for his careful critique and helpful suggestions.

FURTHER READINGS

CHUNG F et al: STOP Questionnaire: A tool to screen patients for obstructive sleep apnea. Anesthesiology 108:812, 2008[PMID: 18431116] [Full Text]

DOUGLAS IS et al: Acute-on-chronic respiratory failure, in *Principles of Critical Care* , 3rd ed, JB Hall et al (eds). McGraw-Hill, 2005

GARDNER WN: The pathophysiology of hyperventilation disorders. Chest 109:516, 1996[PMID: 8620731] [Full Text]

LAFFEY JG, KAVANAGH BP: Hypocapnia. N Engl J Med 347:43, 2002[PMID: 12097540] [Full Text]

LITTLETON SW, MOKHLESI B: The Pickwickian syndrome-obesity hypoventilation syndrome. Clin Chest Med 30:467, 2009[PMID: 19700045] [Full Text]

SIMONDS AK: Recent advances in respiratory care for neuromuscular disease. Chest 130:1879, 2006[PMID: 17167012] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OBSTRUCTIVE SLEEP APNEA

Obstructive sleep apnea/hypopnea syndrome (OSAHS) is one of the most important medical conditions identified in the last 50 years. It is a major cause of morbidity, a significant cause of mortality, and the most common medical cause of daytime sleepiness. Central sleep apnea is a rare clinical problem. **Other sleep disorders are discussed in Chap. 27.**

Definition

OSAHS is defined as the coexistence of unexplained excessive daytime sleepiness with at least five obstructed breathing events (apnea or hypopnea) per hour of sleep (**Table 265-1**). This event threshold may have to be increased in the elderly. *Apneas* are defined in adults as breathing pauses lasting ≥10 s and hypopneas as events ≥10 s in which there is continued breathing but ventilation is reduced by at least 50% from the previous baseline during sleep. As a syndrome, OSAHS is the association of a clinical picture with specific abnormalities on testing; asymptomatic individuals with abnormal breathing during sleep should not be labeled as having OSAHS.

Table 265-1 Clinical Indicators in the Sleepy Patient

	OSAHS	Narcolepsy	IHS
Age of onset (years)	35–60	10–30	10–30
Cataplexy	No	Yes	No
Night sleep			
Duration	Normal	Normal	Long
Awakenings	Occasional	Frequent	Rare
Snoring	Yes, loud	Occasional	Occasional
Morning drunkenness	Occasional	Occasional	Common
Daytime naps			
Frequency	Usually few	Many	Few
Time of day	Afternoon/evening	Afternoon/evening	Morning
Duration	<1 h	<1 h	>1 h

Note: Features suggesting obstructive sleep apnea/hypopnea syndrome (OSAHS), narcolepsy, or idiopathic hypersomnolence (IHS).

Mechanism of Obstruction

Apneas and hypopneas are caused by the airway being sucked closed on inspiration during sleep. This occurs as the upper-airway dilating muscles—like all striated muscles—relax during sleep. In patients with OSAHS, the dilating muscles fail to oppose negative pressure within the airway during inspiration. The primary defect is not in the upper-airway muscles, which function normally in OSAHS patients when awake. These patients have narrow upper airways already during wakefulness, but when they are awake, their airway dilating muscles have increased activity, which ensures airway patency. However, during sleep, muscle tone falls and the airway narrows; snoring may commence before the airway occludes, and apnea results. Apneas and hypopneas terminate when the subject arouses, i.e., wakens briefly, from sleep. This arousal is sometimes too subtle to be seen on the electroencephalogram but may be detected by cardiac acceleration, blood pressure elevation, or increase in sympathetic tone. The arousal results in return of upper-airway dilating muscle tone, and thus airway patency is resumed.

Factors predisposing to OSAHS by narrowing the pharynx include obesity—in Western populations around 50% of OSAHS patients have a body mass index (BMI) >30 kg/m^2—and shortening of the mandible and/or maxilla. This change in jaw shape may be subtle and can be familial. Hypothyroidism and acromegaly predispose to OSAHS by narrowing the upper airway with tissue infiltration. Other predisposing factors for OSAHS include male sex and middle age (40–65 years), myotonic dystrophy, Ehlers-Danlos syndrome, and, perhaps, smoking.

Epidemiology

OSAHS occurs in around 1–4% of middle-aged males and is about half as common in women. The syndrome also occurs in childhood—usually associated with tonsil or adenoid enlargement—and in the elderly, although the frequency is slightly lower in old age. Irregular breathing during sleep *without* daytime sleepiness is much more common, occurring in perhaps a quarter of the middle-aged male population. As these individuals are asymptomatic, they do not have OSAHS, but there is increasing epidemiologic evidence of an association of irregular breathing during sleep with increased vascular risk even in the nonsleepy.

Clinical Features

Randomized controlled treatment trials have shown that OSAHS causes daytime sleepiness; impaired vigilance, cognitive performance and driving; depression; disturbed sleep; and hypertension. Daytime sleepiness may range from mild to irresistible and can be indistinguishable from that in narcolepsy (Chap. 27). The sleepiness may cause inability to work effectively, damage interpersonal relationships, and prevent socializing. The somnolence is dangerous, with a three- to sixfold risk of road accidents. Experiments with normal subjects repeatedly aroused from sleep indicate that the sleepiness results, at least in part, from the repetitive sleep disruption associated with the breathing abnormality. Other symptoms include difficulty concentrating, unrefreshing nocturnal sleep, nocturnal choking, nocturia, and decreased libido. Partners report nightly loud snoring in all postures, which may be punctuated by the silence of apneas.

CARDIOVASCULAR AND CEREBROVASCULAR EVENTS

OSAHS raises 24-h mean blood pressure. The increase is greater in those with recurrent nocturnal hypoxemia, is at least 4–5 mmHg, and may be as great as 10 mmHg in those with >20% arterial oxygen desaturations per hour of sleep. This rise probably results from a combination of surges in blood pressure accompanying each arousal at apnea/hypopnea termination and from the associated 24-h increases in sympathetic tone.

Epidemiologic data in normal populations indicate that this rise in blood pressure would increase the risk of myocardial infarction by around 20% and that of stroke by about 40%. Although there are no long-term randomized controlled trials to indicate whether this is true in OSAHS patients—such studies would be unethical—observational studies suggest an increase in cardiovascular and stroke risk in patients with untreated OSAHS. Furthermore, epidemiologic studies suggest increased vascular risk in normal subjects with raised apneas and hypopneas during sleep. Patients with recent stroke have a high frequency of apneas and hypopneas during sleep. These seem largely to be a consequence, not a cause, of the stroke and to decline over the weeks after the vascular event. There is no evidence that treating the apneas and hypopneas improves stroke outcome.

DIABETES MELLITUS

The association of OSAHS with diabetes mellitus is not due only to the fact that obesity is common in both conditions. Increased apneas and hypopneas during sleep are associated with insulin resistance independent of obesity. In addition, uncontrolled trials suggest OSAHS can aggravate diabetes and that treating OSAHS in patients who also have diabetes decreases their insulin requirements.

LIVER

Hepatic dysfunction also has been associated with irregular breathing during sleep. Non-alcohol-drinking subjects with apneas and hypopneas during sleep were found to have raised liver enzymes and more steatosis and fibrosis on liver biopsy, independent of body weight.

ANESTHESTIC RISK

Patients with OSAHS are at increased risk perioperatively as their upper airways may obstruct during the recovery period or as a consequence of sedation. Patients whom anesthesiologists have difficulty intubating are much more likely to have irregular breathing during sleep. Anesthesiologists should thus take preoperative sleep histories and take the appropriate precautions with patients who might have OSAHS.

DIFFERENTIAL DIAGNOSIS

(See also Chap. 27) Causes of sleepiness that may need to be distinguished include (Table 265-1) the following:

- *Insufficient sleep*: this usually can be diagnosed by history.
- *Shift work*: a major cause of sleepiness, especially in those >40 years old.
- *Psychological psychiatric causes:* depression is a major cause of sleepiness.
- *Drugs*: both stimulant and sedative drugs can produce sleepiness.
- *Narcolepsy*: around 50 times less common than OSAHS, narcolepsy is usually evident from childhood or the teens and is associated with cataplexy.
- *Idiopathic hypersomnolence*: this is an ill-defined condition typified by long sleep duration and sleepiness.

- *Phase alteration syndromes*: both the phase delay and the less common phase advancement syndromes are characterized by sleepiness at the appropriate time of day.

WHO TO REFER FOR DIAGNOSIS

Anyone whose troublesome sleepiness is not readily explained and rectified by considering the differential diagnosis above should be referred to a sleep specialist. The guideline the author uses for patients with troublesome sleepiness includes those with an Epworth Sleepiness Score >11 (**Table 265-2**) and also those whose sleepiness during work or driving poses problems. The Epworth Score is not a perfect measure for detecting sleepiness, as many whose lives are troubled by frequently fighting sleepiness but who never doze will correctly give themselves a low Epworth Score. The patient and his or her partner often give divergent scores for the patient's sleepiness, and in such cases the higher of the two scores should be used.

Table 265-2 Epworth Sleepiness Score

How often are you likely to doze off or fall asleep in the following situations, in contrast to feeling just tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently, try to work out how they would have affected you. Use the following scale to choose the *most appropriate number* for each situation:	
0 = would *never* doze	
1 = *slight* chance of dozing	
2 = *moderate* chance of dozing	
3 = *high* chance of dozing	
Sitting and reading	.....
Watching TV	.....
Sitting, inactive in a public place (e.g., a theater or a meeting)	.....
As a passenger in a car for an hour without a break	.....
Lying down to rest in the afternoon when circumstances permit	.....
Sitting and talking to someone Sitting quietly after lunch without alcohol	....
In a car, while stopped for a few minutes in traffic	.....
TOTAL	.....

Source: From MW Johns: Sleep 14:540, 1991.

DIAGNOSIS

OSAHS requires lifelong treatment, and the diagnosis has to be made or excluded with certainty. This will hinge on obtaining a good sleep history from the patient and partner, with both completing sleep questionnaires, including the Epworth Sleepiness Score (Table 265-2). Physical examination must include assessment of obesity, jaw structure, the upper airway, blood pressure, and possible predisposing causes, including hypothyroidism and acromegaly (see above).

In those with appropriate clinical features, the diagnostic test must demonstrate recurrent breathing pauses during sleep. This may be full polysomnography with recording of multiple respiratory and neurophysiologic signals during sleep (Chap. 27). Increasingly, especially outside the United States, most diagnostic tests are "limited studies"—recording respiratory and oxygenation patterns overnight without neurophysiologic recording. Such approaches in expert hands produce good patient outcomes and are cost-effective. It is sensible to use such limited sleep studies as the first-line diagnostic test and then allow positively diagnosed patients to proceed to treatment. A reasonable approach at present is for patients with troublesome sleepiness but negative limited studies to have polysomnography to exclude or confirm OSAHS.

Treatment: Obstructive Sleep Apnea

WHOM TO TREAT

There is evidence from robust randomized controlled trials (RCTs) that treatment improves symptoms, sleepiness, driving, cognition, mood, quality of life, and blood pressure in patients who have an Epworth Score >11, troublesome sleepiness while driving or working, and >15 apneas + hypopneas per hour of sleep. For those with similar degrees of sleepiness and 5–15 events per hour of sleep, RCTs indicate improvements in symptoms, including subjective sleepiness, with less strong evidence indicating gains in cognition and quality of life. There is no evidence of blood pressure improvements in this group. There is no robust evidence that treating nonsleepy subjects improves their symptoms, function, or blood pressure, and so treatment cannot be advocated for this large group, although this may change with further RCTs or less obtrusive therapy.

HOW TO TREAT

All patients diagnosed with OSAHS should have the condition and its significance explained to them and their partners. This should be accompanied by written and/or Web-based information and a discussion of the implications of the local driving regulations. Rectifiable predispositions should be discussed; this often includes weight loss and alcohol reduction both to reduce weight and because alcohol acutely decreases upper-airway dilating muscle tone, thus predisposing to obstructed breathing. Sedative drugs, which also impair airway tone, should be carefully withdrawn.

Continuous Positive Airway Pressure (CPAP)

CPAP therapy works by blowing the airway open during sleep, usually with pressures of 5–20 mmHg. CPAP has been shown in randomized placebo-controlled trials to improve breathing during sleep, sleep quality, sleepiness, blood pressure, vigilance, cognition, and driving ability as well as mood and quality of life in patients with OSAHS. However, this is obtrusive therapy, and care must be taken to explain the need for the treatment to patients and their partners and to intensively support patients on CPAP with telephone or Web support and regular follow-up. Initiation should include finding the most comfortable mask from the ranges of several manufacturers and trying the system for at least 30 min during the day to prepare for the overnight trial. An overnight monitored trial of CPAP is used to identify the pressure required to keep the patient's airway patent. The development of pressure-varying CPAP machines has made an in-lab CPAP night trial unnecessary, but treatment must be initiated in a supportive environment. Thereafter, patients can be treated with fixed-pressure CPAP machines set at the determined pressure or with a self-adjusting intelligent CPAP device. The main side effect of CPAP is airway drying, which can be countered by using an integral heated humidifier. CPAP use is imperfect, but around 94% of patients with severe OSAHS are still using their therapy after 5 years on objective monitoring.

Mandibular Repositioning Splint (Mrs)

Also called oral devices, MRSs work by holding the lower jaw and the tongue forward, thereby widening the pharyngeal airway. MRSs have been shown in RCTs to improve OSAHS patients' breathing during sleep, daytime somnolence, and blood pressure. As there are many devices with differing designs with unknown relative efficacy, these results cannot be generalized to all MRSs. Self-reports of the use of devices long-term suggest high dropout rates.

Surgery

Four forms of surgery have a role in OSAHS, although it must be remembered that these patients have a raised perioperative risk. Bariatric surgery can be curative in the morbidly obese. Tonsillectomy can be highly effective in children but rarely in adults. Tracheostomy is curative but rarely used because of the associated morbidity rate but should not be overlooked in severe cases. Jaw advancement surgery —particularly maxillomandibular osteotomy—is effective in those with retrognathia (posterior displacement of the mandible) and should be considered particularly in young and thin patients. There is no robust evidence that pharyngeal surgery, including uvulopalatopharyngoplasty (whether by scalpel, laser, or thermal techniques) helps OSAHS patients.

Drugs

Unfortunately, no drugs are clinically useful in the prevention or reduction of apneas and hypopneas. A marginal improvement in sleepiness in patients who remain sleepy despite CPAP can be produced by modafinil, but the clinical value is debatable and the financial cost is significant.

CHOICE OF TREATMENT

CPAP and MRS are the two most widely used and best evidence-based therapies. Direct comparisons in RCTs indicate better outcomes with CPAP in terms of apneas and hypopneas, nocturnal oxygenation, symptoms, quality of life, mood, and vigilance. Adherence to CPAP is generally better than that to an MRS, and there is evidence that CPAP improves driving, whereas there are no such data on MRSs. Thus, CPAP is the current treatment of choice. However, MRSs are evidence-based second-line therapy in those who fail CPAP. In younger, thinner patients, maxillomandibular advancement should be considered.

HEALTH RESOURCES

Untreated OSAHS patients are heavy users of health care and dangerous drivers; they also work beneath their potential. Treatment of OSAHS with CPAP is cost-effective in terms of reducing the health care costs of associated illness and associated accidents.

CENTRAL SLEEP APNEA

Central sleep apneas (CSAs) are respiratory pauses caused by lack of respiratory effort. They occur occasionally in normal subjects, particularly at sleep onset and in rapid eye movement (REM) sleep, and are transiently increased after ascent to altitude. Recurrent CSA is most commonly found in the presence of cardiac failure or neurologic disease, especially stroke. Spontaneous central sleep syndrome is rare and can be classified on the basis of the arterial P_{co2}.

Hypercapnic CSA occurs in conjunction with diminished ventilatory drive in Ondine's curse (central

alveolar hypoventilation). Patients with normocapnic spontaneous CSA have a normal or low arterial P_{CO2} when awake, with brisk ventilatory responses to hypercapnia. This combination results in unstable ventilatory control, with subjects breathing close to or below their apneic threshold for P_{CO2} during sleep; this apneic tendency is compounded by cycles of arousal-induced hyperventilation, inducing further hypocapnia.

Clinical Features

Patients may present with sleep maintenance insomnia, which is relatively unusual in OSAHS. Daytime sleepiness may occur.

Investigation

Many apneas previously labeled central because of absent thoracoabdominal movement are actually obstructive, identification of movement being particularly difficult in the very obese. CSA can be identified with certainty only if either esophageal pressure or respiratory muscle electromyography is recorded and shown to be absent during the events.

Treatment: Central Sleep Apnea

Patients with underlying cardiac failure should have their failure treated appropriately. CPAP may improve outcome but is difficult to initiate and has not been shown to improve survival. Patients with spontaneous normocapnic CSA may be treated with acetazolamide. In a minority of patients, CPAP is effective, perhaps because in some patients with OSAHS, pharyngeal collapse initiates reflex inhibition of respiration, and this is prevented by CPAP. Oxygen and nocturnal nasal positive-pressure ventilation also may be tried.

FURTHER READINGS

BRADLEY TD et al: Obstructive sleep apnoea and its cardiovascular consequences. Lancet 373:82, 2009[PMID: 19101028] [Full Text]

ECKERT DJ et al: Central sleep apnea: Pathophysiology and treatment. Chest 131:595, 2007[PMID: 17296668] [Full Text]

ENGLEMAN HM et al: Randomized crossover trial of two treatments for sleep apnea/hypopnea syndrome: Continuous positive airway pressure and mandibular repositioning splint. Am J Respir Crit Care Med 165:855, 2002

MARIN JM et al: Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: An observational study. Lancet 365:1046, 2005[PMID: 15781100] [Full Text]

PACK AL et al: Risk factors for excessive sleepiness in older adults. Ann Neurol 59:893, 2006[PMID: 16718691] [Full Text]

SOMERS VK et al: Sleep apnea and vascular disease. Circulation 118:1080, 2008[PMID: 18725495] [Full Text]

SUNDARAM S et al: Surgery for obstructive sleep apnoea in adults. Cochrane Database Syst Rev 2008, CD001004

WHITELAW WA et al: Clinical usefulness of home oximetry compared with polysomnography for assessment of sleep apnea. Am J Respir Crit Care Med 171:188, 2005[PMID: 15486338] [Full Text]

YAGGI HK et al: Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med 353:2034, 2005[PMID: 16282178] [Full Text]

YOUNG T et al: Sleep disordered breathing and mortality: Eighteen-year follow-up of the Wisconsin sleep cohort. Sleep 31:1071, 2008[PMID: 18714778] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LUNG TRANSPLANTATION: INTRODUCTION

Lung transplantation is a therapeutic consideration for many patients with nonmalignant end-stage lung disease, and it prolongs survival and improves quality of life in appropriately selected recipients. Since 1985 more than 25,000 procedures have been recorded worldwide, and ~2200 transplants have been reported annually in recent years.

Indications

The indications span the gamut of lung diseases. The most common indications in the last few years have been chronic obstructive pulmonary disease (COPD), ~30%; idiopathic pulmonary fibrosis (IPF), ~30%; cystic fibrosis (CF), ~15%; α_1-antitrypsin deficiency emphysema, ~3%; and idiopathic pulmonary arterial hypertension (IPAH), ~2%. Other diseases have accounted for the balance of primary indications, and retransplantation has accounted for ~4% of procedures.

Recipient Selection

Transplantation should be considered when other therapeutic options have been exhausted and when the patient's prognosis is expected to improve as a result of the procedure. Survival rates after transplantation can be compared with predictive indices for the patient's disease, but each patient's clinical course must be integrated into the assessment, too. Moreover, quality of life is a primary motive for transplantation for many patients, and the prospect of improved quality-adjusted survival is often attractive even if the survival advantage itself may be marginal.

Disease-specific consensus guidelines for referring patients for evaluation and for proceeding with transplantation are summarized in **Table 266-1** and are linked to clinical, physiologic, radiographic, and pathologic features that influence the prognosis of the respective diseases. Candidates for lung transplantation are also thoroughly screened for comorbidities that might affect the outcome adversely. Conditions such as systemic hypertension, diabetes mellitus, gastroesophageal reflux, and osteoporosis are not unusual, but if uncomplicated and adequately managed, they do not disqualify patients from transplantation. The upper age limit is ~65–70 years at most centers.

Table 266-1 Disease-Specific Guidelines for Referral and Transplantation

Chronic Obstructive Pulmonary Disease
Referral BODE index >5
Transplantation BODE index 7–10 *or* any of the following criteria: Hospitalization for exacerbation, with $PaCO_2$ >50 mmHg Pulmonary hypertension or cor pulmonale despite oxygen therapy FEV_1<20% with either DLCO <20% or diffuse emphysema
Cystic Fibrosis/Bronchiectasis
Referral FEV_1<30% or rapidly declining FEV_1 Hospitalization in ICU for exacerbation Increasing frequency of exacerbations Refractory or recurrent pneumothorax Recurrent hemoptysis not controlled by bronchial artery embolization
Transplantation

Oxygen-dependent respiratory failure
Hypercapnia
Pulmonary hypertension
Idiopathic Pulmonary Fibrosis
Referral
Pathologic or radiographic evidence of UIP regardless of vital capacity
Transplantation
Pathologic or radiographic evidence of UIP
and
any of the following criteria
DLCO <39%
Decrement in FVC ≥10% during 6 months of follow-up
Decrease in SpO_2 below 88% during a 6-min walk test
Honeycombing on HRCT (fibrosis score >2)
Idiopathic Pulmonary Arterial Hypertension
Referral
NYHA functional class III or IV regardless of therapy
Rapidly progressive disease
Transplantation
Failing therapy with intravenous epoprostenol (or equivalent drug)
Persistent NYHA functional class III or IV on maximal medical therapy
Low (<350 m) or declining 6-min walk test
Cardiac index <2 L/min/m^2
Right atrial pressure >15 mmHg

Abbreviations: BODE, body-mass index (B), airflow obstruction (O), dyspnea (D), exercise capacity (E); FVC, forced vital capacity; FEV_1, forced expiratory volume in 1 s; DLCO, diffusing capacity for carbon monoxide; $PaCO_2$, partial pressure of carbon dioxide in arterial blood; SpO_2, arterial oxygen saturation by pulse oximetry; ICU, intensive care unit; UIP, usual interstitial pneumonitis; HRCT, high-resolution computed tomography; NYHA, New York Heart Association.

Source: Summarized from Orens et al. For BODE index, BR Celli et al: N Engl J Med 350:1005, 2004.

Standard exclusions include HIV infection, chronic active hepatitis B or C infection, uncontrolled or untreatable pulmonary or extrapulmonary infection, uncured malignancy, active cigarette smoking, drug or alcohol dependency, irreversible physical deconditioning, chronic nonadherence with medical care, significant disease of another vital organ (e.g., heart, liver, or kidney), and psychiatric or psychosocial situations that could substantially interfere with posttransplantation management. Other problems that may compromise the outcome constitute relative contraindications. Some typical issues are ventilator-dependent respiratory failure, previous thoracic surgical procedures, obesity, and coronary artery disease. Chronic infection with antibiotic-resistant *Pseudomonas* species, *Burkholderia* species, *Aspergillus* species, or nontuberculous mycobacteria is a unique concern in some patients with CF. The potential impact of these and other factors has to be judged in clinical context to determine an individual candidate's suitability for transplantation.

Waiting List and Organ Allocation

Organ allocation policies are influenced by medical, ethical, geographic, and political factors, with systems varying from country to country. Regardless of the system, potential recipients are placed on a waiting list and must be matched for blood group compatibility and, with some latitude, for lung size with an acceptable donor. Most lungs are procured from deceased donors after brain death, but only ~15–17% of brain-death organ donors yield either one or two lungs suitable for transplantation. Lungs from donors after cardiac death have been utilized to a limited extent.

A priority algorithm for allocating donor lungs was implemented in the United States in 2005. A lung allocation score that is based on the patient's risk of death on the waiting list and likelihood of survival after transplantation determines priority. The score can range from zero to 100, and precedence for transplantation is ranked from highest to lowest scores. Both the lung disease and its severity affect a patient's score; parameters in the score must be updated biannually but can be submitted for

calculation of a new score whenever the patient's condition changes. The median score for all candidates on the waiting list is usually ~34–35, but the distribution of scores tends to be higher among patients with IPF and CF than among patients with COPD and IPAH.

Under this priority system, the median waiting time for transplantation has fallen to <6 months, and the annual number of deaths on the waiting list has dropped by ~50%. The main indication for transplantation has also shifted from COPD to IPF. Overall survival rates in the first two years after transplantation have not changed substantially under this system; however, recipients with lung allocation scores >60 have had lower survival rates in the first two years compared with recipients with lower scores.

Transplant Procedure

Bilateral transplantation is mandatory for CF and other forms of bronchiectasis because the risk of spillover infection from a remaining native lung precludes single-lung transplantation. Heart-lung transplantation is obligatory for Eisenmenger syndrome with complex anomalies that cannot be readily repaired in conjunction with lung transplantation and for concomitant end-stage lung and heart disease. However, cardiac replacement is not necessary for cor pulmonale because right ventricular function will recover when pulmonary vascular afterload is normalized by lung transplantation.

Either bilateral or single-lung transplantation is an option for other diseases unless there is a special consideration, but bilateral transplantation has been utilized increasingly for most indications. Recently, ~65% of procedures in the United States have been bilateral, and in the international registry, ~55% of transplants for COPD, ~50% for IPF, and ~90% for IPAH have been bilateral.

Living donor lobar transplantation has a limited role in adult lung transplantation. It has been performed predominantly in teenagers or young adults with CF, and it usually has been reserved for patients who were unlikely to survive the wait for a deceased organ donor.

Posttransplantation Management

Induction therapy with an antilymphocyte globulin or an interleukin 2 receptor antagonist is utilized by ~50% of centers, and a three-drug maintenance immunosuppressive regimen that includes a calcineurin inhibitor (cyclosporine or tacrolimus), a purine synthesis antagonist (azathioprine or a mycophenolic acid precursor), and prednisone is traditional. Subsequently, other drugs, such as sirolimus, may be substituted into the regimen for various reasons. Prophylaxis for *Pneumocystis jiroveci* pneumonia is standard, and prophylaxis against cytomegalovirus (CMV) infection and fungal infection is part of many protocols. The dose of cyclosporine, tacrolimus, and sirolimus is adjusted by blood-level monitoring. All these agents are metabolized by the hepatic cytochrome P450 system, and interactions with medications that affect this pathway can significantly alter the clearance and blood level of these drugs.

Routine management focuses on monitoring the allograft, regulating immunosuppressive therapy, and detecting problems or complications expeditiously. Regular contact with a nurse coordinator, physician follow-up, chest radiographs, blood tests, and spirometry are customary, and periodic surveillance bronchoscopies are employed in some programs. If recovery is uncomplicated, lung function rapidly improves and then stabilizes by 3–6 months after transplantation. Subsequently, the variation in spirometric measurements is small, and a sustained decline of ≥10–15% signals a potentially significant problem.

Outcomes

SURVIVAL

Major registries publish survival (**Table 266-2**) and other outcomes annually (*www.ishlt.org; www.ustransplant.org*). In the international registry, survival half-life for the main indications is in the range of 4–6 years; however, age and transplant procedure have a significant impact on outcome. For recipients 18–59 years of age, the survival half-life is 5–6 years, but it decreases to 4 years for those 60–65 years old and to 3 years for those >65 years old. Survival over 10 years has been significantly better after bilateral transplantation than after unilateral transplantation for COPD, α_1-antitrypsin deficiency emphysema, IPF, and IPAH.

Table 266-2 Recipient Survival, by Pretransplantation Diagnosis (1990–2006)

		Survival Rate, %				
Diagnosis	***n***	**3 Months**	**1 Year**	**3 Years**	**5 Years**	**10 Years**
Chronic obstructive pulmonary disease	2444	93	85	69	57	31
Bilateral	5316	90	81	63	47	19
Single						
α_1-Antitrypsin deficiency emphysema	956	88	79	67	58	36

Bilateral	969	87	77	61	51	28
Single						
Cystic fibrosis	3275	90	82	66	56	39
Idiopathic pulmonary fibrosis	1290	81	72	59	48	28
Bilateral	2641	85	73	56	43	19
Single						
Idiopathic pulmonary arterial hypertension	710	75	69	59	51	33
Bilateral	260	71	61	51	41	24
Single						
Sarcoidosis	506	83	70	56	51	31

Source: Data from *www.ishlt.org/registries/slides.asp?slides=heartLungRegistry*.

The main sources of perioperative mortality include technical complications of the operation, primary graft dysfunction, and infections. Acute rejection and CMV infection are common problems in the first year, but neither is usually fatal. Beyond the first year, chronic rejection and non-CMV infections cause the majority of deaths.

FUNCTION

Regardless of the disease, successful transplantation impressively restores cardiopulmonary function. After bilateral transplantation, pulmonary function tests are typically normal; after unilateral transplantation, a mild abnormality characteristic of the remaining diseased lung is still apparent. Formal exercise testing usually demonstrates some impairment in maximum work rate and maximum oxygen uptake, but few recipients report any limitation to activities of daily living.

QUALITY OF LIFE

Both overall and health-related quality of life are enhanced. With multidimensional profiles, improvements extend across most domains and are sustained longitudinally unless chronic rejection or another complication develops. Other problems that detract from quality of life include renal dysfunction and drug side effects.

COST

The cost of transplantation depends on the health care system, other health care policies, and economic factors that vary from country to country. In the United States in 2008 the average billed charge per transplant for the period 30 days before transplantation through 180 days after the transplant admission was $450,400 for single-lung transplantation and $657,800 for bilateral lung transplantation. For bilateral transplantation, the total cost included the following charges: all care during 30 days before transplantation, $20,700; donor organ procurement, $96,500; hospital transplant admission, $344,700; physician fees during transplant admission, $59,300; all inpatient and outpatient care during 180 days after transplant admission, $113,800; and all outpatient drugs, including immunosuppressants, from discharge for the transplant to 180 days after transplant admission, $22,800.

COMPLICATIONS

Lung transplantation can be complicated by a variety of problems (**Table 266-3**). Aside from predicaments that are unique to transplantation, side effects and toxicities of the immunosuppressive medications can cause new medical problems or aggravate preexisting conditions.

Table 266-3 Major Potential Complications of Lung Transplantation and Immunosuppression

Category	Complication
Allograft	Primary graft dysfunction; anastomotic dehiscence or stenosis; ischemic airway injury with bronchostenosis or bronchomalacia; rejection; infection; recurrence of primary disease (sarcoidosis, lymphangioleiomyomatosis, giant cell interstitial pneumonitis, diffuse panbronchiolitis, pulmonary alveolar proteinosis, Langerhans cell histiocytosis)
Thoracic	Phrenic nerve injury—diaphragmatic dysfunction; recurrent laryngeal nerve injury—vocal cord dysfunction; cervical ganglia injury—Horner's syndrome; pneumothorax; pleural effusion; chylothorax; empyema
Cardiovascular	Intraoperative or perioperative air embolism; postoperative pericarditis; perioperative myocardial injury/infarction; venous thromboembolism; supraventricular dysrhythmias; systemic hypertension
Gastrointestinal	Esophagitis [especially *Candida*, herpes, or cytomegalovirus (CMV)]; gastroparesis;

	gastroesophageal reflux; diarrhea (*Clostridium difficile*; medications, especially mycophenolate mofetil and sirolimus); colitis (*C. difficile*; CMV?)
Hepatobiliary	Hepatitis (especially CMV or medications); acalculous cholecystitis
Renal	Calcineurin inhibitor nephropathy; hemolytic-uremic syndrome (thrombotic microangiopathy)
Neurologic	Tremors; seizures; reversible posterior leukoencephalopathy; headaches
Musculoskeletal	Steroid myopathy; rhabdomyolysis (cyclosporine + HMG-CoA reductase inhibitor treatment); osteoporosis; avascular necrosis
Metabolic	Obesity; diabetes mellitus; hyperlipidemia; idiopathic hyperammonemia
Hematologic	Anemia; leukopenia; thrombocytopenia; thrombotic microangiopathy
Oncologic	Lymphoproliferative disease and lymphoma; skin cancers; other malignancies

GRAFT DYSFUNCTION

Primary graft dysfunction (PGD) is an acute lung injury that is a manifestation of multiple potential insults to the donor organ inherent in the transplantation process. The principal clinical features are diffuse pulmonary infiltrates and hypoxemia within 72 h of transplantation; however, the presentation can be mimicked by pulmonary venous obstruction, hyperacute rejection, pulmonary edema, and pneumonia.

The severity is variable, and a standardized grading system has been established. Up to 50% of recipients may have some degree of PGD, and ~10–20% have severe PGD. The treatment follows the conventional, supportive paradigm for acute lung injury. Inhaled nitric oxide and extracorporeal membrane oxygenation have been used in severe cases; retransplantation also has been performed, but retransplantation in the first 30 days has a poor survival rate (~30% at 1 year). Most recipients with mild PGD recover, but the mortality rate for severe PGD has been ~40–60%. PGD also is associated with longer postoperative ventilator support, longer intensive care unit and hospital stays, higher costs, and excess morbidity rates and severe PGD is probably a risk factor for the later development of chronic rejection.

AIRWAY COMPLICATIONS

The bronchial blood supply to the donor lung is disrupted during procurement. Bronchial revascularization during transplantation is technically feasible in some cases, but it is not widely practiced. Consequently, after implantation the donor bronchus is dependent on retrograde bronchial blood flow from the pulmonary circulation and is vulnerable to ischemia.

The spectrum of airway problems includes anastomotic necrosis and dehiscence, occlusive granulation tissue, anastomotic or bronchial stenosis, and bronchomalacia. The incidence has been in the range of 7–18%, but the associated mortality rate has been low. These problems usually can be managed bronchoscopically with techniques such as simple endoscopic debridement, laser photoresection, balloon dilatation, and bronchial stenting.

REJECTION

Rejection is the main limitation to better medium- and long-term survival. It is an immunologic response to alloantigen recognition, and both cell-mediated and antibody-mediated (humoral) cascades can play a role. Cellular rejection is effected by T lymphocyte interactions with donor alloantigens, mainly in the major histocompatibility complex (MHC), whereas humoral rejection is driven by antibodies to donor MHC alloantigens or possibly to non-MHC antigens on epithelial or endothelial cells.

Rejection often is categorized as acute or chronic without mention of the mechanism. Acute rejection is cell-mediated, and its incidence is highest in the first 6–12 months after transplantation. In contrast, chronic rejection generally emerges later, and both alloimmune and nonalloimmune fibroproliferative reactions may contribute to its pathogenesis.

ACUTE CELLULAR REJECTION

With current immunosuppressive regimens, ~25–40% of recipients have acute rejection in the first year. Acute cellular rejection (ACR) can be clinically silent, or it can be manifested by nonspecific symptoms or signs that may include cough, low-grade fever, dyspnea, hypoxemia, inspiratory crackles, interstitial infiltrates, and declining lung function; however, clinical impression is not reliable. The diagnosis is confirmed by transbronchial biopsies showing the characteristic lymphocytic infiltrates around arterioles or bronchioles, and a standardized pathologic scheme is used to grade the biopsies.

Minimal ACR on a surveillance biopsy in a clinically stable recipient often is left untreated, but higher grades generally are treated regardless of the clinical situation. Treatment usually includes a short course of high-dose steroid therapy and adjustment of the maintenance immunosuppressive regimen. Most episodes respond to this approach; however, more intensive therapy is sometimes necessary for persistent or recurrent episodes.

CHRONIC REJECTION

This complication is the main impediment to better long-term survival rates, and it is the source of substantial morbidity rates because of its impact on lung function and quality of life. Clinically, it is characterized physiologically by airflow limitation and pathologically by bronchiolitis obliterans; the process is denoted bronchiolitis obliterans syndrome (BOS). Transbronchial biopsies are relatively insensitive for detecting bronchiolitis obliterans, and pathologic confirmation is not required for diagnosis. Thus, after other causes of graft dysfunction have been excluded, the diagnosis of BOS is based primarily on a sustained decrement (≥20%) in forced expiratory volume in 1s (FEV_1), although smaller declines in FEV_1 (≥10%) or in forced expiratory flow $FEF_{25–75\%}$ may presage BOS. Spirometric criteria for diagnosis and staging of BOS have been standardized.

The prevalence of BOS approaches 50% by 5 years after transplantation. Antecedent ACR is the main risk factor, but PGD, CMV pneumonitis, other community-acquired respiratory viral infections, and gastroesophageal reflux have been implicated as well. BOS can present acutely and imitate infectious bronchitis, or it can manifest as an insidious decline in lung function. The chest radiograph is typically unchanged; computed tomography may reveal mosaic perfusion, air trapping, ground-glass opacities, or bronchiolectasis. Bronchoscopy is indicated to eliminate other processes, but transbronchial biopsies identify bronchiolitis obliterans in a minority of cases.

BOS usually is treated with augmented immunosuppression, but there is no consensus about therapy. Strategies include changes in the maintenance drug regimen, including the addition of azithromycin, antilymphocyte globulin, photopheresis, and total lymphoid irradiation. Although therapy may stabilize lung function, the overall results of treatment have been disappointing; median survival after onset has been ~3–4 years. Retransplantation is a consideration if clinical circumstances and other comorbidities are not prohibitive, but survival has been inferior to that with primary transplantation.

HUMORAL REJECTION

The role of antibody-mediated rejection is still evolving. Hyperacute rejection is caused by preformed human leukocyte (HLA) antibodies in the recipient, but it is minimized by pretransplantation antibody screening coupled with virtual or direct cross-matching with any potential donor. Donor-specific HLA antibodies develop after transplantation in up to 50% of recipients, and their presence has been associated with an increased risk of both ACR and BOS and with poorer overall survival. However, the mechanisms by which these antibodies could contribute to ACR or BOS or could otherwise be detrimental have not been unraveled. Formal criteria for antibody-mediated rejection have been defined for renal transplantation, but few cases in lung transplantation fulfill them. Nonetheless, episodes of acute lung allograft dysfunction occasionally have been attributed directly to antibody-mediated injury. If treatment is indicated, therapies that may deplete antibodies include plasmapheresis, intravenous immune globulin, and rituximab.

INFECTION

The lung allograft is especially susceptible to infection, which has been one of the leading causes of death. In addition to a blunted immune response from the immunosuppressive drugs, other normal defenses are compromised: the cough reflex is diminished, and mucociliary clearance is impaired in the transplanted lung. The spectrum of infections includes both opportunistic and nonopportunistic pathogens.

Bacterial bronchitis or pneumonia can occur at any time, but it is very common in the perioperative period. Later, bronchitis occurs frequently in recipients with BOS, and *Pseudomonas aeruginosa* or methicillin-resistant *Staphylococcus aureus* is often the culprit.

CMV is the most common viral infection. Although gastroenteritis, colitis, and hepatitis can occur, CMV viremia and CMV pneumonia are the main illnesses. Most episodes occur in the first 6 months, and treatment with ganciclovir is effective unless resistance develops. Other community-acquired viruses such as influenza, parainfluenza, and respiratory syncytial virus also contribute to respiratory complications. The most problematic fungal infections are caused by *Aspergillus* species. The spectrum encompasses simple pulmonary colonization, tracheobronchitis, invasive pulmonary aspergillosis, and disseminated aspergillosis, and the clinical scenario dictates treatment.

OTHER COMPLICATIONS

Other potential complications are listed in Table 266-3. Many of them are related to side effects or toxicities of the immunosuppressive drugs. Management of these general medical problems is guided by standard practices, but the complex milieu of transplantation requires close collaboration and good communication among health care providers.

FURTHER READINGS

BELPERIO JA et al: Chronic lung allograft rejection: Mechanisms and therapy. Proc Am Thorac Soc 6:108, 2009[PMID: 19131536] [Full Text]

BHORADE SM, STERN E: Immunosuppression for lung transplantation. Proc Am Thorac Soc 6:47, 2009[PMID: 19131530] [Full Text]

BOASQUEVISQUE CHR et al: Surgical techniques: Lung transplant and lung volume reduction. Proc Am Thorac Soc 6:66, 2009[PMID: 19131532] [Full Text]

CHAN KM et al: Nonmedical therapy for chronic obstructive pulmonary disease. Proc Am Thorac Soc 6:137, 2009[PMID: 19131539] [Full Text]

CHRISTIE JD et al: The Registry of the International Society for Heart and Lung Transplantation: Twenty-sixth Official Adult Lung and Heart-Lung Transplantation Report-2009. J Heart Lung Transplant 28:1031, 2009[PMID: 19782285] [Full Text]

KREIDER M, KOTLOFF RM: Selection of candidates for lung transplantation. Proc Am Thorac Soc 6:20, 2009[PMID: 19131527] [Full Text]

LEE JC, CHRISTIE JD: Primary graft dysfunction. Proc Am Thorac Soc 6: 39, 2009[PMID: 20046372] [Full Text]

LYU DM, ZAMORA MR: Medical complications of lung transplantation. Proc Am Thorac Soc 6:101, 2009[PMID: 19131535] [Full Text]

MARTINU T et al: Acute rejection and humoral sensitization in lung transplant recipients. Proc Am Thorac Soc 6:54, 2009[PMID: 19131531] [Full Text]

ORENS JB, GARRITY ER JR: General overview of lung transplantation and review of organ allocation. Proc Am Thorac Soc 6:13, 2009[PMID: 19131526] [Full Text]

———— et al: International guidelines for the selection of lung transplant candidates: 2006 update—a consensus report from the pulmonary scientific council of the International Society for Heart and Lung Transplantation. J Heart Lung Transplant 25:745, 2006

REMUND KF et al: Infections relevant to lung transplantation. Proc Am Thorac Soc 6:94, 2009[PMID: 19131534] [Full Text]

SANTACRUZ JF, MEHTA AC: Airway complications and management after lung transplantation: Ischemia, dehiscence, and stenosis. Proc Am Thorac Soc 6:79, 2009[PMID: 19131533] [Full Text]

SWEET SC: Pediatric lung transplantation. Proc Am Thorac Soc 6:122, 2009[PMID: 19131537] [Full Text]

VAN RAEMDONCK D et al: Lung donor selection and management. Proc Am Thorac Soc 6:28, 2009

YUSEN RD: Technology and outcomes assessment in lung transplantation. Proc Am Thorac Soc 6:128, 2009[PMID: 19131538] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH CRITICAL ILLNESS: INTRODUCTION

The care of critically ill patients requires a thorough understanding of pathophysiology and is centered initially on resuscitation of patients at extremes of physiologic deterioration. This resuscitation is often fast-paced and occurs early without a detailed awareness of the patients chronic medical problems. While physiologic stabilization is taking place, intensivists attempt to gather important background medical information to supplement the real-time assessment of the patients current physiologic conditions. Numerous tools are available to assist intensivists in the accurate assessment of pathophysiology and management to incipient organ failure, offering a window of opportunity for diagnosing and treating underlying disease(s) in a stabilized patient. Indeed, the use of invasive interventions such as mechanical ventilation and renal replacement therapy is commonplace in the intensive care unit. An appreciation of the risks and benefits of such aggressive and often invasive interventions is vital to assure an optimal patient outcome. Nonetheless, intensivists must recognize when patients chances for recovery are remote or impossible and counsel and comfort dying patients and their significant others. Critical care physicians often must redirect the goals of care from resuscitation and cure to comfort when the resolution of an underlying illness is not possible.

Assessment of Severity of Illness

Categorization of a patients illness into grades of severity occurs frequently in the intensive care unit (ICU). Numerous severity-of-illness (SOI) scoring systems have been developed and validated over the last two decades. Although these scoring systems have been validated as tools to assess populations of critically ill patients, their utility in predicting individual patient outcomes is not clear.

SOI scoring systems are important for defining populations of critically ill patients. This allows effective comparison of groups of patients enrolled in clinical trials. To be assured that a purported benefit of a therapy is real, investigators must be assured that different groups involved in a clinical trial have similar illness severities. SOI scores are also useful in guiding hospital administrative policies. Allocation of resources such as nursing and ancillary care can be directed by such scoring systems. SOI scoring systems also can assist in the assessment of quality of ICU care over time. Scoring system validations are based on the premise that increasing age, the presence of chronic medical illnesses, and increasingly severe derangements from normal physiology are associated with increased mortality rates. All currently existing SOI scoring systems are derived from patients who already have been admitted to the ICU. There are no established scoring systems available that purport to direct clinicians decision-making regarding criteria for admission to an ICU.

Currently, the most commonly utilized scoring systems are the APACHE (acute physiology and chronic health evaluation) system and the SAPS (simplified acute physiology score) system. These systems were designed to predict outcomes in critical illness and use common variables that include age; vital signs; assessments of respiratory, renal, and neurologic function; and an evaluation of chronic medical illnesses.

Apache II Scoring System

The APACHE II system is the most commonly used SOI scoring system in North America. Age, type of ICU admission (after elective surgery vs. nonsurgical or after emergency surgery), a chronic health problem score, and 12 physiologic variables (the most severely abnormal of each in the first 24 h of ICU admission) are used to derive a score. The predicted hospital mortality is derived from a formula that takes into account the APACHE II score, the need for emergency surgery, and a weighted, disease-specific diagnostic category **(Table 267-1)**. The relationship between APACHE II score and mortality is illustrated in **Fig. 267-1**. Updated versions of the APACHE scoring system (APACHE III and APACHE IV) have been published. APACHE III is derived from a larger database than APACHE II and utilizes a daily clinical update protocol to provide daily modification of predicted mortality. APACHE IV uses a modified statistical model of logistic regression; it is the most recently released version of this scoring system.

Table 267–1. Calculation of Acute Physiology and Chronic Health Evaluation II (APACHE II)[a]

Score	4	3	2	1	0	1	2	3	4
Rectal temperature, °C	≥41	39.0–40.9		38.5–38.9	36.0–38.4	34.0–35.9	32.0–33.9	30.0–31.9	≤29.9

Mean blood pressure, mmHg	≥160	130–159	110–129		70–109		50–69		≤49
Heart rate	≥180	140–179	110–139		70–109		55–69	40–54	≤39
Respiratory rate	≥50	35–49		25–34	12–24	10–11	6–9		≤5
Arterial pH	≥7.70	7.60–7.69		7.50–7.59	7.33–7.49		7.25–7.32	7.15–7.24	<7.15
Oxygenation									
If FI_{O2} > 0.5, use (A - a) D_{O2}	≥500	350–499	200–349		<200				
If F_{O2} ≤0.5, use Pa_{O2}					≥70	61–70		55–60	<55
Serum sodium, meq/L	≥180	160–179	155–159	150–154	130–149		120–129	111–119	≤110
Serum potassium, meq/L	≥7.0	6.0–6.9		5.5–5.9	3.5–5.4	3.0–3.4	2.5–2.9		<2.5
Serum creatinine, mg/dL	≥3.5	2.0–3.4	1.5–1.9		0.6–1.4		<0.6		
Hematocrit	≥60		50–59.9	46–49.9	30–45.9		20–29.9		<20
WBC count, 10^3/mL	≥40		20–39.9	15–19.9	3–14.9		1–2.9		<1

Glasgow Coma Score[b,c]

Eye Opening	**Verbal (Nonintubated)**	**Verbal (Intubated)**	**Motor Activity**
4—Spontaneous	5—Oriented and talks	5—Seems able to talk	6—Verbal command
3—Verbal stimuli	4—Disoriented and talks	3—Questionable ability to talk	5—Localizes to pain
2—Painful stimuli	3—Inappropriate words	1—Generally unresponsive	4—Withdraws to pain
1—No response	2—Incomprehensible sounds		3—Decorticate
	1—No response		2—Decerebrate
			1—No response

Points Assigned to Age and Chronic Disease as Part of the APACHE II Score

Age, Years	**Score**
<45	0
45–54	2
55–64	3
65–74	5
≥75	6
Chronic Health (History of Chronic Conditions)[d]	**Score**
None	0
If patient is admitted after elective surgery	2
If patient is admitted after emergency	5

surgery or for reason other than after elective surgery	

a APACHE II score is the sum of the acute physiology score (vital signs, oxygenation, laboratory values), Glasgow coma score, age, and chronic health points. Worst values during first 24 h in the ICU should be used.

b Glasgow coma score (GCS) = eye-opening score + verbal (intubated or nonintubated) score + motor score.

c For GCS component of acute physiology score, subtract GCS from 15 to obtain points assigned.

d Chronic health conditions: liver, cirrhosis with portal hypertension or encephalopathy; cardiovascular, class IV angina (at rest or with minimal self-care activities); pulmonary, chronic hypoxemia or hypercapnia, polycythemia, ventilator dependent; kidney, chronic peritoneal or hemodialysis; immune, immunocompromised host.

Note: (A - a) D_{O2}, alveolar-arterial oxygen difference; WBC, white blood (cell) count.

Figure 267-1

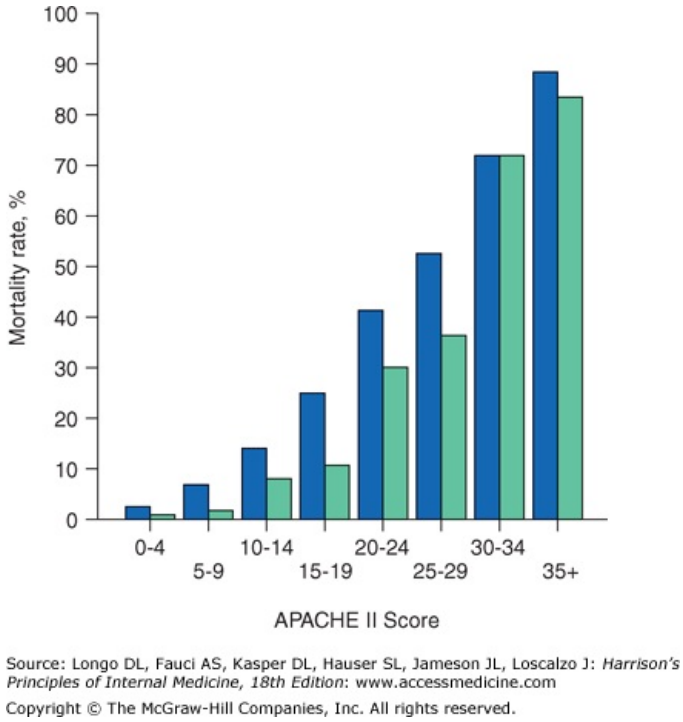

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

PACHE II survival curve. Blue, nonoperative; green, postoperative.

The Saps Scoring System

The SAPS II score, used more frequently in Europe, was derived in a manner similar to the APACHE scores. This score is not disease-specific but, rather, incorporates three underlying disease variables (AIDS, metastatic cancer, and hematologicmalignancy).

Severity of illness scoring systems cannot be used to predict survival in individual patients. Accordingly, the use of these scoring systems to direct therapy and clinical decision-making cannot be recommended at present. Instead, these tools should be used as important data to complement clinical bedside decision-making.

SHOCK

(See also Chap. 270)

Initial Evaluation

Shock is a common condition necessitating admission to the ICU or occurring in the course of critical care. Shock is defined by the presence of multisystem end organ hypoperfusion. Clinical indicators include reduced mean arterial pressure (MAP), tachycardia, tachypnea, cool skin and extremities, acute altered mental status, and oliguria. Hypotension is usually, though not always, present. The end result of multiorgan hypoperfusion is tissue hypoxia, often clinically manifested by lactic acidosis. Since the MAP is the product of cardiac output and systemic vascular resistance (SVR), reductions in blood

pressure can be caused by decreased cardiac output and/or decreased SVR. Accordingly, the initial evaluation of a hypotensive patient should include an assessment of the adequacy of cardiac output; this should be part of the earliest assessment of the patient by the clinician at the bedside once shock is contemplated **(Fig. 267-2)**. Clinical evidence of *diminished* cardiac output includes a narrow pulse pressure—a marker that correlates with stroke volume—and cool extremities with delayed capillary refill. Signs of *increased* cardiac output include a widened pulse pressure (particularly with a reduced diastolic pressure), warm extremities with bounding pulses, and rapid capillary refill. If a hypotensive patient has clinical signs of increased cardiac output, one can infer that the reduced blood pressure is a result of decreased SVR.

Figure 267-2

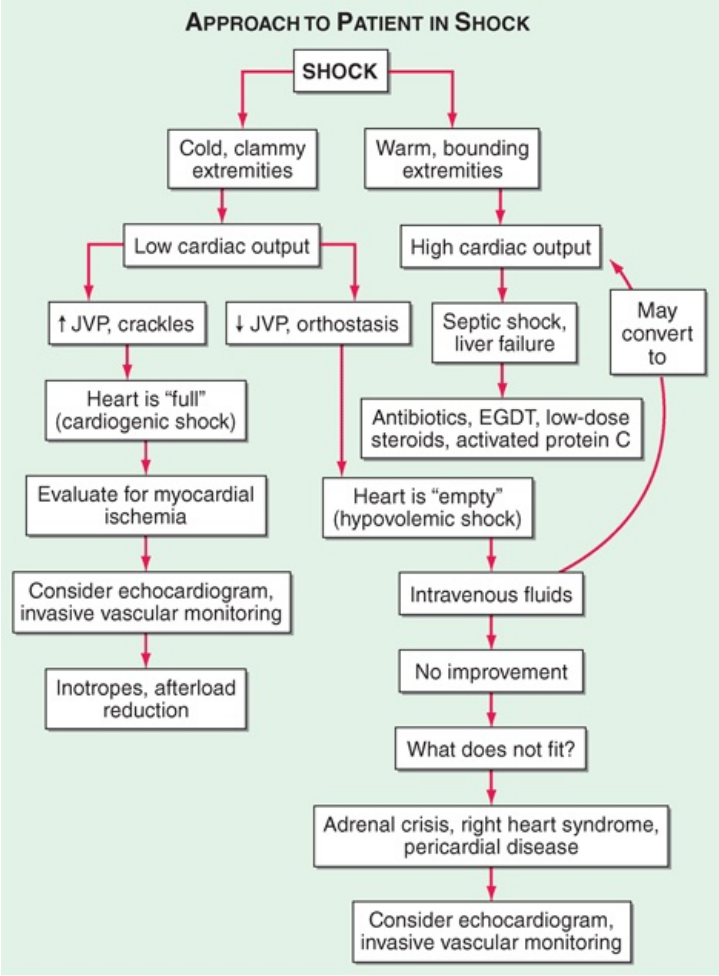

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to patient in shock. EGDT, early goal-directed therapy; JVP, jugular venous pulse.

In hypotensive patients with signs of a reduced cardiac output, an assessment of intravascular and cardiac volume status is appropriate. A hypotensive patient with decreased intravascular volume status may have a history suggesting hemorrhage or other volume losses (e.g., vomiting, diarrhea, polyuria). The jugular venous pressure (JVP) may be reduced in such a patient, although the change in right atrial pressure as a function of spontaneous respiration is a better predictor of fluid responsiveness **(Fig. 267-3)**. Patients with fluid-responsive (i.e., hypovolemic) shock also may manifest large changes in pulse pressure as a function of respiration during positive-pressure mechanical ventilation **(Fig. 267-4)**. A hypotensive patient with increased intravascular volume status and cardiac dysfunction may have S_3 and/or S4 gallops on examination, increased JVP, extremity edema, and crackles on lung auscultation. The chest x-ray may show cardiomegaly, widening of the vascular pedicle, Kerley B lines, and pulmonary edema. Chest pain and electrocardiographic changes consistent with ischemia may be noted (Chap. 272).

Figure 267-3.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Right atrial pressure change during spontaneous respiration in a patient with shock who will increase cardiac output in response to intravenous fluid administration. The right atrial pressure decreases from 7 mmHg to 4 mmHg. The horizontal bar marks the time of spontaneous inspiration.

Figure 267-4.

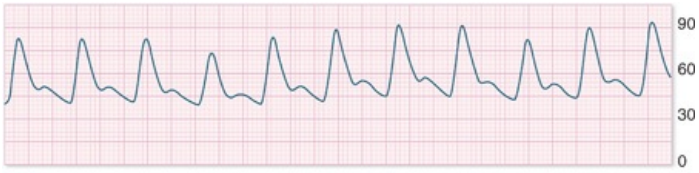

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pulse pressure change during mechanical ventilation in a patient with shock who will increase cardiac output in response to intravenous fluid administration. The pulse pressure (systolic minus diastolic blood pressure) changes during mechanical ventilation in a patient with septic shock.

In hypotensive patients with clinical evidence of increased cardiac output, a search for causes of decreased SVR is appropriate. The most common cause of high cardiac output hypotension is sepsis (Chap. 271). Other causes of high cardiac output hypotension include liver failure, severe pancreatitis, burns and other trauma that elicit the systemic inflammatory response syndrome (SIRS), anaphylaxis, thyrotoxicosis, and peripheral arteriovenous shunts.

In summary, the most common categories of shock are hypovolemic, cardiogenic, and high cardiac output with decreased SVR (high-output hypotension). Certainly these categories may overlap and occur simultaneously (e.g., hypovolemic and septic shock).

The initial assessment of a patient in shock as outlined above should take only a few minutes. It is important that aggressive, early resuscitation is instituted based on the initial assessment, particularly since early resuscitation of septic and cardiogenic shock may improve survival (see below). If the initial bedside assessment yields equivocal or confounding data, more objective assessments such as echocardiography and/or invasive vascular monitoring may be useful. The goal of early resuscitation is to reestablish adequate tissue perfusion to prevent or minimize end organ injury.

Mechanical Ventilatory Support

(See also Chap. 269) During the initial resuscitation of patients in shock, principles of advanced cardiac life support should be followed. Since patients in shock may be obtunded and unable to protect the airway, an early assessment of the patients airway is mandatory during resuscitation from shock. Early intubation and mechanical ventilation often are required. Reasons for the institution of endotracheal intubation and mechanical ventilation include acute hypoxemic respiratory failure and ventilatory failure, which frequently accompany shock. Acute hypoxemic respiratory failure may occur in patients with cardiogenic shock and pulmonary edema (Chap. 272) as well as in those in septic shock with pneumonia or acute respiratory distress syndrome (ARDS) (Chaps. 268 and 271). Ventilatory failure often occurs as a result of an increased load on the respiratory system. This load may present in the form of acute metabolic acidosis (often lactic acidosis) or decreased compliance of the lungs ("stiff" lungs) as a result of pulmonary edema. Inadequate perfusion to respiratory muscles in the setting of shock may be another reason for early intubation and mechanical ventilation. Normally, the respiratory muscles receive a very small percentage of the cardiac output. However, in patients who are in shock with respiratory distress for the reasons listed above, the percentage of cardiac output dedicated to respiratory muscles may increase tenfold or more. Lactic acid production from inefficient respiratory muscle activity presents an additional ventilatory load.

Mechanical ventilation may relieve the patient of the work of breathing and allow redistribution of a limited cardiac output to other vital organs, often with an improvement in lactic acidosis. Patients demonstrate signs of respiratory distress with a number of clinical signs, including inability to speak full sentences, accessory use of respiratory muscles, paradoxical abdominal muscle activity, extreme

tachypnea (>40 breaths/min), and decreasing respiratory rate despite an increasing drive to breathe. When patients with shock are treated with mechanical ventilation, a major goal of ventilator settings is to assume all or the majority of work of breathing, facilitating a state of minimal respiratory muscle work. With the institution of mechanical ventilation for shock, further declines in MAP are frequently seen. The reasons for this include impeded venous return with positive-pressure ventilation, reduced endogenous catecholamine secretion once the stress associated with respiratory failure abates, and the actions of drugs used to facilitate endotracheal intubation (e.g., barbiturates, benzodiazepines, opiates), all of which may result in hypotension. Accordingly, hypotension should be anticipated after endotracheal intubation and positive-pressure ventilation. Many of these patients have a component of hypovolemia, which may respond to IV volume administration. Fig. 267-2 summarizes the diagnosis and treatment of different types of shock. **For further discussion of individual forms of shock, see Chaps. 270, 271, and 272.**

RESPIRATORY FAILURE

Respiratory failure is one of the most common reasons patients are admitted to the ICU. In some ICUs, ≥75% of patients require mechanical ventilation during their stay. Respiratory failure can be categorized mechanistically, based on pathophysiologic derangements in respiratory function. Accordingly, four different types of respiratory failure can be described, based on these pathophysiologic derangements.

TYPE I, ACUTE HYPOXEMIC RESPIRATORY FAILURE

This occurs when alveolar flooding and subsequent intrapulmonary shunt physiology occur. Alveolar flooding may be a consequence of pulmonary edema, pneumonia, or alveolar hemorrhage. Pulmonary edema can be further categorized as occurring due to elevated pulmonary microvascular pressures as seen in heart failure and intravascular volume overload or ARDS ("low-pressure pulmonary edema"; Chap. 268) and represents an extreme degree of lung injury. This syndrome is defined by diffuse bilateral airspace edema seen on chest radiography, the absence of left atrial hypertension, and profound shunt physiology (**Fig. 267-5**) in a clinical setting in which this syndrome is known to occur, including sepsis, gastric aspiration, pneumonia, near-drowning, multiple blood transfusions, and pancreatitis. The mortality rate of patients with ARDS was traditionally very high (50–70%), although recent changes in ventilator management strategy have led to reports of mortality rates closer to 30% (see below).

Figure 267-5.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiograph of a patient with ARDS. ARDS, acute respiratory distress syndrome.

For many years, physicians have suspected that mechanical ventilation of patients with acute lung injury and ARDS may propagate lung injury. Cyclical collapse and reopening of alveoli may be partly

responsible for this. As seen in **Fig. 267-6**, the pressure-volume relationship of the lung in ARDS is not linear. Alveoli may collapse at very low lung volumes. Animal studies have suggested that stretching and overdistention of injured alveoli during mechanical ventilation can further injure the lung. Concern over this alveolar overdistention, termed *ventilator-induced "volutrauma,"* led to a multicenter, randomized, prospective trial to compare traditional ventilator strategies for acute lung injury and ARDS (large tidal volume—12 mL/kg ideal body weight) to a low tidal volume (6 mL/kg ideal body weight). This study showed a dramatic reduction in mortality rate in the low tidal volume group (large tidal volume—39.8% mortality rate versus low tidal volume—31% mortality rate) and confirmed that ventilator management could affect outcomes in these patients. In addition, a "fluid conservative" management strategy [maintaining a relatively low central venous pressure (CVP) or pulmonary capillary wedge pressure (PCWP)] is associated with the need for fewer days of mechanical ventilation compared with a "fluid liberal" management strategy (maintaining a relatively high CVP or PCWP) in acute lung injury and ARDS.

Figure 267-6.

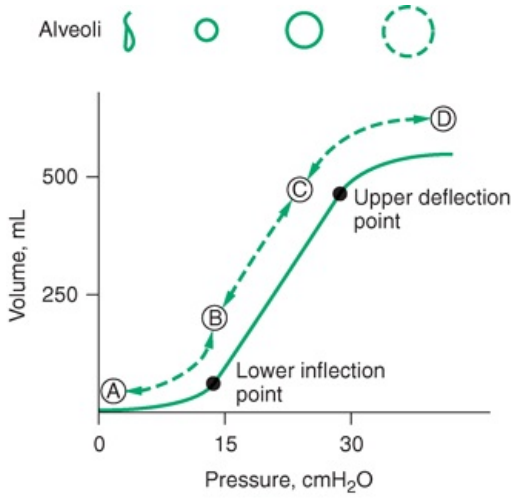

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pressure-volume relationship of the lungs of a patient with ARDS. At the lower inflection point, collapsed alveoli begin to open, and the lung compliance changes. At the upper deflection point, alveoli become overdistended. The shape and size of alveoli are illustrated at the top. ARDS, acute respiratory distress syndrome.

TYPE II RESPIRATORY FAILURE

This type of respiratory failure occurs as a result of alveolar hypoventilation and results in the inability to eliminate carbon dioxide effectively. Mechanisms by which this occurs are categorized by impaired central nervous system (CNS) drive to breathe, impaired strength with failure of neuromuscular function in the respiratory system, and increased load(s) on the respiratory system. Reasons for diminished CNS drive to breathe include drug overdose, brainstem injury, sleep-disordered breathing, and hypothyroidism. Reduced strength can be due to impaired neuromuscular transmission (e.g., myasthenia gravis, Guillain-Barré syndrome, amyotrophic lateral sclerosis, phrenic nerve injury) or respiratory muscle weakness (e.g., myopathy, electrolyte derangements, fatigue).

The overall load on the respiratory system can be subclassified into increased resistive loads (e.g., bronchospasm), loads due to reduced lung compliance [e.g., alveolar edema, atelectasis, intrinsic positive end-expiratory pressure (autoPEEP)—see below], loads due to reduced chest wall compliance (e.g., pneumothorax, pleural effusion, abdominal distention), and loads due to increased minute ventilation requirements (e.g., pulmonary embolus with increased dead space fraction, sepsis).

The mainstays of therapy for type II respiratory failure are treatments directed at reversing the underlying cause(s) of ventilatory failure. Noninvasive positive-pressure ventilation using a mechanical ventilator with a tight-fitting face or nasal mask that avoids endotracheal intubation often can stabilize these patients. This approach has been shown to be beneficial in treating patients with exacerbations of chronic obstructive pulmonary disease. Noninvasive ventilation has been tested less extensively in other types of type II respiratory failure but may be attempted nonetheless in the absence of contraindications (hemodynamic instability, inability to protect airway, respiratory arrest).

TYPE III RESPIRATORY FAILURE

This form of respiratory failure occurs as a result of lung atelectasis. Because atelectasis occurs so commonly in the perioperative period, this is also called perioperative respiratory failure. After general

anesthesia, decreases in functional residual capacity lead to collapse of dependent lung units. Such atelectasis can be treated by frequent changes in position, chest physiotherapy, upright positioning, and aggressive control of incisional and/or abdominal pain. Noninvasive positive-pressure ventilation may also be used to reverse regional atelectasis.

TYPE IV RESPIRATORY FAILURE

This form results from hypoperfusion of respiratory muscles in patients in shock. Normally, respiratory muscles consume <5% of the total cardiac output and O_2 delivery. Patients in shock often experience respiratory distress due to pulmonary edema (e.g., patients in cardiogenic shock), lactic acidosis, and anemia. In this setting, up to 40% of the cardiac output may be distributed to the respiratory muscles. Intubation and mechanical ventilation can allow redistribution of the cardiac output away from the respiratory muscles and back to vital organs while the shock is treated.

CARE OF THE MECHANICALLY VENTILATED PATIENT

(See also Chap. 269) Whereas a thorough understanding of the pathophysiology of respiratory failure is essential to optimize patient care, recognition of a patients readiness to be liberated from mechanical ventilation is similarly important. Several studies have shown that subjecting patients to daily spontaneous breathing trials can identify those ready for extubation. Accordingly, all intubated, mechanically ventilated patients should undergo a daily screening of respiratory function. If oxygenation is stable (i.e., Pa_{O2}/FI_{O2} >200 and PEEP ≤5 cmH_2O), cough and airway reflexes are intact, and no vasopressor agents or sedatives are being administered, the patient has passed the screening test and should undergo a spontaneous breathing trial. This trial consists of a period of breathing through the endotracheal tube without ventilator support [both continuous positive airway pressure (CPAP) of 5 cmH_2O and an open T-piece breathing system can be used] for 30–120 min. The spontaneous breathing trial is declared a failure and stopped if *any* of the following occur: (1) respiratory rate >35/min for >5 min, (2) O_2 saturation <90%, (3) heart rate >140/min or a 20% increase or decrease from baseline, (4) systolic blood pressure <90 mmHg or >180 mmHg, (5) increased anxiety or diaphoresis. If, at the end of the spontaneous breathing trial, the ratio of the respiratory rate and tidal volume in liters (f/V_T) is <105, the patient can be extubated. Such protocol-driven approaches to patient care can have an important impact on the duration of mechanical ventilation and length of stay in the ICU. In spite of such a careful approach to liberation from mechanical ventilation, up to 10% of patients will develop respiratory distress after extubation and may require resumption of mechanical ventilation. Many of these patients will require reintubation. The use of noninvasive ventilation in patients who fail extubation may be associated with worse outcomes compared with immediate reintubation.

Mechanically ventilated patients frequently require sedatives and analgesics. Most patients undergoing mechanical ventilation experience pain, which can be elicited by the presence of the endotracheal tube and endotracheal suctioning. Accordingly, early attention to pain control is extremely important. Opiates are the mainstay of therapy for pain control in mechanically ventilated patients. After adequate pain control has been assured, additional indications for sedation for mechanically ventilated patients include anxiolysis; treatment of subjective dyspnea; psychosis; facilitation of nursing care; reduction of autonomic hyperactivity, which may precipitate myocardial ischemia; and reduction of total O_2 consumption (V_{O2}).

Neuromuscular blocking agents are occasionally needed to facilitate mechanical ventilation in patients with profound dyssynchrony with the ventilator despite optimal sedation. The use of neuromuscular blocking agents may result in prolonged weakness—a myopathy known as the *postparalytic syndrome*. For this reason, these agents typically are used as a last resort when aggressive sedation fails to achieve patient-ventilator synchrony. Because neuromuscular blocking agents result in pharmacologic paralysis without altering mental status, sedative-induced amnesia is mandatory when these agents are administered.

Amnesia can be achieved reliably with benzodiazepines such as lorazepam and midazolam as well as the IV anesthetic agent propofol. Outside the setting of pharmacologic paralysis, there are few data supporting the idea that amnesia is mandatory in all patients who require intubation and mechanical ventilation. Since many of these patients have impaired hepatic and renal function, sedatives and opiates may accumulate in critically ill patients when they are given for prolonged periods. A protocol-driven approach to sedation of mechanically ventilated patients with daily interruption of sedative infusions paired with daily spontaneous breathing trials has been shown to prevent excessive drug accumulation and shorten the duration of mechanical ventilation and length of stay in the ICU.

MULTIORGAN SYSTEM FAILURE

The syndrome of multiorgan system failure is a common problem associated with critical illness. This syndrome is defined by the simultaneous presence of physiologic dysfunction and/or failure of two or more organs. Typically, this occurs in the setting of severe sepsis, shock of any kind, severe inflammatory conditions such as pancreatitis, and trauma. The fact that multiorgan system failure

occurs commonly in the ICU is a testament to our current ability to stabilize and support single-organ failure. The ability to support single-organ failure aggressively (e.g., mechanical ventilation for respiratory failure, renal replacement therapy for acute renal failure) has affected rates of early mortality in critical illness greatly. As a result, it is uncommon for critically ill patients to die in the initial stages of resuscitation. Instead, many patients succumb to critical illness later in the ICU stay, after the initial presenting problem has been stabilized.

Although there is debate regarding specific definitions of organ failure, several general principles governing the syndrome of multiorgan system failure apply. First, organ failure, no matter how defined, must persist beyond 24 h. Second, mortality risk increases as patients accrue additional failing. Third, prognosis is worsened by increased duration of organ failure. These observations remain true across various critical care settings (e.g., medical versus surgical). SIRS is a common basis for multiorgan system failure. Although infection is a common cause of SIRS, "sterile" triggers such as pancreatitis, trauma, and burns often are invoked to explain multiorgan system failure.

MONITORING IN THE ICU

Because respiratory and circulatory failure occurs commonly in critically ill patients, monitoring of the respiratory and cardiovascular systems is undertaken frequently in the ICU. Evaluation of respiratory gas exchange is routine in critical illness. The "gold standard" remains arterial blood-gas analysis, in which pH, partial pressures of O_2 and CO_2, and O_2 saturation are measured directly. With arterial blood-gas analysis, the two main functions of the lung—oxygenation of arterial blood and elimination of CO_2—can be assessed directly. Importantly, the blood pH, which has a profound effect on the drive to breathe, can be assessed only by sampling arterial blood. Though sampling of arterial blood is generally safe, it may be painful and cannot provide continuous information for clinicians routinely. In light of these limitations, noninvasive monitoring of respiratory function is often employed in the critical care setting.

Pulse Oximetry

This is the most commonly utilized noninvasive monitor of respiratory function. This technique takes advantage of differences in the absorptive properties of oxygenated and deoxygenated hemoglobin. At wavelengths of 660 nm, oxyhemoglobin reflects light more effectively than does deoxyhemoglobin, whereas the reverse is true in the infrared spectrum (940 nm). A pulse oximeter passes both wavelengths of light through a perfused digit such as a finger, and the relative intensity of light transmission at these two wavelengths is recorded. This allows the derivation of the relative percentage of oxyhemoglobin. Since arterial pulsations produce phasic changes in the intensity of transmitted light, the pulse oximeter is designed to detect only light of alternating intensity. This allows distinction of arterial and venous blood O_2 saturations.

RESPIRATORY SYSTEM MECHANICS

These can be measured in patients during mechanical ventilation (Chap. 269). When volume-controlled modes of mechanical ventilation are used, accompanying airway pressures can be easily measured, assuming the patient is passive. The peak airway pressure is determined by two variables: airway resistance and respiratory system compliance. At the end of inspiration, inspiratory flow can be stopped transiently. This end-inspiratory pause (*plateau pressure*) is a static measurement, affected only by respiratory system compliance, not airway resistance. Therefore, during volume-controlled ventilation, the difference between the peak (airway resistance + respiratory system compliance) and plateau (respiratory system compliance only) airway pressures provides a quantitative assessment of airway resistance. Accordingly, during volume-controlled ventilation, patients with increases in airway resistance typically have increased peak airway pressures as well as abnormally high gradients between peak and plateau airway pressures (typically >15 cmH_2O). The compliance of the respiratory system is defined by the change in pressure of the respiratory system per unit change in volume.

The respiratory system can be divided further into two components: the lungs and the chest wall. Normally, respiratory system compliance is ~100 mL per cmH_2O. Pathophysiologic processes such as pleural effusions, pneumothorax, and increased abdominal girth from ascites all reduce chest wall compliance. Lung compliance may be reduced by pneumonia, pulmonary edema from any cause, or autoPEEP. Accordingly, patients with abnormalities in compliance of the respiratory system (lungs and/or chest wall) typically have elevated peak *and* plateau airway pressures but a normal gradient between peak and plateau airway pressures. AutoPEEP occurs when there is insufficient time for emptying of alveoli before the next inspiratory cycle. Since the alveoli have not decompressed completely, alveolar pressure remains positive at end exhalation (functional residual capacity). This phenomenon results most commonly from critical narrowing of distal airways in disease processes such as asthma and chronic obstructive pulmonary disease. AutoPEEP with resulting alveolar overdistention may result in diminished lung compliance, reflected by abnormally increased plateau airway pressures. Modern mechanical ventilators allow breath-to-breath display of pressure and flow, which may allow detection of problems such as patient-ventilator dyssynchrony, airflow obstruction, and autoPEEP **(Fig. 267-7)**.

Figure 267-7.

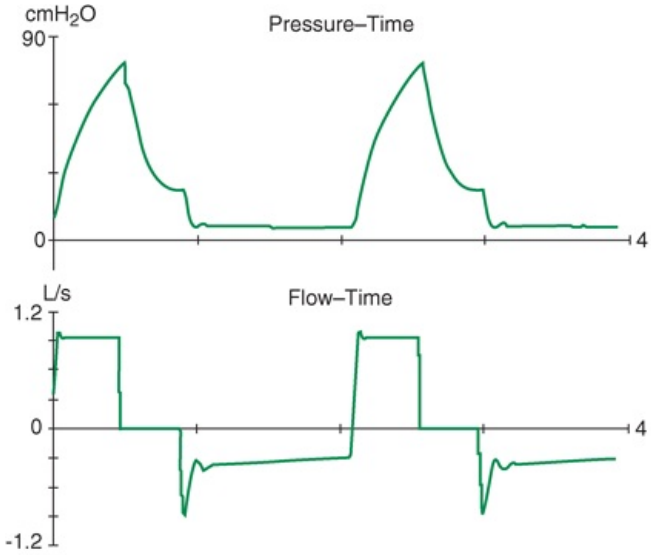

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Increased airway resistance with autoPEEP. The top waveform (airway pressure vs. time) shows a large difference between the peak airway pressure (80 cmH_2O) and the plateau airway pressure (20 cmH_2O). The bottom waveform (flow vs. time) demonstrates airflow throughout expiration (reflected by the flow tracing on the negative portion of the abscissa) that persists up to the next inspiratory effort.

Circulatory Status

Oxygen delivery (Q_{O2}) is a function of cardiac output and the content of O_2 in the arterial blood (Ca_{O2}). The Ca_{O2} is determined by the hemoglobin concentration, the arterial hemoglobin saturation, and dissolved O_2 not bound to hemoglobin. For normal adults:

$$\begin{aligned} Q_{O_2} &= 50\ \text{dL/min} \times [1.39 \times 15\ \text{g/dL (hemoglobin concentration)} \\ &\quad \times 1.0\ \text{(hemoglobin \% saturation)} + 0.0031 \times 100\ (Pa_{O_2})] \\ &= 50\ \text{dL/min (cardiac output)} \times 21.16\ \text{mL}\ O_2\ \text{per dL blood}\ (Ca_{O_2}) \\ &= 1058\ \text{mL}\ O_2\ \text{per min} \end{aligned}$$

It is apparent that the vast majority of O_2 delivered to tissues is bound to hemoglobin and that the dissolved O_2 (PaO_2) contributes very little to O_2 content in arterial blood or O_2 delivery. Normally, the content of O_2 in mixed venous blood ($C\bar{v}_{O2}$) is 15.76 mL O_2 per dL blood, since the mixed venous blood is 75% saturated. Therefore, the normal tissue extraction ratio for O_2 is $Ca_{O2} - C\bar{v}_{O2}/Ca_{O2}$ ([21.16–15.76]/21.16) or ~25%. A pulmonary artery catheter allows measurements of O_2 delivery and O_2 extraction ratio.

The mixed venous O_2 saturation allows assessment of global tissue perfusion. A reduced mixed venous O_2 saturation may be caused by inadequate cardiac output, reduced hemoglobin concentration, and/or reduced arterial O_2 saturation. An abnormally high O_2 consumption (V_{O2}) may also lead to a reduced mixed venous O_2 saturation if O_2 delivery is not concomitantly increased. Abnormally increased V_{O2} by peripheral tissues may be caused by a multitude of problems, such as fever, agitation, shivering, and thyrotoxicosis.

The pulmonary artery catheter originally was designed as a tool to guide therapy in acute myocardial infarction but is currently used in the ICU for evaluation and treatment of a variety of other conditions, such as ARDS, septic shock, congestive heart failure, and acute renal failure. This device has never been validated as a tool associated with reduction in morbidity and mortality rates. Indeed, despite numerous prospective studies, there has been no report of mortality or morbidity rate benefit associated with the use of the pulmonary artery catheter in any setting. Accordingly, it appears that the routine use of pulmonary artery catheterization is not indicated as a monitor to characterize the circulatory status in most critically ill patients.

Recent data suggest that static measurements of circulatory parameters (e.g., CVP, PCWP) do not provide reliable information on the circulatory status of critically ill patients. In contrast, dynamic assessments measuring the impact of breathing on the circulation are more reliable predictors of responsiveness to IV fluid administration. A decrease in CVP of >1 mmHg during inspiration in a

spontaneously breathing patient has been shown to predict an increase in cardiac output after IV fluid administration. Similarly, a changing pulse pressure during mechanical ventilation has been shown to predict an increase in cardiac output after IV fluid administration in patients with septic shock.

PREVENTION OF COMPLICATIONS OF CRITICAL ILLNESS

SEPSIS IN THE CRITICAL CARE UNIT

(See also Chap. 271) Sepsis is a significant problem in the care of critically ill patients. It is the leading cause of death in noncoronary ICUs in the United States, with case rates expected to increase as the population ages with a greater percentage of people vulnerable to infection.

Many therapeutic interventions in the ICU are invasive and predispose patients to infectious complications. These interventions include endotracheal intubation, indwelling vascular catheters, nasally placed enteral feeding tubes, transurethral bladder catheters, and other catheters placed into sterile body cavities (e.g., tube thoracostomy, percutaneous intraabdominal drainage catheters). The longer such devices remain in place, the more prone to these infections patients become. For example, ventilator-associated pneumonia (VAP) correlates strongly with the duration of intubation and mechanical ventilation. Therefore, an important aspect of preventive care is the timely removal of invasive devices as soon as they are no longer needed. Multidrug-resistant organisms are commonplace in the ICU.

An important aspect of critical care is infection control in the ICU. Simple measures such as frequent hand washing are effective but underutilized strategies. Protective isolation of patients with colonization or infection by drug-resistant organisms is another frequently used strategy in the critical care setting. A recent study utilizing silver-coated endotracheal tubes reported a significant reduction in VAP incidence. Studies evaluating multifaceted, evidence-based strategies to decrease catheter-related bloodstream infections have shown improved outcomes from using measures such as hand washing, full-barrier precautions during insertion, chlorhexidine skin preparation, avoidance of the femoral site, and timely catheter removal.

DEEP VENOUS THROMBOSES (DVTS)

All ICU patients are at high risk for this complication because of their predilection for being immobile. Therefore, all should receive some form of prophylaxis against DVT. The most commonly employed forms of prophylaxis are subcutaneous low-dose heparin injections and sequential compression devices for the lower extremities. Observational studies report an alarming incidence of the occurrence of DVTs despite the use of these standard prophylactic regimens. Heparin prophylaxis may result in heparin-induced thrombocytopenia (HIT), another relatively common nosocomial complication in critically ill patients.

Low-molecular-weight heparins such as enoxaparin are more effective than unfractionated heparin for DVT prophylaxis in high-risk patients, such as those undergoing orthopedic surgery, and they have a lower incidence of HIT. Fondaparinux, a selective factor Xa inhibitor, is even more effective than enoxaparin in high-risk orthopedic patients.

STRESS ULCERS

Prophylaxis against stress ulcers is frequently administered in most ICUs; typically, histamine-2 antagonists are given. Currently available data suggest that high-risk patients, such as those with coagulopathy, shock, or respiratory failure requiring mechanical ventilation, benefit from such prophylactic treatment.

NUTRITION AND GLYCEMIC CONTROL

These are important issues in critically ill patients that may be associated with respiratory failure, impaired wound healing, and dysfunctional immune response. Early enteral feeding is reasonable, though no data are available to suggest that this improves patient outcome per se. Certainly, enteral feeding, if possible, is preferred over parenteral nutrition, which is associated with numerous complications, including hyperglycemia, fatty liver, cholestasis, and sepsis. In addition, enteral feeding may prevent bacterial translocation across the gut mucosa. Tight glucose control is another area of controversy in critical care. Although one study showed a significant mortality benefit when glucose levels were aggressively normalized in a large group of surgical ICU patients, more recent data suggest that tight glucose control in a large population of both medical and surgical ICU patients resulted in increased rates of mortality.

ICU-ACQUIRED WEAKNESS

This occurs frequently in patients who survive critical illness. It is particularly common in those with SIRS and/or sepsis. Neuropathies and myopathies both have been described, most commonly after ~1 week in the ICU. The mechanisms behind ICU-acquired weakness syndromes are poorly understood. Intensive insulin therapy may reduce polyneuropathy of critical illness. A recent study of very early physical and occupational therapy in mechanically ventilated critically ill patients reported significant improvements in functional independence at hospital discharge, as well as reduced duration of mechanical ventilation and delirium.

ANEMIA

This is a common problem in critically ill patients. Studies have shown that the vast majority of ICU patients are anemic. Furthermore, most have anemia of chronic inflammation. Phlebotomy contributes significantly to anemia in ICU patients. Studies have demonstrated that erythropoietin levels are inappropriately reduced in most ICU patients and that exogenous erythropoietin administration may reduce transfusion requirements in the ICU. The hemoglobin level that merits transfusion in critically ill patients has been a long-standing area of controversy. A large, multicenter study involving patients in many different ICU settings challenged the conventional notion that a hemoglobin level of 100 g/L (10 g/dL) is needed in critically ill patients. Red blood cell transfusion is associated with impairment of immune function and increased risk of infections as well as acute lung injury and volume overload, all of which may explain the findings in this study. A conservative transfusion strategy should be the rule in managing critically ill patients who are not actively hemorrhaging.

ACUTE RENAL FAILURE

(See also Chap. 279) This occurs in a significant percentage of critically ill patients. The most common underlying etiology is acute tubular necrosis, usually precipitated by hypoperfusion and/or nephrotoxic agents. Currently, there are no pharmacologic agents available for prevention of renal injury in critical illness. A recent study showed convincingly that low-dose dopamine is *not* effective in protecting the kidneys from acute injury.

NEUROLOGIC DYSFUNCTION IN CRITICALLY ILL PATIENTS

DELIRIUM

(See also Chaps. 25 and 274) This state is defined by (1) an acute onset of changes or fluctuations in the course of mental status, (2) inattention, (3) disorganized thinking, and (4) an altered level of consciousness (i.e., other than alert). Delirium is reported to occur in over 80% of mechanically ventilated ICU patients and can be detected by the Confusion Assessment Method (CAM)-ICU. This assessment asks patients to answer simple questions and perform simple tasks and can be completed by the bedside nurse in ~2 min. The differential diagnosis of delirium in ICU patients is broad and includes infectious etiologies (including sepsis), medications (particularly sedatives and analgesics), drug withdrawal, metabolic/electrolyte derangements, intracranial pathology (e.g., stroke, intracranial hemorrhage), seizures, hypoxia, hypertensive crisis, shock, and vitamin deficiencies (particularly thiamine). Patients with ICU delirium have increases in hospital length of stay, time on mechanical ventilation, cognitive impairment at hospital discharge, and 6-month mortality rate. Interventions to reduce ICU delirium have been described recently. The use of the novel sedative dexmedetomidine was associated with reduced ICU delirium compared with midazolam. In addition, as mentioned above in the section "ICU-Acquired Weakness," very early physical and occupational therapy in mechanically ventilated patients also has been demonstrated to reduce delirium.

ANOXIC CEREBRAL INJURY

(See also Chap. 275) This condition is common after cardiac arrest and often results in severe and permanent brain injury in patients whose cardiac arrest is resuscitated. Active cooling of patients after cardiac arrest has been shown to improve neurologic outcomes. Therefore, patients who present to the ICU after circulatory arrest from ventricular fibrillation or pulseless ventricular tachycardia should be actively cooled with cooling blankets and ice packs if necessary to achieve a core body temperature of 32–34°C.

STROKE

(See also Chap. 370) This is a common cause of neurologic critical illness. Hypertension must be managed carefully, since abrupt reductions in blood pressure may be associated with further brain ischemia and injury. Acute ischemic stroke treated with tissue plasminogen activator (tPA) has an improved neurologic outcome when treatment is given within 3 h of onset of symptoms. The mortality rate is not improved when tPA is compared with placebo, despite the improved neurologic outcome. Cerebral hemorrhage is significantly higher in patients given tPA. A treatment benefit is not seen when tPA therapy is given beyond 3 h. Heparin has not been shown to demonstrate improved outcomes convincingly in patients with acute ischemic stroke.

SUBARACHNOID HEMORRHAGE

(See also Chap. 370) This may occur secondary to aneurysm rupture and is often complicated by cerebral vasospasm, rebleeding, and hydrocephalus. Vasospasm can be detected by either transcranial Doppler assessment or cerebral angiography; it is typically treated with the calcium channel blocker nimodipine, aggressive IV fluid administration, and therapy aimed at increasing blood pressure, typically with vasoactive drugs such as phenylephrine. The IV fluids and vasoactive drugs (hypertensive hypervolemic therapy) are used to overcome the cerebral vasospasm. Early surgical clipping of aneurysms is advocated by most authorities to prevent complications related to rebleeding. Hydrocephalus, typically heralded by a decreased level of consciousness, may require ventriculostomy drainage.

STATUS EPILEPTICUS

(See also Chap. 369) Recurrent or relentless seizure activity is a medical emergency. Cessation of seizure activity is required to prevent irreversible neurologic injury. Lorazepam is the most effective benzodiazepine for treating status epilepticus and is the treatment of choice for controlling seizures acutely. Phenytoin or fosphenytoin should be given concomitantly since lorazepam has a short half-life. Other drugs, such as gabapentin, carbamazepine, and phenobarbital, should be reserved for patients with contraindications to phenytoin (e.g., allergy or pregnancy) or ongoing seizures despite phenytoin.

BRAIN DEATH

(See also Chap. 275) Though critically ill patients usually die from irreversible cessation of circulatory and respiratory function, a diagnosis of death also may be established by irreversible cessation of all functions of the entire brain, including the brainstem, even if circulatory and respiratory function remains intact on artificial life support. Patients must demonstrate absence of cerebral function (unresponsive to all external stimuli) and brainstem functions [e.g., unreactive pupils, absent ocular movement to head turning or ice water irrigation of ear canals, positive apnea test (no drive to breathe)]. Absence of brain function must have an established cause and be permanent without possibility of recovery (e.g., must confirm the absence of sedative effect, hypothermia, hypoxemia, neuromuscular paralysis, or severe hypotension). If there is uncertainty about the cause of coma, studies of cerebral blood flow and electroencephalography should be performed.

WITHHOLDING AND WITHDRAWING CARE

(See also Chap. 9) Withholding and withdrawing of care occurs commonly in the ICU setting. The Task Force on Ethics of the Society of Critical Care Medicine reported that it is ethically sound to withhold or withdraw care if a patient or surrogate makes such a request or if the goals of therapy are not achievable according to the physician. Since all medical treatments are justified by their expected benefits, the loss of such an expectation justifies the act of withdrawing or withholding such treatment. Thus, the act of withdrawing care is fundamentally similar to the act of withholding care. An underlying stipulation derived from this report is that an informed patient should have his or her wishes respected with regard to life-sustaining therapy. Implicit in this stipulation is the need to ensure that patients are thoroughly and accurately informed regarding the plausibility and expected results of various therapies.

The act of informing patients and/or surrogate decision makers is the responsibility of the physician and other health care providers. If a patient or surrogate desires therapy deemed futile by the treating physician, the physician is not obligated ethically to provide such treatment. Rather, arrangements may be made to transfer the patients care to another care provider. Whether the decision to withdraw life support should be initiated by the physician or left to surrogate decision makers is not clear. A recent study reported that slightly more than half of surrogate decision makers preferred to receive such a recommendation, whereas the rest did not. Critical care providers should meet regularly with patients and/or surrogates to discuss prognosis when the withholding or withdrawal of care is being considered. After a consensus among caregivers has been reached regarding withholding or withdrawal of care, this should be relayed to the patient and/or surrogate decision maker. If a decision to withhold or withdraw life-sustaining care for a patient has been reached, aggressive attention to analgesia and anxiolysis is needed. Opiates and benzodiazepines are typically used to achieve these goals.

FURTHER READINGS

Abraham E et al: Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. N Engl J Med 353:1332, 2005[PMID: 16192478]

The Acute Respiratory Distress Syndrome Network: Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 342:1301, 2000

Cook DJ et al: Deep venous thrombosis in medical-surgical critically ill patients: Prevalence, incidence, and risk factors. Crit Care Med 33:1565, 2005[PMID: 16003063]

Esteban A et al: Noninvasive positive-pressure ventilation for respiratory failure after extubation. N Engl J Med 350:2452, 2004[PMID: 15190137]

Girard T et al: Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): A randomised controlled trial. Lancet 371:126, 2008[PMID: 18191684]

The Hypothermia after Cardiac Arrest Study Group: Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. N Engl J Med 346:549, 2002

The National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network: Comparison of two fluid-management strategies in acute lung injury. N Engl J Med 354:2564, 2006

The Nice-Sugar Study Investigators: Intensive versus glucose control in critically ill patients. N Engl J Med 360:1283, 2009

Pronovost P et al: An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 355:2725, 2006[PMID: 17192537]

Schweikert WD et al: Early physical and occupational therapy in mechanically ventilated, critically ill patients: A randomised controlled trial. Lancet 373:1874, 2009

Shah MR et al: Impact of the pulmonary artery catheter in critically ill patients: Meta-analysis of randomized clinical trials. JAMA 294:1664, 2005[PMID: 16204666]

Task Force on Ethics of the Society of Critical Care Medicine: Consensus report on the ethics of forgoing life-sustaining treatments in the critically ill. Crit Care Med 18:1424, 1990

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACUTE RESPIRATORY DISTRESS SYNDROME: INTRODUCTION

Acute respiratory distress syndrome (ARDS) is a clinical syndrome of severe dyspnea of rapid onset, hypoxemia, and diffuse pulmonary infiltrates leading to respiratory failure. ARDS is caused by diffuse lung injury from many underlying medical and surgical disorders. The lung injury may be direct, as occurs in toxic inhalation, or indirect, as occurs in sepsis (Table 268-1). The clinical features of ARDS are listed in Table 268-2. Acute lung injury (ALI) is a less severe disorder but has the potential to evolve into ARDS (Table 268-2). The arterial (a) PO_2 (in mmHg)/FIO_2 (inspiratory O_2 fraction) <200 mmHg is characteristic of ARDS, while a PaO_2/FIO_2 between 200 and 300 identifies patients with ALI who are likely to benefit from aggressive therapy.

Table 268–1. Clinical Disorders Commonly Associated with ARDS

Direct Lung Injury	Indirect Lung Injury
Pneumonia	Sepsis
Aspiration of gastric contents	Severe trauma
Pulmonary contusion	Multiple bone fractures
Near-drowning	Flail chest
Toxic inhalation injury	Head trauma
	Burns
	Multiple transfusions
	Drug overdose
	Pancreatitis
	Postcardiopulmonary bypass

Table 268–2. Diagnostic Criteria for ALI and ARDS

Oxygenation	Onset	Chest Radiograph	Absence of Left Atrial Hypertension
ALI: PaO_2/FIO^2 ≤ 300 mmHg	Acute	Bilateral alveolar or interstitial infiltrates	PCWP ≤ 18 mmHg *or* no clinical evidence of increased left atrialpressure
ARDS: PaO_2/FIO_2 ≤ 200 mmHg			

Abbreviations: ALI, acute lung injury; ARDS, acute respiratory distress syndrome; FIO_2, inspired O_2 percentage; PaO_2, arterial partial pressure of O2; PCWP, pulmonary capillary wedge pressure.

The annual incidences of ALI and ARDS are estimated to be up to 80/100,000 and 60/100,000, respectively. Approximately 10% of all intensive care unit (ICU) admissions suffer from acute respiratory failure, with ~20% of these patients meeting criteria for ALI or ARDS.

Etiology

While many medical and surgical illnesses have been associated with the development of ALI and ARDS, most cases (>80%) are caused by a relatively small number of clinical disorders, namely, severe sepsis syndrome and/or bacterial pneumonia (~40–50%), trauma, multiple transfusions, aspiration of gastric contents, and drug overdose. Among patients with trauma, pulmonary contusion, multiple bone fractures, and chest wall trauma/flail chest are the most frequently reported surgical conditions in ARDS, whereas head trauma, near-drowning, toxic inhalation, and burns are rare causes. The risks of developing ARDS are increased in patients suffering from more than one predisposing medical or surgical condition (e.g., the risk for ARDS increases from 25% in patients with severe trauma to 56% in patients with trauma and sepsis).

Several other clinical variables have been associated with the development of ARDS. These include

older age, chronic alcohol abuse, metabolic acidosis, and severity of critical illness. Trauma patients with an acute physiology and chronic health evaluation (APACHE) II score ≥16 (Chap. 267) have a 2.5-fold increase in the risk of developing ARDS, and those with a score >20 have an incidence of ARDS that is more than threefold greater than those with APACHE II scores ≤9.

Clinical Course and Pathophysiology

The natural history of ARDS is marked by three phases—exudative, proliferative, and fibrotic—each with characteristic clinical and pathologic features (Fig. 268-1).

Figure 268-1

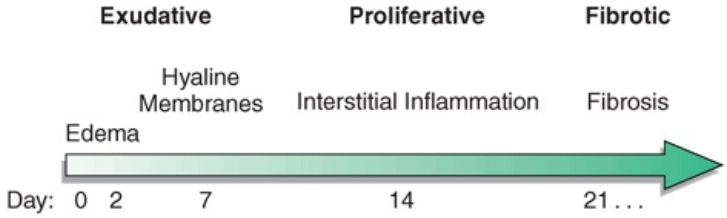

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram illustrating the time course for the development and resolution of ARDS. The exudative phase is notable for early alveolar edema and neutrophil-rich leukocytic infiltration of the lungs with subsequent formation of hyaline membranes from diffuse alveolar damage. Within 7 days, a proliferative phase ensues with prominent interstitial inflammation and early fibrotic changes. Approximately 3 weeks after the initial pulmonary injury, most patients recover. However, some patients enter the fibrotic phase, with substantial fibrosis and bullae formation.

EXUDATIVE PHASE

(Figure. 268-2) In this phase, alveolar capillary endothelial cells and type I pneumocytes (alveolar epithelial cells) are injured, leading to the los of the normally tight alveolar barrier to fluid and macromolecules. Edema fluid that is rich in protein accumulates in the interstitial and alveolar spaces. Significant concentrations of cytokines (e.g., interleukin 1, interleukin 8, and tumor necrosis factor α) and lipid mediators (e.g., leukotriene B_4) are present in the lung in this acute phase. In response to proinflammatory mediators, leukocytes (especially neutrophils) traffic into the pulmonary interstitium and alveoli. In addition, condensed plasma proteins aggregate in the air spaces with cellular debris and dysfunctional pulmonary surfactant to form hyaline membrane whorls. Pulmonary vascular injury also occurs early in ARDS, with vascular obliteration by microthrombi and fibrocellular proliferation (Fig. 268-3).

Figure 268-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A representative anteroposterior (AP) chest x-ray in the exudative phase of ARDS that shows diffuse interstitial and alveolar infiltrates, that can be difficult to distinguish from left ventricular failure.

Figure 268-3

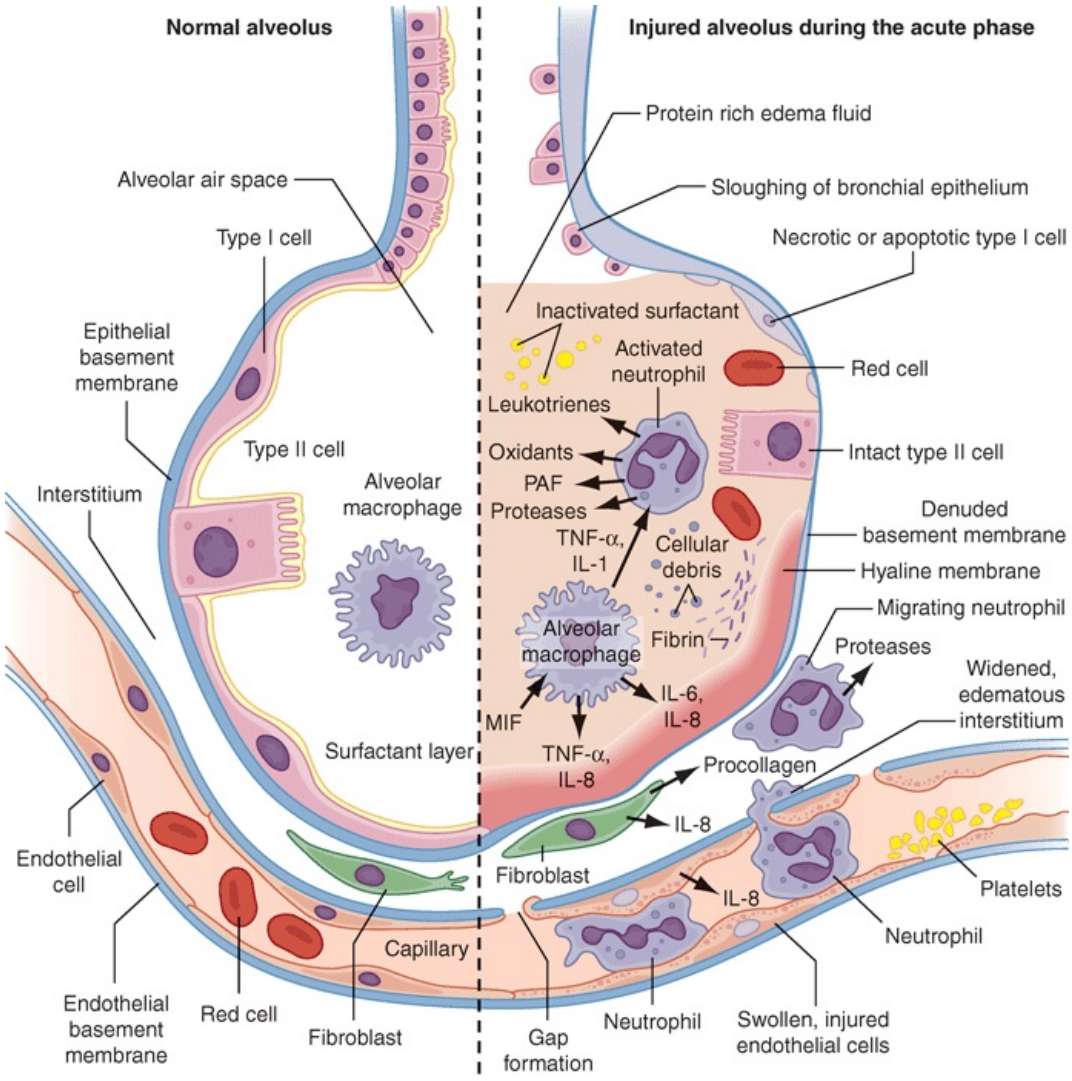

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The normal alveolus (left-hand side) and the injured alveolus in the acute phase of acute lung injury and the acute respiratory distress syndrome (right-hand side). In the acute phase of the syndrome (right-hand side), there is sloughing of both the bronchial and alveolar epithelial cells, with the formation of protein-rich hyaline membranes on the denuded basement membrane. Neutrophils are shown adhering to the injured capillary endothelium and marginating through the interstitium into the air space, which is filled with protein-rich edema fluid. In the air space, an alveolar macrophage is secreting cytokines, interleukins 1, 6, 8, and 10 (IL-1, -6, -8, and -10) and tumor necrosis factor α (TNF-α), that act locally to stimulate chemotaxis and activate neutrophils. Macrophages also secrete other cytokines, including IL-1, -6, and -10. IL-1 can also stimulate the production of extracellular matrix by fibroblasts. Neutrophils can release oxidants, proteases, leukotrienes, and other proinflammatory molecules, such as platelet-activating factor (PAF). A number of antiinflammatory mediators are also present in the alveolar milieu, including IL-1–receptor antagonist, soluble TNF-α receptor, autoantibodies against IL-8, and cytokines such as IL-10 and -11 (not shown). The influx of protein-rich edema fluid into the alveolus has led to the inactivation of surfactant. MIF, macrophage inhibitory factor. (*From Ware and Matthay, with permission.*)

Alveolar edema predominantly involves *dependent* portions of the lung, leading to diminished aeration and atelectasis. Collapse of large sections of dependent lung markedly decreases lung compliance. Consequently, intrapulmonary shunting and hypoxemia develop and the work of breathing rises, leading to dyspnea. The pathophysiologic alterations in alveolar spaces are exacerbated by microvascular occlusion that leads to reductions in pulmonary arterial blood flow to ventilated portions of the lung, increasing the dead space, and to pulmonary hypertension. Thus, in addition to severe hypoxemia, hypercapnia secondary to an increase in pulmonary dead space is also prominent in early ARDS.

The exudative phase encompasses the first 7 days of illness after exposure to a precipitating ARDS risk factor, with the patient experiencing the onset of respiratory symptoms. Although usually present within 12–36 hours after the initial insult, symptoms can be delayed by 5–7 days. Dyspnea develops with a sensation of rapid shallow breathing and an inability to get enough air. Tachypnea and increased work of breathing frequently result in respiratory fatigue and ultimately in respiratory failure. Laboratory values are generally nonspecific and primarily indicative of underlying clinical disorders. The chest radiograph usually reveals alveolar and interstitial opacities involving at least three-quarters of the lung fields (Fig. 268-2). While characteristic for ARDS or ALI, these radiographic findings are not specific and can be indistinguishable from cardiogenic pulmonary edema (Chap. 272). Unlike the latter, however, the chest x-ray in ARDS rarely shows cardiomegaly, pleural effusions, or pulmonary vascular redistribution. Chest computed tomographic (CT) scanning in ARDS reveals extensive heterogeneity of lung involvement (Fig. 268-4).

Figure 268-4

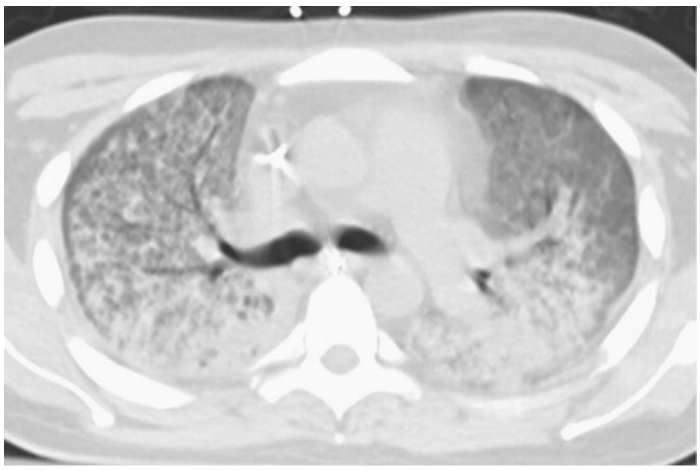

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A representative computed tomographic scan of the chest during the exudative phase of ARDS in which *dependent* alveolar edema and atelectasis predominate.

Because the early features of ARDS and ALI are nonspecific, alternative diagnoses must be considered. In the differential diagnosis of ARDS, the most common disorders are cardiogenic pulmonary edema, diffuse pneumonia, and alveolar hemorrhage. Less frequent diagnoses to consider include acute interstitial lung diseases [e.g., acute interstitial pneumonitis (Chap. 261)], acute immunologic injury [e.g., hypersensitivity pneumonitis (Chap. 255)], toxin injury (e.g., radiation pneumonitis), and neurogenic pulmonary edema.

PROLIFERATIVE PHASE

This phase of ARDS usually lasts from day 7 to day 21. Most patients recover rapidly and are liberated from mechanical ventilation during this phase. Despite this improvement, many still experience dyspnea, tachypnea, and hypoxemia. Some patients develop progressive lung injury and early changes of pulmonary fibrosis during the proliferative phase. Histologically, the first signs of resolution are often evident in this phase with the initiation of lung repair, organization of alveolar exudates, and a shift from a neutrophil- to a lymphocyte-predominant pulmonary infiltrate. As part of the reparative process, there is a proliferation of type II pneumocytes along alveolar basement membranes. These specialized epithelial cells synthesize new pulmonary surfactant and differentiate into type I pneumocytes. The presence of alveolar type III procollagen peptide, a marker of pulmonary fibrosis, is associated with a protracted clinical course and increased mortality from ARDS.

FIBROTIC PHASE

While many patients with ARDS recover lung function 3–4 weeks after the initial pulmonary injury, some will enter a fibrotic phase that may require long-term support on mechanical ventilators and/or supplemental oxygen. Histologically, the alveolar edema and inflammatory exudates of earlier phases are now converted to extensive alveolar duct and interstitial fibrosis. Acinar architecture is markedly disrupted, leading to emphysema-like changes with large bullae. Intimal fibroproliferation in the pulmonary microcirculation leads to progressive vascular occlusion and pulmonary hypertension. The physiologic consequences include an increased risk of pneumothorax, reductions in lung compliance, and increased pulmonary dead space. Patients in this late phase experience a substantial burden of excess morbidity. Lung biopsy evidence for pulmonary fibrosis in any phase of ARDS is associated with increased mortality.

Treatment: Acute Respiratory Distress Syndrome

GENERAL PRINCIPLES

Recent reductions in ARDS/ALI mortality are largely the result of general advances in the care of critically ill patients (Chap. 267). Thus, caring for these patients requires close attention to (1) the recognition and treatment of the underlying medical and surgical disorders (e.g., sepsis, aspiration, trauma); (2) minimizing procedures and their complications; (3) prophylaxis against venous thromboembolism, gastrointestinal bleeding, aspiration, excessive sedation, and central venous catheter infections; (4) prompt recognition of nosocomial infections; and (5) provision of adequate nutrition.

MANAGEMENT OF MECHANICAL VENTILATION

(See also Chap. 269) Patients meeting clinical criteria for ARDS frequently fatigue from increased work of breathing and progressive hypoxemia, requiring mechanical ventilation for support.

Ventilator-Induced Lung Injury

Despite its life-saving potential, mechanical ventilation can aggravate lung injury. Experimental models have demonstrated that ventilator-induced lung injury appears to require two processes: repeated alveolar overdistention and recurrent alveolar collapse. Clearly evident by chest CT (Fig. 268-4), ARDS is a heterogeneous disorder, principally involving dependent portions of the lung with relative sparing of other regions. Because of their differing compliance, attempts to fully inflate the consolidated lung may lead to overdistention and injury to the more "normal" areas of the lung. Ventilator-induced injury can be demonstrated in experimental models of ALI, with high tidal volume (V_T) ventilation resulting in additional, synergistic alveolar damage. These findings led to the hypothesis that ventilating patients suffering from ALI or ARDS with lower V_Ts would protect against ventilator-induced lung injury and improve clinical outcomes.

A large-scale, randomized controlled trial sponsored by the National Institutes of Health and conducted by the ARDS Network compared low V_T (6 mL/kg predicted body weight) ventilation to conventional V_T (12 mL/kg predicted body weight) ventilation. Mortality was significantly lower in the low V_T patients (31%) compared to the conventional V_T patients (40%). This improvement in survival represents the most substantial benefit in ARDS mortality demonstrated for *any* therapeutic intervention in ARDS to date.

Prevention of Alveolar Collapse

In ARDS, the presence of alveolar and interstitial fluid and the loss of surfactant can lead to a marked reduction of lung compliance. Without an increase in end-expiratory pressure, significant alveolar collapse can occur at end-expiration, impairing oxygenation. In most clinical settings, positive end-expiratory pressure (PEEP) is empirically set to minimize FIO_2 and maximize PaO_2. On most modern mechanical ventilators, it is possible to construct a static pressure–volume curve for the respiratory system. The lower inflection point on the curve represents alveolar opening (or "recruitment"). The pressure at this point, usually 12–15 mmHg in ARDS, is a theoretical "optimal PEEP" for alveolar recruitment. Titration of the PEEP to the lower inflection point on the static pressure–volume curve has been hypothesized to keep the lung open, improving oxygenation and protecting against lung injury. Three large randomized trials have investigated the utility of PEEP-based strategies to keep the lung open. In all three trials, improvement in lung function was evident but there were no significant differences in overall mortality. Until more data become available on the clinical utility of high PEEP, it is advisable to set PEEP to minimize FIO_2 and optimize PaO_2 (Chap. 269). Measurement of esophageal pressures to estimate transpulmonary pressure may help identify an optimal PEEP in some patients.

Oxygenation can also be improved by increasing mean airway pressure with "inverse ratio ventilation." In this technique, the inspiratory (*I*) time is lengthened so that it is longer than the expiratory (*E*) time (*I*:*E* > 1:1). With diminished time to exhale, dynamic hyperinflation leads to increased end-expiratory pressure, similar to ventilator-prescribed PEEP. This mode of ventilation has the advantage of improving oxygenation with lower peak pressures than conventional ventilation. Although inverse ratio ventilation can improve oxygenation and help reduce FIO_2 to ≤0.60 to avoid possible oxygen toxicity, no mortality benefit in ARDS has been demonstrated. Recruitment maneuvers that transiently increase PEEP to "recruit" atelectatic lung can also increase oxygenation, but a mortality benefit has not been established.

In several randomized trials, mechanical ventilation in the prone position improved arterial oxygenation, but its effect on survival and other important clinical outcomes remains uncertain. Moreover, unless the critical-care team is experienced in "proning," repositioning critically ill patients can be hazardous, leading to accidental endotracheal extubation, loss of central venous catheters, and orthopedic injury. Until validation of its efficacy, prone-position ventilation should be reserved for only the most critically ill ARDS patients.

Other Strategies in Mechanical Ventilation

Several additional mechanical ventilation strategies that utilize specialized equipment have been tested in ARDS patients, most with mixed or disappointing results in adults. These include high-frequency ventilation (HFV) [i.e., ventilating at extremely high respiratory rates (5–20 cycles per second) and low V_Ts (1–2 mL/kg)]. Partial liquid ventilation (PLV) with perfluorocarbon, an inert, high-density liquid that easily solubilizes oxygen and carbon dioxide, has revealed promising preliminary data on pulmonary function in patients with ARDS but also without survival benefit. Lung-replacement therapy with extracorporeal membrane oxygenation (ECMO), which provides a clear survival benefit in neonatal respiratory distress syndrome, may also have utility in select adult patients with ARDS.

Data in support of the efficacy of "adjunctive" ventilator therapies (e.g., high PEEP, inverse ratio ventilation, recruitment maneuvers, prone positioning, HFV, ECMO, and PLV) remain incomplete, so these modalities are not routinely used.

FLUID MANAGEMENT

(See also Chap. 267) Increased pulmonary vascular permeability leading to interstitial and alveolar edema rich in protein is a central feature of ARDS. In addition, impaired vascular integrity augments the normal increase in extravascular lung water that occurs with increasing left atrial pressure. Maintaining a normal or low left atrial filling pressure minimizes pulmonary edema and prevents further decrements in arterial oxygenation and lung compliance, improves pulmonary mechanics, shortens ICU stay and the duration of mechanical ventilation, and is associated with a lower mortality in both medical and surgical ICU patients. Thus, aggressive attempts to reduce left atrial filling pressures with fluid restriction and diuretics should be an important aspect of ARDS management, limited only by hypotension and hypoperfusion of critical organs such as the kidneys.

GLUCOCORTICOIDS

Inflammatory mediators and leukocytes are abundant in the lungs of patients with ARDS. Many attempts have been made to treat both early and late ARDS with glucocorticoids to reduce this potentially deleterious pulmonary inflammation. Few studies have shown any benefit. Current evidence does *not* support the use of high-dose glucocorticoids in the care of ARDS patients.

OTHER THERAPIES

Clinical trials of surfactant replacement and multiple other medical therapies have proved disappointing. Inhaled nitric oxide (NO) can transiently improve oxygenation but does not improve survival or decrease time on mechanical ventilation. Therefore, the use of NO is *not* currently recommended in ARDS.

RECOMMENDATIONS

Many clinical trials have been undertaken to improve the outcome of patients with ARDS; most have been unsuccessful in modifying the natural history. The large number and uncertain clinical efficacy of ARDS therapies can make it difficult for clinicians to select a rational treatment plan, and these patients' critical illnesses can tempt physicians to try unproven and potentially harmful therapies. While results of large clinical trials must be judiciously administered to *individual* patients, evidence-based recommendations are summarized in Table 268-3, and an algorithm for the initial therapeutic goals and limits in ARDS management is provided in Fig. 268-5.

Table 268–3. Evidence-Based Recommendations for ARDS Therapies

Treatment	Recommendation*
Mechanical ventilation:	
Low tidal volume	A
Minimize left atrial filling pressures	B
High-PEEP or "open lung"	C
Prone position	C
Recruitment maneuvers	C
ECMO	C
High-frequency ventilation	D
Glucocorticoids	D
Surfactant replacement, inhaled nitric oxide, and other anti-inflammatory therapy (e.g., ketoconazole, PGE_1, NSAIDs)	D

*A, recommended therapy based on strong clinical evidence from randomized clinical trials; B, recommended therapy based on supportive but limited clinical data; C, indeterminate evidence: recommended only as alternative therapy; D, not recommended based on clinical evidence against

efficacy of therapy.

Abbreviations: ARDS, acute respiratory distress syndrome; ECMO, extracorporeal membrane oxygenation; NSAIDs, nonsteroidal anti-inflammatory drugs; PEEP, positive end-expiratory pressure; PGE1, prostaglandin E1.

Figure 268-5

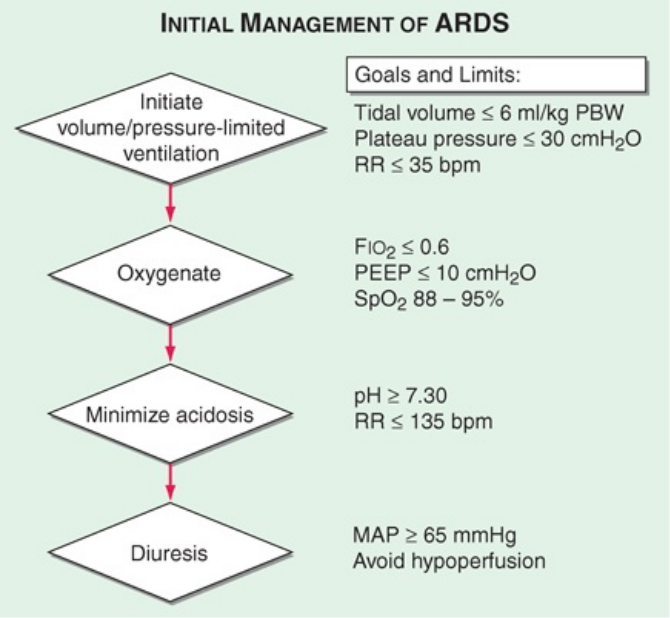

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the initial management of ARDS. Clinical trials have provided evidence-based therapeutic goals for a stepwise approach to the early mechanical ventilation, oxygenation, and correction of acidosis and diuresis of critically ill patients with ARDS.

Prognosis

MORTALITY

Recent mortality estimates for ARDS range from 26 to 44%. There is substantial variability, but a trend toward improved ARDS outcomes appears evident. Of interest, mortality in ARDS is largely attributable to nonpulmonary causes, with sepsis and nonpulmonary organ failure accounting for >80% of deaths. Thus, improvement in survival is likely secondary to advances in the care of septic/infected patients and those with multiple organ failure (Chap. 267).

Several risk factors for mortality to help estimate prognosis have been identified. Similar to the risk factors for developing ARDS, the major risk factors for ARDS mortality are also nonpulmonary. Advanced age is an important risk factor. Patients >75 years of age have a substantially increased mortality (~60%) compared to those <45 (~20%). Also, patients >60 years of age with ARDS and sepsis have a threefold higher mortality compared to those <60. Preexisting organ dysfunction from chronic medical illness is an important additional risk factor for increased mortality. In particular, chronic liver disease, cirrhosis, chronic alcohol abuse, chronic immunosuppression, sepsis, chronic renal disease, any nonpulmonary organ failure, and increased APACHE III scores (Chap. 267) have also been linked to increased ARDS mortality. Several factors related to the presenting clinical disorders also increase the risk for ARDS mortality. Patients with ARDS from direct lung injury (including pneumonia, pulmonary contusion, and aspiration; Table 268-1) have nearly twice the mortality of those with indirect causes of lung injury, while surgical and trauma patients with ARDS, especially those without direct lung injury, have a better survival rate than other ARDS patients.

Surprisingly, there is little value in predicting ARDS mortality from the PaO_2/FIO_2 ratio and any of the following measures of the severity of lung injury: the level of PEEP used in mechanical ventilation, the respiratory compliance, the extent of alveolar infiltrates on chest radiography, and the lung injury score (a composite of all these variables). However, recent data indicate that an early (within 24 hours of presentation) elevation in dead space and the oxygenation index may predict increased mortality from ARDS.

FUNCTIONAL RECOVERY IN ARDS SURVIVORS

While it is common for patients with ARDS to experience prolonged respiratory failure and remain dependent on mechanical ventilation for survival, it is a testament to the resolving powers of the lung

that the majority of patients recover nearly normal lung function. Patients usually recover their maximum lung function within 6 months. One year after endotracheal extubation, more than one-third of ARDS survivors have normal spirometry values and diffusion capacity. Most of the remaining patients have only mild abnormalities in their pulmonary function. Unlike the risk for mortality, recovery of lung function is strongly associated with the extent of lung injury in early ARDS. Low static respiratory compliance, high levels of required PEEP, longer durations of mechanical ventilation, and high lung injury scores are all associated with worse recovery of pulmonary function. When caring for ARDS survivors, it is important to be aware of the potential for a substantial burden of emotional and respiratory symptoms. There are significant rates of depression and posttraumatic stress disorder in ARDS survivors.

Acknowledgment

The authors acknowledge the contribution to this chapter by the previous author, Dr. Steven D. Shapiro.

FURTHER READINGS

Fan E et al: Ventilatory management of acute lung injury and acute respiratory distress syndrome. JAMA 294:2889, 2005[PMID: 16352797]

Tomashefski JF Jr.: Pulmonary pathology of acute respiratory distress syndrome. Clin Chest Med 21:435, 2000[PMID: 11019719]

Ware LB, Matthay MA: The acute respiratory distress syndrome. N Engl J Med 342:1334, 2000[PMID: 10793167]

Wheeler AP, Bernard GR: Acute lung injury and the acute respiratory distress syndrome: A clinical review. Lancet 369:1553, 2007[PMID: 17482987]

WEBSITES

ARDS Support Center for patient-oriented education: *www.ards.org*

ARDS Network clinical trials information: *www.ardsnet.org*

ARDS Foundation: *www.ardsusa.org*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MECHANICAL VENTILATORY SUPPORT

Mechanical ventilation is a therapeutic method that is used to assist or replace spontaneous breathing. The primary indication for initiation of mechanical ventilation is respiratory failure, of which there are two basic types: *hypoxemic* respiratory failure, which is present when arterial O_2 saturation (SaO_2) <90% occurs despite an increased inspired O_2 fraction, and *hypercarbic* respiratory failure, which is characterized by arterial PCO_2 values >50 mmHg. When it is chronic, neither of the two types is obligatorily treated with mechanical ventilation, but when acute, mechanical ventilation may be lifesaving.

Indications

The most common reasons for instituting mechanical ventilation are acute respiratory failure with hypoxemia (acute respiratory distress syndrome, heart failure with pulmonary edema, pneumonia, sepsis, complications of surgery and trauma), which accounts for ~65% of all ventilated cases, followed by causes of hypercarbic ventilatory failure such as coma (15%), exacerbations of chronic obstructive pulmonary disease (13%), and neuromuscular diseases (5%). The primary objectives of mechanical ventilation are to decrease the work of breathing, thus avoiding respiratory muscle fatigue, and to reverse life-threatening hypoxemia and progressive respiratory acidosis.

In some cases, mechanical ventilation is used as an adjunct to other forms of therapy, such as its use in reducing cerebral blood flow in patients with increased intracranial pressure. Mechanical ventilation also is used frequently in conjunction with endotracheal intubation to prevent aspiration of gastric contents in otherwise unstable patients during gastric lavage for suspected drug overdose or during gastrointestinal endoscopy. In critically ill patients, intubation and mechanical ventilation may be indicated before essential diagnostic or therapeutic studies if it appears that respiratory failure may occur during those maneuvers.

Types of Mechanical Ventilation

In its broadest sense, there are two distinct methods for ventilating patients: noninvasive ventilation (NIV) and invasive ventilation or conventional mechanical ventilation (MV).

NONINVASIVE VENTILATION

Noninvasive ventilation has been gaining more acceptance because it is effective in certain conditions, such as acute or chronic respiratory failure, and is associated with fewer complications, namely, pneumonia and tracheolaryngeal trauma. Noninvasive ventilation usually is provided by using a tight-fitting face mask or nasal mask similar to the masks traditionally used for treatment of sleep apnea. Noninvasive ventilation has proved highly effective in patients with respiratory failure from acute exacerbations of chronic obstructive pulmonary disease and is most frequently implemented by using bilevel positive airway pressure ventilation or pressure support ventilation. In both of these modes, a preset positive pressure is applied during inspiration and a lower pressure is applied during expiration at the mask. Both modes are well tolerated by a conscious patient and optimize patient-ventilator synchrony. The major limitation to its widespread application has been patient intolerance because the tight-fitting mask required for NIV can cause both physical and emotional discomfort. In addition, NIV has had limited success in patients with acute hypoxemic respiratory failure, for whom endotracheal intubation and conventional MV remain the ventilatory method of choice.

The most important group of patients who benefit from a trial of NIV are those with acute exacerbations of chronic obstructive pulmonary disease (COPD) leading to respiratory acidosis (pH <7.35). Experience from several well-conducted randomized trials has shown that in patients with ventilatory failure characterized by blood pH levels between 7.25 and 7.35, NIV is associated with low failure rates (15–20%) and good outcomes (intubation rate, length of stay in intensive care, and in some series mortality rates). In more severely ill patients with pH <7.25, the rate of NIV failure is inversely related to the severity of respiratory acidosis, with greater failure as the pH decreases. In patients with milder acidosis (pH >7.35), NIV is not better than conventional therapy that includes controlled oxygen delivery and pharmacotherapy for exacerbations of COPD (systemic corticosteroids, bronchodilators, and, if needed, antibiotics).

Despite its benign outcomes, NIV is not useful in the majority of cases of respiratory failure and is contraindicated in patients with the conditions listed in Table 269-1. Experience shows that NIV can delay lifesaving ventilatory support in those cases and actually results in aspiration or hypoventilation.

Once NIV is initiated, patients should be monitored; a reduction in respiratory frequency and a decrease in the use of accessory muscles (scalene, sternomastoid, and intercostals) are good clinical indicators of adequate therapeutic benefit. Arterial blood gases should be obtained at least within hours of the initiation of therapy to ensure that NIV is having the desired effect and that it is safe to continue its application. Lack of benefit within that time frame should alert one to the possible need for conventional MV.

Table 269–1. Contraindications for Noninvasive Ventilation

Cardiac or respiratory arrest
Severe encephalopathy
Severe gastrointestinal bleed
Hemodynamic instability
Unstable angina and myocardial infarction
Facial surgery or trauma
Upper airway obstruction
High-risk aspiration and/or inability to protect airways
Inability to clear secretions

CONVENTIONAL MECHANICAL VENTILATION

Conventional mechanical ventilation is implemented once a cuffed tube is inserted into the trachea to allow conditioned gas (warmed, oxygenated, and humidified) to be delivered to the airways and lungs at pressures above atmospheric pressure. Great care has to be taken during the act of intubation to avoid brain-damaging hypoxia. In some patients, intubation can be achieved without added sedation. In most patients, the administration of mild sedation may help facilitate the procedure. Opiates and benzodiazepines are good choices but can have a deleterious effect on hemodynamics in patients with depressed cardiac function or low systemic vascular resistance. Morphine can promote histamine release from tissue mast cells and may worsen bronchospasm in patients with asthma; fentanyl, sufentanil, and alfentanil are acceptable alternatives. Ketamine may increase systemic arterial pressure and has been associated with hallucinatory responses; it should be used with caution in patients with hypertensive crisis or a history of psychiatric disorders. Newer agents such as etomidate and propofol have been used for both induction and maintenance of anesthesia in ventilated patients. They are shorter-acting, and etomidate has fewer adverse hemodynamic effects, but both agents are significantly more expensive than older agents. Great care must be taken to avoid the use of neuromuscular paralysis during intubation; in particular, the use of agents whose mechanism of action includes depolarization at the neuromuscular junction, such as succinylcholine chloride, should be avoided in patients with renal failure, tumor lysis syndrome, crush injuries, medical conditions associated with elevated serum potassium levels, and muscular dystrophy syndromes.

Principles of Mechanical Ventilation

Once the patient has been intubated, the basic principles of applying MV are *to optimize oxygenation while avoiding overstretch and collapse/rerecruitment ventilator-induced lung injury (VILI)*. This concept, which is illustrated in Fig. 269-1, has gained acceptance because of important empirical and experimental evidence linking high airway pressures and volumes and overstretching the lung with collapse/rerecruitment with poor outcomes. Although normalization of pH through elimination of CO_2 is desirable, the risk of lung damage associated with the large volume and high pressures needed to achieve this goal has led to the acceptance of permissive hypercapnia. This approach has been found to be well tolerated when care is taken to avoid excess acidosis by pH buffering.

Figure 269-1.

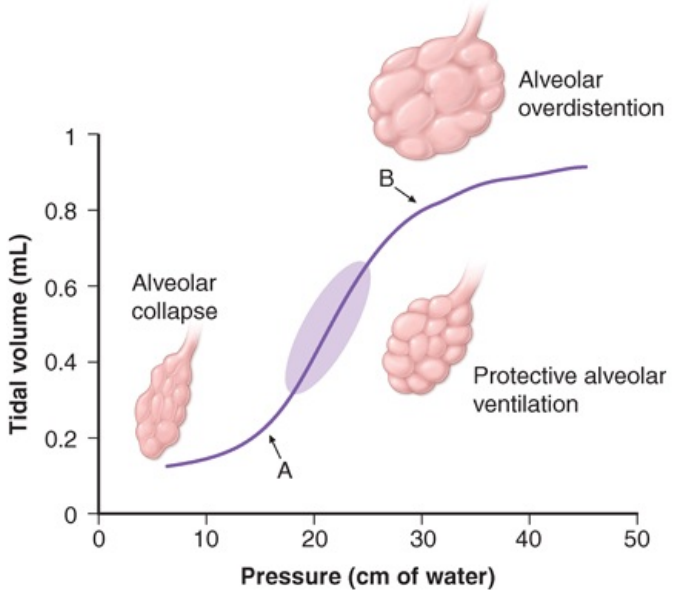

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypothetical pressure-volume curve of the lung in a patient on MV. Alveoli tend to close if the distending pressure falls below the lower inflection point (A), whereas they overstretch if the pressure within them is higher than that of the upper inflection point (B). Collapse and opening of ventilated alveoli are associated with poor outcomes in patients with acute respiratory failure. Protective ventilation (*hatched lines*), using lower tidal volume (6 mL/kg of ideal body weight) and maintaining positive end-expiratory pressure to prevent overstretching and collapse/opening of alveoli, has resulted in improved survival in patients on MV.

Modes of Ventilation

Mode refers to the manner in which ventilator breaths are triggered, cycled, and limited. The *trigger*, either an inspiratory effort or a time-based signal, defines what the ventilator senses to initiate an assisted breath. *Cycle* refers to the factors that determine the end of inspiration. For example, in volume-cycled ventilation, inspiration ends when a specific tidal volume is delivered. Other types of cycling include pressure cycling and time cycling. The *limiting factors* are operator-specified values, such as airway pressure, that are monitored by transducers internal to the ventilator circuit throughout the respiratory cycle; if the specified values are exceeded, inspiratory flow is terminated, and the ventilator circuit is vented to atmospheric pressure or the specified pressure at the end of expiration [positive end-expiratory pressure (PEEP)]. Most patients are ventilated with assist control ventilation, intermittent mandatory ventilation, or pressure-support ventilation, with the latter two modes often used simultaneously (Table 269-2).

Table 269–2. Characteristics of the Most Commonly Used Forms of Mechanical Ventilation

Ventilatory Mode	Variables Set by User (Independent)	Variables Monitored by User (Dependent)	Trigger Cycle Limit	Advantages	Disadvantages
ACMV (assist control ventilation)	Tidal volume Ventilator rate FIO_2 PEEP level Pressure limit	Peak, mean, and plateau airway pressures VE ABG I/E ratio	Patient effort Timer Pressure limit	Patient control Guaranteed ventilation	Potential to hyperventilate Barotrauma and volume trauma Every effective breath generates a ventilator volume
IMV (intermittent mandatory ventilation)	Tidal volume Mandatory Ventilator Rate FIO_2 PEEP level Pressure limit	Peak, mean, and plateau airway pressures VE ABG I/E ratio	Patient effort Timer Pressure limit	Patient control Comfort from spontaneous breaths Guaranteed ventilation	Potential dysynchrony May result in hypoventilation

	Between breaths patients can breathe spontaneously				
PSV (pressure support ventilation)	Inspiratory pressure level FIO_2 PEEP Pressure limit	Tidal volume Respiratory rate VE ABG	Pressure limit Inspiratory flow	Patient control Comfort Assures synchrony	No timer backup May result in hypoventilation
NIV (noninvasive ventilation)	Inspiratory and expiratory pressure level FIO_2	Tidal volume Respiratory rate VE ABG	Pressure limit Inspiratory flow	Patient control	Mask interface may cause discomfort and facial bruising Leaks are common Hypoventilation

Abbreviations: ABG, arterial blood gases; FIO_2, fraction of inspired oxygen; PEEP, positive end-expiratory pressure; I/E, inspiratory to expiratory time ratio; VE, minute ventilation.

ASSIST CONTROL VENTILATION (ACMV)

This is the most widely used mode of ventilation. In this mode, an inspiratory cycle is initiated either by the patient's inspiratory effort or, if none is detected within a specified time window, by a timer signal within the ventilator. Every breath delivered, whether patient- or timer-triggered, consists of the operator-specified tidal volume. Ventilatory rate is determined either by the patient or by the operator-specified backup rate, whichever is of higher frequency. ACMV commonly is used for initiation of mechanical ventilation because it ensures a backup minute ventilation in the absence of an intact respiratory drive and allows for synchronization of the ventilator cycle with the patient's inspiratory effort.

Problems can arise when ACMV is used in patients with tachypnea due to nonrespiratory or nonmetabolic factors, such as anxiety, pain, and airway irritation. Respiratory alkalemia may develop and trigger myoclonus or seizures. Dynamic hyperinflation leading to increased intrathoracic pressures (so-called auto-PEEP) may occur if the patient's respiratory mechanics are such that inadequate time is available for complete exhalation between inspiratory cycles. Auto-PEEP can limit venous return, decrease cardiac output, and increase airway pressures, predisposing to barotrauma.

INTERMITTENT MANDATORY VENTILATION (IMV)

With this mode, the operator sets the number of mandatory breaths of fixed volume to be delivered by the ventilator; between those breaths, the patient can breathe spontaneously. In the most frequently used synchronized mode (SIMV), mandatory breaths are delivered in synchrony with the patient's inspiratory efforts at a frequency determined by the operator. If the patient fails to initiate a breath, the ventilator delivers a fixed-tidal-volume breath and resets the internal timer for the next inspiratory cycle. SIMV differs from ACMV in that only the preset number of breaths is ventilator-assisted.

SIMV allows patients with an intact respiratory drive to exercise inspiratory muscles between assisted breaths, making it useful for both supporting and weaning intubated patients. SIMV may be difficult to use in patients with tachypnea because they may attempt to exhale during the ventilator-programmed inspiratory cycle. When this occurs, the airway pressure may exceed the inspiratory pressure limit, the ventilator-assisted breath will be aborted, and minute volume may drop below that programmed by the operator. In this setting, if the tachypnea is in response to respiratory or metabolic acidosis, a change in ACMV will increase minute ventilation and help normalize the pH while the underlying process is further evaluated and treated.

PRESSURE-SUPPORT VENTILATION (PSV)

This form of ventilation is patient-triggered, flow-cycled, and pressure-limited. It provides graded assistance and differs from the other two modes in that the operator sets the pressure level (rather than the volume) to augment every spontaneous respiratory effort. The level of pressure is adjusted by observing the patient's respiratory frequency. During PSV, the inspiration is terminated when inspiratory airflow falls below a certain level; in most ventilators, this flow rate cannot be adjusted by the operator. When PSV is used, patients receive ventilator assistance only when the ventilator detects an inspiratory effort. PSV frequently is used in combination with SIMV to ensure volume-cycled backup for patients whose respiratory drive is depressed. PSV frequently is well tolerated by most patients who are being weaned; PSV parameters can be set to provide full or nearly full ventilatory support and can be withdrawn to load the respiratory muscles gradually.

There are other modes of ventilation, and each has its own acronym, making it very difficult to

understand for those unfamiliar with the terms. All these modes are modifications of the manner and duration in which pressure is applied to the airway and lungs and of the interaction between the mechanical assistance provided by the ventilator and the patient's respiratory effort. Although their use in acute respiratory failure is limited, the following have been used with varying levels of enthusiasm and adoption.

Pressure-Control Ventilation (PCV)

This form of ventilation is time-triggered, time-cycled, and pressure-limited. During the inspiratory phase, a specified pressure is imposed at the airway opening throughout inspiration. Since the inspiratory airway pressure is specified by the operator, tidal volume and inspiratory flow rate are *dependent*, rather than *independent*, variables and are not operator-specified. PCV is the preferred mode of ventilation for patients in whom it is desirable to regulate peak airway pressures, such as those with preexisting barotrauma, and postoperative thoracic surgical patients, in whom the shear forces across a fresh suture line should be limited. When PCV is used, minute ventilation and tidal volume must be monitored closely; minute ventilation is altered through changes in rate or in the pressure-control value, which changes tidal volume.

Inverse Ratio Ventilation (IRV)

This mode of ventilation is a variant of PCV that incorporates the use of a prolonged inspiratory time with the appropriate shortening of the expiratory time. It has been used in patients with severe hypoxemic respiratory failure. This approach increases mean distending pressures without increasing peak airway pressures. It is thought to work in conjunction with PEEP to open collapsed alveoli and improve oxygenation, although there are no conclusive data showing that IRV improves outcomes in clinical trials.

Continuous Positive Airway Pressure (CPAP)

This is not a true support mode of ventilation because all ventilation occurs through the patient's spontaneous efforts. The ventilator provides fresh gas to the breathing circuit with each inspiration and sets the circuit to a constant, operator-specified pressure. CPAP is used to assess extubation potential in patients who have been effectively weaned and require little ventilator support and patients with intact respiratory system function who require an endotracheal tube for airway protection.

Nonconventional Ventilatory Strategies

Several nonconventional ventilator strategies have been evaluated for their ability to improve oxygenation and reduce mortality rates in patients with advanced hypoxemic respiratory failure. These strategies include high-frequency oscillatory ventilation (HFOV), airway pressure release ventilation (APRV), extracorporeal membrane oxygenation (BCMO), and partial liquid ventilation (PLV) using perfluorocarbons. Although case reports and small uncontrolled cohort studies have shown benefit, randomized controlled trials have failed to demonstrate consistent improvements in outcome with any of these strategies. Currently, these approaches should be considered "salvage" techniques and considered for patients with hypoxemia refractory to conventional therapy. Prone positioning of patients with refractory hypoxemia has been explored because in theory it would tend to improve ventilation-perfusion matching. Although this is conceptually appealing and simple to implement, several randomized trial in patients with acute lung injury did not demonstrate a survival advantage with prone positioning despite demonstration of a transient physiologic benefit. The administration of nitric oxide (NO) gas, which has bronchodilator and pulmonary vasodilator effects when delivered through the airways and has been shown to improve arterial oxygenation in many patients with advanced hypoxemic respiratory failure, also failed to improve outcomes in patients with advanced hypoxemic respiratory failure.

Newer, promising strategies are intended to improve patient-ventilator synchrony, a major practical problem during MV. Currently, the more advanced new ventilators allow patients to trigger the ventilator with their own effort while also incorporating flow algorithms that allow termination of cycles once certain preset criteria are reached; this approach has greatly improved patient-ventilator synchrony and comfort. More recently, new modes of ventilation that synchronize not only the timing but also the levels of assistance to match the patient's effort have been developed. Proportional assist ventilation (PAV) and neurally adjusted ventilatory assist ventilation (NAV) are two modes that are designed to deliver assisted breaths through algorithms incorporating not only pressure, volume, and time but also overall respiratory resistance and compliance in the case of PAV and neural activation of the diaphragm in the case of NAV. Although these modes result in better patient-ventilator synchrony, their practical use in the everyday management of patients on MV needs further study.

Protective Ventilatory Strategy

Whichever mode of MV is used, in acute respiratory failure the evidence from several important controlled trials indicates that the use of a protective ventilation approach guided by the principles outlined below and summarized in Fig. 269-1 is safe and offers the best chance of a good outcome:

1. Set a target tidal volume close to 6 mL/kg of ideal body weight.

2. Prevent plateau pressure (static pressure in the airway at the end of inspiration) over 30 cmH_2O.
3. Use the lowest possible fraction of inspired oxygen (FIO_2) to keep SaO_2 ≥90%.
4. Adjust the PEEP to maintain alveolar patency while preventing overdistention and closure/reopening.

With the application of these techniques, the mortality rate among patients with acute hypoxemic respiratory failure has decreased to ~30% from close to 50% a decade ago.

Patient Management

Once the patient has been stabilized with respect to gas exchange, definitive therapy for the underlying process responsible for respiratory failure is initiated. Subsequent modifications in ventilator therapy must be provided in parallel with changes in the patient's clinical status. As improvement in respiratory function is noted, the first priority is to reduce the level of mechanical ventilator support. Patients on full ventilator support should be monitored frequently with the goal of switching to a mode that allows for weaning as soon as possible. Protocols and guidelines that can be applied by paramedical personnel when physicians are not readily available have proved to be of value in shortening ventilator and intensive care unit (ICU) time, with very good outcomes. Patients whose condition continues to deteriorate after ventilator support is initiated may require increased O_2, PEEP, or one of the alternative modes of ventilation.

GENERAL SUPPORT DURING VENTILATION

Patients started on mechanical ventilation usually require sedation and analgesia to maintain an acceptable level of comfort. Often, this consists of a combination of a benzodiazepine and an opiate administered intravenously. Medications commonly used for this purpose include lorazepam, midazolam, diazepam, morphine, and fentanyl. The use of oversedation must be avoided in the ICU. Indeed, recent trials evaluating the effect of daily interruption of sedation in patients with improved ventilatory status show that this results in shorter time on the ventilator and shorter ICU stay.

Immobilized patients in the ICU who are on mechanical ventilator support are at increased risk for deep venous thrombosis and decubitus ulcers. To prevent venous thrombosis, prophylaxis in the form of subcutaneous heparin and/or pneumatic compression boots is prescribed frequently. Fractionated low-molecular-weight heparin appears to be equally effective for this purpose. To help prevent decubitus ulcers, frequent changes in body position and the use of soft mattress overlays and air mattresses are employed. Prophylaxis against diffuse gastrointestinal mucosal injury is indicated for patients on MV. Histamine-receptor antagonists (H_2-receptor antagonists), antacids, and cytoprotective agents such as Carafate (sucralfate) have all been used for this purpose and appear to be effective. Nutrition support by enteral feeding through either a nasogastric or an orogastric tube should be initiated and maintained whenever possible. Delayed gastric emptying is common in critically ill patients on sedative medications but often responds to promotility agents such as metoclopramide. Parenteral nutrition is an alternative to enteral nutrition in patients with severe gastrointestinal pathology who need prolonged MV.

Complications of Mechanical Ventilation

Endotracheal intubation and mechanical ventilation have direct and indirect effects on the lung and upper airways, the cardiovascular system, and the gastrointestinal system. Pulmonary complications include barotrauma, nosocomial pneumonia, oxygen toxicity, tracheal stenosis, and deconditioning of respiratory muscles. Barotrauma and volutrauma overdistend and disrupt lung tissue; may be clinically manifest by interstitial emphysema, pneumomediastinum, subcutaneous emphysema, or pneumothorax; and can result in the liberation of cytokines from overdistended tissues, further promoting tissue injury. Clinically significant pneumothorax requires tube thoracostomy. Intubated patients are at high risk for ventilator-associated pneumonia (VAP) as a result of aspiration from the upper airways through small leaks around the endotracheal tube cuff; the most common organisms responsible for this condition are *Pseudomonas aeruginosa*, enteric gram-negative rods, and *Staphylococcus aureus*. Because this condition is associated with high mortality rates, early initiation of empirical antibiotics directed against likely pathogens is recommended. *Hypotension* resulting from elevated intrathoracic pressures with decreased venous return is almost always responsive to intravascular volume repletion. In patients who are judged to have respiratory failure on the basis of alveolar edema but in whom the cardiac or pulmonary origin of the edema is unclear, hemodynamic monitoring with a pulmonary arterial catheter may be of value in helping to clarify the cause of the edema.

Gastrointestinal effects of positive-pressure ventilation include stress ulceration and mild to moderate cholestasis.

Weaning from Mechancal Ventilation

It is important to consider discontinuation of mechanical ventilation once the underlying respiratory disease begins to reverse. Although the predictive capacities of multiple clinical and physiologic

variables have been explored, the consensus from a weaning task force includes the following recommendations: (1) lung injury is stable/resolving, (2) gas exchange is adequate with low $PEEP/FIO_2$ (<8 cmH_2O and FIO <0.5), (3) hemodynamic variables are stable (patient off vasopressors), and (4) patient is capable of initiating spontaneous breaths. This "screen" should be done at least daily. If the patient is deemed capable of beginning weaning, the recommendation of the task force is to perform a spontaneous breathing trial (SBT) because several randomized trials support the value of this approach (Fig. 269-2). The SBT involves an integrated patient assessment during spontaneous breathing with little or no ventilator support. The SBT is usually implemented with a T-piece using 1–5 cmH_2O CPAP or a T-piece with 5–7 cmH_2O or PSV from the ventilator to offset the resistance from the endotracheal tube. Once it is determined that the patient can breath spontaneously, a decision must be made about the removal of the artificial airway; this should be done only when it is concluded that the patient has the ability to protect the airway, is able to cough and clear secretions, and is alert enough to follow commands. In addition, other factors must be taken into account, such as the possible difficulty in replacing the tube if that is anticipated. If upper airway difficulty is suspected, an evaluation using a "cuff leak" test (assessing the presence of air movement around a deflated endotracheal tube cuff) is supported by some internists. Despite the application of all of these methods, ~10–15% of extubated patients require reintubation. Several studies suggest that NIV can be used to avert reintubation; this has been particularly useful in patients with ventilatory failure secondary to COPD exacerbation. In this group, earlier extubation with the use of prophylactic NIV has shown good results. The use of NIV to facilitate weaning in other causes of respiratory failure is not currently indicated.

Figure 269-2.

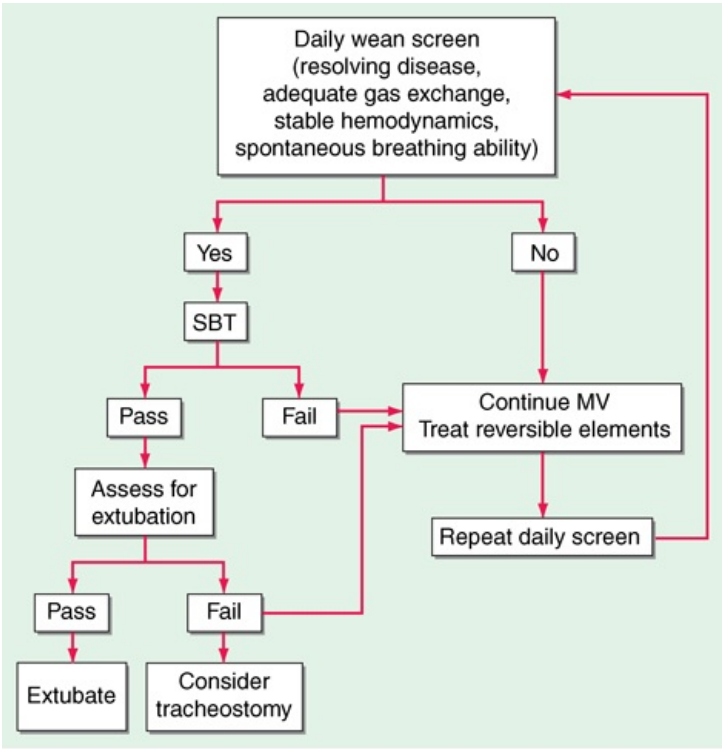

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Flow chart to guide daily approach to managing patients considered for weaning. If the patient fails attempts at extubation, a tracheostomy should be considered.

PROLONGED MECHANICAL VENTILATION AND TRACHEOSTOMY

From 5 to 13% of patients on MV will go on to require prolonged MV (>21 days). In these patients, critical care personnel have to make a decision about whether and when to perform a tracheostomy. This decision is individually based on the risk and benefits of tracheostomy and prolonged intubation and the patient's preferences and expected clinical outcomes. A tracheostomy is thought to be more comfortable, require less sedation, and provide a more secure airway, and it seems to reduce weaning time. However, tracheostomy carries the risk of complications, which occur in 5–40% of the procedures and include bleeding, cardiopulmonary arrest, hypoxia due to airway loss, structural damage, postoperative pneumothorax, pneumomediastinum, and wound infection. In patients with long-term tracheostomy, tracheal stenosis, granulation, and the erosion of the innominate artery are complex complications. It is generally agreed that if a patient is in need of MV for more than 10–14

days, a tracheostomy is indicated and should be planned under optimal conditions. Whether it is completed at the bedside or as an operative procedure depends on the local resources and experience. Some 5–10% of patients are deemed unable to wean in the ICU. These patients may benefit from transfer to special units where a multidisciplinary approach, including nutrition optimization, physical therapy with rehabilitation, and slower weaning methods, including SIMV with PSV, results in up to 30% successful weaning. Unfortunately, close to 2% of ventilated patients may ultimately remain unable to wean and become dependent on ventilatory support to maintain life. Most of these patients remain in chronic care institutions, although some who have strong social, economical, and family support may achieve a fulfilling life with home mechanical ventilation.

Acknowledgment

Portions of this chapter were retained from the work of the previous author, Edward Ingenito, MD.

FURTHER READINGS

The Acute Respiratory Distress Syndrome Network: Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 342:1301, 2000

Boles JM et al: Weaning from mechanical ventilation. Eur Respir J 29:1033, 2007[PMID: 17470624]

Esteban A et al: Non-invasive ventilation for respiratory failure after extubation. N Engl J Med 350:2452, 2004[PMID: 15190137]

MacIntyre N (ed): *Controversies in Mechanical Ventilation*. Clinics in Chest Medicine. Philadelphia, Elsevier Saunders, 2008

Mercat A et al: Positive and end-expiratory pressure setting in adults with acute lung injury and acute respiratory distress syndrome: A randomized controlled trial. JAMA 299:646, 2008[PMID: 18270353]

Scalise PI, Votto J: Weaning from long term mechanical ventilation. Chron Respir Dis 2:99, 2005[PMID: 16279157]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH SHOCK: INTRODUCTION

Shock is the clinical syndrome that results from inadequate tissue perfusion. Irrespective of cause, the hypoperfusion-induced imbalance between the delivery of and requirements for oxygen and substrate leads to cellular dysfunction. The cellular injury created by the inadequate delivery of oxygen and substrates also induces the production and release of damage-associated molecular patterns (DAMPs or "danger signals") and inflammatory mediators that further compromise perfusion through functional and structural changes within the microvasculature. This leads to a vicious cycle in which impaired perfusion is responsible for cellular injury that causes maldistribution of blood flow, further compromising cellular perfusion; the latter ultimately causes multiple organ failure (MOF) and, if the process is not interrupted, leads to death. The clinical manifestations of shock are also the result, in part, of autonomic neuroendocrine responses to hypoperfusion as well as the breakdown in organ function induced by severe cellular dysfunction **(Fig. 270-1)**.

Figure 270-1

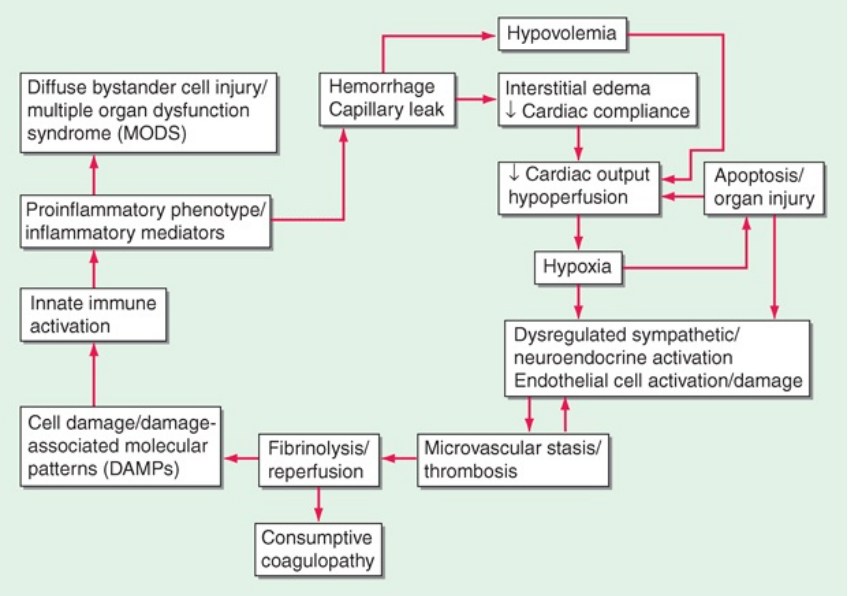

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Shock-induced vicious cycle.

When very severe and/or persistent, inadequate oxygen delivery leads to irreversible cell injury; only rapid restoration of oxygen delivery can reverse the progression of the shock state. The fundamental approach to management, therefore, is to recognize overt and impending shock in a timely fashion and to intervene emergently to restore perfusion. This often requires the expansion or reexpansion of intravascular blood volume. Control of any inciting pathologic process (e.g., continued hemorrhage, impairment of cardiac function, or infection), must occur simultaneously.

Clinical shock is usually accompanied by hypotension (i.e., a mean arterial pressure (MAP) <60 mmHg in previously normotensive persons). Multiple classification schemes have been developed in an attempt to synthesize the seemingly dissimilar processes leading to shock. Strict adherence to a classification scheme may be difficult from a clinical standpoint because of the frequent combination of two or more causes of shock in any individual patient, but the classification shown in **Table 270-1** provides a useful reference point from which to discuss and further delineate the underlying processes.

Table 270–1. Classification of Shock

Hypovolemic	Septic
Traumatic	Hyperdynamic (early)
Cardiogenic	Hypodynamic (late)
Intrinsic	Neurogenic
Compressive	Hypoadrenal

PATHOGENESIS AND ORGAN RESPONSE

Microcirculation

Normally when cardiac output falls, systemic vascular resistance rises to maintain a level of systemic pressure that is adequate for perfusion of the heart and brain at the expense of other tissues such as muscle, skin, and especially the gastrointestinal (GI) tract. Systemic vascular resistance is determined primarily by the luminal diameter of arterioles. The metabolic rates of the heart and brain are high, and their stores of energy substrate are low. These organs are critically dependent on a continuous supply of oxygen and nutrients, and neither tolerates severe ischemia for more than brief periods (minutes). Autoregulation (i.e., the maintenance of blood flow over a wide range of perfusion pressures), is critical in sustaining cerebral and coronary perfusion despite significant hypotension. However, when MAP drops to ≤60 mmHg, blood flow to these organs falls, and their function deteriorates.

Arteriolar vascular smooth muscle has both α- and β-adrenergic receptors. The α_1 receptors mediate vasoconstriction, while the β_2 receptors mediate vasodilation. Efferent sympathetic fibers release norepinephrine, which acts primarily on α_1 receptors as one of the most fundamental compensatory responses to reduced perfusion pressure. Other constrictor substances that are increased in most forms of shock include angiotensin II, vasopressin, endothelin 1, and thromboxane A_2. Both norepinephrine and epinephrine are released by the adrenal medulla, and the concentrations of these catecholamines in the bloodstream rise. Circulating vasodilators in shock include prostacyclin [prostaglandin (PG) I_2], nitric oxide (NO), and, importantly, products of local metabolism such as adenosine that match flow to the tissue's metabolic needs. The balance between these various vasoconstrictors and vasodilators influences acting upon the microcirculation determines local perfusion.

Transport to cells depends on microcirculatory flow; capillary permeability; the diffusion of oxygen, carbon dioxide, nutrients, and products of metabolism through the interstitium; and the exchange of these products across cell membranes. Impairment of the microcirculation that is central to the pathophysiologic responses in the late stages of all forms of shock, results in the derangement of cellular metabolism that is ultimately responsible for organ failure.

The endogenous response to mild or moderate hypovolemia is an attempt at restitution of intravascular volume through alterations in hydrostatic pressure and osmolarity. Constriction of arterioles leads to reductions in both the capillary hydrostatic pressure and the number of capillary beds perfused, thereby limiting the capillary surface area across which filtration occurs. When filtration is reduced while intravascular oncotic pressure remains constant or rises, there is net reabsorption of fluid into the vascular bed, in accord with Starling's law of capillary interstitial liquid exchange. Metabolic changes (including hyperglycemia and elevations in the products of glycolysis, lipolysis, and proteolysis) raise extracellular osmolarity, leading to an osmotic gradient that increases interstitial and intravascular volume at the expense of intracellular volume.

Cellular Responses

Interstitial transport of nutrients is impaired in shock, leading to a decline in intracellular high-energy phosphate stores. Mitochondrial dysfunction and uncoupling of oxidative phosphorylation are the most likely causes for decreased amounts of adenosine triphosphate (ATP). As a consequence, there is an accumulation of hydrogen ions, lactate, and other products of anaerobic metabolism. As shock progresses, these vasodilator metabolites override vasomotor tone, causing further hypotension and hypoperfusion. Dysfunction of cell membranes is thought to represent a common end-stage pathophysiologic pathway in the various forms of shock. Normal cellular transmembrane potential falls, and there is an associated increase in intracellular sodium and water, leading to cell swelling that interferes further with microvascular perfusion. In a preterminal event, homeostasis of calcium via membrane channels is lost with flooding of calcium intracellularly and a concomitant extracellular hypocalcemia. There is also evidence for a widespread but selective apoptotic (programmed cell-death) loss of cells, contributing to organ and immune failure.

Neuroendocrine Response

Hypovolemia, hypotension, and hypoxia are sensed by baroreceptors and chemoreceptors that contribute to an autonomic response that attempts to restore blood volume, maintain central perfusion, and mobilize metabolic substrates. Hypotension disinhibits the vasomotor center, resulting in increased adrenergic output and reduced vagal activity. Release of norepinephrine from adrenergic neurons induces significant peripheral and splanchnic vasoconstriction, a major contributor to the maintenance of central organ perfusion, while reduced vagal activity increases the heart rate and cardiac output. Loss of vagal activity is also recognized to upregulate the innate immunity inflammatory response. The effects of circulating epinephrine released by the adrenal medulla in shock are largely metabolic, causing increased glycogenolysis and gluconeogenesis and reduced pancreatic insulin release. However, epinephrine also inhibits production and release of inflammatory mediators through stimulation of β-adrenergic receptors on innate immune cells.

Severe pain or other stresses cause the hypothalamic release of adrenocorticotropic hormone (ACTH). This stimulates cortisol secretion that contributes to decreased peripheral uptake of glucose and amino acids, enhances lipolysis, and increases gluconeogenesis. Increased pancreatic secretion of glucagon during stress accelerates hepatic gluconeogenesis and further elevates blood glucose concentration. These hormonal actions act synergistically to increase blood glucose for both selective tissue metabolism and the maintenance of blood volume. Many critically ill patients have recently been shown to exhibit low plasma cortisol levels and an impaired response to ACTH stimulation, which is linked to a decrease in survival. The importance of the cortisol response to stress is illustrated by the profound circulatory collapse that occurs in patients with adrenal cortical insufficiency (Chap. 342).

Renin release is increased in response to adrenergic discharge and reduced perfusion of the juxtaglomerular apparatus in the kidney. Renin induces the formation of angiotensin I that is then converted to angiotensin II, an extremely potent vasoconstrictor and stimulator of aldosterone release by the adrenal cortex and of vasopressin by the posterior pituitary. Aldosterone contributes to the maintenance of intravascular volume by enhancing renal tubular reabsorption of sodium, resulting in the excretion of a low-volume, concentrated, sodium-free urine. Vasopressin has a direct action on vascular smooth muscle, contributing to vasoconstriction, and acts on the distal renal tubules to enhance water reabsorption.

Cardiovascular Response

Three variables—ventricular filling (preload), the resistance to ventricular ejection (afterload), and myocardial contractility—are paramount in controlling stroke volume (Chap. 224). Cardiac output, the major determinant of tissue perfusion, is the product of stroke volume and heart rate. Hypovolemia leads to decreased ventricular preload that in turn reduces the stroke volume. An increase in heart rate is a useful but limited compensatory mechanism to maintain cardiac output. A shock-induced reduction in myocardial compliance is frequent, reducing ventricular end-diastolic volume and hence stroke volume at any given ventricular filling pressure. Restoration of intravascular volume may return stroke volume to normal but only at elevated filling pressures. Increased filling pressures stimulate release of brain natriuretic peptide (BNP) to secrete sodium and volume to relieve the pressure on the heart. Levels of BNP correlate with outcome following severe stress. In addition, sepsis, ischemia, myocardial infarction (MI), severe tissue trauma, hypothermia, general anesthesia, prolonged hypotension, and acidemia may all also impair myocardial contractility and reduce the stroke volume at any given ventricular end-diastolic volume. The resistance to ventricular ejection is significantly influenced by the systemic vascular resistance, which is elevated in most forms of shock. However, resistance is depressed in the early hyperdynamic stage of septic shock (Chap. 271) or neurogenic shock, thereby initially allowing the cardiac output to be maintained or elevated.

The venous system contains nearly two-thirds of the total circulating blood volume, most in the small veins, and serves as a dynamic reservoir for autoinfusion of blood. Active venoconstriction as a consequence of α-adrenergic activity is an important compensatory mechanism for the maintenance of venous return and therefore of ventricular filling during shock. On the other hand, venous dilation, as occurs in neurogenic shock, reduces ventricular filling and hence stroke volume and potentially cardiac output.

Pulmonary Response

The response of the pulmonary vascular bed to shock parallels that of the systemic vascular bed, and the relative increase in pulmonary vascular resistance, particularly in septic shock, may exceed that of the systemic vascular resistance, leading to right heart failure. Shock-induced tachypnea reduces tidal volume and increases both dead space and minute ventilation. Relative hypoxia and the subsequent tachypnea induce a respiratory alkalosis. Recumbency and involuntary restriction of ventilation secondary to pain reduce functional residual capacity and may lead to atelectasis. Shock and, in particular, resuscitation-induced oxidant radical generation, is recognized as a major cause of acute lung injury and subsequent acute respiratory distress syndrome (ARDS; Chap. 268). These disorders are characterized by noncardiogenic pulmonary edema secondary to diffuse pulmonary capillary endothelial and alveolar epithelial injury, hypoxemia, and bilateral diffuse pulmonary infiltrates. Hypoxemia results from perfusion of underventilated and nonventilated alveoli. Loss of surfactant and lung volume in combination with increased interstitial and alveolar edema reduces lung compliance. The work of breathing and the oxygen requirements of respiratory muscles increase.

Renal Response

Acute kidney injury (Chap. 279), a serious complication of shock and hypoperfusion, occurs less frequently than heretofore because of early aggressive volume repletion. Acute tubular necrosis is now more frequently seen as a result of the interactions of shock, sepsis, the administration of nephrotoxic agents (such as aminoglycosides and angiographic contrast media), and rhabdomyolysis; the latter may be particularly severe in skeletal muscle trauma. The physiologic response of the kidney to hypoperfusion is to conserve salt and water. In addition to decreased renal blood flow, increased afferent arteriolar resistance accounts for diminished glomerular filtration rate (GFR) that together with increased aldosterone and vasopressin is responsible for reduced urine formation. Toxic injury causes necrosis of tubular epithelium and tubular obstruction by cellular debris with back leak of filtrate. The depletion of renal ATP stores that occurs with prolonged renal hypoperfusion contributes to subsequent impairment of renal function.

Metabolic Derangements

During shock, there is disruption of the normal cycles of carbohydrate, lipid, and protein metabolism. Through the citric acid cycle, alanine in conjunction with lactate, which is converted from pyruvate in the periphery in the presence of oxygen deprivation enhances the hepatic production of glucose. With reduced availability of oxygen, the breakdown of glucose to pyruvate, and ultimately lactate, represents an inefficient cycling of substrate with minimal net energy production. An elevated plasma lactate/pyruvate ratio is preferable to lactate alone as a measure of anaerobic metabolism and reflects inadequate tissue perfusion. Decreased clearance of exogenous triglycerides coupled with increased hepatic lipogenesis causes a significant rise in serum triglyceride concentrations. There is increased protein catabolism as energy substrate, a negative nitrogen balance, and, if the process is prolonged, severe muscle wasting.

Inflammatory Responses

Activation of an extensive network of proinflammatory mediator systems by the innate immune system plays a significant role in the progression of shock and contributes importantly to the development of multiple organ injury, dysfunction (MOD), and failure (MOF) **(Fig. 270-2)**. In those surviving the acute insult, there is a prolonged endogenous counterregulatory response to "turn off" or balance the excessive proinflammatory response. If balance is restored, the patient does well. If the response is excessive, adaptive immunity is suppressed and the patient is highly susceptible to secondary nosocomial infections, which may then drive the inflammatory response and lead to delayed MOF.

Figure 270-2

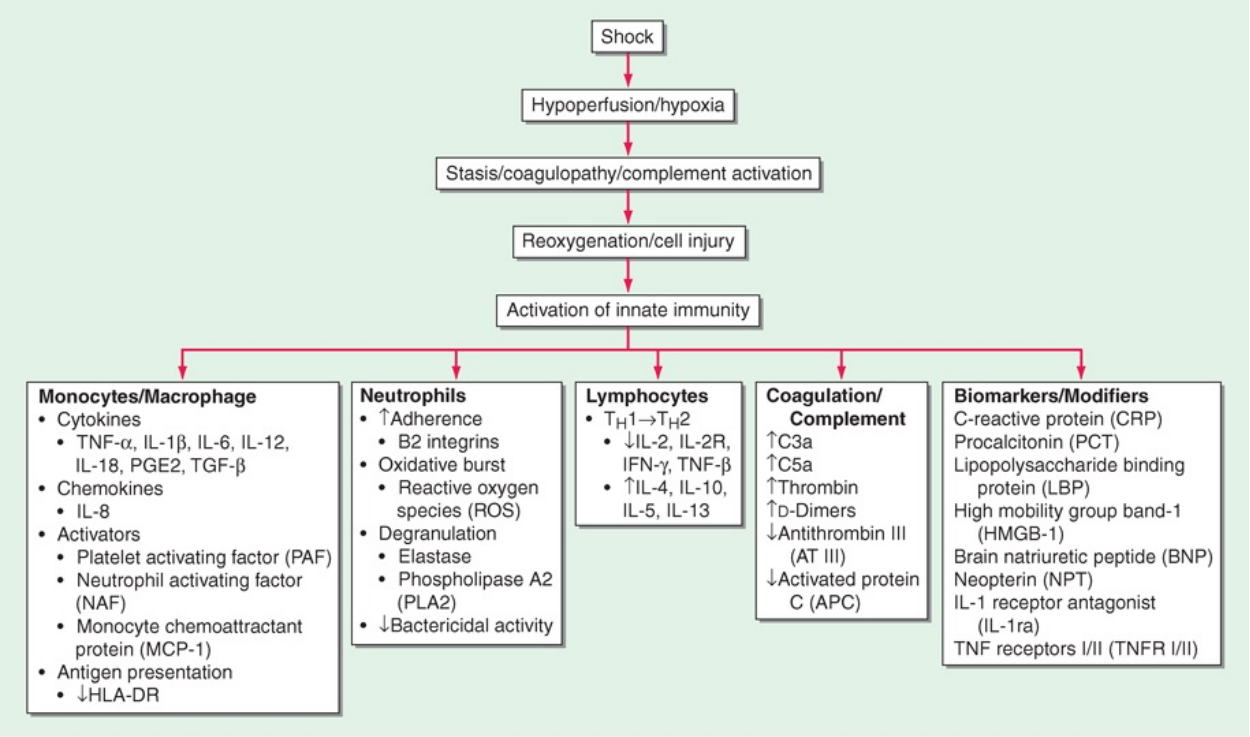

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A schematic of the host immunoinflammatory response to shock.

Multiple humoral mediators are activated during shock and tissue injury. The complement cascade, activated through both the classic and alternate pathways, generates the anaphylatoxins C3a and C5a (Chap. 314). Direct complement fixation to injured tissues can progress to the C5-C9 attack complex, causing further cell damage. Activation of the coagulation cascade (Chap. 116) causes microvascular thrombosis, with subsequent fibrinolysis leading to repeated episodes of ischemia and reperfusion. Components of the coagulation system (e.g., thrombin), are potent proinflammatory mediators that cause expression of adhesion molecules on endothelial cells and activation of neutrophils, leading to microvascular injury. Coagulation also activates the kallikrein-kininogen cascade, contributing to hypotension.

Eicosanoids are vasoactive and immunomodulatory products of arachidonic acid metabolism that include cyclooxygenase-derived prostaglandins (PGs) and thromboxane A_2, as well as lipoxygenase-derived leukotrienes and lipoxins. Thromboxane A_2 is a potent vasoconstrictor that contributes to the pulmonary hypertension and acute tubular necrosis of shock. PGI2 and PGE2 are potent vasodilators that enhance capillary permeability and edema formation. The cysteinyl leukotrienes LTC_4 and LTD_4 are pivotal mediators of the vascular sequelae of anaphylaxis, as well as of shock states resulting from sepsis or tissue injury. LTB_4 is a potent neutrophil chemoattractant and secretagogue that stimulates the formation of reactive oxygen species. Platelet-activating factor, an ether-linked, arachidonyl-containing phospholipid mediator, causes pulmonary vasoconstriction, bronchoconstriction, systemic vasodilation, increased capillary permeability, and the priming of macrophages and neutrophils to produce enhanced levels of inflammatory mediators.

Tumor necrosis factor α (TNF-α), produced by activated macrophages, reproduces many components of the shock state, including hypotension, lactic acidosis, and respiratory failure. Interleukin 1β (IL-1β), originally defined as "endogenous pyrogen" and produced by tissue-fixed macrophages, is critical to the inflammatory response. Both are significantly elevated immediately following trauma and shock. IL-6, also produced predominantly by the macrophage, has a slightly delayed peak response but is the best single predictor of prolonged recovery and development of MOF following shock. Chemokines such as IL-8 are potent neutrophil chemoattractants and activators that upregulate adhesion molecules on the neutrophil to enhance aggregation, adherence, and damage to the vascular endothelium. While the endothelium normally produces low levels of NO, the inflammatory response stimulates the inducible isoform of NO synthase (iNOS), which is overexpressed and produces toxic nitrosyl- and oxygen-derived free radicals that contribute to the hyperdynamic cardiovascular response and tissue injury in sepsis.

Multiple inflammatory cells, including neutrophils, macrophages, and platelets, are major contributors to inflammation-induced injury. Margination of activated neutrophils in the microcirculation is a common pathologic finding in shock, causing secondary injury due to the release of toxic oxygen radicals, lipases, primarily PLA2, and proteases. Release of high levels of reactive oxygen intermediates/species (ROI/ROS) rapidly consumes endogenous essential antioxidants and generates diffuse oxygen radical damage. Newer efforts to control ischemia/reperfusion injury include treatment with carbon monoxide, hydrogen sulfide, or other agents to reduce oxidant stress. Tissue-fixed macrophages produce virtually all major mediators of the inflammatory response and orchestrate the progression and duration of the inflammatory response. A major source of activation of the monocyte/macrophage is through the highly conserved membrane toll-like receptors (TLRs) that recognize DAMPs such as HMGB-1, and pathogen-associated molecular patterns (PAMPs) such as endotoxins released following tissue injury, and by pathogenic microbial organisms, respectively. Toll-like receptors also appear important for the

chronic inflammation seen in Crohn's disease, ulcerative colitis, and transplant rejection. The variability in individual responses is a genetic predisposition that, in part, is due to single nucleotide polymorphisms (SNPs) in genetic sequences affecting the function and production of various inflammatory mediators.

Treatment: Shock

MONITORING

Patients in shock require care in an ICU. Careful and continuous assessment of the physiologic status is necessary. Arterial pressure through an indwelling line, pulse, and respiratory rate should be monitored continuously; a Foley catheter should be inserted to follow urine flow; and mental status should be assessed frequently. Sedated patients should be allowed to awaken ("drug holiday") daily to assess their neurologic status and to shorten duration of ventilator support.

There is ongoing debate as to the indications for using the flow-directed pulmonary artery catheter (PAC, Swan-Ganz catheter). Most patients in the ICU can be safely managed without the use of a PAC. However, in shock with significant ongoing blood loss, fluid shifts, and underlying cardiac dysfunction, a PAC may be useful. The PAC is placed percutaneously via the subclavian or jugular vein through the central venous circulation and right heart into the pulmonary artery. There are ports both proximal in the right atrium and distal in the pulmonary artery to provide access for infusions and for cardiac output measurements. Right atrial and pulmonary artery pressures (PAPs) are measured, and the pulmonary capillary wedge pressure (PCWP) serves as an approximation of the left atrial pressure. Normal hemodynamic parameters are shown in Table 230-2 and **Table 270-2**.

Table 270–2. Normal Hemodynamic Parameters

Parameter	Calculation	Normal Values
Cardiac output (CO)	SV x HR	4–8 L/min
Cardiac index (CI)	CO/BSA	2.6–4.2 (L/min)/m^2
Stroke volume (SV)	CO/HR	50–100 mL/beat
Systemic vascular resistance (SVR)	[(MAP – RAP)/CO] x 80	700–1600 dynes · s/cm^5
Pulmonary vascular resistance (PVR)	[(PAP_m – PCWP)/CO] x 80	20–130 dynes · s/cm^5
Left ventricular stroke work (LVSW)	SV(MAP – PCWP) x 0.0136	60–80 g-m/beat
Right ventricular stroke work (RVSW)	SV(PAP_m – RAP)	10–15 g-m/beat

Abbreviations: BSA, body surface area; HR, heart rate; MAP, mean arterial pressure; PAP_m, pulmonary artery pressure—mean; PCWP, pulmonary capillary wedge pressure; RAP, right atrial pressure.

Cardiac output is determined by the thermodilution technique, and high-resolution thermistors can also be used to determine right ventricular end-diastolic volume to monitor further the response of the right heart to fluid resuscitation. A PAC with an oximeter port offers the additional advantage of on-line monitoring of the mixed venous oxygen saturation, an important index of overall tissue perfusion. Systemic and pulmonary vascular resistances are calculated as the ratio of the pressure drop across these vascular beds to the cardiac output (Chap. 230). Determinations of oxygen content in arterial and venous blood, together with cardiac output and hemoglobin concentration, allow calculation of oxygen delivery, oxygen consumption, and oxygen-extraction ratio **(Table 270-3)**. The hemodynamic patterns associated with the various forms of shock are shown in **Table 270-4**.

Table 270–3. Oxygen Transport Calculations

Parameter	Calculation	Normal Values
Oxygen-carrying capacity of hemoglobin		1.39 mL/g
Plasma O_2 concentration		P_{O2} x 0.0031
Arterial O_2 concentration (Ca_{O2})	1.39 Sa_{O2} + 0.0031 Pa_{O2}	20 vol%
Venous O_2 concentration (Cv_{O2})	1.39 Sv_{O2} + 0.0031 Pv_{O2}	15.5 vol%
Arteriovenous O_2 difference (Ca_{O2} – Cv_{O2})	1.39 (Sa_{O2} – Sv_{O2}) 0.0031 (Pa_{O2} – Pv_{O2})	3.5 vol%
Oxygen delivery (D_{O2})	Ca_{O2} x CO (L/min) x 10 (dL/L) 1.39 Sa_{O2} x CO x 10	800–1600 mL/min
Oxygen uptake (V_{O2})	(Ca_{O2} – Cv_{O2}) x CO x 10 1.39 (Sa_{O2} – Sv_{O2}) x CO x 10	150–400 mL/min
Oxygen delivery index (D_{O2} I)	D_{O2}/BSA	520–720 (mL/min)/m^2
Oxygen uptake index (V_{O2}I)	V_{O2}/BSA	115–165 (mL/min)/m^2

Oxygen extraction ratio (O_2ER)	$[1 - (\dot{V}O_2/\dot{D}O_2)] \times 100$	22–32%

Abbreviations: BSA, body surface area; CO, cardiac output; P_{O2}, partial pressure of oxygen; Pa_{O2}, partial pressure of O_2 in arterial blood; Pv_{O2}, partial pressure of O_2 in venous blood; Sa_{O2}, saturation of hemoglobin with O_2 in arterial blood; Sv_{O2}, saturation of hemoglobin with O_2 in venous blood.

Table 270–4. Physiologic Characteristics of the Various Forms of Shock

Type of Shock	CVP and PCWP	Cardiac Output	Systemic Vascular Resistance	Venous O_2 Saturation
Hypovolemic	↓	↓	↑	↓
Cardiogenic	↑	↓	↑	↓
Septic				
Hyperdynamic	↓↑	↑	↓	↑
Hypodynamic	↓↑	↓	↑	↑↓
Traumatic	↓	↓↑	↑↓	↓
Neurogenic	↓	↓	↓	↓
Hypoadrenal	↓↑	↓	=↓	↓

Abbreviations: CVP, central venous pressure; PCWP, pulmonary capillary wedge pressure.

In resuscitation from shock, it is critical to restore tissue perfusion and optimize oxygen delivery, hemodynamics, and cardiac function rapidly. A reasonable goal of therapy is to achieve a normal mixed venous oxygen-saturation and arteriovenous oxygen-extraction ratio. To enhance oxygen delivery, red cell mass, arterial oxygen saturation, and cardiac output may be augmented singly or simultaneously. An increase in oxygen delivery not accompanied by an increase in oxygen consumption implies that oxygen availability is adequate and that oxygen consumption is not flow dependent. Conversely, an elevation of oxygen consumption with increased delivery implies that the oxygen supply was inadequate. However, cautious interpretation is required due to the link among increased oxygen delivery, cardiac work, and oxygen consumption. A reduction in systemic vascular resistance accompanying an increase in cardiac output indicates that compensatory vasoconstriction is reversing due to improved tissue perfusion. The determination of stepwise expansion of blood volume on cardiac performance allows identification of the optimum preload (Starling's law). An algorithm for the resuscitation of the patient in shock is shown in **Fig. 270-3**.

Figure 270-3

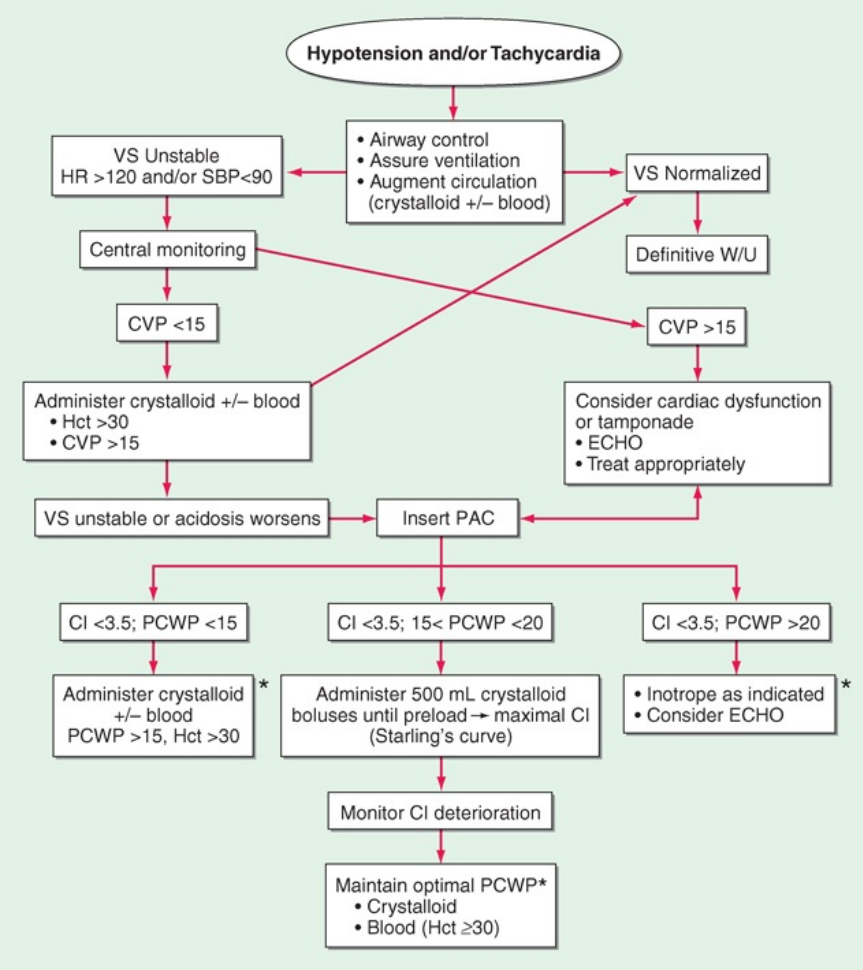

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

An algorithm for the resuscitation of the patient in shock. *Monitor SvO_2, SVRI, and RVEDVI as additional markers of correction for perfusion and hypovolemia. Consider age-adjusted CI. SvO_2, saturation of hemoglobin with O_2 in venous blood; SVRI, systemic vascular resistance index; RVEDVI, right-ventricular end-diastolic volume index. CI, cardiac index in (L/min) per m^2; CVP, central venous pressure; ECHO, echocardiogram; Hct, hematocrit; HR, heart rate; PAC, pulmonary artery catheter; PCWP, pulmonary capillary wedge pressure in mmHg; SBP, systolic blood pressure; VS, vital signs; W/U, work up.

SPECIFIC FORMS OF SHOCK

Hypovolemic Shock

This most common form of shock results either from the loss of red blood cell mass and plasma from hemorrhage or from the loss of plasma volume alone due to extravascular fluid sequestration or GI, urinary, and insensible losses. The signs and symptoms of nonhemorrhagic hypovolemic shock are the same as those of hemorrhagic shock, although they may have a more insidious onset. The normal physiologic response to hypovolemia is to maintain perfusion of the brain and heart while attempting to restore an effective circulating blood volume. There is an increase in sympathetic activity, hyperventilation, collapse of venous capacitance vessels, release of stress hormones, and an attempt to replace the loss of intravascular volume through the recruitment of interstitial and intracellular fluid and by reduction of urine output.

Mild hypovolemia (≤20% of the blood volume) generates mild tachycardia but relatively few external signs, especially in a supine young patient **(Table 270-5)**. With moderate hypovolemia (~20–40% of the blood volume), the patient becomes increasingly anxious and tachycardic; although normal blood pressure may be maintained in the supine position, there may be significant postural hypotension and tachycardia. If hypovolemia is severe (≥40% of the blood volume), the classic signs of shock appear; the blood pressure declines and becomes unstable even in the supine position, and the patient develops marked tachycardia, oliguria, and agitation or confusion. Perfusion of the central nervous system is well maintained until shock becomes severe. Hence, mental obtundation is an ominous clinical sign. The transition from mild to severe hypovolemic shock can be insidious or extremely rapid. If severe shock is not reversed rapidly, especially in elderly patients and those with comorbid illnesses, death is imminent. A very narrow time frame separates the derangements found in severe shock that can be reversed with aggressive resuscitation from those of progressive decompensation and irreversible cell injury.

Table 270–5. Hypovolemic Shock

Mild (<20% Blood Volume)	Moderate (20–40% Blood Volume)	Severe (>40% Blood Volume)
Cool extremities	Same, plus:	Same, plus:

Increased capillary refill time	Tachycardia	Hemodynamic instability
Diaphoresis	Tachypnea	Marked tachycardia
Collapsed veins	Oliguria	Hypotension
Anxiety	Postural changes	Mental status deterioration (coma)

DIAGNOSIS

Hypovolemic shock is readily diagnosed when there are signs of hemodynamic instability and the source of volume loss is obvious. The diagnosis is more difficult when the source of blood loss is occult, as into the GI tract, or when plasma volume alone is depleted. Even after acute hemorrhage, hemoglobin and hematocrit values do not change until compensatory fluid shifts have occurred or exogenous fluid is administered. Thus, an initial normal hematocrit does not disprove the presence of significant blood loss. Plasma losses cause hemoconcentration, and free water loss leads to hypernatremia. These findings should suggest the presence of hypovolemia.

It is essential to distinguish between hypovolemic and cardiogenic shock (Chap. 272) because, while both may respond to volume initially, definitive therapy differs significantly. Both forms are associated with a reduced cardiac output and a compensatory sympathetic mediated response characterized by tachycardia and elevated systemic vascular resistance. However, the findings in cardiogenic shock of jugular venous distention, rales, and an S_3 gallop distinguish it from hypovolemic shock and signify that ongoing volume expansion is undesirable and may cause further organ dysfunction.

TREATMENT: HYPOVOLEMIC SHOCK

Initial resuscitation requires rapid reexpansion of the circulating intravascular blood volume along with interventions to control ongoing losses. In accordance with Starling's law (Chap. 224), stroke volume and cardiac output rise with the increase in preload. After resuscitation, the compliance of the ventricles may remain reduced due to increased interstitial fluid in the myocardium. Therefore, elevated filling pressures are frequently required to maintain adequate ventricular performance.

Volume resuscitation is initiated with the rapid infusion of either isotonic saline (although care must be taken to avoid hyperchloremic acidosis from loss of bicarbonate buffering capacity and replacement with excess chloride) or a balanced salt solution such as Ringer's lactate (being cognizant of the presence of potassium and potential renal dysfunction) through large-bore intravenous lines. Data, particularly on severe traumatic brain injury (TBI), regarding benefits of small volumes of hypertonic saline that more rapidly restore blood pressure are variable, but tend to show improved survival thought to be linked to immunomodulation. No distinct benefit from the use of colloid has been demonstrated, and in trauma patients it is associated with a higher mortality, particularly in patients with TBI. The infusion of 2–3 L of salt solution over 20–30 min should restore normal hemodynamic parameters. Continued hemodynamic instability implies that shock has not been reversed and/or there are significant ongoing blood or other volume losses. Continuing acute blood loss, with hemoglobin concentrations declining to ≤100 g/L (10 g/dL), should initiate blood transfusion, preferably as fully cross-matched recently banked (<14 days old) blood. Resuscitated patients are often coagulopathic due to deficient clotting factors in crystalloids and banked packed red blood cells (PRBCs). Early administration of component therapy during massive transfusion [fresh-frozen plasma (FFP) and platelets] approaching a 1:1 ratio of PRBC/FFP appears to improve survival. In extreme emergencies, type-specific or O-negative packed red cells may be transfused. Following severe and/or prolonged hypovolemia, inotropic support with norepinephrine, vasopressin, or dopamine may be required to maintain adequate ventricular performance *but only after* blood volume has been restored. Increases in peripheral vasoconstriction with inadequate resuscitation leads to tissue loss and organ failure. Once hemorrhage is controlled and the patient has stabilized, blood transfusions should not be continued unless the hemoglobin is <~7g/dL. Studies have demonstrated an increased survival in patients treated with this restrictive blood transfusion protocol.

Successful resuscitation also requires support of respiratory function. Supplemental oxygen should always be provided, and endotracheal intubation may be necessary to maintain arterial oxygenation. Following resuscitation from isolated hemorrhagic shock, end-organ damage is frequently less than following septic or traumatic shock. This may be due to the absence of massive activation of the inflammatory innate immune response and consequent nonspecific organ injury and failure.

Traumatic Shock

Shock following trauma is, in large measure, due to hemorrhage. However, even when hemorrhage has been controlled, patients can continue to suffer loss of plasma volume into the interstitium of injured tissues. These fluid losses are compounded by injury-induced inflammatory responses that which contribute to the secondary microcirculatory injury. Proinflammatory mediators are induced by DAMPs released from injured tissue and are recognized by the highly conserved membrane receptors of the TLR family (see "Inflammatory Responses" above). These receptors on cells of the innate immune system, particularly the circulating monocyte, tissue-fixed macrophage, and dendritic cell, are potent activators of an excessive proinflammatory phenotype in response to cellular injury. This causes secondary tissue injury and maldistribution of blood flow, intensifying tissue ischemia and leading to multiple organ system failure. In addition, direct structural injury to the heart, chest, or head can also contribute to shock. For example, pericardial tamponade or tension pneumothorax impairs ventricular filling, while myocardial contusion depresses myocardial contractility.

TREATMENT: TRAUMATIC SHOCK

Inability of the patient to maintain a systolic blood pressure ≥90 mmHg after trauma-induced hypovolemia is associated with a mortality rate up to –50%. To prevent this decompensation of homeostatic mechanisms, therapy must be promptly administered.

The initial management of the seriously injured patient requires attention to the "ABCs" of resuscitation: assurance of an *airway* (A), adequate ventilation (*breathing*, B), and establishment of an adequate blood volume to support the *circulation* (C). Control of ongoing hemorrhage requires immediate attention. Early stabilization of fractures, debridement of devitalized or contaminated tissues, and evacuation of hematomata all reduce the subsequent

inflammatory response to the initial insult and minimize damaged-tissue release of DAMPs and subsequent diffuse organ injury. Supplementation of depleted endogenous antioxidants also reduces subsequent organ failure and mortality.

Cardiogenic Shock

See Chap. 272.

Compressive Cardiogenic Shock

With extrinsic compression, the heart and surrounding structures are less compliant, and therefore normal filling pressures generate inadequate diastolic filling and stroke volume. Blood or fluid within the poorly distensible pericardial sac may cause tamponade (Chap. 239). Any cause of increased intrathoracic pressure such as tension pneumothorax, herniation of abdominal viscera through a diaphragmatic hernia, or excessive positive-pressure ventilation to support pulmonary function, can also cause compressive cardiogenic shock while simultaneously impeding venous return and preload. Although initially responsive to increased filling pressures produced by volume expansion, as compression increases, cardiogenic shock recurs. The window of opportunity gained by volume loading may be very brief until irreversible shock recurs. Diagnosis and intervention must occur urgently.

The diagnosis of compressive cardiogenic shock is most frequently based on clinical findings, the chest radiograph, and an echocardiogram. The diagnosis of compressive cardiac shock may be more difficult to establish in the setting of trauma when hypovolemia and cardiac compression are present simultaneously. The classic findings of pericardial tamponade include the triad of hypotension, neck vein distention, and muffled heart sounds (Chap. 239). Pulsus paradoxus (i.e., an inspiratory reduction in systolic pressure >10 mmHg), may also be noted. The diagnosis is confirmed by echocardiography, and treatment consists of immediate pericardiocentesis or open subxiphoid pericardial window. A tension pneumothorax produces ipsilateral decreased breath sounds, tracheal deviation away from the affected thorax, and jugular venous distention. Radiographic findings include increased intrathoracic volume, depression of the diaphragm of the affected hemithorax, and shifting of the mediastinum to the contralateral side. Chest decompression must be carried out immediately, and, ideally, should occur based on clinical findings rather than awaiting a chest radiograph. Release of air and restoration of normal cardiovascular dynamics are both diagnostic and therapeutic.

Septic Shock

See Chap. 271.

Neurogenic Shock

Interruption of sympathetic vasomotor input after a high cervical spinal cord injury, inadvertent cephalad migration of spinal anesthesia, or devastating head injury may result in neurogenic shock. In addition to arteriolar dilation, venodilation causes pooling in the venous system, which decreases venous return and cardiac output. The extremities are often warm, in contrast to the usual sympathetic vasoconstriction-induced coolness in hypovolemic or cardiogenic shock. Treatment involves a simultaneous approach to the relative hypovolemia and to the loss of vasomotor tone. Excessive volumes of fluid may be required to restore normal hemodynamics if given alone. Once hemorrhage has been ruled out, norepinephrine or a pure α-adrenergic agent (phenylephrine) may be necessary to augment vascular resistance and maintain an adequate mean arterial pressure.

Hypoadrenal Shock

(See also Chap. 342) The normal host response to the stress of illness, operation, or trauma requires that the adrenal glands hypersecrete cortisol in excess of that normally required. Hypoadrenal shock occurs in settings in which unrecognized adrenal insufficiency complicates the host response to the stress induced by acute illness or major surgery. Adrenocortical insufficiency may occur as a consequence of the chronic administration of high doses of exogenous glucocorticoids. In addition, recent studies have shown that critical illness, including trauma and sepsis, may also induce a relative hypoadrenal state. Other, less common causes include adrenal insufficiency secondary to idiopathicatrophy, use of etomidate for intubation, tuberculosis, metastatic disease, bilateral hemorrhage, and amyloidosis. The shock produced by adrenal insufficiency is characterized by loss of homeostasis with reductions in systemic vascular resistance, hypovolemia, and reduced cardiac output. The diagnosis of adrenal insufficiency may be established by means of an ACTH stimulation test but is inconsistent.

TREATMENT: HYPOADRENAL SHOCK

In the persistently hemodynamically unstable patient, dexamethasone sodium phosphate, 4 mg, should be given intravenously. This agent is preferred if empiric therapy is required because, unlike hydrocortisone, it does not interfere with the ACTH stimulation test. If the diagnosis of absolute or relative adrenal insufficiency is established as shown by nonresponse to corticotropin stimulation (cortisol $\leq$9 μg/dL change poststimulation), the patient has a reduced risk of death if treated with hydrocortisone, 100 mg every 6–8 h, and tapered as the patient achieves hemodynamic stability. Simultaneous volume resuscitation and pressor support are required. The need for simultaneous mineralocoid is unclear.

ADJUNCTIVE THERAPIES

The sympathomimetic amines dobutamine, dopamine, and norepinephrine are widely used in the treatment of all forms of shock. Dobutamine is inotropic with simultaneous afterload reduction, thus minimizing cardiac-oxygen consumption increases as cardiac output increases. Dopamine is an inotropic and chronotropic agent that also supports vascular resistance in those whose blood pressure will not tolerate peripheral vascular dilation. Norepinephrine primarily supports blood pressure through vasoconstriction and increases myocardial oxygen consumption while placing marginally perfused tissues such as extremities and splanchnic organs, at risk for ischemia or necrosis, but it is also inotropic without chronotropy. Arginine-vasopressin (antidiuretic hormone) is being used increasingly to increase afterload and may better protect vital organ blood flow and prevent pathologic vasodilation.

Rewarming

Hypothermia is a frequent adverse consequence of massive volume resuscitation (Chap. 19). The infusion of large

volumes of refrigerated blood products and room temperature crystalloid solutions can rapidly drop core temperatures if fluid is not run through warming devices. Hypothermia may depress cardiac contractility and thereby further impair cardiac output and oxygen delivery/utilization. Hypothermia, particularly temperatures <35°C (<95°F), directly impairs the coagulation pathway, sometimes causing a significant coagulopathy. Rapid rewarming to >35°C (>95°F) significantly decreases the requirement for blood products and produces an improvement in cardiac function. The most effective method for rewarming is endovascular countercurrent warmers through femoral vein cannulation. This process does not require a pump and can rewarm from 30° to 35°C (86° to 95°F) in 30–60 minutes.

FURTHER READINGS

ARDS Clinical Trials Network: Pulmonary-artery versus central venous catheter to guide treatment of acute lung injury. N Engl J Med 354:2213, 2006

Egi M et al: Selecting a vasopressor drug for vasoplegic shock after adult cardiac surgery: A systematic literature review. Ann Thor Surg 83:715, 2007[PMID: 17258030]

Englehart MS et al: Measurement of acid-base resuscitation endpoints: Lactate, base deficit, bicarbonate or what? Curr Opin Crit Care 12:569, 2006[PMID: 17077689]

Gonzales EA et al: Fresh frozen plasma should be given earlier to patients receiving massive transfusion. J Trauma 62:112, 2007

Hebert PC et al: Clinical consequence of anemia and red cell transfusion in the critically ill. Crit Care Clin 20:225, 2004[PMID: 15135462]

Jones AE et al: Goal-directed hemodynamic optimization in the post-cardiac arrest syndrome: A systematic review. Resuscitation 77:26, 2008[PMID: 18164117]

Matsuda N et al: Systemic inflammatory response syndrome (SIRS): Molecular pathophysiology and gene therapy. J Pharmacol Sci 101:189, 2006[PMID: 16823257]

Rivers EP et al: The influence of early hemodynamic optimization on biomarker patterns of severe sepsis and septic shock. Crit Care Med 35:2016, 2007[PMID: 17855815]

Sprung CL et al: Hydrocortisone therapy for patients with septic shock. N Engl J Med 358:111, 2008[PMID: 18184957]

The Safe Study Investigators: Saline or albumin for fluid resuscitation in patients with traumatic brain injury. N Engl J Med 537:874, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SEVERE SEPSIS AND SEPTIC SHOCK: INTRODUCTION

Definitions

(See **Table 271-1**) Animals mount both local and systemic responses to microbes that traverse their epithelial barriers and enter underlying tissues. Fever or hypothermia, leukocytosis or leukopenia, tachypnea, and tachycardia are the cardinal signs of the systemic response, that is often called the *systemic inflammatory response syndrome* (SIRS). SIRS may have an infectious or a noninfectious etiology. If infection is suspected or proven, a patient with SIRS is said to have *sepsis*. When sepsis is associated with dysfunction of organs distant from the site of infection, the patient has *severe sepsis*. Severe sepsis may be accompanied by hypotension or evidence of hypoperfusion. When hypotension cannot be corrected by infusing fluids, the diagnosis is *septic shock*. These definitions were developed by consensus conference committees in 1992 and 2001 and have been widely used; there is evidence that the different stages may form a continuum.

Table 271–1. Definitions Used to Describe the Condition of Septic Patients

Bacteremia	Presence of bacteria in blood, as evidenced by positive blood cultures
Septicemia	Presence of microbes or their toxins in blood
Systemic inflammatory response syndrome (SIRS)	Two or more of the following conditions: (1) fever (oral temperature >38°C) or hypothermia (<36°C); (2) tachypnea (>24 breaths/min); (3) tachycardia (heart rate >90 beats/min); (4) leukocytosis (>12,000/μL), leucopenia (<4,000/μL), or >10% bands; may have a noninfectious etiology
Sepsis	SIRS that has a proven or suspected microbial etiology
Severe sepsis (similar to "sepsis syndrome")	Sepsis with one or more signs of organ dysfunction—for example:
	1. *Cardiovascular:* Arterial systolic blood pressure ≤90 mmHg or mean arterial pressure ≤70 mmHg that responds to administration of intravenous fluid
	2. *Renal:* Urine output <0.5 mL/kg per hour for 1 h despite adequate fluid resuscitation
	3. *Respiratory:* Pa_{O2}/FI_{O2} ≤250 or, if the lung is the only dysfunctional organ, ≤200
	4. *Hematologic:* Platelet count <80,000/μL or 50% decrease in platelet count from highest value recorded over previous 3 days
	5. *Unexplained metabolic acidosis:* A pH ≤7.30 or a base deficit ≥5.0 mEq/L and a plasma lactate level >1.5 times upper limit of normal for reporting lab
	6. *Adequate fluid resuscitation:* Pulmonary artery wedge pressure ≥12 mmHg or central venous pressure ≥8 mmHg
Septic shock	Sepsis with hypotension (arterial blood pressure <90 mmHg systolic, or 40 mmHg less than patient's normal blood pressure) for at least 1 h despite adequate fluid resuscitation;
	or
	Need for vasopressors to maintain systolic blood pressure ≥90 mmHg *or* mean arterial pressure ≥70 mmHg
Refractory septic shock	Septic shock that lasts for >1 h and does not respond to fluid or pressor administration
Multiple-organ dysfunction syndrome (MODS)	Dysfunction of more than one organ, requiring intervention to maintain homeostasis
Predisposition–infection–response–organ dysfunction (PIRO)	A grading system that stratifies patients according to four key aspects of illness; attempts to define subgroups of patients, reducing heterogeneity in clinical trials

Critical illness–related corticosteroid insufficiency (CIRCI)	Inadequate corticosteroid activity for the patient's severity of illness; should be suspected when hypotension is not relieved by fluid administration

Source: Adapted from the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference Committee.

Etiology

Sepsis can be a response to any class of microorganism. Microbial invasion of the bloodstream is not essential, since local inflammation can also elicit distant organ dysfunction and hypotension. In fact, blood cultures yield bacteria or fungi in only –20–40% of cases of severe sepsis and 40–70% of cases of septic shock. Individual gram-negative or gram-positive bacteria account for –70% of these isolates; the remainder are fungi or a mixture of microorganisms **(Table 271-2)**. In patients whose blood cultures are negative, the etiologic agent is often established by culture or microscopic examination of infected material from a local site; specific identification of microbial DNA or RNA in blood or tissue samples is also used. In some case series, a majority of patients with a clinical picture of severe sepsis or septic shock have had negative microbiologic data.

Table 271–2. Microorganisms Involved in Episodes of Severe Sepsis at Eight Academic Medical Centers

Microorganism	Episodes with Bloodstream Infection, % (*n* = 436)	Episodes with Documented Infection but No Bloodstream Infection, % (*n* = 430)	Total Episodes, % (*n* = 866)
Gram-negative bacteria[a]	35	44	40
Gram-positive bacteria[b]	40	24	31
Fungi	7	5	6
Polymicrobial	11	21	16
Classic pathogens[c]	<5	<5	<5

[a]Enterobacteriaceae, pseudomonads, *Haemophilus* spp., other gram-negative bacteria.

[b]*Staphylococcus aureus*, coagulase-negative staphylococci, enterococci, *Streptococcus pneumoniae*, other streptococci, other gram-positive bacteria.

[c]Such as *Neisseria meningitidis*, *S. pneumoniae*, *Haemophilus influenzae*, and *Streptococcus pyogenes*.

Source: Adapted from Sands et al., 1997.

Epidemiology

Severe sepsis is a contributing factor in >200,000 deaths per year in the United States. The incidence of severe sepsis and septic shock has increased over the past 30 years, and the annual number of cases is now >700,000 (~3 per 1000 population). Approximately two-thirds of the cases occur in patients with significant underlying illness. Sepsis-related incidence and mortality rates increase with age and preexisting comorbidity. The rising incidence of severe sepsis in the United States is attributable to the aging of the population, the increasing longevity of patients with chronic diseases, and the relatively high frequency with which sepsis develops in patients with AIDS. The widespread use of immunosuppressive drugs, indwelling catheters, and mechanical devices also plays a role.

Invasive bacterial infections are prominent causes of death around the world, particularly among young children. In sub-Saharan Africa, for example, careful screening for positive blood cultures found that community-acquired bacteremia accounted for at least one-fourth of deaths of children >1 year of age. Nontyphoidal *Salmonella* species, *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Escherichia coli* were the most commonly isolated bacteria. Bacteremic children often had HIV infection or were severely malnourished.

Pathophysiology

Most cases of severe sepsis are triggered by bacteria or fungi that do not ordinarily cause systemic

disease in immunocompetent hosts (Table 271-2). To survive within the human body, these microbes often exploit deficiencies in host defenses, indwelling catheters or other foreign matter, or obstructed fluid drainage conduits. Microbial pathogens, in contrast, can circumvent innate defenses because they (1) lack molecules that can be recognized by host receptors (see below) or (2) elaborate toxins or other virulence factors. In both cases, the body can mount a vigorous inflammatory reaction that results in severe sepsis yet fails to kill the invaders. The septic response may also be induced by microbial exotoxins that act as superantigens (e.g., toxic shock syndrome toxin 1; Chap. 135) as well as by many pathogenic viruses.

HOST MECHANISMS FOR SENSING MICROBES

Animals have exquisitely sensitive mechanisms for recognizing and responding to certain highly conserved microbial molecules. Recognition of the lipid A moiety of lipopolysaccharide (LPS, also called *endotoxin*; Chap. 120) is the best-studied example. A host protein (LPS-binding protein) binds lipid A and transfers the LPS to CD14 on the surfaces of monocytes, macrophages, and neutrophils. LPS then is passed to MD-2, that is bound to toll-like receptor (TLR) 4 to form a molecular complex that transduces the LPS recognition signal to the interior of the cell. This signal rapidly triggers the production and release of mediators, such as tumor necrosis factor (TNF; see below), that amplify the LPS signal and transmit it to other cells and tissues. Bacterial peptidoglycan and lipopeptides elicit responses in animals that are generally similar to those induced by LPS; whereas these molecules also may be transferred by CD14, they interact with different TLRs. Having numerous TLR-based receptor complexes (11 different TLRs have been identified so far in humans) allows animals to recognize many conserved microbial molecules; others include lipopeptides (TLR2/1, TLR2/6), flagellin (TLR5), undermethylated DNA sequences (TLR9), and double-stranded RNA (TLR3, TLR7). The ability of some TLRs to serve as receptors for host ligands (e.g., hyaluronans, heparan sulfate, saturated fatty acids) raises the possibility that they also play a role in producing noninfectious sepsis-like states. Other host pattern-recognition proteins that are important for sensing microbial invasion include the intracellular NOD1 and NOD2 proteins, which recognize discrete fragments of bacterial peptidoglycan; early complement components (principally in the alternative pathway); and mannose-binding lectin and C-reactive protein, which activate the classic complement pathway.

A host's ability to recognize certain microbial molecules may influence both the potency of its own defenses and the pathogenesis of severe sepsis. For example, MD-2–TLR4 best senses LPS that has a hexaacyl lipid A moiety (i.e., one with six fatty acyl chains). Most of the commensal aerobic and facultatively anaerobic gram-negative bacteria that trigger severe sepsis and shock (including *E. coli*, *Klebsiella*, and *Enterobacter*) make this lipid A structure. When they invade human hosts, often through breaks in an epithelial barrier, they are typically confined to the subepithelial tissue by a localized inflammatory response. Bacteremia, if it occurs, is intermittent and low-grade, as these bacteria are efficiently cleared from the bloodstream by TLR4-expressing Kupffer cells and splenic macrophages. These mucosal commensals seem to induce severe sepsis most often by triggering severe local tissue inflammation rather than by circulating within the bloodstream. One exception is *Neisseria meningitidis*. Its hexaacyl LPS seems to be shielded from host recognition by its polysaccharide capsule. This protection may allow meningococci to transit undetected from the nasopharyngeal mucosa into the bloodstream, where they can infect vascular endothelial cells and release large amounts of endotoxin. Host recognition of lipid A may nonetheless influence pathogenesis, as meningococci that produce pentaacyl LPS were isolated from the blood of patients with less severe coagulopathy than was found in patients whose isolates produced hexaacyl lipid A. In contrast, gram-negative bacteria that make lipid A with fewer than six acyl chains (*Yersinia pestis*, *Francisella tularensis*, *Vibrio vulnificus*, *Pseudomonas aeruginosa*, and *Burkholderia pseudomallei*, among others) are poorly recognized by MD-2–TLR4. When these bacteria enter the body, they may initially induce relatively little inflammation. When they do trigger severe sepsis, it is often after they have multiplied to high density in tissues and blood. The importance of LPS recognition in disease pathogenesis has been shown by engineering a virulent strain of *Y. pestis*, which makes tetraacyl LPS at 37°C, to produce hexaacyl LPS; unlike its virulent parent, the mutant strain stimulates local inflammation and is rapidly cleared from tissues. For at least one large class of microbes—gram-negative aerobic bacteria—the pathogenesis of sepsis thus depends, at least in part, upon whether the bacterium's major signal molecule, LPS, can be sensed by the host.

LOCAL AND SYSTEMIC HOST RESPONSES TO INVADING MICROBES

Recognition of microbial molecules by tissue phagocytes triggers the production and/or release of numerous host molecules (cytokines, chemokines, prostanoids, leukotrienes, and others) that increase blood flow to the infected tissue, enhance the permeability of local blood vessels, recruit neutrophils to the site of infection, and elicit pain. These reactions are familiar elements of local inflammation, the body's frontline innate immune mechanism for eliminating microbial invaders. Systemic responses are activated by neural and/or humoral communication with the hypothalamus and brainstem; these responses enhance local defenses by increasing blood flow to the infected area, augmenting the number of circulating neutrophils, and elevating blood levels of numerous molecules (such as the microbial recognition proteins discussed above) that have anti-infective functions.

Cytokines and Other Mediators

Cytokines can exert endocrine, paracrine, and autocrine effects (Chap. 314). TNF-α stimulates leukocytes and vascular endothelial cells to release other cytokines (as well as additional TNF-α), to express cell-surface molecules that enhance neutrophil-endothelial adhesion at sites of infection, and to increase prostaglandin and leukotriene production. Whereas blood levels of TNF-α are not elevated in individuals with localized infections, they increase in most patients with severe sepsis or septic shock. Moreover, IV infusion of TNF-α can elicit the characteristic abnormalities of SIRS. In animals, larger doses of TNF-α induce shock and death.

Although TNF-α is a central mediator, it is only one of many proinflammatory molecules that contribute to innate host defense. Chemokines, most prominently interleukin (IL)-8 and IL-17, attract circulating neutrophils to the infection site. IL-1β exhibits many of the same activities as TNF-α. TNF-α, IL-1β, interferon (IFN) γ, IL-12, IL-17, and other proinflammatory cytokines probably interact synergistically with one another and with additional mediators. The nonlinearity and multiplicity of these interactions have made it difficult to interpret the roles played by individual mediators in both tissues and blood.

Coagulation Factors

Intravascular thrombosis, a hallmark of the local inflammatory response, may help wall off invading microbes and prevent infection and inflammation from spreading to other tissues. IL-6 and other mediators promote intravascular coagulation initially by inducing blood monocytes and vascular endothelial cells to express tissue factor (Chap. 58). When tissue factor is expressed on cell surfaces, it binds to factor VIIa to form an active complex that can convert factors X and IX to their enzymatically active forms. The result is activation of both extrinsic and intrinsic clotting pathways, culminating in the generation of fibrin. Clotting is also favored by impaired function of the protein C–protein S inhibitory pathway and depletion of antithrombin and proteins C and S, while fibrinolysis is prevented by increased plasma levels of plasminogen activator inhibitor 1. Thus, there may be a striking propensity toward intravascular fibrin deposition, thrombosis, and bleeding; this propensity has been most apparent in patients with intravascular endothelial infections such as meningococcemia (Chap. 143). Evidence points to tissue factor–expressing microparticles derived from leukocytes as a potential trigger for intravascular coagulation. Contact-system activation occurs during sepsis but contributes more to the development of hypotension than to disseminated intravascular coagulation (DIC).

Control Mechanisms

Elaborate control mechanisms operate within both local sites of inflammation and the systemic compartment.

Local Control Mechanisms

Host recognition of invading microbes within subepithelial tissues typically ignites immune responses that rapidly kill the invader and then subside to allow tissue recovery. The anti-inflammatory forces that put out the fire and clean up the battleground include molecules that neutralize or inactivate microbial signals. Among these molecules are intracellular factors (e.g., suppressor of cytokine signaling 3 and IL-1 receptor–associated kinase 3) that diminish the production of proinflammatory mediators by neutrophils and macrophages; anti-inflammatory cytokines (IL-10, IL-4); and molecules derived from essential polyunsaturated fatty acids (lipoxins, resolvins, and protectins) that promote tissue restoration. Enzymatic inactivation of microbial signal molecules (e.g., LPS) may be required to restore homeostasis; a leukocyte enzyme, acyloxyacyl hydrolase, has been shown to prevent prolonged inflammation by inactivating LPS in mice.

Systemic Control Mechanisms

The signaling apparatus that links microbial recognition to cellular responses in tissues is less active in the blood. For example, whereas LPS-binding protein plays a role in recognizing the presence of LPS, in plasma it also prevents LPS signaling by transferring LPS molecules into plasma lipoprotein particles that sequester the lipid A moiety so that it cannot interact with cells. At the high concentrations found in blood, LPS-binding protein also inhibits monocyte responses to LPS, and the soluble (circulating) form of CD14 strips off LPS that has bound to monocyte surfaces.

Systemic responses to infection also diminish cellular responses to microbial molecules. Circulating levels of anti-inflammatory cytokines (e.g., IL-10) increase even in patients with mild infections. Glucocorticoids inhibit cytokine synthesis by monocytes in vitro; the increase in blood cortisol levels early in the systemic response presumably plays a similarly inhibitory role. Epinephrine inhibits the TNF-α response to endotoxin infusion in humans while augmenting and accelerating the release of IL-10; prostaglandin E2 has a similar "reprogramming" effect on the responses of circulating monocytes to LPS and other bacterial agonists. Cortisol, epinephrine, IL-10, and C-reactive protein reduce the ability of neutrophils to attach to vascular endothelium, favoring their demargination and thus contributing to leukocytosis while preventing neutrophil-endothelial adhesion in uninflamed organs. The available evidence thus suggests that the body's systemic responses to injury and infection normally prevent inflammation within organs distant from a site of infection. There is also evidence that these responses may be immunosuppressive.

The acute-phase response increases the blood concentrations of numerous molecules that have anti-inflammatory actions. Blood levels of IL-1 receptor antagonist often greatly exceed those of circulating IL-1β, for example, and this excess may inhibit the binding of IL-1β to its receptors. High levels of soluble TNF receptors neutralize TNF-α that enters the circulation. Other acute-phase proteins are protease inhibitors or antioxidants; these may neutralize potentially harmful molecules released from neutrophils and other inflammatory cells. Increased hepatic production of hepcidin promotes the sequestration of iron in hepatocytes, intestinal epithelial cells, and erythrocytes; this effect reduces iron acquisition by invading microbes while contributing to the normocytic, normochromic anemia associated with inflammation.

It can thus be concluded that both local and systemic responses to infectious agents benefit the host in important ways. Most of these responses and the molecules responsible for them have been highly conserved during animal evolution and therefore may be adaptive. Elucidating how they contribute to lethality—i.e., become maladaptive—remains a major challenge for sepsis research.

ORGAN DYSFUNCTION AND SHOCK

As the body's responses to infection intensify, the mixture of circulating cytokines and other molecules becomes very complex: elevated blood levels of more than 50 molecules have been found in patients with septic shock. Although high concentrations of both pro- and anti-inflammatory molecules are found, the net mediator balance in the plasma of these extremely sick patients seems to be anti-inflammatory. For example, blood leukocytes from patients with severe sepsis are often hyporesponsive to agonists such as LPS. In patients with severe sepsis, persistence of leukocyte hyporesponsiveness has been associated with an increased risk of dying. Apoptotic death of B cells, follicular dendritic cells, and CD4+ T lymphocytes also may contribute significantly to the immunosuppressive state.

Endothelial Injury

Many investigators have favored widespread vascular endothelial injury as the major mechanism for multiorgan dysfunction. In keeping with this idea, one study found high numbers of vascular endothelial cells in the peripheral blood of septic patients. Leukocyte-derived mediators and platelet-leukocyte-fibrinthrombi may contribute to vascular injury, but the vascular endothelium also seems to play an active role. Stimuli such as TNF-α induce vascular endothelial cells to produce and release cytokines, procoagulant molecules, platelet-activating factor, nitric oxide, and other mediators. In addition, regulated cell-adhesion molecules promote the adherence of neutrophils to endothelial cells. While these responses can attract phagocytes to infected sites and activate their antimicrobial arsenals, endothelial cell activation can also promote increased vascular permeability, microvascular thrombosis, DIC, and hypotension.

Tissue oxygenation may decrease as the number of functional capillaries is reduced by luminal obstruction due to swollen endothelial cells, decreased deformability of circulating erythrocytes, leukocyte-platelet-fibrinthrombi, or compression by edema fluid. On the other hand, studies using orthogonal polarization spectral imaging of the microcirculation in the tongue found that sepsis-associated derangements in capillary flow could be reversed by applying acetylcholine to the surface of the tongue or by giving nitroprusside intravenously; these observations suggest a neuroendocrine basis for the loss of capillary filling. Oxygen utilization by tissues may also be impaired by a state of "hibernation" in which ATP production is diminished as oxidative phosphorylation decreases; nitric oxide may be responsible for inducing this response.

Remarkably, poorly functioning "septic" organs usually appear normal at autopsy. There is typically very little necrosis or thrombosis, and apoptosis is largely confined to lymphoid organs and the gastrointestinal tract. Moreover, organ function usually returns to normal if patients recover. These points suggest that organ dysfunction during severe sepsis has a basis that is principally biochemical, not structural.

Septic Shock

The hallmark of septic shock is a decrease in peripheral vascular resistance that occurs despite increased levels of vasopressor catecholamines. Before this vasodilatory phase, many patients experience a period during which oxygen delivery to tissues is compromised by myocardial depression, hypovolemia, and other factors. During this "hypodynamic" period, the blood lactate concentration is elevated and central venous oxygen saturation is low. Fluid administration is usually followed by the hyperdynamic, vasodilatory phase during which cardiac output is normal (or even high) and oxygen consumption declines despite adequate oxygen delivery. The blood lactate level may be normal or increased, and normalization of central venous oxygen saturation may reflect either improved oxygen delivery or left-to-right shunting.

Prominent hypotensive molecules include nitric oxide, β-endorphin, bradykinin, platelet-activating factor, and prostacyclin. Agents that inhibit the synthesis or action of each of these mediators can prevent or reverse endotoxic shock in animals. However, in clinical trials, neither a platelet-activating factor receptor antagonist nor a bradykinin antagonist improved survival rates among patients with septic shock, and a nitric oxide synthase inhibitor, L-N^G-methylarginine HCl, actually increased the

mortality rate. Remarkably, recent findings indicate that exogenous nitrite can protect mice from challenge with TNF or LPS. Nitrite provides a storage pool from which nitric oxide can be generated in hypoxic and/or acidic conditions. These findings should renew interest in the possibility of exploiting nitric oxide metabolism to improve survival rates among septic patients.

SEVERE SEPSIS: A SINGLE PATHOGENESIS?

In some cases, circulating bacteria and their products almost certainly elicit multiorgan dysfunction and hypotension by directly stimulating inflammatory responses within the vasculature. In patients with fulminant meningococcemia, for example, mortality rates have correlated directly with blood levels of endotoxin and bacterial DNA and with the occurrence of DIC (Chap. 143). In most patients infected with other gram-negative bacteria, in contrast, circulating bacteria or bacterial molecules may reflect uncontrolled infection at a local tissue site and have little or no direct impact on distant organs; in these patients, inflammatory mediators or neural signals arising from the local site seem to be the key triggers for severe sepsis and septic shock. In a large series of patients with positive blood cultures, the risk of developing severe sepsis was strongly related to the site of primary infection: bacteremia arising from a pulmonary or abdominal source was eightfold more likely to be associated with severe sepsis than was bacteremic urinary tract infection, even after the investigators controlled for age, the kind of bacteria isolated from the blood, and other factors. A third pathogenesis may be represented by severe sepsis due to superantigen-producing *Staphylococcus aureus* or *Streptococcus pyogenes*; the T cell activation induced by these toxins produces a cytokine profile that differs substantially from that elicited by gram-negative bacterial infection. Further evidence for different pathogenetic pathways has come from observations that the pattern of mRNA expression in peripheral-blood leukocytes from children with sepsis is different for gram-positive, gram-negative, and viral pathogens.

The pathogenesis of severe sepsis thus may differ according to the infecting microbe, the ability of the host's innate defense mechanisms to sense it, the site of the primary infection, the presence or absence of immune defects, and the prior physiologic status of the host. Genetic factors are probably important as well, yet despite much study only a few allelic polymorphisms (e.g., in the IL-1β gene) have been associated with sepsis severity in more than one or two analyses. Further studies in this area are needed.

Clinical Manifestations

The manifestations of the septic response are superimposed on the symptoms and signs of the patient's underlying illness and primary infection. The rate at which severe sepsis develops may differ from patient to patient, and there are striking individual variations in presentation. For example, some patients with sepsis are normo- or hypothermic; the absence of fever is most common in neonates, in elderly patients, and in persons with uremia or alcoholism.

Hyperventilation is often an early sign of the septic response. Disorientation, confusion, and other manifestations of encephalopathy may also develop early on, particularly in the elderly and in individuals with preexisting neurologic impairment. Focal neurologic signs are uncommon, although preexisting focal deficits may become more prominent.

Hypotension and DIC predispose to acrocyanosis and ischemic necrosis of peripheral tissues, most commonly the digits. Cellulitis, pustules, bullae, or hemorrhagic lesions may develop when hematogenous bacteria or fungi seed the skin or underlying soft tissue. Bacterial toxins may also be distributed hematogenously and elicit diffuse cutaneous reactions. On occasion, skin lesions may suggest specific pathogens. When sepsis is accompanied by cutaneous petechiae or purpura, infection with *N. meningitidis* (or, less commonly, *H. influenzae*) should be suspected (Fig. e7-42); in a patient who has been bitten by a tick while in an endemic area, petechial lesions also suggest Rocky Mountain spotted fever (Fig. 174-1). A cutaneous lesion seen almost exclusively in neutropenic patients is ecthyma gangrenosum, usually caused by *P. aeruginosa*. It is a bullous lesion, surrounded by edema, that undergoes central hemorrhage and necrosis (Fig. 152-1). Histopathologic examination shows bacteria in and around the wall of a small vessel, with little or no neutrophilic response. Hemorrhagic or bullous lesions in a septic patient who has recently eaten raw oysters suggest *V. vulnificus* bacteremia, while such lesions in a patient who has recently suffered a dog bite may indicate bloodstream infection due to *Capnocytophaga canimorsus* or *C. cynodegmi*. Generalized erythroderma in a septic patient suggests the toxic shock syndrome due to *S. aureus* or *S. pyogenes*.

Gastrointestinal manifestations such as nausea, vomiting, diarrhea, and ileus may suggest acute gastroenteritis. Stress ulceration can lead to upper gastrointestinal bleeding. Cholestatic jaundice, with elevated levels of serum bilirubin (mostly conjugated) and alkaline phosphatase, may precede other signs of sepsis. Hepatocellular or canalicular dysfunction appears to underlie most cases, and the results of hepatic function tests return to normal with resolution of the infection. Prolonged or severe hypotension may induce acute hepatic injury or ischemic bowel necrosis.

Many tissues may be unable to extract oxygen normally from the blood, so that anaerobic metabolism occurs despite near-normal mixed venous oxygen saturation. Blood lactate levels rise early because of increased glycolysis as well as impaired clearance of the resulting lactate and pyruvate by the liver and kidneys. The blood glucose concentration often increases, particularly in patients with diabetes,

although impaired gluconeogenesis and excessive insulin release on occasion produce hypoglycemia. The cytokine-driven acute-phase response inhibits the synthesis of transthyretin while enhancing the production of C-reactive protein, fibrinogen, and complement components. Protein catabolism is often markedly accelerated. Serum albumin levels decline as a result of decreased hepatic synthesis and the movement of albumin into interstitial spaces.

Major Complications

CARDIOPULMONARY COMPLICATIONS

Ventilation-perfusion mismatching produces a fall in arterial P_{O2} early in the course. Increasing alveolar epithelial injury and capillary permeability result in increased pulmonary water content, which decreases pulmonary compliance and interferes with oxygen exchange. In the absence of pneumonia or heart failure, progressive diffuse pulmonary infiltrates and arterial hypoxemia (Pa_{O2}/FI_{O2}, <300) indicate the development of acute lung injury; more severe hypoxemia (Pa_{O2}/FI_{O2}, <200) denotes the acute respiratory distress syndrome (ARDS). Acute lung injury or ARDS develops in ~50% of patients with severe sepsis or septic shock. Respiratory muscle fatigue can exacerbate hypoxemia and hypercapnia. An elevated pulmonary capillary wedge pressure (>18 mmHg) suggests fluid volume overload or cardiac failure rather than ARDS. Pneumonia caused by viruses or by *Pneumocystis* may be clinically indistinguishable from ARDS.

Sepsis-induced hypotension (see "Septic Shock," above) usually results initially from a generalized maldistribution of blood flow and blood volume and from hypovolemia that is due, at least in part, to diffuse capillary leakage of intravascular fluid. Other factors that may decrease effective intravascular volume include dehydration from antecedent disease or insensible fluid losses, vomiting or diarrhea, and polyuria. During early septic shock, systemic vascular resistance is usually elevated and cardiac output may be low. After fluid repletion, in contrast, cardiac output typically increases and systemic vascular resistance falls. Indeed, normal or increased cardiac output and decreased systemic vascular resistance distinguish septic shock from cardiogenic, extracardiac obstructive, and hypovolemic shock; other processes that can produce this combination include anaphylaxis, beriberi, cirrhosis, and overdoses of nitroprusside or narcotics.

Depression of myocardial function, manifested as increased end-diastolic and systolic ventricular volumes with a decreased ejection fraction, develops within 24 h in most patients with severe sepsis. Cardiac output is maintained despite the low ejection fraction because ventricular dilatation permits a normal stroke volume. In survivors, myocardial function returns to normal over several days. Although myocardial dysfunction may contribute to hypotension, refractory hypotension is usually due to low systemic vascular resistance, and death results from refractory shock or the failure of multiple organs rather than from cardiac dysfunction per se.

ADRENAL INSUFFICIENCY

The diagnosis of adrenal insufficiency may be very difficult in critically ill patients. Whereas a plasma cortisol level of ≤15 μg/mL (≤10 μg/mL if the serum albumin concentration is <2.5 mg/dL) indicates adrenal insufficiency (inadequate production of cortisol), many experts now feel that the ACTH (CoSyntropin®) stimulation test is not useful for detecting less profound degrees of corticosteroid deficiency in patients who are critically ill. The concept of critical illness–related corticosteroid insufficiency (CIRCI; Table 271-1) was proposed to encompass the different mechanisms that may produce corticosteroid activity that is inadequate for the severity of a patient's illness. Although CIRCI may result from structural damage to the adrenal gland, it is more commonly due to reversible dysfunction of the hypothalamic-pituitary axis or to tissue corticosteroid resistance resulting from abnormalities of the glucocorticoid receptor or increased conversion of cortisol to cortisone. The major clinical manifestation of CIRCI is hypotension that is refractory to fluid replacement and requires pressor therapy. Some classic features of adrenal insufficiency, such as hyponatremia and hyperkalemia, are usually absent; others, such as eosinophilia and modest hypoglycemia, may sometimes be found. Specific etiologies include fulminant *N. meningitidis* bacteremia, disseminated tuberculosis, AIDS (with cytomegalovirus, *Mycobacterium avium-intracellulare*, or *Histoplasma capsulatum* disease), or the prior use of drugs that diminish glucocorticoid production, such as glucocorticoids, megestrol, etomidate, or ketoconazole.

RENAL COMPLICATIONS

Oliguria, azotemia, proteinuria, and nonspecific urinary casts are frequently found. Many patients are inappropriately polyuric; hyperglycemia may exacerbate this tendency. Most renal failure is due to acute tubular necrosis induced by hypotension or capillary injury, although some patients also have glomerulonephritis, renal cortical necrosis, or interstitial nephritis. Drug-induced renal damage may complicate therapy, particularly when hypotensive patients are given aminoglycoside antibiotics.

COAGULOPATHY

Although thrombocytopenia occurs in 10–30% of patients, the underlying mechanisms are not understood. Platelet counts are usually very low (<50,000/μL) in patients with DIC; these low counts

may reflect diffuse endothelial injury or microvascular thrombosis, yet thrombi have only infrequently been found upon biopsy of septic organs.

NEUROLOGIC COMPLICATIONS

When the septic illness lasts for weeks or months, "critical illness" polyneuropathy may prevent weaning from ventilatory support and produce distal motor weakness. Electrophysiologic studies are diagnostic. Guillain-Barré syndrome, metabolic disturbances, and toxin activity must be ruled out.

Immunosuppression

Patients with severe sepsis are often profoundly immunosuppressed. Manifestations include loss of delayed-type hypersensitivity reactions to common antigens, failure to control the primary infection, and increased risk for secondary infections (e.g., by opportunists such as *Stenotrophomonas maltophilia*, *Acinetobacter calcoaceticus-baumannii*, and *Candida albicans*). Approximately one-third of patients experience reactivation of herpes simplex virus, varicella-zoster virus, or cytomegalovirus infections; the latter are thought to contribute to adverse outcomes in some instances.

Laboratory Findings

Abnormalities that occur early in the septic response may include leukocytosis with a left shift, thrombocytopenia, hyperbilirubinemia, and proteinuria. Leukopenia may develop. The neutrophils may contain toxic granulations, Döhle bodies, or cytoplasmic vacuoles. As the septic response becomes more severe, thrombocytopenia worsens (often with prolongation of the thrombin time, decreased fibrinogen, and the presence of d-dimers, suggesting DIC), azotemia and hyperbilirubinemia become more prominent, and levels of aminotransferases rise. Active hemolysis suggests clostridial bacteremia, malaria, a drug reaction, or DIC; in the case of DIC, microangiopathic changes may be seen on a blood smear.

During early sepsis, hyperventilation induces respiratory alkalosis. With respiratory muscle fatigue and the accumulation of lactate, metabolic acidosis (with increased anion gap) typically supervenes. Evaluation of arterial blood gases reveals hypoxemia that is initially correctable with supplemental oxygen but whose later refractoriness to 100% oxygen inhalation indicates right-to-left shunting. The chest radiograph may be normal or may show evidence of underlying pneumonia, volume overload, or the diffuse infiltrates of ARDS. The electrocardiogram may show only sinus tachycardia or nonspecific ST–T-wave abnormalities.

Most diabetic patients with sepsis develop hyperglycemia. Severe infection may precipitate diabetic ketoacidosis that may exacerbate hypotension (Chap. 344). Hypoglycemia occurs rarely. The serum albumin level declines as sepsis continues. Hypocalcemia is rare.

Diagnosis

There is no specific diagnostic test for the septic response. Diagnostically sensitive findings in a patient with suspected or proven infection include fever or hypothermia, tachypnea, tachycardia, and leukocytosis or leukopenia (Table 271-1); acutely altered mental status, thrombocytopenia, an elevated blood lactate level, or hypotension also should suggest the diagnosis. The septic response can be quite variable, however. In one study, 36% of patients with severe sepsis had a normal temperature, 40% had a normal respiratory rate, 10% had a normal pulse rate, and 33% had normal white blood cell counts. Moreover, the systemic responses of uninfected patients with other conditions may be similar to those characteristic of sepsis. Noninfectious etiologies of SIRS (Table 271-1) include pancreatitis, burns, trauma, adrenal insufficiency, pulmonary embolism, dissecting or ruptured aortic aneurysm, myocardial infarction, occult hemorrhage, cardiac tamponade, postcardiopulmonary bypass syndrome, anaphylaxis, tumor-associated lactic acidosis, and drug overdose.

Definitive etiologic diagnosis requires isolation of the microorganism from blood or a local site of infection. At least two blood samples should be obtained (from two different venipuncture sites) for culture; in a patient with an indwelling catheter, one sample should be collected from each lumen of the catheter and another via venipuncture. In many cases, blood cultures are negative; this result can reflect prior antibiotic administration, the presence of slow-growing or fastidious organisms, or the absence of microbial invasion of the bloodstream. In these cases, Gram's staining and culture of material from the primary site of infection or from infected cutaneous lesions may help establish the microbial etiology. Identification of microbial DNA in peripheral-blood or tissue samples by polymerase chain reaction may also be definitive. The skin and mucosae should be examined carefully and repeatedly for lesions that might yield diagnostic information. With overwhelming bacteremia (e.g., pneumococcal sepsis in splenectomized individuals; fulminant meningococcemia; or infection with *V. vulnificus*, *B. pseudomallei*, or *Y. pestis*), microorganisms are sometimes visible on buffy coat smears of peripheral blood.

Treatment: Severe Sepsis and Septic Shock

Patients in whom sepsis is suspected must be managed expeditiously. This task is best accomplished by personnel who are experienced in the care of the critically ill. Successful management requires urgent measures to treat the infection, to provide hemodynamic and respiratory support, and to

eliminate the offending microorganisms. These measures should be initiated within 1 h of the patient's presentation with severe sepsis or septic shock. Rapid assessment and diagnosis are therefore essential.

ANTIMICROBIAL AGENTS

Antimicrobial chemotherapy should be started as soon as samples of blood and other relevant sites have been obtained for culture. A large retrospective review of patients who developed septic shock found that the interval between the onset of hypotension and the administration of appropriate antimicrobial chemotherapy was the major determinant of outcome; a delay of as little as 1 h was associated with lower survival rates. Use of inappropriate antibiotics, defined on the basis of local microbial susceptibilities and published guidelines for empirical therapy (see below), was associated with fivefold lower survival rates, even among patients with negative cultures.

It is therefore very important to promptly initiate empirical antimicrobial therapy that is effective against both gram-positive and gram-negative bacteria **(Table 271-3)**. Maximal recommended doses of antimicrobial drugs should be given intravenously, with adjustment for impaired renal function when necessary. Available information about patterns of antimicrobial susceptibility among bacterial isolates from the community, the hospital, and the patient should be taken into account. When culture results become available, the regimen can often be simplified, as a single antimicrobial agent is usually adequate for the treatment of a known pathogen. Meta-analyses have concluded that, with one exception, combination antimicrobial therapy is not superior to monotherapy for treating gram-negative bacteremia; the exception is that aminoglycoside monotherapy for *P. aeruginosa* bacteremia is less effective than the combination of an aminoglycoside with an antipseudomonal β-lactam agent. Empirical antifungal therapy should be strongly considered if the septic patient is already receiving broad-spectrum antibiotics or parenteral nutrition, has been neutropenic for ≥5 days, has had a long-term central venous catheter, or has been hospitalized in an intensive care unit for a prolonged period. The chosen antimicrobial regimen should be reconsidered daily in order to provide maximal efficacy with minimal resistance, toxicity, and cost.

Table 271–3. Initial Antimicrobial Therapy for Severe Sepsis with No Obvious Source in Adults with Normal Renal Function

Clinical Condition	Antimicrobial Regimens (Intravenous Therapy)
Immunocompetent adult	The many acceptable regimens include (1) piperacillin-tazobactam (3.375 g q4–6h); (2) imipenem-cilastatin (0.5 g q6h) or meropenem (1 g q8h); or (3) cefepime (2 g q12h). If the patient is allergic to β-lactam agents, use ciprofloxacin (400 mg q12h) or levofloxacin (500–750 mg q12h) plus clindamycin (600 mg q8h). Vancomycin (15 mg/kg q12h) should be added to each of the above regimens.
Neutropenia (<500 neutrophils/μL)	Regimens include (1) imipenem-cilastatin (0.5 g q6h) or meropenem (1 g q8h) or cefepime (2 g q8h); (2) piperacillintazobactam (3.375 g q4h) plus tobramycin (5–7 mg/kg q24h). Vancomycin (15 mg/kg q12h) should be added if the patient has an indwelling vascular catheter, has received quinolone prophylaxis, or has received intensive chemotherapy that produces mucosal damage; if staphylococci are suspected; if the institution has a high incidence of MRSA infections; or if there is a high prevalence of MRSA isolates in the community. Empirical antifungal therapy with an echinocandin (for caspofungin: a 70-mg loading dose, then 50 mg daily) or a lipid formulation of amphotericin B should be added if the patient is hypotensive or has been receiving broad-spectrum antibacterial drugs.
Splenectomy	Cefotaxime (2 g q6–8h) or ceftriaxone (2 g q12h) should be used. If the local prevalence of cephalosporin-resistant pneumococci is high, add vancomycin. If the patient is allergic to β-lactam drugs, vancomycin (15 mg/kg q12h) plus either moxifloxacin (400 mg q24h) or levofloxacin (750 mg q24h) or aztreonam (2 g q8h) should be used.
IV drug user	Vancomycin (15 mg/kg q12h)
AIDS	Cefepime (2 g q8h) or piperacillin-tazobactam (3.375 g q4h) plus tobramycin (5–7 mg/kg q24h) should be used. If the patient is allergic to β-lactam drugs, ciprofloxacin (400 mg q12h) or levofloxacin (750 mg q12h) plus vancomycin (15 mg/kg q12h) plus tobramycin should be used.

Abbreviation: MRSA, methicillin-resistant *Staphylococcus aureus*.

Source: Adapted in part from WT Hughes et al: Clin Infect Dis 25:551, 1997; and DN Gilbert et al: The Sanford Guide to Antimicrobial Therapy, 2009.

Most patients require antimicrobial therapy for at least 1 week. The duration of treatment is typically influenced by factors such as the site of tissue infection, the adequacy of surgical drainage, the patient's underlying disease, and the antimicrobial susceptibility of the microbial isolate(s). The absence of an identified microbial pathogen is not necessarily an indication for discontinuing

antimicrobial therapy, since "appropriate" antimicrobial regimens seem to be beneficial in both culture-negative and culture-positive cases.

REMOVAL OF THE SOURCE OF INFECTION

Removal or drainage of a focal source of infection is essential. In one series, a focus of ongoing infection was found in ~80% of surgical intensive care patients who died of severe sepsis or septic shock. Sites of occult infection should be sought carefully, particularly in the lungs, abdomen, and urinary tract. Indwelling IV or arterial catheters should be removed and the tip rolled over a blood agar plate for quantitative culture; after antibiotic therapy has been initiated, a new catheter should be inserted at a different site. Foley and drainage catheters should be replaced. The possibility of paranasal sinusitis (often caused by gram-negative bacteria) should be considered if the patient has undergone nasal intubation. Even in patients without abnormalities on chest radiographs, CT of the chest may identify unsuspected parenchymal, mediastinal, or pleural disease. In the neutropenic patient, cutaneous sites of tenderness and erythema, particularly in the perianal region, must be carefully sought. In patients with sacral or ischial decubitus ulcers, it is important to exclude pelvic or other soft tissue pus collections with CT or MRI. In patients with severe sepsis arising from the urinary tract, sonography or CT should be used to rule out ureteral obstruction, perinephric abscess, and renal abscess. Sonographic or CT imaging of the upper abdomen may disclose evidence of cholecystitis, bile duct dilatation, and pus collections in the liver, subphrenic space, or spleen.

HEMODYNAMIC, RESPIRATORY, AND METABOLIC SUPPORT

The primary goals are to restore adequate oxygen and substrate delivery to the tissues as quickly as possible and to improve tissue oxygen utilization and cellular metabolism. Adequate organ perfusion is thus essential. Circulatory adequacy is assessed by measurement of arterial blood pressure and monitoring of parameters such as mentation, urine output, and skin perfusion. Indirect indices of oxygen delivery and consumption, such as central venous oxygen saturation, may also be useful. Initial management of hypotension should include the administration of IV fluids, typically beginning with 1–2 L of normal saline over 1–2 h. To avoid pulmonary edema, the central venous pressure should be maintained at 8–12 cmH_2O. The urine output rate should be kept at >0.5 mL/kg per hour by continuing fluid administration; a diuretic such as furosemide may be used if needed. In about one-third of patients, hypotension and organ hypoperfusion respond to fluid resuscitation; a reasonable goal is to maintain a mean arterial blood pressure of >65 mmHg (systolic pressure >90 mmHg). If these guidelines cannot be met by volume infusion, vasopressor therapy is indicated (Chap. 272). Titrated doses of norepinephrine or dopamine should be administered through a central catheter. If myocardial dysfunction produces elevated cardiac filling pressures and low cardiac output, inotropic therapy with dobutamine is recommended.

In patients with septic shock, plasma vasopressin levels increase transiently but then decrease dramatically. Early studies found that vasopressin infusion can reverse septic shock in some patients, reducing or eliminating the need for catecholamine pressors. More recently, a randomized clinical trial that compared vasopressin plus norepinephrine with norepinephrine alone in 776 patients with pressor-dependent septic shock found no difference between treatment groups in the primary study outcome, 28-day mortality. Although vasopressin may have benefited patients who required less norepinephrine, its role in the treatment of septic shock seems to be a minor one overall.

CIRCI should be strongly considered in patients who develop hypotension that does not respond to fluid replacement therapy. Hydrocortisone (50 mg IV every 6 h) should be given; if clinical improvement occurs over 24–48 h, most experts would continue hydrocortisone therapy for 5–7 days before slowly tapering and discontinuing it. Meta-analyses of recent clinical trials have concluded that hydrocortisone therapy hastens recovery from septic shock without increasing long-term survival.

Ventilator therapy is indicated for progressive hypoxemia, hypercapnia, neurologic deterioration, or respiratory muscle failure. Sustained tachypnea (respiratory rate, >30 breaths/min) is frequently a harbinger of impending respiratory collapse; mechanical ventilation is often initiated to ensure adequate oxygenation, to divert blood from the muscles of respiration, to prevent aspiration of oropharyngeal contents, and to reduce the cardiac afterload. The results of recent studies favor the use of low tidal volumes (6 mL/kg of ideal body weight, or as low as 4 mL/kg if the plateau pressure exceeds 30 cmH_2O). Patients undergoing mechanical ventilation require careful sedation, with daily interruptions; elevation of the head of the bed helps to prevent nosocomial pneumonia. Stress-ulcer prophylaxis with a histamine H_2-receptor antagonist may decrease the risk of gastrointestinal hemorrhage in ventilated patients.

Erythrocyte transfusion is generally recommended when the blood hemoglobin level decreases to ≤7 g/dL, with a target level of 9 g/dL in adults. Erythropoietin is not used to treat sepsis-related anemia. Bicarbonate is sometimes administered for severe metabolic acidosis (arterial pH <7.2), but there is little evidence that it improves either hemodynamics or the response to vasopressor hormones. DIC, if complicated by major bleeding, should be treated with transfusion of fresh-frozen plasma and platelets. Successful treatment of the underlying infection is essential to reverse both acidosis and DIC. Patients who are hypercatabolic and have acute renal failure may benefit greatly from intermittent hemodialysis

or continuous veno-venous hemofiltration.

GENERAL SUPPORT

In patients with prolonged severe sepsis (i.e., lasting more than 2 or 3 days), nutritional supplementation may reduce the impact of protein hypercatabolism; the available evidence, which is not strong, favors the enteral delivery route. Prophylactic heparinization to prevent deep venous thrombosis is indicated for patients who do not have active bleeding or coagulopathy; when heparin is contraindicated, compression stockings or an intermittent compression device should be used. Recovery is also assisted by prevention of skin breakdown, nosocomial infections, and stress ulcers.

The role of tight control of the blood glucose concentration in recovery from critical illness has been addressed in numerous controlled trials. Meta-analyses of these trials have concluded that use of insulin to lower blood glucose levels to 100–120 mg/dL is potentially harmful and does not improve survival rates. Most experts now recommend using insulin only if it is needed to maintain the blood glucose concentration below ~150 mg/dL. Patients receiving intravenous insulin must be monitored frequently (every 1–2 h) for hypoglycemia.

OTHER MEASURES

Despite aggressive management, many patients with severe sepsis or septic shock die. Numerous interventions have been tested for their ability to improve survival rates among patients with severe sepsis. The list includes endotoxin-neutralizing proteins, inhibitors of cyclooxygenase or nitric oxide synthase, anticoagulants, polyclonal immunoglobulins, glucocorticoids, a phospholipid emulsion, and antagonists to TNF-α, IL-1, platelet-activating factor, and bradykinin. Unfortunately, none of these agents has improved rates of survival among patients with severe sepsis/septic shock in more than one large-scale, randomized, placebo-controlled clinical trial. Many factors have contributed to this lack of reproducibility, including (1) heterogeneity in the patient populations studied, the primary infection sites, the preexisting illnesses, and the inciting microbes; and (2) the nature of the "standard" therapy also used. A dramatic example of this problem was seen in a trial of tissue factor pathway inhibitor **(Fig. 271-1)**. Whereas the drug appeared to improve survival rates after 722 patients had been studied (p = .006), it did not do so in the next 1032 patients, and the overall result was negative. This inconsistency argues that the results of a clinical trial may not apply to individual patients, even within a carefully selected patient population. It also suggests that, at a minimum, a sepsis intervention should show a significant survival benefit in more than one placebo-controlled, randomized clinical trial before it is accepted as routine clinical practice. In one prominent attempt to reduce patient heterogeneity in clinical trials, experts have called for changes that would restrict these trials to patients who have similar underlying diseases (e.g., major trauma) and inciting infections (e.g., pneumonia). The goal of the predisposition–infection–response–organ dysfunction (PIRO) grading system for classification of septic patients (Table 271-1) is similar. Other investigators have used specific biomarkers, such as IL-6 levels in blood or the expression of HLA-DR on peripheral-blood monocytes, to identify the patients most likely to benefit from certain interventions. Multivariate risk stratification based on easily measurable clinical variables should be used with each of these approaches.

Figure 271-1

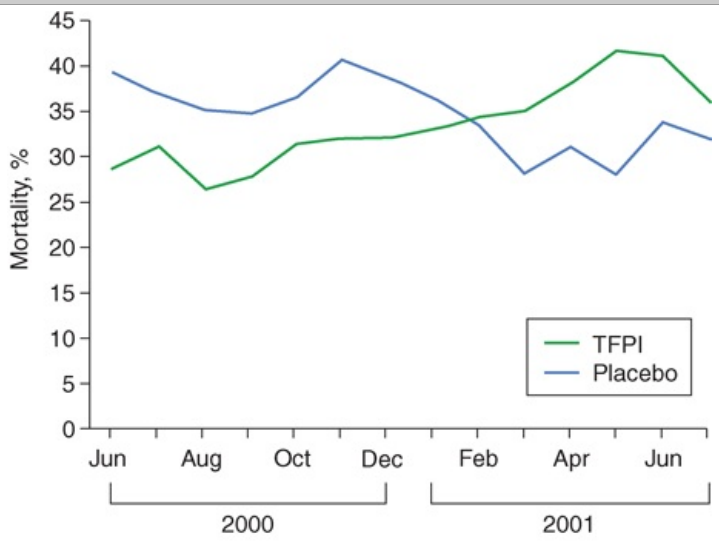

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mortality rates among patients who received tissue factor pathway inhibitor (TFPI) or placebo, shown as the running average over the course of the clinical trial. The drug seemed highly

placebo, shown as the running average over the course of the clinical trial. The drug seemed highly efficacious at the interim analysis in December 2000, but this trend reversed later in the trial. Demonstrating that therapeutic agents for sepsis have consistent, reproducible efficacy has been extremely difficult, even within well-defined patient populations. *(Reprinted with permission from Abraham et al.)*

Recombinant activated protein C (aPC) was the first drug to be approved by the U.S. Food and Drug Administration for the treatment of patients with severe sepsis or septic shock. Approval was based on the results of a single randomized controlled trial in which the drug was given within 24 h of the patient's first sepsis-related organ dysfunction; the 28-day survival rate was significantly higher among aPC recipients who were very sick (APACHE II score, ≥25) before infusion of the protein than among placebo-treated controls. Subsequent trials failed to show a benefit of aPC treatment in patients who were less sick (APACHE II score, <25) or in children. A second trial of aPC in high-risk patients is now under way in Europe. Given the drug's known toxicity (increased risk of severe bleeding) and uncertain performance in clinical practice, many experts are awaiting the results of the European trial before recommending further use of aPC. Other agents in ongoing or planned clinical trials include intravenous immunoglobulin, a small-molecule endotoxin antagonist (eritoran), and granulocyte-macrophage colony-stimulating factor that was recently reported to restore monocyte immunocompetence in patients with sepsis-associated immunosuppression.

A careful retrospective analysis found that the apparent efficacy of all sepsis therapeutics studied to date has been greatest among the patients at greatest risk of dying before treatment; conversely, use of many of these drugs has been associated with increased mortality rates among patients who are less ill. The authors proposed that neutralizing one of many different mediators may help patients who are very sick, whereas disrupting the mediator balance may be harmful to patients whose adaptive defense mechanisms are working well. This analysis suggests that if more aggressive early resuscitation improves survival rates among sicker patients, it will become more difficult to obtain additional benefit from other therapies; that is, if an intervention improves patients' risk status, moving them into a "less severe illness" category, it will be harder to show that adding another agent to the therapeutic regimen is beneficial.

THE SURVIVING SEPSIS CAMPAIGN

An international consortium has advocated "bundling" multiple therapeutic maneuvers into a unified algorithmic approach that will become the standard of care for severe sepsis. In theory, such a strategy could improve care by mandating measures that seem to bring maximal benefit, such as the rapid administration of appropriate antimicrobial therapy; on the other hand, this approach would deemphasize physicians' experience and judgment and minimize the consideration of potentially important differences between patients. Bundling multiple therapies into a single package also obscures the efficacy and toxicity of the individual measures. Caution should be engendered by the fact that two of the key elements of the initial algorithm have now been withdrawn for lack of evidence, while a third remains unproven and controversial.

Prognosis

Approximately 20–35% of patients with severe sepsis and 40–60% of patients with septic shock die within 30 days. Others die within the ensuing 6 months. Late deaths often result from poorly controlled infection, immunosuppression, complications of intensive care, failure of multiple organs, or the patient's underlying disease. Case-fatality rates are similar for culture-positive and culture-negative severe sepsis. Prognostic stratification systems such as APACHE II indicate that factoring in the patient's age, underlying condition, and various physiologic variables can yield estimates of the risk of dying of severe sepsis. Age and prior health status are probably the most important risk factors (**Fig. 271-2**). In patients with no known preexisting morbidity, the case-fatality rate remains below 10% until the fourth decade of life, after which it gradually increases to exceed 35% in the very elderly. Death is significantly more likely in severely septic patients with preexisting illness, especially during the third to fifth decades. Septic shock is also a strong predictor of short- and long-term mortality.

Figure 271-2

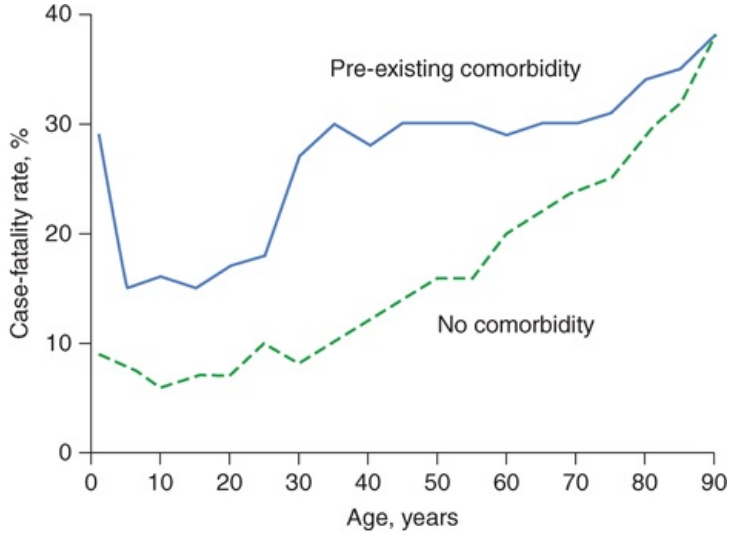

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Influence of age and prior health status on outcome from severe sepsis. With modern therapy, fewer than 10% of previously healthy young individuals (below 35 years of age) die with severe sepsis; the case-fatality rate then increases slowly through middle and old age. The most commonly identified etiologic agents in patients who die are *Staphylococcus aureus*, *Streptococcus pyogenes*, *S. pneumoniae*, and *Neisseria meningitidis*. Individuals with preexisting comorbidities are at greater risk of dying of severe sepsis at any age. The etiologic agents in these cases are likely to be *S. aureus*, *Pseudomonas aeruginosa*, various Enterobacteriaceae, enterococci, or fungi. *(Adapted from Angus et al., 2001.)*

Prevention

Prevention offers the best opportunity to reduce morbidity and mortality from severe sepsis. In developed countries, most episodes of severe sepsis and septic shock are complications of nosocomial infections. These cases might be prevented by reducing the number of invasive procedures undertaken, by limiting the use (and duration of use) of indwelling vascular and bladder catheters, by reducing the incidence and duration of profound neutropenia (<500 neutrophils/μL), and by more aggressively treating localized nosocomial infections. Indiscriminate use of antimicrobial agents and glucocorticoids should be avoided, and optimal infection-control measures (Chap. 131) should be used. Studies indicate that 50–70% of patients who develop nosocomial severe sepsis or septic shock have experienced a less severe stage of the septic response (e.g., SIRS, sepsis) on at least one previous day in the hospital. Research is needed to identify patients at increased risk and to develop adjunctive agents that can modulate the septic response before organ dysfunction or hypotension occurs.

FURTHER READINGS

Abraham E et al: Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis: A randomized controlled trial. JAMA 290:238, 2003[PMID: 12851279]

Angus DC et al: Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. Crit Care Med 29:1303, 2001[PMID: 11445675]

Barochia AV et al: Bundled care for septic shock: Analysis of clinical trials. Crit Care Med 38:668, 2010[PMID: 20029343]

DeBacker D et al: Comparison of dopamine and norepinephrine in the treatment of shock. N Engl J Med 362:779, 2010

Kumar A et al: Initiation of inappropriate antimicrobial therapy results in a 5-fold reduction of survival in human septic shock. Chest 136:1237, 2009[PMID: 19696123]

Marik PE: Critical illness-related corticosteroid insufficiency. Chest 135:181, 2009[PMID: 19136406]

Munford RS: Sensing gram-negative bacterial lipopolysaccharides: A human disease determinant? Infect Immun 76:454, 2008[PMID: 18086818]

Sands KE et al: Epidemiology of sepsis syndrome in 8 academic medical centers. JAMA 278:234, 1997[PMID: 9218672]

Toussaint S, Gerlach H: Activated protein C for sepsis. N Engl J Med 361:2646, 2009[PMID: 20042756]

Vincent J-L et al: Evolving concepts in sepsis definitions. Crit Care Clin 25:665, 2009[PMID: 19892246]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CARDIOGENIC SHOCK AND PULMONARY EDEMA: INTRODUCTION

Cardiogenic shock and pulmonary edema are life-threatening conditions that should be treated as medical emergencies. The most common etiology for both is severe left ventricular (LV) dysfunction that leads to pulmonary congestion and/or systemic hypoperfusion **(Fig. 272-1)**. **The pathophysiology of pulmonary edema and shock is discussed in Chaps. 33 and 270, respectively.**

Figure 272-1

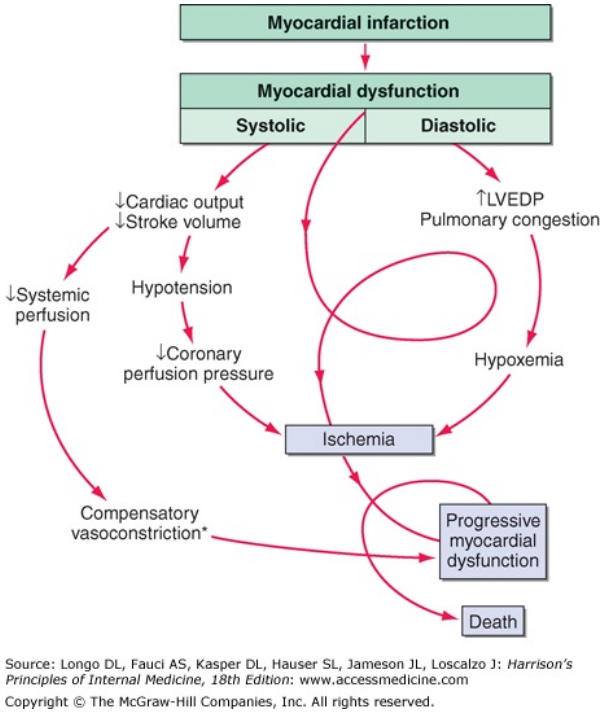

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathophysiology of cardiogenic shock. Systolic and diastolic myocardial dysfunction results in a reduction in cardiac output and often pulmonary congestion. Systemic and coronary hypoperfusion occur, resulting in progressive ischemia. Although a number of compensatory mechanisms are activated in an attempt to support the circulation, these compensatory mechanisms may become maladaptive and produce a worsening of hemodynamics. *Release of inflammatory cytokines after myocardial infarction may lead to inducible nitric oxide expression, excess nitric oxide, and inappropriate vasodilation. This causes further reduction in systemic and coronary perfusion. A vicious spiral of progressive myocardial dysfunction occurs that ultimately results in death if it is not interrupted. LVEDP, left ventricular end-diastolic pressure. (*From SM Hollenberg et al: Ann Intern Med 131:47, 1999* .)

CARDIOGENIC SHOCK

Cardiogenic shock (CS) is characterized by systemic hypoperfusion due to severe depression of the cardiac index [<2.2 (L/min)/m^2] and sustained systolic arterial hypotension (<90 mmHg) despite an elevated filling pressure [pulmonary capillary wedge pressure (PCWP) >18 mmHg]. It is associated with in-hospital mortality rates >50%. The major causes of CS are listed in **Table 272-1**. Circulatory failure based on cardiac dysfunction may be caused by primary myocardial failure, most commonly secondary to acute myocardial infarction (MI) (Chap. 245), and less frequently by cardiomyopathy or myocarditis (Chap. 238), cardiac tamponade (Chap. 239), or critical valvular heart disease (Chap. 237).

Table 272–1. Etiologies of Cardiogenic Shock (CS)[a] and Cardiogenic Pulmonary Edema

Etiologies of Cardiogenic Shock or Pulmonary Edema
Acute myocardial infarction/ischemia
LV failure
VSR
Papillary muscle/chordal rupture—severe MR
Ventricular free wall rupture with subacute tamponade
Other conditions complicating large MIs
Hemorrhage
Infection
Excess negative inotropic or vasodilator medications
Prior valvular heart disease

Hyperglycemia/ketoacidosis
Post-cardiac arrest
Post-cardiotomy
Refractory sustained tachyarrhythmias
Acute fulminant myocarditis
End-stage cardiomyopathy
Left ventricular apical ballooning
Takotsubo's cardiomyopathy
Hypertrophic cardiomyopathy with severe outflow obstruction
Aortic dissection with aortic insufficiency or tamponade
Pulmonary embolus
Severe valvular heart disease
Critical aortic or mitral stenosis
Acute severe aortic or MR
Toxic-metabolic
Beta-blocker or calcium channel antagonist overdose
Other Etiologies of Cardiogenic Shock[b]
RV failure due to:
Acute myocardial infarction
Acute coronary pulmonale
Refractory sustained bradyarrhythmias
Pericardial tamponade
Toxic/metabolic
Severe acidosis, severe hypoxemia

[a]The etiologies of CS are listed. Most of these can cause pulmonary edema instead of shock or pulmonary edema with CS.

[b]These cause CS but not pulmonary edema.

Abbreviations: LV, left ventricular; VSR, ventricular septal rupture; MR, mitral regurgitation; MI, myocardial infarction; RV, right ventricular.

INCIDENCE

CS is the leading cause of death of patients hospitalized with MI. Early reperfusion therapy for acute MI decreases the incidence of CS. The rate of CS complicating acute MI was 20% in the 1960s, stayed at ~8% for >20 years, but decreased to 5–7% in the first decade of this millennium. Shock typically is associated with ST elevation MI (STEMI) and is less common with non-ST elevation MI (Chap. 245).

LV failure accounts for ~80% of cases of CS complicating acute MI. Acute severe mitral regurgitation (MR), ventricular septal rupture (VSR), predominant right ventricular (RV) failure, and free wall rupture or tamponade account for the remainder.

PATHOPHYSIOLOGY

CS is characterized by a vicious circle in which depression of myocardial contractility, usually due to ischemia, results in reduced cardiac output and arterial pressure (BP), which result in hypoperfusion of the myocardium and further ischemia and depression of cardiac output (Fig. 272-1). Systolic myocardial dysfunction reduces stroke volume and, together with diastolic dysfunction, leads to elevated LV end-diastolic pressure and PCWP as well as to pulmonary congestion. Reduced coronary perfusion leads to worsening ischemia and progressive myocardial dysfunction and a rapid downward spiral, which, if uninterrupted, is often fatal. A systemic inflammatory response syndrome may accompany large infarctions and shock. Inflammatory cytokines, inducible nitric oxide synthase, and excess nitric oxide and peroxynitrite may contribute to the genesis of CS as they do to that of other forms of shock (Chap. 270). Lactic acidosis from poor tissue perfusion and hypoxemia from pulmonary edema may result from pump failure and then contribute to the vicious circle by worsening myocardial ischemia and hypotension. Severe acidosis (pH <7.25) reduces the efficacy of endogenous and exogenously administered catecholamines. Refractory sustained ventricular or atrial tachyarrhythmias can cause or exacerbate CS.

PATIENT PROFILE

In patients with acute MI, older age, female sex, prior MI, diabetes, and anterior MI location are all associated with an increased risk of CS. Shock associated with a first inferior MI should prompt a search for a mechanical cause. Reinfarction soon after MI increases the risk of CS. Two-thirds of patients with CS have flow-limiting stenoses in all three major coronary arteries, and 20% have stenosis of the left main coronary artery. CS may rarely occur in the absence of significant stenosis, as seen in LV apical ballooning/Takotsubo's cardiomyopathy.

TIMING

Shock is present on admission in only one-quarter of patients who develop CS complicating MI; one-quarter develop it rapidly thereafter, within 6 h of MI onset. Another quarter develop shock later on the first day. Subsequent onset of CS may be due to reinfarction, marked infarct expansion, or a mechanical complication.

DIAGNOSIS

Due to the unstable condition of these patients, supportive therapy must be initiated simultaneously with diagnostic evaluation **(Fig. 272-2)**. A focused history and physical examination should be performed, blood specimens sent to the laboratory, and an electrocardiogram (ECG) and chest x-ray obtained.

Figure 272-2

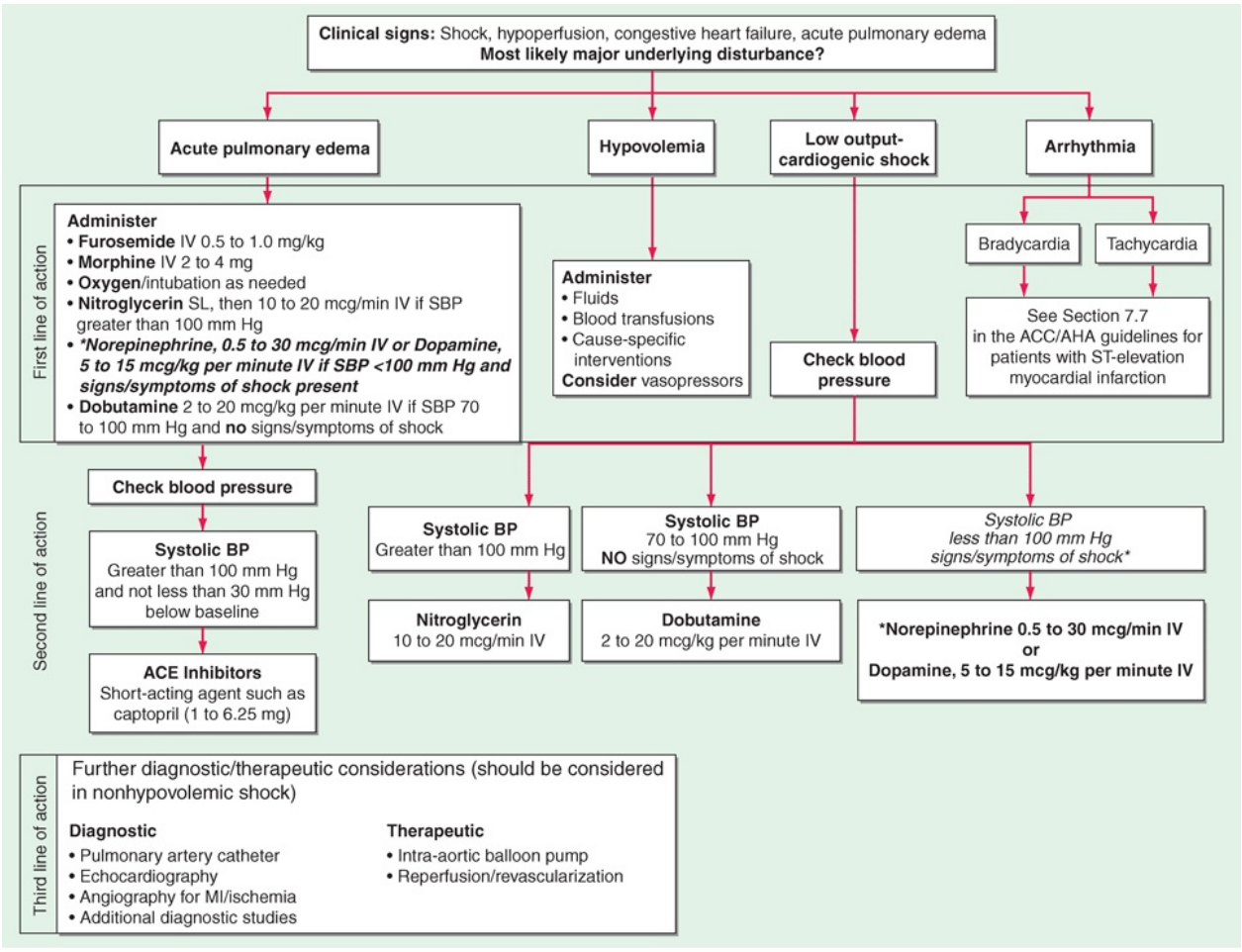

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The emergency management of patients with cardiogenic shock, acute pulmonary edema, or both is outlined. *Furosemide: <0.5 mg/kg for new-onset acute pulmonary edema without hypervolemia; 1 mg/kg for acute on chronic volume overload, renal insufficiency. †For management of bradycardia and tachycardia, see Chaps. 232 and 233. *Indicates modification from published guidelines. ACE, angiotensin-converting enzyme; BP, blood pressure; MI, myocardial infarction. (*Modified from Guidelines 2000 for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Part 7:The era of reperfusion: Section 1: Acute coronary syndromes (acute myocardial infarction). The American Heart Association in collaboration with the International Liaison Committee on Resuscitation. Circulation 102:I172, 2000.*)

Echocardiography is an invaluable diagnostic tool in patients suspected of CS.

Clinical Findings

Most patients have continuing chest pain and dyspnea and appear pale, apprehensive, and diaphoretic. Mentation may be altered, with somnolence, confusion, and agitation. The pulse is typically weak and rapid, often in the range of 90–110 beats/min, or severe bradycardia due to high-grade heart block may be present. Systolic blood pressure is reduced (<90 mmHg) with a narrow pulse pressure (<30 mmHg), but occasionally BP may be maintained by very high systemic vascular resistance. Tachypnea, Cheyne-Stokes respirations, and jugular venous distention may be present. The precordium is typically quiet, with a weak apical pulse. S_1 is usually soft, and an S_3 gallop may be audible. Acute, severe MR and VSR usually are associated with characteristic systolic murmurs (Chap. 245). Rales are audible in most patients with LV failure causing CS. Oliguria (urine output <30 mL/h) is common.

Laboratory Findings

The white blood cell count is typically elevated with a left shift. In the absence of prior renal insufficiency, renal function is initially normal, but blood urea nitrogen and creatinine rise progressively. Hepatic transaminases may be markedly elevated due to liver hypoperfusion. Poor tissue perfusion may result in an anion-gap acidosis and elevation of the lactic acid level. Before support with supplemental O_2, arterial blood gases usually demonstrate hypoxemia and metabolic acidosis, which may be compensated by respiratory alkalosis. Cardiac markers, creatine phosphokinase and its MB fraction, and troponins I and T are markedly elevated.

Electrocardiogram

In CS due to acute MI with LV failure, Q waves and/or >2-mm ST elevation in multiple leads or left bundle branch block are usually present. More than one-half of all infarcts associated with shock are anterior. Global ischemia due to severe left main stenosis usually is accompanied by severe (e.g., >3 mm) ST depressions in multiple leads.

Chest Roentgenogram

The chest x-ray typically shows pulmonary vascular congestion and often pulmonary edema, but these findings may be absent in up to a third of patients. The heart size is usually normal when CS results from a first MI but is enlarged when it occurs in a patient with a previous MI.

Echocardiogram

A two-dimensional echocardiogram with color-flow Doppler (Chap. 229) should be obtained promptly in patients with suspected CS to help define its etiology. Doppler mapping demonstrates a left-to-right shunt in patients with VSR and the severity of MR when the latter is present. Proximal aortic dissection with aortic regurgitation or tamponade may be visualized, or evidence for pulmonary embolism may be obtained (Chap. 262).

Pulmonary Artery Catheterization

There is controversy regarding the use of pulmonary artery (Swan-Ganz) catheters in patients with established or suspected CS (Chaps. 230 and 267). Their use is generally recommended for measurement of filling pressures and cardiac output to confirm the diagnosis and optimize the use of IV fluids, inotropic agents, and vasopressors in persistent shock **(Table 272-2)**. Blood samples for O_2 saturation measurement should be obtained from the right atrium, right ventricle, and pulmonary artery to rule out a left-to-right shunt. Mixed venous O_2 saturations are low and arteriovenous (AV) O_2 differences are elevated, reflecting low cardiac index and high fractional O_2 extraction. However, when a systemic inflammatory response syndrome accompanies CS, AV O_2 differences may not be elevated (Chap. 270). The PCWP is elevated. However, use of sympathomimetic amines may return these measurements and the systemic BP to normal. Systemic vascular resistance may be low, normal, or elevated in CS. Equalization of right- and left-sided filling pressures (right atrial and PCWP) suggests cardiac tamponade as the cause of CS (Chap. 239).

Table 272–2. Hemodynamic Patterns[a]

	RA, mmHg	RVS, mmHg	RVD, mmHg	PAS, mmHg	PAD, mmHg	PCW, mmHg	CI, (L/min)/m^2	SVR, (dyn · s)/cm^5
Normal values	<6	<25	0–12	<25	0–12	<6–12	≥2.5	(800–1600)
MI without pulmonary edema[b]	—	—	—	—	—	~13 (5–18)	~2.7 (2.2–4.3)	—
Pulmonary edema	↔↑	↔↑	↔↑	↑	↑	↑	↔↓	↑
Cardiogenic shock								
LV failure	↔↑	↔↑	↔↑	↔↑	↑	↑	↓	↔↑
RV failure[c]	↑	↓↔↑[d]	↑	↓↔↑[d]	↔↓↑[d]	↓↔↑[d]	↓	↑
Cardiac tamponade	↑	↔↑	↑	↔↑	↔↑	↔↑	↓	↑
Acute mitral regurgitation	↔↑	↑	↔↑	↑	↑	↑	↔↓	↔↑
Ventricular septal rupture	↑	↔↑	↑	↔↑	↔↑	↔↑	↑PBF↓SBF	↔↑
Hypovolemic shock	↓	↔↓	↔↓	↓	↓	↓	↓	↑
Septic shock	↓	↔↓	↔↓	↓	↓	↓	↑	↓

[a]There is significant patient-to-patient variation. Pressure may be normalized if cardiac output is low.

[b]Forrester et al classified nonreperfused MI patients into four hemodynamic subsets. (From Forrester JS et al: N Engl J Med 295:1356, 1976.) PCWP and CI in clinically stable subset 1 patients are shown. Values in parentheses represent range.

[c]"Isolated" or predominant RV failure.

[d]PCW and PA pressures may rise in RV failure after volume loading due to RV dilation, right-to-left shift of the interventricular septum, resulting in impaired LV filling. When biventricular failure is present, the patterns are similar to those shown for LV failure.

Abbreviations: CI, cardiac index; MI, myocardial infarction; P/SBF, pulmonary/systemic blood flow; PAS/D, pulmonary artery systolic/diastolic; PCW, pulmonary capillary wedge; RA, right atrium; RVS/D, right ventricular systolic/diastolic; SVR, systemic vascular resistance.

Source: Table prepared with the assistance of Krishnan Ramanathan, MD.

Left Heart Catheterization and Coronary Angiography

Measurement of LV pressure and definition of the coronary anatomy provide useful information and are indicated in most patients with CS complicating MI. Cardiac catheterization should be performed when there is a plan and capability for immediate coronary intervention (see below) or when a definitive diagnosis has not been made by other tests.

Treatment: Acute Myocardial Infarction

GENERAL MEASURES

(Fig. 272-2) In addition to the usual treatment of acute MI (Chap. 245), initial therapy is aimed at maintaining adequate systemic and coronary perfusion by raising systemic BP with vasopressors and adjusting volume status to a level that ensures optimum LV filling pressure. There is interpatient variability, but the values that generally are associated with adequate perfusion are systolic BP ~90 mmHg or mean BP >60 mmHg and PCWP >20 mmHg. Hypoxemia and acidosis must be corrected; most patients require ventilatory support with either endotracheal intubation or noninvasive ventilation to correct these abnormalities and reduce the work of breathing (see "Pulmonary Edema," below). Negative inotropic agents should be discontinued, and the doses of renally cleared medications adjusted. Hyperglycemia should be controlled with insulin. Bradyarrhythmias may require transvenous pacing. Recurrent ventricular tachycardia or rapid atrial fibrillation may require immediate treatment (Chap. 233).

VASOPRESSORS

Various IV drugs may be used to augment BP and cardiac output in patients with CS. All have important disadvantages, and none has been shown to change the outcome in patients with established shock. *Norepinephrine* is a potent

vasoconstrictor and inotropic stimulant that is useful for patients with CS. As first line of therapy norepinephrine was associated with fewer adverse events, including arrhythmias, compared to a dopamine randomized trial of patients with several eteologies of circulatory shock. Although it did not significantly improve survival compared to dopamine, its relative safety suggests that norepinephrine is reasonable as initial vasopressor therapy. Norepinephrine should be started at a dose of 2 to 4 μg/min and titrated upward as necessary. If systemic perfusion or systolic pressure cannot be maintained at >90 mmHg with a dose of 15 μg/min, it is unlikely that a further increase will be beneficial.

Dopamine has varying hemodynamic effects based on the dose; at low doses (≤ 2 μg/kg per min), it dilates the renal vascular bed, although its outcome benefits at this low dose have not been demonstrated conclusively; at moderate doses (2–10 μg/kg per min), it has positive chronotropic and inotropic effects as a consequence of β-adrenergic receptor stimulation. At higher doses, a vasoconstrictor effect results from α-receptor stimulation. It is started at an infusion rate of 2–5 μg/kg per min, and the dose is increased every 2–5 min to a maximum of 20–50 μg/kg per min. *Dobutamine* is a synthetic sympathomimetic amine with positive inotropic action and minimal positive chronotropic activity at low doses (2.5 μg/kg per min) but moderate chronotropic activity at higher doses. Although the usual dose is up to 10 μg/kg per min, its vasodilating activity precludes its use when a vasoconstrictor effect is required.

AORTIC COUNTERPULSATION

In CS, mechanical assistance with an intraaortic balloon pumping (IABP) system capable of augmenting both arterial diastolic pressure and cardiac output is helpful in rapidly stabilizing patients. A sausage-shaped balloon is introduced percutaneously into the aorta via the femoral artery; the balloon is automatically inflated during early diastole, augmenting coronary blood flow. The balloon collapses in early systole, reducing the afterload against which the LV ejects. IABP improves hemodynamic status temporarily in most patients. In contrast to vasopressors and inotropic agents, myocardial O_2 consumption is reduced, ameliorating ischemia. IABP is useful as a stabilizing measure in patients with CS before and during cardiac catheterization and percutaneous coronary intervention (PCI) or before urgent surgery. IABP is contraindicated if aortic regurgitation is present or aortic dissection is suspected. Ventricular assist devices may be considered for eligible young patients with refractory shock as a bridge to cardiac transplantation (Chap. 235).

REPERFUSION-REVASCULARIZATION

The rapid establishment of blood flow in the infarct-related artery is essential in the management of CS and forms the centerpiece of management. The randomized SHOCK Trial demonstrated that 132 lives were saved per 1000 patients treated with early revascularization with PCI or coronary artery bypass graft (CABG) compared with initial medical therapy including IABP with fibrinolytics followed by delayed revascularization. The benefit is seen across the risk strata and is sustained up to 11 years after an MI. Early revascularization with PCI or CABG is a class I recommendation for patients age <75 years with ST elevation or left bundle branch block MI who develop CS within 36 h of MI and who can be revascularized within 18 h of the development of CS. When mechanical revascularization is not possible, IABP and fibrinolytic therapy are recommended. Older patients who are suitable candidates for aggressive care also should be offered early revascularization.

PROGNOSIS

Within this high-risk condition, there is a wide range of expected death rates based on age, severity of hemodynamic abnormalities, severity of the clinical manifestations of hypoperfusion, and the performance of early revascularization.

Shock Secondary to Right Ventricular Infarction

Although transient hypotension is common in patients with RV infarction and inferior MI (Chap. 245), persistent CS due to RV failure accounts for only 3% of CS complicating MI. The salient features of RV shock are absence of pulmonary congestion, high right atrial pressure (which may be seen only after volume loading), RV dilation and dysfunction, only mildly or moderately depressed LV function, and predominance of single-vessel proximal right coronary artery occlusion. Management includes IV fluid administration to optimize right atrial pressure (10–15 mmHg); avoidance of excess fluids, which cause a shift of the interventricular septum into the LV; sympathomimetic amines; IABP; and the early reestablishment of infarct-artery flow.

Mitral Regurgitation

(See also Chap. 245) Acute severe MR due to papillary muscle dysfunction and/or rupture may complicate MI and result in CS and/or pulmonary edema. This complication most often occurs on the first day, with a second peak several days later. The diagnosis is confirmed by echo-Doppler. Rapid stabilization with IABP is recommended, with administration of dobutamine as needed to raise cardiac output. Reducing the load against which the LV pumps (afterload) reduces the volume of regurgitant flow of blood into the left atrium. Mitral valve surgery is the definitive therapy and should be performed early in the course in suitable candidates.

Ventricular Septal Rupture

(See also Chap. 245) Echo-Doppler demonstrates shunting of blood from the left to the right ventricle and may visualize the opening in the interventricular septum. Timing and management are similar to those for MR with IABP support and surgical correction for suitable candidates.

Free Wall Rupture

Myocardial rupture is a dramatic complication of STEMI that is most likely to occur during the first week after the onset of symptoms; its frequency increases with the age of the patient. The clinical presentation typically is a sudden loss of pulse, blood pressure, and consciousness but sinus rhythm on ECG (pulseless electrical activity) due to cardiac tamponade (Chap. 239). Free wall rupture may also result in CS due to subacute tamponade when the pericardium temporarily seals the rupture sites. Definitive surgical repair is required.

Acute Fulminant Myocarditis

(See also Chap. 238) Myocarditis can mimic acute MI with ST deviation or bundle branch block on the ECG and marked elevation of cardiac markers. Acute myocarditis causes CS in a small proportion of cases. These patients are typically younger than those with CS due to acute MI and often do not have typical ischemic chest pain. Echocardiography usually shows global LV dysfunction. Initial management is the same as for CS complicating acute MI (Fig. 272-2) but does not involve coronary revascularization.

PULMONARY EDEMA

The etiologies and pathophysiology of pulmonary edema are discussed in Chap. 33.

DIAGNOSIS

Acute pulmonary edema usually presents with the rapid onset of dyspnea at rest, tachypnea, tachycardia, and severe hypoxemia. Rales and wheezing due to airway compression from peribronchial cuffing may be audible. Hypertension is usually present due to release of endogenous catecholamines.

It is often difficult to distinguish between cardiogenic and noncardiogenic causes of acute pulmonary edema. *Echocardiography* may identify systolic and diastolic ventricular dysfunction and valvular lesions. Pulmonary edema associated with electrocardiographic ST elevation and evolving Q waves is usually diagnostic of acute MI and should prompt immediate institution of MI protocols and coronary artery reperfusion therapy (Chap. 245). Brain natriuretic peptide levels, when substantially elevated, support heart failure as the etiology of acute dyspnea with pulmonary edema (Chap. 234).

The use of a *Swan-Ganz catheter* permits measurement of PCWP and helps differentiate high-pressure (cardiogenic) from normal-pressure (noncardiogenic) causes of pulmonary edema. Pulmonary artery catheterization is indicated when the etiology of the pulmonary edema is uncertain, when it is refractory to therapy, or when it is accompanied by hypotension. Data derived from use of a catheter often alter the treatment plan, but the impact on mortality rates has not been demonstrated.

Treatment: Pulmonary Edema

The treatment of pulmonary edema depends on the specific etiology. In light of the acute, life-threatening nature of the condition, a number of measures must be applied immediately to support the circulation, gas exchange, and lung mechanics. In addition, conditions that frequently complicate pulmonary edema, such as infection, acidemia, anemia, and renal failure, must be corrected.

SUPPORT OF OXYGENATION AND VENTILATION

Patients with acute cardiogenic pulmonary edema generally have an identifiable cause of acute LV failure—such as arrhythmia, ischemia/infarction, or myocardial decompensation (Chap. 234)—that can be rapidly treated, with improvement in gas exchange. In contrast, noncardiogenic edema usually resolves much less quickly, and most patients require mechanical ventilation.

OXYGEN THERAPY

Support of oxygenation is essential to ensure adequate O_2 delivery to peripheral tissues, including the heart.

Positive-Pressure Ventilation

Pulmonary edema increases the work of breathing and the O_2 requirements of this work, imposing a significant physiologic stress on the heart. When oxygenation or ventilation is not adequate in spite of supplemental O_2, positive-pressure ventilation by face or nasal mask or by endotracheal intubation should be initiated. Noninvasive ventilation (Chap. 269) can rest the respiratory muscles, improve oxygenation and cardiac function, and reduce the need for intubation. In refractory cases, mechanical ventilation can relieve the work of breathing more completely than can noninvasive ventilation. Mechanical ventilation with positive end-expiratory pressure can have multiple beneficial effects on pulmonary edema: (1) decreases both preload and afterload, thereby improving cardiac function, (2) redistributes lung water from the intraalveolar to the extraalveolar space, where the fluid interferes less with gas exchange, and (3) increases lung volume to avoid atelectasis.

REDUCTION OF PRELOAD

In most forms of pulmonary edema, the quantity of extravascular lung water is determined by both the PCWP and the intravascular volume status.

Diuretics

The "loop diuretics" furosemide, bumetanide, and torsemide are effective in most forms of pulmonary edema, even in the presence of hypoalbuminemia, hyponatremia, or hypochloremia. Furosemide is also a venodilator that reduces preload rapidly, before any diuresis, and is the diuretic of choice. The initial dose of furosemide should be ≤0.5 mg/kg, but a higher dose (1 mg/kg) is required in patients with renal insufficiency, chronic diuretic use, or hypervolemia or after failure of a lower dose.

Nitrates

Nitroglycerin and isosorbide dinitrate act predominantly as venodilators but have coronary vasodilating properties as well. They are rapid in onset and effective when administered by a variety of routes. Sublingual nitroglycerin (0.4 mg x 3 every 5 min) is first-line therapy for acute cardiogenic pulmonary edema. If pulmonary edema persists in the absence of hypotension, sublingual may be followed by IV nitroglycerin, commencing at 5–10 μg/min. IV nitroprusside (0.1–5 μg/kg per min) is a potent venous and arterial vasodilator. It is useful for patients with pulmonary edema and hypertension but is not recommended in states of reduced coronary artery perfusion. It requires close monitoring and titration using an arterial catheter for continuous BP measurement.

Morphine

Given in 2- to 4-mg IV boluses, morphine is a transient venodilator that reduces preload while relieving dyspnea and anxiety. These effects can diminish stress, catecholamine levels, tachycardia, and ventricular afterload in patients with pulmonary edema and systemic hypertension.

Angiotensin-Converting Enzyme (ACE) Inhibitors

ACE inhibitors reduce both afterload and preload and are recommended for hypertensive patients. A low dose of a short-acting agent may be initiated and followed by increasing oral doses. In acute MI with heart failure, ACE inhibitors reduce short- and long-term mortality rates.

Other Preload-Reducing Agents

IV recombinant brain natriuretic peptide (nesiritide) is a potent vasodilator with diuretic properties and is effective in the treatment of cardiogenic pulmonary edema. It should be reserved for refractory patients and is not recommended in the setting of ischemia or MI.

Physical Methods

Reduction of venous return reduces preload. Patients without hypotension should be maintained in the sitting position with the legs dangling along the side of the bed.

Inotropic and Inodilator Drugs

The sympathomimetic amines dopamine and dobutamine (see above) are potent inotropic agents. The bipyridine phosphodiesterase-3 inhibitors (inodilators), such as milrinone (50 μg/kg followed by 0.25–0.75 μg/kg per min), stimulate myocardial contractility while promoting peripheral and pulmonary vasodilation. Such agents are indicated in patients with cardiogenic pulmonary edema and severe LV dysfunction.

Digitalis Glycosides

Once a mainstay of treatment because of their positive inotropic action (Chap. 234), digitalis glycosides are rarely used at present. However, they may be useful for control of ventricular rate in patients with rapid atrial fibrillation or flutter and LV dysfunction, since they do not have the negative inotropic effects of other drugs that inhibit atrioventricular nodal conduction.

Intraaortic Counterpulsation

IABP may help relieve cardiogenic pulmonary edema. It is indicated as a stabilizing measure when acute severe mitral regurgitation or ventricular septal rupture causes refractory pulmonary edema, especially in preparation for surgical repair. IABP or LV-assist devices (Chap. 235) are useful as bridging therapy to cardiac transplantation in patients with refractory pulmonary edema secondary to myocarditis or cardiomyopathy.

Treatment of Tachyarrhythmias and Atrial-Ventricular Resynchronization

(See also Chap. 233) Sinus tachycardia or atrial fibrillation can result from elevated left atrial pressure and sympathetic stimulation. Tachycardia itself can limit LV filling time and raise left atrial pressure further. Although relief of pulmonary congestion will slow the sinus rate or ventricular response in atrial fibrillation, a primary tachyarrhythmia may require cardioversion. In patients with reduced LV function and without atrial contraction or with lack of synchronized atrioventricular contraction, placement of an atrioventricular sequential pacemaker should be considered (Chap. 232).

Stimulation of Alveolar Fluid Clearance

Recent mechanistic studies on alveolar epithelial ion transport have defined a variety of ways to upregulate the clearance of solute and water from the alveolar space. In patients with acute lung injury (noncardiogenic pulmonary edema), IV β-adrenergic agonist treatment decreases extravascular lung water, but the outcome benefit is uncertain.

Special Considerations

The Risk of Iatrogenic Cardiogenic Shock

In the treatment of pulmonary edema vasodilators lower BP, and, particularly when used in combination, their use may lead to hypotension, coronary artery hypoperfusion, and shock (Fig. 272-1). In general, patients with a *hypertensive* response to pulmonary edema tolerate and benefit from these medications. In normotensive patients, low doses of single agents should be instituted sequentially, as needed.

Acute Coronary Syndromes

(See also Chap. 245) Acute STEMI complicated by pulmonary edema is associated with in-hospital mortality rates of 20–40%. After immediate stabilization, coronary artery blood flow must be reestablished rapidly. When available, primary PCI is preferable; alternatively, a fibrinolytic agent should be administered. Early coronary angiography and revascularization by PCI or CABG also are indicated for patients with non-ST elevation acute coronary syndrome. IABP use may be required to stabilize patients for coronary angiography if hypotension develops or for refractory pulmonary edema in patients with LV failure who are candidates for revascularization.

Unusual Types of Edema

Specific etiologies of pulmonary edema may require particular therapy. Reexpansion pulmonary edema can develop after removal of air or fluid that has been in the pleural space for some time. These patients may develop hypotension or oliguria resulting from rapid fluid shifts into the lung. Diuretics and preload reduction are contraindicated, and intravascular volume repletion often is needed while supporting oxygenation and gas exchange.

High-altitude pulmonary edema often can be prevented by use of dexamethasone, calcium channel-blocking drugs, or long-acting inhaled β_2-adrenergic agonists. Treatment includes descent from altitude, bed rest, oxygen, and, if feasible, inhaled nitric oxide; nifedipine may also be effective.

For pulmonary edema resulting from upper airway obstruction, recognition of the obstructing cause is key, since treatment then is to relieve or bypass the obstruction.

FURTHER READINGS

Antman EM: Treatment of ST elevation myocardial infarction, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

——— et al: ACC/AHA Guidelines for the management of patients with acute myocardial infarction: A report of the American College of Cardiology/American Heart Association task force on practice guidelines (Committee on Management of Acute Myocardial Infarction). J Am Coll Cardiol 44:671, 2004

Goldberg RJ, et al: Thirty-year trends (1975 to 2005) in the magnitude of, management of, and hospital death rates associated with cardiogenic shock in patients with acute myocardial infarction: A population-based perspective. Circulation 119:1211, 2009[PMID: 19237658]

Hochman JS, Ohman EM: *Cardiogenic Shock*, American Heart Association Clinical Series, E Antman (ed). New York, Wiley–Blackwell, 2008

——— et al: Early revascularization and long-term survival in cardiogenic shock complicating acute myocardial infarction. JAMA 295:2511, 2006

Matthay MA, Ingbar DH (eds): *Pulmonary Edema. Lung Biology in Health and Disease* , vol 116. New York, Marcel Dekker, 1998

Okuda M: A multidisciplinary overview of cardiogenic shock. Shock 25:557, 2006[PMID: 16721262]

Reynolds HR, Hochman JS: Cardiogenic shock: Current concepts and improving outcomes. Circulation 117:686, 2008[PMID: 18250279]

Ware LB, Matthay MA: Clinical practice: Acute pulmonary edema. N Engl J Med 353:2788, 2005[PMID: 16382065]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OVERVIEW AND DEFINITIONS

Sudden cardiac death (SCD) is defined *as natural death due to cardiac causes* in a person who may or may not have previously recognized heart disease but in whom the time and mode of death are *unexpected*. In the context of time, "sudden" is defined for most clinical and epidemiologic purposes as *1 h or less* between a change in clinical status heralding the onset of the terminal clinical event and the cardiac arrest itself. An exception is unwitnessed deaths, in which pathologists may expand the definition of time to 24 h after the victim was last seen to be alive and stable.

The overwhelming majority of natural deaths are caused by cardiac disorders. However, it is common for underlying heart diseases—often far advanced—to go unrecognized before the fatal event. As a result, up to two-thirds of all SCDs occur as the first clinical expression of previously undiagnosed disease or in patients with known heart disease, the extent of which suggests low risk. The magnitude of sudden *cardiac* death as a public health problem is highlighted by the estimate that ~50% of all cardiac deaths are sudden and unexpected, accounting for a total SCD burden estimated to range from <200,000 to >450,000 deaths each year in the United States. SCD is a direct consequence of cardiac arrest, which may be reversible if addressed promptly. Since resuscitation techniques and emergency rescue systems are available to respond to victims of out-of-hospital cardiac arrest, which was uniformly fatal in the past, understanding the SCD problem has practical clinical importance.

Because of community-based interventions, victims may remain biologically alive for days or even weeks after a cardiac arrest that has resulted in irreversible central nervous system damage. Confusion in terms can be avoided by adhering strictly to definitions of cardiovascular collapse, cardiac arrest, and death **(Table 273-1)**. Although cardiac arrest is often potentially reversible by appropriate and timely interventions, death is biologically, legally, and literally an absolute and irreversible event. Death may be delayed in a survivor of cardiac arrest, but "survival after sudden death" is an irrational term. When biologic death of a cardiac arrest victim is delayed because of interventions, the relevant pathophysiologic event remains the sudden and unexpected cardiac arrest that leads ultimately to death, even though delayed by interventions. The language used should reflect the fact that the index event was a cardiac arrest and that death was due to its delayed consequences. Accordingly, for statistical purposes, deaths that occur during hospitalization or within 30 days after resuscitated cardiac arrest are counted as sudden deaths.

Table 273–1. Distinction between Cardiovascular Collapse, Cardiac Arrest, and Death

Term	Definition	Qualifiers	Mechanisms
Cardiovascular collapse	Sudden loss of effective blood flow due to cardiac and/or peripheral vascular factors that may reverse spontaneously (e.g., neurocardiogenic syncope, vasovagal syncope) or require interventions (e.g., cardiac arrest)	Nonspecific term: includes cardiac arrest and its consequences and transient events that characteristically revert spontaneously	Same as "Cardiac Arrest," plus vasodepressor syncope or other causes of transient loss of blood flow
Cardiac arrest	Abrupt cessation of cardiac mechanical function, which may be reversible by a prompt intervention but will lead to death in its absence	Rare spontaneous reversions; likelihood of successful intervention relates to mechanism of arrest, clinical setting, and prompt return of circulation	Ventricular fibrillation, ventricular tachycardia, asystole, bradycardia, pulseless electrical activity, mechanical factors
Sudden cardiac death	Sudden, irreversible cessation of all biological functions	None	

Source: Modified from Myerburg and Castellanos, 2008, with permission of publisher.

Clinical Definition of Forms of Cardiovascular Collapse

Cardiovascular collapse is a general term connoting loss of sufficient cerebral blood flow to maintain consciousness due to acute dysfunction of the heart and/or peripheral vasculature. It may be caused by vasodepressor syncope (vasovagal syncope, postural hypotension with syncope, neurocardiogenic syncope; Chap. 20), a transient severe bradycardia, or cardiac arrest. The latter is distinguished from the transient forms of cardiovascular collapse in that it usually requires an intervention to restore spontaneous blood flow. In contrast, vasodepressor syncope and other primary bradyarrhythmic syncopal events are

transient and non-life-threatening, with spontaneous return of consciousness.

The most common electrical mechanism for cardiac arrest is ventricular fibrillation (VF), which is responsible for 50–80% of cardiac arrests. Severe persistent bradyarrhythmias, asystole, and pulseless electrical activity [PEA: organized electrical activity, unusually slow, without mechanical response, formerly called electromechanical dissociation (EMD)] cause another 20–30%. Pulseless sustained ventricular tachycardia (a rapid arrhythmia distinct from PEA) is a less common mechanism. Acute low cardiac output states, having a precipitous onset, also may present clinically as a cardiac arrest. These hemodynamic causes include massive acute pulmonary emboli, internal blood loss from a ruptured aortic aneurysm, intense anaphylaxis, and cardiac rupture with tamponade after myocardial infarction (MI). Sudden deaths due to these causes are not included in the SCD category.

ETIOLOGY, INITIATING EVENTS, AND CLINICAL EPIDEMIOLOGY

Clinical, epidemiologic, and pathologic studies have provided information on the underlying *structural abnormalities* in victims of SCD and identified subgroups at high risk for SCD. In addition, studies of clinical physiology have begun to identify *transient functional factors* that may convert a long-standing underlying structural abnormality from a stable to an unstable state, leading to the onset of cardiac arrest **(Table 273-2)**.

Table 273–2. Cardiac Arrest and Sudden Cardiac Death

Structural Associations and Causes
I. Coronary heart disease
A. Coronary artery abnormalities
1. Chronic atherosclerotic lesions
2. Acute (active) lesions (plaque fissuring, platelet aggregation, acute thrombosis)
3. Anomalous coronary artery anatomy
B. Myocardial Infarction
1. Healed
2. Acute
II. Myocardial hypertrophy
A. Secondary
B. Hypertrophic cardiomyopathy
1. Obstructive
2. Nonobstructive
III. Dilated cardiomyopathy—primary muscle disease
IV. Inflammatory and infiltrative disorders.
A. Myocarditis
B. Noninfectious inflammatory diseases
C. Infiltrative diseases
V. Valvular heart disease
VI. Electrophysiologic abnormalities, structural
A. Anomalous pathways in Wolff-Parkinson-White syndrome
B. Conducting system disease
VII. Inherited disorders associated with electrophysiological abnormalities (congenital long QT syndromes, right ventricular dysplasia, Brugada syndrome, catecholaminergic polymorphic ventricular tachycardia, etc.)
Functional Contributing Factors
I. Alterations of coronary blood flow
A. Transient ischemia
B. Reperfusion after ischemia
II. Low cardiac output states
A. Heart failure
1. Chronic
2. Acute decompensation
B. Shock
III. Systemic metabolic abnormalities
A. Electrolyte imbalance (e.g., hypokalemia)
B. Hypoxemia, acidosis
IV. Neurologic disturbances

A. Autonomic fluctuations: central, neural, humoral
B. Receptor function
V. Toxic responses
A. Proarrhythmic drug effects
B. Cardiac toxins (e.g., cocaine, digitalis intoxication)
C. Drug interactions

Cardiac disorders constitute the most common causes of sudden *natural* death. After an initial peak incidence of sudden death between birth and 6 months of age [the sudden infant death syndrome (SIDS)], the incidence of sudden death declines sharply and remains low through childhood and adolescence. Among adolescents and young adults, the incidence of SCD is approximately 1 per 100,000 population per year. The incidence begins to increase in adults over age 30 years, reaching a second peak in the age range 45–75 years, when it approximates 1–2 per 1000 per year among the unselected adult population. Increasing age within this range is associated with increasing risk for sudden *cardiac* death **(Fig. 273-1*A*)**. From 1 to 13 years of age, only one of five sudden *natural* deaths is due to cardiac causes. Between 14 and 21 years of age, the proportion increases to 30%, and it rises to 88% in the middle-aged and elderly.

Figure 273-1

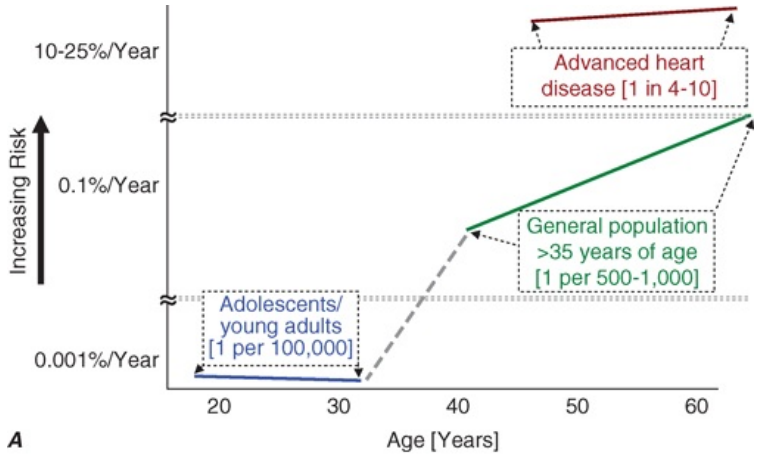

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

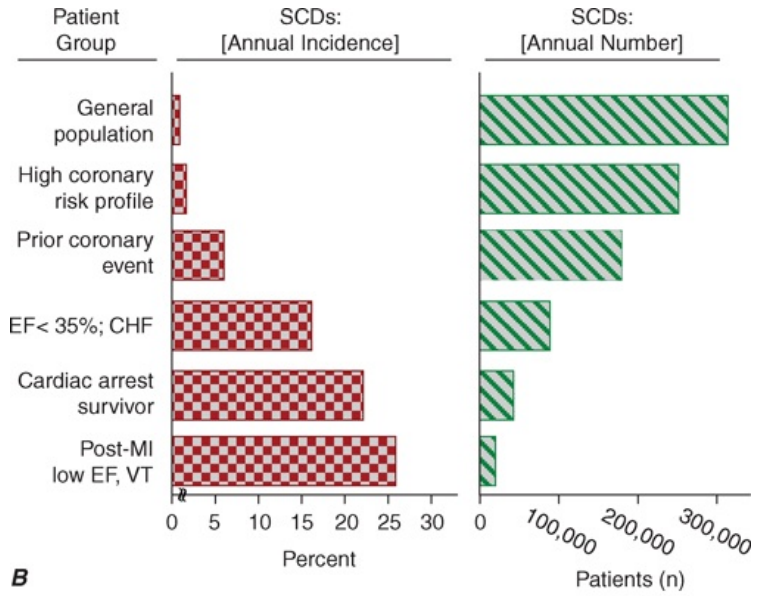

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***Panel A* demonstrates age-related risk for SCD.** For the general population age 35 years and older, SCD risk is 0.1–0.2 percent per year (1 per 500–1000 population). Among the general population of adolescents and adults younger than age 30 years, the overall risk of SCD is 1 per 100,000 population, or 0.001% per year. The risk of SCD increases dramatically beyond age 35 years. The greatest rate of increase is between 40 and 65 years (vertical axis is discontinuous). Among patients older than 30 years of age, with advanced structural heart disease and markers of high risk for cardiac arrest, the event rate may exceed 25% per year, and age-related risk attenuates. *(Modified from Myerburg and Castellanos*

2008, with permission of publisher.)

***Panel B*demonstrates the incidence of SCD in population subgroups** and the relation of total number of events per year to incidence figures. Approximations of subgroup incidence figures and the related population pool from which they are derived are presented. Approximately 50% of all cardiac deaths are sudden and unexpected. The incidence triangle on the left ("Percent/Year") indicates the approximate percentage of sudden and nonsudden deaths in each of the population subgroups indicated, ranging from the lowest percentage in unselected adult populations (0.1–2% per year) to the highest percentage in patients with VT or VF during convalescence after an MI (approximately 50% per year). The triangle on the right indicates the total number of events per year in each of these groups to reflect incidence in context with the size of the population subgroups. The highest risk categories identify the smallest number of total annual events, and the lowest incidence category accounts for the largest number of events per year. EF, ejection fraction; VT, ventricular tachycardia; VF, ventricular fibrillation; MI, myocardial infarction. *(After RJ Myerburg et al: Circulation 85:2, 1992.)*

Young and middle-aged men and women have different susceptibilities to SCD, but the sex differences decrease with advancing age. The difference in risk for SCD parallels the differences in age-related risks for other manifestations of coronary heart disease (CHD) between men and women. As the gender gap for manifestations of CHD closes in the sixth to eighth decades of life, the excess risk of SCD in males progressively narrows. Despite the lower incidence among younger women, coronary risk factors such as cigarette smoking, diabetes, hyperlipidemia, and hypertension are highly influential, and SCD remains an important clinical and epidemiologic problem. The incidence of SCD among the African-American population appears to be higher than it is among the white population; the reasons remain uncertain.

Genetic factors contribute to the risk of acquiring CHD and its expression as acute coronary syndromes, including SCD. In addition, however, there are data suggesting a familial predisposition to SCD as a specific form of expression of CHD. A parental history of SCD as an initial coronary event increases the probability of a similar expression in the offspring. In a number of less common syndromes, such as hypertrophic cardiomyopathy, congenital long QT interval syndromes, right ventricular dysplasia, and the syndrome of right bundle branch block and nonischemic ST-segment elevations (Brugada syndrome), there is a specific inherited risk of ventricular arrhythmias and SCD (Chap. 233).

The structural causes of and functional factors contributing to the SCD syndrome are listed in Table 273-2. Worldwide, and especially in Western cultures, coronary atherosclerotic heart disease is the most common structural abnormality associated with SCD in middle-aged and older adults. Up to 80% of all SCDs in the United States are due to the consequences of coronary atherosclerosis. The nonischemic cardiomyopathies (dilated and hypertrophic, collectively; Chap. 231) account for another 10–15% of SCDs, and all the remaining diverse etiologies cause only 5–10% of all SCDs. The inherited arrhythmia syndromes (see above and Table 273-2) are proportionally more common causes in adolescents and young adults. For some of these syndromes, such as hypertrophic cardiomyopathy (Chap. 238), the risk of SCD increases significantly after the onset of puberty.

Transient ischemia in a previously scarred or hypertrophied heart, hemodynamic and fluid and electrolyte disturbances, fluctuations in autonomic nervous system activity, and transient electrophysiologic changes caused by drugs or other chemicals (e.g., proarrhythmia) have all been implicated as mechanisms responsible for the transition from electrophysiologic stability to instability. In addition, reperfusion of ischemic myocardium may cause transient electrophysiologic instability and arrhythmias.

Pathology

Data from postmortem examinations of SCD victims parallel the clinical observations on the prevalence of CHD as the major structural etiologic factor. More than 80% of SCD victims have pathologic findings of CHD. The pathologic description often includes a combination of long-standing, extensive atherosclerosis of the epicardial coronary arteries and unstable coronary artery lesions, which include various permutations of eroded, fissured, or ruptured plaques; platelet aggregates; hemorrhage; and/or thrombosis. As many as 70–75% of males who die suddenly have preexisting healed MIs, whereas only 20–30% have recent acute MIs, despite the prevalence of unstable plaques and thrombi. The latter suggests transient ischemia as the mechanism of onset. Regional or global left ventricular (LV) hypertrophy often coexists with prior MIs.

PREDICTION AND PREVENTION OF CARDIAC ARREST AND SUDDEN CARDIAC DEATH

SCD accounts for approximately one-half the total number of cardiovascular deaths. As shown in Fig. 273-1*B*, the very high risk subgroups provide more focused populations ("percent per year") for predicting cardiac arrest or SCD, but the representation of such subgroups within the overall population burden of SCD, indicated by the absolute number of events ("events per year"), is relatively small. The requirements for achieving a major population impact are effective prevention of underlying diseases and/or new epidemiologic probes that will allow better resolution of specific high-risk subgroups within the large general populations.

Strategies for predicting and preventing SCD are classified as primary and secondary. *Primary prevention*, as defined in various implantable defibrillator trials, refers to the attempt to identify individual patients at specific risk for SCD and institute preventive strategies. *Secondary prevention* refers to measures taken to prevent recurrent cardiac arrest or death in individuals who have survived a previous cardiac arrest. A third category consists of interventions intended to abort sudden cardiac arrests, thus avoiding their progression to death. This focuses primarily on out-of-hospital response strategies.

The primary prevention strategies currently used depend on the magnitude of risk among the various

population subgroups. Because the annual incidence of SCD among the unselected adult population is limited to 1–2 per 1000 population per year (Fig. 273-1) and >30% of all SCDs due to coronary artery disease occur as the first clinical manifestation of the disease **(Fig. 273-2*A*)**, the only currently practical strategies are profiling for risk of developing CHD and risk factor control (Fig. 273-2*B*). The most powerful long-term risk factors include age, cigarette smoking, elevated serum cholesterol, diabetes mellitus, elevated blood pressure, LV hypertrophy, and nonspecific electrocardiographic abnormalities. Markers of inflammation (e.g., levels of C-reactive protein) that may predict plaque destabilization have been added to risk classifications. The presence of multiple risk factors progressively increases incidence, but not sufficiently or specifically enough to warrant therapies targeted to potentially fatal arrhythmias (Fig. 273-1*A*). However, recent studies offer the hope that genetic markers for specific risk may become available. These studies suggest that a family history of SCD associated with acute coronary syndromes predicts a higher likelihood of cardiac arrest as the initial manifestation of coronary artery disease in first-degree family members.

Figure 273-2

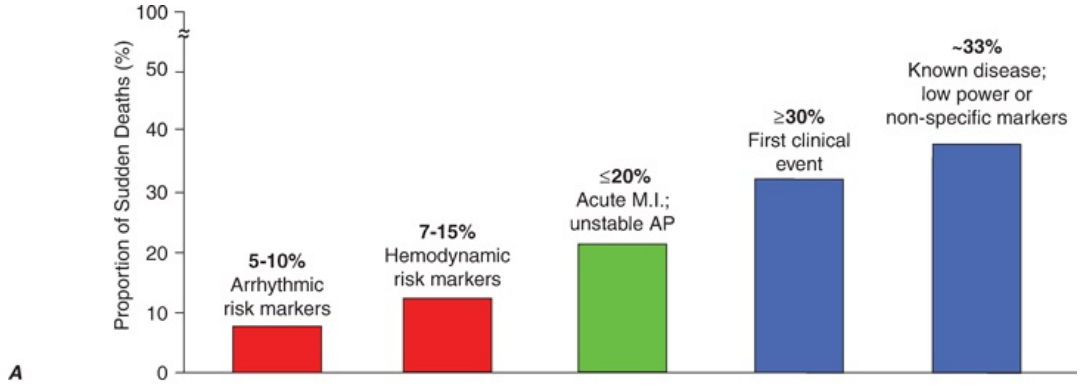

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

	Target	*Examples*	*Goal*	*Sensitivity*
	• ASHD risk factors	• Framingham risk index	• Predict evolution of disease	• Very low
	• Anatomic screening	• CT imaging	• Identify CAD	• Moderate for anatomy
	• Clinical markers	• EF; angiography	• Define extent of disease	• High for extent of disease; variable for specificity of risk
		• AM; EPS	• Identify arrhythmia markers	• Low-to-intermediate for screening
		• History of heart failure	• Define high risk subgroups	• High for specific groups
	• Transient risk predictors	• EP and hemodynamic variations	• Clinical markers of instability	• Primary predictive value unknown
		• Autonomic fluctuations	• Quantify autonomic triggers	• Uncertain; some measures useful (?)
		• Predictors of ischemia	• Predict unstable plaques	• Unknown; potentially high
B	• Individual risk predictors	• Familial/genetic profiles	• Predict specific SCD risk before disease expression	• High potential for future profiling

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Population subsets, risk predictors, and distribution of sudden cardiac deaths (SCDs) according to clinical circumstances. ***A.*** The population subset with high-risk arrhythmia markers in conjunction with low ejection fraction is a group at high risk of SCD but accounts for <10% of the total SCD burden attributable to coronary artery disease. In contrast, nearly two-thirds of all SCD victims present with SCD as the first and only manifestation of underlying disease or have known disease but are considered relatively low risk because of the absence of high-risk markers. ***B.*** Risk profile for prediction and prevention of SCD is difficult. The highest absolute numbers of events occur among the general population who may have risk factors for coronary heart disease or expressions of disease that do not predict high risk. This results in a low sensitivity for predicting and preventing SCD. New approaches that include epidemiologic modeling of transient risk factors and methods of predicting individual patient risk offer hope for greater sensitivity in the future. AP, angina pectoris; ASHD, arteriosclerotic heart disease; CAD, coronary artery disease; EPS, electrophysiologic study; HRV, heart rate variability. *(Modified from Myerburg RJ:J Cardiovasc Electrophysiol 2001; 12:369–381, reproduced with permission of the publisher.)*

After coronary artery disease has been identified in a patient, additional strategies for risk profiling become available (Fig. 273-2*B*), but the majority of SCDs occur among the large unselected groups rather than in the specific high-risk subgroups that become evident among populations with established disease (compare events per year with percent per year in Fig. 273-1*B*). After a major cardiovascular event, such as acute coronary syndromes, recent onset of heart failure, and survival after out-of-hospital cardiac arrest, the highest risk of death occurs during the initial 6–18 months after the event and then plateaus toward the baseline risk associated with the extent of underlying disease. However, many of the early deaths are nonsudden, diluting the potential benefit of strategies targeted specifically to SCD. Thus, although post-MI

beta-blocker therapy has an identifiable benefit for both early SCD and nonsudden mortality risk, a total mortality benefit for ICD therapy early after MI has not been observed.

Among patients in the acute, convalescent, and chronic phases of myocardial infarction (Chap. 245), subgroups at high absolute risk of SCD can be identified. During the acute phase, the potential risk of cardiac arrest from onset through the first 48 h may be as high as 15%, emphasizing the importance for patients to respond promptly to the onset of symptoms. Those who survive acute-phase VF, however, are not at continuing risk for recurrent cardiac arrest indexed to that event. During the convalescent phase after MI (3 days to ~6 weeks), an episode of sustained ventricular tachycardia (VT) or VF, which is usually associated with a large infarct, predicts a natural history mortality risk of >25% at 12 months. At least one-half of the deaths are sudden. Aggressive intervention techniques may reduce this incidence.

After passage into the chronic phase of MI, the longer-term risk for total mortality and SCD mortality is predicted by a number of factors (Fig. 273-2*B*). The most important for both SCD and nonsudden death is the extent of myocardial damage sustained as a result of the acute MI. This is measured by the magnitude of reduction of the ejection fraction (EF) and/or the occurrence of heart failure. Various studies have demonstrated that ventricular arrhythmias identified by ambulatory monitoring contribute significantly to this risk, especially in patients with an EF <40%. In addition, inducibility of VT or VF during electrophysiologic testing of patients who have ambient ventricular arrhythmias [premature ventricular contractions (PVCs) and nonsustained VT] and an EF <35 or 40% is a strong predictor of SCD risk. Patients in this subgroup are now considered candidates for implantable cardioverter defibrillators (ICDs) (see below). Risk falls off sharply with EFs >40% after MI and the absence of ambient arrhythmias and conversely is high with EFs <30% even without the ambient arrhythmia markers.

The cardiomyopathies (dilated and hypertrophic, Chap. 238) are the second most common category of diseases associated with risk of SCD, after CHD (Table 273-2). Some risk factors have been identified, largely related to extent of disease, documented ventricular arrhythmias, and symptoms of arrhythmias (e.g., syncope). The less common causes of SCD include valvular heart disease (primarily aortic) and inflammatory and infiltrative disorders of the myocardium. The latter include viral myocarditis, sarcoidosis, and amyloidosis.

Among adolescents and young adults, rare inherited disorders such as hypertrophic cardiomyopathy, the long QT interval syndromes, right ventricular dysplasia, and the Brugada syndrome have received attention as important causes of SCD because of advances in genetics and the ability to identify some of those at risk before a fatal event. The subgroup of young competitive athletes has received special attention. The incidence of SCD among athletes appears to be higher than it is for the general adolescent and young adult population, perhaps up to 1 in 75,000. Hypertrophic cardiomyopathy (Chap. 238) is the most common cause in the United States, compared with Italy, where more comprehensive screening programs remove potential victims from the population of athletes.

Secondary prevention strategies should be applied to surviving victims of a cardiac arrest that was not associated with an acute MI or a transient risk of SCD (e.g., drug exposures, correctable electrolyte imbalances). Multivessel coronary artery disease and dilated cardiomyopathy, especially with markedly reduced left ventricular EF predict a 1- to 2-year risk of recurrence of an SCD or cardiac arrest of up to 30% in the absence of specific interventions (see below). The presence of life-threatening arrhythmias with long QT syndromes or right ventricular dysplasia are also associated with increased risks.

CLINICAL CHARACTERISTICS OF CARDIAC ARREST

Prodrome, Onset, Arrest, Death

SCD may be presaged by days to months of increasing angina, dyspnea, palpitations, easy fatigability, and other nonspecific complaints. However, these *prodromal symptoms* are generally predictive of any major cardiac event; they are not specific for predicting SCD.

The *onset of the clinical transition* , leading to cardiac arrest, is defined as an acute change in cardiovascular status preceding cardiac arrest by up to 1 h. When the onset is instantaneous or abrupt, the probability that the arrest is cardiac in origin is >95%. Continuous electrocardiographic (ECG) recordings fortuitously obtained at the onset of a cardiac arrest commonly demonstrate changes in cardiac electrical activity during the minutes or hours before the event. There is a tendency for the heart rate to increase and for advanced grades of PVCs to evolve. Most cardiac arrests that are caused by VF begin with a run of nonsustained or sustained VT, which then degenerates into VF.

The probability of achieving successful resuscitation from cardiac arrest is related to the interval from onset of loss of circulation to institution of resuscitative efforts, the setting in which the event occurs, the mechanism (VF, VT, PEA, asystole), and the clinical status of the patient before the cardiac arrest. Return of circulation and survival rates as a result of defibrillation decrease almost linearly from the first minute to 10 min. After 5 min, survival rates are no better than 25–30% in out-of-hospital settings. Those settings in which it is possible to institute prompt cardiopulmonary resuscitation (CPR) followed by prompt defibrillation provide a better chance of a successful outcome. However, the outcome in intensive care units and other in-hospital environments is heavily influenced by the patient's preceding clinical status. The immediate outcome is good for cardiac arrest occurring in the intensive care unit in the presence of an acute cardiac event or transient metabolic disturbance, but survival among patients with far-advanced chronic cardiac disease or advanced noncardiac diseases (e.g., renal failure, pneumonia, sepsis, diabetes, cancer) is low and not much better in the in-hospital than in the out-of-hospital setting. Survival from unexpected cardiac

arrest in unmonitored areas in a hospital is not much better than that it is for witnessed out-of-hospital arrests. Since implementation of community response systems, survival from out-of-hospital cardiac arrest has improved although it still remains low, under most circumstances. Survival probabilities in public sites exceed those in the home environment.

The success rate for initial resuscitation and survival to hospital discharge after an out-of-hospital cardiac arrest depends heavily on the mechanism of the event. When the mechanism is pulseless VT, the outcome is best; VF is the next most successful; and asystole and PEA generate dismal outcome statistics. Advanced age also adversely influences the chances of successful resuscitation.

Progression to biologic death is a function of the mechanism of cardiac arrest and the length of the delay before interventions. VF or asystole without CPR within the first 4–6 min has a poor outcome even if defibrillation is successful because of superimposed brain damage; there are few survivors among patients who had no life support activities for the first 8 min after onset. Outcome statistics are improved by lay bystander intervention (basic life support—see below) before definitive interventions (advanced life support) especially when followed by early successful defibrillation. In regard to the latter, evaluations of deployment of automatic external defibrillators (AEDs) in communities (e.g., police vehicles, large buildings, airports, and stadiums) are beginning to generate encouraging data. Increased deployment is to be encouraged.

Death during the hospitalization after a successfully resuscitated cardiac arrest relates closely to the severity of central nervous system injury. Anoxic encephalopathy and infections subsequent to prolonged respirator dependence account for 60% of the deaths. Another 30% occur as a consequence of low cardiac output states that fail to respond to interventions. Recurrent arrhythmias are the least common cause of death, accounting for only 10% of in-hospital deaths.

In the setting of acute MI (Chap. 245), it is important to distinguish between primary and secondary cardiac arrests. *Primary cardiac arrests* are those which occur in the absence of hemodynamic instability, and *secondary cardiac arrests* are those which occur in patients in whom abnormal hemodynamics dominate the clinical picture before cardiac arrest. The success rate for immediate resuscitation in primary cardiac arrest during acute MI in a monitored setting should exceed 90%. In contrast, as many as 70% of patients with secondary cardiac arrest succumb immediately or during the same hospitalization.

Treatment: Cardiac Arrest

An individual who collapses suddenly is managed in five stages: (1) initial evaluation and basic life support if arrest is confirmed, (2) public access defibrillation (when available), (3) advanced life support, (4) postresuscitation care, and (5) long-term management. The initial response, including confirmation of loss of circulation, followed by basic life support and public access defibrillation, can be carried out by physicians, nurses, paramedical personnel, and trained laypersons. There is a requirement for increasingly specialized skills as the patient moves through the stages of advanced life support, postresuscitation care, and long-term management.

INITIAL EVALUATION AND BASIC LIFE SUPPORT

Confirmation that a sudden collapse is indeed due to a cardiac arrest includes prompt observations of the state of consciousness, respiratory movements, skin color, and the presence or absence of pulses in the carotid or femoral arteries. For lay responders, the pulse check is no longer recommended. As soon as a cardiac arrest is suspected, confirmed, or even considered to be impending, calling an emergency rescue system (e.g., 911) is the immediate priority. With the development of AEDs that are easily used by nonconventional emergency responders, an additional layer for response has evolved (see below).

Agonal respiratory movements may persist for a short time after the onset of cardiac arrest, but it is important to observe for severe stridor with a persistent pulse as a clue to aspiration of a foreign body or food. If this is suspected, a Heimlich maneuver (see below) may dislodge the obstructing body. A precordial blow, or "thump," delivered firmly with a clenched fist to the junction of the middle and lower thirds of the sternum may occasionally revert VT or VF, but there is concern about converting VT to VF. Therefore, it is recommended to use precordial thumps as a life support technique only when monitoring and defibrillation are available. This conservative application of the technique remains controversial.

The third action during the initial response is to clear the airway. The head is tilted back and the chin lifted so that the oropharynx can be explored to clear the airway. Dentures or foreign bodies are removed, and the Heimlich maneuver is performed if there is reason to suspect that a foreign body is lodged in the oropharynx. If respiratory arrest precipitating cardiac arrest is suspected, a second precordial thump is delivered after the airway is cleared.

Basic life support, more popularly known as CPR, is intended to maintain organ perfusion until definitive interventions can be instituted. The elements of CPR are the maintenance of ventilation of the lungs and compression of the chest. Mouth-to-mouth respiration may be used if no specific rescue equipment is immediately available (e.g., plastic oropharyngeal airways, esophageal obturators, masked Ambu bag). Conventional ventilation techniques during single-responder CPR require that the lungs be inflated twice in succession after every 30 chest compressions. Recent data suggest that interrupting chest compressions to perform mouth-to-mouth respiration may be less effective than a continuous chest compression strategy.

Chest compression is based on the assumption that cardiac compression allows the heart to maintain a pump function by sequential filling and emptying of its chambers, with competent valves maintaining forward direction of flow. The palm of one hand is placed over the lower sternum, with the heel of the other

resting on the dorsum of the lower hand. The sternum is depressed, with the arms remaining straight, at a rate of approximately 100 per minute. Sufficient force is applied to depress the sternum 4–5 cm, and relaxation is abrupt.

AUTOMATED EXTERNAL DEFIBRILLATION (AED)

AEDs that are easily used by nonconventional responders, such as nonparamedic firefighters, police officers, ambulance drivers, trained security guards, and minimally trained or untrained laypersons, have been developed. This advance has inserted another level of response into the cardiac arrest paradigm. A number of studies have demonstrated that AED use by nonconventional responders in strategic response systems and public access lay responders can improve cardiac arrest survival rates. This strategy is based on shortening the time to the first defibrillation attempt while awaiting the arrival of advanced life support.

ADVANCED CARDIAC LIFE SUPPORT (ACLS)

ACLS is intended to achieve adequate ventilation, control cardiac arrhythmias, stabilize blood pressure and cardiac output, and restore organ perfusion. The activities carried out to achieve these goals include (1) defibrillation/cardioversion and/or pacing, (2) intubation with an endotracheal tube, and (3) insertion of an intravenous line. The speed with which defibrillation/cardioversion is achieved is an important element in successful resuscitation, both for restoration of spontaneous circulation and for protection of the central nervous system. Immediate defibrillation should precede intubation and insertion of an intravenous line; CPR should be carried out while the defibrillator is being charged. As soon as a diagnosis of VF or VT is established, a shock of at least 300 J should be delivered when one is using a monophasic waveform device or 120–150 J with a biphasic waveform. Additional shocks are escalated to a maximum of 360 J monophasic (200 J biphasic) if the initial shock does not successfully revert VT or VF. However, it is now recommended that five cycles of CPR be carried out before repeated shocks, if the first shock fails to restore an organized rhythm, or 60–90 s of CPR before the first shock if 5 min has elapsed between the onset of cardiac arrest and ability to deliver a shock (see 2005 update of guidelines for cardiopulmonary resuscitation and emergency cardiac care at *http://circ.ahajournals.org/content/vol112/_suppl*).

Epinephrine, 1 mg intravenously, is given after failed defibrillation, and attempts to defibrillate are repeated. The dose of epinephrine may be repeated after intervals of 3–5 min **(Fig. 273-3*A*)**. Vasopressin (a single 40-unit dose given IV) has been suggested as an alternative to epinephrine.

Figure 273-3

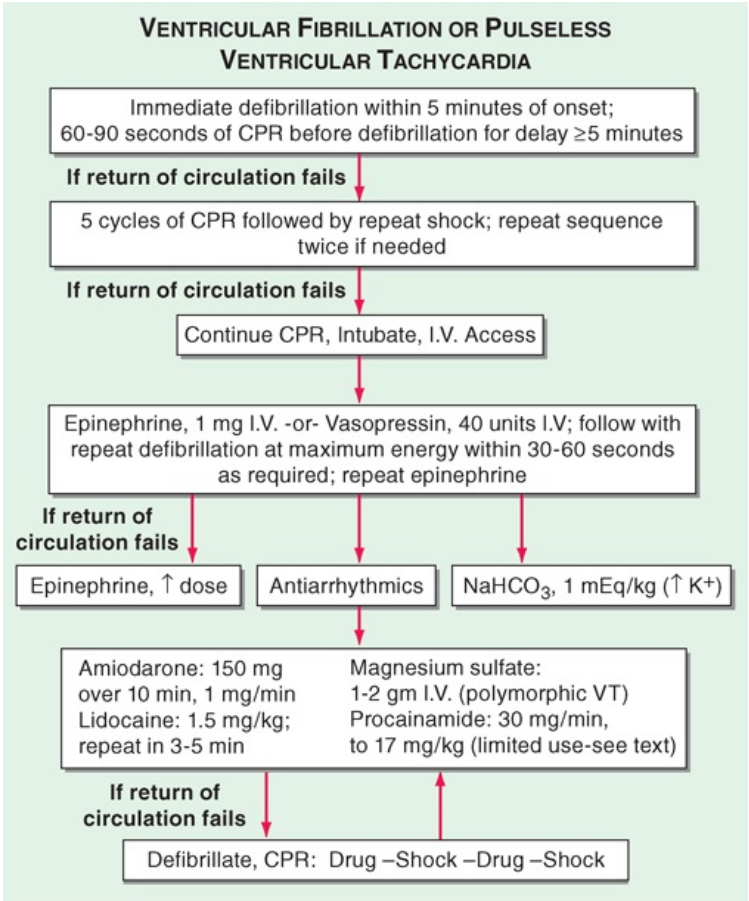

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

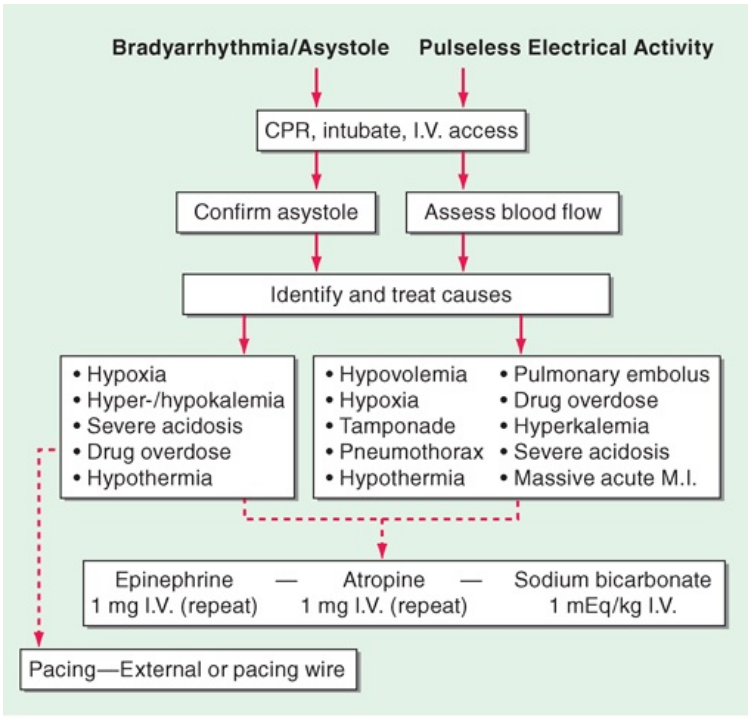

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. **The algorithm of ventricular fibrillation** or pulseless ventricular tachycardia begins with defibrillation attempts. If that fails, it is followed by epinephrine and then antiarrhythmic drugs. See text for details. ***B.*** **The algorithms for bradyarrhythmia/asystole** (*left*) or pulseless electrical activity (*right*) are dominated first by continued life support and a search for reversible causes. Subsequent therapy is nonspecific and is accompanied by a low success rate. See text for details. CPR, cardiopulmonary resuscitation; MI, myocardial infarction.

If the patient is less than fully conscious upon reversion or if two or three attempts fail, prompt intubation, ventilation, and arterial blood gas analysis should be carried out. Ventilation with O_2 (room air if O_2 is not immediately available) may promptly reverse hypoxemia and acidosis. A patient who is persistently acidotic after successful defibrillation and intubation should be given 1 meq/kg NaHCO3 initially and an additional 50% of the dose repeated every 10–15 min. However, it should not be used routinely.

After initial unsuccessful defibrillation attempts or with persistent/recurrent electrical instability, antiarrhythmic therapy should be instituted. Intravenous amiodarone has emerged as the initial treatment of choice (150 mg over 10 min, followed by 1 mg/min for up to 6 h and 0.5 mg/min thereafter) (Fig. 273-3*A*). For cardiac arrest due to VF in the early phase of an acute coronary syndrome, a bolus of 1 mg/kg of lidocaine may be given intravenously as an alternative, and the dose may be repeated in 2 min. It also may be tried in patients in whom amiodarone is unsuccessful. Intravenous procainamide (loading infusion of 100 mg/5 min to a total dose of 500–800 mg, followed by continuous infusion at 2–5 mg/min) is now rarely used in this setting but may be tried for persisting, hemodynamically stable arrhythmias. Intravenous calcium gluconate is no longer considered safe or necessary for routine administration. It is used only in patients in whom acute hyperkalemia is known to be the triggering event for resistant VF, in the presence of known hypocalcemia, or in patients who have received toxic doses of calcium channel antagonists.

Cardiac arrest due to bradyarrhythmias or asystole (B/A cardiac arrest) is managed differently (Fig. 273-3*B*). The patient is promptly intubated, CPR is continued, and an attempt is made to control hypoxemia and acidosis. Epinephrine and/or atropine are given intravenously or by an intracardiac route. External pacing devices are used to attempt to establish a regular rhythm. The success rate may be good when B/A arrest is due to acute inferior wall myocardial infarction or to correctable airway obstruction or drug-induced respiratory depression or with prompt resuscitation efforts. For acute airway obstruction, prompt removal of foreign bodies by the Heimlich maneuver or, in hospitalized patients, by intubation and suctioning of obstructing secretions in the airway is often successful. The prognosis is generally very poor in other causes of this form of cardiac arrest, such as end-stage cardiac or noncardiac diseases. Treatment of PEA is similar to that for bradyarrhythmias, but its outcome is also dismal.

POSTRESUSCITATION CARE

This phase of management is determined by the clinical setting of the cardiac arrest. *Primary VF* in acute MI (not accompanied by low-output states) (Chap. 245) is generally very responsive to life support techniques and easily controlled after the initial event. In the in-hospital setting, respirator support is usually not necessary or is needed for only a short time, and hemodynamics stabilize promptly after defibrillation or cardioversion. In *secondary VF* in acute MI (those events in which hemodynamic abnormalities predispose to the potentially fatal arrhythmia), resuscitative efforts are less often successful, and in patients who are

successfully resuscitated, the recurrence rate is high. The clinical picture and outcome are dominated by hemodynamic instability and the ability to control hemodynamic dysfunction. Bradyarrhythmias, asystole, and PEA are commonly secondary events in hemodynamically unstable patients. The in-hospital phase of care of an out-of-hospital cardiac arrest survivor is dictated by specific clinical circumstances. The most difficult is the presence of anoxic encephalopathy, which is a strong predictor of in-hospital death. A recent addition to the management of this condition is induced hypothermia to reduce metabolic demands and cerebral edema.

The outcome after in-hospital cardiac arrest associated with noncardiac diseases is poor, and in the few successfully resuscitated patients, the postresuscitation course is dominated by the nature of the underlying disease. Patients with end-stage cancer, renal failure, acute central nervous system disease, and uncontrolled infections, as a group, have a survival rate of <10% after in-hospital cardiac arrest. Some major exceptions are patients with transient airway obstruction, electrolyte disturbances, proarrhythmic effects of drugs, and severe metabolic abnormalities, most of whom may have a good chance of survival if they can be resuscitated promptly and stabilized while the transient abnormalities are being corrected.

LONG-TERM MANAGEMENT AFTER SURVIVAL OF OUT-OF-HOSPITAL CARDIAC ARREST

Patients who survive cardiac arrest without irreversible damage to the central nervous system and who achieve hemodynamic stability should have diagnostic testing to define appropriate therapeutic interventions for their long-term management. This aggressive approach is driven by the fact that survival after out-of-hospital cardiac arrest is followed by a 10–25% mortality rate during the first 2 years after the event, and there are data suggesting that significant survival benefits can be achieved by prescription of an implantable cardioverter-defibrillator (ICD).

Among patients in whom an acute ST elevation MI, or transient and reversible myocardial ischemia, is identified as the specific mechanism triggering an out-of-hospital cardiac arrest, the management is dictated in part by the transient nature of life-threatening arrhythmia risk during the acute coronary syndrome (ACS) and in part by the extent of permanent myocardial damage that results. Cardiac arrest during the acute ischemic phase is not an ICD indication, but survivors of cardiac arrest not associated with an ACS do benefit. In addition, patients who survive MI with an ejection fraction less than 30–35% appear to benefit from ICDs.

For patients with cardiac arrest determined to be due to a treatable transient ischemic mechanism, particularly with higher EFs, catheter interventional, surgical, and/or pharmacologic anti-ischemic therapy is generally accepted for long-term management.

Survivors of cardiac arrest due to other categories of disease, such as the hypertrophic or dilated cardiomyopathies and the various rare inherited disorders (e.g., right ventricular dysplasia, long QT syndrome, Brugada syndrome, catecholaminergic polymorphic VT, and so-called idiopathic VF), are all considered ICD candidates.

PREVENTION OF SCD IN HIGH-RISK INDIVIDUALS WITHOUT PRIOR CARDIAC ARREST

Post-MI patients with EFs <35% and other markers of risk such as ambient ventricular arrhythmias, inducible ventricular tachyarrhythmias in the electrophysiology laboratory, and a history of heart failure are considered candidates for ICDs 30 days or more after the MI. Total mortality benefits in the range of a 20–35% reduction over 2–5 years have been observed in a series of clinical trials. One study suggested that an EF <30% was a sufficient marker of risk to indicate ICD benefit, and another demonstrated benefit for patients with Functional Class 2 or 3 heart failure and ejection fractions ≤35%, regardless of etiology (ischemic or nonischemic) or the presence of ambient or induced arrhythmias (see Chaps. 233 and 234). There appears to be a gradient of increasing ICD benefit with EFs ranging lower than the threshold indications. However, patients with very low EFs (e.g., <20%) may receive less benefit.

Decision making for primary prevention in disorders other than coronary artery disease and dilated cardiomyopathy is generally driven by observational data and judgment based on clinical observations. Controlled clinical trials providing evidence-based indicators for ICDs are lacking for these smaller population subgroups. In general, for the rare disorders listed above, indicators of arrhythmic risk such as syncope, documented ventricular tachyarrhythmias, aborted cardiac arrest or a family history of premature SCD in some conditions, and a number of other clinical or ECG markers may be used as indicators for ICDs.

FURTHER READINGS

Huikuri H et al: Sudden death due to cardiac arrhythmias. N Engl J Med 345:1473, 2001[PMID: 11794197]

International Liaison Committee on Resuscitation: 2005 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science with Treatment Recommendations. Circulation 112(Suppl III):III-1–III-136, 2005, *http://circ.ahajournals.org/content/vol112/24_suppl/*

Kokolis S et al: Ventricular arrhythmias and sudden cardiac death. Prog Cardiovasc Dis 48:426, 2006[PMID: 16714162]

Marenco JP et al: Improving survival from sudden cardiac arrest: The role of the automated external defibrillator. JAMA 285:1193, 2001[PMID: 11231750]

Maron BJ, Pelliccia A: The heart of trained athletes: Cardiac remodeling and the risks of sports, including sudden death. Circulation 114:1633, 2006[PMID: 17030703]

Myerburg RJ, Castellanos A: Cardiac arrest and sudden cardiac death, in *Braunwald's Heart Disease*, 8th ed, P Libby et al (eds). Philadelphia, Saunders, 2008

——— et al: Indications for implantable cardioverter-defibrillators based on evidence and judgment. J Am Coll Cardiol 54:747, 2009

Noseworthy PA, Newton-Cheh C: Genetic determinants of sudden cardiac death. Circulation 118:1854, 2008[PMID: 18955676]

Wik L et al: Delaying defibrillation to give basic cardiopulmonary resuscitation to patients with out-of-hospital ventricular fibrillation: A randomized trial. JAMA 289:1389, 2003[PMID: 12636461]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

COMA: INTRODUCTION

Coma is among the most common and striking problems in general medicine. It accounts for a substantial portion of admissions to emergency wards and occurs on all hospital services. Coma demands immediate attention and requires an organized approach.

There is a continuum of states of reduced alertness, the most severe form being *coma*, defined as a deep sleeplike state from which the patient cannot be aroused. *Stupor* refers to a higher degree of arousability in which the patient can be transiently awakened only by vigorous stimuli, accompanied by motor behavior that leads to avoidance of uncomfortable or aggravating stimuli. *Drowsiness*, which is familiar to all persons, simulates light sleep and is characterized by easy arousal and the persistence of alertness for brief periods. Drowsiness and stupor are usually accompanied by some degree of confusion (Chap. 25). A precise narrative description of the level of arousal and of the type of responses evoked by various stimuli as observed at the bedside is preferable to ambiguous terms such as *lethargy*, *semicoma*, or *obtundation*.

Several other conditions that render patients unresponsive and thereby simulate coma are considered separately because of their special significance. The *vegetative state* signifies an awake but nonresponsive state in a patient who has emerged from coma. In the vegetative state, the eyelids may open, giving the appearance of wakefulness. Respiratory and autonomic functions are retained. Yawning, coughing, swallowing, as well as limb and head movements persist and the patient may follow visually presented objects, but there are few, if any, meaningful responses to the external and internal environment—in essence, an "awake coma." The term *vegetative* is unfortunate as it is subject to misinterpretation. There are always accompanying signs that indicate extensive damage in both cerebral hemispheres, e.g., decerebrate or decorticate limb posturing and absent responses to visual stimuli (see below). In the closely related but less severe *minimally conscious state*, the patient has rudimentary vocal or motor behaviors, often spontaneous, but some in response to touch, visual stimuli, or command. Cardiac arrest with cerebral hypoperfusion and head injuries are the most common causes of the vegetative and minimally conscious states (Chaps. 273 and 275). The prognosis for regaining mental faculties once the vegetative state has supervened for several months is very poor, and after a year, almost nil, hence the term *persistent vegetative state*. Most reports of dramatic recovery, when investigated carefully, are found to yield to the usual rules for prognosis but there have been rare instances in which recovery has occurred to a severely disabled condition and, in rare childhood cases, to an even better state. The possibility of incorrectly attributing meaningful behavior to patients in the vegetative and minimally conscious states has created inordinate problems and anguish for families. On the other hand, the question of whether these patients lack any capability for cognition has been reopened by functional imaging studies demonstrating, in a small proportion of posttraumatic cases, cerebral activation in response to external stimuli.

Apart from the above conditions, several syndromes that affect alertness are prone to be misinterpreted as stupor or coma. *Akinetic mutism* refers to a partially or fully awake state in which the patient is able to form impressions and think, as demonstrated by later recounting of events, but remains virtually immobile and mute. The condition results from damage in the regions of the medial thalamic nuclei or the frontal lobes (particularly lesions situated deeply or on the orbitofrontal surfaces) or from extreme hydrocephalus. The term *abulia* describes a milder form of akinetic mutism characterized by mental and physical slowness and diminished ability to initiate activity. It is also usually the result of damage to the frontal lobes and its connections (Chap. 26). *Catatonia* is a curious hypomobile and mute syndrome that occurs as part of a major psychosis, usually schizophrenia or major depression. Catatonic patients make few voluntary or responsive movements, although they blink, swallow, and may not appear distressed. There are nonetheless signs that the patient is responsive, although it may take ingenuity on the part of the examiner to demonstrate them. For example, eyelid elevation is actively resisted, blinking occurs in response to a visual threat, and the eyes move concomitantly with head rotation, all of which are inconsistent with the presence of a brain lesion causing unresponsiveness. It is characteristic but not invariable in catatonia for the limbs to retain the postures in which they have been placed by the examiner ("waxy flexibility," or catalepsy). With recovery, patients often have some memory of events that occurred during their catatonic stupor. Catatonia is superficially similar to akinetic mutism, but clinical evidence of cerebral damage such as Babinski signs and hypertonicity of the limbs is lacking. The special problem of coma in brain death is discussed below.

The *locked-in state* describes yet another type of pseudocoma in which an awake patient has no means of producing speech or volitional movement but retains voluntary vertical eye movements and lid elevation, thus allowing the patient to signal with a clear mind. The pupils are normally reactive. Such individuals have written entire treatises using Morse code. The usual cause is an infarction or hemorrhage of the ventral pons that transects all descending motor (corticospinal and corticobulbar) pathways. A similar awake but de-efferented state occurs as a result of total paralysis of the musculature in severe cases of Guillain-Barré syndrome (Chap. 385), critical illness neuropathy (Chap. 275), and pharmacologic neuromuscular blockade.

The Anatomy and Physiology of Coma

Almost all instances of diminished alertness can be traced to widespread abnormalities of the cerebral hemispheres or to reduced activity of a special thalamocortical alerting system termed the *reticular activating system* (RAS). The proper functioning of this system, its ascending projections to the cortex, and the cortex itself are required to maintain alertness and coherence of thought. It follows that the principal causes of coma are (1) lesions that damage the RAS in the upper midbrain or its projections; (2) destruction of large portions of both cerebral hemispheres; and (3) suppression of reticulocerebral function by drugs, toxins, or metabolic derangements such as hypoglycemia, anoxia, uremia, and hepatic failure.

The proximity of the RAS to midbrain structures that control pupillary function and eye movements permits clinical localization of the cause of coma in many cases. Pupillary enlargement with loss of light reaction and loss of vertical and adduction movements of the eyes suggests that the lesion is in the upper brainstem. Conversely, preservation of pupillary light reactivity and of eye movements absolves the upper brainstem and indicates that widespread structural lesions or metabolic suppression of the cerebral hemispheres is responsible for coma.

COMA DUE TO CEREBRAL MASS LESIONS AND HERNIATIONS

The cranial cavity is separated into compartments by infoldings of the dura. The two cerebral hemispheres are separated by the falx, and the anterior and posterior fossae by the tentorium. Herniation refers to displacement of brain tissue into a compartment that it normally does not occupy. Coma and many of its associated signs can be attributed to these tissue shifts, and certain clinical features are characteristic of specific herniations **(Fig. 274-1)**. They are in essence "false localizing" signs since they derive from compression of brain structures at a distance from the mass.

Figure 274-1

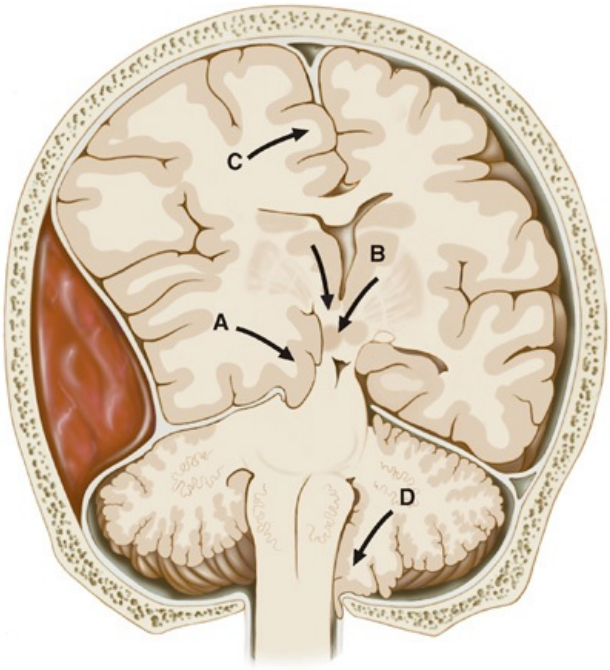

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Types of cerebral herniation. (***A***) uncal; (***B***) central; (***C***) transfalcial; (***D***) foraminal.

The most common herniations are those in which a part of the brain is displaced from the supratentorial to the infratentorial compartment through the tentorial opening; this is referred to as

transtentorial herniation. *Uncal transtentorial* herniation refers to impaction of the anterior medial temporal gyrus (the uncus) into the tentorial opening just anterior to and adjacent to the midbrain (Fig. 274-1A). The uncus compresses the third nerve as it traverses the subarachnoid space, causing enlargement of the ipsilateral pupil (putatively because the fibers subserving parasympathetic pupillary function are located peripherally in the nerve). The coma that follows is due to compression of the midbrain against the opposite tentorial edge by the displaced parahippocampal gyrus **(Fig. 274-2)**. Lateral displacement of the midbrain may compress the opposite cerebral peduncle, producing a Babinski's sign and hemiparesis contralateral to the original hemiparesis (the Kernohan-Woltman sign). Herniation may also compress the anterior and posterior cerebral arteries as they pass over the tentorial reflections, with resultant brain infarction. The distortions may also entrap portions of the ventricular system, resulting in hydrocephalus.

Figure 274-2

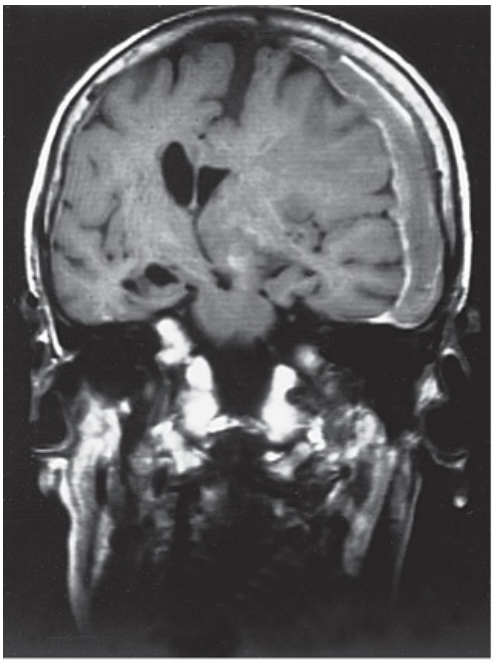

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

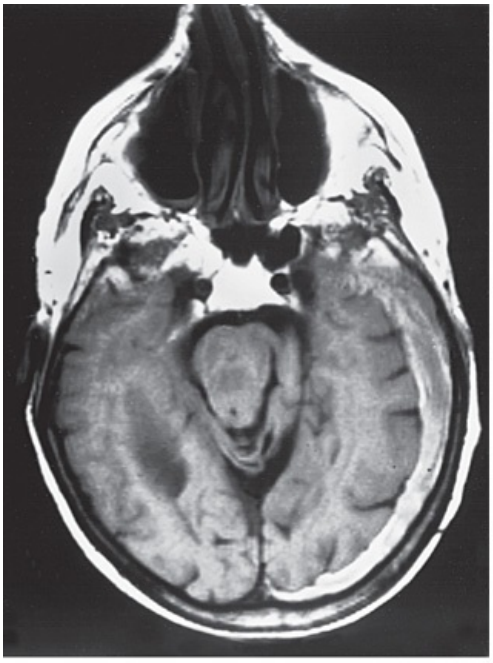

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronal (*A*) and axial (*B*) magnetic resonance images from a stuporous patient with a left third nerve palsy as a result of a large left-sided subdural hematoma (seen as a gray-white rim). The upper midbrain and lower thalamic regions are compressed and displaced horizontally away from the mass, and there is transtentorial herniation of the medial temporal lobe structures, including the uncus anteriorly. The lateral ventricle opposite to the hematoma has become enlarged as a result of compression of the third ventricle.

Central transtentorial herniation denotes a symmetric downward movement of the thalamic medial structures through the tentorial opening with compression of the upper midbrain (Fig. 274-1B). Miotic pupils and drowsiness are the heralding signs. Both temporal and central transtentorial herniations have been considered causes of progressive compression of the brainstem, with initial damage to the midbrain, then the pons, and finally the medulla. The result is a sequence of neurologic signs that corresponds to each affected level. Other forms of herniation are *transfalcial herniation* (displacement of the cingulate gyrus under the falx and across the midline, Fig. 274-1C), and *foraminal herniation* (downward forcing of the cerebellar tonsils into the foramen magnum, Fig. 274-1D), which causes compression of the medulla, respiratory arrest, and death.

A direct relationship between the various configurations of transtentorial herniation and coma is not always found. Drowsiness and stupor can occur with moderate horizontal displacement of the diencephalon (thalamus), before transtentorial herniation is evident. This lateral shift may be quantified on axial images of CT and MRI scans (Fig. 274-2). In cases of *acutely appearing masses*, horizontal displacement of the pineal calcification of 3–5 mm is generally associated with drowsiness, 6–8 mm with stupor, and >9 mm with coma. Intrusion of the medial temporal lobe into the tentorial opening is also apparent on MRI and CT scans as obliteration of the cisterna that surround the upper brainstem.

COMA DUE TO METABOLIC DISORDERS

Many systemic metabolic abnormalities cause coma by interrupting the delivery of energy substrates (e.g., hypoxia, ischemia, hypoglycemia) or by altering neuronal excitability (drug and alcohol intoxication, anesthesia, and epilepsy). The same metabolic abnormalities that produce coma may, in milder forms, induce an acute confusional state. Thus, in metabolic encephalopathies, clouded consciousness and coma are in a continuum.

Cerebral neurons are fully dependent on cerebral blood flow (CBF) and the delivery of oxygen and glucose. CBF is ~75 mL per 100g/min in gray matter and 30 mL per 100 g/min in white matter (mean ~ 55 mL per 100 g/min); oxygen consumption is 3.5 mL per 100 g/min, and glucose utilization is 5 mg per 100 g/min. Brain stores of glucose provide energy for ~2 minutes after blood flow is interrupted, and oxygen stores last 8–10 seconds after the cessation of blood flow. Simultaneous hypoxia and ischemia exhaust glucose more rapidly. The electroencephalogram (EEG) rhythm in these circumstances becomes diffusely slowed, typical of metabolic encephalopathies, and as conditions of substrate delivery worsen, eventually brain electrical activity ceases.

Unlike hypoxia-ischemia, which causes neuronal destruction, most metabolic disorders such as hypoglycemia, hyponatremia, hyperosmolarity, hypercapnia, hypercalcemia, and hepatic and renal

failure cause only minor neuropathologic changes. The causes of the reversible effects of these conditions on the brain are not understood but may result from impaired energy supplies, changes in ion fluxes across neuronal membranes, and neurotransmitter abnormalities. For example, the high ammonia concentration of hepatic coma interferes with cerebral energy metabolism and with the Na^+, K^+-ATPase pump, increases the number and size of astrocytes, and causes increased concentrations of potentially toxic products of ammonia metabolism; it may also affect neurotransmitters, including the production of putative "false" neurotransmitters that are active at receptor sites. Apart from hyperammonemia, which of these mechanisms is of critical importance is not clear. The mechanism of the encephalopathy of renal failure is also not known. Unlike ammonia, urea does not produce central nervous system (CNS) toxicity and a multifactorial causation has been proposed for the encephalopathy, including increased permeability of the blood-brain barrier to toxic substances such as organic acids and an increase in brain calcium and cerebrospinal fluid (CSF) phosphate content.

Coma and seizures are common accompaniments of large shifts in sodium and water balance in the brain. These changes in osmolarity arise from systemic medical disorders, including diabetic ketoacidosis, the nonketotic hyperosmolar state, and hyponatremia from any cause (e.g., water intoxication, excessive secretion of antidiuretic hormone, or atrial natriuretic peptides). Sodium levels <125 mmol/L induce confusion, and <115 mmol/L are associated with coma and convulsions. In hyperosmolar coma, the serum osmolarity is generally >350 mosmol/L. Hypercapnia depresses the level of consciousness in proportion to the rise in carbon dioxide (co_2) tension in the blood. *In all of these metabolic encephalopathies, the degree of neurologic change depends to a large extent on the rapidity with which the serum changes occur*. The pathophysiology of other metabolic encephalopathies such as hypercalcemia, hypothyroidism, vitamin B12 deficiency, and hypothermia are incompletely understood but must also reflect derangements of CNS biochemistry, membrane function, and neurotransmitters.

EPILEPTIC COMA

Generalized electrical discharges of the cortex (*seizures*) are associated with coma, even in the absence of epileptic motor activity (*convulsions*). The self-limited coma that follows a seizure, the *postictal state*, may be due to exhaustion of energy reserves or effects of locally toxic molecules that are the by-product of seizures. The postictal state produces a pattern of continuous, generalized slowing of the background EEG activity similar to that of other metabolic encephalopathies.

TOXIC (INCLUDING DRUG–INDUCED) COMA

This common class of encephalopathy is in large measure reversible and leaves no residual damage provided there has not been cardiorespiratory failure. Many drugs and toxins are capable of depressing nervous system function. Some produce coma by affecting both the brainstem nuclei, including the RAS, and the cerebral cortex. The combination of cortical and brainstem signs, which occurs in certain drug overdoses, may lead to an incorrect diagnosis of structural brainstem disease. Overdose of medications that have atropinic actions produces signs such as dilated pupils, tachycardia, and dry skin; opiate overdose produces pinpoint pupils <1 mm in diameter.

COMA DUE TO WIDESPREAD DAMAGE TO THE CEREBRAL HEMISPHERES

This category, comprising a number of unrelated disorders, results from widespread structural cerebral damage, thereby simulating a metabolic disorder of the cortex. Hypoxia-ischemia is perhaps the best known and one in which it is not possible initially to distinguish the acute reversible effects of hypoperfusion and oxygen deprivation of the brain from the subsequent effects of neuronal damage. Similar bihemispheral damage is produced by disorders that occlude small blood vessels throughout the brain; examples include cerebral malaria, thrombotic thrombocytopenic purpura, and hyperviscosity. Diffuse white matter damage from cranial trauma or inflammatory demyelinating diseases causes a similar syndrome of coma.

Approach to the Patient: Coma

Acute respiratory and cardiovascular problems should be attended to prior to neurologic assessment. In most instances, a complete medical evaluation, except for vital signs, funduscopy, and examination for nuchal rigidity, may be deferred until the neurologic evaluation has established the severity and nature of coma. The approach to the patient with coma from cranial trauma is discussed in Chap. 378.

HISTORY

In many cases, the cause of coma is immediately evident (e.g., trauma, cardiac arrest, or reported drug ingestion). In the remainder, certain points are especially useful: (1) the circumstances and rapidity with which neurologic symptoms developed; (2) the antecedent symptoms (confusion, weakness, headache, fever, seizures, dizziness, double vision, or vomiting); (3) the use of medications, illicit drugs, or alcohol; and (4) chronic liver, kidney, lung, heart, or other medical disease. Direct interrogation of family, observers, and ambulance technicians on the scene, in person or by telephone, is an important part of the evaluation.

GENERAL PHYSICAL EXAMINATION

Fever suggests a systemic infection, bacterial meningitis, encephalitis, heat stroke, neuroleptic malignant syndrome, malignant hyperthermia due to anesthetics or anticholinergic drug intoxication; only rarely is it attributable to a lesion that has disturbed hypothalamic temperature-regulating centers ("*central fever*"). A slight elevation in temperature may follow vigorous convulsions. Hypothermia is observed with exposure; alcoholic, barbiturate, sedative, or phenothiazine intoxication; hypoglycemia; peripheral circulatory failure; or extreme hypothyroidism. Hypothermia itself causes coma only when the temperature is <31°C (87.8°F). Tachypnea may indicate systemic acidosis or pneumonia or rarely infiltration of the brain with lymphoma. Aberrant respiratory patterns that reflect brainstem disorders are discussed below. Marked hypertension suggests hypertensive encephalopathy, but it may also be secondary to a rapid rise in intracranial pressure (ICP) (the Cushing response) most often after cerebral hemorrhage or head injury. Hypotension is characteristic of coma from alcohol or barbiturate intoxication, internal hemorrhage, myocardial infarction, sepsis, profound hypothyroidism, or Addisonian crisis.

The funduscopic examination can detect subarachnoid hemorrhage (subhyaloid hemorrhages), hypertensive encephalopathy (exudates, hemorrhages, vessel-crossing changes, papilledema), and increased ICP (papilledema). Cutaneouspetechiae suggest thrombotic thrombocytopenic purpura, meningococcemia, or a bleeding diathesis associated with an intracerebral hemorrhage. Cyanosis, reddish or anemic skin coloration are other indications of an underlying systemic disease responsible for the coma.

NEUROLOGIC EXAMINATION

The patient should first be observed without intervention by the examiner. Tossing about in the bed, reaching up toward the face, crossing legs, yawning, swallowing, coughing, or moaning reflect a drowsy state that is close to normal awakeness. Lack of restless movements on one side or an outturned leg suggests a hemiplegia. Intermittent twitching movements of a foot, finger, or facial muscle may be the only sign of seizures. Multifocal myoclonus almost always indicates a metabolic disorder, particularly uremia, anoxia, drug intoxication (especially with lithium or haloperidol), or a prion disease (Chap. 383). In a drowsy and confused patient, bilateral *asterixis* is a certain sign of metabolic encephalopathy or drug intoxication.

Decorticate rigidity and *decerebrate rigidity*, or "posturing," describe stereotyped arm and leg movements occurring spontaneously or elicited by sensory stimulation. Flexion of the elbows and wrists and supination of the arm (decortication) suggests bilateral damage rostral to the midbrain, whereas extension of the elbows and wrists with pronation (decerebration) indicates damage to motor tracts in the midbrain or caudal diencephalon. The less frequent combination of arm extension with leg flexion or flaccid legs is associated with lesions in the pons. These concepts have been adapted from animal work and cannot be applied with precision to coma in humans. In fact, acute and widespread disorders of any type, regardless of location, frequently cause limb extension, and almost all extensor posturing becomes predominantly flexor as time passes.

LEVEL OF AROUSAL

A sequence of increasingly intense stimuli is used to determine the threshold for arousal and the motor response of each side of the body. The results of testing may vary from minute to minute, and serial examinations are most useful. Tickling the nostrils with a cotton wisp is a moderate stimulus to arousal—all but deeply stuporous and comatose patients will move the head away and arouse to some degree. An even greater degree of responsiveness is present if the patient uses his hand to remove an offending stimulus. Pressure on the knuckles or bony prominences and pinprick stimulation are humane forms of noxious stimuli; pinching the skin causes unsightly ecchymoses and is generally not necessary but may be useful in eliciting abduction withdrawal movements of the limbs. Posturing in response to noxious stimuli indicates severe damage to the corticospinal system, whereas abduction-avoidance movement of a limb is usually purposeful and denotes an intact corticospinal system. Posturing may also be unilateral and coexists with purposeful limb movements, reflecting incomplete damage to the motor system.

BRAINSTEM REFLEXES

Assessment of brainstem function is essential to localization of the lesion in coma **(Fig. 274-3)**. The brainstem reflexes that are conveniently examined are pupillary size and reaction to light, spontaneous and elicited eye movements, corneal responses, and the respiratory pattern. As a rule, coma is due to bilateral hemispheral disease when these brainstem activities are preserved, particularly the pupillary reactions and eye movements. However, the presence of abnormal brainstem signs does not always indicate that the primary lesion is in the brainstem because hemispheral masses can cause secondary brainstem pathology by transtentorial herniation.

Figure 274-3

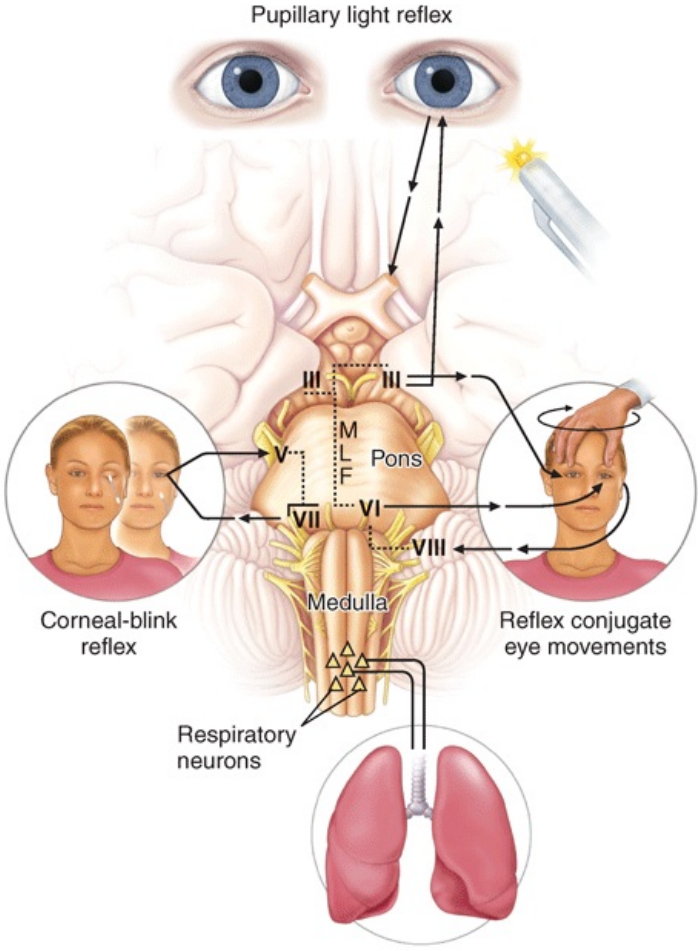

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examination of brainstem reflexes in coma. Midbrain and third nerve function are tested by pupillary reaction to light, pontine function by spontaneous and reflex eye movements and corneal responses, and medullary function by respiratory and pharyngeal responses. Reflex conjugate, horizontal eye movements are dependent on the medial longitudinal fasciculus (MLF) interconnecting the sixth and contralateral third nerve nuclei. Head rotation (oculocephalic reflex) or caloric stimulation of the labyrinths (oculovestibular reflex) elicits contraversive eye movements (for details see text).

Pupillary Signs

Pupillary reactions are examined with a bright, diffuse light (not an ophthalmoscope). Reactive and round pupils of midsize (2.5–5 mm) essentially exclude midbrain damage, either primary or secondary to compression. A response to light may be difficult to appreciate in pupils <2 mm in diameter, and bright room lighting mutes pupillary reactivity. One enlarged and poorly reactive pupil (>6 mm) signifies compression or stretching of the third nerve from the effects of a cerebral mass above. Enlargement of the pupil contralateral to a hemispheral mass may occur but is infrequent. An oval and slightly eccentric pupil is a transitional sign that accompanies early midbrain–third nerve compression. The most extreme pupillary sign, bilaterally dilated and unreactive pupils, indicates severe midbrain damage, usually from compression by a supratentorial mass. Ingestion of drugs with anticholinergic activity, the use of mydriatic eye drops, and direct ocular trauma are among the causes of misleading pupillary enlargement.

Unilateral miosis in coma has been attributed to dysfunction of sympathetic efferents originating in the posterior hypothalamus and descending in the tegmentum of the brainstem to the cervical cord. It is therefore of limited localizing value but is an occasional finding in patients with a large cerebral hemorrhage that affects the thalamus. Reactive and bilaterally small (1–2.5 mm) but not pinpoint pupils are seen in metabolic encephalopathies or in deep bilateral hemispheral lesions such as hydrocephalus or thalamic hemorrhage. Even smaller reactive pupils (<1 mm) characterize narcotic or barbiturate overdoses but also occur with extensive pontine hemorrhage. The response to naloxone and the presence of reflex eye movements (see below) assist in distinguishing these.

Ocular Movements

The eyes are first observed by elevating the lids and noting the resting position and spontaneous movements of the globes. Lid tone, tested by lifting the eyelids and noting their resistance to opening and the speed of closure, is progressively reduced as unresponsiveness progresses. Horizontal

divergence of the eyes at rest is normal in drowsiness. As coma deepens, the ocular axes may become parallel again.

Spontaneous eye movements in coma often take the form of conjugate horizontal roving. This finding alone exonerates damage in the midbrain and pons and has the same significance as normal reflex eye movements (see below). Conjugate horizontal ocular deviation to one side indicates damage to the pons on the opposite side or alternatively, to the frontal lobe on the same side. This phenomenon is summarized by the following maxim: *The eyes look toward a hemispheral lesion and away from a brainstem lesion.* Seizures also drive the eyes to one side but usually with superimposed clonic movements of the globes. The eyes may occasionally turn paradoxically away from the side of a deep hemispheral lesion ("wrong-way eyes"). The eyes turn down and inward with thalamic and upper midbrain lesions, typically thalamic hemorrhage. "Ocular bobbing" describes brisk downward and slow upward movements of the eyes associated with loss of horizontal eye movements and is diagnostic of bilateral pontine damage, usually from thrombosis of the basilar artery. "Ocular dipping" is a slower, arrhythmic downward movement followed by a faster upward movement in patients with normal reflex horizontal gaze; it indicates diffuse cortical anoxic damage.

The oculocephalic reflexes, elicited by moving the head from side to side or vertically and observing eye movements in the direction opposite to the head movement, depend on the integrity of the ocular motor nuclei and their interconnecting tracts that extend from the midbrain to the pons and medulla (Fig. 274-3). The movements, called somewhat inappropriately "doll's eyes" (which refers more accurately to the reflex elevation of the eyelids with flexion of the neck), are normally suppressed in the awake patient. The ability to elicit them therefore indicates both reduced cortical influence on the brainstem and intact brainstem pathways, indicating that coma is caused by a lesion or dysfunction in the cerebral hemispheres. The opposite, an absence of reflex eye movements, usually signifies damage within the brainstem but can result infrequently from overdoses of certain drugs. Normal pupillary size and light reaction distinguishes most drug-induced comas from structural brainstem damage.

Thermal, or "caloric," stimulation of the vestibular apparatus (oculovestibular response) provides a more intense stimulus for the oculocephalic reflex but provides essentially the same information. The test is performed by irrigating the external auditory canal with cool water in order to induce convection currents in the labyrinths. After a brief latency, the result is tonic deviation of both eyes to the side of cool-water irrigation and nystagmus in the opposite direction. (The acronym "COWS" has been used to remind generations of medical students of the direction of nystagmus—"*c*old water *o*pposite, *w*arm water *s*ame.") The loss of induced conjugate ocular movements indicates brainstem damage. The presence of corrective nystagmus indicates that the frontal lobes are functioning and connected to the brainstem; thus functional or hysterical coma is likely.

By touching the cornea with a wisp of cotton, a response consisting of brief bilateral lid closure is normally observed. The corneal reflex depends on the integrity of pontine pathways between the fifth (afferent) and both seventh (efferent) cranial nerves; in conjunction with reflex eye movements, it is a useful test of pontine function. CNS-depressant drugs diminish or eliminate the corneal responses soon after reflex eye movements are paralyzed but before the pupils become unreactive to light. The corneal (and pharyngeal) response may be lost for a time on the side of an acute hemiplegia.

Respiratory Patterns

These are of less localizing value in comparison to other brainstem signs. Shallow, slow, but regular breathing suggests metabolic or drug depression. Cheyne-Stokes respiration in its classic cyclic form, ending with a brief apneic period, signifies bihemispheral damage or metabolic suppression and commonly accompanies light coma. Rapid, deep (Kussmaul) breathing usually implies metabolic acidosis but may also occur with pontomesencephalic lesions. Tachypnea occurs with lymphoma of the CNS. Agonal gasps are the result of lower brainstem (medullary) damage and are recognized as the terminal respiratory pattern of severe brain damage. A number of other cyclic breathing variations have been described but are of lesser significance.

Laboratory Studies and Imaging

The studies that are most useful in the diagnosis of coma are: chemical-toxicologic analysis of blood and urine, cranial CT or MRI, EEG, and CSF examination. Arterial blood gas analysis is helpful in patients with lung disease and acid-base disorders. The metabolic aberrations commonly encountered in clinical practice require measurement of electrolytes, glucose, calcium, osmolarity, and renal (blood urea nitrogen) and hepatic (NH_3) function. Toxicologic analysis is necessary in any case of acute coma where the diagnosis is not immediately clear. However, the presence of exogenous drugs or toxins, especially alcohol, does not exclude the possibility that other factors, particularly head trauma, are also contributing to the clinical state. An ethanol level of 43 mmol/L (0.2 g/dL) in nonhabituated patients generally causes impaired mental activity; a level of >65 mmol/L (0.3 g/dL) is associated with stupor. The development of tolerance may allow the chronic alcoholic to remain awake at levels >87 mmol/L (0.4 g/dL).

The availability of CT and MRI has focused attention on causes of coma that are detectable by

imaging (e.g., hemorrhage, tumor, or hydrocephalus). Resorting primarily to this approach, although at times expedient, is imprudent because most cases of coma (and confusion) are metabolic or toxic in origin. Furthermore, the notion that a normal CT scan excludes anatomic lesion as the cause of coma is erroneous. Bilateral hemisphere infarction, acute brainstem infarction, encephalitis, meningitis, mechanical shearing of axons as a result of closed head trauma, sagittal sinus thrombosis, and subdural hematoma isodense to adjacent brain are some of the disorders that may not be detected. Nevertheless, if the source of coma remains unknown, a scan should be obtained.

The EEG (Chap. e45) is useful in metabolic or drug-induced states but is rarely diagnostic, except when coma is due to clinically unrecognized seizure, to herpesvirus encephalitis, or to prion (Creutzfeldt-Jakob) disease. The amount of background slowing of the EEG is a reflection of the severity of an encephalopathy. Predominant high-voltage slowing (δ or triphasic waves) in the frontal regions is typical of metabolic coma, as from hepatic failure, and widespread fast (β) activity implicates sedative drugs (e.g., diazepines, barbiturates). A special pattern of "alpha coma," defined by widespread, variable 8- to 12-Hz activity, superficially resembles the normal α rhythm of waking but, unlike normal α activity, is not altered by environmental stimuli. Alpha coma results from pontine or diffuse cortical damage and is associated with a poor prognosis. Normal α activity on the EEG, which is suppressed by stimulating the patient, also alerts the clinician to the locked-in syndrome or to hysteria or catatonia. The most important use of EEG recordings in coma is to reveal clinically inapparent epileptic discharges.

Lumbar puncture is performed less frequently than in the past for coma diagnosis because neuroimaging effectively excludes intracerebral and extensive subarachnoid hemorrhage. However, examination of the CSF remains indispensable in the diagnosis of meningitis and encephalitis. For patients with an altered level of consciousness, it is generally recommended that an imaging study be performed prior to lumbar puncture to exclude a large intracranial mass lesion. Blood culture and antibiotic administration usually precede the imaging study if meningitis is suspected (Chap. e46).

Differential Diagnosis of Coma

(Table 274-1) The causes of coma can be divided into three broad categories: those without focal neurologic signs (e.g., metabolic and toxic encephalopathies); meningitis syndromes, characterized by fever or stiff neck and an excess of cells in the spinal fluid (e.g., bacterial meningitis, subarachnoid hemorrhage); and conditions associated with prominent focal signs (e.g., stroke, cerebral hemorrhage). In most instances, coma is part of an obvious medical problem, such as drug ingestion, hypoxia, stroke, trauma, or liver or kidney failure. Conditions that cause sudden coma include drug ingestion, cerebral hemorrhage, trauma, cardiac arrest, epilepsy, or basilar artery embolism. Coma that appears subacutely is usually related to a preexisting medical or neurologic problem or, less often, to secondary brain swelling of a mass such as tumor or cerebral infarction.

Table 274-1 Differential Diagnosis of Coma

1. Diseases that cause no focal or lateralizing neurologic signs, usually with normal brainstem functions; CT scan and cellular content of the CSF are normal
   a. Intoxications: alcohol, sedative drugs, opiates, etc.
   b. Metabolic disturbances: anoxia, hyponatremia, hypernatremia, hypercalcemia, diabetic acidosis, nonketotic hyperosmolar hyperglycemia, hypoglycemia, uremia, hepatic coma, hypercarbia, addisonian crisis, hypo- and hyperthyroid states, profound nutritional deficiency
   c. Severe systemic infections: pneumonia, septicemia, typhoid fever, malaria, Waterhouse-Friderichsen syndrome
   d. Shock from any cause
   e. Postseizure states, status epilepticus, subclinical epilepsy
   f. Hypertensive encephalopathy, eclampsia
   g. Severe hyperthermia, hypothermia
   h. Concussion
   i. Acute hydrocephalus

2. Diseases that cause meningeal irritation with or without fever, and with an excess of WBCs or RBCs in the CSF, usually without focal or lateralizing cerebral or brainstem signs; CT or MRI shows no mass lesion
   a. Subarachnoid hemorrhage from ruptured aneurysm, arteriovenous malformation, trauma
   b. Acute bacterial meningitis
   c. Viral encephalitis
   d. Miscellaneous: fat embolism, cholesterol embolism, carcinomatous and lymphomatous meningitis, etc.

3. Diseases that cause focal brainstem or lateralizing cere-bral signs, with or without changes in the CSF; CT and MRI are abnormal

a. Hemispheral hemorrhage (basal ganglionic, thalamic) or infarction (large middle cerebral artery territory) with secondary brainstem compression

b. Brainstem infarction due to basilar artery thrombosis or embolism

c. Brain abscess, subdural empyema

d. Epidural and subdural hemorrhage, brain contusion

e. Brain tumor with surrounding edema

f. Cerebellar and pontine hemorrhage and infarction

g. Widespread traumatic brain injury

h. Metabolic coma (see above) with preexisting focal damage

i. Miscellaneous: Cortical vein thrombosis, herpes simplex encephalitis, multiple cerebral emboli due to bacterial endocarditis, acute hemorrhagic leukoencephalitis, acute disseminated (postinfectious) encephalomyelitis, thrombotic thrombocytopenic purpura, cerebral vasculitis, gliomatosis cerebri, pituitary apoplexy, intravascular lymphoma, etc.

Abbreviations: CSF, cerebrospinal fluid; RBCs, red blood cells; WBCs, white blood cells.

When cerebrovascular disease is the cause of coma, diagnosis can be difficult (Chap. 370). The most common diseases are (1) basal ganglia and thalamic hemorrhage (acute but not instantaneous onset, vomiting, headache, hemiplegia, and characteristic eye signs); (2) pontine hemorrhage (sudden onset, pinpoint pupils, loss of reflex eye movements and corneal responses, ocular bobbing, posturing, hyperventilation, and excessive sweating); (3) cerebellar hemorrhage (occipital headache, vomiting, gaze paresis, and inability to stand); (4) basilar artery thrombosis (neurologic prodrome or warning spells, diplopia, dysarthria, vomiting, eye movement and corneal response abnormalities, and asymmetric limb paresis); and (5) subarachnoid hemorrhage (precipitous coma after headache and vomiting). The most common stroke, infarction in the territory of the middle cerebral artery, does not generally cause coma, but edema surrounding large infarctions may expand during the first few days and act as a mass.

The syndrome of acute hydrocephalus accompanies many intracranial diseases, particularly subarachnoid hemorrhage. It is characterized by headache and sometimes vomiting that may progress quickly to coma with extensor posturing of the limbs, bilateral Babinski signs, small unreactive pupils, and impaired oculocephalic movements in the vertical direction.

If the history and examination do not indicate the cause of coma, then information obtained from CT or MRI is needed. The majority of medical causes of coma can be established without a neuroimaging study. Sometimes imaging results can be misleading such as when small subdural hematomas or old strokes are found, but the patient's coma is due to intoxication.

Brain Death

This is a state of cessation of cerebral function with preservation of cardiac activity and maintenance of somatic function by artificial means. It is the only type of brain damage recognized as equivalent to death. Several sets of criteria have been advanced for the diagnosis of brain death and it is essential to adhere to those standards endorsed by the local medical community. Ideal criteria are simple, can be assessed at the bedside, and allow no chance of diagnostic error. They contain three essential elements: (1) widespread cortical destruction that is reflected by deep coma and unresponsiveness to all forms of stimulation; (2) global brainstem damage demonstrated by absent pupillary light reaction and by the loss of oculovestibular and corneal reflexes; and (3) destruction of the medulla, manifested by complete apnea. The heart rate is invariant and unresponsive to atropine. Diabetes insipidus is often present but may only develop hours or days after the other clinical signs of brain death. The pupils are usually midsized but may be enlarged; they should not, however, be small. Loss of deep tendon reflexes is not required because the spinal cord remains functional. Babinski signs are generally absent and the toe response is often flexor.

Demonstration that apnea is due to irreversible medullary damage requires that the Pco_2 be high enough to stimulate respiration during a test of spontaneous breathing. *Apnea testing* can be done safely by the use of diffusion oxygenation prior to removing the ventilator. This is accomplished by preoxygenation with 100% oxygen, which is then sustained during the test by oxygen administered through a tracheal cannula. co_2 tension increases ~0.3–0.4 kPa/min (2–3 mmHg/min) during apnea. At the end of a period of observation, typically several minutes, arterial Pco_2 should be at least >6.6–8.0 kPa (50–60 mmHg) for the test to be valid. Apnea is confirmed if no respiratory effort has been observed in the presence of a sufficiently elevated Pco_2. Other techniques, including the administration of co_2 to accelerate the test, are used in special circumstances. The test is usually stopped if there is serious cardiovascular instability.

An isoelectric EEG may be used as a confirmatory test for total cerebral damage. Radionuclide brain scanning, cerebral angiography, or transcranial Doppler measurements may also be included to demonstrate the absence of CBF but they have not been extensively correlated with pathologic changes.

The possibility of profound drug-induced or hypothermic depression of the nervous system should be excluded, and some period of observation, usually 6–24 hours, is desirable, during which the clinical signs of brain death are sustained. It is advisable to delay clinical testing for at least 24 hours if a cardiac arrest has caused brain death or if the inciting disease is not known.

Although it is largely accepted in Western society that the respirator can be disconnected from a brain-dead patient, problems frequently arise because of poor communication and inadequate preparation of the family by the physician. Reasonable medical practice, ideally with the agreement of the family, allows the removal of support or transfer out of an intensive care unit of patients who are not brain dead but whose neurologic conditions are nonetheless hopeless.

Treatment: Coma

The immediate goal in a comatose patient is prevention of further nervous system damage. Hypotension, hypoglycemia, hypercalcemia, hypoxia, hypercapnia, and hyperthermia should be corrected rapidly. An oropharyngeal airway is adequate to keep the pharynx open in a drowsy patient who is breathing normally. Tracheal intubation is indicated if there is apnea, upper airway obstruction, hypoventilation, or emesis, or if the patient is liable to aspirate because of coma. Mechanical ventilation is required if there is hypoventilation or a need to induce hypocapnia in order to lower ICP. IV access is established, and naloxone and dextrose are administered if narcotic overdose or hypoglycemia are possibilities; thiamine is given along with glucose to avoid provoking Wernicke's disease in malnourished patients. In cases of suspected basilar thrombosis with brainstem ischemia, IV heparin or a thrombolytic agent is often utilized, after cerebral hemorrhage has been excluded by a neuroimaging study. Physostigmine may awaken patients with anticholinergic-type drug overdose but should be used only with careful monitoring; many physicians believe that it should only be used to treat anticholinergic overdose–associated cardiac arrhythmias. The use of benzodiazepine antagonists offers some prospect of improvement after overdose of soporific drugs and has transient benefit in hepatic encephalopathy.

Administration of hypotonic solutions should be monitored carefully in any serious acute brain illness because of the potential for exacerbating brain swelling. Cervical spine injuries must not be overlooked, particularly before attempting intubation or evaluation of oculocephalic responses. Fever and meningismus indicate an urgent need for examination of the CSF to diagnose meningitis. If the lumbar puncture in a case of suspected meningitis is delayed, an antibiotic such as a third-generation cephalosporin may be administered, preferably after obtaining blood cultures. The management of raised ICP is discussed in Chap. 275.

Prognosis

One hopes to avoid the anguishing outcome of a patient who is left severely disabled or vegetative. The uniformly poor outcome of the persistent vegetative state has already been mentioned. Children and young adults may have ominous early clinical findings such as abnormal brainstem reflexes and yet recover; temporization in offering a prognosis in this group of patients is wise. Metabolic comas have a far better prognosis than traumatic ones. All systems for estimating prognosis in adults should be taken as approximations, and medical judgments must be tempered by factors such as age, underlying systemic disease, and general medical condition. In an attempt to collect prognostic information from large numbers of patients with head injury, the Glasgow Coma Scale was devised; empirically, it has predictive value in cases of brain trauma (Table 378-2). For anoxic and metabolic coma, clinical signs such as the pupillary and motor responses after 1 day, 3 days, and 1 week have been shown to have predictive value (Fig. 275-4). Other studies suggest that the absence of corneal responses may have the most discriminative value. The absence of the cortical waves of the somatosensory evoked potentials has also proved a strong indicator of poor outcome in coma from any cause.

There have been recent advances using functional imaging that demonstrate some preserved cognitive abilities of vegetative and minimally conscious patients. In one series, about 10% of such patients could be trained to activate the frontal or temporal lobes in response to requests by an examiner to imagine certain visuospatial tasks. In one case, a rudimentary form of one-way communication could be established. There are also reports in a limited number of patients of improvement in cognitive function with the implantation of thalamic-stimulating electrodes. It is prudent to avoid generalizations from these experiments.

FURTHER READINGS

Booth CM et al: Is this patient dead, vegetative, or severely impaired? *JAMA* 291:870, 2004[PMID: 14970067]

Monti MM et al: Willful modulation of brain activity in disorders of consciousness. N Engl J Med 362:579, 2010[PMID: 20130250]

Posner JB et al: *Plum and Posner's Diagnosis of Stupor and Coma* , 4th ed. New York, Oxford University Press, 2007

Ropper AH et al: *Neurological and Neurosurgical Intensive Care* , 4th ed. New York, Lippincott Williams & Wilkins, 2004

Wijdicks EF et al: Neuropathology of brain death in the modern transplant era. Neurology 70:1234, 2008[PMID: 18256368]

Young GB: Clinical Practice. Neurologic prognosis after cardiac arrest. N Engl J Med 361:605, 2009[PMID: 19657124]

——— Wijdicks EF: Disorders of Consciousness, in *Handbook of Clinical Neurology* , v 90, 3rd series, MJ Aminoff et al (eds). Edinburg, Elsevier, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NEUROLOGIC CRITICAL CARE, INCLUDING HYPOXIC-ISCHEMIC ENCEPHALOPATHY, AND SUBARACHNOID HEMORRHAGE: INTRODUCTION

Life-threatening neurologic illness may be caused by a primary disorder affecting any region of the neuraxis or may occur as a consequence of a systemic disorder such as hepatic failure, multisystem organ failure, or cardiac arrest **(Table 275-1)**. Neurologic critical care focuses on preservation of neurologic tissue and prevention of secondary brain injury caused by ischemia, edema, and elevated intracranial pressure (ICP). Management of other organ systems proceeds concurrently and may need to be modified in order to maintain the overall focus on neurologic issues.

Table 275-1 Neurologic Disorders in Critical Illness

Localization Along Neuroaxis	Syndrome
Central Nervous System	
Brain: Cerebral hemispheres	Global encephalopathy Delirium Sepsis Organ failure—hepatic, renal Medication related—sedatives, hypnotics, analgesics, H_2 blockers, antihypertensives Drug overdose Electrolyte disturbance—hyponatremia, hypoglycemia Hypotension/hypoperfusion Hypoxia Meningitis Subarachnoid hemorrhage Wernicke's disease Seizure—postictal or nonconvulsive status Hypertensive encephalopathy Hypothyroidism—myxedema
	Focal deficits Ischemic stroke Tumor Abscess, subdural empyema Subdural/epidural hematoma
Brainstem	Mass effect and compression Ischemic stroke, intraparenchymal hemorrhage Hypoxia
Spinal cord	Mass effect and compression Disk herniation Epidural hematoma Ischemia—hypotension/embolic Epidural abscess
	Trauma, central cord syndrome

Peripheral Nervous System	
Peripheral nerve	
Axonal	Critical illness polyneuropathy Possible neuromuscular blocking agent complication Metabolic disturbances, uremia, hyperglycemia Medication effects—chemotherapeutic, antiretroviral
Demyelinating	Guillain-Barré syndrome Chronic inflammatory demyelinating polyneuropathy
Neuromuscular junction	Prolonged effect of neuromuscular blockade Medication effects—aminoglycosides Myasthenia gravis, Lambert-Eaton sydrome
Muscle	Critical illness myopathy Septic myopathy Cachectic myopathy—with or without disuse atrophy Electrolyte disturbances—hypokalemia/hyperkalemia, hypophosphatemia Acute quadriplegic myopathy

Pathophsiology

BRAIN EDEMA

Swelling, or edema, of brain tissue occurs with many types of brain injury. The two principal types of edema are vasogenic and cytotoxic. *Vasogenic edema* refers to the influx of fluid and solutes into the brain through an incompetent blood-brain barrier (BBB). In the normal cerebral vasculature, endothelial tight junctions associated with astrocytes create an impermeable barrier (the BBB), through which access into the brain interstitium is dependent upon specific transport mechanisms. The BBB may be compromised in ischemia, trauma, infection, and metabolic derangements. Typically, vasogenic edema develops rapidly following injury. *Cytotoxic edema* refers to cellular swelling and occurs in a variety of settings, including brain ischemia and trauma. Early astrocytic swelling is a hallmark of ischemia. Brain edema that is clinically significant usually represents a combination of vasogenic and cellular components. Edema can lead to increased ICP as well as tissue shifts and brain displacement from focal processes (Chap. 274). These tissue shifts can cause injury by mechanical distention and compression in addition to the ischemia of impaired perfusion consequent to the elevated ICP.

ISCHEMIC CASCADE AND CELLULAR INJURY

When delivery of substrates, principally oxygen and glucose, is inadequate to sustain cellular function, a series of interrelated biochemical reactions known as the *ischemic cascade* is initiated (see Fig. 370-2). The release of excitatory amino acids, especially glutamate, leads to influx of calcium and sodium ions, which disrupt cellular homeostasis. An increased intracellular calcium concentration may activate proteases and lipases, which then lead to lipid peroxidation and free radical–mediated cell membrane injury. Cytotoxic edema ensues, and ultimately necrotic cell death and tissue infarction occur. This pathway to irreversible cell death is common to ischemic stroke, global cerebral ischemia, and traumatic brain injury. *Penumbra* refers to areas of ischemic brain tissue that have not yet undergone irreversible infarction, implying that these regions are potentially salvageable if ischemia can be reversed. Factors that may exacerbate ischemic brain injury include systemic hypotension and hypoxia, which further reduce substrate delivery to vulnerable brain tissue, and fever, seizures, and hyperglycemia, which can increase cellular metabolism, outstripping compensatory processes. Clinically, these events are known as *secondary brain insults* because they lead to exacerbation of the primary brain injury. Prevention, identification, and treatment of secondary brain insults are fundamental goals of management.

An alternative pathway of cellular injury is *apoptosis*. This process implies programmed cell death, which may occur in the setting of ischemic stroke, global cerebral ischemia, traumatic brain injury, and possibly intracerebral hemorrhage. Apoptotic cell death can be distinguished histologically from the necrotic cell death of ischemia and is mediated through a different set of biochemical pathways. At present, interventions for prevention and treatment of apoptotic cell death remain less well defined than those for ischemia. **Excitotoxicity and mechanisms of cell death are discussed in more detail in Chap. 366.**

CEREBRAL PERFUSION AND AUTOREGULATION

Brain tissue requires constant perfusion in order to ensure adequate delivery of substrate. The

hemodynamic response of the brain has the capacity to preserve perfusion across a wide range of systemic blood pressures. Cerebral perfusion pressure (CPP), defined as the mean systemic arterial pressure (MAP) minus the ICP, provides the driving force for circulation across the capillary beds of the brain. *Autoregulation* refers to the physiologic response whereby cerebral blood flow (CBF) is regulated via alterations in cerebrovascular resistance in order to maintain perfusion over wide physiologic changes such as neuronal activation or changes in hemodynamic function. If systemic blood pressure drops, cerebral perfusion is preserved through vasodilation of arterioles in the brain; likewise, arteriolar vasoconstriction occurs at high systemic pressures to prevent hyperperfusion, resulting in fairly constant perfusion across a wide range of systemic blood pressures **(Fig. 275-1)**. At the extreme limits of MAP or CPP (high or low), flow becomes directly related to perfusion pressure. These autoregulatory changes occur in the microcirculation and are mediated by vessels below the resolution of those seen on angiography. CBF is also strongly influenced by pH and Pa_{CO2}. CBF increases with hypercapnia and acidosis and decreases with hypocapnia and alkalosis. This forms the basis for the use of hyperventilation to lower ICP, and this effect on ICP is mediated through a decrease in intracranial blood volume. Cerebral autoregulation is a complex process critical to the normal homeostatic functioning of the brain, and this process may be disordered focally and unpredictably in disease states such as traumatic brain injury and severe focal cerebral ischemia.

Figure 275-1

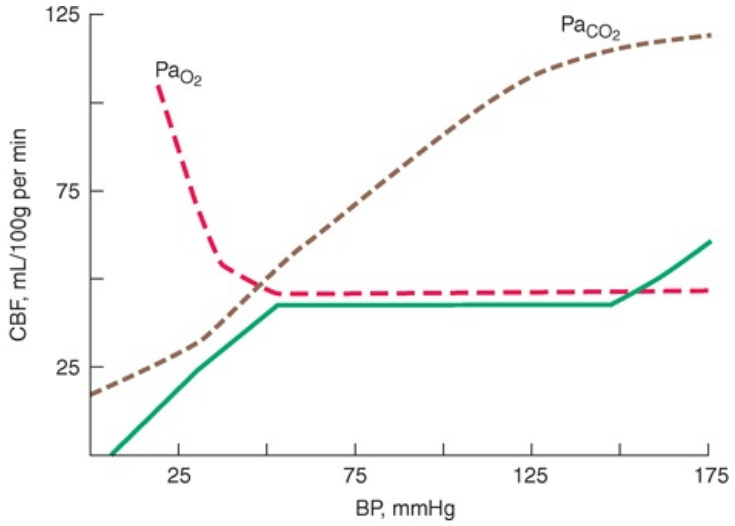

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Autoregulation of cerebral blood flow (*solid line*). Cerebral perfusion is constant over a wide range of systemic blood pressure. Perfusion is increased in the setting of hypoxia or hypercarbia. BP, blood pressure; CBF, cerebral blood flow. *(Reprinted with permission from HM Shapiro: Anesthesiology 43:447, 1975. Copyright 1975, Lippincott Company.*)

CEREBROSPINAL FLUID AND INTRACRANIAL PRESSURE

The cranial contents consist essentially of brain, cerebrospinal fluid (CSF), and blood. CSF is produced principally in the choroid plexus of each lateral ventricle, exits the brain via the foramens of Luschka and Magendi, and flows over the cortex to be absorbed into the venous system along the superior sagittal sinus. Approximately 150 mL of CSF are contained within the ventricles and surrounding the brain and spinal cord; the cerebral blood volume is also ~150 mL. The bony skull offers excellent protection for the brain but allows little tolerance for additional volume. Significant increases in volume eventually result in increased ICP. Obstruction of CSF outflow, edema of cerebral tissue, or increases in volume from tumor or hematoma may increase ICP. Elevated ICP diminishes cerebral perfusion and can lead to tissue ischemia. Ischemia in turn may lead to vasodilation via autoregulatory mechanisms designed to restore cerebral perfusion. However, vasodilation also increases cerebral blood volume, which in turn then increases ICP, lowers CPP, and provokes further ischemia **(Fig. 275-2)**. This vicious cycle is commonly seen in traumatic brain injury, massive intracerebral hemorrhage, and large hemispheric infarcts with significant tissue shifts.

Figure 275-2

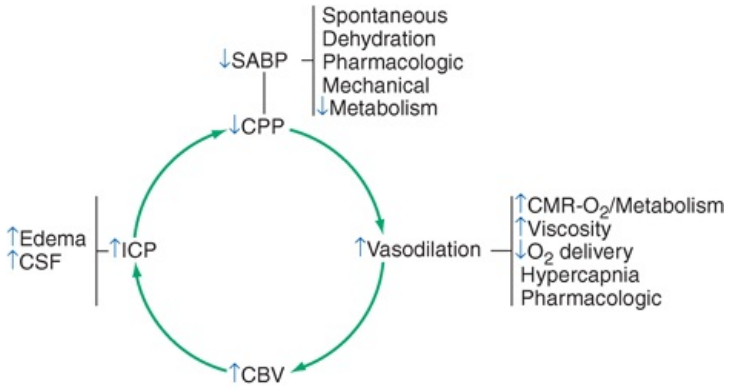

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ischemia and vasodilatation. Reduced cerebral perfusion pressure (CPP) leads to increased ischemia, vasodilation, increased intracranial pressure (ICP), and further reductions in CPP, a cycle leading to further neurologic injury. CBV, cerebral blood volume; CMR, cerebral metabolic rate; CSF, cerebrospinal fluid; SABP, systolic arterial blood pressure. *(Adapted from MJ Rosner et al: J Neurosurg 83:949, 1995; with permission.*)

Approach to the Patient: Severe CNS Dysfunction

Critically ill patients with severe central nervous system dysfunction require rapid evaluation and intervention in order to limit primary and secondary brain injury. Initial neurologic evaluation should be performed concurrent with stabilization of basic respiratory, cardiac, and hemodynamic parameters. Significant barriers may exist to neurologic assessment in the critical care unit, including endotracheal intubation and the use of sedative or paralytic agents to facilitate procedures.

An impaired level of consciousness is common in critically ill patients. The essential first task in assessment is to determine whether the cause of dysfunction is related to a diffuse, usually metabolic, process or whether a focal, usually structural, process is implicated. Examples of diffuse processes include metabolic encephalopathies related to organ failure, drug overdose, or hypoxia-ischemia. Focal processes include ischemic and hemorrhagic stroke and traumatic brain injury, especially with intracranial hematomas. Since these two categories of disorders have fundamentally different causes, treatments, and prognoses, the initial focus is on making this distinction rapidly and accurately. **The approach to the comatose patient is discussed in Chap. 274; etiologies are listed in Table 274-1.**

Minor focal deficits may be present on the neurologic examination in patients with metabolic encephalopathies. However, the finding of prominent focal signs such as pupillary asymmetry, hemiparesis, gaze palsy, or paraplegia should suggest the possibility of a structural lesion. All patients with a decreased level of consciousness associated with focal findings should undergo an urgent neuroimaging procedure, as should all patients with coma of unknown etiology. CT scanning is usually the most appropriate initial study because it can be performed quickly in critically ill patients and demonstrates hemorrhage, hydrocephalus, and intracranial tissue shifts well. MRI may provide more specific information in some situations, such as acute ischemic stroke (diffusion-weighted imaging, DWI) and cerebral venous sinus thrombosis (magnetic resonance venography, MRV). Any suggestion of trauma from the history or examination should alert the examiner to the possibility of cervical spine injury and prompt an imaging evaluation using plain x-rays, CT, or MRI.

Other diagnostic studies are best utilized in specific circumstances, usually when neuroimaging studies fail to reveal a structural lesion and the etiology of the altered mental state remains uncertain. Electroencephalography (EEG) can be important in the evaluation of critically ill patients with severe brain dysfunction. The EEG of metabolic encephalopathy typically reveals generalized slowing. One of the most important uses of EEG is to help exclude inapparent seizures, especially nonconvulsive status epilepticus. Untreated continuous or frequently recurrent seizures may cause neuronal injury, making the diagnosis and treatment of seizures crucial in this patient group. Lumbar puncture (LP) may be necessary to exclude infectious processes, and an elevated opening pressure may be an important clue to cerebral venous sinus thrombosis. In patients with coma or profound encephalopathy, it is preferable to perform a neuroimaging study prior to LP. If bacterial meningitis is suspected, an LP may be performed first or antibiotics may be empirically administered before the diagnostic studies are completed. Standard laboratory evaluation of critically ill patients should include assessment of serum electrolytes (especially sodium and calcium), glucose, renal and hepatic function, complete blood count, and coagulation. Serum or urine toxicology screens should be performed in patients with encephalopathy of unknown cause. EEG, LP, and other specific laboratory tests are most useful when the mechanism of the altered level of consciousness is uncertain; they are not routinely performed in clear-cut cases of stroke or traumatic brain injury.

Monitoring of ICP can be an important tool in selected patients. In general, patients who should be

considered for ICP monitoring are those with primary neurologic disorders, such as stroke or traumatic brain injury, who are at significant risk for secondary brain injury due to elevated ICP and decreased CPP. Included are patients with the following: severe traumatic brain injury [Glasgow Coma Scale (GCS) score ≤ 8 (Table 378-2)]; large tissue shifts from supratentorial ischemic or hemorrhagic stroke; or hydrocephalus from subarachnoid hemorrhage (SAH), intraventricular hemorrhage, or posterior fossa stroke. An additional disorder in which ICP monitoring can add important information is fulminant hepatic failure, in which elevated ICP may be treated with barbiturates or, eventually, liver transplantation. In general, ventriculostomy is preferable to ICP monitoring devices that are placed in the brain parenchyma, because ventriculostomy allows CSF drainage as a method of treating elevated ICP. However, parenchymal ICP monitoring is most appropriate for patients with diffuse edema and small ventricles (which may make ventriculostomy placement more difficult) or any degree of coagulopathy (in which ventriculostomy carries a higher risk of hemorrhagic complications) **(Fig 275-3)**.

Figure 275-3

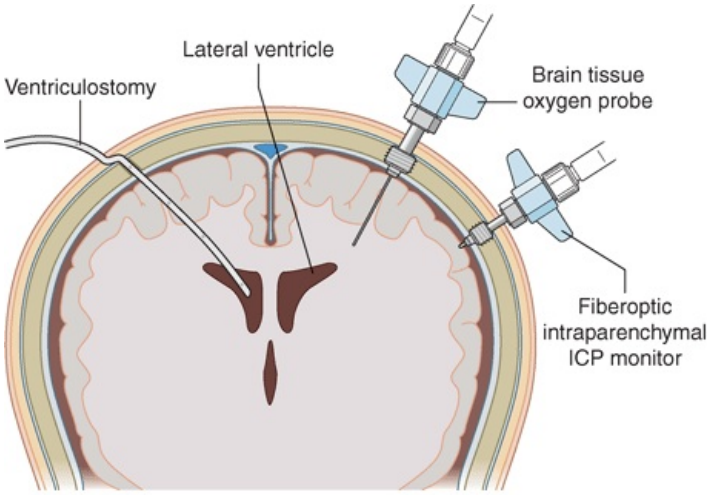

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Intracranial pressure and brain tissue oxygen monitoring. A ventriculostomy allows for drainage of cerebrospinal fluid to treat elevated intracranial pressure (ICP). Fiberoptic ICP and brain tissue oxygen monitors are usually secured using a screwlike skull bolt. Cerebral blood flow and microdialysis probes (not shown) may be placed in a manner similar to the brain tissue oxygen probe.

TREATMENT OF ELEVATED ICP

Elevated ICP may occur in a wide range of disorders, including head trauma, intracerebral hemorrhage, SAH with hydrocephalus, and fulminant hepatic failure. Because CSF and blood volume can be redistributed initially, by the time elevated ICP occurs, intracranial compliance is severely impaired. At this point, any small increase in the volume of CSF, intravascular blood, edema, or a mass lesion may result in a significant increase in ICP and a decrease in cerebral perfusion. This is a fundamental mechanism of secondary ischemic brain injury and constitutes an emergency that requires immediate attention. In general, ICP should be maintained at <20 mmHg and CPP should be maintained at ≥60 mmHg.

Interventions to lower ICP are ideally based on the underlying mechanism responsible for the elevated ICP **(Table 275-2)**. For example, in hydrocephalus from SAH, the principal cause of elevated ICP is impairment of CSF drainage. In this setting, ventricular drainage of CSF is likely to be sufficient and most appropriate. In head trauma and stroke, cytotoxic edema may be most responsible, and the use of osmotic agents such as mannitol or hypertonic saline becomes an appropriate early step. As described above, elevated ICP may cause tissue ischemia, and, if cerebral autoregulation is intact, the resulting vasodilation can lead to a cycle of worsening ischemia. Paradoxically, administration of vasopressor agents to increase mean arterial pressure may actually lower ICP by improving perfusion, thereby allowing autoregulatory vasoconstriction as ischemia is relieved and ultimately decreasing intracranial blood volume.

Table 275-2 Stepwise Approach to Treatment of Elevated Intracranial Pressure*

Insert ICP monitor—ventriculostomy versus parenchymal device
General goals: maintain ICP <20 mmHg and CPP ≥60 mmHg
For ICP >20–25 mmHg for >5 min:

1. Drain CSF via ventriculostomy (if in place)
2. Elevate head of the bed; midline head position
3. Osmotherapy—mannitol 25–100 g q4h as needed (maintain serum osmolality <320 mosmol) or hypertonic saline (30 mL, 23.4% NaCl bolus)
4. Glucocorticoids—dexamethasone 4 mg q6h for vasogenic edema from tumor, abscess (avoid glucocorticoids in head trauma, ischemic and hemorrhagic stroke)
5. Sedation (e.g., morphine, propofol, or midazolam); add neuromuscular paralysis if necessary (patient will require endotracheal intubation and mechanical ventilation at this point, if not before)
6. Hyperventilation—to $PaCO_2$ 30–35 mmHg
7. Pressor therapy—phenylephrine, dopamine, or norepinephrine to maintain adequate MAP to ensure CPP ≥ 60 mmHg (maintain euvolemia to minimize deleterious systemic effects of pressors)
8. Consider second-tier therapies for refractory elevated ICP a. High-dose barbiturate therapy ("pentobarb coma") b. Aggressive hyperventilation to $PaCO_2$ <30 mmHg c. Hypothermia d. Hemicraniectomy

*Throughout ICP treatment algorithm, consider repeat head CT to identify mass lesions amenable to surgical evacuation.

Abbreviations: CPP, cerebral perfusion pressure; CSF, cerebrospi-nal fluid; MAP, mean arterial pressure; $PaCO_2$, arterial partial pressure of carbon dioxide.

Early signs of elevated ICP include drowsiness and a diminished level of consciousness. Neuroimaging studies may reveal evidence of edema and mass effect. Hypotonic IV fluids should be avoided, and elevation of the head of the bed is recommended. Patients must be carefully observed for risk of aspiration and compromise of the airway as the level of alertness declines. Coma and unilateral pupillary changes are late signs and require immediate intervention. Emergent treatment of elevated ICP is most quickly achieved by intubation and hyperventilation, which causes vasoconstriction and reduces cerebral blood volume. In order to avoid provoking or worsening cerebral ischemia, hyperventilation is best used for short periods of time until a more definitive treatment can be instituted. Furthermore, the effects of hyperventilation on ICP are short-lived, often lasting only for several hours because of the buffering capacity of the cerebral interstitium, and rebound elevations of ICP may accompany abrupt discontinuation of hyperventilation. As the level of consciousness declines to coma, the ability to follow the neurologic status of the patient by examination deteriorates and measurement of ICP assumes greater importance. If a ventriculostomy device is in place, direct drainage of CSF to reduce ICP is possible. Finally, high-dose barbiturates, decompressive hemicraniectomy, or hypothermia are sometimes used for refractory elevations of ICP, although these have significant side effects and have not been proven to improve outcome.

SECONDARY BRAIN INSULTS

Patients with primary brain injuries, whether due to trauma or stroke, are at risk for ongoing secondary ischemic brain injury. Because secondary brain injury can be a major determinant of a poor outcome, strategies for minimizing secondary brain insults are an integral part of the critical care of all patients. While elevated ICP may lead to secondary ischemia, most secondary brain injury is mediated through other clinical events that exacerbate the ischemic cascade already initiated by the primary brain injury. Episodes of secondary brain insults are usually not associated with apparent neurologic worsening. Rather, they lead to cumulative injury limiting eventual recovery, which manifests as higher mortality rate or worsened long-term functional outcome. Thus, close monitoring of vital signs is important, as is early intervention to prevent secondary ischemia. Avoiding hypotension and hypoxia is critical, as significant hypotensive events (systolic blood pressure < 90 mmHg) as short as 10 min in duration have been shown to adversely influence outcome after traumatic brain injury. Even in patients with stroke or head trauma who do not require ICP monitoring, close attention to adequate cerebral perfusion is warranted. Hypoxia (pulse oximetry saturation < 90%), particularly in combination with hypotension, also leads to secondary brain injury. Likewise, fever and hyperglycemia both worsen experimental ischemia and have been associated with worsened clinical outcome after stroke and head trauma. Aggressive control of fever with a goal of normothermia is warranted but may be difficult to achieve with antipyretic medications and cooling blankets. The value of newer surface or intravascular temperature control devices for the management of refractory fever is under investigation. The use of IV insulin infusion is encouraged for control of hyperglycemia as this allows better regulation of serum glucose levels than SC insulin. A reasonable goal is to maintain the serum glucose level at <7.8 mmol/L (<140 mg/dL), although episodes of hypoglycemia appear equally detrimental and the optimal targets remain uncertain. New cerebral monitoring tools that allow

continuous evaluation of brain tissue oxygen tension, CBF, and metabolism (via microdialysis) may further improve the management of secondary brain injury.

CRITICAL CARE DISORDERS OF THE CENTRAL NERVOUS SYSTEM

Hypoxic-Ischemic Encephalopathy

This occurs from lack of delivery of oxygen to the brain because of hypotension or respiratory failure. Causes include myocardial infarction, cardiac arrest, shock, asphyxiation, paralysis of respiration, and carbon monoxide or cyanide poisoning. In some circumstances, hypoxia may predominate. Carbon monoxide and cyanide poisoning are termed *histotoxic hypoxia* since they cause a direct impairment of the respiratory chain.

CLINICAL MANIFESTATIONS

Mild degrees of pure hypoxia, such as occur at high altitudes, cause impaired judgment, inattentiveness, motor incoordination, and, at times, euphoria. However, with hypoxia-ischemia, such as occurs with circulatory arrest, consciousness is lost within seconds. If circulation is restored within 3–5 min, full recovery may occur, but if hypoxia-ischemia lasts beyond 3–5 min, some degree of permanent cerebral damage usually results. Except in extreme cases, it may be difficult to judge the precise degree of hypoxia-ischemia, and some patients make a relatively full recovery after even 8–10 min of global cerebral ischemia. The distinction between pure hypoxia and hypoxia-ischemia is important, since a Pa_{O2} as low as 20 mmHg (2.7 kPa) can be well tolerated if it develops gradually and normal blood pressure is maintained, but short durations of very low or absent cerebral circulation may result in permanent impairment.

Clinical examination at different time points after a hypoxic-ischemic insult (especially cardiac arrest) is useful in assessing prognosis for long-term neurologic outcome. The prognosis is better for patients with intact brainstem function, as indicated by normal pupillary light responses and intact oculocephalic (doll's eyes), oculovestibular (caloric), and corneal reflexes (Fig. 275-4) . Absence of these reflexes and the presence of persistently dilated pupils that do not react to light are grave prognostic signs. A uniformly dismal prognosis from hypoxic-ischemic coma is conveyed by an absent pupillary light reflex or extensor or absent motor response to pain on day 3 following the injury. Electrophysiologically, the bilateral absence of the N20 component of the somatosensory evoked potential (SSEP) in the first several days also conveys a poor prognosis. A very elevated serum level (>33 μg/L) of the biochemical marker neuron-specific enolase (NSE) is indicative of brain damage after resuscitation from cardiac arrest and predicts a poor outcome. However, at present, SSEPs and NSE levels may be difficult to obtain in a timely fashion, with SSEP testing requiring substantial expertise in interpretation and NSE measurements not yet standardized. Whether administration of mild hypothermia after cardiac arrest (see “Treatment”) will alter the usefulness of these clinical and electrophysiologic predictors is unknown. Long-term consequences of hypoxic-ischemic encephalopathy include persistent coma or a vegetative state **(Chap. 274)**, dementia, visual agnosia (Chap. 26), parkinsonism, choreoathetosis, cerebellar ataxia, myoclonus, seizures, and an amnestic state, which may be a consequence of selective damage to the hippocampus.

Figure 275-4

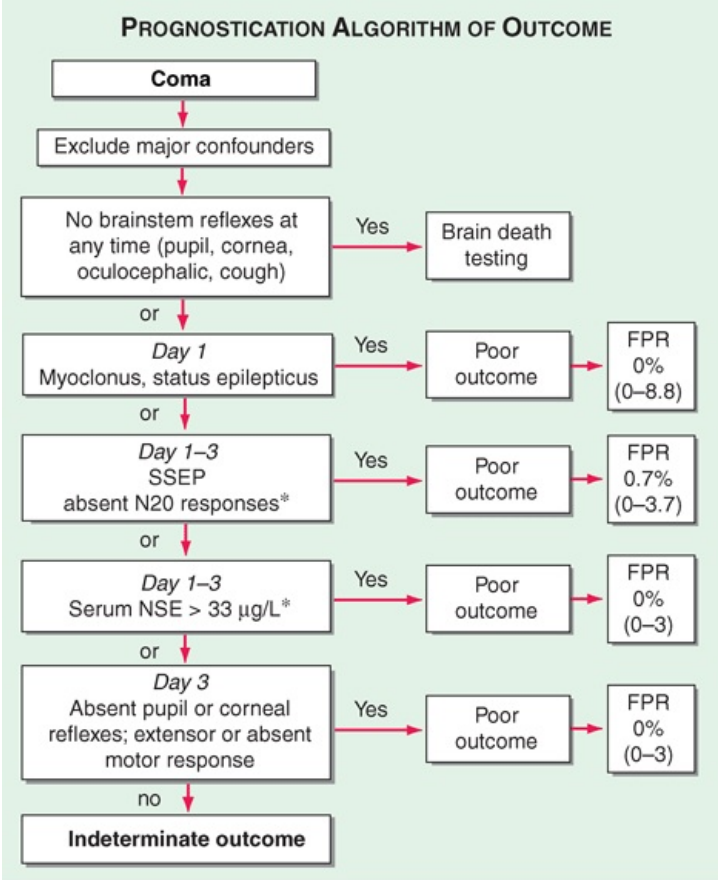

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prognostication of outcome in comatose survivors of cardiopulmonary resuscitation. Numbers in parentheses are 95% confidence intervals. Confounders could include use of sedatives or neuromuscular blocking agents, hypothermia therapy, organ failure, or shock. Tests denoted with an asterisk (*) may not be available in a timely and standardized manner. SSEP, somatosensory evoked potentials; NSE, neuron-specific enolase; FPR, false-positive rate (Chap. 3). *(From Wijdicks et al, with permission.*)

PATHOLOGY

Principal histologic findings are extensive multifocal or diffuse laminar cortical necrosis **(Fig. 275-5)**, with almost invariable involvement of the hippocampus. The hippocampal CA1 neurons are vulnerable to even brief episodes of hypoxia-ischemia, perhaps explaining why selective persistent memory deficits may occur after brief cardiac arrest. Scattered small areas of infarction or neuronal loss may be present in the basal ganglia, hypothalamus, or brainstem. In some cases, extensive bilateral thalamic scarring may affect pathways that mediate arousal, and this pathology may be responsible for the persistent vegetative state. A specific form of hypoxic-ischemic encephalopathy, so-called watershed infarcts, occurs at the distal territories between the major cerebral arteries and can cause cognitive deficits, including visual agnosia, and weakness that is greater in proximal than in distal muscle groups.

Figure 275-5

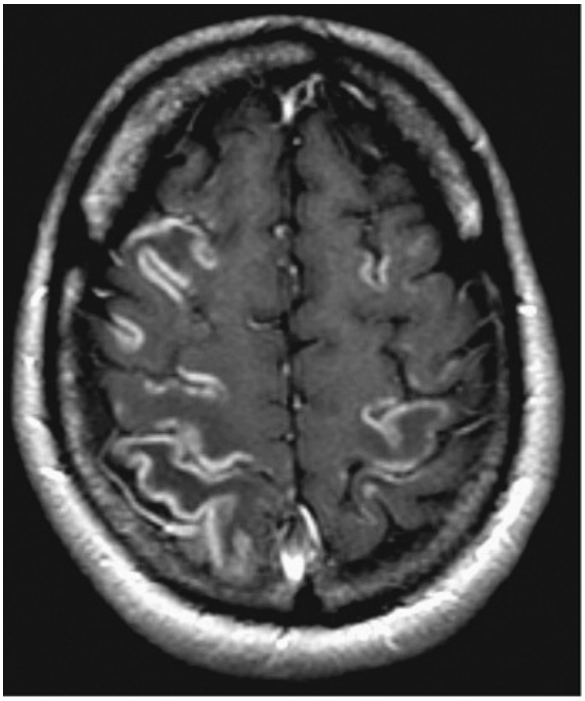

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cortical laminar necrosis in hypoxic-ischemic encephalopathy. T1-weighted postcontrast MRI shows cortical enhancement in a watershed distribution consistent with laminar necrosis.

DIAGNOSIS

Diagnosis is based upon the history of a hypoxic-ischemic event such as cardiac arrest. Blood pressure <70 mmHg systolic or Pao_2 <40 mmHg is usually necessary, although both absolute levels as well as duration of exposure are important determinants of cellular injury. Carbon monoxide intoxication can be confirmed by measurement of carboxyhemoglobin and is suggested by a cherry red color of the skin, although the latter is an inconsistent clinical finding.

TREATMENT: HYPOXIC-ISCHEMIC ENCEPHALOPATHY

Treatment should be directed at restoration of normal cardiorespiratory function. This includes securing a clear airway, ensuring adequate oxygenation and ventilation, and restoring cerebral perfusion, whether by cardiopulmonary resuscitation, fluid, pressors, or cardiac pacing. Hypothermia may target the neuronal cell injury cascade and has substantial neuroprotective properties in experimental models of brain injury. In two trials, mild hypothermia (33°C) improved functional outcome in patients who remained comatose after resuscitation from a cardiac arrest. Treatment was initiated within minutes of cardiac resuscitation and continued for 12 h in one study and 24 h in the other. Potential complications of hypothermia include coagulopathy and an increased risk of infection. Based upon these studies, the International Liaison Committee on Resuscitation issued the following advisory statement in 2003: "Unconscious adult patients with spontaneous circulation after out-of-hospital cardiac arrest should be cooled to 32°–34°C for 12–24 h when the initial rhythm was ventricular fibrillation."

Severe carbon monoxide intoxication may be treated with hyperbaric oxygen. Anticonvulsants may be needed to control seizures, although these are not usually given prophylactically. Posthypoxic myoclonus may respond to oral administration of clonazepam at doses of 1.5–10 mg daily or valproate at doses of 300–1200 mg daily in divided doses. Myoclonic status epilepticus within 24 h after a primary circulatory arrest generally portends a very poor prognosis, even if seizures are controlled.

Carbon monoxide and cyanide intoxication can also cause a delayed encephalopathy. Little clinical impairment is evident when the patient first regains consciousness, but a parkinsonian syndrome characterized by akinesia and rigidity without tremor may develop. Symptoms can worsen over months, accompanied by increasing evidence of damage in the basal ganglia as seen on both CT and MRI.

Metabolic Encephalopathies

Altered mental states, variously described as confusion, delirium, disorientation, and encephalopathy, are present in many patients with severe illness in an intensive care unit (ICU). Older patients are particularly vulnerable to delirium, a confusional state characterized by disordered perception, frequent hallucinations, delusions, and sleep disturbance. This is often attributed to medication effects, sleep deprivation, pain, and anxiety. The presence of delirium is associated with worsened outcome in critically ill patients, even in those without an identifiable central nervous system pathology such as

stroke or brain trauma. In these patients, the cause of delirium is often multifactorial, resulting from organ dysfunction, sepsis, and especially the use of medications given to treat pain, agitation, or anxiety. Critically ill patients are often treated with a variety of sedative and analgesic medications, including opiates, benzodiazepines, neuroleptics, and sedative-anesthetic medications, such as propofol. Recent studies suggest that in critically ill patients requiring sedation, the use of the centrally acting α_2 agonist dexmedetomidine reduces delirium and shortens the duration of mechanical ventilation compared to the use of benzodiazepines such as lorazepam or midazolam. The presence of family members in the ICU may also help to calm and orient agitated patients, and in severe cases, low doses of neuroleptics (e.g., haloperidol 0.5–1 mg) can be useful. Current strategies focus on limiting the use of sedative medications when this can be done safely.

In the ICU setting, several metabolic causes of an altered level of consciousness predominate. Hypercarbic encephalopathy can present with headache, confusion, stupor, or coma. Hypoventilation syndrome occurs most frequently in patients with a history of chronic CO_2 retention who are receiving oxygen therapy for emphysema or chronic pulmonary disease (Chap. 264). The elevated Pa_{CO2} leading to CO_2 narcosis may have a direct anesthetic effect, and cerebral vasodilation from increased Pa_{CO2} can lead to increased ICP. Hepatic encephalopathy is suggested by asterixis and can occur in chronic liver failure or acute fulminant hepatic failure. Both hyperglycemia and hypoglycemia can cause encephalopathy, as can hypernatremia and hyponatremia. Confusion, impairment of eye movements, and gait ataxia are the hallmarks of acute Wernicke's disease (see below).

Sepsis-Associated Encephalopathy

PATHOGENESIS

In patients with sepsis, the systemic response to infectious agents leads to the release of circulating inflammatory mediators that appear to contribute to encephalopathy. Critical illness, in association with the systemic inflammatory response syndrome (SIRS), can lead to multisystem organ failure. This syndrome can occur in the setting of apparent sepsis, severe burns, or trauma, even without clear identification of an infectious agent. Many patients with critical illness, sepsis, or SIRS develop encephalopathy without obvious explanation. This condition is broadly termed *sepsis-associated encephalopathy*. While the specific mediators leading to neurologic dysfunction remain uncertain, it is clear that the encephalopathy is not simply the result of metabolic derangements of multiorgan failure. The cytokines tumor necrosis factor, interleukin (IL)-1, IL-2, and IL-6 are thought to play a role in this syndrome.

DIAGNOSIS

Sepsis-associated encephalopathy presents clinically as a diffuse dysfunction of the brain without prominent focal findings. Confusion, disorientation, agitation, and fluctuations in level of alertness are typical. In more profound cases, especially with hemodynamic compromise, the decrease in level of alertness can be more prominent, at times resulting in coma. Hyperreflexia and frontal release signs such as a grasp or snout reflex (Chap. 26) can be seen. Abnormal movements such as myoclonus, tremor, or asterixis can occur. Sepsis-associated encephalopathy is quite common, occurring in the majority of patients with sepsis and multisystem organ failure. Diagnosis is often difficult because of the multiple potential causes of neurologic dysfunction in critically ill patients and requires exclusion of structural, metabolic, toxic, and infectious (e.g., meningitis or encephalitis) causes. The mortality rate of patients with sepsis-associated encephalopathy severe enough to produce coma approaches 50%, although this principally reflects the severity of the underlying critical illness and is not a direct result of the encephalopathy. Patients dying from severe sepsis or septic shock may have elevated levels of the serum brain injury biomarker S-100β and neuropathologic findings of neuronal apoptosis and cerebral ischemic injury. However, successful treatment of the underlying critical illness almost always results in complete resolution of the encephalopathy, with profound long-term cognitive disability being uncommon.

Central Pontine Myelinolysis

This disorder typically presents in a devastating fashion as quadriplegia and pseudobulbar palsy. Predisposing factors include severe underlying medical illness or nutritional deficiency; most cases are associated with rapid correction of hyponatremia or with hyperosmolar states. The pathology consists of demyelination without inflammation in the base of the pons, with relative sparing of axons and nerve cells. MRI is useful in establishing the diagnosis **(Fig. 275-6)** and may also identify partial forms that present as confusion, dysarthria, and/or disturbances of conjugate gaze without quadriplegia. Occasional cases present with lesions outside of the brainstem. Therapeutic guidelines for the restoration of severe hyponatremia should aim for gradual correction, i.e., by $\leq$10 mmol/L (10 meq/L) within 24 h and 20 mmol/L (20 meq/L) within 48 h.

Figure 275-6

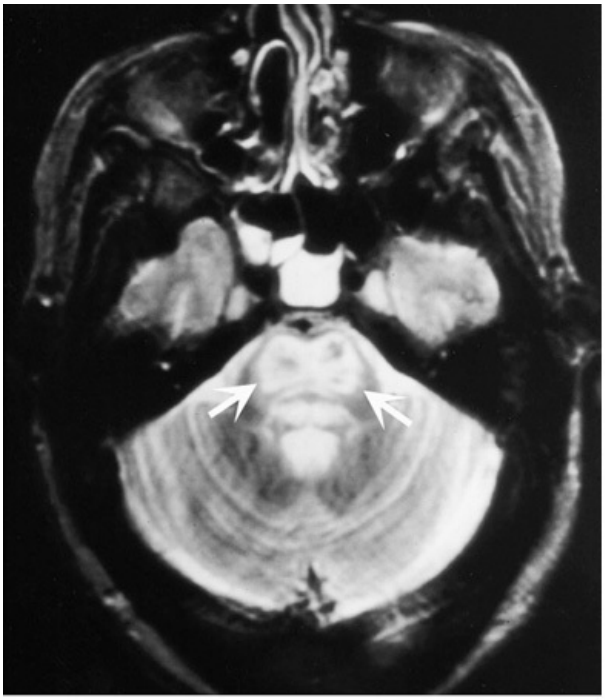

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Central pontine myelinolysis. Axial T2-weighted MR scan through the pons reveals a symmetric area of abnormal high signal intensity within the basis pontis (*arrows*).

Wernicke's Disease

Wernicke's disease is a common and preventable disorder due to a deficiency of thiamine (Chap. 74). In the United States, alcoholics account for most cases, but patients with malnutrition due to hyperemesis, starvation, renal dialysis, cancer, AIDS, or rarely gastric surgery are also at risk. The characteristic clinical triad is that of ophthalmoplegia, ataxia, and global confusion. However, only one-third of patients with acute Wernicke's disease present with the classic clinical triad. Most patients are profoundly disoriented, indifferent, and inattentive, although rarely they have an agitated delirium related to ethanol withdrawal. If the disease is not treated, stupor, coma, and death may ensue. Ocular motor abnormalities include horizontal nystagmus on lateral gaze, lateral rectus palsy (usually bilateral), conjugate gaze palsies, and rarely ptosis. Gait ataxia probably results from a combination of polyneuropathy, cerebellar involvement, and vestibular paresis. The pupils are usually spared, but they may become miotic with advanced disease.

Wernicke's disease is usually associated with other manifestations of nutritional disease, such as polyneuropathy. Rarely, amblyopia or myelopathy occurs. Tachycardia and postural hypotension may be related to impaired function of the autonomic nervous system or to the coexistence of cardiovascular beriberi. Patients who recover show improvement in ocular palsies within hours after the administration of thiamine, but horizontal nystagmus may persist. Ataxia improves more slowly than the ocular motor abnormalities. Approximately half recover incompletely and are left with a slow, shuffling, wide-based gait and an inability to tandem walk. Apathy, drowsiness, and confusion improve more gradually. As these symptoms recede, an amnestic state with impairment in recent memory and learning may become more apparent (*Korsakoff's psychosis*). Korsakoff's psychosis is frequently persistent; the residual mental state is characterized by gaps in memory, confabulation, and disordered temporal sequencing.

PATHOLOGY

Periventricular lesions surround the third ventricle, aqueduct, and fourth ventricle, with petechial hemorrhages in occasional acute cases and atrophy of the mamillary bodies in most chronic cases. There is frequently endothelial proliferation, demyelination, and some neuronal loss. These changes may be detected by MRI scanning **(Fig. 275-7)**. The amnestic defect is related to lesions in the dorsal medial nuclei of the thalamus.

Figure 275-7

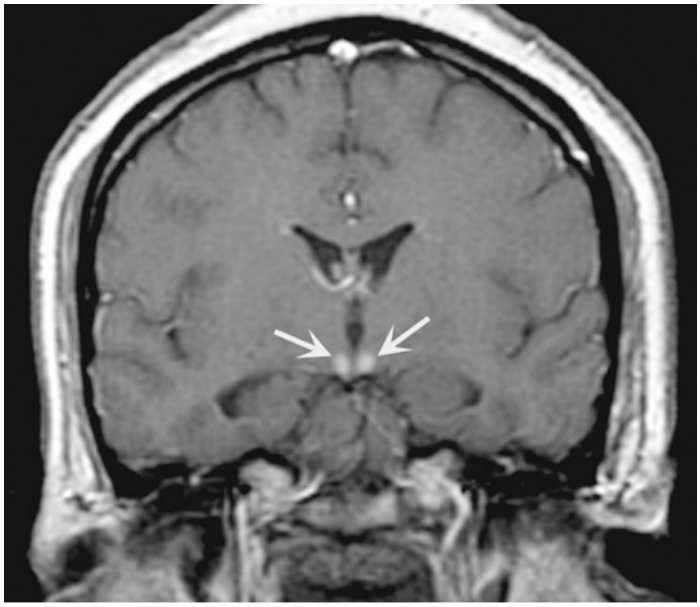

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Wernicke's disease. Coronal T1-weighted postcontrast MRI reveals abnormal enhancement of the mammillary bodies (*arrows*), typical of acute Wernicke's encephalopathy.

PATHOGENESIS

Thiamine is a cofactor of several enzymes, including transketolase, pyruvate dehydrogenase, and α-ketoglutarate dehydrogenase. Thiamine deficiency produces a diffuse decrease in cerebral glucose utilization and results in mitochondrial damage. Glutamate accumulates owing to impairment of α-ketoglutarate dehydrogenase activity and, in combination with the energy deficiency, may result in excitotoxic cell damage.

TREATMENT: WERNICKE'S DISEASE

Wernicke's disease is a medical emergency and requires immediate administration of thiamine, in a dose of 100 mg either IV or IM. The dose should be given daily until the patient resumes a normal diet and should be begun prior to treatment with IV glucose solutions. Glucose infusions may precipitate Wernicke's disease in a previously unaffected patient or cause a rapid worsening of an early form of the disease. For this reason, thiamine should be administered to all alcoholic patients requiring parenteral glucose.

CRITICAL CARE DISORDERS OF THE PERIPHERAL NERVOUS SYSTEM

Critical illness with disorders of the peripheral nervous system (PNS) arises in two contexts: (1) primary neurologic diseases that require critical care interventions such as intubation and mechanical ventilation, and (2) secondary PNS manifestations of systemic critical illness, often involving multisystem organ failure. The former include acute polyneuropathies such as Guillain-Barré syndrome (Chap. 385), neuromuscular junction disorders including myasthenia gravis (Chap. 386) and botulism (Chap. 141), and primary muscle disorders such as polymyositis (Chap. 388). The latter result either from the systemic disease itself or as a consequence of interventions.

General principles of respiratory evaluation in patients with PNS involvement, regardless of cause, include assessment of pulmonary mechanics, such as maximal inspiratory force (MIF) and vital capacity (VC), and evaluation of strength of bulbar muscles. Regardless of the cause of weakness, endotracheal intubation should be considered when the MIF falls to <–25 cmH_2O or the VC is <1 L.

Also, patients with severe palatal weakness may require endotracheal intubation in order to prevent acute upper airway obstruction or recurrent aspiration. Arterial blood gases and oxygen saturation from pulse oximetry are used to follow patients with potential respiratory compromise from PNS dysfunction. However, intubation and mechanical ventilation should be undertaken based on clinical assessment rather than waiting until oxygen saturation drops or CO_2 retention develops from hypoventilation.

Noninvasive mechanical ventilation may be considered initially in lieu of endotracheal intubation but is generally insufficient in patients with severe bulbar weakness or ventilatory failure with hypercarbia. **Principles of mechanical ventilation are discussed in Chap. 269.**

Neuropathy

While encephalopathy may be the most obvious neurologic dysfunction in critically ill patients, dysfunction of the PNS is also quite common. It is typically present in patients with prolonged critical illnesses lasting several weeks and involving sepsis; clinical suspicion is aroused when there is failure

to wean from mechanical ventilation despite improvement of the underlying sepsis and critical illness. *Critical illness polyneuropathy* refers to the most common PNS complication related to critical illness; it is seen in the setting of prolonged critical illness, sepsis, and multisystem organ failure. Neurologic findings include diffuse weakness, decreased reflexes, and distal sensory loss. Electrophysiologic studies demonstrate a diffuse, symmetric, distal axonal sensorimotor neuropathy, and pathologic studies have confirmed axonal degeneration. The precise mechanism of critical illness polyneuropathy remains unclear, but circulating factors such as cytokines, which are associated with sepsis and SIRS, are thought to play a role. It has been reported that up to 70% of patients with the sepsis syndrome have some degree of neuropathy, although far fewer have a clinical syndrome profound enough to cause severe respiratory muscle weakness requiring prolonged mechanical ventilation or resulting in failure to wean. Aggressive glycemic control with insulin infusions appears to decrease the risk of critical illness polyneuropathy. Treatment is otherwise supportive, with specific intervention directed at treating the underlying illness. While spontaneous recovery is usually seen, the time course may extend over weeks to months and necessitate long-term ventilatory support and care even after the underlying critical illness has resolved.

Disorders of Neuromuscular Transmission

A defect in neuromuscular transmission may be a source of weakness in critically ill patients. Myasthenia gravis may be a consideration; however, persistent weakness secondary to impaired neuromuscular junction transmission is almost always due to administration of drugs. A number of medications impair neuromuscular transmission; these include antibiotics, especially aminoglycosides, and beta-blocking agents. In the ICU, the nondepolarizing neuromuscular blocking agents (nd-NMBAs), also known as muscle relaxants, are most commonly responsible. Included in this group of drugs are such agents as pancuronium, vecuronium, rocuronium, and atracurium. They are often used to facilitate mechanical ventilation or other critical care procedures, but with prolonged use persistent neuromuscular blockade may result in weakness even after discontinuation of these agents hours or days earlier. Risk factors for this prolonged action of neuromuscular blocking agents include female sex, metabolic acidosis, and renal failure.

Prolonged neuromuscular blockade does not appear to produce permanent damage to the PNS. Once the offending medications are discontinued, full strength is restored, although this may take days. In general, the lowest dose of neuromuscular blocking agent should be used to achieve the desired result and, when these agents are used in the ICU, a peripheral nerve stimulator should be used to monitor neuromuscular junction function.

Myopathy

Critically ill patients, especially those with sepsis, frequently develop muscle wasting, often in the face of seemingly adequate nutritional support. The assumption has been that this represents a catabolic myopathy brought about as a result of multiple factors, including elevated cortisol and catecholamine release and other circulating factors induced by the SIRS. In this syndrome, known as *cachectic myopathy*, serum creatine kinase levels and electromyography (EMG) are normal. Muscle biopsy shows type II fiber atrophy. Panfascicular muscle fiber necrosis may also occur in the setting of profound sepsis. This so-called *septic myopathy* is characterized clinically by weakness progressing to a profound level over just a few days. There may be associated elevations in serum creatine kinase and urine myoglobin. Both EMG and muscle biopsy may be normal initially but eventually show abnormal spontaneous activity and panfascicular necrosis with an accompanying inflammatory reaction. Both of these myopathic syndromes may be considered under the broader heading of *critical illness myopathy*.

Acute quadriplegic myopathy describes a clinical syndrome of severe weakness seen in the setting of glucocorticoid and nd-NMBA use. The most frequent scenario in which this is encountered is the asthmatic patient who requires high-dose glucocorticoids and nd-NMBA to facilitate mechanical ventilation. This muscle disorder is not due to prolonged action of nd-NMBAs at the neuromuscular junction but, rather, is an actual myopathy with muscle damage; it has occasionally been described with high-dose glucocorticoid use alone. Clinically this syndrome is most often recognized when a patient fails to wean from mechanical ventilation despite resolution of the primary pulmonary process. Pathologically, there may be vacuolar changes in both type I and type II muscle fibers with evidence of regeneration. Acute quadriplegic myopathy has a good prognosis. If patients survive their underlying critical illness, the myopathy invariably improves and most patients return to normal. However, because this syndrome is a result of true muscle damage, not just prolonged blockade at the neuromuscular junction, this process may take weeks or months, and tracheotomy with prolonged ventilatory support may be necessary. Some patients do have residual long-term weakness, with atrophy and fatigue limiting ambulation. At present, it is unclear how to prevent this myopathic complication, except by avoiding use of nd-NMBAs, a strategy not always possible. Monitoring with a peripheral nerve stimulator can help to avoid the overuse of these agents. However, this is more likely to prevent the complication of prolonged neuromuscular junction blockade than it is to prevent this myopathy.

SUBARACHNOID HEMORRHAGE

Subarachnoid hemorrhage (SAH) renders the brain critically ill from both primary and secondary brain insults. Excluding head trauma, the most common cause of SAH is rupture of a saccular aneurysm. Other causes include bleeding from a vascular malformation (arteriovenous malformation or dural arterial-venous fistula) and extension into the subarachnoid space from a primary intracerebral hemorrhage. Some idiopathic SAHs are localized to the perimesencephalic cisterns and are benign; they probably have a venous or capillary source, and angiography is unrevealing.

SACCULAR ("BERRY") ANEURYSM

Autopsy and angiography studies have found that about 2% of adults harbor intracranial aneurysms, for a prevalence of 4 million persons in the United States; the aneurysm will rupture, producing SAH, in 25,000–30,000 cases per year. For patients who arrive alive at hospital, the mortality rate over the next month is about 45%. Of those who survive, more than half are left with major neurologic deficits as a result of the initial hemorrhage, cerebral vasospasm with infarction, or hydrocephalus. If the patient survives but the aneurysm is not obliterated, the rate of rebleeding is about 20% in the first 2 weeks, 30% in the first month, and about 3% per year afterwards. Given these alarming figures, the major therapeutic emphasis is on preventing the predictable early complications of the SAH.

Unruptured, asymptomatic aneurysms are much less dangerous than a recently ruptured aneurysm. The annual risk of rupture for aneurysms <10 mm in size is ~0.1%, and for aneurysms ≥10 mm in size is ~0.5–1%; the surgical morbidity rate far exceeds these percentages. Because of the longer length of exposure to risk of rupture, younger patients with aneurysms >10 mm in size may benefit from prophylactic treatment. As with the treatment of asymptomatic carotid stenosis, this risk-benefit strongly depends on the complication rate of treatment.

Giant aneurysms, those >2.5 cm in diameter, occur at the same sites (see below) as small aneurysms and account for 5% of cases. The three most common locations are the terminal internal carotid artery, middle cerebral artery (MCA) bifurcation, and top of the basilar artery. Their risk of rupture is ~6% in the first year after identification and may remain high indefinitely. They often cause symptoms by compressing the adjacent brain or cranial nerves.

Mycotic aneurysms are usually located distal to the first bifurcation of major arteries of the circle of Willis. Most result from infected emboli due to bacterial endocarditis causing septic degeneration of arteries and subsequent dilation and rupture. Whether these lesions should be sought and repaired prior to rupture or left to heal spontaneously is controversial.

Pathophysiology

Saccular aneurysms occur at the bifurcations of the large- to medium-sized intracranial arteries; rupture is into the subarachnoid space in the basal cisterns and often into the parenchyma of the adjacent brain. Approximately 85% of aneurysms occur in the anterior circulation, mostly on the circle of Willis. About 20% of patients have multiple aneurysms, many at mirror sites bilaterally. As an aneurysm develops, it typically forms a neck with a dome. The length of the neck and the size of the dome vary greatly and are important factors in planning neurosurgical obliteration or endovascular embolization. The arterial internal elastic lamina disappears at the base of the neck. The media thins, and connective tissue replaces smooth-muscle cells. At the site of rupture (most often the dome) the wall thins, and the tear that allows bleeding is often ≤0.5 mm long. Aneurysm size and site are important in predicting risk of rupture. Those >7 mm in diameter and those at the top of the basilar artery and at the origin of the posterior communicating artery are at greater risk of rupture.

Clinical Manifestations

Most unruptured intracranial aneurysms are completely asymptomatic. Symptoms are usually due to rupture and resultant SAH, although some unruptured aneurysms present with mass effect on cranial nerves or brain parenchyma. At the moment of aneurysmal rupture with major SAH, the ICP suddenly rises. This may account for the sudden transient loss of consciousness that occurs in nearly half of patients. Sudden loss of consciousness may be preceded by a brief moment of excruciating headache, but most patients first complain of headache upon regaining consciousness. In 10% of cases, aneurysmal bleeding is severe enough to cause loss of consciousness for several days. In ~45% of cases, severe headache associated with exertion is the presenting complaint. The patient often calls the headache "the worst headache of my life"; however, the most important characteristic is sudden onset. Occasionally, these ruptures may present as headache of only moderate intensity or as a change in the patient's usual headache pattern. The headache is usually generalized, often with neck stiffness, and vomiting is common.

Although sudden headache in the absence of focal neurologic symptoms is the hallmark of aneurysmal rupture, focal neurologic deficits may occur. Anterior communicating artery or MCA bifurcation aneurysms may rupture into the adjacent brain or subdural space and form a hematoma large enough to produce mass effect. The deficits that result can include hemiparesis, aphasia, and abulia.

Occasionally, prodromal symptoms suggest the location of a progressively enlarging unruptured aneurysm. A third cranial nerve palsy, particularly when associated with pupillary dilation, loss of ipsilateral (but retained contralateral) light reflex, and focal pain above or behind the eye, may occur with an expanding aneurysm at the junction of the posterior communicating artery and the internal

carotid artery. A sixth nerve palsy may indicate an aneurysm in the cavernous sinus, and visual field defects can occur with an expanding supraclinoid carotid or anterior cerebral artery aneurysm. Occipital and posterior cervical pain may signal a posterior inferior cerebellar artery or anterior inferior cerebellar artery aneurysm (Chap. 370). Pain in or behind the eye and in the low temple can occur with an expanding MCA aneurysm. Thunderclap headache is a variant of migraine that simulates an SAH. Before concluding that a patient with sudden, severe headache has thunderclap migraine, a definitive workup for aneurysm or other intracranial pathology is required.

Aneurysms can undergo small ruptures and leaks of blood into the subarachnoid space, so-called *sentinel bleeds*. Sudden unexplained headache at any location should raise suspicion of SAH and be investigated, because a major hemorrhage may be imminent.

The initial clinical manifestations of SAH can be graded using the Hunt-Hess or World Federation of Neurosurgical Societies classification schemes **(Table 275-3)**. For ruptured aneurysms, prognosis for good outcomes falls as the grade increases. For example, it is unusual for a Hunt-Hess grade 1 patient to die if the aneurysm is treated, but the mortality rate for grade 4 and 5 patients may be as high as 80%.

Table 275-3 Grading Scales for Subarachnoid Hemorrhage

Grade	Hunt-Hess Scale	World Federation of Neurosurgical Societies (WFNS) Scale
1	Mild headache, normal mental status, no cranial nerve or motor findings	GCS* score 15, no motor deficits
2	Severe headache, normal mental status, may have cranial nerve deficit	GCS score 13–14, no motor deficits
3	Somnolent, confused, may have cranial nerve or mild motor deficit	GCS score 13–14, with motor deficits
4	Stupor, moderate to severe motor deficit, may have intermittent reflex posturing	GCS score 7–12, with or without motor defi-cits
5	Coma, reflex posturing or flaccid	GCS score 3–6, with or without motor deficits

*Glasgow Coma Scale: See Table 378-2.

Delayed Neurologic Deficits

There are four major causes of delayed neurologic deficits: rerupture, hydrocephalus, vasospasm, and hyponatremia.

1. *Rerupture*. The incidence of rerupture of an untreated aneurysm in the first month following SAH is ~30%, with the peak in the first 7 days. Rerupture is associated with a 60% mortality rate and poor outcome. Early treatment eliminates this risk.
2. *Hydrocephalus*. Acute hydrocephalus can cause stupor and coma and can be mitigated by placement of an external ventricular drain. More often, subacute hydrocephalus may develop over a few days or weeks and causes progressive drowsiness or slowed mentation (abulia) with incontinence. Hydrocephalus is differentiated from cerebral vasospasm with a CT scan, CT angiogram, transcranial Doppler (TCD) ultrasound, or conventional x-ray angiography. Hydrocephalus may clear spontaneously or require temporary ventricular drainage. Chronic hydrocephalus may develop weeks to months after SAH and manifest as gait difficulty, incontinence, or impaired mentation. Subtle signs may be a lack of initiative in conversation or a failure to recover independence.
3. *Vasospasm*. Narrowing of the arteries at the base of the brain following SAH causes symptomatic ischemia and infarction in ~30% of patients and is the major cause of delayed morbidity and death. Signs of ischemia appear 4–14 days after the hemorrhage, most often at 7 days. The severity and distribution of vasospasm determine whether infarction will occur.

   Delayed vasospasm is believed to result from direct effects of clotted blood and its breakdown products on the arteries within the subarachnoid space. In general, the more blood that surrounds the arteries, the greater the chance of symptomatic vasospasm. Spasm of major arteries produces symptoms referable to the appropriate vascular territory (Chap. 370). All of these focal symptoms may present abruptly, fluctuate, or develop over a few days. In most cases, focal spasm is preceded by a decline in mental status.

   Vasospasm can be detected reliably with conventional x-ray angiography, but this invasive procedure is expensive and carries the risk of stroke and other complications. TCD ultrasound is based on the principle that the velocity of blood flow within an artery will rise as the lumen diameter is narrowed. By directing the probe along the MCA and proximal anterior cerebral artery (ACA), carotid terminus, and vertebral and basilar arteries on a daily or every-other-day basis, vasospasm can be reliably detected and treatments initiated to prevent cerebral ischemia (see

below). CT angiography is another method that can detect vasospasm.

Severe cerebral edema in patients with infarction from vasospasm may increase the ICP enough to reduce cerebral perfusion pressure. Treatment may include mannitol, hyperventilation, and hemicraniectomy; moderate hypothermia may have a role as well.

4. *Hyponatremia*. Hyponatremia may be profound and can develop quickly in the first 2 weeks following SAH. There is both natriuresis and volume depletion with SAH, so that patients become both hyponatremic and hypovolemic. Both atrial natriuretic peptide and brain natriuretic peptide have a role in producing this "cerebral salt-wasting syndrome." Typically, it clears over the course of 1–2 weeks and, in the setting of SAH, should not be treated with free-water restriction as this may increase the risk of stroke (see below).

Laboratory Evaluation and Imaging

(Fig. 275-8) The hallmark of aneurysmal rupture is blood in the CSF. More than 95% of cases have enough blood to be visualized on a high-quality noncontrast CT scan obtained within 72 h. If the scan fails to establish the diagnosis of SAH and no mass lesion or obstructive hydrocephalus is found, a lumbar puncture should be performed to establish the presence of subarachnoid blood. Lysis of the red blood cells and subsequent conversion of hemoglobin to bilirubin stains the spinal fluid yellow within 6–12 h. This xanthochromic spinal fluid peaks in intensity at 48 h and lasts for 1–4 weeks, depending on the amount of subarachnoid blood.

Figure 275-8

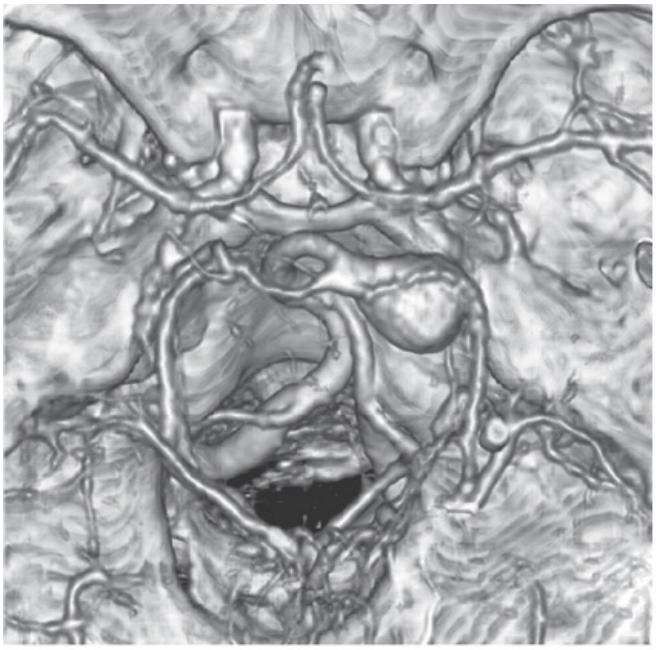

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

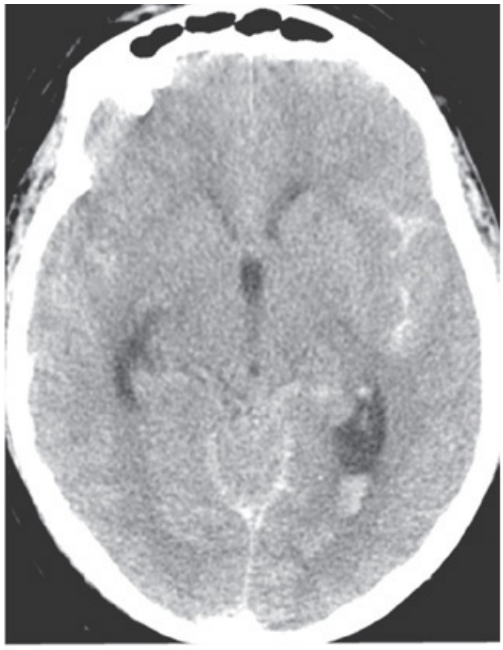

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Subarachnoid hemorrhage. ***A.*** CT angiography revealing an aneurysm of the left superior cerebellar artery. ***B.*** Noncontrast CT scan at the level of the third ventricle revealing subarachnoid blood (*bright*) in the left sylvian fissure and within the left lateral ventricle. ***C.*** Conventional anteroposterior x-ray angiogram of the right vertebral and basilar artery showing the large aneurysm. ***D.*** Conventional angiogram following coil embolization of the aneurysm, whereby the aneurysm body is filled with platinum coils delivered through a microcatheter navigated from the femoral artery into the aneurysm neck.

The extent and location of subarachnoid blood on noncontrast CT scan help locate the underlying aneurysm, identify the cause of any neurologic deficit, and predict delayed vasospasm. A high incidence of symptomatic vasospasm in the MCA and ACA has been found when early CT scans show subarachnoid clots >5 x 3 mm in the basal cisterns or layers of blood >1 mm thick in the cerebral fissures. CT scans less reliably predict vasospasm in the vertebral, basilar, or posterior cerebral arteries.

Lumbar puncture prior to an imaging procedure is indicated only if a CT scan is not available at the time of the suspected SAH. Once the diagnosis of hemorrhage from a ruptured saccular aneurysm is suspected, four-vessel conventional x-ray angiography (both carotids and both vertebrals) is generally performed to localize and define the anatomic details of the aneurysm and to determine if other unruptured aneurysms exist (Fig. 275-8*C*). At some centers, the ruptured aneurysm can be treated using endovascular techniques at the time of the initial angiogram as a way to expedite treatment and minimize the number of invasive procedures. CT angiography is an alternative method for locating the aneurysm and may be sufficient to plan definitive therapy.

Close monitoring (daily or twice daily) of electrolytes is important because hyponatremia can occur precipitously during the first 2 weeks following SAH (see above).

The electrocardiogram (ECG) frequently shows ST-segment and T-wave changes similar to those associated with cardiac ischemia. Prolonged QRS complex, increased QT interval, and prominent "peaked" or deeply inverted symmetric T waves are usually secondary to the intracranial hemorrhage. There is evidence that structural myocardial lesions produced by circulating catecholamines and excessive discharge of sympathetic neurons may occur after SAH, causing these ECG changes and a reversible cardiomyopathy sufficient to cause shock or congestive heart failure. Echocardiography reveals a pattern of regional wall motion abnormalities that follow the distribution of sympathetic nerves rather than the major coronary arteries, with relative sparing of the ventricular wall apex. The sympathetic nerves themselves appear to be injured by direct toxicity from the excessive catecholamine release. An asymptomatic troponin elevation is common. Serious ventricular dysrhythmias are unusual.

Treatment: Subarachnoid Hemorrhage

Early aneurysm repair prevents rerupture and allows the safe application of techniques to improve blood flow (e.g., induced hypertension and hypervolemia) should symptomatic vasospasm develop. An aneurysm can be "clipped" by a neurosurgeon or "coiled" by an endovascular surgeon. Surgical repair

involves placing a metal clip across the aneurysm neck, thereby immediately eliminating the risk of rebleeding. This approach requires craniotomy and brain retraction, which is associated with neurologic morbidity. Endovascular techniques involve placing platinum coils, or other embolic material, within the aneurysm via a catheter that is passed from the femoral artery. The aneurysm is packed tightly to enhance thrombosis and over time is walled off from the circulation (Fig. 275-8*D*). The only prospective randomized trial of surgery versus endovascular treatment for ruptured aneurysm, the International Subarachnoid Aneurysm Trial (ISAT), was terminated early when 24% of patients treated with endovascular therapy were dead or dependent at 1 year compared to 31% treated with surgery, a significant 23% relative reduction. After 5 years, risk of death was lower in the coiling group, although the proportion of survivors who were independent was the same in both groups. Risk of rebleeding was low, but more common in the coiling group. Also, because some aneurysms have a morphology that is not amenable to endovascular treatment, surgery remains an important treatment option. Centers that combine both endovascular and neurosurgical expertise likely offer the best outcomes for patients, and there are reliable data showing that centers that specialize in aneurysm treatment have improved mortality rates.

The medical management of SAH focuses on protecting the airway, managing blood pressure before and after aneurysm treatment, preventing rebleeding prior to treatment, managing vasospasm, treating hydrocephalus, treating hyponatremia, and preventing pulmonary embolus.

Intracranial hypertension following aneurysmal rupture occurs secondary to subarachnoid blood, parenchymal hematoma, acute hydrocephalus, or loss of vascular autoregulation. Patients who are stuporous should undergo emergent ventriculostomy to measure ICP and to treat high ICP in order to prevent cerebral ischemia. Medical therapies designed to combat raised ICP (e.g., mild hyperventilation, mannitol, and sedation) can also be used as needed. High ICP refractory to treatment is a poor prognostic sign.

Prior to definitive treatment of the ruptured aneurysm, care is required to maintain adequate cerebral perfusion pressure while avoiding excessive elevation of arterial pressure. If the patient is alert, it is reasonable to lower the blood pressure to normal using nicardipine, labetolol, or esmolol. If the patient has a depressed level of consciousness, ICP should be measured and the cerebral perfusion pressure targeted to 60–70 mmHg. If headache or neck pain is severe, mild sedation and analgesia are prescribed. Extreme sedation is avoided because it can obscure changes in neurologic status. Adequate hydration is necessary to avoid a decrease in blood volume predisposing to brain ischemia.

Seizures are uncommon at the onset of aneurysmal rupture. The quivering, jerking, and extensor posturing that often accompany loss of consciousness with SAH are probably related to the sharp rise in ICP rather than seizure. However, anticonvulsants are sometimes given as prophylactic therapy since a seizure could theoretically promote rebleeding.

Glucocorticoids may help reduce the head and neck ache caused by the irritative effect of the subarachnoid blood. There is no good evidence that they reduce cerebral edema, are neuroprotective, or reduce vascular injury, and their routine use therefore is not recommended.

Antifibrinolytic agents are not routinely prescribed but may be considered in patients in whom aneurysm treatment cannot proceed immediately. They are associated with a reduced incidence of aneurysmal rerupture but may also increase the risk of delayed cerebral infarction and deep-vein thrombosis (DVT).

Vasospasm remains the leading cause of morbidity and mortality following aneurysmal SAH. Treatment with the calcium channel antagonist nimodipine (60 mg PO every 4 h) improves outcome, perhaps by preventing ischemic injury rather than reducing the risk of vasospasm. Nimodipine can cause significant hypotension in some patients, which may worsen cerebral ischemia in patients with vasospasm. Symptomatic cerebral vasospasm can also be treated by increasing the cerebral perfusion pressure by raising mean arterial pressure through plasma volume expansion and the judicious use of IV vasopressor agents, usually phenylephrine or norepinephrine. Raised perfusion pressure has been associated with clinical improvement in many patients, but high arterial pressure may promote rebleeding in unprotected aneurysms. Treatment with induced hypertension and hypervolemia generally requires monitoring of arterial and central venous pressures; it is best to infuse pressors through a central venous line as well. Volume expansion helps prevent hypotension, augments cardiac output, and reduces blood viscosity by reducing the hematocrit. This method is called "triple-H" (hypertension, hemodilution, and hypervolemic) therapy.

If symptomatic vasospasm persists despite optimal medical therapy, intraarterial vasodilators and percutaneous transluminal angioplasty are considered. Vasodilatation by direct angioplasty appears to be permanent, allowing triple-H therapy to be tapered sooner. The pharmacologic vasodilators (verapamil and nicardipine) do not last more than about 24 h, and therefore multiple treatments may be required until the subarachnoid blood is reabsorbed. Although intraarterial papaverine is an effective vasodilator, there is evidence that papaverine may be neurotoxic, so its use should generally be avoided.

Acute hydrocephalus can cause stupor or coma. It may clear spontaneously or require temporary

ventricular drainage. When chronic hydrocephalus develops, ventricular shunting is the treatment of choice.

Free-water restriction is contraindicated in patients with SAH at risk for vasospasm because hypovolemia and hypotension may occur and precipitate cerebral ischemia. Many patients continue to experience a decline in serum sodium despite receiving parenteral fluids containing normal saline. Frequently, supplemental oral salt coupled with normal saline will mitigate hyponatremia, but often patients also require hypertonic saline. Care must be taken not to correct serum sodium too quickly in patients with marked hyponatremia of several days' duration, as central pontine myelinolysis may occur.

All patients should have pneumatic compression stockings applied to prevent pulmonary embolism. Unfractionated heparin administered subcutaneously for DVT prophylaxis can be initiated immediately following endovascular treatment and within days following craniotomy and surgical clipping and is a useful adjunct to pneumatic compression stockings. Treatment of pulmonary embolus depends on whether the aneurysm has been treated and whether or not the patient has had a craniotomy. Systemic anticoagulation with heparin is contraindicated in patients with ruptured and untreated aneurysms. It is a relative contraindication following craniotomy for several days, and it may delay thrombosis of a coiled aneurysm. Following craniotomy, use of inferior vena cava filters is preferred to prevent further pulmonary emboli, while systemic anticoagulation with heparin is preferred following successful endovascular treatment.

FURTHER READINGS

Latronico N et al: Neuromuscular sequelae of critical illness. Curr Opin Crit Care 11:381, 2005[PMID: 16015120]

Liou AK et al: To die or not to die for neurons in ischemia, traumatic brain injury and epilepsy: A review on the stress-activated signaling pathways and apoptotic pathways. Prog Neurobiol 69:103, 2003[PMID: 12684068]

Molyneux A et al: Risk of recurrent subarachnoid haemorrhage, death, or dependence and standardised mortality ratios after clipping or coiling of an intracranial aneurysm in the International Subarachnoid Aneurysm Trial (ISAT): Long-term follow-up. Lancet Neurol 8:427, 2009[PMID: 19329361]

Morandi A et al: Delirium in the intensive care unit. Int Rev Psych 21:43, 2009[PMID: 19219712]

Nolan JP et al: Therapeutic hypothermia after cardiac arrest: An advisory statement by the Advanced Life Support Task Force of the International Liaison Committee on Resuscitation. Circulation 108:118, 2003[PMID: 12847056]

Posner JB et al: *Plum and Posner's Diagnosis of Stupor and Coma* , 4th ed. New York, Oxford University Press, 2007

Rossetti AO et al: Prognostication after cardiac arrest and hypothermia: A prospective study. Ann Neurol 67:301, 2010[PMID: 20373341]

Wijdicks EFM et al: Practice parameter: Prediction of outcome in comatose survivors after cardiopulmonary resuscitation (an evidence-based review). Neurology 67:203, 2006[PMID: 16864809]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ONCOLOGIC EMERGENCIES: INTRODUCTION

Emergencies in patients with cancer may be classified into three groups: pressure or obstruction caused by a space-occupying lesion, metabolic or hormonal problems (paraneoplastic syndromes, Chap. 100), and treatment-related complications.

STRUCTURAL-OBSTRUCTIVE ONCOLOGIC EMERGENCIES

Superior Vena Cava Syndrome

Superior vena cava syndrome (SVCS) is the clinical manifestation of superior vena cava (SVC) obstruction, with severe reduction in venous return from the head, neck, and upper extremities. Malignant tumors, such as lung cancer, lymphoma, and metastatic tumors, are responsible for the majority of SVCS cases. With the expanding use of intravascular devices (e.g., permanent central venous access catheters, pacemaker/defibrillator leads), the prevalence of benign causes of SVCS is increasing now, accounting for at least 40% of cases. Lung cancer, particularly of small cell and squamous cell histologies, accounts for approximately 85% of all cases of malignant origin. In young adults, malignant lymphoma is a leading cause of SVCS. Hodgkin's lymphoma involves the mediastinum more commonly than other lymphomas but rarely causes SVCS. When SVCS is noted in a young man with a mediastinal mass, the differential diagnosis is lymphoma vs primary mediastinal germ cell tumor. Metastatic cancers to the mediastinum, such as testicular and breast carcinomas, account for a small proportion of cases. Other causes include benign tumors, aortic aneurysm, thyromegaly, thrombosis, and fibrosing mediastinitis from prior irradiation, histoplasmosis, or Behçet's syndrome. SVCS as the initial manifestation of Behçet's syndrome may be due to inflammation of the SVC associated with thrombosis.

Patients with SVCS usually present with neck and facial swelling (especially around the eyes), dyspnea, and cough. Other symptoms include hoarseness, tongue swelling, headaches, nasal congestion, epistaxis, hemoptysis, dysphagia, pain, dizziness, syncope, and lethargy. Bending forward or lying down may aggravate the symptoms. The characteristic physical findings are dilated neck veins; an increased number of collateral veins covering the anterior chest wall; cyanosis; and edema of the face, arms, and chest. More severe cases include proptosis, glossal and laryngeal edema, and obtundation. The clinical picture is milder if the obstruction is located above the azygos vein. Symptoms are usually progressive, but in some cases they may improve as collateral circulation develops.

Signs and symptoms of cerebral and/or laryngeal edema, though rare, are associated with a poorer prognosis and require urgent evaluation. Seizures are more likely related to brain metastases than to cerebral edema from venous occlusion. Patients with small cell lung cancer and SVCS have a higher incidence of brain metastases than those without SVCS.

Cardiorespiratory symptoms at rest, particularly with positional changes, suggest significant airway and vascular obstruction and limited physiologic reserve. Cardiac arrest or respiratory failure can occur, particularly in patients receiving sedatives or undergoing general anesthesia.

Rarely, esophageal varices may develop. These are "downhill" varices based on the direction of blood flow from cephalad to caudad (in contrast to "uphill" varices associated with caudad to cephalad flow from portal hypertension). If the obstruction to the SVC is proximal to the azygous vein, varices develop in the upper one-third of the esophagus. If the obstruction involves or is distal to the azygous vein, varices occur in the entire length of the esophagus. Variceal bleeding may be a late complication of chronic SVCS.

The diagnosis of SVCS is a clinical one. The most significant chest radiographic finding is widening of the superior mediastinum, most commonly on the right side. Pleural effusion occurs in only 25% of patients, often on the right side. The majority of these effusions are exudative and occasionally chylous. However, a normal chest radiograph is still compatible with the diagnosis if other characteristic findings are present. CT provides the most reliable view of the mediastinal anatomy. The diagnosis of SVCS requires diminished or absent opacification of central venous structures with prominent collateral venous circulation. MRI has no advantages over CT. Invasive procedures, including bronchoscopy, percutaneous needle biopsy, mediastinoscopy, and even thoracotomy, can be performed by a skilled clinician without any major risk of bleeding. For patients with a known cancer, a detailed workup usually is not necessary, and appropriate treatment may be started after obtaining a CT scan of the thorax. For those with no history of malignancy, a detailed evaluation is essential to rule out benign causes and determine a specific diagnosis to direct the appropriate therapy.

TREATMENT: SUPERIOR VENA CAVA SYNDROME

The one potentially life-threatening complication of a superior mediastinal mass is tracheal obstruction. Upper airway obstruction demands emergent therapy. Diuretics with a low-salt diet, head elevation, and oxygen may produce temporary symptomatic relief. Glucocorticoids may be useful at shrinking lymphoma masses; they are of no benefit in patients with lung cancer.

Radiation therapy is the primary treatment for SVCS caused by non-small cell lung cancer and other metastatic solid tumors. Chemotherapy is effective when the underlying cancer is small cell carcinoma of the lung, lymphoma, or germ cell tumor. SVCS recurs in 10–30% of patients; it may be palliated with the use of intravascular self-expanding stents **(Fig. 276-1)**. Early stenting may be necessary in patients with severe symptoms; however, the prompt increase in venous return after stenting may precipitate heart failure and pulmonary edema. Surgery may

provide immediate relief for patients in whom a benign process is the cause.

Figure 276-1

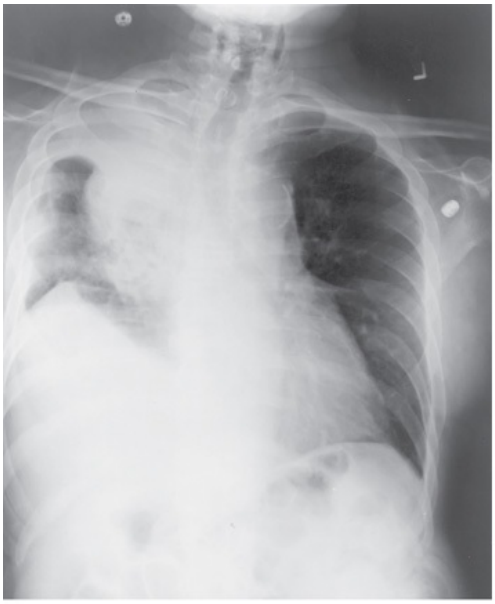

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

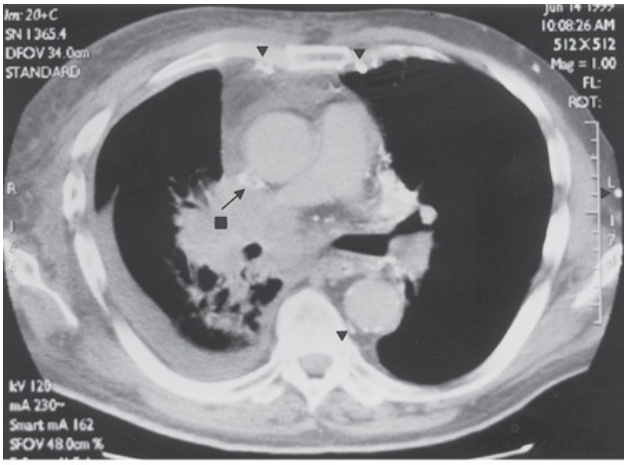

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

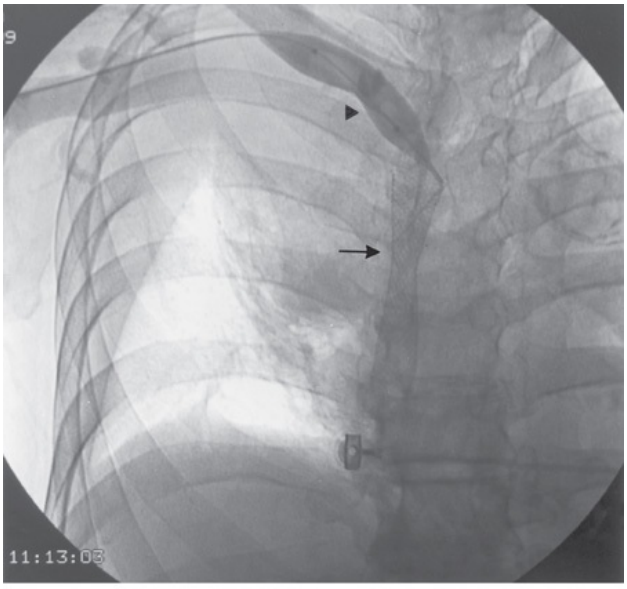

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Superior vena cava syndrome.*A.* Chest radiographs of a 59-year-old man with recurrent SVCS caused by non-small cell lung cancer showing right paratracheal mass with right pleural effusion. ***B.*** CT of same patient demonstrating obstruction of SVC with thrombosis (*arrow*) by the lung cancer (*square*) and collaterals (*arrowheads*). ***C.*** Balloon angioplasty (*arrowhead*) with Wallstent (*arrow*) in same patient.

Clinical improvement occurs in most patients, although this improvement may be due to the development of adequate collateral circulation. The mortality associated with SVCS does not relate to caval obstruction but rather to the underlying cause.

Svcs and Central Venous Catheters in Adults

The use of long-term central venous catheters has become common practice in patients with cancer. Major vessel thrombosis may occur. In these cases, catheter removal should be combined with anticoagulation to prevent embolization. SVCS in this setting, if detected early, can be treated by fibrinolytic therapy without sacrificing the catheter. The routine use of low-dose warfarin or low-molecular-weight heparin to prevent thrombosis related to permanent central venous access catheters in cancer patients is not recommended.

Pericardial Effusion/Tamponade

Malignant pericardial disease is found at autopsy in 5–10% of patients with cancer, most frequently with lung cancer, breast cancer, leukemias, and lymphomas. Cardiac tamponade as the initial presentation of extrathoracic malignancy is rare. The origin is not malignancy in about 50% of cancer patients with symptomatic pericardial disease, but it can be related to irradiation, drug-induced pericarditis, hypothyroidism, idiopathic pericarditis, infection, or autoimmune diseases. Two types of radiation pericarditis occur: an acute inflammatory, effusive pericarditis occurring within months of irradiation, which usually resolves spontaneously, and a chronic effusive pericarditis that may appear up to 20 years after radiation therapy and is accompanied by a thickened pericardium.

Most patients with pericardial metastasis are asymptomatic. However, the common symptoms are dyspnea, cough, chest pain, orthopnea, and weakness. Pleural effusion, sinus tachycardia, jugular venous distention, hepatomegaly, peripheral edema, and cyanosis are the most frequent physical findings. Relatively specific diagnostic findings, such as paradoxical pulse, diminished heart sounds, pulsus alternans (pulse waves alternating between those of greater and lesser amplitude with successive beats), and friction rub are less common than with nonmalignant pericardial disease. Chest radiographs and ECG reveal abnormalities in 90% of patients, but half of these abnormalities are nonspecific. Echocardiography is the most helpful diagnostic test. Pericardial fluid may be serous, serosanguineous, or hemorrhagic, and cytologic examination of pericardial fluid is diagnostic in most patients. Cancer patients with pericardial effusion containing malignant cells on cytology have a very poor survival, about 7 weeks.

TREATMENT: PERICARDIAL EFFUSION/TAMPONADE

Pericardiocentesis with or without the introduction of sclerosing agents, the creation of a pericardial window, complete pericardial stripping, cardiac irradiation, or systemic chemotherapy are effective treatments. Acute pericardial tamponade with life-threatening hemodynamic instability requires immediate drainage of fluid. This can be quickly achieved by pericardiocentesis. The recurrence rate after percutaneous catheter drainage is about 20%. Sclerotherapy (pericardial instillation of bleomycin, mitomycin C, or tetracycline) may decrease recurrences. Alternatively, subxiphoid pericardiotomy can be performed in 45 min under local anesthesia. Thoracoscopic pericardial fenestration can be employed for benign causes; however, 60% of malignant pericardial effusions recur after this procedure.

Intestinal Obstruction

Intestinal obstruction and reobstruction are common problems in patients with advanced cancer, particularly colorectal or ovarian carcinoma. However, other cancers, such as lung or breast cancer and melanoma, can metastasize within the abdomen, leading to intestinal obstruction. Typically, obstruction occurs at multiple sites in

peritoneal carcinomatosis. Melanoma has a predilection to involve the small bowel; this involvement may be isolated and resection may result in prolonged survival. Intestinal pseudoobstruction is caused by infiltration of the mesentery or bowel muscle by tumor, involvement of the celiac plexus, or paraneoplastic neuropathy in patients with small cell lung cancer. Paraneoplastic neuropathy is associated with IgG antibodies reactive to neurons of the myenteric and submucosal plexuses of the jejunum and stomach. Ovarian cancer can lead to authentic luminal obstruction or to pseudoobstruction that results when circumferential invasion of a bowel segment arrests the forward progression of peristaltic contractions.

The onset of obstruction is usually insidious. Pain is the most common symptom and is usually colicky in nature. Pain can also be due to abdominal distention, tumor masses, or hepatomegaly. Vomiting can be intermittent or continuous. Patients with complete obstruction usually have constipation. Physical examination may reveal abdominal distention with tympany, ascites, visible peristalsis, high-pitched bowel sounds, and tumor masses. Erect plain abdominal films may reveal multiple air-fluid levels and dilation of the small or large bowel. Acute cecal dilation to >12–14 cm is considered a surgical emergency because of the high likelihood of rupture. CT scan is useful in differentiating benign from malignant causes of obstruction in patients who have undergone surgery for malignancy. Malignant obstruction is suggested by a mass at the site of obstruction or prior surgery, adenopathy, or an abrupt transition zone and irregular bowel thickening at the obstruction site. Benign obstruction is more likely when CT shows mesenteric vascular changes, a large volume of ascites, or a smooth transition zone and smooth bowel thickening at the obstruction site. The prognosis for the patient with cancer who develops intestinal obstruction is poor; median survival is 3–4 months. About 25–30% of patients are found to have intestinal obstruction due to causes other than cancer. Adhesions from previous operations are a common benign cause. Ileus induced by vinca alkaloids, narcotics, or other drugs is another reversible cause.

TREATMENT: INTESTINAL OBSTRUCTION

The management of intestinal obstruction in patients with advanced malignancy depends on the extent of the underlying malignancy and the functional status of the major organs. The initial management should include surgical evaluation. Operation is not always successful and may lead to further complications with a substantial mortality rate (10–20%). Laparoscopy can diagnose and treat malignant bowel obstruction in some cases. Self-expanding metal stents placed in the gastric outlet, duodenum, proximal jejunum, colon, or rectum may palliate obstructive symptoms at those sites without major surgery. Patients known to have advanced intraabdominal malignancy should receive a prolonged course of conservative management, including nasogastric decompression. Percutaneous endoscopic or surgical gastrostomy tube placement is an option for palliation of nausea and vomiting, the so-called "venting gastrostomy." Treatment with antiemetics, antispasmodics, and analgesics may allow patients to remain outside the hospital. Octreotide may relieve obstructive symptoms through its inhibitory effect on gastrointestinal secretion.

Urinary Obstruction

Urinary obstruction may occur in patients with prostatic or gynecologic malignancies, particularly cervical carcinoma; metastatic disease from other primary sites such as carcinomas of the breast, stomach, lung, colon, and pancreas; or lymphomas. Radiation therapy to pelvic tumors may cause fibrosis and subsequent ureteral obstruction. Bladder outlet obstruction is usually due to prostate and cervical cancers and may lead to bilateral hydronephrosis and renal failure.

Flank pain is the most common symptom. Persistent urinary tract infection, persistent proteinuria, or hematuria in patients with cancer should raise suspicion of ureteral obstruction. Total anuria and/or anuria alternating with polyuria may occur. A slow, continuous rise in the serum creatinine level necessitates immediate evaluation. Renal ultrasound is the safest and cheapest way to identify hydronephrosis. The function of an obstructed kidney can be evaluated by a nuclear scan. CT scan can reveal the point of obstruction and identify a retroperitoneal mass or adenopathy.

TREATMENT: URINARY OBSTRUCTION

Obstruction associated with flank pain, sepsis, or fistula formation is an indication for immediate palliative urinary diversion. Internal ureteral stents can be placed under local anesthesia. Percutaneous nephrostomy offers an alternative approach for drainage. In the case of bladder outlet obstruction due to malignancy, a suprapubic cystostomy can be used for urinary drainage.

Malignant Biliary Obstruction

This common clinical problem can be caused by a primary carcinoma arising in the pancreas, ampulla of Vater, bile duct, or liver or by metastatic disease to the periductal lymph nodes or liver parenchyma. The most common metastatic tumors causing biliary obstruction are gastric, colon, breast, and lung cancers. Jaundice, light-colored stools, dark urine, pruritus, and weight loss due to malabsorption are usual symptoms. Pain and secondary infection are uncommon in malignant biliary obstruction. Ultrasound, CT scan, or percutaneous transhepatic or endoscopic retrograde cholangiography will identify the site and nature of the biliary obstruction.

TREATMENT: MALIGNANT BILIARY OBSTRUCTION

Palliative intervention is indicated only in patients with disabling pruritus resistant to medical treatment, severe malabsorption, or infection. Stenting under radiographic control, surgical bypass, or radiation therapy with or without chemotherapy may alleviate the obstruction. The choice of therapy should be based on the site of obstruction (proximal vs distal), the type of tumor (sensitive to radiotherapy, chemotherapy, or neither), and the general condition of the patient. In the absence of pruritus, biliary obstruction may be a largely asymptomatic cause of death.

Spinal Cord Compression

Malignant spinal cord compression (MSCC) is defined as compression of the spinal cord and/or cauda equina by an extradural tumor mass. The minimum radiologic evidence for cord compression is indentation of the theca at the level of clinical features. Spinal cord compression occurs in 5–10% of patients with cancer. Epidural tumor is the first manifestation of malignancy in about 10% of patients. The underlying cancer is usually identified during the

initial evaluation; lung cancer is the most common cause of MSCC.

Metastatic tumor involves the vertebral column more often than any other part of the bony skeleton. Lung, breast, and prostate cancer are the most frequent offenders. Multiple myeloma also has a high incidence of spine involvement. Lymphomas, melanoma, renal cell cancer, and genitourinary cancers also cause cord compression. The thoracic spine is the most common site (70%), followed by the lumbosacral spine (20%) and the cervical spine (10%). Involvement of multiple sites is most frequent in patients with breast and prostate carcinoma. Cord injury develops when metastases to the vertebral body or pedicle enlarge and compress the underlying dura. Another cause of cord compression is direct extension of a paravertebral lesion through the intervertebral foramen. These cases usually involve a lymphoma, myeloma, or pediatric neoplasm. Parenchymal spinal cord metastasis due to hematogenous spread is rare. Intramedullary metastases can be seen in lung cancer, multiple myeloma, renal cell cancer, and breast cancer and are frequently associated with brain metastases and leptomeningeal disease.

Expanding extradural tumors induce injury through several mechanisms. Obstruction of the epidural venous plexus leads to edema. Local production of inflammatory cytokines enhances blood flow and edema formation. Compression compromises blood flow leading to ischemia. Production of vascular endothelial growth factor is associated with spinal cord hypoxia and has been implicated as a potential cause of damage after spinal cord injury.

The most common initial symptom in patients with spinal cord compression is localized back pain and tenderness due to involvement of vertebrae by tumor. Pain is usually present for days or months before other neurologic findings appear. It is exacerbated by movement and by coughing or sneezing. It can be differentiated from the pain of disk disease by the fact that it worsens when the patient is supine. Radicular pain is less common than localized back pain and usually develops later. Radicular pain in the cervical or lumbosacral areas may be unilateral or bilateral. Radicular pain from the thoracic roots is often bilateral and is described by patients as a feeling of tight, band-like constriction around the thorax and abdomen. Typical cervical radicular pain radiates down the arm; in the lumbar region, the radiation is down the legs. *Lhermitte's*, a tingling or electric sensation down the back and upper and lower limbs upon flexing or extending the neck, may be an early sign of cord compression. Loss of bowel or bladder control may be the presenting symptom but usually occurs late in the course. Occasionally patients present with ataxia of gait without motor and sensory involvement due to involvement of the spinocerebellar tract.

On physical examination, pain induced by straight leg raising, neck flexion, or vertebral percussion may help to determine the level of cord compression. Patients develop numbness and paresthesias in the extremities or trunk. Loss of sensibility to pinprick is as common as loss of sensibility to vibration or position. The upper limit of the zone of sensory loss is often one or two vertebrae below the site of compression. Motor findings include weakness, spasticity, and abnormal muscle stretching. An extensor plantar reflex reflects significant compression. Deep tendon reflexes may be brisk. Motor and sensory loss usually precedes sphincter disturbance. Patients with autonomic dysfunction may present with decreased anal tonus, decreased perineal sensibility, and a distended bladder. The absence of the anal wink reflex or the bulbocavernosus reflex confirms cord involvement. In doubtful cases, evaluation of postvoiding urinary residual volume can be helpful. A residual volume of >150 mL suggests bladder dysfunction. Autonomic dysfunction is an unfavorable prognostic factor. Patients with progressive neurologic symptoms should have frequent neurologic examinations and rapid therapeutic intervention. Other illnesses that may mimic cord compression include osteoporotic vertebral collapse, disk disease, pyogenic abscess or vertebral tuberculosis, radiation myelopathy, neoplastic leptomeningitis, benign tumors, epidural hematoma, and spinal lipomatosis.

Cauda equina syndrome is characterized by low back pain; diminished sensation over the buttocks, posterior-superior thighs, and perineal area in a saddle distribution; rectal and bladder dysfunction; sexual impotence; absent bulbocavernous, patellar, and Achilles' reflexes; and variable amount of lower-extremity weakness. This reflects compression of nerve roots as they form the cauda equina after leaving the spinal cord.

Patients with cancer who develop back pain should be evaluated for spinal cord compression as quickly as possible **(Fig. 276-2)**. Treatment is more often successful in patients who are ambulatory and still have sphincter control at the time treatment is initiated. Patients should have a neurologic examination and plain films of the spine. Those whose physical examination suggests cord compression should receive dexamethasone (6 mg intravenously every 6 h), starting immediately.

Figure 276-2

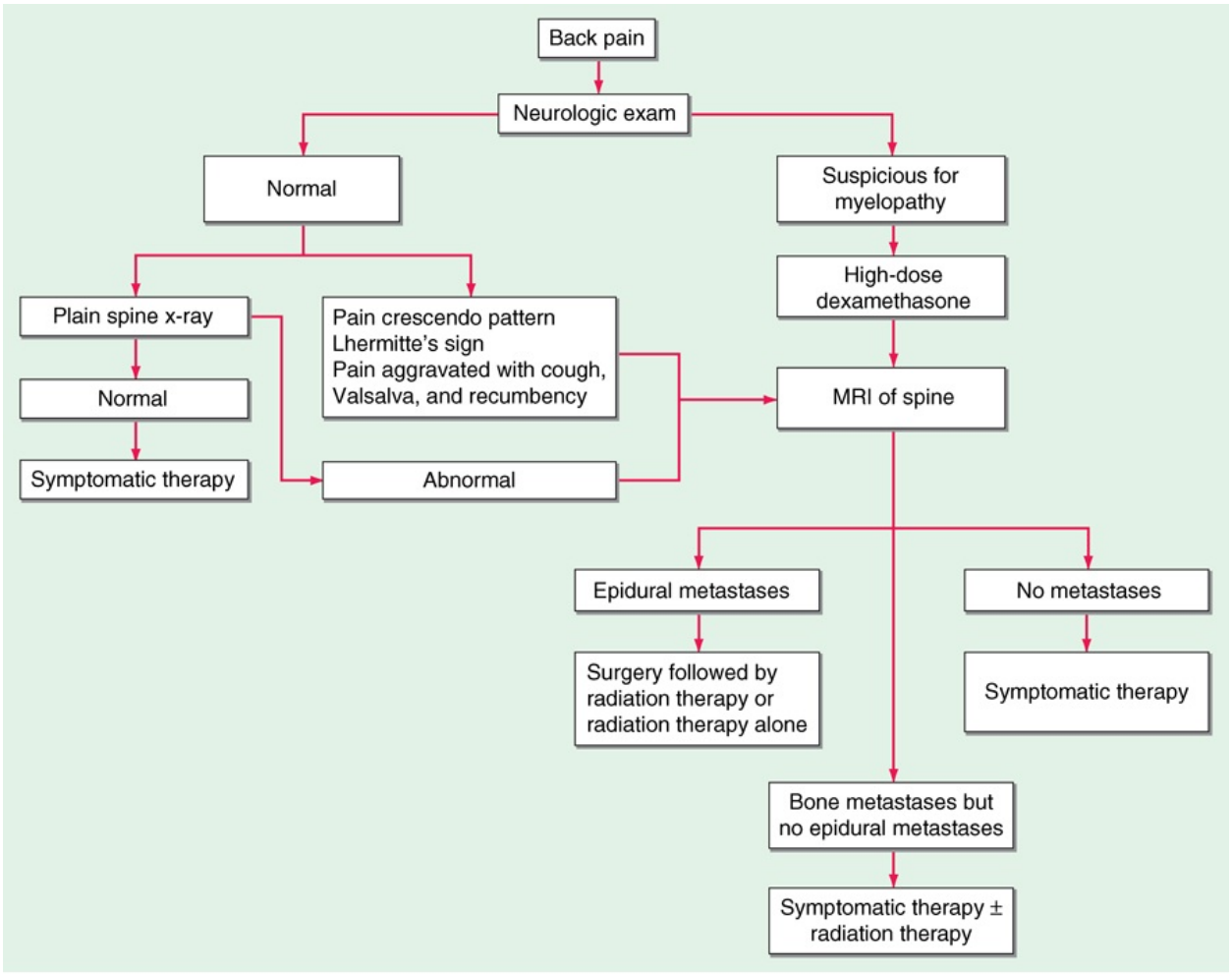

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of cancer patients with back pain.

Erosion of the pedicles (the "winking owl" sign) is the earliest radiologic finding of vertebral tumor. Other radiographic changes include increased intrapedicular distance, vertebral destruction, lytic or sclerotic lesions, scalloped vertebral bodies, and vertebral body collapse. Vertebral collapse is not a reliable indicator of the presence of tumor; about 20% of cases of vertebral collapse, particularly those in older patients and postmenopausal women, are due not to cancer but to osteoporosis. Also, a normal appearance on plain films of the spine does not exclude the diagnosis of cancer. The role of bone scans in the detection of cord compression is not clear; this method is sensitive but less specific than spinal radiography.

The full-length image of the cord provided by MRI is the imaging procedure of choice. Multiple epidural metastases are noted in 25% of patients with cord compression, and their presence influences treatment plans. On T1-weighted images, good contrast is noted between the cord, cerebrospinal fluid, and extradural lesions. Owing to its sensitivity in demonstrating the replacement of bone marrow by tumor, MRI can show which parts of a vertebra are involved by tumor. MRI also visualizes intraspinal extradural masses compressing the cord. T2-weighted images are most useful for the demonstration of intramedullary pathology. Gadolinium-enhanced MRI can help to delineate intramedullary disease. MRI is as good as or better than myelography plus postmyelogram CT scan in detecting metastatic epidural disease with cord compression. Myelography should be reserved for patients who have poor MR images or who cannot undergo MRI promptly. CT scan in conjunction with myelography enhances the detection of small areas of spinal destruction.

In patients with cord compression and an unknown primary tumor, a simple workup including chest radiography, mammography, measurement of prostate-specific antigen, and abdominal CT usually reveals the underlying malignancy.

TREATMENT: SPINAL CORD COMPRESSION

The treatment of patients with spinal cord compression is aimed at relief of pain and restoration/preservation of neurologic function (Fig. 276-2).

Radiation therapy plus glucocorticoids is generally the initial treatment of choice for most patients with spinal cord compression. Up to 75% of patients treated when still ambulatory remain ambulatory, but only 10% of patients with paraplegia recover walking capacity. Indications for surgical intervention include unknown etiology, failure of radiation therapy, a radioresistant tumor type (e.g., melanoma or renal cell cancer), pathologic fracture dislocation, and rapidly evolving neurologic symptoms. Laminectomy is done for tissue diagnosis and for the removal of posteriorly localized epidural deposits in the absence of vertebral body disease. Because most cases of epidural spinal cord compression are due to anterior or anterolateral extradural disease, resection of the anterior vertebral

body along with the tumor, followed by spinal stabilization, has achieved good results. A randomized trial showed that patients who underwent an operation followed by radiotherapy (within 14 days) retained the ability to walk significantly longer than those treated with radiotherapy alone. Surgically treated patients also maintained continence and neurologic function significantly longer than patients in the radiation group. The length of survival was not significantly different in the two groups, although there was a trend toward longer survival in the surgery group. The study drew some criticism for the poorer than expected results in the patients who did not go to surgery. However, patients should be evaluated for surgery if they are expected to survive longer than 3 months. Conventional radiotherapy has a role after surgery. Chemotherapy may have a role in patients with chemosensitive tumors who have had prior radiotherapy to the same region and who are not candidates for surgery. Most patients with prostate cancer who develop cord compression have already had hormonal therapy; however, for those who have not, androgen deprivation is combined with surgery and radiotherapy.

Patients with metastatic vertebral tumors may benefit from percutaneous vertebroplasty or kyphoplasty, the injection of acrylic cement into a collapsed vertebra to stabilize the fracture. Pain palliation is common, and local antitumor effects have been noted. Cement leakage may cause symptoms in about 10% of patients. Bisphosphonates may be helpful in prevention of SCC in patients with bony involvement.

The histology of the tumor is an important determinant of both recovery and survival. Rapid onset and progression of signs and symptoms are poor prognostic features.

Increased Intracranial Pressure

About 25% of patients with cancer die with intracranial metastases. The cancers that most often metastasize to the brain are lung and breast cancers and melanoma. Brain metastases often occur in the presence of systemic disease, and they frequently cause major symptoms, disability, and early death. The initial presentation of brain metastases from a previously unknown primary cancer is common. Lung cancer is most commonly the primary malignancy. Chest CT scans and brain MRI as the initial diagnostic studies can identify a biopsy site in most patients.

The signs and symptoms of a metastatic brain tumor are similar to those of other intracranial expanding lesions: headache, nausea, vomiting, behavioral changes, seizures, and focal, progressive neurologic changes. Occasionally the onset is abrupt, resembling a stroke, with the sudden appearance of headache, nausea, vomiting, and neurologic deficits. This picture is usually due to hemorrhage into the metastasis. Melanoma, germ cell tumors, and renal cell cancers have a particularly high incidence of intracranial bleeding. The tumor mass and surrounding edema may cause obstruction of the circulation of cerebrospinal fluid, with resulting hydrocephalus. Patients with increased intracranial pressure may have papilledema with visual disturbances and neck stiffness. As the mass enlarges, brain tissue may be displaced through the fixed cranial openings, producing various herniation syndromes.

CT scan and MRI are equally effective in the diagnosis of brain metastases. CT scan with contrast should be used as a screening procedure. The CT scan shows brain metastases as multiple enhancing lesions of various sizes with surrounding areas of low-density edema. If a single lesion or no metastases are visualized by contrast-enhanced CT, MRI of the brain should be performed. Gadolinium-enhanced MRI is more sensitive than CT at revealing meningeal involvement and small lesions, particularly in the brainstem or cerebellum.

Intracranial hypertension secondary to tretinoin therapy has been reported.

TREATMENT: INCREASED INTRACRANIAL PRESSURE

Dexamethasone is the best initial treatment for all symptomatic patients with brain metastases. If signs and symptoms of brain herniation (particularly headache, drowsiness, and papilledema) are present, the patient should be intubated and hyperventilated to maintain PCO_2 between 25 and 30 mmHg and should receive infusions of mannitol (1–1.5 g/kg) every 6 h. Other measures include head elevation, fluid restriction, and hypertonic saline with diuretics. Patients with multiple lesions should receive whole-brain radiation. Patients with a single brain metastasis and with controlled extracranial disease may be treated with surgical excision followed by whole-brain radiation therapy, especially if they are younger than 60 years. Radioresistant tumors should be resected if possible. Stereotactic radiosurgery is an effective treatment for inaccessible or recurrent lesions. With a gamma knife or linear accelerator, multiple small, well-collimated beams of ionizing radiation destroy lesions seen on MRI. Some patients with increased intracranial pressure associated with hydrocephalus may benefit from shunt placement. If neurologic deterioration is not reversed with medical therapy, ventriculotomy to remove cerebrospinal fluid (CSF) or craniotomy to remove tumors or hematomas may be necessary.

Neoplastic Meningitis

Tumor involving the leptomeninges is a complication of both primary central nervous system (CNS) tumors and tumors that metastasize to the CNS. The incidence is estimated at 3–8% of patients with cancer. Melanoma, breast and lung cancer, lymphoma (including AIDS-associated), and acute leukemia are the most common causes. Synchronous intraparenchymal brain metastases are evident in 11–31% of patients with neoplastic meningitis.

Patients typically present with multifocal neurologic signs and symptoms, including headache, gait abnormality, mental changes, nausea, vomiting, seizures, back or radicular pain, and limb weakness. Signs include cranial nerve palsies, extremity weakness, paresthesia, and decreased deep tendon reflexes.

Diagnosis is made by demonstrating malignant cells in the CSF; however, up to 40% of patients may have false-negative CSF cytology. An elevated CSF protein level is nearly always present (except in HTLV-1–associated adult T cell leukemia). Patients with neurologic signs and symptoms consistent with neoplastic meningitis who have a negative CSF cytology but an elevated CSF protein level should have the spinal tap repeated at least three times for cytologic examination before the diagnosis is rejected. MRI findings suggestive of neoplastic meningitis include leptomeningeal, subependymal, dural, or cranial nerve enhancement; superficial cerebral lesions; and communicating hydrocephalus. Spinal cord imaging by MRI is a necessary component of the evaluation of nonleukemia neoplastic meningitis as ~20% of patients have cord abnormalities, including intradural enhancing nodules that are diagnostic for leptomeningeal involvement. Cauda equina lesions are common, but lesions may be

seen anywhere in the spinal canal. Radiolabeled CSF flow studies are abnormal in up to 70% of patients with neoplastic meningitis; ventricular outlet obstruction, abnormal flow in the spinal canal, or impaired flow over the cerebral convexities may affect distribution of intrathecal chemotherapy resulting in decreased efficacy or increased toxicity. Radiation therapy may correct CSF flow abnormalities before use of intrathecal chemotherapy. Neoplastic meningitis can also lead to intracranial hypertension and hydrocephalus. Placement of a ventriculoperitoneal shunt may effectively palliate symptoms in these patients.

The development of neoplastic meningitis usually occurs in the setting of uncontrolled cancer outside the CNS; thus, prognosis is poor (median survival 10–12 weeks). However, treatment of the neoplastic meningitis may successfully alleviate symptoms and control the CNS spread.

TREATMENT: NEOPLASTIC MENINGITIS

Intrathecal chemotherapy, usually methotrexate, cytarabine, or thiotepa, is delivered by lumbar puncture or by an intraventricular reservoir (Ommaya) three times a week until the CSF is free of malignant cells. Injections are given twice a week for a month and then weekly for a month. An extended-release preparation of cytarabine (Depocyte) has a longer half-life and is more effective than other formulations. Among solid tumors, breast cancer responds best to therapy. Patients with neoplastic meningitis from either acute leukemia or lymphoma may be cured of their CNS disease if the systemic disease can be eliminated.

Seizures

Seizures occurring in a patient with cancer can be caused by the tumor itself, by metabolic disturbances, by radiation injury, by cerebral infarctions, by chemotherapy-related encephalopathies, or by CNS infections. Metastatic disease to the CNS is the most common cause of seizures in patients with cancer. However, seizures occur more frequently in primary brain tumors than in metastatic brain lesions. Seizures are a presenting symptom of CNS metastasis in 6–29% of cases. Approximately 10% of patients with CNS metastasis eventually develop seizures. Tumors that affect the frontal, temporal, and parietal lobes are more commonly associated with seizures than are occipital lesions. The presence of frontal lesions correlates with early seizures, and the presence of hemispheric symptoms increases the risk for late seizures. Both early and late seizures are uncommon in patients with posterior fossa and sellar lesions. Seizures are common in patients with CNS metastases from melanoma and low-grade primary brain tumors. Very rarely, cytotoxic drugs such as etoposide, busulfan, and chlorambucil cause seizures. Another cause of seizures related to drug therapy is reversible posterior leukoencephalopathy syndrome (RPLS). RPLS is associated with administration of cisplatin, 5-fluorouracil, bleomycin, vinblastine, vincristine, etoposide, paclitaxel, ifosfamide, cyclophosphamide, doxorubicin, cytarabine, methotrexate, oxaliplatin, cyclosporine, tacrolimus, and bevacizumab. RPLS is characterized by headache, altered consciousness, generalized seizures, visual disturbances, hypertension, and posterior cerebral white matter vasogenic edema on CT/MRI. Seizures may begin focally but are typically generalized.

TREATMENT: SEIZURES

Patients in whom seizures due to CNS metastases have been demonstrated should receive anticonvulsive treatment with phenytoin. Prophylactic anticonvulsant therapy is not recommended unless the patient is at high risk for late seizures (melanoma primary, hemorrhagic metastases, treatment with radiosurgery). Serum phenytoin levels should be monitored closely and the dosage adjusted according to serum levels. Phenytoin induces the hepatic metabolism of dexamethasone, reducing its half-life, while dexamethasone may decrease phenytoin levels. Most antiseizure medications induce CYP450, which alters the metabolism of antitumor agents, including irinotecan, taxanes, and etoposide as well as molecular targeted agents, including imatinib, gefitinib, erlotinib, and tipifarnib. Levetiracetam and topiramate are anticonvulsant agents not metabolized by the hepatic cytochrome P450 system and do not alter the metabolism of antitumor agents.

Pulmonary and Intracerebral Leukostasis

Hyperleukocytosis and the leukostasis syndrome associated with it is a potentially fatal complication of acute leukemia (particularly myeloid leukemia) that can occur when the peripheral blast cell count is >100,000/mL. The frequency of hyperleukocytosis is 5–13% in acute myelid leukemia (AML) and 10–30% in acute lymphoid leukemia; however, leukostasis is rare in lymphoid leukemia. At such high blast cell counts, blood viscosity is increased, blood flow is slowed by aggregates of tumor cells, and the primitive myeloid leukemic cells are capable of invading through the endothelium and causing hemorrhage. Brain and lung are most commonly affected. Patients with brain leukostasis may experience stupor, headache, dizziness, tinnitus, visual disturbances, ataxia, confusion, coma, or sudden death. Administration of 600 cGy of whole-brain irradiation can protect against this complication and can be followed by rapid institution of antileukemic therapy. Hydroxyurea, 3-5 grams, can rapidly reduce a high blast cell count while the accurate diagnostic workup is in progress. Pulmonary leukostasis may present as respiratory distress, hypoxemia, and progress to respiratory failure. Chest radiographs may be normal but usually show interstitial or alveolar infiltrates. Arterial blood gas results should be interpreted cautiously. Rapid consumption of plasma oxygen by the markedly increased number of white blood cells can cause spuriously low arterial oxygen tension. Pulse oximetry is the most accurate way of assessing oxygenation in patients with hyperleukocytosis. Leukapheresis may be helpful in decreasing circulating blast counts. Treatment of the leukemia can result in pulmonary hemorrhage from lysis of blasts in the lung, called *leukemic cell lysis pneumopathy*. Intravascular volume depletion and unnecessary blood transfusions may increase blood viscosity and worsen the leukostasis syndrome. Leukostasis is very rarely a feature of the high white cell counts associated with chronic lymphoid or chronic myeloid leukemia.

When acute promyelocytic leukemia is treated with differentiating agents like tretinoin and arsenic trioxide, cerebral or pulmonary leukostasis may occur as tumor cells differentiate into mature neutrophils. This complication can be largely avoided by using cytotoxic chemotherapy together with the differentiating agents.

Hemoptysis

Hemoptysis may be caused by nonmalignant conditions, but lung cancer accounts for a large proportion of cases. Up to 20% of patients with lung cancer have hemoptysis some time in their course. Endobronchial metastases from carcinoid tumors, breast cancer, colon cancer, kidney cancer, and melanoma may also cause hemoptysis. The

volume of bleeding is often difficult to gauge. Massive hemoptysis is defined as >200–600 mL of blood produced in 24 h. However, any hemoptysis should be considered massive if it threatens life. When respiratory difficulty occurs, hemoptysis should be treated emergently. The first priorities are to maintain the airway, optimize oxygenation, and stabilize the hemodynamic status. *Often patients can tell where the bleeding is occurring.* They should be placed bleeding side down and given supplemental oxygen. If large-volume bleeding continues or the airway is compromised, the patient should be intubated and undergo emergency bronchoscopy. If the site of bleeding is detected, either the patient undergoes a definitive surgical procedure or the lesion is treated with a neodymium:yttrium-aluminum-garnet (Nd:YAG) laser. The surgical option is preferred. Bronchial artery embolization may control brisk bleeding in 75–90% of patients, permitting the definitive surgical procedure to be done more safely. Embolization without definitive surgery is associated with rebleeding in 20–50% of patients. Recurrent hemoptysis usually responds to a second embolization procedure. A postembolization syndrome characterized by pleuritic pain, fever, dysphagia, and leukocytosis may occur; it lasts 5–7 days and resolves with symptomatic treatment. Bronchial or esophageal wall necrosis, myocardial infarction, and spinal cord infarction are rare complications.

Pulmonary hemorrhage with or without hemoptysis in hematologicmalignancies is often associated with fungal infections, particularly *Aspergillus* sp. After granulocytopenia resolves, the lung infiltrates in aspergillosis may cavitate and cause massive hemoptysis. Thrombocytopenia and coagulation defects should be corrected, if possible. Surgical evaluation is recommended in patients with aspergillosis-related cavitary lesions.

Bevacizumab, an antibody to vascular endothelial growth factor (VEGF) that inhibits angiogenesis, has been associated with life-threatening hemoptysis in patients with non-small cell lung cancer, particularly squamous cell histology. Non-small cell lung cancer patients with cavitary lesions have higher risk for pulmonary hemorrhage.

Airway Obstruction

Airway obstruction refers to a blockage at the level of the mainstem bronchi or above. It may result either from intraluminal tumor growth or from extrinsic compression of the airway. The most common cause of malignant upper airway obstruction is invasion from an adjacent primary tumor, most commonly lung cancer, followed by esophageal, thyroid, and mediastinal malignancies. Extrathoracic primary tumors such as renal, colon, or breast cancer can cause airway obstruction through endobronchial and/or mediastinal lymph node metastases. Patients may present with dyspnea, hemoptysis, stridor, wheezing, intractable cough, postobstructive pneumonia, or hoarseness. Chest radiographs usually demonstrate obstructing lesions. CT scans reveal the extent of tumor. Cool, humidified oxygen, glucocorticoids, and ventilation with a mixture of helium and oxygen (Heliox) may provide temporary relief. If the obstruction is proximal to the larynx, a tracheostomy may be lifesaving. For more distal obstructions, particularly intrinsic lesions incompletely obstructing the airway, bronchoscopy with laser treatment, photodynamic therapy, or stenting can produce immediate relief in most patients **(Fig. 276-3)**. However, radiation therapy (either external-beam irradiation or brachytherapy) given together with glucocorticoids may also open the airway. Symptomatic extrinsic compression may be palliated by stenting. Patients with primary airway tumors such as squamous cell carcinoma, carcinoid tumor, adenocystic carcinoma, or non-small cell lung cancer should have surgery.

Figure 276-3

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

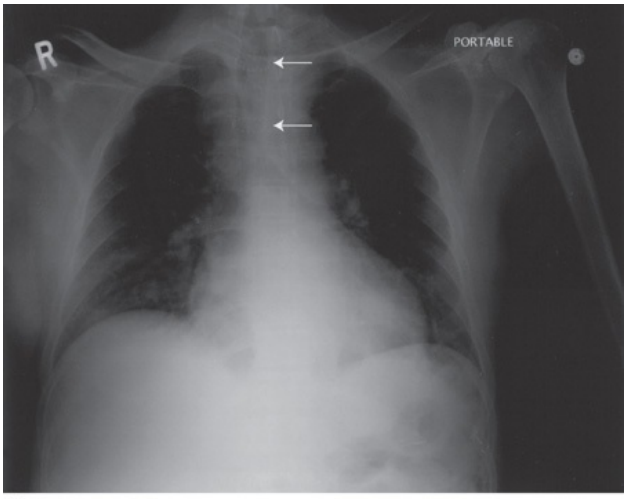

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Airway obstruction. *A.* CT scan of a 62-year-old man with tracheal obstruction caused by renal carcinoma showing paratracheal mass (***A***) with tracheal invasion/obstruction (*arrow*). ***B.*** Chest x-ray of same patient after stent (*arrows*) placement.

METABOLIC EMERGENCIES

Hypercalcemia

Hypercalcemia is the most common paraneoplastic syndrome. **Its pathogenesis and management are discussed fully in Chaps. 100 and 353.**

Syndrome of Inappropriate Secretion of Antidiuretic Hormone (SIADH)

Hyponatremia is a common electrolyte abnormality in cancer patients, and SIADH is the most common cause among patients with cancer. **SIADH is discussed fully in Chaps. 100 and 339**.

Lactic Acidosis

Lactic acidosis is a rare and potentially fatal metabolic complication of cancer. The body produces about 1500 mmols of lactic acid per day, most of which is metabolized by the liver. Normally, this lactate is generated by the skin (25%), muscle (25%), red cells (20%), brain (20%), and gut (10%). Lactic acidosis may occur as a consequence of increased production or decreased hepatic metabolism. Normal venous levels of lactate are 0.5 –2.2 mmol/L (4.5–19.8 mg/dL). Lactic acidosis associated with sepsis and circulatory failure is a common preterminal event in many malignancies. Lactic acidosis in the absence of hypoxemia may occur in patients with leukemia, lymphoma, or solid tumors. Extensive involvement of the liver by tumor is often present. In most cases, decreased metabolism and increased production by the tumor both contribute to lactate accumulation. Tumor cell overexpression of certain glycolytic enzymes and mitochondrial dysfunction can contribute to its increased lactate production. HIV-infected patients have an increased risk of aggressive lymphoma; lactic acidosis that occurs in such patients may be related either to the rapid growth of the tumor or from toxicity of nucleoside reverse transcriptase inhibitors. Symptoms of lactic acidosis include tachypnea, tachycardia, change of mental status, and hepatomegaly. The serum level of lactic acid may reach 10–20 mmol/L (90–180 mg/dL). Treatment is aimed at the underlying disease. *The danger from lactic acidosis is from the acidosis, not the lactate* . Sodium bicarbonate should be added if acidosis is very severe or if hydrogen ion production is very rapid and uncontrolled. The prognosis is poor.

Hypoglycemia

Persistent hypoglycemia is occasionally associated with tumors other than pancreatic islet cell tumors. Usually these tumors are large; tumors of mesenchymal origin, hepatomas, or adrenocortical tumors may cause hypoglycemia. Mesenchymal tumors are usually located in the retroperitoneum or thorax. Obtundation, confusion, and behavioral aberrations occur in the postabsorptive period and may precede the diagnosis of the tumor. These tumors often secrete incompletely processed insulin-like growth factor II (IGF-II), a hormone capable of activating insulin receptors and causing hypoglycemia. Tumors secreting incompletely processed big IGF-II are characterized by an increased IGF-II to IGF-I ratio, suppressed insulin and C-peptide level, and inappropriately low growth hormone and β-hydroxybutyrate concentrations. Rarely, hypoglycemia is due to insulin secretion by a non-islet cell carcinoma. The development of hepatic dysfunction from liver metastases and increased glucose consumption by the tumor can contribute to hypoglycemia. If the tumor cannot be resected, hypoglycemia symptoms may be relieved by the administration of glucose, glucocorticoids, or glucagon.

Hypoglycemia can be artifactual; hyperleukocytosis from leukemia, myeloproliferative diseases, leukemoid reactions, or colony-stimulating factor treatment can increase glucose consumption in the test tube after blood is drawn, leading to pseudohypoglycemia.

Adrenal Insufficiency

In patients with cancer, adrenal insufficiency may go unrecognized because the symptoms, such as nausea, vomiting, anorexia, and orthostatic hypotension, are nonspecific and may be mistakenly attributed to progressive cancer or to therapy. Primary adrenal insufficiency may develop owing to replacement of both glands by

metastases (lung, breast, colon, or kidney cancer; lymphoma), to removal of both glands, or to hemorrhagic necrosis in association with sepsis or anticoagulation. Impaired adrenal steroid synthesis occurs in patients being treated for cancer with mitotane, ketoconazole, or aminoglutethimide or undergoing rapid reduction in glucocorticoid therapy. Rarely, metastatic replacement causes primary adrenal insufficiency as the first manifestation of an occult malignancy. Metastasis to the pituitary or hypothalamus is found at autopsy in up to 5% of patients with cancer, but associated secondary adrenal insufficiency is rare. Megestrol acetate, used to manage cancer and HIV-related cachexia, may suppress plasma levels of cortisol and adrenocorticotropic hormone (ACTH). Patients taking megestrol may develop adrenal insufficiency, and even those whose adrenal dysfunction is not symptomatic may have inadequate adrenal reserve if they become seriously ill. Paradoxically, some patients may develop Cushing's syndrome and/or hyperglycemia because of the glucocorticoid-like activity of megestrol acetate. Cranial irradiation for childhood brain tumors may affect the hypothalamus-pituitary-adrenal axis, resulting in secondary adrenal insufficiency.

Acute adrenal insufficiency is potentially lethal. Treatment of suspected adrenal crisis is initiated after the sampling of serum cortisol and ACTH levels (Chap. 342).

TREATMENT-RELATED EMERGENCIES

Tumor Lysis Syndrome

Tumor lysis syndrome (TLS) is characterized by hyperuricemia, hyperkalemia, hyperphosphatemia, and hypocalcemia and is caused by the destruction of a large number of rapidly proliferating neoplastic cells. Acidosis may also develop. Acute renal failure occurs frequently.

TLS is most often associated with the treatment of Burkitt's lymphoma, acute lymphoblastic leukemia, and other rapidly proliferating lymphomas, but it also may be seen with chronic leukemias and, rarely, with solid tumors. This syndrome has been seen in patients with chronic lymphocytic leukemia after treatment with nucleosides like fludarabine. TLS has been observed with administration of glucocorticoids, hormonal agents such as letrozole and tamoxifen, and monoclonal antibodies such as rituximab and gemtuzumab. TLS usually occurs during or shortly (1–5 days) after chemotherapy. Rarely, spontaneous necrosis of malignancies causes TLS.

Hyperuricemia may be present at the time of chemotherapy. Effective treatment kills malignant cells and leads to increased serum uric acid levels from the turnover of nucleic acids. Owing to the acidic local environment, uric acid can precipitate in the tubules, medulla, and collecting ducts of the kidney, leading to renal failure. Lactic acidosis and dehydration may contribute to the precipitation of uric acid in the renal tubules. The finding of uric acid crystals in the urine is strong evidence for uric acid nephropathy. The ratio of urinary uric acid to urinary creatinine is >1 in patients with acute hyperuricemic nephropathy and <1 in patients with renal failure due to other causes.

Hyperphosphatemia, which can be caused by the release of intracellular phosphate pools by tumor lysis, produces a reciprocal depression in serum calcium, which causes severe neuromuscular irritability and tetany. Deposition of calcium phosphate in the kidney and hyperphosphatemia may cause renal failure. Potassium is the principal intracellular cation, and massive destruction of malignant cells may lead to hyperkalemia. Hyperkalemia in patients with renal failure may rapidly become life-threatening by causing ventricular arrhythmias and sudden death.

The likelihood that TLS will occur in patients with Burkitt's lymphoma is related to the tumor burden and renal function. Hyperuricemia and high serum levels of lactate dehydrogenase (LDH >1500 U/L), both of which correlate with total tumor burden, also correlate with the risk of TLS. In patients at risk for TLS, pretreatment evaluations should include a complete blood count, serum chemistry evaluation, and urine analysis. High leukocyte and platelet counts may artificially elevate potassium levels ("pseudohyperkalemia") due to lysis of these cells after the blood is drawn. In these cases, plasma potassium instead of serum potassium should be followed. In pseudohyperkalemia, no electrocardiographic abnormalities are present. In patients with abnormal baseline renal function, the kidneys and retroperitoneal area should be evaluated by sonography and/or CT to rule out obstructive uropathy. Urine output should be watched closely.

TREATMENT: TUMOR LYSIS SYNDROME

Recognition of risk and prevention are the most important steps in the management of this syndrome **(Fig. 276-4)**. The standard preventive approach consists of allopurinol, urinary alkalinization, and aggressive hydration. Intravenous allopurinol may be given in patients who cannot tolerate oral therapy. In some cases, uric acid levels cannot be lowered sufficiently with the standard preventive approach. Rasburicase (recombinant urate oxidase) can be effective in these instances. Urate oxidase is missing from primates and catalyzes the conversion of poorly soluble uric acid to readily soluble allantoin. Rasburicase acts rapidly, decreasing uric acid levels within hours; however, it may cause hypersensitivity reactions such as bronchospasm, hypoxemia, and hypotension. Rasburicase should also be administered to high-risk patients for TLS prophylaxis. Rasburicase is contraindicated in patients with glucose-6-phosphate dehydrogenase deficiency who are unable to break down hydrogen peroxide, an end product of the urate oxidase reaction. Despite aggressive prophylaxis, TLS and/or oliguric or anuric renal failure may occur. Care should be taken to prevent worsening of symptomatic hypocalcemia by induction of alkalosis during bicarbonate infusion. Administration of sodium bicarbonate may also lead to urinary precipitation of calcium phosphate, which is less soluble at alkaline pH. Dialysis is often necessary and should be considered early in the course. Hemodialysis is preferred. Hemofiltration offers a gradual, continuous method of removing cellular by-products and fluid. The prognosis is excellent, and renal function recovers after the uric acid level is lowered to ≤10 mg/dL.

Figure 276-4

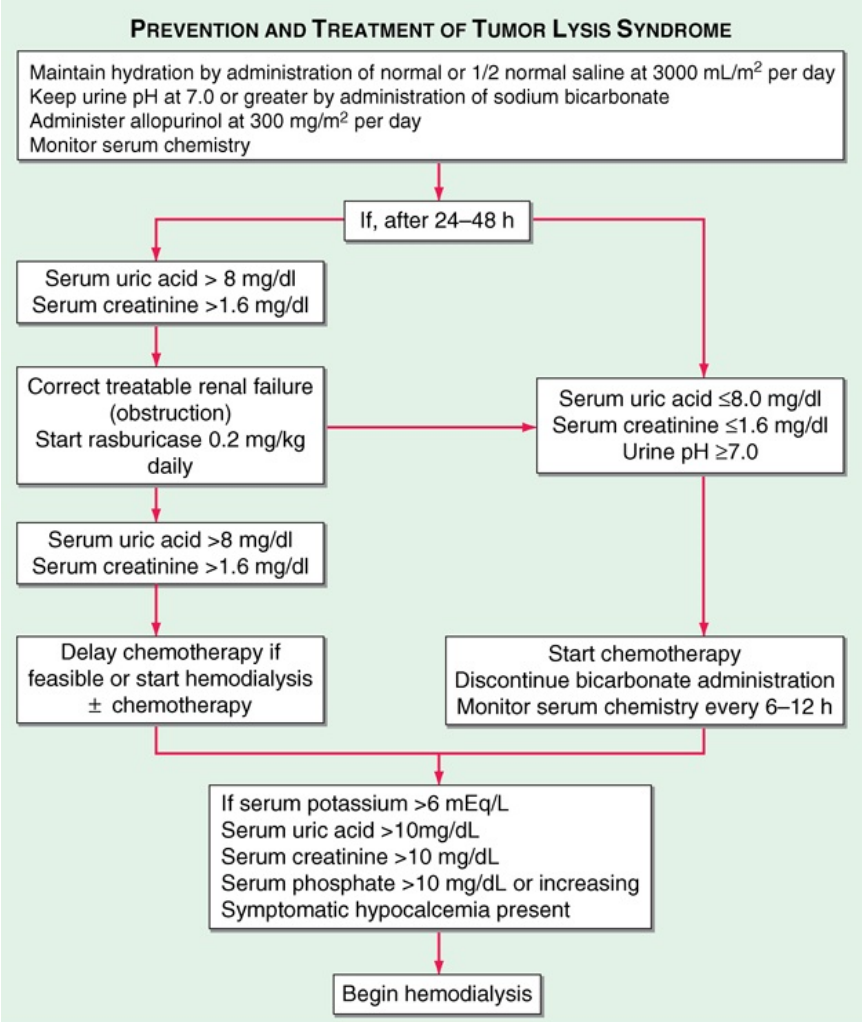

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of patients at high risk for the tumor lysis syndrome.

Human Antibody Infusion Reactions

The initial infusion of human or humanized antibodies (e.g., rituximab, gemtuzumab, trastuzumab) is associated with fever, chills, nausea, asthenia, and headache in up to half of treated patients. Bronchospasm and hypotension occur in 1% of patients. Severe manifestations including pulmonary infiltrates, acute respiratory distress syndrome, and cardiogenic shock occur rarely. Laboratory manifestations include elevated hepatic aminotransferase levels, thrombocytopenia, and prolongation of prothrombin time. The pathogenesis is thought to be activation of immune effector processes (cells and complement) and release of inflammatory cytokines, such as tumor necrosis factor α and interleukin 6 (cytokine release syndrome). Severe reactions from rituximab have occurred with high numbers (more than 50 x 10^9 lymphocytes) of circulating cells bearing the target antigen (CD 20) and have been associated with a rapid fall in circulating tumor cells, mild electrolyte evidence of TLS, and very rarely, with death. In addition, increased liver enzymes, d-dimer, LDH, and prolongation of the prothrombin time may occur. Diphenhydramine, hydrocortisone, and acetaminophen can often prevent or suppress the infusion-related symptoms. If they occur, the infusion is stopped and restarted at half the initial infusion rate after the symptoms have abated. Severe "cytokine release syndrome" may require intensive support for acute respiratory distress syndrome (ARDS) and resistant hypotension.

Hemolytic-Uremic Syndrome

Hemolytic-uremic syndrome (HUS) and, less commonly, thrombotic thrombocytopenic purpura (TTP) (Chap. 286) may rarely occur after treatment with antineoplastic drugs including mitomycin, cisplatin, bleomycin, and gemcitabine. It occurs most often in patients with gastric, lung, colorectal, pancreatic, and breast carcinoma. In one series, 35% of patients were without evident cancer at the time this syndrome appeared. Secondary HUS/TTP has also been reported as a rare but sometimes fatal complication of bone marrow transplantation.

HUS usually has its onset 4–8 weeks after the last dose of chemotherapy, but it is not rare to detect it several months later. HUS is characterized by microangiopathic hemolytic anemia, thrombocytopenia, and renal failure. Dyspnea, weakness, fatigue, oliguria, and purpura are also common initial symptoms and findings. Systemic hypertension and pulmonary edema frequently occur. Severe hypertension, pulmonary edema, and rapid worsening of hemolysis and renal function may occur after a blood or blood product transfusion. Cardiac findings include atrial arrhythmias, pericardial friction rub, and pericardial effusion. Raynaud's phenomenon is part of the syndrome in patients treated with bleomycin.

Laboratory findings include severe to moderate anemia associated with red blood cell fragmentation and numerous schistocytes on peripheral smear. Reticulocytosis, decreased plasma haptoglobin, and an LDH level document hemolysis. The serum bilirubin level is usually normal or slightly elevated. The Coombs' test is negative. The white cell count is usually normal, and thrombocytopenia (<100,000/μL) is almost always present. Most patients have a normal coagulation profile, although some have mild elevations in thrombin time and in levels of fibrin degradation products. The serum creatinine level is elevated at presentation and shows a pattern of subacute worsening within weeks of the initial azotemia. The urinalysis reveals hematuria, proteinuria, and granular or hyaline casts; and circulating immune complexes may be present.

The basic pathologic lesion appears to be deposition of fibrin in the walls of capillaries and arterioles, and these deposits are similar to those seen in HUS due to other causes. These microvascular abnormalities involve mainly the kidneys and rarely occur in other organs. The pathogenesis of chemotherapy-related HUS is unknown. Other forms of HUS/TTP are related to a decrease in processing of von Willebrand factor by a protease called ADAMTS13.

The case fatality rate is high; most patients die within a few months. There is no consensus on the optimal treatment for chemotherapy-induced HUS. Treatment modalities for HUS/TTP including immunocomplex removal (plasmapheresis, immunoadsorption, or exchange transfusion), antiplatelet/anticoagulant therapies, immunosuppressive therapies, and plasma exchange employed varying degrees of success. Rituximab is successfully used in patients with chemotherapy-induced HUS as well as in ADAMTS13-deficient TTP.

Neutropenia and Infection

These remain the most common serious complications of cancer therapy. **They are covered in detail in Chap. 86.**

Pulmonary Infiltrates

Patients with cancer may present with dyspnea associated with diffuse interstitial infiltrates on chest radiographs. Such infiltrates may be due to progression of the underlying malignancy, treatment-related toxicities, infection, and/or unrelated diseases. The cause may be multifactorial; however, most commonly they occur as a consequence of treatment. Infiltration of the lung by malignancy has been described in patients with leukemia, lymphoma, and breast and other solid cancers. Pulmonary lymphatics may be involved diffusely by neoplasm (pulmonary lymphangitic carcinomatosis), resulting in a diffuse increase in interstitial markings on chest radiographs. The patient is often mildly dyspneic at the onset, but pulmonary failure develops over a period of weeks. In some patients, dyspnea precedes changes on the chest radiographs and is accompanied by a nonproductive cough. This syndrome is characteristic of solid tumors. In patients with leukemia, diffuse microscopic neoplastic peribronchial and peribronchiolar infiltration is frequent but may be asymptomatic. However, some patients present with diffuse interstitial infiltrates, an alveolar capillary block syndrome, and respiratory distress. In these situations, glucocorticoids can provide symptomatic relief, but specific chemotherapy should always be started promptly.

Several cytotoxic agents, such as bleomycin, methotrexate, busulfan, nitrosoureas, gemcitabine, mitomycin, vinorelbine, docetaxel, and ifosfamide may cause pulmonary damage. The most frequent presentations are interstitial pneumonitis, alveolitis, and pulmonary fibrosis. Some cytotoxic agents, including methotrexate and procarbazine, may cause an acute hypersensitivity reaction. Cytosine arabinoside has been associated with noncardiogenic pulmonary edema. Administration of multiple cytotoxic drugs, as well as radiotherapy and preexisting lung disease, may potentiate the pulmonary toxicity. Supplemental oxygen may potentiate the effects of drugs and radiation injury. Patients should always be managed with the lowest FIo_2 that is sufficient to maintain hemoglobin saturation.

The onset of symptoms may be insidious, with symptoms including dyspnea, nonproductive cough, and tachycardia. Patients may have bibasilar crepitant rales, end-inspiratory crackles, fever, and cyanosis. The chest radiograph generally shows an interstitial and sometimes an intraalveolar pattern that is strongest at the lung bases and may be symmetric. A small effusion may occur. Hypoxemia with decreased carbon monoxide diffusing capacity is always present. Glucocorticoids may be helpful in patients in whom pulmonary toxicity is related to radiation therapy or to chemotherapy. Treatment is otherwise supportive.

Molecular targeted agents, imatinib, erlotinib, and gefitinib are potent inhibitors of tyrosine kinases. These drugs may cause interstitial lung disease. In the case of gefitinib, preexisting fibrosis, poor performance status, and prior thoracic irradiation are independent risk factors; this complication has a high fatality rate. In Japan, incidence of interstitial lung disease associated with gefitinib was about 4.5% compared to 0.5% in the United States. Temsirolimus, a derivative of rapamycin, is an agent that blocks the effects of mTOR, an enzyme that has an important role in regulating the synthesis of proteins that control cell division. It may cause ground-glass opacities in the lung with or without diffuse interstitial disease and lung parenchymal consolidation.

Radiation pneumonitis and/or fibrosis is a relatively frequent side effect of thoracic radiation therapy. It may be acute or chronic. Radiation-induced lung toxicity is a function of the irradiated lung volume, dose per fraction, and radiation dose. The larger the irradiated lung field, the higher the risk for radiation pneumonitis. The use of concurrent chemoradiation increases pulmonary toxicity. Radiation pneumonitis usually develops from 2 to 6 months after completion of radiotherapy. The clinical syndrome, which varies in severity, consists of dyspnea, cough with scanty sputum, low-grade fever, and an initial hazy infiltrate on chest radiographs. The infiltrate and tissue damage usually are confined to the radiation field. The patients subsequently may develop a patchy alveolar infiltrate and air bronchograms, which may progress to acute respiratory failure that is sometimes fatal. A lung biopsy may be necessary to make the diagnosis. Asymptomatic infiltrates found incidentally after radiation therapy need not be treated. However, prednisone should be administered to patients with fever or other symptoms. The dosage should be tapered slowly after the resolution of radiation pneumonitis, as abrupt withdrawal of glucocorticoids may cause an exacerbation of pneumonia. Delayed radiation fibrosis may occur years after radiation therapy and is signaled by dyspnea on exertion. Often it is mild, but it can progress to chronic respiratory failure. Therapy is supportive.

Classical radiation pneumonitis that leads to pulmonary fibrosis is due to radiation-induced production of local cytokines such as platelet-derived growth factor ß, tumor necrosis factor, interleukins, and transforming growth factor ß in the radiation field. An immunologically mediated sporadic radiation pneumonitis occurs in about 10% of patients; bilateral alveolitis mediated by T cells results in infiltrates outside the radiation field. This form of radiation pneumonitis usually resolves without sequelae.

Pneumonia is a common problem in patients undergoing treatment for cancer. Bacterial pneumonia typically causes a localized infiltrate on chest radiographs. Therapy is tailored to the causative organism. When diffuse interstitial infiltrates appear in a febrile patient, the differential diagnosis is extensive and includes pneumonia due to infection with *Pneumocystis carinii*; viral infections including cytomegalovirus, adenovirus, herpes simplex virus, herpes zoster, respiratory syncytial virus, or intracellular pathogens such as *Mycoplasma* and *Legionella*; effects of drugs or radiation; tumor progression; nonspecific pneumonitis; and fungal disease. Detection of opportunistic pathogens in pulmonary infections is still a challenge. Diagnostic tools include chest radiographs, CT scans, bronchoscopy with bronchoalveolar lavage, brush cytology, transbronchial biopsy, fine-needle aspiration, and open lung biopsy. In addition to the culture, evaluation of bronchoalveolar lavage fluid for *P. carinii* by polymerase chain reaction (PCR) and serum galactomannan test improve the diagnostic yield. Patients with cancer who are neutropenic and have fever and local infiltrates on chest radiograph should be treated initially with broad-spectrum antibiotics. A new or persistent focal infiltrate not responding to broad-spectrum antibiotics argues for initiation of empiric antifungal therapy. When diffuse bilateral infiltrates develop in patients with febrile neutropenia, broad-spectrum antibiotics plus trimethoprim-sulfamethoxazole, with or without erythromycin, should be initiated. Addition of an antiviral agent is necessary in some settings, such as patients undergoing allogeneic hematopoietic stem cell transplantation. If the patient does not improve in 4 days, open lung biopsy is the procedure of choice. Bronchoscopy with bronchoalveolar lavage may be used in patients who are poor candidates for surgery.

In patients with pulmonary infiltrates who are afebrile, heart failure and multiple pulmonary emboli are in the differential diagnosis.

Neutropenic Enterocolitis

Neutropenic enterocolitis (typhlitis) is the inflammation and necrosis of the cecum and surrounding tissues that may complicate the treatment of acute leukemia. Nevertheless, it may involve any segment of the gastrointestinal tract including small intestine, appendix, and colon. This complication has also been seen in patients with other forms of cancer treated with taxanes and in patients receiving high-dose chemotherapy **(Fig. 276-5)**. The patient develops right lower quadrant abdominal pain, often with rebound tenderness and a tense, distended abdomen, in a setting of fever and neutropenia. Watery diarrhea (often containing sloughed mucosa) and bacteremia are common, and bleeding may occur. Plain abdominal films are generally of little value in the diagnosis; CT scan may show marked bowel wall thickening, particularly in the cecum, with bowel wall edema, mesenteric stranding, and ascites. Patients with bowel wall thickness >10 mm on ultrasonogram have higher mortality rates. However, bowel wall thickening is significantly more prominent in patients with *Clostridium difficile* colitis. Pneumatosis intestinalis is a more specific finding, seen only in those with neutropenic enterocolitis and ischemia. The combined involvement of the small and large bowel suggests a diagnosis of neutropenic enterocolitis. Rapid institution of broad-spectrum antibiotics and nasogastric suction may reverse the process. Surgical intervention is reserved for severe cases of neutropenic enterocolitis with evidence of perforation, peritonitis, gangrenous bowel, or gastrointestinal hemorrhage despite correction of any coagulopathy.

Figure 276-5

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Abdominal CT scans of a 72-year-old woman with neutropenic enterocolitis secondary to chemotherapy.*A.* Air in inferior mesenteric vein (*arrow*) and bowel wall with pneumatosis intestinalis. ***B.*** CT scans of upper abdomen demonstrating air in portal vein (*arrows*).

C. difficile colitis is increasing in incidence. Newer strains of *C. difficile* produce about 20 times more of toxins A and B compared to previously studied strains. *C. difficile* risk is also increased with chemotherapy. Antibiotic coverage for *C. difficile* should be added if pseudomembranous colitis cannot be excluded.

Hemorrhagic Cystitis

Hemorrhagic cystitis can develop in patients receiving cyclophosphamide or ifosfamide. Both drugs are metabolized to acrolein, which is a strong chemical irritant that is excreted in the urine. Prolonged contact or high concentrations may lead to bladder irritation and hemorrhage. Symptoms include gross hematuria, frequency, dysuria, burning, urgency, incontinence, and nocturia. The best management is prevention. Maintaining a high rate of urine flow minimizes exposure. In addition, 2-mercaptoethanesulfonate (mesna) detoxifies the metabolites and can be coadministered with the instigating drugs. Mesna usually is given three times on the day of ifosfamide administration in doses that are each 20% of the total ifosfamide dose. If hemorrhagic cystitis develops, the maintenance of a high urine flow may be sufficient supportive care. If conservative management is not effective, irrigation of the bladder with a 0.37–0.74% formalin solution for 10 min stops the bleeding in most cases. *N*-acetylcysteine may also be an effective irrigant. Prostaglandin (carboprost) can inhibit the process. In extreme cases, ligation of the hypogastric arteries, urinary diversion, or cystectomy may be necessary.

Hemorrhagic cystitis also occurs in patients who undergo bone marrow transplantation (BMT). In the BMT setting, early-onset hemorrhagic cystitis is related to drugs in the treatment regimen (e.g., cyclophosphamide), and late-onset hemorrhagic cystitis is usually due to the polyoma virus BKV or adenovirus type 11. BKV load in urine alone or in combination with acute graft-versus-host disease correlate with development of hemorrhagic cystitis. Viral causes are usually detected by PCR-based diagnostic tests. Treatment of viral hemorrhagic cystitis is largely supportive, with reduction in doses of immunosuppressive agents, if possible. No antiviral therapy is approved, though cidofovir is reported to be effective in small series.

Hypersensitivity Reactions to Antineoplastic Drugs

Many antineoplastic drugs may cause hypersensitivity reaction (HSR). These reactions are unpredictable and potentially life-threatening. Most reactions occur during or within hours of parenteral drug administration. Taxanes, platinum compounds, asparaginase, etoposide, and biologic agents, including rituximab, bevacizumab, trastuzumab, gemtuzumab, cetuximab, and alemtuzumab, are more commonly associated with acute HSR than are other agents. Acute hypersensitivity reactions to some drugs, such as taxanes, occur during the first or second dose administered. HSR from platinum compounds occurs after prolonged exposure. Skin testing may identify patients with high risk for HSR after carboplatin exposure. Premedication with histamine H_1 and H_2 receptor antagonists and glucocorticoids reduce the incidence of hypersensitivity reaction to taxanes, particularly paclitaxel. Despite premedication, HSR may still occur. In these cases, re-treatment may be attempted with care, but use of alternative agents may be required.

FURTHER READINGS

Albanell J, Baselga J: Systemic therapy emergencies. Semin Oncol 27:347, 2000[PMID: 10864222]

Coiffier B, Riouffol C: Management of tumor lysis syndrome in adults. Expert Rev Anticancer Ther 7:233, 2007[PMID: 17288532]

Davis MP et al: Modern management of cancer-related intestinal obstruction. Curr Oncol Rep 2:343, 2000[PMID: 11122863]

Gleissner B et al: Neoplastic meningitis. Lancet Neurol 5:443, 2006[PMID: 16632315]

Loblaw DA et al: Systematic review of the diagnosis and management of malignant extradural spinal cord compression: The Cancer Care Ontario Practice Guidelines Initiative's neuro-oncology disease site group. J Clin Oncol 23:2028, 2005[PMID: 15774794]

Maggi E et al: Acute infusion reactions induced by monoclonal antibody therapy. Expert Rev Clin Immunol 7:55, 2011[PMID: 21162650]

Ripamonti C et al: Respiratory problems in advanced cancer. Support Care Cancer 10:204, 2002[PMID: 11904785]

Samphao S et al: Oncological emergencies: Clinical importance and principles of management. Eur J Cancer Care (Engl) 19:707, 2010[PMID: 20030695]

Wan JF, Bezjak A: Superior vena cava syndrome. Emerg Med Clin North Am 27:243, 2009[PMID: 19447309]

Wright FC et al: Predictors of survival in patients with non-curative stage IV cancer and malignant bowel obstruction. J Surg Oncol 101:425, 2010[PMID: 20112263]

Zanotti KM et al: Prevention and management of antineoplastic-induced hypersensitivity reactions; Drug Saf 24:767, 2001[PMID: 11676304]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CELLULAR AND MOLECULAR BIOLOGY OF THE KIDNEY: INTRODUCTION

The kidney is one of the most highly differentiated organs in the body. At the conclusion of embryologic development, nearly 30 different cell types form a multitude of filtering capillaries and segmented nephrons enveloped by a dynamic interstitium. This cellular diversity modulates a variety of complex physiologic processes. Endocrine functions, the regulation of blood pressure and intraglomerular hemodynamics, solute and water transport, acid-base balance, and removal of drug metabolites are all accomplished by intricate mechanisms of renal response. This breadth of physiology hinges on the clever ingenuity of nephron architecture that evolved as complex organisms came out of water to live on land.

EMBRYOLOGIC DEVELOPMENT

Kidneys develop from intermediate mesoderm under the timed or sequential control of a growing number of genes, described in Fig. 277-1. The transcription of these genes is guided by morphogenic cues that invite two ureteric buds to each penetrate bilateral metanephric blastema, where they induce primary mesenchymal cells to form early nephrons. This induction involves a number of complex signaling pathways mediated by Pax2, Six2, WT-1, Wnt9b, c-Met, fibroblast growth factor, transforming growth factor β, glial cell-derived neurotrophic factor, hepatocyte growth factor, and epidermal growth factor. The two ureteric buds emerge from posterior nephric ducts and mature into separate collecting systems that eventually form a renal pelvis and ureter. Induced mesenchyme undergoes mesenchymal epithelial transitions to form comma-shaped bodies at the proximal end of each ureteric bud leading to the formation of S-shaped nephrons that cleft and enjoin with penetrating endothelial cells derived from sprouting angioblasts. Under the influence of vascular endothelial growth factor A (VEGF-A), these penetrating cells form capillaries with surrounding mesangial cells that differentiate into a glomerular filter for plasma water and solute. The ureteric buds branch, and each branch produces a new set of nephrons. The number of branching events ultimately determines the total number of nephrons in each kidney. There are approximately 900,000 glomeruli in each kidney in normal birth weight adults and as few as 225,000 in low birth weight adults—producing the latter in numerous comorbid risks.

Figure 277-1

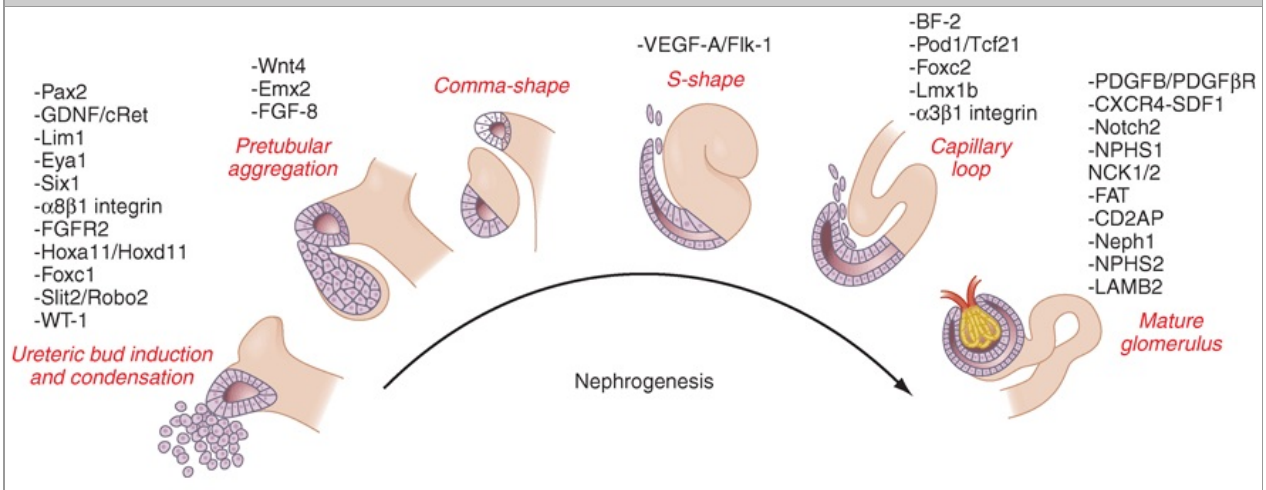

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Genes controlling renal nephrogenesis. A growing number of genes have been identified at various stages of glomerulotubular development in the mammalian kidney. The genes listed have been tested in various genetically modified mice, and their location corresponds to the classical stages of kidney development postulated by Saxen in 1987. GDNF, giant cell line-derived neutrophilic factor; fgFR2, fibroblast growth factor receptor 2; WT-1, Wilms' tumor gene 1; fgF-8, fibroblast growth factor 8; VEGF–A/Flk-1, vascular endothelial growth factor–A/fetal liver kinase-1; PDGFβ, platelet-derived growth factor β; PDGFβR, PDGFβ receptor; SDF-1, stromal-derived factor 1; NPHS1, nephrin; NCK1/2, NCK-adaptor protein; CD2AP, CD2-associated protein; NPHS2, podocin; LAMB2, laminin beta-2.

Glomeruli evolve as complex capillary filters with fenestrated endothelia under the guiding influence of VEGF-A and angiopoietin-1 secreted by adjacently developing podocytes. Epithelial podocytes facing the urinary space envelop the exterior basement membrane supporting these emerging endothelial capillaries. Podocytes are partially polarized and periodically fall off into the urinary space by epithelial-mesenchymal transition, and to a lesser extent apoptosis, only to be replenished by migrating parietal epithelia from Bowman's capsule. Failing replenishment results in heavy proteinuria. Podocytes attach to the basement membrane by special foot

processes and share a slit-pore membrane with their neighbor. The slit-pore membrane forms a filter for plasma water and solute by the synthetic interaction of nephrin, annexin-4, CD2AP, FAT, ZO-1, P-cadherin, podocin, TRPC6, PLCE1, and neph 1–3 proteins. Mutations in many of these proteins also result in heavy proteinuria. The glomerular capillaries are embedded in a mesangial matrix shrouded by parietal and proximal tubular epithelia forming Bowman's capsule. Mesangial cells have an embryonic lineage consistent with arteriolar or juxtaglomerular cells and contain contractile actin-myosin fibers. These mesangial cells make contact with glomerular capillary loops, and their local matrix holds them in condensed arrangement.

Between nephrons lies the renal interstitium. This region forms a functional space surrounding glomeruli and their downstream tubules, which are home to resident and trafficking cells such as fibroblasts, dendritic cells, occasional lymphocytes, and lipid-laden macrophages. The cortical and medullary capillaries, which siphon off solute and water following tubular reclamation of glomerular filtrate, are also part of the interstitial fabric as well as a web of connective tissue that supports the kidney's emblematic architecture of folding tubules. The relational precision of these structures determines the unique physiology of the kidney.

Each nephron is partitioned during embryologic development into a proximal tubule, descending and ascending limbs of the loop of Henle, distal tubule, and the collecting duct. These classic tubular segments build from subsegments lined by highly unique epithelia serving regional physiology. All nephrons have the same structural components, but there are two types whose structure depend on their location within the kidney. The majority of nephrons are cortical, with glomeruli located in the mid-to-outer cortex. Fewer nephrons are juxtamedullary, with glomeruli at the boundary of the cortex and outer medulla. Cortical nephrons have short loops of Henle, whereas juxtamedullary nephrons have long loops of Henle. There are critical differences in blood supply as well. The peritubular capillaries surrounding cortical nephrons are shared among adjacent nephrons. By contrast, juxtamedullary nephrons depend on individual capillaries called *vasa recta*. Cortical nephrons perform most of the glomerular filtration because there are more of them and because their afferent arterioles are larger than their respective efferent arterioles. The juxtamedullary nephrons, with longer loops of Henle, create a hyperosmolar gradient for concentrating urine. How developmental instructions specify the differentiation of all these unique epithelia among various tubular segments is still unknown.

DETERMINANTS AND REGULATION OF GLOMERULAR FILTRATION

Renal blood flow normally drains approximately 20% of the cardiac output, or 1000 mL/min. Blood reaches each nephron through the afferent arteriole leading into a glomerular capillary where large amounts of fluid and solutes are filtered to form the tubular fluid. The distal ends of the glomerular capillaries coalesce to form an efferent arteriole leading to the first segment of a second capillary network (cortical peritubular capillaries or medullary vasa recta) surrounding the tubules **(Fig. 277-2*A*)**. Thus, nephrons have two capillary beds arranged in a series separated by the efferent arteriole that regulates the hydrostatic pressure in both capillary beds. The distal capillaries empty into small venous branches that coalesce into larger veins to eventually form the renal vein.

Figure 277-2

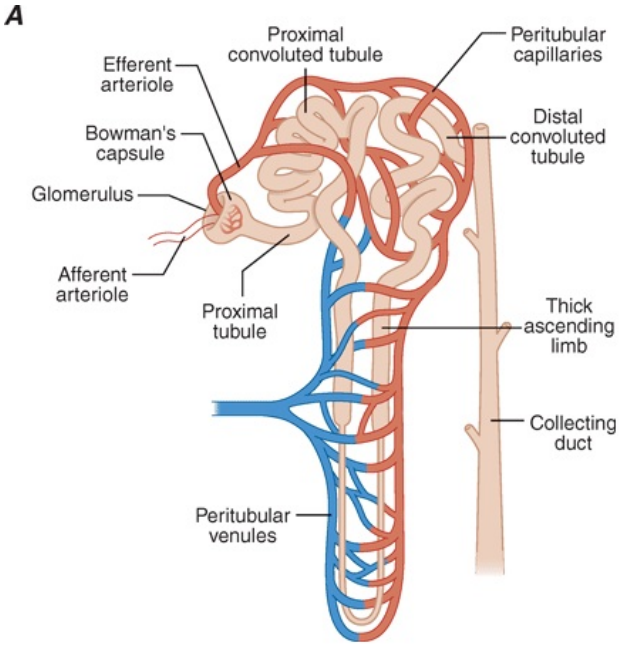

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

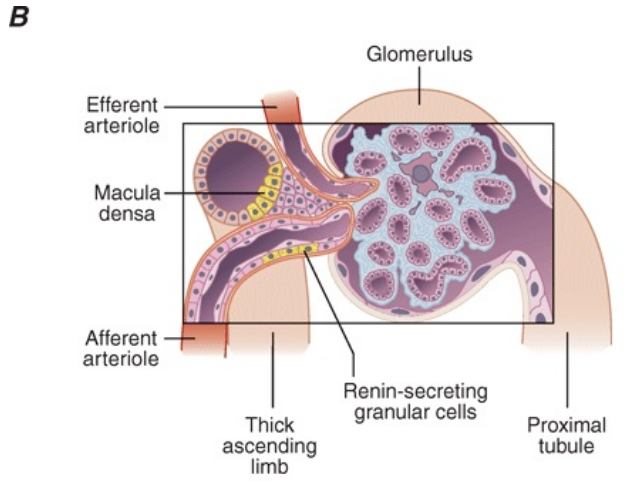

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

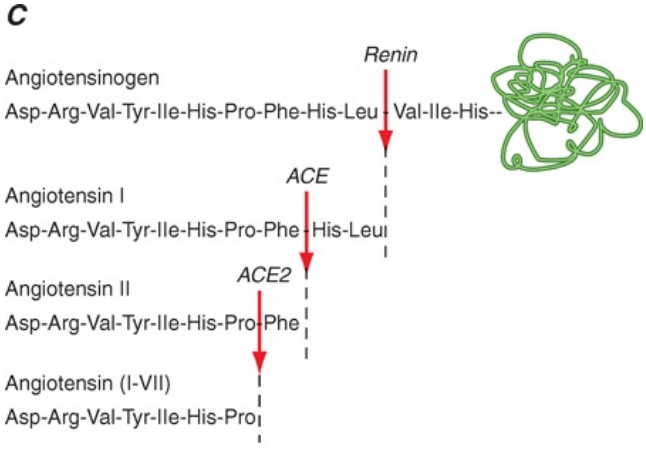

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Renal microcirculation and the renin-angiotensin system.*A.* Diagram illustrating relationships of the nephron with glomerular and peritubular capillaries. ***B.*** Expanded view of the glomerulus with its juxtaglomerular apparatus including the macula densa and adjacent afferent arteriole. ***C.*** Proteolytic processing steps in the generation of angiotensins.

The hydrostatic pressure gradient across the glomerular capillary wall is the primary driving force for glomerular filtration. Oncotic pressure within the capillary lumen, determined by the concentration of unfiltered plasma proteins, partially offsets the hydrostatic pressure gradient and opposes filtration. As the oncotic pressure rises along the length of the glomerular capillary, the driving force for filtration falls to zero on reaching the efferent arteriole. Approximately 20% of the renal plasma flow is filtered into Bowman's space, and the ratio of glomerular filtration rate (GFR) to renal plasma flow determines the filtration fraction. Several factors, mostly hemodynamic, contribute to the regulation of filtration under physiologic conditions.

Although glomerular filtration is affected by renal artery pressure, this relationship is not linear across the range of physiologic blood pressures due to autoregulation of GFR. Autoregulation of glomerular filtration is the result of three major factors that modulate either afferent or efferent arteriolar tone: these include an autonomous vasoreactive (myogenic) reflex in the afferent arteriole, *tubuloglomerular feedback* , and angiotensin II –mediated vasoconstriction of the efferent arteriole. The myogenic reflex is a first line of defense against fluctuations in renal blood flow. Acute changes in renal perfusion pressure evoke reflex constriction or dilatation of the afferent arteriole in response to increased or decreased pressure, respectively. This phenomenon helps protect the glomerular capillary from sudden changes in systolic pressure.

Tubuloglomerular feedback changes the rate of filtration and tubular flow by reflex vasoconstriction or dilatation of the afferent arteriole. Tubuloglomerular feedback is mediated by specialized cells in the thick ascending limb of the loop of Henle called the *macula densa* that act as sensors of solute concentration and tubular flow rate. With high tubular flow rates, a proxy for an inappropriately high filtration rate, there is increased solute delivery to the macula densa **(Fig.277-2*B*)** that evokes vasoconstriction of the afferent arteriole causing GFR to return toward normal. One component of the soluble signal from the macula densa is adenosine triphosphate (ATP) released by the cells during increased NaCl reabsorption. ATP is metabolized in the extracellular space to generate adenosine, a potent vasoconstrictor of the afferent arteriole. During conditions associated with a fall in filtration rate, reduced solute delivery to the macula densa attenuates the tubuloglomerular response, allowing afferent arteriolar dilatation and restoring glomerular filtration to normal levels. Angiotensin II and reactive oxygen species enhance, while nitric oxide (NO) blunts, tubuloglomerular feedback.

The third component underlying autoregulation of GFR involves angiotensin II. During states of reduced renal

blood flow, renin is released from granular cells within the wall of the afferent arteriole near the macula densa in a region called the juxtaglomerular apparatus (Fig.277-2*B*). Renin, a proteolyticenzyme, catalyzes the conversion of angiotensinogen to angiotensin I, that is subsequently converted to angiotensin II by angiotensin-converting enzyme (ACE) **(Fig.277-2*C*)**. Angiotensin II evokes vasoconstriction of the efferent arteriole, and the resulting increased glomerular hydrostatic pressure elevates filtration to normal levels.

MECHANISMS OF RENAL TUBULAR TRANSPORT

The renal tubules are composed of highly differentiated epithelia that vary dramatically in morphology and function along the nephron **(Fig.277-3)**. The cells lining the various tubular segments form monolayers connected to one another by a specialized region of the adjacent lateral membranes called the *tight junction*. Tight junctions form an occlusive barrier that separates the lumen of the tubule from the interstitial spaces surrounding the tubule and also apportions the cell membrane into discrete domains: the apical membrane facing the tubular lumen and the basolateral membrane facing the interstitium. This regionalization allows cells to allocate membrane proteins and lipids asymmetrically. Owing to this feature, renal epithelial cells are said to be *polarized*. The asymmetric assignment of membrane proteins, especially proteins mediating transport processes, provides the machinery for directional movement of fluid and solutes by the nephron.

Figure 277-3

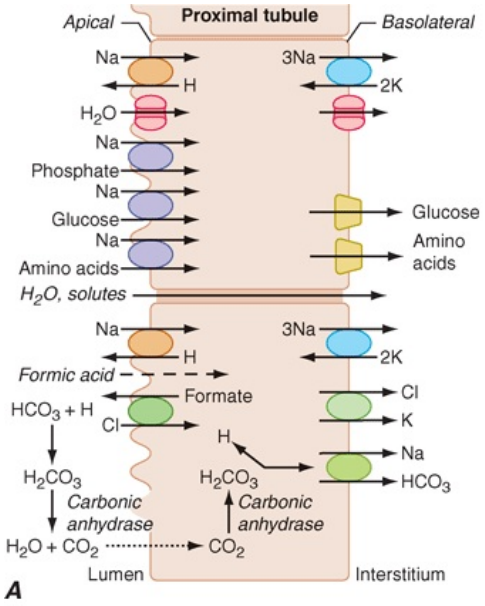

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

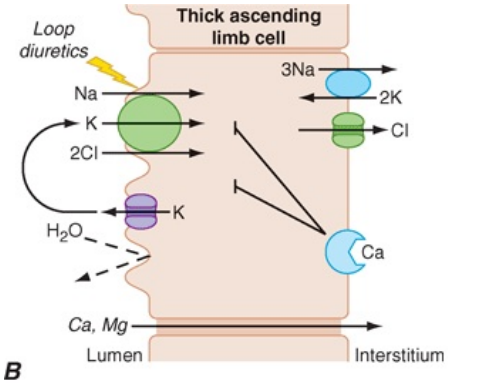

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

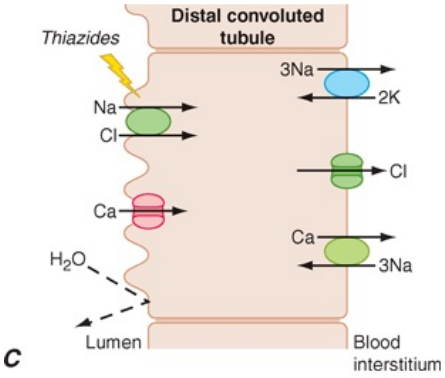

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

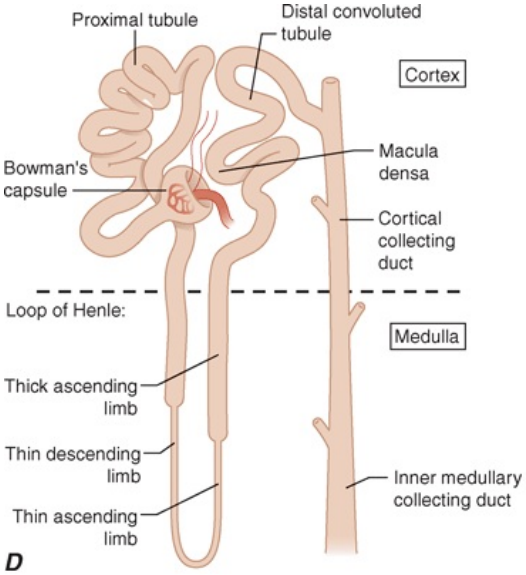

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

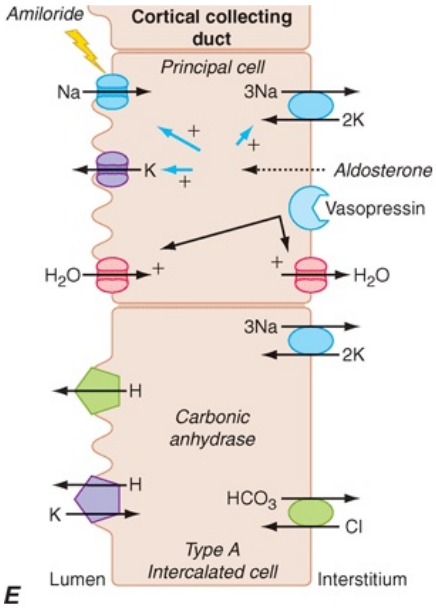

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

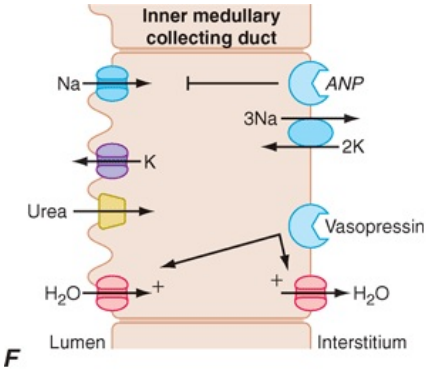

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transport activities of the major nephron segments. Representative cells from five major tubular segments are illustrated with the lumen side (apical membrane) facing left and interstitial side (basolateral membrane) facing right. ***A.*** Proximal tubular cells. ***B.*** Typical cell in the thick ascending limb of the loop of Henle. ***C.*** Distal convoluted tubular cell. ***D.*** Overview of entire nephron. ***E.*** Cortical collecting duct cells. ***F.*** Typical cell in the inner medullary collecting duct. The major membrane transporters, channels, and pumps are drawn with arrows indicating the direction of solute or water movement. For some events, the stoichiometry of transport is indicated by numerals preceding the solute. Targets for major diuretic agents are labeled. The actions of hormones are illustrated by arrows with plus signs for stimulatory effects and lines with perpendicular ends for inhibitory events. Dotted lines indicate free diffusion across cell membranes. The dashed line indicates water impermeability of cell membranes in the thick ascending limb and distal convoluted tubule.

Epithelial Solute Transport

There are two types of epithelial transport. Movement of fluid and solutes sequentially across the apical and basolateral cell membranes (or vice versa) mediated by transporters, channels, or pumps is called *cellular transport.* By contrast, movement of fluid and solutes through the narrow passageway between adjacent cells is called *paracellular transport.* Paracellular transport occurs through tight junctions, indicating that they are not completely "tight." Indeed, some epithelial cell layers allow rather robust paracellular transport to occur (*leaky epithelia*), whereas other epithelia have more effective tight junctions (*tight epithelia*). In addition, because the ability of ions to flow through the paracellular pathway determines the electrical resistance across the epithelial monolayer, leaky and tight epithelia are also referred to as low or high-resistance epithelia, respectively. The proximal tubule contains leaky epithelia, whereas distal nephron segments such as the collecting duct, contain tight epithelia. Leaky epithelia are most well suited for bulk fluid reabsorption, whereas tight epithelia allow for more refined control and regulation of transport.

Membrane Transport

Cell membranes are composed of hydrophobic lipids that repel water and aqueous solutes. The movement of solutes and water across cell membranes is made possible by discrete classes of integral membrane proteins, including channels, pumps, and transporters. These different mechanisms mediate specific types of transport activities, including *active transport* (pumps), *passive transport* (channels), *facilitated diffusion* (transporters), and *secondary active transport* (cotransporters). Active transport requires metabolic energy generated by the hydrolysis of ATP. Active transport pumps are ion-translocating ATPases, including the ubiquitous Na^+/K^+-ATPase, the H^+-ATPases, and Ca^{2+}-ATPases. Active transport creates asymmetric ion concentrations across a cell membrane and can move ions against a chemical gradient. The potential energy stored in a concentration gradient of an ion such as Na^+ can be utilized to drive transport through other mechanisms (secondary active transport). Pumps are often *electrogenic*, meaning they can create an asymmetric distribution of electrostatic charges across the membrane and establish a voltage or membrane potential. The movement of solutes through a membrane protein by simple diffusion is called passive transport. This activity is mediated by channels created by selectively permeable membrane proteins, and it allows solute or water to move across a membrane driven by favorable *concentration gradients* or *electrochemical potential*. Examples in the kidney include water channels (aquaporins), K^+ channels, epithelial Na^+ channels, and Cl^- channels. Facilitated diffusion is a specialized type of passive transport mediated by simple transporters called *carriers* or *uniporters*. For example, hexose transporters such as GLUT2 mediate glucose transport by tubular cells. These transporters are driven by the concentration gradient for glucose that is highest in extracellular fluids and lowest in the cytoplasm due to rapid metabolism. Many other transporters operate by translocating two or more ions/solutes in concert either in the same direction (*symporters* or *cotransporters*) or in opposite directions (*antiporters* or *exchangers*) across the cell membrane. The movement of two or more ions/solutes may produce no net change in the balance of electrostatic charges across the membrane (*electroneutral*), or a transport event may alter the balance of charges (*electrogenic*). Several inherited disorders of renal tubular solute and water transport occur as a consequence of mutations in genes encoding a variety of channels, transporter proteins, and their regulators **(Table 277-1)**.

Table 277-1 Inherited Disorders Affecting Renal Tubular Ion and Solute Transport

Disease or Syndrome	Gene	OMIM*

Disorders Involving the Proximal Tubule		
Proximal renal tubular acidosis	Sodium bicarbonate cotransporter (*SLC4A4*, 4q21)	604278
Fanconi-Bickel syndrome	Glucose transporter, GLUT2 (*SLC2A2*, 3q26.2)	227810
Isolated renal glycosuria	Sodium glucose cotransporter (*SLC5A2*, 16p11.2)	233100
Cystinuria		
Type I	Cystine, dibasic and neutral amino acid transporter (*SLC3A1*, 2p16.3)	220100
Nontype I	Amino acid transporter, light subunit (*SLC7A9*, 19q13.1)	600918
Lysinuric protein intolerance	Amino acid transporter (*SLC7A7*, 4q11.2)	222700
Hartnup disorder	Neutral amino acid transporter (*SLC6A19*, 5p15.33)	34500
Hereditary hypophosphatemic rickets with hypercalcemia	Sodium phosphate cotransporter (*SLC34A3*, 9q34)	241530
Renal hypouricemia		
Type 1	Urate-anion exchanger (*SLC22A12*, 11q13)	220150
Type 2	Urate transporter, GLUT9 (*SLC2A9*, 4p16.1)	612076
Dent's disease	Chloride channel, ClC-5 (*CLCN5*, Xp11.22)	300009
X-linked recessive nephrolithiasis with renal failure	Chloride channel, ClC-5 (*CLCN5*, Xp11.22)	310468
X-linked recessive hypophosphatemic rickets	Chloride channel, ClC-5 (*CLCN5*, Xp11.22)	307800
Disorders Involving the Loop of Henle		
Bartter's syndrome		
Type 1	Sodium, potassium chloride cotransporter (*SLC12A1*, 15q21.1)	241200
Type 2	Potassium channel, ROMK (*KCNJ1*, 11q24)	601678
Type 3	Chloride channel, ClC-Kb (*CLCNKB*, 1p36)	602023
with sensorineural deafness	Chloride channel accessory subunit, Barttin (*BSND*, 1p31)	602522
Autosomal dominant hypocalcemia with Bartter-like syndrome	Calcium-sensing receptor (*CASR*, 3q13.33))	601199
Familial hypocalciuric hypercalcemia	Calcium-sensing receptor (*CASR*, 3q13.33)	145980
Primary hypomagnesemia	Claudin-16 or paracellin-1 (*CLDN16* or *PCLN1*, 3q27)	248250
Isolated renal magnesium loss	Sodium potassium ATPase, γ_1-subunit (*ATP1G1*, 11q23)	154020

Disorders Involving the Distal Tubule and Collecting Duct		
Gitelman's syndrome	Sodium chloride cotransporter (*SLC12A3*, 16q13)	263800
Primary hypomagnesemia with secondary hypocalcemia	Melastatin-related transient receptor potential cation channel 6 (*TRPM6*, 9q22)	602014
Pseudoaldosteronism (Liddle's syndrome)	Epithelial sodium channel β and γ subunits (*SCNN1B*, *SCNN1G*, 16p12.1)	177200
Recessive pseudohypoaldosteronism Type 1	Epithelial sodium channel, α, β, and γ subunits (*SCNN1A*, 12p13; *SCNN1B*, *SCNN1G*, 16pp12.1)	264350
Pseudohypoaldosteronism Type 2 (Gordon's hyperkalemia-hypertension syndrome)	Kinases WNK-1, WNK-4 (*WNK1*, 12p13; *WNK4*, 17q21.31)	145260
X-linked nephrogenic diabetes insipidus	Vasopressin V2 receptor (AVPR2, Xq28)	304800
Nephrogenic diabetes insipidus (autosomal)	Water channel, aquaporin-2 (*AQP2*, 12q13)	125800
Distal renal tubular acidosis		
autosomal dominant	Anion exchanger-1 (*SLC4A1*, 17q21.31)	179800
autosomal recessive	Anion exchanger-1 (*SLC4A1*, 17q21.31)	602722
with neural deafness	Proton ATPase, β1 subunit (*ATP6V1B1*, 2p13.3)	192132
with normal hearing	Proton ATPase, 116-kD subunit (*ATP6V0A4*, 7q34)	602722

*Online Mendelian Inheritance in Man database (*http://www.ncbi.nlm.nih.gov/Omim*).

SEGMENTAL NEPHRON FUNCTIONS

Each anatomic segment of the nephron has unique characteristics and specialized functions enabling selective transport of solutes and water (Fig. 277-3). Through sequential events of reabsorption and secretion along the nephron, tubular fluid is progressively conditioned into urine. Knowledge of the major tubular mechanisms responsible for solute and water transport is critical for understanding hormonal regulation of kidney function and the pharmacologic manipulation of renal excretion.

Proximal Tubule

The proximal tubule is responsible for reabsorbing ~60% of filtered NaCl and water, as well as ~90% of filtered bicarbonate and most critical nutrients such as glucose and amino acids. The proximal tubule utilizes both cellular and paracellular transport mechanisms. The apical membrane of proximal tubular cells has an expanded surface area available for reabsorptive work created by a dense array of microvilli called the *brush border*, and leaky tight junctions enable high-capacity fluid reabsorption.

Solute and water pass through these tight junctions to enter the lateral intercellular space where absorption by the peritubular capillaries occurs. Bulk fluid reabsorption by the proximal tubule is driven by high oncotic pressure and low hydrostatic pressure within the peritubular capillaries. Physiologic adjustments in GFR made by changing efferent arteriolar tone cause proportional changes in reabsorption, a phenomenon known as *glomerulotubular balance.* For example, vasoconstriction of the efferent arteriole by angiotensin II will increase glomerular capillary hydrostatic pressure but lower pressure in the peritubular capillaries. At the same time, increased GFR and filtration fraction raise oncotic pressure near the end of the glomerular capillary. These changes, a lowered hydrostatic and increased oncotic pressure, increase the driving force for fluid absorption by the peritubular capillaries.

Cellular transport of most solutes by the proximal tubule is coupled to the Na^+ concentration gradient established by the activity of a basolateral Na^+/K^+-ATPase (Fig.277-3*A*). This active transport mechanism maintains a steep Na^+ gradient by keeping intracellular Na^+ concentrations low. Solute reabsorption is coupled to the Na^+ gradient by Na^+-dependent transporters such as Na^+-glucose and Na^+-phosphate cotransporters. In addition to the paracellular route, water reabsorption also occurs through the cellular pathway enabled by constitutively active water channels (aquaporin-1) present on both apical and basolateral membranes. Small,

local *osmotic gradients* close to plasma membranes generated by cellular Na^+ reabsorption are likely responsible for driving directional water movement across proximal tubule cells, but reabsorption along the proximal tubule does not produce a net change in tubular fluid osmolality.

Proximal tubular cells reclaim bicarbonate by a mechanism dependent on carbonic anhydrases. Filtered bicarbonate is first titrated by protons delivered to the lumen by Na^+/H^+ exchange. The resulting carbonic acid (H_2CO_3) is metabolized by brush border carbonic anhydrase to water and carbon dioxide. Dissolved carbon dioxide then diffuses into the cell, where it is enzymatically hydrated by cytoplasmic carbonic anhydrase to re-form carbonic acid. Finally, intracellular carbonic acid dissociates into free protons and bicarbonate anions, and bicarbonate exits the cell through a basolateral Na^+/HCO_3^- cotransporter. This process is saturable, resulting in urinary bicarbonate excretion when plasma levels exceed the physiologically normal range (24–26 meq/L). Carbonic anhydrase inhibitors such as acetazolamide, a class of weak diuretic agents, block proximal tubule reabsorption of bicarbonate and are useful for alkalinizing the urine.

Chloride is poorly reabsorbed throughout the first segment of the proximal tubule, and a rise in Cl^- concentration counterbalances the removal of bicarbonate anion from tubular fluid. In later proximal tubular segments, cellular Cl^- reabsorption is initiated by apical exchange of cellular formate for higher luminal concentrations of Cl^-. Once in the lumen, formate anions are titrated by H^+ (provided by Na^+/H^+ exchange) to generate neutral formic acid, which can diffuse passively across the apical membrane back into the cell where it dissociates a proton and is recycled. Basolateral Cl^- exit is mediated by a K^+/Cl^- cotransporter.

Reabsorption of glucose is nearly complete by the end of the proximal tubule. Cellular transport of glucose is mediated by apical Na^+-glucose cotransport coupled with basolateral, facilitated diffusion by a glucose transporter. This process is also saturable, leading to glycosuria when plasma levels exceed 180–200 mg/dL, as seen in untreated diabetes mellitus.

The proximal tubule possesses specific transporters capable of secreting a variety of organic acids (carboxylate anions) and bases (mostly primary amine cations). Organic anions transported by these systems include urate, ketoacid anions, and several protein-bound drugs not filtered at the glomerulus (penicillins, cephalosporins, and salicylates). Probenecid inhibits renal organic anion secretion and can be clinically useful for raising plasma concentrations of certain drugs like penicillin and oseltamivir. Organic cations secreted by the proximal tubule include various biogenic amine neurotransmitters (dopamine, acetylcholine, epinephrine, norepinephrine, and histamine) and creatinine. The ATP-dependent transporter P-glycoprotein is highly expressed in brush border membranes and secretes several medically important drugs, including cyclosporine, digoxin, tacrolimus, and various cancer chemotherapeutic agents. Certain drugs like cimetidine and trimethoprim compete with endogenous compounds for transport by the organic cation pathways. While these drugs elevate serum creatinine levels, there is no change in the actual GFR.

The proximal tubule, through distinct classes of Na^+-dependent and Na^+-independent transport systems, reabsorbs amino acids efficiently. These transporters are specific for different groups of amino acids. For example, cystine, lysine, arginine, and ornithine are transported by a system comprising two proteins encoded by the *SLC3A1* and *SLC7A9* genes. Mutations in either *SLC3A1* or *SLC7A9* impair reabsorption of these amino acids and cause the disease cystinuria. Peptide hormones such as insulin and growth hormone, $ß_2$-microglobulin, albumin, and other small proteins, are taken up by the proximal tubule through a process of absorptive endocytosis and are degraded in acidified endocytic lysosomes. Acidification of these vesicles depends on a vacuolar H^+-ATPase and Cl^- channel. Impaired acidification of endocytic vesicles because of mutations in a Cl^- channel gene (*CLCN5*) causes low molecular weight proteinuria in Dent's disease. Renal ammoniagenesis from glutamine in the proximal tubule provides a major tubular fluid buffer to ensure excretion of secreted H^+ ion as NH_4^+ by the collecting duct. Cellular K^+ levels inversely modulate ammoniagenesis, and in the setting of high serum K^+ from hypoaldosteronism, reduced ammoniagenesis facilitates the appearance of Type IV renal tubular acidosis.

Loop of Henle

The loop of Henle consists of three major segments: descending thin limb, ascending thin limb, and ascending thick limb. These divisions are based on cellular morphology and anatomic location, but also correlate with specialization of function. Approximately 15–25% of filtered NaCl is reabsorbed in the loop of Henle, mainly by the thick ascending limb. The loop of Henle has an important role in urinary concentration by contributing to the generation of a hypertonic medullary interstitium in a process called *countercurrent multiplication*. The loop of Henle is the site of action for the most potent class of diuretic agents (loop diuretics) and also contributes to reabsorption of calcium and magnesium ions.

The descending thin limb is highly water permeable owing to dense expression of constitutively active aquaporin-1 water channels. By contrast, water permeability is negligible in the ascending limb. In the thick ascending limb, there is a high level of secondary active salt transport enabled by the $Na^+/K^+/2Cl^-$ cotransporter on the apical membrane in series with basolateral Cl^- channels and Na^+/K^+-ATPase (Fig.277-3*B*). The $Na^+/K^+/2Cl^-$ cotransporter is the primary target for loop diuretics. Tubular fluid K^+ is the limiting substrate for this cotransporter (tubular concentration of K^+ is similar to plasma, about 4 meq/L), but transporter activity is maintained by K^+ recycling through an apical potassium channel. An inherited disorder of the thick

ascending limb, Bartter's syndrome, also results in a salt-wasting renal disease associated with hypokalemia and metabolic alkalosis; loss-of-function mutations in one of five distinct genes encoding components of the $Na^+/K^+/2Cl^-$ cotransporter (*NKCC2*), apical K^+ channel (*KCNJ1*), or basolateral Cl^- channel (*CLCNKB*, *BSND*), or calcium-sensing receptor (*CASR*) can cause Bartter's syndrome.

Potassium recycling also contributes to a positive electrostatic charge in the lumen relative to the interstitium that promotes divalent cation (Mg^{2+} and Ca^{2+}) reabsorption through a paracellular pathway. A Ca^{2+}-sensing, G-protein–coupled receptor (CaSR) on basolateral membranes regulates NaCl reabsorption in the thick ascending limb through dual signaling mechanisms utilizing either cyclic AMP or eicosanoids. This receptor enables a steep relationship between plasma Ca^{2+} levels and renal Ca^{2+} excretion. Loss-of-function mutations in CaSR cause familial hypercalcemic hypocalciuria because of a blunted response of the thick ascending limb to extracellular Ca^{2+}. Mutations in *CLDN16* encoding paracellin-1, a transmembrane protein located within the tight junction complex, leads to familial hypomagnesemia with hypercalcuria and nephrocalcinosis, suggesting that the ion conductance of the paracellular pathway in the thick limb is regulated.

The loop of Henle contributes to urine-concentrating ability by establishing a *hypertonic medullary interstitium* that promotes water reabsorption by the downstream inner medullary collecting duct. *Countercurrent multiplication* produces a hypertonic medullary interstitium using two countercurrent systems: the loop of Henle (opposing descending and ascending limbs) and the vasa recta (medullary peritubular capillaries enveloping the loop). The countercurrent flow in these two systems helps maintain the hypertonic environment of the inner medulla, but NaCl reabsorption by the thick ascending limb is the primary initiating event. Reabsorption of NaCl without water dilutes the tubular fluid and adds new osmoles to medullary interstitial fluid. Because the descending thin limb is highly water permeable, osmotic equilibrium occurs between the descending limb tubular fluid and the interstitial space, leading to progressive solute trapping in the inner medulla. Maximum medullary interstitial osmolality also requires partial recycling of urea from the collecting duct.

Distal Convoluted Tubule

The distal convoluted tubule reabsorbs ~5% of the filtered NaCl. This segment is composed of a tight epithelium with little water permeability. The major NaCl-transporting pathway utilizes an apical membrane, electroneutral thiazide-sensitive Na^+/Cl^- cotransporter in tandem with basolateral Na^+/K^+-ATPase and Cl^- channels (Fig.277-3*C*). Apical Ca^{2+}-selective channels (TRPV5) and basolateral Na^+/Ca^{2+} exchange mediate calcium reabsorption in the distal convoluted tubule. Ca^{2+} reabsorption is inversely related to Na^+ reabsorption and is stimulated by parathyroid hormone. Blocking apical Na^+/Cl^- cotransport will reduce intracellular Na^+, favoring increased basolateral Na^+/Ca^{2+} exchange and passive apical Ca^{2+} entry. Loss-of-function mutations of *SLC12A3* encoding the apical Na^+/Cl^- cotransporter cause Gitelman's syndrome, a salt-wasting disorder associated with hypokalemic alkalosis and hypocalciuria. Mutations in genes encoding WNK kinases, WNK-1 and WNK-4, cause pseudohypoaldosteronism type II or Gordon's syndrome characterized by familial hypertension with hyperkalemia. WNK kinases influence the activity of several tubular ion transporters. Mutations in this disorder lead to overactivity of the apical Na^+/Cl^- cotransporter in the distal convoluted tubule as the primary stimulus for increased salt reabsorption, extracellular volume expansion, and hypertension. Hyperkalemia may be caused by diminished activity of apical K^+ channels in the collecting duct, a primary route for K^+ secretion. Mutations in *TRPM6* encoding Mg^{2+} permeable ion channels also cause familial hypomagnesemia with hypocalcemia. A molecular complex of TRPM6 and TRPM7 proteins is critical for Mg^{2+} reabsorption in the distal convoluted tubule.

Collecting Duct

The collecting duct modulates the final composition of urine. The two major divisions, the cortical collecting duct and inner medullary collecting duct, contribute to reabsorbing ~4–5% of filtered Na^+ and are important for hormonal regulation of salt and water balance. The cortical collecting duct contains *high-resistance epithelia* with two cell types. Principal cells are the main water, Na^+-reabsorbing, and K^+-secreting cells, and the site of action of aldosterone, K^+-sparing diuretics, and mineralocorticoid receptor antagonists such as spironolactone. The other cells are type A and B intercalated cells. Type A intercalated cells mediate acid secretion and bicarbonate reabsorption also under the influence of aldosterone. Type B intercalated cells mediate bicarbonate secretion and acid reabsorption.

Virtually all transport is mediated through the cellular pathway for both principal cells and intercalated cells. In principal cells, passive apical Na^+ entry occurs through the amiloride-sensitive, epithelial Na^+ channel (ENaC) with basolateral exit via the Na^+/K^+-ATPase (Fig.277-3*E*). This Na^+ reabsorptive process is tightly regulated by aldosterone and is physiologically activated by a variety of proteolytic enzymes that cleave extracellular domains of ENaC; plasmin in the tubular fluid of nephrotic patients, for example, activates ENaC, leading to sodium retention. Aldosterone enters the cell across the basolateral membrane, binds to a cytoplasmic mineralocorticoid receptor, and then translocates into the nucleus, where it modulates gene transcription, resulting in increased Na^+ reabsorption and K^+ secretion. Activating mutations in ENaC increase Na^+ reclamation and produce hypokalemia, hypertension, and metabolic alkalosis (Liddle's syndrome). The potassium-sparing diuretics amiloride and triamterene block ENaC, causing reduced Na^+ reabsorption.

Principal cells secrete K^+ through an apical membrane potassium channel. Several forces govern the secretion of K^+. Most importantly, the high intracellular K^+ concentration generated by Na^+/K^+-ATPase creates a favorable concentration gradient for K^+ secretion into tubular fluid. With reabsorption of Na^+ without an accompanying anion, the tubular lumen becomes negative relative to the cell interior, creating a favorable electrical gradient for secretion of potassium. When Na^+ reabsorption is blocked, the electrical component of the driving force for K^+ secretion is blunted and this explains lack of excess urinary K^+ loss during treatment with potassium-sparing diuretics. K^+ secretion is also promoted by aldosterone actions that increase regional Na^+ transport favoring more electronegativity and by increasing the number and activity of potassium channels. Fast tubular fluid flow rates that occur during volume expansion or diuretics acting "upstream" of the cortical collecting duct also increase K^+ secretion, as does the presence of relatively nonreabsorbable anions (including bicarbonate and semisynthetic penicillins) that contribute to the lumen-negative potential. Off-target effects of certain antibiotics such as trimethoprim and pentamidine, block ENaCs and predispose to hyperkalemia, especially when renal K^+ handling is impaired for other reasons. Principal cells, as described below, also participate in water reabsorption by increased water permeability in response to vasopressin.

Intercalated cells do not participate in Na^+ reabsorption but instead mediate acid-base secretion. These cells perform two types of transport: active H^+ transport mediated by H^+-ATPase (proton pump), and Cl^-/HCO_3^- exchange. Intercalated cells arrange the two transport mechanisms on opposite membranes to enable either acid or base secretion. Type A intercalated cells have an apical proton pump that mediates acid secretion and a basolateral Cl^-/HCO_3^- anion exchanger for bicarbonate reabsorption (Fig. 277-3*E*); aldosterone increases the number of H^+-ATPase pumps, sometimes contributing to the development of metabolic alkalosis. By contrast, type B intercalated cells have the anion exchanger on the apical membrane to mediate bicarbonate secretion while the proton pump resides on the basolateral membrane to enable acid reabsorption. Under conditions of acidemia, the kidney preferentially uses type A intercalated cells to secrete the excess H^+ and generate more HCO_3^-. The opposite is true in states of bicarbonate excess with alkalemia where the type B intercalated cells predominate. An extracellular protein called *hensin* mediates this adaptation.

Inner medullary collecting duct cells share many similarities with principal cells of the cortical collecting duct. They have apical Na^+ and K^+ channels that mediate Na^+ reabsorption and K^+ secretion, respectively (Fig.277-3*F*). Inner medullary collecting duct cells also have vasopressin-regulated water channels (aquaporin-2 on the apical membrane, aquaporin-3 and -4 on the basolateral membrane). The antidiuretic hormone vasopressin binds to the V2 receptor on the basolateral membrane and triggers an intracellular signaling cascade through G-protein–mediated activation of adenylyl cyclase, resulting in an increase in the cellular levels of cyclic AMP. This signaling cascade stimulates the insertion of water channels into the apical membrane of the inner medullary collecting duct cells to promote increased water permeability. This increase in permeability enables water reabsorption and production of concentrated urine. In the absence of vasopressin, inner medullary collecting duct cells are water impermeable, and urine remains dilute.

Sodium reabsorption by inner medullary collecting duct cells is also inhibited by the natriuretic peptides called *atrial natriuretic peptide* or *renal natriuretic peptide* (urodilatin); the same gene encodes both peptides but uses different posttranslational processing of a common preprohormone to generate different proteins. Atrial natriuretic peptides are secreted by atrial myocytes in response to volume expansion, whereas urodilatin is secreted by renal tubular epithelia. Natriuretic peptides interact with either apical (urodilatin) or basolateral (atrial natriuretic peptides) receptors on inner medullary collecting duct cells to stimulate guanylyl cyclase and increase levels of cytoplasmic cGMP. This effect in turn reduces the activity of the apical Na^+ channel in these cells and attenuates net Na^+ reabsorption, producing natriuresis.

The inner medullary collecting duct transports urea out of the lumen, returning urea to the interstitium, where it contributes to the hypertonicity of the medullary interstitium. Urea is recycled by diffusing from the interstitium into the descending and ascending limbs of the loop of Henle.

HORMONAL REGULATION OF SODIUM AND WATER BALANCE

The balance of solute and water in the body is determined by the amounts ingested, distributed to various fluid compartments, and excreted by skin, bowel, and kidneys. *Tonicity*, the osmolar state determining the volume behavior of cells in a solution, is regulated by water balance **(Fig.277-4*A*)**, and *extracellular blood volume* is regulated by Na^+ balance **(Fig.277-4*B*)**. The kidney is a critical modulator of both physiologic processes.

Figure 277-4

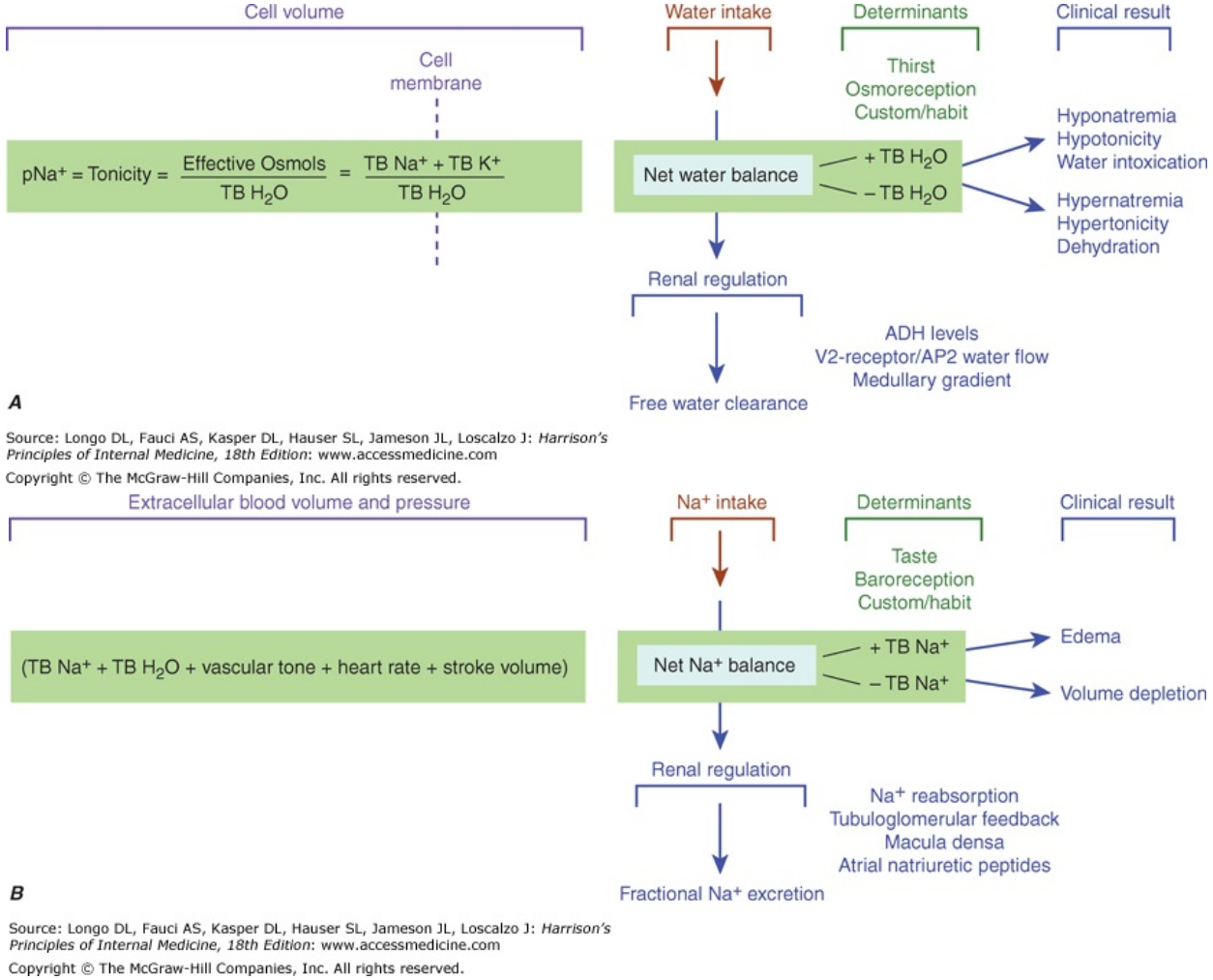

Determinants of sodium and water balance. ***A.*** Plasma Na^+ concentration is a surrogate marker for plasma tonicity, the volume behavior of cells in a solution. Tonicity is determined by the number of effective osmols in the body divided by the total body H_2O (TB H_2O), which translates simply into the total body Na (TB Na^+) and anions outside the cell separated from the total body K (TB K^+) inside the cell by the cell membrane. Net water balance is determined by the integrated functions of thirst, osmoreception, Na reabsorption, vasopressin release, and the strength of the medullary gradient in the kidney, keeping tonicity within a narrow range of osmolality around 280 mosmol/L. When water metabolism is disturbed and total body water increases, hyponatremia, hypotonicity, and water intoxication occur; when total body water decreases, hypernatremia, hypertonicity, and dehydration occur. ***B.*** Extracellular blood volume and pressure are an integrated function of total body Na^+ (TB Na^+), total body H_2O (TB H_2O), vascular tone, heart rate, and stroke volume that modulates volume and pressure in the vascular tree of the body. This extracellular blood volume is determined by net Na balance under the control of taste, baroreception, habit, Na^+ reabsorption, macula densa/tubuloglomerular feedback, and natriuretic peptides. When Na^+ metabolism is disturbed and total body Na^+ increases, edema occurs; when total body Na^+ is decreased, volume depletion occurs. ADH, antidiuretic hormone; AQP2, aquaporin-2.

Water Balance

Tonicity depends on the variable concentration of *effective osmoles* inside and outside the cell causing water to move in either direction across its membrane. Classic effective osmoles, like Na^+, K^+, and their anions, are solutes trapped on either side of a cell membrane, where they collectively partition and obligate water to move and find equilibrium in proportion to retained solute; Na^+/K^+-ATPase keeps most K^+ inside cells and most Na^+ outside. Normal tonicity (~280 mosmol/L) is rigorously defended by osmoregulatory mechanisms that control water balance to protect tissues from inadvertent *dehydration* (cell shrinkage) or *water intoxication* (cell swelling), both of which are deleterious to cell function (Fig. 277-4*A*).

The mechanisms that control osmoregulation are distinct from those governing extracellular volume, although there is some shared physiology in both processes. While cellular concentrations of K^+ have a determinant role in any level of tonicity, the routine surrogate marker for assessing clinical tonicity is the concentration of serum Na^+. Any reduction in total body water, which raises the Na^+ concentration, triggers a brisk sense of thirst and conservation of water by decreasing renal water excretion mediated by release of vasopressin from the posterior pituitary. Conversely, a decrease in plasma Na^+ concentration triggers an increase in renal water excretion by suppressing the secretion of vasopressin. While all cells expressing mechanosensitive TRPV1, 2, or 4 channels, among potentially other sensors, respond to changes in tonicity by altering their volume and

Ca^{2+} concentration, only $TRPV^{+}$ neuronal cells connected to the organum vasculosum of the lamina terminalis are *osmoreceptive*. Only these cells, because of their neural connectivity and adjacency to a minimal blood-brain barrier, modulate the downstream release of vasopressin by the posterior lobe of the pituitary gland. Secretion is stimulated primarily by changing tonicity and secondarily by other nonosmotic signals such as variable blood volume, stress, pain, nausea, and some drugs. The release of vasopressin by the posterior pituitary increases linearly as plasma tonicity rises above normal, although this varies, depending on the perception of extracellular volume (one form of cross-talk between mechanisms that adjudicate blood volume and osmoregulation). Changing the intake or excretion of water provides a means for adjusting plasma tonicity; thus, osmoregulation governs water balance.

The kidneys play a vital role in maintaining water balance through the regulation of renal water excretion. The ability to concentrate urine to an osmolality exceeding that of plasma enables water conservation, while the ability to produce urine more dilute than plasma promotes excretion of excess water. For water to enter or exit a cell, the cell membrane must express aquaporins. In the kidney, aquaporin 1 is constitutively active in all water-permeable segments of the proximal and distal tubules, while vasopressin-regulated aquaporins 2, 3, and 4 in the inner medullary collecting duct promote rapid water permeability. Net water reabsorption is ultimately driven by the osmotic gradient between dilute tubular fluid and a hypertonic medullary interstitium.

Sodium Balance

The perception of *extracellular blood volume* is determined, in part, by the integration of arterial tone, cardiac stroke volume, heart rate, and the water and solute content of extracellular fluid. Na^{+} and accompanying anions are the most abundant extracellular effective osmols and together support a blood volume around which pressure is generated. Under normal conditions, this volume is regulated by sodium balance (Fig.277-4*B*), and the balance between daily Na^{+} intake and excretion is under the influence of *baroreceptors* in regional blood vessels and vascular hormone sensors modulated by atrial natriuretic peptides, the renin-angiotensin-aldosterone system, Ca^{2+} signaling, adenosine, vasopressin, and the neural adrenergic axis. If Na^{+} intake exceeds Na^{+} excretion (positive Na^{+} balance), then an increase in blood volume will trigger a proportional increase in urinary Na^{+} excretion. Conversely, when Na^{+} intake is less than urinary excretion (negative Na^{+} balance), blood volume will decrease and trigger enhanced renal Na^{+} reabsorption, leading to decreased urinary Na^{+} excretion.

The renin-angiotensin-aldosterone system is the best-understood hormonal system modulating renal Na^{+} excretion. Renin is synthesized and secreted by granular cells in the wall of the afferent arteriole. Its secretion is controlled by several factors, including β_1-adrenergic stimulation to the afferent arteriole, input from the macula densa, and prostaglandins. Renin and ACE activity eventually produce angiotensin II that directly or indirectly promotes renal Na^{+} and water reabsorption. Stimulation of proximal tubular Na^{+}/H^{+} exchange by angiotensin II directly increases Na^{+} reabsorption. Angiotensin II also promotes Na^{+} reabsorption along the collecting duct by stimulating aldosterone secretion by the adrenal cortex. Constriction of the efferent glomerular arteriole by angiotensin II indirectly increases the filtration fraction and raises peritubular capillary oncotic pressure to promote tubular Na^{+} reabsorption. Finally, angiotensin II inhibits renin secretion through a negative feedback loop. Alternative metabolism of angiotensin by ACE2 generates the vasodilatory peptide angiotensin 1-7 that acts through Mas receptors to counterbalance several actions of angiotensin II on blood pressure and renal function (Fig. 277-2*C*).

Aldosterone is synthesized and secreted by granulosa cells in the adrenal cortex. It binds to cytoplasmic mineralocorticoid receptors in the collecting duct principal cells that increase activity of ENaC, apical membrane K^{+} channel, and basolateral Na^{+}/K^{+}-ATPase. These effects are mediated in part by aldosterone-stimulated transcription of the gene encoding serum/glucocorticoid-induced kinase 1 (SGK1). The activity of ENaC is increased by SGK1-mediated phosphorylation of Nedd4-2, a protein that promotes recycling of the Na^{+} channel from the plasma membrane. Phosphorylated Nedd4-2 has impaired interactions with ENaC, leading to increased channel density at the plasma membrane and increased capacity for Na^{+} reabsorption by the collecting duct.

Chronic exposure to aldosterone causes a decrease in urinary Na^{+} excretion lasting only a few days, after which Na^{+} excretion returns to previous levels. This phenomenon, called *aldosterone escape*, is explained by decreased proximal tubular Na^{+} reabsorption following blood volume expansion. Excess Na^{+} that is not reabsorbed by the proximal tubule overwhelms the reabsorptive capacity of more distal nephron segments. This escape may be facilitated by atrial natriuretic peptides that lose their effectiveness in the clinical settings of heart failure, nephrotic syndrome, and cirrhosis, leading to severe Na^{+} retention and volume overload.

FURTHER READINGS

Bhalla V, Hallows KR: Mechanisms of ENaC regulation and clinical implications. J Am Soc Nephrol 19:1845, 2008[PMID: 18753254]

Kopan R et al: Molecular insights into segmentation along the proximal–distal axis of the nephron. J Am Soc Nephrol 18:2014, 2007[PMID: 17568016]

Mange K et al: Language guiding therapy: The case of dehydration versus volume depletion. Ann Intern Med 127:848, 1997[PMID: 9382413]

Mathieson PW: Update on the podocyte. Curr Opin Nephrol Hypertens 18:206, 2009[PMID: 19374006]

Ribes D et al: Transcriptional control of epithelial differentiation during kidney development. J Am Soc Nephrol 14:S9, 2003

Schlöndorff D, Banas B: The mesangial cell revisited: No cell is an island. J Am Soc Nephrol 20: 1179, 2009[PMID: 21690147]

Schrier RT: Decreased effective blood volume in edematous disorders: What does this mean? J Am Soc Nephrol 18:2028, 2007[PMID: 17568020]

Schrier RW, Ecder T: Gibbs memorial lecture: Unifying hypothesis of body fluid volume regulation. Mt Sinai J Med 68:350, 2001[PMID: 11687862]

Vaughan mr, Quaggin SE: How do mesangial and endothelial cells form the glomerular tuft? J Am Soc Nephrol 19:24, 2008[PMID: 18178797]

Verbalis JG: How does the brain sense osmolality? J Am Soc Nephrol 18:3056, 2007[PMID: 18003769]

Wang WH, Giebisch G: Regulation of potassium (K) handling in the renal collecting duct. Pflugers Arch 458:157, 2009[PMID: 18839206]

Wiseman AC, Linas S: Disorders of potassium and acid-base balance. Am J Kidney Dis 45:941, 2005[PMID: 15861362]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ADAPTION OF THE KIDNEY TO RENAL INJURY: INTRODUCTION

The size of a kidney and the total number of nephrons formed late in embryologic development depend on the degree to which the ureteric bud undergoes branching morphogenesis. Humans have between 225,000 and 900,000 nephrons in each kidney, a number that mathematically hinges on whether ureteric branching goes to completion or is terminated prematurely by one or two cycles. Although the signaling mechanisms regulating cycle number are incompletely understood, these final rounds of branching likely determine how well the kidney will adapt to the physiologic demands of blood pressure and body size, various environmental stresses, or unwanted inflammation leading to chronic renal failure.

One of the intriguing generalities regarding chronic renal failure is that residual nephrons hyperfunction to compensate for the loss of those nephrons succumbing to primary disease. This compensation depends on adaptive changes produced by renal hypertrophy and adjustments in *tubuloglomerular feedback* and *glomerulotubular balance*, as advanced in the *intact nephron hypothesis* by Neal Bricker in 1969. Some physiologic adaptations to nephron loss also produce unintended clinical consequences explained by Bricker's *trade-off hypothesis* in 1972, and eventually some adaptations accelerate the deterioration of residual nephrons, as described by Barry Brenner in his *hyperfiltration hypothesis* in 1982. These three important notions regarding chronic renal failure form a conceptual basis for understanding common pathophysiology leading to uremia.

COMMON MECHANISMS OF PROGRESSIVE RENAL DISEASE

When the initial complement of nephrons is reduced by a sentinel event, such as unilateral nephrectomy, the remaining kidney adapts by enlarging and increasing its glomerular filtration rate. If the kidneys were initially normal, the filtration rate usually returns to 80% of normal for two kidneys. The remaining kidney grows by *compensatory renal hypertrophy* with very little cellular proliferation. This unique event is accomplished by increasing the size of each cell along the nephron, which is accommodated by the elasticity or growth of interstitial spaces under the renal capsule. The mechanism of this *compensatory renal hypertrophy* is only partially understood; studies suggest roles for angiotensin II transactivation of heparin-binding epithelial growth factor, PI3K, and $p27^{kip1}$, a cell cycle protein that prevents tubular cells exposed to angiotensin II from proliferating, and the mammalian target of rapamycin (mTOR), which mediates new protein synthesis.

Hyperfiltration during pregnancy or in humans born with one kidney or who lose one to trauma or transplantation generally produces no ill consequences. By contrast, experimental animals that undergo resection of 80% of their renal mass, or humans who have persistent injury that destroys a comparable amount of renal tissue, progress to end-stage disease **(Fig. 278-1)**. Clearly, there is a critical amount of primary nephron loss that produces maladaptive deterioration in remaining nephrons. This maladaptive response is referred to clinically as *renal progression,* and the pathologic correlate of renal progression is the relentless advance of tubular atrophy and tissue fibrosis. The mechanism for this maladaptive response is the focus of intense investigation. A unified theory of *renal progression* is just starting to emerge, and most importantly, this progression follows a final common pathway regardless of whether renal injury begins in glomeruli or within the tubulointerstitium.

Figure 278-1

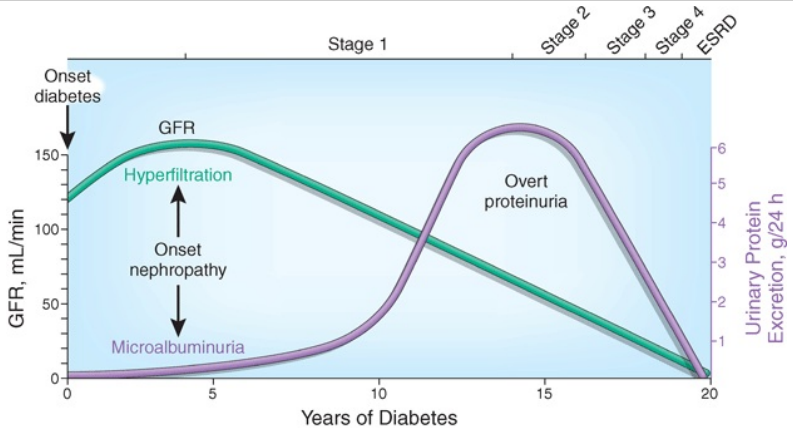

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Progression of chronic renal injury. Although various types of renal injury have their own unique rates of progression, one of the best understood is that associated with type I diabetic nephropathy. Notice the early increase in glomerular filtration rate, followed by inexorable decline associated with increasing proteinuria. Also indicated is the National Kidney Foundation K/DOQI classification of the stages of chronic kidney disease.

There are six mechanisms that hypothetically unify this final common pathway. If injury begins in glomeruli, these sequential steps build on each other: (1) persistent glomerular injury produces local hypertension in capillary tufts, increases their single-nephron glomerular filtration rate and engenders protein leak into the tubular fluid; (2) significant glomerular proteinuria, accompanied by increases in the local production of angiotensin II, facilitates a downstream cytokine bath that induces the accumulation of interstitial mononuclear cells; (3) the initial appearance of interstitial neutrophils is quickly replaced by a gathering of macrophages and T lymphocytes, which form a nephritogenic immune response producing interstitial nephritis; (4) some tubular epithelia respond to this inflammation by disaggregating from their basement membrane and adjacent sister cells to undergo *epithelial-mesenchymal transitions* forming new interstitial fibroblasts; and finally (5) surviving fibroblasts lay down

a collagenous matrix that disrupts adjacent capillaries and tubular nephrons, eventually leaving an acellular scar. The details of these complex events are outlined below **(Fig. 278-2)**.

Figure 278-2

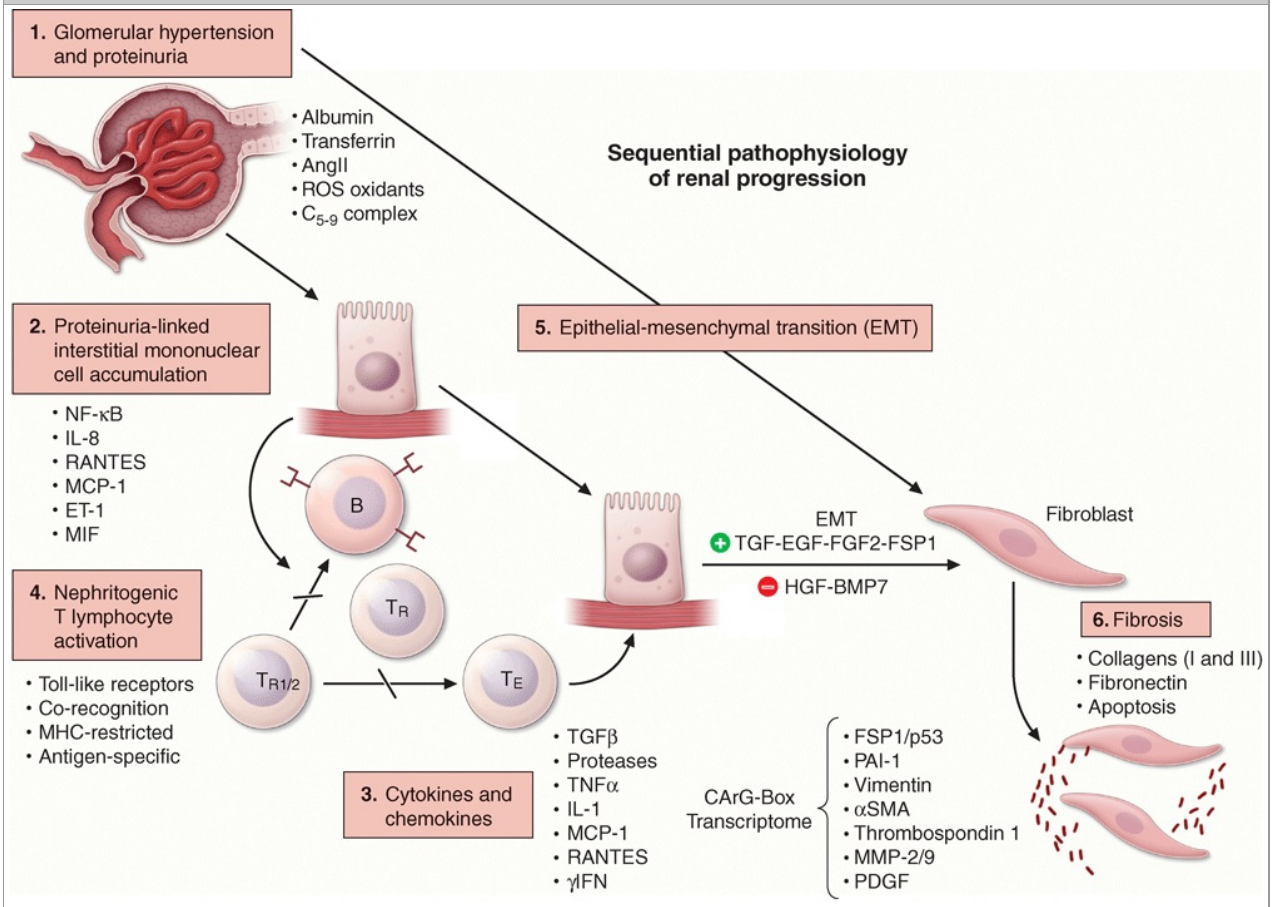

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanisms of renal progression. The general mechanisms of renal progression advance sequentially through six stages that include hyperfiltration, proteinuria, cytokine bath, mononuclear cell infiltration, epithelial-mesenchymal transition, and fibrosis. *(Modified from Harris and Neilson.)*

Significant ablation of renal mass results in *hyperfiltration* characterized by an increase in the rate of *single-nephron glomerular filtration.* The remaining nephrons lose their ability to autoregulate, and systemic hypertension is transmitted to the glomerulus. Both the *hyperfiltration* and *intraglomerular* hypertension stimulate the eventual appearance of glomerulosclerosis. Angiotensin II acts as an essential mediator of increased *intraglomerular* capillary pressure by selectively increasing efferent arteriolar vasoconstriction relative to afferent arteriolar tone. Angiotensin II impairs glomerular size selectivity, induces protein ultrafiltration, and increases intracellular Ca^{2+} in podocytes, which alters podocyte function. Diverse vasoconstrictor mechanisms, including blockade of nitric oxide synthase and activation of angiotensin II and thromboxane receptors, can also induce oxidative stress in surrounding renal tissue. Finally, the effects of aldosterone on increasing renal vascular resistance and glomerular capillary pressure, or stimulating plasminogen activator inhibitor-1, facilitate fibrogenesis and may complement the detrimental activity of angiotensin II.

On occasion, inflammation that begins in the renal interstitium disables tubular reclamation of filtered protein, producing mild nonselective proteinuria. Renal inflammation that initially damages glomerular capillaries often spreads to the tubulointerstitium in association with heavier proteinuria. Many clinical observations support the association of worsening *glomerular proteinuria* with *renal progression*. The simplest explanation for this expansion of mononuclear cells is that increasingly severe proteinuria triggers a downstream inflammatory cascade in tubular epithelial cells, producing interstitial nephritis, fibrosis, and tubular atrophy. As albumin is an abundant polyanion in plasma and can bind a variety of cytokines, chemokines, and lipid mediators, it is likely these small albumin-carried molecules initiate the tubular inflammation brought on by proteinuria. Furthermore, glomerular injury either adds activated mediators to the proteinuric filtrate or alters the balance of cytokine inhibitors and activators such that attainment of a critical level of activated cytokines eventually damages downstream tubular nephron.

Tubular epithelia bathed in these complex mixtures of proteinuric cytokines respond by increasing their secretion of chemokines and relocating NF-kappa;B to the nucleus to induce the proinflammatory release of transforming growth factor β (TGF-β), platelet-derived growth factor–BB (PDGF-BB), and fibroblast growth factor 2 (FGF-2). Inflammatory cells are drawn into the renal interstitium by this cytokine milieu. This interstitial spreading reduces the likelihood that the kidney will survive. The immunologic mechanisms for spreading include loss of tolerance to parenchymal self, immune deposits that share cross-reactive epitopes in either compartment or glomerular injury that reveals a new interstitial epitope. Drugs, infection, and metabolic defects also induce autoimmunity through toll-like receptors (TLRs) that bind to ligands with an immunologically distinct molecular pattern. Bacterial and viral ligands activate TLRs but interestingly so do Tamm-Horsfall protein, bacterial CpG repeats, and RNA that is released nonspecifically from injured tubular cells. Dendritic cells and macrophages are subsequently activated, and circulating T cells engage in the formal cellular immunologic response.

Nephritogenic interstitial T cells are a mix of $CD4^+$ helper, $CD17^+$ effector, and $CD8^+$ cytotoxic lymphocytes. Presumptive evidence of antigen-driven T cells found by examining the DNA sequence of T cell receptors suggests a polyclonal expansion

responding to multiple epitopes. Some experimental interstitial lesions are histologically analogous to a cutaneous delayed-type hypersensitivity reaction, and more intense reactions sometimes induce granuloma formation. The cytotoxic activity of antigen-reactive T cells probably accounts for tubular cell destruction and atrophy. Cytotoxic T cells synthesize proteins with serine esterase activity as well as pore-forming proteins, which can effect membrane damage much like the activated membrane attack complex of the complement cascade. Such enzymatic activity provides a structural explanation for target cell lysis.

One long-term consequence of tubular epithelia and adjacent endothelia exposed to cytokines is the profibrotic activation of *epithelial/endothelial-mesenchymal transition* (EMT). Persistent cytokine activity during renal inflammation and disruption of underlying basement membranes by local proteases initiates the process of transition. Rather than collapsing into the tubular lumens and dying, some epithelia become fibroblasts while translocating back into the interstitial space behind deteriorating tubules through holes in the ruptured basement membrane; the contribution of endothelial cells from interstitial vessels may be equally important. Wnt proteins, integrin-linked kinases, insulin-like growth factors, EGF, FGF-2, and TGF-ß are among the classic modulators of EMT. Fibroblasts that deposit collagen during fibrogenesis also replicate locally at sites of persistent inflammation. Estimates indicate that more than half of the total fibroblasts found in fibrotic renal tissues are products of the proliferation of newly transitioned or preexisting fibroblasts. Fibroblasts are stimulated to multiply by activation of cognate cell-surface receptors for PDGF and TGF-ß.

Tubulointerstitial scars are composed principally of fibronectin, collagen types I and III, and tenascin, but other glycoproteins such as thrombospondin, SPARC, osteopontin, and proteoglycan may be also important. Although tubular epithelia can synthesize collagens I and III and are modulated by a variety of growth factors, these epithelia disappear through transition and tubular atrophy, leaving fibroblasts as the major contributor to matrix production. After fibroblasts acquire a synthetic phenotype, expand their population, and locally migrate around areas of inflammation, they begin to deposit fibronectin, which provides a scaffold for interstitial collagens. When fibroblasts outdistance their survival factors, they die from apoptosis, leaving an acellular scar.

RESPONSE TO REDUCTION IN NUMBERS OF FUNCTIONING NEPHRONS

As mentioned above, the response to the loss of many functioning nephrons produces an increase in renal blood flow with glomerular *hyperfiltration*, which is the result of increased vasoconstriction in postglomerular efferent arterioles relative to preglomerular afferent arterioles, increasing the *intraglomerular* capillary pressure and filtration fraction. Persistent intraglomerular hypertension is associated with progressive nephron destruction. Although the hormonal and metabolic factors mediating *hyperfiltration* are not fully understood, a number of vasoconstrictive and vasodilatory substances have been implicated, chief among them being angiotensin II. Angiotensin II incrementally vasoconstricts the efferent arteriole, and studies in animals and humans demonstrate that interruption of the renin-angiotensin system with either angiotensin-converting inhibitors or angiotensin II receptor blockers will decrease *intraglomerular* capillary pressure, decrease proteinuria, and slow the rate of nephron destruction. The vasoconstrictive agent endothelin has also been implicated in *hyperfiltration*, and increases in afferent vasodilatation have been attributed, at least in part, to local prostaglandins and release of endothelium-derived nitric oxide. Finally, hyperfiltration may be mediated in part by a resetting of the kidney's intrinsic autoregulatory mechanism of glomerular filtration by a *tubuloglomerular feedback system*. This feedback originates from the macula densa and modulates renal blood flow and glomerular filtration (see Chap. 277).

Even with the loss of functioning nephrons, there is some continued maintenance of *glomerulotubular balance*, by which the residual tubules adapt to increases in *single nephron glomerular filtration* with appropriate alterations in reabsorption or excretion of filtered water and solutes in order to maintain homeostasis. *Glomerulotubular balance* results both from tubular hypertrophy and from regulatory adjustments in tubular oncotic pressure or solute transport along the proximal tubule. Some studies indicate these alterations in tubule size and function may themselves be maladaptive, and as a trade-off, predispose to further tubule injury.

TUBULAR FUNCTION IN CHRONIC RENAL FAILURE

Sodium

Na^+ ions are reclaimed along many parts of the nephron by various transport mechanisms (see Chapter 278). This transport function and its contribution to maintaining extracellular blood volume usually remains near normal until limitations from advanced renal disease inadequately excrete dietary Na^+ intake. Prior to this point and throughout renal progression, increasing the fractional excretion of Na^+ in final urine at progressively reduced rates of glomerular filtration provides a mechanism of early adaptation. Na^+ excretion increases predominantly by decreasing Na^+ reabsorption in the loop of Henle and distal nephron. An increase in the osmotic obligation of residual nephrons increases tubular water and lowers the concentration of Na^+ in tubular fluid, reducing efficient Na^+ reclamation; increased excretion of inorganic and organic anions also obligates more Na^+ excretion. In addition, hormonal influences, notably increased expression of atrial natriuretic peptides that increase distal Na^+ excretion, play an important role in maintaining net Na^+ excretion. Although many details of these adjustments are only understood conceptually, it is an example of a trade-off by which initial adjustments following the loss of functioning nephrons leads to compensatory responses that maintain homeostasis. Eventually, with advancing nephron loss, the atrial natriuretic peptides lose their effectiveness and Na^+ retention results in intravascular volume expansion, edema, and worsening hypertension.

Urinary Dilution and Concentration

Patients with progressive renal injury gradually lose the capacity either to dilute or concentrate their urine, and urine osmolality becomes relatively fixed about 350 mOsm/L (specific gravity ~1.010). Although the absolute ability of a single nephron to excrete water free of solute may not be impaired, the reduced number of functioning nephrons obligates increased fractional solute excretion by residual nephrons, and this greater obligation impairs the ability to dilute tubular fluid maximally. Similarly, urinary concentrating ability falls as more water is needed to hydrate an increasing solute load. Tubulointerstitial damage also creates insensitivity to the antidiuretic effects of vasopressin along the collecting duct or loss of the medullary gradient that eventually disturbs control of variation in urine osmolality. Patients with moderate degrees of chronic renal failure often complain of *nocturia* as a manifestation of this fixed urine osmolality, and they are prone to extracellular volume depletion if they do not keep up with the persistent loss of Na^+ or to hypotonicity if they drink too much water.

Potassium

Renal excretion is the major pathway for reducing excess total body K^+. Normally, the kidney excretes 90% of dietary K^+, while 10% is excreted in the stool with a trivial amount lost to sweat. Although the colon possesses some capacity to increase K^+ excretion—up to 30% of ingested K^+ may be excreted in the stool of patients with worsening renal failure, the majority of the

K^+ load continues to be excreted by the kidneys due to elevation in levels of serum K^+ that increase filtered load. Aldosterone also regulates collecting duct Na^+ reabsorption and K^+ secretion. Aldosterone is released from the adrenal cortex not only in response to the renin-angiotensin system but also in direct response to elevated levels of serum K^+ and, for a while, a compensatory increase in the capacity of the collecting duct to secrete K^+ keeps up with *renal progression.* As serum K^+ levels rise with renal failure, circulating levels of aldosterone also increase over what is required to maintain normal levels of blood volume.

Acid-Base Regulation

The kidneys excrete 1 meq/kg/day of noncarbonic H^+ ion on a normal diet. To do this, all of the filtered HCO_3^{2-} needs to be reabsorbed proximally so that H^+ pumps in the intercalated cells of the collecting duct can secrete H^+ ions that are subsequently trapped by urinary buffers, particularly phosphates and ammonia (see Chap. 277). While remaining nephrons increase their solute load with loss of renal mass, the ability to maintain total body H^+ excretion is often impaired by the gradual loss of H^+ pumps or with reductions in ammoniagenesis leading to development of a non-delta acidosis. Although hypertrophy of the proximal tubules initially increases their ability to reabsorb filtered HCO_3^{2-} and increases ammoniagenesis, with progressive loss of nephrons this compensation is eventually overwhelmed. In addition, with advancing renal failure, ammoniagenesis is further inhibited by elevation in levels of serum K^+, producing type IV renal tubular acidosis. Once the glomerular filtration rate falls below 25 mL/min, noncarbonic organic acids accumulate, producing a delta metabolic acidosis. Hyperkalemia can also inhibit tubular HCO_3^{2-} reabsorption, as can extracellular volume expansion and elevated levels of parathyroid hormone. Eventually, as the kidneys fail, the level of serum HCO_3^{2-} falls severely, reflecting the exhaustion of all body buffer systems, including bone.

Calcium and Phosphate

The kidney and gut play an important role in the regulation of serum levels of Ca^{2+} and PO_4^{2-}. With decreasing renal function and the appearance of tubulointerstitial nephritis, the expression of 1α-hydroxylase by the proximal tubule is reduced, lowering levels of calcitriol and Ca^{2+} absorption by the gut. Loss of nephron mass with progressive renal failure also gradually reduces the excretion of PO_4^{2-} and Ca^{2+}, and elevations in serum PO_4^{2-} further lower serum levels of Ca^{2+}, causing sustained secretion of parathyroid hormone. Unregulated increases in levels of parathyroid hormone cause Ca^{2+} mobilization from bone, Ca^{2+}/PO_4^{2-} precipitation in vascular tissues, abnormal bone remodeling, decreases in tubular bicarbonate reabsorption, and increases in renal PO_4^{2-} excretion. While elevated serum levels of parathyroid hormone initially maintain serum PO_4^{2-} near normal, with progressive nephron destruction, the capacity for renal PO_4^{2-} excretion is overwhelmed, the serum PO_4^{2-} elevates, and bone is progressively demineralized from secondary hyperparathyroidism. These adaptations evoke another classic functional trade-off **(Fig. 278-3)**.

Figure 278-3

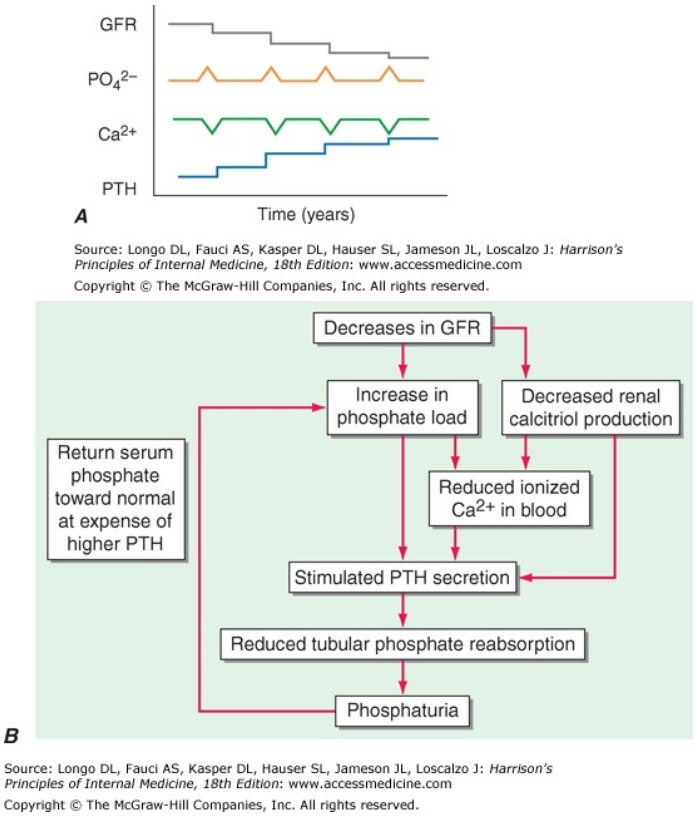

The "Trade-off Hypothesis" for Ca^{2+}/PO_4^{2-} homeostasis with progressively declining renal function. ***A.*** How adaptation to maintain Ca^{2+}/PO_4^{2-} homeostasis leads to increasing levels of parathyroid hormone *("classic" presentation from Slatopolsky et al.: Kidney Int 4:141, 1973).* ***B.*** current understanding of the underlying mechanisms for this Ca^{2+}/PO_4^{2-}

trade-off.

MODIFIERS INFLUENCING THE PROGRESSION OF RENAL DISEASE

Well-described risk factors for the progressive loss of renal function include systemic hypertension, diabetes, and activation of the renin-angiotensin-aldosterone system (Table 278-1). Poor glucose control will aggravate *renal progression* in both diabetic and nondiabetic renal disease. Angiotensin II produces *intraglomerular* hypertension and stimulates fibrogenesis. Aldosterone also serves as an independent fibrogenic mediator of progressive nephron loss apart from its role in modulating Na^+ and K^+ homeostasis. Genetic factors also play a role. There is recent, exciting evidence that risk alleles for *APOL1* underlies the increased susceptibility of African Americans to development of progressive kidney injury.

Table 278-1 Potential Modifiers of Renal Disease Progression

Hypertension	Hyperlipidemia
Renin-angiotensin system activation	Abnormal calcium/phosphorus homeostasis
Angiotensin II	Cigarette smoking
Aldosterone	Intrinsic paucity in nephron number
Diabetes	Prematurity/low birth weight
Obesity	Genetic predisposition
Excessive dietary protein	Genetic factors

Lifestyle choices also affect the progression of renal disease. Cigarette smoking either predisposes or accelerates the progression of nephron loss. Whether the effect of cigarettes is related to systemic hemodynamic alterations or specific damage to the renal microvasculature and/or tubules is unclear. Increases in fetuin-A, decreases in adiponectin, and increases in lipid oxidation associated with obesity also accelerate cardiovascular disease and progressive renal damage. Recent epidemiologic studies also confirm an association between high protein diets and progression of renal disease. Progressive nephron loss in experimental animals, and possibly in humans, is slowed by adherence to a low protein diet. Although a large multicenter trial, the Modification of Diet in Renal Disease, did not provide conclusive evidence that dietary protein restriction could retard progression to renal failure in humans, secondary analyses and a number of meta-analyses suggest a renoprotective effect from supervised low-protein diets in the range of 0.6–0.75 g/kg/day. Repair of chronic low serum bicarbonate levels during renal progression increases kidney survival. Abnormal Ca^{2+} and PO_4^{2-} metabolism in chronic kidney disease also plays a role in renal progression, and administration of calcitriol or its analogues can attenuate progression in a variety of models of chronic kidney disease.

An intrinsic paucity in the number of functioning nephrons predisposes to the development of renal disease. A reduced number of nephrons leads to permanent hypertension, either through direct renal damage or *hyperfiltration* producing glomerulosclerosis, or by primary induction of systemic hypertension that further exacerbates glomerular barotrauma. Younger individuals with hypertension who die suddenly as a result of trauma have 47% fewer glomeruli per kidney than age-matched controls.

A consequence of low birth weight is a relative deficit in the number of total nephrons, and low birth weight associates in adulthood with more hypertension and renal failure, among other abnormalities. In this regard, in addition or instead of a genetic predisposition to development of a specific disease or condition such as low birth weight, different epigenetic phenomena may produce varying clinical phenotypes from a single genotype depending upon maternal exposure to different environmental stimuli during gestation, a phenomena known as *developmental plasticity.* A specific clinical phenotype can also be selected in response to an adverse environmental exposure during critical periods of intrauterine development, also known as *fetal programming*. In the United States, there is at least a twofold increased incidence of low birth weight among blacks compared with whites, much but not all of which can be attributed to maternal age, health, or socioeconomic status.

As in other conditions producing nephron loss, the glomeruli of low-birth-weight individuals enlarge and associate with early *hyperfiltration* to maintain normal levels of renal function. With time, the resulting *intraglomerular* hypertension initiates a progressive decline in residual hyperfunctioning nephrons, ultimately accelerating renal failure. In African Americans, as well as other populations at increased risk for kidney failure, such as Pima Indians and Australian aborigines, large glomeruli are seen at early stages of kidney disease. An association between low birth weight and the development of albuminuria and nephropathy is reported for both diabetic and nondiabetic renal disease.

FURTHER READINGS

Brenner BM: Remission of renal disease: Recounting the challenge, acquiring the goal. J Clin Invest 110:1753, 2002[PMID: 12488422]

de Brito-Ashurst I et al: Bicarbonate supplementation slows progression of CKD and improves nutritional status. J Am Soc Nephrol 20: 2075, 2009

Genovese G et al: Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science 329:841, 2010[PMID: 20647424]

Harris RC, Neilson EG: Towards a unified theory of renal progression. Annu Rev Med 57:365, 2006[PMID: 16409155]

Ix JH, Sharma K: Mechanisms linking obesity, chronic kidney disease, and fatty liver disease: The roles of fetuin-A, adiponectin, and AMPK. J Am Soc Nephrol 21:406, 2010[PMID: 20150538]

Kao WH et al: MYH9 is associated with nondiabetic end-stage renal disease in African Americans. Nat Genet 40:1185, 2008[PMID: 18794854]

Luyckx VA, Brenner BM: The clinical importance of nephron mass. J Am Soc Nephrol 21:898, 2010[PMID: 20150537]

Peti-Peterdi J, Harris RC: Macula densa sensing and signaling mechanisms of renin release. J Am Soc Nephrol 21:1093, 2010[PMID: 20360309]

Slatopolsky E, Bricker NS: The role of phosphorus restriction in the prevention of secondary hyperparathyroidism in chronic

renal disease. Kidney Int 4:141, 1973[PMID: 4355426]

———— et al: Calcium, phosphorus and vitamin D disorders in uremia. Contr Nephrol 149:261, 2005

Zandi-Nejad K et al: Adult hypertension and kidney disease: The role of fetal programming. Hypertension 47:502, 2006[PMID: 16415374]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACUTE KIDNEY INJURY: INTRODUCTION

Acute kidney injury (AKI), previously known as acute renal failure, is characterized by the sudden impairment of kidney function resulting in the retention of nitrogenous and other waste products normally cleared by the kidneys. AKI is not a single disease but, rather, a designation for a heterogeneous group of conditions that share common diagnostic features: specifically, an increase in the blood urea nitrogen (BUN) concentration and/or an increase in the plasma or serum creatinine (SCr) concentration, often associated with a reduction in urine volume. AKI can range in severity from asymptomatic and transient changes in laboratory parameters of glomerular filtration rate (GFR), to overwhelming and rapidly fatal derangements in effective circulating volume regulation and electrolyte and acid-base composition of the plasma.

Changing the name of a syndrome as well known as "acute renal failure" does not occur frequently. We will summarize some of the reasons why the name was changed to "acute kidney injury." The term *failure* reflects only part of the spectrum of damage to the kidney that occurs clinically. In most cases of damage, the reduction in kidney function is modest. Nevertheless, this modest change has been documented to be associated with negative effects on outcome, albeit not nearly as ominous as seen with large decreases in kidney function associated with frank kidney failure that often requires acute dialysis therapies. Furthermore, the term *renal* is not well understood in the general population and this makes communication with patients and family more challenging; hence "kidney" has replaced "renal."

EPIDEMIOLOGY

AKI complicates 5–7% of acute care hospital admissions and up to 30% of admissions to the intensive care unit. AKI is also a major medical complication in the developing world, particularly in the setting of diarrheal illnesses, infectious diseases like malaria and leptospirosis, and natural disasters such as earthquakes. The incidence of AKI has grown by more than fourfold in the United States since 1988 and is estimated to have a yearly incidence of 500 per 100,000 population, higher than the yearly incidence of stroke. AKI is associated with a markedly increased risk of death in hospitalized individuals, particularly in those admitted to the ICU where in-hospital mortality rates may exceed 50%.

AKI IN THE DEVELOPING WORLD

The epidemiology of AKI differs tremendously between developed and developing countries, owing to differences in demographics, economics, geography, and comorbid disease burden. While certain features of AKI are common to both —particularly since urban centers of some developing countries increasingly resemble those in the developed world—many etiologies for AKI are region-specific such as envenomations from snakes, spiders, caterpillars, and bees; infectious causes such as malaria and leptospirosis; and crush injuries and resultant rhabdomyolysis from earthquakes.

ETIOLOGY AND PATHOPHYSIOLOGY

The causes of AKI have traditionally been divided into three broad categories: prerenal azotemia, intrinsic renal parenchymal disease, and postrenal obstruction (Fig. 279-1).

Figure 279-1

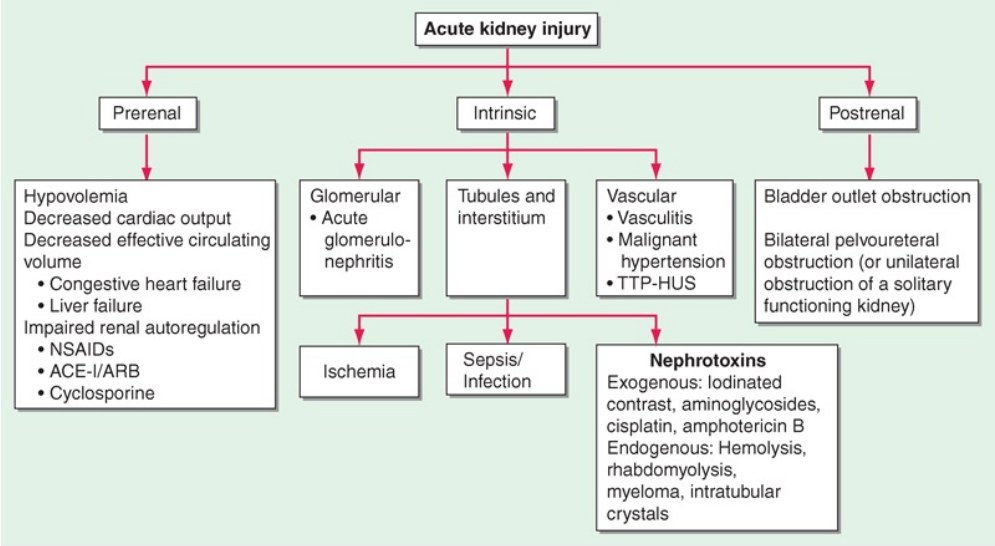

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Classification of the major causes of acute kidney injury. ACE-1, angiotensin-converting enzyme 1; ARB, angiotensin receptor blocker; NSAIDs, nonsteroidal anti-inflammatory drugs; TTP-HUS, thrombotic thrombocytopenic purpura-hemolytic uremic syndrome.

Prerenal Azotemia

Prerenal azotemia (from "azo," meaning nitrogen, and "-emia") is the most common form of AKI. It is the designation for a rise in SCr or BUN concentration due to inadequate renal plasma flow and intraglomerular hydrostatic pressure to support normal glomerular filtration. The most common clinical conditions associated with prerenal azotemia are hypovolemia, decreased cardiac output, and medications that interfere with renal autoregulatory responses such as nonsteroidal anti-inflammatory drugs

(NSAIDs) and inhibitors of angiotensin II (Fig. 279-2). Prerenal azotemia may coexist with other forms of intrinsic AKI. Prolonged periods of prerenal azotemia may lead to ischemic injury, often termed acute tubular necrosis, or ATN. By definition, prerenal azotemia involves no parenchymal damage to the kidney and is rapidly reversible once intraglomerular hemodynamics are restored.

Figure 279-2

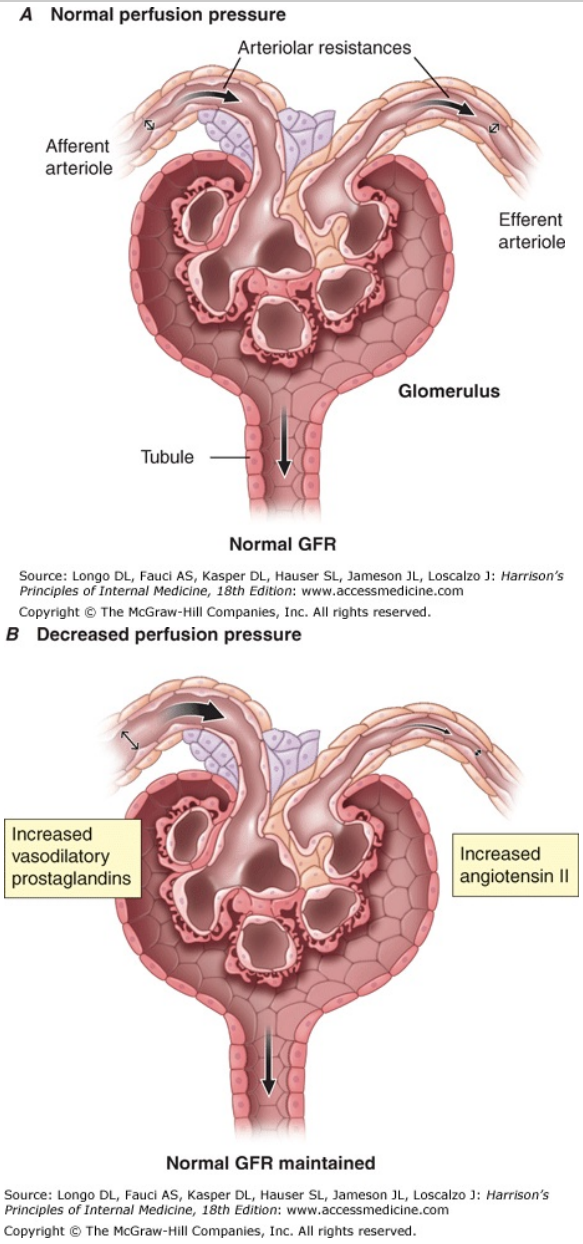

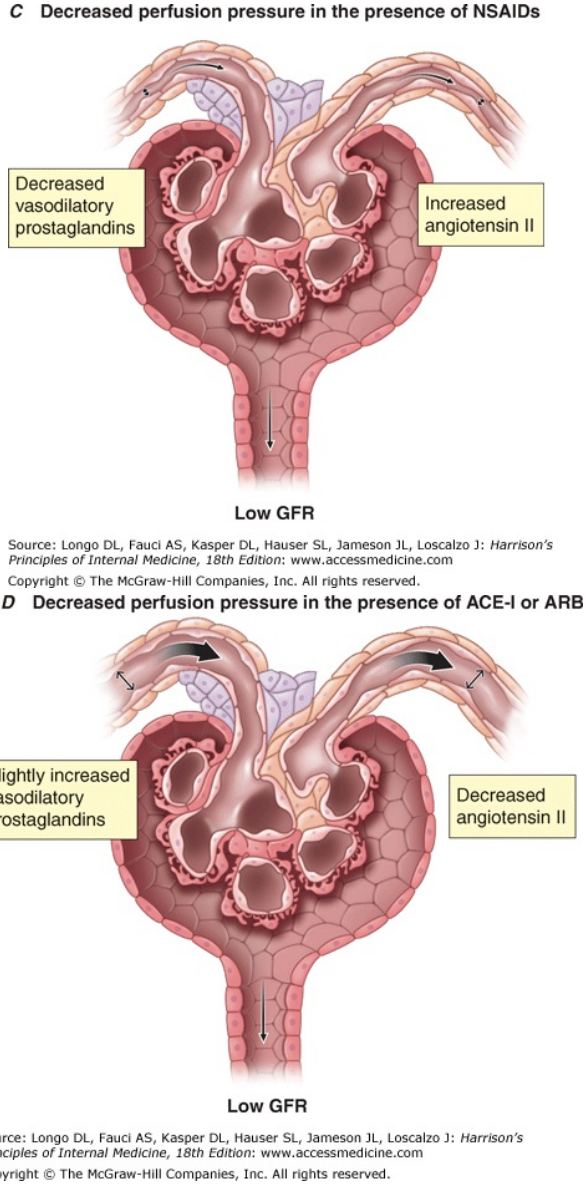

Intrarenal Mechanisms for Autoregulation of the Glomerular Filtration Rate (GFR) under Decreased Perfusion Pressure and Reduction of the GFR by Drugs. Panel A shows normal conditions and a normal GFR. Panel B shows reduced perfusion pressure within the autoregulatory range. Normal glomerular capillary pressure is maintained by afferent vasodilatation and efferent vasoconstriction. Panel C shows reduced perfusion pressure with a nonsteroidal anti-inflammatory drug (NSAID). Loss of vasodilatory prostaglandins increases afferent resistance; this causes the glomerular capillary pressure to drop below normal values and the GFR to decrease. Panel D shows reduced perfusion pressure with an angiotensin-converting enzyme (ACE-I) inhibitor or an angiotensin receptor blocker (ARB). Loss of angiotensin II action reduces efferent resistance; this causes the glomerular capillary pressure to drop below normal values and the GFR to decrease. *(From N Engl J Med 2007;357:797-805; with permission.*)

Normal GFR is maintained in part by the relative resistances of the afferent and efferent renal arterioles, which determine the glomerular plasma flow and the transcapillary hydraulic pressure gradient that drive glomerular ultrafiltration. Mild degrees of hypovolemia and reductions in cardiac output elicit compensatory renal physiologic changes. Because renal blood flow accounts for 20% of the cardiac output, renal vasoconstriction and salt and water reabsorption occur as a homeostatic response to decreased effective circulating volume or cardiac output in order to maintain blood pressure and increase intravascular volume to sustain perfusion to the cerebral and coronary vessels. Mediators of this response include angiotensin II, norepinephrine, and vasopressin (also termed antidiuretic hormone). Glomerular filtration can be maintained despite reduced renal blood flow by angiotensin II–mediated renal efferent vasoconstriction, which maintains glomerular capillary hydrostatic pressure closer to normal and thereby prevents marked reductions in GFR if renal blood flow reduction is not excessive.

In addition, a myogenic reflex within the afferent arteriole leads to dilation in the setting of low perfusion pressure, thereby maintaining glomerular perfusion. Intrarenal biosynthesis of vasodilator prostaglandins (prostacyclin, prostaglandin E2), kallikrein and kinins, and possibly nitric oxide (NO) also increase in response to low renal perfusion pressure. Autoregulation is also accomplished by tubuloglomerular feedback, in which decreases in solute delivery to the macula densa (specialized cells within the proximal tubule) elicit dilation of the juxtaposed afferent arteriole in order to maintain glomerular perfusion, a mechanism mediated, in part, by NO. There is a limit, however, to the ability of these counterregulatory mechanisms to maintain GFR in the face of systemic hypotension. Even in healthy adults, renal autoregulation usually fails once the systolic blood pressure falls below 80 mmHg.

A number of factors determine the robustness of the autoregulatory response and, thereby, the risk of prerenal azotemia. Atherosclerosis, long-standing hypertension, and older age can lead to hyalinosis and myointimal hyperplasia, causing structural narrowing of the intrarenal arterioles and impaired capacity for renal afferent vasodilation. In chronic kidney disease, renal afferent vasodilation may be operating at maximal capacity in order to maximize GFR in response to reduced functional renal mass. Drugs can affect the compensatory changes evoked to maintain GFR. NSAIDs inhibit renal prostaglandin production, limiting renal afferent vasodilation. ACE inhibitors and angiotensin receptor blockers (ARBs) limit renal efferent vasoconstriction; this effect is particularly pronounced in patients with bilateral renal artery stenosis or unilateral renal artery stenosis (in the case

of a solitary functioning kidney) because renal efferent vasoconstriction is needed to maintain GFR due to low renal perfusion. The combined use of nonsteroidal anti-inflammatory agents with ACE inhibitors or ARBs poses a particularly high risk for developing prerenal azotemia.

Many individuals with advanced cirrhosis exhibit a unique hemodynamic profile that resembles prerenal azotemia despite total body volume overload. Systemic vascular resistance is markedly reduced due to primary arterial vasodilation in the splanchnic circulation, resulting ultimately in activation of vasoconstrictor responses similar to those seen in hypovolemia. AKI is a common complication in this setting, and it can be triggered by volume depletion and spontaneous bacterial peritonitis. A particularly poor prognosis is seen in the case of type 1 hepatorenal syndrome, in which AKI without an alternate cause (e.g., infection, shock, nephrotoxic drugs) persists despite volume administration and withholding of diuretics. Type 2 hepatorenal syndrome is a less severe form characterized mainly by refractory ascites.

Intrinsic Aki

The most common causes of intrinsic AKI are sepsis, ischemia, and nephrotoxins, both endogenous and exogenous (Fig. 279-3). In many cases, prerenal azotemia advances to tubular injury. Although classically termed "acute tubular necrosis," human biopsy confirmation of tubular necrosis is, in general, lacking in cases of sepsis and ischemia; indeed, processes such as inflammation, apoptosis, and altered regional perfusion may be more relevant pathophysiologically. Other causes of intrinsic AKI are less common and can be conceptualized anatomically according to the major site of renal parenchymal damage: glomeruli, tubulointerstitium, and vessels.

Figure 279-3

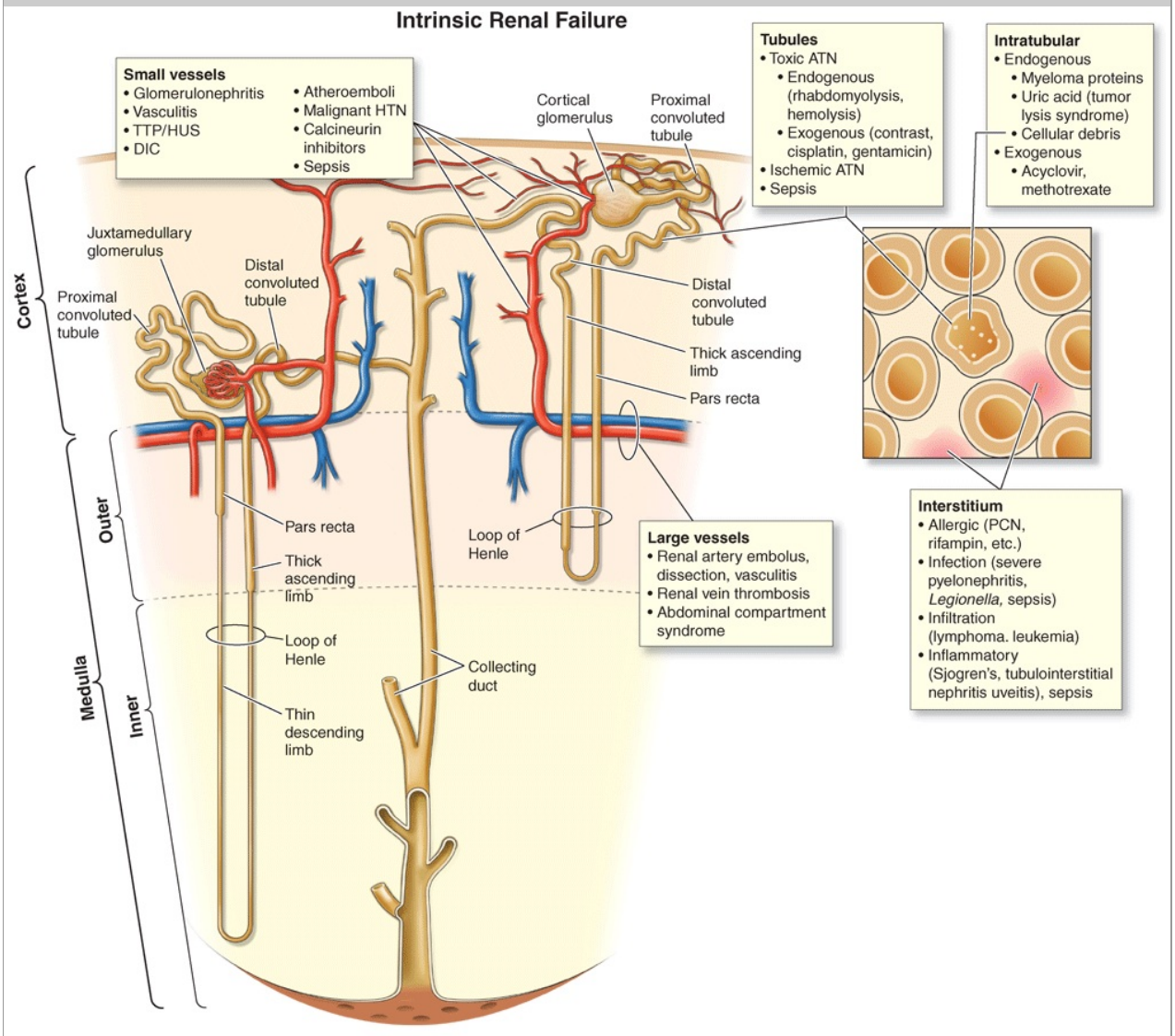

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Major causes of intrinsic acute kidney injury. ATN, acute tubular necrosis; DIC, disseminated intravascular coagulation; HTN, hypertensive nephropathy; MTX, methotrexate; PCN, penicillin; TTP/HUS, thrombotic thrombocytopenic purpura/hemolytic uremic syndrome; TINU, tubulointerstitial nephritis-uveitis.

SEPSIS-ASSOCIATED AKI

In the United States, more than 700,000 cases of sepsis occur each year. AKI complicates more than 50% of cases of severe sepsis, and greatly increases the risk of death. Sepsis is also a very important cause of AKI in the developing world. Decreases in GFR with sepsis can occur even in the absence of overt hypotension, although most cases of severe AKI typically occur in the setting of hemodynamic collapse requiring vasopressor support. While there is clearly tubular injury associated with AKI in sepsis as manifest by the presence of tubular debris and casts in the urine, postmortem examinations of kidneys from individuals with severe sepsis suggest that other factors, perhaps related to inflammation and interstitial edema, must be considered in the pathophysiology of sepsis-induced AKI.

The hemodynamic effects of sepsis—arising from generalized arterial vasodilation, mediated in part by cytokines that upregulate the expression of inducible NO synthase in the vasculature—can lead to a reduction in GFR. The operative mechanisms may be excessive efferent arteriole vasodilation, particularly early in the course of sepsis, or renal vasoconstriction from activation of the sympathetic nervous system, the renin-angiotensin-aldosterone system, vasopressin, and endothelin. Sepsis may lead to endothelial damage, which results in microvascular thrombosis, activation of reactive oxygen species, and leukocyte adhesion and migration, all of which may injure renal tubular cells.

Ischemia-Associated Aki

Healthy kidneys receive 20% of the cardiac output and account for 10% of resting oxygen consumption, despite constituting only 0.5% of the human body mass. The kidneys are also the site of one of the most hypoxic regions in the body, the renal medulla. The outer medulla is particularly vulnerable to ischemic damage because of the architecture of the blood vessels that supply oxygen and nutrients to the tubules. Enhanced leukocyte-endothelial interactions in the small vessels lead to inflammation and reduced local blood flow to the metabolically very active S3 segment of the proximal tubule, which depends on oxidative metabolism for survival. Ischemia alone in a normal kidney is usually not sufficient to cause severe AKI, as evidenced by the relatively low risk of severe AKI even after total interruption of renal blood flow during suprarenal aortic clamping or cardiac arrest. Clinically, AKI more commonly develops when ischemia occurs in the context of limited renal reserve (e.g., chronic kidney disease or older age) or coexisting insults such as sepsis, vasoactive or nephrotoxic drugs, rhabdomyolysis, and the systemic inflammatory states associated with burns and pancreatitis. Prerenal azotemia and ischemia-associated AKI represent a continuum of the manifestations of renal hypoperfusion. Persistent preglomerular vasoconstriction may be a common underlying cause of the reduction in GFR seen in AKI; implicated factors for vasoconstriction include activation of tubuloglomerular feedback from enhanced delivery of solute to the macula densa following proximal tubule injury, increased basal vascular tone and reactivity to vasoconstrictive agents, and decreased vasodilator responsiveness. Other contributors to low GFR include backleak of filtrate across ischemic and denuded tubular epithelium and mechanical obstruction of tubules from necrotic debris (Fig. 279-4).

Figure 279-4

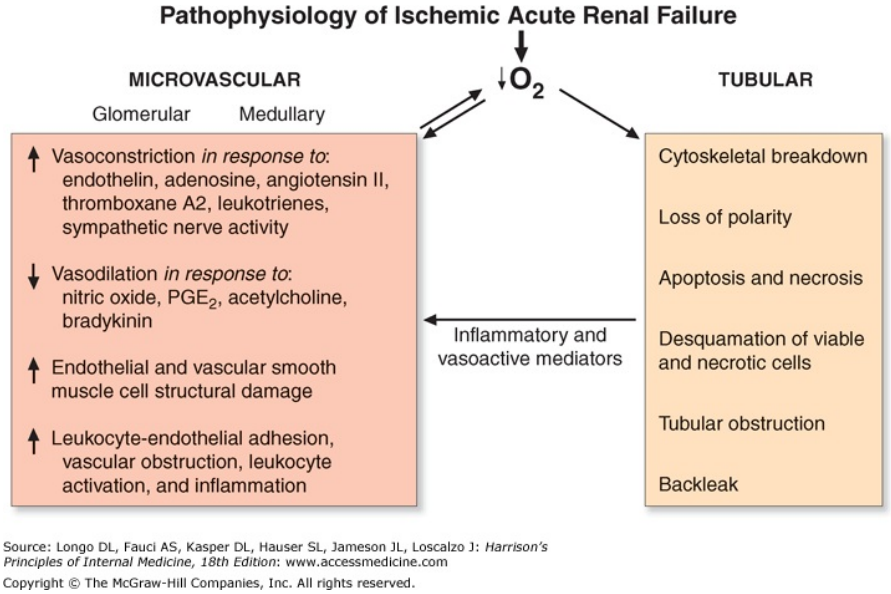

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Interacting microvascular and tubular events contributing to the pathophysiology of ischemic acute kidney injury. *PGE2, prostaglandin E2. (From J Am Soc Nephrol 14:2199, 2003.)*

POSTOPERATIVE AKI

Ischemia-associated AKI is a serious complication in the postoperative period, especially after major operations involving significant blood loss and intraoperative hypotension. The procedures most commonly associated with AKI are cardiac surgery with cardiopulmonary bypass (particularly for combined valve and bypass procedures), vascular procedures with aortic cross clamping, and intraperitoneal procedures. Severe AKI requiring dialysis occurs in approximately 1% of cardiac and vascular surgery procedures. The risk of severe AKI has been less well studied for major intra peritoneal procedures, but appears to be of comparable magnitude. Common risk factors for postoperative AKI include underlying chronic kidney disease, older age, diabetes mellitus, congestive heart failure, and emergency procedures. The pathophysiology of AKI following cardiac surgery is multifactorial. Major AKI risk factors are common in the population undergoing cardiac surgery. The use of nephrotoxic agents including iodinated contrast for cardiac imaging prior to surgery may increase the risk of AKI. Cardiopulmonary bypass is a unique hemodynamic state characterized by nonpulsatile flow and exposure of the circulation to extracorporeal circuits. Longer duration of cardiopulmonary bypass is a risk factor for AKI. In addition to ischemic injury from sustained hypoperfusion, cardiopulmonary bypass may cause AKI through a number of mechanisms including extracorporeal circuit activation of leukocytes and inflammatory processes, hemolysis with resultant pigment nephropathy (see below), and aortic injury with resultant atheroemboli. AKI from atheroembolic disease, which can also occur following percutaneous catheterization of the aorta, or spontaneously, is due to cholesterol crystal embolization resulting in partial or total occlusion of multiple small arteries within the kidney. Over time, a foreign body reaction can result in intimal proliferation, giant cell formation, and further narrowing of the vascular lumen, accounting for the generally subacute (over a period of weeks rather than days) decline in renal function.

BURNS AND ACUTE PANCREATITIS

Extensive fluid losses into the extravascular compartments of the body frequently accompany severe burns and acute pancreatitis. AKI is an ominous complication of burns, affecting 25% of individuals with more than 10% total body surface area involvement. In addition to severe hypovolemia resulting in decreased cardiac output and increased neurohormonal activation, burns and acute pancreatitis both lead to dysregulated inflammation and an increased risk of sepsis and acute lung injury, all of which may facilitate the development and progression of AKI. Individuals undergoing massive fluid resuscitation for trauma, burns, and acute pancreatitis can also develop the abdominal compartment syndrome, where markedly elevated intraabdominal pressures, usually higher than 20 mmHg, lead to renal vein compression and reduced GFR.

DISEASES OF THE MICROVASCULATURE LEADING TO ISCHEMIA

Microvascular causes of AKI include the thrombotic microangiopathies [antiphospholipid antibody syndrome, radiation nephritis, malignant nephrosclerosis, and thrombotic thrombocytopenic purpura/hemolytic uremic syndrome (TTP-HUS)], scleroderma, and atheroembolic disease. Large vessel diseases associated with AKI include renal artery dissection, thromboembolism, thrombosis, and renal vein compression or thrombosis.

Nephrotoxin-Associated Aki

The kidney has very high susceptibility to nephrotoxicity due to extremely high blood perfusion and concentration of circulating substances along the nephron where water is reabsorbed and in the medullary interstitium; this results in high-concentration exposure of toxins to tubular, interstitial, and endothelial cells. Nephrotoxic injury occurs in response to a number of pharmacologic compounds with diverse structures, endogenous substances, and environmental exposures. All structures of the kidney are vulnerable to toxic injury, including the tubules, interstitium, vasculature, and collecting system. As with other forms of AKI, risk factors for nephrotoxicity include older age, chronic kidney disease (CKD), and prerenal azotemia. Hypoalbuminemia may increase the risk of some forms of nephrotoxin-associated AKI due to increased free circulating drug concentrations.

CONTRAST AGENTS

Iodinated contrast agents used for cardiovascular and CT imaging are a leading cause of AKI. The risk of AKI, or "contrast nephropathy," is negligible in those with normal renal function but increases markedly in the setting of chronic kidney disease, particularly diabetic nephropathy. The most common clinical course of contrast nephropathy is characterized by a rise in SCr beginning 24–48 hours following exposure, peaking within 3–5 days, and resolving within 1 week. More severe, dialysis-requiring AKI is uncommon except in the setting of significant preexisting chronic kidney disease, often in association with congestive heart failure or other coexisting causes for ischemia-associated AKI. Patients with multiple myeloma and renal disease are particularly susceptible. Low fractional excretion of sodium and relatively benign urinary sediment without features of tubular necrosis (see below) are common findings. Contrast nephropathy is thought to occur from a combination of factors, including (1) hypoxia in the renal outer medulla due to perturbations in renal microcirculation and occlusion of small vessels; (2) cytotoxic damage to the tubules directly or via the generation of oxygen free radicals, especially since the concentration of the agent within the tubule is markedly increased; and (3) transient tubule obstruction with precipitated contrast material. Other diagnostic agents implicated as a cause of AKI are high-dose gadolinium used for MRI and oral sodium phosphate solutions used as bowel purgatives.

ANTIBIOTICS

Several antimicrobial agents are commonly associated with AKI. *Aminoglycosides andamphotericin B* both cause tubular necrosis. Nonoliguric AKI (i.e., without a significant reduction in urine volume) accompanies 10–30% of courses of aminoglycoside antibiotics, even when plasma levels are in the therapeutic range. Aminoglycosides are freely filtered across the glomerulus and then accumulate within the renal cortex, where concentrations can greatly exceed those of the plasma. AKI typically manifests after 5–7 days of therapy and can present even after the drug has been discontinued. Hypomagnesemia is a common finding.

Amphotericin B causes renal vasoconstriction from an increase in tubuloglomerular feedback as well as direct tubular toxicity mediated by reactive oxygen species. Nephrotoxicity from amphotericin B is dose and duration dependent. This drug binds to tubular membrane cholesterol and introduces pores. Clinical features of amphotericin B nephrotoxicity include polyuria, hypomagnesemia, hypocalcemia, and nongap metabolic acidosis.

Vancomycin may be associated with AKI, particularly when trough levels are high, but a causal relationship with AKI has not been definitively established. *Acyclovir* can precipitate in tubules and cause AKI by tubular obstruction, particularly when given as an intravenous bolus at high doses (500 mg/m^2) or in the setting of hypovolemia. *Foscarnet*, *pentamidine*, and *cidofovir* (less commonly prescribed antimicrobials) are also frequently associated with AKI due to tubular toxicity. AKI secondary to acute interstitial nephritis can occur as a consequence of a number of antibiotics, including *penicillins*, *cephalosporins*, *quinolones*, *sulfonamides*, and *rifampin*.

CHEMOTHERAPEUTIC AGENTS

Cisplatin and *carboplatin* are accumulated by proximal tubular cells and cause necrosis and apoptosis. Intensive hydration regimens have reduced the incidence of cisplatin nephrotoxicity, but it remains a dose-limiting toxicity. *Ifosfamide* may cause hemorrhagic cystitis and tubular toxicity, manifested as Type II renal tubular acidosis (Fanconi's syndrome), polyuria, hypokalemia, and a modest decline in GFR. Antiangiogenesis agents such as *bevacizumab*, can cause proteinuria and hypertension via injury to the glomerular microvasculature (thrombotic microangiopathy). Other antineoplastic agents such as mitomycin C and gemcitabine may cause thrombotic microangiopathy with resultant AKI.

TOXIC INGESTIONS

Ethylene glycol, present in automobile antifreeze, is metabolized to oxalic acid, glycolaldehyde, and glyoxylate, which may cause AKI through direct tubular injury. Diethylene glycol is an industrial agent that has been the cause of outbreaks of severe AKI around the world due to adulteration of pharmaceutical preparations. The metabolite 2-hydroxyethoxyacetic acid (HEAA) is thought to be responsible for tubular injury. Melamine contamination of foodstuffs has led to nephrolithiasis and AKI, either through intratubular obstruction or possibly direct tubular toxicity. Aristolochic acid was found to be the cause of "Chinese herb nephropathy" and "Balkan nephropathy" due to contamination of medicinal herbs or farming. The list of environmental toxins is likely to grow and contribute to a better understanding of previously catalogued "idiopathic" chronic tubular interstitial disease, a common diagnosis in both the developed and developing world.

ENDOGENOUS TOXINS

AKI may be caused by a number of endogenous compounds, including myoglobin, hemoglobin, uric acid, and myeloma light chains. Myoglobin can be released by injured muscle cells, and hemoglobin can be released during massive hemolysis leading to pigment nephropathy. Rhabdomyolysis may result from traumatic crush injuries, muscle ischemia during vascular or orthopedic surgery, compression during coma or immobilization, prolonged seizure activity, excessive exercise, heat stroke or malignant hyperthermia, infections, metabolic disorders (e.g., hypophosphatemia, severe hypothyroidism), and myopathies (drug-induced, metabolic, or inflammatory). Pathogenic factors for AKI include intrarenal vasoconstriction, direct proximal tubular toxicity, and mechanical obstruction of the distal nephron lumen when myoglobin or hemoglobin precipitates with Tamm-Horsfall protein (uromodulin, the most common protein in urine and produced in the thick ascending limb of the loop of Henle), a process favored by acidic urine. Tumor lysis syndrome may follow initiation of cytotoxic therapy in patients with high-grade lymphomas and acute lymphoblastic leukemia; massive release of uric acid (with serum levels often exceeding 15 mg/dL) leads to precipitation of uric acid in the renal tubules and AKI (Chap. 276). Other features of tumor lysis syndrome include hyperkalemia and hyperphosphatemia. The tumor lysis syndrome can also occasionally occur spontaneously or with treatment for solid tumors or multiple myeloma. Myeloma light chains can also cause AKI by direct tubular toxicity and by binding to Tamm-Horsfall protein to form obstructing intratubular casts. Hypercalcemia, which can also be seen in multiple myeloma, may cause AKI by intense renal vasoconstriction and volume depletion.

ALLERGIC ACUTE TUBULOINTERSTITIAL DISEASE AND OTHER CAUSES OF INTRINSIC AKI

While many of the ischemic and toxic causes of AKI previously described result in tubulointerstitial disease, many drugs are also associated with the development of an allergic response characterized by an inflammatory infiltrate and often peripheral and urinary eosinophilia. AKI may be caused by severe infections and infiltrative diseases. Diseases of the glomeruli or vasculature can lead to AKI by compromising blood flow within the renal circulation. Glomerulonephritis or vasculitis are relatively uncommon but potentially severe causes of AKI that may necessitate timely treatment with immunosuppressive agents or therapeutic plasma exchange.

Postrenal Acute Kidney Injury

(See also Chap. 289) Postrenal AKI occurs when the normally unidirectional flow of urine is acutely blocked either partially or totally, leading to increased retrograde hydrostatic pressure and interference with glomerular filtration. Obstruction to urinary flow may be caused by functional or structural derangements anywhere from the renal pelvis to the tip of the urethra (Fig. 279-5). Normal urinary flow rate does not rule out the presence of partial obstruction, since the GFR is normally two orders of magnitude higher than the urinary flow rate. For AKI to occur in healthy individuals, obstruction must affect both kidneys unless only one kidney is functional, in which case unilateral obstruction can cause AKI. Unilateral obstruction may cause AKI in the setting of significant underlying CKD or in rare cases from reflex vasospasm of the contralateral kidney. Bladder neck obstruction is a common cause of postrenal AKI and can be due to prostate disease (benign prostatic hypertrophy or prostate cancer), neurogenic bladder, or therapy with anticholinergic drugs. Obstructed Foley catheters can cause postrenal AKI if not recognized and relieved. Other causes of lower tract obstruction are blood clots, calculi, and urethral strictures. Ureteric obstruction can occur from intraluminal obstruction (e.g., calculi, blood clots, sloughed renal papillae), infiltration of the ureteric wall (e.g., neoplasia), or external compression (e.g., retroperitoneal fibrosis, neoplasia, abscess, or inadvertent surgical damage). The pathophysiology of postrenal AKI involves hemodynamic alterations triggered by an abrupt increase in intratubular pressures. An initial period of hyperemia from afferent arteriolar dilation is followed by intrarenal vasoconstriction from the generation of angiotensin II, thromboxane A2, and vasopressin, and a reduction in NO production. Reduced GFR is due to underperfusion of glomeruli and, possibly, changes in the glomerular ultrafiltration coefficient.

Figure 279-5

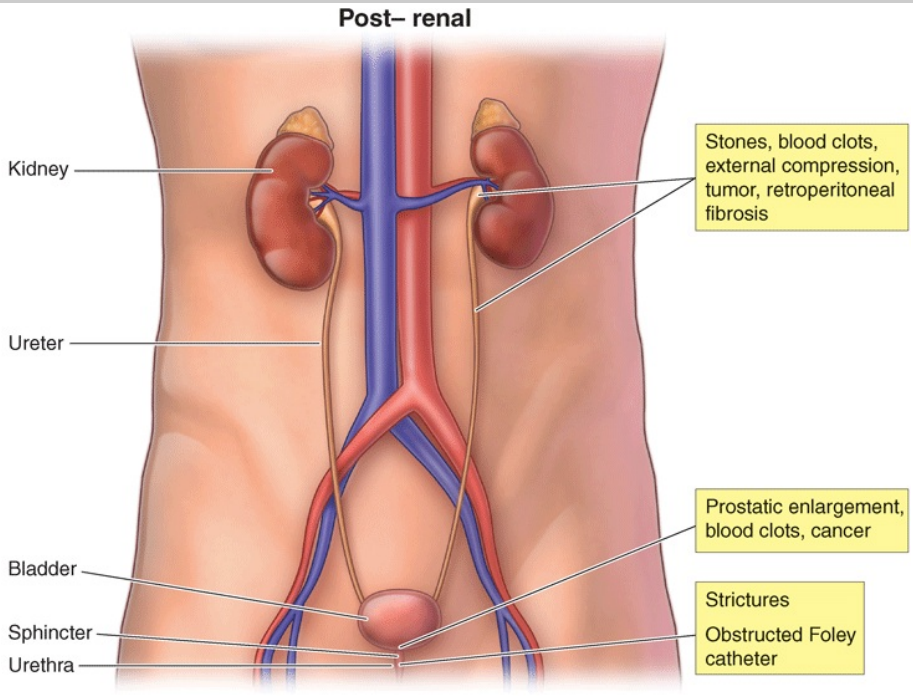

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anatomic sites and causes of obstruction leading to postrenal acute kidney injury.

DIAGNOSTIC EVALUATION (TABLE 279-1)

The presence of AKI is usually inferred by an elevation in the SCr concentration. AKI is currently defined by a rise of at least 0.3 mg/dL or 50% higher than baseline within a 24–48-hours period or a reduction in urine output to 0.5 mL/kg per hour for longer than 6 hours. It is important to recognize that given this definition, some patients with AKI will not have tubular or glomerular damage (e.g., prerenal azotemia). The distinction between AKI and chronic kidney disease is important for proper diagnosis and treatment. The distinction is straightforward when a recent baseline SCr concentration is available, but more difficult in the many instances in which the baseline is unknown. In such cases, clues suggestive of chronic kidney disease can come from radiologic studies (e.g., small, shrunken kidneys with cortical thinning on renal ultrasound, or evidence of renal osteodystrophy) or laboratory tests such as normocytic anemia or secondary hyperparathyroidism with hyperphosphatemia and hypocalcemia, consistent with CKD. No set of tests, however, can rule out AKI superimposed on CKD since AKI is a frequent complication in patients with CKD, further complicating the distinction. Serial blood tests showing continued substantial rise of SCr is clear evidence of AKI. Once the diagnosis of AKI is established, its cause needs to be determined.

Table 279-1 Major Causes, Clinical Features, and Diagnostic Studies for Acute Kidney Injury

Etiology	Clinical Features	Laboratory Features	Comments
Prerenal azotemia	History of poor fluid intake or fluid loss (hemorrhage, diarrhea, vomiting, sequestration into extravascular space);	BUN/creatinine ratio above 20, FeNa <1%, hyaline casts in urine	Low FeNa, high specific gravity and osmolality may not be seen in the setting of

	NSAID/ACE-I/ARB; heart failure; evidence of volume depletion (tachycardia, absolute or postural hypotension, low jugular venous pressure, dry mucous membranes), decreased effective circulatory volume (cirrhosis, heart failure)	sediment, urine specific gravity >1.018, urine osmolality >500 mOsm/kg	CKD, diuretic use; BUN elevation out of proportion to creatinine may alternatively indicate upper GI bleed or increased catabolism. Response to restoration of hemodynamics is most diagnostic.
Sepsis-associated AKI	Sepsis, sepsis syndrome, or septic shock. Overt hypotension not always seen in mild to moderate AKI	Positive culture from normally sterile body fluid; urine sediment often contains granular casts, renal tubular epithelial cell casts	FeNa may be low (<1%), particularly early in the course, but is usually >1% and osmolality <500 mOsm/kg
Ischemia-associated AKI	Systemic hypotension, often superimposed upon sepsis and/or reasons for limited renal reserve such as older age, CKD	Urine sediment often contains granular casts, renal tubular epithelial cell casts. FeNa typically >1%.	
Nephrotoxin-Associated AKI: Endogenous			
Rhabdomyolysis	Traumatic crush injuries, seizures, immobilization	Elevated myoglobin, creatine kinase; urinalysis heme positive with few red blood cells	FeNa may be low (<1%)
Hemolysis	Recent blood transfusion with transfusion reaction	Anemia, elevated LDH, low haptoglobin	FeNa may be low (<1%); evaluation for transfusion reaction
Tumor lysis	Recent chemotherapy	Hyperphosphatemia, hypocalcemia, hyperuricemia	
Multiple myeloma	Age >60 years, constitutional symptoms, bone pain	Monoclonal spike in urine or serum electrophoresis; low anion gap	Bone marrow or renal biopsy can be diagnostic
Contrast nephropathy	Exposure to iodinated contrast	Characteristic course is rise in SCr within 1–2 d, peak within 3–5 d, recovery within 7 d	FeNa may be low (<1%)
Nephrotoxin-Associated AKI: Exogenous			
Tubular injury	Aminoglycoside antibiotics, cisplatin, tenofovir, zoledronate	Urine sediment often contains granular casts, renal tubular epithelial cell casts. FeNa typically >1%.	
Interstitial nephritis	Recent medication exposure; can have fever, rash arthralgias	Eosinophilia, sterile pyuria; often nonoliguric	Urine eosinophils have limited diagnostic accuracy; systemic signs of drug reaction often absent; kidney biopsy may be helpful
Other Causes of Intrinsic AKI			
Glomerulonephritis/vasculitis	Variable **(Chap. 283)** features include skin rash, arthralgias, sinusitis (AGBM disease), lung hemorrhage (AGBM, ANCA, lupus), recent skin infection or pharyngitis (poststreptococcal)	ANA, ANCA, AGBM antibody, hepatitis serologies, cryoglobulins, blood culture, decreased complement levels, ASO titer (abnormalities of these tests depending on etiology)	Kidney biopsy may be necessary
Interstitial nephritis	Nondrug-related causes include tubulointerstitial nephritis-uveitis (TINU) syndrome, *Legionella* infection	Eosinophilia, sterile pyuria; often nonoliguric	Urine eosinophils have limited diagnostic accuracy; kidney biopsy may be necessary
TTP/HUS	Recent GI infection or use of calcineurin inhibitors	Schistocytes on peripheral blood smear, elevated LDH, anemia, thrombocytopenia	Kidney biopsy may be necessary
Atheroembolic disease	Recent manipulation of the aorta or other large vessels; may occur spontaneously or after anticoagulation; retinal plaques, palpable purpura, livedo reticularis, GI bleed	Hypocomplementemia, eosinophiluria (variable), variable amounts of proteinuria	Skin or kidney biopsy can be diagnostic
Postrenal AKI	History of kidney stones, prostate disease, obstructed bladder catheter, retroperitoneal or pelvic neoplasm	No specific findings other than AKI; may have pyuria or hematuria	Imaging with computed tomography or ultrasound

Abbreviations: ACE-1, angiotensin-converting enzyme-1; AGBM, antiglomerular basement membrane; AKI, acute kidney injury; ANA, antinuclear antibody; ANCA, antineutrophilic cytoplasmic antibody; ARB, angiotensin receptor blocker; ASO, antistreptolysin O; BUN, blood urea nitrogen; CKD, chronic kidney disease; FeNa, fractional excretion of sodium; GI, gastrointestinal; LDH, lactate dehydrogenase; NSAID, nonsteroidal anti-inflammatory drug; TTP/HUS, thrombotic

thrombocytopenic purpura/hemolytic uremic syndrome.

History and Physical Examination

The clinical context, careful history taking, and physical examination often narrow the differential diagnosis for the cause of AKI. Prerenal azotemia should be suspected in the setting of vomiting, diarrhea, glycosuria causing polyuria, and several medications including diuretics, NSAIDs, ACE inhibitors, and ARBs. Physical signs of orthostatic hypotension, tachycardia, reduced jugular venous pressure, decreased skin turgor, and dry mucous membranes are often present in prerenal azotemia. A history of prostatic disease, nephrolithiasis, or pelvic or paraaortic malignancy would suggest the possibility of postrenal AKI. Whether or not symptoms are present early during obstruction of the urinary tract depends on the location of obstruction. Colicky flank pain radiating to the groin suggests acute ureteric obstruction. Nocturia and urinary frequency or hesitancy can be seen in prostatic disease. Abdominal fullness and suprapubic pain can accompany massive bladder enlargement. Definitive diagnosis of obstruction requires radiologic investigations.

A careful review of all medications is imperative in the evaluation of an individual with AKI. Not only are medications frequently a cause of AKI but doses of administered medications must be adjusted for estimated GFR. Idiosyncratic reactions to a wide variety of medications can lead to allergic interstitial nephritis, which may be accompanied by fever, arthralgias, and a pruritic erythematous rash. The absence of systemic features of hypersensitivity, however, does not exclude the diagnosis of interstitial nephritis.

AKI accompanied by palpable purpura, pulmonary hemorrhage, or sinusitis raises the possibility of systemic vasculitis with glomerulonephritis. Atheroembolic disease can be associated with livedo reticularis and other signs of emboli to the legs. A tense abdomen should prompt consideration of acute abdominal compartment syndrome, which requires measurement of bladder pressure. Signs of limb ischemia may be clues to the diagnosis of rhabdomyolysis.

Urine Findings

Complete anuria early in the course of AKI is uncommon except in the following situations: complete urinary tract obstruction, renal artery occlusion, overwhelming septic shock, severe ischemia (often with cortical necrosis), or severe proliferative glomerulonephritis or vasculitis. A reduction in urine output (oliguria, defined as <400 mL/24 h) usually denotes more significant AKI (i.e., lower GFR) than when urine output is preserved. Oliguria is associated with worse clinical outcomes. Preserved urine output can be seen in nephrogenic diabetes insipidus characteristic of longstanding urinary tract obstruction, tubulointerstitial disease, or nephrotoxicity from cisplatin or aminoglycosides, among other causes. Red or brown urine may be seen with or without gross hematuria; if the color persists in the supernatant after centrifugation, then pigment nephropathy from rhabdomyolysis or hemolysis should be suspected.

The urinalysis and urine sediment examination are invaluable tools, but they require clinical correlation because of generally limited sensitivity and specificity (Fig. 279-6) (Chap. e14). In the absence of preexisting proteinuria from CKD, AKI from ischemia or nephrotoxins leads to mild proteinuria (<1 g/d). Greater proteinuria in AKI suggests damage to the glomerular ultrafiltration barrier or excretion of myeloma light chains; the latter are not detected with conventional urine dipsticks (which detect albumin) and require the sulfosalicylic acid test or immunoelectrophoresis. Atheroemboli can cause a variable degree of proteinuria. Extremely heavy proteinuria ("nephrotic range," >3.5 g/d) can occasionally be seen in glomerulonephritis, vasculitis, or interstitial nephritis (particularly from NSAIDs). AKI can also complicate cases of minimal change disease, a cause of the nephrotic syndrome (Chap. 277). If the dipstick is positive for hemoglobin but few red blood cells are evident in the urine sediment, then rhabdomyolysis or hemolysis should be suspected.

Figure 279-6

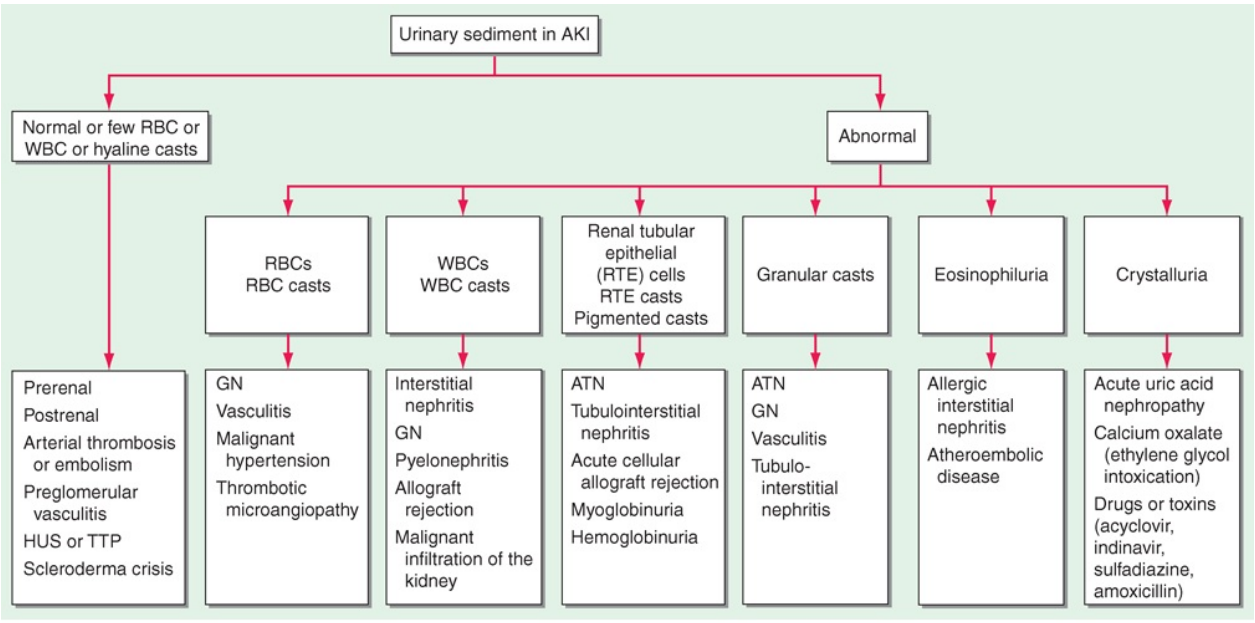

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Interpretation of urinary sediment findings in acute kidney injury. GN, glomerulonephritis; RTE, renal tubular epithelial. [Adapted from L Yang, JV Bonventre: Diagnosis and clinical evaluation of acute kidney injury. In Comprehensive Nephrology, 4th ed. J Floege et al (eds). Philadelphia, Elsevier, 2010.]

Prerenal azotemia may present with hyaline casts or an unremarkable urine sediment exam. Postrenal AKI may also lead to an unremarkable sediment, but hematuria and pyuria may be seen depending on the cause of obstruction. AKI from ATN due to ischemic injury, sepsis, or certain nephrotoxins has characteristic urine sediment findings: pigmented "muddy brown" granular casts and tubular epithelial cell casts. These findings may be absent in more than 20% of cases, however. Glomerulonephritis may lead to dysmorphic red blood cells or red blood cell casts. Interstitial nephritis may lead to white blood cell casts. The urine sediment findings overlap somewhat in glomerulonephritis and interstitial nephritis, and a diagnosis is not always possible on the

basis of the urine sediment alone. Urine eosinophils have a limited role in differential diagnosis; they can be seen in interstitial nephritis, pyelonephritis, cystitis, atheroembolic disease, or glomerulonephritis. Crystalluria may be important diagnostically. The finding of oxalate crystals in AKI should prompt an evaluation for ethylene glycol toxicity. Abundant uric acid crystals may be seen in the tumor lysis syndrome.

Blood Laboratory Findings

Certain forms of AKI are associated with characteristic patterns in the rise and fall of SCr. Prerenal azotemia typically leads to modest rises in SCr that return to baseline with improvement in hemodynamic status. Contrast nephropathy leads to a rise in SCr within 24–48 hours, peak within 3–5 days, and resolution within 5–7 days. In comparison, atheroembolic disease usually manifests with more subacute rises in SCr, although severe AKI with rapid increases in SCr can occur in this setting. With many of the epithelial cell toxins such as aminoglycoside antibiotics and cisplatin, the rise in SCr is characteristically delayed for 4–5 days to 2 weeks after initial exposure.

A complete blood count may provide diagnostic clues. Anemia is common in AKI and is usually multifactorial in origin. It is not related to an effect of AKI solely on production of red blood cells since this effect in isolation takes longer to manifest. Severe anemia in the absence of bleeding may reflect hemolysis, multiple myeloma, or thrombotic microangiopathy (e.g., HUS or TTP). Other laboratory findings of thrombotic microangiopathy include thrombocytopenia, schistocytes on peripheral blood smear, elevated lactate dehydrogenase level, and low haptoglobin content. Peripheral eosinophilia can accompany interstitial nephritis, atheroembolic disease, polyarteritis nodosa, and Churg-Strauss vasculitis.

AKI often leads to hyperkalemia, hyperphosphatemia, and hypocalcemia. Marked hyperphosphatemia with accompanying hypocalcemia, however, suggests rhabdomyolysis or the tumor lysis syndrome. Creatinine phosphokinase levels and serum uric acid are elevated in rhabdomyolysis, while tumor lysis syndrome shows normal or marginally elevated creatine kinase and markedly elevated serum uric acid. The anion gap may be increased with any cause of uremia due to retention of anions such as phosphate, hippurate, sulfate, and urate. The co-occurrence of an increased anion gap and an osmolal gap may suggest ethylene glycol poisoning, which may also cause oxalate crystalluria. Low anion gap may provide a clue to the diagnosis of multiple myeloma due to the presence of unmeasured cationic proteins. Laboratory blood tests helpful for the diagnosis of glomerulonephritis and vasculitis include depressed complement levels and high titers of antinuclear antibodies (ANAs), antineutrophilic cytoplasmic antibodies (ANCAs), antiglomerular basement membrane (AGBM) antibodies, and cryoglobulins.

Radiologic Evaluation

Postrenal AKI should always be considered in the differential diagnosis of AKI because treatment is usually successful if instituted early. Simple bladder catheterization can rule out urethral obstruction. Imaging of the urinary tract with renal ultrasound or CT should be undertaken to investigate obstruction in individuals with AKI unless an alternate diagnosis is apparent. Findings of obstruction include dilation of the collecting system and hydroureteronephrosis. Obstruction can be present without radiologic abnormalities in the setting of volume depletion, retroperitoneal fibrosis, encasement with tumor, and also early in the course of obstruction. If a high clinical index of suspicion for obstruction persists despite normal imaging, antegrade or retrograde pyelography should be performed. Imaging may also provide additional helpful information about kidney size and echogenicity to assist in the distinction between acute versus CKD. Large kidneys observed in these studies suggest the possibility of diabetic nephropathy, HIV-associated nephropathy, infiltrative diseases, or occasionally acute interstitial nephritis. Vascular imaging may be useful if venous or arterial obstruction is suspected, but the risks of contrast administration should be kept in mind. MRI with gadolinium-based contrast agents should be avoided if possible in severe AKI due to the possibility of inducing nephrogenic system fibrosis, a rare but serious complication seen most commonly in patients with end-stage renal disease.

Renal Failure Indices

Several indices have been used to help differentiate prerenal azotemia from intrinsic AKI when the tubules are malfunctioning. The low tubular flow rate and increased recycling of urea seen in prerenal azotemia may cause a disproportionate elevation of the BUN compared to creatinine. Other causes of disproportionate BUN elevation need to be kept in mind, however, including upper gastrointestinal bleeding, hyperalimentation, increased tissue catabolism, and glucocorticoid use.

The fractional excretion of sodium (FeNa) is the fraction of the filtered sodium load that is reabsorbed by the tubules and is a measure of both the kidney's ability to reabsorb sodium as well as endogenously and exogenously administered factors that affect tubular reabsorption. As such, it depends on sodium intake, effective intravascular volume, GFR, and intact tubular reabsorptive mechanisms. With prerenal azotemia, the FeNa may be below 1%, suggesting avid tubular sodium reabsorption. In patients with CKD, a FeNa significantly above 1% can still be present despite a prerenal state. The FeNa may also be above 1% despite hypovolemia due to treatment with diuretics. Low FeNa is often seen in glomerulonephritis (and other disorders), and, hence, should not be taken as prima facie evidence of prerenal azotemia. Low FeNa is therefore suggestive but not synonymous with effective intravascular volume depletion, and should not be used as the sole guide for volume management. The response of urine output to crystalloid or colloid fluid administration may be both diagnostic and therapeutic in prerenal azotemia. In ischemic AKI, the FeNa is frequently above 1% because of tubular injury and resultant inability to reabsorb sodium. Several causes of ischemia-associated and nephrotoxin-associated AKI can present with FeNa below 1%, however, including sepsis (often early in the course), rhabdomyolysis, and contrast nephropathy.

The ability of the kidneys to produce a concentrated urine is dependent upon many factors and reliant on good tubular function in multiple regions of the kidney. In the patient not taking diuretics and with good baseline kidney function, urine osmolality may be above 500 mOsm/kg in prerenal azotemia, consistent with an intact medullary gradient and elevated serum vasopressin levels causing water reabsorption resulting in concentrated urine. In elderly patients and those with CKD, however, baseline concentrating defects may exist, making urinary osmolality unreliable in many instances. Loss of concentrating ability is common in septic or ischemic AKI, resulting in urine osmolality below 350 mOsm/kg, but the finding is not specific.

Kidney Biopsy

If the cause of AKI is not apparent based on the clinical context, physical examination, and laboratory studies, kidney biopsy should be considered. The results of kidney biopsy can provide definitive diagnostic and prognostic information about acute and CKDs. The procedure is most often used in AKI when prerenal azotemia, postrenal AKI, and ischemic or nephrotoxic AKI have been deemed unlikely, and other possible diagnoses are being considered such as glomerulonephritis, vasculitis, interstitial nephritis, myeloma kidney, HUS and TTP, and allograft dysfunction. Kidney biopsy is associated with a risk of bleeding, which can be severe and organ or life threatening in patients with thrombocytopenia or coagulopathy.

Novel Biomarkers

BUN and creatinine are functional biomarkers of glomerular filtration rather than tissue injury biomarkers and, therefore, may be suboptimal for the diagnosis of actual parenchymal kidney damage. BUN and creatinine are also relatively slow to rise after kidney injury. Several novel kidney injury biomarkers have been investigated and show great promise for the early and accurate diagnosis of AKI. *Kidney injury molecule-1 (KIM-1)* is a type 1 transmembrane protein that is abundantly expressed in proximal tubular cells injured by ischemia or nephrotoxins such as cisplatin. KIM-1 is not expressed in appreciable quantities in the absence of tubular injury or in extrarenal tissues. KIM-1's functional role may be to confer phagocytic properties to tubular cells, enabling them to clear debris from the tubular lumen after kidney injury. KIM-1 can be detected shortly after ischemic or

nephrotoxic injury in the urine and, therefore, may be an easily tested biomarker in the clinical setting. *Neutrophil gelatinase associated lipocalin* (*NGAL*, also known as lipocalin-2 or siderocalin) is another leading novel biomarker of AKI. NGAL was first discovered as a protein in granules of human neutrophils. NGAL can bind to iron siderophore complexes and may have tissue-protective effects in the proximal tubule. NGAL is highly upregulated after inflammation and kidney injury and can be detected in the plasma and urine within 2 hours of cardiopulmonary bypass–associated AKI. Other injury markers that are being studied in an attempt to increase the early recognition of injury and predict the outcome in AKI are listed in Table 279-2.

Table 279-2 Biomarkers of Acute Kidney Injury

Biomarker	Comments	Detection	Species
Alanine aminopeptidase (AAP)	1. Proximal tubule brush border enzyme 2. Instability may limit clinical utility	Colorimetry	Rat, dog, human
Alkaline phosphatase (AP)	1. Proximal tubule brush border enzyme. Human intestinal alkaline phosphatase is specific for proximal tubular S3 segment; human tissue nonspecific alkaline phosphatase is specific for S1 and S2 segments 2. Levels may not correlate with extent of functional injury 3. Instability may limit clinical utility	Colorimetry	Rat, human
α-Glutathione-*S*-transferase (α-GST)	1. Proximal tubule cytosolic enzyme 2. Requires stabilization buffer for specimen storage and processing 3. Upregulated in AKI and renal cell carcinoma	ELISA	Mouse, rat, human
γ-Glutamyl transpeptidase (γGT)	1. Proximal tubule brush border enzyme 2. Instability requires samples to be analyzed quickly after collection, limiting clinical utility	Colorimetry	Rat, human
N-Acetyl-β-(D) glucosaminidase (NAG)	1. Proximal tubule lysosomal enzyme 2. More stable than other urinary enzymes 3. Extensive preclinical and clinical data in a variety of conditions (nephrotoxicant exposure, cardiopulmonary bypass, delayed renal allograft function, etc.) 4. Endogenous urea may inhibit activity	Colorimetry	Mouse, rat, human
β_2-Microglobulin	1. Light chain of the MHC I molecule expressed on the cell surface of all nucleated cells 2. Monomeric form is filtered by the glomerulus and reabsorbed by the proximal tubule cells 3. Early marker of tubular dysfunction in a variety of conditions 4. Instability in acidic urine limits clinical utility	ELISA Nephelometry	Mouse, rat, human
α_1-Microglobulin	1. Synthesized by the liver 2. Filtered by the glomerulus and reabsorbed by proximal tubule cells 3. Early marker of tubular dysfunction; high levels may predict poorer outcome 4. Stable across physiologic urinary pH	ELISA Nephelometry	Mouse, rat, human
Retinol-binding protein	1. Synthesized by liver, involved in vitamin A transport 2. Filtered by glomerulus and reabsorbed by proximal tubule cells 3. Early marker of tubular dysfunction 4. Increased stability in acidic urine when compared to β_2-microglobulin	ELISA Nephelometry	Mouse, rat, human
Cystatin C	1. Important extracellular inhibitor of cysteineproteases 2. Filtered by the glomerulus and reabsorbed by proximal tubule cells 3. Elevated urinary levels reflect tubular dysfunction; high levels may predict poorer outcome	ELISA Nephelometry	Mouse, rat, human
Microalbumin	1. Established marker for monitoring progression of chronic kidney disease 2. Elevated urinary levels may be indicative of proximal tubular damage 3. Lack of specificity for AKI may limit its utility	ELISA Immunoturbidimetry	Mouse, rat, dog, monkey, human
Kidney injury molecule-1 (KIM-1)	1. Type-1 cell membrane glycoprotein upregulated in dedifferentiated proximal tubule epithelial cells 2. Ectodomain is shed and can be quantitated in urine following AKI in preclinical and clinical studies 3. Elevated urinary levels are highly sensitive and specific for AKI 4. Upregulated following various models of preclinical and clinical AKI, fibrosis, renal cell carcinoma, and polycystic kidney disease	ELISA, Luminex®-based assay	Zebrafish, mouse, rat, dog, monkey, human
Clusterin	1. Expressed on dedifferentiated proximal tubular epithelial cells 2. Elevated kidney and urinary levels are very sensitive for AKI in preclinical models 3. Upregulated in various rodent models of AKI, fibrosis, renal cell carcinoma, and polycystic kidney disease 4. No clinical study demonstrating its use	ELISA	Mouse, rat, dog, monkey, human
Neutrophil	1. Initially identified bound to gelatinase in specific granules of the	ELISA	Mouse,

gelatinase associated lipocalin (NGAL)	neutrophils, but also may be induced in epithelial cells in the setting of inflammation or malignancy 2. Expression upregulated in kidney proximal tubule cells and urine following ischemic or cisplatin induced renal injury 3. Found to be an early indicator of AKI following cardiopulmonary bypass 4. Specificity for AKI in setting of sepsis and pyuria need to be further established	Luminex®-based assay	rat, human
Interleukin-18 (IL-18)	1. Cytokine with broad immunomodulatory properties, particularly in setting of ischemic injury 2. Constitutively expressed in distal tubules; strong immunoreactivity in proximal tubules with transplant rejection 3. Elevated urinary levels found to be early marker of AKI and independent predictor of mortality in critically ill patients	ELISA Luminex®-based assay	Mouse, rat, human
Cysteine-rich protein (CYR-61)	1. Induced in proximal straight tubules of kidney and secreted in the urine within 3–6 h following ischemic kidney injury 2. Urinary levels decrease rapidly in spite of progression of injury indicating stability issue 3. No clinical study demonstrating its use 4. No quantitative method established	Western blot	Mouse, rat, human
Osteopontin	1. Upregulated in various rodent models of AKI 2. The induction correlates with inflammation and tubulointerstitial fibrosis 3. No clinical study demonstrating its use	ELISA	Mouse, rat, monkey, human
Liver fatty acid –binding protein (L-FABP)	1. Expressed in proximal tubule epithelial cells 2. Current evidence suggests clinical utility as a biomarker in CKD and diabetic nephropathy 3. Additional studies necessary to determine utility in setting of preclinical and clinical AKI	ELISA	Mouse, rat, human
Sodium/hydrogen exchanger isoform (NHE3)	1. Most abundant sodium transporter in the renal tubule 2. Urinary levels found to discriminate between prerenal azotemia and AKI in ICU patients 3. Samples require considerable processing, limiting assay throughput	Immunoblotting	Mouse, rat, human
Exosomal fetuin-A	1. Acute phase protein synthesized in the liver and secreted into the circulation 2. Levels in proximal tubule cell cytoplasm correspond to degree of injury 3. Urinary levels found to be much higher in ICU patients with AKI compared to ICU patients without AKI and healthy volunteers 4. Samples require considerable processing, limiting assay throughput 5. Additional studies necessary to determine utility in setting of preclinical and clinical AKI	Immunoblotting	Rat, human

Abbreviations: AKI, acute kidney injury; ELISA, enzyme-linked immunosorbent assay; ICU, intensive care unit; RCC, renal cell carcinoma.

COMPLICATIONS

The kidney plays a central role in homeostatic control of volume status, blood pressure, plasma electrolyte composition, and acid-base balance, and for excretion of nitrogenous and other waste products. Complications associated with AKI are, therefore, protean, and depend on the severity of AKI and other associated conditions. Mild to moderate AKI may be entirely asymptomatic, particularly early in the course.

Uremia

Buildup of nitrogenous waste products, manifested as an elevated BUN concentration, is a hallmark of AKI. BUN itself poses little direct toxicity at levels below 100 mg/dL. At higher concentrations, mental status changes and bleeding complications can arise. Other toxins normally cleared by the kidney may be responsible for the symptom complex known as uremia. Few of the many possible uremic toxins have been definitively identified. The correlation of BUN and SCr concentrations with uremic symptoms is extremely variable, due in part to differences in urea and creatinine generation rates across individuals.

Hypervolemia and Hypovolemia

Expansion of extracellular fluid volume is a major complication of oliguric and anuric AKI, due to impaired salt and water excretion. The result can be weight gain, dependent edema, increased jugular venous pressure, and pulmonary edema; the latter can be life threatening. Pulmonary edema can also occur from volume overload and hemorrhage in pulmonary renal syndromes. AKI may also induce or exacerbate acute lung injury characterized by increased vascular permeability and inflammatory cell infiltration in lung parenchyma. Recovery from AKI can sometimes be accompanied by polyuria, which, if untreated, can lead to significant volume depletion. The polyuric phase of recovery may be due to an osmotic diuresis from retained urea and other waste products as well as delayed recovery of tubular reabsorptive functions.

Hyponatremia

Administration of excessive hypotonic crystalloid or isotonic dextrose solutions can result in hypoosmolality and hyponatremia, which, if severe, can cause neurologic abnormalities, including seizures.

Hyperkalemia

Abnormalities in plasma electrolyte composition can be mild or life threatening. Frequently the most concerning complication of AKI is hyperkalemia. Marked hyperkalemia is particularly common in rhabdomyolysis, hemolysis, and tumor lysis syndrome due to release of intracellular potassium from damaged cells. Potassium affects the cellular membrane potential of cardiac and

neuromuscular tissues. Muscle weakness may be a symptom of hyperkalemia. The more serious complication of hyperkalemia is due to effects on cardiac conduction, leading to potentially fatal arrhythmias.

Acidosis

Metabolic acidosis, usually accompanied by an elevation in the anion gap, is common in AKI, and can further complicate acid-base and potassium balance in individuals with other causes of acidosis, including sepsis, diabetic ketoacidosis, or respiratory acidosis.

Hyperphosphatemia and Hypocalcemia

AKI can lead to hyperphosphatemia, particularly in highly catabolic patients or those with AKI from rhabdomyolysis, hemolysis, and tumor lysis syndrome. Metastatic deposition of calcium phosphate can lead to hypocalcemia. AKI-associated hypocalcemia may also arise from derangements in the vitamin D–parathyroid axis. Hypocalcemia is often asymptomatic but can lead to perioral paresthesias, muscle cramps, seizures, carpopedal spasms, and prolongation of the QT interval on electrocardiography. Calcium levels should be corrected for the degree of hypoalbuminemia, if present, or ionized calcium levels should be followed. Mild, asymptomatic hypocalcemia does not require treatment.

Bleeding

Hematologic complications of AKI include anemia and bleeding, both of which are exacerbated by coexisting disease processes such as sepsis, liver disease, and disseminated intravascular coagulation. Direct hematologic effects from AKI-related uremia include decreased erythropoiesis and platelet dysfunction.

Infections

Infections are a common precipitant of AKI and also a dreaded complication of AKI. Impaired host immunity has been described in end-stage renal disease and may be operative in severe AKI.

Cardiac Complications

The major cardiac complications of AKI are arrhythmias, pericarditis, and pericardial effusion.

Malnutrition

AKI is often a severely hypercatabolic state, and, therefore, malnutrition is a major complication.

Treatment: Acute Kidney Injury

PREVENTION AND TREATMENT

The management of individuals with and at risk for AKI varies according to the underlying cause (Table 279-3). Common to all are several principles. Optimization of hemodynamics, correction of fluid and electrolyte imbalances, discontinuation of nephrotoxic medications, and dose adjustment of administered medications are all critical. Common causes of AKI such as sepsis and ischemic ATN, do not yet have specific therapies once injury is established, but meticulous clinical attention is needed to support the patient until (if) AKI resolves. The kidney possesses remarkable capacity to repair itself after even severe, dialysis-requiring AKI. However, some patients with AKI do not recover fully and may remain dialysis dependent.

Table 279-3 Management of Ischemia- and Nephrotoxin-Associated Aki

General Issues
1. Optimization of systemic and renal hemodynamics through volume resuscitation and judicious use of vasopressors
2. Elimination of nephrotoxic agents (e.g., ACE inhibitors, ARBs, NSAIDs, aminoglycosides) if possible
3. Initiation of renal replacement therapy when indicated
Specific Issues
1. Nephrotoxin-specific
a. Rhabdomyolysis: consider forced alkaline diuresis
b. Tumor lysis syndrome: allopurinol or rasburicase
2. Volume overload
a. Salt and water restriction
b. Diuretics
c. Ultrafiltration
3. Hyponatremia
a. Restriction of enteral free water intake, minimization of hypotonic intravenous solutions including those containing dextrose
4. Hyperkalemia
a. Restriction of dietary potassium intake
b. Discontinuation of potassium-sparing diuretics, ACE inhibitors, ARBs, NSAIDs
c. Loop diuretics to promote urinary potassium loss
d. Potassium binding ion-exchange resin (sodium polystyrene sulfonate)
e. Insulin (10 units regular) and glucose (50 mL of 50% dextrose) to promote entry of potassium intracellularly
f. Inhaled beta-agonist therapy to promote entry of potassium intracellularly
g. Calcium gluconate or calcium chloride (1 g) to stabilize the myocardium
5. Metabolic acidosis
a. Sodium bicarbonate (if pH <7.2 to keep serum bicarbonate >15 mmol/L)
b. Administration of other bases e.g., THAM
c. Renal replacement therapy
6. Hyperphosphatemia
a. Restriction of dietary phosphate intake
b. Phosphate binding agents (calcium acetate, sevelamer hydrochloride, aluminum hydroxide—taken with meals)
7. Hypocalcemia

a. Calcium carbonate or calcium gluconate if symptomatic

8. Hypermagnesemia

a. Discontinue Mg^{2+} containing antacids

9. Hyperuricemia

a. Acute treatment is usually not required except in the setting of tumor lysis syndrome (see above)

10. Nutrition

a. Sufficient protein and calorie intake to avoid negative nitrogen balance

11. Drug dosing

a. Careful attention to dosages and frequency of administration of drugs, adjustment for degree of renal failure

Abbreviations: ACE, angiotensin-converting enzyme; ARBs, angiotensin receptor blocker; NSAIDs, nonsteroidal anti-inflammatory drug; TRAM, tris (hydroxymethyl) aminomethane.

Prerenal Azotemia

Prevention and treatment of prerenal azotemia requires optimization of renal perfusion. The composition of replacement fluids should be targeted to the type of fluid lost. Severe acute blood loss should be treated with packed red blood cells. Isotonic crystalloid and/or colloid should be used for less severe acute hemorrhage or plasma loss in the case of burns and pancreatitis. Crystalloid solutions are less expensive and probably equally efficacious as colloid solutions. Crystalloid has been reported to be preferable to albumin in the setting of traumatic brain injury. Isotonic crystalloid (e.g., 0.9% saline) or colloid should be used for volume resuscitation in severe hypovolemia, whereas hypotonic crystalloids (e.g., 0.45% saline) suffice for less severe hypovolemia. Excessive chloride administration from 0.9% saline may lead to hyperchloremic metabolic acidosis. Use of bicarbonate-containing solutions (e.g., dextrose water with 150 mEq sodium bicarbonate) should be used if metabolic acidosis is a concern.

Optimization of cardiac function in the cardiorenal syndrome (i.e., renal hypoperfusion from poor cardiac output) may require use of inotropic agents, preload- and afterload-reducing agents, antiarrhythmic drugs, and mechanical aids such as an intraaortic balloon pump. Invasive hemodynamic monitoring to guide therapy may be necessary.

Cirrhosis and Hepatorenal Syndrome

Fluid management in individuals with cirrhosis, ascites, and AKI is challenging because of the frequent difficulty in ascertaining intravascular volume status. Administration of intravenous fluids as a volume challenge may be required diagnostically as well as therapeutically. Excessive volume administration may, however, result in worsening ascites and pulmonary compromise in the setting of hepatorenal syndrome or AKI due to superimposed spontaneous bacterial peritonitis. Peritonitis should be ruled out by culture of ascitic fluid. Albumin may prevent AKI in those treated with antibiotics for spontaneous bacterial peritonitis. The definitive treatment of the hepatorenal syndrome is orthotopic liver transplantation. Bridge therapies that have shown promise include terlipressin (a vasopressin analog), combination therapy with octreotide (a somatostatin analog) and midodrine (an α1-adrenergic agonist), and norepinephrine, all in combination with intravenous albumin (25–50 mg per day, maximum 100 g/d).

Intrinsic Aki

Several agents have been tested and have failed to show benefit in the treatment of ischemic acute tubular injury. These include atrial natriuretic peptide, low-dose dopamine, endothelin antagonists, loop diuretics, calcium channel blockers, α-adrenergic receptor blockers, prostaglandin analogs, antioxidants, antibodies against leukocyte adhesion molecules, and insulin-like growth factor and many others. Most studies have enrolled patients with severe and well-established AKI, and treatment may have been initiated too late. Novel kidney injury biomarkers may provide an opportunity to test agents earlier in the course of AKI.

AKI due to acute glomerulonephritis or vasculitis may respond to immunosuppressive agents and/or plasmapheresis (Chap. 277). Allergic interstitial nephritis due to medications requires discontinuation of the offending agent. Glucocorticoids have been used, but not tested in randomized trials, in cases where AKI persists or worsens despite discontinuation of the suspected medication. AKI due to scleroderma (scleroderma renal crisis) should be treated with ACE inhibitors.

Early and aggressive volume repletion is mandatory in patients with rhabdomyolysis, who may require 10 L of fluid per day. Alkaline fluids (e.g., 75 mmol sodium bicarbonate added to 0.45% saline) may be beneficial in preventing tubular injury and cast formation, but carry the risk of worsening hypocalcemia. Diuretics may be used if fluid repletion is adequate but unsuccessful in achieving urinary flow rates of 200–300 mL/h. There is no specific therapy for established AKI in rhabdomyolysis, other than dialysis in severe cases or general supportive care to maintain fluid and electrolyte balance and tissue perfusion. Careful attention must be focused on calcium and phosphate status because of precipitation in damaged tissue and released when the tissue heals.

Postrenal Aki

Prompt recognition and relief of urinary tract obstruction can forestall the development of permanent structural damage induced by urinary stasis. The site of obstruction defines the treatment approach. Transurethral or suprapubic bladder catheterization may be all that is needed initially for urethral strictures or functional bladder impairment. Ureteric obstruction may be treated by percutaneous nephrostomy tube placement or ureteral stent placement. Relief of obstruction is usually followed by an appropriate diuresis for several days. In rare cases, severe polyuria persists due to tubular dysfunction and may require continued administration of intravenous fluids and electrolytes for a period of time.

SUPPORTIVE MEASURES

Volume Management

Hypervolemia in oliguric or anuric AKI may be life threatening due to acute pulmonary edema, especially since many patients have coexisting pulmonary disease, and AKI likely increases pulmonary vascular permeability. Fluid and sodium should be restricted, and diuretics may be used to increase the urinary flow rate. There is no evidence that increasing urine output itself improves the natural history of AKI, but diuretics may help to avoid the need for dialysis in some cases. In severe cases of volume overload, furosemide may be given as a bolus (200 mg) followed by an intravenous drip (10–40 mg/h), with or without a thiazide diuretic. Diuretic therapy should be stopped if there is no response. Dopamine in low doses may transiently increase salt and water excretion by the kidney in prerenal states, but clinical trials have failed to show any benefit in patients with intrinsic AKI. Because of the risk of arrhythmias and potential bowel ischemia, it has been argued that the risks of dopamine may outweigh the benefits in the treatment or prevention of AKI.

Electrolyte and Acid-Base Abnormalities

The treatment of dysnatremias and hyperkalemia is described in Chap. 45. Metabolic acidosis is not treated unless severe (pH

<7.20 and serum bicarbonate <15 mmol/L). Acidosis can be treated with oral or intravenous sodium bicarbonate (Chap. 47), but overcorrection should be avoided because of the possibility of metabolic alkalosis, hypocalcemia, hypokalemia, and volume overload. Hyperphosphatemia is common in AKI and can usually be treated by limiting intestinal absorption of phosphate using phosphate binders (calcium carbonate, calcium acetate, sevelamer, or aluminum hydroxide). Hypocalcemia does not usually require therapy unless symptoms are present.

Malnutrition

Protein energy wasting is common in AKI, particularly in the setting of multisystem organ failure. Inadequate nutrition may lead to starvation ketoacidosis and protein catabolism. Excessive nutrition may increase the generation of nitrogenous waste and lead to worsening azotemia. Total parenteral nutrition requires large volumes of fluid administration and may complicate efforts at volume control.

Anemia

The anemia seen in AKI is usually multifactorial and is not improved by erythropoiesis stimulating agents, due to their delayed onset of action and the presence of bone marrow resistance in critically ill patients. Uremic bleeding may respond to desmopressin or estrogens, but may require dialysis in the case of longstanding or severe uremia. Gastrointestinal prophylaxis with proton pump inhibitors or histamine (H_2) receptor blockers is required. Venous thromboembolism prophylaxis is important and should be tailored to the clinical setting; low-molecular-weight heparins and factor Xa inhibitors have unpredictable pharmacokinetics in severe AKI and should be avoided.

Dialysis Indications and Modalities

(See also Chap. 281) Dialysis is indicated when medical management fails to control volume overload, hyperkalemia, acidosis, in some toxic ingestions, and when there are severe complications of uremia (asterixis, pericardial rub or effusion, encephalopathy, uremic bleeding). The timing of dialysis is still a matter of debate. Late initiation of dialysis carries the risk of avoidable volume, electrolyte, and metabolic complications of AKI. On the other hand, initiating dialysis too early may unnecessarily expose individuals to intravenous lines and invasive procedures, with the attendant risks of infection, bleeding, procedural complications, and hypotension. The initiation of dialysis should not await the development of a life-threatening complication of renal failure. Many nephrologists initiate dialysis for AKI empirically when the BUN exceeds 100 mg/dL in patients without clinical signs of recovery of kidney function.

The available modes for renal replacement therapy in AKI require either access to the peritoneal cavity (for peritoneal dialysis) or the large blood vessels (for hemodialysis, hemofiltration, and other hybrid procedures). Small solutes are removed across a semipermeable membrane down their concentration gradient ("diffusive" clearance) and/or along with the movement of plasma water ("convective" clearance). The choice of modality is often dictated by the immediate availability of technology and the expertise of medical staff. Peritoneal dialysis is performed through a temporary intraperitoneal catheter. It is rarely used in the United States for AKI in adults but has enjoyed widespread use internationally, particularly when hemodialysis technology is not available. Dialysate solution is instilled into and removed from the peritoneal cavity at regular intervals in order to achieve diffusive and convective clearance of solutes across the peritoneal membrane; ultrafiltration of water is achieved by the presence of an osmotic gradient across the peritoneal membrane, typically due to high concentrations of dextrose in the dialysate. Being a continuous procedure, it is often better tolerated than intermittent procedures like hemodialysis in hypotensive patients. Peritoneal dialysis may not be sufficient for hypercatabolic patients due to inherent limitations in dialysis efficacy.

Hemodialysis can be employed intermittently or continuously, and can be done through convective clearance, diffusive clearance, or a combination of the two. Vascular access is through the femoral, internal jugular, or subclavian veins. Hemodialysis is an intermittent procedure that removes solutes through diffusive and convective clearance. Hemodialysis is performed 3–4 h per d, three to four times per week, and is the most common form of renal replacement therapy for AKI. One of the major complications of hemodialysis is hypotension, particularly in the critically ill.

Continuous intravascular procedures were developed in the early 1980s to treat hemodynamically unstable patients without inducing the rapid shifts of volume, osmolarity, and electrolytes characteristic of intermittent hemodialysis. Continuous renal replacement therapy (CRRT) can be performed by convective clearance [continuous venovenous hemofiltration (CVVH)], in which large volumes of plasma water (and accompanying solutes) are forced across the semipermeable membrane by means of hydrostatic pressure; the plasma water is then replaced by a physiologic crystalloid solution. CRRT can also be performed by diffusive clearance [continuous venovenous hemodialysis (CVVHD)], a technology similar to hemodialysis except at lower blood flow and dialysate flow rates. A hybrid therapy combines both diffusive and convective clearance [continuous venovenous hemodiafiltration (CVVHDF)]. To achieve some of the advantages of CRRT without the need for 24-h staffing of the procedure, a newer form of therapy has been introduced, termed slow low-efficiency dialysis (SLED) or extended daily dialysis (EDD). In this therapy, blood flow and dialysate flow are higher than in CVVHD, but the treatment time is reduced to 12 h.

The optimal dose of dialysis for AKI is not clear. Daily intermittent hemodialysis and high-dose CRRT do not confer a demonstrable survival or renal recovery advantage, but care should be taken to avoid undertreatment. Studies have failed to show that continuous therapies are superior to intermittent therapies. If available, CRRT is often preferred in patients with severe hemodynamic instability, cerebral edema, or significant volume overload.

OUTCOME AND PROGNOSIS

The development of AKI is associated with a significantly increased risk of in-hospital and long-term mortality, longer length of stay, and increased costs. Prerenal azotemia, with the exception of the cardiorenal and hepatorenal syndromes, and postrenal azotemia carry a better prognosis than most cases of intrinsic AKI. The kidneys may recover even after severe, dialysis-requiring AKI. Survivors of an episode of AKI requiring temporary dialysis, however, are at extremely high risk for progressive CKD, and up to 10% may develop end-stage renal disease. Postdischarge care under the supervision of a nephrologist for aggressive secondary prevention of kidney disease is prudent. Patients with AKI are more likely to die prematurely after they leave the hospital even if their kidney function has recovered.

FURTHER READINGS

Bonventre JV: Pathophysiology of AKI: Injury and normal and abnormal repair. Contrib Nephrol 165:9, 2010[PMID: 20427950]

——— et al: Next-generation biomarkers for detecting kidney toxicity. Nat Biotechnol 28:436, 2010

Chertow GM et al: Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. J Am Soc Nephrol 16:3365, 2005[PMID: 16177006]

Coca SG et al: Biomarkers for the diagnosis and risk stratification of acute kidney injury: a systematic review. Kidney Int 73:1008, 2008[PMID: 18094679]

Devarajan P: Update on mechanisms of ischemic acute kidney injury. J Am Soc Nephrol 17:1503, 2006[PMID: 16707563]

Waikar SS, Bonventre JV: Creatinine kinetics and the definition of acute kidney injury. J Am Soc Nephrol 20:672, 2009[PMID: 19244578]

Wald R: Chronic dialysis and death among survivors of acute kidney injury requiring dialysis. JAMA 302:1532, 2009.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHRONIC KIDNEY DISEASE: INTRODUCTION

Chronic kidney disease (CKD) encompasses a spectrum of different pathophysiologic processes associated with abnormal kidney function and a progressive decline in glomerular filtration rate (GFR). **Table 280-1** provides a widely accepted classification, based on guidelines of the National Kidney Foundation [Kidney Dialysis Outcomes Quality Initiative (KDOQI)], in which stages of CKD are defined according to the estimated GFR.

Table 280-1 Classification of Chronic Kidney Disease (CKD)

Stage	GFR, mL/min per 1.73 m^2
0	>90[a]
1	≥90[b]
2	60–89
3	30–59
4	15–29
5	<15

[a]With risk factors for CKD (see text). [b]With demonstrated kidney damage (e.g., persistent proteinuria, abnormal urine sediment, abnormal blood and urine chemistry, abnormal imaging studies). *Abbreviation:* GFR, glomerular filtration rate. *Source:* Modified from National Kidney Foundation. K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, classification and stratification. Am J Kidney Dis 39:suppl 1, 2002.

The term *chronic renal failure* applies to the process of continuing significant irreversible reduction in nephron number and typically corresponds to CKD stages 3–5. The pathophysiologic processes and adaptations associated with chronic renal failure will be the focus of this chapter. The dispiriting term *end-stage renal disease* represents a stage of CKD where the accumulation of toxins, fluid, and electrolytes normally excreted by the kidneys results in the *uremic syndrome*. This syndrome leads to death unless the toxins are removed by renal replacement therapy, using dialysis or kidney transplantation. These latter interventions are discussed in Chaps. 281 and 282. *End-stage renal disease* will be supplanted in this chapter by the term *stage 5 CKD*.

Pathophysiology of Chronic Kidney Disease

The pathophysiology of CKD involves two broad sets of mechanisms of damage: (1) initiating mechanisms specific to the underlying etiology (e.g., genetically determined abnormalities in kidney development or integrity, immune complex deposition and inflammation in certain types of glomerulonephritis, or toxin exposure in certain diseases of the renal tubules and interstitium) and (2) a set of progressive mechanisms, involving hyperfiltration and hypertrophy of the remaining viable nephrons, that are a common consequence following long-term reduction of renal mass, irrespective of underlying etiology (Chap. 278). The responses to reduction in nephron number are mediated by vasoactive hormones, cytokines, and growth factors. Eventually, these short-term adaptations of hypertrophy and hyperfiltration become maladaptive as the increased pressure and flow predisposes to distortion of glomerular architecture, associated with sclerosis and dropout of the remaining nephrons **(Fig. 280-1)**. Increased intrarenal activity of the renin-angiotensin axis appears to contribute both to the initial adaptive hyperfiltration and to the subsequent maladaptive hypertrophy and sclerosis, the latter, in part, owing to the stimulation of transforming growth factor β (TGF-β). This process explains why a reduction in renal mass from an isolated insult may lead to a progressive decline in renal function over many years (Fig. 280-2).

Figure 280-1

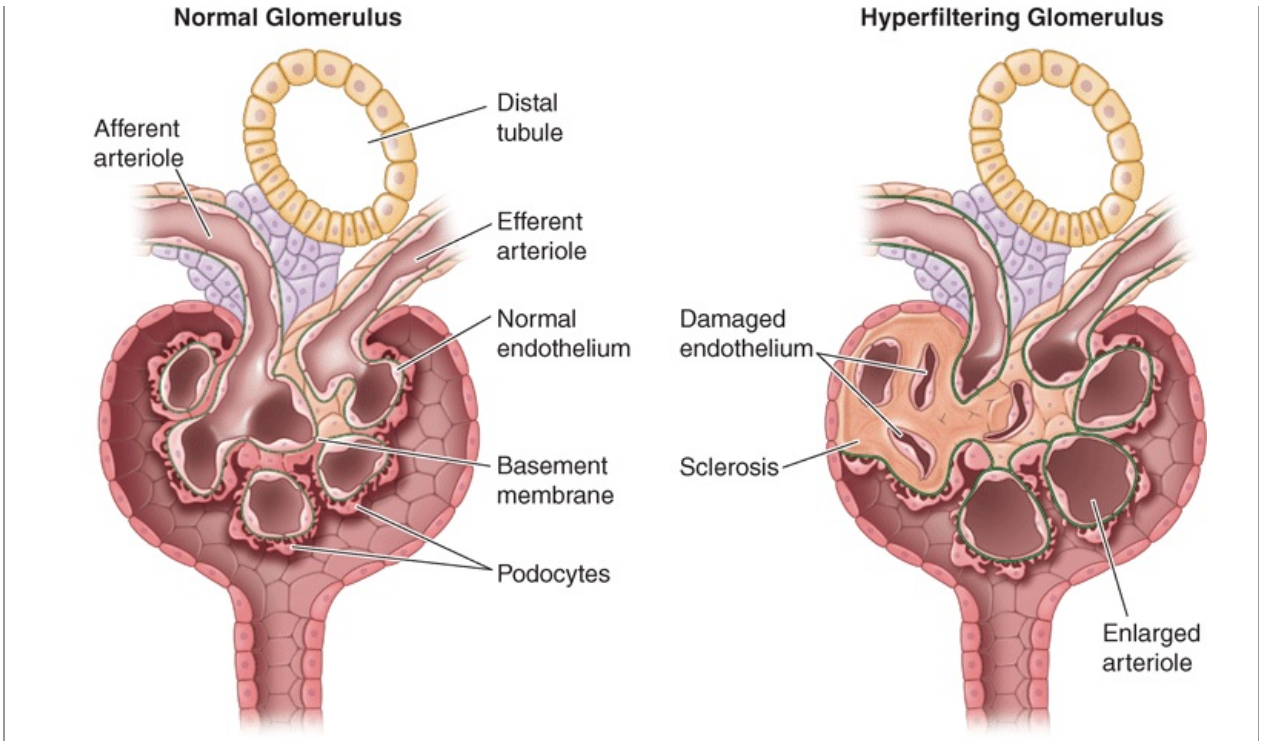

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left: Schema of the normal glomerular architecture. Right: Secondary glomerular changes associated with a reduction in nephron number, including enlargement of capillary lumens and focal adhesions, which are thought to occur consequent to compensatory hyperfiltration and hypertrophy in the remaining nephrons. (*Modified from JR Ingelfinger: N Engl J Med 348:99, 2003* .)

Identification of Risk Factors and Staging of CKD

It is important to identify factors that increase the risk for CKD, even in individuals with normal GFR. Risk factors include hypertension, diabetes mellitus, autoimmune disease, older age, African ancestry, a family history of renal disease, a previous episode of acute kidney injury, and the presence of proteinuria, abnormal urinary sediment, or structural abnormalities of the urinary tract.

Recent research in the genetics of predisposition to common complex diseases (Chap. 61) has revealed DNA sequence variants at a number of genetic loci that are associated with common forms of CKD. A striking example is the finding of allelic versions of the *APOL1* gene, of West African population ancestry, which contributes to the several-fold higher frequency of certain common etiologies of CKD (e.g., focal segmental glomerulosclerosis) observed among African and Hispanic Americans. The prevalence in West African populations seems to have an evolutionary basis, since these same variants offer protection from tropical pathogens.

In order to stage CKD, it is necessary to estimate the GFR. Two equations commonly used to estimate GFR are shown in **Table 280-2** and incorporate the measured plasma creatinine concentration, age, sex, and ethnic origin. Many laboratories now report an estimated GFR, or "eGFR," using one of these equations.

Table 280-2 Recommended Equations for Estimation of Glomerular Filtration Rate (GFR) Using Serum Creatinine Concentration (P_{Cr}), Age, Sex, Race, and Body Weight

1. Equation from the Modification of Diet in Renal Disease study*

   Estimated GFR (mL/min per 1.73 m^2) = 1.86 x $(P_{Cr})^{-1.154}$ x $(age)^{-0.203}$

   Multiply by 0.742 for women

   Multiply by 1.21 for African Americans
2. Cockcroft-Gault equation

   Estimated creatinine clearance (mL/min)

   $= \frac{(140 - age) \times body\ weight\ (kg)}{72 \times P_{Cr}\ (mg/dL)}$

Multiply by 0.85 for women

*Equation is *available* in hand-held calculators and in tabular form.

Source: Adapted from AS Levey et al: Am J Kidney Dis 39: S1, 2002, with permission.

The normal annual mean decline in GFR with age from the peak GFR (~120 mL/min per 1.73 m^2) attained during the third decade of life is ~1 mL/min per year per 1.73 m^2, reaching a mean value of 70 mL/min per 1.73 m^2 at age 70. The mean GFR is lower in women than in men. For example, a woman in her 80s with a normal serum creatinine may have a GFR of just 50 mL/min per 1.73 m^2. Thus, even a mild elevation in serum creatinine concentration [e.g., 130 μmol/L (1.5 mg/dL)] often signifies a substantial reduction in GFR in most individuals.

Measurement of albuminuria is also helpful for monitoring nephron injury and the response to therapy in many forms of CKD, especially chronic glomerular diseases. While an accurate 24-h urine collection is the criterion standard for measurement of albuminuria, the measurement of protein-to-creatinine ratio in a spot first-morning urine sample is often more practical to obtain and correlates well, but not perfectly, with 24-h urine collections. Persistence in the urine of >17 mg of albumin per gram of creatinine in adult males and 25 mg albumin per gram of creatinine in adult females usually signifies chronic renal damage. *Microalbuminuria* refers to the excretion of amounts of albumin too small to detect by urinary dipstick or conventional measures of urine protein. It is a good screening test for early detection of renal disease, and may be a marker for the presence of microvascular disease in general. If a patient has a large amount of excreted albumin, there is no reason to test for microalbuminuria.

Stages 1 and 2 CKD are usually not associated with any symptoms arising from the decrement in GFR. However, there may be symptoms from the underlying renal disease itself, such as edema in patients with nephrotic syndrome or signs of hypertension secondary to the renal parenchymal disease in patients with polycystic kidney disease, some forms of glomerulonephritis, and many other parenchymal and vascular renal diseases, even with well-preserved GFR. If the decline in GFR progresses to stages 3 and 4, clinical and laboratory complications of CKD become more prominent. Virtually all organ systems are affected, but the most evident complications include anemia and associated easy fatigability; decreasing appetite with progressive malnutrition; abnormalities in calcium, phosphorus, and mineral-regulating hormones, such as $1,25(OH)_2D3$ (calcitriol), parathyroid hormone (PTH), and fibroblast growth factor 23 (FGF-23); and abnormalities in sodium, potassium, water, and acid-base homeostasis. Many patients, especially the elderly, will have eGFR values compatible with stage 2 or 3 CKD. However, the majority of these patients will show no further deterioration of renal function. The primary care physician is advised to recheck kidney function, and if it is stable and not associated with proteinuria, the patient can usually be managed in this setting. However, if there is evidence of decline of GFR and uncontrolled hypertension or proteinuria, referral to a nephrologist is appropriate. If the patient progresses to stage 5 CKD, toxins accumulate such that patients usually experience a marked disturbance in their activities of daily living, well-being, nutritional status, and water and electrolyte homeostasis, eventuating in the *uremic syndrome*. As noted, this state will culminate in death unless renal replacement therapy (dialysis or transplantation) is instituted.

Etiology and Epidemiology

It has been estimated from population survey data that at least 6% of the adult population in the United States has CKD at stages 1 and 2. An unknown subset of this group will progress to more advanced stages of CKD. An additional 4.5% of the U.S. population is estimated to have stages 3 and 4 CKD. **Table 280-3** lists the five most frequent categories of causes of CKD, cumulatively accounting for greater than 90% of the CKD disease burden worldwide. The relative contribution of each category varies among different geographic regions. The most frequent cause of CKD in North America and Europe is diabetic nephropathy, most often secondary to type 2 diabetes mellitus. Patients with newly diagnosed CKD often also present with hypertension. When no overt evidence for a primary glomerular or tubulointerstitial kidney disease process is present, CKD is often attributed to hypertension. However, it is now appreciated that such individuals can be considered in two categories. The first includes patients with a silent primary glomerulopathy, such as focal segmental glomerulosclerosis, without the overt nephrotic or nephritic manifestations of glomerular disease (Chap. 283). The second includes patients in whom progressive nephrosclerosis and hypertension is the renal correlate of a systemic vascular disease, often also involving large- and small-vessel cardiac and cerebral pathology. This latter combination is especially common in the elderly, in whom chronic renal ischemia as a cause of CKD may be underdiagnosed. The increasing incidence of CKD in the elderly has been ascribed, in part, to decreased mortality rate from the cardiac and cerebral complications of atherosclerotic vascular disease, enabling a greater segment of the population to manifest the renal component of generalized vascular disease. Nevertheless, it should be appreciated that the vast majority of such patients with early stages of CKD will succumb to the cardiovascular and cerebrovascular consequences of the vascular disease before they can progress to the most

advanced stages of CKD. Indeed, even a minor decrement in GFR or the presence of albuminuria is now recognized as a major risk factor for cardiovascular disease.

Table 280-3 Leading Categories of Etiologies of CKD*

- Diabetic glomerular disease
- Glomerulonephritis
- Hypertensive nephropathy
  - Primary glomerulopathy with hypertension
  - Vascular and ischemic renal disease
- Autosomal dominant polycystic kidney disease
- Other cystic and tubulointerstitial nephropathy

*Relative contribution of each category varies with geographic region.

Pathophysiology and Biochemistry of Uremia

Although serum urea and creatinine concentrations are used to measure the excretory capacity of the kidneys, accumulation of these two molecules themselves do not account for the many symptoms and signs that characterize the uremic syndrome in advanced renal failure. Hundreds of toxins that accumulate in renal failure have been implicated in the uremic syndrome. These include water-soluble, hydrophobic, protein-bound, charged, and uncharged compounds. Additional categories of nitrogenous excretory products include guanidino compounds, urates and hippurates, products of nucleic acid metabolism, polyamines, myoinositol, phenols, benzoates, and indoles. Compounds with a molecular mass between 500 and 1500 Da, the so-called middle molecules, are also retained and contribute to morbidity and mortality. It is thus evident that the serum concentrations of urea and creatinine should be viewed as being readily measured, but incomplete, surrogate markers for these compounds, and monitoring the levels of urea and creatinine in the patient with impaired kidney function represents a vast oversimplification of the uremic state.

The uremic syndrome and the disease state associated with advanced renal impairment involve more than renal excretory failure. A host of metabolic and endocrine functions normally performed by the kidneys is also impaired or suppressed, and this results in anemia, malnutrition, and abnormal metabolism of carbohydrates, fats, and proteins. Furthermore, plasma levels of many hormones, including PTH, FGF-23, insulin, glucagon, steroid hormones including vitamin D and sex hormones, and prolactin, change with renal failure as a result of urinary retention, decreased degradation, or abnormal regulation. Finally, progressive renal impairment is associated with worsening systemic inflammation. Elevated levels of C-reactive protein are detected along with other acute-phase reactants, while levels of so-called negative acute-phase reactants, such as albumin and fetuin, decline with progressive renal impairment, even in nonproteinuric kidney disease. Thus, the inflammation associated with renal impairment is important in the malnutrition-inflammation-atherosclerosis/calcification syndrome, which contributes in turn to the acceleration of vascular disease and comorbidity rate associated with advanced kidney disease.

In summary, the pathophysiology of the uremic syndrome can be divided into manifestations in three spheres of dysfunction: (1) those consequent to the accumulation of toxins that normally undergo renal excretion, including products of protein metabolism; (2) those consequent to the loss of other renal functions, such as fluid and electrolyte homeostasis and hormone regulation; and (3) progressive systemic inflammation and its vascular and nutritional consequences.

CLINICAL AND LABORATORY MANIFESTATIONS OF CHRONIC KIDNEY DISEASE AND UREMIA

Uremia leads to disturbances in the function of virtually every organ system. Chronic dialysis can reduce the incidence and severity of many of these disturbances, so that the overt and florid manifestations of uremia have largely disappeared in the modern health setting. However, as indicated in **Table 280-4**, even optimal dialysis therapy is not completely effective as renal replacement therapy, because some disturbances resulting from impaired renal function fail to respond to dialysis.

Table 280-4 Clinical Abnormalities in Uremia[a]

Fluid and electrolyte disturbances	Neuromuscular disturbances	Dermatologic disturbances
Volume expansion (I)	Fatigue (I)[b]	Pallor (I)[b]
Hyponatremia (I)	Sleep disorders (P)	Hyperpigmentation (I, P, or D)
Hyperkalemia (I)	Headache (P)	Pruritus (P)

Hyperkalemia (I)	Impaired mentation (I)[b]	Ecchymoses (I)
Hyperphosphatemia (I)	Lethargy (I)[b]	Nephrogenic fibrosing dermopathy (D)
Endocrine-metabolic disturbances	Asterixis (I)	Uremic frost (I)
Secondary hyperparathyroidism (I or P)	Muscular irritability	**Gastrointestinal disturbances**
Adynamic bone (D)	Peripheral neuropathy (I or P)	Anorexia (I)
Vitamin D–deficient osteomalacia (I)	Restless legs syndrome (I or P)	Nausea and vomiting (I)
Carbohydrate resistance (I)	Myoclonus (I)	Gastroenteritis (I)
Hyperuricemia (I or P)	Seizures (I or P)	Peptic ulcer (I or P)
Hypertriglyceridemia (I or P)	Coma (I)	Gastrointestinal bleeding (I, P, or D)
Increased Lp(a) level (P)	Muscle cramps (P or D)	Idiopathicascites (D)
Decreased high-density lipoprotein level (P)	Dialysis disequilibrium syndrome (D)	Peritonitis (D)
Protein-energy malnutrition (I or P)	Myopathy (P or D)	**Hematologic and immunologic disturbances**
Impaired growth and development (P)	**Cardiovascular and pulmonary disturbances**	Anemia (I)[b]
Infertility and sexual dysfunction (P)	Arterial hypertension (I or P)	Lymphocytopenia (P)
Amenorrhea (I/P)	Congestive heart failure or pulmonary edema (I)	Bleeding diathesis (I or D)[b]
β_2-Microglobulin–associated amyloidosis (P or D)	Pericarditis (I)	Increased susceptibility to infection (I or P)
	Hypertrophic or dilated cardiomyopathy (I, P, or D)	Leukopenia (D)
	Uremic lung (I)	Thrombocytopenia (D)
	Accelerated atherosclerosis (P or D)	
	Hypotension and arrhythmias (D)	
	Vascular calcification (P or D)	

[a]Virtually all abnormalities in this table are completely reversed in time by successful renal transplantation. The response of these abnormalities to hemodialysis or peritoneal dialysis therapy is more variable. (I) denotes an abnormality that usually improves with an optimal program of dialysis and related therapy; (P) denotes an abnormality that tends to persist or even progress, despite an optimal program; (D) denotes an abnormality that develops only after initiation of dialysis therapy.

[b]Improves with dialysis and erythropoietin therapy.

Abbreviation: Lp(a), lipoprotein A.

Fluid, Electrolyte, and Acid-Base Disorders

SODIUM AND WATER HOMEOSTASIS

In most patients with stable CKD, the total-body content of sodium and water is modestly increased, although this may not be apparent on clinical examination. Normal renal function guarantees that the tubular reabsorption of filtered sodium and water is adjusted so that urinary excretion matches intake. Many forms of renal disease (e.g., glomerulonephritis) disrupt this glomerulotubular balance such that dietary intake of sodium exceeds its urinary excretion, leading to sodium retention and attendant extracellular fluid volume (ECFV) expansion. This expansion may contribute to hypertension, which itself can accelerate the nephron injury. As long as water intake does not exceed the capacity for water clearance, the ECFV expansion will be isotonic and the patient will have a normal plasma sodium concentration and effective osmolality (Chap. 278). Hyponatremia is not commonly seen in CKD patients but, when present, can respond to water restriction. If the patient has evidence of ECFV expansion (peripheral edema, sometimes hypertension poorly responsive to therapy), he or she should be counseled regarding salt restriction. Thiazide diuretics have limited utility in stages 3–5 CKD, such that administration of loop diuretics, including furosemide, bumetanide, or torsemide, may also be needed. Resistance to loop diuretics in renal failure often mandates use of higher doses than those used in patients with near-normal kidney function. The combination of loop diuretics with metolazone, which inhibits the sodium chloride co-transporter of the distal convoluted tubule, can help effect renal salt excretion. Ongoing diuretic resistance with intractable edema and hypertension in advanced CKD may serve as an indication to initiate dialysis.

In addition to problems with salt and water excretion, some patients with CKD may instead have impaired renal conservation of sodium and water. When an extrarenal cause for fluid loss, such as

gastrointestinal (GI) loss, is present, these patients may be prone to ECFV depletion because of the inability of the failing kidney to reclaim filtered sodium adequately. Furthermore, depletion of ECFV, whether due to GI losses or overzealous diuretic therapy, can further compromise kidney function through underperfusion, or a "prerenal" basis, leading to acute-on-chronic kidney failure. In this setting, cautious volume repletion with normal saline may return the ECFV to normal and restore renal function to baseline without having to intervene with dialysis.

POTASSIUM HOMEOSTASIS

In CKD, the decline in GFR is not necessarily accompanied by a parallel decline in urinary potassium excretion, which is predominantly mediated by aldosterone-dependent secretory events in distal nephron segments. Another defense against potassium retention in these patients is augmented potassium excretion in the GI tract. Notwithstanding these two homeostatic responses, hyperkalemia may be precipitated in certain settings. These include increased dietary potassium intake, protein catabolism, hemolysis, hemorrhage, transfusion of stored red blood cells, and metabolic acidosis. In addition, a host of medications can inhibit renal potassium excretion. The most important medications in this respect include the angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers (ARBs), and spironolactone and other potassium-sparing diuretics such as amiloride, eplerenone, and triamterene.

Certain causes of CKD can be associated with earlier and more severe disruption of potassium-secretory mechanisms in the distal nephron, out of proportion to the decline in GFR. These include conditions associated with hyporeninemic hypoaldosteronism, such as diabetes, and renal diseases that preferentially affect the distal nephron, such as obstructive uropathy and sickle cell nephropathy.

Hypokalemia is not common in CKD and usually reflects markedly reduced dietary potassium intake, especially in association with excessive diuretic therapy or concurrent GI losses. Hypokalemia can also occur as a result of primary renal potassium wasting in association with other solute transport abnormalities, such as Fanconi's syndrome, renal tubular acidosis, or other forms of hereditary or acquired tubulointerstitial disease. However, even with these conditions, as the GFR declines, the tendency to hypokalemia diminishes and hyperkalemia may supervene. Therefore, the use of potassium supplements and potassium-sparing diuretics should be constantly reevaluated as GFR declines.

METABOLIC ACIDOSIS

Metabolic acidosis is a common disturbance in advanced CKD. The majority of patients can still acidify the urine, but they produce less ammonia and, therefore, cannot excrete the normal quantity of protons in combination with this urinary buffer. Hyperkalemia, if present, further depresses ammonia production. The combination of hyperkalemia and hyperchloremic metabolic acidosis is often present, even at earlier stages of CKD (stages 1–3), in patients with diabetic nephropathy or in those with predominant tubulointerstitial disease or obstructive uropathy; this is a non-anion-gap metabolic acidosis. Treatment of hyperkalemia may increase renal ammonia production, improve renal generation of bicarbonate, and improve the metabolic acidosis.

With worsening renal function, the total urinary net daily acid excretion is usually limited to 30–40 mmol, and the anions of retained organic acids can then lead to an anion-gap metabolic acidosis. Thus, the non-anion-gap metabolic acidosis that can be seen in earlier stages of CKD may be complicated by the addition of an anion-gap metabolic acidosis as CKD progresses. In most patients, the metabolic acidosis is mild; the pH is rarely <7.35 and can usually be corrected with oral sodium bicarbonate supplementation. Animal and human studies have suggested that even modest degrees of metabolic acidosis may be associated with the development of protein catabolism. Alkali supplementation may attenuate the catabolic state and possibly slow CKD progression and accordingly is recommended when the serum bicarbonate concentration falls below 20–23 mmol/L. The concomitant sodium load mandates careful attention to volume status and the potential need for diuretic agents.

TREATMENT: FLUID, ELECTROLYTE, AND ACID-BASE DISORDERS

Adjustments in the dietary intake of salt and use of loop diuretics, occasionally in combination with metolazone, may be needed to maintain euvolemia. In contrast, overzealous salt restriction or diuretic use can lead to ECFV depletion and precipitate a further decline in GFR. The rare patient with salt-losing nephropathy may require a sodium-rich diet or salt supplementation. Water restriction is indicated only if there is a problem with hyponatremia. Otherwise, patients with CKD and an intact thirst mechanism may be instructed to drink fluids in a quantity that keeps them just ahead of their thirst. Intractable ECFV expansion, despite dietary salt restriction and diuretic therapy, may be an indication to start renal replacement therapy. Hyperkalemia often responds to dietary restriction of potassium, avoidance of potassium supplements (including occult sources, such as dietary salt substitutes) as well as potassium-retaining medications (especially ACE inhibitors or ARBs), or the use of kaliuretic diuretics. Kaliuretic diuretics promote urinary potassium excretion, while potassium-binding resins, such as calcium resonium or sodium polystyrene, can promote potassium loss through the GI tract and may reduce the incidence of hyperkalemia in CKD patients. Intractable hyperkalemia is an

indication (although uncommon) to consider institution of dialysis in a CKD patient. The renal tubular acidosis and subsequent anion-gap metabolic acidosis in progressive CKD will respond to alkali supplementation, typically with sodium bicarbonate. Recent studies suggest that this replacement should be considered when the serum bicarbonate concentration falls below 20–23 mmol/L to avoid the protein catabolic state seen with even mild degrees of metabolic acidosis and to slow the progression of CKD.

Disorders of Calcium and Phosphate Metabolism

The principal complications of abnormalities of calcium and phosphate metabolism in CKD occur in the skeleton and the vascular bed, with occasional severe involvement of extraosseous soft tissues. It is likely that disorders of bone turnover and disorders of vascular and soft tissue calcification are related to each other **(Fig. 280-2**).

Figure 280-2

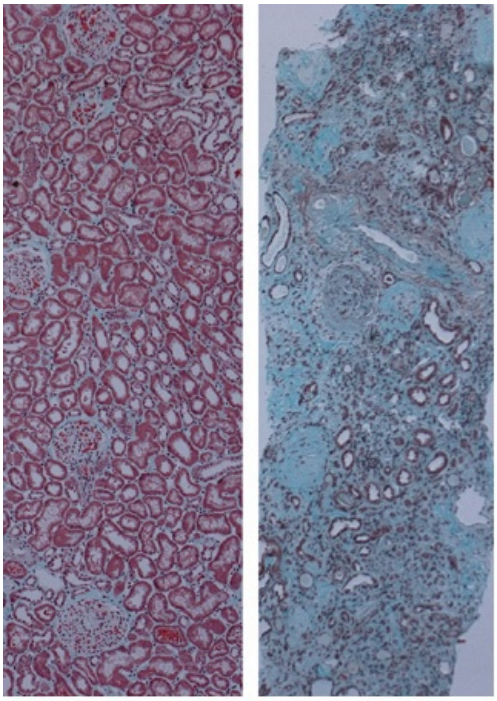

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left: Low-power photomicrograph of a normal kidney showing normal glomeruli and healthy tubulointerstitium without fibrosis. Right: Low-power photomicrograph of chronic kidney disease with sclerosis of many glomeruli and severe tubulointerstitial fibrosis (*Masson trichrome, x 40 magnification. Slides courtesy of the late Dr. Andrew Herzenberg.*)

BONE MANIFESTATIONS OF CKD

The major disorders of bone disease can be classified into those associated with high bone turnover with increased PTH levels (including osteitis fibrosa cystica, the classic lesion of secondary hyperparathyroidism) and low bone turnover with low or normal PTH levels (adynamic bone disease and osteomalacia).

The pathophysiology of secondary hyperparathyroidism and the consequent high-turnover bone disease is related to abnormal mineral metabolism through the following events: (1) declining GFR leads to reduced excretion of phosphate and, thus, phosphate retention; (2) the retained phosphate stimulates increased synthesis of PTH and growth of parathyroid gland mass; and (3) decreased levels of ionized calcium, resulting from diminished calcitriol production by the failing kidney as well as phosphate retention, also stimulate PTH production. Low calcitriol levels contribute to hyperparathyroidism, both by leading to hypocalcemia and also by a direct effect on PTH gene transcription. These changes start to occur when the GFR falls below 60 mL/min.

Fibroblast growth factor 23 (FGF-23) is part of a family of phosphatonins that promotes renal phosphate excretion. Recent studies have shown that levels of this hormone, secreted by osteocytes, increases early in the course of CKD. It may defend normal serum phosphorus in at least three ways: (1) increased renal phosphate excretion; (2) stimulation of PTH, which also increases renal phosphate excretion; and (3) suppression of the formation of 1,25(OH)2D3, leading to diminished phosphorus absorption from the gastrointestinal tract. Interestingly, high levels of FGF-23 are also an independent risk factor for left ventricular hypertrophy and mortality in dialysis patients. Moreover, elevated levels of

FGF-23 may indicate the need for therapeutic intervention (e.g., phosphate restriction), even when serum phosphate levels are within the normal range.

Hyperparathyroidism stimulates bone turnover and leads to *osteitis fibrosa cystica*. Bone histology shows abnormal osteoid, bone and bone marrow fibrosis, and in advanced stages, the formation of bone cysts, sometimes with hemorrhagic elements so that they appear brown in color, hence the term *brown tumor*. Clinical manifestations of severe hyperparathyroidism include bone pain and fragility, brown tumors, compression syndromes, tumors, and erythropoietin resistance in part related to the bone marrow fibrosis. Furthermore, PTH itself is considered a uremic toxin, and high levels are associated with muscle weakness, fibrosis of cardiac muscle, and nonspecific constitutional symptoms.

Low-turnover bone disease can be grouped into two categories—adynamic bone disease and osteomalacia. In the latter condition, there is accumulation of unmineralized bone matrix that may be caused by a number of processes, including vitamin D deficiency, metabolic acidosis, and in the past aluminum deposition. Adynamic bone disease is increasing in prevalence, especially among diabetics and the elderly. It is characterized by reduced bone volume and mineralization and may result from excessive suppression of PTH production, chronic inflammation, or both. Suppression of PTH can result from the use of vitamin D preparations or from excessive calcium exposure in the form of calcium-containing phosphate binders or high-calcium dialysis solutions. Complications of adynamic bone disease include an increased incidence of fracture and bone pain and an association with increased vascular and cardiac calcification.

CALCIUM, PHOSPHORUS, AND THE CARDIOVASCULAR SYSTEM

Recent epidemiologic evidence has shown a strong association between hyperphosphatemia and increased cardiovascular mortality rate in patients with stage 5 CKD and even in patients with earlier stages of CKD. Hyperphosphatemia and hypercalcemia are associated with increased vascular calcification, but it is unclear whether the excessive mortality rate is mediated by this mechanism. Studies using CT and electron-beam CT scanning show that CKD patients have calcification of the media in coronary arteries and even heart valves that appear to be orders of magnitude greater than that in patients without renal disease. The magnitude of the calcification is proportional to age and hyperphosphatemia and is also associated with low PTH levels and low bone turnover. It is possible that in patients with advanced kidney disease, ingested calcium cannot be deposited in bones with low turnover and, therefore, is deposited at extraosseous sites, such as the vascular bed and soft tissues. It is interesting in this regard that there is also an association between osteoporosis and vascular calcification in the general population. Finally, there is recent evidence indicating that hyperphosphatemia can induce a change in gene expression in vascular cells to an osteoblast-like profile, leading to vascular calcification and even ossification.

OTHER COMPLICATIONS OF ABNORMAL MINERAL METABOLISM

Calciphylaxis (calcific uremic arteriolopathy) is a devastating condition seen almost exclusively in patients with advanced CKD. It is heralded by livedo reticularis and advances to patches of ischemic necrosis, especially on the legs, thighs, abdomen, and breasts (Fig. 280-3). Pathologically, there is evidence of vascular occlusion in association with extensive vascular and soft tissue calcification. It appears that this condition is increasing in incidence. Originally it was ascribed to severe abnormalities in calcium and phosphorus control in dialysis patients, usually associated with advanced hyperparathyroidism. However, more recently, calciphylaxis has been seen with increasing frequency in the absence of severe hyperparathyroidism. Other etiologies have been suggested, including the increased use of oral calcium as a phosphate binder. Warfarin is commonly used in hemodialysis patients, and one of the effects of warfarin therapy is to decrease the vitamin K–dependent regeneration of matrix GLA protein. This latter protein is important in preventing vascular calcification. Thus, warfarin treatment is considered a risk factor for calciphylaxis, and if a patient develops this syndrome, this medication should be discontinued and replaced with alternative forms of anticoagulation.

Figure 280-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Calciphylaxis. This peritoneal dialysis patient was on chronic warfarin therapy for prophylactic anticoagulation for a mechanical heart valve. She slept with the dialysis catheter pressed between her legs. A small abrasion was followed by progressive skin necrosis along the catheter tract on her inner thighs. Despite treatment with hyperbaric oxygen, intravenous thiosulfate, and discontinuation of warfarin, she succumbed to systemic complications of the necrotic process.

TREATMENT: DISORDERS OF CALCIUM AND PHOSPHATE METABOLISM

The optimal management of secondary hyperparathyroidism and osteitis fibrosa is prevention. Once the parathyroid gland mass is very large, it is difficult to control the disease. Careful attention should be paid to the plasma phosphate concentration in CKD patients, who should be counseled on a low-phosphate diet as well as the appropriate use of phosphate-binding agents. These are agents that are taken with meals and complex the dietary phosphate to limit its GI absorption. Examples of phosphate binders are calcium acetate and calcium carbonate. A major side effect of calcium-based phosphate binders is total-body calcium accumulation and hypercalcemia, especially in patients with low-turnover bone disease. Sevelamer and lanthanum are non-calcium-containing polymers that also function as phosphate binders; they do not predispose CKD patients to hypercalcemia and may attenuate calcium deposition in the vascular bed.

Calcitriol exerts a direct suppressive effect on PTH secretion and also indirectly suppresses PTH secretion by raising the concentration of ionized calcium. However, calcitriol therapy may result in hypercalcemia and/or hyperphosphatemia through increased GI absorption of these minerals. Certain analogues of calcitriol are available (e.g., paricalcitol) that suppress PTH secretion with less attendant hypercalcemia.

Recognition of the role of the extracellular calcium-sensing receptor has led to the development of calcimimetic agents that enhance the sensitivity of the parathyroid cell to the suppressive effect of calcium. This class of drug, which includes cinacalcet, produces a dose-dependent reduction in PTH and plasma calcium concentration in some patients.

Current KDOQI guidelines recommend a target PTH level between 150 and 300 pg/mL, recognizing that very low PTH levels are associated with adynamic bone disease and possible consequences of fracture and ectopic calcification.

Cardiovascular Abnormalities

Cardiovascular disease is the leading cause of morbidity and mortality in patients at every stage of CKD. The incremental risk of cardiovascular disease in those with CKD compared to the age- and sex-matched general population ranges from 10- to 200-fold, depending on the stage of CKD. Between 30 and 45% of patients reaching stage 5 CKD already have advanced cardiovascular complications. As a result, most patients with CKD succumb to cardiovascular disease **(Fig. 280-4)** before ever reaching stage 5 CKD. Thus, the focus of patient care in earlier CKD stages should be directed to prevention of cardiovascular complications.

Figure 280-4

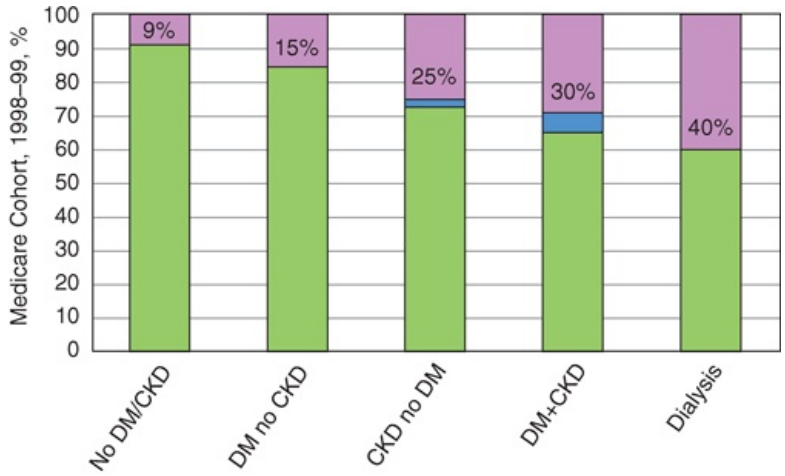

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

U.S. Renal Data System showing increased likelihood of dying rather than starting dialysis or reaching stage 5 chronic kidney disease (CKD). 1, Death; 2, ESRD/D; 3, event-free. DM; diabetes mellitus. (*Adapted from RN Foley et al: J Am Soc Nephrol 16:489-95, 2005.*)

ISCHEMIC VASCULAR DISEASE

The presence of any stage of CKD is a major risk factor for ischemic cardiovascular disease, including occlusive coronary, cerebrovascular, and peripheral vascular disease. The increased prevalence of vascular disease in CKD patients derives from both traditional ("classic") and nontraditional (CKD-related) risk factors. Traditional risk factors include hypertension, hypervolemia, dyslipidemia, sympathetic overactivity, and hyperhomocysteinemia. The CKD-related risk factors comprise anemia, hyperphosphatemia, hyperparathyroidism, sleep apnea, and generalized inflammation. The inflammatory state associated with a reduction in kidney function is reflected in increased circulating acute-phase reactants, such as inflammatory cytokines and C-reactive protein, with a corresponding fall in the "negative acute-phase reactants," such as serum albumin and fetuin. The inflammatory state appears to accelerate vascular occlusive disease, and low levels of fetuin may permit more rapid vascular calcification, especially in the face of hyperphosphatemia. Other abnormalities seen in CKD may augment myocardial ischemia, including left ventricular hypertrophy and microvascular disease. Coronary reserve, defined as the increase in coronary blood flow in response to greater demand, is also attenuated. There is diminished availability of nitric oxide because of increased concentration of asymmetric dimethyl-1-arginine and increased scavenging by reactive oxygen species. In addition, hemodialysis, with its attendant episodes of hypotension and hypovolemia, may further aggravate coronary ischemia. Interestingly, however, the largest increment in cardiovascular mortality rate in dialysis patients is not necessarily directly associated with documented acute myocardial infarction but, instead, presents with congestive heart failure and all of its manifestations, including sudden death.

Cardiac troponin levels are frequently elevated in CKD without evidence of acute ischemia. The elevation complicates the diagnosis of acute myocardial infarction in this population. Serial measurements may be needed, and if the level is unchanged, it is possible that there is no acute myocardial ischemia. On the other hand, an increase upon subsequent testing suggests cardiac injury. Therefore, the trend in levels over the hours after presentation may be more informative than a single, elevated level. Interestingly, consistently elevated levels are an independent prognostic factor for adverse cardiovascular events in this population.

HEART FAILURE

Abnormal cardiac function secondary to myocardial ischemia, left ventricular hypertrophy, and frank cardiomyopathy, in combination with the salt and water retention that can be seen with CKD, often results in heart failure or even episodes of pulmonary edema. Heart failure can be a consequence of diastolic or systolic dysfunction, or both. A form of "low-pressure" pulmonary edema can also occur in advanced CKD, manifesting as shortness of breath and a "bat wing" distribution of alveolar edema fluid on the chest x-ray. This finding can occur even in the absence of ECFV overload and is associated with normal or mildly elevated pulmonary capillary wedge pressure. This process has been ascribed to increased permeability of alveolar capillary membranes as a manifestation of the uremic state, and it responds to dialysis. Other CKD-related risk factors, including anemia and sleep apnea, may contribute to the risk of heart failure.

HYPERTENSION AND LEFT VENTRICULAR HYPERTROPHY

Hypertension is one of the most common complications of CKD. It usually develops early during the course of CKD and is associated with adverse outcomes, including the development of ventricular hypertrophy and a more rapid loss of renal function. Many studies have shown a relationship between

the level of blood pressure and the rate of progression of diabetic and non-diabetic kidney disease. Left ventricular hypertrophy and dilated cardiomyopathy are among the strongest risk factors for cardiovascular morbidity and mortality in patients with CKD and are thought to be related primarily, but not exclusively, to prolonged hypertension and ECFV overload. In addition, anemia and the placement of an arteriovenous fistula for hemodialysis can generate a high cardiac output state and consequent heart failure.

The absence of hypertension may signify the presence of a salt-wasting form of renal disease, the effect of antihypertensive therapy, or volume depletion or may signify poor left ventricular function. Indeed, in epidemiologic studies of dialysis patients, low blood pressure actually carries a worse prognosis than does high blood pressure. This mechanism, in part, accounts for the "reverse causation" seen in dialysis patients, wherein the presence of traditional risk factors, such as hypertension, hyperlipidemia, and obesity, appear to portend a better prognosis. Importantly, these observations derive from cross-sectional studies of late-stage CKD patients and should not be interpreted to discourage appropriate management of these risk factors in CKD patients, especially at early stages. In contrast to the general population, it is possible that in late-stage CKD, low blood pressure, reduced body mass index, and hypolipidemia indicate the presence of an advanced malnutrition-inflammation state, with poor prognosis.

The use of exogenous erythropoiesis-stimulating agents can increase blood pressure and the requirement for antihypertensive drugs. Chronic ECFV overload is also a contributor to hypertension, and improvement in blood pressure can often be seen with the use of dietary sodium restriction, diuretics, and fluid removal with dialysis. Nevertheless, because of activation of the renin-angiotensin-aldosterone axis and other disturbances in the balance of vasoconstrictors and vasodilators, some patients remain hypertensive despite careful attention to ECFV status.

TREATMENT: CARDIOVASCULAR ABNORMALITIES

Management of Hypertension

There are two overall goals of therapy for hypertension in these patients: to slow the progression of the kidney disease itself, and to prevent the extrarenal complications of high blood pressure, such as cardiovascular disease and stroke. In all patients with CKD, blood pressure should be controlled to levels recommended by national guideline panels. In CKD patients with diabetes or proteinuria >1 g per 24 h, blood pressure should be reduced to 125/75, if achievable without prohibitive adverse effects. Salt restriction should be the first line of therapy. When volume management alone is not sufficient, the choice of antihypertensive agent is similar to that in the general population. The ACE inhibitors and ARBs slow the rate of decline of kidney function, but occasionally can precipitate an episode of acute kidney injury, especially when used in patients with ischemic renovascular disease. The use of ACE inhibitors and ARBs may also be complicated by the development of hyperkalemia. Often the concomitant use of a kaliuretic diuretic, such as metolazone, can improve potassium excretion in addition to improving blood pressure control. Potassium-sparing diuretics should be used with caution or avoided altogether in most patients. Indeed, increased usage of spironolactone for the management of heart failure has resulted in an increase in serious hyperkalemic events in patients with reduced kidney function.

Management of Cardiovascular Disease

There are many strategies available to treat the traditional and nontraditional risk factors in CKD patients. While these have been proved effective in the general population, there is little evidence for their benefit in patients with advanced CKD, especially those on dialysis. Certainly hypertension, elevated serum levels of homocysteine, and dyslipidemia promote atherosclerotic disease and are treatable complications of CKD. Renal disease complicated by nephrotic syndrome is associated with a very atherogenic lipid profile and hypercoagulability, which increases the risk of occlusive vascular disease. Since diabetes mellitus and hypertension are the two most frequent causes of advanced CKD, it is not surprising that cardiovascular disease is the most frequent cause of death in dialysis patients. The role of "inflammation" may be quantitatively more important in patients with kidney disease, and the treatment of more traditional risk factors may result in only modest success. However, modulation of traditional risk factors may be the only weapon in the therapeutic armamentarium for these patients until the nature of inflammation in CKD and its treatment are better understood.

Lifestyle changes, including regular exercise, should be advocated but are not often implemented. Hyperlipidemia in patients with CKD should be managed according to national guidelines. If dietary measures are not sufficient, preferred lipid-lowering medications, such as statins, should be used. Again, the use of these agents has not been of proven benefit for patients with advanced CKD.

PERICARDIAL DISEASE

Chest pain with respiratory accentuation, accompanied by a friction rub, is diagnostic of pericarditis. Classic electrocardiographic abnormalities include PR-interval depression and diffuse ST-segment elevation. Pericarditis can be accompanied by pericardial effusion that is seen on echocardiography and can rarely lead to tamponade. However, the pericardial effusion can be asymptomatic, and pericarditis can be seen without significant effusion.

Pericarditis is observed in advanced uremia, and with the advent of timely initiation of dialysis, is not as common as it once was. It is now more often observed in underdialyzed, nonadherent patients than in those starting dialysis.

Treatment: Pericardial Disease

Uremic pericarditis is an absolute indication for the urgent initiation of dialysis or for intensification of the dialysis prescription in those already receiving dialysis. Because of the propensity to hemorrhage in pericardial fluid, hemodialysis should be performed without heparin. A pericardial drainage procedure should be considered in patients with recurrent pericardial effusion, especially with echocardiographic signs of impending tamponade. Nonuremic causes of pericarditis and effusion include viral, malignant, tuberculous, and autoimmune etiologies. It may also be seen after myocardial infarction and as a complication of treatment with the antihypertensive drug minoxidil.

Hematologic Abnormalities

ANEMIA

A normocytic, normochromic anemia is observed as early as stage 3 CKD and is almost universal by stage 4. The primary cause in patients with CKD is insufficient production of erythropoietin (EPO) by the diseased kidneys. Additional factors include iron deficiency, acute and chronic inflammation with impaired iron utilization ("anemia of chronic disease"), severe hyperparathyroidism with consequent bone marrow fibrosis, and shortened red cell survival in the uremic environment. In addition, comorbid conditions such as hemoglobinopathy can worsen the anemia **(Table 280-5)**.

Table 280-5 Causes of Anemia in CKD

Relative deficiency of erythropoietin
Diminished red blood cell survival
Bleeding diathesis
Iron deficiency
Hyperparathyroidism/bone marrow fibrosis
"Chronic inflammation"
Folate or vitamin B12 deficiency
Hemoglobinopathy
Comorbid conditions: hypo/hyperthyroidism, pregnancy, HIV-associated disease, autoimmune disease, immunosuppressive drugs

The anemia of CKD is associated with a number of adverse pathophysiologic consequences, including decreased tissue oxygen delivery and utilization, increased cardiac output, ventricular dilation, and ventricular hypertrophy. Clinical manifestations include fatigue and diminished exercise tolerance, angina, heart failure, decreased cognition and mental acuity, and impaired host defense against infection. In addition, anemia may play a role in growth retardation in children with CKD. While many studies in CKD patients have found that anemia and resistance to exogenous EPO are associated with a poor prognosis, the relative contribution to a poor outcome of the low hematocrit itself, versus inflammation as a cause of the anemia, remains unclear.

Treatment: Anemia

The availability of recombinant human EPO and modified EPO products, such as darbepoetin-alpha, has been one of the most significant advances in the care of renal patients since the introduction of dialysis and renal transplantation. The routine use of these products has obviated the need for regular blood transfusions in severely anemic CKD patients, thus dramatically reducing the incidence of transfusion-associated infections and iron overload. Frequent blood transfusions in dialysis patients also leads to the development of allo-antibodies that could sensitize the patient to donor kidney antigens and make renal transplantation more problematic.

Adequate bone marrow iron stores should be available before treatment with EPO is initiated. Iron supplementation is usually essential to ensure an optimal response to EPO in patients with CKD because the demand for iron by the marrow frequently exceeds the amount of iron that is immediately available for erythropoiesis (measured by percent transferrin saturation), as well as the amount in iron stores (measured by serum ferritin). For the CKD patient not yet on dialysis or the patient treated with peritoneal dialysis, oral iron supplementation should be attempted. If there is GI intolerance, the patient may have to undergo IV iron infusion. For patients on hemodialysis, IV iron can be administered during dialysis. In addition to iron, an adequate supply of other major substrates and cofactors for red cell production must be ensured, including vitamin B12 and folate. Anemia resistant to recommended doses of EPO in the face of adequate iron stores may be due to some combination of the following: acute or chronic inflammation, inadequate dialysis, severe hyperparathyroidism, chronic blood loss or

hemolysis, chronic infection, or malignancy. Patients with a hemoglobinopathy, such as sickle cell disease or thalassemia, will usually not respond normally to exogenous EPO; however, an increase in hemoglobin concentration is still seen in many of these patients. Blood transfusions increase the risk of hepatitis, iron overload, and transplant sensitization; they should be avoided unless the anemia fails to respond to EPO and the patient is symptomatic.

At least three randomized, controlled trials of erythropoietin-stimulating agents in CKD have failed to show an improvement in cardiovascular outcomes with this therapy. Indeed, there has been an indication that the use of EPO in CKD may be associated with an increased risk of stroke in those with type 2 diabetes, an increase in thromboembolic events, and perhaps a faster progression to the need for dialysis. Therefore, any benefit in terms of improvement of anemic symptoms needs to be balanced against the potential cardiovascular risk of EPO therapy in CKD. While further studies are needed, it is quite clear that complete normalization of the hemoglobin concentration has not been demonstrated to be of incremental benefit to CKD patients. Current practice is to target a hemoglobin concentration of 100–115 g/L.

ABNORMAL HEMOSTASIS

Patients with later stages of CKD may have a prolonged bleeding time, decreased activity of platelet factor III, abnormal platelet aggregation and adhesiveness, and impaired prothrombin consumption. Clinical manifestations include an increased tendency to bleeding and bruising, prolonged bleeding from surgical incisions, menorrhagia, and spontaneous GI bleeding. Interestingly, CKD patients also have a greater susceptibility to thromboembolism, especially if they have renal disease that includes nephrotic-range proteinuria. The latter condition results in hypoalbuminemia and renal loss of anticoagulant factors, which can lead to a thrombophilic state.

Treatment: Abnormal Hemostasis

Abnormal bleeding time and coagulopathy in patients with renal failure may be reversed temporarily with desmopressin (DDAVP), cryoprecipitate, IV conjugated estrogens, blood transfusions, and EPO therapy. Optimal dialysis will usually correct a prolonged bleeding time.

Given the coexistence of bleeding disorders and a propensity to thrombosis that is unique in the CKD patient, decisions about anticoagulation that have a favorable risk-benefit profile in the general population may not be applicable to the patient with advanced CKD. One example is warfarin anticoagulation for atrial fibrillation: the decision to anticoagulate should be made on an individual basis in the CKD patient, as there appears to be a greater risk of bleeding complications.

Certain anticoagulants, such as fractionated low-molecular-weight heparin, may need to be avoided or dose adjusted in these patients, with monitoring of factor Xa activity where available. It is often more prudent to use conventional high-molecular-weight heparin, titrated to the measured partial thromboplastin time, in hospitalized patients requiring an alternative to warfarin anticoagulation.

Neuromuscular Abnormalities

Central nervous system (CNS), peripheral, and autonomic neuropathy as well as abnormalities in muscle structure and function are all well-recognized complications of CKD. Retained nitrogenous metabolites and middle molecules, including PTH, contribute to the pathophysiology of neuromuscular abnormalities. Subtle clinical manifestations of uremic neuromuscular disease usually become evident at stage 3 CKD. Early manifestations of CNS complications include mild disturbances in memory and concentration and sleep disturbance. Neuromuscular irritability, including hiccups, cramps, and fasciculations or twitching of muscles, becomes evident at later stages. In advanced untreated kidney failure, asterixis, myoclonus, seizures, and coma can be seen.

Peripheral neuropathy usually becomes clinically evident after the patient reaches stage 4 CKD, although electrophysiologic and histologic evidence occurs earlier. Initially, sensory nerves are involved more than motor, lower extremities more than upper, and distal parts of the extremities more than proximal. The "restless leg syndrome" is characterized by ill-defined sensations of sometimes debilitating discomfort in the legs and feet relieved by frequent leg movement. If dialysis is not instituted soon after onset of sensory abnormalities, motor involvement follows, including muscle weakness. Evidence of peripheral neuropathy without another cause (e.g., diabetes mellitus) is a firm indication for starting renal replacement therapy. Many of the complications described above will resolve with dialysis, although subtle nonspecific abnormalities may persist. Successful renal transplantation may reverse residual neurologic changes.

Gastrointestinal and Nutritional Abnormalities

Uremic fetor, a urine-like odor on the breath, derives from the breakdown of urea to ammonia in saliva and is often associated with an unpleasant metallic taste (dysgeusia). Gastritis, peptic disease, and mucosal ulcerations at any level of the GI tract occur in uremic patients and can lead to abdominal pain, nausea, vomiting, and GI bleeding. These patients are also prone to constipation, which can be worsened by the administration of calcium and iron supplements. The retention of uremic toxins also leads to anorexia, nausea, and vomiting.

Protein restriction may be useful to decrease nausea and vomiting; however, it may put the patient at

risk for malnutrition and should be carried out, if possible, in consultation with a registered dietitian specializing in the management of CKD patients. Protein-energy malnutrition, a consequence of low protein and caloric intake, is common in advanced CKD and is often an indication for initiation of renal replacement therapy. In addition to diminished intake, these patients are resistant to the anabolic actions of insulin and other hormones and growth factors. Metabolic acidosis and the activation of inflammatory cytokines can promote protein catabolism. Assessment for protein-energy malnutrition should begin at stage 3 CKD. A number of indices are useful in this assessment and include dietary history, including food diary and subjective global assessment; edema-free body weight; and measurement of urinary protein nitrogen appearance. Dual-energy x-ray absorptiometry is now widely used to estimate lean body mass versus ECFV. Adjunctive tools include clinical signs, such as skin-fold thickness, mid-arm muscle circumference, and additional laboratory tests such as serum pre-albumin and cholesterol levels. Nutritional guidelines for patients with CKD are summarized in the "Treatment" section below.

Endocrine-Metabolic Disturbances

Glucose metabolism is impaired in CKD, as evidenced by a slowing of the rate at which blood glucose levels decline after a glucose load. However, fasting blood glucose is usually normal or only slightly elevated, and the mild glucose intolerance does not require specific therapy. Because the kidney contributes to insulin removal from the circulation, plasma levels of insulin are slightly to moderately elevated in most uremic patients, both in the fasting and postprandial states. Because of this diminished renal degradation of insulin, patients on insulin therapy may need progressive reduction in dose as their renal function worsens. Many hypoglycemic agents require dose reduction in renal failure, and some, such as metformin, are contraindicated when the GFR is less than half of normal.

In women with CKD, estrogen levels are low, and menstrual abnormalities and inability to carry pregnancies to term are common. When the GFR has declined to ~40 mL/min, pregnancy is associated with a high rate of spontaneous abortion, with only ~20% of pregnancies leading to live births, and pregnancy may hasten the progression of the kidney disease itself. Women with CKD who are contemplating pregnancy should consult first with a nephrologist in conjunction with an obstetrician specializing in high-risk pregnancy. Men with CKD have reduced plasma testosterone levels, and sexual dysfunction and oligospermia may supervene. Sexual maturation may be delayed or impaired in adolescent children with CKD, even among those treated with dialysis. Many of these abnormalities improve or reverse with intensive dialysis or most importantly with successful renal transplantation.

Dermatologic Abnormalities

Abnormalities of the skin are prevalent in progressive CKD. Pruritus is quite common and one of the most vexing manifestations of the uremic state. In advanced CKD, even on dialysis, patients may become more pigmented, and this is felt to reflect the deposition of retained pigmented metabolites, or *urochromes*. Although many of the cutaneous abnormalities improve with dialysis, pruritus is often tenacious. The first lines of management are to rule out unrelated skin disorders, such as scabies, and to treat hyperphosphatemia, which can cause itch. EPO therapy was initially reported to improve uremic pruritus, although that is not always the case. Local moisturizers, mild topical glucocorticoids, oral antihistamines, and ultraviolet radiation have been reported to be helpful.

A skin condition unique to CKD patients called *nephrogenic fibrosing dermopathy* consists of progressive subcutaneous induration, especially on the arms and legs. The condition is similar to scleromyxedema and is seen in patients with CKD who have been exposed to the magnetic resonance contrast agent gadolinium. Current recommendations are that patients with CKD stage 2 (GFR 30–59 mL/min) should minimize exposure to gadolinium, and those with CKD stages 3–5 (GFR < 30 mL/min) should avoid the use of gadolinium agents unless it is medically necessary. Concomitant liver disease appears to be a risk factor. However, no patient should be denied an imaging investigation that is critical to clinical management, and under such circumstances, rapid removal of gadolinium by hemodialysis (even in patients not yet receiving renal replacement therapy) shortly after the imaging procedure may mitigate this sometimes devastating complication

EVALUATION AND MANAGEMENT OF PATIENTS WITH CKD

Initial Approach

HISTORY AND PHYSICAL EXAMINATION

Symptoms and overt signs of kidney disease are often subtle or absent until renal failure supervenes. Thus, the diagnosis of kidney disease often surprises patients and may be a cause of skepticism and denial. Particular aspects of the history that are germane to renal disease include a history of hypertension (which can cause CKD or more commonly be a consequence of CKD), diabetes mellitus, abnormal urinalyses, and problems with pregnancy such as preeclampsia or early pregnancy loss. A careful drug history should be elicited: patients may not volunteer use of analgesics, for example. Other drugs to consider include nonsteroidal anti-inflammatory agents, gold, penicillamine, antimicrobials, chemotherapeutic agents, antiretroviral agents, proton pump inhibitors, phosphate-containing bowel cathartics, and lithium, as well as prior exposure to medical imaging radiocontrast

agents. In evaluating the uremic syndrome, questions about appetite, weight loss, nausea, hiccups, peripheral edema, muscle cramps, pruritus, and restless legs are especially helpful. A careful family history of kidney disease, together with assessment of manifestations in other organ systems such as auditory, visual, integumentary and others, may lead to the diagnosis of a heritable form of CKD (e.g., Alport or Fabry syndrome, cystinuria, among others) or shared environmental exposure to nephrotoxic agents (e.g., heavy metals, aristolochic acid). It should be noted that clustering of CKD, sometimes of different etiologies, is often observed within families.

The physical examination should focus on blood pressure and target organ damage from hypertension. Thus, funduscopy and precordial examination (left ventricular heave, a fourth heart sound) should be carried out. Funduscopy is important in the diabetic patient, as it may show evidence of diabetic retinopathy, which is associated with nephropathy. Other physical examination manifestations of CKD include edema and sensory polyneuropathy. The finding of asterixis or a pericardial friction rub not attributable to other causes usually signifies the presence of the uremic syndrome.

LABORATORY INVESTIGATION

Laboratory studies should focus on a search for clues to an underlying causative or aggravating disease process and on the degree of renal damage and its consequences. Serum and urine protein electrophoresis, looking for multiple myeloma, should be obtained in all patients >35 years with unexplained CKD, especially if there is associated anemia and elevated, or even inappropriately normal, serum calcium concentration in the face of renal insufficiency. In the presence of glomerulonephritis, autoimmune diseases such as lupus and underlying infectious etiologies such as hepatitis B and C and HIV should be assessed. Serial measurements of renal function should be obtained to determine the pace of renal deterioration and ensure that the disease is truly chronic rather than acute or subacute and hence potentially reversible. Serum concentrations of calcium, phosphorus, vitamin D, and PTH should be measured to evaluate metabolic bone disease. Hemoglobin concentration, iron, B_{12}, and folate should also be evaluated. A 24-h urine collection may be helpful, as protein excretion >300 mg may be an indication for therapy with ACE inhibitors or ARBs.

IMAGING STUDIES

The most useful imaging study is a renal ultrasound, which can verify the presence of two kidneys, determine if they are symmetric, provide an estimate of kidney size, and rule out renal masses and evidence of obstruction. Since it takes time for kidneys to shrink as a result of chronic disease, the finding of bilaterally small kidneys supports the diagnosis of CKD of long-standing duration, with an irreversible component of scarring. If the kidney size is normal, it is possible that the renal disease is acute or subacute. The exceptions are diabetic nephropathy (where kidney size is increased at the onset of diabetic nephropathy before CKD with loss of GFR supervenes), amyloidosis, and HIV nephropathy, where kidney size may be normal in the face of CKD. Polycystic kidney disease that has reached some degree of renal failure will almost always present with enlarged kidneys with multiple cysts (Chap. 284). A discrepancy >1 cm in kidney length suggests either a unilateral developmental abnormality or disease process or renovascular disease with arterial insufficiency affecting one kidney more than the other. The diagnosis of renovascular disease can be undertaken with different techniques, including Doppler sonography, nuclear medicine studies, or CT or MRI studies. If there is a suspicion of reflux nephropathy (recurrent childhood urinary tract infection, asymmetric renal size with scars on the renal poles), a voiding cystogram may be indicated. However, in most cases by the time the patient has CKD, the reflux has resolved, and even if still present, repair does not improve renal function. Radiographic contrast imaging studies are not particularly helpful in the investigation of CKD. Intravenous or intraarterial dye should be avoided where possible in the CKD patient, especially with diabetic nephropathy, because of the risk of radiographic contrast dye–induced renal failure. When unavoidable, appropriate precautionary measures include avoidance of hypovolemia at the time of contrast exposure, minimization of the dye load, and choice of radiographic contrast preparations with the least nephrotoxic potential. Additional measures thought to attenuate contrast-induced worsening of renal function include judicious administration of sodium bicarbonate–containing solutions and *N*-acetyl-cysteine.

RENAL BIOPSY

In the patient with bilaterally small kidneys, renal biopsy is not advised because (1) it is technically difficult and has a greater likelihood of causing bleeding and other adverse consequences, (2) there is usually so much scarring that the underlying disease may not be apparent, and (3) the window of opportunity to render disease-specific therapy has passed. Other contraindications to renal biopsy include uncontrolled hypertension, active urinary tract infection, bleeding diathesis (including ongoing anticoagulation), and severe obesity. Ultrasound-guided percutaneous biopsy is the favored approach, but a surgical or laparoscopic approach can be considered, especially in the patient with a single kidney where direct visualization and control of bleeding are crucial. In the CKD patient in whom a kidney biopsy is indicated (e.g., suspicion of a concomitant or superimposed active process such as interstitial nephritis or in the face of accelerated loss of GFR), the bleeding time should be measured, and, if increased, desmopressin should be administered immediately prior to the procedure.

A brief run of hemodialysis (without heparin) may also be considered prior to renal biopsy to normalize

the bleeding time.

Establishing the Diagnosis and Etiology of CKD

The most important initial diagnostic step in the evaluation of a patient presenting with elevated serum creatinine is to distinguish newly diagnosed CKD from acute or subacute renal failure because the latter two conditions may respond to therapy specific to the disease. Previous measurements of serum creatinine concentration are particularly helpful in this regard. Normal values from recent months or even years suggest that the current extent of renal dysfunction could be more acute, and hence reversible, than might otherwise be appreciated. In contrast, elevated serum creatinine concentration in the past suggests that the renal disease represents the progression of a chronic process. Even if there is evidence of chronicity, there is the possibility of a superimposed acute process (e.g., ECFV depletion, urinary infection or obstruction, or nephrotoxin exposure) supervening on the chronic condition. If the history suggests multiple systemic manifestations of recent onset (e.g., fever, polyarthritis, and rash), it should be assumed that renal insufficiency is part of the acute process.

Some of the laboratory tests and imaging studies outlined above can be helpful. Evidence of metabolic bone disease with hyperphosphatemia, hypocalcemia, and elevated PTH and bone alkaline phosphatase levels suggests chronicity. Normochromic, normocytic anemia suggests that the process has been ongoing for some time. The finding of bilaterally reduced kidney size (<8.5 cm in all but the smallest adults) favors CKD.

While renal biopsy can usually be performed in early CKD (stages 1–3), it is not always indicated. For example, in a patient with a history of type 1 diabetes mellitus for 15–20 years with retinopathy, nephrotic-range proteinuria, and absence of hematuria, the diagnosis of diabetic nephropathy is very likely and biopsy is usually not necessary. However, if there were some other finding not typical of diabetic nephropathy, such as hematuria or white blood cell casts, or absence of diabetic retinopathy, some other disease may be present and a biopsy may be indicated. Ischemic nephropathy is usually diagnosed clinically by the presence of long-standing hypertension, evidence of ischemic disease elsewhere (e.g., cardiac or peripheral vascular disease), and the finding of nonnephrotic proteinuria in the absence of urinary blood or red cell casts. It is important to consider progressive ischemic nephropathy because a small subset of these patients may respond to revascularization procedures, although this remains controversial.

In the absence of a clinical diagnosis, renal biopsy may be the only recourse to establish an etiology in early-stage CKD. However, as noted above, once the CKD is advanced and the kidneys are small and scarred, there is little utility and significant risk in attempting to arrive at a specific diagnosis.

Treatment: Chronic Kidney Disease

Treatments aimed at specific causes of CKD are discussed elsewhere. Among others, these include optimized glucose control in diabetes mellitus, immunomodulatory agents for glomerulonephritis, and emerging specific therapies to retard cytogenesis in polycystic kidney disease. The optimal timing of both specific and nonspecific therapy is usually well before there has been a measurable decline in GFR and certainly before CKD is established **(Table 280-6)**. It is helpful to sequentially measure and plot the rate of decline of GFR in all patients. Any acceleration in the rate of decline should prompt a search for superimposed acute or subacute processes that may be reversible. These include ECFV depletion, uncontrolled hypertension, urinary tract infection, new obstructive uropathy, exposure to nephrotoxic agents [such as nonsteroidal anti-inflammatory drugs (NSAIDs) or radiographic dye], and reactivation or flare of the original disease, such as lupus or vasculitis.

Table 280-6 Clinical Action Plan

Stage	Description	GFR, mL/min per 1.73 m^2	Action*
1	Kidney damage with normal or ↑ GFR	≥90	Diagnosis and treatment, treatment of comorbid conditions, slowing progression, CVD risk reduction
2	Kidney damage with mild ↓ GFR	60–89	Estimating progression
3	Moderate ↓ GFR	30–59	Evaluating and treating complications
4	Severe ↓ GFR	15–29	Preparation for kidney replacement therapy
5	Kidney failure	<15 (or dialysis)	Kidney replacement (if uremia present)

*Includes actions from preceding stages.

Abbreviations: CVD, cardiovascular disease; GFR, glomerular filtration rate.

Source: National Kidney Foundation: Am J Kidney Dis 39:S1, 2002.

SLOWING THE PROGRESSION OF CKD

There is variation in the rate of decline of GFR among patients with CKD. However, the following interventions should be considered in an effort to stabilize or slow the decline of renal function.

Reducing Intraglomerular Hypertension and Proteinuria

Increased intraglomerular filtration pressures and glomerular hypertrophy develop as a response to loss of nephron number from different kidney diseases. This response is maladaptive, as it promotes the ongoing decline of kidney function even if the inciting process has been treated or spontaneously resolved. Control of systemic and glomerular hypertension is important in slowing the progression of CKD. Therefore, in addition to reduction of cardiovascular disease risk, antihypertensive therapy in patients with CKD also aims to slow the progression of nephron injury by reducing intraglomerular hypertension. Elevated blood pressure increases proteinuria by increasing its flux across the glomerular capillaries. Conversely, the renoprotective effect of antihypertensive medications is gauged through the consequent reduction of proteinuria. Thus, the more effective a given treatment is in lowering protein excretion, the greater the subsequent impact on protection from decline in GFR. This observation is the basis for the treatment guideline establishing 125/75 mmHg as the target blood pressure in proteinuric CKD patients.

ACE inhibitors and ARBs inhibit the angiotensin-induced vasoconstriction of the efferent arterioles of the glomerular microcirculation. This inhibition leads to a reduction in both intraglomerular filtration pressure and proteinuria. Several controlled studies have shown that these drugs are effective in slowing the progression of renal failure in patients with advanced stages of both diabetic and nondiabetic CKD. This slowing in progression of CKD is strongly associated with the proteinuria-lowering effect. In the absence of an anti-proteinuric response with either agent alone, combined treatment with both ACE inhibitors and ARBs has been considered. The combination is associated with a greater reduction in proteinuria compared to either agent alone. Insofar as reduction in proteinuria is a surrogate for improved renal outcome, the combination would appear to be advantageous. However, a recent randomized, controlled study found a greater incidence of acute renal failure and adverse cardiac events from such combination therapy. It is uncertain, therefore, whether the ACE plus ARB therapy can be advised routinely. Adverse effects from these agents include cough and angioedema with ACE inhibitors, anaphylaxis, and hyperkalemia with either class. A progressive increase in serum creatinine concentration with these agents may suggest the presence of renovascular disease within the large or small arteries. Development of these side effects may mandate the use of second-line antihypertensive agents instead of the ACE inhibitors or ARBs. Among the calcium channel blockers, diltiazem and verapamil may exhibit superior antiproteinuric and renoprotective effects compared to the dihydropyridines. At least two different categories of response can be considered: one in which progression is strongly associated with systemic and intraglomerular hypertension and proteinuria (e.g., diabetic nephropathy, glomerular diseases) and in which ACE inhibitors and ARBs are likely to be the first choice; and another in which proteinuria is mild or absent initially (e.g., adult polycystic kidney disease and other tubulointerstitial diseases), where the contribution of intraglomerular hypertension is less prominent, and other antihypertensive agents can be useful for control of systemic hypertension.

SLOWING PROGRESSION OF DIABETIC RENAL DISEASE

(See also Chap. 344) Diabetic nephropathy is now the leading cause of CKD requiring renal replacement therapy in many parts of the world, and its prevalence is increasing disproportionately in the developing world. Furthermore, the prognosis of diabetic patients on dialysis is poor, with survival comparable to many forms of cancer. Accordingly, it is mandatory to develop strategies whose aim is to prevent or slow the progression of diabetic nephropathy in these patients.

Control of Blood Glucose

Excellent glycemic control reduces the risk of kidney disease and its progression in both type 1 and type 2 diabetes mellitus. It is recommended that plasma values for preprandial glucose be kept in the 5.0–7.2 mmol/L (90–130 mg/dL) range and hemoglobin A_{1C} should be < 7%. As the GFR decreases with progressive nephropathy, the use and dose of oral hypoglycemics needs to be reevaluated. For example, chlorpropamide may be associated with prolonged hypoglycemia in patients with decreased renal function; metformin can cause lactic acidosis in the patient with renal impairment and should be discontinued when the GFR is reduced; and the thiazolidinediones (e.g., rosiglitazone, pioglitazone, and others), may increase renal salt and water absorption and aggravate volume-overload states, and contribute to adverse cardiovascular events. Finally, as renal function declines, renal degradation of administered insulin will also decline, so that less insulin may be required for glycemic control.

Control of Blood Pressure and Proteinuria

Hypertension is found in the majority of type 2 diabetic patients at diagnosis. This finding correlates with the presence of albuminuria and is a strong predictor of cardiovascular events and nephropathy. Microalbuminuria, the finding of albumin in the urine not detectable by the urine dipstick, precedes the decline in GFR and heralds renal and cardiovascular complications. Testing for microalbumin is recommended in all diabetic patients, at least annually. If the patient already has established proteinuria, then testing for microalbumin is not necessary. Antihypertensive treatment reduces

albuminuria and diminishes its progression even in normotensive diabetic patients. In addition to treatment of hypertension in general, the use of ACE inhibitors and ARBs in particular is associated with additional renoprotection. These salutary effects are mediated by reducing intraglomerular pressure and inhibition of angiotensin-driven sclerosing pathways, in part through inhibition of TGF-β-mediated pathways. Recent studies have highlighted the benefit of renin-angiotensin axis blockade on the development of retinopathy in diabetic patients, but do not necessarily show a clear-cut renal benefit of intervention at the very earliest stages, prior to onset of overt proteinuria. In any case, as noted above, the combination of ACE inhibitors with ARBs is not recommended.

Protein Restriction

While protein restriction has been advocated to reduce symptoms associated with uremia, it may also slow the rate of renal decline at earlier stages of renal disease. This concept is based on clinical and experimental evidence that protein-mediated hyperfiltration contributes to ongoing decline in renal function in many different forms of renal disease. A number of studies have shown that protein restriction may be effective in slowing the progression of CKD, especially proteinuric and diabetic renal diseases. However, the Modification of Diet in Renal Disease study was unable to demonstrate a robust benefit in delaying progression to advanced stages of CKD with dietary restriction of protein intake. Nonetheless, restriction of dietary protein intake has been recommended for CKD patients. KDOQI clinical practice guidelines include a daily protein intake of between 0.60 and 0.75 g/kg per day, depending upon patient adherence, comorbid disease, presence of proteinuria, and nutritional status. It is further advised that at least 50% of the protein intake be of high biologic value. As patients approach stage 5 CKD, spontaneous protein intake tends to decrease, and patients may enter a state of protein-energy malnutrition. In these circumstances, a protein intake of up to 0.90 g/kg per day might be recommended, again with an emphasis on proteins of high biologic value.

Sufficient energy intake is important to prevent protein-calorie malnutrition, and 35 kcal/kg is recommended. Monitoring of parameters of nutritional status must accompany the dietary intervention, using the parameters outlined above in the section on GI and nutritional abnormalities

MANAGING OTHER COMPLICATIONS OF CHRONIC KIDNEY DISEASE

Medication Dose Adjustment

Although the loading dose of most drugs is not affected by CKD because no renal elimination is used in the calculation, the maintenance doses of many drugs will need to be adjusted. For those agents in which >70% excretion is by a nonrenal route, such as hepatic elimination, dose adjustment may not be needed. Some drugs that should be avoided include metformin, meperidine, and oral hypoglycemics that are eliminated by the kidney. NSAIDs should be avoided because of the risk of further worsening of kidney function. Many antibiotics, antihypertensives, and antiarrhythmics may require a reduction in dosage or change in the dose interval. Several online web-based databases for dose adjustment of medications according to stage of CKD or estimated GFR are available (e.g., *http://www.globalrph.com/renaldosing2.htm*). Nephrotoxic medical imaging radiocontrast agents and gadolinium should be avoided or used according to strict guidelines when medically necessary as described above.

Preparation for Renal Replacement Therapy

(See also Chap. 282) Temporary relief of symptoms and signs of impending uremia, such as anorexia, nausea, vomiting, lassitude, and pruritus, may sometimes be achieved with protein restriction. However, this carries a significant risk of protein-energy malnutrition, and thus plans for more long-term management should be in place.

Maintenance dialysis and kidney transplantation has extended the lives of hundreds of thousands of patients with CKD worldwide. Clear indications for initiation of renal replacement therapy for patients with CKD include uremic pericarditis, encephalopathy, intractable muscle cramping, anorexia, and nausea not attributable to reversible causes such as peptic ulcer disease, evidence of malnutrition, and fluid and electrolyte abnormalities, principally hyperkalemia or ECF volume overload, that are refractory to other measures.

Recommendations for the optimal time for initiation of renal replacement therapy have been established by the National Kidney Foundation in their KDOQI Guidelines and are based on recent evidence demonstrating that delaying initiation of renal replacement therapy until patients are malnourished or have severe uremic complications leads to a worse prognosis on dialysis or with transplantation. Because of the interindividual variability in the severity of uremic symptoms and renal function, it is ill-advised to assign an arbitrary urea nitrogen or creatinine level to the need to start dialysis. Moreover, patients may become accustomed to chronic uremia and deny symptoms, only to find that they feel better with dialysis and realize in retrospect how poorly they were feeling before its initiation.

Previous studies suggested that starting dialysis before the onset of severe symptoms and signs of uremia were associated with prolongation of survival. This led to the concept of "healthy" start and is congruent with the philosophy that it is better to keep patients feeling well all along rather than allowing them to become ill with uremia before trying to return them to better health with dialysis or

transplantation. Although recent studies have not confirmed a clear association of early-start dialysis with improved patient survival, there is still merit in this approach. On a practical level, advanced preparation may help to avoid problems with the dialysis process itself (e.g., a poorly functioning fistula for hemodialysis or malfunctioning peritoneal dialysis catheter) and thus preempt the morbidity associated with resorting to the insertion of temporary hemodialysis access with its attendant risks of sepsis, bleeding, and thrombosis.

Patient Education

Social, psychological, and physical preparation for the transition to renal replacement therapy and the choice of the optimal initial modality are best accomplished with a gradual approach involving a multidisciplinary team. Along with conservative measures discussed in the sections above, it is important to prepare patients with an intensive educational program, explaining the likelihood and timing of initiation of renal replacement therapy and the various forms of therapy available. The more knowledgeable that patients are about hemodialysis (both in-center and home-based), peritoneal dialysis, and kidney transplantation, the easier and more appropriate will be their decisions. Patients who are provided with educational programs are more likely to choose home-based dialysis therapy. This approach is of societal benefit because home-based therapy is less expensive and is associated with improved quality of life. The educational programs should be commenced no later than stage 4 CKD so that the patient has sufficient time and cognitive function to learn the important concepts, to make informed choices, and implement preparatory measures for renal replacement therapy.

Exploration of social service support is also important. In those who may perform home dialysis or undergo preemptive renal transplantation, early education of family members for selection and preparation of a home dialysis helper, or a biologically or emotionally related potential living kidney donor, should occur long before the onset of symptomatic renal failure.

Kidney transplantation (Chap. 282) offers the best potential for complete rehabilitation, because dialysis replaces only a small fraction of the kidneys' filtration function and none of the other renal functions, including endocrine and anti-inflammatory effects. Generally, kidney transplantation follows a period of dialysis treatment, although preemptive kidney transplantation (usually from a living donor) can be carried out if it is certain that the renal failure is irreversible.

Implications for Global Health

In distinction to the natural decline and successful eradication of many devastating infectious diseases, there is rapid growth in the prevalence of metabolic and vascular disease in developing countries. Diabetes mellitus is becoming increasingly prevalent in these countries, perhaps due in part to change in dietary habits, diminished physical activity, and weight gain. Therefore, it follows that there will be a proportionate increase in vascular and renal disease. Healthcare agencies must plan for improved screening for early detection, prevention, and treatment plans in these nations and must start considering options for improved availability of renal replacement therapies.

FURTHER READINGS

Abboud H, Heinrich WL: Clinical practice: Stage IV chronic kidney disease. N Engl J Med 362:56, 2010[PMID: 20054047]

Eknoyan G et al: The burden of kidney disease: Improving global outcomes. Kidney Int 66:1310, 2004[PMID: 15458424]

Go A et al: Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 351:1296, 2004[PMID: 15385656]

Golestaneh L et al: Uremic memory: The role of acute kidney injury in long-term outcomes. Kidney Int 76:813, 2009[PMID: 19789540]

Ketteler M et al: Calcification and cardiovascular health: New insights into an old phenomenon. Hypertension 47:1027, 2006[PMID: 16618842]

Kottgen A: Genome-wide association studies in nephrology. Am J Kidney Dis 56:743, 2010[PMID: 20728256]

Kronenberg F: Emerging risk factors and markers of chronic kidney disease progression. Nat Rev Nephrol 5:677, 2009[PMID: 19935815]

Levey AS et al: CKD: Common, harmful and treatable—World Kidney Day 2007. Am J Kidney Dis 2:401, 2007[PMID: 17699441]

Meyer TW et al: Uremia. N Engl J Med 357:1316, 2007[PMID: 17898101]

National kidney foundation: Kidney Disease Outcomes Quality Initiative Clinical Practice Guidelines for Nutrition in Chronic Renal Failure. Am J Kidney Dis 35 (suppl 2): S1, 2000

———: K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, classification and stratification. Am J Kidney Dis 39 (suppl 1), 2000

Sarnak M et al: Kidney disease as a risk factor for development of cardiovascular disease: A statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention. Circulation 108:2154, 2003[PMID: 14581387]

Strong K et al: Preventing chronic disease: How many lives can we save? Lancet 366:1578, 2005[PMID: 16257345]

Taal MW et al: Predicting initiation and progression of chronic kidney disease: Developing renal risk scores. Kidney Int 70:1694, 2006[PMID: 16969387]

Tonelli: M et al: Using proteinuria and estimated glomerular filtration rate to classify risk in patients with chronic kidney disease: a cohort study. Ann Intern Med 154:12, 2011

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIALYSIS IN THE TREATMENT OF RENAL FAILURE: INTRODUCTION

Dialysis may be required for the treatment of either acute or chronic kidney disease. The use of continuous renal replacement therapies (CRRTs) and slow low-efficiency dialysis (SLED) is specific to the management of acute renal failure and is discussed in Chap. 279. These modalities are performed continuously (CRRT) or over 6–12 hours per session (SLED), in contrast to the 3–4 hours of an intermittent hemodialysis session. **Advantages and disadvantages of CRRT and SLED are discussed in Chap. 279.**

Peritoneal dialysis is rarely used in developed countries for the treatment of acute renal failure because of the increased risk of infection and (as will be discussed in more detail below) less efficient clearance per unit of time. The focus of the majority of this chapter will be on the use of peritoneal and hemodialysis for end-stage renal disease (ESRD).

With the widespread availability of dialysis, the lives of hundreds of thousands of patients with ESRD have been prolonged. In the United States alone, there are now approximately 530,000 patients with ESRD, the vast majority of whom require dialysis. The incidence rate for ESRD is 350 cases per million population per year. The incidence of ESRD is disproportionately higher in African Americans (approximately 1000 per million population per year) as compared with white Americans (275 per million population per year). In the United States, the leading cause of ESRD is diabetes mellitus, currently accounting for nearly 55% of newly diagnosed cases of ESRD. Approximately one-third (33%) of patients have ESRD that has been attributed to hypertension, although it is unclear whether in these cases hypertension is the cause or a consequence of vascular disease or other unknown causes of kidney failure. Other prevalent causes of ESRD include glomerulonephritis, polycystic kidney disease, and obstructive uropathy.

Globally, mortality rates for patients with ESRD are lowest in Europe and Japan but very high in the developing world because of the limited availability of dialysis. In the United States, the mortality rate of patients on dialysis is approximately 18–20% per year, with a 5-year survival rate of approximately 30–35%. Deaths are due mainly to cardiovascular diseases and infections (approximately 50 and 15% of deaths, respectively). Older age, male sex, nonblack race, diabetes mellitus, malnutrition, and underlying heart disease are important predictors of death.

TREATMENT OPTIONS FOR ESRD PATIENTS

Commonly accepted criteria for initiating patients on maintenance dialysis include the presence of uremic symptoms, the presence of hyperkalemia unresponsive to conservative measures, persistent extracellular volume expansion despite diuretic therapy, acidosis refractory to medical therapy, a bleeding diathesis, and a creatinine clearance or estimated glomerular filtration rate (GFR) below 10 mL/min per 1.73 m^2 (see Chap. 280 for estimating equations). Timely referral to a nephrologist for advanced planning and creation of a dialysis access, education about ESRD treatment options, and management of the complications of advanced chronic kidney disease (CKD), including hypertension, anemia, acidosis, and secondary hyperparathyroidism, is advisable. Recent data have suggested that a sizable fraction of ESRD cases result following episodes of acute renal failure, particularly among persons with underlying CKD.

In ESRD, treatment options include hemodialysis (in center or at home); peritoneal dialysis, as either continuous ambulatory peritoneal dialysis (CAPD) or continuous cyclic peritoneal dialysis (CCPD); or transplantation (Chap. 282). Although there are significant geographic variations and differences in practice patterns, hemodialysis remains the most common therapeutic modality for ESRD (>90% of patients) in the United States. In contrast to hemodialysis, peritoneal dialysis is continuous, but much less efficient, in terms of solute clearance. While no large-scale clinical trials have been completed comparing outcomes among patients randomized to either hemodialysis or peritoneal dialysis, outcomes associated with both therapies are similar in most reports, and the decision of which modality to select is often based on personal preferences and quality-of-life considerations.

HEMODIALYSIS

Hemodialysis relies on the principles of solute diffusion across a semipermeable membrane. Movement of metabolic waste products takes place down a concentration gradient from the circulation into the dialysate. The rate of diffusive transport increases in response to several factors, including the magnitude of the concentration gradient, the membrane surface area, and the mass transfer coefficient of the membrane. The latter is a function of the porosity and thickness of the membrane, the size of the solute molecule, and the conditions of flow on the two sides of the membrane. According to laws of diffusion, the larger the molecule, the slower its rate of transfer across the membrane. A small molecule, such as urea (60 Da), undergoes substantial clearance, whereas a larger molecule, such as creatinine (113 Da), is cleared less efficiently. In addition to diffusive clearance, movement of waste products from the circulation into the dialysate may occur as a result of ultrafiltration. Convective clearance occurs because of solvent drag, with solutes being swept along with water across the semipermeable dialysis membrane.

The Dialyzer

There are three essential components to hemodialysis: the dialyzer, the composition and delivery of the dialysate, and the blood delivery system **(Fig. 281-1)**. The dialyzer is a plastic chamber with the ability to perfuse blood and dialysate compartments simultaneously at very high flow rates. The surface area of modern dialysis membranes in adult patients is usually in the range of 1.5–2.0 m^2. The hollow-fiber dialyzer is the most common in use in the United States. These dialyzers are composed of bundles of capillary tubes through which blood circulates while dialysate travels on the outside of the fiber bundle.

Figure 281-1

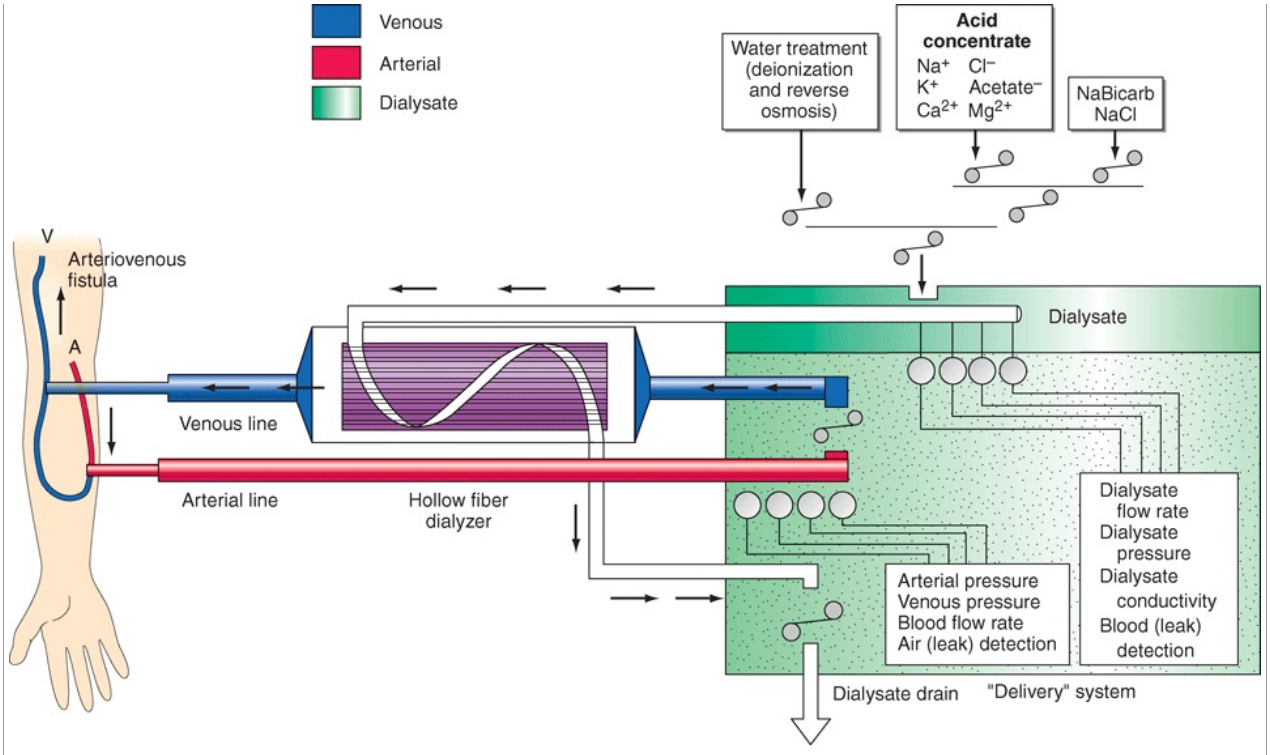

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schema for hemodialysis.

Recent advances have led to the development of many different types of membrane material. Broadly, there are four categories of dialysis membranes: cellulose, substituted cellulose, cellulosynthetic, and synthetic. Over the past three decades, there has been a gradual switch from cellulose-derived to synthetic membranes, because the latter are more "biocompatible." *Bioincompatibility* is generally defined as the ability of the membrane to activate the complement cascade. Cellulosic membranes are bioincompatible because of the presence of free hydroxyl groups on the membrane surface. In contrast, with the substituted cellulose membranes (e.g., cellulose acetate) or the cellulosynthetic membranes, the hydroxyl groups are chemically bound to either acetate or tertiary amino groups, resulting in limited complement activation. Synthetic membranes, such as polysulfone, polymethylmethacrylate, and polyacrylonitrile membranes, are even more biocompatible because of the absence of these hydroxyl groups. The majority of dialyzers now manufactured in the United States are derived from polysulfone or newer derivatives (polyarylethersulfone).

Reprocessing and reuse of hemodialyzers are often employed for patients on maintenance hemodialysis in the United States. However, as the manufacturing costs for disposable dialyzers have declined, more and more outpatient dialysis facilities are no longer reprocessing dialyzers. In most centers employing reuse, only the dialyzer unit is reprocessed and reused, whereas in the developing world blood lines are also frequently reused. The reprocessing procedure can be either manual or automated. It consists of the sequential rinsing of the blood and dialysate compartments with water, a chemical cleansing step with reverse ultrafiltration from the dialysate to the blood compartment, the testing of the patency of the dialyzer, and, finally, disinfection of the dialyzer. Formaldehyde, peracetic acid–hydrogen peroxide, glutaraldehyde, and bleach have all been used as reprocessing agents.

Dialysate

The potassium concentration of dialysate may be varied from 0 to 4 mmol/L depending on the predialysis serum potassium concentration. The usual dialysate calcium concentration in U.S. hemodialysis centers is 1.25 mmol/L (2.5 meq/L, although modification may be required in selected settings (e.g., higher dialysate calcium concentrations may be used in patients with hypocalcemia associated with secondary hyperparathyroidism or following parathyroidectomy). The usual dialysate sodium concentration is 140 mmol/L. Lower dialysate sodium concentrations are associated with a higher frequency of hypotension, cramping, nausea, vomiting, fatigue, and dizziness in some patients, although may attenuate thirst. In patients who frequently develop hypotension during their dialysis run, "sodium modeling" to counterbalance urea-related osmolar gradients is often employed. With sodium modeling, the dialysate sodium concentration is gradually lowered from the range of 145–155 mmol/L to isotonic concentrations (140 mmol/L) near the end of the dialysis treatment, typically declining either in steps or in a linear or exponential fashion. Higher dialysate sodium concentrations and sodium modeling may predispose patients to positive sodium balance; thus, these strategies to ameliorate intradialytic hypotension may be undesirable in hypertensive patients or in patients with large interdialytic weight gains. Because patients are exposed to approximately 120 L of water during each dialysis treatment, water used for the dialysate is subjected to filtration, softening, deionization, and, ultimately, reverse osmosis. During the reverse osmosis process, water is forced through a semipermeable membrane at very high pressure to remove microbiologic contaminants and >90% of dissolved ions.

Blood Delivery System

The blood delivery system is composed of the extracorporeal circuit in the dialysis machine and the dialysis access. The dialysis machine consists of a blood pump, dialysis solution delivery system, and various safety monitors. The blood pump moves blood from the access site, through the dialyzer, and back to the patient. The blood flow rate may range from 250–500 mL/min, depending largely on the type and integrity of the vascular access. Negative hydrostatic pressure on the dialysate side can be manipulated to achieve desirable fluid removal or *ultrafiltration.* Dialysis membranes have different ultrafiltration coefficients (i.e., mL removed/min per mmHg) so that along with hydrostatic changes, fluid removal

can be varied. The dialysis solution delivery system dilutes the concentrated dialysate with water and monitors the temperature, conductivity, and flow of dialysate.

Dialysis Access

The fistula, graft, or catheter through which blood is obtained for hemodialysis is often referred to as a *dialysis access.* A native fistula created by the anastomosis of an artery to a vein (e.g., the Brescia-Cimino fistula, in which the cephalic vein is anastomosed end-to-side to the radial artery) results in arterialization of the vein. This facilitates its subsequent use in the placement of large needles (typically 15 gauge) to access the circulation. Although fistulas have the highest long-term patency rate of all dialysis access options, fistulas are created in a minority of patients in the United States. Many patients undergo placement of an arteriovenous graft (i.e., the interposition of prosthetic material, usually polytetrafluoroethylene, between an artery and a vein) or a tunneled dialysis catheter. In recent years, nephrologists, vascular surgeons, and health care policy makers in the United States have encouraged creation of arteriovenous fistulas in a larger fraction of patients (the "fistula first" initiative). Unfortunately, even when created, arteriovenous fistulas may not mature sufficiently to provide reliable access to the circulation, or they may thrombose early in their development. Novel surgical approaches (e.g., brachiobasilic fistula creation with transposition of the basilic vein fistula to the arm surface) have increased options for "native" vascular access.

Grafts and catheters tend to be used among persons with smaller-caliber veins or persons whose veins have been damaged by repeated venipuncture, or after prolonged hospitalization. The most important complication of arteriovenous grafts is thrombosis of the graft and graft failure, due principally to intimal hyperplasia at the anastomosis between the graft and recipient vein. When grafts (or fistulas) fail, catheter-guided angioplasty can be used to dilate stenoses; monitoring of venous pressures on dialysis and of access flow, although not routinely performed, may assist in the early recognition of impending vascular access failure. In addition to an increased rate of access failure, grafts and (in particular) catheters are associated with much higher rates of infection than fistulas.

Intravenous large-bore catheters are often used in patients with acute and chronic kidney disease. For persons on maintenance hemodialysis, tunneled catheters (either two separate catheters or a single catheter with two lumens) are often used when arteriovenous fistulas and grafts have failed or are not feasible due to anatomic considerations. These catheters are tunneled under the skin; the tunnel reduces bacterial translocation from the skin, resulting in a lower infection rate than with nontunneled temporary catheters. Most tunneled catheters are placed in the internal jugular veins; the external jugular, femoral, and subclavian veins may also be used.

Nephrologists, interventional radiologists, and vascular surgeons generally prefer to avoid placement of catheters into the subclavian veins; while flow rates are usually excellent, subclavian stenosis is a frequent complication and, if present, will likely prohibit permanent vascular access (i.e., a fistula or graft) in the ipsilateral extremity. Infection rates may be higher with femoral catheters. For patients with multiple vascular access complications and no other options for permanent vascular access, tunneled catheters may be the last "lifeline" for hemodialysis. Translumbar or transhepatic approaches into the inferior vena cava may be required if the superior vena cava or other central veins draining the upper extremities are stenosed or thrombosed.

Goals of Dialysis

The hemodialysis procedure is targeted at removing both low- and high-molecular-weight solutes. The procedure consists of pumping heparinized blood through the dialyzer at a flow rate of 300–500 mL/min, while dialysate flows in an opposite *counter-current* direction at 500–800 mL/min. The efficiency of dialysis is determined by blood and dialysate flow through the dialyzer as well as dialyzer characteristics (i.e., its efficiency in removing solute). The *dose* of dialysis, which is currently defined as a derivation of the fractional urea clearance during a single dialysis treatment, is further governed by patient size, residual kidney function, dietary protein intake, the degree of anabolism or catabolism, and the presence of comorbid conditions.

Since the landmark studies of Sargent and Gotch relating the measurement of the dose of dialysis using urea concentrations with morbidity in the National Cooperative Dialysis Study, the *delivered* dose of dialysis has been measured and considered as a quality assurance and improvement tool. While the fractional removal of urea nitrogen and derivations thereof are considered to be the standard methods by which "adequacy of dialysis" is measured, a large multicenter randomized clinical trial (the HEMO Study) failed to show a difference in mortality associated with a large difference in urea clearance. Still, multiple observational studies and widespread expert opinion have suggested that higher dialysis dose is warranted; current targets include a urea reduction ratio (the fractional reduction in blood urea nitrogen per hemodialysis session) of >65–70% and a body water–indexed clearance x time product (KT/V) above 1.2 or 1.05, depending on whether urea concentrations are "equilibrated." For the majority of patients with ESRD, between 9 and 12 h of dialysis are required each week, usually divided into three equal sessions. Several studies have suggested that longer hemodialysis session lengths may be beneficial (independent of urea clearance), although these studies are confounded by a variety of patient characteristics, including body size and nutritional status. Hemodialysis "dose" should be individualized, and factors other than the urea nitrogen should be considered, including the adequacy of ultrafiltration or fluid removal and control of hyperkalemia, hyperphosphatemia, and metabolic acidosis. Several authors have highlighted improved intermediate outcomes associated with more frequent hemodialysis (i.e., more than three times a week), although these studies are also confounded by multiple factors. A randomized clinical trial is currently under way to test whether more frequent dialysis results in differences in a variety of physiologic and functional markers.

Complications during Hemodialysis

Hypotension is the most common acute complication of hemodialysis, particularly among patients with diabetes mellitus. Numerous factors appear to increase the risk of hypotension, including excessive ultrafiltration with inadequate compensatory vascular filling, impaired vasoactive or autonomic responses, osmolar shifts, overzealous use of antihypertensive agents, and reduced cardiac reserve. Patients with arteriovenous fistulas and grafts may develop high output cardiac failure due to shunting of blood through the dialysis access; on rare occasions, this may necessitate ligation of the fistula or graft. Because of the vasodilatory and cardiodepressive effects of acetate, its use as the buffer in dialysate was once a common cause of hypotension. Since the introduction of bicarbonate-containing dialysate, dialysis-associated hypotension has become less common. The management of hypotension during dialysis consists of discontinuing ultrafiltration, the administration of 100–250 mL of isotonic saline or 10 mL of 23% saturated hypertonic saline, or administration of salt-poor albumin. Hypotension during dialysis can frequently be prevented by careful evaluation of the dry weight and by ultrafiltration modeling, such that more fluid is removed at the beginning rather than the end of the dialysis procedure. Additional maneuvers include the performance of sequential ultrafiltration followed by dialysis, cooling of the dialysate during dialysis treatment; and avoiding heavy meals during dialysis. Midodrine, a

selective α_1 adrenergic pressor agent, has been advocated by some practitioners, although there is insufficient evidence of its safety and efficacy to support its routine use.

Muscle cramps during dialysis are also a common complication of the procedure. The etiology of dialysis-associated cramps remains obscure. Changes in muscle perfusion because of excessively aggressive volume removal, particularly below the estimated dry weight, and the use of low-sodium–containing dialysate, have been proposed as precipitants of dialysis-associated cramps. Strategies that may be used to prevent cramps include reducing volume removal during dialysis, ultrafiltration profiling, and the use of higher concentrations of sodium in the dialysate or sodium modeling (see above).

Anaphylactoid reactions to the dialyzer, particularly on its first use, have been reported most frequently with the bioincompatible cellulosic-containing membranes. With the gradual phasing out of cuprophane membranes in the United States, dialyzer reactions have become uncommon. Dialyzer reactions can be divided into two types, A and B. Type A reactions are attributed to an IgE-mediated intermediate hypersensitivity reaction to ethylene oxide used in the sterilization of new dialyzers. This reaction typically occurs soon after the initiation of a treatment (within the first few minutes) and can progress to full-blown anaphylaxis if the therapy is not promptly discontinued. Treatment with steroids or epinephrine may be needed if symptoms are severe. The type B reaction consists of a symptom complex of nonspecific chest and back pain, which appears to result from complement activation and cytokine release. These symptoms typically occur several minutes into the dialysis run and typically resolve over time with continued dialysis.

Cardiovascular disease constitutes the major cause of death in patients with ESRD. Cardiovascular mortality and event rates are higher in dialysis patients than in patients posttransplantation, although rates are extraordinarily high in both populations. The underlying cause of cardiovascular disease is unclear but may be related to shared risk factors (e.g., diabetes mellitus, hypertension, atherosclerotic and arteriosclerotic vascular disease), chronic inflammation, massive changes in extracellular volume (especially with high interdialytic weight gains), inadequate treatment of hypertension, dyslipidemia, anemia, dystrophic vascular calcification, hyperhomocysteinemia, and, perhaps, alterations in cardiovascular dynamics during the dialysis treatment. Few studies have targeted cardiovascular risk reduction in ESRD patients; none have demonstrated consistent benefit. Two clinical trials of statin agents in ESRD demonstrated significant reductions in low-density lipoprotein (LDL) cholesterol concentrations but no significant reductions in death or cardiovascular events [Die Deutsche Diabetes Dialyse Studie (4D) and AURORA studies]. Nevertheless, most experts recommend conventional cardioprotective strategies (e.g., lipid-lowering agents, aspirin, β-adrenergic antagonists) in dialysis patients based on the patients' cardiovascular risk profile, which appears to be increased by more than an order of magnitude relative to persons unaffected by kidney disease.

PERITONEAL DIALYSIS

In peritoneal dialysis, 1.5–3 L of a dextrose-containing solution is infused into the peritoneal cavity and allowed to dwell for a set period of time, usually 2–4 h. As with hemodialysis, toxic materials are removed through a combination of convective clearance generated through ultrafiltration and diffusive clearance down a concentration gradient. The clearance of solutes and water during a peritoneal dialysis exchange depends on the balance between the movement of solute and water into the peritoneal cavity versus absorption from the peritoneal cavity. The rate of diffusion diminishes with time and eventually stops when equilibration between plasma and dialysate is reached. Absorption of solutes and water from the peritoneal cavity occurs across the peritoneal membrane into the peritoneal capillary circulation and via peritoneal lymphatics into the lymphatic circulation. The rate of peritoneal solute transport varies from patient to patient and may be altered by the presence of infection (peritonitis), drugs, and physical factors such as position and exercise.

Forms of Peritoneal Dialysis

Peritoneal dialysis may be carried out as CAPD, CCPD, or a combination of both. In CAPD, dialysis solution is manually infused into the peritoneal cavity during the day and exchanged three to five times daily. A nighttime dwell is frequently instilled at bedtime and remains in the peritoneal cavity through the night. The drainage of spent dialysate is performed manually with the assistance of gravity to move fluid out of the abdomen. In CCPD, exchanges are performed in an automated fashion, usually at night; the patient is connected to an automated cycler that performs a series of exchange cycles while the patient sleeps. The number of exchange cycles required to optimize peritoneal solute clearance varies by the peritoneal membrane characteristics; as with hemodialysis, experts suggest careful tracking of solute clearances to ensure dialysis "adequacy."

Peritoneal dialysis solutions are available in volumes typically ranging from 1.5 to 3 L. Lactate is the preferred buffer in peritoneal dialysis solutions. The most common additives to peritoneal dialysis solutions are heparin to prevent obstruction of the dialysis catheter lumen with fibrin and antibiotics during an episode of acute peritonitis. Insulin may also be added in patients with diabetes mellitus.

Access to the Peritoneal Cavity

Access to the peritoneal cavity is obtained through a peritoneal catheter. Catheters used for maintenance peritoneal dialysis are flexible, being made of silicone rubber with numerous side holes at the distal end. These catheters usually have two Dacron cuffs to promote fibroblast proliferation, granulation, and invasion of the cuff. The scarring that occurs around the cuffs anchors the catheter and seals it from bacteria tracking from the skin surface into the peritoneal cavity; it also prevents the external leakage of fluid from the peritoneal cavity. The cuffs are placed in the preperitoneal plane and ~2 cm from the skin surface.

The *peritoneal equilibrium test* is a formal evaluation of peritoneal membrane characteristics that measures the transfer rates of creatinine and glucose across the peritoneal membrane. Patients are classified as low, low–average, high–average, and high transporters. Patients with rapid equilibration (i.e., high transporters) tend to absorb more glucose and lose efficiency of ultrafiltration with long daytime dwells. High transporters also tend to lose larger quantities of albumin and other proteins across the peritoneal membrane. In general, patients with rapid transporting characteristics require more frequent, shorter dwell time exchanges, nearly always obligating use of a cycler for feasibility. Slower (low and low–average) transporters tend to do well with fewer exchanges. The efficiency of solute clearance also depends on the volume of dialysate infused. Larger volumes allow for greater solute clearance, particularly with CAPD in patients with low and low–average transport characteristics. Interestingly, solute clearance also increases with physical activity, presumably related to more efficient flow dynamics within the peritoneal cavity.

As with hemodialysis, the optimal dose of peritoneal dialysis is unknown. Several observational studies have suggested that higher rates of urea and creatinine clearance (the latter generally measured in L/week) are associated with lower mortality rates and fewer uremic complications. However, a randomized clinical trial [Adequacy of Peritoneal Dialysis in

Mexico (ADEMEX)] failed to show a significant reduction in mortality or complications with a relatively large increment in urea clearance. In general, patients on peritoneal dialysis do well when they retain residual kidney function. The rates of technique failure increase with years on dialysis and have been correlated with loss of residual function to a greater extent than loss of peritoneal membrane capacity. Recently, a nonabsorbable carbohydrate (icodextrin) has been introduced as an alternative osmotic agent. Studies have demonstrated more efficient ultrafiltration with icodextrin than with dextrose-containing solutions. Icodextrin is typically used as the "last fill" for patients on CCPD or for the longest dwell in patients on CAPD. For some patients in whom CCPD does not provide sufficient solute clearance, a hybrid approach can be adopted where one or more daytime exchanges are added to the CCPD regimen. While this approach can enhance solute clearance and prolong a patient's capacity to remain on peritoneal dialysis, the burden of the hybrid approach can be overwhelming to some.

Complications during Peritoneal Dialysis

The major complications of peritoneal dialysis are peritonitis, catheter-associated nonperitonitis infections, weight gain and other metabolic disturbances, and residual uremia (especially among patients with no residual kidney function).

Peritonitis typically develops when there has been a break in sterile technique during one or more of the exchange procedures. Peritonitis is usually defined by an elevated peritoneal fluid leukocyte count ($100/mm^3$, of which at least 50% are polymorphonuclear neutrophils); these cutoffs are lower than in spontaneous bacterial peritonitis because of the presence of dextrose in peritoneal dialysis solutions and rapid bacterial proliferation in this environment without antibiotic therapy. The clinical presentation typically consists of pain and cloudy dialysate, often with fever and other constitutional symptoms. The most common culprit organisms are gram-positive cocci, including *Staphylococcus,* reflecting the origin from the skin. Gram-negative rod infections are less common; fungal and mycobacterial infections can be seen in selected patients, particularly after antibacterial therapy. Most cases of peritonitis can be managed either with intraperitoneal or oral antibiotics, depending on the organism; many patients with peritonitis do not require hospitalization. In cases where peritonitis is due to hydrophilic gram negative rods (e.g., *Pseudomonas* sp.) or yeast, antimicrobial therapy is usually not sufficient, and catheter removal is required to ensure complete eradication of infection. Nonperitonitis catheter-associated infections (often termed *tunnel infections*) vary widely in severity. Some cases can be managed with local antibiotic or silver nitrate administration, while others are severe enough to require parenteral antibiotic therapy and catheter removal.

Peritoneal dialysis is associated with a variety of metabolic complications. As noted above, albumin and other proteins can be lost across the peritoneal membrane in concert with the loss of metabolic wastes. The hypoproteinemia induced by peritoneal dialysis obligates a higher dietary protein intake in order to maintain nitrogen balance. Hyperglycemia and weight gain are also common complications of peritoneal dialysis. Several hundred calories in the form of dextrose are absorbed each day, depending on the concentration employed. Peritoneal dialysis patients, particularly those with type II diabetes mellitus, are then prone to other complications of insulin resistance, including hypertriglyceridemia. On the positive side, the continuous nature of peritoneal dialysis usually allows for a more liberal diet, due to continuous removal of potassium and phosphorus—two major dietary components whose accumulation can be hazardous in ESRD.

GLOBAL PERSPECTIVE

The incidence of ESRD is increasing worldwide with longer life expectancies and improved care of infectious and cardiovascular diseases. The management of ESRD varies widely by country and within country by region, and it is influenced by economic and other major factors. In general, peritoneal dialysis is more commonly performed in poorer countries owing to its lower expense and the high cost of establishing in-center hemodialysis units.

ACKNOWLEDGMENT

We are grateful to Dr. Ajay Singh and Dr. Barry Brenner, authors of "Dialysis in the Treatment of Renal Failure" in the 16th edition of Harrison's, for contributions to this chapter.

FURTHER READINGS

Burkart JM et al: Peritoneal dialysis, in *Brenner and Rector's The Kidney,* 7th ed, BM Brenner (ed). Philadelphia, Saunders, 2004

Eknoyan G et al: Effect of dialysis dose and membrane flux in maintenance hemodialysis. N Engl J Med 346:2010, 2002

Fellstrom BC et al: Rosuvastatin and cardiovascular events in patients undergoing hemodialysis. N Engl J Med 360:1395, 2009[PMID: 19332456]

Forni LG, Hilton PJ: Current concepts: Continuous hemofiltration in the treatment of acute renal failure. N Engl J Med 336:1303, 1997[PMID: 9113935]

Himmelfarb J, Kliger AS: End-stage renal disease measures of quality. Annu Rev Med 58:387, 2007[PMID: 17217331]

National Kidney Foundation: Kidney Disease Quality Initiative Clinical Practice Guidelines: Hemodialysis and peritoneal dialysis adequacy, 2001. Available online: *http://www.kidney.org/professionals/kdoqi/guidelines.cfm*

Paniagua R et al: Effects of increased peritoneal clearances on mortality rates in peritoneal dialysis: ADEMEX, a prospective, randomized, controlled trial. J Am Soc Nephrol 13:1307, 2002[PMID: 11961019]

Rayner HC et al: Vascular access results from the Dialysis Outcomes and Practice Patterns Study (DOPPS): Performance against Kidney Disease Outcomes Quality Initiative (K/DOQI) Clinical Practice Guidelines. Am J Kidney Dis 44:S22, 2004

U.S. Renal Data System: USRDS 2009 Annual Data Report: Atlas of End-Stage Renal Disease in the United States. Bethesda, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Disease, 2009

Wanner C et al: Atorvastatin in patients with type 2 diabetes mellitus undergoing hemodialysis. N Engl J Med 353: 238, 2005[PMID: 16034009]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TRANSPLANTATION IN THE TREATMENT OF RENAL FAILURE: INTRODUCTION

Transplantation of the human kidney is the treatment of choice for advanced chronic renal failure. Worldwide, tens of thousands of these procedures have been performed. When azathioprine and prednisone initially were used as immunosuppressive drugs in the 1960s, the results with properly matched familial donors were superior to those with organs from deceased donors: 75–90% compared with 50–60% graft survival rates at 1 year. During the 1970s and 1980s, the success rate at the 1-year mark for deceased-donor transplants rose progressively. Currently, deceased-donor grafts have an 89% 1-year survival and living-donor grafts have a 95% 1-year survival. Although there has been improvement in long-term survival, it has not been as impressive as the short-term survival, and currently the "average" ($t_{1/2}$) life expectancy of a living-donor graft is around 20 years and that of a deceased-donor graft is close to 14 years.

Mortality rates after transplantation are highest in the first year and are age-related: 2% for ages 18–34 years, 3% for ages 35–49 years, and 6.8% for ages ≥50–60 years. These rates compare favorably with those in the chronic dialysis population even after risk adjustments for age, diabetes, and cardiovascular status. Occasionally, acute irreversible rejection may occur after many months of good function, especially if the patient neglects to take the prescribed immunosuppressive drugs. Most grafts, however, succumb at varying rates to a chronic process consisting of interstitial fibrosis, tubular atrophy, vasculopathy, and glomerulopathy, the pathogenesis of which is incompletely understood. Overall, transplantation returns most patients to an improved lifestyle and an improved life expectancy compared with patients on dialysis. There are at least 100,000 patients with functioning kidney transplants in the United States, and when one adds in the numbers of kidney transplants in centers around the world, the total activity is doubled.

RECENT ACTIVITY AND RESULTS

In 2008 there were more than 10,500 deceased-donor kidney transplants and 6000 living-donor transplants in the United States, with the ratio of deceased to living donors being stable over the last few years. The backlog of patients with end-stage renal disease (ESRD) has been increasing every year, and it always lags behind the number of available donors. As the number of patients with end-stage kidney disease increases, the demand for kidney transplants continues to increase. In 2008, 33,051 new registrants were added to the waiting list and under 17,000 patients were transplanted. This imbalance is set to worsen over the coming years with the predicted increased rates of obesity and diabetes worldwide. In an attempt to increase utilization of deceased-donor kidneys and reduce discard rates of organs, criteria for the use of so-called expanded criteria donor (ECD) kidneys and kidneys from donors after cardiac death (DCD) have been developed (**Table 282-1**). ECD kidneys are usually used for older patients who are expected to fare less well on dialysis.

Table 282-1 Definition of an Expanded Criteria Donor (ECD) and a Non-Heart-Beating Donor [Donation after Cardiac Death (DCD)]

Expanded Criteria Donor (ECD)
Deceased donor >60 years
Deceased donor >50 years and hypertension and creatinine >1.5 mg/dL
Deceased donor >50 years and hypertension and death caused by cerebrovascular accident (CVA)
Deceased donor >50 years and death caused by CVA and creatinine >1.5 mg/dL
Donation after Cardiac Death[a] (DCD)
I Brought in dead
II Unsuccessful resuscitation
III Awaiting cardiac arrest
IV Cardiac arrest after brainstem death
V Cardiac arrest in a hospital patient

[a]Kidneys can be used for transplantation from categories II–V but are commonly only used from categories III and IV. The survival of these kidneys has not been shown to be inferior to that of deceased-donor kidneys.

Note: Kidneys can be bought ECD and DCD. ECD kidneys have been shown to have a poorer survival, and there is a separate shorter waiting list for ECD kidneys. They are generally utilized for patients for whom the benefits of being transplanted earlier outweigh the associated risks of using an ECD kidney.

The overall results of transplantation are presented in **Table 282-2** as the survival of grafts and of patients. At the 1-year mark, graft survival is higher for living-donor recipients, most likely because those grafts are not subject to as much ischemic injury. The more powerful drugs now in use for immunosuppression have almost equalized the risk of graft rejection in all patients for the first year. At 5 and 10 years, however, there has been a steeper decline in survival of those with deceased-donor kidneys.

Table 282-2 Mean Rates of Graft and Patient Survival for Kidneys Transplanted in the United States from 1992 to 2002[a]

	1-Year Follow-Up		5-Year Follow-Up		10-Year Follow-Up	
	Grafts, %	**Patients, %**	**Grafts, %**	**Patients, %**	**Grafts, %**	**Patients, %**
Deceased donor	89	95	67	81	41	61
Living donor	95	98	80	90	56	76

[a]All patients transplanted are included, and the follow-up unadjusted survival data from the 1-, 5-, and 10-year periods are presented to show the attrition rates over time within the two types of organ donors.

Source: Data from Summary Tables, 2004 and 2005 Annual Reports, Scientific Registry of Transplant Recipients.

RECIPIENT SELECTION

There are few absolute contraindications to renal transplantation. The transplant procedure is relatively noninvasive, as the organ is placed in the inguinal fossa without entering the peritoneal cavity. Recipients without perioperative complications often can be discharged from the hospital in excellent condition within 5 days of the operation.

Virtually all patients with ESRD who receive a transplant have a higher life expectancy than do risk-matched patients who remain on dialysis. Even though diabetic patients and older candidates have a higher mortality rate than other transplant recipients, their survival is improved with transplantation compared with remaining on dialysis. This global benefit of transplantation as a treatment modality poses substantial ethical issues for policy makers, as the number of deceased kidneys available is far from sufficient to meet the current needs of the candidates. The current standard of care is that the candidate should have a life expectancy of >5 years to be put on a deceased organ wait list. Even for living donation, the candidate should have >5 years of life expectancy. This standard has been established because the benefits of kidney transplantation over dialysis are realized only after a perioperative period in which the mortality rate is higher in transplanted patients than in dialysis patients with comparable risk profiles.

All candidates must have a thorough risk-versus-benefit evaluation before being approved for transplantation. In particular, an aggressive approach to diagnosis of correctable coronary artery disease, presence of latent or indolent infection (HIV, hepatitis B or C, tuberculosis), and neoplasm should be a routine part of the candidate workup. Most transplant centers consider overt AIDS and active hepatitis absolute contraindications to transplantation because of the high risk of opportunistic infection. Some centers are now transplanting individuals with hepatitis and even HIV infection under strict protocols to determine whether the risks and benefits favor transplantation over dialysis.

Among the few absolute "immunologic" contraindications to transplantation is the presence of a potentially harmful antibody against the donor kidney at the time of the anticipated transplant. Harmful antibodies that can cause very early graft loss include natural antibodies against the ABO blood group antigens and antibodies against human leukocyte antigen (HLA) class I (A, B, C) or class II (DR) antigens. These antibodies are routinely excluded by proper screening of the candidate's ABO compatibility, HLA typing of donor and recipient, and direct cross-matching of candidate serum with lymphocytes of the donor.

DONOR SELECTION

Donors can be deceased or volunteer living donors. The latter are usually family members selected to have at least partial compatibility for HLA antigens. Living volunteer donors should be normal on

physical examination and of the same major ABO blood group, because crossing major blood group barriers prejudices survival of the allograft. It is possible, however, to transplant a kidney of a type O donor into an A, B, or AB recipient. Selective renal arteriography should be performed on donors to rule out the presence of multiple or abnormal renal arteries because the surgical procedure is difficult and the ischemic time of the transplanted kidney is long when there are vascular abnormalities. Transplant surgeons are now using a laparoscopic method to isolate and remove the living donor's kidney. This operation has the advantage of less evident surgical scars, and, because there is less tissue trauma, the laparoscopic donors have a substantially shorter hospital stay and less discomfort than those who have the traditional surgery. Deceased donors should be free of malignant neoplastic disease, hepatitis, and HIV because of possible transmission to the recipient. Increased risk of graft failure exists when the donor is elderly or has renal failure and when the kidney has a prolonged period of ischemia and storage.

In the United States, there is a coordinated national system of regulations, allocation support, and outcomes analysis for kidney transplantation called the Organ Procurement Transplant Network. It is now possible to remove deceased-donor kidneys and maintain them for up to 48 h on cold pulsatile perfusion or simple flushing and cooling. This approach permits adequate time for typing, cross-matching, transportation, and selection problems to be solved.

TISSUE TYPING AND CLINICAL IMMUNOGENETICS

Matching for antigens of the HLA major histocompatibility gene complex (Chap. 315) is an important criterion for selection of donors for renal allografts. Each mammalian species has a single chromosomal region that encodes the strong, or major, transplantation antigens, and this region on the human sixth chromosome is called *HLA*. HLA antigens have been classically defined by serologic techniques, but methods to define specific nucleotide sequences in genomic DNA are increasingly being used. Other "minor" antigens may play crucial roles, in addition to the ABH(O) blood groups and endothelial antigens that are not shared with lymphocytes. The Rh system is not expressed on graft tissue. Evidence for designation of HLA as the genetic region that encodes major transplantation antigens comes from the success rate in living related donor renal and bone marrow transplantation, with superior results in HLA-identical sibling pairs. Nevertheless, 5% of HLA-identical renal allografts are rejected, often within the first weeks after transplantation. These failures represent states of prior sensitization to non-HLA antigens. Non-HLA minor antigens are relatively weak when initially encountered and are, therefore, suppressible by conventional immunosuppressive therapy. Once priming has occurred, however, secondary responses are much more refractory to treatment.

Living Donors

When first-degree relatives are donors, graft survival rates at 1 year are 5–7% greater than those for deceased-donor grafts. The 5-year survival rates still favor a partially matched (3/6 HLA mismatched) family donor over a randomly selected cadaver donor **(Table 282-3)**. In addition, living donors provide the advantage of immediate availability. For both living and deceased donors, the 5-year outcomes are poor if there is a complete (6/6) HLA mismatch.

Table 282-3 Effect of HLA-A, -B, -DR Mismatching on Kidney Graft Survival

Degree of Donor Mismatch	1-Year Survival, %	5-Year Survival, %
Cadaver donor (all)	89.2	61.3
0/6-HLA mismatch	91.3	68.2
3/6-HLA mismatch	90.1	60.8
6/6-HLA mismatch	85.2	55.3
Living related donor (all)	94.7	76.0
0/6-HLA mismatch	96.7	87.0
3/6-HLA mismatch	94.3	73.2
6/6-HLA mismatch	92.7	57.7

Note: 0-mismatched related donor transplants are virtually all from HLA-identical siblings, whereas 3/6-mismatched transplants can be one haplotype mismatched (1-A, 1-B, and 1-DR antigen) from a parent, child, or sibling; 6/6-HLA-mismatched living related kidneys are derived from siblings or relatives outside the nuclear family.

The survival rate of living unrelated renal allografts is as high as that of perfectly HLA-matched cadaver renal transplants and comparable to that of kidneys from living relatives. This outcome is probably a consequence of both short cold ischemia time and the extra care taken to document that the condition and renal function of the donor are optimal before proceeding with a living unrelated donation. It is illegal in the United States to purchase organs for transplantation.

Concern has been expressed about the potential risk to a volunteer kidney donor of premature renal

failure after several years of increased blood flow and hyperfiltration per nephron in the remaining kidney. There are a few reports of the development of hypertension, proteinuria, and even lesions of focal segmental sclerosis in donors over long-term follow-up. Difficulties in donors followed for ≥20 years are unusual, however, and it may be that having a single kidney becomes significant only when another condition, such as hypertension, is superimposed. It is also desirable to consider the risk of development of type 1 diabetes mellitus in a family member who is a potential donor to a diabetic renal failure patient. Anti-insulin and anti-islet cell antibodies should be measured, and glucose tolerance tests should be performed in such donors to exclude a prediabetic state.

Presensitization

A positive cytotoxic cross-match of recipient serum with donor T lymphocytes representing anti-HLA class I is usually predictive of an acute vasculitic event termed *hyperacute rejection*. Patients with anti-HLA antibodies can be transplanted safely if careful cross-matching of donor blood lymphocytes with recipient serum is performed. The known sources of such sensitization are blood transfusion, a prior transplant, and pregnancy. Patients sustained by dialysis often show fluctuating antibody titers and specificity patterns. At the time of assignment of a cadaveric kidney, cross-matches are performed with at least a current serum. Previously analyzed antibody specificities and additional cross-matches are performed accordingly. The minimal purpose for the cross-match is avoidance of hyperacute rejection mediated by recipient antibodies to donor HLA class I antigens. More sensitive tests, such as the use of flow cytometry, can be useful for avoidance of accelerated, and often untreatable, early graft rejection in patients receiving second or third transplants. Donor T lymphocytes, which express only class I antigens, are used as targets for detection of anti–class I (HLA-A and -B) antibodies.

Preformed anti–class II (HLA-DR) antibodies against the donor carry a higher risk of graft loss as well, particularly in recipients who have suffered early loss of a prior kidney transplant. B lymphocytes expressing both class I and class II antigens are used in these assays. Non-HLA antigens restricted in expression to endothelium and sometimes monocytes have been described, but clinical relevance is not well established. A series of minor histocompatibility antigens do not elicit antibodies, and sensitization to these antigens is detectable only by cytotoxic T cells, an assay too cumbersome for routine use. Desensitization before transplantation by reducing the level of antidonor antibodies via plasmapheresis of blood and administration of pooled immunoglobulin or both has been useful in reducing the hazard of hyperacute rejection.

IMMUNOLOGY OF REJECTION

Both cellular and humoral (antibody-mediated) effector mechanisms can play roles in kidney transplant rejection. Antibodies can also initiate a form of antibody-dependent but cell-mediated cytotoxicity by recipient cells that bear receptors for the Fc portion of immunoglobulin.

Cellular rejection is mediated by lymphocytes that respond to HLA antigens expressed within the organ. The CD4+ lymphocyte responds to class II (HLA-DR) incompatibility by proliferating and releasing proinflammatory cytokines that augment the proliferative response of both CD4+ and CD8+ cells. CD8+ cytotoxic lymphocyte precursors respond primarily to class I (HLA-A, -B) antigens and mature into cytotoxic effector cells. The cytotoxic effector ("killer") T cells cause organ damage through direct contact and lysis of donor target cells. The natural role of HLA antigen molecules is to present processed peptide fragments of antigen to T lymphocytes, the fragments residing in a "groove" of the HLA molecule distal to the cell surface. T cells can be directly stimulated by intact nonself HLA molecules expressed on donor parenchymal cells and residual donor leukocytes residing in the kidney interstitium. In addition, donor HLA molecules can be processed by a variety of donor or recipient cells capable of antigen presentation of peptides and then presented to T cells in the same manner as most other antigens. The former mode of stimulation is sometimes called *direct presentation,* and the latter mode *indirect presentation***(Fig. 282-1)**. There is evidence that non-HLA antigens can also play a role in renal transplant rejection episodes. Recipients who receive a kidney from an HLA-identical sibling can have rejection episodes and require maintenance immunosuppression, whereas identical twin transplants require no immunosuppression. There are documented non-HLA antigens, such as an endothelial-specific antigen system with limited polymorphism and a tubular antigen, that can be targets of humoral or cellular rejection responses, respectively.

Figure 282-1

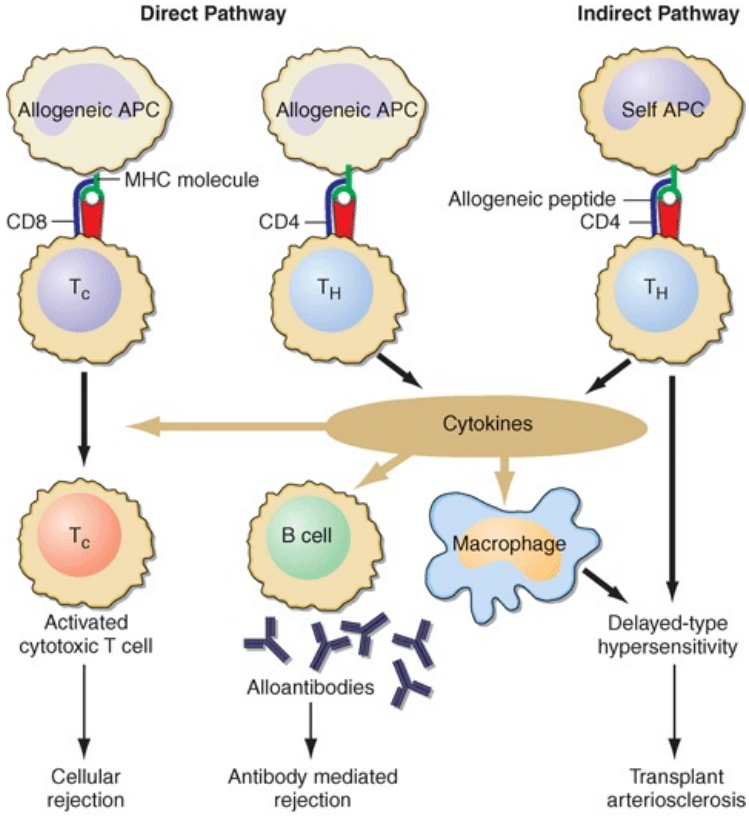

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Recognition pathways for major histocompatibility complex (MHC) antigens. Graft rejection is initiated by CD4 helper T lymphocytes (T_H) having antigen receptors that bind to specific complexes of peptides and MHC class II molecules on antigen-presenting cells (APC). In transplantation, in contrast to other immunologic responses, there are two sets of T cell clones involved in rejection. In the direct pathway the class II MHC of donor allogeneic APCs is recognized by CD4 T_H cells that bind to the intact MHC molecule, and class I MHC allogeneic cells are recognized by CD8 T cells. The latter generally proliferate into cytotoxic cells (T_C). In the indirect pathway, the incompatible MHC molecules are processed into peptides that are presented by the self-APCs of the recipient. The indirect, but not the direct, pathway is the normal physiologic process in T cell recognition of foreign antigens. Once T_H cells are activated, they proliferate and, by secretion of cytokines and direct contact, exert strong helper effects on macrophages, T_C, and B cells. *(From MH Sayegh, LH Turka, N Engl J Med, 338:1813, 1998. Copyright 1998, Massachusetts Medical Society. All rights reserved.)*

IMMUNOSUPPRESSIVE TREATMENT

Immunosuppressive therapy, as currently available, generally suppresses all immune responses, including those to bacteria, fungi, and even malignant tumors. In the 1950s, when clinical renal transplantation began, sublethal total-body irradiation was employed. We have now reached the point where sophisticated pharmacologic immunosuppression is available, but it still has the hazard of promoting infection and malignancy. In general, all clinically useful drugs are more selective to primary than to memory immune responses. Agents to suppress the immune response are discussed in the following paragraphs, and those currently in clinical use are listed in **Table 282-4**.

Table 282-4 Maintenance Immunosuppressive Drugs

Agent	Pharmacology	Mechanisms	Side Effects
Glucocorticoids	Increased bioavailability with hypoalbuminemia and liver disease; prednisone/prednisolone generally used	Binds cytosolic receptors and heat shock proteins. Blocks transcription of IL-1,-2,-3,-6, TNF-α, and IFN-γ	Hypertension, glucose intolerance, dyslipidemia, osteoporosis
Cyclosporine (CsA)	Lipid-soluble polypeptide, variable absorption, microemulsion more predictable	Trimolecular complex with cyclophilin and calcineurin → block in cytokine (e.g., IL-2) production; however, stimulates TGF-β production	Nephrotoxicity, hypertension, dyslipidemia, glucose intolerance, hirsutism/hyperplasia of gums

Tacrolimus (FK506)	Macrolide, well absorbed	Trimolecular complex with FKBP-12 and calcineurin → block in cytokine (e.g., IL-2) production; may stimulate TGF-β production	Similar to CsA, but hirsutism/hyperplasia of gums unusual, and diabetes more likely
Azathioprine	Mercaptopurine analogue	Hepatic metabolites inhibit purine synthesis	Marrow suppression (WBC > RBC > platelets)
Mycophenolate mofetil (MMF)	Metabolized to mycophenolic acid	Inhibits purine synthesis via inosine monophosphate dehydrogenase	Diarrhea/cramps; dose-related liver and marrow suppression is uncommon
Sirolimus	Macrolide, poor oral bioavailability	Complexes with FKBP-12 and then blocks p70 S6 kinase in the IL-2 receptor pathway for proliferation	Hyperlipidemia, thrombocytopenia

Abbreviations: FKBP-12, FK506 binding protein 12; IFN, interferon; IL, interleukin; RBC, red blood cells; TGF, transforming growth factor; TNF, tumor necrosis factor; WBC, white blood cells.

Drugs

Azathioprine, an analogue of mercaptopurine, was for two decades the keystone to immunosuppressive therapy in humans but has given way to more effective agents. This agent can inhibit synthesis of DNA, RNA, or both. Therapy with azathioprine in doses of 1.5–2 mg/kg per d is generally added to cyclosporine as a means of decreasing the requirements for the latter. Reduction in the dose is required because of leukopenia and occasionally thrombocytopenia. Excessive amounts of azathioprine may also cause jaundice, anemia, and alopecia. If it is essential to administer allopurinol concurrently, the azathioprine dose must be reduced. As inhibition of xanthine oxidase delays degradation, this combination is best avoided.

Mycophenolate mofetil or mycophenolate sodium, both of which are metabolized to mycophenolic acid, is now used in place of azathioprine in most centers. It has a similar mode of action and a mild degree of gastrointestinal toxicity but produces minimal bone marrow suppression. Its advantage is its increased potency in preventing or reversing rejection. Patients with hyperuricemia can be given allopurinol without adjustment of the mycophenolic acid dose. The usual dose is 2–3 g/d in divided doses.

Glucocorticoids are important adjuncts to immunosuppressive therapy. Among all the agents employed, prednisone has effects that are easiest to assess, and in large doses it is usually effective for the reversal of rejection. In general, 200–300 mg prednisone is given immediately before or at the time of transplantation, and the dose is reduced to 30 mg within a week. The side effects of the glucocorticoids, particularly impairment of wound healing and predisposition to infection, make it desirable to taper the dose as rapidly as possible in the immediate postoperative period. Many centers now have protocols for early discontinuance or avoidance of steroids because of long-term adverse effects on bone, skin, and glucose metabolism. For treatment of acute rejection, methylprednisolone, 0.5–1 g IV, is administered immediately upon diagnosis of beginning rejection and continued once daily for 3 days. When the drug is effective, the results are usually apparent within 96 h. Such "pulse" doses are not effective in chronic rejection. Most patients whose renal function is stable after 6 months or a year do not require large doses of prednisone; maintenance doses of 10–15 mg/d are the rule. Many patients tolerate an alternate-day course of steroids without an increased risk of rejection. A major effect of steroids is on the monocyte-macrophage system, preventing the release of interleukin (IL) 6 and IL-1.

Cyclosporine is a fungal peptide with potent immunosuppressive activity. It acts on the calcineurin pathway to block transcription of mRNA for IL-2 and other proinflammatory cytokines, thereby inhibiting T cell proliferation. Although it works alone, cyclosporine is more effective in conjunction with glucocorticoids and mycophenolate. Clinical results with tens of thousands of renal transplants have been impressive. Among its toxic effects (nephrotoxicity, hepatotoxicity, hirsutism, tremor, gingival hyperplasia, diabetes), only nephrotoxicity presents a serious management problem and is further discussed below.

Tacrolimus (previously called FK506) is a fungal macrolide that has the same mode of action as cyclosporine as well as a similar side-effect profile; it does not, however, produce hirsutism or gingival hyperplasia. De novo diabetes mellitus is more common with tacrolimus. The drug was first used in liver transplantation and may substitute for cyclosporine entirely or be tried as an alternative in renal patients whose rejections are poorly controlled by cyclosporine.

Sirolimus (previously called rapamycin) is another fungal macrolide but has a different mode of action; i.e., it inhibits T cell growth factor signaling pathways, preventing the response to IL-2 and other cytokines. Sirolimus can be used in conjunction with cyclosporine or tacrolimus, or with mycophenolic acid, to avoid calcineurin inhibitors. Its use with tacrolimus alone shows promise as a steroid-sparing

regimen, especially in patients who would benefit from pancreatic islet cell transplantation, where steroids have an adverse effect on islet survival.

Antibodies to Lymphocytes

When serum from animals made immune to host lymphocytes is injected into the recipient, a marked suppression of cellular immunity to the tissue graft results. The action on cell-mediated immunity is greater than the action on humoral immunity. A globulin fraction of serum [antilymphocyte globulin (ALG)] is the agent generally employed. For use in humans, peripheral human lymphocytes, thymocytes, or lymphocytes from spleens or thoracic duct fistulas have been injected into horses, rabbits, or goats to produce antilymphocyte serum, from which the globulin fraction is then separated. A rabbit antithymocyte globulin (thymoglobulin) is the agent most commonly in use currently. Monoclonal antibodies against defined lymphocyte subsets offer a more precise and standardized form of therapy. OKT3 is directed to the CD3 molecules that form a portion of the T cell antigen-receptor complex and is thus expressed on all mature T cells.

Another approach to more selective therapy is to target the 55-kDa alpha chain of the IL-2 receptor, which is expressed only on T cells that have been recently activated. Two such antibodies to the IL-2 receptor, in which either a chimeric protein has been made between mouse Fab with human Fc (basiliximab) or the antibody has been "humanized" by splicing the combining sites of the mouse into a molecule that is 90% human IgG (daclizumab), are in use for prophylaxis of acute rejection in the immediate posttransplant period. They are effective at decreasing the acute rejection rate and have few adverse side effects.

More recently, two new strategies have involved administration of engineered biologic agents: a depleting T cell antibody (alemtuzumab) as induction therapy to minimize maintenance immunosuppression and a fusion protein (Belatacept) to block B7 T cell costimulatory signals. The latter has shown promise in phase 2 trials and is currently being tested in phase 3 trials in kidney transplantation. Both of these new biologics as well as antilymphocyte globulin are increasingly being used as "induction" therapy at the time of transplantation to minimize or eliminate the use of either steroids or calcineurin inhibitors because of their perceived toxicities. The next step in the evolution of this therapeutic strategy, which has already been achieved in the short term in small numbers of immunologically well-matched patients, is the elimination of all maintenance immunosuppression therapy altogether.

CLINICAL COURSE AND MANAGEMENT OF THE RECIPIENT

Adequate hemodialysis should be performed within 48 h of surgery, and care should be taken that the serum potassium level is not markedly elevated so that intraoperative cardiac arrhythmias can be averted. The diuresis that commonly occurs postoperatively must be carefully monitored. In some instances, it may be massive, reflecting the inability of ischemic tubules to regulate sodium and water excretion; with large diureses, massive potassium losses may occur. Most chronically uremic patients have some excess of extracellular fluid, and it is useful to maintain an expanded fluid volume in the immediate postoperative period. Acute tubular necrosis (ATN) may cause immediate oliguria or may follow an initial short period of graft function. ATN is most likely when cadaveric donors have been underperfused or if the interval between cessation of blood flow and organ harvest (warm ischemic time) is more than a few minutes. Recovery usually occurs within 3 weeks, although periods as long as 6 weeks have been reported. Superimposition of rejection on ATN is common, and the differential diagnosis may be difficult without a graft biopsy. Cyclosporine therapy prolongs ATN, and some patients do not diurese until the dose is reduced drastically. Many centers avoid starting cyclosporine for the first several days, using ALG or a monoclonal antibody along with mycophenolic acid and prednisone until renal function is established. **Figure 282-2** illustrates an algorithm followed by many transplant centers for early posttransplant management of recipients at high or low risk of early renal dysfunction.

Figure 282-2

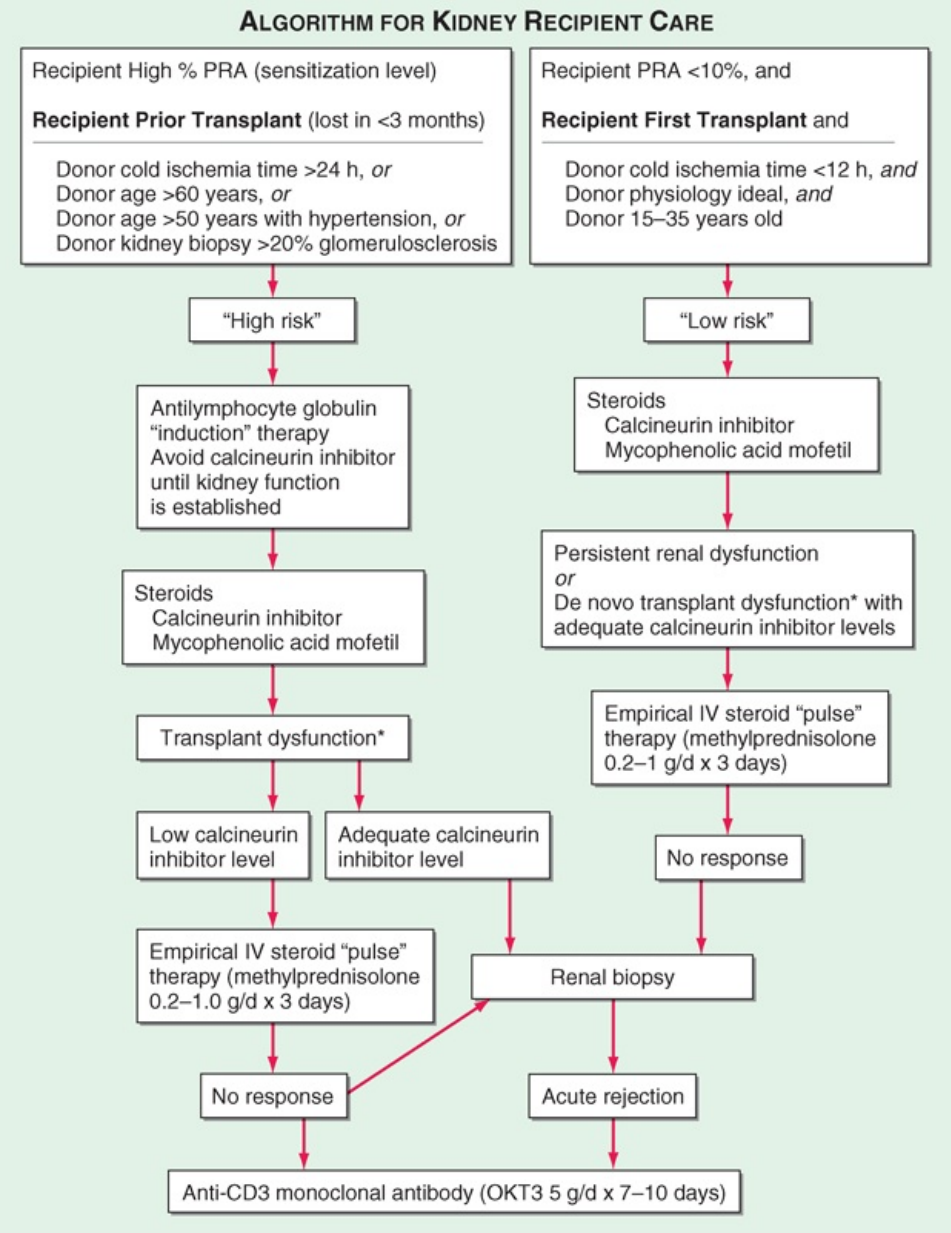

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A typical algorithm for early posttransplant care of a kidney recipient. If any of the recipient or donor "high-risk" factors exist, more aggressive management is called for. Low-risk patients can be treated with a standard immunosuppressive regimen. Patients at higher risk of rejection or early ischemic and nephrotoxic transplant dysfunction are often induced with an antilymphocyte globulin to provide more potent early immunosuppression or to spare calcineurin nephrotoxicity. *When there is early transplant dysfunction, prerenal, obstructive, and vascular causes must be ruled out by ultrasonographic examination. The panel reactive antibody (PRA) is a quantitation of how much antibody is present in a candidate against a panel of cells representing the distribution of antigens in the donor pool. APC, antigen-presenting cell; MHC, major histocompatibility complex.

The Rejection Episode

Early diagnosis of rejection allows prompt institution of therapy to preserve renal function and prevent irreversible damage. Clinical evidence of rejection is rarely characterized by fever, swelling, and tenderness over the allograft. Rejection may present only with a rise in serum creatinine, with or without a reduction in urine volume. The focus should be on ruling out other causes of functional deterioration.

Doppler ultrasonography or magnetic resonance angiography may be useful in ascertaining changes in the renal vasculature and in renal blood flow, even in the absence of changes in urinary flow. Thrombosis of the renal vein occurs rarely; it may be reversible if it is caused by technical factors and intervention is prompt. Diagnostic ultrasound is the procedure of choice to rule out urinary obstruction or to confirm the presence of perirenal collections of urine, blood, or lymph. When renal function has been good initially, a rise in the serum creatinine level is the most sensitive and reliable indicator of

possible rejection and may be the only sign.

Calcineurin inhibitors (cyclosporine and tacrolimus) may cause deterioration in renal function in a manner similar to a rejection episode. In fact, rejection processes tend to be more indolent with these inhibitors, and the only way to make a diagnosis may be by renal biopsy. Calcineurin inhibitors have an afferent arteriolar constrictor effect on the kidney and may produce permanent vascular and interstitial injury after sustained high-dose therapy. The addition of angiotensin-converting enzyme (ACE) inhibitors or nonsteroidal anti-inflammatory drugs is likely to raise serum creatinine levels. The former are generally safe to use after the early months, whereas the latter are best avoided in all renal transplant patients. There is no universally accepted lesion that makes a diagnosis of calcineurin inhibitor toxicity, although interstitial fibrosis, isometric tubular vacuolization, and thickening of arteriolar walls have been noted by some. Basically, if the biopsy does not reveal moderate and active cellular rejection activity, the serum creatinine most likely will respond to a reduction in dose. Blood levels of drug can be useful if they are very high or very low but do not correlate precisely with renal function, although serial changes in a patient can be useful. If rejection activity is present in the biopsy, appropriate therapy is indicated. The first rejection episode is usually treated with IV administration of methylprednisolone, 500–1000 mg daily for 3 days. Failure to respond is an indication for antibody therapy, usually with OKT3 or antithymocyte globulin.

Biopsy may be necessary to confirm the presence of rejection; when evidence of antibody-mediated injury is present with endothelial injury and deposition of complement component C4d is detected by fluorescence labeling, one can usually detect the antibody in recipient blood. The prognosis is poor, and aggressive use of plasmapheresis, immunoglobulin infusions, or anti-CD20 monoclonal antibody (rituximab) that targets B lymphocytes is indicated.

Management Problems

The usual clinical manifestations of infection in the posttransplant period are blunted by immunosuppressive therapy. The major toxic effect of azathioprine is bone marrow suppression, which is less likely with mycophenolic acid, whereas calcineurin inhibitors have no marrow effects. All drugs predispose to unusual opportunistic infections, however. The typical times after transplantation when the most common opportunistic infections occur are shown in **Table 282-5**. The signs and symptoms of infection may be masked or distorted. Fever without obvious cause is common, and only after days or weeks may it become apparent that it has a viral or fungal origin. Bacterial infections are most common during the first month after transplantation. The importance of blood cultures in such patients cannot be overemphasized because systemic infection without obvious foci is common, although wound infections with or without urinary fistulas are the most common. Particularly ominous are rapidly occurring pulmonary lesions, which may result in death within 5 days of onset. When these lesions become apparent, immunosuppressive agents should be discontinued, except for maintenance doses of prednisone.

Table 282-5 The Most Common Opportunistic Infections in Renal Transplant Recipients

Peritransplant (<1 month)	Late (>6 months)
Wound infections	*Aspergillus Nocardia*
Herpesvirus	BK virus (polyoma)
Oral candidiasis	Herpes zoster
Urinary tract infection	Hepatitis B
Early (1–6 months)	Hepatitis C
Pneumocystis carinii	
Cytomegalovirus	
Legionella Listeria	
Hepatitis B	
Hepatitis C	

Aggressive diagnostic procedures, including transbronchial and open-lung biopsy, are frequently indicated. In the case of *Pneumocystis carinii* (Chap. 207) infection, trimethoprim-sulfamethoxazole (TMP-SMX) is the treatment of choice; amphotericin B has been used effectively in systemic fungal infections. Prophylaxis against *P. carinii* with daily or alternate-day low-dose TMP-SMX is very effective. Involvement of the oropharynx with *Candida* (Chap. 203) may be treated with local nystatin. Tissue-invasive fungal infections require treatment with systemic agents such as fluconazole. Small doses (a total of 300 mg) of amphotericin given over a period of 2 weeks may be effective in fungal infections refractory to fluconazole. Macrolide antibiotics, especially ketoconazole and erythromycin, and some calcium channel blockers (diltiazem, verapamil) compete with calcineurin inhibitors for P450

catabolism and cause elevated levels of these immunosuppressive drugs. Analeptics, such as phenytoin and carbamazepine, will increase catabolism to result in low levels. *Aspergillus* (Chap. 204), *Nocardia* (Chap. 162), and especially cytomegalovirus (CMV) (Chap. 182) infections also occur.

CMV is a common and dangerous DNA virus in transplant recipients. It does not generally appear until the end of the first posttransplant month. Active CMV infection is sometimes associated, or occasionally confused, with rejection episodes. Patients at highest risk for severe CMV disease are those without anti-CMV antibodies who receive a graft from a CMV antibody–positive donor (15% mortality). Valganciclovir is a cost-effective and bioavailable oral form of ganciclovir that has been proved effective in both prophylaxis and treatment of CMV disease. Early diagnosis in a febrile patient with clinical suspicion of CMV disease can be made by determining CMV viral load in the blood. A rise in IgM antibodies to CMV is also diagnostic. Culture of CMV from blood may be less sensitive. Tissue invasion of CMV is common in the gastrointestinal tract and lungs. CMV retinopathy occurs late in the course, if untreated. Treatment of active CMV disease with valganciclovir is always indicated. In many patients immune to CMV, viral activation can occur with major immunosuppressive regimens.

The polyoma group (BK, JC, SV40) is another class of DNA viruses that can become dormant in kidneys and can be activated by immunosuppression. When reactivation occurs with BK, there is a 50% chance of progressive fibrosis and loss of the graft within 1 year by the activated virus. Risk of infection is associated with the overall degree of immunosuppression rather than the individual immunosuppressive drugs used. Renal biopsy is necessary for the diagnosis. There have been promising results with leflunomide, cidofovir, and quinolone anitibiotics (which are effective against polyoma helicase), but it is most important to reduce the immunosuppressive load.

The complications of glucocorticoid therapy are well known and include gastrointestinal bleeding, impairment of wound healing, osteoporosis, diabetes mellitus, cataract formation, and hemorrhagic pancreatitis. The treatment of unexplained jaundice in transplant patients should include cessation or reduction of immunosuppressive drugs if hepatitis or drug toxicity is suspected. Therapy in such circumstances often does not result in rejection of a graft, at least for several weeks. Acyclovir is effective in therapy for herpes simplex virus infections.

Chronic Lesions of the Transplanted Kidney

Although 1-year transplant survival is excellent, most recipients experience progressive decline in kidney function over time thereafter. Chronic renal transplant dysfunction can be caused by recurrent disease, hypertension, cyclosporine or tacrolimusnephrotoxicity, chronic immunologic rejection, secondary focal glomerulosclerosis, or a combination of these pathophysiologies. Chronic vascular changes with intimal proliferation and medial hypertrophy are commonly found. Control of systemic and intrarenal hypertension with ACE inhibitors is thought to have a beneficial influence on the rate of progression of chronic renal transplant dysfunction. Renal biopsy can distinguish subacute cellular rejection from recurrent disease or secondary focal sclerosis.

MALIGNANCY

The incidence of tumors in patients on immunosuppressive therapy is 5–6%, or approximately 100 times greater than that in the general population in the same age range. The most common lesions are cancer of the skin and lips and carcinoma in situ of the cervix, as well as lymphomas such as non-Hodgkin's lymphoma. The risks are increased in proportion to the total immunosuppressive load administered and the time elapsed since transplantation. Surveillance for skin and cervical cancers is necessary.

Other Complications

Hypercalcemia after transplantation may indicate failure of hyperplastic parathyroid glands to regress. Aseptic necrosis of the head of the femur is probably due to preexisting hyperparathyroidism, with aggravation by glucocorticoid treatment. With improved management of calcium and phosphorus metabolism during chronic dialysis, the incidence of parathyroid-related complications has fallen dramatically. Persistent hyperparathyroid activity may require subtotal parathyroidectomy.

Hypertension may be caused by (1) native kidney disease, (2) rejection activity in the transplant, (3) renal artery stenosis if an end-to-end anastomosis was constructed with an iliac artery branch, and (4) renal calcineurin inhibitor toxicity. This toxicity may improve with reduction in dose. Whereas ACE inhibitors may be useful, calcium channel blockers are more frequently used initially. Amelioration of hypertension to the range of 120–130/70–80 mmHg should be the goal in all patients.

Although most transplant patients have robust production of erythropoietin and normalization of hemoglobin, *anemia* is commonly seen in the posttransplant period. Often the anemia is attributable to bone marrow–suppressant immunosuppressive medications such as azathioprine, mycophenolic acid, and sirolimus. Gastrointestinal bleeding is a common side effect of high-dose and long-term steroid administration. Many transplant patients have creatinine clearances of 30–50 mL/min and can be considered in the same way as other patients with chronic renal insufficiency for anemia management, including supplemental erythropoietin.

Chronic hepatitis, particularly when due to hepatitis B virus, can be a progressive, fatal disease over a decade or so. Patients who are persistently hepatitis B surface antigen–positive are at higher risk, according to some studies, but the presence of hepatitis C virus is also a concern when one embarks on a course of immunosuppression in a transplant recipient.

Both chronic dialysis and renal transplant patients have a higher incidence of death from myocardial infarction and stroke than does the population at large, and this is particularly true in diabetic patients. Contributing factors are the use of glucocorticoids and sirolimus, as well as hypertension. Recipients of renal transplants have a high prevalence of coronary artery and peripheral vascular diseases. The percentage of deaths from these causes has been slowly rising as the numbers of transplanted diabetic patients and the average age of all recipients increase. More than 50% of renal recipient mortality is attributable to cardiovascular disease. In addition to strict control of blood pressure and blood lipid levels, close monitoring of patients for indications of further medical or surgical intervention is an important part of management.

FURTHER READINGS

Chandraker A et al: Transplantation immunobiology, in *Brenner and Rector's The Kidney* , 8th ed, B Brenner (ed). Philadelphia, Saunders, 2008, pp 2101–2126

Denton MD et al: Immunosuppressive strategies in renal transplantation. Lancet 353:1083, 1999[PMID: 10199367]

Kawai T et al: HLA-mismatched renal transplantation without maintenance immunosuppression. N Engl J Med 358:353, 2008[PMID: 18216355]

Li XC et al: Costimulatory pathways in transplantation: Challenges and new developments. Immunol Rev 229:271, 2009[PMID: 19426228]

Mahoney RJ et al: B-cell crossmatching and kidney allograft outcome in 9031 United States transplant recipients. Hum Immunol 63:324, 2002[PMID: 12039415]

Pescovitz MD, Govani m: Sirolimus and mycophenolate mofetil for calcineurin-free immunosuppression in renal transplant recipients. Am J Kidney Dis 38:S16, 2001

Sayegh MH, Carpenter CB: Transplantation 50 years later—progress, challenges and promises. N Engl J Med 351:2761, 2004[PMID: 15616214]

Vicenti f, Kirk AD: What's next in the pipeline. Am J Transplant 8:1972, 2008

Vo AA et al: Rituximab and intravenous immune globulin for desensitization during renal transplantation. N Engl J Med 359:242, 2008[PMID: 18635429]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GLOMERULAR DISEASES: INTRODUCTION

Two human kidneys harbor nearly 1.8 million glomerular capillary tufts. Each glomerular tuft resides within Bowman's space. The capsule circumscribing this space is lined by parietal epithelial cells that transition into tubular epithelia forming the proximal nephron or migrate into the tuft to replenish podocytes. The glomerular capillary tuft derives from an afferent arteriole that forms a branching capillary bed embedded in mesangial matrix (Fig. 283-1). This capillary network funnels into an efferent arteriole, which passes filtered blood into cortical peritubular capillaries or medullary vasa recta that supply and exchange with a folded tubular architecture. Hence the glomerular capillary tuft, fed and drained by arterioles, represents an arteriolar portal system. Fenestrated endothelial cells resting on a glomerular basement membrane (GBM) line glomerular capillaries. Delicate foot processes extending from epithelial podocytes shroud the outer surface of these capillaries, and podocytes interconnect to each other by slit-pore membranes forming a selective filtration barrier.

Figure 283-1

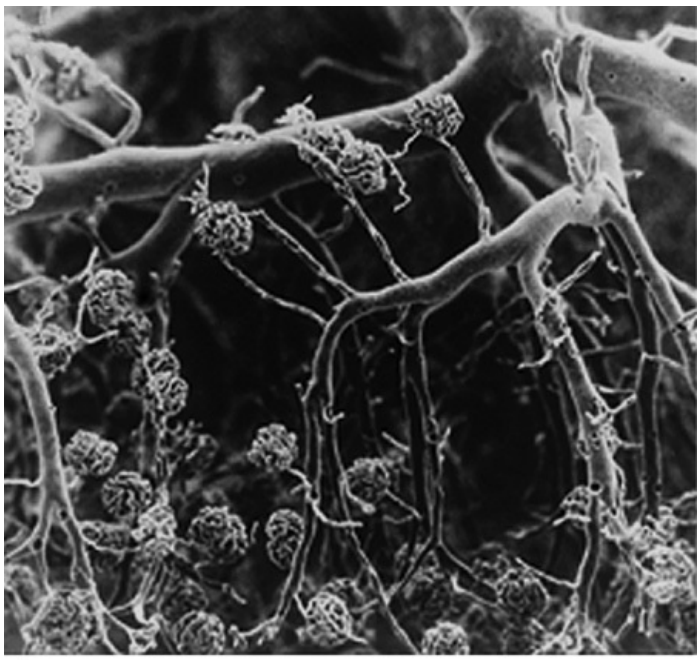

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

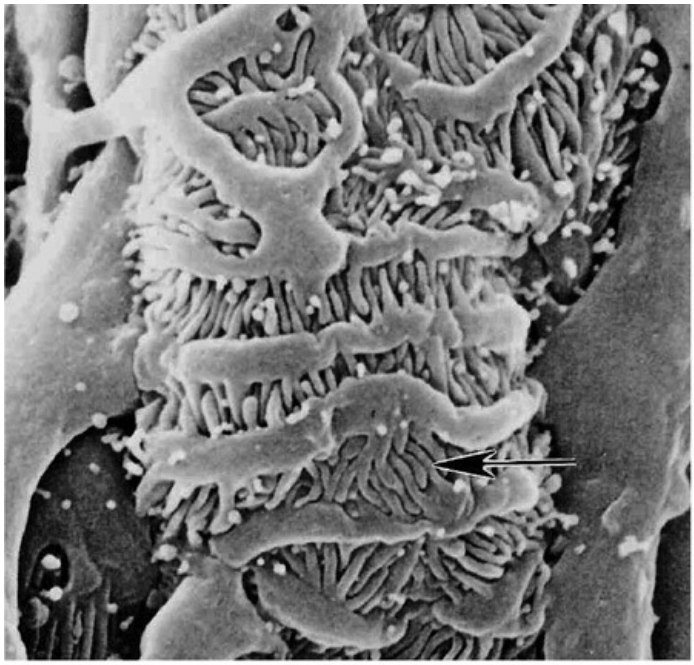

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

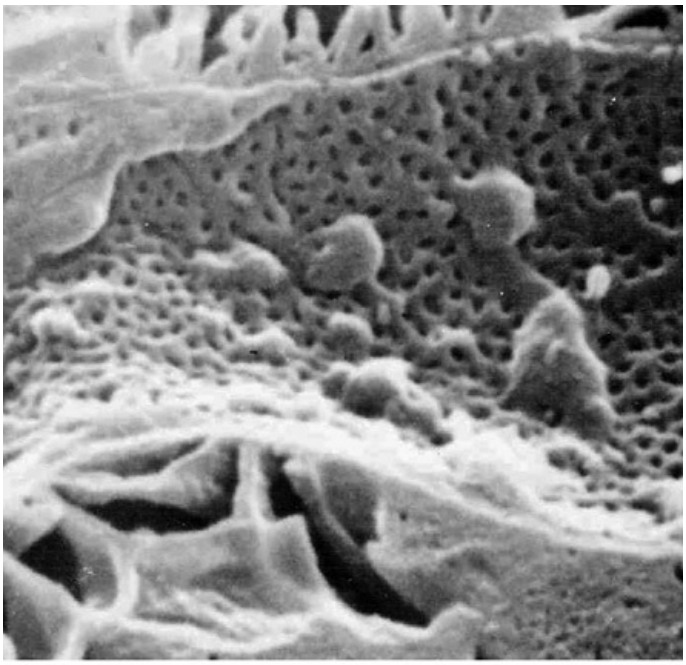

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

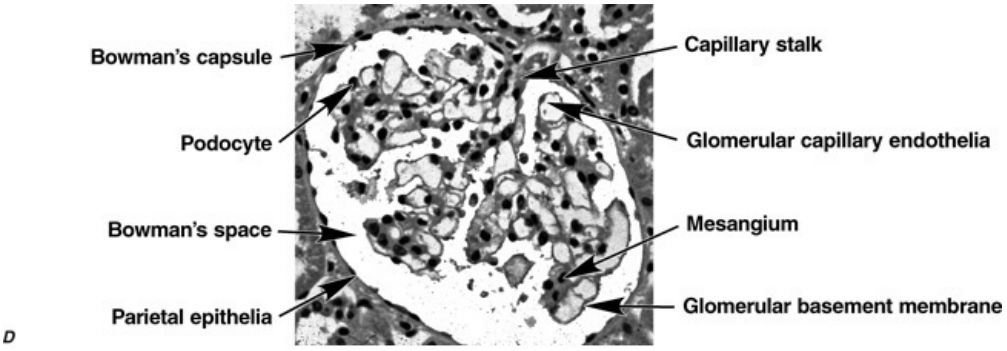

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Glomerular architecture. ***A.*** The glomerular capillaries form from a branching network of renal arteries, arterioles, leading to an afferent arteriole, glomerular capillary bed (tuft), and a draining efferent arteriole. (From VH Gattone II et al: Hypertension 5:8, 1983.) ***B.*** Scanning electron micrograph of podocytes that line the outer surface of the glomerular capillaries (arrow shows foot process). ***C.*** Scanning electron micrograph of the fenestrated endothelia lining the glomerular capillary. ***D.*** The various normal regions of the glomerulus on light microscopy *(A–C, courtesy of Dr. Vincent Gattone, Indiana University; with permission).*

The glomerular capillaries filter 120–180 L/d of plasma water containing various solutes for reclamation or discharge by downstream tubules. Most large proteins and all cells are excluded from filtration by a physicochemical barrier governed by pore size and negative electrostatic charge. The mechanics of filtration and reclamation are quite complicated for many solutes (Chap. 271). For example, in the case of serum albumin, the glomerulus is an imperfect barrier. Although albumin has a negative charge, which would tend to repel the negatively charged GBM, it only has a physical radius of 3.6 nm, while pores in the GBM and slit-pore membranes have a radius of 4 nm. Consequently, variable amounts of albumin inevitably cross the filtration barrier to be reclaimed by megalin and cubilin receptors along the proximal tubule. Remarkably, humans with normal nephrons do not excrete more than 8–10 mg of albumin in daily voided urine, approximately 20–60% of total excreted protein. This amount of albumin, and other proteins, can rise to gram quantities following glomerular injury.

The breadth of diseases affecting the glomerulus is expansive because the glomerular capillaries can be injured in a variety of ways, producing many different lesions and several unique changes to urinalysis. Some order to this vast subject is brought by grouping all of these diseases into a smaller number of clinical syndromes.

PATHOGENESIS OF GLOMERULAR DISEASE

There are many forms of glomerular disease with pathogenesis variably linked to the presence of genetic mutations, infection, toxin exposure, autoimmunity, atherosclerosis, hypertension, emboli, thrombosis, or diabetes mellitus. Even after careful study, however, the cause often remains unknown, and the lesion is called *idiopathic.* Specific or unique features of pathogenesis are mentioned with the description of each of the glomerular diseases later in this chapter.

Some glomerular diseases result from genetic mutations producing familial disease or a founder effect: congenital nephrotic syndrome from mutations in *NPHS1* (nephrin) and *NPHS2* (podocin) affect the slit-pore membrane at birth, and *TRPC6* cation channel mutations produce *focal segmental glomerulosclerosis(FSGS)* in adulthood; polymorphisms in the gene encoding apolipoprotein L1, APOL1 are a major risk for nearly 70% of African Americans with nondiabetic end-stage renal disease (ESRD), particularly FSGS; mutations in complement factor H associate with *membranoproliferative glomerulonephritis(MPGN)* or *atypical hemolytic uremic syndrome (aHUS),* type II partial lipodystrophy from mutations in genes encoding lamin A/C, or PPARγ cause a metabolic syndrome associated with MPGN, which is sometimes accompanied by dense deposits and C3 nephritic factor; Alport's syndrome, from mutations in the genes encoding for the α3, α4, or α5 chains of type IV collagen, produces *split-basement membranes* with *glomerulosclerosis;* and lysosomal storage diseases, such as α-galactosidase A deficiency causing Fabry's disease and *N*-acetylneuraminic acid hydrolase deficiency causing nephrosialidosis, produce FSGS.

Systemic hypertension and atherosclerosis can produce pressure stress, ischemia, or lipid oxidants that lead to *chronic glomerulosclerosis.Malignant hypertension* can quickly complicate glomerulosclerosis with fibrinoid necrosis of arterioles and glomeruli, thrombotic microangiopathy, and acute renal failure. *Diabetic nephropathy* is an acquired sclerotic injury associated with thickening of the GBM secondary to the long-standing effects of hyperglycemia, advanced glycosylation end products, and reactive oxygen species.

Inflammation of the glomerular capillaries is called *glomerulonephritis.* Most glomerular or mesangial antigens involved in *immune-mediated glomerulonephritis* are unknown (Fig. 283-2). Glomerular

epithelial or mesangial cells may shed or express epitopes that mimic other immunogenic proteins made elsewhere in the body. Bacteria, fungi, and viruses can directly infect the kidney producing their own antigens. Autoimmune diseases like idiopathic *membranous glomerulonephritis (MGN)* or MPGN are confined to the kidney, while systemic inflammatory diseases like *lupus nephritis* or *granulomatosis with polyangiitis (Wegener's)* spread to the kidney, causing secondary glomerular injury. *Antiglomerular basement membrane disease* producing Goodpasture's syndrome primarily injures both the lung and kidney because of the narrow distribution of the α3 NC1 domain of type IV collagen that is the target antigen.

Figure 283-2

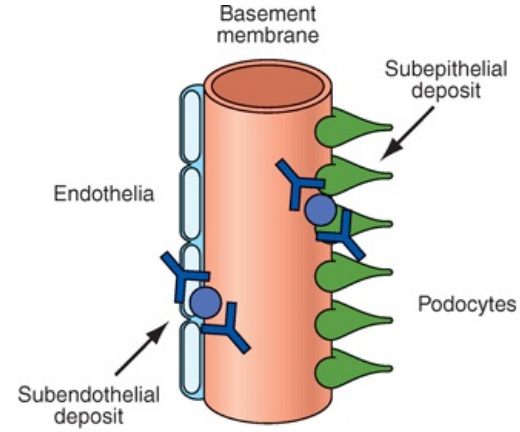

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

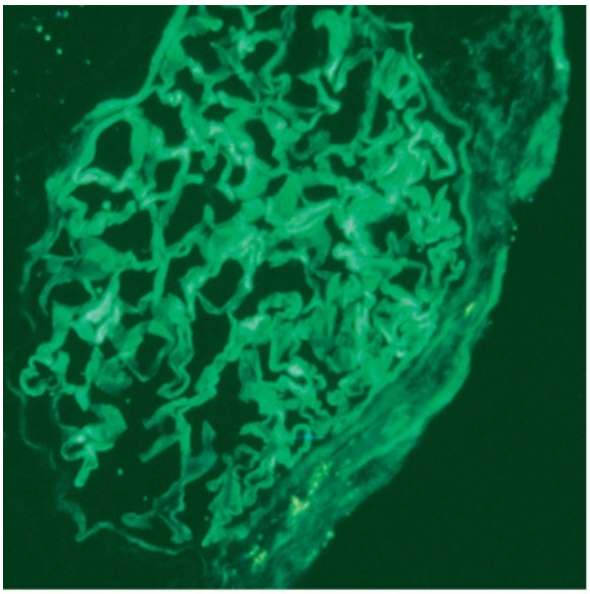

Linear IgG staining

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

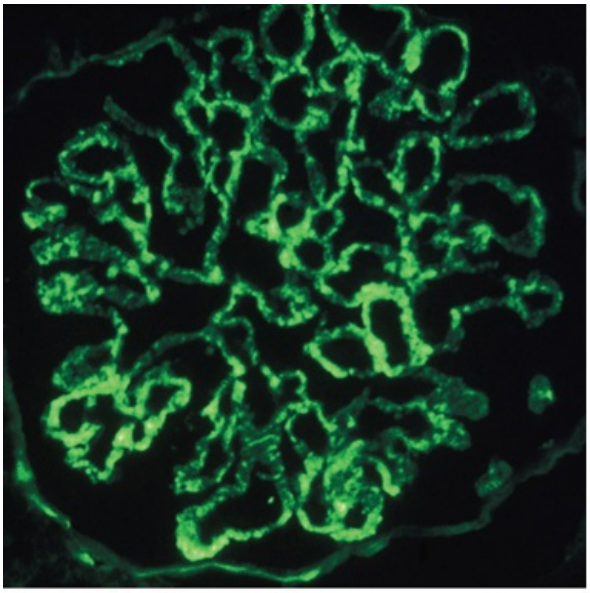

IgG Lumpy-bumpy staining

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

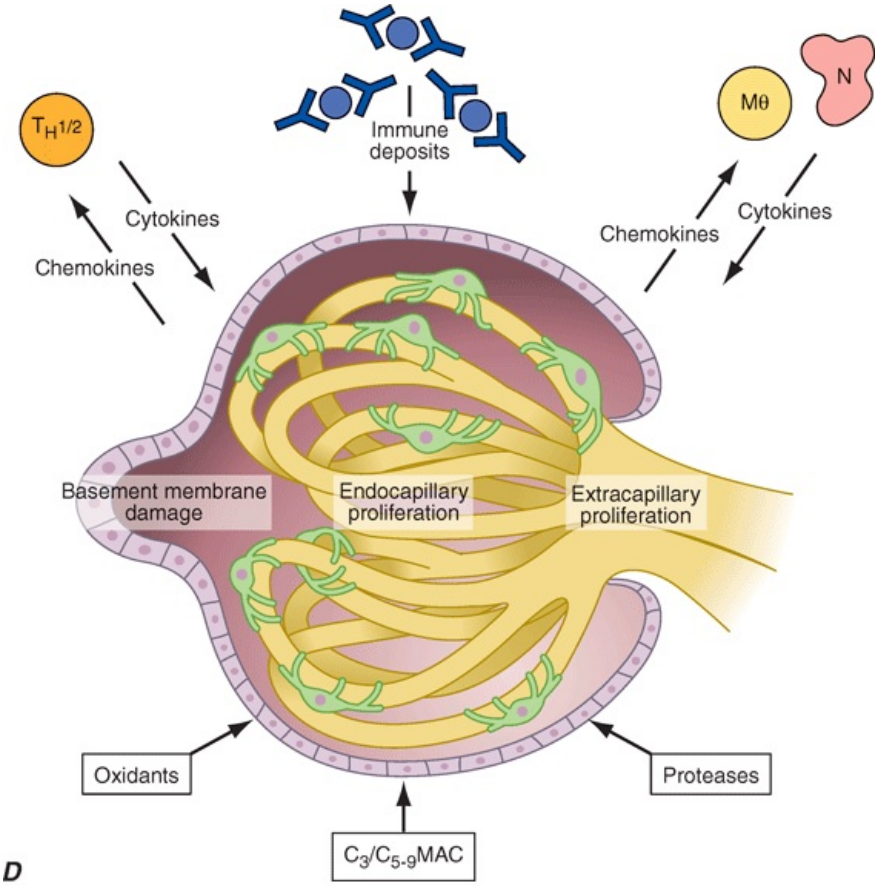

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The glomerulus is injured by a variety of mechanisms. ***A.*** Preformed immune deposits can precipitate from the circulation and collect along the glomerular basement membrane (GBM) in the subendothelial space or can form in situ along the subepithelial space. ***B.***Immunofluorescent staining of glomeruli with labeled anti-IgG demonstrating linear staining from a patient with anti-GBM disease or immune deposits from a patient with membranous glomerulonephritis. ***C.*** The mechanisms of glomerular injury have a complicated pathogenesis. Immune deposits and complement deposition classically draw macrophages and neutrophils into the glomerulus. T lymphocytes may follow to participate in the injury pattern as well. ***D.*** Amplification mediators as locally derived oxidants and

participate in the injury pattern as well. DNA replication mediators as locally derived oxidants and proteases expand this inflammation, and, depending on the location of the target antigen and the genetic polymorphisms of the host, basement membranes are damaged with either endocapillary or extracapillary proliferation.

Local activation of Toll-like receptors on glomerular cells, deposition of immune complexes, or complement injury to glomerular structures induces mononuclear cell infiltration, which subsequently leads to an adaptive immune response attracted to the kidney by local release of chemokines. Neutrophils, macrophages, and T cells are drawn by chemokines into the glomerular tuft, where they react with antigens and epitopes on or near somatic cells or their structures, producing more cytokines and proteases that damage the mesangium, capillaries, and/or the GBM. While the adaptive immune response is similar to that of other tissues, early T cell activation plays an important role in the mechanism of glomerulonephritis. Antigens presented by class II major histocompatibility complex (MHC) molecules on macrophages and dendritic cells in conjunction with associative recognition molecules engage the CD4/8 T cell repertoire.

Mononuclear cells by themselves can injure the kidney, but autoimmune events that damage glomeruli classically produce a humoral immune response. *Poststreptococcal glomerulonephritis, lupus nephritis,* and idiopathic *membranous nephritis* typically are associated with immune deposits along the GBM, while anti-GBM antibodies produce the linear binding of anti-GBM disease. Preformed circulating immune complexes can precipitate along the subendothelial side of the GBM, while other immune deposits form in situ on the subepithelial side. These latter deposits accumulate when circulating autoantibodies find their antigen trapped along the subepithelial edge of the GBM. Immune deposits in the glomerular mesangium may result from the deposition of preformed circulating complexes or in situ antigen-antibody interactions. Immune deposits stimulate the release of local proteases and activate the complement cascade, producing C_{5-9} attack complexes. In addition, local oxidants damage glomerular structures, producing proteinuria and effacement of the podocytes. Overlapping etiologies or pathophysiologic mechanisms can produce similar glomerular lesions, suggesting that downstream molecular and cellular responses often converge toward common patterns of injury.

PROGRESSION OF GLOMERULAR DISEASE

Persistent glomerulonephritis that worsens renal function is always accompanied by interstitial nephritis, renal fibrosis, and tubular atrophy (Fig. e14-27). What is not so obvious, however, is that renal failure in glomerulonephritis best correlates histologically with the appearance of tubulointerstitial nephritis rather than with the type of inciting glomerular injury.

Loss of renal function due to interstitial damage is explained hypothetically by several mechanisms. The simplest explanation is that urine flow is impeded by tubular obstruction as a result of interstitial inflammation and fibrosis. Thus, obstruction of the tubules with debris or by extrinsic compression results in aglomerular nephrons. A second mechanism suggests that interstitial changes, including interstitial edema or fibrosis, alter tubular and vascular architecture and thereby compromise the normal tubular transport of solutes and water from tubular lumen to vascular space. This failure increases the solute and water content of the tubule fluid, resulting in isosthenuria and polyuria. Adaptive mechanisms related to tubuloglomerular feedback also fail, resulting in a reduction of renin output from the juxtaglomerular apparatus trapped by interstitial inflammation. Consequently, the local vasoconstrictive influence of angiotensin II on the glomerular arterioles decreases, and filtration drops owing to a generalized decrease in arteriolar tone. A third mechanism involves changes in vascular resistance due to damage of peritubular capillaries. The cross-sectional volume of these capillaries is decreased by interstitial inflammation, edema, or fibrosis. These structural alterations in vascular resistance affect renal function through two mechanisms. First, tubular cells are very metabolically active, and, as a result, decreased perfusion leads to ischemic injury. Second, impairment of glomerular arteriolar outflow leads to increased intraglomerular hypertension in less-involved glomeruli; this selective intraglomerular hypertension aggravates and extends *mesangial sclerosis* and *glomerulosclerosis* to less-involved glomeruli. Regardless of the exact mechanism, early *acute tubulointerstitial nephritis* (Fig. e14-27) suggests potentially recoverable renal function, while the development of *chronic interstitial fibrosis* prognosticates permanent loss (Fig. e14-30).

Persistent damage to glomerular capillaries spreads to the tubulointerstitium in association with proteinuria. There is an untested hypothesis that efferent arterioles leading from inflamed glomeruli carry forward inflammatory mediators, which induces downstream interstitial nephritis, resulting in fibrosis. Glomerular filtrate from injured glomerular capillaries adherent to Bowman's capsule may also be misdirected to the periglomerular interstitium. Most nephrologists believe, however, that proteinuric glomerular filtrate forming tubular fluid is the primary route to downstream tubulointerstitial injury, although none of these hypotheses are mutually exclusive.

The simplest explanation for the effect of proteinuria on the development of interstitial nephritis is that increasingly severe proteinuria, carrying activated cytokines and lipoproteins producing reactive oxygen species, triggers a downstream inflammatory cascade in and around epithelial cells lining the tubular nephron. These effects induce T lymphocyte and macrophage infiltrates in the interstitial

spaces along with fibrosis and tubular atrophy.

Tubules disaggregate following direct damage to their basement membranes, leading to epithelial-mesenchymal transitions forming more interstitial fibroblasts at the site of injury. Transforming growth factor β (TGF-β), fibroblast growth factor 2 (FGF-2), hypoxemia-inducible factor 1α (HIF-1α), and platelet-derived growth factor (PDGF) are particularly active in this transition. With persistent nephritis, fibroblasts multiply and lay down tenascin and a fibronectin scaffold for the polymerization of new interstitial collagen types I/III. These events form scar tissue through a process called fibrogenesis. In experimental studies, bone morphogenetic protein 7 and hepatocyte growth factor can reverse early fibrogenesis and preserve tubular architecture. When fibroblasts outdistance their survival factors, apoptoses occurs, and the permanent renal scar becomes acellular, leading to irreversible renal failure.

Approach to the Patient: Glomerular Disease

HEMATURIA, PROTEINURIA, AND PYURIA

Patients with glomerular disease usually have some hematuria with varying degrees of proteinuria. Hematuria is typically asymptomatic. As few as three to five red blood cells in the spun sediment from first-voided morning urine is suspicious. The diagnosis of glomerular injury can be delayed because patients will not realize they have *microscopic hematuria,* and only rarely with the exception of IgA nephropathy and sickle cell disease is *gross hematuria* present. When working up microscopic hematuria, perhaps accompanied by minimal proteinuria (<500 mg/24 h), it is important to exclude anatomic lesions, such as malignancy of the urinary tract, particularly in older men. Microscopic hematuria may also appear with the onset of benign prostatic hypertrophy, interstitial nephritis, papillary necrosis, hypercalciuria, renal stones, cystic kidney diseases, or renal vascular injury. However, when red blood cell casts (Fig. e14-34) or dysmorphic red blood cells are found in the sediment, glomerulonephritis is likely.

Sustained proteinuria >1–2 g/24 h is also commonly associated with glomerular disease. Patients often will not know they have proteinuria unless they become edematous or notice foaming urine on voiding. *Sustained proteinuria* has to be distinguished from lesser amounts of so-called *benign proteinuria* in the normal population (Table 283-1). This latter class of proteinuria is nonsustained, generally <1 g/24 h, and is sometimes called *functional* or *transient proteinuria.* Fever, exercise, obesity, sleep apnea, emotional stress, and congestive heart failure can explain transient proteinuria. Proteinuria only seen with upright posture is called *orthostatic proteinuria* and has a benign prognosis. Isolated proteinuria sustained over multiple clinic visits is found in diabetic nephropathy, *nil lesion, mesangioproliferative glomerulonephritis,* and FSGS. Proteinuria in most adults with glomerular disease is *nonselective,* containing albumin and a mixture of other serum proteins, while in children with nil lesion from *minimal change disease,* the proteinuria is *selective* and composed largely of albumin.

Table 283-1 Urine Assays for Albuminuria/Proteinuria

	24-Hour Albumin[a] (mg/24 h)	Albumin[a]/Creatinine Ratio (mg/g)	Dipstick Proteinuria	24-Hour Urine Protein[b] (mg/24 h)
Normal	8–10	<30	–	<150
Microalbuminuria	30–300	30–300	–/Trace/1+	–
Proteinuria	>;300	>;300	Trace–3+	>;150

[a]Albumin detected by radioimmunoassay.

[b]Albumin represents 30–70% of the total protein excreted in the urine.

Some patients with inflammatory glomerular disease, such as acute poststreptococcal glomerulonephritis or MPGN, have *pyuria* characterized by the presence of considerable numbers of leukocytes. This latter finding has to be distinguished from urine infected with bacteria.

CLINICAL SYNDROMES

Various forms of glomerular injury can also be parsed into several distinct syndromes on clinical grounds (Table 283-2). These syndromes, however, are not always mutually exclusive. There is an *acute nephritic syndrome* producing 1–2 g/24 h of proteinuria, hematuria with red blood cell casts, pyuria, hypertension, fluid retention, and a rise in serum creatinine associated with a reduction in glomerular filtration. If glomerular inflammation develops slowly, the serum creatinine will rise gradually over many weeks, but if the serum creatinine rises quickly, particularly over a few days, acute nephritis is sometimes called *rapidly progressive glomerulonephritis* (RPGN); the histopathologic term *crescentic glomerulonephritis* is the pathologic equivalent of the clinical presentation of RPGN. When

patients with RPGN present with lung hemorrhage from Goodpasture's syndrome, antineutrophil cytoplasmic antibodies (ANCA)-associated small-vessel vasculitis, lupus erythematosus, or cryoglobulinemia, they are often diagnosed as having a *pulmonary-renal syndrome. Nephrotic syndrome* describes the onset of heavy proteinuria (>3.0 g/24 h), hypertension, hypercholesterolemia, hypoalbuminemia, edema/anasarca, and microscopic hematuria; if only large amounts of proteinuria are present without clinical manifestations, the condition is sometimes called *nephrotic-range proteinuria.* The glomerular filtration rate (GFR) in these patients may initially be normal or, rarely, higher than normal, but with persistent hyperfiltration and continued nephron loss, it typically declines over months to years. Patients with a *basement membrane syndrome* either have genetically abnormal basement membranes (Alport's syndrome) or an autoimmune response to basement membrane collagen IV (Goodpasture's syndrome) associated with microscopic hematuria, mild to heavy proteinuria, and hypertension with variable elevations in serum creatinine. *Glomerular-vascular syndrome* describes patients with vascular injury producing hematuria and moderate proteinuria. Affected individuals can have vasculitis, thrombotic microangiopathy, antiphospholipid syndrome, or, more commonly, a systemic disease such as atherosclerosis, cholesterol emboli, hypertension, sickle cell anemia, and autoimmunity. *Infectious disease-associated syndrome* is most important if one has an international perspective. Save for subacute bacterial endocarditis in the Western Hemisphere, malaria and schistosomiasis may be the most common causes of glomerulonephritis throughout the world, closely followed by HIV and chronic hepatitis B and C. These infectious diseases produce a variety of inflammatory reactions in glomerular capillaries, ranging from nephrotic syndrome to acute nephritic injury, and urinalyses that demonstrate a combination of hematuria and proteinuria.

Table 283-2 Patterns of Clinical Glomerulonephritis

Glomerular Syndromes	Proteinuria	Hematuria	Vascular Injury
Acute Nephritic Syndromes			
Poststreptococcal glomerulonephritis[a]	+/++	++/+++	–
Subacute bacterial endocarditis[a]	+/++	++	–
Lupus nephritis[a]	+/++	++/+++	–
Antiglomerular basement membrane disease[a]	++	++/+++	–
IgA nephropathy[a]	+/++	++/+++[c]	–
ANCA small-vessel vasculitis[a]			
Granulomatosis with polyangiitis (Wegener's)	+/++	++/+++	++++
Microscopic polyangiitis	+/++	++/+++	++++
Churg-Strauss syndrome	+/++	++/+++	++++
Henoch-Schönlein purpura[a]	+/++	++/+++	++++
Cryoglobulinemia[a]	+/++	++/+++	++++
Membranoproliferative glomerulonephritis[a]	++	++/+++	–
Mesangioproliferative glomerulonephritis	+	+/++	–
Pulmonary-Renal Syndromes			
Goodpasture's syndrome[a]	++	++/+++	–
ANCA small-vessel vasculitis[a]			
Granulomatosis with polyangiitis (Wegener's)	+/++	++/+++	++++
Microscopic polyangiitis	+/++	++/+++	++++
Churg-Strauss syndrome	+/++	++/+++	++++
Henoch-Schönlein purpura[a]	+/++	++/+++	++++

Cryoglobulinemia[a]	+/++	++/+++	++++
Nephrotic Syndromes			
Minimal change disease	++++	–	–
Focal segmental glomerulosclerosis	+++/++++	+	–
Membranous glomerulonephritis	++++	+	–
Diabetic nephropathy	++/++++	–/+	–
AL and AA amyloidosis	+++/++++	+	+/++
Light-chain deposition disease	+++	+	–
Fibrillary-immunotactoid disease	+++/++++	+	+
Fabry's disease	+	+	–
Basement Membrane Syndromes			
Anti-GBM disease[a]	++	++/+++	–
Alport's syndrome	++	++	–
Thin basement membrane disease	+	++	–
Nail-patella syndrome	++/+++	++	–
Glomerular Vascular Syndromes			
Atherosclerotic nephropathy	+	+	+++
Hypertensive nephropathy[b]	+/++	+/++	++
Cholesterol emboli	+/++	++	+++
Sickle cell disease	+/++	++[c]	+++
Thrombotic microangiopathies	++	++	+++
Antiphospholipid syndrome	++	++	+++
ANCA small-vessel vasculitis[a]			
Granulomatosis with polyangiitis (Wegener's)	+/++	++/+++	++++
Microscopic polyangiitis	+/++	++/+++	++++
Churg-Strauss syndrome	+++	++/+++	++++
Henoch-Schönlein purpura[a]	+/++	++/+++	++++
Cryoglobulinemia[a]	+/++	++/+++	++++
AL and AA amyloidosis	+++/++++	+	+/++
Infectious Disease–Associated Syndromes			
Poststreptococcal glomerulonephritis[a]	+/++	++/+++	–
Subacute bacterial endocarditis[a]	+/++	++	–
HIV	+++	+/++	–
Hepatitis B and C	+++	+/++	–
Syphilis	+++	+	–
Leprosy	+++	+	–
Malaria	+++	+/++	–
Schistosomiasis	+++	+/++	–

[a]Can present as rapidly progressive glomerulonephritis (RPGN); sometimes called crescentic glomerulonephritis.

[b]Can present as a malignanthypertensive crisis producing an aggressive fibrinoid necrosis in arterioles and small arteries with microangiopathic hemolytic anemia.

[c]Can present with gross hematuria.

Abbreviations: AA, amyloid A; AL, amyloid L; ANCA, antineutrophil cytoplasmic antibodies; GBM, glomerular basement membrane.

These six general categories of syndromes are usually determined at the bedside with the help of a history and physical examination, blood chemistries, renal ultrasound, and urinalysis. These initial studies help frame further diagnostic workup that typically involves some testing of the serum for the presence of various proteins (HIV and hepatitis B and C antigens), antibodies [anti-GBM, antiphospholipid, antistreptolysin O (ASO), anti-DNAse, antihyaluronidase, ANCA, anti-DNA, cryoglobulins, anti-HIV, and anti-hepatitis B and C antibodies] or depletion of complement components (C_3 and C_4). The bedside history and physical examination can also help determine whether the glomerulonephritis is isolated to the kidney (*primary glomerulonephritis*) or is part of a systemic disease (*secondary glomerulonephritis*).

When confronted with an abnormal urinalysis and elevated serum creatinine, with or without edema or congestive heart failure, one must consider whether the glomerulonephritis is *acute* or *chronic.* This assessment is best made by careful history (last known urinalysis or serum creatinine during pregnancy or insurance physical, evidence of infection, or use of medication or recreational drugs); the size of the kidneys on renal ultrasound examination; and how the patient feels at presentation. Chronic glomerular disease often presents with decreased kidney size. Patients who quickly develop renal failure are fatigued and weak; feel miserable; often have uremic symptoms associated with nausea, vomiting, fluid retention, and somnolence. Primary glomerulonephritis presenting with renal failure that has progressed slowly, however, can be remarkably asymptomatic, as are patients with acute glomerulonephritis without much loss in renal function. Once this initial information is collected, selected patients who are clinically stable, have adequate blood clotting parameters, and are willing and able to receive treatment are encouraged to have a renal biopsy. Biopsies can be done safely with an ultrasound-guided biopsy gun.

Renal Pathology

A renal biopsy in the setting of glomerulonephritis quickly identifies the type of glomerular injury and often suggests a course of treatment. The biopsy is processed for light microscopy using stains for *hematoxylin and eosin (H&E)* to assess cellularity and architecture, *periodic acid–Schiff (PAS)* to stain carbohydrate moieties in the membranes of the glomerular tuft and tubules, *Jones-methenamine silver* to enhance basement membrane structure, *Congo red* for amyloid deposits, and *Masson's trichrome* to identify collagen deposition and assess the degree of glomerulosclerosis and interstitial fibrosis. Biopsies are also processed for direct immunofluorescence using conjugated antibodies against IgG, IgM, and IgA to detect the presence of "lumpy-bumpy" immune deposits or "linear" IgG or IgA antibodies bound to GBM, antibodies against trapped complement proteins (C_3 and C_4), or specific antibodies against a relevant antigen. High-resolution electron microscopy can clarify the principal location of immune deposits and the status of the basement membrane.

Each region of a renal biopsy is assessed separately. By light microscopy, glomeruli (at least 10 and ideally 20) are reviewed individually for discrete lesions; <50% involvement is considered *focal,* and >50% is *diffuse.* Injury in each glomerular tuft can be *segmental,* involving a portion of the tuft, or *global,* involving most of the glomerulus. Glomeruli having *proliferative* characteristics show increased cellularity. When cells in the capillary tuft proliferate, it is called *endocapillary,* and when cellular proliferation extends into Bowman's space, it is called *extracapillary.Synechiae* are formed when epithelial podocytes attach to Bowman's capsule in the setting of glomerular injury; *crescents,* which in some cases may be the extension of synechiae, develop when fibrocellular/fibrin collections fill all or part of Bowman's space; and *sclerotic* glomeruli show acellular, amorphous accumulations of proteinaceous material throughout the tuft with loss of functional capillaries and normal mesangium. Since *age-related glomerulosclerosis* is common in adults, one can estimate the background percentage of sclerosis by dividing the patient's age in half and subtracting 10. Immunofluorescent and electron microscopy can detect the presence and location of *subepithelial,subendothelial, or mesangial* immune deposits, or *reduplication* or *splitting* of the basement membrane. In the other regions of the biopsy, the vasculature surrounding glomeruli and tubules can show *angiopathy, vasculitis,* the presence of *fibrils,* or *thrombi.* The tubules can be assessed for adjacency to one another; separation can be the result of edema, tubular dropout, or collagen deposition resulting from interstitial fibrosis. Interstitial fibrosis is an ominous sign of irreversibility and progression to renal failure.

ACUTE NEPHRITIC SYNDROMES

Acute nephritic syndromes classically present with hypertension, hematuria, red blood cell casts, pyuria, and mild to moderate proteinuria. Extensive inflammatory damage to glomeruli causes a fall in GFR and eventually produces uremic symptoms with salt and water retention, leading to edema and hypertension.

Poststreptococcal Glomerulonephritis

Poststreptococcal glomerulonephritis is prototypical for *acute endocapillary proliferative glomerulonephritis.* The incidence of poststreptococcal glomerulonephritis has dramatically decreased in developed countries and in these locations is typically sporadic; epidemics are less common. Acute poststreptococcal glomerulonephritis in underdeveloped countries usually affects children between the ages of 2 and 14 years, but in developed countries is more typical in the elderly, especially in association with debilitating conditions. It is more common in males, and the familial or cohabitant incidence is as high as 40%. Skin and throat infections with particular M types of streptococci (nephritogenic strains) antedate glomerular disease; M types 47, 49, 55, 2, 60, and 57 are seen following impetigo and M types 1, 2, 4, 3, 25, 49, and 12 with pharyngitis. Poststreptococcal glomerulonephritis due to impetigo develops 2–6 weeks after skin infection and 1–3 weeks after streptococcal pharyngitis.

The renal biopsy in poststreptococcal glomerulonephritis demonstrates hypercellularity of mesangial and endothelial cells, glomerular infiltrates of polymorphonuclear leukocytes, granular subendothelial immune deposits of IgG, IgM, C_3, C_4, and $C_{5\text{-}9}$, and subepithelial deposits (which appear as "humps") (Fig. e14-6). (See Glomerular Schematic 1.) Poststreptococcal glomerulonephritis is an immune-mediated disease involving putative streptococcal antigens, circulating immune complexes, and activation of complement in association with cell-mediated injury. Many candidate antigens have been proposed over the years; candidates from nephritogenic streptococci of interest at the moment are: a cationic cysteine proteinase known as streptococcal pyrogenic exotoxin B (SPEB) that is generated by proteolysis of a zymogen precursor (zSPEB) and, NAPlr, the nephritis-associated plasmin receptor. These two antigens have biochemical affinity for plasmin, bind as complexes facilitated by this relationship, and both activate the alternate complement pathway. The nephritogenic antigen, SPEB, has been demonstrated inside the subepithelial "humps" on biopsy.

The classic presentation is an acute nephritic picture with hematuria, pyuria, red blood cell casts, edema, hypertension, and oliguric renal failure, which may be severe enough to appear as RPGN. Systemic symptoms of headache, malaise, anorexia, and flank pain (due to swelling of the renal capsule) are reported in as many as 50% of cases. Five percent of children and 20% of adults have proteinuria in the nephrotic range. In the first week of symptoms, 90% of patients will have a depressed CH_{50} and decreased levels of C_3 with normal levels of C_4. Positive rheumatoid factor (30–40%), cryoglobulins and circulating immune complexes (60–70%), and ANCA against myeloperoxidase (10%) are also reported. Positive cultures for streptococcal infection are inconsistently present (10–70%), but increased titers of ASO (30%), anti-DNAse, (70%), or antihyaluronidase antibodies (40%) can help confirm the diagnosis. Consequently, the diagnosis of poststreptococcal glomerulonephritis rarely requires a renal biopsy. A subclinical disease is reported in some series to be four to five times as common as clinical nephritis, and these latter cases are characterized by asymptomatic microscopic hematuria with low serum C_3 complement levels.

Treatment is supportive, with control of hypertension, edema, and dialysis as needed. Antibiotic treatment for streptococcal infection should be given to all patients and their cohabitants. There is no role for immunosuppressive therapy, even in the setting of crescents. Recurrent poststreptococcal glomerulonephritis is rare despite repeated streptococcal infections. Early death is rare in children but does occur in the elderly. Overall, the prognosis is good, with permanent renal failure being very uncommon, less than 1% in children. Complete resolution of the hematuria and proteinuria in the majority of children occurs within 3–6 weeks of the onset of nephritis but 3–10% of children may have persistent microscopic hematuria, non-nephrotic proteinuria, or hypertension. The prognosis in elderly patients is worse with a high incidence of azotemia (up to 60%), nephrotic-range proteinuria, and end-stage renal disease.

Subacute Bacterial Endocarditis

Endocarditis-associated glomerulonephritis is typically a complication of subacute bacterial endocarditis, particularly in patients who remain untreated for a long time, have negative blood cultures, or have right-sided endocarditis. Glomerulonephritis is unusual in acute bacterial endocarditis because it takes 10–14 days to develop immune complex–mediated injury, by which time the patient has been treated, often with emergent surgery. Grossly, the kidneys in subacute bacterial endocarditis have subcapsular hemorrhages with a "flea-bitten" appearance, and microscopy on renal biopsy reveals focal proliferation around foci of necrosis associated with abundant mesangial, subendothelial, and subepithelial immune deposits of IgG, IgM, and C_3. Patients who present with a clinical picture of RPGN have crescents. Embolic infarcts or septic abscesses may also be present. The pathogenesis hinges on the renal deposition of circulating immune complexes in the kidney with complement activation. Patients present with gross or microscopic hematuria, pyuria, and mild proteinuria or, less commonly, RPGN with rapid loss of renal function. A normocytic anemia, elevated erythrocyte sedimentation rate, hypocomplementemia, high titers of rheumatoid factor, type III cryoglobulins, and circulating immune complexes are often present. Levels of serum creatinine may be elevated at diagnosis, but with modern therapy there is little progression to chronic renal failure. Primary treatment is eradication of the infection with 4–6 weeks of antibiotics, and if accomplished expeditiously, the prognosis for renal recovery is good. ANCA-associated vasculitis sometimes accompanies or is

confused with subacute bacterial endocarditis (SBE) and should be ruled out, as the treatment is different.

As variants of persistent bacterial infection in blood, glomerulonephritis can occur in patients with ventriculoatrial and ventriculoperitoneal shunts; pulmonary, intraabdominal, pelvic, or cutaneous infections; and infected vascular prostheses. The clinical presentation of these conditions is variable and includes proteinuria, microscopic hematuria, and acute renal failure. Blood cultures are usually positive and serum complement levels low, and there may be elevated levels of C-reactive proteins, rheumatoid factor, antinuclear antibodies, and cryoglobulins. Renal lesions include membranoproliferative glomerulonephritis (MPGN), diffuse proliferative glomerulonephritis (DPGN), or mesangioproliferative glomerulonephritis, sometimes leading to RPGN. Treatment focuses on eradicating the infection, with most patients treated as if they have endocarditis.

Lupus Nephritis

Lupus nephritis is a common and serious complication of systemic lupus erythematosus (SLE) and most severe in African-American female adolescents. Thirty to fifty percent of patients will have clinical manifestations of renal disease at the time of diagnosis, and 60% of adults and 80% of children develop renal abnormalities at some point in the course of their disease. Lupus nephritis results from the deposition of circulating immune complexes, which activate the complement cascade leading to complement-mediated damage, leukocyte infiltration, activation of procoagulant factors, and release of various cytokines. In situ immune complex formation following glomerular binding of nuclear antigens, particularly necrotic nucleosomes, also plays a role in renal injury. The presence of antiphospholipid antibodies may also trigger a thrombotic microangiopathy in a minority of patients.

The clinical manifestations, course of disease, and treatment of lupus nephritis are closely linked to renal pathology. The most common clinical sign of renal disease is proteinuria, but hematuria, hypertension, varying degrees of renal failure, and active urine sediment with red blood cell casts can all be present. Although significant renal pathology can be found on biopsy even in the absence of major abnormalities in the urinalysis, most nephrologists do not biopsy patients until the urinalysis is convincingly abnormal. The extrarenal manifestations of lupus are important in establishing a firm diagnosis of systemic lupus because, while serologic abnormalities are common in lupus nephritis, they are not diagnostic. Anti-dsDNA antibodies that fix complement correlate best with the presence of renal disease. Hypocomplementemia is common in patients with acute lupus nephritis (70–90%) and declining complement levels may herald a flare. Although urinary biomarkers of lupus nephritis are being identified to assist in predicting renal flares, renal biopsy is the only reliable method of identifying the morphologic variants of lupus nephritis.

The World Health Organization (WHO) workshop in 1974 first outlined several distinct patterns of lupus-related glomerular injury; these were modified in 1982. In 2004 the International Society of Nephrology in conjunction with the Renal Pathology Society again updated the classification. This latest version of lesions seen on biopsy (Table 283-3) best defines clinicopathologic correlations, provides valuable prognostic information, and forms the basis for modern treatment recommendations. Class I nephritis describes normal glomerular histology by any technique or normal light microscopy with minimal mesangial deposits on immunofluorescent or electron microscopy. Class II designates mesangial immune complexes with *mesangial proliferation.* Both class I and II lesions are typically associated with minimal renal manifestation and normal renal function; nephrotic syndrome is rare. Patients with lesions limited to the renal mesangium have an excellent prognosis and generally do not need therapy for their lupus nephritis.

Table 283-3 Classification for Lupus Nephritis

Class I	Minimal mesangial	Normal histology with mesangial deposits
Class II	Mesangial proliferation	Mesangial hypercellularity with expansion of the mesangial matrix
Class III	Focal nephritis	Focal endocapillary ± extracapillary proliferation with focal subendothelial immune deposits and mild mesangial expansion
Class IV	Diffuse nephritis	Diffuse endocapillary ± extracapillary proliferation with diffuse subendothelial immune deposits and mesangial alterations
Class V	Membranous nephritis	Thickened basement membranes with diffuse subepithelial immune deposits; may occur with class III or IV lesions and is sometimes called mixed membranous and proliferative nephritis
Class VI	Sclerotic nephritis	Global sclerosis of nearly all glomerular capillaries

Note: Revised in 2004 by the International Society of Nephrology-Renal Pathology Society Study Group.

The subject of lupus nephritis is presented under acute nephritic syndromes because of the aggressive and important proliferative lesions seen in class III–V renal disease. Class III describes *focal lesions with proliferation or scarring,* often involving only a segment of the glomerulus (Fig. e14-12). Class III lesions have the most varied course. Hypertension, an active urinary sediment, and proteinuria are common with nephrotic-range proteinuria in 25–33% of patients. Elevated serum creatinine is present in 25% of patients. Patients with mild proliferation involving a small percentage of glomeruli respond well to therapy with steroids alone, and fewer than 5% progress to renal failure over 5 years. Patients with more severe proliferation involving a greater percentage of glomeruli have a far worse prognosis and lower remission rates. Treatment of those patients is the same as that for class IV lesions. Most nephrologists believe that class III lesions are simply an early presentation of class IV disease. Others believe severe class III disease is a discrete lesion also requiring aggressive therapy. Class IV describes *global, diffuse proliferative lesions* involving the vast majority of glomeruli. Patients with class IV lesions commonly have high anti-DNA antibody titers, low serum complement, hematuria, red blood cell casts, proteinuria, hypertension, and decreased renal function; 50% of patients have nephrotic-range proteinuria. Patients with crescents on biopsy often have a rapidly progressive decline in renal function (Fig. e14-12). Without treatment, this aggressive lesion has the worst renal prognosis. However, if a remission—defined as a return to near-normal renal function and proteinuria ≤330 mg/dL per day—is achieved with treatment, renal outcomes are excellent. Current evidence suggests that inducing a remission with administration of high-dose steroids and either cyclophosphamide or mycophenolate mofetil for 2–6 months, followed by maintenance therapy with lower doses of steroids and mycophenolate mofetil, best balances the likelihood of successful remission with the side effects of therapy. There is no consensus on use of high-dose intravenous methylprednisolone versus oral prednisone, monthly intravenous cyclophosphamide versus daily oral cyclophosphamide, or other immunosuppressants such as cyclosporine, tacrolimus, rituximab, or azathioprine. Nephrologists tend to avoid prolonged use of cyclophosphamide in patients of childbearing age without first banking eggs or sperm.

The class V lesion describes subepithelial immune deposits producing a *membranous pattern;* a subcategory of class V lesions is associated with proliferative lesions and is sometimes called *mixed membranous and proliferative disease* (Fig. e14-11); this category of injury is treated like class IV glomerulonephritis. Sixty percent of patients present with nephrotic syndrome or lesser amounts of proteinuria. Patients with lupus nephritis class V, like patients with *idiopathic membranous nephropathy,* are predisposed to renal-vein thrombosis and other thrombotic complications. A minority of patients with class V will present with hypertension and renal dysfunction. There are conflicting data on the clinical course, prognosis, and appropriate therapy for patients with class V disease, which may reflect the heterogeneity of this group of patients. Patients with severe nephrotic syndrome, elevated serum creatinine, and a progressive course will probably benefit from therapy with steroids in combination with other immunosuppressive agents. Therapy with inhibitors of the renin-angiotensin system also may attenuate the proteinuria. Antiphospholipid antibodies present in lupus may result in glomerular microthromboses and complicate the course in up to 20% of lupus nephritis patients. The renal prognosis is worse even with anticoagulant therapy.

Patients with any of the above lesions also can transform to another lesion; hence patients often require reevaluation, including repeat renal biopsy. Lupus patients with class VI lesions have greater than 90% *sclerotic glomeruli* and end-stage renal disease with interstitial fibrosis. As a group, approximately 20% of patients with lupus nephritis will reach end-stage disease, requiring dialysis or transplantation. Systemic lupus tends to become quiescent once there is renal failure, perhaps due to the immunosuppressant effects of uremia. Renal transplantation in renal failure from lupus, usually performed after approximately 6 months of inactive disease, results in allograft survival rates comparable to patients transplanted for other reasons.

Antiglomerular Basement Membrane Disease

Patients who develop autoantibodies directed against glomerular basement antigens frequently develop a glomerulonephritis termed *antiglomerular basement membrane (anti-GBM) disease.* When they present with lung hemorrhage and glomerulonephritis, they have a pulmonary-renal syndrome called *Goodpasture's syndrome.* The target epitopes for this autoimmune disease lie in the quaternary structure of α3 NC1 domain of collagen IV. MHC-restricted T cells initiate the autoantibody response because humans are not tolerant to the epitopes created by this quaternary structure. The epitopes are normally sequestered in the collagen IV hexamer and can be exposed by infection, smoking, oxidants, or solvents. Goodpasture's syndrome appears in two age groups: in young men in their late 20s and in men and women in their 60–70s. Disease in the younger age group is usually explosive, with hemoptysis, a sudden fall in hemoglobin, fever, dyspnea, and hematuria. Hemoptysis is largely confined to smokers, and those who present with lung hemorrhage as a group do better than older populations who have prolonged, asymptomatic renal injury; presentation with oliguria is often associated with a particularly bad outcome. The performance of an urgent kidney biopsy is important in suspected cases of Goodpasture's syndrome to confirm the diagnosis and assess prognosis. Renal biopsies typically show *focal or segmental necrosis* that later, with aggressive destruction of the capillaries by cellular proliferation, leads to crescent formation in Bowman's space (Fig. e14-14). As

these lesions progress, there is concomitant interstitial nephritis with fibrosis and tubular atrophy. The presence of anti-GBM antibodies and complement is recognized on biopsy by linear immunofluorescent staining for IgG (rarely IgA). In testing serum for anti-GBM antibodies, it is particularly important that the α3 NC1 domain of collagen IV alone be used as the target. This is because nonnephritic antibodies against the α1 NC1 domain are seen in paraneoplastic syndromes and cannot be discerned from assays that use whole basement membrane fragments as the binding target. Between 10 and 15% of sera from patients with Goodpasture's syndrome also contain ANCA antibodies against myeloperoxidase. This subset of patients has a vasculitis-associated variant, which has a surprisingly good prognosis with treatment. Prognosis at presentation is worse if there are >50% crescents on renal biopsy with advanced fibrosis, if serum creatinine is >5–6 mg/dL, if oliguria is present, or if there is a need for acute dialysis. Although frequently attempted, most of these latter patients will not respond to plasmapheresis and steroids. Patients with advanced renal failure who present with hemoptysis should still be treated for their lung hemorrhage, as it responds to plasmapheresis and can be lifesaving. Treated patients with less severe disease typically respond to 8–10 treatments of plasmapheresis accompanied by oral prednisone and cyclophosphamide in the first 2 weeks. Kidney transplantation is possible, but because there is risk of recurrence, experience suggests that patients should wait for 6 months and until serum antibodies are undetectable.

IgA Nephropathy

Berger first described the glomerulonephritis now termed *IgA nephropathy.* It is classically characterized by episodichematuria associated with the deposition of IgA in the mesangium. IgA nephropathy is one of the most common forms of glomerulonephritis worldwide. There is a male preponderance, a peak incidence in the second and third decades of life, and rare familial clustering. There are geographic differences in the prevalence of IgA nephropathy, with 30% prevalence along the Asian and Pacific Rim and 20% in southern Europe, compared to a much lower prevalence in northern Europe and North America. It was initially hypothesized that variation in detection, in part, accounted for regional differences. With clinical care in nephrology becoming more uniform, this variation in prevalence more likely reflects true differences among racial and ethnic groups.

IgA nephropathy is predominantly a sporadic disease but susceptibility to it has been shown uncommonly to have a genetic component depending on geography and the existence of "founder effects." Familial forms of IgA nephropathy are more common in northern Italy and eastern Kentucky. No single causal gene has been identified. Clinical and laboratory evidence suggests close similarities between Henoch-Schönlein purpura and IgA nephropathy. Henoch-Schönlein purpura is distinguished clinically from IgA nephropathy by prominent systemic symptoms, a younger age (<20 years old), preceding infection, and abdominal complaints. Deposits of IgA are also found in the glomerular mesangium in a variety of systemic diseases, including chronic liver disease, Crohn's disease, gastrointestinal adenocarcinoma, chronic bronchiectasis, idiopathic interstitial pneumonia, dermatitis herpetiformis, mycosis fungoides, leprosy, ankylosing spondylitis, relapsing polychondritis, and Sjögren's syndrome. IgA deposition in these entities is not usually associated with clinically significant glomerular inflammation or renal dysfunction and thus is not called IgA nephropathy.

IgA nephropathy is an immune complex–mediated glomerulonephritis defined by the presence of diffuse mesangial IgA deposits often associated with mesangial hypercellularity. (See Glomerular Schematic 2.) IgM, IgG, C_3, or immunoglobulin light chains may be codistributed with IgA. IgA deposited in the mesangium is typically polymeric and of the IgA1 subclass, the pathogenic significance of which is not clear. Abnormalities have been described in IgA production by plasma cells, particularly secretory IgA; in IgA clearance, predominately by the liver; in mesangial IgA clearance and receptors for IgA; and in growth factor and cytokine-mediated events. Currently, however, abnormalities in the *O*-glycosylation of the hinge region of IgA seem to best account for the pathogenesis of sporadic IgA nephropathy. Despite the presence of elevated serum IgA levels in 20–50% of patients, IgA deposition in skin biopsies in 15–55% of patients, or elevated levels of secretory IgA and IgA-fibronectin complexes, a renal biopsy is necessary to confirm the diagnosis. Although the immunofluorescent pattern of IgA on renal biopsy defines IgA nephropathy in the proper clinical context, a variety of histologic lesions may be seen on light microscopy (Fig. e14-8), including DPGN; *segmental sclerosis;* and, rarely, *segmental necrosis with cellular crescent formation,* which typically presents as RPGN.

The two most common presentations of IgA nephropathy are recurrent episodes of macroscopic hematuria during or immediately following an upper respiratory infection often accompanied by proteinuria or persistent asymptomatic microscopic hematuria. Nephrotic syndrome, however, is uncommon. Proteinuria can also first appear late in the course of the disease. Rarely patients present with acute renal failure and a rapidly progressive clinical picture. IgA nephropathy is a benign disease for the majority of patients, and 5–30% of patients may go into a complete remission, with others having hematuria but well preserved renal function. In the minority of patients who have progressive disease, progression is slow, with renal failure seen in only 25–30% of patients with IgA nephropathy over 20–25 years. This risk varies considerably among populations. Cumulatively, risk factors for the loss of renal function identified thus far account for less than 50% of the variation in observed outcome

but include the presence of hypertension or proteinuria, the absence of episodes of macroscopic hematuria, male age, older age of onset, and extensive glomerulosclerosis or interstitial fibrosis on renal biopsy. Several analyses in large populations of patients found persistent proteinuria for 6 months or longer to have the greatest predictive power for adverse renal outcomes.

There is no agreement on optimal treatment. Both large studies that include patients with multiple glomerular diseases and small studies of patients with IgA nephropathy support the use of angiotensin-converting enzyme (ACE) inhibitors in patients with proteinuria or declining renal function. Tonsillectomy, steroid therapy, and fish oil have all been suggested in small studies to benefit select patients with IgA nephropathy. When presenting as RPGN, patients typically receive steroids, cytotoxic agents, and plasmapheresis.

ANCA Small-Vessel Vasculitis

A group of patients with small-vessel vasculitis (arterioles, capillaries, and venules; rarely small arteries) and glomerulonephritis have serum ANCA; the antibodies are of two types, anti-proteinase 3 (PR3) or anti-myeloperoxidase (MPO) (Chap. 326); Lamp-2 antibodies have also been reported experimentally as potentially pathogenic. ANCA are produced with the help of T cells and activate leukocytes and monocytes, which together damage the walls of small vessels. Endothelial injury also attracts more leukocytes and extends the inflammation. Granulomatosis with polyangiitis (Wegener's), microscopic polyangiitis, and Churg-Strauss syndrome belong to this group because they are ANCA-positive and have a *pauci-immune glomerulonephritis* with few immune complexes in small vessels and glomerular capillaries. Patients with any of these three diseases can have any combination of the above serum antibodies, but anti-PR3 antibodies are more common in granulomatosis with polyangiitis (Wegener's) and anti-MPO antibodies are more common in microscopic polyangiitis or Churg-Strauss. While each of these diseases have some unique clinical features, most features do not predict relapse or progression, and as a group they are generally treated in the same way. Since mortality is high without treatment, virtually all patients receive urgent treatment. Induction therapy usually includes some combination of plasmapheresis, methylprednisolone, and cyclophosphamide. The benefit of plasmapheresis in this setting is uncertain. Monthly "pulse" IV cyclophosphamide to induce remission of ANCA-associated vasculitis is as effective as daily oral cyclophosphamide and results in reduced cumulative adverse events but may be associated with increased relapses. Steroids are tapered soon after acute inflammation subsides, and patients are maintained on cyclophosphamide or azathioprine for up to a year to minimize the risk of relapse.

GRANULOMATOSIS WITH POLYANGIITIS (WEGENER'S)

Patients with this disease classically present with fever, purulent rhinorrhea, nasal ulcers, sinus pain, polyarthralgias/arthritis, cough, hemoptysis, shortness of breath, microscopic hematuria, and 0.5–1 g/24 h of proteinuria; occasionally there may be cutaneous purpura and mononeuritis multiplex. Presentation without renal involvement is termed *limited granulomatosis with polyangiitis (Wegener's),* although some of these patients will show signs of renal injury later. Chest x-ray often reveals nodules and persistent infiltrates, sometimes with cavities. Biopsy of involved tissue will show a small-vessel vasculitis and adjacent noncaseating granulomas. Renal biopsies during active disease demonstrate *segmental necrotizing glomerulonephritis* without immune deposits (Fig. e14-13). The cause of granulomatosis with polyangiitis (Wegener's) is unknown. In case-controlled studies there is greater risk associated with exposure to silica dust. The disease is also more common in patients with α_1-antitrypsin deficiency, which is an inhibitor of PR3. Relapse after achieving remission is more common in patients with granulomatosis with polyangiitis (Wegener's) than the other ANCA-associated vasculitis, necessitating diligent follow-up care.

MICROSCOPIC POLYANGIITIS

Clinically, these patients look somewhat similar to those with granulomatosis with polyangiitis (Wegener's), except they rarely have significant lung disease or destructive sinusitis. The distinction is made on biopsy, where the vasculitis in microscopic polyangiitis is without granulomas. Some patients will also have injury limited to the capillaries and venules.

CHURG-STRAUSS SYNDROME

When small-vessel vasculitis is associated with peripheral eosinophilia, cutaneous purpura, mononeuritis, asthma, and allergic rhinitis, a diagnosis of Churg-Strauss syndrome is considered. Hypergammaglobulinemia, elevated levels of serum IgE, or the presence of rheumatoid factor sometimes accompanies the allergic state. Lung inflammation, including fleeting cough and pulmonary infiltrates, often precedes the systemic manifestations of disease by years; lung manifestations are rarely absent. A third of patients may have exudative pleural effusions associated with eosinophils. Small-vessel vasculitis and *focal segmental necrotizing glomerulonephritis* can be seen on renal biopsy, usually absent eosinophils or granulomas. The cause of Churg-Strauss syndrome is autoimmune, but the inciting factors are unknown. Interestingly, some asthma patients treated with leukotriene receptor antagonists will develop this vasculitis.

Membranoproliferative Glomerulonephritis

MPGN is sometimes called *mesangiocapillary glomerulonephritis* or *lobar glomerulonephritis.* It is an immune-mediated glomerulonephritis characterized by thickening of the GBM with mesangioproliferative changes; 70% of patients have hypocomplementemia. MPGN is rare in African Americans, and idiopathic disease usually presents in childhood or young adulthood. MPGN is subdivided pathologically into type I, type II, and type III disease. *Type I MPGN* is commonly associated with persistent hepatitis C infections, autoimmune diseases like lupus or cryoglobulinemia, or neoplastic diseases (Table 283-4). *Types II and III MPGN* are usually idiopathic, except in patients with complement factor H deficiency, in the presence of C_3 nephritic factor and/or in partial lipodystrophy producing type II disease, or complement receptor deficiency in type III disease.

Table 283-4 Membranoproliferative Glomerulonephritis

Type I Disease (Most Common)
Idiopathic Subacute bacterial endocarditis
Systemic lupus erythematosus
Hepatitis C ± cryoglobulinemia
Mixed cryoglobulinemia
Hepatitis B
Cancer: Lung, breast, and ovary (germinal)
Type II Disease (Dense Deposit Disease)
Idiopathic
C_3 nephritic factor–associated
Partial lipodystrophy
Type III Disease
Idiopathic
Complement receptor deficiency

Type I MPGN, the most proliferative of the three types, shows mesangial proliferation with lobular segmentation on renal biopsy and mesangial interposition between the capillary basement membrane and endothelial cells, producing a double contour sometimes called *tram-tracking* (Fig. e14-9). (See Glomerular Schematic 3.) Subendothelial deposits with low serum levels of C_3 are typical, although 50% of patients have normal levels of C_3 and occasional intramesangial deposits. Low serum C_3 and a dense thickening of the GBM containing ribbons of dense deposits and C_3 characterize type II MPGN, sometimes called *dense deposit disease* (Fig. e14-10). Classically, the glomerular tuft has a lobular appearance; intramesangial deposits are rarely present and subendothelial deposits are generally absent. Proliferation in type III MPGN is less common than the other two types and is often focal; mesangial interposition is rare, and subepithelial deposits can occur along widened segments of the GBM that appear laminated and disrupted.

Type I MPGN is secondary to glomerular deposition of circulating immune complexes or their in situ formation. Types II and III MPGN may be related to "nephritic factors," which are autoantibodies that stabilize C_3 convertase and allow it to activate serum C_3. Patients with MPGN present with proteinuria, hematuria, and pyuria (30%), systemic symptoms of fatigue and malaise that are most common in children with type I disease, or an acute nephritic picture with RPGN and a speedy deterioration in renal function in up to 25% of patients. Low serum C_3 levels are common. Fifty percent of patients with MPGN develop end-stage disease 10 years after diagnosis, and 90% have renal insufficiency after 20 years. Nephrotic syndrome, hypertension, and renal insufficiency all predict poor outcome. In the presence of proteinuria, treatment with inhibitors of the renin-angiotensin system is prudent. Evidence for treatment with dipyridamole, Coumadin (warfarin), or cyclophosphamide is not strongly established. There is some evidence supporting the efficacy of treatment of *primary MPGN* with steroids, particularly in children, as well as reports of efficacy with plasma exchange and other immunosuppressive drugs. In *secondary MPGN,* treating the associated infection, autoimmune disease, or neoplasms is of demonstrated benefit. In particular, pegylated interferon and ribavirin are useful in reducing viral load. Although all primary renal diseases can recur over time in transplanted renal allografts, patients with MPGN are well known to be at risk for not only a histologic recurrence but also a clinically significant recurrence with loss of graft function.

Mesangioproliferative Glomerulonephritis

Mesangioproliferative glomerulonephritis is characterized by expansion of the mesangium, sometimes associated with mesangial hypercellularity; thin, single contoured capillary walls; and mesangial

immune deposits. Clinically, it can present with varying degrees of proteinuria and, commonly, hematuria. Mesangioproliferative disease may be seen in IgA nephropathy, *Plasmodium falciparum* malaria, resolving postinfectious glomerulonephritis, and class II nephritis from lupus, all of which can have a similar histologic appearance. With these secondary entities excluded, the diagnosis of *primary mesangioproliferative glomerulonephritis* is made in less than 15% of renal biopsies. As an immune-mediated renal lesion with deposits of IgM, C1q, and C_3, the clinical course is variable. Patients with isolated hematuria may have a very benign course, and those with heavy proteinuria occasionally progress to renal failure. There is little agreement on treatment, but some clinical reports suggest benefit from use of inhibitors of the renin-angiotensin system, steroid therapy, and even cytotoxic agents.

NEPHROTIC SYNDROME

Nephrotic syndrome classically presents with heavy proteinuria, minimal hematuria, hypoalbuminemia, hypercholesterolemia, edema, and hypertension. If left undiagnosed or untreated, some of these syndromes will progressively damage enough glomeruli to cause a fall in GFR, producing renal failure.

Therapies for various causes of nephrotic syndrome are noted under individual disease headings below. In general, all patients with hypercholesterolemia secondary to nephrotic syndrome should be treated with lipid-lowering agents because they are at increased risk for cardiovascular disease. Edema secondary to salt and water retention can be controlled with the judicious use of diuretics, avoiding intravascular volume depletion. Venous complications secondary to the hypercoagulable state associated with nephrotic syndrome can be treated with anticoagulants. The losses of various serum binding proteins, such as thyroid-binding globulin, lead to alterations in functional tests. Lastly, proteinuria itself is hypothesized to be nephrotoxic, and treatment of proteinuria with inhibitors of the renin-angiotensin system can lower urinary protein excretion.

Minimal Change Disease

Minimal change disease (MCD), sometimes known as *nil lesion,* causes 70–90% of nephrotic syndrome in childhood but only 10–15% of nephrotic syndrome in adults. Minimal change disease usually presents as a primary renal disease but can be associated with several other conditions, including Hodgkin's disease, allergies, or use of nonsteroidal anti-inflammatory agents; significant interstitial nephritis often accompanies cases associated with nonsteroidal use. Minimal change disease on renal biopsy shows no obvious glomerular lesion by light microscopy and is negative for deposits by immunofluorescent microscopy, or occasionally shows small amounts of IgM in the mesangium (Fig. e14-1). (See Glomerular Schematic 4.) Electron microscopy, however, consistently demonstrates an effacement of the foot process supporting the epithelial podocytes with weakening of slit-pore membranes. The pathophysiology of this lesion is uncertain. Most agree there is a circulating cytokine, perhaps related to a T cell response that alters capillary charge and podocyte integrity. The evidence for cytokine-related immune injury is circumstantial and is suggested by the presence of preceding allergies, altered cell-mediated immunity during viral infections, and the high frequency of remissions with steroids.

Minimal change disease presents clinically with the abrupt onset of edema and nephrotic syndrome accompanied by acellular urinary sediment. Average urine protein excretion reported in 24 hours is 10 grams with severe hypoalbuminemia. Less common clinical features include hypertension (30% in children, 50% in adults), microscopic hematuria (20% in children, 33% in adults), atopy or allergic symptoms (40% in children, 30% in adults), and decreased renal function (<5% in children, 30% in adults). The appearance of acute renal failure in adults is often seen more commonly in patients with low serum albumin and intrarenal edema (nephrosarca) that is responsive to intravenous albumin and diuretics. This presentation must be distinguished from acute renal failure secondary to hypovolemia. Acute tubular necrosis and interstitial inflammation is also reported. In children, the abnormal urine principally contains albumin with minimal amounts of higher-molecular-weight proteins, and is sometimes called *selective proteinuria.* Although up to 30% of children have a spontaneous remission, all children today are treated with steroids; only children who are nonresponders are biopsied in this setting. Primary responders are patients who have a complete remission (<0.2 mg/24 h of proteinuria) after a single course of prednisone; steroid-dependent patients relapse as their steroid dose is tapered. Frequent relapsers have two or more relapses in the 6 months following taper, and steroid-resistant patients fail to respond to steroid therapy. Adults are not considered steroid-resistant until after 4 months of therapy. Ninety to 95% of children will develop a complete remission after 8 weeks of steroid therapy, and 80–85% of adults will achieve complete remission, but only after a longer course of 20–24 weeks. Patients with steroid resistance may have FSGS on repeat biopsy. Some hypothesize that if the first renal biopsy does not have a sample of deeper corticomedullary glomeruli, then the correct early diagnosis of FSGS may be missed.

Relapses occur in 70–75% of children after the first remission, and early relapse predicts multiple subsequent relapses. The frequency of relapses decreases after puberty, although there is an increased risk of relapse following the rapid tapering of steroids in all groups. Relapses are less common in adults but are more resistant to subsequent therapy. Prednisone is first-line therapy, either

given daily or on alternate days. Other immunosuppressive drugs, such as cyclophosphamide, chlorambucil, and mycophenolate mofetil, are saved for frequent relapsers, steroid-dependent, or steroid-resistant patients. Cyclosporine can induce remission, but relapse is also common when cyclosporine is withdrawn. The long-term prognosis in adults is less favorable when acute renal failure or steroid resistance occurs.

Focal Segmental Glomerulosclerosis

Focal segmental glomerulosclerosis (FSGS) refers to a pattern of renal injury characterized by segmental glomerular scars that involve some but not all glomeruli; the clinical findings of FSGS largely manifest as proteinuria. When the secondary causes of FSGS are eliminated (Table 283-5), the remaining patients are considered to have primary FSGS. The incidence of this disease is increasing, and it now represents up to one-third of cases of nephrotic syndrome in adults and one-half of cases of nephrotic syndrome in African Americans, in whom it is seen more commonly. The pathogenesis of FSGS is probably multifactorial. Possible mechanisms include a T cell–mediated circulating permeability factor, TGF-β–mediated cellular proliferation and matrix synthesis, and podocyte abnormalities associated with genetic mutations. Risk polymorphisms at the *APOL1* locus encoding apolipoprotein L1 expressed in podocytes substantially explain the increased burden of FSGS among African Americans with or without HIV-associated disease.

Table 283-5 Focal Segmental Glomerulosclerosis

Primary focal segmental glomerulosclerosis
Secondary focal segmental glomerulosclerosis
Viruses: HIV/Hepatitis B/Parvovirus
Hypertensive nephropathy
Reflux nephropathy
Cholesterol emboli
Drugs: Heroin/analgesics/pamidronate
Oligomeganephronia
Renal dysgenesis
Alport's syndrome
Sickle cell disease
Lymphoma
Radiation nephritis
Familial podocytopathies
NPHS1 mutation/nephrin
NPHS2 mutation/podocin
TRPC6 mutation/cation channel
ACTN4 mutation/actinin
α-Galactosidase A deficiency/Fabry's disease
N-acetylneuraminic acid hydrolase deficiency/nephrosialidosis

The pathologic changes of FSGS are most prominent in glomeruli located at the corticomedullary junction (Fig. e14-2), so if the renal biopsy specimen is from superficial tissue, the lesions can be missed, which sometimes leads to a misdiagnosis of MCD. In addition to focal and segmental scarring, other variants have been described, including cellular lesions with *endocapillary hypercellularity* and heavy proteinuria; *collapsing glomerulopathy* (Fig. e14-3) with segmental or global glomerular collapse and a rapid decline in renal function; a hilar stalk lesion **(**Fig. e14-4) or the *glomerular tip lesion***(**Fig. e14-5), which may have a better prognosis. (See Glomerular Schematic 5.)

FSGS can present with hematuria, hypertension, any level of proteinuria or renal insufficiency. Nephrotic-range proteinuria, African-American race, and renal insufficiency are associated with a poor outcome, with 50% of patients reaching renal failure in 6–8 years. FSGS rarely remits spontaneously, but treatment-induced remission of proteinuria significantly improves prognosis. Treatment of patients with *primary FSGS* should include inhibitors of the renin-angiotensin system. Based on retrospective studies, patients with nephrotic-range proteinuria can be treated with steroids but respond far less often and after a longer course of therapy than patients with MCD. Proteinuria remits in only 20–45% of patients receiving a course of steroids over 6–9 months. Limited evidence suggests the use of cyclosporine in steroid-responsive patients helps ensure remissions. Relapse frequently occurs after

cessation of cyclosporine therapy, and cyclosporine itself can lead to a deterioration of renal function due to its nephrotoxic effects. A role for other agents that suppress the immune system has not been established. Primary FSGS recurs in 25–40% of patients given allografts at end-stage disease, leading to graft loss in half of those cases. The treatment of *secondary FSGS* typically involves treating the underlying cause and controlling proteinuria. There is no role for steroids or other immunosuppressive agents in secondary FSGS.

Membranous Glomerulonephritis

Membranous glomerulonephritis (MGN), or *membranous nephropathy* as it is sometimes called, accounts for approximately 30% of cases of nephrotic syndrome in adults, with a peak incidence between the ages of 30 and 50 years and a male to female ratio of 2:1. It is rare in childhood and the most common cause of nephrotic syndrome in the elderly. In 25–30% of cases, MGN is associated with a malignancy (solid tumors of the breast, lung, colon), infection (hepatitis B, malaria, schistosomiasis), or rheumatologic disorders like lupus or rarely rheumatoid arthritis (Table 283-6).

Table 283-6 Membranous Glomerulonephritis

Primary/idiopathic membranous glomerulonephritis
Secondary membranous glomerulonephritis
Infection: Hepatitis B and C, syphilis, malaria, schistosomiasis, leprosy, filariasis
Cancer: Breast, colon, lung, stomach, kidney, esophagus, neuroblastoma
Drugs: gold, mercury, penicillamine, nonsteroidal anti-inflammatory agents, probenecid
Autoimmune diseases: systemic lupus erythematosus, rheumatoid arthritis, primary biliary cirrhosis, dermatitis herpetiformis, bullous pemphigoid, myasthenia gravis, Sjögren's syndrome, Hashimoto's thyroiditis
Other systemic diseases: Fanconi's syndrome, sickle cell anemia, diabetes, Crohn's disease, sarcoidosis, Guillain-Barré syndrome, Weber-Christian disease, angiofollicular lymph node hyperplasia

Uniform thickening of the basement membrane along the peripheral capillary loops is seen by light microscopy on renal biopsy (Fig. e14-7); this thickening needs to be distinguished from that seen in diabetes and amyloidosis. (See Glomerular Schematic 6.) Immunofluorescence demonstrates diffuse granular deposits of IgG and C_3, and electron microscopy typically reveals electron-dense subepithelial deposits. While different stages (I–V) of progressive membranous lesions have been described, some published analyses indicate the degree of tubular atrophy or interstitial fibrosis is more predictive of progression than is the stage of glomerular disease. The presence of subendothelial deposits or the presence of tubuloreticular inclusions strongly points to a diagnosis of membranous lupus nephritis, which may precede the extrarenal manifestations of lupus. Work in Heyman nephritis, an animal model of MGN, suggests that glomerular lesions result from in situ formation of immune complexes with megalin receptor–associated protein as the putative antigen. This antigen is not found in human podocytes, but human antibodies have been described against neutral endopeptidase expressed by podocytes, hepatitis antigens B/C, *Helicobacter pylori* antigens, and tumor antigens. In a newer study, autoantibodies against the M-type phospholipase A_2 receptor (PLA_2R) circulate and bind to a conformational epitope present in the receptor on human podocytes, producing in situ deposits characteristic of idiopathic membranous nephropathy. Other renal diseases and secondary membranous nephropathy do not appear to involve such autoantibodies. Eighty percent of patients with MGN present with nephrotic syndrome and nonselective proteinuria. Microscopic hematuria is seen in up to 50% of patients but is seen less commonly than in IgA nephropathy or FSGS. Spontaneous remissions occur in 20–33% of patients and often occur late in the course after years of nephrotic syndrome, which make treatment decisions difficult. One-third of patients continue to have relapsing nephrotic syndrome but maintain normal renal function, and approximately another third of patients develop renal failure or die from the complications of nephrotic syndrome. Male gender, older age, hypertension, and the persistence of proteinuria are associated with worse prognosis. Although thrombotic complications are a feature of all nephrotic syndromes, MGN has the highest reported incidences of renal vein thrombosis, pulmonary embolism, and deep vein thrombosis. Prophylactic anticoagulation is controversial but has been recommended for patients with severe or prolonged proteinuria in the absence of risk factors for bleeding.

In addition to the treatment of edema, dyslipidemia, and hypertension, inhibition of the renin-angiotensin system is recommended. Therapy with immunosuppressive drugs is also recommended for patients with primary MGN and persistent proteinuria (>3.0 g/24 h). The choice of immunosuppressive drugs for therapy is controversial, but current recommendations based on small clinical studies are to treat with steroids and cyclophosphamide, chlorambucil, mycophenolate mofetil, or cyclosporine. In patients who relapse or fail to respond to this therapy there are case reports of beneficial effects with

the use of rituximab, an anti-CD20 antibody directed at B cells, or with synthetic adrenocorticotropic hormone.

Diabetic Nephropathy

Diabetic nephropathy is the single most common cause of chronic renal failure in the United States, accounting for 45% of patients receiving renal replacement therapy, and is a rapidly growing problem worldwide. The dramatic increase in the number of patients with diabetic nephropathy reflects the epidemic increase in obesity, metabolic syndrome, and type 2 diabetes mellitus. Approximately 40% of patients with types 1 or 2 diabetes develop nephropathy, but due to the higher prevalence of type 2 diabetes (90%) compared to type 1 (10%), the majority of patients with diabetic nephropathy have type 2 disease. Renal lesions are more common in African-American, Native American, Polynesian, and Maori populations. Risk factors for the development of diabetic nephropathy include hyperglycemia, hypertension, dyslipidemia, smoking, a family history of diabetic nephropathy, and genepolymorphisms affecting the activity of the renin-angiotensin-aldosterone axis.

Within 1–2 years after the onset of clinical diabetes, morphologic changes appear in the kidney. Thickening of the GBM is a sensitive indicator for the presence of diabetes but correlates poorly with the presence or absence of clinically significant nephropathy. The composition of the GBM is altered notably with a loss of heparan sulfate moieties that form the negatively charged filtration barrier. This change results in increased filtration of serum proteins into the urine, predominately negatively charged albumin. The expansion of the mesangium due to the accumulation of extracellular matrix correlates with the clinical manifestations of diabetic nephropathy (see stages in Fig. e14-20). This expansion in mesangial matrix is associated with the development of *mesangial sclerosis.* Some patients also develop eosinophilic, PAS+ nodules called *nodular glomerulosclerosis* or *Kimmelstiel-Wilson nodules.* Immunofluorescence microscopy often reveals the nonspecific deposition of IgG (at times in a linear pattern) or complement staining without immune deposits on electron microscopy. Prominent vascular changes are frequently seen with hyaline and hypertensive arteriosclerosis. This is associated with varying degrees of chronic glomerulosclerosis and tubulointerstitial changes. Renal biopsies from patients with types 1 and 2 diabetes are largely indistinguishable.

These pathologic changes are the result of a number of postulated factors. Multiple lines of evidence support an important role for increases in glomerular capillary pressure (intraglomerular hypertension) in alterations in renal structure and function. Direct effects of hyperglycemia on the actin cytoskeleton of renal mesangial and vascular smooth-muscle cells as well as diabetes-associated changes in circulating factors such as atrial natriuretic factor, angiotensin II, and insulin-like growth factor (IGF) may account for this. Sustained glomerular hypertension increases matrix production, alterations in the GBM with disruption in the filtration barrier (and hence proteinuria), and glomerulosclerosis. A number of factors have also been identified that alter matrix production, including the accumulation of advanced glycosylation end products, circulating factors including growth hormone, IGF-I, angiotensin II, connective tissue growth factor, TGF-ß, and dyslipidemia.

The natural history of diabetic nephropathy in patients with types 1 and 2 diabetes is similar. However, since the onset of Type 1 diabetes is readily identifiable and the onset of type 2 diabetes is not, a patient newly diagnosed with type 2 diabetes may have renal disease for many years before nephropathy is discovered and presents as *advanceddiabetic nephropathy.* At the onset of diabetes, renal hypertrophy and glomerular hyperfiltration are present. The degree of glomerular hyperfiltration correlates with the subsequent risk of clinically significant nephropathy. In the approximately 40% of patients with diabetes who develop diabetic nephropathy, the earliest manifestation is an increase in albuminuria detected by sensitive radioimmunoassay (Table 283-1). Albuminuria in the range of 30 –300 mg/24 h is called *microalbuminuria.* In patients with types 1 or 2 diabetes, microalbuminuria appears 5–10 years after the onset of diabetes. It is currently recommended to test patients with type 1 disease for microalbuminuria 5 years after diagnosis of diabetes and yearly thereafter, and, because the time of onset of type 2 diabetes is often unknown, to test type 2 patients at the time of diagnosis of diabetes and yearly thereafter.

Patients with small rises in albuminuria increase their levels of urinary albumin excretion, typically reaching dipstick positive levels of proteinuria (>300 mg albuminuria) 5–10 years after the onset of early albuminuria. Microalbuminuria is a potent risk factor for cardiovascular events and death in patients with type 2 diabetes. Many patients with type 2 diabetes and microalbuminuria succumb to cardiovascular events before they progress to proteinuria or renal failure. Proteinuria in frank diabetic nephropathy can be variable, ranging from 500 mg to 25 g/24 h, and is often associated with nephrotic syndrome. More than 90% of patients with type 1 diabetes and nephropathy have diabetic retinopathy, so the absence of retinopathy in type 1 patients with proteinuria should prompt consideration of a diagnosis other than diabetic nephropathy; only 60% of patients with type 2 diabetes with nephropathy have diabetic retinopathy. There is a highly significant correlation between the presence of retinopathy and the presence of Kimmelstiel-Wilson nodules (Fig. e14-20). Also, characteristically, patients with advanced diabetic nephropathy have normal to enlarged kidneys, in contrast to other glomerular diseases where kidney size is usually decreased. Using the above epidemiologic and clinical data, and in the absence of other clinical or serologic data suggesting another disease, diabetic nephropathy is

usually diagnosed without a renal biopsy. After the onset of proteinuria, renal function inexorably declines, with 50% of patients reaching renal failure over another 5–10 years; thus, from the earliest stages of microalbuminuria, it usually takes 10–20 years to reach end-stage renal disease. Hypertension may predict which patients develop diabetic nephropathy, as the presence of hypertension accelerates the rate of decline in renal function. Once renal failure appears, however, survival on dialysis is far shorter for patients with diabetes compared to other dialysis patients. Survival is best for patients with type 1 diabetes who receive a transplant from a living related donor.

Good evidence supports the benefits of blood sugar and blood pressure control as well as inhibition of the renin-angiotensin system in retarding the progression of diabetic nephropathy. In patients with type 1 diabetes, intensive control of blood sugar clearly prevents the development or progression of diabetic nephropathy. The evidence for benefit of intensive blood glucose control in patients with type 2 diabetes is less certain, with current studies reporting conflicting results. Some, but not all, trials have reported increased mortality rate associated with intensive blood glucose control and the safety of $HgbA_1C$ goals less than 7% in patients with type 2 diabetes is currently unclear.

Controlling systemic blood pressure decreases renal and cardiovascular adverse events in this high-risk population. The vast majority of patients with diabetic nephropathy require three or more antihypertensive drugs to achieve this goal. Drugs that inhibit the renin-angiotensin system, independent of their effects on systemic blood pressure, have been shown in numerous large clinical trials to slow the progression of diabetic nephropathy at early (microalbuminuria) and late (proteinuria with reduced glomerular filtration) stages, independent of any effect they may have on systemic blood pressure. Since angiotensin II increases efferent arteriolar resistance and, hence, glomerular capillary pressure, one key mechanism for the efficacy of ACE inhibitors or angiotensin receptor blockers (ARBs) is reducing glomerular hypertension. Patients with type 1 diabetes for 5 years, who develop albuminuria or declining renal function should be treated with ACE inhibitors. Patients with type 2 diabetes and microalbuminuria or proteinuria may be treated with ACE inhibitors or ARBs. Less compelling evidence supports therapy with a combination of two drugs (ACE inhibitors, ARBs, renin inhibitors, or aldosterone antagonists) that suppress several components of the renin-angiotensin system.

Glomerular Deposition Diseases

Plasma cell dyscrasias producing excess light chain immunoglobulin sometimes lead to the formation of glomerular and tubular deposits that cause heavy proteinuria and renal failure; the same is true for the accumulation of serum amyloid A protein fragments seen in several inflammatory diseases. This broad group of proteinuric patients has *glomerular deposition disease.*

LIGHT CHAIN DEPOSITION DISEASE

The biochemical characteristics of nephrotoxic light chains produced in patients with light chain malignancies often confer a specific pattern of renal injury; that of either cast nephropathy (Fig. e14-17), which causes renal failure but not heavy proteinuria or amyloidosis, or light chain deposition disease (Fig. e14-16), which produces nephrotic syndrome with renal failure. These latter patients produce kappa light chains that do not have the biochemical features necessary to form amyloid fibrils. Instead, they self-aggregate and form granular deposits along the glomerular capillary and mesangium, tubular basement membrane, and Bowman's capsule. When predominant in glomeruli, nephrotic syndrome develops, and about 70% of patients progress to dialysis. Light-chain deposits are not fibrillar and do not stain with Congo red, but they are easily detected with anti–light chain antibody using immunofluorescence or as granular deposits on electron microscopy. A combination of the light chain rearrangement, self-aggregating properties at neutral pH, and abnormal metabolism probably contribute to the deposition. Treatment for light chain deposition disease is treatment of the primary disease. As so many patients with light chain deposition disease progress to renal failure, the overall prognosis is grim.

RENAL AMYLOIDOSIS

Most *renal amyloidosis* is either the result of primary fibrillar deposits of immunoglobulin light chains known as amyloid L (AL), or secondary to fibrillar deposits of serum amyloid A (AA) protein fragments (Chap. 112). Even though both occur for different reasons, their clinicopathophysiology is quite similar and will be discussed together. Amyloid infiltrates the liver, heart, peripheral nerves, carpal tunnel, upper pharynx, and kidney, producing restrictive cardiomyopathy, hepatomegaly, macroglossia, and heavy proteinuria sometimes associated with renal vein thrombosis. In systemic AL amyloidosis, also called *primary amyloidosis,* light chains produced in excess by clonal plasma cell dyscrasias are made into fragments by macrophages so they can self-aggregate at acid pH. A disproportionate number of these light chains (75%) are of the *lambda* class. About 10% of these patients have overt myeloma with lytic bone lesions and infiltration of the bone marrow with >30% plasma cells; nephrotic syndrome is common, and about 20% of patients progress to dialysis. AA amyloidosis is sometimes called *secondary amyloidosis* and also presents as nephrotic syndrome. It is due to deposition of β-pleated sheets of serum amyloid A protein, an acute phase reactant whose physiologic functions include cholesterol transport, immune cell attraction, and metalloproteases activation. Forty percent of patients

with AA amyloid have rheumatoid arthritis, and another 10% have ankylosing spondylitis or psoriatic arthritis; the rest derive from other lesser causes. Less common in Western countries but more common in Mediterranean regions, particularly in Sephardic and Iraqi Jews, is familial Mediterranean fever (FMF). FMF is caused by a mutation in the gene encoding pyrin, while Muckle-Wells syndrome, a related disorder, results from a mutation in cryopyrin; both proteins are important in the apoptosis of leukocytes early in inflammation; such proteins with pyrin domains are part of a new pathway called the *inflammasome*. Receptor mutations in tumor necrosis factor receptor 1 (TNFR1)-associated periodic syndrome also produce chronic inflammation and secondary amyloidosis. Fragments of serum amyloid A protein increase and self-aggregate by attaching to receptors for advanced glycation end products in the extracellular environment; nephrotic syndrome is common, and about 40–60% of patients progress to dialysis. AA and AL amyloid fibrils are detectable with Congo red or in more detail with electron microscopy (Fig. e14-15). Currently developed serum free light chain nephelometry assays are useful in the early diagnosis and follow-up of disease progression. Biopsy of involved liver or kidney is diagnostic 90% of the time when the pretest probability is high; abdominal fat pad aspirates are positive about 70% of the time, but apparently less so when looking for AA amyloid. Amyloid deposits are distributed along blood vessels and in the mesangial regions of the kidney. The treatment for primary amyloidosis is not particularly effective; melphalan and autologous hematopoietic stem cell transplantation can delay the course of disease in about 30% of patients. Secondary amyloidosis is also relentless unless the primary disease can be controlled. Some new drugs in development that disrupt the formation of fibrils may be available in the future.

FIBRILLARY-IMMUNOTACTOID GLOMERULOPATHY

Fibrillary-immunotactoid glomerulopathy is a rare (<1.0% of renal biopsies) morphologically defined disease characterized by glomerular accumulation of nonbranching randomly arranged fibrils. Some classify amyloid and nonamyloid fibril-associated renal disease all as fibrillary glomerulopathies with immunotactoid glomerulopathy reserved for nonamyloid fibrillary disease not associated with a systemic illness. Others define fibrillary glomerulonephritis as a nonamyloid fibrillary disease with fibrils 12–24 nm and immunotactoid glomerulonephritis with fibrils >30 nm. In either case, fibrillar/microtubular deposits of oligoclonal or oligotypic immunoglobulins and complement appear in the mesangium and along the glomerular capillary wall. Congo red stains are negative. The cause of this "nonamyloid" glomerulopathy is mostly idiopathic; reports of immunotactoid glomerulonephritis describe an occasional association with chronic lymphocytic leukemia or B cell lymphoma. Both disorders appear in adults in the fourth decade with moderate to heavy proteinuria, hematuria, and a wide variety of histologic lesions, including DPGN, MPGN, MGN, or mesangioproliferative glomerulonephritis. Nearly half of patients develop renal failure over a few years. There is no consensus on treatment of this uncommon disorder. The disease has been reported to recur following renal transplantation in a minority of cases.

Fabry's Disease

Fabry's disease is an X-linked inborn error of globotriaosylceramide metabolism secondary to deficient lysosomal α-galactosidase A activity, resulting in excessive intracellular storage of globotriaosylceramide. Affected organs include the vascular endothelium, heart, brain, and kidneys. Classically, Fabry's disease presents in childhood in males with acroparesthesias, angiokeratoma, and hypohidrosis. Over time male patients develop cardiomyopathy, cerebrovascular disease, and renal injury, with an average age of death around 50 years of age. Hemizygotes with hypomorphic mutations sometimes present in the fourth to sixth decade with single-organ involvement. Rarely, dominant-negative α-galactosidase A mutations or female heterozygotes with unfavorable X inactivation present with mild single-organ involvement. Rare females develop severe manifestations including renal failure but do so later in life than males. Renal biopsy reveals enlarged glomerular visceral epithelial cells packed with small clear vacuoles containing globotriaosylceramide; vacuoles may also be found in parietal and tubular epithelia (Fig. e14-18). These vacuoles of electron-dense materials in parallel arrays (zebra bodies) are easily seen on electron microscopy. Ultimately, renal biopsies reveal FSGS. The nephropathy of Fabry's disease typically presents in the third decade as mild to moderate proteinuria, sometimes with microscopic hematuria or nephrotic syndrome. Urinalysis may reveal oval fat bodies and birefringent glycolipid globules under polarized light (Maltese cross). Renal biopsy is necessary for definitive diagnosis. Progression to renal failure occurs by the fourth or fifth decade. Treatment with inhibitors of the renin-angiotensin system is recommended. Treatment with recombinant α-galactosidase A clears microvascular endothelial deposits of globotriaosylceramide from the kidneys, heart, and skin. The degree of organ involvement at the time when enzyme replacement is initiated is crucial. In patients with advanced organ involvement, progression of disease occurs despite enzyme replacement therapy. Variable responses to enzyme therapy may be due to the occurrence of neutralizing antibodies or differences in uptake of the enzyme. Graft and patient survival following renal transplantation in patients with Fabry's are similar to other causes of end-stage renal disease.

PULMONARY-RENAL SYNDROMES

Several diseases can present with catastrophic hemoptysis and glomerulonephritis associated with

varying degrees of renal failure. The usual causes include Goodpasture's syndrome, granulomatosis with polyangiitis (Wegener's), microscopic polyangiitis, Churg-Strauss vasculitis, and, rarely, Henoch-Schönlein purpura or cryoglobulinemia. Each of these diseases can also present without hemoptysis and are discussed in detail in "Acute Nephritic Syndromes," above. (See Glomerular Schematic 7.) Pulmonary bleeding in this setting is life-threatening and often results in airway intubation, and acute renal failure requires dialysis. Diagnosis is difficult initially because biopsies and serologic testing take time. Treatment with plasmapheresis and methylprednisolone is often empirical and temporizing until results of testing are available.

BASEMENT MEMBRANE SYNDROMES

All kidney epithelia, including podocytes, rest on basement membranes assembled into a planar surface through the interweaving of collagen IV with laminins, nidogen, and sulfated proteoglycans. Structural abnormalities in GBM associated with hematuria are characteristic of several familial disorders related to the expression of collagen IV genes. The extended family of collagen IV contains six chains, which are expressed in different tissues at different stages of embryonic development. All epithelial basement membranes early in human development are composed of interconnected triple-helical protomers rich in α1.α1.α2(IV) collagen. Some specialized tissues undergo a developmental switch replacing α1.α1.α2(IV) protomers with an α3.α4.α5(IV) collagen network; this switch occurs in the kidney (glomerular and tubular basement membrane), lung, testis, cochlea, and eye, while an α5.α5.α6(IV) network appears in skin, smooth muscle, and esophagus and along Bowman's capsule in the kidney. This switch probably occurs because the α3.α4.α5(IV) network is more resistant to proteases and ensures the structural longevity of critical tissues. When basement membranes are the target of glomerular disease, they produce moderate proteinuria, some hematuria, and progressive renal failure.

Anti-GBM Disease

Autoimmune disease where antibodies are directed against the α3 NC1 domain of collagen IV produces an *anti-GBM disease* often associated with RPGN and/or a pulmonary-renal syndrome called *Goodpasture's syndrome.* Discussion of this disease is covered in "Acute Nephritic Syndromes," above.

Alport's Syndrome

Classically, patients with Alport's syndrome develop hematuria, thinning and splitting of the GBMs, mild proteinuria (<1–2 g/24 h), which appears late in the course, followed by chronic glomerulosclerosis leading to renal failure in association with sensorineural deafness. Some patients develop lenticonus of the anterior lens capsule, "dot and fleck" retinopathy, and rarely, mental retardation or leiomyomatosis. Approximately 85% of patients with Alport's syndrome have an X-linked inheritance of mutations in the α5(IV) collagen chain on chromosome Xq22–24. Female carriers have variable penetrance depending on the type of mutation or the degree of mosaicism created by X inactivation. Fifteen percent of patients have autosomal recessive disease of the α3(IV) or α4(IV) chains on chromosome 2q35–37. Rarely, some kindred have an autosomal dominant inheritance of dominant-negative mutations in α3(IV) or α4(IV) chains.

Pedigrees with the X-linked syndrome are quite variable in their rate and frequency of tissue damage leading to organ failure. Seventy percent of patients have the juvenile form with nonsense or missense mutations, reading frame shifts, or large deletions and generally develop renal failure and sensorineural deafness by age 30. Patients with splice variants, exon skipping, or missense mutations of α-helical glycines generally deteriorate after the age of 30 (adult form) with mild or late deafness. Early severe deafness, lenticonus, or proteinuria suggests a poorer prognosis. Usually females from X-linked pedigrees have only microhematuria, but up to 25% of carrier females have been reported to have more severe renal manifestations. Pedigrees with the autosomal recessive form of the disease have severe early disease in both females and males with asymptomatic parents.

Clinical evaluation should include a careful eye examination and hearing tests. However, the absence of extrarenal symptoms does not rule out the diagnosis. Since α5(IV) collagen is expressed in the skin, some X-linked Alport patients can be diagnosed with a skin biopsy revealing the lack of the α5(IV) collagen chain on immunofluorescent analysis. Other patients with suspected disease require a renal biopsy. Alport's patients early in their disease typically have thin basement membranes on renal biopsy (Fig. e14-19), which thicken over time into multilamellations surrounding lucent areas that often contain granules of varying density—the so-called split basement membrane. In any Alport kidney there are areas of thinning mixed with splitting of the GBM. Tubules drop out, glomeruli scar, and the kidney eventually succumbs to interstitial fibrosis. Primary treatment is control of systemic hypertension and use of ACE inhibitors to slow renal progression. Although patients who receive renal allografts usually develop anti-GBM antibodies directed toward the collagen epitopes absent in their native kidney, overt Goodpasture's syndrome is rare and graft survival is good.

Thin Basement Membrane Disease

Thin basement membrane disease (TBMD) characterized by persistent or recurrent hematuria is not typically associated with proteinuria, hypertension, or loss of renal function or extrarenal disease.

Although not all cases are familial (perhaps a founder effect), it usually presents in childhood in multiple family members and is also called *benign familial hematuria*. Cases of TBMD have genetic defects in type IV collagen but in contrast to Alport behave as an autosomal dominant disorder that in ~40% of families segregates with the *COL(IV)* α*3/COL(IV)* α*4* loci. Mutations in these loci can result in a spectrum of disease ranging from TBMD to autosomal dominant or recessive Alport's. The GBM shows diffuse thinning compared to normal values for the patient's age in otherwise normal biopsies (Fig. e14-19). The vast majority of patients have a benign course.

Nail-Patella Syndrome

Patients with nail-patella syndrome develop iliac horns on the pelvis and dysplasia of the dorsal limbs involving the patella, elbows, and nails, variably associated with neural-sensory hearing impairment, glaucoma, and abnormalities of the GBM and podocytes, leading to hematuria, proteinuria, and FSGS. The syndrome is autosomal dominant, with haploinsufficiency for the LIM homeodomain transcription factor LMX1B; pedigrees are extremely variable in the penetrance for all features of the disease. LMX1B regulates the expression of genes encoding α3 and α4 chains of collagen IV, interstitial type III collagen, podocin, and CD2AP that help form the slit-pore membranes connecting podocytes. Mutations in the LIM domain region of LMX1B associate with glomerulopathy, and renal failure appears in as many as 30% of patients. Proteinuria or isolated hematuria is discovered throughout life, but usually by the third decade, and is inexplicably more common in females. On renal biopsy there is lucent damage to the lamina densa of the GBM, an increase in collagen III fibrils along glomerular capillaries and in the mesangium, and damage to the slit-pore membrane, producing heavy proteinuria not unlike that seen in congenital nephrotic syndrome. Patients with renal failure do well with transplantation.

GLOMERULAR-VASCULAR SYNDROMES

A variety of diseases result in classic vascular injury to the glomerular capillaries. Most of these processes also damage blood vessels elsewhere in the body. The group of diseases discussed here lead to vasculitis, renal endothelial injury, thrombosis, ischemia, and/or lipid-based occlusions.

Atherosclerotic Nephropathy

Aging in the developed world is commonly associated with the occlusion of coronary and systemic blood vessels. The reasons for this include obesity, insulin resistance, smoking, hypertension, and diets rich in lipids that deposit in the arterial and arteriolar circulation, producing local inflammation and fibrosis of small blood vessels. When the renal arterial circulation is involved, the glomerular microcirculation is damaged, leading to *chronic nephrosclerosis.* Patients with GFRs <60 mL/min have more cardiovascular events and hospitalizations than those with higher filtration rates. Several aggressive lipid disorders can accelerate this process, but most of the time atherosclerotic progression to chronic nephrosclerosis is associated with poorly controlled hypertension. Approximately 10% of glomeruli are normally sclerotic by age 40, rising to 20% by age 60 and 30% by age 80. Serum lipid profiles in humans are greatly affected by *apolipoprotein E* polymorphisms; the E4 allele is accompanied by increases in serum cholesterol and is more closely associated with atherogenic profiles in patients with renal failure. Mutations in E2 alleles, particularly in Japanese patients, produce a specific renal abnormality called *lipoprotein glomerulopathy* associated with glomerular lipoprotein thrombi and capillary dilation.

Hypertensive Nephrosclerosis

Uncontrolled systemic hypertension causes permanent damage to the kidneys in about 6% of patients with elevated blood pressure. As many as 27% of patients with end-stage kidney disease have hypertension as a primary cause. Although there is not a clear correlation between the extent or duration of hypertension and the risk of end-organ damage, *hypertensive nephrosclerosis* is fivefold more frequent in African Americans than whites. Risk alleles associated with *APOL1,* a functional gene for apolipoprotein L1 expressed in podocytes substantially explains the increased burden of end-stage renal disease among African Americans. Associated risk factors for progression to end-stage kidney disease include age, sex, race, smoking, hypercholesterolemia, duration of hypertension, low birth weight, and preexisting renal injury. Kidney biopsies in patients with hypertension, microhematuria, and moderate proteinuria demonstrate arteriolosclerosis, chronic nephrosclerosis, and interstitial fibrosis in the absence of immune deposits (Fig. e14-21). Today, based on a careful history, physical examination, urinalysis, and some serologic testing, the diagnosis of chronic nephrosclerosis is usually inferred without a biopsy. Treating hypertension is the best way to avoid progressive renal failure; most guidelines recommend lowering blood pressure to <130/80 mmHg if there is preexisting diabetes or kidney disease. In the presence of kidney disease, most patients begin therapy with two drugs, classically a thiazide diuretic and an ACE inhibitor; most will require three drugs. There is strong evidence in African Americans with hypertensive nephrosclerosis that therapy initiated with an ACE inhibitor can slow the rate of decline in renal function independent of effects on systemic blood pressure. Malignant acceleration of hypertension complicates the course of chronic nephrosclerosis, particularly in the setting of scleroderma or cocaine use (Fig. e14-24). The hemodynamic stress of malignant hypertension leads to fibrinoid necrosis of small blood vessels, thrombotic

microangiography, a nephritic urinalysis, and acute renal failure. In the setting of renal failure, chest pain, or papilledema, the condition is treated as a hypertensive emergency. Slightly lowering the blood pressure often produces an immediate reduction in GFR that improves as the vascular injury attenuates and autoregulation of blood vessel tone is restored.

Cholesterol Emboli

Aging patients with clinical complications from atherosclerosis sometimes shower cholesterol crystals into the circulation—either spontaneously or, more commonly, following an endovascular procedure with manipulation of the aorta—or with use of systemic anticoagulation. Spontaneous emboli may shower acutely or shower subacutely and somewhat more silently. Irregular emboli trapped in the microcirculation produce ischemic damage that induces an inflammatory reaction. Depending on the location of the atherosclerotic plaques releasing these cholesterol fragments, one may see cerebral transient ischemic attacks; livedo reticularis in the lower extremities; Hollenhorst plaques in the retina with visual field cuts; necrosis of the toes; and acute glomerular capillary injury leading to *focal segmental glomerulosclerosis* sometimes associated with hematuria, mild proteinuria, and loss of renal function, which typically progresses over a few years. Occasional patients have fever, eosinophilia, or eosinophiluria. A skin biopsy of an involved area may be diagnostic. Since tissue fixation dissolves the cholesterol, one typically sees only residual, biconvex clefts in involved vessels (Fig. e14-22). There is no therapy to reverse embolic occlusions, and steroids do not help. Controlling blood pressure and lipids and cessation of smoking are usually recommended for prevention.

Sickle Cell Disease

Although individuals with SA-hemoglobin are usually asymptomatic, most will gradually develop hyposthenuria due to subclinical infarction of the renal medulla, thus predisposing them to volume depletion; interestingly, there is an unexpectedly high prevalence of sickle trait among dialysis patients who are African American. Patients with homozygous SS-sickle cell disease develop chronic vasoocclusive disease in many organs. Polymers of deoxygenated SS-hemoglobin distort the shape of red blood cells. These cells attach to endothelia and obstruct small blood vessels, producing frequent, random, and painful sickle cell crises over time. Vessel occlusions in the kidney produce glomerular hypertension, FSGS, interstitial nephritis, and renal infarction associated with hyposthenuria, microscopic hematuria, and even gross hematuria; some patients also present with MPGN. By the second or third decade of life, persistent vasoocclusive disease in the kidney leads to varying degrees of renal failure, and some patients end up on dialysis. Treatment is directed to reducing the frequency of painful crises and administering ACE inhibitors in the hope of delaying a progressive decline in renal function. In sickle cell patients undergoing renal transplantation, renal graft survival is comparable to African Americans in the general transplant population.

Thrombotic Microangiopathies

Thrombotic thrombocytopenic purpura (TTP) and *hemolytic-uremic syndrome* (HUS) represent a spectrum of thrombotic microangiopathies. Thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome share the general features of idiopathic thrombocytopenic purpura, hemolytic anemia, fever, renal failure, and neurologic disturbances. When patients, particularly children, have more evidence of renal injury, their condition tends to be called HUS. In adults with neurologic disease, it is considered to be TTP. In adults there is often a mixture of both, which is why they are often called TTP/HUS. On examination of kidney tissue there is evidence of *glomerular capillary endotheliosis* associated with platelet thrombi, damage to the capillary wall, and formation of fibrin material in and around glomeruli (Fig. e14-23). These tissue findings are similar to what is seen in preeclampsia/HELLP (hemolysis, elevated liver enzymes, and low platelet count syndrome), malignanthypertension, and the antiphospholipid syndrome. Thrombotic thrombocytopenic purpura/hemolytic-uremic syndrome is also seen in pregnancy; with the use of oral contraceptives or quinine; in renal transplant patients given OKT3 for rejection; in patients taking the calcineurin inhibitors, cyclosporine and tacrolimus, or in patients taking the antiplatelet agents, ticlopidine and clopidogrel; or following HIV infection.

Although there is no agreement on how much they share a final common pathophysiology, two general groups of patients are recognized: childhood HUS associated with enterohemorrhagic diarrhea and TTP/HUS in adults. Childhood HUS is caused by a toxin released by *Escherichia coli* 0157:H7 and occasionally by *Shigella dysenteriae.* This shiga toxin (verotoxin) directly injures endothelia, enterocytes, and renal cells, causing apoptosis, platelet clumping, and intravascular hemolysis by binding to the glycolipid receptors (Gb3). These receptors are more abundant along endothelia in children compared to adults. Shiga toxin also inhibits the endothelial production of ADAMTS13. In familial cases of adult TTP/HUS, there is a genetic deficiency of the ADAMTS13 metalloprotease that cleaves large multimers of von Willebrand's factor. Absent ADAMTS13, these large multimers cause platelet clumping and intravascular hemolysis. An antibody to ADAMTS13 is found in many sporadic cases of adult TTP/HUS, but not all; many patients also have antibodies to the thrombospondin receptor on selected endothelial cells in small vessels or increased levels of plasminogen-activator inhibitor 1 (PAI-1). Some children with complement protein deficiencies express atypical HUS (aHUS), which can be treated with liver transplant. The treatment of adult TTP/HUS is daily plasmapheresis,

which can be lifesaving. Plasmapheresis is given until the platelet count rises, but in relapsing patients it normally is continued well after the platelet count improves, and in resistant patients twice-daily exchange may be helpful. Most patients respond within 2 weeks of daily plasmapheresis. Since TTP/HUS often has an autoimmune basis, there is an anecdotal role in relapsing patients for using splenectomy, steroids, immunosuppressive drugs, or rituximab, an anti-CD20 antibody. Patients with childhood HUS from infectious diarrhea are not given antibiotics, as antibiotics are thought to accelerate the release of the toxin and the diarrhea is usually self-limited. No intervention appears superior to supportive therapy in children with postdiarrheal HUS.

Antiphospholipid Antibody Syndrome (see Chap. 320)

INFECTIOUS DISEASE–ASSOCIATED SYNDROMES

A number of infectious diseases will injure the glomerular capillaries as part of a systemic reaction producing an immune response or from direct infection of renal tissue. Evidence of this immune response is collected by glomeruli in the form of immune deposits that damage the kidney, producing moderate proteinuria and hematuria. Some of these infectious diseases represent the most common causes of glomerulonephritis in many parts of the world.

Post-Streptococcal Glomerulonephritis

This form of glomerulonephritis is one of the classic complications of streptococcal infection. The discussion of this disease can be found in the section on *acute nephritic syndromes.*

Subacute Bacterial Endocarditis

Renal injury from persistent bacteremia absent the continued presence of a foreign body, regardless of cause, is treated presumptively as if the patient has endocarditis. The discussion of this disease can be found in the section on *acute nephritic syndromes.*

Human Immunodeficiency Virus

Renal disease is an important complication of HIV disease. The risk of development of end-stage renal disease is much higher in HIV-infected African Americans than in HIV-infected whites. About 50% of HIV-infected patients with kidney disease have HIV-associated nephropathy (HIVAN) on biopsy. The lesion in HIVAN is FSGS, characteristically revealing a collapsing glomerulopathy (Fig. e14-3) with visceral epithelial cell swelling, microcystic dilatation of renal tubules, and tubuloreticular inclusion. Renal epithelial cells express replicating HIV virus, but host immune responses also play a role in the pathogenesis. MPGN and DPGN have also been reported but more commonly in HIV-infected whites and in patients coinfected with hepatitis B or C. HIV-associated TTP has also been reported. Other renal lesions include DPGN, IgA nephropathy, and MCD. Renal biopsy may be indicated to distinguish between these lesions.

HIV patients with FSGS typically present with nephrotic-range proteinuria and hypoalbuminemia, but unlike patients with other etiologies for nephrotic syndrome, they do not commonly have hypertension, edema, or hyperlipidemia. Renal ultrasound also reveals large, echogenic kidneys despite the finding that renal function in some patients declines rapidly. Treatment with inhibitors of the renin-angiotensin system decreases the proteinuria. Effective antiretroviral therapy benefits both the patient and the kidney and improves survival of HIV-infected patient with chronic kidney disease (CKD) or end-stage renal disease. In HIV-infected patients not yet on therapy, the presence of HIVAN is an indication to initiate therapy. Following the introduction of antiretroviral therapy, survival on dialysis for the HIV-infected patient has improved dramatically and is equivalent in patients treated with hemodialysis or peritoneal dialysis. Renal transplants in HIV-infected patients without detectable viral loads or histories of opportunistic infections have a better survival benefit over dialysis. Following transplantation, patient and graft survival are similar to the general transplant population despite frequent rejections.

Hepatitis B and C

Typically infected patients present with microscopic hematuria, nonnephrotic or nephrotic-range proteinuria, and hypertension. There is a close association between hepatitis B infection and polyarteritis nodosa with vasculitis appearing generally in the first 6 months following infection. Renal manifestations include renal artery aneurysms, renal infarction, and ischemic scars. Alternatively, the hepatitis B carrier state can produce a MGN that is more common in children than adults, or MPGN that is more common in adults than in children. Renal histology is indistinguishable from idiopathic MGN or type I MPGN. Viral antigens are found in the renal deposits. There are no good treatment guidelines, but interferon α-2b and lamivudine have been used to some effect in small studies. Children have a good prognosis, with 60–65% achieving spontaneous remission within 4 years. In contrast, 30% of adults have renal insufficiency and 10% have renal failure 5 years after diagnosis.

Up to 30% of patients with chronic hepatitis C infection have some renal manifestations. Patients often present with type II mixed cryoglobulinemia, nephrotic syndrome, microscopic hematuria, abnormal liver function tests, depressed C3 levels, anti–hepatitis C virus (HCV) antibodies, and viral RNA in the blood. The renal lesions most commonly seen, in order of decreasing frequency, are *cryoglobulinemic glomerulonephritis, MGN, and type I MPGN.* Treatment with pegylated interferon and ribavirin is

typical to reduce the viral load.

Other Viruses

Other viral infections are occasionally associated with glomerular lesions, but cause and effect are not well established. These viral infections and their respective glomerular lesions include: cytomegalovirus producing MPGN; influenza and anti-GBM disease; measles-associated endocapillary proliferative glomerulonephritis, with measles antigen in the capillary loops and mesangium; parvovirus causing mild proliferative or mesangioproliferative glomerulonephritis or FSGS; mumps and mesangioproliferative glomerulonephritis; Epstein-Barr virus producing MPGN, diffuse proliferative nephritis, or IgA nephropathy; dengue hemorrhagic fever causing endocapillary proliferative glomerulonephritis; and coxsackievirus producing *focal glomerulonephritis* or DPGN.

Syphilis

Secondary syphilis, with rash and constitutional symptoms, develops weeks to months after the chancre first appears and occasionally presents with the nephrotic syndrome from MGN caused by subepithelial immune deposits containing treponemal antigens. Other lesions have also rarely been described including interstitial syphilitic nephritis. The diagnosis is confirmed with nontreponemal and treponemal tests for *Treponema pallidum.* The renal lesion responds to treatment with penicillin or an alternative drug, if allergic. Additional testing for other sexually transmitted diseases is an important part of disease management.

Leprosy

Despite aggressive eradication programs, approximately 400,000 new cases of leprosy appear annually worldwide. The diagnosis is best made in patients with multiple skin lesions accompanied by sensory loss in affected areas, using skin smears showing paucibacillary or multibacillary infection (WHO criteria). Leprosy is caused by infection with *Mycobacterium leprae* and can be classified by Ridley-Jopling criteria into various types: tuberculoid, borderline tuberculoid, mid-borderline and borderline lepromatous, and lepromatous. Renal involvement in leprosy is related to the quantity of bacilli in the body, and the kidney is one of the target organs during splanchnic localization. In some series, all cases with borderline lepromatous and lepromatous types of leprosy have various forms of renal involvement including FSGS, mesangioproliferative glomerulonephritis, or renal amyloidosis; much less common are the renal lesions of DPGN and MPGN. Treatment with dapsone, rifampicin, and clofazimine can eradicate the infection in nearly all patients.

Malaria

There are 300–500 million incident cases of malaria each year worldwide, and the kidney is commonly involved. Glomerulonephritis is due to immune complexes containing malarial antigens that are implanted in the glomerulus. In malaria from *P. falciparum,* mild proteinuria is associated with subendothelial deposits, mesangial deposits, and mesangioproliferative glomerulonephritis that usually resolve with treatment. In quartan malaria from infection with *P. malariae,* children are more commonly affected and renal involvement is more severe. Transient proteinuria and microscopic hematuria can resolve with treatment of the infection. However, resistant nephrotic syndrome with progression to renal failure over 3–5 years does happen, as <50% of patients respond to steroid therapy. Affected patients with nephrotic syndrome have thickening of the glomerular capillary walls, with subendothelial deposits of IgG, IgM, and C_3 associated with a sparse membranoproliferative lesion. The rare mesangioproliferative glomerulonephritis reported with *P. vivax* or *P. ovale* typically has a benign course.

Schistosomiasis

Schistosomiasis affects more than 300 million people worldwide and primarily involves the urinary and gastrointestinal tracts. Glomerular involvement varies with the specific strain of schistosomiasis; *Schistosoma mansoni* is most commonly associated with clinical renal disease, and the glomerular lesions can be classified: Class I is a *mesangioproliferative glomerulonephritis;* class II is an *extracapillary proliferative glomerulonephritis;* class III is a *membranoproliferative glomerulonephritis;* class IV is a *focal segmental glomerulonephritis;* and class V lesions have *amyloidosis.* Classes I–II often remit with treatment of the infection, but classes III and IV lesions are associated with IgA immune deposits and progress despite antiparasitic and/or immunosuppressive therapy.

Other Parasites

Renal involvement with toxoplasmosis infections is rare. When it occurs, patients present with nephrotic syndrome and have a histologic picture of MPGN. Fifty percent of patients with leishmaniasis will have mild to moderate proteinuria and microscopic hematuria, but renal insufficiency is rare. Acute DPGN, MGN, and mesangioproliferative glomerulonephritis have all been observed on biopsy. Filariasis and trichinosis are caused by nematodes and are sometimes associated with glomerular injury presenting with proteinuria, hematuria, and a variety of histologic lesions that typically resolve with eradication of the infection.

FURTHER READINGS

Beck lh Jr, Salant DJ: Glomerular and tubulointerstitial diseases. Prim Care 35:265, 2008[PMID: 18486716]

Berthoux FC et al: Natural history of primary IgA nephropathy. Semin Nephrol 28:4, 2008[PMID: 18222341]

Bomback AS, Appel GB: Updates on the treatment of lupus nephritis. J Am Soc Nephrol 21:2028, 2010[PMID: 21051743]

Falk RJ, Jennette JC: ANCA disease: Where is this field heading? J Am Soc Nephrol 21:745, 2010[PMID: 20395376]

Freedman BI, Sedor Jr: Hypertension-associated kidney disease: perhaps no more. J Am Soc Nephrol 19: 2047, 2008[PMID: 18923054]

Hudson BG et al: Alport and Goodpasture syndromes and the type IV collagen family. N Engl J Med 348:2543, 2003[PMID: 12815141]

Peti-Peterdi J, Sipos A: A high-powered view of the filtration barrier. J Am Soc Nephrol 21:1835, 2010[PMID: 20576805]

Rodriguez-Iturbe B, Musser JM: The current state of poststreptococcal glomerulonephritis. J Am Soc Nephrol 19: 1855, 2008[PMID: 18667731]

Smith Rjh et al: New approaches to the treatment of dense deposit disease. J Am Soc Nephrol 18: 2447, 2007[PMID: 17675665]

Tervaert TW et al: Pathologic classification of diabetic nephropathy. J Am Soc Nephrol 21:556, 2010[PMID: 20167701]

Tryggvason K et al: Hereditary proteinuria syndromes and mechanisms of proteinuria. N Engl J Med 354:1387, 2006[PMID: 16571882]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

The polycystic kidney diseases are among the most common life-threatening inherited diseases worldwide and frequently cause kidney failure. Autosomal dominant polycystic kidney disease (ADPKD) is seen predominantly in adults (**Fig. 284-1**), whereas autosomal recessive polycystic kidney disease (ARPKD) is mainly a disease of childhood. Renal cysts also are seen in several other hereditary kidney diseases (**Table 284-1**), some of which may have defects in a common signaling pathway with ADPKD and ARPKD. Other inherited tubular diseases manifest primarily with alterations in fluid, electrolyte, acid-base, and mineral balance (Table 284-2).

Table 284-1 Inherited Cystic Kidney Diseases

Disease (OMIM)	Mode of Inheritance	Locus	Gene	Protein	Renal Abnormalities	Extrarenal Abnormalities
Autosomal dominant polycystic kidney disease (601313, 173910)	AD	16p13	*PKD1*	Polycystin-1	Cortical and medullary cysts	Cerebral aneurysms; liver cysts, other[a]
	AD	4q21	*PKD2*	Polycystin-2	Cortical and medullary cysts	Cerebral aneurysms; liver cysts, other[a]
Autosomal recessive polycystic kidney disease (263200)	AR	6p21	*PKHD1*	Fibrocystin (polyductin)	Distal tubule and collecting duct cysts	Hepatic fibrosis; Caroli's disease
Nephronophthisis I (juvenile/adolescent, 256100)[b]	AR	2q13	*NPHP1*	Nephrocystin	Small fibrotic kidneys; medullary cysts	Retinitis pigmentosa
Nephronophthisis II (infantile, 602088)[b]	AR	9q31	*NPHP2 (INVS)*	Inversin	Large kidneys; widespread cysts	Situs inversus
Nephronophthisis III (juvenile/adolescent, 604387)[b]	AR	3q22	*NPHP3*	Nephrocystin-3	Small fibrotic kidneys; medullary cysts	Retinitis pigmentosa; hepatic fibrosis
Medullary cystic kidney disease (174000, 603860)	AD	1q21	*MCKD1*	Unknown	Small fibrotic kidneys; medullary cysts	None
	AD	16p12	*MCKD2 (UMOD)*	Uromodulin (Tamm-Horsfall protein)	Small fibrotic kidneys; medullary cysts	Hyperuricemia and gout
Tuberous sclerosis (191100)	AD	9q34	*TSC1*	Hamartin	Renal cysts; angiomyolipomas; renal cell carcinoma	Facial angiofibromas; CNS hamartomas
	AD	16p13	*TSC2*	Tuberin	Renal cysts; angiomyolipomas; renal cell carcinoma	Facial angiofibromas; CNS hamartomas
Von Hippel-Lindau disease (608537)	AD	3p26-p25	*VHL*	pVHL	Renal cysts; renal cell carcinoma	Retinal angiomas; CNS hemangioblastomas; pheochromocytomas

[a] See text for details.

[b] The three variants of nephronophthisis listed in the table are the most prevalent of the currently described 11 forms of nephronophthisis. Each variant has similar renal abnormalities but varying extrarenal phenotypes.

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; OMIM, online Mendelian inheritance in man.

Figure 284-1

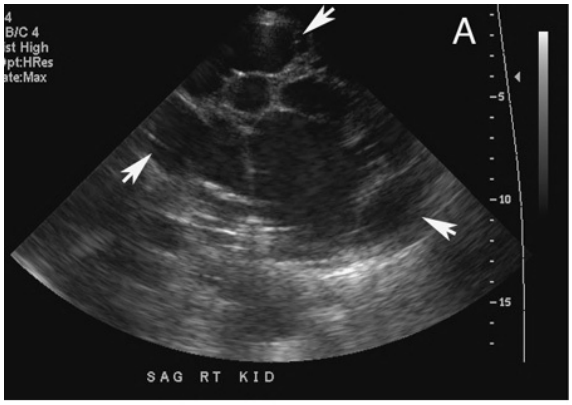

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

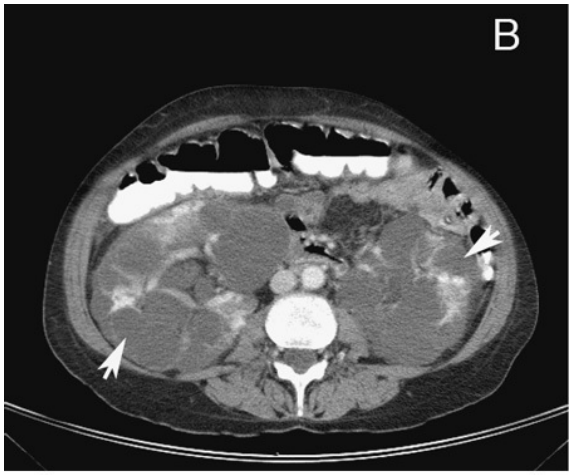

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

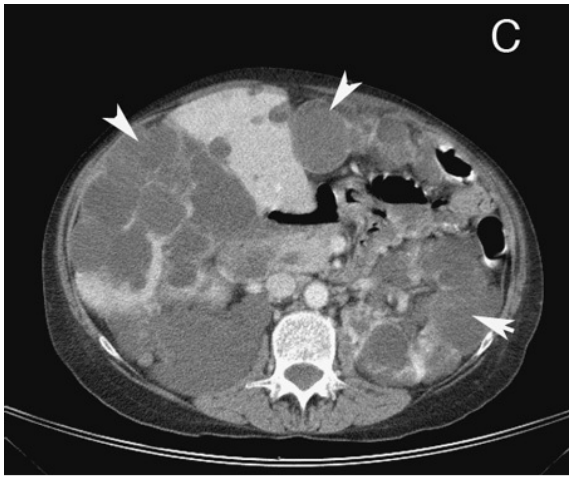

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Renal ultrasonogram and contrast-enhanced abdominal CT scan in a 56-year-old woman with autosomal dominant polycystic kidney disease. ***A***. Sonogram of the right kidney showing numerous cysts of varying sizes (*arrows*). ***B***. Abdominal CT scan demonstrating bilaterally enlarged kidneys with large cysts (*arrows*). ***C***. Multiple liver cysts (*arrowheads*) and renal cysts (*arrow*) are seen in an upper abdominal image.

Autosomal Dominant Polycystic Kidney Disease

ETIOLOGY AND PATHOGENESIS

ADPKD is a systemic disorder resulting from mutations in either the *PKD-1* or the *PKD-2* gene. The *PKD-1*-encoded protein, polycystin-1, is a large receptor-like molecule, whereas the *PKD-2* gene product, polycystin-2, has features of a calcium channel protein. Both are transmembrane proteins that are present throughout all segments of the nephron. They have been localized to the luminal surface of tubular cells in primary cilia, where they appear to serve as flow sensors; on the basal surface in focal adhesion complexes; and on the lateral surface in adherens junctions. The proteins are thought to function independently, or as a complex, to regulate fetal and adult epithelial cell gene transcription, apoptosis, differentiation, and cell-matrix interactions. Disruption of these processes leads to epithelial dedifferentiation, unregulated proliferation and apoptosis, altered cell polarity, disorganization of surrounding extracellular matrix, excessive fluid secretion, and abnormal expression of several genes, including some that encode growth factors. Vasopressin-mediated elevation of cyclic AMP levels in cystepithelia plays a major role in cystogenesis by stimulating cell proliferation and fluid secretion into the cyst lumen through apical chloride

and aquaporin channels. Cyst formation begins in utero from any point along the nephron, although <5% of total nephrons are thought to be involved. As the cysts accumulate fluid, they enlarge, separate entirely from the nephron, compress the neighboring renal parenchyma, and progressively compromise renal function.

Genetic Considerations

ADPKD occurs in 1:400–1:1000 individuals worldwide and accounts for ~4% of end-stage renal disease (ESRD) in the United States. ADPKD is equally prevalent in all ethnic and racial groups. Over 90% of cases are inherited as an autosomal dominant trait, with the remainder probably representing spontaneous mutations. Mutations in the *PKD-1* gene on chromosome 16 (ADPKD-1) account for 85% of cases, and mutations in the *PKD-2* gene on chromosome 4 (ADPKD-2) account for the remainder. A few families appear to have a defect at a site that is different from either of these loci. Direct mutation analysis of isolated cysts suggests that there is loss of heterozygosity, whereby a somatic mutation in the normal allele of a small number of tubular epithelial cells leads to unregulated clonal proliferation of the cells that ultimately form the cyst lining.

CLINICAL FEATURES

Phenotypic heterogeneity is a hallmark of ADPKD, as evidenced by family members who have the same mutation but have a different clinical course. Affected individuals are often asymptomatic into the fourth or fifth decade. Presenting symptoms and signs include abdominal discomfort, hematuria, urinary tract infection, incidental discovery of hypertension, abdominal masses, elevated serum creatinine, and cystic kidneys on imaging studies (Fig. 284-1*A* and *B*). Frequently, the diagnosis is made before the onset of symptoms, when asymptomatic members in affected families request screening. In most patients, renal function declines progressively over the course of 10–20 years from the time of diagnosis, but not everyone with ADPKD develops ESRD; it occurs in about 60% of these patients by age 70. Those with ADPKD-2 tend to have later onset and slower progression. Hypertension is common and often precedes renal dysfunction, perhaps mediated by increased activity of the renin-angiotensin system. There is only mild proteinuria, and impaired urinary concentrating ability manifests early as polyuria and nocturia. Risk factors for progressive kidney disease include younger age at diagnosis, black race, male sex, presence of polycystin-1 mutation, and hypertension. There is a close correlation between the rate of kidney expansion, as measured by magnetic resonance imaging (MRI) scanning, and the rate of decline in kidney function. Dull, persistent flank and abdominal pain and early satiety are common due to the mass effect of the enlarged kidneys or liver. Cyst rupture or hemorrhage into a cyst may produce acute flank pain or symptoms and signs of localized peritonitis. Gross hematuria may result from cyst rupture into the collecting system or from uric acid or calcium oxalate kidney stones. Nephrolithiasis occurs in about 20% of patients. Urinary tract infection, including acute pyelonephritis, occurs with increased frequency in ADPKD. Infection in a kidney cyst is a particularly serious complication. It is most often due to Gram-negative bacteria and presents with flank pain, fever, and chills. Blood cultures are frequently positive, but urine culture may be negative because infected kidney cysts do not communicate directly with the collecting system. Distinguishing between infection and cyst hemorrhage is often challenging, and the diagnosis relies mainly on clinical and bacteriologic findings. Radiologic and nuclear imaging studies are generally not helpful.

Numerous extrarenal manifestations of ADPKD highlight the systemic nature of the disease. Patients with ADPKD have a twofold to fourfold increased risk of subarachnoid or cerebral hemorrhage from a ruptured intracranial aneurysm compared with the general population. Saccular aneurysms of the anterior cerebral circulation may be detected in up to 10% of asymptomatic patients on magnetic resonance angiography (MRA) screening, but most are small, have a low risk of spontaneous rupture, and do not merit the risk of intervention. In general, hemorrhage tends to occur before age 50 years, in patients with a family history of intracranial hemorrhage, and in those who have survived a previous bleed, have aneurysms >10 mm, and have uncontrolled hypertension. Other vascular abnormalities include aortic root and annulus dilation. Cardiac valvular abnormalities occur in 25% of patients, most commonly mitral valve prolapse and aortic regurgitation. Although most valvular lesions are asymptomatic, some may progress over time and warrant valve replacement. The incidence of hepatic cysts is 83% by MRI in patients age 15–46 years. Most patients are asymptomatic with normal liver function tests, but hepatic cysts may bleed, become infected, rupture, and cause pain. Although the frequency of liver cysts is equal between the sexes, women are more likely to have massive cysts (Fig. 284-1*C*). Colonic diverticulae are common, with a higher incidence of perforation in patients with ADPKD. Abdominal wall and inguinal hernias also occur with a higher frequency than in the general population.

DIAGNOSIS AND SCREENING

Most often, the diagnosis of ADPKD is made from a positive family history and imaging studies showing large kidneys with multiple bilateral cysts and possibly liver cysts (Fig. 284-1). Criteria for the diagnosis of ADPKD by ultrasonography in asymptomatic individuals account for the later onset of ADPKD-2 and assume that the genotype of the individual and family being tested is unknown. The presence of three or more cysts in one or both kidneys is required to diagnose ADPKD in patients age 15–39 years with a specificity and positive predictive value of 100%; sensitivity varies from 82 to 96% for persons age 15–29 and 30–39 years, respectively. The presence of two or more cysts in each kidney is associated with a sensitivity and specificity of 90% and 100%, respectively, in patients age 40–59 years. In subjects older than 60 years, the presence of four or more cysts in each kidney is required to diagnose ADPKD because of the increased frequency of benign simple cysts, whereas fewer than two renal cysts in at-risk individuals age ≥40 years is sufficient to exclude the disease. Computed tomography (CT) scan and T2-weighted MRI are more sensitive for detecting presymptomatic disease in young patients. Genetic linkage analysis and mutational screening for *ADPKD-1* and *ADPKD-2* is available for equivocal cases, especially when a young adult from an affected family is being considered as a potential kidney donor. Genetic counseling is essential for those being screened. Screening for asymptomatic intracranial aneurysms should be restricted to patients with a personal or family history of intracranial hemorrhage and those in high-risk occupations. Intervention should be limited to aneurysms larger than 10 mm.

TREATMENT: AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

No treatment has been proved to prevent cyst growth or the decline in kidney function. Hypertension control with a target blood pressure of 130/80 mmHg or less is recommended according to Joint National Committee (JNC) VII guidelines. A multidrug approach that includes agents to inhibit the renin-angiotensin system is frequently required. Studies are investigating the role of angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs) in slowing growth of kidney volume and loss of glomerular filtration rate (GFR). Lipid-soluble antimicrobials such as trimethoprim-sulfamethoxazole and fluoroquinolones that have good cyst penetration are the preferred therapy for infected kidney and liver cysts. Pain management occasionally requires cyst drainage by percutaneous aspiration, sclerotherapy with alcohol, or, rarely, surgical drainage. Patients with ADPKD appear to have a survival advantage on either peritoneal or hemodialysis compared with patients with other causes of ESRD. Those undergoing kidney transplantation may require bilateral nephrectomy if the kidneys are massively enlarged or have been the site of infected cysts. Posttransplantation survival rates are similar to those of patients with other causes of kidney failure, but these patients remain at risk for the extrarenal complications of ADPKD. Studies in animal models of inherited cystic diseases have identified promising therapeutic strategies, including vasopressin V2 receptor antagonists that suppress cyst growth by lowering intracellular cAMP, and inhibitors of cell dedifferentiation and proliferation that target the epidermal growth factor receptor tyrosine kinase and the mammalian target of rapamycin (mTOR). Clinical trials of these agents are ongoing.

Autosomal Recessive Polycystic Kidney Disease

GENETIC CONSIDERATIONS

ARPKD is primarily a disease of infants and children. The incidence is 1:20,000 births. The kidneys are enlarged, with small cysts, <5 mm, limited to the collecting tubules. The ARPKD gene on chromosome 6p21, *PKHD1*, encodes several alternatively spliced transcripts (Table 284-1). The largest transcript produces a multidomain transmembrane protein termed *fibrocystin (polyductin)* that is found in the cortical and medullary collecting ducts and the thick ascending limb of Henle's loop in the kidney as well as in biliary and pancreatic duct epithelia. Like the polycystins, fibrocystin has receptor-like features and may be involved in cell-cell and cell-matrix interactions. Fibrocystin, the polycystins, and several proteins involved in animal models of PKD are located in association with primary cilia on the tubular epithelial cell apical surface; this suggests that they may cooperate in a mechanosensory pathway. A large number of different mutations have been identified throughout *PKHD1* and are unique to individual families. Most patients are compound heterozygotes. Those with two truncating mutations frequently die shortly after birth, whereas those who survive beyond the neonatal period generally have at least one missense mutation. Mutations in *PKHD1* have also been identified in about 30% of children with congenital hepatic fibrosis (Caroli's syndrome) without evident kidney involvement.

CLINICAL FEATURES

The clinical presentation of ARPKD is highly variable. Up to 50% of affected neonates die of pulmonary hypoplasia, the result of oligohydramnios from severe intrauterine kidney disease. About 80% of those who survive the neonatal period are still alive after 10 years; however, one-third will have developed ESRD. Enlarged kidneys may be detected soon after birth as bilateral abdominal masses. Impaired urinary concentrating ability and metabolic acidosis ensue as tubular function deteriorates. Hypertension often occurs in the first few years of life. Kidney function deteriorates progressively from childhood into early adult life. Longer-term survivors frequently develop complications of portal hypertension from periportal fibrosis.

DIAGNOSIS

Ultrasonography reveals large, echogenic kidneys. The diagnosis can be made in utero after 24 weeks of gestation in severe cases, but cysts generally become visible only after birth. The absence of renal cysts in either parent on ultrasonography helps distinguish ARPKD from ADPKD in older patients. The wide range of different mutations and the large size of the gene complicate molecular diagnosis, although prenatal diagnosis is possible by gene linkage to the *PKHD1* locus in families with a previous confirmed ARPKD birth.

TREATMENT: AUTOSOMAL RECESSIVE POLYCYSTIC KIDNEY DISEASE

There is no specific therapy for ARPKD. Improvements in neonatal intensive care, blood pressure management, dialysis, and kidney transplantation have led to survival into adulthood. Complications of hepatic fibrosis may necessitate liver transplantation. Future therapies may target aberrant cell signaling mechanisms, as in ADPKD.

Nephronophthisis

GENETICS AND PATHOGENESIS

Nephronophthisis (NPHP) is the most common genetic cause of ESRD in childhood and adolescence. Eleven distinct genetic mutations with autosomal recessive inheritance have been identified to date and produce different renal and extrarenal manifestations of NPHP (Table 284-1). Although their precise functions are unclear, the defective protein products, named nephrocystins and inversin, localize to the primary cilium and associated basal body of renal epithelial cells, similar to the polycystins and fibrocystin. NPHP is classified into infantile, juvenile, and adolescent forms based on the age of ESRD onset. In juvenile NPHP, the most common form, the kidneys are shrunken and histology shows tubular atrophy, thickening of tubular basement membranes, diffuse interstitial fibrosis, and microscopic medullary cysts. In the infantile form, the kidneys are large with histology similar to that of the juvenile form except that medullary cysts are more prominent and develop earlier.

CLINICAL FEATURES

In juvenile NPHP symptoms typically appear after 1 year of age. Impaired tubular function causes salt wasting and defective urinary concentration and acidification. Patients may present with polyuria, polydipsia, volume depletion, or systemic acidosis. Hypertension is usually absent due to salt wasting. Progressive kidney failure and volume depletion lead to growth retardation. On average, ESRD occurs by age 3 in the infantile form, age 13 in the juvenile form, and age 19 in the adolescent form. Up to 15% of patients with juvenile NPHP have extrarenal manifestations (Table 284-1), most commonly retinitis pigmentosa (Senior-Loken syndrome). Other abnormalities include blindness from amaurosis, oculomotor apraxia, cerebellar ataxia (Joubert syndrome), polydactyly, mental retardation, hepatic fibrosis, and ventricular septal defect. Situs inversus is seen in some cases of infantile NPHP, consistent with mutation in *INVS (NPHP2)*, a gene critical for left-right patterning in the embryo.

DIAGNOSIS

The diagnosis of NPHP should be considered in patients with a family history of kidney disease, early-onset progressive renal failure, and a bland urine sediment with minimal proteinuria. Ultrasonography reveals small hyperechoic kidneys in juvenile NPHP and large kidneys with cysts in the infantile form.

TREATMENT: NEPHRONOPHTHISIS

There is no specific therapy to prevent loss of kidney function in NPHP. Salt and water replacement are required for patients with salt wasting and polyuria. Therapy should include sodium bicarbonate or citrate for acidosis, management of chronic kidney disease, and timely institution of dialysis and kidney transplantation. NPHP does not recur in transplanted kidneys.

Medullary Cystic Kidney Disease

GENETIC CONSIDERATIONS

The medullary cystic kidney diseases (MCKDs) generally present in young adults. Two genetic loci have been defined, both with autosomal dominant transmission (Table 284-1). The locus for MCKD1 has been mapped to chromosome 1q21. Mutations in the uromodulin gene (*UMOD*) that encodes the Tamm-Horsfall mucoprotein on chromosome 16p12 have been identified in MCKD2.

CLINICAL FEATURES

As with NPHP, patients with MCKD have atrophic kidneys with diffuse interstitial fibrosis, cysts restricted to the renal medulla, salt wasting, and polyuria. Disease onset is later than in NPHP. Consequently, there is no growth retardation, salt wasting is milder, and ESRD occurs later, usually between ages 20 and 70. There are no extrarenal manifestations in MCKD1, but most patients with MCKD2 have severe hyperuricemia and precocious onset of gout.

DIAGNOSIS

MCKD should be considered in young adults with a family history suggesting dominant inheritance of kidney disease who present with progressive renal failure, bland urinalysis with little or no proteinuria, and small dense kidneys with medullary cysts on radiographic imaging. The presence of hyperuricemia and gout is a further clue to the diagnosis of MCKD2, which can be confirmed by mutation analysis of *UMOD*.

TREATMENT: MEDULLARY CYSTIC KIDNEY DISEASE

There is no specific therapy for MCKD. Allopurinol is indicated for patients with gout and is reasonable for those with asymptomatic hyperuricemia, although there is no evidence that it prevents progressive renal failure in MCKD2. Dialysis and transplantation outcomes appear to be favorable. The disease does not recur in transplanted kidneys.

Tuberous Sclerosis

Tuberous sclerosis (TS) is an autosomal dominant disorder that affects 1 in 6000 people. It results from mutations in either the *TSC1* gene encoding hamartin or the *TSC2* gene encoding tuberin (Table 284-1). Hamartin and tubulin form a complex that is thought to negatively regulate cell growth and proliferation through inhibition of mTOR. The presence of either mutation produces uncontrolled proliferation in numerous tissues, including the kidneys, skin, central nervous system, and heart. The kidneys are affected in 80% of patients. Renal TS occurs in three forms: renal angiomyolipomas, renal cysts, and renal cell carcinoma. Angiomyolipomas are the most common renal abnormality. They occur bilaterally, are often multiple, and are usually asymptomatic; however, they may cause spontaneous bleeding, flank pain, hematuria, and life-threatening retroperitoneal hemorrhage. Large lesions, >4 cm, are more likely to be symptomatic and may require transcatheter arterial embolization or surgical excision. Cysts are usually asymptomatic and are not evident on imaging studies until adulthood. Rarely, cysts may be large and numerous, sometimes leading to ESRD and producing a clinical scenario that can be confused with ADPKD, especially if there are few other systemic manifestations of TS. Multicentric renal cell carcinomas occur with increased frequency in TS. Patients with TS should be screened for renal involvement at initial diagnosis with ultrasonography or CT. Those with cysts or angiomyolipomas require regular imaging to monitor for the development of renal cell carcinoma.

Von Hippel-Lindau Disease

Von Hippel-Lindau disease (VHL) is a rare autosomal dominant disease characterized by abnormal angiogenesis with benign and malignant tumors that affect multiple tissues. The disease is inherited as a mutation in one allele of the *VHL* tumor-suppressor gene. Somatic mutation of the normal allele leads to retinal angiomas, central nervous system (CNS) hemangioblastomas, pheochromocytomas and multicentric clear cell cysts, hemangiomas, and adenomas of the kidney. The kidneys are affected in three-quarters of patients, and half these patients develop clear cell carcinomas in the renal cysts. It is noteworthy that *VHL* mutations also account for 60% of spontaneous clear cell carcinomas of the kidney. The mean age of diagnosis of renal cell carcinoma in VHL disease is 44 years, and 70% of patients who survive to age 60 develop renal cell carcinoma. The high risk of renal cell carcinoma mandates periodic surveillance (usually yearly in adults) by CT or MRI. Routine screening and awareness of the natural history of lesions has enabled renal-sparing approaches to disease management. Tumors <3 cm in size require careful monitoring for growth, whereas partial nephrectomy is indicated in those >3 cm in the absence of metastasis. Nonsurgical renal-sparing strategies, including percutaneous radio frequency ablation and selective arterial embolization, have shown promise in short-term trials.

Medullary Sponge Kidney

PATHOLOGY AND CLINICAL FEATURES

Medullary sponge kidney (MSK) is a relatively common benign condition of unknown cause characterized by ectasia of the papillary collecting ducts of one or both kidneys. Urinary stasis in the dilated ducts, hypocitraturia, and occasionally incomplete distal renal tubular acidosis (dRTA) contribute to the formation of small calcium-containing calculi. Most cases are asymptomatic or are discovered during investigation of hematuria. Symptomatic patients typically present as young adults with renal colic and nephrolithiasis or recurrent urinary tract infections; however, MSK also may affect children. Most cases are sporadic, although MSK has been found rarely in association with other congenital anomalies of the urinary tract and with congenital hepatic ductal ectasia (Caroli's disease).

DIAGNOSIS

MSK is characteristically seen as hyperdense papillae with clusters of small stones on renal ultrasonography or abdominal x-ray (**Fig. 284-2**). The classical "paintbrush-like" features of MSK, representing the ectatic collecting ducts, are best seen on intravenous urography. However, this procedure has been supplanted by contrast-enhanced, high-resolution helical CT with digital reconstruction (Fig. 284-2).

Figure 284-2

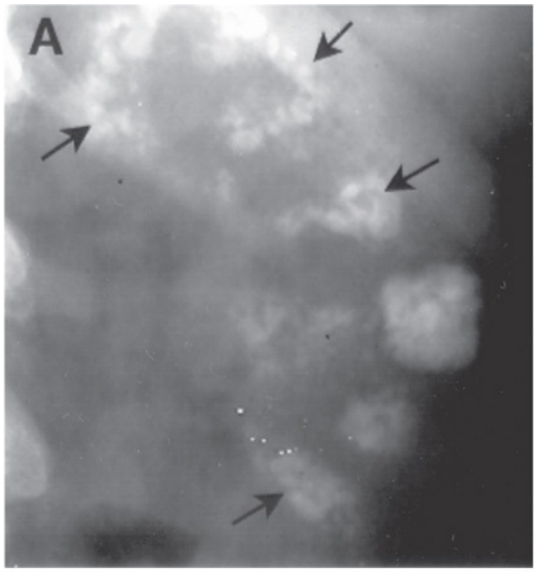

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

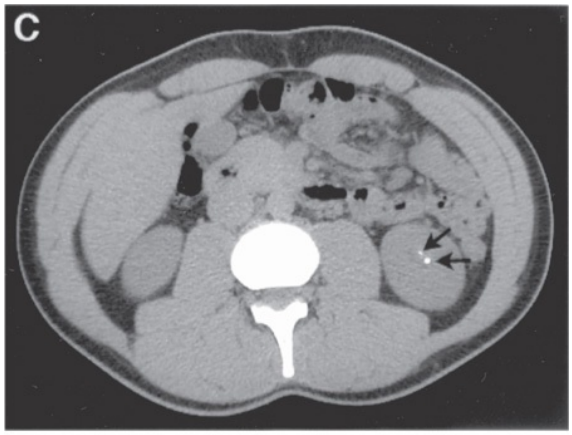

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

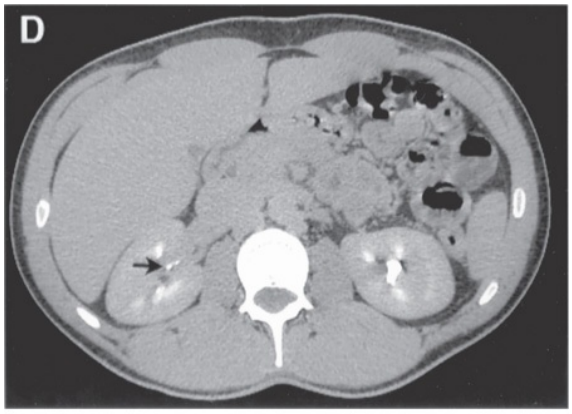

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

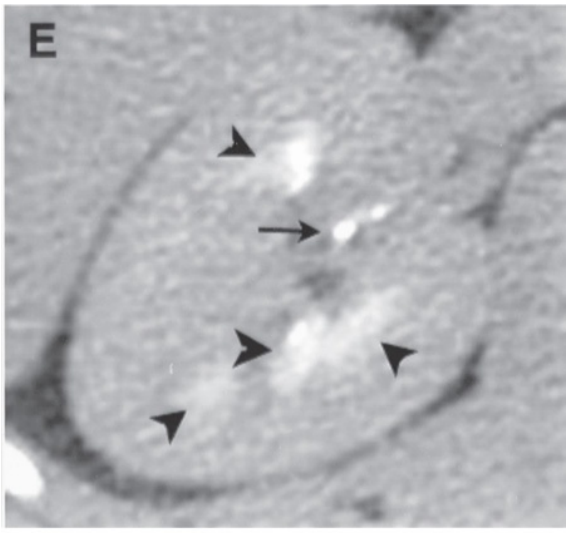

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiographs of medullary sponge kidney disease. *A.* Plain x-ray film of a patient with a history of recurrent nephrolithiasis showing clusters of stones in the papillae (arrows). ***B–E***. CT scan of an 18-year-old male patient investigated for persistent microscopic hematuria. ***B*** and ***C***. CT without contrast showing a few small stones in the papillae (*arrows*). ***D*** and ***E***. Contrast-enhanced CT of the same region shown in *B*. In addition to the stone (*arrow*), a blush of contrast is seen filling the ectatic collecting ducts (*arrowheads*).

TREATMENT: MEDULLARY SPONGE KIDNEY

No treatment is necessary in asymptomatic individuals, aside from maintaining high fluid intake to reduce the risk of nephrolithiasis. Recurrent stone formation should prompt a metabolic evaluation and treatment as in any stone former (Chap. 287). In patients with hypocitraturia and incomplete dRTA, treatment with potassium citrate helps prevent new stone formation. Urinary tract infections should be treated promptly.

HEREDITARY DISORDERS OF SODIUM, POTASSIUM, AND MAGNESIUM HANDLING WITHOUT HYPERTENSION

Inherited forms of hypochloremic metabolic alkalosis and hypokalemia without hypertension are due to genetic mutations of various ion transporters and channels of the thick ascending limb of Henle's loop (TAL) and distal convoluted tubule (DCT) (**Table 284-2** and **Fig. 284-3**). In 1962 Bartter described two patients with a syndrome of metabolic alkalosis, hypovolemia, and failure to thrive associated with juxtaglomerular apparatus hyperplasia, hyperaldosteronism, and normal blood pressure. Subsequently, Gitelman identified a similar but milder syndrome accompanied by hypomagnesemia from urinary magnesium wasting and presenting in later childhood and adolescence. These disorders are now known to occur sporadically or result from genetically heterogeneous loss-of-function autosomal recessive mutations that cause salt-losing tubulopathy.

Table 284-2 Inherited Tubular Disorders

Disease (OMIM)	Mode of Inheritance	Locus	Gene	Protein	Renal Abnormalities	Extrarenal Abnormalities
Bartter's syndrome Type 1 (601678)	AR	15q15	*SLC12A1*	NKCC2	Salt wasting; hypokalemia	
Type 2 (241200)	AR	11q24	*KCNJ1*	ROMK	Salt wasting; hypokalemia	
Type 3 (607364)	AR	1p36	*ClCNKb*	CLC-Kb	Salt wasting; hypokalemia	
Type 4 (602023)	AR	1p31	*BSND*	Barttin	Salt wasting; hypokalemia	Sensorineural deafness
Type 5 (601199)	AD	3q13	*CASR*	CASR	Salt wasting; hypokalemia	
Gitelman's syndrome (263800)	AR	16q13	*SLC12A3*	NCCT	Salt wasting; hypokalemia; hypomagnesemia	
Pseudohypoaldosteronism type I (264350, 177735)	AR	16p13 16p13 12p13	*SCNN1B* *SCNN1G* *SCNN1A*	α, β, or γ subunit of ENaC	Hyperkalemia; salt wasting	Increased lung secretions and lung infections
Familial hypomagnesemia with hypercalciuria and nephrocalcinosis (FHHNC) (248250, 248190)	AD AR	4q31 3q27 1p34	*NR3C2* *CLDN16* *CLDN19*	Mineralocorticoid receptor (type I) Claudin 16 Claudin 19	Hyperkalemia; salt wasting Hypomagnesemia; nephrocalcinosis	Ocular abnormalities (claudin 19 defect only)
Hypomagnesemia with secondary hypocalcemia (HSH) (602014)	AR	9q22	*TRPM6*	TRPM6	Hypomagnesemia; hypocalcemia	
Autosomal dominant hypomagnesemia (154020)	AD	11q23	*FXYD2*	γ subunit of basolateral Na/K-ATPase of DCT	Hypomagnesemia; hypocalciuria	
Autosomal dominant	AD	3q13	*CASR*	CASR	Hypocalcemia;	

hypoparathyroidism (601199)					hypercalciuria; hypomagnesemia	
Isolated autosomal recessive hypomagnesemia (611718)	AR	4q25	*EGF*	EGF	Hypomagnesemia	
Liddle's syndrome (177200)	AD	16p13	*SCNN1B* *SCNN1G*	β and γ subunits of ENaC	Hypertension; hypokalemia; alkalosis	
Pseudohypoaldosteronism type II (Gordon's syndrome, 145260)	AD	12p13 17q21	*WNK1* *WNK4*	WNK 1 WNK 4	Hypertension; hyperkalemia	
Nephrogenic DI type 1 (304800)	XL	Xq28	*AVPR2*	AVPR2	Renal concentrating defect	
Nephrogenic DI type 2 (125800)	AR, AD	12q13	*AQP2*	AQP2	Renal concentrating defect	
Nephrogenic syndrome of inappropriate antidiuresis (300539)	XL	Xq28	*AVPR2*	AVPR2	Hyponatremia	
Distal renal tubular acidosis (267300, 602722, 259730, 179800)	AR	2cenq13 7q33	*ATP6V1B1* *ATP6VOA4*	H^+-ATPase (B1)H^+-ATPase (α4)	Hyperchloremic metabolic acidosis; nephrocalcinosis	Sensorineural deafness (B1 defect only); growth retardation
	AR	8q22	*CA2*	CA2	Proximal and distal RTA	Osteopetrosis, short stature, mental retardation
	AD	17q21	*SLC4A1*	AE1	Distal RTA	
Proximal renal tubular acidosis (604278)	AR	4q21	*SLC4A4*	NBC-1	Moderate hyperchloremic metabolic acidosis	Glaucoma; band keratopathy
Cystinuria (220100)	AR	2p16 19q13	*SLC3A1* *SLC7A9*	rBAT$b^{o,+}$AT1	Cystine stones; dibasic aminoaciduria	
Hartnup disease (234500)	AR	5p15	*SLC6A19*	B^oAT1	Neutral aminoaciduria	Dermatitis, ataxia; dementia
Dent's disease (300009)	XL	Xp11	*CLCN5*	CLC-5	Fanconi syndrome; nephrocalcinosis	Osteomalacia; rickets
Cystinosis (219800)	AR	17p13	*CTNS*	Cystinosin	Fanconi syndrome; progressive kidney failure	Ocular, muscular, liver, gonadal, and thyroid involvement
Renal glucosuria (233100)	AR	16p11	*SLC5A2*	SGLT2	Glucosuria	
Hereditary hypophosphatemic rickets with hypercalciuria (HHRH, 241530)	AR	9q34	*SLC34A3*	Sodium-phosphate co-transporter	Hypophoshatemia; hypercalciuria	Rickets
Vitamin D–dependent rickets type I (VDDR I, 264700)	AR	12q14.1	*CYP27B1*	25-vitamin D3-1-α-hydroxylase	Hypocalcemia	Rickets

Abbreviations: AD, autosomal dominant; AE1, anion exchanger 1; AR, autosomal recessive; AT1, amino acid transporter; AVPR2, arginine vasopressin receptor 2; CA2, carbonic anhydrase II; CASR, calcium-sensing receptor; CLC-5, chloride channel 5; CLC-Kb, chloride channel Kb; DI, diabetes insipidus; ENaC, amiloride-sensitive epithelial sodium channel; NBC, sodium-bicarbonate co-transporter; NCCT, thiazide-senstitive Na-Cl co-transporter; NKCC2, Na-K-2Cl co-transporter; OMIM, online Mendelian inheritance in man; rBAT, renal basic amino acid transport glycoprotein; ROMK, renal outer medullary potassium channel; SGLT2, sodium/glucose co-transporter.TRPM6, transient receptor potential cation channel, subfamily M, member 6; WNK, with no lysine (K); XL, X-linked;

Figure 284-3

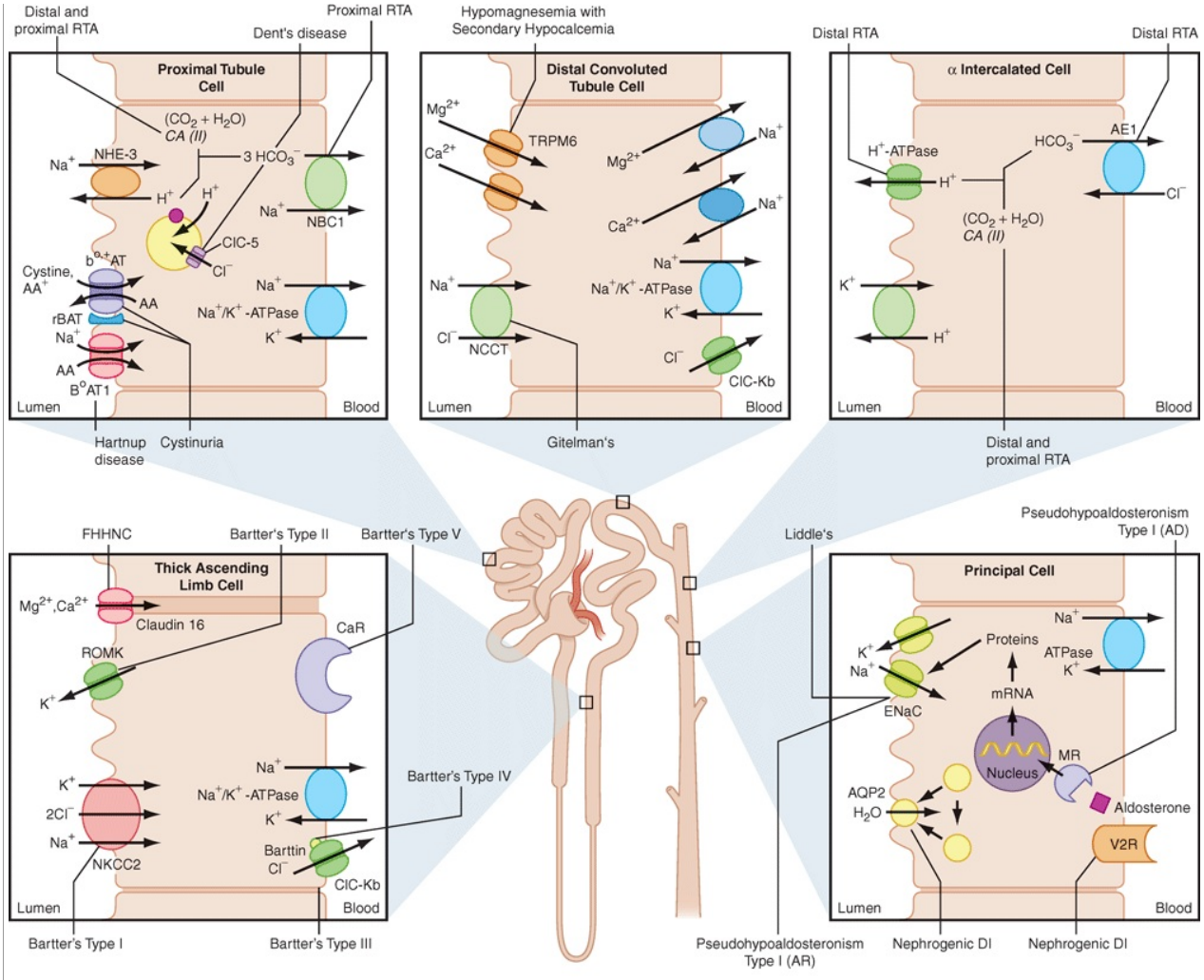

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of channels, transporters, and enzymes associated with hereditary renal tubular disorders. AD, autosomal dominant; AR, autosomal recessive; DI, diabetes insipidus; NKCC2, Na-K-2Cl co-transporter; ROMK, renal outer medullary potassium channel; AQP2, aquaporin-2; CLC-Kb, chloride channel Kb; CaR, calcium-sensing receptor; NCCT, thiazide-senstitive Na-Cl co-transporter; ENaC, amiloride-sensitive epithelial sodium channel; TRPM6, transient receptor potential cation channel, subfamily M, member 6; WNK, with no lysine (K); V2R, arginine vasopressin receptor 2; MR, mineralocorticoid receptor; RTA, renal tubular acidosis; CA (II), carbonic anhydrase II; AE1, anion exchanger 1; NBC1; sodium-bicarbonate co-transporter; rBAT, renal basic amino acid transport glycoprotein; AT1, amino acid transporter; CLC-5, chloride channel 5; AA, amino acids.

Bartter's Syndrome and Gitelman's Syndrome

GENETICS AND PATHOGENESIS

Bartter's syndrome may result from mutations affecting any of five ion transport proteins in the TAL. The proteins affected include the apical loop diuretic–sensitive sodium-potassium-chloride co-transporter NKCC2 (type 1), the apical potassium channel ROMK (type 2), and the basolateral chloride channel ClC-Kb (type 3). Bartter's type 4 results from mutations in barttin, an essential subunit of ClC-Ka and ClC-Kb that enables transport of the chloride channels to the cell surface. Barttin is also expressed in the inner ear; this accounts for the deafness invariably associated with Bartter's type 4. A Bartter-like phenotype (type 5) with associated hypocalcemia has been described in patients with autosomal dominant gain-of-function mutations in the extracellular calcium-sensing receptor (CaSR). Unregulated activation of this G protein–coupled receptor inhibits sodium reabsorption in the TAL. The TAL transporters function in an integrated manner to maintain both the electrical potential difference and the sodium gradient between the lumen and the cell (Fig. 284-3). Loss of the lumen-positive electrical transport potential that normally drives the paracellular reabsorption of sodium, calcium, and magnesium causes NaCl wasting, hypercalciuria, and mild hypomagnesemia. As expected, the clinical syndrome mimics the effects of chronic ingestion of a loop diuretic.

Gitelman's syndrome is due to mutations in the thiazide-sensitive Na-Cl co-transporter, NCCT, in the DCT. Defects in NCCT in Gitelman's syndrome impair sodium and chloride reabsorption in the DCT (Fig. 284-3) and thus resemble the effects of thiazide diuretics. It remains unclear how this defect leads to severe magnesium wasting.

In both Bartter's and Gitelman's syndromes, hypovolemia from impaired sodium and chloride reabsorption in either the TAL or the DCT activates the renin-angiotensin-aldosterone system (RAAS). The consequent hyperaldosteronism, together with increased distal flow and sodium delivery, stimulates increased sodium reabsorption in the collecting tubules via the epithelial sodium channel (ENaC). This promotes increased potassium and hydrogen ion secretion, causing hypokalemia and metabolic alkalosis. Additionally, in Bartter's syndrome, RAAS activation causes increased levels of cyclooxygenase 2 (COX-2) and marked overproduction of renal prostaglandins (PGE2), and this exacerbates the polyuria and electrolyte abnormalities.

CLINICAL FEATURES

Bartter's Syndrome

Bartter's syndrome is a rare disease that most often presents in the neonatal period or early childhood with polyuria, polydipsia,

salt craving, and growth retardation. Blood pressure is normal or low. Metabolic abnormalities include hypokalemia, hypochloremic metabolic alkalosis, decreased urinary concentrating and diluting ability, hypercalciuria with nephrocalcinosis, mild hypomagnesemia, and increased urinary prostaglandin excretion. Hyperprostaglandin E syndrome is a particularly severe form of Bartter's syndrome in which neonates present with pronounced volume depletion and failure to thrive, as well as fever, vomiting, and diarrhea from PGE2 overproduction. In the antenatal period, fetal polyuria may cause maternal polyhydramnios and premature labor. Sensorineural deafness occurs in patients with barttin gene mutations (type 4). Patients with severe Bartter's syndrome who survive early childhood may develop chronic kidney disease from nephrocalcinosis or from tubular atrophy and interstitial fibrosis from severe persistent hypokalemia. Patients with Bartter's syndrome type 3 have a phenotype intermediate between those of Bartter's and Gitelman's syndromes, consistent with mutation of the ClC-Kb chloride channel in both the TAL and the DCT with preservation of the ClC-Ka chloride channel in the TAL. This disease occurs predominantly in African-American patients and resembles most closely the classic syndrome described by Bartter. Onset is generally later in childhood, patients have mild or no nephrocalcinosis, and prostaglandin excretion is normal.

Gitelman's Syndrome

Gitelman's syndrome is more common than Bartter's syndrome and has a generally milder clinical course with a later age of presentation. It is characterized by prominent neuromuscular symptoms and signs, including fatigue, weakness, carpopedal spasm, cramps, and tetany.

DIAGNOSIS

Hypokalemia and hypochloremic metabolic alkalosis without hypertension are more often due to surreptitious vomiting or diuretic abuse than to Bartter's or Gitelman's syndrome. In contrast to Bartter's and Gitelman's syndromes, urinary chloride levels are very low in patients with surreptitious vomiting. Diuretic abuse can be diagnosed by screening the urine for the offending agents. Gitelman's syndrome is distinguished from most forms of Bartter's syndrome by the presence of severe hypomagnesemia and hypocalciuria.

TREATMENT: BARTTER'S SYNDROME AND GITELMAN'S SYNDROME

Both conditions require lifelong therapy with potassium and magnesium supplements and liberal salt intake. High doses of spironolactone or amiloride treat the hypokalemia, alkalosis, and magnesium wasting. Nonsteroidal anti-inflammatory drugs (NSAIDs) reduce the polyuria and salt wasting in Bartter's syndrome but are ineffective in Gitelman's syndrome. They may be lifesaving in hyperprostaglandin E syndrome and can be given in the form of a COX-2 inhibitor to avoid the gastrointestinal side effects of long-term high-dose NSAIDs. In Gitelman's syndrome, magnesium repletion is essential to correct the hypokalemia and control muscle weakness, tetany, and metabolic alkalosis; however, it may prove difficult in patients wasting large amounts of magnesium.

Pseudohypoaldosteronism Type 1

Patients with type 1 pseudohypoaldosteronism present with severe renal salt wasting and hyperkalemia. Although these findings resemble mineralocorticoid deficiency, plasma renin activity and aldosterone levels are elevated. Defective salt handling is the result of autosomal recessive loss-of-function mutations of the α, β, or γ subunit of the ENaC or autosomal dominant mutations of one allele of the mineralocorticoid receptor (Table 284-2 and Fig. 284-3). The autosomal recessive form is a multisystem disorder with a severe phenotype, often manifesting in the neonatal period with renal salt wasting, vomiting, hyponatremia, hyperkalemia, acidosis, and failure to thrive. Impaired channel activity in the skin and lungs can produce excess sodium and chloride loss in sweat, excess fluid in the airways, and a propensity for lower respiratory tract infections that mimic cystic fibrosis. In contrast, the autosomal dominant form has a more benign course that is limited mainly to renal salt wasting and hyperkalemia. Aggressive salt replacement and management of hyperkalemia can lead to survival into adulthood, and symptoms may become less severe with time, especially in the dominant form. In the latter, high-dose fludrocortisone or carbenoxolone provides additional benefit by increasing mineralocorticoid activity and partly restoring the functional defect in the mutant receptor.

Magnesium Wasting Disorders

In addition to Gitelman's syndrome, several hereditary disorders cause urinary magnesium wasting (Table 284-2 and Fig. 284-3). These disorders include autosomal recessive familial hypomagnesemia with hypercalciuria and nephrocalcinosis (FHHNC), autosomal recessive hypomagnesemia with secondary hypocalcemia (HSH), autosomal dominant hypomagnesemia, autosomal dominant hypoparathyroidism, and isolated autosomal recessive hypomagnesemia (Chap. 353). Common clinical features are the early onset of spasms, tetany, and seizures as well as associated or secondary disturbances in calcium homeostasis.

FAMILIAL HYPOMAGNESEMIA WITH HYPERCALCIURIA AND NEPHROCALCINOSIS

FHHNC is the first example of a disorder attributable to a defective protein involved in paracellular ion transport. *CLDCN16* encodes claudin 16 (previously known as paracellin-1), a member of the claudin family of proteins that are involved in tight junction formation. Claudin 16 is expressed in the TAL of Henle's loop and the DCT. Claudin 16 is thought to be an essential component of the paracellular pathway for Mg, and Ca to a lesser extent, reabsorption in the TAL. Clinical manifestations begin in infancy and include hypomagnesemia that is refractory to oral supplementation, hypercalciuria, and nephrocalcinosis. Recurrent urinary tract infections and nephrolithiasis have also been observed. Patients with mutations of claudin 19 have a similar phenotype but also manifest ocular defect**s,** including corneal calcifications and chorioretinitis.

HYPOMAGNESEMIA WITH SECONDARY HYPOCALCEMIA

Hypomagnesemia in HSH results from a defect in the TRPM6 channel, a member of the transient receptor potential (TRP) family of cation transport channels. TRPM6 is expressed in intestinal epithelia and the DCT and is thought to mediate transepithelial magnesium transport. Symptoms are attributable to hypomagnesemia with secondary impairment of parathyroid function and hypocalcemia. Seizures and muscle spasms occur in infancy, and restoration of magnesium and calcium levels requires high doses of oral magnesium supplementation.

OTHER HEREDITARY HYPOMAGNESEMIC DISORDERS

Mutations of the sodium-potassium-ATPase γ subunit can cause an autosomal dominant hypomagnesemia. Activating mutations of the CaSR in autosomal dominant hypoparathyroidism primarily manifest as hypocalcemia, but hypomagnesemia has been reported in 50% of these patients. Mutations in epidermal growth factor (EGF) result in isolated autosomal recessive hypomagnesemia due to impaired activation of the EGF receptor and consequent failure to activate TRPM6.

HEREDITARY TUBULAR DISORDERS CAUSING HYPERTENSION DUE TO SALT RETENTION

Liddle's Syndrome

Liddle's syndrome mimics a state of aldosterone excess by the presence of early and severe hypertension, often accompanied by hypokalemia and metabolic alkalosis, but plasma aldosterone and renin levels are low. This disorder is due to unregulated sodium reabsorption by an overactive ENaC in the cortical collecting duct (Fig. 284-3). Deletional mutations of the intracellular domain of the β or γ subunit of ENaC (Table 284-2) prevent binding of the ubiquitin ligase Nedd4-2 that normally targets the

channel for proteasomal degradation. This results in an inability to downregulate the number of channels despite a high intracellular sodium concentration. Increased potassium and hydrogen ion secretion follow the lumen-negative electrical potential that results from chloride-independent sodium reabsorption. Amiloride or triamterene blocks ENaC and, combined with salt restriction, provides effective therapy for the hypertension and hypokalemia.

Pseudohypoaldosteronism Type II (Familial Hyperkalemic Hypertension; Gordon's Syndrome)

Pseudohypoaldosteronism type II is a rare autosomal dominant disease that manifests in adolescence or early adulthood with thiazide-responsive low-renin hypertension, hyperkalemia, and metabolic acidosis with normal renal function. Mutations have been identified in the WNK kinases 1 and 4 that lead to increased activity of the thiazide-sensitive sodium chloride channel, NCCT. This causes hypertension from enhanced salt reabsorption in the DCT and impaired distal secretion of potassium and hydrogen ion, all of which can be corrected with thiazide diuretics.

INHERITED DISORDERS OF WATER HANDLING

Hereditary Nephrogenic Diabetes Insipidus

Hereditary nephrogenic diabetes insipidus (NDI) is a rare monogenic disease that usually presents in infancy with severe vasopressin-resistant polyuria, dehydration, failure to thrive, and dilute urine despite the presence of hypernatremia.

GENETICS AND PATHOGENESIS

Vasopressin [antidiuretic hormone (ADH)]-stimulated water reabsorption in the collecting duct is mediated by the type 2 vasopressin receptor (V2R) on the basal surface of principal cells. Activation of the adenylyl cyclase–cAMP pathway phosphorylates vesicle-associated aquaporin-2 (AQP2) water channels and stimulates their insertion into the apical plasma membrane. Water enters the cells from the tubular lumen through AQP2 and exits along an osmotic gradient into the hypertonic medulla and vasa rectae via basal AQP3/4 channels (Fig. 284-3). X-linked mutations of *AVPR2,* the gene that encodes V2R, account for about 90% of cases of hereditary NDI, such that the expression of the receptor on the cell surface is impaired. The remaining cases are due to various autosomal recessive or dominant mutations of *AQP2* that cause the water channels to be retained within the cytosol (Table 284-2). The effect of these mutations is an inability to concentrate the urine and conserve water despite high plasma levels of vasopressin. Penetrance is variable in heterozygous female carriers of X-linked NDI, and some have a moderate concentrating defect that may be exacerbated in pregnancy due to placental vasopressinase.

CLINICAL FEATURES

Whereas NDI in adults is most often acquired from lithium therapy, hypercalcemia, and partial chronic urinary obstruction, hereditary NDI typically presents in infancy. Unlike other polyuric syndromes such as Bartter's and Gitelman's syndromes, conservation of electrolytes is normal, and hypernatremia is entirely from the loss of water. Recurrent episodes of dehydration and hypernatremia can lead to seizures and mental retardation. Although renal function is otherwise normal, chronically high urine flow causes dilation of the ureters and bladder and may cause bladder dysfunction and obstructive uropathy.

DIAGNOSIS

The diagnosis in infants and children with hereditary NDI is usually apparent from the family history and clinical presentation. The diagnosis can be confirmed by the presence of high plasma levels of vasopressin in the face of polyuria and hypotonic urine. This may be especially useful in adults with partial NDI to distinguish the polyuric state from central diabetes insipidus or psychogenic polydipsia (Chap. 340). Genetic screening for mutations in *AVPR2* and *AQP2* is available in research centers and can be performed to identify affected infants from families at risk of NDI to begin treatment and avoid dehydration and its consequences.

TREATMENT: HEREDITARY NEPHROGENIC DIABETES INSIPIDUS

Early diagnosis and treatment with abundant water intake have enabled many patients to live to adulthood with normal mental and physical development. Exogenous vasopressin is ineffective, and because these patients can excrete up to 20 L of urine per day, maintaining adequate water intake is challenging. Thiazide diuretics and salt restriction can reduce urine output by inducing a state of mild volume contraction, thereby promoting increased proximal reabsorption of isotonic fluid and inhibiting the delivery of free water to the collecting duct. A combination thiazide-amiloride formulation will avoid thiazide-induced hypokalemia, and indomethacin may further reduce urine output by inhibiting prostaglandin synthesis.

Nephrogenic Syndrome of Inappropriate Antidiuresis

Activating mutations of the V2R result in hyponatremia, inappropriately elevated urinary osmolality, and undetectable arginine vasopressin (AVP) levels in affected males. This syndrome results from missense mutations of *AVPR2* on the X chromosome, which cause constitutive activation of V2R and inappropriate water reabsorption. Heterozygous female carriers may be at risk of hyponatremia when exposed to large volumes of hypotonic fluids.

INHERITED RENAL TUBULAR ACIDOSIS

Non-anion-gap (hyperchloremic) metabolic acidosis from proximal tubular bicarbonate wasting or impaired distal net acid excretion may be a primary (sporadic or inherited) tubular disorder or may be acquired secondary to a variety of conditions (Chap. 47). There are three forms of renal tubular acidosis (RTA). Types 1 and 2 may be acquired or primary, whereas the most common form, type 4 RTA, usually is acquired in association with moderate renal dysfunction and is characterized by hyperkalemia.

Type 1 (Distal) Rta

CLINICAL FEATURES AND DIAGNOSIS

In distal RTA the kidneys are unable to acidify the urine to pH <5.5 in the presence of systemic metabolic acidosis or after acid loading as a result of impaired hydrogen ion secretion or bicarbonate reabsorption in the distal nephron. Other features include hypokalemia, hypocitraturia, hypercalciuria, nephrocalcinosis, and/or nephrolithiasis. Chronic untreated acidosis may cause rickets or osteomalacia. Inheritance of primary dRTA includes autosomal dominant and autosomal recessive forms with a broad spectrum of clinical expression. Autosomal recessive dRTA most often presents in infancy with severe acidosis, failure to thrive, impaired growth, and impaired kidney function from nephrocalcinosis. Many patients with autosomal dominant dRTA, and some with recessive disease, are asymptomatic, and RTA is discovered incidentally in adolescence or adulthood during evaluation for kidney stones. In the absence of systemic acidosis, the diagnosis of incomplete dRTA is suggested by hypocitraturia and hypercalciuria and can be confirmed by failure of the urine to acidify to pH <5.5 after acid loading.

GENETICS AND PATHOPHYSIOLOGY

Primary dRTA may be hereditary or sporadic with autosomal recessive and autosomal dominant forms. Several kindreds with autosomal recessive dRTA have been identified in Southeast Asia and in areas of the world where parental consanguinity is high. The cellular basis for dRTA lies in dysfunction at the level of the α type intercalated cell of the cortical collecting duct (Fig. 284-3). Mutations affecting subunits of the H^+-ATPase proton pump on the luminal surface impair hydrogen ion secretion and

account for most forms of autosomal recessive dRTA and are often associated with early-onset sensorineural hearing loss (Table 284-2). Autosomal dominant dRTA results from mutations involving the chloride-bicarbonate exchanger, AE1, on the basolateral membrane. Anion exchange by the mutant AE1 is normal, but aberrant targeting of AE1 from the basal to the apical plasma membrane is believed to cause bicarbonate loss into the urine instead of reabsorption. Bi-allelic mutations of AE1 may impair transport activity and account for some cases of recessive disease (Table 284-2). A syndrome of osteopetrosis, short stature, and mental retardation, so-called marble-brain disease with dRTA, is due to mutations in carbonic anhydrase II. Urinary potassium wasting and defective urinary concentration are characteristic of dRTA. Calcium is released from bone in the process of buffering of acid and results in hypercalciuria. Enhanced proximal citrate absorption accounts for hypocitraturia and, together with hypercalciuria, predisposes to nephrocalcinosis and formation of calcium phosphate stones.

TREATMENT: TYPE 1 (DISTAL) RTA

Early initiation of alkali replacement at doses equivalent to 1–3 mmol/kg per day of bicarbonate in divided doses will usually correct the acidosis, hypokalemia, and hypocitraturia, maintaining growth and preventing bone disease in early-onset dRTA. Citrate is generally tolerated better than sodium bicarbonate and can be given as the potassium or sodium salt, depending on the severity of hypokalemia. In patients who present later with kidney stones, large fluid intake and sufficient alkali to restore normal acid-base balance correct the hypocitraturia and reduce hypercalciuria, thereby inhibiting the formation of new stones.

Type 2 (Proximal) Rta

Proximal RTA (pRTA) is the result of impaired bicarbonate reabsorption in the proximal tubule, where the bulk of filtered bicarbonate is recovered (Fig. 284-3). It is most often secondary to various autoimmune, drug-induced, infiltrative, or other tubulopathies (Chap. 47) or results from tubular injury from inherited diseases in which endogenous metabolites accumulate and produce tubular injury. Such inherited disorders include Wilson's disease, cystinosis, tyrosinemia, galactosemia, hereditary fructose intolerance, glycogen storage disease type I, and Lowe's syndrome. In this situation pRTA is only one of several abnormalities that constitute Fanconi syndrome. Other features are hyperphosphaturia, hyperuricosuria, hypercalciuria, nonselective aminoaciduria, and glycosuria. In addition to hyperchloremic acidosis, rickets and osteomalacia are the predominant effects of Fanconi syndrome.

A rare infantile form of primary pRTA with isolated proximal tubular bicarbonate wasting is due to homozygous mutations of the proximal tubule basolateral sodium-bicarbonate co-transporter NBC1 (Table 284-2). This co-transporter is the main mechanism by which bicarbonate moves from the proximal tubule cell back into the blood. Other manifestations include short stature and mental retardation. An ocular phenotype that includes bilateral glaucoma, cataracts, and band keratopathy reflects a role of NBC1 in maintaining normal fluid balance in the eye and clarity of the lens.

TREATMENT: TYPE 2 (PROXIMAL) RTA

It is difficult to restore normal acid-base balance in patients with pRTA despite large amounts of alkali. This is the case because they continue to waste bicarbonate (fractional excretion >15%) until the serum level falls below a threshold, usually about 15 –17 mmol/L, at which time bicarbonate is completely reabsorbed distally and the urine is maximally acidified with pH <5.5. When the serum bicarbonate concentration is raised above the threshold with alkali therapy, bicarbonate wasting recurs and causes hypokalemia as potassium is secreted to maintain luminal electroneutrality. Thus, treatment of pRTA requires 5–15 mmol/kg per day of bicarbonate together with supplemental potassium.

OTHER MONOGENIC DISORDERS OF PROXIMAL TUBULAR FUNCTION (FIG. 284-3)

Cystinuria

Cystinuria is an autosomal recessive disorder of cystine and dibasic amino acid (ornithine, arginine, and lysine) transport in the proximal tubule and intestinal epithelial cells. With a prevalence of about 1 in 10,000, it represents one of the more common heritable diseases. Impaired tubular absorption leads to high concentrations of cystine, which is insoluble in the acid environment of the renal tubules. Clinical severity varies from asymptomatic cystine crystalluria in heterozygous carriers to the frequent passage of gravel and cystine stones, ureteral obstruction, recurrent urinary infections, formation of staghorn calculi, and progressive kidney failure in homozygotes. The median onset of nephrolithiasis is 12 years. The disease is due to mutations in one of two genes, *SLC3A1* and *SLC7A9* (Table 284-2). *SLC3A1* encodes rBAT, a high-affinity, sodium-independent transporter for dibasic amino acids. The protein product of *SLC7A9*, $b^{0,+}$AT, is a catalytic subunit that associates with rBAT to form the active transporter. Diagnosis of cystinuria is established by a positive family history, the finding of hexagonal cystine crystals on urinalysis, and 24-h urinary cystine excretion that exceeds 400 mg (normal less than 30 mg/d).

TREATMENT: CYSTINURIA

The mainstay of treatment is hydration to achieve a urine output of 2.5 L/d or more to reduce urine cystine concentration to <300 mg/L, together with urine alkalinization to pH 7.0–7.5 with potassium citrate, and sodium restriction. Cystine is an oxidized dimer that is formed by linking two cysteine residues via a disulfide bond between the -SH groups. Thus, in intractable cases, thiol derivatives such as penicillamine, tiopronin, and captopril may be added as chelation therapy to dissociate the cystine molecule into more soluble disulfide compounds. Various stone removal and urinary drainage procedures are often required.

Hartnup Disease

Hartnup disease is an autosomal recessive condition caused by a defect in intestinal and renal transport of neutral amino acids. The major clinical manifestations are cerebellar ataxia and a pellagra-like skin rash. Other than aminoaciduria, the kidneys are unaffected. The defective gene, *SLC6A19,* encodes a sodium-dependent and chloride-independent neutral amino acid transporter (B^{0}AT1), which is expressed predominately in the small intestine and proximal tubule of the kidney (Table 284-2). Amino acids such as tryptophan that are retained in the intestinal lumen are converted to indole compounds that are toxic to the CNS. Abnormal tryptophan metabolism also leads to a niacin deficiency that accounts for the skin manifestations. Symptoms are aggravated by a protein-deficient diet and are alleviated with a high-protein diet and nicotinamide supplements.

Dent's Disease

Dent's disease and X-linked recessive nephrolithiasis are unusual forms of Fanconi syndrome due to X-linked mutations of the gene encoding CLC-5, a voltage-gated chloride channel (Table 284-2). The disorders are characterized by childhood onset of low-molecular-weight proteinuria, hypercalciuria, nephrocalcinosis, and nephrolithiasis. Rickets or osteomalacia occurs in 25% of patients, and progressive renal failure from interstitial fibrosis, tubular atrophy, and glomerulosclerosis commonly develops in adulthood. CLC-5 serves to maintain the electrical gradient and acid environment established in proximal tubular cell endosomes by proton-ATPase, which is necessary for the degradation of low-molecular-weight proteins normally filtered by the glomerulus. Defects in CLC-5 appear to disrupt this process and lead to tubular cell dysfunction.

TREATMENT: DENT'S DISEASE

Treatment is directed at controlling hypercalciuria by dietary salt restriction and thiazide diuretics, which favor calcium reabsorption. Restriction of dietary calcium is not recommended.

Cystinosis

Cystinosis is a rare multisystem autosomal recessive disease caused by mutations of cystinosin, a hydrogen ion–driven transporter responsible for exporting cystine from lysosomes. Accumulation of insoluble cystine leads to crystal formation in proximal tubular cells and other organs. Cystinosis occurs in infantile (nephropathic), adolescent, and adult forms. The nephropathic form is the most common, with clinical signs developing between 3 and 6 months of age, including Fanconi syndrome, salt and water wasting, growth retardation, rickets, vomiting, constipation, and unexplained fever. End-stage renal disease occurs by age 10 in the infantile form of the disease but after age 15 in the intermediate form. Extrarenal manifestations result from cystine accumulation in organs and include photophobia and blindness, muscular weakness from carnitine deficiency, hepatomegaly, hypothyroidism, delayed pubertal development, and late-onset neurologic disease. The adult form of cystinosis is largely asymptomatic except for photophobia. The diagnosis is made by measuring elevated cystine content of peripheral blood leukocytes.

TREATMENT: CYSTINOSIS

Treatment includes replacement of fluid and electrolyte losses related to Fanconi syndrome and polyuria. l-Carnitine supplementation is recommended to achieve normal plasma carnitine levels. Cysteamine provides a direct treatment of the disease that converts cystine to cysteine, which can exit the lysosome. Cysteamine should be started promptly upon the diagnosis of cystinosis as it preserves kidney function, prevents hypothyroidism, and improves growth. Kidney transplantation is the treatment of choice for ESRD as cystinosis does not recur in transplanted kidneys, but the extrarenal manifestations persist and may progress.

Renal Glucosuria

Isolated glucosuria in the presence of a normal blood glucose concentration is due to mutations in *SLC5A2*, the gene that encodes the high-capacity sodium-glucose co-transporter SGLT2 in the proximal renal tubule (Table 284-2). Subjects with this disorder are usually asymptomatic and do not have other features of proximal tubular dysfunction. Depending on the severity of the defect, the tubular maximum for glucose reabsorption may fall well within normal blood glucose levels and lead to >50 g/d of glucosuria. Such patients may have polyuria from osmotic diuresis.

Renal Phosphate Wasting

Renal phosphate wasting resulting in hypophosphatemia and rickets or osteomalacia may be part of a generalized disorder of proximal tubular function, as in Fanconi syndrome, or an isolated phenomenon. Isolated phosphaturia is most often due to the inhibition of renal tubular phosphate reabsorption by one or another phosphaturic hormone, with FGF-23 playing a major role (Chap. 352). An exception is hereditary hypophosphatemic rickets with hypercalciuria (HHRH), an autosomal recessive disorder due to mutations in *SLC34A3*, the gene encoding the proximal tubule sodium-phosphate co-transporter, NPT-2c (Table 284-2). Defective phosphate reabsorption causes renal phosphate wasting and stunted growth from rickets. The low serum phosphorus levels stimulate 1-hydroxylation of vitamin D, which increases intestinal calcium absorption, suppresses parathyroid hormone (PTH) secretion, and results in hypercalciuria. High levels of 1,25-dihydroxyvitamin D help distinguish HHRH from hormonal causes of hyperphosphaturia (Chap. 352). Treatment is directed at phosphate repletion.

Vitamin D–Dependent Rickets

Vitamin D–dependent rickets exists in two forms that manifest with hypocalcemia, hypophosphatemia, elevated PTH levels, and the skeletal abnormalities of rickets and osteomalacia. Tetany may be present in severe cases. Vitamin D–dependent rickets type I is an autosomal recessive disease that results from mutations in *CYP27B1*, the gene that encodes 25(OH)D3-1α-hydroxylase, an enzyme in the proximal tubule that catalyzes the hydroxylation and activation of $25(OH)D_3$ into $1,25(OH)_2D_3$ (Table 284-2). It can be treated with physiologic replacement doses of $1,25(OH)_2D_3$. In contrast, autosomal recessive vitamin D–dependent rickets type II is due to end organ resistance to $1,25(OH)_2D_3$ as a result of mutations in the vitamin D receptor (Chap. 353).

FURTHER READINGS

Camargo SM et al: Aminoacidurias: Clinical and molecular aspects. Kidney Int 73:918, 2008[PMID: 18200002]

Crino PB et al: The tuberous sclerosis complex. N Engl J Med 355:1345, 2006[PMID: 17005952]

Devonald MA, Karet FE: Renal epithelial traffic jams and one-way streets. J Am Soc Nephrol 15:1370, 2004[PMID: 15153548]

Grantham JJ: Autosomal dominant polycystic kidney disease. N Engl J Med 359:1477, 2008[PMID: 18832246]

Guay-Woodford L: Other cystic kidney diseases, in *Comprehensive Clinical Nephrology*, 2nd ed, RJ Johnson, J Feehally (eds). New York, Saunders, 2003

Hart TC et al: Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. J Med Genet 39:882, 2002[PMID: 12471200]

Hildebrandt F et al: Nephronophthisis: Disease mechanisms of a ciliopathy. J Am Soc Nephrol 11:1937, 2000; 20:23, 2009[PMID: 19118152]

Jeck N et al: Salt handling in the distal nephron: Lessons learned from inherited human disorders. Am J Physiol Regul Integr Comp Physiol 288:R782, 2005

Knoers Nvam: Inherited forms of renal hypomagnesemia: An update. Pediatr Nephrol 24:697, 2009[PMID: 18818955]

Laing CM et al: Renal tubular acidosis: Developments in our understanding of the molecular basis. Int J Biochem Cell Biol 37:1151, 2005[PMID: 15778079]

Sands JM, Bichet DG: Nephrogenic diabetes insipidus. Ann Intern Med 144:186, 2006[PMID: 16461963]

Saunier S et al: Nephronophthisis. Curr Opin Genet Dev 15:324, 2005[PMID: 15917209]

Siroky BJ: Renal involvement in tuberous sclerosis complex and von Hippel-Lindau disease: Shared disease mechanisms? Nat Clin Pract Nephrol 5:143, 2009[PMID: 19240728]

Torres VE, Harris PC: Mechanisms of disease: Autosomal dominant and recessive polycystic kidney diseases. Nat Clin Pract

Nephrol 2:40, 2006[PMID: 16932388]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TUBULOINTERSTITIAL DISEASES OF THE KIDNEY: INTRODUCTION

Inflammation or fibrosis of the renal interstitium and atrophy of the tubular compartment are common consequences of diseases that target the glomeruli or vasculature. Distinct from these secondary phenomena, however, are a group of disorders that primarily affect the tubules and interstitium, with relative sparing of the glomeruli and renal vessels. Such disorders are conveniently divided into acute and chronic tubulointerstitial nephritis (TIN) (Table 285-1).

Table 285-1 Classification of the Causes of Tubulointerstitial Diseases of the Kidney

Acute Tubulointerstitial Disorders
Acute Interstitial Nephritis
Therapeutic agents
• Antibiotics (β-lactams, sulfonamides, quinolones, vancomycin, erythromycin, minocycline, rifampin, ethambutol, acyclovir)
• Nonsteroidal anti-inflammatory drugs, COX-2 inhibitors
• Diuretics (rarely thiazides, loop diuretics, triamterene)
• Anticonvulsants (phenytoin, valproate, carbamazepine, phenobarbital)
• Miscellaneous (proton pump inhibitors, H_2 blockers, captopril, mesalazine, indinavir, allopurinol)
Infection
• Bacteria (*Streptococcus, Staphylococcus, Legionella, Salmonella, Brucella, Yersinia, Corynebacterium diphtheriae*)
• Viruses (EBV, CMV, hantavirus, polyomavirus, HIV)
• Miscellaneous (*Leptospira, Rickettsia, Mycoplasma*)
Autoimmune
• Tubulointerstitial nephritis with uveitis (TINU)
• Sjögren's syndrome
• Systemic lupus erythematosus
• Granulomatous interstitial nephritis
• IgG4-related systemic disease
• Idiopathic autoimmune interstitial nephritis
Acute obstructive disorders
• Light chain cast nephropathy ("myeloma kidney")
• Acute phosphate nephropathy
• Acute urate nephropathy
Chronic Tubulointerstitial Disorders
• Vesicoureteral reflux/reflux nephropathy
• Sickle cell disease
• Chronic exposure to toxins or therapeutic agents
• Analgesics, especially those containing phenacetin
• Lithium
• Heavy metals (lead, cadmium)
• Aristolochic acid (Chinese herbal and Balkan endemic nephropathies)
• Calcineurin inhibitors (cyclosporine, tacrolimus)
Metabolic Disturbances
• Hypercalcemia and/or nephrocalcinosis
• Hyperuricemia

• Prolonged hypokalemia
• Hyperoxaluria
• Cystinosis (see Chap. 284)
Cystic and Hereditary Disorders (see Chap. 284)
• Polycystic kidney disease
• Nephronophthisis
• Adult medullary cystic disease
• Medullary sponge kidney
Miscellaneous
• Aging
• Chronic glomerulonephritis
• Chronic urinary tract obstruction
• Ischemia and vascular disease
• Radiation nephritis (rare)

Abbreviations: CMV, cytomegalovirus; COX, cyclooxygenase; EBV, Epstein-Barr virus.

Acute TIN most often presents with acute renal failure (Chap. 279). The acute nature of this group of disorders may be caused by aggressive inflammatory infiltrates that lead to tissue edema, tubular cell injury, and compromised tubular flow, or by frank obstruction of the tubules with casts, cellular debris, or crystals. There is sometimes flank pain due to distention of the renal capsule. Urinary sediment is often active with leukocytes and cellular casts, but depends on the exact nature of the disorder in question.

The clinical features of chronic TIN are more indolent and may manifest with disorders of tubular function, including polyuria from impaired concentrating ability (nephrogenic diabetes insipidus), defective proximal tubular reabsorption leading to features of Fanconi syndrome [glycosuria, phosphaturia, aminoaciduria, hypokalemia- and type II renal tubular acidosis (RTA) from bicarbonaturia], or non-anion-gap metabolic acidosis and hyperkalemia (type IV RTA) due to impaired ammoniagenesis, as well as progressive azotemia [rising creatinine and blood urea nitrogen (BUN)]. There is often modest proteinuria (rarely >2 g/d) attributable to decreased tubular reabsorption of filtered proteins; however, nephrotic-range albuminuria may occur in some conditions due to the development of secondary focal segmental glomerulosclerosis (FSGS). Renal ultrasonography may reveal changes of "medical renal disease," such as increased echogenicity of the renal parenchyma with loss of corticomedullary differentiation, prominence of the renal pyramids, and cortical scarring in some conditions. The predominant pathology in chronic TIN is interstitial fibrosis with patchy mononuclear cell infiltration and widespread tubular atrophy, luminal dilation, and thickening of tubular basement membranes. Because of the nonspecific nature of the histopathology, biopsy specimens rarely provide a specific diagnosis. Thus, diagnosis relies on careful analysis of history, drug or toxin exposure, associated symptoms, and imaging studies.

ACUTE INTERSTITIAL NEPHRITIS

In 1897 Councilman reported on eight cases of acute interstitial nephritis (AIN) in the Medical and Surgical Reports of the Boston City Hospital; three as a postinfectious complication of scarlet fever and two from diphtheria. Later, he described the lesion as "an acute inflammation of the kidney characterized by cellular and fluid exudation in the interstitial tissue, accompanied by, but not dependant on, degeneration of the epithelium; the exudation is not purulent in character, and the lesions may be both diffuse and focal." Today AIN is far more often encountered as an allergic reaction to a drug (Table 285-1). Immune-mediated AIN may also occur as part of a known autoimmune syndrome, but in some cases there is no identifiable cause despite features suggestive of an immunological etiology (Table 285-1).

Allergic Interstitial Nephritis

Although biopsy-proven AIN accounts for no more than ~15% of cases of unexplained acute renal failure, this is likely a substantial underestimate of the true incidence. This is because potentially offending medications are more often identified and empirically discontinued in a patient noted to have a rising serum creatinine, without the benefit of a renal biopsy to establish the diagnosis of AIN.

CLINICAL FEATURES

The classic presentation of AIN, namely, fever, rash, peripheral eosinophilia, and oliguric renal failure occurring after 7–10 days of treatment with methicillin or another β-lactam antibiotic is the exception rather than the rule. More often, patients are found incidentally to have a rising serum creatinine or present with symptoms attributable to acute renal failure (Chap. 279). Atypical reactions can occur; most notably nonsteroidal anti-inflammatory drug (NSAID)-induced AIN, in which fever, rash and

eosinophilia are rare, but acute renal failure with heavy proteinuria is common. A particularly severe and rapid-onset AIN may occur upon reintroduction of rifampin after a drug-free period. More insidious reactions to the agents listed in Table 285-1 may lead to progressive tubulointerstitial damage. Examples include proton pump inhibitors and, rarely, sulfonamide and 5-aminosalicylate (mesalazine and sulfasalazine) derivatives and antiretrovirals.

DIAGNOSIS

Finding otherwise unexplained renal failure with or without oliguria and exposure to a potentially offending agent usually points to the diagnosis. Peripheral blood eosinophilia adds supporting evidence but is present in only a minority of patients. Urinalysis reveals pyuria with white blood cell casts and hematuria. Urinary eosinophils are neither sensitive nor specific for AIN; therefore, testing is not recommended. Renal biopsy is generally not required for diagnosis but reveals extensive interstitial and tubular infiltration of leukocytes, including eosinophils.

TREATMENT: ALLERGIC INTERSTITIAL NEPHRITIS

Discontinuation of the offending agent often leads to reversal of the renal injury. However, depending on the duration of exposure and degree of tubular atrophy and interstitial fibrosis that has occurred, the renal damage may not be completely reversible. Glucocorticoid therapy may accelerate renal recovery, but does not appear to impact long-term renal survival. It is best reserved for those cases with severe renal failure in which dialysis is imminent or if renal function continues to deteriorate despite stopping the offending drug (Fig. 285-1 and Table 285-2).

Table 285-2 Indications for Corticosteroids and Immunosuppressives in Interstitial Nephritis

Absolute Indications
• Sjögren's syndrome
• Sarcoidosis
• SLE interstitial nephritis
• Adults with TINU
• Idiopathic and other granulomatous interstitial nephritis
Relative Indications
• Drug-induced or idiopathic AIN with: Rapid progression of renal failure emsp; Diffuse infiltrates on biopsy emsp; Impending need for dialysis emsp; Delayed recovery • Children with TINU
• Postinfectious AIN with delayed recovery (?)

Abbreviations: AIN, acute interstitial nephritis; SLE, systemic lupus erythematosus; TINU, tubulointerstitial nephritis with uveitis.

Source: Modified from S Reddy, DJ Salant: Ren Fail 20:829, 1998.

Figure 285-1

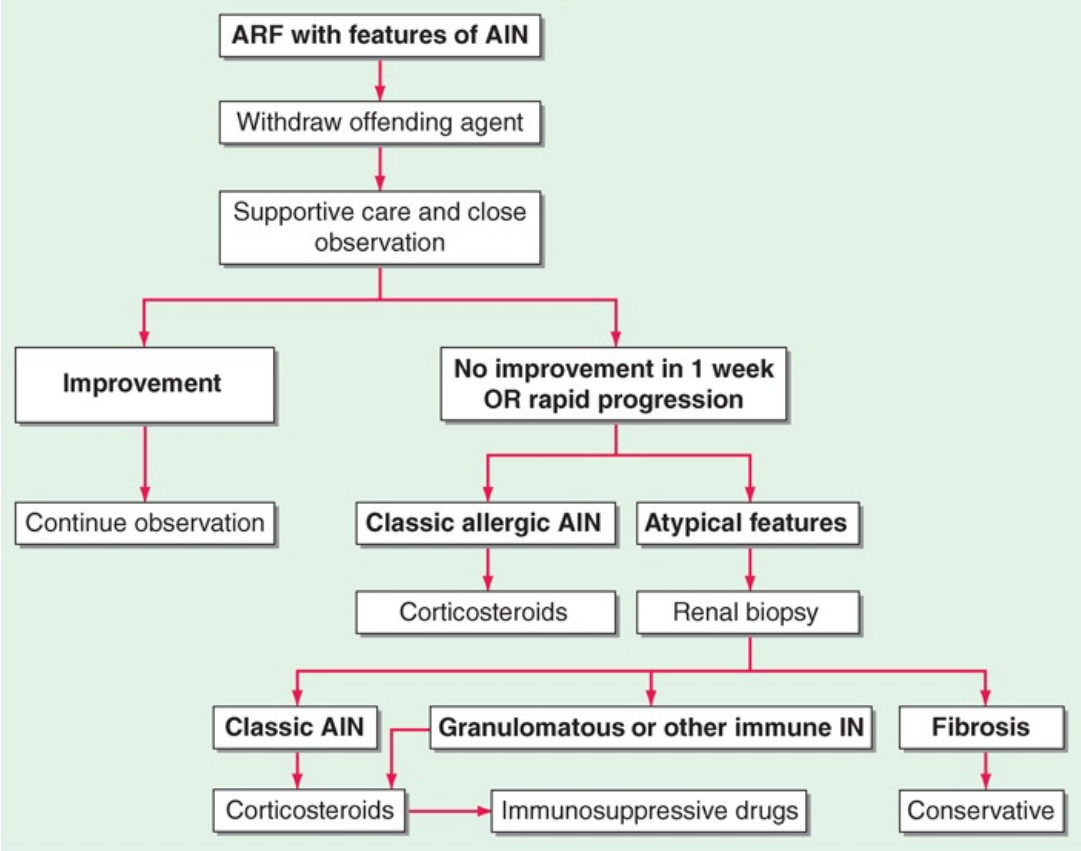

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the treatment of allergic and other immune-mediated acute interstitial nephritis (AIN). ARF, acute renal failure. See text for immunosuppressive drugs used for refractory or relapsing AIN. *(Modified from S Reddy, DJ Salant: Ren Fail 20:829, 1998.)*

Sjögren's Syndrome

Sjögren's syndrome is a systemic autoimmune disorder that primarily targets the exocrine glands, especially the lacrimal and salivary glands, and thus results in symptoms, such as dry eyes and mouth, that constitute the "sicca syndrome" (Chap. 324). Tubulointerstitial nephritis with a predominant lymphocytic infiltrate is the most common renal manifestation of Sjögren's syndrome and can be associated with distal RTA, nephrogenic diabetes insipidus, and moderate renal failure. Diagnosis is strongly supported by positive serologic testing for anti-Ro (SS-A) and anti-La (SS-B) antibodies. A large proportion of patients with Sjögren's syndrome also have polyclonal hypergammaglobulinemia. Treatment is initially with glucocorticoids, although patients may require maintenance therapy with azathioprine or mycophenolate mofetil to prevent relapse (Fig. 285-1 and Table 285-2).

Tubulointerstitial Nephritis with Uveitis (Tinu)

TINU is a systemic autoimmune disease of unknown etiology. It accounts for fewer than 5% of all cases of AIN, affects females three times more often than males, and has a median age of onset of 15 years. Its hallmark feature, in addition to a lymphocyte-predominant interstitial nephritis (Fig. 285-2), is a painful anterior uveitis, often bilateral and accompanied by blurred vision and photophobia. Diagnosis is often confounded by the fact that the ocular symptoms precede or accompany the renal disease in only one-third of cases. Additional extrarenal features include fever, anorexia, weight loss, abdominal pain, and arthralgia. The presence of such symptoms as well as elevated creatinine, sterile pyuria, mild proteinuria, features of the Fanconi syndrome, and elevated erythrocyte sedimentation rate should raise suspicion for this disorder. Serologies suggestive of the more common autoimmune diseases are usually negative, and TINU is often a diagnosis of exclusion after other causes of uveitis and renal disease, such as Sjögren's syndrome, Behöet's disease, sarcoidosis, and systemic lupus erythematosus, have been considered. Clinical symptoms are typically self-limited in children, but are more apt to follow a relapsing course in adults. The renal and ocular manifestations generally respond well to oral glucocorticoids, although maintenance therapy with agents such as methotrexate, azathioprine, or mycophenolate may be necessary to prevent relapses (Fig. 285-1 and Table 285-2).

Figure 285-2

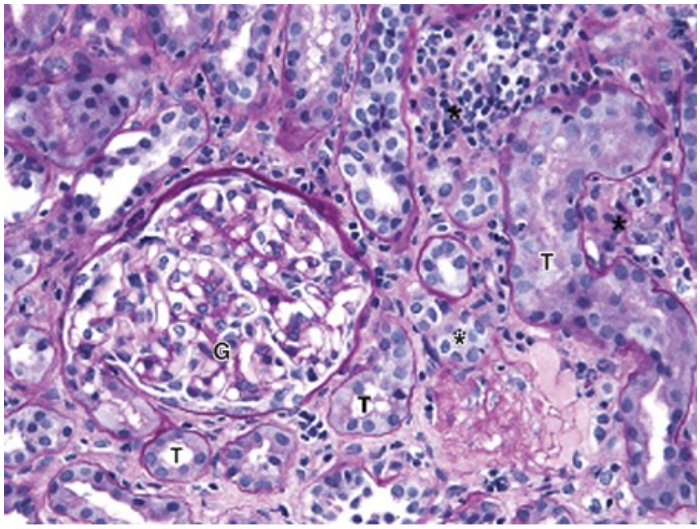

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute interstitial nephritis (AIN) in a patient who presented with acute iritis, low-grade fever, erythrocyte sedimentation rate of 103, pyuria and cellular casts on urinalysis, and a newly elevated serum creatinine of 2.4 mg/dL. Both the iritis and AIN improved after intravenous methylprednisolone. This PAS-stained renal biopsy shows a mononuclear cell interstitial infiltrate (asterisks) and edema separating the tubules (T) and a normal glomerulus (G). Some of the tubules contain cellular debris and infiltrating inflammatory cells. The findings in this biopsy are indistinguishable from those that would be seen in a case of drug-induced AIN. PAS, Periodic acid –Schiff.

Systemic Lupus Erythematosus

An interstitial mononuclear cell inflammatory reaction often accompanies the glomerular lesion in most cases of class III or IV lupus nephritis (Chap. 283), and deposits of immune complexes can be identified in tubule basement membranes in about 50%. Occasionally, however, the tubulointerstitial inflammation predominates and may manifest with azotemia and type IV RTA rather than features of glomerulonephritis.

Granulomatous Interstitial Nephritis

Some patients may present with features of AIN but follow a protracted and relapsing course. Renal biopsy in such patients reveals a more chronic inflammatory infiltrate with granulomas and multinucleated giant cells. Most often, no associated disease or cause is found; however, some of these cases may have or subsequently develop the pulmonary, cutaneous, or other systemic manifestations of *sarcoidosis* such as hypercalcemia. Most patients experience some improvement in renal function if treated early with glucocorticoids before the development of significant interstitial fibrosis and tubular atrophy (Table 285-2). Other immunosuppressive agents may be required for those who relapse frequently upon steroid withdrawal (Fig. 285-1). Tuberculosis should be ruled out before starting treatment because this too is a rare cause of granulomatous interstitial nephritis.

Igg4-Related Systemic Disease

A form of AIN characterized by a dense inflammatory infiltrate containing IgG4-expressing plasma cells can occur as a part of a recently described syndrome known as IgG4-related systemic disease. Autoimmune pancreatitis, sclerosing cholangitis, retroperitoneal fibrosis, and a chronic sclerosing sialadenitis (mimicking Sjögren's syndrome) may variably be present as well. Fibrotic lesions that form pseudotumors in the affected organs soon replace the initial inflammatory infiltrates and often lead to biopsy or excision for fear of true malignancy. Although the involvement of IgG4 in the pathogenesis is not understood, glucocorticoids have been successfully used as first-line treatment in this group of disorders, once they are correctly diagnosed.

Idiopathic Ain

Some patients present with typical clinical and histologic features of AIN but have no evidence of drug exposure or clinical or serologic features of an autoimmune disease. The presence in some cases of autoantibodies to a tubular antigen, similar to that identified in rats with an induced form of interstitial nephritis, suggests that an autoimmune response may be involved. Like TINU and granulomatous interstitial nephritis, idiopathic AIN is responsive to glucocorticoid therapy but may follow a relapsing course requiring maintenance treatment with another immunosuppressive agent (Fig. 285-1 and Table 285-2).

Infection-Associated Ain

AIN may also occur as a local inflammatory reaction to microbial infection (Table 285-1) and should be distinguished from acute bacterial pyelonephritis (Chap. 288). Acute bacterial pyelonephritis does not generally cause acute renal failure unless it affects both kidneys or causes septic shock. Presently, infection-associated AIN is most often seen in immunocompromised patients, particularly renal transplant recipients with reactivation of polyomavirus BK (Chaps. 132 and 282).

Crystal Deposition Disorders and Obstructive Tubulopathies

Acute renal failure may occur when crystals of various types are deposited in tubular cells and interstitium or when they obstruct tubules. Oliguric acute renal failure, often accompanied by flank pain from tubular obstruction, may occur in patients treated with sulfadiazine for toxoplasmosis, indinavir for HIV, and intravenous acyclovir for severe herpesvirus infections. Urinalysis reveals "sheaf of wheat" sulfonamide crystals, individual or parallel clusters of needle-shaped indinavir crystals, or red-green birefringement needle-shaped crystals of acyclovir. This adverse effect is generally precipitated by hypovolemia and is reversible with saline volume repletion and drug withdrawal. Distinct from the obstructive disease, a frank AIN from indinavir crystal deposition has also been reported.

Acute tubular obstruction is also the cause of oliguric renal failure in patients with *acute urate nephropathy.* It typically results from severe hyperuricemia from tumor lysis syndrome in patients with lympho- or myeloproliferative disorders treated with cytotoxic agents, but also may occur spontaneously before the treatment has been initiated (Chap. 276). Uric acid crystallization in the tubules and collecting system leads to partial or complete obstruction of the collecting ducts, renal pelvis, or ureter. A dense precipitate of birefringent uric acid crystals is found in the urine, usually in association with microscopic or gross hematuria. Prophylactic allopurinol reduces the risk of uric acid nephropathy but is of no benefit once tumor lysis has occurred. Once oliguria has developed, attempts to increase tubular flow and solubility of uric acid with alkaline diuresis may be of some benefit; however, emergent treatment with hemodialysis or rasburicase, a recombinant urate oxidase, is usually required to rapidly lower uric acid levels and restore renal function.

Calcium oxalate crystal deposition in tubular cells and interstitium may lead to permanent renal dysfunction in patients who survive ethylene glycol intoxication, in patients with enteric hyperoxaluria from ileal resection or small-bowel bypass surgery, and in patients with hereditary hyperoxaluria (Chap. 287). *Acute phosphate nephropathy* is an uncommon but serious complication of oral Phospho-soda used as a laxative or for bowel preparation for colonoscopy. It is the result of calcium phosphate crystal deposition in tubules and interstitium and occurs especially in subjects with underlying renal impairment and hypovolemia. Consequently, Phospho-soda should be avoided in patients with chronic kidney disease.

Light Chain Cast Nephropathy

Patients with multiple myeloma may develop acute renal failure in the setting of hypovolemia, infection, or hypercalcemia or after exposure to NSAIDs or radiographic contrast media. The diagnosis of light chain cast nephropathy (LCCN)—commonly known as *myeloma kidney*—should be considered in patients who fail to recover when the precipitating factor is corrected or in any elderly patient with otherwise unexplained acute renal failure.

In this disorder, filtered monoclonal immunoglobulin light chains (Bence-Jones proteins) form intratubular aggregates with secreted Tamm-Horsfall protein in the distal tubule. Casts, in addition to obstructing the tubular flow in affected nephrons, incite a giant cell or foreign body reaction and can lead to tubular rupture, resulting in interstitial fibrosis (Fig. 285-3). Although LCCN generally occurs in patients with known multiple myeloma and a large plasma cell burden, the disorder should also be considered as a possible diagnosis in patients who have known monoclonal gammopathy even in the absence of frank myeloma. Filtered monoclonal light chains may also cause less pronounced renal manifestations in the absence of obstruction, due to direct toxicity to proximal tubular cells and intracellular crystal formation. This may result in isolated tubular disorders such as RTA or the full Fanconi syndrome.

Figure 285-3

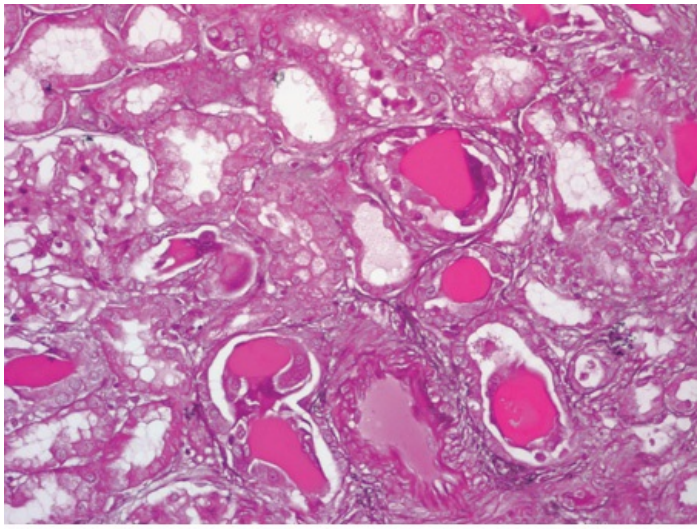

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histologic appearance of myeloma cast nephropathy. A hematoxylin-eosin-stained kidney biopsy shows many atrophic tubules filled with eosinophilic casts (consisting of Bence-Jones protein), which are surrounded by giant cell reactions. *(Courtesy of Dr. Michael N. Koss, University of Southern California Keck School of Medicine; with permission.)*

DIAGNOSIS

Clinical clues to the diagnosis include anemia, bone pain, hypercalcemia, and an abnormally narrow anion gap due to hypoalbuminemia and hypergammaglobulinemia. Urinary dipsticks detect albumin but not immunoglobulin light chains; however, laboratory detection of increased amounts of protein in a spot urine specimen and a negative dipstick result are highly suggestive that the urine contains Bence-Jones protein. Serum and urine should both be sent for protein electrophoresis and for immunofixation for the detection and identification of a potential monoclonal band. A sensitive method is now clinically available to detect urine and serum free light chains.

TREATMENT: LIGHT CHAIN CAST NEPHROPATHY

The goals of treatment are to correct precipitating factors such as hypovolemia and hypercalcemia, discontinue potential nephrotoxic agents, and treat the underlying plasma cell dyscrasia (Chap. 111); plasmapheresis to remove light chains is of questionable value for LCCN.

Lymphomatous Infiltration of the Kidney

Interstitial infiltration by malignant B lymphocytes is a common autopsy finding in patients dying of chronic lymphocytic leukemia and non-Hodgkin's lymphoma; however, this is usually an incidental finding. Rarely, such infiltrates may cause massive enlargement of the kidneys and oliguric acute renal failure. Although high-dose glucocorticoids and subsequent chemotherapy often results in recovery of renal function, the prognosis in such cases is generally poor.

CHRONIC TUBULOINTERSTITIAL DISEASES

Improved occupational and public health measures, together with the banning of over-the-counter phenacetin-containing analgesics, has led to a dramatic decline in the incidence of chronic interstitial nephritis (CIN) from heavy metal—particularly lead and cadmium—exposure and analgesic nephropathy in North America. Today, CIN is most often the result of renal ischemia or secondary to a primary glomerular disease (Chap. 283). Other important forms of CIN are the result of developmental anomalies or inherited diseases such as reflux nephropathy or sickle cell nephropathy and may not be recognized until adolescence or adulthood. Whereas it is impossible to reverse damage that has already occurred, further deterioration may be prevented or at least slowed in such cases by treating glomerular hypertension, a common denominator in the development of secondary FSGS and progressive loss of functioning nephrons. Therefore, awareness and early detection of patients at risk may prevent them from developing end-stage renal disease (ESRD).

Vesicoureteral Reflux and Reflux Nephropathy

Reflux nephropathy is the consequence of vesicoureteral reflux (VUR) or other urologic anomalies in early childhood. It was previously called *chronic pyelonephritis* because it was believed to result from recurrent urinary tract infections (UTIs) in childhood. VUR stems from abnormal retrograde urine flow from the bladder into one or both ureters and kidneys because of mislocated and incompetent ureterovesical valves (Fig. 285-4). Although high-pressure sterile reflux may impair normal growth of the kidneys, when coupled with recurrent UTIs in early childhood, the result is patchy interstitial scarring and tubular atrophy. Loss of functioning nephrons leads to hypertrophy of the remnant

glomeruli and eventual secondary FSGS. Reflux nephropathy often goes unnoticed until early adulthood when chronic kidney disease is detected during routine evaluation or during pregnancy. Affected adults are frequently asymptomatic, but may give a history of prolonged bed-wetting or recurrent UTIs during childhood, and exhibit variable renal insufficiency, hypertension, mild to moderate proteinuria, and unremarkable urine sediment. When both kidneys are affected, the disease often progresses inexorably over several years to end-stage kidney disease, despite the absence of ongoing urinary infections or reflux. A single affected kidney may go undetected, except for the presence of hypertension. Renal ultrasound in adults characteristically shows asymmetric small kidneys with irregular outlines, thinned cortices, and regions of compensatory hypertrophy (Fig. 285-4).

Figure 285-4

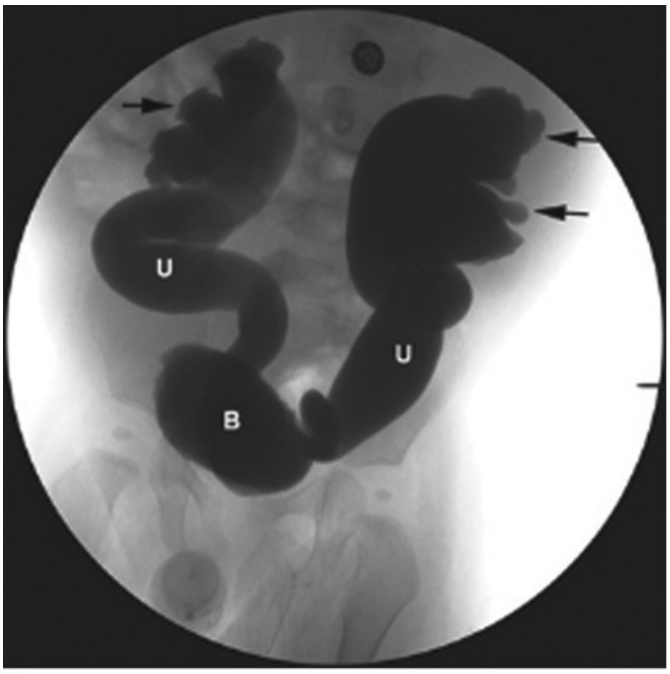

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

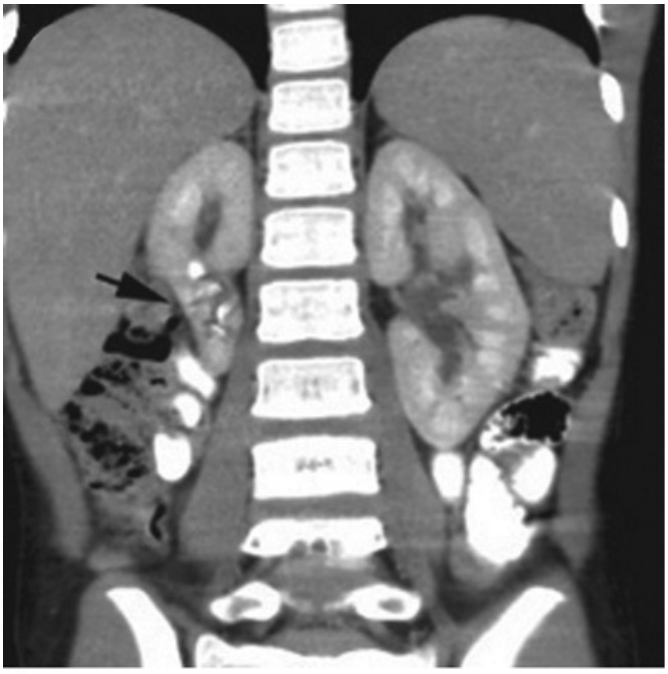

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

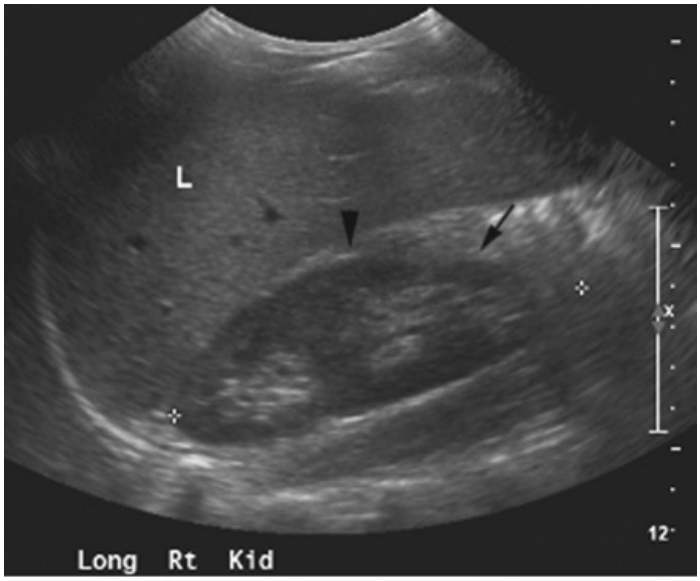

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiographs of vesicoureteral reflux (VUR) and reflux nephropathy.*A*. Voiding cystourethrogram in a 7-month-old baby with bilateral high-grade VUR evidenced by clubbed calyces (arrows) and dilated tortuous ureters (U) entering the bladder (B). ***B***. Abdominal CT scan (coronal plane reconstruction) in a child showing severe scarring of the lower portion of the right kidney (arrow). ***C***. Sonogram of the right kidney showing loss of parenchyma at the lower pole due to scarring (arrow) and hypertrophy of the mid-region (arrowhead). *(Courtesy of Dr. George Gross, University of Maryland Medical Center; with permission.)*

TREATMENT: VESICOURETERAL REFLUX AND REFLUX NEPHROPATHY

Maintenance of a sterile urine in childhood has been shown to limit scarring of the kidneys. Surgical reimplantation of the ureters into the bladder to restore competency is indicated in young children with persistent high-grade reflux but is ineffective and is not indicated in adolescents or adults after scarring has occurred. Aggressive control of blood pressure with an angiotensin-converting enzyme inhibitor (ACEI) or angiotensin receptor blocker (ARB) and other agents is effective in reducing proteinuria and may significantly forestall further deterioration of renal function.

Sickle Cell Nephropathy

The pathogenesis and clinical manifestations of sickle cell nephropathy are described in Chap. 286. Evidence of tubular injury may be evident in childhood and early adolescence in the form of polyuria due to decreased concentrating ability or type IV renal tubular acidosis years before there is significant nephron loss and proteinuria from secondary FSGS. Early recognition of these subtle renal abnormalities or development of microalbuminuria in a child with sickle cell disease may warrant consultation with a nephrologist and/or therapy with low-dose ACEIs. Papillary necrosis may result from ischemia due to sickling of red cells in the relatively hypoxemic and hypertonic medullary vasculature and present with gross hematuria and ureteric obstruction by sloughed ischemic papillae (Table 285-3).

Table 285-3 Major Causes of Papillary Necrosis
Analgesic nephropathy
Sickle cell nephropathy
Diabetes with urinary tract infection
Prolonged NSAID use (rare)

Abbreviation: NSAID, nonsteroidal anti-inflammatory.

Tubulointerstitial Abnormalities Associated with Glomerulonephritis

Primary glomerulopathies are often associated with damage to tubules and interstitium. This may occasionally be due to the same pathologic process affecting the glomerulus and tubulointerstitium, as is the case with immune-complex deposition in lupus nephritis. More often, however, chronic tubulointerstitial changes occur as a secondary consequence of prolonged glomerular dysfunction. Potential mechanisms by which glomerular disease might cause tubulointerstitial injury include proteinuria-mediated damage to the epithelial cells, activation of tubular cells by cytokines and

complement, or reduced peritubular blood flow leading to downstream tubulointerstitial ischemia, especially in the case of glomeruli that are globally obsolescent due to severe glomerulonephritis. It is often difficult to discern the initial cause of injury by renal biopsy in a patient who presents with advanced renal disease in this setting.

Analgesic Nephropathy

Analgesic nephropathy results from the long-term use of compound analgesic preparations containing phenacetin (banned in the United States since 1983), aspirin, and caffeine. In its classic form, analgesic nephropathy is characterized by renal insufficiency, papillary necrosis (Table 285-3) attributable to the presumed concentration of the drug to toxic levels in the inner medulla, and a radiographic constellation of small, scarred kidneys with papillary calcifications best appreciated by computed tomography (Fig. 285-5). Patients may also have polyuria due to impaired concentrating ability and non-anion-gap metabolic acidosis from tubular damage. Shedding of a sloughed necrotic papilla can cause gross hematuria and ureteric colic due to ureteral obstruction. Individuals with end-stage kidney disease as a result of analgesic nephropathy are at increased risk of a urothelial malignancy compared to patients with other causes of renal failure. Recent cohort studies in individuals with normal baseline renal function suggest that the moderate chronic use of current analgesic preparations available in the United States, including acetaminophen and NSAIDs, does not seem to cause the constellation of findings known as analgesic nephropathy, although volume-depleted individuals and those with chronic kidney disease are at higher risk of NSAID-related renal toxicity. Nonetheless, it is recommended that heavy users of acetaminophen and NSAIDs be screened for evidence of renal disease.

Figure 285-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiologic appearance of analgesic nephropathy. A non-contrast CT scan shows an atrophic left kidney with papillary calcifications in a garland pattern. *(Reprinted by permission from Macmillan Publishers, Ltd., MM Elseviers et al., Kidney International 48:1316, 1995.)*

Chinese Herbal Nephropathy and Balkan Nephropathy

Nontraditional (alternative or herbal) medications can also lead to progressive tubulointerstitial disease. In Chinese herbal nephropathy, first described in young women taking Chinese herbal preparations as part of a weight-loss regimen, one of the offending agents has been identified as aristolochic acid, a botanical product and known rodent carcinogen from the plant Aristolochia. This chemical, after prolonged exposure, produces renal interstitial fibrosis with a relative paucity of cellular infiltrates. The urine sediment is bland, with rare leukocytes and only mild proteinuria. Anemia may be disproportionately severe relative to the level of renal dysfunction. Like analgesic nephropathy, Chinese herbal nephropathy has been associated with a long-term increased risk of bladder and ureteral malignancies. There is recent evidence that Balkan endemic nephropathy, a chronic tubulointerstitial nephritis found primarily in towns along the tributaries of the Danube River, may also be linked to aristolochic acid as a result of contamination of local grain preparations. Although other environmental factors for Balkan endemic nephropathy, such as the mycotoxin ochratoxin A or water-soluble hydrocarbons leached from the coal deposits in the area, have been proposed as causative agents, current evidence appears to be strongest for aristolochic acid. It has been proposed that Chinese herbal nephropathy and Balkan endemic nephropathy be collectively known as aristolochic acid nephropathy.

Lithium-Associated Nephropathy

The use of lithium salts for the treatment of manic-depressive illness may have several renal sequelae, the most common of which is nephrogenic diabetes insipidus manifesting as polyuria and polydipsia. Lithium accumulates in principal cells of the collecting duct by entering through the epithelial sodium channel (ENaC), where it inhibits glycogen synthase kinase 3β and downregulates vasopressin-regulated aquaporin water channels. Less frequently, chronic tubulointerstitial nephritis develops after prolonged (greater than 10–20 years) lithium use and is most likely to occur in patients that have experienced repeated episodes of toxic lithium levels. Findings on renal biopsy include interstitial fibrosis and tubular atrophy that are out of proportion to the degree of glomerulosclerosis or vascular disease, a sparse lymphocytic infiltrate, and small cysts or dilation of the distal tubule and collecting duct that are highly characteristic of this disorder. The degree of interstitial fibrosis correlates both with duration and cumulative dose of lithium. Individuals with lithium-associated nephropathy are typically asymptomatic, with minimal proteinuria, few urinary leukocytes, and normal blood pressure. Some patients develop more severe proteinuria due to secondary FSGS, which may contribute to further loss of renal function.

TREATMENT: LITHIUM-ASSOCIATED NEPHROPATHY

Renal function should be followed regularly in patients taking lithium, and caution should be exercised in patients with underlying renal disease. The use of amiloride to inhibit lithium entry via ENaC has been effective to prevent and treat lithium-induced nephrogenic diabetes insipidus, but it is not clear if it will prevent lithium-induced CIN. Once lithium-associated nephropathy is detected, the discontinuation of lithium in attempt to forestall further renal deterioration can be problematic, as lithium is an effective mood stabilizer that is often incompletely substituted by other agents. Furthermore, despite discontinuation of lithium, chronic renal disease in such patients is often irreversible and can slowly progress to end-stage kidney disease. The most prudent approach is to monitor lithium levels frequently and adjust dosing to avoid toxic levels (preferably <1 meq/L). This is especially important as lithium is cleared less effectively as renal function declines. In those cases that develop significant proteinuria, ACEI or ARB treatment should be initiated.

Calciurin-Inhibitor Nephrotoxicity

The calcineurin inhibitor (CNI) immunosuppressive agents cyclosporine and tacrolimus can cause both acute and chronic renal injury. Acute forms can result from vascular causes such as vasoconstriction or the development of thrombotic microangiopathy, or can be due to a toxic tubulopathy. Chronic CNI-induced renal injury is typically seen in solid organ (including heart-lung and liver) transplant recipients and manifests with a slow but irreversible reduction of glomerular filtration rate, with mild proteinuria and arterial hypertension. Hyperkalemia is a relatively common complication and is caused, in part, by tubular resistance to aldosterone. The histologic changes in renal tissue include patchy interstitial fibrosis and tubular atrophy, often in a "striped" pattern. In addition, the intrarenal vasculature often demonstrates hyalinosis, and focal glomerulosclerosis can be present as well. Similar changes may occur in patients receiving CNIs for autoimmune diseases, although the doses are generally lower than those used for organ transplantation. Dose reduction or CNI avoidance appears to mitigate the chronic tubulointerstitial changes, but may increase the risk of rejection and graft loss.

Heavy Metal (Lead) Nephropathy

Heavy metals, such as lead or cadmium, can lead to a chronic tubulointerstitial process after prolonged exposure. The disease entity is no longer commonly diagnosed, because such heavy metal exposure has been greatly reduced due to the known health risks from lead and the consequent removal of lead from most commercial products and fuels. Nonetheless, occupational exposure is possible in workers involved in the manufacture or destruction of batteries, removal of lead paint, or manufacture of alloys and electrical equipment (cadmium) in countries where industrial regulation is less stringent. In addition, ingestion of moonshine whiskey distilled in lead-tainted containers has been one of the more frequent sources of lead exposure.

Early signs of chronic lead intoxication are attributable to proximal tubule dysfunction, particularly hyperuricemia as a result of diminished urate secretion. The triad of "saturnine gout," hypertension, and renal insufficiency should prompt a practitioner to ask specifically about lead exposure. Unfortunately, evaluating lead burden is not as straightforward as ordering a blood test; the preferred methods involve measuring urinary lead after infusion of a chelating agent or by radiographic fluoroscopy of bone. Several recent studies have shown an association between chronic low-level lead exposure and decreased renal function, although either of these two factors may have been the primary event. In those patients who have CIN of unclear origin and an elevated total body lead burden, repeated treatments of lead chelation therapy have been shown to slow the decline in renal function.

METABOLIC DISORDERS

Disorders leading to excessively high or low levels of certain electrolytes and products of metabolism can also lead to chronic kidney disease if untreated.

Chronic Uric Acid Nephropathy

The constellation of pathologic findings that represent *gouty nephropathy* are very uncommon nowadays and are more of historical interest than clinical importance, as gout is typically well managed with allopurinol and other agents. However, there is emerging evidence that hyperuricemia is an independent risk factor for the development of chronic kidney disease, perhaps through endothelial damage. The complex interactions of hyperuricemia, hypertension, and renal failure are still incompletely understood.

Presently, gouty nephropathy is most likely to be encountered in patients with severe tophaceous gout and prolonged hyperuricemia from a hereditary disorder of purine metabolism (Chap. 359). Histologically, the distinctive feature of gouty nephropathy is the presence of crystalline deposits of uric acid and monosodium urate salts in the kidney parenchyma. These deposits not only cause intrarenal obstruction but also incite an inflammatory response, leading to lymphocytic infiltration, foreign-body giant cell reaction, and eventual fibrosis, especially in the medullary and papillary regions of the kidney. Since patients with gout frequently suffer from hypertension and hyperlipidemia, degenerative changes of the renal arterioles may constitute a striking feature of the histologic abnormality, out of proportion to the other morphologic defects. Clinically, gouty nephropathy is an insidious cause of chronic kidney disease. Early in its course, glomerular filtration rate may be near normal, often despite morphologic changes in medullary and cortical interstitium, proteinuria, and diminished urinary concentrating ability. Treatment with allopurinol and urine alkalinization is generally effective in preventing uric acid nephrolithiasis and the consequences of recurrent kidney stones; however, gouty nephropathy may be intractable to such measures. Furthermore, the use of allopurinol in asymptomatic hyperuricemia has not been consistently shown to improve renal function.

Hypercalcemic Nephropathy

(See also Chap. 353) Chronic hypercalcemia, as occurs in primary hyperparathyroidism, sarcoidosis, multiple myeloma, vitamin D intoxication, or metastatic bone disease, can cause tubulointerstitial disease and progressive renal failure. The earliest lesion is a focal degenerative change in renal epithelia, primarily in collecting ducts, distal tubules, and loops of Henle. Tubular cell necrosis leads to nephron obstruction and stasis of intrarenal urine, favoring local precipitation of calcium salts and infection. Dilation and atrophy of tubules eventually occurs, as does interstitial fibrosis, mononuclear leukocyte infiltration, and interstitial calcium deposition (nephrocalcinosis). Calcium deposition may also occur in glomeruli and the walls of renal arterioles.

Clinically, the most striking defect is an inability to maximally concentrate the urine, due to reduced collecting duct responsiveness to AVP and defective transport of sodium and chloride in the loop of Henle. Reductions in both glomerular filtration rate and renal blood flow can occur, both in acute and in prolonged hypercalcemia. Eventually, uncontrolled hypercalcemia leads to severe tubulointerstitial damage and overt renal failure. Abdominal x-rays may demonstrate nephrocalcinosis as well as nephrolithiasis, the latter due to the hypercalciuria that often accompanies hypercalcemia.

Treatment consists of reducing the serum calcium concentration toward normal and correcting the primary abnormality of calcium metabolism (Chap. 353). Renal dysfunction of acute hypercalcemia may be completely reversible. Gradual progressive renal insufficiency related to chronic hypercalcemia, however, may not improve even with correction of the calcium disorder.

Hypokalemic Nephropathy

Patients with prolonged and severe hypokalemia from chronic laxative or diuretic abuse, surreptitious vomiting, or primary aldosteronism may develop a reversible tubular lesion characterized by vacuolar degeneration of proximal and distal tubular cells. Eventually, tubular atrophy and cystic dilation accompanied by interstitial fibrosis may ensue, leading to irreversible chronic kidney disease. Timely correction of the hypokalemia will prevent further progression, but persistent hypokalemia can cause ESRD.

GLOBAL PERSPECTIVE

The causes of acute and chronic interstitial nephritis vary widely across the globe. Analgesic nephropathy continues to be seen in countries where phenacetin-containing compound analgesic preparations are readily available. Adulterants in unregulated herbal and traditional medicaments pose a threat of toxic interstitial nephritis, as exemplified by aristolochic acid contamination of herbal slimming preparations. Contamination of food sources with toxins, such as the recent outbreak of nephrolithiasis and acute renal failure from melamine contamination of infant milk formula, poses a continuing risk. Likewise, Balkan endemic nephropathy appears likely to be the result of aristolochic acid contamination of grain preparations. While industrial exposure to lead and cadmium has largely disappeared as a cause of chronic interstitial nephritis in developed nations, it remains a risk for nephrotoxicity in countries where such exposure is less well controlled. Conversely, the widespread use of proton pump inhibitors for gastroesophageal reflux disease (GERD) and Phospho-soda prior to screening colonoscopy has introduced a new spectrum of drug-induced kidney diseases to wealthier nations.

ACKNOWLEDGMENT

We are grateful to Drs. Alan Yu and Barry Brenner, authors of "Tubulointerstitial Diseases of the Kidney" in the 17th edition of *Harrison's*, for contributions to this chapter.

FURTHER READINGS

Appel GB: The treatment of acute interstitial nephritis: More data at last. Kidney Int 73:905, 2008[PMID: 18379525]

Batuman V: Proximal tubular injury in myeloma. Contrib Nephrol 153:87, 2007[PMID: 17075225]

Beck LH Jr, Salant DJ: Glomerular and tubulointerstitial diseases. Prim Care 35:265, 2008[PMID: 18486716]

Becker GJ, Kincaid-Smith P: Reflux nephropathy: The glomerular lesion and progression of renal failure. Pediatr Nephrol 7:365, 1993[PMID: 8398643]

Brewster UC, Perazella MA: Proton pump inhibitors and the kidney: Critical review. Clin Nephrol 68:65, 2007[PMID: 17722704]

De Broe ME, Elseviers MM: Over-the-counter analgesic use. J Am Soc Nephrol 20:2098, 2009

Geevasinga N et al: Proton pump inhibitors and acute interstitial nephritis. Clin Gastroenterol Hepatol 4:597, 2006[PMID: 16630752]

Grunfeld JP, Rossier BC: Lithium nephrotoxicity revisited. Nat Rev Nephrol 5:270, 2009[PMID: 19384328]

Henrich WL et al: Non-contrast-enhanced computerized tomography and analgesic-related kidney disease: Report of the National Analgesic Nephropathy Study. J Am Soc Nephrol 17:1472, 2006[PMID: 16611714]

Isnard Bagnis C et al: Herbs and the kidney. Am J Kidney Dis 44:1, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VASCULAR INJURY TO THE KIDNEY: INTRODUCTION

The renal vasculature is unusually complex with rich arteriolar flow to the cortex in excess of metabolic requirements, consistent with its primary function as a filtering organ. After delivering blood to cortical glomeruli, the postglomerular circulation supplies deeper medullary segments that support energy-dependent solute transport at multiple levels of the renal tubule. These postglomerular vessels carry less blood, and high oxygen consumption leaves the deeper medullary regions at the margin of hypoxemia. Vascular disorders that commonly threaten the blood supply of the kidney include large vessel atherosclerosis, fibromuscular diseases, and embolic, inflammatory, and primary hematologic disorders that produce microvascular injury.

ATHEROSCLEROSIS AND KIDNEY CIRCULATION

Microvascular Disease

The glomerular capillary endothelium shares susceptibility to oxidative stress, pressure injury, and inflammation with other vascular territories. Rates of urinary albumin excretion (UAE) are predictive of systemic atherosclerotic disease events. Increased UAE may develop years before cardiovascular events. UAE and the risk of cardiovascular events are both reduced with pharmacologic therapy such as statins. Experimental studies demonstrate functional changes and rarefaction of renal microvessels under conditions of accelerated atherosclerosis and/or compromise of proximal perfusion pressures with large vessel disease **(Fig. 286-1).**

Figure 286-1

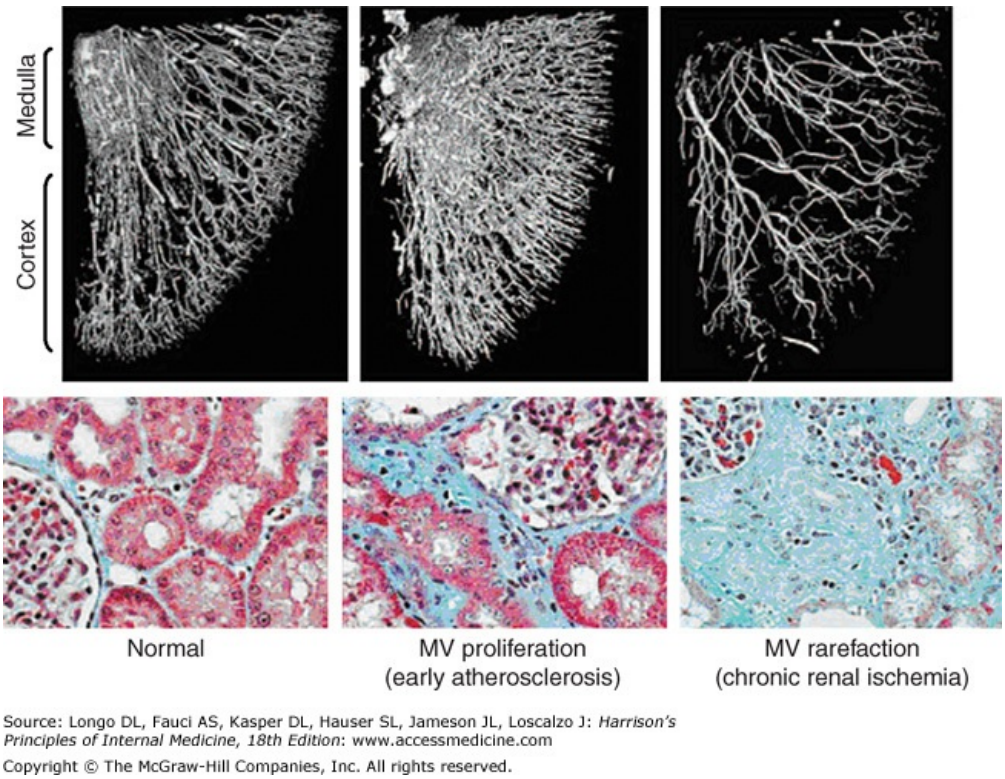

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examples of micro-CT images from vessels defined by radiopaque casts injected into the renal vasculature. These illustrate the complex, dense cortical capillary network supplying the kidney cortex that can either proliferate or succumb to rarefaction under the influence of atherosclerosis and/or occlusive disease. Changes in blood supply are followed by tubulointerstitial fibrosis and loss of kidney function. *MV, microvascular (From LO Lerman, AR Chade: Curr Opin Nephrol Hyper 18:160, 2009, with permission.)*

Macrovascular Disease

Large-vessel renal artery occlusive disease can result from extrinsic compression of the vessel, fibromuscular dysplasias, or most commonly, from atherosclerotic disease. Any disorder that reduces perfusion pressure to the kidney can activate mechanisms that tend to restore renal pressures at the expense of developing systemic hypertension. Because restoration of perfusion pressures can reverse these pathways, renal artery stenosis is considered a specifically treatable "secondary" cause of hypertension.

Renal artery stenosis is common and often has only minor hemodynamic effects. Fibromuscular dysplasia (FMD) is reported in 3–5% of normal subjects presenting as potential kidney donors without hypertension. It may present clinically with hypertension in younger individuals (between age 15 and 50), most often women. FMD does not often threaten kidney function, but sometimes produces total occlusion and can be associated with renal artery aneurysms. Atherosclerotic renal artery stenosis (ARAS) is common in the general population (6.8% of a community-based sample above age 65), a prevalence that increases with age and for patients with other vascular conditions such as coronary artery disease (18–23%) and/or peripheral aortic or lower extremity disease (more than 30%). If untreated, ARAS progresses in nearly 50% of cases over a 5-year period, sometimes to total occlusion. Intensive treatment of arterial blood pressure and statin therapy appear to slow these rates and improve clinical outcomes.

Critical levels of stenosis lead to a reduction in perfusion pressure that activates the renin-angiotensin system, reduces sodium excretion, and activates sympathetic adrenergic pathways. These events lead to systemic hypertension characterized by angiotensin dependence in the early stages, widely varying pressures, loss of circadian blood pressure (BP) rhythms, and accelerated target organ injury, including left ventricular hypertrophy and renal fibrosis. Renovascular hypertension can be treated with agents that block the renin-angiotensin system and other drugs that modify these pressor pathways. It can also be treated with restoration of renal blood flow by either endovascular or surgical revascularization. In most cases, patients require continued antihypertensive drug therapy because revascularization alone rarely lowers BP to normal.

ARAS and systemic hypertension tend to affect both the poststenotic and contralateral kidneys, reducing overall glomerular filtration rate (GFR) in ARAS. When kidney function is threatened by large vessel disease primarily, it has been labeled *ischemic nephropathy*. Unlike FMD, ARAS develops in patients with other risk factors for atherosclerosis and is commonly superimposed upon preexisting small vessel disease in the kidney resulting from hypertension, aging, and diabetes. Nearly 85% of patients considered for renal revascularization have stage 3–5 chronic kidney disease (CKD) with GFR below 60 mL/min per 1.73 m^2. The presence of ARAS is a strong predictor of morbidity- and mortality-related cardiovascular events, independent of whether renal revascularization is undertaken.

Diagnostic approaches to renal artery stenosis depend partly on the specific issues to be addressed. Noninvasive characterization of the renal vasculature may be achieved by several techniques summarized in **(Table 286-1).** Although activation of the renin-angiotensin system is a key step in developing renovascularhypertension, it is transient. Levels of renin activity are therefore subject to timing, the effects of drugs and sodium intake, and do not reliably predict the response to vascular therapy. Renal artery velocities by Doppler ultrasound above 200 cm/s generally predict hemodynamically important lesions (above 60% vessel lumen occlusion), although treatment trials require velocity above 300 cm/s to avoid false positives. The renal resistive index has predictive value regarding the viability of the kidney. It remains operator- and institution- dependent, however. Captopril-enhanced renography has a strong negative predictive value when entirely normal. Magnetic resonance angiography (MRA) is now less often used, as gadolinium contrast has been associated with nephrogenic systemic fibrosis. Contrast-enhanced CT with vascular reconstruction provides excellent vascular images and functional assessment, but carries a small risk of contrast toxicity.

Table 286-1 Summary of Imaging Modalities for Evaluating the Kidney Vasculature

Perfusion Studies to Assess Differential Renal Blood Flow			
Captopril renography with technetium ^{99m}Tc mertiatide (^{99m}Tc MAG3)	Captopril-mediated fall in filtration pressure amplifies differences in renal perfusion	Normal study excludes renovascularhypertension	Multiple limitations in patients with advanced atherosclerosis or creatinine >;2.0 mg/dL (177 μmol/L)
Nuclear imaging with technetium mertiatide or technetium-labeled pentetic acid (DTPA) to estimate fractional flow to each kidney	Estimates fractional flow to each kidney	Allows calculation of single-kidney glomerular filtration rate	Results may be influenced by other conditions, e.g., obstructive uropathy
Vascular Studies to Evaluate the Renal Arteries			
Duplex ultrasonography	Shows the renal	Inexpensive; widely	Heavily dependent on

	arteries and measures flow velocity as a means of assessing the severity of stenosis	available	operator's experience; less useful than invasive angiography for the diagnosis of fibromuscular dysplasia and abnormalities in accessory renal arteries
Magnetic resonance angiography	Shows the renal arteries and perirenal aorta	Not nephrotoxic, but concerns for gadolinium toxicity exclude use in GFR <30 mL/min/1.73 m^2; provides excellent images	Expensive; gadolinium excluded in renal failure, unable to visualize stented vessels
Computed tomographic angiography	Shows the renal arteries and perirenal aorta	Provides excellent images; stents do not cause artifacts	Expensive, moderate volume of contrast required, potentially nephrotoxic
Intra-arterial angiography	Shows location and severity of vascular lesion	Considered "gold standard" for diagnosis of large vessel disease, usually performed simultaneous with planned intervention	Expensive, associated hazard of atheroemboli, contrast toxicity, procedure-related complications, e.g., dissection

Abbrevations: DTPA, diethylenetriamine pentaacetic acid (pentetic acid); GFR, glomerular filtration rate.

Treatment: Renal Artery Stenosis

While restoring renal blood flow and perfusion seems intuitively beneficial for high-grade occlusive lesions, revascularization procedures also pose hazards and expense. Patients with FMD are commonly younger females with otherwise normal vessels and a long life expectancy. These patients often respond well to percutaneous renal artery angioplasty. If blood pressure can be controlled to goal levels and kidney function remains stable in patients with ARAS, it may be argued that medical therapy with follow-up for disease progression is equally effective. Prospective trials up to now fail to identify compelling benefits for interventional procedures regarding short-term results of blood pressure and renal function, although long-term studies regarding cardiovascular outcomes such as stroke, congestive heart failure, myocardial infarction, and end-stage renal failure are not yet complete. Medical therapy should include blockade of the renin-angiotensin system, attainment of goal blood pressures, cessation of tobacco, statins, and aspirin.

Techniques of renal revascularization are improving. With experienced operators, major complications develop in about 9% of cases, including renal artery dissection, capsular perforation, hemorrhage, and occasional atheroembolic disease. Although not common, atheroembolic disease can be catastrophic and accelerate both hypertension and kidney failure, exactly the events that revascularization is intended to prevent. Although renal blood flow usually can be restored by endovascular stenting, recovery of renal function is limited to about 25% of cases, with no change in 50%, and some deterioration evident in others. When hypertension is refractory to effective therapy, revascularization offers real benefits. **Table 286-2** summarizes currently accepted guidelines for considering renal revascularization.

Table 286-2 Clinical Factors Favoring Medical Therapy and Revascularization or Surveillance for Renal Artery Stenosis

Factors Favoring Medical Therapy and Revascularization for Renal Artery Stenosis
• Progressive decline in GFR during treatment of systemic hypertension
• Failure to achieve adequate blood pressure control with optimal medical therapy (medical failure)
• Rapid or recurrent decline in the GFR in association with a reduction in systemic pressure • Decline in the GFR during therapy with ACE inhibitors or ARBs
• Recurrent congestive heart failure in a patient in whom the adequacy of left ventricular function does not explain a cause
Factors Favoring Medical Therapy and Surveillance of Renal Artery Disease
• Controlled blood pressure with stable renal function (e.g., stable renal insufficiency)
• Stable renal artery stenosis without progression on surveillance studies (e.g., serial duplex ultrasound)
• Very advanced age and/or limited life expectancy
• Extensive comorbidity that make revascularization too risky

• High risk for or previous experience with atheroembolic disease
• Other concomitant renal parenchymal diseases that cause progressive renal dysfunction (e.g., interstitial nephritis, diabetic nephropathy)

Abbreviations: ACE, angiotensin-converting enzyme; ARBs, angiotensin receptor blockers; GFR, glomerular filtration rate.

ATHEROEMBOLIC RENAL DISEASE

Emboli to the kidneys arise most frequently as a result of cholesterol crystals breaking free of atherosclerotic vascular plaque and lodging in downstream microvessels. Most clinical atheroembolic events follow angiographic procedures, often of the coronary vessels. It has been argued that nearly all vascular interventional procedures lead to plaque fracture and release of microemboli, but clinical manifestations develop only in a fraction of these. The incidence of clinical atheroemboli has been increasing with more vascular procedures and longer life spans. Atheroembolic renal disease is suspected in more than 3% of end-stage renal disease (ESRD) in elderly subjects and is likely underdiagnosed. It is more frequent in males with history of diabetes, hypertension, and ischemic cardiac disease. Atheroemboli in the kidney are strongly associated with aortic aneurysmal disease and renal artery stenosis. Most clinical cases can be associated with precipitating events, such as angiography, vascular surgery, anticoagulation with heparin, thrombolytic therapy, or trauma. Clinical manifestations of this syndrome commonly develop between 1 and 14 days after an inciting event and may continue to develop for weeks thereafter. Systemic embolic disease manifestations, such as fever, abdominal pain, and weight loss are present in less than half of patients, although cutaneous manifestations including livedo reticularis and localized toe gangrene may be more common. Worsening hypertension and deteriorating kidney function are common, sometimes reaching a malignant phase. Progressive renal failure can occur and require dialytic support. These cases often develop after a stuttering onset over many weeks and have an ominous prognosis. Mortality rate after 1 year reaches 38%, and although some may eventually recover sufficiently to no longer require dialysis, many do not.

Beyond the clinical manifestations above, laboratory findings include rising creatinine, eosinophilia (60–80%), elevated sedimentation rate, and hypocomplementemia (15%). Establishing this diagnosis can be difficult and is often by exclusion. Definitive diagnosis depends upon kidney biopsy demonstrating microvessel occlusion with cholesterol crystals that leave a "cleft" in the vessel. Biopsies obtained from patients undergoing surgical revascularization of the kidney indicate that silent cholesterol emboli are frequently present before any further manipulation is performed.

No effective therapy is available for atheroembolic disease once it has developed. Withdrawal of anticoagulation is recommended. Late recovery of kidney function after supportive measures sometimes occurs, and statin therapy may improve outcome. The role of embolic protection devices in the renal circulation is unclear, but a few prospective trials have failed to demonstrate major benefits. These devices are limited to distal protection during the endovascular procedure and offer no protection from embolic debris after removal.

THROMBOEMBOLIC RENAL DISEASE

Thrombotic occlusion of renal vessels or branch arteries can lead to declining renal function and hypertension. It is difficult to diagnose and is often overlooked, especially in elderly patients. Thrombosis can develop as a result of local vessel abnormalities, such as local dissection, trauma, or inflammatory vasculitis. While hypercoagulability conditions sometimes present as renal artery thrombosis, this is rare. It can also derive from distant embolic events, e.g., the left atrium in patients with atrial fibrillation or from fat emboli originating from traumatized tissue, most commonly large bone fractures. Cardiac sources include vegetations from subacute bacterial endocarditis. Systemic emboli to the kidneys may also arise from the venous circulation if right-to-left shunting occurs, e.g., through a patent foramen ovale.

Clinical manifestations vary depending upon the rapidity of onset and extent of occlusion. Acute arterial thrombosis may produce flank pain, fever, leukocytosis, nausea, and vomiting. If kidney infarction results, enzymes such as lactate dehydrogenase (LDH) rise to extreme levels. If both kidneys are affected, renal function will decline precipitously with a drop in urine output. If a single kidney is involved, renal functional changes may be minor. Hypertension related to sudden release of renin from ischemic tissue can develop rapidly, so long as some viable tissue in the "peri-infarct" border zone remains. If the infarct zone demarcates precisely, the rise in blood pressure and renin activity may resolve. Diagnosis of renal infarction may be established by vascular imaging with MR, CT angiography, or arteriography **(Figs. 286-2A and B).**

Figure 286-2

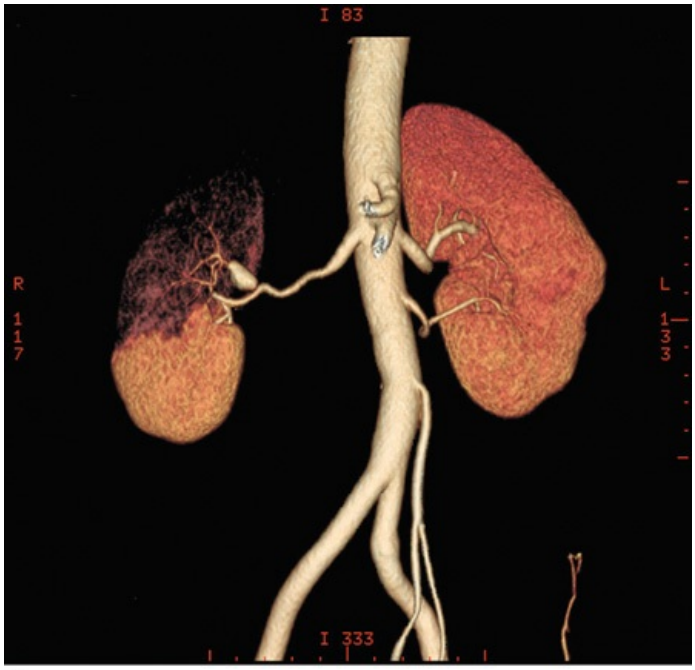

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

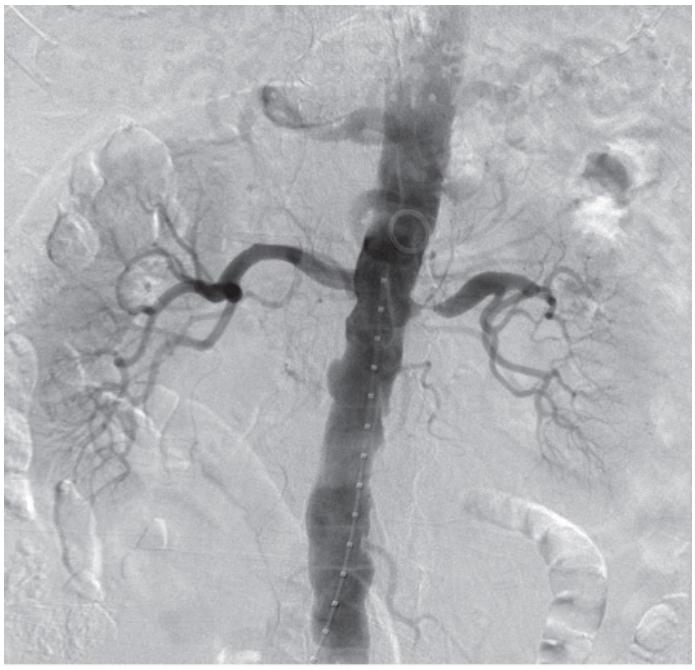

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. **CT angiogram illustrating loss of circulation to the upper pole of the right kidney in a patient with fibromuscular disease and a renal artery aneurysm.** Activation of the renin-angiotensin system produced rapidly developing hypertension. ***B.*** Angiogram illustrating high-grade renal artery stenosis affecting the left kidney. This lesion is often part of widespread atherosclerosis and sometimes is an extension of aortic plaque. This lesion develops in older individuals with preexisting atherosclerotic risk factors.

Management of Arterial Thrombosis of the Kidney

Options for interventions of newly detected arterial occlusion include surgical reconstruction, anticoagulation, thrombolytic therapy, endovascular procedures, and supportive care, particularly antihypertensive drug therapy. Application of these methods depends upon the patient's overall condition, the precipitating factors (e.g., local trauma or systemic illness), the magnitude of renal tissue and function at risk, and the likelihood of recurrent events in the future. For unilateral disease, e.g., arterial dissection with thrombosis, supportive care with anticoagulation may suffice. Acute, bilateral

occlusion is potentially catastrophic, producing anuric renal failure. Depending upon the precipitating event, surgical or thrombolytic therapies can sometimes restore kidney viability.

MICROVASCULAR INJURY AND HYPERTENSION

Arteriolonephrosclerosis

"MALIGNANT" HYPERTENSION

Although BP rises with age, it has long been recognized that some individuals develop rapidly progressive BP elevations with target organ injury including retinal hemorrhages, encephalopathy, and declining kidney function. Placebo arms during the controlled trials of hypertension therapy identified progression to severe levels in 20% of subjects over 5 years. If untreated, patients with target organ injury including papilledema and declining kidney function suffered mortality rates in excess of 50% over 6–12 months, hence the designation "malignant." Postmortem studies of such patients identified vascular lesions, designated "fibrinoid necrosis," with breakdown of the vessel wall, deposition of eosinophilic material including fibrin, and a perivascular cellular infiltrate. A separate lesion was identified in the larger interlobular arteries in many patients with hyperplastic proliferation of the vascular wall cellular elements, deposition of collagen, and separation of layers, designated the "onionskin" lesion. For many of these patients, fibrinoid necrosis led to obliteration of glomeruli and loss of tubular structures. Progressive kidney failure ensued and without dialysis support led to early mortality in untreated malignant-phase hypertension. These vascular changes could develop with pressure-related injury from a variety of hypertensive pathways, including but not limited to activation of the renin-angiotensin system, and severe vasospasm associated with catecholamine release. Occasionally endothelial injury is sufficient to induce microangiopathic hemolysis as discussed below.

Antihypertensive therapy is the mainstay of therapy for malignanthypertension. With effective BP reduction, manifestations of vascular injury including microangiopathic hemolysis and renal dysfunction can improve over time. Whereas mortality in series reported before the era of drug therapy suggested that 1-year mortality rates exceeded 90%, current survival over 5 years exceeds 50%.

Malignanthypertension is less common in Western countries, although it persists in parts of the world where medical care and antihypertensive drug therapy are less available. It most commonly develops in patients with treated hypertension that neglect to take medications, or who may use vasospastic drugs, such as cocaine. Renal abnormalities typically include rising serum creatinine, occasionally hematuria and proteinuria. Biochemical findings may include evidence of hemolysis (anemia, schistocytes, and reticulocytosis) and changes associated with kidney failure. African-American males are more likely to develop rapidly progressive hypertension and kidney failure than are whites in the United States. Genetic polymorphisms (MYH9) that are common in the African-American population and predispose to subtle focal sclerosing glomerular disease may be responsible, with hypertension developing secondary to renal disease in this instance.

"HYPERTENSIVE NEPHROSCLEROSIS"

Based on experience with malignanthypertension and epidemiologic evidence linking BP with long-term risks of kidney failure, it has long been assumed that lesser degrees of hypertension induce less severe, but prevalent changes in kidney vessels and loss of kidney function. As a result, a large portion of patients reaching ESRD without a specific etiologic diagnosis are assigned the designation "hypertensive nephrosclerosis." Pathologic examination commonly identifies afferent arteriolar thickening with deposition of homogeneous eosinophilic material (*hyaline arteriolosclerosis*) associated with narrowing of vascular lumina. Clinical manifestations include retinal vessel changes associated with hypertension (arteriolar narrowing, crossing changes), left ventricular hypertrophy, and elevated blood pressure. The role of these vascular changes in kidney function is unclear. Postmortem and biopsy samples from normotensive kidney donors demonstrate similar vessel changes associated with aging, dyslipidemia, and glucose intolerance. While BP reduction does slow progression of proteinuric kidney diseases and is warranted to reduce the excessive cardiovascular risks associated with CKD, antihypertensive therapy does not alter the course of kidney dysfunction identified specifically as hypertensive nephrosclerosis.

THROMBOTIC MICROANGIOPATHY

Thrombotic microangiopathy (TMA) refers to injured endothelial cells that are thickened, swollen, or detached mainly from arterioles and capillaries. Platelet and hyaline thrombi causing partial or complete occlusion are integral to the histopathology. TMA is the histologic result of microangiopathic hemolytic anemia (MAHA), which consumes platelets and erythrocytes and is characterized by thrombocytopenia and schistocytes. In the kidney, TMA is characterized by swelling of the endocapillary cells (endotheliosis), fibrin thrombi, platelet plugs, arterial intimal fibrosis, and membranoproliferative changes. In severe cases, the fibrin thrombi may extend into the arteriolar vascular pole producing glomerular collapse and sometimes cortical necrosis. Secondary focal segmental glomerulosclerosis may be seen in individuals who recover from acute TMA. Diseases that present with this lesion include thrombotic thrombocytopenia (TTP), hemolytic-uremia syndrome (HUS), malignanthypertension, scleroderma renal crisis, antiphospholipid syndrome,

preeclampsia/HELLP (hemolysis, elevated liver enzymes, low platelet count) syndrome, HIV infection, and radiation nephropathy.

Hemolytic-Uremic Syndrome (HUS)/Thrombotic Thrombocytopenic Purpura (TTP)

HUS and TTP are the prototypes of MAHA. Whether they represent a spectrum of the same disease or two distinct entities continues to be debated. Histologically, the diseases are inseparable, but they differ regarding epidemiology and pathophysiology. Typical HUS usually affects children (most under the age of 5) and is preceded by hemorrhagic diarrhea. Typical TTP affects individuals in their thirties and forties. Neurologic symptoms are more common in TTP and have significant morbidity and mortality rates if not treated with plasma exchange, while plasma exchange is ineffective in most HUS. The argument is strengthened with the discovery of a disintegrin and metalloproteinase with a thrombospondin type 1 motif member 13 (ADAMTS13), a von Willebrand factor (vWF) cleaving protease that is either absent or inactive in TTP but not in HUS. However, neurologic symptoms can occur in HUS, and low ADAMTS13 activity has been identified in HUS cases. Furthermore, plasma infusion/exchange is effective in some HUS. As a result, the distinction between the two is blurred, and they are often identified simply as HUS/TTP.

Hemolytic-Uremic Syndrome

There are at least four variants of HUS. The most common is D+ HUS referring to its association with bacterial gastroenteritis. This typically affects young children (<5 years), but adults are also susceptible. More than 80% of cases are preceded within a week by diarrhea, often bloody. Gastrointestinal symptoms include abdominal pain, cramping, and vomiting. Fever is typically absent. Neurologic symptoms are common and may include lethargy, encephalopathy, seizures, and even cerebral infarction. The pathogenic agent linked to D+ HUS is the shiga toxin, also referred to as *verotoxin*. This toxin is produced by certain strains of *Escherichia coli* and *Shigella dysenteriae.* In the United States and Europe, the most common shiga-toxigenic *E. coli* (STEC) strain is the 0157:H7.

Other strains such as 0157/H$^-$, 0111:H$^-$, 026:H11/H$^-$, and 0145:H28 can also produce shiga toxin. Once shiga toxin enters the circulation, it binds to neutrophils and preferentially localizes in the kidney, where it causes damage to the endothelial cells. This results in platelet aggregation, which initiates the microangiopathic process. Another bacterium associated with HUS is *Streptococcus pneumoniae.* This bacterium produces a neuraminidase that cleaves the *N*-acetyl neuraminic acid moieties that cover the Thomsen-Friedenreich antigen on platelets and endothelial cells. Exposure of this normally cryptic antigen to preformed IgM results in severe MAHA.

Another variant produces atypical HUS (aHUS), caused by congenital complement dysregulation rather than a toxin. These patients have low C3 levels, a characteristic of alternative pathway activation. The most common cause is a deficiency of factor H, which has been linked to families with aHUS. Factor H competes with factor B to prevent the formation of C3b,Bb and acts as a cofactor for factor I, which proteolytically degrades C3b. More than 70 mutations of the factor H gene have been identified. Most are missense mutations that produce normal levels of factor H with abnormalities mainly in the C-terminus region, which affect its binding to C3b. Other mutations result in low or complete absence of the protein. Deficiencies in other complement regulatory proteins such as factor I, factor B, membrane cofactor protein or MCP (CD46), C3, complement factor H–related protein 1 (CFHR1), CFHR3, and CFHR5 have also been described. Finally, an autoimmune variant of HUS has been discovered. Deficient for CHFR protein and factor H autoantibody–positive (DEAP), HUS occurs when an autoantibody is formed against factor H. DEAP-HUS is often associated with a deletion of an 84-kb fragment of the chromosome that encodes for CFHR1 and CFHR3. The autoantibody blocks the binding of factor H to C3b and surface-bound C3 convertase.

Thrombotic Thrombocytopenic Purpura

Traditionally TTP is characterized by the pentad (hemolytic anemia, thrombocytopenia, neurologic symptoms, fever, and renal failure). Classic TTP is differentiated from HUS by neurologic involvement. However, in practice, differentiation between TTP and HUS is unreliable due to overlap in clinical manifestations. TTP has been linked with the absence or marked decreased activity in the metalloprotease ADAMTS13 specific for vWF, although this is not universally present. Even complete absence of ADAMTS13 alone does not produce TTP. Most often, an additional trigger (such as infection, surgery, pancreatitis, or pregnancy) initiates clinical TTP.

Data from the Oklahoma TTP/HUS Registry reveal an incidence rate of 11.3 per 10^6 patients. The median age of the patients was 40 years. Higher frequency was noted among blacks, with an incidence more than nine times higher than non-blacks. Women have nearly three times the incidence, similar to the demographics for systemic lupus erythematosus. If untreated, TTP has a mortality rate exceeding 90%. Even with modern therapy, 20% of the patients die within the first month from complications of microvascular thrombosis.

Several subtypes of TTP have been described. The classic form is acquired or idiopathic TTP, which usually follows an infection, malignancy, or an intense inflammatory reaction such as pancreatitis. This variant typically occurs with deficiency of ADAMTS13 or its activity and is the result of an autoantibody.

The autoantibody (IgG or IgM) can either increase clearance of ADAMTS13 or inhibit its activity. A hereditary form with congenital deficiency of ADAMTS13 is seen in patients with Upshaw-Schulman characterized by MAHA and thrombocytopenia. TTP in these patients can start within the first weeks of life, but in others, may not start until several years of age. Environmental and genetic factors are thought to influence the development of TTP. Plasma transfusion is effective as a prevention and treatment during the TTP episodes.

Drug-induced TTP/TMA is a recognized complication of chemotherapeutic agents, immunosuppressive agents, antiplatelet agents, and quinine. Two mechanisms are responsible for drug-induced TMA. With chemotherapeutic agents (mitomycin C, gemcitabine, etc.) and immunosuppressive agents (cyclosporine, tacrolimus, and sirolimus), endothelial damage is the main cause of the MAHA. This process is usually dose-dependent. Alternatively, drugs can induce autoantibodies that produce TMA. Suppression of ADAMTS13 activity and formation of an autoantibody has been detected in patients exposed to ticlopidine. Quinine appears to induce autoantibodies against granulocytes, lymphocytes, endothelial cells, and platelet glycoprotein IbB/IX or IIb/IIIa complexes but not to ADAMTS13. Quinine-associated TTP is more common in women. Autoantibody-associated TTP can occur after a single dose in patients who had previous exposure to the drug. Most patients developing TTP from clopidogrel do not have either autoantibodies or decreased ADAMTS13 activity. Drugs that inhibit vascular endothelial growth factor (VEGF) sometimes produce TMA. The mechanism is not fully understood.

Treatment of HUS/TTP should be based on the pathophysiologic pathways that are identified. Autoantibody-mediated TTP and DEAP HUS should be treated with plasma exchange or plasmapheresis. In addition to removing the autoantibodies, plasma exchange replaces ADAMTS13. Twice-daily plasma exchanges, vincristine, and rituximab occasionally have been found to be effective in refractory cases. Plasma infusion is usually sufficient for congenital TTP such as Upshaw-Schulman syndrome. Plasma exchange should be considered if larger volumes of plasma are necessary. TTP secondary to drug-induced autoantibodies responds well to plasma exchange, while drugs that cause endothelial damage may not. D+ HUS should be treated with supportive measures. Plasma exchange has not been found to be effective. Antimotility agents and antibiotics increase the incidence of HUS and should be avoided. Conversely, plasma infusion/exchange may be beneficial in aHUS by repleting complement regulatory proteins. Antibiotics and washed red cells should be given in neuraminidase-associated HUS. Plasma and whole blood should be avoided since they contain IgM, which would exacerbate the MAHA. The coexistence of factor H and ADAMTS13 deficiency can exacerbate TTP and make it less responsive to plasma infusion, illustrating the complexity of managing these disorders.

Transplantation-Associated Thrombotic Microangiopathy (Ta-Tma)

TA-TMA can develop after hematopoietic stem cell transplantation (HSCT) with an incidence of 8.2%. Etiologic factors include conditioning regimens, immunosuppression, infections, and graft-versus-host disease. Other risk factors include female sex, age, and human leukocyte antigen (HLA)-mismatched donor grafts. TA-TMA usually occurs within the first 100 days after HSCT. Table 286-3 lists definitions of TA-TMA currently used for clinical trials. A firm diagnosis may be difficult because thrombocytopenia, anemia, and renal insufficiency are common in the posttransplant period. TA-TMA carries a high mortality rate (75% within 3 months). Plasma exchange is beneficial in less than 50% of patients, most of whom have more than 5% ADAMTS13 activity. Calciuria inhibitors should be discontinued, and substitution with daclizumab [antibody to the interleukin 2 (IL-2) receptor] is recommended. Treatment with rituximab and defibrotide may also be helpful.

Table 286-3 Criteria for Establishing Microangiopathic Kidney Injury Associated with Hematopoietic Stem Cell Transplantation

International Working Group	Blood and Marrow Transplant Clinical Trials Network Toxicity Committee
>;4% schistocytes in the blood	RBC fragmentation and at least 2 schistocytes per high-power field
De novo, prolonged, or progressive thrombocytopenia	Concurrent increase in LDH above baseline
A sudden and persistent increase in LDH	Negative direct and indirect Coombs test
Decrease in hemoglobin or increased RBC transfusion requirement	Concurrent renal and/or neurologic dysfunction without other explanations
Decrease in haptoglobin concentration	

Note: These features underscore the need to identify pathways of hemolysis and thrombocytopenia that accompany deterioration of kidney function.

Abbreviations: LDH, lactate dehydrogenase; RBC, red blood cell.

HIV-Related Tma

TMA is mainly a complication encountered before widespread use of highly active retroviral therapy for HIV. It is seen in patients with advanced AIDS and low CD4 count, although it occasionally can be the first manifestation of HIV infection. The presence of MAHA thrombocytopenia and renal failure are suggestive, but renal biopsy is required to establish the diagnosis since HIV is associated with several other renal diseases. The median platelet count is 77,000/μL with a range of 10,000 to 160,000/μL, which may prohibit a renal biopsy in some patients. Cytomegalovirus (CMV) coinfection may also be a risk factor. The mechanism of injury is unclear, but HIV may induce apoptosis in endothelial cells. Plasma exchange is effective and is recommended in conjunction with antiviral therapy.

Radiation Nephropathy

Radiation can produce microangiopathic injury after either local or total body irradiation. The kidney is one of the most radiosensitive organs, and injury can result with as little as 4–5 Gy exposure. It is characterized by renal insufficiency, proteinuria, and hypertension usually presenting 6 months or longer after radiation exposure. Renal biopsy reveals classic TMA in the kidney with damage to glomerular, tubular, and vascular cells. Systemic evidence for MAHA is rare. Because of its high incidence after allogeneic HSCT, it is often referred to as bone marrow transplant (BMT) nephropathy. No specific therapy is available, although some evidence supports treatment with renin-angiotensin system blockade.

Scleroderma (Progressive Systemic Sclerosis)

Scleroderma commonly affects the kidney, with 52% of subjects with widespread scleroderma having renal involvement sometime during the follow-up period. Of these, 19% were due to scleroderma renal crisis (SRC). Other renal manifestations in scleroderma include transient (prerenal) or medication-related forms of acute kidney injury [e.g., d-penicillamine, nonsteroidal anti-inflammatory drugs (NSAIDs), or cyclosporine]. SRC occurs in patients with diffuse systemic sclerosis (12 vs. 2% in limited systemic sclerosis). SRC is the most severe manifestation, characterized by accelerated hypertension, a rapid decline in renal function, nephrotic proteinuria, and hematuria. Retinopathy and encephalopathy may accompany the hypertension. Salt and water retention with microvascular injury can lead to pulmonary edema. Other manifestations include myocarditis, pericarditis, and arrhythmias, which denote an especially poor prognosis. Although MAHA is present in over half of the patients, coagulopathy is rare.

The renal lesion in SRC is characterized by arcuate artery intimal and medial proliferation with luminal narrowing. This lesion is described as *onionskinning* and can be accompanied by glomerular collapse due to reduced blood flow. Histologically it is indistinguishable from malignanthypertension. Fibrinoid necrosis and thrombosis are common. Before the availability of angiotensin-converting enzyme (ACE) inhibitors, the mortality rate for SRC at 1 month was greater than 90%. Introduction of renin-angiotensin system blockade has lowered the mortality rate to 30% at 3 years. Nearly two-thirds of patients with SRC require dialysis support. Half of those needing dialysis as a result of SRC will recover renal function (median time = 1 year). Glomerulonephritis and vasculitis associated with antineutrophil cytoplasmic antibodies (ANCAs) and systemic lupus erythematosus have been described in patients with scleroderma. An association has been found with antinuclear antibodies' (ANAs) speckled pattern and anti-RNA polymerase antibodies (I and III). Anti-U3-RNP may identify young patients at risk for SRC. Anticentromere antibody (ACA), however, is a negative predictor of SRC. Because of the overlap between SRC and other autoimmune disorders, a renal biopsy is recommended for patients with atypical renal involvement, especially if hypertension is absent.

Treatment with ACE inhibition is the first-line therapy unless contraindicated. The goal of therapy is to reduce systolic blood pressure by 20 mmHg and diastolic by 10 mmHg every 24 hours until blood pressure is normalized. Additional antihypertensive therapy may be added once the ACE inhibition is maximized. Both ACE inhibitors and angiotensin II receptor antagonists are effective, although published data show that treatment is superior with ACE inhibitors. ACE inhibition alone does not prevent SRC, although it reduces the role of hypertension. Intravenous iloprost has been used in Europe for blood pressure management and improvement of renal perfusion. Kidney transplantation is not recommended for 2 years after the start of dialysis, since delayed recovery may occur.

Antiphospholipid Syndrome (Aps)

Antiphospholipid syndrome (Chap. 320) can be either primary or secondary to systemic lupus erythematosus. It is characterized by systemic thrombosis (arterial and venous) and fetal morbidity mediated by antiphospholipid antibodies (aPLs). The aPLs are mainly anticardiolipin (aCL) antibodies, which can be IgG, IgM, or IgA, lupus anticoagulant (LA), and anti-ß-2 glycoprotein I antibodies (antiß 2GPI). Patients with both aCL and antiß2GPI appear to have the highest risk of thrombosis. The vascular compartment within the kidney is the main site of renal involvement. Arteriosclerosis is commonly present in the arcuate and intralobular arteries. In the intralobular arteries, fibrous intimal hyperplasia characterized by intimal thickening secondary to intense myofibroblastic intimal cellular proliferation with extracellular matrix deposition is frequently seen along with onionskinning. Arterial and arteriolar fibrous and fibrocellular occlusions are present in over two-thirds of the biopsies. Cortical necrosis and focal cortical atrophy may result from vascular occlusion. TMA is commonly

present in the renal biopsies, although signs of MAHA and platelet consumption are usually absent. TMA is especially common in the catastrophic variant of APS. In patients with secondary antiphospholipid syndrome (APS), other glomerulopathies may be present including membranous nephropathy, minimal change disease, focal segmental glomerulosclerosis, and pauci-immune crescentic glomerulonephritis.

Large vessels can be involved in APS and may form the proximal nidus near the ostium for thrombosis of the renal artery. Renal vein thrombosis can occur and should be suspected in patients with lupus anticoagulant LA who develop nephrotic range proteinuria. Progression to end-stage renal disease can occur, and thrombosis may form in the vascular access and the renal allografts. Hypertension is common. Treatment entails lifelong anticoagulation. Glucocorticoids may be beneficial in accelerated hypertension. Immunosuppression and plasma exchange may be helpful for catastrophic episodes of APS, but themselves do not reduce recurrent thrombosis.

HELLP Syndrome

HELLP (hemolysis, elevated liver enzymes, low platelets) syndrome is a dangerous complication of pregnancy. Occurring in 0.5–0.9% of all pregnancies and 10–20% of cases with severe preeclampsia, it has a mortality rate that ranges between 7.4 and 34%. Most commonly occurring in the third trimester, 10% of cases occur before week 27 and 30% postpartum. Although most consider HELLP to be a severe form of preeclampsia, nearly 20% are not preceded by preeclampsia. HELLP patients have increased inflammatory markers [C-reactive protein (CRP), IL-1Ra, and IL-6] as compared to preeclampsia alone.

Renal failure occurs in half of patients with HELLP, although the etiology is not well understood. Limited data suggest renal failure is the result of a combination of preeclampsia and acute tubular necrosis from HELLP. Renal histologic findings are those of TMA with endothelial cell swelling and occlusion of the capillary lumens, but luminal thrombi are typically absent. However, thrombi become more common in severe eclampsia and HELLP. Although renal failure is common, the organ that defines this syndrome is the liver. Subcapsular hepatic hematomas sometimes produce spontaneous rupture of the liver and can be a life-threatening complication. Neurologic complications such as strokes, cerebral infarcts, cerebral and brainstem hemorrhage, and cerebral edema are other major potentially life-threatening complications. Nonfatal complications include placental abruption, permanent vision loss due to Purtscher-like (hemorrhagic and vasoocclusive vasculopathy) retinopathy, pulmonary edema, bleeding, and fetal demise.

The HELLP syndrome shares many features with other forms of MAHA. Distinguishing the specific disorders is complicated by the fact that both aHUS and TTP flares can be triggered by pregnancy. Patients with antiphospholipid syndrome have a higher risk of HELLP. A history of episodes of MAHA before pregnancy is helpful. Serum levels of ADAMTS13 activity is reduced (30–60%) in HELLP but not to the levels seen in TTP (<5%). Some authors suggest using LDH to AST ratio for diagnosis. Patients with HELLP and preeclampsia have an LDH to AST ratio of 13 to 1 versus 29 to 1 in patients without preeclampsia. Other markers such as antithrombin III (decreased in HELLP but not in TTP) and d-dimer (elevated in HELLP but not in TTP) may aid in the diagnosis. HELLP syndrome resolves spontaneously in most cases after delivery, although a portion of HELLP occurs postpartum. Glucocorticoids may decrease inflammatory markers, although two randomized, controlled trials failed to confirm beneficial effects. Plasma exchange should be considered if the hemolysis is refractory to glucocorticoids and/or delivery, especially if TTP had not been ruled out.

Sickle Cell Nephropathy

Renal complications in sickle cell disease (SCD) result from occlusion of the vasa recta in the renal medulla. The low partial pressure of oxygen and high osmolarity predispose to hemoglobin S polymerization and erythrocyte sickling. Sequelae include hyposthenuria, hematuria, and papillary necrosis. The kidney responds by increasing blood flow and GFR mediated by prostaglandins. This dependence on prostaglandins may explain why patients with SCD experience greater reduction of GFR by NSAIDs than others. The glomeruli are typically enlarged. Intracapillary fragmentation and phagocytosis of sickled erythrocytes are thought to be responsible for the membranoproliferative glomerulonephritis-like lesion, and focal segmental glomerulosclerosis is sometimes seen. Proteinuria is present in 20–30% of the patients, and nephrotic range proteinuria is associated with renal failure. ACE inhibitors reduce proteinuria, although data are lacking on prevention of renal failure. Patients with SCD are also more prone to acute renal failure. The cause is thought to reflect microvascular occlusion associated with nontraumatic rhabdomyolysis, high fever, infection, and generalized sickling. Chronic kidney disease is present in 12–20% of patients. Despite the frequency of renal disease, hypertension is uncommon in patients with SCD.

RENAL VEIN THROMBOSIS

Renal vein thrombosis (RVT) can either present with flank pain, tenderness, hematuria, rapid decline in renal function, and proteinuria or it can be silent. Occasionally, RVT is identified during workup for pulmonary embolism. The left renal vein is more commonly involved and two-thirds of cases are bilateral. Etiologies can be divided into three broad categories: endothelial damage, venous stasis, and

hypercoagulable states. Homocystinuria, endovascular intervention, and surgery can produce vascular endothelial damage. Dehydration, which is more common in males, is a common cause of stasis in the pediatric population. Stasis also can result from compression and kinking of the renal veins from retroperitoneal processes such as retroperitoneal fibrosis and abdominal neoplasms. Thrombosis can occur throughout the renal circulation with antiphospholipid antibody syndrome. RVT can also be secondary to nephrotic syndrome, particularly membranous nephropathy. Other hypercoagulable states associated with RVT include proteins C and S, antithrombin deficiency, factor V Leiden, disseminated malignancy, and oral contraceptives.

Diagnostic screening can be performed with Doppler ultrasound, which is more sensitive than ultrasound alone. The most sensitive test is CT angiography, which is nearly 100% sensitive. MR angiography is another option but is more expensive and requires sedation in pediatric patients. Treatment for RVT is anticoagulation and therapy for the underlying cause. Endovascular thrombolysis may be considered in severe cases. Occasionally nephrectomy may be undertaken for life-threatening complications. Vena caval filters are often used to prevent migration of the thrombi.

FURTHER READINGS

de Mast Q, Beutler JJ: The prevalence of atherosclerotic renal artery stenosis in risk groups: A systematic literature review. J Hypertens 27:1333, 2009

Denton CP: Renal manifestations of systemic sclerosis—clinical features and outcome assessment. Rheumatology 47(Suppl 5): v54, 2008

Garovic V, Textor SC: Renovascular hypertension and ischemic nephropathy. Circulation 112:1362, 2005[PMID: 16129817]

George JN: The thrombotic thrombocytopenic purpura and hemolytic uremic syndromes: Overview of pathogenesis (Experience of the Oklahoma TTP-HUS Registry, 1989–2007). Kidney Int Suppl 112:S8, 2009

Nochy D et al: The intrarenal vascular lesions associated with primary antiphospholipid syndrome. J Am Soc Nephrol 10:507, 1999[PMID: 10073601]

Penn H, Denton CP: Diagnosis, management and prevention of scleroderma renal disease. Curr Opin Rheumatol 20:692, 2008[PMID: 18946330]

Slovut DP, Olin JW: Current concepts: Fibromuscular dysplasia. N Engl J Med 350:1862, 2004[PMID: 15115832]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NEPHROLITHIASIS: INTRODUCTION

Kidney stones are one of the most common urologic problems. In the United States, ~13% of men and 7% of women will develop a kidney stone during their lifetimes, and the prevalence is increasing throughout the industrialized world.

Types of Stones

Calcium salts, uric acid, cystine, and struvite are the constituents of most kidney stones in the western hemisphere (Chap. e14). Calcium oxalate and calcium phosphate stones make up 75–85% of the total **(Table 287-1)** and those constituents may be admixed in the same stone. Calcium phosphate in stones is usually hydroxyapatite [$Ca_5(PO_4)_3OH$] or, less commonly, brushite ($CaHPO_4H_2O$), although the incidence of brushite stones is increasing.

Table 287-1 Major Causes of Renal Stones

Stone Type and Causes	Percent of all Stones[a]	Percent Occurrence of Specific Causes[a]	Ratio of Males to Females	Etiology	Diagnosis	Treatment
Calcium stones	75–85		2:1 to 3:1			
Idiopathic hypercalciuria		50–55	2:1	? Hereditary	Normocalcemia, unexplained hypercalciuria[b]	Low-sodium, low-protein diet; thiazide diuretics
Hyperuricosuria		20	4:1	Diet	Urine uric acid >750 mg per 24 h (women), >800 mg per 24 h (men)	Allopurinol or low-purine diet
Primary hyperparathyroidism		3–5	3:10	Neoplasia	Hypercalcemia with nonsuppressed parathyroid hormone	Surgery
Distal renal tubular acidosis		Rare	1:1	Hereditary or acquired	Hyperchloremic acidosis, minimum urine pH >5.5	Alkali replacement
Dietary hyperoxaluria		10–30	1:1	High-oxalate diet or low-calcium diet	Urine oxalate >40 mg per 24 h	Low-oxalate, normal-calcium diet
Enteric hyperoxaluria		~1–2	1:1	Bowel surgery	Urine oxalate >75 mg per 24 h	Low-oxalate diet and oral calcium pills
Primary hyperoxaluria		Rare	1:1	Hereditary	Urine oxalate and glycolic or l-glyceric acid increased	Fluids, pyridoxine, citrate and neutral phosphate
Hypocitraturia		20–40	1:1 to 2:1	? Hereditary, diet	Urine citrate <320 mg per 24 h	Alkali supplements
Idiopathic stone disease		20	2:1	Unknown	None of the above present	Oral phosphate, fluids
Uric acid stones	5–10					

Metabolic syndrome		~30	1:1	Diet	Glucose intolerance, obesity, hyperlipidemia	Alkali and allopurinol if daily urine uric acid >1000 mg
Gout		~30	3:1 to 4:1	Hereditary	Clinical diagnosis	Alkali and allopurinol
Idiopathic		~30	1:1	? Hereditary	Uric acid stones, no gout	Alkali and allopurinol if daily urine uric acid >1000 mg
Dehydration		?	1:1	Intestinal, habit	History, intestinal fluid loss	Alkali, fluids, reversal of cause
Lesch-Nyhan syndrome		Rare	Males only	Hereditary	Reduced hypoxanthine-guanine phosphoribosyltransferase level	Allopurinol
Cystine stones	1		1:1	Hereditary	Stone type; elevated cystine excretion	Massive fluids, alkali, D-penicillamine if needed
Struvite stones	5		1:3	Infection	Stone type	Antimicrobial agents and judicious surgery

[a]Values are percentages of patients who form a particular type of stone and who display each specific cause of stones.

[b]Urine calcium >300 mg/24 h (men), 250 mg/24 h (women), or 4 mg/kg per 24 h either sex. Hyperthyroidism, Cushing's syndrome, sarcoidosis, malignant tumors, immobilization, vitamin D intoxication, rapidly progressive bone disease, and Paget's disease all cause hypercalciuria and must be excluded in diagnosis of idiopathic hypercalciuria.

Calcium stones are more common in men; the average age of onset is the third to fourth decade. Approximately 50% of people who form a single calcium stone form another within the next 10 years, and some form multiple recurrent stones. The average rate of new stone formation in recurrent stone formers is about one stone every 3 years. *Uric acid stones* account for 5–10% of kidney stones and are also more common in men. Five percent of stones are *struvite*, whereas *cystine stones* are uncommon, accounting for ~1% of cases in most series of nephrolithiasis.

Manifestations of Stones

As stones grow on the surfaces of the renal papillae or within the collecting system, they do not necessarily produce symptoms. Asymptomatic stones may be discovered during the course of radiographic studies undertaken for unrelated reasons. Stones are a common cause of isolated hematuria. Stones become symptomatic when they enter the ureter or occlude the ureteropelvic junction, causing pain and obstruction.

STONE PASSAGE

A stone can traverse the ureter without symptoms, but passage usually produces pain and bleeding. The pain begins gradually, usually in the flank, but increases over the next 20–60 min to become so severe that narcotics may be needed for its control. The pain may remain in the flank or spread downward and anteriorly toward the ipsilateral loin, testis, or vulva. A stone in the portion of the ureter within the bladder wall causes frequency, urgency, and dysuria that may be confused with urinary tract infection. The vast majority of ureteral stones <0.5 cm in diameter pass spontaneously.

Helical computed tomography (CT) scanning without radiocontrast enhancement is now the standard radiologic procedure for diagnosis of nephrolithiasis. The advantages of CT include detection of uric acid stones in addition to the traditional radiopaque stones, no exposure to the risk of radiocontrast agents, and possible diagnosis of other causes of abdominal pain in a patient suspected of having renal colic from stones. Ultrasound is not as sensitive as CT in detecting renal or ureteral stones. Standard abdominal x-rays may be used to monitor patients for formation and growth of kidney stones, as they are less expensive and provide less radiation exposure than CT scans. Calcium, cystine, and struvite stones are all radiopaque on standard x-rays, whereas uric acid stones are radiolucent.

OTHER SYNDROMES

Staghorn Calculi

Struvite, cystine, and uric acid stones often grow too large to enter the ureter. They gradually fill the renal pelvis and may extend outward through the infundibula to the calyces themselves. Very large staghorn stones can have surprisingly few symptoms and may lead to the eventual loss of kidney function.

Nephrocalcinosis

Calcium stones grow on the papillae. Most break loose and cause colic, but they may remain in place so that multiple papillary calcifications are found by x-ray, a condition termed *nephrocalcinosis*. Papillary nephrocalcinosis is common in hereditary distal renal tubular acidosis (RTA) and in other types of severe hypercalciuria. In medullary sponge kidney disease (Chap. 284), calcification may occur in dilated distal collecting ducts.

INFECTION

Although urinary tract infection is not a direct consequence of stone disease, it can occur after instrumentation and surgery of the urinary tract, which are used frequently in the treatment of stone disease. Stone disease and urinary tract infection can enhance their respective seriousness and interfere with treatment. Obstruction of an infected kidney by a stone may lead to sepsis and extensive damage of renal tissue, since it converts the urinary tract proximal to the obstruction into a closed space that can become an abscess. Stones may harbor bacteria in the stone matrix, leading to recurrent urinary tract infection, and infection due to bacteria that have the enzyme urease can cause stones composed of struvite.

ACTIVITY OF STONE DISEASE

In active disease, new stones are forming or preformed stones are growing. Sequential radiographs are needed to document the growth or appearance of new stones and ensure that passed stones are actually newly formed, not preexistent.

Pathogenesis of Stones

Urinary stones usually arise because of the breakdown of a delicate balance between solubility and precipitation of salts. The kidneys must conserve water, but they must excrete materials that have low solubility. These two opposing requirements must be balanced during adaptation to diet, climate, and activity. The problem is mitigated to some extent by the fact that urine contains substances such as pyrophosphate, citrate, and glycoproteins that inhibit crystallization. These protective mechanisms are less than perfect. When urine becomes supersaturated with insoluble materials, because excretion rates are excessive and/or because water conservation is extreme, crystals form and may grow and aggregate to form a stone.

SUPERSATURATION

A solution in equilibrium with a solid phase is said to be saturated with respect to that substance. If the concentration of a substance in a solution is above the saturation point, the solution is said to be supersaturated and can support the growth of crystals, and if supersaturation is excessive, new crystals can begin to develop spontaneously. Excessive supersaturation is common in stone formation.

Calcium, oxalate, and phosphate form many soluble complexes among themselves and with other substances in urine, such as citrate. As a result, their free ion activities are below their chemical concentrations. Reduction in ligands such as citrate can increase ion activity and therefore supersaturation. Urine supersaturation can be increased by dehydration or by overexcretion of calcium, oxalate, phosphate, cystine, or uric acid. Urine pH is also important; phosphate and uric acid are acids that dissociate readily over the physiologic range of urine pH. Alkaline urine contains more dibasic phosphate, favoring deposits of brushite and apatite. Below a urine pH of 5.5, uric acid crystals predominate, whereas phosphate crystals are rare. The solubility of calcium oxalate is not influenced by changes in urine pH. Measurements of supersaturation in a 24-h urine sample probably underestimate the risk of precipitation. Transient dehydration, variation of urine pH, and postprandial bursts of overexcretion may cause spikes in supersaturation.

CRYSTALLIZATION

When urine supersaturation is excessive, crystals begin to nucleate. Once formed, crystal nuclei will grow in size if urine is supersaturated with respect to that crystal phase. Multiple crystals can then aggregate to form a kidney stone. For a kidney stone to form, crystals must be retained in the renal pelvis long enough to grow and aggregate to a clinically significant size. Recent studies have shown that common calcium oxalate kidney stones form as overgrowths on apatite plaques in the renal papillae. These plaques, called Randall's plaques, provide an excellent surface for heterogeneous nucleation of calcium oxalate salts. The Randall's plaques begin in the deep medulla in the basement membrane of the thin limb of the loop of Henle and then spread through the interstitium to the basement membrane of the papillary urothelium. If the urothelium becomes damaged, the plaque is exposed to the urine, and calcium oxalate crystals form on the plaque, accumulating a clinically significant mass to form a stone. Calcium phosphate stone formers, particularly formers of brushite, do

not follow this pattern. Inner medullary collecting ducts are plugged with apatite crystals, and stones form as extension of those plugs. Unlike in calcium oxalate stone formers, renal papillae are often fibrotic and deformed.

Evaluation and Treatment of Patients with Nephrolithiasis

Most patients with nephrolithiasis have remediable metabolic disorders that cause stones and can be detected by chemical analyses of serum and urine. Adults with recurrent kidney stones and children with even a single kidney stone should be evaluated. A practical outpatient evaluation consists of two 24-h urine collections, with a corresponding blood sample; measurements of serum and urine calcium, uric acid, electrolytes, and creatinine, along with urine pH, volume, oxalate, and citrate should be made. Since stone risks vary with diet, activity, and environment, at least one urine collection should be made on a weekend when the patient is at home and another on a workday. When possible, the composition of kidney stones should be determined because treatment depends on stone type (Table 287-1). No matter what disorders are found, every patient should be counseled to avoid dehydration and drink copious amounts of water. The efficacy of high fluid intake was confirmed in a prospective study of first-time stone formers. Increasing urine volume to 2.5 L per day resulted in a 50% reduction of stone recurrence compared with the control group.

Treatment: Nephrolithiasis

The management of stones already present in the kidneys or urinary tract requires a combined medical and surgical approach. The specific treatment depends on the location of the stone, the extent of obstruction, the nature of the stone, the function of the affected and unaffected kidneys, the presence or absence of urinary tract infection, the progress of stone passage, and the risks of operation or anesthesia in light of the clinical state of the patient. Medical therapy can enhance passage of ureteral stones. Oral α_1-adrenergic blockers relax ureteral muscle and have been shown to reduce time to stone passage and the need for surgical removal of small stones. Severe obstruction, infection, intractable pain, and serious bleeding are indications for removal of a stone.

Advances in urologic technology have rendered open surgery for stones a rare event. There are now three alternatives for stone removal. *Extracorporeal lithotripsy* causes the in situ fragmentation of stones in the kidney, renal pelvis, or ureter by exposing them to shock waves. After multiple shock waves, most stones are reduced to powder that moves through the ureter into the bladder. *Percutaneous nephrolithotomy* requires the passage of a nephroscope into the renal pelvis through a small incision in the flank. Stones are then disrupted by a small ultrasound transducer or holmium laser. The third method is *ureteroscopy* with stone disruption using a holmium laser. Ureteroscopy generally is used for stones in the ureter, but some surgeons are now using ureteroscopy for stones in the renal pelvis as well.

CALCIUM STONES

Idiopathic Hypercalciuria

This condition is the most common metabolic abnormality found in patients with nephrolithiasis (Table 287-1). It is familial and is probably a polygenic trait, although there are some rare monogenic causes of hypercalciuria and kidney stones such as Dent's disease, which is an X-linked disorder characterized by hypercalciuria, nephrocalcinosis, and progressive kidney failure. Idiopathic hypercalciuria is diagnosed by the presence of hypercalciuria without hypercalcemia and the absence of other systemic disorders known to affect mineral metabolism. Vitamin D overactivity through either high calcitriol levels or excess vitamin D receptor is a likely explanation for the hypercalciuria in many patients. Recent studies have shown that a polymorphism (Arg990Gly) of the calcium-sensing receptor, which leads to activation of the receptor, is more common in hypercalciuric subjects and probably contributes to higher urine calcium excretion. Hypercalciuria contributes to stone formation by raising urine saturation with respect to calcium oxalate and calcium phosphate.

Treatment: Hypercalciuria

For many years the standard therapy for hypercalciuria was dietary calcium restriction. However, studies have shown that low-calcium diets increase the risk of incident stone formation, perhaps by reducing the amount of calcium in the intestine to bind oxalate, thereby increasing urine oxalate levels. A 5-year prospective trial compared the efficacy of a low-calcium diet to a low-protein, low-sodium, normal-calcium diet in preventing stone recurrence in male calcium stone formers. The group on the low-calcium diet had a significantly greater rate of stone relapse. In addition, hypercalciuric stone formers have reduced bone mineral density and an increased risk of fracture compared with the non-stone-forming population. Low calcium intake probably contributes to the low bone mineral density. In sum, low-calcium diets are of unknown efficacy in preventing stone formation and carry a long-term risk of bone disease, making low-sodium and low-protein diets a superior treatment option. If diet therapy is not sufficient to prevent stones, thiazide diuretics may be used. Thiazide diuretics lower urine calcium and are effective in preventing the formation of stones. Three 3-year randomized trials have shown a 50% decrease in stone formation in the thiazide-treated groups compared with the placebo-treated controls. The drug effect requires slight contraction of the extracellular fluid volume, and high dietary NaCl intake reduces its therapeutic effect. Thiazide-induced hypokalemia should be

treated aggressively since hypokalemia will reduce urine citrate, an important inhibitor of calcium crystallization.

Hyperuricosuria

About 20% of calcium oxalate stone formers are hyperuricosuric, primarily because of an excessive intake of purine from meat and fish. The mechanism of stone formation probably involves salting out calcium oxalate by urate. A low-purine diet is desirable but difficult for many patients to achieve. The alternative is allopurinol, which has been shown to be effective in a randomized, controlled trial.

Primary Hyperparathyroidism

(See also Chap. 353) The diagnosis of this condition is established by documenting that hypercalcemia that cannot be otherwise explained is accompanied by inappropriately elevated serum concentrations of parathyroid hormone. Hypercalciuria, which usually is present, raises the urine supersaturation of calcium phosphate and/or calcium oxalate (Table 287-1). Calcium oxalate stones form on interstitial apatite plaque, whereas calcium phosphate stones form on apatite crystals, obstructing collecting ducts. In patients who have hyperparathyroidism, the Arg990Gly polymorphism of the calcium-sensing receptor leads to higher urine calcium excretion and an increased risk of nephrolithiasis. Prompt diagnosis of hyperparathyroidism is important because parathyroidectomy should be carried out before recurrent stones or renal damage occurs.

Distal Renal Tubular Acidosis

(See also Chap. 284) The defect in this condition seems to reside in the distal nephron, which cannot establish a normal pH gradient between urine and blood, leading to hyperchloremic acidosis. The diagnosis is suggested by a minimum urine pH >5.5 in the presence of systemic acidosis. Hypercalciuria, an alkaline urine, and a low urine citrate level increase urine saturation with respect to calcium phosphate. Calcium phosphate stones form, nephrocalcinosis is common, and osteomalacia or rickets may occur. Apatite deposits form in inner medullary collecting ducts and cause extensive medullary tubular interstitial nephropathy, which can lead to reduced kidney function. Renal tubular acidosis may be genetic, secondary to a systemic disease, or caused by a medication. Topiramate, a drug commonly used for seizures and migraines, inhibits the enzyme carbonic anhydrase and may cause calcium nephrolithiasis.

Treatment with supplemental alkali reduces hypercalciuria and limits the production of new stones. The preferred form of alkali is potassium citrate, which is given at a dose of 0.5–2.0 meq/kg body weight in two to three divided doses per day. In incomplete distal renal tubular acidosis, systemic acidosis is absent, but urine pH cannot be lowered below 5.5 after an exogenous acid load. Incomplete RTA may develop in some patients who form calcium oxalate stones because of idiopathic hypercalciuria; the importance of RTA in producing stones in this situation is uncertain, and thiazide treatment is a reasonable alternative. Alkali also can be used in incomplete RTA. In treating patients with alkali, it is prudent to monitor changes in urine citrate and pH. If urine pH increases without an increase in citrate, calcium phosphate supersaturation will increase and stone disease may worsen.

Hyperoxaluria

Oxalate is a metabolic end product in humans. Urine oxalate comes from diet and endogenous metabolic production, with ~40–50% originating from dietary sources. The upper limit of normal for oxalate excretion is generally considered to be 40–50 mg per day. Mild hyperoxaluria (50–80 mg/d) usually is caused by excessive intake of high-oxalate foods such as spinach, nuts, and chocolate. In addition, low-calcium diets may promote hyperoxaluria as there is less calcium available to bind oxalate in the intestine. Enteric hyperoxaluria is a consequence of small-bowel disease, resulting in fat malabsorption. Oxalate excretion is often >100 mg per day. Enteric hyperoxaluria may be caused by jejunoileal bypass for obesity, pancreatic insufficiency, or extensive small-intestine involvement from Crohn's disease. With fat malabsorption, calcium in the bowel lumen is bound by fatty acids instead of oxalate, which is left free for absorption in the colon. Delivery of unabsorbed fatty acids and bile salts to the colon injures the colonic mucosa and enhances oxalate absorption. Recent studies have shown that modern bariatric surgery for obesity that involves bypassing intestinal segments, such as Roux-en-Y gastric bypass and biliopancreatic diversions, may lead to hyperoxaluria that can cause kidney failure as well as kidney stones. The mechanism of hyperoxaluria has not been well studied.

Primary hyperoxaluria is a rare autosomal recessive disease that causes severe hyperoxaluria. Patients usually present with recurrent calcium oxalate stones during childhood. Primary hyperoxaluria type 1 is due to a deficiency in the peroxisomal enzyme alanine:glyoxylate aminotransferase. Type 2 is due to a deficiency of D-glyceric dehydrogenase. Severe hyperoxaluria from any cause can lead to stone formation and produce tubulointerstitial nephropathy (Chap. 285).

Treatment: Hyperoxaluria

Patients with mild to moderate hyperoxaluria should be treated with a diet low in oxalate and with a normal intake of calcium and magnesium to reduce oxalate absorption. Enteric hyperoxaluria can be treated with a low-fat, low-oxalate diet and calcium supplements, given with meals, to bind oxalate in the gut lumen. The oxalate-binding resin cholestyramine provides an additional form of therapy.

Treatment for primary hyperoxaluria includes a high fluid intake, neutral phosphate, potassium citrate, and pyridoxine (25–200 mg/d). Even with aggressive therapy, irreversible renal failure may occur. Liver transplantation to correct the enzyme defect, combined with kidney transplantation, has been successfully utilized in patients with primary hyperoxaluria.

Hypocitraturia

Urine citrate prevents calcium stone formation by creating a soluble complex with calcium, effectively reducing free urine calcium. Hypocitraturia is found in 20–40% of stone formers either as a single disorder or in combination with other metabolic abnormalities. It can be secondary to systemic disorders such as RTA, chronic diarrheal illness, and hypokalemia, or it may be a primary disorder, in which case it is called *idiopathic hypocitraturia*.

Treatment: Hypocitraturia

Treatment is with alkali, which increases urine citrate excretion; generally, bicarbonate or citrate salts are used. Potassium salts are preferred as sodium loading increases urinary excretion of calcium, reducing the effectiveness of treatment. Two randomized, placebo-controlled trials have demonstrated the effectiveness of citrate supplements in calcium oxalate stone formers. Lemonade and other citrate-rich beverages have been used to treat hypocitraturia, although the increase in urine citrate is not as great as is seen with pharmacologic dosing of citrate salts.

Idiopathic Calcium Lithiasis

Some patients have no metabolic cause for stones despite a thorough metabolic evaluation (Table 287-1). The best treatment appears to be high fluid intake so that the urine specific gravity remains at ≤ 1.005 throughout the day and night. Thiazide diuretics and citrate therapy may help reduce crystallization of calcium salts, but there have been no prospective trials in this patient population. Oral phosphate at a dose of 2 g phosphorus daily may lower urine calcium and increase urine pyrophosphate, thereby reducing the rate of recurrence. Orthophosphate causes mild nausea and diarrhea, but tolerance may improve with continued intake.

URIC ACID STONES

Persistently acidic urine is the major risk factor for uric acid stone formation. When urine pH is low, the protonated form of uric acid predominates and is soluble in urine at concentrations of 100 mg/L. Concentrations above this level represent supersaturation that causes crystals and stones to form. Common causes of acidic urine and uric acid stones include metabolic syndrome, chronic diarrheal states, gout, and idiopathic uric acid lithiasis. As the prevalence of obesity increases, metabolic syndrome is becoming an increasingly important cause of uric acid stone formation, as insulin resistance leads to a decrease in ammoniagenesis, requiring that the metabolic acid load be excreted as titratable acid. Hyperuricosuria, when present, increases supersaturation, but urine of low pH can be supersaturated with uric acid even though the daily excretion rate is normal. Myeloproliferative syndromes, chemotherapy for malignant tumors, and Lesch-Nyhan syndrome cause such massive production of uric acid and consequent hyperuricosuria that stones and uric acid sludge form even at a normal urine pH. Obstruction of the renal tubules by uric acid crystals can cause acute renal failure.

Treatment: Uric Acid Lithiasis

The two goals of treatment are to raise urine pH and lower excessive urine uric acid excretion to <1 g/d. Supplemental alkali, 1–3 meq/kg of body weight per day, should be given in three or four divided doses, one of which should be given at bedtime. The goal of treatment should be a urine pH between 6 and 6.5 in a 24-h urine collection. Increasing urine pH above 6.5 will not provide additional benefit in preventing uric acid crystallization but increases the risk of calcium phosphate stone formation. The form of the alkali may be important. Potassium citrate may reduce the risk of calcium salts crystallizing when urine pH is increased, whereas sodium alkali salts may increase the risk. A low-purine diet should be instituted in uric acid stone formers with hyperuricosuria. Patients who continue to form uric acid stones despite treatment with fluids, alkali, and a low-purine diet should have allopurinol added to their regimen.

CYSTINURIA AND CYSTINE STONES

(See also Chap. 364) In this inherited disorder, proximal tubular and jejunal transport of the dibasic amino acids cystine, lysine, arginine, and ornithine is defective, and excessive amounts are lost in the urine. Clinical disease is due solely to the insolubility of cystine. Cystine crystals plug terminal collecting ducts, and stones may grow as an extension of those plugs. Damage to the papillae and medulla from crystal obstruction is the probable reason why kidney function is reduced in cystinuria compared with routine stone disease.

Pathogenesis

Cystinuria occurs because of defective transport of dibasic amino acids by the brush borders of renal tubule and intestinal epithelial cells. Disease-causing mutations have been identified in both the heavy and light chains of a heteromeric amino acid transporter found in the proximal tubule of the kidney. Cystinuria is classified into two main types, based on the urinary excretion of cystine in obligate heterozygotes. In type I cystinuria, heterozygotes have normal urine cystine excretion; thus, type I has

an autosomal recessive pattern of inheritance. A gene on chromosome 2 that has been designated *SLC3A1* encodes the heavy chain of the transporter and has been found to be abnormal in type I. In non-type I cystinuria, heterozygotes have moderately elevated urine cystine excretion, with homozygotes having a much higher urine cystine excretion. Non-type I is inherited as a dominant trait with incomplete penetrance. Non-type I is due to mutations in the *SLC7A9* gene on chromosome 19, which encodes the light chain of the heteromeric transporter. In rare cases, mutations of the *SLC7A9* gene can lead to a type I phenotype.

Diagnosis

Cystine stones are formed only by patients with cystinuria, but 10% of stones in cystinuric patients do not contain cystine; therefore, every stone former should be screened for the disease. The sediment from a first morning urine specimen in many patients with homozygous cystinuria reveals typical hexagonal, platelike cystine crystals. Cystinuria can also be detected by using the urine sodium nitroprusside test. Because the test is sensitive, it is positive for cystinuria in many asymptomatic heterozygotes. A positive nitroprusside test or the finding of cystine crystals in the urine sediment should be evaluated by measurement of daily cystine excretion. Cystine stones seldom form in adults unless urine excretion is at least 300 mg/d.

Treatment: Cystinuria and Cystine Stones

High fluid intake, even at night, is the cornerstone of therapy. Daily urine volume should exceed 3 L. Raising urine pH with alkali is helpful provided that the urine pH exceeds 7.5. A low-salt diet (100 mmol/d) can reduce cystine excretion up to 40%. Because side effects are common, drugs such as penicillamine and tiopronin, which form mixed soluble disulfide cysteine-drug complexes, should be used only when fluid loading, salt reduction, and alkali therapy are ineffective. Low-methionine diets have not proved to be practical for clinical use, but patients should avoid protein gluttony.

STRUVITE STONES

These stones are a result of urinary infection with bacteria, usually *Proteus* species, which possess urease, an enzyme that degrades urea to NH_3 and CO_2. The NH_3 hydrolyzes to NH_4^+ and raises urine pH to 8 or 9. The NH_4^+ precipitates PO_4^{3-} and Mg^{2+} to form $MgNH_4PO_4$ (struvite). Struvite does not form in urine in the absence of infection, because NH_4^+ concentration is low in urine that is alkaline in response to physiologic stimuli. Chronic *Proteus* infection can occur because of impaired urinary drainage, urologic instrumentation or surgery, and especially with chronic antibiotic treatment, which can favor the dominance of *Proteus* in the urinary tract. The presence of struvite crystals in urine, rectangular prisms that resemble coffin lids, indicates infection with urease-producing organisms.

Treatment: Struvite Stones

Complete removal of the stone with subsequent sterilization of the urinary tract is the treatment of choice for patients who can tolerate the procedures. Percutaneous nephrolithotomy is the preferred surgical approach for most patients. At times, extracorporeal lithotripsy may be used in combination with a percutaneous approach. Open surgery is rarely required. Irrigation of the renal pelvis and calyces with hemiacidrin, a solution that dissolves struvite, can reduce recurrence after surgery. Stone-free rates of 50–90% have been reported after surgical intervention. Antimicrobial treatment is best reserved for dealing with acute infection and for maintenance of a sterile urine after surgery. Urine cultures and culture of stone fragments removed at surgery should guide the choice of antibiotic. For patients who are not candidates for surgical removal of a stone, acetohydroxamic acid, an inhibitor of urease, can be used. Unfortunately, acetohydroxamic acid has many side effects, such as headache, tremor, and thrombophlebitis, that limit its use.

FURTHER READINGS

Al-Ansari A et al: Efficacy of tamsulosin in the management of lower ureteral stones: A randomized double-blind placebo-controlled study of 100 patients. Urology 75:4, 2010[PMID: 20109697]

Asplin JR: Evaluation of the kidney stone patient. Semin Nephrol 28:99, 2008[PMID: 18359391]

Cameron MA, Sakhaee K: Uric acid nephrolithiasis. Urol Clin North Am 34:335, 2007[PMID: 17678984]

Miller NL et al: A formal test of the hypothesis that idiopathic calcium oxalate stones grow on Randall's plaque. BJU Int 103:966, 2009[PMID: 19021625]

Patel BN et al: Prevalence of hyperoxaluria after bariatric surgery. J Urol 181:161, 2009[PMID: 19013593]

Vezzoli G et al: R990G polymorphism of calcium-sensing receptor does produce a gain-of-function and predispose to primary hypercalciuria. Kidney Int 71:1155, 2007[PMID: 17332735]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

URINARY TRACT INFECTIONS, PYELONEPHRITIS, AND PROSTATITIS: INTRODUCTION

Urinary tract infection (UTI) is a common and painful human illness that, fortunately, is rapidly responsive to modern antibiotic therapy. In the preantibiotic era, UTI caused significant morbidity. Hippocrates, writing about a disease that appears to have been acute cystitis, said that the illness could last for a year before either resolving or worsening to involve the kidneys. When chemotherapeutic agents used to treat UTI were introduced in the early twentieth century, they were relatively ineffective, and persistence of infection after 3 weeks of therapy was common. Nitrofurantoin, which became available in the 1950s, was the first tolerable and effective agent for the treatment of UTI.

Since the most common manifestation of UTI is acute cystitis and since acute cystitis is far more prevalent among women than among men, most clinical research on UTI has involved women. Many studies have enrolled women from college campuses or large health maintenance organizations in the United States. Therefore, when reviewing the literature and recommendations concerning UTI, clinicians must consider whether the findings are applicable to their patient populations.

Definitions

UTI may be asymptomatic (subclinical infection) or symptomatic (disease). Thus, the term *UTI* encompasses a variety of clinical entities, including asymptomatic bacteriuria (ABU), cystitis, prostatitis, and pyelonephritis. The distinction between symptomatic UTI and ABU has major clinical implications. Both UTI and ABU connote the presence of bacteria in the urinary tract, usually accompanied by white blood cells and inflammatory cytokines in the urine. However, ABU occurs in the absence of symptoms attributable to the bacteria in the urinary tract and does not usually require treatment, while UTI has more typically been assumed to imply symptomatic disease that warrants antimicrobial therapy. Much of the literature concerning UTI, particularly catheter-associated infection, does not differentiate between UTI and ABU. In this chapter, the term *UTI* denotes symptomatic disease; *cystitis*, symptomatic infection of the bladder; and *pyelonephritis*, symptomatic infection of the kidneys. *Uncomplicated UTI* refers to acute cystitis or pyelonephritis in nonpregnant outpatient women without anatomic abnormalities or instrumentation of the urinary tract; *complicated UTI* is a catch-all term that encompasses all other types of UTI. *Recurrent UTI* is not necessarily complicated; individual episodes can be uncomplicated and treated as such. *Catheter-associated bacteriuria* can be either symptomatic (CAUTI) or asymptomatic.

Epidemiology and Risk Factors

Except among infants and the elderly, UTI occurs far more commonly in females than in males. During the neonatal period, the incidence of UTI is slightly higher among males than among females because male infants more commonly have congenital urinary tract anomalies. After 50 years of age, obstruction from prostatic hypertrophy becomes common in men, and the incidence of UTI is almost as high among men as among women. Between 1 year and ~50 years of age, UTI and recurrent UTI are predominantly diseases of females. The prevalence of ABU is ~5% among women between ages 20 and 40 and may be as high as 40–50% among elderly women and men.

As many as 50–80% of women in the general population acquire at least one UTI during their lifetime —uncomplicated cystitis in most cases. Recent use of a diaphragm with spermicide, frequent sexual intercourse, and a history of UTI are independent risk factors for acute cystitis. Cystitis is temporally related to recent sexual intercourse, with a sixtyfold increase in the relative odds of acute cystitis in the 48 h after intercourse. In healthy postmenopausal women, sexual activity, diabetes mellitus, and incontinence are risk factors for UTI.

Many factors predisposing women to cystitis also increase the risk of pyelonephritis. Factors independently associated with pyelonephritis in young healthy women include frequent sexual intercourse, a new sexual partner, a UTI in the previous 12 months, a maternal history of UTI, diabetes, and incontinence. The common risk factors for cystitis and pyelonephritis are not surprising given that pyelonephritis typically arises through the ascent of bacteria from the bladder to the upper urinary tract. However, pyelonephritis can occur without clear antecedent cystitis.

About 20–30% of women who have had one episode of UTI will have recurrent episodes. Early recurrence (within 2 weeks) is usually regarded as relapse rather than reinfection and may indicate the need to evaluate the patient for a sequestered focus. Intracellular pods of infecting organisms within the bladder epithelium have been demonstrated in animal models of UTI, but the importance of this phenomenon in humans is not yet clear. The rate of recurrence ranges from 0.3 to 7.6 infections per patient per year, with an average of 2.6 infections per year. It is not uncommon for multiple recurrences to follow an initial infection, resulting in clustering of episodes. Clustering may be related temporally to the presence of a new risk factor or to the sloughing of the protective outer bladder epithelial layer in response to bacterial attachment during acute cystitis. The likelihood of a recurrence decreases with increasing time since the last infection. A case-control study of predominantly white premenopausal women with recurrent UTI identified frequent sexual intercourse, use of spermicide, a new sexual partner, a first UTI before 15 years of age, and a maternal history of UTI as independent risk factors for recurrent UTI.

The only consistently documented behavioral risk factors for recurrent UTI include frequent sexual intercourse and spermicide use. In postmenopausal women, anatomic factors affecting bladder emptying, such as cystoceles,

urinary incontinence, and residual urine, are most strongly associated with recurrent UTI.

In pregnant women, ABU has clinical consequences, and both screening for and treatment of this condition are indicated. Specifically, ABU during pregnancy is associated with preterm birth and perinatal mortality for the fetus and with pyelonephritis for the mother. A Cochrane meta-analysis found that treatment of ABU in pregnant women decreased the risk of pyelonephritis by 75%.

The majority of men with UTI have a functional or anatomic abnormality of the urinary tract, most commonly urinary obstruction secondary to prostatic hypertrophy. That said, not all men with UTI have detectable urinary abnormalities; this point is particularly relevant for men ≤45 years of age. Lack of circumcision is also associated with an increased risk of UTI, because *Escherichia coli* is more likely to colonize the glans and prepuce and subsequently migrate into the urinary tract.

Women—but not men—with diabetes have a two- to threefold higher rate of ABU and UTI than women without diabetes. Increased duration of diabetes and the use of insulin rather than oral medication are also associated with a higher risk of UTI among women with diabetes. Poor bladder function, obstruction in urinary flow, and incomplete voiding are additional factors commonly found in patients with diabetes that increase the risk of UTI. Impaired cytokine secretion may contribute to ABU in diabetic women.

Etiology

The uropathogens causing UTI vary by clinical syndrome but are usually enteric gram-negative rods that have migrated to the urinary tract. The susceptibility patterns of these organisms vary by clinical syndrome and by geography. In acute uncomplicated cystitis in the United States, the etiologic agents are highly predictable: *E. coli* accounts for 75–90% of isolates; *Staphylococcus saprophyticus* for 5–15% (with particularly frequent isolation from younger women); and *Klebsiella* species, *Proteus* species, *Enterococcus* species, *Citrobacter species*, and other organisms for 5–10%. Similar etiologic agents are found in Europe and Brazil. The spectrum of agents causing uncomplicated pyelonephritis is similar, with *E. coli* predominating. In complicated UTI (e.g., CAUTI), *E. coli* remains the predominant organism, but other aerobic gram-negative rods, such as *Klebsiella* species, *Proteus* species, *Citrobacter* species, *Acinetobacter* species, *Morganella* species, and *Pseudomonas aeruginosa*, also are frequently isolated. Gram-positive bacteria (e.g., enterococci and *Staphylococcus aureus*), and yeasts are also important pathogens in complicated UTI. Data on etiology and resistance are generally obtained from laboratory surveys and should be understood in the context that organism identification is performed only in cases in which urine is sent for culture—i.e., typically when complicated UTI or pyelonephritis is suspected. The available data demonstrate a worldwide increase in the resistance of *E. coli* to antibiotics commonly used to treat UTI. North American and European surveys of *E. coli* isolates from women with acute cystitis have documented rates of resistance to trimethoprim-sulfamethoxazole (TMP-SMX) greater than 20% and rates of resistance to ciprofloxacin between 5% and 10% in some regions. Since resistance rates vary by local geographic region, with individual patient characteristics, and over time, it is important to use current and local data when choosing a treatment regimen.

Pathogenesis

The urinary tract can be viewed as an anatomic unit united by a continuous column of urine extending from the urethra to the kidneys. In the majority of UTIs, bacteria establish infection by ascending from the urethra to the bladder. Continuing ascent up the ureter to the kidney is the pathway for most renal parenchymal infections. However, introduction of bacteria into the bladder does not inevitably lead to sustained and symptomatic infection. The interplay of host, pathogen, and environmental factors determines whether tissue invasion and symptomatic infection will ensue **(Fig. 288-1)**. For example, bacteria often enter the bladder after sexual intercourse, but normal voiding and innate host defense mechanisms in the bladder eliminate these organisms. Any foreign body in the urinary tract, such as a urinary catheter or stone, provides an inert surface for bacterial colonization. Abnormal micturition and/or significant residual urine volume promotes true infection. In the simplest of terms, anything that increases the likelihood of bacteria entering the bladder and staying there increases the risk of UTI.

Figure 288-1

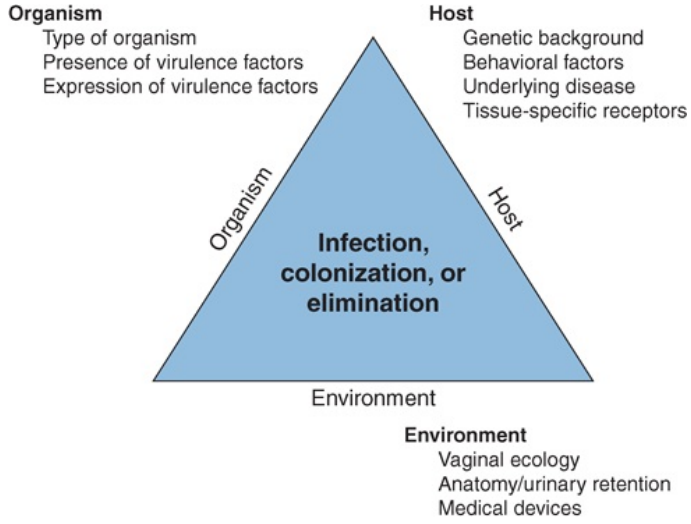

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis of urinary tract infection. The relationship between specific host, pathogen, and environmental factors determines the clinical outcome.

Bacteria can also gain access to the urinary tract through the bloodstream. However, hematogenous spread accounts for <2% of documented UTIs and usually results from bacteremia caused by relatively virulent organisms, such as *Salmonella* and *S. aureus*. Indeed, the isolation of either of these pathogens from a patient without a catheter or other instrumentation warrants a search for a bloodstream source. Hematogenous infections may produce focal abscesses or areas of pyelonephritis within a kidney and result in positive urine cultures. The pathogenesis of candiduria is distinct in that the hematogenous route is common. The presence of *Candida* in the urine of a noninstrumented immunocompetent patient implies either genital contamination or potentially widespread visceral dissemination.

ENVIRONMENTAL FACTORS

Vaginal Ecology

In women, vaginal ecology is an important environmental factor affecting the risk of UTI. Colonization of the vaginal introitus and perirurethral area with organisms from the intestinal flora (usually *E. coli*) is the critical initial step in the pathogenesis of UTI. Sexual intercourse is associated with an increased risk of vaginal colonization with *E. coli* and thereby increases the risk of UTI. Nonoxynol-9 in spermicide is toxic to the normal vaginal microflora and thus is likewise associated with an increased risk of *E. coli* vaginal colonization and bacteriuria. In postmenopausal women, the previously predominant vaginal lactobacilli are replaced with gram-negative colonization. The use of topical estrogens to prevent UTI in postmenopausal women is controversial; given the side effects of systemic hormone replacement, oral estrogens should not be used to prevent UTI.

Anatomic and Functional Abnormalities

Any condition that permits urinary stasis or obstruction predisposes the individual to UTI. Foreign bodies such as stones or urinary catheters provide an inert surface for bacterial colonization and formation of a persistent biofilm. Thus, vesicoureteral reflux, ureteral obstruction secondary to prostatic hypertrophy, neurogenic bladder, and urinary diversion surgery create an environment favorable to UTI. In persons with such conditions, *E. coli* strains lacking typical urinary virulence factors are often the cause of infection. Inhibition of ureteral peristalsis and decreased ureteral tone leading to vesicoureteral reflux are important in the pathogenesis of pyelonephritis in pregnant women. Anatomic factors—specifically, the distance of the urethra from the anus—are considered to be the primary reason why UTI is predominantly an illness of young women rather than of young men.

HOST FACTORS

The genetic background of the host influences the individual's susceptibility to recurrent UTI, at least among women. A familial disposition to UTI and to pyelonephritis is well documented. Women with recurrent UTI are more likely to have had their first UTI before age 15 years and to have a maternal history of UTI. A component of the underlying pathogenesis of this familial predisposition to recurrent UTI may be persistent vaginal colonization with *E. coli*, even during asymptomatic periods. Vaginal and periurethral mucosal cells from women with recurrent UTI bind threefold more uropathogenic bacteria than do mucosal cells from women without recurrent infection. Epithelial cells from susceptible women may possess specific types or greater numbers of receptors to which *E. coli* can bind, thereby facilitating colonization and invasion. Mutations in host response genes (e.g., those coding for Toll-like receptors and the interleukin 8 receptor) have also been linked to recurrent UTI and pyelonephritis. Polymorphisms in the interleukin 8–specific receptor gene CXCR1 are associated with increased susceptibility to pyelonephritis. Lower-level expression of CXCR1 on the surface of neutrophils impairs neutrophil-dependent host defense against bacterial invasion of the renal parenchyma.

MICROBIAL FACTORS

An anatomically normal urinary tract presents a stronger barrier to infection than a compromised urinary tract. Thus, strains of *E. coli* that cause invasive symptomatic infection of the urinary tract in otherwise normal hosts often possess and express genetic virulence factors, including surface adhesins that mediate binding to specific receptors on the surface of uroepithelial cells. The best-studied adhesins are the P fimbriae, hairlike protein

structures that interact with a specific receptor on renal epithelial cells. (The letter *P* denotes the ability of these fimbriae to bind to blood group antigen P, which contains a d-galactose-d-galactose residue.) P fimbriae are important in the pathogenesis of pyelonephritis and subsequent bloodstream invasion from the kidney.

Another adhesin is the type 1 pilus (fimbria), which all *E. coli* strains possess but not all *E. coli* strains express. Type 1 pili are thought to play a key role in initiating *E. coli* bladder infection; they mediate binding to uroplakins on the luminal surface of bladder uroepithelial cells. The binding of type 1 fimbriae of *E. coli* to receptors on uroepithelial cells initiates a complex series of signaling events that leads to apoptosis and exfoliation of uroepithelial cells, with the attached *E. coli* organisms carried away in the urine.

Approach to the Patient: Clinical Manifestations

The most important issue to be addressed when a UTI is suspected is the characterization of the clinical syndrome as ABU, uncomplicated cystitis, pyelonephritis, prostatitis, or complicated UTI. This information will shape the diagnostic and therapeutic approach.

ASYMPTOMATIC BACTERIURIA

A diagnosis of ABU can be considered only when the patient does not have local or systemic symptoms referable to the urinary tract. The clinical presentation is usually that of a patient who undergoes a screening urine culture for a reason unrelated to the genitourinary tract and is incidentally found to have bacteriuria. The presence of systemic signs or symptoms such as fever, altered mental status, and leukocytosis in the setting of a positive urine culture does not merit a diagnosis of symptomatic UTI unless other potential etiologies have been considered.

CYSTITIS

The typical symptoms of cystitis are dysuria, urinary frequency, and urgency. Nocturia, hesitancy, suprapubic discomfort, and gross hematuria are often noted as well. Unilateral back or flank pain is generally an indication that the upper urinary tract is involved. Fever is also an indication of invasive infection of either the kidney or the prostate.

PYELONEPHRITIS

Mild pyelonephritis can present as low-grade fever with or without lower-back or costovertebral-angle pain, whereas severe pyelonephritis can manifest as high fever, rigors, nausea, vomiting, and flank and/or loin pain. Symptoms are generally acute in onset, and symptoms of cystitis may not be present. Fever is the main feature distinguishing cystitis and pyelonephritis. The fever of pyelonephritis typically exhibits a high, spiking "picket-fence" pattern and resolves over 72 h of therapy. Bacteremia develops in 20–30% of cases of pyelonephritis. Patients with diabetes may present with obstructive uropathy associated with acute papillary necrosis when the sloughed papillae obstruct the ureter. Papillary necrosis may also be evident in some cases of pyelonephritis complicated by obstruction, sickle cell disease, analgesic nephropathy, or combinations of these conditions. In the rare cases of bilateral papillary necrosis, a rapid rise in the serum creatinine level may be the first indication of the condition. *Emphysematous* pyelonephritis is a particularly severe form of the disease that is associated with the production of gas in renal and perinephric tissues and occurs almost exclusively in diabetic patients **(Fig. 288-2)**. *Xanthogranulomatous* pyelonephritis occurs when chronic urinary obstruction (often by staghorn calculi), together with chronic infection, leads to suppurative destruction of renal tissue **(Fig. 288-3)**. On pathologic examination, the residual renal tissue frequently has a yellow coloration with infiltration by lipid-laden macrophages. Pyelonephritis can also be complicated by intraparenchymal abscess formation; this situation should be suspected when a patient has continued fever and/or bacteremia despite antibacterial therapy.

Figure 288-2

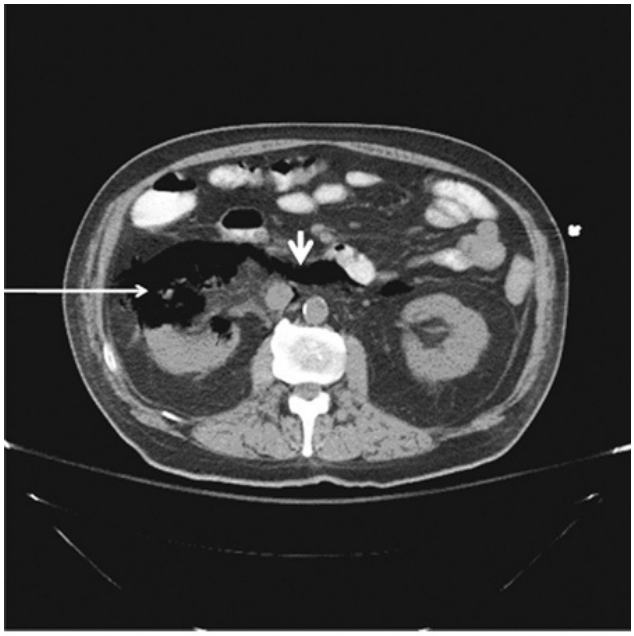

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Emphysematous pyelonephritis. Infection of the right kidney of a diabetic man by *Escherichia coli*, a gas-forming, facultative anaerobic uropathogen, has led to destruction of the renal parenchyma (*arrow*) and tracking of gas through the retroperitoneal space (*arrowhead*).

Figure 288-3

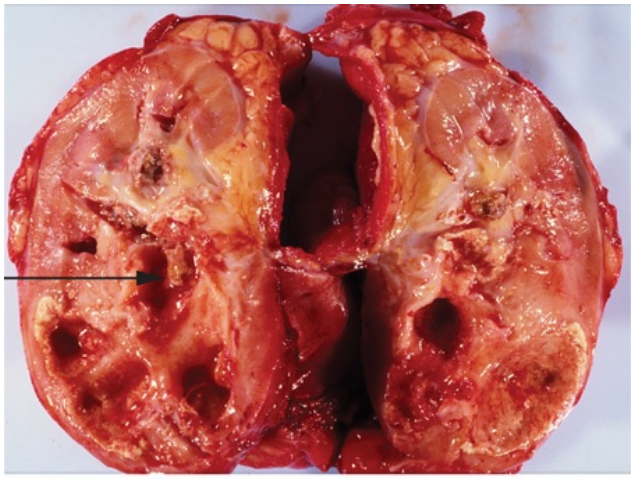

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

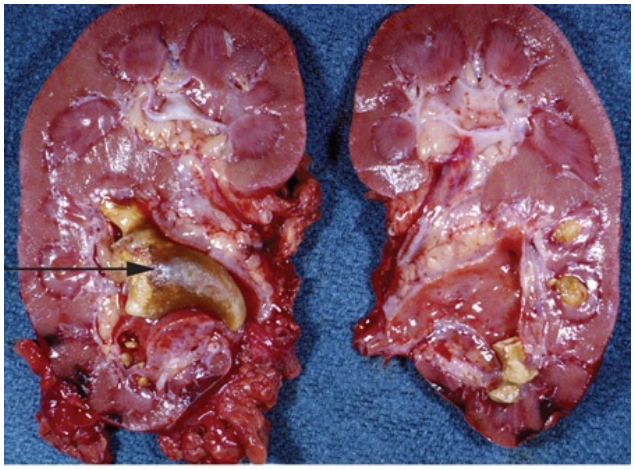

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Xanthogranulomatous pyelonephritis.*A.* This photograph shows extensive destruction of renal parenchyma due to long-standing suppurative inflammation. The precipitating factor was obstruction by a staghorn calculus, which has been removed, leaving a depression (*arrow*). The mass effect of xanthogranulomatous pyelonephritis can mimic renal malignancy. ***B.*** A large staghorn calculus (*arrow*) is seen obstructing the renal pelvis and calyceal system. The lower pole of the kidney shows areas of hemorrhage and necrosis with collapse of cortical areas. *(Both images: Courtesy of Dharam M. Ramnani, MD, Virginia Urology Pathology Laboratory, Richmond, VA.)*

PROSTATITIS

Prostatitis includes both infectious and noninfectious abnormalities of the prostate gland. Infections can be acute or chronic, are almost always bacterial in nature, and are far less common than the noninfectious entity of chronic pelvic pain syndrome (formerly known as chronic prostatitis). Acute bacterial prostatitis presents as dysuria, frequency, and pain in the prostatic, pelvic, or perineal area. Fever and chills are usually present, and symptoms of bladder outlet obstruction are common. Chronic bacterial prostatitis presents more insidiously as recurrent episodes of cystitis, sometimes with associated pelvic and perineal pain. Men who present with recurrent cystitis should be evaluated for a prostatic focus.

COMPLICATED UTI

Complicated UTI presents as a symptomatic episode of cystitis or pyelonephritis in a man or woman with an anatomic predisposition to infection, with a foreign body in the urinary tract, or with factors predisposing to a delayed response to therapy.

Diagnostic Tools

HISTORY

The diagnosis of any the UTI syndromes or ABU begins with a detailed history **(Fig. 288-4)**. The history given by the patient has a high predictive value in uncomplicated cystitis. A meta-analysis evaluating the probability of acute UTI on the basis of history and physical findings concluded that, in women presenting with at least one symptom of UTI (dysuria, frequency, hematuria, or back pain) and without complicating factors, the probability of acute cystitis or pyelonephritis is 50%. The even higher rates of accuracy of self-diagnosis among women with recurrent UTI probably account for the success of patient-initiated treatment of recurrent cystitis. If vaginal discharge and complicating factors are absent and risk factors for UTI are present, then the probability of UTI is close to 90%, and no laboratory evaluation is needed. Similarly, a combination of dysuria and urinary frequency in the absence of vaginal discharge increases the probability of UTI to 96%. Further laboratory evaluation with dipstick testing or urine culture is not necessary in such patients before the initiation of definitive therapy.

Figure 288-4

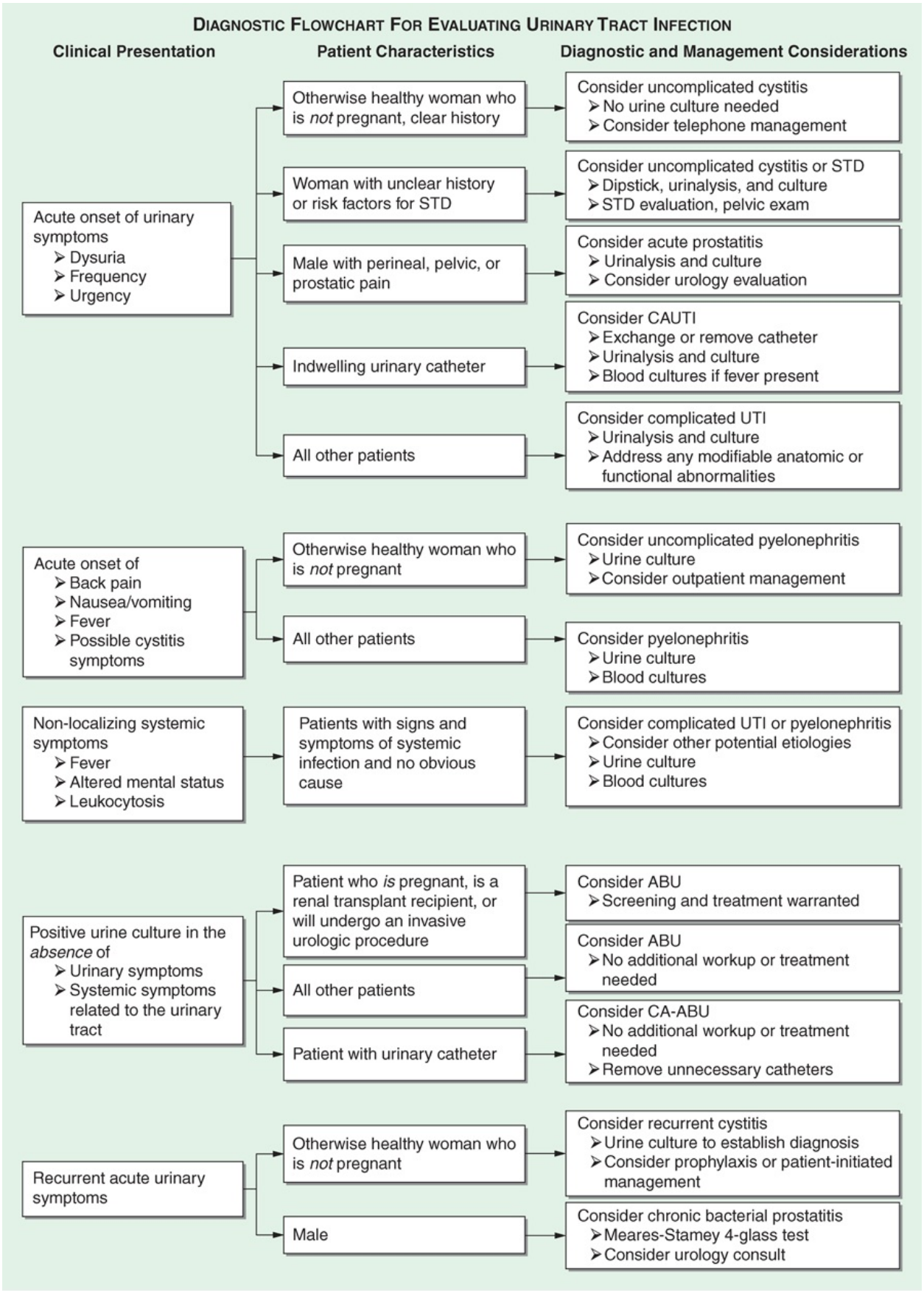

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagnostic approach to urinary tract infection. STD, sexually transmitted disease; CAUTI, catheter-associated UTI; ABU, asymptomatic bacteriuria; CA-ABU, catheter-associated ABU.

When the patient's history is applied as a diagnostic tool, it is important to recall that the studies included in the meta-analysis cited above did not enroll children, adolescents, pregnant women, men, or patients with

complicated UTI. One significant concern is that sexually transmitted disease—that caused by *Chlamydia trachomatis* in particular—may be inappropriately treated as UTI. This concern is particularly relevant for female patients under the age of 25. The differential diagnosis to be considered when women present with dysuria includes cervicitis (*C. trachomatis*, *Neisseria gonorrhoeae*), vaginitis (*Candida albicans*, *Trichomonas vaginalis*), herpetic urethritis, interstitial cystitis, and noninfectious vaginal or vulvar irritation. Women with more than one sexual partner and inconsistent use of condoms are at high risk for both UTI and sexually transmitted disease, and symptoms alone do not always distinguish between these conditions.

THE URINE DIPSTICK TEST, URINALYSIS, AND URINE CULTURE

Useful diagnostic tools include the urine dipstick test and urinalysis, both of which provide point-of-care information, and the urine culture, which can retrospectively confirm a prior diagnosis. Understanding the parameters of the dipstick test is important in interpreting its results. Only members of the family Enterobacteriaceae convert nitrate to nitrite, and enough nitrite must accumulate in the urine to reach the threshold of detection. If a woman with acute cystitis is forcing fluids and voiding frequently, the dipstick test for nitrite is less likely to be positive, even when *E. coli* is present. The leukocyte esterase test detects this enzyme in the host's polymorphonuclear leukocytes in the urine, whether the cells are intact or lysed. Many reviews have attempted to describe the diagnostic accuracy of dipstick testing. The bottom line for clinicians is that a urine dipstick test can confirm the diagnosis of uncomplicated cystitis in a patient with a reasonably high pretest probability of this disease. Either nitrite or leukocyte esterase positivity can be interpreted as a positive result. Blood in the urine may also suggest a diagnosis of UTI. A dipstick test negative for both nitrite and leukocyte esterase in the same type of patient should prompt consideration of other explanations for the patient's symptoms and collection of urine for culture. A negative dipstick test is not sufficiently sensitive to rule out bacteriuria in pregnant women, in whom it is important to detect all episodes of bacteriuria. Performance characteristics of the dipstick test differ in men (highly specific) and in noncatheterized nursing home residents (highly sensitive).

Urine microscopy reveals pyuria in nearly all cases of cystitis and hematuria in ~30% of cases. In current practice, most hospital laboratories use an automated system rather than manual examination for urine microscopy. A machine aspirates a sample of the urine and then classifies the particles in the urine by size, shape, contrast, light scatter, volume, and other properties. These automated systems can be overwhelmed by high numbers of dysmorphic red blood cells, white blood cells, or crystals; in general, counts of bacteria are less accurate than are counts of red and white blood cells. Our clinical recommendation is that the patient's symptoms and presentation should outweigh an incongruent result on automated urinalysis.

The detection of bacteria in a urine culture is the diagnostic "gold standard" for UTI; unfortunately, however, culture results do not become available until 24 h after the patient's presentation. Identifying specific organism(s) can require an additional 24 h. Studies of women with symptoms of cystitis have found that a colony count threshold of $>10^2$ bacteria/mL is more sensitive (95%) and specific (85%) than a threshold of 10^5/mL for the diagnosis of acute cystitis in women. In men, the minimal level indicating infection appears to be 10^3/mL. Urine specimens frequently become contaminated with the normal microbial flora of the distal urethra, vagina, or skin. These contaminants can grow to high numbers if the collected urine is allowed to stand at room temperature. In most instances, a culture that yields mixed bacterial species is contaminated except in settings of long-term catheterization, chronic urinary retention, or the presence of a fistula between the urinary tract and the gastrointestinal or genital tract.

Diagnosis

The approach to diagnosis is influenced by which of the clinical UTI syndromes is suspected (Fig. 288-4).

UNCOMPLICATED CYSTITIS IN WOMEN

Uncomplicated cystitis in women can be treated on the basis of history alone. However, if the symptoms are not specific or if a reliable history cannot be obtained, then a urine dipstick test should be performed. A positive nitrite or leukocyte esterase result in a woman with one symptom of UTI increases the probability of UTI from 50% to ~80%, and empirical treatment can be considered without further testing. In this setting, a negative dipstick result does not rule out UTI, and a urine culture, close clinical follow-up, and possibly a pelvic examination are recommended. These recommendations are made with the caveat that factors associated with complicated UTI, such as pregnancy, are not present.

CYSTITIS IN MEN

The signs and symptoms of cystitis in men are similar to those in women, but this disease differs in several important ways in the male population. Collection of urine for culture is strongly recommended when a man has symptoms of UTI, as the documentation of bacteriuria can differentiate the less common syndromes of acute and chronic bacterial prostatitis from the very common entity of chronic pelvic pain syndrome, which is not associated with bacteriuria and thus is not usually responsive to antibacterial therapy. If the diagnosis is unclear, localization cultures using the two- or four-glass Meares-Stamey test (urine collection after prostate massage) should be undertaken to differentiate between bacterial and nonbacterial prostatic syndromes, and the patient should be referred to a urologist. Men with febrile UTI often have an elevated serum level of prostate-specific antigen as well as an enlarged prostate and enlarged seminal vesicles on ultrasound—findings indicative of prostate involvement. In 85 men with febrile UTI, symptoms of urinary retention, early recurrence of UTI, hematuria at follow-up, and voiding difficulties were predictive of surgically correctable disorders. Men with none of these symptoms had normal upper and lower urinary tracts on urologic workup.

ASYMPTOMATIC BACTERIURIA

The diagnosis of ABU involves both microbiologic and clinical criteria. The microbiologic criterion is usually $\geq 10^5$ bacterial cfu/mL except in catheter-associated disease, in which case $\geq 10^2$ cfu/mL is the cutoff. The clinical criterion is that the person has no signs or symptoms referable to UTI.

Treatment: Urinary Tract Infections

Antimicrobial therapy is warranted for any symptomatic UTI. The choice of antimicrobial agent and the dose and duration of therapy depend on the site of infection and the presence or absence of complicating conditions. Each category of UTI warrants a different approach based on the particular clinical syndrome.

UNCOMPLICATED CYSTITIS IN WOMEN

Since the species and antimicrobial susceptibilities of the bacteria that cause acute uncomplicated cystitis are highly predictable, many episodes of uncomplicated cystitis can be managed over the telephone (Fig. 288-4). Most patients with other UTI syndromes require further diagnostic evaluation. Although the risk of serious complications with telephone management appears to be low, studies of telephone management algorithms generally have involved otherwise healthy white women who are at low risk for complications of UTI.

In 1999, TMP-SMX was recommended as the first-line agent for treatment of uncomplicated UTI in the published guidelines of the Infectious Diseases Society of America. Antibiotic resistance among uropathogens causing uncomplicated cystitis has since increased, appreciation of the importance of collateral damage (as defined below) has increased, and newer agents have been studied. Unfortunately, there is no longer a single best agent for acute uncomplicated cystitis.

Collateral damage refers to the adverse ecologic effects of antimicrobial therapy, including killing of the normal flora and selection of drug-resistant organisms. Outbreaks of *Clostridium difficile* infection offer an example of collateral damage in the hospital environment. The implication of collateral damage in this context is that a drug that is highly efficacious for the treatment of UTI is not necessarily the optimal first-line agent if it also has pronounced secondary effects on the normal flora or is likely to change resistance patterns. Drugs used for UTI that have a minimal effect on fecal flora include pivmecillinam, fosfomycin, and nitrofurantoin. In contrast, trimethoprim, TMP-SMX, quinolones, and ampicillin affect the fecal flora more significantly; these drugs are notably the agents for which rising resistance levels have been documented.

Several effective therapeutic regimens are available for acute uncomplicated cystitis in women **(Table 288-1)**. Well-studied first-line agents include TMP-SMX and nitrofurantoin. Second-line agents include fluoroquinolone and β-lactam compounds. Single-dose fosfomycin treatment for acute cystitis is widely used in Europe but has produced mixed results in randomized trials. Pivmecillinam is not currently available in the United States or Canada but is a popular agent in some European countries. The pros and cons of other therapies are discussed briefly below.

Table 288-1 Treatment Strategies for Acute Uncomplicated Cystitis

Drug and Dose	Estimated Clinical Efficacy (%)	Estimated Bacterial Efficacy (%)	Common Side Effects
Nitrofurantoin, 100 mg bid x 5–7 d	84–95	86–92	Nausea, headache
TMP-SMX, 1 DS tablet bid x 3 d	90–100	91–100	Rash, urticaria, nausea, vomiting, hematologic abnormalities
Fosfomycin, 3-g single-dose sachet	70–91	78–83	Diarrhea, nausea, headache
Pivmecillinam, 400 mg bid x 3–7 d	55–82	74–84	Nausea, vomiting, diarrhea
Fluoroquinolones, dose varies by agent; 3-d regimen	85–95	81–98	Nausea, vomiting, diarrhea, headache, drowsiness, insomnia
β-Lactams, dose varies by agent; 5- to 7-d regimen	79–98	74–98	Diarrhea, nausea, vomiting, rash, urticaria

Note: Efficacy rates are averages or ranges calculated from the data and studies included in the 2010 Infectious Diseases Society of America/European Society of Clinical Microbiology and Infectious Diseases Guideline for Treatment of Uncomplicated UTI. TMP-SMX, trimethoprim-sulfamethoxazole; DS, double-strength.

Traditionally, TMP-SMX has been recommended as first-line treatment for acute cystitis, and it remains appropriate to consider the use of this drug in regions with resistance rates not exceeding 20%. TMP-SMX resistance has clinical significance: in TMP-SMX-treated patients with resistant isolates, the time to symptom resolution is longer and rates of both clinical and microbiologic failure are higher. Individual host factors associated with an elevated risk of UTI caused by a strain of *E. coli* resistant to TMP-SMX include recent use of TMP-SMX or another antimicrobial agent and recent travel to an area with high rates of TMP-SMX resistance. The optimal setting for empirical use of TMP-SMX is uncomplicated UTI in a female patient who has an established relationship with the practitioner and who can thus seek further care if her symptoms do not respond promptly.

Resistance to nitrofurantoin remains low despite >60 years of use. Since this drug affects bacterial metabolism in multiple pathways, several mutational steps are required for the development of resistance. Nitrofurantoin remains highly active against *E. coli* and most non–*E. coli* isolates. *Proteus*, *Pseudomonas*, *Serratia*, *Enterobacter*, and yeasts are all intrinsically resistant to this drug. Although nitrofurantoin has traditionally been prescribed as a 7-day regimen, similar microbiologic and clinical efficacies are noted with a 5-day course of nitrofurantoin or a 3-day course of TMP-SMX for treatment of women with acute cystitis; 3-day courses of nitrofurantoin are not recommended for acute cystitis. Nitrofurantoin does not reach significant levels in tissue and cannot be used to treat pyelonephritis.

Most fluoroquinolones are highly effective for short-course therapy for cystitis; the exception is moxifloxacin, which

does not achieve adequate urinary levels. The fluoroquinolones commonly used for UTI include ofloxacin, ciprofloxacin, and levofloxacin. The main concern about fluoroquinolone use for acute cystitis is the propagation of fluoroquinolone resistance, not only among uropathogens but also among other organisms causing more serious and difficult-to-treat infections at other sites. Fluoroquinolone use is also a factor driving the emergence of *C. difficile* outbreaks in hospital settings. Most experts now call for restricting fluoroquinolones to specific instances of uncomplicated cystitis in which other antimicrobial agents are not suitable. Quinolone use in the elderly has been associated with an increased risk of Achilles tendon rupture.

Except for pivmecillinam, β-lactam agents generally have not performed as well as TMP-SMX or fluoroquinolones in acute cystitis. Rates of pathogen eradication are lower and relapse rates are higher with β-lactam drugs. The generally accepted explanation is that β-lactams fail to eradicate uropathogens from the vaginal reservoir. A proposed role for intracellular biofilm communities is intriguing. Many strains of *E. coli* that are resistant to TMP-SMX are also resistant to amoxicillin and cephalexin; thus, these drugs should be used only for patients infected with susceptible strains.

Urinary analgesics are appropriate in certain situations to speed resolution of bladder discomfort. The urinary tract analgesic phenazopyridine is widely used but can cause significant nausea. Combination analgesics containing urinary antiseptics (methenamine, methylene blue), a urine-acidifying agent (sodium phosphate), and an antispasmodic agent (hyoscyamine) are also available.

PYELONEPHRITIS

Since patients with pyelonephritis have tissue-invasive disease, the treatment regimen chosen should have a very high likelihood of eradicating the causative organism and should reach therapeutic blood levels quickly. High rates of TMP-SMX-resistant *E. coli* in patients with pyelonephritis have made fluoroquinolones the first-line therapy for acute uncomplicated pyelonephritis. Whether the fluoroquinolones are given orally or parenterally depends on the patient's tolerance for oral intake. A randomized clinical trial demonstrated that a 7-day course of therapy with oral ciprofloxacin (500 mg twice daily, with or without an initial IV 400-mg dose) was highly effective for the initial management of pyelonephritis in the outpatient setting. Oral TMP-SMX (one double-strength tablet twice daily for 14 days) is also effective for treatment of acute uncomplicated pyelonephritis if the uropathogen is known to be susceptible. If the pathogen's susceptibility is not known and TMP-SMX is used, an initial IV 1-g dose of ceftriaxone is recommended. Oral β-lactam agents are less effective than the fluoroquinolones and should be used with caution and close follow-up. Options for parenteral therapy for uncomplicated pyelonephritis include fluoroquinolones, an aminoglycoside with or without ampicillin, an extended-spectrum cephalosporin with or without an aminoglycoside, or a carbapenem. Combinations of a β-lactam and a β-lactamase inhibitor (e.g., ampicillin-sulbactam, ticarcillin-clavulanate, and piperacillin-tazobactam) or imipenem-cilastatin can be used in patients with more complicated histories, previous episodes of pyelonephritis, or recent urinary tract manipulations; in general, the treatment of such patients should be guided by urine culture results. Once the patient has responded clinically, oral therapy should be substituted for parenteral therapy.

UTI IN PREGNANT WOMEN

Nitrofurantoin, ampicillin, and the cephalosporins are considered relatively safe in early pregnancy. One retrospective case-control study suggesting an association between nitrofurantoin and birth defects awaits confirmation. Sulfonamides should clearly be avoided both in the first trimester (because of possible teratogenic effects) and near term (because of a possible role in the development of kernicterus). Fluoroquinolones are avoided because of possible adverse effects on fetal cartilage development. Ampicillin and the cephalosporins have been used extensively in pregnancy and are the drugs of choice for the treatment of asymptomatic or symptomatic UTI in this group of patients. For pregnant women with overt pyelonephritis, parenteral β-lactam therapy with or without aminoglycosides is the standard of care.

UTI IN MEN

Since the prostate is involved in the majority of cases of febrile UTI in men, the goal in these patients is to eradicate the prostatic infection as well as the bladder infection. In men with apparently uncomplicated UTI, a 7- to 14-day course of a fluoroquinolone or TMP-SMX is recommended. If acute bacterial prostatitis is suspected, antimicrobial therapy should be initiated after urine and blood are obtained for cultures. Therapy can be tailored to urine culture results and should be continued for 2–4 weeks. For documented chronic bacterial prostatitis, a 4- to 6-week course of antibiotics is often necessary. Recurrences, which are not uncommon in chronic prostatitis, often warrant a 12-week course of treatment.

COMPLICATED UTI

Complicated UTI (other than that discussed above) occurs in a heterogeneous group of patients with a wide variety of structural and functional abnormalities of the urinary tract and kidneys. The range of species and their susceptibility to antimicrobial agents are likewise heterogeneous. As a consequence, therapy for complicated UTI must be individualized and guided by urine culture results. Frequently, a patient with complicated UTI will have prior urine culture data that can be used to guide empirical therapy while current culture results are awaited. Xanthogranulomatous pyelonephritis is treated with nephrectomy. Percutaneous drainage can be used as the initial therapy in emphysematous pyelonephritis and can be followed by elective nephrectomy as needed. Papillary necrosis with obstruction requires intervention to relieve the obstruction and to preserve renal function.

ASYMPTOMATIC BACTERIURIA

Treatment of ABU does not decrease the frequency of symptomatic infections or complications except in pregnant women, persons undergoing urologic surgery, and perhaps neutropenic patients and renal transplant recipients. Treatment of ABU in pregnant women and patients undergoing urologic procedures should be directed by urine culture results. In all other populations, screening for and treatment of ABU are discouraged. The majority of cases of catheter-associated bacteriuria are asymptomatic and do not warrant antimicrobial therapy.

CATHETER-ASSOCIATED UTI

Multiple institutions have released guidelines for the treatment of CAUTI, which is defined by bacteriuria and symptoms in a catheterized patient. The signs and symptoms either are localized to the urinary tract or can include otherwise unexplained systemic manifestations, such as fever. The accepted threshold for bacteriuria varies from $\geq 10^3$ cfu/mL to $\geq 10^5$ cfu/mL.

The formation of biofilm—a living layer of uropathogens—on the urinary catheter is central to the pathogenesis of CAUTI and affects both therapeutic and preventive strategies. Organisms in a biofilm are relatively resistant to killing by antibiotics, and eradication of a catheter-associated biofilm is difficult without removal of the device itself. Furthermore, because catheters provide a conduit for bacteria to enter the bladder, bacteriuria is inevitable with long-term catheter use.

The typical signs and symptoms of UTI, including pain, urgency, dysuria, fever, peripheral leukocytosis, and pyuria, have less predictive value for the diagnosis of infection in catheterized patients. Furthermore, the presence of bacteria in the urine of a patient who is febrile and catheterized does not necessarily predict CAUTI, and other explanations for the fever should be considered.

The etiology of CAUTI is diverse, and urine culture results are essential to guide treatment. Fairly good evidence supports the practice of catheter change during treatment for CAUTI. The goal is to remove biofilm-associated organisms that could serve as a nidus for reinfection. Pathology studies reveal that many patients with long-term catheters have occult pyelonephritis. A randomized trial in persons with spinal cord injury who were practicing intermittent catheterization found that relapse was more common after 3 days of therapy than after 14 days. In general, a 7- to 14-day course of antibiotics is recommended, but further studies on the optimal duration of therapy are needed.

In the setting of long-term catheter use, systemic antibiotics, bladder-acidifying agents, antimicrobial bladder washes, topical disinfectants, and antimicrobial drainage-bag solutions have all been ineffective at preventing the onset of bacteriuria and have been associated with the emergence of resistant organisms. The best strategy for prevention of CAUTI is to avoid insertion of unnecessary catheters and to remove catheters once they are no longer necessary. Evidence is insufficient to recommend suprapubic catheters and condom catheters as alternatives to indwelling urinary catheters as a means to prevent CAUTI. However, intermittent catheterization may be preferable to long-term indwelling urethral catheterization in certain populations (e.g., spinal cord–injured persons) to prevent both infectious and anatomic complications. Antimicrobial catheters impregnated with silver or nitrofurazone have not been shown to provide significant clinical benefit in terms of reducing rates of symptomatic UTI.

CANDIDURIA

The appearance of *Candida* in the urine is an increasingly common complication of indwelling catheterization, particularly for patients in the intensive care unit, those taking broad-spectrum antimicrobial drugs, and those with underlying diabetes mellitus. *C. albicans* is still the most common isolate, although *C. glabrata* and other non-*albicans* species are also isolated frequently. The clinical presentation varies from an asymptomatic laboratory finding to pyelonephritis and even sepsis. In asymptomatic patients, removal of the urethral catheter results in resolution of candiduria in more than one-third of cases. Treatment is recommended for patients who have symptomatic cystitis or pyelonephritis and for those who are at high risk for disseminated disease. High-risk patients include those with neutropenia, those who are undergoing urologic manipulation, and low-birth-weight infants. Fluconazole (200–400 mg/d for 14 days) achieves high levels in urine and is the first-line regimen for *Candida* infections of the urinary tract. The newer azoles and echinocandins are characterized by only low-level urinary excretion and thus are not recommended, although cases of successful eradication of candiduria with some of these agents have been reported. For *Candida* isolates with high levels of resistance to fluconazole, oral flucytosine and/or parenteral amphotericin B are options. Bladder irrigation with amphotericin B generally is not recommended.

Prevention of Recurrent UTI in Women

Recurrence of uncomplicated cystitis in reproductive-age women is common, and a preventive strategy is indicated if recurrent UTIs are interfering with a patient's lifestyle. The threshold of two or more symptomatic episodes per year is not absolute; decisions about interventions should take the patient's preferences into account.

Three prophylactic strategies are available: continuous, postcoital, or patient-initiated therapy. Continuous prophylaxis and postcoital prophylaxis usually entail low doses of TMP-SMX, a fluoroquinolone, or nitrofurantoin. These regimens are all highly effective during the period of active antibiotic intake. Typically, a prophylactic regimen is prescribed for 6 months and then discontinued, at which point the rate of recurrent UTI often returns to baseline. If bothersome infections recur, the prophylactic program can be reinstituted for a longer period.

Patient-initiated therapy involves supplying the patient with materials for urine culture and self-medication with a course of antibiotics at the first symptoms of infection. The urine culture is refrigerated and delivered to the physician's office for confirmation of the diagnosis. When an established and reliable patient-provider relationship exists, the urine culture can be omitted as long as the symptomatic episodes respond completely to short-course therapy and are not followed by relapse.

Prognosis

Cystitis is a risk factor for recurrent cystitis and pyelonephritis. ABU is common among elderly and catheterized patients but does not in itself increase the risk of death. The relationships among recurrent UTI, chronic pyelonephritis, and renal insufficiency have been widely studied. In the absence of anatomic abnormalities, recurrent infection in children and adults does not lead to chronic pyelonephritis or to renal failure. Moreover, infection does not play a primary role in chronic interstitial nephritis; the primary etiologic factors in this condition are analgesic abuse, obstruction, reflux, and toxin exposure. In the presence of underlying renal abnormalities (particularly obstructing stones), infection as a secondary factor can accelerate renal parenchymal damage. In

spinal cord–injured patients, use of a long-term indwelling bladder catheter is a well-documented risk factor for bladder cancer. Chronic bacteriuria resulting in chronic inflammation is one possible explanation for this observation.

FURTHER READINGS

Bent S et al: Does this woman have acute uncomplicated urinary tract infection? JAMA 287:2701, 2002[PMID: 12020306]

Fihn SD: Clinical practice. Acute uncomplicated urinary tract infection in women. N Engl J Med 349:259, 2003[PMID: 12867610]

Gupta K et al: International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: A 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. Clin Infect Dis 52:e103, 2011

Gupta K et al: Patient-initiated treatment of uncomplicated recurrent urinary tract infections in young women. Ann Intern Med 135:9, 2001[PMID: 11434727]

Hooton TM et al: Diagnosis, prevention, and treatment of catheter-associated urinary tract infection in adults: 2009 international clinical practice guidelines from the Infectious Diseases Society of America. Clin Infect Dis 50:625, 2010[PMID: 20175247]

Johnson JR et al: Systematic review: Antimicrobial urinary catheters to prevent catheter-associated urinary tract infection in hospitalized patients. Ann Intern Med 144:116, 2006[PMID: 16418411]

Le BV, Schaeffer AJ: Genitourinary pain syndromes, prostatitis, and lower urinary tract symptoms. Urol Clin North Am 36:527, 2009[PMID: 19942050]

Nicolle LE et al: Infectious Diseases Society of America guidelines for the diagnosis and treatment of asymptomatic bacteriuria in adults. Clin Infect Dis 40:643, 2005[PMID: 15714408]

Saint S et al: The effectiveness of a clinical practice guideline for the management of presumed uncomplicated urinary tract infection in women. Am J Med 106:636, 1999[PMID: 10378621]

Trautner BW: Management of catheter-associated urinary tract infection. Curr Opin Infect Dis 23:76, 2010[PMID: 19926986]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTERSTITIAL CYSTITIS/PAINFUL BLADDER SYNDROME: INTRODUCTION

Most clinicians with outpatient practices see undiagnosed cases of interstitial cystitis/painful bladder syndrome (IC/PBS). This chronic condition is characterized by pain perceived to be from the urinary bladder, urinary urgency and frequency, and nocturia. As currently diagnosed, the great majority of cases occur in women. Symptoms wax and wane for months or years or possibly even for the patient's lifetime. The spectrum of symptom intensity is broad. The pain can be excruciating, urgency can be distressing, frequency can be up to 60 times per 24 h, and nocturia can be causative of sleep deprivation. These symptoms can be disabling in terms of daily activities, work schedules, and personal relationships; patients with IC/PBS report less life satisfaction than do those with end-stage renal disease. The etiology of IC/PBS is unknown. It is not a new disease, having first been described in the late nineteenth century in a patient with the symptoms described above and a single ulcer visible on cystoscopy (now called Hunner's ulcer after the urologist who first reported it). Over the ensuing decades, it became clear that many patients with similar symptoms had no ulcer. It is now appreciated that ≥10% of patients with IC/PBS have a Hunner's ulcer.

The definition of IC/PBS, its diagnostic features, and even its name continue to evolve. The International Continence Society, a body devoted to studying disorders of the lower urinary tract and pelvic floor, has defined PBS as "the complaint of suprapubic pain related to bladder filling, accompanied by other symptoms such as increased daytime and night-time frequency, in the absence of proven urinary infection or other obvious pathology." In practice, clinicians have interpreted this definition to include any chronic pelvic pain that increases with bladder filling and/or decreases with voiding and that cannot be explained by reference to another identifiable disease.

Many patients with IC/PBS also have other syndromes, such as fibromyalgia, chronic fatigue syndrome, irritable bowel syndrome, vulvodynia, and migraine. These syndromes collectively are known as functional somatic syndromes (FSSs): chronic conditions in which pain and fatigue are prominent features but laboratory tests and histologic findings are normal. Like IC/PBS, the FSSs often are associated with depression and anxiety. The majority of cases affect women, and more than one FSS can affect a single patient. Because of its similar features and comorbidity, IC/PBS sometimes is considered an FSS.

Epidemiology

Contemporary population studies of IC/PBS in the United States indicate prevalences of 2–3% among women and 1–2% among men. For decades, it was thought that 90% of IC/PBS cases occurred in women. These prevalence findings, however, have generated research aimed at determining the proportion of men with symptoms usually diagnosed as chronic prostatitis (now known as chronic prostatitis/chronic pelvic pain syndrome) who actually have IC/PBS. Well-designed studies of the incidence of IC/PBS have not been reported.

Among women, the average age at onset of IC/PBS symptoms is the early forties, but the range is from childhood through the early sixties. Risk factors (antecedent features that distinguish cases from controls) primarily have been FSSs. Indeed, the odds of IC/PBS increase with the number of such syndromes present. Surgery was long thought to be a risk factor for IC/PBS, but analyses adjusting for FSSs refuted that association. A minority of patients have a bacterial urinary tract infection (UTI) at the onset of IC/PBS.

The natural history of IC/PBS is not known. Although studies from urology and urogynecology practices have been interpreted as showing that IC/PBS lasts for the lifetime of the patient, population studies suggest that some individuals with IC/PBS do not consult specialists and may not seek medical care at all, and most prevalence studies do not show an upward trend with age—a pattern that would be expected with incident cases throughout adulthood followed by lifetime persistence of a nonfatal disease. It may be reasonable to conclude that patients in a urology practice represent those with the most severe and recalcitrant IC/PBS.

Pathology

For the ≥10% of IC/PBS patients who have a Hunner's ulcer, the term *interstitial cystitis* may indeed describe the histopathologic picture. Most of these patients have substantive inflammation, mast cells, and granulation tissue. However, in the 90% of patients without such ulcers, the bladder mucosa is relatively normal, with scant inflammation.

Etiology

Numerous theories about the pathogenesis of IC/PBS have been put forward. It is not surprising that most early theories focused on the bladder. For instance, IC/PBS has been investigated as a chronic bladder infection. Sophisticated technologies have not identified a causative organism in urine or in bladder tissue; however, the patients studied by these methods had IC/PBS of chronic duration, and the results do not preclude the possibility that infection may trigger the syndrome or may be a feature of early IC/PBS. Other inflammatory factors, including a role for mast cells, have been postulated, but as was noted above, the 90% of patients without a Hunner's ulcer have little bladder inflammation and do not exhibit a prominence of mast cells. Autoimmunity has been considered, but autoantibodies are low in titer, nonspecific, and thought to be a result rather than a cause of IC/PBS. Increased permeability of the bladder mucosa due to defective epithelium or glycosaminoglycan (the bladder's mucous coating) has been studied frequently, but the findings have been inconclusive.

Investigations of causes outside the bladder have been prompted by the presence of comorbid FSSs. Many patients with FSSs have abnormal pain sensitivity as evidenced by (1) low pain thresholds in body areas unrelated to the diagnosed syndrome, (2) dysfunctional descending neurologic control of tactile signals, and (3) enhanced brain responses to touch in functional neuroimaging studies. Moreover, in patients with IC/PBS, body surfaces remote from the bladder are more sensitive to pain than is the case in individuals without IC/PBS. All these findings are consistent with upregulation of sensory processing in the brain. Indeed, the prevailing theory is that these concomitantly occurring syndromes have in common an abnormality of brain processing of sensory input. However, antecedence is a critical criterion for causality, and no study has demonstrated that abnormal pain sensitivity precedes either IC/PBS or FSSs.

Clinical Presentation

In some patients, IC/PBS has a gradual onset and/or the cardinal symptoms of pain, urgency, frequency, and nocturia appear sequentially in no consistent order. Other patients can identify the exact date of onset of IC/PBS symptoms. More than half of the latter patients describe burning or pain with urination beginning on that date. This symptom, usually termed *dysuria*, is characteristic of a UTI. As was stated above, only a minority of IC/PBS patients who obtain medical care soon after symptom onset have uropathogenic bacteria or leukocytes in urine. These patients and many others with new-onset IC/PBS are treated with antibiotics for a presumptive UTI. Men presenting with such symptoms generally are treated for chronic bacterial prostatitis. Persistent or recurring symptoms without bacteriuria eventually prompt a differential diagnosis, and IC/PBS is considered. Traditionally, the diagnosis of IC/PBS has been delayed (sometimes for years), but recent interest in the disease has shortened this interval.

The pain of IC/PBS includes suprapubic prominence and changes with the voiding cycle. Two-thirds of women with IC/PBS report two or more sites of pain. The most common site (involved in 80% of women) and generally the one with the most severe pain is the suprapubic area. About 35% of female patients have pain in the urethra, 25% in other parts of the vulva, and 30% in nonurogenital areas, mostly the low back and also the anterior or posterior thighs or the buttocks. The pain of IC/PBS most commonly is described as aching, pressing, throbbing, tender, and/or piercing. What may distinguish IC/PBS from other pelvic pain is that in 95% of patients bladder filling exacerbates the pain and/or bladder emptying relieves it. Almost as many patients report a puzzling pattern in which certain dietary substances worsen the pain of IC/PBS. Smaller proportions—but still the majority—of patients report that their IC/PBS pain is worsened by menstruation, stress, tight clothing, exercise, and riding in a car as well as during or after vaginal intercourse.

The urethral and vulvar pain of IC/PBS merits special mention. In addition to the descriptive adjectives for IC/PBS mentioned above, this pain commonly is described as burning, stinging, and sharp and as being worsened by touch, tampons, and vaginal intercourse. Patients report that urethral pain increases during urination and generally lessens afterward. These characteristics have commonly caused the urethral pain of IC/PBS to be diagnosed as chronic urethral syndrome and the vulvar pain as vulvodynia.

In many patients with IC/PBS, there is a link between pain and urinary urgency; that is, two-thirds of patients describe the urge to urinate as a desire to relieve pain. Only 20% report that the urge stems from a desire to prevent incontinence; indeed, very few patients with IC/PBS are incontinent. Urinary frequency can be severe, with ~85% of patients voiding >10 times per 24 h and some as often as 60 times. Voiding continues through the night, and nocturia is common, frequent, and often associated with symptoms of sleep deprivation.

Beyond these common symptoms of IC/PBS, additional urinary and other symptoms may be present. Among the urinary symptoms are difficulty in starting urine flow and perceptions of difficulty in emptying the bladder and bladder spasms. Among the other symptoms are the manifestations of comorbid FSSs as well as symptoms that do not constitute recognized syndromes, such as numbness, muscle spasms, dizziness, ringing in the ears, and blurred vision.

The pain, urgency, and frequency of IC/PBS can be debilitating. Proximity to a bathroom is a

continuous focus, and patients report difficulties in the workplace, leisure activities, travel, or simply leaving home. Familial and sexual relationships can be strained.

Diagnosis

Traditionally, IC/PBS has been considered a rare condition that is diagnosed by urologists at cystoscopy. However, this disorder is much more common than once was thought; it is now being considered earlier in its course and is being diagnosed and managed more often by primary care clinicians. Results of physical examination, urinalysis, and urologic procedures are insensitive and/or nonspecific. Thus, diagnosis is based on the presence of appropriate symptoms and the exclusion of diseases with a similar presentation.

Three categories of disorders can be considered in the differential diagnosis of IC/PBS. The first represents other identifiable diseases that manifest as pelvic pain and/or urinary symptoms, including common infections (recurrent UTI, vaginitis, genital herpes); cystitis caused by irradiation, cyclophosphamide treatment, or tuberculosis; neurogenic bladder, bladder stones, or urethral diverticulum; cancer of the bladder, uterus, cervix, vagina, or urethra; and, in men, prostate cancer. Overactive bladder is a chronic condition of women and men that manifests as urgency and frequency and that can be distinguished from IC/PBS by the patient's history: pain is not a feature of overactive bladder, and its urgency arises from the need to avoid incontinence. Endometriosis is a special case: it can be asymptomatic or can cause pelvic pain, dysmenorrhea, and dyspareunia—i.e., types of pain that mimic IC/PBS. Endometrial implants on the bladder (although uncommon) can cause urinary symptoms, and the resulting syndrome can mimic IC/PBS. Even if endometriosis is identified, it is difficult in the absence of bladder implants to determine whether it is causative of or incidental to the symptoms of IC/PBS in a specific woman.

The second category of disorders encompasses the FSSs that can accompany IC/PBS. IC/PBS can be misdiagnosed as chronic pelvic pain, irritable bowel syndrome, or fibromyalgia. The correct diagnosis may be entertained only when changes of pain with altered bladder volume or urinary symptoms appear or become more prominent.

The third category involves syndromes that IC/PBS mimics by way of its referred pain. These syndromes include vulvodynia and chronic urethral syndrome. In studies of adults with chronic urinary symptoms, the distribution of symptoms is similar in men and women, a finding that calls into question the distinctions among IC/PBS, overactive bladder, benign prostatic hyperplasia, and chronic prostatitis /chronic pelvic pain syndrome.

Therefore, IC/PBS should be considered in the differential diagnosis of persistent or recurrent "UTI" but sterile urine cultures; "overactive bladder" with pain; chronic pelvic pain, endometriosis, vulvodynia, or FSSs with urinary symptoms; and "chronic prostatitis." As was mentioned above, important clues to the diagnosis of IC/PBS are the exacerbation of pain with bladder filling or with the consumption of certain foods or drinks and the alleviation of pain after urination.

Cystoscopy under anesthesia formerly was thought to be necessary for the diagnosis because of its capacity to reveal a Hunner's ulcer or—in the 90% of patients without an ulcer—petechial hemorrhages after bladder distention. However, because these findings are nonspecific, many experts find this procedure unnecessary. Because other urologic diagnostic procedures are not particularly helpful, the indications for urologic referral have evolved toward the need to rule out other diseases or administer more advanced treatment.

A typical patient presents to the primary clinician after days, weeks, or months of pain, urgency, frequency, and/or nocturia. The presence of urinary nitrites, leukocytes, or uropathogenic bacteria should prompt treatment for a UTI and bacterial prostatitis in women and men, respectively. Persistence or recurrence of symptoms in the absence of bacteriuria should prompt a pelvic examination for women, an assay for serum prostate-specific antigen for men, and urine cytology and inclusion of IC/PBS in the differential diagnosis in both sexes.

In the diagnosis of IC/PBS, inquiries about pain, pressure, and discomfort are useful; IC/PBS should be considered if any of those sensations are noted in one or more anterior or posterior sites between the umbilicus and the upper thighs. Nondirective questions about the effect of bladder volume changes include "As your next urination approaches, does this pain get better, get worse, or stay the same?" and "After you urinate, does this pain get better, get worse, or stay the same?" Establishing that the pain is exacerbated by the consumption of certain foods and drinks can not only support the diagnosis of IC/PBS but also serve as the basis for one of the first steps in managing this syndrome. A nondirective way to ask about urgency is to describe it to the patient as a compelling urge to urinate that is difficult to postpone; follow-up questions can determine whether this urge is intended to relieve pain or prevent incontinence. To assess severity and provide quantitative baseline measures, pain and urgency should be estimated by the patient on a scale of 0–10, with 0 being none and 10 the worst imaginable. Frequency per 24-h period should be determined and nocturia assessed as the number of times per night the patient is awakened by the need to urinate.

About half of patients with IC/PBS have intermittent or persistent microscopic hematuria; this manifestation and the need to exclude bladder stones or cancer require urologic or urogynecologic

referral. Initiation of therapy for IC/PBS does not hamper subsequent urologic evaluation.

TREATMENT

The goal of therapy is to relieve the symptoms of IC/PBS; the challenge lies in the fact that no treatment is uniformly successful. However, most patients eventually obtain relief, generally with a multifaceted approach. The correct strategy is to begin with less invasive therapies and move to more invasive measures only if necessary and under the supervision of a urologist or urogynecologist. Tactics include education, dietary changes, medications, pelvic-floor physical therapy, and treatment of associated FSSs.

Months or even years may have passed since the onset of symptoms, and the patient's life may have been disrupted continually, with repeated medical visits provoking frustration and dismay in both the patient and physicians. In this circumstance, simply giving a name to the syndrome is beneficial. The physician should discuss the disease, the diagnostic and therapeutic strategies, and the prognosis with the patient and with the spouse and/or other pertinent family members, who may need to be made aware that although IC/PBS has no visible manifestations, the patient is undergoing substantial pain and suffering. This information is particularly important for sexual partners, as exacerbation of pain during and after intercourse is a common feature of IC/PBS. The Interstitial Cystitis Association (*http://www.ichelp.com*) and the Interstitial Cystitis Network (*http://www.ic-network.com*) can be useful in this educational process.

Over time, many patients identify particular foods and drinks that exacerbate their symptoms. Common among these are chilies, chocolate, citrus fruits, tomatoes, alcohol, caffeinated drinks, and carbonated beverages; full lists of common trigger foods are available at the websites cited above. To construct a benign diet, some patients find it useful to exclude all possible offenders and add those items back into the diet one at a time to identify the ones that worsen IC/PBS symptoms. Patients also should experiment with fluid volumes; some find relief with less fluid, others with more.

Among oral medications, nonsteroidal anti-inflammatory drugs are commonly used first but are often unsuccessful. A small randomized controlled trial suggested that amitriptyline is a reasonable choice for the next agent. This drug is used not for its antidepressant activity but because of its proven effects on neuropathic pain. The initial dose of 25 mg at bedtime is increased weekly by 25 mg up to 100 mg (or less if adequate relief of symptoms is obtained with a lower dose). Side effects can be expected and include dry mouth, weight gain, sedation, and constipation. If this regimen does not control symptoms adequately, pentosan polysulfate, a semisynthetic polysaccharide, can be added at a dose of 100 mg three times a day. Its theoretical effect is to replenish a possibly defective glycosaminoglycan layer over the bladder mucosa, but randomized clinical trials suggest only a modest benefit over placebo. Adverse reactions are uncommon and include gastrointestinal symptoms, headache, and alopecia. Pentosan polysulfate has weak anticoagulant effects and perhaps should be avoided by patients with coagulation abnormalities.

Tenderness of the pelvic floor often is reported by IC/PBS patients. A small but cleverly designed randomized clinical trial suggested that weekly physical therapy directed at muscles and soft tissues of the pelvis yields significantly more relief than a similar schedule of general body massage. This intervention can be initiated under the direction of a knowledgeable physical therapist while trials of medications are under way.

Anecdotal reports suggest that successful therapy for one FSS is accompanied by diminished symptoms of other FSSs. As has been noted here, IC/PBS often is associated with one or several FSSs. Thus, it seems reasonable to hope that, to the extent that accompanying FSSs can be treated successfully, the symptoms of IC/PBS may be relieved as well.

If several months of these therapies in combination do not relieve symptoms adequately, the patient should be referred to a urologist or urogynecologist who has access to additional modalities. Cystoscopy under anesthesia allows distention of the bladder with water, a procedure that provides ~ 40% of patients with several months of relief and can be repeated. For those few patients with a Hunner's ulcer, fulguration may offer relief. Bladder instillation of solutions containing lidocaine can be repeated. A small randomized clinical trial indicates that bladder instillation of dimethyl sulfoxide is effective in significantly more patients than is placebo. Physicians experienced in the care of IC/PBS patients have used anticonvulsants, narcotics, and cyclosporine as components of therapy. Pain specialists can be of assistance. Sacral neuromodulation can be tested with a temporary percutaneous electrode and, if effective, can be administered with an implanted device. In a very small number of patients with recalcitrant symptoms, surgeries, including cystoplasty, partial or total cystectomy, and urinary diversion, can be valuable.

FURTHER READINGS

FitzGerald MP et al: Randomized multicenter feasibility trial of myofascial physical therapy for the treatment of urological chronic pelvic pain syndromes. J Urol 182:570, 2009[PMID: 19535099]

Hanno P et al: Bladder Pain Syndrome Committee of the International Consultation on Incontinence.

Neurourol Urodyn 29:191, 2010[PMID: 20025029]

Marinkovic SP et al: The management of interstitial cystitis or painful bladder syndrome in women. BMJ 339:b2707, 2009

Phatak S, Foster HE Jr: The management of interstitial cystitis: An update. Nat Clin Pract Urol 3:45, 2006[PMID: 16474494]

van Ophoven A et al: A prospective, randomized, placebo controlled, double-blind study of amitriptyline for the treatment of interstitial cystitis. J Urol 172:533, 2004

Warren JW et al: Antecedent nonbladder syndromes in case-control study of interstitial cystitis/painful bladder syndrome. Urology 73:52, 2009[PMID: 18995888]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

URINARY TRACT OBSTRUCTION: INTRODUCTION

Obstruction to the flow of urine, with attendant stasis and elevation in urinary tract pressure, impairs renal and urinary conduit functions and is a common cause of acute and chronic kidney disease (obstructive nephropathy). With early relief of obstruction, the defects in function usually disappear completely. However, chronic obstruction may produce permanent loss of renal mass (renal atrophy) and excretory capability, as well as enhanced susceptibility to local infection and stone formation. Early diagnosis and prompt therapy are, therefore, essential to minimize the otherwise devastating effects of obstruction on kidney structure and function.

ETIOLOGY

Obstruction to urine flow can result from *intrinsic* or *extrinsic mechanical blockade* as well as from *functional defects* not associated with fixed occlusion of the urinary drainage system. Mechanical obstruction can occur at any level of the urinary tract, from the renal calyces to the external urethral meatus. Normal points of narrowing, such as the ureteropelvic and ureterovesical junctions, bladder neck, and urethral meatus, are common sites of obstruction. When obstruction is above the level of the bladder, unilateral dilatation of the ureter (*hydroureter*) and renal pyelocalyceal system (*hydronephrosis*) occur; lesions at or below the level of the bladder cause bilateral involvement.

Common forms of obstruction are listed in Table 289-1. Childhood causes include *congenital malformations*, such as narrowing of the ureteropelvic junction and abnormal insertion of the ureter into the bladder, the most common cause. Vesicoureteral reflux in the absence of urinary tract infection or bladder neck obstruction often resolves with age. Reinsertion of the ureter into the bladder is indicated if reflux is severe and unlikely to improve spontaneously, if renal function deteriorates, or if urinary tract infections recur despite chronic antimicrobial therapy. Vesicoureteral reflux may cause prenatal hydronephrosis and, if severe, can lead to recurrent urinary infections and renal scarring in childhood. Posterior urethral valves are the most common cause of bilateral hydronephrosis in boys. In adults, urinary tract obstruction (UTO) is due mainly to *acquired defects*. Pelvic tumors, calculi, and urethral stricture predominate. Ligation of, or injury to, the ureter during pelvic or colonic surgery can lead to hydronephrosis which, if unilateral, may remain undetected. Obstructive uropathy may also result from extrinsic neoplastic (carcinoma of cervix or colon) or inflammatory disorders. Lymphomas and pelvic or colonic neoplasms with retroperitoneal involvement are causes of ureteral obstruction.

Table 289-1 Common Mechanical Causes of Urinary Tract Obstruction

Ureter	Bladder Outlet	Urethra
Congenital		
Ureteropelvic junction narrowing or obstruction	Bladder neck obstruction	Posterior urethral valves
Ureterovesical junction narrowing or obstruction and reflux	Ureterocele	Anterior urethral valves
Ureterocele		Stricture
Retrocaval ureter		Meatal stenosis
		Phimosis
Acquired Intrinsic Defects		
Calculi	Benign prostatic hyperplasia	Stricture
Inflammation	Cancer of prostate	Tumor
Infection	Cancer of bladder	Calculi
Trauma	Calculi	Trauma
Sloughed papillae	Diabetic neuropathy	Phimosis
Tumor	Spinal cord disease	
Blood clots	Anticholinergic drugs and α-adrenergic antagonists	

Acquired Extrinsic Defects		
Pregnant uterus Retroperitoneal fibrosis Aortic aneurysm Uterine leiomyomata Carcinoma of uterus, prostate, bladder, colon, rectum Lymphoma Pelvic inflammatory disease, endometriosis Accidental surgical ligation	Carcinoma of cervix, colon Trauma	Trauma

Functional impairment of urine flow usually results from disorders that involve both the ureter and bladder. Causes include neurogenic bladder, often with adynamic ureter, and vesicoureteral reflux. Reflux in children may result in severe unilateral or bilateral hydroureter and hydronephrosis. Urinary retention may be the consequence of alpha-adrenergic and anticholinergic agents, as well as opiates. Hydronephrosis in pregnancy is due to relaxational effects of progesterone on smooth muscle of the renal pelvis, as well as ureteral compression by the enlarged uterus.

CLINICAL FEATURES AND PATHOPHYSIOLOGY

The pathophysiology and clinical features of UTO are summarized in Table 289-2. *Pain*, the symptom that most commonly leads to medical attention, is due to distention of the collecting system or renal capsule. Pain severity is influenced more by the rate at which distention develops than by the degree of distention. Acute supravesical obstruction, as from a stone lodged in a ureter (see Chap. 287), is associated with excruciating pain, known as *renal colic*. This pain often radiates to the lower abdomen, testes, or labia. By contrast, more insidious causes of obstruction, such as chronic narrowing of the ureteropelvic junction, may produce little or no pain and yet result in total destruction of the affected kidney. Flank pain that occurs only with micturition is pathognomonic of vesicoureteral reflux. Hesitancy and straining to initiate the urinary stream, postvoid dribbling, urinary frequency, and incontinence are common with obstruction at or below the level of the bladder.

Table 289-2 Pathophysiology of Bilateral Ureteral Obstruction

Hemodynamic Effects	Tubule Effects	Clinical Features
Acute		
↑Renal blood flow ↓GFR ↓Medullary blood flow ↑Vasodilator prostaglandins, nitric oxide	↑Ureteral and tubule pressures ↑Reabsorption of Na^+, urea, water	Pain (capsule distention) Azotemia Oliguria or anuria
Chronic		
↓ Renal blood flow ↓↓GFR ↑Vasoconstrictor prostaglandins ↑Renin-angiotensin production	↓Medullary osmolarity ↓Concentrating ability Structural damage; parenchymal atrophy ↓Transport functions for Na^+, K^+, H^+	Azotemia Hypertension AVP-insensitive polyuria Natriuresis Hyperkalemic, hyperchloremic acidosis
Release of Obstruction		
Slow ↑ in GFR (variable)	↓Tubule pressure ↑Solute load per nephron (urea, NaCl) Natriuretic factors present	Postobstructive diuresis Potential for volume depletion and electrolyte imbalance due to losses of Na^+, K^+, PO_4^{2-}, Mg^{2+}, and water

Abbreviations: AVP, arginine vasopressin; GFR, glomerular filtration rate.

Obstruction of urine flow results in an increase in hydrostatic pressures proximal to the site of obstruction. It is this buildup of pressure that leads to the accompanying pain, the distention of the collecting system in the kidney, and elevated intratubular pressures that initiate tubular dysfunction. As the increased hydrostatic pressure is expressed in the urinary space of the glomeruli, further filtration decreases or stops completely.

Azotemia develops when overall excretory function is impaired, often in the setting of bladder outlet obstruction, bilateral renal pelvic or ureteric obstruction, or unilateral disease in a patient with a solitary functioning kidney. Complete bilateral obstruction should be suspected when acute renal failure is accompanied by anuria. Any patient with renal failure otherwise unexplained, or with a history of nephrolithiasis, hematuria, diabetes mellitus, prostatic enlargement, pelvic surgery, trauma, or tumor should be evaluated for UTO.

In the acute setting, partial, bilateral obstruction may mimic prerenal azotemia with concentrated urine and sodium retention. However, with more prolonged obstruction, symptoms of *polyuria* and *nocturia* commonly accompany partial UTO and result from diminished renal concentrating ability. Impairment of transcellular salt reabsorption in the proximal tubule, medullary thick ascending limb of Henle, and collecting duct cells is due to down-regulation of transport proteins including the Na^+, K^+ adenosine triphosphatase (ATPase), NaK_2Cl cotransporter (NKCC) in the thick ascending limb, and the epithelial Na+ channel (ENaC) in collecting duct cells. Consequences include failure to produce urine free of salt (natriuresis), and loss of medullary hypertonicity producing a urinary concentrating defect. In addition to direct effects on renal transport mechanisms, increased PGE2 (due to induction of COX-2), angiotensin-II (with its downregulation of Na^+ transporters), and atrial natriuretic peptide (ANP) (due to volume expansion in the azotemic patient) contribute to the decreased salt reabsorption along the nephron.

Dysregulation of aquaporin-2 water channels in the collecting duct contributes to the polyuria. The defect usually does not improve with administration of vasopressin and is therefore a form of acquired nephrogenic diabetes insipidus.

Wide fluctuations in urine output in a patient with azotemia should always raise the possibility of intermittent or partial UTO. If fluid intake is inadequate, severe dehydration and hypernatremia may develop. However, as with other causes of poor renal function, excesses of salt and water intake may result in edema and hyponatremia.

Partial bilateral UTO often results in *acquired distal renal tubular acidosis, hyperkalemia* , and *renal salt wasting*. The H^+-ATPase, situated on the apical membrane of the intercalated cells of the collecting duct, is critical for distal H^+ secretion. The trafficking of intracellular H^+ pumps from the cytoplasm to the cell membrane is disrupted in UTO. The decreased function of the ENaC, in the apical membrane of neighboring collecting duct principal cells, contributes to decreased Na+ reabsorption (salt-wasting), decreased electronegativity of the tubule lumen, and therefore decreased K^+ secretion via K^+ channels (hyperkalemia) and H^+ secretion via the H^+-ATPases [distal renal tubular acidosis (RTA)]. Proximal tubule ammoniagenesis, important to the elimination of H^+ as $NH4^+$, is impaired. These defects in tubule function are often accompanied by renal tubulointerstitial damage. Azotemia with hyperkalemia and metabolic acidosis should prompt consideration of UTO.

The renal interstitium becomes edematous and infiltrated with mononuclear inflammatory cells early in UTO. Later, interstitial fibrosis and atrophy of the papillae and medulla occur and precede these processes in the cortex. The increase in angiotensin-II noted in UTO contributes to the inflammatory response and fibroblast accumulation through mechanisms involving profibrotic cytokines. With time, this process leads to chronic kidney damage.

UTO must always be considered in patients with urinary tract infections or urolithiasis. Urinary stasis encourages the growth of organisms. Urea-splitting bacteria are associated with magnesium ammonium phosphate (struvite) calculi. *Hypertension* is frequent in acute and subacuteunilateral obstruction and is usually a consequence of increased release of renin by the involved kidney. Chronic kidney disease from bilateral UTO, often associated with extracellular volume expansion, may result in significant hypertension. *Erythrocytosis*, an infrequent complication of obstructive uropathy, is secondary to increased erythropoietin production.

DIAGNOSIS

A history of difficulty in voiding, pain, infection, or change in urinary volume is common. Evidence for distention of the kidney or urinary bladder can often be obtained by palpation and percussion of the abdomen. A careful rectal and genital examination may reveal enlargement or nodularity of the prostate, abnormal rectal sphincter tone, or a rectal or pelvic mass.

Urinalysis may reveal hematuria, pyuria, and bacteriuria. The urine sediment is often normal, even

when obstruction leads to marked azotemia and extensive structural damage. An abdominal scout film may detect nephrocalcinosis or a radiopaque stone. As indicated in Fig. 289-1, if UTO is suspected, a bladder catheter should be inserted. Abdominal ultrasonography should be performed to evaluate renal and bladder size, as well as pyelocalyceal contour. Ultrasonography is approximately 90% specific and sensitive for detection of hydronephrosis. False-positive results are associated with diuresis, renal cysts, or the presence of an extrarenal pelvis, a normal congenital variant. Congenital ureteropelvic junction (UPJ) obstruction may be mistaken for renal cystic disease. Hydronephrosis may be absent on ultrasound when obstruction is less than 48 hours in duration or associated with volume contraction, staghorn calculi, retroperitoneal fibrosis, or infiltrative renal disease. Duplex Doppler ultrasonography may detect an increased resistive index in urinary obstruction.

Figure 289-1

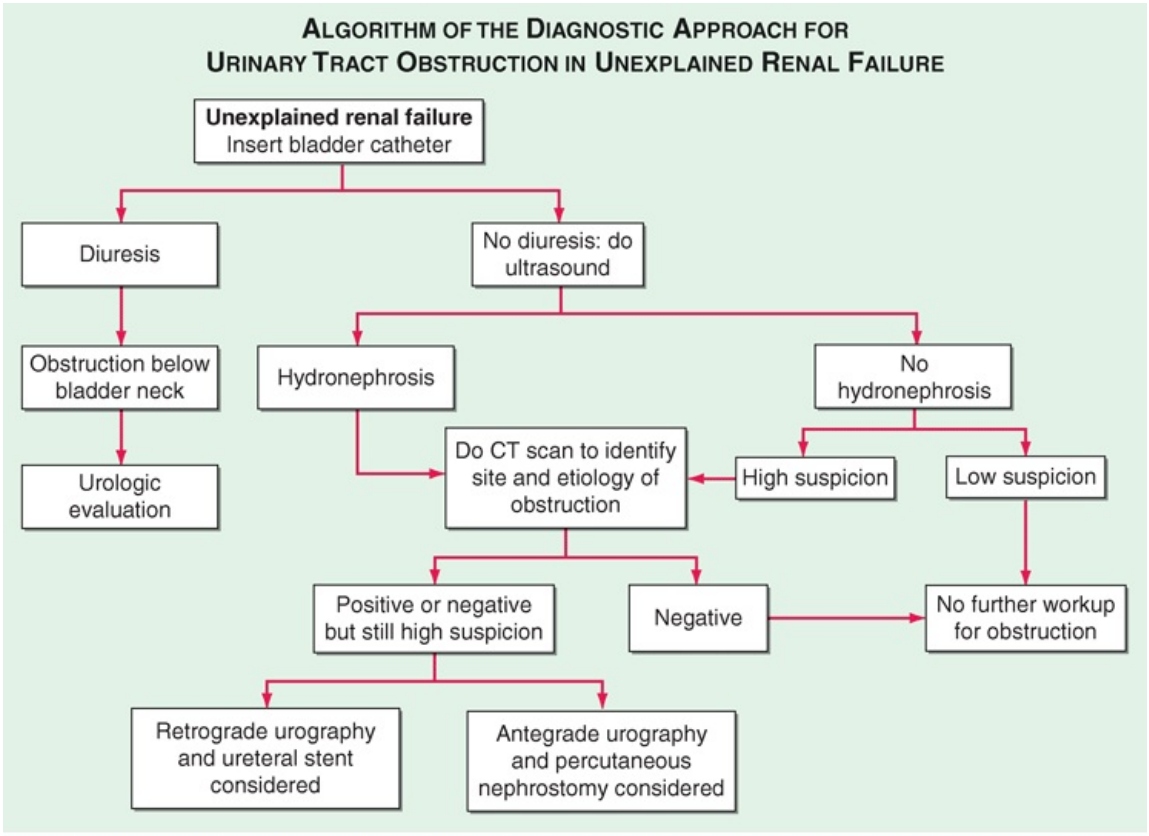

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagnostic approach for urinary tract obstruction in unexplained renal failure. CT, computed tomography.

Recent advances in technology have led to alternatives to the once standard intravenous urogram in the further evaluation of UTO. The high-resolution multidetector row CT scan in particular has advantages of visualizing the retroperitoneum, as well as identifying both intrinsic and extrinsic sites of obstruction. Noncontrast CT scans improve visualization of the urinary tract in the patient with renal impairment and are safer for patients at risk for contrast nephropathy. MR urography is a promising technique but at this time not superior to the CT scan and carries the risk of certain gadolinium agents in patients with renal insufficiency, i.e., nephrogenic systemic fibrosis. The intravenous urogram may define the site of obstruction and demonstrate dilatation of the calyces, renal pelvis, and ureter above the obstruction. The ureter may be tortuous in chronic obstruction. Radionuclide scans are able to give differential renal function but give less anatomic detail than CT or intravenous urography (IVU).

To facilitate visualization of a suspected lesion in a ureter or renal pelvis, *retrograde* or *antegrade urography* should be attempted. These procedures do not carry risk of contrast-induced acute renal failure in patients with renal insufficiency. The retrograde approach involves catheterization of the involved ureter under cystoscopic control, while the antegrade technique necessitates percutaneous placement of a catheter into the renal pelvis. While the antegrade approach may provide immediate decompression of a unilateral obstructing lesion, many urologists initially attempt the retrograde approach unless the catheterization is unsuccessful.

Voiding cystourethrography is of value in the diagnosis of vesicoureteral reflux and bladder neck and urethral obstructions. Postvoiding films reveal residual urine. Endoscopic visualization by the urologist

often permits precise identification of lesions involving the urethra, prostate, bladder, and ureteral orifices.

Treatment: Urinary Tract Obstruction

UTO complicated by infection requires immediate relief of obstruction to prevent development of generalized sepsis and progressive renal damage. Sepsis necessitates prompt urologic intervention. Drainage may be achieved by nephrostomy, ureterostomy, or ureteral, urethral, or suprapubic catheterization. Prolonged antibiotic treatment may be necessary. Chronic or recurrent infections in a poorly functioning obstructed kidney may necessitate nephrectomy. When infection is not present, surgery is often delayed until acid-base, fluid, and electrolyte status is restored. Nevertheless, the site of obstruction should be ascertained as soon as feasible. Elective relief of obstruction is usually recommended in patients with urinary retention, recurrent urinary tract infections, persistent pain, or progressive loss of renal function. Benign prostatic hypertrophy may be treated medically with alpha adrenergic blockers and 5α-reductase inhibitors. Functional obstruction secondary to neurogenic bladder may be decreased with the combination of frequent voiding and cholinergic drugs.

PROGNOSIS

With relief of obstruction, the prognosis regarding return of renal function depends largely on whether irreversible renal damage has occurred. When obstruction is not relieved, the course will depend mainly on whether the obstruction is complete or incomplete and bilateral or unilateral, as well as whether or not urinary tract infection is also present. Complete obstruction with infection can lead to total destruction of the kidney within days. Partial return of glomerular filtration rate may follow relief of complete obstruction of 1 and 2 weeks' duration, but after 8 weeks of obstruction, recovery is unlikely. In the absence of definitive evidence of irreversibility, every effort should be made to decompress the obstruction in the hope of restoring renal function at least partially. A renal radionuclide scan, performed after a prolonged period of decompression, may be used to predict the reversibility of renal dysfunction.

POSTOBSTRUCTIVE DIURESIS

Relief of bilateral, but not unilateral, complete obstruction commonly results in polyuria, which may be massive. The urine is usually hypotonic and may contain large amounts of sodium chloride, potassium, phosphate, and magnesium. The natriuresis is due in part to the excretion of retained urea (osmotic diuresis), natriuretic factors accumulated during uremia and depressed salt and water reabsorption when urine flow is reestablished. In the majority of patients this diuresis results in the *appropriate* excretion of the excesses of retained salt and water. When extracellular volume and composition return to normal, the diuresis usually abates spontaneously. Occasionally, iatrogenic expansion of extracellular volume is responsible for, or sustains, the diuresis observed in the postobstructive period. Replacement with intravenous fluids in amounts less than urinary losses usually avoids this complication. More aggressive fluid management is required in the setting of hypovolemia, hypotension, or disturbances in serum electrolyte concentrations.

The loss of electrolyte-free water with urea may result in hypernatremia. Serum and urine sodium and osmolal concentrations should guide the use of appropriate intravenous replacement. Often replacement with 0.45% saline is required. Relief of obstruction may be followed by urinary salt and water losses severe enough to provoke profound dehydration and vascular collapse. In these patients, decreased tubule reabsorptive capacity is probably responsible for the marked diuresis. Appropriate therapy in such patients includes intravenous administration of salt-containing solutions to replace sodium and volume deficits.

FURTHER READINGS

Campbell SC, Walsh PC: Pathophysiology of urinary tract obstruction, in *Campbell-Walsh Urology*, vol 2, 9th ed, J Wein (ed). Philadelphia, Saunders, 2007, pp 1195–1226

Frokiaer J, Zeidel ML: Urinary tract obstruction, in *Brenner and Rector's The Kidney* , 8th ed, BM Brenner (ed). Philadelphia, Saunders, 2007, pp 1239–1259

Klahr S: Urinary tract obstruction, in *Diseases of the Kidney* , 7th ed, RW Schrier, CW Gottschalk (eds). Boston, Little, Brown, 2001, pp 751–787

Williams B et al: Pathophysiology and treatment of ureteropelvic junction obstruction. Curr Urol Rep 8:111, 2007[PMID: 17303015]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ANATOMIC CONSIDERATIONS

The gastrointestinal (GI) tract extends from the mouth to the anus and is composed of several organs with distinct functions. Specialized independently controlled thickened sphincters that assist in gut compartmentalization separate the organs. The gut wall is organized into well-defined layers that contribute to functional activities in each region. The mucosa is a barrier to luminal contents or as a site for transfer of fluids or nutrients. Gut smooth muscle mediates propulsion from one region to the next. Many GI organs possess a serosal layer that provides a supportive foundation but that also permits external input.

Interactions with other organ systems serve the needs both of the gut and the body. Pancreaticobiliary conduits deliver bile and enzymes into the duodenum. A rich vascular supply is modulated by GI tract activity. Lymphatic channels assist in gut immune activities. Intrinsic gut wall nerves provide the basic controls for propulsion and fluid regulation. Extrinsic neural input provides volitional or involuntary control to degrees that are specific for each gut region.

FUNCTIONS OF THE GASTROINTESTINAL TRACT

The GI tract serves two main functions—assimilating nutrients and eliminating waste. The gut anatomy is organized to serve these functions. In the mouth, food is processed, mixed with salivary amylase, and delivered to the gut lumen. The esophagus propels the bolus into the stomach; the lower esophageal sphincter prevents oral reflux of gastric contents. The esophageal mucosa has a protective squamous histology, which does not permit significant diffusion or absorption. Propulsive esophageal activities are exclusively aboral and coordinate with relaxation of the upper and lower esophageal sphincters on swallowing.

The stomach furthers food preparation by triturating and mixing the bolus with pepsin and acid. Gastric acid also sterilizes the upper gut. The proximal stomach serves a storage function by relaxing to accommodate the meal. The distal stomach exhibits phasic contractions that propel solid food residue against the pylorus, where it is repeatedly propelled proximally for further mixing before it is emptied into the duodenum. Finally, the stomach secretes intrinsic factor for vitamin B12 absorption.

The small intestine serves most of the nutrient absorptive function of the gut. The intestinal mucosa exhibits villus architecture to provide maximal surface area for absorption and is endowed with specialized enzymes and transporters. Triturated food from the stomach mixes with pancreatic juice and bile in the duodenum to facilitate digestion. Pancreatic juice contains the main enzymes for carbohydrate, protein, and fat digestion as well as bicarbonate to optimize the pH for activation of these enzymes. Bile secreted by the liver and stored in the gallbladder is essential for intestinal lipid digestion. The proximal intestine is optimized for rapid absorption of nutrient breakdown products and most minerals, while the ileum is better suited for absorption of vitamin B12 and bile acids. The small intestine also aids in waste elimination. Bile contains by-products of erythrocyte degradation, toxins, metabolized and unmetabolized medications, and cholesterol. Motor function of the small intestine delivers indigestible food residue and sloughed enterocytes into the colon for further processing. The small intestine terminates in the ileocecal junction, a sphincteric structure that prevents coloileal reflux and maintains small-intestinal sterility.

The colon prepares the waste material for controlled evacuation. The colonic mucosa dehydrates the stool, decreasing daily fecal volumes from 1000–1500 mL delivered from the ileum to 100–200 mL expelled from the rectum. The colonic lumen possesses a dense bacterial colonization that ferments undigested carbohydrates and short-chain fatty acids. Whereas transit times in the esophagus are on the order of seconds and times in the stomach and small intestine range from minutes to a few hours, propagation through the colon takes more than one day in most individuals. Colonic motor patterns exhibit a to-and-fro character that facilitates slow fecal desiccation. The proximal colon serves to mix and absorb fluid, while the distal colon exhibits peristaltic contractions and mass actions that function to expel the stool. The colon terminates in the anus, a structure with volitional and involuntary controls to permit retention of the fecal bolus until it can be released in a socially convenient setting.

EXTRINSIC MODULATION OF GUT FUNCTION

GI function is modified by influences outside of the gut. Unlike other organ systems, the gut is in continuity with the outside environment. Thus, protective mechanisms are vigilant against deleterious effects of foods, medications, toxins, and infectious organisms. Mucosal immune mechanisms include

chronic lymphocyte and plasma cell populations in the epithelial layer and lamina propria backed up by lymph node chains to prevent noxious agents from entering the circulation. All substances absorbed into the bloodstream are filtered through the liver via the portal venous circulation. In the liver, many drugs and toxins are detoxified by a variety of mechanisms. Although intrinsic nerves control most basic gut activities, extrinsic neural input modulates many functions. Two activities under voluntary control are swallowing and defecation. Many normal GI reflexes involve extrinsic vagus or splanchnic nerve pathways. The brain gut axis further alters function in regions not under volitional regulation. As an example, stress has potent effects on gut motor, secretory, and sensory functions.

OVERVIEW OF GASTROINTESTINAL DISEASES

GI diseases develop as a result of abnormalities within or outside of the gut and range in severity from those that produce mild symptoms and no long-term morbidity to those with intractable symptoms or adverse outcomes. Diseases may be localized to one organ or exhibit diffuse involvement at many sites.

Classification of GI Diseases

GI diseases are manifestations of alterations in nutrient assimilation or waste evacuation or in the activities supporting these main functions.

IMPAIRED DIGESTION AND ABSORPTION

Diseases of the stomach, intestine, biliary tree, and pancreas can disrupt digestion and absorption. The most common intestinal maldigestion syndrome, lactase deficiency, produces gas and diarrhea after dairy products and has no adverse outcomes. Other intestinal enzyme deficiencies produce similar symptoms after ingestion of other simple sugars. Conversely, celiac disease, bacterial overgrowth, infectious enteritis, Crohn's ileitis, and radiation damage, which affect digestion and/or absorption more diffusely, produce anemia, dehydration, electrolyte disorders, or malnutrition. Gastric hypersecretory conditions such as Zollinger-Ellison syndrome damage the intestinal mucosa, impair pancreatic enzyme activation, and accelerate transit due to excess gastric acid. Biliary obstruction from stricture or neoplasm impairs fat digestion. Impaired pancreatic enzyme release in chronic pancreatitis or pancreatic cancer decreases intraluminal digestion and can lead to malnutrition.

ALTERED SECRETION

Selected GI diseases result from dysregulation of gut secretion. Gastric acid hypersecretion occurs in Zollinger-Ellison syndrome, G cell hyperplasia, retained antrum syndrome, and some individuals with duodenal ulcers. Conversely, patients with atrophic gastritis or pernicious anemia release little or no gastric acid. Inflammatory and infectious small-intestinal and colonic diseases produce fluid loss through impaired absorption or enhanced secretion. Common intestinal and colonic hypersecretory conditions cause diarrhea and include acute bacterial or viral infection, chronic *Giardia* or cryptosporidia infections, small-intestinal bacterial overgrowth, bile salt diarrhea, microscopic colitis, diabetic diarrhea, and abuse of certain laxatives. Less common causes include large colonic villus adenomas and endocrine neoplasias with tumor overproduction of secretagogue transmitters like vasoactive intestinal polypeptide.

ALTERED GUT TRANSIT

Impaired gut transit may be secondary to mechanical obstruction. Esophageal occlusion often results from acid-induced stricture or neoplasm. Gastric outlet obstruction develops from peptic ulcer disease or gastric cancer. Small-intestinal obstruction most commonly results from adhesions but may also occur with Crohn's disease, radiation- or drug-induced strictures, and less likely malignancy. The most common cause of colonic obstruction is colon cancer, although inflammatory strictures develop in patients with inflammatory bowel disease, after certain infections such as diverticulitis, or with some drugs.

Retardation of propulsion also develops from disordered motor function. Achalasia is characterized by impaired esophageal body peristalsis and incomplete lower esophageal sphincter relaxation. Gastroparesis is the symptomatic delay in gastric emptying of meals due to impaired gastric motility. Intestinal pseudoobstruction causes marked delays in small-bowel transit due to enteric nerve or intestinal smooth-muscle injury. Slow-transit constipation is produced by diffusely impaired colonic propulsion. Constipation also is produced by outlet abnormalities such as rectal prolapse, intussusception, or dyssynergia—a failure of anal or puborectalis relaxation upon attempted defecation.

Disorders of rapid propulsion are less common than those with delayed transit. Rapid gastric emptying occurs in postvagotomy dumping syndrome, with gastric hypersecretion, and in some cases of functional dyspepsia and cyclic vomiting syndrome. Exaggerated intestinal or colonic motor patterns may be responsible for diarrhea in irritable bowel syndrome. Accelerated transit with hyperdefecation is noted in hyperthyroidism.

IMMUNE DYSREGULATION

Many inflammatory GI conditions are consequences of altered gut immune function. The mucosal inflammation of celiac disease results from dietary ingestion of gluten-containing grains. Some patients

with food allergy also exhibit altered immune populations. Eosinophilic esophagitis and eosinophilic gastroenteritis are inflammatory disorders with prominent mucosal eosinophils. Ulcerative colitis and Crohn's disease are disorders of uncertain etiology that produce mucosal injury primarily in the lower gut. The microscopic colitides, lymphocytic and collagenous colitis, exhibit colonic subepithelial infiltrates without visible mucosal damage. Bacterial, viral, and protozoal organisms may produce ileitis or colitis in selected patient populations.

IMPAIRED GUT BLOOD FLOW

Different GI regions are at variable risk for ischemic damage from impaired blood flow. Rare cases of gastroparesis result from blockage of the celiac and superior mesenteric arteries. More commonly encountered are intestinal and colonic ischemia that are consequences of arterial embolus, arterial thrombosis, venous thrombosis, or hypoperfusion from dehydration, sepsis, hemorrhage, or reduced cardiac output. These may produce mucosal injury, hemorrhage, or even perforation. Some cases of radiation enterocolitis exhibit reduced mucosal blood flow.

NEOPLASTIC DEGENERATION

All GI regions are susceptible to malignant degeneration to varying degrees. In the United States, colorectal cancer is most common and usually presents after age 50 years. Worldwide, gastric cancer is prevalent especially in certain Asian regions. Esophageal cancer develops with chronic acid reflux or after an extensive alcohol or tobacco use history. Small-intestinal neoplasms are rare and occur with underlying inflammatory disease. Anal cancers arise after prior anal infection or inflammation. Pancreatic and biliary cancers elicit severe pain, weight loss, and jaundice and have poor prognoses. Hepatocellular carcinoma usually arises in the setting of chronic viral hepatitis or cirrhosis secondary to other causes. Most GI cancers exhibit carcinomatous histology; however, lymphomas and other cell types also are observed.

DISORDERS WITHOUT OBVIOUS ORGANIC ABNORMALITIES

The most common GI disorders show no abnormalities on biochemical or structural testing and include irritable bowel syndrome, functional dyspepsia, functional chest pain, and functional heartburn. These disorders exhibit altered gut motor function; however, the pathogenic relevance of these abnormalities is uncertain. Exaggerated visceral sensory responses to noxious stimulation may cause discomfort in these disorders. Symptoms in other patients result from altered processing of visceral pain sensations in the central nervous system. Functional bowel patients with severe symptoms may exhibit significant emotional disturbances on psychometric testing. Subtle immunologic defects may contribute to functional symptoms as well.

GENETIC INFLUENCES

Although many GI diseases result from environmental factors, others exhibit hereditary components. Family members of inflammatory bowel disease patients show a genetic predisposition to disease development themselves. Colonic and esophageal malignancies arise in certain inherited disorders. Rare genetic dysmotility syndromes are described. Familial clustering is even observed in the functional bowel disorders, although this may be secondary learned familial illness behavior rather than a true hereditary factor.

Symptoms of Gastrointestinal Disease

The most common GI symptoms are abdominal pain, heartburn, nausea and vomiting, altered bowel habits, GI bleeding, and jaundice **(Table 290-1)**. Others are dysphagia, anorexia, weight loss, fatigue, and extraintestinal symptoms.

Table 290-1 Common Causes of Common GI Symptoms

Abdominal Pain	Nausea and Vomiting	Diarrhea	GI Bleeding	Obstructive Jaundice
Appendicitis	Medications	Infection	Ulcer disease	Bile duct stones
Gallstone disease	GI obstruction	Poorly absorbed sugars	Esophagitis	Cholangiocarcinoma
Pancreatitis	Motor disorders	Inflammatory bowel disease	Varices	Cholangitis
Diverticulitis	Functional bowel disorder	Microscopic colitis	Vascular lesions	Sclerosing cholangitis
Ulcer disease	Enteric infection	Functional bowel disorder	Neoplasm	Ampullary stenosis
Esophagitis	Pregnancy	Celiac disease	Diverticula	Ampullary carcinoma
GI obstruction	Endocrine disease	Pancreatic insufficiency	Hemorrhoids	Pancreatitis
Inflammatory bowel disease	Motion sickness	Hyperthyroidism	Fissures	Pancreatic tumor
Functional bowel disorder	Central nervous system disease		Inflammatory bowel disease	
Vascular disease			Infectious colitis	

vascular disease Gynecologic causes Renal stone	Ischemia Endocrine tumor		

ABDOMINAL PAIN

Abdominal pain results from GI disease and extraintestinal conditions involving the genitourinary tract, abdominal wall, thorax, or spine. Visceral pain generally is midline in location and vague in character, while parietal pain is localized and precisely described. Common inflammatory diseases with pain include peptic ulcer, appendicitis, diverticulitis, inflammatory bowel disease, and infectious enterocolitis. Other intraabdominal causes of pain include gallstone disease and pancreatitis. Noninflammatory visceral sources include mesenteric ischemia and neoplasia. The most common causes of abdominal pain are irritable bowel syndrome and functional dyspepsia.

HEARTBURN

Heartburn, a burning substernal sensation, is reported intermittently by at least 40% of the population. Classically, heartburn is felt to result from excess gastroesophageal reflux of acid. However, some cases exhibit normal esophageal acid exposure and may result from reflux of nonacidic material or heightened sensitivity of esophageal mucosal nerves.

NAUSEA AND VOMITING

Nausea and vomiting are caused by GI diseases, medications, toxins, acute and chronic infection, endocrine disorders, labyrinthine conditions, and central nervous system disease. The best-characterized GI etiologies relate to mechanical obstruction of the upper gut; however, disorders of propulsion including gastroparesis and intestinal pseudoobstruction also elicit prominent symptoms. Nausea and vomiting also are commonly reported by patients with irritable bowel syndrome and functional disorders of the upper gut (including chronic idiopathic nausea and functional vomiting).

ALTERED BOWEL HABITS

Altered bowel habits are common complaints of patients with GI disease. Constipation is reported as infrequent defecation, straining with defecation, passage of hard stools, or a sense of incomplete fecal evacuation. Causes of constipation include obstruction, motor disorders of the colon, medications, and endocrine diseases such as hypothyroidism and hyperparathyroidism. Diarrhea is reported as frequent defecation, passage of loose or watery stools, fecal urgency, or a similar sense of incomplete evacuation. The differential diagnosis of diarrhea is broad and includes infections, inflammatory causes, malabsorption, and medications. Irritable bowel syndrome produces constipation, diarrhea, or an alternating bowel pattern. Fecal mucus is common in irritable bowel syndrome, while pus characterizes inflammatory disease. Steatorrhea develops with malabsorption.

GI BLEEDING

Hemorrhage may develop from any gut organ. Most commonly, upper GI bleeding presents with melena or hematemesis, whereas lower GI bleeding produces passage of bright red or maroon stools. However, briskly bleeding upper sites can elicit voluminous red rectal bleeding, while slowly bleeding ascending colon sites may produce melena. Chronic slow GI bleeding may present with iron deficiency anemia. The most common upper GI causes of bleeding are ulcer disease, gastroduodenitis, and esophagitis. Other etiologies include portal hypertensive causes, malignancy, tears across the gastroesophageal junction, and vascular lesions. The most prevalent lower GI sources of hemorrhage include hemorrhoids, anal fissures, diverticula, ischemic colitis, and arteriovenous malformations. Other causes include neoplasm, inflammatory bowel disease, infectious colitis, drug-induced colitis, and other vascular lesions.

JAUNDICE

Jaundice results from prehepatic, intrahepatic, or posthepatic disease. Posthepatic causes of jaundice include biliary diseases such as choledocholithiasis, acute cholangitis, primary sclerosing cholangitis, other strictures, and neoplasm and pancreatic disorders, such as acute and chronic pancreatitis, stricture, and malignancy.

OTHER SYMPTOMS

Other symptoms are manifestations of GI disease. Dysphagia, odynophagia, and unexplained chest pain suggest esophageal disease. A globus sensation is reported with esophagopharyngeal conditions, but also occurs with functional GI disorders. Weight loss, anorexia, and fatigue are nonspecific symptoms of neoplastic, inflammatory, gut motility, pancreatic, small-bowel mucosal, and psychiatric conditions. Fever is reported with inflammatory illness, but malignancies also evoke febrile responses. GI disorders also produce extraintestinal symptoms. Inflammatory bowel disease is associated with hepatobiliary dysfunction, skin and eye lesions, and arthritis. Celiac disease may present with dermatitis herpetiformis. Jaundice can produce pruritus. Conversely, systemic diseases can have GI consequences. Systemic lupus may cause gut ischemia, presenting with pain or bleeding.

Overwhelming stress or severe burns may lead to gastric ulcer formation.

EVALUATION OF THE PATIENT WITH GASTROINTESTINAL DISEASE

Evaluation of the patient with GI disease begins with a careful history and examination. Subsequent investigation with a variety of tools designed to test gut structure or function are indicated in selected cases. Some patients exhibit normal findings on diagnostic testing. In these individuals, validated symptom profiles are employed to confidently diagnose a functional bowel disorder.

History

The history of the patient with suspected GI disease has several components. Symptom timing suggests specific etiologies. Symptoms of short duration commonly result from acute infection, toxin exposure, or abrupt inflammation or ischemia. Long-standing symptoms point to underlying chronic inflammatory or neoplastic conditions or functional bowel disorders. Symptoms from mechanical obstruction, ischemia, inflammatory bowel disease, and functional bowel disorders are worsened by meals. Conversely, ulcer symptoms may be relieved by eating or antacids. Symptom patterns and duration may suggest underlying etiologies. Ulcer pain occurs at intermittent intervals lasting weeks to months, while biliary colic has a sudden onset and lasts up to several hours. Pain from acute inflammation as with acute pancreatitis is severe and persists for days to weeks. Meals elicit diarrhea in some cases of inflammatory bowel disease and irritable bowel syndrome. Defecation relieves discomfort in inflammatory bowel disease and irritable bowel syndrome. Functional bowel disorders are exacerbated by stress. Sudden awakening from sound sleep suggests organic rather than functional disease. Diarrhea from malabsorption usually improves with fasting, while secretory diarrhea persists without oral intake.

Symptom relation to other factors narrows the list of diagnostic possibilities. Obstructive symptoms with prior abdominal surgery raise concern for adhesions, whereas loose stools after gastrectomy or gallbladder excision suggest dumping syndrome or postcholecystectomy diarrhea. Symptom onset after travel prompts a search for enteric infection. Medications may produce pain, altered bowel habits, or GI bleeding. Lower GI bleeding likely results from neoplasms, diverticula, or vascular lesions in an older person and from anorectal abnormalities or inflammatory bowel disease in a younger individual. Celiac disease is prevalent in people of northern European descent, while inflammatory bowel disease is more common in certain Jewish populations. A sexual history may raise concern for sexually transmitted diseases or immunodeficiency.

For more than two decades, working groups have been convened to devise symptom criteria to improve the confident diagnosis of functional bowel disorders and to minimize the numbers of unnecessary diagnostic tests performed. The most widely accepted symptom-based criteria are the Rome criteria. When tested against findings of structural investigations, the Rome criteria exhibit diagnostic specificities exceeding 90% for many of the functional bowel disorders.

Physical Examination

The physical exam complements information from the history. Abnormal vital signs provide diagnostic clues and determine the need for acute intervention. Fever suggests inflammation or neoplasm. Orthostasis is found with significant blood loss, dehydration, sepsis, or autonomic neuropathy. Skin, eye, or joint findings may point to specific diagnoses. Neck exam with swallowing assessment evaluates dysphagia. Cardiopulmonary disease may present with abdominal pain or nausea, thus lung and cardiac exams are important. Pelvic examination tests for a gynecologic source of abdominal pain. Rectal exam may detect blood, indicating gut mucosal injury or neoplasm or a palpable inflammatory mass in appendicitis. Metabolic conditions and gut motor disorders have associated peripheral neuropathy.

Inspection of the abdomen may reveal distention from obstruction, tumor, or ascites or vascular abnormalities with liver disease. Ecchymoses develop with severe pancreatitis. Auscultation can detect bruits or friction rubs from vascular disease or hepatic tumors. Loss of bowel sounds signifies ileus, while high-pitched, hyperactive sounds characterize intestinal obstruction. Percussion assesses liver size and can detect shifting dullness from ascites. Palpation assesses for hepatosplenomegaly as well as neoplastic or inflammatory masses. Abdominal exam is helpful in evaluating unexplained pain. Intestinal ischemia elicits severe pain but little tenderness. Patients with visceral pain may exhibit generalized discomfort, while those with parietal pain or peritonitis have directed pain, often with involuntary guarding, rigidity, or rebound. Patients with musculoskeletal abdominal wall pain may note tenderness exacerbated by Valsalva or straight-leg lift maneuvers.

Tools for Patient Evaluation

Laboratory, radiographic, and functional tests can assist in diagnosis of suspected GI disease. The GI tract also is amenable to internal evaluation with upper and lower endoscopy and to examination of luminal contents. Histopathologic exams of GI tissues complement these tests.

LABORATORY

Selected laboratory tests facilitate the diagnosis of GI disease. Iron-deficiency anemia suggests

mucosal blood loss, while vitamin B12 deficiency results from small-intestinal, gastric, or pancreatic disease. Either also can result from inadequate oral intake. Leukocytosis and increased sedimentation rates and C-reactive proteins are found in inflammatory conditions, while leukopenia is seen in viremic illness. Severe vomiting or diarrhea elicits electrolyte disturbances, acid-base abnormalities, and elevated blood urea nitrogen. Pancreaticobiliary or liver disease is suggested by elevated pancreatic or liver chemistries. Thyroid chemistries, cortisol, and calcium levels are obtained to exclude endocrinologic causes of GI symptoms. Pregnancy testing is considered for women with unexplained nausea. Serologic tests can screen for celiac disease, inflammatory bowel disease, rheumatologic diseases like lupus or scleroderma, and paraneoplastic dysmotility syndromes. Hormone levels are obtained for suspected endocrine neoplasia. Intraabdominal malignancies produce other tumor markers including the carcinoembryonic antigen CA 19-9 and α-fetoprotein. Blood testing also monitors medication therapy in some diseases, as with thiopurine metabolite levels in inflammatory bowel disease. Other body fluids are sampled under certain circumstances. Ascitic fluid is analyzed for infection, malignancy, or findings of portal hypertension. Cerebrospinal fluid is obtained for suspected central nervous system causes of vomiting. Urine samples screen for carcinoid, porphyria, and heavy metal intoxication.

LUMINAL CONTENTS

Luminal contents can be examined for diagnostic clues. Stool samples are cultured for bacterial pathogens, examined for leukocytes and parasites, or tested for *Giardia* antigen. Duodenal aspirates can be examined for parasites or cultured for bacterial overgrowth. Fecal fat is quantified in possible malabsorption. Stool electrolytes can be measured in diarrheal conditions. Laxative screens are done when laxative abuse is suspected. Gastric acid is quantified to rule out Zollinger-Ellison syndrome. Esophageal pH testing is done for refractory symptoms of acid reflux, whereas impedance techniques assess for nonacidic reflux. Pancreatic juice is analyzed for enzyme or bicarbonate content to exclude pancreatic exocrine insufficiency.

ENDOSCOPY

The gut is accessible with endoscopy, which can provide the diagnosis of the causes of bleeding, pain, nausea and vomiting, weight loss, altered bowel function, and fever. **Table 290-2** lists the most common indications for the major endoscopic procedures. Upper endoscopy evaluates the esophagus, stomach, and duodenum, while colonoscopy assesses the colon and distal ileum. Upper endoscopy is advocated as the initial structural test performed in patients with suspected ulcer disease, esophagitis, neoplasm, malabsorption, and Barrett's metaplasia because of its ability to directly visualize as well as biopsy the abnormality. Colonoscopy is the procedure of choice for colon cancer screening and surveillance as well as diagnosis of colitis secondary to infection, ischemia, radiation, and inflammatory bowel disease. Sigmoidoscopy examines the colon up to the splenic flexure and is currently used to exclude distal colonic inflammation or obstruction in young patients not at significant risk for colon cancer. For elusive GI bleeding secondary to arteriovenous malformations or superficial ulcers, small-intestinal examination is performed with push enteroscopy, capsule endoscopy, or double-balloon enteroscopy. Capsule endoscopy also can visualize small-intestinal Crohn's disease in individuals with negative barium radiography. Endoscopic retrograde cholangiopancreaticography (ERCP) provides diagnoses of pancreatic and biliary disease. Endoscopic ultrasound is useful for evaluating extent of disease in GI malignancy as well as exclusion of choledocholithiasis, evaluation of pancreatitis, drainage of pancreatic pseudocysts, and assessment of anal continuity.

Table 290-2 Common Indications for Endoscopy

Upper Endoscopy	Colonoscopy	Endoscopic Retrograde Cholangiopancreatography	Endoscopic Ultrasound	Capsule Endoscopy	Double Balloon Endoscopy
Dyspepsia despite treatment Dyspepsia with signs of organic disease Refractory vomiting Dysphagia Upper GI bleeding Anemia Weight loss Malabsorption	Cancer screening Lower GI bleeding Anemia Diarrhea Polypectomy Obstruction Biopsy radiologic abnormality Cancer surveillance: family history prior	Jaundice Postbiliary surgery complaints Cholangitis Gallstone pancreatitis Pancreatic/biliary/ampullary tumor Unexplained pancreatitis Pancreatitis with unrelenting pain Fistulas Biopsy radiologic abnormality Pancreaticobiliary drainage	Staging of malignancy Characterize and biopsy submucosal mass Bile duct stones Chronic pancreatitis Drain pseudocyst Large gastric folds Anal	Obscure GI bleeding Suspected Crohn's disease of the small intestine	Ablation of small-intestinal bleeding sources Biopsy of suspicious small-intestinal masses/ulcers

Malabsorption Biopsy radiologic abnormality Polypectomy Place gastrostomy Barrett's surveillance Palliate neoplasm Sample duodenal tissue/fluid Remove foreign body Endoscopic mucosal resection or ablation of dysplastic Barrett's mucosa Place stent across stenosis	prior polyp/cancer, colitis Palliate neoplasm Remove foreign body Place stent across stenosis	Sample bile Sphincter of Oddi manometry	continuity		

RADIOGRAPHY/NUCLEAR MEDICINE

Radiographic tests evaluate diseases of the gut and extraluminal structures. Oral or rectal contrast agents like barium provide mucosal definition from the esophagus to the rectum. Contrast radiography also assesses gut transit and pelvic floor dysfunction. Barium swallow is the initial procedure for evaluation of dysphagia to exclude subtle rings or strictures and assess for achalasia, whereas small-bowel contrast radiology reliably diagnoses intestinal tumors and Crohn's ileitis. Contrast enemas are performed when colonoscopy is unsuccessful or contraindicated. Ultrasound and computed tomography (CT) evaluate regions not accessible by endoscopy or contrast studies, including the liver, pancreas, gallbladder, kidneys, and retroperitoneum. These tests are useful for diagnosis of mass lesions, fluid collections, organ enlargement, and in the case of ultrasound gallstones. CT and magnetic resonance (MR) colonography are being evaluated as alternatives to colonoscopy for colon cancer screening. MR imaging assesses the pancreaticobiliary ducts to exclude neoplasm, stones, and sclerosing cholangitis, and the liver to characterize benign and malignant tumors. Specialized CT or MR enterography can assess intensity of inflammatory bowel disease. Angiography excludes mesenteric ischemia and determines spread of malignancy. Angiographic techniques also access the biliary tree in obstructive jaundice. CT and MR techniques can be used to screen for mesenteric occlusion, thereby limiting exposure to angiographic dyes. Positron emission tomography can facilitate distinguishing malignant from benign disease in several organ systems.

Scintigraphy both evaluates structural abnormalities and quantifies luminal transit. Radionuclide bleeding scans localize bleeding sites in patients with brisk hemorrhage so that therapy with endoscopy, angiography, or surgery may be directed. Radiolabeled leukocyte scans can search for intraabdominal abscesses not visualized on CT. Biliary scintigraphy is complementary to ultrasound in the assessment of cholecystitis. Scintigraphy to quantify esophageal and gastric emptying are well established, while techniques to measure small intestinal or colonic transit are less widely used.

HISTOPATHOLOGY

Gut mucosal biopsies obtained at endoscopy evaluate for inflammatory, infectious, and neoplastic disease. Deep rectal biopsies assist with diagnosis of Hirschsprung's disease or amyloid. Liver biopsy is indicated in cases with abnormal liver chemistries, unexplained jaundice, following liver transplant to exclude rejection, and to characterize the degree of inflammation in patients with chronic viral hepatitis prior to initiating antiviral therapy. Biopsies obtained during CT or ultrasound can evaluate for other intraabdominal conditions not accessible by endoscopy.

FUNCTIONAL TESTING

Tests of gut function provide important data when structural testing is nondiagnostic. In addition to gastric acid and pancreatic function testing, functional testing of motor activity is provided by manometric techniques. Esophageal manometry is useful for suspected achalasia, whereas small-intestinal manometry tests for pseudoobstruction. A wireless motility capsule is now available to

measure transit and contractile activity in the stomach, small intestine, and colon in a single test. Anorectal manometry with balloon expulsion testing is employed for unexplained incontinence or constipation from outlet dysfunction. Anorectal manometry and electromyography also assess anal function in fecal incontinence. Biliary manometry tests for sphincter of Oddi dysfunction with unexplained biliary pain. Measurement of breath hydrogen while fasting and after oral mono- or oligosaccharide challenge can screen for carbohydrate intolerance and small-intestinal bacterial overgrowth.

Treatment: Gastrointestinal Disease

Management options for the patient with GI disease depend on the cause of symptoms. Available treatments include modifications in dietary intake, medications, interventional endoscopy or radiology techniques, surgery, and therapies directed to external influences.

NUTRITIONAL MANIPULATION

Dietary modifications for GI disease include treatments that only reduce symptoms, therapies that correct pathologic defects, and measures that replace normal food intake with enteral or parenteral formulations. Changes that improve symptoms but do not reverse an organic abnormality include lactose restriction for lactase deficiency, liquid meals in gastroparesis, carbohydrate restrictions with dumping syndrome, and high-fiber diets in irritable bowel syndrome. The gluten-free diet for celiac disease exemplifies a modification that serves as primary therapy to reduce mucosal inflammation. Enteral medium-chain triglycerides replace normal fats with short-gut syndrome or severe ileal disease. Perfusion of liquid meals through a gastrostomy is performed in those who cannot swallow safely. Enteral feeding through a jejunostomy is considered for gastric dysmotility syndromes that preclude feeding into the stomach. Intravenous hyperalimentation is employed for individuals with generalized gut malfunction who cannot tolerate or who cannot be sustained with enteral nutrition.

PHARMACOTHERAPY

Several medications are available to treat GI diseases. Considerable health care resources are expended on over-the-counter remedies. Many prescription drug classes are offered as short-term or continuous therapy of GI illness. A plethora of alternative treatments have gained popularity in GI conditions for which traditional therapies provide incomplete relief.

Over-the-Counter Agents

Over-the-counter agents are reserved for mild GI symptoms. Antacids and histamine H_2 antagonists decrease symptoms in gastroesophageal reflux and dyspepsia, whereas antiflatulents and adsorbents reduce gaseous symptoms. More potent acid inhibitors such as proton pump inhibitors are now available over the counter for treatment of chronic gastroesophageal reflux disease (GERD). Fiber supplements, stool softeners, enemas, and laxatives are used for constipation. Laxatives are categorized as stimulants, osmotic agents (including isotonic preparations containing polyethylene glycol), and poorly absorbed sugars. Nonprescription antidiarrheal agents include bismuth subsalicylate, kaolin-pectin combinations, and loperamide. Supplemental enzymes include lactase pills for lactose intolerance and bacterial α-galactosidase to treat excess gas. In general, use of a nonprescription preparation for more than a short time for chronic persistent symptoms should be supervised by a health care provider.

Prescription Drugs

Prescription drugs for GI diseases are a major focus of attention from pharmaceutical companies. Potent acid suppressants including drugs that inhibit the proton pump are advocated for acid reflux when over-the-counter preparations are inadequate. Cytoprotective agents rarely are used for upper gut ulcers. Prokinetic drugs stimulate GI propulsion in gastroparesis and pseudoobstruction. Prosecretory drugs are prescribed for constipation refractory to other agents. Prescription antidiarrheals include opiate drugs, anticholinergic antispasmodics, tricyclics, bile acid binders, and serotonin antagonists. Antispasmodics and antidepressants also are useful for functional abdominal pain, whereas narcotics are used for pain control in organic conditions such as disseminated malignancy and chronic pancreatitis. Antiemetics in several classes reduce nausea and vomiting. Potent pancreatic enzymes decrease malabsorption and pain from pancreatic disease. Antisecretory drugs such as the somatostatin analogue octreotide treat hypersecretory states. Antibiotics treat ulcer disease secondary to *Helicobacter pylori* , infectious diarrhea, diverticulitis, intestinal bacterial overgrowth, and Crohn's disease. Some cases of irritable bowel syndrome (especially those with diarrhea) respond to nonabsorbable antibiotic therapy. Anti-inflammatory and immunosuppressive drugs are used in ulcerative colitis, Crohn's disease, microscopic colitis, refractory celiac disease, and gut vasculitis. Chemotherapy with or without radiotherapy is offered for GI malignancies. Most GI carcinomas respond poorly to such therapy, whereas lymphomas may be cured with such intervention.

Alternative Therapies

Alternative treatments are marketed to treat selected GI symptoms. Ginger, acupressure, and acustimulation have been advocated for nausea, while pyridoxine has been investigated for nausea of first-trimester pregnancy. Probiotics containing active bacterial cultures are used as adjuncts in some

cases of infectious diarrhea and irritable bowel syndrome. Probiotics that selectively nourish benign luminal bacteria may ultimately show benefit in functional disorders as well. Low-potency pancreatic enzyme preparations are sold as general digestive aids but have little evidence to support their efficacy.

ENTERIC THERAPIES/INTERVENTIONAL ENDOSCOPY AND RADIOLOGY

Simple luminal interventions are commonly performed for GI diseases. Nasogastric tube suction decompresses the upper gut in ileus or mechanical obstruction. Nasogastric lavage of saline or water in the patient with upper GI hemorrhage determines the rate of bleeding and helps evacuate blood prior to endoscopy. Enteral feedings can be initiated through a nasogastric or nasoenteric tube. Enemas relieve fecal impaction or assist in gas evacuation in acute colonic pseudoobstruction. A rectal tube can be left in place to vent the distal colon in colonic pseudoobstruction and other colonic distention disorders.

In addition to its diagnostic role, endoscopy has therapeutic capabilities in certain settings. Cautery techniques can stop hemorrhage from ulcers, vascular malformations, and tumors. Injection with vasoconstrictor substances or sclerosants is used for bleeding ulcers, vascular malformations, varices, and hemorrhoids. Endoscopic encirclement of varices and hemorrhoids with constricting bands stops hemorrhage from these sites, while endoscopically placed clips can occlude arterial bleeding sites. Endoscopy can remove polyps or debulk lumen-narrowing malignancies. Endoscopic mucosal resection and radio frequency techniques can remove or ablate some cases of Barrett's esophagus with dysplasia. Endoscopic sphincterotomy of the ampulla of Vater relieves symptoms of choledocholithiasis. Obstructions of the gut lumen and pancreaticobiliary tree are relieved by endoscopic dilatation or placement of plastic or expandable metal stents. In cases of acute colonic pseudoobstruction, colonoscopy is employed to withdraw luminal gas. Finally, endoscopy is commonly used to insert feeding tubes.

Radiologic measures also are useful in GI disease. Angiographic embolization or vasoconstriction decreases bleeding from sites not amenable to endoscopic intervention. Dilatation or stenting with fluoroscopic guidance relieves luminal strictures. Contrast enemas can reduce volvulus and evacuate air in acute colonic pseudoobstruction. CT and ultrasound help drain abdominal fluid collections, in many cases obviating the need for surgery. Percutaneous transhepatic cholangiography relieves biliary obstruction when ERCP is contraindicated. Lithotripsy can fragment gallstones in patients who are not candidates for surgery. In some instances, radiologic approaches offer advantages over endoscopy for gastroenterostomy placement. Finally, central venous catheters for parenteral nutrition may be placed using radiographic techniques.

SURGERY

Surgery is performed to cure disease, control symptoms without cure, maintain nutrition, or palliate unresectable neoplasm. Medication-unresponsive ulcerative colitis, diverticulitis, cholecystitis, appendicitis, and intraabdominal abscess are curable with surgery, while only symptom control without cure is possible with Crohn's disease. Surgery is mandated for ulcer complications such as bleeding, obstruction, or perforation and intestinal obstructions that persist after conservative care. Fundoplication of the gastroesophageal junction is performed for severe ulcerative esophagitis and drug-refractory symptomatic acid reflux. Achalasia responds to operations to relieve lower esophageal sphincter pressure. Operations for motor disorders have been introduced including implanted electrical stimulators for gastroparesis and electrical devices and artificial sphincters for fecal incontinence. Surgery may be needed to place a jejunostomy for long-term enteral feedings. The threshold for performing surgery depends on the clinical setting. In all cases, the benefits of operation must be weighed against the potential for postoperative complications.

THERAPY DIRECTED TO EXTERNAL INFLUENCES

In some conditions, GI symptoms respond to treatments directed outside the gut. Psychological therapies including psychotherapy, behavior modification, hypnosis, and biofeedback have shown efficacy in functional bowel disorders. Patients with significant psychological dysfunction and those with little response to treatments targeting the gut are likely to benefit from this form of therapy.

FURTHER READINGS

KAHRILAS PJ et al: American Gastroenterological Association Institute technical review on the management of gastroesophageal reflux disease. Gastroenterology 135:1392, 2008[PMID: 18801365] [Full Text]

LICHTENSTEIN GR et al: American Gastroenterological Association Institute technical review on corticosteroids, immunomodulators, and infliximab in inflammatory bowel disease. Gastroenterology 130:940, 2006[PMID: 16530532] [Full Text]

LONGSTRETH GF et al: Functional bowel disorders. Gastroenterology 130:1480, 2006[PMID: 16678561] [Full Text]

REX DK et al: American College of Gastroenterology guidelines for colorectal cancer screening 2008. Am J Gastroenterol 104:739, 2009[PMID: 19240699] [Full Text]

YAMADA T (ed): *Textbook of Gastroenterology* , 5th ed. Hoboken, NJ, Wiley-Blackwell, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GASTROINTESTINAL ENDOSCOPY: INTRODUCTION

Gastrointestinal endoscopy has been attempted for over 200 years, but the introduction of semirigid gastroscopes in the middle of the twentieth century marked the dawn of the modern endoscopic era. Since then, rapid advances in endoscopic technology have led to dramatic changes in the diagnosis and treatment of many digestive diseases. Innovative endoscopic devices and new endoscopic treatment modalities continue to expand the use of endoscopy in patient care.

Flexible endoscopes provide either an optical image (transmitted over fiberoptic bundles) or an electronic video image (generated by a charge-coupled device in the tip of the endoscope). Operator controls permit deflection of the endoscope tip; fiberoptic bundles bring light to the tip of the endoscope; and working channels allow washing, suctioning, and the passage of instruments. Progressive changes in the diameter and stiffness of endoscopes have improved the ease and patient tolerance of endoscopy.

ENDOSCOPIC PROCEDURES

Upper Endoscopy

Upper endoscopy, also referred to as esophagogastroduodenoscopy (EGD), is performed by passing a flexible endoscope through the mouth into the esophagus, stomach, bulb, and second duodenum. The procedure is the best method of examining the upper gastrointestinal mucosa. While the upper gastrointestinal radiographic series has similar accuracy for diagnosis of duodenal ulcer **(Fig. 291-1)**, EGD is superior for detection of gastric ulcers **(Fig. 291-2)** and flat mucosal lesions such as Barrett's esophagus **(Fig. 291-3)**, and it permits directed biopsy and endoscopic therapy. Intravenous conscious sedation is given to most patients in the United States to ease the anxiety and discomfort of the procedure, although in many countries EGD is routinely performed with topical pharyngeal anesthesia only. Patient tolerance of unsedated EGD is improved by the use of an ultrathin, 5-mm diameter endoscope that can be passed transorally or transnasally.

Figure 291-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

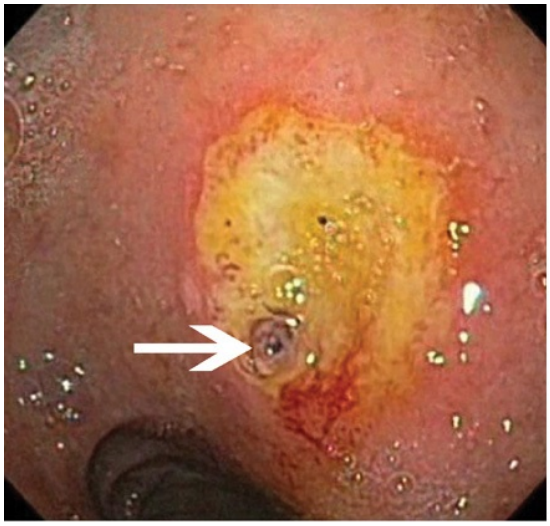

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Duodenal ulcers. Ulcer with a clean base. Ulcer with a visible vessel (*arrow*) in a patient with recent hemorrhage.

Figure 291-2

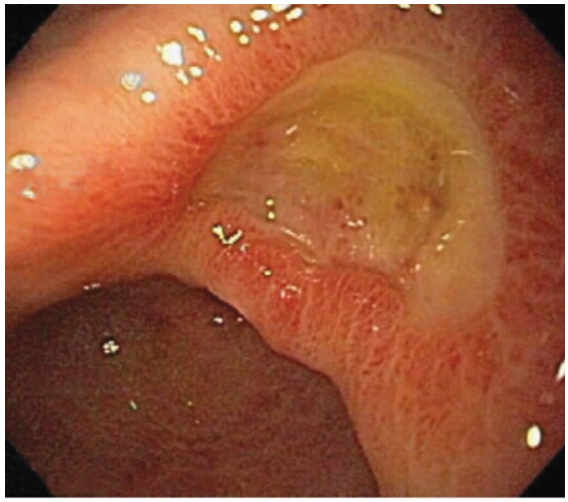

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

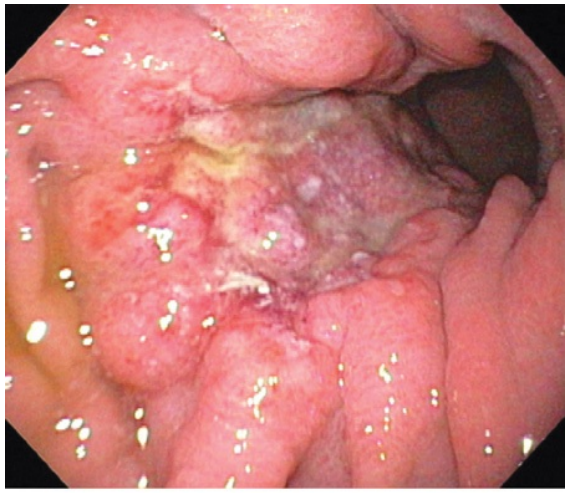

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gastric ulcers.*A.* Benign gastric ulcer. ***B.*** Malignant gastric ulcer involving greater curvature of stomach.

Figure 291-3

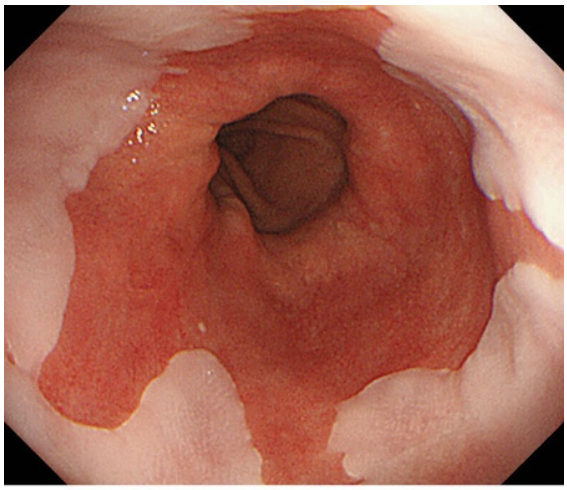

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

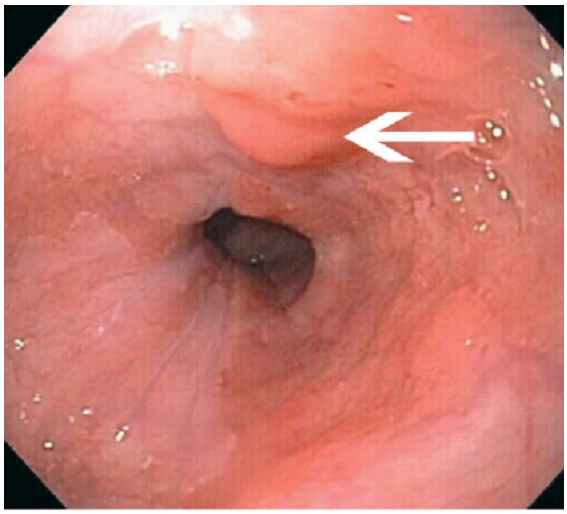

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

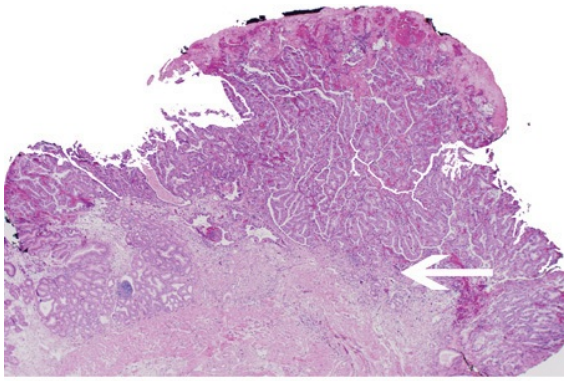

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

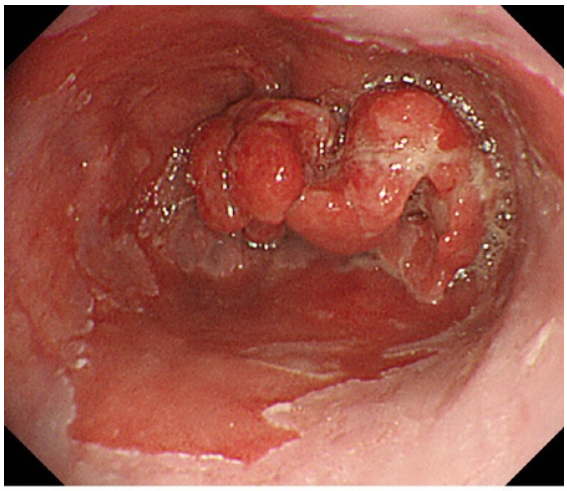

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Barrett's esophagus. Pink tongues of Barrett's mucosa extending proximally from the gastroesophageal junction. Barrett's esophagus with a suspicious nodule (*arrow*) identified during endoscopic surveillance. Histologic finding of intramucosal adenocarcinoma in the endoscopically resected nodule. Tumor extends into the esophageal submucosa (*arrow*). Barrett's esophagus with locally advanced adenocarcinoma.

Colonoscopy

Colonoscopy is performed by passing a flexible colonoscope through the anal canal into the rectum and colon. The cecum is reached in >95% of cases, and the terminal ileum can often be examined. Colonoscopy is the gold standard for diagnosis of colonic mucosal disease. Colonoscopy has greater sensitivity than barium enema for colitis **(Fig. 291-4)**, polyps **(Fig. 291-5)**, and cancer **(Fig. 291-6)**. CT colonography is an emerging technology that rivals colonoscopy's accuracy for detection of polyps and cancer. Conscious sedation is usually given before colonoscopy in the United States, although a willing patient and a skilled examiner can complete the procedure without sedation in many cases.

Figure 291-4

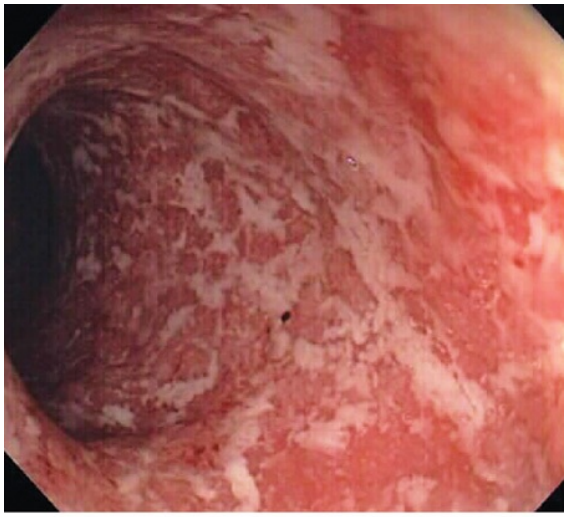

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

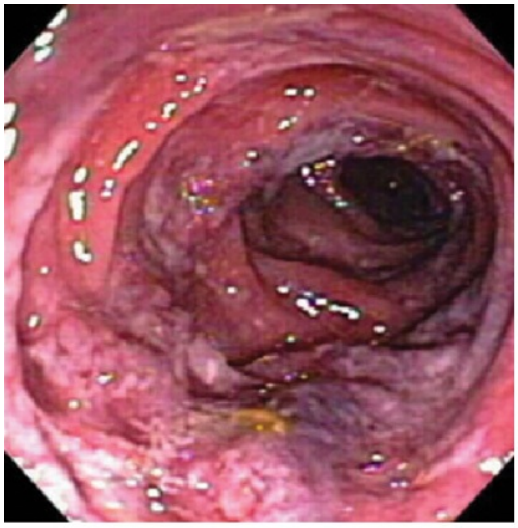

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Causes of colitis.*A.* Chronic ulcerative colitis with diffuse ulcerations and exudates. ***B.*** Severe Crohn's colitis with deep ulcers. ***C.*** Pseudomembranous colitis with yellow, adherent pseudomembranes. ***D.*** Ischemic colitis with patchy mucosal edema, subepithelial hemorrhage, and cyanosis.

Figure 291-5

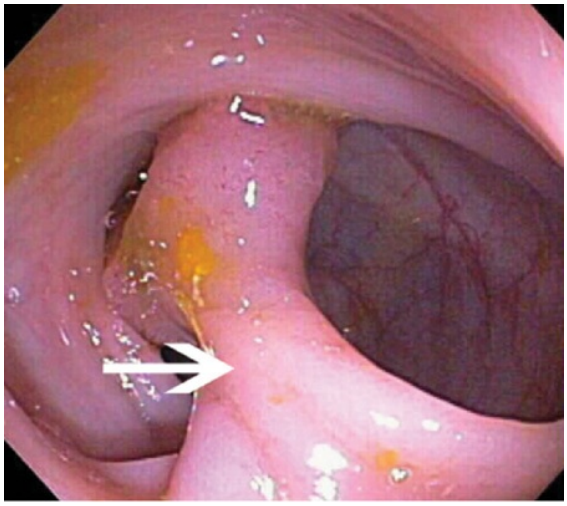

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

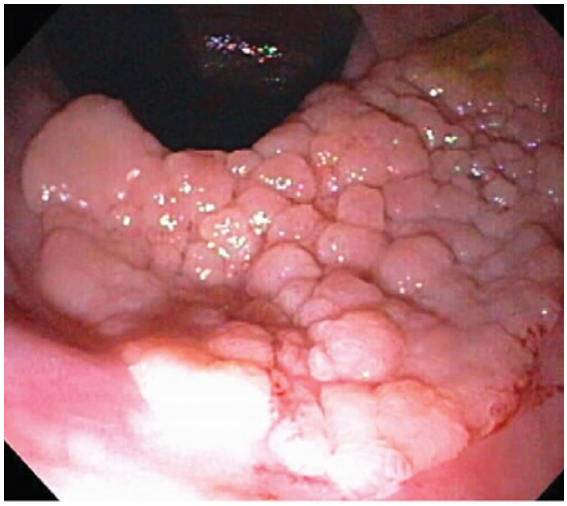

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Colonic polyps. Pedunculated colon polyp on a thick stalk covered with normal mucosa (*arrow*). Sessile rectal polyp.

Figure 291-6

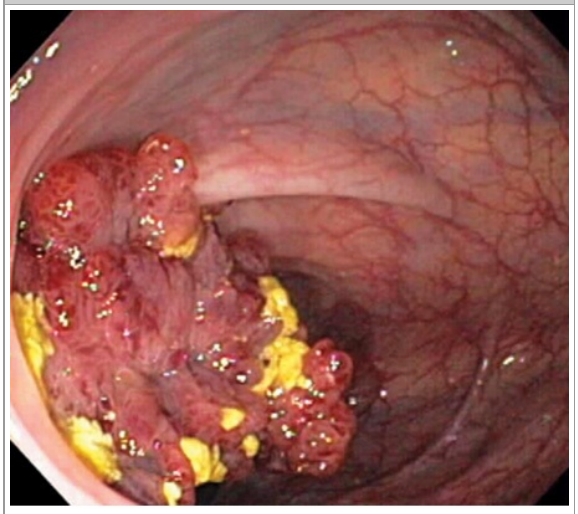

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Colon adenocarcinoma growing into the lumen.

Flexible Sigmoidoscopy

Flexible sigmoidoscopy is similar to colonoscopy but visualizes only the rectum and a variable portion of the left colon, typically to 60 cm from the anal verge. This procedure causes abdominal cramping, but it is brief and is usually performed without sedation. Flexible sigmoidoscopy is primarily used for evaluation of diarrhea and rectal outlet bleeding.

Small-Bowel Endoscopy

Three techniques are currently used to evaluate the small intestine, most often in patients presenting with presumed small-bowel bleeding. For *capsule endoscopy* the patient swallows a disposable capsule that contains a complementary metal oxide silicon (CMOS) chip camera. Color still images **(Fig. 291-7)** are transmitted wirelessly to an external receiver at several frames per second until the capsule's battery is exhausted or it is passed into the toilet. Although capsule endoscopy enables visualization of the jejunal and ileal mucosa beyond the reach of a conventional endoscope, it remains solely a diagnostic procedure at present.

Figure 291-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Capsule endoscopy image of jejunal vascular ectasia.

Push enteroscopy is performed with a long endoscope similar in design to an upper endoscope. The enteroscope is pushed down the small bowel, sometimes with the help of a stiffening overtube that extends from the mouth to the small intestine. The proximal to mid-jejunum is usually reached, and the endoscope's instrument channel allows for biopsies or endoscopic therapy.

Deeper insertion into the small bowel can be accomplished by *single-* or *double-balloon enteroscopy* or *spiral enteroscopy* **(Fig. 291-8)**. These instruments enable pleating of the small intestine onto an overtube **(Video e36-1)**. With balloon-assisted enteroscopy, the entire small bowel can be visualized in some patients when both the oral and anal routes of insertion are used. Biopsies and endoscopic therapy can be performed throughout the visualized small bowel **(Fig. 291-9)**.

Figure 291-8

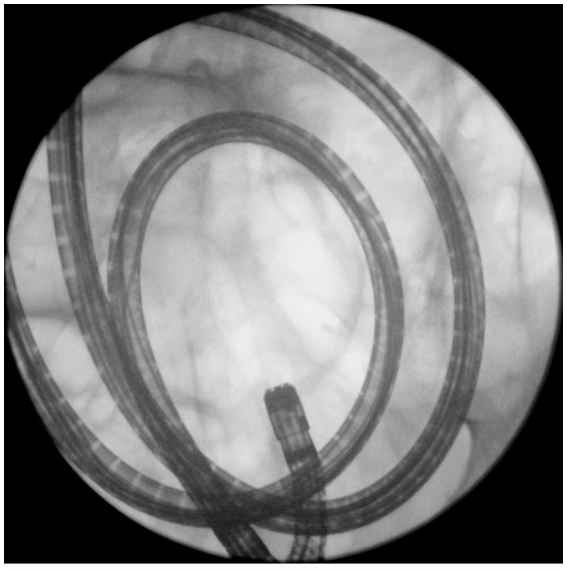

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiograph of a double-balloon enteroscope in the small intestine.

Figure 291-9

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

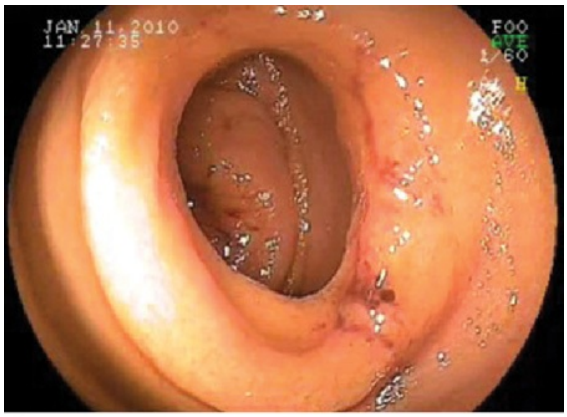

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nonsteroidal anti-inflammatory (NSAID)-induced proximal ileal stricture diagnosed by double-balloon endoscopy.*A.* Ileal stricture causing obstructive symptoms. ***B.*** Balloon dilatation of the ileal stricture. ***C.*** Appearance of stricture after dilatation.

Endoscopic Retrograde Cholangiopancreatography (ERCP)

During ERCP, a side-viewing endoscope is passed through the mouth to the duodenum, the ampulla of Vater is identified and cannulated with a thin plastic catheter, and radiographic contrast material is injected into the bile duct and pancreatic duct under fluoroscopic guidance **(Fig. 291-10)**. When indicated, the sphincter of Oddi can be opened using the technique of endoscopic sphincterotomy **(Fig. 291-11)**. Stones can be retrieved from the ducts, biopsies can be performed, strictures can be dilated and/or stented **(Fig. 291-12)**, and ductal leaks can be stented **(Fig. 291-13)**. ERCP is often performed for therapy but remains important in diagnosis, especially for ductal strictures and bile duct stones.

Figure 291-10

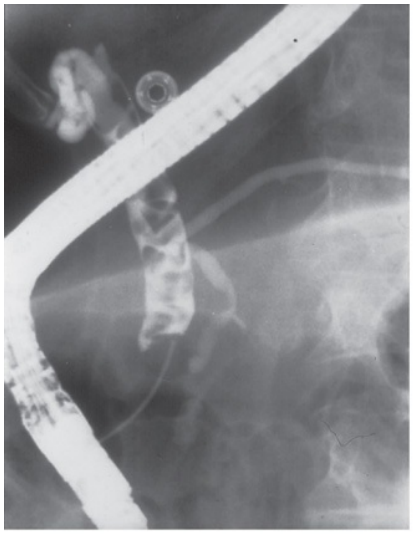

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

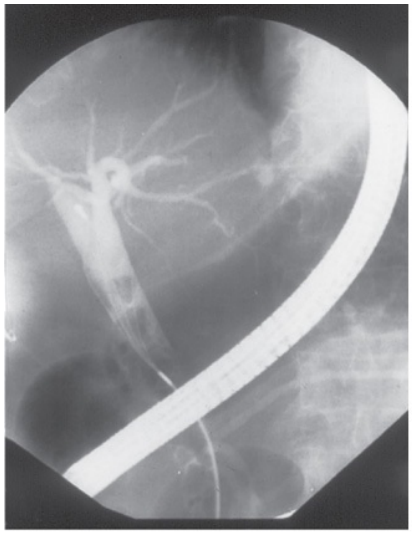

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Endoscopic retrograde cholangiopancreatography (ERCP) for bile duct stones with cholangitis. Faceted bile duct stones are demonstrated in the common bile duct. After endoscopic sphincterotomy, the stones are extracted with a Dormia basket. A small abscess communicates with the left hepatic duct.

Figure 291-11

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

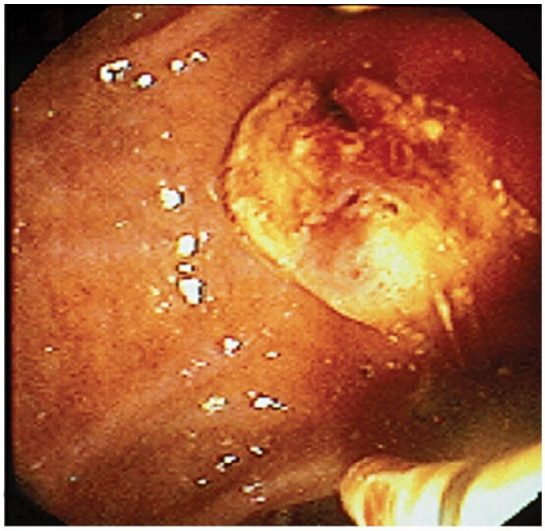

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Endoscopic sphincterotomy.*A.* A normal-appearing ampulla of Vater. ***B.*** Sphincterotomy is performed with electrocautery. ***C.*** Bile duct stones are extracted with a balloon catheter. ***D.*** Final appearance of the sphincterotomy.

Figure 291-12

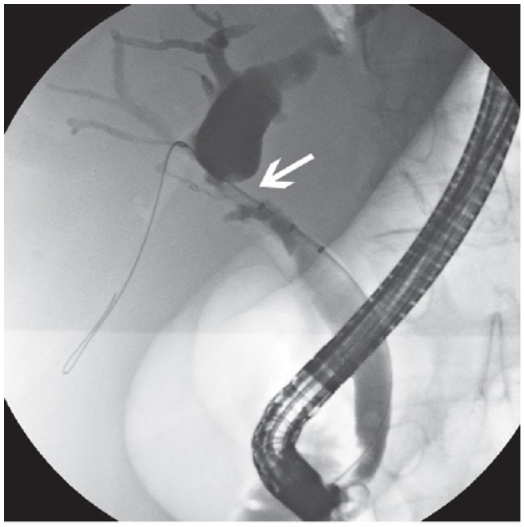

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

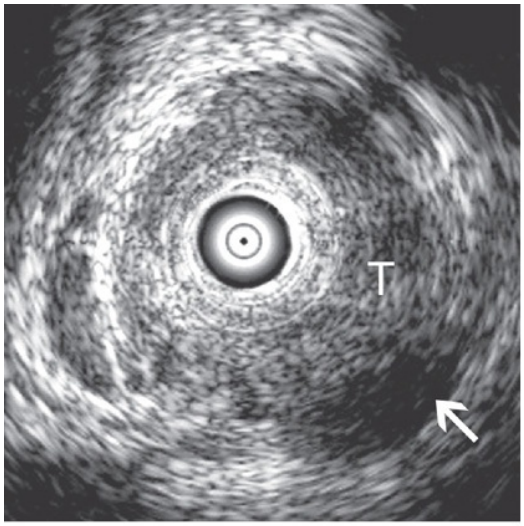

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

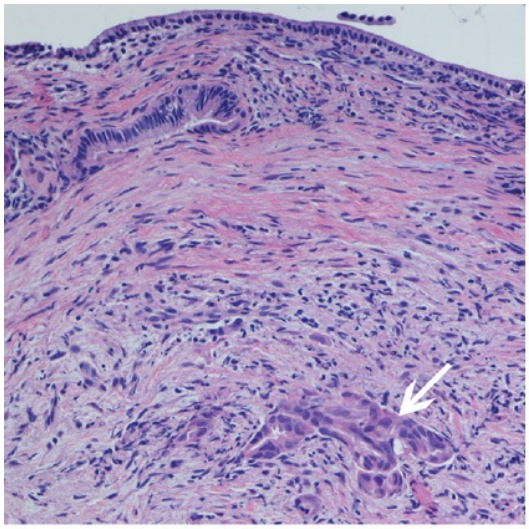

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

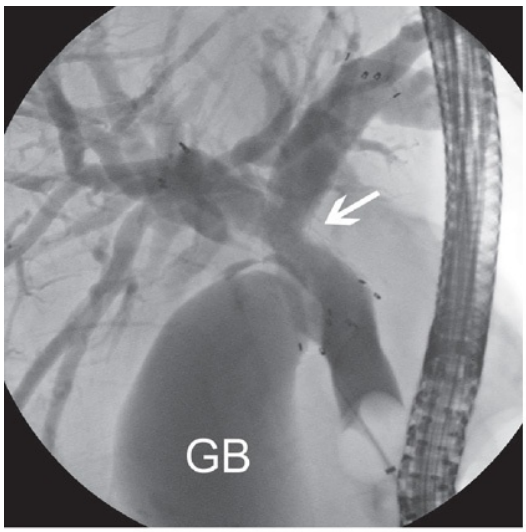

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Endoscopic diagnosis, staging, and palliation of hilar cholangiocarcinoma. Endoscopic retrograde cholangiopancreatography (ERCP) in a patient with obstructive jaundice demonstrates a malignant-appearing stricture of the biliary confluence extending into the left and right intrahepatic ducts. Intraductal ultrasound of the biliary stricture demonstrates

marked bile duct wall thickening due to tumor (T) with partial encasement of the hepatic artery (*arrow*). Intraductal biopsy obtained during ERCP demonstrates malignant cells infiltrating the submucosa of the bile duct wall (*arrow*). Endoscopic placement of bilateral self-expanding metal stents (arrow) relieves the biliary obstruction. GB, gallbladder. (*Image C courtesy of Dr. Thomas Smyrk; with permission.)*

Figure 291-13

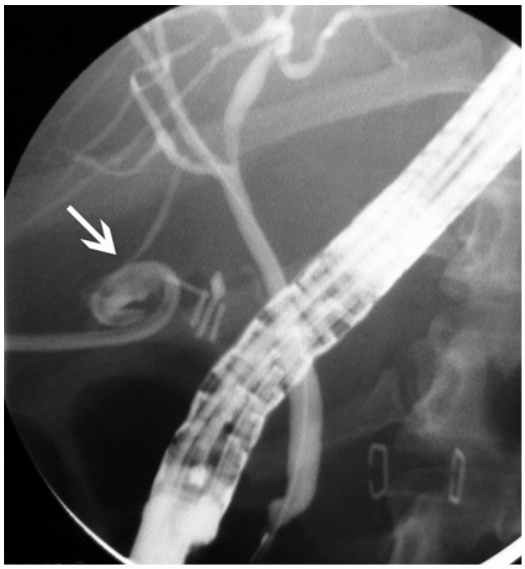

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bile leak (*arrow*) from a duct of Luschka after laparoscopic cholecystectomy. Contrast leaks from a small right intrahepatic duct into the gallbladder fossa, then flows into the pigtail of a percutaneous drainage catheter.

Endoscopic Ultrasound (Eus)

EUS utilizes high-frequency ultrasound transducers incorporated into the tip of a flexible endoscope. Ultrasound images are obtained of the gut wall and adjacent organs, vessels, and lymph nodes. By sacrificing depth of ultrasound penetration and bringing the ultrasound transducer close to the area of interest via endoscopy, high-resolution images are obtained. EUS provides the most accurate preoperative local staging of esophageal, pancreatic, and rectal malignancies **(Fig. 291-14)**, although it does not detect most distant metastases. EUS is also useful for diagnosis of bile duct stones, gallbladder disease, submucosal gastrointestinal lesions, and chronic pancreatitis. Fine-needle aspirates and core biopsies of masses and lymph nodes in the posterior mediastinum, abdomen, pancreas, retroperitoneum, and pelvis can be obtained under EUS guidance **(Fig. 291-15)**.

Figure 291-14

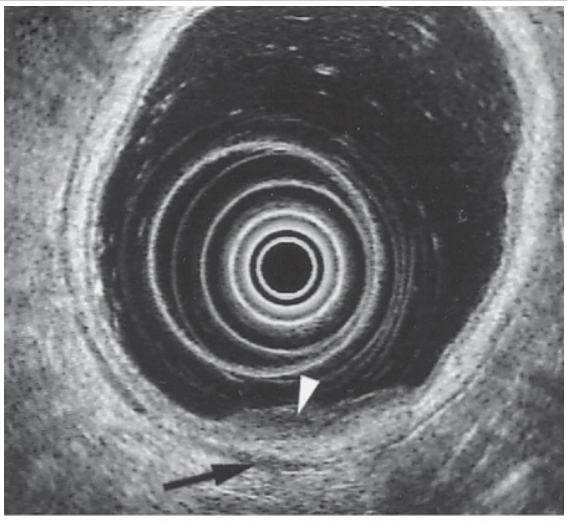

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

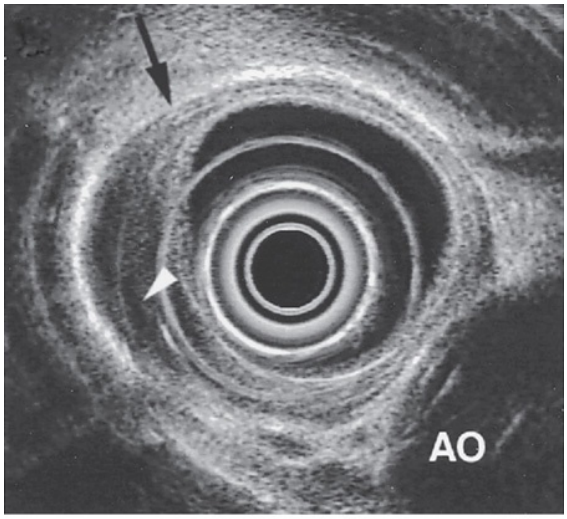

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

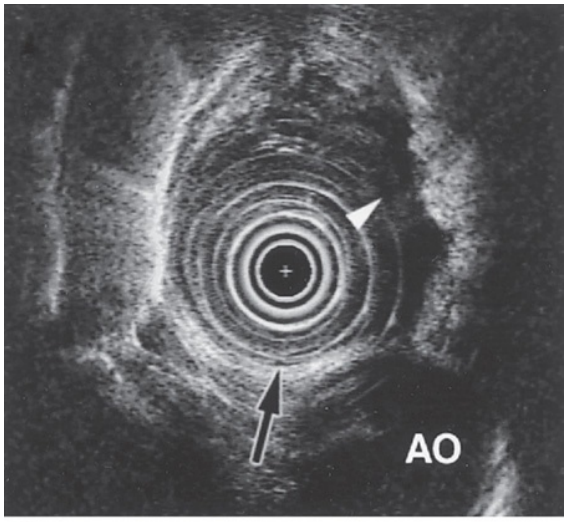

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Local staging of gastrointestinal cancers with endoscopic ultrasound. In each example the white arrowhead marks the primary tumor and the black arrow indicates the muscularis propria (mp) of the intestinal wall. ***A.*** T1 gastric cancer. The tumor does not invade the mp. ***B.*** T2 esophageal cancer. The tumor invades the mp. ***C.*** T3 esophageal cancer. The tumor extends through the mp into the surrounding tissue, and focally abuts the aorta. AO, aorta.

Figure 291-15

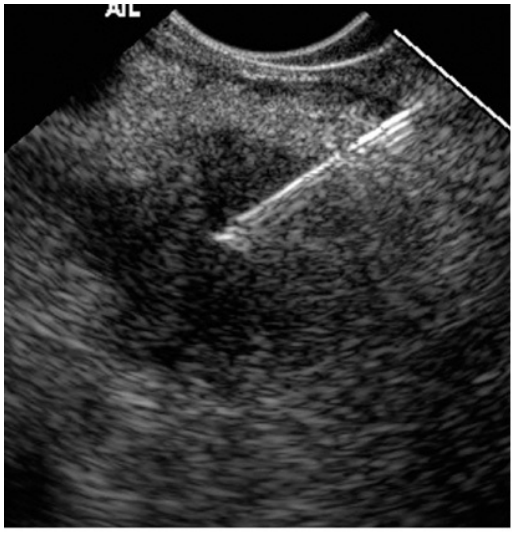

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

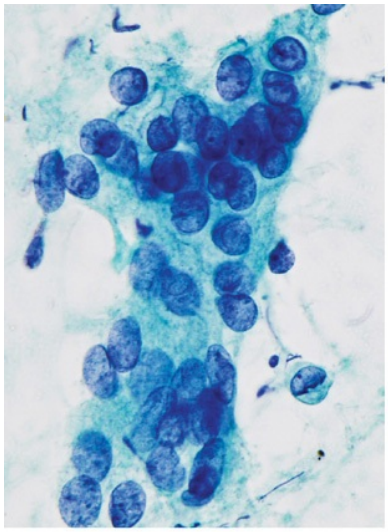

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Endoscopic ultrasound (EUS)-guided fine-needle aspiration (FNA). ***A.*** Ultrasound image of a 22-gauge needle passed through the duodenal wall and positioned in a hypoechoic pancreatic head mass. ***B.*** Micrograph of aspirated malignant cells. *(Image B courtesy of Dr. Michael R. Henry; with permission.)*

Natural Orifice Transluminal Endoscopic Surgery (Notes)

NOTES is an evolving collection of endoscopic methods that entail passage of an endoscope or its accessories through the wall of the gastrointestinal tract (e.g., stomach) to perform diagnostic or therapeutic interventions. Some NOTES procedures, such as percutaneous endoscopic gastrostomy (PEG) or endoscopic necrosectomy of pancreatic necrosis, are established clinical procedures **(Video e36-2)**; others, such as endoscopic appendectomy, cholecystectomy, and tubal ligation, are in development, and their ultimate clinical application is presently unclear. NOTES is currently an area of intense innovation and endoscopic research.

RISKS OF ENDOSCOPY

Medications used during conscious sedation may cause respiratory depression or allergic reactions. All endoscopic procedures carry some risk of bleeding and gastrointestinal perforation. These risks are quite low with diagnostic upper endoscopy and colonoscopy (<1:1000 procedures), although the risk is as high as 2:100 when therapeutic procedures such as polypectomy, control of hemorrhage, or stricture dilatation are performed. Bleeding and perforation are rare with flexible sigmoidoscopy. The risks for diagnostic EUS (without needle aspiration) are similar to the risks for diagnostic upper endoscopy.

Infectious complications are unusual with most endoscopic procedures. Some procedures carry a higher incidence of postprocedure bacteremia, and prophylactic antibiotics may be indicated **(Table 291-1)**.

Table 291-1 Antibiotic Prophylaxis for Endoscopic Procedures

Patient Condition	Procedure Contemplated	Goal of Prophylaxis	Periprocedural Antibiotic Prophylaxis
All cardiac conditions	Any endoscopic procedure	Prevention of infective endocarditis	Not indicated

Bile-duct obstruction in the absence of cholangitis	ERCP with complete drainage	Prevention of cholangitis	Not recommended[a]
Bile-duct obstruction in absence of cholangitis	ERCP with anticipated incomplete drainage (e.g., PSC, hilar strictures)	Prevention of cholangitis	Recommended; continue antibiotics after the procedure
Sterile pancreatic fluid collection (e.g., pseudocyst, necrosis), which communicates with pancreatic duct	ERCP	Prevention of cyst infection	Recommended
Sterile pancreatic fluid collection	Transmural drainage	Prevention of cyst infection	Recommended
Solid lesion along upper GI tract	EUS-FNA	Prevention of local infection	Not recommended
Solid lesion along lower GI tract	EUS-FNA	Prevention of local infection	Insufficient data to make firm recommendation[b]
Cystic lesions along GI tract (including mediastinum)	EUS-FNA	Prevention of cyst infection	Recommended
All patients	Percutaneous endoscopic feeding tube placement	Prevention of peristomal infection	Recommended
Cirrhosis with acute GI bleeding	Required for all patients, regardless of endoscopic procedures	Prevention of infectious complications and reduction of mortality	Upon admission[c]
Synthetic vascular graft and other nonvalvular cardiovascular devices	Any endoscopic procedure	Prevention of graft and device infection	Not recommended[d]
Prosthetic joints	Any endoscopic procedure	Prevention of septic arthritis	Not recommended[e]

[a]Low rates of bacteremia and local infection.

[b]Endoscopists may choose on a case by case basis.

[c]Risk for bacterial infection associated with cirrhosis and GI bleeding is well established.

[d]No reported cases of infection associated with endoscopy.

[e]Very low risk of infection.

Abbreviations: ERCP, endoscopic retrograde cholangiopancreatography; EUS-FNA, endoscopic ultrasound–fine-needle aspiration; PSC, primary sclerosing cholangitis.

Source: Adapted from S Banerjee et al: Gastrointest Endosc 67:719, 2008; with permission from Elsevier.

ERCP carries additional risks. Pancreatitis occurs in about 5% of patients undergoing ERCP and in up to 25% of patients with sphincter of Oddi dysfunction. Young anicteric patients with normal ducts are at increased risk. Post-ERCP pancreatitis is usually mild and self-limited but may rarely result in prolonged hospitalization, surgery, diabetes, or death. Bleeding occurs in 1% of endoscopic sphincterotomies. Ascending cholangitis, pseudocyst infection, retroperitoneal perforation, and abscess may occur as a result of ERCP.

Percutaneous gastrostomy tube placement during EGD is associated with a 10–15% incidence of complications, most often wound infections. Fasciitis, pneumonia, bleeding, buried bumper syndrome, and colonic injury may result from gastrostomy tube placement.

URGENT ENDOSCOPY

Acute Gastrointestinal Hemorrhage

Endoscopy is an important diagnostic and therapeutic technique for patients with acute gastrointestinal hemorrhage. Although gastrointestinal bleeding stops spontaneously in most cases, some patients will have persistent or recurrent hemorrhage that may be life-threatening. Clinical predictors of rebleeding help identify patients most likely to benefit from urgent endoscopy and endoscopic, angiographic, or surgical hemostasis.

INITIAL EVALUATION

The initial evaluation of the bleeding patient focuses on the magnitude of hemorrhage as reflected by the postural vital signs, the frequency of hematemesis or melena, and (in some cases) findings on nasogastric lavage. Decreases in hematocrit and hemoglobin lag behind the clinical course and are not reliable gauges of the magnitude of acute bleeding. This initial evaluation, completed well before the bleeding source is confidently identified, guides immediate supportive care of the patient and helps determine the timing of endoscopy. The severity of the initial hemorrhage is the most important indication for urgent endoscopy, since a large initial bleed increases the likelihood of ongoing or recurrent bleeding. Patients with resting hypotension, repeated hematemesis, bloody nasogastric aspirate that does not clear with large volume lavage, or orthostatic change in vital signs, or those requiring blood transfusions, should be considered for urgent endoscopy. In addition, patients with cirrhosis, coagulopathy, respiratory or renal failure, and those over 70 years of age are more likely to have significant rebleeding.

Bedside evaluation also suggests an upper or lower gastrointestinal source of bleeding in most patients. Over 90% of patients with melena are bleeding proximal to the ligament of Treitz, and about 90% of patients with hematochezia are bleeding from the colon. Melena can result from bleeding in the small bowel or right colon, especially in older patients with slow colonic transit. Conversely, some patients with massive hematochezia may be bleeding from an upper gastrointestinal source, such as a gastric Dieulafoy's lesion or duodenal ulcer, with rapid intestinal transit. Early upper endoscopy should be considered in such patients.

Endoscopy should be performed after the patient has been resuscitated with intravenous fluids and transfusions as necessary. Marked coagulopathy or thrombocytopenia is usually treated before endoscopy, since correction of these abnormalities may lead to resolution of bleeding, and techniques for endoscopic hemostasis are limited in such patients. Metabolic derangements should also be addressed. Tracheal intubation for airway protection should be considered before upper endoscopy in patients with repeated recent hematemesis and suspected variceal hemorrhage.

Most patients with impressive hematochezia can undergo colonoscopy after a rapid colonic purge with a polyethylene glycol solution; the preparation fluid may be administered via a nasogastric tube. Colonoscopy has a higher diagnostic yield than radionuclide bleeding scans or angiography in lower gastrointestinal bleeding, and endoscopic therapy can be applied in some

cases. In a minority of cases, endoscopic assessment is hindered by poor visualization due to persistent vigorous bleeding with recurrent hemodynamic instability, and other techniques (such as angiography or emergent subtotal colectomy) must be employed. In such patients, massive bleeding originating from an upper gastrointestinal source should also be considered and excluded by upper endoscopy. The anal and rectal mucosa should be visualized endoscopically early in the course of massive rectal bleeding, as bleeding lesions in or close to the anal canal may be identified that are amenable to endoscopic or surgical transanal hemostatic techniques.

PEPTIC ULCER

The endoscopic appearance of peptic ulcers provides useful prognostic information and guides the need for endoscopic therapy in patients with acute hemorrhage **(Fig. 291-16)**. A clean-based ulcer is associated with a low, 3–5% risk of rebleeding; patients with melena and a clean-based ulcer are often discharged home from the emergency room or endoscopy suite if they are young, reliable, and otherwise healthy. Flat pigmented spots and adherent clots covering the ulcer base have a 10 and 20% risk of rebleeding, respectively. Endoscopic therapy is often considered for an ulcer with an adherent clot. When a platelet plug is seen protruding from a vessel wall in the base of an ulcer (so-called sentinel clot or visible vessel), the risk of rebleeding from the ulcer is 40%. This finding generally leads to endoscopic therapy to decrease the rebleeding rate. Occasionally, active spurting from an ulcer is seen with >90% risk of ongoing bleeding without therapy.

Figure 291-16

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

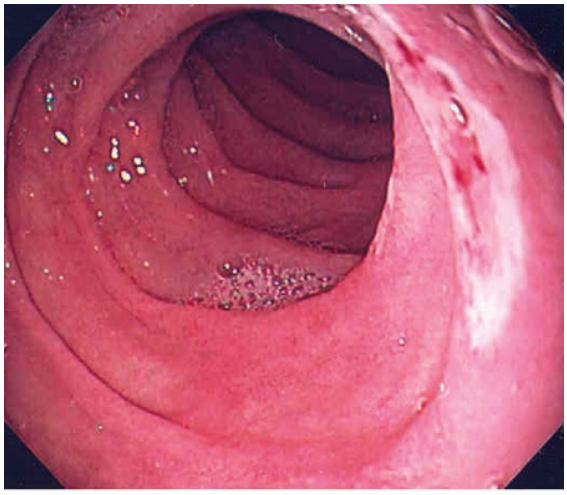

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

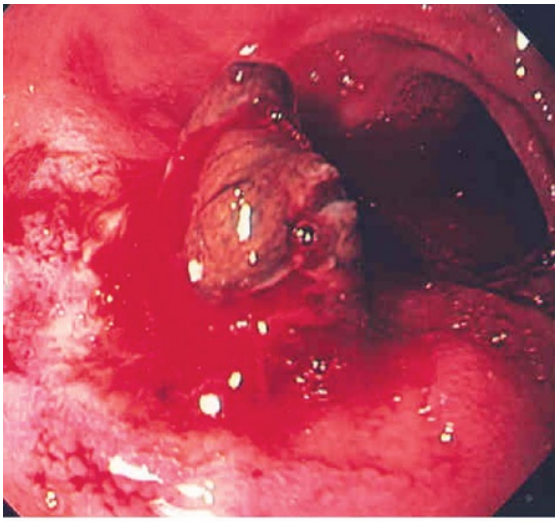

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

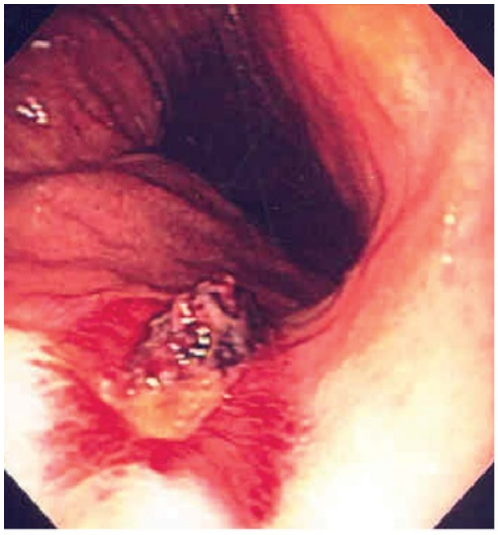

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

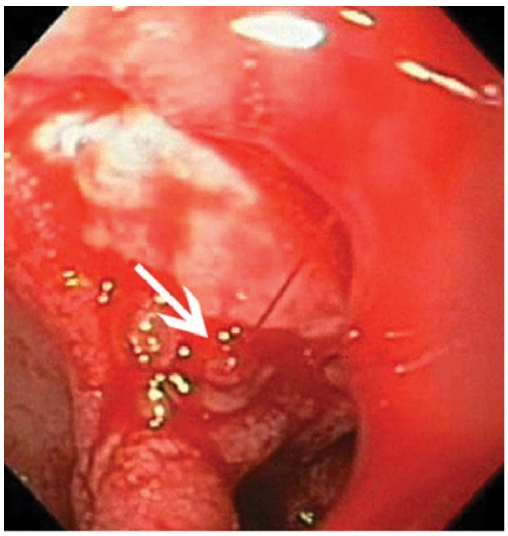

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stigmata of hemorrhage in peptic ulcers.*A.* Gastric antral ulcer with a clean base. ***B.*** Duodenal ulcer with flat pigmented spots. ***C.*** Duodenal ulcer with a dense adherent clot. ***D.*** Gastric ulcer with a pigmented protuberance/visible vessel. ***E.*** Duodenal ulcer with active spurting (*arrow*).

Endoscopic therapy of ulcers with high-risk stigmata typically lowers the rebleeding rate to 5–10%. Several hemostatic techniques are available, including injection of epinephrine or a sclerosant into and around the vessel, "coaptive coagulation" of the vessel in the base of the ulcer using a thermal probe that is pressed against the site of bleeding, placement of hemoclips, or a combination of these modalities **(Video e36-3)**. In conjunction with endoscopic therapy, the administration of a proton pump inhibitor decreases the risk of rebleeding and improves patient outcome.

VARICES

Two complementary strategies guide therapy of bleeding varices: local treatment of the bleeding varices and treatment of the underlying portal hypertension. Local therapies, including endoscopic variceal sclerotherapy, endoscopic variceal band ligation, and balloon tamponade with a Sengstaken-Blakemore tube, effectively control acute hemorrhage in most patients, although therapies that decrease portal pressure (pharmacologic treatment, surgical shunts, or radiologically placed intrahepatic portosystemic shunts) also play an important role.

Endoscopic variceal ligation (EVL) is indicated for the prevention of a first bleed from large esophageal varices **(Fig. 291-17)**, particularly in patients in whom beta blockers are contraindicated or not tolerated (primary prophylaxis). EVL is also the preferred endoscopic therapy for control of active esophageal variceal bleeding and for subsequent eradication of esophageal varices (secondary prophylaxis). During EVL, a varix is suctioned into a cap fitted on the end of the endoscope, and a rubber band is released from the cap, ligating the varix **(Video e36-4)**. EVL controls acute hemorrhage in up to 90% of patients. Complications of EVL, such as postbanding ulcer bleeding and esophageal stenosis, are uncommon. Endoscopic variceal sclerotherapy (EVS) involves the injection of a sclerosing, thrombogenic solution into or next to the esophageal varices. EVS also controls acute hemorrhage in most patients but has a higher complication rate than EVL. These techniques are used when varices are actively bleeding during endoscopy or (more commonly) when varices are the only identifiable cause of acute hemorrhage. Bleeding from large gastric fundic varices **(Fig. 291-18)** is best treated with endoscopic cyanoacrylate ("glue") injection **(Video e36-5)**, since EVL or EVS of these varices are associated with a high rebleeding rate. Complications of cyanoacrylate injection include infection and glue embolization to other organs, such as the lungs, brain, and spleen.

Figure 291-17

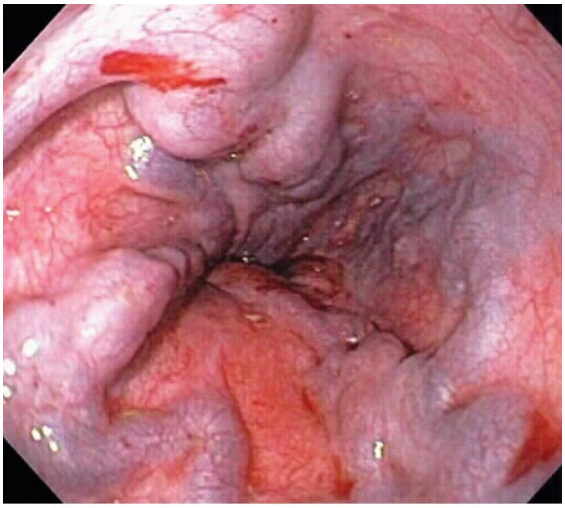

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Esophageal varices.

Figure 291-18

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gastric fundic varices.

After treatment of the acute hemorrhage, an elective course of endoscopic therapy can be undertaken with the goal of eradicating esophageal varices and preventing rebleeding months to years later. However, this chronic therapy is less successful, preventing long-term rebleeding in ~50% of patients. Pharmacologic therapies that decrease portal pressure have similar efficacy, and the two modalities may be combined.

DIEULAFOY'S LESION

This lesion, also called *persistent caliber artery,* is a large-caliber arteriole that runs immediately beneath the gastrointestinal mucosa and bleeds through a pinpoint mucosal erosion **(Fig. 291-19)**. Dieulafoy's lesion is seen most commonly on the lesser curvature of the proximal stomach, causes impressive arterial hemorrhage, and may be difficult to diagnose; it is often recognized only after repeated endoscopy for recurrent bleeding. Endoscopic therapy, such as thermal coagulation, is typically effective for control of bleeding and ablation of the underlying vessel once the lesion has been identified **(Video e36-6)**. Rescue therapies, such as angiographic embolization or surgical oversewing, are considered in situations where endoscopic therapy has failed.

Figure 291-19

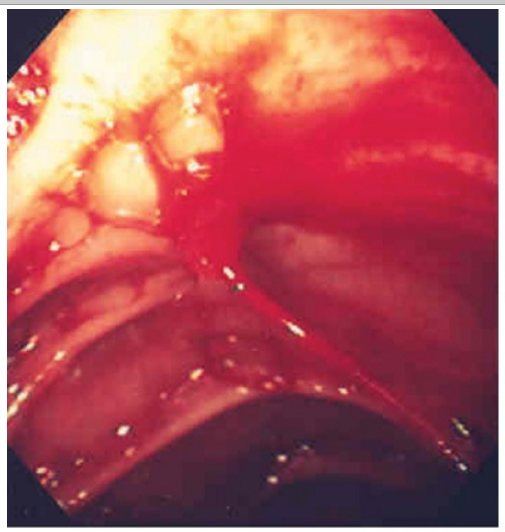

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

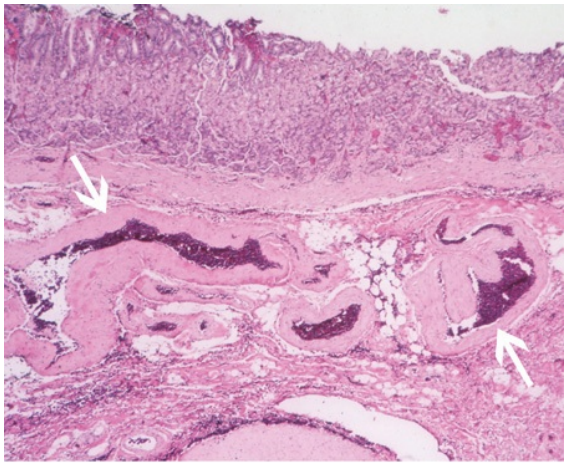

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dieulafoy's lesion.*A.* Actively spurting jejunal Dieulafoy's lesion. There is no underlying mucosal lesion. ***B.*** Histology of a gastric Dieulafoy's lesion. A persistent caliber artery (*arrows*) is present in the gastric submucosa, immediately beneath the mucosa.

MALLORY-WEISS TEAR

A Mallory-Weiss tear is a linear mucosal rent near or across the gastroesophageal junction that is often associated with retching or vomiting **(Fig. 291-20)**. When the tear disrupts a submucosal arteriole, brisk hemorrhage may result. Endoscopy is the best method of diagnosis, and an actively bleeding tear can be treated endoscopically with epinephrine injection, coaptive coagulation, band ligation, or hemoclips **(Video e36-7)**. Unlike peptic ulcer, a Mallory-Weiss tear with a nonbleeding sentinel clot in its base rarely rebleeds and thus does not necessitate endoscopic therapy.

Figure 291-20

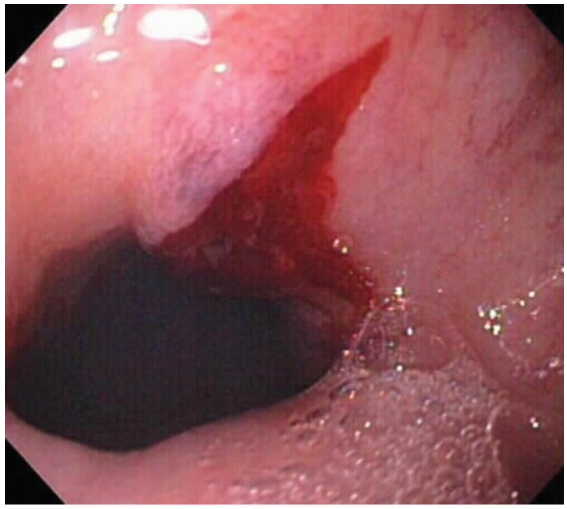

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mallory-Weiss tear at the gastroesophageal junction.

VASCULAR ECTASIAS

Vascular ectasias are flat mucosal vascular anomalies that are best diagnosed by endoscopy. They usually cause slow intestinal blood loss and occur either in a sporadic fashion or in a well-defined pattern of distribution [e.g., gastric antral vascular ectasia (GAVE) or "watermelon stomach"] **(Fig. 291-21)**. Cecal vascular ectasias (senile lesions), GAVE, and radiation-induced rectal ectasias are often responsive to local endoscopic ablative therapy, such as argon plasma coagulation **(Video e36-8)**. Patients with diffuse small-bowel vascular ectasias (associated with chronic renal failure and with hereditary hemorrhagic telangiectasia) may continue to bleed despite endoscopic treatment of easily accessible lesions by conventional endoscopy. These patients may benefit from deep enteroscopy with endoscopic therapy, pharmacologic treatment with octreotide or estrogen/progesterone therapy, or intraoperative enteroscopy.

Figure 291-21

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

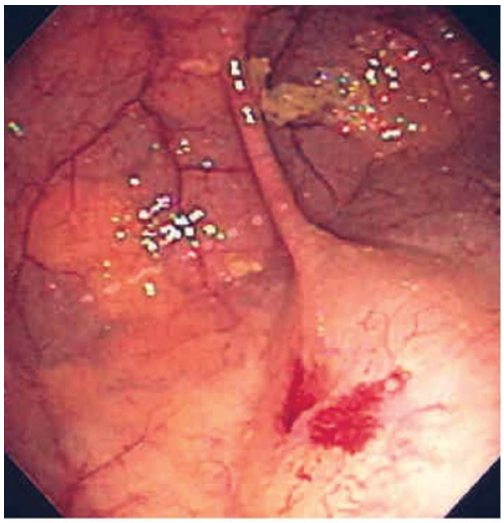

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gastrointestinal vascular ectasias.*A.* Gastric antral vascular ectasia ("watermelon stomach") characterized by stripes of prominent flat or raised vascular ectasias. ***B.*** Cecal vascular ectasias. ***C.*** Radiation-induced vascular ectasias of the rectum in a patient previously treated for prostate cancer.

COLONIC DIVERTICULA

Diverticula form where nutrient arteries penetrate the muscular wall of the colon en route to the colonic mucosa **(Fig. 291-22)**. The artery found in the base of a diverticulum may bleed, causing painless and impressive hematochezia. Colonoscopy is indicated in patients with hematochezia and suspected diverticular hemorrhage, since other causes of bleeding (such as vascular ectasias, colitis, and colon cancer) must be excluded. In addition, an actively bleeding diverticulum may be seen and treated during colonoscopy **(Video e36-9)**.

Figure 291-22

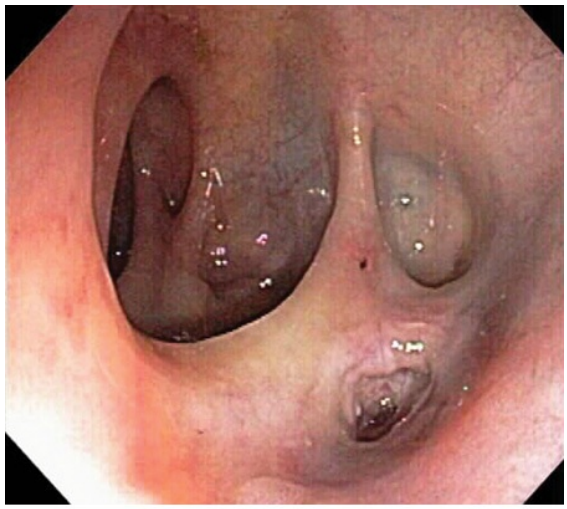

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Colonic diverticula.

Gastrointestinal Obstruction and Pseudoobstruction

Endoscopy is useful for evaluation and treatment of some forms of gastrointestinal obstruction. An important exception is small bowel obstruction due to surgical adhesions, which is generally not diagnosed or treated endoscopically. Esophageal, gastroduodenal, and colonic obstruction or pseudoobstruction can all be diagnosed and often managed endoscopically.

ACUTE ESOPHAGEAL OBSTRUCTION

Esophageal obstruction by impacted food **(Fig. 291-23)** or an ingested foreign body is a potentially life-threatening event and represents an endoscopic emergency. Left untreated, the patient may develop esophageal ulceration, ischemia, and perforation. Patients with persistent esophageal obstruction often have hypersalivation and are usually unable to swallow water; endoscopy is generally the best initial test in such patients, since endoscopic removal of the obstructing material is usually possible, and the presence of an underlying esophageal pathology can often be determined. Radiographs of the chest and neck should be considered before endoscopy in patients with fever, obstruction for ge;24 h, or ingestion of a sharp object such as a fishbone. Radiographic contrast studies interfere with subsequent endoscopy and are not advisable in most patients with a clinical picture of esophageal obstruction. Occasionally, sublingual nifedipine or nitrates, or intravenous glucagon, may resolve an esophageal food impaction, but in most patients an underlying web, ring, or stricture is present and endoscopic removal of the obstructing food bolus is necessary.

Figure 291-23

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Esophageal food (meat) impaction.

GASTRIC OUTLET OBSTRUCTION

Obstruction of the gastric outlet is commonly caused by gastric, duodenal, or pancreatic malignancy, or chronic peptic ulceration with stenosis of the pylorus. Patients vomit partially digested food many hours after eating. Gastric decompression with a

nasogastric tube and subsequent lavage for removal of retained material is the first step in treatment. The diagnosis can then be confirmed with a saline load test, if desired. Endoscopy is useful for diagnosis and treatment. Patients with benign pyloric stenosis may be treated with endoscopic balloon dilatation of the pylorus, and a course of endoscopic dilatation results in long-term relief of symptoms in about 50% of patients. Malignant gastric outlet obstruction can be relieved with endoscopically placed expandable stents **(Fig. 291-24)** in patients with inoperable malignancy.

Figure 291-24

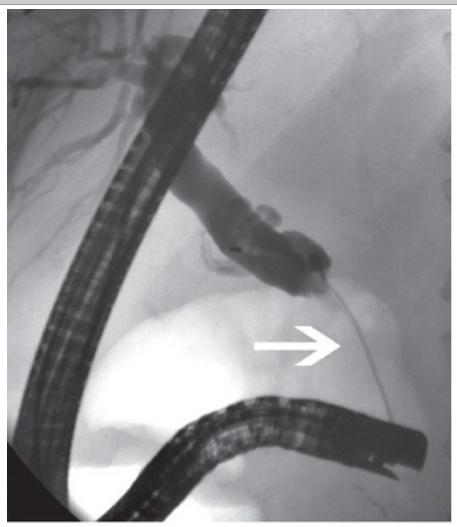

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

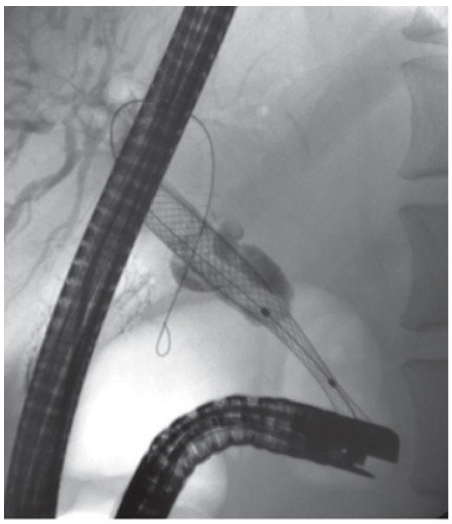

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

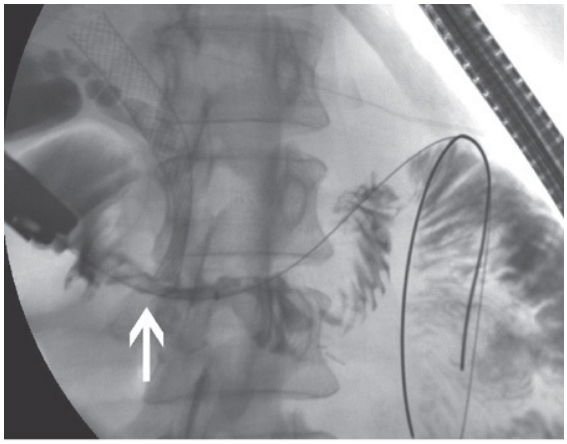

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

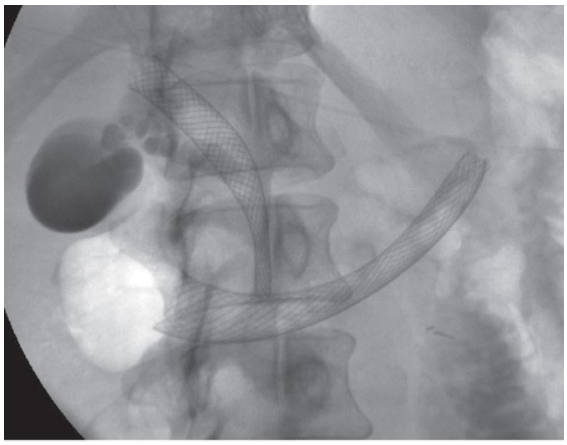

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Biliary and duodenal self-expanding metal stents (SEMS) for obstruction caused by pancreatic cancer.*A.* Endoscopic retrograde cholangiopancreatography (ERCP) demonstrates a distal bile duct stricture (*arrow*). ***B.*** A biliary SEMS is placed. ***C.*** Contrast injection demonstrates a duodenal stricture (*arrow*). ***D.*** Biliary and duodenal SEMS in place.

COLONIC OBSTRUCTION AND PSEUDOOBSTRUCTION

These both present with abdominal distention and discomfort; tympany; and a dilated, air-filled colon on plain abdominal radiography. The radiographic appearance can be characteristic of a particular condition, such as sigmoid volvulus **(Fig. 291-25)**. Both structural obstruction and pseudoobstruction may lead to colonic perforation if untreated. Acute colonic pseudoobstruction is a form of colonic ileus that is usually attributable to electrolyte disorders, narcotic and anticholinergic medications, immobility (as after surgery), and retroperitoneal hemorrhage or mass. Multiple causative factors are often present. Colonoscopy, water-soluble contrast enema, or CT may be used to look for an obstructing lesion and differentiate obstruction from pseudoobstruction. One of these diagnostic studies should be strongly considered if the patient does not have clear risk factors for pseudoobstruction, if radiographs do not show air in the rectum, or if the patient fails to improve when underlying causes of pseudoobstruction have been addressed. The risk of cecal perforation in pseudoobstruction rises when the cecal diameter exceeds 12 cm, and decompression of the colon may be achieved using intravenous neostigmine or via colonoscopic decompression **(Fig. 291-26)**. Most patients should receive a trial of conservative therapy (with correction of electrolyte disorders, removal of offending medications, and increased mobilization) before undergoing an invasive decompressive procedure for colonic pseudoobstruction.

Figure 291-25

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sigmoid volvulus with the characteristic radiologic appearance of a "bent inner tube."

Figure 291-26

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

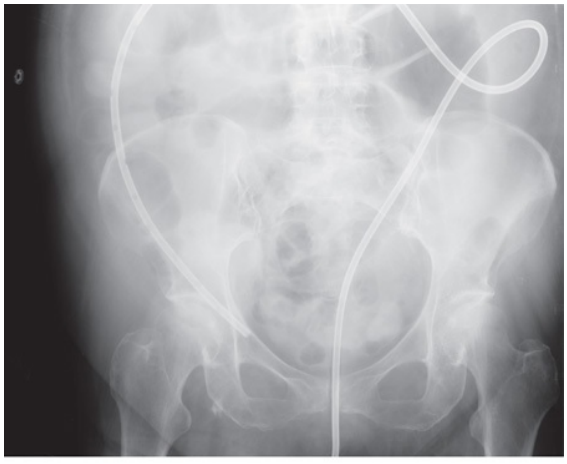

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute colonic pseudoobstruction.*A.* Acute colonic dilatation occurring in a patient soon after knee surgery. ***B.*** Colonoscopic placement of decompression tube with marked improvement in colonic dilatation.

Colonic obstruction is an indication for urgent intervention. Emergent diverting colostomy may be performed with a subsequent second operation after bowel preparation to treat the underlying cause of obstruction. Colonoscopic placement of an expandable stent is an alternative that can relieve malignant obstruction without emergency surgery and permit bowel preparation for an elective one-stage operation **(Fig. 291-27)**.

Figure 291-27

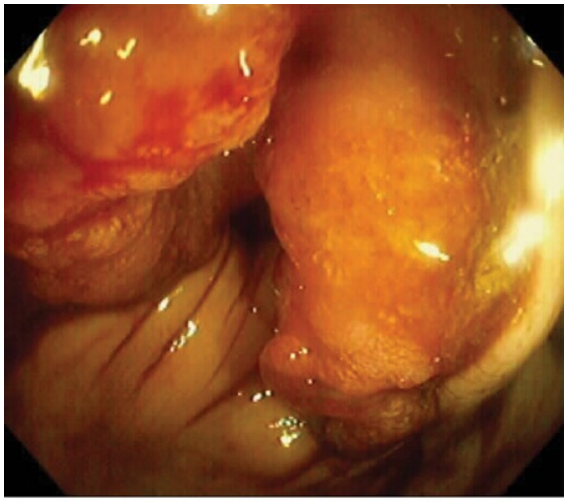

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

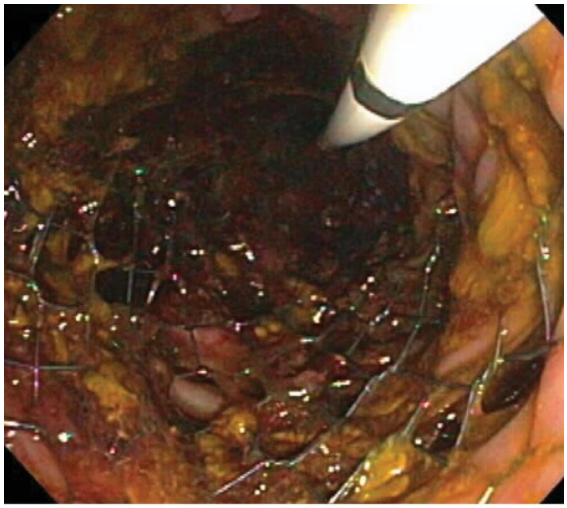

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

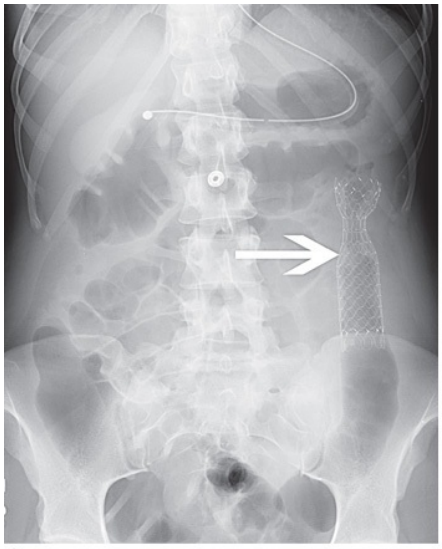

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Obstructing colonic carcinoma.*A.* Colonic adenocarcinoma causing marked luminal narrowing of the descending colon. ***B.*** Endoscopic placement of a self-expanding metal stent. ***C.*** Radiograph of expanded stent across the obstructing tumor with a residual waist (*arrow*). *(Image A courtesy of Dr. Glenn Alexander; with permission.)*

Acute Biliary Obstruction

The steady, severe pain that occurs when a gallstone acutely obstructs the common bile duct often brings patients to a hospital. The diagnosis of a ductal stone is suspected when the patient is jaundiced or when serum liver tests or pancreatic enzyme levels are elevated; it is confirmed by direct cholangiography (performed endoscopically, percutaneously, or during surgery). ERCP is currently the primary means of diagnosing and treating common bile duct stones in most hospitals in the United States **(Figs. 291-10** and **291-11)**.

BILE DUCT IMAGING

While transabdominal ultrasound diagnoses only a minority of bile duct stones, magnetic resonance cholangiopancreatography (MRCP) and EUS are >90% accurate and have an important role in diagnosis. Examples of these modalities are shown in **Fig. 291-28**.

Figure 291-28

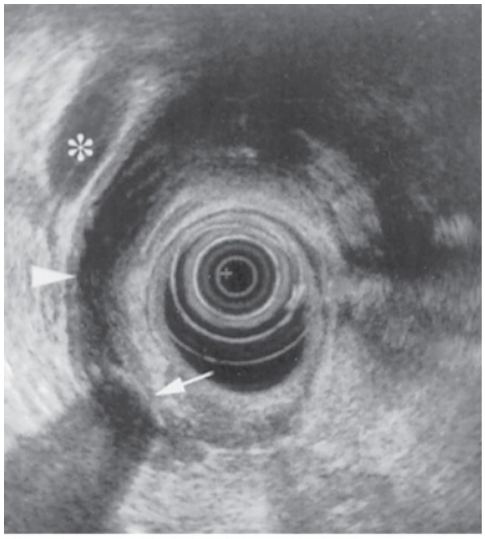

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

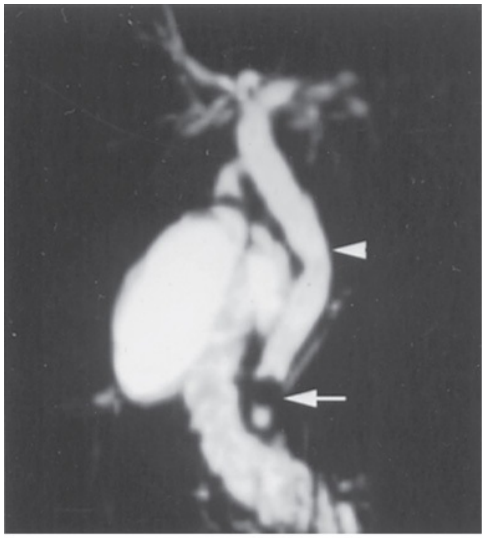

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

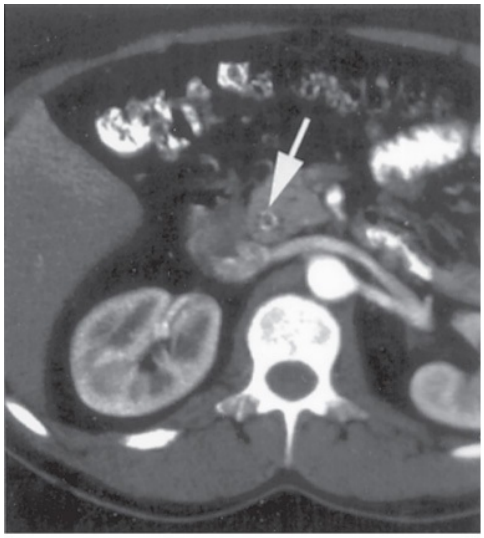

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Methods of bile duct imaging. Arrows mark bile duct stones. Arrowheads indicate the common bile duct, and the asterisk marks the portal vein. ***A.*** Endoscopic ultrasound (EUS). ***B.*** Magnetic resonance cholangiopancreatography (MRCP). ***C.*** Helical computed tomography (CT).

If the suspicion for a bile duct stone is high and urgent treatment is required (as in a patient with obstructive jaundice and biliary sepsis), ERCP is the procedure of choice, since it remains the gold standard for diagnosis and allows for immediate treatment **(Video e36-10)**. If a persistent bile duct stone is unlikely (as in a patient with gallstone pancreatitis), ERCP may be supplanted by less invasive imaging techniques, such as EUS or MRCP.

ASCENDING CHOLANGITIS

Charcot's triad of jaundice, abdominal pain, and fever is present in about 70% of patients with ascending cholangitis and biliary sepsis. These patients are managed initially with fluid resuscitation and intravenous antibiotics. Abdominal ultrasound is often performed to assess for gallbladder stones and bile duct dilation. However, the bile duct may not be dilated early in the course of acute biliary obstruction. Medical management usually improves the patient's clinical status, providing a window of approximately 24 h during which biliary drainage should be established, typically by ERCP. Undue delay can result in recrudescence of overt sepsis and increased morbidity and mortality rates. In addition to Charcot's triad, the additional presence of shock and confusion (Reynolds's pentad) is associated with high mortality rate and should prompt an urgent intervention to restore biliary drainage.

GALLSTONE PANCREATITIS

Gallstones may cause acute pancreatitis as they pass through the ampulla of Vater. The occurrence of gallstone pancreatitis usually implies passage of a stone into the duodenum, and only about 20% of patients harbor a persistent stone in the ampulla or the common bile duct. Retained stones are more common in patients with jaundice, rising serum liver tests following hospitalization, severe pancreatitis, or superimposed ascending cholangitis.

Urgent ERCP decreases the morbidity rate of gallstone pancreatitis in a subset of patients with retained bile duct stones. It is unclear whether the benefit of ERCP is mainly attributable to treatment and prevention of ascending cholangitis or to relief of pancreatic duct obstruction. ERCP is warranted early in the course of gallstone pancreatitis if ascending cholangitis is suspected, especially in a jaundiced patient. Urgent ERCP also appears to benefit patients predicted to have severe pancreatitis using a clinical index of severity such as the Glasgow or Ranson score. Since the benefit of ERCP is limited to patients with a retained bile duct stone, a strategy of initial MRCP or EUS for diagnosis decreases the utilization of ERCP in gallstone

pancreatitis and improves clinical outcomes by limiting the occurrence of ERCP-related complications.

ELECTIVE ENDOSCOPY

Dyspepsia

Dyspepsia is a chronic or recurrent burning discomfort or pain in the upper abdomen that may be caused by diverse processes such as gastroesophageal reflux, peptic ulcer disease, and "nonulcer dyspepsia," a heterogeneous category that includes disorders of motility, sensation, and somatization. Gastric and esophageal malignancies are less common causes of dyspepsia. Careful history-taking allows accurate differential diagnosis of dyspepsia in only about half of patients. In the remainder, endoscopy can be a useful diagnostic tool, especially in patients whose symptoms are not resolved by an empirical trial of symptomatic treatment. Endoscopy should be performed at the outset in patients with dyspepsia and alarm features, such as weight loss or iron deficiency anemia.

Gastroesophageal Reflux Disease (GERD)

When classic symptoms of gastroesophageal reflux are present, such as water brash and substernal heartburn, presumptive diagnosis and empirical treatment are often sufficient. Endoscopy is a sensitive test for diagnosis of esophagitis **(Fig. 291-29)**, but will miss nonerosive reflux disease (NERD) since some patients have symptomatic reflux without esophagitis. The most sensitive test for diagnosis of GERD is 24-h ambulatory pH monitoring. Endoscopy is indicated in patients with reflux symptoms refractory to antisecretory therapy; in those with alarm symptoms such as dysphagia, weight loss, or gastrointestinal bleeding; and in those with recurrent dyspepsia after treatment that is not clearly due to reflux on clinical grounds alone. Endoscopy may be considered in patients with long-standing (ge;10 years) GERD with frequent symptoms, as they have a sixfold increased risk of harboring Barrett's esophagus compared to a patient with <1 year of reflux symptoms. Patients with Barrett's esophagus **(Fig. 291-3)** generally undergo a surveillance program of periodic endoscopy with biopsies to detect dysplasia or early carcinoma.

Figure 291-29

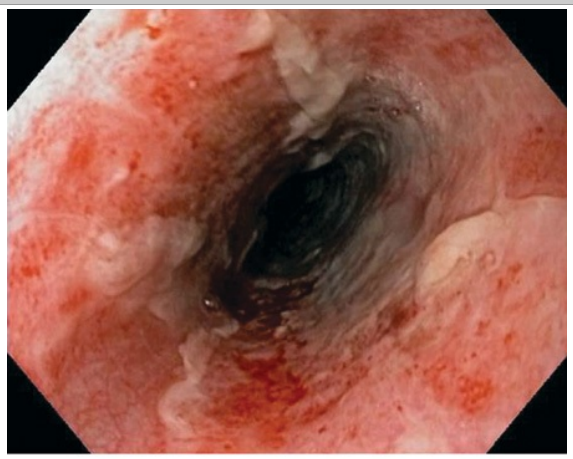

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

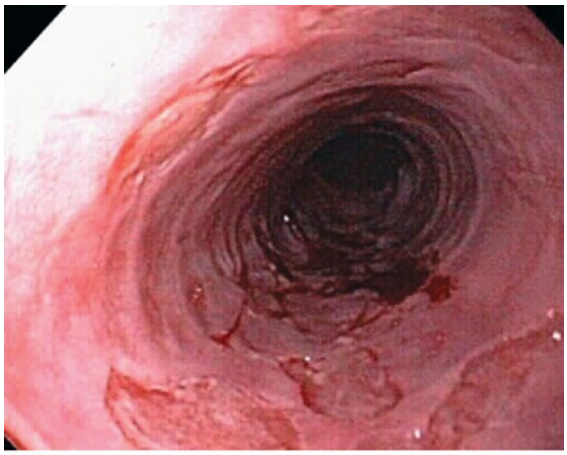

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

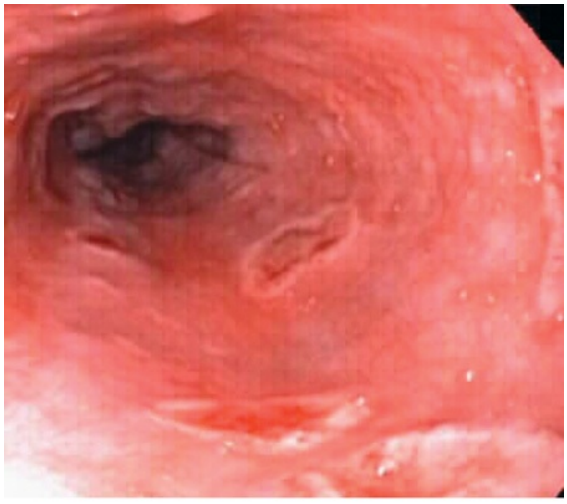

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

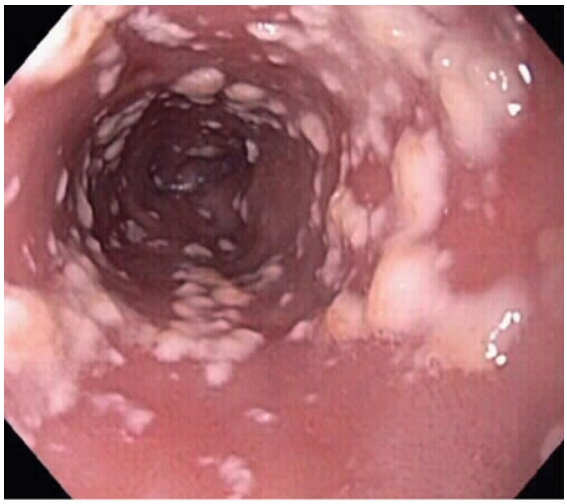

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Causes of esophagitis.*A.* Severe reflux esophagitis with mucosal ulceration and friability. ***B.*** Cytomegalovirus esophagitis. ***C.*** Herpes simplex virus esophagitis with target-type shallow ulcerations. ***D.*** Candida esophagitis with white plaques adherent to the esophageal mucosa.

BARRETT'S ESOPHAGUS

Barrett's esophagus is specialized columnar metaplasia that replaces the normal squamous mucosa of the distal esophagus in some persons with GERD. Barrett's epithelium is a major risk factor for adenocarcinoma of the esophagus and is readily detected endoscopically, due to proximal displacement of the squamocolumnar junction **(Fig. 291-3)**. A screening EGD for Barrett's esophagus may be considered in patients with a chronic history (>10 year) of GERD symptoms. Endoscopic biopsy is the gold standard for confirmation of Barrett's esophagus, and for dysplasia or cancer arising in Barrett's mucosa. Endoscopic therapies such as endoscopic mucosal resection (EMR) **(Video e36-11)**, endoscopic submucosal dissection (ESD), photodynamic therapy (PDT), and radiofrequency ablation (RFA) are effective modalities for treatment of high-grade dysplasia and intramucosal cancer in Barrett's esophagus.

Peptic Ulcer

Peptic ulcer classically causes epigastric gnawing or burning, often occurring nocturnally and promptly relieved by food or antacids. Although endoscopy is the most sensitive diagnostic test for peptic ulcer, it is not a cost-effective strategy in young patients with ulcer-like dyspeptic symptoms unless endoscopy is available at low cost. Patients with suspected peptic ulcer should be evaluated for *Helicobacter pylori* infection. Serology (past or present infection), urea breath testing (current infection), and stool tests are noninvasive and less costly than endoscopy with biopsy. Patients with alarm symptoms and those with persistent symptoms despite treatment should undergo endoscopy to exclude gastric malignancy and other etiologies.

Nonulcer Dyspepsia

Nonulcer dyspepsia may be associated with bloating and, unlike peptic ulcer, tends not to remit and recur. Most patients describe marginal relief on acid-reducing, prokinetic, or anti-*Helicobacter* therapy, and are referred for endoscopy to exclude a refractory ulcer and assess for other causes. Although endoscopy is useful for excluding other diagnoses, its impact on the treatment of patients with nonulcer dyspepsia is limited.

Dysphagia

About 50% of patients presenting with difficulty swallowing have a mechanical obstruction; the remainder has a motility disorder, such as achalasia or diffuse esophageal spasm. Careful history-taking often points to a presumptive diagnosis and leads to the appropriate use of diagnostic tests. Esophageal strictures **(Fig. 291-30)** typically cause progressive dysphagia, first for solids,

then for liquids; motility disorders often cause intermittent dysphagia for both solids and liquids. Some underlying disorders have characteristic historic features: Schatzki's ring **(Fig. 291-31)** causes episodic dysphagia for solids, typically at the beginning of a meal; oropharyngeal motor disorders typically present with difficulty initiating deglutition (*transfer dysphagia*) and nasal reflux or coughing with swallowing; and achalasia may cause nocturnal regurgitation of undigested food.

Figure 291-30

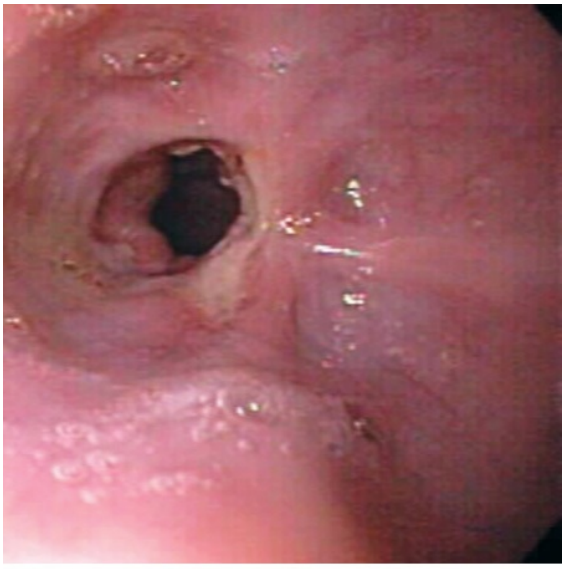

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Peptic esophageal stricture associated with ulceration and scarring of the distal esophagus.

Figure 291-31

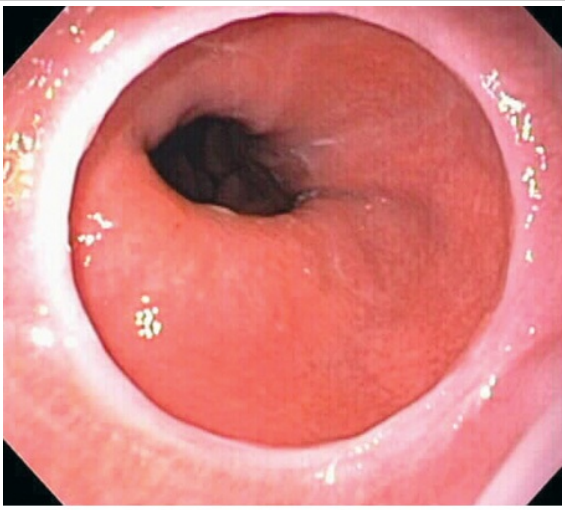

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schatzki's ring at the gastroesophageal junction.

When mechanical obstruction is suspected, endoscopy is a useful initial diagnostic test, since it permits immediate biopsy and/or dilatation of strictures, masses, or rings. The presence of linear furrows and multiple corrugated rings throughout a narrowed esophagus (*feline esophagus*) should raise suspicion for eosinophilic esophagitis, an increasingly recognized cause for recurrent dysphagia and food impaction **(Fig. 291-32)**. Blind or forceful passage of an endoscope may lead to perforation in a patient with stenosis of the cervical esophagus or a Zenker's diverticulum, but gentle passage of an endoscope under direct visual guidance is reasonably safe. Endoscopy can miss a subtle stricture or ring in some patients.

Figure 291-32

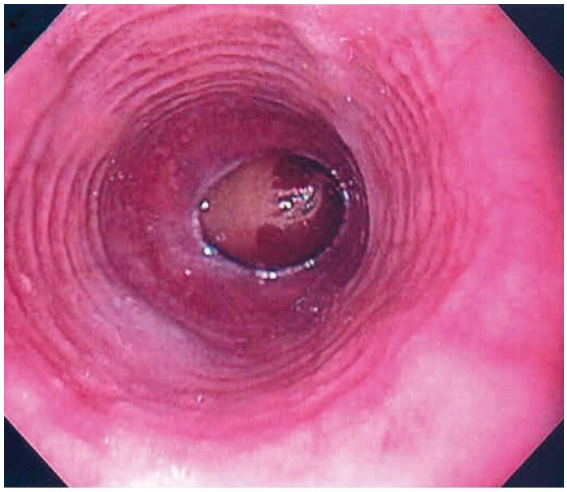

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Eosinophilic esophagitis with multiple circular rings of the esophagus creating a corrugated appearance, and an impacted grape at the narrowed esophagogastric junction. The diagnosis requires biopsy with histologic finding of >15–20 eosinophils/high power field.

When transfer dysphagia is evident or an esophageal motility disorder is suspected, esophageal radiography and/or a video-swallow study are the best initial diagnostic tests. The oropharyngeal swallowing mechanism, esophageal peristalsis, and the lower esophageal sphincter can all be assessed. In some disorders, subsequent esophageal manometry may also be important for diagnosis.

Anemia and Occult Blood in the Stool

Iron-deficiency anemia may be attributed to poor iron absorption (as in celiac sprue) or, more commonly, chronic blood loss. Intestinal bleeding should be strongly suspected in men and postmenopausal women with iron-deficiency anemia, and colonoscopy is indicated in such patients, even in the absence of detectable occult blood in the stool. Approximately 30% will have large colonic polyps, 10% will have colorectal cancer, and a few additional patients will have colonic vascular lesions. When a convincing source of blood loss is not found in the colon, upper gastrointestinal endoscopy should be considered; if no lesion is found, duodenal biopsies should be obtained to exclude sprue **(Fig. 291-33)**. Small bowel evaluation with capsule endoscopy or deep enteroscopy may be appropriate if both EGD and colonoscopy are unrevealing **(Fig. 291-34)**.

Figure 291-33

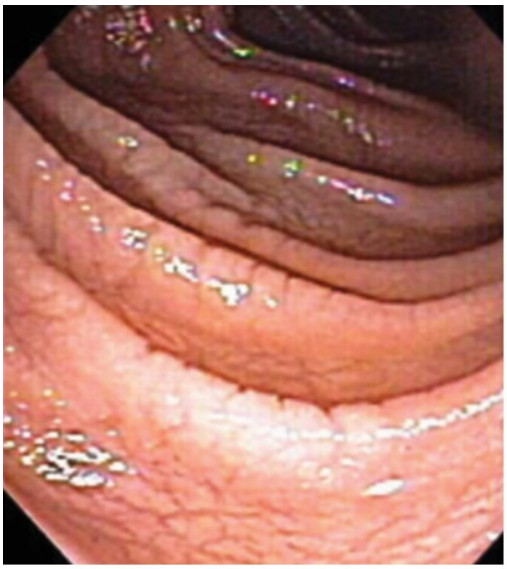

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scalloped duodenal folds in a patient with celiac sprue.

Figure 291-34

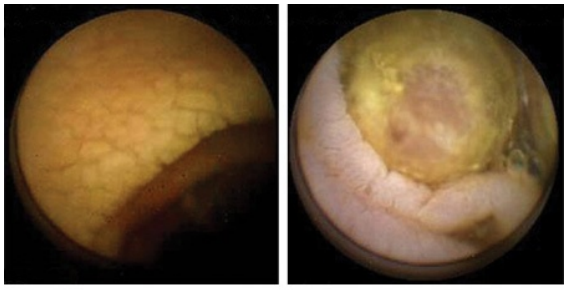

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Capsule endoscopy images of a mildly scalloped jejunal fold (left) and an ileal tumor (right) in a patient with celiac sprue.*(Images courtesy of Dr. Elizabeth Rajan; with permission.)*

Tests for occult blood in the stool detect hemoglobin or the heme moiety and are most sensitive for colonic blood loss, although they will also detect larger amounts of upper gastrointestinal bleeding. Patients over age 50 with occult blood in normal-appearing stool should undergo colonoscopy to diagnose or exclude colorectal neoplasia. The diagnostic yield is lower than in iron-deficiency anemia. Whether upper endoscopy is also indicated depends on the patient's symptoms.

The small intestine may be the source of chronic intestinal bleeding, especially if colonoscopy and upper endoscopy are not diagnostic. The utility of small bowel evaluation varies with the clinical setting and is most important in patients in whom bleeding causes chronic or recurrent anemia. In contrast to the low diagnostic yield of small bowel radiography, positive findings on capsule endoscopy are seen in 50–70% of patients with suspected small-intestinal bleeding. The most common finding is mucosal vascular ectasias. Deep enteroscopy may follow capsule endoscopy for biopsy of lesions or to provide specific therapy, such as argon plasma coagulation of vascular ectasias **(Fig. 291-35, Video e36-8)**.

Figure 291-35

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Mid-jejunal vascular ectasia identified by double-balloon endoscopy. ***B.*** Ablation of vascular ectasia with argon plasma coagulation.

Colorectal Cancer Screening

The majority of colon cancers develop from preexisting colonic adenomas, and colorectal cancer can be largely prevented by the detection and removal of adenomatous polyps **(Video e36-12)**. The choice of screening strategy for an asymptomatic

person depends on personal and family history. Individuals with inflammatory bowel disease, a history of colorectal polyps or cancer, family members with adenomatous polyps or cancer, or certain familial cancer syndromes **(Fig. 291-36)** are at increased risk for colorectal cancer. An individual without these factors is generally considered at average risk.

Figure 291-36

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Innumerable colon polyps of various sizes in a patient with familial adenomatous polyposis syndrome.

Screening strategies are summarized in **Table 291-2**. While stool tests for occult blood have been shown to decrease mortality rate from colorectal cancer, they do not detect some cancers and many polyps, and direct visualization of the colon is a more effective screening strategy. Either sigmoidoscopy or colonoscopy may be used for cancer screening in asymptomatic average-risk individuals. The use of sigmoidoscopy was based on the historical finding that the majority of colorectal cancers occurred in the rectum and left colon, and that patients with right-sided colon cancers had left-sided polyps. Over the past several decades, however, the distribution of colon cancers has changed, with proportionally fewer rectal and left-sided cancers than in the past. Large studies of colonoscopy for screening of average-risk individuals show that cancers are roughly equally distributed between left and right colon and half of patients with right-sided lesions have no polyps in the left colon. Visualization of the entire colon thus appears to be the optimal strategy for colorectal cancer screening and prevention.

Table 291-2 Colorectal Cancer Screening Strategies

	Choices/Recommendations	Comments
Average-Risk Patients		
Asymptomatic individuals ≥50 years of age (≥45 years of age for African Americans)	Colonoscopy every 10 years*	Preferred cancer prevention strategy
	Annual fecal immunochemical test (FIT) for occult bleeding, fecal DNA testing every 3 years	Cancer detection strategy; fails to detect many polyps and some cancers
	CT colonography every 5 years	Evolving technology (see text)
	Flexible sigmoidoscopy every 5 years	Fails to detect proximal colon polyps and cancers
	Double-contrast barium enema every 5 years	Less sensitive than colonoscopy or CT colonography, misses some rectosigmoid polyps and cancers
Personal History of Polyps or Colorectal Cancer		
1 or 2 small (<1 cm) adenomas with low-grade dysplasia	Repeat colonoscopy in 5 years	Assuming complete polyp resection
3 to 9 adenomas, or any adenoma ≥1 cm or containing high-grade dysplasia or villus features	Repeat colonoscopy in 3 years; subsequent colonoscopy based on findings	Assuming complete polyp resection
≥10 adenomas	Colonoscopy in <3 years based on clinical judgment	Consider evaluation for FAP or HNPCC; see recommendations below
Piecemeal removal of a sessile polyp	Exam in 2 to 6 months to verify complete removal	
Small (<1 cm) hyperplastic polyps of sigmoid and rectum	Colonoscopy in 10 years	
>2 serrated polyps, or any serrated or hyperplastic polyp ≥1 cm	Repeat colonoscopy in 3 years	
Incompletely removed serrated polyp ≥1 cm	Exam in 2 to 6 months to verify complete removal	
Colon cancer	Evaluate entire colon around the time of resection, then repeat colonoscopy in 3 years	

Inflammatory Bowel Disease		
Long-standing (>8 years) ulcerative colitis or Crohn's colitis, or left-sided ulcerative colitis of >15 years' duration	Colonoscopy with biopsies every 1 to 3 years	
Family History of Polyps or Colorectal Cancer		
First-degree relatives with only small tubular adenomas	Same as average risk	
Single first-degree relative with CRC or advanced adenoma at age ≥60 years	Same as average risk	
Single first-degree relative with CRC or advanced adenoma at age <60 years, OR two first-degree relatives with CRC or advanced adenomas at any age	Colonoscopy every 5 years beginning at age 40 years or 10 years younger than age at diagnosis of the youngest affected relative	
FAP	Sigmoidoscopy or colonoscopy annually, beginning at age 10–12 years	Consider genetic counseling and testing
HNPCC	Colonoscopy every 2 years beginning at age 20–25 years until age 40, then annually thereafter	Consider histologic evaluation for microsatellite instability in tumor specimens of patients who meet Bethesda criteria; consider genetic counseling and testing

*Assumes good colonic preparation and complete exam to cecum.

Abbreviations: CRC, colorectal cancer; FAP, familial adenomatous polyposis; HNPCC, hereditary nonpolyposis colorectal cancer.

Source: Adapted from Winawer SJ et al: Gastroenterology 130:1872, 2006 and Levin B et al: CA Cancer J Clin 58:130, 2008.

Virtual colonoscopy (VC) is a radiologic technique that images the colon with CT following rectal insufflation of the colonic lumen. Computer rendering of CT images generates an electronic display of a virtual "flight" along the colonic lumen, simulating colonoscopy **(Fig. 291-37)**. Comparative studies of virtual and routine colonoscopy have shown conflicting results, but technical refinements have improved the performance characteristics of VC. The use of VC for colorectal cancer screening may become more widespread in the future, particularly at institutions with demonstrated skill with this technique. Findings detected during virtual colonoscopy often require subsequent conventional colonoscopy for confirmation and treatment.

Figure 291-37

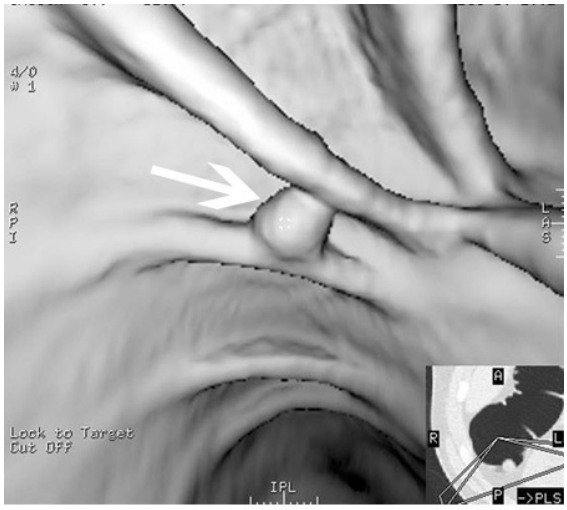

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Virtual colonoscopy image of a colon polyp (*arrow*). *(Image courtesy of Dr. Jeff Fidler; with permission.)*

Diarrhea

Most cases of diarrhea are acute, self-limited, and due to infections or medication. Chronic diarrhea (lasting >6 weeks) is more often due to a primary inflammatory, malabsorptive, or motility disorder; is less likely to resolve spontaneously; and generally requires diagnostic evaluation. Patients with chronic diarrhea or severe, unexplained acute diarrhea often undergo endoscopy if stool tests for pathogens are unrevealing. The choice of endoscopic testing depends on the clinical setting.

Patients with colonic symptoms and findings such as bloody diarrhea, tenesmus, fever, or leukocytes in stool generally undergo sigmoidoscopy or colonoscopy to assess for colitis (Fig. 291-4). Sigmoidoscopy is an appropriate initial test in most patients. Conversely, patients with symptoms and findings suggesting small-bowel disease, such as large-volume watery stools, substantial weight loss, and malabsorption of iron, calcium, or fat may undergo upper endoscopy with duodenal aspirates for assessment of bacterial overgrowth and biopsies for assessment of mucosal diseases, such as celiac sprue.

Many patients with chronic diarrhea do not fit either of these patterns. In the setting of a long-standing history of alternating constipation and diarrhea dating to early adulthood, without findings such as blood in the stool or anemia, a diagnosis of irritable bowel syndrome may be made without direct visualization of the bowel. Steatorrhea and upper abdominal pain may prompt evaluation of the pancreas rather than the gut. Patients whose chronic diarrhea is not easily categorized often undergo initial colonoscopy to examine the entire colon and terminal ileum for inflammatory or neoplastic disease **(Fig. 291-38)**.

Figure 291-38

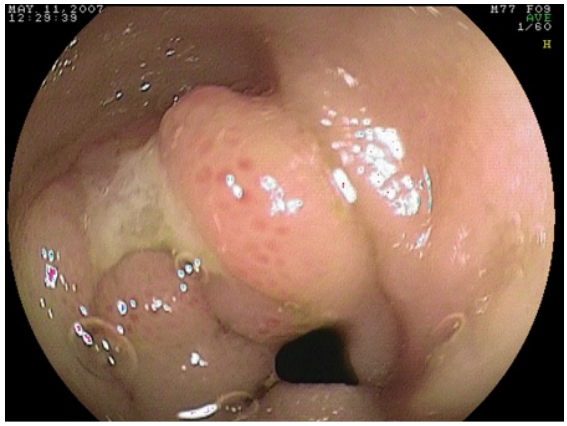

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ulcerated ileal carcinoid tumor.

Minor Hematochezia

Bright red blood passed with or on formed brown stool usually has a rectal, anal, or distal sigmoid source **(Fig. 291-39)**. Patients with even trivial amounts of hematochezia should be investigated with flexible sigmoidoscopy and anoscopy to exclude polyps or cancers in the distal colon. Patients reporting red blood on the toilet tissue only, without blood in the toilet or on the stool, are generally bleeding from a lesion in the anal canal. Careful external inspection, digital examination, and proctoscopy with anoscopy are sufficient for diagnosis in most cases.

Figure 291-39

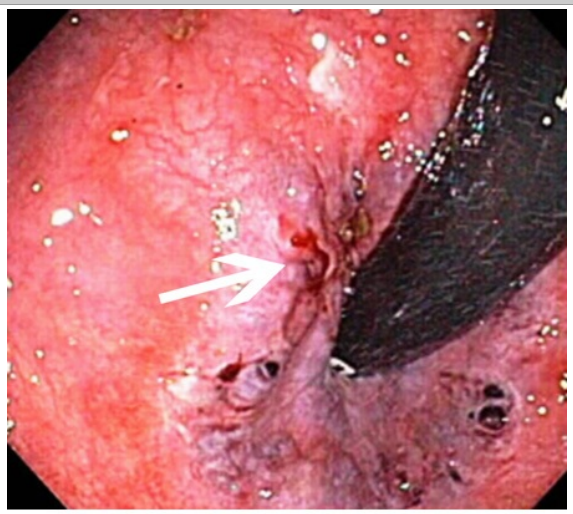

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Internal hemorrhoids with bleeding (*arrow*) as seen on a retroflexed view of the rectum.

Pancreatitis

About 20% of patients with pancreatitis have no identified cause after routine clinical investigation (including a review of medication and alcohol use, measurement of serum triglyceride and calcium levels, abdominal ultrasonography, and CT). Endoscopic assessment leads to a specific diagnosis in the majority of such patients, often altering clinical management. Endoscopic investigation is particularly appropriate if the patient has had more than one episode of pancreatitis.

Microlithiasis, or the presence of microscopic crystals in bile, is a leading cause of previously unexplained acute pancreatitis and is sometimes seen during abdominal ultrasonography as layering sludge or flecks of floating, echogenic material in the gallbladder. Gallbladder bile can be obtained for microscopic analysis by administering a cholecystokinin analogue during endoscopy, causing contraction of the gallbladder. Bile is suctioned from the duodenum as it drains from the papilla, and the darkest fraction is examined for cholesterol crystals or bilirubinate granules. The combination of EUS of the gallbladder and bile microscopy is probably the most sensitive means of diagnosing microlithiasis.

Previously undetected chronic pancreatitis, pancreatic malignancy, or pancreas divisum may be diagnosed by either ERCP or EUS. Sphincter of Oddi dysfunction or stenosis is a potential cause for pancreatitis and can be diagnosed by manometric studies performed during ERCP. Autoimmune pancreatitis may require EUS-guided pancreatic biopsy for histologic diagnosis.

Severe pancreatitis often results in pancreatic fluid collections. Both pseudocysts and areas of organized pancreatic necrosis can be drained into the stomach or duodenum endoscopically, using transpapillary and transmural endoscopic techniques. Pancreatic necrosis can be treated by direct endoscopic necrosectomy **(Video e36-2)**.

Cancer Staging

Local staging of esophageal, gastric, pancreatic, bile duct, and rectal cancers can be obtained with EUS (Fig. 291-14). EUS with fine-needle aspiration (Fig. 291-15) currently provides the most accurate preoperative assessment of local tumor and nodal staging, but it does not detect most distant metastases. Details of the local tumor stage can guide treatment decisions including resectability and need for neoadjuvant therapy. EUS with transesophageal needle biopsy may also be used to assess the

presence of non-small cell lung cancer in mediastinal nodes.

OPEN-ACCESS ENDOSCOPY

Direct scheduling of endoscopic procedures by primary care physicians without preceding gastroenterology consultation, or *open-access endoscopy,* is common. When the indications for endoscopy are clear-cut and appropriate, the procedural risks are low, and the patient understands what to expect, open-access endoscopy streamlines patient care and decreases costs.

Patients referred for open-access endoscopy should have a recent history, physical examination, and medication review. A copy of such an evaluation should be available when the patient comes to the endoscopy suite. Patients with unstable cardiovascular or respiratory conditions should not be referred directly for open-access endoscopy. Patients with particular conditions and undergoing certain procedures should be prescribed prophylactic antibiotics prior to endoscopy (Table 291-1). In addition, patients taking anticoagulants and/or antiplatelet drugs may require adjustment of these agents before endoscopy based on the procedure risk for bleeding and condition risk for a thromboembolic event **(Figs. 291-40 and 291-41)**. Common indications for open-access EGD include dyspepsia resistant to a trial of appropriate therapy; dysphagia; gastrointestinal bleeding; and persistent anorexia or early satiety. Open-access colonoscopy is often requested in men or postmenopausal women with iron-deficiency anemia, in patients over age 50 with occult blood in the stool, in patients with a previous history of colorectal adenomatous polyps or cancer, and for colorectal cancer screening. Flexible sigmoidoscopy is commonly performed as an open-access procedure.

Figure 291-40

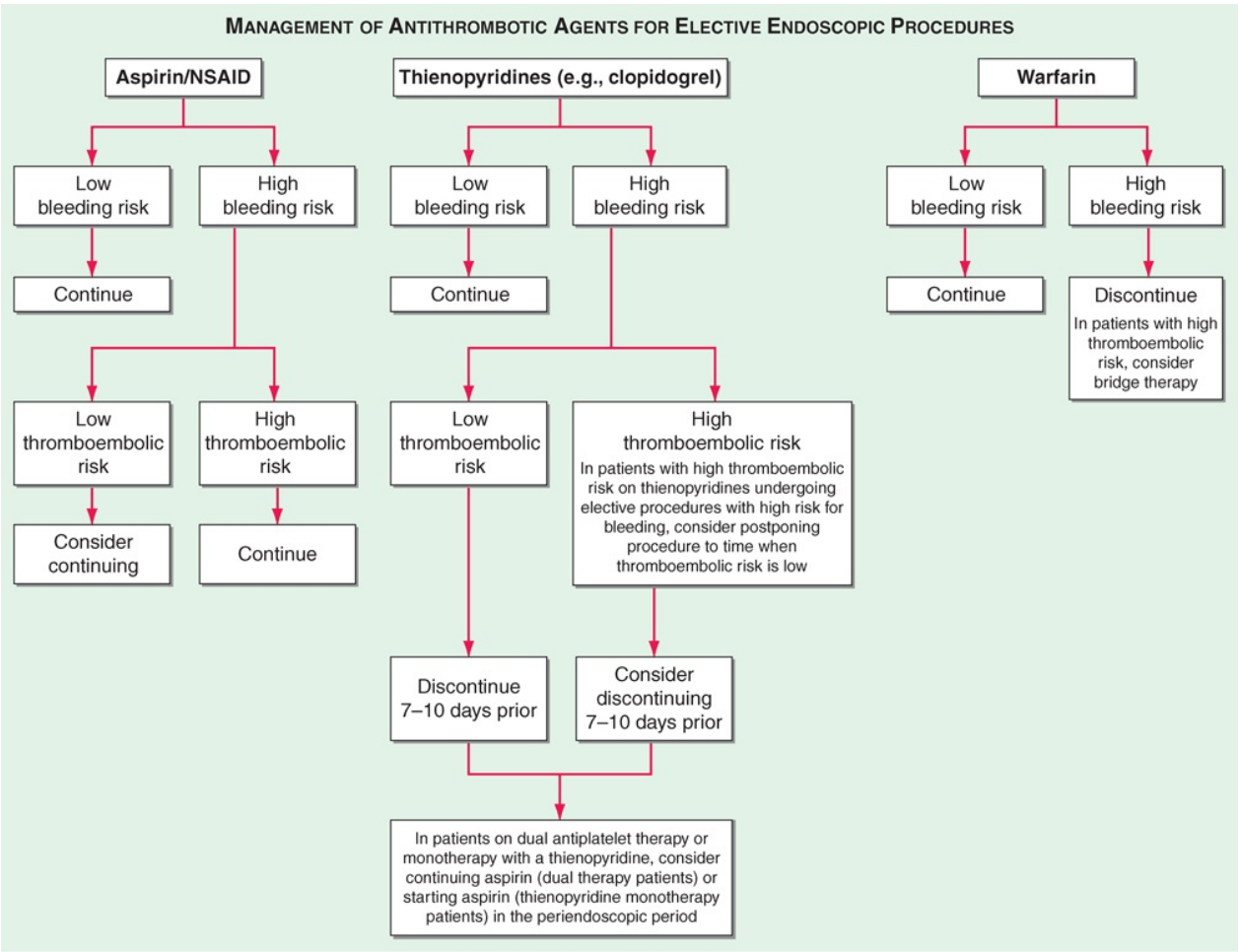

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of antithrombotic agents for elective endoscopic procedures. Higher-risk procedures for bleeding: Ploypectomy, biliary or pancreatic sphincterotomy, therapeutic balloon-assisted enteroscopy, PEG placement, pneumatic or bougie dilatation, treatment of varices, endoscopic hemostasis, tumor ablation by any technique, cystogastrostomy, EUS with FNA. Low-risk procedures for bleeding: Diagnostic (EGD, colonoscopy, flexible sigmoidoscopy) including biopsy, ERCP without sphincterotomy, EUS without FNA, enteroscopy and diagnostic balloon-assisted enteroscopy, capsule endoscopy, enteral stent deployment (without dilatation). Higher-risk conditions for thromboembolic event: Atrial fibrillation associated with valvular heart disease, prosthetic valves, active congestive heart failure, left ventricular ejection fraction <35%, a history of a thromboembolic event, hypertension, diabetes mellitus, or age >75 y; mechanical valve in the mitral position; mechanical valve in any position and previous thromboembolic event; recently (<1 y) placed coronary stent; acute coronary syndrome; non-stented percutaneous coronary intervention after myocardial infarction. Low-risk conditions for thromboembolic event: Uncomplicated or paroxysmal nonvalvular atrial fibrillation; bioprosthetic valve; mechanical valve in the aortic position; deep vein thrombosis. *(Adapted from MA Anderson et al: Gastrointest Endosc 70:1060, 2009; with permission from Elsevier.)*

Figure 291-41

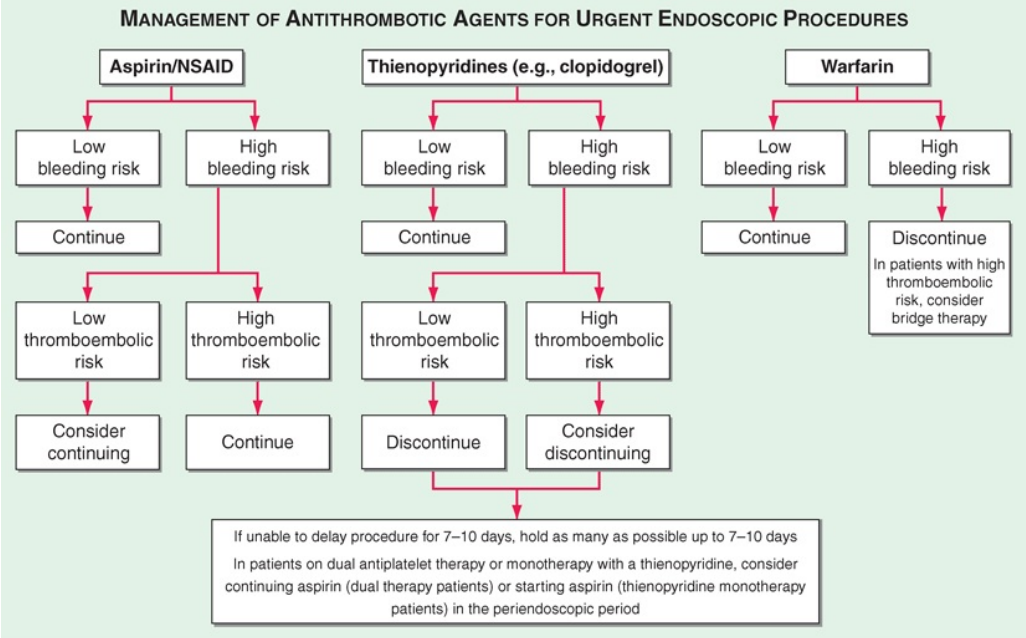

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of antithrombotic agents for urgent endoscopic procedures. Higher-risk procedures for bleeding: Polypectomy, biliary or pancreatic sphincterotomy, therapeutic balloon-assisted entero-scopy, PEG placement, pneumatic or bougie dilatation, treatment of varices, endoscopic hemostasis, tumor ablation by any technique, cystogastrostomy, EUS with FNA. Low-risk procedures for bleeding: Diagnostic (EGD, colonoscopy, flexible sigmoidoscopy) including biopsy, ERCP without sphincterotomy, EUS without FNA, enteroscopy and diagnostic balloon-assisted enteroscopy, capsule endoscopy, enteral stent deployment (without dilatation). Higher-risk conditions for thromboembolic event: Atrail fibrillation associated with valvular heart disease, prosthetic valves, active congestive heart failure, left ventricular ejection fraction <35%, a history of a thromboembolic event, hypertension, diabetes mellitus, or age > 75 y; mechanical valve in the mitral position; mechanical valve in any position and previous thromboembolic event; recently (>1y) placed coronary stent; acute coronary syndrome; non-stented percutaneous coronary intervention after myocardial infarction. Low-risk conditions for thromboembolic event: Uncomplicated or paroxysmal nonvalvular atrial fibrillation; bioprosthetic valve; mechanical valve in the aortic position, deep vein thrombosis. *(Adapted from MA Anderson et al: Gastrointest Endosc 70:1060, 2009; with permission from Elsevier.)*

When patients are referred for open-access colonoscopy, the primary care provider may need to choose a colonic preparation. Commonly used oral preparations include polyethylene glycol lavage solution, with or without citric acid. A "split-dose" regimen improves the quality of colonic preparation. Sodium phosphate purgatives may cause fluid and electrolyte abnormalities and renal toxicity, especially in patients with renal failure or congestive heart failure and those over 70 years of age.

FURTHER READINGS

BANERJEE S et al: The role of endoscopy in the management of patients with peptic ulcer disease. Gastrointest Endosc 71:663, 2010[PMID: 20363407] [Full Text]

BARKUN AN et al: International consensus recommendations on the management of patients with nonvariceal upper gastrointestinal bleeding. Ann Intern Med 152:101, 2010[PMID: 20083829] [Full Text]

GARCIA-TSAO G et al: Management of varices and variceal hemorrhage in cirrhosis. N Engl J Med 362:823, 2010[PMID: 20200386] [Full Text]

HARRISON ME et al: The role of endoscopy in the management of patients with known and suspected colonic obstruction and pseudo-obstruction. Gastrointest Endosc 71:669, 2010[PMID: 20363408] [Full Text]

IKENBERRY SO et al: The role of endoscopy in dyspepsia. Gastrointest Endosc 66:1071, 2007[PMID: 18028927] [Full Text]

RAJU GS et al: American Gastroenterological Association (AGA) Institute technical review on obscure gastrointestinal bleeding. Gastroenterology 133:1697, 2007[PMID: 17983812] [Full Text]

SHARMA P: Clinical practice. Barrett's esophagus. N Engl J Med 361:2548, 2009[PMID: 20032324] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VIDEO ATLAS OF GASTROINTESTINAL ENDOSCOPY

Gastrointestinal endoscopy is an increasingly important method for diagnosis and treatment of disease. This atlas demonstrates endoscopic findings in a variety of gastrointestinal infectious, inflammatory, vascular, and neoplastic conditions. Cancer screening and prevention are common indications for gastrointestinal endoscopy, and the premalignant conditions of Barrett's esophagus and colonic polyps are illustrated.

Endoscopic treatment modalities for gastrointestinal bleeding, polyps, and biliary stones are demonstrated in video clips. The images shown in this atlas are also found in Chap. 291 of the book.

Video e36-1

Methods of deep enteroscopy. *(Animations courtesy of Dr. Mark Stark and Dr. Jonathan Leighton; with permission.)*

Video e36-2

Pancreatic necrosis treated by transduodenal endoscopic drainage and necrosectomy.

Video e36-3

Actively bleeding duodenal ulcer treated with dilute epinephrine injection, thermal probe application, and hemoclips. *(Video courtesy of Dr. Navtej Buttar; with permission.*)

Video e36-4

Actively bleeding esophageal varices treated with endoscopic band ligation.

Video e36-5

Large, bleeding gastric varix treated with endoscopic cyanoacrylate injection.

Video e36-6

Dieulafoy's lesion treated endoscopically.

Video e36-7

Bleeding Mallory-Weiss tear treated with hemoclip placement.

Video e36-8

Radiation proctopathy treated with argon plasma coagulation.

Video e36-9

Actively bleeding colonic diverticulum treated with dilute epinephrine injection and band ligation.

Video e36-10

Bile duct stones removed after endoscopic sphincterotomy.

Video e36-11

Barrett's esophagus with high-grade dysplasia treated with endoscopic mucosal resection.

Video e36-12

Pedunculated and sessile colonic polyps removed with snare cautery during colonoscopy.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ESOPHAGEAL STRUCTURE AND FUNCTION

The esophagus is a hollow muscular tube coursing through the posterior mediastinum joining the hypopharynx to the stomach with a sphincter at each end. It functions to transport food and fluid between these ends, otherwise remaining empty. The physiology of swallowing, esophageal motility, and oral and pharyngeal dysphagia are described in **Chap. 38**. Esophageal diseases can be manifested by impaired function or pain. Key functional impairments are swallowing disorders and excessive gastroesophageal reflux. Pain, sometimes indistinguishable from cardiac chest pain, can result from inflammation, infection, dysmotility, or neoplasm.

SYMPTOMS OF ESOPHAGEAL DISEASE

The clinical history remains central to the evaluation of esophageal symptoms. A thoughtfully obtained history will often expedite management. Important details include weight gain or loss, gastrointestinal bleeding, dietary habits including the timing of meals, smoking, and alcohol consumption. The major esophageal symptoms are heartburn, regurgitation, chest pain, dysphagia, odynophagia, and globus sensation.

Heartburn (pyrosis), the most common esophageal symptom, is characterized by a discomfort or burning sensation behind the sternum that arises from the epigastrium and may radiate toward the neck. Heartburn is an intermittent symptom, most commonly experienced after eating, during exercise, and while lying recumbent. The discomfort is relieved with drinking water or antacid but can occur frequently and interfere with normal activities including sleep. The association between heartburn and gastroesophageal reflux disease (GERD) is so strong that empirical therapy for GERD has become accepted management. However, the term "heartburn" is often misused and/or referred to with other terms such as "indigestion" or "repeating," making it important to clarify the intended meaning.

Regurgitation is the effortless return of food or fluid into the pharynx without nausea or retching. Patients report a sour or burning fluid in the throat or mouth that may also contain undigested food particles. Bending, belching, or maneuvers that increase intraabdominal pressure can provoke regurgitation. A clinician needs to discriminate among regurgitation, vomiting, and rumination. *Vomiting* is preceded by nausea and accompanied by retching. *Rumination* is a behavior in which recently swallowed food is regurgitated and then reswallowed repetitively for up to an hour. Although there is some linkage between rumination and mental deficiency, the behavior is also exhibited by unimpaired individuals who sometimes even find it pleasurable.

Chest pain is a common esophageal symptom with characteristics similar to cardiac pain, sometimes making this distinction difficult. Esophageal pain is usually experienced as a pressure type sensation in the mid chest, radiating to the mid back, arms, or jaws. The similarity to cardiac pain is likely because the two organs share a nerve plexus and the nerve endings in the esophageal wall have poor discriminative ability among stimuli. Esophageal distention or even chemostimulation (e.g., with acid) will often be perceived as chest pain. Gastroesophageal reflux is the most common cause of esophageal chest pain.

Esophageal *dysphagia* (see also **Chap. 38**) is often described as a feeling of food "sticking" or even lodging in the chest. Important distinctions are between uniquely solid food dysphagia as opposed to liquid and solid, episodic versus constant dysphagia, and progressive versus static dysphagia. If the dysphagia is for liquids as well as solid food, it suggests a motility disorder such as achalasia. Conversely, uniquely solid food dysphagia is suggestive of a stricture, ring, or tumor. Of note, a patient's localization of food hang-up in the esophagus is notoriously imprecise. Approximately 30% of distal esophageal obstructions are perceived as cervical dysphagia. In such instances, the absence of concomitant symptoms generally associated with oropharyngeal dysphagia such as aspiration, nasopharyngeal regurgitation, cough, drooling, or obvious neuromuscular compromise should suggest an esophageal etiology.

Odynophagia is pain either caused by or exacerbated by swallowing. Odynophagia is more common with pill or infectious esophagitis than with reflux esophagitis and should prompt a search for these entities. When odynophagia does occur in GERD, it is likely related to an esophageal ulcer or deep erosion.

Globus sensation, alternatively labeled "globus hystericus," is the perception of a lump or fullness in the throat that is felt irrespective of swallowing. Although such patients are frequently referred for an

evaluation of dysphagia, globus sensation is often relieved by the act of swallowing. As implied by its alternative name (globus hystericus), globus sensation often occurs in the setting of anxiety or obsessive-compulsive disorders. Clinical experience teaches that it is often attributable to GERD.

Water brash is excessive salivation resulting from a vagal reflex triggered by acidification of the esophageal mucosa. This is not a common symptom. Afflicted individuals will describe the unpleasant sensation of the mouth rapidly filling with salty thin fluid, often in the setting of concomitant heartburn.

DIAGNOSTIC STUDIES

Endoscopy

Endoscopy, also known as esophagogastroduodenoscopy (EGD) is the best test for the evaluation of the proximal gastrointestinal tract. Modern instruments produce high-quality color images of the esophageal, gastric, and duodenal lumen. Endoscopes also have an instrumentation channel through which biopsy forceps, sclerotherapy catheters, balloon dilators, or cautery devices can be utilized. The key advantages of endoscopy over barium radiography are: (1) increased sensitivity for the detection of mucosal lesions, (2) vastly increased sensitivity for the detection of abnormalities mainly identifiable by an abnormal color such as Barrett's metaplasia, (3) the ability to obtain biopsy specimens for histologic examination of suspected abnormalities, and (4) the ability to dilate strictures during the examination. The main disadvantage of endoscopy is that it usually necessitates the use of conscious sedation with medicines such as midazolam (Versed), meperidine (Demerol), or fentanyl.

Radiography

Contrast radiography of the esophagus, stomach, and duodenum can demonstrate barium reflux, hiatal hernia, mucosal granularity, erosions, ulcerations, and strictures. The sensitivity of radiography compared with endoscopy for detecting esophagitis reportedly ranges from 22–95%, with higher grades of esophagitis (i.e., ulceration or stricture) exhibiting greater detection rates. Conversely, the sensitivity of barium radiography for detecting esophageal strictures is greater than that of endoscopy, especially when the study is done in conjunction with barium-soaked bread or a 13-mm barium tablet. Barium studies also provide an assessment of esophageal function and morphology that may be undetected on endoscopy. Hypopharyngeal pathology and disorders of the cricopharyngeal muscle are better appreciated on radiographic examination, particularly with videofluoroscopic recording. The major shortcoming of barium radiography is that it rarely obviates the need for endoscopy. Either a positive or a negative study is usually followed by an endoscopic evaluation either to clarify findings in the case of a positive examination or to add a level of certainty in the case of a negative one.

Endoscopic Ultrasound

Endoscopic ultrasound (EUS) instruments combine an endoscope with an ultrasound transducer to create a transmural image of the tissue surrounding the endoscope tip. The key advantage of EUS over alternative radiologic imaging techniques is much greater resolution attributable to the proximity of the ultrasound transducer to the area being examined. Available devices can provide either radial imaging (360-degree, cross-sectional) or a curved linear image that can guide fine-needle aspiration of imaged structures such as lymph nodes or tumors. Major esophageal applications of EUS are to stage esophageal cancer, to evaluate dysplasia in Barrett's esophagus, and to assess submucosal tumors.

Esophageal Manometry

Esophageal manometry, or motility testing, entails positioning a pressure sensing catheter within the esophagus and then observing the contractility following test swallows. The upper and lower esophageal sphincters appear as zones of high pressure that relax on swallowing while the intersphincteric esophagus exhibits peristaltic contractions. Manometry is used to diagnose motility disorders (achalasia, diffuse esophageal spasm) and to assess peristaltic integrity prior to the surgery for reflux disease. Technological advances have rebranded esophageal manometry as high-resolution esophageal pressure topography **(Fig. 292-1)**. Manometry can also be combined with intraluminal impedance monitoring. Impedance recordings utilize a catheter with a series of paired electrodes. Esophageal luminal contents in contact with the electrodes decrease (liquid) or increase (air) the impedance signal allowing detection of anterograde or retrograde transit of esophageal bolus transit.

Figure 292-1

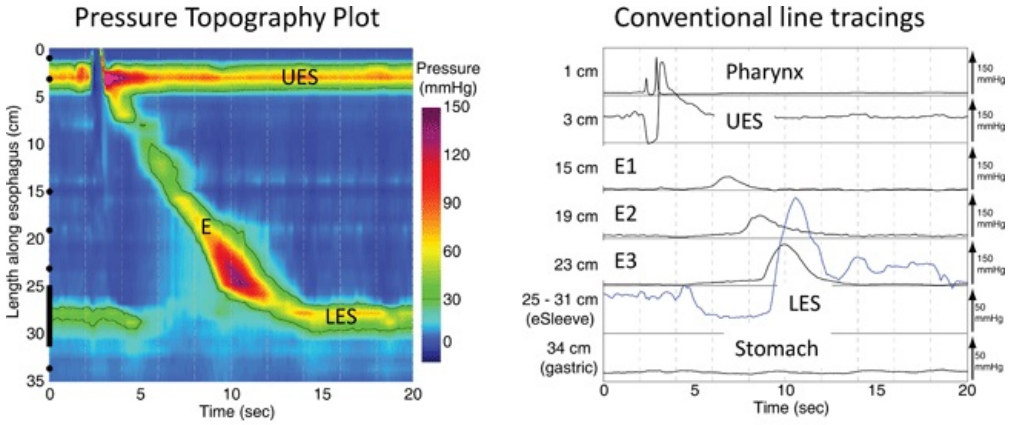

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

High-resolution esophageal pressure topography *(right)* and conventional manometry *(left)* of a normal swallow. LES, lower esophageal sphincter; E, esophageal body; UES, upper esophageal sphincter.

Reflux Testing

GERD is often diagnosed in the absence of endoscopic esophagitis, which would otherwise define the disease. This occurs in the settings of partially treated disease, an abnormally sensitive esophageal mucosa, or without obvious explanation. In such instances, reflux testing can demonstrate excessive esophageal exposure to refluxed gastric juice, the physiologic abnormality of GERD. This can be done by ambulatory 24- to 48-hour esophageal pH recording using either a wireless pH-sensitive transmitter that is anchored to the esophageal mucosa or with a transnasally positioned wire electrode with the tip stationed in the distal esophagus. Either way, the outcome is expressed as the percentage of the day that the pH was less than 4 (indicative of recent acid reflux), with values exceeding 5% indicative of GERD. Reflux testing is useful with atypical symptoms or an inexplicably poor response to therapy. Intraluminal impedance monitoring can be added to pH monitoring to detect reflux events irrespective of whether or not they are acidic, potentially increasing the sensitivity of the study.

STRUCTURAL DISORDERS

Hiatal Hernia

Hiatus hernia is a herniation of viscera, most commonly the stomach, into the mediastinum through the esophageal hiatus of the diaphragm. Four types of hiatus hernia are distinguished with type I, or sliding hiatal hernia comprising at least 95% of the overall total. A sliding hiatal hernia is one in which the gastroesophageal junction and gastric cardia slide upward as a result of weakening of the phrenoesophageal ligament attaching the gastroesophageal junction to the diaphragm at the hiatus. True to its name, sliding hernias enlarge with increased intraabdominal pressure, swallowing, and respiration. The incidence of sliding hernias increases with age and conceptually, results from wear and tear: increased intraabdominal pressure from abdominal obesity, pregnancy, etc., and hereditary factors predisposing to the condition. The main significance of sliding hernias is the propensity of affected individuals to have GERD.

Types II, III, and IV hiatal hernias are all subtypes of paraesophageal hernia in which the herniation into the mediastinum includes a visceral structure other than the gastric cardia. With type II and III paraesophageal hernias, the gastric fundus also herniates with the distinction being that in type II, the gastroesophageal junction remains fixed at the hiatus, while type III is a mixed sliding/paraesophageal hernia. With type IV hiatal hernias, viscera other than the stomach herniate into the mediastinum, most commonly the colon. With type II and III paraesophageal hernias, the stomach inverts as it herniates and large paraesophageal hernias can lead to an upside down stomach, gastric volvulus, and even strangulation of the stomach. Because of this risk, surgical repair is often advocated for large paraesophageal hernias.

Rings and Webs

A lower esophageal mucosal ring, also called a *B ring*, is a thin membranous narrowing at the squamocolumnar mucosal junction **(Fig. 292-2)**. Its origin is unknown but B rings are demonstrable in about 15% of people and are usually asymptomatic. When the lumen diameter is less than 13 mm, distal rings are usually associated with episodic solid food dysphagia and are called *Schatzki rings*. Patients typically present older than 40 years, consistent with an acquired rather than congenital origin. Schatzki ring is one of the most common causes of intermittent food impaction, also known as "steakhouse syndrome" as meat is a typical instigator. Symptomatic rings are easily treated by dilatation.

Figure 292-2

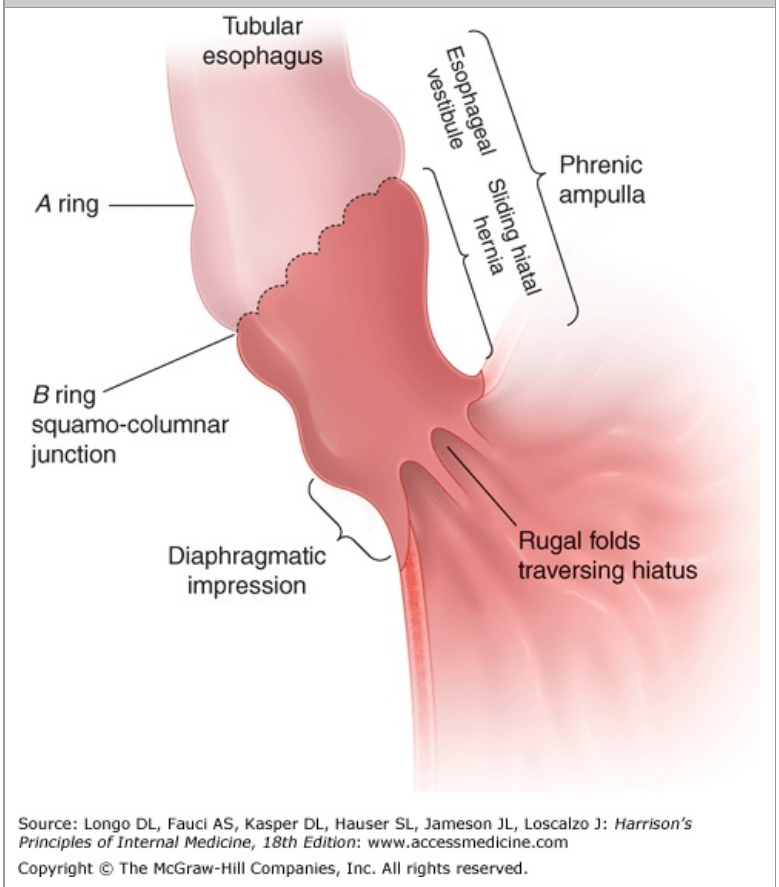

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiographic anatomy of the gastroesophageal junction.

Web-like constrictions higher in the esophagus can be of congenital or inflammatory origin. Asymptomatic cervical esophageal webs are demonstrated in about 10% of people and typically originate along the anterior aspect of the esophagus. When circumferential, they can cause intermittent dysphagia to solids similar to Schatzki rings and are similarly treated with dilatation. The combination of symptomatic proximal esophageal webs and iron-deficiency anemia in middle-aged women constitutes Plummer-Vinson syndrome.

Diverticula

Esophageal diverticula are categorized by location with the most common being epiphrenic, hypopharyngeal (Zenker's), and mid esophageal. Epiphrenic and Zenker's diverticula are false diverticula involving herniation of the mucosa and submucosa through the muscular layer of the esophagus. These lesions result from increased intraluminal pressure associated with distal obstruction. In the case of Zenker's, the obstruction is a stenotic cricopharyngeus muscle (upper esophageal sphincter) and the hypopharyngeal herniation most commonly occurs in an area of natural weakness known as *Killian's triangle***(Fig. 292-3)**. Small Zenker's diverticula are usually asymptomatic but when they enlarge sufficiently to retain food and saliva they can be associated with dysphagia, halitosis, and aspiration. Treatment is by surgical diverticulectomy and cricopharyngeal myotomy or a marsupialization procedure in which an endoscopic stapling device is used to divide the cricopharyngeus.

Figure 292-3

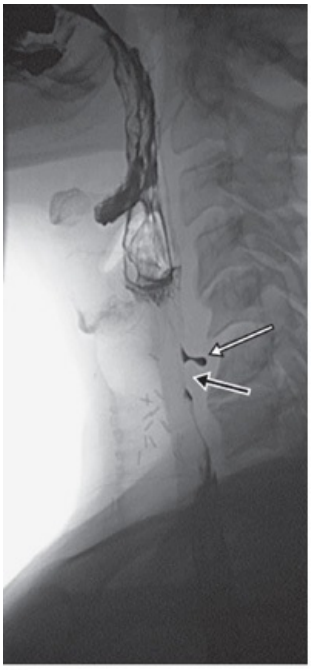

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

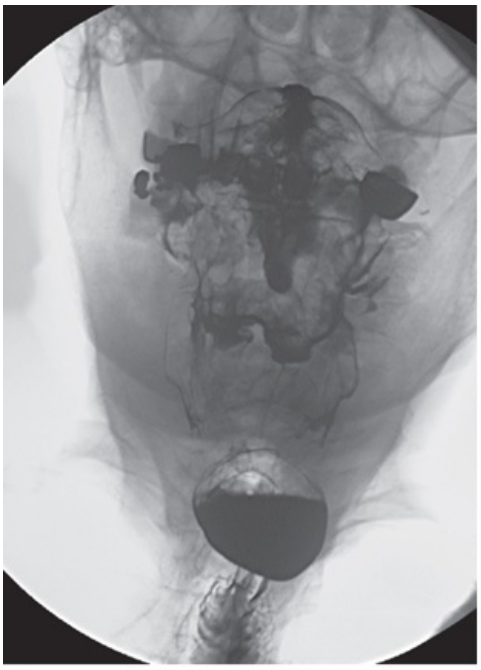

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examples of small (*left*) and large (*middle, right*) Zenker's diverticulum arising from Killian's triangle in the distal hypopharynx. Smaller diverticula are evident only during the swallow, whereas larger ones retain food and fluid.

Epiphrenic diverticula are usually associated with achalasia or a distal esophageal stricture. Mid-esophageal diverticula may be caused by traction from adjacent inflammation (classically tuberculosis) in which case they are true diverticula involving all layers of the esophageal wall, or by pulsion associated with esophageal motor disorders. Mid-esophageal and epiphrenic diverticula are usually asymptomatic until they enlarge sufficiently to retain food and cause dysphagia and regurgitation. Symptoms attributable to the diverticula tend to correlate more with the underlying esophageal disorder than the size of the diverticula. Large diverticula can be removed surgically, usually in conjunction with a myotomy if the underlying cause is achalasia. Diffuse intramural esophageal diverticulosis is a rare entity that results from dilatation of the excretory ducts of submucosal esophageal glands **(Fig. 292-4)**. Esophageal candidiasis and proximal esophageal strictures are commonly found in association with this disorder.

Figure 292-4

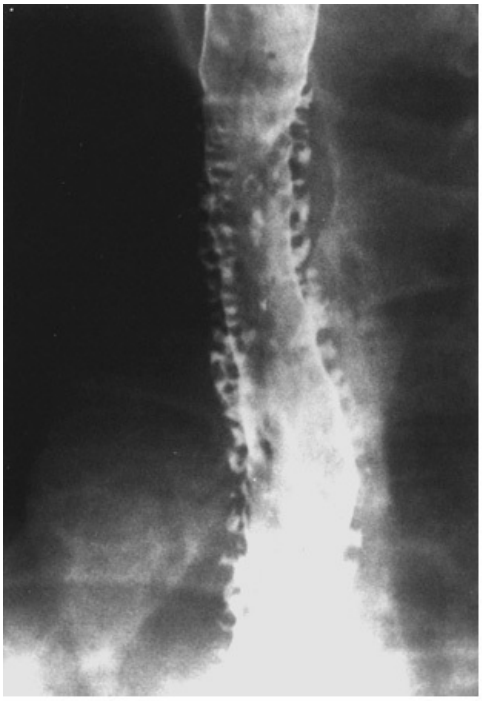

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Intramural esophageal pseudodiverticulosis associated with chronic obstruction. Invaginations of contrast into the esophageal wall outline deep esophageal glands.

Tumors

Esophageal cancer occurs in about 4.5:100,000 people in the United States with the associated mortality being only slightly less at 4.4:100,000. It is about 10 times less common than colorectal cancer but kills about one-quarter as many patients. These statistics emphasize both the rarity and lethality of esophageal cancer. One notable trend is the shift of dominant esophageal cancer type from squamous cell to adenocarcinoma, strongly linked to reflux disease and Barrett's metaplasia. Other distinctions between cell types are the predilection for adenocarcinoma to affect white males in the distal esophagus and squamous cell to affect black males in the more proximal esophagus with the added risk factors of smoking, alcohol consumption, caustic injury, and human papilloma virus infection **(Chap. 91)**.

The typical presentation of esophageal cancer is of progressive solid food dysphagia and weight loss. Associated symptoms may include odynophagia, iron deficiency, and, with mid-esophageal tumors, hoarseness from left recurrent laryngeal nerve injury. Generally, these are indications of locally invasive or even metastatic disease manifest by tracheoesophageal fistulas, and vocal cord paralysis. Even when detected as a small lesion, esophageal cancer has poor survival because of the abundant esophageal lymphatics leading to regional lymph node metastases.

Benign esophageal tumors are uncommon and usually discovered incidentally. In decreasing frequency of occurrence, cell types include leiomyomas, fibrovascular polyps, squamous papillomas, granular cell, lipomas, neurofibromas, and inflammatory fibroid polyps. These generally become symptomatic only when they are associated with dysphagia and merit removal only under the same circumstances.

CONGENITAL ANOMALIES

The most common congenital esophageal anomaly is esophageal atresia, occurring in about 1 in 5,000 live births. Atresia can occur in several permutations, the common denominator being developmental failure of fusion between the proximal and distal esophagus associated with a tracheoesophageal fistula, most commonly with the distal segment excluded. Alternatively, there can be an H-type configuration in which esophageal fusion has occurred, but with a tracheoesophageal fistula. Esophageal atresia is usually recognized and corrected surgically within the first few days of life. Later life complications include dysphagia from anastomotic strictures or absent peristalsis and reflux, which can be severe. Less common developmental anomalies include congenital esophageal stenosis, webs, and duplications.

Dysphagia can also result from congenital abnormalities that cause extrinsic compression of the esophagus. In dysphagia lusoria, the esophagus is compressed by an aberrant right subclavian artery arising from the descending aorta and passing behind the esophagus. Alternatively vascular rings may surround and constrict the esophagus.

Heterotopic gastric mucosa, also known as an esophageal inlet patch, is a focus of gastric type epithelium in the proximal cervical esophagus; the estimated prevalence is 4.5%. The inlet patch is thought to result from incomplete replacement of embryonic columnar epithelium with squamous epithelium. The majority of patches are asymptomatic, but acid production can occur as most contain fundic type gastric epithelium with parietal cells.

ESOPHAGEAL MOTALITY DISORDERS

Esophageal motility disorders are diseases attributable to esophageal neuromuscular dysfunction commonly associated with dysphagia, chest pain, or heartburn. The major entities are achalasia, diffuse esophageal spasm (DES), and GERD. Motility disorders can also be secondary to broader disease processes as is the case with pseudoachalasia, Chagas' disease, and scleroderma. Not included in this discussion are diseases affecting the pharynx and proximal esophagus, impairment of which is almost always part of a more global neuromuscular disease process.

Achalasia

Achalasia is a rare disease caused by loss of ganglion cells within the esophageal myenteric plexus with a population incidence of about 1:100,000 and usually presenting between age 25 and 60. With long-standing disease, virtual aganglionosis is noted. Excitatory (cholinergic) ganglionic neurons are variably affected and inhibitory (nitric oxide) ganglionic neurons are necessarily involved. Functionally, inhibitory neurons mediate deglutitive lower esophageal sphincter (LES) relaxation and the sequential propagation of peristalsis. Their absence leads to impaired deglutitive LES relaxation and absent peristalsis. Increasing evidence suggests that the ultimate cause of ganglion cell degeneration in achalasia is an autoimmune process attributable to a latent infection with human herpes simplex virus 1 combined with genetic susceptibility.

Long-standing achalasia is characterized by progressive dilatation and sigmoid deformity of the esophagus with hypertrophy of the LES. Clinical manifestations may include dysphagia, regurgitation, chest pain, and weight loss. Most patients report solid and liquid food dysphagia. Regurgitation occurs when food, fluid, and secretions are retained in the dilated esophagus. Patients with advanced achalasia are at risk for bronchitis, pneumonia, or lung abscess from chronic regurgitation and aspiration. Chest pain is frequent early in the course of achalasia, thought to result from esophageal spasm. Patients describe a squeezing, pressure-like retrosternal pain, sometimes radiating to the neck, arms, jaw, and back. Paradoxically, some patients complain of heartburn that may be a chest pain equivalent. Treatment of achalasia is less effective in relieving chest pain than it is in relieving dysphagia or regurgitation.

The differential diagnosis of achalasia includes DES, Chagas' disease, and pseudoachalasia. Chagas' disease is endemic in areas of central Brazil, Venezuela, and northern Argentina, spread by the bite of the reduvid (kissing) bug that transmits the protozoan, *Trypanosoma cruzi*. The chronic phase of the disease develops years after infection and results from destruction of autonomic ganglion cells throughout the body, including the heart, gut, urinary tract, and respiratory tract. Tumor infiltration, most commonly seen with carcinoma in the gastric fundus or distal esophagus can mimic idiopathic achalasia. The resultant "pseudoachalasia" accounts for up to 5% of suspected cases and is more likely with advanced age, abrupt onset of symptoms (<1 year), and weight loss. Hence, endoscopy should be part of the evaluation of achalasia. When the clinical suspicion for pseudoachalasia is high and endoscopy nondiagnostic, CT scanning or endoscopic ultrasonography may be of value. Rarely, pseudoachalasia can result from a paraneoplastic syndrome with circulating antineuronal antibodies.

Achalasia is diagnosed by barium swallow x-ray and/or esophageal manometry; endoscopy has a relatively minor role other than to exclude pseudoachalasia. The barium swallow x-ray appearance is of a dilated esophagus with poor emptying, an air-fluid level, and tapering at the LES giving it a beak-like appearance **(Fig. 292-5)**. Occasionally, an epiphrenic diverticulum is observed. In long-standing achalasia, the esophagus may assume a sigmoid configuration. The diagnostic criteria for achalasia with esophageal manometry are impaired LES relaxation and absent peristalsis. High-resolution manometry has somewhat advanced this diagnosis; three subtypes of achalasia are differentiated based on the pattern of pressurization in the nonperistaltic esophagus **(Fig. 292-6)**. Because manometry identifies early disease before esophageal dilatation and food retention, it is the most sensitive diagnostic test.

Figure 292-5

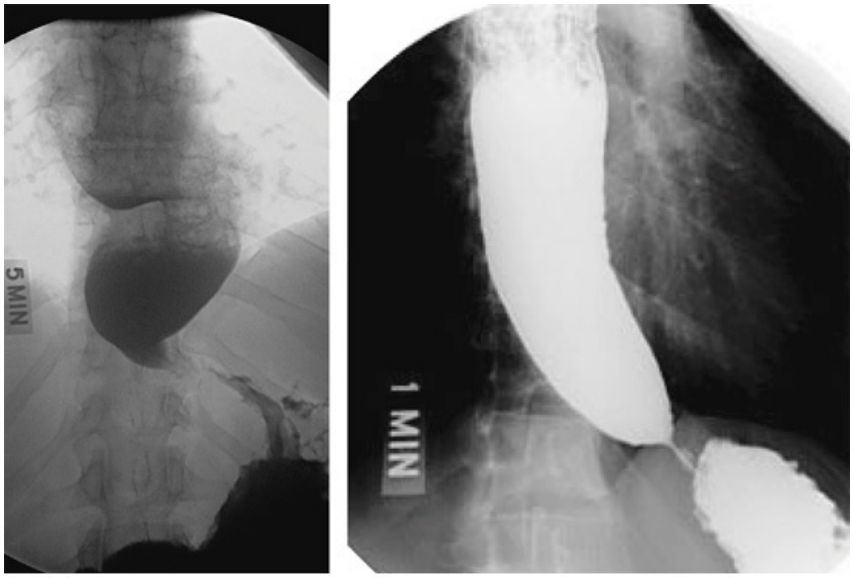

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Achalasia with esophageal dilatation, tapering at the gastroesophageal junction and an air-fluid level within the esophagus. The example on the left shows sigmoid deformity with very advanced disease.

Figure 292-6

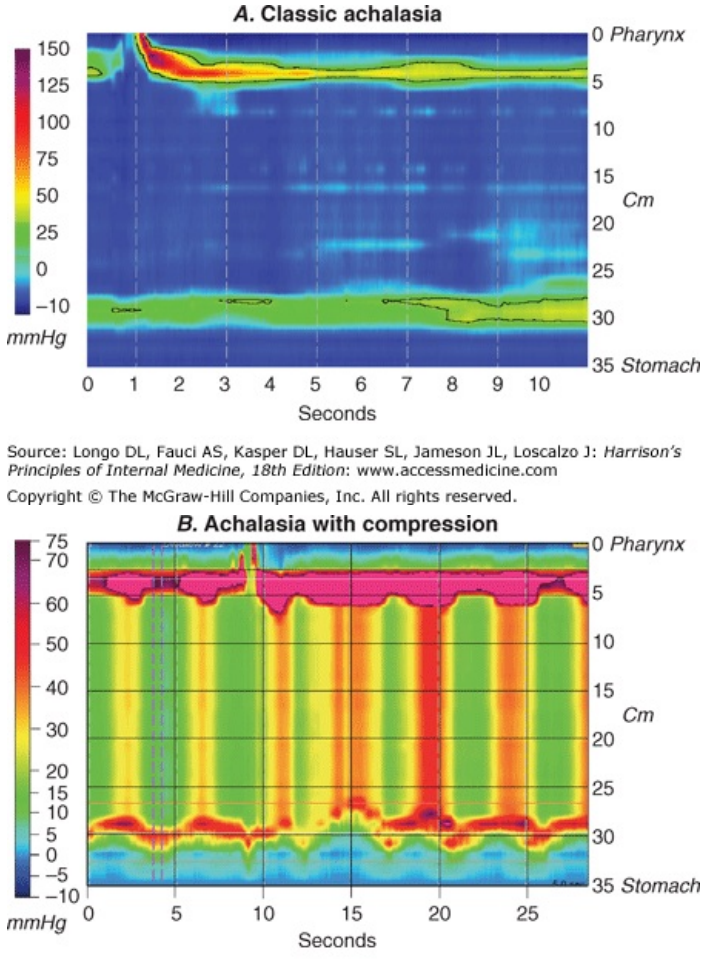

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

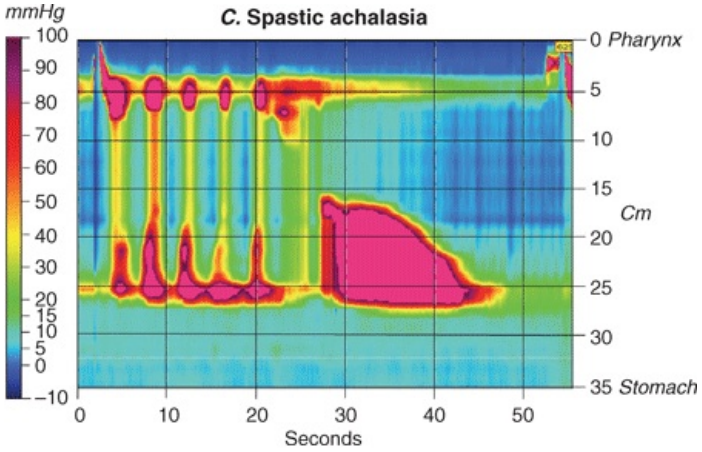

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Three subtypes of achalasia: classic (*Panel A*), with esophageal compression (*Panel B*), and spastic achalasia (*Panel C*) imaged with pressure topography. All are characterized by impaired lower esophageal sphincter (LES) relaxation and absent peristalsis. However, classic achalasia has minimal pressurization of the esophageal body while substantial fluid pressurization is observed in achalasia with esophageal compression and spastic esophageal contractions are observed with spastic achalasia.

There is no known way of preventing or reversing achalasia. Therapy is directed at reducing LES pressure so that gravity and esophageal pressurization promote esophageal emptying. Peristalsis rarely, if ever, returns. LES pressure can be reduced by pharmacologicals therapy, forceful dilatation, or surgical myotomy. No large, controlled trials of the therapeutic alternatives exist and the optimal approach is debated. Pharmacologicals therapies are relatively ineffective but are often used as temporizing therapies. Nitrates or calcium channel blockers are administered before eating, advising caution because of their effects on blood pressure. Botulinum toxin, injected into the LES under endoscopic guidance, inhibits acetylcholine release from nerve endings and improves dysphagia in about 66% of cases for at least 6 months. Sildenafil, or alternative phosphodiesterase inhibitors, effectively decrease LES pressure, but practicalities limit their clinical use in achalasia.

The only durable therapies for achalasia are pneumatic dilatation and Heller myotomy. Pneumatic dilatation, with a reported efficacy ranging from 32–98%, is an endoscopic technique using a noncompliant, cylindrical balloon dilator positioned across the LES and inflated to a diameter of 3–4 cm. The major complication is perforation with a reported incidence of 1–5%. The most common surgical procedure for achalasia is laparoscopic Heller myotomy, usually performed in conjunction with an antireflux procedure (partial fundoplication); good to excellent results are reported in 62–100% of cases. Occasionally, patients with advanced disease fail to respond to pneumatic dilatation or Heller myotomy. In such refractory cases, esophageal resection with gastric pull-up or interposition of a segment of transverse colon may be the only option other than gastrostomy feeding.

In untreated or inadequately treated achalasia, esophageal dilatation predisposes to stasis esophagitis. Prolonged stasis esophagitis is the likely explanation for the association between achalasia and esophageal squamous cell cancer. Tumors develop after years of achalasia, usually in the setting of a greatly dilated esophagus with the overall squamous cell cancer risk increased 17-fold compared to controls.

Diffuse Esophageal Spasm (Des)

DES is manifested by episodes of dysphagia and chest pain attributable to abnormal esophageal contractions with normal deglutitive LES relaxation. Beyond that, there is little consensus. The pathophysiology and natural history of DES are ill defined. Radiographically, DES has been characterized by tertiary contractions or a "corkscrew esophagus" **(Fig. 292-7)**, but in many instances these abnormalities are actually indicative of achalasia. Manometrically, a variety of defining features have been proposed including uncoordinated ("spastic") activity in the distal esophagus, spontaneous and repetitive contractions, or high amplitude and prolonged contractions. Greatest consensus exists with the concept that simultaneous contractions define DES. All of these definitions lead to patients with a variety of disorders being diagnosed as DES. In fact, high-resolution manometry suggests that DES, when defined in a restrictive fashion **(Fig. 292-8)**, is actually much less common than achalasia and suspected cases are often incorrectly categorized achalasia.

Figure 292-7

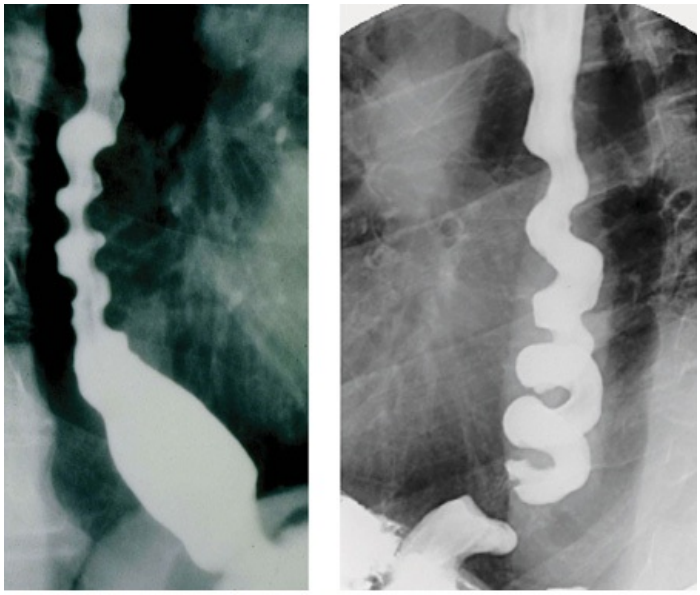

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diffuse esophageal spasm. The characteristic "corkscrew" esophagus results from spastic contraction of the circular muscle in the esophageal wall; more precisely, this is actually a helical array of muscle. These findings are also seen with spastic achalasia.

Figure 292-8

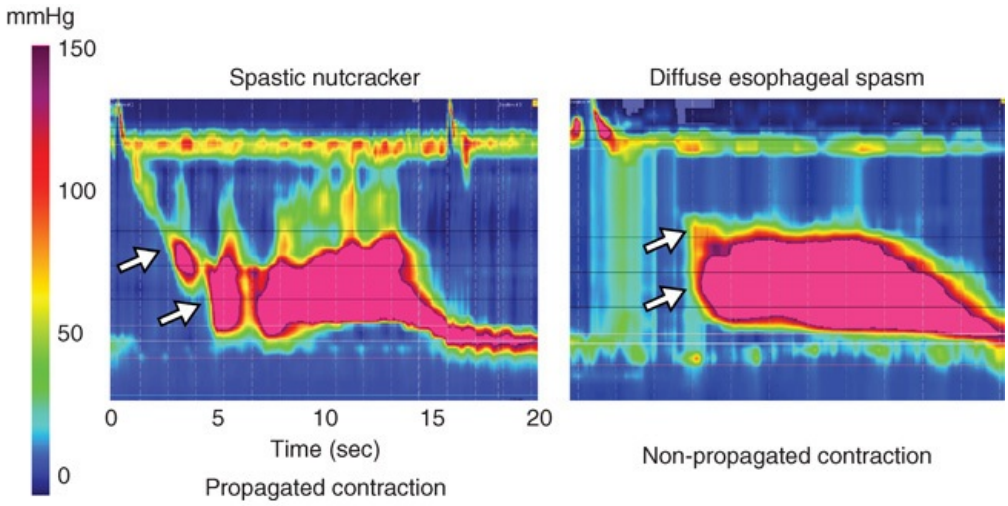

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Esophageal pressure topography of the two major variants of esophageal spasm: spastic nutcracker (*left*) and diffuse esophageal spasm (*right*). Spastic nutcracker is defined by the extraordinarily vigorous and repetitive contractions with normal peristaltic onset. Diffuse esophageal spasm is similar but primarily defined by a rapid propagation at the onset of the contraction.

Esophageal chest pain closely mimics angina pectoris. Features suggesting esophageal pain include pain that is nonexertional, prolonged, interrupts sleep, is meal-related, is relieved with antacids, and is accompanied by heartburn, dysphagia, or regurgitation. However, all of these features exhibit overlap with cardiac pain, which still must be the primary consideration. Furthermore, even within the spectrum of esophageal diseases, both chest pain and dysphagia are also characteristic of peptic or infectious esophagitis. Only after these more common entities have been excluded by evaluation and/or treatment should a diagnosis of DES be pursued.

Although the defining criteria are currently disputed, DES is diagnosed by manometry. Endoscopy is useful to identify alternative structural and inflammatory lesions that may cause chest pain. Radiographically, a "corkscrew esophagus," "rosary bead esophagus," pseudodiverticula, or curling can be indicative of DES, but these are also found with spastic achalasia. Given these vagaries of defining DES, and the resultant heterogeneity of patients identified for inclusion in therapeutic trials, it

is not surprising that trial results have been disappointing. Only small, uncontrolled trials exist, reporting response to nitrates, calcium channel blockers, hydralazine, botulinum toxin, and anxiolytics. The only controlled trial showing efficacy was with an anxiolytic. Surgical therapy (long myotomy or even esophagectomy) should be considered only with severe weight loss or unbearable pain. These indications are extremely rare.

Nonspecific Manometric Findings

Manometric studies done to evaluate chest pain and/or dysphagia often report minor abnormalities (hypertensive or hypotensive peristalsis, hypertensive LES, etc.) that are insufficient to diagnose either achalasia or DES. These findings are of unclear significance. Reflux and psychiatric diagnoses, particularly anxiety and depression, are common among such individuals. A lower visceral pain threshold and symptoms of irritable bowel syndrome are noted in more than half of such patients. Consequently, therapy for these individuals should either target the most common esophageal disorder, GERD, or more global conditions such as depression or somatization neurosis that are found to be coexistent.

GASTROESOPHAGEAL REFLUX DISEASE (GERD)

The current conception of GERD is to encompass a family of conditions with the commonality that they are caused by the gastroesophageal reflux resulting in either troublesome symptoms or an array of potential esophageal and extraesophageal manifestations. It is estimated that 15% of adults in the United States are affected by GERD, although such estimates are based only on self-reported chronic heartburn. With respect to the esophagus, the spectrum of injury includes esophagitis, stricture, Barrett's esophagus, and adenocarcinoma **(Fig. 292-9)**. Of particular concern is the rising incidence of esophageal adenocarcinoma, an epidemiologic trend that parallels the increasing incidence of GERD. There were about 8,000 incident cases of esophageal adenocarcinoma in the United States in 2010 (half of all esophageal cancers); it is estimated that this disease burden has increased two- to sixfold in the last 20 years.

Figure 292-9

A Erosive esophagitis

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B Esophageal stricture with chronic erosive esophagitis

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

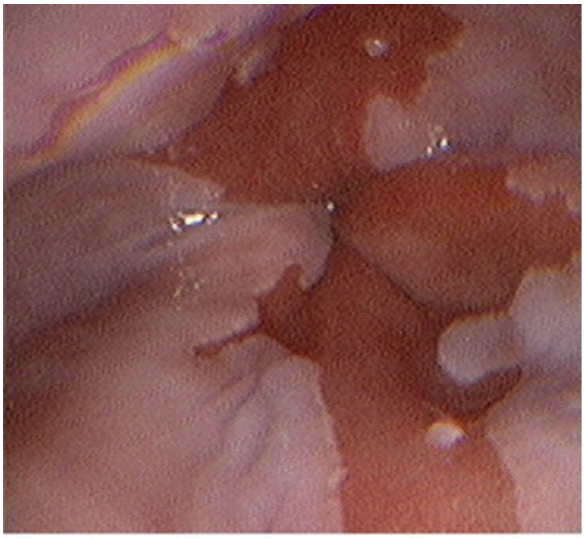

C Barrett's esophagus

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D Esophageal adenocarcinoma with Barrett's esophagus

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Endoscopic appearance of (*A*) peptic esophagitis, (*B*) a peptic stricture, (*C*) Barrett's metaplasia, and (*D*) adenocarcinoma developing within an area of Barrett's esophagus.

Pathophysiology

The best defined subset of GERD patients, albeit a minority overall, have esophagitis. Esophagitis occurs when refluxed gastric acid and pepsin cause necrosis of the esophageal mucosa causing erosions and ulcers. Note that some degree of gastroesophageal reflux is normal, physiologically intertwined with the mechanism of belching (transient LES relaxation), but esophagitis results from excessive reflux, often accompanied by impaired clearance of the refluxed gastric juice. Restricting reflux to that which is physiologically intended depends on the anatomic and physiologic integrity of the esophagogastric junction, a complex sphincter comprised of both the LES and the surrounding crural diaphragm. Three dominant mechanisms of esophagogastric junction incompetence are recognized: (1) transient LES relaxations (a vagovagal reflex in which LES relaxation is elicited by gastric distention), (2) LES hypotension, or (3) anatomic distortion of the esophagogastric junction inclusive of hiatus hernia. Of note, the third factor, esophagogastric junction anatomic disruption, is both significant unto itself and also because it interacts with the first two mechanisms. Transient LES relaxations account for at least 90% of reflux in normal subjects or GERD patients without hiatus hernia, but patients with hiatus hernia have a more heterogeneous mechanistic profile. Factors tending to exacerbate reflux regardless of mechanism are abdominal obesity, pregnancy, gastric hypersecretory states, delayed gastric emptying, disruption of esophageal peristalsis, and gluttony.

After acid reflux, peristalsis returns the refluxed fluid to the stomach and acid clearance is completed by titration of the residual acid by bicarbonate contained in swallowed saliva. Consequently, two causes of prolonged acid clearance are impaired peristalsis and reduced salivation. Impaired peristaltic emptying can be attributable to disrupted peristalsis or superimposed reflux associated with a hiatal hernia. With superimposed reflux, fluid retained within a sliding hiatal hernia refluxes back into the esophagus during swallow-related LES relaxation, a phenomenon that does not normally occur.

Inherent in the pathophysiologic model of GERD is that gastric juice is harmful to the esophageal epithelium. However, gastric acid hypersecretion is usually not a dominant factor in the development of esophagitis. An obvious exception is with Zollinger-Ellison syndrome, which is associated with severe esophagitis in about 50% of patients. Another caveat is with chronic *H. pylori* gastritis, which may have a protective effect by inducing atrophic gastritis with concomitant hypoacidity. Pepsin, bile, and pancreatic enzymes within gastric secretions can also injure the esophageal epithelium, but their noxious properties are either lessened in an acidic environment or dependent on acidity for activation. Bile warrants attention because it persists in refluxate despite acid-suppressing medications. Bile can transverse the cell membrane, imparting severe cellular injury in a weakly acidic environment, and has also been invoked as a cofactor in the pathogenesis of Barrett's metaplasia and adenocarcinoma. Hence, the causticity of gastric refluxate extends beyond hydrochloric acid.

Symptoms

Heartburn and regurgitation are the typical symptoms of GERD. Somewhat less common are dysphagia and chest pain. In each case, multiple potential mechanisms for symptom genesis operate that extend beyond the basic concepts of mucosal erosion and activation of afferent sensory nerves.

Specifically, hypersensitivity and functional pain are increasingly recognized as confounding factors. Nonetheless the dominant clinical strategy is of empirical treatment with acid inhibitors, reserving further evaluation for those who fail to respond. Important exceptions to this are patients with chest pain or persistent dysphagia, each of which may be indicative of more morbid conditions. With chest pain, cardiac disease must be carefully considered. In the case of persistent dysphagia, chronic reflux can lead to the development of a peptic stricture or adenocarcinoma, each of which benefits from early detection and/or specific therapy.

Extraesophageal syndromes with an established association to GERD include chronic cough, laryngitis, asthma, and dental erosions. A multitude of other conditions including pharyngitis, chronic bronchitis, pulmonary fibrosis, chronic sinusitis, cardiac arrhythmias, sleep apnea, and recurrent aspiration pneumonia have proposed associations with GERD. However, in both cases it is important to emphasize the word association as opposed to causation. In many instances the disorders likely coexist because of shared pathogenetic mechanisms rather than strict causality. Potential mechanisms for extraesophageal GERD manifestations are of either regurgitation with direct contact between the refluxate and supra-esophageal structures or via a vagovagal reflex wherein reflux activation of esophageal afferent nerves triggers efferent vagal reflexes such as bronchospasm, cough, or arrhythmias.

Differential Diagnosis

Although generally quite characteristic, symptoms from GERD need to be distinguished from symptoms related to infectious, pill, or eosinophilic esophagitis, peptic ulcer disease, dyspepsia, biliary colic, coronary artery disease, and esophageal motility disorders. It is especially important that coronary artery disease be given early consideration because of its potentially lethal implications. The remaining elements of the differential diagnosis can be addressed by endoscopy, upper gastrointestinal series, or biliary tract ultrasonography as appropriate. The distinction among etiologies of esophagitis is usually easily made by endoscopy with mucosal biopsies, which are necessary to evaluate for eosinophilic inflammation. In terms of endoscopic appearance, infectious esophagitis is diffuse and tends to involve the proximal esophagus far more frequently than does reflux esophagitis. The ulcerations seen in peptic esophagitis are usually solitary and distal, whereas infectious ulcerations are punctate and diffuse. Eosinophilic esophagitis characteristically exhibits multiple esophageal rings, linear furrows, or white punctate exudate. Esophageal ulcerations from pill esophagitis are usually singular and deep at points of luminal narrowing, especially near the carina, with sparing of the distal esophagus.

Complications

The complications of GERD are related to chronic esophagitis (bleeding and stricture) and the relationship between GERD and esophageal adenocarcinoma. However, both esophagitis and peptic strictures have become increasingly rare in the era of potent antisecretory medications. Conversely, the most severe histologic consequence of GERD is Barrett's metaplasia with the associated risk of esophageal adenocarcinoma, and the incidence of these lesions has increased, not decreased in the era of potent acid suppression. Barrett's metaplasia, endoscopically recognized by tongues of reddish mucosa extending proximally from the gastroesophageal junction **(Fig. 292-9)** or histopathologically by the finding of specialized columnar metaplasia, is associated with at least a 20-fold increased risk for development of esophageal adenocarcinoma.

Barrett's metaplasia can progress to adenocarcinoma through the intermediate stages of low- and high-grade dysplasia **(Fig. 292-10)**. Owing to this risk, areas of Barrett's and especially any included areas of mucosal irregularity should be extensively biopsied. The rate of cancer development is estimated at 0.5% per year, but vagaries in definition and of the extent of Barrett's metaplasia requisite to establish the diagnosis have contributed to variability and inconsistency in this risk assessment. The group at greatest risk is obese white males in their sixth decade of life. However, despite common practice, the utility of endoscopic screening and surveillance programs intended to control the adenocarcinoma risk has not been established. Also of note, no high-level evidence confirms that aggressive antisecretory therapy or antireflux surgery causes regression of Barrett's esophagus or prevents adenocarcinoma.

Figure 292-10

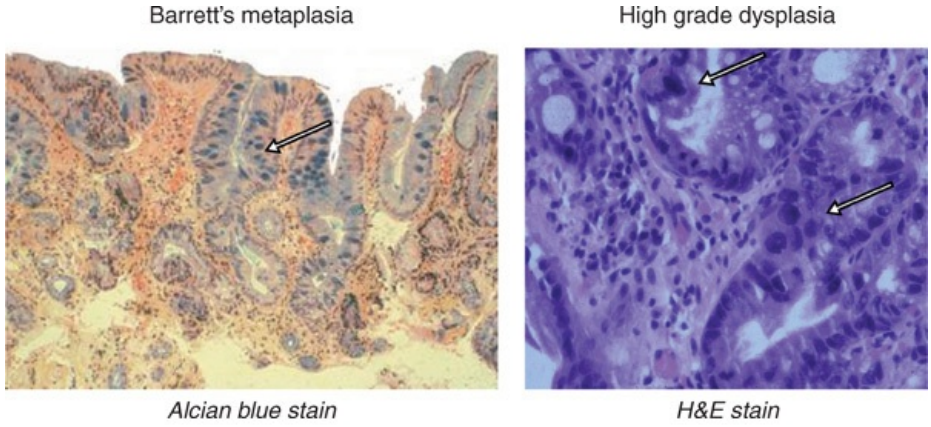

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histopathology of Barrett's metaplasia and Barrett's with high-grade dysplasia. H&E, hematoxylin and eosin.

Although the management of Barrett's esophagus remains controversial, the finding of dysplasia in Barrett's, particularly high-grade dysplasia, mandates further intervention. In addition to the high rate of progression to adenocarcinoma, there is also a high prevalence of unrecognized coexisting cancer with high-grade dysplasia. Nonetheless, treatment remains controversial. Esophagectomy, intensive endoscopic surveillance, and mucosal ablation have all been advocated. Currently, most experts advocate esophagectomy as treatment for high-grade dysplasia in an otherwise healthy patient with minimal surgical risk. However, esophagectomy has a mortality ranging from 3–10%, along with substantial morbidity. That, along with increasing evidence of the effectiveness of endoscopic therapy with purpose-built radio frequency ablation devices, has led many to favor this therapy as a preferable alternative.

Treatment: Gastroesophageal Reflux Disease (GERD)

Lifestyle modifications are routinely advocated as GERD therapy. Broadly speaking, these fall into three categories:(1) avoidance of foods that reduce lower esophageal sphincter pressure, making them "refluxogenic" (these commonly include fatty foods, alcohol, spearmint, peppermint, tomato-based foods, possibly coffee and tea); (2) avoidance of acidic foods that are inherently irritating; and (3) adoption of behaviors to minimize reflux and/or heartburn. In general, minimal evidence supports the efficacy of these measures. However, clinical experience dictates that subsets of patients are benefitted by specific recommendations, based on their unique history and symptom profile. A patient with sleep disturbance from nighttime heartburn is likely to benefit from elevation of the head of the bed and avoidance of eating before retiring, but those recommendations are superfluous for a patient without nighttime symptoms. The most broadly applicable recommendation is for weight reduction. Even though the benefit with respect to reflux cannot be assured, the strong epidemiologic association between obesity and GERD and the secondary health gains of weight reduction are beyond dispute.

The dominant pharmacologic approach to GERD management is with inhibitors of gastric acid secretion and abundant data support the effectiveness of this approach. Pharmacologically reducing the acidity of gastric juice does not prevent reflux, but it ameliorates reflux symptoms and allows esophagitis to heal. The hierarchy of effectiveness among pharmaceuticals parallels their antisecretory potency. Proton pump inhibitors (PPIs) are more efficacious than histamine$_2$ receptor antagonists (H_2RAs), and both are superior to placebo. No major differences exist among PPIs and only modest gain is achieved by increased dosage.

Paradoxically, the perceived frequency and severity of heartburn correlate poorly with the presence or severity of esophagitis. When GERD treatments are assessed in terms of resolving heartburn, both efficacy and differences among pharmaceuticals are less clear-cut than with the objective of healing esophagitis. Although the same overall hierarchy of effectiveness exists, observed efficacy rates are lower and vary widely, likely reflective of patient heterogeneity.

Reflux symptoms tend to be chronic, irrespective of esophagitis. Thus, a common management strategy is indefinite treatment with PPIs or H_2RAs as necessary for symptom control. The side effects of PPI therapy are generally minimal. Vitamin B12, calcium, and iron absorption may be compromised and susceptibility to enteric infections, particularly *Clostridium difficile* colitis increased with treatment. Consequently, as with any medication, dosage should be minimized to that necessary.

Laparoscopic Nissen fundoplication, wherein the proximal stomach is wrapped around the distal esophagus to create an antireflux barrier, is a surgical alternative to the management of chronic GERD. Just as with PPI therapy, evidence on the utility of fundoplication is strongest for treating esophagitis and controlled trials suggest similar efficacy to PPI therapy. However, the benefits of

fundoplication must be weighed against potential deleterious effects, including surgical morbidity and mortality, postoperative dysphagia, failure or breakdown requiring reoperation, an inability to belch, and increased bloating, flatulence, and bowel symptoms after surgery.

EOSINOPHILIC ESOPHAGITIS

Eosinophilic esophagitis (EoE) is increasingly recognized in adults and children around the world. Population-based studies suggest the prevalence to be in excess of 1:1000 with a predilection for white males. The increasing prevalence of EoE is attributable to a combination of an increasing incidence and a growing awareness of the condition. There is also an incompletely understood, but important, overlap between EoE and GERD that delays or confuses diagnosis of the disease in many cases.

EoE is diagnosed based on the combination of typical esophageal symptoms and esophageal mucosal biopsies demonstrating esophageal squamous epithelial infiltration with eosinophils. Secondary etiologies of esophageal eosinophilia including GERD, drug hypersensitivity, connective tissue disorders, hypereosinophilic syndrome and infection are excluded. Current evidence indicates that EoE is an allergic disorder induced by antigen sensitization in susceptible individuals. Studies have demonstrated an important role for dietary allergens in both the pathogenesis and treatment of EoE. Aeroallergens may also contribute but there is much less evidence in this regard. The natural history of the disorder is uncertain as are the consequences of not treating asymptomatic or minimally symptomatic patients.

EoE should be strongly considered in children and adults with dysphagia and food impactions, regardless of the presence or absence of heartburn. Other symptoms may include atypical chest pain and heartburn, particularly heartburn that is refractory to PPI therapy. An atopic history of food allergy, asthma, eczema, or allergic rhinitis is present in the majority of patients. Cytokines such as IL-5, eotaxin, and thymus and activation-related chemokine (TARC) may be elevated in the serum. The characteristic endoscopic findings include multiple esophageal rings, linear furrows, and punctate exudates **(Fig. 292-11)**. Histologic confirmation is made with the demonstration of increased eosinophils in the esophagal mucosa (generally ge; 15 eosinophils per high-power field) **(Fig. 292-12)**. Fibrosis, narrow caliber esophagus, and stricture can occur with EoE, but the predictive variables for these are not known. Complications of disease include food impaction and esophageal perforation.

Figure 292-11

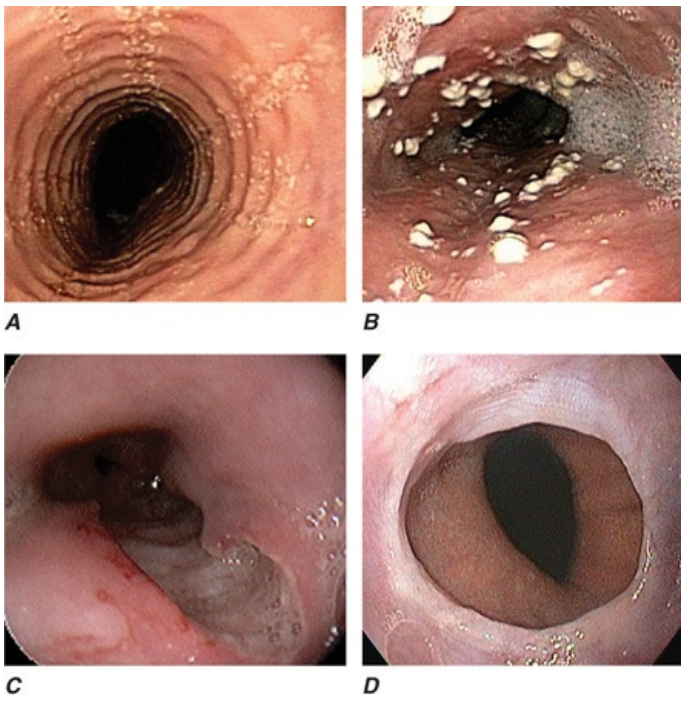

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Endoscopic features of (*A*) eosinophilic esophagitis (EoE), (*B*) *Candida* esophagitis, (*C*) giant ulcer associated with HIV, (*D*) and a Schatzki ring.

Figure 292-12

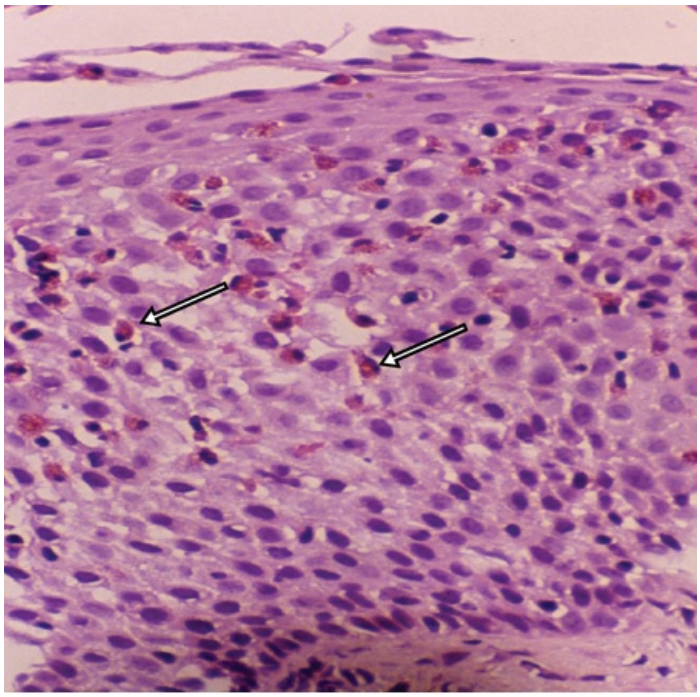

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histopathology of eosinophilic esophagitis (EoE) showing dense infiltration of the esophageal squamous epithelium with eosinophils. Eosinophilic inflammation can also be seen with gastroesophageal reflux disease (GERD); the optimal discriminatory threshold for EoE is greater than 15 eosinophils per high-power field.

Treatments for EoE include dietary restrictions, PPIs, systemic or topical glucocorticoids, montelukast, immunomodulators, and endoscopic dilatation of strictures. Notably, allergy testing [radioallergosorbent test (RAST), skin prick testing] has demonstrated limited specificity in the identification of causative foods. Once esophageal eosinophilia is demonstrated, patients in whom GERD may be a confounding factor should undergo a trial of PPI therapy to determine if this results in clinical or histologic improvement. If symptoms and eosinophilia persist despite PPI therapy, other treatment options should be pursued. Topical glucocorticoids (fluticasone propionate or budesonide) are the most commonly used treatment in adults, but dietary restriction has proven effective primarily in pediatric studies. Systemic glucocorticoids are reserved for severely afflicted patients refractory to less morbid treatments. Esophageal dilation should be approached cautiously in patients with stricturing because of concerns for increased risk of esophageal mural disruption and perforation.

INFECTIOUS ESOPHAGITIS

With the increased use of immunosuppression for organ transplantation as well as chronic inflammatory diseases and chemotherapy along with the AIDSepidemic, infections with *Candida* species, *Herpesvirus,* and cytomegalovirus (CMV) have become relatively common. Although rare, infectious esophagitis also occurs among the nonimmunocompromised, with herpes simplex and *Candida albicans* being the most common pathogens. Among AIDS patients, infectious esophagitis becomes more common as the CD4 count declines; cases are rare with the CD4 count >200 and common when <100. HIV itself may also be associated with a self-limited syndrome of acute esophageal ulceration with oral ulcers and a maculopapular skin rash at the time of seroconversion. Additionally, some patients with advanced disease have deep, persistent esophageal ulcers treated with oral glucocorticoids or thalidomide. However, with the widespread use of protease inhibitors, a reduction in these HIV complications has been noted.

Regardless of the infectious agent, odynophagia is a characteristic symptom of infectious esophagitis; dysphagia, chest pain, and hemorrhage are also common. Odynophagia is uncommon with reflux esophagitis, so its presence should always raise suspicion of an alternative etiology.

Candida Esophagitis

Candida is normally found in the throat, but can become pathogenic and produce esophagitis in a compromised host; *C. albicans* is most common. *Candida* esophagitis also occurs with esophageal stasis secondary to esophageal motor disorders and diverticula. Patients complain of odynophagia and dysphagia. If oral thrush is present, empirical therapy is appropriate, but coinfection is common, and persistent symptoms should lead to prompt endoscopy with biopsy, which is the most useful diagnostic evaluation. *Candida* esophagitis has a characteristic appearance of white plaques with

friability. Rarely, *Candida* esophagitis is complicated by bleeding, perforation, stricture, or systemic invasion. Oral fluconazole (200 mg on the first day, followed by 100 mg daily) for 7–14 days is the preferred treatment. Patients refractory to fluconazole may respond to itraconazole. Alternatively, poorly responsive patients or those who cannot swallow medications can be treated with an intravenous echinocandin (caspofungin 50 mg daily for 7–21 days). Amphotericin B (10–15 mg IV infusion for 6 h daily to a total dose of 300–500 mg) is used in severe cases.

Herpetic Esophagitis

Herpes simplex virus type 1 or 2 may cause esophagitis. Vesicles on the nose and lips may coexist and are suggestive of a herpetic etiology. Varicella-zoster virus can also cause esophagitis in children with chickenpox or adults with zoster. The characteristic endoscopic findings are vesicles and small, punched-out ulcerations. Because herpes simplex infections are limited to squamous epithelium, biopsies from the ulcer margins are most likely to reveal the characteristic ground glass nuclei, eosinophilic Cowdry's type A inclusion bodies, and giant cells. Culture or polymerase chain reaction (PCR) assays are helpful to identify acyclovir-resistant strains. The infection is often self-limited after a 1–2 week period. Acyclovir (400 mg orally 5 times a day for 14–21 days) or valacyclovir (1 g orally tid for 7 days) reduces this morbidity. In patients with severe odynophagia, intravenous acyclovir, 5 mg/kg every 8 h for 7–14 days, foscarnet (90 mg/kg intravenously bid for 2–4 weeks) or oral famciclovir are used.

Cytomegalovirus

CMV esophagitis occurs only in immunocompromised patients, particularly transplant recipients. CMV is usually activated from a latent stage or may be acquired from transfusions. Endoscopically, CMV lesions appear as serpiginous ulcers in an otherwise normal mucosa, particularly in the distal esophagus. Biopsies of the ulcer bases have the highest diagnostic yield for finding the pathognomonic large nuclear or cytoplasmic inclusion bodies. Immunohistology with monoclonal antibodies to CMV and in situ hybridization tests are useful for early diagnosis. Ganciclovir, 5 mg/kg every 12 h intravenously, is the treatment of choice. Valganciclovir (900 mg bid), an oral formulation of ganciclovir, or foscarnet (90 mg/kg every 12 h intravenously) can also be used. Therapy is continued until healing, which may take 3–6 weeks.

MECHANICAL TRAUMA AND IATROGENIC INJURY

Esophageal Perforation

Most cases of esophageal perforation are from instrumentation of the esophagus or trauma. Alternatively, forceful vomiting or retching can lead to spontaneous rupture at the gastroesophageal junction (Boerhaave's syndrome). More rarely, corrosive esophagitis or neoplasms lead to perforation. Instrumental perforation from endoscopy or nasogastric tube placement typically occurs in the hypopharynx or at the gastroesophageal junction. Perforation may also result at the site of stricture in the setting of endoscopic food disimpaction or esophageal dilation. Esophageal perforation causes pleuritic retrosternal pain that can be associated with pneumomediastinum and subcutaneous emphysema. Mediastinitis is a major complication of esophageal perforation, and prompt recognition is key to optimizing outcome. CT of the chest is most sensitive in detecting mediastinal air. Esophageal perforation is confirmed by a contrast swallow; usually Gastrografin followed by thin barium. Treatment includes nasogastric suction and parenteral broad-spectrum antibiotics with prompt surgical drainage and repair in noncontained leaks. Conservative therapy with NPO status and antibiotics without surgery may be appropriate in cases of minor instrumental perforation that are detected early. Endoscopic clipping or stent placement may be indicated in nonoperable cases such as perforated tumors.

Mallory-Weiss Tear

Vomiting, retching, or vigorous coughing can cause a nontransmural tear at the gastroesophageal junction that is a common cause of upper gastrointestinal bleeding. Most patients present with hematemesis. Antecedent vomiting is anticipated but not always evident. Bleeding usually abates spontaneously, but protracted bleeding may respond to local epinephrine or cauterization therapy, endoscopic clipping, or angiographic embolization. Surgery is rarely needed.

Radiation Esophagitis

Radiation esophagitis can complicate treatment for thoracic cancers, especially breast and lung, with the risk proportional to radiation dosage. Radiosensitizing drugs such as doxorubicin, bleomycin, cyclophosphamide, and cisplatin also increase the risk. Dysphagia and odynophagia may last weeks to months after therapy. The esophageal mucosa becomes erythematous, edematous, and friable. Submucosal fibrosis and degenerative tissue changes and stricturing may occur years after the radiation exposure. Radiation exposure in excess of 5000 cGY has been associated with increased risk of esophageal stricture. Treatment for acute radiation esophagitis is supportive. Chronic strictures are managed with esophageal dilation.

Corrosive Esophagitis

Caustic esophageal injury from ingestion of alkali or, less commonly, acid can be accidental or from attempted suicide. Absence of oral injury does not exclude possible esophageal involvement. Thus, early endoscopic evaluation is recommended to assess and grade the injury to the esophageal mucosa. Severe corrosive injury may lead to esophageal perforation, bleeding, stricture, and death. Glucocorticoids have not been shown to improve the clinical outcome of acute corrosive esophagitis and are not recommended. Healing of more severe grades of caustic injury is commonly associated with severe stricture formation and often requires repeated dilatation.

Pill Esophagitis

Pill-induced esophagitis occurs when a swallowed pill fails to traverses the entire esophagus and lodges within the lumen. Generally, this is attributed to poor "pill taking habits": inadequate liquid with the pill, or lying down immediately after taking a pill. The most common location for the pill to lodge is in the mid-esophagus near the crossing of the aorta or carina. Extrinsic compression from these structures halts the movement of the pill or capsule. Since initially reported in 1970, more than 1000 cases of pill esophagitis have been reported, suggesting that this is not an unusual occurrence. A wide variety of medications are implicated with the most common being doxycycline, tetracycline, quinidine, phenytoin, potassium chloride, ferrous sulfate, nonsteroidal anti-inflammatory drugs (NSAIDs), and bisphosphonates. However, virtually any pill can result in pill esophagitis if taken carelessly.

Typical symptoms of pill esophagitis are the sudden onset of chest pain and odynophagia. Characteristically, the pain will develop over a period of hours or will awaken the individual from sleep. A classic history in the setting of ingestion of recognized pill offenders obviates the need for diagnostic testing in most patients. When endoscopy is performed, localized ulceration or inflammation is evident. Histologically, acute inflammation is typical. Chest CT imaging will sometimes reveal esophageal thickening consistent with transmural inflammation. Although the condition usually resolves within days to weeks, symptoms may persist for months and stricture can develop in severe cases. No specific therapy is known to hasten the healing process, but antisecretory medications are frequently prescribed to remove concomitant reflux as an aggravating factor. When healing results in stricture formation, dilatation is indicated.

Foreign Bodies and Food Impaction

Food or foreign bodies may lodge in the esophagus causing complete obstruction, causing an inability to handle secretions (foaming at the mouth) and severe chest pain. Food impaction may occur due to stricture, carcinoma, Schatzki ring, eosinophilic esophagitis, or simply inattentive eating. If it does not spontaneously resolve, impacted food is dislodged endoscopically. Use of meat tenderizer enzymes to facilitate passage of a meat bolus is discouraged because of potential esophageal injury. Glucagon (1 mg IV) is sometimes tried before endoscopic dislodgement. After emergent treatment patients should be evaluated for potential causes of the impaction with treatment rendered as indicated.

ESOPHAGEAL MANIFESTATIONS OF SYSTEMIC DISEASE

Scleroderma and Collagen Vascular Diseases

Scleroderma esophagus (hypotensive LES and absent esophageal peristalsis) was initially described as a manifestation of scleroderma or other collagen vascular diseases and thought to be specific for these disorders. However, this nomenclature subsequently proved unfortunate and has been discarded because an estimated half of qualifying patients do not have an identifiable systemic disease, and reflux disease is often the only identifiable association. When scleroderma esophagus occurs as a manifestation of a collagen vascular disease, the histopathologic findings are of infiltration and destruction of the esophageal muscularis propria with collagen deposition and fibrosis. The pathogenesis of absent peristalsis and LES hypotension in the absence of a collagen vascular disease is unknown. Regardless of the underlying cause, the manometric abnormalities predispose patients to severe GERD due to inadequate LES barrier function combined with poor esophageal clearance of refluxed acid. Dysphagia may also be manifest but is generally mild and alleviated by eating in an upright position and using liquids to facilitate solid emptying.

Dermatologic Diseases

A host of dermatologic disorders (pemphigus vulgaris, bullous pemphigoid, cicatricial pemphigoid, Behçet's syndrome, epidermolysis bullosa) can affect the oropharynx and esophagus, particularly the proximal esophagus with blisters, bullae, webs, and strictures. Glucocorticoid treatment is usually effective. Erosive lichen planus, Stevens-Johnson syndrome, and graft-versus-host disease can also involve the esophagus. Esophageal dilatation may be necessary to treat strictures.

FURTHER READINGS

Furuta GT et al: Eosinophilic esophagitis in children and adults: A systematic review and consensus recommendations for diagnosis and treatment. Gastroenterology 133:1342, 2007[PMID: 17919504] [Full Text]

Jacobsen B et al: Body-mass index and symptoms of gastroesophageal reflux in women. N Engl J Med

354:2340, 2006

Kahrilas PJ: Clinical practice. Gastroesophageal reflux disease. N Engl J Med 359:1700, 2008[PMID: 18923172] [Full Text]

———: Esophageal motor disorders in terms of high resolution esophageal pressure topography: What has changed? Am J Gastroenterol 105:981, 2010

———, Pandolfino JE: Hiatus hernia. GI Motility Online, *http://www.nature.com/gimo/contents/pt1/full/gimo48.html,* 2006

——— et al: AGAI medical position statement: Management of gastroesophageal reflux disease. Gastroenterology 135:1383, 2008

Sampliner RE, Sharma P (eds): *Barrett's Esophagus and Esophageal Adenocarcinoma,* 2nd ed. Malden, MA, Blackwell, 2006

Shaheen NJ et al: Radiofrequency ablation in Barrett's esophagus with dysplasia. N Engl J Med 360:2277, 2009[PMID: 19474425] [Full Text]

Walzer N, Hirano I: Achalasia. Gastroenterol Clin North Am 37:807, 2008[PMID: 19028319] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PEPTIC ULCER DISEASE

Burning epigastric pain exacerbated by fasting and improved with meals is a symptom complex associated with peptic ulcer disease (PUD). An *ulcer* is defined as disruption of the mucosal integrity of the stomach and/or duodenum leading to a local defect or excavation due to active inflammation. Ulcers occur within the stomach and/or duodenum and are often chronic in nature. Acid peptic disorders are very common in the United States, with 4 million individuals (new cases and recurrences) affected per year. Lifetime prevalence of PUD in the United States is ~12% in men and 10% in women. Moreover, an estimated 15,000 deaths per year occur as a consequence of complicated PUD. The financial impact of these common disorders has been substantial, with an estimated burden on direct and indirect health care costs of ~$10 billion per year in the United States.

Gastric Physiology

Despite the constant attack on the gastroduodenal mucosa by a host of noxious agents (acid, pepsin, bile acids, pancreatic enzymes, drugs, and bacteria), integrity is maintained by an intricate system that provides mucosal defense and repair.

GASTRIC ANATOMY

The gastric epithelial lining consists of rugae that contain microscopic gastric pits, each branching into four or five gastric glands made up of highly specialized epithelial cells. The makeup of gastric glands varies with their anatomic location. Glands within the gastric cardia comprise <5% of the gastric gland area and contain mucous and endocrine cells. The 75% of gastric glands are found within the oxyntic mucosa and contain mucous neck, parietal, chief, endocrine, enterochromaffin, and enterochromaffin-like (ECL) cells **(Fig. 293-1)**. Pyloric glands contain mucous and endocrine cells (including gastrin cells) and are found in the antrum.

Figure 293-1

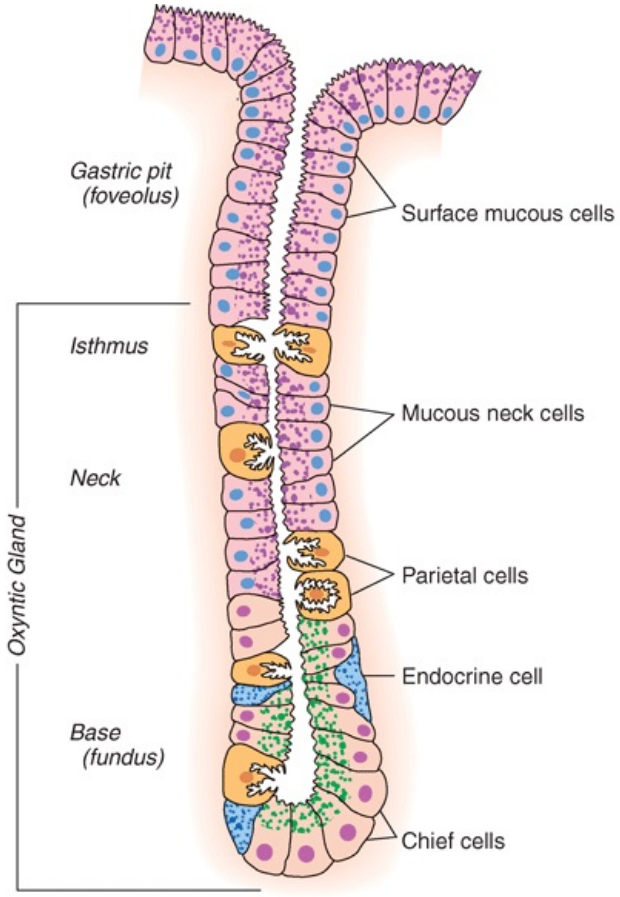

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagrammatic representation of the oxyntic gastric gland. *(Adapted from S Ito, RJ Winchester: Cell Biol 16:541, 1963. © The Rockefeller University Press.)*

The parietal cell, also known as the oxyntic cell, is usually found in the neck, or isthmus, or in the oxyntic gland. The resting, or unstimulated, parietal cell has prominent cytoplasmic tubulovesicles and intracellular canaliculi containing short microvilli along its apical surface **(Fig. 293-2).** H^+,K^+-adenosine triphosphatase (ATPase) is expressed in the tubulovesicle membrane; upon cell stimulation, this membrane, along with apical membranes, transforms into a dense network of apical intracellular canaliculi containing long microvilli. Acid secretion, a process requiring high energy, occurs at the apical canalicular surface. Numerous mitochondria (30–40% of total cell volume) generate the energy required for secretion.

Figure 293-2

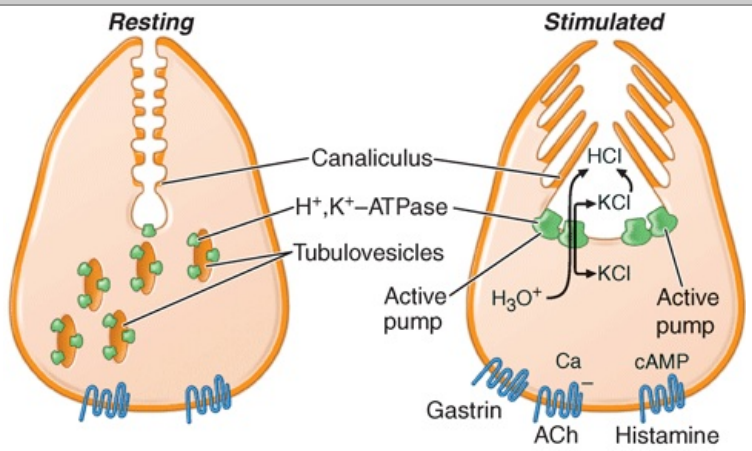

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gastric parietal cell undergoing transformation after secretagogue-mediated stimulation. cAMP, cyclic adenosine monophosphate. *(Adapted from SJ Hersey, G Sachs: Physiol Rev 75:155, 1995.)*

GASTRODUODENAL MUCOSAL DEFENSE

The gastric epithelium is under constant assault by a series of endogenous noxious factors, including hydrochloric acid (HCl), pepsinogen/pepsin, and bile salts. In addition, a steady flow of exogenous substances such as medications, alcohol, and bacteria encounter the gastric mucosa. A highly intricate biologic system is in place to provide defense from mucosal injury and to repair any injury that may occur.

The mucosal defense system can be envisioned as a three-level barrier, composed of preepithelial, epithelial, and subepithelial elements **(Fig. 293-3)**. The first line of defense is a mucus-bicarbonate-phospholipid layer, which serves as a physicochemical barrier to multiple molecules, including hydrogen ions. Mucus is secreted in a regulated fashion by gastroduodenal surface epithelial cells. It consists primarily of water (95%) and a mixture of phospholipids and glycoproteins (mucin). The mucous gel functions as a nonstirred water layer impeding diffusion of ions and molecules such as pepsin. Bicarbonate, secreted in a regulated manner by surface epithelial cells of the gastroduodenal mucosa into the mucous gel, forms a pH gradient ranging from 1 to 2 at the gastric luminal surface and reaching 6 to 7 along the epithelial cell surface.

Figure 293-3

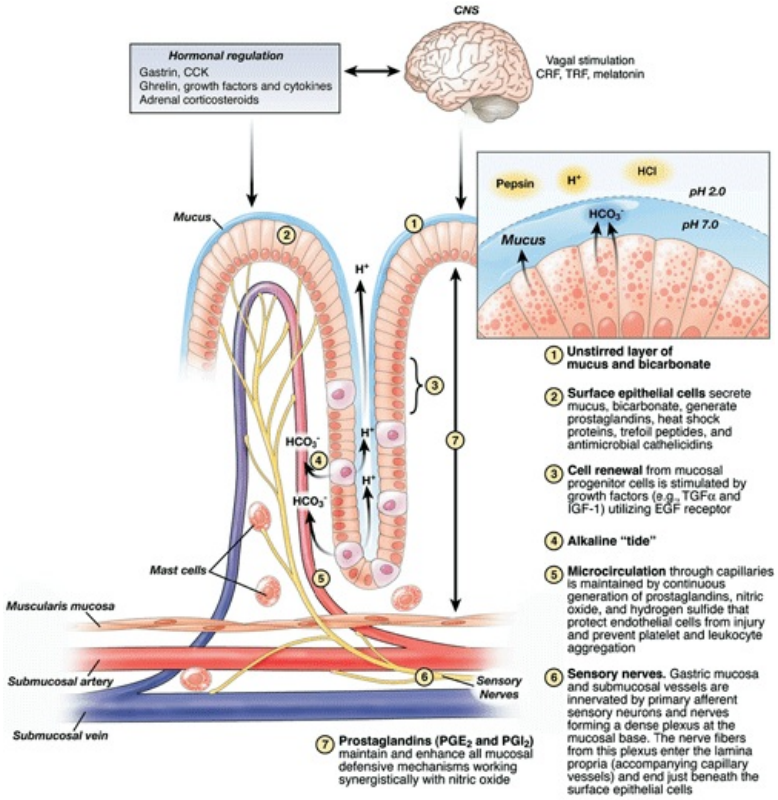

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Components involved in providing gastroduodenal mucosal defense and repair. CCK, cholecystokinin; CRF, corticotropin-releasing factor; EGF, epidermal growth factor; HCl, hydrochloride; IGF, insulin-like growth factor; TGFα, transforming growth factor α; TRF, thyrotropin releasing factor. *(Modified and updated from Tarnawski A. Cellular and molecular mechanisms of mucosal defense and repair. In: Yoshikawa T, Arakawa T. Bioregulation and Its Disorders in the Gastrointestinal Tract. Tokyo, Japan: Blackwell Science, 1998:3–17.)*

Surface epithelial cells provide the next line of defense through several factors, including mucus production, epithelial cell ionic transporters that maintain intracellular pH and bicarbonate production, and intracellular tight junctions. Surface epithelial cells generate heat shock proteins that prevent protein denaturation and protect cells from certain factors such as increased temperature, cytotoxic agents, or oxidative stress. Epithelial cells also generate trefoil factor family peptides and cathelicidins, which also play a role in surface cell protection and regeneration. If the preepithelial barrier were breached, gastric epithelial cells bordering a site of injury can migrate to restore a damaged region (*restitution*). This process occurs independent of cell division and requires uninterrupted blood flow

and an alkaline pH in the surrounding environment. Several growth factors, including epidermal growth factor (EGF), transforming growth factor (TGF) α, and basic fibroblast growth factor (FGF), modulate the process of restitution. Larger defects that are not effectively repaired by restitution require cell proliferation. Epithelial cell regeneration is regulated by prostaglandins and growth factors such as EGF and TGF-α. In tandem with epithelial cell renewal, formation of new vessels (*angiogenesis*) within the injured microvascular bed occurs. Both FGF and vascular endothelial growth factor (VEGF) are important in regulating angiogenesis in the gastric mucosa.

An elaborate microvascular system within the gastric submucosal layer is the key component of the subepithelial defense/repair system, providing HCO_3^-, which neutralizes the acid generated by the parietal cell. Moreover, this microcirculatory bed provides an adequate supply of micronutrients and oxygen while removing toxic metabolic by-products.

Prostaglandins play a central role in gastric epithelial defense/repair **(Fig. 293-4)**. The gastric mucosa contains abundant levels of prostaglandins that regulate the release of mucosal bicarbonate and mucus, inhibit parietal cell secretion, and are important in maintaining mucosal blood flow and epithelial cell restitution. Prostaglandins are derived from esterified arachidonic acid, which is formed from phospholipids (cell membrane) by the action of phospholipase A_2. A key enzyme that controls the rate-limiting step in prostaglandin synthesis is cyclooxygenase (COX), which is present in two isoforms (COX-1, COX-2), each having distinct characteristics regarding structure, tissue distribution, and expression. COX-1 is expressed in a host of tissues, including the stomach, platelets, kidneys, and endothelial cells. This isoform is expressed in a constitutive manner and plays an important role in maintaining the integrity of renal function, platelet aggregation, and gastrointestinal (GI) mucosal integrity. In contrast, the expression of COX-2 is inducible by inflammatory stimuli, and it is expressed in macrophages, leukocytes, fibroblasts, and synovial cells. The beneficial effects of nonsteroidal anti-inflammatory drugs (NSAIDs) on tissue inflammation are due to inhibition of COX-2; the toxicity of these drugs (e.g., GI mucosal ulceration and renal dysfunction) is related to inhibition of the COX-1 isoform. The highly COX-2–selective NSAIDs have the potential to provide the beneficial effect of decreasing tissue inflammation while minimizing toxicity in the GI tract. Selective COX-2 inhibitors have had adverse effects on the cardiovascular system, leading to increased risk of myocardial infarction. Therefore, the FDA has removed two of these agents (valdecoxib and rofecoxib) from the market (see below).

Figure 293-4

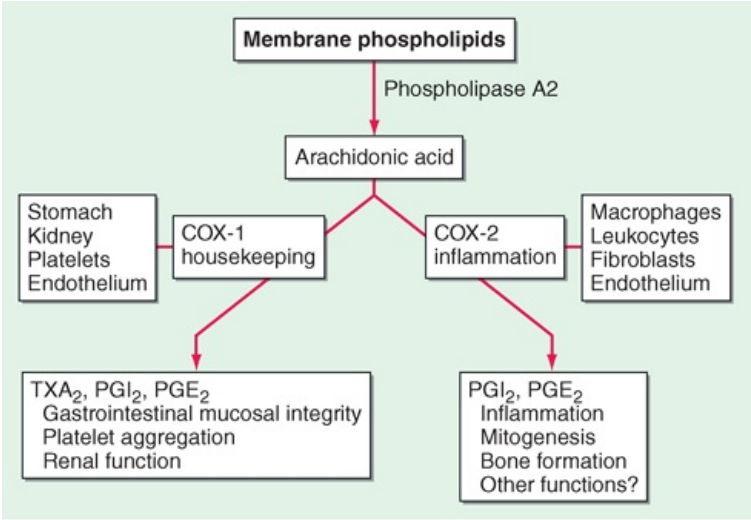

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of the steps involved in synthesis of prostaglandin E2 (PGE2) and prostacyclin (PGI2). Characteristics and distribution of the cyclooxygenase (COX) enzymes 1 and 2 are also shown. TXA_2, thromboxane A_2.

Nitric oxide (NO) is important in the maintenance of gastric mucosal integrity. The key enzyme NO synthase is constitutively expressed in the mucosa and contributes to cytoprotection by stimulating gastric mucus, increasing mucosal blood flow and maintaining epithelial cell barrier function. The central nervous system (CNS) and hormonal factors also play a role in regulating mucosal defense through multiple pathways (Fig. 293-3).

PHYSIOLOGY OF GASTRIC SECRETION

Hydrochloric acid and pepsinogen are the two principal gastric secretory products capable of inducing

mucosal injury. Gastric acid and pepsinogen play a physiologic role in protein digestion, absorption of iron and vitamin B12 as well as killing ingested bacteria. Acid secretion should be viewed as occurring under basal and stimulated conditions. Basal acid production occurs in a circadian pattern, with highest levels occurring during the night and lowest levels during the morning hours. Cholinergic input via the vagus nerve and histaminergic input from local gastric sources are the principal contributors to basal acid secretion. Stimulated gastric acid secretion occurs primarily in three phases based on the site where the signal originates (cephalic, gastric, and intestinal). Sight, smell, and taste of food are the components of the cephalic phase, which stimulates gastric secretion via the vagus nerve. The gastric phase is activated once food enters the stomach. This component of secretion is driven by nutrients (amino acids and amines) that directly stimulate the G cell to release gastrin, which in turn activates the parietal cell via direct and indirect mechanisms. Distention of the stomach wall also leads to gastrin release and acid production. The last phase of gastric acid secretion is initiated as food enters the intestine and is mediated by luminal distention and nutrient assimilation. A series of pathways that inhibit gastric acid production are also set into motion during these phases. The GI hormone somatostatin is released from endocrine cells found in the gastric mucosa (D cells) in response to HCl. Somatostatin can inhibit acid production by both direct (parietal cell) and indirect mechanisms (decreased histamine release from ECL cells and gastrin release from G cells). Additional neural (central and peripheral) and humoral [amylin, atrial natriuretic peptide (ANP), cholecystokinin, ghrelin, obestatin, secretin, and serotonin] factors play a role in counterbalancing acid secretion. Under physiologic circumstances, these phases occur simultaneously. Ghrelin, the appetite-regulating hormone expressed in stomach, may stimulate gastric acid secretion through a vagal-mediated mechanism, but this remains to be confirmed.

The acid-secreting parietal cell is located in the oxyntic gland, adjacent to other cellular elements (ECL cell, D cell) important in the gastric secretory process **(Fig. 293-5)**. This unique cell also secretes intrinsic factor (IF). The parietal cell expresses receptors for several stimulants of acid secretion, including histamine (H_2), gastrin (cholecystokinin B/gastrin receptor), and acetylcholine (muscarinic, M_3). Binding of histamine to the H_2 receptor leads to activation of adenylate cyclase and an increase in cyclic adenosine monophosphate (AMP). Activation of the gastrin and muscarinic receptors results in activation of the protein kinase C/phosphoinositide signaling pathway. Each of these signaling pathways in turn regulates a series of downstream kinase cascades that control the acid-secreting pump, H^+,K^+-ATPase. The discovery that different ligands and their corresponding receptors lead to activation of different signaling pathways explains the potentiation of acid secretion that occurs when histamine and gastrin or acetylcholine are combined. More importantly, this observation explains why blocking one receptor type (H_2) decreases acid secretion stimulated by agents that activate a different pathway (gastrin, acetylcholine). Parietal cells also express receptors for ligands that inhibit acid production (prostaglandins, somatostatin, and EGF). Histamine also stimulates gastric acid secretion indirectly by activating the histamine H_3 receptor on D-cells, which inhibits somatostatin release.

Figure 293-5

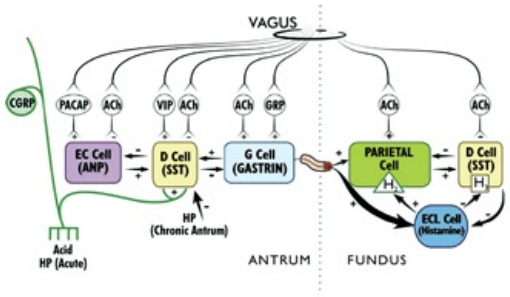

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Regulation of gastric acid secretion at the cellular level. ACh, acetylcholine; ANP, atrial natriuretic peptide; CGRP, calcitoningene-related peptide; EC, enterochromaffin; ECL, enterochromaffin-like; GRP, gastrin-releasing peptide; PACAP, pituitary adenylate-cyclase activating peptide; SST, somatostatin; VIP, vasoactive intestinal peptide.

The enzyme H^+,K^+-ATPase is responsible for generating the large concentration of H^+. It is a membrane-bound protein that consists of two subunits, α and β. The active catalytic site is found within the α subunit; the function of the β subunit is unclear. This enzyme uses the chemical energy of adenosine triphosphate (ATP) to transfer H^+ ions from parietal cell cytoplasm to the secretory canaliculi in exchange for K^+. The H^+,K^+-ATPase is located within the secretory canaliculus and in nonsecretory cytoplasmic tubulovesicles. The tubulovesicles are impermeable to K^+, which leads to an

inactive pump in this location. The distribution of pumps between the nonsecretory vesicles and the secretory canaliculus varies according to parietal cell activity (Fig. 293-2). Proton pumps are recycled back to the inactive state in cytoplasmic vesicles once parietal cell activation ceases.

The chief cell, found primarily in the gastric fundus, synthesizes and secretes pepsinogen, the inactive precursor of the proteolyticenzyme pepsin. The acid environment within the stomach leads to cleavage of the inactive precursor to pepsin and provides the low pH (<2) required for pepsin activity. Pepsin activity is significantly diminished at a pH of 4 and irreversibly inactivated and denatured at a pH of ≥7. Many of the secretagogues that stimulate acid secretion also stimulate pepsinogen release. The precise role of pepsin in the pathogenesis of PUD remains to be established.

Pathophysiologic Basis of Peptic Ulcer Disease

PUD encompasses both gastric and duodenal ulcers. *Ulcers* are defined as breaks in the mucosal surface >5 mm in size, with depth to the submucosa. Duodenal ulcers (DUs) and gastric ulcers (GUs); share many common features in terms of pathogenesis, diagnosis, and treatment, but several factors distinguish them from one another.

EPIDEMIOLOGY

Duodenal Ulcers

DUs are estimated to occur in 6–15% of the Western population. The incidence of DUs declined steadily from 1960 to 1980 and has remained stable since then. The death rates, need for surgery, and physician visits have decreased by >50% over the past 30 years. The reason for the reduction in the frequency of DUs is likely related to the decreasing frequency of *Helicobacter pylori.* Before the discovery of *H. pylori,* the natural history of DUs was typified by frequent recurrences after initial therapy. Eradication of *H. pylori* has greatly reduced these recurrence rates.

Gastric Ulcers

GUs tend to occur later in life than duodenal lesions, with a peak incidence reported in the sixth decade. More than one-half of GUs occur in males and are less common than DUs, perhaps due to the higher likelihood of GUs being silent and presenting only after a complication develops. Autopsy studies suggest a similar incidence of DUs and GUs.

PATHOLOGY

Duodenal Ulcers

DUs occur most often in the first portion of the duodenum (>95%), with ~90% located within 3 cm of the pylorus. They are usually ≤1 cm in diameter but can occasionally reach 3–6 cm (giant ulcer). Ulcers are sharply demarcated, with depth at times reaching the muscularis propria. The base of the ulcer often consists of a zone of eosinophilic necrosis with surrounding fibrosis. Malignant DUs are extremely rare.

Gastric Ulcers

In contrast to DUs, GUs can represent a malignancy and should be biopsied upon discovery. Benign GUs are most often found distal to the junction between the antrum and the acid secretory mucosa. Benign GUs are quite rare in the gastric fundus and are histologically similar to DUs. Benign GUs associated with *H. pylori* are also associated with antral gastritis. In contrast, NSAID-related GUs are not accompanied by chronic active gastritis but may instead have evidence of a chemical gastropathy, typified by foveolar hyperplasia, edema of the lamina propria, and epithelial regeneration in the absence of *H. pylori.* Extension of smooth-muscle fibers into the upper portions of the mucosa, where they are not typically found, may also occur.

PATHOPHYSIOLOGY

Duodenal Ulcers

H. pylori and NSAID-induced injury account for the majority of DUs. Many acid secretory abnormalities have been described in DU patients. Of these, average basal and nocturnal gastric acid secretion appears to be increased in DU patients as compared to controls; however, the level of overlap between DU patients and control subjects is substantial. The reason for this altered secretory process is unclear, but *H. pylori* infection may contribute. Accelerated gastric emptying of liquids has been noted in some DU patients, but its role in DU formation, if any, is unclear. Bicarbonate secretion is significantly decreased in the duodenal bulb of patients with an active DU as compared to control subjects. *H. pylori* infection may also play a role in this process (see below).

Gastric Ulcers

As in DUs, the majority of GUs can be attributed to either *H. pylori* or NSAID-induced mucosal damage. GUs that occur in the prepyloric area or those in the body associated with a DU or a duodenal scar are similar in pathogenesis to DUs. Gastric acid output (basal and stimulated) tends to be normal or decreased in GU patients. When GUs develop in the presence of minimal acid levels, impairment of mucosal defense factors may be present. Gastric ulcers have been classified based on

their location: Type I occur in the gastric body and tend to be associated with low gastric acid production; type II occur in the antrum and gastric acid can vary from low to normal; type III occur within 3 cm of the pylorus and are commonly accompanied by duodenal ulcers and normal or high gastric acid production; and type IV are found in the cardia and are associated with low gastric acid production.

Abnormalities in resting and stimulated pyloric sphincter pressure with a concomitant increase in duodenal gastric reflux have been implicated in some GU patients. Although bile acids, lysolecithin, and pancreatic enzymes may injure gastric mucosa, a definite role for these in GU pathogenesis has not been established. Delayed gastric emptying of solids has been described in GU patients but has not been reported consistently

H. Pylori and Acid Peptic Disorders

Gastric infection with the bacterium *H. pylori* accounts for the majority of PUD **(Chap. 151)**. This organism also plays a role in the development of gastric mucosa-associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma. Although the entire genome of *H. pylori* has been sequenced, it is still not clear how this organism, which resides in the stomach, causes ulceration in the duodenum, or whether its eradication will lead to a decrease in gastric cancer.

The Bacterium

The bacterium, initially named *Campylobacter pyloridis,* is a gram-negative microaerophilic rod found most commonly in the deeper portions of the mucous gel coating the gastric mucosa or between the mucous layer and the gastric epithelium. It may attach to gastric epithelium but under normal circumstances does not appear to invade cells. It is strategically designed to live within the aggressive environment of the stomach. It is S-shaped (~0.5–3 μm in size) and contains multiple sheathed flagella. Initially, *H. pylori* resides in the antrum but, over time, migrates toward the more proximal segments of the stomach. The organism is capable of transforming into a coccoid form, which represents a dormant state that may facilitate survival in adverse conditions. The genome of *H. pylori* (1.65 million base pairs) encodes ~1500 proteins. Among this multitude of proteins there are factors that are essential determinants of *H. pylori*–mediated pathogenesis and colonization such as the outer membrane protein (Hop proteins), urease, and the vacuolating cytotoxin (Vac A). Moreover, the majority of *H. pylori* strains contain a genomic fragment that encodes the cag pathogenicity island (cag-PAI). Several of the genes that make up cag-PAI encode components of a type IV secretion island that translocates Cag A into host cells. Once in the cell, Cag A activates a series of cellular events important in cell growth and cytokine production. *H. pylori* also has extensive genetic diversity that in turn enhances its ability to promote disease. The first step in infection by *H. pylori* is dependent on the bacteria's motility and its ability to produce urease. Urease produces ammonia from urea, an essential step in alkalinizing the surrounding pH. Additional bacterial factors include catalase, lipase, adhesins, platelet-activating factor, and pic B (induces cytokines). Multiple strains of *H. pylori* exist and are characterized by their ability to express several of these factors (Cag A, Vac A, etc.). It is possible that the different diseases related to *H. pylori* infection can be attributed to different strains of the organism with distinct pathogenic features.

Epidemiology

The prevalence of *H. pylori* varies throughout the world and depends largely on the overall standard of living in the region. In developing parts of the world, 80% of the population may be infected by the age of 20, whereas the prevalence is 20–50% in industrialized countries. In contrast, in the United States this organism is rare in childhood. The overall prevalence of *H. pylori* in the United States is ~30%, with individuals born before 1950 having a higher rate of infection than those born later. About 10% of Americans <30 years of age are colonized with the bacteria. The rate of infection with *H. pylori* in industrialized countries has decreased substantially in recent decades. The steady increase in the prevalence of *H. pylori* noted with increasing age is due primarily to a cohort effect, reflecting higher transmission during a period in which the earlier cohorts were children. It has been calculated through mathematical models that improved sanitation during the latter half of the nineteenth century dramatically decreased transmission of *H. pylori.* Moreover, with the present rate of intervention, the organism will be ultimately eliminated from the United States. Two factors that predispose to higher colonization rates include poor socioeconomic status and less education. These factors, not race, are responsible for the rate of *H. pylori* infection in blacks and Hispanic Americans being double the rate seen in whites of comparable age. Other risk factors for *H. pylori* infection are (1) birth or residence in a developing country, (2) domestic crowding, (3) unsanitary living conditions, (4) unclean food or water, and (5) exposure to gastric contents of an infected individual.

Transmission of *H. pylori* occurs from person to person, following an oral-oral or fecal-oral route. The risk of *H. pylori* infection is declining in developing countries. The rate of infection in the United States has fallen by >50% when compared to 30 years ago.

Pathophysiology

H. pylori infection is virtually always associated with a chronic active gastritis, but only 10–15% of

infected individuals develop frank peptic ulceration. The basis for this difference is unknown, but is likely due to a combination of host and bacterial factors some of which are outlined below. Initial studies suggested that >90% of all DUs were associated with *H. pylori,* but *H. pylori* is present in only 30–60% of individuals with GUs and 50–70% of patients with DUs. The pathophysiology of ulcers not associated with *H. pylori* or NSAID ingestion [or the rare Zollinger-Ellison syndrome (ZES)] is becoming more relevant as the incidence of *H. pylori* is dropping, particularly in the Western world (see below).

The particular end result of *H. pylori* infection (gastritis, PUD, gastric MALT lymphoma, gastric cancer) is determined by a complex interplay between bacterial and host factors **(Fig. 293-6)**.

1. *Bacterial factors:H. pylori* is able to facilitate gastric residence, induce mucosal injury, and avoid host defense. Different strains of *H. pylori* produce different virulence factors. A specific region of the bacterial genome, the pathogenicity island (cag-PAI), encodes the virulence factors Cag A and pic B. Vac A also contributes to pathogenicity, although it is not encoded within the pathogenicity island. These virulence factors, in conjunction with additional bacterial constituents, can cause mucosal damage, in part through their ability to target the host immune cells. For example, Vac A targets human CD4 T cells, inhibiting their proliferation and in addition can disrupt normal function of B cells, CD8 T cells, macrophages and mast cells. Multiple studies have demonstrated that *H. pylori* strains that are cag-PAI positive are associated with a higher risk of peptic ulcer disease, premalignant gastric lesions and gastric cancer than are strains that lack the cag-PAI. Urease, which allows the bacteria to reside in the acidic stomach, generates NH_3, which can damage epithelial cells. The bacteria produce surface factors that are chemotactic for neutrophils and monocytes, which in turn contribute to epithelial cell injury (see below). *H. pylori* makes proteases and phospholipases that break down the glycoprotein lipid complex of the mucous gel, thus reducing the efficacy of this first line of mucosal defense. *H. pylori* expresses adhesins (OMPs like BabA), which facilitate attachment of the bacteria to gastric epithelial cells. Although lipopolysaccharide (LPS) of gram-negative bacteria often plays an important role in the infection, *H. pylori* LPS has low immunologic activity compared to that of other organisms. It may promote a smoldering chronic inflammation.
2. *Host factors:* Studies in twins suggest that there may be genetic predisposition to acquire *H. pylori.* The inflammatory response to *H. pylori* includes recruitment of neutrophils, lymphocytes (T and B), macrophages, and plasma cells. The pathogen leads to local injury by binding to class II major histocompatability complex (MHC) molecules expressed on gastric epithelial cells, leading to cell death (*apoptosis*). Moreover, bacterial strains that encode cag-PAI can introduce Cag A into the host cells, leading to further cell injury and activation of cellular pathways involved in cytokine production. Elevated concentrations of multiple cytokines are found in the gastric epithelium of *H. pylori*–infected individuals, including interleukin (IL) 1α/β, IL-2, IL-6, IL-8, tumor necrosis factor (TNF) α, and interferon (IFN-γ). *H. pylori* infection also leads to both a mucosal and a systemic humoral response, which does not lead to eradication of the bacteria but further compounds epithelial cell injury. Additional mechanisms by which *H. pylori* may cause epithelial cell injury include (1) activated neutrophil-mediated production of reactive oxygen or nitrogen species and enhanced epithelial cell turnover and (2) apoptosis related to interaction with T cells (T helper 1, or T_H1, cells) and IFN-γ.

Figure 293-6

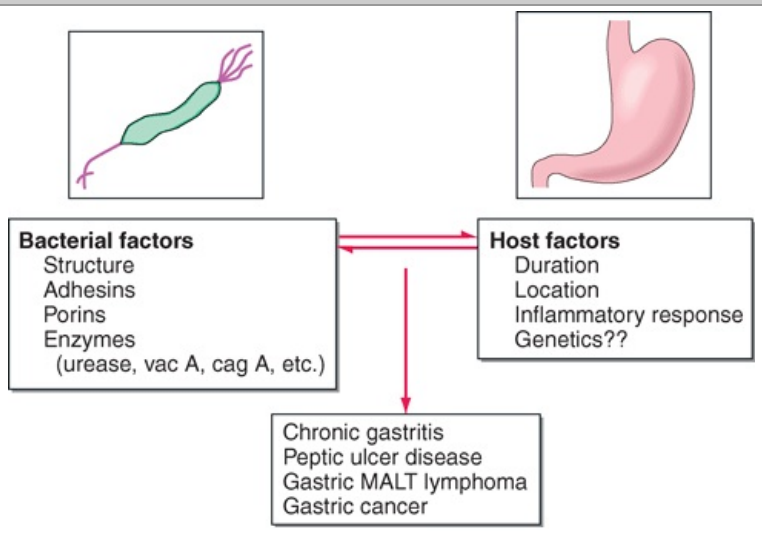

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Outline of the bacterial and host factors Important in determining *H. pylori*–induced

gastrointestinal disease. MALT, mucosalassociated lymphoid tissue.

The reason for *H. pylori*–mediated duodenal ulceration remains unclear. Studies suggest that *H. pylori* associated with duodenal ulceration may be more virulent. In addition, certain specific bacterial factors such as the duodenal ulcer-promoting gene A (*dupA*), may be associated with the development of duodenal ulcers. Another potential contributing factor is that gastric metaplasia in the duodenum of DU patients, which may be due to high acid exposure (see below), permits *H. pylori* to bind to it and produce local injury secondary to the host response. Another hypothesis is that *H. pylori* antral infection could lead to increased acid production, increased duodenal acid, and mucosal injury. Basal and stimulated [meal, gastrin-releasing peptide (GRP)] gastrin release are increased in *H. pylori*–infected individuals, and somatostatin-secreting D cells may be decreased. *H. pylori* infection might induce increased acid secretion through both direct and indirect actions of *H. pylori* and proinflammatory cytokines (IL-8, TNF, and IL-1) on G, D, and parietal cells **(Fig. 293-7)**. Gastric ulcers, in contrast, are associated with *H. pylori* induced pangastritis and normal or low gastric acid secretion. *H. pylori* infection has also been associated with decreased duodenal mucosal bicarbonate production. Data supporting and contradicting each of these interesting theories have been demonstrated. Thus, the mechanism by which *H. pylori* infection of the stomach leads to duodenal ulceration remains to be established.

Figure 293-7

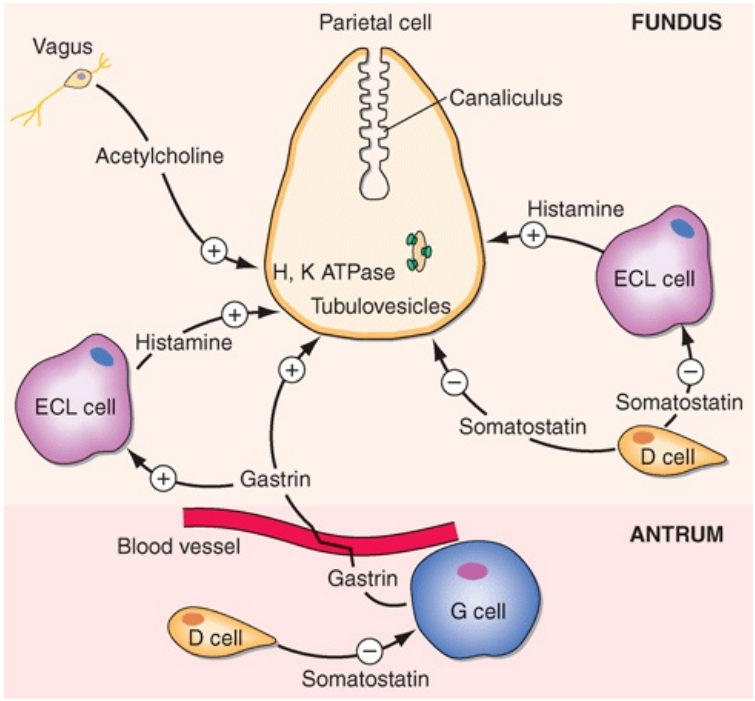

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Summary of potential mechanisms by which *H.pylori* may lead to gastric secretory abnormalities. D, somatostatin cell; ECL, enterochromaffin-like cell; G, G cell. *(Adapted from J Calam et al: Gastroenterology 113:543, 1997.)*

In summary, the final effect of *H. pylori* on the GI tract is variable and determined by microbial and host factors. The type and distribution of gastritis correlate with the ultimate gastric and duodenal pathology observed. Specifically, the presence of antral-predominant gastritis is associated with DU formation; gastritis involving primarily the corpus predisposes to the development of GUs, gastric atrophy, and ultimately gastric carcinoma **(Fig. 293-8)**.

Figure 293-8

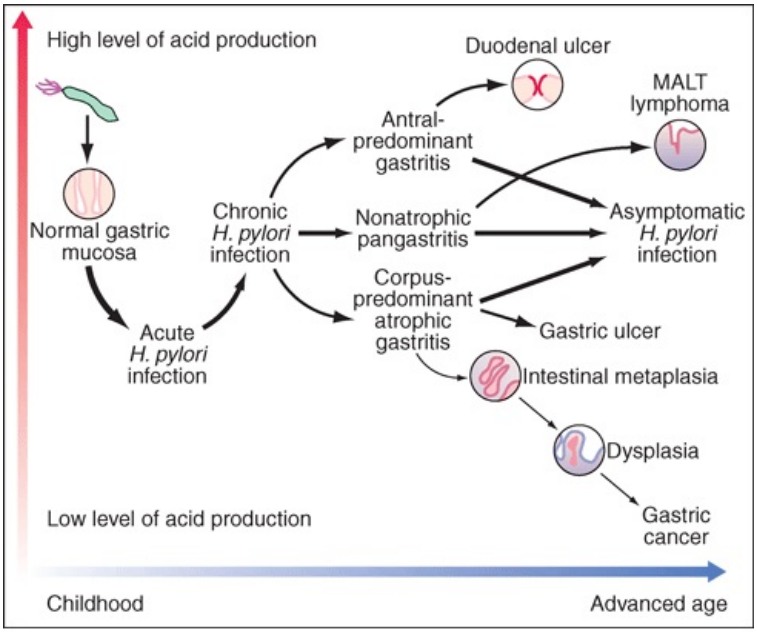

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Natural history of *H. pylori*-infection.*(Used with permission from S Suerbaum, P Michetti: N Engl J Med 347:1175, 2002.)*

NSAID-Induced Disease

Epidemiology

NSAIDs represent a group of the most commonly used medications in the United States. More than 30 billion over-the-counter tablets and over 100 million prescriptions are sold yearly in the United States alone. In fact, after the introduction of COX-2 inhibitors in the year 2000, the number of prescriptions written for NSAIDs was >111 million at a cost of $4.8 billion. Side effects and complications due to NSAIDs are considered the most common drug-related toxicities in the United States. The spectrum of NSAID-induced morbidity ranges from nausea and dyspepsia (prevalence reported as high as 50–60%) to a serious GI complication such as endoscopy-documented peptic ulceration (15–30% of individuals taking NSAIDs regularly) complicated by bleeding or perforation in as many as 1.5% of users per year. It is estimated that NSAID-induced GI bleeding accounts for 60,000 to 120,000 hospital admissions per year, and deaths related to NSAID-induced toxicity may be as high as 16,000 per year in the United States. Approximately 4–5% of patients develop symptomatic ulcers within 1 year. Unfortunately, dyspeptic symptoms do not correlate with NSAID-induced pathology. Over 80% of patients with serious NSAID-related complications did not have preceding dyspepsia. In view of the lack of warning signs, it is important to identify patients who are at increased risk for morbidity and mortality related to NSAID usage. Even 75 mg/d of aspirin may lead to serious GI ulceration; thus, no dose of NSAID is completely safe. Established risk factors include advanced age, history of ulcer, concomitant use of glucocorticoids, high-dose NSAIDs, multiple NSAIDs, concomitant use of anticoagulants, clopidogrel, and serious or multisystem disease. Possible risk factors include concomitant infection with *H. pylori,* cigarette smoking, and alcohol consumption.

Pathophysiology

Prostaglandins play a critical role in maintaining gastroduodenal mucosal integrity and repair. It therefore follows that interruption of prostaglandin synthesis can impair mucosal defense and repair, thus facilitating mucosal injury via a systemic mechanism. Animal studies have demonstrated that neutrophil adherence to the gastric microcirculation plays an essential role in the initiation of NSAID-induced mucosal injury. A summary of the pathogenetic pathways by which systemically administered NSAIDs may lead to mucosal injury is shown in **Fig. 293-9**.

Figure 293-9

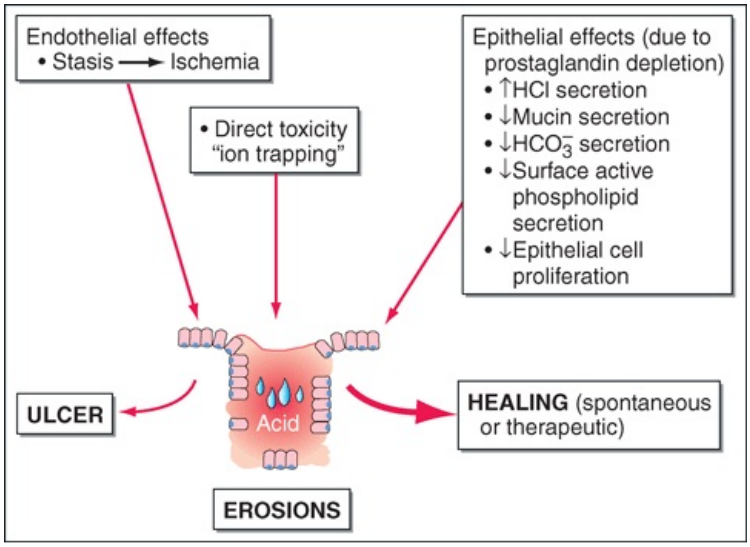

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanisms by which NSAIDs may induce mucosal injury.*(Adapted from J Scheiman et al: J Clin Outcomes Management 3:23, 1996. Copyright 2003 Turner White Communications, Inc., www.turner-white.com. Used with permission.)*

Injury to the mucosa also occurs as a result of the topical encounter with NSAIDs. Aspirin and many NSAIDs are weak acids that remain in a nonionized lipophilic form when found within the acid environment of the stomach. Under these conditions, NSAIDs migrate across lipid membranes of epithelial cells, leading to cell injury once trapped intracellularly in an ionized form. Topical NSAIDs can also alter the surface mucous layer, permitting back diffusion of H^+ and pepsin, leading to further epithelial cell damage. Moreover, enteric-coated or buffered preparations are also associated with risk of peptic ulceration.

The interplay between *H. pylori* and NSAIDs in the pathogenesis of PUD is complex. Meta-analysis supports the conclusion that each of these aggressive factors is independent and synergistic risk factors for PUD and its complications such as GI bleeding. For example, eradication of *H. pylori* reduces the likelihood of GI complications in high-risk individuals to levels observed in individuals with average risk of NSAID-induced complications.

Pathogenetic Factors Unrelated to *H. Pylori* and NSAID in Acid Peptic Disease

Cigarette smoking has been implicated in the pathogenesis of PUD. Not only have smokers been found to have ulcers more frequently than do nonsmokers, but smoking appears to decrease healing rates, impair response to therapy, and increase ulcer-related complications such as perforation. The mechanism responsible for increased ulcer diathesis in smokers is unknown. Theories have included altered gastric emptying, decreased proximal duodenal bicarbonate production, increased risk for *H. pylori* infection, and cigarette-induced generation of noxious mucosal free radicals. Genetic predisposition may play a role in ulcer development. First-degree relatives of DU patients are three times as likely to develop an ulcer; however, the potential role of *H. pylori* infection in contacts is a major consideration. Increased frequency of blood group O and of the nonsecretor status have also been implicated as genetic risk factors for peptic diathesis. However, *H. pylori* preferentially binds to group O antigens. Psychological stress has been thought to contribute to PUD, but studies examining the role of psychological factors in its pathogenesis have generated conflicting results. Although PUD is associated with certain personality traits (neuroticism), these same traits are also present in individuals with nonulcer dyspepsia (NUD) and other functional and organic disorders.

Diet has also been thought to play a role in peptic diseases. Certain foods and beverages can cause dyspepsia, but no convincing studies indicate an association between ulcer formation and a specific diet. Specific chronic disorders have been shown to have a strong association with PUD: (1) systemic mastocytosis, (2) chronic pulmonary disease, (3) chronic renal failure, (4) cirrhosis, (5) nephrolithiasis, and (6) α_1-antitrypsin deficiency. Those with a possible association are (1) hyperparathyroidism, (2) coronary artery disease, (3) polycythemia vera, and (4) chronic pancreatitis.

Multiple factors play a role in the pathogenesis of PUD. The two predominant causes are *H. pylori* infection and NSAID ingestion. PUD not related to *H. pylori* or NSAIDs is increasing. Other less common causes of PUD are shown in **Table 293-1**. These etiologic agents should be considered as the incidence of *H. pylori* is decreasing. Independent of the inciting or injurious agent, peptic ulcers develop as a result of an imbalance between mucosal protection/repair and aggressive factors. Gastric acid plays an essential role in mucosal injury.

Table 293-1 Causes of Ulcers Not Caused by *Helicobacter Pylori* and NSAIDs
Pathogenesis of Non-Hp and Non-NSAID Ulcer Disease
Infection
Cytomegalovirus
Herpes simplex virus
H. heilmannii
Drug/Toxin
Bisphosphonates
Chemotherapy
Clopidogrel
Crack cocaine
Glucocorticoids (when combined with NSAIDs)
Mycophenolate mofetil
Potassium chloride
Miscellaneous
Basophilia in myeloproliferative disease
Duodenal obstruction (e.g., annular pancreas)
Infiltrating disease
Ischemia
Radiation therapy
Sarcoidosis
Crohn's disease
Idiopathic hypersecretory state

Abbreviations: Hp, *H. pylori;* NSAIDs, nonsteroidal anti-inflammatory drugs.

Clinical Features

HISTORY

Abdominal pain is common to many GI disorders, including DU and GU, but has a poor predictive value for the presence of either DU or GU. Up to 10% of patients with NSAID-induced mucosal disease can present with a complication (bleeding, perforation, and obstruction) without antecedent symptoms. Despite this poor correlation, a careful history and physical examination are essential components of the approach to a patient suspected of having peptic ulcers.

Epigastric pain described as a burning or gnawing discomfort can be present in both DU and GU. The discomfort is also described as an ill-defined, aching sensation or as hunger pain. The typical pain pattern in DU occurs 90 minutes to 3 hours after a meal and is frequently relieved by antacids or food. Pain that awakes the patient from sleep (between midnight and 3 A.M.) is the most discriminating symptom, with two-thirds of DU patients describing this complaint. Unfortunately, this symptom is also present in one-third of patients with NUD. The pain pattern in GU patients may be different from that in DU patients, where discomfort may actually be precipitated by food. Nausea and weight loss occur more commonly in GU patients. Endoscopy detects ulcers in <30% of patients who have dyspepsia.

The mechanism for development of abdominal pain in ulcer patients is unknown. Several possible explanations include acid-induced activation of chemical receptors in the duodenum, enhanced duodenal sensitivity to bile acids and pepsin, or altered gastroduodenal motility.

Variation in the intensity or distribution of the abdominal pain, as well as the onset of associated symptoms such as nausea and/or vomiting, may be indicative of an ulcer complication. Dyspepsia that becomes constant, is no longer relieved by food or antacids, or radiates to the back may indicate a penetrating ulcer (pancreas). Sudden onset of severe, generalized abdominal pain may indicate perforation. Pain worsening with meals, nausea, and vomiting of undigested food suggest gastric outlet obstruction. Tarry stools or coffee-ground emesis indicate bleeding.

PHYSICAL EXAMINATION

Epigastric tenderness is the most frequent finding in patients with GU or DU. Pain may be found to the right of the midline in 20% of patients. Unfortunately, the predictive value of this finding is rather low. Physical examination is critically important for discovering evidence of ulcer complication. Tachycardia and orthostasis suggest dehydration secondary to vomiting or active GI blood loss. A severely tender, board like abdomen suggests a perforation. Presence of a succussion splash indicates retained fluid in the stomach, suggesting gastric outlet obstruction.

PUD-RELATED COMPLICATIONS

Gastrointestinal Bleeding

GI bleeding is the most common complication observed in PUD. It occurs in ~15% of patients and more often in individuals >60 years of age. The mortality rate is as high as 5–10%. The higher incidence in the elderly is likely due to the increased use of NSAIDs in this group. Up to 20% of patients with ulcer-related hemorrhage bleed without any preceding warning signs or symptoms.

Perforation

The second most common ulcer-related complication is perforation, being reported in as many as 6 –7% of PUD patients. As in the case of bleeding, the incidence of perforation in the elderly appears to be increasing secondary to increased use of NSAIDs. *Penetration* is a form of perforation in which the ulcer bed tunnels into an adjacent organ. DUs tend to penetrate posteriorly into the pancreas, leading to pancreatitis, whereas GUs tend to penetrate into the left hepatic lobe. Gastrocolic fistulas associated with GUs have also been described.

Gastric Outlet Obstruction

Gastric outlet obstruction is the least common ulcer-related complication, occurring in 1–2% of patients. A patient may have relative obstruction secondary to ulcer-related inflammation and edema in the peripyloric region. This process often resolves with ulcer healing. A fixed, mechanical obstruction secondary to scar formation in the peripyloric areas is also possible. The latter requires endoscopic (balloon dilation) or surgical intervention. Signs and symptoms relative to mechanical obstruction may develop insidiously. New onset of early satiety, nausea, vomiting, increase of postprandial abdominal pain, and weight loss should make gastric outlet obstruction a possible diagnosis.

DIFFERENTIAL DIAGNOSIS

The list of gastrointestinal and nongastrointestinal disorders that can mimic ulceration of the stomach or duodenum is quite extensive. The most commonly encountered diagnosis among patients seen for upper abdominal discomfort is NUD. NUD, also known as *functional dyspepsia* or *essential dyspepsia,* refers to a group of heterogeneous disorders typified by upper abdominal pain without the presence of an ulcer. Dyspepsia has been reported to occur in up to 30% of the U.S. population. Up to 60% of patients seeking medical care for dyspepsia have a negative diagnostic evaluation. The etiology of NUD is not established, and the potential role of *H. pylori* in NUD remains controversial.

Several additional disease processes that may present with "ulcer-like" symptoms include proximal GI tumors, gastroesophageal reflux, vascular disease, pancreaticobiliary disease (biliary colic, chronic pancreatitis), and gastroduodenal Crohn's disease.

DIAGNOSTIC EVALUATION

In view of the poor predictive value of abdominal pain for the presence of a gastroduodenal ulcer and the multiple disease processes that can mimic this disease, the clinician is often confronted with having to establish the presence of an ulcer. Documentation of an ulcer requires either a radiographic (barium study) or an endoscopic procedure. However, a large percentage of patients with symptoms suggestive of an ulcer have NUD; empirical therapy is appropriate for individuals who are otherwise healthy and <45 years of age, before embarking on a diagnostic evaluation **(Chap. 39)**.

Barium studies of the proximal GI tract are still commonly used as a first test for documenting an ulcer. The sensitivity of older single-contrast barium meals for detecting a DU is as high as 80%, with a double-contrast study providing detection rates as high as 90%. Sensitivity for detection is decreased in small ulcers (<0.5 cm), presence of previous scarring, or in postoperative patients. A DU appears as a well-demarcated crater, most often seen in the bulb **(Fig. 293-10*A*)**. A GU may represent benign or malignant disease. Typically, a benign GU also appears as a discrete crater with radiating mucosal folds originating from the ulcer margin **(Fig. 293-10*B*)**. Ulcers >3 cm in size or those associated with a mass are more often malignant. Unfortunately, up to 8% of GUs that appear to be benign by radiographic appearance are malignant by endoscopy or surgery. Radiographic studies that show a GU must be followed by endoscopy and biopsy.

Figure 293-10

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

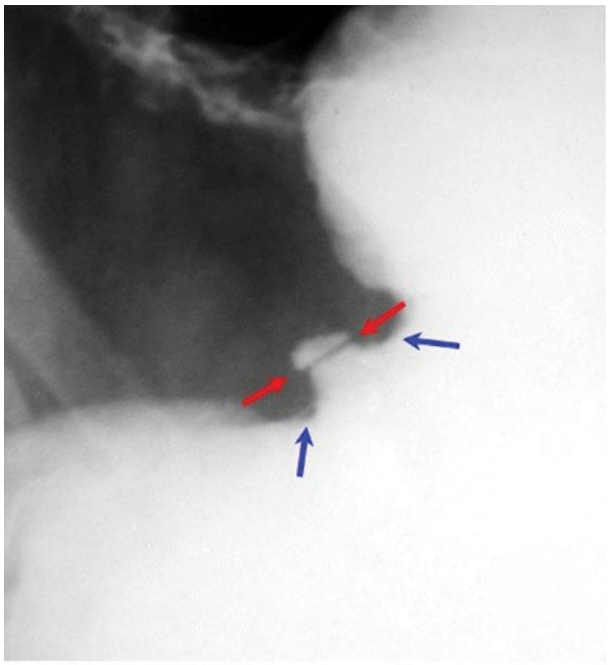

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Barium study demonstrating:*A.* a benign duodenal ulcer; ***B.*** a benign gastric ulcer.

Endoscopy provides the most sensitive and specific approach for examining the upper GI tract **(Fig. 293-11)**. In addition to permitting direct visualization of the mucosa, endoscopy facilitates photographic documentation of a mucosal defect and tissue biopsy to rule out malignancy (GU) or *H. pylori.* Endoscopic examination is particularly helpful in identifying lesions too small to detect by radiographic examination, for evaluation of atypical radiographic abnormalities, or to determine if an ulcer is a source of blood loss.

Figure 293-11

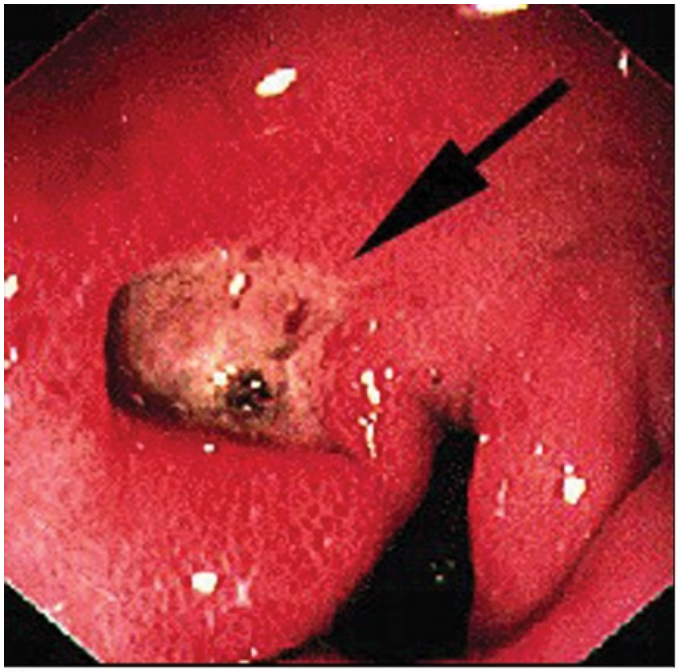

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Endoscopy demonstrating:*A.* a benign duodenal ulcer; ***B.*** a benign gastric ulcer.

Although the methods for diagnosing *H. pylori* are outlined in **Chap. 144**, a brief summary will be included here **(Table 293-2)**. Several biopsy urease tests have been developed (PyloriTek, CLOtest, Hpfast, Pronto Dry) that have a sensitivity and specificity of >90–95%. Several noninvasive methods for detecting this organism have been developed. Three types of studies routinely used include serologic testing, the ^{13}C- or ^{14}C-urea breath test, and the fecal *H. pylori* (Hp) antigen test. A urinary Hp antigen test, as well as a refined monoclonal antibody stool antigen test, appears promising.

Table 293-2 Tests for Detection of *H. Pylori*

Test	Sensitivity/Specificity, %	Comments

Invasive (Endoscopy/Biopsy Required)		
Rapid urease	80–95/95–100	Simple, false negative with recent use of PPIs, antibiotics, or bismuth compounds
Histology	80–90/>95	Requires pathology processing and staining; provides histologic information
Culture	—/—	Time-consuming, expensive, dependent on experience; allows determination of antibiotic susceptibility
Noninvasive		
Serology	>80/>90	Inexpensive, convenient; not useful for early follow-up
Urea breath test	>90/>90	Simple, rapid; useful for early follow-up; false negatives with recent therapy (see rapid urease test); exposure to low-dose radiation with ^{14}C test
Stool antigen	>90/>90	Inexpensive, convenient; not established for eradication but promising

Abbreviation: PPIs, proton pump inhibitors.

Occasionally, specialized testing such as serum gastrin and gastric acid analysis or sham feeding may be needed in individuals with complicated or refractory PUD [see "Zollinger-Ellison Syndrome (ZES)," below]. Screening for aspirin or NSAIDs (blood or urine) may also be necessary in refractory *H. pylori*–negative PUD patients.

Treatment: Peptic Ulcer Disease

Before the discovery of *H. pylori,* the therapy of PUD was centered on the old dictum by Schwartz of "no acid, no ulcer." Although acid secretion is still important in the pathogenesis of PUD, eradication of *H. pylori* and therapy/prevention of NSAID-induced disease is the mainstay of treatment. A summary of commonly used drugs for treatment of acid peptic disorders is shown in **Table 293-3**.

Table 293-3 Drugs Used in the Treatment of Peptic Ulcer Disease

Drug Type/Mechanism	Examples	Dose
Acid-suppressing drugs		
Antacids	Mylanta, Maalox, Tums, Gaviscon	100–140 meq/L 1 and 3 h after meals and hs
H_2 receptor antagonists	Cimetidine	400 mg bid
	Ranitidine	300 mg hs
	Famotidine	40 mg hs
	Nizatidine	300 mg hs
Proton pump inhibitors	Omeprazole	20 mg/d
	Lansoprazole	30 mg/d
	Rabeprazole	20 mg/d
	Pantoprazole	40 mg/d
	Esomeprazole	20 mg/d
Mucosal protective agents		
Sucralfate	Sucralfate	1 g qid
Prostaglandin analogue	Misoprostol	200 μg qid
Bismuth-containing compounds	Bismuth subsalicylate (BSS)	See anti- regimens (Table 293-4)

Abbreviation: hs, at bedtime (*hora somni*).

ACID NEUTRALIZING/INHIBITORY DRUGS

Antacids

Before we understood the important role of histamine in stimulating parietal cell activity, neutralization of secreted acid with antacids constituted the main form of therapy for peptic ulcers. They are now rarely, if ever, used as the primary therapeutic agent but instead are often used by patients for symptomatic relief of dyspepsia. The most commonly used agents are mixtures of aluminum hydroxide and magnesium hydroxide. Aluminum hydroxide can produce constipation and phosphate depletion; magnesium hydroxide may cause loose stools. Many of the commonly used antacids (e.g., Maalox,

Mylanta) have a combination of both aluminum and magnesium hydroxide in order to avoid these side effects. The magnesium-containing preparation should not be used in chronic renal failure patients because of possible hypermagnesemia, and aluminum may cause chronic neurotoxicity in these patients.

Calcium carbonate and sodium bicarbonate are potent antacids with varying levels of potential problems. The long-term use of calcium carbonate (converts to calcium chloride in the stomach) can lead to milk-alkali syndrome (hypercalcemia, hyperphosphatemia with possible renal calcinosis and progression to renal insufficiency). Sodium bicarbonate may induce systemic alkalosis.

H_2 Receptor Antagonists

Four of these agents are presently available (cimetidine, ranitidine, famotidine, and nizatidine), and their structures share homology with histamine. Although each has different potency, all will significantly inhibit basal and stimulated acid secretion to comparable levels when used at therapeutic doses. Moreover, similar ulcer-healing rates are achieved with each drug when used at the correct dosage. Presently, this class of drug is often used for treatment of active ulcers (4–6 weeks) in combination with antibiotics directed at eradicating *H. pylori* (see below).

Cimetidine was the first H_2 receptor antagonist used for the treatment of acid peptic disorders. The initial recommended dosing profile for cimetidine was 300 mg qid. Subsequent studies have documented the efficacy of using 800 mg at bedtime for treatment of active ulcer, with healing rates approaching 80% at 4 weeks. Cimetidine may have weak antiandrogenic side effects resulting in reversible gynecomastia and impotence, primarily in patients receiving high doses for prolonged periods of time (months to years, as in ZES). In view of cimetidine's ability to inhibit cytochrome P450, careful monitoring of drugs such as warfarin, phenytoin, and theophylline is indicated with long-term usage. Other rare reversible adverse effects reported with cimetidine include confusion and elevated levels of serum aminotransferases, creatinine, and serum prolactin. Ranitidine, famotidine, and nizatidine are more potent H_2 receptor antagonists than cimetidine. Each can be used once a day at bedtime for ulcer prevention, which was commonly done before the discovery of *H. pylori* and the development of proton pump inhibitors (PPIs). Patients may develop tolerance to H_2 blockers, a rare event with PPIs (see below). Comparable nighttime dosing regimens are ranitidine 300 mg, famotidine 40 mg, and nizatidine 300 mg.

Additional rare, reversible systemic toxicities reported with H_2 receptor antagonists include pancytopenia, neutropenia, anemia, and thrombocytopenia, with a prevalence rate varying from 0.01–0.2%. Cimetidine and ranitidine (to a lesser extent) can bind to hepatic cytochrome P450; famotidine and nizatidine do not.

Proton Pump (H^+,K^+-ATPase) Inhibitors

Omeprazole, esomeprazole, lansoprazole, rabeprazole, and pantoprazole are substituted benzimidazole derivatives that covalently bind and irreversibly inhibit H^+,K^+-ATPase. Esomeprazole, the newest member of this drug class, is the S-enantiomer of omeprazole, which is a racemic mixture of both S- and R-optical isomers. These are the most potent acid inhibitory agents available. Omeprazole and lansoprazole are the PPIs that have been used for the longest time. Both are acid-labile and are administered as enteric-coated granules in a sustained-release capsule that dissolves within the small intestine at a pH of 6. Lansoprazole is available in an orally disintegrating tablet that can be taken with or without water, an advantage for individuals who have significant dysphagia. Absorption kinetics are similar to the capsule. In addition, a lansoprazole-naproxen combination preparation that has been made available is targeted at decreasing NSAID-related GI injury (see below). Omeprazole is available as nonenteric-coated granules mixed with sodium bicarbonate in a powder form that can be administered orally or via gastric tube. The sodium bicarbonate has two purposes: to protect the omeprazole from acid degradation and to promote rapid gastric alkalinization and subsequent proton pump activation, which facilitates rapid action of the PPI. Pantoprazole and rabeprazole are available as enteric-coated tablets. Pantoprazole is also available as a parenteral formulation for intravenous use. These agents are lipophilic compounds; upon entering the parietal cell, they are protonated and trapped within the acid environment of the tubulovesicular and canalicular system. These agents potently inhibit all phases of gastric acid secretion. Onset of action is rapid, with a maximum acid inhibitory effect between 2 and 6 hours after administration and duration of inhibition lasting up to 72–96 hours. With repeated daily dosing, progressive acid inhibitory effects are observed, with basal and secretagogue-stimulated acid production being inhibited by >95% after 1 week of therapy. The half-life of PPIs is ~18 hours; thus, it can take between 2 and 5 days for gastric acid secretion to return to normal levels once these drugs have been discontinued. Because the pumps need to be activated for these agents to be effective, their efficacy is maximized if they are administered before a meal (except for the immediate-release formulation of omeprazole) (e.g., in the morning before breakfast). Mild to moderate hypergastrinemia has been observed in patients taking these drugs. Carcinoid tumors developed in some animals given the drugs preclinically; however, extensive experience has failed to demonstrate gastric carcinoid tumor development in humans. Serum

gastrin levels return to normal levels within 1–2 weeks after drug cessation. Rebound gastric acid hypersecretion has been described in *H. pylori* negative individuals after discontinuation of PPIs. It occurs even after relatively short-term usage (2 months) and may last for up to 2 months after the PPI has been discontinued. The mechanism involves gastrin-induced hyperplasia and hypertrophy of histamine-secreting ECL cells. The clinical relevance of this observation is that individuals may have worsening symptoms of gastroesophageal reflux disease (GERD) or dyspepsia upon stopping the PPI. Gradual tapering of the PPI and switching to an H_2 receptor antagonist may prevent this from occurring. *H. pylori*-induced inflammation and concomitant decrease in acid production may explain why this does not occur in *H. pylori*-positive patients. IF production is also inhibited, but vitamin B12-deficiency anemia is uncommon, probably because of the large stores of the vitamin. As with any agent that leads to significant hypochlorhydria, PPIs may interfere with absorption of drugs such as ketoconazole, ampicillin, iron, and digoxin. Hepatic cytochrome P450 can be inhibited by the earlier PPIs (omeprazole, lansoprazole). Rabeprazole, pantoprazole, and esomeprazole do not appear to interact significantly with drugs metabolized by the cytochrome P450 system. The overall clinical significance of this observation is not definitely established. Caution should be taken when using theophylline, warfarin, diazepam, atazanavir, and phenytoin concomitantly with PPIs. Long-term acid suppression, especially with PPIs, has been associated with a higher incidence of community-acquired pneumonia as well as community and hospital acquired *Clostridium difficile*-associated disease. These observations require confirmation but should alert the practitioner to take caution when recommending these agents for long-term use, especially in elderly patients at risk for developing pneumonia or *C. difficile* infection. A population-based study revealed that long-term use of PPIs was associated with the development of hip fractures in older women. The absolute risk of fracture remained low despite an observed increase associated with the dose and duration of acid suppression. The mechanism for this observation is not clear and this finding must be confirmed before making broad recommendations regarding the discontinuation of these agents in patients who benefit from them. PPIs may exert a negative effect on the anti-platelet effect of clopidogrel. Although the evidence is mixed and inconclusive, a small increase in mortality and readmission rate for coronary events is seen in patients receiving a PPI while on clopidogrel. The mechanism involves the competition of the PPI and clopidogrel with the same cytochrome p450 (CYP2C19). Whether this is a class effect of PPIs is unclear; there appears to be at least a theoretical advantage of pantoprazole over the other PPIs, but this has not been confirmed. This drug interaction is particularly relevant in light of the common use of aspirin and clopidogrel for prevention of coronary events and the efficacy of PPIs in preventing GI bleeding in these patients. The FDA has made several recommendations while awaiting further evidence to clarify the impact of PPI therapy on clopidogrel use. Health care providers should continue to prescribe clopidogrel to patients who require it and should reevaluate the need for starting or continuing treatment with a PPI. From a practical standpoint additional recommendations to consider include: Patients taking clopidogrel with aspirin, especially with other GI risk factors for bleeding, should receive GI protective therapy. Although high-dose H_2 blockers have been considered an option these do not appear to be as effective as PPIs. If PPIs are to be given, there should be a 12-h separation between administration of the PPI and clopidogrel to minimize competition of the two agents with the involved cytochrome p450. One option is to give the PPI 30 min before breakfast and the clopidogrel at bedtime. Insufficient data are available to firmly recommend one PPI over another.

Two new formulations of acid inhibitory agents are being developed. Tenatoprazole is a PPI containing an imidazopyridine ring instead of a benzimidazole ring, which promotes irreversible proton pump inhibition. This agent has a longer half-life than the other PPIs and may be beneficial for inhibiting nocturnal acid secretion, which has significant relevance in GERD. A second new class of agents is the potassium-competitive acid pump antagonists (P-CABs). These compounds inhibit gastric acid secretion via potassium competitive binding of the H^+,K^+-ATPase.

CYTOPROTECTIVE AGENTS

Sucralfate

Sucralfate is a complex sucrose salt in which the hydroxyl groups have been substituted by aluminum hydroxide and sulfate. This compound is insoluble in water and becomes a viscous paste within the stomach and duodenum, binding primarily to sites of active ulceration. Sucralfate may act by several mechanisms: serving as a physicochemical barrier, promoting a trophic action by binding growth factors such as EGF, enhancing prostaglandin synthesis, stimulating mucus and bicarbonate secretion, and enhancing mucosal defense and repair. Toxicity from this drug is rare, with constipation being most common (2–3%). It should be avoided in patients with chronic renal insufficiency to prevent aluminum-induced neurotoxicity. Hypophosphatemia and gastric bezoar formation have also been reported rarely. Standard dosing of sucralfate is 1 g qid.

Bismuth-Containing Preparations

Sir William Osler considered bismuth-containing compounds the drug of choice for treating PUD. The resurgence in the use of these agents is due to their effect against *H. pylori*. Colloidal bismuth subcitrate (CBS) and bismuth subsalicylate (BSS, Pepto-Bismol) are the most widely used

preparations. The mechanism by which these agents induce ulcer healing is unclear. Adverse effects with short-term usage include black stools, constipation, and darkening of the tongue. Long-term usage with high doses, especially with the avidly absorbed CBS, may lead to neurotoxicity. These compounds are commonly used as one of the agents in an anti-*H. pylori* regimen (see below).

Prostaglandin Analogues

In view of their central role in maintaining mucosal integrity and repair, stable prostaglandin analogues were developed for the treatment of PUD. The mechanism by which this rapidly absorbed drug provides its therapeutic effect is through enhancement of mucosal defense and repair. The most common toxicity noted with this drug is diarrhea (10–30% incidence). Other major toxicities include uterine bleeding and contractions; misoprostol is contraindicated in women who may be pregnant, and women of childbearing age must be made clearly aware of this potential drug toxicity. The standard therapeutic dose is 200 mu;g qid.

Miscellaneous Drugs

A number of drugs including anticholinergic agents and tricyclic antidepressants were used for treating acid peptic disorders but in light of their toxicity and the development of potent antisecretory agents, these are rarely, if ever, used today.

THERAPY OF *H. PYLORI*

Extensive effort has been made in determining who of the many individuals with *H. pylori* infection should be treated. The common conclusion arrived at by multiple consensus conferences around the world is that *H. pylori* should be eradicated in patients with documented PUD. This holds true independent of time of presentation (first episode or not), severity of symptoms, presence of confounding factors such as ingestion of NSAIDs, or whether the ulcer is in remission. Some have advocated treating patients with a history of documented PUD who are found to be *H. pylori*–positive by serology or breath testing. Over one-half of patients with gastric MALT lymphoma experience complete remission of the tumor in response to *H. pylori* eradication. Treating patients with NUD, to prevent gastric cancer or patients with GERD requiring long-term acid suppression, remains controversial. Guidelines from the American College of Gastroenterology suggest eradication of *H. pylori* in patients who have undergone resection of early gastric cancer. The role of *H. pylori* eradication as a means to prevent gastric cancer is still controversial although data suggest a benefit of early eradication of *H. pylori* for prevention of gastric cancer in patients with peptic ulcer disease.

Multiple drugs have been evaluated in the therapy of *H. pylori.* No single agent is effective in eradicating the organism. Combination therapy for 14 days provides the greatest efficacy. A shorter course administration (7–10 days), although attractive, has not proved as successful as the 14-days regimens. The agents used with the greatest frequency include amoxicillin, metronidazole, tetracycline, clarithromycin, and bismuth compounds.

The physician's goal in treating PUD is to provide relief of symptoms (pain or dyspepsia), promote ulcer healing, and ultimately prevent ulcer recurrence and complications. The greatest impact of understanding the role of *H. pylori* in peptic disease has been the ability to prevent recurrence. Documented eradication of *H. pylori* in patients with PUD is associated with a dramatic decrease in ulcer recurrence to <10–20% as compared to 59% in GU patients and 67% in DU patients when the organism is not eliminated. Eradication of the organism may lead to diminished recurrent ulcer bleeding. The impact of its eradication on ulcer perforation is unclear.

Suggested treatment regimens for *H. pylori* are outlined in **Table 293-4**. Choice of a particular regimen will be influenced by several factors, including efficacy, patient tolerance, existing antibiotic resistance, and cost of the drugs. The aim for initial eradication rates should be 85–90%. Dual therapy [PPI plus amoxicillin, PPI plus clarithromycin, ranitidine bismuth citrate (Tritec) plus clarithromycin] are not recommended in view of studies demonstrating eradication rates of <80–85%. The combination of bismuth, metronidazole, and tetracycline was the first triple regimen found effective against *H. pylori.* The combination of two antibiotics plus either a PPI, H_2 blocker, or bismuth compound has comparable success rates. Addition of acid suppression assists in providing early symptom relief and may enhance bacterial eradication.

Table 293-4 Regimens Recommended for Eradication of *H. Pylori* Infection

Drug	Dose
Triple Therapy	
1. Bismuth subsalicylate*plus*	2 tablets qid
Metronidazole*plus*	250 mg qid
Tetracycline[a]	500 mg qid
2. Ranitidine bismuth citrate *plus*	400 mg bid

Tetracycline*plus*	500 mg bid
Clarithromycin or metronidazole	500 mg bid
3. Omeprazole (lansoprazole) *plus*	20 mg bid (30 mg bid)
Clarithromycin*plus*	250 or 500 mg bid
Metronidazole[b] *or*	500 mg bid
Amoxicillin[c]	1 g bid
Quadruple Therapy	
Omeprazole (lansoprazole)	20 mg (30 mg) daily
Bismuth subsalicylate	2 tablets qid
Metronidazole	250 mg qid
Tetracycline	500 mg qid

[a]Alternative: use prepacked Helidac (see text).

[b]Alternative: use prepacked Prevpac (see text).

[c]Use either metronidazole or amoxicillin, not both.

Triple therapy, although effective, has several drawbacks, including the potential for poor patient compliance and drug-induced side effects. Compliance is being addressed by simplifying the regimens so that patients can take the medications twice a day. Simpler (dual therapy) and shorter regimens (7 and 10 days) are not as effective as triple therapy for 14 days. Two anti-*H. pylori* regimens are available in prepackaged formulation: Prevpac (lansoprazole, clarithromycin, and amoxicillin) and Helidac (BSS, tetracycline, and metronidazole). The contents of the Prevpac are to be taken twice per day for 14 days, whereas Helidac constituents are taken four times per day with an antisecretory agent (PPI or H_2 blocker), also for at least 14 days.

Side effects have been reported in up to 20–30% of patients on triple therapy. Bismuth may cause black stools, constipation, or darkening of the tongue. The most feared complication with amoxicillin is pseudomembranous colitis, but this occurs in <1–2% of patients. Amoxicillin can also lead to antibiotic-associated diarrhea, nausea, vomiting, skin rash, and allergic reaction. Tetracycline has been reported to cause rashes and, very rarely, hepatotoxicity and anaphylaxis.

One important concern with treating patients who may not need therapy is the potential for development of antibiotic-resistant strains. The incidence and type of antibiotic-resistant *H. pylori* strains vary worldwide. Strains resistant to metronidazole, clarithromycin, amoxicillin, and tetracycline have been described, with the latter two being uncommon. Antibiotic-resistant strains are the most common cause for treatment failure in compliant patients. Unfortunately, in vitro resistance does not predict outcome in patients. Culture and sensitivity testing of *H. pylori* is not performed routinely. Although resistance to metronidazole has been found in as many as 30% of isolates in North America and 80% in developing countries, triple therapy is effective in eradicating the organism in >50% of patients infected with a resistant strain. Clarithromycin resistance is seen in 13% of individuals in the United States, with resistance to amoxicillin being <1% and resistance to both metronidazole and clarithromycin in the 5% range.

Failure of *H. pylori* eradication with triple therapy in a compliant patient is usually due to infection with a resistant organism. Quadruple therapy (Table 293-4), where clarithromycin is substituted for metronidazole (or vice versa), should be the next step. The combination of pantoprazole, amoxicillin, and rifabutin for 10 days has also been used successfully (86% cure rate) in patients infected with resistant strains. Additional regimens considered for second-line therapy include levofloxacin-based triple therapy (levofloxacin, amoxicillin, PPI) for 10 days and furazolidone-based triple therapy (furazolidone, amoxicillin, PPI) for 14 days. Unfortunately, there is no universally accepted treatment regimen recommended for patients who have failed two courses of antibiotics. If eradication is still not achieved in a compliant patient, then culture and sensitivity of the organism should be considered. Additional factors that may lower eradication rates include the patient's country of origin (higher in Northeast Asia than other parts of Asia or Europe) and cigarette smoking. In addition, meta-analysis suggests that even the most effective regimens (quadruple therapy including PPI, bismuth, tetracycline, and metronidazole and triple therapy including PPI, clarithromycin, and amoxicillin) may have suboptimal eradication rates (<80%), thus demonstrating the need for the development of more efficacious treatments.

In view of the observation that 15–25% of patients treated with first-line therapy may still remain infected with the organism, new approaches to treatment have been explored. One promising approach is sequential therapy. This regimen consists of 5 days of amoxicillin and a PPI, followed by an additional 5 days of PPI plus tinidazole and clarithromycin. Initial studies have demonstrated eradication rates of >90% with good patient tolerance. Confirmation of these findings and applicability

of this approach in the United States are needed.

Reinfection after successful eradication of *H. pylori* is rare in the United States (<1% per year). If recurrent infection occurs within the first 6 months after completing therapy, the most likely explanation is recrudescence as opposed to reinfection.

THERAPY OF NSAID-RELATED GASTRIC OR DUODENAL INJURY

Medical intervention for NSAID-related mucosal injury includes treatment of an active ulcer and primary prevention of future injury. Recommendations for the treatment and primary prevention of NSAID-related mucosal injury are listed in **Table 293-5**. Ideally, the injurious agent should be stopped as the first step in the therapy of an active NSAID-induced ulcer. If that is possible, then treatment with one of the acid inhibitory agents (H_2 blockers, PPIs) is indicated. Cessation of NSAIDs is not always possible because of the patient's severe underlying disease. Only PPIs can heal GUs or DUs, independent of whether NSAIDs are discontinued.

Table 293-5 Recommendations for Treatment of NSAID-Related Mucosal Injury

Clinical Setting	Recommendation
Active ulcer	
NSAID discontinued	H_2 receptor antagonist or PPI
NSAID continued	PPI
Prophylactic therapy	Misoprostol
	PPI
	Selective COX-2 inhibitor
H. pylori infection	Eradication if active ulcer present or there is a past history of peptic ulcer disease

Abbreviations: COX-2, isoenzyme of cyclooxygenase; PPI, proton pump inhibitor.

The approach to primary prevention has included avoiding the agent, using NSAIDs that are theoretically less injurious, and/or the use of concomitant medical therapy to prevent NSAID-induced injury. Several nonselective NSAIDs that are associated with a lower likelihood of GI toxicity include diclofenac, aceclofenac, and ibuprofen, although the beneficial effect may be eliminated if higher dosages of the agents are used. Primary prevention of NSAID-induced ulceration can be accomplished by misoprostol (200 μg qid) or a PPI. High-dose H_2 blockers (famotidine, 40 mg bid) have also shown some promise in preventing endoscopically documented ulcers, although PPIs are superior. The highly selective COX-2 inhibitors, celecoxib and rofecoxib, are 100 times more selective inhibitors of COX-2 than standard NSAIDs, leading to gastric or duodenal mucosal injury that is comparable to placebo; their utilization led to an increase in cardiovascular events and withdrawal from the market. Additional caution was engendered when the CLASS study demonstrated that the advantage of celecoxib in preventing GI complications was offset when low-dose aspirin was used simultaneously. Therefore, gastric protection therapy is required in individuals taking COX-2 inhibitors and aspirin prophylaxis. Finally, much of the work demonstrating the benefit of COX-2 inhibitors and PPIs on GI injury has been performed in individuals of average risk; it is unclear if the same level of benefit will be achieved in high-risk patients. For example, concomitant use of warfarin and a COX-2 inhibitor was associated with rates of GI bleeding similar to those observed in patients taking nonselective NSAIDs. A combination of factors, including withdrawal of the majority of COX-2 inhibitors from the market, the observation that low-dose aspirin appears to diminish the beneficial effect of COX-2 selective inhibitors, and the growing use of aspirin for prophylaxis of cardiovascular events, have significantly altered the approach to gastric protective therapy during the use of NSAIDs. A set of guidelines for the approach to the use of NSAIDs was published by the American College of Gastroenterology and is shown in **Table 293-6**. Individuals who are not at risk for cardiovascular events, do not use aspirin, and are without risk for GI complications can receive nonselective NSAIDs without gastric protection. In those without cardiovascular risk factors but with a high potential risk (prior GI bleeding or multiple GI risk factors) for NSAID-induced GI toxicity, cautious use of a selective COX-2 inhibitor and co-therapy with misoprostol or high-dose PPI is recommended. Individuals at moderate GI risk without cardiac risk factors can be treated with a COX-2 inhibitor alone or with a nonselective NSAID with misoprostol or a PPI. Individuals with cardiovascular risk factors, who require low-dose aspirin and have low potential for NSAID-induced toxicity, should be considered for a non-NSAID agent or use of a traditional NSAID in combination with gastric protection, if warranted. Finally, individuals with cardiovascular and GI risks, who require aspirin must be considered for non-NSAID therapy, but if that is not an option, then gastric protection with any type of NSAID must be considered. Any patient, regardless of risk status, who is being considered for long-term traditional NSAID therapy, should also be considered for *H. pylori* testing and treatment if positive.

Table 293-6 Guide to NSAID Therapy

	No/Low NSAID GI Risk	NSAID GI Risk
No CV risk (no aspirin)	Traditional NSAID	Coxib *or*
		Traditional NSAID + PPI *or* misoprostol
		Consider non-NSAID therapy
CV risk (consider aspirin)	Traditional NSAID + PPI *or* misoprostol if GI risk warrants gastroprotection	A gastroprotective agent must be added if a traditional NSAID is prescribed
	Consider non-NSAID therapy	Consider non-NSAID therapy

Abbreviations: CV, cardiovascular; GI, gastrointestinal; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor.

Source: Adapted from AM Fendrick: Am J Manag Care 10:740, 2004. Reproduced with permission of INTELLISPHERE, LLC via Copyright Clearance Center.

APPROACH AND THERAPY: SUMMARY

Controversy continues regarding the best approach to the patient who presents with dyspepsia **(Chap. 39)**. The discovery of *H. pylori* and its role in pathogenesis of ulcers has added a new variable to the equation. Previously, if a patient <50 years of age presented with dyspepsia and without alarming signs or symptoms suggestive of an ulcer complication or malignancy, an empirical therapeutic trial with acid suppression was commonly recommended. Although this approach is practiced by some today, an approach presently gaining approval for the treatment of patients with dyspepsia is outlined in **Fig. 293-12**. The referral to a gastroenterologist is for the potential need of endoscopy and subsequent evaluation and treatment if the endoscopy is negative.

Figure 293-12

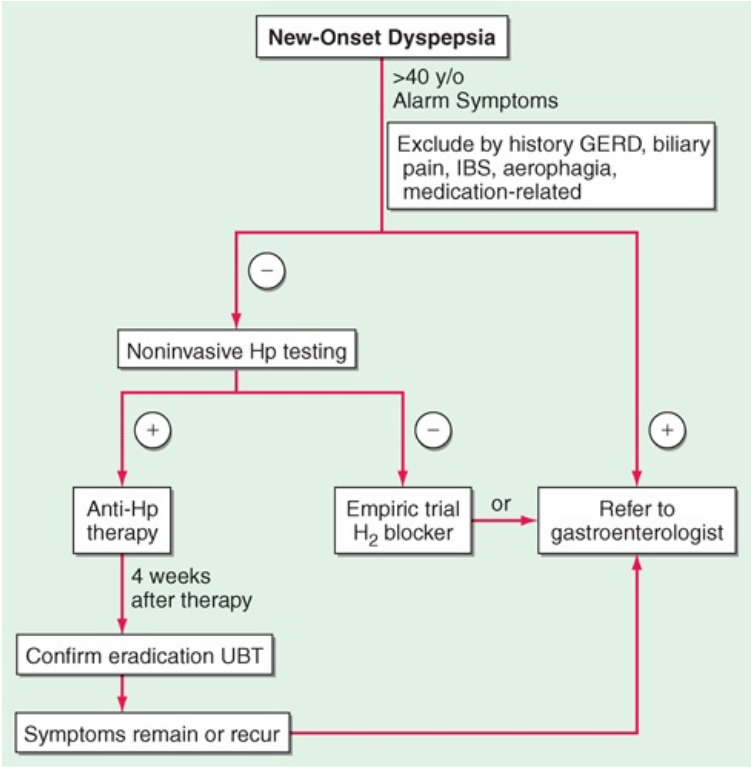

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Overview of new-onset dyspepsia. Hp, *Helicobacter pylori;* UBT, urea breath test; IBS, irritable bowel syndrome. *(Adapted from BS Anand and DY Graham: Endoscopy 31:215, 1999.)*

Once an ulcer (GU or DU) is documented, the main issue at stake is whether *H. pylori* or an NSAID is involved. With *H. pylori* present, independent of the NSAID status, triple therapy is recommended for 14 days, followed by continued acid-suppressing drugs (H_2 receptor antagonist or PPIs) for a total of 4–6 weeks. Selection of patients for documentation of *H. pylori* eradication (organisms gone at least 4

weeks after completing antibiotics) is an area of some debate. The test of choice for documenting eradication is the urea breath test (UBT). The stool antigen assay may also hold promise for this purpose, but the data have not been as clear cut as in the case of using the stool antigen test for primary diagnosis, especially if one considers patients who live in areas of low *H. pylori* prevalence. Further studies are warranted, but if the UBT is not available, a stool antigen should be considered to document eradication. The patient must be off antisecretory agents when being tested for eradication of *H. pylori* with UBT or stool antigen. Serologic testing is not useful for the purpose of documenting eradication since antibody titers fall slowly and often do not become undetectable. Two approaches toward documentation of eradication exist: (1) Test for eradication only in individuals with a complicated course or in individuals who are frail or with multisystem disease who would do poorly with an ulcer recurrence, and (2) test all patients for successful eradication. Some recommend that patients with complicated ulcer disease, or who are frail, should be treated with long-term acid suppression, thus making documentation of *H. pylori* eradication a moot point. In view of this discrepancy in practice, it would be best to discuss with the patient the different options available.

Several issues differentiate the approach to a GU versus a DU. GUs, especially of the body and fundus, have the potential of being malignant. Multiple biopsies of a GU should be taken initially; even if these are negative for neoplasm, repeat endoscopy to document healing at 8–12 weeks should be performed, with biopsy if the ulcer is still present. About 70% of GUs eventually found to be malignant undergo significant (usually incomplete) healing.

The majority (>90%) of GUs and DUs heal with the conventional therapy outlined above. A GU that fails to heal after 12 weeks and a DU that does not heal after 8 weeks of therapy should be considered refractory. Once poor compliance and persistent *H. pylori* infection have been excluded, NSAID use, either inadvertent or surreptitious, must be excluded. In addition, cigarette smoking must be eliminated. For a GU, malignancy must be meticulously excluded. Next, consideration should be given to a gastric acid hypersecretory state such as ZES (see "Zollinger-Ellison Syndrome," below) or the idiopathic form, which can be excluded with gastric acid analysis. Although a subset of patients have gastric acid hypersecretion of unclear etiology as a contributing factor to refractory ulcers, ZES should be excluded with a fasting gastrin or secretin stimulation test (see below). More than 90% of refractory ulcers (either DUs or GUs) heal after 8 weeks of treatment with higher doses of PPI (omeprazole, 40 mg/d; lansoprazole 30–60 mg/d). This higher dose is also effective in maintaining remission. Surgical intervention may be a consideration at this point; however, other rare causes of refractory ulcers must be excluded before recommending surgery. Rare etiologies of refractory ulcers that may be diagnosed by gastric or duodenal biopsies include ischemia, Crohn's disease, amyloidosis, sarcoidosis, lymphoma, eosinophilic gastroenteritis, or infection [cytomegalovirus (CMV), tuberculosis, or syphilis].

SURGICAL THERAPY

Surgical intervention in PUD can be viewed as being either elective, for treatment of medically refractory disease, or as urgent/emergent, for the treatment of an ulcer-related complication. The development of pharmacologic and endoscopic approaches for the treatment of peptic disease and its complications has led to a substantial decrease in the number of operations needed for this disorder. Refractory ulcers are an exceedingly rare occurrence. Surgery is more often required for treatment of an ulcer-related complication.

Hemorrhage is the most common ulcer-related complication, occurring in ~15–25% of patients. Bleeding may occur in any age group but is most often seen in older patients (sixth decade or beyond). The majority of patients stop bleeding spontaneously, but endoscopic therapy **(Chap. 291)** is necessary in some. Parenterally and orally administered PPIs also decrease ulcer rebleeding in patients who have undergone endoscopic therapy. Patients unresponsive or refractory to endoscopic intervention will require surgery (~5% of transfusion-requiring patients).

Free peritoneal perforation occurs in ~2–3% of DU patients. As in the case of bleeding, up to 10% of these patients will not have antecedent ulcer symptoms. Concomitant bleeding may occur in up to 10% of patients with perforation, with mortality being increased substantially. Peptic ulcer can also penetrate into adjacent organs, especially with a posterior DU, which can penetrate into the pancreas, colon, liver, or biliary tree.

Pyloric channel ulcers or DUs can lead to gastric outlet obstruction in ~2–3% of patients. This can result from chronic scarring or from impaired motility due to inflammation and/or edema with pylorospasm. Patients may present with early satiety, nausea, vomiting of undigested food, and weight loss. Conservative management with nasogastric suction, intravenous hydration/nutrition, and antisecretory agents is indicated for 7–10 days with the hope that a functional obstruction will reverse. If a mechanical obstruction persists, endoscopic intervention with balloon dilation may be effective. Surgery should be considered if all else fails.

SPECIFIC OPERATIONS FOR DUODENAL ULCERS

Surgical treatment is designed to decrease gastric acid secretion. Operations most commonly performed include (1) vagotomy and drainage (by pyloroplasty, gastroduodenostomy, or gastrojejunostomy), (2) highly selective vagotomy (which does not require a drainage procedure), and

(3) vagotomy with antrectomy. The specific procedure performed is dictated by the underlying circumstances: elective vs. emergency, the degree and extent of duodenal ulceration, and the expertise of the surgeon. Moreover, the trend has been toward minimally invasive and anatomy-preserving operations.

Vagotomy is a component of each of these procedures and is aimed at decreasing acid secretion through ablating cholinergic input to the stomach. Unfortunately, both truncal and selective vagotomy (preserves the celiac and hepatic branches) result in gastric atony despite successful reduction of both basal acid output (BAO, decreased by 85%) and maximal acid output (MAO, decreased by 50%). Drainage through pyloroplasty or gastroduodenostomy is required in an effort to compensate for the vagotomy-induced gastric motility disorder. This procedure has an intermediate complication rate and a 10% ulcer recurrence rate. To minimize gastric dysmotility, highly selective vagotomy (also known as parietal cell, super-selective, or proximal vagotomy) was developed. Only the vagal fibers innervating the portion of the stomach that contains parietal cells is transected, thus leaving fibers important for regulating gastric motility intact. Although this procedure leads to an immediate decrease in both BAO and stimulated acid output, acid secretion recovers over time. By the end of the first postoperative year, basal and stimulated acid output are ~30 and 50%, respectively, of preoperative levels. Ulcer recurrence rates are higher with highly selective vagotomy (≥10%), although the overall complication rates are the lowest of the three procedures.

The procedure that provides the lowest rates of ulcer recurrence (1%) but has the highest complication rate is vagotomy (truncal or selective) in combination with antrectomy. Antrectomy is aimed at eliminating an additional stimulant of gastric acid secretion, gastrin. Two principal types of reanastomoses are used after antrectomy: gastroduodenostomy (Billroth I) or gastrojejunostomy (Billroth II) **(Fig. 293-13)**. Although Billroth I is often preferred over II, severe duodenal inflammation or scarring may preclude its performance. Prospective, randomized studies confirm that partial gastrectomy followed by Roux-en-Y reconstruction leads to a significantly better clinical, endoscopic, and histologic outcome than Billroth II reconstruction.

Figure 293-13

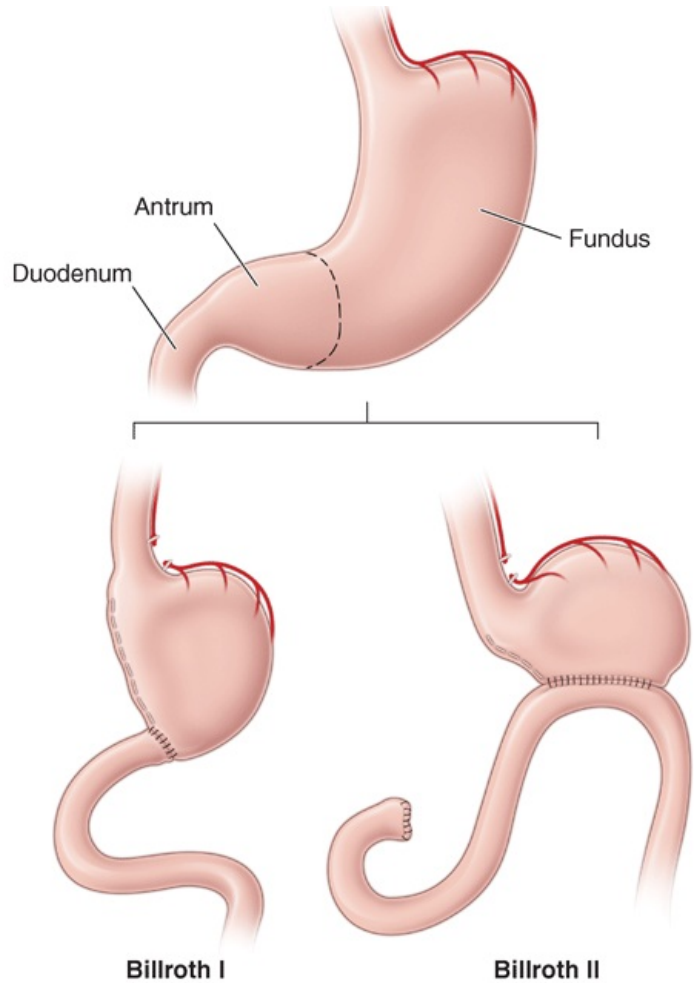

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of Billroth I and II procedures.

Of these procedures, highly selective vagotomy may be the one of choice in the elective setting, except in situations where ulcer recurrence rates are high (prepyloric ulcers and those refractory to medical therapy). Selection of vagotomy and antrectomy may be more appropriate in these circumstances.

These procedures have been traditionally performed by standard laparotomy. The advent of laparoscopic surgery has led several surgical teams to successfully perform highly selective vagotomy, truncal vagotomy/pyloroplasty, and truncal vagotomy/antrectomy through this approach. An increase in the number of laparoscopic procedures for treatment of PUD has occurred. Laparoscopic repair of perforated peptic ulcers is safe, feasible for the experienced surgeon and is associated with decreased postoperative pain although it does take longer than an open approach. Moreover, no difference between the two approaches is noted in postoperative complications or length of hospital stay.

Specific Operations for Gastric Ulcers

The location and the presence of a concomitant DU dictate the operative procedure performed for a GU. Antrectomy (including the ulcer) with a Billroth I anastomosis is the treatment of choice for an antral ulcer. Vagotomy is performed only if a DU is present. Although ulcer excision with vagotomy and drainage procedure has been proposed, the higher incidence of ulcer recurrence makes this a less desirable approach. Ulcers located near the esophagogastric junction may require a more radical approach, a subtotal gastrectomy with a Roux-en-Y esophagogastrojejunostomy (Csende's procedure). A less aggressive approach, including antrectomy, intraoperative ulcer biopsy, and vagotomy (Kelling-Madlener procedure), may be indicated in fragile patients with a high GU. Ulcer recurrence approaches 30% with this procedure.

Surgery-Related Complications

Complications seen after surgery for PUD are related primarily to the extent of the anatomic modification performed. Minimal alteration (highly selective vagotomy) is associated with higher rates of ulcer recurrence and less GI disturbance. More aggressive surgical procedures have a lower rate of ulcer recurrence but a greater incidence of GI dysfunction. Overall, morbidity and mortality related to these procedures are quite low. Morbidity associated with vagotomy and antrectomy or pyloroplasty is ≤5%, with mortality ~1%. Highly selective vagotomy has lower morbidity and mortality rates of 1 and 0.3%, respectively.

In addition to the potential early consequences of any intraabdominal procedure (bleeding, infection, thromboembolism), gastroparesis, duodenal stump leak, and efferent loop obstruction can be observed.

Recurrent Ulceration

The risk of ulcer recurrence is directly related to the procedure performed. Ulcers that recur after partial gastric resection tend to develop at the anastomosis (stomal or marginal ulcer). Epigastric abdominal pain is the most frequent presenting complaint (>90%). Severity and duration of pain tend to be more progressive than observed with DUs before surgery.

Ulcers may recur for several reasons, including incomplete vagotomy, inadequate drainage, retained antrum, and, less likely, persistent or recurrent *H. pylori* infection. ZES should have been excluded preoperatively. Surreptitious use of NSAIDs is an important reason for recurrent ulcers after surgery, especially if the initial procedure was done for an NSAID-induced ulcer. Once *H. pylori* and NSAIDs have been excluded as etiologic factors, the question of incomplete vagotomy or retained gastric antrum should be explored. For the latter, fasting plasma gastrin levels should be determined. If elevated, retained antrum or ZES (see below) should be considered. Incomplete vagotomy can be ruled out by gastric acid analysis coupled with sham feeding. In this test, gastric acid output is measured while the patient sees, smells, and chews a meal (without swallowing). The cephalic phase of gastric secretion, which is mediated by the vagus, is being assessed with this study. An increase in gastric acid output in response to sham feeding is evidence that the vagus nerve is intact. A rise in serum pancreatic polypeptide >50% within 30 min of sham feeding is also suggestive of an intact vagus nerve.

Medical therapy with H_2 blockers will heal postoperative ulceration in 70–90% of patients. The efficacy of PPIs has not been fully assessed in this group, but one may anticipate greater rates of ulcer healing compared to those obtained with H_2 blockers. Repeat operation (complete vagotomy, partial gastrectomy) may be required in a small subgroup of patients who have not responded to aggressive medical management.

Afferent Loop Syndromes

Two types of afferent loop syndrome can occur in patients who have undergone partial gastric resection with Billroth II anastomosis. The more common of the two is bacterial overgrowth in the afferent limb secondary to stasis. Patients may experience postprandial abdominal pain, bloating, and diarrhea with concomitant malabsorption of fats and vitamin B12. Cases refractory to antibiotics may

require surgical revision of the loop. The less common afferent loop syndrome can present with severe abdominal pain and bloating that occur 20–60 minutes after meals. Pain is often followed by nausea and vomiting of bile-containing material. The pain and bloating may improve after emesis. The cause of this clinical picture is theorized to be incomplete drainage of bile and pancreatic secretions from an afferent loop that is partially obstructed. Cases refractory to dietary measures may need surgical revision.

Dumping Syndrome

Dumping syndrome consists of a series of vasomotor and GI signs and symptoms and occurs in patients who have undergone vagotomy and drainage (especially Billroth procedures). Two phases of dumping, early and late, can occur. Early dumping takes place 15–30 minutes after meals and consists of crampy abdominal discomfort, nausea, diarrhea, belching, tachycardia, palpitations, diaphoresis, light-headedness, and, rarely, syncope. These signs and symptoms arise from the rapid emptying of hyperosmolar gastric contents into the small intestine, resulting in a fluid shift into the gut lumen with plasma volume contraction and acute intestinal distention. Release of vasoactive GI hormones (vasoactive intestinal polypeptide, neurotensin, motilin) is also theorized to play a role in early dumping.

The late phase of dumping typically occurs 90 min to 3 h after meals. Vasomotor symptoms (light-headedness, diaphoresis, palpitations, tachycardia, and syncope) predominate during this phase. This component of dumping is thought to be secondary to hypoglycemia from excessive insulin release.

Dumping syndrome is most noticeable after meals rich in simple carbohydrates (especially sucrose) and high osmolarity. Ingestion of large amounts of fluids may also contribute. Up to 50% of postvagotomy and drainage patients will experience dumping syndrome to some degree. Signs and symptoms often improve with time, but a severe protracted picture can occur in up to 1% of patients.

Dietary modification is the cornerstone of therapy for patients with dumping syndrome. Small, multiple (six) meals devoid of simple carbohydrates coupled with elimination of liquids during meals is important. Antidiarrheals and anticholinergic agents are complementary to diet. Guar and pectin, which increase the viscosity of intraluminal contents, may be beneficial in more symptomatic individuals. Acarbose, an α-glucosidase inhibitor that delays digestion of ingested carbohydrates, has also been shown to be beneficial in the treatment of the late phases of dumping. The somatostatin analogue octreotide has been successful in diet-refractory cases. This drug is administered subcutaneously (50 μg tid), titrated according to clinical response. A long-acting depot formulation of octreotide can be administered once every 28 days and provides symptom relief comparable to the short-acting agent. In addition, patient weight gain and quality of life appear to be superior with the long-acting form.

Postvagotomy Diarrhea

Up to 10% of patients may seek medical attention for the treatment of postvagotomy diarrhea. This complication is most commonly observed after truncal vagotomy. Patients may complain of intermittent diarrhea that occurs typically 1–2 hours after meals. Occasionally the symptoms may be severe and relentless. This is due to a motility disorder from interruption of the vagal fibers supplying the luminal gut. Other contributing factors may include decreased absorption of nutrients (see below), increased excretion of bile acids, and release of luminal factors that promote secretion. Diphenoxylate or loperamide is often useful in symptom control. The bile salt–binding agent cholestyramine may be helpful in severe cases. Surgical reversal of a 10-cm segment of jejunum may yield a substantial improvement in bowel frequency in a subset of patients.

Bile Reflux Gastropathy

A subset of postpartial gastrectomy patients who present with abdominal pain, early satiety, nausea, and vomiting will have mucosal erythema of the gastric remnant as the only finding. Histologic examination of the gastric mucosa reveals minimal inflammation but the presence of epithelial cell injury. This clinical picture is categorized as bile or alkaline reflux gastropathy/gastritis. Although reflux of bile is implicated as the reason for this disorder, the mechanism is unknown. Prokinetic agents, cholestyramine, and sucralfate have been somewhat effective treatments. Severe refractory symptoms may require using either nuclear scanning with ^{99m}Tc-HIDA to document reflux or an alkaline challenge test, where 0.1 *N* NaOH is infused into the stomach in an effort to reproduce the patient's symptoms. Surgical diversion of pancreaticobiliary secretions away from the gastric remnant with a Roux-en-Y gastrojejunostomy consisting of a long (50–60 cm) Roux limb has been used in severe cases. Bilious vomiting improves, but early satiety and bloating may persist in up to 50% of patients.

Maldigestion and Malabsorption

Weight loss can be observed in up to 60% of patients after partial gastric resection. A significant component of this weight reduction is due to decreased oral intake. However, mild steatorrhea can also develop. Reasons for maldigestion/malabsorption include decreased gastric acid production, rapid gastric emptying, decreased food dispersion in the stomach, reduced luminal bile concentration, reduced pancreatic secretory response to feeding, and rapid intestinal transit.

Decreased serum vitamin B12 levels can be observed after partial gastrectomy. This is usually not due

to deficiency of IF, since a minimal amount of parietal cells (source of IF) are removed during antrectomy. Reduced vitamin B_{12} may be due to competition for the vitamin by bacterial overgrowth or inability to split the vitamin from its protein-bound source due to hypochlorhydria.

Iron-deficiency anemia may be a consequence of impaired absorption of dietary iron in patients with a Billroth II gastrojejunostomy. Absorption of iron salts is normal in these individuals; thus, a favorable response to oral iron supplementation can be anticipated. Folate deficiency with concomitant anemia can also develop in these patients. This deficiency may be secondary to decreased absorption or diminished oral intake.

Malabsorption of vitamin D and calcium resulting in osteoporosis and osteomalacia is common after partial gastrectomy and gastrojejunostomy (Billroth II). Osteomalacia can occur as a late complication in up to 25% of postpartial gastrectomy patients. Bone fractures occur twice as commonly in men after gastric surgery as in a control population. It may take years before x-ray findings demonstrate diminished bone density. Elevated alkaline phosphatase, reduced serum calcium, bone pain, and pathologic fractures may be seen in patients with osteomalacia. The high incidence of these abnormalities in this subgroup of patients justifies treating them with vitamin D and calcium supplementation indefinitely. Therapy is especially important in females.

Gastric Adenocarcinoma

The incidence of adenocarcinoma in the gastric stump is increased 15 years after resection. Some have reported a four- to fivefold increase in gastric cancer 20–25 years after resection. The pathogenesis is unclear but may involve alkaline reflux, bacterial proliferation, or hypochlorhydria. The role of endoscopic screening is not clear, and most guidelines do not support its use.

RELATED CONDITIONS

Zollinger–Ellison Syndrome

Severe peptic ulcer diathesis secondary to gastric acid hypersecretion due to unregulated gastrin release from a non-β cell endocrine tumor (gastrinoma) defines the components of ZES. Initially, ZES was typified by aggressive and refractory ulceration in which total gastrectomy provided the only chance for enhancing survival. Today it can be cured by surgical resection in up to 30% of patients.

EPIDEMIOLOGY

The incidence of ZES varies from 0.1–1% of individuals presenting with PUD. Males are more commonly affected than females, and the majority of patients are diagnosed between ages 30 and 50. Gastrinomas are classified into sporadic tumors (more common) and those associated with multiple endocrine neoplasia (MEN) type I (see below). The widespread availability and use of PPIs has led to a decreased patient referral for gastrinoma evaluation, delay in diagnosis, and an increase in false-positive diagnoses of ZES.

PATHOPHYSIOLOGY

Hypergastrinemia originating from an autonomous neoplasm is the driving force responsible for the clinical manifestations in ZES. Gastrin stimulates acid secretion through gastrin receptors on parietal cells and by inducing histamine release from ECL cells. Gastrin also has a trophic action on gastric epithelial cells. Long-standing hypergastrinemia leads to markedly increased gastric acid secretion through both parietal cell stimulation and increased parietal cell mass. The increased gastric acid output leads to peptic ulcer diathesis, erosive esophagitis, and diarrhea.

TUMOR DISTRIBUTION

Although early studies suggested that the vast majority of gastrinomas occurred within the pancreas, a significant number of these lesions are extrapancreatic. Over 80% of these tumors are found within the hypothetical gastrinoma triangle (confluence of the cystic and common bile ducts superiorly, junction of the second and third portions of the duodenum inferiorly, and junction of the neck and body of the pancreas medially). Duodenal tumors constitute the most common nonpancreatic lesion; between 50 and 75% of gastrinomas are found here. Duodenal tumors are smaller, slower growing, and less likely to metastasize than pancreatic lesions. Less-common extrapancreatic sites include stomach, bones, ovaries, heart, liver, and lymph nodes. More than 60% of tumors are considered malignant, with up to 30–50% of patients having multiple lesions or metastatic disease at presentation. Histologically, gastrin-producing cells appear well-differentiated, expressing markers typically found in endocrine neoplasms (chromogranin, neuron-specific enolase).

CLINICAL MANIFESTATIONS

Gastric acid hypersecretion is responsible for the signs and symptoms observed in patients with ZES. Peptic ulcer is the most common clinical manifestation, occurring in >90% of gastrinoma patients. Initial presentation and ulcer location (duodenal bulb) may be indistinguishable from common PUD. Clinical situations that should create suspicion of gastrinoma are ulcers in unusual locations (second part of the duodenum and beyond), ulcers refractory to standard medical therapy, ulcer recurrence after acid-reducing surgery, ulcers presenting with frank complications (bleeding, obstruction, and perforation), or

ulcers in the absence of *H. pylori* or NSAID ingestion. Symptoms of esophageal origin are present in up to two-thirds of patients with ZES, with a spectrum ranging from mild esophagitis to frank ulceration with stricture and Barrett's mucosa.

Diarrhea, the next most common clinical manifestation, is found in up to 50% of patients. Although diarrhea often occurs concomitantly with acid peptic disease, it may also occur independent of an ulcer. Etiology of the diarrhea is multifactorial, resulting from marked volume overload to the small bowel, pancreatic enzyme inactivation by acid, and damage of the intestinal epithelial surface by acid. The epithelial damage can lead to a mild degree of maldigestion and malabsorption of nutrients. The diarrhea may also have a secretory component due to the direct stimulatory effect of gastrin on enterocytes or the co-secretion of additional hormones from the tumor such as vasoactive intestinal peptide.

Gastrinomas can develop in the presence of MEN I syndrome **(Chaps. 350, 351)** in ~25% of patients. This autosomal dominant disorder involves primarily three organ sites: the parathyroid glands (80–90%), pancreas (40–80%), and pituitary gland (30–60%). The genetic defect in MEN I is in the long arm of chromosome 11 (11q11-q13). In view of the stimulatory effect of calcium on gastric secretion, the hyperparathyroidism and hypercalcemia seen in MEN I patients may have a direct effect on ulcer disease. Resolution of hypercalcemia by parathyroidectomy reduces gastrin and gastric acid output in gastrinoma patients. An additional distinguishing feature in ZES patients with MEN I is the higher incidence of gastric carcinoid tumor development (as compared to patients with sporadic gastrinomas). Gastrinomas tend to be smaller, multiple, and located in the duodenal wall more often than is seen in patients with sporadic ZES. Establishing the diagnosis of MEN I is critical not only from the standpoint of providing genetic counseling to the patient and his or her family but also to the surgical approach recommended.

DIAGNOSIS

The first step in the evaluation of a patient suspected of having ZES is to obtain a fasting gastrin level. A list of clinical scenarios that should arouse suspicion regarding this diagnosis is shown in **Table 293-7**. Fasting gastrin levels are usually <150 pg/mL. Virtually all gastrinoma patients will have a gastrin level >150–200 pg/mL. Measurement of fasting gastrin should be repeated to confirm the clinical suspicion.

Table 293-7 When to Obtain a Fasting Serum Gastrin Level

Multiple ulcers
Ulcers in unusual locations; associated with severe esophagitis; resistant to therapy with frequent recurrences; in the absence of NSAID ingestion or *H. pylori* infection
Ulcer patients awaiting surgery
Extensive family history for peptic ulcer disease
Postoperative ulcer recurrence
Basal hyperchlorhydria
Unexplained diarrhea or steatorrhea
Hypercalcemia
Family history of pancreatic islet, pituitary, or parathyroid tumor
Prominent gastric or duodenal folds

Multiple processes can lead to an elevated fasting gastrin level: gastric hypochlorhydria or achlorhydria (the most frequent), with or without pernicious anemia; retained gastric antrum; G cell hyperplasia; gastric outlet obstruction; renal insufficiency; massive small-bowel obstruction; and conditions such as rheumatoid arthritis, vitiligo, diabetes mellitus, and pheochromocytoma. Gastric acid induces feedback inhibition of gastrin release. A decrease in acid production will subsequently lead to failure of the feedback inhibitory pathway, resulting in net hypergastrinemia. Gastrin levels will thus be high in patients using antisecretory agents for the treatment of acid peptic disorders and dyspepsia. *H. pylori* infection can also cause hypergastrinemia. Although a fasting gastrin >10 times normal is highly suggestive of ZES, two-thirds of patients will have fasting gastrin levels that overlap with levels found in the more common disorders outlined above.

The next step in establishing a biochemical diagnosis of gastrinoma is to assess acid secretion. Nothing further needs to be done if decreased acid output is observed. In contrast, normal or elevated gastric acid output suggests a need for additional tests. Up to 12% of patients with common PUD may have comparable levels of acid secretion. A BAO/MAO ratio >0.6 is highly suggestive of ZES, but a ratio <0.6 does not exclude the diagnosis. Pentagastrin is no longer available in the United States, making measurement of MAO virtually impossible. An endoscopic method for measuring gastric acid output has been developed but requires further validation. If the technology for measuring gastric acid secretion is not available, a basal gastric pH ≥3 virtually excludes a gastrinoma.

Gastrin provocative tests have been developed in an effort to differentiate between the causes of hypergastrinemia and are especially helpful in patients with indeterminate acid secretory studies. The tests are the secretin stimulation test and the calcium infusion study. The most sensitive and specific gastrin provocative test for the diagnosis of gastrinoma is the secretin study. An increase in gastrin of ≥ 120 pg within 15 minutes of secretin injection has a sensitivity and specificity of >90% for ZES. PPI induced hypochlorhydria or achlorhydria may lead to a false-positive secretin test, thus this agent must be stopped for 1 week before testing.

The calcium infusion study is less sensitive and specific than the secretin test, which coupled with it being a more cumbersome study with greater potential for adverse effects, relegates it to rare utilization in the cases where the patient's clinical characteristics are highly suggestive of ZES, but the secretin stimulation is inconclusive.

TUMOR LOCALIZATION

Once the biochemical diagnosis of gastrinoma has been confirmed, the tumor must be located. Multiple imaging studies have been utilized in an effort to enhance tumor localization **(Table 293-8)**. The broad range of sensitivity is due to the variable success rates achieved by the different investigative groups. Endoscopic ultrasound (EUS) permits imaging of the pancreas with a high degree of resolution (<5 mm). This modality is particularly helpful in excluding small neoplasms within the pancreas and in assessing the presence of surrounding lymph nodes and vascular involvement, but it is not very sensitive for finding duodenal lesions. Several types of endocrine tumors express cell-surface receptors for somatostatin. This permits the localization of gastrinomas by measuring the uptake of the stable somatostatin analogue ^{111}In-pentreotide (OctreoScan) with sensitivity and specificity rates of >85%.

Table 293-8 Sensitivity of Imaging Studies in Zollinger-Ellison Syndrome

	Sensitivity, %	
Study	**Primary Gastrinoma**	**Metastatic Gastrinoma**
Ultrasound	21–28	14
CT scan	55–70	>85
Selective angiography	35–68	33–86
Portal venous sampling	70–90	N/A
SASI	55–78	41
MRI	55–70	>85
OctreoScan	67–86	80–100
EUS	80–100	N/A

Note: CT, computed tomography; EUS, endoscopic ultrasonography; MRI, magnetic resonance imaging; OctreoScan, imaging with ^{111}In-pentreotide; SASI, selective arterial secretin injection.

Up to 50% of patients have metastatic disease at diagnosis. Success in controlling gastric acid hypersecretion has shifted the emphasis of therapy toward providing a surgical cure. Detecting the primary tumor and excluding metastatic disease are critical in view of this paradigm shift. Once a biochemical diagnosis has been confirmed, the patient should first undergo an abdominal CT scan, MRI, or OctreoScan (depending on availability) to exclude metastatic disease. Once metastatic disease has been excluded, an experienced endocrine surgeon may opt for exploratory laparotomy with intraoperative ultrasound or transillumination. In other centers, careful examination of the peripancreatic area with EUS, accompanied by endoscopic exploration of the duodenum for primary tumors, will be performed before surgery. Selective arterial secretin injection may be a useful adjuvant for localizing tumors in a subset of patients.

TREATMENT: ZOLLINGER-ELLISON SYNDROME

Treatment of functional endocrine tumors is directed at ameliorating the signs and symptoms related to hormone overproduction, curative resection of the neoplasm, and attempts to control tumor growth in metastatic disease.

PPIs are the treatment of choice and have decreased the need for total gastrectomy. Initial PPI doses tend to be higher than those used for treatment of GERD or PUD. The initial dose of omeprazole, lansoprazole, rabeprazole or esomeprazole should be in the range of 60 mg in divided doses in a 24-hour period. Dosing can be adjusted to achieve a BAO <10 meq/h (at the drug trough) in surgery-naive patients and to <5 meq/h in individuals who have previously undergone an acid-reducing operation. Although the somatostatin analogue has inhibitory effects on gastrin release from receptor-bearing tumors and inhibits gastric acid secretion to some extent, PPIs have the advantage of reducing parietal cell activity to a greater degree. Despite this, octreotide may be considered as adjunctive therapy to the PPI in patients with tumors that express somatostatin receptors and have peptic symptoms that are

difficult to control with high-dose PPI.

The ultimate goal of surgery would be to provide a definitive cure. Improved understanding of tumor distribution has led to immediate cure rates as high as 60% with 10-year disease-free intervals as high as 34% in sporadic gastrinoma patients undergoing surgery. A positive outcome is highly dependent on the experience of the surgical team treating these rare tumors. Surgical therapy of gastrinoma patients with MEN I remains controversial because of the difficulty in rendering these patients disease-free with surgery. In contrast to the encouraging postoperative results observed in patients with sporadic disease, only 6% of MEN I patients are disease free 5 years after an operation. Some groups suggest surgery only if a clearly identifiable, nonmetastatic lesion is documented by structural studies. Others advocate a more aggressive approach, where all patients free of hepatic metastasis are explored and all detected tumors in the duodenum are resected; this is followed by enucleation of lesions in the pancreatic head, with a distal pancreatectomy to follow. The outcome of the two approaches has not been clearly defined. Laparoscopic surgical interventions may provide attractive approaches in the future.

Therapy of metastatic endocrine tumors in general remains suboptimal; gastrinomas are no exception. In light of the observation that in many instances tumor growth is indolent and that many individuals with metastatic disease remain relatively stable for significant periods of time, many advocate not instituting systemic tumor targeted therapy until evidence of tumor progression or refractory symptoms not controlled with PPIs are noted. Medical approaches including biological therapy (IFN-α, long acting somatostatin analogues, peptide receptor radionuclides), systemic chemotherapy (streptozotocin, 5-fluorouracil, and doxorubicin), and hepatic artery embolization may lead to significant toxicity without a substantial improvement in overall survival. ^{111}In-pentetreotide has been used in the therapy of metastatic neuroendocrine tumors; further studies are needed. Several novel therapies are being explored, including radiofrequency or cryoablation of liver lesions and use of agents that block the vascular endothelial growth receptor pathway (bevacizumab, sunitinib) or the mammalian target of rapamycin **(Chap. 350)**.

Surgical approaches including debulking surgery and liver transplantation for hepatic metastasis have also produced limited benefit.

The overall 5- and 10-year survival rates for gastrinoma patients are 62–75% and 47–53%, respectively. Individuals with the entire tumor resected or those with a negative laparotomy have 5- and 10-year survival rates >90%. Patients with incompletely resected tumors have 5- and 10-year survival rates of 43% and 25%, respectively. Patients with hepatic metastasis have <20% survival at 5 years. Favorable prognostic indicators include primary duodenal wall tumors, isolated lymph node tumor, and undetectable tumor upon surgical exploration. Poor outcome is seen in patients with shorter disease duration; higher gastrin levels (>10,000 pg/mL); large pancreatic primary tumors (>3 cm); metastatic disease to lymph nodes, liver, and bone; and Cushing's syndrome. Rapid growth of hepatic metastases is also predictive of poor outcome.

Stress-Related Mucosal Injury

Patients suffering from shock, sepsis, massive burns, severe trauma, or head injury can develop acute erosive gastric mucosal changes or frank ulceration with bleeding. Classified as stress-induced gastritis or ulcers, injury is most commonly observed in the acid-producing (fundus and body) portions of the stomach. The most common presentation is GI bleeding, which is usually minimal but can occasionally be life threatening. Respiratory failure requiring mechanical ventilation and underlying coagulopathy are risk factors for bleeding, which tends to occur 48–72 hours after the acute injury or insult.

Histologically, stress injury does not contain inflammation or *H. pylori;* thus, "gastritis" is a misnomer. Although elevated gastric acid secretion may be noted in patients with stress ulceration after head trauma (Cushing's ulcer) and severe burns (Curling's ulcer), mucosal ischemia and breakdown of the normal protective barriers of the stomach also play an important role in the pathogenesis. Acid must contribute to injury in view of the significant drop in bleeding noted when acid inhibitors are used as prophylaxis for stress gastritis.

Improvement in the general management of intensive care unit patients has led to a significant decrease in the incidence of GI bleeding due to stress ulceration. The estimated decrease in bleeding is from 20–30% to <5%. This improvement has led to some debate regarding the need for prophylactic therapy. The limited benefit of medical (endoscopic, angiographic) and surgical therapy in a patient with hemodynamically compromising bleeding associated with stress ulcer/gastritis supports the use of preventive measures in high-risk patients (mechanically ventilated, coagulopathy, multiorgan failure, or severe burns). Maintenance of gastric pH >3.5 with continuous infusion of H_2 blockers or liquid antacids administered every 2–3 hours are viable options. Tolerance to the H_2 blocker is likely to develop; thus, careful monitoring of the gastric pH and dose adjustment is important if H_2 blockers are used. Sucralfate slurry (1 g every 4–6 hours) has also been somewhat successful but requires a gastric tube and may lead to constipation and aluminum toxicity. Sucralfate use in endotracheal intubated patients has also been associated with aspiration pneumonia. PPIs are the treatment of

choice for stress prophylaxis. Oral PPI is the best option if the patient can tolerate enteral administration. Pantoprazole is available as an intravenous formulation for individuals in whom enteral administration is not possible. If bleeding occurs despite these measures, endoscopy, intraarterial vasopressin, or embolization are options. If all else fails, then surgery should be considered. Although vagotomy and antrectomy may be used, the better approach would be a total gastrectomy, which has an exceedingly high mortality rate in this setting.

Gastritis

The term *gastritis* should be reserved for histologically documented inflammation of the gastric mucosa. Gastritis is not the mucosal erythema seen during endoscopy and is not interchangeable with "dyspepsia." The etiologic factors leading to gastritis are broad and heterogeneous. Gastritis has been classified based on time course (acute versus chronic), histologic features, and anatomic distribution or proposed pathogenic mechanism **(Table 293-9)**.

Table 293-9 Classification of Gastritis

I. Acute gastritis
A. Acute *H. pylori* infection
B. Other acute infectious gastritides
1. Bacterial (other than *H. pylori*)
2. *H. heilmannii*
3. Phlegmonous
4. Mycobacterial
5. Syphilitic
6. Viral
7. Parasitic
8. Fungal
II. Chronic atrophic gastritis
A. Type A: Autoimmune, body-predominant
B. Type B: *H. pylori*–related, antral-predominant
C. Indeterminant
III. Uncommon forms of gastritis
A. Lymphocytic
B. Eosinophilic
C. Crohn's disease
D. Sarcoidosis
E. Isolated granulomatous gastritis

The correlation between the histologic findings of gastritis, the clinical picture of abdominal pain or dyspepsia, and endoscopic findings noted on gross inspection of the gastric mucosa is poor. Therefore, there is no typical clinical manifestation of gastritis.

ACUTE GASTRITIS

The most common causes of acute gastritis are infectious. Acute infection with *H. pylori* induces gastritis. However, *H. pylori* acute gastritis has not been extensively studied. It is reported as presenting with sudden onset of epigastric pain, nausea, and vomiting, and limited mucosal histologic studies demonstrate a marked infiltrate of neutrophils with edema and hyperemia. If not treated, this picture will evolve into one of chronic gastritis. Hypochlorhydria lasting for up to 1 year may follow acute *H. pylori* infection.

Bacterial infection of the stomach or phlegmonous gastritis is a rare, potentially life-threatening disorder characterized by marked and diffuse acute inflammatory infiltrates of the entire gastric wall, at times accompanied by necrosis. Elderly individuals, alcoholics, and AIDS patients may be affected. Potential iatrogenic causes include polypectomy and mucosal injection with India ink. Organisms associated with this entity include streptococci, staphylococci, *Escherichia coli,Proteus,* and *Haemophilus* species Failure of supportive measures and antibiotics may result in gastrectomy.

Other types of infectious gastritis may occur in immunocompromised individuals such as AIDS patients. Examples include herpetic (herpes simplex) or CMV gastritis. The histologic finding of intranuclear inclusions would be observed in the latter.

CHRONIC GASTRITIS

Chronic gastritis is identified histologically by an inflammatory cell infiltrate consisting primarily of lymphocytes and plasma cells, with very scant neutrophil involvement. Distribution of the inflammation may be patchy, initially involving superficial and glandular portions of the gastric mucosa. This picture may progress to more severe glandular destruction, with atrophy and metaplasia. Chronic gastritis has been classified according to histologic characteristics. These include superficial atrophic changes and gastric atrophy.

The early phase of chronic gastritis is *superficial gastritis.* The inflammatory changes are limited to the lamina propria of the surface mucosa, with edema and cellular infiltrates separating intact gastric glands. The next stage is *atrophic gastritis.* The inflammatory infiltrate extends deeper into the mucosa, with progressive distortion and destruction of the glands. The final stage of chronic gastritis is *gastric atrophy.* Glandular structures are lost, and there is a paucity of inflammatory infiltrates. Endoscopically, the mucosa may be substantially thin, permitting clear visualization of the underlying blood vessels.

Gastric glands may undergo morphologic transformation in chronic gastritis. Intestinal metaplasia denotes the conversion of gastric glands to a small intestinal phenotype with small-bowel mucosal glands containing goblet cells. The metaplastic changes may vary in distribution from patchy to fairly extensive gastric involvement. Intestinal metaplasia is an important predisposing factor for gastric cancer **(Chap. 91)**.

Chronic gastritis is also classified according to the predominant site of involvement. Type A refers to the body-predominant form (autoimmune) and type B is the antral-predominant form (*H. pylori*–related). This classification is artificial in view of the difficulty in distinguishing between these two entities. The term *AB gastritis* has been used to refer to a mixed antral/body picture.

Type a Gastritis

The less common of the two forms involves primarily the fundus and body, with antral sparing. Traditionally, this form of gastritis has been associated with pernicious anemia **(Chap. 105)** in the presence of circulating antibodies against parietal cells and IF; thus, it is also called *autoimmune gastritis.H. pylori* infection can lead to a similar distribution of gastritis. The characteristics of an autoimmune picture are not always present.

Antibodies to parietal cells have been detected in >90% of patients with pernicious anemia and in up to 50% of patients with type A gastritis. The parietal cell antibody is directed against H^+,K^+-ATPase. T cells are also implicated in the injury pattern of this form of gastritis. A subset of patients infected with *H. pylori* develop antibodies against H^+,K^+-ATPase, potentially leading to the atrophic gastritis pattern seen in some patients infected with this organism. The mechanism is thought to involve molecular mimicry between *H. pylori* LPS and H^+,K^+-ATPase.

Parietal cell antibodies and atrophic gastritis are observed in family members of patients with pernicious anemia. These antibodies are observed in up to 20% of individuals over age 60 and in ~20% of patients with vitiligo and Addison's disease. About one-half of patients with pernicious anemia have antibodies to thyroid antigens, and about 30% of patients with thyroid disease have circulating antiparietal cell antibodies. Anti-IF antibodies are more specific than parietal cell antibodies for type A gastritis, being present in ~40% of patients with pernicious anemia. Another parameter consistent with this form of gastritis being autoimmune in origin is the higher incidence of specific familial histocompatibility haplotypes such as HLA-B8 and HLA-DR3.

The parietal cell–containing gastric gland is preferentially targeted in this form of gastritis, and achlorhydria results. Parietal cells are the source of IF, the lack of which will lead to vitamin B12 deficiency and its sequelae (megaloblastic anemia, neurologic dysfunction).

Gastric acid plays an important role in feedback inhibition of gastrin release from G cells. Achlorhydria, coupled with relative sparing of the antral mucosa (site of G cells), leads to hypergastrinemia. Gastrin levels can be markedly elevated (>500 pg/mL) in patients with pernicious anemia. ECL cell hyperplasia with frank development of gastric carcinoid tumors may result from gastrin trophic effects. Hypergastrinemia and achlorhydria may also be seen in nonpernicious anemia–associated type A gastritis.

Type B Gastritis

Type B, or antral-predominant, gastritis is the more common form of chronic gastritis. *H. pylori* infection is the cause of this entity. Although described as "antral-predominant," this is likely a misnomer in view of studies documenting the progression of the inflammatory process toward the body and fundus of infected individuals. The conversion to a pangastritis is time-dependent–estimated to require 15–20 years. This form of gastritis increases with age, being present in up to 100% of persons over age 70. Histology improves after *H. pylori* eradication. The number of *H. pylori* organisms decreases dramatically with progression to gastric atrophy, and the degree of inflammation correlates with the level of these organisms. Early on, with antral-predominant findings, the quantity of *H. pylori* is highest and a dense chronic inflammatory infiltrate of the lamina propria is noted, accompanied by

epithelial cell infiltration with polymorphonuclear leukocytes **(Fig. 293-14)**.

Figure 293-14

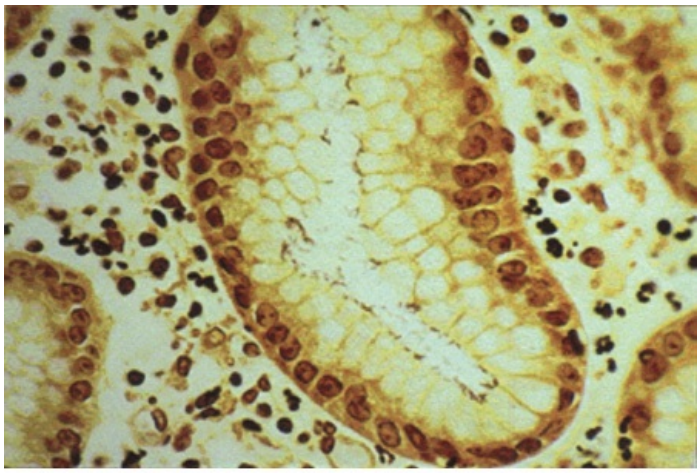

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic gastritis and *H. pylori* organisms. Steiner silver stain of superficial gastric mucosa, showing abundant darkly stained microorganisms layered over the apical portion of the surface epithelium. Note that there is no tissue invasion.

Multifocal atrophic gastritis, gastric atrophy with subsequent metaplasia, has been observed in chronic *H. pylori*–induced gastritis. This may ultimately lead to development of gastric adenocarcinoma (Fig. 293-8; **Chap. 91**). *H. pylori* infection is now considered an independent risk factor for gastric cancer. Worldwide epidemiologic studies have documented a higher incidence of *H. pylori* infection in patients with adenocarcinoma of the stomach as compared to control subjects. Seropositivity for *H. pylori* is associated with a three- to sixfold increased risk of gastric cancer. This risk may be as high as ninefold after adjusting for the inaccuracy of serologic testing in the elderly. The mechanism by which *H. pylori* infection leads to cancer is unknown, but it appears to be related to the chronic inflammation induced by the organism. Eradication of *H. pylori* as a general preventative measure for gastric cancer is being evaluated but is not yet recommended.

Infection with *H. pylori* is also associated with development of a low-grade B cell lymphoma, gastric MALT lymphoma **(Chap. 110)**. The chronic T cell stimulation caused by the infection leads to production of cytokines that promote the B cell tumor. The tumor should be initially staged with a CT scan of the abdomen and EUS. Tumor growth remains dependent on the presence of *H. pylori,* and its eradication is often associated with complete regression of the tumor. The tumor may take more than a year to regress after treating the infection. Such patients should be followed by EUS every 2–3 months. If the tumor is stable or decreasing in size, no other therapy is necessary. If the tumor grows, it may have become a high-grade B cell lymphoma. When the tumor becomes a high-grade aggressive lymphoma histologically, it loses responsiveness to *H. pylori* eradication.

Treatment: Chronic Gastritis

Treatment in chronic gastritis is aimed at the sequelae and not the underlying inflammation. Patients with pernicious anemia will require parenteral vitamin B12 supplementation on a long-term basis. Eradication of *H. pylori* is not routinely recommended unless PUD or a low-grade MALT lymphoma is present.

MISCELLANEOUS FORMS OF GASTRITIS

Lymphocytic gastritis is characterized histologically by intense infiltration of the surface epithelium with lymphocytes. The infiltrative process is primarily in the body of the stomach and consists of mature T cells and plasmacytes. The etiology of this form of chronic gastritis is unknown. It has been described in patients with celiac sprue, but whether there is a common factor associating these two entities is unknown. No specific symptoms suggest lymphocytic gastritis. A subgroup of patients have thickened folds noted on endoscopy. These folds are often capped by small nodules that contain a central depression or erosion; this form of the disease is called *varioliform gastritis.H. pylori* probably plays no significant role in lymphocytic gastritis. Therapy with glucocorticoids or sodium cromoglycate has obtained unclear results.

Marked eosinophilic infiltration involving any layer of the stomach (mucosa, muscularis propria, and serosa) is characteristic of *eosinophilic gastritis.* Affected individuals will often have circulating eosinophilia with clinical manifestation of systemic allergy. Involvement may range from isolated gastric

disease to diffuse eosinophilic gastroenteritis. Antral involvement predominates, with prominent edematous folds being observed on endoscopy. These prominent antral folds can lead to outlet obstruction. Patients can present with epigastric discomfort, nausea, and vomiting. Treatment with glucocorticoids has been successful.

Several systemic disorders may be associated with *granulomatous gastritis.* Gastric involvement has been observed in Crohn's disease. Involvement may range from granulomatous infiltrates noted only on gastric biopsies to frank ulceration and stricture formation. Gastric Crohn's disease usually occurs in the presence of small-intestinal disease. Several rare infectious processes can lead to granulomatous gastritis, including histoplasmosis, candidiasis, syphilis, and tuberculosis. Other unusual causes of this form of gastritis include sarcoidosis, idiopathic granulomatous gastritis, and eosinophilic granulomas involving the stomach. Establishing the specific etiologic agent in this form of gastritis can be difficult, at times requiring repeat endoscopy with biopsy and cytology. Occasionally, a surgically obtained full-thickness biopsy of the stomach may be required to exclude malignancy.

MéNéTrier's Disease

Ménétrier's disease is a rare entity characterized by large, tortuous gastric mucosal folds. The differential diagnosis of large gastric folds includes ZES, malignancy, infectious etiologies (CMV, histoplasmosis, syphilis), and infiltrative disorders such as sarcoidosis. The mucosal folds in Ménétrier's disease are often most prominent in the body and fundus. Histologically, massive foveolar hyperplasia (hyperplasia of surface and glandular mucous cells) is noted, which replaces most of the chief and parietal cells. This hyperplasia produces the prominent folds observed. The pits of the gastric glands elongate and may become extremely tortuous. Although the lamina propria may contain a mild chronic inflammatory infiltrate, Ménétrier's disease is not considered a form of gastritis. The etiology of this unusual clinical picture is unknown. Overexpression of growth factors such as TGF-α may be involved in the process.

Epigastric pain, at times accompanied by nausea, vomiting, anorexia, and weight loss, are signs and symptoms in patients with Ménétrier's disease. Occult GI bleeding may occur, but overt bleeding is unusual and, when present, is due to superficial mucosal erosions. Twenty to 100% of patients (depending on time of presentation) develop a protein-losing gastropathy accompanied by hypoalbuminemia and edema. Gastric acid secretion is usually reduced or absent because of the replacement of parietal cells. Large gastric folds are readily detectable by either radiographic (barium meal) or endoscopic methods. Endoscopy with deep mucosal biopsy (and cytology) is required to establish the diagnosis and exclude other entities that may present similarly. A nondiagnostic biopsy may lead to a surgically obtained full-thickness biopsy to exclude malignancy.

TREATMENT: MÉNÉTRIER'S DISEASE

Medical therapy with anticholinergic agents, prostaglandins, PPIs, prednisone, and H_2 receptor antagonists yields varying results. Anticholinergics decrease protein loss. A high-protein diet should be recommended to replace protein loss in patients with hypoalbuminemia. Ulcers should be treated with a standard approach. Severe disease with persistent and substantial protein loss may require total gastrectomy. Subtotal gastrectomy is performed by some; it may be associated with higher morbidity and mortality secondary to the difficulty in obtaining a patent and long-lasting anastomosis between normal and hyperplastic tissues.

ACKNOWLEDGMENTS

The author acknowledges the contribution of material to this chapter by Dr. Lawrence Friedman and Dr. Walter Peterson from their chapter on this subject in the 14th edition of Harrison's.

FURTHER READINGS

Ali A et al: Long-term safety concerns with proton pump inhibitors. Am J Med 122:896, 2009[PMID: 19786155] [Full Text]

Amieva M, El-Omar E: Host-bacterial interactions in *Helicobacter pylori* infection. Gastroenterology 134:306, 2008[PMID: 18166359] [Full Text]

Atherton J, Blaser M: Coadaptation of *Helicobacter pylori* and humans: Ancient history, modern implications. J Clin Invest 119:2475, 2009[PMID: 19729845] [Full Text]

Cover T, Blaser M: *Helicobacter pylori* in health and disease. Gastroenterology 136:1863, 2009[PMID: 19457415] [Full Text]

Fox JG, Wang TC: Inflammation, atrophy, and gastric cancer. J Clin Invest 117:60, 2007[PMID: 17200707] [Full Text]

Jones R et al: Gastrointestinal and cardiovascular risks of nonsteroidal anti-inflammatory drugs. Am J

Med 121:464, 2008[PMID: 18501223] [Full Text]

Juhász M et al: Current standings of the proton pump inhibitor and clopidogrel co-therapy: Review on an evolving field with the eyes of the gastroenterologist. Digestion 81:10, 2010[PMID: 21429865] [Full Text]

Juurlink D: Proton pump inhibitors and clopidogrel: Putting the interaction in perspective. Circulation 120:2310, 2009[PMID: 19933929] [Full Text]

Lanas A: Nonsteroidal anti-inflammatory drugs and cyclooxygenase inhibition in the gastrointestinal tract: A trip from peptic ulcer to colon cancer. Am J Med Sci 338:96, 2009[PMID: 19680014] [Full Text]

Lanza F et al: Guidelines for prevention of NSAID-related ulcer complications. Am J Gastroenterol 104:728, 2009[PMID: 19240698] [Full Text]

Malfertheiner P et al: Peptic ulcer disease. Lancet 374:1449, 2009[PMID: 19683340] [Full Text]

Metz D, Jensen R: Gastrointestinal neuroendocrine tumors: pancreatic endocrine tumors. Gastroenterology 135:1469, 2008[PMID: 18703061] [Full Text]

Schubert M, Peura D: Control of gastric acid secretion in health and disease. Gastroenterology 134:1842, 2008[PMID: 18474247] [Full Text]

Wu C et al: Early *Helicobacter pylori* eradication decreases risk of gastric cancer in patients with peptic ulcer disease. Gastroenterology 137:1641, 2009[PMID: 19664631] [Full Text]

Zagari R et al: Investigating dyspepsia. BMJ 337:a1400, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF ABSORPTION: INTRODUCTION

Disorders of absorption constitute a broad spectrum of conditions with multiple etiologies and varied clinical manifestations. Almost all of these clinical problems are associated with *diminished* intestinal absorption of one or more dietary nutrients and are often referred to as the *malabsorption syndrome.* This term is not ideal as it represents a pathophysiologic state, does *not* provide an etiologic explanation for the underlying problem, and should not be considered an adequate final diagnosis. The only clinical situations in which absorption is *increased* are hemochromatosis and Wilson's disease, in which absorption of iron and copper, respectively, are increased.

Most, but not all, malabsorption syndromes are associated with *steatorrhea,* an increase in stool fat excretion of >6% of dietary fat intake. Some malabsorption disorders are not associated with steatorrhea: primary lactase deficiency, a congenital absence of the small intestinal brush border disaccharidase enzyme lactase, is associated with lactose "malabsorption," and pernicious anemia is associated with a marked decrease in intestinal absorption of cobalamin (vitamin B12) due to an absence of gastric parietal cell intrinsic factor required for cobalamin absorption.

Disorders of absorption must be included in the differential diagnosis of diarrhea **(Chap. 40)**. First, diarrhea is frequently associated with and/or is a consequence of the diminished absorption of one or more dietary nutrients. The diarrhea may be secondary either to the intestinal process that is responsible for the steatorrhea or to steatorrhea per se. Thus, celiac disease (see below) is associated with both extensive morphologic changes in the small intestinal mucosa and reduced absorption of several dietary nutrients; in contrast, the diarrhea of steatorrhea is the result of the effect of nonabsorbed dietary fatty acids on intestinal, usually colonic, ion transport. For example, oleic acid and ricinoleic acid (a bacterially hydroxylated fatty acid that is also the active ingredient in castor oil, a widely used laxative) induce active colonic Cl ion secretion, most likely secondary to increasing intracellular Ca. In addition, diarrhea per se may result in mild steatorrhea (<11 g fat excretion while on a 100-g fat diet). Second, most patients will indicate that they have diarrhea, not that they have fat malabsorption. Third, many intestinal disorders that have diarrhea as a prominent symptom (e.g., ulcerative colitis, traveler's diarrhea secondary to an enterotoxin produced by *Escherichia coli*) do not necessarily have diminished absorption of any dietary nutrient.

Diarrhea as a *symptom* (i.e., when used by patients to describe their bowel movement pattern) may be a decrease in stool consistency, an increase in stool volume, an increase in number of bowel movements, or any combination of these three changes. In contrast, diarrhea as a *sign* is a quantitative increase in stool water or weight of >200–225 mL or gram per 24 h, when a Western-type diet is consumed. Individuals consuming a diet with higher fiber content may normally have a stool weight of up to 400 g/24 h. Thus, the clinician must clarify what an individual patient means by diarrhea. Some 10% of patients referred to gastroenterologists for further evaluation of unexplained diarrhea do not have an increase in stool water when it is determined quantitatively. Such patients may have small, frequent, somewhat loose bowel movements with stool urgency that is indicative of proctitis but do not have an increase in stool weight or volume.

It is also critical to establish whether a patient's diarrhea is secondary to diminished absorption of one or more dietary nutrients, in contrast to diarrhea that is due to small- and/or large-intestinal fluid and electrolyte secretion. The former has often been termed *osmotic diarrhea,* while the latter has been referred to as *secretory diarrhea.* Unfortunately, both secretory and osmotic elements can be present simultaneously in the same disorder; thus, this separation is not always precise. Nonetheless, two studies—determination of stool electrolytes and observation of the effect of a fast on stool output—can help make this distinction.

The demonstration of the effect of prolonged (>24 h) fasting on stool output can be very effective in suggesting that a *dietary nutrient* is responsible for the individual's diarrhea. A secretory diarrhea associated with enterotoxin-induced traveler's diarrhea would not be affected by prolonged fasting, as enterotoxin-induced stimulation of intestinal fluid and electrolyte secretion is not altered by eating. In contrast, diarrhea secondary to lactose malabsorption in primary lactase deficiency would undoubtedly cease during a prolonged fast. Thus, a substantial decrease in stool output while fasting during a quantitative stool collection of at least 24 h is presumptive evidence that the diarrhea is related to malabsorption of a dietary nutrient. The persistence of stool output while fasting indicates that the diarrhea is likely secretory and that the cause of diarrhea is *not* a dietary nutrient. Either a luminal

(e.g., *E. coli* enterotoxin) or circulating (e.g., vasoactive intestinal peptide) secretagogue could be responsible for the patient's diarrhea persisting unaltered during a prolonged fast. The observed effects of fasting can be compared and correlated with stool electrolyte and osmolality determinations.

Measurement of stool electrolytes and osmolality requires the comparison of stool Na^+ and K^+ concentrations determined in liquid stool to the stool osmolality to determine the presence or absence of a so-called stool osmotic gap. The following formula is used: 2 x (stool [Na^+] + stool [K^+]) ≤ stool osmolality

The cation concentrations are doubled to estimate stool anion concentrations. The presence of a significant osmotic gap suggests the presence in stool water of a substance (or substances) other than Na/K anions that is presumably responsible for the patient's diarrhea. Originally, stool osmolality was measured, but it is almost invariably greater than the required 290–300 mosmol/kg H_2O, reflecting bacterial degradation of nonabsorbed carbohydrate either immediately before defecation or in the stool jar while awaiting chemical analysis, even when the stool is refrigerated. As a result, the stool osmolality should be assumed to be 300 mosmol/kg H_2O. A low stool osmolality (<290 mosmol/kg H_2O) reflects the addition of either dilute urine or water indicating either collection of urine and stool together or so-called factitious diarrhea, a form of Münchausen's syndrome. When the calculated difference is >50, an osmotic gap is present, suggesting that the diarrhea is due to a nonabsorbed dietary nutrient, e.g., a fatty acid and/or carbohydrate. When this difference is <25, it is presumed that a dietary nutrient is not responsible for the diarrhea. Since elements of both osmotic (i.e., malabsorption of a dietary nutrient) and secretory diarrhea may be present, this separation at times is less clear-cut at the bedside than when used as a teaching example. Ideally, the presence of an osmotic gap will be associated with a marked decrease in stool output during a prolonged fast, while the absence of an osmotic gap will likely be present in an individual whose stool output had not been reduced substantially during a period of fasting.

NUTRIENT DIGESTION AND ABSORPTION

The lengths of the small intestine and colon are ~300 cm and ~80 cm, respectively. However, the effective functional surface area is approximately 600-fold greater than that of a hollow tube as a result of the presence of folds, villi (in the small intestine), and microvilli. The functional surface area of the small intestine is somewhat greater than that of a doubles tennis court. In addition to nutrient digestion and absorption, the intestinal epithelia have several other functions:

1. *Barrier and immune defense.* The intestine is exposed to a large number of potential antigens and enteric and invasive microorganisms, and it is extremely effective preventing the entry of almost all these agents. The intestinal mucosa also synthesizes and secretes secretory IgA.
2. *Fluid and electrolyte absorption and secretion.* The intestine absorbs ~7–8 L of fluid daily, comprising dietary fluid intake (1–2 L/d) and salivary, gastric, pancreatic, biliary, and intestinal fluid (6–7 L/d). Several stimuli, especially bacteria and bacterial enterotoxins, induce fluid and electrolyte secretion that may lead to diarrhea **(Chap. 128)**.
3. *Synthesis and secretion of several proteins.* The intestinal mucosa is a major site for the production of proteins, including apolipoproteins.
4. *Production of several bioactive amines and peptides.* The intestine is one of the largest endocrine organs in the body and produces several amines (e.g., 5-hydroxytryptophan) and peptides that serve as paracrine and hormonal mediators of intestinal function.

The small and large intestines are distinct anatomically (villi are present in the small intestine but are absent in the colon) and functionally (nutrient digestion and absorption take place in the small intestine but not in the colon). No precise anatomic characteristics separate duodenum, jejunum, and ileum, although certain nutrients are absorbed exclusively in specific areas of the small intestine. However, villous cells in the small intestine (and surface epithelial cells in the colon) and crypt cells have distinct anatomic and functional characteristics. Intestinal epithelial cells are continuously renewed, with new proliferating epithelial cells at the base of the crypt migrating over 48–72 h to the tip of the villus (or surface of the colon), where they are well-developed epithelial cells with digestive and absorptive function. This high rate of cell turnover explains the relatively rapid resolution of diarrhea and other digestive tract side effects during chemotherapy as new cells not exposed to these toxic agents are produced. Equally important is the paradigm of separation of villous/surface cell and crypt cell function: Digestive hydrolytic enzymes are present primarily in the brush border of villous epithelial cells. Absorptive and secretory functions are also separated, with villous/surface cells primarily, but not exclusively, being the site for absorptive function, while secretory function is present in crypts of both the small and large intestine.

Nutrients, minerals, and vitamins are absorbed by one or more active transport mechanisms. Active transport mechanisms are energy-dependent and mediated by membrane transport proteins. These processes will result in the *net* movement of a substance against or in the absence of an electrochemical concentration gradient. Intestinal absorption of amino acids and monosaccharides,

e.g., glucose, is also a specialized form of active transport—*secondary active transport.* The movement of these actively transported nutrients against a concentration gradient is Na^+-dependent and is due to a Na^+ gradient across the apical membrane. The Na^+ gradient is maintained by Na^+, K^+-adenosine triphosphatase (ATPase), the so-called Na^+ pump located on the basolateral membrane, which extrudes Na^+ and maintains low intracellular [Na] as well as the Na^+ gradient across the apical membrane. As a result, active glucose absorption and glucose-stimulated Na^+ absorption require both the apical membrane transport protein, SGLT1, and the basolateral Na^+, K^+-ATPase. In addition to glucose absorption being Na^+-dependent, glucose also stimulates Na^+ and fluid absorption, which is the physiologic basis of oral rehydration therapy for the treatment of diarrhea **(Chap. 40)**.

The mechanisms of intestinal fluid and electrolyte absorption and secretion are discussed in Chap. 40.

Although the intestinal epithelial cells are crucial mediators of absorption and ion and water flow, the several cell types in the lamina propria (e.g., mast cells, macrophages, myofibroblasts) and the enteric nervous system interact with the epithelium to regulate mucosal cell function. The function of the intestine is the result of the integrated responses of and interactions between both intestinal epithelial cells and intestinal muscle.

Enterohepatic Circulation of Bile Acids

Bile acids are not present in the diet but are synthesized in the liver by a series of enzymatic steps that also include cholesterol catabolism. Indeed, interruption of the enterohepatic circulation of bile acids can reduce serum cholesterol levels by 10% before a new steady state is established. Bile acids are either primary or secondary: Primary bile acids are synthesized in the liver from cholesterol, and secondary bile acids are synthesized from primary bile acids in the intestine by colonic bacterial enzymes. The two primary bile acids in humans are cholic acid and chenodeoxycholic acid; the two most abundant secondary bile acids are deoxycholic acid and lithocholic acid. Approximately 500 mg bile acids are synthesized in the liver daily, conjugated to either taurine or glycine to form tauroconjugated or glycoconjugated bile acids, respectively, and are secreted into the duodenum in bile. The primary functions of bile acids are (1) to promote bile flow, (2) to solubilize cholesterol and phospholipid in the gallbladder by mixed micelle formation, and (3) to enhance dietary lipid digestion and absorption by forming mixed micelles in the proximal small intestine.

Bile acids are primarily absorbed by an active, Na^+-dependent process that is located exclusively in the ileum, though bile acids can also be absorbed to a lesser extent by non-carrier-mediated transport processes in the jejunum, ileum, and colon. Conjugated bile acids that enter the colon are deconjugated by colonic bacterial enzymes to unconjugated bile acids and are rapidly absorbed by nonionic diffusion. Colonic bacterial enzymes also dehydroxylate bile acids to secondary bile acids.

Bile acids absorbed from the intestine return to the liver via the portal vein where they are re-secreted **(Fig. 294-1)**. Bile acid synthesis is largely autoregulated by 7α-hydroxylase, the initial enzyme in cholesterol degradation. A decrease in the amount of bile acids returning to the liver from the intestine is associated with an increase in bile acid synthesis/cholesterol catabolism, which helps keep the bile acid pool size relatively constant. However, the capacity to increase bile acid synthesis is limited to about two to two and a half-fold (see below). The bile acid pool size is approximately 4 g and is circulated via the enterohepatic circulation about twice during each meal, or six to eight times during a 24-h period. A relatively small quantity of bile acids is not absorbed and is excreted in stool daily; this fecal loss is matched by hepatic bile acid synthesis.

Figure 294-1

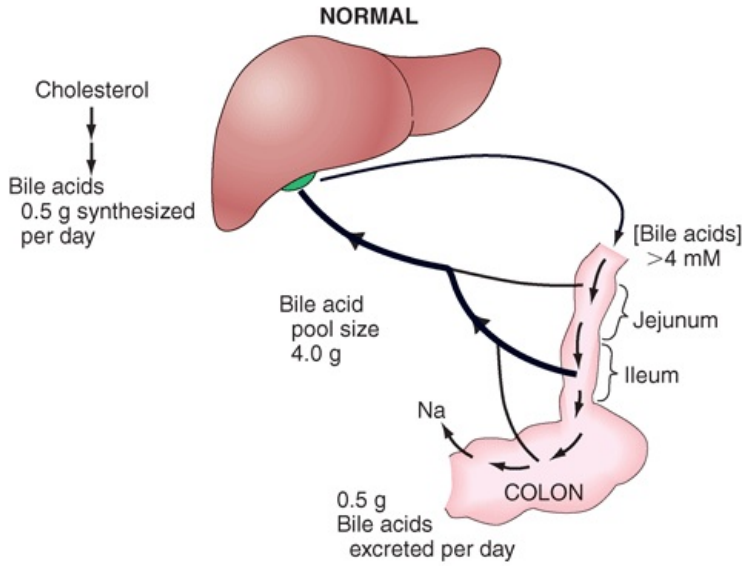

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of the enterohepatic circulation of bile acids. Bile acid synthesis is cholesterol catabolism and occurs in the liver. Bile acids are secreted in bile and are stored in the gallbladder between meals and at night. Food in the duodenum induces the release of cholecystokinin, a potent stimulus for gallbladder contraction resulting in bile acid entry into the duodenum. Bile acids are primarily absorbed via a Na-dependent transport process that is located only in the ileum. A relatively small quantity of bile acids (~500 mg) is not absorbed in a 24-h period and is lost in stool. Fecal bile acid losses are matched by bile acid synthesis. The bile acid pool (the total amount of bile acids in the body) is ~4 g and is circulated twice during each meal or six to eight times in a 24-h period.

Defects in any of the steps of the enterohepatic circulation of bile acids can result in a decrease in duodenal concentration of conjugated bile acids and, as a result, steatorrhea. Thus, steatorrhea can be caused by abnormalities in bile acid synthesis and excretion, their physical state in the intestinal lumen, and reabsorption **(Table 294-1)**.

Table 294-1 Defects in Enterohepatic Circulation of Bile Acids

Process	PathophysiologicDefect	Disease Example
Synthesis	Decreased hepatic function	Cirrhosis
Biliary secretion	Altered canalicular function	Primary biliary cirrhosis
Maintenance of conjugated bile acids	Bacterial overgrowth	Jejunal diverticulosis
Reabsorption	Abnormal ileal function	Crohn's disease

SYNTHESIS

Decreased bile acid synthesis and steatorrhea have been demonstrated in chronic liver disease, but steatorrhea is often not a major component of the illness of these patients.

SECRETION

Although bile acid secretion may be reduced or absent in biliary obstruction, steatorrhea is rarely a significant medical problem in these patients. In contrast, primary biliary cirrhosis represents a defect in canalicular excretion of organic anions, including bile acids, and not infrequently is associated with steatorrhea and its consequences, e.g., chronic bone disease. Thus, the osteopenia/osteomalacia and other chronic bone abnormalities often present in patients with primary biliary cirrhosis and other cholestatic syndromes are secondary to steatorrhea that then leads to calcium and vitamin D malabsorption as well as to the effects of cholestasis (e.g., bile acids and inflammatory cytokines).

MAINTENANCE OF CONJUGATED BILE ACIDS

In bacterial overgrowth syndromes associated with diarrhea, steatorrhea, and macrocytic anemia, a colonic type of bacterial flora is increased in the small intestine. The steatorrhea is primarily a result of the decrease in conjugated bile acids secondary to their deconjugation by colonic-type bacteria. Two complementary explanations account for the resulting impairment of micelle formation: (1) unconjugated bile acids are rapidly absorbed in the jejunum by nonionic diffusion, resulting in a reduced concentration of duodenal bile acids; and (2) the critical micellar concentration (CMC) of unconjugated bile acids is higher than that of conjugated bile acids, and therefore unconjugated bile acids are less effective than conjugated bile acids in micelle formation.

REABSORPTION

Ileal dysfunction caused by either Crohn's disease or surgical resection results in a decrease in bile acid reabsorption in the ileum and an *increase* in the delivery of bile acids to the large intestine. The resulting clinical consequences—diarrhea with or without steatorrhea—are determined by the *degree* of ileal dysfunction and the *response* of the enterohepatic circulation to bile acid losses **(Table 294-2)**. Patients with limited ileal disease or resection will often have diarrhea but not steatorrhea. The diarrhea, a result of bile acids in the colon stimulating active Cl secretion, has been called *bile acid diarrhea,* or choleretic enteropathy, and responds promptly to cholestyramine, an anion-binding resin. Such patients do not develop steatorrhea because hepatic synthesis of bile acids increases to compensate for the rate of fecal bile acid losses, resulting in maintenance of both the bile acid pool size and the intraduodenal concentrations of bile acids. In contrast, patients with greater degrees of ileal disease and/or resection will often have diarrhea and steatorrhea that do not respond to cholestyramine. In this situation, ileal disease is also associated with increased amounts of bile acids entering the colon; however, hepatic synthesis can no longer increase sufficiently to maintain the bile acid pool size. As a consequence, the intraduodenal concentration of bile acids is also reduced to less than the CMC, resulting in impaired micelle formation and steatorrhea. This second situation is often called *fatty acid diarrhea.* Cholestyramine may not be effective (and may even increase the diarrhea by further depleting the intraduodenal bile acid concentration); however, a low-fat diet to reduce fatty acids entering the colon can be effective. Two clinical features, the length of ileum removed and the degree of steatorrhea, can predict whether an individual patient will respond to cholestyramine. Unfortunately, these predictors are imperfect, and a therapeutic trial of cholestyramine is often necessary to establish whether an individual patient will benefit from cholestyramine. Table 294-2 contrasts the characteristics of bile acid diarrhea (small ileal dysfunction) and fatty acid diarrhea (large ileal dysfunction).

Table 294-2 Comparison of Bile Acid and Fatty Acid Diarrhea

	Bile Acid Diarrhea	Fatty Acid Diarrhea
Extent of ileal disease	Limited	Extensive
Ileal bile acid absorption	Reduced	Reduced
Fecal bile acid excretion	Increased	Increased
Fecal bile acid loss compensated by hepatic synthesis	Yes	No
Bile acid pool size	Normal	Reduced
Intraduodenal [bile acid]	Normal	Reduced
Steatorrhea	None or mild	>20 g
Response to cholestyramine	Yes	No
Response to low-fat diet	No	Yes

Lipids

Steatorrhea is caused by one or more defects in the digestion and absorption of dietary fat. Average intake of dietary fat in the United States is approximately 120–150 g/d, and fat absorption is linear to dietary fat intake. The total load of fat presented to the small intestine is considerably greater, as substantial amounts of lipid are secreted in bile each day. (See above for discussion of enterohepatic circulation of bile acids.) Three types of fatty acids compose fats: long-chain fatty acids (LCFAs), medium-chain fatty acids (MCFAs), and short-chain fatty acids (SCFAs) **(Table 294-3)**. Dietary fat is exclusively composed of long-chain triglycerides (LCTs), i.e., glycerol that is bound via ester-linkages to three LCFAs. While the majority of dietary LCFAs have carbon chain lengths of 16 or 18, fatty acids of carbon chain length >12 are metabolized in the same manner; saturated and unsaturated fatty acids are handled identically.

Table 294-3 Comparison of Different Types of Fatty Acids

	Long-Chain	Medium-Chain	Short-Chain
Carbon chain length	>12	8–12	<8
Present in diet	In large amounts	In small amounts	No
Origin	In diet as triglycerides	Only in small amounts in diet as triglycerides	Bacterial degradation in colon of nonabsorbed carbohydrate to fatty acids
Primary site of absorption	Small intestine	Small intestine	Colon
Requires	Yes	No	No

pancreatic lipolysis			
Requires micelle formation	Yes	No	No
Presence in stool	Minimal	No	Substantial

Assimilation of dietary lipid requires three integrated processes: (1) an intraluminal, or digestive, phase; (2) a mucosal, or absorptive, phase; and (3) a delivery, or postabsorptive, phase. An abnormality at any site of this process can cause steatorrhea **(Table 294-4)**. Therefore, it is essential that any patient with steatorrhea be evaluated to identify the specific physiologic defect in overall lipid digestion-absorption, as therapy will be determined by the specific cause of the steatorrhea.

Table 294-4 Defects in Lipid Digestion and Absorption in Steatorrhea

Phase: Process	Pathophysiologic Defect	Disease Example
Digestive		
Lipolysis formation	Decreased lipase secretion	Chronic pancreatitis
Micelle formation	Decreased intraduodenal bile acids	See Table 294-1
Absorptive		
Mucosal uptake and reesterification	Mucosal dysfunction	Celiac disease
Postabsorptive		
Chylomicron formation	Absent betalipoproteins	Abetalipoproteinemia
Delivery from intestine	Abnormal lymphatics	Intestinal lymphangiectasia

The digestive phase has two components, *lipolysis* and *micellar formation.* Although dietary lipid is in the form of LCTs, the intestinal mucosa does not absorb triglycerides; they must first be hydrolyzed **(Fig. 294-2)**. The initial step in lipid digestion is the formation of emulsions of finely dispersed lipid, which is accomplished by mastication and gastric contractions. Lipolysis, the hydrolysis of triglycerides to free fatty acids, monoglycerides, and glycerol by lipase, is initiated in the stomach by lingual and gastric lipases that have a pH optimum of 4.5–6.0. About 20–30% of total lipolysis occurs in the stomach. Lipolysis is completed in the duodenum and jejunum by pancreatic lipase, which is inactivated by a pH <7.0. Pancreatic lipolysis is greatly enhanced by the presence of a second pancreatic enzyme, colipase, which facilitates the movement of lipase to the triglyceride.

Figure 294-2

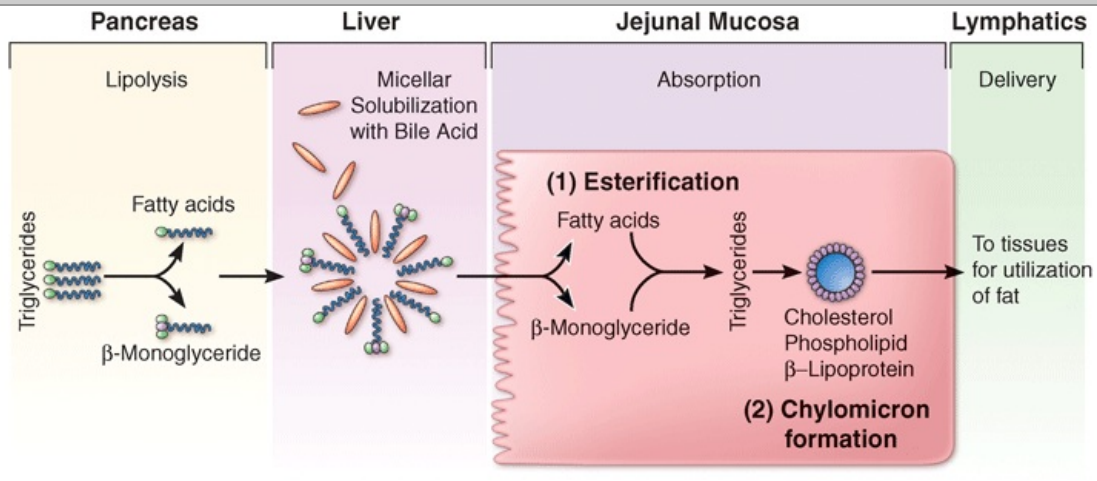

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of lipid digestion and absorption. Dietary lipid is in the form of long-chain triglycerides (LCTs). The overall process can be divided into (1) a digestive phase that includes both lipolysis and micelle formation requiring pancreatic lipase and conjugated bile acids, respectively, in the duodenum; (2) an absorptive phase for mucosal uptake and reesterification; and (3) a postabsorptive phase that includes chylomicron formation and exit from the intestinal epithelial cell via lymphatics. *(Courtesy of John M. Dietschy, MD; with permission.)*

Impaired lipolysis can lead to steatorrhea and can occur in the presence of pancreatic insufficiency

due to chronic pancreatitis in adults or cystic fibrosis in children and adolescents. Normal lipolysis can be maintained by approximately 5% of maximal pancreatic lipase secretion; thus, steatorrhea is a late manifestation of these disorders. A reduction in intraduodenal pH can also result in altered lipolysis as pancreatic lipase is inactivated at pH <7. Thus, ~15% of patients with gastrinoma **(Chap. 293)** with substantial increases in gastric acid secretion from ectopic production of gastrin (usually from an islet cell adenoma) have diarrhea, and some will have steatorrhea believed secondary to acid-inactivation of pancreatic lipase. Similarly, patients with chronic pancreatitis (who have reduced lipase secretion) often have a decrease in pancreatic bicarbonate secretion, which will also result in a decrease in intraduodenal pH and inactivation of endogenous pancreatic lipase or of therapeutically administered lipase.

Overlying the microvillus membrane of the small intestine is the so-called unstirred water layer, a relatively stagnant aqueous phase that must be traversed by the products of lipolysis that are primarily water-insoluble. Water-soluble mixed micelles provide a mechanism for the water-insoluble products of lipolysis to reach the luminal plasma membrane of villous epithelial cells, the site for lipid absorption. Mixed micelles are molecular aggregates composed of fatty acids, monoglycerides, phospholipids, cholesterol, and conjugated bile acids. Mixed micelles are formed when the concentration of conjugated bile acids is greater than its CMC, which differs among the several bile acids present in the small intestinal lumen. Conjugated bile acids, synthesized in the liver and excreted into the duodenum in bile, are regulated by the enterohepatic circulation (see above). Steatorrhea can result from impaired movement of fatty acids across the unstirred aqueous fluid layer in two situations: (1) an increase in the relative thickness of the unstirred water layer that occurs in bacterial overgrowth syndromes (see below) secondary to functional stasis (e.g., scleroderma); and (2) a decrease in the *duodenal* concentration of conjugated bile acids below its CMC, resulting in impaired micelle formation. Thus, steatorrhea can be caused by one or more defects in the enterohepatic circulation of bile acids.

Uptake and reesterification constitute the *absorptive phase* of lipid digestion-absorption. Although passive diffusion has been thought responsible, a carrier-mediated process may mediate fatty acid and monoglyceride uptake. Regardless of the uptake process, fatty acids and monoglycerides are reesterified by a series of enzymatic steps in the endoplasmic reticulum to form triglycerides, the form in which lipid exits from the intestinal epithelial cell. Impaired lipid absorption as a result of either mucosal inflammation (e.g., celiac disease) and/or intestinal resection can also lead to steatorrhea.

The reesterified triglycerides require the formation of *chylomicrons* to permit their exit from the small-intestinal epithelial cell and their delivery to the liver via the *lymphatics.* Chylomicrons are composed of β-lipoprotein and contain triglycerides, cholesterol, cholesterol esters, and phospholipids and enter the lymphatics, not the portal vein. Defects in the *postabsorptive phase* of lipid digestion-absorption can also result in steatorrhea, but these disorders are uncommon. Abetalipoproteinemia, or acanthocytosis, is a rare disorder of impaired synthesis of β-lipoprotein associated with abnormal erythrocytes (acanthocytes), neurologic problems, and steatorrhea. Lipolysis, micelle formation, and lipid uptake are all normal in patients with abetalipoproteinemia, but the reesterified triglyceride cannot exit from the epithelial cell because of the failure to produce chylomicrons. Small-intestinal biopsies of these rare patients in the postprandial state reveal lipid-laden small-intestinal epithelial cells that become perfectly normal in appearance following a 72–96 h fast. Similarly, abnormalities of intestinal lymphatics (e.g., intestinal lymphangiectasia) may also be associated with steatorrhea as well as protein loss (see below). Steatorrhea can result from defects at any of the several steps in lipid digestion-absorption.

The mechanism of lipid digestion-absorption outlined above is limited to *dietary* lipid that is almost exclusively in the form of LCTs (Table 294-3). Medium-chain triglycerides (MCTs), composed of fatty acids with carbon chain lengths of 8–12, are present in large amounts in coconut oil and are used as a nutritional supplement. MCTs can be digested and absorbed by a different pathway from LCTs and at one time held promise as an important treatment of steatorrhea of almost all etiologies. Unfortunately, their therapeutic effects have been less than expected because their use is often not associated with an increase in body weight for reasons that are not completely understood.

MCTs, in contrast to LCTs, do not require pancreatic lipolysis as the triglyceride can be absorbed intact by the intestinal epithelial cell. Further, micelle formation is not necessary for the absorption of MCTs or medium-chain fatty acids, if hydrolyzed by pancreatic lipase. MCTs are absorbed more efficiently than LCTs for the following reasons: (1) The rate of MCT absorption is greater than that of long-chain fatty acids; (2) medium-chain fatty acids following absorption are not reesterified; (3) following absorption, MCTs are hydrolyzed to medium-chain fatty acids; (4) MCTs do not require chylomicron formation for their exit from the intestinal epithelial cells; and (5) their route of exit is via the portal vein and not via lymphatics. Thus, the absorption of MCTs is greater than that of LCTs in pancreatic insufficiency, conditions with reduced intraduodenal bile acid concentrations, small-intestinal mucosal disease, abetalipoproteinemia, and intestinal lymphangiectasia.

SCFAs are not dietary lipids but are synthesized by colonic bacterial enzymes from nonabsorbed carbohydrate and are the anions in highest concentration in stool (between 80 and 130 m*M*). The SCFAs present in stool are primarily acetate, propionate, and butyrate, whose carbon chain lengths

are 2, 3, and 4, respectively. Butyrate is the primary nutrient for colonic epithelial cells, and its deficiency may be associated with one or more colitides. SCFAs conserve calories and carbohydrate, because carbohydrates not completely absorbed in the small intestine will not be absorbed in the large intestine due to the absence of both disaccharidases and SGLT1, the transport protein that mediates monosaccharide absorption. In contrast, SCFAs are rapidly absorbed and stimulate colonic Na-Cl and fluid absorption. Most non–*Clostridium difficile* antibiotic-associated diarrhea is due to antibiotic suppression of colonic microbiota, with a resulting decrease in SCFA production. As *C. difficile* accounts for about 15–20% of all antibiotic-associated diarrhea, a relative decrease in colonic production of SCFA is likely the cause of most antibiotic-associated diarrhea.

The clinical manifestations of steatorrhea are a consequence of both the underlying disorder responsible for the development of steatorrhea and steatorrhea per se. Depending on the degree of steatorrhea and the level of dietary intake, significant fat malabsorption may lead to weight loss. Steatorrhea per se can be responsible for diarrhea; if the primary cause of the steatorrhea has not been identified, a low-fat diet can often ameliorate the diarrhea by decreasing fecal fat excretion. Steatorrhea is often associated with fat-soluble vitamin deficiency, which will require replacement with water-soluble preparations of these vitamins.

Disorders of absorption may also be associated with malabsorption of other dietary nutrients, most often carbohydrates, with or without a decrease in dietary lipid digestion and absorption. Therefore, knowledge of the mechanism of the digestion and absorption of carbohydrates, proteins, and other minerals and vitamins is useful in the evaluation of patients with altered intestinal nutrient absorption.

Carbohydrates

Carbohydrates in the diet are present in the form of starch, disaccharides (sucrose and lactose), and glucose. Carbohydrates are absorbed only in the small intestine and only in the form of monosaccharides. Therefore, before their absorption, starch and disaccharides must first be digested by pancreatic amylase and intestinal brush border disaccharidases to monosaccharides. Monosaccharide absorption occurs by a Na-dependent process mediated by the brush border transport protein SGLT1.

Lactose malabsorption is the only clinically important disorder of carbohydrate absorption. Lactose, the disaccharide present in milk, requires digestion by brush border lactase to its two constituent monosaccharides, glucose and galactose. Lactase is present in almost all species in the postnatal period but then disappears throughout the animal kingdom, except in humans. Lactase activity persists in many individuals throughout life. Two different types of lactase deficiency exist—primary and secondary. In *primary lactase deficiency,* a genetically determined decrease or absence of lactase is noted, while all other aspects of both intestinal absorption and brush border enzymes are normal. In a number of non-white groups, primary lactase deficiency is common in adulthood. **Table 294-5** presents the incidence of primary lactase deficiency in several ethnic groups. Northern European and North American whites are the only groups to maintain small-intestinal lactase activity throughout adult life. The persistence of lactase is the abnormality due to a defect in the regulation of its maturation. In contrast, *secondary lactase deficiency* occurs in association with small-intestinal mucosal disease with abnormalities in both structure and function of other brush border enzymes and transport processes. Secondary lactase deficiency is often seen in celiac disease.

Table 294-5 Primary Lactase Deficiency in Different Adult Ethnic Groups

Ethnic Group	Prevalence of Lactase Deficiency, %
Northern European	5–15
Mediterranean	60–85
African black	85–100
American black	45–80
American white	10–25
Native American	50–95
Mexican American	40–75
Asian	90–100

Source: From FJ Simons: Am J Dig Dis 23:963, 1978.

As lactose digestion is rate-limiting compared to glucose/galactose absorption, lactase deficiency is associated with significant lactose malabsorption. Some individuals with lactose malabsorption develop symptoms such as diarrhea, abdominal pain, cramps, and/or flatus. Most individuals with primary lactase deficiency do not have symptoms. Since lactose intolerance may be associated with symptoms suggestive of irritable bowel syndrome, persistence of such symptoms in an individual with lactose intolerance while on a strict lactose-free diet would suggest that the individual's symptoms were related to irritable bowel syndrome.

Development of symptoms of lactose intolerance is related to several factors:

1. *Amount of lactose in the diet* .
2. *Rate of gastric emptying.* Symptoms are more likely when gastric emptying is rapid than when gastric emptying is slower. Therefore, it is more likely that skim milk will be associated with symptoms of lactose intolerance than will whole milk, as the rate of gastric emptying following skim milk intake is more rapid. Similarly, the diarrhea observed following subtotal gastrectomy is often a result of lactose intolerance, as gastric emptying is accelerated in patients with a gastrojejunostomy.
3. *Small-intestinal transit time.* Although the small and large intestine contribute to the development of symptoms, many of the symptoms of lactase deficiency are related to the interaction of colonic bacteria and nonabsorbed lactose. More rapid small-intestinal transit makes symptoms more likely.
4. *Colonic compensation by production of SCFAs from nonabsorbed lactose.* Reduced levels of colonic microflora, which can occur following antibiotic use, will also be associated with increased symptoms following lactose ingestion, especially in a lactase-deficient individual.

Glucose-galactose or monosaccharide malabsorption may also be associated with diarrhea and is due to a congenital absence of SGLT1. Diarrhea is present when individuals with this disorder ingest carbohydrates that contain actively transported monosaccharides (e.g., glucose, galactose) but not monosaccharides that are not actively transported (e.g., fructose). Fructose is absorbed by the brush border transport protein GLUT 5, a facilitated diffusion process that is not Na-dependent and is distinct from SGLT1. In contrast, some individuals develop diarrhea as a result of consuming large quantities of sorbitol, a sugar used in diabetic candy; sorbitol is only minimally absorbed due to the absence of an intestinal absorptive transport mechanism for sorbitol.

Proteins

Protein is present in food almost exclusively as polypeptides and requires extensive hydrolysis to di- and tripeptides and amino acids before absorption. Proteolysis occurs in both the stomach and small intestine; it is mediated by pepsin secreted as pepsinogen by gastric chief cells and trypsinogen and other peptidases from pancreatic acinar cells. These proenzymes, pepsinogen and trypsinogen, must be activated to pepsin (by pepsin in the presence of a pH <5) and to trypsin (by the intestinal brush border enzyme enterokinase and subsequently by trypsin), respectively. Proteins are absorbed by separate transport systems for di- and tripeptides and for different types of amino acids, e.g., neutral and dibasic. Alterations in either protein or amino acid digestion and absorption are rarely observed clinically, even in the presence of extensive small-intestinal mucosal inflammation. However, three rare genetic disorders involve protein digestion-absorption: (1) *Enterokinase deficiency* is due to an absence of the brush border enzyme that converts the proenzyme trypsinogen to trypsin and is associated with diarrhea, growth retardation, and hypoproteinemia. (2) *Hartnup syndrome,* a defect in neutral amino acid transport, is characterized by a pellagra-like rash and neuropsychiatric symptoms. (3) *Cystinuria,* a defect in dibasic amino acid transport, is associated with renal calculi and chronic pancreatitis.

Approach to the Patient: Malabsorption

The clues provided by the history, symptoms, and initial preliminary observations will serve to limit extensive, ill-focused, and expensive laboratory and imaging studies. For example, a clinician evaluating a patient with symptoms suggestive of malabsorption, who recently had extensive small-intestinal resection for mesenteric ischemia, should direct the initial assessment almost exclusively to define whether a short bowel syndrome might explain the entire clinical picture. Similarly, the development of a pattern of bowel movements suggestive of steatorrhea in a patient with long-standing alcohol abuse and chronic pancreatitis should lead toward assessing pancreatic exocrine function.

The classic picture of malabsorption is rarely seen today in most parts of the United States. As a consequence, diseases with malabsorption must be suspected in individuals with less severe symptoms and signs and with subtle evidence of the altered absorption of only a *single* nutrient rather than obvious evidence of the malabsorption of multiple nutrients.

Although diarrhea can be caused by changes in fluid and electrolyte movement in either the small or the large intestine, dietary nutrients are absorbed almost exclusively in the small intestine. Therefore, the demonstration of diminished absorption of a dietary nutrient provides unequivocal evidence of small-intestinal disease, although colonic dysfunction may also be present (e.g., Crohn's disease may involve both small and large intestine). Dietary nutrient absorption may be segmental or diffuse along the small intestine and is site-specific. Thus, for example, calcium, iron, and folic acid are exclusively absorbed by active transport processes in the proximal small intestine, especially the duodenum; in contrast, the active transport mechanisms for both cobalamin and bile acids are present only in the ileum. Therefore, in an individual who years previously had had an intestinal resection, the details of which are not presently available, a presentation with evidence of calcium, folic acid, and/or iron

malabsorption but without cobalamin deficiency would make it likely that the duodenum and proximal jejunum, but not ileum, had been resected.

Some nutrients, e.g., glucose, amino acids, and lipids, are absorbed throughout the small intestine, though their rate of absorption is greater in the proximal than in the distal segments. However, following segmental resection of the small intestine, the remaining segments undergo both morphologic and functional "adaptation" to enhance absorption. Such adaptation is secondary to the presence of luminal nutrients and hormonal stimuli and may not be complete in humans for several months following the resection. Adaptation is critical for survival in individuals who have undergone massive resection of the small intestine and/or colon.

Establishing the presence of steatorrhea and identifying its specific cause are often quite difficult. The "gold standard" still remains a timed, quantitative stool fat determination. On a practical basis, stool collections are invariably difficult and often incomplete, as nobody wants to handle stool. A qualitative test—Sudan III stain—has long been available to establish the presence of an increase in stool fat. This test is rapid and inexpensive but, as a qualitative test, does not establish the degree of fat malabsorption and is best used as a preliminary screening study. Many of the blood, breath, and isotopic tests that have been developed (1) do not directly measure fat absorption; (2) have excellent sensitivity when steatorrhea is obvious and severe but have poor sensitivity when steatorrhea is mild (e.g., stool chymotrypsin, elastase, that can potentially distinguish pancreatic from nonpancreatic etiologies of steatorrhea); or (3) have not survived the transition from the research laboratory to commercial application.

Despite this situation, the use of routine laboratory studies (i.e., complete blood count, prothrombin time, serum protein determination, alkaline phosphatase) may suggest the presence of dietary nutrient depletion, especially iron, folate, cobalamin, and vitamins D and K. Additional studies include measurement of serum carotene, cholesterol, albumin, iron, folate, and cobalamin levels. The serum carotene level can also be reduced if the patient has poor dietary intake of leafy vegetables.

If steatorrhea and/or altered absorption of other nutrients are suspected, the history, clinical observations, and laboratory testing can help detect deficiency of a nutrient, especially the fat-soluble vitamins A, D, E, or K. Thus, evidence of metabolic bone disease with elevated alkaline phosphatase and/or reduced serum calcium levels would suggest vitamin D malabsorption. A deficiency of vitamin K would be suggested by an elevated prothrombin time in an individual without liver disease who was not taking anticoagulants. Macrocytic anemia would lead to evaluation of whether cobalamin or folic acid malabsorption was present. The presence of iron-deficiency anemia in the absence of occult bleeding from the gastrointestinal tract in either a male or a nonmenstruating female would require evaluation of iron malabsorption and the exclusion of celiac disease, as iron is absorbed exclusively in the proximal small intestine.

At times, however, a timed (72 h) quantitative stool collection, preferably on a defined diet, must be obtained to determine stool fat content and establish the presence of steatorrhea. The presence of steatorrhea then requires further assessment to establish the pathophysiologic process(es) responsible for the defect in dietary lipid digestion-absorption (Table 294-4). Some of the other studies include the Schilling test, D-xylose test, duodenal mucosal biopsy, small-intestinal radiologic examination, and tests of pancreatic exocrine function.

THE SCHILLING TEST

This test is performed to determine the cause of cobalamin malabsorption. Unfortunately, the Schilling test has not been available commercially in the United States for the past few years. Since understanding the physiology and pathophysiology of cobalamin absorption is very valuable to enhance one's understanding of aspects of gastric, pancreatic, and ileal function, discussion of the Schilling test is provided in **Chap. e37**.

URINARY D-XYLOSE TEST

The urinary D-xylose test for carbohydrate absorption provides an assessment of proximal small-intestinal mucosal function. D-Xylose, a pentose, is absorbed almost exclusively in the proximal small intestine. The D-xylose test is usually performed by giving 25 g D-xylose and collecting urine for 5 h. An abnormal test (<4.5 g excretion) primarily reflects the presence of duodenal/jejunal mucosal disease. The D-xylose test can also be abnormal in patients with blind loop syndrome (as a consequence primarily of abnormal intestinal mucosa) and, as a false-positive study, in patients with large collections of fluid in a third space (i.e., ascites, pleural fluid). The ease of obtaining a mucosal biopsy of the small intestine by endoscopy and the false-negative rate of the D-xylose test have led to its diminished use. When small-intestinal mucosal disease is suspected, a small-intestinal mucosal biopsy should be performed.

RADIOLOGIC EXAMINATION

Radiologic examination of the small intestine using barium contrast (small-bowel series or study) can provide important information in the evaluation of the patient with presumed or suspected malabsorption. These studies are most often performed in conjunction with the examination of the esophagus, stomach, and duodenal bulb, and insufficient barium is given to the patient to permit an

adequate examination of the small-intestinal mucosa, especially the ileum. As a result, many gastrointestinal radiologists alter the procedure of a barium contrast examination of the small intestine by performing either a small-bowel series in which a large amount of barium is given by mouth without concurrent examination of the esophagus and stomach or an enteroclysis study in which a large amount of barium is introduced into the duodenum via a fluoroscopically placed tube. In addition, many of the diagnostic features initially described by radiologists to denote the presence of small-intestinal disease (e.g., flocculation, segmentation) are rarely seen with current barium suspensions. Nonetheless, in skilled hands barium contrast examination of the small intestine can yield important information. For example, with extensive mucosal disease, dilation of intestine can be seen, as dilution of barium from increased intestinal fluid secretion **(Fig. 294-3)**. A normal barium contrast study does *not* exclude the possibility of small-intestinal disease. However, a small-bowel series remains a useful examination to look for anatomic abnormalities, such as strictures and fistulas (as in Crohn's disease) or blind loop syndrome (e.g., multiple jejunal diverticula), and to define the extent of a previous surgical resection. Other imaging studies to assess the integrity of small intestinal morphology are CT enteroclysis and magnetic resonance (MR) enteroclysis, while capsule endoscopy and double-barrel enteroscopy are other useful aids in the diagnostic assessment of small intestinal pathology.

Figure 294-3

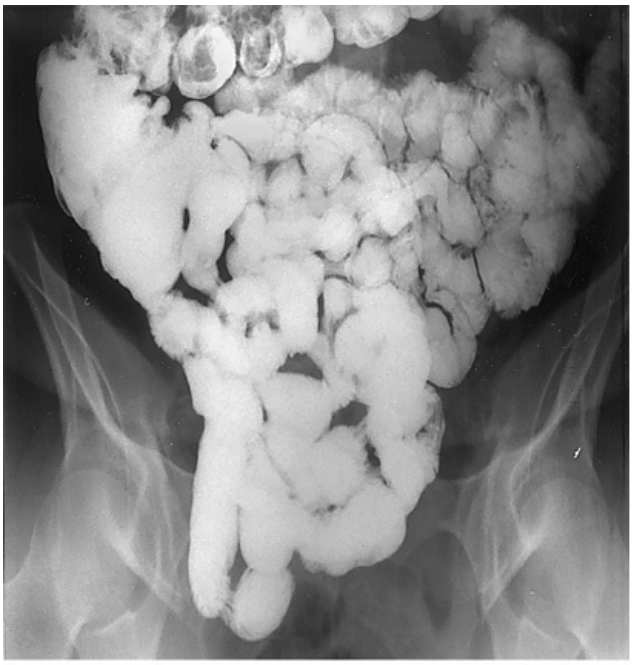

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Barium contrast small-intestinal radiologic examinations. ***A.*** Normal individual. ***B.*** Celiac sprue. ***C.*** Jejunal diverticulosis. ***D.*** Crohn's disease. *(Courtesy of Morton Burrell, MD, Yale University; with permission.)*

BIOPSY OF SMALL-INTESTINAL MUCOSA

A small-intestinal mucosal biopsy is essential in the evaluation of a patient with documented steatorrhea or chronic diarrhea (lasting >3 weeks) **(Chap. 40)**. The ready availability of endoscopic equipment to examine the stomach and duodenum has led to its almost uniform use as the preferred method to obtain histologic material of proximal small-intestinal mucosa. The primary indications for a small-intestinal biopsy are (1) evaluation of a patient either with documented or suspected steatorrhea or with chronic diarrhea, and (2) diffuse or focal abnormalities of the small intestine defined on a small-intestinal series. Lesions seen on small-bowel biopsy can be classified into three different categories **(Table 294-6)**:

1. *Diffuse, specific lesions.* Relatively few diseases associated with altered nutrient absorption have specific histopathologic abnormalities on small-intestinal mucosal biopsy, and they are uncommon. Whipple's disease is characterized by the presence of periodic acid–Schiff (PAS) –positive macrophages in the lamina propria, while the bacilli that are also present may require electron-microscopic examination for identification **(Fig. 294-4)**. *Abetalipoproteinemia* is characterized by a normal mucosal appearance except for the presence of mucosal absorptive cells that contain lipid postprandially and disappear following a prolonged period of either fat-free intake or fasting. *Immune globulin deficiency* is associated with a variety of histopathologic findings on small-intestinal mucosal biopsy. The characteristic feature is the absence of or substantial reduction in the number of plasma cells in the lamina propria; the mucosal architecture may be either perfectly normal or flat (i.e., villous atrophy). As patients with immune globulin deficiency are often infected with *Giardia lamblia,Giardia* trophozoites may also be seen in the biopsy.
2. *Patchy, specific lesions.* Several diseases show abnormal small-intestinal mucosa with a patchy distribution. As a result, biopsies obtained randomly or in the absence of abnormalities visualized endoscopically may not reveal the diagnostic features. Intestinal *lymphoma* can at times be diagnosed on mucosal biopsy by the identification of malignant lymphoma cells in the lamina propria and submucosa **(Chap. 110)**. The presence of dilated lymphatics in the submucosa and sometimes in the lamina propria indicates the presence of *lymphangiectasia* associated with hypoproteinemia secondary to protein loss into the intestine. *Eosinophilic gastroenteritis* comprises a heterogeneous group of disorders with a spectrum of presentations and symptoms with an eosinophilic infiltrate of the lamina propria, with or without peripheral eosinophilia. The patchy nature of the infiltrate as well as its presence in the submucosa often leads to an absence of histopathologic findings on mucosal biopsy. As the involvement of the duodenum in *Crohn's disease* is also submucosal and not necessarily continuous, mucosal biopsies are not the most direct approach to the diagnosis of duodenal Crohn's disease **(Chap. 295)**. Amyloid deposition can be identified by Congo Red stain in some patients with *amyloidosis* involving the duodenum **(Chap. 111)**.

3. Several microorganisms can be identified on small-intestinal biopsies, establishing a correct diagnosis. At times the small biopsy is performed to establish the diagnosis of the infection, e.g., Whipple's disease or giardiasis. In most other instances the infection is picked up incidentally during the workup of diarrhea or other abdominal symptoms. Many of these infections occur in immunocompromised patients with diarrhea and include *Cryptosporidium, Isospora belli, Microsporidia, Cyclospora, Toxoplasma,* cytomegalovirus, adenovirus, *Mycobacterium avium-intracellulare,* and *G. lamblia.* In immunocompromised patients, when *Candida, Aspergillus, Cryptococcus,* or *Histoplasma* organisms are seen on duodenal biopsy, their presence generally reflects systemic infection. Apart from Whipple's disease and infections in the immunocompromised host, the small bowel biopsy is seldom used as the primary mode to diagnose infection. Even giardiasis is more easily diagnosed with duodenal aspirates and/or stool antigen studies than by duodenal biopsy.
4. *Diffuse, nonspecific lesions.Celiac disease* presents with a characteristic mucosal appearance on duodenal/proximal jejunal mucosal biopsy that is *not* diagnostic of the disease. The diagnosis of celiac disease is established by clinical, histologic, and immunologic response to a gluten-free diet. *Tropical sprue* is associated with histologic findings similar to those of celiac disease after a tropical or subtropical exposure but does not respond to gluten restriction; most often symptoms improve with antibiotics and folate administration.

Table 294-6 Disease that Can Be Diagnosed by Small-Intestinal Mucosal Biopsies

Lesions	Pathologic Findings
Diffuse, Specific	
Whipple's disease	Lamina propria contains macrophages containing PAS+ material
Agammaglobulinemia	No plasma cells; either normal or absent villi ("flat mucosa")
Abetalipoproteinemia	Normal villi; epithelial cells vacuolated with fat postprandially
Patchy, Specific	
Intestinal lymphoma	Malignant cells in lamina propria and submucosa
Intestinal lymphangiectasia	Dilated lymphatics; clubbed villi
Eosinophilic gastroenteritis	Eosinophil infiltration of lamina propria and mucosa
Amyloidosis	Amyloid deposits
Crohn's disease	Noncaseating granulomas
Infection by one or more microorganisms (see text)	Specific organisms
Mastocytosis	Mast cell infiltration of lamina propria
Diffuse, Nonspecific	
Celiac disease	Short or absent villi; mononuclear infiltrate; epithelial cell damage; hypertrophy of crypts
Tropical sprue	Similar to celiac disease
Bacterial overgrowth	Patchy damage to villi; lymphocyte infiltration
Folate deficiency	Short villi; decreased mitosis in crypts; megalocytosis
Vitamin B12 deficiency	Similar to folate deficiency
Radiation enteritis	Similar to folate deficiency
Zollinger-Ellison syndrome	Mucosal ulceration and erosion from acid
Protein-calorie malnutrition	Villous atrophy; secondary bacterial overgrowth
Drug-induced enteritis	Variable histology

Abbreviation: PAS+, periodic acid–Schiff positive.

Figure 294-4

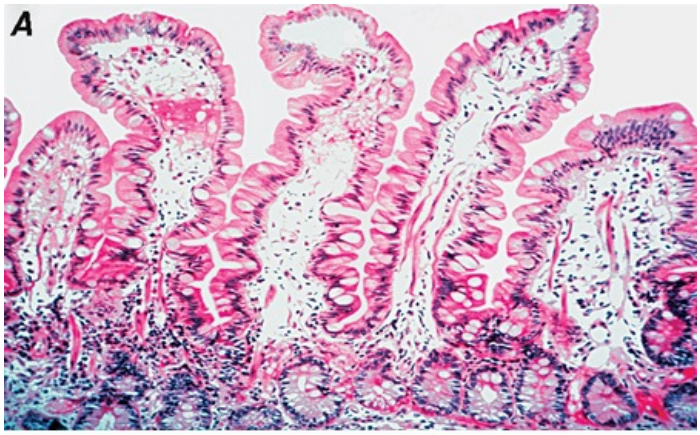

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

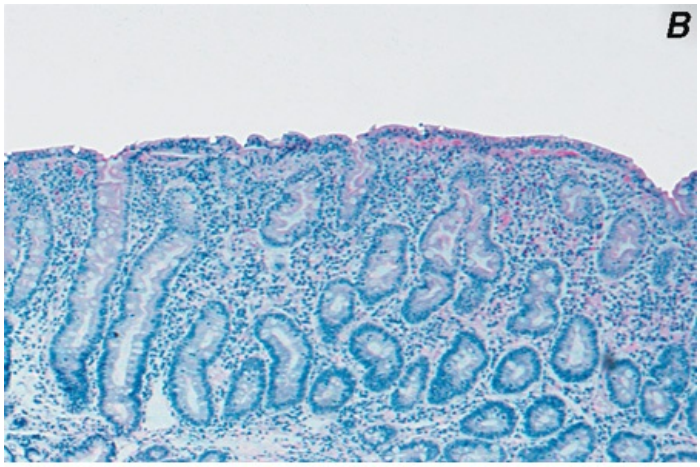

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

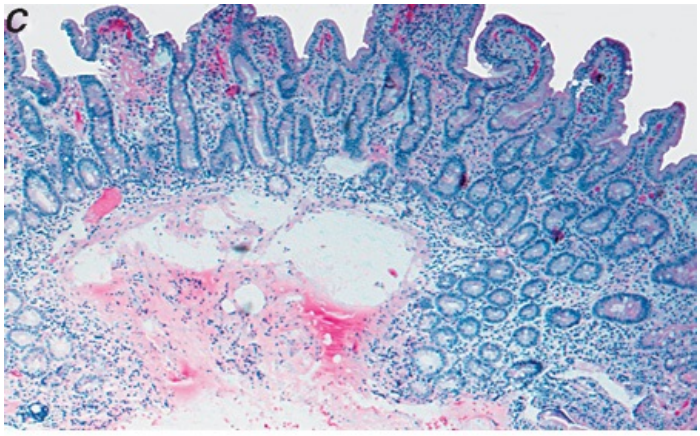

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

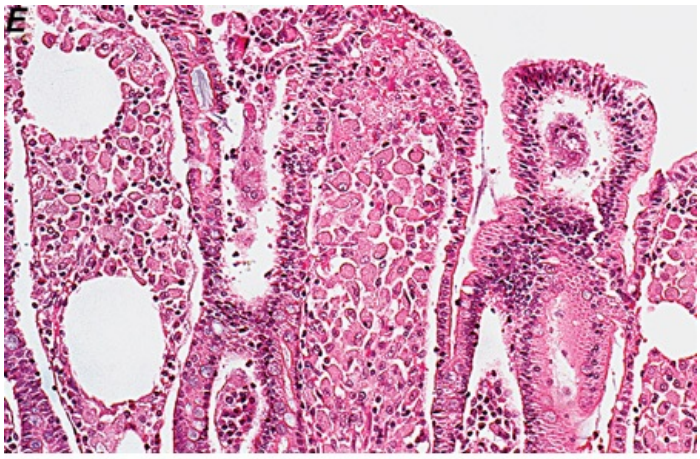

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

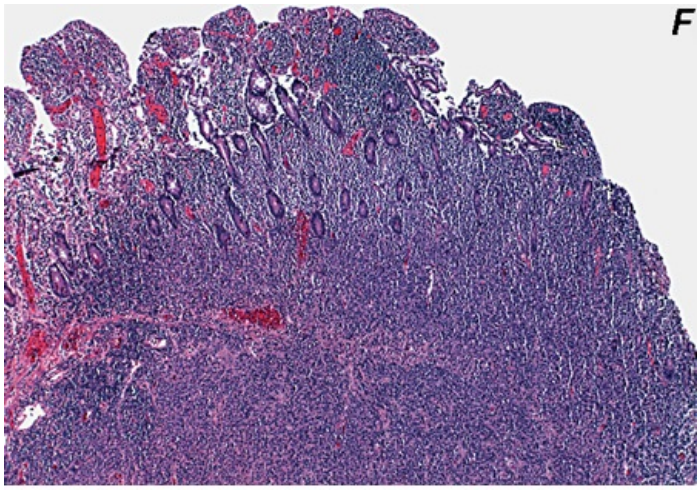

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

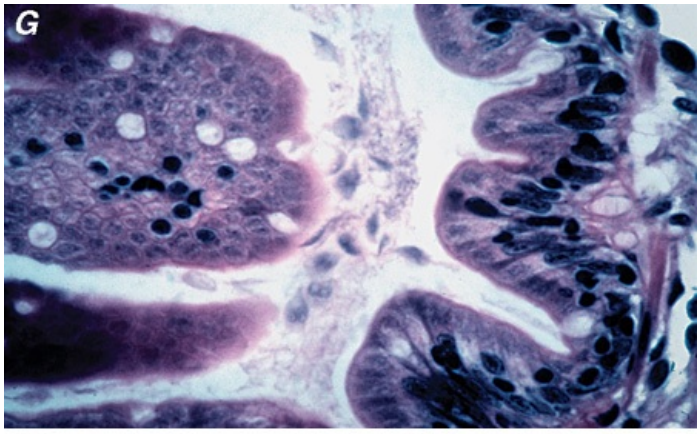

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Small-intestinal mucosal biopsies. ***A.*** Normal individual. ***B.*** Untreated celiac sprue. ***C.*** Treated celiac sprue. ***D.*** Intestinal lymphangiectasia. ***E.*** Whipple's disease. ***F.*** Lymphoma. ***G.*** Giardiasis. *(Courtesy of Marie Robert, MD, Yale University; with permission.)*

Patients with steatorrhea require assessment of *pancreatic exocrine function,* which is often abnormal in chronic pancreatitis. The secretin test that collects pancreatic secretions by duodenal intubation following intravenous administration of secretin is the only test that directly measures pancreatic exocrine function but is available only at few specialized centers. Endoscopic approaches provide excellent assessment of pancreatic duct anatomy but do *not* assess exocrine function **(Chap. 312)**.

Table 294-7 summarizes the results of the D-xylose test, Schilling test, and small-intestinal mucosal biopsy in patients with five different causes of steatorrhea.

Table 294-7 Results of Diagnostic Studies in Different Causes of Steatorrhea

	D-Xylose Test	Schilling Test	Duodenal Mucosal Biopsy
Chronic pancreatitis	Normal	50% abnormal; if abnormal, normal with pancreatic enzymes	Normal
Bacterial overgrowth syndrome	Normal or only modestly abnormal	Often abnormal; if abnormal, normal after antibiotics	Usually normal
Ileal disease	Normal	Abnormal	Normal
Celiac disease	Decreased	Normal	Abnormal: probably "flat"
Intestinal lymphangiectasia	Normal	Normal	Abnormal: "dilated lymphatics"

SPECIFIC DISEASE ENTITIES

Celiac Disease

Celiac disease is a common cause of malabsorption of one or more nutrients. Though originally considered largely a disease in whites, especially those of European descent, recent observations have established that celiac disease is a common disease with protean manifestations, a worldwide distribution, and an estimated incidence in the United States that is as high as 1:113 people. Its incidence has increased over the past 50 years. Celiac disease has had several other names, including nontropical sprue, celiac sprue, adult celiac disease, and gluten-sensitive enteropathy. The etiology of celiac disease is not known, but environmental, immunologic, and genetic factors are important. Celiac disease is considered an "iceberg" disease with a small number of individuals with classical symptoms and manifestations related to nutrient malabsorption, and a varied natural history, with the onset of symptoms occurring at ages ranging from the first year of life through the eighth decade. A much larger number of individuals have manifestations that are not obviously related to intestinal malabsorption, e.g., anemia, osteopenia, infertility, neurologic symptoms ("atypical celiac disease"); while an even larger group are essentially asymptomatic though with abnormal small intestinal histopathology and serologies (see below) and are referred to as "silent' celiac disease.

The hallmark of celiac disease is the presence of an abnormal small-intestinal biopsy (Fig. 294-4) and the response of the condition—symptoms and the histologic changes on the small-intestinal biopsy—to the elimination of gluten from the diet. The histologic changes have a proximal-to-distal intestinal

distribution of severity, which probably reflects the exposure of the intestinal mucosa to varied amounts of dietary gluten; the symptoms do not necessarily correlate with histologic changes especially as many newly diagnosed patients with celiac disease may be asymptomatic.

The symptoms of celiac disease may appear with the introduction of cereals in an infant's diet, although spontaneous remissions often occur during the second decade of life that may be either permanent or followed by the reappearance of symptoms over several years. Alternatively, the symptoms of celiac disease may first become evident at almost any age throughout adulthood. In many patients, frequent spontaneous remissions and exacerbations occur. The symptoms range from significant malabsorption of multiple nutrients, with diarrhea, steatorrhea, weight loss, and the consequences of nutrient depletion (i.e., anemia and metabolic bone disease), to the absence of any gastrointestinal symptoms but with evidence of the depletion of a single nutrient (e.g., iron or folate deficiency, osteomalacia, edema from protein loss). Asymptomatic relatives of patients with celiac disease have been identified as having this disease either by small-intestinal biopsy or by serologic studies [e.g., antiendomysial antibodies, tissue transglutaminase (tTG)]. The availability of these "celiac serologies" has led to a substantial increase in the diagnosis of celiac disease, and the diagnosis is now being made primarily in patients without "classic" symptoms but with atypical and subclinical presentations.

ETIOLOGY

The etiology of celiac disease is not known, but environmental, immunologic, and genetic factors all appear to contribute to the disease. One *environmental* factor is the clear association of the disease with gliadin, a component of gluten that is present in wheat, barley, and rye. In addition to the role of gluten restriction in treatment, the instillation of gluten into both normal-appearing rectum and distal ileum of patients with celiac disease results in morphologic changes within hours.

An *immunologic* component in the pathogenesis of celiac disease is critical and involves both adaptive and innate immune responses. Serum antibodies—IgA antigliadin, IgA antiendomysial, and IgA anti-tTG antibodies—are present, but it is not known whether such antibodies are primary or secondary to the tissue damage. The antiendomysial antibody has 90–95% sensitivity and 90–95% specificity; the antigen recognized by the antiendomysial antibody is tTG, which deaminates gliadin, which is presented to HLA-DQ2 or HLA-DQ8 (see below). Antibody studies are frequently used to identify patients with celiac disease; patients with these antibodies should undergo duodenal biopsy. This autoantibody has not been linked to a pathogenetic mechanism (or mechanisms) responsible for celiac disease. Nonetheless, this antibody is useful in establishing the true prevalence of celiac disease in the general population. A 4-week treatment with prednisolone of a patient with celiac disease who continues to eat gluten will induce a remission and convert the "flat" abnormal duodenal biopsy to a more normal-appearing one. In addition, gliadin peptides interact with gliadin-specific T cells that mediate tissue injury and induce the release of one or more cytokines (e.g., IFN-ɣ) that cause tissue injury.

Genetic factor(s) are also involved in celiac disease. The incidence of symptomatic celiac disease varies widely in different population groups (high in whites, low in blacks and Asians) and is 10% in first-degree relatives of celiac disease patients; however, serologic studies provide clear evidence that celiac disease is present worldwide. Furthermore, all patients with celiac disease express the HLA-DQ2 or HLA-DQ8 allele, though only a minority of people expressing DQ2/DQ8 have celiac disease. Absence of DQ2/DQ8 excludes the diagnosis of celiac disease.

DIAGNOSIS

A small-intestinal biopsy is required to establish a diagnosis of celiac disease (Fig. 294-4). A biopsy should be performed in patients with symptoms and laboratory findings suggestive of nutrient malabsorption and/or deficiency and with a positive endomysial antibody test. Since the presentation of celiac disease is often subtle, without overt evidence of malabsorption or nutrient deficiency, a relatively low threshold to perform a biopsy is important. It is more prudent to perform a biopsy than to obtain another test of intestinal absorption, which can never completely exclude or establish this diagnosis.

The diagnosis of celiac disease requires the presence of characteristic histologic changes on small-intestinal biopsy together with a prompt clinical and histologic response following the institution of a gluten-free diet. If serologic studies have detected the presence of IgA antiendomysial or tTG antibodies, they too should disappear after a gluten-free diet is started. With the increase in number of patients diagnosed with celiac disease that have been largely identified by serologic studies, the spectrum of histologic changes seen on duodenal biopsy has increased and includes findings that are not as severe as the classic changes shown in Fig. 294-4. The classical changes seen on duodenal/jejunal biopsy are restricted to the mucosa and include (1) an increase in the number of intraepithelial lymphocytes; (2) absence or reduced height of villi, resulting in a flat appearance with increased crypt cell proliferation, resulting in crypt hyperplasia and loss of villous structure, with consequent villous, but not mucosal, atrophy; (3) cuboidal appearance and nuclei that are no longer oriented basally in surface epithelial cells; and (4) increased lymphocytes and plasma cells in the lamina propria (Fig. 294-4*B*). Although these features are characteristic of celiac disease, they are *not*

diagnostic because a similar appearance can be seen in tropical sprue, eosinophilic enteritis, and milk-protein intolerance in children and occasionally in lymphoma, bacterial overgrowth, Crohn's disease, and gastrinoma with acid hypersecretion. However, the presence of a characteristic histologic appearance that reverts toward normal following the initiation of a gluten-free diet establishes the diagnosis of celiac disease (Fig. 294-4*C*). Readministration of gluten with or without an additional small-intestinal biopsy is not necessary.

FAILURE TO RESPOND TO GLUTEN RESTRICTION

The most common cause of persistent symptoms in a patient who fulfills all the criteria for the diagnosis of celiac disease is continued intake of gluten. Gluten is ubiquitous, and significant effort must be made to exclude all gluten from the diet. Use of rice in place of wheat flour is very helpful, and several support groups provide important aid to patients with celiac disease and to their families. More than 90% of patients who have the characteristic findings of celiac disease will respond to complete dietary gluten restriction. The remainder constitute a heterogeneous group (whose condition is often called *refractory celiac disease or refractory sprue*) that includes some patients who (1) respond to restriction of other dietary protein, e.g., soy; (2) respond to glucocorticoids; (3) are "temporary" (i.e., the clinical and morphologic findings disappear after several months or years); or (4) fail to respond to all measures and have a fatal outcome, with or without documented complications of celiac disease, such as development of intestinal T cell lymphoma.

MECHANISM OF DIARRHEA

The diarrhea in celiac disease has several pathogenetic mechanisms. Diarrhea may be secondary to (1) steatorrhea, which is primarily a result of the changes in jejunal mucosal function; (2) secondary lactase deficiency, a consequence of changes in jejunal brush border enzymatic function; (3) bile acid malabsorption resulting in bile acid–induced fluid secretion in the colon, in cases with more extensive disease involving the ileum; and (4) endogenous fluid secretion resulting from crypt hyperplasia. Patients with more severe involvement with celiac disease may obtain temporary improvement with *dietary lactose and fat restriction* while awaiting the full effects of total gluten restriction, which is primary therapy.

ASSOCIATED DISEASES

Celiac disease is associated with dermatitis herpetiformis (DH), though the association has not been explained. Patients with DH have characteristic papulovesicular lesions that respond to dapsone. Almost all patients with DH have histologic changes in the small intestine consistent with celiac disease, although usually much milder and less diffuse in distribution. Most patients with DH have mild or no gastrointestinal symptoms. In contrast, relatively few patients with celiac disease have DH.

Celiac disease is also associated with diabetes mellitus type 1; IgA deficiency; Down syndrome; and Turner's syndrome. The clinical importance of the association with diabetes is that although severe watery diarrhea without evidence of malabsorption is most often diagnosed as "diabetic diarrhea" **(Chap. 344)**, assay of antiendomysial antibodies and/or a small-intestinal biopsy must be considered to exclude celiac disease.

COMPLICATIONS

The most important complication of celiac disease is the development of cancer. An increased incidence of both gastrointestinal and nongastrointestinal neoplasms as well as intestinal lymphoma exists in patients with celiac disease. For unexplained reasons the occurrence of lymphoma in patients with celiac disease is higher in Ireland and the United Kingdom than in the United States. The possibility of lymphoma must be considered whenever a patient with celiac disease previously doing well on a gluten-free diet is no longer responsive to gluten restriction or a patient who presents with clinical and histologic features consistent with celiac disease does not respond to a gluten-free diet. Other complications of celiac disease include the development of intestinal ulceration independent of lymphoma and so-called refractory sprue (see above) and collagenous sprue. In *collagenous sprue,* a layer of collagen-like material is present beneath the basement membrane; patients with collagenous sprue generally do not respond to a gluten-free diet and often have a poor prognosis.

Tropical Sprue

Tropical sprue is a poorly understood syndrome that affects both expatriates and natives in certain but not all tropical areas and is manifested by chronic diarrhea, steatorrhea, weight loss, and nutritional deficiencies, including those of both folate and cobalamin. This disease affects 5–10% of the population in some tropical areas.

Chronic diarrhea in a tropical environment is most often caused by infectious agents including *G. lamblia, Yersinia enterocolitica, C. difficile, Cryptosporidium parvum,* and *Cyclospora cayetanensis,* among other organisms. Tropical sprue should not be entertained as a possible diagnosis until the presence of cysts and trophozoites has been excluded in three stool samples.

Chronic infections of the gastrointestinal tract and diarrhea in patients with or without AIDS are discussed in Chaps. 128 and 189.

The small-intestinal mucosa in individuals living in tropical areas is not identical to that of individuals who reside in temperate climates. Biopsies reveal a mild alteration of villous architecture with a modest increase in mononuclear cells in the lamina propria, which on occasion can be as severe as that seen in celiac disease. These changes are observed both in native residents and in expatriates living in tropical regions and are usually associated with mild decreases in absorptive function, but they revert to "normal" when an individual moves or returns to a temperate area. Some have suggested that the changes seen in tropical enteropathy and in tropical sprue represent different ends of the spectrum of a single entity, but convincing evidence to support this concept is lacking.

ETIOLOGY

Because tropical sprue responds to antibiotics, the consensus is that it may be caused by one or more infectious agents. Nonetheless, the etiology and pathogenesis of tropical sprue are uncertain. First, its occurrence is not evenly distributed in all tropical areas; rather, it is found in specific locations, including southern India, the Philippines, and several Caribbean islands (e.g., Puerto Rico, Haiti), but is rarely observed in Africa, Jamaica, or Southeast Asia. Second, an occasional individual will not develop symptoms of tropical sprue until long after having left an endemic area. This is the reason why the original term for celiac disease (often referred to as celiac sprue) was *nontropical sprue* to distinguish it from tropical sprue. Third, multiple microorganisms have been identified on jejunal aspirate with relatively little consistency among studies. *Klebsiella pneumoniae,Enterobacter cloacae,* or *E. coli* have been implicated in some studies of tropical sprue, while other studies have favored a role for a toxin produced by one or more of these bacteria. Fourth, the incidence of tropical sprue appears to have decreased substantially during the past two or three decades, perhaps related to improved sanitation in many tropical countries during this time. One speculation for the reduced occurrence is the wider use of antibiotics in acute diarrhea, especially in travelers to tropical areas from temperate countries. Fifth, the role of folic acid deficiency in the pathogenesis of tropical sprue requires clarification. Folic acid is absorbed exclusively in the duodenum and proximal jejunum, and most patients with tropical sprue have evidence of folate malabsorption and depletion. Although folate deficiency can cause changes in small-intestinal mucosa that are corrected by folate replacement, several earlier studies reporting that tropical sprue could be cured by folic acid did not provide an explanation for the "insult" that was initially responsible for folate malabsorption.

The clinical pattern of tropical sprue varies in different areas of the world (e.g., India vs. Puerto Rico). Not infrequently, individuals in South India initially will report the occurrence of an acute enteritis before the development of steatorrhea and malabsorption. In contrast, in Puerto Rico a most insidious onset of symptoms and a more dramatic response to antibiotics is seen when compared to some other locations. Tropical sprue in different areas of the world may not be the same disease, and similar clinical entities may have different etiologies.

DIAGNOSIS

The diagnosis of tropical sprue is best made by the presence of an abnormal small-intestinal mucosal biopsy in an individual with chronic diarrhea and evidence of malabsorption who is either residing or has recently lived in a tropical country. The small-intestinal biopsy in tropical sprue does not have pathognomonic features but resembles, and can often be indistinguishable from, that seen in celiac disease (Fig. 294-4). The biopsy in tropical sprue will have less villous architectural alteration and more mononuclear cell infiltrate in the lamina propria. In contrast to celiac disease, the histologic features of tropical sprue are present with a similar degree of severity throughout the small intestine, and a gluten-free diet does not result in either clinical or histologic improvement in tropical sprue.

TREATMENT: TROPICAL SPRUE

Broad-spectrum antibiotics and folic acid are most often curative, especially if the patient leaves the tropical area and does not return. Tetracycline should be used for up to 6 months and may be associated with improvement within 1–2 weeks. Folic acid alone will induce a hematologicremission as well as improvement in appetite, weight gain, and some morphologic changes in small intestinal biopsy. Because of the presence of marked folate deficiency, folic acid is most often given together with antibiotics.

Short Bowel Syndrome

This is a descriptive term for the myriad clinical problems that occur following resection of varying lengths of small intestine; or on rare occasions may be congenital, e.g., microvillous inclusion disease. The factors that determine both the type and degree of symptoms include (1) the specific segment (jejunum vs. ileum) resected, (2) the length of the resected segment, (3) the integrity of the ileocecal valve, (4) whether any large intestine has also been removed, (5) residual disease in the remaining small and/or large intestine (e.g., Crohn's disease, mesenteric artery disease), and (6) the degree of adaptation in the remaining intestine. Short bowel syndrome can occur at any age from neonates through the elderly. *Intestinal failure* is the inability to maintain nutrition without parenteral support.

Three different situations in adults demand intestinal resections: (1) mesenteric vascular disease, including atherosclerosis, thrombotic phenomena, and vasculitides; (2) primary mucosal and submucosal disease, e.g., Crohn's disease; and (3) operations without preexisting small intestinal

disease, such as trauma.

Following resection of the small intestine, the residual intestine undergoes adaptation of both structure and function that may last for up to 6–12 months. Continued intake of dietary nutrients and calories is required to stimulate adaptation via direct contact with intestinal mucosa, the release of one or more intestinal hormones, and pancreatic and biliary secretions. Thus, enteral nutrition with calorie administration must be maintained, especially in the early postoperative period, even if an extensive intestinal resection requiring parenteral nutrition (PN) had been performed. The subsequent ability of such patients to absorb nutrients will not be known for several months, until adaptation is completed.

Multiple factors besides the absence of intestinal mucosa (required for lipid, fluid, and electrolyte absorption) contribute to the diarrhea and steatorrhea in these patients. Removal of the ileum and especially the ileocecal valve is often associated with more severe diarrhea than jejunal resection. Without part or all of the ileum, diarrhea can be caused by an increase in bile acids entering the colon, leading to their stimulation of colonic fluid and electrolyte secretion. Absence of the ileocecal valve is also associated with a decrease in intestinal transit time and bacterial overgrowth from the colon. The presence of the colon (or a major portion) is associated with substantially less diarrhea and lower likelihood of intestinal failure as a result of fermentation of nonabsorbed carbohydrates to SCFAs. The latter are absorbed in the colon and stimulate Na and water absorption, improving overall fluid balance. Lactose intolerance as a result of the removal of lactase-containing mucosa as well as gastric hypersecretion may also contribute to the diarrhea.

In addition to diarrhea and/or steatorrhea, a range of nonintestinal symptoms is also observed in some patients. A significant increase in renal calcium oxalate calculi is observed in patients with a small-intestinal resection with an intact colon and is due to an increase in oxalate absorption by the large intestine, with subsequent hyperoxaluria (called *enteric hyperoxaluria*). Two possible mechanisms for the increase in oxalate absorption in the colon have been suggested: (1) bile acids and fatty acids that increase colonic mucosal permeability, resulting in increased oxalate absorption; and (2) increased fatty acids that bind calcium, resulting in increased soluble oxalate that is then absorbed. Since oxalate is high in relatively few foods (e.g., spinach, rhubarb, tea), dietary restrictions alone are not adequate treatment. Cholestyramine, an anion-binding resin, and calcium have proved useful in reducing the hyperoxaluria. Similarly, an increase in cholesterol gallstones is related to a decrease in the bile acid pool size, which results in the generation of cholesterol supersaturation in gallbladder bile. Gastric hypersecretion of acid occurs in many patients following large resections of the small intestine. The etiology is unclear but may be related to either reduced hormonal inhibition of acid secretion or increased gastrin levels due to reduced small-intestinal catabolism of circulating gastrin. The resulting gastric acid secretion may be an important factor contributing to the diarrhea and steatorrhea. A reduced pH in the duodenum can inactivate pancreatic lipase and/or precipitate duodenal bile acids, thereby increasing steatorrhea, and an increase in gastric secretion can create a volume overload relative to the reduced small-intestinal absorptive capacity. Inhibition of gastric acid secretion with proton pump inhibitors can help in reducing the diarrhea and steatorrhea but only for the first six months.

TREATMENT: SHORT BOWEL SYNDROME

Treatment of short bowel syndrome depends on the severity of symptoms and whether the individual is able to maintain caloric and electrolyte balance with oral intake alone. Initial treatment includes judicious use of opiates (including codeine) to reduce stool output and to establish an effective diet. An initial diet should be low-fat and high-carbohydrate, if the colon is in situ, to minimize the diarrhea from fatty acid stimulation of colonic fluid secretion. MCTs (see above), a low-lactose diet, and various soluble fiber-containing diets should also be tried. In the absence of an ileocecal valve, the possibility of bacterial overgrowth must be considered and treated. If gastric acid hypersecretion is contributing to the diarrhea and steatorrhea, a proton pump inhibitor may be helpful. Usually none of these therapeutic approaches will provide an instant solution, but they can reduce disabling diarrhea.

The patient's vitamin and mineral status must also be monitored; replacement therapy should be initiated if indicated. Fat-soluble vitamins, folate, cobalamin, calcium, iron, magnesium, and zinc are the most critical factors to monitor on a regular basis. If these approaches are not successful, home PN is an established therapy that can be maintained for many years. Small intestinal transplantation is becoming established as a possible approach for individuals with extensive intestinal resection who cannot be maintained without PN, i.e., "intestinal failure." Considerable attention has been directed to the potential effectiveness of trophic hormones, e.g., glucagon-like peptide 2 (GLP-2), to improve absorptive function.

Bacterial Overgrowth Syndrome

Bacterial overgrowth syndrome comprises a group of disorders with diarrhea, steatorrhea, and macrocytic anemia whose common feature is the proliferation of colonic-type bacteria within the small intestine. This bacterial proliferation is due to stasis caused by impaired peristalsis (*functional stasis*), changes in intestinal anatomy (*anatomic stasis*), or direct communication between the small and large intestine. These conditions have also been referred to as *stagnant bowel syndrome* or *blind loop*

syndrome.

PATHOGENESIS

The manifestations of bacterial overgrowth syndromes are a direct consequence of the presence of increased amounts of a colonic-type bacterial flora, such as *E. coli* or *Bacteroides,* in the small intestine. *Macrocytic anemia* is due to cobalamin, not folate, deficiency. Most bacteria require cobalamin for growth, and increasing concentrations of bacteria use up the relatively small amounts of dietary cobalamin. *Steatorrhea* is due to impaired micelle formation as a consequence of a reduced intraduodenal concentration of conjugated bile acids and the presence of unconjugated bile acids. Certain bacteria, e.g., *Bacteroides,* deconjugate conjugated bile acids to unconjugated bile acids. Unconjugated bile acids will be absorbed more rapidly than conjugated bile acids, and, as a result, the intraduodenal concentration of bile acids will be reduced. In addition, the CMC of unconjugated bile acids is higher than that of conjugated bile acids, resulting in a decrease in micelle formation. *Diarrhea* is due, at least in part, to the steatorrhea, when it is present. However, some patients manifest diarrhea *without* steatorrhea, and it is assumed that the colonic-type bacteria in these patients are producing one or more bacterial enterotoxins that are responsible for fluid secretion and diarrhea.

ETIOLOGY

The etiology of these different disorders is bacterial proliferation in the small intestinal lumen secondary to either anatomic or functional stasis or to a communication between the relatively sterile small intestine and the colon with its high levels of aerobic and anaerobic bacteria. Several examples of *anatomic* stasis have been identified: (1) one or more diverticula (both duodenal and jejunal) (Fig. 294-3*C*); (2) fistulas and strictures related to Crohn's disease (Fig. 294-3*D*); (3) a proximal duodenal afferent loop following a subtotal gastrectomy and gastrojejunostomy; (4) a bypass of the intestine, e.g., jejunoileal bypass for obesity; and (5) dilation at the site of a previous intestinal anastomosis. These anatomic derangements are often associated with the presence of a segment (or segments) of intestine out of continuity of propagated peristalsis, resulting in stasis and bacterial proliferation. Bacterial overgrowth syndromes can also occur in the *absence* of an anatomic blind loop when *functional* stasis is present. Impaired peristalsis and bacterial overgrowth in the absence of a blind loop occur in scleroderma, where motility abnormalities exist in both the esophagus and small intestine **(Chap. 323)**. Functional stasis and bacterial overgrowth can also occur in association with diabetes mellitus and in the small intestine when a direct connection exists between the small and large intestine, including an ileocolonic resection, or occasionally following an enterocolic anastomosis that permits entry of bacteria into the small intestine as a result of bypassing the ileocecal valve.

DIAGNOSIS

The diagnosis may be suspected from the combination of a low serum cobalamin level and an elevated serum folate level, as enteric bacteria frequently produce folate compounds that will be absorbed in the duodenum. Ideally, the diagnosis of the bacterial overgrowth syndrome is the demonstration of increased levels of aerobic and/or anaerobic colonic-type bacteria in a jejunal aspirate obtained by intubation. This specialized test is rarely available. Breath hydrogen testing with lactulose (a nondigestible disaccharide) administration has also been used to detect bacterial overgrowth. The Schilling test can also diagnose bacterial overgrowth (see supplementary material) but is also not available routinely. Often the diagnosis is suspected clinically and confirmed by response to treatment.

TREATMENT: BACTERIAL OVERGROWTH SYNDROME

Primary treatment should be directed, if at all possible, to the surgical correction of an anatomic blind loop. In the absence of functional stasis, it is important to define the anatomic relationships responsible for stasis and bacterial overgrowth. For example, bacterial overgrowth secondary to strictures, one or more diverticula, or a proximal afferent loop can potentially be cured by surgical correction of the anatomic state. In contrast, the functional stasis of scleroderma or certain anatomic stasis states (e.g., multiple jejunal diverticula) cannot be corrected surgically, and these conditions should be treated with broad-spectrum antibiotics. Tetracycline used to be the initial treatment of choice; due to increasing resistance, however, other antibiotics such as metronidazole, amoxicillin/clavulanic acid, and cephalosporins have been employed. The antibiotic should be given for approximately 3 weeks or until symptoms remit. Although the natural history of these conditions is chronic, antibiotics should not be given continuously. Symptoms usually remit within 2–3 weeks of initial antibiotic therapy. Therapy need not be repeated until symptoms recur. In the presence of frequent recurrences, several treatment strategies exist, but the use of antibiotics for 1 week per month, whether or not symptoms are present, is often most effective.

Unfortunately, therapy for bacterial overgrowth syndrome is largely empirical, with an absence of clinical trials on which to base rational decisions regarding the antibiotic choice, the duration of treatment, and/or the best approach for treating recurrences. Bacterial overgrowth may also occur as a component of another chronic disease, e.g., Crohn's disease, radiation enteritis, or short bowel syndrome. Treatment of the bacterial overgrowth in these settings will not cure the underlying problem but may be very important in ameliorating a subset of clinical problems that are related to bacterial

overgrowth.

Whipple's Disease

Whipple's disease is a chronic multisystem disease associated with diarrhea, steatorrhea, weight loss, arthralgia, and central nervous system (CNS) and cardiac problems; it is caused by the bacteria *Tropheryma whipplei.* Until the identification of *T. whipplei* by polymerase chain reaction, the hallmark of Whipple's disease had been the presence of PAS-positive macrophages in the small intestine (Fig. 294-4*E*) and other organs with evidence of disease.

ETIOLOGY

Whipple's disease is caused by a small gram-positive bacillus, *T. whipplei.* The bacillus, an Actinobacteria, has low virulence but high infectivity, and relatively minimal symptoms are observed compared to the extent of the bacilli in multiple tissues.

CLINICAL PRESENTATION

The onset of Whipple's disease is insidious and is characterized by diarrhea, steatorrhea, abdominal pain, weight loss, migratory large-joint arthropathy, and fever as well as ophthalmologic and CNS symptoms. The development of dementia is a relatively late symptom and an extremely poor prognostic sign, especially in patients who relapse following the induction of a remission with antibiotics. For unexplained reasons, the disease occurs primarily in middle-aged white men. The steatorrhea in these patients is generally believed secondary to both small-intestinal mucosal injury and lymphatic obstruction secondary to the increased number of PAS-positive macrophages in the lamina propria of the small intestine.

DIAGNOSIS

The diagnosis of Whipple's disease is suggested by a multisystem disease in a patient with diarrhea and steatorrhea. Obtaining tissue biopsies from the small intestine and/or other organs that may be involved (e.g., liver, lymph nodes, heart, eyes, CNS, or synovial membranes), based on the patient's symptoms, is the primary approach to establish the diagnosis of Whipple's disease. The presence of PAS-positive macrophages containing the characteristic small (0.25—1–2 mm) bacilli is suggestive of this diagnosis. However, Whipple's disease can be confused with the PAS-positive macrophages containing *M. avium* complex, which may be a cause of diarrhea in AIDS. The presence of the *T. whipplei* bacillus outside of macrophages is a more important indicator of active disease than is their presence within the macrophages. *T. whipplei* has now been successfully grown in culture.

TREATMENT: WHIPPLE'S DISEASE

The treatment for Whipple's disease is prolonged use of antibiotics. The current drug of choice is double-strength trimethoprim/sulfamethoxazole for approximately 1 year. PAS-positive macrophages can persist following successful treatment, and the presence of bacilli outside of macrophages is indicative of persistent infection or an early sign of recurrence. Recurrence of disease activity, especially with dementia, is an extremely poor prognostic sign and requires an antibiotic that crosses the blood-brain barrier. If trimethoprim/sulfamethoxazole is not tolerated, chloramphenicol is an appropriate second choice.

Protein-Losing Enteropathy

Protein-losing enteropathy is not a specific disease but rather a group of gastrointestinal and nongastrointestinal disorders with hypoproteinemia and edema in the absence of either proteinuria or defects in protein synthesis, e.g., chronic liver disease. These diseases are characterized by excess protein loss into the gastrointestinal tract. Normally, about 10% of total protein catabolism occurs via the gastrointestinal tract. Evidence of increased protein loss into the gastrointestinal tract occurs in more than 65 different diseases, which can be classified into three groups: (1) mucosal ulceration, such that the protein loss primarily represents exudation across damaged mucosa, e.g., ulcerative colitis, gastrointestinal carcinomas, and peptic ulcer; (2) nonulcerated mucosa, but with evidence of mucosal damage so that the protein loss represents loss across epithelia with altered permeability, e.g., celiac disease and Ménétrier's disease in the small intestine and stomach, respectively; and (3) lymphatic dysfunction, representing either primary lymphatic disease or secondary to partial lymphatic obstruction that may occur as a result of enlarged lymph nodes or cardiac disease.

DIAGNOSIS

The diagnosis of protein-losing enteropathy is suggested by the presence of peripheral edema and low serum albumin and globulin levels in the absence of renal and hepatic disease. An individual with protein-losing enteropathy only rarely has selective loss of *only* albumin or *only* globulins. Therefore, marked reduction of serum albumin with normal serum globulins should not initiate an evaluation for protein-losing enteropathy but should suggest the presence of renal and/or hepatic disease. Likewise, reduced serum globulins with normal serum albumin levels are more likely a result of reduced globulin synthesis rather than enhanced globulin loss into the intestine. Documentation of an increase in protein loss into the gastrointestinal tract has been established by the administration of one of several radiolabeled proteins and its quantitation in stool during a 24- or 48-h period. Unfortunately, none of

these radiolabeled proteins is available for routine clinical use. α_1-Antitrypsin, a protein that accounts for ~4% of total serum proteins and is resistant to proteolysis, can be used to document enhanced rates of serum protein loss into the intestinal tract but cannot be used to assess gastric protein loss due to its degradation in an acid milieu. α_1-Antitrypsin clearance is measured by determining stool volume and both stool and plasma α_1-antitrypsin concentrations. In addition to the loss of protein via abnormal and distended lymphatics, peripheral lymphocytes may also be lost via lymphatics, resulting in a relative lymphopenia. Thus, the presence of lymphopenia in a patient with hypoproteinemia supports the presence of increased loss of protein into the gastrointestinal tract.

Patients with increased protein loss into the gastrointestinal tract from lymphatic obstruction often have steatorrhea and diarrhea. The steatorrhea is a result of altered lymphatic flow as lipid-containing chylomicrons exit from intestinal epithelial cells via intestinal lymphatics (Table 294-4; Fig. 294-4). In the absence of mechanical or anatomic lymphatic obstruction, intrinsic intestinal lymphatic dysfunction, with or without lymphatic dysfunction in the peripheral extremities, has been named *intestinal lymphangiectasia.* Similarly, about 50% of individuals with intrinsic peripheral lymphatic disease (Milroy's disease) will also have intestinal lymphangiectasia and hypoproteinemia. Other than steatorrhea and enhanced protein loss into the gastrointestinal tract, all other aspects of intestinal absorptive function are normal in intestinal lymphangiectasia.

OTHER CAUSES

Patients who appear to have idiopathic protein-losing enteropathy without any evidence of gastrointestinal disease should be examined for cardiac disease—especially right-sided valvular disease and chronic pericarditis **(Chaps. 237 and 239)**. On occasion, hypoproteinemia can be the only presentation for these two types of heart disease. Ménétrier's disease (also called *hypertrophic gastropathy*) is an uncommon entity that involves the body and fundus of the stomach and is characterized by large gastric folds, reduced gastric acid secretion, and, at times, enhanced protein loss into the stomach.

TREATMENT: PROTEIN-LOSING ENTEROPATHY

As excess protein loss into the gastrointestinal tract is most often secondary to a specific disease, treatment should be directed primarily to the underlying disease process and not to the hypoproteinemia. For example, if significant hypoproteinemia with resulting peripheral edema is secondary to either celiac disease or ulcerative colitis, a gluten-free diet or mesalamine, respectively, would be the initial therapy. When enhanced protein loss is secondary to lymphatic obstruction, it is critical to establish the nature of this obstruction. Identification of mesenteric nodes or lymphoma may be possible by imaging studies. Similarly, it is important to exclude cardiac disease as a cause of protein-losing enteropathy either by echosonography or, on occasion, by a right-heart catheterization.

The increased protein loss that occurs in intestinal lymphangiectasia is a result of distended lymphatics associated with lipid malabsorption. Treatment of the hypoproteinemia is accomplished by a low-fat diet and the administration of MCTs (Table 294-3), which do not exit from the intestinal epithelial cells via lymphatics but are delivered to the body via the portal vein.

SUMMARY

A pathophysiologic classification of the many conditions that can produce malabsorption is given in **Table 294-8**. A summary of the pathophysiology of the various clinical manifestations of malabsorption is given in **Table 294-9**.

Table 294-8 Classification of Malabsorption Syndromes

Inadequate digestion
Postgastrectomy[a]
Deficiency or inactivation of pancreatic lipase
Exocrine pancreatic insufficiency
Chronic pancreatitis
Pancreatic carcinoma Cystic fibrosis
Pancreatic insufficiency—congenital or acquired
Gastrinoma—acid inactivation of lipase[a]
Drugs—orlistat
Reduced intraduodenal bile acid concentration/impaired micelle formation
Liver disease
Parenchymal liver disease

Cholestatic liver disease
Bacterial overgrowth in small intestine:
Anatomic stasis
Afferent loop Stasis/blind
Loop/strictures/fistulae
Functional stasis
Diabetes[a]
Scleroderma[a]
Intestinal pseudoobstruction
Interrupted enterohepatic circulation of bile salts
Ileal resection
Crohn's disease[a]
Drugs (bind or precipitate bile salts)—neomycin, cholestyramine, calcium carbonate
Impaired mucosal absorption/mucosal loss or defect
Intestinal resection or bypass[a]
Inflammation, infiltration, or infection:
Crohn's disease[a]
Amyloidosis
Scleroderma[a]
Lymphoma[a]
Eosinophilic enteritis
Mastocytosis
Tropical sprue Celiac disease
Collagenous sprue
Whipple's disease[a]
Radiation enteritis[a]
Folate and vitamin B12 deficiency
Infections—giardiasis
Graft versus host disease
Genetic disorders
Disaccharidase deficiency
Agammaglobulinemia
Abetalipoproteinemia
Hartnup's disease Cystinuria
Impaired nutrient delivery to and/or from intestine:
Lymphatic obstruction
Lymphoma[a]
Lymphangiectasia
Circulatory disorders
Congestive heart failure
Constrictive pericarditis
Mesenteric artery atherosclerosis
Vasculitis
Endocrine and metabolic disorders
Diabetes[a]
Hypoparathyroidism
Adrenal insufficiency

Hyperthyroidism
Carcinoid syndrome

Table 294-9 Pathophysiology of Clinical Manifestations of Malabsorption Disorders

Symptom or Sign	Mechanism
Weight loss/malnutrition	Anorexia, malabsorption of nutrients
Diarrhea	Impaired absorption or secretion of water and electrolytes; colonic fluid secretion secondary to unabsorbed dihydroxy bile acids and fatty acids
Flatus	Bacterial fermentation of unabsorbed carbohydrate
Glossitis, cheilosis, stomatitis	Deficiency of iron, vitamin B12, folate, and vitamin A
Abdominal pain	Bowel distention or inflammation, pancreatitis
Bone pain	Calcium, vitamin D malabsorption, protein deficiency, osteoporosis
Tetany, paresthesia	Calcium and magnesium malabsorption
Weakness	Anemia, electrolyte depletion (particularly K^+)
Azotemia, hypotension	Fluid and electrolyte depletion
Amenorrhea, decreased libido	Protein depletion, decreased calories, secondary hypopituitarism
Anemia	Impaired absorption of iron, folate, vitamin B12
Bleeding	Vitamin K malabsorption, hypoprothrombinemia
Night blindness/xerophthalmia	Vitamin A malabsorption
Peripheral neuropathy	Vitamin B12 and thiamine deficiency
Dermatitis	Deficiency of vitamin A, zinc, and essential fatty acid

[a]Malabsorption caused by more than one mechanism.

FURTHER READINGS

American Gastroenterological Association: AGA technical review on the evaluation and management of chronic diarrhea. Gastroenterology 116:1464, 1999

Binder HJ: Role of colonic short-chain fatty acid transport in diarrhea. Annu Rev Physiol 72:297, 2010[PMID: 20148677] [Full Text]

Buchman AL: Etiology and initial management of short bowel syndrome. Gastroenterology 130:S5, 2006

Fenoller F et al: Whipple's disease. N Engl J Med 356:55, 2007

Green PH, Cellier C: Celiac disease. N Engl J Med 357:1731, 2007[PMID: 17960014] [Full Text]

Hammer HF: Pancreatic exocrine insufficiency: Diagnostic evaluation and replacement therapy with pancreatic enzymes. Dig Dis 28:339, 2010[PMID: 20814209] [Full Text]

Kagnoff MF: Celiac disease: Pathogenesis of a model immunogenetic disease. J Clin Invest 117:41, 2007[PMID: 17200705] [Full Text]

Schulzke JD et al: Disorders of intestinal secretion and absorption. Best Pract Res Clin Gastroenterol 23:395, 2009[PMID: 19505667] [Full Text]

Suchy FJ et al: National Institutes of Health Consensus Development Conference: Lactose intolerance and health. Ann Intern Med 152:792, 2010[PMID: 20404261] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE SCHILLING TEST: INTRODUCTION

The Schilling test is performed to determine the cause for cobalamin malabsorption. Unfortunately, this test has not been available commercially in the United States for the last few years. Since understanding the physiology and pathophysiology of cobalamin absorption is very valuable for enhancing one's understanding of aspects of gastric, pancreatic, and ileal function, discussion of the Schilling test is provided as supplemental information to Chap. 294. Since cobalamin absorption requires multiple steps, including gastric, pancreatic, and ileal processes, the Schilling test also can be used to assess the integrity of those other organs (Chap. 105). Cobalamin is present primarily in meat. Except in strict vegans, *dietary* cobalamin deficiency is exceedingly uncommon. Dietary cobalamin is bound in the stomach to a glycoprotein called *R-binder protein*, which is synthesized in both the stomach and the salivary glands. This cobalamin–R binder complex is formed in the acid milieu of the stomach. Cobalamin absorption has an absolute requirement for intrinsic factor, another glycoprotein synthesized and released by gastric parietal cells, to promote its uptake by specific cobalamin receptors on the brush border of ileal enterocytes. Pancreatic protease enzymes split the cobalamin–R binder complex to release cobalamin in the proximal small intestine, where cobalamin then is bound by intrinsic factor.

As a consequence, cobalamin absorption may be abnormal in the following:

1. *Pernicious anemia*, a disease in which immunologically mediated atrophy of gastric parietal cells leads to an absence of both gastric acid and intrinsic factor secretion.
2. *Chronic pancreatitis* as a result of deficiency of pancreatic proteases to split the cobalamin–R binder complex. Although 50% of patients with chronic pancreatitis have been reported to have an abnormal Schilling test that was corrected by pancreatic enzyme replacement, the presence of a cobalamin-responsive macrocytic anemia in chronic pancreatitis is extremely rare. Although this probably reflects a difference in the digestion/absorption of cobalamin in food versus that in a crystalline form, the Schilling test still can be used to assess pancreatic exocrine function.
3. *Achlorhydria*, or absence of another factor secreted with acid that is responsible for splitting cobalamin away from the proteins in food to which it is bound. Up to one-third of individuals>60 years of age have marginal vitamin B12 absorption because of the inability to release cobalamin from food; these people have no defects in absorbing crystalline vitamin B_{12}.
4. *Bacterial overgrowth syndromes*, which are most often secondary to stasis in the small intestine, leading to bacterial utilization of cobalamin (often referred to as *stagnant bowel syndrome*; see below).
5. *Ileal dysfunction* (as a result of either inflammation or prior intestinal resection) due to impaired function of the mechanism of cobalamin–intrinsic factor uptake by ileal intestinal epithelial cells.

The Schilling test is performed by administering ^{58}Co-labeled cobalamin orally and collecting urine for 24 h, and it is dependent on normal renal and bladder function. Urinary excretion of cobalamin will reflect cobalamin absorption provided that intrahepatic binding sites for cobalamin are fully occupied. To ensure saturation of hepatic cobalamin binding sites so that all absorbed radiolabeled cobalamin will be excreted in urine, 1 mg of cobalamin is administered intramuscularly 1 h after ingestion of the radiolabeled cobalamin. The Schilling test may be abnormal (usually defined as <10% excretion in 24 h) in pernicious anemia, chronic pancreatitis, blind loop syndrome, and ileal disease **(Table e37-1)**.

Therefore, whenever an abnormal Schilling test is found, ^{58}Co-labeled cobalamin should be administered on another occasion bound to intrinsic factor, with pancreatic enzymes, or after a 5-day course of antibiotics (often tetracycline). A variation of the Schilling test can detect failure to split cobalamin from food proteins. The labeled cobalamin is cooked together with a scrambled egg and administered orally. People with achlorydria will excrete <10% of the labeled cobalamin in the urine. In addition to establishing the etiology for cobalamin deficiency, the Schilling test can be used to help delineate the pathologic process responsible for steatorrhea by assessing ileal, pancreatic, and small-intestinal luminal function. Unfortunately, the Schilling test is performed infrequently because of the unavailability of human intrinsic factor.

Table e37-1 Differential Results of Schilling Test in Several Diseases Associated with Cobalamin (Cbl) Malabsorption

	^{58}Co-Cbl	With Intrinsic Factor	With Pancreatic Enzymes	After 5 Days of Antibiotics
Pernicious anemia	Reduced	Normal	Reduced	Reduced
Chronic pancreatitis	Reduced	Reduced	Normal	Reduced
Bacterialovergrowth	Reduced	Reduced	Reduced	Normal
Ileal disease	Reduced	Reduced	Reduced	Reduced

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFLAMMATORY BOWEL DISEASE: INTRODUCTION

Inflammatory bowel disease (IBD) is an immune-mediated chronic intestinal condition. Ulcerative colitis (UC) and Crohn's disease (CD) are the two major types of IBD.

EPIDEMIOLOGY

The incidence of IBD varies within different geographic areas. CD and UC both occur at the highest incidence in Europe, the United Kingdom, and North America. In North America, incidence rates range from 2.2–14.3 cases per 100,000 person-years for UC and from 3.1–14.6 cases per 100,000 person-years for CD **(Table 295-1)**. Prevalence ranges from 37–246 cases per 100,000 person-years for UC and from 26–199 cases per 100,000 person-years for CD. In Europe, incidence ranges from 1.5–20.3 cases per 100,000 person-years for UC and from 0.7–9.8 cases for CD; prevalence ranges from 21.4–243 cases for UC and from 8.3–214 cases per 100,000 person-years for CD. IBD has been rare in other areas except Israel, Australia, and South Africa. The incidence of IBD, especially UC, is rising in Japan, South Korea, Singapore, northern India, and Latin America, areas previously thought to have low incidence. The incidence of UC has increased sixfold in the past two decades in Hong Kong. Reports from the United States, Poland, Denmark, and South Korea indicate that the incidence of pediatric IBD is increasing rapidly as well. The highest mortality is during the first years of disease and in long-duration disease due to the risk of colon cancer. In a Danish population study, the standardized mortality ratios for CD and UC were 1.31 and 1.1, respectively.

Table 295-1 Epidemiology of IBD

	Ulcerative Colitis	Crohn's Disease
Incidence (North America) per person-years	2.2–14.3:100,000	3.1–14.6:100,000
Age of onset	15–30 & 60–80	15–30 & 60–80
Ethnicity	Jewish > non-Jewish white > African American > Hispanic > Asian	
Male/female ratio	1:1	1.1–1.8:1
Smoking	May prevent disease	May cause disease
Oral contraceptives	No increased risk	Odds ratio 1.4
Appendectomy	Protective	Not protective
Monozygotic twins	6% concordance	58% concordance
Dizygotic twins	0% concordance	4% concordance

The peak age of onset of UC and CD is between 15 and 30 years. A second peak occurs between the ages of 60 and 80. The male to female ratio for UC is 1:1 and for CD is 1.1–1.8:1. UC and CD have two- to fourfold increased frequency in Jewish populations in the United States, Europe, and South Africa. Furthermore, disease frequency differs within the Jewish populations. The prevalence of IBD in Ashkenazi Jews is about twice that of Israeli-born, Sephardic, or Asian Jews. The prevalence decreases progressively in non-Jewish white, African-American, Hispanic, and Asian populations. Urban areas have a higher prevalence of IBD than rural areas, and high socioeconomic classes have a higher prevalence than lower socioeconomic classes.

The effects of cigarette smoking are different in UC and CD. The risk of UC in smokers is 40% that of nonsmokers. Additionally, former smokers have a 1.7-fold increased risk for UC than people who have never smoked. In contrast, smoking is associated with a twofold increased risk of CD. Oral contraceptives are also linked to CD; the odds ratio of CD for oral contraceptive users is about 1.4. Appendectomy is protective against UC but is associated with an increased risk of CD. This elevated risk in CD is observed early after an appendectomy, which is diminished thereafter, making it likely that it reflects diagnostic problems in patients with incipient CD.

IBD is a familial disease in 5–10% of patients. Some of these patients may exhibit early onset disease during the first decade of life and, in CD, a concordance of anatomic site and clinical type within families. In the remainder of patients, IBD is observed in the absence of a family history (i.e., sporadic disease). If a patient has IBD, the lifetime risk that a first-degree relative will be affected is ~10%. If two parents have IBD, each child has a 36% chance of being affected. In twin studies, 58% of monozygotic twins are concordant for CD and 6% are concordant for UC, whereas 4% of dizygotic twins are concordant for CD and none are concordant for UC. In a recent twin study from Germany, the relative risk of a monozygotic twin developing Crohn's disease if his or her twin was affected was 738. The risks of developing IBD are higher in first-degree relatives of Jewish versus non-Jewish patients: 7.8% versus 5.2% for CD and 4.5% versus 1.6% for UC.

Additional evidence for genetic predisposition to IBD comes from its association with certain genetic

syndromes. UC and CD are both associated with Turner's syndrome, and Hermansky-Pudlak syndrome is associated with granulomatous colitis. Glycogen storage disease type 1b can present with Crohn's-like lesions of the large and small bowel. Severe immunodeficiency disorders such as Wiskott-Aldrich syndrome and chronic granulomatous disease are associated with IBD. Immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome is associated with a severe enteropathy and autoimmunity **(Table 295-2)**. Other immunodeficiency disorders such as hypogammaglobulinemia, selective IgA deficiency, and hereditary angioedema, also exhibit an increased association with IBD.

Table 295-2 Primary Genetic Disorders Associated with Ibd

Name	Genetic Association	Phenotype
Turner's syndrome	Loss of part or all of X chromosome	Associated with UC and colonic CD
Hermansky-Pudlak	Autosomal recessive chromosome 10q23	Granulomatous colitis, oculocutaneous albinism, platelet dysfunction, pulmonary fibrosis
Wiskott-Aldrich syndrome (WAS)	X-linked recessive disorder, loss of WAS protein function	Colitis, immunodeficiency, severely dysfunctional platelets, and thrombocytopenia
Glycogen Storage disease	Deficiency of the glucose-6-phosphate transport protein type B1	Granulomatous colitis, presents in infancy with hypoglycemia, growth failure, hepatomegaly, and neutropenia
Immune dysregulation polyendocrinopathy, enteropathy X-linked (IPEX)	Loss of FoxP3 transcription factor and T regulatory cell function	UC-like autoimmune enteropathy, with endocrinopathy (neonatal type 1 diabetes or thyroiditis), dermatitis
Early onset IBD	Deficient IL-10 receptor function	Severe, refractory IBD in early life

Abbreviations: CD, Crohn's disease; IBD, inflammatory bowel disease; IL, interleukin; UC, ulcerative colitis; WASP, Wiskott-Aldrich syndrome protein.

ETIOLOGY AND PATHOGENESIS

A consensus hypothesis is that in genetically predisposed individuals, both exogenous factors (e.g., composition of normal intestinal microbiota) and endogenous host factors (e.g., intestinal epithelial cell barrier function, innate and adaptive immune function) interact to cause a chronic state of dysregulated mucosal immune function that is further modified by specific environmental factors (e.g., smoking, enteropathogens). Although chronic activation of the mucosal immune system may represent an appropriate response to an unidentified infectious agent, a search for such an agent has thus far been unrewarding in IBD. As such, IBD is currently considered an inappropriate immune response to the endogenous commensal microbiota within the intestines, with or without some component of autoimmunity. Importantly, the normal intestines contain a large number of immune cells in a chronic state of so-called physiologic inflammation, in which the gut is restrained from full immunologic responses to the commensal microbiota and dietary antigens by very powerful regulatory pathways that function within the immune system (e.g., $FoxP3^+$ T regulatory cells). During the course of infections in the normal host, full activation of the gut-associated lymphoid tissues occurs but is rapidly superseded by dampening of the immune response and tissue repair. In IBD this process may not be regulated normally.

GENETIC CONSIDERATIONS

IBD is a polygenic disorder that gives rise to multiple clinical subgroups within UC and CD. A variety of genetic approaches including candidate gene studies, linkage analysis and genome-wide association studies that focus on the identification of disease-associated, single-nucleotide polymorphisms (SNP) within the human genome have identified Approximately 100 disease-associated loci on many different chromosomes **(Table 295-3)**. About one-third of these genetic risk factors are shared between CD and UC accounting for the overlapping immunopathogenesis and consequently epidemiologic observations of both diseases in the same families and similarities in response to therapies. Because the specific causal variants for each gene or locus are largely unknown, it is not clear whether the similarities in the genetic risk factors associated with CD and UC that are observed are shared at structural or functional levels. Similarly, many of the genetic risk factors identified are also observed to be associated with risk for other immune-mediated diseases suggesting that related immunogenetic pathways are involved in the pathogenesis of multiple different disorders accounting for the common responsiveness to similar types of biologic therapies (e.g., anti-tumor necrosis factor therapies) and possibly the simultaneous occurrence of these disorders. The diseases and the genetic risk factors that are shared with IBD include rheumatoid arthritis (*TNFAIP3*), psoriasis (*IL23R,IL12B*), ankylosing spondylitis (*IL23R*), type 1 diabetes mellitus (*IL10,PTPN2*), asthma (*ORMDL3*), and systemic lupus erythematosus (*TNFAIP3,IL10*).

Table 295-3 Genetic Loci Associated with CD and/or UC

Chr	Putative Gene	Gene Name	Protein Function	CD	UC
Innate Immunity and Autophagy					

1q23	*ITLN1*	Intelectin 1	Bacterial binding	+	
2q37	*ATG16L1*	ATG16 autophagy related 16-like 1	Autophagy	+	
5q33	*IRGM*	Immunity-related GTPase family, M	Autophagy	+	
9p24	*JAK2*	Janus kinase 2	IL-6R & IL-23R signaling	+	+
12q12	*LRRK2*	Leucine-rich repeat kinase 2	Autophagy ?	+	
16q12	*NOD2*	Nucleotide-binding oligomerization domain containing 2	Bacterial sensing	+	
17q21	*STAT3*	Signal transducer and activator of transcription 3	IL-6R, IL-23R & IL-10R signaling	+	+
ER Stress and Metabolism					
5q31	*SLC22A5*	Solute carrier family 22, member 5	ß carnitine transporter	+	
7p21	*AGR2*	Anterior gradient 2	ER stress	+	+
17q21	*ORMDL3*	Orosomucoid related member 1-like 3	ER stress	+	+
22q12	*XBP1*	X-box binding protein 1	ER stress	+	+
Adaptive Immunity					
1p31	*IL23R*	Interleukin 23 receptor	Th17 cell stimulation	+	+
1q32	*IL10*	Interleukin-10	Treg associated cytokine		+
5q33	*IL12B*	Interleukin 12B	IL-12 p40 chain of IL-12/IL-23	+	+
18p11	*PTPN2*	Protein tyrosine phosphatase, nonreceptor type 2	T cell regulation	+	
Inflammation					
3p21	*MST1*	Macrophage stimulating 1	Macrophage activation	+	+
5p13	*PTGER4*	Prostaglandin E receptor 4	PGE2 receptor	+	+
6q23	*TNFAIP3*	Tumor necrosis factor, alpha-induced protein 3 (A20)	Toll-like receptor regulation	+	
6q27	*CCR6*	Chemokine (C-C motif) receptor 6	Dendritic cell migration	+	

Abbreviations: CD, Crohn's disease; ER, endoplasmic reticulum; GTPase, guanosine triphosphatase; IL, interleukin; UC, ulcerative colitis.

Source: Adapted from Kaser et al, Ann Rev Immunol 2010

The genetic factors defined to date that are recognized to mediate risk for IBD have highlighted the importance of several common mechanisms of disease (Table 295-3). These include the following: Those genes that are associated with innate immunity and autophagy (e.g., *NOD2,ATG16L1,IRGM, JAK2, STAT3*) that function in innate immune cells (both parenchymal and hematopoietic) to respond to and clear bacteria, mycobacteria and viruses; those that are associated with endoplasmic reticulum (ER) and metabolic stress (e.g., *XBP1,ORMDL3,OCTN*), which serve to regulate the secretory activity of cells involved in responses to the commensal microbiota such as Paneth and goblet cells and the manner in which intestinal cells respond to the metabolic products of bacteria; those that are associated with the regulation of adaptive immunity (e.g., *IL23R,IL12B,IL10,PTPN2*), which regulate the balance between inflammatory and regulatory cytokines; and, finally, those that are involved in the development and resolution of inflammation (e.g., *MST1,CCR6, TNFAIP3, PTGER4*) and ultimately leukocyte recruitment and inflammatory mediator production. Some of these loci are associated with specific subtypes of disease such as the association between *NOD2* polymorphisms and fibrostenosing CD, especially within the ileum. However, the clinical utility of these genetic risk factors for the diagnosis or determination of prognosis and therapeutic responses remains to be defined.

Defective Immune Regulation in Ibd

The mucosal immune system is normally unreactive to luminal contents due to oral (mucosal) tolerance. When soluble antigens are administered orally rather than subcutaneously or intramuscularly, antigen-specific nonresponsiveness is induced. Multiple mechanisms are involved in the induction of oral tolerance and include deletion or anergy of antigen-reactive T cells or induction of $CD4^+$ T cells that suppress gut inflammation (e.g., T regulatory cells expressing the FoxP3 transcription factor) that secrete anti-inflammatory cytokines such as interleukin (IL) 10 and transforming growth factor ß (TGF-ß). Oral tolerance may be responsible for the lack of immune responsiveness to dietary antigens and the commensal microbiota in the intestinal lumen. In IBD this suppression of inflammation is altered, leading to uncontrolled inflammation. The mechanisms of this regulated immune suppression are incompletely known.

Gene knockout ($^{-/-}$) or transgenic (Tg) mouse models of IBD have revealed that deleting specific cytokines (e.g., IL-2, IL-10, TGF-ß) or their receptors, deleting molecules associated with T cell antigen recognition (e.g., T cell antigen receptors) or interfering with intestinal epithelial cell barrier function and the regulation of responses to commensal bacteria (e.g., XBP1, N-cadherin, mucus glycoprotein or $NF_{\kappa}B$) leads to spontaneous colitis or enteritis. In the majority of circumstances, intestinal inflammation in these animal models requires the presence of the commensal microbiota. Thus, a variety of specific alterations can lead to immune activation by commensal microbiota and inflammation directed at the intestines in mice. How these relate to human IBD

remains to be defined but are consistent with inappropriate responses of the genetically susceptible host to the commensal bacteria.

In both UC and CD, an inflammatory pathway thus likely emerges from the genetic predisposition that is associated with inappropriate innate immune sensing and reactivity to commensal bacteria together with inadequate regulatory pathways that lead to activated $CD4^+$ T cells in the lamina propria that secrete excessive quantities of inflammatory cytokines relative to anti-inflammatory cytokines. Some cytokines activate other inflammatory cells (macrophages and B cells) and others act indirectly to recruit other lymphocytes, inflammatory leukocytes, and mononuclear cells from the bloodstream into the gut through interactions between homing receptors on leukocytes (e.g., $\alpha^4\beta7$ integrin) and addressins on vascular endothelium (e.g., MadCAM1). $CD4^+$ T helper (T_H) cells that promote inflammation are of three major types, all of which may be associated with colitis in animal models and perhaps humans: T_H1 cells [secrete interferon (IFN) gamma;], T_H2 cells (secrete IL-4, IL-5, IL-13), and T_H17 cells (secrete IL-17, IL-21). T_H1 cells induce transmural granulomatous inflammation that resembles CD, T_H2 cells, and related natural killer T cells that secrete IL-13 induce superficial mucosal inflammation resembling UC, and T_H17 cells may be responsible for neutrophilic recruitment. Each of these T cell subsets cross-regulate each other. The T_H1 cytokine pathway is initiated by IL-12, a key cytokine in the pathogenesis of experimental models of mucosal inflammation. IL-4 and IL-23, together with IL-6 and TGF-β, induce T_H2 and T_H17 cells, respectively. Activated macrophages secrete tumor necrosis factor (TNF and IL-6). Thus, use of antibodies to block proinflammatory cytokines (e.g., anti-TNF, anti-IL-12, anti-IL-23, anti-IL-6, anti-IFN-gamma;) or molecules associated with leukocyte recruitment (e.g., anti-$\alpha^4\beta7$) or use of cytokines that inhibit inflammation and promote regulatory T cells (e.g., IL-10) or promote intestinal barrier function may be beneficial to humans with intestinal inflammation.

The Inflammatory Cascade in Ibd

Once initiated in IBD by abnormal innate immune sensing of bacteria by parenchymal cells (e.g., intestinal epithelial cells) and hematopoietic cells (e.g., dendritic cells), the immune inflammatory response is perpetuated by T-cell activation. A sequential cascade of inflammatory mediators extends the response; each step is a potential target for therapy. Inflammatory cytokines such as IL-1, IL-6, and TNF, have diverse effects on tissues. They promote fibrogenesis, collagen production, activation of tissue metalloproteinases, and the production of other inflammatory mediators; they also activate the coagulation cascade in local blood vessels (e.g., increased production of von Willebrand's factor). These cytokines are normally produced in response to infection but are usually turned off or inhibited at the appropriate time to limit tissue damage. In IBD their activity is not regulated, resulting in an imbalance between the proinflammatory and anti-inflammatory mediators. Therapies such as the 5-aminosalicylic acid (5-ASA) compounds are potent inhibitors of these inflammatory mediators through inhibition of transcription factors such as $NF_\kappa B$ that regulate their expression.

Exogenous Factors

IBD may have an as yet undefined infectious etiology. Observational studies suggest that multiple pathogens (e.g., *Salmonella*, *Shigella*, *Campylobacter*, *Clostridium difficile* spp.) may initiate IBD by triggering an inflammatory response that the mucosal immune system may fail to control. However, in an IBD patient, the normal microbiota is likely perceived inappropriately as if it were a pathogen. Alterations in the composition of the commensal microbiota are observed in both CD and UC. However, whether these changes are primary or secondary to inflammation is unknown. Anaerobic organisms, particularly *Bacteroides* and *Clostridia* species, and some aerobic species such as *Escherichia* may be responsible for the induction of inflammation. This notion is supported by the immune response in patients with CD to a number of bacterial antigens. In addition, agents that alter the intestinal flora such as metronidazole, ciprofloxacin, and elemental diets, may improve CD. CD also responds to fecal diversion, demonstrating the ability of luminal contents to exacerbate disease. Conversely, other organisms, so-called probiotics (e.g., *Faecalibacterium prausnitzii,Lactobacillus*, *Bifidobacterium*, *Taenia suis,* and *Saccharomyces boulardii* spp.), may inhibit inflammation in animal models and humans.

Psychosocial factors can contribute to worsening of symptoms. Major life events such as illness or death in the family, divorce or separation, interpersonal conflict, or other major loss are associated with an increase in IBD symptoms such as pain, bowel dysfunction, and bleeding. Acute daily stress can worsen bowel symptoms even after controlling for major life events. When measured with validated psychological scales, patients with active IBD have lower psychological well-being and mastery as well as higher distress than non-IBD controls.

PATHOLOGY

Ulcerative Colitis: Macroscopic Features

UC is a mucosal disease that usually involves the rectum and extends proximally to involve all or part of the colon. About 40–50% of patients have disease limited to the rectum and rectosigmoid, 30–40% have disease extending beyond the sigmoid but not involving the whole colon, and 20% have a total colitis. Proximal spread occurs in continuity without areas of uninvolved mucosa. When the whole colon is involved, the inflammation extends 2–3 cm into the terminal ileum in 10–20% of patients. The endoscopic changes of *backwash ileitis* are superficial and mild and are of little clinical significance. Although variations in macroscopic activity may suggest skip areas, biopsies from normal-appearing mucosa are usually abnormal. Thus, it is important to obtain multiple biopsies from apparently uninvolved mucosa, whether proximal or distal, during endoscopy. One caveat is that effective medical therapy can change the appearance of the mucosa such that either skip areas or the entire colon can be microscopically normal.

With mild inflammation, the mucosa is erythematous and has a fine granular surface that resembles sandpaper. In more severe disease, the mucosa is hemorrhagic, edematous, and ulcerated **(Fig. 295-1)** In long-standing disease, inflammatory polyps (pseudopolyps) may be present as a result of epithelial regeneration. The mucosa may appear normal in remission, but in patients with many years of disease it appears atrophic and featureless, and the entire colon becomes narrowed and shortened. Patients with fulminant disease can develop a toxic colitis or megacolon where the bowel wall thins and the mucosa is severely ulcerated; this may lead to perforation.

Figure 295-1

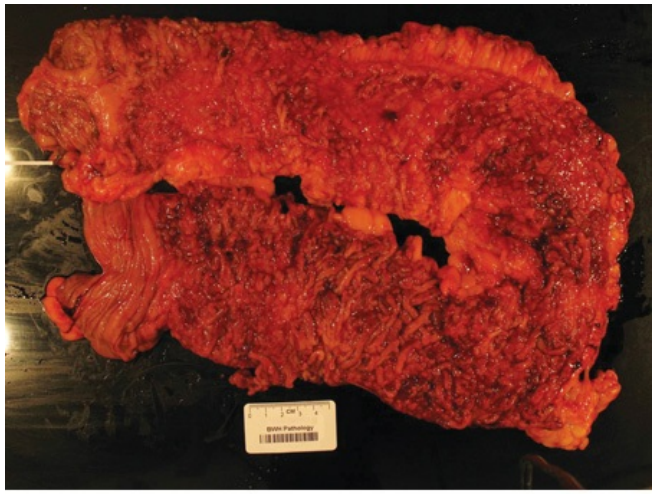

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ulcerative colitis. Diffuse (nonsegmental) mucosal disease, with broad areas of ulceration. The bowel wall is not thickened, and there is no cobblestoning. *(Courtesy of Dr. R Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

Ulcerative Colitis: Microscopic Features

Histologic findings correlate well with the endoscopic appearance and clinical course of UC. The process is limited to the mucosa and superficial submucosa, with deeper layers unaffected except in fulminant disease. In UC, two major histologic features suggest chronicity and help distinguish it from infectious or acute self-limited colitis. First, the crypt architecture of the colon is distorted; crypts may be bifid and reduced in number, often with a gap between the crypt bases and the muscularis mucosae. Second, some patients have basal plasma cells and multiple basal lymphoid aggregates. Mucosal vascular congestion, with edema and focal hemorrhage, and an inflammatory cell infiltrate of neutrophils, lymphocytes, plasma cells, and macrophages may be present. The neutrophils invade the epithelium, usually in the crypts, giving rise to cryptitis and, ultimately, to crypt abscesses **(Fig. 295-2)**. Ileal changes in patients with backwash ileitis include villous atrophy and crypt regeneration with increased inflammation, increased neutrophil and mononuclear inflammation in the lamina propria, and patchy cryptitis and crypt abscesses.

Figure 295-2

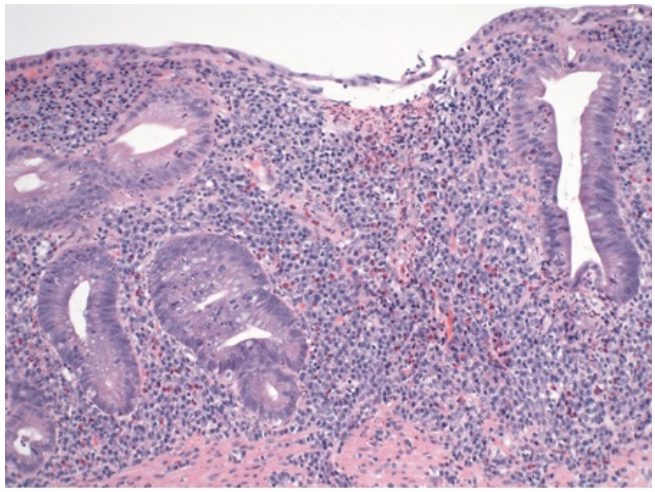

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Medium power view of colonic mucosa in ulcerative colitis showing diffuse mixed inflammation, basal lymphoplasmacytosis, crypt atrophy and irregularity and superficial erosion. These features are typical of chronic active ulcerative colitis. *(Courtesy of Dr. R Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

Crohn's Disease: Macroscopic Features

CD can affect any part of the gastrointestinal (GI) tract from the mouth to the anus. Some 30–40% of patients have smallbowel disease alone, 40–55% have disease involving both the small and large intestines, and 15–25% have colitis alone. In the 75% of patients with smallintestinal disease, the terminal ileum is involved in 90%. Unlike UC, which almost always involves the rectum, the rectum is often spared in CD. CD is segmental with skip areas in the midst of diseased intestine **(Fig. 295-3)** Perirectal fistulas, fissures, abscesses, and anal stenosis are present in one-third of patients with CD, particularly those with colonic involvement. Rarely, CD may also involve the liver and the pancreas.

Figure 295-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Crohn's disease of the colon showing thickening of the wall, with stenosis, linear serpiginous ulcers and cobblestoning of the mucosa. *(Courtesy of Dr. R Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

Unlike UC, CD is a transmural process. Endoscopically, aphthous or small superficial ulcerations characterize mild disease; in more active disease, stellate ulcerations fuse longitudinally and transversely to demarcate islands of mucosa that frequently are histologically normal. This "cobblestone" appearance is characteristic of CD, both endoscopically and by barium radiography. As in UC, pseudopolyps can form in CD.

Active CD is characterized by focal inflammation and formation of fistula tracts, which resolve by fibrosis and stricturing of the bowel. The bowel wall thickens and becomes narrowed and fibrotic, leading to chronic, recurrent bowel obstructions. Projections of thickened mesentery encase the bowel ("creeping fat"), and serosal and mesenteric inflammation promotes adhesions and fistula formation.

Crohn's Disease: Microscopic Features

The earliest lesions are aphthoid ulcerations and focal crypt abscesses with loose aggregations of macrophages, which form noncaseating granulomas in all layers of the bowel wall **(Fig. 295-4)**. Granulomas can be seen in lymph nodes, mesentery, peritoneum, liver, and pancreas. Although granulomas are a pathognomonic feature of CD, they are rarely found on mucosal biopsies. Surgical resection reveals granulomas in about one-half of cases. Other histologic features of CD include submucosal or subserosal lymphoid aggregates, particularly away from areas of ulceration, gross and microscopic skip areas, and transmural inflammation that is accompanied by fissures that penetrate deeply into the bowel wall and sometimes form fistulous tracts or local abscesses.

Figure 295-4

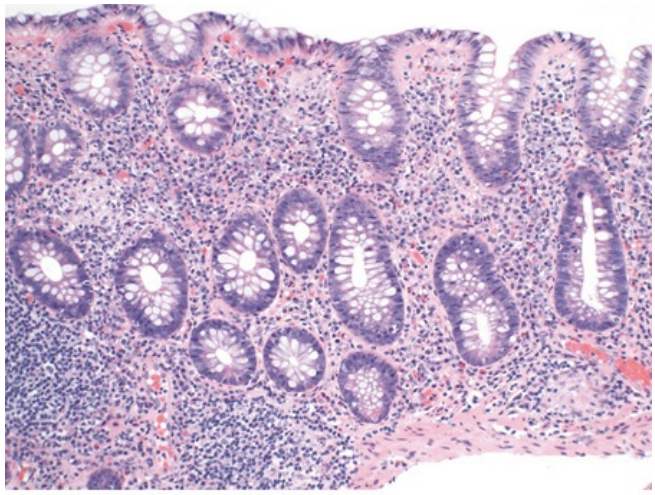

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Medium power view of Crohn's colitis showing mixed acute and chronic inflammation, crypt atrophy, and multiple small epithelioid granulomas in the mucosa. *(Courtesy of Dr. R Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

CLINICAL PRESENTATION

Ulcerative Colitis

SIGNS AND SYMPTOMS

The major symptoms of UC are diarrhea, rectal bleeding, tenesmus, passage of mucus, and crampy abdominal pain. The severity of symptoms correlates with the extent of disease. Although UC can present acutely, symptoms usually have been present for weeks to months. Occasionally, diarrhea and bleeding are so intermittent and mild that the patient does not seek medical attention.

Patients with proctitis usually pass fresh blood or blood-stained mucus, either mixed with stool or streaked onto the surface of a normal or hard stool. They also have tenesmus, or urgency with a feeling of incomplete evacuation, but rarely have abdominal pain. With proctitis or proctosigmoiditis, proximal transit slows, which may account for the constipation commonly seen in patients with distal disease.

When the disease extends beyond the rectum, blood is usually mixed with stool or grossly bloody diarrhea may be noted. Colonic motility is altered by inflammation with rapid transit through the inflamed intestine. When the disease is severe, patients pass a liquid stool containing blood, pus, and fecal matter. Diarrhea is often nocturnal and/or postprandial. Although severe pain is not a prominent symptom, some patients with active disease may experience vague lower abdominal discomfort or mild central abdominal cramping. Severe cramping and abdominal pain can occur with severe attacks of the disease. Other symptoms in moderate to severe disease include anorexia, nausea, vomiting, fever, and weight loss.

Physical signs of proctitis include a tender anal canal and blood on rectal examination. With more extensive disease, patients have tenderness to palpation directly over the colon. Patients with a toxic colitis have severe pain and bleeding, and those with megacolon have hepatic tympany. Both may have signs of peritonitis if a perforation has occurred. The classification of disease activity is shown in **Table 295-4**.

Table 295-4 Ulcerative Colitis: Disease Presentation

	Mild	Moderate	Severe
Bowel movements	<4 per day	4–6 per day	>6 per day
Blood in stool	Small	Moderate	Severe
Fever	None	<37.5°C mean (<99.5°F)	>37.5°C mean (>99.5°F)
Tachycardia	None	<90 mean pulse	>90 mean pulse
Anemia	Mild	>75%	≤75%
Sedimentation rate	<30 mm		>30 mm
Endoscopic appearance	Erythema, decreased vascular pattern, fine granularity	Marked erythema, coarse granularity, absent vascular markings, contact bleeding, no ulcerations	Spontaneous bleeding, ulcerations

LABORATORY, ENDOSCOPIC, AND RADIOGRAPHIC FEATURES

Active disease can be associated with a rise in acute-phase reactants [C-reactive protein (CRP)], platelet

count, erythrocyte sedimentation rate (ESR), and a decrease in hemoglobin. Fecal lactoferrin is a highly sensitive and specific marker for detecting intestinal inflammation. Fecal calprotectin levels correlate well with histologic inflammation, predict relapses, and detect pouchitis. In severely ill patients, the serum albumin level will fall rather quickly. Leukocytosis may be present but is not a specific indicator of disease activity. Proctitis or proctosigmoiditis rarely causes a rise in CRP. Diagnosis relies upon the patient's history; clinical symptoms; negative stool examination for bacteria, *C. difficile* toxin, and ova and parasites; sigmoidoscopic appearance (see **Fig. 291-4***A*); and histology of rectal or colonic biopsy specimens.

Sigmoidoscopy is used to assess disease activity and is usually performed before treatment. If the patient is not having an acute flare, colonoscopy is used to assess disease extent and activity **(Fig. 295-5)**. Endoscopically mild disease is characterized by erythema, decreased vascular pattern, and mild friability. Moderate disease is characterized by marked erythema, absent vascular pattern, friability and erosions, and severe disease by spontaneous bleeding and ulcerations. Histologic features change more slowly than clinical features but can also be used to grade disease activity.

Figure 295-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Colonoscopy with acute ulcerative colitis: Severe colon inflammation with erythema, friability, and exudates. *(Courtesy of Dr. M. Hamilton, Gastroenterology Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

The earliest radiologic change of UC seen on single-contrast barium enema is a fine mucosal granularity. With increasing severity, the mucosa becomes thickened, and superficial ulcers are seen. Deep ulcerations can appear as "collar-button" ulcers, which indicate that the ulceration has penetrated the mucosa. Haustral folds may be normal in mild disease, but as activity progresses they become edematous and thickened. Loss of haustration can occur, especially in patients with long-standing disease. In addition, the colon becomes shortened and narrowed. Polyps in the colon may be postinflammatory polyps or pseudopolyps, adenomatous polyps, or carcinoma.

CT scanning is not as helpful as endoscopy and barium enema in making the diagnosis of UC, but typical findings include mild mural thickening (<1.5 cm), inhomogeneous wall density, absence of small bowel thickening, increased perirectal and presacral fat, target appearance of the rectum, and adenopathy.

COMPLICATIONS

Only 15% of patients with UC present initially with catastrophic illness. Massive hemorrhage occurs with severe attacks of disease in 1% of patients, and treatment for the disease usually stops the bleeding. However, if a patient requires 6–8 units of blood within 24–48 hours, colectomy is indicated. *Toxic megacolon* is defined as a transverse or right colon with a diameter of >6 cm, with loss of haustration in patients with severe attacks of UC. It occurs in about 5% of attacks and can be triggered by electrolyte abnormalities and narcotics. About 50% of acute dilations will resolve with medical therapy alone, but urgent colectomy is required for those that do not improve. Perforation is the most dangerous of the local complications, and the physical signs of peritonitis may not be obvious, especially if the patient is receiving glucocorticoids. Although perforation is rare, the mortality rate for perforation complicating a toxic megacolon is about 15%. In addition, patients can develop a toxic colitis and such severe ulcerations that the bowel may perforate without first dilating.

Strictures occur in 5–10% of patients and are always a concern in UC because of the possibility of underlying neoplasia. Although benign strictures can form from the inflammation and fibrosis of UC, strictures that are impassable with the colonoscope should be presumed malignant until proven otherwise. A stricture that prevents passage of the colonoscope is an indication for surgery. UC patients occasionally develop anal fissures, perianal abscesses, or hemorrhoids, but the occurrence of extensive perianal lesions should suggest CD.

Crohn's Disease

SIGNS AND SYMPTOMS

Although CD usually presents as acute or chronic bowel inflammation, the inflammatory process evolves toward one of two patterns of disease: a fibrostenotic obstructing pattern or a penetrating fistulous pattern, each with different treatments and prognoses. The site of disease influences the clinical manifestations.

Ileocolitis

Because the most common site of inflammation is the terminal ileum, the usual presentation of ileocolitis is a chronic history of recurrent episodes of right lower quadrant pain and diarrhea. Sometimes the initial presentation mimics acute appendicitis with pronounced right lower quadrant pain, a palpable mass, fever, and leukocytosis. Pain is usually colicky; it precedes and is relieved by defecation. A low-grade fever is usually noted. High-spiking fever suggests intraabdominal abscess formation. Weight loss is common—typically 10–20% of body weight—and develops as a consequence of diarrhea, anorexia, and fear of eating.

An inflammatory mass may be palpated in the right lower quadrant of the abdomen. The mass is composed of inflamed bowel, adherent and indurated mesentery, and enlarged abdominal lymph nodes. Extension of the mass can cause obstruction of the right ureter or bladder inflammation, manifested by dysuria and fever. Edema, bowel wall thickening, and fibrosis of the bowel wall within the mass account for the radiographic "string sign" of a narrowed intestinal lumen.

Bowel obstruction may take several forms. In the early stages of disease, bowel wall edema and spasm produce intermittent obstructive manifestations and increasing symptoms of postprandial pain. Over several years, persistent inflammation gradually progresses to fibrostenotic narrowing and stricture. Diarrhea will decrease and be replaced by chronic bowel obstruction. Acute episodes of obstruction occur as well, precipitated by bowel inflammation and spasm or sometimes by impaction of undigested food or medication. These episodes usually resolve with intravenous fluids and gastric decompression.

Severe inflammation of the ileocecal region may lead to localized wall thinning, with microperforation and fistula formation to the adjacent bowel, the skin, or the urinary bladder, or to an abscess cavity in the mesentery. Enterovesical fistulas typically present as dysuria or recurrent bladder infections or, less commonly, as pneumaturia or fecaluria. Enterocutaneous fistulas follow tissue planes of least resistance, usually draining through abdominal surgical scars. Enterovaginal fistulas are rare and present as dyspareunia or as a feculent or foul-smelling, often painful vaginal discharge. They are unlikely to develop without a prior hysterectomy.

Jejunoileitis

Extensive inflammatory disease is associated with a loss of digestive and absorptive surface, resulting in malabsorption and steatorrhea. Nutritional deficiencies can also result from poor intake and enteric losses of protein and other nutrients. Intestinal malabsorption can cause anemia, hypoalbuminemia, hypocalcemia, hypomagnesemia, coagulopathy, and hyperoxaluria with nephrolithiasis in patients with an intact colon. Many patients need to take oral and often intravenous iron. Vertebral fractures are caused by a combination of vitamin D deficiency, hypocalcemia, and prolonged glucocorticoid use. Pellagra from niacin deficiency can occur in extensive small bowel disease, and malabsorption of vitamin B12 can lead to megaloblastic anemia and neurologic symptoms. Other important nutrients to measure and replete if low are folate and vitamins A, E, and K. Levels of minerals such as zinc, selenium, copper, and magnesium are often low in patients with extensive small bowel inflammation or resections and these should be repleted as well. Most patients should take a daily multivitamin, calcium, and vitamin D supplements.

Diarrhea is characteristic of active disease; its causes include (1) bacterial overgrowth in obstructive stasis or fistulization, (2) bile-acid malabsorption due to a diseased or resected terminal ileum, and (3) intestinal inflammation with decreased water absorption and increased secretion of electrolytes.

Colitis and Perianal Disease

Patients with colitis present with low-grade fevers, malaise, diarrhea, crampy abdominal pain, and sometimes hematochezia. Gross bleeding is not as common as in UC and appears in about one-half of patients with exclusively colonic disease. Only 1–2% bleed massively. Pain is caused by passage of fecal material through narrowed and inflamed segments of the large bowel. Decreased rectal compliance is another cause for diarrhea in Crohn's colitis patients. Toxic megacolon is rare but may be seen with severe inflammation and short duration disease.

Stricturing can occur in the colon in 4–16% of patients and produce symptoms of bowel obstruction. If the endoscopist is unable to traverse a stricture in Crohn's colitis, surgical resection should be considered, especially if the patient has symptoms of chronic obstruction. Colonic disease may fistulize into the stomach or duodenum, causing feculent vomiting, or to the proximal or mid-small bowel, causing malabsorption by "short circuiting" and bacterial overgrowth. Ten percent of women with Crohn's colitis will develop a rectovaginal fistula.

Perianal disease affects about one-third of patients with Crohn's colitis and is manifested by incontinence, large hemorrhoidal tags, anal strictures, anorectal fistulae, and perirectal abscesses. Not all patients with perianal fistula will have endoscopic evidence of colonic inflammation.

Gastroduodenal Disease

Symptoms and signs of upper GI tract disease include nausea, vomiting, and epigastric pain. Patients usually have an *Helicobacter pylori* –negative gastritis. The second portion of the duodenum is more commonly involved than the bulb. Fistulas involving the stomach or duodenum arise from the small or large bowel and do

not necessarily signify the presence of upper GI tract involvement. Patients with advanced gastroduodenal CD may develop a chronic gastric outlet obstruction.

LABORATORY, ENDOSCOPIC, AND RADIOGRAPHIC FEATURES

Laboratory abnormalities include elevated ESR and CRP. In more severe disease, findings include hypoalbuminemia, anemia, and leukocytosis.

Endoscopic features of CD include rectal sparing, aphthous ulcerations, fistulas, and skip lesions. Colonoscopy allows examination and biopsy of mass lesions or strictures and biopsy of the terminal ileum. Upper endoscopy is useful in diagnosing gastroduodenal involvement in patients with upper tract symptoms. Ileal or colonic strictures may be dilated with balloons introduced through the colonoscope. Strictures le; 4 cm and those at a anastomotic sites respond better to endoscopic dilation. The perforation rate is as high as 10%. Most endoscopists dilate only fibrotic strictures and not those associated with active inflammation. Wireless capsule endoscopy (WCE) allows direct visualization of the entire small bowel mucosa **(Fig. 295-6)**. The diagnostic yield of detecting lesions suggestive of active CD is higher with WCE than CT enterography or small bowel series. WCE cannot be used in the setting of a small bowel stricture. Capsule retention occurs in <1% of patients with suspected CD, but retention rates of 4–6% are seen in patients with established CD.

Figure 295-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Wireless capsule endoscopy image in a patient with Crohn's disease of the ileum shows ulcerations and narrowing of the intestinal lumen. *(Courtesy of Dr. S Reddy, Gastroenterology Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

In CD, early radiographic findings in the small bowel include thickened folds and aphthous ulcerations. "Cobblestoning" from longitudinal and transverse ulcerations most frequently involves the small bowel. In more advanced disease, strictures, fistulas, inflammatory masses, and abscesses may be detected. The earliest macroscopic findings of colonic CD are aphthous ulcers. These small ulcers are often multiple and separated by normal intervening mucosa. As the disease progresses, aphthous ulcers become enlarged, deeper, and occasionally connected to one another, forming longitudinal stellate, serpiginous, and linear ulcers (see **Fig. 291-4***B*).

The transmural inflammation of CD leads to decreased luminal diameter and limited distensibility. As ulcers progress deeper, they can lead to fistula formation. The radiographic "string sign" represents long areas of circumferential inflammation and fibrosis, resulting in long segments of luminal narrowing. The segmental nature of CD results in wide gaps of normal or dilated bowel between involved segments.

CT enterography combines the improved spatial and temporal resolution of multidetector-row CT with large volumes of ingested neutral enteric contrast material to permit visualization of the entire small bowel and lumen. Unlike routine CT, which is used to detect the extraenteric complications of CD such as fistula and abscess, CT enterography clearly depicts the small bowel inflammation associated with CD by displaying mural hyperenhancement, stratification, and thickening; engorged vasa recta; and perienteric inflammatory changes **(Figs. 295-7** and **295-8)**. CT enterography is the first-line test for the evaluation of suspected CD and its complications. As an initial test in children or in adults with multiple radiation exposures, MR enterography is comparable to CT in diagnostic accuracy. Pelvic MRI is superior to CT for demonstrating pelvic lesions such as ischiorectal abscesses and perianal fistulae **(Fig. 295-9)**.

Figure 295-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronal contrast-enhanced multidetector computed tomography (MDCT) image obtained after oral administration of 1350 cc of neutral oral contrast material shows dilation of small bowel loops, segmental mucosal hyperenhancement, and interloop sinus tracts (*white arrow*) and mesenteric fat stranding. *(Courtesy of Dr. K Mortele, Gastrointestinal Radiology, Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

Figure 295-8

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronal contrast-enhanced multidetector computed tomography (MDCT) image obtained after oral administration of 1350 cc of neutral oral contrast material shows mucosal hyperenhancement of the terminal ileum with narrowing and mild prestenotic dilatation. *(Courtesy of Dr. K Mortele, Gastrointestinal Radiology, Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

Figure 295-9

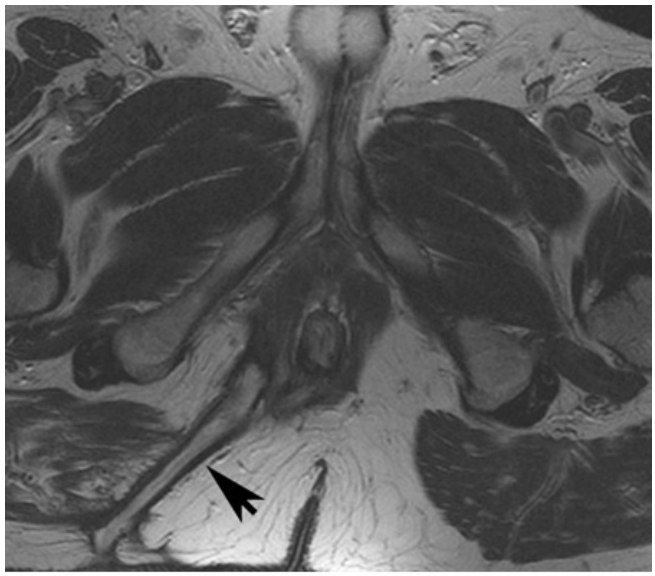

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial T2-weighted MR image obtained in a 37-year-old man with Crohn's disease shows a linear fluid-filled perianal fistula (***arrow***) in the right ischioanal fossa. *(Courtesy of Dr. K Mortele, Gastrointestinal Radiology, Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

COMPLICATIONS

Because CD is a transmural process, serosal adhesions develop that provide direct pathways for fistula formation and reduce the incidence of free perforation. Perforation occurs in 1–2% of patients, usually in the ileum but occasionally in the jejunum or as a complication of toxic megacolon. The peritonitis of free perforation, especially colonic, may be fatal. Intraabdominal and pelvic abscesses occur in 10–30% of patients with Crohn's disease at some time in the course of their illness. CT-guided percutaneous drainage of the abscess is standard therapy. Despite adequate drainage, most patients need resection of the offending bowel segment. Percutaneous drainage has an especially high failure rate in abdominal wall abscesses. Systemic glucocorticoid therapy increases the risk of intraabdominal and pelvic abscesses in CD patients who have never had an operation. Other complications include intestinal obstruction in 40%, massive hemorrhage, malabsorption, and severe perianal disease.

SEROLOGIC MARKERS

Patients with Crohn's disease show a wide variation in the way they present and progress over time. Some patients present with mild disease activity and do well with generally safe and mild medications, but many others exhibit more severe disease and can develop serious complications that will require surgery. Current and developing biologic therapies can help halt progression of disease and give patients with moderate to severe Crohn's disease a better quality of life. There are potential risks of biologic therapies such as infection and malignancy, and it would be optimal to determine at the time of diagnosis which patients will require more aggressive medical therapy. This same argument holds true for UC patients as well.

Subsets of patients with differing immune responses to microbial antigens have been described. These include antibodies to *Escherichia coli* (*E. coli*) outer membrane porin protein C (OmpC), which is found in 55% of CD patients, an antibody to I_2, a homologue of the bacterial transcription-factor families from a *Pseudomonas fluorescens*–associated sequence that is found in 50–54% of CD patients, as well as anti-*Saccharomyces cerevisiae* (ASCA) and autoantigens [perinuclear antineutrophil antibody (pANCA)]. A novel immune response, antiflagellin (anti-CBir1) has been identified in approximately 50% of Crohn's patients and has been suggested to represent a unique subgroup of CD patients.

Unfortunately, these serologic markers are only marginally useful in helping to make the diagnosis of UC or CD and in predicting the course of disease. For success in diagnosing IBD and in differentiating between CD and UC, the efficacy of these serologic tests depends upon the prevalence of IBD in a specific population. pANCA positivity is found in about 60–70% of UC patients and 5–10% of CD patients; 5–15% of first-degree relatives of UC patients are pANCA positive, whereas only 2–3% of the general population is pANCA positive.

Sixty to seventy percent of CD patients, 10–15% of UC patients, and up to 5% of non-IBD controls are ASCA-positive. In a patient population with a combined prevalence of UC and CD of 62%, pANCA/ASCA serology showed a sensitivity of 64% and a specificity of 94%. Positive and negative predictive values (PPVs and NPVs) for pANCA/ASCA also vary based on the prevalence of IBD in a given population. For the patient population with a prevalence of IBD of 62%, the PPV is 94%, and the NPV is 63%.

Combining these diagnostic assays may improve the ability to diagnose CD. In a referral population of CD patients, 85% had an antibody to at least one antigen (pANCA, ASCA, OmpC, and I_2); only 4% responded to

all four. Some evidence suggests that antibody positivity may help predict disease phenotype. ASCA positivity is associated with an increased rate of early CD complications; OmpC–positive patients are more likely to have internal perforating disease; and I_2 positive patients are more likely to have fibrostenosing disease. Patients positive for I_2, OmpC, and ASCA are the most likely to have undergone small bowel surgery.

Anti-Cbir1 expression is associated with small-bowel disease, fibrostenosing, and internal penetrating disease. Children with CD positive for all four immune responses (ASCA+, OmpC+, I_2+, and anti-Cbir1+) may have more aggressive disease and a shorter time to progression to internal perforating and/or stricturing disease. However, larger prospective studies in both children and adults have not yet been performed and compared to CRP or other markers.

Clinical factors described at diagnosis are more helpful than serologies at predicting the natural history of Crohn's disease. The initial requirement for glucocorticoid use, an age at diagnosis below 40 years and the presence of perianal disease at diagnosis, have been shown to be independently associated with subsequent disabling CD after 5 years. Except in special circumstances [such as before consideration of an ileoanal pouch anastomosis (IPAA) in a patient with indeterminate colitis], serologic markers have only minimal clinical utility.

DIFFERENTIAL DIAGNOSIS OF UC AND CD

UC and CD have similar features to many other diseases. In the absence of a key diagnostic test, a combination of features is used **(Table 295-5)**. Once a diagnosis of IBD is made, distinguishing between UC and CD is impossible initially in up to 15% of cases. These are termed *indeterminate colitis*. Fortunately, in most cases, the true nature of the underlying colitis becomes evident later in the course of the patient's disease. Approximately 5% (range 1–20%) of colon resection specimens are difficult to classify as either UC or CD because they exhibit overlapping histologic features.

Table 295-5 Different Clinical, Endoscopic, and Radiographic Features

	Ulcerative Colitis	Crohn's Disease
Clinical		
Gross blood in stool	Yes	Occasionally
Mucus	Yes	Occasionally
Systemic symptoms	Occasionally	Frequently
Pain	Occasionally	Frequently
Abdominal mass	Rarely	Yes
Significant perineal disease	No	Frequently
Fistulas	No	Yes
Small intestinal obstruction	No	Frequently
Colonic obstruction	Rarely	Frequently
Response to antibiotics	No	Yes
Recurrence after surgery	No	Yes
ANCA-positive	Frequently	Rarely
ASCA-positive	Rarely	Frequently
Endoscopic		
Rectal sparing	Rarely	Frequently
Continuous disease	Yes	Occasionally
"Cobblestoning"	No	Yes
Granuloma on biopsy	No	Occasionally
Radiographic		
Small bowel significantly abnormal	No	Yes
Abnormal terminal ileum	No	Yes
Segmental colitis	No	Yes
Asymmetric colitis	No	Yes
Stricture	Occasionally	Frequently

Abbreviations: ANCA, antineutrophil cytoplasm antibody; ASCA, anti-*Saccharomycescerevisiae* antibody.

Infectious Disease

Infections of the small intestines and colon can mimic CD or UC. They may be bacterial, fungal, viral, or protozoal in origin **(Table 295-6)**. *Campylobacter* colitis can mimic the endoscopic appearance of severe UC and can cause a relapse of established UC. *Salmonella* can cause watery or bloody diarrhea, nausea, and vomiting. Shigellosis causes watery diarrhea, abdominal pain, and fever followed by rectal tenesmus and by the passage of blood and mucus per rectum. All three are usually self-limited, but 1% of patients infected with *Salmonella* become asymptomatic carriers. *Yersinia enterocolitica* infection occurs mainly in the terminal ileum and causes mucosal ulceration, neutrophil invasion, and thickening of the ileal wall. Other bacterial

infections that may mimic IBD include *C. difficile,* which presents with watery diarrhea, tenesmus, nausea, and vomiting; and *E. coli,* three categories of which can cause colitis. These are enterohemorrhagic, enteroinvasive, and enteroadherent *E. coli,* all of which can cause bloody diarrhea and abdominal tenderness. Diagnosis of bacterial colitis is made by sending stool specimens for bacterial culture and *C. difficile* toxin analysis. Gonorrhea, *Chlamydia*, and syphilis can also cause proctitis.

Table 295-6 Diseases that Mimic Ibd

Infectious Etiologies		
Bacterial	**Mycobacterial**	**Viral**
Salmonella	Tuberculosis	Cytomegalovirus
Shigella	*Mycobacterium avium*	Herpes simplex
Toxigenic	**Parasitic**	HIV
Escherichia coli	Amebiasis	**Fungal**
Campylobacter	*Isospora*	Histoplasmosis
Yersinia	*Trichuris trichiura*	*Candida*
Clostridium difficile	Hookworm	*Aspergillus*
Gonorrhea	*Strongyloides*	
Chlamydia trachomatis		
Noninfectious Etiologies		
Inflammatory	**Neoplastic**	**Drugs and Chemicals**
Appendicitis	Lymphoma	NSAIDs
Diverticulitis	Metastatic carcinoma	Phosphosoda
Diversion colitis	Carcinoma of the ileum	Cathartic colon
Collagenous/lymphocytic colitis	Carcinoid	Gold
Ischemic colitis	Familial polyposis	Oral contraceptives
Radiation colitis/enteritis		Cocaine
Solitary rectal ulcer syndrome		Chemotherapy
Eosinophilic gastroenteritis		
Neutropenic colitis		
Behçet's syndrome		
Graft-versus-host disease		

Abbreviation: NSAIDs, nonsteroidal anti-inflammatory drugs.

GI involvement with mycobacterial infection occurs primarily in the immunosuppressed patient but may occur in patients with normal immunity. Distal ileal and cecal involvement predominates, and patients present with symptoms of small bowel obstruction and a tender abdominal mass. The diagnosis is made most directly by colonoscopy with biopsy and culture. *Mycobacterium avium-intracellulare* complex infection occurs in advanced stages of HIV infection and in other immunocompromised states; it usually manifests as a systemic infection with diarrhea, abdominal pain, weight loss, fever, and malabsorption. Diagnosis is established by acid-fast smear and culture of mucosal biopsies.

Although most of the patients with viral colitis are immunosuppressed, cytomegalovirus (CMV) and herpes simplex proctitis may occur in immunocompetent individuals. CMV occurs most commonly in the esophagus, colon, and rectum but may also involve the small intestine. Symptoms include abdominal pain, bloody diarrhea, fever, and weight loss. With severe disease, necrosis and perforation can occur. Diagnosis is made by identification of characteristic intranuclear inclusions in mucosal cells on biopsy. Herpes simplex infection of the GI tract is limited to the oropharynx, anorectum, and perianal areas. Symptoms include anorectal pain, tenesmus, constipation, inguinal adenopathy, difficulty with urinary voiding, and sacral paresthesias. Diagnosis is made by rectal biopsy with identification of characteristic cellular inclusions and viral culture. HIV itself can cause diarrhea, nausea, vomiting, and anorexia. Small intestinal biopsies show partial villous atrophy; small bowel bacterial overgrowth and fat malabsorption may also be noted.

Protozoan parasites include *Isospora belli,* which can cause a self-limited infection in healthy hosts but causes a chronic profuse, watery diarrhea, and weight loss in AIDS patients. *Entamoeba histolytica* or related species infect about 10% of the world's population; symptoms include abdominal pain, tenesmus, frequent loose stools containing blood and mucus, and abdominal tenderness. Colonoscopy reveals focal punctate ulcers with normal intervening mucosa; diagnosis is made by biopsy or serum amebic antibodies. Fulminant amebic colitis is rare but has a mortality rate of >50%.

Other parasitic infections that may mimic IBD include hookworm (*Necator americanus*), whipworm (*T. trichiura*), and *Strongyloides stercoralis.* In severely immunocompromised patients, *Candida* or *Aspergillus* can be identified in the submucosa. Disseminated histoplasmosis can involve the ileocecal area.

Noninfectious Disease

Diverticulitis can be confused with CD clinically and radiographically. Both diseases cause fever, abdominal

pain, tender abdominal mass, leukocytosis, elevated ESR, partial obstruction, and fistulas. Perianal disease or ileitis on small bowel series favors the diagnosis of CD. Significant endoscopic mucosal abnormalities are more likely in CD than in diverticulitis. Endoscopic or clinical recurrence following segmental resection favors CD. Diverticular-associated colitis is similar to CD, but mucosal abnormalities are limited to the sigmoid and descending colon.

Ischemic colitis is commonly confused with IBD. The ischemic process can be chronic and diffuse, as in UC, or segmental, as in CD. Colonic inflammation due to ischemia may resolve quickly or may persist and result in transmural scarring and stricture formation. Ischemic bowel disease should be considered in the elderly following abdominal aortic aneurysm repair or when a patient has a hypercoagulable state or a severe cardiac or peripheral vascular disorder. Patients usually present with sudden onset of left lower quadrant pain, urgency to defecate, and the passage of bright red blood per rectum. Endoscopic examination often demonstrates a normal-appearing rectum and a sharp transition to an area of inflammation in the descending colon and splenic flexure.

The effects of radiotherapy on the GI tract can be difficult to distinguish from IBD. Acute symptoms can occur within 1–2 weeks of starting radiotherapy. When the rectum and sigmoid are irradiated, patients develop bloody, mucoid diarrhea and tenesmus, as in distal UC. With small bowel involvement, diarrhea is common. Late symptoms include malabsorption and weight loss. Stricturing with obstruction and bacterial overgrowth may occur. Fistulas can penetrate the bladder, vagina, or abdominal wall. Flexible sigmoidoscopy reveals mucosal granularity, friability, numerous telangiectasias, and occasionally discrete ulcerations. Biopsy can be diagnostic.

Solitary rectal ulcer syndrome is uncommon and can be confused with IBD. It occurs in persons of all ages and may be caused by impaired evacuation and failure of relaxation of the puborectalis muscle. Single or multiple ulcerations may arise from anal sphincter overactivity, higher intrarectal pressures during defecation, and digital removal of stool. Patients complain of constipation with straining and pass blood and mucus per rectum. Other symptoms include abdominal pain, diarrhea, tenesmus, and perineal pain. Ulceration as large as 5 cm in diameter is usually seen anteriorly or anterior-laterally 3–15 cm from the anal verge. Biopsies can be diagnostic.

Several types of colitis are associated with nonsteroidal anti-inflammatory drugs (NSAIDs), including de novo colitis, reactivation of IBD, and proctitis caused by use of suppositories. Most patients with NSAID-related colitis present with diarrhea and abdominal pain, and complications include stricture, bleeding, obstruction, perforation, and fistulization. Withdrawal of these agents is crucial, and in cases of reactivated IBD, standard therapies are indicated.

The Atypical Colitides

Two atypical colitides—collagenous colitis and lymphocytic colitis—have completely normal endoscopic appearances. Collagenous colitis has two main histologic components: increased subepithelial collagen deposition and colitis with increased intraepithelial lymphocytes. The female to male ratio is 9:1, and most patients present in the sixth or seventh decades of life. The main symptom is chronic watery diarrhea. Treatments range from sulfasalazine or mesalamine and Lomotil to bismuth to budesonide to prednisone for refractory disease.

Lymphocytic colitis has features similar to collagenous colitis, including age at onset and clinical presentation, but it has almost equal incidence in men and women and no subepithelial collagen deposition on pathologic section. However, intraepithelial lymphocytes are increased. The frequency of celiac disease is increased in lymphocytic colitis and ranges from 9 to 27%. Celiac disease should be excluded in all patients with lymphocytic colitis, particularly if diarrhea does not respond to conventional therapy. Treatment is similar to that of collagenous colitis with the exception of a gluten-free diet for those who have celiac disease.

Diversion colitis is an inflammatory process that arises in segments of the large intestine that are excluded from the fecal stream. It usually occurs in patients with ileostomy or colostomy when a mucus fistula or a Hartmann's pouch has been created. Clinically, patients have mucus or bloody discharge from the rectum. Erythema, granularity, friability, and, in more severe cases, ulceration can be seen on endoscopy. Histopathology shows areas of active inflammation with foci of cryptitis and crypt abscesses. Crypt architecture is normal, which differentiates it from UC. It may be impossible to distinguish from CD. Short-chain fatty acid enemas may help in diversion colitis, but the definitive therapy is surgical reanastomosis.

EXTRAINTESTINAL MANIFESTATIONS

Up to one-third of IBD patients have at least one extraintestinal disease manifestation.

Dermatologic

Erythema nodosum (EN) occurs in up to 15% of CD patients and 10% of UC patients. Attacks usually correlate with bowel activity; skin lesions develop after the onset of bowel symptoms, and patients frequently have concomitant active peripheral arthritis. The lesions of EN are hot, red, tender nodules measuring 1–5 cm in diameter and are found on the anterior surface of the lower legs, ankles, calves, thighs, and arms. Therapy is directed toward the underlying bowel disease.

Pyoderma gangrenosum (PG) is seen in 1–12% of UC patients and less commonly in Crohn's colitis. Although it usually presents after the diagnosis of IBD, PG may occur years before the onset of bowel symptoms, run a course independent of the bowel disease, respond poorly to colectomy, and even develop years after proctocolectomy. It is usually associated with severe disease. Lesions are commonly found on the dorsal surface of the feet and legs but may occur on the arms, chest, stoma, and even the face. PG usually begins as a pustule and then spreads concentrically to rapidly undermine healthy skin. Lesions then ulcerate, with

violaceous edges surrounded by a margin of erythema. Centrally, they contain necrotic tissue with blood and exudates. Lesions may be single or multiple and grow as large as 30 cm. They are sometimes very difficult to treat and often require intravenous (IV) antibiotics, intravenous, glucocorticoids, dapsone, azathioprine, thalidomide, IV cyclosporine, or infliximab.

Other dermatologic manifestations include pyoderma vegetans, which occurs in intertriginous areas; pyostomatitis vegetans, which involves the mucous membranes; Sweet's syndrome, a neutrophilic dermatosis; and metastatic CD, a rare disorder defined by cutaneous granuloma formation. Psoriasis affects 5–10% of patients with IBD and is unrelated to bowel activity consistent with the potential shared immunogenetic basis of these diseases. Perianal skin tags are found in 75–80% of patients with CD, especially those with colon involvement. Oral mucosal lesions, seen often in CD and rarely in UC, include aphthous stomatitis and "cobblestone" lesions of the buccal mucosa.

Rheumatologic

Peripheral arthritis develops in 15–20% of IBD patients, is more common in CD, and worsens with exacerbations of bowel activity. It is asymmetric, polyarticular, and migratory and most often affects large joints of the upper and lower extremities. Treatment is directed at reducing bowel inflammation. In severe UC, colectomy frequently cures the arthritis.

Ankylosing spondylitis (AS) occurs in about 10% of IBD patients and is more common in CD than UC. About two-thirds of IBD patients with AS express the HLA-B27 antigen. The AS activity is not related to bowel activity and does not remit with glucocorticoids or colectomy. It most often affects the spine and pelvis, producing symptoms of diffuse low-back pain, buttock pain, and morning stiffness. The course is continuous and progressive, leading to permanent skeletal damage and deformity. Infliximab reduces spinal inflammation and improves functional status and quality of life.

Sacroiliitis is symmetric, occurs equally in UC and CD, is often asymptomatic, does not correlate with bowel activity, and does not always progress to AS. Other rheumatic manifestations include hypertrophic osteoarthropathy, pelvic/femoral osteomyelitis, and relapsing polychondritis.

Ocular

The incidence of ocular complications in IBD patients is 1–10%. The most common are conjunctivitis, anterior uveitis/iritis, and episcleritis. Uveitis is associated with both UC and Crohn's colitis, may be found during periods of remission, and may develop in patients following bowel resection. Symptoms include ocular pain, photophobia, blurred vision, and headache. Prompt intervention, sometimes with systemic glucocorticoids, is required to prevent scarring and visual impairment. Episcleritis is a benign disorder that presents with symptoms of mild ocular burning. It occurs in 3–4% of IBD patients, more commonly in Crohn's colitis, and is treated with topical glucocorticoids.

Hepatobiliary

Hepatic steatosis is detectable in about one-half of the abnormal liver biopsies from patients with CD and UC; patients usually present with hepatomegaly. Fatty liver usually results from a combination of chronic debilitating illness, malnutrition, and glucocorticoid therapy. Cholelithiasis occurs in 10–35% of CD patients with ileitis or ileal resection. Gallstone formation is caused by malabsorption of bile acids, resulting in depletion of the bile salt pool and the secretion of lithogenic bile.

Primary sclerosing cholangitis (PSC) is a disorder characterized by both intrahepatic and extrahepatic bile duct inflammation and fibrosis, frequently leading to biliary cirrhosis and hepatic failure; approximately 5% of patients with UC have PSC, but 50–75% of patients with PSC have IBD. PSC occurs less often in patients with CD. Although it can be recognized after the diagnosis of IBD, PSC can be detected earlier or even years after proctocolectomy. Consistent with this, the immunogenetic basis for PSC appears to be overlapping but distinct from UC based upon genome-wide association studies (GWAS) although both IBD and PSC are commonly pANCA positive. Most patients have no symptoms at the time of diagnosis; when symptoms are present, they consist of fatigue, jaundice, abdominal pain, fever, anorexia, and malaise. The traditional gold-standard diagnostic test is endoscopic retrograde cholangiopancreatography (ERCP), but magnetic resonance cholangiopancreatography (MRCP) is also sensitive and specific. MRCP is reasonable as an initial diagnostic test in children and can visualize irregularities, multifocal strictures, and dilatations of all levels of the biliary tree. In patients with PSC, both ERCP and MRCP demonstrate multiple bile duct strictures alternating with relatively normal segments.

The bile acid ursodeoxycholic acid (ursodiol) may reduce alkaline phosphatase and serum aminotransferase levels, but histologic improvement has been marginal. High doses (25–30 mg/kg per day) may decrease the risk of colorectal dysplasia and cancer in patients with UC and PSC. Endoscopic stenting may be palliative for cholestasis secondary to bile duct obstruction. Patients with symptomatic disease develop cirrhosis and liver failure over 5–10 years and eventually require liver transplantation. PSC patients have a 10–15% lifetime risk of developing cholangiocarcinoma and then cannot be transplanted. Patients with IBD and PSC are at increased risk of colon cancer and should be surveyed yearly by colonoscopy and biopsy.

In addition, cholangiography is normal in a small percentage of patients who have a variant of PSC known as *small duct primary sclerosing cholangitis* . This variant (sometimes referred to as "pericholangitis") is probably a form of PSC involving small caliber bile ducts. It has similar biochemical and histologic features to classic PSC. It appears to have a significantly better prognosis than classic PSC, although it may evolve into classic PSC. Granulomatous hepatitis and hepatic amyloidosis are much rarer extraintestinal manifestations of IBD.

Urologic

The most frequent genitourinary complications are calculi, ureteral obstruction, and ileal bladder fistulas. The highest frequency of nephrolithiasis (10–20%) occurs in patients with CD following small bowel resection. Calcium oxalate stones develop secondary to hyperoxaluria, which results from increased absorption of dietary oxalate. Normally, dietary calcium combines with luminal oxalate to form insoluble calcium oxalate, which is eliminated in the stool. In patients with ileal dysfunction, however, nonabsorbed fatty acids bind calcium and leave oxalate unbound. The unbound oxalate is then delivered to the colon, where it is readily absorbed, especially in the presence of inflammation.

Metabolic Bone Disorders

Low bone mass occurs in 3–30% of IBD patients. The risk is increased by glucocorticoids, cyclosporine, methotrexate and total parenteral nutrition (TPN). Malabsorption and inflammation mediated by IL-1, IL-6, TNF and other inflammatory mediators also contribute to low bone density. An increased incidence of hip, spine, wrist, and rib fractures has been noted: 36% in CD and 45% in UC. The absolute risk of an osteoporotic fracture is about 1% per person per year. Fracture rates, particularly in the spine and hip, were highest among the elderly (age >60). One study noted an odds ratio of vertebral fracture to be 1.72 and hip fracture 1.59. The disease severity predicted the risk of a fracture. Only 13% of IBD patients who had a fracture were on any kind of antifracture treatment. Up to 20% of bone mass can be lost per year with chronic glucocorticoid use. The effect is dosage-dependent. Budesonide may also suppress the pituitary-adrenal axis and thus carries a risk of causing osteoporosis.

Osteonecrosis is characterized by death of osteocytes and adipocytes and eventual bone collapse. The pain is aggravated by motion and swelling of the joints. It affects the hips more often than knees and shoulders, and in one series 4.3% of patients developed osteonecrosis within 6 months of starting glucocorticoids. Diagnosis is made by bone scan or MRI, and treatment consists of pain control, cord decompression, osteotomy, and joint replacement.

Thromboembolic Disorders

Patients with IBD have an increased risk of both venous and arterial thrombosis even if the disease is not active. Factors responsible for the hypercoagulable state have included abnormalities of the platelet-endothelial interaction, hyperhomocysteinemia, alterations in the coagulation cascade, impaired fibrinolysis, involvement of tissue factor-bearing microvesicles, disruption of the normal coagulation system by autoantibodies, as well as a genetic predisposition. A spectrum of vasculitides involving small, medium, and large vessels has also been observed.

Other Disorders

More common cardiopulmonary manifestations include endocarditis, myocarditis, pleuropericarditis, and interstitial lung disease. A secondary or reactive amyloidosis can occur in patients with long-standing IBD, especially in patients with CD. Amyloid material is deposited systemically and can cause diarrhea, constipation, and renal failure. The renal disease can be successfully treated with colchicine. Pancreatitis is a rare extraintestinal manifestation of IBD and results from duodenal fistulas; ampullary CD; gallstones; PSC; drugs such as 6-mercaptopurine, azathioprine, or, very rarely, 5-ASA agents; autoimmune pancreatitis; and primary CD of the pancreas.

Treatment: Inflammatory Bowel Disease Treatment

5-ASA AGENTS

The mainstay of therapy for mild to moderate UC is sulfasalazine and the other 5-ASA agents. These agents are effective at inducing and maintaining remission in UC. They may have a limited role in inducing remission in CD but no clear role in maintenance of CD. The most convincing evidence for the use of sulfasalazine is treatment of active Crohn's disease involving the colon. Sulfasalazine was originally developed to deliver both antibacterial (sulfapyridine) and anti-inflammatory (5-ASA) therapy into the connective tissues of joints and the colonic mucosa. The molecular structure provides a convenient delivery system to the colon by allowing the intact molecule to pass through the small intestine after only partial absorption, and to be broken down in the colon by bacterial azo reductases that cleave the azo bond linking the sulfa and 5-ASA moieties. Sulfasalazine is effective treatment for mild to moderate UC, but its high rate of side effects limits its use. Although sulfasalazine is more effective at higher doses, at 6 or 8 g/d up to 30% of patients experience allergic reactions or intolerable side effects such as headache, anorexia, nausea, and vomiting that are attributable to the sulfapyridine moiety. Hypersensitivity reactions, independent of sulfapyridine levels, include rash, fever, hepatitis, agranulocytosis, hypersensitivity pneumonitis, pancreatitis, worsening of colitis, and reversible sperm abnormalities. Sulfasalazine can also impair folate absorption, and patients should be given folic acid supplements.

Newer sulfa-free aminosalicylate preparations deliver increased amounts of the pharmacologically active ingredient of sulfasalazine (5-ASA, mesalamine) to the site of active bowel disease while limiting systemic toxicity. Peroxisome proliferator activated receptor gamma; (PPAR-gamma;) may mediate 5-ASA therapeutic action by decreasing nuclear localization of NF κB. Sulfa-free aminosalicylate formulations include alternative azo-bonded carriers, 5-ASA dimers, pH-dependent tablets, delayed-release and controlled-release preparations. Each has the same efficacy as sulfasalazine when equimolar concentrations are used. Olsalazine is composed of two 5-ASA radicals linked by an azo bond, which is split in the colon by bacterial reduction, and two 5-ASA molecules are released. Olsalazine is similar in effectiveness to sulfasalazine in treating UC, but up to 17% of patients experience nonbloody diarrhea caused by increased secretion of fluid in the small bowel. Balsalazide contains an azo bond binding mesalamine to the carrier molecule 4-aminobenzoyl-β-alanine; it is effective in the colon.

Asacol is an enteric-coated form of mesalamine with the 5-ASA being released at pH >7. The disintegration of

Asacol is variable, with complete breakup of the tablet occurring in many different parts of the gut ranging from the small intestine to the splenic flexure; it has increased gastric residence when taken with a meal. Pentasa is another mesalamine formulation that uses an ethylcellulose coating to allow water absorption into small beads containing the mesalamine. Water dissolves the 5-ASA, which then diffuses out of the bead into the lumen. Disintegration of the capsule occurs in the stomach. The microspheres then disperse throughout the entire tract from the small intestine through the distal colon in both fasted and fed conditions. Additional formulations of mesalamine continue to be developed. A once-a-day formulation of mesalamine [Multi-Matrix System (MMX), marketed in the United States as Lialda] is designed to release mesalamine in the colon. The MMX technology incorporates mesalamine into a lipophilic matrix within a hydrophilic matrix encapsulated in a polymer resistant to degradation at a low pH (<7) to delay release throughout the colon. The safety profile appears to be comparable to other 5-ASA formulations. A formulation containing encapsulated mesalamine granules (Apriso) was approved for use in the United States. Apriso delivers mesalamine to the terminal ileum and colon via a proprietary extended-release mechanism (Intellicor). The outer coating (Eudragit L) dissolves at a pH >6. In addition, there is a polymer matrix core that aids in sustained release throughout the colon. Since Lialda and Apriso are given once daily, an anticipated benefit is improved compliance compared with two to four daily doses required for other mesalamine preparations. Unencapsulated versions of mesalamine (Salofalkreg; Granu-Stix) have been in use in Europe for induction and maintenance of remission for several years.

Appropriate doses of Asacol and other 5-ASA compounds are shown in **Table 295-7**. Some 50–75% of patients with mild to moderate UC improve when treated with 5-ASA doses equivalent to 2 g/d of mesalamine; the dose response continues up to at least 4.8 g/d. As a general rule, 5-ASA agents act within 2–4 weeks. 5-ASA doses equivalent to 1.5–4 g/d of mesalamine maintain remission in 50–75% of patients with UC.

Table 295-7 Oral 5-ASA Preparations

Preparation	Formulation	Delivery	Dosing Per Day
Azo-Bond			
Sulfasalazine (500 mg) (Azulfidine)	Sulfapyridine-5-ASA	Colon	3–6 g (acute)2–4 g (maintenance)
Olsalazine (250 mg) (Dipentum)	5-ASA-5-ASA	Colon	1–3 g
Balsalazide (750 mg) (Colazal)	Aminobenzoyl-alanine-5-ASA	Colon	6.75–9 g
Delayed-Release			
Mesalamine (400, 800 mg) (Asacol)	Eudragit S (pH 7)	Distal ileum-colon	2.4–4.8 g (acute)1.6–4.8 g (maintenance)
Mesalamine (1.2 g) (Lialda)	MMX mesalamine (SPD476)	Ileum-colon	2.4–4.8 g
Controlled-Release			
Mesalamine (250, 500, 1000 mg) (Pentasa)	Ethylcellulose microgranules	Stomach-colon	2–4 g (acute)1.5–4 g (maintenance)
Delayed and Extended-Release			
Mesalamine (.375 g) (Apriso)	Intellicor extended-release mechanism	Ileum-colon	1.5 g (maintenance)

Topical mesalamine enemas are effective in mild-to-moderate distal UC. Clinical response occurs in up to 80% of UC patients with colitis distal to the splenic flexure. Combination therapy with mesalamine in both oral and enema form is more effective than either treatment alone for both distal and extensive UC. Mesalamine suppositories are effective in treating proctitis.

GLUCOCORTICOIDS

The majority of patients with moderate-to-severe UC benefit from oral or parenteral glucocorticoids. Prednisone is usually started at doses of 40–60 mg/d for active UC that is unresponsive to 5-ASA therapy. Parenteral glucocorticoids may be administered as hydrocortisone, 300 mg/d, or methylprednisolone, 40–60 mg/d. Topically applied glucocorticoids are also beneficial for distal colitis and may serve as an adjunct in those who have rectal involvement plus more proximal disease. Hydrocortisone enemas or foam may control active disease, although they have no proven role as maintenance therapy. These glucocorticoids are significantly absorbed from the rectum and can lead to adrenal suppression with prolonged administration. Topical 5-ASA therapy is more effective than topical steroid therapy in the treatment of distal UC.

Glucocorticoids are also effective for treatment of moderate-to-severe CD and induce a 60–70% remission rate compared to a 30% placebo response. The systemic effects of standard glucocorticoid formulations have led to the development of more potent formulations that are less well-absorbed and have increased first-pass metabolism. Controlled ileal-release budesonide has been nearly equal to prednisone for ileocolonic CD with fewer glucocorticoid side effects. Budesonide is used for 2–3 months at a dose of 9 mg/d, then tapered. Budesonide 6 mg/d is effective in reducing relapse rates at 3–6 months but not at 12 months in CD patients with a medically induced remission.

Glucocorticoids play no role in maintenance therapy in either UC or CD. Once clinical remission has been induced, they should be tapered according to the clinical activity, normally at a rate of no more than 5 mg/week. They can usually be tapered to 20 mg/d within 4–5 weeks but often take several months to be discontinued altogether. The side effects are numerous, including fluid retention, abdominal striae, fat redistribution, hyperglycemia, subcapsular cataracts, osteonecrosis, osteoporosis, myopathy, emotional

disturbances, and withdrawal symptoms. Most of these side effects, aside from osteonecrosis, are related to the dose and duration of therapy.

ANTIBIOTICS

Antibiotics have no role in the treatment of active or quiescent UC. However, pouchitis, which occurs in about a third of UC patients after colectomy and IPAA, usually responds to treatment with metronidazole and/or ciprofloxacin.

Metronidazole is effective in active inflammatory, fistulous, and perianal CD and may prevent recurrence after ileal resection. The most effective dose is 15–20 mg/kg per day in three divided doses; it is usually continued for several months. Common side effects include nausea, metallic taste, and disulfiram-like reaction. Peripheral neuropathy can occur with prolonged administration (several months) and on rare occasions is permanent despite discontinuation. Ciprofloxacin (500 mg bid) is also beneficial for inflammatory, perianal, and fistulous CD but has recently been associated with Achilles tendinitis and rupture. Both ciprofloxacin and metronidazole antibiotics can be used as first-line drugs for short periods of time in active inflammatory, fistulizing and perianal CD.

AZATHIOPRINE AND 6-MERCAPTOPURINE

Azathioprine and 6-mercaptopurine (6-MP) are purine analogues commonly employed in the management of glucocorticoid-dependent IBD. Azathioprine is rapidly absorbed and converted to 6-MP, which is then metabolized to the active end product, thioinosinic acid, an inhibitor of purine ribonucleotide synthesis and cell proliferation. These agents also inhibit the immune response. Efficacy can be seen as early as 3–4 weeks but can take up to 4–6 months. Adherence can be monitored by measuring the levels of 6-thioguanine and 6-methyl-mercaptopurine, end products of 6-MP metabolism. Azathioprine (2–3 mg/kg per day) or 6-MP (1–1.5 mg/kg per day) have been employed successfully as glucocorticoid-sparing agents in up to two-thirds of UC and CD patients previously unable to be weaned from glucocorticoids. The role of these immunomodulators as maintenance therapy in UC and CD and for treating active perianal disease and fistulas in CD appears promising. In addition, 6-MP or azathioprine is effective for postoperative prophylaxis of CD.

Although azathioprine and 6-MP are usually well tolerated, pancreatitis occurs in 3–4% of patients, typically presents within the first few weeks of therapy, and is completely reversible when the drug is stopped. Other side effects include nausea, fever, rash, and hepatitis. Bone marrow suppression (particularly leukopenia) is dose-related and often delayed, necessitating regular monitoring of the complete blood cell count (CBC). Additionally, 1 in 300 individuals lacks thiopurine methyltransferase, the enzyme responsible for drug metabolism; an additional 11% of the population are heterozygotes with intermediate enzyme activity. Both are at increased risk of toxicity because of increased accumulation of thioguanine metabolites. Although 6-thioguanine and 6-methylmercaptopurine levels can be followed to determine correct drug dosing and reduce toxicity, weight-based dosing is an acceptable alternative. CBCs and liver function tests should be monitored frequently regardless of dosing strategy. IBD patients treated with azathioprine/6-MP are at a fourfold increased risk of developing a lymphoma. This increased risk could be a result of the medications, the underlying disease, or both.

METHOTREXATE

Methotrexate (MTX) inhibits dihydrofolate reductase, resulting in impaired DNA synthesis. Additional anti-inflammatory properties may be related to decreased IL-1 production. Intramuscular (IM) or subcutaneous (SC) MTX (25 mg/week) is effective in inducing remission and reducing glucocorticoid dosage; 15 mg/week is effective in maintaining remission in active CD. Potential toxicities include leukopenia and hepatic fibrosis, necessitating periodic evaluation of CBCs and liver enzymes. The role of liver biopsy in patients on long-term MTX is uncertain but is probably limited to those with increased liver enzymes. Hypersensitivity pneumonitis is a rare but serious complication of therapy.

CYCLOSPORINE

Cyclosporine (CSA) is a lipophilic peptide with inhibitory effects on both the cellular and humoral immune systems. CSA blocks the production of IL-2 by T-helper lymphocytes. CSA binds to cyclophilin, and this complex inhibits calcineurin, a cytoplasmic phosphatase enzyme involved in the activation of T cells. CSA also indirectly inhibits B cell function by blocking helper T cells. CSA has a more rapid onset of action than 6-MP and azathioprine.

CSA is most effective when given at 2–4 mg/kg per day IV in severe UC that is refractory to IV glucocorticoids, with 82% of patients responding. CSA can be an alternative to colectomy. The long-term success of oral CSA is not as dramatic, but if patients are started on 6-MP or azathioprine at the time of hospital discharge, remission can be maintained. For the 2 mg/kg dose, levels as measured by monoclonal radioimmunoassay or by the high performance liquid chromatography assay should be maintained between 150 and 350 ng/mL.

CSA may cause significant toxicity; renal function should be monitored frequently. Hypertension, gingival hyperplasia, hypertrichosis, paresthesias, tremors, headaches, and electrolyte abnormalities are common side effects. Creatinine elevation calls for dose reduction or discontinuation. Seizures may also complicate therapy, especially if the patient is hypomagnesemic or if serum cholesterol levels are <3.1 mmol/L (<120 mg/dL). Opportunistic infections, most notably *Pneumocystis carinii* pneumonia, may occur with combination immunosuppressive treatment; prophylaxis should be given. Major adverse events occurred in 15% of patients in one large study including nephrotoxicity not responding to dose adjustment, serious infections, seizures, anaphylaxis, and death of two patients. This high incidence suggests that vigorous monitoring by experienced clinicians at tertiary care centers may be required.

TACROLIMUS

Tacrolimus is a macrolide antibiotic with immunomodulatory properties similar to CSA. It is 100 times as potent as CSA and is not dependent on bile or mucosal integrity for absorption. These pharmacologic properties enable tacrolimus to have good oral absorption despite proximal small bowel Crohn's involvement. It has shown efficacy in children with refractory IBD and in adults with extensive involvement of the small bowel. It is also effective in adults with steroid-dependent or refractory UC and CD as well as refractory fistulizing CD.

BIOLOGIC THERAPIES

Biologic therapy is often reserved for moderately to severely ill patients with Crohn's disease, who have failed other therapies. Patients who respond to biologic therapies enjoy an improvement in clinical symptoms, a better quality of life, less disability, fatigue and depression, and fewer surgeries and hospitalizations.

Anti-TNF Therapy

The first biologic therapy approved for Crohn's disease was infliximab, a chimeric IgG1 antibody against TNF-alpha, which is now also approved for treatment of moderately to severely active ulcerative colitis. Of active CD patients refractory to glucocorticoids, 6-MP, or 5-ASA, 65% will respond to IV infliximab (5 mg/kg); one-third will enter complete remission. The ACCENT I (A Crohn's Disease Clinical Trial Evaluating Infliximab in a New Long Term Treatment Regimen) study showed that of the patients who experience an initial response, 40% of these will maintain remission for at least 1 year with repeated infusions of infliximab every 8 weeks.

Infliximab is also effective in CD patients with refractory perianal and enterocutaneous fistulas, with the ACCENT II trial showing a 68% response rate (50% reduction in fistula drainage) and a 50% complete remission rate. Reinfusion, typically every 8 weeks, is necessary to continue therapeutic benefits in many patients.

The development of antibodies to infliximab (ATI) is associated with an increased risk of infusion reactions and a decreased response to treatment. Current practice does not include giving on-demand or episodic infusions rather than periodic (every 8 weeks) infusions because patients are more likely to develop ATI. ATI are generally present when the quality of response or the response duration to infliximab infusion decreases. Decreasing the dosing intervals or increasing the dosage to 10 mg/kg may restore the efficacy of the drug.

The SONIC (Study of Biologic and Immunomodulator-Naïve Patients with Crohn's Disease) Trial compared infliximab plus azathioprine, infliximab alone and azathioprine alone in immunomodulator and biologic naïve patients with moderate-to-severe Crohn's disease. At one year, of 508 randomized patients, the infliximab plus azathioprine group exhibited a steroid-free remission rate of 46% compared with 35% (infliximab alone) and 24% (azathioprine alone). There was also increased complete mucosal healing at week 26 with the combined approach relative to either infliximab or azathioprine alone (44% vs. 30% vs. 17%). The adverse events were equal between groups.

The annual risk of lymphoma (e.g., Hodgkin's and non-Hodgkin's lymphoma) with infliximab has been estimated to be anywhere from 5:10,000 to 20:10,000. The annual risk of lymphoma in the general population is 2:10,000. Forty-eight cases of malignancy were identified by the FDA in children and adolescents with the use of TNF blockers. Etanercept and infliximab were the only TNF blockers included in the analysis. Of the 48 cases, about 50% were lymphomas. Other malignancies such as leukemia, melanoma, and solid organ tumors were reported; malignancies rarely seen in children (e.g., leiomyosarcoma, hepatic malignancies, and renal cell carcinoma) were also observed. Of note, most of these cases (88%) were receiving other immunosuppressive medications (e.g., azathioprine and methotrexate).

Hepatosplenic T cell lymphoma is a nearly universally fatal lymphoma in patients with Crohn's disease. At least 12 cases involved immunomodulators alone, and 19 cases received combination therapy. There have been three reports in patients taking adalimumab alone without an immunomodulator. Patients tend to be young and almost all male.

The FDA also reviewed 147 postmarketing reports of leukemia (including acute myeloid leukemia, chronic lymphocytic leukemia, and chronic myeloid leukemia) and 69 cases of new-onset psoriasis (including pustular, palmoplantar) occurring in patients using TNF blockers. The FDA concluded that there is a possible association with both leukemia and new-onset psoriasis with the use of TNF blockers.

Other morbidities of infliximab include acute infusion reactions and severe serum sickness. All of the anti-TNF drugs are associated with an increased risk of infections, particularly reactivation of latent tuberculosis and opportunistic fungal infections including disseminated histoplasmosis and coccidioidomycosis. Rarely, infliximab and the other anti-TNF drugs have been associated with optic neuritis, seizures, new-onset or exacerbation of clinical symptoms, and radiographic evidence of central nervous system demyelinating disorders, including multiple sclerosis. They may exacerbate symptoms in patients with New York Heart Association functional class III/IV heart failure.

Infliximab has also shown efficacy in UC. In two large randomized, placebo-controlled trials, 37–49% of patients responded to infliximab and 22% and 20% of patients were able to maintain remission after 30 and 54 weeks, respectively. Patients received infliximab at 0, 2, and 6 weeks and then every 8 weeks until the end of the study.

Some patients losing response or not tolerating infliximab can be switched to adalimumab or certolizumab pegol. The GAIN (Gauging Infliximab Efficacy in Infliximab Non-Responders) trial evaluated patients who were previously treated with infliximab and became intolerant or who initially responded and lost response. Three-hundred and twenty-five patients were randomized to adalimumab or placebo. At 4 weeks, 21% of the adalimumab group and 7% of the placebo group were in remission. In clinical practice, this remission rate in the adalimumab group increases over time with a dose increase to 40 mg weekly instead of every other week.

Adalimumab is a recombinant human monoclonal IgG1 antibody containing only human peptide sequences

and is injected subcutaneously. Adalimumab binds TNF and neutralizes its function by blocking the interaction between TNF and its cell-surface receptor. Therefore, it seems to have a similar mechanism of action to infliximab but with less immunogenicity. Adalimumab has been approved for treatment of moderate to severe CD. CHARM (Crohn's Trial of the Fully Human Adalimumab for Remission Maintenance) is an adalimumab maintenance study in patients who responded to adalimumab induction therapy. About 50% of the patients in this trial were previously treated with infliximab. Remission rates ranged from 42–48% of infliximab naïve patients at 1 year compared with remission rates of 31–34% in the patients who had previously received infliximab. Certolizumab pegol is a PEGylated form of an anti-TNF antibody administered SC once monthly. SC certolizumab pegol was effective for induction of clinical response in patients with active inflammatory CD. In the PRECISE II (The PEGylated Antibody Fragment Evaluation in Crohn's Disease) trial of maintenance therapy with certolizumab in patients who responded to certolizumab induction, the results were similar to the CHARM trial. At week 26, the subgroup of patients who were infliximab naïve had a response of 69% as compared to 44% in patients who had previously received infliximab.

At least one-third of patients do not respond to infliximab, but no published controlled trial has examined infliximab nonresponders for a response to other anti-TNF agents. If a patient does not have an initial response to any anti-TNF therapy, currently it must be considered futile to try another. Before the approval of natalizumab, the only option for this group of patients was surgery.

Natalizumab

Integrins are expressed on the surface of leukocytes and serve as mediators of leukocyte adhesion to vascular endothelium. Alpha4 (α4) integrin along with its beta1 (β1) or beta7 (β7) subunit interact with endothelial ligands, termed adhesion molecules or vascular addressins. Interaction between α4β7 and mucosal addressin cellular adhesion molecule (MAdCAM-1) is important in lymphocyte tracking to gut mucosa. Natalizumab is a recombinant humanized immunoglobulin G4 antibody against α4 integrin that is effective in the induction and maintenance of remission in CD patients. It was approved February 2008 for the treatment of patients with CD refractory or intolerant to anti-TNF therapy. In the ENACT-2 (Evaluation of Natalizumab in Active Crohn's Disease Therapy) study, 354 patients who had a response to natalizumab in ENACT-1 were enrolled into maintenance therapy with an infusion of natalizumab or placebo every 4 weeks through week 56. Natalizumab patients were more likely to have a response (61% vs. 28% placebo) and remission 44% versus 26% placebo. However, 3 cases of progressive multifocal leukoencephalopathy (PML) associated with the JC polyoma virus in the clinical trials and 102 cases in the postmarketing setting have been reported to date. One case occurred in a patient with Crohn's disease and 104 in patients with multiple sclerosis. The annual risk of PML associated with natalizumab is approximately 1:1000. Patients and caregivers must now adhere to the TOUCH Treatment Program, which details strict criteria including no concomitant 6-MP, azathioprine or MTX, no glucocorticoids for longer than 6 months, the signing of consent forms and a monthly check by nurses for symptoms of PML.

THERAPIES IN DEVELOPMENT

Other therapies currently in development include monoclonal antibodies against IL-12 and IL-23. IL-12, derived from intestinal antigen presenting cells, initiates T_H1 mediated inflammation. IL-23 is a cytokine composed of a unique p19 subunit together with the p40 subunit of IL-12, which is also upregulated in CD mucosa and promotes T_H17 cells and inhibits T regulatory cells. Therefore, both IL-12 and IL-23 biologic activity can be inhibited by neutralizing IL-12 p40 with specific antibodies. The discovery of IL-23R as an IBD susceptibility gene strengthens the case for the use of IL-23 directed immunotherapy in IBD. Clinical trials are under way. Other promising therapies included those directed at IL-6 and the class of selective adhesion molecule inhibitors (e.g., anti-α4β7 and anti-MadCAM1 antibodies).

NUTRITIONAL THERAPIES

Dietary antigens may stimulate the mucosal immune response. Patients with active CD respond to bowel rest, along with TPN. Bowel rest and TPN are as effective as glucocorticoids at inducing remission of active CD but are not effective as maintenance therapy. Enteral nutrition in the form of elemental or peptide-based preparations is also as effective as glucocorticoids or TPN, but these diets are not palatable. Enteral diets may provide the small intestine with nutrients vital to cell growth and do not have the complications of TPN. In contrast to CD, dietary intervention does not reduce inflammation in UC. Standard medical management of UC and CD is shown in **Fig. 295-10**.

Figure 295-10

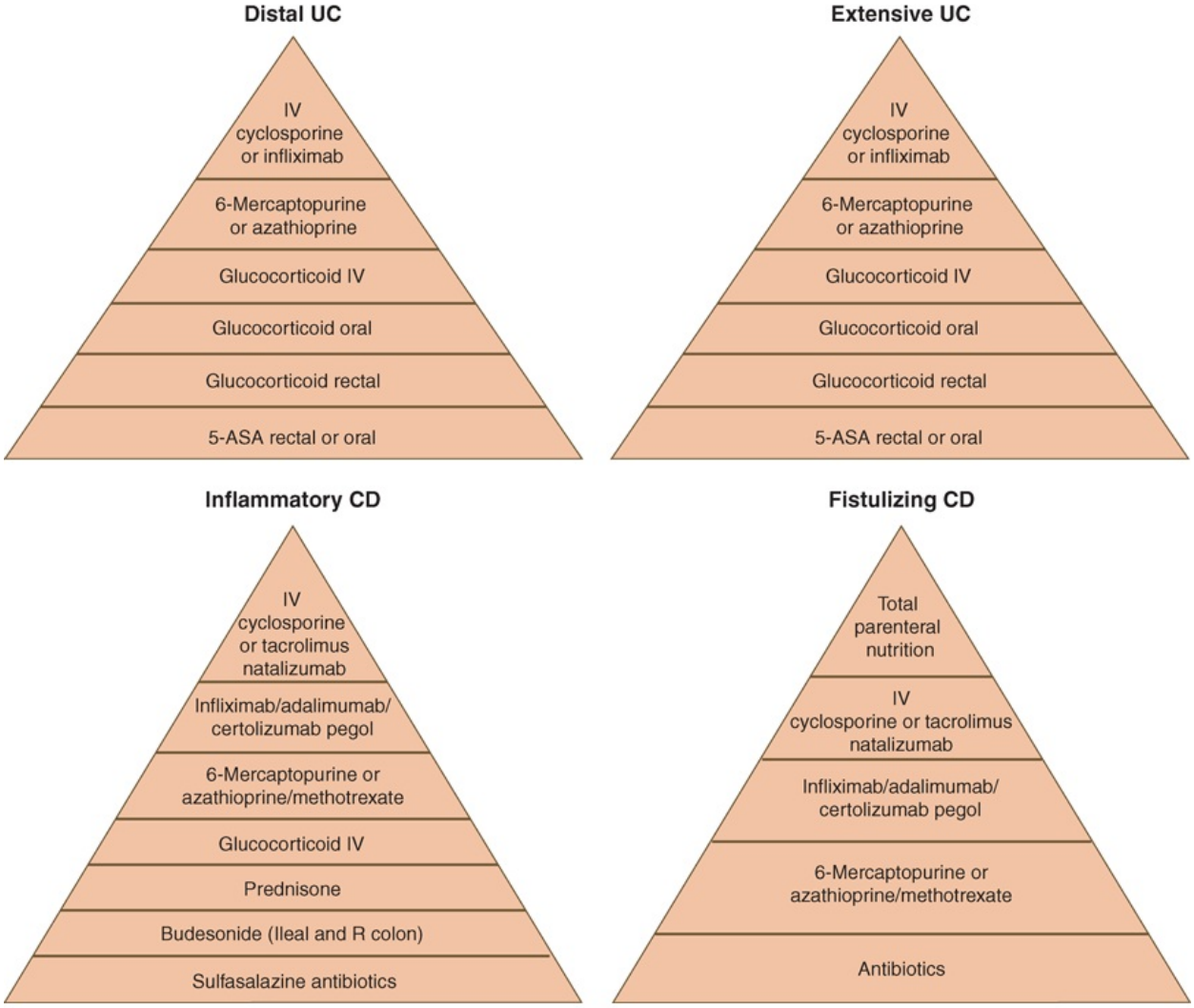

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Medical management of IBD. 5-ASA, 5-aminosalicylic acid; CD, Crohn's disease; UC, ulcerative colitis.

Surgical Therapy

ULCERATIVE COLITIS

Nearly one-half of patients with extensive chronic UC undergo surgery within the first 10 years of their illness. The indications for surgery are listed in **Table 295-8**. Morbidity is about 20% in elective, 30% for urgent, and 40% for emergency proctocolectomy. The risks are primarily hemorrhage, contamination and sepsis, and neural injury. The operation of choice is an IPAA.

Table 295-8 Indications for Surgery

Ulcerative Colitis	Crohn's Disease
Intractable disease	Small Intestine
Fulminant disease	Stricture and obstruction unresponsive to medical therapy
Toxic megacolon	Massive hemorrhage
Colonic perforation	Refractory fistula
Massive colonic hemorrhage	Abscess
Extracolonic disease	Colon and rectum
Colonic obstruction	Intractable disease
Colon cancer prophylaxis	Fulminant disease
Colon dysplasia or cancer	Perianal disease unresponsive to medical therapy
	Refractory fistula
	Colonic obstruction
	Cancer prophylaxis
	Colon dysplasia or cancer

Because UC is a mucosal disease, the rectal mucosa can be dissected and removed down to the dentate line of the anus or about 2 cm proximal to this landmark. The ileum is fashioned into a pouch that serves as a neorectum. This ileal pouch is then sutured circumferentially to the anus in an end-to-end fashion. If performed carefully, this operation preserves the anal sphincter and maintains continence. The overall operative morbidity is 10%, with the major complication being bowel obstruction. Pouch failure necessitating conversion to permanent ileostomy occurs in 5–10% of patients. Some inflamed rectal mucosa is usually left behind, and thus endoscopic surveillance is necessary. Primary dysplasia of the ileal mucosa of the pouch has occurred rarely.

Patients with IPAA usually have about 6–10 bowel movements a day. On validated quality-of-life indices, they report better performance in sports and sexual activities than ileostomy patients. The most frequent complication of IPAA is pouchitis in about 30–50% of patients with UC. This syndrome consists of increased stool frequency, watery stools, cramping, urgency, nocturnal leakage of stool, arthralgias, malaise, and fever. Pouch biopsies may distinguish true pouchitis from underlying CD. Although pouchitis usually responds to antibiotics, 3–5% of patients remain refractory and may require steroids, immunomodulators, anti-TNF therapy or even pouch removal. A highly concentrated probiotic preparation with four strains of *Lactobacillus,* three strains of *Bifidobacterium,* and one strain of *Streptococcus salivarius* can prevent the recurrence of pouchitis when taken daily.

Crohn's Disease

Most patients with CD require at least one operation in their lifetime. The need for surgery is related to duration of disease and the site of involvement. Patients with small-bowel disease have an 80% chance of requiring surgery. Those with colitis alone have a 50% chance. Surgery is an option only when medical treatment has failed or complications dictate its necessity. The indications for surgery are shown in Table 295-8.

Small Intestinal Disease

Because CD is chronic and recurrent, with no clear surgical cure, as little intestine as possible is resected. Current surgical alternatives for treatment of obstructing CD include resection of the diseased segment and strictureplasty. Surgical resection of the diseased segment is the most frequently performed operation, and in most cases primary anastomosis can be done to restore continuity. If much of the small bowel has already been resected and the strictures are short, with intervening areas of normal mucosa, strictureplasties should be done to avoid a functionally insufficient length of bowel. The strictured area of intestine is incised longitudinally and the incision sutured transversely, thus widening the narrowed area. Complications of strictureplasty include prolonged ileus, hemorrhage, fistula, abscess, leak, and restricture.

There is evidence that mesalamine, nitro-imidazole antibiotics, 6-MP/azathioprine and infliximab are all superior to placebo for the prevention of postoperative recurrence of Crohn's disease. Mesalamine is the least effective and the side effects of the nitro-imidazole antibiotics limit their use. Risk factors for early recurrence of disease include cigarette smoking, penetrating disease (internal fistulas, abscesses or other evidence of penetration through the wall of the bowel), early recurrence since a previous surgery, multiple surgeries or a young age at the time of the first surgery. Aggressive postoperative treatment with 6-MP/azathioprine or infliximab should be considered for this group of patients. It is also recommended to evaluate for endoscopic recurrence of Crohn's disease via a colonoscopy, if possible, 6 months after surgery.

Colorectal Disease

A greater percentage of patients with Crohn's colitis require surgery for intractability, fulminant disease, and anorectal disease. Several alternatives are available, ranging from the use of a temporary loop ileostomy to resection of segments of diseased colon or even the entire colon and rectum. For patients with segmental involvement, segmental colon resection with primary anastomosis can be performed. In 20–25% of patients with extensive colitis, the rectum is spared sufficiently to consider rectal preservation. Most surgeons believe that an IPAA is contraindicated in CD due to the high incidence of pouch failure. A diverting colostomy may help heal severe perianal disease or rectovaginal fistulas, but disease almost always recurs with reanastomosis. These patients often require a total proctocolectomy and ileostomy.

INFLAMMATORY BOWEL DISEASE AND PREGNANCY

Patients with quiescent UC and CD have normal fertility rates; the fallopian tubes can be scarred by the inflammatory process of CD, especially on the right side because of the proximity of the terminal ileum. In addition, perirectal, perineal, and rectovaginal abscesses and fistulae can result in dyspareunia. Infertility in men can be caused by sulfasalazine but reverses when treatment is stopped. In women who have had pouch surgery, most studies show that the fertility rate is reduced to about one-third of normal. This is due to scarring or occlusion of the fallopian tubes secondary to pelvic inflammation.

In mild or quiescent UC and CD, fetal outcome is nearly normal. Spontaneous abortions, stillbirths, and developmental defects are increased with increased disease activity, not medications. The courses of CD and UC during pregnancy mostly correlate with disease activity at the time of conception. Patients should be in remission for 6 months before conceiving. Most CD patients can deliver vaginally, but cesarean section may be the preferred route of delivery for patients with anorectal and perirectal abscesses and fistulas to reduce the likelihood of fistulas developing or extending into the episiotomy scar.

Sulfasalazine, mesalamine, and balsalazide are safe for use in pregnancy and nursing, but additional folate supplementation must be given with sulfasalazine. Topical 5-ASA agents are also safe during pregnancy and nursing. Glucocorticoids are generally safe for use during pregnancy and are indicated for patients with

moderate to severe disease activity. The amount of glucocorticoids received by the nursing infant is minimal. The safest antibiotics to use for CD in pregnancy for short periods of time (weeks, not months) are ampicillin and cephalosporin. Metronidazole can be used in the second or third trimester. Ciprofloxacin causes cartilage lesions in immature animals and should be avoided because of the absence of data on its effects on growth and development in humans.

6-MP and azathioprine pose minimal or no risk during pregnancy, but experience is limited. If the patient cannot be weaned from the drug or has an exacerbation that requires 6-MP/azathioprine during pregnancy, she should continue the drug with informed consent. Breast milk contained negligible levels of 6-MP/azathioprine when measured in a limited number of patients.

Little data exist on CSA in pregnancy. In a small number of patients with severe IBD treated with IV CSA during pregnancy, 80% of pregnancies were successfully completed without development of renal toxicity, congenital malformations, or developmental defects. However, because of the lack of data, CSA should probably be avoided unless the patient would otherwise require surgery. Methotrexate is contraindicated in pregnancy and nursing. No increased risk of stillbirths, miscarriages, or spontaneous abortions has been seen with infliximab, adalimumab or certolizumab, all class B drugs. The anti-TNF drugs are relatively safe in nursing as well because they do not pass into breast milk. Natalizumab is a class C drug, and there is limited data on pregnancy.

Surgery in UC should be performed only for emergency indications, including severe hemorrhage, perforation, and megacolon refractory to medical therapy. Total colectomy and ileostomy carry a 50–60% risk of postoperative spontaneous abortion. Fetal mortality is also high in CD requiring surgery. Patients with IPAAs have increased nighttime stool frequency during pregnancy that resolves postpartum. Transient small bowel obstruction or ileus has been noted in up to 8% of patients with ileostomies.

CANCER IN INFLAMMATORY BOWEL DISEASE

Ulcerative Colitis

Patients with long-standing UC are at increased risk for developing colonic epithelial dysplasia and carcinoma **(Fig. 295-11)**.

Figure 295-11

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Medium power view of low-grade dysplasia in a patient with chronic ulcerative colitis. Low-grade dysplastic crypts are interspersed among regenerating crypts. *(Courtesy of Dr. R Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)*

The risk of neoplasia in chronic UC increases with duration and extent of disease. The risk of cancer, as measured in tertiary referral centers, rises 0.5–1% per year after 8–10 years of disease in patients with pancolitis. The only prospective surveillance study reported a lower rate of cancer; 2.5% at 20 years of disease, 7.6% at 30 years of disease, and 10.8% at 40 years. The rates of colon cancer are higher than in the general population, and colonoscopic surveillance is the standard of care.

Annual or biennial colonoscopy with multiple biopsies is recommended for patients with >8–10 years of pancolitis or 12–15 years of left-sided colitis and has been widely employed to screen and survey for subsequent dysplasia and carcinoma. Risk factors for cancer in UC include long-duration disease, extensive disease, family history of colon cancer, PSC, a colon stricture, and the presence of postinflammatory pseudopolyps on colonoscopy.

Crohn's Disease

Risk factors for developing cancer in Crohn's colitis are long-duration and extensive disease, bypassed colon segments, colon strictures, PSC, and family history of colon cancer. The cancer risks in CD and UC are probably equivalent for similar extent and duration of disease. In patients with extensive Crohn's colitis, the cumulative risk of detecting an initial finding of any definite dysplasia or cancer after a negative screening colonoscopy is 25% by the tenth surveillance examination. The cumulative risk of detecting an initial finding of flat high-grade dysplasia (HGD) or cancer after a negative screening colonoscopy is 7% by the ninth

surveillance examination. Thus, the same endoscopic surveillance strategy used for UC is recommended for patients with chronic Crohn's colitis. A pediatric colonoscope can be used to pass narrow strictures in CD patients, but surgery should be considered in symptomatic patients with impassable strictures.

Management of Dysplasia and Cancer

Dysplasia can be flat or polypoid. If flat HGD is encountered on colonoscopic surveillance, the usual treatment for UC is colectomy and for CD is either colectomy or segmental resection. If flat low-grade dysplasia (LGD) is found (Fig. 295-11), most investigators recommend immediate colectomy. Adenomas may occur coincidently in UC and CD patients with chronic colitis and can be removed endoscopically provided that biopsies of the surrounding mucosa are free of dysplasia. New techniques such as high definition and magnification colonoscopes and dye sprays have increased the rate of dysplasia detection. In the future, endoscopists may be able to do targeted rather than segmental biopsies in patients with chronic Crohn's or ulcerative colitis.

IBD patients are also at greater risk for other malignancies. Patients with CD may have an increased risk of non-Hodgkin's lymphoma, leukemia, and myelodysplastic syndromes. Severe chronic, complicated perianal disease in CD patients may be associated with an increased risk of cancer in the lower rectum and anal canal (squamous cell cancers). Although the absolute risk of small-bowel adenocarcinoma in CD is low (2.2% at 25 years in one study), patients with long-standing, extensive, small-bowel disease should consider screening.

FURTHER READINGS

Farraye FA et al: AGA medical position statement on the diagnosis and management of colorectal neoplasia in inflammatory bowel disease. Gastroenterology 138:738, 2010[PMID: 20141808] [Full Text]

Colombel JF et al: Infliximab, azathioprine, or combination therapy for Crohn's disease. N Engl J Med 362:1383, 2010[PMID: 20393175] [Full Text]

Glocker EO et al: Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. N Engl J Med 361:2033, 2009[PMID: 19890111] [Full Text]

Kane SV, Acquah LA: Placental transport of immunoglobulins: A clinical review for gastroenterologists who prescribe therapeutic monoclonal antibodies to women during conception and pregnancy. Am J Gastroenterol 104:228, 2009[PMID: 19098873] [Full Text]

Kaser A et al: Inflammatory bowel disease. Annu Rev Immunol 28:573, 2010[PMID: 20192811] [Full Text]

Kornbluth A, Sachar DB: Ulcerative colitis practice guidelines in adults: American College of Gastroenterology, Practice Parameters Committee. Am J Gastroenterol 105:501, 2010[PMID: 20068560] [Full Text]

Lichtenstein GR et al: Practice Parameters Committee of American College of Gastroenterology. Management of Crohn's disease in adults. Am J Gastroenterol 104:465, 2009[PMID: 19174807] [Full Text]

Moscandrew ME, Loftus EV Jr: Diagnostic advances in inflammatory bowel disease (imaging and laboratory). Curr Gastroenterol Rep 11:488, 2009[PMID: 19903425] [Full Text]

Rutgeerts P et al: Biological therapies for inflammatory bowel diseases. Gastroenterology 136:1182, 2009[PMID: 19249397] [Full Text]

Rutter MD et al: Thirty-year analysis of a colonoscopic surveillance program for neoplasia in ulcerative colitis. Gastroenterology 130:1030, 2006[PMID: 16618396] [Full Text]

Targan SR et al: International Efficacy of Natalizumab in Crohn's Disease Response and Remission (ENCORE) Trial Group. Natalizumab for the treatment of active Crohn's disease: Results of the ENCORE Trial. Gastroenterology 132:1672, 2007[PMID: 17484865] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

IRRITABLE BOWEL SYNDROME: INTRODUCTION

Irritable bowel syndrome (IBS) is a functional bowel disorder characterized by abdominal pain or discomfort and altered bowel habits in the absence of detectable structural abnormalities. No clear diagnostic markers exist for IBS, thus the diagnosis of the disorder is based on clinical presentation. In 2006, the Rome II criteria for the diagnosis of IBS were revised **(Table 296-1)**. Throughout the world, about 10–20% of adults and adolescents have symptoms consistent with IBS, and most studies show a female predominance. IBS symptoms tend to come and go over time and often overlap with other functional disorders such as fibromyalgia, headache, backache, and genitourinary symptoms. Severity of symptoms varies and can significantly impair quality of life, resulting in high health care costs. Advances in basic, mechanistic, and clinical investigations have improved our understanding of this disorder and its physiologic and psychosocial determinants. Altered gastrointestinal (GI) motility, visceral hyperalgesia, disturbance of brain-gut interaction, abnormal central processing, autonomic and hormonal events, genetic and environmental factors, and psychosocial disturbances are variably involved, depending on the individual. This progress may result in improved methods of treatment.

Table 296-1 Diagnostic Criteria for Irritable Bowel Syndrome[a]

Recurrent abdominal pain or discomfort[b] at least 3 days per month in the last 3 months associated with *two or more* of the following:

1. Improvement with defecation
2. Onset associated with a change in frequency of stool
3. Onset associated with a change in form (appearance) of stool

[a]Criteria fulfilled for the last 3 months with symptom onset at least 6 months prior to diagnosis.
[b]Discomfort means an uncomfortable sensation not described as pain. In pathophysiology research and clinical trials, a pain/discomfort frequency of at least 2 days a week during screening evaluation is required for subject eligibility. ***Source:*** Adapted from Longstreth et al.

Clinical Features

IBS is a disorder that affects all ages, although most patients have their first symptoms before age 45. Older individuals have a lower reporting frequency. Women are diagnosed with IBS two to three times as often as men and make up 80% of the population with severe IBS. As indicated in Table 296-1, pain or abdominal discomfort is a key symptom for the diagnosis of IBS. These symptoms should be improved with defecation and/or have their onset associated with a change in frequency or form of stool. Painless diarrhea or constipation does not fulfill the diagnostic criteria to be classified as IBS. Supportive symptoms that are not part of the diagnostic criteria include defecation straining, urgency or a feeling of incomplete bowel movement, passing mucus, and bloating.

ABDOMINAL PAIN

According to the current IBS diagnostic criteria, abdominal pain or discomfort is a prerequisite clinical feature of IBS. Abdominal pain in IBS is highly variable in intensity and location. It is frequently episodic and crampy, but it may be superimposed on a background of constant ache. Pain may be mild enough to be ignored or it may interfere with daily activities. Despite this, malnutrition due to inadequate caloric intake is exceedingly rare with IBS. Sleep deprivation is also unusual because abdominal pain is almost uniformly present only during waking hours. However, patients with severe IBS frequently wake repeatedly during the night; thus, nocturnal pain is a poor discriminating factor between organic and functional bowel disease. Pain is often exacerbated by eating or emotional stress and improved by passage of flatus or stools. In addition, female patients with IBS commonly report worsening symptoms during the premenstrual and menstrual phases.

ALTERED BOWEL HABITS

Alteration in bowel habits is the most consistent clinical feature in IBS. The most common pattern is constipation alternating with diarrhea, usually with one of these symptoms predominating. At first, constipation may be episodic, but eventually it becomes continuous and increasingly intractable to

treatment with laxatives. Stools are usually hard with narrowed caliber, possibly reflecting excessive dehydration caused by prolonged colonic retention and spasm. Most patients also experience a sense of incomplete evacuation, thus leading to repeated attempts at defecation in a short time span. Patients whose predominant symptom is constipation may have weeks or months of constipation interrupted with brief periods of diarrhea. In other patients, diarrhea may be the predominant symptom. Diarrhea resulting from IBS usually consists of small volumes of loose stools. Most patients have stool volumes of <200 mL. Nocturnal diarrhea does not occur in IBS. Diarrhea may be aggravated by emotional stress or eating. Stool may be accompanied by passage of large amounts of mucus. Bleeding is not a feature of IBS unless hemorrhoids are present, and malabsorption or weight loss does not occur.

Bowel pattern subtypes are highly unstable. In a patient population with ~33% prevalence rates of IBS-diarrhea predominant (IBS-D), IBS-constipation predominant (IBS-C), and IBS-mixed (IBS-M) forms, 75% of patients change subtypes and 29% switch between IBS-C and IBS-D over 1 year. The heterogeneity and variable natural history of bowel habits in IBS increase the difficulty of conducting pathophysiology studies and clinical trials.

GAS AND FLATULENCE

Patients with IBS frequently complain of abdominal distention and increased belching or flatulence, all of which they attribute to increased gas. Although some patients with these symptoms actually may have a larger amount of gas, quantitative measurements reveal that most patients who complain of increased gas generate no more than a normal amount of intestinal gas. Most IBS patients have impaired transit and tolerance of intestinal gas loads. In addition, patients with IBS tend to reflux gas from the distal to the more proximal intestine, which may explain the belching.

Some patients with bloating may also experience visible distention with increase in abdominal girth. Both symptoms are more common among female patients and in those with higher overall Somatic Symptom Checklist scores. IBS patients who experienced bloating alone have been shown to have lower thresholds for pain and desire to defecate compared to those with concomitant distention irrespective of bowel habit. When patients were grouped according to sensory threshold, hyposensitive individuals had distention significantly more than those with hypersensitivity and this was observed more in the constipation subgroup. This suggests that the pathogenesis of bloating and distention may not be the same.

UPPER GASTROINTESTINAL SYMPTOMS

Between 25 and 50% of patients with IBS complain of dyspepsia, heartburn, nausea, and vomiting. This suggests that other areas of the gut apart from the colon may be involved. Prolonged ambulant recordings of small-bowel motility in patients with IBS show a high incidence of abnormalities in the small bowel during the diurnal (waking) period; nocturnal motor patterns are not different from those of healthy controls. The overlap between dyspepsia and IBS is great. The prevalence of IBS is higher among patients with dyspepsia (31.7%) than among those who reported no symptoms of dyspepsia (7.9%). Conversely, among patients with IBS, 55.6% reported symptoms of dyspepsia. In addition, the functional abdominal symptoms can change over time. Those with predominant dyspepsia or IBS can flux between the two. Although the prevalence of functional gastrointestinal disorders is stable over time, the turnover in symptom status is high. Many episodes of symptom disappearance are due to subjects changing symptoms rather than total symptom resolution. Thus it is conceivable that functional dyspepsia and IBS are two manifestations of a single, more extensive digestive system disorder. Furthermore, IBS symptoms are prevalent in noncardiac chest pain patients, suggesting overlap with other functional gut disorders.

Pathophysiology

The pathogenesis of IBS is poorly understood, although roles of abnormal gut motor and sensory activity, central neural dysfunction, psychological disturbances, mucosal inflammation, stress, and luminal factors have been proposed.

GASTROINTESTINAL MOTOR ABNORMALITIES

Studies of colonic myoelectrical and motor activity under unstimulated conditions have not shown consistent abnormalities in IBS. In contrast, colonic motor abnormalities are more prominent under stimulated conditions in IBS. IBS patients may exhibit increased rectosigmoid motor activity for up to 3 h after eating. Similarly, inflation of rectal balloons both in IBS-D and IBS-C patients leads to marked and prolonged distention-evoked contractile activity. Recordings from the transverse, descending, and sigmoid colon showed that the motility index and peak amplitude of high-amplitude propagating contractions (HAPCs) in diarrhea-prone IBS patients were greatly increased compared to those in healthy subjects and were associated with rapid colonic transit and accompanied by abdominal pain.

VISCERAL HYPERSENSITIVITY

As with studies of motor activity, IBS patients frequently exhibit exaggerated sensory responses to visceral stimulation. Postprandial pain has been temporally related to entry of the food bolus into the cecum in 74% of patients. Rectal balloon inflation produces nonpainful and painful sensations at lower

volumes in IBS patients than in healthy controls without altering rectal tension, suggestive of visceral afferent dysfunction in IBS. Similar studies show gastric and esophageal hypersensitivity in patients with nonulcer dyspepsia and noncardiac chest pain, raising the possibility that these conditions have a similar pathophysiologic basis. Lipids lower the thresholds for the first sensation of gas, discomfort, and pain in IBS patients. Hence, postprandial symptoms in IBS patients may be explained in part by a nutrient-dependent exaggerated sensory component of the gastrocolonic response. In contrast to enhanced gut sensitivity, IBS patients do not exhibit heightened sensitivity elsewhere in the body. Thus, the afferent pathway disturbances in IBS appear to be selective for visceral innervation with sparing of somatic pathways. The mechanisms responsible for visceral hypersensitivity are still under investigation. It has been proposed that these exaggerated responses may be due to (1) increased end-organ sensitivity with recruitment of "silent" nociceptors; (2) spinal hyperexcitability with activation of nitric oxide and possibly other neurotransmitters; (3) endogenous (cortical and brainstem) modulation of caudad nociceptive transmission; and (4) over time, the possible development of long-term hyperalgesia due to development of neuroplasticity, resulting in permanent or semipermanent changes in neural responses to chronic or recurrent visceral stimulation **(Table 296-2)**.

Table 296-2 Proposed Mechanisms for Visceral Hypersensitivity

End-organ sensitivity	Long-term hyperalgesia
"Silent" nociceptors	Tonic cortical regulation
CNS modulation	Neuroplasticity
Cortex	
Brainstem	

Abbreviation: CNS, central nervous system.

CENTRAL NEURAL DYSREGULATION

The role of central nervous system (CNS) factors in the pathogenesis of IBS is strongly suggested by the clinical association of emotional disorders and stress with symptom exacerbation and the therapeutic response to therapies that act on cerebral cortical sites. Functional brain imaging studies such as MRI have shown that in response to distal colonic stimulation, the mid-cingulate cortex—a brain region concerned with attention processes and response selection—shows greater activation in IBS patients. Modulation of this region is associated with changes in the subjective unpleasantness of pain. In addition, IBS patients also show preferential activation of the prefrontal lobe, which contains a vigilance network within the brain that increases alertness. These may represent a form of cerebral dysfunction leading to the increased perception of visceral pain.

ABNORMAL PSYCHOLOGICAL FEATURES

Abnormal psychiatric features are recorded in up to 80% of IBS patients, especially in referral centers; however, no single psychiatric diagnosis predominates. Most of these patients demonstrated exaggerated symptoms in response to visceral distention, and this abnormality persists even after exclusion of psychological factors.

Psychological factors influence pain thresholds in IBS patients, as stress alters sensory thresholds. An association between prior sexual or physical abuse and development of IBS has been reported. Abuse is associated with greater pain reporting, psychological distress, and poor health outcome. Brain functional MRI studies show greater activation of the posterior and middle dorsal cingulate cortex, which is implicated in affect processing in IBS patients with a past history of sexual abuse.

Thus, patients with IBS frequently demonstrate increased motor reactivity of the colon and small bowel to a variety of stimuli and altered visceral sensation associated with lowered sensation thresholds. These may result from CNS—enteric nervous system dysregulation **(Fig. 296-1)**.

Figure 296-1

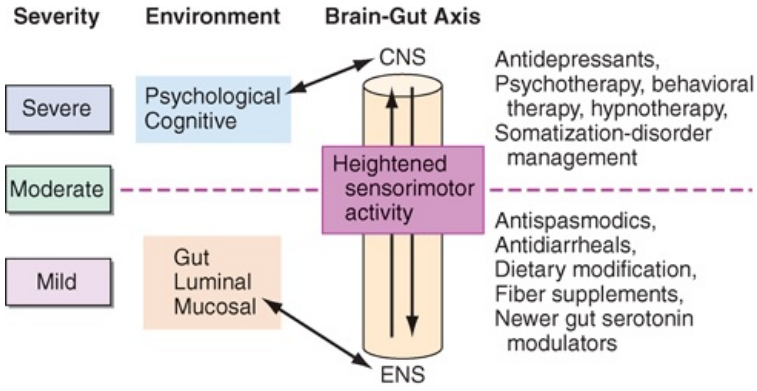

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Therapeutic targets for irritable bowel syndrome. Patients with mild to moderate symptoms usually have intermittent symptoms that correlate with altered gut physiology. Treatments include gut-acting pharmacologic agents such as antispasmodics, antidiarrheals, fiber supplements, and gut serotonin modulators. Patients who have severe symptoms usually have constant pain and psychosocial difficulties. This group of patients is best managed with antidepressants and other psychosocial treatments. CNS, central nervous system; ENS, enteric nervous system.

POST-INFECTIOUS IBS

IBS may be induced by GI infection. In an investigation of 544 patients with confirmed bacterial gastroenteritis, one-quarter developed IBS subsequently. Conversely, about a third of IBS patients experienced an acute "gastroenteritis-like" illness at the onset of their chronic IBS symptomatology. This group of "postinfective" IBS occurs more commonly in females and affects younger rather than older patients. Risk factors for developing post-infectious IBS include, in order of importance, prolonged duration of initial illness, toxicity of infecting bacterial strain, smoking, mucosal markers of inflammation, female gender, depression, hypochondriasis, and adverse-life events in the preceding 3 months. Age older than 60 years might protect against post-infectious IBS, whereas treatment with antibiotics has been associated with increased risk. The microbes involved in the initial infection are *Campylobacter*, *Salmonella*, and *Shigella*. Those patients with *Campylobacter* infection who are toxin-positive are more likely to develop postinfective IBS. Increased rectal mucosal enteroendocrine cells, T lymphocytes, and increased gut permeability are acute changes following *Campylobacter* enteritis that could persist for more than a year and may contribute to postinfective IBS.

IMMUNE ACTIVATION AND MUCOSAL INFLAMMATION

Some patients with IBS display persistent signs of low-grade mucosal inflammation with activated lymphocytes, mast cells, and enhanced expression of proinflammatory cytokines. These abnormalities may contribute to abnormal epithelial secretion and visceral hypersensitivity. Interestingly, clinical studies have shown increased intestinal permeability in patients with IBS-D. Psychological stress and anxiety can increase the release of proinflammatory cytokine and this in turn may alter intestinal permeability. This provides a functional link between psychological stress, immune activation, and symptom generation in patients with IBS.

ALTERED GUT FLORA

A high prevalence of small intestinal bacterial overgrowth in IBS patients has been noted based on positive lactulose hydrogen breath test. This finding, however, has been challenged by a number of other studies that found no increased incidence of bacterial overgrowth based on jejunal aspirate culture. Abnormal H_2 breath test can occur because of small bowel rapid transit and may lead to erroneous interpretation. Hence, the role of testing for small intestinal bacterial overgrowth in IBS patients remains unclear.

A number of studies found significant differences between the molecular profile of the fecal microbiota of IBS patients compared with that of healthy subjects. Several bacterial genera with *Lactobacillus* sequence appear to be absent from IBS, and *Collinsella* sequences were greatly reduced in this group of patients. Currently it is unclear whether such changes are causal, consequential, or merely the result of constipation and diarrhea. In addition, the stability of the changes in the microbiota needs to be determined.

ABNORMAL SEROTONIN PATHWAYS

The serotonin (5HT)-containing enterochromaffin cells in the colon are increased in a subset of IBS-D patients compared to healthy individuals or patients with ulcerative colitis. Furthermore, postprandial plasma 5HT levels were significantly higher in this group of patients compared to healthy controls. Since serotonin plays an important role in the regulation of GI motility and visceral perception, the increased release of serotonin may contribute to the postprandial symptoms of these patients and

provides a rationale for the use of serotonin antagonists in the treatment of this disorder.

Approach to the Patient: Irritable Bowel Syndrome

Because IBS is a disorder for which no pathognomonic abnormalities have been identified, its diagnosis relies on recognition of positive clinical features and elimination of other organic diseases. A careful history and physical examination are frequently helpful in establishing the diagnosis. Clinical features suggestive of IBS include the following: recurrence of lower abdominal pain with altered bowel habits over a period of time without progressive deterioration, onset of symptoms during periods of stress or emotional upset, absence of other systemic symptoms such as fever and weight loss, and small-volume stool without any evidence of blood.

On the other hand, the appearance of the disorder for the first time in old age, progressive course from time of onset, persistent diarrhea after a 48-h fast, and presence of nocturnal diarrhea or steatorrheal stools argue against the diagnosis of IBS.

Because the major symptoms of IBS—abdominal pain, abdominal bloating, and alteration in bowel habits—are common complaints of many GI organic disorders, the list of differential diagnoses is a long one. The quality, location, and timing of pain may be helpful to suggest specific disorders. Pain due to IBS that occurs in the epigastric or periumbilical area must be differentiated from biliary tract disease, peptic ulcer disorders, intestinal ischemia, and carcinoma of the stomach and pancreas. If pain occurs mainly in the lower abdomen, the possibility of diverticular disease of the colon, inflammatory bowel disease (including ulcerative colitis and Crohn's disease), and carcinoma of the colon must be considered. Postprandial pain accompanied by bloating, nausea, and vomiting suggests gastroparesis or partial intestinal obstruction. Intestinal infestation with *Giardia lamblia* or other parasites may cause similar symptoms. When diarrhea is the major complaint, the possibility of lactase deficiency, laxative abuse, malabsorption, celiac sprue, hyperthyroidism, inflammatory bowel disease, and infectious diarrhea must be ruled out. On the other hand, constipation may be a side effect of many different drugs, such as anticholinergic, antihypertensive, and antidepressant medications. Endocrinopathies such as hypothyroidism and hypoparathyroidism must also be considered in the differential diagnosis of constipation, particularly if other systemic signs or symptoms of these endocrinopathies are present. In addition, acute intermittent porphyria and lead poisoning may present in a fashion similar to IBS, with painful constipation as the major complaint. These possibilities are suspected on the basis of their clinical presentations and are confirmed by appropriate serum and urine tests.

Few tests are required for patients who have typical IBS symptoms and no alarm features. Unnecessary investigations may be costly and even harmful. The American Gastroenterological Association has delineated factors to be considered when determining the aggressiveness of the diagnostic evaluation. These include the duration of symptoms, the change in symptoms over time, the age and sex of the patient, the referral status of the patient, prior diagnostic studies, a family history of colorectal malignancy, and the degree of psychosocial dysfunction. Thus, a younger individual with mild symptoms requires a minimal diagnostic evaluation, while an older person or an individual with rapidly progressive symptoms should undergo a more thorough exclusion of organic disease. Most patients should have a complete blood count and sigmoidoscopic examination; in addition, stool specimens should be examined for ova and parasites in those who have diarrhea. In patients with persistent diarrhea not responding to simple anti-diarrhea agents, a sigmoid colon biopsy should be performed to rule out microscopic colitis. In those aged >40 years, an air-contrast barium enema or colonoscopy should also be performed. If the main symptoms are diarrhea and increased gas, the possibility of lactase deficiency should be ruled out with a hydrogen breath test or with evaluation after a 3-week lactose-free diet. Some patients with IBS-D may have undiagnosed celiac sprue. Because the symptoms of celiac sprue respond to a gluten-free diet, testing for celiac sprue in IBS may prevent years of morbidity and attendant expense. Decision-analysis studies show that serology testing for celiac sprue in patients with IBS-D has an acceptable cost when the prevalence of celiac sprue is >1% and is the dominant strategy when the prevalence is >8%. In patients with concurrent symptoms of dyspepsia, upper GI radiographs or esophagogastroduodenoscopy may be advisable. In patients with postprandial right upper quadrant pain, an ultrasonogram of the gallbladder should be obtained. Laboratory features that argue against IBS include evidence of anemia, elevated sedimentation rate, presence of leukocytes or blood in stool, and stool volume >200–300 mL/d. These findings would necessitate other diagnostic considerations.

Treatment: Irritable Bowel Syndrome

PATIENT COUNSELING AND DIETARY ALTERATIONS

Reassurance and careful explanation of the functional nature of the disorder and of how to avoid obvious food precipitants are important first steps in patient counseling and dietary change. Occasionally, a meticulous dietary history may reveal substances (such as coffee, disaccharides, legumes, and cabbage) that aggravate symptoms. Excessive fructose and artificial sweeteners, such as sorbitol or mannitol, may cause diarrhea, bloating, cramping or flatulence. As a therapeutic trial, patients should be encouraged to eliminate any foodstuffs that appear to produce symptoms. However patients should avoid nutritionally depleted diets. Patients with IBS-D anecdotally report symptom

improvement after initiating a low-carbohydrate diet. A prospective study has shown marked symptomatic improvement in stool frequency, consistency, pain scores, and quality of life following 4 weeks of a very-low-carbohydrate (CHO) diet (20 g CHO/day). This diet may be tried in IBS patients who report intolerance to certain carbohydrates.

Stool-Bulking Agents

High-fiber diets and bulking agents, such as bran or hydrophilic colloid, are frequently used in treating IBS. The water-holding action of fibers may contribute to increased stool bulk because of the ability of fiber to increase fecal output of bacteria. Fiber also speeds up colonic transit in most persons. In diarrhea-prone patients, whole-colonic transit is faster than average; however, dietary fiber can delay transit. Furthermore, because of their hydrophilic properties, stool-bulking agents bind water and thus prevent both excessive hydration and dehydration of stool. The latter observation may explain the clinical experience that a high-fiber diet relieves diarrhea in some IBS patients. Fiber supplementation with psyllium has been shown to reduce perception of rectal distention, indicating that fiber may have a positive effect on visceral afferent function.

The beneficial effects of dietary fiber on colonic physiology suggest that dietary fiber should be an effective treatment for IBS patients, but controlled trials of dietary fiber have produced variable results. This is not surprising since IBS is a heterogeneous disorder, with some patients being constipated and other having predominant diarrhea. Most investigations report increases in stool weight, decreases in colonic transit times, and improvement in constipation. Others have noted benefits in patients with alternating diarrhea and constipation, pain, and bloating. However, most studies observe no responses in patients with diarrhea- or pain-predominant IBS. It is possible that different fiber preparations may have dissimilar effects on selected symptoms in IBS. A cross-over comparison of different fiber preparations found that psyllium produced greater improvements in stool pattern and abdominal pain than bran. Furthermore, psyllium preparations tend to produce less bloating and distention. Despite the equivocal data regarding efficacy, most gastroenterologists consider stool-bulking agents worth trying in patients with IBS-C.

Antispasmodics

Clinicians have observed that anticholinergic drugs may provide temporary relief for symptoms such as painful cramps related to intestinal spasm. Although controlled clinical trials have produced mixed results, evidence generally supports beneficial effects of anticholinergic drugs for pain. A meta-analysis of 26 double-blind clinical trials of antispasmodic agents in IBS reported better global improvement (62%) and abdominal pain reductions (64%) compared to placebo (35% and 45%, respectively), suggesting efficacy in some patients. The drugs are most effective when prescribed in anticipation of predictable pain. Physiologic studies demonstrate that anticholinergic drugs inhibit the gastrocolic reflex; hence, postprandial pain is best managed by giving antispasmodics 30 min before meals so that effective blood levels are achieved shortly before the anticipated onset of pain. Most anticholinergics contain natural belladonna alkaloids, which may cause xerostomia, urinary hesitancy and retention, blurred vision, and drowsiness. They should be used in the elderly with caution. Some physicians prefer to use synthetic anticholinergics such as dicyclomine that have less effect on mucous membrane secretions and produce fewer undesirable side effects.

Antidiarrheal Agents

Peripherally acting opiate-based agents are the initial therapy of choice for IBS-D. Physiologic studies demonstrate increases in segmenting colonic contractions, delays in fecal transit, increases in anal pressures, and reductions in rectal perception with these drugs. When diarrhea is severe, especially in the painless diarrhea variant of IBS, small doses of loperamide, 2–4 mg every 4–6 h up to a maximum of 12 g/d, can be prescribed. These agents are less addictive than paregoric, codeine, or tincture of opium. In general, the intestines do not become tolerant of the antidiarrheal effect of opiates, and increasing doses are not required to maintain antidiarrheal potency. These agents are most useful if taken before anticipated stressful events that are known to cause diarrhea. However, not infrequently, a high dose of loperamide may cause cramping because of increases in segmenting colonic contractions. Another anti-diarrhea agent that may be used in IBS patients is the bile acid binder cholestyramine resin.

Antidepressant Drugs

In addition to their mood-elevating effects, antidepressant medications have several physiologic effects that suggest they may be beneficial in IBS. In IBS-D patients, the tricyclic antidepressant imipramine slows jejunal migrating motor complex transit propagation and delays orocecal and whole-gut transit, indicative of a motor inhibitory effect. Some studies also suggest that tricyclic agents may alter visceral afferent neural function.

A number of studies indicate that tricyclic antidepressants may be effective in some IBS patients. In a 2-month study of desipramine, abdominal pain improved in 86% of patients compared to 59% given placebo. Another study of desipramine in 28 IBS patients showed improvement in stool frequency, diarrhea, pain, and depression. When stratified according to the predominant symptoms, improvements were observed in IBS-D patients, with no improvement being noted in IBS-C patients. The beneficial

effects of the tricyclic compounds in the treatment of IBS appear to be independent of their effects on depression. The therapeutic benefits for the bowel symptoms occur faster and at a lower dosage. The efficacy of antidepressant agents in other chemical classes in the management of IBS is less well evaluated. In contrast to tricyclic agents, the selective serotonin reuptake inhibitor (SSRI) paroxetine accelerates orocecal transit, raising the possibility that this drug class may be useful in IBS-C patients. The SSRI citalopram blunts perception of rectal distention and reduces the magnitude of the gastrocolonic response in healthy volunteers. A small placebo-controlled study of citalopram in IBS patients reported reductions in pain. However, these findings could not be confirmed in another randomized controlled trial which showed that citalopram at 20 mg/day for 4 weeks was not superior to placebo in treating non-depressed IBS patients. Hence, the efficacy of SSRIs in the treatment of IBS needs further confirmation.

Antiflatulence Therapy

The management of excessive gas is seldom satisfactory, except when there is obvious aerophagia or disaccharidase deficiency. Patients should be advised to eat slowly and not chew gum or drink carbonated beverages. Bloating may decrease if an associated gut syndrome such as IBS or constipation is improved. If bloating is accompanied by diarrhea and worsens after ingesting dairy products, fresh fruits, vegetables, or juices, further investigation or a dietary exclusion trial may be worthwhile. Avoiding flatogenic foods, exercising, losing excess weight, and taking activated charcoal are safe but unproven remedies. Data regarding the use of surfactants such as simethicone are conflicting. Antibiotics may help in a subgroup of IBS patients with predominant symptoms of bloating. Beano, an over-the-counter oral β-glycosidase solution, may reduce rectal passage of gas without decreasing bloating and pain. Pancreatic enzymes reduce bloating, gas, and fullness during and after high-calorie, high-fat meal ingestion.

Modulation of Gut Flora

Antibiotic treatment benefits a subset of IBS patients. In a double-blind randomized placebo controlled study, neomycin dosed at 500 mg twice daily for 10 days was more effective than placebo at improving symptom scores among IBS patients. The non-absorbed oral antibiotic rifaximin is the most thoroughly studied antibiotic for the treatment of IBS. Patients receiving rifaximin at a dose of 400 mg three times daily experienced substantial improvement of global IBS symptoms over placebo. Rifaximin is the only antibiotic with demonstrated sustained benefit beyond therapy cessation in IBS patients. The drug has a favorable safety and tolerability profile compared with systemic antibiotics. However, currently there is still insufficient data to recommend routine use of this antibiotic in the treatment of IBS.

Since altered colonic flora may contribute to the pathogenesis of IBS, this has led to great interest in using probiotics to naturally alter the flora. *Bifidobacterium infantis* 35624 showed significant improvement in the composite score for abdominal pain, bloating/distention, and/or bowel movement compared with placebo in two placebo-controlled trials. Currently, there are inadequate data to comment on the efficacy of other probiotics.

Serotonin Receptor Agonist and Antagonists

Serotonin receptor antagonists have been evaluated as therapies for IBS-D. Serotonin acting on 5-HT3 receptors enhances the sensitivity of afferent neurons projecting from the gut. In humans, a 5-HT3 receptor antagonist such as alosetron reduces perception of painful visceral stimulation in IBS. It also induces rectal relaxation, increases rectal compliance, and delays colonic transit. Meta-analysis of 14 randomized controlled trials of alosetron or cilansetron showed that these antagonists are more effective than placebo in achieving global improvement in IBS symptoms and relief of abdominal pain and discomfort. These agents are more likely to cause constipation in IBS patients with diarrhea alternating with constipation. 0.2% of patients using $5HT_3$ antagonist developed ischemic colitis versus none in the control group. In postrelease surveillance, 84 cases of ischemic colitis were observed, including 44 cases that required surgery and 4 deaths. As a consequence, the medication was voluntarily withdrawn by the manufacturer in 2000. Alosetron has been reintroduced under a new risk-management program where patients have to sign a patient-physician agreement. This has significantly limited its usage.

Novel 5-HT4 receptor agonists such as tegaserod exhibit prokinetic activity by stimulating peristalsis. In IBS patients with constipation, tegaserod accelerated intestinal and ascending colon transit. Clinical trials involving >4000 IBS-C patients reported reductions in discomfort and improvements in constipation and bloating, compared to placebo. Diarrhea is the major side effect. However, tegaserod has been withdrawn from the market; a meta-analysis revealed an increase in serious cardiovascular events.

Chloride Channel Activators

Lubiprostone is a bicyclic fatty acid that stimulates chloride channels in the apical membrane of intestinal epithelial cells. Chloride secretion induces passive movement of sodium and water into the bowel lumen and improves bowel function. Oral lubiprostone was effective in the treatment of patients with constipation-predominant IBS in large phase II and phase III randomized double-blinded placebo-controlled multicenter trials. Responses were significantly greater in patients receiving lubiprostone 8 μ

g twice daily for 3 months than in those receiving placebo. In general, the drug was quite well tolerated. The major side effects are nausea and diarrhea. Lubiprostone is a new class of compounds for treatment of chronic constipation with or without IBS.

SUMMARY

The treatment strategy of IBS depends on the severity of the disorder **(Table 296-3)**. Most of the IBS patients have mild symptoms. They are usually cared for in primary care practices, have little or no psychosocial difficulties, and do not seek health care often. Treatment usually involves education, reassurance, and dietary/lifestyle changes. A smaller portion have moderate symptoms that are usually intermittent and correlate with altered gut physiology, e.g., worsened with eating or stress and relieved by defecation. Treatments include gut-acting pharmacologic agents such as antispasmodics, antidiarrheals, fiber supplements, and the newer gut serotonin modulators **(Table 296-4)**. A small proportion of IBS patients have severe and refractory symptoms, are usually seen in referral centers, and frequently have constant pain and psychosocial difficulties (Fig. 296-1). This group of patients is best managed with antidepressants and other psychological treatments (Table 296-4).

Table 296-3 Spectrum of Severity in IBS

	Mild	Moderate	Severe
Clinical Features			
Prevalence	70%	25%	5%
Correlations with gut physiology	+++	++	+
Symptoms constant	0	+	+++
Psychosocial difficulties	0	+	+++
Health care issues	+	++	+++
Practice type	Primary	Specialty	Referral

Table 296-4 Possible Drugs for a Dominant Symptom in IBS

Symptom	Drug	Dose
Diarrhea	Loperamide	2–4 mg when necessary/maximum 12 g/d
	Cholestyramine resin	4 g with meals
	Alosetron*	0.5–1 mg bid (for severe IBS, women)
Constipation	Psyllium husk	3–4 g bid with meals, then adjust
	Methylcellulose	2 g bid with meals, then adjust
	Calciumpolycarbophil	1 g qd to qid
	Lactulose syrup 70% sorbitol	10–20 g bid 15 mL bid
	Polyethylene glycol 3350	17 g in 250 mL water qd
	Lubiprostone (Amitiza)	24 mg bid
	Magnesium hydroxide	30–60 mL qd
Abdominal pain	Smooth-muscle relaxant	qd to qid ac
	Tricyclic antidepressants	Start 25–50 mg hs, then adjust
	Selective serotonin reuptake inhibitors	Begin small dose, increase as needed

*Available only in the United States.

Source: Adapted from Longstreth et al.

FURTHER READINGS

Ford AC et al: Efficacy of antidepressants and psychological therapies in irritable bowel syndrome: Systematic review and meta-analysis. Gut 58:367, 2009[PMID: 19001059] [Full Text]

——— et al: Yield of diagnostic tests for celiac disease in individuals with symptoms suggestive of irritable bowel syndrome. Arch Intern Med 169:651, 2009

Gershon MD, Jack J: The serotonin signaling system: From basic understanding to drug development for functional GI disorders. Gastroenterology 132:397, 2007[PMID: 17241888] [Full Text]

Kassinen A et al: The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. Gastroenterology 133:24, 2007[PMID: 17631127] [Full Text]

Longstreth GF et al: Functional bowel disorders. Gastroenterology 130:1480, 2006[PMID: 16678561] [Full Text]

Mayer EA et al: Neuroimaging of the brain-gut axis: From basic understanding to treatment of functional GI disorders. Gastroenterology 131:1925, 2006[PMID: 17188960] [Full Text]

Owyang C: Irritable bowel syndrome, in *Textbook of Gastroenterology* , 5th ed, T Yamada (ed). Oxford, UK: Wiley-Blackwell, 2009, pp 1536–73

Preidis GA, Versalovic J: Targeting the human microbiome with antibiotics, probiotics, and prebiotics: Gastrocenterology enters the metagenomics era. Gastroenterology 136:2015, 2009[PMID: 19462507] [Full Text]

Spiller R, Garsed K: Postinfectious irritable bowel syndrome. Gastroenterology 136:1979, 2009[PMID: 19457422] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIVERTICULAR DISEASE

INCIDENCE AND EPIDEMIOLOGY

Among Western populations, diverticulosis of the colon affects nearly one-half of individuals older than age 60 years. Fortunately, only 20% of patients with diverticulosis develop symptomatic disease. However, in the United States, diverticular disease results in >200,000 hospitalizations annually, making it the fifth most costly gastrointestinal disorder. The incidence of the disease is on the rise, mainly among young patients. The mean age at presentation of the disease is 59 years. Although the prevalence among females and males is similar, males tend to present at a younger age. Diverticulosis is rare in underdeveloped countries, where diets include more fiber and roughage. However, shortly following migration to the United States, immigrants will develop diverticular disease at the same rate as U.S. natives.

ANATOMY AND PATHOPHYSIOLOGY

Two types of diverticula occur in the intestine: true and false (or pseudodiverticula). A true diverticulum is a saclike herniation of the entire bowel wall, whereas a pseudodiverticulum involves only a protrusion of the mucosa through the muscularis propria of the colon **(Fig. 297-1)**. The type of diverticulum affecting the colon is the pseudodiverticulum. The protrusion occurs at the point where the nutrient artery, or *vasa recti*, penetrates through the muscularis propria, resulting in a break in the integrity of the colonic wall. Diverticula commonly affect the sigmoid colon; only 5% of persons exhibit pancolonic diverticula. This anatomic restriction may be a result of the relative high-pressure zone within the muscular sigmoid colon. Thus, higher-amplitude contractions combined with constipated, high-fat-content stool within the sigmoid lumen results in the creation of these diverticula. *Diverticulitis* is inflammation of a diverticulum. The cause is not well understood and is probably multifactorial. The predominant theory is the retention of particulate material within the diverticular sac and the formation of a fecalith. Consequently, the vasa recti is either compressed or eroded, leading to either perforation or bleeding.

Figure 297-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gross and microscopic view of sigmoid diverticular disease. Arrows mark an inflamed diverticulum with the diverticular wall made up only of mucosa.

PRESENTATION, EVALUATION, AND MANAGEMENT OF DIVERTICULAR BLEEDING

Hemorrhage from a colonic diverticulum is the most common cause of hematochezia in patients >60 years, yet only 20% of patients with diverticulosis will have gastrointestinal bleeding. Patients at increased risk for bleeding tend to be hypertensive, have atherosclerosis, and regularly use nonsteroidal anti-inflammatory agents. Most bleeds are self-limited and stop spontaneously with bowel rest. The lifetime risk of rebleeding is 25%.

Localization of diverticular bleeding should include colonoscopy, which may be both diagnostic and therapeutic in the management of mild to moderate diverticular bleeding. If the patient is stable, massive bleeding is best managed by angiography. Mesenteric angiography can localize the bleeding site and occlude the bleeding vessel successfully with a coil in 80% of cases. The patient can then be followed closely with repetitive colonoscopy, if necessary, looking for evidence of colonic ischemia. Alternatively, a segmental resection of the colon can be undertaken to eliminate the risk of further bleeding. This may be advantageous in patients on chronic blood thinners. However, with newer techniques of highly selective coil embolization, the rate of colonic ischemia is <10% and the risk of acute rebleeding is <25%. Long-term results (40 months) indicate that more than 50% of patients with acute diverticular bleeds have had definitive treatment with highly selective angiography.

As another alternative, a selective infusion of vasopressin can be given to stop the hemorrhage, although this has been associated with significant complications, including myocardial infarction and intestinal ischemia. Furthermore, bleeding recurs in 50% of patients once the infusion is stopped. Localization studies indicate that bleeding as a result of colonic diverticulosis is more often seen from the right colon. For this reason, patients with presumed bleeding from diverticular disease requiring emergent surgery without localization should undergo a total abdominal colectomy. If the patient is unstable or has had a 6-unit bleed within 24 h, current recommendations are that surgery should be performed. In patients without severe comorbidities, surgical resection can be performed with a primary anastomosis. A higher anastomotic leak rate has been reported in patients who received >10 units of blood.

PRESENTATION, EVALUATION, AND STAGING OF DIVERTICULITIS

Acute uncomplicated diverticulitis characteristically presents with fever, anorexia, left lower quadrant abdominal pain, and obstipation **(Table 297-1)**. In <25% of cases, patients may present with generalized peritonitis indicating the presence of a diverticular perforation. If a pericolonic abscess has formed, the patient may have abdominal distention and signs of localized peritonitis. Laboratory investigations will demonstrate a leukocytosis. Rarely, a patient may present with an air-fluid level in the left lower quadrant on plain abdominal film. This is a giant diverticulum of the sigmoid colon and is managed with resection to avoid impending perforation.

Table 297-1 Presentation of Diverticular Disease

Uncomplicated Diverticular Disease—75%
Abdominal pain
Fever
Leukocytosis
Anorexia/obstipation
Complicated Diverticular Disease—25%
Abscess 16%
Perforation 10% Stricture 5%
Fistula 2%

The diagnosis of diverticulitis is best made on CT with the following findings: sigmoid diverticula, thickened colonic wall >4 mm, and inflammation within the pericolic fat ± the collection of contrast material or fluid. In 16% of patients, an abdominal abscess may be present. Symptoms of irritable bowel syndrome (IBS) may mimic those of diverticulitis. Therefore, suspected diverticulitis that does not meet CT criteria or is not associated with a leukocytosis or fever is not diverticular disease. Other conditions that can mimic diverticular disease include an ovarian cyst, endometriosis, acute appendicitis, and pelvic inflammatory disease.

Barium enema or colonoscopy should not be performed in the acute setting because of the higher risk of colonic perforation associated with insufflation or insertion of barium-based contrast material under pressure. A sigmoid malignancy can masquerade as diverticular disease. Therefore, a colonoscopy should be performed ~6 weeks after an attack of diverticular disease.

Complicated diverticular disease is defined as diverticular disease associated with an abscess or perforation and less commonly with a fistula (Table 297-1). Perforated diverticular disease is staged using the Hinchey classification system **(Fig. 297-2)**. This staging system was developed to predict outcomes following the surgical management of complicated diverticular disease. In complicated diverticular disease with fistula formation, common locations include cutaneous, vaginal, or vesicle fistulas. These conditions present with either passage of stool through the skin or vagina or the presence of air in the urinary stream (pneumaturia). Colovaginal fistulas are more common in women who have undergone a hysterectomy.

Figure 297-2

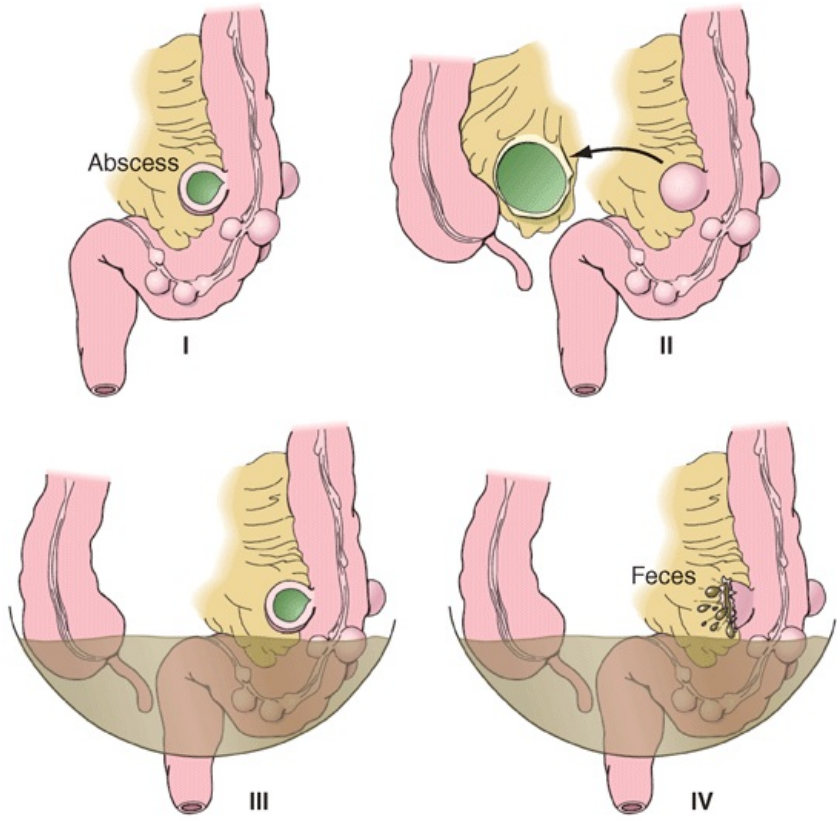

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hinchey classification of diverticulitis. Stage I: Perforated diverticulitis with a confined paracolic abscess. Stage II: Perforated diverticulitis that has closed spontaneously with distant abscess formation. Stage III: Noncommunicating perforated diverticulitis with fecal peritonitis (the diverticular neck is closed off and therefore contrast will not freely expel on radiographic images). Stage IV: Perforation and free communication with the peritoneum, resulting in fecal peritonitis.

Treatment: Diverticular Disease

MEDICAL MANAGEMENT

Asymptomatic diverticular disease discovered on imaging studies or at the time of colonoscopy is best managed by diet alterations. Patients should be instructed to eat a fiber-enriched diet that includes 30 g of fiber each day. Supplementary fiber products such as Metamucil, Fibercon, or Citrucel are useful. The incidence of complicated diverticular disease appears to be increased in patients who smoke. Therefore, patients should be encouraged to refrain from smoking. The historical recommendation to avoid eating nuts is not based on more than anecdotal data.

Symptomatic uncomplicated diverticular disease with confirmation of inflammation and infection within the colon should be treated initially with antibiotics and bowel rest. Nearly 75% of patients hospitalized for acute diverticulitis will respond to nonoperative treatment with a suitable antimicrobial regimen. The current recommended antimicrobial coverage is trimethoprim/sulfamethoxazole or ciprofloxacin and metronidazole targeting aerobic gram-negative rods and anaerobic bacteria. Unfortunately, these agents do not cover enterococci, and the addition of ampicillin to this regimen for nonresponders is recommended. Alternatively, single-agent therapy with a third-generation penicillin such as IV piperacillin or oral penicillin/clavulanic acid may be effective. The usual course of antibiotics is 7–10 days. Patients should remain on a limited diet until their pain resolves.

For long-term medical management of uncomplicated diverticular disease, rifaximin (a poorly absorbed broad-spectrum antibiotic), when compared to fiber alone, is associated with 30% less frequent recurrent symptoms from uncomplicated diverticular disease. Furthermore, the use of probiotics has been shown to decrease the incidence of recurrent attacks. Culture data from patients on probiotics noted a decrease in the presence of *Clostridium* species and an increase in *Lactobacillus* and *Bifidobacterium* strains.

SURGICAL MANAGEMENT

Preoperative risk factors influencing postoperative mortality rates include higher American Society of Anesthesiologists (ASA) physical status class **(Table 297-2)** and preexisting organ failure. In patients who are low risk (ASA P1 and P2), surgical therapy can be offered to those who do not rapidly improve on medical therapy. For uncomplicated diverticular disease, studies indicate that medical therapy can

be continued beyond two attacks without an increased risk of perforation requiring a colostomy. However, patients on immunosuppressive therapy, in chronic renal failure, or with a collagen-vascular disease have a fivefold greater risk of perforation during recurrent attacks. Surgical therapy is indicated in all low-surgical-risk patients with complicated diverticular disease.

Table 297-2 American Society of Anesthesiologists Physical Status Classification System

P1	A normal healthy patient
P2	A patient with mild systemic disease
P3	A patient with severe systemic disease
P4	A patient with severe systemic disease that is a constant threat to life
P5	A moribund patient who is not expected to survive without the operation
P6	A declared brain-dead patient whose organs are being removed for donor purposes

The goals of surgical management of diverticular disease include controlling sepsis, eliminating complications such as fistula or obstruction, removing the diseased colonic segment, and restoring intestinal continuity. These goals must be obtained while minimizing morbidity rate, length of hospitalization, and cost in addition to maximizing survival and quality of life. **Table 297-3** lists the operations most commonly indicated based upon Hinchey classification and the predicted morbidity and mortality rates. Surgical objectives include removal of the diseased sigmoid down to the rectosigmoid junction. Failure to do this may result in recurrent disease. The current options for uncomplicated diverticular disease include an open sigmoid resection or a laparoscopic sigmoid resection. The benefits of laparoscopic resection over open surgical techniques include early discharge (by at least 1 day), less narcotic use, less postoperative complications, and an earlier return to work.

Table 297-3 Outcome Following Surgical Therapy for Complicated Diverticular Disease

Hinchey Stage	Operative Procedure	Anastomotic Leak Rate, %	Overall Morbidity rate, %
I	Resection with primary anastomosis without diverting stoma	3.8	22
II	Resection with primary anastomosis +/– diversion	3.8	30
III	Hartmann's procedure vs. diverting colostomy and omental pedal graft	—	0 vs. 6 mortality
IV	Hartmann's procedure vs. diverting colostomy and omental pedicle graft	—	6 vs. 2 mortality

The options for the surgical management of complicated diverticular disease **(Fig. 297-3)** include the following: (1) proximal diversion of the fecal stream with an ileostomy or colostomy and sutured omental patch with drainage, (2) resection with colostomy and mucous fistula or closure of distal bowel with formation of a Hartmann's pouch, (3) resection with anastomosis (coloproctostomy), or (4) resection with anastomosis and diversion (coloproctostomy with loop ileostomy or colostomy). Laparoscopic techniques have been employed for complicated diverticular disease; however, higher conversion rates to open techniques have been reported.

Figure 297-3

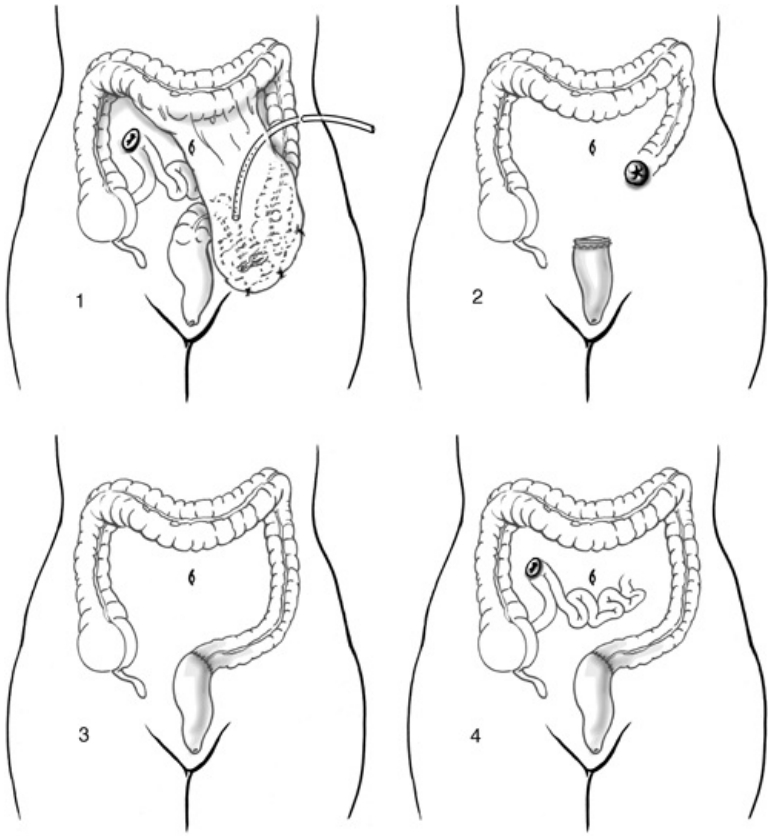

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Methods of surgical management of complicated diverticular disease. ***(1)*** Drainage, omental pedicle graft, and proximal diversion. ***(2)*** Hartmann's procedure. ***(3)*** Sigmoid resection with coloproctostomy. ***(4)*** Sigmoid resection with coloproctostomy and proximal diversion.

Patients with Hinchey stages I and II disease are managed with percutaneous drainage followed by resection with anastomosis about 6 weeks later. Percutaneous drainage is recommended for abscesses ≥5 cm with a well-defined wall that is accessible. Paracolic abscesses <5 cm in size may resolve with antibiotics alone. Contraindications to percutaneous drainage are no percutaneous access route, pneumoperitoneum, and fecal peritonitis. Urgent operative intervention is undertaken if patients develop generalized peritonitis, and most will need to be managed with a Hartmann's procedure. In selected cases, nonoperative therapy may be considered. In one nonrandomized study, nonoperative management of isolated paracolic abscesses (Hinchey stage I) was associated with only a 20% recurrence rate at 2 years. More than 80% of patients with distant abscesses (Hinchey stage II) required surgical resection for recurrent symptoms.

Hinchey stage III disease is managed with a Hartmann's procedure or with primary anastomosis and proximal diversion. If the patient has significant comorbidities, making operative intervention risky, a limited procedure including intraoperative peritoneal lavage (irrigation), omental patch to the oversewn perforation, and proximal diversion of the fecal stream with either an ileostomy or transverse colostomy can be performed. No anastomosis of any type should be attempted in Hinchey stage IV disease. A limited approach to these patients is associated with a decreased mortality rate.

RECURRENT SYMPTOMS

Recurrent abdominal symptoms following surgical resection for diverticular disease occurs in 10% of patients. Recurrent diverticular disease develops in patients following inadequate surgical resection. A retained segment of diseased rectosigmoid colon is associated with twice the incidence of recurrence. IBS may also cause recurrence of initial symptoms. Patients undergoing surgical resection for presumed diverticulitis and symptoms of abdominal cramping and irregular loose bowel movements consistent with IBS have functionally poorer outcomes.

COMMON DISEASES OF THE ANORECTUM

Rectal Prolapse (Procidentia)

INCIDENCE AND EPIDEMIOLOGY

Rectal prolapse is six times more common in women than in men. The incidence of rectal prolapse peaks in women >60 years. Women with rectal prolapse have a higher incidence of associated pelvic floor disorders including urinary incontinence, rectocele, cystocele, and enterocele. About 20% of

children with rectal prolapse will have cystic fibrosis. All children presenting with prolapse should undergo a sweat chloride test. Less common associations include Ehlers-Danlos syndrome, solitary rectal ulcer syndrome, congenital hypothyroidism, and Hirschsprung's disease.

ANATOMY AND PATHOPHYSIOLOGY

Rectal prolapse (procidentia) is a circumferential, full-thickness protrusion of the rectal wall through the anal orifice. It is often associated with a redundant sigmoid colon, pelvic laxity, and a deep rectovaginal septum (pouch of Douglas). Initially, rectal prolapse was felt to be the result of early internal rectal intussusception, which occurs in the upper to mid rectum. This was considered to be the first step in an inevitable progression to full-thickness external prolapse. However, only 1 of 38 patients with internal prolapse followed for >5 years developed full-thickness prolapse. Others have suggested that full-thickness prolapse is the result of damage to the nerve supply to the pelvic floor muscles or pudendal nerves from repeated stretching with straining to defecate. Damage to the pudendal nerves would weaken the pelvic floor muscles, including the external anal sphincter muscles. Bilateral pudendal nerve injury is more significantly associated with prolapse and incontinence than unilateral injury.

PRESENTATION AND EVALUATION

In external prolapse, the majority of patient complaints include anal mass, bleeding per rectum, and poor perianal hygiene. Prolapse of the rectum usually occurs following defecation and will spontaneously reduce or require the patient to manually reduce the prolapse. Constipation occurs in ~30–67% of patients with rectal prolapse. Differing degrees of fecal incontinence occur in 50–70% of patients. Patients with internal rectal prolapse will present with symptoms of both constipation and incontinence. Other associated findings include outlet obstruction (anismus) in 30%, colonic inertia in 10%, and solitary rectal ulcer syndrome in 12%.

Office evaluation is best performed after the patient has been given an enema, which enables the prolapse to protrude. An important distinction should be made between full-thickness rectal prolapse and isolated mucosal prolapse associated with hemorrhoidal disease **(Fig. 297-4)**. Mucosal prolapse is known for radial grooves rather than circumferential folds around the anus and is due to increased laxity of the connective tissue between the submucosa and underlying muscle of the anal canal. The evaluation of prolapse should also include cystoproctography and colonoscopy. These examinations evaluate for associated pelvic floor disorders and rule out a malignancy or a polyp as the lead point for prolapse. If rectal prolapse is associated with chronic constipation, the patient should undergo a defecating proctogram and a sitzmark study. This will evaluate for the presence of anismus or colonic inertia. Anismus is the result of attempting to defecate against a closed pelvic floor and is also known as *nonrelaxing puborectalis*. This can be seen when straightening of the rectum fails to occur on fluoroscopy while the patient is attempting to defecate. In colonic inertia, a sitzmark study will demonstrate retention of >20% of markers on abdominal x-ray 5 days after swallowing. For patients with fecal incontinence, endoanal ultrasound and manometric evaluation, including pudendal nerve testing of their anal sphincter muscles, may be performed before surgery for prolapse (see "Fecal Incontinence," below).

Figure 297-4

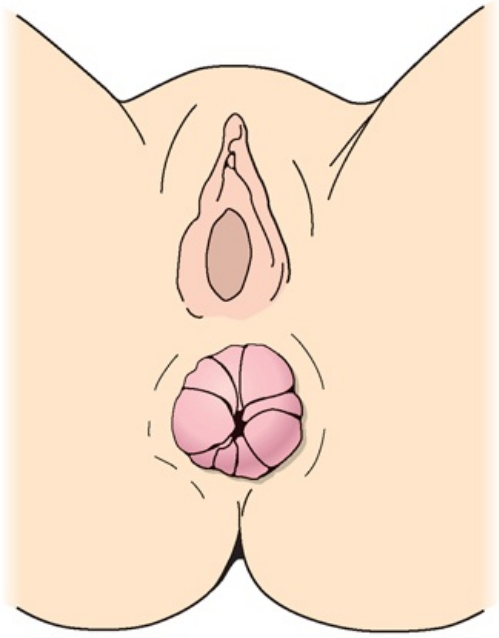

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

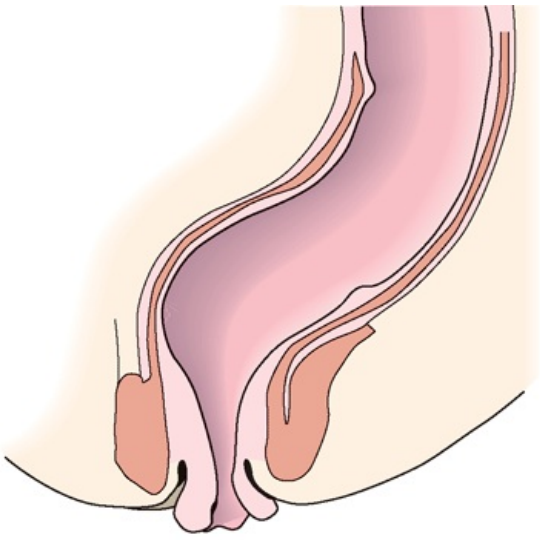

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

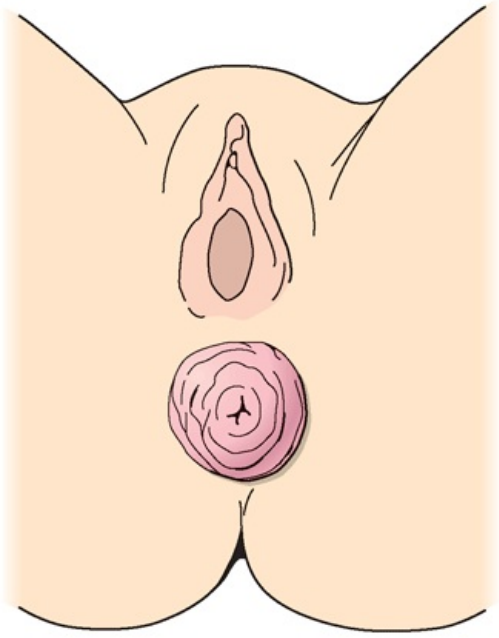

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

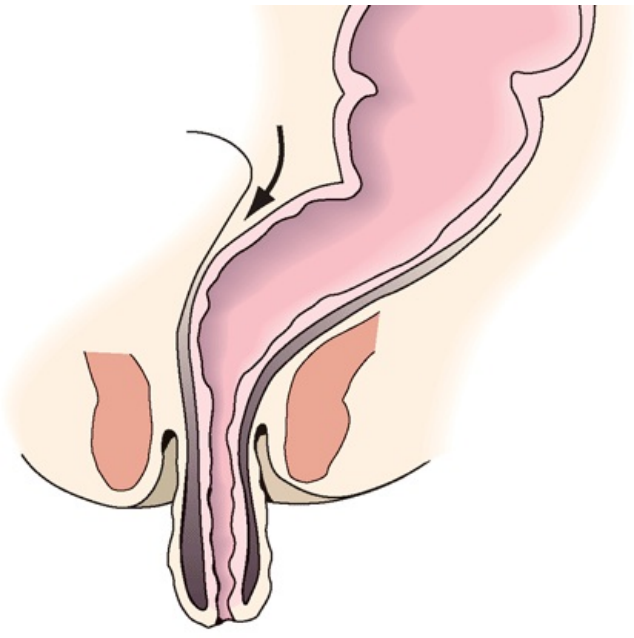

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Degree of rectal prolapse. Mucosal prolapse only (***A, B,*** sagittal view). Full-thickness prolapse associated with redundant rectosigmoid and deep pouch of Douglas (***C, D,*** sagittal view).

TREATMENT: RECTAL PROLAPSE

The medical approach to the management of rectal prolapse is limited and includes stool-bulking agents or fiber supplementation to ease the process of evacuation. Surgical correction of rectal prolapse is the mainstay of therapy. Two approaches are commonly considered, transabdominal and transperineal. Transabdominal approaches have been associated with lower recurrence rates, but some patients with significant comorbidities are better served by a transperineal approach.

Common transperineal approaches include a transanal proctectomy (Altmeier procedure), mucosal proctectomy (Delorme procedure), or placement of a Tirsch wire encircling the anus. The goal of the transperineal approach is to remove the redundant rectosigmoid colon. Common transabdominal approaches include presacral suture or mesh rectopexy (Ripstein) with (Frykman-Goldberg) or without

resection of the redundant sigmoid. Transabdominal procedures can be performed effectively with laparoscopic techniques without increased incidence of recurrence. The goal of the transabdominal approach is to restore normal anatomy by removing redundant bowel and reattaching the supportive tissue of the rectum to the presacral fascia. The final alternative is abdominal proctectomy with end-sigmoid colostomy. Colon resection, in general, is reserved for patients with constipation and outlet obstruction. If total colonic inertia is present, as defined by a history of constipation and a positive sitzmark study, a subtotal colectomy with an ileosigmoid or rectal anastomosis may be required at the time of rectopexy.

Previously, the presence of internal rectal prolapse identified on imaging studies has been considered a nonsurgical disorder and biofeedback was recommended. However, only one-third of patients will have successful resolution of symptoms from biofeedback. Two surgical procedures have been shown to be more effective than biofeedback. The STARR (stapled transanal rectal resection) procedure **(Fig. 297-5)** is performed through the anus in patients with internal prolapse. A circular stapling device is inserted through the anus; the internal prolapse is identified and ligated with the stapling device. The Laparoscopic Ventral Rectopexy (LVR) **(Fig. 297-6)** is performed by creating an opening in the peritoneum on the left side of the rectosigmoid and carrying this opening down anterior on the rectum into the pouch of Douglas. No rectal mobilization is performed, thus avoiding any autonomic nerve injury. Mesh is secured to the anterior and lateral portion of the rectum, the vaginal fornix, and the sacral promontory, allowing for closure of the rectovaginal septum and correction of the internal prolapse. In both procedures, recurrence at 1 year was low (<10%) and symptoms improved in more than three-fourths of patients.

Figure 297-5

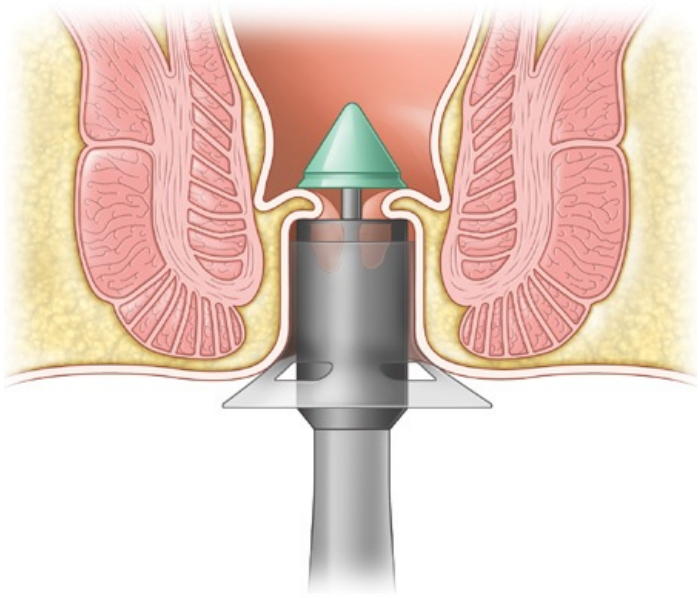

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stapled transanal rectal resection. Schematic of placement of the circular stapling device.

Figure 297-6

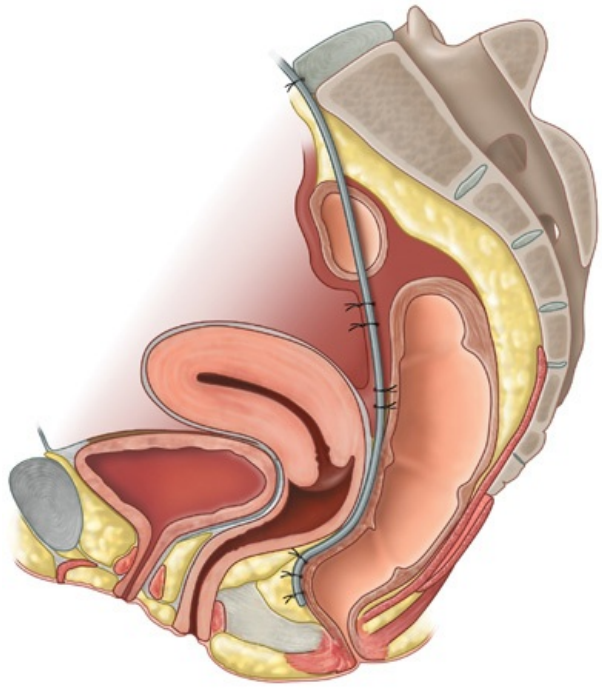

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Laparoscopic Ventral Rectopexy (LVR). To reduce the internal prolapse and close any rectovaginal septal defect, the pouch of Douglas is opened and mesh is secured to the anterolateral rectum, vaginal fornix, and sacrum. (*From D'Hoore et al: Br J Surg 91:1500, 2004.*)

Fecal Incontinence

INCIDENCE AND EPIDEMIOLOGY

Fecal incontinence is the involuntary passage of fecal material >10 mL for at least 1 month. The prevalence of fecal incontinence in the United States is 0.5–11%. The majority of patients are women. A higher incidence of incontinence is seen among parous women. One-half of patients with fecal incontinence also suffer from urinary incontinence. The majority of incontinence is a result of obstetric injury to the pelvic floor, either while carrying a fetus or during the delivery. An anatomic sphincter defect may occur in up to 32% of women following childbirth regardless of visible damage to the perineum. Risk factors at the time of delivery include prolonged labor, the use of forceps, and the need for an episiotomy. Medical conditions known to contribute to the development of fecal incontinence are listed in **Table 297-4**.

Table 297-4 Medical Conditions that Contribute to Symptoms of Fecal Incontinence

Neurologic Disorders
▪ Dementia ▪ Brain tumor ▪ Stroke ▪ Multiple sclerosis ▪ Tabes dorsalis ▪ Cauda equina lesions
Skeletal Muscle Disorders
▪ Myasthenia gravis ▪ Myopathies, muscular dystrophy
Miscellaneous
▪ Hypothyroidism ▪ Irritable bowel syndrome ▪ Sedation ▪ Severe diarrhea

ANATOMY AND PATHOPHYSIOLOGY

The anal sphincter complex is made up of the internal and external anal sphincter. The internal sphincter is smooth muscle and a continuation of the circular fibers of the rectal wall. It is innervated by the intestinal myenteric plexus and is therefore not under voluntary control. The external anal sphincter is formed in continuation with the levator ani muscles and is under voluntary control. The pudendal nerve supplies motor innervation to the external anal sphincter. Obstetric injury may result in tearing of the muscle fibers anteriorly at the time of the delivery. This results in an obvious anterior defect on endoanal ultrasound. Injury may also be the result of stretching of the pudendal nerves. The majority of patients who suffer from fecal incontinence following obstetric injury do so several years following the birth of their last child.

PRESENTATION AND EVALUATION

Patients may suffer with varying degrees of fecal incontinence. Minor incontinence includes incontinence to flatus and occasional seepage of liquid stool. Major incontinence is frequent inability to control solid waste. As a result of fecal incontinence, patients suffer from poor perianal hygiene. Beyond the immediate problems associated with fecal incontinence, these patients are often withdrawn and suffer from depression. For this reason, quality-of-life measures have become an important component in the evaluation of patients with fecal incontinence.

The evaluation of fecal incontinence should include a thorough history and physical examination, anal manometry, pudendal nerve terminal motor latency (PNTML), and endoanal ultrasound. Unfortunately, all of these investigations are user-dependent. Centers that care for patients with fecal incontinence will have an anorectal physiology laboratory that uses standardized methods of evaluating anorectal physiology. Anal manometry measures resting and squeeze pressures within the anal canal using an intraluminal water-perfused catheter. Pudendal nerve studies evaluate the function of the nerves innervating the anal canal using a finger electrode placed in the anal canal. Stretch injuries to these nerves will result in a delayed response of the sphincter muscle to a stimulus, indicating a prolonged latency. Finally, ultrasound will evaluate the extent of the injury to the sphincter muscles before surgical repair. Only PNTML has been shown to consistently predict outcome following surgical intervention.

Rarely does a pelvic floor disorder exist alone. The majority of patients with fecal incontinence will have a degree of urinary incontinence. Similarly, fecal incontinence is a part of the spectrum of pelvic organ prolapse. For this reason, patients may present with symptoms of obstructed defecation as well as fecal incontinence. Careful evaluation including cinedefecography should be performed to search for other associated defects. Surgical repair of incontinence without attention to other associated defects may decrease the success of the repair.

TREATMENT: FECAL INCONTINENCE

The "gold standard" for the treatment of fecal incontinence with an isolated sphincter defect is overlapping sphincteroplasty. The external anal sphincter muscle and scar tissue as well as any identifiable internal sphincter muscle are dissected free from the surrounding adipose and connective tissue and then an overlapping repair is performed in an attempt to rebuild the muscular ring and restore its function. Other newer approaches include radio frequency therapy to the anal canal to aid in the development of collagen fibers and provide tensile strength to the sphincter muscles. Sacral nerve stimulation and the artificial bowel sphincter are both adaptations of procedures developed for the management of urinary incontinence. Sacral nerve stimulation is ideally suited for patients with intact but weak anal sphincters. A temporary nerve stimulator is placed on the third sacral nerve. If there is at least a 50% improvement in symptoms, a permanent nerve stimulator is placed under the skin. The artificial bowel sphincter is a cuff and reservoir apparatus that allows for manual inflation of a cuff placed around the anus, increasing anal tone. This allows the patient to manually close off the anal canal until defecation is necessary.

Long-term results following overlapping sphincteroplasty show about a 50% failure rate over 5 years. Poorer outcome has been seen in patients with prolonged PNTML. Long-term results for sacral stimulation have been promising; however, the indications for this procedure are presently limited in the United States. Unfortunately, the artificial bowel sphincter has been associated with a 30% infection rate.

Hemorrhoidal Disease

INCIDENCE AND EPIDEMIOLOGY

Symptomatic hemorrhoids affect >1 million individuals in the Western world per year. The prevalence of hemorrhoidal disease is not selective for age or sex. However, age is known to have a deleterious effect on the anal canal. The prevalence of hemorrhoidal disease is less in underdeveloped countries. The typical low-fiber, high-fat Western diet is associated with constipation and straining and the development of symptomatic hemorrhoids.

ANATOMY AND PATHOPHYSIOLOGY

Hemorrhoidal cushions are a normal part of the anal canal. The vascular structures contained within this tissue aid in continence by preventing damage to the sphincter muscle. Three main hemorrhoidal complexes traverse the anal canal—the left lateral, the right anterior, and the right posterior. Engorgement and straining leads to prolapse of this tissue into the anal canal. Over time, the anatomic support system of the hemorrhoidal complex weakens, exposing this tissue to the outside of the anal canal where it is susceptible to injury. Hemorrhoids are commonly classified as internal or external. Although small external cushions do exist, the standard classification of hemorrhoidal disease is based on the progression of the disease from their normal internal location to the prolapsing external position **(Table 297-5)**.

Table 297-5 The Staging and Treatment of Hemorrhoids

Stage	Description of Classification	Treatment
I	Enlargement with bleeding	Fiber supplementation Cortisone suppository Sclerotherapy
II	Protrusion with spontaneous reduction	Fiber supplementation Cortisone suppository
III	Protrusion requiring manual reduction	Fiber supplementation Cortisone suppository Banding Operative hemorrhoidectomy (stapled or traditional)
IV	Irreducible protrusion	Fiber supplementation Cortisone suppository Operative hemorrhoidectomy

PRESENTATION AND EVALUATION

Patients commonly present to a physician for two reasons: bleeding and protrusion. Pain is less common than with fissures and, if present, is described as a dull ache from engorgement of the hemorrhoidal tissue. Severe pain may indicate a thrombosed hemorrhoid. Hemorrhoidal bleeding is described as bright red blood seen either in the toilet or upon wiping. Occasional patients can present with significant bleeding, which may be a cause of anemia; however, the presence of a colonic neoplasm must be ruled out. Patients who present with a protruding mass complain about inability to maintain perianal hygiene and are often concerned about the presence of a malignancy.

The diagnosis of hemorrhoidal disease is made on physical examination. Inspection of the perianal region for evidence of thrombosis or excoriation is performed, followed by a careful digital examination. Anoscopy is performed paying particular attention to the known position of hemorrhoidal disease. The patient is asked to strain. If this is difficult for the patient, the maneuver can be performed while sitting on a toilet. The physician is notified when the tissue prolapses. It is important to differentiate the circumferential appearance of a full-thickness rectal prolapse from the radial nature of prolapsing hemorrhoids (see "Rectal Prolapse," above). The stage and location of the hemorrhoidal complexes are defined.

TREATMENT: HEMORRHOIDAL DISEASE

The treatment for bleeding hemorrhoids is based upon the stage of the disease (Table 297-5). In all patients with bleeding, the possibility of other causes must be considered. In young patients without a family history of colorectal cancer, the hemorrhoidal disease may be treated first and a colonoscopic examination performed if the bleeding continues. Older patients who have not had colorectal cancer screening should undergo colonoscopy or flexible sigmoidoscopy.

With rare exceptions, the acutely thrombosed hemorrhoid can be excised within the first 72 h by performing an elliptical excision. Sitz baths, fiber, and stool softeners are prescribed. Additional therapy for bleeding hemorrhoids includes banding, sclerotherapy, excisional hemorrhoidectomy, and stapled hemorrhoidectomy. Sensation begins at the dentate line; therefore, banding or sclerotherapy can be performed without discomfort in the office. Bands are placed around the engorged tissue, causing ischemia and fibrosis. This aids in fixing the tissue proximally in the anal canal. Patients may complain of a dull ache for 24 h following band application. During sclerotherapy, 1–2 mL of a sclerosant (usually sodium tetradecyl sulfate) is injected using a 25-gauge needle into the submucosa of the hemorrhoidal complex. Care must be taken not to inject the anal canal circumferentially, or stenosis may occur. The sutured and stapled hemorrhoidectomies are equally effective in the treatment of symptomatic third- and fourth-degree hemorrhoids. However, because the sutured hemorrhoidectomy involves the removal of redundant tissue down to the anal verge, unpleasant anal skin tags are removed as well. The stapled hemorrhoidectomy is associated with less discomfort; however, this procedure does not remove anal skin tags. No procedures on hemorrhoids should be done in patients who are immunocompromised or who have active proctitis. Furthermore, emergent hemorrhoidectomy for bleeding hemorrhoids is associated with a higher complication rate.

Acute complications associated with the treatment of hemorrhoids include pain, infection, recurrent bleeding, and urinary retention. Care should be taken to place bands properly and to avoid overhydration in patients undergoing operative hemorrhoidectomy. Late complications include fecal

incontinence as a result of injury to the sphincter during the dissection. Anal stenosis may develop from overzealous excision, with loss of mucosal skin bridges for reepithelialization. Finally, an *ectropion* (prolapse of rectal mucosa from the anal canal) may develop. Patients with an ectropion complain of a "wet" anus as a result of inability to prevent soiling once the rectal mucosa is exposed below the dentate line.

Anorectal Abscess

INCIDENCE AND EPIDEMIOLOGY

The development of a perianal abscess is more common in men than women by a ratio of 3:1. The peak incidence is in the third to fifth decade of life. Perianal pain associated with the presence of an abscess accounts for 15% of office visits to a colorectal surgeon. The disease is more prevalent in immunocompromised patients such as those with diabetes, hematologic disorders, or inflammatory bowel disease (IBD) and persons who are HIV. positive. These disorders should be considered in patients with recurrent perianal infections.

ANATOMY AND PATHOPHYSIOLOGY

An anorectal abscess is an abnormal fluid-containing cavity in the anorectal region. Anorectal abscess results from an infection involving the glands surrounding the anal canal. Normally, these glands release mucus into the anal canal, which aids in defecation. When stool accidentally enters the anal glands, the glands become infected and an abscess develops. Anorectal abscesses are perianal in 40 –50% of patients, ischiorectal in 20–25%, intersphincteric in 2–5%, and supralevator in 2.5% **(Fig. 297-7)**.

Figure 297-7

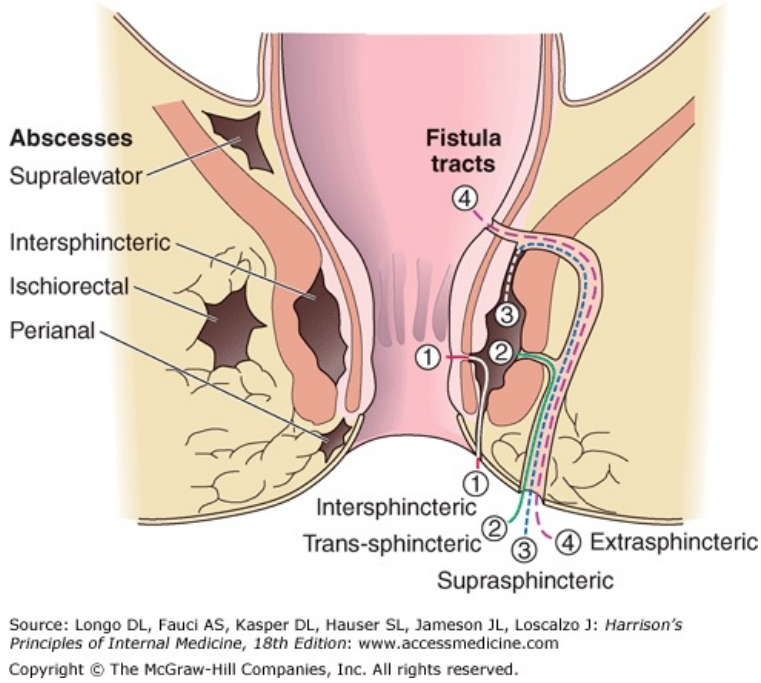

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Common locations of anorectal abscess *(left)* and fistula in ano *(right).*

PRESENTATION AND EVALUATION

Perianal pain and fever are the hallmarks of an abscess. Patients may have difficulty voiding and have blood in the stool. A prostatic abscess may present with similar complaints, including dysuria. Patients with a prostatic abscess will often have a history of recurrent sexually transmitted diseases. On physical examination, a large fluctuant area is usually readily visible. Routine laboratory evaluation shows an elevated white blood cell count. Diagnostic procedures are rarely necessary unless evaluating a recurrent abscess. A CT scan or MRI has an accuracy of 80% in determining incomplete drainage. If there is a concern about the presence of IBD, a rigid or flexible sigmoidoscopic examination may be done at the time of drainage to evaluate for inflammation within the rectosigmoid region. A more complete evaluation for Crohn's disease would include a full colonoscopy and small-bowel series.

TREATMENT: ANORECTAL ABSCESS

Office drainage of an uncomplicated anorectal abscess may suffice. A small incision close to the anal verge is made and a Mallenkot drain is advanced into the abscess cavity. For patients who have a complicated abscess or who are diabetic or immunocompromised, drainage should be performed in an

operating room under anesthesia. These patients are at greater risk for developing necrotizing fasciitis. The course of antibiotics is controversial but should be at least 2 weeks in patients who are immunocompromised or have prosthetic heart valves, artificial joints, diabetes, or IBD.

Fistula in Ano

INCIDENCE AND EPIDEMIOLOGY

The incidence and prevalence of fistulating perianal disease parallels the incidence of anorectal abscess. Some 30–40% of abscesses will give rise to fistula in ano. While the majority of the fistulas are cryptoglandular in origin, 10% are associated with IBD, tuberculosis, malignancy, and radiation.

ANATOMY AND PATHOPHYSIOLOGY

A fistula in ano is defined as a communication of an abscess cavity with an identifiable internal opening within the anal canal. This identifiable opening is most commonly located at the dentate line where the anal glands enter the anal canal. Patients experiencing continuous drainage following the treatment of a perianal abscess likely have a fistula in ano. These fistulas are classified by their relationship to the anal sphincter muscles, with 70% being intersphincteric, 23% transsphincteric, 5% suprasphincteric, and 2% extrasphincteric (Fig. 297-7).

PRESENTATION AND EVALUATION

A patient with a fistula in ano will complain of constant drainage from the perianal region. The drainage may increase with defecation. Perianal hygiene is difficult to maintain. Examination under anesthesia is the best way to evaluate a fistula. At the time of the examination, anoscopy is performed to look for an internal opening. Diluted hydrogen peroxide will aid in identifying such an opening. In lieu of anesthesia, MRI with an endoanal coil will also identify tracts in 80% of the cases. After drainage of an abscess with insertion of a Mallenkot catheter, a fistulagram through the catheter can be obtained in search of an occult fistula tract. Goodsall's rule states that a posterior external fistula will enter the anal canal in the posterior midline, whereas an anterior fistula will enter at the nearest crypt. A fistula exiting >3 cm from the anal verge may have a complicated upward extension and may not obey Goodsall's rule.

TREATMENT: FISTULA IN ANO

A newly diagnosed draining fistula is best managed with placement of a seton, a vessel loop or silk tie placed through the fistula tract, which maintains the tract open and quiets down the surrounding inflammation that occurs from repeated blockage of the tract. Once the inflammation is less, the exact relationship of the fistula tract to the anal sphincters can be ascertained. A simple fistulotomy can be performed for intersphincteric and low (less than one-third of the muscle) transsphincteric fistulas without compromising continence. For a higher transsphincteric fistula, an anorectal advancement flap in combination with a drainage catheter or fibrin glue may be used. Very long (>2 cm) and narrow tracts respond better to fibrin glue than shorter tracts. Simple ligation of the internal fistula tract (LIFT procedure) has also been used in the management of simple fistula with good success.

Patients should be maintained on stool-bulking agents, nonnarcotic pain medication, and sitz baths following surgery for a fistula. Early complications from these procedures include urinary retention and bleeding. Later complications are rare (<10%) and include temporary and permanent incontinence. Recurrence following fistulotomy is 0–18% and following anorectal advancement flap and the LIFT procedure is 20–30%.

Anal Fissure

INCIDENCE AND EPIDEMIOLOGY

Anal fissures occur at all ages but are more common in the third through the fifth decades. A fissure is the most common cause of rectal bleeding in infancy. The prevalence is equal in males and females. It is associated with constipation, diarrhea, infectious etiologies, perianal trauma, and Crohn's disease.

ANATOMY AND PATHOPHYSIOLOGY

Trauma to the anal canal occurs following defecation. This injury occurs in the anterior or, more commonly, the posterior anal canal. Irritation caused by the trauma to the anal canal results in an increased resting pressure of the internal sphincter. The blood supply to the sphincter and anal mucosa enters laterally. Therefore, increased anal sphincter tone results in a relative ischemia in the region of the fissure and leads to poor healing of the anal injury. A fissure that is not in the posterior or anterior position should raise suspicion for other causes, including tuberculosis, syphilis, Crohn's disease, and malignancy.

PRESENTATION AND EVALUATION

A fissure can be easily diagnosed on history alone. The classic complaint is pain, which is strongly associated with defecation and is relentless. The bright red bleeding that can be associated with a fissure is less extensive than that associated with hemorrhoids. On examination, most fissures are located in either the posterior or anterior position. A lateral fissure is worrisome as it may have a less benign nature, and systemic disorders should be ruled out. A chronic fissure is indicated by the

presence of a hypertrophied anal papilla at the proximal end of the fissure and a sentinel pile or skin tag at the distal end. Often the circular fibers of the hypertrophied internal sphincter are visible within the base of the fissure. If anal manometry is performed, elevation in anal resting pressure and a sawtooth deformity with paradoxical contractions of the sphincter muscles are pathognomonic.

TREATMENT: ANAL FISSURE

The management of the acute fissure is conservative. Stool softeners for those with constipation, increased dietary fiber, topical anesthetics, glucocorticoids, and sitz baths are prescribed and will heal 60–90% of fissures. Chronic fissures are those present for >6 weeks. These can be treated with modalities aimed at decreasing the anal canal resting pressure including nifedipine or nitroglycerin ointment applied three times a day, and botulinum toxin type A, up to 20 units, injected into the internal sphincter on each side of the fissure. Surgical management includes anal dilatation and lateral internal sphincterotomy. Usually, one-third of the internal sphincter muscle is divided; it is easily identified because it is hypertrophied. Recurrence rates from medical therapy are higher, but this is offset by a risk of incontinence following sphincterotomy. Lateral internal sphincterotomy may lead to incontinence more commonly in women.

ACKNOWLEDGMENT

We would like to thank Cory Sandore for providing some illustrations for this chapter. Gregory Bulkley, MD, contributed to this chapter in the 16th edition.

FURTHER READINGS

Collinson R et al: Laparoscopic ventral rectopexy for internal prolapse: Short-term functional results. Colorectal Dis 12:97, 2010[PMID: 19788493] [Full Text]

Etzioni DA et al: Diverticulitis in California from 1995 to 2006: Increased rates of treatment for younger patients. Am Surg 75:981, 2009[PMID: 19886149] [Full Text]

Klarenbeek BR et al: Indications for elective sigmoid resection in diverticular disease. Ann Surg 251:670, 2010[PMID: 20224374] [Full Text]

Latella G, Scarpignato C: Rifaximin in the management of colonic diverticular disease. Expert Rev Gastroenterol Hepatol 3:585, 2009[PMID: 19929580] [Full Text]

Pinto R, Sands D: Surgery and sacral nerve stimulation for constipation and fecal incontinence. Gastrointest Endosc Clin North Am 19:83, 2009[PMID: 19232283] [Full Text]

Russ AJ et al: Laparoscopy improves short-term outcomes after surgery for diverticular disease. Gastroenterology 138:2213, 2010

Sajid MS et al: Open vs laparoscopic repair of full-thickness rectal prolapse: A re-meta-analysis. Colorectal Dis 12:515, 2010[PMID: 20557324] [Full Text]

Schwander O et al: Assessing the safety, effectiveness, and quality of life after the STARR procedure for obstructed defecation: Results of the German STARR registry. Langenbecks Arch Surg 395:505, 2010

Shanwani A et al: Ligation of the intersphincteric fistula tract (LIFT): A sphincter-saving technique for fistula-in-ano. Dis Colon Rectum 53:39, 2010[PMID: 20010348] [Full Text]

Touzios J, Dozois E: Diverticulosis and acute diverticulitis. Gastroenterol Clin North Am 38:513, 2009[PMID: 19699411] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTESTINAL ISCHEMIA

Incidence and Epidemiology

Intestinal ischemia is an uncommon vascular disease associated with a high mortality. It is categorized according to etiology: (1) arterioocclusive mesenteric ischemia (AOMI), (2) nonocclusive mesenteric ischemia (NOMI), and (3) mesenteric venous thrombosis (MVT). Acute intestinal ischemia is more common than its counterpart, chronic arterial ischemia. Risk factors for acute arterial ischemia include atrial fibrillation, recent myocardial infarction, valvular heart disease, and recent cardiac or vascular catheterization. The increased incidence of intestinal ischemia seen in Western countries parallels the incidence of atherosclerosis and the aging population. With the exception of strangulated small-bowel obstruction, ischemic colitis is the most common form of acute ischemia and the most prevalent gastrointestinal disease complicating cardiovascular surgery. The incidence of ischemic colitis following elective aortic repair is 5–9%, and the incidence triples in patients following emergent repair. Other less common forms of intestinal ischemia include chronic mesenteric angina associated with atherosclerotic disease and MVT. The latter is associated with the presence of a hypercoagulable state including protein C or S deficiency, antithrombin III deficiency, polycythemia vera, and carcinoma.

Anatomy and Pathophysiology

Intestinal ischemia occurs when insufficient perfusion to intestinal tissue produces ischemic tissue injury. The blood supply to the intestines is depicted in **Fig. 298-1**. To prevent ischemic injury, extensive collateralization occurs between major mesenteric trunks and branches of the mesenteric arcades **(Table 298-1)**. Collateral vessels within the small bowel are numerous and meet within the duodenum and the bed of the pancreas. Collateral vessels within the colon meet at the splenic flexure and descending/sigmoid colon. These areas, which are inherently at risk for decreased blood flow, are known as *Griffiths' point* and *Sudeck's point*, respectively, and are the most common locations for colonic ischemia (Fig. 298-1, shaded areas). The splanchnic circulation can receive up to 30% of the cardiac output. Protective responses to prevent intestinal ischemia include abundant collateralization, autoregulation of blood flow, and the ability to increase oxygen extraction from the blood.

Figure 298-1

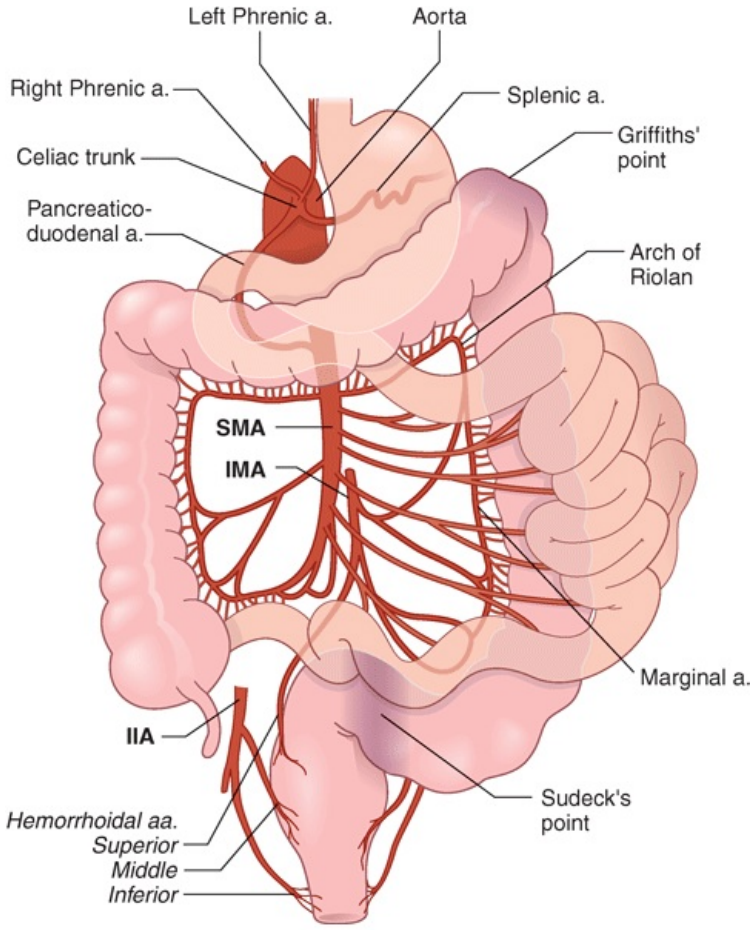

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Blood supply to the intestines includes the celiac artery, superior mesenteric artery (SMA), inferior mesenteric artery (IMA), and branches of the internal iliac artery (IIA). Griffiths' and Sudeck's points, indicated by shaded areas, are watershed areas within the colonic blood supply and common locations for ischemia.

Table 298-1 Collateral Arterial Intestinal Blood Flow

Involved Circulation	Mesenteric Artery	Adjoining Artery	Collateral Artery
Systemic	Celiac	Descending aorta	Phrenic
Systemic	IMA	Hypogastric	Middle hemorrhoidal
Mesenteric	Celiac	SMA	Superior/inferior pancreaticoduodenal
Mesenteric	SMA	IMA	Arch of Riolan
Mesenteric	SMA	Celiac/IMA	Intramesenteric
Mesenteric	SMA	IMA	Marginal

Abbreviations: IMA, inferior mesenteric artery; SMA, superior mesenteric artery.

Occlusive ischemia is a result of disruption of blood flow by an embolus or progressive thrombosis in a major artery supplying the intestine. Emboli originate from the heart in >75% of cases and lodge preferentially just distal to the origin of the middle colic artery from the superior mesenteric artery. Progressive thrombosis of at least two of the major vessels supplying the intestine is required for the development of chronic intestinal angina. Nonocclusive ischemia is disproportionate mesenteric vasoconstriction (arteriolar vasospasm) in response to a severe physiologic stress such as dehydration or shock. If left untreated, early mucosal stress ulceration will progress to full-thickness injury.

Presentation, Evaluation, and Management

Intestinal ischemia remains one of the most challenging diagnoses. The mortality rate is >50%. The most significant indicator of survival is the timeliness of diagnosis and treatment. An overview of diagnosis and management of each form of intestinal ischemia is given in **Table 298-2**.

Table 298-2 Overview of the Management of Acute Intestinal Ischemia

Condition	Key to Early Diagnosis	Treatment of Underlying Cause	Treatment of Specific Lesion	Treatment of Systemic Consequences
Arterial embolus	Early laparotomy	Anticoagulation Cardioversion Proximal thrombectomy Aneurysmectomy	Laparotomy Embolectomy Vascular bypass Assess viability and resect dead bowel	Ensure hydration Give antibiotics Reverse acidosis Optimize oxygen delivery Support cardiac output Treat other embolic sites Avoid vasoconstrictors
Arterial thrombosis	Duplex ultrasound Angiography	Anticoagulation Hydration	Endovascular stent Endarterectomy/thrombectomy or vascular bypass Assess viability and resect dead bowel	Give antibiotics Reverse acidosis Optimize oxygen delivery Support cardiac output Avoid vasoconstrictors
Venous thrombosis	Spiral CT	Anticoagulation Massive hydration	Anticoagulation ± laparotomy/ thrombectomy/portasystemic shunt Assess viability and resect dead bowel	Give antibiotics Reverse acidosis Optimize oxygen delivery Support cardiac output Avoid vasoconstrictors
Nonocclusive mesenteric ischemia	Vasospasm: Angiography Hypoperfusion: Spiral CT or colonoscopy	Ensure hydration Support cardiac output Avoid vasoconstrictors Ablate renin-angiotensin axis	Vasospasm Intraarterial vasodilators Hypoperfusion Delayed laparotomy Assess viability and resect dead bowel	Ensure hydration Give antibiotics Reverse acidosis Optimize oxygen delivery Support cardiac output Avoid vasoconstrictors

Source: Modified from GB Bulkley, in JL Cameron (ed): *Current Surgical Therapy*, 2nd ed. Toronto, BC Decker, 1986.

Acute mesenteric ischemia resulting from arterial embolus or thrombosis presents with severe acute, nonremitting abdominal pain strikingly out of proportion to the physical findings. Associated symptoms may include nausea and vomiting, transient diarrhea, and bloody stools. With the exception of minimal abdominal distention and hypoactive bowel sounds, early abdominal examination is unimpressive. Later findings will demonstrate peritonitis and cardiovascular collapse. In the evaluation of acute intestinal ischemia, routine laboratory tests should be obtained, including complete blood count, serum chemistry, coagulation profile, arterial blood gas, amylase, lipase, lactic acid, blood type and cross match, and cardiac enzymes. Regardless of the need for urgent surgery, emergent admission to a monitored bed or intensive care unit is recommended for resuscitation and further evaluation. If the diagnosis of intestinal ischemia is being considered, consultation with a surgical service is necessary.

Other diagnostic modalities that may be useful in diagnosis but should not delay surgical therapy include electrocardiogram (ECG), abdominal radiographs, CT, and mesenteric angiography. More recently, mesentery duplex scanning and visible light spectroscopy during colonoscopy have been demonstrated to be beneficial. The ECG may demonstrate an arrhythmia, indicating the possible

source of the emboli. A plain abdominal film may show evidence of free intraperitoneal air, indicating a perforated viscus and the need for emergent exploration. Earlier features of intestinal ischemia seen on abdominal radiographs include bowel-wall edema, known as "thumbprinting." If the ischemia progresses, air can be seen within the bowel wall (*pneumatosis intestinalis*) and within the portal venous system. Other features include calcifications of the aorta and its tributaries, indicating atherosclerotic disease. With the administration of oral and IV contrast, dynamic CT with three-dimensional reconstruction is a highly sensitive test for intestinal ischemia. In acute embolic disease, mesenteric angiography is best performed intraoperatively. A mesenteric duplex scan demonstrating a high peak velocity of flow in the superior mesenteric artery (SMA) is associated with an ~80% positive predictive value of mesenteric ischemia. More significantly, a negative duplex scan virtually precludes the diagnosis of mesenteric ischemia. Duplex imaging serves as a screening test; further investigations with angiography are needed. Endoscopic techniques using visible light spectroscopy can be used in the diagnosis of chronic ischemia.

The "gold standard" for the diagnosis and management of acute arterial occlusive disease is laparotomy. Surgical exploration should not be delayed if suspicion of acute occlusive mesenteric ischemia is high or evidence of clinical deterioration or frank peritonitis is present. The goal of operative exploration is to resect compromised bowel and restore blood supply. Intraoperative or preoperative arteriography and systemic heparinization may assist the vascular surgeon in restoring blood supply to the compromised bowel. The entire length of the small and large bowel beginning at the ligament of Treitz should be evaluated. The pattern of intestinal ischemia may indicate the level of arterial occlusion. In the case of SMA occlusion where the embolus usually lies just proximal to the origin of the middle colic artery, the proximal jejunum is often spared while the remainder of the small bowel to the transverse colon will be ischemic. The surgical management of acute mesenteric ischemia of the small bowel is attempted embolectomy via intraoperative angiography or arteriotomy. Although more commonly applied to chronic disease, acute thrombosis may be managed with angioplasty, with or without endovascular stent placement. If this is unsuccessful, a bypass from the aorta to the superior mesenteric artery is performed.

Nonocclusive or vasospastic mesenteric ischemia presents with generalized abdominal pain, anorexia, bloody stools, and abdominal distention. Often these patients are obtunded, and physical findings may not assist in the diagnosis. The presence of a leukocytosis, metabolic acidosis, elevated amylase or creatinine phosphokinase levels, and/or lactic acidosis are useful in support of the diagnosis of advanced intestinal ischemia; however, these markers may not be indicative of either reversible ischemia or frank necrosis. Investigational markers for intestinal ischemia include D-dimer, glutathione S-transferase, platelet-activating factor (PAF), and mucosal pH monitoring. Regardless of the need for urgent surgery, emergent admission to a monitored bed or intensive care unit is recommended for resuscitation and further evaluation. Early manifestations of intestinal ischemia include fluid sequestration within the bowel wall leading to a loss of interstitial volume. Aggressive fluid resuscitation may be necessary. To optimize oxygen delivery, nasal O_2 and blood transfusions may be given. Broad-spectrum antibiotics should be given to provide sufficient coverage for enteric pathogens, including gram-negative and anaerobic organisms. Frequent monitoring of the patient's vital signs, urine output, blood gases, and lactate levels is paramount, as is frequent abdominal examination. All vasoconstricting agents should be avoided; fluid resuscitation is the intervention of choice to maintain hemodynamics.

If ischemic colitis is a concern, colonoscopy should be performed to assess the integrity of the colon mucosa. Visualization of the rectosigmoid region may demonstrate decreased mucosal integrity, associated more commonly with nonocclusive mesenteric ischemia, or, on occasion, occlusive disease as a result of acute loss of inferior mesenteric arterial flow following aortic surgery. Ischemia of the colonic mucosa is graded as *mild* with minimal mucosal erythema or as *moderate* with pale mucosal ulcerations and evidence of extension to the muscular layer of the bowel wall. *Severe* ischemic colitis presents with severe ulcerations resulting in black or green discoloration of the mucosa, consistent with full-thickness bowel-wall necrosis. The degree of reversibility can be predicted from the mucosal findings: Mild erythema is nearly 100% reversible, moderate ~50%, and frank necrosis is simply dead bowel. Follow-up colonoscopy can be performed to rule out progression of ischemic colitis.

Laparotomy for nonocclusive mesenteric ischemia is warranted for signs of peritonitis or worsening endoscopic findings and if the patient's condition does not improve with aggressive resuscitation. Ischemic colitis is optimally treated with resection of the ischemic bowel and formation of a proximal stoma. Primary anastomosis should not be performed in patients with acute intestinal ischemia.

Patients with MVT may present with a gradual or sudden onset. Symptoms include vague abdominal pain, nausea, and vomiting. Examination findings include abdominal distention with mild to moderate tenderness and signs of dehydration. The diagnosis of mesenteric thrombosis is frequently made on abdominal spiral CT with oral and IV contrast. Findings on CT include bowel-wall thickening and ascites. Intravenous contrast will demonstrate a delayed arterial phase and clot within the superior mesenteric vein. The goal of management is to optimize hemodynamics and correct electrolyte abnormalities with massive fluid resuscitation. Intravenous antibiotics as well as anticoagulation should be initiated. If laparotomy is performed and MVT is suspected, heparin anticoagulation is immediately

initiated and compromised bowel is resected. Of all acute intestinal disorders, mesenteric venous insufficiency is associated with the best prognosis.

Chronic intestinal ischemia presents with intestinal angina or abdominal pain associated with need for increased blood flow to the intestine. Patients report abdominal cramping and pain following ingestion of a meal. Weight loss and chronic diarrhea may also be noted. Abdominal pain without weight loss is not chronic mesenteric angina. Physical examination will often reveal the presence of an abdominal bruit as well as other manifestations of atherosclerosis. Duplex ultrasound evaluation of the mesenteric vessels has gained in popularity. In the absence of obesity and an increased bowel gas pattern, the radiologist may be able to identify flow disturbances within the vessels or the lack of a vasodilation response to feeding. This tool is frequently used as a screening test for patients with symptoms suggestive of chronic mesenteric ischemia. The gold standard for confirmation of mesenteric arterial occlusion is mesenteric angiography. Evaluation with mesenteric angiography allows for identification and possible intervention for the treatment of thrombus within the vessel lumen and will also evaluate the patency of remaining mesenteric vessels. The use of mesenteric angiography may be limited in the presence of renal failure or contrast allergy. Magnetic resonance angiography is an alternative if the administration of contrast dye is contraindicated.

The management of chronic intestinal ischemia includes medical management of atherosclerotic disease by lipid-lowering medications, exercise, and cessation of smoking. A full cardiac evaluation should be performed before intervention. Newer endovascular procedures may avoid an operative intervention in selected patient populations. Angioplasty with endovascular stenting in the treatment of chronic mesenteric ischemia is associated with an 80% long-term success rate. In patients requiring surgical exploration, the approach used is determined by the mesenteric angiogram. The entire length of the small and large bowel should be evaluated, beginning at the ligament of Treitz. Restoration of blood flow at the time of laparotomy is accomplished with mesenteric bypass.

Determination of intestinal viability intraoperatively in patients with suspected intestinal ischemia can be challenging. After revascularization, the bowel wall should be observed for return of a pink color and peristalsis. Palpation of major arterial vessels can be performed as well as applying a doppler flowmeter to the antimesenteric border of the bowel wall, but neither is a definitive indicator of viability. In equivocal cases, 1 g of IV sodium fluorescein is administered and the pattern of bowel reperfusion is observed under ultraviolet illumination with a standard (3600 A) Wood's lamp. An area of nonfluorescence >5 mm in diameter suggests nonviability. If doubt persists, reexploration performed 24–48 h following surgery will allow demarcation of nonviable bowel. Primary intestinal anastomosis in patients with ischemic bowel is always worrisome, and reanastomosis should be deferred to the time of second-look laparotomy.

ACKNOWLEDGMENTS

We thank Cory Sandore for providing some illustrations for this chapter. Gregory Bulkley contributed to this chapter in the 16th edition.

FURTHER READINGS

Hsu H et al: Impact of etiological factors and APACHE II and POSSUM scores in management and clinical outcomes of acute intestinal ischemic disease after surgical treatment. World J Surg 30:2152, 2006[PMID: 17103101] [Full Text]

Matsumoto AH et al: Percutaneous transluminal angioplasty and stenting in the treatment of chronic mesenteric ischemia: Results and long-term follow-up. J Am Coll Surg 194:S22, 2002

Mitchell EL, Moneta GL: Mesenteric duplex scanning. Perspect Vasc Surg Endovasc Ther 18:175, 2006[PMID: 17060239] [Full Text]

Shih MC et al: CTA and MRA in mesenteric ischemia: Part 2, normal findings and complications after surgical and endovascular treatment. AJR Am J Roentgenol 188:462, 2007[PMID: 17242256] [Full Text]

Sise MJ: Mesenteric ischemia: the whole spectrum. Scand J Surg 99:106, 2010[PMID: 20679047] [Full Text]

Wyers MC: Acute mesenteric ischemia: Diagnostic approach and surgical treatment. Semin Vasc Surg 23:9, 2010[PMID: 20298945] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ETIOLOGY AND CLASSIFICATION

In 75% of patients, acute intestinal obstruction results from adhesive bands or internal hernias secondary to previous abdominal surgery or from external hernias. The incidence of acute intestinal obstruction requiring hospital admission within the first few postoperative weeks is 5–25%, and 10–50% of these patients will require surgical intervention. The incidence of postoperative intestinal obstruction may be lower following laparoscopic surgery than open procedures. However, the laparoscopic gastric bypass procedure may be associated with an unexpected high rate of intestinal obstruction, with a higher reoperative rate. Other causes of intestinal obstruction not related to previous abdominal surgery include lesions *intrinsic* to the wall of the intestine, e.g., diverticulitis, carcinoma, and regional enteritis; and luminal obstruction, e.g., gallstone obstruction, intussusception.

Two other conditions that must be differentiated from acute intestinal obstruction include *adynamic ileus* and *primary intestinal pseudo-obstruction*. Adynamic ileus is mediated via the hormonal component of the sympathoadrenal system and may occur after any peritoneal insult; its severity and duration will be dependent to some degree on the type of peritoneal injury. Hydrochloric acid, colonic contents, and pancreatic enzymes are among the most irritating to the peritoneum, whereas blood and urine are less so. Adynamic ileus occurs to some degree after any abdominal operation. Retroperitoneal hematoma, particularly associated with vertebral fracture, may cause severe adynamic ileus, and the latter may occur with other retroperitoneal conditions, such as ureteral calculus or severe pyelonephritis. Thoracic diseases, including lower-lobe pneumonia, fractured ribs, and myocardial infarction, frequently produce adynamic ileus, as do electrolyte disturbances, particularly potassium depletion. Finally, intestinal ischemia, whether from vascular occlusion or intestinal distention itself, may perpetuate an adynamic ileus. Intestinal pseudo-obstruction is a chronic motility disorder that frequently mimics mechanical obstruction. This condition is often exacerbated by narcotic use.

Pathophysiology

Distention of the intestine is caused by the accumulation of gas and fluid proximal to and within the obstructed segment. Between 70 and 80% of intestinal gas consists of swallowed air, and because this is composed mainly of nitrogen, which is poorly absorbed from the intestinal lumen, removal of air by continuous gastric suction is a useful adjunct in the treatment of intestinal distention. The accumulation of fluid proximal to the obstructing mechanism results not only from ingested fluid, swallowed saliva, gastric juice, and biliary and pancreatic secretions but also from interference with normal sodium and water transport. During the first 12–24 h of obstruction, a marked depression of flux from lumen to blood of sodium and water occurs in the distended proximal intestine. After 24 h, sodium and water move into the lumen, contributing further to the distention and fluid losses. Intraluminal pressure rises from a normal of 2–4 cmH_2O to 8–10 cmH_2O. The loss of fluids and electrolytes may be extreme, and unless replacement is prompt, hypovolemia, renal insufficiency, and shock may result. Vomiting, accumulation of fluids within the lumen, and the sequestration of fluid into the edematous intestinal wall and peritoneal cavity as a result of impairment of venous return from the intestine all contribute to massive loss of fluid and electrolytes.

A "closed loop" is the most feared complication of acute intestinal obstruction. Closed-loop obstruction results when the lumen is occluded at two points by a single mechanism such as a fascial hernia or adhesive band, thus producing a closed loop, the blood supply of which is also often occluded by the hernia or band. During peristalsis, when a "closed loop" is present, pressures reach 30–60 cmH_2O. Strangulation of the closed loop is common in association with marked distention proximal to the involved loop. A form of closed-loop obstruction is encountered when complete obstruction of the colon exists in the presence of a competent ileocecal valve (85% of individuals). Although the blood supply of the colon is not entrapped within the obstructing mechanism, distention of the cecum is extreme because of its greater diameter (Laplace's law), and impairment of the intramural blood supply is considerable, with consequent gangrene of the cecal wall. Once impairment of blood supply to the gastrointestinal tract occurs, bacterial invasion supervenes, and peritonitis develops. The systemic effects of extreme distention include elevation of the diaphragm with restricted ventilation and subsequent atelectasis. Venous return via the inferior vena cava may also be impaired.

Symptoms

Mechanical intestinal obstruction is characterized by cramping midabdominal pain, which tends to be

more severe the higher the obstruction. The pain occurs in paroxysms, and the patient is relatively comfortable in the intervals between the pains. Audible borborygmi are often noted by the patient simultaneously with the paroxysms of pain. The pain may become less severe as distention progresses, probably because motility is impaired in the edematous intestine. When strangulation is present, the pain is usually more localized and may be steady and severe without a colicky component, a fact that often causes delay in diagnosis of obstruction. Vomiting is almost invariable, and it is earlier and more profuse the higher the obstruction. The vomitus initially contains bile and mucus and remains as such if the obstruction is high in the intestine. With low ileal obstruction, the vomitus becomes feculent, i.e., orange-brown in color with a foul odor, which results from the overgrowth of bacteria proximal to the obstruction. Hiccups (singultus) are common. Obstipation and failure to pass gas by rectum are invariably present when the obstruction is complete, although some stool and gas may be passed spontaneously or after an enema shortly after onset of the complete obstruction. Diarrhea is occasionally observed in partial obstruction. Blood in the stool is rare but does occur in cases of intussusception.

In *adynamic ileus* as well as *colonic pseudo-obstruction*, colicky pain is absent and only discomfort from distention is evident. Vomiting may be frequent but is rarely profuse. Complete obstipation may or may not occur. Singultus (hiccups) is common.

Physical Findings

Abdominal distention is the hallmark of all forms of intestinal obstruction. It is least marked in cases of obstruction high in the small intestine and most marked in colonic obstruction. In early obstruction of the small and large intestine, tenderness and rigidity are usually minimal; the temperature is rarely >37.8°C (100°F). The appearance of shock, tenderness, rigidity, and fever indicates that contamination of the peritoneum with infected intestinal content has occurred. Hernial orifices should always be carefully examined for the presence of a mass. Auscultation may reveal loud, high-pitched borborygmi coincident with colicky pain, but this finding is often absent late in strangulating or nonstrangulating obstruction. A quiet abdomen does not eliminate the possibility of obstruction, nor does it necessarily establish the diagnosis of adynamic ileus. The presence of a palpable abdominal mass usually signifies a closed-loop strangulating small-bowel obstruction; the tense fluid-filled loop is the palpable lesion.

Laboratory and X-Ray Findings

Laboratory and radiographic studies are used to help differentiate the two important clinical aspects of this disorder: strangulation vs. nonstrangulation and partial vs. complete obstruction. Leukocytosis, with shift to the left, usually occurs when strangulation is present, but a normal white blood cell count does not exclude strangulation. Elevation of the serum amylase level is encountered occasionally in all forms of intestinal obstruction. Radiographic images demonstrating distention of fluid- and gas-filled loops of small intestine usually arranged in a "stepladder" pattern with air-fluid levels and an absence or paucity of colonic gas are pathognomonic for small-bowel obstruction. Complete obstruction is suggested when passage of gas or stool per rectum has ceased and when gas is absent in the distal intestine by x-ray. A general haze due to peritoneal fluid and sometimes a "coffee bean"–shaped mass are seen in strangulating closed loop obstruction. A thin barium upper gastrointestinal series may help to differentiate partial from complete obstruction. However, thick barium given by mouth should be avoided when the obstruction is considered to be high grade or complete since retained barium sulfate may become inspissated and either make an incomplete obstruction complete or be aspirated into the tracheobronchial tree. CT is the most commonly used modality to evaluate patients for intestinal obstruction but differentiating adynamic ileus, partial obstruction, and complete obstruction may be difficult **(Fig. 299-1)**. The sensitivity and specificity of CT for strangulating obstruction are low (50 and 80%, respectively).

Figure 299-1

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

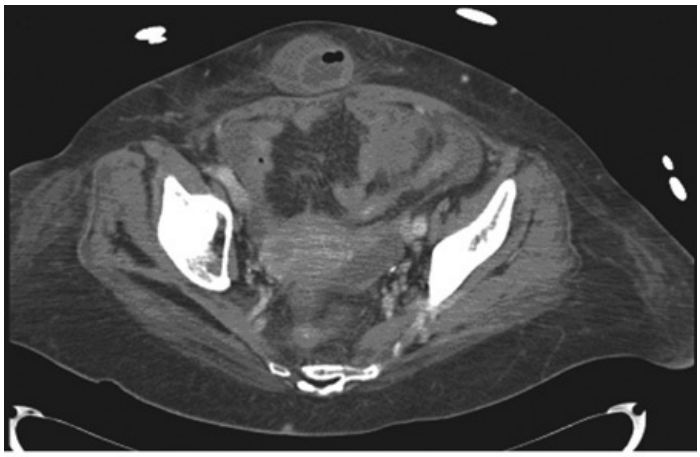

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

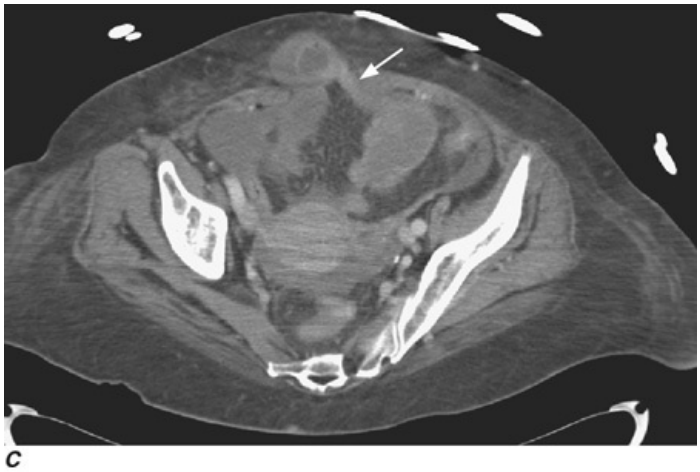

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT with oral and intravenous contrast demonstrating **(*A*)** evidence of small-bowel dilatation with air-fluid levels consistent with a small-bowel obstruction; **(*B*)** a partial small-bowel obstruction from an incarcerated ventral hernia (*arrow*); and **(*C*)** decompressed bowel seen distal to the hernia (*arrow*).

Common causes of colonic obstruction can be seen on abdominal radiographic series. These films may demonstrate a "bird's beak" sign when a sigmoid volvulus has occurred or an enlarged cecum when a cecal torsion or bascule is present. Colonic obstruction with a competent ileocecal valve is easily recognized because distention with gas is mainly confined to the colon. Gastrografin enema may help in demonstrating a complete colonic obstruction. Furthermore, *barium should never be given by*

mouth to a patient with a possible colonic obstruction until that possibility has been excluded.

Treatment: Acute Intestinal Obstruction

SMALL-INTESTINAL OBSTRUCTION

The overall mortality rate for obstruction of the small intestine is about 10%. While the mortality rate for nonstrangulating obstruction is 5–8%, the mortality rate for a strangulating obstruction ranges from 20 to 75%. Since strangulating small-bowel obstruction is always complete, surgical interventions should always be undertaken in such patients after suitable preparation. Before operating, fluid and electrolyte balance should be restored and decompression instituted by means of a nasogastric tube. Replacement of potassium is especially important because intake is nil and losses in vomitus are large. A long intestinal tube is not indicated. Operative intervention may be undertaken successfully by laparoscopic techniques with a decreased incidence of wound complications. However, laparoscopic lysis of adhesions is associated with a longer operative time and higher conversion to open rate when compared to other laparoscopic procedures. Alternatively, lysis of adhesions can be achieved through an open abdominal incision. In general, >50% of adhesions that occur are found at the previous incision site. Purely nonoperative therapy is safe only in the presence of incomplete obstruction and is best used in patients without increasing abdominal pain or leukocytosis. The overall recurrence of small-bowel obstruction is 16%. Population-based studies show that although the surgical management of small-bowel obstruction is associated with longer hospital stays, the rate of readmission for obstruction is lower. However, regardless of treatment type, following the index admission, only 20% of patients required readmission within a 5-year follow-up period.

COLONIC OBSTRUCTION

The mortality rate for colonic obstruction is about 20%. As in small-bowel obstruction, nonoperative treatment is contraindicated unless the obstruction is incomplete. Incomplete obstruction can be treated with colonoscopic decompression and placement of a metallic stent if a malignant lesion is present. The success rate approaches 90% depending on the location of the obstruction, with left-sided lesions being more successfully stented than right-sided lesions. In general, the colonic stent is considered to be a temporary solution or a "bridge to surgery," which allows for colonic preparation before surgical intervention. When obstruction is complete, early operation is mandatory, especially when the ileocecal valve is competent, because of the concern for cecal perforation. Cecal perforation is more likely if the cecal diameter is >10 cm on plain abdominal film.

Decisions regarding the operative management of colonic obstruction are based on the cause of the obstruction and the patient's overall well-being. For obstruction on the left side of the colon, operative management strategies include either decompression by cecostomy or transverse colostomy or resection with end-colostomy formation (Hartmann's procedure). Primary resection of obstructing left-sided lesions with on-table washout of the colon has also been accomplished safely. For a lesion of the right or transverse colon, primary resection and anastomosis can be performed safely because distention of the ileum with consequent discrepancy in size and hazard in suture are usually not present. Furthermore, the bacterial and stool content is less on the right side of the colon, decreasing the chance of infection.

ADYNAMIC ILEUS

This type of ileus usually responds to nonoperative decompression and treatment of the primary disease. The prognosis is usually good. Correction of electrolyte abnormalities should be instituted (i.e., potassium, magnesium). Successful decompression of a colonic ileus has been accomplished by repetitive colonoscopy. Neostigmine is also effective in cases of colonic ileus that have not responded to other conservative treatment. Rarely, adynamic colonic distention may become so great that cecostomy is required if cecal gangrene is feared.

FURTHER READINGS

Dubois A et al: Postoperative ileus: Physiopathology, etiology and treatment. Ann Surg 178:781, 1973[PMID: 4759410] [Full Text]

Eskelinen M et al: Contributions of history-taking, physical examination, and computer assistance to diagnosis of acute small-bowel obstruction. A prospective study of 1333 patients with acute abdominal pain. Scand J Gastroenterol 29:715, 1994[PMID: 7973431] [Full Text]

Fevang BT et al: Complications and death after surgical treatment of small bowel obstruction: A 35-year institutional experience. Ann Surg 231:529, 2000[PMID: 10749614] [Full Text]

Jackson BR: The diagnosis of colonic obstruction. Dis Colon Rectum 25:603, 1982[PMID: 7117069] [Full Text]

Kucukmetin A et al: Palliative surgery versus medical management for bowel obstruction in ovarian cancer. Cochrane Database Syst Rev 7:CD007792, 2010

Silen W: *Cope's Early Diagnosis of the Acute Abdomen* , 21st ed. London, Oxford, 2005

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACUTE APPENDICITIS

Incidence and Epidemiology

With more than 250,000 appendectomies performed annually, appendicitis is the most common abdominal surgical emergency in the United States. The peak incidence of acute appendicitis is in the second and third decades of life; it is relatively rare at the extremes of age. However, perforation is more common in infancy and in the elderly, during which periods mortality rates are highest. Males and females are equally affected, except between puberty and age 25, when males predominate in a 3:2 ratio. The incidence of appendicitis has remained stable in the United States over the last 30 years, while the incidence of appendicitis is much lower in underdeveloped countries, especially parts of Africa, and in lower socioeconomic groups. The mortality rate in the United States decreased eightfold between 1941 and 1970 but has remained at <1 per 100,000 since then.

Pathogenesis

Appendicitis is believed to occur as a result of appendiceal luminal obstruction. Obstruction is most commonly caused by a fecalith, which results from accumulation and inspissation of fecal matter around vegetable fibers. Enlarged lymphoid follicles associated with viral infections (e.g., measles), inspissated barium, worms (e.g., pinworms, *Ascaris*, and *Taenia*), and tumors (e.g., carcinoid or carcinoma) may also obstruct the lumen. Other common pathologic findings include appendiceal ulceration. The cause of the ulceration is unknown, although a viral etiology has been postulated. Infection with *Yersinia* organisms may cause the disease, since high complement fixation antibody titers have been found in up to 30% of cases of proven appendicitis. Luminal bacteria multiply and invade the appendiceal wall as venous engorgement and subsequent arterial compromise result from the high intraluminal pressures. Finally, gangrene and perforation occur. If the process evolves slowly, adjacent organs such as the terminal ileum, cecum, and omentum may wall off the appendiceal area so that a localized abscess will develop, whereas rapid progression of vascular impairment may cause perforation with free access to the peritoneal cavity. Subsequent rupture of primary appendiceal abscesses may produce fistulas between the appendix and bladder, small intestine, sigmoid, or cecum. Occasionally, acute appendicitis may be the first manifestation of Crohn's disease.

While chronic infection of the appendix with tuberculosis, amebiasis, and actinomycosis may occur, a useful clinical aphorism states that *chronic appendiceal inflammation is not usually the cause of prolonged abdominal pain of weeks' or months' duration*. In contrast, recurrent acute appendicitis does occur, often with complete resolution of inflammation and symptoms between attacks. Recurrent acute appendicitis may also occur if a long appendiceal stump is left after initial appendectomy.

CLINICAL MANIFESTATIONS

The sequence of abdominal discomfort and anorexia associated with acute appendicitis is pathognomonic. The pain is described as being located in the periumbilical region initially and then migrating to the right lower quadrant. This classic sequence of symptoms occurs in only 66% of patients. The differential diagnoses for periumbilical and right lower quadrant pain is listed in **Table 300-1**. The periumbilical abdominal pain is of the visceral type, resulting from distention of the appendiceal lumen. This pain is carried on slow-conducting C fibers and is usually poorly localized in the periumbilical or epigastric region. In general, this visceral pain is mild, often cramping and usually lasting 4–6 h, but it may not be noted by stoic individuals. As inflammation spreads to the parietal peritoneal surfaces, the pain becomes somatic, steady, and more severe and aggravated by motion or cough. Parietal afferent nerves are A delta fibers, which are fast-conducting and unilateral. These fibers localize the pain to the *right lower quadrant* . *Anorexia* is very common; a hungry patient almost invariably does not have acute appendicitis. *Nausea* and *vomiting* occur in 50–60% of cases, but vomiting is usually self-limited. Change in bowel habit is of little diagnostic value, since any or no alteration may be observed, although the presence of diarrhea caused by an inflamed appendix in juxtaposition to the sigmoid may cause diagnostic difficulties. Urinary frequency and dysuria occur if the appendix lies adjacent to the bladder.

Table 300-1 The Anatomic Origin of Periumbilical and Right Lower Quadrant Pain in the Differential Diagnosis of Appendicitis

Periumbilical

Appendicitis	
Small-bowel obstruction Gastroenteritis	
Mesenteric ischemia	
Right Lower Quadrant	
Gastrointestinal causes	Gynecologic causes
Appendicitis	Ovarian tumor/torsion
Inflammatory bowel disease	Pelvic inflammatory disease
Right-sided diverticulitis	Renal causes
Gastroenteritis Inguinal hernia	Pyelonephritis Perinephritic abscess
	Nephrolithiasis

Physical findings vary with time after onset of the illness and according to the location of the appendix, which may be situated deep in the pelvic cul-de-sac; in the right lower quadrant in any relation to the peritoneum, cecum, and small intestine; in the right upper quadrant (especially during pregnancy); or even in the left lower quadrant. *The diagnosis cannot be established unless tenderness can be elicited*. While tenderness is sometimes absent in the early visceral stage of the disease, it ultimately always develops and is found in any location corresponding to the position of the appendix. Typically, tenderness to palpation will often occur at McBurney's point, anatomically located on a line one-third of the way between the anterior iliac spine and the umbilicus. Abdominal tenderness may be completely absent if a retrocecal or pelvic appendix is present, in which case the sole physical finding may be tenderness in the flank or on rectal or pelvic examination. Referred rebound tenderness is often present and is most likely to be absent early in the illness. Flexion of the right hip and guarded movement by the patient are due to parietal peritoneal involvement. Hyperesthesia of the skin of the right lower quadrant and a positive psoas or obturator sign are often late findings and are rarely of diagnostic value.

The temperature is usually normal or slightly elevated [37.2°–38°C (99°–100.5°F)], but a temperature >38.3°C (101°F) should suggest perforation. Tachycardia is commensurate with the elevation of the temperature. Rigidity and tenderness become more marked as the disease progresses to perforation and localized or diffuse peritonitis. Distention is rare unless severe diffuse peritonitis has developed. A mass may develop if localized perforation has occurred but will not usually be detectable before 3 days after onset. Earlier presence of a mass suggests carcinoma of the cecum or Crohn's disease. Perforation is rare before 24 h after onset of symptoms, but the rate may be as high as 80% after 48 h.

Although moderate leukocytosis of 10,000–18,000 cells/μL is frequent (with a concomitant left shift), the absence of leukocytosis does not rule out acute appendicitis. Leukocytosis of >20,000 cells/μL suggests probable perforation. Anemia and blood in the stool suggest a primary diagnosis of carcinoma of the cecum, especially in elderly individuals. The urine may contain a few white or red blood cells without bacteria if the appendix lies close to the right ureter or bladder. Urinalysis is most useful in excluding genitourinary conditions that may mimic acute appendicitis.

Radiographs are rarely of value except when an opaque fecalith (5% of patients) is observed in the right lower quadrant (especially in children). Consequently, abdominal films are not routinely obtained unless other conditions such as intestinal obstruction or ureteral calculus may be present. The diagnosis may also be established by the ultrasonic demonstration of an enlarged and thick-walled appendix. Ultrasound is most useful to exclude ovarian cysts, ectopic pregnancy, or tuboovarian abscess. Several studies have recently demonstrated the benefit of contrast-enhanced or nonenhanced CT over ultrasound and plain radiographs in the diagnosis of acute appendicitis. The findings on CT will include a thickened appendix with periappendiceal stranding and often the presence of a fecalith **(Figs. 300-1and300-2)**. The reported positive predictive value of CT is 95–97% and the overall accuracy is 90–98%. Furthermore, nonvisualization of the appendix on CT is associated with the findings of a normal appendix 98% of the time. Free peritoneal air is uncommon, even in perforated appendicitis.

Figure 300-1

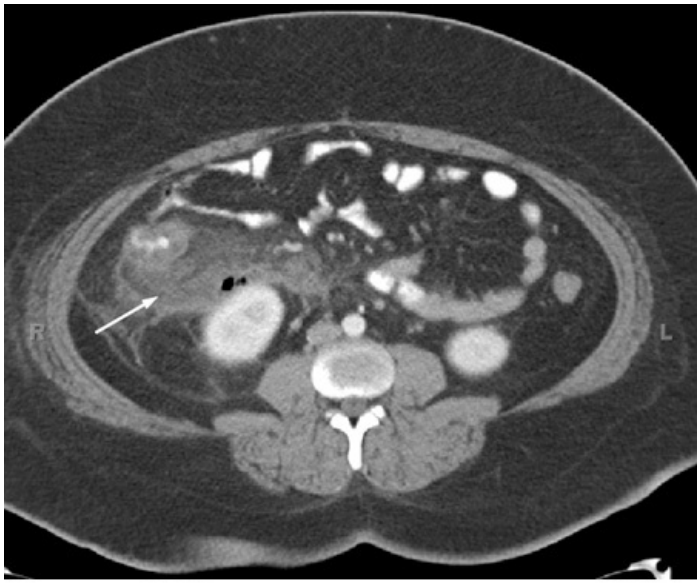

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT with oral and intravenous contrast of acute appendicitis. There is thickening of the wall of the appendix and periappendiceal stranding *(arrow)*.

Figure 300-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Appendiceal fecolith*(arrow)*.

While the typical historic sequence and physical findings are present in 50–60% of cases, a wide variety of atypical patterns of disease are encountered, especially at the age extremes and during pregnancy. Infants under 2 years of age have a 70–80% incidence of perforation and generalized peritonitis. This is thought to be the result of a delay in diagnosis. Any infant or child with diarrhea, vomiting, and abdominal pain is highly suspect. Fever is much more common in this age group, and abdominal distention is often the only physical finding. In the elderly, pain and tenderness are often blunted, and thus the diagnosis is also frequently delayed and leads to a 30% incidence of perforation in patients older than age 70 years. Elderly patients often present initially with a slightly painful mass (a primary appendiceal abscess) or with adhesive intestinal obstruction 5 or 6 days after a previously undetected perforated appendix.

Appendicitis occurs about once in every 500–2000 pregnancies and is the most common extrauterine condition requiring abdominal operation. The diagnosis may be missed or delayed because of the

frequent occurrence of mild abdominal discomfort and nausea and vomiting during pregnancy, and because of the gradual shift of the appendix from the right lower quadrant to the right upper quadrant during the second and third trimester of pregnancy. Appendicitis tends to be most common during the second trimester. The diagnosis is best made with ultrasound, which is 80% accurate; however, if perforation has already occurred, the accuracy of ultrasound decreases to 30%. Early intervention is warranted because the incidence of fetal loss with a normal appendix is 1.5%. With perforation, the incidence of fetal loss is 20–35%.

Differential Diagnosis

Acute appendicitis has been labeled the *masquerader*, and the diagnosis is often more difficult to make in young females. Obtaining a good history, including sexual activity and the presence of a vaginal discharge, will help differentiate acute appendicitis from pelvic inflammatory disease (PID). The presence of a malodorous vaginal discharge and gram-negative intracellular diplococci are pathognomonic for PID. Pain on movement of the cervix is also more specific for PID but may occur in appendicitis if perforation has occurred or if the appendix lies adjacent to the uterus or adnexa. *Rupture of a graafian follicle* (mittelschmerz) occurs at midcycle and will produce pain and tenderness more diffuse and usually of a less severe degree than in appendicitis. *Rupture of a corpus luteumcyst* is identical clinically to rupture of a graafian follicle but develops about the time of menstruation. The presence of an adnexal mass, evidence of blood loss, and a positive pregnancy test help differentiate *ruptured tubal pregnancy*. *Twisted ovarian cyst* and *endometriosis* are occasionally difficult to distinguish from appendicitis. In all these female conditions, ultrasonography and laparoscopy may be of great value.

Acute mesenteric lymphadenitis and *acute gastroenteritis* are the diagnoses usually given when enlarged, slightly reddened lymph nodes at the root of the mesentery and a normal appendix are encountered at operation in a patient who usually has right lower quadrant tenderness. Retrospectively, these patients may have had a higher temperature, diarrhea, more diffuse pain and abdominal tenderness, and a lymphocytosis. Between cramps, the abdomen is completely relaxed. Children seem to be affected more frequently than adults. Some of these patients have infection with *Y. pseudotuberculosis* or *Y. enterocolitica*, in which case the diagnosis can be established by culture of the mesenteric nodes or by serologic titers **(Chap. 159)**. In *Salmonella* gastroenteritis, the abdominal findings are similar, although the pain may be more severe and more localized, and fever and chills are common. The occurrence of similar symptoms among other members of the family may be helpful. *Regional enteritis* (Crohn's disease) is usually associated with a more prolonged history, often with previous exacerbations regarded as episodes of gastroenteritis unless the diagnosis has been established previously. Often an inflammatory mass is palpable. In addition, acute cholecystitis, perforated ulcer, acute pancreatitis, acute diverticulitis, strangulating intestinal obstruction, ureteral calculus, and pyelonephritis may present diagnostic difficulties.

Treatment: Acute Appendicitis

If the diagnosis is in question, 4–6 h of observation with serial abdominal exams is always more beneficial than harmful. Antibiotics should not be administered when the diagnosis is in question, since they will only mask the perforation. The treatment of presumed acute appendicitis is early operation and appendectomy as soon as the patient can be prepared. Appendectomy is frequently accomplished laparoscopically and is associated with less postoperative narcotic use and earlier discharge. It is acceptable to have a 15–20% incidence of a normal appendix at the time of appendectomy to avoid perforation. The use of early laparoscopy instead of close clinical observation has not shown a clinical benefit in the management of patients with nonspecific abdominal pain.

A different approach is indicated if a palpable mass is found 3–5 days after the onset of symptoms. This finding usually represents the presence of a phlegmon or abscess, and complications from attempted surgical excision are frequent. Such patients treated with broad-spectrum antibiotics, drainage of abscesses >3 cm, parenteral fluids, and bowel rest usually show resolution of symptoms within 1 week. *Interval appendectomy* can be performed safely 6–12 weeks later. A randomized clinical trial has demonstrated that antibiotics alone can effectively treat acute, nonperforated appendicitis in 86% of male patients. However, antibiotics alone were associated with a higher recurrence rate than when followed by surgical intervention. If the mass enlarges or the patient becomes more toxic, the abscess should be drained. Free perforation is associated with generalized peritonitis and its complications, including subphrenic, pelvic, or other abscesses, and can be avoided by early diagnosis. The mortality rate for nonperforated appendicitis is 0.1%, little more than the risk of general anesthesia; for perforated appendicitis, mortality is 3% (and can reach 15% in the elderly).

ACUTE PERITONITIS

Peritonitis is an inflammation of the peritoneum; it may be localized or diffuse in location, acute or chronic in natural history, and infectious or aseptic in pathogenesis. Acute peritonitis is most often infectious and is usually related to a perforated viscus (and called *secondary peritonitis*). When no intraabdominal source is identified, infectious peritonitis is called *primary* or *spontaneous*. Acute peritonitis is associated with decreased intestinal motor activity, resulting in distention of the intestinal

lumen with gas and fluid (adynamic ileus). The accumulation of fluid in the bowel together with the lack of oral intake leads to rapid intravascular volume depletion with effects on cardiac, renal, and other systems.

Etiology

Infectious agents gain access to the peritoneal cavity through a perforated viscus, a penetrating wound of the abdominal wall, or external introduction of a foreign object that is or becomes infected (e.g., a chronic peritoneal dialysis catheter). In the absence of immune compromise, host defenses are capable of eradicating small contaminations. The conditions that most commonly result in the introduction of bacteria into the peritoneum are ruptured appendix, ruptured diverticulum, perforated peptic ulcer, incarcerated hernia, gangrenous gall bladder, volvulus, bowel infarction, cancer, inflammatory bowel disease, or intestinal obstruction. However, a wide range of mechanisms may play a role **(Table 300-2)**. Bacterial peritonitis can also occur in the apparent absence of an intraperitoneal source of bacteria (primary or spontaneous bacterial peritonitis). This condition occurs in the setting of ascites and liver cirrhosis in 90% of the cases, usually in patients with ascites with low protein concentration (<1 g/L) **(Chap. 308)**. **Bacterial peritonitis is discussed in detail in Chap. 127**.

Table 300-2 Conditions Leading to Secondary Bacterial Peritonitis

Perforations of bowel	**Perforations or leaking of other organs**
Trauma, blunt or penetrating	Pancreas—pancreatitis
Inflammation	Gallbladder—cholecystitis
Appendicitis	Urinary bladder—trauma, rupture
Diverticulitis	Liver—bile leak after biopsy
Peptic ulcer disease	Fallopian tubes—salpingitis
Inflammatory bowel disease	Bleeding into the peritoneal cavity
Iatrogenic	**Disruption of integrity of peritoneal cavity**
Endoscopic perforation	Trauma
Anastomotic leaks	Continuous ambulatory peritoneal dialysis (indwelling catheter)
Catheter perforation	Intraperitoneal chemotherapy
Vascular	Perinephric abscess
Embolus	Iatrogenic—postoperative, foreign body
Ischemia	
Obstructions	
Adhesions	
Strangulated hernias	
Volvulus	
Intussusception	
Neoplasms	
Ingested foreign body, toothpick, fish bone	

Aseptic peritonitis may be due to peritoneal irritation by abnormal presence of physiologic fluids (e.g., gastric juice, bile, pancreatic enzymes, blood, or urine) or sterile foreign bodies (e.g., surgical sponges or instruments, starch from surgical gloves) in the peritoneal cavity or as a complication of rare systemic diseases such as lupus erythematosus, porphyria, or familial Mediterranean fever **(Chap. 330)**. Chemical irritation of the peritoneum is greatest for acidic gastric juice and pancreatic enzymes. Secondary bacterial infection is common in chemical peritonitis.

Clinical Features

The cardinal manifestations of peritonitis are acute abdominal pain and tenderness, usually with fever. The location of the pain depends on the underlying cause and whether the inflammation is localized or generalized. Localized peritonitis is most common in uncomplicated appendicitis and diverticulitis, and physical findings are limited to the area of inflammation. Generalized peritonitis is associated with widespread inflammation and diffuse abdominal tenderness and rebound. Rigidity of the abdominal wall is common in both localized and generalized peritonitis. Bowel sounds are usually but not always absent. Tachycardia, hypotension, and signs of dehydration are common. Leukocytosis and marked acidosis are common laboratory findings. Plain abdominal films may show dilation of large and small bowel with edema of the bowel wall. Free air under the diaphragm is associated with a perforated

viscus. CT and/or ultrasonography can identify the presence of free fluid or an abscess. When ascites is present, diagnostic paracentesis with cell count (>250 neutrophils/μL is usual in peritonitis), protein and lactate dehydrogenase levels, and culture is essential. In elderly and immunosuppressed patients, signs of peritoneal irritation may be more difficult to detect.

Therapy and Prognosis

Treatment relies on rehydration, correction of electrolyte abnormalities, antibiotics, and surgical correction of the underlying defect. Mortality rates are <10% for uncomplicated peritonitis associated with a perforated ulcer or ruptured appendix or diverticulum in an otherwise healthy person. Mortality rates of ≥40% have been reported for elderly people, those with underlying illnesses, and when peritonitis has been present for >48 h.

FURTHER READINGS

Anderson RE: The natural history and traditional management of appendicitis revisited: Spontaneous resolution and predominance of prehospital perforations imply that a correct diagnosis is more important than an early diagnosis. World J Surg 31:86, 2007

Flum DR et al: Has misdiagnosis of appendicitis decreased over time? A population-based analysis. JAMA 286:1748, 2001[PMID: 11594900] [Full Text]

Merlin MA et al: Evidence-based appendicitis: The initial work-up. Postgrad Med 122:189, 2010[PMID: 20463429] [Full Text]

Morino M et al: Acute non-specific abdominal pain: A randomized controlled study comparing early laparoscopy vs. clinical observation. Ann Surg 241:881, 2006

Solomkin JS, Mazuski J: Intra-abdominal sepsis: Newer interventional and antimicrobial therapies. Infect Dis Clin North Am 23:595, 2009

Styrud J et al: Appendectomy vs. antibiotic treatment in acute appendectomy: A prospective multicenter randomized controlled trial. World J Surg 30:1033, 2006[PMID: 16736333] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH LIVER DISEASE: INTRODUCTION

A diagnosis of liver disease usually can be made accurately by a careful history, physical examination, and application of a few laboratory tests. In some circumstances, radiologic examinations are helpful or, indeed, diagnostic. Liver biopsy is considered the criterion standard in evaluation of liver disease but is now needed less for diagnosis than for grading and staging of disease. This chapter provides an introduction to diagnosis and management of liver disease, briefly reviewing the structure and function of the liver; the major clinical manifestations of liver disease; and the use of clinical history, physical examination, laboratory tests, imaging studies, and liver biopsy.

Liver Structure and Function

The liver is the largest organ of the body, weighing 1–1.5 kg and representing 1.5–2.5% of the lean body mass. The size and shape of the liver vary and generally match the general body shape—long and lean or squat and square. The liver is located in the right upper quadrant of the abdomen under the right lower rib cage against the diaphragm and projects for a variable extent into the left upper quadrant. The liver is held in place by ligamentous attachments to the diaphragm, peritoneum, great vessels, and upper gastrointestinal organs. It receives a dual blood supply; ~20% of the blood flow is oxygen-rich blood from the hepatic artery, and 80% is nutrient-rich blood from the portal vein arising from the stomach, intestines, pancreas, and spleen.

The majority of cells in the liver are hepatocytes, which constitute two-thirds of the mass of the liver. The remaining cell types are Kupffer cells (members of the reticuloendothelial system), stellate (Ito or fat-storing) cells, endothelial cells and blood vessels, bile ductular cells, and supporting structures. Viewed by light microscopy, the liver appears to be organized in lobules, with portal areas at the periphery and central veins in the center of each lobule. However, from a functional point of view, the liver is organized into acini, with both hepatic arterial and portal venous blood entering the acinus from the portal areas (zone 1) and then flowing through the sinusoids to the terminal hepatic veins (zone 3); the intervening hepatocytes constituting zone 2. The advantage of viewing the acinus as the physiologic unit of the liver is that it helps to explain the morphologic patterns and zonality of many vascular and biliary diseases not explained by the lobular arrangement.

Portal areas of the liver consist of small veins, arteries, bile ducts, and lymphatics organized in a loose stroma of supporting matrix and small amounts of collagen. Blood flowing into the portal areas is distributed through the sinusoids, passing from zone 1 to zone 3 of the acinus and draining into the terminal hepatic veins ("central veins"). Secreted bile flows in the opposite direction, in a countercurrent pattern from zone 3 to zone 1. The sinusoids are lined by unique endothelial cells that have prominent fenestrae of variable size, allowing the free flow of plasma but not cellular elements. The plasma is thus in direct contact with hepatocytes in the subendothelial space of Disse.

Hepatocytes have distinct polarity. The basolateral side of the hepatocyte lines the space of Disse and is richly lined with microvilli; it demonstrates endocytotic and pinocytotic activity, with passive and active uptake of nutrients, proteins, and other molecules. The apical pole of the hepatocyte forms the canalicular membranes through which bile components are secreted. The canaliculi of hepatocytes form a fine network, which fuses into the bile ductular elements near the portal areas. Kupffer cells usually lie within the sinusoidal vascular space and represent the largest group of fixed macrophages in the body. The stellate cells are located in the space of Disse but are not usually prominent unless activated, when they produce collagen and matrix. Red blood cells stay in the sinusoidal space as blood flows through the lobules, but white blood cells can migrate through or around endothelial cells into the space of Disse and from there to portal areas, where they can return to the circulation through lymphatics.

Hepatocytes perform numerous and vital roles in maintaining homeostasis and health. These functions include the synthesis of most essential serum proteins (albumin, carrier proteins, coagulation factors, many hormonal and growth factors), the production of bile and its carriers (bile acids, cholesterol, lecithin, phospholipids), the regulation of nutrients (glucose, glycogen, lipids, cholesterol, amino acids), and metabolism and conjugation of lipophilic compounds (bilirubin, anions, cations, drugs) for excretion in the bile or urine. Measurement of these activities to assess liver function is complicated by the multiplicity and variability of these functions. The most commonly used liver "function" tests are measurements of serum bilirubin, albumin, and prothrombin time. The serum bilirubin level is a measure of hepatic conjugation and excretion, and the serum albumin level and prothrombin time are measures of protein synthesis. Abnormalities of bilirubin, albumin, and prothrombin time are typical of

hepatic dysfunction. Frank liver failure is incompatible with life, and the functions of the liver are too complex and diverse to be subserved by a mechanical pump; dialysis membrane; or concoction of infused hormones, proteins, and growth factors.

LIVER DISEASES

While there are many causes of liver disease **(Table 301-1)**, they generally present clinically in a few distinct patterns, usually classified as hepatocellular, cholestatic (obstructive), or mixed. In *hepatocellular diseases* (such as viral hepatitis or alcoholic liver disease), features of liver injury, inflammation, and necrosis predominate. In *cholestatic diseases* (such as gallstone or malignant obstruction, primary biliary cirrhosis, some drug-induced liver diseases), features of inhibition of bile flow predominate. In a mixed pattern, features of both hepatocellular and cholestatic injury are present (such as in cholestatic forms of viral hepatitis and many drug-induced liver diseases). The pattern of onset and prominence of symptoms can rapidly suggest a diagnosis, particularly if major risk factors are considered such as the age and sex of the patient and a history of exposure or risk behaviors.

Table 301-1 Liver Diseases

Inherited hyperbilirubinemia	**Liver involvement in systemic diseases**
Gilbert's syndrome	Sarcoidosis
Crigler-Najjar syndrome, types I and II	Amyloidosis
Dubin-Johnson syndrome	Glycogen storage diseases
Rotor syndrome	Celiac disease
Viral hepatitis	Tuberculosis
Hepatitis A	*Mycobacterium avium intracellulare*
Hepatitis B	**Cholestatic syndromes**
Hepatitis C	Benign postoperative cholestasis
Hepatitis D	Jaundice of sepsis
Hepatitis E	Total parenteral nutrition (TPN)-induced jaundice
Others (mononucleosis, herpes, adenovirus hepatitis)	Cholestasis of pregnancy
Cryptogenic hepatitis	Cholangitis and cholecystitis
Immune and autoimmune liver diseases	Extrahepatic biliary obstruction (stone, stricture, cancer)
Primary biliary cirrhosis	Biliary atresia
Autoimmune hepatitis	Caroli's disease
Sclerosing cholangitis	Cryptosporidiosis
Overlap syndromes	**Drug-induced liver disease**
Graft-versus-host disease	Hepatocellular patterns (isoniazid, acetaminophen)
Allograft rejection	Cholestatic patterns (methyltestosterone)
Genetic liver diseases	Mixed patterns (sulfonamides, phenytoin)
α_1 Antitrypsin deficiency	Micro- and macrovesicular steatosis (methotrexate, fialuridine)
Hemochromatosis	**Vascular injury**
Wilson's disease	Venoocclusive disease
Benign recurrent intrahepatic cholestasis (BRIC)	Budd-Chiari syndrome
Progressive familial intrahepatic cholestasis (PFIC), types I-III	Ischemic hepatitis
Others (galactosemia, tyrosinemia, cystic fibrosis, Newman-Pick disease, Gaucher's disease)	Passive congestion
Alcoholic liver disease	Portal vein thrombosis
Acute fatty liver	Nodular regenerative hyperplasia
Acute alcoholic hepatitis	**Mass lesions**
Laënnec's cirrhosis	Hepatocellular carcinoma
Nonalcoholic fatty liver	Cholangiocarcinoma
Steatosis	
Steatohepatitis	
Acute fatty liver of pregnancy	

	Adenoma Focal nodular hyperplasia Metastatic tumors Abscess Cysts Hemangioma

Typical presenting symptoms of liver disease include jaundice, fatigue, itching, right upper quadrant pain, nausea, poor appetite, abdominal distention, and intestinal bleeding. At present, however, many patients are diagnosed with liver disease who have no symptoms and who have been found to have abnormalities in biochemical liver tests as a part of a routine physical examination or screening for blood donation or for insurance or employment. The wide availability of batteries of liver tests makes it relatively simple to demonstrate the presence of liver injury as well as to rule it out in someone suspected of liver disease.

Evaluation of patients with liver disease should be directed at (1) establishing the etiologic diagnosis, (2) estimating the disease severity (grading), and (3) establishing the disease stage (staging). *Diagnosis* should focus on the category of disease such as hepatocellular, cholestatic, or mixed injury, as well as on the specific etiologic diagnosis. *Grading* refers to assessing the severity or activity of disease—active or inactive, and mild, moderate, or severe. *Staging* refers to estimating the place in the course of the natural history of the disease, whether acute or chronic; early or late; precirrhotic, cirrhotic, or end-stage.

The goal of this chapter is to introduce general, salient concepts in the evaluation of patients with liver disease that help lead to the diagnoses discussed in subsequent chapters.

Clinical History

The clinical history should focus on the symptoms of liver disease—their nature, patterns of onset, and progression—and on potential risk factors for liver disease. The symptoms of liver disease include constitutional symptoms such as fatigue, weakness, nausea, poor appetite, and malaise and the more liver-specific symptoms of jaundice, dark urine, light stools, itching, abdominal pain, and bloating. Symptoms can also suggest the presence of cirrhosis, end-stage liver disease, or complications of cirrhosis such as portal hypertension. Generally, the constellation of symptoms and their patterns of onset rather than a specific symptom points to an etiology.

Fatigue is the most common and most characteristic symptom of liver disease. It is variously described as lethargy, weakness, listlessness, malaise, increased need for sleep, lack of stamina, and poor energy. The fatigue of liver disease typically arises after activity or exercise and is rarely present or severe in the morning after adequate rest (afternoon versus morning fatigue). Fatigue in liver disease is often intermittent and variable in severity from hour to hour and day to day. In some patients, it may not be clear whether fatigue is due to the liver disease or to other problems such as stress, anxiety, sleep disturbance, or a concurrent illness.

Nausea occurs with more severe liver disease and may accompany fatigue or be provoked by odors of food or eating fatty foods. Vomiting can occur but is rarely persistent or prominent. Poor appetite with weight loss occurs commonly in acute liver diseases but is rare in chronic disease, except when cirrhosis is present and advanced. Diarrhea is uncommon in liver disease, except with severe jaundice, where lack of bile acids reaching the intestine can lead to steatorrhea.

Right upper quadrant discomfort or ache ("liver pain") occurs in many liver diseases and is usually marked by tenderness over the liver area. The pain arises from stretching or irritation of Glisson's capsule, which surrounds the liver and is rich in nerve endings. Severe pain is most typical of gallbladder disease, liver abscess, and severe venoocclusive disease but is an occasional accompaniment of acute hepatitis.

Itching occurs with acute liver disease, appearing early in obstructive jaundice (from biliary obstruction or drug-induced cholestasis) and somewhat later in hepatocellular disease (acute hepatitis). Itching also occurs in chronic liver diseases, typically the cholestatic forms such as primary biliary cirrhosis and sclerosing cholangitis where it is often the presenting symptom, occurring before the onset of jaundice. However, itching can occur in any liver disease, particularly once cirrhosis is present.

Jaundice is the hallmark symptom of liver disease and perhaps the most reliable marker of severity. Patients usually report darkening of the urine before they notice scleral icterus. Jaundice is rarely detectable with a bilirubin level <43 μmol/L (2.5 mg/dL). With severe cholestasis there will also be lightening of the color of the stools and steatorrhea. Jaundice without dark urine usually indicates indirect (unconjugated) hyperbilirubinemia and is typical of hemolytic anemia and the genetic disorders of bilirubin conjugation, the common and benign form being Gilbert's syndrome and the rare and severe form being Crigler-Najjar syndrome. Gilbert's syndrome affects up to 5% of the population; the

jaundice is more noticeable after fasting and with stress.

Major risk factors for liver disease that should be sought in the clinical history include details of alcohol use, medications (including herbal compounds, birth control pills, and over-the-counter medications), personal habits, sexual activity, travel, exposure to jaundiced or other high-risk persons, injection drug use, recent surgery, remote or recent transfusion with blood and blood products, occupation, accidental exposure to blood or needlestick, and familial history of liver disease.

For assessing the risk of viral hepatitis, a careful history of sexual activity is of particular importance and should include the number of lifetime sexual partners and, for men, a history of having sex with men. Sexual exposure is a common mode of spread of hepatitis B but is rare for hepatitis C. A family history of hepatitis, liver disease, and liver cancer is also important. Maternal-infant transmission occurs with both hepatitis B and C. Vertical spread of hepatitis B can now be prevented by passive and active immunization of the infant at birth. Vertical spread of hepatitis C is uncommon, but there are no reliable means of prevention. Transmission is more common in HIV-co-infected mothers and is also linked to prolonged and difficult labor and delivery, early rupture of membranes, and internal fetal monitoring. A history of injection drug use, even in the remote past, is of great importance in assessing the risk for hepatitis B and C. Injection drug use is now the single most common risk factor for hepatitis C. Transfusion with blood or blood products is no longer an important risk factor for acute viral hepatitis. However, blood transfusions received before the introduction of sensitive enzyme immunoassays for antibody to hepatitis C virus (anti-HCV) in 1992 is an important risk factor for chronic hepatitis C. Blood transfusion before 1986, when screening for antibody to hepatitis B core antigen (anti-HBc) was introduced, is also a risk factor for hepatitis B. Travel to an underdeveloped area of the world, exposure to persons with jaundice, and exposure to young children in day-care centers are risk factors for hepatitis A. Hepatitis E is is one of the more common causes of jaundice in Asia and Africa but is uncommon in developed nations, although mild cases have been associated with eating raw or undercooked pork or game (deer and wild boars). Tattooing and body piercing (for hepatitis B and C) and eating shellfish (for hepatitis A) are frequently mentioned but are actually quite rare types of exposure for acquiring hepatitis.

A history of alcohol intake is important in assessing the cause of liver disease and also in planning management and recommendations. In the United States, for example, at least 70% of adults drink alcohol to some degree, but significant alcohol intake is less common; in population-based surveys, only 5% have more than two drinks per day, the average drink representing 11–15 g alcohol. Alcohol consumption associated with an increased rate of alcoholic liver disease is probably more than two drinks (22–30 g) per day in women and three drinks (33–45 g) in men. Most patients with alcoholic cirrhosis have a much higher daily intake and have drunk excessively for ≥10 years before onset of liver disease. In assessing alcohol intake, the history should also focus on whether alcohol abuse or dependence is present. Alcoholism is usually defined by the behavioral patterns and consequences of alcohol intake, not on the basis of the amount of alcohol intake. *Abuse* is defined by a repetitive pattern of drinking alcohol that has adverse effects on social, family, occupational, or health status. *Dependence* is defined by alcohol-seeking behavior, despite its adverse effects. Many alcoholics demonstrate both dependence and abuse, and dependence is considered the more serious and advanced form of alcoholism. A clinically helpful approach to diagnosis of alcohol dependence and abuse is the use of the CAGE questionnaire **(Table 301-2)**, which is recommended in all medical history-taking.

Table 301-2 Cage Questions*

Acronym	Question
C	Have you ever felt you ought to *C* ut down on your drinking?
A	Have people *A*nnoyed you by criticizing your drinking?
G	Have you ever felt *G*uilty or bad about your drinking?
E	Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (*E*yeopener)?

*One "yes" response should raise suspicion of an alcohol use problem, and more than one is a strong indication that abuse or dependence exists.

Family history can be helpful in assessing liver disease. Familial causes of liver disease include Wilson's disease; hemochromatosis and α1 antitrypsin (α1AT) deficiency; and the more uncommon inherited pediatric liver diseases of familial intrahepatic cholestasis, benign recurrent intrahepatic cholestasis, and Alagille syndrome. Onset of severe liver disease in childhood or adolescence with a family history of liver disease or neuropsychiatric disturbance should lead to investigation for Wilson's disease. A family history of cirrhosis, diabetes, or endocrine failure and the appearance of liver

disease in adulthood should suggest hemochromatosis and lead to investigation of iron status. Adult patients with abnormal iron studies warrant genotyping of the *HFE* gene for the C282Y and H63D mutations typical of genetic hemochromatosis. In children and adolescents with iron overload, other non-*HFE* causes of hemochromatosis should be sought. A family history of emphysema should provoke investigation of α1AT levels and, if low, for Pi genotype.

Physical Examination

The physical examination rarely demonstrates evidence of liver dysfunction in a patient without symptoms or laboratory findings, nor are most signs of liver disease specific to one diagnosis. Thus, the physical examination complements rather than replaces the need for other diagnostic approaches. In many patients, the physical examination is normal unless the disease is acute or severe and advanced. Nevertheless, the physical examination is important in that it can be the first evidence for the presence of hepatic failure, portal hypertension, and liver decompensation. In addition, the physical examination can reveal signs that point to a specific diagnosis, either in risk factors or in associated diseases or findings.

Typical physical findings in liver disease are icterus, hepatomegaly, hepatic tenderness, splenomegaly, spider angiomata, palmar erythema, and excoriations. Signs of advanced disease include muscle wasting, ascites, edema, dilated abdominal veins, hepatic fetor, asterixis, mental confusion, stupor, and coma. In males with cirrhosis, particularly when related to alcohol, signs of hyperestrogenemia such as gynecomastia, testicular atrophy, and loss of male-pattern hair distribution may be found.

Icterus is best appreciated by inspecting the sclera under natural light. In fair-skinned individuals, a yellow color of the skin may be obvious. In dark-skinned individuals, the mucous membranes below the tongue can demonstrate jaundice. Jaundice is rarely detectable if the serum bilirubin level is <43 μmol/L (2.5 mg/dL) but may remain detectable below this level during recovery from jaundice (because of protein and tissue binding of conjugated bilirubin).

Spider angiomata and palmar erythema occur in both acute and chronic liver disease and may be especially prominent in persons with cirrhosis, but they can occur in normal individuals and are frequently present during pregnancy. Spider angiomata are superficial, tortuous arterioles and, unlike simple telangiectases, typically fill from the center outward. Spider angiomata occur only on the arms, face, and upper torso; they can be pulsatile and may be difficult to detect in dark-skinned individuals.

Hepatomegaly is not a very reliable sign of liver disease, because of the variability of the size and shape of the liver and the physical impediments to assessing liver size by percussion and palpation. Marked hepatomegaly is typical of cirrhosis, venoocclusive disease, infiltrative disorders such as amyloidosis, metastatic or primary cancers of the liver, and alcoholic hepatitis. Careful assessment of the liver edge may also demonstrate unusual firmness, irregularity of the surface, or frank nodules. Perhaps the most reliable physical finding in examining the liver is hepatic tenderness. Discomfort on touching or pressing on the liver should be carefully sought with percussive comparison of the right and left upper quadrants.

Splenomegaly occurs in many medical conditions but can be a subtle but significant physical finding in liver disease. The availability of ultrasound (US) assessment of the spleen allows for confirmation of the physical finding.

Signs of advanced liver disease include muscle-wasting and weight loss as well as hepatomegaly, bruising, ascites, and edema. Ascites is best appreciated by attempts to detect shifting dullness by careful percussion. US examination will confirm the finding of ascites in equivocal cases. Peripheral edema can occur with or without ascites. In patients with advanced liver disease, other factors frequently contribute to edema formation, including hypoalbuminemia, venous insufficiency, heart failure, and medications.

Hepatic failure is defined as the occurrence of signs or symptoms of hepatic encephalopathy in a person with severe acute or chronic liver disease. The first signs of hepatic encephalopathy can be subtle and nonspecific—change in sleep patterns, change in personality, irritability, and mental dullness. Thereafter, confusion, disorientation, stupor, and eventually coma supervene. In acute liver failure, excitability and mania may be present. Physical findings include asterixis and flapping tremors of the body and tongue. *Fetor hepaticus* refers to the slightly sweet, ammoniacal odor that can occur in patients with liver failure, particularly if there is portal-venous shunting of blood around the liver. Other causes of coma and disorientation should be excluded, mainly electrolyte imbalances, sedative use, and renal or respiratory failure. The appearance of hepatic encephalopathy during acute hepatitis is the major criterion for diagnosis of fulminant hepatitis and indicates a poor prognosis. In chronic liver disease, encephalopathy is usually triggered by a medical complication such as gastrointestinal bleeding, over-diuresis, uremia, dehydration, electrolyte imbalance, infection, constipation, or use of narcotic analgesics.

A helpful measure of hepatic encephalopathy is a careful mental status examination and use of the trail-making test, which consists of a series of 25 numbered circles that the patient is asked to connect as rapidly as possible using a pencil. The normal range for the connect-the-dot test is 15–30 seconds; it is considerably delayed in patients with early hepatic encephalopathy. Other tests include drawing

abstract objects or comparison of a signature to previous examples. More sophisticated testing such as with electroencephalography and visual evoked potentials can detect mild forms of encephalopathy, but are rarely clinically useful.

Other signs of advanced liver disease include umbilical hernia from ascites, hydrothorax, prominent veins over the abdomen, and *caput medusa*, which consists of collateral veins seen radiating from the umbilicus and resulting from the recanulation of the umbilical vein. Widened pulse pressure and signs of a hyperdynamic circulation can occur in patients with cirrhosis as a result of fluid and sodium retention, increased cardiac output, and reduced peripheral resistance. Patients with long-standing cirrhosis and portal hypertension are prone to develop the hepatopulmonary syndrome, defined by the triad of liver disease, hypoxemia, and pulmonary arteriovenous shunting. The hepatopulmonary syndrome is characterized by platypnea and orthodeoxia, representing shortness of breath and oxygen desaturation that occur paradoxically upon assuming an upright position. Measurement of oxygen saturation by pulse oximetry is a reliable screening test for the presence of hepatopulmonary syndrome.

Several skin disorders and changes occur commonly in liver disease. Hyperpigmentation is typical of advanced chronic cholestatic diseases such as primary biliary cirrhosis and sclerosing cholangitis. In these same conditions, xanthelasma and tendon xanthomata occur as a result of retention and high serum levels of lipids and cholesterol. A slate-gray pigmentation to the skin also occurs with hemochromatosis if iron levels are high for a prolonged period. Mucocutaneous vasculitis with palpable purpura, especially on the lower extremities, is typical of cryoglobulinemia of chronic hepatitis C but can also occur in chronic hepatitis B.

Some physical signs point to specific liver diseases. Kayser-Fleischer rings occur in Wilson's disease and consist of a golden-brown copper pigment deposited in Descemet's membrane at the periphery of the cornea; they are best seen by slit-lamp examination. Dupuytren contracture and parotid enlargement are suggestive of chronic alcoholism and alcoholic liver disease. In metastatic liver disease or primary hepatocellular carcinoma, signs of cachexia and wasting may be prominent, as well as firm hepatomegaly and a hepatic bruit.

Laboratory Testing

Diagnosis in liver disease is greatly aided by the availability of reliable and sensitive tests of liver injury and function. A typical battery of blood tests used for initial assessment of liver disease includes measuring levels of serum alanine and aspartate aminotransferases (ALT and AST), alkaline phosphatase (AlkP), direct and total serum bilirubin, and albumin and assessing prothrombin time. The pattern of abnormalities generally points to hepatocellular versus cholestatic liver disease and will help to decide whether the disease is acute or chronic and whether cirrhosis and hepatic failure are present. Based on these results, further testing over time may be necessary. Other laboratory tests may be helpful, such as γ-glutamyl transpeptidase (gGT) to define whether alkaline phosphatase elevations are due to liver disease; hepatitis serology to define the type of viral hepatitis; and autoimmune markers to diagnose primary biliary cirrhosis (antimitochondrial antibody; AMA), sclerosing cholangitis (peripheral antineutrophil cytoplasmic antibody; P-ANCA), and autoimmune hepatitis (antinuclear, smooth-muscle, and liver-kidney microsomal antibody). A simple delineation of laboratory abnormalities and common liver diseases is given in **Table 301-3**.

Table 301-3 Important Diagnostic Tests in Common Liver Diseases

Disease	Diagnostic Test
Hepatitis A	Anti-HAV IgM
Hepatitis B	
Acute	HBsAg and anti-HBc IgM
Chronic	HBsAg and HBeAg and/or HBV DNA
Hepatitis C	Anti-HCV and HCV RNA
Hepatitis D (delta)	HBsAg and anti-HDV
Hepatitis E	Anti-HEV
Autoimmune hepatitis	ANA or SMA, elevated IgG levels, and compatible histology
Primary biliary cirrhosis	Mitochondrial antibody, elevated IgM levels, and compatible histology
Primary sclerosing cholangitis	P-ANCA, cholangiography
Drug-induced liver	History of drug ingestion

disease	
Alcoholic liver disease	History of excessive alcohol intake and compatible histology
Nonalcoholic steatohepatitis	Ultrasound or CT evidence of fatty liver and compatible histology
α1 Antitrypsin disease	Reduced α1 antitrypsin levels, phenotypes PiZZ or PiSZ
Wilson's disease	Decreased serum ceruloplasmin and increased urinary copper; increased hepatic copper level
Hemochromatosis	Elevated iron saturation and serum ferritin; genetic testing for *HFE* gene mutations
Hepatocellular cancer	Elevated α-fetoprotein level >500; US or CT image of mass

Abbreviations: HAV, HBV, HCV, HDV, HEV: hepatitis A, B, C, D, or E virus; HBsAg, hepatitis B surface antigen; anti-HBc, antibody to hepatitis B core (antigen); HBeAg, hepatitis e antigen; ANA, antinuclear antibodies; SMA, smooth-muscle antibody; P-ANCA, peripheral antineutrophil cytoplasmic antibody.

The use and interpretation of liver function tests is summarized in Chap. 302.

Diagnostic Imaging

There have been great advances made in hepatic imaging, although no method is suitably accurate in demonstrating underlying cirrhosis. There are many modalities available for imaging the liver. US, CT, and MRI are the most commonly employed and are complementary to each other. In general, US and CT have a high sensitivity for detecting biliary duct dilatation and are the first-line options for investigating the patient with suspected obstructive jaundice. All three modalities can detect a fatty liver, which appears bright on imaging studies. Modifications of CT and MRI can be used to quantify liver fat, which may ultimately be valuable in monitoring therapy in patients with fatty liver disease. Magnetic resonance cholangiopancreatography (MRCP) and endoscopic retrograde cholangiopancreatography (ERCP) are the procedures of choice for visualization of the biliary tree. MRCP offers several advantages over ERCP; there is no need for contrast media or ionizing radiation, images can be acquired faster, it is less operator dependent, and it carries no risk of pancreatitis. MRCP is superior to US and CT for detecting choledocholithiasis but less specific. It is useful in the diagnosis of bile duct obstruction and congenital biliary abnormalities, but ERCP is more valuable in evaluating ampullary lesions and primary sclerosing cholangitis. ERCP allows for biopsy, direct visualization of the ampulla and common bile duct, and intraductal ultrasonography. It also provides several therapeutic options in patients with obstructive jaundice such as sphincterotomy, stone extraction, and placement of nasobiliary catheters and biliary stents. Doppler US and MRI are used to assess hepatic vasculature and hemodynamics and to monitor surgically or radiologically placed vascular shunts such as transjugular intrahepatic portosystemic shunts. CT and MRI are indicated for the identification and evaluation of hepatic masses, staging of liver tumors, and preoperative assessment. With regard to mass lesions, sensitivity of hepatic imaging continues to increase; unfortunately, specificity remains a problem, and often two and sometimes three studies are needed before a diagnosis can be reached. Recently, methods using elastography have been developed to measure hepatic stiffness as a means of assessing hepatic fibrosis. US and MR elastrography are now undergoing evaluation for their ability to detect different degrees of hepatic fibrosis and to obviate the need for liver biopsy in assessing disease stage. If found to be reliable, hepatic elastography may be an appropriate means of monitoring fibrosis and disease progression. Finally, interventional radiologic techniques allow the biopsy of solitary lesions, performance of radiofrequency ablation and chemoembolization of cancerous lesions, insertion of drains into hepatic abscesses, measurement of portal pressure, and creation of vascular shunts in patients with portal hypertension. Which modality to use depends on factors such as availability, cost, and experience of the radiologist with each technique.

LIVER BIOPSY

Liver biopsy remains the criterion standard in the evaluation of patients with liver disease, particularly in patients with chronic liver diseases. In selected instances, liver biopsy is necessary for diagnosis but is more often useful in assessing the severity (grade) and stage of liver damage, in predicting prognosis, and in monitoring response to treatment. The size of the liver biopsy is an important determinant of its reliability; a length of 1.5–2 cm being necessary for accurate assessment of fibrosis. In the future, noninvasive means of assessing disease activity (batteries of blood tests) and fibrosis (elastography and fibrosis markers) may replace liver biopsy in assessing stage and grade of disease.

Diagnosis of Liver Disease

The major causes of liver disease and key diagnostic features are outlined in Table 301-3, and an algorithm for evaluation of the patient with suspected liver disease is given in **Fig. 301-1**. Specifics of

diagnosis are discussed in later chapters. The most common causes of acute liver disease are viral hepatitis (particularly hepatitis A, B, and C), drug-induced liver injury, cholangitis, and alcoholic liver disease. Liver biopsy is usually not needed in the diagnosis and management of acute liver disease, exceptions being situations where the diagnosis remains unclear despite thorough clinical and laboratory investigation. Liver biopsy can be helpful in the diagnosis of drug-induced liver disease and in establishing the diagnosis of acute alcoholic hepatitis.

Figure 301-1

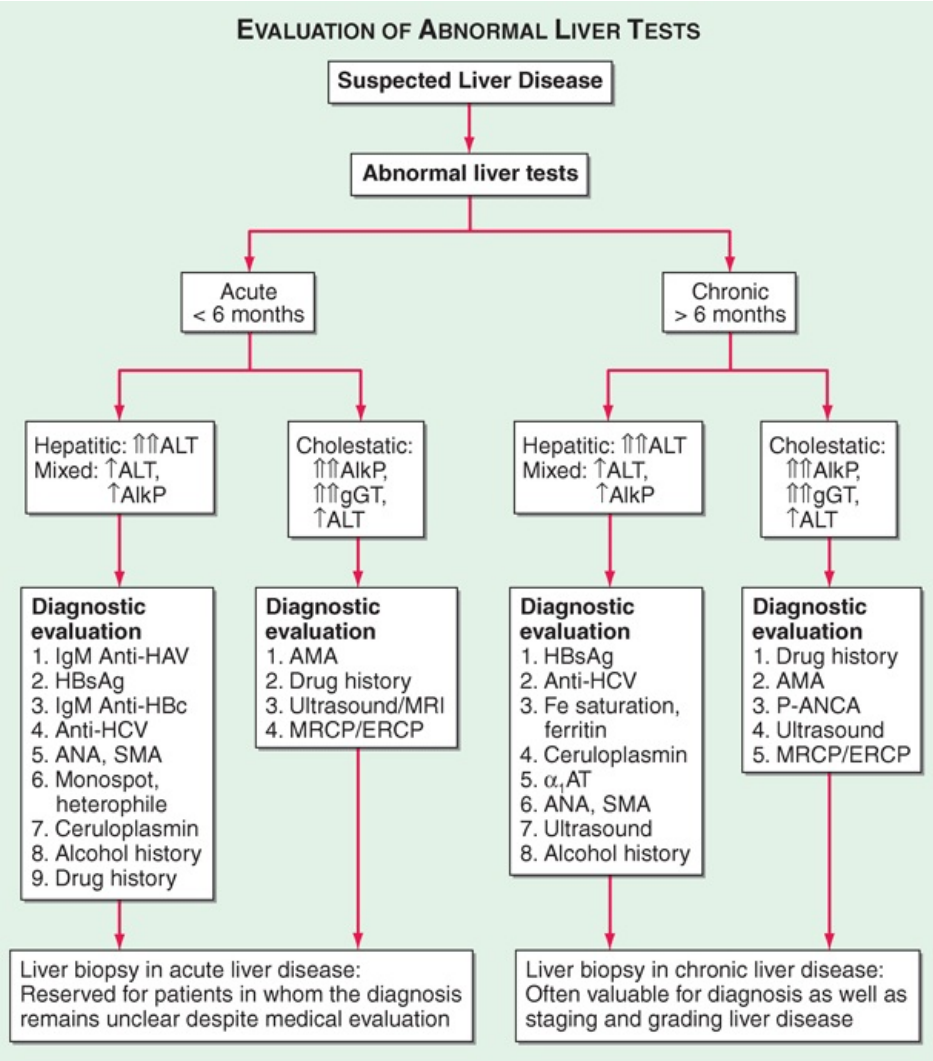

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for evaluation of abnormal liver tests. For patients with suspected liver disease, an appropriate approach to evaluation is initial testing for routine liver tests such as bilirubin, albumin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), and alkaline phosphatase (AlkP). These results (sometimes complemented by testing of γ-glutamyl transpeptidase; gGT) will establish whether the pattern of abnormalities is hepatic, cholestatic, or mixed. In addition, the duration of symptoms or abnormalities will show whether the disease is acute or chronic. If the disease is acute and if history, laboratory tests, and imaging studies do not reveal a diagnosis, liver biopsy is appropriate to help establish the diagnosis. If the disease is chronic, liver biopsy can be helpful not only for diagnosis but also to grade the activity and stage the progression of disease. This approach is largely applicable to patients without immune deficiency. In patients with HIV infection or after bone marrow or solid organ transplantation, diagnostic evaluation should also include evaluation of opportunistic infections (adenovirus, cytomegalovirus, coccidioidomycosis, etc.) as well as vascular and immunologic conditions (venoocclusive disease, graft-versus-host disease). HAV, HCV: hepatitis A or C virus; HBsAg, hepatitis B surface antigen; anti-HBc, antibody to hepatitis B core (antigen); ANA, antinuclear antibodies; SMA, smooth-muscle antibody; MRCP, magnetic resonance cholangiopancreatography; ERCP, endoscopic retrograde cholangiopancreatography; α1AT, α1 antitrypsin; AMA; antimitochondrial antibody; P-ANCA, peripheral antineutrophil cytoplasmic antibody.

The most common causes of chronic liver disease in general order of frequency are chronic hepatitis C, alcoholic liver disease, nonalcoholic steatohepatitis, chronic hepatitis B, autoimmune hepatitis, sclerosing cholangitis, primary biliary cirrhosis, hemochromatosis, and Wilson's disease. Strict

diagnostic criteria have not been developed for most liver diseases, but liver biopsy plays an important role in the diagnosis of autoimmune hepatitis, primary biliary cirrhosis, nonalcoholic and alcoholic steatohepatitis, and Wilson's disease (with a quantitative hepatic copper level).

Grading and Staging of Liver Disease

Grading refers to an assessment of the severity or activity of liver disease, whether acute or chronic; active or inactive; and mild, moderate, or severe. Liver biopsy is the most accurate means of assessing severity, particularly in chronic liver disease. Serum aminotransferase levels are used as convenient and noninvasive means to follow disease activity, but aminotransferase levels are not always reliable in reflecting disease severity. Thus, normal serum aminotransferase levels in patients with hepatitis B surface antigen (HBsAg) in serum may indicate the inactive HBsAg carrier state or may reflect mild chronic hepatitis B or hepatitis B with fluctuating disease activity. Serum testing for hepatitis B e antigen and hepatitis B virus DNA can help resolve these different patterns, but these markers can also fluctuate and change over time. Similarly, in chronic hepatitis C, serum aminotransferase levels can be normal despite moderate activity of disease. Finally, in both alcoholic and nonalcoholic steatohepatitis, aminotransferase levels are quite unreliable in reflecting severity. In these conditions, liver biopsy is helpful in guiding management and recommending therapy, particularly if therapy is difficult, prolonged, and expensive as is often the case in chronic viral hepatitis. There are several well-verified numerical scales for grading activity in chronic liver disease, the most common being the histology activity index and the Ishak histology scale.

Liver biopsy is also the most accurate means of assessing stage of disease as early or advanced, precirrhotic, and cirrhotic. Staging of disease pertains largely to chronic liver diseases in which progression to cirrhosis and end-stage liver disease can occur, but which may require years or decades to develop. Clinical features, biochemical tests, and hepatic imaging studies are helpful in assessing stage but generally become abnormal only in the middle to late stages of cirrhosis. Noninvasive tests that suggest advanced fibrosis include mild elevations of bilirubin, prolongation of prothrombin time, slight decreases in serum albumin, and mild thrombocytopenia (which is often the first indication of worsening fibrosis). Combinations of blood test results have been used to create models for predicting advanced liver disease, but these are not reliable enough to use on a regular basis and they only separate advanced from early disease. Recently, elastography and noninvasive breath tests using 13C-labeled compounds have been proposed as a means of detecting early stages of fibrosis and liver dysfunction, but their reliability and reproducibility remain to be proven. Thus, at present, mild to moderate stages of hepatic fibrosis are detectable only by liver biopsy. In assessing stage, the degree of fibrosis is usually used as its quantitative measure. The amount of fibrosis is generally staged on a 0 to 4+ (Metavir scale) or 0 to 6+ scale (Ishak scale). The importance of staging relates primarily to prognosis and to guiding management of complications. Patients with cirrhosis are candidates for screening and surveillance for esophageal varices and hepatocellular carcinoma. Patients without advanced fibrosis need not undergo screening.

Cirrhosis can also be staged clinically. A reliable staging system is the modified Child-Pugh classification with a scoring system of 5–15: scores of 5 and 6 being Child-Pugh class A (consistent with "compensated cirrhosis"), scores of 7–9 indicating class B, and 10–15 indicating class C **(Table 301-4)**. This scoring system was initially devised to stratify patients into risk groups prior to undergoing portal decompressive surgery. The Child-Pugh score is a reasonably reliable predictor of survival in many liver diseases and predicts the likelihood of major complications of cirrhosis such as bleeding from varices and spontaneous bacterial peritonitis. It was used to assess prognosis in cirrhosis and to provide the standard criteria for listing liver transplantation (Child-Pugh class B). Recently the Child-Pugh system has been replaced by the model for end-stage liver disease (MELD) score for assessing the need for liver transplantation. The MELD score is a prospectively derived scoring system designed to predict prognosis of patients with liver disease and portal hypertension. It is calculated using three noninvasive variables—the prothrombin time expressed as international normalized ratio (INR), serum bilirubin, and serum creatinine (http://www.unos.org/resources/meldPeldCalculator.asp).

Table 301-4 Child-Pugh Classification of Cirrhosis

Factor	Units	1	2	3
Serum bilirubin	μmol/L	<34	34-51	>51
	mg/dL	<2.0	2.0-3.0	>3.0
Serum albumin	g/L	>35	30-35	<30
	g/dL	>3.5	3.0-3.5	<3.0
Prothrombin time	seconds	0-4	4-6	>6
	prolonged	<1.7	1.7-2.3	>2.3
	INR			

Ascites		None	Easily controlled	Poorly controlled
Hepatic encephalopathy		None	Minimal	Advanced

Note: The Child-Pugh score is calculated by adding the scores of the five factors and can range from 5 to 15. Child-Pugh class can be A (a score of 5-6), B (7-9), or C (10 or above). Decompensation indicates cirrhosis with a Child-Pugh score of >7 (class B). This level has been the accepted criterion for listing liver transplantation.

MELD provides a more objective means of assessing disease severity and has less center-to-center variation than the Child-Pugh score and has a wider range of values. MELD is currently used to establish priority listing for liver transplantation in the United States. A similar system using bilirubin, INR, serum albumin, age, and nutritional status is used for children below the age of 12 years [pediatric end-stage liver disease (PELD)].

Thus, liver biopsy is helpful not only in diagnosis but also in management of chronic liver disease and assessment of prognosis. Because liver biopsy is an invasive procedure and not without complications, it should be used only when it will contribute materially to management and therapeutic decisions.

NONSPECIFIC ISSUES IN MANAGEMENT OF PATIENTS WITH LIVER DISEASE

Specifics on management of different forms of acute or chronic liver disease are given in subsequent chapters, but certain issues are applicable to any patient with liver disease. These include advice regarding alcohol use, medications, vaccination, and surveillance for complications of liver disease. Alcohol should be used sparingly, if at all, by patients with liver disease. Abstinence from alcohol should be encouraged for all patients with alcohol-related liver disease and in patients with cirrhosis and those receiving interferon-based therapy for hepatitis B or C. Regarding vaccinations, all patients with liver disease should receive hepatitis A vaccine and those with risk factors should receive hepatitis B vaccination as well. Influenza and pneumococcal vaccination should also be encouraged. Patients with liver disease should be careful using any medications, other than the most necessary. Drug-induced hepatotoxicity can mimic many forms of liver disease and can cause exacerbations of chronic hepatitis and cirrhosis; drugs should be suspected in any situation in which the cause of exacerbation is unknown. Finally, consideration should be given to surveillance for complications of chronic liver disease such as variceal hemorrhage and hepatocellular carcinoma. Patients with cirrhosis warrant upper endoscopy to assess the presence of varices and should be given chronic therapy with beta blockers or offered endoscopic obliteration if large varices are found. Patients with cirrhosis also warrant screening and long-term surveillance for development of hepatocellular carcinoma. While the optimal regimen for such surveillance has not been established, an appropriate approach is US of the liver at 6- to 12-month intervals.

FURTHER READINGS

Boyer TD et al (eds): *Zakim and Boyer's Hepatology: A Textbook of Liver Disease* , 6th ed. Philadelphia, Saunders, 2011

Castera L: Transient elastography and other noninvasive tests to assess hepatic fibrosis in patients with viral hepatitis. J Viral Hepat 16:300, 2009[PMID: 19254351] [Full Text]

Freeman RB JR: Model for end-stage liver disease (MELD) for liver allocation: A 5-year score card. Hepatology 47:1052, 2008[PMID: 18161047] [Full Text]

Hamer OW et al: Technology insight: Advances in liver imaging. Nature Gastroenterol Hepatol 4:216: 2007

Kaplowitz N, Deleve LD (eds): *Drug-Induced Liver Disease* . New York, Informa Healthcare, 2007

Rockey DC et al: Liver biopsy. Hepatology 49: 1017, 2009[PMID: 19243014] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

EVALUATION OF LIVER FUNCTION: INTRODUCTION

Several biochemical tests are useful in the evaluation and management of patients with hepatic dysfunction. These tests can be used to (1) detect the presence of liver disease, (2) distinguish among different types of liver disorders, (3) gauge the extent of known liver damage, and (4) follow the response to treatment.

Liver tests have shortcomings. They can be normal in patients with serious liver disease and abnormal in patients with diseases that do not affect the liver. Liver tests rarely suggest a specific diagnosis; rather, they suggest a general category of liver disease, such as hepatocellular or cholestatic, which then further directs the evaluation.

The liver carries out thousands of biochemical functions, most of which cannot be easily measured by blood tests. Laboratory tests measure only a limited number of these functions. In fact, many tests, such as the aminotransferases or alkaline phosphatase, do not measure liver function at all. Rather, they detect liver cell damage or interference with bile flow. Thus, no one test enables the clinician to accurately assess the liver's total functional capacity.

To increase both the sensitivity and the specificity of laboratory tests in the detection of liver disease, it is best to use them as a battery. Tests usually employed in clinical practice include the bilirubin, aminotransferases, alkaline phosphatase, albumin, and prothrombin time tests. When more than one of these tests provide abnormal findings or the findings are persistently abnormal on serial determinations, the probability of liver disease is high. When all test results are normal, the probability of missing occult liver disease is low.

When evaluating patients with liver disorders, it is helpful to group these tests into general categories. The classification we have found most useful is given below.

Tests Based on Detoxification and Excretory Functions

SERUM BILIRUBIN

(See also Chap. 42) Bilirubin, a breakdown product of the porphyrin ring of heme-containing proteins, is found in the blood in two fractions—conjugated and unconjugated. The unconjugated fraction, also termed the *indirect fraction*, is insoluble in water and is bound to albumin in the blood. The conjugated (direct) bilirubin fraction is water soluble and can therefore be excreted by the kidney. When measured by modifications of the original van den Bergh method, normal values of total serum bilirubin are reported between 1 and 1.5 mg/dL with 95% of a normal population falling between 0.2 and 0.9 mg/dL. If the direct-acting fraction is less than 15% of the total, the bilirubin can be considered to all be indirect. The most frequently reported upper limit of normal for conjugated bilirubin is 0.3 mg/dL.

Elevation of the unconjugated fraction of bilirubin is rarely due to liver disease. An isolated elevation of unconjugated bilirubin is seen primarily in hemolytic disorders and in a number of genetic conditions such as Crigler-Najjar and Gilbert's syndromes **(Chap. 42)**. Isolated unconjugated hyperbilirubinemia (bilirubin elevated but <15% direct) should prompt a workup for hemolysis **(Fig. 302-1)**. In the absence of hemolysis, an isolated, unconjugated hyperbilirubinemia in an otherwise healthy patient can be attributed to Gilbert's syndrome, and no further evaluation is required.

Figure 302-1

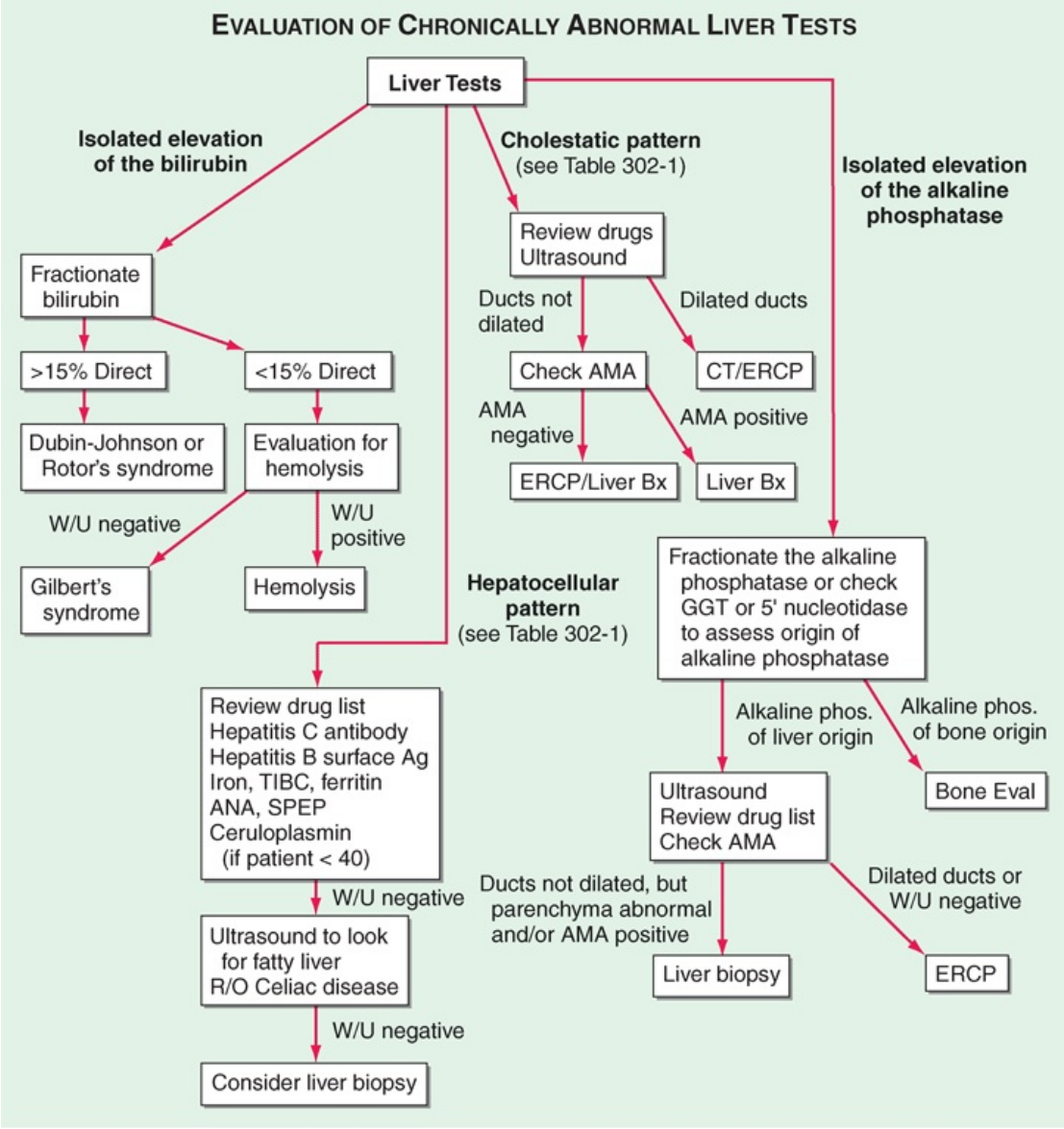

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

is an algorithm for the evaluation of chronically abnormal liver tests. ERCP, endoscopic retrograde cholangiopancreatography; CT, computed tomography; AMA, antimitochondrial antibody; ANA, antinuclear antibody; SPEP, serum protein electrophoresis; TIBC, total iron-binding capacity; GGT, α glutamyl transpeptidase; W/U, work up.

In contrast, conjugated hyperbilirubinemia almost always implies liver or biliary tract disease. The rate-limiting step in bilirubin metabolism is not conjugation of bilirubin, but rather the transport of conjugated bilirubin into the bile canaliculi. Thus, elevation of the conjugated fraction may be seen in any type of liver disease. In most liver diseases, both conjugated and unconjugated fractions of the bilirubin tend to be elevated. Except in the presence of a purely unconjugated hyperbilirubinemia, fractionation of the bilirubin is rarely helpful in determining the cause of jaundice.

While the degree of elevation of the serum bilirubin has not been critically assessed as a prognostic marker, it important in a number of conditions. In viral hepatitis, the higher the serum bilirubin, the greater the hepatocellular damage. Total serum bilirubin correlates with poor outcomes in alcoholic hepatitis. It is also a critical component of the Model for Endstage Liver Disease (MELD) score, a tool used to estimate survival of patients with end-stage liver disease. An elevated total serum bilirubin in patients with drug-induced liver disease indicates more severe injury.

URINE BILIRUBIN

Unconjugated bilirubin always binds to albumin in the serum and is not filtered by the kidney. Therefore, any bilirubin found in the urine is conjugated bilirubin; the presence of bilirubinuria implies the presence of liver disease. A urine dipstick test can theoretically give the same information as fractionation of the serum bilirubin. This test is almost 100% accurate. Phenothiazines may give a false-positive reading with the Ictotest tablet. In patients recovering from jaundice, the urine bilirubin

clears prior to the serum bilirubin.

BLOOD AMMONIA

Ammonia is produced in the body during normal protein metabolism and by intestinal bacteria, primarily those in the colon. The liver plays a role in the detoxification of ammonia by converting it to urea, which is excreted by the kidneys. Striated muscle also plays a role in detoxification of ammonia, which is combined with glutamic acid to form glutamine. Patients with advanced liver disease typically have significant muscle wasting, which likely contributes to hyperammonemia in these patients. Some physicians use the blood ammonia for detecting encephalopathy or for monitoring hepatic synthetic function, although its use for either of these indications has problems. There is very poor correlation between either the presence or the severity of acute encephalopathy and elevation of blood ammonia; it can be occasionally useful for identifying occult liver disease in patients with mental status changes. There is also a poor correlation of the blood serum ammonia and hepatic function. The ammonia can be elevated in patients with severe portal hypertension and portal blood shunting around the liver even in the presence of normal or near-normal hepatic function. Elevated arterial ammonia levels have been shown to correlate with outcome in fulminant hepatic failure.

SERUM ENZYMES

The liver contains thousands of enzymes, some of which are also present in the serum in very low concentrations. These enzymes have no known function in the serum and behave like other serum proteins. They are distributed in the plasma and in interstitial fluid and have characteristic half-lives, which are usually measured in days. Very little is known about the catabolism of serum enzymes, although they are probably cleared by cells in the reticuloendothelial system. The elevation of a given enzyme activity in the serum is thought to primarily reflect its increased rate of entrance into serum from damaged liver cells.

Serum enzyme tests can be grouped into three categories: (1) enzymes whose elevation in serum reflects damage to hepatocytes, (2) enzymes whose elevation in serum reflects cholestasis, and (3) enzyme tests that do not fit precisely into either pattern.

Enzymes that Reflect Damage to Hepatocytes

The aminotransferases (transaminases) are sensitive indicators of liver cell injury and are most helpful in recognizing acute hepatocellular diseases such as hepatitis. They include the aspartate aminotransferase (AST) and the alanine aminotransferase (ALT). AST is found in the liver, cardiac muscle, skeletal muscle, kidneys, brain, pancreas, lungs, leukocytes, and erythrocytes in decreasing order of concentration. ALT is found primarily in the liver and is therefore a more specific indicator of liver injury. The aminotransferases are normally present in the serum in low concentrations. These enzymes are released into the blood in greater amounts when there is damage to the liver cell membrane resulting in increased permeability. Liver cell necrosis is not required for the release of the aminotransferases, and there is a poor correlation between the degree of liver cell damage and the level of the aminotransferases. Thus, the absolute elevation of the aminotransferases is of no prognostic significance in acute hepatocellular disorders.

The normal range for aminotransferases varies widely among laboratories, but generally ranges from 10-40 U/L. The inter-laboratory variation in normal range is due to technical reasons; no reference standards exist to establish upper limits of normal for ALT and AST. Some have recommended revisions of normal limits of the aminotransferases adjustments for sex and BMI, but others have noted the potential costs and unclear benefits of implementing this change.

Any type of liver cell injury can cause modest elevations in the serum aminotransferases. Levels of up to 300 U/L are nonspecific and may be found in any type of liver disorder. Minimal ALT elevations in asymptomatic blood donors rarely indicate severe liver disease; studies have shown that fatty liver disease is the most likely explanation. Striking elevations—i.e., aminotransferases >1000 U/L—occur almost exclusively in disorders associated with extensive hepatocellular injury such as (1) viral hepatitis, (2) ischemic liver injury (prolonged hypotension or acute heart failure), or (3) toxin- or drug-induced liver injury.

The pattern of the aminotransferase elevation can be helpful diagnostically. In most acute hepatocellular disorders, the ALT is higher than or equal to the AST. While the AST:ALT ratio is typically less than 1 in patients with chronic viral hepatitis and non-alcoholic fatty liver disease, a number of groups have noted that as cirrhosis develops this ratio rises to greater than 1. An AST:ALT ratio >2:1 is suggestive, while a ratio >3:1 is highly suggestive of alcoholic liver disease. The AST in alcoholic liver disease is rarely >300 U/L, and the ALT is often normal. A low level of ALT in the serum is due to an alcohol-induced deficiency of pyridoxal phosphate.

The aminotransferases are usually not greatly elevated in obstructive jaundice. One notable exception occurs during the acute phase of biliary obstruction caused by the passage of a gallstone into the common bile duct. In this setting, the aminotransferases can briefly be in the 1000–2000 U/L range. However, aminotransferase levels decrease quickly, and the liver-function tests rapidly evolve into one typical of cholestasis.

Enzymes that Reflect Cholestasis

The activities of three enzymes—alkaline phosphatase, 5'-nucleotidase, and γ-glutamyl transpeptidase (GGT)—are usually elevated in cholestasis. Alkaline phosphatase and 5'-nucleotidase are found in or near the bile canalicular membrane of hepatocytes, while GGT is located in the endoplasmic reticulum and in bile duct epithelial cells. Reflecting its more diffuse localization in the liver, GGT elevation in serum is less specific for cholestasis than are elevations of alkaline phosphatase or 5'-nucleotidase. Some have advocated the use of GGT to identify patients with occult alcohol use. Its lack of specificity makes its use in this setting questionable.

The normal serum alkaline phosphatase consists of many distinct isoenzymes found in the liver; bone; placenta; and, less commonly, small intestine. Patients over age 60 can have a mildly elevated alkaline phosphatase @@@(1–11/2 times normal), while individuals with blood types O and B can have an elevation of the serum alkaline phosphatase after eating a fatty meal due to the influx of intestinal alkaline phosphatase into the blood. It is also nonpathologically elevated in children and adolescents undergoing rapid bone growth, because of bone alkaline phosphatase, and late in normal pregnancies due to the influx of placental alkaline phosphatase.

Elevation of liver-derived alkaline phosphatase is not totally specific for cholestasis, and a less than threefold elevation can be seen in almost any type of liver disease. Alkaline phosphatase elevations greater than four times normal occur primarily in patients with cholestatic liver disorders, infiltrative liver diseases such as cancer and amyloidosis, and bone conditions characterized by rapid bone turnover (e.g., Paget's disease). In bone diseases, the elevation is due to increased amounts of the bone isoenzymes. In liver diseases, the elevation is almost always due to increased amounts of the liver isoenzyme.

If an elevated serum alkaline phosphatase is the only abnormal finding in an apparently healthy person, or if the degree of elevation is higher than expected in the clinical setting, identification of the source of elevated isoenzymes is helpful (Fig. 302-1). This problem can be approached in several ways. First, and most precise, is the fractionation of the alkaline phosphatase by electrophoresis. The second approach is based on the observation that alkaline phosphatases from individual tissues differ in susceptibility to inactivation by heat. The finding of an elevated serum alkaline phosphatase level in a patient with a heat-stable fraction strongly suggests that the placenta or a tumor is the source of the elevated enzyme in serum. Susceptibility to inactivation by heat increases, respectively, for the intestinal, liver, and bone alkaline phosphatases, bone being by far the most sensitive. The third, best substantiated, and most available approach involves the measurement of serum 5'-nucleotidase or GGT. These enzymes are rarely elevated in conditions other than liver disease.

In the absence of jaundice or elevated aminotransferases, an elevated alkaline phosphatase of liver origin often, but not always, suggests early cholestasis and, less often, hepatic infiltration by tumor or granulomata. Other conditions that cause isolated elevations of the alkaline phosphatase include Hodgkin's disease, diabetes, hyperthyroidism, congestive heart failure, amyloidosis, and inflammatory bowel disease.

The level of serum alkaline phosphatase elevation is not helpful in distinguishing between intrahepatic and extrahepatic cholestasis. There is essentially no difference among the values found in obstructive jaundice due to cancer, common duct stone, sclerosing cholangitis, or bile duct stricture. Values are similarly increased in patients with intrahepatic cholestasis due to drug-induced hepatitis; primary biliary cirrhosis; rejection of transplanted livers; and, rarely, alcohol-induced steatohepatitis. Values are also greatly elevated in hepatobiliary disorders seen in patients with AIDS (e.g., AIDS cholangiopathy due to cytomegalovirus or cryptosporidial infection and tuberculosis with hepatic involvement).

Tests that Measure Biosynthetic Function of the Liver

SERUM ALBUMIN

Serum albumin is synthesized exclusively by hepatocytes. Serum albumin has a long half-life: 18–20 days, with ~4% degraded per day. Because of this slow turnover, the serum albumin is not a good indicator of acute or mild hepatic dysfunction; only minimal changes in the serum albumin are seen in acute liver conditions such as viral hepatitis, drug-related hepatoxicity, and obstructive jaundice. In hepatitis, albumin levels <3 g/dL should raise the possibility of chronic liver disease. Hypoalbuminemia is more common in chronic liver disorders such as cirrhosis and usually reflects severe liver damage and decreased albumin synthesis. One exception is the patient with ascites in whom synthesis may be normal or even increased, but levels are low because of the increased volume of distribution. However, hypoalbuminemia is not specific for liver disease and may occur in protein malnutrition of any cause, as well as protein-losing enteropathies, nephrotic syndrome, and chronic infections that are associated with prolonged increases in levels of serum interleukin 1 and/or tumor necrosis factor, cytokines that inhibit albumin synthesis. Serum albumin should not be measured for screening in patients in whom there is no suspicion of liver disease. A general medical clinic study of consecutive patients in whom no indications were present for albumin measurement showed that while 12% of patients had abnormal test results, the finding was of clinical importance in only 0.4%.

SERUM GLOBULINS

Serum globulins are a group of proteins made up of γ globulins (immunoglobulins) produced by B lymphocytes and α and β globulins produced primarily in hepatocytes. γ Globulins are increased in chronic liver disease, such as chronic hepatitis and cirrhosis. In cirrhosis, the increased serum gamma globulin concentration is due to the increased synthesis of antibodies, some of which are directed against intestinal bacteria. This occurs because the cirrhotic liver fails to clear bacterial antigens that normally reach the liver through the hepatic circulation.

Increases in the concentration of specific isotypes of γ globulins are often helpful in the recognition of certain chronic liver diseases. Diffuse polyclonal increases in IgG levels are common in autoimmune hepatitis; increases >100% should alert the clinician to this possibility. Increases in the IgM levels are common in primary biliary cirrhosis, while increases in the IgA levels occur in alcoholic liver disease.

Coagulation Factors

With the exception of factor VIII, which is produced by vascular endothelial cells, the blood clotting factors are made exclusively in hepatocytes. Their serum half-lives are much shorter than albumin, ranging from 6 h for factor VII to 5 days for fibrinogen. Because of their rapid turnover, measurement of the clotting factors is the single best acute measure of hepatic synthetic function and helpful in both the diagnosis and assessing the prognosis of acute parenchymal liver disease. Useful for this purpose is the *serum prothrombin time*, which collectively measures factors II, V, VII, and X. Biosynthesis of factors II, VII, IX, and X depends on vitamin K. The international normalized ratio (INR) is used to express the degree of anticoagulation on warfarin therapy. The INR standardizes prothrombin time measurement according to the characteristics of the thromboplastin reagent used in a particular lab which is expressed as an International Sensitivity Index (ISI); the ISI is then used in calculating the INR. Because the ISI is validated only for patients on vitamin K antagonists, there has been concern regarding the validity of using it for patients with chronic liver disease.

The prothrombin time may be elevated in hepatitis and cirrhosis as well as in disorders that lead to vitamin K deficiency such as obstructive jaundice or fat malabsorption of any kind. Marked prolongation of the prothrombin time, >5 s above control and not corrected by parenteral vitamin K administration, is a poor prognostic sign in acute viral hepatitis and other acute and chronic liver diseases. The INR, along with the total serum bilirubin and creatinine, are components of the MELD score which is used to allocate organs for liver transplantation.

Other Diagnostic Tests

While tests may direct the physician to a category of liver disease, additional radiologic testing and procedures are often necessary to make the proper diagnosis, as shown in Fig. 302-1. The two most commonly used ancillary tests are reviewed here.

PERCUTANEOUS LIVER BIOPSY

Percutaneous biopsy of the liver is a safe procedure that can be easily performed at the bedside with local anesthesia and ultrasound guidance. Liver biopsy is of proven value in the following situations: (1) hepatocellular disease of uncertain cause, (2) prolonged hepatitis with the possibility of chronic active hepatitis, (3) unexplained hepatomegaly, (4) unexplained splenomegaly, (5) hepatic filling defects by radiologic imaging, (6) fever of unknown origin, (7) staging of malignant lymphoma. Liver biopsy is most accurate in disorders causing diffuse changes throughout the liver and is subject to sampling error in focal infiltrative disorders such as hepatic metastases. Liver biopsy should not be the initial procedure in the diagnosis of cholestasis. The biliary tree should first be assessed for signs of obstruction. Contraindications to performing a percutaneous liver biopsy include significant ascites and prolonged INR. Under these circumstances, the biopsy can be performed via the transjugular approach.

ULTRASONOGRAPHY

Ultrasonography is the first diagnostic test to use in patients whose liver tests suggest cholestasis, to look for the presence of a dilated intrahepatic or extrahepatic biliary tree or to identify gallstones. In addition, it shows space-occupying lesions within the liver, enables the clinician to distinguish between cystic and solid masses, and helps direct percutaneous biopsies. Ultrasound with Doppler imaging can detect the patency of the portal vein, hepatic artery, and hepatic veins and determine the direction of blood flow. This is the first test ordered in patients suspected of having Budd-Chiari syndrome.

Use of Liver Tests

As previously noted, the best way to increase the sensitivity and specificity of laboratory tests in the detection of liver disease is to employ a battery of tests that includes the aminotransferases, alkaline phosphatase, bilirubin, albumin, and prothrombin time along with the judicious use of the other tests described in this chapter. **Table 302-1** shows how patterns of liver tests can lead the clinician to a category of disease that will direct further evaluation. However, it is important to remember that no single set of liver tests will necessarily provide a diagnosis. It is often necessary to repeat these tests on several occasions over days to weeks for a diagnostic pattern to emerge.

Table 302-1 Liver Test Patterns in Hepatobiliary Disorders

Type of Disorder	Bilirubin	Aminotransferases	Alkaline Phosphatase	Albumin	Prothrombin Time
Hemolysis/Gilbert's syndrome	Normal to 86 μmol/L (5 mg/dL) 85% due to indirect fractions No bilirubinuria	Normal	Normal	Normal	Normal
Acute hepatocellular necrosis (viral and drug hepatitis, hepatotoxins, acute heart failure)	Both fractions may be elevated Peak usually follows aminotransferases Bilirubinuria	Elevated, often >500 IU ALT >AST	Normal to <3 times normal elevation	Normal	Usually normal. If >5X above control and not corrected by parenteral vitamin K, suggests poor prognosis
Chronic hepatocellular disorders	Both fractions may be elevated Bilirubinuria	Elevated, but usually <300 IU	Normal to <3 times normal elevation	Often decreased	Often prolonged Fails to correct with parenteral vitamin K
Alcoholic hepatitis Cirrhosis	Both fractions may be elevated Bilirubinuria	AST:ALT > 2 suggests alcoholic hepatitis or cirrhosis	Normal to <3 times normal elevation	Often decreased	Often prolonged Fails to correct with parenteral vitamin K
Intra- and extra-hepatic cholestasis (Obstructive jaundice)	Both fractions may be elevated Bilirubinuria	Normal to moderate elevation Rarely >500 IU	Elevated, often >4 times normal elevation	Normal, unless chronic	Normal If prolonged, will correct with parenteral vitamin K
Infiltrative diseases (tumor, granulomata); partial bile duct obstruction	Usually normal	Normal to slight elevation	Elevated, often >4 times normal elevation Fractionate, or confirm liver origin with 5' nucleotidase or γ glutamyl transpeptidase	Normal	Normal

FURTHER READINGS

Bhatia V et al: Predictive value of arterial ammonia for complications and outcome in acute liver failure. Gut 55:98, 2006[PMID: 16024550] [Full Text]

Bosma PJ et al: The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N Engl J Med 333:1171, 1995[PMID: 7565971] [Full Text]

Kim HC et al: Normal serum aminotransferase concentration and risk of mortality from liver diseases: Prospective cohort study. BMJ 328:983, 2004[PMID: 15028636] [Full Text]

Prati D et al: Updated definitions of healthy ranges for serum alanine aminotranferase levels. Ann Intern Med 137:1, 2002[PMID: 12093239] [Full Text]

Pratt DS, Kaplan MM: Evaluation of abnormal liver-enzyme tests in the asymptomatic patient. N Engl J Med 342:1266, 2000[PMID: 10781624] [Full Text]

———: Laboratory tests, in *Schiff's Diseases of the Liver* , 10th ed, ER Schiff et al (eds). Philadelphia, Lippincott Williams & Wilkins, 2006

Torezan-Filho MA et al: Clinical significance of elevated alanine aminotransferase in blood donors: A follow-up study. Liver Int 24:575, 2004[PMID: 15566507] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BILIRUBIN METABOLISM

The details of bilirubin metabolism are presented in **Chap. 42**. However, the hyperbilirubinemias are best understood in terms of perturbations of specific aspects of bilirubin metabolism and transport, and these will be briefly reviewed here as depicted in **Fig. 303-1**.

Figure 303-1

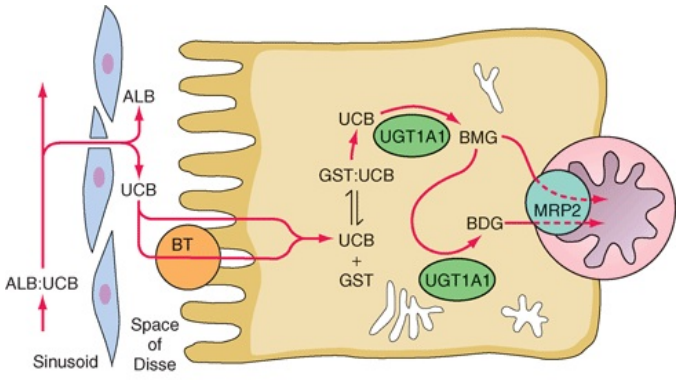

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hepatocellular bilirubin transport. Albumin-bound bilirubin in sinusoidal blood passes through endothelial cell fenestrae to reach the hepatocyte surface, entering the cell by both facilitated and simple diffusional processes. Within the cell it is bound to glutathione-S-transferases and conjugated by bilirubin-UDP-glucuronosyltransferase (UGT1A1) to monoand diglucuronides, which are actively transported across the canalicular membrane into the bile. ALB, albumin; BDG, bilirubin diglucuronide; BMG, bilirubin monoglucuronide; BT, proposed bilirubin transporter; GST, glutathione-S-transferase; MRP2, multidrug resistance–associated protein 2; UCB, unconjugated bilirubin; UGT1A1, bilirubin-UDP-glucuronosyltransferase.

Bilirubin is the end product of heme degradation. Some 70–90% of bilirubin is derived from degradation of the hemoglobin of senescent red blood cells. Bilirubin produced in the periphery is transported to the liver within the plasma, where, due to its insolubility in aqueous solutions, it is tightly bound to albumin. Under normal circumstances, bilirubin is removed from the circulation rapidly and efficiently by hepatocytes. Transfer of bilirubin from blood to bile involves four distinct but interrelated steps (Fig. 303-1).

1. *Hepatocellular uptake:* Uptake of bilirubin by the hepatocyte has carrier-mediated kinetics. Although a number of candidate bilirubin transporters have been proposed, the actual transporter remains elusive.
2. *Intracellular binding:* Within the hepatocyte, bilirubin is kept in solution by binding as a nonsubstrate ligand to several of the glutathione-S-transferases, formerly called ligandins.
3. *Conjugation:* Bilirubin is conjugated with one or two glucuronic acid moieties by a specific UDP-glucuronosyltransferase to form bilirubin mono- and diglucuronide, respectively. Conjugation disrupts the internal hydrogen bonding that limits aqueous solubility of bilirubin, and the resulting glucuronide conjugates are highly soluble in water. Conjugation is obligatory for excretion of bilirubin across the bile canalicular membrane into bile. The UDP-glucuronosyltransferases have been classified into gene families based on the degree of homology among the mRNAs for the various isoforms. Those that conjugate bilirubin and certain other substrates have been designated the *UGT1* family. These are expressed from a single gene complex by alternative promoter usage. This gene complex contains multiple substrate-specific first exons, designated A1, A2, etc. **(Fig. 303-2)**, each with its own promoter and each encoding the amino-terminal half of a specific isoform. In addition, there are four common exons (exons 2–5) that encode the shared carboxyl-terminal half of all of the *UGT1* isoforms. The various first exons encode the specific aglycone substrate binding sites for each isoform, while the shared exons encode the binding site for the sugar donor, UDP-glucuronic acid, and the transmembrane domain. Exon A1 and the four common exons, collectively designated the *UGT1A1* gene (Fig. 303-2), encode the physiologically critical enzyme bilirubin-UDP-glucuronosyltransferase (UGT1A1). A functional corollary of the organization of the *UGT1* gene is that a mutation in one of the first exons will affect only a single enzyme isoform. By contrast, a mutation in exons 2–5 will alter all isoforms encoded by the *UGT1* gene complex.
4. *Biliary excretion:* Bilirubin mono- and diglucuronides are excreted across the canalicular plasma membrane into the bile canaliculus by an ATP-dependent transport process mediated by a canalicular membrane protein called *multidrug resistance–associated protein 2* (MRP2). Mutations of MRP2 result in the Dubin-Johnson syndrome (see below).

Figure 303-2

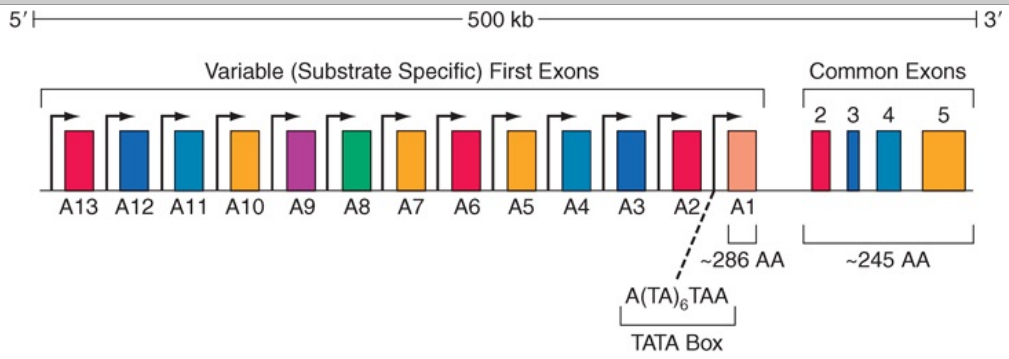

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Structural organization of the human UGT1 gene complex. This large complex on chromosome 2 contains at least 13 substrate-specific first exons (A1, A2, etc.). Since four of these are pseudogenes, nine UGT1 isoforms with differing substrate specificities are expressed. Each exon 1 has its own promoter and encodes the amino-terminal substrate-specific ~286 amino acids of the various UGT1 -encoded isoforms, and common exons 2–5 that encode the 245 carboxyl-terminal amino acids common to all of the isoforms. mRNAs for specific isoforms are assembled by splicing a particular first exon such as the bilirubin-specific exon A1 to exons 2 to 5. The resulting message encodes a complete enzyme, in this particular case bilirubin-UDP-glucuronosyltransferase (UGT1A1). Mutations in a first exon affect only a single isoform. Those in exons 2–5 affect all enzymes encoded by the UGT1 complex.

Extrahepatic Aspects of Bilirubin Disposition

BILIRUBIN IN THE GUT

Following secretion into bile, conjugated bilirubin reaches the duodenum and passes down the gastrointestinal tract without reabsorption by the intestinal mucosa. An appreciable fraction is converted by bacterial metabolism in the gut to the water-soluble colorless compound urobilinogen. Urobilinogen undergoes enterohepatic cycling. Urobilinogen not taken up by the liver reaches the systemic circulation, from which some is cleared by the kidneys. Unconjugated bilirubin ordinarily does not reach the gut except in neonates or, by ill-defined alternative pathways, in the presence of severe unconjugated hyperbilirubinemia [e.g., Crigler-Najjar syndrome, type I (CN-I)]. Unconjugated bilirubin that reaches the gut is partly reabsorbed, amplifying any underlying hyperbilirubinemia. Recent reports suggest that oral administration of calcium phosphate with or without the lipase inhibitor orlistat may be an efficient means to interrupt bilirubin enterohepatic cycling to reduce serum bilirubin levels in this situation. Although orlistat administration for 4–6 weeks to 16 patients with Crigler-Najjar syndrome was associated with a 10-20% decrease in serum bilirubin in 7 patients, the cost and side effects (i.e., diarrhea) may obviate the small benefit achievable with this treatment.

RENAL EXCRETION OF BILIRUBIN CONJUGATES

Unconjugated bilirubin is not excreted in urine, as it is too tightly bound to albumin for effective glomerular filtration and there is no tubular mechanism for its renal secretion. In contrast, the bilirubin conjugates are readily filtered at the glomerulus and can appear in urine in disorders characterized by increased bilirubin conjugates in the circulation.

DISORDERS OF BILIRUBIN METABOLISM LEADING TO UNCONJUGATED HYPERBILIRUBINEMIA

Increased Bilirubin Production

HEMOLYSIS

Increased destruction of erythrocytes leads to increased bilirubin turnover and unconjugated hyperbilirubinemia; the hyperbilirubinemia is usually modest in the presence of normal liver function. In particular, the bone marrow is only capable of a sustained eightfold increase in erythrocyte production in response to a hemolytic stress. Therefore, hemolysis alone cannot result in a sustained hyperbilirubinemia of more than ~68 μmol/L (4 mg/dL). Higher values imply concomitant hepatic dysfunction. When hemolysis is the only abnormality in an otherwise healthy individual, the result is a purely unconjugated hyperbilirubinemia, with the direct-reacting fraction as measured in a typical clinical laboratory being ≤15% of the total serum bilirubin. In the presence of systemic disease, which may include a degree of hepatic dysfunction, hemolysis may produce a component of conjugated hyperbilirubinemia in addition to an elevated unconjugated bilirubin concentration. Prolonged hemolysis may lead to the precipitation of bilirubin salts within the gallbladder or biliary tree, resulting in the formation of gallstones in which bilirubin, rather than cholesterol, is the major component. Such pigment stones may lead to acute or chronic cholecystitis, biliary obstruction, or any other biliary tract consequence of calculous disease.

INEFFECTIVE ERYTHROPOIESIS

During erythroid maturation, small amounts of hemoglobin may be lost at the time of nuclear extrusion, and a fraction of developing erythroid cells is destroyed within the marrow. These processes normally account for a small proportion of bilirubin that is produced. In various disorders, including thalassemia major, megaloblastic anemias due to folate or vitamin B12 deficiency, congenital erythropoietic porphyria, lead poisoning, and various congenital and acquired dyserythropoietic anemias, the fraction of total bilirubin production derived from ineffective erythropoiesis is increased, reaching as much as 70% of the total. This may be sufficient to produce modest degrees of unconjugated hyperbilirubinemia.

MISCELLANEOUS

Degradation of the hemoglobin of extravascular collections of erythrocytes, such as those seen in massive tissue

infarctions or large hematomas, may lead transiently to unconjugated hyperbilirubinemia.

Decreased Hepatic Bilirubin Clearance

DECREASED HEPATIC UPTAKE

Decreased hepatic bilirubin uptake is believed to contribute to the unconjugated hyperbilirubinemia of Gilbert's syndrome (GS), although the molecular basis for this finding remains unclear (see below). Several drugs, including flavaspidic acid, novobiocin, and rifampin, as well as various cholecystographic contrast agents, have been reported to inhibit bilirubin uptake. The resulting unconjugated hyperbilirubinemia resolves with cessation of the medication.

IMPAIRED CONJUGATION

Physiologic Neonatal Jaundice

Bilirubin produced by the fetus is cleared by the placenta and eliminated by the maternal liver. Immediately after birth, the neonatal liver must assume responsibility for bilirubin clearance and excretion. However, many hepatic physiologic processes are incompletely developed at birth. Levels of UGT1A1 are low, and alternative excretory pathways allow passage of unconjugated bilirubin into the gut. Since the intestinal flora that convert bilirubin to urobilinogen are also undeveloped, an enterohepatic circulation of unconjugated bilirubin ensues. As a consequence, most neonates develop mild unconjugated hyperbilirubinemia between days 2 and 5 after birth. Peak levels are typically <85–170 μmol/L (5–10 mg/dL) and decline to normal adult concentrations within 2 weeks, as mechanisms required for bilirubin disposition mature. Prematurity, often associated with more profound immaturity of hepatic function and hemolysis, can result in higher levels of unconjugated hyperbilirubinemia. A rapidly rising unconjugated bilirubin concentration, or absolute levels >340 μmol/L (20 mg/dL), puts the infant at risk for bilirubin encephalopathy, or kernicterus. Under these circumstances, bilirubin crosses an immature blood-brain barrier and precipitates in the basal ganglia and other areas of the brain. The consequences range from appreciable neurologic deficits to death. Treatment options include phototherapy, which converts bilirubin into water-soluble photoisomers that are excreted directly into bile, and exchange transfusion. The canalicular mechanisms responsible for bilirubin excretion are also immature at birth, and their maturation may lag behind that of UGT1A1; this can lead to transient conjugated neonatal hyperbilirubinemia, especially in infants with hemolysis.

Acquired Conjugation Defects

A modest reduction in bilirubin-conjugating capacity may be observed in advanced hepatitis or cirrhosis. However, in this setting, conjugation is better preserved than other aspects of bilirubin disposition, such as canalicular excretion. Various drugs, including pregnanediol, novobiocin, chloramphenicol, and gentamicin, may produce unconjugated hyperbilirubinemia by inhibiting UGT1A1 activity. Bilirubin conjugation may be inhibited by certain fatty acids that are present in breast milk but not serum of mothers whose infants have excessive neonatal hyperbilirubinemia (*breast milk jaundice*). Alternatively, there may be increased enterohepatic circulation of bilirubin in these infants. A recent study has correlated epidermal growth factor (EGF) content of breast milk with elevated bilirubin levels in these infants; however, a cause and effect relationship remains to be established. The pathogenesis of breast milk jaundice appears to differ from that of transient familial neonatal hyperbilirubinemia (Lucey-Driscoll syndrome), in which there is a UGT1A1 inhibitor in maternal serum.

Hereditary Defects in Bilirubin Conjugation

Three familial disorders characterized by differing degrees of unconjugated hyperbilirubinemia have long been recognized. The defining clinical features of each are described below **(Table 303-1)**. While these disorders have been recognized for decades to reflect differing degrees of deficiency in the ability to conjugate bilirubin, recent advances in the molecular biology of the *UGT1* gene complex have elucidated their interrelationships and clarified previously puzzling features.

Table 303-1 Principal Differential Characteristics of Gilbert's and Crigler-Najjar Syndromes

Feature	Crigler-Najjar Syndrome		Gilbert's Syndrome
	Type I	Type II	
Total serum bilirubin, μmol/L (mg/dL)	310–755 (usually >345) [18–45 (usually >20)]	100–430 (usually ≤ 345) [6–25 (usually ≤20)]	Typically ≤70 μmol/L (≤4 mg/dL) in absence of fasting or hemolysis
Routine liver tests	Normal	Normal	Normal
Response to Phenobarbital	None	Decreases bilirubin by >25%	Decreases bilirubin to normal
Kernicterus	Usual	Rare	No
Hepatic histology	Normal	Normal	Usually normal; increased lipofuscin pigment in some
Bile characteristics			
Color	Pale or colorless	Pigmented	Normal dark color
Bilirubin fractions	>90% unconjugated	Largest fraction (mean: 57%) monoconjugates	Mainly diconjugates but monoconjugates increased (mean: 23%)

Bilirubin UDP-glucuronosyltransferase activity	Typically absent; traces in some patients	Markedly reduced: 0–10% of normal	Reduced: typically 10–33% of normal
Inheritance (all autosomal)	Recessive	Predominantly recessive	Promoter mutation: recessive Missense mutations: 7 of 8 dominant; 1 reportedly recessive

CRIGLER-NAJJAR SYNDROME, TYPE I

CN-I is characterized by striking unconjugated hyperbilirubinemia of about 340–765 μmol/L (20–45 mg/dL) that appears in the neonatal period and persists for life. Other conventional hepatic biochemical tests such as serum aminotransferases and alkaline phosphatase are normal, and there is no evidence of hemolysis. Hepatic histology is also essentially normal except for the occasional presence of bile plugs within canaliculi. Bilirubin glucuronides are virtually absent from the bile, and there is no detectable constitutive expression of UGT1A1 activity in hepatic tissue. Neither UGT1A1 activity nor the serum bilirubin concentration responds to administration of phenobarbital or other enzyme inducers. In the absence of conjugation, unconjugated bilirubin accumulates in plasma, from which it is eliminated very slowly by alternative pathways that include direct passage into the bile and small intestine. These account for the small amounts of urobilinogen found in feces. No bilirubin is found in the urine. First described in 1952, the disorder is rare (estimated prevalence, 0.6–1.0 per million). Many patients are from geographically or socially isolated communities in which consanguinity is common, and pedigree analyses show an autosomal recessive pattern of inheritance. The majority of patients (type IA) exhibit defects in the glucuronide conjugation of a spectrum of substrates in addition to bilirubin, including various drugs and other xenobiotics. These individuals have mutations in one of the common exons (2–5) of the *UGT1* gene (Fig. 303-2). In a smaller subset (type IB), the defect is limited largely to bilirubin conjugation, and the causative mutation is in the bilirubin-specific exon A1. Estrogen glucuronidation is mediated by UGT1A1 and is defective in all CN-I patients. More than 30 different genetic lesions of *UGT1A1* responsible for CN-I have been identified, including deletions, insertions, alterations in intron splice donor and acceptor sites, exon skipping, and point mutations that introduce premature stop codons or alter critical amino acids. Their common feature is that they all encode proteins with absent or, at most, traces of bilirubin-UDP-glucuronosyltransferase enzymatic activity.

Prior to the availability of phototherapy, most patients with CN-I died of bilirubin encephalopathy (*kernicterus*) in infancy or early childhood. A few lived as long as early adult life without overt neurologic damage, although more subtle testing usually indicated mild but progressive brain damage. In the absence of liver transplantation, death eventually supervened from late-onset bilirubin encephalopathy, which often followed a nonspecific febrile illness. Although isolated hepatocyte transplantation has been used in a small number of cases of CN-I, early liver transplantation **(Chap. 310)** remains the best hope to prevent brain injury and death.

CRIGLER-NAJJAR SYNDROME, TYPE II (CN-II)

This condition was recognized as a distinct entity in 1962 and is characterized by marked unconjugated hyperbilirubinemia in the absence of abnormalities of other conventional hepatic biochemical tests, hepatic histology, or hemolysis. It differs from CN-I in several specific ways (Table 303-1): (1) Although there is considerable overlap, average bilirubin concentrations are lower in CN-II; (2) accordingly, CN-II is only infrequently associated with kernicterus; (3) bile is deeply colored, and bilirubin glucuronides are present, with a striking, characteristic increase in the proportion of monoglucuronides; (4) UGT1A1 in liver is usually present at reduced levels (typically ≤10% of normal) but may be undetectable by older, less sensitive assays; (5) while typically detected in infancy, hyperbilirubinemia was not recognized in some cases until later in life and, in one instance, at age 34. As with CN-I, most CN-II cases exhibit abnormalities in the conjugation of other compounds, such as salicylamide and menthol, but in some instances the defect appears limited to bilirubin. Reduction of serum bilirubin concentrations by >25% in response to enzyme inducers such as phenobarbital distinguishes CN-II from CN-I, although this response may not be elicited in early infancy and often is not accompanied by measurable UGT1A1 induction. Bilirubin concentrations during phenobarbital administration do not return to normal but are typically in the range of 51–86 μmol/L (3–5 mg/dL). Although the incidence of kernicterus in CN-II is low, instances have occurred, not only in infants but also in adolescents and adults, often in the setting of an intercurrent illness, fasting, or another factor that temporarily raises the serum bilirubin concentration above baseline and reduces serum albumin levels. For this reason, phenobarbital therapy is widely recommended, a single bedtime dose often sufficing to maintain clinically safe plasma bilirubin concentrations.

Over 77 different mutations in the *UGT1* gene have been identified as causing CN-I or CN-II. It was found that missense mutations are more common in CN-II patients, as would be expected in this less severe phenotype. Their common feature is that they encode for a bilirubin-UDP-glucuronosyltransferase with markedly reduced, but detectable, enzymatic activity. The spectrum of residual enzyme activity explains the spectrum of phenotypic severity of the resulting hyperbilirubinemia. Molecular analysis has established that a large majority of CN-II patients are either homozygotes or compound heterozygotes for CN-II mutations and that individuals carrying one mutated and one entirely normal allele have normal bilirubin concentrations.

GILBERT'S SYNDROME

This syndrome is characterized by mild unconjugated hyperbilirubinemia, normal values for standard hepatic biochemical tests, and normal hepatic histology other than a modest increase of lipofuscin pigment in some patients. Serum bilirubin concentrations are most often <51 μmol/L (<3 mg/dL), although both higher and lower values are frequent. The clinical spectrum of hyperbilirubinemia fades into that of CN-II at serum bilirubin concentrations of 86–136 μmol/L (5–8 mg/dL). At the other end of the scale, the distinction between mild cases of GS and a normal state is often blurred. Bilirubin concentrations may fluctuate substantially in any given individual, and at least 25% of patients will exhibit temporarily normal values during prolonged follow-up. More elevated values are associated with stress, fatigue, alcohol use, reduced caloric intake, and intercurrent illness, while increased caloric intake or administration of enzyme-inducing agents produces lower bilirubin levels. GS is most often diagnosed at or shortly after puberty or in adult life during routine examinations that include multichannel

biochemical analyses. UGT1A1 activity is typically reduced to 10–35% of normal, and bile pigments exhibit a characteristic increase in bilirubin monoglucuronides. Studies of radiobilirubin kinetics indicate that hepatic bilirubin clearance is reduced to an average of one-third of normal. Administration of phenobarbital normalizes both the serum bilirubin concentration and hepatic bilirubin clearance; however, failure of UGT1A1 activity to improve in many such instances suggests the possible coexistence of an additional defect. Compartmental analysis of bilirubin kinetic data suggests that GS patients have a defect in bilirubin uptake as well as in conjugation. Defect(s) in the hepatic uptake of other organic anions that at least partially share an uptake mechanism with bilirubin, such as sulfobromophthalein and indocyanine green (ICG), are observed in a minority of patients. The metabolism and transport of bile acids that do not utilize the bilirubin uptake mechanism, are normal. The magnitude of changes in the plasma bilirubin concentration induced by provocation tests such as 48 hours of fasting or the IV administration of nicotinic acid have been reported to be of help in separating GS patients from normal individuals. Other studies dispute this assertion. Moreover, on theoretical grounds, the results of such studies should provide no more information than simple measurements of the baseline plasma bilirubin concentration. Family studies indicate that GS and hereditary hemolytic anemias such as hereditary spherocytosis, glucose-6-phosphate dehydrogenase deficiency, and β-thalassemia trait sort independently. Reports of hemolysis in up to 50% of GS patients are believed to reflect better case finding, since patients with both GS and hemolysis have higher bilirubin concentrations, and are more likely to be jaundiced, than patients with either defect alone.

GS is common, with many series placing its prevalence at ≥8%. Males predominate over females by reported ratios ranging from 1.5:1 to >7:1. However, these ratios may have a large artifactual component since normal males have higher mean bilirubin levels than normal females, but the diagnosis of GS is often based on comparison to normal ranges established in men. The high prevalence of GS in the general population may explain the reported frequency of mild unconjugated hyperbilirubinemia in liver transplant recipients. The disposition of most xenobiotics metabolized by glucuronidation appears to be normal in GS, as is oxidative drug metabolism in the majority of reported studies. The principal exception is the metabolism of the antitumor agent irinotecan (CPT-11), whose active metabolite (SN-38) is glucuronidated specifically by bilirubin-UDP-glucuronosyltransferase. Administration of CPT-11 to patients with GS has resulted in several toxicities, including intractable diarrhea and myelosuppression. Some reports also suggest abnormal disposition of menthol, estradiol benzoate, acetaminophen, tolbutamide, and rifamycin SV. Although some of these studies have been disputed, and there have been no reports of clinical complications from use of these agents in GS, prudence should be exercised in prescribing them, or any agents metabolized primarily by glucuronidation, in this condition. It should also be noted that the HIVprotease inhibitors indinavir and atazanavir **(Chap. 189)** can inhibit UGT1A1, resulting in hyperbilirubinemia that is most pronounced in patients with preexisting GS.

Most older pedigree studies of GS were consistent with autosomal dominant inheritance with variable expressivity. However, studies of the *UGT1* gene in GS have indicated a variety of molecular genetic bases for the phenotypic picture and several different patterns of inheritance. Studies in Europe and the United States found that nearly all patients had normal coding regions for UGT1A1 but were homozygous for the insertion of an extra TA (i.e., A[TA]$_7$TAA rather than A[TA]$_6$TAA) in the promoter region of the first exon. This appeared to be necessary, but not sufficient, for clinically expressed GS, since 15% of normal controls were also homozygous for this variant. While normal by standard criteria, these individuals had somewhat higher bilirubin concentrations than the rest of the controls studied. Heterozygotes for this abnormality had bilirubin concentrations identical to those homozygous for the normal A[TA]$_6$TAA allele. The prevalence of the A[TA]$_7$TAA allele in a general Western population is 30%, in which case 9% would be homozygotes. This is slightly higher than the prevalence of GS based on purely phenotypic parameters. It was suggested that additional variables, such as mild hemolysis or a defect in bilirubin uptake, might be among the factors enhancing phenotypic expression of the defect.

Phenotypic expression of GS due solely to the A[TA]$_7$TAA promoter abnormality is inherited as an autosomal recessive trait. A number of CN-II kindreds have been identified in whom there is also an allele containing a normal coding region but the A[TA]$_7$TAA promoter abnormality. CN-II heterozygotes who have the A[TA]$_6$TAA promoter are phenotypically normal, whereas those with the A[TA]$_7$TAA promoter express the phenotypic picture of GS. GS in such kindreds may also result from homozygosity for the A[TA]$_7$TAA promoter abnormality. Seven different missense mutations in the *UGT1* gene that reportedly cause GS with dominant inheritance have been found in Japanese individuals. Another Japanese patient with mild unconjugated hyperbilirubinemia was homozygous for a missense mutation in exon 5. GS in her family appeared to be recessive. Missense mutations causing GS have not been reported outside of certain Asian populations.

DISORDERS OF BILIRUBIN METABOLISM LEADING TO MIXED OR PREDOMINANTLY CONJUGATED HYPERBILIRUBINEMIA

In hyperbilirubinemia due to acquired liver disease (e.g., acute hepatitis, common bile duct stone), there are usually elevations in the serum concentrations of both conjugated and unconjugated bilirubin. Although biliary tract obstruction or hepatocellular cholestatic injury may present on occasion with a predominantly conjugated hyperbilirubinemia, it is generally not possible to differentiate intrahepatic from extrahepatic causes of jaundice based on the serum levels or relative proportions of unconjugated and conjugated bilirubin. The major reason for determining the amounts of conjugated and unconjugated bilirubin in the serum is for the initial differentiation of hepatic parenchymal and obstructive disorders (mixed conjugated and unconjugated hyperbilirubinemia) from the inheritable and hemolytic disorders discussed above that are associated with unconjugated hyperbilirubinemia.

Familial Defects in Hepatic Excretory Function

DUBIN-JOHNSON SYNDROME (DJS)

This benign, relatively rare disorder is characterized by low-grade, predominantly conjugated hyperbilirubinemia **(Table 303-2)**. Total bilirubin concentrations are typically between 34 and 85 μmol/L (2 and 5 mg/dL) but on occasion can be in the normal range or as high as 340–430 μmol/L (20–25 mg/dL) and can fluctuate widely in any given patient. The degree of hyperbilirubinemia may be increased by intercurrent illness, oral contraceptive use,

and pregnancy. As the hyperbilirubinemia is due to a predominant rise in conjugated bilirubin, bilirubinuria is characteristically present. Aside from elevated serum bilirubin levels, other routine laboratory tests are normal. Physical examination is usually normal except for jaundice, although an occasional patient may have hepatosplenomegaly.

Table 303-2 Principal Differential Characteristics of Inheritable Disorders of Bile Canalicular Function

	DJS	Rotor	PFIC1	BRIC1	PFIC2	BRIC2	PFIC3
Gene	*ABCCA*	?	*ATP8B1*	*ATP8B1*	*ABCB11*	*ABCB11*	*ABCB4*
Protein	MRP2	?	FIC1	FIC1	BSEP	BSEP	MDR3
Cholestasis	No	No	Yes	Episodic	Yes	Episodic	Yes
Serum γ-GT	Normal	Normal	Normal	Normal	Normal	Normal	↑↑
Serum bile acids	Normal	Normal ↑	↑↑	↑↑ during episodes	↑↑	↑↑ during episodes	↑↑
Clinical features	Mild conjugated hyperbilirubinemia; otherwise normal liver function; dark pigment in liver; characteristic pattern of urinary coproporphyrins	Mild conjugated hyperbilirubinemia; otherwise normal liver function; liver without abnormal pigmentation	Severe cholestasis beginning in childhood	Recurrent episodes of cholestasis beginning at any age	Severe cholestasis beginning in childhood	Recurrent episodes of cholestasis beginning at any age	Severe cholestasis beginning in childhood; decreased phospholipids in bile

Patients with DJS are usually asymptomatic, although some may have vague constitutional symptoms. These latter patients have usually undergone extensive and often unnecessary diagnostic examinations for unexplained jaundice and have high levels of anxiety. In women, the condition may be subclinical until the patient becomes pregnant or receives oral contraceptives, at which time chemical hyperbilirubinemia becomes frank jaundice. Even in these situations, other routine liver function tests, including serum alkaline phosphatase and transaminase activities, are normal.

A cardinal feature of DJS is the accumulation in the lysosomes of centrilobular hepatocytes of dark, coarsely granular pigment. As a result, the liver may be grossly black in appearance. This pigment is thought to be derived from epinephrine metabolites that are not excreted normally. The pigment may disappear during bouts of viral hepatitis, only to reaccumulate slowly after recovery.

Biliary excretion of a number of anionic compounds is compromised in DJS. These include various cholecystographic agents, as well as sulfobromophthalein (Bromsulphalein, BSP), a synthetic dye formerly used in a test of liver function. In this test, the rate of disappearance of BSP from plasma was determined following bolus IV administration. BSP is conjugated with glutathione in the hepatocyte; the resulting conjugate is normally excreted rapidly into the bile canaliculus. Patients with DJS exhibit characteristic rises in plasma concentrations at 90 minutes after injection, due to reflux of conjugated BSP into the circulation from the hepatocyte. Dyes such as ICG that are taken up by hepatocytes but are not further metabolized prior to biliary excretion do not show this reflux phenomenon. Continuous BSP infusion studies suggest a reduction in the t_{max} for biliary excretion. Bile acid disposition, including hepatocellular uptake and biliary excretion, is normal in DJS. These patients have normal serum and biliary bile acid concentrations and do not have pruritus.

By analogy with findings in several mutant rat strains, the selective defect in biliary excretion of bilirubin conjugates and certain other classes of organic compounds, but not of bile acids, that characterizes DJS in humans was found to reflect defective expression of MRP2, an ATP-dependent canalicular membrane transporter. Several different mutations in the *MRP2* gene produce the Dubin-Johnson phenotype, which has an autosomal recessive pattern of inheritance. Although MRP2 is undoubtedly important in the biliary excretion of conjugated bilirubin, the fact that this pigment is still excreted in the absence of MRP2 suggests that other, as yet uncharacterized, transport proteins may serve in a secondary role in this process.

Patients with DJS also have a diagnostic abnormality in urinary coproporphyrin excretion. There are two naturally occurring coproporphyrin isomers, I and III. Normally, ~75% of the coproporphyrin in urine is isomer III. In urine from DJS patients, total coproporphyrin content is normal, but >80% is isomer I. Heterozygotes for the syndrome show an intermediate pattern. The molecular basis for this phenomenon remains unclear.

ROTOR SYNDROME

This benign, autosomal recessive disorder is clinically similar to DJS (Table 303-2), although it is seen even less frequently. A major phenotypic difference is that the liver in patients with Rotor syndrome has no increased pigmentation and appears totally normal. The only abnormality in routine laboratory tests is an elevation of total serum bilirubin, due to a predominant rise in conjugated bilirubin. This is accompanied by bilirubinuria. Several additional features differentiate Rotor syndrome from DJS. In Rotor syndrome, the gallbladder is usually visualized on oral cholecystography, in contrast to the nonvisualization that is typical of DJS. The pattern of urinary coproporphyrin excretion also differs. The pattern in Rotor syndrome resembles that of many acquired disorders of hepatobiliary function, in which coproporphyrin I, the major coproporphyrin isomer in bile, refluxes from the hepatocyte back into the circulation and is excreted in urine. Thus, total urinary coproporphyrin excretion is substantially increased in Rotor syndrome, in contrast to the normal levels seen in DJS. Although the fraction of coproporphyrin I in urine is elevated, it is usually <70% of the total, compared with ≥80% in DJS. The disorders also can be distinguished by their patterns of BSP excretion. Although clearance of BSP from plasma is delayed in

Rotor syndrome, there is no reflux of conjugated BSP back into the circulation as seen in DJS. Kinetic analysis of plasma BSP infusion studies suggests the presence of a defect in intrahepatocellular storage of this compound. This has never been demonstrated directly, and the molecular basis of Rotor syndrome remains unknown.

BENIGN RECURRENT INTRAHEPATIC CHOLESTASIS (BRIC)

This rare disorder is characterized by recurrent attacks of pruritus and jaundice. The typical episode begins with mild malaise and elevations in serum aminotransferase levels, followed rapidly by rises in alkaline phosphatase and conjugated bilirubin and onset of jaundice and itching. The first one or two episodes may be misdiagnosed as acute viral hepatitis. The cholestatic episodes, which may begin in childhood or adulthood, can vary in duration from several weeks to months, followed by a complete clinical and biochemical resolution. Intervals between attacks may vary from several months to years. Between episodes, physical examination is normal, as are serum levels of bile acids, bilirubin, transaminases, and alkaline phosphatase. The disorder is familial and has an autosomal recessive pattern of inheritance. BRIC is considered a benign disorder in that it does not lead to cirrhosis or end-stage liver disease. However, the episodes of jaundice and pruritus can be prolonged and debilitating, and some patients have undergone liver transplantation to relieve the intractable and disabling symptoms. Treatment during the cholestatic episodes is symptomatic; there is no specific treatment to prevent or shorten the occurrence of episodes.

A gene termed *FIC1* was recently identified and found to be mutated in patients with BRIC. Curiously, this gene is expressed strongly in the small intestine but only weakly in the liver. The protein encoded by *FIC1* shows little similarity to those that have been shown to play a role in bile canalicular excretion of various compounds. Rather, it appears to be a member of a P-type ATPase family that transports aminophospholipids from the outer to the inner leaflet of a variety of cell membranes. Its relationship to the pathobiology of this disorder remains unclear. A second phenotypically identical form of BRIC, termed BRIC type 2, has been described resulting from mutations in the bile salt excretory protein (BSEP), the protein that is defective in progressive familial intrahepatic cholestasis type 2 (Table 303-2). How some mutations in this protein result in the episodic BRIC phenotype is unknown.

PROGRESSIVE FAMILIAL INTRAHEPATIC CHOLESTASIS (FIC)

This name is applied to three phenotypically related syndromes (Table 303-2). Progressive FIC type 1 (Byler disease) presents in early infancy as cholestasis that may be initially episodic. However, in contrast to BRIC, Byler's disease progresses to malnutrition, growth retardation, and end-stage liver disease during childhood. This disorder is also a consequence of an *FIC1* mutation. The functional relationship of the FIC1 protein to the pathogenesis of cholestasis in these disorders is unknown. Two other types of progressive FIC (types 2 and 3) have been described. Progressive FIC type 2 is associated with a mutation in the protein named *sister of p-glycoprotein*, which is the major bile canalicular exporter of bile acids and is also known as *bile salt excretory protein*. As noted above, some mutations of this protein are associated with BRIC type 2, rather than the progressive FIC type 2 phenotype. Progressive FIC type 3 has been associated with a mutation of MDR3, a protein that is essential for normal hepatocellular excretion of phospholipids across the bile canaliculus. Although all three types of progressive FIC have similar clinical phenotypes, only type 3 is associated with high serum levels of γ-glutamyltransferase activity. In contrast, activity of this enzyme is normal or only mildly elevated in symptomatic BRIC and progressive FIC types 1 and 2.

FURTHER READINGS

Davit-Spraul A et al: ATP8B1 and ABCB11 analysis in 62 children with normal gamma-glutamyl transferase progressive familial intrahepatic cholestasis (PFIC): Phenotypic differences between PFIC1 and PFIC2 and natural history. Hepatology 51: 1645, 2010[PMID: 20232290] [Full Text]

Fevery J: Bilirubin in clinical practice: A review. Liver Int 28:592, 2008[PMID: 18433389] [Full Text]

Kagawa T et al: Phenotypic differences in PFIC2 and BRIC2 correlate with protein stability of mutant BSEP and impaired taurocholate secretion in MDCK II cells. Am J Physiol Gastrointest Liver Physiol 294:G58, 2008

Strassburg, CP: Pharmacogenetics of Gilbert's syndrome. Pharmacogenomics 9:703, 2008.

Strauss KA et al: Management of hyperbilirubinemia and prevention of kernicterus in 20 patients with Crigler-Najjar disease. Eur J Pediatr 165:306, 2006[PMID: 16435131] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACUTE VIRAL HEPATITIS: INTRODUCTION

Acute viral hepatitis is a systemic infection affecting the liver predominantly. Almost all cases of acute viral hepatitis are caused by one of five viral agents: hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the HBV-associated delta agent or hepatitis D virus (HDV), and hepatitis E virus (HEV). Other transfusion-transmitted agents (e.g., "hepatitis G" virus and "TT" virus, have been identified but do not cause hepatitis). All these human hepatitis viruses are RNA viruses, except for hepatitis B, which is a DNA virus. Although these agents can be distinguished by their molecular and antigenic properties, all types of viral hepatitis produce clinically similar illnesses. These range from asymptomatic and inapparent to fulminant and fatal acute infections common to all types, on the one hand, and from subclinical persistent infections to rapidly progressive chronic liver disease with cirrhosis and even hepatocellular carcinoma, common to the bloodborne types (HBV, HCV, and HDV), on the other.

Virology and Etiology

HEPATITIS A

Hepatitis A virus is a nonenveloped 27-nm, heat-, acid-, and ether-resistant RNA virus in the *Hepatovirus* genus of the picornavirus family **(Fig. 304-1)**. Its virion contains four capsid polypeptides, designated VP1 to VP4, which are cleaved posttranslationally from the polyprotein product of a 7500-nucleotide genome. Inactivation of viral activity can be achieved by boiling for 1 minute, by contact with formaldehyde and chlorine, or by ultraviolet irradiation. Despite nucleotide sequence variation of up to 20% among isolates of HAV, and despite the recognition of four genotypes affecting humans, all strains of this virus are immunologically indistinguishable and belong to one serotype. Hepatitis A has an incubation period of ~4 weeks. Its replication is limited to the liver, but the virus is present in the liver, bile, stools, and blood during the late incubation period and acute preicteric phase of illness. Despite persistence of virus in the liver, viral shedding in feces, viremia, and infectivity diminish rapidly once jaundice becomes apparent. HAV can be cultivated reproducibly in vitro.

Figure 304-1

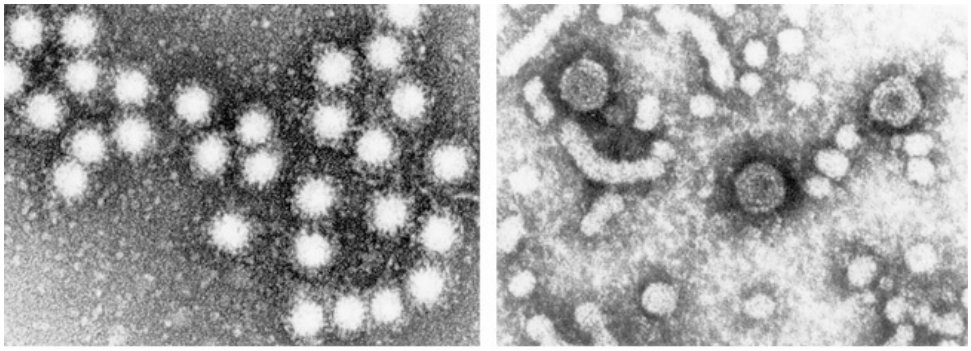

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Electron micrographs of hepatitis A virus particles and serum from a patient with hepatitis B. *Left*: 27-nm hepatitis A virus particles purified from stool of a patient with acute hepatitis A and aggregated by antibody to hepatitis A virus. ***Right***: Concentrated serum from a patient with hepatitis B, demonstrating the 42-nm virions, tubular forms, and spherical 22-nm particles of hepatitis B surface antigen. 132,000x. (Hepatitis D resembles 42-nm virions of hepatitis B but is smaller, 35–37 nm; hepatitis E resembles hepatitis A virus but is slightly larger, 32–34 nm; hepatitis C has been visualized as a 55-nm particle.)

Antibodies to HAV (anti-HAV) can be detected during acute illness when serum aminotransferase activity is elevated and fecal HAV shedding is still occurring. This early antibody response is predominantly of the IgM class and persists for several months, rarely for 6–12 months. During convalescence, however, anti-HAV of the IgG class becomes the predominant antibody **(Fig. 304-2)**. Therefore, the diagnosis of hepatitis A is made during acute illness by demonstrating anti-HAV of the IgM class. After acute illness, anti-HAV of the IgG class remains detectable indefinitely, and patients with serum anti-HAV are immune to reinfection. Neutralizing antibody activity parallels the appearance of anti-HAV, and the IgG anti-HAV present in immune globulin accounts for the protection it affords against HAV infection.

Figure 304-2

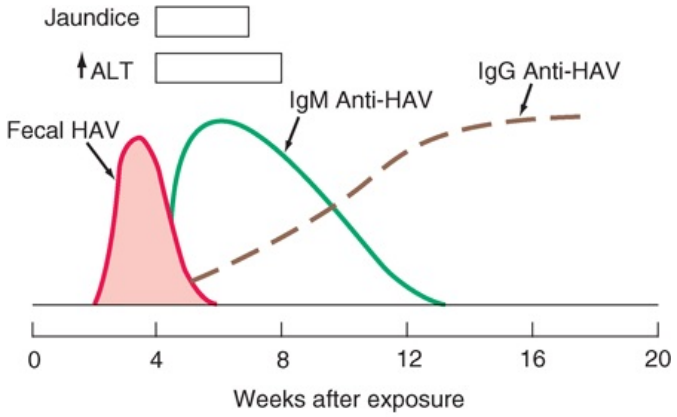

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scheme of typical clinical and laboratory features of hepatitis A.

HEPATITIS B

Hepatitis B virus is a DNA virus with a remarkably compact genomic structure; despite its small, circular, 3200-bp size, HBV DNA codes for four sets of viral products with a complex, multiparticle structure. HBV achieves its genomic economy by relying on an efficient strategy of encoding proteins from four overlapping genes: S, C, P, and X **(Fig. 304-3)**, as detailed below. Once thought to be unique among viruses, HBV is now recognized as one of a family of animal viruses, hepadnaviruses (hepatotropic DNA viruses), and is classified as hepadnavirus type 1. Similar viruses infect certain species of woodchucks, ground and tree squirrels, and Pekin ducks, to mention the most carefully characterized. Like HBV, all have the same distinctive three morphologic forms, have counterparts to the envelope and nucleocapsid virus antigens of HBV, replicate in the liver but exist in extrahepatic sites, contain their own endogenous DNA polymerase, have partially double-strand and partially single-strand genomes, are associated with acute and chronic hepatitis and hepatocellular carcinoma, and rely on a replicative strategy unique among DNA viruses but typical of retroviruses. Instead of DNA replication directly from a DNA template, hepadnaviruses rely on reverse transcription (effected by the DNA polymerase) of minus-strand DNA from a "pregenomic" RNA intermediate. Then plus-strand DNA is transcribed from the minus-strand DNA template by the DNA-dependent DNA polymerase and converted in the hepatocyte nucleus to a covalently closed circular DNA, which serves as a template for messenger RNA and pregenomic RNA. Viral proteins are translated by the messenger RNA, and the proteins and genome are packaged into virions and secreted from the hepatocyte. Although HBV is difficult to cultivate in vitro in the conventional sense from clinical material, several cell lines have been transfected with HBV DNA. Such transfected cells support in vitro replication of the intact virus and its component proteins.

Figure 304-3

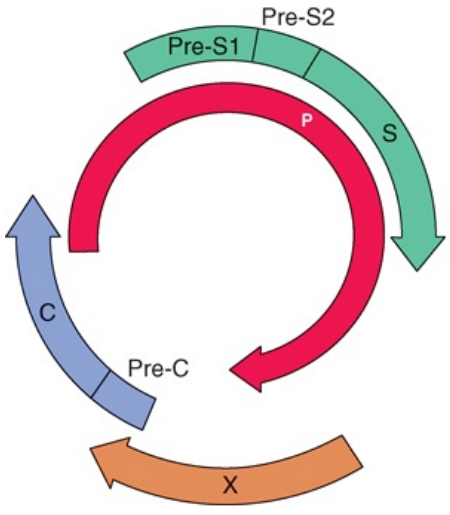

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Compact genomic structure of HBV. This structure, with overlapping genes, permits HBV to code for multiple proteins. The S gene codes for the "major" envelope protein, HBsAg. Pre-S1 and pre-S2

multiple proteins. The S gene codes for the "major" envelope protein, HBsAg; pre-S1 and pre-S2, upstream of S, combine with S to code for two larger proteins, "middle" protein, the product of pre-S2 + S, and "large" protein, the product of pre-S1 + pre-S2 + S. The largest gene, P, codes for DNA polymerase. The C gene codes for two nucleocapsid proteins, HBeAg, a soluble, secreted protein (initiation from the pre-C region of the gene) and HBcAg, the intracellular core protein (initiation after pre-C). The X gene codes for HBxAg, which can transactivate the transcription of cellular and viral genes; its clinical relevance is not known, but it may contribute to carcinogenesis by binding to p53.

Viral Proteins and Particles

Of the three particulate forms of HBV **(Table 304-1)**, the most numerous are the 22-nm particles, which appear as spherical or long filamentous forms; these are antigenically indistinguishable from the outer surface or envelope protein of HBV and are thought to represent excess viral envelope protein. Outnumbered in serum by a factor of 100 or 1000 to 1 compared with the spheres and tubules are large, 42-nm, double-shelled spherical particles, which represent the intact hepatitis B virion (Fig. 304-1). The envelope protein expressed on the outer surface of the virion and on the smaller spherical and tubular structures is referred to as *hepatitis B surface antigen* (HBsAg). The concentration of HBsAg and virus particles in the blood may reach 500 μg/mL and 10 trillion particles per milliliter, respectively. The envelope protein, HBsAg, is the product of the S gene of HBV.

Table 304-1 Nomenclature and Features of Hepatitis Viruses

Hepatitis Type	Virus Particle, nm	Morphology	Genome*	Classification	Antigen(s)	Antibodies	Remarks
HAV	27	Icosahedral nonenveloped	7.5-kb RNA, linear, ss, +	Hepatovirus	HAV	Anti-HAV	Early fecal shedding Diagnosis: IgM anti-HAV Previous infection: IgG anti-HAV
HBV	42	Double-shelled virion (surface and core) spherical	3.2-kb DNA, circular, ss/ds	Hepadnavirus	HBsAg HBcAg HBeAg	Anti-HBs Anti-HBc Anti-HBe	Bloodborne virus; carrier state Acute diagnosis: HBsAg, IgM anti-HBc Chronic diagnosis: IgG anti-HBc, HBsAg Markers of replication: HBeAg, HBV DNA Liver, lymphocytes, other organs
	27	Nucleocapsid core			HBcAg HBeAg	Anti-HBc Anti-HBe	Nucleocapsid contains DNA and DNA polymerase; present in hepatocyte nucleus; HBcAg does not circulate; HBeAg (soluble, nonparticulate) and HBV DNA circulate—correlate with infectivity and complete virions
	22	Spherical and filamentous; represents excess virus coat material			HBsAg	Anti-HBs	HBsAg detectable in >95% of patients with acute hepatitis B; found in serum, body fluids, hepatocyte cytoplasm; anti-HBs appears following infection

							—protective antibody
HCV	Approx. 40–60	Enveloped	9.4-kb RNA, linear, ss, +	Hepacivirus	HCV C100-3 C33c C22-3 NS5	Anti-HCV	Bloodborne agent, formerly labeled non-A, non-B hepatitis Acute diagnosis: anti-HCV (C33c, C22-3, NS5), HCV RNA Chronic diagnosis: anti-HCV (C100-3, C33c, C22-3, NS5) and HCV RNA; cytoplasmic location in hepatocytes
HDV	35–37	Enveloped hybrid particle with HBsAg coat and HDV core	1.7-kb RNA, circular, ss, –	Resembles viroids and plant satellite viruses	HBsAg HDV antigen	Anti-HBs Anti-HDV	Defective RNA virus, requires helper function of HBV (hepadnaviruses); HDV antigen present in hepatocyte nucleus Diagnosis: anti-HDV, HDV RNA; HBV/HDV coinfection—IgM anti-HBc and anti-HDV; HDV superinfection—IgG anti-HBc and anti-HDV
HEV	32–34	Nonenveloped icosahedral	7.6-kb RNA, linear, ss, +	Hepevirus	HEV antigen	Anti-HEV	Agent of enterically transmitted hepatitis; rare in USA; occurs in Asia, Mediterranean countries, Central America Diagnosis: IgM/IgG anti-HEV (assays not routinely available); virus in stool, bile, hepatocyte cytoplasm

*ss, single-strand; ss/ds, partially single-strand, partially double-strand; –, minus-strand; +, plus-strand.

A number of different HBsAg subdeterminants have been identified. There is a common group-reactive antigen, *a*, shared by all HBsAg isolates. In addition, HBsAg may contain one of several subtype-specific antigens—namely, *d* or *y*, *w* or *r*—as well as other more recently characterized specificities. Hepatitis B isolates fall into one of at least eight subtypes and eight genotypes (A–H). Geographic distribution of genotypes and subtypes varies; genotypes A (corresponding to subtype *adw*) and D (*ayw*) predominate in the United States and Europe, while genotypes B (*adw*) and C (*adr*) predominate in Asia. Clinical course and outcome are independent of subtype, but preliminary reports suggest that genotype B is associated with less rapidly progressive liver disease and a lower likelihood, or delayed appearance, of hepatocellular carcinoma than genotype C. Patients with genotype A appear to be more likely to clear circulating viremia and to achieve HBsAg seroconversion, both spontaneously and in response to antiviral therapy. In addition, "precore" mutations are favored by certain genotypes (see below).

Upstream of the S gene are the pre-S genes (Fig. 304-3), which code for pre-S gene products, including receptors on the HBV surface for polymerized human serum albumin and for hepatocyte membrane proteins. The pre-S region actually consists of both pre-S1 and pre-S2. Depending on where translation is

initiated, three potential HBsAg gene products are synthesized. The protein product of the S gene is HBsAg (*major protein*), the product of the S region plus the adjacent pre-S2 region is the *middle protein*, and the product of the pre-S1 plus pre-S2 plus S regions is the *large protein*. Compared with the smaller spherical and tubular particles of HBV, complete 42-nm virions are enriched in the large protein. Both pre-S proteins and their respective antibodies can be detected during HBV infection, and the period of pre-S antigenemia appears to coincide with other markers of virus replication, as detailed below.

The intact 42-nm virion contains a 27-nm nucleocapsid core particle. Nucleocapsid proteins are coded for by the C gene. The antigen expressed on the surface of the nucleocapsid core is referred to as *hepatitis B core antigen* (HBcAg), and its corresponding antibody is anti-HBc. A third HBV antigen is *hepatitis B e antigen* (HBeAg), a soluble, nonparticulate, nucleocapsid protein that is immunologically distinct from intact HBcAg but is a product of the same C gene. The C gene has two initiation codons, a precore and a core region (Fig. 304-3). If translation is initiated at the precore region, the protein product is HBeAg, which has a signal peptide that binds it to the smooth endoplasmic reticulum and leads to its secretion into the circulation. If translation begins with the core region, HBcAg is the protein product; it has no signal peptide, it is not secreted, but it assembles into nucleocapsid particles, which bind to and incorporate RNA, and which, ultimately, contain HBV DNA. Also packaged within the nucleocapsid core is a DNA polymerase, which directs replication and repair of HBV DNA. When packaging within viral proteins is complete, synthesis of the incomplete plus strand stops; this accounts for the single-strand gap and for differences in the size of the gap. HBcAg particles remain in the hepatocyte, where they are readily detectable by immunohistochemical staining, and are exported after encapsidation by an envelope of HBsAg. Therefore, naked core particles do not circulate in the serum. The secreted nucleocapsid protein, HBeAg, provides a convenient, readily detectable, qualitative marker of HBV replication and relative infectivity.

HBsAg-positive serum containing HBeAg is more likely to be highly infectious and to be associated with the presence of hepatitis B virions (and detectable HBV DNA, see below) than HBeAg-negative or anti-HBe-positive serum. For example, HBsAg carrier mothers who are HBeAg-positive almost invariably (>90%) transmit hepatitis B infection to their offspring, whereas HBsAg carrier mothers with anti-HBe rarely (10–15%) infect their offspring.

Early during the course of acute hepatitis B, HBeAg appears transiently; its disappearance may be a harbinger of clinical improvement and resolution of infection. Persistence of HBeAg in serum beyond the first three months of acute infection may be predictive of the development of chronic infection, and the presence of HBeAg during chronic hepatitis B is associated with ongoing viral replication, infectivity, and inflammatory liver injury.

The third of the HBV genes is the largest, the P gene (Fig. 304-3), which codes for the DNA polymerase; as noted above, this enzyme has both DNA-dependent DNA polymerase and RNA-dependent reverse transcriptase activities. The fourth gene, X, codes for a small, nonparticulate protein, *hepatitis B x antigen* (HBxAg), that is capable of transactivating the transcription of both viral and cellular genes (Fig. 304-3). In the cytoplasm, HBxAg effects calcium release (possibly from mitochondria), which activates signal-transduction pathways that lead to stimulation of HBV reverse transcription and HBV DNA replication. Such transactivation may enhance the replication of HBV, leading to the clinical association observed between the expression of HBxAg and antibodies to it in patients with severe chronic hepatitis and hepatocellular carcinoma. The transactivating activity can enhance the transcription and replication of other viruses besides HBV, such as HIV. Cellular processes transactivated by X include the human interferon γ gene and class I major histocompatibility genes; potentially, these effects could contribute to enhanced susceptibility of HBV-infected hepatocytes to cytolytic T cells. The expression of X can also induce programmed cell death (apoptosis).

Serologic and Virologic Markers

After a person is infected with HBV, the first virologic marker detectable in serum within 1–12 weeks, usually between 8–12 weeks, is HBsAg **(Fig. 304-4)**. Circulating HBsAg precedes elevations of serum aminotransferase activity and clinical symptoms by 2–6 weeks and remains detectable during the entire icteric or symptomatic phase of acute hepatitis B and beyond. In typical cases, HBsAg becomes undetectable 1–2 months after the onset of jaundice and rarely persists beyond 6 months. After HBsAg disappears, antibody to HBsAg (anti-HBs) becomes detectable in serum and remains detectable indefinitely thereafter. Because HBcAg is intracellular and, when in the serum, sequestered within an HBsAg coat, naked core particles do not circulate in serum and, therefore, HBcAg is not detectable routinely in the serum of patients with HBV infection. By contrast, anti-HBc is readily demonstrable in serum, beginning within the first 1–2 weeks after the appearance of HBsAg and preceding detectable levels of anti-HBs by weeks to months. Because variability exists in the time of appearance of anti-HBs after HBV infection, occasionally a gap of several weeks or longer may separate the disappearance of HBsAg and the appearance of anti-HBs. During this "gap" or "window" period, anti-HBc may represent the only serologic evidence of current or recent HBV infection, and blood containing anti-HBc in the absence of HBsAg and anti-HBs has been implicated in the development of transfusion-associated hepatitis B. In part because the sensitivity of immunoassays for HBsAg and anti-HBs has increased, however, this window period is rarely encountered. In some persons, years after HBV infection, anti-HBc may persist in the circulation longer than anti-HBs. Therefore, isolated anti-HBc does not necessarily indicate active virus replication; most instances of isolated anti-HBc represent hepatitis B infection in the remote past. Rarely, however, isolated anti-HBc represents low-level hepatitis B viremia, with HBsAg below the detection threshold; occasionally, isolated anti-HBc represents a cross-reacting or false-positive immunologic

specificity. Recent and remote HBV infections can be distinguished by determination of the immunoglobulin class of anti-HBc. Anti-HBc of the IgM class (IgM anti-HBc) predominates during the first six months after acute infection, whereas IgG anti-HBc is the predominant class of anti-HBc beyond six months. Therefore, patients with current or recent acute hepatitis B, including those in the anti-HBc window, have IgM anti-HBc in their serum. In patients who have recovered from hepatitis B in the remote past as well as those with chronic HBV infection, anti-HBc is predominantly of the IgG class. Infrequently, in ≤1–5% of patients with acute HBV infection, levels of HBsAg are too low to be detected; in such cases, the presence of IgM anti-HBc establishes the diagnosis of acute hepatitis B. When isolated anti-HBc occurs in the rare patient with chronic hepatitis B whose HBsAg level is below the sensitivity threshold of contemporary immunoassays (a low-level carrier), the anti-HBc is of the IgG class. Generally, in persons who have recovered from hepatitis B, anti-HBs and anti-HBc persist indefinitely.

Figure 304-4

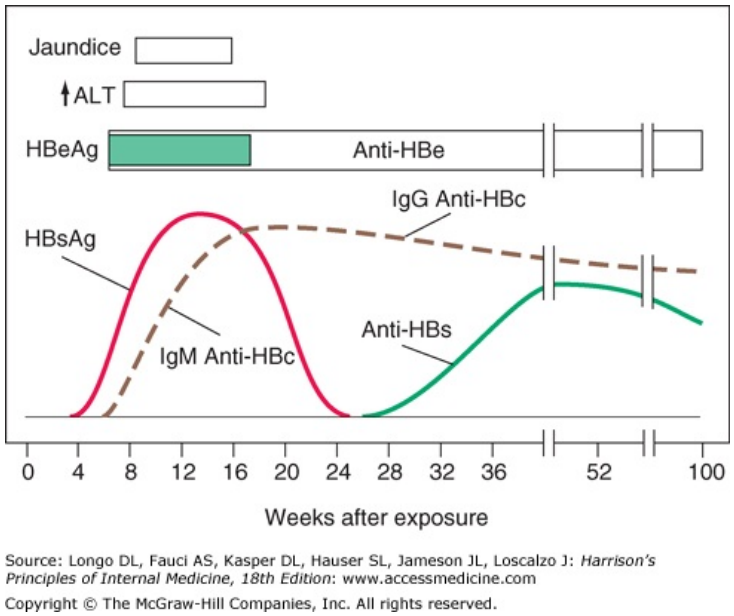

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scheme of typical clinical and laboratory features of acute hepatitis B.

The temporal association between the appearance of anti-HBs and resolution of HBV infection as well as the observation that persons with anti-HBs in serum are protected against reinfection with HBV suggests that *anti-HBs is the protective antibody*. Therefore, strategies for prevention of HBV infection are based on providing susceptible persons with circulating anti-HBs (see below). Occasionally, in 10–20% of patients with chronic hepatitis B, low-level, low-affinity anti-HBs can be detected. This antibody is directed against a subtype determinant different from that represented by the patient's HBsAg; its presence is thought to reflect the stimulation of a related clone of antibody-forming cells, but it has no clinical relevance and does not signal imminent clearance of hepatitis B. These patients with HBsAg and such nonneutralizing anti-HBs should be categorized as having chronic HBV infection.

The other readily detectable serologic marker of HBV infection, HBeAg, appears concurrently with or shortly after HBsAg. Its appearance coincides temporally with high levels of virus replication and reflects the presence of circulating intact virions and detectable HBV DNA (with the notable exception of patients with precore mutations who cannot synthesize HBeAg—see "Molecular Variants"). Pre-S1 and pre-S2 proteins are also expressed during periods of peak replication, but assays for these gene products are not routinely available. In self-limited HBV infections, HBeAg becomes undetectable shortly after peak elevations in aminotransferase activity, before the disappearance of HBsAg, and anti-HBe then becomes detectable, coinciding with a period of relatively lower infectivity (Fig. 304-4). Because markers of HBV replication appear transiently during acute infection, testing for such markers is of little clinical utility in typical cases of acute HBV infection. In contrast, markers of HBV replication provide valuable information in patients with protracted infections.

Departing from the pattern typical of acute HBV infections, in chronic HBV infection, HBsAg remains detectable beyond six months, anti-HBc is primarily of the IgG class, and anti-HBs is either undetectable or detectable at low levels (see "Laboratory Features") **(Fig. 304-5)**. During early chronic HBV infection, HBV DNA can be detected both in serum and in hepatocyte nuclei, where it is present in free or episomal form. This *replicative stage* of HBV infection is the time of maximal infectivity and liver injury; HBeAg is a qualitative marker and HBV DNA a quantitative marker of this replicative phase, during which all three forms of HBV circulate, including intact virions. Over time, the replicative phase of chronic HBV infection gives way to a relatively *nonreplicative phase*. This occurs at a rate of ~10% per year and is accompanied by seroconversion from HBeAg-positive to anti-HBe-positive. In most cases, this seroconversion coincides with a transient, acute hepatitis-like elevation in aminotransferase activity, believed to reflect cell-mediated

immune clearance of virus-infected hepatocytes. In the nonreplicative phase of chronic infection, when HBV DNA is demonstrable in hepatocyte nuclei, it tends to be integrated into the host genome. In this phase, only spherical and tubular forms of HBV, *not intact virions*, circulate, and liver injury tends to subside. Most such patients would be characterized as *inactiveHBV carriers*. In reality, the designations *replicative* and *nonreplicative* are only relative; even in the so-called nonreplicative phase, HBV replication can be detected at levels of ~≤10^3 virions with highly sensitive amplification probes such as the polymerase chain reaction (PCR); below this replication threshold, liver injury and infectivity of HBV are limited to negligible. Still, the distinctions are pathophysiologically and clinically meaningful. Occasionally, nonreplicative HBV infection converts back to replicative infection. Such spontaneous reactivations are accompanied by re-expression of HBeAg and HBV DNA, and sometimes of IgM anti-HBc, as well as by exacerbations of liver injury. Because high-titer IgM anti-HBc can reappear during acute exacerbations of chronic hepatitis B, relying on IgM anti-HBc versus IgG anti-HBc to distinguish between acute and chronic hepatitis B infection, respectively, may not always be reliable; in such cases, patient history is invaluable in helping to distinguish de novo acute hepatitis B infection from acute exacerbation of chronic hepatitis B infection.

Figure 304-5

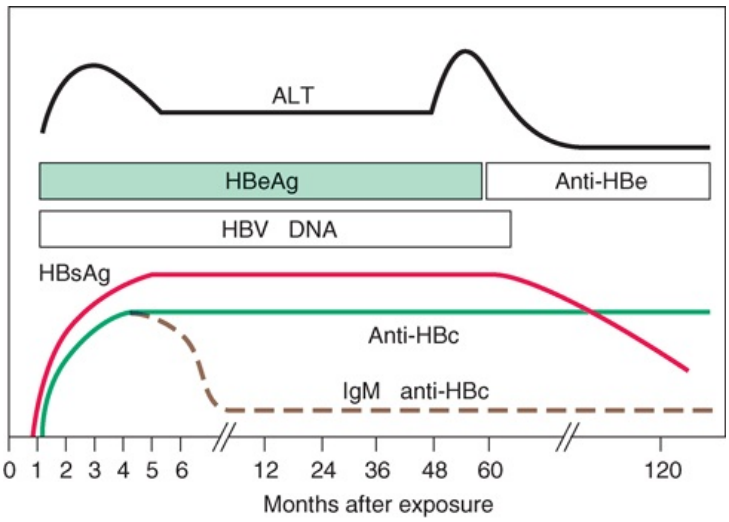

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scheme of typical laboratory features of wild-type chronic hepatitis B. HBeAg and HBV DNA can be detected in serum during the *replicative phase* of chronic infection, which is associated with infectivity and liver injury. Seroconversion from the replicative phase to the *nonreplicative phase* occurs at a rate of ~10% per year and is heralded by an acute hepatitis- like elevation of ALT activity; during the nonreplicative phase, infectivity and liver injury are limited. In HBeAg-negative chronic hepatitis B associated with mutations in the precore region of the HBV genome, replicative chronic hepatitis B occurs in the absence of HBeAg.

Molecular Variants

Variation occurs throughout the HBV genome, and clinical isolates of HBV that do not express typical viral proteins have been attributed to mutations in individual or even multiple gene locations. For example, variants have been described that lack nucleocapsid proteins, envelope proteins, or both. Two categories of naturally occurring HBV variants have attracted the most attention. One of these was identified initially in Mediterranean countries among patients with an unusual serologic clinical profile. They have severe chronic HBV infection and detectable HBV DNA but with anti-HBe instead of HBeAg. These patients were found to be infected with an HBV mutant that contained an alteration in the precore region rendering the virus incapable of encoding HBeAg. Although several potential mutation sites exist in the pre-C region, the region of the C gene necessary for the expression of HBeAg (see "Virology and Etiology"), the most commonly encountered in such patients is a single base substitution, from G to A, which occurs in the second to last codon of the pre-C gene at nucleotide 1896. This substitution results in the replacement of the TGG tryptophan codon by a stop codon (T*A*G), which prevents the translation of HBeAg. Another mutation, in the core-promoter region, prevents transcription of the coding region for HBeAg and yields an HBeAg-negative phenotype. Patients with such mutations in the precore region and who are unable to secrete HBeAg tend to have severe liver disease that progresses more rapidly to cirrhosis or, alternatively, they are identified clinically later in the course of the natural history of chronic hepatitis B, when the disease is more advanced. Both "wild-type" HBV and precore-mutant HBV can coexist in the same patient, or mutant HBV may arise late during wild-type HBV infection. In addition, clusters of fulminant hepatitis B in Israel and Japan have been attributed to common-source infection with a precore mutant. Fulminant hepatitis B in North America and western Europe, however, occurs in patients infected with wild-type HBV, in the absence of precore mutants, and both precore mutants and other mutations throughout the HBV

genome occur commonly, even in patients with typical, self-limited, milder forms of HBV infection. HBeAg-negative chronic hepatitis with mutations in the precore region is now the most frequently encountered form of hepatitis B in Mediterranean countries and in Europe. In the United States, where HBV genotype A (less prone to G1896A mutation) is prevalent, precore-mutant HBV is much less common; however, as a result of immigration from Asia and Europe, the proportion of HBeAg-negative hepatitis B–infected individuals has increased in the United States, and they now represent approximately one-third of patients with chronic hepatitis B. Characteristic of such HBeAg-negative chronic hepatitis B are lower levels of HBV DNA (usually $\leq 10^5$ copies/mL) and one of several patterns of aminotransferase activity—persistent elevations, periodic fluctuations above the normal range, and periodic fluctuations between the normal and elevated range.

The second important category of HBV mutants consists of *escape mutants*, in which a single amino acid substitution, from glycine to arginine, occurs at position 145 of the immunodominant *a* determinant common to all subtypes of HBsAg. This change in HBsAg leads to a critical conformational change that results in a loss of neutralizing activity by anti-HBs. This specific HBV/*a* mutant has been observed in two situations, active and passive immunization, in which humoral immunologic pressure may favor evolutionary change ("escape") in the virus—in a small number of hepatitis B vaccine recipients who acquired HBV infection despite the prior appearance of neutralizing anti-HBs and in liver transplant recipients who underwent the procedure for hepatitis B and who were treated with a high-potency human monoclonal anti-HBs preparation. Although such mutants have not been recognized frequently, their existence raises a concern that may complicate vaccination strategies and serologic diagnosis.

Different types of mutations emerge during antiviral therapy of chronic hepatitis B with nucleoside analogues; such "YMDD" and similar mutations in the polymerase motif of HBV are described in Chap. 306.

Extrahepatic Sites

Hepatitis B antigens and HBV DNA have been identified in extrahepatic sites, including lymph nodes, bone marrow, circulating lymphocytes, spleen, and pancreas. Although the virus does not appear to be associated with tissue injury in any of these extrahepatic sites, its presence in these "remote" reservoirs has been invoked (but is not necessary) to explain the recurrence of HBV infection after orthotopic liver transplantation. A more complete understanding of the clinical relevance of extrahepatic HBV remains to be defined.

HEPATITIS D

The delta hepatitis agent, or HDV, the only member of the genus *Deltavirus*, is a defective RNA virus that coinfects with and requires the helper function of HBV (or other hepadnaviruses) for its replication and expression. Slightly smaller than HBV, delta is a formalin-sensitive, 35- to 37-nm virus with a hybrid structure. Its nucleocapsid expresses delta antigen, which bears no antigenic homology with any of the HBV antigens, and contains the virus genome. The delta core is "encapsidated" by an outer envelope of HBsAg, indistinguishable from that of HBV except in its relative compositions of major, middle, and large HBsAg component proteins. The genome is a small, 1700-nucleotide, circular, single-strand RNA of negative polarity that is nonhomologous with HBV DNA (except for a small area of the polymerase gene) but that has features and the rolling circle model of replication common to genomes of plant satellite viruses or viroids. HDV RNA contains many areas of internal complementarity; therefore, it can fold on itself by internal base pairing to form an unusual, very stable, rodlike structure that contains a very stable, self-cleaving and self-ligating ribozyme. HDV RNA requires host RNA polymerase II for its replication via RNA-directed RNA synthesis by transcription of genomic RNA to a complementary antigenomic (plus strand) RNA; the antigenomic RNA, in turn, serves as a template for subsequent genomic RNA synthesis. HDV RNA has only one open reading frame, and delta antigen (HDAg), a product of the antigenomic strand is the only known HDV protein; HDAg exists in two forms: a small, 195-amino-acid species, which plays a role in facilitating HDV RNA replication, and a large, 214-amino-acid species, which appears to suppress replication but is required for assembly of the antigen into virions. Delta antigens have been shown to bind directly to RNA polymerase II, resulting in stimulation of transcription. Although complete hepatitis D virions and liver injury require the cooperative helper function of HBV, intracellular replication of HDV RNA can occur without HBV. Genomic heterogeneity among HDV isolates has been described; however, pathophysiologic and clinical consequences of this genetic diversity have not been recognized. The clinical spectrum of hepatitis D is common to all seven genotypes identified, the predominant of which is genotype 1.

HDV can either infect a person simultaneously with HBV (*co-infection*) or superinfect a person already infected with HBV (*super-infection*); when HDV infection is transmitted from a donor with one HBsAg subtype to an HBsAg-positive recipient with a different subtype, the HDV agent assumes the HBsAg subtype of the recipient, rather than the donor. Because HDV relies absolutely on HBV, the duration of HDV infection is determined by the duration of (and cannot outlast) HBV infection. HDV antigen is expressed primarily in hepatocyte nuclei and is occasionally detectable in serum. During acute HDV infection, anti-HDV of the IgM class predominates, and 30–40 days may elapse after symptoms appear before anti-HDV can be detected. In self-limited infection, anti-HDV is low-titer and transient, rarely remaining detectable beyond the clearance of HBsAg and HDV antigen. In chronic HDV infection, anti-HDV circulates in high titer, and both IgM and IgG anti-HDV can be detected. HDV antigen in the liver and HDV RNA in serum and liver can be detected during HDV replication.

HEPATITIS C

Hepatitis C virus, which, before its identification was labeled "non-A, non-B hepatitis," is a linear, single-strand, positive-sense, 9600-nucleotide RNA virus, the genome of which is similar in organization to that of flaviviruses and pestiviruses; HCV is the only member of the genus *Hepacivirus* in the family Flaviviridae. The HCV genome contains a single, large open reading frame (gene) that codes for a virus polyprotein of ~3000 amino acids, which is cleaved after translation to yield 10 viral proteins. The 5' end of the genome consists of an untranslated region (containing an internal ribosomal entry site) adjacent to the genes for four structural proteins, the nucleocapsid core protein, C; two envelope glycoproteins, E1 and E2; and a membrane protein p7. The 5' untranslated region and core gene are highly conserved among genotypes, but the envelope proteins are coded for by the hypervariable region, which varies from isolate to isolate and may allow the virus to evade host immunologic containment directed at accessible virus-envelope proteins. The 3' end of the genome also includes an untranslated region and contains the genes for six nonstructural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A, and NS5B. The NS2 cysteineprotease cleaves NS3 from NS2, and the NS3-4A serine protease cleaves all the downstream proteins from the polyprotein. Important NS proteins involved in virus replication include the NS3 helicase, NS3-NS4A serine protease, and the NS5B RNA-dependent RNA polymerase **(Fig. 304-6)**. Because HCV does not replicate via a DNA intermediate, it does not integrate into the host genome. Because HCV tends to circulate in relatively low titer, 10^3–10^7 virions/mL, visualization of virus particles, estimated to be 40–60 nm in diameter, remains difficult. Still, the replication rate of HCV is very high, 10^{12} virions per day; its half-life is 2.7 hours. The chimpanzee is a helpful but cumbersome animal model. Although a robust, reproducible, small animal model is lacking, HCV replication has been documented in an immunodeficient mouse model containing explants of human liver and in transgenic mouse and rat models. Although in vitro replication has been difficult, hepatocellular carcinoma–derived cell lines have been described (replicon systems) that support replication of genetically manipulated, truncated, or full-length HCV RNA (but not intact virions). Recently, complete replication of HCV and intact 55-nm virions have been described in cell culture systems. HCV gains entry into the hepatocyte via the nonliver-specific CD81 receptor and the liver-specific tight junction protein claudin-1. Relying on the same assembly and secretion pathway as low-density lipoproteins (LPLs), HCV masquerades as a lipoprotein, which may limit its visibility to the adaptive immune system and which may explain its ability to evade immune containment and clearance.

Figure 304-6

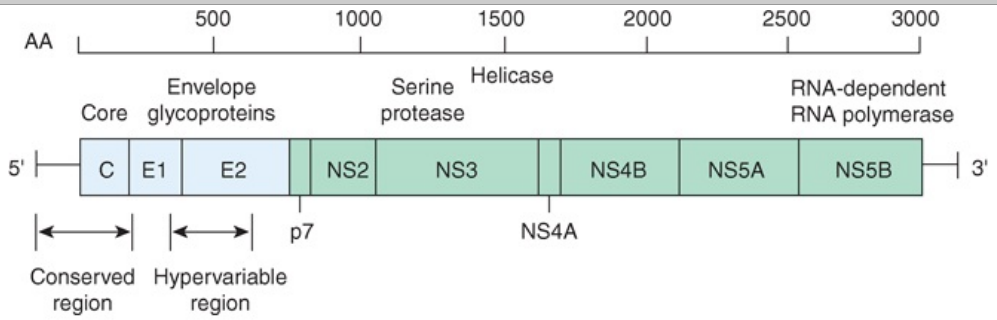

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Organization of the hepatitis C virus genome and its associated, 3000 amino-acid (AA) proteins. The three structural genes at the 5' end are the core region, C, which codes for the nucleocapsid, and the envelope regions, E1 and E2, which code for envelope glycoproteins. The 5' untranslated region and the C region are highly conserved among isolates, while the envelope domain E2 contains the hypervariable region. Adjacent to the structural proteins is p7, a membrane protein that appears to function as an ion channel. At the 3' end are six nonstructural (NS) regions, NS2, which codes for a cysteineprotease; NS3, which codes for a serine protease and an RNA helicase; NS4 and NS4B; NS5A; and NS5B, which codes for an RNA-dependent RNA polymerase. After translation of the entire polyprotein, individual proteins are cleaved by both host and viral proteases.

At least six distinct major genotypes, as well as >50 subtypes within genotypes, of HCV have been identified by nucleotide sequencing. Genotypes differ one from another in sequence homology by ≥30%. Because divergence of HCV isolates within a genotype or subtype and, within the same host, may vary insufficiently to define a distinct genotype, these intragenotypic differences are referred to as *quasispecies* and differ in sequence homology by only a few percent. The genotypic and quasispecies diversity of HCV, resulting from its high mutation rate, interferes with effective humoral immunity. Neutralizing antibodies to HCV have been demonstrated, but they tend to be short lived, and HCV infection does not induce lasting immunity against reinfection with different virus isolates or even the same virus isolate. Thus, neither *heterologous* nor *homologous* immunity appears to develop commonly after acute HCV infection. Some HCV genotypes are distributed worldwide, while others are more geographically confined (see "Epidemiology and Global Features"). In addition, differences exist among genotypes in responsiveness to antiviral therapy; however, early reports of differences in pathogenicity among genotypes have not been corroborated.

Currently available, third-generation immunoassays, which incorporate proteins from the core, NS3, and NS5 regions, detect anti-HCV antibodies during acute infection. The most sensitive indicator of HCV infection is the presence of HCV RNA, which requires molecular amplification by PCR or transcription-mediated amplification (TMA) **(Fig. 304-7)**. To allow standardization of the quantification of HCV RNA among laboratories and commercial assays, HCV RNA is reported as international units (IUs) per milliliter; quantitative assays are available that allow detection of HCV RNA with a sensitivity as low as 5 IU/mL. HCV RNA can be detected within a few days of exposure to HCV—well before the appearance of anti-HCV— and tends to persist for the duration of HCV infection; however, occasionally in patients with chronic HCV infection, HCV RNA may be detectable only intermittently. Application of sensitive molecular probes for HCV RNA has revealed the presence of replicative HCV in peripheral blood lymphocytes of infected persons; however, as is the case for HBV in lymphocytes, the clinical relevance of HCV lymphocyte infection is not known.

HEPATITIS E

Previously labeled *epidemic* or *enterically transmitted non-A, non-B hepatitis*, HEV is an enterically transmitted virus that occurs primarily in India, Asia, Africa, and Central America; in those geographic areas, HEV is the most common cause of acute hepatitis. This agent, with epidemiologic features resembling those of hepatitis A, is a 32- to 34-nm, nonenveloped, HAV-like virus with a 7600-nucleotide, single-strand, positive-sense RNA genome. HEV has three open reading frames (ORF) (genes), the largest of which, *ORF1*, encodes nonstructural proteins involved in virus replication. A middle-sized gene, *ORF2*, encodes the nucleocapsid protein, the major nonstructural protein, and the smallest, *ORF3*, encodes a structural protein whose function remains undetermined. All HEV isolates appear to belong to a single serotype, despite genomic heterogeneity of up to 25% and the existence of five genotypes, only four of which have been detected in humans; genotypes 1 and 2 appear to be more virulent, while genotypes 3 and 4 are more attenuated and account for subclinical infections. Contributing to the perpetuation of this virus are animal reservoirs, most notably in swine. There is no genomic or antigenic homology, however, between HEV and HAV or other picornaviruses; and HEV, although resembling caliciviruses, is sufficiently distinct from any known agent to merit a new classification of its own as a unique genus, *Hepevirus*, within the family Hepeviridae. The virus has been detected in stool, bile, and liver and is excreted in the stool during the late incubation period; immune responses to viral antigens occur very early during the course of acute infection. Both IgM anti-HEV and IgG anti-HEV can be detected, but both fall rapidly after acute infection, reaching low levels within 9–12 months. Currently, serologic testing for HEV infection is not available routinely.

Figure 304-7

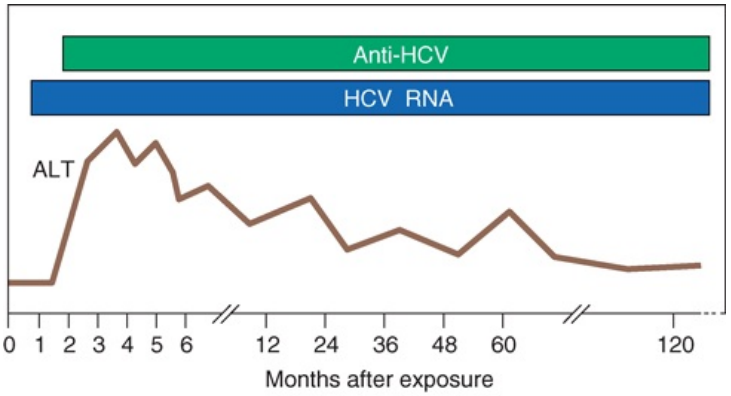

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scheme of typical laboratory features during acute hepatitis C progressing to chronicity. HCV RNA is the first detectable event, preceding alanine aminotrans ferase (ALT) elevation and the appearance of anti-HCV.

Pathogenesis

Under ordinary circumstances, none of the hepatitis viruses is known to be directly cytopathic to hepatocytes. Evidence suggests that the clinical manifestations and outcomes after acute liver injury associated with viral hepatitis are determined by the immunologic responses of the host. Among the viral hepatitides, the immunopathogenesis of hepatitis B and C have been studied most extensively.

HEPATITIS B

For HBV, the existence of inactive hepatitis B carriers with normal liver histology and function suggests that the virus is not directly cytopathic. The fact that patients with defects in cellular immune competence are more likely to remain chronically infected rather than to clear HBV supports the role of cellular immune responses in the pathogenesis of hepatitis B–related liver injury. The model that has the most experimental support involves cytolytic T cells sensitized specifically to recognize host and hepatitis B viral antigens on the liver cell surface. Nucleocapsid proteins (HBcAg and possibly HBeAg), present on the cell membrane

in minute quantities, are the viral target antigens that, with host antigens, invite cytolytic T cells to destroy HBV-infected hepatocytes. Differences in the robustness and broad polyclonality of CD8+ cytolytic T cell responsiveness and in the elaboration of antiviral cytokines by T cells have been invoked to explain differences in outcomes between those who recover after acute hepatitis, and those who progress to chronic hepatitis, or between those with mild and those with severe (fulminant) acute HBV infection.

Although a robust cytolytic T cell response occurs and eliminates virus-infected liver cells during acute hepatitis B, >90% of HBV DNA has been found in experimentally infected chimpanzees to disappear from the liver and blood before maximal T cell infiltration of the liver and before most of the biochemical and histologic evidence of liver injury. This observation suggests that components of the innate immune system and inflammatory cytokines, independent of cytopathic antiviral mechanisms, participate in the early immune response to HBV infection; this effect has been shown to represent elimination of HBV replicative intermediates from the cytoplasm and covalently closed circular viral DNA from the nucleus of infected hepatocytes. Ultimately, HBV-HLA-specific cytolytic T cell responses of the adaptive immune system are felt to be responsible for recovery from HBV infection.

Debate continues over the relative importance of viral and host factors in the pathogenesis of HBV-associated liver injury and its outcome. As noted above, precore genetic mutants of HBV have been associated with the more severe outcomes of HBV infection (severe chronic and fulminant hepatitis), suggesting that, under certain circumstances, relative pathogenicity is a property of the virus, not the host. The fact that concomitant HDV and HBV infections are associated with more severe liver injury than HBV infection alone and the fact that cells transfected in vitro with the gene for HDV (delta) antigen express HDV antigen and then become necrotic in the absence of any immunologic influences are also consistent with a viral effect on pathogenicity. Similarly, in patients who undergo liver transplantation for end-stage chronic hepatitis B, occasionally, rapidly progressive liver injury appears in the new liver. This clinical pattern is associated with an unusual histologic pattern in the new liver, *fibrosing cholestatic hepatitis*, which, ultrastructurally, appears to represent a choking of the cell with overwhelming quantities of HBsAg. This observation suggests that, under the influence of the potent immunosuppressive agents required to prevent allograft rejection, HBV may have a direct cytopathic effect on liver cells, independent of the immune system.

Although the precise mechanism of liver injury in HBV infection remains elusive, studies of nucleocapsid proteins have shed light on the profound immunologic tolerance to HBV of babies born to mothers with highly replicative (HBeAg-positive), chronic HBV infection. In HBeAg-expressing transgenic mice, in utero exposure to HBeAg, which is sufficiently small to traverse the placenta, induces T cell tolerance to both nucleocapsid proteins. This, in turn, may explain why, when infection occurs so early in life, immunologic clearance does not occur, and protracted, lifelong infection ensues.

An important distinction should be drawn between HBV infection acquired at birth, common in endemic areas, such as the Far East, and infection acquired in adulthood, common in the west. Infection in the neonatal period is associated with the acquisition of immunologic tolerance to HBV, absence of an acute hepatitis illness, but the almost invariable establishment of chronic, often lifelong infection. Neonatally acquired HBV infection can culminate decades later in cirrhosis and hepatocellular carcinoma (see "Complications and Sequelae"). In contrast, when HBV infection is acquired during adolescence or early adulthood, the host immune response to HBV infected hepatocytes tends to be robust, an acute hepatitis-like illness is the rule, and failure to recover is the exception. After adulthood acquired infection, chronicity is uncommon, and the risk of hepatocellular carcinoma is very low. Based on these observations, some authorities categorize HBV infection into an "immunotolerant" phase, an "immunoreactive" phase, and an "inactive" phase. This somewhat simplistic formulation does not apply at all to the typical adult in the west with self-limited acute hepatitis B, in whom no period of immunologic tolerance occurs. Even among those with neonatally acquired HBV infection, in whom immunologic tolerance is established definitively, intermittent bursts of hepatic necroinflammatory activity punctuate the period during the early decades of life during which liver injury appears to be quiescent (labeled by some as the "immunotolerant" phase). In addition, even when clinically apparent, liver injury and progressive fibrosis emerge during later decades (the so-called immunoreactive, or immunointolerant phase), the level of immunologic tolerance to HBV remains substantial. More accurately, in patients with neonatally acquired HBV infection, a dynamic equilibrium exists between tolerance and intolerance, the outcome of which determines the clinical expression of chronic infection. Those individuals who are infected as neonates tend to have a relatively higher level of immunologic tolerance during the early decades of life and a relatively lower level (but only rarely a loss) of tolerance in the later decades of life.

HEPATITIS C

Cell-mediated immune responses and elaboration by T cells of antiviral cytokines contribute to the containment of infection and pathogenesis of liver injury associated with hepatitis C. Perhaps HCV infection of lymphoid cells plays a role in moderating immune responsiveness to the virus, as well. Intrahepatic HLA class I restricted cytolytic T cells directed at nucleocapsid, envelope, and nonstructural viral protein antigens have been demonstrated in patients with chronic hepatitis C; however, such virus-specific cytolytic T cell responses do not correlate adequately with the degree of liver injury or with recovery. Yet, a consensus has emerged supporting a role in the pathogenesis of HCV-associated liver injury of virus-activated CD4 helper T cells that stimulate, via the cytokines they elaborate, HCV-specific CD8 cytotoxic T cells. These responses appear to be more robust (higher in number, more diverse in viral antigen specificity, more functionally effective, and more long lasting) in those who recover from HCV than

in those who have chronic infection. Several HLA alleles have been linked with self-limited hepatitis C, the most convincing of which is the C/C haplotype of the IL28B gene. Although attention has focused on adaptive immunity, HCV proteins have been shown to interfere with innate immunity by resulting in blocking of type 1 interferon responses and inhibition of interferon signaling and effector molecules in the interferon signaling cascade. Also shown to contribute to limiting HCV infection are natural killer cells of the innate immune system that function when HLA class 1 molecules required for successful adaptive immunity are underexpressed. Of note, the emergence of substantial viral quasispecies diversity and HCV sequence variation allow the virus to evade attempts by the host to contain HCV infection by both humoral and cellular immunity.

Finally, cross-reactivity between viral antigens (HCV NS3 and NS5A) and host autoantigens (cytochrome P450 2D6) has been invoked to explain the association between hepatitis C and a subset of patients with autoimmune hepatitis and antibodies to liver-kidney microsomal (LKM) antigen (anti-LKM) **(Chap. 306)**.

Extrahepatic Manifestations

Immune complex–mediated tissue damage appears to play a pathogenetic role in the extrahepatic manifestations of acute hepatitis B. The occasional prodromal serum sickness–like syndrome observed in acute hepatitis B appears to be related to the deposition in tissue blood vessel walls of HBsAg-anti-HBs circulating immune complexes, leading to activation of the complement system and depressed serum complement levels.

In patients with chronic hepatitis B, other types of immune-complex disease may be seen. Glomerulonephritis with the nephrotic syndrome is observed occasionally; HBsAg, immunoglobulin, and C3 deposition has been found in the glomerular basement membrane. While generalized vasculitis (polyarteritis nodosa) develops in considerably fewer than 1% of patients with chronic HBV infection, 20–30% of patients with polyarteritis nodosa have HBsAg in serum **(Chap. 326)**. In these patients, the affected small- and medium-size arterioles contain HBsAg, immunoglobulins, and complement components. Another extrahepatic manifestation of viral hepatitis, essential mixed cryoglobulinemia (EMC), was reported initially to be associated with hepatitis B. The disorder is characterized clinically by arthritis; cutaneous vasculitis (palpable purpura); and, occasionally, with glomerulonephritis and serologically by the presence of circulating cryoprecipitable immune complexes of more than one immunoglobulin class **(Chaps. 283 and 326)**. Many patients with this syndrome have chronic liver disease, but the association with HBV infection is limited; instead, a substantial proportion has chronic HCV infection, with circulating immune complexes containing HCV RNA. Immune-complex glomerulonephritis is another recognized extrahepatic manifestation of chronic hepatitis C.

Pathology

The typical morphologic lesions of all types of viral hepatitis are similar and consist of panlobular infiltration with mononuclear cells, hepatic cell necrosis, hyperplasia of Kupffer cells, and variable degrees of cholestasis. Hepatic cell regeneration is present, as evidenced by numerous mitotic figures, multinucleated cells, and "rosette" or "pseudoacinar" formation. The mononuclear infiltration consists primarily of small lymphocytes, although plasma cells and eosinophils occasionally are present. Liver cell damage consists of hepatic cell degeneration and necrosis, cell dropout, ballooning of cells, and acidophilic degeneration of hepatocytes (forming so-called Councilman or apoptotic bodies). Large hepatocytes with a ground-glass appearance of the cytoplasm may be seen in chronic but not in acute HBV infection; these cells contain HBsAg and can be identified histochemically with orcein or aldehyde fuchsin. In uncomplicated viral hepatitis, the reticulin framework is preserved.

In hepatitis C, the histologic lesion is often remarkable for a relative paucity of inflammation, a marked increase in activation of sinusoidal lining cells, lymphoid aggregates, the presence of fat (more frequent in genotype 3 and linked to increased fibrosis), and, occasionally, bile duct lesions in which biliary epithelial cells appear to be piled up without interruption of the basement membrane. Occasionally, microvesicular steatosis occurs in hepatitis D. In hepatitis E, a common histologic feature is marked cholestasis. A cholestatic variant of slowly resolving acute hepatitis A also has been described.

A more severe histologic lesion, *bridging hepatic necrosis*, also termed *subacute* or *confluent necrosis* or *interface hepatitis*, is observed occasionally in acute hepatitis. "Bridging" between lobules results from large areas of hepatic cell dropout, with collapse of the reticulin framework. Characteristically, the bridge consists of condensed reticulum, inflammatory debris, and degenerating liver cells that span adjacent portal areas, portal to central veins, or central vein to central vein. This lesion had been thought to have prognostic significance; in many of the originally described patients with this lesion, a subacute course terminated in death within several weeks to months, or severe chronic hepatitis and cirrhosis developed; however, the association between bridging necrosis and a poor prognosis in patients with acute hepatitis has not been upheld. Therefore, although demonstration of this lesion in patients with chronic hepatitis has prognostic significance **(Chap. 306)**, its demonstration during acute hepatitis is less meaningful, and liver biopsies to identify this lesion are no longer undertaken routinely in patients with acute hepatitis. In *massive hepatic necrosis* (fulminant hepatitis, "acute yellow atrophy"), the striking feature at postmortem examination is the finding of a small, shrunken, soft liver. Histologic examination reveals massive necrosis and dropout of liver cells of most lobules with extensive collapse and condensation of the reticulin framework. When histologic documentation is required in the management of fulminant or very severe hepatitis, a biopsy can be done by the angiographically guided transjugular route, which permits the performance of this invasive procedure in the presence of severe coagulopathy.

Immunohistochemical and electron-microscopic studies have localized HBsAg to the cytoplasm and plasma membrane of infected liver cells. In contrast, HBcAg predominates in the nucleus, but, occasionally, scant amounts are also seen in the cytoplasm and on the cell membrane. HDV antigen is localized to the hepatocyte nucleus, while HAV, HCV, and HEV antigens are localized to the cytoplasm.

Epidemiology and Global Features

Before the availability of serologic tests for hepatitis viruses, all viral hepatitis cases were labeled either as "infectious" or "serum" hepatitis. Modes of transmission overlap, however, and *a clear distinction among the different types of viral hepatitis cannot be made solely on the basis of clinical or epidemiologic features*(**Table 304-2**). The most accurate means to distinguish the various types of viral hepatitis involves specific serologic testing.

Table 304-2 Clinical and Epidemiologic Features of Viral Hepatitis

Feature	HAV	HBV	HCV	HDV	HEV
Incubation (days)	15–45, mean 30	30–180, mean 60–90	15–160, mean 50	30–180, mean 60–90	14–60, mean 40
Onset	Acute	Insidious or acute	Insidious	Insidious or acute	Acute
Age preference	Children, young adults	Young adults (sexual and percutaneous), babies, toddlers	Any age, but more common in adults	Any age (similar to HBV)	Young adults (20–40 years)
Transmission					
Fecal-oral	+++	–	–	–	+++
Percutaneous	Unusual	+++	+++	+++	–
Perinatal	–	+++	±[a]	+	–
Sexual	±	++	±[a]	++	–
Clinical					
Severity	Mild	Occasionally severe	Moderate	Occasionally severe	Mild
Fulminant	0.1%	0.1–1%	0.1%	5–20%[b]	1–2%[e]
Progression to chronicity	None	Occasional (1–10%) (90% of neonates)	Common (85%)	Common[d]	None
Carrier	None	0.1–30%[c]	1.5–3.2%	Variable[f]	None
Cancer	None	+(Neonatal infection)	+	±	None
Prognosis	Excellent	Worse with age, debility	Moderate	Acute, good Chronic, poor	Good
Prophylaxis	IG, inactivated vaccine	HBIG, recombinant vaccine	None	HBV vaccine (none for HBV carriers)	Vaccine
Therapy	None	Interferon Lamivudine Adefovir Pegylated interferon Entecavir Telbivudine Tenofovir	Pegylated interferon plus ribavirin telaprevir boceprevir	Interferon ±	None

[a]Primarily with HIV co-infection and high-level viremia in index case; risk ~5%.

[b]Up to 5% in acute HBV/HDV co-infection; up to 20% in HDV superinfection of chronic HBV infection.

[c]Varies considerably throughout the world and in subpopulations within countries; see text.

[d] In acute HBV/HDV co-infection, the frequency of chronicity is the same as that for HBV; in HDV superinfection, chronicity is invariable.

[e] 10–20% in pregnant women.

[f] Common in Mediterranean countries, rare in North America and western Europe.

Abbreviation: HBIG, hepatitis B immunoglobulin.

HEPATITIS A

This agent is transmitted almost exclusively by the fecal-oral route . Person-to-person spread of HAV is enhanced by poor personal hygiene and overcrowding; large outbreaks as well as sporadic cases have been traced to contaminated food, water, milk, frozen raspberries and strawberries, green onions imported from Mexico, and shellfish. Intrafamily and intrainstitutional spread are also common. Early epidemiologic observations supported a predilection for hepatitis A to occur in late fall and early winter. In temperate zones, epidemic waves have been recorded every 5–20 years as new segments of nonimmune population appeared; however, in developed countries, the incidence of hepatitis A has been declining, presumably as a function of improved sanitation, and these cyclic patterns are no longer observed. No HAV carrier state has been identified after acute hepatitis A; perpetuation of the virus in nature depends presumably on nonepidemic, inapparent subclinical infection, ingestion of contaminated food or water in, or imported from, endemic areas, and/or contamination linked to environmental reservoirs.

In the general population, anti-HAV, a marker for previous HAV infection, increases in prevalence as a function of increasing age and of decreasing socioeconomic status. In the 1970s, serologic evidence of prior hepatitis A infection occurred in ~40% of urban populations in the United States, most of whose members never recalled having had a symptomatic case of hepatitis. In subsequent decades, however, the prevalence of anti-HAV has been declining in the United States. In developing countries, exposure, infection, and subsequent immunity are almost universal in childhood. As the frequency of subclinical childhood infections declines in developed countries, a susceptible cohort of adults emerges. Hepatitis A tends to be more symptomatic in adults; therefore, paradoxically, as the frequency of HAV infection declines, the likelihood of clinically apparent, even severe, HAV illnesses increases in the susceptible adult population. Travel to endemic areas is a common source of infection for adults from nonendemic areas. More recently recognized epidemiologic foci of HAV infection include child-care centers, neonatal intensive care units, promiscuous men who have sex with men, and injection drug users. Although hepatitis A is rarely bloodborne, several outbreaks have been recognized in recipients of clotting-factor concentrates. In the United States, the introduction of hepatitis A vaccination programs among children from high-incidence states has resulted in a >70% reduction in the annual incidence of new HAV infections and has shifted the burden of new infections from children to young adults.

HEPATITIS B

Percutaneous inoculation has long been recognized as a major route of hepatitis B transmission, but the outmoded designation "serum hepatitis" is an inaccurate label for the epidemiologic spectrum of HBV infection recognized today. As detailed below, most of the hepatitis transmitted by blood transfusion is not caused by HBV; moreover, in approximately two-thirds of patients with acute type B hepatitis, no history of an identifiable percutaneous exposure can be elicited. We now recognize that many cases of hepatitis B result from less obvious modes of nonpercutaneous or covert percutaneous transmission. HBsAg has been identified in almost every body fluid from infected persons, and at least some of these body fluids —most notably semen and saliva—are infectious, albeit less so than serum, when administered percutaneously or nonpercutaneously to experimental animals. Among the nonpercutaneous modes of HBV transmission, oral ingestion has been documented as a potential but inefficient route of exposure. By contrast, the two nonpercutaneous routes considered to have the greatest impact are intimate (especially sexual) contact and perinatal transmission.

In sub-Saharan Africa, intimate contact among toddlers is considered instrumental in contributing to the maintenance of the high frequency of hepatitis B in the population. Perinatal transmission occurs primarily in infants born to HBsAg carrier mothers or mothers with acute hepatitis B during the third trimester of pregnancy or during the early postpartum period. Perinatal transmission is uncommon in North America and western Europe but occurs with great frequency and is the most important mode of HBV perpetuation in the Far East and developing countries. Although the precise mode of perinatal transmission is unknown, and although ~10% of infections may be acquired in utero, epidemiologic evidence suggests that most infections occur approximately at the time of delivery and are not related to breast-feeding. The likelihood of perinatal transmission of HBV correlates with the presence of HBeAg and high-level viral replication; 90% of HBeAg-positive mothers but only 10–15% of anti-HBe-positive mothers transmit HBV infection to their offspring. In most cases, acute infection in the neonate is clinically asymptomatic, but the child is very likely to remain chronically infected.

The >350–400 million HBsAg carriers in the world constitute the main reservoir of hepatitis B in human beings. Whereas serum HBsAg is infrequent (0.1–0.5%) in normal populations in the United States and western Europe, a prevalence of up to 5–20% has been found in the Far East and in some tropical countries; in persons with Down's syndrome, lepromatous leprosy, leukemia, Hodgkin's disease, polyarteritis nodosa; in patients with chronic renal disease on hemodialysis; and in injection drug users.

Other groups with high rates of HBV infection include spouses of acutely infected persons; sexually

promiscuous persons (especially promiscuous men who have sex with men); health care workers exposed to blood; persons who require repeated transfusions especially with pooled blood-product concentrates (e.g., hemophiliacs); residents and staff of custodial institutions for the developmentally handicapped; prisoners; and, to a lesser extent, family members of chronically infected patients. In volunteer blood donors, the prevalence of anti-HBs, a reflection of previous HBV infection, ranges from 5–10%, but the prevalence is higher in lower socioeconomic strata, older age groups, and persons—including those mentioned above—exposed to blood products. Because of highly sensitive virologic screening of donor blood, the risk of acquiring HBV infection from a blood transfusion is 1 in 230,000.

Prevalence of infection, modes of transmission, and human behavior conspire to mold geographically different epidemiologic patterns of HBV infection. In the Far East and Africa, hepatitis B, a disease of the newborn and young children, is perpetuated by a cycle of maternal-neonatal spread. In North America and western Europe, hepatitis B is primarily a disease of adolescence and early adulthood, the time of life when intimate sexual contact as well as recreational and occupational percutaneous exposures tend to occur. To some degree, however, this dichotomy between high-prevalence and low-prevalence geographic regions has been minimized by immigration from high-prevalence to low-prevalence areas. The introduction of hepatitis B vaccine in the early 1980s and adoption of universal childhood vaccination policies in many countries resulted in a dramatic, ~90%, decline in the incidence of new HBV infections in those countries as well as in the dire consequences of chronic infection, including hepatocellular carcinoma. Populations and groups for whom HBV-infection screening is recommended are listed in **Table 304-3**.

Table 304-3 High-Risk Populations for Which HBV-Infection Screening Is Recommended

Persons born in countries/regions with a high (>8%) and intermediate (>2%) prevalence of HBV infection including immigrants and adopted children and including persons born in the United States who were not vaccinated as infants and whose parents immigrated from areas of high HBV endemicity.
Household and sexual contacts of persons with hepatitis B
Persons who have used injection drugs
Persons with multiple sexual contacts or a history of sexually transmitted disease
Men who have sex with men
Inmates of correctional facilities
Persons with elevated alanine or aspartate aminotransferase levels
Persons with HCV or HIV infection
Hemodialysis patients
Pregnant women
Persons who require immunosuppressive or cytotoxic therapy

HEPATITIS D

Infection with HDV has a worldwide distribution, but two epidemiologic patterns exist. In Mediterranean countries (northern Africa, southern Europe, the Middle East), HDV infection is endemic among those with hepatitis B, and the disease is transmitted predominantly by nonpercutaneous means, especially close personal contact. In nonendemic areas, such as the United States and northern Europe, HDV infection is confined to persons exposed frequently to blood and blood products, primarily injection drug users and hemophiliacs. HDV infection can be introduced into a population through drug users or by migration of persons from endemic to nonendemic areas. Thus, patterns of population migration and human behavior facilitating percutaneous contact play important roles in the introduction and amplification of HDV infection. Occasionally, the migrating epidemiology of hepatitis D is expressed in explosive outbreaks of severe hepatitis, such as those that have occurred in remote South American villages as well as in urban centers in the United States. Ultimately, such outbreaks of hepatitis D—either of co-infections with acute hepatitis B or of superinfections in those already infected with HBV—may blur the distinctions between endemic and nonendemic areas. On a global scale, HDV infection declined at the end of the 1990s. Even in Italy, an HDV-endemic area, public health measures introduced to control HBV infection resulted during the 1990s in a 1.5%/year reduction in the prevalence of HDV infection. Still, the frequency of HDV infection during the first decade of the twenty-first century has not fallen below levels reached during the 1990s; the reservoir has been sustained by survivors infected during 1970–1980 and recent immigrants from still-endemic to less-endemic countries.

HEPATITIS C

Routine screening of blood donors for HBsAg and the elimination of commercial blood sources in the early 1970s reduced the frequency of, but did not eliminate, transfusion-associated hepatitis. During the 1970s, the likelihood of acquiring hepatitis after transfusion of voluntarily donated, HBsAg-screened blood was ~ 10% per patient (up to 0.9% per unit transfused); 90–95% of these cases were classified, based on serologic exclusion of hepatitis A and B, as "non-A, non-B" hepatitis. For patients requiring transfusion of pooled products, such as clotting factor concentrates, the risk was even higher, up to 20–30%.

During the 1980s, voluntary self-exclusion of blood donors with risk factors for AIDS and then the introduction of donor screening for anti-HIV reduced further the likelihood of transfusion-associated hepatitis to <5%. During the late 1980s and early 1990s, the introduction first of "surrogate" screening tests for non-A, non-B hepatitis [alanine aminotransferase (ALT) and anti-HBc, both shown to identify blood donors with a higher likelihood of transmitting non-A, non-B hepatitis to recipients] and, subsequently, after the discovery of HCV, first-generation immunoassays for anti-HCV reduced the frequency of transfusion-associated hepatitis even further. A prospective analysis of transfusion-associated hepatitis conducted between 1986 and 1990 showed that the frequency of transfusion-associated hepatitis at one urban university hospital fell from a baseline of 3.8% per patient (0.45% per unit transfused) to 1.5% per patient (0.19% per unit) after the introduction of surrogate testing and to 0.6% per patient (0.03% per unit) after the introduction of first-generation anti-HCV assays. The introduction of second-generation anti-HCV assays reduced the frequency of transfusion-associated hepatitis C to almost imperceptible levels—1 in 100,000—and these gains were reinforced by the application of third-generation anti-HCV assays and of automated PCR testing of donated blood for HCV RNA, which has resulted in a reduction in the risk of transfusion-associated HCV infection to 1 in 2.3 million transfusions.

In addition to being transmitted by transfusion, hepatitis C can be transmitted by other percutaneous routes, such as injection drug use. In addition, this virus can be transmitted by occupational exposure to blood, and the likelihood of infection is increased in hemodialysis units. Although the frequency of transfusion-associated hepatitis C fell as a result of blood-donor screening, the overall frequency of hepatitis C remained the same until the early 1990s, when the overall frequency fell by 80%, in parallel with a reduction in the number of new cases in injection drug users. After the exclusion of anti-HCV-positive plasma units from the donor pool, rare, sporadic instances have occurred of hepatitis C among recipients of immunoglobulin (IG) preparations for intravenous (but not intramuscular) use.

Serologic evidence for HCV infection occurs in 90% of patients with a history of transfusion-associated hepatitis (almost all occurring before 1992, when second-generation HCV-screening tests were introduced); hemophiliacs and others treated with clotting factors; injection drug users; 60–70% of patients with sporadic "non-A, non-B" hepatitis who lack identifiable risk factors; 0.5% of volunteer blood donors; and, in the most recent survey conducted in the United States between 1999 and 2000, 1.6% of the general population in the United States, which translates into 4.1 million persons (3.2 million with viremia). Comparable frequencies of HCV infection occur in most countries around the world, with 170 million persons infected worldwide, but extraordinarily high prevalences of HCV infection occur in certain countries such as Egypt, where >20% of the population in some cities is infected. The high frequency in Egypt is attributable to contaminated equipment used for medical procedures and unsafe injection practices in the 1970s. In the United States, African Americans and Mexican Americans have higher frequencies of HCV infection than whites. Between 1988 and 1994, 30- to 40-year-old adult males had the highest prevalence of HCV infection; however, in a survey conducted between 1999 and 2000, the peak age decile had shifted to those age 40–49 years; an increase in hepatitis C–related mortality has paralleled this secular trend, increasing since 1995 predominantly in the 55- to 64-year age group. Thus, despite an 80% reduction in new HCV infections during the 1990s, the prevalence of HCV infection in the population was sustained by an aging cohort that had acquired their infections 2 to 3 decades earlier, during the 1960s and 1970s, as a result predominantly of self-inoculation with recreational drugs. Hepatitis C accounts for 40% of chronic liver disease, is the most frequent indication for liver transplantation, and is estimated to account for 8000–10,000 deaths per year in the United States.

The distribution of HCV genotypes varies in different parts of the world. Worldwide, genotype 1 is the most common. In the United States, genotype 1 accounts for 70% of HCV infections, while genotypes 2 and 3 account for the remaining 30%; among African Americans, the frequency of genotype 1 is even higher (i.e., 90%). Genotype 4 predominates in Egypt; genotype 5 is localized to South Africa, and genotype 6 to Hong Kong.

Most asymptomatic blood donors found to have anti-HCV and ~20–30% of persons with reported cases of acute hepatitis C do not fall into a recognized risk group; however, many such blood donors do recall risk-associated behaviors when questioned carefully.

As a bloodborne infection, HCV potentially can be transmitted sexually and perinatally; however, both of these modes of transmission are inefficient for hepatitis C. Although 10–15% of patients with acute hepatitis C report having potential sexual sources of infection, most studies have failed to identify sexual transmission of this agent. The chances of sexual and perinatal transmission have been estimated to be ~ 5%, well below comparable rates for HIV and HBV infections. Moreover, sexual transmission appears to be confined to such subgroups as persons with multiple sexual partners and sexually transmitted diseases; transmission of HCV infection is rare between stable, monogamous sexual partners. Breast-feeding does not increase the risk of HCV infection between an infected mother and her infant. Infection of health workers is not dramatically higher than among the general population; however, health workers are more likely to acquire HCV infection through accidental needle punctures, the efficiency of which is ~3%. Infection of household contacts is rare as well.

Other groups with an increased frequency of HCV infection include patients who require hemodialysis and organ transplantation, those who require transfusions in the setting of cancer chemotherapy, HIV-infected persons, and persons with unexplained serum aminotransferase elevations. In immunosuppressed individuals, levels of anti-HCV may be undetectable, and a diagnosis may require testing for HCV RNA. Although new acute cases of hepatitis C are rare, newly diagnosed cases are common among otherwise

healthy persons who experimented briefly with injection drugs, as noted above, 2 or 3 decades earlier. Such instances usually remain unrecognized for years, until unearthed by laboratory screening for routine medical examinations, insurance applications, and attempted blood donation. Populations groups for whom HCV-infection screening is recommended are listed in **Table 304-4**.

Table 304-4 High-Risk Populations for Which HCV-Infection Screening Is Recommended

Persons who have used injection drugs or those who have used illicit drugs by noninjection routes
Persons with HIV infection
Hemophiliacs treated with clotting factor concentrates prior to 1987
Hemodialysis patients
Persons with unexplained elevations of aminotransferase levels
Transfusion or transplantation recipients prior to July 1992
Children born to women with hepatitis C
Health care, public safety, and emergency medical personnel following needle injury or mucosal exposure to HCV-contaminated blood
Sexual partners of persons with hepatitis C infection

HEPATITIS E

This type of hepatitis, identified in India, Asia, Africa, the Middle East, and Central America, resembles hepatitis A in its primarily enteric mode of spread. The commonly recognized cases occur after contamination of water supplies such as after monsoon flooding, but sporadic, isolated cases occur. An epidemiologic feature that distinguishes HEV from other enteric agents is the rarity of secondary person-to-person spread from infected persons to their close contacts. Infections arise in populations that are immune to HAV and favor young adults. In endemic areas, the prevalence of antibodies to HEV is ≤40%. In nonendemic areas of the world, such as the United States, clinically apparent acute hepatitis E is extremely rare; however, the prevalence of antibodies to HEV can be as high as 20% in such areas. In nonendemic areas, HEV does not account for any of the sporadic "non-A, non-B" cases of hepatitis; however, cases imported from endemic areas have been found in the United States. Several reports suggest a zoonotic reservoir for HEV in swine.

Clinical and Laboratory Features

SYMPTOMS AND SIGNS

Acute viral hepatitis occurs after an incubation period that varies according to the responsible agent. Generally, incubation periods for hepatitis A range from 15–45 days (mean, 4 weeks), for hepatitis B and D from 30–180 days (mean, 8–12 weeks), for hepatitis C from 15–160 days (mean, 7 weeks), and for hepatitis E from 14–60 days (mean, 5–6 weeks). The *prodromal symptoms* of acute viral hepatitis are systemic and quite variable. Constitutional symptoms of anorexia, nausea and vomiting, fatigue, malaise, arthralgias, myalgias, headache, photophobia, pharyngitis, cough, and coryza may precede the onset of jaundice by 1–2 weeks. The nausea, vomiting, and anorexia are frequently associated with alterations in olfaction and taste. A low-grade fever between 38° and 39°C (100°–102°F) is more often present in hepatitis A and E than in hepatitis B or C, except when hepatitis B is heralded by a serum sickness–like syndrome; rarely, a fever of 39.5°–40°C (103°–104°F) may accompany the constitutional symptoms. Dark urine and clay-colored stools may be noticed by the patient from 1–5 days before the onset of clinical jaundice.

With the onset of *clinical jaundice*, the constitutional prodromal symptoms usually diminish, but in some patients mild weight loss (2.5–5 kg) is common and may continue during the entire icteric phase. The liver becomes enlarged and tender and may be associated with right upper quadrant pain and discomfort. Infrequently, patients present with a cholestatic picture, suggesting extrahepatic biliary obstruction. Splenomegaly and cervical adenopathy are present in 10–20% of patients with acute hepatitis. Rarely, a few spider angiomas appear during the icteric phase and disappear during convalescence. During the *recovery phase*, constitutional symptoms disappear, but usually some liver enlargement and abnormalities in liver biochemical tests are still evident. The duration of the posticteric phase is variable, ranging 2–12 weeks, and is usually more prolonged in acute hepatitis B and C. Complete clinical and biochemical recovery is to be expected 1–2 months after all cases of hepatitis A and E and 3–4 months after the onset of jaundice in three-quarters of uncomplicated, self-limited cases of hepatitis B and C (among healthy adults, acute hepatitis B is self-limited in 95–99% while hepatitis C is self-limited in only ~15%). In the remainder, biochemical recovery may be delayed. A substantial proportion of patients with viral hepatitis never become icteric.

Infection with HDV can occur in the presence of acute or chronic HBV infection; the duration of HBV infection determines the duration of HDV infection. When acute HDV and HBV infection occur simultaneously, clinical and biochemical features may be indistinguishable from those of HBV infection alone, although occasionally they are more severe. As opposed to patients with *acute* HBV infection, patients with *chronic* HBV infection can support HDV replication indefinitely. This can happen when acute

HDV infection occurs in the presence of a nonresolving acute HBV infection. More commonly, acute HDV infection becomes chronic when it is superimposed on an underlying chronic HBV infection. In such cases, the HDV superinfection appears as a clinical exacerbation or an episode resembling acute viral hepatitis in someone already chronically infected with HBV. Superinfection with HDV in a patient with chronic hepatitis B often leads to clinical deterioration (see below).

In addition to superinfections with other hepatitis agents, acute hepatitis-like clinical events in persons with chronic hepatitis B may accompany spontaneous HBeAg to anti-HBe seroconversion or spontaneous reactivation (i.e., reversion from nonreplicative to replicative infection). Such reactivations can occur as well in therapeutically immunosuppressed patients with chronic HBV infection when cytotoxic/immunosuppressive drugs are withdrawn; in these cases, restoration of immune competence is thought to allow resumption of previously checked cell-mediated immune cytolysis of HBV-infected hepatocytes. Occasionally, acute clinical exacerbations of chronic hepatitis B may represent the emergence of a precore mutant (see "Virology and Etiology"), and the subsequent course in such patients may be characterized by periodic exacerbations.

LABORATORY FEATURES

The serum aminotransferases aspartate aminotransferase (AST) and ALT (previously designated SGOT and SGPT) show a variable increase during the prodromal phase of acute viral hepatitis and precede the rise in bilirubin level (Figs. 304-2 and 304-4). The acute level of these enzymes, however, does not correlate well with the degree of liver cell damage. Peak levels vary from 400–4000 IU or more; these levels are usually reached at the time the patient is clinically icteric and diminish progressively during the recovery phase of acute hepatitis. The diagnosis of anicteric hepatitis is based on clinical features and on aminotransferase elevations.

Jaundice is usually visible in the sclera or skin when the serum bilirubin value is >43 μmol/L (2.5 mg/dL). When jaundice appears, the serum bilirubin typically rises to levels ranging from 85–340 μmol/L (5–20 mg/dL). The serum bilirubin may continue to rise despite falling serum aminotransferase levels. In most instances, the total bilirubin is equally divided between the conjugated and unconjugated fractions. Bilirubin levels >340 μmol/L (20 mg/dL) extending and persisting late into the course of viral hepatitis are more likely to be associated with severe disease. In certain patients with underlying hemolytic anemia, however, such as glucose-6-phosphate dehydrogenase deficiency and sickle cell anemia, a high serum bilirubin level is common, resulting from superimposed hemolysis. In such patients, bilirubin levels >513 μmol/L (30 mg/dL) have been observed and are not necessarily associated with a poor prognosis.

Neutropenia and lymphopenia are transient and are followed by a relative lymphocytosis. Atypical lymphocytes (varying between 2 and 20%) are common during the acute phase. Measurement of the prothrombin time (PT) is important in patients with acute viral hepatitis, for a prolonged value may reflect a severe hepatic synthetic defect, signify extensive hepatocellular necrosis, and indicate a worse prognosis. Occasionally, a prolonged PT may occur with only mild increases in the serum bilirubin and aminotransferase levels. Prolonged nausea and vomiting, inadequate carbohydrate intake, and poor hepatic glycogen reserves may contribute to hypoglycemia noted occasionally in patients with severe viral hepatitis. Serum alkaline phosphatase may be normal or only mildly elevated, while a fall in serum albumin is uncommon in uncomplicated acute viral hepatitis. In some patients, mild and transient steatorrhea has been noted as well as slight microscopic hematuria and minimal proteinuria.

A diffuse but mild elevation of the γ globulin fraction is common during acute viral hepatitis. Serum IgG and IgM levels are elevated in about one-third of patients during the acute phase of viral hepatitis, but the serum IgM level is elevated more characteristically during acute hepatitis A. During the acute phase of viral hepatitis, antibodies to smooth muscle and other cell constituents may be present, and low titers of rheumatoid factor, nuclear antibody, and heterophil antibody can also be found occasionally. In hepatitis C and D, antibodies to LKM may occur; however, the species of LKM antibodies in the two types of hepatitis are different from each other as well as from the LKM antibody species characteristic of autoimmune hepatitis type 2 **(Chap. 306)**. The autoantibodies in viral hepatitis are nonspecific and can also be associated with other viral and systemic diseases. In contrast, virus-specific antibodies, which appear during and after hepatitis virus infection, are serologic markers of diagnostic importance.

As described above, serologic tests are available with which to establish a diagnosis of hepatitis A, B, D, and C. Tests for fecal or serum HAV are not routinely available. Therefore, a diagnosis of hepatitis A is based on detection of IgM anti-HAV during acute illness (Fig. 304-2). Rheumatoid factor can give rise to false-positive results in this test.

A diagnosis of HBV infection can usually be made by detection of HBsAg in serum. Infrequently, levels of HBsAg are too low to be detected during acute HBV infection, even with contemporary, highly sensitive immunoassays. In such cases, the diagnosis can be established by the presence of IgM anti-HBc.

The titer of HBsAg bears little relation to the severity of clinical disease. Indeed, an inverse correlation exists between the serum concentration of HBsAg and the degree of liver cell damage. For example, titers are highest in immunosuppressed patients, lower in patients with chronic liver disease (but higher in mild chronic than in severe chronic hepatitis), and very low in patients with acute fulminant hepatitis. These observations suggest that, in hepatitis B, the degree of liver cell damage and the clinical course are related to variations in the patient's immune response to HBV rather than to the amount of circulating HBsAg. In immunocompetent persons, however, a correlation exists between markers of HBV *replication* and liver injury (see below).

Another serologic marker that may be of value in patients with hepatitis B is HBeAg. Its principal clinical usefulness is as an indicator of relative infectivity. Because HBeAg is invariably present during early acute hepatitis B, HBeAg testing is indicated primarily during follow-up of chronic infection.

In patients with hepatitis B surface antigenemia of unknown duration (e.g., blood donors found to be HBsAg-positive and referred to a physician for evaluation), testing for IgM anti-HBc may be useful to distinguish between acute or recent infection (IgM anti-HBc-positive) and chronic HBV infection (IgM anti-HBc-negative, IgG anti-HBc-positive). A false-positive test for IgM anti-HBc may be encountered in patients with high-titer rheumatoid factor.

Anti-HBs is rarely detectable in the presence of HBsAg in patients with *acute* hepatitis B, but 10–20% of persons with *chronic* HBV infection may harbor low-level anti-HBs. This antibody is directed not against the common group determinant, *a*, but against the heterotypic subtype determinant (e.g., HBsAg of subtype *ad* with anti-HBs of subtype *y*). In most cases, this serologic pattern cannot be attributed to infection with two different HBV subtypes, and the presence of this antibody is not a harbinger of imminent HBsAg clearance. When such antibody is detected, its presence is of no recognized clinical significance (see "Virology and Etiology").

After immunization with hepatitis B vaccine, which consists of HBsAg alone, anti-HBs is the only serologic marker to appear. The commonly encountered serologic patterns of hepatitis B and their interpretations are summarized in **Table 304-5**. Tests for the detection of HBV DNA in liver and serum are now available. Like HBeAg, serum HBV DNA is an indicator of HBV replication, but tests for HBV DNA are more sensitive and quantitative. First-generation hybridization assays for HBV DNA had a sensitivity of 10^5–10^6 virions/mL, a relative threshold below which infectivity and liver injury are limited and HBeAg is usually undetectable. Currently, testing for HBV DNA has shifted from insensitive hybridization assays to amplification assays (e.g., the PCR-based assay, which can detect as few as 10 or 100 virions/mL); among the commercially available PCR assays, the most useful are those with the highest sensitivity (5–10 IU/mL) and the largest dynamic range (10^0–10^9 IU/mL). With increased sensitivity, amplification assays remain reactive well below the threshold for infectivity and liver injury. These markers are useful in following the course of HBV replication in patients with chronic hepatitis B receiving antiviral chemotherapy (e.g., with interferon or nucleoside analogues) **(Chap. 306)**. In immunocompetent persons with chronic hepatitis B, a general correlation does appear to exist between the level of HBV replication, as reflected by the level of HBV DNA in serum, and the degree of liver injury. High-serum HBV DNA levels, increased expression of viral antigens, and necroinflammatory activity in the liver go hand in hand unless immunosuppression interferes with cytolytic T cell responses to virus-infected cells; reduction of HBV replication with antiviral drugs tends to be accompanied by an improvement in liver histology. Among patients with chronic hepatitis B, high levels of HBV DNA increase the risk of cirrhosis, hepatic decompensation, and hepatocellular carcinoma (see "Complications and Sequelae").

Table 304-5 Commonly Encountered Serologic Patterns of Hepatitis B Infection

HBsAg	Anti-HBs	Anti-HBc	HBeAg	Anti-HBe	Interpretation
+	–	IgM	+	–	Acute hepatitis B, high infectivity
+	–	IgG	+	–	Chronic hepatitis B, high infectivity
+	–	IgG	–	+	Late acute or chronic hepatitis B, low infectivityHBeAg–negative ("precore–mutant") hepatitis B (chronic or, rarely, acute)
+	+	+	+/–	+/–	1. HBsAg of one subtype and heterotypic anti–HBs (common) 2. Process of seroconversion from HBsAg to anti– HBs (rare)
–	–	IgM	+/–	+/–	1. Acute hepatitis B 2. Anti–HBc "window"
–	–	IgG	–	+/–	1. Low–level hepatitis B carrier 2. Hepatitis B in remote past
–	+	IgG	–	+/–	Recovery from hepatitis B
–	+	–	–	–	1. Immunization with HBsAg (after vaccination) 2. Hepatitis B in the remote past (?) 3. False–positive

In patients with hepatitis C, an episodic pattern of aminotransferase elevation is common. A specific serologic diagnosis of hepatitis C can be made by demonstrating the presence in serum of anti-HCV.

When contemporary immunoassays are used, anti-HCV can be detected in acute hepatitis C during the initial phase of elevated aminotransferase activity. This antibody may never become detectable in 5–10% of patients with acute hepatitis C, and levels of anti-HCV may become undetectable after recovery (albeit rare) from acute hepatitis C. In patients with chronic hepatitis C, anti-HCV is detectable in >95% of cases. Nonspecificity can confound immunoassays for anti-HCV, especially in persons with a low prior probability of infection, such as volunteer blood donors, or in persons with circulating rheumatoid factor, which can bind nonspecifically to assay reagents; testing for HCV RNA can be used in such settings to distinguish between true-positive and false-positive anti-HCV determinations. Assays for HCV RNA are the most sensitive tests for HCV infection and represent the "gold standard" in establishing a diagnosis of hepatitis C. HCV RNA can be detected even before acute elevation of aminotransferase activity and before the appearance of anti-HCV in patients with acute hepatitis C. In addition, HCV RNA remains detectable indefinitely, continuously in most but intermittently in some, in patients with chronic hepatitis C (detectable as well in some persons with normal liver tests (i.e., inactive carriers). In the small minority of patients with hepatitis C who lack anti-HCV, a diagnosis can be supported by detection of HCV RNA. If all these tests are negative and the patient has a well-characterized case of hepatitis after percutaneous exposure to blood or blood products, a diagnosis of hepatitis caused by an unidentified agent can be entertained.

Amplification techniques are required to detect HCV RNA, and two types are available. One is a branched-chain complementary DNA (bDNA) assay, in which the detection signal (a colorimetrically detectable enzyme bound to a complementary DNA probe) is amplified. The other involves target amplification (i.e., synthesis of multiple copies of the viral genome). This can be done by PCR or TMA, in which the viral RNA is reverse transcribed to complementary DNA and then amplified by repeated cycles of DNA synthesis. Both can be used as quantitative assays and a measurement of relative "viral load"; PCR and TMA, with a sensitivity of 10–10^2 IU/mL, are more sensitive than bDNA, with a sensitivity of 10^3 IU/mL; assays are available with a wide dynamic range (10–10^7 IU/mL). Determination of HCV RNA level is not a reliable marker of disease severity or prognosis but is helpful in predicting relative responsiveness to antiviral therapy. The same is true for determinations of HCV genotype **(Chap. 306)**.

A proportion of patients with hepatitis C have isolated anti-HBc in their blood, a reflection of a common risk in certain populations of exposure to multiple bloodborne hepatitis agents. The anti-HBc in such cases is almost invariably of the IgG class and usually represents HBV infection in the remote past (HBV DNA undetectable), rarely current HBV infection with low-level virus carriage.

The presence of HDV infection can be identified by demonstrating intrahepatic HDV antigen or, more practically, an anti-HDV seroconversion (a rise in titer of anti-HDV or de novo appearance of anti-HDV). Circulating HDV antigen, also diagnostic of acute infection, is detectable only briefly, if at all. Because anti-HDV is often undetectable once HBsAg disappears, retrospective serodiagnosis of acute self-limited, simultaneous HBV and HDV infection is difficult. Early diagnosis of acute infection may be hampered by a delay of up to 30–40 days in the appearance of anti-HDV.

When a patient presents with acute hepatitis and has HBsAg and anti-HDV in serum, determination of the class of anti-HBc is helpful in establishing the relationship between infection with HBV and HDV. Although IgM anti-HBc does not distinguish *absolutely* between acute and chronic HBV infection, its presence is a reliable indicator of recent infection and its absence a reliable indicator of infection in the remote past. In simultaneous acute HBV and HDV infections, IgM anti-HBc will be detectable, while in acute HDV infection superimposed on chronic HBV infection, anti-HBc will be of the IgG class.

Tests for the presence of HDV RNA are useful for determining the presence of ongoing HDV replication and relative infectivity. Diagnostic tests for hepatitis E are commercially available in several countries outside the United States; in the United States, diagnostic assays can be performed at the Centers for Disease Control and Prevention.

Liver biopsy is rarely necessary or indicated in acute viral hepatitis, except when the diagnosis is questionable or when clinical evidence suggests a diagnosis of chronic hepatitis.

A diagnostic algorithm can be applied in the evaluation of cases of acute viral hepatitis. A patient with acute hepatitis should undergo four serologic tests, HBsAg, IgM anti-HAV, IgM anti-HBc, and anti-HCV **(Table 304-6)**. The presence of HBsAg, with or without IgM anti-HBc, represents HBV infection. If IgM anti-HBc is present, the HBV infection is considered acute; if IgM anti-HBc is absent, the HBV infection is considered chronic. A diagnosis of acute hepatitis B can be made in the absence of HBsAg when IgM anti-HBc is detectable. A diagnosis of acute hepatitis A is based on the presence of IgM anti-HAV. If IgM anti-HAV coexists with HBsAg, a diagnosis of simultaneous HAV and HBV infections can be made; if IgM anti-HBc (with or without HBsAg) is detectable, the patient has simultaneous acute hepatitis A and B, and if IgM anti-HBc is undetectable, the patient has acute hepatitis A superimposed on chronic HBV infection. The presence of anti-HCV supports a diagnosis of acute hepatitis C. Occasionally, testing for HCV RNA or repeat anti-HCV testing later during the illness is necessary to establish the diagnosis. Absence of all serologic markers is consistent with a diagnosis of "non-A, non-B, non-C" hepatitis, if the epidemiologic setting is appropriate.

Table 304-6 Simplified Diagnostic Approach in Patients Presenting with Acute Hepatitis

	Serologic Tests of Patient's Serum			

HBsAg	IgM Anti-HAV	IgM Anti-HBc	Anti-HCV	Diagnostic Interpretation
+	–	+	–	Acute hepatitis B
+	–	–	–	Chronic hepatitis B
+	+	–	–	Acute hepatitis A superimposed on chronic hepatitis B
+	+	+	–	Acute hepatitis A and B
–	+	–	–	Acute hepatitis A
–	+	+	–	Acute hepatitis A and B (HBsAg below detection threshold)
–	–	+	–	Acute hepatitis B (HBsAg below detection threshold)
–	–	–	+	Acute hepatitis C

In patients with chronic hepatitis, initial testing should consist of HBsAg and anti-HCV. Anti-HCV supports and HCV RNA testing establishes the diagnosis of chronic hepatitis C. If a serologic diagnosis of chronic hepatitis B is made, testing for HBeAg and anti-HBe is indicated to evaluate relative infectivity. Testing for HBV DNA in such patients provides a more quantitative and sensitive measure of the level of virus replication and, therefore, is very helpful during antiviral therapy **(Chap. 306)**. In patients with chronic hepatitis B and normal aminotransferase activity in the absence of HBeAg, serial testing over time is often required to distinguish between inactive carriage and HBeAg-negative chronic hepatitis B with fluctuating virologic and necroinflammatory activity. In persons with hepatitis B, testing for anti-HDV is useful in those with severe and fulminant disease, with severe chronic disease, with chronic hepatitis B and acute hepatitis-like exacerbations, with frequent percutaneous exposures, and from areas where HDV infection is endemic.

Prognosis

Virtually all previously healthy patients with hepatitis A recover completely with no clinical sequelae. Similarly, in acute hepatitis B, 95–99% of previously healthy adults have a favorable course and recover completely. Certain clinical and laboratory features, however, suggest a more complicated and protracted course. Patients of advanced age and with serious underlying medical disorders may have a prolonged course and are more likely to experience severe hepatitis. Initial presenting features such as ascites, peripheral edema, and symptoms of hepatic encephalopathy suggest a poorer prognosis. In addition, a prolonged PT, low-serum albumin level, hypoglycemia, and very high-serum bilirubin values suggest severe hepatocellular disease. Patients with these clinical and laboratory features deserve prompt hospital admission. The case fatality rate in hepatitis A and B is very low (~0.1%) but is increased by advanced age and underlying debilitating disorders. Among patients ill enough to be hospitalized for acute hepatitis B, the fatality rate is 1%. Hepatitis C is less severe during the acute phase than hepatitis B and is more likely to be anicteric; fatalities are rare, but the precise case fatality rate is not known. In outbreaks of waterborne hepatitis E in India and Asia, the case fatality rate is 1–2% and up to 10–20% in pregnant women. Patients with simultaneous acute hepatitis B and hepatitis D do not necessarily experience a higher mortality rate than do patients with acute hepatitis B alone; however, in several recent outbreaks of acute simultaneous HBV and HDV infection among injection drug users, the case fatality rate has been ~5%. In the case of HDV superinfection of a person with chronic hepatitis B, the likelihood of fulminant hepatitis and death is increased substantially. Although the case fatality rate for hepatitis D has not been defined adequately, in outbreaks of severe HDV superinfection in isolated populations with a high hepatitis B carrier rate, the mortality rate has been recorded in excess of 20%.

Complications and Sequelae

A small proportion of patients with hepatitis A experience *relapsing hepatitis* weeks to months after apparent recovery from acute hepatitis. Relapses are characterized by recurrence of symptoms, aminotransferase elevations, occasionally jaundice, and fecal excretion of HAV. Another unusual variant of acute hepatitis A is *cholestatic hepatitis*, characterized by protracted cholestatic jaundice and pruritus. Rarely, liver test abnormalities persist for many months, even up to a year. Even when these complications occur, hepatitis A remains self-limited and does not progress to chronic liver disease. During the prodromal phase of acute hepatitis B, a serum sickness-like syndrome characterized by arthralgia or arthritis, rash, angioedema, and rarely, hematuria and proteinuria may develop in 5–10% of patients. This syndrome occurs before the onset of clinical jaundice, and these patients are often diagnosed erroneously as having rheumatologic diseases. The diagnosis can be established by measuring serum aminotransferase levels, which are almost invariably elevated, and serum HBsAg. As noted above, EMC is an immune-complex disease that can complicate chronic hepatitis C and is part of a spectrum of B cell lymphoproliferative disorders, which, in rare instances, can evolve to B cell lymphoma **(Chap. 110)**. Attention has been drawn as well to associations between hepatitis C and such cutaneous disorders as porphyria cutanea tarda and lichen planus. A mechanism for these associations is unknown. Finally, related to the reliance of HCV on lipoprotein secretion and assembly pathways and on interactions of HCV with glucose metabolism, HCV

infection may be complicated by hepatic steatosis, hypercholesterolemia, insulin resistance (and other manifestations of the metabolic syndrome), and type 2 diabetes mellitus; both hepatic steatosis and insulin resistance appear to accelerate hepatic fibrosis and blunt responsiveness to antiviral therapy **(Chap. 306)**.

The most feared complication of viral hepatitis is *fulminant hepatitis* (massive hepatic necrosis); fortunately, this is a rare event. Fulminant hepatitis is primarily seen in hepatitis B and D, as well as hepatitis E, but rare fulminant cases of hepatitis A occur primarily in older adults and in persons with underlying chronic liver disease, including, according to some reports, chronic hepatitis B and C. Hepatitis B accounts for >50% of fulminant cases of viral hepatitis, a sizable proportion of which are associated with HDV infection and another proportion with underlying chronic hepatitis C. Fulminant hepatitis is hardly ever seen in hepatitis C, but hepatitis E, as noted above, can be complicated by fatal fulminant hepatitis in 1–2% of all cases and in up to 20% of cases in pregnant women. Patients usually present with signs and symptoms of encephalopathy that may evolve to deep coma. The liver is usually small and the PT excessively prolonged. The combination of rapidly shrinking liver size, rapidly rising bilirubin level, and marked prolongation of the PT, even as aminotransferase levels fall, together with clinical signs of confusion, disorientation, somnolence, ascites, and edema, indicates that the patient has hepatic failure with encephalopathy. Cerebral edema is common; brainstem compression, gastrointestinal bleeding, sepsis, respiratory failure, cardiovascular collapse, and renal failure are terminal events. The mortality rate is exceedingly high (>80% in patients with deep coma), but patients who survive may have a complete biochemical and histologic recovery. If a donor liver can be located in time, liver transplantation may be life-saving in patients with fulminant hepatitis **(Chap. 310)**.

Documenting the disappearance of HBsAg after apparent clinical recovery from acute hepatitis B is particularly important. Before laboratory methods were available to distinguish between acute hepatitis and acute hepatitis-like exacerbations (*spontaneous reactivations*) of chronic hepatitis B, observations suggested that ~10% of previously healthy patients remained HBsAg-positive for >6 months after the onset of clinically apparent acute hepatitis B. One-half of these persons cleared the antigen from their circulations during the next several years, but the other 5% remained chronically HBsAg-positive. More recent observations suggest that the true rate of chronic infection after clinically apparent acute hepatitis B is as low as 1% in normal, immunocompetent, young adults. Earlier, higher estimates may have been confounded by inadvertent inclusion of acute exacerbations in chronically infected patients; these patients, chronically HBsAg-positive before exacerbation, were unlikely to seroconvert to HBsAg-negative thereafter. Whether the rate of chronicity is 10% or 1%, such patients have anti-HBc in serum; anti-HBs is either undetected or detected at low titer against the opposite subtype specificity of the antigen (see "Laboratory Features"). These patients may (1) be inactive carriers; (2) have low-grade, mild chronic hepatitis; or (3) have moderate to severe chronic hepatitis with or without cirrhosis. The likelihood of remaining chronically infected after acute HBV infection is especially high among neonates, persons with Down's syndrome, chronically hemodialyzed patients, and immunosuppressed patients, including persons with HIV infection.

Chronic hepatitis is an important late complication of acute hepatitis B occurring in a small proportion of patients with acute disease but more common in those who present with chronic infection without having experienced an acute illness, as occurs typically after neonatal infection or after infection in an immunosuppressed host **(Chap. 306)**. Certain clinical and laboratory features suggest progression of acute hepatitis to chronic hepatitis: (1) lack of complete resolution of clinical symptoms of anorexia, weight loss, fatigue, and the persistence of hepatomegaly; (2) the presence of bridging/interface or multilobular hepatic necrosis on liver biopsy during protracted, severe acute viral hepatitis; (3) failure of the serum aminotransferase, bilirubin, and globulin levels to return to normal within 6–12 months after the acute illness; and (4) the persistence of HBeAg for >3 months or HBsAg for >6 months after acute hepatitis.

Although acute hepatitis D infection does not increase the likelihood of chronicity of simultaneous acute hepatitis B, hepatitis D has the potential for contributing to the severity of chronic hepatitis B. Hepatitis D superinfection can transform inactive or mild chronic hepatitis B into severe, progressive chronic hepatitis and cirrhosis; it also can accelerate the course of chronic hepatitis B. Some HDV superinfections in patients with chronic hepatitis B lead to fulminant hepatitis. As defined in longitudinal studies over 3 decades, the annual rates of cirrhosis and hepatocellular carcinoma in patients with chronic hepatitis D are 4% and 2.8%, respectively. Although HDV and HBV infections are associated with severe liver disease, mild hepatitis and even inactive carriage have been identified in some patients, and the disease may become indolent beyond the early years of infection.

After acute HCV infection, the likelihood of remaining chronically *infected* approaches 85–90%. Although many patients with chronic hepatitis C have no symptoms, cirrhosis may develop in as many as 20% within 10–20 years of acute illness; in some series of cases reported by referral centers, cirrhosis has been reported in as many as 50% of patients with chronic hepatitis C. Although chronic hepatitis C accounts for at least 40% of cases of chronic liver disease and of patients undergoing liver transplantation for end-stage liver disease in the United States and Europe, in the majority of patients with chronic hepatitis C, morbidity and mortality are limited during the initial 20 years after the onset of infection. Progression of chronic hepatitis C may be influenced by age of acquisition, duration of infection, immunosuppression, coexisting excessive alcohol use, concomitant hepatic steatosis, other hepatitis virus infection, or HIV co-infection. In fact, instances of severe and rapidly progressive chronic hepatitis B and C are being recognized with increasing frequency in patients with HIV infection **(Chap. 189)**. In contrast, neither HAV nor HEV causes chronic liver disease.

Rare complications of viral hepatitis include pancreatitis, myocarditis, atypical pneumonia, aplastic anemia, transverse myelitis, and peripheral neuropathy. Persons with chronic hepatitis B, particularly those infected in infancy or early childhood and especially those with HBeAg and/or high-level HBV DNA, have an enhanced risk of hepatocellular carcinoma. The risk of hepatocellular carcinoma is increased as well in patients with chronic hepatitis C, almost exclusively in patients with cirrhosis, and almost always after at least several decades, usually after 3 decades of disease **(Chap. 92)**. In children, hepatitis B may present rarely with anicteric hepatitis, a nonpruritic papular rash of the face, buttocks, and limbs, and lymphadenopathy (papular acrodermatitis of childhood or Gianotti-Crosti syndrome).

Rarely, autoimmune hepatitis **(Chap. 306)** can be triggered by a bout of otherwise self-limited acute hepatitis, as reported after acute hepatitis A, B, and C.

Differential Diagnosis

Viral diseases such as infectious mononucleosis; those due to cytomegalovirus, herpes simplex, and coxsackieviruses; and toxoplasmosis may share certain clinical features with viral hepatitis and cause elevations in serum aminotransferase and, less commonly, in serum bilirubin levels. Tests such as the differential heterophile and serologic tests for these agents may be helpful in the differential diagnosis if HBsAg, anti-HBc, IgM anti-HAV, and anti-HCV determinations are negative. Aminotransferase elevations can accompany almost any systemic viral infection; other rare causes of liver injury confused with viral hepatitis are infections with *Leptospira*, *Candida*, *Brucella*, *Mycobacteria*, and *Pneumocystis*. A complete drug history is particularly important, for many drugs and certain anesthetic agents can produce a picture of either acute hepatitis or cholestasis **(Chap. 305)**. Equally important is a past history of unexplained "repeated episodes" of acute hepatitis. This history should alert the physician to the possibility that the underlying disorder is chronic hepatitis. Alcoholic hepatitis must also be considered, but usually the serum aminotransferase levels are not as markedly elevated and other stigmata of alcoholism may be present. The finding on liver biopsy of fatty infiltration, a neutrophilic inflammatory reaction, and "alcoholic hyaline" would be consistent with alcohol-induced rather than viral liver injury. Because acute hepatitis may present with right upper quadrant abdominal pain, nausea and vomiting, fever, and icterus, it is often confused with acute cholecystitis, common duct stone, or ascending cholangitis. Patients with acute viral hepatitis may tolerate surgery poorly; therefore, it is important to exclude this diagnosis, and in confusing cases, a percutaneous liver biopsy may be necessary before laparotomy. Viral hepatitis in the elderly is often misdiagnosed as obstructive jaundice resulting from a common duct stone or carcinoma of the pancreas. Because acute hepatitis in the elderly may be quite severe and the operative mortality high, a thorough evaluation including biochemical tests, radiographic studies of the biliary tree, and even liver biopsy may be necessary to exclude primary parenchymal liver disease. Another clinical constellation that may mimic acute hepatitis is right ventricular failure with passive hepatic congestion or hypoperfusion syndromes, such as those associated with shock, severe hypotension, and severe left ventricular failure. Also included in this general category is any disorder that interferes with venous return to the heart, such as right atrial myxoma, constrictive pericarditis, hepatic vein occlusion (Budd-Chiari syndrome), or veno-occlusive disease. Clinical features are usually sufficient to distinguish among these vascular disorders and viral hepatitis. Acute fatty liver of pregnancy, cholestasis of pregnancy, eclampsia, and the HELLP (*h*emolysis, *e*levated *l*iver tests, and *l*ow *p*latelets) syndrome can be confused with viral hepatitis during pregnancy. Very rarely, malignancies metastatic to the liver can mimic acute or even fulminant viral hepatitis. Occasionally, genetic or metabolic liver disorders (e.g., Wilson's disease, α_1-antitrypsin deficiency) as well as nonalcoholic fatty liver disease are confused with viral hepatitis.

Treatment: Acute Viral Hepatitis

In hepatitis B, among previously healthy adults who present with clinically apparent acute hepatitis, recovery occurs in ~99%; therefore, antiviral therapy is not likely to improve the rate of recovery and is not required. In rare instances of severe acute hepatitis B, treatment with a nucleoside analogue at oral doses used to treat chronic hepatitis B **(Chap. 306)** has been attempted successfully. Although clinical trials have not been done to establish the efficacy of this approach; although severe acute hepatitis B is not an approved indication for therapy; and although the duration of therapy has not been determined; nonetheless, most authorities would recommend institution of antiviral therapy with a nucleoside analogue for severe, but not mild–moderate, acute hepatitis B. In typical cases of acute hepatitis C, recovery is rare, progression to chronic hepatitis is the rule, and meta-analyses of small clinical trials suggest that antiviral therapy with interferon alfa monotherapy (3 million units SC three times a week) is beneficial, reducing the rate of chronicity considerably by inducing sustained responses in 30–70% of patients. In a German multicenter study of 44 patients with acute symptomatic hepatitis C, initiation of intensive interferon alfa therapy (5 million units SC daily for 4 weeks, then three times a week for another 20 weeks) within an average of 3 months after infection resulted in a sustained virologic response rate of 98%. Although treatment of acute hepatitis C is recommended, the optimum regimen, duration of therapy, and time to initiate therapy remain to be determined. Many authorities now opt for a 24-week course (beginning within 2–3 months after onset) of the best regimen identified for the treatment of chronic hepatitis C, long-acting pegylated interferon plus the nucleoside analogue ribavirin, although the value of adding ribavirin has not been demonstrated **(**see Chap. 306**for doses**). Because of the marked reduction over the past 2 decades in the frequency of acute hepatitis C, opportunities to identify and treat patients with acute hepatitis C are rare, except in injection drug users. Hospital epidemiologists, however, will encounter health workers who sustain hepatitis C-contaminated needle sticks; when monitoring for ALT elevations and HCV, RNA after these accidents identifies acute hepatitis C (risk only ~3%), therapy should be initiated.

Notwithstanding these specific therapeutic considerations, in most cases of typical acute viral hepatitis, specific treatment generally is not necessary. Although hospitalization may be required for clinically severe illness, most patients do not require hospital care. Forced and prolonged bed rest is not essential for full recovery, but many patients will feel better with restricted physical activity. A high-calorie diet is desirable, and because many patients may experience nausea late in the day, the major caloric intake is best tolerated in the morning. Intravenous feeding is necessary in the acute stage if the patient has persistent vomiting and cannot maintain oral intake. Drugs capable of producing adverse reactions such as cholestasis and drugs metabolized by the liver should be avoided. If severe pruritus is present, the use of the bile salt-sequestering resin cholestyramine is helpful. Glucocorticoid therapy has no value in acute viral hepatitis, even in severe cases associated with *bridging necrosis*, and may be deleterious, even increasing the risk of chronicity (e.g., of acute hepatitis B).

Physical isolation of patients with hepatitis to a single room and bathroom is rarely necessary except in the case of fecal incontinence for hepatitis A and E or uncontrolled, voluminous bleeding for hepatitis B (with or without concomitant hepatitis D) and hepatitis C. Because most patients hospitalized with hepatitis A excrete little, if any, HAV, the likelihood of HAV transmission from these patients during their hospitalization is low. Therefore, burdensome *enteric precautions are no longer recommended* . Although gloves should be worn when the bedpans or fecal material of patients with hepatitis A are handled, these precautions do not represent a departure from sensible procedure and contemporary universal precautions for all hospitalized patients. For patients with hepatitis B and hepatitis C, emphasis should be placed on blood precautions (i.e., avoiding direct, ungloved hand contact with blood and other body fluids). Enteric precautions are unnecessary. The importance of simple hygienic precautions such as hand washing cannot be overemphasized. Universal precautions that have been adopted for all patients apply to patients with viral hepatitis.

Hospitalized patients may be discharged following substantial symptomatic improvement, a significant downward trend in the serum aminotransferase and bilirubin values, and a return to normal of the PT. Mild aminotransferase elevations should not be considered contraindications to the gradual resumption of normal activity.

In *fulminant hepatitis*, the goal of therapy is to support the patient by maintenance of fluid balance, support of circulation and respiration, control of bleeding, correction of hypoglycemia, and treatment of other complications of the comatose state in anticipation of liver regeneration and repair. Protein intake should be restricted, and oral lactulose or neomycin administered. Glucocorticoid therapy has been shown in controlled trials to be ineffective. Likewise, exchange transfusion, plasmapheresis, human cross-circulation, porcine liver cross-perfusion, hemoperfusion, and extracorporeal liver-assist devices have not been proven to enhance survival. Meticulous intensive care that includes prophylactic antibiotic coverage is the one factor that does appear to improve survival. Orthotopic liver transplantation is resorted to with increasing frequency, with excellent results, in patients with fulminant hepatitis **(Chap. 310)**.

Prophylaxis

Because application of therapy for acute viral hepatitis is limited and because antiviral therapy for chronic viral hepatitis is cumbersome and costly but effective in only a proportion of patients **(Chap. 306)**, emphasis is placed on prevention through immunization. The prophylactic approach differs for each of the types of viral hepatitis. In the past, immunoprophylaxis relied exclusively on passive immunization with antibody-containing globulin preparations purified by cold ethanol fractionation from the plasma of hundreds of normal donors. Currently, for hepatitis A and B, active immunization with vaccines is the preferable approach to prevention.

HEPATITIS A

Both passive immunization with IG and active immunization with killed vaccines are available. All preparations of IG contain anti-HAV concentrations sufficient to be protective. When administered before exposure or during the early incubation period, IG is effective in preventing clinically apparent hepatitis A. For postexposure prophylaxis of intimate contacts (household, sexual, institutional) of persons with hepatitis A, the administration of 0.02 mL/kg is recommended as early after exposure as possible; it may be effective even when administered as late as 2 weeks after exposure. Prophylaxis is not necessary for those who have already received hepatitis A vaccine, casual contacts (office, factory, school, or hospital), for most elderly persons, who are very likely to be immune, or for those known to have anti-HAV in their serum. In day-care centers, recognition of hepatitis A in children or staff should provide a stimulus for immunoprophylaxis in the center and in the children's family members. By the time most common-source outbreaks of hepatitis A are recognized, it is usually too late in the incubation period for IG to be effective; however, prophylaxis may limit the frequency of secondary cases. For travelers to tropical countries, developing countries, and other areas outside standard tourist routes, IG prophylaxis had been recommended before a vaccine became available. When such travel lasted <3 months, 0.02 mL/kg was given; for longer travel or residence in these areas, a dose of 0.06 mL/kg every 4–6 months was recommended. Administration of plasma-derived globulin is safe; all contemporary lots of IG are subjected to viral inactivation steps and must be free of HCV RNA as determined by PCR testing. Administration of IM lots of IG has not been associated with transmission of HBV, HCV, or HIV.

Formalin-inactivated vaccines made from strains of HAV attenuated in tissue culture have been shown to be safe, immunogenic, and effective in preventing hepatitis A. Hepatitis A vaccines are approved for use in persons who are at least one year old and appear to provide adequate protection beginning 4 weeks after

a primary inoculation. If it can be given within 4 weeks of an expected exposure, such as by travel to an endemic area, hepatitis A vaccine is the preferred approach to *pre-exposure* immunoprophylaxis. If travel is more imminent, IG (0.02 mL/kg) should be administered at a different injection site, along with the first dose of vaccine. Because vaccination provides long-lasting protection (protective levels of anti-HAV should last 20 years after vaccination), persons whose risk will be sustained (e.g., frequent travelers or those remaining in endemic areas for prolonged periods) should be vaccinated, and vaccine should supplant the need for repeated IG injections. Shortly after its introduction, hepatitis A vaccine was recommended for children living in communities with a high incidence of HAV infection; in 1999, this recommendation was extended to include all children living in states, counties, and communities with high rates of HAV infection. As of 2006, the Advisory Committee on Immunization Practices of the U.S. Public Health Service recommended *routine hepatitis A vaccination of all children* . Other groups considered to be at increased risk for HAV infection and who are candidates for hepatitis A vaccination include military personnel, populations with cyclic outbreaks of hepatitis A (e.g., Alaskan natives), employees of day-care centers, primate handlers, laboratory workers exposed to hepatitis A or fecal specimens, and patients with chronic liver disease. Because of an increased risk of fulminant hepatitis A—observed in some experiences but not confirmed in others—among patients with chronic hepatitis C, patients with chronic hepatitis C are candidates for hepatitis A vaccination, as are persons with chronic hepatitis B. Other populations whose recognized risk of hepatitis A is increased should be vaccinated, including men who have sex with men, injection drug users, persons with clotting disorders who require frequent administration of clotting-factor concentrates, persons traveling from the United States to countries with high or intermediate hepatitis A endemicity, postexposure prophylaxis for contacts of persons with hepatitis A, and household members and other close contacts of adopted children arriving from countries with high and moderate hepatitis A endemicity. Recommendations for dose and frequency differ for the two approved vaccine preparations **(Table 304-7)**; all injections are IM. Hepatitis A vaccine has been reported to be effective in preventing secondary household cases of acute hepatitis A, but its role in other instances of postexposure prophylaxis remains to be demonstrated. In the United States, reported mortality resulting from hepatitis A declined in parallel with hepatitis A vaccine-associated reductions in the annual incidence of new infections.

Table 304-7 Hepatitis A Vaccination Schedules

Age, years	No. of Doses	Dose	Schedule, months
HAVRIX (GlaxoSmithKline)[a]			
1–18	2	720 ELU[b] (0.5 mL)	0, 6–12
≥19	2	1440 ELU (1 mL)	0, 6–12
VAQTA (Merck)			
1–18	2	25 units (0.5 mL)	0, 6–18
≥19	2	50 units (1 mL)	0, 6–18

[a] A combination of this hepatitis A vaccine and hepatitis B vaccine, TWINRIX, is licensed for simultaneous protection against both of these viruses among adults (age ≥ 18 years). Each 1-mL dose contains 720 ELU of hepatitis A vaccine and 20 μg of hepatitis B vaccine. These doses are recommended at months 0, 1, and 6.

[b] Enzyme-linked immunoassay units.

Abbreviation: ELU, enzyme-linked immunoassay unit.

HEPATITIS B

Until 1982, prevention of hepatitis B was based on *passive* immunoprophylaxis either with standard IG, containing modest levels of anti-HBs, or hepatitis B immunoglobulin (HBIG), containing high-titer anti-HBs. The efficacy of standard IG has never been established and remains questionable; even the efficacy of HBIG, demonstrated in several clinical trials, has been challenged, and its contribution appears to be in reducing the frequency of clinical *illness*, not in preventing *infection*. The first vaccine for *active* immunization, introduced in 1982, was prepared from purified, noninfectious 22-nm spherical forms of HBsAg derived from the plasma of healthy HBsAg carriers. In 1987, the plasma-derived vaccine was supplanted by a genetically engineered vaccine derived from recombinant yeast. The latter vaccine consists of HBsAg particles that are nonglycosylated but are otherwise indistinguishable from natural HBsAg; two recombinant vaccines are licensed for use in the United States. Current recommendations can be divided into those for pre-exposure and postexposure prophylaxis.

For *pre-exposure* prophylaxis against hepatitis B in settings of frequent exposure (health workers exposed to blood; hemodialysis patients and staff; residents and staff of custodial institutions for the developmentally handicapped; injection drug users; inmates of long-term correctional facilities; persons with multiple sexual partners; persons such as hemophiliacs who require long-term, high-volume therapy with blood derivatives; household and sexual contacts of HBsAg carriers; persons living in or traveling

extensively in endemic areas; unvaccinated children under the age of 18; and unvaccinated children who are Alaskan natives, Pacific Islanders, or residents in households of first-generation immigrants from endemic countries), three IM (deltoid, not gluteal) injections of hepatitis B vaccine are recommended at 0, 1, and 6 months (other, optional schedules are summarized in Table 304-8). Pregnancy is *not* a contraindication to vaccination. In areas of low HBV endemicity such as the United States, despite the availability of safe and effective hepatitis B vaccines, a strategy of vaccinating persons in high-risk groups has not been effective. The incidence of new hepatitis B cases continued to increase in the United States after the introduction of vaccines; <10% of all targeted persons in high-risk groups have actually been vaccinated, and ~30% of persons with sporadic acute hepatitis B do not fall into any high-risk-group category. Therefore, to have an impact on the frequency of HBV infection in an area of low endemicity such as the United States, universal hepatitis B vaccination in childhood has been recommended. For unvaccinated children born after the implementation of universal infant vaccination, vaccination during early adolescence, at age 11–12 years, was recommended, and this recommendation has been extended to include all unvaccinated children age 0–19 years. In HBV-hyperendemic areas (e.g., Asia), universal vaccination of children has resulted in a marked 10- to 15-year decline in hepatitis B and its complications, including hepatocellular carcinoma.

Table 304-8 Preexposure Hepatitis B Vaccination Schedules

Target Group	No. of Doses	Dose	Schedule, months
RECOMBIVAX-HB (Merck) [a]			
Infants, children (<1–10 years)	3	5 μg (0.5 mL)	0, 1–2, 4–6
Adolescents (11–19 years)	3 or 4	5 μg (0.5 mL)	0–2, 1–4, 4–6 *or* 0, 12, 24 *or* 0, 1, 2, 12
	or		
	2	10 μg (1 mL)	0, 4–6 (age 11–15)
Adults (≥20 years)	3	10 μg (1 mL)	0–2, 1–4, 4–6
Hemodialysis patients[b]			
<20 years	3	5 μg (0.5 mL)	0, 1, 6
≥20 years	3	40 μg (4 mL)	0, 1, 6
ENGERIX-B (GlaxoSmithKline)[c]			
Infants, children (<1–10 years)	3 or 4	10 μg (0.5 mL)	0, 1–2, 4–6 *or* 0, 1, 2,12
Adolescents (10–19 years)	3 or 4	10 μg (0.5 mL)	0, 1–2, 4–6 *or* 0, 12, 24 *or* 0, 1, 2, 12
Adults (≥20 years)	3 or 4	20 μg (1 mL)	0–2, 1–4, 4–60, 1, 2, 12
Hemodialysis patients[b]			
<20 years	4	10 μg (0.5 mL)	0, 1, 2, 6
≥20 years	4	40 μg (2 mL)	0, 1, 2, 6

[a] This manufacturer produces a licensed combination of hepatitis B vaccine and vaccines against *Haemophilus influenzae* type b and *Neisseria meningitides*, Comvax, for use in infants and young children. Please consult product insert for dose and schedule.

[b] This group also includes other immunocompromised persons.

[c] This manufacturer produces two licensed combination hepatitis B vaccines: (1) Twinrix, recombinant hepatitis B vaccine plus inactivated hepatitis A vaccine, is licensed for simultaneous protection against both of these viruses among adults (age ≥ 18 years). Each 1-mL dose contains 720 ELU of hepatitis A vaccine and 20 μg of hepatitis B vaccine. These doses are recommended at months 0, 1, and 6. (2) Pediatrix, recombinant hepatitis B vaccine plus diphtheria and tetanus toxoid, pertussis, and inactivated poliovirus, is licensed for use in infants and young children. Please consult product insert for doses and schedules.

The two available recombinant hepatitis B vaccines are comparable, one containing 10 μg of HBsAg (Recombivax-HB) and the other containing 20 μg of HBsAg (Engerix-B), and recommended doses for each injection vary for the two preparations (Table 304-8). Combinations of hepatitis B vaccine with other childhood vaccines are available as well (Table 304-8).

For unvaccinated persons sustaining an exposure to HBV, *postexposure* prophylaxis with a combination of

HBIG (for rapid achievement of high-titer circulating anti-HBs) and hepatitis B vaccine (for achievement of long-lasting immunity as well as its apparent efficacy in attenuating clinical illness after exposure) is recommended. For *perinatal* exposure of infants born to HBsAg-positive mothers, a single dose of HBIG, 0.5 mL, should be administered IM in the thigh *immediately after birth* , followed by a complete course of three injections of recombinant hepatitis B vaccine (see doses above) to be started within the first 12 hours of life. For those experiencing a direct percutaneous inoculation or transmucosal exposure to HBsAg-positive blood or body fluids (e.g., accidental *needle stick*, other mucosal penetration, or ingestion), a single IM dose of HBIG, 0.06 mL/kg, administered as soon after exposure as possible, is followed by a complete course of hepatitis B vaccine to begin within the first week. For those exposed by *sexual* contact to a patient with acute hepatitis B, a single IM dose of HBIG, 0.06 mL/kg, should be given within 14 days of exposure, to be followed by a complete course of hepatitis B vaccine. When both HBIG and hepatitis B vaccine are recommended, they may be given at the same time but at separate sites.

The precise duration of protection afforded by hepatitis B vaccine is unknown; however, ~80–90% of immunocompetent vaccinees retain protective levels of anti-HBs for at least 5 years, and 60–80% for 10 years. Thereafter and even after anti-HBs becomes undetectable, protection persists against clinical hepatitis B, hepatitis B surface antigenemia, and chronic HBV infection. Currently, *booster* immunizations are not recommended routinely, except in immunosuppressed persons who have lost detectable anti-HBs or immunocompetent persons who sustain percutaneous HBsAg-positive inoculations after losing detectable antibody. Specifically, for hemodialysis patients, annual anti-HBs testing is recommended after vaccination; booster doses are recommended when anti-HBs levels fall to <10 mIU/mL. As noted above, for persons at risk of both hepatitis A and B, a combined vaccine is available containing 720 enzyme-linked immunoassay units (ELUs) of inactivated HAV and 20 μg of recombinant HBsAg (at 0, 1, and 6 months).

HEPATITIS D

Infection with hepatitis D can be prevented by vaccinating susceptible persons with hepatitis B vaccine. No product is available for immunoprophylaxis to prevent HDV superinfection in HBsAg carriers; for them, avoidance of percutaneous exposures and limitation of intimate contact with persons who have HDV infection are recommended.

HEPATITIS C

IG is ineffective in preventing hepatitis C and is no longer recommended for postexposure prophylaxis in cases of perinatal, needle stick, or sexual exposure. Although prototype vaccines that induce antibodies to HCV envelope proteins have been developed, currently, hepatitis C vaccination is not feasible practically. Genotype and quasispecies viral heterogeneity, as well as rapid evasion of neutralizing antibodies by this rapidly mutating virus, conspire to render HCV a difficult target for immunoprophylaxis with a vaccine. Prevention of transfusion-associated hepatitis C has been accomplished by the following successively introduced measures: exclusion of commercial blood donors and reliance on a volunteer blood supply; screening donor blood with surrogate markers such as ALT (no longer recommended) and anti-HBc, markers that identify segments of the blood donor population with an increased risk of bloodborne infections; exclusion of blood donors in high-risk groups for AIDS and the introduction of anti-HIV screening tests; and progressively sensitive serologic and virologic screening tests for HCV infection.

In the absence of active or passive immunization, prevention of hepatitis C includes behavior changes and precautions to limit exposures to infected persons. Recommendations designed to identify patients with clinically inapparent hepatitis as candidates for medical management have as a secondary benefit the identification of persons whose contacts could be at risk of becoming infected. A so-called look-back program has been recommended to identify persons who were transfused before 1992 with blood from a donor found subsequently to have hepatitis C. In addition, anti-HCV testing is recommended for anyone who received a blood transfusion or a transplanted organ before the introduction of second-generation screening tests in 1992, those who ever used injection drugs (or took other illicit drugs by noninjection routes), chronically hemodialyzed patients, persons with clotting disorders who received clotting factors made before 1987 from pooled blood products, persons with elevated aminotransferase levels, health workers exposed to HCV-positive blood or contaminated needles, persons with HIV infection, health care and public safety personnel following a needle-stick or other nonpercutaneous exposure to HCV-infected material, sexual partners of persons with hepatitis C, and children born to HCV-positive mothers (Table 304-4).

For stable, monogamous sexual partners, sexual transmission of hepatitis C is unlikely, and sexual barrier precautions are not recommended. For persons with multiple sexual partners or with sexually transmitted diseases, the risk of sexual transmission of hepatitis C is increased, and barrier precautions (latex condoms) are recommended. A person with hepatitis C should avoid sharing such items as razors, toothbrushes, and nail clippers with sexual partners and family members. No special precautions are recommended for babies born to mothers with hepatitis C, and breast-feeding does not have to be restricted.

HEPATITIS E

Whether IG prevents hepatitis E remains undetermined. A safe and effective recombinant vaccine has been developed and is available in endemic areas but not in the United States.

FURTHER READINGS

Blum HE, Marcellin P: EASL Consensus Conference on Hepatitis B. J Hepatol 39(Suppl 1):1, 2003

Centers For Disease Control and Prevention: Updated U.S. Public Health Service guidelines for the management of occupational exposures to HBV, HCV, and HIV and recommendations for postexposure prophylaxis. MMWR 50(RR-11):1, 2001

——: A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP) part 1: Immunization of infants, children, and adolescents. MMWR 54 (RR-16):1, 2005

——: A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP) part II: Immunization of adults. MMWR 55(RR-16):1, 2006

——: Prevention of hepatitis A through active or passive immunization: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR 55 (RR-7):1, 2006

——: Recommendations for identification and public health management of persons with chronic hepatitis B virus infection. MMWR 57(RR-8):1, 2008

——: Updated recommendations from the Advisory Committee on Immunization Practices (ACIP) for use of hepatitis A vaccine in close contacts of newly arriving international adoptees. MMWR 58:1006, 2009

Chen CJ et al: Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. JAMA 295:65, 2006[PMID: 16391218] [Full Text]

Dienstag JL: Hepatitis B virus infection. N Engl J Med 359:1486, 2008[PMID: 18832247] [Full Text]

Dienstag JL, McHutchison JG: American Gastroenterological Association technical review on the management of hepatitis C. Gastroenterology 130:231, 2006[PMID: 16401486] [Full Text]

Ferrari C et al: Immunopathogenesis of hepatitis B. J Hepatol 39(Suppl 1):S36, 2003

Ganem D, Prince AM: Hepatitis B virus infection—natural history and clinical consequences. N Engl J Med 350:1118, 2004[PMID: 15014185] [Full Text]

Ghany MG et al: Diagnosis, management, and treatment of hepatitis C: An update. Hepatology 49:1335, 2009[PMID: 19330875] [Full Text]

Gish RG, Locarnini SA: Chronic hepatitis B: Current testing strategies. Clin Gastroenterol Hepatol 4:666, 2006[PMID: 16765304] [Full Text]

Hadziyannis SJ, Vassilopoulos D: Hepatitis B e antigen-negative chronic hepatitis B. Hepatology 34:617, 2001[PMID: 11584355] [Full Text]

Jaeckel E et al: Treatment of acute hepatitis C with interferon alfa-2b. N Engl J Med 345:1452, 2001[PMID: 11794193] [Full Text]

Lauer GM, Walker BD: Medical progress: Hepatitis C virus infection. N Engl J Med 345:41, 2001[PMID: 11439948] [Full Text]

Lok Asf, McMahon BJ: Chronic hepatitis B: Hepatology 45:507, 2007[PMID: 17256718] [Full Text]

——: Chronic hepatitis B: Update 2009. Hepatology 50: 661, 2009

National Institutes of Health Consensus Development Conference: Management of hepatitis B. Hepatology 49 (Suppl 5):S1, 2009

Pawlotsky J-M: Molecular diagnosis of viral hepatitis. Gastroenterology 122:1554, 2002[PMID: 12016423] [Full Text]

Rehermann B: Immune response in hepatitis B virus infection. Semin Liver Dis 23:21, 2003[PMID: 12616448] [Full Text]

Rizzetto M: Hepatitis D: Thirty years after. J Hepatol 50:1043, 2009[PMID: 19285743] [Full Text]

Romeo R et al: A 28-year study of the course of hepatitis Delta infection: A risk factor for cirrhosis and hepatocellular carcinoma. Gastroenterology 136:1629, 2009[PMID: 19208358] [Full Text]

Schreiber GB et al: The risk of transfusion-transmitted viral infection. N Engl J Med 334:1685, 1996[PMID: 8637512] [Full Text]

Thimme R et al: Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease. Proc Natl Acad Sci USA 99:15661, 2002[PMID: 12441397] [Full Text]

Thomas D, Seeff LB: Natural history of hepatitis C. Clin Liver Dis 9:383, 2005[PMID: 16023972] [Full Text]

—— et al: Genetic variation in IL28B and spontaneous clearance of hepatitis C virus Nature 461:798, 2009

Von Hahn T et al: Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection in vivo. Gastroenterology 132:667, 2007

Yang H-I et al: Hepatitis B e antigen and the risk of hepatocellular carcinoma. N Engl J Med 347:168, 2002[PMID: 12124405] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TOXIC AND DRUG-INDUCED HEPATITIS: INTRODUCTION

Liver injury may follow the inhalation, ingestion, or parenteral administration of a number of pharmacologic and chemical agents. These include industrial toxins (e.g., carbon tetrachloride, trichloroethylene, and yellow phosphorus); the heat-stable toxic bicyclic octapeptides of certain species of *Amanita* and *Galerina* (hepatotoxic mushroom poisoning); and, more commonly, pharmacologic agents used in medical therapy. Among patients with acute liver failure, drug-induced liver injury is the cause in a majority of all cases, and liver toxicity accounts for the abandonment of many new drugs during their development. It is essential that any patient presenting with jaundice or altered biochemical liver tests be questioned carefully about exposure to chemicals used in work or at home, drugs taken by prescription or bought over the counter, and herbal or alternative medicines. Hepatotoxic drugs can injure the hepatocyte directly (e.g., via a free-radical or metabolic intermediate that causes peroxidation of membrane lipids and that results in liver cell injury). Alternatively, the drug or its metabolite can distort cell membranes or other cellular molecules, bind covalently to intracellular proteins, activate apoptotic pathways, interfere with bile salt export proteins, or block biochemical pathways or cellular integrity **(Figure 305-1)**. Interference with bile canalicular pumps can allow endogenous bile acids, which can injure the liver, to accumulate. Such injuries, in turn, may lead to necrosis of hepatocytes; injure bile ducts, producing cholestasis; or block pathways of lipid movement, inhibit protein synthesis, or impair mitochondrial oxidation of fatty acids, resulting in lactic acidosis and intracellular triglyceride accumulation (expressed histologically as microvesicular steatosis). In some cases, drug metabolites sensitize hepatocytes to toxic cytokines, and differences between susceptible and nonsusceptible drug recipients may be attributable to polymorphisms in elaboration of competing, protective cytokines, as has been suggested for acetaminophen hepatotoxicity (see below). Immunologically mediated liver injury has been postulated to represent another mechanism of drug hepatotoxicity (see below). In addition, a role has been shown for activation of nuclear transporters, such as the constitutive androstane receptor (CAR), in the induction of drug hepatotoxicity.

Figure 305-1

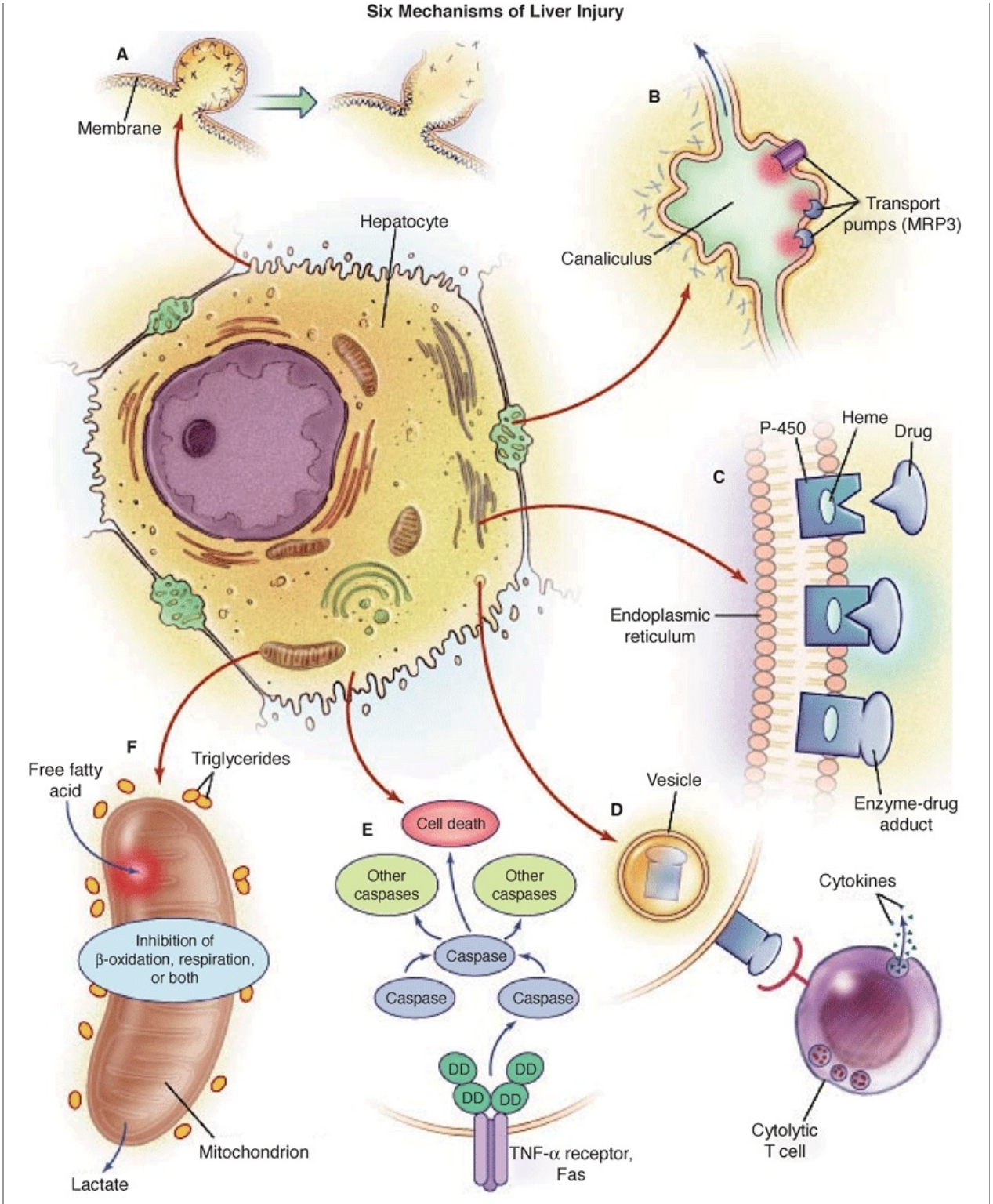

A. Rupture of cell membrane.
B. Injury of bile canaliculus (disruption of transport pumps).
C. P-450-drug covalent binding (drug adducts).
D. Drug adducts targeted by CTLs/cytokines.
E. Activation of apoptotic pathway by TNFα/Fas.
F. Inhibition of mitochondrial function.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Potential mechanisms of drug-induced liver injury. The normal hepatocyte may be affected adversely by drugs through **A.** disruption of intracellular calcium homeostasis that leads to the disassembly of actin fibrils at the surface of the hepatocyte, resulting in blebbing of the cell membrane, rupture, and cell lysis; **B.** disruption of actin filaments next to the canaliculus (the specialized portion of the cell responsible for bile excretion), leading to loss of villous processes and interruption of transport pumps such as multidrug resistance–associated protein 3 (MRP3), which, in turn, prevents the excretion of bilirubin and other organic compounds; **C.** covalent binding of the heme-containing cytochrome P-450 enzyme to the drug, thus creating nonfunctioning adducts; **D.** migration of these enzyme-drug adducts to the cell surface in vesicles to serve as target immunogens for cytolytic attack by T cells, stimulating an immune response involving cytolytic T cells and cytokines; **E.** activation of apoptotic pathways by tumor necrosis factor α (TNF-α) receptor or Fas (DD denotes death domain), triggering the cascade of intercellular caspases, resulting in programmed cell death;

or **F.** inhibition of mitochondrial function by a dual effect on both β -oxidation and the respiratory-chain enzymes, leading to failure of free fatty acid metabolism, a lack of aerobic respiration, and accumulation of lactate and reactive oxygen species (which may disrupt mitochondrial DNA). Toxic metabolites excreted in bile may damage bile-duct epithelium (not shown). CTLs, cytolytic T lymphocytes. *(Reproduced from LEE WM: Drug-induced hepatotoxicity. N Engl J Med 349:474, 2003, with permission.)*

Most drugs, which are water-insoluble, undergo a series of hepatic metabolic transformation steps, culminating in a water-soluble form appropriate for renal or biliary excretion. This process begins with oxidation or methylation initially mediated by the microsomal mixed-function oxygenases cytochrome P450 (phase I reaction) followed by glucuronidation or sulfation (phase II reaction) or inactivation by glutathione. Most drug hepatotoxicity is mediated by a phase I toxic metabolite, but glutathione depletion, precluding inactivation of harmful compounds by glutathione S-transferase, can contribute as well.

In general, two major types of chemical hepatotoxicity have been recognized: (1) direct toxic and (2) idiosyncratic. As shown in **Table 305-1**, direct toxic hepatitis occurs with predictable regularity in individuals exposed to the offending agent and is dose-dependent. The latent period between exposure and liver injury is usually short (often several hours), although clinical manifestations may be delayed for 24–48 hours. Agents producing toxic hepatitis are generally systemic poisons or are converted in the liver to toxic metabolites. The direct hepatotoxins result in morphologic abnormalities that are reasonably characteristic and reproducible for each toxin. For example, carbon tetrachloride and trichloroethylene characteristically produce a centrilobular zonal necrosis, whereas yellow phosphorus poisoning typically results in periportal injury. The hepatotoxic octapeptides of *Amanita phalloides* usually produce massive hepatic necrosis; the lethal dose of the toxin is ~ 10 mg, the amount found in a single deathcap mushroom. Tetracycline, when administered in IV doses >1.5 g daily, leads to microvesicular fat deposits in the liver. Liver injury, which is often only one facet of the toxicity produced by the direct hepatotoxins, may go unrecognized until jaundice appears.

Table 305-1 Some Features of Toxic and Drug-Induced Hepatic Injury

	Direct Toxic Effect*		Idiosyncratic*			Other*
Features	**(Carbon Tetrachloride)**	**(Acetaminophen)**	**(Halothane)**	**(Isoniazid)**	**(Chlorpromazine)**	**(Oral Contraceptive Agents)**
Predictable and dose-related toxicity	+	+	0	0	0	+
Latent period	Short	Short	Variable	Variable	Variable	Variable
Arthralgia, fever, rash, eosinophilia	0	0	+	0	+	0
Liver morphology	Necrosis, fatty infiltration	Centrilobular necrosis	Similar to viral hepatitis	Similar to viral hepatitis	Cholestasis *with* portal inflammation	Cholestasis *without* portal inflammation, vascular lesions

* The drugs listed are typical samples.

In idiosyncratic drug reactions, the occurrence of hepatitis is usually infrequent (1 in 10^3–10^5 patients) and unpredictable; the response is not as clearly dose-dependent as is injury associated with direct hepatotoxins, and liver injury may occur at any time during or shortly after exposure to the drug. Adding to the difficulty of predicting or identifying idiosyncratic drug hepatotoxicity is the occurrence of mild, transient, nonprogressive serum aminotransferase elevations that resolve with continued drug use. Such "adaptation," the mechanism of which is unknown, occurs in such drugs as isoniazid, valproate, phenytoin, and HMG-CoA reductase inhibitors (statins). Extrahepatic manifestations of hypersensitivity, such as rash, arthralgias, fever, leukocytosis, and eosinophilia, occur in about one-quarter of patients with idiosyncratic hepatotoxic drug reactions; this observation and the unpredictability of idiosyncratic drug hepatotoxicity contributed to the hypothesis that this category of drug reactions is immunologically mediated. More recent evidence, however, suggests that, in most cases, even idiosyncratic reactions represent direct hepatotoxicity but are caused by drug metabolites rather than by the intact compound. Even the prototypes of idiosyncratic hepatotoxicity reactions, halothane hepatitis and isoniazid hepatotoxicity, associated frequently with hypersensitivity manifestations, are now recognized to be mediated by toxic metabolites that damage liver cells directly. Currently, most idiosyncratic reactions are thought to result from differences in metabolic reactivity to specific agents; host susceptibility is mediated by the kinetics of toxic metabolite generation, which differs among individuals, probably mediated by genetic polymorphisms in drug-metabolizing pathways (e.g., differences in cytochrome C450 enzyme isotypes or in acetylation). Associations between certain HLAhaplotypes have been drawn with hepatotoxicity of such drugs as amoxicillin/clavulanate, statins, halothane, nitrofurantoin, chlorpromazine, and flucloxacillin. Occasionally, however, the clinical features of an allergic reaction (prominent tissue eosinophilia, autoantibodies, etc.) are difficult to ignore. In vitro models have been described in which lymphocyte

cytotoxicity can be demonstrated against rabbit hepatocytes altered by incubation with the potential offending drug. Furthermore, several instances of drug hepatotoxicity are associated with the appearance of autoantibodies, including a class of antibodies to liver-kidney microsomes, anti-LKM2, directed against a cytochrome P450 enzyme. Similarly, in selected cases, a drug or its metabolite has been shown to bind to a host cellular component forming a hapten; the immune response to this "neoantigen" is postulated to play a role in the pathogenesis of liver injury. Therefore, some authorities subdivide idiosyncratic drug hepatotoxicity into hypersensitivity (allergic) and "metabolic" categories. Several unusual exceptions notwithstanding, true drug allergy is difficult to support in most cases of idiosyncratic drug-induced liver injury.

Idiosyncratic reactions lead to a morphologic pattern that is more variable than those produced by direct toxins; a single agent is often capable of causing a variety of lesions, although certain patterns tend to predominate. Depending on the agent involved, idiosyncratic hepatitis may result in a clinical and morphologic picture indistinguishable from that of viral hepatitis (e.g., halothane) or may simulate extrahepatic bile duct obstruction clinically with morphologic evidence of cholestasis. Drug-induced cholestasis ranges from mild to increasingly severe: (1) bland cholestasis with limited hepatocellular injury (e.g., estrogens, 17,α-substituted androgens); (2) inflammatory cholestasis [e.g., phenothiazines, amoxicillin-clavulanic acid (the most frequently implicated antibiotic among cases of drug-induced liver injury), oxacillin, erythromycin estolate]; (3) sclerosing cholangitis (e.g., after intrahepatic infusion of the chemotherapeutic agent floxuridine for hepatic metastases from a primary colonic carcinoma); (4) disappearance of bile ducts, "ductopenic" cholestasis, similar to that observed in chronic rejection following liver transplantation (e.g., carbamazepine, chlorpromazine, tricyclic antidepressant agents). Cholestasis may result from binding of drugs to canalicular membrane transporters, accumulation of toxic bile acids resulting from canalicular pump failure, or genetic defects in canalicular transporter proteins. Morphologic alterations may also include bridging hepatic necrosis (e.g., methyldopa), or, infrequently, hepatic granulomas (e.g., sulfonamides). Some drugs result in macrovesicular or microvesicular steatosis or steatohepatitis, which, in some cases, has been linked to mitochondrial dysfunction and lipid peroxidation. Severe hepatotoxicity associated with steatohepatitis, most likely a result of mitochondrial toxicity, is being recognized with increasing frequency among patients receiving antiretroviral therapy with reverse transcriptase inhibitors (e.g., zidovudine, didanosine) or protease inhibitors (e.g., indinavir, ritonavir) for HIV infection **(Chap. 189)**. Generally, such mitochondrial hepatotoxicity of these antiretroviral agents is reversible, but dramatic, nonreversible hepatotoxicity associated with mitochondrial injury (inhibition of DNA polymerase γ) was the cause of acute liver failure encountered during early clinical trials of now-abandoned fialuridine, a fluorinated pyrimidine analogue with potent antiviral activity against hepatitis B virus. Another potential target for idiosyncratic drug hepatotoxicity is sinusoidal lining cells; when these are injured, such as by high-dose chemotherapeutic agents (e.g., cyclophosphamide, melphalan, busulfan) administered prior to bone marrow transplantation, venoocclusive disease can result.

Not all adverse hepatic drug reactions can be classified as either toxic or idiosyncratic in type. For example, oral contraceptives, which combine estrogenic and progestational compounds, may result in impairment of hepatic tests and, occasionally, jaundice; however, they do not produce necrosis or fatty change, manifestations of hypersensitivity are generally absent, and susceptibility to the development of oral contraceptive–induced cholestasis appears to be genetically determined. Such estrogen-induced cholestasis is more common in women with cholestasis of pregnancy, a disorder linked to genetic defects in multidrug resistance–associated canalicular transporter proteins. Other instances of genetically determined drug hepatotoxicity have been identified. For example, ~10% of the population have an autosomal recessive trait associated with the absence of cytochrome P450 enzyme 2D6 and have impaired debrisoquine-4-hydroxylase enzyme activity. As a result, they cannot metabolize, and are at increased risk of hepatotoxicity resulting from certain compounds such as desipramine, propranolol, and quinidine.

Some forms of drug hepatotoxicity are so rare (e.g., occurring in <1:10,000 recipients), that they do not become apparent during clinical trials, involving only several thousand recipients, conducted to obtain drug registration. An example of such rare, but serious, idiosyncratic drug hepatotoxicity followed the approval and generalized use of troglitazone, a peroxisomal, proliferator activator–receptor γ agonist, the first introduced example of a thiazolidinedione insulin-sensitizing agent. This instance of drug hepatotoxicity was not recognized until well after the drug was introduced, underlining the importance of postmarketing surveillance in identifying toxic drugs and in leading to their withdrawal from use. Fortunately, such hepatotoxicity is not characteristic of the second-generation thiazolidinedione insulin-sensitizing agents rosiglitazone and pioglitazone; in clinical trials, the frequency of aminotransferase elevations in patients treated with these medications did not differ from that in placebo recipients, and isolated reports of liver injury among recipients are extremely rare.

Because drug-induced hepatitis is often a presumptive diagnosis and many other disorders produce a similar clinicopathologic picture, evidence of a causal relationship between the use of a drug and subsequent liver injury may be difficult to establish. The relationship is most convincing for the direct hepatotoxins, which lead to a high frequency of hepatic impairment after a short latent period. Idiosyncratic reactions may be reproduced, in some instances, when rechallenge, after an asymptomatic period, results in a recurrence of signs, symptoms, and morphologic and biochemical abnormalities. Rechallenge, however, is often ethically unfeasible, because severe reactions may occur. Causality-assessment methodologies [scoring systematically based on a checklist of such variables as index of suspicion, time of onset, clinical-biochemical features, type of injury (direct, idiosyncratic), extrahepatic features, course, histologic features, drug serum levels, genetic markers and polymorphisms, and exclusion of other potential causes] have been adopted to add objectivity to diagnoses of drug-induced liver injury; however, even these approaches have their limitations and yield residual uncertainty.

Generally, drug hepatotoxicity is not more frequent in persons with underlying chronic liver disease. Reported exceptions include hepatotoxicity of aspirin, methotrexate, isoniazid (only in certain experiences), and

antiretroviral therapy for HIV infection.

Treatment: Toxic and Drug-Induced Hepatic Disease

Treatment is largely supportive, except in acetaminophen hepatotoxicity (see below). In patients with fulminant hepatitis resulting from drug hepatotoxicity, liver transplantation may be lifesaving **(Chap. 310)**. Withdrawal of the suspected agent is indicated at the first sign of an adverse reaction. In the case of the direct toxins, liver involvement should not divert attention from renal or other organ involvement, which may also threaten survival. Glucocorticoids for drug hepatotoxicity with allergic features, silibinin for hepatotoxic mushroom poisoning, and ursodeoxycholic acid for cholestatic drug hepatotoxicity have never been shown to be effective and are not recommended.

In **Table 305-2**, several classes of chemical agents are listed together with examples of the pattern of liver injury produced by them. Certain drugs appear to be responsible for the development of chronic as well as acute hepatic injury. For example, oxyphenisatin, methyldopa, and isoniazid have been associated with moderate to severe chronic hepatitis, and halothane and methotrexate have been implicated in the development of cirrhosis. A syndrome resembling primary biliary cirrhosis has been described following treatment with chlorpromazine, methyl testosterone, tolbutamide, and other drugs. Portal hypertension in the absence of cirrhosis may result from alterations in hepatic architecture produced by vitamin A or arsenicintoxication, industrial exposure to vinyl chloride, or administration of thorium dioxide. The latter three agents have also been associated with angiosarcoma of the liver. Oral contraceptives have been implicated in the development of hepatic adenoma and, rarely, hepatocellular carcinoma and hepatic vein occlusion (Budd-Chiari syndrome). Another unusual lesion, peliosis hepatis (blood cysts of the liver), has been observed in some patients treated with anabolic steroids. The existence of these hepatic disorders expands the spectrum of liver injury induced by chemical agents and emphasizes the need for a thorough drug history in all patients with liver dysfunction.

Table 305-2 Principal Alterations of Hepatic Morphology Produced by Some Commonly Used Drugs and Chemicals[a]

Principal Morphologic Change	Class of Agent	Example
Cholestasis	Anabolic steroid	Methyl testosterone
	Antibiotic	Erythromycin estolate, nitrofurantoin, rifampin, amoxicillinclavulanic acid, oxacillin
	Anticonvulsant	Carbamazine
	Antidepressant	Duloxetine, mirtazapine, tricyclic antidepressants
	Anti-inflammatory	Sulindac
	Antiplatelet	Clopidogrel
	Antihypertensive	Irbesartan, fosinopril
	Antithyroid	Methimazole
	Calcium channel blocker	Nifedipine, verapamil
	Immunosuppressive	Cyclosporine
	Lipid-lowering	Ezetimibe
	Oncotherapeutic	Anabolic steroids, busulfan, tamoxifen, irinotecan, cytarabine
	Oral contraceptive	Norethynodrel with mestranol
	Oral hypoglycemic	Chlorpropamide
	Tranquilizer	Chlorpromazine[b]
Fatty liver	Antiarrhythmic	Amiodarone
	Antibiotic	Tetracycline (high-dose, IV)
	Anticonvulsant	Valproic acid
	Antiviral	Dideoxynucleosides (e.g., zidovudine), protease inhibitors (e.g., indinavir, ritonavir)
	Oncotherapeutic	Asparaginase, methotrexate
Hepatitis	Anesthetic	Halothane[c]

	Antiandrogen	Flutamide
	Antibiotic	Isoniazid,[c] rifampicin, nitrofurantoin, telithromycin, minocycline,[d] pyrazinamide, trovafloxacin[e]
	Anticonvulsant	Phenytoin, carbamazine, valproic acid, phenobarbital
	Antidepressant	Iproniazid, amitriptyline, imipramine, trazodone, venlafaxine, fluoxetine, paroxetine, duloxetine, sertraline, nefazodone,[e] bupropion
	Antifungal	Ketoconazole, fluconazole, itraconazole
	Antihypertensive	Methyldopa,[c] captopril, enalapril, lisinopril, losartan
	Anti-inflammatory	Ibuprofen, indomethacin, diclofenac, sulindac, bromfenac
	Antipsychotic	Risperidone
	Antiviral	Zidovudine, didanosine, stavudine, nevirapine, ritonavir, indinavir, tipranavir, zalcitabine
	Calcium channel blocker	Nifedipine, verapamil, diltiazem
	Cholinesterase inhibitor	Tacrine
	Diuretic	Chlorothiazide
	Laxative	Oxyphenisatin[c,e]
	Norepinephrine-reuptake inhibitor	Atomoxetine
	Oral hypoglycemic	Troglitazone,[e] acarbose
Mixed hepatitis/cholestatic	Antibiotic	Amoxicillin-clavulanic acid, trimethoprim-sulfamethoxazole
	Antibacterial	Clindamycin
	Antifungal	Terbinafine
	Antihistamine	Cyproheptadine
	Immunosuppressive	Azathioprine
	Lipid-lowering	Nicotinic acid, lovastatin, ezetimide
Toxic (necrosis)	Analgesic	Acetaminophen
	Hydrocarbon	Carbon tetrachloride
	Metal	Yellow phosphorus
	Mushroom	*Amanita phalloides*
	Solvent	Dimethylformamide
Granulomas	Antiarrhythmic	Quinidine, diltiazem
	Antibiotic	Sulfonamides
	Anticonvulsant	Carbamazine
	Anti-inflammatory	Phenylbutazone
	Xanthine oxidase inhibitor	Allopurinol

[a] Several agents cause more than one type of liver lesion and appear under more than one category.

[b] Rarely associated with primary biliary cirrhosis-like lesion.

[c] Occasionally associated with chronic hepatitis or bridging hepatic necrosis or cirrhosis.

[d] Associated with an autoimmune hepatitis-like syndrome.

[e] Withdrawn from use because of severe hepatotoxicity.

The following are patterns of adverse hepatic reactions for some prototypic agents.

Acetaminophen Hepatotoxicity (Direct Toxin)

Acetaminophen can cause severe centrilobular hepatic necrosis when ingested in large amounts in suicide attempts or accidentally by children. In the United States and England, acetaminophen hepatotoxicity is the most common culprit among patients presenting with acute liver failure and the leading indication for liver transplantation among patients with drug-induced liver failure. A single dose of 10–15 g, occasionally less, may produce clinical evidence of liver injury. Fatal fulminant disease is usually (although not invariably) associated with ingestion of ≥25 g. Blood levels of acetaminophen correlate with the severity of hepatic injury (levels >300 μg/mL 4 h after ingestion are predictive of the development of severe damage; levels <150 μg/mL suggest that hepatic injury is highly unlikely). Nausea, vomiting, diarrhea, abdominal pain, and shock are early manifestations occurring 4–12 hours after ingestion. Then 24–48 hours later, when these features are abating, hepatic injury becomes apparent. Maximal abnormalities and hepatic failure may not be evident until 4–6 days after ingestion, and aminotransferase levels approaching 10,000 units are not uncommon (i.e., levels far exceeding those in patients with viral hepatitis). Renal failure and myocardial injury may be present.

Acetaminophen is metabolized predominantly by a phase II reaction to innocuous sulfate and glucuronide metabolites; however, a small proportion of acetaminophen is metabolized by a phase I reaction to a hepatotoxic metabolite formed from the parent compound by the cytochrome P450 CYP2E1. This metabolite, *N*-acetyl-benzoquinone-imine (NAPQI), is detoxified by binding to "hepatoprotective" glutathione to become harmless, water-soluble mercapturic acid, which undergoes renal excretion. When excessive amounts of NAPQI are formed, or when glutathione levels are low, glutathione levels are depleted and overwhelmed, permitting covalent binding to nucleophilic hepatocyte macromolecules forming acetaminophen-protein "adducts." These adducts, which can be measured in serum by high-performance liquid chromatography, hold promise as diagnostic markers of acetaminophen hepatotoxicity. The binding of acetaminophen to hepatocyte macromolecules is believed to lead to hepatocyte necrosis; the precise sequence and mechanism are unknown. Hepatic injury may be potentiated by prior administration of alcohol, phenobarbital, isoniazid, or other drugs; by conditions that stimulate the mixed-function oxidase system; or by conditions such as starvation that reduce hepatic glutathione levels. The xenobiotic (environmental, exogenous substance) receptor CAR has been shown in a mouse model of acetaminophen hepatotoxicity to induce acetaminophen-metabolizing enzymes and, thereby, regulate and increase hepatotoxicity. Cimetidine, which inhibits P450 enzymes, has the potential to reduce generation of the toxic metabolite. Alcohol induces cytochrome P450 CYP2E1; consequently, increased levels of the toxic metabolite NAPQI are produced in chronic alcoholics after acetaminophen ingestion. In addition, alcohol suppresses hepatic glutathione production. Therefore, in chronic alcoholics, the toxic dose of acetaminophen may be as low as 2 g, and alcoholic patients should be warned specifically about the dangers of even standard doses of this commonly used drug. Such "therapeutic misadventures" also occur occasionally in patients with severe, febrile illnesses or pain syndromes; in such a setting, several days of anorexia and near-fasting coupled with regular administration of extra-strength acetaminophen formulations result in a combination of glutathione depletion and relatively high NAPQI levels in the absence of a history of recognized acetaminophen overdose. In a 2006 study, aminotransferase elevations were identified in 31–44% of normal subjects treated for 14 days with the maximal recommended dose of acetaminophen, 4 g daily (administered alone or as part of an acetaminophen/opioid combination); because these changes were transient and never associated with bilirubin elevation, the clinical relevance of these findings remains to be determined. Although underlying HCV infection was found to be associated with an increased risk of acute liver injury in patients hospitalized for acetaminophen overdose, generally, in patients with nonalcoholic liver disease, acetaminophen taken in recommended doses, may be the safest analgesic/antipyretic. In this vein, acetaminophen use in cirrhotic patients has not been associated with hepatic decompensation. On the other hand, because of the link between acetaminophen use and liver injury, and because of the limited safety margin between safe and toxic doses, the Food and Drug Administration (FDA) has recommended that the daily dose of acetaminophen be reduced from 4 g to 3.25 g (even lower for persons with chronic alcohol use), that all acetaminophen-containing products be labeled prominently as containing acetaminophen, and that the potential for liver injury be prominent in the packaging of acetaminophen and acetaminophen-containing products.

TREATMENT: ACETAMINOPHEN OVERDOSAGE

Treatment includes gastric lavage, supportive measures, and oral administration of activated charcoal or cholestyramine to prevent absorption of residual drug. Neither charcoal nor cholestyramine appears to be effective if given >30 min after acetaminophen ingestion; if they are used, the stomach lavage should be done before other agents are administered orally. The chances of possible, probable, and high-risk hepatotoxicity can be derived from a nomogram plot **(Fig. 305-2)**, readily available in emergency departments as a function of measuring acetaminophen plasma levels 8 h after ingestion. In patients with high acetaminophen blood levels (>200 μg/mL measured at 4 h or >100 μg/mL at 8 h after ingestion), the administration of sulfhydryl compounds (e.g., cysteamine, cysteine, or *N*-acetylcysteine) reduces the severity of hepatic necrosis. These agents appear to act by providing a reservoir of sulfhydryl groups to bind the toxic metabolites or by stimulating synthesis and repletion of hepatic glutathione. Therapy should be begun within 8 h of ingestion but may be effective even if given as late as 24–36 h after overdose. Later administration of sulfhydryl compounds is of uncertain value. Routine use of *N*-acetylcysteine has substantially reduced the occurrence of fatal acetaminophen hepatotoxicity. When given orally, *N*-acetylcysteine is diluted to yield a 5% solution. A loading dose of 140 mg/kg is given, followed by 70 mg/kg every 4 h for 15–20 doses. Whenever a patient with potential acetaminophen hepatotoxicity is encountered, a local poison control center should be contacted. Treatment can be stopped when plasma acetaminophen levels indicate that the risk of liver damage is low. If

signs of hepatic failure (e.g., progressive jaundice, coagulopathy, confusion) occur despite *N*-acetylcysteine therapy for acetaminophen hepatotoxicity, liver transplantation may be the only option. Early arterial blood lactate levels among such patients with acute liver failure may distinguish patients highly likely to require liver transplantation (lactate levels >3.5 mmol/L) from those likely to survive without liver replacement.

Figure 305-2

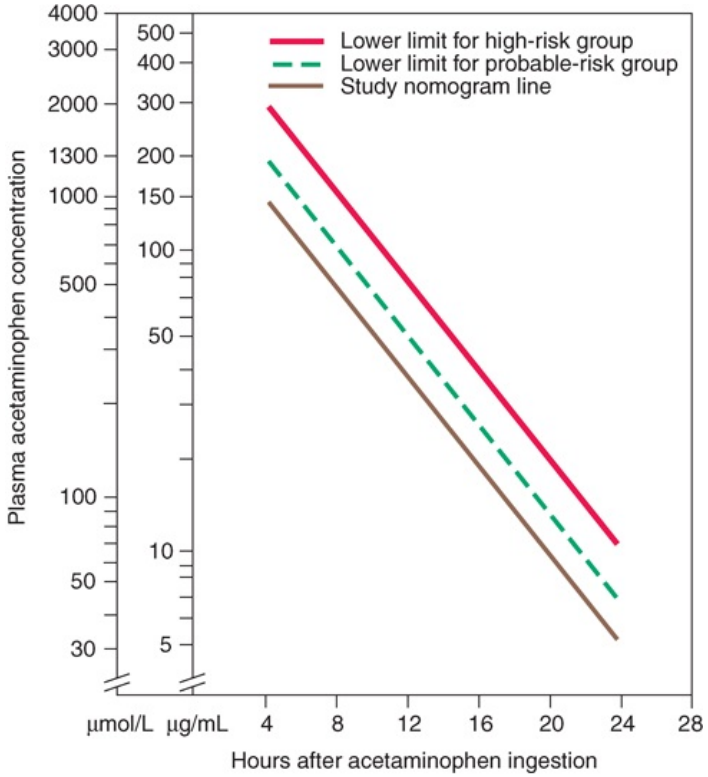

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nomogram to define risk of acetaminophen hepatotoxicity according to initial plasma acetaminophen concentration. (*After BH Rumack, H Matthew: Pediatrics 55:871, 1975.)*

Survivors of acute acetaminophen overdose usually have no hepatic sequelae. In a few patients, prolonged or repeated administration of acetaminophen in therapeutic doses appears to have led to the development of chronic hepatitis and cirrhosis.

Halothane Hepatotoxicity (Idiosyncratic Reaction)

Although, currently, halothane anesthesia is administered in only rare situations, halothane hepatotoxicity was one of the prototypical, and most intensively studied, examples of idiosyncratic drug hepatotoxicity. Administration of halothane, a nonexplosive fluorinated hydrocarbon anesthetic agent that is structurally similar to chloroform, results in severe hepatic necrosis in a small number of individuals, many of whom have previously been exposed to this agent. The failure to produce similar hepatic lesions reliably in animals, the rarity of hepatic impairment in human beings, and the delayed appearance of hepatic injury suggest that halothane is not a direct hepatotoxin but rather a sensitizing agent; however, manifestations of hypersensitivity are seen in <25% of cases. A genetic predisposition leading to an idiosyncratic metabolic reactivity has been postulated and appears to be the most likely mechanism of halothane hepatotoxicity. Adults (rather than children), obese people, and women appear to be particularly susceptible. Fever, moderate leukocytosis, and eosinophilia may occur in the first week following halothane administration. Jaundice is usually noted 7–10 days after exposure but may occur earlier in previously exposed patients. Nausea and vomiting may precede the onset of jaundice. Hepatomegaly is often mild, but liver tenderness is common, and serum aminotransferase levels are elevated. The pathologic changes at autopsy are indistinguishable from massive hepatic necrosis resulting from viral hepatitis. The case-fatality rate of halothane hepatitis is not known but may vary from 20–40% in cases with severe liver involvement. Patients in whom unexplained spiking fever, especially delayed fever, or jaundice develops after halothane anesthesia should not receive this agent again. Because cross-reactions between halothane and methoxyflurane have been reported, the latter agent should not be used after halothane reactions. Later-generation halogenated hydrocarbon anesthetics that have supplanted halothane except in rare instances (e.g., certain types of thoracic surgery), are believed to be associated with a lower risk of hepatotoxicity.

Methyldopa Hepatotoxicity (Toxic and Idiosyncratic Reaction)

Minor alterations in liver tests are reported in ~5% of patients treated with this antihypertensive agent. These trivial abnormalities typically resolve despite continued drug administration. In <1% of patients, acute liver injury resembling viral or chronic hepatitis or, rarely, a cholestatic reaction is seen 1–20 weeks after

methyldopa is started. In 50% of cases the interval is <4 weeks. A prodrome of fever, anorexia, and malaise may be noted for a few days before the onset of jaundice. Rash, lymphadenopathy, arthralgia, and eosinophilia are rare. Serologic markers of autoimmunity are detected infrequently, and <5% of patients have a Coombs-positive hemolytic anemia. In ~15% of patients with methyldopa hepatotoxicity, the clinical, biochemical, and histologic features are those of moderate to severe chronic hepatitis, with or without bridging necrosis and macronodular cirrhosis. With discontinuation of the drug, the disorder usually resolves. Although methyldopa is currently used infrequently, its hepatotoxicity is very well characterized. Among the currently popular antihypertensive agents, angiotensin-converting enzyme (ACE) inhibitors, such as captopril and enalapril, have been blamed, albeit rarely, for hepatotoxicity (primarily cholestasis and cholestatic hepatitis, but also hepatocellular injury). Angiotensin-II receptor antagonists, such as losartan, are unlikely hepatotoxins, although rare reports of liver injury in their recipients have appeared.

Isoniazid Hepatotoxicity (Toxic and Idiosyncratic Reaction)

In ~10% of adults treated with the antituberculosis agent isoniazid, elevated serum aminotransferase levels develop during the first few weeks of therapy; this appears to represent an adaptive response to a toxic metabolite of the drug. Whether or not isoniazid is continued, these values (usually <200 units) return to normal in a few weeks. In ~1% of treated patients, an illness develops that is indistinguishable from viral hepatitis; approximately one-half of these cases occur within the first 2 months of treatment; in the remainder, clinical disease may be delayed for many months. Liver biopsy reveals morphologic changes similar to those of viral hepatitis or bridging hepatic necrosis. The disease may be severe, with a case-fatality rate of 10%. Important liver injury appears to be age-related, increasing substantially after age 35; the highest frequency is in patients over age 50, the lowest under the age of 20. Even for patients >50 years of age monitored carefully during therapy, hepatotoxicity occurs in only ~2%, well below the risk estimate derived from earlier experiences. Isoniazid hepatotoxicity is enhanced by alcohol, rifampin, and pyrazinamide. Fever, rash, eosinophilia, and other manifestations of drug allergy are distinctly unusual. A reactive metabolite of acetylhydrazine, a metabolite of isoniazid, may be responsible for liver injury, and patients who are rapid acetylators would be more prone to such injury. Counterintuitively, in some reports, the opposite is true; slow acetylators are more likely to experience hepatotoxicity and more severe hepatotoxicity than rapid acetylators. Contrary to past reports, more recent studies suggest that hepatotoxicity due to isoniazid as well as to combination antituberculous therapy that includes isoniazid is more likely in patients with underlying chronic hepatitis B. A picture resembling chronic hepatitis has been observed in a few patients. Careful liver-test monitoring is advisable in patients being treated with isoniazid.

Sodium Valproate Hepatotoxicity (Toxic and Idiosyncratic Reaction)

Sodium valproate, an anticonvulsant useful in the treatment of petit mal and other seizure disorders, has been associated with the development of severe hepatic toxicity and, rarely, fatalities, predominantly in children but also in adults. Among children listed as candidates for liver transplantation, valproate is the most common antiepileptic drug implicated. Asymptomatic elevations of serum aminotransferase levels have been recognized in as many as 45% of treated patients. These "adaptive" changes, however, appear to have no clinical importance, because major hepatotoxicity is not seen in the majority of patients despite continuation of drug therapy. In the rare patients in whom jaundice, encephalopathy, and evidence of hepatic failure are found, examination of liver tissue reveals microvesicular fat and bridging hepatic necrosis, predominantly in the centrilobular zone. Bile duct injury may also be apparent. Most likely, sodium valproate is not directly hepatotoxic, but its metabolite, 4-pentenoic acid, may be responsible for hepatic injury. Valproate hepatotoxicity is more common in persons with mitochondrial enzyme deficiencies and may be ameliorated by IV administration of carnitine, which valproate therapy can deplete.

Phenytoin Hepatotoxicity (Idiosyncratic Reaction)

Phenytoin, formerly diphenylhydantoin, a mainstay in the treatment of seizure disorders, has been associated in rare instances with the development of severe hepatitis-like liver injury leading to fulminant hepatic failure. In many patients, the hepatitis is associated with striking fever, lymphadenopathy, rash (Stevens-Johnson syndrome or exfoliative dermatitis), leukocytosis, and eosinophilia, suggesting an immunologically mediated hypersensitivity mechanism. Despite these observations, evidence suggests that metabolic idiosyncrasy may be responsible for hepatic injury. In the liver, phenytoin is converted by cytochrome P450 to metabolites, including the highly reactive electrophilic arene oxides. These metabolites are normally metabolized further by epoxide hydrolases. A defect (genetic or acquired) in epoxide hydrolase activity could permit covalent binding of arene oxides to hepatic macromolecules, thereby leading to hepatic injury. Hepatic injury is usually manifest within the first 2 months after beginning phenytoin therapy. With the exception of an abundance of eosinophils in the liver, the clinical, biochemical, and histologic picture resembles that of viral hepatitis. In rare instances, bile duct injury may be the salient feature of phenytoin hepatotoxicity, with striking features of intrahepatic cholestasis. Asymptomatic elevations of aminotransferase and alkaline phosphatase levels have been observed in a sizable proportion of patients receiving long-term phenytoin therapy. These liver changes are believed by some authorities to represent the potent hepatic enzyme-inducing properties of phenytoin and are accompanied histologically by swelling of hepatocytes in the absence of necroinflammatory activity or evidence of chronic liver disease.

Amiodarone Hepatotoxicity (Toxic and Idiosyncratic Reaction)

Therapy with this potent antiarrhythmic drug is accompanied in 15–50% of patients by modest elevations of serum aminotransferase levels that may remain stable or diminish despite continuation of the drug. Such abnormalities may appear days to many months after beginning therapy. A proportion of those with elevated aminotransferase levels have detectable hepatomegaly, and clinically important liver disease develops in <5% of patients. Features that represent a direct effect of the drug on the liver and that are common to the majority

of long-term recipients are ultrastructural phospholipidosis, unaccompanied by clinical liver disease, and interference with hepatic mixed-function oxidase metabolism of other drugs. The cationic amphiphilic drug and its major metabolite desethylamiodarone accumulate in hepatocyte lysosomes and mitochondria and in bile duct epithelium. The relatively common elevations in aminotransferase levels are also considered a predictable, dose-dependent, direct hepatotoxic effect. On the other hand, in the rare patient with clinically apparent, symptomatic liver disease, liver injury resembling that seen in alcoholic liver disease is observed. The so-called pseudoalcoholic liver injury can range from steatosis to alcoholic hepatitis-like neutrophilic infiltration and Mallory's hyaline to cirrhosis. Electron-microscopic demonstration of phospholipid-laden lysosomal lamellar bodies can help to distinguish amiodarone hepatotoxicity from typical alcoholic hepatitis. This category of liver injury appears to be a metabolic idiosyncrasy that allows hepatotoxic metabolites to be generated. Rarely, an acute idiosyncratic hepatocellular injury resembling viral hepatitis or cholestatic hepatitis occurs. Hepatic granulomas have occasionally been observed. Because amiodarone has a long half-life, liver injury may persist for months after the drug is stopped.

Erythromycin Hepatotoxicity (Cholestatic Idiosyncratic Reaction)

The most important adverse effect associated with erythromycin, more common in children than adults, is the infrequent occurrence of a cholestatic reaction. Although most of these reactions have been associated with erythromycin estolate, other erythromycins may also be responsible. The reaction usually begins during the first 2 or 3 weeks of therapy and includes nausea, vomiting, fever, right upper quadrant abdominal pain, jaundice, leukocytosis, and moderately elevated aminotransferase and alkaline phosphatase levels. The clinical picture can resemble acute cholecystitis or bacterial cholangitis. Liver biopsy reveals variable cholestasis; portal inflammation comprising lymphocytes, polymorphonuclear leukocytes, and eosinophils; and scattered foci of hepatocyte necrosis. Symptoms and laboratory findings usually subside within a few days of drug withdrawal, and evidence of chronic liver disease has not been found on follow-up. The precise mechanism remains ill-defined.

Oral Contraceptive Hepatotoxicity (Cholestatic Reaction)

The administration of oral contraceptive combinations of estrogenic and progestational steroids leads to intrahepatic cholestasis with pruritus and jaundice in a small number of patients weeks to months after taking these agents. Especially susceptible seem to be patients with recurrent idiopathic jaundice of pregnancy, severe pruritus of pregnancy, or a family history of these disorders. With the exception of liver biochemical tests, laboratory studies are normal, and extrahepatic manifestations of hypersensitivity are absent. Liver biopsy reveals cholestasis with bile plugs in dilated canaliculi and striking bilirubin staining of liver cells. In contrast to chlorpromazine-induced cholestasis, portal inflammation is absent. The lesion is reversible on withdrawal of the agent. The two steroid components appear to act synergistically on hepatic function, although the estrogen may be primarily responsible. Oral contraceptives are contraindicated in patients with a history of recurrent jaundice of pregnancy. Primarily benign, but rarely malignant, neoplasms of the liver, hepatic vein occlusion, and peripheral sinusoidal dilatation have also been associated with oral contraceptive therapy. Focal nodular hyperplasia of the liver is not more frequent among users of oral contraceptives.

17, α-Alkyl-Substituted Anabolic Steroids (Cholestatic Reaction)

In the majority of patients receiving these agents, used therapeutically mainly in the treatment of bone marrow failure but used surreptitiously and without medical indication (or unknowingly when included in nutritional supplements) by athletes to improve their performance, mild hepatic dysfunction develops. Impaired excretory function is the predominant defect, but the precise mechanism is uncertain. Jaundice, which appears to be dose-related, develops in only a minority of patients and may be the sole clinical manifestation of hepatotoxicity, although anorexia, nausea, and malaise may occur. Pruritus is not a prominent feature. Serum aminotransferase levels are usually <100 units, and serum alkaline phosphatase levels are normal; mildly elevated; or, in <5% of patients, three or more times the upper limit of normal. Examination of liver tissue reveals cholestasis without inflammation or necrosis. Hepatic sinusoidal dilatation and peliosis hepatis have been found in a few patients. The cholestatic disorder is usually reversible on cessation of treatment, although fatalities have been linked to peliosis. An association with hepatic adenoma and hepatocellular carcinoma has been reported.

Trimethoprim-Sulfamethoxazole Hepatotoxicity (Idiosyncratic Reaction)

This antibiotic combination is used routinely for urinary tract infections in immunocompetent persons and for prophylaxis against and therapy of *Pneumocystis carinii* pneumonia in immunosuppressed persons (transplant recipients, patients with AIDS). With its increasing use, its occasional hepatotoxicity is being recognized with growing frequency. Its likelihood is unpredictable, but when it occurs, trimethoprim-sulfamethoxazole hepatotoxicity follows a relatively uniform latency period of several weeks and is often accompanied by eosinophilia, rash, and other features of a hypersensitivity reaction. Biochemically and histologically, acute hepatocellular necrosis predominates, but cholestatic features are quite frequent. Occasionally, cholestasis without necrosis occurs, and, very rarely, a severe cholangiolytic pattern of liver injury is observed. In most cases, liver injury is self-limited, but rare fatalities have been recorded. The hepatotoxicity is attributable to the sulfamethoxazole component of the drug and is similar in features to that seen with other sulfonamides; tissue eosinophilia and granulomas may be seen. The risk of trimethoprim-sulfamethoxazole hepatotoxicity is increased in persons with HIV infection.

HMG-CoA Reductase Inhibitors (Statins) (Idiosyncratic Mixed Hepatocellular and Cholestatic Reaction)

Between 1 and 2% of patients taking lovastatin, simvastatin, pravastatin, fluvastatin, or one of the newer statin drugs for the treatment of hypercholesterolemia experience asymptomatic, reversible elevations (>threefold) of

aminotransferase activity. Acute hepatitis-like histologic changes, centrilobular necrosis, and centrilobular cholestasis have been described in several cases. In a larger proportion, minor aminotransferase elevations appear during the first several weeks of therapy. Careful laboratory monitoring can distinguish between patients with minor, transitory changes; who may continue therapy; and those with more profound and sustained abnormalities, who should discontinue therapy. Because clinically meaningful aminotransferase elevations are so rare after statin use and do not differ in meta-analyses from the frequency of such laboratory abnormalities in placebo recipients, a panel of liver experts recommended to the National Lipid Association's Safety Task Force that liver-test monitoring was not necessary in patients treated with statins and that statin therapy need not be discontinued in patients found to have asymptomatic isolated aminotransferase elevations during therapy. Statin hepatotoxicity is not increased in patients with chronic hepatitis C, hepatic steatosis, or other underlying liver diseases, and statins can be used safely in these patients.

Total Parenteral Nutrition (Steatosis, Cholestasis)

Total parenteral nutrition (TPN) is often complicated by cholestatic hepatitis attributable to either steatosis, cholestasis, or gallstones (or gallbladder sludge). Steatosis or steatohepatitis may result from the excess carbohydrate calories in these nutritional supplements and is the predominant form of TPN-associated liver disorder in adults. The frequency of this complication has been reduced substantially by the introduction of balanced TPN formulas that rely on lipid as an alternative caloric source. Cholestasis and cholelithiasis, caused by the absence of stimulation of bile flow and secretion resulting from the lack of oral intake, is the predominant form of TPN-associated liver disease in infants, especially in premature neonates. Often, cholestasis in such neonates is multifactorial, contributed to by other factors such as sepsis, hypoxemia, and hypotension; occasionally, TPN-induced cholestasis in neonates culminates in chronic liver disease and liver failure. When TPN-associated liver-test abnormalities occur in adults, balancing the TPN formula with more lipid is the intervention of first recourse. In infants with TPN-associated cholestasis, the addition of oral feeding may ameliorate the problem. Therapeutic interventions suggested, but not shown, to be of proven benefit, include cholecystokinin, ursodeoxycholic acid, *S*-adenosyl methionine, and taurine.

"Alternative and Complementary Medicines" (Idiosyncratic Hepatitis, Steatosis)

The misguided popularity of herbal medications that are of scientifically unproven efficacy and that lack prospective safety oversight by regulatory agencies has resulted in occasional instances of hepatotoxicity. Included among the herbal remedies associated with toxic hepatitis are Jin Bu Huan, xiao-chai-hu-tang, germander, chaparral, senna, mistletoe, skullcap, gentian, comfrey (containing pyrrolizidine alkaloids), Ma huang, bee pollen, valerian root, pennyroyal oil, kava, celandine, Impila (*Callilepsis laureaola*), LipoKinetix, Hyroxycut, herbal nutritional supplements, and herbal teas. Well characterized are the acute hepatitis-like histologic lesions following Jin Bu Huan use: focal hepatocellular necrosis, mixed mononuclear portal tract infiltration, coagulative necrosis, apoptotic hepatocyte degeneration, tissue eosinophilia, and microvesicular steatosis. Megadoses of vitamin A can injure the liver, as can pyrrolizidine alkaloids, which often contaminate Chinese herbal preparations and can cause a venoocclusive injury leading to sinusoidal hepatic vein obstruction. Because some alternative medicines induce toxicity via active metabolites, alcohol and drugs that stimulate cytochrome P450 enzymes may enhance the toxicity of some of these products. Conversely, some alternative medicines also stimulate cytochrome P450 and may result in or amplify the toxicity of recognized drug hepatotoxins. Given the widespread use of such poorly defined herbal preparations, hepatotoxicity is likely to be encountered with increasing frequency; therefore, a drug history in patients with acute and chronic liver disease should include use of "alternative medicines" and other nonprescription preparations sold in so-called health food stores.

Highly Active Antiretroviral Therapy (Haart) for HIV Infection (Mitochondrial Toxic, Idiosyncratic, Steatosis; Hepatocellular, Cholestatic, and Mixed)

The recognition of drug hepatotoxicity in persons with HIV infection is complicated in this population by the many alternative causes of liver injury (chronic viral hepatitis, fatty infiltration, infiltrative disorders, mycobacterial infection, etc.), but drug hepatotoxicity associated with HAART is an emerging and common type of liver injury in HIV-infected persons **(Chap. 189)**. Although no one antiviral agent is recognized as a potent hepatotoxin, combination regimens including reverse transcriptase and protease inhibitors cause hepatotoxicity in ~10% of treated patients. Implicated most frequently are combinations including nucleoside analogue reverse transcriptase inhibitors zidovudine, didanosine, and, to a lesser extent, stavudine; protease inhibitors ritonavir and indinavir (and amprenavir when used together with ritonavir) as well as tipranavir; and nonnucleoside reverse transcriptase inhibitors nevirapine and, to a lesser extent, efavirenz. These drugs cause predominantly hepatocellular injury but cholestatic injury as well, and prolonged (>6 months) use of reverse transcriptase inhibitors has been associated with mitochondrial injury, steatosis, and lactic acidosis. Indirect hyperbilirubinemia, resulting from direct inhibition of bilirubin-conjugating activity by UDP-glucuronosyltransferase, usually without elevation of aminotransferase or alkaline phosphatase activities, occurs in ~10% of patients treated with the protease inhibitor indinavir. Distinguishing the impact of HAART hepatotoxicity in patients with HIV and hepatitis virus co-infection is made challenging by the following: (1) both chronic hepatitis B and hepatitis C can affect the natural history of HIV infection and the response to HAART, and (2) HAART can have an impact on chronic viral hepatitis. For example, immunologic reconstitution with HAART can result in immunologically mediated liver-cell injury in patients with chronic hepatitis B co-infection if treatment with an antiviral agent for hepatitis B, e.g., the nucleoside analogue lamivudine, is withdrawn or if nucleoside analogue resistance emerges. Infection with HIV, especially with low CD4+ T cell counts, has been reported to increase the rate of hepatic fibrosis associated with chronic hepatitis C, and HAART therapy can increase levels of serum aminotransferases and hepatitis C virus RNA in patients with hepatitis C co-infection. Didanosine or stavudine should not be used with ribavirin in patients with HIV/hepatitis C virus co-infection, because of an increased risk of severe mitochondrial toxicity and lactic acidosis.

Acknowledgement

Kurt J. Isselbacher, MD, contributed to this chapter in previous editions of Harrison's.

FURTHER READINGS

Chalasani et al: Causes, clinical features, and outcomes from a prospective study of drug-induced liver injury in the United States. Gastroenterology 135:1924, 2008[PMID: 18955056] [Full Text]

Chang CY, Schiano TD: Review article: drug hepatotoxicity. Aliment Pharmacol Ther 25:1135, 2007[PMID: 17451560] [Full Text]

Daly AK et al: *HLA-B*5701* genotype is a major determinant of drug-induced liver injury due to flucloxacillin. Nature Genetics 41:816, 2009[PMID: 19483685] [Full Text]

Kaplowitz N, Deleve LD (eds): *Drug-Induced Liver Disease* . 2nd ed, New York, Informa Healthcare, 2007

Lee WM: Drug-induced hepatotoxicity. N Engl J Med 349:474, 2003[PMID: 12890847] [Full Text]

Lucena MI et al: Mitochondrial superoxide dismutase and glutathione peroxidase in idiosyncratic drug-induced liver injury. Hepatology 52:303, 2010[PMID: 20578157] [Full Text]

Mindikoglu AL et al: Outcome of liver transplantation for drug-induced acute liver failure in the United States: Analysis of the United Network for Organ Sharing database. Liver Transpl 15:719, 2009[PMID: 19562705] [Full Text]

Navarro VJ, Senior JR: Drug-related hepatotoxicity. N Engl J Med 354:731, 2006[PMID: 16481640] [Full Text]

Nunez M: Hepatotoxicity of antiretrovirals: Incidence, mechanisms and management. J Hepatol 44:S132, 2006

Schmidt LE et al: Acute versus chronic alcohol consumption in acetaminophen-induced hepatotoxicity. Hepatology 35:876, 2002[PMID: 11915034] [Full Text]

Zhang J et al: Modulation of acetaminophen-induced hepatotoxicity by the xenobiotic receptor CAR. Science 298:422, 2002[PMID: 12376703] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHRONIC HEPATITIS: INTRODUCTION

Chronic hepatitis represents a series of liver disorders of varying causes and severity in which hepatic inflammation and necrosis continue for at least 6 months. Milder forms are nonprogressive or only slowly progressive, while more severe forms may be associated with scarring and architectural reorganization, which, when advanced, lead ultimately to cirrhosis. Several categories of chronic hepatitis have been recognized. These include chronic viral hepatitis, druginduced chronic hepatitis **(Chap. 305)**, and autoimmune chronic hepatitis. In many cases, clinical and laboratory features are insufficient to allow assignment into one of these three categories; these "idiopathic" cases are also believed to represent autoimmune chronic hepatitis. Finally, clinical and laboratory features of chronic hepatitis are observed occasionally in patients with such hereditary/metabolic disorders as Wilson's disease (copper overload) **(Chaps. 308 and 360)** and nonalcoholic fatty liver disease **(Chap. 309)** and even occasionally in patients with alcoholic liver injury **(Chap. 307)**. Although all types of chronic hepatitis share certain clinical, laboratory, and histopathologic features, chronic viral and chronic autoimmune hepatitis are sufficiently distinct to merit separate discussions. **For discussion of acute hepatitis, see Chap. 304**.

CLASSIFICATION OF CHRONIC HEPATITIS

Common to all forms of chronic hepatitis are histopathologic distinctions based on localization and extent of liver injury. These vary from the milder forms, previously labeled *chronic persistent hepatitis* and *chronic lobular hepatitis*, to the more severe form, formerly called *chronic active hepatitis*. When first defined, these designations were believed to have prognostic implications, which have been challenged by more recent observations. Categorization of chronic hepatitis based primarily on histopathologic features has been replaced by a more informative classification based on a combination of clinical, serologic, and histologic variables. Classification of chronic hepatitis is based on (1) its *cause*; (2) its histologic activity, or *grade*; and (3) its degree of progression, or *stage*. Thus, neither clinical features alone nor histologic features—requiring liver biopsy—alone are sufficient to characterize and distinguish among the several categories of chronic hepatitis.

Classification by Cause

Clinical and serologic features allow the establishment of a diagnosis of *chronic viral hepatitis*, caused by hepatitis B, hepatitis B plus D, or hepatitis C; *autoimmune hepatitis*, including several subcategories, I and II (perhaps III), based on serologic distinctions; *drug-associated chronic hepatitis*; and a category of unknown cause, or *cryptogenic chronic hepatitis***(Table 306-1)**. These are addressed in more detail below.

Table 306-1 Clinical and Laboratory Features of Chronic Hepatitis

Type of hepatitis	Diagnostic test(s)	Autoantibodies	Therapy
Chronic hepatitis B	HBsAg, IgG anti-HBc, HBeAg, HBV DNA	Uncommon	IFN-α, PEG IFN-α lamivudine adefovir entecavir telbivudine tenofovir
Chronic hepatitis C	Anti-HCV, HCV RNA	Anti-LKM1[a]	PEG IFN-α ribavirin Telaprevir[d] Boceprevir[d]
Chronic hepatitis D	Anti-HDV, HDV RNA, HBsAg, IgG anti-HBc	Anti-LKM3	IFN-α, PEG IFN-α[c]
Autoimmune hepatitis	ANA[b] (homogeneous), anti-LKM1 (±) Hyperglobulinemia	ANA, anti-LKM1 anti-SLA[e]	Prednisone, azathioprine
Drug-associated	—	Uncommon	Withdraw drug
Cryptogenic	All negative	None	Prednisone (?), azathioprine (?)

[a] Antibodies to liver-kidney microsomes type 1 (autoimmune hepatitis type II and some cases of hepatitis C)

[b] Antinuclear antibody (autoimmune hepatitis type I)

[c] Clinical trials suggest benefit of IFN-α therapy; PEG IFN-α is as effective, if not more so.

[d] Expected approval date 2011.

[e] Antibodies to soluble liver antigen (autoimmune hepatitis type III)

Abbreviations: HBc, hepatitis B core; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis D virus; IFN-α, interferon-α; IgG, immunoglobulin G; LKM, liver-kidney microsome;

PEG-IFN-α, pegylated interferon-α; SLA, soluble liver antigen.

Classification by Grade

Grade, a histologic assessment of necroinflammatory activity, is based on examination of the liver biopsy. An assessment of important histologic features includes the degree of *periportal necrosis* and the disruption of the limiting plate of periportal hepatocytes by inflammatory cells (so-called *piecemeal necrosis* or *interface hepatitis*); the degree of confluent necrosis that links or forms bridges between vascular structures —between portal tract and portal tract or even more important bridges between portal tract and central vein—referred to as *bridging necrosis*; the degree of hepatocyte degeneration and focal necrosis within the lobule; and the degree of *portal inflammation*. Several scoring systems that take these histologic features into account have been devised, and the most popular are the histologic activity index (HAI), used commonly in the United States, and the METAVIR score, used in Europe **(Table 306-2)**. Based on the presence and degree of these features of histologic activity, chronic hepatitis can be graded as mild, moderate, or severe.

Table 306-2 Histologic Grading and Staging of Chronic Hepatitis

		Histologic Activity Index (HAI)[a]		METAVIR[b]	
Histologic Feature		**Severity**	**Score**	**Severity**	**Score**
Necroinflammatory Activity (grade)					
Periportal necrosis, including piecemeal necrosis and/or bridging necrosis (BN)		None	0	None	0
		Mild	1	Mild	1
		Mild/Moderate	2	Moderate	2
		Moderate	3	Severe	3
		Severe	4		

				Bridging necrosis	Yes No
Intralobular necrosis	Confluent	–none –focal –Zone 3 some –Zone 3 most –Zone 3 + BN few –Zone 3 + BN multiple –Panacinar/multiacinar	0 1 2 3 4 5 6	None or mild Moderate Severe	0 1 2
	Focal	–none –≤1 focus/10x field –2–4 foci/10x field –5–10 foci/10x field –>10 foci/10x field	0 1 2 3 4		
Portal Inflammation		None Mild Moderate Moderate/marked Marked Total	0 1 2 3 4 0–18		A0–A3[c]
Fibrosis (stage)					
None			0		F0
Portal fibrosis–some			1		F1
Portal fibrosis–most			2		F1
Bridging fibrosis–few			3		F2
Bridging fibrosis–many			4		F3
Incomplete cirrhosis			5		F4
Cirrhosis			6		F4
		Total	6		4

[a] J Hepatol 22:696, 1995

[b] Hepatology 24:289, 1996

[c] Necroinflammatory grade: A0 = none; A1 = mild; A2 = moderate; A3 = severe

Classification by Stage

The stage of chronic hepatitis, which reflects the level of progression of the disease, is based on the degree of hepatic fibrosis. When fibrosis is so extensive that fibrous septa surround parenchymal nodules and alter the normal architecture of the liver lobule, the histologic lesion is defined as *cirrhosis*. Staging is based on the degree of fibrosis as categorized on a numerical scale from 0–6 (HAI) or 0–4 (METAVIR) (Table 306-2).

CHRONIC VIRAL HEPATITIS

Both the enterically transmitted forms of viral hepatitis, hepatitis A and E, are self-limited and do not cause chronic hepatitis (rare reports notwithstanding in which acute hepatitis A serves as a trigger for the onset of autoimmune hepatitis in genetically susceptible patients). In contrast, the entire clinicopathologic spectrum of chronic hepatitis occurs in patients with chronic viral hepatitis B and C as well as in patients with chronic hepatitis D superimposed on chronic hepatitis B.

Chronic Hepatitis B

The likelihood of chronicity after acute hepatitis B varies as a function of age. Infection at birth is associated with clinically silent acute infection but a 90% chance of chronic infection, while infection in young adulthood in immunocompetent persons is typically associated with clinically apparent acute hepatitis but a risk of chronicity of only approximately 1%. Most cases of chronic hepatitis B among adults, however, occur in patients who never had a recognized episode of clinically apparent acute viral hepatitis. The degree of liver injury (grade) in patients with chronic hepatitis B is variable, ranging from none in inactive carriers to mild to moderate to severe. Among adults with chronic hepatitis B, histologic features are of prognostic importance. In one long-term study of patients with chronic hepatitis B, investigators found a 5-year survival rate of 97% for patients with mild chronic hepatitis, 86% for patients with moderate to severe chronic hepatitis, and only 55% for patients with chronic hepatitis and postnecrotic cirrhosis. The 15-year survival in these cohorts was 77, 66, and 40%, respectively. On the other hand, more recent observations do not allow us to be so sanguine about the prognosis in patients with mild chronic hepatitis; among such patients followed for 1–13 years, progression to more severe chronic hepatitis and cirrhosis has been observed in more than a quarter of cases.

More important to consider than histology alone in patients with chronic hepatitis B is the degree of hepatitis B virus (HBV) replication. As reviewed in **Chap. 304**, chronic HBV infection can occur in the presence or absence of serum hepatitis B e antigen (HBeAg), and generally, for both HBeAg-reactive and HBeAgnegative chronic hepatitis B, the level of HBV DNA correlates with the level of liver injury and risk of progression. In *HBeAg-reactive chronic hepatitis B*, two phases have been recognized based on the relative level of HBV replication. The relatively *replicative phase* is characterized by the presence in the serum of HBeAg and HBV DNA levels well in excess of 10^5 –10^6 virions/mL, by the presence in the liver of detectable intrahepatocyte nucleocapsid antigens [primarily hepatitis B core antigen (HBcAg)], by high infectivity, and by accompanying liver injury. In contrast, the relatively *nonreplicative phase* is characterized by the absence of the conventional serum marker of HBV replication (HBeAg), the appearance of anti-HBe, levels of HBV DNA below a threshold of ~10^3 virions/mL, the absence of intrahepatocytic HBcAg, limited infectivity, and minimal liver injury. Those patients in the replicative phase tend to have more severe chronic hepatitis, while those in the nonreplicative phase tend to have minimal or mild chronic hepatitis or to be inactive hepatitis B carriers; however, distinctions in HBV replication and in histologic category do not always coincide. The likelihood in a patient with HBeAg-reactive chronic hepatitis B of converting spontaneously from relatively replicative to nonreplicative infection is approximately 10–15% per year. In patients with HBeAg-reactive chronic HBV infection, especially when acquired at birth or in early childhood, as recognized commonly in Asian countries, a dichotomy is common between very high levels of HBV replication and negligible levels of liver injury. Yet despite the relatively immediate, apparently benign nature of liver disease for many decades in this population, patients with childhood-acquired HBV infection are the ones at ultimately increased risk later in life of cirrhosis and hepatocellular carcinoma (HCC) **(Chap. 92)**. **A discussion of the pathogenesis of liver injury in patients with chronic hepatitis B appears in Chap. 304.**

HBeAg-negative chronic hepatitis B [i.e., chronic HBV infection with active virus replication, readily detectable HBV DNA but without HBeAg (anti-HBe-reactive)], is more common than HBeAg-reactive chronic hepatitis B in Mediterranean and European countries and in Asia (and, correspondingly, in HBV genotypes other than A). Compared to patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative

chronic hepatitis B have levels of HBV DNA that are several orders of magnitude lower (no more than 10^5 –10^6 virions/mL) than those observed in the HBeAg-reactive subset. Most such cases represent precore or core-promoter mutations acquired late in the natural history of the disease (mostly early-life onset; age range 40–55 years, older than that for HBeAg-reactive chronic hepatitis B); these mutations prevent translation of HBeAg from the precore component of the HBV genome (precore mutants) or are characterized by down-regulated transcription of precore mRNA (core-promoter mutants; **Chap. 304**). Although their levels of HBV DNA tend to be lower than among patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative chronic hepatitis B can have progressive liver injury (complicated by cirrhosis and HCC) and experience episodic reactivation of liver disease reflected in fluctuating levels of aminotransferase activity ("flares"). The biochemical and histologic activity of HBeAg-negative disease tends to correlate closely with levels of HBV replication, unlike the case mentioned above of Asian patients with HBeAg-reactive chronic hepatitis B during the early decades of their HBV infection. An important point worth reiterating is the observation that the level of HBV replication is the most important risk factor for the ultimate development of cirrhosis and HCC in both HBeAg-reactive and HBeAg-negative patients. Although levels of HBV DNA are lower and more readily suppressed by therapy to undetectable levels in HBeAg-negative (compared to HBeAg-reactive) chronic hepatitis B, achieving sustained responses that permit discontinuation of antiviral therapy is less likely in HBeAg-negative patients (see below). Inactive carriers are patients with circulating hepatitis B surface antigen (HBsAg), normal serum aminotransferase levels, undetectable HBeAg, and levels of HBV DNA that are either undetectable or present at levels $\leq 10^3$ virions/mL. This serologic profile can occur not only in inactive carriers but also in patients with HBeAg-negative chronic hepatitis B during periods of relative inactivity; distinguishing between the two requires sequential biochemical and virologic monitoring over many months.

The spectrum of *clinical features* of chronic hepatitis B is broad, ranging from asymptomatic infection to debilitating disease or even end-stage, fatal hepatic failure. As noted above, the onset of the disease tends to be insidious in most patients, with the exception of the very few in whom chronic disease follows failure of resolution of clinically apparent acute hepatitis B. **The clinical and laboratory features associated with progression from acute to chronic hepatitis B are discussed in Chap. 304.**

Fatigue is a common symptom, and persistent or intermittent *jaundice* is a common feature in severe or advanced cases. Intermittent deepening of jaundice and recurrence of malaise and anorexia, as well as worsening fatigue, are reminiscent of acute hepatitis; such exacerbations may occur spontaneously, often coinciding with evidence of virologic reactivation; may lead to progressive liver injury; and, when superimposed on well-established cirrhosis, may cause hepatic decompensation. Complications of cirrhosis occur in end-stage chronic hepatitis and include ascites, edema, bleeding gastroesophageal varices, hepatic encephalopathy, coagulopathy, or hypersplenism. Occasionally, these complications bring the patient to initial clinical attention. Extrahepatic complications of chronic hepatitis B, similar to those seen during the prodromal phase of acute hepatitis B, are associated with deposition of circulating hepatitis B antigen–antibody immune complexes. These include arthralgias and arthritis, which are common, and the more rare purpuric cutaneous lesions (leukocytoclastic vasculitis), immune-complex glomerulonephritis, and generalized vasculitis (polyarteritis nodosa) **(Chaps. 304 and 326)**.

Laboratory features of chronic hepatitis B do not distinguish adequately between histologically mild and severe hepatitis. Aminotransferase elevations tend to be modest for chronic hepatitis B but may fluctuate in the range of 100–1000 units. As is true for acute viral hepatitis B, alanine aminotransferase (ALT) tends to be more elevated than aspartate aminotransferase (AST); however, once cirrhosis is established, AST tends to exceed ALT. Levels of alkaline phosphatase activity tend to be normal or only marginally elevated. In severe cases, moderate elevations in serum bilirubin [51.3–171 μmol/L (3–10 mg/dL)] occur. Hypoalbuminemia and prolongation of the prothrombin time occur in severe or end-stage cases. Hyperglobulinemia and detectable circulating autoantibodies are distinctly absent in chronic hepatitis B (in contrast to autoimmune hepatitis). **Viral markers of chronic HBV infection are discussed in Chaps. 304.**

TREATMENT: CHRONIC HEPATITIS B

Although progression to cirrhosis is more likely in severe than in mild or moderate chronic hepatitis B, all forms of chronic hepatitis B can be progressive, and progression occurs primarily in patients with active HBV replication. Moreover, in populations of patients with chronic hepatitis B who are at risk for HCC **(Chap. 92)**, the risk is highest for those with continued, high-level HBV replication and lower for persons in whom initially high-level HBV DNA falls spontaneously over time. Therefore, management of chronic hepatitis B is directed at suppressing the level of virus replication. Although clinical trials tend to focus on clinical endpoints achieved over 1–2 years (e.g., suppression of HBV DNA to undetectable levels, loss of HBeAg/HBsAg, improvement in histology, normalization of ALT), these short-term gains translate into reductions in the risk of clinical progression, hepatic decompensation, and death. To date, seven drugs have been approved for treatment of chronic hepatitis B: injectable interferon (IFN) α pegylated interferon [long-acting IFN bound to polyethylene glycol (PEG), known as *PEG IFN*]; and the oral agents lamivudine, adefovir dipivoxil, entecavir, telbivudine, and tenofovir.

Antiviral therapy for hepatitis B has evolved rapidly since the mid-1990s, as has the sensitivity of tests for HBV DNA. When IFN and lamivudine were evaluated in clinical trials, HBV DNA was measured by insensitive hybridization assays with detection thresholds of 10^5 –10^6 virions/mL; when adefovir, entecavir, telbivudine, tenofovir, and PEG IFN were studied in clinical trials, HBV DNA was measured by sensitive amplification assays [polymerase chain reaction [(PCR)] with detection thresholds of 10^1 –10^3 virions/mL. Recognition of these distinctions is helpful when comparing results of clinical trials that established the efficacy of these therapies (reviewed below in chronological order of publication of these efficacy trials).

Interferon

IFN-α was the first approved therapy for chronic hepatitis B. Although it is no longer used to treat hepatitis B, standard IFN is important historically, having provided important lessons about antiviral therapy in general. For immunocompetent adults with HBeAg-reactive chronic hepatitis B [who tend to have high-level HBV DNA (>10^5 –10^6 virions/mL) and histologic evidence of chronic hepatitis on liver biopsy], a 16-week course of IFN given subcutaneously at a daily dose of 5 million units, or three times a week at a dose of 10 million units, results in a loss of HBeAg and hybridization-detectable HBV DNA (i.e., a reduction to levels below 10^5 –10^6 virions/mL) in ~30% of patients, with a concomitant improvement in liver histology. Seroconversion from HBeAg to anti-HBe occurs in approximately 20%, and, in early trials, approximately 8% lost HBsAg. Successful IFN therapy and seroconversion are often accompanied by an acute hepatitis-like elevation in aminotransferase activity, which has been postulated to result from enhanced cytolytic T cell clearance of HBV-infected hepatocytes. Relapse after successful therapy is rare (1 or 2%). The likelihood of responding to IFN is higher in patients with lower levels of HBV DNA and substantial elevations of ALT. Although children can respond as well as adults, IFN therapy has not been effective in very young children infected at birth. Similarly, IFN therapy has not been effective in immunosuppressed persons, Asian patients with minimal-to-mild ALT elevations, or patients with decompensated chronic hepatitis B (in whom such therapy can actually be detrimental, sometimes precipitating decompensation, often associated with severe adverse effects). Among patients with HBeAg loss during therapy, long-term follow-up has demonstrated that 80% experience eventual loss of HBsAg [i.e., all serologic markers of infection, and normalization of ALT over a 9-year posttreatment period]. In addition, improved longterm and complication-free survival as well as a reduction in the frequency of HCC have been documented among interferon responders, supporting the conclusion that successful interferon therapy improves the natural history of chronic hepatitis B.

Initial trials of brief-duration IFN therapy in patients with *HBeAg-negative chronic hepatitis B* were disappointing, suppressing HBV replication transiently during therapy but almost never resulting in sustained antiviral responses. In subsequent IFN trials among patients with HBeAg-negative chronic hepatitis B, however, more protracted courses, lasting up to 11/2 years, have been reported to result in sustained remissions documented to last for several years, with suppressed HBV DNA and aminotransferase activity, in ~20%.

Complications of IFN therapy include systemic "flu-like" symptoms; marrow suppression; emotional lability (irritability, depression, anxiety); autoimmune reactions (especially autoimmune thyroiditis); and miscellaneous side effects such as alopecia, rashes, diarrhea, and numbness and tingling of the extremities. With the possible exception of autoimmune thyroiditis, all these side effects are reversible upon dose lowering or cessation of therapy.

Although no longer competitive with the newer generation of antivirals, IFN did represent the first successful antiviral approach and set a standard against which to measure subsequent drugs in the achievement of durable virologic, serologic, biochemical, and histologic responses; consolidation of virologic and biochemical benefit in the ensuing years after therapy; and improvement in the natural history of chronic hepatitis B. Standard IFN has been supplanted by long-acting PEG IFN (see below), and IFN nonresponders are now treated with one of the newer oral nucleoside analogues.

Lamivudine

The first of the nucleoside analogues to be approved, the dideoxynucleoside lamivudine, inhibits reverse transcriptase activity of both HIV and HBV and is a potent and effective agent for patients with chronic hepatitis B. Although generally superseded by newer, more potent agents, lamivudine is still used in regions of the world where newer agents are not yet approved are or not affordable. In clinical trials among patients with HBeAg-reactive chronic hepatitis B, lamivudine therapy at daily doses of 100 mg for 48–52 weeks suppressed HBV DNA by a median of

approximately 5.5 $\log_{10}$ copies/mL and to undetectable levels, as measured by PCR amplification assays, in approximately 40% of patients. Therapy was associated with HBeAg loss in 32–33%; HBeAg seroconversion (i.e., conversion from HBeAg-reactive to anti-HBe-reactive) in 16–21%; normalization of ALT in 40–75%; improvement in histology in 50–60%; retardation in fibrosis in 20–30%; and prevention of progression to cirrhosis. HBeAg responses can occur even in subgroups who are resistant to IFN (e.g., those with high-level HBV DNA) or who failed in the past to respond to it. As is true for IFN therapy of chronic hepatitis B, patients with near-normal ALT activity tend not to experience HBeAg responses (despite suppression of HBV DNA), and those with ALT levels exceeding five times the upper limit of normal can expect 1-year HBeAg seroconversion rates of 50–60%. Generally, HBeAg seroconversions are confined to patients who achieve suppression of HBV DNA to $<10^4$ genomes/mL. Among patients who undergo HBeAg responses during a year-long course of therapy and in whom the response is sustained for 4–6 months after cessation of therapy, the response is durable thereafter in the vast majority, >80%; therefore, the achievement of an HBeAg response represents a viable stopping point in therapy. Reduced durability has been reported in some Asian experiences; however, in most western and Asian patient study populations, long-term durability of HBeAg responses is the rule, which, at least in western patients, is accompanied by a posttreatment HBsAg seroconversion rate comparable to that seen after IFN-induced HBeAg responses. To support the durability of HBeAg responses, patients receive a period of consolidation therapy (at least 6 months in western patients, at least 1 year in Asian patients) after HBeAg seroconversion; close posttreatment monitoring is necessary to identify HBV reactivation promptly and to resume therapy. If HBeAg is unaffected by lamivudine therapy, the current approach is to continue therapy until an HBeAg response occurs, but longterm therapy may be required to suppress HBV replication and, in turn, limit liver injury; HBeAg seroconversions can increase to a level of 50% after 5 years of therapy. Histologic improvement continues to accrue with therapy beyond the first year; after a cumulative course of 3 years of lamivudine therapy, necroinflammatory activity is reduced in the majority of patients, and even cirrhosis has been shown to regress to precirrhotic stages.

Losses of HBsAg have been few during the first year of lamivudine therapy, and this observation had been cited as an advantage of IFN-based over lamivudine therapy; however, in head-to-head comparisons between standard IFN and lamivudine monotherapy, HBsAg losses were rare in both groups. Trials in which lamivudine and IFN were administered in combination failed to show a benefit of combination therapy over lamivudine monotherapy for either treatment-naïve patients or prior IFN nonresponders.

In patients with *HBeAg-negative chronic hepatitis B* (i.e., in those with precore and core-promoter HBV mutations), 1 year of lamivudine therapy results in HBV DNA suppression and normalization of ALT in three-quarters of patients and in histologic improvement in approximately two-thirds. Therapy has been shown to suppress HBV DNA by approximately 4.5 $\log_{10}$ copies/mL (baseline HBV DNA levels are lower than in patients with HBeAg-reactive hepatitis B) and to undetectable levels in approximately 70%, as measured by sensitive PCR amplification assays. Lacking HBeAg at the outset, patients with HBeAg-negative chronic hepatitis B cannot achieve an HBeAg response—a stopping point in HBeAg-reactive patients; almost invariably, when therapy is discontinued, reactivation is the rule. Therefore, these patients require long-term therapy; with successive years, the proportion with suppressed HBV DNA and normal ALT increases.

Clinical and laboratory side effects of lamivudine are negligible, indistinguishable from those observed in placebo recipients. Still, lamivudine doses should be reduced in patients with reduced creatinine clearance. During lamivudine therapy, transient ALT elevations, resembling those seen during IFN therapy and during spontaneous HBeAg-to-anti-HBe seroconversions, occur in one-fourth of patients. These ALT elevations may result from restored cytolytic T cell activation permitted by suppression of HBV replication. Similar ALT elevations, however, occur at an identical frequency in placebo recipients, but ALT elevations associated with HBeAg seroconversion are confined to lamivudine-treated patients. When therapy is stopped after a year of therapy, two- to threefold ALT elevations occur in 20–30% of lamivudine-treated patients, representing renewed liver-cell injury as HBV replication returns. Although these posttreatment flares are almost always transient and mild, rare severe exacerbations, especially in cirrhotic patients, have been observed, mandating close and careful clinical and virologic monitoring after discontinuation of treatment. Many authorities caution against discontinuing therapy in patients with cirrhosis, in whom posttreatment flares could precipitate decompensation.

Long-term monotherapy with lamivudine is associated with methionine-to-valine (M204V) or methionine-to- isoleucine (M204I) mutations, primarily at amino acid 204 in the tyrosine-methionine-aspartate-aspartate (YMDD) motif of HBV DNA polymerase, analogous to mutations that occur in HIV-infected patients treated with this drug. During a year of therapy, YMDD mutations occur in 15–30% of patients; the frequency increases with each year of therapy, reaching 70% at year 5. Ultimately, patients with YMDD mutants experience degradation of clinical, biochemical, and histologic responses; therefore, if treatment is begun with lamivudine monotherapy, the emergence of lamivudine resistance, reflected clinically by a breakthrough from suppressed levels of HBV DNA and ALT, is managed by adding another antiviral to which YMDD variants are sensitive (e.g., adefovir, tenofovir; see below).

Currently, although lamivudine is very safe and still used widely in other parts of the world, in the United States and Europe, lamivudine has been eclipsed by more potent antivirals that have superior resistance profiles (see below). Still, as the first successful oral antiviral agent for use in hepatitis B, lamivudine has provided proof of the concept that polymerase inhibitors can achieve virologic, serologic, biochemical, and histologic benefits. In addition, lamivudine has been shown to be effective in the treatment of patients with decompensated hepatitis B (for whom IFN is contraindicated), in some of whom decompensation can be reversed. Moreover, among patients with cirrhosis or advanced fibrosis, lamivudine has been shown to be effective in reducing the risk of progression to hepatic decompensation and, marginally, the risk of HCC.

Because lamivudine monotherapy can result universally in the rapid emergence of YMDD variants in persons with HIV infection, patients with chronic hepatitis B should be tested for anti-HIV prior to therapy; if HIV infection is identified, lamivudine monotherapy at the HBV daily dose of 100 mg is contraindicated. These patients should be treated for both HIV and HBV with an HIV drug regimen that includes or is supplemented by at least two drugs active against HBV; highly active antiretroviral therapy (HAART) often contains two drugs with antiviral activity against HBV (e.g., tenofovir and emtricitabine), but if lamivudine is part of the regimen, the daily dose should be 300 mg **(Chap. 189)**. The safety of lamivudine during pregnancy has not been established; however, the drug is not teratogenic in rodents and has been used safely in pregnant women with HIV infection and with HBV infection. Limited data even suggest that administration of lamivudine during the last months of pregnancy to mothers with high-level hepatitis B viremia ($\geq 10^8$ IU/ml) can reduce the likelihood of perinatal transmission of hepatitis B.

Adefovir Dipivoxil

At an oral daily dose of 10 mg, the acyclic nucleotide analogue adefovir dipivoxil, the prodrug of adefovir, reduces HBV DNA by approximately 3.5–4 $\log_{10}$ copies/mL and is equally effective in treatment-naïve patients and IFN nonresponders. In HBeAg-reactive chronic hepatitis B, a 48-week course of adefovir dipivoxil was shown to achieve histologic improvement (and reduce the progression of fibrosis) and normalization of ALT in just over one-half of patients, HBeAg seroconversion in 12%, HBeAg loss in 23%, and suppression to an undetectable level of HBV DNA in 13–21%, as measured by PCR. Similar to IFN and lamivudine, adefovir dipivoxil is more likely to achieve an HBeAg response in patients with high baseline ALT (e.g., among adefovir-treated patients with ALT level >5 times the upper limit of normal), HBeAg seroconversions occurred in 25%. The durability of adefovir-induced HBeAg responses is high (91% in one study); therefore, HBeAg response can be relied upon as a stopping point for adefovir therapy, after a period of consolidation therapy, as outlined above. Although data on the impact of additional therapy beyond 1 year are limited, biochemical, serologic, and virologic outcomes improve progressively as therapy is continued.

In patients with *HBeAg-negative chronic hepatitis B*, a 48-week course of 10 mg/d of adefovir dipivoxil resulted in histologic improvement in two-thirds, normalization of ALT in threefourths, and suppression of HBV DNA to PCR-undetectable levels in one-half to two-thirds. As was true for lamivudine, because HBeAg responses—a potential stopping point—cannot be achieved in this group, reactivation is the rule when adefovir therapy is discontinued, and indefinite, long-term therapy is required. Treatment beyond the first year consolidates the gain of the first year; after 5 years of therapy, improvement in hepatic inflammation and regression of fibrosis was observed in threefourths of patients, ALT was normal in 70%, and HBV DNA was undetectable in almost 70%.

Adefovir contains a flexible acyclic linker instead of the L-nucleoside ring of lamivudine, avoiding steric hindrance by mutated amino acids. In addition, the molecular structure of phosphorylated adefovir is very similar to that of its natural substrate; therefore mutations to adefovir would also affect binding of the natural substrate, dATP. Hypothetically, these are among the reasons that resistance to adefovir dipivoxil is much less likely than resistance to lamivudine; no resistance was encountered in 1 year of clinical-trial therapy. In subsequent years, however, adefovir resistance begins to emerge [asparagine to threonine at amino acid 236 (N236T) and alanine to valine or threonine at amino acid 181 (A181V/T), primarily], occurring in 2.5% after 2 years, but in 29% after 5 years of therapy (reported in HBeAg-negative patients). Among patients coinfected with HBV and HIV and who have normal CD4+ T cell counts, adefovir dipivoxil is effective in suppressing HBV dramatically (by 5 logs 10 in one study). Moreover, adefovir dipivoxil is effective in lamivudine-resistant, YMDD-mutant HBV and can be used when such lamivudine-induced variants emerge. When lamivudine resistance occurs, adding adefovir (i.e., maintaining lamivudine to preempt the emergence of adefovir resistance), is superior to switching to adefovir. Almost invariably, patients with adefovir-mutant HBV respond to lamivudine (or newer agents, such as entecavir, see below). When, in the past, adefovir had been evaluated as therapy for HIV infection, doses of 60–120 mg were required to suppress HIV, and, at these doses, the drug was nephrotoxic. Even at 30 mg/d, creatinine elevations of 44 μmol/L (0.5 mg/dL) occur in 10% of patients; however, at the HBV-effective dose of 10 mg, such elevations of creatinine are rarely encountered. If any nephrotoxicity does occur, it

rarely appears before 6–8 months of therapy. Although renal tubular injury is a rare potential side effect, and although creatinine monitoring is recommended during treatment, the therapeutic index of adefovir dipivoxil is high, and the nephrotoxicity observed in clinical trials at higher doses was reversible. For patients with underlying renal disease, frequency of administration of adefovir dipivoxil should be reduced to every 48 h for creatinine clearances of 20–49 mL/min; to every 72 h for creatinine clearances of 10–19 mL/min; and once a week, following dialysis, for patients undergoing hemodialysis. Adefovir dipivoxil is very well tolerated, and ALT elevations during and after withdrawal of therapy are similar to those observed and described above in clinical trials of lamivudine. An advantage of adefovir is its relatively favorable resistance profile; however, it is not as potent as the other approved oral agents, it does not suppress HBV DNA as rapidly or as uniformly as the others, it is the least likely of all agents to result in HBeAg seroconversion, and 20–50% of patients fail to suppress HBV DNA by 2 $\log_{10}$ ("primary nonresponders"). For these reasons, adefovir has been supplanted in both treatment-naïve and lamivudine-resistant patients by the more potent, less resistance-prone nucleotide analogue tenofovir (see below).

Pegylated Interferon

After long-acting PEG IFN was shown to be effective in the treatment of hepatitis C (see below), this more convenient drug was evaluated in the treatment of chronic hepatitis B. Once-a-week PEG IFN is more effective than the more frequently administered, standard IFN, and several large-scale trials of PEG IFN versus oral nucleoside analogues have been conducted among patients with HBeAgreactive and HBeAg-negative chronic hepatitis B.

In HBeAg-reactive chronic hepatitis B, two large-scale studies were done, one with PEG IFN-α 2b (100 μg weekly for 32 weeks, then 50 μg weekly for another 20 weeks for a total of 52 weeks, with a comparison arm of combination PEG IFN with oral lamivudine) in 307 subjects; the other involved PEG IFN-α 2a (180 μg weekly for 48 weeks) in 814 primarily Asian patients, three-fourths of whom had ALT ≥2 x the upper limit of normal, with comparison arms of lamivudine monotherapy and combination PEG IFN plus lamivudine. At the end of therapy (48–52 weeks) in the PEG IFN monotherapy arms, HBeAg loss occurred in approximately 30%, HBeAg seroconversion in 22–27%, undetectable HBV DNA (<400 copies/mL by PCR) in 10–25%, normal ALT in 34–39%, and a mean reduction in HBV DNA of 2 log 10 copies/mL (PEG IFN-α 2b) to 4.5 $\log_{10}$ copies/mL (PEG IFN-α 2a). Six months after completing PEG IFN monotherapy in these trials, HBeAg losses were present in approximately 35%, HBeAg seroconversion in approximately 30%, undetectable HBV DNA in 7–14%, normal ALT in 32–41%, and a mean reduction in HBV DNA of 2–2.4 $\log_{10}$ copies/mL. Although the combination of PEG IFN and lamivudine was superior at the end of therapy in one or more serologic, virologic, or biochemical outcomes, neither the combination arm (in both studies) nor the lamivudine monotherapy arm (in the PEG IFN-α 2a trial) demonstrated any benefit compared to the PEG IFN monotherapy arms 6 months after therapy. Moreover, HBsAg seroconversion occurred in 3–7% of PEG IFN recipients (with or without lamivudine); some of these seroconversions were identified by the end of therapy, but many were identified during the posttreatment follow-up period. The likelihood of HBeAg loss in PEG IFN–treated HBeAg-reactive patients is associated with HBV genotype A > B > C > D (shown for PEG IFN α-2b but not for α-2a).

Based on these results, some authorities concluded that PEG IFN monotherapy should be the first-line therapy of choice in HBeAg-reactive chronic hepatitis B; however, this conclusion has been challenged. Although a finite, 1-year course of PEG IFN results in a higher rate of sustained response (6 months after treatment) than is achieved with oral nucleoside/nucleotide analogue therapy, the comparison is confounded by the fact that oral agents are not discontinued at the end of 1 year. Instead, taken orally and free of side effects, therapy with oral agents is extended indefinitely or until after the occurrence of an HBeAg response. The rate of HBeAg responses after 2 years of oralagent nucleoside analogue therapy is at least as high as, if not higher than, that achieved with PEG IFN after 1 year; favoring oral agents is the absence of injections and difficult-to-tolerate side effects as well as lower direct and indirect medical costs and inconvenience. The association of HBsAg responses with PEG IFN therapy occurs in such a small proportion of patients that subjecting everyone to PEG IFN for the marginal gain of HBsAg responses during or immediately after therapy in such a very small minority is questionable. Moreover, HBsAg responses occur in a comparable proportion of patients treated with earlygeneration nucleoside/nucleotide analogues in the years after therapy, and, with the newer, more potent nucleoside analogues, the frequency of HBsAg loss during the first year of therapy equals that of PEG IFN and is exceeded during year 2 (see below). Of course, resistance is not an issue during PEG IFN therapy, but the risk of resistance is much lower with new agents (≤1% up to 3–5 years in previously treatment-naïve, entecavir-treated and tenofovir-treated patients; see below). Finally, the level of HBV DNA inhibition that can be achieved with the newer agents, and even with lamivudine, exceeds that which can be achieved with PEG IFN, in some cases by several orders of magnitude.

In HBeAg-negative chronic hepatitis B, a trial of PEG IFN-α 2a (180 μg weekly for 48 weeks versus comparison arms of lamivudine monotherapy and of combination therapy) in 564 patients showed that PEG IFN monotherapy resulted at the end of therapy in suppression of HBV DNA by a mean of 4.1 log 10 copies/mL, undetectable HBV DNA (<400 copies/mL by PCR) in 63%, normal ALT in 38%, and loss of HBsAg in 4%. Although lamivudine monotherapy and combination lamivudine–PEG IFN therapy were both superior to PEG IFN at the end of therapy, no advantage of lamivudine monotherapy or combination therapy was apparent over PEG IFN monotherapy 6 months after therapy— suppression of HBV DNA by a mean of 2.3 log 10 copies/mL, undetectable HBV DNA in 19%, and normal ALT in 59%. In subjects involved in this trial followed for up to 5 years, among the two-thirds followed who had been treated initially with PEG IFN, 17% maintained HBV DNA suppression to <400 copies/ml, but ALT remained normal in only 22%; HBsAg loss increased gradually to 12%. Among the half followed who had been treated initially with lamivudine monotherapy, HBV DNA remained <400 copies/ml in 7% and ALT normal in 16%; by year 5, 3.5% had lost HBsAg. As was the case for standard IFN therapy in HBeAgnegative patients, longer after PEG IFN treatment, although a small subset maintained their responses, the proportion who benefited was very small, raising questions about the relative value of a finite period of PEG IFN, versus a longer course with a potent, low-resistance oral nucleoside analogue in these patients.

Entecavir

Entecavir, an oral cyclopentyl guanosine analogue polymerase inhibitor, appears to be the most potent of the HBV antivirals and is just as well tolerated as lamivudine. In a 709-subject clinical trial among HBeAg-reactive patients, oral entecavir, 0.5 mg daily, was compared to lamivudine, 100 mg daily. At 48 weeks, entecavir was superior to lamivudine in suppression of HBV DNA, mean 6.9 versus 5.5 log 10 copies/mL and in percent with undetectable HBV DNA (<300 copies/mL by PCR), 67% versus 36%; histologic improvement (≥2-point improvement in necroinflammatory HAI score), 72% versus 62%; and normal ALT (68% versus 60%). The two treatments were indistinguishable in percent with HBeAg loss (22% versus 20%) and seroconversion (21% versus 18%). Among patients treated with entecavir for 96 weeks, HBV DNA was undetectable cumulatively in 80% (versus 39% for lamivudine), and HBeAg seroconversions had occurred in 31% (versus 26% for lamivudine); the HBeAg seroconversion rate after 3 years of entecavir in this cohort was 39%. Similarly, in a 638-subject clinical trial among HBeAg-negative patients, at week 48, oral entecavir, 0.5 mg daily, was superior to lamivudine, 100 mg daily, in suppression of HBV DNA, mean 5.0 versus 4.5 log 10 copies/mL and in percent with undetectable HBV DNA, 90% versus 72%; histologic improvement, 70% versus 61% and normal ALT, 78% versus 71%. No resistance mutations were encountered in previously treatment-naïve, entecavir-treated patients during 96 weeks of therapy, and in a cohort of subjects treated for up to 5 years, resistance emerged in 1.2%. Its high barrier to resistance coupled with its high potency renders entecavir a first-line drug for patients with chronic hepatitis B.

Entecavir is also effective against lamivudine-resistant HBV infection. In a trial of 286 lamivudine-resistant patients, entecavir, at a higher daily dose of 1 mg, was superior to lamivudine, as measured at week 48, in achieving suppression of HBV DNA (mean 5.1 versus 0.48 log 10 copies/mL); undetectable HBV DNA, in 72% versus 19%; normal ALT, in 61% versus 15%; HBeAg loss, in 10% versus 3%; and HBeAg seroconversion, in 8% versus 3%. In this population of lamivudine-experienced patients, however, entecavir resistance emerged in 7% at 48 weeks. Although entecavir resistance requires both a YMDD mutation and a second mutation at one of several other sites (e.g., T184A, S202G/I, or M250V), resistance to entecavir in lamivudineresistant chronic hepatitis B has been recorded to increase progressively to 43% at 4 years; therefore, entecavir is not as attractive a choice as adefovir or tenofovir for patients with lamivudine-resistant hepatitis B.

At the end of 2 years of entecavir therapy in clinical trials among HBeAg-reactive patients, HBsAg seroconversion was observed in 5% (≤2% during the first year). In addition, ontreatment and posttreatment ALT flares are relatively uncommon and relatively mild in entecavir-treated patients. In clinical trials, entecavir has had an excellent safety profile; doses should be reduced for patients with reduced creatinine clearance. Entecavir does have low-level antiviral activity against HIV and cannot be used as monotherapy to treat HBV infection in HIVHBV co-infected persons.

Telbivudine

Telbivudine, a cytosine analogue, appears to be similar in efficacy to entecavir; however, it is slightly less potent in suppressing HBV DNA (a slightly more profound median 6.4 $\log_{10}$ reduction in HBeAg-reactive disease, a similar 5.2 $\log_{10}$ reduction in HBeAg-negative disease). In its registration trial, telbivudine at an oral daily dose of 600 mg suppressed HBV DNA to <300 copies/ml in 60% of HBeAg-positive and 88% of HBeAg-negative patients, reduced ALT to normal in 77% of HBeAg-positive and 74% of HBeAg-negative patients, and improved histology in 65% of HBeAg-positive and 67% HBeAgnegative patients. Although resistance to telbivudine (M204I, not M204V mutations) was less frequent than resistance to lamivudine at the end of 1 year, resistance mutations after 2 years of treatment occurred in up to 22%. Generally well tolerated,

telbivudine has been associated with a low frequency of asymptomatic creatine kinase elevations and with a very low frequency of peripheral neuropathy; frequency of administration should be reduced for patients with impaired creatinine clearance. Its excellent potency notwithstanding, the inferior resistance profile of telbivudine has limited its appeal; telbivudine is neither recommended as first-line therapy nor widely used.

Tenofovir

Tenofovir disoproxil fumarate, an acyclic nucleotide analogue and potent antiretroviral agent used to treat HIV infection, is similar to adefovir but more potent in suppressing HBV DNA and inducing HBeAg responses; it is highly active against both wild-type and lamivudine-resistant HBV and active in patients whose response to adefovir is slow and/or limited. At an oral once-daily dose of 300 mg for 48 weeks, tenofovir suppressed HBV DNA by 6.2 $\log_{10}$ [to undetectable levels (<400 copies/ml) in 76%] in HBeAg-positive and 4.6 $\log_{10}$ (to undetectable levels in 93%) in HBeAg-negative patients; reduced ALT to normal in 68% of HBeAg-positive and 76% of HBeAg-negative patients; and improved histology in 74% of HBeAg-positive and 72% of HBeAg-negative patients. In HBeAg-positive patients, HBeAg seroconversions occurred in 21% by the end of year 1 and in 27% by the end of year 2 of tenofovir treatment; HBsAg loss occurred in 3% by the end of year 1 and 6% by the end of year 2. The safety (negligible renal toxicity and mild reduction in bone density) and resistance profile (none recorded through 3 years) of tenofovir are very favorable as well; therefore, tenofovir has supplanted adefovir both as first-line therapy for chronic hepatitis B and as add-on therapy for lamivudineresistant chronic hepatitis B. Frequency of tenofovir administration should be reduced for patients with impaired creatinine clearance.

A comparison of the six antiviral therapies in current use appears in Table 306-3; their relative potencies in suppressing HBV DNA are shown in Fig. 306-1.

Table 306-3 Comparison of Pegylated Interferon (Peg IFN), Lamivudine, Adefovir, Entecavir, Telbivudine, and Tenofovir Therapy for Chronic Hepatitis B[a]

Feature	PEG IFN[b]	Lamivudine	Adefovir	Entecavir	Telbivudine	Tenofovir
Route of administration	Subcutaneous Injection	Oral	Oral	Oral	Oral	Oral
Duration of therapy[c]	48–52 weeks	≥52 weeks	≥48 weeks	≥48 weeks	≥52 weeks	≥48 weeks
Tolerability	Poorly tolerated	Well tolerated	Well tolerated; creatinine monitoring recommended	Well tolerated	Well tolerated	Well tolerated creatinine monitoring recommended
HBeAg seroconversion						
1 yr Rx	18–20%	16–21%	12%	21%	22%	21%
>1 yr Rx	NA	up to 50% @ 5 yrs	43% @ 3 yrs[d]	31% @ 2 yrs 39% @ 3 yrs	30% @ 2 yrs	27% @ 2 yrs
$\log_{10}$ HBV DNA reduction (mean copies/ml)						
HBeAg-reactive	4.5	5.5	median 3.5–5	6.9	6.4	6.2
HBeAg-negative	4.1	4.4–4.7	median 3.5–3.9	5.0	5.2	4.6
HBV DNA PCR negative (<300–400 copies/ml; <1,000 copies/ml for adefovir) end of yr 1						
HBeAg-reactive	10–25%	36–44%	13–21%	67% (91% @ 4 yrs)	60%	76%
HBeAg-negative	63%	60–73%	48–77%	90%	88%	93%
ALT normalization at end of yr 1						
HBeAg-reactive	39%	41–75%	48–61%	68%	77%	68%
HBeAg-negative	34–38%	62–79%	48–77%	78%	74%	76%
HBsAg loss yr 1	3–4%	≥1%	0%	2%	<1%	3%
yr 2	12% 5 yr after 1 yr of Rx	no data	5% at yr 5	5%	no data	6%
Histologic improvement (≥2 point reduction in HAI) at yr 1						
HBeAg-reactive	38% 6 months after	49–62%	53–68%	72%	65%	74%
HBeAg-negative	48% 6 months after	61–66%	64%	70%	67%	72%
Viral resistance	None	15-30% @ 1 yr 70% @ 5 yrs	None @ 1 yr 29% at 5 yrs	≥1% @ 1 yr[e] 1.2% @ 5 yr[e]	up to 5% @ yr 1 up to 22% @ yr 2	0% @ yr 1 0% through yr 3
Cost (US$) for 1 yr	~$18,000	~$2,500	~$6,500	~$8,700f	~$6,000	~$6,000

[a] Generally, these comparisons are based on data on each drug tested individually versus placebo in registration clinical trials; because, with rare exception, these comparisons are not based on head-to-head testing of these drugs, relative advantages and disadvantages should be interpreted cautiously.

[b] Although standard interferon α administered daily or three times a week is approved as therapy for chronic hepatitis B, it has been supplanted by PEG IFN, which is administered once a week and is more effective. Standard interferon has no advantages over PEG IFN.

[c] Duration of therapy in clinical efficacy trials; use in clinical practice may vary.

[d] Because of a computer-generated randomization error that resulted in misallocation of drug versus placebo during the second year of clinical-trial treatment, the frequency of HBeAg seroconversion beyond the first year is an estimate (Kaplan-Meier analysis) based on the small subset in whom adefovir was administered correctly.

[e] 7% during a year of therapy (43% at year 4) in lamivudine-resistant patients.

[f] ~17,400 for lamivudine-refractory patients.

Abbreviations: ALT, alanine aminotransferase; HAI, histologic activity index; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; NA, not applicable; PEG IFN, pegylated interferon; PCR, polymerase chain reaction; Rx, therapy; yr, year.

Figure 306-1

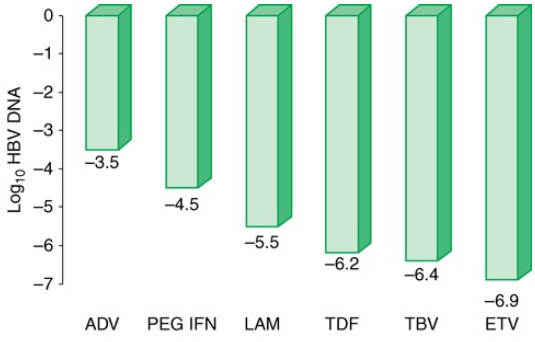

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relative potency of antiviral drugs for hepatitis B, as reflected by median log 10 HBV DNA reduction in HBeAg-positive chronic hepatitis B. These data are from individual reports of large, randomized controlled registration trials that were the basis for approval of the drugs. In most instances, these data do not represent direct comparisons among the drugs, because study populations were different, baseline patient variables were not always uniform, and the sensitivity and dynamic range of the HBV DNA assays used in the trials varied. ADV, adefovir dipivoxil; PEG IFN, pegylated interferon α-2a; LAM, lamivudine; TDF, tenofovir disoproxil fumarate; TBV, telbivudine; ETV, entecavir.

Combination Therapy

Although the combination of lamivudine and PEG IFN suppresses HBV DNA more profoundly during therapy than does monotherapy with either drug alone (and is much less likely to be associated with lamivudine resistance), this combination used for a year is no better than a year of PEG IFN in achieving sustained responses. To date, combinations of oral nucleoside/nucleotide agents have not achieved an enhancement in virologic, serologic, or biochemical efficacy over that achieved by the more potent of the combined drugs given individually. On the other hand, combining agents that are not cross-resistant (e.g., lamivudine and adefovir or tenofovir) has the potential to reduce the risk or perhaps even to preempt entirely the emergence of drug resistance. In the future, the treatment paradigm may shift from the current approach of sequential monotherapy to preemptive combination therapy; however, designing and executing clinical trials that demonstrate superior efficacy and resistance profile of combination therapy over monotherapy with entecavir or tenofovir will be very challenging.

Novel Antivirals and Strategies

In addition to the seven approved antiviral drugs for hepatitis B, emtricitabine, a fluorinated cytosine analogue very similar to lamivudine in structure, efficacy, and resistance profile, offers no advantage over lamivudine. A combination of emtricitabine and tenofovir is approved for the treatment of HIV infection and is an appealing combination therapy for hepatitis B; however, neither emtricitabine nor the combination are approved yet for hepatitis B. Several initially promising antiviral agents have been abandoned because of toxicity (e.g., clevudine, which was linked to myopathy during its clinical development). Because direct-acting antivirals have been so successful in the management of chronic hepatitis B, more unconventional approaches—e.g., immunologic or genetic manipulation—are not likely to be competitive. Finally, initial emphasis in the development of antiviral therapy for hepatitis B was placed on monotherapy; whether combination regimens will yield additive or synergistic efficacy remains to be determined).

Treatment Recommendations

Several learned societies and groups of expert physicians have issued treatment recommendations for patients with chronic hepatitis B; the most authoritative and updated (and free of financial support by pharmaceutical companies) are those of the American Association for the Study of Liver Diseases (AASLD) and of the European Association for the Study of the Liver (EASL). Although the recommendations differ slightly, a consensus has emerged on most of the important points **(Table 306-4)**. No treatment is recommended or available for inactive "nonreplicative" hepatitis B carriers (undetectable HBeAg with normal ALT and HBV DNA ≤10^3 IU/ml documented serially over time). In patients with detectable HBeAg and HBV DNA levels >2 x10^4 IU/ml, treatment is recommended by the AASLD for those with ALT levels above 2 x the upper limit of normal. (The EASL recommends treatment in HBeAg-positive patients for HBV DNA levels >2 x 10^3 IU/ml and ALT above the upper limit of normal.) For HBeAgpositive patients with ALT ≤2 x the upper limit of normal, in whom sustained responses are not likely and who would require multiyear therapy, antiviral therapy is not recommended currently. This pattern is common during the early decades of life among Asian patients infected at birth; even in this group, therapy would be considered for those >40 years of age, ALT persistently at the high end of the twofold range, and/or with a family history of hepatocellular carcinoma, especially if the liver biopsy shows moderate to severe necroinflammatory activity or fibrosis. In this group, when, eventually, ALT becomes elevated later in life, antiviral therapy should be instituted. For patients with HBeAg-negative chronic hepatitis B, ALT >2 x the upper limit of normal (above the upper limit of normal according to EASL) and HBV DNA >2 x 10^3 IU/ml, antiviral therapy is recommended. If HBV DNA is >2 x 10^3 IU/ml and ALT is 1 to >2 x the upper limit of normal, liver biopsy should be considered to help in arriving at a decision to treat if substantial liver injury is present (treatment in this subset would be recommended according to EASL guidelines, because ALT is elevated).

Table 306-4 Recommendations for Treatment of Chronic Hepatitis B[a]

HBeAg status	Clinical	HBV DNA (IU/ml)	ALT	Recommendation
HBeAg-reactive	*b*	>2 x10^4	<2 x ULN[c]	No treatment; monitor. In patients >40, with family history of hepatocellular carcinoma, and/or ALT persistently at the high end of the twofold range, liver biopsy may help in decision to treat
	Chronic hepatitis	>2 x 10^{4d}	>2 x ULN[d]	Treat[e]
	Cirrhosis compensated	>2 x 10^3	< or > ULN	Treat[e] with oral agents, not PEG IFN
		<2 x 10^3	>ULN	Consider treatment[f]
	Cirrhosis decompensated	Detectable	< or > ULN	Treat[e] with oral agents[g], not PEG IFN; refer for liver transplantation
		Undetectable	< or > ULN	Observe; refer for liver transplantation
HBeAg-negative	*b*	≤2 x 10^3	≤ULN	Inactive carrier; treatment not necessary
	Chronic hepatitis	>10^3	1->2 x ULN[d]	Consider liver biopsy; treat[h] if biopsy shows moderate to severe inflammation or fibrosis
	Chronic hepatitis	>10^4	>2 x ULN[d]	Treat[h,i]
	Cirrhosis	>2 x 10^3	< or >	Treat[e] with oral agents, not PEG IFN

compensated	>2 x 10	ULN	Treat with oral agents, not PEG IFN
	<2 x 10^3	>ULN	Consider treatment[f]
Cirrhosis decompensated	Detectable	< or > ULN	Treat[h] with oral agents[g], not PEG IFN; refer for liver transplantation
	Undetectable	< or > ULN	Observe; refer for liver transplantation

[a] Based on practice guidelines of the American Association for the Study of Liver Diseases (AASLD). Except as indicated in footnotes, these guidelines are similar to those issued by the European Association for the Study of the Liver (EASL).

[b] Liver disease tends to be mild or inactive clinically; most such patients do not undergo liver biopsy.

[c] This pattern is common during early decades of life in Asian patients infected at birth.

[d] According to the EASL guidelines, treat if HBV DNA is >2 x 10^3 IU/ml and ALT >ULN.

[e] One of the potent oral drugs with a high barrier to resistance (entecavir or tenofovir) or PEG IFN can be used as first-line therapy (see text). These oral agents, but not PEG IFN, should be used for interferon-refractory/intolerant and immunocompromised patients. PEG IFN is administered weekly by subcutaneous injection for a year; the oral agents are administered daily for at least a year and continued indefinitely or until at least 6 months after HBeAg seroconversion.

[f] According to EASL guidelines, patients with compensated cirrhosis and detectable HBV DNA at any level, even with normal ALT, are candidates for therapy. Most authorities would treat indefinitely, even in HBeAg-positive disease after HBeAg seroconversion.

[g] Because the emergence of resistance can lead to loss of antiviral benefit and further deterioration in decompensated cirrhosis, a low-resistance regimen is recommended-entecavir or tenofovir monotherapy or combination therapy with the more resistance-prone lamivudine (or telbivudine) plus adefovir. Therapy should be instituted urgently.

[h] Because HBeAg seroconversion is not an option, the goal of therapy is to suppress HBV DNA and maintain a normal ALT. PEG IFN is administered by subcutaneous injection weekly for a year; caution is warranted in relying on a 6-month posttreatment interval to define a sustained response, because the majority of such responses are lost thereafter. Oral agents, entecavir or tenofovir, are administered daily, usually indefinitely or, until as very rarely occurs, virologic and biochemical responses are accompanied by HBsAg seroconversion.

[i] For older patients and those with advanced fibrosis, consider lowering the HBV DNA threshold to >2 x 10^3 IU/ml.

Abbreviations: ALT, alanine aminotransferase; AASLD, American Association for the Study of Liver Diseases; EASL, European Association for the Study of the Liver; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; PEG IFN, pegylated interferon; ULN, upper limits of normal.

For patients with compensated cirrhosis, because antiviral therapy has been shown to retard clinical progression, treatment is recommended regardless of HBeAg status and ALT as long as HBV DNA is detectable at >2 x 10^3 IU/ml (detectable at any level according to the EASL); monitoring without therapy is recommended for those with HBV DNA <2 x 10^3 IU/ml, unless ALT is elevated. For patients with decompensated cirrhosis, treatment is recommended regardless of serologic and biochemical status, as long as HBV DNA is detectable. Patients with decompensated cirrhosis should be evaluated as candidates for liver transplantation.

Among the seven available drugs for hepatitis B, PEG IFN has supplanted standard IFN, entecavir has supplanted lamivudine, and tenofovir has supplanted adefovir. PEG IFN, entecavir, or tenofovir are recommended as first-line therapy (Table 306-3). PEG IFN requires finite-duration therapy, achieves the highest rate of HBeAg responses after a year of therapy, and does not support viral mutations, but it requires subcutaneous injections and is associated with inconvenience and intolerability. Oral nucleoside analogues require long-term therapy in most patients, and when used alone, lamivudine and telbivudine foster the emergence of viral mutations, adefovir somewhat less so, and entecavir (except in lamivudine-experienced patients) and tenofovir rarely at all. Oral agents do not require injections, are very well tolerated, lead to improved histology in 50–90% of patients, suppress HBV DNA more profoundly than PEG IFN, and are effective even in patients who fail to respond to IFN-based therapy. Although oral agents are less likely to result in HBeAg responses during the first year of therapy, as compared to PEG IFN, treatment with oral agents tends to be extended beyond the first year and, by the end of the second year, yields HBeAg responses (and even HBsAg responses) comparable in frequency to those achieved after 1 year of PEG IFN (and without the associated side effects) **(Table 306-5)**. Although adefovir and tenofovir are safe, creatinine monitoring is recommended. Substantial experience with lamivudine during pregnancy (see above) has identified no teratogenicity. Although interferons do not appear to cause congenital anomalies, interferons have antiproliferative properties and should not be used during pregnancy. Adefovir during pregnancy has not been associated with birth defects; however, there may be an increased risk of spontaneous abortion. Data on the safety of entecavir during pregnancy have not been published. Sufficient data in animals and limited data in humans suggest that telbivudine and tenofovir can be used safely during pregnancy. In general, except perhaps for lamivudine, and until additional data become available, the other antivirals for hepatitis B should be avoided or used with extreme caution during pregnancy.

Table 306-5 Pegylated Interferon versus Oral Nucleoside Analogues for the Treatment of Chronic Hepatitis B

	PEG IFN	Nucleoside Analogues
Administration	Weekly injection	Daily, orally
Tolerability	Poorly tolerated, intensive monitoring	Well tolerated, limited monitoring
Duration of therapy	Finite 48 weeks	≥1 year, indefinite in most patients
Maximum mean HBV DNA suppression	4.5 $\log_{10}$	6.9 $\log_{10}$
Effective in high-level HBV DNA (≥10^9 IU/ml)	No	Yes
HBeAg seroconversion		
During 1 year of therapy	~30%	~20%
During >1 year of therapy	Not applicable	30% (year 2)–50% (year 5)
HBeAg-negative posttreatment HBV DNA suppression	17% @ 5 years	7% @ 4 years (lamivudine)
HBsAg loss		
During 1 year of therapy	3–4%	0–3%
During >1 year of therapy	Not applicable	3–6% @ 2 years of therapy
After 1 year of therapy–HBeAg-negative	12% @ 5 years	3.5% @ 5 years
Antiviral resistance	None	Lamivudine: ~30% year 1, ~70% year 5 Adefovir: 0% year 1, ~30% year 5 Telbivudine: up to 4% year 1, 22% year 2 Entecavir: ≤1.2% through year 5

		Tenofovir: 0% through year 3
Use in cirrhosis, transplantation, immunosuppressed	No	Yes
Cost, 1 year of therapy	++++	+ to ++

Abbreviations: HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; IU/ml, international units per milliliter; PEG IFN, pegylated interferon.

As noted above, some physicians prefer to begin with PEG IFN, while other physicians and patients prefer oral agents as first-line therapy. For patients with decompensated cirrhosis, the emergence of resistance can result in further deterioration and loss of antiviral effectiveness. Therefore, in this patient subset, the threshold for relying on therapy with a very favorable resistance profile (e.g., entecavir or tenofovir) or on combination therapy (e.g., lamivudine or telbivudine with adefovir) is low. PEG IFN should not be used in patients with compensated or decompensated cirrhosis.

For patients with end-stage chronic hepatitis B who undergo liver transplantation, reinfection of the new liver is almost universal in the absence of antiviral therapy. The majority of patients become high-level viremic carriers with minimal liver injury. Before the availability of antiviral therapy, an unpredictable proportion experienced severe hepatitis B–related liver injury, sometimes a fulminant-like hepatitis, sometimes a rapid recapitulation of the original severe chronic hepatitis B **(Chap. 304)**. Currently, however, prevention of recurrent hepatitis B after liver transplantation has been achieved definitively by *combining* hepatitis B immune globulin with one of the oral nucleoside or nucleotide analogues **(Chap. 310)**.

For patients treated with the more resistance-prone (lamivudine, telbivudine) or less potent (adefovir) oral agents, assessment of response at 24 weeks (48 weeks for adefovir) can identify those at high risk for inadequate response and breakthrough resistance (i.e., the presence of residual detectable viremia). When such inadequate responses are identified, a second, noncross-resistant agent can be added or the initial agent can be replaced by a more potent agent. This "roadmap" approach has been rendered irrelevant by the use of the current generation of highly potent, low-resistant agents entecavir and tenofovir. Still, at 24 weeks, if HBV DNA exceeds 2×10^3 IU/ml, switching to a different agent or adding a second agent is advisable.

Patients with HBV-HIV co-infection can have progressive HBV-associated liver disease and, occasionally, a severe exacerbation of hepatitis B resulting from immunologic reconstitution following highly active antiretroviral therapy. Lamivudine should never be used as monotherapy in patients with HBVHIV infection because HIV resistance emerges rapidly to both viruses. Adefovir has been used successfully to treat chronic hepatitis B in HBV-HIV co-infected patients but is no longer considered a first-line agent for HBV. Entecavir has low-level activity against HIV and can result in selection of HIV resistance; therefore, it should be avoided in HBV-HIV co-infection. Tenofovir and the combination of tenofovir and emtricitabine in one pill are approved therapies for HIV and represent excellent choices for treating HBV infection in HBV-HIV co-infected patients. Generally, even for HBV-HIV co-infected patients who do not yet meet treatment criteria for HIV infection, treating for both HBV and HIV is recommended.

Patients with chronic hepatitis B who undergo cytotoxic chemotherapy for treatment of malignancies as well as patients treated with immunosuppressive, anticytokine, or antitumor necrosis factor therapies experience enhanced HBV replication and viral expression on hepatocyte membranes during chemotherapy coupled with suppression of cellular immunity. When chemotherapy is withdrawn, such patients are at risk for reactivation of hepatitis B, often severe and occasionally fatal. Such rebound reactivation represents restoration of cytolytic T cell function against a target organ enriched in HBV expression. Preemptive treatment with lamivudine prior to the initiation of chemotherapy has been shown to reduce the risk of such reactivation. In all likelihood, the newer, more potent oral antiviral agents will work as well and with a lower risk of antiviral drug resistance. The optimal duration of antiviral therapy after completion of chemotherapy is not known, but a suggested approach is 6 months for inactive hepatitis B carriers and longer-duration therapy in patients with baseline HBV DNA levels $>2 \times 10^3$ IU/ml, until standard clinical endpoints are met (Table 306-4).

Chronic Hepatitis D (Delta Hepatitis)

Chronic hepatitis D (HDV) may follow acute co-infection with HBV but at a rate no higher than the rate of chronicity of acute hepatitis B. That is, although HDV co-infection can increase the severity of acute hepatitis B, HDV does not increase the likelihood of progression to chronic hepatitis B. When, however, HDV super-infection occurs in a person who is already chronically infected with HBV, long-term HDV infection is the rule and a worsening of the liver disease the expected consequence. Except for severity, chronic hepatitis B plus D has similar clinical and laboratory features to those seen in chronic hepatitis B alone. Relatively severe and progressive chronic hepatitis, with or without cirrhosis, is the rule, and mild chronic hepatitis is the exception. Occasionally, mild hepatitis or even, rarely, inactive carriage occurs in patients with chronic hepatitis B plus D, and the disease may become indolent after several years of infection. A distinguishing serologic feature of chronic hepatitis D is the presence in the circulation of antibodies to liver-kidney microsomes (anti-LKM); however, the anti-LKM seen in hepatitis D, anti-LKM3, are directed against uridine diphosphate glucuronosyltransferase and are distinct from anti-LKM1 seen in patients with autoimmune hepatitis and in a subset of patients with chronic hepatitis C (see below). **The clinical and laboratory features of chronic HDV infection are summarized in Chaps. 304.**

TREATMENT: CHRONIC HEPATITIS D

Management is not well defined. Glucocorticoids are ineffective and are not used. Preliminary experimental trials of IFN-α suggested that conventional doses and durations of therapy lower levels of HDV RNA and aminotransferase activity only transiently during treatment but have no impact on the natural history of the disease. In contrast, high-dose IFN-α (9 million units three times a week) for 12 months may be associated with a sustained loss of HDV replication and clinical improvement in up to 50% of patients. Moreover, the beneficial impact of treatment has been observed to persist for 15 years and to be associated with a reduction in grade of hepatic necrosis and inflammation, reversion of advanced fibrosis (improved stage), and clearance of HDV RNA in some patients. A suggested approach to therapy has been high-dose, long-term IFN for at least a year and, in responders, extension of therapy until HDV RNA and HBsAg clearance. PEG IFN has also been shown to be effective in the treatment of chronic hepatitis D and is likely to become a more convenient replacement for standard IFN. None of the nucleoside analogue antiviral agents for hepatitis B is effective in hepatitis D. In patients with end-stage liver disease secondary to chronic hepatitis D, liver transplantation has been effective. If hepatitis D recurs in the new liver without the expression of hepatitis B (an unusual serologic profile in immunocompetent persons but common in transplant patients), liver injury is limited. In fact, the outcome of transplantation for chronic hepatitis D is superior to that for chronic hepatitis B; in such patients, combination hepatitis B immune globulin and nucleoside analogue therapy for hepatitis B is indicated **(Chap. 310)**.

Chronic Hepatitis C

Regardless of the epidemiologic mode of acquisition of hepatitis C virus (HCV) infection, chronic hepatitis follows acute hepatitis C in 50–70% of cases; chronic infection is common even in those with a return to normal in aminotransferase levels after acute hepatitis C, adding up to an 85% likelihood of chronic HCV infection after acute hepatitis C. Few clues had emerged to explain host differences associated with chronic infection until recently, when variation in a single nucleotide polymorphism (SNP) on chromosome 19, IL28B (which codes for interferon-λ3), was identified that distinguished between responders and nonresponders to antiviral therapy (see below). The same variants correlated with spontaneous resolution after acute infection: 53% in genotype C/C, 30% in genotype C/T, but only 23% in genotype T/T.

In patients with chronic hepatitis C followed for 20 years, progression to cirrhosis occurs in about 20–25%. Such is the case even for patients with relatively clinically mild chronic hepatitis, including those without symptoms, with only modest elevations of aminotransferase activity and with mild chronic hepatitis on liver biopsy. Even in cohorts of well-compensated patients with chronic hepatitis C referred for clinical research trials (no complications of chronic liver disease and with normal hepatic synthetic function), the prevalence of cirrhosis may be as high as 50%. Most cases of hepatitis C are identified initially in asymptomatic patients who have no history of acute hepatitis C (e.g., those discovered while attempting to donate blood, while undergoing lab testing as part of an application for life insurance, or as a result of routine laboratory tests). The source of HCV infection in many of these cases is not defined, although a long-forgotten percutaneous exposure in the remote past can be elicited in a substantial proportion and probably accounts for most infections; most of these infections were acquired in the 1960s and 1970s, coming to clinical attention decades later.

Approximately one-third of patients with chronic hepatitis C have normal or near-normal aminotransferase activity; although one-third to one-half of these patients have chronic hepatitis on liver biopsy, the grade of liver injury and stage of fibrosis tend to be mild in the vast majority. In some cases, more severe liver injury has been reported—even, rarely, cirrhosis, most likely the result of previous histologic activity. Among patients with persistent normal aminotransferase activity sustained over $\geq$5–10 years, histologic progression has been shown to be rare; however, approximately one-fourth of patients with normal aminotransferase activity experience subsequent aminotransferase elevations, and histologic injury can be progressive once abnormal biochemical activity resumes. Therefore, continued clinical monitoring is indicated, even for patients with

normal aminotransferase activity.

Despite this substantial rate of progression of chronic hepatitis C, and despite the fact that liver failure can result from end-stage chronic hepatitis C, the long-term prognosis for chronic hepatitis C in a majority of patients is relatively benign. Mortality over 10–20 years among patients with transfusion-associated chronic hepatitis C has been shown not to differ from mortality in a matched population of transfused patients in whom hepatitis C did not develop. Although death in the hepatitis group is more likely to result from liver failure, and although hepatic decompensation may occur in ~15% of such patients over the course of a decade, the majority (almost 60%) of patients remain asymptomatic and well compensated, with no clinical sequelae of chronic liver disease. Overall, chronic hepatitis C tends to be very slowly and insidiously progressive, if at all, in the vast majority of patients, while in approximately one-fourth of cases, chronic hepatitis C will progress eventually to end-stage cirrhosis. In fact, because HCV infection is so prevalent, and because a proportion of patients progress inexorably to endstage liver disease, hepatitis C is the most frequent indication for liver transplantation **(Chap. 310)**. Referral bias may account for the more severe outcomes described in cohorts of patients reported from tertiary care centers (20-year progression of ≥20%) versus the more benign outcomes in cohorts of patients monitored from initial blood-product-associated acute hepatitis or identified in community settings (20-year progression of only 4–7%). Still unexplained, however, are the wide ranges in reported progression to cirrhosis, from 2% over 17 years in a population of women with hepatitis C infection acquired from contaminated anti-D immune globulin to 30% over ≤11 years in recipients of contaminated intravenous immune globulin.

Progression of liver disease in patients with chronic hepatitis C has been reported to be more likely in patients with older age, longer duration of infection, advanced histologic stage and grade, genotype 1, more complex quasispecies diversity, increased hepatic iron, concomitant other liver disorders (alcoholic liver disease, chronic hepatitis B, hemochromatosis, α_1 -antitrypsin deficiency, and steatohepatitis), HIV infection, and obesity. Among these variables, however, duration of infection appears to be one of the most important, and some of the others probably reflect disease duration to some extent (e.g., quasispecies diversity, hepatic iron accumulation). No other epidemiologic or clinical features of chronic hepatitis C (e.g., severity of acute hepatitis, level of aminotransferase activity, level of HCV RNA, presence or absence of jaundice during acute hepatitis) are predictive of eventual outcome. Despite the relatively benign nature of chronic hepatitis C over time in many patients, cirrhosis following chronic hepatitis C has been associated with the late development, after several decades, of HCC **(Chap. 88)**; the annual rate of HCC in cirrhotic patients with hepatitis C is 1–4%, occurring primarily in patients who have had HCV infection for 30 years or more.

Perhaps the best prognostic indicator in chronic hepatitis C is liver histology; the rate of hepatic fibrosis may be slow, moderate, or rapid. Patients with mild necrosis and inflammation as well as those with limited fibrosis have an excellent prognosis and limited progression to cirrhosis. In contrast, among patients with moderate to severe necroinflammatory activity or fibrosis, including septal or bridging fibrosis, progression to cirrhosis is highly likely over the course of 10–20 years. Among patients with compensated cirrhosis associated with hepatitis C, the 10-year survival rate is close to 80%; mortality occurs at a rate of 2–6% per year; decompensation at a rate of 4–5% per year; and, as noted above, HCC at a rate of 1–4% per year. **A discussion of the pathogenesis of liver injury in patients with chronic hepatitis C appears in Chaps. 304.**

Clinical features of chronic hepatitis C are similar to those described above for chronic hepatitis B. Generally, fatigue is the most common symptom; jaundice is rare. Immune complex–mediated extrahepatic complications of chronic hepatitis C are less common than in chronic hepatitis B (despite the fact that assays for immune complexes are often positive in patients with chronic hepatitis C), with the exception of essential mixed cryoglobulinemia **(Chap. 304)**, which is linked to cutaneousvasculitis and membranoproliferative glomerulonephritis as well as lymphoproliferative disorders such as B-cell lymphoma and unexplained monoclonal gammopathy. In addition, chronic hepatitis C has been associated with extrahepatic complications unrelated to immune-complex injury. These include Sjögren's syndrome, lichen planus, porphyria cutanea tarda, type-II diabetes mellitus and the metabolic syndrome (including insulin resistance and steatohepatitis).

Laboratory features of chronic hepatitis C are similar to those in patients with chronic hepatitis B, but aminotransferase levels tend to fluctuate more (the characteristic episodic pattern of aminotransferase activity) and to be lower, especially in patients with long-standing disease. An interesting and occasionally confusing finding in patients with chronic hepatitis C is the presence of autoantibodies. Rarely, patients with autoimmune hepatitis (see below) and hyperglobulinemia have false-positive immunoassays for anti-HCV. On the other hand, some patients with serologically confirmable chronic hepatitis C have circulating anti-LKM. These antibodies are anti-LKM1, as seen in patients with autoimmune hepatitis type 2 (see below), and are directed against a 33-aminoacid sequence of cytochrome P450 IID6. The occurrence of anti-LKM1 in some patients with chronic hepatitis C may result from the partial sequence homology between the epitope recognized by anti-LKM1 and two segments of the HCV polyprotein. In addition, the presence of this autoantibody in some patients with chronic hepatitis C suggests that autoimmunity may be playing a role in the pathogenesis of chronic hepatitis C.

Histopathologic features of chronic hepatitis C, especially those that distinguish hepatitis C from hepatitis B, are described in Chaps. 304.

TREATMENT: CHRONIC HEPATITIS C

Therapy for chronic hepatitis C has evolved substantially in the two decades since IFN-α was introduced for this indication. When first approved, IFN-α was administered via subcutaneous injection three times a week for 6 months but achieved a sustained virologic response, SVR **(Fig. 306-2)** (a reduction of HCV RNA to undetectable levels by PCR when measured ≥6 months after completion of therapy) below 10%. Doubling the duration of therapy—but not increasing the dose or changing IFN preparations—increased the SVR rate to ~20%, and addition to the regimen of daily ribavirin, an oral guanosine nucleoside, increased the SVR rate to 40%. When used alone, ribavirin is ineffective and does not reduce HCV RNA levels, but ribavirin enhances the efficacy of IFN by reducing the likelihood of virologic relapse after the achievement of an end-treatment response (ETR) (Fig. 306-2) (response measured during, and maintained to the end of, treatment). Proposed mechanisms to explain the role of ribavirin include subtle direct reduction of HCV replication, inhibition of host inosine monophosphate dehydrogenase activity (and associated depletion of guanosine pools), immune modulation, induction of virologic mutational catastrophe, and enhancement of interferon-stimulated gene expression. Interferon therapy results in activation of the JAKSTAT signal transduction pathway, which culminates in the intracellular elaboration of genes and their protein products that have antiviral properties. Hepatitis C proteins inhibit JAK-STAT signaling at several steps along the pathway, and exogenous interferon restores expression of interferon-stimulated genes and their antiviral effects.

Figure 306-2

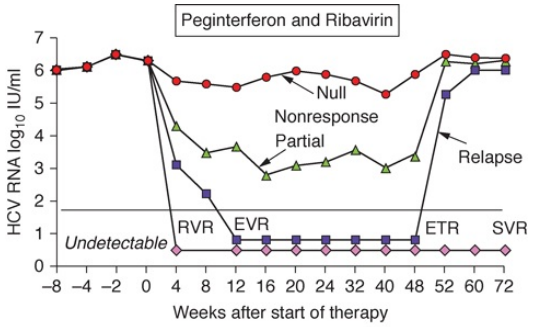

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Virologic responses during a 48-week course of antiviral therapy in patients with hepatitis C, genotype 1 or 4 (for genotypes 2 or 3, the course would be 24 weeks). Nonresponders can be classified as null responders (HCV RNA reduction of <2 $\log_{10}$ IU/ml) or partial responders (HCV RNA reduction ≥2 $\log_{10}$ IU/ml but not suppressed to undetectable) by week 24 of therapy. In responders, HCV RNA can become undetectable, as shown with sensitive amplification assays, within 4 weeks (RVR, rapid virologic response); can be reduced by ≥2 $\log_{10}$ IU/ml within 12 weeks (early virologic response, EVR; if HCV RNA is undetectable at 12 weeks, the designation is "complete" EVR); or at the end of therapy, 48 weeks (ETR, end-treatment response). In responders, if HCV RNA remains undetectable for 24 weeks after ETR, week 72, the patient has a sustained virologic response (SVR), but if HCV RNA becomes detectable again, the patient is considered to have relapsed. *(Reproduced with permission, courtesy of Marc G. Ghany, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health and the American Association for the Study of Liver Diseases Hepatology 49:1335, 2009.)*

The current standard of care is the combination of longacting pegylated IFN (PEG IFN) and ribavirin, which has increased responsiveness (frequency of SVR) to as high as 55% overall, >40% in genotypes 1 and 4 and to >80% in genotypes 2 and 3. Still, many important lessons about antiviral therapy for chronic hepatitis C were learned from the experience with IFN monotherapy and combination IFN-ribavirin therapy. Even in the absence of biochemical and virologic responses, histologic improvement occurs in approximately three-fourths of all treated patients. In chronic hepatitis C, unlike the case in hepatitis B, responses to therapy are not accompanied by transient, acute hepatitis-like aminotransferase elevations. Instead, ALT levels fall precipitously during therapy. Up to 90% of virologic responses are achieved within the first 12 weeks of therapy; responses thereafter are rare. Most relapses occur within the first 12 weeks after treatment. Sustained virologic responses are very durable; normal ALT, improved histology, and absence of HCV RNA in serum and liver have been documented a decade after successful therapy, and "relapses" 2 years after sustained responses are almost unheard of. Thus, an SVR to antiviral therapy of chronic hepatitis C is tantamount to a cure.

Patient variables that tend to correlate with sustained virologic responsiveness to IFN-based therapy include favorable genotype (genotypes 2 and 3 as opposed to genotypes 1 and 4), low baseline HCV RNA level (<2 million copies/mL, which is equivalent to ~800,000 IU/ml, the current convention of quantitation), histologically mild hepatitis and minimal fibrosis, age <40, absence of obesity as well as insulin resistance and type-II diabetes mellitus, and female gender. Patients with cirrhosis can respond, but they are less likely to do so. Studies of combination IFN-ribavirin therapy have shown that in patients with genotype 1, therapy should last a full 48 weeks, while in those with genotypes 2 and 3, a 24-week course of therapy suffices (although more recent observations allow refined tailoring of treatment duration based on rapidity of response, see below). The response rate in African Americans is disappointingly low for reasons that are not fully understood. Potentially contributing to, but not explaining entirely, low responsiveness in African Americans are a higher proportion with genotype 1, slower early viral kinetics during therapy, impaired HCV-specific immunity, and recently recognized host genetic differences in IL28B alleles, described below. The response rate in Latino patients is also low, despite the fact that the frequency of the favorable IL28B C allele is as common in Hispanic patients as in whites. Moreover, the likelihood of a sustained response is best if adherence to the treatment regimen is high (i.e., if patients receive ≥80% of the IFN and ribavirin doses and if they continue treatment for ≥80% of the anticipated duration of therapy). Other variables reported to correlate with increased responsiveness include brief duration of infection, low HCV quasispecies diversity, immunocompetence, absence of hepatic steatosis and insulin resistance, and low liver iron levels. High levels of HCV RNA, more histologically advanced liver disease, and high quasispecies diversity all go hand in hand with advanced duration of infection, which may be the single most important clinical variable determining IFN responsiveness. The ironic fact, then, is that patients whose disease is least likely to progress are the ones *most* likely to respond to interferon and vice versa.

Genetic changes in the virus may explain differences in treatment responsiveness in some patients (e.g., among patients with genotype 1b, responsiveness to IFN is enhanced in those with amino-acid-substitution mutations in the nonstructural protein 5A gene). As described above in the discussion of spontaneous recovery from acute hepatitis C, interferon gene variants discovered recently in gene-wide association studies have been shown to have a substantial impact on responsiveness of patients with genotype 1 to antiviral therapy. In studies of patients treated with PEG IFN and ribavirin, variants of the IL28B SNP that code for IFN-λ3 (a type-III IFN, the receptors for which are more discretely distributed than IFN α receptors and more concentrated in hepatocytes) correlate significantly with responsiveness. Patients homozygous for the C allele at this locus have the highest frequency of achieving an SVR (~80%), those homozygous for the T allele at this locus are least likely to achieve an SVR (~25%), and those heterozygous at this locus (C/T) have an intermediate level of responsiveness (SVRs in ~35%). The fact that C/C is common in whites of European ancestry and even more so in Japanese persons but rare in African Americans helps explain the differences in observed responsiveness among these population groups.

Side effects of IFN therapy are described above in the section on treatment of chronic hepatitis B. The most pronounced side effect of ribavirin therapy is hemolysis; a reduction in hemoglobin of up to 2–3 g or in hematocrit of 5–10% can be anticipated. A small, unpredictable proportion of patients experience profound, brisk hemolysis, resulting in symptomatic anemia; therefore, close monitoring of blood counts is crucial, and ribavirin should be avoided in patients with anemia or hemoglobinopathies and in patients with coronary artery disease or cerebrovascular disease, in whom anemia can precipitate an ischemic event. When symptomatic anemia occurs, ribavirin dose reductions or addition of erythropoietin to boost red blood cell levels may be required; erythropoietin has been shown to improve patients' quality of life but not the likelihood of achieving an SVR. If ribavirin is stopped during therapy, SVR rates fall, but responsiveness can be maintained as long as the ribavirin is not stopped and the total ribavirin dose exceed 60% of the planned dose. In addition, ribavirin, which is renally excreted, should not be used in patients with renal insufficiency; the drug is teratogenic, precluding its use during pregnancy and mandating the scrupulous use of efficient contraception during therapy (interferons, too, because of their antiproliferative properties, are contraindicated during pregnancy).

Ribavirin can also cause nasal and chest congestion, pruritus, and precipitation of gout. Combination IFN-ribavirin therapy is more difficult to tolerate than IFN monotherapy. In one large clinical trial of combination therapy versus monotherapy, among those in the 1-year treatment group, 21% of the combination group (but only 14% of the monotherapy group) had to discontinue treatment, while 26% of the combination group (but only 9% of the monotherapy group) required dose reductions.

Studies of viral kinetics have shown that despite a virion halflife in serum of only 2–3 h, the level of HCV is maintained by a high replication rate of 10^{12} hepatitis C virions per day. IFN-α blocks virion production or release with an efficacy that increases with increasing drug doses; moreover, the calculated death rate for infected cells during IFN therapy is inversely related to viral load; patients with the most rapid death rate of infected hepatocytes are more likely to achieve undetectable HCV RNA at 3 months; in practice, failure to achieve an early virologic response (EVR), a ≥2-$\log_{10}$ reduction in HCV RNA by week 12, predicts failure to experience a subsequent SVR. Similarly, patients in whom HCV RNA becomes undetectable within 4 weeks [i.e., who achieve a rapid virologic response (RVR)], have a very high likelihood of achieving a sustained virologic response (Fig. 306-2). Therefore, to achieve rapid viral clearance from serum and the liver, *high-dose induction therapy* has been advocated. In practice, however, high-dose induction with IFN-based therapy has not yielded higher sustained response rates.

Treatment of Choice

For the treatment of chronic hepatitis C, standard IFNs have now been supplanted by PEG IFNs. These have elimination times up to sevenfold longer than standard IFNs (i.e., a substantially longer half-life), and achieve prolonged concentrations, permitting administration once (rather than three times) a week. Instead of the frequent drug peaks (linked to side effects) and troughs (when drug is absent) associated with frequent administration of short-acting IFNs, administration of PEG IFNs results in drug concentrations that are more stable and sustained over time. Once-a-week PEG IFN monotherapy is twice as effective as monotherapy with its standard IFN counterpart, approaches the efficacy of combination standard IFN plus ribavirin, and is as well tolerated as standard IFNs, without more difficult-to-manage thrombocytopenia and leukopenia than standard IFNs. The current standard of care, however, is a combination of PEG IFN plus ribavirin.

Two PEG IFNs are available: PEG IFN α-2b and α-2a. PEG IFN α-2b consists of a 12-kD, linear PEG molecule bound to IFN α-2b, while PEG IFN α-2a consists of a larger, 40kD, branched PEG molecule bound to IFN α-2a; because of its larger size and smaller volume of extravascular distribution, PEG IFN α-2a can be given at a uniform dose independent of weight, while the dose of the smaller PEG IFN α-2b, which has a much wider volume distribution, must be weight-based **(Table 306-6)**. In the registration trial for PEG IFN α-2b plus ribavirin, the best regimen was 48 weeks of 1.5 μg/kg of PEG IFN once a week plus 800 mg of ribavirin daily. A post hoc analysis suggested that weight-based dosing of ribavirin would have been more effective than the fixed 800-mg dose used in the study. In the first registration trial for PEG IFN α-2a plus ribavirin, the best regimen was 48 weeks of 180 μg of PEG IFN plus 1000 mg (for patients <75 kg) to 1200 mg (for patients ≥75 kg) of ribavirin. Sustained virologic responses of 54 and 56% were reported in these two studies, respectively. A subsequent study of PEG IFN α-2a plus ribavirin showed that, for patients with genotypes 2 and 3, a duration of 24 weeks and a ribavirin dose of 800 mg was sufficient. Among the three studies, for patients in the optimal treatment arm, SVR rates for patients with genotype 1 were 42–51% and for patients with genotypes 2 and 3 rates were 76–82%. Between genotypes 2 and 3, genotype 3 is somewhat more refractory, and some authorities would extend therapy for a full 48 weeks in patients with genotype 3, especially if they have advanced hepatic fibrosis or cirrhosis and/or high-level HCV RNA.

Table 306-6 Pegylated Interferon α-2a and α-2b for Chronic Hepatitis C

	PEG IFN α-2b	PEG IFN α-2a
PEG size	12 kD linear	40 kD branched
Elimination half-life	54 hours	65 hours
Clearance	725 mL/hour	60 mL/hour
Dose	1.5 μg/kg (weight-based)	180 μg
Storage	Room temperature	Refrigerated

Ribavirin dose		
Genotype 1	800–1400 mg[a]	1000–1200 mg[b]
Genotype 2/3	800 mg	800 mg
Duration of therapy		
Genotype 1	48 weeks	48 weeks
Genotype 2/3	48 weeks[c]	24 weeks
Efficacy of combination Rx[d]	54%	56%
Genotype 1	40–42%	41–51%
Genotype 2/3	82%	76–78%

[a] In the registration trial for PEG IFN α-2b plus ribavirin, the optimal regimen was 1.5 μg of PEG IFN plus 800 mg of ribavirin; however, a posthoc analysis of this study suggested that higher ribavirin doses are better. In subsequent trials of PEG IFN α-2b with ribavirin in patients with genotype 1, the following daily ribavirin doses have been validated: 800 mg for patients weighing <65 kg, 1000 mg for patients weighing >65-85 kg, 1200 for patients weighing >85–105 kg, and 1400 mg for patients weighing >105 kg.

[b] 1000 mg for patients weighing <75 kg; 1200 mg for patients weighing ≥75 kg.

[c] In the registration trial for PEG IFN α-2b plus ribavirin, all patients were treated for 48 weeks; however, data from other trials of standard interferons and the other PEG IFN demonstrated that 24 weeks suffices for patients with genotypes 2 and 3. For patients with genotype 3 who have advanced fibrosis/cirrhosis and/or high-level HCV RNA, a full 48 weeks is preferable.

[d] Attempts to compare the two PEG IFN preparations based on the results of registration clinical trials are confounded by differences between trials of the two agents in methodological details (different ribavirin doses, different methods for recording depression, and other side effects) and study-population composition (different proportion with bridging fibrosis/cirrhosis, proportion from the United States versus international, mean weight, proportion with genotype 1, and proportion with high-level HCV RNA). In the head-to-head comparison of the two PEG IFN preparations in the "IDEAL" trial reported in 2009, the two drugs were comparable in tolerability and efficacy. PEG IFN α-2b was administered at [a] weekly weight-based dose of 1.0 μg/kg or 1.5 μg/kg, and PEG IFN α-2a was administered at a weekly fixed dose of 180 μg. For PEG IFN α-2b, daily ribavirin weight-based doses ranged between 800–1400 mg based on weight criteria (see footnote a, above), while for PEG IFN α-2a, daily ribavirin weight-based doses ranged between 1000–1200 mg (footnote b, above). For the two PEG IFN α-2b study arms, ribavirin dose reductions for ribavirin-associated adverse effects were done in 200–400-mg decrements; for PEG IFN α-2a, the ribavirin dose was reduced to 600 mg for intolerability. Sustained virologic responses occurred in 38.0% of the low-dose PEG IFN α-2b group, 39.8% of the standard, full-dose PEG IFN α-2b group, and 40.9% of the PEG IFN α-2a group.

Abbreviations: PEG, polyethylene glycol; PEG IFN, pegylated interferon; HCV RNA, hepatitis C virus RNA.

In the initial registration trials for combination PEG IFN plus ribavirin, both combination PEG IFN regimens were compared to standard IFN α-2b plus ribavirin. Side effects of the combination PEG IFN α-2b regimen were comparable to those for the combination standard IFN regimen; however, when the combination PEG IFN α-2a regimen was compared to the combination standard IFN α-2b regimen, flu-like symptoms and depression were less common in the combination PEG IFN group. Although ascertainment of side effects differed between studies of the two drugs, when each was tested against standard IFN α-2b plus ribavirin, combination PEG IFN α-2a plus ribavirin appeared to be better tolerated. In a recent head-to-head trial of the two PEG IFNs (the "IDEAL" trial), the two PEG IFNs were found to be comparable in efficacy (achievement of SVR) **(Fig. 306-3)** and tolerability, although headache, nausea, fever, myalgia, depression, and drug discontinuation for any reason were less frequent in patients treated with PEG IFN α-2a than standard-dose PEG IFN α-2b. In contrast, neutropenia and rash were more frequent in patients treated with PEG IFN α-2a than standarddose PEG IFN α-2b. In two subsequent head-to-head trials and a systematic review of randomized trials, PEG IFN α–2a was more effective than α–2b (SVR in genotype 1-4: 48–55% versus 32–40%, respectively). In trials of PEG IFN α-2b among patients with HCV genotype 1, a broader range of weight-based daily ribavirin doses has been validated: 800 mg for weight <65 kg, 1000 mg for weight 65–85 kg, 1200 mg for weight >85–105 kg, and 1400 mg for weight >105 kg. Recommended doses for the two PEG IFNs plus ribavirin and other comparisons between the two therapies are shown in Table 306-6.

Figure 306-3

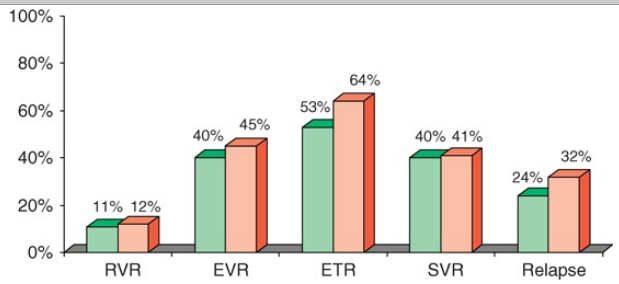

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Head-to-head comparison of standard-dose PEG IFN α-2b 1.5 μg/kg weekly and PEG IFN α-2a 180 μg weekly administered with daily ribavirin in the "IDEAL" trial. Percent achieving treatment milestones for PEG IFN α-2b (green boxes) and PEG IFN α-2a (orange boxes). RVR, rapid virologic response, HCV RNA undetectable at week 4; EVR, early virologic response, HCV RNA undetectable at week 12; ETR, end-treatment response, HCV RNA undetectable at end of treatment week 48; SVR, sustained virologic response, HCV RNA remaining undetectable 24 weeks after completing 48 weeks of therapy. Relapse, reappearance of detectable HCV RNA by week 72 in patients with an end-treatment response at week 48. PEG IFN α-2a suppressed HCV RNA in a higher proportion of patients at weeks 12 and 48 but, because of a higher relapse rate at week 72, resulted in the same SVR rate as PEG IFN α-2b.

Unless ribavirin is contraindicated (see above), combination PEG IFN plus ribavirin is the recommended course of therapy—24 weeks for genotypes 2 and 3 and 48 weeks for genotype 1. Measurement of quantitative HCV RNA levels at 12 weeks is helpful in guiding therapy; if a 2-log 10 drop in HCV RNA has not been achieved by this time, chances for an SVR are negligible. If the 12-week HCV RNA has fallen by two logs 10 (EVR), the chances for an SVR at the end of therapy are approximately two-thirds; if the 12-week HCV RNA is undetectable ("complete" EVR), the chances for a sustained virologic response exceed 80% (Fig. 306-2). Because absence of an EVR is such a strong predictor of the absence of an ultimate sustained virologic response, failure to achieve a 12-week 2-log 10 drop in HCV RNA (EVR) may be used as a signal to discontinue therapy.

Studies have suggested that the frequency of an SVR to PEG IFN/ribavirin therapy can be increased in patients with baseline variables weighing against a response (e.g., HCV RNA >8 x 10^5 IU/ml, weight >85 kg) by raising the dose of PEG IFN (e.g., to as high as 270 μg of PEG IFN α-2a) and/or the dose of ribavirin to as high as 1600 mg daily (if tolerated or supplemented by erythropoietin) or by tailoring treatment based on viral response to prolong the duration of viral clearance before discontinuing therapy, i.e., extending therapy from 48 to 72 weeks for patients with genotype 1 and a slow virologic response, i.e., those whose HCV RNA has not fallen rapidly to undetectable levels within 4 weeks (absence of "rapid virologic response"). Tailoring therapy based on the kinetics of HCV RNA reduction has also been applied to abbreviating the duration of therapy in patients with genotype 1 (and 4). The results of several clinical trials suggest that, in patients with genotype 1 (and 4) who have a 4-week RVR (which occurs in ≤20%), especially in the subset with a baseline low level of HCV RNA, 24 weeks of therapy with PEG IFN and weight-

based ribavirin suffices, yielding SVR rates of ~90% and comparable to those achieved in this cohort with 48 weeks of therapy. Although initial reports suggested that, for patients with genotype 2 and (somewhat less so) genotype 3, in rapid virologic responders with undetectable HCV RNA at week 4, the total duration of therapy required to achieve an SVR could be as short as 12–16 weeks, a very sizable, definitive subsequent trial showed that relapse is increased if treatment duration is curtailed and that a full 24 weeks is superior for these genotypes (except for the minority with very low baseline levels of HCV RNA).

Persons with chronic HCV infection have been shown to suffer increased liver-related mortality. On the other hand, successful antiviral therapy of chronic hepatitis C resulting in an SVR has been shown to improve survival, to lower the risk of liver failure and liver-related death, to slow the progression of chronic hepatitis C, and to reverse fibrosis and even cirrhosis. Although successful treatment reduces mortality in cirrhotic patients (and those with advanced fibrosis) and reduces the likelihood of hepatocellular carcinoma, the risk of decompensation, death, and liver cancer persists, albeit at a much reduced level, necessitating continued clinical monitoring and cancer surveillance after SVR in cirrhotics. On the other hand, in the absence of an SVR, IFNbased therapy does not reduce the risk of hepatocellular carcinoma. Similarly, for nonresponders to PEG IFN/ribavirin therapy, three trials of long-term maintenance therapy with PEG IFN have shown no benefit in reducing the risk of histologic progression or clinical decompensation, including the development of hepatocellular carcinoma. For PEG IFN/ribavirin nonresponders who have had a full, adequate course of therapy, the benefit of retreatment—with higher doses or a longer course of the original PEG IFN regimen or the alternative PEG IFN regimen or with a different type of IFN preparation (e.g., consensus IFN)—is marginal at best.

Indications for Antiviral Therapy

Patients with chronic hepatitis C who have detectable HCV RNA in serum, whether or not aminotransferase levels are increased, and chronic hepatitis of at least moderate grade and stage (portal or bridging fibrosis) are candidates for antiviral therapy with PEG IFN plus ribavirin. Most authorities recommend 800 mg of ribavirin for patients with genotypes 2 and 3 for both types of PEG IFN and weight-based 1000–1200 mg (when used with PEG IFN α-2a) or 800–1400 mg (when used with PEG IFN α-2b) ribavirin for patients with genotype 1 (and 4), unless ribavirin is contraindicated **(Table 306-7)**. Although patients with persistently normal ALT activity tend to progress histologically very slowly or not at all, they respond to antiviral therapy just as well as do patients with elevated ALT levels; therefore, while observation without therapy is an option, such patients are potential candidates for antiviral therapy. As noted above, therapy with IFN has been shown to improve survival and complication-free survival and to slow progression of fibrosis.

Table 306-7 Indications and Recommendations for Antiviral Therapy of Chronic Hepatitis C

Standard Indications for Therapy
Detectable HCV RNA (with or without elevated ALT)
Portal/bridging fibrosis or moderate to severe hepatitis on liver biopsy (the necessity of a pretreatment biopsy is being debated).
These indications apply to adults as well as to children aged 2–17, in whom treatment may be considered at reduced weight-based doses (see product inserts).
Retreatment Recommended
Relapsers after a previous course of standard interferon monotherapy or combination standard interferon/ribavirin therapy.
A course of PEG IFN plus ribavirin (retreatment not recommended with PEG IFN/ribavirin if relapse occurred after a full course of PEG IFN/ribavirin).
Nonresponders to a previous course of standard IFN monotherapy or combination standard IFN/ribavirin therapy.
A course of PEG IFN plus ribavirin—more likely to achieve a sustained virologic response in white patients without previous ribavirin therapy, with low baseline HCV RNA levels, with a ≥2-$\log_{10}$ reduction in HCV RNA during previous therapy, with genotypes 2 and 3, and without reduction in ribavirin dose. (Retreatment not recommended with PEG IFN/ribavirin if nonresponse occurred to a full course of PEG IFN/ribavirin.)
Antiviral Therapy not Recommended Routinely but Management Decisions Made on an Individual Basis
Age >60
Mild hepatitis on liver biopsy.
Persons with severe renal insufficiency (glomerular filtration rate <60 ml/min) who do not require hemodialysis (reduced-dose PEG IFN and ribavirin). Antiviral therapy in patients requiring hemodialysis is more complicated, less successful, and associated with more adverse effects; if treatment is pursued, either standard doses of standard interferon 3 times a week or reduced doses of weekly PEG IFN in combination with reduced doses of daily ribavirin should be used.
Long-Term Maintenance Therapy Recommended
Cutaneousvasculitis and glomerulonephritis associated with chronic hepatitis C.
Long-Term Maintenance Therapy in Nonresponders not Recommended
Antiviral Therapy not Recommended
Decompensated cirrhosis (except, perhaps, in transplantation centers with experience in graded escalation, low-dose treatment to achieve undetectable HCV RNA prior to transplantation; results are mixed).
Pregnancy (teratogenicity of ribavirin).
Contraindications to use of interferon or ribavirin.
Standard Therapeutic Regimens
First-line treatment: PEG IFN subcutaneously once a week plus daily ribavirin orally
HCV genotypes 1 and 4—48 weeks of therapy
PEG IFN α-2a 180 μg weekly plus ribavirin 1000 mg/day (weight <75 kg) to 1200 mg/day (weight <75 kg) or
PEG IFN α-2b 1.5 μg/kg weekly plus daily oral ribavirin 800 mg for weight <65 kg, 1000 mg for weight 65–85 kg, 1200 mg for weight >85–105 kg, and 1400 mg for weight >105 kg
HCV genotypes 2 and 3—24 weeks of therapy
PEG IFN α-2a 180 μg weekly plus ribavirin 800 mg/day or
PEG IFN α-2b 1.5 μg/kg weekly plus ribavirin 800 mg/day (For patients with genotype 3 who have advanced fibrosis and/or high-level HCV RNA, a full 48 weeks of therapy may be preferable.)
Alternative regimen: PEG IFN (α-2a 180 μg or α-2b 1.0 μg/kg) subcutaneously once a week (primarily for patients in whom ribavirin is contraindicated or not tolerated) for 24 (genotypes 2 and 3) or 48 (genotypes 1 and 4) weeks.
Early discontinuation: Failure to achieve an EVR, i.e., ≥2 $\log^{10}$ HCV RNA reduction by week 12 or, if EVR is achieved, failure to achieve suppression of HCV RNA to undetectable by week 24.
"Tailored" Therapeutic Regimens Based on Rapid Treatment Milestones
HCV genotypes 1 and 4.
For RVR, i.e., undetectable HCV RNA at week 4, especially in patients with low baseline HCV RNA, consider truncating the course of therapy to 24 weeks.
For patients with slow, delayed response, i.e., who clear detectable HCV RNA between weeks 12 and 24, consider prolonging the course of therapy to 72 weeks.
"Tailored" Therapeutic Regimens Based on Baseline Variables Associated with Reduced Responsiveness

HCV genotypes 1 and 4 For patients with HCV RNA >8 x 10^5 IU/ml and weighing >85 kg, consider increasing the weekly PEG IFN dose (e.g., for PEG IFN α-2a up to 270 μg) and the daily ribavirin dose (e.g., up to 1600 mg). *For HCV-HIV co-infected patients:* 48 weeks, regardless of genotype, of weekly PEG IFN α-2a (180 μg) or PEG IFN α-2b (1.5 μg/kg) plus a daily ribavirin dose of at least 600–800 mg, up to full weight-based dosing, at doses comparable to those for HCV-monoinfected patients, if tolerated.
Features Associated with Reduced Responsiveness
Single nucleotide polymorphism (SNP) T allele (as opposed to C allele) at IL28B locus
Genotype 1
High-level HCV RNA (>2 x 10^6 copies/ml or >8 x 10^5 IU/ml)
Advanced fibrosis (bridging fibrosis, cirrhosis)
Long-duration disease
Age >40
High HCV quasispecies diversity
Immunosuppression
African-American ethnicity
Latino ethnicity
Obesity
Hepatic steatosis
Insulin resistance, type-II diabetes mellitus
Reduced adherence (lower drug doses and reduced duration of therapy)

Abbreviations: ALT, alanine aminotransferase; HCV, hepatitis C virus; IFN, interferon; PEG IFN, pegylated interferon; IU, international units (1 IU/ml is equivalent to ~2.5 copies/ml).

Prior to therapy, HCV genotype should be determined, and the genotype dictates the duration of therapy: 48 weeks for patients with genotype 1, 24 weeks for those with genotypes 2 and 3. For patients with genotype 1 (and 4), especially those with low baseline HCV RNA, 24 weeks of PEG IFN/ribavirin therapy may suffice if HCV RNA becomes undetectable within 4 weeks (RVR); for patients with genotypes 2 and 3, a full, 24-week course is most effective, although the duration may be reduced to 12–16 weeks for patients with genotype 2, a low baseline level of viremia, and an RVR, especially to be considered for patients who tolerate therapy poorly. As noted above, the absence of a 2-log 10 drop in HCV RNA at week 12 (EVR) weighs heavily against the likelihood of an SVR; therefore, measuring HCV RNA at 12 weeks is recommended routinely (Fig. 306-2), especially for patients with genotype 1, and therapy can be discontinued if an EVR is not achieved. Among patients with an EVR (≥2-$\log_{10}$ HCV RNA reduction) but with HCV RNA still detectable at week 24, an SVR is unlikely, and therapy can be discontinued. Although response rates are lower in patients with certain pretreatment variables, selection for treatment should not be based on symptoms, genotype, HCV RNA level, mode of acquisition of hepatitis C, or advanced hepatic fibrosis. Patients with cirrhosis can respond and should not be excluded as candidates for therapy.

Patients who have relapsed (Fig. 306-2) after a course of IFN monotherapy are candidates for retreatment with PEG IFN plus ribavirin (i.e., a more effective treatment regimen is required). For nonresponders to a prior course of IFN monotherapy, retreatment with IFN monotherapy or combination IFN plus ribavirin therapy is unlikely to achieve a sustained virologic response; however, a trial of combination PEG IFN plus ribavirin may be worthwhile. End-treatment virologic responses as high as 40% can occur in this setting, but an SVR is the outcome in <15–20% of patients. Sustained virologic responses to retreatment of nonresponders are more frequent in those who had never received ribavirin in the past, those with genotypes 2 and 3, those with low pretreatment HCV RNA levels, and noncirrhotics, but less frequent in African Americans, those who failed to achieve a substantial reduction in HCV RNA during their previous course of therapy (null responders, Fig. 306-2), and those who required ribavirin-dose reductions. Potential approaches to improving responsiveness to PEG IFN/ribavirin in prior nonresponders include longer duration of treatment; higher doses of either PEG IFN, ribavirin, or both; and switching to a different IFN preparation; however, as noted above, none of these approaches achieves more than a marginal benefit.

Early treatment is indicated for persons with acute hepatitis C **(Chap. 304)**. In patients with biochemically and histologically mild chronic hepatitis C, the rate of progression is slow, and monitoring without therapy is an option; however, such patients respond just as well to combination PEG IFN plus ribavirin therapy as those with elevated ALT and more histologically severe hepatitis. Therefore, therapy for these patients should be considered and the decision made based on such factors as patient motivation, genotype, stage of fibrosis, age, and comorbid conditions. A pretreatment liver biopsy to assess histologic grade and stage provides substantial information about progression of hepatitis C in the past, has prognostic value for future progression, and can identify such histologic factors as steatosis and stage of fibrosis, which can influence responsiveness to therapy. As therapy has improved for patients with a broad range of histologic severity, and as noninvasive laboratory markers and imaging correlates of fibrosis have gained popularity, some authorities, especially in Europe, have placed less value on, and do not recommend, pretreatment liver biopsies. On the other hand, serum markers of fibrosis are not considered sufficiently accurate, and histologic findings provide important prognostic information to physician and patient. Therefore, although the contemporary role of a pretreatment liver biopsy commands less of a consensus, a pretreatment liver biopsy still provides useful information and should be considered.

Patients with compensated cirrhosis can respond to therapy, although their likelihood of a sustained response is lower than in noncirrhotics; moreover, survival has been shown to improve after successful antiviral therapy in cirrhotics. Similarly, although several retrospective studies have suggested that antiviral therapy in cirrhotics with chronic hepatitis C, independent of treatment outcome per se, reduces the frequency of HCC, less advanced disease in the treated cirrhotics, not treatment itself (i.e., lead-time bias), may have accounted for the reduced frequency of HCC observed in the treated cohorts in these reports; prospective studies to address this question have failed to demonstrate benefit, unless a sustained virologic response is achieved. Patients with decompensated cirrhosis are not candidates for IFN-based antiviral therapy but should be referred for liver transplantation. Some liver-transplantation centers have evaluated progressively escalated, low-dose antiviral therapy in an attempt to eradicate hepatitis C viremia prior to transplantation; however, such therapy has been shown to reduce but not to prevent the risk of HCV reinfection after transplantation. After liver transplantation for end-stage liver disease caused by hepatitis C, recurrent hepatitis C is the rule, and the pace of disease progression is more accelerated than in immunocompetent patients **(Chap. 310)**. Current therapy with PEG IFN and ribavirin after liver transplantation is unsatisfactory in most patients, but attempts to minimize immunosuppression are beneficial. The cutaneous and renal vasculitis of HCV-associated essential mixed cryoglobulinemia **(Chap. 304)** may respond to antiviral therapy, but sustained responses are rare after discontinuation of therapy; therefore, prolonged, perhaps indefinite, therapy is recommended in this group. Anecdotal reports suggest that antiviral therapy may be effective in porphyria cutanea tarda or lichen planus associated with hepatitis C.

In patients with HCV/HIV co-infection, hepatitis C is more progressive and severe than in HCV-monoinfected patients. Although patients with HCV/HIV co-infection respond to antiviral therapy for hepatitis C, they do not respond as well as patients with HCV infection alone. Four large national and international trials of antiviral therapy among patients with HCV/HIV co-infection have shown that PEG IFN (both α-2a and α-2b) plus ribavirin (daily doses ranging from flat-dosed 600–800 mg to weight-based 1000/1200 mg) is superior to standard IFN regimens; however, SVR rates were lower than in HCV-monoinfected patients, ranging from 14 to 38% for patients with genotypes 1 and 4 and from 44 to 73% for patients with genotypes 2 and 3. In the three largest trials, all patients, including those with genotypes 2 and 3, were treated for a full 48 weeks. In addition, tolerability of therapy was lower than in HCV-monoinfected patients; therapy was discontinued because of side effects in 12–39% of patients in these clinical trials. Based on these trials, weekly PEG IFN plus daily ribavirin at a daily dose of at least 600–800 mg, up to full weight-based doses, at doses recommended for HCV-monoinfected patients, if tolerated, is recommended for a full 48 weeks, regardless of genotype. An alternative recommendation for ribavirin doses was issued by a European Consensus Conference and consisted of standard, weight-based 1000–1200 mg for genotypes 1 and 4, but 800 mg for genotypes 2 and 3. A head-to-head trial of combination PEG IFN/ribavirin therapy in HCV/HIV co-infection demonstrated statistically indistinguishable efficacy of the two types of PEG IFN, despite a small advantage for PEG IFN α-2a: for PEG IFN α-2b and α-2a, SVRs occurred in 28% versus 32%, respectively, of patients with genotypes 1 and 4 and in 62% versus 71%, respectively, of patients with genotypes 2 and 3. In HCV/HIV-infected patients, ribavirin can potentiate the toxicity of didanosine (e.g., lactic acidosis) and the

lipoatrophy of stavudine, and zidovudine can exacerbate ribavirin-associated hemolytic anemia; therefore, these drug combinations should be avoided.

Patients with a history of injection-drug use and alcoholism can be treated successfully for chronic hepatitis C, preferably in conjunction with drug- and alcohol-treatment programs. Because ribavirin is excreted renally, patients with end-stage renal disease, including those undergoing dialysis (which does not clear ribavirin), are not ideal candidates for ribavirin therapy. Rare reports suggest that reduced-dose ribavirin can be used, but the frequency of anemia is very high and data on efficacy are limited. If patients with renal failure (glomerular filtration rate <60 ml/min) are treated, the PEG IFN α-2a dose should be reduced from 180 to 135 μg weekly and the PEG IFN α-2b dose reduced from 1.5 to 1 μg/kg weekly; similarly, the daily ribavirin dose in this population should be reduced to 200–800 mg (but not used or used cautiously at very low doses) if hemodialysis is required. Neither the optimal regimen nor the efficacy of therapy is well established in this population.

Novel Antivirals

To date, attempts to develop better-tolerated ribavirin successors or improved types of IFN α or longer acting IFNs than PEG IFN have not been successful. The demonstration that responsiveness to antiviral therapy is influenced by genetic variation in IL28B, which codes for IFN-λ (as noted above), raises the possibility that IFN-λ might be an effective or even more effective IFN for treating hepatitis C; early trials are in progress. Among the most exciting new approaches to antiviral therapy are orally administered direct antivirals that target HCV polymerase or protease. Two protease inhibitors that are in late stages of development, are expected to be approved in 2011. The NS3-4A serine protease inhibitors telaprevir and boceprevir suppress HCV RNA profoundly and, when used together with PEG IFN and ribavirin in patients with genotype-1 HCV infection, can increase RVR rates to as high as 80% (telaprevir) and SVR rates from those achieved with current standard-of-care therapy by 20–30% to ~65–75%, in most patients with only half the duration of current therapy. These triple-drug combinations appear to yield even higher rates of SVR in >50% of prior relapsers (>70–90%) but also to achieve SVR in prior nonresponders, even in null responders to PEG IFN/ribavirin therapy (~30%). Although these new drugs add elements of additional toxicity (severe rash in ~5% of telaprevir-treated patients and anemia in half of boceprevir-treated patients), they represent an opportunity for curing a substantially larger proportion of patients with shorter treatment courses. Because resistance to these oral agents used alone has been both anticipated and observed, polymerase and protease inhibitors are being evaluated in combinations with PEG IFN and ribavirin to preempt the emergence of resistance. Potentially, in the future, combinations of direct antiviral agents will be used in drug cocktails that may replace IFN-based regimens entirely.

AUTOIMMUNE HEPATITIS

Definition

Autoimmune hepatitis is a chronic disorder characterized by continuing hepatocellular necrosis and inflammation, usually with fibrosis, which can progress to cirrhosis and liver failure. When fulfilling criteria of severity, this type of chronic hepatitis, when untreated, may have a 6-month mortality of as high as 40%. Based on contemporary estimates of the natural history of treated autoimmune hepatitis, the 10-year survival is 80–90%. The prominence of extrahepatic features of autoimmunity as well as seroimmunologic abnormalities in this disorder supports an autoimmune process in its pathogenesis; this concept is reflected in the labels *lupoid*, *plasma cell*, or *autoimmune hepatitis*. Autoantibodies and other typical features of autoimmunity, however, do not occur in all cases; among the broader categories of "idiopathic" or cryptogenic chronic hepatitis, many, perhaps the majority, are probably autoimmune in origin. Cases in which hepatotropic viruses, metabolic/genetic derangements, and hepatotoxic drugs have been excluded represent a spectrum of heterogeneous liver disorders of unknown cause, a proportion of which are most likely autoimmune hepatitis.

Immunopathogenesis

The weight of evidence suggests that the progressive liver injury in patients with autoimmune hepatitis is the result of a cell-mediated immunologic attack directed against liver cells. In all likelihood, predisposition to autoimmunity is inherited, while the liver specificity of this injury is triggered by environmental (e.g., chemical or viral) factors. For example, patients have been described in whom apparently self-limited cases of acute hepatitis A, B, or C led to autoimmune hepatitis, presumably because of genetic susceptibility or predisposition. Evidence to support an autoimmune pathogenesis in this type of hepatitis includes the following: (1) In the liver, the histopathologic lesions are composed predominantly of cytotoxic T cells and plasma cells; (2) circulating autoantibodies (nuclear, smooth muscle, thyroid, etc.; see below), rheumatoid factor, and hyperglobulinemia are common; (3) other autoimmune disorders—such as thyroiditis, rheumatoid arthritis, autoimmune hemolytic anemia, ulcerative colitis, membranoproliferative glomerulonephritis, juvenile diabetes mellitus, celiac disease, and Sjögren's syndrome—occur with increased frequency in patients and in their relatives who have autoimmune hepatitis; (4) histocompatibility haplotypes associated with autoimmune diseases, such as HLA-B1, -B8, -DR3, and -DR4 as well as extended haplotype DRB1 alleles, are common in patients with autoimmune hepatitis; and (5) this type of chronic hepatitis is responsive to glucocorticoid/immunosuppressive therapy, effective in a variety of autoimmune disorders.

Cellular immune mechanisms appear to be important in the pathogenesis of autoimmune hepatitis. In vitro studies have suggested that in patients with this disorder, lymphocytes are capable of becoming sensitized to hepatocyte membrane proteins and of destroying liver cells. Abnormalities of immunoregulatory control over cytotoxic lymphocytes (impaired regulatory CD4+CD25+ T cell influences) may play a role as well. Studies of genetic predisposition to autoimmune hepatitis demonstrate that certain haplotypes are associated with the disorder, as enumerated above. The precise triggering factors, genetic influences, and cytotoxic and immunoregulatory mechanisms involved in this type of liver injury remain incompletely defined.

Intriguing clues into the pathogenesis of autoimmune hepatitis come from the observation that circulating autoantibodies are prevalent in patients with this disorder. Among the autoantibodies described in these patients are antibodies to nuclei [so-called antinuclear antibodies (ANAs), primarily in a homogeneous pattern] and smooth muscle (so-called anti-smooth-muscle antibodies, directed at actin), anti-LKM (see below), antibodies to "soluble liver antigen/liver pancreas antigen" (directed against a uracil-guanineadenine transfer RNA suppressor protein), as well as antibodies to the liver-specific asialoglycoprotein receptor (or "hepatic lectin") and other hepatocyte membrane proteins. Although some of these provide helpful diagnostic markers, their involvement in the pathogenesis of autoimmune hepatitis has not been established.

Humoral immune mechanisms have been shown to play a role in the extrahepatic manifestations of autoimmune and idiopathic hepatitis. Arthralgias, arthritis, cutaneousvasculitis, and glomerulonephritis occurring in patients with autoimmune hepatitis appear to be mediated by the deposition of circulating immune complexes in affected tissue vessels, followed by complement activation, inflammation, and tissue injury. While specific viral antigen-antibody complexes can be identified in acute and chronic viral hepatitis, the nature of the immune complexes in autoimmune hepatitis has not been defined.

Many of the *clinical features* of autoimmune hepatitis are similar to those described for chronic viral hepatitis. The onset of disease may be insidious or abrupt; the disease may present initially like, and be confused with, acute viral hepatitis; a history of recurrent bouts of what had been labeled *acute hepatitis* is not uncommon. A subset of patients with autoimmune hepatitis has distinct features. Such patients are predominantly young to middle-aged women with marked hyperglobulinemia and high-titer circulating ANAs. This is the group with positive lupus erythematosus (LE) preparations (initially labeled "lupoid" hepatitis) in whom other autoimmune features are common. Fatigue, malaise, anorexia, amenorrhea, acne, arthralgias, and jaundice are common. Occasionally arthritis, maculopapular eruptions (including cutaneousvasculitis), erythema nodosum, colitis, pleurisy, pericarditis, anemia, azotemia, and sicca syndrome (keratoconjunctivitis, xerostomia) occur. In some patients, complications of cirrhosis, such as ascites and edema (associated with hypoalbuminemia), encephalopathy, hypersplenism, coagulopathy, or variceal bleeding may bring the patient to initial medical attention.

The course of autoimmune hepatitis may be variable. In those with mild disease or limited histologic lesions (e.g., piecemeal necrosis without bridging), progression to cirrhosis is limited. In those with severe symptomatic autoimmune hepatitis (aminotransferase levels >10 times normal, marked hyperglobulinemia, "aggressive" histologic lesions—bridging necrosis or multilobular collapse, cirrhosis), the 6-month mortality without therapy may be as high as 40%. Such severe disease accounts for only 20% of cases; the natural history of milder disease is variable, often accentuated by spontaneous remissions and exacerbations. Especially poor prognostic signs include the presence histologically of multilobular collapse at the time of initial presentation and failure of the bilirubin to improve after 2 weeks of therapy. Death may result from hepatic failure, hepatic coma, other complications of cirrhosis (e.g., variceal hemorrhage), and intercurrent infection. In patients with established cirrhosis, HCC may be a late complication **(Chap. 92)** but occurs less frequently than in cirrhosis associated with viral hepatitis.

Laboratory features of autoimmune hepatitis are similar to those seen in chronic viral hepatitis. Liver biochemical tests are invariably abnormal but may not correlate with the clinical severity or histopathologic features in individual cases. Many patients with autoimmune hepatitis have normal serum bilirubin, alkaline phosphatase, and globulin levels with only minimal aminotransferase elevations. Serum AST and ALT levels are increased and fluctuate in the range of 100–1000 units. In severe cases, the serum bilirubin level is moderately elevated [51–171 μmol/L (3–10 mg/dL)]. Hypoalbuminemia occurs in patients with very active or advanced disease. Serum alkaline phosphatase levels may be moderately elevated or near normal. In a small proportion of patients, marked elevations of alkaline phosphatase activity occur; in such patients, clinical and laboratory features overlap with those of primary biliary cirrhosis **(Chap. 308)**. The prothrombin time is often prolonged, particularly late in the disease or

during active phases.

Hypergammaglobulinemia (>2.5 g/dL) is common in autoimmune hepatitis. Rheumatoid factor is common as well. As noted above, circulating autoantibodies are also prevalent. The most characteristic are ANAs in a homogeneous staining pattern. Smoothmuscle antibodies are less specific, seen just as frequently in chronic viral hepatitis. Because of the high levels of globulins achieved in the circulation of some patients with autoimmune hepatitis, occasionally the globulins may bind nonspecifically in solid-phase binding immunoassays for viral antibodies. This has been recognized most commonly in tests for antibodies to hepatitis C virus, as noted above. In fact, studies of autoantibodies in autoimmune hepatitis have led to the recognition of new categories of autoimmune hepatitis. *Type I autoimmune hepatitis* is the classic syndrome occurring in young women, associated with marked hyperglobulinemia, lupoid features, circulating ANAs, and HLA-DR3 or HLA-DR4 (especially B8-DRB1*03). Also associated with type I autoimmune hepatitis are autoantibodies against actin as well as atypical perinuclear antineutrophilic cytoplasmic antibodies (pANCA).

Type II autoimmune hepatitis, often seen in children, more common in Mediterranean populations, and linked to HLA-DRB1 and HLA-DQB1 haplotypes, is associated not with ANA but with anti-LKM. Actually, anti-LKM represent a heterogeneous group of antibodies. In type II autoimmune hepatitis, the antibody is anti-LKM1, directed against cytochrome P450 2D6. This is the same anti-LKM seen in some patients with chronic hepatitis C. Anti-LKM2 is seen in drug-induced hepatitis, and anti-LKM3 is seen in patients with chronic hepatitis D. Another autoantibody observed in type II autoimmune hepatitis is directed against liver cytosol formiminotransferase cyclodeaminase (anti-liver cytosol 1). More controversial is whether or not a third category of autoimmune hepatitis exists, *type III autoimmune hepatitis*. These patients lack ANA and anti-LKM1 but have circulating antibodies to soluble liver antigen/liver pancreas antigen. Most of these patients are women and have clinical features similar to, perhaps more severe than, those of patients with type I autoimmune hepatitis. Type III autoimmune hepatitis does not appear to represent a distinct category but, instead, is part of the spectrum of type I autoimmune hepatitis; this subcategory has not been adopted by a consensus of international experts.

Liver biopsy abnormalities are similar to those described for chronic viral hepatitis. Expanding portal tracts and extending beyond the plate of periportal hepatocytes into the parenchyma (designated *interface hepatitis* or *piecemeal necrosis*) is a mononuclear cell infiltrate that, in autoimmune hepatitis, may include the presence of plasma cells. Necroinflammatory activity characterizes the lobular parenchyma, and evidence of hepatocellular regeneration is reflected by "rosette" formation, the occurrence of thickened liver cell plates, and regenerative "pseudolobules." Septal fibrosis, bridging fibrosis, and cirrhosis are frequent. Bile duct injury and granulomas are uncommon; however, a subgroup of patients with autoimmune hepatitis has histologic, biochemical, and serologic features overlapping those of primary biliary cirrhosis **(Chap. 308)**.

Diagnostic Criteria

An international group has suggested a set of criteria for establishing a diagnosis of autoimmune hepatitis. Exclusion of liver disease caused by genetic disorders, viral hepatitis, drug hepatotoxicity, and alcohol are linked with such inclusive diagnostic criteria as hyperglobulinemia, autoantibodies, and characteristic histologic features. This international group has also suggested a comprehensive diagnostic scoring system that, rarely required for typical cases, may be helpful when typical features are not present. Factors that weigh in favor of the diagnosis include female gender; predominant aminotransferase elevation; presence and level of globulin elevation; presence of nuclear, smooth muscle, LKM1, and other autoantibodies; concurrent other autoimmune diseases; characteristic histologic features (interface hepatitis, plasma cells, rosettes); HLA DR3 or DR4 markers; and response to treatment (see below). Weighing against the diagnosis are predominant alkaline phosphatase elevation, mitochondrial antibodies, markers of viral hepatitis, history of hepatotoxic drugs or excessive alcohol, histologic evidence of bile duct injury, or such atypical histologic features as fatty infiltration, iron overload, and viral inclusions.

Differential Diagnosis

Early during the course of chronic hepatitis, autoimmune hepatitis may resemble typical *acute viral hepatitis***(Chap. 304)**. Without histologic assessment, severe chronic hepatitis cannot be readily distinguished based on clinical or biochemical criteria from mild chronic hepatitis. In adolescence, *Wilson's disease***(Chaps. 308 and 360)** may present with features of chronic hepatitis long before neurologic manifestations become apparent and before the formation of Kayser-Fleischer rings. In this age group, serum ceruloplasmin and serum and urinary copper determinations plus measurement of liver copper levels will establish the correct diagnosis. *Postnecrotic* or *cryptogenic cirrhosis* and *primary biliary cirrhosis***(Chap. 308)** share clinical features with autoimmune hepatitis, and both alcoholic hepatitis **(Chap. 307)** and nonalcoholic steatohepatitis **(Chap. 309)** may present with many features common to autoimmune hepatitis; historic, biochemical, serologic, and histologic assessments are usually sufficient to allow these entities to be distinguished from autoimmune hepatitis. Of course, the distinction between autoimmune and chronic viral hepatitis is not always straightforward, especially when viral antibodies occur in patients with autoimmune disease or when autoantibodies occur in patients with viral disease. Furthermore, the presence of extrahepatic features such as arthritis, cutaneousvasculitis, or pleuritis—not to mention the presence of circulating autoantibodies—may cause confusion with *rheumatologic disorders* such as rheumatoid arthritis and systemic lupus erythematosus. The existence of clinical and biochemical features of progressive necroinflammatory liver disease distinguishes chronic hepatitis from these other disorders, which are not associated with severe liver disease.

Finally, occasionally, features of autoimmune hepatitis overlap with features of autoimmune biliary disorders such as primary biliary cirrhosis, primary sclerosing cholangitis **(Chaps. 308 and 311)**, or, even more rarely, mitochondrial antibody-negative autoimmune cholangitis. Such overlap syndromes are difficult to categorize, and often response to therapy may be the distinguishing factor that establishes the diagnosis.

Treatment: Autoimmune Hepatitis

The mainstay of management in autoimmune hepatitis is glucocorticoid therapy. Several controlled clinical trials have documented that such therapy leads to symptomatic, clinical, biochemical, and histologic improvement as well as increased survival. A therapeutic response can be expected in up to 80% of patients. Unfortunately, therapy has not been shown to prevent ultimate progression to cirrhosis; however, instances of reversal of fibrosis and cirrhosis have been reported in patients responding to treatment. Although some advocate the use of prednisolone (the hepatic metabolite of prednisone), prednisone is just as effective and is favored by most authorities. Therapy may be initiated at 20 mg/d, but a popular regimen in the United States relies on an initiation dose of 60 mg/d. This high dose is tapered successively over the course of a month down to a maintenance level of 20 mg/d. An alternative, but equally effective, approach is to begin with half the prednisone dose (30 mg/d) along with azathioprine (50 mg/d). With azathioprine maintained at 50 mg/d, the prednisone dose is tapered over the course of a month down to a maintenance level of 10 mg/d. The advantage of the combination approach is a reduction, over the span of an 18-month course of therapy, in serious, life-threatening complications of steroid therapy from 66% down to under 20%. In combination regimens, 6-mercaptopurine may be substituted for its prodrug azathioprine, but this is rarely required. Azathioprine alone, however, is not effective in achieving remission, nor is alternateday glucocorticoid therapy. Limited experience with budesonide in noncirrhotic patients suggests that this steroid side effect–sparing drug may be effective. Although therapy has been shown to be effective for severe autoimmune hepatitis (AST ≥10 times the upper limit of normal or ≥5 times the upper limit of normal in conjunction with serum globulin greater than or equal to twice normal; bridging necrosis or multilobular necrosis on liver biopsy; presence of symptoms), therapy is not indicated for mild forms of chronic hepatitis, and the efficacy of therapy in mild or asymptomatic autoimmune hepatitis has not been established.

Improvement of fatigue, anorexia, malaise, and jaundice tends to occur within days to several weeks; biochemical improvement occurs over the course of several weeks to months, with a fall in serum bilirubin and globulin levels and an increase in serum albumin. Serum aminotransferase levels usually drop promptly, but improvements in AST and ALT alone do not appear to be reliable markers of recovery in individual patients; histologic improvement, characterized by a decrease in mononuclear infiltration and in hepatocellular necrosis, may be delayed for 6–24 months. Still, if interpreted cautiously, aminotransferase levels are valuable indicators of relative disease activity, and many authorities do *not* advocate for serial liver biopsies to assess therapeutic success or to guide decisions to alter or stop therapy. Rapidity of response is more common in older patients (≥69 years) and those with HLA DBR1*04; although rapid responders may progress less slowly to cirrhosis and liver transplantation, they are no less likely than slower responders to relapse after therapy. Therapy should continue for at least 12–18 months. After tapering and cessation of therapy, the likelihood of relapse is at least 50%, even if posttreatment histology has improved to show mild chronic hepatitis, and the majority of patients require therapy at maintenance doses indefinitely. Continuing azathioprine alone (2 mg/kg body weight daily) after cessation of prednisone therapy may reduce the frequency of relapse.

In medically refractory cases, an attempt should be made to intensify treatment with high-dose glucocorticoid monotherapy (60 mg daily) or combination glucocorticoid (30 mg daily) plus high-dose azathioprine (150 mg daily) therapy. After a month, doses of prednisone can be reduced by 10 mg a month, and doses of azathioprine can be reduced by 50 mg a month toward ultimate, conventional maintenance doses. Patients refractory to this regimen may be treated with cyclosporine, tacrolimus, or mycophenolate mofetil; however, to date, only limited anecdotal reports support these approaches. If medical therapy fails, or when chronic hepatitis progresses to cirrhosis and is associated with life-threatening complications of liver decompensation, liver transplantation is the only recourse **(Chap. 310)**; failure of the bilirubin to improve after 2 weeks of therapy should prompt early consideration of the patient for liver transplantation. Recurrence of autoimmune hepatitis in the new liver occurs rarely in most experiences but in as many as 35–40% of cases in others.

ACKNOWLEDGMENT

Kurt J. Isselbacher, MD, contributed to this chapter in previous editions of Harrison's.

FURTHER READINGS

Benhamou Y, Salmon D (guest ed): Proceedings of the 1 st European consensus conference on the treatment of chronic hepatitis B and C in HIV co-infected patients. J Hepatol 44(Suppl 1): S1, 2006

Chang T-T et al: A comparison of entecavir and lamivudine for HBeAg-positive chronic hepatitis B. N Engl J Med 354:1001, 2006[PMID: 16525137] [Full Text]

Czaja AJ, Freese DK: AASLD practice guidelines: Diagnosis and treatment of autoimmune hepatitis. Hepatology 36:479, 2002[PMID: 12143059] [Full Text]

——, Manns MP: Advances in the diagnosis, pathogenesis, and management of autoimmune hepatitis. Gastroenterology 139:58, 2010

Di Bisceglie AM et al: Prolonged therapy of advanced chronic hepatitis C with low dose peginterferon. N Engl J Med 359:2429, 2008

Dienstag JL, Mchutchison JG: American Gastroenterological Association medical position statement on the management of hepatitis C. Gastroenterology 130:225, 2006[PMID: 16401485] [Full Text]

——: American Gastroenterological Association technical review on the management of hepatitis C. Gastroenterology 130:231, 2006

——: Hepatitis B virus infection. N Engl J Med 359:1486, 2008

European Association for the Study of the Liver: EASL Consensus Conference on Hepatitis B. J Hepatol 39(Suppl 1):S1, 2003

———: EASL practice guidelines: Management of chronic hepatitis B. J Hepatol 50:227, 2009

Fung SK, Lok ASF: Treatment of chronic hepatitis B: Who to treat, what to use, and for how long? Clin Gastroenterol Hepatol 2:839, 2004[PMID: 15476146] [Full Text]

Gale MJ, Foy EM: Evasion of intracellular host defence by hepatitis C virus. Nature 436:939, 2005[PMID: 16107833] [Full Text]

Ge D et al: Genetic Variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461:399, 2009[PMID: 19684573] [Full Text]

Ghany MG et al: AASLD practice guidelines: Diagnosis, management, and treatment of hepatitis C: An update. Hepatology 49:1335, 2009[PMID: 19330875] [Full Text]

Hézode C et al: Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. N Engl J Med 360:1839, 2009[PMID: 21698804] [Full Text]

Krawitt EL: Autoimmune hepatitis. N Engl J Med 354:54, 2006[PMID: 16394302] [Full Text]

Lai C-L et al: Entecavir versus lamivudine for patients with HBeAg-negative chronic hepatitis B. N Engl J Med 354:1011, 2006[PMID: 16525138] [Full Text]

Liaw Y-F et al: Asian-Pacific consensus statement on the management of chronic hepatitis B: A 2008 update. Hepatol International 2:263, 2008[PMID: 19669255] [Full Text]

Locarnini SA (guest ed): The control of hepatitis B: The role for chemoprevention. Semin Liver Dis 26:1, 2006

Lok ASF, McMahon BJ: AASLD practice guidelines: Chronic hepatitis B: An update. Hepatology 45:507, 2007[PMID: 17256718] [Full Text]

——: AASLD practice guideline update: Chronic hepatitis B: Update 2009. Hepatology 50:661, 2009 (full version accessible at http://publish.aasld.org/practiceguidelines/Documents/Bookmarked%20Practice%20Guidelines/Chronic_Hep_B_Update_2009%208_24_2009.pdf, accessed September 20, 2009)

McHutchison JG et al: Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. N Engl J Med 361:580, 2009[PMID: 19625712] [Full Text]

——: Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. N Engl J Med 360:1827, 2009

National Institutes of Health Consensus Development Conference: Management of hepatitis C. Hepatology 36:1S, 2002

——: Management of hepatitis B. Hepatology 49 L:S1, 2009

O'Brien TR: Interferon-alfa, interferon-α and hepatitis C. Nature Genet 41:1048, 2009[PMID: 21660345] [Full Text]

Pawlotsky J-M et al: The hepatitis C virus life cycle as a target for new antiviral therapies. Gastroenterology 132:1979, 2007[PMID: 17484890] [Full Text]

Pearlman BL: Chronic hepatitis C therapy: Changing the rules of duration. Clin Gastroenterol Hepatol 4:963, 2006[PMID: 16843732] [Full Text]

Thomas DL et al: Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 461:798, 2009[PMID: 19759533] [Full Text]

Toy M et al: Potential impact of long-term nucleoside therapy on the mortality and morbidity of active chronic hepatitis B. Hepatology 50:743, 2009[PMID: 19585616] [Full Text]

Wong WW-S et al: Surrogate end points and long-term outcome in patients with chronic hepatitis B. Clin Gastroenterol Hepatol 7:1113, 2009[PMID: 19500693] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ALCOHOLIC LIVER DISEASE: INTRODUCTION

Chronic and excessive alcohol ingestion is one of the major causes of liver disease. Per capita, alcohol consumption and cirrhosis have risen in the last decade in United Kingdom and Russia but has decreased in many developed countries including the United States. The pathology of alcoholic liver disease consists of three major lesions, with the injury rarely existing in a pure form: (1) fatty liver, (2) alcoholic hepatitis, and (3) cirrhosis. Fatty liver is present in >90% of binge and chronic drinkers. A much smaller percentage of heavy drinkers will progress to alcoholic hepatitis, thought to be a precursor to cirrhosis. The prognosis of severe alcoholic liver disease is dismal; the mortality of patients with alcoholic hepatitis concurrent with cirrhosis is nearly 60% at 4 years. Although alcohol is considered a direct hepatotoxin, only between 10 and 20% of alcoholics will develop alcoholic hepatitis. The explanation for this apparent paradox is unclear but involves the complex interaction of facilitating factors, such as intake frequency, diet, and gender.

Etiology and Pathogenesis

Quantity and duration of alcohol intake are the most important risk factors involved in the development of alcoholic liver disease **(Table 307-1)**. The roles of beverage type(s), i.e. wine, beer, or spirits, and pattern of drinking (daily versus binge drinking) are less clear. Progress of the hepatic injury beyond the fatty liver stage seems to require additional risk factors that remain incompletely defined. Although there are genetic predispositions for alcoholism **(Chapter 392)**, and candidate genes for liver steatosis and fibrosis, gender is a strong determinant for alcoholic liver disease. Women are more susceptible to alcoholic liver injury when compared to men. They develop advanced liver disease with substantially less alcohol intake. In general, the time it takes to develop liver disease is directly related to the amount of alcohol consumed. It is useful in estimating alcohol consumption to understand that one beer, four ounces of wine, or one ounce of 80% spirits all contain ~12 g of alcohol. The threshold for developing alcoholic liver disease in men is an intake of >60–80 g/d of alcohol for 10 years, while women are at increased risk for developing similar degrees of liver injury by consuming 20–40 g/d. Ingestion of 160 g/d is associated with a 25-fold increased risk of developing alcoholic cirrhosis. Gender-dependent differences result from poorly understood effects of estrogen and the metabolism of alcohol. Diet, particularly an increase in liver injury from high fat or the protective effect of coffee, has been postulated to play a part in the development of the pathogenic process.

Table 307-1 Risk Factors for Alcoholic Liver Disease

Risk Factor	Comment
Quantity	In men, 40–80 g/d of ethanol produces fatty liver; 160 g/d for 10–20 years causes hepatitis or cirrhosis. Only 15% of alcoholics develop alcoholic liver disease.
Gender	Women exhibit increased susceptibility to alcoholic liver disease at amounts >20 g/d; two drinks per day is probably safe.
Hepatitis C	HCV infection concurrent with alcoholic liver disease is associated with younger age for severity, more advanced histology, decreased survival.
Genetics	Genepolymorphisms may include alcohol dehydrogenase, cytochrome P4502E1, and those associated with alcoholism (twin studies).
Malnutrition	Alcohol injury does not require malnutrition, but obesity and fatty liver from the effect of carbohydrate on the transcriptional control of lipid synthesis and transport may be factors. Patients should receive vigorous attention to nutritional support.

Chronic infection with hepatitis C (HCV) **(Chap. 306)** is an important comorbidity in the progression of alcoholic liver disease to cirrhosis in chronic and excessive drinkers. Even moderate alcohol intake of 20–50 g/d increases the risk of cirrhosis and hepatocellular cancer in HCV-infected individuals. Patients with both alcoholic liver injury and HCV infection develop decompensated liver disease at a younger age and have poorer overall survival. Increased liver iron stores and, rarely, porphyria cutanea tarda can occur as a consequence of the overlapping injurious processes secondary to

alcohol abuse and HCV infection. In addition, alcohol intake of >50 g/d by HCV-infected patients decreases the efficacy of interferon-based antiviral therapy.

Our understanding of the pathogenesis of alcoholic liver injury is incomplete. Alcohol is a direct hepatotoxin, but ingestion of alcohol initiates a variety of metabolic responses that influence the final hepatotoxic response. The initial concept of malnutrition as the major pathogenic mechanism has been replaced by the understanding that the hepatic metabolism of alcohol initiates a pathogenic process including the production of toxic protein-aldehyde adducts, the generation of reducing equivalents that promotes lipogenesis, and the inhibition of fatty-acid oxidation. Endotoxins, oxidative stress, immunologic activity, and pro-inflammatory cytokine release contribute to the resulting liver injury **(Fig. 307-1)**. The complex interaction of intestinal and hepatic cells is crucial to alcohol-mediated liver injury. Tumor necrosis factor α (TNF-α) and intestine-derived endotoxemia facilitate hepatocyte apoptosis and necrosis. Stellate cell activation and collagen production are key events in hepatic fibrogenesis. The resulting fibrosis determines the architectural derangement of the liver following chronic alcohol ingestion.

Figure 307-1

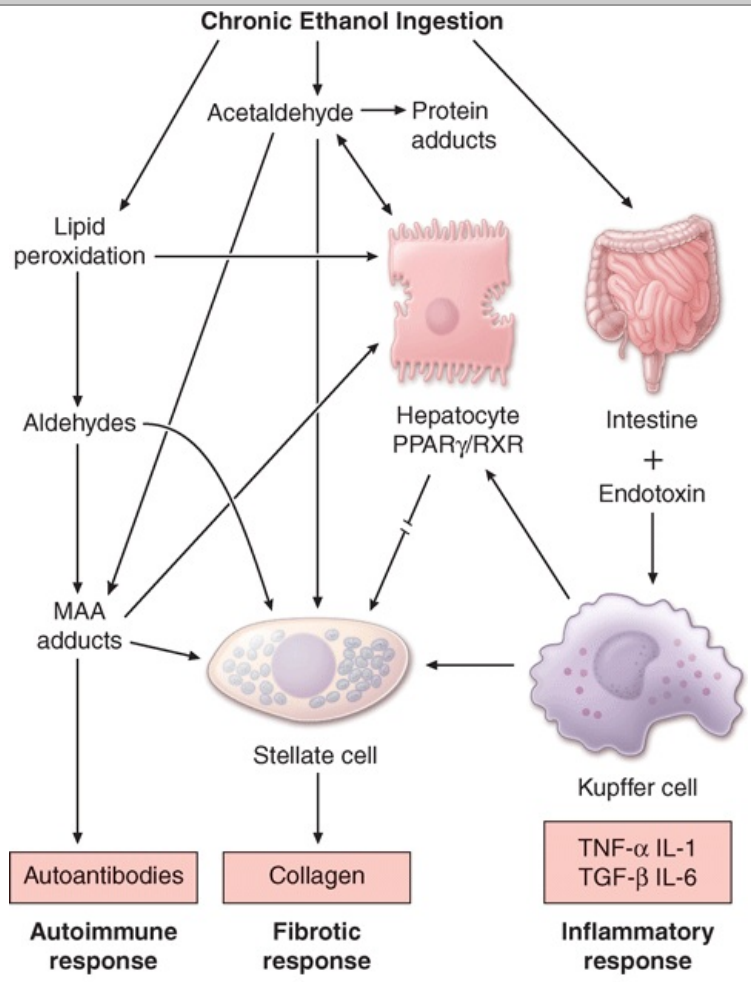

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Biomedical and cellular pathogenesis of liver injury secondary to chronic ethanol ingestion. MAA, malondialdehyde-acetaldehyde; TNF, tumor necrosis factor; TGF, transforming growth factor; IL, interleukin; PPAR, peroxisome proliferator-activated receptor; RXR, retinoid X receptor.

Pathology

The liver has a limited repertoire in response to injury. Fatty liver is the initial and most common histologic response to hepatotoxic stimuli, including excessive alcohol ingestion. The accumulation of fat within the perivenular hepatocytes coincides with the location of alcohol dehydrogenase, the major enzyme responsible for alcohol metabolism. Continuing alcohol ingestion results in fat accumulation throughout the entire hepatic lobule. Despite extensive fatty change and distortion of the hepatocytes with macrovesicular fat, the cessation of drinking results in normalization of hepatic architecture and fat content within the liver. Alcoholic fatty liver has traditionally been regarded as entirely benign, but similar to the spectrum of nonalcoholic fatty-liver disease **(Chap. 309)**, the appearance of

steatohepatitis and certain pathologic features such as giant mitochondria, perivenular fibrosis, and macrovesicular fat may be associated with progressive liver injury.

The transition between fatty liver and the development of alcoholic hepatitis is blurred. The hallmark of alcoholic hepatitis is hepatocyte injury characterized by ballooning degeneration, spotty necrosis, polymorphonuclear infiltrate, and fibrosis in the perivenular and perisinusoidal space of Disse. Mallory bodies are often present in florid cases but are neither specific nor necessary to establishing the diagnosis. Alcoholic hepatitis is thought to be a precursor to the development of cirrhosis. However, like fatty liver, it is potentially reversible with cessation of drinking. Cirrhosis is present in up to 50% of patients with biopsy-proven alcoholic hepatitis and its regression is uncertain, even with abstention.

Clinical Features

The clinical manifestations of alcoholic fatty liver are subtle and characteristically detected as a consequence of the patient's visit for a seemingly unrelated matter. Previously unsuspected hepatomegaly is often the only clinical finding. Occasionally, patients with fatty liver will present with right upper quadrant discomfort, nausea, and, rarely, jaundice. Differentiation of alcoholic fatty liver from nonalcoholic fatty liver is difficult unless an accurate drinking history is ascertained. In every instance where liver disease is present, a thoughtful and sensitive drinking history should be obtained. Standard, validated questions accurately detect alcohol-related problems **(Chap. 392)**. Alcoholic hepatitis is associated with a wide gamut of clinical features. Fever, spider nevi, jaundice, and abdominal pain simulating an acute abdomen represent the extreme end of the spectrum, while many patients will be entirely asymptomatic. Portal hypertension, ascites, or variceal bleeding can occur in the absence of cirrhosis. Recognition of the clinical features of alcoholic hepatitis is central to the initiation of an effective and appropriate diagnostic and therapeutic strategy. It is important to recognize that patients with alcoholic cirrhosis often exhibit clinical features identical to other causes of cirrhosis.

Laboratory Features

Patients with alcoholic liver disease are often identified through routine screening tests. The typical laboratory abnormalities seen in fatty liver are nonspecific and include modest elevations of the aspartate aminotransferase (AST), alanine aminotransferase (ALT), and γ-glutamyl transpeptidase (GGTP), accompanied by hypertriglyceridemia, hypercholesterolemia, and, occasionally, hyperbilirubinemia. In alcoholic hepatitis and in contrast to other causes of fatty liver, the AST and ALT are usually elevated two- to sevenfold. They are rarely >400 IU, and the AST/ALT ratio >1 **(Table 307-2)**. Hyperbilirubinemia is common and is accompanied by modest increases in the alkaline phosphatase level. Derangement in hepatocyte synthetic function indicates more serious disease. Hypoalbuminemia and coagulopathy are common in advanced liver injury. Ultrasonography is useful in detecting fatty infiltration of the liver and determining liver size. The demonstration by ultrasound of portal vein flow reversal, ascites, and intraabdominal collaterals indicates serious liver injury with less potential for complete reversal of liver disease.

Table 307-2 Laboratory Diagnosis of Alcoholic Fatty Liver and Alcoholic Hepatitis

Test	Comment
AST	Increased two- to sevenfold, <400 U/L, greater than ALT
ALT	Increased two- to sevenfold, <400 U/L
AST/ALT	Usually >1
GGTP	Not specific to alcohol, easily inducible, elevated in all forms of fatty liver
Bilirubin	May be markedly increased in alcoholic hepatitis despite modest elevation in alkaline phosphatase
PMN	If >5500/μL, predicts severe alcoholic hepatitis when discriminant function >32

Note: AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGTP, gamma-glutamyl transpeptidase; PMN, polymorphonuclear cells.

Prognosis

Critically ill patients with alcoholic hepatitis have short-term (30-day) mortality rates >50%. Severe alcoholic hepatitis is heralded by coagulopathy (prothrombin time increased >5 s), anemia, serum albumin concentrations <25 g/L (2.5 mg/dL), serum bilirubin levels >137 μmol/L (8 mg/dL), renal failure, and ascites. A discriminant function calculated as 4.6 X [the prolongation of the prothrombin time above control (seconds)] + serum bilirubin (mg/dL) can identify patients with a poor prognosis (discriminant function >32). A Model for End-Stage Liver Disease score **(MELD, Chap. 310)** ≥21 also is associated with significant mortality in alcoholic hepatitis. The presence of ascites, variceal hemorrhage, deep encephalopathy, or hepatorenal syndrome predicts a dismal prognosis. The

pathologic stage of the injury can be helpful in predicting prognosis. Liver biopsy should be performed whenever possible to confirm the diagnosis, to establish potential reversibility of the liver disease, and to guide the therapeutic decisions.

Treatment: Alcoholic Liver Disease

Complete abstinence from alcohol is the cornerstone in the treatment of alcoholic liver disease. Improved survival and the potential for reversal of histologic injury regardless of the initial clinical presentation are associated with total avoidance of alcohol ingestion. Referral of patients to experienced alcohol counselors and/or alcohol treatment programs should be routine in the management of patients with alcoholic liver disease. Attention should be directed to the nutritional and psychosocial states during the evaluation and treatment periods. Because of data suggesting that the pathogenic mechanisms in alcoholic hepatitis involve cytokine release and the perpetuation of injury by immunologic processes, glucocorticoids have been extensively evaluated in the treatment of alcoholic hepatitis. Patients with severe alcoholic hepatitis, defined as a discriminant function >32 or MELD >20, should be given prednisone, 40 mg/d, or prednisolone, 32 mg/d, for 4 weeks, followed by a steroid taper **(Fig. 307-2)**. Exclusion criteria include active gastrointestinal bleeding, renal failure, or pancreatitis. Women with encephalopathy from severe alcoholic hepatitis may be particularly good candidates for glucocorticoids. A Lille score >0.45, at http://www.lillemodel.com, uses pretreatment variables plus the change in total bilirubin at day seven of glucocorticoids to identify patients unresponsive to therapy.

Figure 307-2

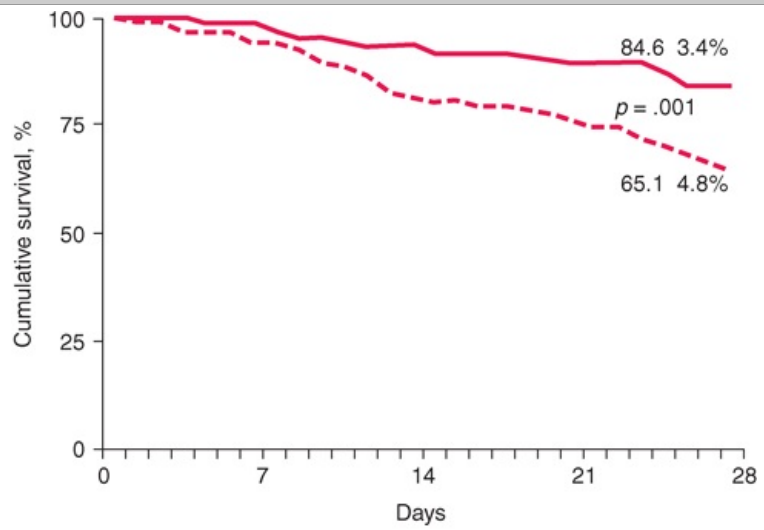

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effect of glucocorticoid therapy of severe alcoholic hepatitis on short-term survival: the result of a meta-analysis of individual data from three studies. Prednisolone, solid line; placebo, dotted line. *(Adapted from Mathurin et al., with permission from Elsevier Science.)*

The role of TNF-α expression and receptor activity in alcoholic liver injury has led to an examination of TNF inhibition as an alternative to glucocorticoids for severe alcoholic hepatitis. The nonspecific TNF inhibitor, pentoxifylline, demonstrated improved survival in the therapy of severe alcoholic hepatitis **(Fig. 307-3)**. Monoclonal antibodies that neutralize serum TNF-α should not be used in alcoholic hepatitis because of recent studies reporting increased deaths secondary to infection and renal failure. Because of inordinate surgical mortality and the high rates of recidivism following transplantation, patients with alcoholic hepatitis are not candidates for immediate liver transplantation. The transplant candidacy of these patients should be reevaluated after a defined period of sobriety.

Figure 307-3

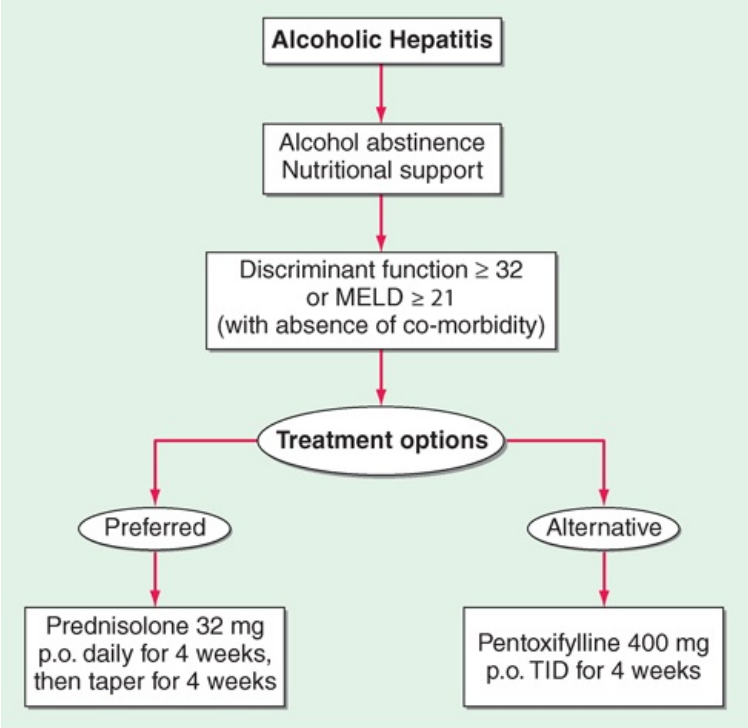

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment algorithm for alcoholic hepatitis. As identified by a calculated discriminant function >32 (see text), patients with severe alcoholic hepatitis, without the presence of gastrointestinal bleeding or infection, would be candidates for either glucocorticoids or pentoxifylline administration.

FURTHER READINGS

Akriviadis E et al: Pentoxifylline improves short-term survival in severe acute alcoholic hepatitis: A double blind placebo controlled trial. Gastroenterology 119:1637, 2000[PMID: 11113085] [Full Text]

Luvet A et al: The Lille model: A new tool for therapeutic strategy in patients with severe alcoholic hepatitis treated with steroids. Hepatology 45: 1348, 2007

Mathurin P et al: Corticosteroids improve short-term survival in patients with severe alcoholic hepatitis (AH): Individual data analysis of the last three randomized placebo controlled double blind trials of corticosteroids in severe AH. J Hepatol 36:480, 2002[PMID: 11943418] [Full Text]

Zakhari S, LI Ting-Kai: Determinants of alcohol use and abuse: Impact of quantity and frequency patterns on liver disease. Hepatology 46: 2032, 2007[PMID: 18046720] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CIRRHOSIS AND ITS COMPLICATIONS: INTRODUCTION

Cirrhosis is a condition that is defined histopathologically and has a variety of clinical manifestations and complications, some of which can be life-threatening. In the past, it has been thought that cirrhosis was never reversible; however, it has become apparent that when the underlying insult that has caused the cirrhosis has been removed, there can be reversal of fibrosis. This is most apparent with the successful treatment of chronic hepatitis C; however, reversal of fibrosis is also seen in patients with hemochromatosis who have been successfully treated and in patients with alcoholic liver disease who have discontinued alcohol use.

Regardless of the cause of cirrhosis, the pathologic features consist of the development of fibrosis to the point that there is architectural distortion with the formation of regenerative nodules. This results in a decrease in hepatocellular mass, and thus function, and an alteration of blood flow. The induction of fibrosis occurs with activation of hepatic stellate cells, resulting in the formation of increased amounts of collagen and other components of the extracellular matrix.

Clinical features of cirrhosis are the result of pathologic changes and mirror the severity of the liver disease. Most hepatic pathologists provide an assessment of grading and staging when evaluating liver biopsy samples. These grading and staging schemes vary between disease states and have been developed for most conditions, including chronic viral hepatitis, nonalcoholic fatty liver disease, and primary biliary cirrhosis. Advanced fibrosis usually includes bridging fibrosis with nodularity designated as stage 3 and cirrhosis designated as stage 4. Patients who have cirrhosis have varying degrees of compensated liver function, and clinicians need to differentiate between those who have stable, compensated cirrhosis and those who have decompensated cirrhosis. Patients who have developed complications of their liver disease and have become decompensated should be considered for liver transplantation. Many of the complications of cirrhosis will require specific therapy. *Portal hypertension* is a significant complicating feature of decompensated cirrhosis and is responsible for the development of ascites and bleeding from esophagogastric varices, two complications that signify decompensated cirrhosis. Loss of hepatocellular function results in jaundice, coagulation disorders, and hypoalbuminemia and contributes to the causes of portosystemic encephalopathy. The complications of cirrhosis are basically the same regardless of the etiology. Nonetheless, it is useful to classify patients by the cause of their liver disease **(Table 308-1)**; patients can be divided into broad groups with alcoholic cirrhosis, cirrhosis due to chronic viral hepatitis, biliary cirrhosis, and other, less-common causes such as cardiac cirrhosis, cryptogenic cirrhosis, and other miscellaneous causes.

Table 308-1 Causes of Cirrhosis

Alcoholism	Chronic viral hepatitis
Cardiac cirrhosis	Inherited metabolic liver disease
Hepatitis B	Hemochromatosis
Hepatitis C	Wilson's disease
Autoimmune hepatitis	α_1 Antitrypsin deficiency
Nonalcoholic steatohepatitis	Cystic fibrosis
Biliary cirrhosis	Cryptogenic cirrhosis
Primary biliary cirrhosis	
Primary sclerosing cholangitis	
Autoimmune cholangiopathy	

ALCOHOLIC CIRRHOSIS

Excessive chronic alcohol use can cause several different types of chronic liver disease, including alcoholic fatty liver, alcoholic hepatitis, and alcoholic cirrhosis. Furthermore, use of excessive alcohol can contribute to liver damage in patients with other liver diseases, such as hepatitis C, hemochromatosis, and those patients who have fatty liver disease related to obesity. Chronic alcohol use can produce fibrosis in the absence of accompanying inflammation and/or necrosis. Fibrosis can

be centrilobular, pericellular, or periportal. When fibrosis reaches a certain degree, there is disruption of the normal liver architecture and replacement of liver cells by regenerative nodules. In alcoholic cirrhosis, the nodules are usually <3 mm in diameter; this form of cirrhosis is referred to as *micronodular*. With cessation of alcohol use, larger nodules may form, resulting in a mixed micronodular and macronodular cirrhosis.

PATHOGENESIS

Alcohol is the most commonly used drug in the United States, and more than two-thirds of adults drink alcohol each year. Thirty percent have had a binge within the past month, and over 7% of adults regularly consume more than two drinks per day. Unfortunately, more than 14 million adults in the United States meet the diagnostic criteria for alcohol abuse or dependence. In the United States, chronic liver disease is the tenth most common cause of death in adults, and alcoholic cirrhosis accounts for approximately 40% of deaths due to cirrhosis.

Ethanol is mainly absorbed by the small intestine and, to a lesser degree, through the stomach. Gastric alcohol dehydrogenase (ADH) initiates alcohol metabolism. Three enzyme systems account for metabolism of alcohol in the liver. These include cytosolic ADH, the microsomal ethanol oxidizing system (MEOS), and peroxisomal catalase. The majority of ethanol oxidation occurs via ADH to form acetaldehyde, which is a highly reactive molecule that may have multiple effects. Ultimately, acetaldehyde is metabolized to acetate by aldehyde dehydrogenase (ALDH). Intake of ethanol increases intracellular accumulation of triglycerides by increasing fatty acid uptake and by reducing fatty acid oxidation and lipoprotein secretion. Protein synthesis, glycosylation, and secretion are impaired. Oxidative damage to hepatocyte membranes occurs due to the formation of reactive oxygen species; acetaldehyde is a highly reactive molecule that combines with proteins to form protein-acetaldehyde adducts. These adducts may interfere with specific enzyme activities, including microtubular formation and hepatic protein trafficking. With acetaldehyde-mediated hepatocyte damage, certain reactive oxygen species can result in Kupffer cell activation. As a result, profibrogenic cytokines are produced that initiate and perpetuate stellate cell activation, with the resultant production of excess collagen and extracellular matrix. Connective tissue appears in both periportal and pericentral zones and eventually connects portal triads with central veins forming regenerative nodules. Hepatocyte loss occurs, and with increased collagen production and deposition, together with continuing hepatocyte destruction, the liver contracts and shrinks in size. This process generally takes from years to decades to occur and requires repeated insults.

CLINICAL FEATURES

The diagnosis of alcoholic liver disease requires an accurate history regarding both amount and duration of alcohol consumption. Patients with alcoholic liver disease can present with nonspecific symptoms such as vague right upper quadrant pain, fever, nausea and vomiting, diarrhea, anorexia, and malaise. Alternatively, they may present with more specific complications of chronic liver disease, including ascites, edema, or upper gastrointestinal (GI) hemorrhage. Many cases present incidentally at the time of autopsy or elective surgery. Other clinical manifestations include the development of jaundice or encephalopathy. The abrupt onset of any of these complications may be the first event prompting the patient to seek medical attention. Other patients may be identified in the course of an evaluation of routine laboratory studies that are found to be abnormal. On physical examination, the liver and spleen may be enlarged, with the liver edge being firm and nodular. Other frequent findings include scleral icterus, palmar erythema **(Fig. 308-1)**, spider angiomas **(Fig. 308-2)**, parotid gland enlargement, digital clubbing, muscle wasting, or the development of edema and ascites. Men may have decreased body hair and gynecomastia as well as testicular atrophy, which may be a consequence of hormonal abnormalities or a direct toxic effect of alcohol on the testes. In women with advanced alcoholic cirrhosis, menstrual irregularities usually occur, and some women may be amenorrheic. These changes are often reversible following cessation of alcohol.

Figure 308-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Palmar erythema. This figure shows palmar erythema in a patient with alcoholic cirrhosis. The erythema is peripheral over the palm with central pallor.

Figure 308-2

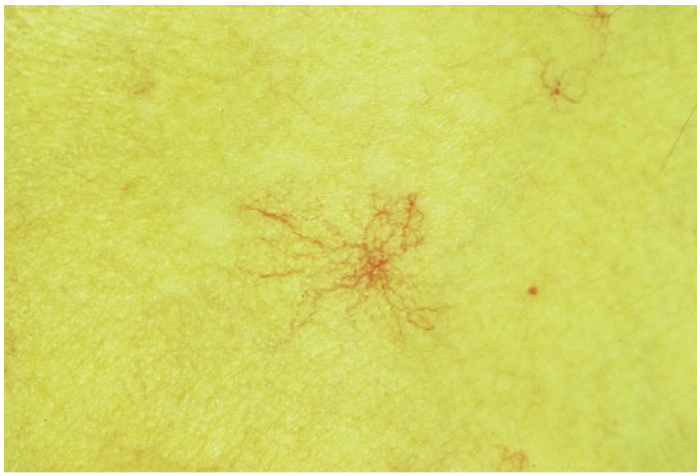

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spider angioma. This figure shows a spider angioma in a patient with hepatitis C cirrhosis. With release of central compression, the arteriole fills from the center and spreads out peripherally.

Laboratory tests may be completely normal in patients with early compensated alcoholic cirrhosis. Alternatively, in advanced liver disease, many abnormalities usually are present. Patients may be anemic either from chronic GI blood loss, nutritional deficiencies, or hypersplenism related to portal hypertension, or as a direct suppressive effect of alcohol on the bone marrow. A unique form of hemolytic anemia (with spur cells and acanthocytes) called *Zieve's syndrome* can occur in patients with severe alcoholic hepatitis. Platelet counts are often reduced early in the disease, reflective of portal hypertension with hypersplenism. Serum total bilirubin can be normal or elevated with advanced disease. Direct bilirubin is frequently mildly elevated in patients with a normal total bilirubin, but the abnormality typically progresses as the disease worsens. Prothrombin times are often prolonged and usually do not respond to administration of parenteral vitamin K. Serum sodium levels are usually normal unless patients have ascites and then can be depressed, largely due to ingestion of excess free water. Serum alanine and aspartate aminotransferases (ALT, AST) are typically elevated, particularly in patients who continue to drink, with AST levels being higher than ALT levels, usually by a 2:1 ratio.

DIAGNOSIS

Patients who have any of the above-mentioned clinical features, physical examination findings, or laboratory studies should be considered to have alcoholic liver disease. The diagnosis, however, requires accurate knowledge that the patient is continuing to use and abuse alcohol. Furthermore, other forms of chronic liver disease (e.g., chronic viral hepatitis or metabolic or autoimmune liver diseases) must be considered or ruled out or, if present, an estimate of relative causality along with the

alcohol use should be determined. Liver biopsy can be helpful to confirm a diagnosis, but generally when patients present with alcoholic hepatitis and are still drinking, liver biopsy is withheld until abstinence has been maintained for at least 6 months to determine residual, nonreversible disease.

In patients who have had complications of cirrhosis and who continue to drink, there is a <50% 5-year survival. In contrast, in those patients who are able to remain abstinent, the prognosis is significantly improved. In patients with advanced liver disease, the prognosis remains poor; however, in those individuals who are able to remain abstinent, liver transplantation is a viable option.

Treatment: Alcoholic Cirrhosis

Abstinence is the cornerstone of therapy for patients with alcoholic liver disease. In addition, patients require good nutrition and long-term medical supervision to manage underlying complications that may develop. Complications such as the development of ascites and edema, variceal hemorrhage, or portosystemic encephalopathy all require specific management and treatment. Glucocorticoids are occasionally used in patients with severe alcoholic hepatitis in the absence of infection. Survival has been shown to improve in certain studies. Treatment is restricted to patients with a discriminant function (DF) value of >32. The DF is calculated as the serum total bilirubin plus the difference in the patient's prothrombin time compared to control (in seconds) multiplied by 4.6. In patients for whom this value is >32, there is improved survival at 28 days with the use of glucocorticoids.

Other therapies that have been used include oral pentoxifylline, which decreases the production of tumor necrosis factor ∝ (TNF-∝) and other proinflammatory cytokines. In contrast to glucocorticoids, with which complications can occur, pentoxifylline is relatively easy to administer and has few if any side effects. A variety of nutritional therapies have been tried with either parenteral or enteral feedings; however, it is unclear whether any of these modalities have significantly improved survival.

Recent studies have used parenterally administered inhibitors of TNF-∝ such as infliximab or etanercept. Early results have shown no adverse events; however, there was no clear-cut improvement in survival. Anabolic steroids, propylthiouracil, antioxidants, colchicine, and penicillamine have all been used but do not show clear-cut benefits and are not recommended.

As mentioned above, the cornerstone to treatment is cessation of alcohol use. Recent experience with medications that reduce craving for alcohol such as acamprosate calcium has been favorable. Patients may take other necessary medications even in the presence of cirrhosis. Acetaminophen use is often discouraged in patients with liver disease; however, if no more than 2 g of acetaminophen per day are consumed, there generally are no problems.

CIRRHOSIS DUE TO CHRONIC VIRAL HEPATITIS B OR C

Of patients exposed to the hepatitis C virus (HCV), approximately 80% develop chronic hepatitis C, and of those, about 20–30% will develop cirrhosis over 20–30 years. Many of these patients have had concomitant alcohol use, and the true incidence of cirrhosis due to hepatitis C alone is unknown. Nonetheless, this represents a significant number of patients. It is expected that an even higher percentage will go on to develop cirrhosis over longer periods of time. In the United States, approximately 5 million people have been exposed to the hepatitis C virus, with about 3.5 to 4 million who are chronically viremic. Worldwide, about 170 million individuals have hepatitis C, with some areas of the world (e.g., Egypt) having up to 15% of the population infected. HCV is a noncytopathic virus, and liver damage is probably immune-mediated. Progression of liver disease due to chronic hepatitis C is characterized by portal-based fibrosis with bridging fibrosis and nodularity developing, ultimately culminating in the development of cirrhosis. In cirrhosis due to chronic hepatitis C, the liver is small and shrunken with characteristic features of a mixed micro- and macronodular cirrhosis seen on liver biopsy. In addition to the increased fibrosis that is seen in cirrhosis due to hepatitis C, an inflammatory infiltrate is found in portal areas with interface hepatitis and occasionally some lobular hepatocellular injury and inflammation. In patients with HCV genotype 3, steatosis is often present.

Similar findings are seen in patients with cirrhosis due to chronic hepatitis B. adult Of patients exposed to hepatitis B, about 5% develop chronic hepatitis B, and about 20% of those patients will go on to develop cirrhosis. Special stains for hepatitis B core (HBc) and hepatitis B surface (HBs) antigen will be positive, and ground-glass hepatocytes signifying hepatitis B surface antigen (HBsAg) may be present. In the United States, there are about 2 million carriers of hepatitis B, whereas in other parts of the world where hepatitis B virus (HBV) is endemic (i.e., Asia, Southeast Asia, sub-Saharan Africa), up to 15% of the population may be infected having acquired the infection vertically at the time of birth. Thus, over 300–400 million individuals are thought to have hepatitis B worldwide. Approximately 25% of these individuals may ultimately develop cirrhosis.

CLINICAL FEATURES AND DIAGNOSIS

Patients with cirrhosis due to either chronic hepatitis C or B can present with the usual symptoms and signs of chronic liver disease. Fatigue, malaise, vague right upper quadrant pain, and laboratory abnormalities are frequent presenting features. Diagnosis requires a thorough laboratory evaluation, including quantitative HCV RNA testing and analysis for HCV genotype, or hepatitis B serologies to include HBsAg, anti-HBs, HBeAg (hepatitis B e antigen), anti-HBe, and quantitative HBV DNA levels.

Treatment: Cirrhosis Due to Chronic Viral Hepatitis B or C

Management of complications of cirrhosis revolves around specific therapy for treatment of whatever complications occur, whether they be esophageal variceal hemorrhage, development of ascites and edema, or encephalopathy. In patients with chronic hepatitis B, numerous studies have shown beneficial effects of antiviral therapy, which is effective at viral suppression, as evidenced by reducing aminotransferase levels and HBV DNA levels, and improving histology by reducing inflammation and fibrosis. Several clinical trials and case series have demonstrated that patients with decompensated liver disease can become compensated with the use of antiviral therapy directed against hepatitis B. Currently available therapy includes lamivudine, adefovir, telbivudine, entecavir, and tenofovir. Interferon α can also be used for treating hepatitis B, but it should not be used in cirrhotics.

Treatment of patients with cirrhosis due to hepatitis C is a little more difficult because the side effects of pegylated interferon and ribavirin therapy are oftentimes difficult to manage. Dose-limiting cytopenias (platelets, white blood cells, red blood cells) or severe side effects can result in discontinuation of treatment. Nonetheless, if patients can tolerate treatment, and if it is successful, the benefit is great and disease progression is reduced.

CIRRHOSIS FROM AUTOIMMUNE HEPATITIS AND NONALCOHOLIC FATTY LIVER DISEASE

Other causes of posthepatitic cirrhosis include autoimmune hepatitis and cirrhosis due to nonalcoholic steatohepatitis. Many patients with autoimmune hepatitis (AIH) present with cirrhosis that is already established. Typically, these patients will not benefit from immunosuppressive therapy with glucocorticoids or azathioprine because the AIH is "burned out." In this situation, liver biopsy does not show a significant inflammatory infiltrate. Diagnosis in this setting requires positive autoimmune markers such as antinuclear antibody (ANA) or anti-smooth-muscle antibody (ASMA). When patients with AIH present with cirrhosis and active inflammation accompanied by elevated liver enzymes, there can be considerable benefit from the use of immunosuppressive therapy.

Patients with nonalcoholic steatohepatitis are increasingly being found to have progressed to cirrhosis. With the epidemic of obesity that continues in Western countries, more and more patients are identified with nonalcoholic fatty liver disease. Of these, a significant subset have nonalcoholic steatohepatitis and can progress to increased fibrosis and cirrhosis. Over the past several years, it has been increasingly recognized that many patients who were thought to have cryptogenic cirrhosis in fact have nonalcoholic steatohepatitis. As their cirrhosis progresses, they become catabolic and then lose the telltale signs of steatosis seen on biopsy. Management of complications of cirrhosis due to either AIH or nonalcoholic steatohepatitis is similar to that for other forms of cirrhosis.

BILIARY CIRRHOSIS

Biliary cirrhosis has pathologic features that are different from either alcoholic cirrhosis or posthepatitic cirrhosis, yet the manifestations of end-stage liver disease are the same. Cholestatic liver disease may result from necroinflammatory lesions, congenital or metabolic processes, or external bile duct compression. Thus, two broad categories reflect the anatomic sites of abnormal bile retention: *intrahepatic* and *extrahepatic*. The distinction is important for obvious therapeutic reasons. Extrahepatic obstruction may benefit from surgical or endoscopic biliary tract decompression, whereas intrahepatic cholestatic processes will not improve with such interventions and require a different approach.

The major causes of chronic cholestatic syndromes are primary biliary cirrhosis (PBC), autoimmune cholangitis (AIC), primary sclerosing cholangitis (PSC), and idiopathic adulthood ductopenia. These syndromes are usually clinically distinguished from each other by antibody testing, cholangiographic findings, and clinical presentation. However, they all share the histopathologic features of chronic cholestasis, such as cholate stasis; copper deposition; xanthomatous transformation of hepatocytes; and irregular, so-called biliary fibrosis. In addition, there may be chronic portal inflammation, interface activity, and chronic lobular inflammation. Ductopenia is a result of this progressive disease as patients develop cirrhosis.

Primary Biliary Cirrhosis

PBC is seen in about 100–200 individuals per million, with a strong female preponderance and a median age of around 50 years at the time of diagnosis. The cause of PBC is unknown; it is characterized by portal inflammation and necrosis of cholangiocytes in small- and medium-sized bile ducts. Cholestatic features prevail, and biliary cirrhosis is characterized by an elevated bilirubin level and progressive liver failure. Liver transplantation is the treatment of choice for patients with decompensated cirrhosis due to PBC. A variety of therapies have been proposed, but ursodeoxycholic acid (UDCA) is the only approved treatment that has some degree of efficacy by slowing the rate of progression of the disease.

Antimitochondrial antibodies (AMA) are present in about 90% of patients with PBC. These autoantibodies recognize intermitochondrial membrane proteins that are enzymes of the pyruvate dehydrogenase complex (PDC), the branched-chain 2-oxoacid dehydrogenase complex, and the 2-

oxogluterate dehydrogenase complex. Most relate to pyruvate dehydrogenase. These autoantibodies are not pathogenic but rather are useful markers for making a diagnosis of PBC.

PATHOLOGY

Histopathologic analyses of liver biopsies of patients with PBC have resulted in identifying four distinct stages of the disease as it progresses. The earliest lesion is termed *chronic nonsuppurative destructive cholangitis* and is a necrotizing inflammatory process of the portal tracts. Medium and small bile ducts are infiltrated with lymphocytes and undergo duct destruction. Mild fibrosis and sometimes bile stasis can occur. With progression, the inflammatory infiltrate becomes less prominent, but the number of bile ducts is reduced and there is proliferation of smaller bile ductules. Increased fibrosis ensues with the expansion of periportal fibrosis to bridging fibrosis. Finally, cirrhosis, which may be micronodular or macronodular, develops.

CLINICAL FEATURES

Currently, most patients with PBC are diagnosed well before the end-stage manifestations of the disease are present, and, as such, most patients are actually asymptomatic. When symptoms are present, they most prominently include a significant degree of fatigue out of proportion to what would be expected for either the severity of the liver disease or the age of the patient. Pruritus is seen in approximately 50% of patients at the time of diagnosis, and it can be debilitating. It might be intermittent and usually is most bothersome in the evening. In some patients, pruritus can develop toward the end of pregnancy, and there are examples of patients having been diagnosed with cholestasis of pregnancy rather than PBC. Pruritus that presents prior to the development of jaundice indicates severe disease and a poor prognosis.

Physical examination can show jaundice and other complications of chronic liver disease, including hepatomegaly, splenomegaly, ascites, and edema. Other features that are unique to PBC include hyperpigmentation, xanthelasma, and xanthomata, which are related to the altered cholesterol metabolism seen in this disease. Hyperpigmentation is evident on the trunk and the arms and is seen in areas of exfoliation and lichenification associated with progressive scratching related to the pruritus. Bone pain resulting from osteopenia or osteoporosis is occasionally seen at the time of diagnosis.

LABORATORY FINDINGS

Laboratory findings in PBC show cholestatic liver enzyme abnormalities with an elevation in γ-glutamyl transpeptidase and alkaline phosphatase (ALP) along with mild elevations in aminotransferases (ALT and AST). Immunoglobulins, particularly IgM, are typically increased. Hyperbilirubinemia usually is seen once cirrhosis has developed. Thrombocytopenia, leukopenia, and anemia may be seen in patients with portal hypertension and hypersplenism. Liver biopsy shows characteristic features as described above and should be evident to any experienced hepatopathologist. Up to 10% of patients with characteristic PBC will have features of AIH as well and are defined as having "overlap" syndrome. These patients are treated as PBC patients and may progress to cirrhosis with the same frequency as typical PBC patients.

DIAGNOSIS

PBC should be considered in patients with chronic cholestatic liver enzyme abnormalities. It is most often seen in middle-aged women. AMA testing may be negative, and it should be remembered that as many as 10% of patients with PBC may be AMA-negative. Liver biopsy is most important in this setting of AMA-negative PBC. In patients who are AMA-negative with cholestatic liver enzymes, PSC should be ruled out by way of cholangiography.

TREATMENT: PRIMARY BILIARY CIRRHOSIS

Treatment of the typical manifestations of cirrhosis are no different for PBC than for other forms of cirrhosis. UDCA has been shown to improve both biochemical and histologic features of the disease. Improvement is greatest when therapy is initiated early; the likelihood of significant improvement with UDCA is low in patients with PBC who present with manifestations of cirrhosis. UDCA is given in doses of 13–15 mg/kg per day; the medication is usually well-tolerated, although some patients have worsening pruritus with initiation of therapy. A small proportion of patients may have diarrhea or headache as a side effect of the drug. UDCA has been shown to slow the rate of progression of PBC, but it does not reverse or cure the disease. Patients with PBC require long-term follow-up by a physician experienced with the disease. Certain patients may need to be considered for liver transplantation should their liver disease decompensate.

The main symptoms of PBC are fatigue and pruritus, and symptom management is important. Several therapies have been tried for treatment of fatigue, but none of them have been successful; frequent naps should be encouraged. Pruritus is treated with antihistamines, narcotic receptor antagonists (naltrexone), and rifampin. Cholestyramine, a bile salt–sequestering agent, has been helpful in some patients but is somewhat tedious and difficult to take. Plasmapheresis has been used rarely in patients with severe intractable pruritus. There is an increased incidence of osteopenia and osteoporosis in patients with cholestatic liver disease, and bone density testing should be performed. Treatment with a bisphosphonate should be instituted when bone disease is identified.

Primary Sclerosing Cholangitis

As in PBC, the cause of PSC remains unknown. PSC is a chronic cholestatic syndrome that is characterized by diffuse inflammation and fibrosis involving the entire biliary tree, resulting in chronic cholestasis. This pathologic process ultimately results in obliteration of both the intra- and extrahepatic biliary tree, leading to biliary cirrhosis, portal hypertension, and liver failure. The cause of PSC remains unknown despite extensive investigation into various mechanisms related to bacterial and viral infections, toxins, genetic predisposition, and immunologic mechanisms, all of which have been postulated to contribute to the pathogenesis and progression of this syndrome.

Pathologic changes that can occur in PSC show bile duct proliferation as well as ductopenia and fibrous cholangitis (pericholangitis). Often, liver biopsy changes in PSC are not pathognomonic, and establishing the diagnosis of PSC must involve imaging of the biliary tree. Periductal fibrosis is occasionally seen on biopsy specimens and can be quite helpful in making the diagnosis. As the disease progresses, biliary cirrhosis is the final, end-stage manifestation of PSC.

CLINICAL FEATURES

The usual clinical features of PSC are those found in cholestatic liver disease, with fatigue, pruritus, steatorrhea, deficiencies of fat-soluble vitamins, and the associated consequences. As in PBC, the fatigue is profound and nonspecific. Pruritus can often be debilitating and is related to the cholestasis. The severity of pruritus does not correlate with the severity of the disease. Metabolic bone disease, as seen in PBC, can occur with PSC and should be treated (see above).

LABORATORY FINDINGS

Patients with PSC typically are identified in the course of an evaluation of abnormal liver enzymes. Most patients have at least a twofold increase in ALP and may have elevated aminotransferases as well. Albumin levels may be decreased, and prothrombin times are prolonged in a substantial proportion of patients at the time of diagnosis. Some degree of correction of a prolonged prothrombin time may occur with parenteral vitamin K. A small subset of patients have aminotransferase elevations greater than five times the upper limit of normal and may have features of AIH on biopsy. These individuals are thought to have an overlap syndrome between PSC and AIH. Autoantibodies are frequently positive in patients with the overlap syndrome but are typically negative in patients who only have PSC. One autoantibody, the perinuclear antineutrophil cytoplasmic antibody (p-ANCA), is positive in about 65% of patients with PSC. Over 50% of patients with PSC also have ulcerative colitis (UC); accordingly, once a diagnosis of PSC is established, colonoscopy should be performed to look for evidence of UC.

DIAGNOSIS

The definitive diagnosis of PSC requires cholangiographic imaging. Over the last several years, MRI with magnetic resonance cholangiopancreatography (MRCP) has been used as the imaging technique of choice for initial evaluation. Once patients are screened in this manner, some investigators feel that endoscopic retrograde cholangiopancreatography (ERCP) should also be performed to be certain whether or not a dominant stricture is present. Typical cholangiographic findings in PSC are multifocal stricturing and beading involving both the intrahepatic and extrahepatic biliary tree. However, though involvement may be of the intrahepatic bile ducts alone or of the extrahepatic bile ducts alone, more commonly, both are involved. These strictures are typically short and with intervening segments of normal or slightly dilated bile ducts that are distributed diffusely, producing the classic beaded appearance. The gallbladder and cystic duct can be involved in up to 15% of cases. Patients with high-grade, diffuse stricturing of the intrahepatic bile ducts have an overall poor prognosis. Gradually, biliary cirrhosis develops, and patients will progress to decompensated liver disease with all the manifestations of ascites, esophageal variceal hemorrhage, and encephalopathy.

TREATMENT: PRIMARY SCLEROSING CHOLANGITIS

There is no specific proven treatment for PSC, although studies are currently ongoing using high-dose (20 mg/kg per day) UDCA to determine its benefit. Endoscopic dilatation of dominant strictures can be helpful, but the ultimate treatment is liver transplantation. A dreaded complication of PSC is the development of cholangiocarcinoma, which is a relative contraindication to liver transplantation. Symptoms of pruritus are common, and the approach is as mentioned previously for this problem in patients with PBC (see above).

CARDIAC CIRRHOSIS

DEFINITION

Patients with long-standing right-sided congestive heart failure may develop chronic liver injury and cardiac cirrhosis. This is an increasingly uncommon, if not rare, cause of chronic liver disease given the advances made in the care of patients with heart failure.

ETIOLOGY AND PATHOLOGY

In the case of long-term right-sided heart failure, there is an elevated venous pressure transmitted via the inferior vena cava and hepatic veins to the sinusoids of the liver, which become dilated and

engorged with blood. The liver becomes enlarged and swollen, and with long-term passive congestion and relative ischemia due to poor circulation, centrilobular hepatocytes can become necrotic, leading to pericentral fibrosis. This fibrotic pattern can extend to the periphery of the lobule outward until a unique pattern of fibrosis causing cirrhosis can occur.

CLINICAL FEATURES

Patients typically have signs of congestive heart failure and will manifest an enlarged firm liver on physical examination. ALP levels are characteristically elevated, and aminotransferases may be normal or slightly increased with AST usually higher than ALT. It is unlikely that patients will develop variceal hemorrhage or encephalopathy.

DIAGNOSIS

The diagnosis is usually made in someone with clear-cut cardiac disease who has an elevated ALP and an enlarged liver. Liver biopsy shows a pattern of fibrosis that can be recognized by an experienced hepatopathologist. Differentiation from Budd-Chiari syndrome (BCS) can be made by seeing extravasation of red blood cells in BCS, but not in cardiac hepatopathy. Venoocclusive disease can also affect hepatic outflow and has characteristic features on liver biopsy. Venoocclusive disease can be seen under the circumstances of conditioning for bone marrow transplant with radiation and chemotherapy; it can also be seen with the ingestion of certain herbal teas as well as pyrrolizidine alkaloids. This is typically seen in Caribbean countries and rarely in the United States. Treatment is based on management of the underlying cardiac disease.

OTHER TYPES OF CIRRHOSIS

There are several other less-common causes of chronic liver disease that can progress to cirrhosis. These include inherited metabolic liver diseases such as hemochromatosis, Wilson's disease, α_1 antitrypsin (α_1AT) deficiency, and cystic fibrosis. For all of these disorders, the manifestations of cirrhosis are similar, with some minor variations, to those seen in other patients with other causes of cirrhosis.

Hemochromatosis is an inherited disorder of iron metabolism that results in a progressive increase in hepatic iron deposition, which, over time, can lead to a portal-based fibrosis progressing to cirrhosis, liver failure, and hepatocellular cancer. While the frequency of hemochromatosis is relatively common, with genetic susceptibility occurring in 1 in 250 individuals, the frequency of end-stage manifestations due to the disease is relatively low, and fewer than 5% of those patients who are genotypically susceptible will go on to develop severe liver disease from hemochromatosis. Diagnosis is made with serum iron studies showing an elevated transferrin saturation and an elevated ferritin level, along with abnormalities identified by *HFE* mutation analysis. Treatment is straightforward, with regular therapeutic phlebotomy.

Wilson's disease is an inherited disorder of copper homeostasis with failure to excrete excess amounts of copper, leading to an accumulation in the liver. This disorder is relatively uncommon, affecting 1 in 30,000 individuals. Wilson's disease typically affects adolescents and young adults. Prompt diagnosis before end-stage manifestations become irreversible can lead to significant clinical improvement. Diagnosis requires determination of ceruloplasmin levels, which are low; 24-hour urine copper levels, which are elevated; typical physical examination findings, including Kayser-Fleischer corneal rings, and characteristic liver biopsy findings. Treatment consists of copper-chelating medications.

α_1AT deficiency results from an inherited disorder that causes abnormal folding of the α_1AT protein, resulting in failure of secretion of that protein from the liver. It is unknown how the retained protein leads to liver disease. Patients with α_1AT deficiency at greatest risk for developing chronic liver disease have the ZZ phenotype, but only about 10–20% of such individuals will develop chronic liver disease. Diagnosis is made by determining α_1AT levels and phenotype. Characteristic periodic acid–Schiff (PAS)-positive, diastase-resistant globules are seen on liver biopsy. The only effective treatment is liver transplantation, which is curative.

Cystic fibrosis is an uncommon inherited disorder affecting Caucasians of Northern European descent. A biliary-type cirrhosis can occur, and some patients derive benefit from the chronic use of UDCA.

MAJOR COMPLICATIONS OF CIRRHOSIS

The clinical course of patients with advanced cirrhosis is often complicated by a number of important sequelae that can occur regardless of the underlying cause of the liver disease. These include portal hypertension and its consequences of gastroesophageal variceal hemorrhage, splenomegaly, ascites, hepatic encephalopathy, spontaneous bacterial peritonitis (SBP), hepatorenal syndrome, and hepatocellular carcinoma **(Table 308-2)**.

Table 308-2 Complications of Cirrhosis

Portal hypertension	Coagulopathy
Gastroesophageal varices	Factor deficiency
Portal hypertensive gastropathy	Fibrinolysis
Splenomegaly, hypersplenism	Thrombocytopenia
Ascites	Bone disease
Spontaneous bacterial peritonitis	Osteopenia
Hepatorenal syndrome	Osteoporosis
Type 1	Osteomalacia
Type 2	Hematologic abnormalities
Hepatic encephalopathy	Anemia
Hepatopulmonary syndrome	Hemolysis
Portopulmonary hypertension	Thrombocytopenia
Malnutrition	Neutropenia

Portal Hypertension

Portal hypertension is defined as the elevation of the hepatic venous pressure gradient (HVPG) to >5 mmHg. Portal hypertension is caused by a combination of two simultaneously occurring hemodynamic processes: (1) increased intrahepatic resistance to the passage of blood flow through the liver due to cirrhosis and regenerative nodules, and (2) increased splanchnic blood flow secondary to vasodilation within the splanchnic vascular bed. Portal hypertension is directly responsible for the two major complications of cirrhosis: variceal hemorrhage and ascites. *Variceal hemorrhage* is an immediate life-threatening problem with a 20–30% mortality rate associated with each episode of bleeding. The portal venous system normally drains blood from the stomach, intestines, spleen, pancreas, and gallbladder, and the portal vein is formed by the confluence of the superior mesenteric and splenic veins. Deoxygenated blood from the small bowel drains into the superior mesenteric vein along with blood from the head of the pancreas, the ascending colon, and part of the transverse colon. Conversely, the splenic vein drains the spleen and the pancreas and is joined by the inferior mesenteric vein, which brings blood from the transverse and descending colon as well as from the superior two-thirds of the rectum. Thus, the portal vein normally receives blood from almost the entire GI tract.

The causes of portal hypertension are usually subcategorized as prehepatic, intrahepatic, and posthepatic **(Table 308-3)**. Prehepatic causes of portal hypertension are those affecting the portal venous system before it enters the liver; they include portal vein thrombosis and splenic vein thrombosis. Posthepatic causes encompass those affecting the hepatic veins and venous drainage to the heart; they include BCS, venoocclusive disease, and chronic right-sided cardiac congestion. Intrahepatic causes account for over 95% of cases of portal hypertension and are represented by the major forms of cirrhosis. Intrahepatic causes of portal hypertension can be further subdivided into presinusoidal, sinusoidal, and postsinusoidal causes. Postsinusoidal causes include venoocclusive disease, while presinusoidal causes include congenital hepatic fibrosis and schistosomiasis. Sinusoidal causes are related to cirrhosis from various causes.

Table 308-3 Classification of Portal Hypertension

Prehepatic
Portal vein thrombosis
Splenic vein thrombosis
Massive splenomegaly (Banti's syndrome)
Hepatic
Presinusoidal
emsp; Schistosomiasis
emsp; Congenital hepatic fibrosis
Sinusoidal
emsp; Cirrhosis—many causes
emsp; Alcoholic hepatitis
Postsinusoidal
Hepatic sinusoidal obstruction (venoocclusive syndrome)

Posthepatic
Budd-Chiari syndrome
Inferior vena caval webs
Cardiac causes
Restrictive cardiomyopathy
Constrictive pericarditis
Severe congestive heart failure

Cirrhosis is the most common cause of portal hypertension in the United States, and clinically significant portal hypertension is present in >60% of patients with cirrhosis. Portal vein obstruction may be idiopathic or can occur in association with cirrhosis or with infection, pancreatitis, or abdominal trauma.

Coagulation disorders that can lead to the development of portal vein thrombosis include polycythemia vera; essential thrombocytosis; deficiencies in protein C, protein S, antithrombin 3, and factor V Leiden; and abnormalities in the gene-regulating prothrombin production. Some patients may have a subclinical myeloproliferative disorder.

CLINICAL FEATURES

The three primary complications of portal hypertension are gastroesophageal varices with hemorrhage, ascites, and hypersplenism. Thus, patients may present with upper GI bleeding, which, on endoscopy, is found to be due to esophageal or gastric varices, with the development of ascites along with peripheral edema, or with an enlarged spleen with associated reduction in platelets and white blood cells on routine laboratory testing.

Esophageal Varices

Over the last decade, it has become common practice to screen known cirrhotics with endoscopy to look for esophageal varices. Such screening studies have shown that approximately one-third of patients with histologically confirmed cirrhosis have varices. Approximately 5–15% of cirrhotics per year develop varices, and it is estimated that the majority of patients with cirrhosis will develop varices over their lifetimes. Furthermore, it is anticipated that roughly one-third of patients with varices will develop bleeding. Several factors predict the risk of bleeding, including the severity of cirrhosis (Child's class, MELD score); the height of wedged-hepatic vein pressure; the size of the varix; the location of the varix; and certain endoscopic stigmata, including red wale signs, hematocystic spots, diffuse erythema, bluish color, cherry red spots, or white-nipple spots. Patients with tense ascites are also at increased risk for bleeding from varices.

DIAGNOSIS

In patients with cirrhosis who are being followed chronically, the development of portal hypertension is usually revealed by the presence of thrombocytopenia; the appearance of an enlarged spleen; or the development of ascites, encephalopathy, and/or esophageal varices with or without bleeding. In previously undiagnosed patients, any of these features should prompt further evaluation to determine the presence of portal hypertension and liver disease. Varices should be identified by endoscopy. Abdominal imaging, either by CT or MRI, can be helpful in demonstrating a nodular liver and in finding changes of portal hypertension with intraabdominal collateral circulation. If necessary, interventional radiologic procedures can be performed to determine wedged and free hepatic vein pressures that will allow for the calculation of a wedged-to-free gradient, which is equivalent to the portal pressure. The average normal wedged-to-free gradient is 5 mmHg, and patients with a gradient >12 mmHg are at risk for variceal hemorrhage.

TREATMENT: VARICEAL HEMORRHAGE

Treatment for variceal hemorrhage as a complication of portal hypertension is divided into two main categories: (1) primary prophylaxis and (2) prevention of re-bleeding once there has been an initial variceal hemorrhage. Primary prophylaxis requires routine screening by endoscopy of all patients with cirrhosis. Once varices that are at increased risk for bleeding are identified, primary prophylaxis can be achieved either through nonselective beta blockade or by variceal band ligation. Numerous placebo-controlled clinical trials of either propranolol or nadolol have been reported in the literature. The most rigorous studies were those that only included patients with significantly enlarged varices or with hepatic vein pressure gradients >12 mmHg. Patients treated with beta blockers have a lower risk of variceal hemorrhage than those treated with placebo over 1 and 2 years of follow-up. There is also a decrease in mortality related to variceal hemorrhage. Unfortunately, overall survival was improved in only one study. Further studies have demonstrated that the degree of reduction of portal pressure is a significant feature to determine success of therapy. Therefore, it has been suggested that repeat measurements of hepatic vein pressure gradients may be used to guide pharmacologic therapy; however, this may be cost-prohibitive. Several studies have evaluated variceal band ligation and

variceal sclerotherapy as methods for providing primary prophylaxis.

Endoscopic variceal ligation (EVL) has achieved a level of success and comfort with most gastroenterologists who see patients with these complications of portal hypertension. Thus, in patients with cirrhosis who are screened for portal hypertension and are found to have large varices, it is recommended that they receive either beta blockade or primary prophylaxis with EVL.

The approach to patients once they have had a variceal bleed is first to treat the acute bleed, which can be life-threatening, and then to prevent further bleeding. Prevention of further bleeding is usually accomplished with repeated variceal band ligation until varices are obliterated. Treatment of acute bleeding requires both fluid and blood-product replacement as well as prevention of subsequent bleeding with EVL.

The medical management of acute variceal hemorrhage includes the use of vasoconstricting agents, usually somatostatin or Octreotide. Vasopressin was used in the past but is no longer commonly used. Balloon tamponade (Sengstaken-Blakemore tube or Minnesota tube) can be used in patients who cannot get endoscopic therapy immediately or who need stabilization prior to endoscopic therapy. Control of bleeding can be achieved in the vast majority of cases; however, bleeding recurs in the majority of patients if definitive endoscopic therapy has not been instituted. Octreotide, a direct splanchnic vasoconstrictor, is given at dosages of 50–100 μg/h by continuous infusion. Endoscopic intervention is employed as first-line treatment to control bleeding acutely. Some endoscopists will use variceal injection therapy (sclerotherapy) as initial therapy, particularly when bleeding is vigorous. Variceal band ligation is used to control acute bleeding in over 90% of cases and should be repeated until obliteration of all varices is accomplished. When esophageal varices extend into the proximal stomach, band ligation is less successful. In these situations, when bleeding continues from gastric varices, consideration for transjugular intrahepatic portosystemic shunt (TIPS) should be made. This technique creates a portosystemic shunt by a percutaneous approach using an expandable metal stent, which is advanced under angiographic guidance to the hepatic veins and then through the substance of the liver to create a direct portocaval shunt. This offers an alternative to surgery for acute decompression of portal hypertension. Encephalopathy can occur in as many as 20% of patients after TIPS and is particularly problematic in elderly patients and in those patients with preexisting encephalopathy. TIPS should be reserved for those individuals who fail endoscopic or medical management or who are poor surgical risks. TIPS can sometimes be used as a bridge to transplantation. Surgical esophageal transsection is a procedure that is rarely used and generally is associated with a poor outcome.

Prevention of Recurrent Bleeding (Fig. 308-3)

Once patients have had an acute bleed and have been managed successfully, attention should be paid to preventing recurrent bleeding. This usually requires repeated variceal band ligation until varices are obliterated. Beta blockade may be of adjunctive benefit in patients who are having recurrent variceal band ligation; however, once varices have been obliterated, the need for beta blockade is lessened. Despite successful variceal obliteration, many patients will still have portal hypertensive gastropathy from which bleeding can occur. Nonselective beta blockade may be helpful to prevent further bleeding from portal hypertensive gastropathy once varices have been obliterated.

Figure 308-3

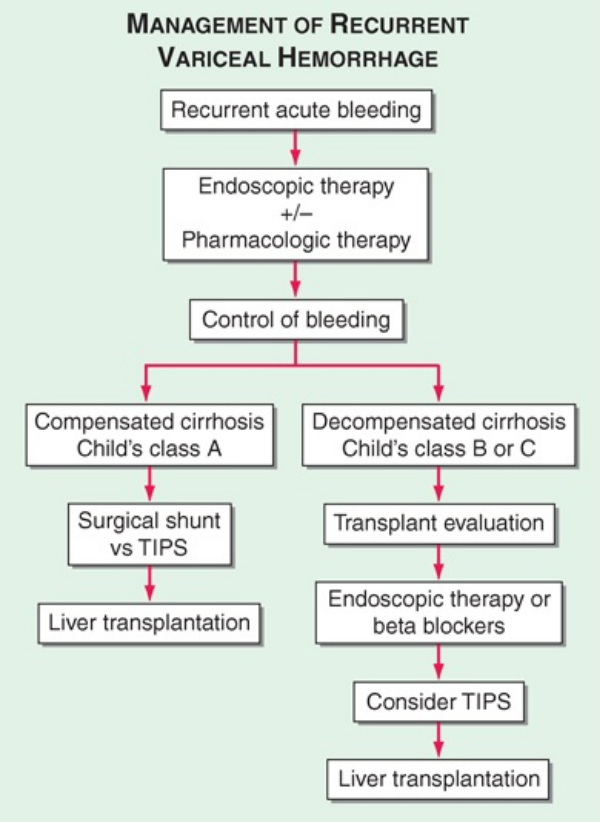

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of recurrent variceal hemorrhage. This algorithm describes an approach to management of patients who have recurrent bleeding from esophageal varices. Initial therapy is generally with endoscopic therapy often supplemented by pharmacologic therapy. With control of bleeding, a decision needs to be made as to whether patients should go on to a surgical shunt or TIPS (if they are Child's class A) and be considered for transplant, or if they should have TIPS and be considered for transplant (if they are Child's class B or C). TIPS, transjugular intrahepatic portosystemic shunt.

Portosystemic shunt surgery is less commonly performed with the advent of TIPS; nonetheless, this procedure should be considered for patients with good hepatic synthetic function who could benefit by having portal decompressive surgery.

Splenomegaly and Hypersplenism

Congestive splenomegaly is common in patients with portal hypertension. Clinical features include the presence of an enlarged spleen on physical examination and the development of thrombocytopenia and leukopenia in patients who have cirrhosis. Some patients will have fairly significant left-sided and left upper quadrant abdominal pain related to an enlarged and engorged spleen. Splenomegaly itself usually requires no specific treatment, although splenectomy can be successfully performed under very special circumstances.

Hypersplenism with the development of thrombocytopenia is a common feature of patients with cirrhosis and is usually the first indication of portal hypertension.

Ascites

DEFINITION

Ascites is the accumulation of fluid within the peritoneal cavity. Overwhelmingly, the most common cause of ascites is portal hypertension related to cirrhosis; however, clinicians should remember that malignant or infectious causes of ascites can be present as well, and careful differentiation of these other causes are obviously important for patient care.

PATHOGENESIS

The presence of portal hypertension contributes to the development of ascites in patients who have cirrhosis **(Fig. 308-4)**. There is an increase in intrahepatic resistance, causing increased portal pressure, but there is also vasodilation of the splanchnic arterial system, which, in turn, results in an increase in portal venous inflow. Both of these abnormalities result in increased production of splanchnic lymph. Vasodilating factors such as nitric oxide are responsible for the vasodilatory effect. These hemodynamic changes result in sodium retention by causing activation of the renin-angiotensin-aldosterone system with the development of hyperaldosteronism. The renal effects of increased

aldosterone leading to sodium retention also contribute to the development of ascites. Sodium retention causes fluid accumulation and expansion of the extracellular fluid volume, which results in the formation of peripheral edema and ascites. Sodium retention is the consequence of a homeostatic response caused by underfilling of the arterial circulation secondary to arterial vasodilation in the splanchnic vascular bed. Because the retained fluid is constantly leaking out of the intravascular compartment into the peritoneal cavity, the sensation of vascular filling is not achieved, and the process continues. Hypoalbuminemia and reduced plasma oncotic pressure also contribute to the loss of fluid from the vascular compartment into the peritoneal cavity. Hypoalbuminemia is due to decreased synthetic function in a cirrhotic liver.

Figure 308-4

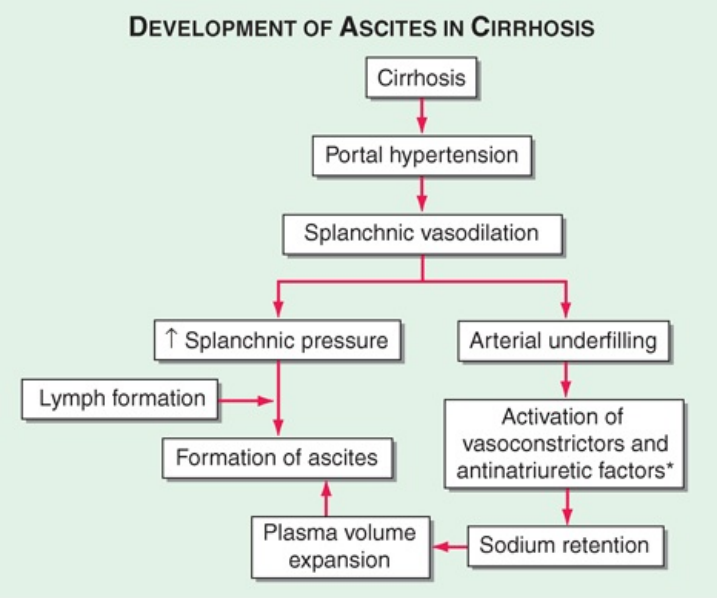

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Development of ascites in cirrhosis. This flow diagram illustrates the importance of portal hypertension with splanchnic vasodilation in the development of ascites. *Antinatriuretic factors include the renin-angiotensin-aldosterone system and the sympathetic nervous system.

CLINICAL FEATURES

Patients typically note an increase in abdominal girth that is often accompanied by the development of peripheral edema. The development of ascites is often insidious, and it is surprising that some patients wait so long and become so distended before seeking medical attention. Patients usually have at least 1–2 L of fluid in the abdomen before they are aware that there is an increase. If ascitic fluid is massive, respiratory function can be compromised, and patients will complain of shortness of breath. Hepatic hydrothorax may also occur in this setting, contributing to respiratory symptoms. Patients with massive ascites are often malnourished and have muscle wasting and excessive fatigue and weakness.

DIAGNOSIS

Diagnosis of ascites is by physical examination and is often aided by abdominal imaging. Patients will have bulging flanks, may have a fluid wave, or may have the presence of shifting dullness. This is determined by taking patients from a supine position to lying on either their left or right side and noting the movement of the dullness to percussion. Subtle amounts of ascites can be detected by ultrasound or CT scanning. Hepatic hydrothorax is more common on the right side and implicates a rent in the diaphragm with free flow of ascitic fluid into the thoracic cavity.

When patients present with ascites for the first time, it is recommended that a diagnostic paracentesis be performed to characterize the fluid. This should include the determination of total protein and albumin content, blood cell counts with differential, and cultures. In the appropriate setting, amylase may be measured and cytology performed. In patients with cirrhosis, the protein concentration of the ascitic fluid is quite low, with the majority of patients having an ascitic fluid protein concentration <1 g/dL. The development of the serum ascites-to-albumin gradient (SAAG) has replaced the description of exudative or transudative fluid. When the gradient between the serum albumin level and the ascitic fluid albumin level is >1.1 g/dL, the cause of the ascites is most likely due to portal hypertension; this is usually in the setting of cirrhosis. When the gradient is <1.1 g/dL, infectious or malignant causes of ascites should be considered. When levels of ascitic fluid proteins are very low, patients are at increased risk for developing SBP. A high level of red blood cells in the ascitic fluid signifies a traumatic

tap or perhaps a hepatocellular cancer or a ruptured omental varix. When the absolute level of polymorphonuclear leukocytes is >250/μL, the question of ascitic fluid infection should be strongly considered. Ascitic fluid cultures should be obtained using bedside inoculation of culture media.

TREATMENT: ASCITES

Patients with small amounts of ascites can usually be managed with dietary sodium restriction alone. Most average diets in the United States contain 6 to 8 g of sodium per day, and if patients eat at restaurants or fast-food outlets, the amount of sodium in their diet can exceed this amount. Thus, it is often extremely difficult to get patients to change their dietary habits to ingest <2 g of sodium per day, which is the recommended amount. Patients are frequently surprised to realize how much sodium is in the standard U.S. diet; thus, it is important to make educational pamphlets available to the patient. Often, a simple recommendation is to eat fresh or frozen foods, avoiding canned or processed foods, which are usually preserved with sodium. When a moderate amount of ascites is present, diuretic therapy is usually necessary. Traditionally, spironolactone at 100–200 mg/d as a single dose is started, and furosemide may be added at 40–80 mg/d, particularly in patients who have peripheral edema. In patients who have never received diuretics before, the failure of the above-mentioned dosages suggests that they are not being compliant with a low-sodium diet. If compliance is confirmed and ascitic fluid is not being mobilized, spironolactone can be increased to 400–600 mg/d and furosemide increased to 120–160 mg/d. If ascites is still present with these dosages of diuretics in patients who are compliant with a low-sodium diet, then they are defined as having *refractory ascites*, and alternative treatment modalities including repeated large-volume paracentesis, or a TIPS procedure should be considered **(Fig. 308-5)**. Recent studies have shown that TIPS, while managing the ascites, does not improve survival in these patients. Unfortunately, TIPS is often associated with an increased frequency of hepatic encephalopathy and must be considered carefully on a case-by-case basis. The prognosis for patients with cirrhosis with ascites is poor, and some studies have shown that <50% of patients survive 2 years after the onset of ascites. Thus, there should be consideration for liver transplantation in patients with the onset of ascites.

Figure 308-5

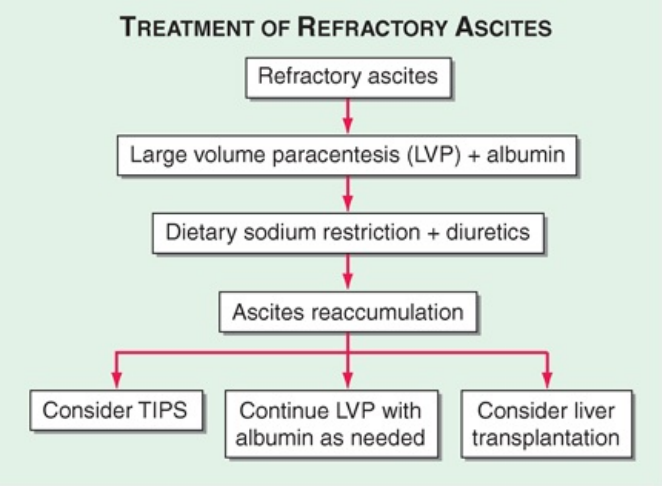

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment of refractory ascites. In patients who develop azotemia in the course of receiving diuretics in the management of their ascites, some will require repeated large-volume paracentesis (LVP), some may be considered for transjugular intrahepatic portosystemic shunt (TIPS), and some would be good candidates for liver transplantation. These decisions are all individualized.

Spontaneous Bacterial Peritonitis

SBP is a common and severe complication of ascites characterized by spontaneous infection of the ascitic fluid without an intraabdominal source. In patients with cirrhosis and ascites severe enough for hospitalization, SBP can occur in up to 30% of individuals and can have a 25% in-hospital mortality rate. Bacterial translocation is the presumed mechanism for development of SBP, with gut flora traversing the intestine into mesenteric lymph nodes, leading to bacteremia and seeding of the ascitic fluid. The most common organisms are *Escherichia coli* and other gut bacteria; however, gram-positive bacteria, including *Streptococcus viridans, Staphylococcus aureus*, and *Enterococcus* sp., can also be found. If more than two organisms are identified, secondary bacterial peritonitis due to a perforated viscus should be considered. The diagnosis of SBP is made when the fluid sample has an absolute neutrophil count >250/μL. Bedside cultures should be obtained when ascitic fluid is tapped. Patients with ascites may present with fever, altered mental status, elevated white blood cell count,

and abdominal pain or discomfort, or they may present without any of these features. Therefore, it is necessary to have a high degree of clinical suspicion, and peritoneal taps are important for making the diagnosis. Treatment is with a second-generation cephalosporin, with cefotaxime being the most commonly used antibiotic. In patients with variceal hemorrhage, the frequency of SBP is significantly increased, and prophylaxis against SBP is recommended when a patient presents with upper GI bleeding. Furthermore, in patients who have had an episode(s) of SBP and recovered, once-weekly administration of antibiotics is used as prophylaxis for recurrent SBP.

Hepatorenal Syndrome

The hepatorenal syndrome (HRS) is a form of functional renal failure without renal pathology that occurs in about 10% of patients with advanced cirrhosis or acute liver failure. There are marked disturbances in the arterial renal circulation in patients with HRS; these include an increase in vascular resistance accompanied by a reduction in systemic vascular resistance. The reason for renal vasoconstriction is most likely multifactorial and is poorly understood. The diagnosis is made usually in the presence of a large amount of ascites in patients who have a stepwise progressive increase in creatinine. Type 1 HRS is characterized by a progressive impairment in renal function and a significant reduction in creatinine clearance within 1–2 weeks of presentation. Type 2 HRS is characterized by a reduction in glomerular filtration rate with an elevation of serum creatinine level, but it is fairly stable and is associated with a better outcome than that of Type 1 HRS.

HRS is often seen in patients with refractory ascites and requires exclusion of other causes of acute renal failure. Treatment has, unfortunately, been difficult, and in the past, dopamine or prostaglandin analogues were used as renal vasodilating medications. Carefully performed studies have failed to show clear-cut benefit from these therapeutic approaches. Currently, patients are treated with midodrine, an α-agonist, along with octreotide and intravenous albumin. The best therapy for HRS is liver transplantation; recovery of renal function is typical in this setting. In patients with either type 1 or type 2 HRS, the prognosis is poor unless transplant can be achieved within a short period of time.

Hepatic Encephalopathy

Portosystemic encephalopathy is a serious complication of chronic liver disease and is broadly defined as an alteration in mental status and cognitive function occurring in the presence of liver failure. In acute liver injury with fulminant hepatic failure, the development of encephalopathy is a requirement for a diagnosis of fulminant failure. Encephalopathy is much more commonly seen in patients with chronic liver disease. Gut-derived neurotoxins that are not removed by the liver because of vascular shunting and decreased hepatic mass get to the brain and cause the symptoms that we know of as hepatic encephalopathy. Ammonia levels are typically elevated in patients with hepatic encephalopathy, but the correlation between severity of liver disease and height of ammonia levels is often poor, and most hepatologists do not rely on ammonia levels to make a diagnosis. Other compounds and metabolites that may contribute to the development of encephalopathy include certain false neurotransmitters and mercaptans.

CLINICAL FEATURES

In acute liver failure, changes in mental status can occur within weeks to months. Brain edema can be seen in these patients, with severe encephalopathy associated with swelling of the gray matter. Cerebral herniation is a feared complication of brain edema in acute liver failure, and treatment is meant to decrease edema with mannitol and judicious use of intravenous fluids.

In patients with cirrhosis, encephalopathy is often found as a result of certain precipitating events such as hypokalemia, infection, an increased dietary protein load, or electrolyte disturbances. Patients may be confused or exhibit a change in personality. They may actually be quite violent and difficult to manage; alternatively, patients may be very sleepy and difficult to rouse. Because precipitating events are so commonly found, they should be sought carefully. If patients have ascites, this should be tapped to rule out infection. Evidence of GI bleeding should be sought, and patients should be appropriately hydrated. Electrolytes should be measured and abnormalities corrected. In patients presenting with encephalopathy, asterixis is often present. Asterixis can be elicited by having patients extend their arms and bend their wrists back. In this maneuver, patients who are encephalopathic have a "liver flap"—i.e., a sudden forward movement of the wrist. This requires patients to be able to cooperate with the examiner and obviously cannot be elicited in patients who are severely encephalopathic or in hepatic coma.

The diagnosis of hepatic encephalopathy is clinical and requires an experienced clinician to recognize and put together all of the various features. Often when patients have encephalopathy for the first time, they are unaware of what is transpiring, but once they have been through the experience for the first time, they can identify when this is developing in subsequent situations and can often self-medicate to impair the development or worsening of encephalopathy.

TREATMENT: HEPATIC ENCEPHALOPATHY

Treatment is multifactorial and includes management of the above-mentioned precipitating factors. Sometimes hydration and correction of electrolyte imbalance is all that is necessary. In the past,

restriction of dietary protein was considered for patients with encephalopathy; however, the negative impact of that maneuver on overall nutrition is thought to outweigh the benefit when treating encephalopathy, and it is thus discouraged. There may be some benefit to replacing animal-based protein with vegetable-based protein in some patients with encephalopathy that is difficult to manage. The mainstay of treatment for encephalopathy, in addition to correcting precipitating factors, is to use lactulose, a nonabsorbable disaccharide, which results in colonic acidification. Catharsis ensues, contributing to the elimination of nitrogenous products in the gut that are responsible for the development of encephalopathy. The goal of lactulose therapy is to promote 2–3 soft stools per day. Patients are asked to titrate their amount of ingested lactulose to achieve the desired effect. Poorly absorbed antibiotics are often used as adjunctive therapies for patients who have had a difficult time with lactulose. The alternating administration of neomycin and metronidazole has commonly been employed to reduce the individual side effects of each: neomycin for renal insufficiency and ototoxicity and metronidazole for peripheral neuropathy. More recently, rifaximin at 550 mg twice daily has been very effective in treating encephalopathy without the known side effects of neomycin or metronidazole. Zinc supplementation is sometimes helpful in patients with encephalopathy and is relatively harmless. The development of encephalopathy in patients with chronic liver disease is a poor prognostic sign, but this complication can be managed in the vast majority of patients.

Malnutrition in Cirrhosis

Because the liver is principally involved in the regulation of protein and energy metabolism in the body, it is not surprising that patients with advanced liver disease are commonly malnourished. Once patients become cirrhotic, they are more catabolic, and muscle protein is metabolized. There are multiple factors that contribute to the malnutrition of cirrhosis, including poor dietary intake, alterations in gut nutrient absorption, and alterations in protein metabolism. Dietary supplementation for patients with cirrhosis is helpful in preventing patients from becoming catabolic.

Abnormalities in Coagulation

Coagulopathy is almost universal in patients with cirrhosis. There is decreased synthesis of clotting factors and impaired clearance of anticoagulants. In addition, patients may have thrombocytopenia from hypersplenism due to portal hypertension. Vitamin K–dependent clotting factors are Factors II, VII, IX, and X. Vitamin K requires biliary excretion for its subsequent absorption; thus, in patients with chronic cholestatic syndromes, vitamin K absorption is frequently diminished. Intravenous or intramuscular vitamin K can quickly correct this abnormality. More commonly, the synthesis of vitamin K–dependent clotting factors is diminished because of a decrease in hepatic mass, and, under these circumstances, administration of parenteral vitamin K does not improve the clotting factors or the prothrombin time. Platelet function is often abnormal in patients with chronic liver disease, in addition to decreases in platelet levels due to hypersplenism.

Bone Disease in Cirrhosis

Osteoporosis is common in patients with chronic cholestatic liver disease because of malabsorption of vitamin D and decreased calcium ingestion. The rate of bone resorption exceeds that of new bone formation in patients with cirrhosis resulting in bone loss. Dual x-ray absorptiometry (DEXA) is a useful method for determining osteoporosis or osteopenia in patients with chronic liver disease. When a DEXA scan shows decreased bone mass, treatment should be administered with bisphosphonates that are effective at inhibiting resorption of bone and efficacious in the treatment of osteoporosis.

Hematologic Abnormalities in Cirrhosis

Numerous hematologic manifestations of cirrhosis are present, including anemia from a variety of causes including hypersplenism, hemolysis, iron deficiency, and perhaps folate deficiency from malnutrition. Macrocytosis is a common abnormality in red blood cell morphology seen in patients with chronic liver disease, and neutropenia may be seen as a result of hypersplenism.

FURTHER READINGS

Arroyo V et al: Pathogenesis and treatment of hepatorenal syndrome. Semin Liver Dis 28:81, 2008[PMID: 18293279] [Full Text]

Bass NM et al: Rifaximin treatment in hepatic encephalopathy. N Engl J Med 362:1071, 2010[PMID: 20335583] [Full Text]

Friedman SL: Mechanisms of hepatic fibrogenesis. Gastroenterology 134:1655, 2008[PMID: 18471545] [Full Text]

Krowka MJ et al: Hepatopulmonary syndrome and portopulmonary hypertension: A report of the multicenter liver transplant database. Liver Transpl 10:174, 2004[PMID: 14762853] [Full Text]

O'Brien AO, Williams R: Nutrition and endstage liver disease: Principles and practice. Gastroenterology 134:1729, 2008[PMID: 21660345] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF LIVER BIOPSIES

Although clinical and laboratory features yield clues to the extent of inflammatory processes (disease grade), the degree of scarring and architectural distortion (disease stage), and the nature of the disease process, the liver biopsy is felt to represent the gold standard for assessing the degree of liver injury and fibrosis. Examination of liver histology provides not only a basis for quantitative scoring of disease activity and progression but also a wealth of qualitative information that can direct and inform diagnosis and management.

A normal liver lobule consists of portal (zone 1), lobular (midzonal or zone 2), and central (zone 3) zones. The portal tract contains the hepatic artery (HA) and portal vein (PV), which represent the dual vascular supply to the liver, as well as the bile duct (BD). The lobular area contains cords of liver cells surrounded by vascular sinusoids, and the central zone consists of the central vein (CV), the terminal branch of the hepatic vein (see figure below).

Included in this atlas of liver biopsies are examples of common morphologic features of acute and chronic liver disorders, some involving the lobular areas (e.g., the lobular inflammatory changes of acute hepatitis, apoptotic hepatocyte degeneration in acute and chronic hepatitis, virus antigen localization in hepatocyte cytoplasm and/or nuclei, viral inclusion bodies, copper or iron deposition, other inclusion bodies), and others involving the portal tracts (e.g., the portal mononuclear infiltrate that expands and spills over beyond the border of periportal hepatocytes in chronic hepatitis C, autoimmune hepatitis, and liver allograft rejection) or centrizonal areas (e.g., acute acetaminophen hepatotoxicity). Other histologic features of importance include hepatic steatosis (observed in alcoholic liver injury, in nonalcoholic fatty liver disorders, in metabolic disorders—including mitochondrial injury—and in patients with chronic viral hepatitis); injury of bile ducts in the portal tract, an important diagnostic hallmark of primary biliary cirrhosis, primary sclerosing cholangitis, as well as of liver allograft rejection; cholestasis in intrahepatic or extrahepatic biliary obstruction or in infiltrative disorders; ductular proliferation in the setting of marked hepatocellular necrosis; plasma cell infiltration common in autoimmune hepatitis; portal inflammation affecting portal veins ("endothelialitis") in liver allograft rejection; and mild-to-severe fibrosis, in varying distribution and pattern, as a consequence of liver injury common to many disorders. (All magnifications reflect the objective lens used.)

Figure e38-1

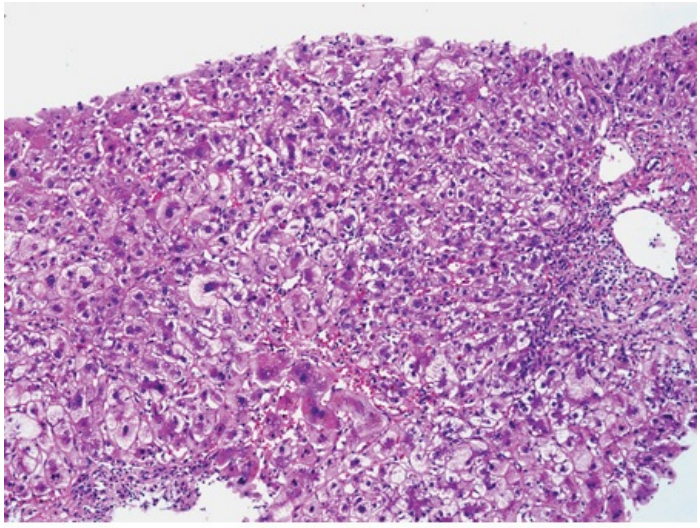

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute hepatitis with lobular inflammation and hepatocellular ballooning (H&E, 10x).

Figure e38-2

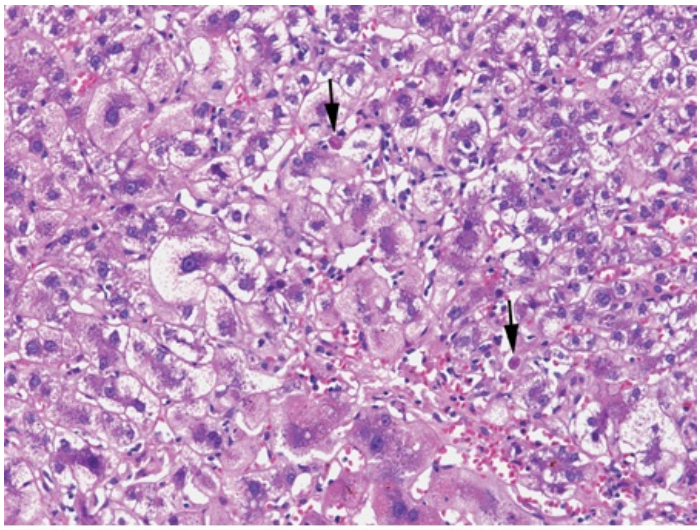

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute hepatitis, higher magnification, showing lobular inflammation, hepatocellular ballooning, and acidophilic bodies (*arrows*) (H&E, 20x).

Figure e38-3

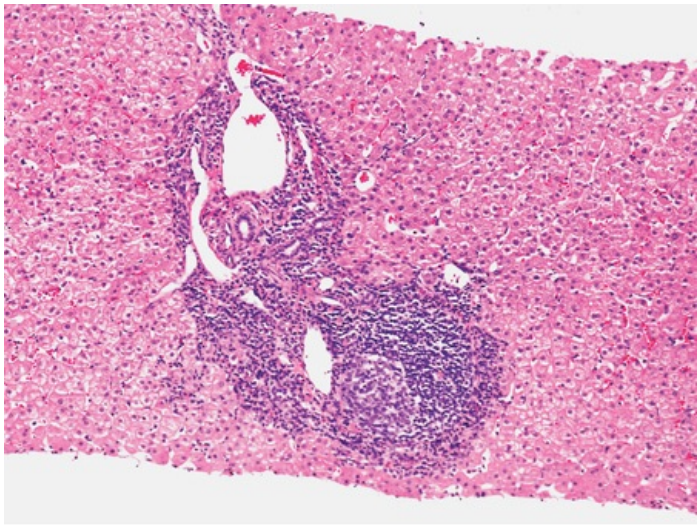

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic hepatitis C with portal lymphoid infiltrate and lymphoid follicle containing germinal center (H&E, 10x).

Figure e38-4

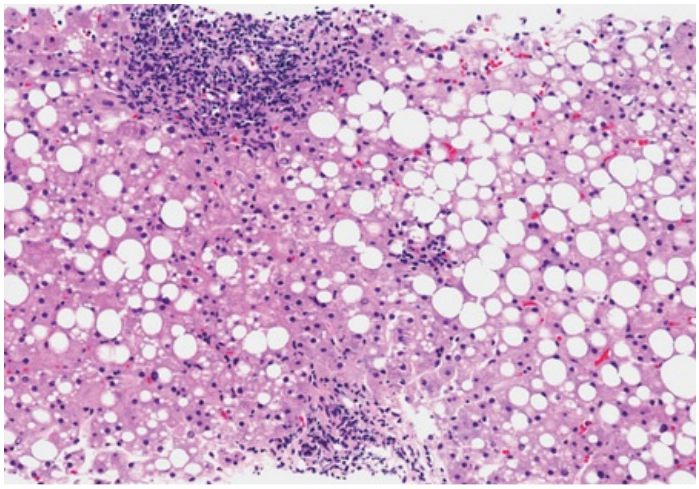

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic hepatitis C with portal and lobular inflammation and steatosis (H&E, 10x).

Figure e38-5

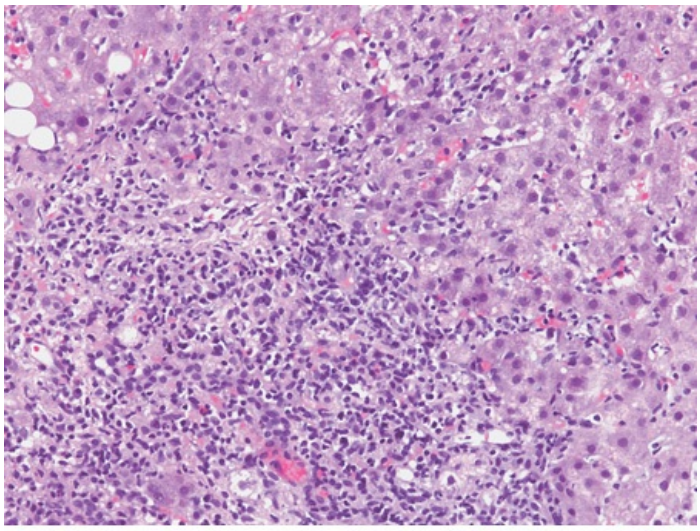

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic hepatitis C with portal inflammation and interface hepatitis (erosion of the limiting plate of periportal hepatocytes by infiltrating mononuclear cells) (H&E, 20x).

Figure e38-6

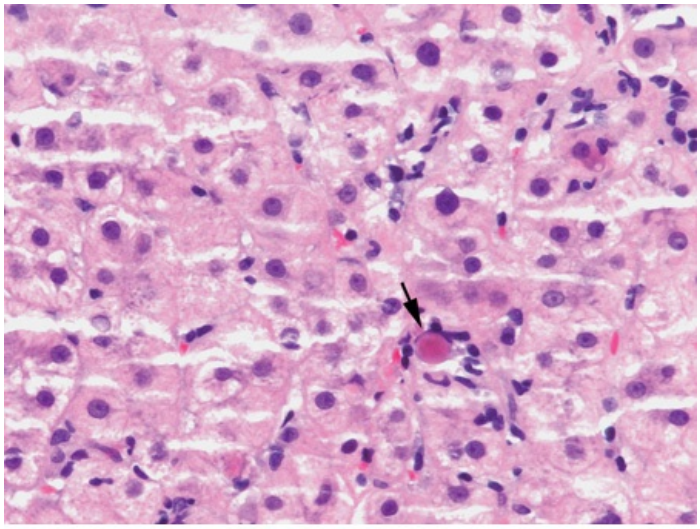

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lobular inflammation with acidophilic body (apoptotic body) surrounded by lymphoid cells (H&E, 40x).

Figure e38-7

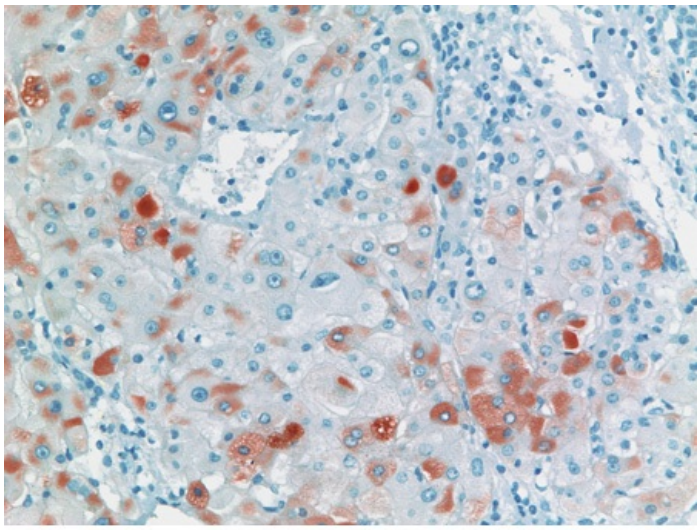

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic hepatitis B with hepatocellular cytoplasmic staining for hepatitis B surface antigen (immunoperoxidase, 20x).

Figure e38-8

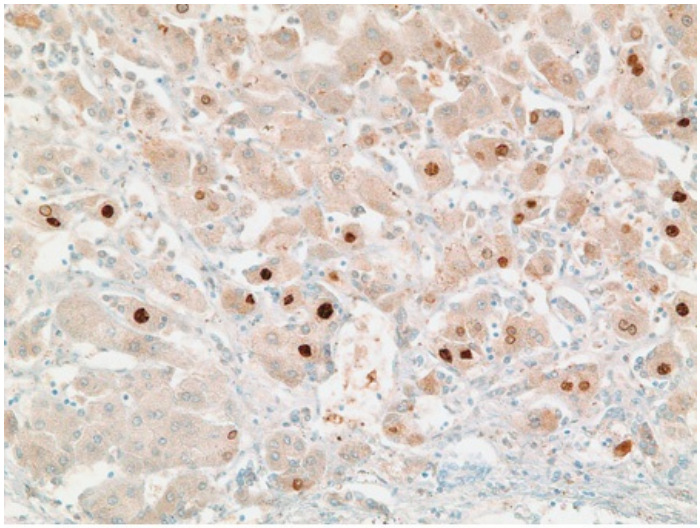

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic hepatitis B with hepatocellular nuclear staining for hepatitis B core antigen (immunoperoxidase, 20x).

Figure e38-9

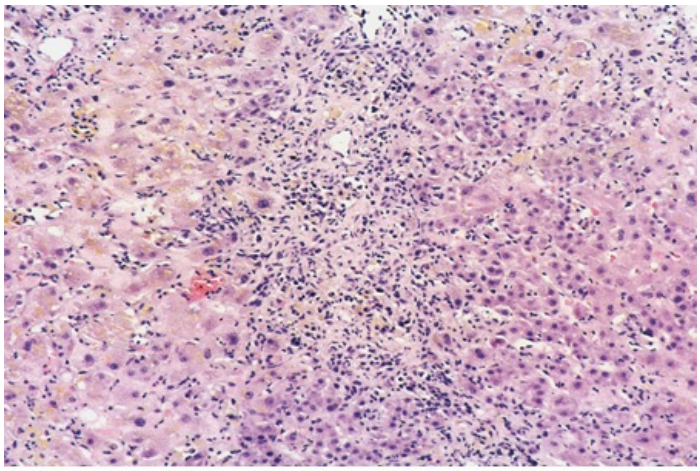

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Autoimmune hepatitis with portal and lobular inflammation, interface hepatitis, and cholestasis (H&E, 10x).

Figure e38-10

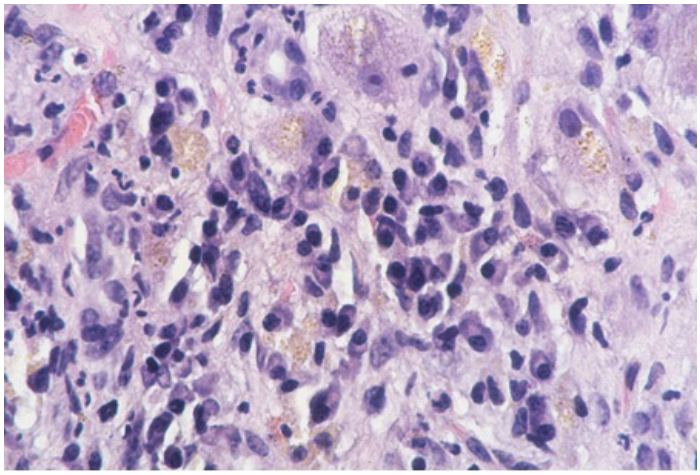

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Autoimmune hepatitis, higher magnification, showing dense plasma cell infiltrate in the portal and periportal regions (H&E, 40x).

Figure e38-11

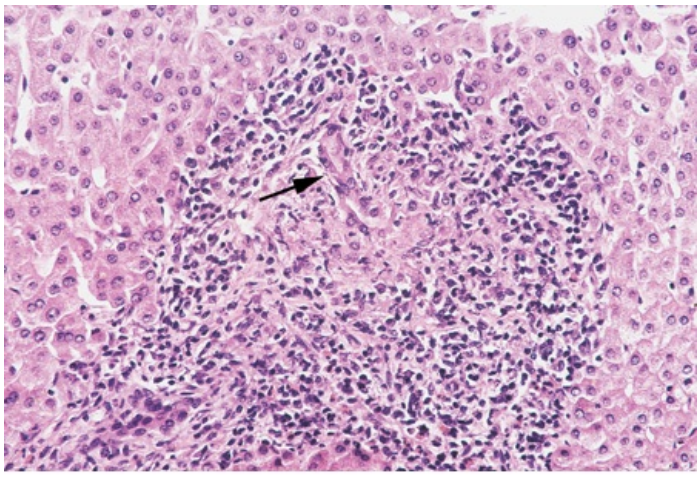

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary biliary cirrhosis with degenerating bile duct epithelium ("florid ductular lesion") (*arrow*) surrounded by epithelioid granulomatous reaction and lymphoplasmacytic infiltrate (H&E, 40x).

Figure e38-12

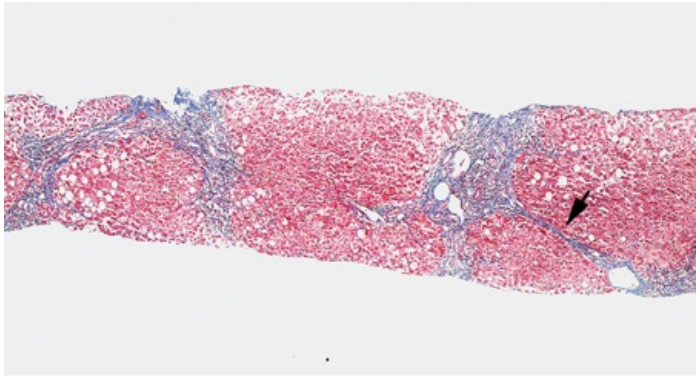

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic hepatitis C with bridging fibrosis (*arrow*) (Masson trichrome, 10x).

Figure e38-13

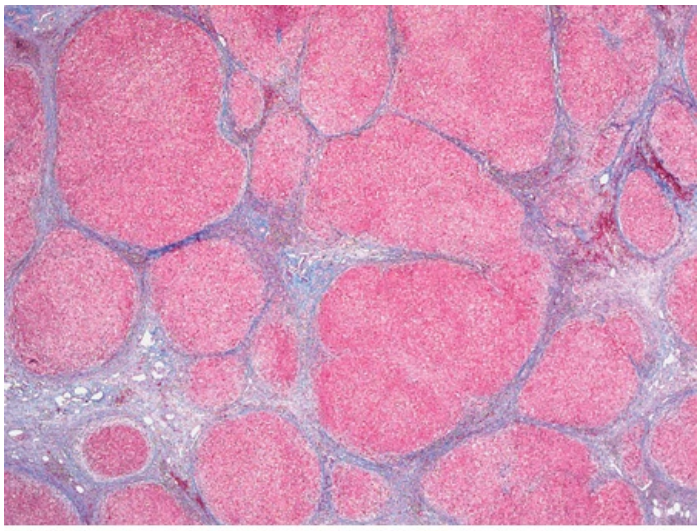

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cirrhosis with architectural alteration resulting from fibrosis and nodular hepatocellular regeneration (Masson trichrome, 2x).

Figure e38-14

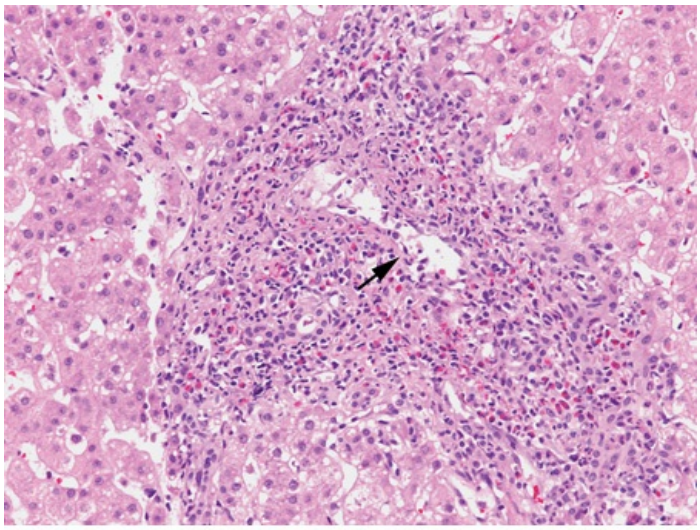

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute cellular rejection of orthotopic liver allograft demonstrating a mixed inflammatory cell infiltrate (lymphoid cells, eosinophils, neutrophils) of the portal tract as well as endothelialitis of the portal vein (*arrow*) and bile duct injury (H&E, 10x).

Figure e38-15

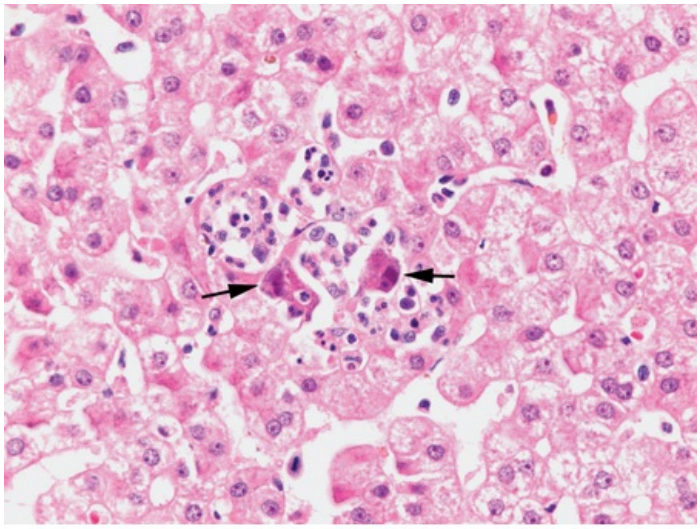

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Liver allograft with cytomegalovirus infection showing hepatocytes with nuclear inclusions (*arrows*) surrounded by a neutrophilic and lymphoid infiltrate (H&E, 10x).

Figure e38-16

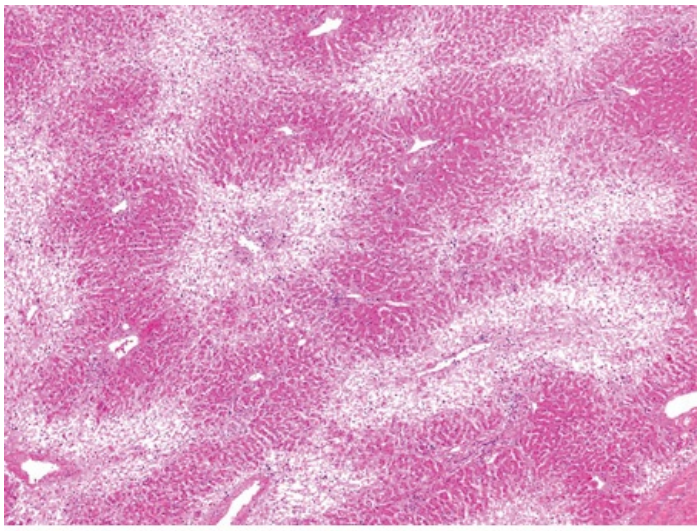

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Combined acetaminophen hepatotoxicity and alcoholic liver injury with extensive centrilobular areas of necrosis (H&E, 4x).

Figure e38-17

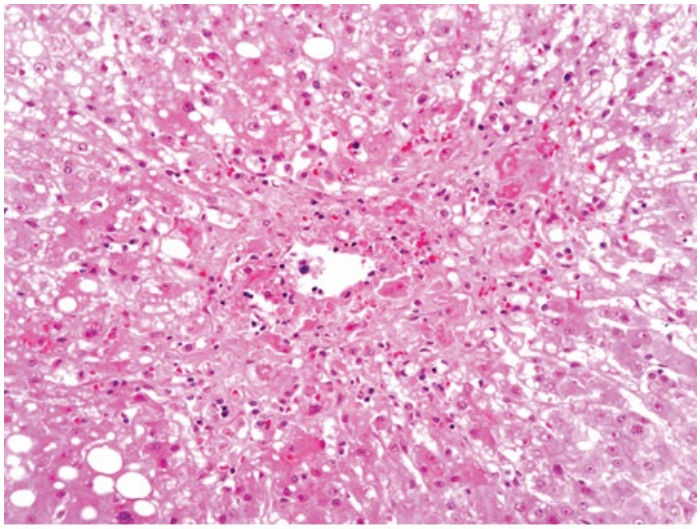

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Combined acetaminophen hepatotoxicity and alcoholic liver injury at higher magnification showing necrotic centrilobular area with Mallory bodies (H&E 20x).

Figure e38-18

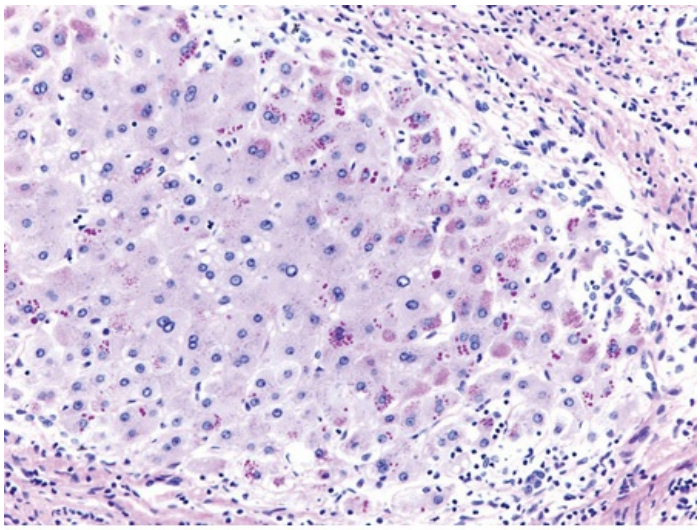

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

α1 antitrypsin deficiency with cytoplasmic periodic acid–Schiff (PAS)-positive, diastase-resistant globules in many hepatocytes, predominantly at the periphery of a cirrhotic nodule (PAS, 20x).

Figure e38-19

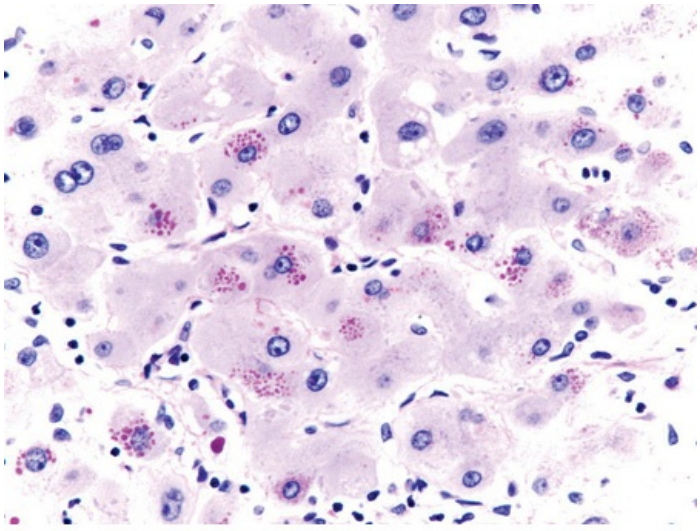

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

α_1 antitrypsin deficiency with higher magnification of PAS-positive, diastase-resistant globules (PAS, 40x).

Figure e38-20

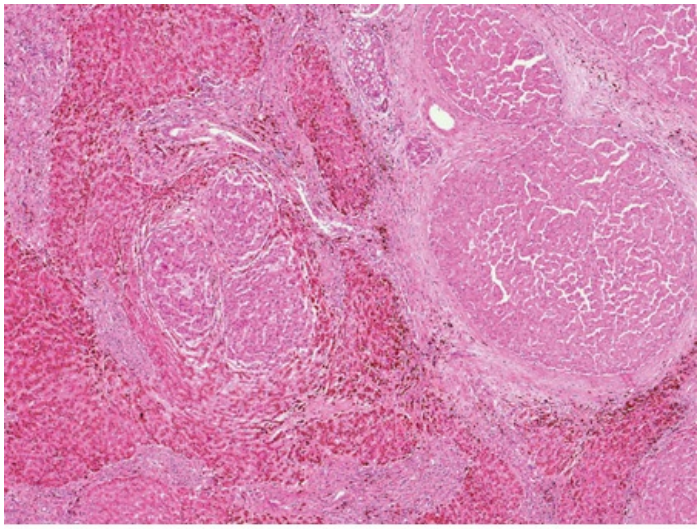

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cirrhosis secondary to hemochromatosis with hepatocellular carcinoma; brown hemosiderin pigment (iron) is present in the cirrhotic liver, while the hepatocellular carcinoma nodules are hemosiderin-free (H&E, 4x).

Figure e38-21

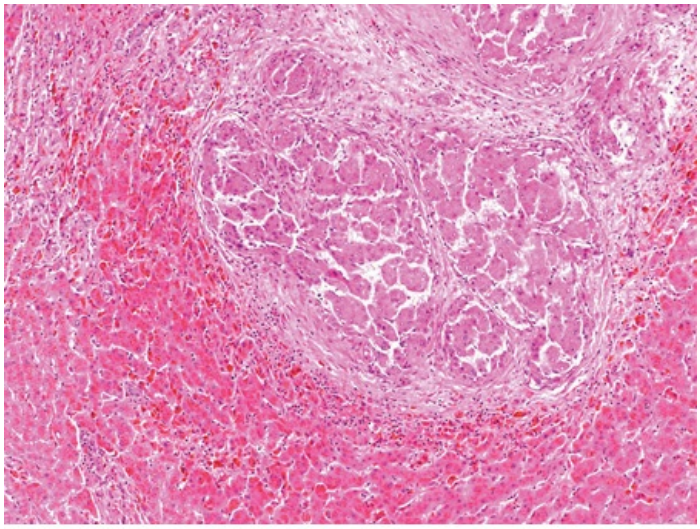

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cirrhosis secondary to hemochromatosis with hepatocellular carcinoma at higher magnification, demonstrating nodules of large malignant cells with highly disorganized architecture (H&E, 10x).

Figure e38-22

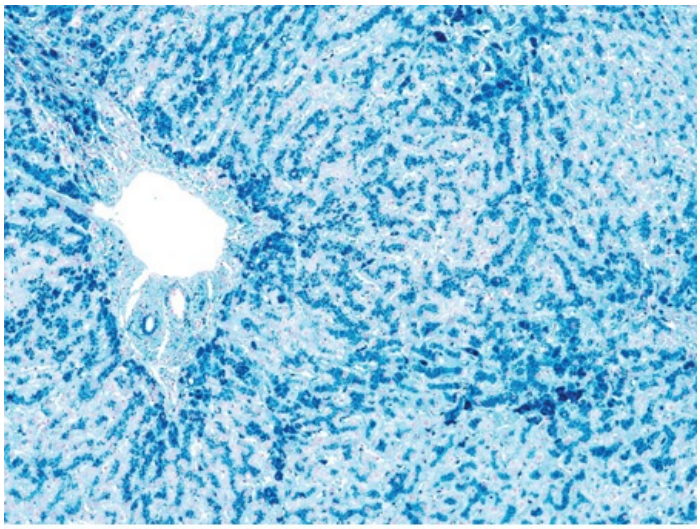

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hemochromatosis with iron stain demonstrating extensive iron deposition and characteristic pattern of pericanalicular distribution of iron (iron stain, 10x).

Figure e38-23

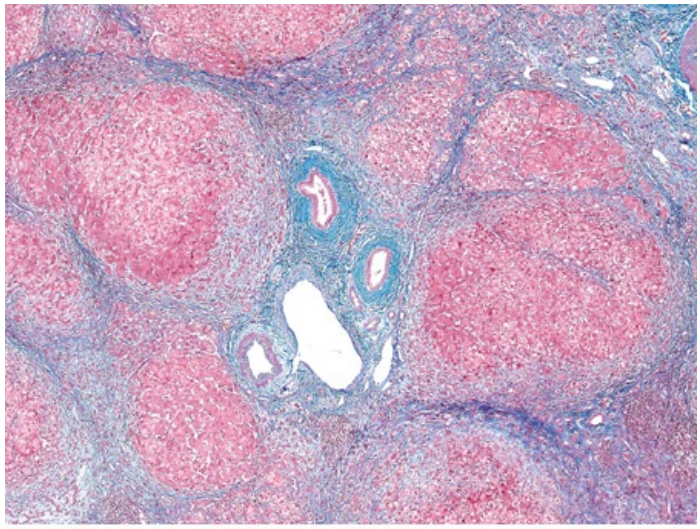

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary sclerosing cholangitis showing cirrhosis and periductular fibrosis (Masson trichrome, 4x).

Figure e38-24

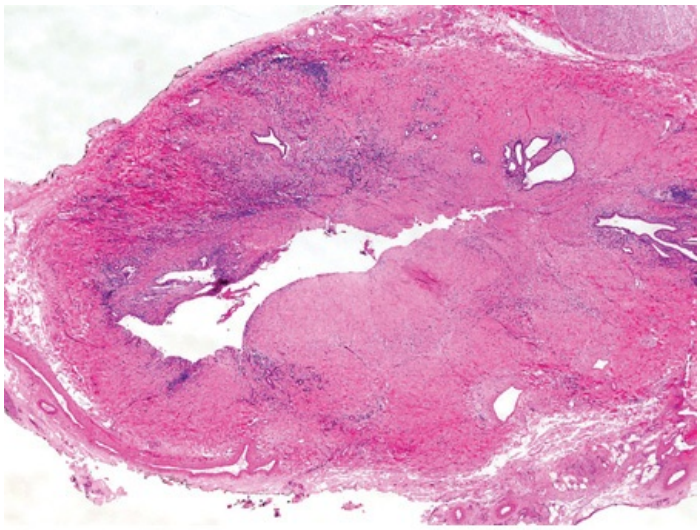

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary sclerosing cholangitis showing the extrahepatic bile duct (in a liver explant obtained at the time of hepatectomy for orthotopic liver transplantation) with marked mural chronic inflammation and fibrosis as well as peribiliary glands (H&E, 2x).

Figure e38-25

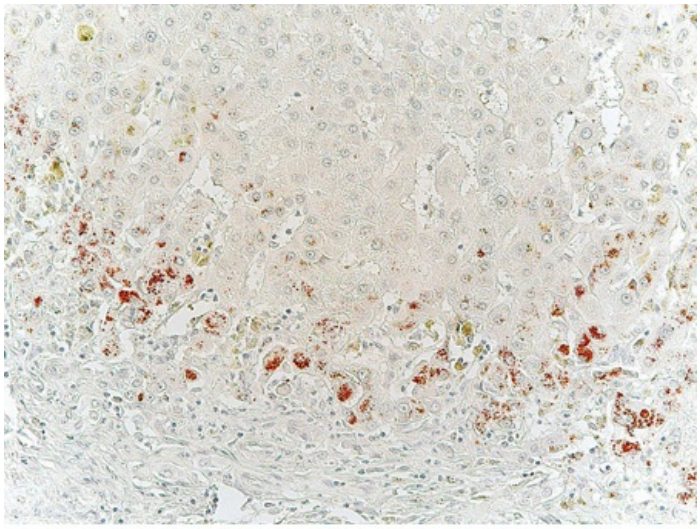

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary sclerosing cholangitis showing peripheral cholestasis (*green*) and cytoplasmic red granular staining of hepatocytes for copper (rhodanine copper stain, 20x).

Figure e38-26

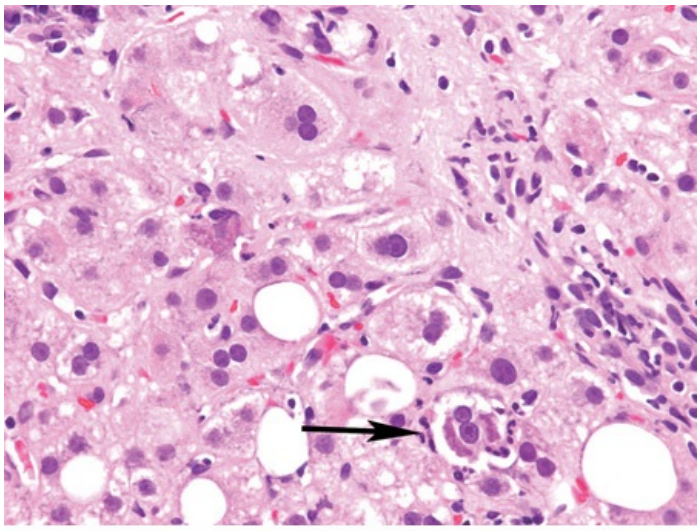

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nonalcoholic steatohepatitis (NASH) showing steatosis, ballooned hepatocytes, and Mallory bodies with surrounding polymorphonuclear leukocytes (*arrow*) (H&E, 20x).

Figure e38-27

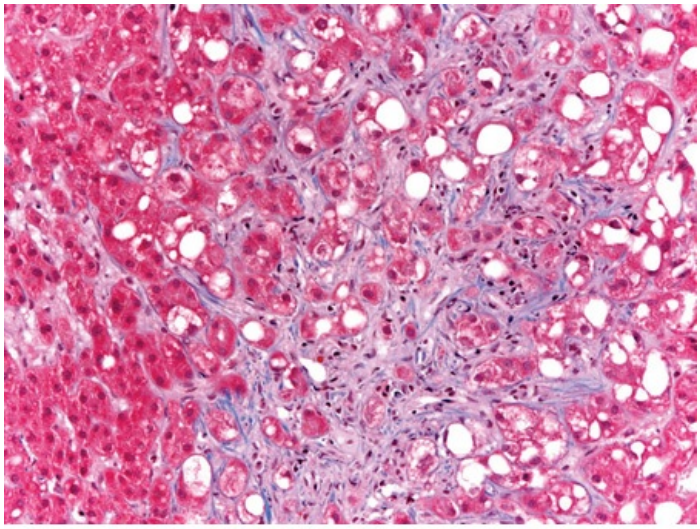

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nonalcoholic steatohepatitis (NASH) showing steatosis with perisinusoidal and pericellular fibrosis (H&E, 20x).

Figure e38-28

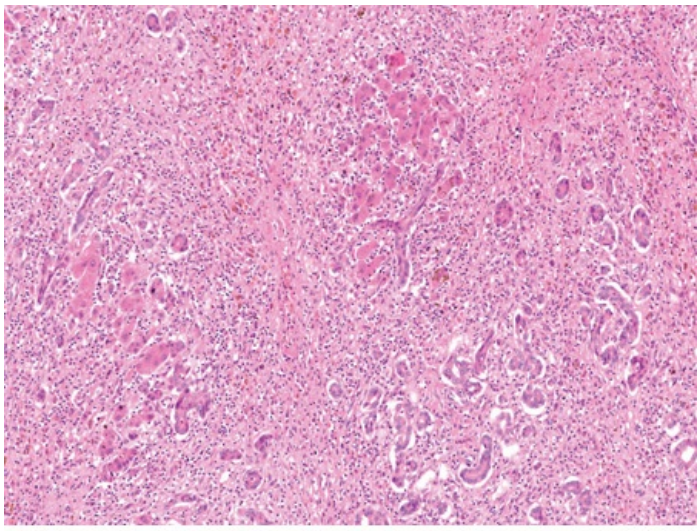

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute hepatitis with submassive hepatic necrosis with marked parenchymal collapse, remnant islands of surviving hepatocytes, and a marked ductular reaction (H&E, 10x).

Figure e38-29

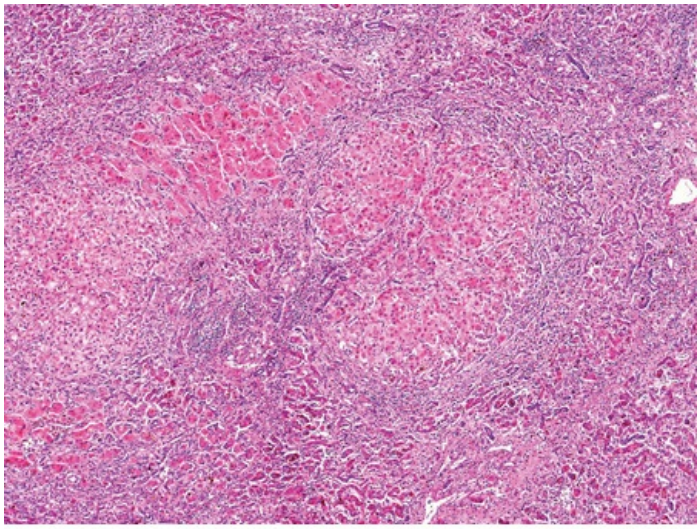

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Wilson's disease showing cirrhosis, extensive collapse, and ductular reaction in a teenager with an acute presentation (H&E, 4x).

Figure e38-30

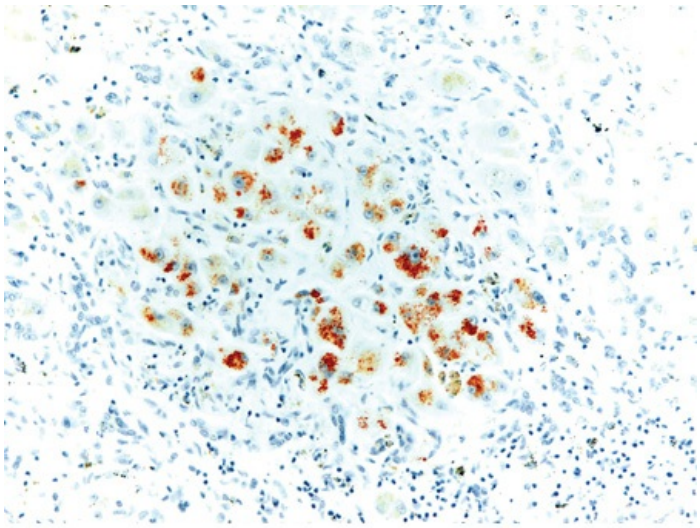

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Wilson's disease showing extensive hepatocyte cytoplasmic red granular staining for copper in a cirrhotic nodule (rhodanine copper stain, 20x).

Copyright © The McGraw-Hill Companies. All rights reserved.
The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GENETIC, METABOLIC, AND INFILTRATIVE DISEASES AFFECTING THE LIVER: INTRODUCTION

There are a number of disorders of the liver that fit within the categories of genetic, metabolic, and infiltrative disorders. Inherited disorders include hemochromatosis, Wilson's disease, α_1 antitrypsin (α_1AT) deficiency, and cystic fibrosis (CF). Hemochromatosis is the most common inherited disorder affecting Caucasian populations, with the genetic susceptibility for the disease being identified in 1 in 250 individuals. Over the past 15 years, it has become increasingly apparent that nonalcoholic fatty liver disease (NAFLD) is the most common cause of elevated liver enzymes found in the U.S. population. With the obesity epidemic in the United States, it is estimated that 20% of the population may have abnormal liver enzymes on the basis of NAFLD and 3% may have nonalcoholic steatohepatitis (NASH). Infiltrative disorders of the liver are relatively rare.

Genetic Liver Diseases

HEREDITARY HEMOCHROMATOSIS

Hereditary hemochromatosis (HH) is a common inherited disorder of iron metabolism **(Chap. 357)**. Our knowledge of the disease and its phenotypic expression has changed since 1996, when the gene for HH, called *HFE*, was identified, allowing for genetic testing for the two major mutations (C282Y and H63D) that are responsible for *HFE*-related HH. Subsequently, several additional genes/proteins involved in the regulation of iron homeostasis have been identified, contributing to a better understanding of cellular iron uptake and release and the characterization of additional causes of inherited iron overload **(Table 309-1)**.

Table 309-1 Classification of Iron Overload Syndromes

Hereditary Hemochromatosis (HH)
HFE-related (type 1) C282Y/C282Y C282Y/H63D Other *HFE* mutations
Non-*HFE*-related Juvenile HH HJV—hemojuvelin (type 2a) HAMP—hepcidin (type 2b) TfR2-related HH (type 3) Ferroportin-related HH (type 4) African iron overload
Secondary Iron Overload
Iron-loading anemias Parenteral iron overload Chronic liver disease
Miscellaneous
Neonataliron overload Aceruloplasminemia Congenital atransferrinemia

Abbreviations: HJV, hemojuvelin; HAMP, hepcidin; TfR2, transferrin receptor 2.

Most patients with HH are asymptomatic; however, when patients present with symptoms, they are frequently nonspecific and include weakness, fatigue, lethargy, and weight loss. Specific, organ-related symptoms include abdominal pain, arthralgias, and symptoms and signs of chronic liver disease. Increasingly, most patients are now identified before they have symptoms, either through family studies or from the performance of screening iron studies. Several prospective population studies have shown that C282Y homozygosity is found in about 1 in 250 individuals of Northern European descent, with the heterozygote frequency seen in approximately 1 in 10 individuals. It is important to consider HH in patients who present with the symptoms and signs known to occur in established HH. When confronted with abnormal serum iron studies, clinicians should not wait for typical symptoms or findings of HH to appear before considering the diagnosis. However, once the diagnosis of HH is considered, either by an evaluation of abnormal screening iron studies, in the context of family studies, in a patient with an abnormal genetic test, or in the evaluation of a patient with any of the typical symptoms or clinical findings, definitive diagnosis is relatively straightforward. Transferrin saturation [serum iron divided by total iron-binding capacity (TIBC) or transferrin, times 100%] and ferritin levels should be obtained. Both of these will be elevated in a symptomatic patient. It must be remembered that ferritin is an acute-phase reactant and can be elevated in a number of other inflammatory disorders, such as rheumatoid arthritis, or in various neoplastic diseases, such as lymphoma or other cancers. Also, serum ferritin is elevated in a majority of patients with NASH, in the absence of iron overload.

At present, if patients have an elevated transferrin saturation or ferritin level, genetic testing should be performed; if they are a C282Y homozygote or a compound heterozygote (C282Y/H63D), the diagnosis is confirmed. If the ferritin is >1000 μg/L, the patient should be considered for liver biopsy because there is an increased frequency of advanced fibrosis in these individuals. If liver biopsy is performed, iron deposition is found in a periportal distribution with a periportal to pericentral gradient; iron is found predominantly in parenchymal cells, and Kupffer cells are spared.

Treatment: Hereditary Hemochromatosis

Treatment of HH is relatively straightforward with weekly phlebotomy aimed to reduce iron stores, recognizing that each unit of blood contains 200 to 250 mg of iron. If patients are diagnosed and treated before the development of hepatic fibrosis, all complications of the disease can be avoided. Maintenance phlebotomy is required in most patients and usually can be achieved with 1 unit of blood removed every 2–3 months. Family studies should be performed with transferrin saturation, ferritin, and genetic testing offered to all first-degree relatives.

WILSON'S DISEASE

Wilson's disease is an inherited disorder of copper homeostasis first described in 1912 **(Chap. 360)**. The Wilson's disease gene was discovered in 1993, with the identification of *ATP7B*. This P-type ATPase is involved in copper transport and is necessary for the export of copper from the hepatocyte. Thus, in patients with mutations in *ATP7B*, copper is retained in the liver, leading to increased copper storage and ultimately liver disease as a result.

The clinical presentation of Wilson's disease is variable and includes chronic hepatitis, hepatic steatosis, and cirrhosis in adolescents and young adults. Neurologic manifestations indicate that liver disease is present and include speech disorders and various movement disorders. Diagnosis includes the demonstration of a reduced ceruloplasmin level, increased urinary excretion of copper, the presence of Kayser-Fleischer rings in the corneas of the eyes, and an elevated hepatic copper level, in the appropriate clinical setting. The genetic diagnosis of Wilson's disease is difficult because >200 mutations in *ATP7B* have been described with different degrees of frequency and penetration in certain populations.

Treatment: Wilson's Disease

Treatment consists of copper-chelating medications such as D-penicillamine and trientine. A role for zinc acetate has also been established. Medical treatment is lifelong, and severe relapses leading to liver failure and death can occur with cessation of therapy. Liver transplantation is curative with respect to the underlying metabolic defect and restores the normal phenotype with respect to copper homeostasis.

α1 ANTITRYPSIN DEFICIENCY

Alpha-1-antitrypsin (AAT) deficiency was first described in the late 1960s in patients with severe pulmonary disease. AAT is a 52 kD glycoprotein produced in hepatocytes, phagocytes, and epithelial cells in the lungs, which inhibits serine proteases, primarily neutrophil elastase. In AAT deficiency, increased amounts of neutrophil elastase can result in progressive lung injury from degradation of elastin leading to premature emphysema. In the 1970s, AAT deficiency was discovered as a cause of neonatal liver disease, so-called "neonatal hepatitis." It is now known to be a cause of liver disease in infancy, early childhood, adolescence, and in adults.

In AAT deficiency, variants in the proteinase inhibitor (Pi) gene located on chromosome 14, alters AAT structure interfering with hepatocellular export. Aggregated, deformed polymers of AAT accumulate in

the hepatocyte endoplasmic reticulum. There are over 75 different AAT variants. Conventional nomenclature identifies normal variants as PiMM; these individuals have normal blood levels of AAT. The most common abnormal variants are called S and Z. Individuals homozygous for the Z mutation (PiZZ) have low levels of AAT (about 15% of normal) and these patients are susceptible to liver and/or lung disease, yet only a proportion (about 25%) of PiZZ patients develop disease manifestations. Null variants have undetectable levels of AAT and are susceptible to premature lung disease.

AAT deficiency has been identified in all populations; however, the disorder is most common in patients of Northern European and Iberian descent. The disorder affects about 1 in 1500 to 2000 individuals in North America. The natural history of AAT deficiency is quite variable because many individuals with the PiZZ variant never develop disease, whereas others can develop childhood cirrhosis leading to liver transplantation.

In adults, the diagnosis often comes in the course of evaluation of patients with abnormal liver test abnormalities or in a work-up for cirrhosis. A hint to diagnosis may be coexistent lung disease at a relatively young age or a family history of liver and/or lung disease. Patients may have symptoms of pulmonary disease with cough and dyspnea. Liver disease may be asymptomatic other than fatigue, or patients may present with complications of decompensated liver disease.

Diagnosis of AAT deficiency is confirmed by blood tests showing reduced levels of serum AAT, accompanied by Pi determinations. Most patients with liver disease have either PiZZ or PiSZ; occasionally, patients with PiMZ have reduced levels of AAT, but they usually do not have a low enough level to cause disease. Liver biopsy is often performed to determine stage of hepatic fibrosis and shows characteristic PAS-positive, diastase-resistant globules in the periphery of the hepatic lobule.

Treatment α_1 Antitrypsin Deficiency

Treatment of AAT deficiency is usually nonspecific and supportive. For patients with liver involvement, other sources of liver injury, such as alcohol, should be avoided. Evidence for other liver diseases (e.g., viral hepatitis B and C, hemochromatosis, NAFLD, etc.) should be sought and treated if possible. Smoking can worsen lung disease progression in AAT and should be discontinued. Patients with lung disease may be eligible to receive infusions of AT, which has been shown to halt further damage to the lungs. If liver disease becomes decompensated, transplantation should be pursued and is curative. Following transplant, patients express the Pi phenotype of the donor. Finally, risk of hepatocellular carcinoma is significantly increased in patients with cirrhosis due to AAT deficiency.

CYSTIC FIBROSIS

CF should also be considered as an inherited form of chronic liver disease, although the principal manifestations of CF include chronic lung disease and pancreatic insufficiency **(Chap. 259)**. A small percentage of patients with CF who survive to adulthood have a form of biliary cirrhosis characterized by cholestatic liver enzyme abnormalities and the development of chronic liver disease. Ursodeoxycholic acid is occasionally helpful in improving liver test abnormalities and in reducing symptoms. The disease is slowly progressive.

Metabolic Liver Diseases

NONALCOHOLIC FATTY LIVER DISEASE

NAFLD was first described in the 1950s when fatty liver was characterized in a group of obese patients. In 1980, Ludwig and colleagues at the Mayo Clinic described 20 obese, diabetic, nonalcoholic patients who had similar findings on liver biopsy to patients with alcoholic liver disease, and the term nonalcoholic steatohepatitis was introduced. The prevalence of NAFLD in the United States and Europe ranges from 14–20%. This increased prevalence relates directly to the obesity epidemic seen in these populations. In the United States, NASH is thought to occur in ~3% of the general population, with fibrosis due to NASH being seen in >40% of obese patients. The spectrum of NAFLD includes simple hepatic steatosis, which, over time, can progress to NASH, with the subsequent development of fibrosis and cirrhosis. Causes of macrovesicular steatosis are listed in **Table 309-2**. It is now known that many patients with hitherto identified "cryptogenic" cirrhosis in fact have liver disease on the basis of NASH, with the resolution of the steatosis once patients become catabolic due to cirrhosis.

Table 309-2 Causes of Macrovesicular Steatosis
Insulin resistance, hyperinsulinemia
Centripetal obesity
Type 2 diabetes
Medications
Glucocorticoids

Estrogens
Tamoxifen
Amiodarone
Nutritional
Starvation
Protein deficiency (Kwashiorkor)
Choline deficiency
Liver disease
Wilson disease
Chronic hepatitis C—genotype 3
Indian childhood cirrhosis
Jejunoileal bypass

Most patients who come to medical attention with NAFLD are identified as a result of incidentally discovered elevated liver enzymes (ALT, AST). When patients are symptomatic, symptoms include fatigue or a vague right upper quadrant discomfort. ALT is generally higher than AST, and aminotransferases are only mildly (1.5–2 times the upper limit of normal) elevated. Recent studies have shown that many patients can have advanced fibrosis with NASH and even cirrhosis due to NASH with normal liver enzymes, indicating that the prevalence of the disease is likely to be even greater than was previously suspected. NASH is frequently seen in conjunction with other components of the metabolic syndrome (hypertension, diabetes mellitus, elevated lipids, and obesity), with NAFLD being considered the hepatic manifestation of this syndrome **(Chap. 242)**. Insulin resistance is the underlying link between these various disorders and numerous studies have shown that virtually all patients with NASH have insulin resistance. Abnormal ferritin values are seen in ~50% of patients with NASH, and an elevated ferritin level may be a marker of insulin resistance in NASH.

The diagnosis of NAFLD requires a careful history to determine the amount of alcohol used. Most investigators in the field of fatty liver disease require that <20 g/d of alcohol be consumed to exclude alcoholic liver disease. Laboratory testing for other liver diseases such as hepatitis B and C, iron studies, ceruloplasmin, α-^{1} antitrypsin levels, and autoimmune serologies should also be determined. Imaging studies can show characteristic features of a fatty liver, but the ultimate diagnosis of either hepatic steatosis or NASH requires liver biopsy. Liver biopsy shows characteristic macrovesicular steatosis with occasional microvesicular fat being identified. A mixed inflammatory infiltrate is found in a lobular distribution. The histologic features of NASH are very similar to those seen in alcoholic liver disease; Mallory's hyaline can be seen in both disorders, although the number of hepatocytes containing Mallory's hyaline and the size of the deposits are frequently greater in alcoholic liver disease than in NASH. The fibrosis that occurs in NASH has a characteristic perivenular and perisinusoidal distribution. Most cross-sectional studies show that up to 30–40% of NASH patients can develop advanced fibrosis, with cirrhosis being identified in 10–15% of individuals in series. Increasingly, patients are being identified with cryptogenic cirrhosis who have most likely had NASH for decades. These patients can develop liver failure and require liver transplantation, and some patients can progress to the development of hepatocellular cancer. Often, when cirrhotic, these patients will not have steatosis on biopsy, but following transplant, NAFLD will frequently recur.

Treatment: Nonalcoholic Fatty Liver Disease

The mainstay of treatment of fatty liver disease is weight loss and exercise, which is often difficult to achieve in this population. As an aid to weight loss, orlistat, which is a reversible inhibitor of gastric and pancreatic lipase, has been shown to result in a small decrease in body weight and is usually fairly well tolerated. This medication is now available over-the-counter. Bariatric surgery has been used and shows striking success, but is obviously a fairly drastic maneuver for induction of weight loss. Recent studies have focused on the presence of insulin resistance at the center of the pathophysiologic mechanisms of NAFLD. The thiazolidinedione medications are PPAR gamma inhibitors, which improve insulin sensitivity within the adipocyte and skeletal muscle by upregulating specific protein kinases involved in decreasing fatty acid synthesis. Two drugs—pioglitizone and rosiglitizone—are currently available and are being evaluated as potential therapeutic options in the treatment of NASH. Antioxidants have also been used, and a recent large multicenter study has shown benefit from vitamin E supplementation. Treatment of hyperlipidemia with statin-type agents has shown improvement in liver enzymes, but they have not been assessed for effects on histology. Ursodeoxycholic acid has been used and improves liver enzymes in patients with many liver diseases, but it has not been definitively helpful for fatty liver disease. At present, efforts should be directed to encouraging patients with NAFLD to lose weight and exercise.

LIPID STORAGE DISEASES

There are a number of rare lipid storage diseases that involve the liver, including the inherited disorders of Gaucher's and Niemann-Pick disease **(Chap. 362)**. Other rare disorders include abetalipoproteinemia, Tangier disease, Fabray's disease, and types I and V hyperlipoproteinemia. Hepatomegaly is present due to increased fat deposition and increased glycogen found in the liver.

PORPHYRIAS

The porphyrias are a group of metabolic disorders in which there are defects in the biosynthesis of heme necessary for incorporation into numerous hemoproteins such as hemoglobin, myoglobin, catalase, and the cytochromes **(Chap. 358)**. Porphyrias can present as either acute or chronic diseases, with the acute disorder causing recurring bouts of abdominal pain, and the chronic disorders characterized by painful skin lesions. Porphyria cutanea tarda (PCT) is the most commonly encountered porphyria. Patients present with characteristic vesicular lesions on sun-exposed areas of the skin, principally the dorsum of the hands, the tips of the ears, or the cheeks. About 40% of patients with PCT have mutations in the gene for hemochromatosis (*HFE*), and ~50% have hepatitis C; thus, iron studies and *HFE* mutation analysis as well as hepatitis C testing should be considered in all patients who present with PCT. PCT is also associated with excess alcohol use and some medications, most notably estrogens.

Treatment: Porphyrias

The mainstay of treatment of PCT is iron reduction by therapeutic phlebotomy, which is successful in reversing the skin lesions in the majority of patients. If hepatitis C is present, this should be treated as well. Acute intermittent porphyria presents with abdominal pain, with the diagnosis made by avoidance of certain precipitating factors such as starvation or certain diets. Intravenous heme as hematin has been used for treatment.

Infiltrative Disorders

AMYLOIDOSIS

Amyloidosis is a metabolic storage disease that results from deposition of insoluble proteins that are aberrantly folded and assembled and then deposited in a variety of tissues **(Chap. 111)**. Amyloidosis is divided into two types, primary and secondary, based on the broad concepts of association with myeloma (primary) or chronic inflammatory illnesses (secondary). The disease is generally considered rare, although, in certain disease states or in certain populations, it can be more common. For example, when associated with familial Mediterranean fever, it is seen in high frequency in Sephardic Jews and Armenians living in Armenia and less frequently in Ashkenazi Jews, Turks, and Arabs. Amyloidosis frequently affects patients suffering from tuberculosis and leprosy and can be seen in upwards of 10–15% of patients with ankylosing spondylitis, rheumatoid arthritis, or Crohn's disease. In one surgical pathology series, amyloid was found in <1% of cases. The liver is commonly involved in cases of systemic amyloidosis, but it is frequently inapparent clinically and only documented at autopsy. Pathologic findings in the liver include positive staining with the Congo red histochemical stain where there is an apple-green birefringence noted under polarizing light.

GRANULOMAS

Granulomas are frequently found in the liver when patients are being evaluated for cholestatic liver enzyme abnormalities. Granulomas can be seen in primary biliary cirrhosis, but there are other characteristic clinical (e.g., pruritus, fatigue) and laboratory findings (cholestatic liver tests, antimitochondrial antibody) that allow for a definitive diagnosis of that disorder. Granulomatous infiltration can also be seen as the principal hepatic manifestation of sarcoidosis, and this is the most common presentation of hepatic granulomas **(Chap. 329)**. The vast majority of these patients do not require any specific treatment other than what would normally be used for treatment of their sarcoidosis. A small subset, however, can develop a particularly bothersome desmoplastic reaction with a significant increase in fibrosis, which can progress to cirrhosis and liver failure. These patients may require treatment with immunosuppressive therapy and may require liver transplantation. In patients who have granulomas in the liver not associated with sarcoidosis, treatment is rarely needed.

Diagnosis requires liver biopsy, and it is important to establish a diagnosis so that a cause for the elevated liver enzymes is carefully identified. Some medications can cause granulomatous infiltration of the liver, the most notable of which is allopurinol.

LYMPHOMA

Involvement of the liver with lymphoma can sometimes be with bulky mass lesions but can also be as a difficult-to-diagnose infiltrative disorder that does not show any characteristic findings on abdominal imaging studies **(Chap. 110)**. Patients may present with severe liver disease, jaundice, hypoalbuminemia, mild to moderately elevated aminotransferases, and an elevated alkaline phosphatase.

A liver biopsy is required for diagnosis and should be considered when routine blood testing does not lead to a diagnosis of the liver dysfunction.

FURTHER READINGS

Pietrangelo A: Hereditary hemochromatosis: Pathogenesis, diagnosis, and treatment. Gastroenterology 139:393, 2010[PMID: 20542038] [Full Text]

Roberts EA et al: Diagnosis and treatment of Wilson disease: An update. Hepatology 47:2089, 2008[PMID: 18506894] [Full Text]

Sanyal AJ et al: Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N Engl J Med 362:1675, 2010[PMID: 20427778] [Full Text]

Silverman EK, SANDHAUS RA: Clinical practice. Alpha 1-antitrypsin deficiency. N Engl J Med 360:2749, 2009[PMID: 19553648] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LIVER TRANSPLANTATION: INTRODUCTION

Liver transplantation—the replacement of the native, diseased liver by a normal organ (allograft)—has matured from an experimental procedure reserved for desperately ill patients to an accepted, lifesaving operation applied more optimally in the natural history of end-stage liver disease. The preferred and technically most advanced approach is *orthotopic transplantation*, in which the native organ is removed and the donor organ is inserted in the same anatomic location. Pioneered in the 1960s by Thomas Starzl at the University of Colorado and, later, at the University of Pittsburgh and by Roy Calne in Cambridge, England, liver transplantation is now performed routinely worldwide. Success measured as 1-year survival has improved from ~30% in the 1970s to ~90% today. These improved prospects for prolonged survival, dating back to the early 1980s, resulted from refinements in operative technique, improvements in organ procurement and preservation, advances in immunosuppressive therapy, and, perhaps most influentially, more enlightened patient selection and timing. Despite the perioperative morbidity and mortality, the technical and management challenges of the procedure, and its costs, liver transplantation has become the approach of choice for selected patients whose chronic or acute liver disease is progressive, life-threatening, and unresponsive to medical therapy. Based on the current level of success, the number of liver transplants has continued to grow each year; in 2009, 6320 patients received liver allografts in the United States. Still, the demand for new livers continues to outpace availability; as of mid-2010, 16,785 patients in the United States were on a waiting list for a donor liver. In response to this drastic shortage of donor organs, many transplantation centers supplement cadaver-organ liver transplantation with living-donor transplantation.

INDICATIONS

Potential candidates for liver transplantation are children and adults who, in the absence of contraindications (see below), suffer from severe, irreversible liver disease for which alternative medical or surgical treatments have been exhausted or are unavailable. *Timing of the operation is of critical importance*. Indeed, improved timing and better patient selection are felt to have contributed more to the increased success of liver transplantation in the 1980s and beyond than all the impressive technical and immunologic advances combined. Although the disease should be advanced, and although opportunities for spontaneous or medically induced stabilization or recovery should be allowed, the procedure should be done sufficiently early to give the surgical procedure a fair chance for success. Ideally, transplantation should be considered in patients with end-stage liver disease who are experiencing or have experienced a life-threatening complication of hepatic decompensation or whose quality of life has deteriorated to unacceptable levels. Although patients with well-compensated cirrhosis can survive for many years, many patients with quasi-stable chronic liver disease have much more advanced disease than may be apparent. As discussed below, the better the status of the patient prior to transplantation, the higher will be the anticipated success rate of transplantation. The decision about *when* to transplant is complex and requires the combined judgment of an experienced team of hepatologists, transplant surgeons, anesthesiologists, and specialists in support services, not to mention the well-informed consent of the patient and the patient's family.

Transplantation in Children

Indications for transplantation in children are listed in **Table 310-1**. The most common is *biliary atresia*. *Inherited or genetic disorders of metabolism* associated with liver failure constitute another major indication for transplantation in children and adolescents. In Crigler-Najjar disease type I and in certain hereditary disorders of the urea cycle and of amino acid or lactate-pyruvate metabolism, transplantation may be the only way to prevent impending deterioration of central nervous system function, despite the fact that the native liver is structurally normal. Combined heart and liver transplantation has yielded dramatic improvement in cardiac function and in cholesterol levels in children with homozygous familial hypercholesterolemia; combined liver and kidney transplantation has been successful in patients with primary hyperoxaluria type I. In hemophiliacs with transfusion-associated hepatitis and liver failure, liver transplantation has been associated with recovery of normal Factor VIII synthesis.

Table 310-1 Indications for Liver Transplantation

Children	Adults

Biliary atresia	Primary biliary cirrhosis
Neonatal hepatitis	Secondary biliary cirrhosis
Congenital hepatic fibrosis	Primary sclerosing cholangitis
Alagille's syndrome[a]	Autoimmune hepatitis
Byler's disease[b]	Caroli's disease[c]
α_1-Antitrypsin deficiency	Cryptogenic cirrhosis
Inherited disorders of metabolism	Chronic hepatitis with cirrhosis
Wilson's disease	Hepatic vein thrombosis
Tyrosinemia	Fulminant hepatitis
Glycogen storage diseases	Alcoholic cirrhosis
Lysosomal storage diseases	Chronic viral hepatitis
Protoporphyria	Primary hepatocellular malignancies
Crigler-Najjar disease type I	Hepatic adenomas
Familial hypercholesterolemia	Nonalcoholic steatohepatitis
Primary hyperoxaluria type I	Familial amyloid polyneuropathy
Hemophilia	

[a] Arteriohepatic dysplasia, with paucity of bile ducts, and congenital malformations, including pulmonary stenosis.

[b] Intrahepatic cholestasis, progressive liver failure, mental and growth retardation.

[c] Multiple cystic dilatations of the intrahepatic biliary tree.

Transplantation in Adults

Liver transplantation is indicated for end-stage *cirrhosis* of all causes (Table 310-1). In *sclerosing cholangitis* and *Caroli's disease* (multiple cystic dilatations of the intrahepatic biliary tree), recurrent infections and sepsis associated with inflammatory and fibrotic obstruction of the biliary tree may be an indication for transplantation. Because prior biliary surgery complicates and is a relative contraindication for liver transplantation, surgical diversion of the biliary tree has been all but abandoned for patients with sclerosing cholangitis. In patients who undergo transplantation for *hepatic vein thrombosis (Budd-Chiari syndrome)*, postoperative anticoagulation is essential; underlying myeloproliferative disorders may have to be treated but are not a contraindication to liver transplantation. If a donor organ can be located quickly, before life-threatening complications —including cerebral edema—set in, patients with acute liver failure are candidates for liver transplantation. Routine candidates for liver transplantation are patients with *alcoholic cirrhosis*, *chronic viral hepatitis*, and *primary hepatocellular malignancies* . Although all three of these categories are considered to be high risk, liver transplantation can be offered to carefully selected patients. Currently, chronic hepatitis C and alcoholic liver disease are the most common indications for liver transplantation, accounting for over 40% of all adult candidates who undergo the procedure. Patients with alcoholic cirrhosis can be considered as candidates for transplantation if they meet strict criteria for abstinence and reform; however, these criteria still do not prevent recidivism in up to a quarter of cases. Patients with chronic hepatitis C have early allograft and patient survival comparable to those of other subsets of patients after transplantation; however, reinfection in the donor organ is universal, recurrent hepatitis C is insidiously progressive, the impact of antiviral therapy is limited, allograft cirrhosis develops in 20–30% at 5 years, and cirrhosis and late organ failure are being recognized with increasing frequency beyond 5 years. In patients with chronic hepatitis B, in the absence of measures to prevent recurrent hepatitis B, survival after transplantation is reduced by approximately 10–20%; however, prophylactic use of hepatitis B immune globulin (HBIg) during and after transplantation increases the success of transplantation to a level comparable to that seen in patients with nonviral causes of liver decompensation. Specific oral antiviral drugs (e.g., lamivudine, adefovir, entecavir, and tenofovir disoproxil fumarate **(Chap. 306)**, can be used both for prophylaxis against and for treatment of recurrent hepatitis B, facilitating further the management of patients undergoing liver transplantation for end-stage hepatitis B; most transplantation centers rely on a combination of HBIg and antiviral drugs to manage patients with hepatitis B. Issues of disease recurrence are discussed in more detail below. Patients with nonmetastatic primary hepatobiliary tumors—primary hepatocellular carcinoma (HCC), cholangiocarcinoma, hepatoblastoma, angiosarcoma, epithelioid hemangioendothelioma, and multiple or massive hepatic adenomata—have undergone liver transplantation; however, for some hepatobiliary malignancies, overall survival is significantly lower than that for other categories of liver disease. Most transplantation centers have

reported 5-year recurrence-free survival rates in patients with unresectable HCC for single tumors <5 cm in diameter or for three or fewer lesions all <3 cm comparable to those seen in patients undergoing transplantation for nonmalignant indications. Consequently, liver transplantation is currently restricted to patients whose hepatic malignancies meet these criteria. Expanded criteria for patients with HCC are being evaluated. Because the likelihood of recurrent cholangiocarcinoma is very high, only highly selected patients with limited disease are being evaluated for transplantation after intensive chemotherapy and radiation.

CONTRAINDICATIONS

Absolute contraindications for transplantation include life-threatening systemic diseases, uncontrolled extrahepatic bacterial or fungal infections, preexisting advanced cardiovascular or pulmonary disease, multiple uncorrectable life-threatening congenital anomalies, metastatic malignancy, and active drug or alcohol abuse **(Table 310-2)**. Because carefully selected patients in their sixties and even seventies have undergone transplantation successfully, advanced age per se is no longer considered an absolute contraindication; however, in older patients a more thorough preoperative evaluation should be undertaken to exclude ischemic cardiac disease and other comorbid conditions. Advanced age (>70 years), however, should be considered a *relative contraindication*—that is, a factor to be taken into account with other relative contraindications. Other relative contraindications include portal vein thrombosis, HIV infection, preexisting renal disease not associated with liver disease (which may prompt consideration of combined liver and kidney transplantation), intrahepatic or biliary sepsis, severe hypoxemia (PO_2 <50 mmHg) resulting from right-to-left intrapulmonary shunts, portopulmonary hypertension with high mean pulmonary artery pressures (>35 mmHg), previous extensive hepatobiliary surgery, any uncontrolled serious psychiatric disorder, and lack of sufficient social supports. Any one of these relative contraindications is insufficient in and of itself to preclude transplantation. For example, the problem of portal vein thrombosis can be overcome by constructing a graft from the donor liver portal vein to the recipient's superior mesenteric vein. Now that highly active antiretroviral therapy has dramatically improved the survival of persons with HIV infection **(Chap. 189)**, and because end-stage liver disease caused by chronic hepatitis C and B has emerged as a serious source of morbidity and mortality in the HIV-infected population, liver transplantation has now been performed successfully in selected HIV-positive persons who have excellent control of HIV infection. Selected patients with CD4+ T cell counts >100/μL and with pharmacologic suppression of HIV viremia have undergone transplantation for end-stage liver disease. HIV-infected persons who have received liver allografts for end-stage liver disease resulting from chronic hepatitis B have experienced survival rates compared to those of HIV-negative persons undergoing transplantation for the same indication. In contrast, recurrent hepatitis C virus (HCV) in the allograft has limited long-term success in persons with HCV-related end-stage liver disease.

Table 310-2 Contraindications to Liver Transplantation

Absolute	Relative
Uncontrolled extrahepatobiliary infection	Age <70
Active, untreated sepsis	Prior extensive hepatobiliary surgery
Uncorrectable, life-limiting congenital anomalies	Portal vein thrombosis
Active substance or alcohol abuse	Renal failure not attributable to liver disease
Advanced cardiopulmonary disease	Previous extrahepatic malignancy (not including nonmelanoma skin cancer)
Extrahepatobiliary malignancy (not including nonmelanoma skin cancer)	Severe obesity
Metastatic malignancy to the liver	Severe malnutrition/wasting
Cholangiocarcinoma	Medical noncompliance
AIDS	HIV seropositivity with failure to control HIV viremia or CD4 <100/μL
Life-threatening systemic diseases	Intrahepatic sepsis
	Severe hypoxemia secondary to right-to-left intrapulmonary shunts (PO_2<50 mmHg)
	Severe pulmonary hypertension (mean pulmonary artery pressure >35 mmHg)

Uncontrolled psychiatric disorder

TECHNICAL CONSIDERATIONS

Cadaver Donor Selection

Cadaver donor livers for transplantation are procured primarily from victims of head trauma. Organs from brain-dead donors up to age 60 are acceptable if the following criteria are met: hemodynamic stability, adequate oxygenation, absence of bacterial or fungal infection, absence of abdominal trauma, absence of hepatic dysfunction, and serologic exclusion of hepatitis B (HBV) and C viruses and HIV. Occasionally, organs from donors with hepatitis B and C are used (e.g., for recipients with prior hepatitis B and C, respectively). Organs from donors with antibodies to hepatitis B core antigen (anti-HBc) can also be used when the need is especially urgent, and recipients of these organs are treated prophylactically with HBIg and other antiviral drugs. Cardiovascular and respiratory functions are maintained artificially until the liver can be removed. Transplantation of organs procured from deceased donors who have succumbed to cardiac death can be performed successfully under selected circumstances, when ischemic time is minimized and liver histology preserved. Compatibility in ABO blood group and organ size between donor and recipient are important considerations in donor selection; however, ABO-incompatible, split liver, or reduced-donor-organ transplants can be performed in emergencies or marked donor scarcity. Tissue typing for human leukocyte antigen (HLA) matching is not required, and preformed cytotoxic HLA antibodies do not preclude liver transplantation. Following perfusion with cold electrolyte solution, the donor liver is removed and packed in ice. The use of University of Wisconsin (UW) solution, rich in lactobionate and raffinose, has permitted the extension of cold ischemic time up to 20 hours; however, 12 hours may be a more reasonable limit. Improved techniques for harvesting multiple organs from the same donor have increased the availability of donor livers, but the availability of donor livers is far outstripped by the demand. Currently in the United States, all donor livers are distributed through a nationwide organ-sharing network [United Network for Organ Sharing (UNOS)] designed to allocate available organs based on regional considerations and recipient acuity. Recipients who have the highest disease severity generally have the highest priority, but allocation strategies that balance highest urgency against best outcomes continue to evolve to distribute cadaver organs most effectively. Allocation based on the Child-Turcotte-Pugh (CTP) score, which uses five clinical variables (encephalopathy stage, ascites, bilirubin, albumin, and prothrombin time) and waiting time, has been replaced by allocation based on urgency alone, calculated by the Model for End-Stage Liver Disease (MELD) score. The MELD score is based on a mathematical model that includes bilirubin, creatinine, and prothrombin time expressed as international normalized ratio (INR) **(Table 310-3)**. Neither waiting time (except as a tie breaker between two potential recipients with the same MELD scores) nor posttransplantation outcome is taken into account, but the MELD score has been shown to reduce waiting list mortality, to reduce waiting time prior to transplantation, to be the best predictor of pretransplantation mortality, to satisfy the prevailing view that medical need should be the decisive determinant, and to eliminate both the subjectivity inherent in the CTP scoring system (presence and degree of ascites and hepatic encephalopathy) and the differences in waiting times among different regions of the country. Recent data indicate that liver recipients with MELD scores <15 experienced higher posttransplantation mortality rates than similarly classified patients who remained on the wait list. This observation has led to the modification of UNOS policy to allocate donor organs to candidates with MELD scores exceeding 15 within the local or regional procurement organization before offering the organ to local patients whose scores are <15. In addition, serum sodium, another important predictor of survival in liver transplantation candidates, is taken into consideration in allocating donor livers.

Table 310-3 United Network for Organ Sharing (Unos) Liver Transplantation Waiting List Criteria

Status 1	Fulminant hepatic failure (including primary graft nonfunction and hepatic artery thrombosis within 7 days after transplantation as well as acute decompensated Wilson's disease)[a]
The Model for End-Stage Liver Disease (MELD) score, on a continuous scale,[b] determines allocation of the remainder of donor organs. This model is based on the following calculation:	
3.78 x $\log_e$ bilirubin (mg/100 mL) + 11.2 x $\log_e$ international normalized ratio (INR) + 9.57 x $\log_e$ creatinine (mg/100 mL) + 6.43 (x 0 for alcoholic and cholestatic liver disease, x 1 for all other types of liver disease).[c,d,e]	
Online calculators to determine MELD scores are available, such as the following: http://optn.transplant.hrsa.gov/resources/professionalresources.asp?index=9.	

[a] For children <18 years of age, Status 1 includes acute or chronic liver failure plus hospitalization in

an intensive care unit or inborn errors of metabolism. Status 1 is retained for those persons with fulminant hepatic failure and supersedes the MELD score.

b The MELD scale is continuous, with 34 levels ranging between 6 and 40. Donor organs usually do not become available unless the MELD score exceeds 20.

c Patients with stage T2 hepatocellular carcinoma receive 22 disease-specific points. An α-fetoprotein level = 500 ng/mL is considered as Stage I hepatocellular carcinoma even without evidence for a tumor on imaging.

d Creatinine is included because renal function is a validated predictor of survival in patients with liver disease. For adults undergoing dialysis twice a week, the creatinine in the equation is set to 4 mg/100 mL.

e For children <18 years of age, the Pediatric End-Stage Liver Disease (PELD) scale is used. This scale is based on albumin, bilirubin, INR, growth failure, and age. Status 1 is retained.

The highest priority (status 1) continues to be reserved for patients with fulminant hepatic failure or primary graft nonfunction. Because candidates for liver transplantation who have HCC may not be sufficiently decompensated to compete for donor organs based on urgency criteria alone, and because protracted waiting for cadaver donor organs often results in tumor growth beyond acceptable limits for transplantation, such patients are assigned disease-specific MELD points (Table 310-3).

Living Donor Transplantation

Occasionally, especially for liver transplantation in children, one cadaver organ can be split between two recipients (one adult and one child). A more viable alternative, transplantation of the right lobe of the liver from a healthy adult donor into an adult recipient, has gained increased popularity. Living donor transplantation of the left lobe (left lateral segment), introduced in the early 1990s to alleviate the extreme shortage of donor organs for small children, accounts currently for approximately one-third of all liver transplantation procedures in children. Driven by the shortage of cadaver organs, living donor transplantation involving the more sizable right lobe is being considered with increasing frequency in adults; however, living donor liver transplantation cannot be expected to solve the donor organ shortage; 219 such procedures were done in 2009, representing only about 4% of all liver transplant operations done in the United States.

Living donor transplantation can reduce waiting time and cold-ischemia time; is done under elective, rather than emergency, circumstances; and may be lifesaving in recipients who cannot afford to wait for a cadaver donor. The downside, of course, is the risk to the healthy donor (a mean of 10 weeks of medical disability; biliary complications in ~5%; postoperative complications such as wound infection, small-bowel obstruction, and incisional hernias in 9–19%; and even, in 0.2–0.4%, death) as well as the increased frequency of biliary (15–32%) and vascular (10%) complications in the recipient. Potential donors must participate voluntarily without coercion, and transplantation teams should go to great lengths to exclude subtle coercive or inappropriate psychological factors as well as outline carefully to both donor and recipient the potential benefits and risks of the procedure. Donors for the procedure should be 18–60 years old; have a compatible blood type with the recipient; have no chronic medical problems or history of major abdominal surgery; be related genetically or emotionally to the recipient; and pass an exhaustive series of clinical, biochemical, and serologic evaluations to unearth disqualifying medical disorders. The recipient should meet the same UNOS criteria for liver transplantation as recipients of a cadaver donor allograft. Comprehensive outcome data on adult-to-adult living donor liver transplantation are being collected (www.nih-a2all.org).

Surgical Technique

Removal of the recipient's native liver is technically difficult, particularly in the presence of portal hypertension with its associated collateral circulation and extensive varices and especially in the presence of scarring from previous abdominal operations. The combination of portal hypertension and coagulopathy (elevated prothrombin time and thrombocytopenia) may translate into large blood product transfusion requirements. After the portal vein and infrahepatic and suprahepatic inferior vena cavae are dissected, the hepatic artery and common bile duct are dissected. Then the native liver is removed and the donor organ inserted. During the anhepatic phase, coagulopathy, hypoglycemia, hypocalcemia, and hypothermia are encountered and must be managed by the anesthesiology team. Caval, portal vein, hepatic artery, and bile duct anastomoses are performed in succession, the last by end-to-end suturing of the donor and recipient common bile ducts or by choledochojejunostomy to a Roux-en-Y loop if the recipient common bile duct cannot be used for reconstruction (e.g., in sclerosing cholangitis). A typical transplant operation lasts 8 hours, with a range of 6–18 hours. Because of excessive bleeding, large volumes of blood, blood products, and volume expanders may be required during surgery; however, blood requirements have fallen sharply with improvements in surgical technique and experience.

As noted above, emerging alternatives to orthotopic liver transplantation include split-liver grafts, in which one donor organ is divided and inserted into two recipients; and living donor procedures, in which part of the left (for children), the left (for children or small adults), or the right (for adults) lobe of

the liver is harvested from a living donor for transplantation into the recipient. In the adult procedure, once the right lobe is removed from the donor, the donor right hepatic vein is anastomosed to the recipient right hepatic vein remnant, followed by donor-to-recipient anastomoses of the portal vein and then the hepatic artery. Finally, the biliary anastomosis is performed, duct-to-duct if practical or via Roux-en-Y anastomosis. Heterotopic liver transplantation, in which the donor liver is inserted without removal of the native liver, has met with very limited success and acceptance, except in a very small number of centers. In attempts to support desperately ill patients until a suitable donor organ can be identified, several transplantation centers are studying extracorporeal perfusion with bioartificial liver cartridges constructed from hepatocytes bound to hollow fiber systems and used as temporary hepatic-assist devices, but their efficacy remains to be established. Areas of research with the potential to overcome the shortage of donor organs include hepatocyte transplantation and xenotransplantation with genetically modified organs of nonhuman origin (e.g., swine).

POSTOPERATIVE COURSE AND MANAGEMENT

Immunosuppressive Therapy

The introduction in 1980 of cyclosporine as an immunosuppressive agent contributed substantially to the improvement in survival after liver transplantation. Cyclosporine, a calcineurin inhibitor (CNI), blocks early activation of T cells and is specific for T cell functions that result from the interaction of the T cell with its receptor and that involve the calcium-dependent signal transduction pathway. As a result, the activity of cyclosporine leads to inhibition of lymphokine gene activation, blocking interleukins 2, 3, and 4, tumor necrosis factor α, and other lymphokines. Cyclosporine also inhibits B cell functions. This process occurs without affecting rapidly dividing cells in the bone marrow, which may account for the reduced frequency of posttransplantation systemic infections. The most common and important side effect of cyclosporine therapy is nephrotoxicity. Cyclosporine causes dose-dependent renal tubular injury and direct renal artery vasospasm. Following renal function is therefore important in monitoring cyclosporine therapy, perhaps even a more reliable indicator than blood levels of the drug. Nephrotoxicity is reversible and can be managed by dose reduction. Other adverse effects of cyclosporine therapy include hypertension, hyperkalemia, tremor, hirsutism, glucose intolerance, and gingival hyperplasia.

Tacrolimus, a macrolide lactone antibiotic isolated from a Japanese soil fungus, *Streptomyces tsukubaensis*, has the same mechanism of action as cyclosporine but is 10–100 times more potent. Initially applied as "rescue" therapy for patients in whom rejection occurred despite the use of cyclosporine, tacrolimus was shown to be associated with a reduced frequency of acute, refractory, and chronic rejection. Although patient and graft survival are the same with these two drugs, the advantage of tacrolimus in minimizing episodes of rejection, reducing the need for additional glucocorticoid doses, and reducing the likelihood of bacterial and cytomegalovirus (CMV) infection has simplified the management of patients undergoing liver transplantation. In addition, the oral absorption of tacrolimus is more predictable than that of cyclosporine, especially during the early postoperative period when T-tube drainage interferes with the enterohepatic circulation of cyclosporine. As a result, in most transplantation centers tacrolimus has now supplanted cyclosporine for primary immunosuppression, and many centers rely on oral rather than IV administration from the outset. For transplantation centers that prefer cyclosporine, a better-absorbed microemulsion preparation is now available.

Although more potent than cyclosporine, tacrolimus is also more toxic and more likely to be discontinued for adverse events. The toxicity of tacrolimus is similar to that of cyclosporine; nephrotoxicity and neurotoxicity are the most commonly encountered adverse effects, and neurotoxicity (tremor, seizures, hallucinations, psychoses, coma) is more likely and more severe in tacrolimus-treated patients. Both drugs can cause diabetes mellitus, but tacrolimus does not cause hirsutism or gingival hyperplasia. Because of overlapping toxicity between cyclosporine and tacrolimus, especially nephrotoxicity, and because tacrolimus reduces cyclosporine clearance, these two drugs should not be used together. Because 99% of tacrolimus is metabolized by the liver, hepatic dysfunction reduces its clearance; in primary graft nonfunction (when, for technical reasons or because of ischemic damage prior to its insertion, the allograft is defective and does not function normally from the outset), tacrolimus doses have to be reduced substantially, especially in children. Both cyclosporine and tacrolimus are metabolized by the cytochrome P450 IIIA system, and, therefore, drugs that induce cytochrome P450 (e.g., phenytoin, phenobarbital, carbamazepine, rifampin) reduce available levels of cyclosporine and tacrolimus; drugs that inhibit cytochrome P450 (e.g., erythromycin, fluconazole, ketoconazole, clotrimazole, itraconazole, verapamil, diltiazem, nicardipine, cimetidine, danazol, metoclopramide, bromocriptine, and the HIVprotease inhibitor ritonavir) increase cyclosporine and tacrolimus blood levels. Indeed, itraconazole is used occasionally to help boost tacrolimus levels. Like azathioprine, cyclosporine and tacrolimus appear to be associated with a risk of lymphoproliferative malignancies (see below), which may occur earlier after cyclosporine or tacrolimus than after azathioprine therapy. Because of these side effects, combinations of cyclosporine or tacrolimus with prednisone and an antimetabolite (azathioprine or mycophenolic acid, see below)—all at reduced doses—are preferable regimens for immunosuppressive therapy.

Mycophenolic acid, a nonnucleoside purine metabolism inhibitor derived as a fermentation product from several *Penicillium* species, is another immunosuppressive drug being used increasingly for patients undergoing liver transplantation. Mycophenolate has been shown to be better than azathioprine, when used with other standard immunosuppressive drugs, in preventing rejection after renal transplantation and has been adopted widely as well for use in liver transplantation. The most common adverse effects of mycophenolate are bone marrow suppression and gastrointestinal complaints.

In patients with pretransplantation renal dysfunction or renal deterioration that occurs intraoperatively or immediately postoperatively, tacrolimus or cyclosporine therapy may not be practical; under these circumstances, induction or maintenance of immunosuppression with antithymocyte globulin (ATG, thymoglobulin) or monoclonal antibodies to T cells, OKT3, may be appropriate. Therapy with these agents has been especially effective in reversing acute rejection in the posttransplant period and is the standard treatment for acute rejection that fails to respond to methylprednisolone boluses. Available data support the use of thymoglobulin induction to delay CNI use and its attendant nephrotoxicity. IV infusions of thymoglobulin may be complicated by fever and chills, which can be ameliorated by premedication with antipyretics and a low dose of glucocorticoids. Infusions of OKT3 may be complicated by fever, chills, and diarrhea, or by pulmonary edema, which can be fatal. Because OKT3 is such a potent immunosuppressive agent, its use is also more likely to be complicated by opportunistic infection or lymphoproliferative disorders; therefore, because of the availability of alternative immunosuppressive drugs, OKT3 is used less often nowadays.

Rapamycin, an inhibitor of later events in T cell activation, is approved for use in kidney transplantation but is not approved for use in liver transplant recipients because of the reported association with an increased frequency of hepatic artery thrombosis in the first month posttransplantation. In patients with CNI-related nephrotoxicity, conversion to rapamycin has been demonstrated to be effective in preventing rejection with accompanying improvements in renal function. Because of its profound antiproliferative effects, rapamycin has also been suggested to be a useful immunosuppressive agent in patients with a prior or current history of malignancy, such as HCC. Side effects include hyperlipidemia, peripheral edema, oral ulcers, and interstitial pneumonitis.

The most important principle of immunosuppression is that the ideal approach strikes a balance between immunosuppression and immunologic competence. In general, given sufficient immunosuppression, acute liver allograft rejection is nearly always reversible. On one hand, incompletely treated acute rejection predisposes to the development of chronic rejection, which can threaten graft survival. On the other hand, if the cumulative dose of immunosuppressive therapy is too large, the patient may succumb to opportunistic infection. In hepatitis C, pulse glucocorticoids or OKT3 use accelerate recurrent allograft hepatitis. Further complicating matters, acute rejection can be difficult to distinguish histologically from recurrent hepatitis C. Therefore, immunosuppressive drugs must be used judiciously, with strict attention to the infectious consequences of such therapy and careful confirmation of the diagnosis of acute rejection. In this vein, efforts have been made to minimize the use of glucocorticoids, a mainstay of immunosuppressive regimens, and steroid-free immunosuppression can be achieved in some instances. Patients who undergo liver transplantation for autoimmune diseases such as primary biliary cirrhosis, autoimmune hepatitis, and primary sclerosing cholangitis are less likely to achieve freedom from glucocorticoids.

Postoperative Complications

Complications of liver transplantation can be divided into nonhepatic and hepatic categories **(Tables 310-4 and 310-5)**. In addition, both immediate postoperative and late complications are encountered. As a rule, patients who undergo liver transplantation have been chronically ill for protracted periods and may be malnourished and wasted. The impact of such chronic illness and the multisystem failure that accompanies liver failure continue to require attention in the postoperative period. Because of the massive fluid losses and fluid shifts that occur during the operation, patients may remain fluid-overloaded during the immediate postoperative period, straining cardiovascular reserve; this effect can be amplified in the face of transient renal dysfunction and pulmonary capillary vascular permeability. Continuous monitoring of cardiovascular and pulmonary function, measures to maintain the integrity of the intravascular compartment and to treat extravascular volume overload, and scrupulous attention to potential sources and sites of infection are of paramount importance. Cardiovascular instability may also result from the electrolyte imbalance that may accompany reperfusion of the donor liver as well as from restoration of systemic vascular resistance following implantation. Pulmonary function may be compromised further by paralysis of the right hemidiaphragm associated with phrenic nerve injury. The hyperdynamic state with increased cardiac output that is characteristic of patients with liver failure reverses rapidly after successful liver transplantation.

Table 310-4 Nonhepatic Complications of Liver Transplantation

Fluid overload	
Cardiovascular instability	Arrhythmias

	Congestive heart failure Cardiomyopathy
Pulmonary compromise	Pneumonia Pulmonary capillary vascular permeability Fluid overload
Renal dysfunction	Prerenal azotemia Hypoperfusion injury (acute tubular necrosis) Drug nephrotoxicity ↓ Renal blood flow secondary to ↑ intraabdominal pressure
Hematologic	Anemia 2° to gastrointestinal and/or intraabdominal bleeding Hemolytic anemia, aplastic anemia Thrombocytopenia
Infection	Bacterial: early, common postoperative infections Fungal/parasitic: late, opportunistic infections Viral: late, opportunistic infections, recurrent hepatitis
Neuropsychiatric	Seizures Metabolic encephalopathy Depression Difficult psychosocial adjustment
Diseases of donor	Infectious Malignant
Malignancy	B cell lymphoma (posttransplantation lymphoproliferative disorders) De novo neoplasms (particularly squamous cell skin carcinoma)

Table 310-5 Hepatic Complications of Liver Transplantation

Hepatic Dysfunction Common after Major Surgery	
Prehepatic	Pigment load Hemolysis Blood collections (hematomas, abdominal collections)
Intrahepatic	
Early	Hepatotoxic drugs and anesthesia Hypoperfusion (hypotension, shock, sepsis) Benign postoperative cholestasis
Late	Transfusion-associated hepatitis Exacerbation of primary hepatic disease
Posthepatic	Biliary obstruction ↓ Renal clearance of conjugated bilirubin (renal dysfunction)
Hepatic Dysfunction Unique to Liver Transplantation	
Primary graft nonfunction	
Vascular compromise	Portal vein obstruction Hepatic artery thrombosis Anastomotic leak with intraabdominal bleeding
Bile duct disorder	Stenosis, obstruction, leak

Rejection	
Recurrent primary hepatic disease	

Other immediate management issues include renal dysfunction. Prerenal azotemia, acute kidney injury associated with hypoperfusion (acute tubular necrosis), and renal toxicity caused by antibiotics, tacrolimus, or cyclosporine are encountered frequently in the postoperative period, sometimes necessitating dialysis. Hemolytic uremic syndrome can be associated with cyclosporine, tacrolimus, or OKT3. Occasionally, postoperative intraperitoneal bleeding may be sufficient to increase intraabdominal pressure, which, in turn, may reduce renal blood flow; this effect is rapidly reversible when abdominal distention is relieved by exploratory laparotomy to identify and ligate the bleeding site and to remove intraperitoneal clot.

Anemia may also result from acute upper gastrointestinal bleeding or from transient hemolytic anemia, which may be autoimmune, especially when blood group O livers are transplanted into blood group A or B recipients. This autoimmune hemolytic anemia is mediated by donor intrahepatic lymphocytes that recognize red blood cell A or B antigens on recipient erythrocytes. Transient in nature, this process resolves once the donor liver is repopulated by recipient bone marrow–derived lymphocytes; the hemolysis can be treated by transfusing blood group O red blood cells and/or by administering higher doses of glucocorticoids. Transient thrombocytopenia is also commonly encountered. Aplastic anemia, a late occurrence, is rare but has been reported in almost 30% of patients who underwent liver transplantation for acute, severe hepatitis of unknown cause.

Bacterial, fungal, or viral infections are common and may be life-threatening postoperatively. Early after transplant surgery, common postoperative infections predominate—pneumonia, wound infections, infected intraabdominal collections, urinary tract infections, and IV line infections—rather than opportunistic infections; these infections may involve the biliary tree and liver as well. Beyond the first postoperative month, the toll of immunosuppression becomes evident, and opportunistic infections —CMV, herpes viruses, fungal infections (*Aspergillus*, *Candida*, cryptococcal disease), mycobacterial infections, parasitic infections (*Pneumocystis*, *Toxoplasma*), bacterial infections (*Nocardia*, *Legionella*, and *Listeria*)—predominate. Rarely, early infections represent those transmitted with the donor liver, either infections present in the donor or infections acquired during procurement processing. De novo viral hepatitis infections acquired from the donor organ or, almost unheard of nowadays, from transfused blood products occur after typical incubation periods for these agents (well beyond the first month). Obviously, infections in an immunosuppressed host demand early recognition and prompt management; prophylactic antibiotic therapy is administered routinely in the immediate postoperative period. Use of sulfamethoxazole with trimethoprim reduces the incidence of postoperative *Pneumocystis carinii* pneumonia. Antiviral prophylaxis for CMV with ganciclovir should be administered in patients at high risk (e.g., when a CMV-seropositive donor organ is implanted into a CMV-seronegative recipient).

Neuropsychiatric complications include seizures (commonly associated with cyclosporine and tacrolimus toxicity), metabolic encephalopathy, depression, and difficult psychosocial adjustment. Rarely, diseases are transmitted by the allograft from the donor to the recipient. In addition to viral and bacterial infections, malignancies of donor origin have occurred. Posttransplantation lymphoproliferative disorders, especially B cell lymphoma, are a recognized complication associated with immunosuppressive drugs such as azathioprine, tacrolimus, and cyclosporine (see above). Epstein-Barr virus has been shown to play a contributory role in some of these tumors, which may regress when immunosuppressive therapy is reduced. De novo neoplasms appear at increased frequency after liver transplantation, particularly squamous cell carcinomas of the skin. Routine screening should be performed.

Long-term complications after liver transplantation attributable primarily to immunosuppressive medications include diabetes mellitus (associated with glucocorticoids) as well as hypertension, hyperlipidemia, and chronic renal insufficiency (associated with cyclosporine and tacrolimus). Monitoring and treating these disorders is a routine component of posttransplantation care; in some cases, they respond to changes in immunosuppressive regimen, while in others, specific treatment of the disorder is introduced.

Hepatic Complications

Hepatic dysfunction after liver transplantation is similar to the hepatic complications encountered after major abdominal and cardiothoracic surgery; however, in addition, hepatic complications include primary graft failure, vascular compromise, failure or stricture of the biliary anastomoses, and rejection. As in nontransplant surgery, postoperative jaundice may result from prehepatic, intrahepatic, and posthepatic sources. *Prehepatic* sources represent the massive hemoglobin pigment load from transfusions, hemolysis, hematomas, ecchymoses, and other collections of blood. *Early intrahepatic* liver injury includes effects of hepatotoxic drugs and anesthesia; hypoperfusion injury associated with hypotension, sepsis, and shock; and benign postoperative cholestasis. *Late intrahepatic* sources of liver injury include posttransfusion hepatitis and exacerbation of primary disease. *Posthepatic* sources

of hepatic dysfunction include biliary obstruction and reduced renal clearance of conjugated bilirubin. Hepatic complications unique to liver transplantation include primary graft failure associated with ischemic injury to the organ during harvesting; vascular compromise associated with thrombosis or stenosis of the portal vein or hepatic artery anastomoses; vascular anastomotic leak; stenosis, obstruction, or leakage of the anastomosed common bile duct; recurrence of primary hepatic disorder (see below); and rejection.

Transplant Rejection

Despite the use of immunosuppressive drugs, rejection of the transplanted liver still occurs in a proportion of patients, beginning 1–2 weeks after surgery. Clinical signs suggesting rejection are fever, right upper quadrant pain, and reduced bile pigment and volume. Leukocytosis may occur, but the most reliable indicators are increases in serum bilirubin and aminotransferase levels. Because these tests lack specificity, distinguishing among rejection, biliary obstruction, primary graft nonfunction, vascular compromise, viral hepatitis, CMV infection, drug hepatotoxicity, and recurrent primary disease may be difficult. Radiographic visualization of the biliary tree and/or percutaneous liver biopsy often help to establish the correct diagnosis. Morphologic features of acute rejection include a mixed portal cellular infiltrate, bile duct injury, and/or endothelial inflammation ("endothelialitis"); some of these findings are reminiscent of graft-versus-host disease, primary biliary cirrhosis, or recurrent allograft hepatitis C. As soon as transplant rejection is suspected, treatment consists of IV methylprednisolone in repeated boluses; if this fails to abort rejection, many centers use thymoglobulin or OKT3. Caution should be exercised when managing acute rejection with pulse glucocorticoids or OKT3 in patients with HCV infection, because of the high risk of triggering recurrent allograft hepatitis C.

Chronic rejection is a relatively rare outcome that can follow repeated bouts of acute rejection or that occurs unrelated to preceding rejection episodes. Morphologically, chronic rejection is characterized by progressive cholestasis, focal parenchymal necrosis, mononuclear infiltration, vascular lesions (intimal fibrosis, subintimal foam cells, fibrinoid necrosis), and fibrosis. This process may be reflected as ductopenia—the vanishing bile duct syndrome. Reversibility of chronic rejection is limited; in patients with therapy-resistant chronic rejection, retransplantation has yielded encouraging results.

OUTCOME

Survival

The survival rate for patients undergoing liver transplantation has improved steadily since 1983. One-year survival rates have increased from ~70% in the early 1980s to 85–90% from 2003 to 2009. Currently the 5-year survival rate exceeds 60%. An important observation is the relationship between clinical status before transplantation and outcome. For patients who undergo liver transplantation when their level of compensation is high (e.g., still working or only partially disabled), a 1-year survival rate of >85% is common. For those whose level of decompensation mandates continuous in-hospital care prior to transplantation, the 1-year survival rate is about 70%, while for those who are so decompensated that they require life support in an intensive care unit, the 1-year survival rate is ~50%. Since UNOS's adoption in 2002 of the MELD system for organ allocation, posttransplantation survival has been found to be affected adversely for candidates with MELD scores >25, considered high disease severity. Thus, irrespective of allocation scheme, high disease severity pretransplantation corresponds to diminished posttransplantation survival. Another important distinction in survival has been drawn between high- and low-risk patient categories. For patients who do not fit any "high-risk" designations, 1-year and 5-year survival rates of 85 and 80%, respectively, have been recorded. In contrast, among patients in high-risk categories—cancer, fulminant hepatitis, age >65, concurrent renal failure, respirator dependence, portal vein thrombosis, and history of a portacaval shunt or multiple right upper quadrant operations—survival statistics fall into the range of 60% at 1 year and 35% at 5 years. Survival after retransplantation for primary graft nonfunction is ~50%. Causes of failure of liver transplantation vary with time. Failures within the first 3 months result primarily from technical complications, postoperative infections, and hemorrhage. Transplant failures after the first 3 months are more likely to result from infection, rejection, or recurrent disease (such as malignancy or viral hepatitis).

Recurrence of Primary Disease

Features of autoimmune hepatitis, primary sclerosing cholangitis, and primary biliary cirrhosis overlap with those of rejection or posttransplantation bile duct injury. Whether autoimmune hepatitis and sclerosing choangitis recur after liver transplantation is controversial; data supporting recurrent autoimmune hepatitis (in up to one-third of patients in some series) are more convincing than those supporting recurrent sclerosing cholangitis. Similarly, reports of recurrent primary biliary cirrhosis after liver transplantation have appeared; however, the histologic features of primary biliary cirrhosis and chronic rejection are virtually indistinguishable and occur as frequently in patients with primary biliary cirrhosis as in patients undergoing transplantation for other reasons. The presence of a florid inflammatory bile duct lesion is highly suggestive of the recurrence of primary biliary cirrhosis, but even this lesion can be observed in acute rejection. Hereditary disorders such as Wilson's disease and α_1 antitrypsin deficiency have not recurred after liver transplantation; however, recurrence of disordered

iron metabolism has been observed in some patients with hemochromatosis. Hepatic vein thrombosis (Budd-Chiari syndrome) may recur; this can be minimized by treating underlying myeloproliferative disorders and by anticoagulation. Because cholangiocarcinoma recurs almost invariably, few centers now offer transplantation to such patients; however, a few highly selected patients with operatively confirmed stage I or II cholangiocarcinoma who undergo liver transplantation combined with neoadjuvant chemoradiation may experience excellent outcomes. In patients with intrahepatic hepatocellular carcinoma who meet criteria for transplantation, 1- and 5-year survivals are similar to those observed in patients undergoing liver transplantation for nonmalignant disease. Finally, metabolic disorders such as nonalcoholic steatohepatitis recur frequently, especially if the underlying metabolic predisposition is not altered. The metabolic syndrome occurs commonly after liver transplantation as a result of recurrent nonalcoholic fatty liver, immunosuppressive medications, and/or, in patients with hepatitis C related to the impact of HCV infection on insulin resistance, diabetes, and fatty liver.

Hepatitis A can recur after transplantation for fulminant hepatitis A, but such acute reinfection has no serious clinical sequelae. In fulminant hepatitis B, recurrence is not the rule; however, in the absence of any prophylactic measures, hepatitis B usually recurs after transplantation for end-stage chronic hepatitis B. Before the introduction of prophylactic antiviral therapy, immunosuppressive therapy sufficient to prevent allograft rejection led inevitably to marked increases in hepatitis B viremia, regardless of pretransplantation levels. Overall graft and patient survival were poor, and some patients experienced a rapid recapitulation of severe injury—severe chronic hepatitis or even fulminant hepatitis—after transplantation. Also recognized in the era before availability of antiviral regimens was *fibrosing cholestatic hepatitis*, rapidly progressive liver injury associated with marked hyperbilirubinemia, substantial prolongation of the prothrombin time (both out of proportion to relatively modest elevations of aminotransferase activity), and rapidly progressive liver failure. This lesion has been suggested to represent a "choking off" of the hepatocyte by an overwhelming density of HBV proteins. Complications such as sepsis and pancreatitis were also observed more frequently in patients undergoing liver transplantation for hepatitis B prior to the introduction of antiviral therapy. The introduction of long-term prophylaxis with HBIg revolutionized liver transplantation for chronic hepatitis B. Preoperative hepatitis B vaccination, preoperative or postoperative interferon (IFN) therapy, or short-term (≤2 months) HBIg prophylaxis has not been shown to be effective, but a retrospective analysis of data from several hundred European patients followed for 3 years after transplantation has shown that long-term (≥6 months) prophylaxis with HBIg is associated with a lowering of the risk of HBV reinfection from ~75% to 35% and a reduction in mortality from ~50% to 20%.

As a result of long-term HBIg use following liver transplantation for chronic hepatitis B, similar improvements in outcome have been observed in the United States, with 1-year survival rates between 75% and 90%. Currently, with HBIg prophylaxis, the outcome of liver transplantation for chronic hepatitis B is indistinguishable from that for chronic liver disease unassociated with chronic hepatitis B; essentially, medical concerns regarding liver transplantation for chronic hepatitis B have been eliminated. Passive immunoprophylaxis with HBIg is begun during the anhepatic stage of surgery, repeated daily for the first 6 postoperative days, then continued with infusions that are given either at regular intervals of 4–6 weeks or, alternatively, when anti-hepatitis B surface (HBs) levels fall below a threshold of 100 mIU/mL. The current approach in most centers is to continue HBIg indefinitely, which can add approximately $20,000 per year to the cost of care; some centers are evaluating regimens that shift to less frequent administration or to IM administration in the late posttransplantation period or, in low-risk patients, maintenance with antiviral therapy (see below) alone. Still, "breakthrough" HBV infection occasionally occurs.

Further improving the outcome of liver transplantation for chronic hepatitis B is the current availability of such antiviral drugs as lamivudine, adefovir, entecavir, and tenofovir disoproxil fumarate **(Chap. 306)**. When these drugs are administered to patients with decompensated liver disease, a proportion improve sufficiently to postpone imminent liver transplantation. In addition, lamivudine can be used to prevent recurrence of HBV infection when administered *prior* to transplantation; to treat hepatitis B that recurs *after* transplantation, including in patients who break through HBIg prophylaxis; and to reverse the course of otherwise fatal fibrosing cholestatic hepatitis. Clinical trials have shown that lamivudine antiviral therapy reduces the level of HBV replication substantially, sometimes even resulting in clearance of hepatitis B surface antigen (HBsAg); reduces alanine aminotransferase (ALT) levels; and improves histologic features of necrosis and inflammation. Long-term use of lamivudine is safe and effective, but after several months a proportion of patients become resistant to lamivudine, resulting from YMDD (tyrosine-methionine-aspartate-aspartate) mutations in the HBV polymerase motif **(Chap. 306)**. In approximately one-half of such resistant patients, hepatic deterioration may ensue. Fortunately, adefovir or tenofovir disoproxil fumarate are available as well and can be used to treat lamivudine-associated YMDD variants, effectively "rescuing" patients experiencing hepatic decompensation after lamivudine breakthrough. Currently, most liver transplantation centers combine HBIg plus lamivudine, adefovir, entecavir, or tenofovir disoproxil fumarate. Clinical trials are underway to define the optimal application of these antiviral agents in the management of patients undergoing liver transplantation for chronic hepatitis B; conceivably, in the future, combinations of oral antiviral

drugs may supplant HBIg.

Prophylactic approaches applied to patients undergoing liver transplantation for chronic hepatitis B are being used as well for patients without hepatitis B who receive organs from donors with anti-hepatitis B core (HBc). Patients who undergo liver transplantation for chronic hepatitis B plus D are less likely to experience recurrent liver injury than patients undergoing liver transplantation for hepatitis B alone; still, such co-infected patients would also be offered standard posttransplantation prophylactic therapy for hepatitis B.

Accounting for up to 40% of all liver transplantation procedures, the most common indication for liver transplantation is end-stage liver disease resulting from chronic hepatitis C. Recurrence of HCV infection after liver transplantation can be documented in almost every patient. The clinical consequences of recurrent hepatitis C are limited during the first 5 years after transplantation. Nonetheless, despite the relative clinical benignity of recurrent hepatitis C in the early years after liver transplantation, and despite the negligible impact on patient survival during these early years, histologic studies have documented the presence of moderate to severe chronic hepatitis in more than one-half of all patients and bridging fibrosis or cirrhosis in ~10%. Moreover, progression to cirrhosis within 5 years is even more common, occurring in up to two-thirds of patients if moderate hepatitis is detected in a 1-year biopsy. Not surprisingly, then, for patients undergoing transplantation for hepatitis C, allograft and patient survival are diminished substantially between 5 and 10 years after transplantation. In a proportion of patients, even during the early posttransplantation period, recurrent hepatitis C may be sufficiently severe biochemically and histologically to merit antiviral therapy. Treatment with pegylated interferon (IFN) can *suppress* HCV-associated liver injury but rarely leads to *sustained* benefit. Sustained virologic responses are the exception, and reduced tolerability is often dose-limiting. Preemptive combination antiviral therapy with pegylated IFN and the nucleoside analogue ribavirin immediately after transplantation does not appear to provide any advantage over therapy introduced after clinical hepatitis has occurred. Similarly, although IFN-based antiviral therapy is not recommended for patients with decompensated liver disease, some centers have experimented with pretransplantation antiviral therapy in an attempt to eradicate HCV replication prior to transplantation; preliminary results are promising, but IFN treatment of patients with end-stage liver disease can lead to worsening of hepatic decompensation, and HCV infection has recurred after transplantation in some of these recipients. Trials of hepatitis C immune globulin preparations to prevent recurrent hepatitis C after liver transplantation have not been successful.

A small number succumb to early HCV-associated liver injury, and a syndrome reminiscent of fibrosing cholestatic hepatitis (see above) has been observed rarely. Because patients with more episodes of rejection receive more immunosuppressive therapy, and because immunosuppressive therapy enhances HCV replication, patients with severe or multiple episodes of rejection are more likely to experience early recurrence of hepatitis C after transplantation. Both high viral levels and older donor age have been linked to recurrent HCV-induced liver disease and to earlier disease recurrence after transplantation.

Patients who undergo liver transplantation for end-stage alcoholic cirrhosis are at risk of resorting to drinking again after transplantation, a potential source of recurrent alcoholic liver injury. Currently, alcoholic liver disease is one of the more common indications for liver transplantation, accounting for 20–25% of all liver transplantation procedures, and most transplantation centers screen candidates carefully for predictors of continued abstinence. Recidivism is more likely in patients whose sobriety prior to transplantation was <6 months. For abstinent patients with alcoholic cirrhosis, liver transplantation can be undertaken successfully, with outcomes comparable to those for other categories of patients with chronic liver disease, when coordinated by a team approach that includes substance abuse counseling.

Posttransplantation Quality of Life

Full rehabilitation is achieved in the majority of patients who survive the early postoperative months and escape chronic rejection or unmanageable infection. Psychosocial maladjustment interferes with medical compliance in a small number of patients, but most manage to adhere to immunosuppressive regimens, which must be continued indefinitely. In one study, 85% of patients who survived their transplant operations returned to gainful activities. In fact, some women have conceived and carried pregnancies to term after transplantation without demonstrable injury to their infants.

FURTHER READINGS

Angus PW et al: A randomized study of adefovir dipivoxil in place of HBIG in combination with lamivudine as post-liver transplantation hepatitis B prophylaxis. Hepatology 48:1460, 2008[PMID: 18925641] [Full Text]

Berenguer M: Systematic review of the treatment of established recurrent hepatitis C with pegylated interferon in combination with ribavirin. J Hepatol 49:274, 2008[PMID: 18571272] [Full Text]

Berg CL et al: Improvement in survival associated with adult-to-adult living donor liver transplantation.

Gastroenterology 133:1806, 2007[PMID: 18054553] [Full Text]

Cholongitas E et al: A systematic review of the performance of the model for end-stage liver disease (MELD) in the setting of liver transplantation. Liver Transpl 12:1049, 2006[PMID: 16799946] [Full Text]

Dharancy S et al: Mycophenolate mofetil monotherapy for severe side effects of calcineurin inhibitors following liver transplantation. Am J Transplant 40:2985, 2008

Fishman JA: Infection in solid-organ transplant recipients. N Engl J Med 357:2601, 2007[PMID: 18094380] [Full Text]

Gallegos-Orozco JF, Vargas HE: Liver transplantation: From Child to MELD. Med Clin North Am 93:931, 2009[PMID: 19577123] [Full Text]

Gane EJ et al: Lamivudine plus low-dose hepatitis B immunoglobulin to prevent recurrent hepatitis B following liver transplantation. Gastroenterology 132:931, 2007[PMID: 17383422] [Full Text]

Kim WR et al: Hyponatremia and mortality among patients on the liver-transplant waiting list. N Engl J Med 359:1018, 2008[PMID: 18768945] [Full Text]

Murray KF, Carithers RL JR: AASLD practice guidelines: Evaluation of the patient for liver transplantation. Hepatology 41:1407, 2005[PMID: 15880505] [Full Text]

O'Leary JG et al: Indications for liver transplantation. Gastroenterology 134:1764, 2008[PMID: 21660345] [Full Text]

Pillai AA, Levitsky J: Overview of immunosuppression in liver transplantation. World J Gastroenterol 15:4225, 2009[PMID: 19750565] [Full Text]

Roland ME, Stock PG: Liver transplantation in HIV-infected recipients. Semin Liver Dis 26:273, 2006[PMID: 16850377] [Full Text]

Starzl TE, Fung JJ: Themes of liver transplantation. Hepatology 51:1869, 2010[PMID: 20235333] [Full Text]

Terrault NA, Berenguer M: Treating hepatitis C infection in liver transplant recipients. Liver Transpl 12:1192, 2006[PMID: 16868944] [Full Text]

Watt KDS, Charlton MR: Metabolic syndrome and liver transplantation: A review and guide to management. J Hepatol 53:199, 2010[PMID: 20451282] [Full Text]

Webb K et al: Transplantation for alcoholic liver disease: Report of a consensus meeting. Liver Transpl 12:301, 2006[PMID: 16447187] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PHYSIOLOGY OF BILE PRODUCTION AND FLOW

Bile Secretion and Composition

Bile formed in the hepatic lobules is secreted into a complex network of canaliculi, small bile ductules, and larger bile ducts that run with lymphatics and branches of the portal vein and hepatic artery in portal tracts situated between hepatic lobules. These interlobular bile ducts coalesce to form larger septal bile ducts that join to form the right and left hepatic ducts, which in turn, unite to form the common hepatic duct. The common hepatic duct is joined by the cystic duct of the gallbladder to form the common bile duct (CBD), which enters the duodenum (often after joining the main pancreatic duct) through the ampulla of Vater.

Hepatic bile is an isotonic fluid with an electrolyte composition resembling blood plasma. The electrolyte composition of gallbladder bile differs from that of hepatic bile because most of the inorganic anions, chloride and bicarbonate, have been removed by reabsorption across the gallbladder epithelium. As a result of water reabsorption, total solute concentration of bile increases from 3–4 g/dL in hepatic bile to 10–15 g/dL in gallbladder bile.

Major solute components of bile by moles percent include bile acids (80%), lecithin and traces of other phospholipids (16%), and unesterified cholesterol (4.0%). In the lithogenic state, the cholesterol value can be as high as 8–10%. Other constituents include conjugated bilirubin; proteins (all immunoglobulins, albumin, metabolites of hormones, and other proteins metabolized in the liver); electrolytes; mucus; and, often, drugs and their metabolites.

The total daily basal secretion of hepatic bile is ~500–600 mL. Many substances taken up or synthesized by the hepatocyte are secreted into the bile canaliculi. The canalicular membrane forms microvilli and is associated with microfilaments of actin, microtubules, and other contractile elements. Prior to their secretion into the bile, many substances are taken up into the hepatocyte, while others, such as phospholipids, a portion of primary bile acids, and some cholesterol are synthesized de novo in the hepatocyte. Three mechanisms are important in regulating bile flow: (1) active transport of bile acids from hepatocytes into the bile canaliculi, (2) active transport of other organic anions, and (3) cholangiocellular secretion. The last is a secretin-mediated and cyclic AMP–dependent mechanism that results in the secretion of a sodium- and bicarbonate-rich fluid into the bile ducts.

Active vectorial secretion of biliary constituents from the portal blood into the bile canaliculi is driven by a set of polarized transport systems at the basolateral (sinusoidal) and the canalicular apical plasma membrane domains of the hepatocyte. Two sinusoidal bile salt uptake systems have been cloned in humans, the Na^+/taurocholate cotransporter (NTCP, SLC10A1) and the organic anion–transporting proteins (OATPs), which also transport a large variety of non-bile salt organic anions. Several ATP-dependent canalicular transport systems, "export pumps," (ATP-binding cassette transport proteins, also known as ABC transporters) have been identified, the most important of which are: the bile salt export pump (BSEP, ABCB11); the anionic conjugate export pump (MRP2, ABCC2), which mediates the canalicular excretion of various amphiphilic conjugates formed by phase II conjugation (e.g., bilirubin mono- and diglucuronides and drugs); the multidrug export pump (MDR1, ABCB1) for hydrophobic cationic compounds; and the phospholipid export pump (MDR3, ABCB4). Two hemitransporters ABCG5/G8, functioning as a couple, constitute the canalicular cholesterol and phytosterol transporter. F1C1 (ATP8B1) is an aminophospholipid transferase ("flippase") essential for maintaining the lipid asymmetry of the canalicular membrane. The canalicular membrane also contains ATP-independent transport systems such as the Cl/HCO_3 anion exchanger isoform 2 (AE2, SLC4A2) for canalicular bicarbonate secretion. For most of these transporters, genetic defects have been identified that are associated with various forms of cholestasis or defects of biliary excretion. F1C1 is defective in progressive familial intrahepatic cholestasis type 1 (PFIC1) and benign recurrent intrahepatic cholestasis type 1 (BRIC1) and results in ablation of all other ATP-dependent transporter functions. BSEP is defective in PFIC2 and BRIC2. Mutations of MRP2 (ABCC2) cause the Dubin-Johnson syndrome, an inherited form of conjugated hyperbilirubinemia **(Chap. 303)**. A defective MDR3 (ABCB4) results in PFIC3. ABCG5/G8, the canalicular half transporters for cholesterol and other neutral sterols, are defective in sitosterolemia. The cystic fibrosis transmembrane regulator (CFTR, ABCC7) located on bile duct epithelial cells but not on canalicular membranes is defective in cystic fibrosis, which is associated with impaired cholangiocellular pH regulation during ductular bile formation and chronic cholestatic liver disease, occasionally resulting in biliary cirrhosis.

The Bile Acids

The primary bile acids, cholic acid and chenodeoxycholic acid (CDCA), are synthesized from cholesterol in the liver, conjugated with glycine or taurine, and secreted into the bile. Secondary bile acids, including deoxycholate and lithocholate, are formed in the colon as bacterial metabolites of the primary bile acids. However, lithocholic acid is much less efficiently absorbed from the colon than deoxycholic acid. Another secondary bile acid, found in low concentration, is ursodeoxycholic acid (UDCA), a stereoisomer of CDCA. In healthy subjects, the ratio of glycine to taurine conjugates in bile is ~3:1.

Bile acids are detergent-like molecules that in aqueous solutions and above a critical concentration of about 2 m*M* form molecular aggregates called *micelles*. Cholesterol alone is sparingly soluble in aqueous environments, and its solubility in bile depends on both the total lipid concentration and the relative molar percentages of bile acids and lecithin. Normal ratios of these constituents favor the formation of solubilizing *mixed micelles*, while abnormal ratios promote the precipitation of cholesterol crystals in bile via an intermediate liquid crystal phase.

In addition to facilitating the biliary excretion of cholesterol, bile acids facilitate the normal intestinal absorption of dietary fats, mainly cholesterol and fat-soluble vitamins, via a micellar transport mechanism **(Chap. 294)**. Bile acids also serve as a major physiologic driving force for hepatic bile flow and aid in water and electrolyte transport in the small bowel and colon.

Enterohepatic Circulation

Bile acids are efficiently conserved under normal conditions. Unconjugated, and to a lesser degree also conjugated, bile acids are absorbed by *passive diffusion* along the entire gut. Quantitatively much more important for bile salt recirculation, however, is the *active transport* mechanism for conjugated bile acids in the distal ileum **(Chap. 294)**. The reabsorbed bile acids enter the portal bloodstream and are taken up rapidly by hepatocytes, reconjugated, and resecreted into bile (enterohepatic circulation).

The normal bile acid pool size is approximately 2–4 g. During digestion of a meal, the bile acid pool undergoes at least one or more enterohepatic cycles, depending on the size and composition of the meal. Normally, the bile acid pool circulates ~5–10 times daily. Intestinal absorption of the pool is about 95% efficient; therefore, fecal loss of bile acids is in the range of 0.2–0.4 g/d. In the steady state, this fecal loss is compensated by an equal daily synthesis of bile acids by the liver, and, thus, the size of the bile acid pool is maintained. Bile acids returning to the liver suppress de novo hepatic synthesis of primary bile acids from cholesterol by inhibiting the rate-limiting enzyme cholesterol 7-hydroxylase. While the loss of bile salts in stool is usually matched by increased hepatic synthesis, the maximum rate of synthesis is ~5 g/d, which may be insufficient to replete the bile acid pool size when there is pronounced impairment of intestinal bile salt reabsorption.

The expression of ABC transporters in the enterohepatic circulation and of the rate-limiting enzymes of bile acid and cholesterol synthesis are regulated in a coordinated fashion by nuclear receptors, which are ligand-activated transcription factors. The hepatic bile salt export pump (BSEP, ABCB11) is upregulated by the farnesoid X receptor (FXR), a bile acid sensor that also represses bile acid synthesis. The expression of the cholesterol transporter, ABCG5/G8, is upregulated by the liver X receptor (LXR), which is an oxysterol sensor.

Gallbladder and Sphincteric Functions

In the fasting state, the sphincter of Oddi offers a high-pressure zone of resistance to bile flow from the CBD into the duodenum. This tonic contraction serves to (1) prevent reflux of duodenal contents into the pancreatic and bile ducts and (2) promote filling of the gallbladder. The major factor controlling the evacuation of the gallbladder is the peptide hormone cholecystokinin (CCK), which is released from the duodenal mucosa in response to the ingestion of fats and amino acids. CCK produces (1) powerful contraction of the gallbladder, (2) decreased resistance of the sphincter of Oddi, and (3) enhanced flow of biliary contents into the duodenum.

Hepatic bile is "concentrated" within the gallbladder by energy-dependent transmucosal absorption of water and electrolytes. Almost the entire bile acid pool may be sequestered in the gallbladder following an overnight fast for delivery into the duodenum with the first meal of the day. The normal capacity of the gallbladder is ~30 mL of bile.

DISEASES OF THE GALLBLADDER

Congenital Anomalies

Anomalies of the biliary tract are not uncommon and include abnormalities in number, size, and shape (e.g., agenesis of the gallbladder, duplications, rudimentary or oversized "giant" gallbladders, and diverticula). *Phrygian cap* is a clinically innocuous entity in which a partial or complete septum (or fold) separates the fundus from the body. Anomalies of position or suspension are not uncommon and include left-sided gallbladder, intrahepatic gallbladder, retrodisplacement of the gallbladder, and "floating" gallbladder. The latter condition predisposes to acute torsion, volvulus, or herniation of the gallbladder.

Gallstones

EPIDEMIOLOGY AND PATHOGENESIS

Gallstones are quite prevalent in most western countries. In the United States, the third National Health and Nutrition Examination Survey (NHANES III) has revealed an overall prevalence of gallstones of 7.9% in men and 16.6% in women. The prevalence was high in Mexican Americans (8.9% in men, 26.7% in women), intermediate for non-Hispanic whites (8.6% in men, 16.6% in women), and low for African Americans (5.3% in men, 13.9% in women).

Gallstones are formed because of abnormal bile composition. They are divided into two major types: cholesterol stones account for more than 80% of the total, with pigment stones comprising less than 20%. Cholesterol gallstones usually contain >50% cholesterol monohydrate plus an admixture of calcium salts, bile pigments, and proteins. Pigment stones are composed primarily of calcium bilirubinate; they contain <20% cholesterol and are classified into "black" and "brown" types, the latter forming secondary to chronic biliary infection.

Cholesterol Stones and Biliary Sludge

Cholesterol is essentially water insoluble and requires aqueous dispersion into either micelles or vesicles, both of which require the presence of a second lipid to solubilize the cholesterol. Cholesterol and phospholipids are secreted into bile as unilamellar bilayered vesicles, which are converted into mixed micelles consisting of bile acids, phospholipids, and cholesterol by the action of bile acids. If there is an excess of cholesterol in relation to phospholipids and bile acids, unstable, cholesterol-rich vesicles remain, which aggregate into large multilamellar vesicles from which cholesterol crystals precipitate **(Fig. 311-1)**.

Figure 311-1

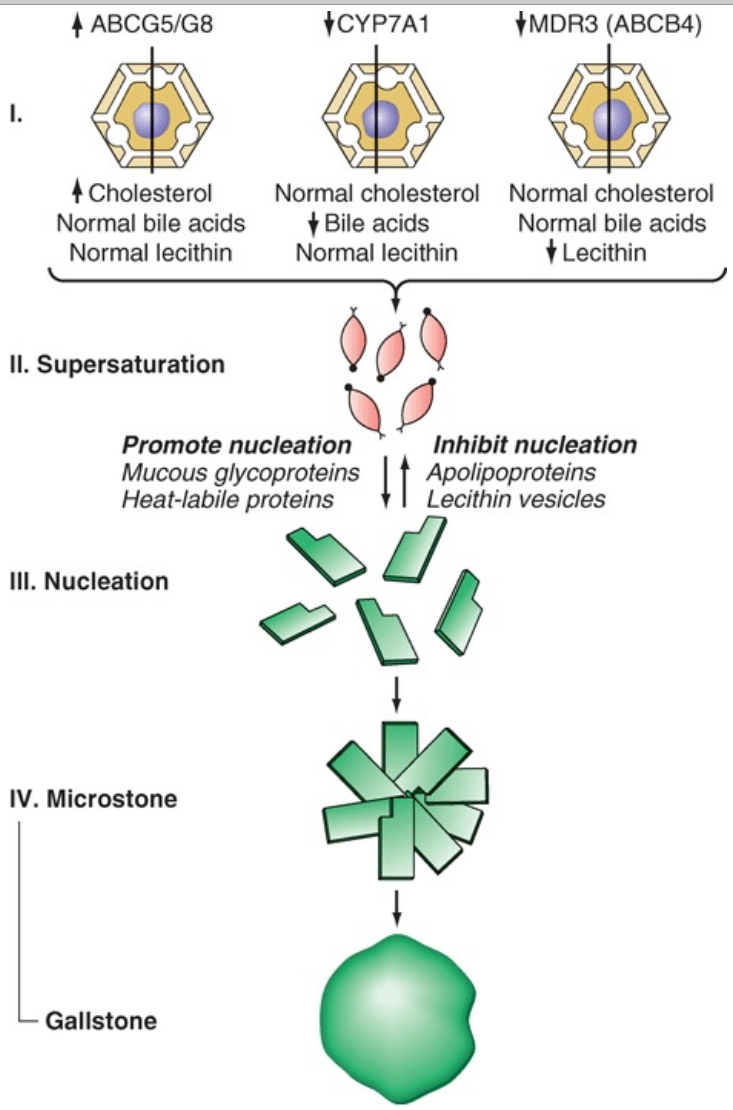

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scheme showing pathogenesis of cholesterol gallstone formation. Conditions or factors that increase the ratio of cholesterol to bile acids and phospholipids (lecithin) favor gallstone formation. ABCB4, ATPbinding cassette transporter; ABCG5/8, ATP-binding cassette (ABC) transporter G5/G8; CYP7A1, cytochrome P-450 7A1; MDR3, multidrug resistance protein 3, also called phospholipid export pump.

There are several important mechanisms in the formation of lithogenic (stone-forming) bile. The most important is increased biliary secretion of cholesterol. This may occur in association with obesity, the metabolic syndrome, high-caloric and cholesterol-rich diets, or drugs (e.g., clofibrate) and may result from increased activity of hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-limiting enzyme of hepatic cholesterol synthesis, and increased hepatic uptake of cholesterol from blood. In patients with gallstones, dietary cholesterol *increases* biliary cholesterol secretion. This does not occur in non-gallstone patients on high-cholesterol diets. In addition to environmental factors such as high-caloric and cholesterol-rich diets, genetic factors play an important role in gallstone disease. A large study of symptomatic gallstones in Swedish twins provided strong evidence for a role of genetic factors in gallstone pathogenesis. Genetic factors accounted for 25%, shared environmental factors for 13%, and individual environmental factors for 62% of the phenotypic variation among monozygotic twins. A single nucleotide polymorphism of the gene encoding the hepatic cholesterol transporter ABCG5/G8 has been found in 21% of patients with gallstones, but only in 9% of the general population. It is thought to cause a gain of function of the cholesterol transporter and to contribute to cholesterol hypersecretion. A high prevalence of gallstones is found among first-degree relatives of gallstone carriers and in certain ethnic populations such as American Indians as well as Chilean Indians and Chilean Hispanics. A common genetic trait has been identified for some of these populations by mitochondrial DNA analysis. In some patients, impaired hepatic conversion of cholesterol to bile acids may also occur, resulting in an increase of the lithogenic cholesterol/bile acid ratio. Although most cholesterol stones have a polygenic basis, there are rare monogenic (mendelian) causes. Recently, a mutation in the *CYP7A1* gene has been described that results in a deficiency of the enzyme cholesterol 7-hydroxylase, which catalyzes the initial step in cholesterol catabolism and bile acid synthesis. The homozygous state is associated with hypercholesterolemia and gallstones. Because the phenotype is expressed in the heterozygote state, mutations in the *CYP7A1* gene may contribute to the susceptibility to cholesterol gallstone disease in the population. Mutations in the *MDR3* (ABCB4) gene, which encodes the phospholipid export pump in the canalicular membrane of the hepatocyte, may cause defective phospholipid secretion into bile, resulting in cholesterol supersaturation of bile and formation of cholesterol gallstones in the gallbladder and in the bile ducts. Thus, an excess of biliary cholesterol in relation to bile acids and phospholipids is primarily due to hypersecretion of cholesterol, but hyposecretion of bile acids or phospholipids may contribute. An additional disturbance of bile acid metabolism that is likely to contribute to supersaturation of bile with cholesterol is enhanced conversion of cholic acid to deoxycholic acid, with replacement of the cholic acid pool by an expanded deoxycholic acid pool. It may result from enhanced dehydroxylation of cholic acid and increased absorption of newly formed deoxycholic acid. An increased deoxycholate secretion is associated with hypersecretion of cholesterol into bile.

While supersaturation of bile with cholesterol is an important prerequisite for gallstone formation, it is generally not sufficient by itself to produce cholesterol precipitation in vivo. Most individuals with supersaturated bile do not develop stones because the time required for cholesterol crystals to nucleate and grow is longer than the time bile spends in the gallbladder.

An important mechanism is *nucleation* of cholesterol monohydrate crystals, which is greatly accelerated in human lithogenic bile. Accelerated nucleation of cholesterol monohydrate in bile may be due to either an *excess of pronucleating factors* or a *deficiency of antinucleating factors* . Mucin and certain non-mucin glycoproteins, principally immunoglobulins, appear to be pronucleating factors, while apolipoproteins A-I and A-II and other glycoproteins appear to be antinucleating factors. Cholesterol monohydrate crystal nucleation and crystal growth probably occur within the mucin gel layer. Vesicle fusion leads to liquid crystals, which, in turn, nucleate into solid cholesterol monohydrate crystals. Continued growth of the crystals occurs by direct nucleation of cholesterol molecules from supersaturated unilamellar or multilamellar biliary vesicles.

A third important mechanism in cholesterol gallstone formation is *gallbladder hypomotility* . If the gallbladder emptied all supersaturated or crystal-containing bile completely, stones would not be able to grow. A high percentage of patients with gallstones exhibit abnormalities of gallbladder emptying. Ultrasonographic studies show that gallstone patients display an increased gallbladder volume during fasting and also after a test meal (residual volume) and that fractional emptying after gallbladder stimulation is decreased.

Biliary sludge is a thick, mucous material that, upon microscopic examination, reveals lecithin-cholesterol liquid crystals, cholesterol monohydrate crystals, calcium bilirubinate, and mucin gels. Biliary sludge typically forms a crescent-like layer in the most dependent portion of the gallbladder and is recognized by characteristic echoes on ultrasonography (see below). The presence of biliary sludge implies two abnormalities: (1) the normal balance between gallbladder mucin secretion and elimination has become deranged and (2) nucleation of biliary solutes has occurred. That biliary sludge may be a

precursor form of gallstone disease is evident from several observations. In one study, 96 patients with gallbladder sludge were followed prospectively by serial ultrasound studies. In 18%, biliary sludge disappeared and did not recur for at least 2 years. In 60%, biliary sludge disappeared and reappeared; in 14%, gallstones (8% asymptomatic, 6% symptomatic) developed; and in 6%, severe biliary pain with or without acute pancreatitis occurred. In 12 patients, cholecystectomies were performed, 6 for gallstone-associated biliary pain and 3 in symptomatic patients with sludge but without gallstones who had prior attacks of pancreatitis; the latter did not recur after cholecystectomy. It should be emphasized that biliary sludge can develop with disorders that cause gallbladder hypomotility; i.e., surgery, burns, total parenteral nutrition, pregnancy, and oral contraceptives—all of which are associated with gallstone formation. However, the presence of biliary sludge implies supersaturation of bile with either cholesterol or calcium bilirubinate.

Two other conditions are associated with cholesterol-stone or biliary-sludge formation: pregnancy and rapid weight reduction through a very low-calorie diet. There appear to be two key changes during pregnancy that contribute to a "cholelithogenic state": (1) a marked increase in cholesterol saturation of bile during the third trimester and (2) sluggish gallbladder contraction in response to a standard meal, resulting in impaired gallbladder emptying. That these changes are related to pregnancy per se is supported by several studies that show reversal of these abnormalities quite rapidly after delivery. During pregnancy, gallbladder sludge develops in 20–30% of women and gallstones in 5–12%. Although biliary sludge is a common finding during pregnancy, it is usually asymptomatic and often resolves spontaneously after delivery. Gallstones, which are less common than sludge and frequently associated with biliary colic, may also disappear after delivery because of spontaneous dissolution related to bile becoming unsaturated with cholesterol postpartum.

Approximately 10–20% of persons with rapid weight reduction achieved through very low calorie dieting develop gallstones. In a study involving 600 patients who completed a 16-week, 520-kcal/d diet, UDCA in a dosage of 600 mg/d proved highly effective in preventing gallstone formation; gallstones developed in only 3% of UDCA recipients, compared to 28% of placebo-treated patients.

To summarize, cholesterol gallstone disease occurs because of several defects, which include (1) bile supersaturation with cholesterol, (2) nucleation of cholesterol monohydrate with subsequent crystal retention and stone growth, and (3) abnormal gallbladder motor function with delayed emptying and stasis. Other important factors known to predispose to cholesterol-stone formation are summarized in **Table 311-1**.

Table 311-1 Predisposing Factors for Cholesterol and Pigment Gallstone Formation

Cholesterol Stones
1. Demographic/genetic factors: Prevalence highest in North American Indians, Chilean Indians, and Chilean Hispanics, greater in Northern Europe and North America than in Asia, lowest in Japan; familial disposition; hereditary aspects
2. Obesity, metabolic syndrome: Normal bile acid pool and secretion but increased biliary secretion of cholesterol
3. Weight loss: Mobilization of tissue cholesterol leads to increased biliary cholesterol secretion while enterohepatic circulation of bile acids is decreased
4. Female sex hormones
a. Estrogens stimulate hepatic lipoprotein receptors, increase uptake of dietary cholesterol, and increase biliary cholesterol secretion
b. Natural estrogens, other estrogens, and oral contraceptives lead to decreased bile salt secretion and decreased conversion of cholesterol to cholesteryl esters
5. Increasing age: Increased biliary secretion of cholesterol, decreased size of bile acid pool, decreased secretion of bile salts
6. Gallbladder hypomotility leading to stasis and formation of sludge
a. Prolonged parenteral nutrition
b. Pregnancy
c. Fasting
d. Drugs such as octreotide
7. Clofibrate therapy: Increased biliary secretion of cholesterol
8. Decreased bile acid secretion
a. Primary biliary cirrhosis
b. Genetic defect of the *CYP7A1* gene
9. Decreased phospholipid secretion: Genetic defect of the *MDR3* gene
10. Miscellaneous
a. High-calorie, high-fat diet

b. Spinal cord injury
Pigment Stones
1. Demographic/genetic factors: Asia, rural setting
2. Chronic hemolysis
3. Alcoholic cirrhosis
4. Pernicious anemia
5. Cystic fibrosis
6. Chronic biliary tract infection, parasite infections
7. Increasing age
8. Ileal disease, ileal resection or bypass

Pigment Stones

Black pigment stones are composed of either pure calcium bilirubinate or polymer-like complexes with calcium and mucin glycoproteins. They are more common in patients who have chronic hemolytic states (with increased conjugated bilirubin in bile), liver cirrhosis, Gilbert's syndrome, or cystic fibrosis. Gallbladder stones in patients with ileal diseases, ileal resection, or ileal bypass generally are also black pigment stones. Enterohepatic recycling of bilirubin in ileal disease states contributes to their pathogenesis. Brown pigment stones are composed of calcium salts of unconjugated bilirubin with varying amounts of cholesterol and protein. They are caused by the presence of increased amounts of unconjugated, insoluble bilirubin in bile that precipitates to form stones. Deconjugation of an excess of soluble bilirubin mono- and diglucuronides may be mediated by endogenous β-glucuronidase but may also occur by spontaneous hydrolysis. Sometimes, the enzyme is also produced when bile is chronically infected by bacteria, and such stones are brown. Pigment stone formation is especially prominent in Asians and is often associated with infections in the gallbladder and biliary tree (Table 311-1).

DIAGNOSIS

Procedures of potential use in the diagnosis of cholelithiasis and other diseases of the gallbladder are detailed in **Table 311-2**. Ultrasonography of the gallbladder is very accurate in the identification of cholelithiasis and has replaced oral cholecystography **(Fig. 311-2***A*). Stones as small as 1.5 mm in diameter may be confidently identified provided that firm criteria are used [e.g., acoustic "shadowing" of opacities that are within the gallbladder lumen and that change with the patient's position (by gravity)]. In major medical centers, the false-negative and false-positive rates for ultrasound in gallstone patients are ~2–4%. Biliary sludge is material of low echogenic activity that typically forms a layer in the most dependent position of the gallbladder. This layer shifts with postural changes but fails to produce acoustic shadowing; these two characteristics distinguish sludges from gallstones. Ultrasound can also be used to assess the emptying function of the gallbladder.

Table 311-2 Diagnostic Evaluation of the Gallbladder

Diagnostic Advantages	Diagnostic Limitations	Comment
Gallbladder Ultrasound		
Rapid Accurate identification of gallstones (>95%) Simultaneous scanning of GB, liver, bile ducts, pancreas "Real-time" scanning allows assessment of GB volume, contractility Not limited by jaundice, pregnancy May detect very small stones	Bowel gas Massive obesity Ascites	Procedure of choice for detection of stones

Plain Abdominal x-ray		
Low cost	Relatively low yield	Pathognomonic findings in: calcified gallstones
Readily available	? Contraindicated in pregnancy	Limey bile, porcelain GB Emphysematous cholecystitis
		Gallstone ileus
Radioisotope Scans (HIDA, DIDA, etc.)		
Accurate identification of cystic duct obstruction	? Contraindicated in pregnancy	Indicated for confirmation of suspected acute cholecystitis; less sensitive and less specific in chronic cholecystitis; useful in diagnosis of acalculous cholecystopathy, especially if given with CCK to assess gallbladder emptying
Simultaneous assessment of bile ducts	Serum bilirubin >103–205 μmol/L (6–12 mg/dL)	
	Cholecystogram of low resolution	

Abbreviations: CCK, cholecystokinin; GB, gallbladder; GBUS, gallbladder ultrasound.

Figure 311-2

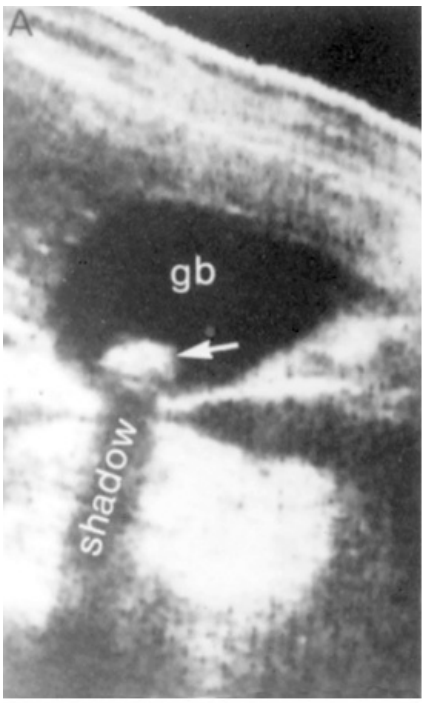

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

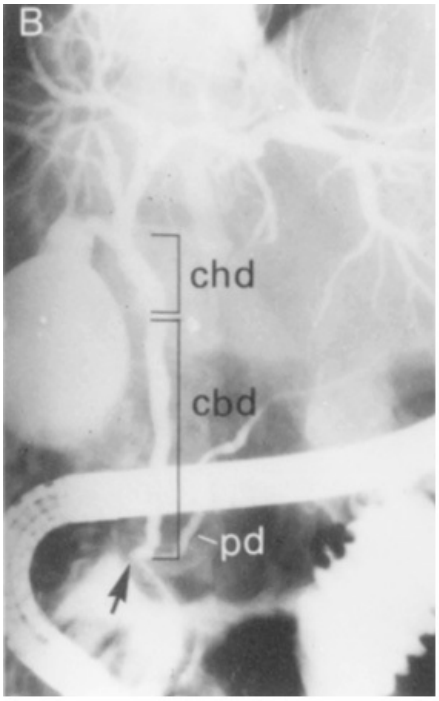

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

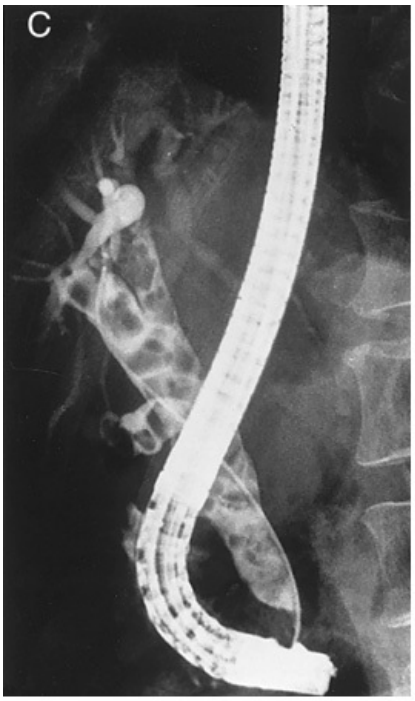

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

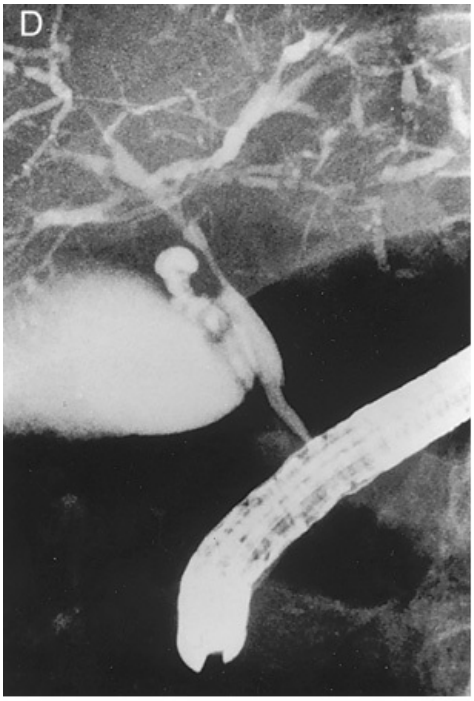

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examples of ultrasound and radiologic studies of the biliary tract. ***A***. An ultrasound study showing a distended gallbladder containing a single large stone (*arrow*), which casts an acoustic shadow. ***B***. Endoscopic retrograde cholangiopancreatogram (ERCP) showing normal biliary tract anatomy. In addition to the endoscope and large vertical gallbladder filled with contrast dye, the common hepatic duct (CHD), common bile duct (CBD), and pancreatic duct (PD) are shown. The arrow points to the ampulla of Vater. ***C***. Endoscopic retrograde cholangiogram (ERC) showing choledocholithiasis. The biliary tract is dilatated and contains multiple radiolucent calculi. ***D***. ERCP showing sclerosing cholangitis. The common bile duct shows areas that are strictured and narrowed.

The plain abdominal film may detect gallstones containing sufficient calcium to be radiopaque (10–15% of cholesterol and ~50% of pigment stones). Plain radiography may also be of use in the diagnosis of emphysematous cholecystitis, porcelain gallbladder, limey bile, and gallstone ileus.

Oral cholecystography (OCG) has historically been a useful procedure for the diagnosis of gallstones but has been replaced by ultrasound and is regarded as obsolete. It may be used to assess the patency of the cystic duct and gallbladder emptying function. Further, OCG can also delineate the size and number of gallstones and determine whether they are calcified.

Radiopharmaceuticals such as ^{99m}Tc-labeled *N*-substituted iminodiacetic acids (HIDA, DIDA, DISIDA, etc.) are rapidly extracted from the blood and are excreted into the biliary tree in high concentration even in the presence of mild to moderate serum bilirubin elevations. Failure to image the gallbladder in the presence of biliary ductal visualization may indicate cystic duct obstruction, acute or chronic cholecystitis, or surgical absence of the organ. Such scans have some application in the diagnosis of acute cholecystitis.

SYMPTOMS OF GALLSTONE DISEASE

Gallstones usually produce symptoms by causing inflammation or obstruction following their migration into the cystic duct or CBD. The most specific and characteristic symptom of gallstone disease is biliary colic that is a constant and often long-lasting pain (see below). Obstruction of the cystic duct or CBD by a stone produces increased intraluminal pressure and distention of the viscus that cannot be relieved by repetitive biliary contractions. The resultant visceral pain is characteristically a severe, steady ache or fullness in the epigastrium or right upper quadrant (RUQ) of the abdomen with frequent radiation to the interscapular area, right scapula, or shoulder.

Biliary colic begins quite suddenly and may persist with severe intensity for 15 min to 5 h, subsiding gradually or rapidly. It is steady rather than intermittent as would be suggested by the word *colic*, which must be regarded as a misnomer, although it is in widespread use. An episode of biliary pain persisting beyond 5 h should raise the suspicion of acute cholecystitis (see below). Nausea and vomiting frequently accompany episodes of biliary pain. An elevated level of serum bilirubin and/or alkaline phosphatase suggests a common duct stone. Fever or chills (rigors) with biliary pain usually imply a complication, i.e., cholecystitis, pancreatitis, or cholangitis. Complaints of vague epigastric fullness, dyspepsia, eructation, or flatulence, especially following a fatty meal, should not be confused with biliary pain. Such symptoms are frequently elicited from patients with or without gallstone disease

but are not specific for biliary calculi. Biliary colic may be precipitated by eating a fatty meal, by consumption of a large meal following a period of prolonged fasting, or by eating a normal meal; it is frequently nocturnal, occurring within a few hours of retiring.

NATURAL HISTORY

Gallstone disease discovered in an asymptomatic patient or in a patient whose symptoms are not referable to cholelithiasis is a common clinical problem. The natural history of "silent," or asymptomatic, gallstones has occasioned much debate. A study of predominantly male silent gallstone patients suggests that the cumulative risk for the development of symptoms or complications is relatively low—10% at 5 years, 15% at 10 years, and 18% at 15 years. Patients remaining asymptomatic for 15 years were found to be unlikely to develop symptoms during further follow-up, and most patients who did develop complications from their gallstones experienced *prior* warning symptoms. Similar conclusions apply to diabetic patients with silent gallstones. Decision analysis has suggested that (1) the cumulative risk of death due to gallstone disease while on expectant management is small, and (2) prophylactic cholecystectomy is not warranted.

Complications requiring cholecystectomy are much more common in gallstone patients who have developed symptoms of biliary pain. Patients found to have gallstones at a young age are more likely to develop symptoms from cholelithiasis than are patients >60 years at the time of initial diagnosis. Patients with diabetes mellitus and gallstones may be somewhat more susceptible to septic complications, but the magnitude of risk of septic biliary complications in diabetic patients is incompletely defined.

Treatment: Gallstones

SURGICAL THERAPY

In asymptomatic gallstone patients, the risk of developing symptoms or complications requiring surgery is quite small (in the range of 1–2% per year). Thus, a recommendation for cholecystectomy in a patient with gallstones should probably be based on assessment of three factors: (1) the presence of symptoms that are frequent enough or severe enough to interfere with the patient's general routine; (2) the presence of a prior complication of gallstone disease, i.e., history of acute cholecystitis, pancreatitis, gallstone fistula, etc.; or (3) the presence of an underlying condition predisposing the patient to increased risk of gallstone complications (e.g., calcified or porcelain gallbladder and/or a previous attack of acute cholecystitis regardless of current symptomatic status). Patients with very large gallstones (>3 cm in diameter) and patients having gallstones in a congenitally anomalous gallbladder might also be considered for prophylactic cholecystectomy. Although young age is a worrisome factor in asymptomatic gallstone patients, few authorities would now recommend routine cholecystectomy in all young patients with silent stones. Laparoscopic cholecystectomy is a minimal-access approach for the removal of the gallbladder together with its stones. Its advantages include a markedly shortened hospital stay, minimal disability, as well as decreased cost, and it is the procedure of choice for most patients referred for elective cholecystectomy.

From several studies involving >4000 patients undergoing laparoscopic cholecystectomy, the following key points emerge: (1) complications develop in ~4% of patients, (2) conversion to laparotomy occurs in 5%, (3) the death rate is remarkably low (i.e., <0.1%), and (4) bile duct injuries are unusual (i.e., 0.2–0.5%) but more frequent than with open cholecystectomy. These data indicate why laparoscopic cholecystectomy has become the "gold standard" for treating symptomatic cholelithiasis.

MEDICAL THERAPY—GALLSTONE DISSOLUTION

Ursodeoxycholic acid (UDCA) decreases cholesterol saturation of bile and also appears to produce a lamellar liquid crystalline phase in bile that allows a dispersion of cholesterol from stones by physical-chemical means. UDCA may also retard cholesterol crystal nucleation. In carefully selected patients with a functioning gallbladder and with radiolucent stones <10 mm in diameter, complete dissolution can be achieved in ~50% of patients within 6 months to 2 years. For good results within a reasonable time period, this therapy should be limited to radiolucent stones smaller than 5 mm in diameter. The dose of UDCA should be 10–15 mg/kg per day. Stones larger than 15 mm in size rarely dissolve. Pigment stones are not responsive to UDCA therapy. The highest success rate (i.e., >70%) occurs in patients with small (<5 mm) floating radiolucent gallstones. Probably ≤10% of patients with *symptomatic* cholelithiasis are candidates for such treatment. However, in addition to the vexing problem of recurrent stones (30–50% over 3–5 years of follow-up), there is also the factor of taking an expensive drug for up to 2 years. The advantages and success of laparoscopic cholecystectomy have largely reduced the role of gallstone dissolution to patients who wish to avoid or are not candidates for elective cholecystectomy. However, patients with cholesterol gallstone disease who develop recurrent choledocholithiasis after cholecystectomy should be on long-term treatment with ursodeoxycholic acid.

Acute and Chronic Cholecystitis

ACUTE CHOLECYSTITIS

Acute inflammation of the gallbladder wall usually follows obstruction of the cystic duct by a stone. Inflammatory response can be evoked by three factors: (1) *mechanical inflammation* produced by

increased intraluminal pressure and distention with resulting ischemia of the gallbladder mucosa and wall, (2) *chemical inflammation* caused by the release of lysolecithin (due to the action of phospholipase on lecithin in bile) and other local tissue factors, and (3) *bacterial inflammation*, which may play a role in 50–85% of patients with acute cholecystitis. The organisms most frequently isolated by culture of gallbladder bile in these patients include *Escherichia coli*, *Klebsiella* spp., *Streptococcus* spp., and *Clostridium* spp.

Acute cholecystitis often begins as an attack of biliary pain that progressively worsens. Approximately 60–70% of patients report having experienced prior attacks that resolved spontaneously. As the episode progresses, however, the pain of acute cholecystitis becomes more generalized in the right upper abdomen. As with biliary colic, the pain of cholecystitis may radiate to the interscapular area, right scapula, or shoulder. Peritoneal signs of inflammation such as increased pain with jarring or on deep respiration may be apparent. The patient is anorectic and often nauseated. Vomiting is relatively common and may produce symptoms and signs of vascular and extracellular volume depletion. Jaundice is unusual early in the course of acute cholecystitis but may occur when edematous inflammatory changes involve the bile ducts and surrounding lymph nodes.

A low-grade fever is characteristically present, but shaking chills or rigors are not uncommon. The RUQ of the abdomen is almost invariably tender to palpation. An enlarged, tense gallbladder is palpable in 25–50% of patients. Deep inspiration or cough during subcostal palpation of the RUQ usually produces increased pain and inspiratory arrest (Murphy's sign). Localized rebound tenderness in the RUQ is common, as are abdominal distention and hypoactive bowel sounds from paralytic ileus, but generalized peritoneal signs and abdominal rigidity are usually lacking, in the absence of perforation.

The diagnosis of acute cholecystitis is usually made on the basis of a characteristic history and physical examination. The triad of sudden onset of RUQ tenderness, fever, and leukocytosis is highly suggestive. Typically, leukocytosis in the range of 10,000–15,000 cells per microliter with a left shift on differential count is found. The serum bilirubin is mildly elevated [<85.5 μmol/L (5 mg/dL)] in fewer than half of patients, while about one-fourth have modest elevations in serum aminotransferases (usually less than a fivefold elevation). Ultrasound will demonstrate calculi in 90–95% of cases and is useful for detection of signs of gallbladder inflammation including thickening of the wall, pericholecystic fluid, and dilation of the bile duct. The radionuclide (e.g., HIDA) biliary scan may be confirmatory if bile duct imaging is seen without visualization of the gallbladder.

Approximately 75% of patients treated medically have remission of acute symptoms within 2–7 days following hospitalization. In 25%, however, a complication of acute cholecystitis will occur despite conservative treatment (see below). In this setting, prompt surgical intervention is required. Of the 75% of patients with acute cholecystitis who undergo remission of symptoms, ~25% will experience a recurrence of cholecystitis within 1 year, and 60% will have at least one recurrent bout within 6 years. In view of the natural history of the disease, acute cholecystitis is best treated by early surgery whenever possible.

Mirizzi's syndrome is a rare complication in which a gallstone becomes impacted in the cystic duct or neck of the gallbladder causing compression of the CBD, resulting in CBD obstruction and jaundice. Ultrasound shows gallstone(s) lying outside the hepatic duct. Endoscopic retrograde cholangiopancreatography (ERCP) **(Fig. 311-2*B*)** or percutaneous transhepatic cholangiography (PTC) or magnetic resonance cholangiopancreatography (MRCP) will usually demonstrate the characteristic extrinsic compression of the CBD. Surgery consists of removing the cystic duct, diseased gallbladder, and the impacted stone. The preoperative diagnosis of Mirizzi's syndrome is important to avoid CBD injury.

Acalculous Cholecystitis

In 5–10% of patients with acute cholecystitis, calculi obstructing the cystic duct are not found at surgery. In >50% of such cases, an underlying explanation for acalculous inflammation is not found. An increased risk for the development of acalculous cholecystitis is especially associated with serious trauma or burns, with the postpartum period following prolonged labor, and with orthopedic and other nonbiliary major surgical operations in the postoperative period. It may possibly complicate periods of prolonged parenteral hyperalimentation. For some of these cases, biliary sludge in the cystic duct may be responsible. Other precipitating factors include vasculitis, obstructing adenocarcinoma of the gallbladder, diabetes mellitus, torsion of the gallbladder, "unusual" bacterial infections of the gallbladder (e.g., *Leptospira*, *Streptococcus*, *Salmonella*, or *Vibrio cholerae*), and parasitic infestation of the gallbladder. Acalculous cholecystitis may also be seen with a variety of other systemic disease processes (sarcoidosis, cardiovascular disease, tuberculosis, syphilis, actinomycosis, etc.).

Although the clinical manifestations of acalculous cholecystitis are indistinguishable from those of calculous cholecystitis, the setting of acute gallbladder inflammation complicating severe underlying illness is characteristic of acalculous disease. Ultrasound, CT, or radionuclide examinations demonstrating a large, tense, static gallbladder without stones and with evidence of poor emptying over a prolonged period may be diagnostically useful in some cases. The complication rate for

acalculous cholecystitis exceeds that for calculous cholecystitis. Successful management of acute acalculous cholecystitis appears to depend primarily on early diagnosis and surgical intervention, with meticulous attention to postoperative care.

Acalculous Cholecystopathy

Disordered motility of the gallbladder can produce recurrent biliary pain in patients without gallstones. Infusion of an octapeptide of CCK can be used to measure the gallbladder ejection fraction during cholescintigraphy. The surgical findings have included abnormalities such as chronic cholecystitis, gallbladder muscle hypertrophy, and/or a markedly narrowed cystic duct. Some of these patients may well have had antecedent gallbladder disease. The following criteria can be used to identify patients with acalculous cholecystopathy: (1) recurrent episodes of typical RUQ pain characteristic of biliary tract pain, (2) abnormal CCK cholescintigraphy demonstrating a gallbladder ejection fraction of <40%, and (3) infusion of CCK reproduces the patient's pain. An additional clue would be the identification of a large gallbladder on ultrasound examination. Finally, it should be noted that sphincter of Oddi dysfunction can also give rise to recurrent RUQ pain and CCK-scintigraphic abnormalities.

Emphysematous Cholecystitis

So-called emphysematous cholecystitis is thought to begin with acute cholecystitis (calculous or acalculous) followed by ischemia or gangrene of the gallbladder wall and infection by gas-producing organisms. Bacteria most frequently cultured in this setting include anaerobes, such as *C. welchii* or *C. perfringens*, and aerobes, such as *E. coli.* This condition occurs most frequently in elderly men and in patients with diabetes mellitus. The clinical manifestations are essentially indistinguishable from those of nongaseous cholecystitis. The diagnosis is usually made on plain abdominal film by finding gas within the gallbladder lumen, dissecting within the gallbladder wall to form a gaseous ring, or in the pericholecystic tissues. The morbidity and mortality rates with emphysematous cholecystitis are considerable. Prompt surgical intervention coupled with appropriate antibiotics is mandatory.

CHRONIC CHOLECYSTITIS

Chronic inflammation of the gallbladder wall is almost always associated with the presence of gallstones and is thought to result from repeated bouts of subacute or acute cholecystitis or from persistent mechanical irritation of the gallbladder wall by gallstones. The presence of bacteria in the bile occurs in >25% of patients with chronic cholecystitis. The presence of infected bile in a patient with *chronic* cholecystitis undergoing elective cholecystectomy probably adds little to the operative risk. Chronic cholecystitis may be asymptomatic for years, may progress to symptomatic gallbladder disease or to acute cholecystitis, or may present with complications (see below).

COMPLICATIONS OF CHOLECYSTITIS

Empyema and Hydrops

Empyema of the gallbladder usually results from progression of acute cholecystitis with persistent cystic duct obstruction to superinfection of the stagnant bile with a pus-forming bacterial organism. The clinical picture resembles that of cholangitis with high fever; severe RUQ pain; marked leukocytosis; and often, prostration. Empyema of the gallbladder carries a high risk of gram-negative sepsis and/or perforation. Emergency surgical intervention with proper antibiotic coverage is required as soon as the diagnosis is suspected.

Hydrops or mucocele of the gallbladder may also result from prolonged obstruction of the cystic duct, usually by a large solitary calculus. In this instance, the obstructed gallbladder lumen is progressively distended, over a period of time, by mucus (mucocele) or by a clear transudate (hydrops) produced by mucosal epithelial cells. A visible, easily palpable, nontender mass sometimes extending from the RUQ into the right iliac fossa may be found on physical examination. The patient with hydrops of the gallbladder frequently remains asymptomatic, although chronic RUQ pain may also occur. Cholecystectomy is indicated, because empyema, perforation, or gangrene may complicate the condition.

Gangrene and Perforation

Gangrene of the gallbladder results from ischemia of the wall and patchy or complete tissue necrosis. Underlying conditions often include marked distention of the gallbladder, vasculitis, diabetes mellitus, empyema, or torsion resulting in arterial occlusion. Gangrene usually predisposes to perforation of the gallbladder, but perforation may also occur in chronic cholecystitis without premonitory warning symptoms. *Localized perforations* are usually contained by the omentum or by adhesions produced by recurrent inflammation of the gallbladder. Bacterial superinfection of the walled-off gallbladder contents results in abscess formation. Most patients are best treated with cholecystectomy, but some seriously ill patients may be managed with cholecystostomy and drainage of the abscess. *Free perforation* is less common but is associated with a mortality rate of ~30%. Such patients may experience a sudden transient relief of RUQ pain as the distended gallbladder decompresses; this is followed by signs of generalized peritonitis.

Fistula Formation and Gallstone Ileus

Fistulization into an adjacent organ adherent to the gallbladder wall may result from inflammation and adhesion formation. Fistulas into the duodenum are most common, followed in frequency by those involving the hepatic flexure of the colon, stomach or jejunum, abdominal wall, and renal pelvis. Clinically "silent" biliary-enteric fistulas occurring as a complication of acute cholecystitis have been found in up to 5% of patients undergoing cholecystectomy. Asymptomatic cholecystoenteric fistulas may sometimes be diagnosed by finding gas in the biliary tree on plain abdominal films. Barium contrast studies or endoscopy of the upper gastrointestinal tract or colon may demonstrate the fistula. Treatment in the symptomatic patient usually consists of cholecystectomy, CBD exploration, and closure of the fistulous tract.

Gallstone ileus refers to mechanical intestinal obstruction resulting from the passage of a large gallstone into the bowel lumen. The stone customarily enters the duodenum through a cholecystoenteric fistula at that level. The site of obstruction by the impacted gallstone is usually at the ileocecal valve, provided that the more proximal small bowel is of normal caliber. The majority of patients do not give a history of either prior biliary tract symptoms or complaints suggestive of acute cholecystitis or fistulization. Large stones, >2.5 cm in diameter, are thought to predispose to fistula formation by gradual erosion through the gallbladder fundus. Diagnostic confirmation may occasionally be found on the plain abdominal film (e.g., small-intestinal obstruction with gas in the biliary tree and a calcified, ectopic gallstone) or following an upper gastrointestinal series (cholecystoduodenal fistula with small-bowel obstruction at the ileocecal valve). Laparotomy with stone extraction (or propulsion into the colon) remains the procedure of choice to relieve obstruction. Evacuation of large stones within the gallbladder should also be performed. In general, the gallbladder and its attachment to the intestines should be left alone.

Limey (Milk of Calcium) Bile and Porcelain Gallbladder

Calcium salts in the lumen of the gallbladder in sufficient concentration may produce calcium precipitation and diffuse, hazy opacification of bile or a layering effect on plain abdominal roentgenography. This so-called limey bile, or milk of calcium bile, is usually clinically innocuous, but cholecystectomy is recommended, especially when it occurs in a hydropic gallbladder. In the entity called *porcelain gallbladder*, calcium salt deposition within the wall of a chronically inflamed gallbladder may be detected on the plain abdominal film. Cholecystectomy is advised in all patients with porcelain gallbladder because in a high percentage of cases this finding appears to be associated with the development of carcinoma of the gallbladder.

Treatment: Acute Cholecystitis

MEDICAL THERAPY

Although surgical intervention remains the mainstay of therapy for acute cholecystitis and its complications, a period of in-hospital stabilization may be required before cholecystectomy. Oral intake is eliminated, nasogastric suction may be indicated, and extracellular volume depletion and electrolyte abnormalities are repaired. Meperidine or nonsteroidal anti-inflammatory drugs (NSAIDs) are usually employed for analgesia because they may produce less spasm of the sphincter of Oddi than drugs such as morphine. Intravenous antibiotic therapy is usually indicated in patients with severe acute cholecystitis, even though bacterial superinfection of bile may not have occurred in the early stages of the inflammatory process. Antibiotic therapy is guided by the most common organisms likely to be present, which are *E. coli*, *Klebsiella* spp., and *Streptococcus* spp. Effective antibiotics include ureidopenicillins such as piperacillin or mezlocillin, ampicillin sulbactam, ciprofloxacin, moxifloxacin, and third-generation cephalosporins. Anaerobic coverage by a drug such as metronidazole should be added if gangrenous or emphysematous cholecystitis is suspected. Imipenem/meropenem represent potent parenteral antibiotics that cover the whole spectrum of bacteria causing ascending cholangitis. They should, however, be reserved for the most severe, life-threatening infections when other regimens have failed **(Chap. 149)**. Postoperative complications of wound infection, abscess formation, or sepsis are reduced in antibiotic-treated patients.

SURGICAL THERAPY

The optimal timing of surgical intervention in patients with acute cholecystitis depends on stabilization of the patient. The clear trend is toward earlier surgery, and this is due in part to requirements for shorter hospital stays. Urgent (emergency) cholecystectomy or cholecystostomy is probably appropriate in most patients in whom a complication of acute cholecystitis such as empyema, emphysematous cholecystitis, or perforation is suspected or confirmed. Patients with uncomplicated acute cholecystitis should undergo early elective laparoscopic cholecystectomy, ideally within 72 hours after diagnosis. The complication rate is not increased in patients undergoing early as opposed to delayed (>6 weeks after diagnosis) cholecystectomy. Delayed surgical intervention is probably best reserved for (1) patients in whom the overall medical condition imposes an unacceptable risk for early surgery and (2) patients in whom the diagnosis of acute cholecystitis is in doubt. Early cholecystectomy (within 72 hours) is the treatment of choice for most patients with acute cholecystitis. Mortality figures for emergency cholecystectomy in most centers approach 3%, while the mortality risk for early elective cholecystectomy is ~0.5% in patients under age 60. Of course, the operative risks increase with age-related diseases of other organ systems and with the presence of long- or short-

term complications of gallbladder disease. Seriously ill or debilitated patients with cholecystitis may be managed with cholecystostomy and tube drainage of the gallbladder. Elective cholecystectomy may then be done at a later date.

POSTCHOLECYSTECTOMY COMPLICATIONS

Early complications following cholecystectomy include atelectasis and other pulmonary disorders, abscess formation (often subphrenic), external or internal hemorrhage, biliary-enteric fistula, and bile leaks. Jaundice may indicate absorption of bile from an intraabdominal collection following a biliary leak or mechanical obstruction of the CBD by retained calculi, intraductal blood clots, or extrinsic compression.

Overall, cholecystectomy is a very successful operation that provides total or near-total relief of preoperative symptoms in 75–90% of patients. The most common cause of persistent postcholecystectomy symptoms is an overlooked symptomatic nonbiliary disorder (e.g., reflux esophagitis, peptic ulceration, pancreatitis, or—most often—irritable bowel syndrome). In a small percentage of patients, however, a disorder of the extrahepatic bile ducts may result in persistent symptomatology. These so-called postcholecystectomy syndromes may be due to (1) biliary strictures, (2) retained biliary calculi, (3) cystic duct stump syndrome, (4) stenosis or dyskinesia of the sphincter of Oddi, or (5) bile salt–induced diarrhea or gastritis.

Cystic Duct Stump Syndrome

In the absence of cholangiographically demonstrable retained stones, symptoms resembling biliary pain or cholecystitis in the postcholecystectomy patient have frequently been attributed to disease in a long (>1 cm) cystic duct remnant (cystic duct stump syndrome). Careful analysis, however, reveals that postcholecystectomy complaints are attributable to other causes in almost all patients in whom the symptom complex was originally thought to result from the existence of a long cystic duct stump. Accordingly, considerable care should be taken to investigate the possible role of other factors in the production of postcholecystectomy symptoms before attributing them to cystic duct stump syndrome.

Papillary Dysfunction, Papillary Stenosis, Spasm of the Sphincter of Oddi, and Biliary Dyskinesia

Symptoms of biliary colic accompanied by signs of recurrent, intermittent biliary obstruction may be produced by papillary stenosis, papillary dysfunction, spasm of the sphincter of Oddi, and biliary dyskinesia. Papillary stenosis is thought to result from acute or chronic inflammation of the papilla of Vater or from glandular hyperplasia of the papillary segment. Five criteria have been used to define papillary stenosis: (1) upper abdominal pain, usually RUQ or epigastric; (2) abnormal liver tests; (3) dilatation of the common bile duct upon ERCP examination; (4) delayed (>45 min) drainage of contrast material from the duct; and (5) increased basal pressure of the sphincter of Oddi, a finding that may be of only minor significance. An alternative to ERCP is magnetic resonance cholangiography (MRC) if ERCP and/or biliary manometry are either unavailable or not feasible. In patients with papillary stenosis, quantitative hepatobiliary scintigraphy has revealed delayed transit from the common bile duct to the bowel, ductal dilatation, and abnormal time-activity dynamics. This technique can also be used before and after sphincterotomy to document improvement in biliary emptying. Treatment consists of endoscopic or surgical sphincteroplasty to ensure wide patency of the distal portions of both the bile and pancreatic ducts. The greater the number of the preceding criteria present, the greater the likelihood that a patient does have a degree of papillary stenosis sufficient to justify correction. The factors usually considered as indications for sphincterotomy include (1) prolonged duration of symptoms, (2) lack of response to symptomatic treatment, (3) presence of severe disability, and (4) the patient's choice of sphincterotomy over surgery (given a clear understanding on his or her part of the risks involved in both procedures).

Criteria for diagnosing dyskinesia of the sphincter of Oddi are even more controversial than those for papillary stenosis. Proposed mechanisms include spasm of the sphincter, denervation sensitivity resulting in hypertonicity, and abnormalities of the sequencing or frequency rates of sphincteric-contraction waves. When thorough evaluation has failed to demonstrate another cause for the pain, and when cholangiographic and manometric criteria suggest a diagnosis of biliary dyskinesia, medical treatment with nitrites or anticholinergics to attempt pharmacologic relaxation of the sphincter has been proposed. Endoscopic biliary sphincterotomy (EBS) or surgical sphincteroplasty may be indicated in patients who fail to respond to a 2- to 3-month trial of medical therapy, especially if basal sphincter of Oddi pressures are elevated. EBS has become the procedure of choice for removing bile duct stones and for other biliary and pancreatic problems.

Bile Salt–Induced Diarrhea and Gastritis

Postcholecystectomy patients may develop symptoms of dyspepsia, which have been attributed to duodenogastric reflux of bile. However, firm data linking these symptoms to bile gastritis after surgical removal of the gallbladder are lacking. Cholecystectomy induces persistent changes in gut transit, and these changes effect a noticeable modification of bowel habits. Cholecystectomy shortens gut transit time by accelerating passage of the fecal bolus through the colon with marked acceleration in the right colon, thus causing an increase in colonic bile acid output and a shift in bile acid composition toward

the more diarrheagenic secondary bile acids. Diarrhea that is severe enough, i.e., three or more watery movements per day, can be classified as postcholecystectomy diarrhea, and this occurs in 5–10% of patients undergoing elective cholecystectomy. Treatment with bile acid–sequestering agents such as cholestyramine or colestipol is often effective in ameliorating troublesome diarrhea.

The Hyperplastic Cholecystoses

The term *hyperplastic cholecystoses* is used to denote a group of disorders of the gallbladder characterized by excessive proliferation of normal tissue components.

Adenomyomatosis is characterized by a benign proliferation of gallbladder surface epithelium with glandlike formations, extramural sinuses, transverse strictures, and/or fundal nodule ("adenoma" or "adenomyoma") formation.

Cholesterolosis is characterized by abnormal deposition of lipid, especially cholesteryl esters within macrophages in the lamina propria of the gallbladder wall. In its diffuse form ("strawberry gallbladder"), the gallbladder mucosa is brick red and speckled with bright yellow flecks of lipid. The localized form shows solitary or multiple "cholesterol polyps" studding the gallbladder wall. Cholesterol stones of the gallbladder are found in nearly half the cases. Cholecystectomy is indicated in both adenomyomatosis and cholesterolosis when symptomatic or when cholelithiasis is present.

The prevalence of gallbladder polyps in the adult population is ~5%, with a marked male predominance. Few significant changes have been found over a 5-year period in asymptomatic patients with gallbladder polyps <10 mm in diameter. Cholecystectomy is recommended in symptomatic patients, as well as in asymptomatic patients >50 years of age, or in those whose polyps are >10 mm in diameter or associated with gallstones or polyp growth on serial ultrasonography.

DISEASES OF THE BILE DUCTS

Congenital Anomalies

BILIARY ATRESIA AND HYPOPLASIA

Atretic and hypoplastic lesions of the extrahepatic and large intrahepatic bile ducts are the most common biliary anomalies of clinical relevance encountered in infancy. The clinical picture is one of severe obstructive jaundice during the first month of life, with pale stools. When biliary atresia is suspected on the basis of clinical, laboratory, and imaging findings the diagnosis is confirmed by surgical exploration and operative cholangiography. Approximately 10% of cases of biliary atresia are treatable with roux-en-Y choledochojejunostomy, with the Kasai procedure (hepatic portoenterostomy) being attempted in the remainder in an effort to restore some bile flow. Most patients, even those having successful biliary-enteric anastomoses, eventually develop chronic cholangitis, extensive hepatic fibrosis, and portal hypertension.

CHOLEDOCHAL CYSTS

Cystic dilatation may involve the free portion of the CBD, i.e., choledochal cyst, or may present as diverticulum formation in the intraduodenal segment. In the latter situation, chronic reflux of pancreatic juice into the biliary tree can produce inflammation and stenosis of the extrahepatic bile ducts leading to cholangitis or biliary obstruction. Because the process may be gradual, ~50% of patients present with onset of symptoms after age 10. The diagnosis may be made by ultrasound, abdominal CT, MRC, or cholangiography. Only one-third of patients show the classic triad of abdominal pain, jaundice, and an abdominal mass. Ultrasonographic detection of a cyst separate from the gallbladder should suggest the diagnosis of choledochal cyst, which can be confirmed by demonstrating the entrance of extrahepatic bile ducts into the cyst. Surgical treatment involves excision of the "cyst" and biliary-enteric anastomosis. Patients with choledochal cysts are at increased risk for the subsequent development of cholangiocarcinoma.

CONGENITAL BILIARY ECTASIA

Dilatation of intrahepatic bile ducts may involve either the major intrahepatic radicles (Caroli's disease), the inter- and intralobular ducts (congenital hepatic fibrosis), or both. In Caroli's disease, clinical manifestations include recurrent cholangitis, abscess formation in and around the affected ducts, and, often, gallstone formation within portions of ectatic intrahepatic biliary radicles. Ultrasound, MRC, and CT are of great diagnostic value in demonstrating cystic dilatation of the intrahepatic bile ducts. Treatment with ongoing antibiotic therapy is usually undertaken in an effort to limit the frequency and severity of recurrent bouts of cholangitis. Progression to secondary biliary cirrhosis with portal hypertension, extrahepatic biliary obstruction, cholangiocarcinoma, or recurrent episodes of sepsis with hepatic abscess formation is common.

Choledocholithiasis

PATHOPHYSIOLOGY AND CLINICAL MANIFESTATIONS

Passage of gallstones into the CBD occurs in ~10–15% of patients with cholelithiasis. The incidence of common duct stones increases with increasing age of the patient, so that up to 25% of elderly patients may have calculi in the common duct at the time of cholecystectomy. Undetected duct stones

are left behind in ~1–5% of cholecystectomy patients. The overwhelming majority of bile duct stones are cholesterol stones formed in the gallbladder, which then migrate into the extrahepatic biliary tree through the cystic duct. Primary calculi arising de novo in the ducts are usually pigment stones developing in patients with (1) hepatobiliary parasitism or chronic, recurrent cholangitis; (2) congenital anomalies of the bile ducts (especially Caroli's disease); (3) dilated, sclerosed, or strictured ducts; or (4) an *MDR3* (ABCB4) gene defect leading to impaired biliary phospholipids secretion (low phospholipid–associated cholelithiasis). Common duct stones may remain asymptomatic for years, may pass spontaneously into the duodenum, or (most often) may present with biliary colic or a complication.

COMPLICATIONS

Cholangitis

Cholangitis may be acute or chronic, and symptoms result from inflammation, which usually is caused by at least partial obstruction to the flow of bile. Bacteria are present on bile culture in ~75% of patients with acute cholangitis early in the symptomatic course. The characteristic presentation of acute cholangitis involves biliary pain, jaundice, and spiking fevers with chills (Charcot's triad). Blood cultures are frequently positive, and leukocytosis is typical. *Nonsuppurative acute cholangitis* is most common and may respond relatively rapidly to supportive measures and to treatment with antibiotics. In *suppurative acute cholangitis*, however, the presence of pus under pressure in a completely obstructed ductal system leads to symptoms of severe toxicity—mental confusion, bacteremia, and septic shock. Response to antibiotics alone in this setting is relatively poor, multiple hepatic abscesses are often present, and the mortality rate approaches 100% unless prompt endoscopic or surgical relief of the obstruction and drainage of infected bile are carried out. Endoscopic management of bacterial cholangitis is as effective as surgical intervention. ERCP with endoscopic sphincterotomy is safe and the preferred initial procedure for both establishing a definitive diagnosis and providing effective therapy.

Obstructive Jaundice

Gradual obstruction of the CBD over a period of weeks or months usually leads to initial manifestations of jaundice or pruritus without associated symptoms of biliary colic or cholangitis. Painless jaundice may occur in patients with choledocholithiasis, but is much more characteristic of biliary obstruction secondary to malignancy of the head of the pancreas, bile ducts, or ampulla of Vater.

In patients whose obstruction is secondary to choledocholithiasis, associated chronic calculous cholecystitis is very common, and the gallbladder in this setting may be relatively indistensible. The absence of a palpable gallbladder in most patients with biliary obstruction from duct stones is the basis for Courvoisier's law, i.e., that the presence of a palpably enlarged gallbladder suggests that the biliary obstruction is secondary to an underlying malignancy rather than to calculous disease. Biliary obstruction causes progressive dilatation of the intrahepatic bile ducts as intrabiliary pressures rise. Hepatic bile flow is suppressed, and reabsorption and regurgitation of conjugated bilirubin into the bloodstream lead to jaundice accompanied by dark urine (bilirubinuria) and light-colored (acholic) stools.

CBD stones should be suspected in any patient with cholecystitis whose serum bilirubin level is >85.5 μmol/L (5 mg/dL). The maximum bilirubin level is seldom >256.5 μmol/L (15.0 mg/dL) in patients with choledocholithiasis unless concomitant hepatic disease or another factor leading to marked hyperbilirubinemia exists. Serum bilirubin levels ≥342.0 μmol/L (20 mg/dL) should suggest the possibility of neoplastic obstruction. The serum alkaline phosphatase level is almost always elevated in biliary obstruction. A rise in alkaline phosphatase often precedes clinical jaundice and may be the only abnormality in routine liver function tests. There may be a two- to tenfold elevation of serum aminotransferases, especially in association with acute obstruction. Following relief of the obstructing process, serum aminotransferase elevations usually return rapidly to normal, while the serum bilirubin level may take 1–2 weeks to return to normal. The alkaline phosphatase level usually falls slowly, lagging behind the decrease in serum bilirubin.

Pancreatitis

The most common associated entity discovered in patients with nonalcoholic acute pancreatitis is biliary tract disease. Biochemical evidence of pancreatic inflammation complicates acute cholecystitis in 15% of cases and choledocholithiasis in >30%, and the common factor appears to be the passage of gallstones through the common duct. Coexisting pancreatitis should be suspected in patients with symptoms of cholecystitis who develop (1) back pain or pain to the left of the abdominal midline, (2) prolonged vomiting with paralytic ileus, or (3) a pleural effusion, especially on the left side. Surgical treatment of gallstone disease is usually associated with resolution of the pancreatitis.

Secondary Biliary Cirrhosis

Secondary biliary cirrhosis may complicate prolonged or intermittent duct obstruction with or without recurrent cholangitis. Although this complication may be seen in patients with choledocholithiasis, it is more common in cases of prolonged obstruction from stricture or neoplasm. Once established,

secondary biliary cirrhosis may be progressive even after correction of the obstructing process, and increasingly severe hepatic cirrhosis may lead to portal hypertension or to hepatic failure and death. Prolonged biliary obstruction may also be associated with clinically relevant deficiencies of the fat-soluble vitamins A, D, E, and K.

DIAGNOSIS AND TREATMENT

The diagnosis of choledocholithiasis is usually made by cholangiography **(Table 311-3)**, either preoperatively by endoscopic retrograde cholangiogram (ERC) **(Fig. 311-2*C*)** or MRCP or intraoperatively at the time of cholecystectomy. As many as 15% of patients undergoing cholecystectomy will prove to have CBD stones. When CBD stones are suspected prior to laparoscopic cholecystectomy, preoperative ERCP with endoscopic papillotomy and stone extraction is the preferred approach. It not only provides stone clearance but also defines the anatomy of the biliary tree in relationship to the cystic duct. CBD stones should be suspected in gallstone patients who have any of the following risk factors: (1) a history of jaundice or pancreatitis, (2) abnormal tests of liver function, and (3) ultrasonographic or MRCP evidence of a dilated CBD or stones in the duct. Alternatively, if intraoperative cholangiography reveals retained stones, postoperative ERCP can be carried out. The need for preoperative ERCP is expected to decrease further as laparoscopic techniques for bile duct exploration improve.

Table 311-3 Diagnostic Evaluation of the Bile Ducts

Diagnostic Advantages	Diagnostic Limitations	Contraindications	Complications	Comment
Hepatobiliary Ultrasound				
Rapid Simultaneous scanning of GB, liver, bile ducts, pancreas Accurate identification of dilated bile ducts Not limited by jaundice, pregnancy Guidance for fine-needle biopsy	Bowel gas Massive obesity Ascites Barium Partial bile duct obstruction Poor visualization of distal CBD	None	None	Initial procedure of choice in investigating possible biliary tract obstruction
Computed Tomography				
Simultaneous scanning of GB, liver, bile ducts, pancreas Accurate identification of dilated bile ducts, masses Not limited by jaundice, gas, obesity, ascites High-resolution image Guidance for fine-needle biopsy	Extreme cachexia Movement artifact Ileus Partial bile duct obstruction	Pregnancy	Reaction to iodinated contrast, if used	Indicated for evaluation of hepatic or pancreatic masses Procedure of choice in investigating possible biliary obstruction if diagnostic limitations prevent HBUS
Magnetic Resonance Cholangiopancreatography				
Useful modality for visualizing pancreatic and biliary ducts Has excellent sensitivity for bile duct dilatation, biliary stricture, and intraductal abnormalities Can identify pancreatic duct dilatation or stricture, pancreatic duct stenosis, and pancreas divisum	Cannot offer therapeutic intervention High cost	Claustrophobia Certain metals (iron)	None	
Endoscopic Retrograde Cholangiopancreatography				
Simultaneous	Gastroduodenal	Pregnancy	Pancreatitis	Cholangiogram

pancreatography Best visualization of distal biliary tract Bile or pancreatic cytology Endoscopic sphincterotomy and stone removal Biliary manometry	obstruction ? Roux-en-Y biliary-enteric anastomosis	? Acute pancreatitis ? Severe cardiopulmonary disease	Cholangitis, sepsis Infected pancreatic pseudocyst Perforation (rare) Hypoxemia, aspiration	of choice in: Absence of dilated ducts ? Pancreatic, ampullary or gastroduodenal disease Prior biliary surgery Endoscopic sphincterotomy a treatment possibility
Percutaneous Transhepatic Cholangiogram				
Extremely successful when bile ducts dilated Best visualization of proximal biliary tract Bile cytology/culture Percutaneous transhepatic drainage	Nondilated or sclerosed ducts	Pregnancy Uncorrectable coagulopathy Massive ascites ? Hepatic abscess	Bleeding Hemobilia Bile peritonitis Bacteremia, sepsis	Indicated when ERCP is contraindicated or failed
Endoscopic Ultrasound				
Most sensitive method to detect ampullary stones				

Abbreviations: CBD, common bile duct; ERCP, endoscopic retrograde cholangiopancreatography; GB, gallbladder; HBUS, hepatobiliary ultrasound.

The widespread use of laparoscopic cholecystectomy and ERCP has decreased the incidence of complicated biliary tract disease and the need for choledocholithotomy and T-tube drainage of the bile ducts. EBS followed by spontaneous passage or stone extraction is the treatment of choice in the management of patients with common duct stones, especially in elderly or poor-risk patients.

Trauma, Strictures, and Hemobilia

Most benign strictures of the extrahepatic bile ducts result from surgical trauma and occur in about 1 in 500 cholecystectomies. Strictures may present with bile leak or abscess formation in the immediate postoperative period or with biliary obstruction or cholangitis as long as 2 years or more following the inciting trauma. The diagnosis is established by percutaneous or endoscopic cholangiography. Endoscopic brushing of biliary strictures may be helpful in establishing the nature of the lesion and is more accurate than bile cytology alone. When positive exfoliative cytology is obtained, the diagnosis of a neoplastic stricture is established. This procedure is especially important in patients with primary sclerosing cholangitis (PSC) who are predisposed to the development of cholangiocarcinomas. Successful operative correction of non-PSC bile duct strictures by a skillful surgeon with duct-to-bowel anastomosis is usually possible, although mortality rates from surgical complications, recurrent cholangitis, or secondary biliary cirrhosis are high.

Hemobilia may follow traumatic or operative injury to the liver or bile ducts, intraductal rupture of a hepatic abscess or aneurysm of the hepatic artery, biliary or hepatic tumor hemorrhage, or mechanical complications of choledocholithiasis or hepatobiliary parasitism. Diagnostic procedures such as liver biopsy, PTC, and transhepatic biliary drainage catheter placement may also be complicated by hemobilia. Patients often present with a classic triad of biliary pain, obstructive jaundice, and melena or occult blood in the stools. The diagnosis is sometimes made by cholangiographic evidence of blood clot in the biliary tree, but selective angiographic verification may be required. Although minor episodes of hemobilia may resolve without operative intervention, surgical ligation of the bleeding vessel is frequently required.

Extrinsic Compression of the Bile Ducts

Partial or complete biliary obstruction may be produced by extrinsic compression of the ducts. The most common cause of this form of obstructive jaundice is carcinoma of the head of the pancreas. Biliary obstruction may also occur as a complication of either acute or chronic pancreatitis or involvement of lymph nodes in the porta hepatis by lymphoma or metastatic carcinoma. The latter should be distinguished from cholestasis resulting from massive replacement of the liver by tumor.

Hepatobiliary Parasitism

Infestation of the biliary tract by adult helminths or their ova may produce a chronic, recurrent pyogenic cholangitis with or without multiple hepatic abscesses, ductal stones, or biliary obstruction. This condition is relatively rare but does occur in inhabitants of southern China and elsewhere in Southeast Asia. The organisms most commonly involved are trematodes or flukes, including *Clonorchis sinensis*, *Opisthorchis viverrini* or *O. felineus*, and *Fasciola hepatica*. The biliary tract also may be involved by intraductal migration of adult *Ascaris lumbricoides* from the duodenum or by intrabiliary rupture of hydatid cysts of the liver produced by *Echinococcus* spp. The diagnosis is made by cholangiography and the presence of characteristic ova on stool examination. When obstruction is present, the treatment of choice is laparotomy under antibiotic coverage, with common duct exploration and a biliary drainage procedure.

Sclerosing Cholangitis

Primary or idiopathic sclerosing cholangitis is characterized by a progressive, inflammatory, sclerosing, and obliterative process affecting the extrahepatic and/or the intrahepatic bile ducts. The disorder occurs up to 75% in association with inflammatory bowel disease, especially ulcerative colitis. It may also be associated with autoimmune pancreatitis; multifocal fibrosclerosis syndromes such as retroperitoneal, mediastinal, and/or periureteral fibrosis; Riedel's struma; or pseudotumor of the orbit.

Immunoglobulin G4–associated cholangitis is a recently described biliary disease of unknown etiology that presents with biochemical and cholangiographic features indistinguishable from PSC, is often associated with autoimmune pancreatitis and other fibrosing conditions, and is characterized by elevated serum IgG4 and infiltration of IgG4-positive plasma cells in bile ducts and liver tissue. In contrast to PSC, it is not associated with inflammatory bowel disease and should be suspected if associated with increased serum IgG4 and unexplained pancreatic disease. Glucocorticoids are regarded as the initial treatment of choice. Relapse is common after steroid withdrawal especially with proximal strictures. Long-term treatment with glucocorticoids and/or azathioprine may be needed after relapse or for inadequate response **(Chap. 313)**.

Patients with primary sclerosing cholangitis often present with signs and symptoms of chronic or intermittent biliary obstruction: RUQ abdominal pain, pruritus, jaundice, or acute cholangitis. Late in the course, complete biliary obstruction, secondary biliary cirrhosis, hepatic failure, or portal hypertension with bleeding varices may occur. The diagnosis is usually established by finding multifocal, diffusely distributed strictures with intervening segments of normal or dilated ducts, producing a beaded appearance on cholangiography **(Fig. 311-2*D*)**. The cholangiographic techniques of choice in suspected cases are MRCP and ERCP. When a diagnosis of sclerosing cholangitis has been established, a search for associated diseases, especially for chronic inflammatory bowel disease, should be carried out.

A recent study describes the natural history and outcome for 305 patients of Swedish descent with primary sclerosing cholangitis; 134 (44%) of the patients were asymptomatic at the time of diagnosis and, not surprisingly, had a significantly higher survival rate. The independent predictors of a bad prognosis were age, serum bilirubin concentration, and liver histologic changes. Cholangiocarcinoma was found in 24 patients (8%). Inflammatory bowel disease was closely associated with primary sclerosing cholangitis and had a prevalence of 81% in this study population.

Small duct PSC is defined by the presence of chronic cholestasis and hepatic histology consistent with PSC but with normal findings on cholangiography. Small duct PSC is found in ~5% of patients with PSC and may represent an earlier stage of PSC associated with a significantly better long-term prognosis. However, such patients may progress to classic PSC and/or end-stage liver disease with consequent necessity of liver transplantation.

In patients with AIDS, cholangiopancreatography may demonstrate a broad range of biliary tract changes as well as pancreatic duct obstruction and occasionally pancreatitis **(Chap. 189)**. Further, biliary tract lesions in AIDS include infection and cholangiopancreatographic changes similar to those of PSC. Changes noted include: (1) diffuse involvement of intrahepatic bile ducts alone, (2) involvement of both intra- and extrahepatic bile ducts, (3) ampullary stenosis, (4) stricture of the intrapancreatic portion of the common bile duct, and (5) pancreatic duct involvement. Associated infectious organisms include *Cryptosporidium*, *Mycobacterium avium-intracellulare*, cytomegalovirus, *Microsporidia*, and *Isospora*. In addition, acalculous cholecystitis occurs in up to 10% of patients. ERCP sphincterotomy, while not without risk, provides significant pain reduction in patients with AIDS-associated papillary stenosis. Secondary sclerosing cholangitis may occur as a long-term complication of choledocholithiasis, cholangiocarcinoma, operative or traumatic biliary injury, or contiguous inflammatory processes.

TREATMENT: SCLEROSING CHOLANGITIS

Therapy with cholestyramine may help control symptoms of pruritus, and antibiotics are useful when cholangitis complicates the clinical picture. Vitamin D and calcium supplementation may help prevent the loss of bone mass frequently seen in patients with chronic cholestasis. Glucocorticoids, methotrexate, and cyclosporine have not been shown to be efficacious in PSC. UDCA in high dosage (20 mg/kg) improves serum liver tests, but an effect on survival has not been documented. In cases

where high-grade biliary obstruction (dominant strictures) has occurred, balloon dilatation or stenting may be appropriate. Only rarely is surgical intervention indicated. Efforts at biliary-enteric anastomosis or stent placement may, however, be complicated by recurrent cholangitis and further progression of the stenosing process. The prognosis is unfavorable, with a median survival of 9 to 12 years following the diagnosis, regardless of therapy. Four variables (age, serum bilirubin level, histologic stage, and splenomegaly) predict survival in patients with PSC and serve as the basis for a risk score. PSC is one of the most common indications for liver transplantation.

FURTHER READINGS

Apstein MD, Carey MC: Pathogenesis of cholesterol gallstones: A parsimonious hypothesis. Eur J Clin Invest 26:343, 1996[PMID: 8796361] [Full Text]

Ghazale A et al: Immunoglobuline G4-associated cholangitis: Clinical profile and response to therapy. Gastroenterol 134:706, 2008[PMID: 18222442] [Full Text]

Johansson M et al: Randomized clinical trial of open versus laparoscopic cholecystectomy for acute cholecystitis. Br J Surg 92:44, 2005[PMID: 15584058] [Full Text]

Lammert F, Miquel J-F: Gallstone disease: From genes to evidence-based therapy. J Hepatol 48:S124, 2008

Papi C et al: Timing of cholecystectomy for acute calculous cholecystitis: a meta-analysis. Am J Gastroenterol 99:147, 2004[PMID: 14687156] [Full Text]

Portincasa P et al: Cholesterol gallstone disease. Lancet 368:230, 2006[PMID: 16844493] [Full Text]

Ransohoff DF, Gracie WA: Treatment of gallstones. Ann Intern Med 119:606, 1993[PMID: 8363172] [Full Text]

Tischendorf JJW et al: Characterization, outcome, and prognosis in 273 patients with primary sclerosing cholangitis: a single center study. Am J Gastroenterol 102:107, 2007[PMID: 17037993] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH PANCREATIC DISEASE: INTRODUCTION

General Considerations

As emphasized in Chap. 313, the etiologies as well as the clinical manifestations of pancreatitis are quite varied. Although it is well-appreciated that pancreatitis is frequently secondary to biliary tract disease and alcohol abuse, it can also be caused by drugs, trauma, and viral infections and is associated with metabolic and connective tissue disorders. In ~30% of patients with acute pancreatitis and 25–40% of patients with chronic pancreatitis, the etiology initially can be obscure.

Although good data exist concerning the incidence of acute pancreatitis (about 5–35/100,000 new cases per year worldwide, with a mortality rate of about 3%), the number of patients who suffer with acute pancreatitis is largely increasing and is now estimated to be 70 hospitalizations/100,000 persons annually, resulting in >200,000 new cases of acute pancreatitis per year in the United States. Only one prospective study on the incidence of chronic pancreatitis is available; it showed an incidence of 8.2 new cases per 100,000 per year and a prevalence of 26.4 cases per 100,000. These numbers probably underestimate considerably the true incidence and prevalence, because non alcohol–induced pancreatitis has been largely ignored. At autopsy, the prevalence of chronic pancreatitis ranges from 0.04 to 5%. The relative inaccessibility of the pancreas to direct examination and the nonspecificity of the abdominal pain associated with pancreatitis make the diagnosis of pancreatitis difficult and usually dependent on elevation of blood amylase and/or lipase levels. Many patients with chronic pancreatitis do not have elevated blood amylase or lipase levels. Some patients with chronic pancreatitis develop signs and symptoms of pancreatic exocrine insufficiency, and, thus, objective evidence for pancreatic disease can be demonstrated. However, there is a very large reservoir of pancreatic exocrine function. More than 90% of the pancreas must be damaged before maldigestion of fat and protein is manifested. Noninvasive, indirect tests of pancreatic exocrine function (fecal elastase) are much more likely to give abnormal results in patients with obvious pancreatic disease (i.e., pancreatic calcification, steatorrhea, or diabetes mellitus, than in patients with occult disease). Thus, the number of patients who have subclinical exocrine dysfunction (<90% loss of function) is unknown.

Tests Useful in the Diagnosis of Pancreatic Disease

Several tests have proved of value in the evaluation of pancreatic disease. Examples of specific tests and their usefulness in the diagnosis of acute and chronic pancreatitis are summarized in **Table 312-1** and **Fig. 312-1**. At some institutions, pancreatic-function tests are available and performed if the diagnosis of pancreatic disease remains a possibility after noninvasive tests [ultrasound, CT, magnetic resonance cholangiopancreatography (MRCP)] or invasive tests [endoscopic retrograde cholangiopancreatography (ERCP), endoscopic ultrasonography (EUS)] have given normal or inconclusive results. In this regard, tests employing *direct* stimulation of the pancreas are the most sensitive.

Table 312-1 Tests Useful in the Diagnosis of Acute and Chronic Pancreatitis and Pancreatic Tumors

Test	Principle	Comment
Pancreatic Enzymes in Body Fluids		
Amylase		
1. Serum	Pancreatic inflammation leads to increased enzyme levels	Simple; 20–40% false negatives and positives; reliable if test results are three times the upper limit of normal
2. Urine	Renal clearance of amylase is increased in acute pancreatitis	Infrequently used
3. Ascitic fluid	Disruption of gland or main pancreatic duct leads to increased amylase concentration	Can help establish diagnosis of acute pancreatitis; false positives occur with intestinal obstruction and perforated ulcer
4. Pleural fluid	Exudative pleural effusion with pancreatitis	False positives occur with carcinoma of the lung and esophageal perforation
Serum lipase	Pancreatic inflammation leads to	New methods have greatly simplified

	increased enzyme levels	determination; positive in 70–85% of cases
Studies Pertaining to Pancreatic Structure		
Radiologic and radionuclide tests		
1. Plain film of the abdomen	Can be abnormal in acute and chronic pancreatitis	Simple; normal in >50% of cases of both acute and chronic pancreatitis
2. Upper GI x rays		Now obsolete
3. Ultrasonography (US)	Can provide information on edema, inflammation, calcification, pseudocysts, and mass lesions	Simple, noninvasive; sequential studies quite feasible; useful in diagnosis of pseudocyst limited by interference by bowel gas
4. CT scan	Permits detailed visualization of pancreas and surrounding structures, pancreatic fluid collection, pseudocyst, degree of necrosis	Useful in the diagnosis of pancreatic calcification, dilated pancreatic ducts, and pancreatic tumors; may not be able to distinguish between inflammatory and neoplastic mass lesions
5. Endoscopic retrograde cholangiopancreatography (ERCP)	Cannulation of pancreatic and common bile duct permits visualization of pancreatic-biliary ductal system	Can provide diagnostic data in 60–85% of cases; differentiation of chronic pancreatitis from pancreatic carcinoma may be difficult; now considered primarily a therapeutic procedure
6. Endoscopic ultrasonography (EUS)	High-frequency transducer employed with EUS can produce very high-resolution images and depict changes in the pancreatic duct and parenchyma with great detail	Can be used to assess chronic pancreatitis and pancreatic carcinoma
7. Magnetic resonance cholangiopancreatography	Three-dimensional rendering has been used to produce very good images of the pancreatic duct by a noninvasive technique	Has largely replaced ERCP as a diagnostic test
Pancreatic biopsy with US or CT guidance	Percutaneous aspiration biopsy with skinny needle and localization of lesion by US	High diagnostic yield; laparotomy avoided; can be done with EUS requires special technical skills
Tests of Exocrine Pancreatic Function		
Direct stimulation of the pancreas with analysis of duodenal contents		
1. Secretin-pancreozymin (CCK) test	Secretin leads to increased output of pancreatic juice and HCO_3^-; CCK leads to increased output of pancreatic enzymes; pancreatic secretory response is related to the functional mass of pancreatic tissue	Sensitive enough to detect occult disease; involves duodenal intubation and fluoroscopy; poorly defined normal enzyme response; overlap in chronic pancreatitis; large secretory reserve capacity of the pancreas, currently done at only a few medical centers
2. Endoscopic secretin—CCK test	Replaces need for tube placement duodenum	Sensitive enough to detect occult disease; avoids intubation and fluoroscopy; requires sedation
Measurement of intraluminal digestion products		
1. Microscopic examination of stool for undigested meat fibers and fat	Lack of proteolytic and lipolytic enzymes causes decreased digestion of meat fibers and triglycerides	Simple, reliable; not sensitive enough to detect milder cases of pancreatic insufficiency
2. Quantitative stool fat determination	Lack of lipolytic enzymes brings about impaired fat digestion	Reliable, reference standard for defining severity of malabsorption; does not distinguish between maldigestion and malabsorption
3. Fecal nitrogen	Lack of proteolytic enzymes leads to impaired protein digestion, resulting in an increase in stool nitrogen	Does not distinguish between maldigestion and malabsorption; low sensitivity
Measurement of pancreatic enzymes in feces		
1. Elastase	Pancreatic secretion of proteolytic enzymes; not degraded in intestine	Good sensitivity if stools not liquid

Abbreviation: CCK, cholecystokinin.

Figure 312-1

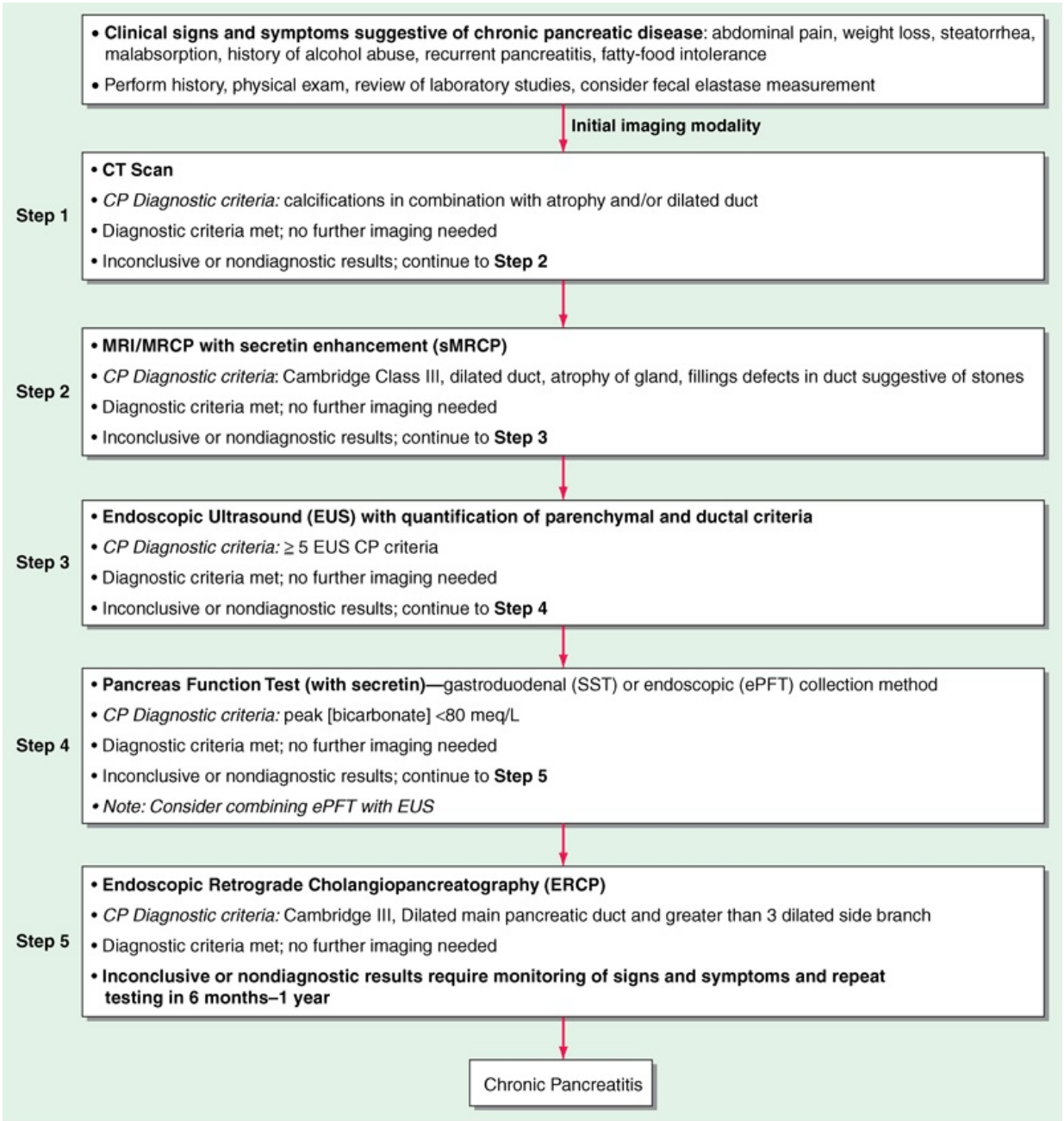

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A step-wise diagnostic approach to the patient with suspected chronic pancreatitis (CP). Endoscopic ultrasonography (EUS) and magnetic resonance cholangiopancreatography (sMRCP/MRCP) are appropriate diagnostic alternatives to endoscopic retrograde cholangiopancreatography (ERCP).

PANCREATIC ENZYMES IN BODY FLUIDS

The serum amylase and lipase levels are widely used as screening tests for acute pancreatitis in the patient with acute abdominal pain or back pain. Values greater than three times the upper limit of normal virtually clinch the diagnosis if gut perforation or infarction is excluded. In acute pancreatitis, the serum amylase and lipase are usually elevated within 24 hours of onset and remains so for 3–7 days. Levels usually return to normal within 7 days unless there is pancreatic ductal disruption, ductal obstruction, or pseudocyst formation. Approximately 85% of patients with acute pancreatitis have a threefold or greater elevated serum amylase and lipase levels. The values may be normal if (1) there is a delay (of 2–5 days) before blood samples are obtained, (2) the underlying disorder is chronic pancreatitis rather than acute pancreatitis, or (3) hypertriglyceridemia is present. Patients with

hypertriglyceridemia and proven pancreatitis have been found to have spuriously low levels of amylase and perhaps lipase activity. In the absence of objective evidence of pancreatitis by abdominal ultrasound, CT scan, MRCP, or EUS, mild to moderate elevations of amylase, and/or lipase are not helpful in making a diagnosis of chronic pancreatitis.

The serum amylase can be elevated in other conditions **(Table 312-2)**, in part because the enzyme is found in many organs. In addition to the pancreas and salivary glands, small quantities of amylase are found in the tissues of the fallopian tubes, lung, thyroid, and tonsils and can be produced by various tumors (carcinomas of the lung, esophagus, breast, and ovary). Urinary amylase measurements, including the amylase/creatinine clearance ratio, are no more sensitive or specific than blood amylase levels and are rarely employed. Isoamylase determinations do not accurately distinguish elevated blood amylase levels due to bona fide pancreatitis from elevated blood amylase levels due to a nonpancreatic source of amylase, especially when the blood amylase level is only moderately elevated.

Table 312-2 Causes of Hyperamylasemia and Hyperamylasuria

Pancreatic Disease	I. Pancreatitis A. Acute B. Chronic: ductal obstruction C. Complications of pancreatitis 1. Pancreatic pseudocyst 2. Pancreatogenous ascites 3. Pancreatic abscess Pancreatic necrosis II. Pancreatic trauma III. Pancreatic carcinoma
Nonpancreatic Disorders I. Renal insufficiency II. Salivary gland lesions A. Mumps B. Calculus C. Irradiation sialadenitis D. Maxillofacial surgery III. "Tumor" hyperamylasemia A. Carcinoma of the lung B. Carcinoma of the esophagus C. Breast carcinoma, ovarian carcinoma IV. Macroamylasemia V. Burns VI. Diabetic ketoacidosis VII. Pregnancy VIII. Renal transplantation IX. Cerebral trauma X. Drugs: morphine	
Other Abdominal Disorders I. Biliary tract disease: cholecystitis, choledocholithiasis II. Intraabdominal disease A. Perforated or penetrating peptic ulcer B. Intestinal obstruction or infarction C. Ruptured ectopic pregnancy D. Peritonitis E. Aortic aneurysm F. Chronic liver disease	

G. Postoperative hyperamylasemia

Elevation of ascitic fluid amylase occurs in acute pancreatitis as well as in (1) pancreatogenous ascites due to disruption of the main pancreatic duct or a leaking pseudocyst and (2) other abdominal disorders that simulate pancreatitis (e.g., intestinal obstruction, intestinal infarction, or perforated peptic ulcer). Elevation of pleural fluid amylase can occur in acute pancreatitis, chronic pancreatitis, carcinoma of the lung, and esophageal perforation.

Lipase may now be the single best enzyme to measure for the diagnosis of acute pancreatitis. Improvements in substrates and technology offer clinicians improved options, especially when a turbidimetric assay is used. The newer lipase assays have colipase as a cofactor and are fully automated.

No single blood test is reliable for the diagnosis of acute pancreatitis in patients with renal failure. Determining whether a patient with renal failure and abdominal pain has pancreatitis remains a difficult clinical problem. One study found that serum amylase levels were elevated in patients with renal dysfunction only when creatinine clearance was <0.8 mL/s (<50 mL/min). In such patients, the serum amylase level was invariably <8.3 μkat/L (<500 IU/L) in the absence of objective evidence of acute pancreatitis. In that study, serum lipase and trypsin levels paralleled serum amylase values. With these limitations in mind, the recommended screening tests for acute pancreatitis are *serum lipase* and *serum amylase levels*. Serum lipase and amylase values greater than three times normal are highly specific.

STUDIES PERTAINING TO PANCREATIC STRUCTURE

Radiologic Tests

Plain films of the abdomen, which once provided useful information in patients with acute and chronic pancreatitis, have been superceded by other detailed imaging procedures (US, EUS, CT, MRCP).

Ultrasonography can provide important information in patients with acute pancreatitis, chronic pancreatitis, pseudocysts, and pancreatic carcinoma. Echographic appearances can indicate the presence of edema, inflammation, and calcification (not obvious on plain films of the abdomen), as well as pseudocysts, mass lesions, and gallstones. In acute pancreatitis, the pancreas is characteristically enlarged. In pancreatic pseudocyst, the usual appearance is primarily that of smooth, round fluid collection. Pancreatic carcinoma distorts the usual landmarks, and mass lesions >3.0 cm are usually detected as localized, solid lesions. Ultrasound is often the initial investigation for most patients with suspected pancreatic disease. However, obesity, excess small- and large-bowel gas can interfere with pancreatic imaging by ultrasound studies.

CT is the best imaging study for initial evaluation of a suspected pancreatic disorder and for the complications of acute and chronic pancreatitis. It is especially useful in the detection of pancreatic and peripancreatic acute fluid collections, fluid-containing lesions such as pseudocysts, walled-off necrosis, calcium deposits (see Chapter 313, Figs. 313-1, 313-2, and 313-4), and pancreatic neoplasms. Most lesions are characterized by (1) enlargement of the pancreatic outline, (2) distortion of the pancreatic contour, and/or (3) a fluid filling that has a different attenuation coefficient than normal pancreas. Oral, water-soluble contrast agents are used to opacify the stomach and duodenum during CT scans; this strategy permits more precise delineation of various organs as well as mass lesions. Dynamic CT (using rapid IV administration of contrast) is useful in estimating the extent of pancreatic necrosis and in predicting morbidity and mortality. Spiral (helical) CT provides clear images much more rapidly and essentially negates artifact caused by patient movement.

EUS produces high-resolution images of the pancreatic parenchyma and pancreatic duct with a transducer fixed to an endoscope that can be directed onto the surface of the pancreas through the stomach or duodenum. EUS and MRCP have largely replaced ERCP for diagnostic purposes in many centers. EUS allows one to obtain information about the pancreatic duct as well as the parenchyma and has few procedure-related complications associated with it, in contrast to the 5–20% of post-ERCP pancreatitis observed. EUS is also helpful in detecting common bile duct stones. Pancreatic masses can be biopsied via EUS and one can deliver nerve-blocking agents through EUS fine-needle injection. Criteria for abnormalities on EUS in severe chronic pancreatic disease have been developed. Currently, chronic pancreatitis is considered diagnosed by EUS if five or more criteria listed in **Table 312-3** are present. Recent studies comparing EUS and ERCP to the secretin test in patients with unexplained abdominal pain suspected of having chronic pancreatitis show equivalent diagnostic accuracy in detecting early changes of chronic pancreatitis. The exact role of EUS versus CT, ERCP, or function testing in the early diagnosis of chronic pancreatitis has yet to be clearly defined.

Table 312-3 Endoscopic Ultrasonographic Criteria for Chronic Pancreatitis

Ductal	Parenchymal
Stones	Echogenic strands

Echogenic ductal walls	Echogenic foci
Irregular ductal walls	Calcifications
Stricture	Lobular contour
Visible side branches Ductular dilatation	Cyst

MRCP/ MRI is now being used to view the bile ducts, pancreatic duct, and the pancreas parenchyma. Non breath–holding and three-dimensional turbo spin-echo techniques are being used to produce superb MRCP images. The main pancreatic duct and common bile duct can be seen well, but there is still a question as to whether changes can be detected consistently in the secondary ducts. The secondary ducts are not visualized in a normal pancreas. MRCP may be particularly useful to evaluate the pancreatic duct in high-risk patients such as the elderly because this is a noninvasive procedure. Secretin enhanced MRCP is currently under investigation but is emerging as a method to better evaluate ductal changes.

Both EUS and MRCP have largely replaced diagnostic ERCP in most patients. As these techniques become more refined, they may well be the diagnostic tests of choice to evaluate the pancreatic duct. ERCP is still needed for treatment of bile duct and pancreatic duct lesions. ERCP is primarily of therapeutic value after CT, EUS, or MRCP have detected abnormalities requiring invasive endoscopic treatment. ERCP can also be helpful at clarification of equivocal findings discovered with other imaging techniques (see Chap. 313, Figs. 313-1*C*, 313-3*D*, and 313-4*B*). Pancreatic carcinoma is characterized by stenosis or obstruction of either the pancreatic duct or the common bile duct; both ductal systems are often abnormal. In chronic pancreatitis, ERCP abnormalities include (1) luminal narrowing; (2) irregularities in the ductal system with stenosis, dilation, sacculation, and ectasia; and (3) blockage of the pancreatic duct by calcium deposits. The presence of ductal stenosis and irregularity can make it difficult to distinguish chronic pancreatitis from carcinoma. It is important to be aware that ERCP changes interpreted as indicating chronic pancreatitis actually may be due to the effects of aging on the pancreatic duct or to the fact that the procedure was performed within several weeks of an attack of acute pancreatitis. Although aging may cause impressive ductal alterations, it does not affect the results of pancreatic function tests (i.e., the secretin test). Elevated serum amylase levels after ERCP have been reported in 25–75% of patients, and clinical pancreatitis in 5–20% of patients. There are no satisfactory means to pharmacologically prevent ERCP-induced pancreatitis, despite many agents such as octreotide and nitroglycerin having been suggested and evaluated. The best way to prevent ERCP-induced pancreatitis is to not perform this procedure for diagnostic purposes in high-risk patients, especially in women with acute relapsing pancreatitis in whom there is no evidence of biliary obstruction and patients with unexplained abdominal pain but no other abnormalities. If no lesion is found in the biliary and/or pancreatic ducts in a patient with repeated attacks of acute pancreatitis, manometric studies of the sphincter of Oddi may be indicated. Such studies, however, do increase the risk of post-ERCP/manometry acute pancreatitis. Such pancreatitis appears to be more common in patients with a nondilated pancreatic duct.

Pancreatic Biopsy with Radiologic Guidance

Percutaneous aspiration biopsy or a trucut biopsy of a pancreatic mass often distinguishes a pancreatic inflammatory mass from a pancreatic neoplasm.

Tests of Exocrine Pancreatic Function

Pancreatic function tests (Table 312-1) can be divided into the following:

1. *1. Direct stimulation of the pancreas* by IV infusion of secretin or secretin plus cholecystokinin (CCK) followed by collection and measurement of duodenal contents
2. Study of *intraluminal digestion products*, such as undigested meat fibers, stool fat, and fecal nitrogen
3. *3. Measurement of fecal pancreatic enzymes* such as elastase

The secretin test, used to detect diffuse pancreatic disease, is based on the physiologic principle that the pancreatic secretory response is directly related to the functional mass of pancreatic tissue. In the standard assay, secretin is given IV in a dose of 0.2 μg/kg of synthetic human secretin as a bolus. Normal values for the standard secretin test are (1) volume output >2 mL/kg per hour, (2) bicarbonate (HCO_3^-) concentration >80 mmol/L, and (3) HCO_3^- output >10 mmol/L in 1 hour. The most reproducible measurement, giving the highest level of discrimination between normal subjects and patients with chronic pancreatic exocrine insufficiency, appears to be the maximal bicarbonate concentration.

There may be a dissociation between the results of the secretin test and other tests of absorptive function. For example, patients with chronic pancreatitis often have abnormally low outputs of HCO_3^-

after secretin but have normal fecal fat excretion. Thus the secretin test measures the secretory capacity of ductular epithelium, while fecal fat excretion indirectly reflects intraluminal lipolytic activity. Steatorrhea does not occur until intraluminal levels of lipase are markedly reduced, underscoring the fact that only small amounts of enzymes are necessary for intraluminal digestive activities. It must be noted that, an abnormal secretin test result suggests only that chronic pancreatic damage is present.

Measurement of *intraluminal digestion products* (i.e., undigested muscle fibers, stool fat, and fecal nitrogen) is discussed in Chap. 294. The amount of human elastase in stool reflects the pancreatic output of this proteolyticenzyme. Decreased elastase activity in stool is an excellent test to detect severe pancreatic exocrine insufficiency in patients with chronic pancreatitis and cystic fibrosis provided that the stool specimen is solid.

Tests useful in the diagnosis of exocrine pancreatic insufficiency and the differential diagnosis of malabsorption are also discussed in Chaps. 294 and 313.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BIOCHEMISTRY AND PHYSIOLOGY OF PANCREATIC EXOCRINE SECRETION

General Considerations

The pancreas secretes 1500–3000 mL of isosmotic alkaline (pH >8) fluid per day containing about 20 enzymes. The pancreatic secretions provide the enzymes needed to effect the major digestive activity of the gastrointestinal tract and provide an optimal pH for the function of these enzymes.

Regulation of Pancreatic Secretion

The exocrine pancreas is influenced by intimately interacting hormonal and neural systems. *Gastric acid* is the stimulus for the release of secretin from the duodenum, which stimulates the secretion of water and electrolytes from pancreatic ductal cells. Release of cholecystokinin (CCK) from the duodenum and proximal jejunum is largely triggered by long-chain fatty acids, certain essential amino acids (tryptophan, phenylalanine, valine, methionine), and gastric acid itself. CCK evokes an enzyme-rich secretion from acinar cells in the pancreas. The *parasympathetic nervous system* (via the vagus nerve) exerts significant control over pancreatic secretion. Secretion evoked by secretin and CCK depends on permissive roles of vagal afferent and efferent pathways. This is particularly true for enzyme secretion, whereas water and bicarbonate secretions are heavily dependent on the hormonal effects of secretin and to a lesser extent CCK. Also, vagal stimulation effects the release of vasoactive intestinal peptide (VIP), a secretin agonist.

Pancreatic exocrine secretion is influenced by inhibitory neuropeptides such as somatostatin, pancreatic polypeptide, peptide YY, neuropeptide Y, enkephalin, pancreastatin, calcitonin gene–related peptides, glucagon, and galanin. Although pancreatic polypeptide and peptide YY may act primarily on nerves outside the pancreas, somatostatin acts at multiple sites. Nitric oxide (NO) is also an important neurotransmitter. The mechanism of action of these various factors has not been fully defined.

Water and Electrolyte Secretion

Bicarbonate is the ion of primary physiologic importance within pancreatic secretion. The ductal cells secrete bicarbonate predominantly derived from plasma (93%) more than from intracellular metabolism (7%). Bicarbonate enters through the sodium bicarbonate cotransporter with depolarization caused by chloride efflux through the cystic fibrosis transmembrane conductance regulator (CFTR). Secretin and VIP, both of which increase intracellular cyclic AMP, act on the ductal cells opening the CFTR in promoting secretion. CCK, acting as a neuromodulator, markedly potentiates the stimulatory effects of secretin. Acetylcholine also plays an important role in ductal cell secretion. Bicarbonate helps neutralize gastric acid and creates the appropriate pH for the activity of pancreatic enzymes and bile salts.

Enzyme Secretion

The acinar cell is highly compartmentalized and is concerned with the secretion of pancreatic enzymes. Proteins synthesized by the rough endoplasmic reticulum are processed in the Golgi and then targeted to the appropriate site, whether that be zymogen granules, lysosomes, or other cell compartments. The pancreas secretes amylolytic, lipolytic, and proteolytic enzymes. *Amylolytic enzymes* such as amylase, hydrolyze starch to oligosaccharides and to the disaccharide maltose. The *lipolytic enzymes* include lipase, phospholipase A_2, and cholesterol esterase. Bile salts inhibit lipase in isolation, but colipase, another constituent of pancreatic secretion, binds to lipase and prevents this inhibition. Bile salts activate phospholipase A and cholesterol esterase. *Proteolytic enzymes* include endopeptidases (trypsin, chymotrypsin), which act on internal peptide bonds of proteins and polypeptides; exopeptidases (carboxypeptidases, aminopeptidases), which act on the free carboxyl- and amino-terminal ends of peptides, respectively; and elastase. The proteolytic enzymes are secreted as inactive precursors and packaged as zymogens. Ribonucleases (deoxyribonucleases, ribonuclease) are also secreted. *Enterokinase*, an enzyme found in the duodenal mucosa, cleaves the lysine-isoleucine bond of trypsinogen to form trypsin. Trypsin then activates the other proteolytic zymogens and phospholipase A_2 in a cascade phenomenon. All pancreatic enzymes have pH optima in the alkaline range. The nervous system initiates pancreatic enzyme secretion. The neurologic stimulation is cholinergic, involving extrinsic innervation by the vagus nerve and subsequent innervation by intrapancreatic cholinergic nerves. The stimulatory neurotransmitters are acetylcholine and gastrin-releasing peptides. These neurotransmitters activate calcium-dependent second messenger systems, resulting in the release of zymogen granules. VIP is present in intrapancreatic

nerves and potentiates the effect of acetylcholine. In contrast to other species, there are no CCK receptors on acinar cells in humans. CCK in physiologic concentrations stimulates pancreatic secretion by stimulating afferent vagal and intrapancreatic nerves.

Autoprotection of the Pancreas

Autodigestion of the pancreas is prevented by the packaging of pancreatic proteases in precursor form and by the synthesis of protease inhibitor [i.e., pancreatic secretory trypsin inhibitor (PSTI) or SPINK1], which can bind and inactivate about 20% of trypsin activity. Mesotrypsin, chymotrypsin c, and enzyme y can also lyse and inactivate trypsin. These protease inhibitors are found in the acinar cell, the pancreatic secretions, and the $_1$- and $_2$-globulin fractions of plasma. In addition, low calcium concentration within the cytosol of acinar cells in the normal pancreas promotes the destruction of spontaneously activated trypsin. Loss of any of these protective mechanisms leads to zymogen activation, autodigestion, and acute pancreatitis.

Exocrine-Endocrine Relationships

Insulin appears to be needed locally for secretin and CCK to promote exocrine secretion; thus, it acts in a permissive role for these two hormones.

Enteropancreatic Axis and Feedback Inhibition

Pancreatic enzyme secretion is controlled, at least in part, by a negative feedback mechanism induced by the presence of active serine proteases in the duodenum. To illustrate, perfusion of the duodenal lumen with phenylalanine causes a prompt result in increased plasma CCK levels as well as increased secretion of chymotrypsin and other pancreatic enzymes. However, simultaneous perfusion with trypsin blunts both responses. Conversely, perfusion of the duodenal lumen with protease inhibitors actually leads to enzyme hypersecretion. The available evidence supports the concept that the duodenum contains a peptide called *CCK-releasing factor* (CCK-RF) that is involved in stimulating CCK release. It appears that serine proteases inhibit pancreatic secretion by inactivating a CCK-releasing peptide in the lumen of the small intestine. Thus, the integrative result of both bicarbonate and enzyme secretion depends on a feedback process for both bicarbonate and pancreatic enzymes. Acidification of the duodenum releases secretin, which stimulates vagal and other neural pathways to activate pancreatic duct cells, which secrete bicarbonate. This bicarbonate then neutralizes the duodenal acid, and the feedback loop is completed. Dietary proteins bind proteases, thereby leading to an increase in free CCK-RF. CCK is then released into the blood in physiologic concentrations, acting primarily through the neural pathways (vagal-vagal). This leads to acetylcholine-mediated pancreatic enzyme secretion. Proteases continue to be secreted from the pancreas until the protein within the duodenum is digested. At this point, pancreatic protease secretion is reduced to basic levels, thus completing this step in the feedback process.

ACUTE PANCREATITIS

General Considerations

Pancreatic inflammatory disease may be classified as (1) acute pancreatitis or (2) chronic pancreatitis. The pathologic spectrum of acute pancreatitis varies from *interstitial pancreatitis*, which is usually a mild and self-limited disorder, to *necrotizing pancreatitis*, in which the extent of pancreatic necrosis may correlate with the severity of the attack and its systemic manifestations.

The incidence of pancreatitis varies in different countries and depends on cause [e.g., alcohol, gallstones, metabolic factors, and drugs **(Table 313-1)**]. The estimated incidence in the United States is increasing and is now estimated to be 70 hospitalizations/100,000 persons annually, thus resulting in >200,000 new cases of acute pancreatitis per year.

Table 313-1 Causes of Acute Pancreatitis

Common Causes
Gallstones (including microlithiasis)
Alcohol (acute and chronic alcoholism) Hypertriglyceridemia
Endoscopic retrograde cholangiopancreatography (ERCP), especially after biliary manometry
Trauma (especially blunt abdominal trauma)
Postoperative (abdominal and nonabdominal operations)
Drugs (azathioprine, 6-mercaptopurine, sulfonamides, estrogens, tetracycline, valproic acid, anti-HIV medications)
Sphincter of Oddi dysfunction
Uncommon Causes
Vascular causes and vasculitis (ischemic-hypoperfusion states after cardiac surgery)
Connective tissue disorders and thrombotic thrombocytopenic

purpura (TTP)
Cancer of the pancreas
Hypercalcemia
Periampullary diverticulum
Pancreas divisum
Hereditary pancreatitis
Cystic fibrosis
Renal failure
Rare Causes
Infections (mumps, coxsackievirus, cytomegalovirus, echovirus, parasites)
Autoimmune (e.g., Sjögren's syndrome)
Causes to Consider in Patients with Recurrent Bouts of Acute Pancreatitis without an Obvious Etiology
Occult disease of the biliary tree or pancreatic ducts, especially microlithiasis, sludge
Drugs
Hypertriglyceridemia
Pancreas divisum
Pancreatic cancer
Sphincter of Oddi dysfunction
Cystic fibrosis
Idiopathic

Etiology and Pathogenesis

There are many causes of acute pancreatitis **(Table 313-1)**, but the mechanisms by which these conditions trigger pancreatic inflammation have not been fully elucidated. Gallstones continue to be the leading cause of acute pancreatitis in most series (30–60%). The risk of acute pancreatitis in patients with at least one gallstone <5 mm in diameter is fourfold greater than that in patients with larger stones. Alcohol is the second most common cause, responsible for 15–30% of cases in the United States. The incidence of pancreatitis in alcoholics is surprisingly low (5/100,000), indicating that in addition to the amount of alcohol ingested unknown factors affect a person's susceptibility to pancreatic injury. The mechanism of injury is incompletely understood. Acute pancreatitis occurs in 5–20% of patients following endoscopic retrograde cholangiopancreatography (ERCP). Despite extensive research into the medical and endoscopic prevention of post-ERCP pancreatitis, there has been little decline in incidence. Use of prophylactic pancreatic duct stent after retrograde pancreatogram or pancreatic sphincterotomy has shown promise in reducing pancreatitis but requires further prospective evaluation. Risk factors for post-ERCP pancreatitis include minor papilla sphincterotomy, sphincter of Oddi dysfunction, prior history of post-ERCP pancreatitis, age <60 years, >2 contrast injections into the pancreatic duct, and endoscopic trainee involvement.
Hypertriglyceridemia is the cause of acute pancreatitis in 1.3–3.8% of cases; serum triglyceride levels are usually >11.3 mmol/L (>1000 mg/dL). Most patients with hypertriglyceridemia, when subsequently examined, show evidence of an underlying derangement in lipid metabolism, probably unrelated to pancreatitis. Such patients are prone to recurrent episodes of pancreatitis. Any factor (e.g., drugs or alcohol) that causes an abrupt increase in serum triglycerides to levels >11 mmol/L (1000 mg/dL) can precipitate a bout of acute pancreatitis. Finally, patients with a deficiency of apolipoprotein CII have an increased incidence of pancreatitis; apolipoprotein CII activates lipoprotein lipase, which is important in clearing chylomicrons from the bloodstream. Patients with diabetes mellitus who have developed ketoacidosis and patients who are on certain medications such as oral contraceptives may also develop high triglyceride levels. Approximately 2–5% of cases of acute pancreatitis are drug related. Drugs cause pancreatitis either by a hypersensitivity reaction or by the generation of a toxic metabolite, although in some cases it is not clear which of these mechanisms is operative (Table 313-1).

Autodigestion is a currently accepted pathogenic theory; according to it, pancreatitis results when proteolytic enzymes (e.g., trypsinogen, chymotrypsinogen, proelastase, and lipolytic enzymes such as phospholipase A_2) are activated in the pancreas rather than in the intestinal lumen. A number of factors (e.g., endotoxins, exotoxins, viral infections, ischemia, anoxia, lysosomal calcium, and direct trauma) are believed to facilitate activation of trypsin. Activated proteolytic enzymes, especially trypsin, not only digest pancreatic and peripancreatic tissues but also can activate other enzymes, such as elastase and phospholipase A_2. Spontaneous activation of trypsin also can occur.

Activation of Pancreatic Enzymes in the Pathogenesis of Acute Pancreatitis

Several recent studies have suggested that pancreatitis is a disease that evolves in three phases. The

initial phase is characterized by intrapancreatic digestive enzyme activation and acinar cell injury. Trypsin activation appears to be mediated by lysosomal hydrolases such as cathepsin B that become colocalized with digestive enzymes in intracellular organelles; it is currently believed that acinar cell injury is the consequence of trypsin activation. The second phase of pancreatitis involves the activation, chemoattraction, and sequestration of leukocytes and macrophages in the pancreas, resulting in an enhanced intrapancreatic inflammatory reaction. Neutrophil depletion induced by prior administration of an antineutrophil serum has been shown to reduce the severity of experimentally induced pancreatitis. There is also evidence to support the concept that neutrophil sequestration can activate trypsinogen. Thus, intrapancreatic acinar cell activation of trypsinogen could be a two-step process (i.e., an early neutrophil-independent and a later neutrophil-dependent phase). The third phase of pancreatitis is due to the effects of activated proteolytic enzymes and cytokines, released by the inflamed pancreas, on distant organs. Activated proteolytic enzymes, especially trypsin, not only digest pancreatic and peripancreatic tissues but also activate other enzymes such as elastase and phospholipase A_2. The active enzymes and cytokines then digest cellular membranes and cause proteolysis, edema, interstitial hemorrhage, vascular damage, coagulation necrosis, fat necrosis, and parenchymal cell necrosis. Cellular injury and death result in the liberation of bradykinin peptides, vasoactive substances, and histamine that can produce vasodilation, increased vascular permeability, and edema with profound effects on many organs, most notably the lung. The systemic inflammatory response syndrome (SIRS) and acute respiratory distress syndrome (ARDS) as well as multiorgan failure may occur as a result of this cascade of local as well as distant effects.

There appear to be a number of genetic factors that can increase the susceptibility and/or modify the severity of pancreatic injury in acute pancreatitis. Four susceptibility genes have been identified: (1) cationic trypsinogen mutations (PRSS1m, R122Hm, and N291), (2) pancreatic secretory trypsin inhibitor (SPINK1), (3) CFTR, and (4) monocyte chemotactic protein (MCP-1). Experimental and clinical data indicate that MCP-1 may be an important inflammatory mediator in the early pathologic process of acute pancreatitis, a determinant of the severity of the inflammatory response, and a promoter of organ failure.

Approach to the Patient: Abdominal Pain

Abdominal pain is the major symptom of acute pancreatitis. Pain may vary from a mild and tolerable discomfort and more commonly to severe, constant, and incapacitating distress. Characteristically, the pain, which is steady and boring in character, is located in the epigastrium and periumbilical region and often radiates to the back as well as to the chest, flanks, and lower abdomen. The pain is frequently more intense when the patient is supine, and patients may obtain some relief by sitting with the trunk flexed and knees drawn up. Nausea, vomiting, and abdominal distention due to gastric and intestinal hypomotility and chemical peritonitis are also frequent complaints.

Physical examination frequently reveals a distressed and anxious patient. Low-grade fever, tachycardia, and hypotension are fairly common. Shock is not unusual and may result from (1) hypovolemia secondary to exudation of blood and plasma proteins into the retroperitoneal space and a "retroperitoneal burn" due to activated proteolytic enzymes; (2) increased formation and release of kinin peptides, which cause vasodilation and increased vascular permeability; and (3) systemic effects of proteolytic and lipolytic enzymes released into the circulation. Jaundice occurs infrequently; when present, it usually is due to edema of the head of the pancreas with compression of the intrapancreatic portion of the common bile duct. Erythematous skin nodules due to subcutaneous fat necrosis may occur. In 10–20% of patients, there are pulmonary findings, including basilar rales, atelectasis, and pleural effusion, the latter most frequently left sided. Abdominal tenderness and muscle rigidity are present to a variable degree, but, compared with the intense pain, these signs may be unimpressive. Bowel sounds are usually diminished or absent. An enlarged pancreas with walled off necrosis or a pseudocyst may be palpable in the upper abdomen later in the disease course (i.e., four to six weeks). A faint blue discoloration around the umbilicus (Cullen's sign) may occur as the result of hemoperitoneum, and a blue-red-purple or green-brown discoloration of the flanks (Turner's sign) reflects tissue catabolism of hemoglobin. The latter two findings, which are uncommon, indicate the presence of a severe necrotizing pancreatitis.

Laboratory Data

The diagnosis of acute pancreatitis is usually established by the detection of an increased level of serum amylase and lipase. Values threefold or more above normal virtually clinch the diagnosis if gut perforation, ischemia, and infarction are excluded. However, there appears to be no definite correlation between the severity of pancreatitis and the degree of serum lipase and amylase elevations. After three to seven days, even with continuing evidence of pancreatitis, total serum amylase values tend to return toward normal. However, pancreatic isoamylase and lipase levels may remain elevated for 7 to 14 days. It will be recalled that amylase elevations in serum and urine occur in many conditions other than pancreatitis **(see Chap. 312, Table 312-2)**. Importantly, patients with *acidemia* (arterial pH ≤7.32) may have spurious elevations in serum amylase. In one study, 12 of 33 patients with acidemia had elevated serum amylase, but only 1 had an elevated lipase value; in 9, salivary-type amylase was the predominant serum isoamylase. This finding explains why patients with diabetic ketoacidosis may have

marked elevations in serum amylase without any other evidence of acute pancreatitis. Serum lipase activity increases in parallel with amylase activity. A threefold elevated serum lipase value is usually diagnostic of acute pancreatitis; these tests are especially helpful in patients with nonpancreatic causes of hyperamylasemia **(see Chap. 312, Table 312-2)**.

Table 313-2 Severe Acute Pancreatitis

Risk Factors for Severity

- Age >60 years
- Obesity, BMI >30
- Comorbid disease

Markers of Severity within 24 Hours

- SIRS [temperature >38° or <36°C (>100.4° or 96.8°F), Pulse >90, Tachypnea >24, ↑ WBC >12,000]
- Hemoconcentration (Hct >44%)
- BISAP
  - (B) Blood urea nitrogen (BUN) >22 mg%
  - (I) Impaired mental status
  - (S) SIRS: 2/4 present
  - (A) Age >60 years
  - (P) Pleural effusion
- Organ Failure
  - Cardiovascular: systolic BP <90 mmHg, heartrate >130
  - Pulmonary: Pao_2 <60 mmHg
  - Renal serum creatinine >2.0 mg%

Markers of Severity during Hospitalization

- Persistent organ failure
- Pancreatic necrosis
- Hospital-acquired infection

Abbreviation: BISAP, bedside index of severity in acute pancreatitis.

Leukocytosis (15,000–20,000 leukocytes per μL) occurs frequently. Patients with more severe disease may show hemoconcentration with hematocrit values >44% and/or azotemia with a blood urea nitrogen (BUN) level >22 mg/dL because of loss of plasma into the retroperitoneal space and peritoneal cavity. Hemoconcentration may be the harbinger of more severe disease (i.e., pancreatic necrosis), while azotemia is a significant risk factor for mortality. *Hyperglycemia* is common and is due to multiple factors, including decreased insulin release, increased glucagon release, and an increased output of adrenal glucocorticoids and catecholamines. *Hypocalcemia* occurs in ~25% of patients, and its pathogenesis is incompletely understood. Although earlier studies suggested that the response of the parathyroid gland to a decrease in serum calcium is impaired, subsequent observations have failed to confirm this phenomenon. Intraperitoneal saponification of calcium by fatty acids in areas of fat necrosis occurs occasionally, with large amounts (up to 6.0 g) dissolved or suspended in ascitic fluid. Such "soap formation" may also be significant in patients with pancreatitis, mild hypocalcemia, and little or no obvious ascites. *Hyperbilirubinemia* [serum bilirubin >68 μmol/L (>4.0 mg/dL)] occurs in ~10% of patients. However, jaundice is transient, and serum bilirubin levels return to normal in four to seven days. Serum alkaline phosphatase and aspartate aminotransferase levels are also transiently elevated and they parallel serum bilirubin values and may point to gallbladder-related disease. Markedly elevated serum lactic dehydrogenase levels [>8.5 μmol/L (>500 U/dL)] suggest a poor prognosis. *Hypertriglyceridemia* occurs in 5–10% of patients, and serum amylase levels in these individuals are often spuriously normal (Chap. 312). Approximately 5–10% of patients have *hypoxemia* (arterial $Po_2 \leq$ 60 mmHg), which may herald the onset of ARDS. Finally, the electrocardiogram is occasionally abnormal in acute pancreatitis with ST-segment and T-wave abnormalities simulating myocardial ischemia.

A CT scan can confirm the clinical impression of acute pancreatitis even with less than a threefold increase in serum amylase and lipase levels. Importantly, CT can be helpful in indicating the severity of acute pancreatitis and the risk of morbidity and mortality and in evaluating the complications of acute

pancreatitis **(Table 313-3**). However, a CT scan obtained within the first several days of symptom onset may underestimate the extent of tissue injury. What may appear to be intestinal pancreatitis on initial CT scan may evolve to pancreatic necrosis on repeat CT scan three to five days later (**Fig. 313-1**). Sonography is useful in acute pancreatitis to evaluate the gallbladder if gallstone disease is suspected. **Radiologic studies useful in the diagnosis of acute pancreatitis are discussed in Chap. 312, and listed in Table 312-1, and depicted in Figs. 313-1 to 313-3**.

Table 313-3 CT Findings and Grading of Acute Pancreatitis [CT Severity Index (Ctsi)]

Grade	Findings	Score
A	Normal pancreas: normal size, sharply defined, smooth contour, homogeneous enhancement, retroperitoneal peripancreatic fat without enhancement	0
B	Focal or diffuse enlargement of the pancreas, contour may show irregularity, enhancement may be inhomogeneous but there is no peripancreatic inflammation	1
C	Peripancreatic inflammation with intrinsic pancreatic abnormalities	2
D	Intrapancreatic or extrapancreatic fluid collections	3
E	Two or more large collections or gas in the pancreas or retroperitoneum	4
Necrosis score based on contrast-enhanced CT		
Necrosis, %		Score
0		0
<33		2
33–50		4
≥50		6

Note: CT severity index equals unenhanced CT score plus necrosis score: maximum = 10; ≥6 = severe disease.

Source: Modified from EJ Balthazar et al: Radiology 1990;174:331.

Figure 313-1

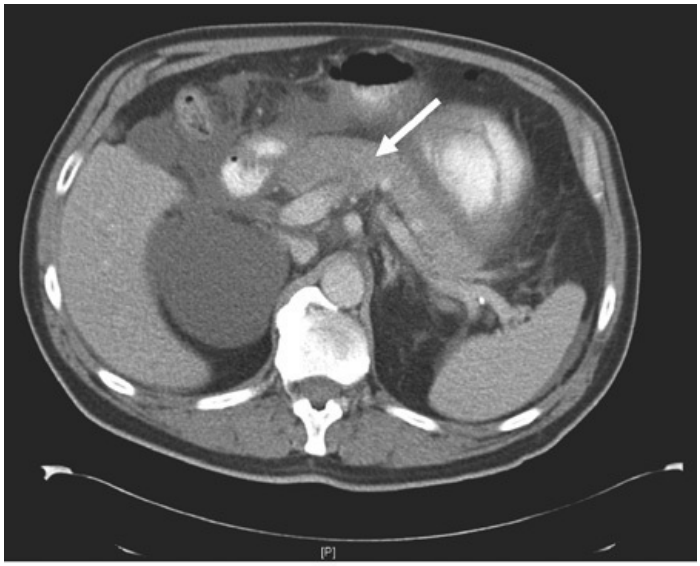

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

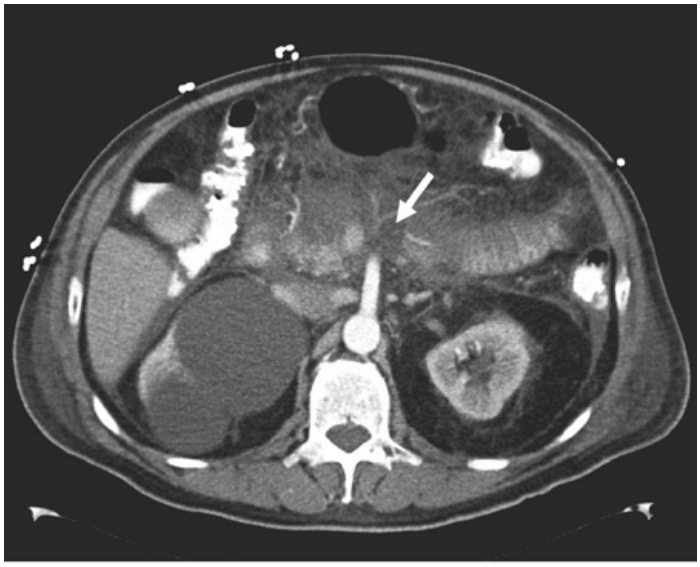

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

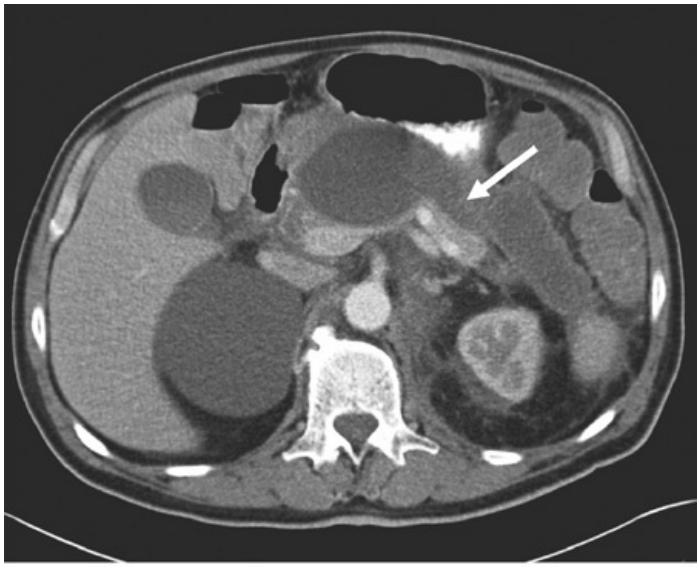

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute pancreatitis: CT evolution.*A.* Contrast-enhanced CT scan of the abdomen performed on admission for a patient with clinical and biochemical parameters suggestive of acute pancreatitis. Note the abnormal enhancement of the pancreatic parenchyma (arrow) suggestive of interstitial pancreatitis. *B.* Contrast-enhanced CT scan of the abdomen performed on the same patient six days later for persistent fever and systemic inflammatory response syndrome. The pancreas now demonstrates significant areas of nonenhancement consistent with development of necrosis, particularly in the body and neck region (arrow). Note that an early CT scan obtained within the first 48 hours of hospitalization may underestimate or miss necrosis. *C.* Contrast-enhanced CT scan of the abdomen performed on the same patient two months after the initial episode of acute pancreatitis. CT now demonstrates evidence of a fluid collection consistent with walled-off pancreatic necrosis (arrow). *(Courtesy of Dr. KJ Mortele, Brigham and Women's Hospital; with permission.)*

Figure 313-2

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

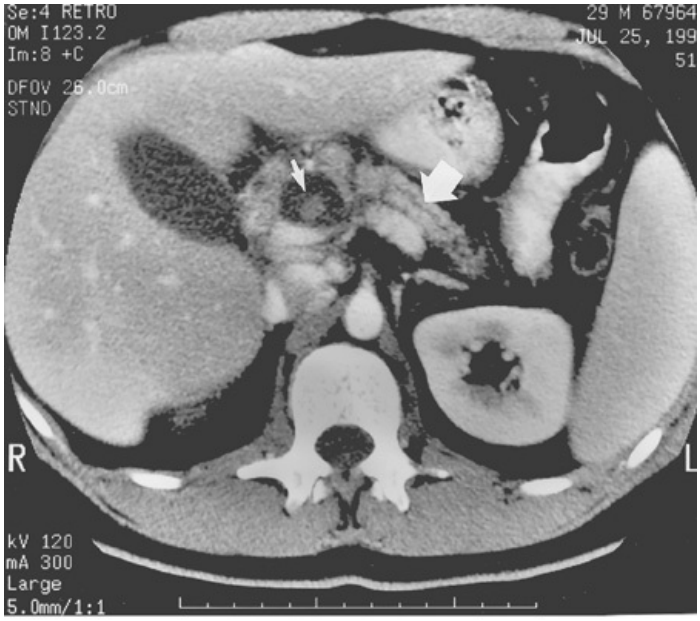

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Acute necrotizing pancreatitis: CT scan. Contrast-enhanced CT scan showing acute pancreatitis with necrosis. Arrow shows partially enhancing body/tail of pancreas surrounded by fluid with decreased enhancement in the neck/body of the pancreas. *B.* Acute fluid collection: CT scan. Contrast-enhanced CT scan showing fluid collection in the retroperitoneum (arrow) compressing the air-filled stomach arising from the pancreas in a patient with asparaginase-induced acute necrotizing pancreatitis. *C.* Walled-off pancreatic necrosis: CT scan. CT scan showing marked walled-off necrosis of the pancreas and peripancreatic area (arrow) in a patient with necrotizing pancreatitis. Addendum: In past years, both of these CT findings (Figs. 313-2B and 313-2C) would have been misinterpreted as pseudocysts. *D.* Spiral CT showing a pseudocyst (small arrow) with a pseudoaneurysm (light area in pseudocyst). Note the demonstration of the main pancreatic duct (big arrow), even though this duct is minimally dilated by ERCP. *(A, B, C, courtesy of Dr. KJ Mortele, Brigham and Women's Hospital; D, courtesy of Dr. PR Ros, Brigham and Women's Hospital; with permission.)*

Figure 313-3

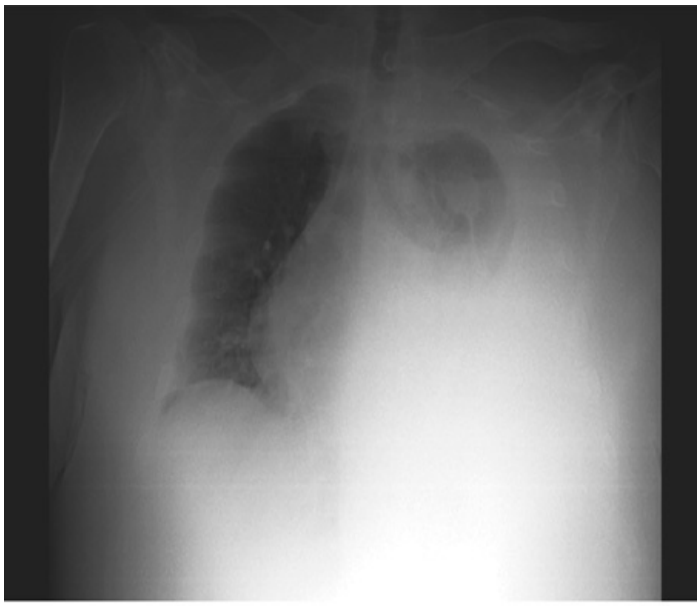

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

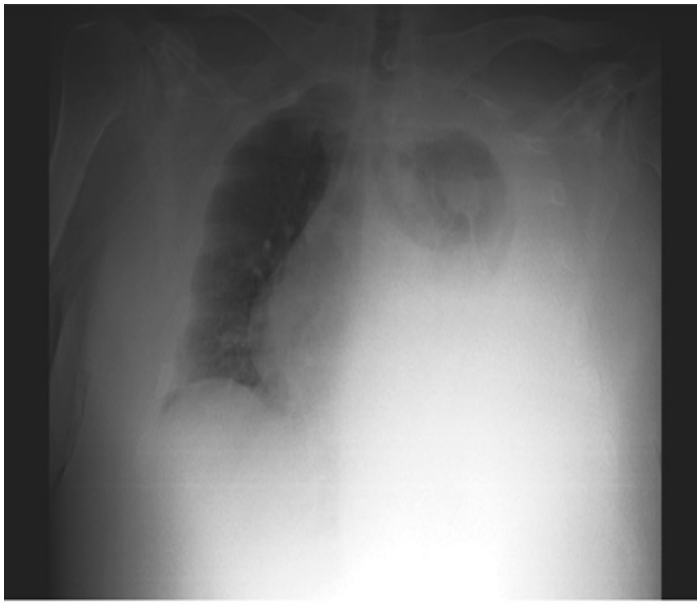

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

*A.***Pancreaticopleural fistula: pancreatic duct leak on ERCP.** Pancreatic duct leak demonstrated (arrow) at the time of retrograde pancreatogram in a patient with acute exacerbation of alcohol-induced acute or chronic pancreatitis. *B.* Pancreaticopleural fistula: CT Scan. Contrast-enhanced CT scan (coronal view) with arrows showing fistula tract from pancreatic duct disruption in the pancreatic pleural fistula. *C.* Pancreaticopleural fistula: Chest x-ray. Large pleural effusion in the left hemithorax from a disrupted pancreatic duct. Analysis of pleural fluid revealed elevated amylase concentration. *(Courtesy of Dr. KJ Mortele, Brigham and Women's Hospital; with permission.)*

Diagnosis

Any severe acute pain in the abdomen or back should suggest the possibility acute pancreatitis. The diagnosis is usually entertained when a patient with a possible predisposition to pancreatitis presents with severe and constant abdominal pain, frequently associated with nausea, emesis, fever, tachycardia, and abnormal findings on abdominal examination. Laboratory studies may reveal leukocytosis, hypocalcemia, and hyperglycemia. The diagnosis of acute pancreatitis requires two of the following: typical abdominal pain, threefold or greater elevation in serum amylase and/or lipase level, and/or confirmatory findings on cross-sectional abdominal imaging. Although not required for diagnosis, markers of severity include hemoconcentration (hematocrit >44%), azotemia (BUN >22 mg/dL), and signs of organ failure (Table 313-2).

The *differential diagnosis* should include the following disorders: (1) perforated viscus, especially peptic ulcer; (2) acute cholecystitis and biliary colic; (3) acute intestinal obstruction; (4) mesenteric vascular occlusion; (5) renal colic; (6) myocardial infarction; (7) dissecting aortic aneurysm; (8) connective tissue disorders with vasculitis; (9) pneumonia; and (10) diabetic ketoacidosis. A penetrating duodenal ulcer can usually be identified by imaging studies or endoscopy. A perforated duodenal ulcer is readily diagnosed by the presence of free intraperitoneal air on abdominal imaging. It may be difficult to differentiate acute cholecystitis from acute pancreatitis, since an elevated serum amylase may be found in both disorders. Pain of biliary tract origin is more right sided or epigastric than periumbilical and can be more severe; ileus is usually absent. Sonography is helpful in establishing the diagnosis of cholelithiasis and cholecystitis. Intestinal obstruction due to mechanical factors can be differentiated from pancreatitis by the history of crescendo-decrescendo pain, findings on abdominal examination, and CT of the abdomen showing changes characteristic of mechanical obstruction. Acute mesenteric vascular occlusion is usually suspected in elderly debilitated patients with brisk leukocytosis, abdominal distention, and bloody diarrhea, confirmed by CT or MR angiography. Systemic lupus erythematosus and polyarteritis nodosa may be confused with pancreatitis, especially since pancreatitis may develop as a complication of these diseases. Diabetic ketoacidosis is often accompanied by abdominal pain and elevated total serum amylase levels, thus closely mimicking acute pancreatitis. However, the serum lipase level is not elevated in diabetic ketoacidosis.

Course of the Disease and Complications

The initial assessment of severity in acute pancreatitis is critical for the appropriate triage and management of patients. The basis for the classification, severity, and complications of acute pancreatitis was initially established at the International Symposium held in Atlanta in 1992. While the definitions have come under greater scrutiny in recent years, it still serves as the common language for clinical care and research in acute pancreatitis. The criteria for severity in acute pancreatitis was defined as organ failure of at least one organ system (defined as a systolic blood pressure <90 mmHg, Pao_2 ≤60 mmHg, creatinine >2.0 mg/dL after rehydration, and gastrointestinal bleeding >500 mL/24 hours) and the presence of a local complication such as necrosis, pseudocyst, and abscess.

Early predictors of severity at 48 hours included ≥3 Ranson's signs and APACHE II score ≥8. Traditional severity indices such as APACHE II and Ranson's criteria have not been clinically useful since they are cumbersome, require collection of a large amount of clinical and laboratory data over time, and do not have acceptable positive and negative predictive value for severe acute pancreatitis. A recent simplified scoring system for the early prediction of mortality was developed from a large cohort of patients with acute pancreatitis. This scoring system, referred to as the Bedside Index of Severity in Acute Pancreatitis (BISAP), incorporates five clinical and laboratory parameters obtained within the first 24 hours of hospitalization: (Table 313-2) (*B*UN >25, *I*mpaired mental status, *S*IRS, *A*ge >60 years, *P*leural effusion on radiography). Presence of three or more of these factors was associated with substantially increased risk for in-hospital mortality among patients with acute pancreatitis.

Apart from the severity indices, there are additional factors that can be used to assess severity in acute pancreatitis. They are best separated into risk factors for severity and markers of severity within 24 hours of admission and during hospitalization. Risk factors for severe acute pancreatitis on admission include older age (>60 years), obesity (BMI ≥30), and comorbid disease. There is also evidence to support initial episode and alcohol use as additional risk factors for severity. At admission and during the first 24 hours, markers of severity in acute pancreatitis include scoring systems such as BISAP score and APACHE II, SIRS, azotemia, hemoconcentration, and organ failure. During hospitalization, markers of severity include persistent organ failure lasting more than 48 hours and pancreatic necrosis.

The course of acute pancreatitis is defined by two phases. In the first phase, which lasts one to two weeks, severity is defined by clinical parameters rather than morphologic findings. The most important clinical parameter is persistent organ failure (i.e., lasting longer than 48 hours), which is the usual cause of death. Severity in the second phase is defined by both clinical parameters and morphologic criteria. The important clinical parameter of severity, as in the first phase, is persistent organ failure. The morphologic criteria of greatest interest is the development of necrotizing pancreatitis, especially when it prolongs hospitalization and/or it requires active intervention such as operative, endoscopic, or percutaneous therapy or requires supportive measures such as renal dialysis, ventilator support, or need for nasoenteric feeding.

The importance of the recognition of interstitial versus necrotizing acute pancreatitis has lead to the development of a CT severity index (Table 313-3) as another measure of severity that is best evaluated three to five days into hospitalization because it may not be possible to distinguish interstitial from necrotizing pancreatitis on contrast-enhanced CT scan on the day of admission. CT identification of local complications, particularly necrosis, is critical because patients with infected and sterile necrosis are at greatest risk of mortality (Figs. 313-1, 313-2). The median prevalence of organ failure is 54% in necrotizing pancreatitis. The prevalence of organ failure is perhaps slightly higher in infected

versus sterile necrosis. With single organ system failure, the mortality is 3–10% but increases to 47% with multisystem organ failure. These data serve to highlight that a patient found to have pancreatic necrosis with multisystem organ failure is the most likely to die.

However, it should be noted that necrotizing pancreatitis is uncommon (10% of all patients with acute pancreatitis), and the far greater proportion of patients presenting in clinical practice have interstitial pancreatitis, which also is associated with organ failure in 10% and death in 3% of cases. This roughly translates to similar absolute mortality figures in the interstitial and necrotizing pancreatitis populations since interstitial disease is far more prevalent.

MILD ACUTE PANCREATITIS

The majority of patients with mild acute pancreatitis and either no organ failure or only transient organ failure will respond to simple supportive care measures that form the hallmark of treatment in acute pancreatitis: bowel rest, intravenous hydration with crystalloid, and analgesia. Oral intake can be resumed once the patient is essentially pain free in the absence of parenteral analgesia, has no nausea or vomiting, normal bowel sounds, and is hungry. Typically, a clear or full liquid diet has been recommended for the initial meal, but a low-fat solid diet is a reasonable choice following recovery from mild acute pancreatitis. Patients with gallstone pancreatitis are at increased risk of recurrence. Therefore, following recovery from mild pancreatitis, consideration should be given to performing a laparoscopic cholecystectomy during the same admission. An alternative for patients who are not surgical candidates would be to perform an endoscopic biliary sphincterotomy.

SEVERE ACUTE PANCREATITIS (SEE FIGS. 313-1, 313-2)

Patients with predictive markers of severity on admission such as obesity or hemoconcentration are also managed with supportive measures outlined as above. It is recommended that vigorous fluid resuscitation take place. Measurement of hematocrit and BUN every 12 hours is recommended to ensure adequacy of fluid resuscitation. A decrease in hematocrit and BUN during the first 12 to 24 hours is strong evidence that sufficient fluids are being administered. If the hematocrit remains elevated or increases further (particularly among those whose hematocrit on admission are >44), fluid resuscitation is inadequate.

Patients with persistent organ failure that does not respond to increased fluids (to counteract hypotension and increased serum creatinine) and/or nasal oxygen to overcome hypoxemia as well as those patients with labored respirations that may herald respiratory failure should be transferred to an intensive care unit for aggressive hydration and close monitoring for the possible need of intubation with mechanical ventilation, hemodialysis, and support of blood pressure.

Treatment: Acute Pancreatitis

In most patients (85–90%) with acute pancreatitis, the disease is self-limited and subsides spontaneously, usually within three to seven days after treatment is instituted. Conventional measures include (1) analgesics for pain, (2) IV fluids and colloids to maintain normal intravascular volume, and (3) no oral alimentation.

Once it is clear that a patient will not be able to tolerate oral feeding (a determination that can usually be made within 48–72 hours), enteral nutrition should be considered [rather than total parenteral nutrition (TPN)] since it maintains gut barrier integrity, thereby preventing bacterial translocation, is less expensive, and has fewer complications than TPN. The route through which enteral feeding is administered is under debate. Nasogastric access is easier to establish and may be as safe as nasojejunal enteral nutrition. However, enteral nutrition that bypasses the stomach and duodenum stimulates pancreatic secretions less and this rationale theoretically supports the use of the nasojejunal route. It has not been demonstrated whether either route is superior in altering morbidity and mortality. When patients with necrotizing pancreatitis begin oral intake of food, consideration should also be given to the addition of pancreatic enzyme supplementation and proton pump inhibitor therapy to assist with fat digestion and reduce gastric acid.

ROLE OF ANTIBIOTICS

There is currently no role for prophylactic antibiotics in either interstitial or necrotizing pancreatitis. Although several early studies suggested a role for prophylactic antibiotics in patients with necrotizing pancreatitis, two recent double-blind, randomized controlled trials failed to demonstrate a reduction in pancreatic infection with use of antibiotic prophylaxis. However, it should also be noted that the overall rate of infected necrosis has been in decline over the past 10–15 years and currently is found in 20% of patients with necrotizing pancreatitis. It is reasonable to start antibiotics in a patient who appears septic while awaiting the results of cultures. If cultures are negative, the antibiotics should be discontinued to minimize the risk of developing fungal superinfection.

Percutaneous aspiration of necrosis with Gram stain and culture should generally not be performed until at least 7–10 days after establishing a diagnosis of necrotizing pancreatitis and only if there are ongoing signs of possible pancreatic infection such as sustained leukocytosis, fever, or organ failure. Once a diagnosis of infected necrosis is established, appropriate antibiotics should be instituted and surgical debridement should be undertaken. There exist minimally invasive alternative therapies such

as endoscopic, percutaneous catheter, and retroperitoneal techniques for necrosectomy. However, there are currently no randomized studies supporting the use of one over another modality. For patients with sterile necrosis, medical management is usually maintained indefinitely unless patients develop serious complications such as compartment syndrome, intestinal perforation, pseudoaneurysms not responding to embolization, or inability to resume oral intake after four to six weeks of treatment (Fig. 313-2).

There are several clearly defined roles for ERCP in acute pancreatitis. Urgent ERCP (within 24 hours) is indicated in patients who have severe acute biliary pancreatitis with organ failure and/or cholangitis. Elective ERCP with sphincterotomy can be considered in patients with persistent or incipient biliary obstruction, those deemed to be poor candidates for cholecystectomy, and for those in whom there is strong suspicion for bile duct stones after cholecystectomy. ERCP with stent placement is also indicated for pancreatic ductal disruptions that occur as part of the inflammatory process and result in peripancreatic fluid collections (Fig. 313-3A).

Several drugs have been evaluated by prospective controlled trials and found ineffective in the treatment of acute pancreatitis. The list, by no means complete, includes glucagon, H_2 blockers, protease inhibitors such as aprotinin, glucocorticoids, calcitonin, nonsteroidal anti-inflammatory drugs (NSAIDs), and lexipafant, a platelet-activating factor inhibitor. A recent meta-analysis of somatostatin, octreotide, and the antiprotease gabexate mesylate in the therapy of acute pancreatitis suggested (1) a reduced mortality rate but no change in complications with octreotide and (2) no effect on the mortality rate but reduced pancreatic damage with gabexate.

A dynamic contrast-enhanced CT (CECT) scan performed three to five days after hospitalization provides valuable information on the severity and prognosis of acute pancreatitis (Fig. 313-1). In particular, a CECT scan allows estimation of the presence and extent of pancreatic necrosis. Recent studies suggest that the likelihood of prolonged pancreatitis or a serious complication is negligible when the CT severity index is 1 or 2 and low with scores of 3–6. However, patients with scores of 7–10 had a 92% morbidity rate and a 17% mortality rate (Table 313-3). A few retrospective studies have raised concern that the use of IV contrast early in the course of acute pancreatitis might intensify pancreatic necrosis. However, since prospective human studies are not available, it is recommended that a CECT scan be obtained only after vigorous initial fluid resuscitation.

Elevation of serum amylase/lipase or persistent inflammatory changes seen on CT scans should not discourage feeding a hungry asymptomatic patient. In this regard, persistence of inflammatory changes on CT scans or persistent elevations in serum amylase/lipase may not resolve for weeks to months. The patient with unremitting severe necrotizing pancreatitis requires vigorous fluid resuscitation and close attention to complications such as cardiovascular collapse, respiratory insufficiency, and pancreatic infection. A useful indicator of severe/complicated forms of acute pancreatitis is the persistence of the systemic SIRS beyond 48 hours. SIRS was defined in 1992 in a joint conference of the American College of Chest Physicians and Society of Critical Care Medicine as a standardized clinical syndrome to indicate the presence of systemic inflammation irrespective of etiology. Several studies have linked persistent SIRS with an increased risk of organ failure and death in acute pancreatitis. Complications from acute pancreatitis should be managed by a combination of radiologic and surgical means (see below). Although sterile necrosis is most often managed conservatively, surgical pancreatic debridement (necrosectomy) should be considered for definitive management of infected necrosis. Such decisions are influenced by response to antibiotic treatment. Multiple operations may be required. A recent study compared the step-up approach, i.e., percutaneous or endoscopic transgastric drainage with open necrosectomy for necrotizing pancreatitis. One third of the patients successfully treated with the step-up approach did not require major abdominal surgery. Enteral-feeding with a nasojejunal tube has been demonstrated to have fewer infectious complications than with total parenteral nutrition (TPN) and is the preferred method of nutritional support. In addition to nutritional support, enteral feeding helps to maintain integrity of the intestinal tract during severe acute pancreatitis.

Patients with severe gallstone-induced pancreatitis, complicated by cholangitis, may improve dramatically if papillotomy is carried out within the first 36–72 hours of the attack. Studies indicate that only those patients with gallstone pancreatitis who are in the very severe group should be considered for urgent ERCP. Finally, the treatment for patients with hypertriglyceridemia-associated pancreatitis includes (1) weight loss to ideal weight, (2) a lipid-restricted diet, (3) exercise, (4) avoidance of alcohol and of drugs that can elevate serum triglycerides (i.e., estrogens, vitamin A, thiazides, and propranolol), and (5) control of diabetes.

RECURRENT PANCREATITIS

Approximately 25% of patients who have had an attack of acute pancreatitis have a recurrence. The two most common etiologic factors are alcohol and cholelithiasis. In patients with recurrent pancreatitis without an obvious cause the differential diagnosis should encompass occult biliary tract disease including microlithiasis, hypertriglyceridemia, drugs, pancreatic cancer, sphincter of Oddi dysfunction, pancreas divisum, cystic fibrosis, and pancreatic cancer (Table 313-1). In one series of 31 patients diagnosed initially as having idiopathic or recurrent acute pancreatitis, 23 were found to have occult

gallstone disease. Thus, approximately two-thirds of patients with recurrent acute pancreatitis without an obvious cause actually have occult gallstone disease due to microlithiasis. Genetic defects as in hereditary pancreatitis can result in recurrent pancreatitis. Other diseases of the biliary tree and pancreatic ducts that can cause acute pancreatitis include choledochocele; ampullary tumors; pancreas divisum; and pancreatic duct stones, stricture, and tumor. Approximately 2–4% of patients with pancreatic carcinoma present with acute pancreatitis.

Infected Pancreatic Necrosis and Pseudocyst

Pancreatic necrosis does not usually become secondarily infected until at least 7–10 days after the onset of acute pancreatitis. Approximately one-half of cases of infected necrosis can be diagnosed between the 7th and 21st day, the remainder after 21 days. The diagnosis of pancreatic infection can be accomplished by CT-guided needle aspiration with Gram stain and culture. The organisms are most frequently gram-negative bacteria of intestinal origin. Clinical clues that should alert the clinician to the possibility of infected necrosis are persistent fever, leukocytosis, and organ failure in a patient with necrotizing pancreatitis. Some reports suggest that patients who have more than 50% pancreatic necrosis are more likely to have infected pancreatic necrosis than those who have lesser amounts of necrosis. Choices of treatment in infected pancreatic necrosis include surgical debridement; endoscopic debridement, if the pancreatic necrosis has been circumscribed into the entity termed walled-off necrosis that affects the posterior wall of the stomach; and, on occasion, radiologic catheter drainage with irrigation in an effort to eliminate at least some infected semisolid material as well as the infected liquid material. Radiologic approach is usually suggested to treat a patient who is too ill to undergo surgical debridement.

WALLED-OFF NECROSIS

In necrotizing pancreatitis, there is invariably an intense inflammatory response involving the fat around the pancreas. This inflammatory process frequently results in peripancreatic necrosis. Eventually, after three to six weeks, there is coalescence of the pancreatic necrosis and peripancreatic fat necrosis into a structure that is encapsulated by fibrous tissue. The name that was originally used to describe this entity was "organized necrosis." New terminology now refers to it as "walled-off necrosis."

The walled-off necrosis contains semisolid necrotic tissue together with a considerable amount of dark fluid representing liquefaction of devitalized pancreatic and peripancreatic tissue as well as some blood.

Walled-off necrosis and a pancreatic pseudocyst may look very similar on first inspection of a contrast-enhanced CT scan. Both show a low attenuation nonenhancing round structure enclosed by a capsule containing fibrous tissue that enhances due to small blood vessels within the capsule. On closer inspection, a distinction can be made. In walled-off necrosis, serial images clearly show that a portion of the pancreas as well as variable amounts of peripancreatic tissue are necrotic. In interstitial pancreatitis, the pancreas enhances normally in response to intravenous contrast, thereby confirming that the process is interstitial pancreatitis. The encapsulated structure is readily seen to be adjacent to the pancreas.

PSEUDOCYSTS

Pseudocysts of the pancreas are extrapancreatic collections of pancreatic fluid containing pancreatic enzymes and a small amount of debris. In contrast to true cysts, pseudocysts do not have an epithelial lining. The walls consist of necrotic tissue, granulation tissue, and fibrous tissue.

A pseudocyst should be distinguished from a postnecrotic fluid collection that contains heterogeneous material including residual necrotic debris. Disruption of the pancreatic ductal system is common. However, the subsequent course of this disruption varies widely, ranging from spontaneous healing to continuous leakage of pancreatic juice, which results in tense ascites. Pseudocysts are preceded by pancreatitis in 90% of cases and by trauma in 10%. Approximately 85% are located in the body or tail of the pancreas and 15% in the head. Some patients have two or more pseudocysts. Abdominal pain, with or without radiation to the back, is the usual presenting complaint. A palpable, tender mass may be found in the middle or left upper abdomen.

On imaging studies, 75% of pseudocysts can be seen to displace some portion of the gastrointestinal tract. Sonography, however, is reliable in detecting pseudocysts. Sonography also permits differentiation between an edematous, inflamed pancreas, which can give rise to a palpable mass, and an actual pseudocyst. Furthermore, serial ultrasound studies will indicate whether a pseudocyst has resolved. CT or MRI complements ultrasonography in the diagnosis of pancreatic pseudocyst, especially when the pseudocyst is infected as suggested by the rare finding of gas within the fluid collection.

In earlier studies with sonography, lesions thought to be pseudocysts were seen to resolve in 25–40% of patients. However, it is now recognized that it is important to distinguish between walled-off necrosis and pseudocysts that typically develop later in the course of acute pancreatitis. Pseudocysts that are >5 cm in diameter may persist for >6 weeks. Recent natural history studies have suggested that noninterventional, expectant management is the best course in selected patients with minimal

symptoms and no evidence of active alcohol use in whom the pseudocyst appears mature by radiography and does not resemble a cystic neoplasm. A significant number of these pseudocysts resolve spontaneously in >6 weeks after their formation. Also, these studies demonstrate that large pseudocyst size is not an absolute indication for interventional therapy and that many peripancreatic fluid collections detected on CT in cases of acute pancreatitis resolve spontaneously. A pseudocyst that does not resolve spontaneously can occasionally lead to serious complications, such as (1) pain caused by expansion of the lesion and pressure on other viscera, (2) rupture, (3) hemorrhage, and (4) abscess. Rupture of a pancreatic pseudocyst is a particularly serious complication. In this case, shock almost always supervenes, and mortality rates range from 14% if the rupture is not associated with hemorrhage to >60% if hemorrhage has occurred. Rupture and hemorrhage are the prime causes of death from pancreatic pseudocyst. A triad of findings—an increase in the size of the mass, a localized bruit over the mass, and a sudden decrease in hemoglobin level and hematocrit without obvious external blood loss—should alert one to the possibility of hemorrhage from a pseudocyst. Thus, in patients who are stable and free of complications and in whom serial ultrasound studies show that the pseudocyst is shrinking, conservative therapy is indicated. Conversely, if the pseudocyst is expanding and is complicated by severe pain, hemorrhage, or abscess, the patient should be operated on. Chronic pseudocysts can be treated safely and drainage can be accomplished by endoscopic, radiologic, or surgical means.

Pseudoaneurysms develop in up to 10% of patients with acute pancreatitis at sites reflecting the distribution of pseudocysts and fluid collections (Fig. 313-2*D*). The splenic artery is most frequently involved, followed by the inferior and superior pancreatic duodenal arteries. This diagnosis should be suspected in patients with pancreatitis who develop upper gastrointestinal bleeding without an obvious cause or in whom thin-cut CT scanning reveals a contrast-enhanced lesion within or adjacent to a suspected pseudocyst. CT angiography can identify the lesion, which can then be treated with angiographic embolization.

The local and systemic complications of acute pancreatitis are summarized in **Table 313-4**. Systemic complications include pulmonary, cardiovascular, hematologic, renal, metabolic, and central nervous system (CNS) abnormalities. *Purtscher's retinopathy*, a relatively unusual complication, is manifested by a sudden and severe loss of vision in a patient with acute pancreatitis. It is characterized by a peculiar funduscopic appearance with cotton-wool spots and hemorrhages confined to an area limited by the optic disc and macula; it is believed to be due to occlusion of the posterior retinal artery with aggregated granulocytes.

Table 313-4 Complications of Acute Pancreatitis

Local
Necrosis
Sterile
Infected
Walled-off necrosis
Pancreatic fluid collections
Pancreatic abscess
Pancreatic pseudocyst
Pain
Rupture
Hemorrhage
Infection
Obstruction of gastrointestinal tract (stomach, duodenum, colon)
Pancreatic ascites
Disruption of main pancreatic duct
Leaking pseudocyst
Involvement of contiguous organs by necrotizing pancreatitis
Massive intraperitoneal hemorrhage
Thrombosis of blood vessels (splenic vein, portal vein)
Bowel infarction
Obstructive jaundice
Systemic
Pulmonary

Pleural effusion
Atelectasis
Mediastinal abscess
Pneumonitis
Acute respiratory distress syndrome
Cardiovascular
Hypotension
Hypovolemia
Sudden death
Nonspecific ST-T changes in electrocardiogram simulating myocardial infarction
Pericardial effusion
Hematologic
Disseminated intravascular coagulation
Gastrointestinal hemorrhage
Peptic ulcer disease
Erosive gastritis
Hemorrhagic pancreatic necrosis with erosion into major blood vessels
Portal vein thrombosis, variceal hemorrhage
Renal
Oliguria
Azotemia
Renal artery and/or renal vein thrombosis
Acute tubular necrosis
Metabolic
Hyperglycemia
Hypertriglyceridemia
Hypocalcemia
Encephalopathy
Sudden blindness (Purtscher's retinopathy)
Central nervous system
Psychosis
Fat emboli
Fat necrosis
Subcutaneous tissues (erythematous nodules)
Bone
Miscellaneous (mediastinum, pleura, nervous system)

PANCREATITIS IN PATIENTS WITH AIDS

The incidence of acute pancreatitis is increased in patients with AIDS for two reasons: (1) the high incidence of infections involving the pancreas such as infections with cytomegalovirus, *Cryptosporidium*, and the *Mycobacterium avium* complex; and (2) the frequent use by patients with AIDS of medications such as didanosine, pentamidine, trimethoprim-sulfamethoxazole, and protease inhibitors (Chap. 189).

Pancreatic Ascites and Pancreatic Pleural Effusions

Pancreatic ascites or pancreatic pleural effusion are initially identified based on CT or MRI imaging and are usually due to disruption of the main pancreatic duct, often by an internal fistula between the duct and the peritoneal cavity or a leaking pseudocyst(Fig. 313-3A). This diagnosis is suggested in a patient with a history of acute pancreatitis in whom the ascites or pleural fluid has both increased levels of albumin [>30 g/L (>3 g/dL)] and a markedly elevated level of amylase. An ERCP or magnetic resonance cholangio pancreatography (MRCP) confirms the clinical suspicion and radiologic findings and often demonstrates passage of contrast material from a disrupted major pancreatic duct or a pseudocyst into the peritoneal cavity. The differential diagnosis of pancreatic ascites should include

intraperitoneal carcinomatosis, tuberculous peritonitis, constrictive pericarditis, and Budd-Chiari syndrome.

TREATMENT: PANCREATIC ASCITES AND PANCREATIC PLEURAL EFFUSIONS

If the pancreatic duct disruption is posterior, an internal fistula may develop between the pancreatic duct and the pleural space, producing a pleural effusion (pancreaticopleural fistula) that is usually left-sided and often massive (Fig. 313-3). If the pancreatic duct disruption is anterior, amylase- and lipase-rich peritoneal fluid accumulate (pancreatic ascites). A leaking, disrupted pancreatic duct is best treated by ERCP and "bridging" stent placement and infrequently requires thoracentesis or chest tube drainage.

Treatment may also require enteral or parenteral alimentation to improve nutrition. If ascites or pleural fluid persists after two to three weeks of medical management, and the disruption is unable to be stented, the patient should be considered for surgical intervention after retrograde pancreatography to define the anatomy of the disrupted duct.

CHRONIC PANCREATITIS AND PANCREATIC EXOCRINE INSUFFICIENCY

Pathophysiology

Chronic pancreatitis is a disease process characterized by irreversible damage to the pancreas as distinct from the reversible changes noted in acute pancreatitis. The condition is best defined by the presence of histologic abnormalities, including chronic inflammation, fibrosis, and progressive destruction of both exocrine and eventually endocrine tissue. A number of etiologies may result in chronic pancreatitis, and may result in the cardinal complications of chronic pancreatitis such as abdominal pain, steatorrhea, weight loss, and diabetes mellitus **(Table 313-5)**.

Table 313-5 Chronic Pancreatitis and Pancreatic Exocrine Insufficiency: Tigar-O Classification System

Toxic-metabolic
Alcoholic
Tobacco smoking
Hypercalcemia
Hyperlipidemia
Chronic renal failure
Medications—phenacetin abuse
Toxins—organotin compounds (e.g., DBTC)
Idiopathic
Early onset
Late onset
Tropical
Genetic
Hereditary pancreatitis
Cationic trypsinogen
$PRSS_1$
$PRSS_2$
CFTR mutations
SPINK1 mutations
Autoimmune
Isolated autoimmune chronic pancreatitis
Autoimmune chronic pancreatitis associated with Sjögren's syndrome
Inflammatory bowel disease
Primary biliary cirrhosis
Recurrent and Severe Acute Pancreatitis
Postnecrotic (severe acute pancreatitis)
Recurrent acute pancreatitis
Vascular diseases/ischemia

Postirradiation
Obstructive
Pancreas divisum
Sphincter of Oddi disorders (controversial)
Duct obstruction (e.g., tumor)
Preampullary duodenal wall cysts
Posttraumatic pancreatic duct scars

Abbreviations: DBTC, dibutylin dichloride; TIGAR-O, toxic-metabolic, idiopathic, genetic, autoimmune, recurrent and severe acute pancreatitis, obstructive.

The events that initiate the inflammatory process in the pancreas are incompletely understood. Current experimental and clinical observations have shown that alcohol has a direct toxic effect on the pancreas. While patients with alcohol-induced pancreatitis generally consume large amounts of alcohol, some consume as little as ≤50 g/d. Prolonged consumption of socially acceptable amounts of alcohol is compatible with the development of chronic pancreatitis. Findings of extensive pancreatic fibrosis in patients who died during their first attack of clinical acute alcohol-induced pancreatitis support the concept that such patients already had chronic pancreatitis.

There is a strong association of smoking and chronic pancreatitis. Cigarette smoke leads to an increased susceptibility to pancreatic self-digestion and predisposes to dysregulation of duct cell CFTR function. It has become increasingly apparent that smoking is an independent, dose-dependent risk factor for chronic pancreatitis and recurrent acute pancreatitis. Smoking is clearly associated with progression of disease in late-onset idiopathic chronic pancreatitis and with increased disease severity in alcohol-induced chronic pancreatitis.

Recent characterization of pancreatic stellate cells (PSC) has added insight to the underlying cellular responses behind development of chronic pancreatitis. Specifically, PSCs are believed to play a role in maintaining normal pancreatic architecture that can shift toward fibrogenesis in the case of chronic pancreatitis. The sentinel acute pancreatitis event (SAPE) hypothesis uniformly describes the events in the pathogenesis of chronic pancreatitis. It is believed that alcohol or additional stimuli lead to matrix metalloproteinase–mediated destruction of normal collagen in pancreatic parenchyma, which later allows for pancreatic remodeling. Proinflammatory cytokines, tumor necrosis factor α (TNF-α), interleukin 1 (IL-1), and interleukin 6 (IL-6) as well as oxidant complexes are able to induce PSC activity with subsequent new collagen synthesis. In addition to being stimulated by cytokines, oxidants, or growth factors, PSCs also possess transforming growth factor β (TGF-β)–mediated self-activating autocrine pathways that may explain disease progression in chronic pancreatitis even after removal of noxious stimuli.

Etiologic Considerations

Among adults in the United States, alcoholism is the most common cause of clinically apparent chronic pancreatitis, while cystic fibrosis is the most frequent cause in children. In up to 25% of adults in the United States with chronic pancreatitis, the cause is not known. That is, they are labeled as *idiopathic chronic pancreatitis*. Recent investigations have indicated that up to 15% of patients with idiopathic pancreatitis may have pancreatitis due to genetic defects (Table 313-5).

Whitcomb and associates studied several large families with hereditary chronic pancreatitis and were able to identify a genetic defect that affects the gene encoding for trypsinogen. Several additional defects of this gene have also been described. The defect prevents the destruction of trypsinogen and allows it to be resistant to the effect of trypsin inhibitor, become spontaneously activated, and to remain activated. It is hypothesized that this continual activation of digestive enzymes within the gland leads to acute injury and, finally, chronic pancreatitis. This group of investigators has also reported that another form of hereditary chronic pancreatitis tends to present later in life, has a female predominance, and frequently leads to chronic pancreatitis.

Several other groups of investigators have documented mutations of *CFTR*. This gene functions as a cyclic AMP–regulated chloride channel. In patients with cystic fibrosis, the high concentration of macromolecules can block the pancreatic ducts. It must be appreciated, however, that there is a great deal of heterogeneity in relationship to the *CFTR* gene defect. More than 1000 putative mutations of the *CFTR* gene have been identified. Attempts to elucidate the relationship between the genotype and pancreatic manifestations have been hampered by the number of mutations. The ability to detect *CFTR* mutations has led to the recognition that the clinical spectrum of the disease is broader than previously thought. Two recent studies have clarified the association between mutations of the *CFTR* gene and another monosymptomatic form of cystic fibrosis (i.e., chronic pancreatitis). It is estimated that in patients with idiopathic pancreatitis, the frequency of a single *CFTR* mutation is 11 times the expected frequency and the frequency of two mutant alleles is 80 times the expected frequency. In these studies, the patients were adults when the diagnosis of pancreatitis was made; none had any

clinical evidence of pulmonary disease, and sweat test results were not diagnostic of cystic fibrosis. The prevalence of such mutations is unclear, and further studies are certainly needed. In addition, the therapeutic and prognostic implication of these findings with respect to managing pancreatitis remains to be determined. Long-term follow-up of affected patients is needed. *CFTR* mutations are common in the general population. It is unclear whether the *CFTR* mutation alone can lead to pancreatitis as an autosomal recessive disease. A recent study evaluated 39 patients with idiopathic chronic pancreatitis to assess the risk associated with these mutations. Patients with two *CFTR* mutations (compound heterozygotes) demonstrated *CFTR* function at a level between that seen in typical cystic fibrosis and cystic fibrosis carriers and had a fortyfold increased risk of pancreatitis. The presence of an *N34S SPINK1* mutation increased the risk twentyfold. A combination of two *CFTR* mutations and an *N34S SPINK1* mutation increased the risk of pancreatitis 900-fold. Table 313-5 lists recognized causes of chronic pancreatitis and pancreatic exocrine insufficiency.

Autoimmune Pancreatitis (Table 313-6)

Table 313-6 Clinical Features of Autoimmune Pancreatitis (Aip)

- Mild symptoms usually abdominal pain, but without frequent attacks of pancreatitis, which are unusual
- Presentation with obstructive jaundice
- Diffuse swelling and enlargement of the pancreas, especially the head, the latter mimicking carcinoma of the pancreas
- Diffuse irregular narrowing of the pancreatic duct in ERCP
- Increased levels of serum gamma globulins especially IgG4
- Presence of other autoantibodies (ANA), rheumatoid factor (RF)
- Can occur with other autoimmune diseases: Sjögren's syndrome, primary sclerosing cholangitis, ulcerative colitis, rheumatoid arthritis
- Extra pancreatic bile duct changes such as stricture of the common bile duct and intrahepatic ducts
- Absence of pancreatic calcifications or cysts
- Pancreatic biopsies reveal extensive fibrosis and lymphoplasmacytic infiltration
- Glucocorticoids are effective in alleviating symptoms, decreasing size of the pancreas, and reversing histopathologic changes
- Two-thirds of patients present with either obstructive jaundice or a "mass" in the head of the pancreas mimicking carcinoma

Autoimmune pancreatitis (AIP) is an uncommon disorder of presumed autoimmune causation with characteristic laboratory, histologic, and morphologic findings. AIP has been described as a primary pancreatic disorder; however, it is also associated with other disorders of presumed autoimmune etiology, including primary sclerosing cholangitis, primary biliary sclerosis, rheumatoid arthritis, Sjögren's syndrome, ulcerative colitis, mediastinal adenopathy, autoimmune thyroiditis, tubulointerstitial nephritis, and retroperitoneal fibrosis. Mild symptoms, usually abdominal pain, are present but attacks of acute pancreatitis are unusual. Furthermore, AIP is not a common cause of idiopathic recurrent pancreatitis. In the United States, 50–75% of patients with AIP present with obstructive jaundice.

Weight loss and new onset of diabetes may also occur. An obstructive pattern on liver tests is common (i.e., disproportionately elevated serum alkaline phosphatase and minimally elevated serum aminotransferases). Elevated serum levels of immunoglobulin G4 (IgG4) provide a marker for the disease, particularly in Western populations. Serum IgG4 normally accounts for only 5–6% of the total IgG4 in healthy patients but is elevated at least twofold higher than 135 mg/dL in those with AIP. CT scans reveal abnormalities in the majority of patients and include diffuse enlargement, focal enlargement, and a distinct enlargement at the head of the pancreas. ERCP or MRCP reveals strictures in the bile duct in more than one-third of patients with AIP; these may be common bile duct strictures, intrahepatic bile duct strictures, or proximal bile duct strictures, with accompanying narrowing of the pancreatic bile duct. This has been termed autoimmune cholangitis. Characteristic histologic findings include extensive lymphoplasmacytic infiltrates with dense fibrosis around pancreatic ducts, as well as a lymphoplasmacytic infiltration, resulting in an obliterative phlebitis.

The Mayo Clinic criteria indicate that AIP can be diagnosed with at least one of three abnormalities: (1) diagnostic histology;(2) characteristic findings on CT and pancreatography combined with elevated IgG4 levels; and (3) response to glucocorticoid therapy, with improvement in pancreatic and extrapancreatic manifestations.

Glucocorticoids have shown efficacy in alleviating symptoms, decreasing the size of the pancreas, and

reversing histopathologic features in patients with AIP. Patients may respond dramatically to glucocorticoid therapy within a two- to four-week period. Prednisone is usually administered at an initial dose of 40 mg/d for four weeks followed by a taper of the daily dosage by 5 mg/week based on monitoring of clinical parameters. Relief of symptoms, serial changes in abdominal imaging of the pancreas and bile ducts, decreased serum γ-globulin and IgG4 levels, and improvements in liver tests are parameters to follow. A poor response to glucocorticoids over a two- to four-week period should raise suspicion of pancreatic cancer or other forms of chronic pancreatitis. In most reports, 50–70% of patients responded to glucocorticoids, but about 25% required a second course of treatment while a smaller number required maintenance treatment with prednisone at a dosage of 5–10 mg/d. Patients with bile duct strictures are less likely to have a sustained response to glucocorticoids and may require immunosuppressive therapy with azathioprine or 6-mercaptopurine.

CLINICAL FEATURES OF CHRONIC PANCREATITIS

Patients with chronic pancreatitis seek medical attention predominantly because of two symptoms: abdominal pain or maldigestion and weight loss. The abdominal pain may be quite variable in location, severity, and frequency. The pain can be constant or intermittent with frequent pain-free intervals. Eating may exacerbate the pain, leading to a fear of eating with consequent weight loss. The spectrum of abdominal pain ranges from mild to quite severe, with narcotic dependence as a frequent consequence. Maldigestion is manifested as chronic diarrhea, steatorrhea, weight loss, and fatigue. Patients with chronic abdominal pain may or may not progress to maldigestion, and ~20% of patients will present with symptoms of maldigestion without a history of abdominal pain. Patients with chronic pancreatitis have significant morbidity and mortality and utilize appreciable amounts of societal resources. Despite the steatorrhea, clinically apparent deficiencies of fat-soluble vitamins are surprisingly uncommon. Physical findings in these patients are usually unimpressive so that there is a disparity between the severity of abdominal pain and the physical signs that usually consist of some mild tenderness.

In contrast to acute pancreatitis, the serum amylase and lipase levels are usually not strikingly elevated in chronic pancreatitis. Elevation of serum bilirubin and alkaline phosphatase may indicate cholestasis secondary to common bile duct stricture caused by chronic inflammation. Many patients have impaired glucose tolerance with elevated fasting blood glucose levels. The diagnostic test with the best sensitivity and specificity is the hormone stimulation test utilizing secretin. It becomes abnormal when ≥60% of the pancreatic exocrine function has been lost. This usually correlates well with the onset of chronic abdominal pain. In earlier studies, approximately 40% of patients with chronic pancreatitis had cobalamin (vitamin B12) malabsorption. This can be corrected by the administration of oral pancreatic enzymes. The fecal elastase-1 and small bowel biopsy are useful in the evaluation of patients with suspected pancreatic steatorrhea. The fecal elastase level will be abnormal and small bowel histology will be normal in such patients. A decrease of fecal elastase level to <100 g per gram of stool strongly suggests severe pancreatic exocrine insufficiency.

Utilizing radiographic techniques **(Fig. 313-4)**, it can be shown that diffuse calcifications noted on plain film of the abdomen usually indicate significant damage to the pancreas. While alcohol is by far the most common cause of pancreatic calcification such calcification may also be noted in hereditary pancreatitis, posttraumatic pancreatitis, hypercalcemic pancreatitis, islet cell tumors, idiopathicchronic pancreatitis, and tropical pancreatitis. Abdominal ultrasonography, CT scanning, and MRCP greatly aid in the diagnosis of pancreatic disease (Fig. 313-4). In addition to excluding a pseudocyst and pancreatic cancer, CT may show calcification, dilated ducts, or an atrophic pancreas. MRCP provides a direct view of the pancreatic duct and is now the diagnostic procedure of choice. The role of endoscopic ultrasonography (EUS) in diagnosing early chronic pancreatitis is still being defined. A total of nine endosonographic features have been described in chronic pancreatitis. The presence of five or more features is considered diagnostic of chronic pancreatitis. EUS complements pancreatic function tests, and a combination of a hormone-stimulation function test and EUS is a modality to evaluate the pancreatic duct morphology, parenchymal architecture, and secretory function for the presence or extent of chronic pancreatitis (Chap. 312). Whether EUS alone can detect early, noncalcific chronic pancreatitis with the same degree of accuracy as the hormone-stimulation test is controversial. Data comparing these modalities head-to-head have indicated that EUS is not a sensitive enough test for detecting early chronic pancreatitis (Chap. 312) and may show positive features in patients who have dyspepsia or even in normal controls. However, recent data suggest that EUS can be combined with endoscopic pancreatic function testing (EUS-ePFT) during a single endoscopy to screen for chronic pancreatitis in patients with chronic abdominal pain.

Figure 313-4

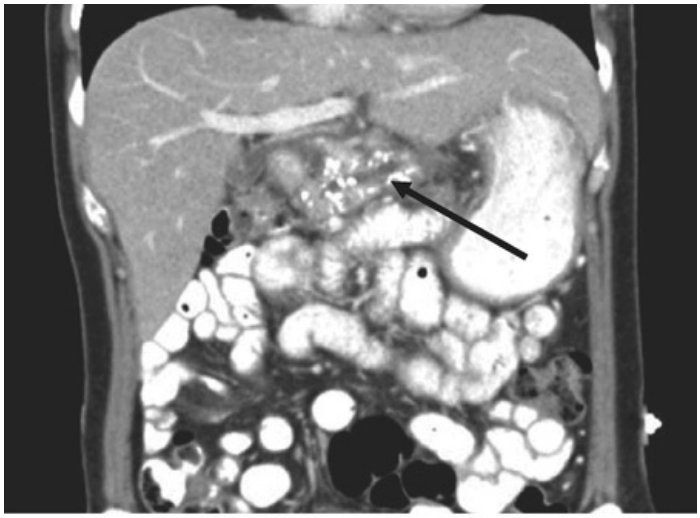

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

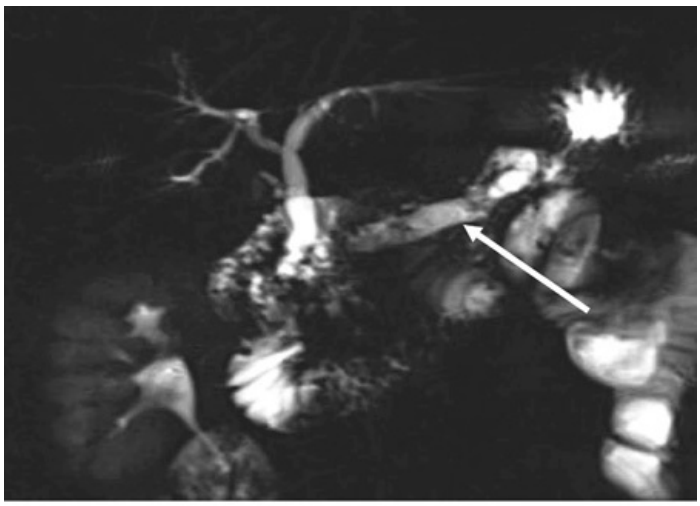

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Chronic pancreatitis and pancreatic calculi: CT scan. In this contrast-enhanced CT scan of the abdomen, there is evidence of an atrophic pancreas with multiple calcifications and stones in the parenchyma and dilated pancreatic duct (arrow). *B.* In this contrast-enhanced CT scan of the abdomen, there is evidence of an atrophic pancreas with multiple calcifications (*arrows*). Note the markedly dilated pancreatic duct seen in this section through the body and tail (*open arrows*). *C.* Chronic pancreatitis on MRCP: dilated duct with filling defects. Gadolinium-enhanced MRI/MRCP

reveals a dilated pancreatic duct (arrow) in chronic pancreatitis with multiple filling defects suggestive of pancreatic duct calculi. *(A, C, courtesy of Dr. KJ Mortele, Brigham and Women's Hospital; with permission.)*

COMPLICATIONS OF CHRONIC PANCREATITIS

The complications of chronic pancreatitis are protean and are listed in **Table 313-7**. Although most patients have impaired glucose tolerance, diabetic ketoacidosis and coma are uncommon. Likewise, end-organ damage (retinopathy, neuropathy, nephropathy) is also uncommon. A nondiabetic retinopathy may be due to either vitamin A and/or zinc deficiency. Gastrointestinal bleeding may occur from peptic ulceration, gastritis, a pseudocyst eroding into the duodenum, or ruptured varices secondary to splenic vein thrombosis due to chronic inflammation of the tail of the pancreas. Jaundice, cholestasis, and biliary cirrhosis may occur from the chronic inflammatory reaction around the intrapancreatic portion of the common bile duct. Twenty years after the diagnosis of calcific chronic pancreatitis, the cumulative risk of pancreatic carcinoma is 4%. Patients with hereditary pancreatitis are at a tenfold higher risk for pancreatic cancer.

Table 313-7 Complications of Chronic Pancreatitis

Narcotic addiction
Impaired glucose tolerance
Gastroparesis
Cobalamin malabsorption
Nondiabetic retinopathy
Effusions with high amylase content Gastrointestinal bleeding
Jaundice
Cholangitis and/or biliary cirrhosis
Subcutaneous fat necrosis
Bone pain
Pancreatic cancer

Treatment: Chronic Pancreatitis

The treatment of steatorrhea with pancreatic enzymes is straightforward even though complete correction of steatorrhea is unusual. Enzyme therapy usually brings diarrhea under control and restores absorption of fat to an acceptable level and effects weight gain. Thus, pancreatic enzymes have been the cornerstone of pancreatic therapy. In treating steatorrhea, it is important to use a potent pancreatic formulation that will deliver sufficient lipase into the duodenum to correct maldigestion and decrease steatorrhea **(Table 313-8)**. In an attempt to standardize the enzyme activity, potency and bioavailability, the Food and Drug Administration (FDA) required that all pancreas enzyme drugs in the United States obtain a New Drug Application (NDA) by April 2008. Table 313-8 lists frequently utilized formulations but availability will be based on compliance with the FDA mandate. Recent data suggests that dosages up to 80,000–100,000 units of lipase per meal may be necessary to normalize nutritional parameters in malnourished chronic pancreatitis patients.

Table 313-8 Frequently Utilized Pancreatic Enzyme Preparations

Enzyme Preparations	Manufacturer, Location	Lipase*	Protease*	Amylase*
Enteric Coated (EC)				
Ultrase [EC microspheres in capsules]	Axcan Pharma, Birmingham, AL			
Ultrase		4,500	25,000	20,000
Ultrase 12		12,000	39,000	39,000
Ultrase 18		18,000	58,500	58,500
Ultrase 20	20,000	65,000	65,000	
Creon [delayed-release capsules containing EC spheres]	Solvay Pharmaceuticals, Marietta, GA			
Creon 6		6,000	19,000	30,000
Creon 12		12,000	38,000	60,000
Creon 24	24,000	76,000	120,000	
Pancrease [EC microtablets in capsule]	Ortho-McNeil			

[illegible]	Pharmaceuticals, Riritan, NJ			
Pancrease MT 4		4,000	12,000	12,000
Pancrease MT 10		10,000	30,000	30,000
Pancrease MT 16		16,000	48,000	48,000
Pancrease MT 20	20,000	44,000	56,000	
Pancreacarb [EC microspheres (buffered) in delayed-release capsule]	Digestive Care, Inc., Bethlehem, PA			
Pancreacarb MS-8		8,000	45,000	40,000
Nonenteric Coated				
Viokase (pancrelipase, USP) Tablets, Powder	Axcan Scandipharm, Birmingham, AL			
Viokase 8		8,000	30,000	30,000
Viokase 16		16,000	60,000	60,000
Viokase Powder: Lactose, sodium chloride, each 0.7 g (1/4 teaspoonful)	16,800	70,000	70,000	
Kuzyme/Ku-trase	UCB Inc., Rochester, NY			
Ku-zyme		1,200	15,000	15,000
Kutrase	1,200	30,000	30,000	

*United States Pharmacopeia (USP) units per tablet or capsule

Note: FDA has mandated all enzyme manufacturers to submit new drug applications (NDAs) for all pancreatic extract drug products after reviewing data that showed substantial variations among currently marketed products. Numerous manufacturers have investigations underway to seek FDA approval for the treatment of exocrine pancreatic insufficiency (EPI) due to cystic fibrosis (CF) or other conditions under the new guidelines for this class of drugs (www.fda.gov).

The management of pain in patients with chronic pancreatitis is problematic.

Recent meta-analyses have shown no consistent benefit of enzyme therapy at reducing pain in chronic pancreatitis. In some patients with idiopathic chronic pancreatitis, conventional nonenteric coated enzyme preparations containing high concentrations of serine proteases may relieve mild abdominal pain or discomfort. The pain relief experienced by these patients actually may be due to improvements in the dyspepsia from maldigestion. Table 313-8 lists the frequently utilized pancreatic enzyme preparations in the United States.

Oxidative stress has also been implicated in the pathophysiology of the pain of chronic pancreatitis. A recent randomized prospective study from India showed antioxidant therapy to be beneficial at reducing pain in mild chronic pancreatitis. Gastroparesis is also quite common in patients with chronic pancreatitis. It is important to recognize this because treatment with enzymes may fail simply because gastroparesis is preventing the appropriate delivery of enzymes into the upper intestine where the enzymes can then act via a feedback inhibition process. In patients with painful chronic pancreatitis, it is important to evaluate gastric emptying and, if gastric emptying is impaired, to effect proper emptying with prokinetic agents. In this setting, enzyme therapy is more apt to be successful.

Endoscopic treatment of chronic pancreatitis pain may involve sphincterotomy, stenting, stone extraction, and drainage of a pancreatic pseudocyst. Therapy directed to the pancreatic duct would seem to be most appropriate in the setting of a dominant stricture, if a ductal stone has led to obstruction. The use of endoscopic stenting for patients with chronic pain, but without a dominant stricture, has not been subjected to any controlled trials. It is now appreciated that significant complications can occur from stenting (i.e., bleeding, cholangitis, stent migration, and stent clogging). All of these may lead to pancreatitis. Importantly, damage to the pancreatic duct and the pancreatic parenchyma can occur following stenting. In patients with large-duct disease usually from alcohol-induced chronic pancreatitis, ductal decompression has been the therapy of choice. Among such patients, 80% seem to obtain immediate relief; however, at the end of three years, one-half the patients have recurrence of pain. Two randomized prospective trials comparing endoscopic to surgical therapy for chronic pancreatitis demonstrated that surgical therapy was superior to endoscopy at decreasing pain and improving quality of life in selected patients with dilated ducts and abdominal pain. This would suggest that chronic pancreatitis patients with dilated ducts and pain should be considered for surgical intervention. The role of preoperative stenting prior to surgery as a predictor of response has yet to be proven.

A Whipple procedure as well as total pancreatectomy and autologous islet cell transplantation have been used in selected patients with chronic pancreatitis and abdominal pain refractory to conventional therapy. The patients who have benefited the most from total pancreatectomy have chronic pancreatitis without prior pancreatic surgery or evidence of islet cell insufficiency. The role of this

procedure remains to be fully defined but may be an option in lieu of ductal decompression surgery or pancreatic resection in patients with intractable, painful small-duct disease, particularly as the standard surgical procedures tend to decrease islet cell yield. Celiac plexus block has not been demonstrated to provide long-lasting pain relief.

Hereditary Pancreatitis

Hereditary pancreatitis is a rare disease that is similar to chronic pancreatitis except for an early age of onset and evidence of hereditary factors (involving an autosomal dominant gene with incomplete penetrance). A genomewide search using genetic linkage analysis identified the hereditary pancreatitis gene on chromosome 7. Mutations in ion codons 29 (exon 2) and 122 (exon 3) of the cationic trypsinogen gene cause autosomal dominant forms of hereditary pancreatitis. The codon 122 mutations lead to a substitution of the corresponding arginine with another amino acid, usually histidine. This substitution, when it occurs, eliminates a fail-safe trypsin self-destruction site necessary to eliminate trypsin that is prematurely activated within the acinar cell. These patients have recurring attacks of severe abdominal pain that may last from a few days to a few weeks. The serum amylase and lipase levels may be elevated during acute attacks but are usually normal. Patients frequently develop pancreatic calcification, diabetes mellitus, and steatorrhea; in addition, they have an increased incidence of pancreatic carcinoma, with the cumulative incidence being as high as 40% by age 70 years. A recent natural history study of hereditary pancreatitis in more than 200 patients from France reported that abdominal pain started in childhood at age 10 years, steatorrhea developed at age 29 years, diabetes at age 38 years, and pancreatic carcinoma at age 55 years. Such patients often require surgical ductal decompression for pain relief. Abdominal complaints in relatives of patients with hereditary pancreatitis should raise the question of pancreatic disease.

PANCREATIC SECRETORY TRYPSIN INHIBITOR (PSTI) GENE MUTATIONS

PSTI, or SPINK1, is a 56-amino-acid peptide that specifically inhibits trypsin by physically blocking its active site. SPINK1 acts as the first line of defense against prematurely activated trypsinogen in the acinar cell. Recently, it has been shown that the frequency of SPINK1 mutations in patients with idiopathicchronic pancreatitis is markedly increased, suggesting that these mutations may be associated with pancreatitis.

Pancreatic Endocrine Tumors

Pancreatic endocrine tumors are discussed in Chap. 350.

OTHER CONDITIONS

Annular Pancreas

When the ventral pancreatic anlage fails to migrate correctly to make contact with the dorsal anlage, the result may be a ring of pancreatic tissue encircling the duodenum. Such an annular pancreas may cause intestinal obstruction in the neonate or the adult. Symptoms of postprandial fullness, epigastric pain, nausea, and vomiting may be present for years before the diagnosis is entertained. The radiographic findings are symmetric dilation of the proximal duodenum with bulging of the recesses on either side of the annular band, effacement but not destruction of the duodenal mucosa, accentuation of the findings in the right anterior oblique position, and lack of change on repeated examinations. The differential diagnosis should include duodenal webs, tumors of the pancreas or duodenum, postbulbar peptic ulcer, regional enteritis, and adhesions. Patients with annular pancreas have an increased incidence of pancreatitis and peptic ulcer. Because of these and other potential complications, the treatment is surgical even if the condition has been present for years. Retrocolic duodenojejunostomy is the procedure of choice, although some surgeons advocate Billroth II gastrectomy, gastroenterostomy, and vagotomy.

Pancreas Divisum

Pancreas divisum occurs when the embryologic ventral and dorsal pancreatic anlagen fail to fuse, so that pancreatic drainage is accomplished mainly through the accessory papilla. Pancreas divisum is the most common congenital anatomic variant of the human pancreas. Current evidence indicates that this anomaly does not predispose to the development of pancreatitis in the great majority of patients who harbor it. However, the combination of pancreas divisum and a small accessory orifice could result in dorsal duct obstruction. The challenge is to identify this subset of patients with dorsal duct pathology. Cannulation of the dorsal duct by ERCP is not as easily done as is cannulation of the ventral duct. Patients with pancreatitis and pancreas divisum demonstrated by MRCP or ERCP should be treated with conservative measures. In many of these patients, pancreatitis is idiopathic and unrelated to the pancreas divisum. Endoscopic or surgical intervention is indicated only if pancreatitis recurs and no other cause can be found. If marked dilation of the dorsal duct can be demonstrated, surgical ductal decompression should be performed. It should be stressed that the ERCP appearance of pancreas divisum (i.e., a small-caliber ventral duct with an arborizing pattern) may be mistaken as representing an obstructed main pancreatic duct secondary to a mass lesion.

Macroamylasemia

In macroamylasemia, amylase circulates in the blood in a polymer form too large to be easily excreted by the kidney. Patients with this condition demonstrate an elevated serum amylase value, a low urinary amylase value, and a C_{am}/C_{cr} ratio of <1%. The presence of macroamylase can be documented by chromatography of the serum. The prevalence of macroamylasemia is 1.5% of the nonalcoholic general adult hospital population. Usually macroamylasemia is an incidental finding and is not related to disease of the pancreas or other organs.

Macrolipasemia has now been documented in a few patients with cirrhosis or non-Hodgkin's lymphoma. In these patients, the pancreas appeared normal on ultrasound and CT examination. Lipase was shown to be complexed with immunoglobulin A. Thus, the possibility of *both* macroamylasemia and macrolipasemia should be considered in patients with elevated blood levels of these enzymes.

ACKNOWLEDGMENTS

This chapter represents a revised version of the chapter by Dr. Norton J. Greenberger and Dr. Phillip P. Toskes that was in the previous editions of Harrison's.

FURTHER READINGS

Banks PA, Freeman M: Practice guidelines in acute pancreatitis. Am J Gastroenterol 101:2379, 2006[PMID: 17032204] [Full Text]

Conwell DL et al: An endoscopic pancreatic function test with synthetic porcine secretin for the evaluation of chronic abdominal pain and suspected chronic pancreatitis. Gastrointest Endosc 57:37, 2003[PMID: 12518128] [Full Text]

Forsmark CE: The early diagnosis of chronic pancreatitis. Clin Gastroenterol Hepatol 6:1291, 2008[PMID: 18986847] [Full Text]

Pandol SJ et al: Acute pancreatitis: Bench to the bedside. Gastroenterology 132:1127, 2007[PMID: 17383433] [Full Text]

Rebours V et al: The natural history of hereditary pancreatitis: A national series. Gut 58:97, 2009[PMID: 18755888] [Full Text]

Thomson A: Enteral versus parenteral nutritional support in acute pancreatitis: A clinical review. J Gastroenterol Hepatol 21:22, 2006[PMID: 16706807] [Full Text]

van Santvoort HC et al: A step-up approach or open necrosectomy for necrotizing pancreatitis. N Engl J Med 362:1491, 2010

Waljee AK et al: Systematic review: Pancreatic enzyme treatment of malabsorption associated with chronic pancreatitis. Aliment Pharmacol Ther 29:235, 2008[PMID: 19035969] [Full Text]

Whitcomb DC et al: Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nat Genet 14:141, 1996[PMID: 8841182] [Full Text]

Witt H et al: Chronic pancreatitis: Challenges and advances in pathogenesis, genetics, diagnosis, and therapy. Gastroenterology 132:1557, 2007[PMID: 17466744] [Full Text]

Wu BU, Conwell DL: Acute pancreatitis part I: Approach to early management. Clin Gastroenterol Hepatol 8:410, 2010[PMID: 19896558] [Full Text]

——— et al: The early prediction of mortality in acute pancreatitis: A large population-based study. Gut 57:1698, 2008

——— et al: Early changes in blood urea nitrogen mortality in acute pancreatitis. Gastroenterology 137:129, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION TO THE IMMUNE SYSTEM

Definitions

- *Adaptive immune system*&mdah;recently evolved system of immune responses mediated by T and B lymphocytes. Immune responses by these cells are based on specific antigen recognition by clonotypic receptors that are products of genes that rearrange during development and throughout the life of the organism. Additional cells of the adaptive immune system include various types of antigen-presenting cells.
- *Antibody*—B cell–produced molecules encoded by genes that rearrange during B cell development consisting of immunoglobulin heavy and light chains that together form the central component of the B cell receptor for antigen. Antibody can exist as B cell–surface antigen-recognition molecules or as secreted molecules in plasma and other body fluids (Table 314-13).
- *Antigens*—foreign or self-molecules that are recognized by the adaptive and innate immune systems resulting in immune cell triggering, T cell activation, and/or B cell antibody production.
- *Antimicrobial peptides*—small peptides <100 amino acids in length that are produced by cells of the innate immune system and have anti-infectious agent activity (Table 314-2).
- *Apoptosis*—the process of *programmed cell death* whereby signaling through various "death receptors" on the surface of cells [e.g., tumor necrosis factor (TNF) receptors, CD95] leads to a signaling cascade that involves activation of the caspase family of molecules and leads to DNA cleavage and cell death. Apoptosis, which does not lead to induction of inordinate inflammation, is to be contrasted with *cell necrosis*, which does lead to induction of inflammatory responses.
- *Autoimmune diseases*—diseases such as systemic lupus erythematosus and rheumatoid arthritis in which cells of the adaptive immune system such as autoreactive T and B cells become overreactive and produce self-reactive T cell and antibody responses.
- *Autoinflammatory diseases*—hereditary disorders such as hereditary periodic fevers (HPFs) characterized by recurrent episodes of severe inflammation and fever due to mutations in controls of the innate inflammatory response, i.e., the inflammasome (see below and Table 314-6). Patients with HPFs also have rashes and serosal and joint inflammation and some can have neurologic symptoms. Autoinflammatory diseases are different from autoimmune diseases in that evidence for activation of adaptive immune cells such as autoreactive B cells is not present.
- *B cell receptor for antigen* —complex of surface molecules that rearrange during postnatal B cell development, made up of surface immunoglobulin (Ig) and associated Ig αβ chain molecules that recognize nominal antigen via Ig heavy- and light-chain variable regions, and signal the B cell to terminally differentiate to make antigen-specific antibody (Fig. 314-8).
- *B lymphocytes*—bone marrow–derived or bursal-equivalent lymphocytes that express surface immunoglobulin (the B cell receptor for antigen) and secrete specific antibody after interaction with antigen (Figs. 314-2 and 314-6).
- *CD classification of human lymphocyte differentiation antigens*—the development of monoclonal antibody technology led to the discovery of a large number of new leukocyte surface molecules. In 1982, the First International Workshop on Leukocyte Differentiation Antigens was held to establish a nomenclature for cell-surface molecules of human leukocytes. From this and subsequent leukocyte differentiation workshops has come the *cluster of differentiation* (CD) classification of leukocyte antigens **(Table 314-1)**.
- *Chemokines*—soluble molecules that direct and determine immune cell movement and circulation pathways.
- *Complement*—cascading series of plasma enzymes and effector proteins whose function is to lyse pathogens and/or target them to be phagocytized by neutrophils and monocyte/macrophage lineage cells of the reticuloendothelial system (Fig. 314-5).
- *Co-stimulatory molecules*—molecules of antigen-presenting cells (such as B7-1 and B7-2 or CD40) that lead to T cell activation when bound by ligands on activated T cells (such as CD28 or CD40 ligand) (Fig. 314-7).
- *Cytokines*—soluble proteins that interact with specific cellular receptors that are involved in the regulation of the growth and activation of immune cells and mediate normal and pathologic inflammatory and immune responses (Tables 314-7, 314-9, and 314-10).
- *Dendritic cells*—myeloid and/or lymphoid lineage antigen-presenting cells of the adaptive immune system. Immature dendritic cells, or dendritic cell precursors, are key components of the innate immune system by responding to infections with production of high levels of cytokines. Dendritic cells are key initiators both of innate immune responses via cytokine production and of adaptive immune responses via presentation of antigen to T lymphocytes (Figs. 314-2 and 314-3, Table 314-5).
- *Inflammasome*—large cytoplasmic complexes of intracellular proteins that link the sensing of microbial products and cellular stress to the proteolytic activation of interleukin (IL)-1β and IL-18 inflammatory cytokines. Activation of molecules in the inflammasome is a key step in the response of the innate immune system for intracellular recognition of microbial and other danger signals in both health and pathologic states (Table 314-6).
- *Innate immune system*—ancient immune recognition system of host cells bearing germ line–encoded pattern recognition receptors (PRRs) that recognize pathogens and trigger a variety of mechanisms of pathogen elimination. Cells of the innate immune system include natural killer cell lymphocytes,

monocytes/macrophages, dendritic cells, neutrophils, basophils, eosinophils, tissue mast cells, and epithelial cells **(Tables 314-2 to 314-5 and 314-12)**.

- *Large granular lymphocytes*—lymphocytes of the innate immune system with azurophilic cytotoxic granules that have natural killer cell activity capable of killing foreign and host cells with few or no self–major histocompatibility complex (MHC) class I molecules (Fig. 314-4).
- *Natural killer cells*—large granular lymphocytes that kill target cells expressing few or no human leukocyte antigen (HLA) class I molecules, such as malignantly transformed cells and virally infected cells. Natural killer cells express receptors that inhibit killer cell function when self–major histocompatibility complex class I is present (Fig. 314-4).
- *Natural killer (NK) T cells*—innate-like lymphocytes that use an invariant T cell receptor (TCR)-α chain combined with a limited set of TCR-β chains and coexpress receptors commonly found on NK cells. NK T cells recognize lipid antigens of bacterial, viral, fungal, and protozoal infectious agents.
- *Pathogen-associated molecular patterns* (PAMPs)—Invariant molecular structures expressed by large groups of microorganisms that are recognized by host cellular pattern recognition receptors in the mediation of innate immunity **(Fig. 314-1)**.
- *Pattern recognition receptors* (PRRs)—germ line–encoded receptors expressed by cells of the innate immune system that recognize pathogen-associated molecular patterns (Table 314-3).
- *Polyreactive natural antibodies*—preexisting low-affinity antibodies produced by innate B cells that cross-react with multiple antigens and are available at the time of infection to bind to and coat the invading pathogen and harness innate responses to slow the infection until an adaptive high-affinity protective antibody response can be made.
- *T cell receptor (TCR) for antigen*—complex of surface molecules that rearrange during postnatal T cell development made up of clonotypic TCR-α and -β chains that are associated with the CD3 complex composed of invariant γ, δ, ε, ζ, and η chains. TCR-α and -β chains recognize peptide fragments of protein antigen physically bound in antigen-presenting cell major histocompatibility complex class I or II molecules, leading to signaling via the CD3 complex to mediate effector functions (Fig. 314-7).
- *T cells*—thymus-derived lymphocytes that mediate adaptive cellular immune responses including T helper, T regulatory, and cytotoxic T lymphocyte effector cell functions (**Figs. 314-2,314-3**, and 314-7).
- *Tolerance*—B and T cell nonresponsiveness to antigens that results from encounter with foreign or self-antigens by B and T lymphocytes in the absence of expression of antigen-presenting cell co-stimulatory molecules. Tolerance to antigens may be induced and maintained by multiple mechanisms either centrally (in the thymus for T cells or bone marrow for B cells) or peripherally at sites throughout the peripheral immune system.

Introduction

The human immune system has evolved over millions of years from both invertebrate and vertebrate organisms to develop sophisticated defense mechanisms to protect the host from microbes and their virulence factors. The normal immune system has three key properties: a highly diverse repertoire of antigen receptors that enables recognition of a nearly infinite range of pathogens; immune memory, to mount rapid recall immune responses; and immunologic tolerance, to avoid immune damage to normal self-tissues. From invertebrates, humans have inherited the *innate immune system*, an ancient defense system that uses germ line–encoded proteins to recognize pathogens. Cells of the innate immune system, such as macrophages, dendritic cells, and natural killer (NK) lymphocytes, recognize pathogen-associated molecular patterns (PAMPs) that are highly conserved among many microbes and use a diverse set of pattern recognition receptor molecules (PRRs). Important components of the recognition of microbes by the innate immune system include (1) recognition by germ line–encoded host molecules, (2) recognition of key microbe virulence factors but not recognition of self-molecules, and (3) nonrecognition of benign foreign molecules or microbes. Upon contact with pathogens, macrophages and NK cells may kill pathogens directly or, in concert with dendritic cells, may activate a series of events that both slow the infection and recruit the more recently evolved arm of the human immune system, the *adaptive immune system*.

Adaptive immunity is found only in vertebrates and is based on the generation of antigen receptors on T and B lymphocytes by gene rearrangements, such that individual T or B cells express unique antigen receptors on their surface capable of specifically recognizing diverse antigens of the myriad infectious agents in the environment. Coupled with finely tuned specific recognition mechanisms that maintain tolerance (nonreactivity) to self-antigens, T and B lymphocytes bring both *specificity* and *immune memory* to vertebrate host defenses.

This chapter describes the cellular components, key molecules (Table 314-1), and mechanisms that make up the innate and adaptive immune systems and describes how adaptive immunity is recruited to the defense of the host by innate immune responses. An appreciation of the cellular and molecular bases of innate and adaptive immune responses is critical to understanding the pathogenesis of inflammatory, autoimmune, infectious, and immunodeficiency diseases.

Table 314–1. Human Leukocyte Surface Antigens—the CD Classification of Leukocyte Differentiation Antigens

Surface Antigen (Other Names)	Family	Molecular Mass, kDa	Distribution	Ligand(s)	Function
CD1a (T6, HTA-1)	Ig	49	CD, cortical thymocytes, Langerhans type of dendritic cells	TCRγδ T cells	CD1 molecules present lipid antigens of intracellular bacteria such as *Mycobacterium leprae* and *M. tuberculosis* to TCRγδ T cells.
CD1b	Ig	45	CD, cortical thymocytes,	TCRγδ T cells	

			Langerhans type of dendritic cells		
CD1c	Ig	43	DC, cortical thymocytes, subset of B cells, Langerhans type of dendritic cells	TCR$\gamma\delta$ T cells	
CD1d	Ig	?	Cortical thymocytes, intestinal epithelium, Langerhans type of dendritic cells	TCR$\gamma\delta$ T cells	
CD2 (T12, LFA-2)	Ig	50	T, NK	CD58, CD48, CD59, CD15	Alternative T cell activation, T cell anergy, T cell cytokine production, T- or NK-mediated cytolysis, T cell apoptosis, cell adhesion
CD3 (T3, Leu-4)	Ig	γ:25–28, δ:21–28, ε:20–25, η:21–22, ζ:16	T	Associates with the TCR	T cell activation and function; ζ is the signal transduction component of the CD3 complex
CD4 (T4, Leu-3)	Ig	55	T, myeloid	MHC-II, HIV, gp120, IL-16, SABP	T cell selection, T cell activation, signal transduction with p56/*ck*, primary receptor for HIV
CD7 (3A1, Leu-9)	Ig	40	T, NK	K-12 (CD7L)	T and NK cell signal transduction and regulation of IFN-γ, TNF-α production
CD8 (T8, Leu-2)	Ig	34	T	MHC-I	T cell selection, T cell activation, signal transduction with p56/*ck*
CD14 (LPS-receptor)	LRG	53–55	M, G (weak), not by myeloid progenitors	Endotoxin (lipopolysaccharide), lipoteichoic acid, PI	TLR4 mediates with LPS and other PAMP activation of innate immunity
CD19 B4	Ig	95	B (except plasma cells), FDC	Not known	Associates with CD21 and CD81 to form a complex involved in signal transduction in B cell development, activation, and differentiation
CD20 (B1)	Unassigned	33–37	B (except plasma cells)	Not known	Cell signaling, may be important for B cell activation and proliferation
CD21 (B2, CR2, EBV-R, C3dR)	RCA	145	Mature B, FDC, subset of thymocytes	C3d, C3dg, iC3b, CD23, EBV	Associates with CD19 and CD81 to form a complex involved in signal transduction in B cell development, activation, and differentiation; Epstein-Barr virus receptor
CD22 (BL-CAM)	Ig	130–140	Mature B	CDw75	Cell adhesion, signaling through association with p72*sky*, p53/56/*yn*, PI3 kinase, SHP1, fLCγ
CD23 (FcεRII, B6, Leu-20, BLAST-2)	C-type lectin	45	B, M, FDC	IgE, CD21, CD11b, CD11c	Regulates IgE synthesis, cytokine release by monocytes
CD28	Ig	44	T, plasma cells	CD80, CD86	Co-stimulatory for T cell activation; involved in the decision between T cell activation and anergy
CD40	TNFR	48–50	B, DC, EC, thymic epithelium, MP, cancers	CD154	B cell activation, proliferation, and differentiation; formation of GCs; isotype switching; rescue from apoptosis
CD45 (LCA, T200, B220)	PTP	180, 200, 210, 220	All leukocytes	Galectin-1, CD2, CD3, CD4	T and B activation, thymocyte development, signal transduction, apoptosis
CD45RA	PTP	210, 220	Subset T, medullary thymocytes, “naive” T	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing exon 4 (A), restricted to a subset of T cells
CD45RB	PTP	200, 210, 220	All leukocytes	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing exon 5 (B)
CD45RC	PTP	210, 220	Subset T, medullary thymocytes, “naive” T	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing exon 6 (C), restricted to a subset of T cells
CD45RO	PTP	180	Subset T, cortical thymocytes, “memory” T	Galectin-1, CD2, CD3, CD4	Isoforms of CD45 containing no differentially spliced exons, restricted to a subset of T cells
CD80 (B7-1, BB1)	Ig	60	Activated B and T, MP, DC	CD28, CD152	Co-regulator of T cell activation; signaling through CD28 stimulates and through CD152 inhibits T cell activation

CD86 (B7-2, B70)	Ig	80	Subset B, DC, EC, activated T, thymic epithelium	CD28, CD152	Co-regulator of T cell activation; signaling through CD28 stimulates and through CD152 inhibits T cell activation
CD95 (APO-1, Fas)	TNFR	135	Activated T and B	Fas ligand	Mediates apoptosis
CD152 (CTLA-4)	Ig	30–33	Activated T	CD80, CD86	Inhibits T cell proliferation
CD154 (CD40L)	TNF	33	Activated CD4+ T, subset CD8+ T, NK, M, basophil	CD40	Co-stimulatory for T cell activation, B cell proliferation and differentiation

Abbreviations: CTLA, cytotoxic T lymphocyte–associated protein; DC, dendritic cells; EBV, Epstein-Barr virus; EC, endothelial cells; ECM, extracellular matrix; FcγRIIIA, low-affinity IgG receptor isoform A; FDC, follicular dendritic cells; G, granulocytes; GC, germinal center; GPI, glycosyl phosphatidylinositol; HTA, human thymocyte antigen; IgG, immunoglobulin G; LCA, leukocyte common antigen; LPS, lipopolysaccharide; MHC-I, major histocompatibility complex class I; MP, macrophages; Mr, relative molecular mass; NK, natural killer cells; P, platelets; PBT, peripheral blood T cells; PI, phosphatidylinositol; PI3K, phosphatidylinositol 3-kinase; PLC, phospholipase C; PTP, protein tyrosine phosphatase; TCR, T cell receptor; TNF, tumor necrosis factor; TNFR, tumor necrosis factor receptor. For an expanded list of cluster of differentiation (CD) human antigens, see Harrison's Online at http://www.accessmedicine.com; and for a full list of CD human antigens from the most recent Human Workshop on Leukocyte Differentiation Antigens (VII), see **http://mpr.nci.nih.gov/prow/**.

Source: Compiled from T Kishimoto et al (eds): Leukocyte Typing VI, New York, Garland Publishing 1997; R Brines et al: Immunology Today 18S:1, 1997; and S Shaw (ed): Protein Reviews on the Web, **http://mpr.nci.nih.gov/prow/**.

The Innate Immune System

All multicellular organisms, including humans, have developed the use of a limited number of surface and intracellular germ line–encoded molecules that recognize large groups of pathogens. Because of the myriad human pathogens, host molecules of the human innate immune system sense "danger signals" and either recognize PAMPs, the common molecular structures shared by many pathogens, or recognize host cell molecules produced in response to infection such as heat shock proteins and fragments of the extracellular matrix. PAMPs must be conserved structures vital to pathogen virulence and survival, such as bacterial endotoxin, so that pathogens cannot mutate molecules of PAMPs to evade human innate immune responses. PRRs are host proteins of the innate immune system that recognize PAMPs as host danger signal molecules (Tables 314-2 and 314-3). Thus, recognition of pathogen molecules by hematopoietic and nonhematopoietic cell types leads to activation/production of the complement cascade, cytokines, and antimicrobial peptides as effector molecules. In addition, pathogen PAMPs as host danger signal molecules activate dendritic cells to mature and to express molecules on the dendritic cell surface that optimize antigen presentation to respond to foreign antigens.

Table 314–2. Major Components of the Innate Immune System

Pattern recognition receptors (PRR)	C-type lectins, leucine-rich proteins, scavenger receptors, pentraxins, lipid transferases, integrins, inflammasome proteins
Antimicrobial peptides	α-Defensins, β-defensins, cathelin, protegrin, granulysin, histatin, secretory leukoprotease inhibitor, and probiotics
Cells	Macrophages, dendritic cells, NK cells, NK-T cells, neutrophils, eosinophils, mast cells, basophils, and epithelial cells
Complement components	Classic and alternative complement pathway, and proteins that bind complement components
Cytokines	Autocrine, paracrine, endocrine cytokines that mediate host defense and inflammation, as well as recruit, direct, and regulate adaptive immune responses

Abbreviation: NK cells, natural killer cells.

Table 314–3. Major Pattern Recognition Receptors (PRR) of the Innate Immune System

PRR Protein Family	Sites of Expression	Examples	Ligands (PAMPs)	Functions of PRR
Toll-like receptors	Multiple cell types	TLR2-10	Bacterial and viral carbohydrates	Activate innate immune cells to respond to multiple pathogens and initiate adaptive immune responses
C-type Lectins	Plasma proteins	Collectins		Opsonization of bacteria and virus, activation of complement
Humoral			Terminal mannose	
Cellular	Macrophages, dendritic cell	Macrophage mannose receptor	Carbohydrate on HLA molecules	Phagocytosis of pathogens
	Natural killer (NK) cells	NKG2-A		Inhibits killing of host cells expressing HLA + self-peptides
Leucine-rich proteins	Macrophages, dendritic cells, epithelial cells	CD14	Lipopolysaccharide (LPS)	Binds LPS and Toll proteins

Scavenger receptors	Macrophage	Macrophage scavenger receptors	Bacterial cell walls	Phagocytosis of bacteria
Pentraxins	Plasma protein	C-reactive proteins	Phosphatidyl choline	Opsonization of bacteria, activation of complement
	Plasma protein	Serum amyloid P	Bacterial cell walls	Opsonization of bacteria, activation of complement
Lipid transferases	Plasma protein	LPS binding protein	LPS	Binds LPS, transfers LPS to CD14
Integrins	Macrophages, dendritic cells, NK cells	CD11b,c; CD18	LPS	Signals cells, activates phagocytosis
NOD-like Receptors	Innate cells	NALP-3	Viral DNA Bacterial muramyl dipeptide	Cytosolic proteins involved in innate sensing

Abbreviation: PAMPs, pathogen-associated molecular patterns.

Source: Adapted from R Medzhitov, CA Janeway: Curr Opin Immunol 9:4, 1997. Copyright 1997, with permission from Elsevier.

Pattern Recognition

Major PRR families of proteins include C-type lectins, leucine-rich proteins, macrophage scavenger receptor proteins, plasma pentraxins, lipid transferases, and integrins (Table 314-3). A major group of PRR collagenous glycoproteins with C-type lectin domains are termed *collectins* and include the serum protein mannose-binding lectin (MBL). MBL and other collectins, as well as two other protein families—the pentraxins (such as C-reactive protein and serum amyloid P) and macrophage scavenger receptors—all have the property of opsonizing (coating) bacteria for phagocytosis by macrophages and can also activate the complement cascade to lyse bacteria. Integrins are cell-surface adhesion molecules that signal after cells bind bacterial lipopolysaccharide (LPS) and activate phagocytic cells to ingest pathogens.

There are multiple connections between the innate and adaptive immune systems; these include (1) a plasma protein, LPS-binding protein, which binds and transfers LPS to the macrophage LPS receptor, CD14; (2) a human family of proteins called *Toll-like receptor proteins* (TLRs), some of which are associated with CD14, bind LPS, and signal epithelial cells, dendritic cells, and macrophages to produce cytokines and upregulate cell-surface molecules that signal the initiation of adaptive immune responses (Fig. 314-1, Tables 314-3 and 314-4), and (3) families of intracellular microbial sensors called NOD-like receptors (NLRs) and RIG-like helicases (RLHs). Proteins in the Toll family can be expressed on macrophages, dendritic cells, and B cells as well as on a variety of nonhematopoietic cell types, including respiratory epithelial cells. Ten TLRs have been identified in humans and 13 TLRs in mice (Tables 314-4 and 314-5). Upon ligation, TLRs activate a series of intracellular events that lead to the killing of bacteria- and viral-infected cells as well as to the recruitment and ultimate activation of antigen-specific T and B lymphocytes (Fig. 314-1). Importantly, signaling by massive amounts of LPS through TLR4 leads to the release of large amounts of cytokines that mediate LPS-induced shock. Mutations in TLR4 proteins in mice protect from LPS shock, and TLR mutations in humans protect from LPS-induced inflammatory diseases such as LPS-induced asthma (Fig. 314-1).

Figure 314-1

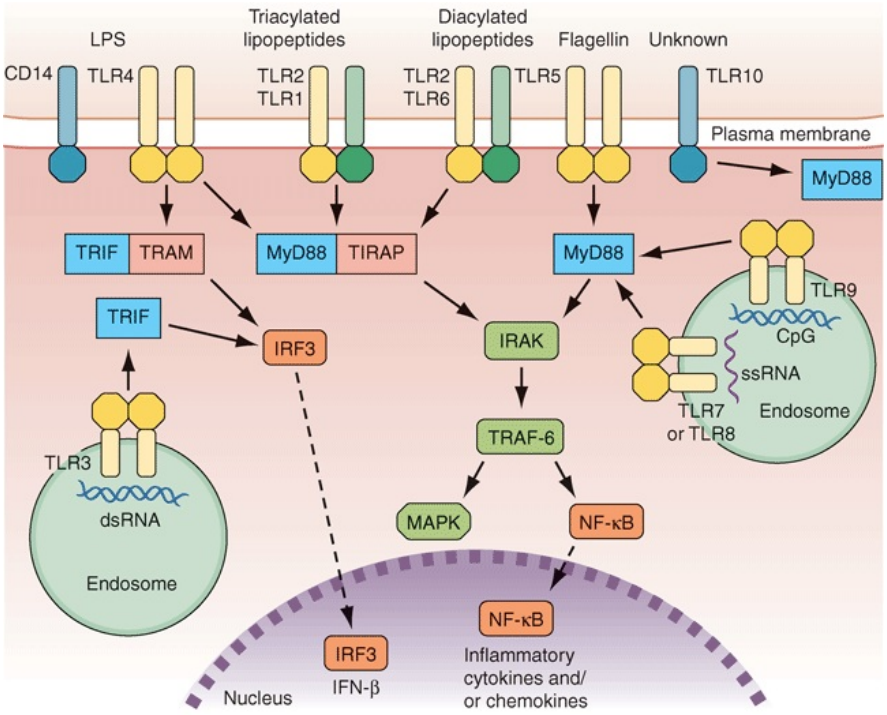

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Overview of major TLR signaling pathways. All TLRs signal through MyD88, with the exception of TLR3. TLR4 and the TLR2 subfamily (TLR1, TLR2, TLR6) also engage TIRAP. TLR3 signals through TRIF. TRIF is also used in conjunction with TRAM in the TLR4–MyD88-independent pathway. Dashed arrrows indicate translocation into the nucleus. LPS, lipopolysaccharide; dsRNA, double-strand RNA; ssRNA, single-strand RNA; MAPK, mitogen-activated protein kinases; NF-κB, nuclear factor-κB; IFN, interferon; IRF3, interferon regulatory factor 3; TLR, Toll-like receptor. (*Adapted from D van Duin et al., with permission.*)

Table 314–4. The Role of Pattern Recognition Receptors (PRRs) in Modulation of Adaptive Immune Responses

PRR Family	PRRs	Ligand	DC or Macrophage Cytokine Response	Adaptive Immune Response
TLRs	TLR2 (heterodimer with TLR1 or 6)	Lipopeptides Pam-3-cys (TLR 2/1) MALP (TLR 2/6)	Low IL-12p70 High IL-10 IL-6	T_H1 T_H2 T regulatory
	TLR3	dsRNA	IL-12p70 IFN-α IL-6	T_H1
	TLR4	*E. coli* LPS	High IL-12p70 Intermediate IL-10 IL-6	T_H1
	TLR5	Flagellin	High IL-12p70 Low IL-12p70	T_H1 T_H2
	TLR7/8	ssRNA Imidazoquinolines	High IL-12p70 IFN-α IL-6	T_H1
	TLR9	CpG DNA	High IL-12p70 Low IL-10 IL-6 IFN-α	T_H1
	TLR10	?	?	?
C-type lectins	DC-SIGN	Env of HIV; core protein of HCV; components of *Mycobacterium tuberculosis; Helicobacter pylori*, Lewis Ag	*H. pylori*, Lewis Ag Suppresses IL-	T_H2 T regulatory

		pylori, Lewis Ag	12p70 Suppresses TLR signaling in DCs	
NOD	NOD2	Muramyl dipeptide of peptidoglycan	Induces IL-10 in DCs	Weak T cell response (tolerogenic?)
Mannose receptor	Mannose receptor	Mannosylated lipoarabinomannans from bacillus Calmette-Guerin and *M. tuberculosis*	Suppresses IL-12 and TLR signaling in DCs	Weak T cell response? (tolerogenic?)

Abbreviations: dsRNA, double-strand RNA; ssRNA, single-strand RNA; LPS, lipopolysaccharide; T_H2, helper T cell; T_H1, helper T cell; CpG, sequences in DNA recognized by TLR-9; MALP, macrophage-activating lipopeptide; DC-SIGN, DC-specific C-type lectin; NOD, NOTCH protein domain; TLR, Toll-like receptor; HCV, hepatitis C.

Source: B Pulendran: J Immunol 174:2457, 2005. Copyright 2005 The American Association of Immunologists, Inc., with permission.

Table 314–5. Cells of the Innate Immune System and Their Major Roles in Triggering Adaptive Immunity

Cell Type	Major Role in Innate Immunity	Major Role in Adaptive Immunity
Macrophages	Phagocytose and kill bacteria; produce antimicrobial peptides; bind (LPS); produce inflammatory cytokines	Produce IL-1 and TNF-α to upregulate lymphocyte adhesion molecules and chemokines to attract antigen-specific lymphocyte. Produce IL-12 to recruit T_H1 T helper cell responses; upregulate co-stimulatory and MHC molecules to facilitate T and B lymphocyte recognition and activation. Macrophages and dendritic cells, after LPS signaling, upregulate co-stimulatory molecules B7-1 (CD80) and B7-2 (CD86) that are required for activation of antigen-specific antipathogen T cells. There are also Toll-like proteins on B cells and dendritic cells that, after LPS ligation, induce CD80 and CD86 on these cells for T cell antigen presentation.
Plasmacytoid dendritic cells (DCs) of lymphoid lineage	Produce large amounts of interferon-α (IFN-α), which has antitumor and antiviral activity, and are found in T cell zones of lymphoid organs; they circulate in blood	IFN-α is a potent activator of macrophage and mature DCs to phagocytose invading pathogens and present pathogen antigens to T and B cells
Myeloid dendritic cells are of two types; interstitial and Langerhans-derived	Interstitial DCs are strong producers of IL-12 and IL-10 and are located in T cell zones of lymphoid organs, circulate in blood, and are present in the interstices of the lung, heart, and kidney; Langerhans DCs are strong producers of IL-12; are located in T cell zones of lymph nodes, skin epithelia, and the thymic medulla; and circulate in blood	Interstitial DCs are potent activators of macrophage and mature DCs to phagocytose invading pathogens and present pathogen antigens to T and B cells
Natural killer (NK) cells	Kill foreign and host cells that have low levels of MHC+ self-peptides. Express NK receptors that inhibit NK function in the presence of high expression of self-MHC.	Produce TNF-α and IFN-γ, which recruit T_H1 helper T cell responses
NK-T cells	Lymphocytes with both T cell and NK surface markers that recognize lipid antigens of intracellular bacteria such as *Mycobacterium tuberculosis* by CD1 molecules and kill host cells infected with intracellular bacteria.	Produce IL-4 to recruit T_H2 helper T cell responses, IgG1 and IgE production
Neutrophils	Phagocytose and kill bacteria, produce antimicrobial peptides	Produce nitric oxide synthase and nitric oxide, which inhibit apoptosis in lymphocytes and can prolong adaptive immune responses
Eosinophils Mast cells and basophils	Kill invading parasites Release TNF-α, IL-6, and IFN-γ in response to a variety of bacterial PAMPs	Produce IL-5, which recruits Ig-specific antibody responses Produce IL-4, which recruits T_H2 helper T cell responses and recruit IgG1- and IgE-specific antibody responses
Epithelial cells	Produce antimicrobial peptides; tissue-specific epithelia produce mediator of local innate immunity; e.g., lung epithelial cells produce surfactant proteins (proteins within the collectin family) that bind and promote clearance of lung-	Produces TGF-β, which triggers IgA-specific antibody responses

	invading microbes	

Abbreviations: LPS, lipopolysaccharide; PAMP, pathogen-associated molecular patterns; TNF-α, tumor necrosis factor-alpha; IL-4, IL-5, IL-6, IL-10, and IL-12, interleukin 4, 5, 6, 10, and 12, respectively.

Source: Adapted from R Medzhitov, CA Janeway: Curr Opinion Immunol 9:4, 1997. Copyright 1997, with permission from Elsevier.

Two other families of intracellular PRRs are the NLRs (NOD-like receptors) and the RLHs (RIG-like helicases). These families, unlike the TLRs, are composed primarily of soluble intracellular proteins that scan the cytoplasm for intracellular pathogens (Tables 314-2 and 314-3).

The intracellular microbial sensors, NLRs, after triggering, form large cytoplasmic complexes termed *inflammasomes*, which are aggregates of molecules including NOD-like receptor pyrin (NLRP) proteins that are members of the NLR family (Table 314-3). Inflammasomes activate inflammatory caspases and IL-1β in the presence of nonbacterial danger signals (cell stress) and bacterial PAMPs. Mutations in inflammasome proteins can lead to chronic inflammation in a group of periodic febrile diseases called *autoinflammatory syndromes***(Table 314-6).**

Table 314–6. Diseases Associated with Inflammasome Activity

Disease	Clinical Features	Gene Mutated	Etiologic Agent	Inflammasome Involvement	*Anakinra Response
Familial cold autoinflammatory syndrome (FCAS)	Fever, arthralgia, cold-induced urticaria	NALP3		Overactive	Yes
Muckle-Wells syndrome (MWS)	Fever, arthralgia, urticaria, sensorineural deafness, amyloidosis	NAPL3		Overactive	Yes
Chronic infantile neurologic cutaneous and articular syndrome (CINCA, NOMID)	Fever, severe arthralgia, urticaria, neurologic problems, severe amyloidosis	NALP3		Overactive	Yes
Familial Mediterranean fever (FMF)	Fever, peritonitis, pleuritis, amyloidosis	Pyrin		Overactive	Partial
Pyogenic arthritis, pyoderma gangrenosum, and acne syndrome (PAPA)	Pyogenic sterile arthritis	PSTPIP1		Overactive	Yes
Hyperimmunoglobulin D syndrome (HIDS)	Arthralgia, abdominal pain, lymphadenopathy	Mevalonate kinase		To be demonstrated	Yes
Tumor necrosis factor receptor-1–associated syndrome (TRAPS)	Fever, abdominal pain, skin lesions	TNF-R1		To be demonstrated	Yes
Systemic onset juvenile idiopathic arthritis (SOJIA)	Chronic joint inflammation		Unknown	To be demonstrated	Yes
Adult-onset Still's disease (AOSD)	Arthralgia, fever		Unknown	To be demonstrated	Yes
Behçet's disease	Arthralgia, uveitis, ulcers		Unknown	To be demonstrated	Yes
Schnitzler's syndrome	Urticaria, fever, arthralgia		Unknown	To be demonstrated	Yes
Gout	Metabolic arthritis, pain		Uric acid (MSU)	Activated	Yes
Pseudogout	Arthritis		CPPD	Activated	Yes
Contact dermatitis	Urticaria		Irritants	Activated	Unknown
Fever syndrome	Fever	NALP12		Unknown	Unknown
Hydatidiform mole	Hydatid mole	NALP7		Unknown	Unknown
Vitiligo	Skin depigmentation, autoimmunity	NALP1		Unknown	Unknown

*Anakinra is a recombinant IL-1 receptor antagonist that functions to block the biologic activity of naturally occuring interleukin-1 (IL-1).

Source: From F Martinon et al: Ann Rev Immunol 27:229, 2009. Copyright 2009. Reproduced with permission from Annual Reviews Inc.

Effector Cells of Innate Immunity

Cells of the innate immune system and their roles in the first line of host defense are listed in Table 314-5. Equally important as their roles in the mediation of innate immune responses are the roles that each cell type plays in recruiting T and B lymphocytes of the adaptive immune system to engage in specific antipathogen responses.

MONOCYTES-MACROPHAGES

Monocytes arise from precursor cells within bone marrow (Fig. 314-2) and circulate with a half-life ranging from 1 to 3 days. Monocytes leave the peripheral circulation by marginating in capillaries and migrating into a vast extravascular pool. Tissue macrophages arise from monocytes that have migrated out of the circulation and by in situ proliferation of macrophage precursors in tissue. Common locations where tissue macrophages (and certain of their specialized forms) are found are lymph node, spleen, bone marrow, perivascular connective tissue, serous

cavities such as the peritoneum, pleura, skin connective tissue, lung (alveolar macrophages), liver (Kupffer cells), bone (osteoclasts), central nervous system (microglia cells), and synovium (type A lining cells).

Figure 314-2

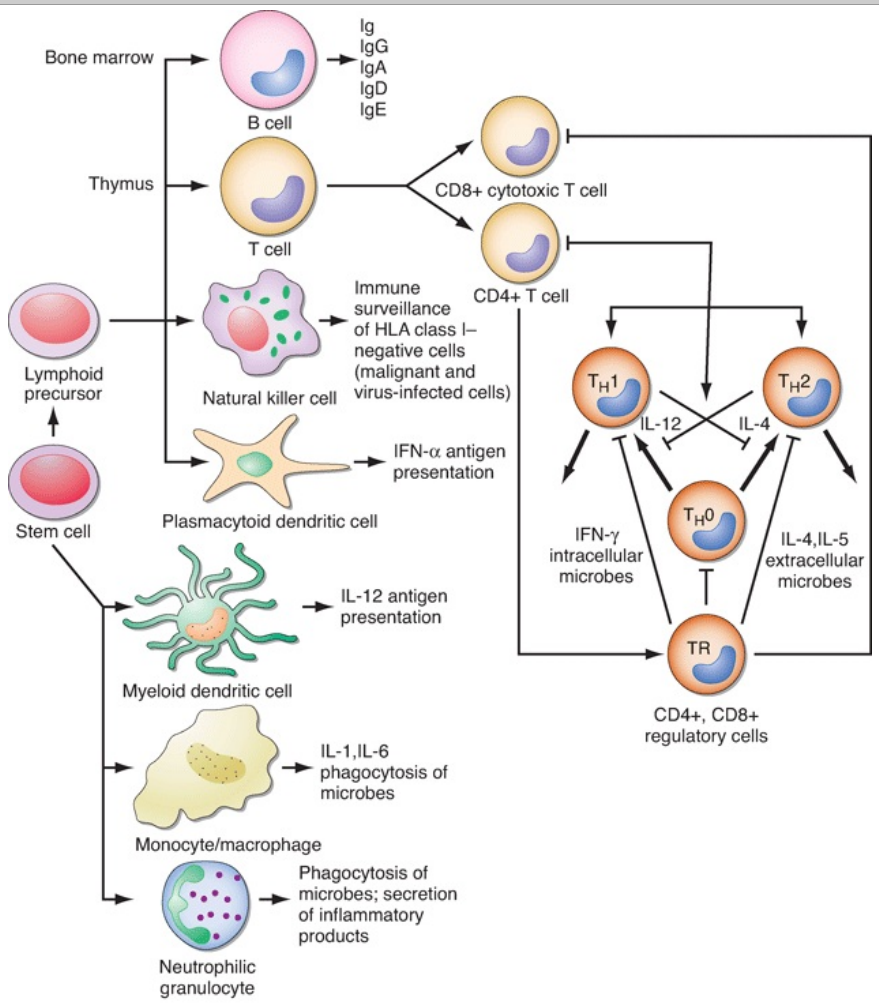

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic model of intercellular interactions of adaptive immune system cells. In this figure, the arrows denote that cells develop from precursor cells or produce cytokines or antibodies; lines ending with bars indicate suppressive intercellular interactions. Stem cells differentiate into either T cells, antigen-presenting dendritic cells, natural killer cells, macrophages, granulocytes, or B cells. Foreign antigen is processed by dendritic cells, and peptide fragments of foreign antigen are presented to CD4+ and/or CD8+ T cells. CD8+ T cell activation leads to induction of cytotoxic T lymphocyte (CTL) or killer T cell generation, as well as induction of cytokine-producing CD8+ cytotoxic T cells. For antibody production against the same antigen, active antigen is bound to sIg within the B cell receptor complex and drives B cell maturation into plasma cells that secrete Ig. T_H1 or T_H2 CD4+ T cells producing interleukin (IL) 4, IL-5, or interferon (IFN)γ regulate the Ig class switching and determine the type of antibody produced. T_H17 cells secrete IL-17, IL-22, IL-26, which contribute to host defense against extracellular bacteria and fungi, particularly at mucosal surfaces. CD4+, CD25+ T regulatory cells produce IL-10 and downregulate T and B cell responses once the microbe has been eliminated. GM-CSF, granulocyte-macrophage colony stimulating factor; TNF, tumor necrosis factor.

In general, monocytes-macrophages are on the first line of defense associated with innate immunity and ingest and destroy microorganisms through the release of toxic products such as hydrogen peroxide (H_2O_2) and nitric oxide (NO). Inflammatory mediators produced by macrophages attract additional effector cells such as neutrophils to the site of infection. Macrophage mediators include prostaglandins; leukotrienes; platelet activating factor; cytokines such as interleukin (IL)-1, tumor necrosis factor (TNF)-α, IL-6, and IL-12; and chemokines (Tables 314-7, 314-8, 314-9, and 314-10).

Table 314–7. Cytokines and Cytokine Receptors

Cytokine	Receptor	Cell Source	Cell Target	Biologic Activity
IL-1α,β	Type I IL-1r, Type II IL-1r	Monocytes/macrophages, B cells, fibroblasts, most epithelial cells including thymic epithelium, endothelial cells	All cells	Upregulates adhesion molecule expression, neutrophil and macrophage emigration, mimics shock, fever, upregulates hepatic acute-phase protein

				production, facilitates hematopoiesis
IL-2	IL-2r α,β, common γ	T cells	T cells, B cells, NK cells, monocytes-macrophages	Promotes T cell activation and proliferation, B cell growth, NK cell proliferation and activation, enhanced monocyte/macrophage cytolytic activity
IL-3	IL-3r, common β	T cells, NK cells, mast cells	Monocytes-macrophages, mast cells, eosinophils, bone marrow progenitors	Stimulates hematopoietic progenitors
IL-4	IL-4r α, common γ	T cells, mast cells, basophils	T cells, B cells, NK cells, monocytes-macrophages, neutrophils, eosinophils, endothelial cells, fibroblasts	Stimulates T_H2 helper T cell differentiation and proliferation. Stimulates B cell Ig class switch to IgG1 and IgE anti-inflammatory action on T cells, monocytes
IL-5	IL-5r α, common γ	T cells, mast cells, eosinophils	Eosinophils, basophils, murine B cells	Regulates eosinophil migration and activation
IL-6	IL-6r, gp130	Monocytes-macrophages, B cells, fibroblasts, most epithelium including thymic epithelium, endothelial cells	T cells, B cells, epithelial cells, hepatocytes, monocytes-macrophages	Induces acute-phase protein production, T and B cell differentiation and growth, myeloma cell growth, and osteoclast growth and activation
IL-7	IL-7r α, common γ	Bone marrow, thymic epithelial cells	T cells, B cells, bone marrow cells	Differentiates B, T, and NK cell precursors, activates T and NK cells
IL-8	CXCR1, CXCR2	Monocytes-macrophages, T cells, neutrophils, fibroblasts, endothelial cells, epithelial cells	Neutrophils, T cells, monocytes-macrophages, endothelial cells, basophils	Induces neutrophil, monocyte, and T cell migration, induces neutrophil adherence to endothelial cells and histamine release from basophils, and stimulates angiogenesis. Suppresses proliferation of hepatic precursors
IL-9	IL-9r α, common γ	T cells	Bone marrow progenitors, B cells, T cells, mast cells	Induces mast cell proliferation and function, synergizes with IL-4 in IgG and IgE production and T cell growth, activation, and differentiation
IL-10	IL-10r	Monocytes-macrophages, T cells, B cells, keratinocytes, mast cells	Monocytes-macrophages, T cells, B cells, NK cells, mast cells	Inhibits macrophage proinflammatory cytokine production, downregulates cytokine class II antigen and B7-1 and B7-2 expression, inhibits differentiation of T_H1 helper T cells, inhibits NK cell function, stimulates mast cell proliferation and function, B cell activation, and differentiation
IL-11	IL-11, gp130	Bone marrow stromal cells	Megakaryocytes, B cells, hepatocytes	Induces megakaryocyte colony formation and maturation, enhances antibody responses, stimulates acute-phase protein production
IL-12 (35-kD and 40-kD subunits)	IL-12r	Activated macrophages, dendritic cells, neutrophils	T cells, NK cells	Induces T_H1 T helper cell formation and lymphokine-activated killer cell formation. Increases CD8+ CTL cytolytic activity; ↓IL-17, ↑IFN-γ.

IL-13	IL-13/IL-4	T cells (T_H2)	Monocytes-macrophages, B cells, endothelial cells, keratinocytes	Upregulates VCAM-1 and C-C chemokine expression on endothelial cells and B cell activation and differentiation, and inhibits macrophage proinflammatory cytokine production
IL-14	Unknown	T cells	Normal and malignant B cells	Induces B cell proliferation
IL-15	IL-15r α, common γ, IL2r β	Monocytes-macrophages, epithelial cells, fibroblasts	T cells, NK cells	Promotes T cell activation and proliferation, angiogenesis, and NK cells
IL-16	CD4	Mast cells, eosinophils, CD8+ T cells, respiratory epithelium	CD4+ T cells, monocytesmacrophages, eosinophils	Promotes chemoattraction of CD4+ T cells, monocytes, and eosinophils. Inhibits HIV replication. Inhibits T cell activation through CD3/T cell receptor
IL-17	IL17r	CD4+ T cells	Fibroblasts, endothelium, epithelium	Enhances cytokine secretion
IL-18	IL-18r (IL-1R-related protein)	Keratinocytes, macrophages	T cells, B cells, NK cells	Upregulates IFN-γ production, enhances NK cell cytotoxicity
IL-21	IL-δγ chain/IL-21R	CD4 T cells	NK cells	Downregulates NK cell –activating molecules, NKG2D/DAP10
IL-23	IL-12Rb1/IL23R	Macrophages, other cell types	T cells	Opposite effects of IL-12 (↑IL-17, ↓IFN-γ)
IFN-α	Type I interferon receptor	All cells	All cells	Promotes antiviral activity. Stimulates T cell, macrophage, and NK cell activity. Direct antitumor effects. Upregulates MHC class I antigen expression. Used therapeutically in viral and autoimmune conditions
IFN-β	Type I interferon receptor	All cells	All cells	Antiviral activity. Stimulates T cell, macrophage, and NK cell activity. Direct antitumor effects. Upregulates MHC class I antigen expression. Used therapeutically in viral and autoimmune conditions
IFN-γ	Type II interferon receptor	T cells, NK cells	All cells	Regulates macrophage and NK cell activations. Stimulates immunoglobulin secretion by B cells. Induction of class II histocompatibility antigens. T_H1 T cell differentiation.
TNF-α	TNFrI, TNFrII	Monocytes-macrophages, mast cells, basophils, eosinophils, NK cells, B cells, T cells, keratinocytes, fibroblasts, thymic epithelial cells	All cells except erythrocytes	Fever, anorexia, shock, capillary leak syndrome, enhanced leukocyte cytotoxicity, enhanced NK cell function, acute phase protein synthesis, proinflammatory cytokine induction.
TNF-β	TNFrI, TNFrII	T cells, B cells	All cells except erythrocytes	Cell cytotoxicity, lymph node and spleen development.
LT-β	LTβR	T cells	All cells except erythrocytes	Cell cytotoxicity, normal lymph node development

G-CSF	G-CSFr; gp130	Monocytes-macrophages, fibroblasts, endothelial cells, thymic epithelial cells, stromal cells	Myeloid cells, endothelial cells	Regulates myelopoiesis. Enhances survival and function of neutrophils. Clinical use in reversing neutropenia after cytotoxic chemotherapy.
GM-CSF	GM-CSFr, common β	T cells, monocytesmacrophages, fibroblasts, endothelial cells, thymic epithelial cells	Monocytes-macrophages, neutrophils, eosinophils, fibroblasts, endothelial cells	Regulates myelopoiesis. Enhances macrophage bactericidal and tumoricidal activity. Mediator of dendritic cell maturation and function. Upregulates NK cell function. Clinical use in reversing neutropenia after cytotoxic chemotherapy.
M-CSF	M-CSFr (*c-fms* protooncogene	Fibroblasts, endothelial cells, monocytesmacrophages, T cells, B cells, epithelial cells including thymic epithelium	Monocytes-macrophages	Regulates monocyte-macrophage production and function.
LIF	LIFr; gp130	Activated T cells, bone marrow stromal cells, thymic epithelium	Megakaryocytes, monocytes, hepatocytes, possibly lymphocyte subpopulations	Induces hepatic acute-phase protein production. Stimulates macrophage differentiation. Promotes growth of myeloma cells and hematopoietic progenitors. Stimulates thrombopoiesis.
OSM	OSMr; LIFr; gp130	Activated monocytesmacrophages and T cells, bone marrow stromal cells, some breast carcinoma cell lines, myeloma cells	Neurons, hepatocytes, monocytes-macrophages, adipocytes, alveolar epithelial cells, embryonic stem cells, melanocytes, endothelial cells, fibroblasts, myeloma cells	Induces hepatic acute-phase protein production. Stimulates macrophage differentiation. Promotes growth of myeloma cells and hematopoietic progenitors. Stimulates thrombopoiesis. Stimulates growth of Kaposi's sarcoma cells.
SCF	SCFr (*c-kit* protooncogene)	Bone marrow stromal cells and fibroblasts	Embryonic stem cells, myeloid and lymphoid precursors, mast cells.	Stimulates hematopoietic progenitor cell growth, mast cell growth, promotes embryonic stem cell migration.
TGF-β (3 isoforms)	Type I, II, III TGF-β receptor	Most cell types	Most cell types	Downregulates T cell, macrophage, and granulocyte responses. Stimulates synthesis of matrix proteins. Stimulates angiogenesis.
Lymphotactin/SCM-1	Unknown	NK cells, mast cells, double negative thymocytes, activated CD8+ T cells	T cells, NK cells	Chemoattractant for lymphocytes. Only known chemokine of C class.
MCP-1	CCR2	Fibroblasts, smoothmuscle cells, activated PBMCs	Monocytes-macrophages, NK cells, memory T cells, basophils	Chemoattractant for monocytes, activated memory T cells, and NK cells. Induces granule release from CD8+ T cells and NK cells. Potent histamine-releasing factor for basophils. Suppresses proliferation of hematopoietic precursors. Regulates monocyte protease production.
MCP-2	CCR1, CCR2	Fibroblasts, activated PBMCs	Monocytes-macrophages, T cells, eosinophils, basophils, NK cells	Chemoattractant for monocytes, memory and naïve T cells, eosinophils, ?NK cells. Activates basophils and eosinophils. Regulates monocyte protease production.

MCP-3	CCR1, CCR2	Fibroblasts, activated PBMCs	Monocytes-macrophages, T cells, eosinophils, basophils, NK cells, dendritic cells	Chemoattractant for monocytes, memory and naïve T cells, dendritic cells, eosinophils, ?NK cells. Activates basophils and eosinophils. Regulates monocyte protease production.
MCP-4	CCR2, CCR3	Lung, colon, small intestinal epithelial cells, activated endothelial cells	Monocytes-macrophages, T cells, eosinophils, basophils	Chemoattractant for monocytes, T cells, eosinophils, and basophils
Eotaxin	CCR3	Pulmonary epithelial cells, heart	Eosinophils, basophils	Potent chemoattractant for eosinophils and basophils. Induces allergic airways disease. Acts in concert with IL-5 to activate eosinophils. Antibodies to eotaxin inhibit airway inflammation.
TARC	CCR4	Thymus, dendritic cells, activated T cells	T cells, NK cells	Chemoattractant for T and NK cells.
MDC	CCR4	Monocytes-macrophages, dendritic cells, thymus	Activated T cells	Chemoattractant for activated T cells. Inhibits infection with T cell tropic HIV.
MIP-1α	CCR1, CCR5	Monocytes-macrophages, T cells	Monocytes-macrophages, T cells, dendritic cells, NK cells, eosinophils, basophils	Chemoattractant for monocytes, T cells, dendritic cells, NK cells, and weak chemoattractant for eosinophils and basophils. Activates NK cell function. Suppresses proliferation of hematopoietic precursors. Necessary for myocarditis associated with Coxsackie virus infection. Inhibits infection with monocytotropic HIV.
MIP-1β	CCR5	Monocytes-macrophages, T cells	Monocytesmacrophages, T cells, NK cells, dendritic cells	Chemoattractant for monocytes, T cells, and NK cells. Activates NK cell function. Inhibits infection with monocytotropic HIV.
RANTES	CCR1, CCR2, CCR5	Monocytes-macrophages, T cells, fibroblasts, eosinophils	Monocytes-macrophages, T cells, NK cells, dendritic cells, eosinophils, basophils	Chemoattractant for monocytesmacrophages, CD4+, CD45Ro+T cells, CD8+ T cells, NK cells, eosinophils, and basophils. Induces histamine release from basophils. Inhibits infections with monocytotropic HIV.
LARC/MIP-3α/Exodus-1	CCR6	Dendritic cells, fetal liver cells, activated T cells	T cells, B cells	Chemoattractant for lymphocytes.
ELC/MIP-3β	CCR7	Thymus, lymph node, appendix	Activated T cells and B cells	Chemoattractant for B and T cells. Receptor upregulated on EBV-infected B cells and HSV-infected T cells.
I-309/TCA-3	CCR8	Activated T cells	Monocytes-macrophages, T cells	Chemoattractant for monocytes. Prevents glucocorticoid-induced apoptosis in some T cell lines.
SLC/TCA-4/Exodus-2	Unknown	Thymic epithelial cells, lymph node, appendix and spleen	T cells	Chemoattractant for T lymphocytes. Inhibits hematopoiesis.
DC-CK1/PARC	Unknown	Dendritic cells in secondary lymphoid tissues	Naïve T cells	May have a role in induction of immune responses.
TECK	Unknown	Dendritic cells, thymus,	T cells,	Thymic dendritic cell

		liver, small intestine	monocytesmacrophages, dendritic cells	–derived cytokine, possibly involved in T cell development
GRO-α/MGSA	CXCR2	Activated granulocytes, monocyte-macrophages, and epithelial cells.	Neutrophils, epithelial cells, ?endothelial cells	Neutrophil chemoattractant and activator. Mitogenic for some melanoma cell lines. Suppresses proliferation of hematopoietic precursors. Angiogenic activity.
GRO-β/MIP-2α	CXCR2	Activated granulocytes and monocyte-macrophages	Neutrophils and ? endothelial cells.	Neutrophil chemoattractant and activator. Angiogenic activity.
NAP-2	CXCR2	Platelets	Neutrophils, basophils	Derived from platelet basic protein. Neutrophil chemoattractant and activator.
IP-10	CXCR3	Monocytes-macrophages, T cells, fibroblasts, endothelial cells, epithelial cells	Activated T cells, tumorinfiltrating lymphocytes, ? endothelial cells, ?NK cells	IFN-γ-inducible protein that is a chemoattractant for T cells. Suppresses proliferation of hematopoietic precursors.
MIG	CXCR3	Monocytes-macrophages, T cells, fibroblasts	Activated T cells, tumorinfiltrating lymphocytes	IFN-γ-inducible protein that is a chemoattractant for T cells. Suppresses proliferation of hematopoietic precursors.
SDF-1	CXCR4	Fibroblasts	T cells, dendritic cells, ? basophils, ?endothelial cells	Low-potency, high-efficacy T cell chemoattractant. Required for B-lymphocyte development. Prevents infection of CD4+, CXCR4+ cells by T cell tropic HIV.
Fractalkine	CX3CR1	Activated endothelial cells	NK cells, T cells, monocytes-macrophages	Cell-surface chemokine/mucin hybrid molecule that functions as a chemoattractant, leukocyte activator, and cell adhesion molecule.
PF-4	Unknown	Platelets, megakaryocytes	Fibroblasts, endothelial cells	Chemoattractant for fibroblasts. Suppresses proliferation of hematopoietic precursors. Inhibits endothelial cell proliferation and angiogenesis.

Abbreviations: IL, interleukin; NK, natural killer; T_H1 and T_H2, helper T cell subsets; Ig, immunoglobulin; CXCR, CXC-type chemokine receptor; B7-1, CD80, B7-2, CD86; PBMC, peripheral blood mononuclear cells; VCAM, vascular cell adhesion molecule; IFN, interferon; MHC, major histocompatibility complex; TNF, tumor necrosis factor; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte-macrophage CSF; M-CSF, macrophage CSF; LIF, leukemia inhibitory factor; OSM, oncostatin M; SCF, stem cell factor; TGF, transforming growth factor; MCP, monocyte chemotactic protein; CCR, CC-type chemokine receptor; TARC, thymus- and activation-regulated chemokine; MDC, macrophage-derived chemokine; MIP, macrophage inflammatory protein; RANTES, regulated on activation, normally T cell–expressed and –secreted; LARC, liver- and activation-regulated chemokine; EBV, Epstein-Barr virus; ELC, EB11 ligand chemokine (MIP-1b); HSV, herpes simplex virus; TCA, T-cell activation protein; DC-CK, dendritic cell chemokine; PARC, pulmonary- and activation-regulated chemokine; SLC, secondary lymphoid tissue chemokine; TECK, thymus expressed chemokine; GRP, growth-related peptide; MGSA, melanoma growth-stimulating activity; NAP, neutrophil-activating protein; IP-10, IFN-γ-inducible protein-10; MIG, monokine induced by IFN-γ ; SDF, stromal cell-derived factor; PF, platelet factor.

Source: Data from JS Sundy et al: Appendix B, *in Inflammation, Basic Principles and Clinical Correlates,* 3rd ed, J Gallin and R Snyderman (eds). Philadelphia, Lippincott Williams and Wilkins, 1999.

Table 314–8. CC, CXC_1, CX_3, C_1 and XC Families of Chemokines and Chemokine Receptors

Chemokine Receptor	Chemokine Ligands	Cell Types	Disease Connection
CCR1	CCL3 (MIP-1α), CCL5 (RANTES), CCL7 (MCP-3), CCL14 (HCC1)	T cells, monocytes, eosinophils, basophils	Rheumatoid arthritis, multiple sclerosis

CCR2	CCL2 (MCP-1), CCL8 (MCP-2), CCL7 (MCP-3), CCL13 (MCP-4), CCL16 (HCC4)	Monocytes, dendritic cells (immature), memory T cells	Atherosclerosis, rheumatoid arthritis, multiple sclerosis, resistance to intracellular pathogens, type 2 diabetes mellitus
CCR3	CCL11 (eotaxin), CCL13 (eotaxin-2), CCL7 (MCP-3), CCL5 (RANTES), CCL8 (MCP-2), CCL13 (MCP-4)	Eosinophils, basophils, mast cells, T_H2, platelets	Allergic asthma and rhinitis
CCR4	CCL17 (TARC), CCL22 (MDC)	T cells (T_H2) dendritic cells (mature), basophils, macrophages, platelets	Parasitic infection, graft rejection, T cell homing to skin
CCR5	CCL3 (MIP-1α), CCL4 (MIP-1α), CCL5 (RANTES), CCL11 (eotaxin), CCL14 (HCC1), CCL16 (HCC4)	T cells, monocytes	HIV-1 co-receptor (T cell–tropic strains), transplant rejection
CCR6	CCL20 (MIP-3α, LARC)	T cells (T regulatory and memory), B cells, dendritic cells	Mucosal humoral immunity, allergic asthma, intestinal T cell homing
CCR7	CCL19 (ELC), CCL21 (SLC)	T cells, dendritic cells (mature)	Transport of T cells and dendritic cells to lymph nodes, antigen presentation, and cellular immunity
CCR8	CCL1 (1309)	T cells (T_H2), monocytes, dendritic cells	Dendritic cell migration to lymph node, type 2 cellular immunity, granuloma formation
CCR9	CCL25 (TECK)	T cells, IgA+ plasma cells	Homing of T cells and IgA+ plasma cells to the intestine, inflammatory bowel disease
CCR10	CCL27 (CTACK), CCL28 (MEC)	T cells	T cell homing to intestine and skin
CXCR1	CXCL8 (interleukin-8), CXCL6 (GCP2)	Neutrophils, monocytes	Inflammatory lung disease, COPD
CXCR2	CXCL8, CXCL1 (GROα), CXCL2 (GROα), CXCL3 (GROα), CXCL5 (ENA-78), CXCL6	Neutrophils, monocytes, microvascular endothelial cells	Inflammatory lung disease, COPD, angiogenic for tumor growth
CXCR3-A	CXCL9 (MIG), CXCL10 (IP-10), CXCL11 (I-TAC)	Type 1 helper cells, mast cells, mesangial cells	Inflammatory skin disease, multiple sclerosis, transplant rejection
CXCR3-B	CXCL4 (PF4), CXCL9 (MIG), CXCL10 (IP-10), CXCL11 (I-TAC)	Microvascular endothelial cells, neoplastic cells	Angiostatic for tumor growth
CXCR4	CXCL12 (SDF-1)	Widely expressed	HIV-1 co-receptor (T cell–tropic), tumor metastases, hematopoiesis
CXCR5	CXCL13 (BCA-1)	B cells, follicular helper T cells	Formation of B cell follicles
CXCR6	CXCL16 (SR-PSOX)	CD8+ T cells, natural killer cells, and memory CD4+ T cells	Inflammatory liver disease, atherosclerosis (CXCL16)
CX_3CR1	CX3CL1 (fractalkine)	Macrophages, endothelial cells, smooth-muscle cells	Atherosclerosis
XCR1	XCL1 (lymphotactin), XCL2	T cells, natural killer cells	Rheumatoid arthritis, IgA nephropathy, tumor response

Abbreviations: MIP, macrophage inflammatory protein; MCP, monocyte chemoattractant protein; HCC, hemofiltrate chemokine; T_H2, type 2 helper T cells; TARC, thymus- and activation-regulated chemokine; MDC, macrophage-derived chemokine; LARC, liver- and activation-regulated chemokine; ELC, Epstein-Barr I1-ligand chemokine; SLC, secondary lymphoid-tissue chemokine; TECK, thymus-expressed chemokine; CTACK, cutaneous T cell–attracting chemokine; MEC, mammary-enriched chemokine; GCP, granulocyte chemotactic protein; COPD, chronic obstructive pulmonary disease; GRO, growth-regulated oncogene; ENA, epithelial-cell–derived neutrophil-activating peptide; MIG, monokine induced by interferon-γ IP-10, interferon inducible 10; I-TAC, interferon-inducible T-cell alpha chemoattractant; PF, platelet factor; SDF, stromal-cell–derived factor; BCA-1, B-cell chemoattractant 1; SR-PSOX, scavenger receptor for phosphatidylserine-containing oxidized lipids.

Source: From IF Charo, RM Ranshohoff: N Engl J Med 354:610, 2006, with permission. Copyright Massachusetts Medical Society. All rights reserved.

Table 314–9. Major Structural Families of Cytokines

Four α-helix-bundle family interleukins	Interleukin-2 (IL-2) Subfamily:
	Interleukins: IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-11, IL-12, IL-13, IL-15, IL-21, IL-23
	Not called interleukins: Colony-stimulating factor-1 (CSF1), granulocyte–macrophage colony-stimulating factor (CSF2), Flt-3 ligand, erythropoietin (EPO), thrombopoietin (THPO), leukocyte

	inhibitory factor (LIF)
	Not interleukins: Growth hormone (GH1), prolactin (PRL), leptin (LEP), cardiotrophin (CTF1), ciliary neurotrophic factor (CNTF), cytokine receptor-like factor 1 (CLC or CLF) Interferon (IFN) subfamily: IFN-β, IFN-α
	IL-10 subfamily: IL-10, IL-19, IL-20, IL-22, IL-24, and IL-26
IL-1 family	IL-1α (IL1A), IL-1β (IL1B), IL-18 (IL18), and paralogues, IL-17A, IL-17B, IL-17C, IL-17D, IL-17E, IL-17F
Chemokines	IL-8, MCP-1, MCP-2, MCP-3, MCP-4, eotaxin, TARC, LARC/MIP-3α, MDC, MIP-1α, MIP-1β, RANTES, MIP-3β, I-309, SLC, PARC, TECK, GROα, GROβ, NAP-2, IP-19, MIG, SDF-1, PF4

Abbreviations: GRO, growth-related peptide; IL, interleukin; IP, INF-γ-inducible protein; LARC, liver- and activation-regulated chemokine; MCP, monocyte chemotactic protein; MDC, macrophage-derived chemokine; MIG, monokine induced by IFN-γ ; MIP, macrophage inflammatory protein; NAP, neutrophil-activating protein; PARC, pulmonary- and activation-regulated chemokine; PF4, platelet factor; RANTES, regulated on activation, normally T cell–expressed and –secreted; SDF, stromal cell–derived factor; SLC, secondary lymphoid tissue.

Source: Adapted from JW Schrader: Trends Immunol 23:573, 2002. Copyright 2002, with permission from Elsevier.

Table 314–10. Cytokine Families Grouped by Structural Similarity

Hematopoietins	IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, IL-16, IL-17, IL-21, IL-23, EPO, LIF, GM-CSF, G-CSF, OSM, CNTF, GH, and TPO TNF-α, LT-α, LT-β, CD40L, CD30L, CD27L, 4-1BBL, OX40, OPG, and FasL
IL-1	IL-1α, IL-1β, IL-1ra, IL-18, bFGF, aFGF, and ECGF
PDGF	PDGF A, PDGF B, and M-CSF
TGF-β	TGF-β and BMPs (1,2,4 etc.)
C-X-C chemokines	IL-8, Gro-α/β/γ, NAP-2, ENA78, GCP-2, PF4, CTAP-3, MIG, and IP-10
C-C chemokines	MCP-1, MCP-2, MCP-3, MIP-1α, MIP-1β, RANTES

Abbreviations: aFGF, acidic fibroblast growth factor; 4-1BBL, 401 BB ligand; bFGF, basic fibroblast growth factor; BMP, bone marrow morphogenetic proteins; C-C, cysteine-cysteine; CD, cluster of differentiation; CNTF, ciliary neurotrophic factor; CTAP, connective tissue–activating peptide; C-X-C, cysteine-x-cysteine; ECGF, endothelial cell growth factor; EPO, erythropoietin; FasL, Fas ligand; GCP-2, granulocyte chemotactic protein 2; G-CSF, granulocyte colony-stimulating factor; GH, growth hormone; GM-CSF, granulocyte-macrophage colony-stimulating factor; Gro, growth-related gene products; IFN, interferon; IL, interleukin; IP, interferon-γ inducible protein; LIF, leukemia inhibitory factor; LT, lymphotoxin; MCP, monocyte chemoattractant; M-CSF, macrophage colony-stimulating factor; MIG, monokine induced by interferon-γ; MIP, macrophage inflammatory protein; NAP-2, neutrophil activating protein 2; OPG, osteoprotegerin; OSM, oncostatin M; PDGF, platelet-derived growth factor; PF, platelet factor; R, receptor; RANTES, regulated on activation, normal T cell–expressed and –secreted; TGF, transforming growth factor; TNF, tumor necrosis factor; TPO, thyroperoxidase.

Although monocytes-macrophages were originally thought to be the major antigen-presenting cells (APCs) of the immune system, it is now clear that cell types called dendritic cells are the most potent and effective APCs in the body (see below). Monocytes-macrophages mediate innate immune effector functions such as destruction of antibody-coated bacteria, tumor cells, or even normal hematopoietic cells in certain types of autoimmune cytopenias. Monocytes-macrophages ingest bacteria or are infected by viruses, and in doing so, they frequently undergo programmed cell death or *apoptosis*. Macrophages that are infected by intracellular infectious agents are recognized by dendritic cells as infected and apoptotic cells and are phagocytosed by dendritic cells. In this manner, dendritic cells "cross-present" infectious agent antigens of macrophages to T cells. Activated macrophages can also mediate antigen-nonspecific lytic activity and eliminate cell types such as tumor cells in the absence of antibody. This activity is largely mediated by cytokines (i.e., TNF-α and IL-1). Monocytes-macrophages express lineage-specific molecules (e.g., the cell-surface LPS receptor, CD14) as well as surface receptors for a number of molecules, including the Fc region of IgG, activated complement components, and various cytokines (Table 314-7).

DENDRITIC CELLS

Human dendritic cells (DCs) are heterogenous and contain several subsets, including myeloid DCs and plasmacytoid DCs. Myeloid DCs can differentiate into either macrophages-monocytes or tissue-specific DCs. In contrast to myeloid DCs, plasmacytoid DCs are inefficient antigen-presenting cells but are potent producers of type I interferon (IFN) (e.g., IFN-α) in response to viral infections. The maturation of DCs is regulated through cell-to-cell contact and soluble factors, and DCs attract immune effectors through secretion of chemokines. When DCs come in contact with bacterial products, viral proteins, or host proteins released as danger signals from distressed host cells (Figs. 314-2 and 314-3), infectious agent molecules bind to various TLRs and activate DCs to release cytokines and chemokines that drive cells of the innate immune system to become activated to respond to the invading organism, and recruit T and B cells of the adaptive immune system to respond. Plasmacytoid DCs produce antiviral IFN-α that activates NK cell killing of pathogen-infected cells; IFN-α also activates T cells to mature into antipathogen cytotoxic (killer) T cells. Following contact with pathogens, both plasmacytoid and myeloid DCs produce chemokines that attract helper and cytotoxic T cells, B cells, polymorphonuclear cells, and naïve and memory T cells as well as regulatory T cells to ultimately dampen the immune response once the pathogen is controlled. TLR engagement on DCs upregulates MHC class II, B7-1 (CD80), and B7-2 (CD86), which enhance DC-specific antigen presentation and induce cytokine production (Table 314-7). Thus, DCs are important bridges between early (innate) and later (adaptive) immunity. DCs also modulate and determine the types of immune responses induced by pathogens via the TLRs expressed on DCs (TLR7-9 on plasmacytoid DCs, TLR4 on monocytoid DCs) and via the TLR adapter proteins that are induced to associate with TLRs (Fig. 314-1, Table 314-4). In addition, other PRRs, such as C-type lectins, NLRs, and mannose receptors, upon ligation by pathogen products, activate cells of the adaptive immune

system and, like TLR stimulation, by a variety of factors, determine the type and quality of the adaptive immune response that is triggered (Table 314-4).

Figure 314-3

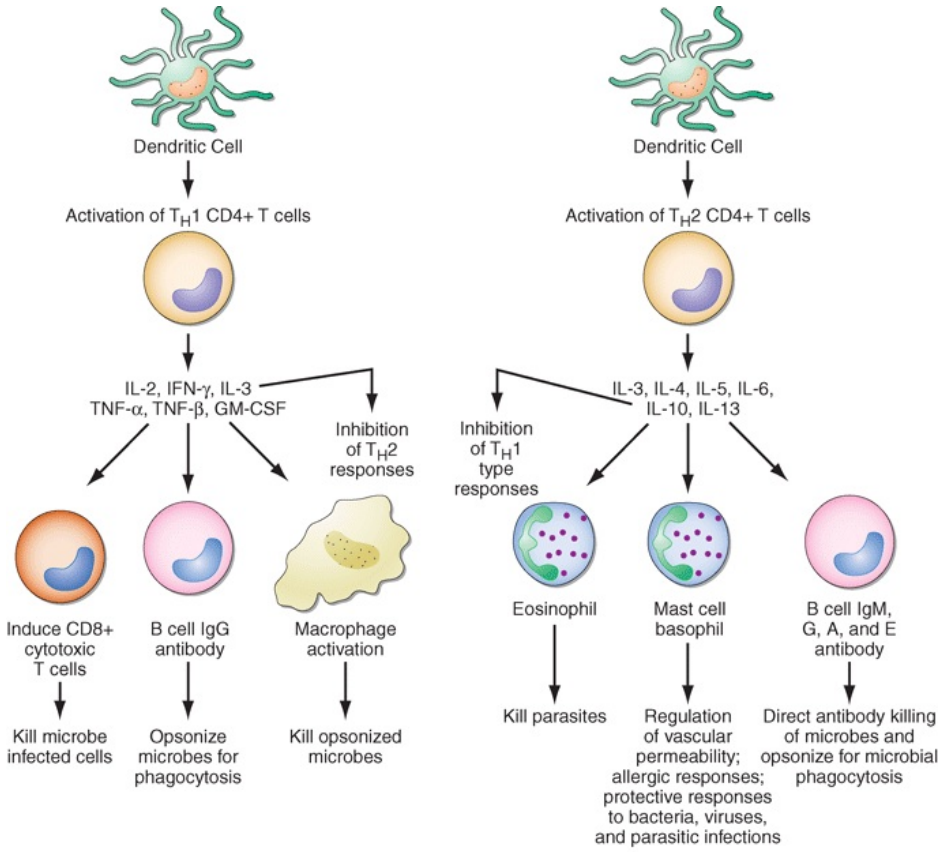

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CD4+ helper T1 (T_H1) cells and T_H2 T cells secrete distinct but overlapping sets of cytokines. T_H1 CD4+ cells are frequently activated in immune and inflammatory reactions against intracellular bacteria or viruses, while T_H2 CD4+ cells are frequently activated for certain types of antibody production against parasites and extracellular encapsulated bacteria; they are also activated in allergic diseases. GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN, interferon; IL, interleukin; TNF, tumor necrosis factor. *(Adapted from Romagnani; with permission.*)

LARGE GRANULAR LYMPHOCYTES/NATURAL KILLER CELLS

Large granular lymphocytes (LGLs) or NK cells account for ~5–15% of peripheral blood lymphocytes. NK cells are nonadherent, nonphagocytic cells with large azurophilic cytoplasmic granules. NK cells express surface receptors for the Fc portion of IgG (CD16) and for NCAM-I (CD56), and many NK cells express T lineage markers, particularly CD8, and proliferate in response to IL-2. NK cells arise in both bone marrow and thymic microenvironments.

Functionally, NK cells share features with both monocytes- macrophages and neutrophils in that they mediate both antibody-dependent cellular cytotoxicity (ADCC) and NK cell activity. ADCC is the binding of an opsonized (antibody-coated) target cell to an Fc receptor–bearing effector cell via the Fc region of antibody, resulting in lysis of the target by the effector cell. NK cell cytotoxicity is the nonimmune (i.e., effector cell never having had previous contact with the target), MHC-unrestricted, non-antibody-mediated killing of target cells, which are usually malignant cell types, transplanted foreign cells, or virus-infected cells. Thus, NK cell cytotoxicity may play an important role in immune surveillance and destruction of malignant and virally infected host cells. NK cell hyporesponsiveness is also observed in patients with *Chédiak-Higashi syndrome*, an autosomal recessive disease associated with fusion of cytoplasmic granules and defective degranulation of neutrophil lysosomes.

NK cells have a variety of surface receptors that have inhibitory or activating functions and belong to two structural families. These families include the immunoglobulin superfamily and the lectin-like type II transmembrane proteins. NK immunoglobulin superfamily receptors include the killer cell immunoglobulin-like activating or inhibitory receptors (KIRs), many of which have been shown to have HLA class I ligands. The KIRs are made up proteins with either two (KIR2D) or three (KIR3D) extracellular immunoglobulin domains (D). Moreover, their nomenclature designates their function as either inhibitory KIRs with a long (L) cytoplasmic tail and immunoreceptor tyrosine-based inhibitory motif (ITIM) (KIRDL) or activating KIRs with a short (S) cytoplasmic tail (KIRDS). NK cell inactivation by KIRs is a central mechanism to prevent damage to normal host cells. Genetic studies have demonstrated the association of KIRs with viral infection outcome and autoimmune disease **(Table 314-11)**.

Table 314–11. Association of KIRs with Disease

Disease	KIR Association	Observation

Psoriatic arthritis	KIR2DS1/KIR2DS2; HLA-Cw group homozygosity	Susceptibility
Spondylarthritides	Increased KIR3DL2 expression	May contribute to disease pathology
	Interaction HLA-B27 homodimers with KIR3DL1/KIR3DL2; independent of peptide	May contribute to disease pathogenesis
Ankylosing spondylitis	KIR3DL1/3DS1; HLA B27 genotypes	Susceptibility
Rheumatoid vasculitis	KIR2DS2; HLA-Cw*03	Susceptibility
	Increased KIR2L2/2DS2 in patients with extra articular manifestations	Clinical manifestations may have different genetic backgrounds with respect to KIR genotype
Rheumatoid arthritis	Decreased KIR2DS1/3DS1 in patients without bone erosions	Susceptibility
	KIR2DS4; HLA-Cw4	Susceptibility
Scleroderma	KIR2DS2+/KIR2DL2-	Susceptibility
Behçet's disease	Altered KIR3DL1 expression	Associated with severe eye disease
Psoriasis vulgaris	2DS1; HLA-Cw*06	Susceptibility
	2DS1; 2DL5; Haplotype B	Susceptibility
IDDM	KIR2DS2; HLA-C1	Susceptibility
Type 1 diabetes	KIR2DS2; HLA-C1 and no HLA-C2, no HLA-Bw4	Increased disease progression
Preeclampsia	KIR2DL1 with fewer KIR2DS (mother); HLA-C2 (fetus)	Increased disease progression
AIDS	KIR3DS1; HLA-Bw4Ile80	Decreased disease progression
	KIR3DS1 homozygous; No HLA-Bw4Ile^{80}	Increased disease progression
HCV infection	KIR2DL3 homozygous; HLA-C1 homozygous	Decreased disease progression
Cervical neoplasia (HPV induced)	KIR3DS1; HLA-C1 homozygous and no HLA-Bw4	Increased disease progression
Malignant melanoma	KIR2DL2 and/or KIR2DL3; HLA-C1	Increased disease progression

Abbreviations: HCV, hepatitis C virus; HLA, human leukocyte antigen; HPV, human papillomavirus; IDDM, insulin-dependent diabetes mellitus; KIR, killer cell immunoglobulin-like receptor.

Source: Adapted from Diaz-Pena et al.

In addition to the KIRs, a second set of immunoglobulin superfamily receptors include the natural cytotoxicity receptors (NCRs), which include NKp46, NKp30, and NKp44. These receptors help to mediate NK cell activation against target cells. The ligands to which NCRs bind on target cells remain largely undefined.

NK cell signaling is, therefore, a highly coordinated series of inhibiting and activating signals that prevent NK cells from responding to uninfected, nonmalignant self-cells; however, they are activated to attack malignant and virally infected cells **(Fig. 314-4)**. Recent evidence suggests that NK cells, although not possessing rearranging immune recognition genes, may be able to mediate recall for NK cell responses to viruses and for immune responses such as contact hypersensitivity.

Figure 314-4

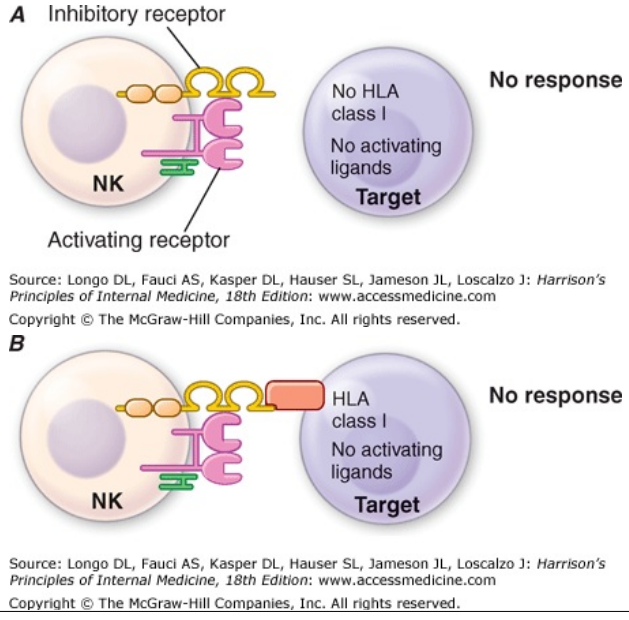

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

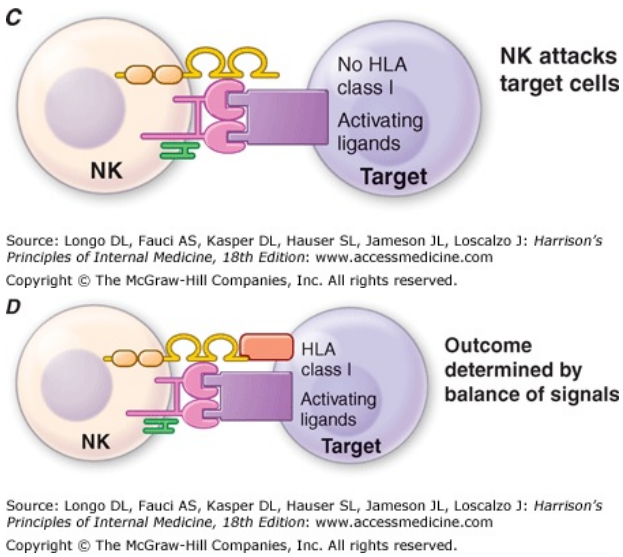

Encounters between NK cells: potential targets and possible outcomes. The amount of activating and inhibitory receptors on the NK cells and the amount of ligands on the target cell, as well as the qualitative differences in the signals transduced, determine the extent of the NK response. ***A.*** When target cells have no HLA class I nor activating ligands, NK cells cannot kill target cells. ***B.*** When target cells bear self-HLA, NK cells cannot kill targets. ***C.*** When target cells are pathogen infected and have downregulated HLA and express activating ligands, NK cells kill target cells. ***D.*** When NK cells encounter targets with both self-HLA and activating receptors, then the level of target killing is determined by the balance of inhibitory and activating signals to the NK cell. HLA, human leukocyte antigen; NK, natural killer. *(Adapted from Lanier; reproduced with permission from Annual Reviews Inc. Copyright 2011 by Annual Reviews Inc.)*

Some NK cells express CD3 and invariant T cell receptor (TCR) alpha chains and are termed *NK T cells*. TCRs of NK T cells recognize lipid molecules of intracellular bacteria when presented in the context of CD1d molecules on APCs. Upon activation, NK T cells secrete effector cytokines such as IL-4 and IFNγ. This mode of recognition of intracellular bacteria such as *Listeria monocytogenes* and *Mycobacterium tuberculosis* by NK T cells leads to induction of activation of DCs and is thought to be an important innate defense mechanism against these organisms.

NEUTROPHILS, EOSINOPHILS, AND BASOPHILS

Granulocytes are present in nearly all forms of inflammation and are amplifiers and effectors of innate immune responses (Figs. 314-2 and 314-3). Unchecked accumulation and activation of granulocytes can lead to host tissue damage, as seen in neutrophil- and eosinophil-mediated *systemic necrotizing vasculitis*. Granulocytes are derived from stem cells in bone marrow. Each type of granulocyte (neutrophil, eosinophil, or basophil) is derived from a different subclass of progenitor cell that is stimulated to proliferate by colony-stimulating factors (Table 314-7). During terminal maturation of granulocytes, class-specific nuclear morphology and cytoplasmic granules appear that allow for histologic identification of granulocyte type.

Neutrophils express Fc receptors for IgG (CD16) and receptors for activated complement components (C3b or CD35). Upon interaction of neutrophils with opsonized bacteria or immune complexes, azurophilic granules (containing myeloperoxidase, lysozyme, elastase, and other enzymes) and specific granules (containing lactoferrin, lysozyme, collagenase, and other enzymes) are released, and microbicidal superoxide radicals (O_2^-) are generated at the neutrophil surface. The generation of superoxide leads to inflammation by direct injury to tissue and by alteration of macromolecules such as collagen and DNA.

Eosinophils express Fc receptors for IgG (CD32) and are potent cytotoxic effector cells for various parasitic organisms. In *Nippostrongylus brasiliensis* helminth infection, eosinophils are important cytotoxic effector cells for removal of these parasites. Key to regulation of eosinophil cytotoxicity to *N. brasiliensis* worms are antigen-specific T helper cells that produce IL-4, thus providing an example of regulation of innate immune responses by adaptive immunity antigen-specific T cells. Intracytoplasmic contents of eosinophils, such as major basic protein, eosinophil cationic protein, and eosinophil-derived neurotoxin, are capable of directly damaging tissues and may be responsible in part for the organ system dysfunction in the *hypereosinophilic syndromes* (Chap. 60). Since the eosinophil granule contains anti-inflammatory types of enzymes (histaminase, arylsulfatase, phospholipase D), eosinophils may homeostatically downregulate or terminate ongoing inflammatory responses.

Basophils and tissue mast cells are potent reservoirs of cytokines such as IL-4 and can respond to bacteria and viruses with antipathogen cytokine production through multiple TLRs expressed on their surface. Mast cells and basophils can also mediate immunity through the binding of antipathogen antibodies. This is a particularly important host defense mechanism against parasitic diseases. Basophils express high-affinity surface receptors for IgE (FcRI) and, upon cross-linking of basophil-bound IgE by antigen, can release histamine, eosinophil chemotactic factor of anaphylaxis, and neutral protease—all mediators of allergic immediate (anaphylaxis) hypersensitivity responses **(Table 314-12)**. In addition, basophils express surface receptors for activated complement components (C3a, C5a), through which mediator release can be directly effected. Thus, basophils, like most cells of the immune system, can be activated in the service of host defense against pathogens, or they can be activated for mediation release and cause pathogenic responses in allergic and inflammatory diseases. **For further discussion of tissue mast cells, see Chap. 317.**

Table 314–12. Examples of Mediators Released from Human Cells and Basophils

Mediator	Actions
Histamine	Smooth-muscle contraction, increased vascular permeability
Slow reacting substance of anaphylaxis (SRSA) (leukotriene C4, D4, E4)	Smooth-muscle contraction
Eosinophil chemotactic factor of anaphylaxis (ECF-A)	Chemotactic attraction of eosinophils
Platelet-activating factor	Activates platelets to secrete serotonin and other mediators: smooth-muscle contraction; induces vascular permeability
Neutrophil chemotactic factor (NCF)	Chemotactic attraction of neutrophils
Leukotactic activity (leukotriene B4)	Chemotactic attraction of neutrophils
Heparin	Anticoagulant
Basophil kallikrein of anaphylaxis (BK-A)	Cleaves kininogen to form bradykinin

THE COMPLEMENT SYSTEM

The complement system, an important soluble component of the innate immune system, is a series of plasma enzymes, regulatory proteins, and proteins that are activated in a cascading fashion, resulting in cell lysis. There are four pathways of the complement system: the classic activation pathway activated by antigen/antibody immune complexes, the MBL (a serum collectin; Table 314-3) activation pathway activated by microbes with terminal mannose groups, the alternative activation pathway activated by microbes or tumor cells, and the terminal pathway that is common to the first three pathways and leads to the membrane attack complex that lyses cells **(Fig. 314-5)**. The series of enzymes of the complement system are serine proteases.

Figure 314-5

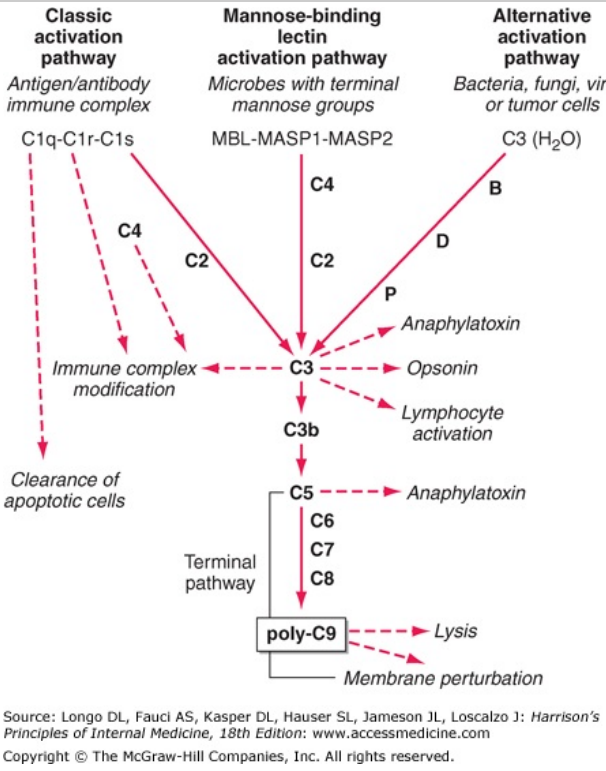

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The four pathways and the effector mechanisms of the complement system. Dashed arrows indicate the functions of pathway components. (*After Morley and Walport; with permission. Copyright Academic Press, London, 2000.)*

Activation of the classic complement pathway via immune complex binding to C1q links the innate and adaptive immune systems via specific antibody in the immune complex. The alternative complement activation pathway is antibody-independent and is activated by binding of C3 directly to pathogens and "altered self" such as tumor cells. In the renal glomerular inflammatory disease *IgA nephropathy*, IgA activates the alternative complement pathway and causes glomerular damage and decreased renal function. Activation of the classic complement pathway via C1, C4, and C2 and activation of the alternative pathway via factor D, C3, and factor B both lead to cleavage and activation of C3. C3 activation fragments, when bound to target surfaces such as bacteria and other foreign antigens, are critical for opsonization (coating by antibody and complement) in preparation for phagocytosis. The MBL pathway substitutes MBL-associated serine proteases (MASPs) 1 and 2 for C1q, C1r, and C1s to activate C4. The MBL activation pathway is activated by mannose on the surface of bacteria and viruses.

The three pathways of complement activation all converge on the final common terminal pathway. C3 cleavage by each pathway results in activation of C5, C6, C7, C8, and C9, resulting in the membrane attack complex that physically inserts into the membranes of target cells or bacteria and lyses them.

Thus, complement activation is a critical component of innate immunity for responding to microbial infection. The functional consequences of complement activation by the three initiating pathways and the terminal pathway are

shown in Fig. 314-5. In general the cleavage products of complement components facilitate microbe or damaged cell clearance (C1q, C4, C3), promote activation and enhancement of inflammation (anaphylatoxins, C3a, C5a), and promote microbe or opsonized cell lysis (membrane attack complex).

Cytokines

Cytokines are soluble proteins produced by a wide variety of hematopoietic and nonhematopoietic cell types (Tables 314-7, 314-8, 314-9, and 314-10). They are critical for both normal innate and adaptive immune responses, and their expression may be perturbed in most immune, inflammatory, and infectious disease states.

Cytokines are involved in the regulation of the growth, development, and activation of immune system cells and in the mediation of the inflammatory response. In general, cytokines are characterized by considerable redundancy; different cytokines have similar functions. In addition, many cytokines are pleiotropic in that they are capable of acting on many different cell types. This pleiotropism results from the expression on multiple cell types of receptors for the same cytokine (see below), leading to the formation of "cytokine networks." The action of cytokines may be (1) autocrine when the target cell is the same cell that secretes the cytokine, (2) paracrine when the target cell is nearby, and (3) endocrine when the cytokine is secreted into the circulation and acts distal to the source.

Cytokines have been named based on presumed targets or based on presumed functions. Those cytokines that are thought to primarily target leukocytes have been named interleukins (IL-1, -2, -3, etc.). Many cytokines that were originally described as having a certain function have retained those names (granulocyte colony-stimulating factor or G-CSF, etc.). Cytokines belong in general to three major structural families: the hematopoietin family; the TNF, IL-1, platelet-derived growth factor (PDGF), and transforming growth factor (TGF) β families; and the CXC and C-C chemokine families (Table 314-10). Chemokines are cytokines that regulate cell movement and trafficking; they act through G protein–coupled receptors and have a distinctive three-dimensional structure. IL-8 is the only chemokine that early on was named an interleukin (Table 314-7).

In general, cytokines exert their effects by influencing gene activation that results in cellular activation, growth, differentiation, functional cell-surface molecule expression, and cellular effector function. In this regard, cytokines can have dramatic effects on the regulation of immune responses and the pathogenesis of a variety of diseases. Indeed, T cells have been categorized on the basis of the pattern of cytokines that they secrete, which results in either humoral immune response (T_H2) or cell-mediated immune response (T_H1). A third type of T helper cell is the T_H17 cell that contributes to host defense against extracellular bacteria and fungi, particularly at mucosal sites (Fig. 314-2).

Cytokine receptors can be grouped into five general families based on similarities in their extracellular amino acid sequences and conserved structural domains. The *immunoglobulin (Ig) superfamily* represents a large number of cell-surface and secreted proteins. The IL-1 receptors (type 1, type 2) are examples of cytokine receptors with extracellular Ig domains.

The hallmark of the *hematopoietic growth factor (type 1) receptor* family is that the extracellular regions of each receptor contain two conserved motifs. One motif, located at the N terminus, is rich in cysteine residues. The other motif is located at the C terminus proximal to the transmembrane region and comprises five amino acid residues, tryptophan-serine-X-tryptophan-serine (WSXWS). This family can be grouped on the basis of the number of receptor subunits they have and on the utilization of shared subunits. A number of cytokine receptors, i.e., IL-6, IL-11, IL-12, and leukemia inhibitory factor, are paired with gp130. There is also a common 150-kDa subunit shared by IL-3, IL-5, and granulocyte-macrophage colony-stimulating factor (GM-CSF) receptors. The gamma chain (γ_c) of the IL-2 receptor is common to the IL-2, IL-4, IL-7, IL-9, and IL-15 receptors. Thus, the specific cytokine receptor is responsible for ligand-specific binding, while the subunits such as gp130, the 150-kDa subunit, and γ_c are important in signal transduction. The γ_c gene is on the X chromosome, and mutations in the γ_c protein result in the *X-linked form of severe combined immune deficiency syndrome (X-SCID)* (Chap. 316).

The members of the *interferon (type II) receptor* family include the receptors for IFN-γ and -β, which share a similar 210-amino-acid binding domain with conserved cysteine pairs at both the amino and carboxy termini. The members of the *TNF (type III) receptor family* share a common binding domain composed of repeated cysteine-rich regions. Members of this family include the p55 and p75 receptors for TNF (TNF-R1 and TNF-R2, respectively); CD40 antigen, which is an important B cell–surface marker involved in immunoglobulin isotype switching; fas/Apo-1, whose triggering induces apoptosis; CD27 and CD30, which are found on activated T cells and B cells; and nerve growth factor receptor.

The common motif for the *seven transmembrane helix family* was originally found in receptors linked to GTP-binding proteins. This family includes receptors for chemokines (Table 314-8), β-adrenergic receptors, and retinal rhodopsin. It is important to note that two members of the chemokine receptor family, CXC chemokine receptor type 4 (CXCR4) and β chemokine receptor type 5 (CCR5), have been found to serve as the two major co-receptors for binding and entry of HIV into CD4-expressing host cells (Chap. 189).

Significant advances have been made in defining the signaling pathways through which cytokines exert their effects intracellularly. The Janus family of protein tyrosine kinases (JAK) is a critical element involved in signaling via the hematopoietin receptors. Four JAK kinases, JAK1, JAK2, JAK3, and Tyk2, preferentially bind different cytokine receptor subunits. Cytokine binding to its receptor brings the cytokine receptor subunits into apposition and allows a pair of JAKs to transphosphorylate and activate one another. The JAKs then phosphorylate the receptor on the tyrosine residues and allow signaling molecules to bind to the receptor, where these molecules become phosphorylated. Signaling molecules bind the receptor because they have domains (SH2, or src homology 2 domains) that can bind phosphorylated tyrosine residues. There are a number of these important signaling molecules that bind the receptor, such as the adapter molecule SHC, which can couple the receptor to the activation of the mitogen-activated protein kinase pathway. In addition, an important class of substrate of the JAKs is the signal transducers and activators of transcription (STAT) family of transcription factors. STATs have SH2 domains that enable them to bind to phosphorylated receptors, where they are then phosphorylated by the JAKs. It appears that different STATs have specificity for different receptor subunits. The STATs then dissociate from the receptor and translocate to the nucleus, bind to DNA motifs that they recognize, and regulate gene expression. The STATs

preferentially bind DNA motifs that are slightly different from one another and thereby control transcription of specific genes. The importance of this pathway is particularly relevant to lymphoid development. Mutations of JAK3 itself also result in a disorder identical to X-SCID; however, since JAK3 is found on chromosome 19 and not on the X chromosome, JAK3 deficiency occurs in boys and girls (Chap. 316).

The Adaptive Immune System

Adaptive immunity is characterized by antigen-specific responses to a foreign antigen or pathogen. A key feature of adaptive immunity is that following the initial contact with antigen (*immunologic priming*), subsequent antigen exposure leads to more rapid and vigorous immune responses (*immunologic memory*). The adaptive immune system consists of dual limbs of cellular and humoral immunity. The principal effectors of cellular immunity are T lymphocytes, while the principal effectors of humoral immunity are B lymphocytes. Both B and T lymphocytes derive from a common stem cell **(Fig. 314-6)**.

Figure 314-6

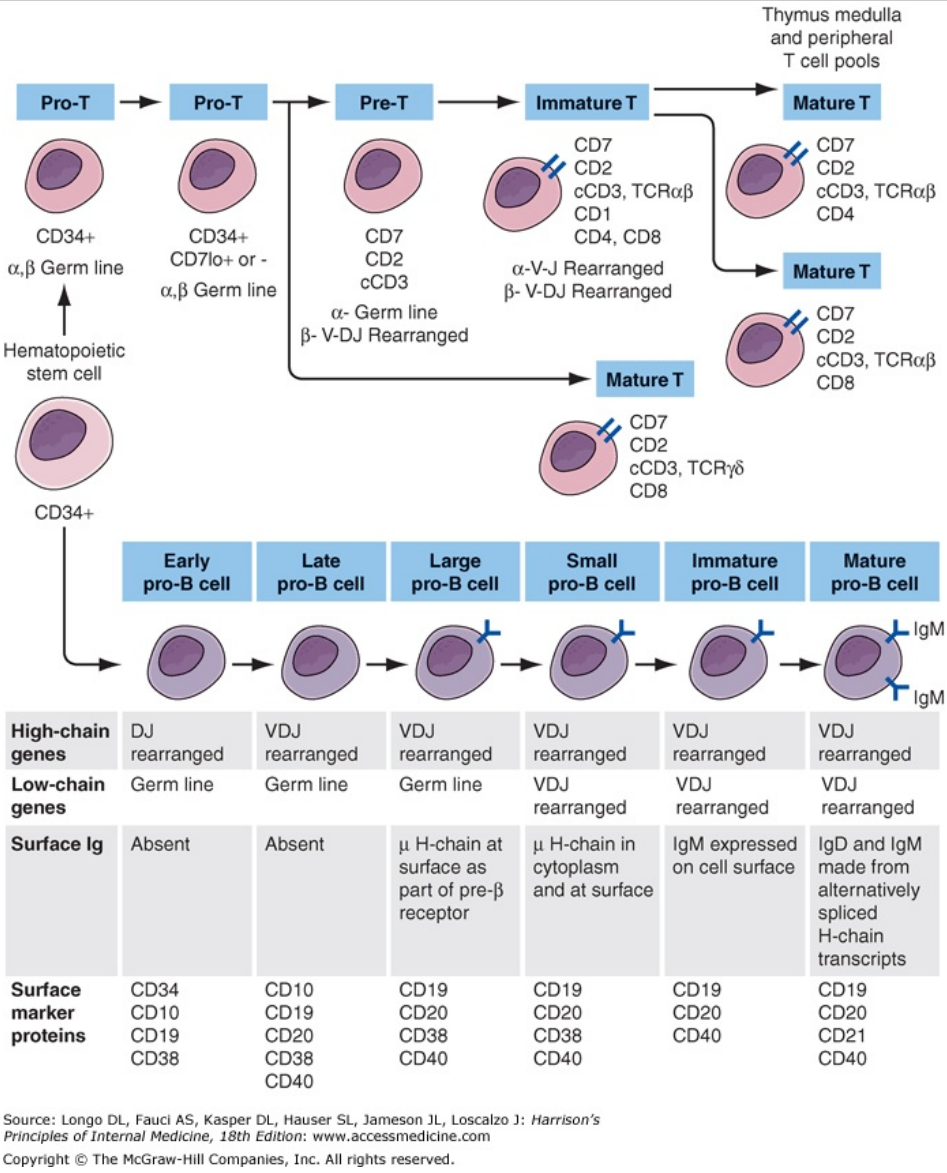

High-chain genes	DJ rearranged	VDJ rearranged	VDJ rearranged	VDJ rearranged	VDJ rearranged	VDJ rearranged
Low-chain genes	Germ line	Germ line	Germ line	VDJ rearranged	VDJ rearranged	VDJ rearranged
Surface Ig	Absent	Absent	μ H-chain at surface as part of pre-β receptor	μ H-chain in cytoplasm and at surface	IgM expressed on cell surface	IgD and IgM made from alternatively spliced H-chain transcripts
Surface marker proteins	CD34 CD10 CD19 CD38	CD10 CD19 CD20 CD38 CD40	CD19 CD20 CD38 CD40	CD19 CD20 CD38 CD40	CD19 CD20 CD40	CD19 CD20 CD21 CD40

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Development stages of T and B cells. Elements of the developing T and B cell receptor for antigen are shown schematically. The classification into the various stages of B cell development is primarily defined by rearrangement of the immunoglobulin (Ig), heavy (H), and light (L) chain genes and by the absence or presence of specific surface markers. *[Adapted from CA Janeway et al (eds): Immunobiology. The Immune Systemic Health and Disease* , 4th ed. New York, Garland, 1999; with permission.] The classification of stages of T cell development is primarily defined by cell-surface marker protein expression (sCD3, surface CD3 expression, cCD3, cytoplasmic CD3 expression; TCR, T cell receptor).

The proportion and distribution of immunocompetent cells in various tissues reflect cell traffic, homing patterns, and functional capabilities. Bone marrow is the major site of maturation of B cells, monocytes-macrophages, dendritic cells, and granulocytes and contains pluripotent stem cells that, under the influence of various colony-stimulating factors, are capable of giving rise to all hematopoietic cell types. T cell precursors also arise from hematopoietic stem cells and home to the thymus for maturation. Mature T lymphocytes, B lymphocytes, monocytes, and dendritic cells enter the circulation and home to peripheral lymphoid organs (lymph nodes, spleen) and mucosal surface-associated lymphoid tissue (gut, genitourinary, and respiratory tracts) as well as the skin and mucous membranes and await activation by foreign antigen.

T CELLS

The pool of effector T cells is established in the thymus early in life and is maintained throughout life both by new T cell production in the thymus and by antigen-driven expansion of virgin peripheral T cells into "memory" T cells that reside in peripheral lymphoid organs. The thymus exports ~2% of the total number of thymocytes per day throughout life, with the total number of daily thymic emigrants decreasing by ~3% per year during the first four decades of life.

Mature T lymphocytes constitute 70–80% of normal peripheral blood lymphocytes (only 2% of the total-body lymphocytes are contained in peripheral blood), 90% of thoracic duct lymphocytes, 30–40% of lymph node cells, and 20–30% of spleen lymphoid cells. In lymph nodes, T cells occupy deep paracortical areas around B cell germinal centers, and in the spleen, they are located in periarteriolar areas of white pulp (Chap. 59). T cells are the primary effectors of cell-mediated immunity, with subsets of T cells maturing into CD8+ cytotoxic T cells capable of lysis of virus-infected or foreign cells (short-lived effector T cells). Two populations of long-lived memory T cells are triggered by infections: effector memory and central memory T cells. Effector memory T cells reside in nonlymphoid organs and respond rapidly to repeated pathogenic infections with cytokine production and cytotoxic functions to kill virus-infected cells. Central memory T cells home to lymphoid organs where they replenish long- and short-lived and effector memory T cells as needed.

In general, CD4+ T cells are also the primary regulatory cells of T and B lymphocyte and monocyte function by the production of cytokines and by direct cell contact (Fig. 314-2). In addition, T cells regulate erythroid cell maturation in bone marrow and, through cell contact (CD40 ligand), have an important role in activation of B cells and induction of Ig isotype switching.

Human T cells express cell-surface proteins that mark stages of intrathymic T cell maturation or identify specific functional subpopulations of mature T cells. Many of these molecules mediate or participate in important T cell functions (Table 314-1, Fig. 314-6).

The earliest identifiable T cell precursors in bone marrow are CD34+ pro-T cells (i.e., cells in which TCR genes are neither rearranged nor expressed). In the thymus, CD34+ T cell precursors begin cytoplasmic (c) synthesis of components of the CD3 complex of TCR-associated molecules (Fig. 314-6). Within T cell precursors, TCR for antigen gene rearrangement yields two T cell lineages, expressing either TCR-αβ chains or TCR-γδ chains. T cells expressing the TCR-αβ chains constitute the majority of peripheral T cells in blood, lymph node, and spleen and terminally differentiate into either CD4+ or CD8+ cells. Cells expressing TCR-γδ chains circulate as a minor population in blood; their functions, although not fully understood, have been postulated to be those of immune surveillance at epithelial surfaces and cellular defenses against mycobacterial organisms and other intracellular bacteria through recognition of bacterial lipids.

In the thymus, the recognition of self-peptides on thymic epithelial cells, thymic macrophages, and dendritic cells plays an important role in shaping the T cell repertoire to recognize foreign antigen (*positive selection*) and in eliminating highly autoreactive T cells (*negative selection*). As immature cortical thymocytes begin to express surface TCR for antigen, autoreactive thymocytes are destroyed (negative selection), thymocytes with TCRs capable of interacting with foreign antigen peptides in the context of self-MHC antigens are activated and develop to maturity (positive selection), and thymocytes with TCRs that are incapable of binding to self-MHC antigens die of attrition (*no selection*). Mature thymocytes that are positively selected are either CD4+ helper T cells or MHC class II –restricted cytotoxic (killer) T cells, or they are CD8+ T cells destined to become MHC class I–restricted cytotoxic T cells. *MHC class I–* or *class II–restricted* means that T cells recognize antigen peptide fragments only when they are presented in the antigen-recognition site of a class I or class II MHC molecule, respectively (Chap. 315).

After thymocyte maturation and selection, CD4 and CD8 thymocytes leave the thymus and migrate to the peripheral immune system. The thymus continues to be a contributor to the peripheral immune system, well into adult life, both normally and when the peripheral T cell pool is damaged, such as occurs in AIDS and cancer chemotherapy.

Molecular Basis of T Cell Recognition of Antigen

The TCR for antigen is a complex of molecules consisting of an antigen-binding heterodimer of either αβ or γδ chains noncovalently linked with five CD3 subunits (γ, δ, ε, ζ, and π) **(Fig. 314-7)**. The CD3 ζ chains are either disulfide-linked homodimers (CD3-ζ_2) or disulfide-linked heterodimers composed of one ζ chain and one π chain. TCR-αβ or TCR-γδ molecules must be associated with CD3 molecules to be inserted into the T cell–surface membrane, TCRα being paired with TCR-β and TCR-γ being paired with TCR-δ. Molecules of the CD3 complex mediate transduction of T cell activation signals via TCRs, while TCR-α and -β or -γ and -δ molecules combine to form the TCR antigen-binding site.

Figure 314-7

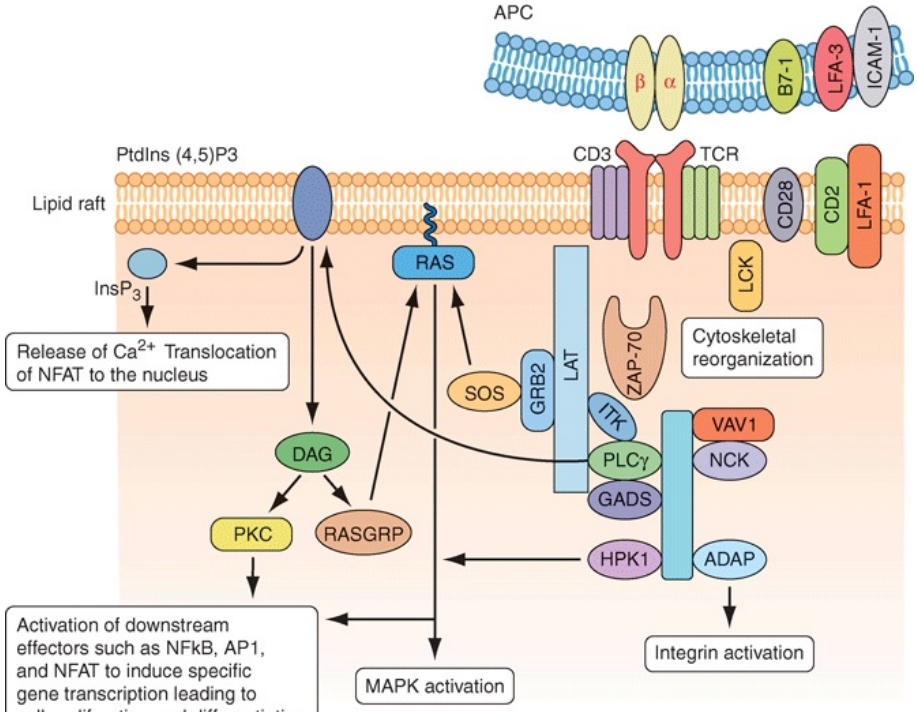

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Signaling through the T cell receptor. Activation signals are mediated via immunoreceptor tyrosine-based activation (ITAM) sequences in LAT and CD3 chains (blue bars) that bind to enzymes and transduce activation signals to the nucleus via the indicated intracellular activation pathways. Ligation of the T-cell receptor (TCR) by MHC complexed with antigen results in sequential activation of LCK and ζ-chain-associated protein kinase of 70 kDa (ZAP-70). ZAP-70 phosphorylates several downstream targets, including LAT (linker for activation of T cells) and SLP76 [SCR homology 2 (SH2) domain-containing leukocyte protein of 76 kDa]. SLP76 is recruited to membrane-bound LAT through its constitutive interaction with GADS (GRB2-related adaptor protein). Together, SLP76 and LAT nucleate a multimolecular signaling complex, which induces a host of downstream responses, including calcium flux, mitogen-activated protein kinase (MAPK) activation, integrin activation, and cytoskeletal reorganization. APC denotes antigen-presenting cell. *(Adapted from Koretzky et al; with permission from Macmillan Publishers Ltd. Copyright 2006.)*

The α, β, γ, and δ TCR for antigen molecules have amino acid sequence homology and structural similarities to immunoglobulin heavy and light chains and are members of the *immunoglobulin gene superfamily* of molecules. The genes encoding TCR molecules are encoded as clusters of gene segments that rearrange during the course of T cell maturation. This creates an efficient and compact mechanism for housing the diversity requirements of antigen receptor molecules. The TCR-α chain is on chromosome 14 and consists of a series of V (variable), J (joining), and C (constant) regions. The TCR-β chain is on chromosome 7 and consists of multiple V, D (diversity), J, and C TCR-β loci. The TCR-γ chain is on chromosome 7, and the TCR-δ chain is in the middle of the TCR-α locus on chromosome 14. Thus, molecules of the TCR for antigen have constant (framework) and variable regions, and the gene segments encoding the α, β, γ, and δ chains of these molecules are recombined and selected in the thymus, culminating in synthesis of the completed molecule. In both T and B cell precursors (see below), DNA rearrangements of antigen receptor genes involve the same enzymes, recombinase activating gene (RAG)1 and RAG2, both DNA-dependent protein kinases.

TCR diversity is created by the different V, D, and J segments that are possible for each receptor chain by the many permutations of V, D, and J segment combinations, by "N-region diversification" due to the addition of nucleotides at the junction of rearranged gene segments, and by the pairing of individual chains to form a TCR dimer. As T cells mature in the thymus, the repertoire of antigen-reactive T cells is modified by selection processes that eliminate many autoreactive T cells, enhance the proliferation of cells that function appropriately with self-MHC molecules and antigen, and allow T cells with nonproductive TCR rearrangements to die.

TCR-αβ cells do not recognize native protein or carbohydrate antigens. Instead, T cells recognize only short (~9–13 amino acids) peptide fragments derived from protein antigens taken up or produced in APCs. Foreign antigens may be taken up by endocytosis into acidified intracellular vesicles or by phagocytosis and degraded into small peptides that associate with MHC class II molecules (exogenous antigen-presentation pathway). Other foreign antigens arise endogenously in the cytosol (such as from replicating viruses) and are broken down into small peptides that associate with MHC class I molecules (endogenous antigen-presenting pathway). Thus, APCs proteolytically degrade foreign proteins and display peptide fragments embedded in the MHC class I or II antigen-recognition site on the MHC molecule surface, where foreign peptide fragments are available to bind to TCR-αβ or TCR-γδ chains of reactive T cells. CD4 molecules act as adhesives and, by direct binding to MHC class II (DR, DQ, or DP) molecules, stabilize the interaction of TCR with peptide antigen (Fig. 314-7). Similarly, CD8 molecules also act as adhesives to stabilize the TCR-antigen interaction by direct CD8 molecule binding to MHC class I (A, B, or C) molecules.

Antigens that arise in the cytosol and are processed via the endogenous antigen-presentation pathway are cleaved into small peptides by a complex of proteases called the *proteasome*. From the proteasome, antigen peptide fragments are transported from the cytosol into the lumen of the endoplasmic reticulum by a heterodimeric complex termed *transporters associated with antigen processing*, or TAP proteins. There, MHC class I molecules in the

endoplasmic reticulum membrane physically associate with processed cytosolic peptides. Following peptide association with class I molecules, peptide–class I complexes are exported to the Golgi apparatus, and then to the cell surface, for recognition by CD8+ T cells.

Antigens taken up from the extracellular space via endocytosis into intracellular acidified vesicles are degraded by vesicle proteases into peptide fragments. Intracellular vesicles containing MHC class II molecules fuse with peptide-containing vesicles, thus allowing peptide fragments to physically bind to MHC class II molecules. Peptide–MHC class II complexes are then transported to the cell surface for recognition by CD4+ T cells (Chap. 315).

Whereas it is generally agreed that the TCR-αβ receptor recognizes peptide antigens in the context of MHC class I or class II molecules, lipids in the cell wall of intracellular bacteria such as *M. tuberculosis* can also be presented to a wide variety of T cells, including subsets of TCR-γδ T cells, and a subset of CD8+ TCR-αβ T cells. Importantly, bacterial lipid antigens are not presented in the context of MHC class I or II molecules, but rather are presented in the context of MHC-related CD1 molecules. Some γδ T cells that recognize lipid antigens via CD1 molecules have very restricted TCR usage, do not need antigen priming to respond to bacterial lipids, and may actually be a form of innate rather than acquired immunity to intracellular bacteria.

Just as foreign antigens are degraded and their peptide fragments presented in the context of MHC class I or class II molecules on APCs, endogenous self-proteins also are degraded and self-peptide fragments are presented to T cells in the context of MHC class I or class II molecules on APCs. In peripheral lymphoid organs, there are T cells that are capable of recognizing self-protein fragments but normally are *anergic* or *tolerant*, i.e., nonresponsive to self-antigenic stimulation, due to lack of self-antigen upregulating APC *co-stimulatory molecules* such as B7-1 (CD80) and B7-2 (CD86) (see below).

Once engagement of mature T cell TCR by foreign peptide occurs in the context of self–MHC class I or class II molecules, binding of non-antigen-specific adhesion ligand pairs such as CD54-CD11/CD18 and CD58-CD2 stabilizes MHC peptide–TCR binding, and the expression of these adhesion molecules is upregulated (Fig. 314-7). Once antigen ligation of the TCR occurs, the T cell membrane is partitioned into *lipid membrane microdomains*, or *lipid rafts*, that coalesce the key signaling molecules TCR/CD3 complex, CD28, CD2, LAT (linker for activation of T cells), intracellular activated (dephosphorylated) src family protein tyrosine kinases (PTKs), and the key CD3ζ-associated protein-70 (ZAP-70) PTK (Fig. 314-7). Importantly, during T cell activation, the CD45 molecule, with protein tyrosine phosphatase activity is partitioned away from the TCR complex to allow activating phosphorylation events to occur. The coalescence of signaling molecules of activated T lymphocytes in *microdomains* has suggested that T cell–APC interactions can be considered *immunologic synapses*, analogous in function to neuronal synapses.

After TCR-MHC binding is stabilized, activation signals are transmitted through the cell to the nucleus and lead to the expression of gene products important in mediating the wide diversity of T cell functions such as the secretion of IL-2. The TCR does not have intrinsic signaling activity but is linked to a variety of signaling pathways via immunoreceptor tyrosine-based activation motifs (ITAMs) expressed on the various CD3 chains that bind to proteins that mediate signal transduction. Each of the pathways results in the activation of particular transcription factors that control the expression of cytokine and cytokine receptor genes. Thus, antigen-MHC binding to the TCR induces the activation of the src family of PTKs, fyn and lck (lck is associated with CD4 or CD8 co-stimulatory molecules); phosphorylation of CD3ζ chain; activation of the related tyrosine kinases ZAP-70 and syk; and downstream activation of the calcium-dependent calcineurin pathway, the ras pathway, and the protein kinase C pathway. Each of these pathways leads to activation of specific families of transcription factors (including NF-AT, fos and jun, and rel/NF-κB) that form heteromultimers capable of inducing expression of IL-2, IL-2 receptor, IL-4, TNF-α, and other T cell mediators.

In addition to the signals delivered to the T cell from the TCR complex and CD4 and CD8, molecules on the T cell such as CD28 and inducible co-stimulator (ICOS) and molecules on dendritic cells such as B7-1 (CD80) and B7-2 (CD86) also deliver important co-stimulatory signals that upregulate T cell cytokine production and are essential for T cell activation. If signaling through CD28 or ICOS does not occur, or if CD28 is blocked, the T cell becomes anergic rather than activated (see "Immune Tolerance and Autoimmunity" below).

T Cell Superantigens

Conventional antigens bind to MHC class I or II molecules in the groove of the αβ heterodimer and bind to T cells via the V regions of the TCR-α and -β chains. In contrast, superantigens bind directly to the lateral portion of the TCR-β chain and MHC class II β chain and stimulate T cells based solely on the Vβ gene segment utilized independent of the D, J, and Vα sequences present. *Superantigens* are protein molecules capable of activating up to 20% of the peripheral T cell pool, whereas conventional antigens activate <1 in 10,000 T cells. T cell superantigens include staphylococcal enterotoxins and other bacterial products. Superantigen stimulation of human peripheral T cells occurs in the clinical setting of *staphylococcal toxic shock syndrome*, leading to massive overproduction of T cell cytokines that leads to hypotension and shock (Chap. 135).

B Cells

Mature B cells constitute 10–15% of human peripheral blood lymphocytes, 20–30% of lymph node cells, 50% of splenic lymphocytes, and ~10% of bone marrow lymphocytes. B cells express on their surface intramembrane immunoglobulin (Ig) molecules that function as B cell receptors (BCRs) for antigen in a complex of Ig-associated α and β signaling molecules with properties similar to those described in T cells **(Fig. 314-8)**. Unlike T cells, which recognize only processed peptide fragments of conventional antigens embedded in the notches of MHC class I and class II antigens of APCs, B cells are capable of recognizing and proliferating to whole unprocessed native antigens via antigen binding to B cell–surface Ig (sIg) receptors. B cells also express surface receptors for the Fc region of IgG molecules (CD32) as well as receptors for activated complement components (C3d or CD21, C3b or CD35). The primary function of B cells is to produce antibodies. B cells also serve as APCs and are highly efficient at antigen processing. Their antigen-presenting function is enhanced by a variety of cytokines. Mature B cells are derived from bone marrow precursor cells that arise continuously throughout life (Fig. 314-6).

Figure 314-8

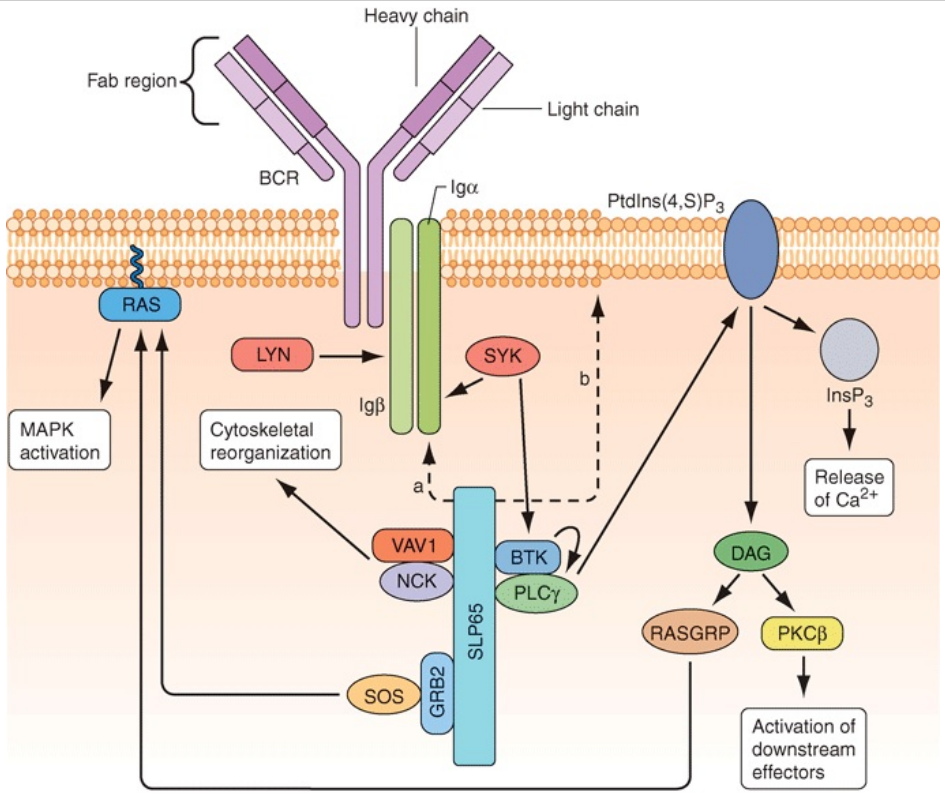

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B cell receptor (BCR) activation results in the sequential activation of protein tyrosine kinases, which results in the formation of a signaling complex and activation of downstream pathways as shown. Whereas SLP76 is recruited to the membrane through GADS and LAT, the mechanism of SLP65 recruitment is unclear. Studies have indicated two mechanisms: (a) direct binding by the SH2 domain of SLP65 to immunoglobulin (Ig) of the BCR complex or (b) membrane recruitment through a leucine zipper in the amino terminus of SLP65 and an unknown binding partner. ADAP, adhesion- and degranulation-promoting adaptor protein; AP1, activator protein 1; BTK, Bruton's tyrosine kinase; DAG, diacylglycerol; GRB2, growth factor receptor–bound protein 2; HPK1, hematopoietic progenitor kinase 1; $InsP_3$, inositol-1,4,5-trisphosphate; ITK, interleukin-2-inducible T cell kinase; NCK, noncatalytic region of tyrosine kinase; NF-B, nuclear factor B; PKC, protein kinase C; PLC, phospholipase C; $PtdIns(4,5)P_2$, phosphatidylinositol-4, 5-bisphosphate; RASGRP, RAS guanyl-releasing protein; SOS, son of sevenless homologue; SYK, spleen tyrosine kinase. *(Adapted from Koretzky et al; with permission from Macmillan Publishers Ltd. Copyright 2006.)*

B lymphocyte development can be separated into antigen- independent and antigen-dependent phases. Antigen-independent B cell development occurs in primary lymphoid organs and includes all stages of B cell maturation up to the sIg+ mature B cell. Antigen-dependent B cell maturation is driven by the interaction of antigen with the mature B cell sIg, leading to memory B cell induction, Ig class switching, and plasma cell formation. Antigen-dependent stages of B cell maturation occur in secondary lymphoid organs, including lymph node, spleen, and gut Peyer's patches. In contrast to the T cell repertoire that is generated intrathymically before contact with foreign antigen, the repertoire of B cells expressing diverse antigen-reactive sites is modified by further alteration of Ig genes after stimulation by antigen—a process called *somatic hypermutation*—which occurs in lymph node germinal centers.

During B cell development, diversity of the antigen-binding variable region of Ig is generated by an ordered set of Ig gene rearrangements that are similar to the rearrangements undergone by TCR α, β, γ, and δ genes. For the heavy chain, there is first a rearrangement of D segments to J segments, followed by a second rearrangement between a V gene segment and the newly formed D-J sequence; the C segment is aligned to the V-D-J complex to yield a functional Ig heavy chain gene (V-D-J-C). During later stages, a functional κ or γ light chain gene is generated by rearrangement of a V segment to a J segment, ultimately yielding an intact Ig molecule composed of heavy and light chains.

The process of Iggene rearrangement is regulated and results in a single antibody specificity produced by each B cell, with each Ig molecule comprising one type of heavy chain and one type of light chain. Although each B cell contains two copies of Ig light and heavy chain genes, only one gene of each type is productively rearranged and expressed in each B cell, a process termed *allelic exclusion*.

There are ~300 $V_κ$ genes and 5 $J_κ$ genes, resulting in the pairing of $V_κ$ and $J_κ$ genes to create >1500 different light chain combinations. The number of distinct κ light chains that can be generated is increased by somatic mutations within the $V_κ$ and $J_κ$ genes, thus creating large numbers of possible specificities from a limited amount of germ-line genetic information. As noted above, in heavy chain Iggene rearrangement, the VH domain is created by the joining of three types of germ-line genes called V_H, D_H, and J_H, thus allowing for even greater diversity in the variable region of heavy chains than of light chains.

The most immature B cell precursors (early pro-B cells) lack cytoplasmic Ig (cIg) and sIg (Fig. 314-6). The large pre-B cell is marked by the acquisition of the surface pre-BCR composed of μ heavy (H) chains and a pre-B light chain,

termed ψLC. ψLC is a surrogate light chain receptor encoded by the nonrearranged V pre-B and the γ5 light chain locus (the pre-BCR). Pro- and pre-B cells are driven to proliferate and mature by signals from bone marrow stroma —in particular, IL-7. Light chain rearrangement occurs in the small pre-B cell stage such that the full BCR is expressed at the immature B cell stage. Immature B cells have rearranged Ig light chain genes and express sIgM. As immature B cells develop into mature B cells, sIgD is expressed as well as sIgM. At this point, B lineage development in bone marrow is complete, and B cells exit into the peripheral circulation and migrate to secondary lymphoid organs to encounter specific antigens.

Random rearrangements of Iggenes occasionally generate self-reactive antibodies, and mechanisms must be in place to correct these mistakes. One such mechanism is BCR editing, whereby autoreactive BCRs are mutated to not react with self-antigens. If receptor editing is unsuccessful in eliminating autoreactive B cells, then autoreactive B cells undergo negative selection in the bone marrow through induction of apoptosis after BCR engagement of self-antigen.

After leaving the bone marrow, B cells populate peripheral B cell sites, such as lymph node and spleen, and await contact with foreign antigens that react with each B cell's clonotypic receptor. Antigen-driven B cell activation occurs through the BCR, and a process known as *somatic hypermutation* takes place whereby point mutations in rearranged H- and L-genes give rise to mutant sIg molecules, some of which bind antigen better than the original sIg molecules. Somatic hypermutation, therefore, is a process whereby memory B cells in peripheral lymph organs have the best binding, or the highest-affinity antibodies. This overall process of generating the best antibodies is called *affinity maturation of antibody* .

Lymphocytes that synthesize IgG, IgA, and IgE are derived from sIgM+, sIgD+ mature B cells. Ig class switching occurs in lymph node and other peripheral lymphoid tissue germinal centers. CD40 on B cells and CD40 ligand on T cells constitute a critical co- stimulatory receptor-ligand pair of immune-stimulatory molecules. Pairs of CD40+ B cells and CD40 ligand+ T cells bind and drive B cell Ig class switching via T cell–produced cytokines such as IL-4 and TGF-β. IL-1, -2, -4, -5, and -6 synergize to drive mature B cells to proliferate and differentiate into Ig-secreting cells.

HUMORAL MEDIATORS OF ADAPTIVE IMMUNITY: IMMUNOGLOBULINS

Immunoglobulins are the products of differentiated B cells and mediate the humoral arm of the immune response. The primary functions of antibodies are to bind specifically to antigen and bring about the inactivation or removal of the offending toxin, microbe, parasite, or other foreign substance from the body. The structural basis of Ig molecule function and Ig gene organization has provided insight into the role of antibodies in normal protective immunity, pathologic immune-mediated damage by immune complexes, and autoantibody formation against host determinants.

All immunoglobulins have the basic structure of two heavy and two light chains (Fig. 314-8). Immunoglobulin isotype (i.e., G, M, A, D, E) is determined by the type of Ig heavy chain present. IgG and IgA isotypes can be divided further into subclasses (G1, G2, G3, G4, and A1, A2) based on specific antigenic determinants on Ig heavy chains. The characteristics of human immunoglobulins are outlined in **Table 314-13**. The four chains are covalently linked by disulfide bonds. Each chain is made up of a V region and C regions (also called *domains*), themselves made up of units of ~110 amino acids. Light chains have one variable (V_L) and one constant (C_L) unit; heavy chains have one variable unit (V_H) and three or four constant (C_H) units, depending on isotype. As the name suggests, the constant, or C, regions of Ig molecules are made up of homologous sequences and share the same primary structure as all other Ig chains of the same isotype and subclass. Constant regions are involved in biologic functions of Ig molecules. The C_H2 domain of IgG and the C_H4 units of IgM are involved with the binding of the C1q portion of C1 during complement activation. The C_H region at the carboxy-terminal end of the IgG molecule, the Fc region, binds to surface Fc receptors (CD16, CD32, CD64) of macrophages, dendritic cells, NK cells, B cells, neutrophils, and eosinophils.

Table 314–13. Physical, Chemical, and Biologic Properties of Human Immunoglobulins

Property	IgG	IgA	IgM	IgD	IgE
Usual molecular form	Monomer	Monomer, dimer	Pentamer, hexamer	Monomer	Monomer
Other chains	None	J chain, SC	J chain	None	None
Subclasses	G1, G2, G3, G4	A1, A2	None	None	None
Heavy chain allotypes	Gm (=30)	No A1, A2m (2)	None	None	None
Molecular mass, kDa	150	160, 400	950, 1150	175	190
Serum level in average adult, mg/mL	9.5–12.5	1.5–2.6	0.7–1.7	0.04	0.0003
Percentage of total serum Ig	75–85	7–15	5–10	0.3	0.019
Serum half-life, days	23	6	5	3	2.5
Synthesis rate, mg/kg per day	33	65	7	0.4	0.016
Antibody valence	2	2,4	10, 12	2	2
Classical complement activation	+(G1, 2?, 3)	–	++	–	–
Alternate	+(G4)	+	–	+	–

complement activation					
Binding cells via Fc	Macrophages, neutrophils, large granular lymphocytes	Lymphocytes	Lymphocytes	None	Mast cells, basophils, B cells
Biologic properties	Placental transfer, secondary Ab for most antipathogen responses	Secretory immunoglobulin	Primary Ab responses	Marker for mature B cells	Allergy, antiparasite responses

Source: After L Carayannopoulos, JD Capra, in WE Paul (ed): *Fundamental Immunology*, 3rd ed. New York, Raven, 1993; with permission.

Variable regions (V_L and V_H) constitute the antibody-binding (Fab) region of the molecule. Within the V_L and V_H regions are hypervariable regions (extreme sequence variability) that constitute the antigen-binding site unique to each Ig molecule. The idiotype is defined as the specific region of the Fab portion of the Ig molecule to which antigen binds. Antibodies against the idiotype portion of an antibody molecule are called *anti-idiotype antibodies*. The formation of such antibodies in vivo during a normal B cell antibody response may generate a negative (or "off") signal to B cells to terminate antibody production.

IgG constitutes ~75–85% of total serum immunoglobulin. The four IgG subclasses are numbered in order of their level in serum, IgG1 being found in greatest amounts and IgG4 the least. IgG subclasses have clinical relevance in their varying ability to bind macrophage and neutrophilFc receptors and to activate complement (Table 314-13). Moreover, selective deficiencies of certain IgG subclasses give rise to clinical syndromes in which the patient is inordinately susceptible to bacterial infections. IgG antibodies are frequently the predominant antibody made after rechallenge of the host with antigen (secondary antibody response).

IgM antibodies normally circulate as a 950-kDa pentamer with 160-kDa bivalent monomers joined by a molecule called the *J chain*, a 15-kDa nonimmunoglobulin molecule that also effects polymerization of IgA molecules. IgM is the first immunoglobulin to appear in the immune response (primary antibody response) and is the initial type of antibody made by neonates. Membrane IgM in the monomeric form also functions as a major antigen receptor on the surface of mature B cells (Fig. 314-13). IgM is an important component of immune complexes in autoimmune diseases. For example, IgM antibodies against IgG molecules (rheumatoid factors) are present in high titers in *rheumatoid arthritis*, other collagen diseases, and some infectious diseases (*subacute bacterial endocarditis*).

IgA constitutes only 7–15% of total serum immunoglobulin but is the predominant class of immunoglobulin in secretions. IgA in secretions (tears, saliva, nasal secretions, gastrointestinal tract fluid, and human milk) is in the form of secretory IgA (sIgA), a polymer consisting of two IgA monomers, a joining molecule, again called the J chain, and a glycoprotein called the *secretory protein*. Of the two IgA subclasses, IgA1 is primarily found in serum, whereas IgA2 is more prevalent in secretions. IgA fixes complement via the alternative complement pathway and has potent antiviral activity in humans by prevention of virus binding to respiratory and gastrointestinal epithelial cells.

IgD is found in minute quantities in serum and, together with IgM, is a major receptor for antigen on the B cell surface. IgE, which is present in serum in very low concentrations, is the major class of immunoglobulin involved in arming mast cells and basophils by binding to these cells via the Fc region. Antigen cross-linking of IgE molecules on basophil and mast cell surfaces results in release of mediators of the immediate hypersensitivity (allergic) response (Table 314-13).

Cellular Interactions in Regulation of Normal Immune Responses

The net result of activation of the humoral (B cell) and cellular (T cell) arms of the adaptive immune system by foreign antigen is the elimination of antigen directly by specific effector T cells or in concert with specific antibody. Figure 314-2 is a simplified schematic diagram of the T and B cell responses indicating some of these cellular interactions.

The expression of adaptive immune cell function is the result of a complex series of immunoregulatory events that occur in phases. Both T and B lymphocytes mediate immune functions, and each of these cell types, when given appropriate signals, passes through stages, from activation and induction through proliferation, differentiation, and ultimately effector functions. The effector function expressed may be at the end point of a response, such as secretion of antibody by a differentiated plasma cell, or it might serve a regulatory function that modulates other functions, such as is seen with CD4+ and CD8+ T lymphocytes that modulate both differentiation of B cells and activation of CD8+ cytotoxic T cells.

CD4 helper T cells can be subdivided on the basis of cytokines produced (Fig. 314-2). Activated T_H1-type helper T cells secrete IL-2, IFN-γ, IL-3, TNF-α, GM-CSF, and TNF-β, while activated T_H2-type helper T cells secrete IL-3, -4, -5, -6, -10, and -13. T_H1 CD4+ T cells, through elaboration of IFN-γ, have a central role in mediating intracellular killing by a variety of pathogens. T_H1 CD4+ T cells also provide T cell help for generation of cytotoxic T cells and some types of opsonizing antibody, and they generally respond to antigens that lead to delayed hypersensitivity types of immune responses for many intracellular viruses and bacteria (such as HIV or *M. tuberculosis*). In contrast, T_H2 cells have a primary role in regulatory humoral immunity and isotype switching. T_H2 cells, through production of IL-4 and IL-10, have a regulatory role in limiting proinflammatory responses mediated by T_H1 cells (Fig. 314-2). In addition, T_H2 CD4+ T cells provide help to B cells for specific Ig production and respond to antigens that require high antibody levels for foreign antigen elimination (extracellular encapsulated bacteria such as *Streptococcus pneumoniae* and certain parasite infections). More recently, a new subset of the T_H family has been described termed T_h17 characterized by these cells to secrete cytokines such as IL-17, -22, and -26. T_h17 cells have been shown to play a role in autoimmune inflammatory disorders in addition to defense against extracellular bacteria and fungi, particularly at mucosal surfaces (Fig. 314-3). In summary, the type of T cell response generated in an immune response is determined by the microbe PAMPs presented to the DCs, the TLRs on the DCs that become activated, the types of DCs that are activated, and the cytokines that are produced (Table 308-4). Commonly, myeloid DCs

produce IL-12 and activate T_H1 T cell responses that result in IFN-γ and cytotoxic T cell induction, and plasmacytoid DCs produce IFN-α and lead to T_H2 responses that result in IL-4 production and enhanced antibody responses.

As shown in Figs. 314-2 and 314-3, upon activation by DCs, T cell subsets that produce IL-2, IL-3, IFN-γ, and/or IL-4, -5, -6, -10, and -13 are generated and exert positive and negative influences on effector T and B cells. For B cells, trophic effects are mediated by a variety of cytokines, particularly T cell–derived IL-3, -4, -5, and -6, that act at sequential stages of B cell maturation, resulting in B cell proliferation, differentiation, and ultimately antibody secretion. For cytotoxic T cells, trophic factors include inducer T cell secretion of IL-2, IFN-γ, and IL-12.

An important type of immunomodulatory T cell that controls immune responses is *CD4+ and CD8+ T regulatory cells*. These cells constitutively express the α chain of the IL-2 receptor (CD25), produce large amounts of IL-10, and can suppress both T and B cell responses. T regulatory cells are induced by immature dendritic cells and play key roles in maintaining tolerance to self-antigens in the periphery. Loss of T regulatory cells is the cause of organ-specific autoimmune disease in mice such as autoimmune thyroiditis, adrenalitis, and oophoritis (see "Immune Tolerance and Autoimmunity" below). T regulatory cells also play key roles in controlling the magnitude and duration of immune responses to microbes. Normally, after the initial immune response to a microbe has eliminated the invader, T regulatory cells are activated to suppress the antimicrobe response and prevent host injury. Some microbes have adapted to induce T regulatory cell activation at the site of infection to promote parasite infection and survival. In *Leishmania* infection, the parasite locally induces T regulatory cell accumulation at skin infection sites that dampens anti-*Leishmania* T cell responses and prevents elimination of the parasite. It is thought that many chronic infections such as by *M. tuberculosis* are associated with abnormal T regulatory cell activation that prevents elimination of the microbe.

Although B cells recognize native antigen via B cell–surface Ig receptors, B cells require T cell help to produce high-affinity antibody of multiple isotypes that are the most effective in eliminating foreign antigen. This T cell dependence likely functions in the regulation of B cell responses and in protection against excessive autoantibody production. T cell–B cell interactions that lead to high-affinity antibody production require (1) processing of native antigen by B cells and expression of peptide fragments on the B cell surface for presentation to T_H cells, (2) the ligation of B cells by both the TCR complex and the CD40 ligand, (3) induction of the process termed *antibody isotype switching* in antigen-specific B cell clones, and (4) induction of the process of affinity maturation of antibody in the germinal centers of B cell follicles of lymph node and spleen.

Naïve B cells express cell-surface IgD and IgM, and initial contact of naïve B cells with antigen is via binding of native antigen to B cell–surface IgM. T cell cytokines, released following T_H2 cell contact with B cells or by a "bystander" effect, induce changes in Iggene conformation that promote recombination of Ig genes. These events then result in the "switching" of expression of heavy chain exons in a triggered B cell, leading to the secretion of IgG, IgA, or, in some cases, IgE antibody with the same V region antigen specificity as the original IgM antibody, for response to a wide variety of extracellular bacteria, protozoa, and helminths. CD40 ligand expression by activated T cells is critical for induction of B cell antibody isotype switching and for B cell responsiveness to cytokines. Patients with mutations in T cell CD40 ligand have B cells that are unable to undergo isotype switching, resulting in lack of memory B cell generation and the immunodeficiency syndrome of *X-linked hyper-IgM syndrome* (Chap. 316).

Immune Tolerance and Autoimmunity

Immune tolerance is defined as the absence of activation of pathogenic autoreactivity. *Autoimmune diseases* are syndromes caused by the activation of T or B cells or both, with no evidence of other causes such as infections or malignancies (Chap. 318). Once thought to be mutually exclusive, immune tolerance and autoimmunity are now both recognized to be present normally in health; when abnormal, they represent extremes from the normal state. For example, it is now known that low levels of autoreactivity of T and B cells with self-antigens in the periphery are critical to their survival. Similarly, low levels of autoreactivity and thymocyte recognition of self-antigens in the thymus are the mechanisms whereby (1) normal T cells are positively selected to survive and leave the thymus to respond to foreign microbes in the periphery and (2) T cells highly reactive to self-antigens are negatively selected and die to prevent overly self-reactive T cells from getting into the periphery (central tolerance). However, not all self-antigens are expressed in the thymus to delete highly self-reactive T cells, and there are mechanisms for peripheral tolerance induction of T cells as well. Unlike the presentation of microbial antigens by mature dendritic cells, the presentation of self-antigens by immature dendritic cells neither activates nor matures the dendritic cells to express high levels of co-stimulatory molecules such as B7-1 (CD80) or B7-2 (CD86). When peripheral T cells are stimulated by dendritic cells expressing self-antigens in the context of HLA molecules, sufficient stimulation of T cells occurs to keep them alive, but otherwise they remain anergic, or nonresponsive, until they contact a dendritic cell with high levels of co-stimulatory molecules expressing microbial antigens. In the latter setting, normal T cells then become activated to respond to the microbe. If B cells have high-self-reactivity BCRs, they normally undergo either deletion in the bone marrow or receptor editing to express a less autoreactive receptor. Although many autoimmune diseases are characterized by abnormal or pathogenic autoantibody production **(Table 314-14)**, most autoimmune diseases are caused by a combination of excess T and B cell reactivity.

Table 314–14. Recombinant or Purified Autoantigens Recognized by Autoantibodies Associated with Human Autoimmune Disorders

Autoantigen	Autoimmune Diseases	Autoantigen	Autoimmune Diseases
Cell- or Organ-Specific Autoimmunity			
Acetylcholine receptor	Myasthenia gravis	Insulin receptor	Type B insulin resistance, acanthosis, systemic lupus erythematosus (SLE)
Actin	Chronic active hepatitis, primary biliary cirrhosis	Intrinsic factor type 1	Pernicious anemia
Adenine nucleotide	Dilated	Leukocyte	Treatment-resistant Lyme arthritis

translator (ANT)	cardiomyopathy, myocarditis	function-associated antigen (LFA-1)	
β-Adrenoreceptor	Dilated cardiomyopathy		
Aromatic L-amino acid decarboxylase	Autoimmune polyendocrine syndrome type 1 (APS-1)	Myelin-associated glycoprotein (MAG)	Polyneuropathy
Asialoglycoprotein receptor	Autoimmune hepatitis	Myelin-basic protein	Multiple sclerosis, demyelinating diseases
Bactericidal/permeability-increasing protein (Bpi)	Cystic fibrosis vasculitides	Myelin oligodendrocyte glycoprotein (MOG)	Multiple sclerosis
Calcium-sensing receptor	Acquired hypoparathyroidism	Myosin	Rheumatic fever
Cholesterol side-chain cleavage enzyme (CYPIIa)	Autoimmune polyglandular syndrome-1	p-80-Collin	Atopic dermatitis
Collagen type IV-α3-chain	Goodpasture syndrome	Pyruvate dehydrogenase complex-E2 (PDC-E2)	Primary biliary cirrhosis
Cytochrome P450 2D6 (CYP2D6)	Autoimmune hepatitis		
Desmin	Crohn's disease, coronary artery disease	Sodium iodide symporter (NIS)	Graves' disease, autoimmune hypothyroidism
Desmoglein 1	Pemphigus foliaceus		
Desmoglein 3	Pemphigus vulgaris	SOX-10	Vitiligo
F-actin	Autoimmune hepatitis	Thyroid and eye muscle shared protein	Thyroid-associated ophthalmopathy
GM gangliosides	Guillain-Barré syndrome		
Glutamate decarboxylase (GAD65)	Type 1 diabetes, stiff-person syndrome	Thyroglobulin	Autoimmune thyroiditis
Glutamate receptor (GLUR)	Rasmussen encephalitis	Thyroid peroxidase	Autoimmune Hashimoto thyroiditis
H/K ATPase	Autoimmune gastritis	Thyrotropin receptor	Graves' disease
17-α-Hydroxylase (CYP17)	Autoimmune polyglandular syndrome-1	Tissue transglutaminase	Celiac disease
21-Hydroxylase (CYP21)	Addison disease	Transcription coactivator p75	Atopic dermatitis
IA-2 (ICA512)	Type 1 diabetes	Tryptophan hydroxylase	Autoimmune polyglandular syndrome-1
Insulin	Type 1 diabetes, insulin hypoglycemic syndrome (Hirata disease)	Tyrosinase Tyrosine hydroxylase	Vitiligo, metastatic melanoma Autoimmune polyglandular syndrome-1
Systemic Autoimmunity			
ACTH	ACTH deficiency	Histone H2A-H2B-DNA	SLE
Aminoacyl-tRAN histidyl synthetase	Myositis, dermatomyositis	IgE receptor	Chronic idiopathic urticaria
Aminoacyl-tRNA synthetase (several)	Polymyositis, dermatomyositis	Keratin	RA
Cardiolipin	SLE, anti-phospholipid syndrome	Ku-DNA-protein kinase	SLE
Carbonic anhydrase II	SLE, Sjögren's syndrome, systemic sclerosis	Ku-nucleoprotein	Connective tissue syndrome
		La phosphoprotein (La 55-B)	Sjögren's syndrome
Collagen (multiple types)	Rheumatoid arthritis (RA), SLE, progressive systemic sclerosis	Myeloperoxidase	Necrotizing and crescentic glomerulonephritis (NCGN), systemic vasculitis
Centromere-associated proteins	Systemic sclerosis	Proteinase 3 (PR3)	Granulomatosis with polyangiitis (Wegener's), Churg-Strauss syndrome
DNA-dependent nucleoside-stimulated ATPase	Dermatomyositis	RNA polymerase I–III (RNP)	Systemic sclerosis, SLE

Fibrillarin	Scleroderma	Signal recognition protein (SRP54)	Polymyositis
Fibronectin	SLE, RA, morphea	Topoisomerase-1 (Scl-70)	Scleroderma, Raynaud syndrome
Glucose-6-phosphate isomerase β2-Glycoprotein I (B2-GPI)	RA Primary antiphospholipid syndrome	Tublin	Chronic liver disease, visceral leishmaniasis
Golgin (95, 97, 160, 180)	Sjögren's syndrome, SLE, RA	Vimentin	Systemic autoimmune disease
Heat shock protein	Various immune-related disorders		
Hemidesmosomal protein 180	Bullous pemphigoid, herpes gestationis, cicatricial pemphigoid		
Plasma Protein and Cytokine Autoimmunity			
C1 inhibitor	Autoimmune C1 deficiency	Glycoprotein IIb/IIIg and Ib/IX	Autoimmune thrombocytopenia purpura
C1q	SLE, membrane proliferative glomerulonephritis (MPGN)	IgA	Immunodeficiency associated with SLE, pernicious anemia, thyroiditis, Sjögren's syndrome and chronic active hepatitis
Cytokines (IL-1α, IL-1β, IL-6, IL-10, LIF)	RA, systemic sclerosis, normal subjects		
Factor II, factor V, factor VII, factor VIII, factor IX, factor X, factor XI, thrombin vWF	Prolonged coagulation time	Oxidized LDL (OxLDL)	Atherosclerosis
Cancer and Paraneoplastic Autoimmunity			
Amphiphysin	Neuropathy, small cell lung cancer	p62 (IGF-II mRNA-binding protein)	Hepatocellular carcinoma (China)
Cyclin B1	Hepatocellular carcinoma	Recoverin	Cancer-associated retinopathy
DNA topoisomerase II	Liver cancer	Ri protein	Paraneoplastic opsoclonus myoclonus ataxia
Desmoplakin	Paraneoplastic pemphigus		
Gephyrin	Paraneoplastic stiff-person syndrome	βIV spectrin	Lower motor neuron syndrome
Hu proteins Neuronal nicotinic acetylcholine receptor	Paraneoplastic encephalomyelitis Subacute autonomic neuropathy, cancer	Synaptotagmin Voltage-gated calcium channels	Lambert-Eaton myasthenic syndrome Lambert-Eaton myasthenic syndrome
p53	Cancer, SLE	Yo protein	Paraneoplastic cerebellar degeneration

Source: From A Lernmark et al: J Clin Invest 108:1091, 2001; with permission.

Multiple factors contribute to the genesis of clinical autoimmune disease syndromes, including genetic susceptibility (Table 314-14), environmental immune stimulants such as drugs [e.g., procainamide and phenytoin (Dilantin) with drug-induced systemic lupus erythematosus], infectious agent triggers (such as Epstein-Barr virus and autoantibody production against red blood cells and platelets), and loss of T regulatory cells (leading to thyroiditis, adrenalitis, and oophoritis).

IMMUNITY AT MUCOSAL SURFACES

Mucosa covering the respiratory, digestive, and urogenital tracts; the eye conjunctiva; the inner ear; and the ducts of all exocrine glands contain cells of the innate and adaptive mucosal immune system that protect these surfaces against pathogens. In the healthy adult, mucosa-associated lymphoid tissue (MALT) contains 80% of all immune cells within the body and constitutes the largest mammalian lymphoid organ system.

MALT has three main functions: (1) to protect the mucous membranes from invasive pathogens; (2) to prevent uptake of foreign antigens from food, commensal organisms, and airborne pathogens and particulate matter; and (3) to prevent pathologic immune responses from foreign antigens if they do cross the mucosal barriers of the body (Fig. 314-9).

Figure 314-9

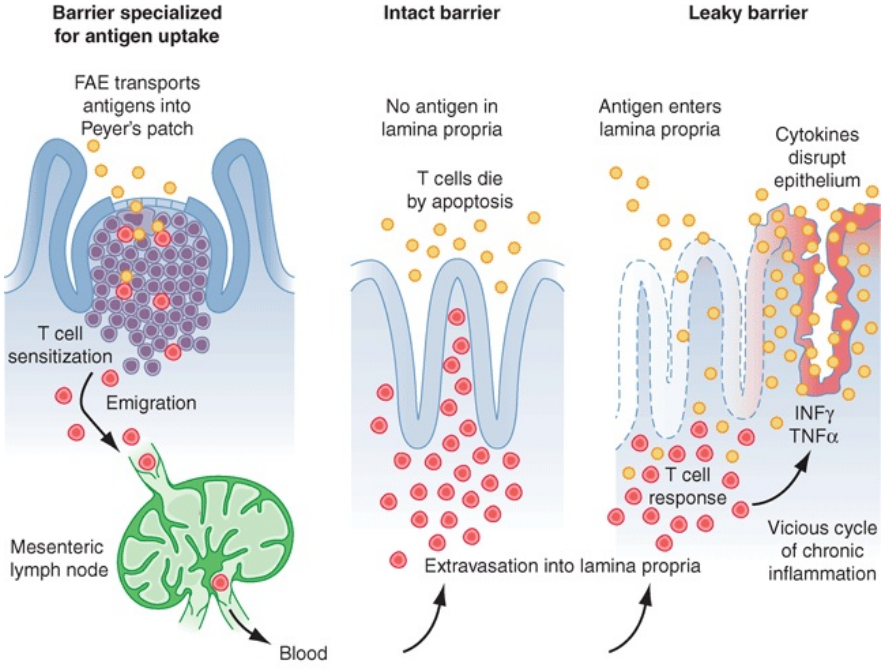

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Increased epithelial permeability may be important in the development of chronic gut T cell–mediated inflammation. CD4 T cells activated by gut antigens in Peyer's patches migrate to the LP. In healthy individuals, these cells die by apoptosis. Increased epithelial permeability may allow sufficient antigen to enter the LP to trigger T cell activation, breaking tolerance mediated by immunosuppressive cytokines and perhaps T regulatory cells. Proinflammatory cytokines then further increase epithelial permeability, setting up a vicious cycle of chronic inflammation. *(From MacDonald and Monteleone; with permission.*)

MALT is a compartmentalized system of immune cells that functions independently from systemic immune organs. Whereas the systemic immune organs are essentially sterile under normal conditions and respond vigorously to pathogens, MALT immune cells are continuously bathed in foreign proteins and commensal bacteria, and they must select those pathogenic antigens that must be eliminated. MALT contains anatomically defined foci of immune cells in the intestine, tonsil, appendix, and peribronchial areas that are inductive sites for mucosal immune responses. From these sites, immune T and B cells migrate to effector sites in mucosal parenchyma and exocrine glands where mucosal immune cells eliminate pathogen-infected cells. In addition to mucosal immune responses, all mucosal sites have strong mechanical and chemical barriers and cleansing functions to repel pathogens.

Key components of MALT include specialized epithelial cells called "membrane" or "M" cells that take up antigens and deliver them to dendritic cells or other APCs. Effector cells in MALT include B cells producing antipathogen neutralizing antibodies of secretory IgA as well as IgG isotype, T cells producing similar cytokines as in systemic immune system response, and T helper and cytotoxic T cells that respond to pathogen-infected cells.

Secretory IgA is produced in amounts of >50 mg/kg of body weight per 24 h and functions to inhibit bacterial adhesion, inhibit macromolecule absorption in the gut, neutralize viruses, and enhance antigen elimination in tissue through binding to IgA and receptor-mediated transport of immune complexes through epithelial cells.

Recent studies have demonstrated the importance of commensal gut and other mucosal bacteria to the health of the human immune system. Normal commensal flora induces anti-inflammatory events in the gut and protects epithelial cells from pathogens through TLRs and other PRR signaling. When the gut is depleted of normal commensal flora, the immune system becomes abnormal, with loss of T_H1 T cell function. Restoration of the normal gut flora can reestablish the balance in T helper cell ratios characteristic of the normal immune system. When the gut barrier is intact, either antigens do not transverse the gut epithelium or, when pathogens are present, a self-limited, protective MALT immune response eliminates the pathogen **(Fig. 314-9)**. However, when the gut barrier breaks down, immune responses to commensal flora antigens can cause inflammatory bowel diseases such as *Crohn's disease* and, perhaps, *ulcerative colitis* (Fig. 314-9) (Chap. 295). Uncontrolled MALT immune responses to food antigens, such as gluten, can cause *celiac disease* (Chap. 295).

The Cellular and Molecular Control of Programmed Cell Death

The process of apoptosis (programmed cell death) plays a crucial role in regulating normal immune responses to antigen. In general, a wide variety of stimuli trigger one of several apoptotic pathways to eliminate microbe-infected cells, eliminate cells with damaged DNA, or eliminate activated immune cells that are no longer needed **(Fig. 314-10)**. The largest known family of "death receptors" is the tumor necrosis factor receptor (TNF-R) family [TNF-R1, TNF-R2, Fas (CD95), death receptor 3 (DR3), death receptor 4 [DR4, TNF-related apoptosis-including ligand receptor 1 (TRAIL-R1)], and death receptor 5 (DR5, TRAIL-R2)]; their ligands are all in the TNF-α family. Binding of ligands to these death receptors leads to a signaling cascade that involves activation of the *caspase* family of molecules that leads to DNA cleavage and cell death. Two other pathways of programmed cell death involve nuclear *p53* in the elimination of cells with abnormal DNA and *mitochondrial cytochrome c* to induce cell death in damaged cells (Fig. 314-10). A number of human diseases have now been described that result from, or are associated with, mutated apoptosis genes **(Table 314-15)**. These include mutations in the Fas and Fas ligand genes in autoimmune and lymphoproliferation syndromes, and multiple associations of mutations in genes in the apoptotic pathway with malignant syndromes.

Figure 314-10

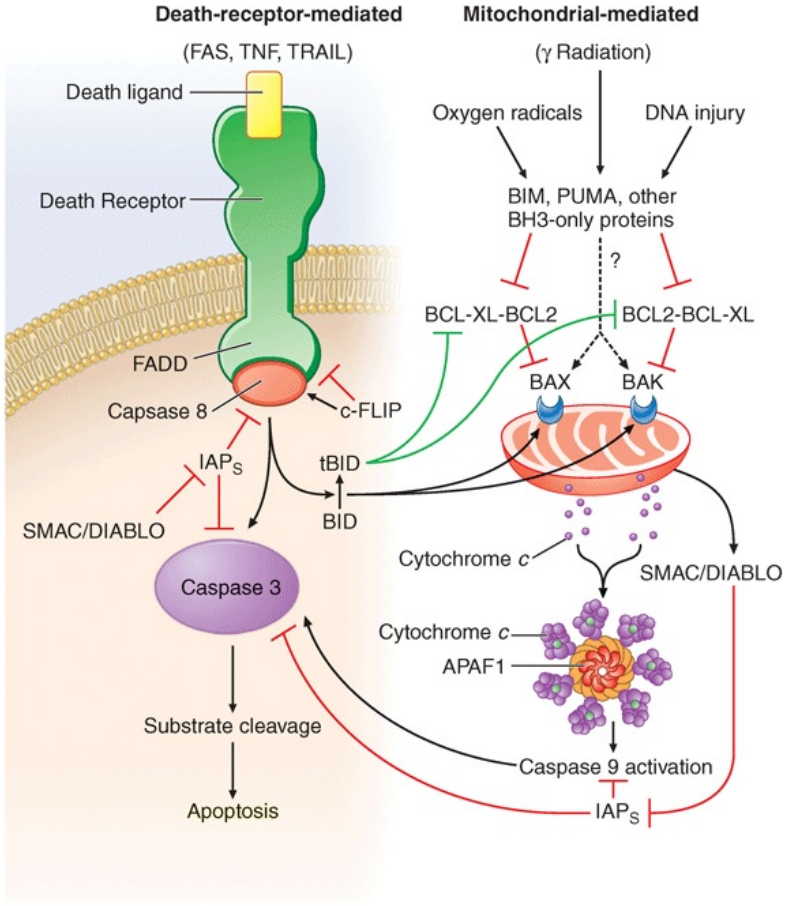

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathways of Cellular Apoptosis. There are two major pathways of apoptosis: the death-receptor pathway, which is mediated by activation of death receptors, and the BCL2-regulated mitochondrial pathway, which is mediated by noxious stimuli that ultimately lead to mitochondrial injury. Ligation of death receptors recruits the adaptor protein FAS-associated death domain (FADD). FADD in turn recruits caspase 8, which ultimately activates caspase 3, the key "executioner" caspase. Cellular FLICE-inhibitory protein (c-FLIP) can either inhibit or potentiate binding of FADD and caspase 8, depending on its concentration. In the intrinsic pathway, proapoptotic BH3 proteins are activated by noxious stimuli, which interact with and inhibit antiapoptotic BCL2 or BCL-XL. Thus, BAX and BAK are free to induce mitochondrial permeabilization with release of cytochrome *c*, which ultimately results in the activation of caspase 9 through the apoptosome. Caspase 9 then activates caspase 3. SMAC/DIABLO is also released after mitochondrial permeabilization and acts to block the action of inhibitors of apoptosis protein (IAPs), which inhibit caspase activation. There is potential cross-talk between the two pathways, which is mediated by the truncated form of BID (tBID) that is produced by caspase 8–mediated BID cleavage; tBID acts to inhibit the BCL2-BCL-XL pathway and to activate BAX and BAK. There is debate (indicated by the question mark) as to whether proapoptotic BH3 molecules (e.g., BIM and PUMA) act directly on BAX and BAK to induce mitochondrial permeability or whether they act only on BCL2-BCL-XL. APAF1, apoptotic protease-activating factor 1; BH3, BCL homologue; TNF, tumor necrosis factor; TRAIL, TNF-related apoptosis-inducing ligand. *(From Hotchkiss et al; with permission.)*

Table 314–15. Immune System Molecule Defects in Animals or Humans that Cause Autoimmune or Malignant Syndromes

Protein	Defect	Disease or Syndrome	Observation in Animal Models or Humans
Cytokines and Signaling Proteins			
Tumor necrosis factor (TNF) α	Overexpression	Inflammatory bowel disease (IBD), arthritis, vasculitis	Mice
TNF-α	Underexpression	Systemic lupus erythematosus (SLE)	Mice
Interleukin-1-receptor antagonist	Underexpression	Arthritis	Mice
IL-2	Overexpression	IBD	Mice
IL-7	Overexpression	IBD	Mice
IL-10	Overexpression	IBD	Mice
IL-2 receptor	Overexpression	IBD	Mice
IL-10 receptor	Overexpression	IBD	Mice

IL-3	Overexpression	Demyelinating syndrome	Mice
Interferon-δ	Overexpression in skin	SLE	Mice
STAT-3	Underexpression	IBD	Mice
STAT-4	Overexpression	IBD	Mice
Transforming growth factor (TGF) β	Underexpression	Systemic wasting syndrome and IBD	Mice
TGF-β receptor in T cells	Underexpression	SLE	Mice
Programmed death (PD-1)	Underexpression	SLE-like syndrome	Mice
Cytotoxic T lymphocyte, antigen-4 (CTLA-4)	Underexpression	Systemic lymphoproliferative disease	Mice
IL-10	Underexpression	IBD (mouse) Type 1 diabetes, thyroid disease, primary (human)	Mice and humans
Major Histocompatibility Locus Molecules*			
HLA B27	Allele expression or overexpression	Inflammatory bowel disease	Rats and humans
Complement deficiency of C1, 2, 3 or 4	Underexpression		Humans
LIGHT (TNF superfamily 14)	Overexpression	Systemic lymphoproliferative (mouse) and autoimmunity	Mice
HLA class II DQB10301, DQB10302	Allele expression	Juvenile-onset diabetes	Humans
HLA class II DQB10401, DQB10402	Allele expression	Rheumatoid arthritis	Humans
HLA class I B27	Allele expression	Ankylosing spondylitis, IBD	Rats and humans
Apoptosis Proteins			
TNF receptor 1 (TNF-R1)	Underexpression	Familial periodic fever syndrome	Humans
Fas (CD95; Apo-1)	Underexpression	Autoimmune lymphoproliferative syndrome type 1 (ALPS 1); malignant lymphoma; bladder cancer	Humans
Fas ligand	Underexpression	SLE (only one case identified)	Humans
Perforin	Underexpression	Familial hemophagocytic lymphohistiocytosis (FHL)	Humans
Caspase 10	Underexpression	Autoimmune lymphoproliferative syndrome type II (ALPS II)	Humans
bcl-10	Underexpression	Non-Hodgkin's lymphoma	Humans
P53	Underexpression	Various malignant neoplasms	Humans
Bax	Underexpression	Colon cancer; hematopoietic malignancies	Humans
bcl-2	Underexpression	Non-Hodgkin's lymphoma	Humans
c-IAP2	Underexpression	Low-grade MALT lymphoma	Humans
NAIP1	Underexpression	Spinal muscular atrophy	Humans

*Many autoimmune diseases are associated with a myriad of major compatibility complex gene allele (HLA) types. They are presented here as examples.

Abbreviation: MALT, mucosa-associated lymphoid tissue.

Source: Adapted from Mullauer and from Davidson and Diamond; with permission.

Mechanisms of Immune-Mediated Damage to Microbes or Host Tissues

Several responses by the host innate and adaptive immune systems to foreign microbes culminate in rapid and efficient elimination of microbes. In these scenarios, the classic weapons of the adaptive immune system (T cells, B cells) interface with cells (macrophages, dendritic cells, NK cells, neutrophils, eosinophils, basophils) and soluble products (microbial peptides, pentraxins, complement and coagulation systems) of the innate immune system (Chaps. 60 and 317).

There are five general phases of host defenses: (1) migration of leukocytes to sites of antigen localization; (2) antigen-nonspecific recognition of pathogens by macrophages and other cells and systems of the innate immune system; (3) specific recognition of foreign antigens mediated by T and B lymphocytes; (4) amplification of the inflammatory response with recruitment of specific and nonspecific effector cells by complement components, cytokines, kinins, arachidonic acid metabolites, and mast cell–basophil products; and (5) macrophage, neutrophil, and lymphocyte participation in destruction of antigen with ultimate removal of antigen particles by phagocytosis (by macrophages or neutrophils) or by direct cytotoxic mechanisms (involving macrophages, neutrophils, DCs, and lymphocytes). Under normal circumstances, orderly progression of host defenses through these phases results in a well-controlled immune and inflammatory response that protects the host from the offending antigen. However,

dysfunction of any of the host defense systems can damage host tissue and produce clinical disease. Furthermore, for certain pathogens or antigens, the normal immune response itself might contribute substantially to the tissue damage. For example, the immune and inflammatory response in the brain to certain pathogens such as *M. tuberculosis* may be responsible for much of the morbidity rate of this disease in that organ system (Chap. 165). In addition, the morbidity rate associated with certain pneumonias such as that caused by *Pneumocystis jiroveci* may be associated more with inflammatory infiltrates than with the tissue-destructive effects of the microorganism itself (Chap. 207).

THE MOLECULAR BASIS OF LYMPHOCYTE–ENDOTHELIAL CELL INTERACTIONS

The control of lymphocyte circulatory patterns between the bloodstream and peripheral lymphoid organs operates at the level of lymphocyte–endothelial cell interactions to control the specificity of lymphocyte subset entry into organs. Similarly, lymphocyte–endothelial cell interactions regulate the entry of lymphocytes into inflamed tissue. Adhesion molecule expression on lymphocytes and endothelial cells regulates the retention and subsequent egress of lymphocytes within tissue sites of antigenic stimulation, delaying cell exit from tissue and preventing reentry into the circulating lymphocyte pool **(Fig. 314-11)**. All types of lymphocyte migration begin with lymphocyte attachment to specialized regions of vessels, termed *high endothelial venules* (HEVs). An important concept is that adhesion molecules do not generally bind their ligand until a conformational change (ligand activation) occurs in the adhesion molecule that allows ligand binding. Induction of a conformation-dependent determinant on an adhesion molecule can be accomplished by cytokines or via ligation of other adhesion molecules on the cell.

Figure 314-11

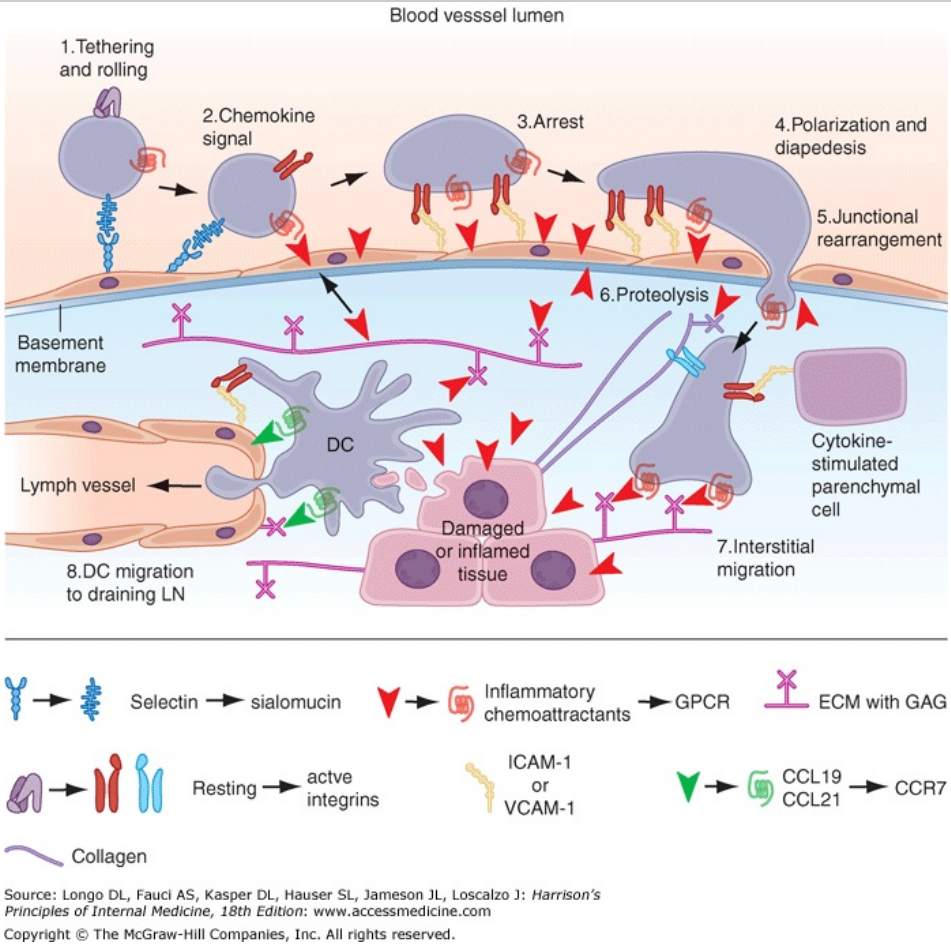

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Key migration steps of immune cells at sites of inflammation. Inflammation due to tissue damage or infection induces the release of cytokines (not shown) and inflammatory chemoattractants (red arrowheads) from distressed stromal cells and "professional" sentinels, such as mast cells and macrophages (not shown). The inflammatory signals induce upregulation of endothelial selectins and immunoglobulin "superfamily" members, particularly ICAM-1 and/or VCAM-1. Chemoattractants, particularly chemokines, are produced by or translocated across venular endothelial cells (red arrow) and are displayed in the lumen to rolling leukocytes. Those leukocytes that express the appropriate set of trafficking molecules undergo a multistep adhesion cascade (steps 1–3) and then polarize and move by diapedesis across the venular wall (steps 4 and 5). Diapedesis involves transient disassembly of endothelial junctions and penetration through the underlying basement membrane (step 6). Once in the extravascular (interstitial) space, the migrating cell uses different integrins to gain "footholds" on collagen fibers and other ECM molecules, such as laminin and fibronectin, and on inflammation-induced ICAM-1 on the surface of parenchymal cells (step 7). The migrating cell receives guidance cues from distinct sets of chemoattractants, particularly chemokines, which may be immobilized on glycosaminoglycans (GAG) that "decorate" many ECM molecules and stromal cells. Inflammatory signals also induce tissue DCs to undergo maturation. Once DCs process material from damaged tissues and invading pathogens, they upregulate CCR7, which allows them to enter draining lymph vessels that express the CCR7 ligand CCL21 (and CCL19). In lymph nodes (LN), these antigen-loaded mature DCs activate naïve T cells and expand pools of effector lymphocytes, which enter the blood and migrate back to the site of inflammation. T cells in tissue also use this CCR7-dependent route to migrate from peripheral sites to draining lymph nodes through afferent lymphatics. *(Adapted from AD Luster et al: Nat Immunol 6:1182, 2005; with permission from Macmillan Publishers Ltd. Copyright 2005.)*

The first stage of lymphocyte–endothelial cell interactions, *attachment and rolling* , occurs when lymphocytes leave the stream of flowing blood cells in a postcapillary venule and roll along venule endothelial cells (Fig. 314-11). Lymphocyte rolling is mediated by the l-selectin molecule (LECAM-1, LAM-1, CD62L) and slows cell transit time through venules, allowing time for activation of adherent cells.

The second stage of lymphocyte–endothelial cell interactions, *firm adhesion with activation-dependent stable arrest*, requires stimulation of lymphocytes by chemoattractants or by endothelial cell–derived cytokines. Cytokines thought to participate in adherent cell activation include members of the IL-8 family, platelet-activation factor, leukotriene B_4, and C5a. In addition, HEVs express chemokines, SLC (CCL21) and ELC (CCL19), which participate in this process. Following activation by chemoattractants, lymphocytes shed l-selectin from the cell surface and upregulate cell CD11b/18 (MAC-1) or CD11a/18 (LFA-1) molecules, resulting in firm attachment of lymphocytes to HEVs.

Lymphocyte homing to peripheral lymph nodes involves adhesion of l-selectin to glycoprotein HEV ligands collectively referred to as *peripheral node addressin (PNAd)*, whereas homing of lymphocytes to intestine Peyer's patches primarily involves adhesion of the α4,β7 integrin to mucosal addressin cell adhesion molecule-1 (MAdCAM-1) on the Peyer's patch HEVs. However, for migration to mucosal Peyer's patch lymphoid aggregates, naïve lymphocytes primarily use l-selectin, whereas memory lymphocytes use α4,β7 integrin. α4,β1 Integrin (CD49d/CD29, VLA-4)–VCAM-1 interactions are important in the initial interaction of memory lymphocytes with HEVs of multiple organs in sites of inflammation **(Table 314-16)**.

Table 314–16. Trafficking Molecules Involved in Inflammatory Disease Processes

		Proposed Leukocyte Receptors for Endothelial Traffic Signals		
Disease	**Key Effector Cell**	**L-Selectin, Ligand**	**GPCR**	**Integrin[a]**
Acute Inflammation				
Myocardial infarction	Neutrophil	PSGL-1	CXCR1, CXCR2, PAFR, BLT1	LFA-1, Mac-1
Stroke	Neutrophil	L-Selectin, PSGL-1	CXCR1, CXCR2, PAFR, BLT1	LFA-1, Mac-1
Ischemia-reperfusion	Neutrophil	PSGL-1	CXCR1, CXCR2, PAFR, BLT1	LFA-1, Mac-1
T_H1 Inflammation				
Atherosclerosis	Monocyte	PSGL-1	CCR1, CCR2, BLT1, CXCR2, CX3CR1	VLA-4
	T_H1	PSGL-1	CXCR3, CCR5	VLA-4
Multiple sclerosis	T_H1	PSGL-1 (?)	CXCR3, CXCR6	VLA-4, LFA-1
	Monocyte	PSGL-1 (?)	CCR2, CCR1	VLA-4, LFA-1
Rheumatoid arthritis	Monocyte	PSGL-1	CCR1, CCR2	VLA-1, VLA-2, VLA-4, LFA-1
	T_H1	PSGL-1	CXCR3, CXCR6	VLA-1, VLA-2, VLA-4, LFA-1
	Neutrophil	L-Selectin, PSGL-1	CXCR2, BLT1	LFA-1[b]
Psoriasis	Skin-homing T_H1	CLA	CCR4, CCR10, CXCR3	VLA-4[c], LFA-1
Crohn disease	Gut-homing T_H1	PSGL-1	CCR9, CXCR3	α4, β7, LFA-1
Type I diabetes	T_H1	PSGL-1 (?)	CCR4, CCR5	VLA-4, LFA-1
	CD8	L-Selectin (?), PSGL-1 (?)	CXCR3	VLA-4, LFA-1
Allograft rejection	CD8	PSGL-1	CXCR3, CX3CR1, BLT1	VLA-4, LFA-1
	B cell	L-Selectin, PSGL-1	CXCR5, CXCR4	VLA-4, LFA-1
Hepatitis	CD8	PSGL-1	CXCR3, CCR5, CXCR6	VLA-4
Lupus	T_H1	None	CXCR6	VLA-4[d]
	Plasmacytoid DC	L-Selectin, CLA	CCR7, CXCR3, ChemR23	LFA-1, Mac-1
	B cell	CLA (?)	CXCR5, CXCR4	LFA-1
T_H2 Inflammation				
Asthma	T_H2	PSGL-1	CCR4, CCR8, BLT1	LFA-1
	Eosinophil	PSGL-1	CCR3, PAFR, BLT1	VLA-4, LFA-1

	Mast cells	PSGL-1	CCR2, CCR3, BLT1	VLA-4, LFA-1
Atopic dermatitis	Skin-homing T_H2	CLA	CCR4, CCR10	VLA-4, LFA-1

[a] Various β_1 integrins have been linked in different ways in basal lamina and interstitial migration of distinct cell types and inflammatory settings.

[b] In some settings, Mac-1 has been linked to transmigration.

[c] CD44 can act in concert with VLA-4 in particular models of leukocyte arrest.

[d] T_H2 cells require VAP-1 to traffic to inflamed liver.

Source: From AD Luster et al: Nat Immunol 6:1182, 2005; with permission from Macmillan Publishers Ltd. Copyright 2005.

The third stage of leukocyte emigration in HEVs is *sticking and arrest* . Sticking of the lymphocyte to endothelial cells and arrest at the site of sticking are mediated predominantly by ligation of $\alpha 1,\beta 2$ integrin LFA-1 to the integrin ligand ICAM-1 on HEVs. While the first three stages of lymphocyte attachment to HEVs take only a few seconds, the fourth stage of lymphocyte emigration, *transendothelial migration*, takes ~10 min. Although the molecular mechanisms that control lymphocyte transendothelial migration are not fully characterized, the HEV CD44 molecule and molecules of the HEV glycocalyx (extracellular matrix) are thought to play important regulatory roles in this process (Fig. 314-11). Finally, expression of matrix metalloproteases capable of digesting the subendothelial basement membrane, rich in nonfibrillar collagen, appears to be required for the penetration of lymphoid cells into the extravascular sites.

Abnormal induction of HEV formation and use of the molecules discussed above have been implicated in the induction and maintenance of inflammation in a number of chronic inflammatory diseases. In animal models of Type 1 diabetes mellitus, MAdCAM-1 and GlyCAM-1 have been shown to be highly expressed on HEVs in inflamed pancreatic islets, and treatment of these animals with inhibitors of l-selectin and $\alpha 4$ integrin function blocked the development of Type 1 diabetes mellitus (Chap. 344). A similar role for abnormal induction of the adhesion molecules of lymphocyte emigration has been suggested in *rheumatoid arthritis* (Chap. 321), *Hashimoto's thyroiditis* (Chap. 341), *Graves' disease* (Chap. 341), *multiple sclerosis* (Chap. 380), *Crohn's disease* (Chap. 295), and *ulcerative colitis* (Chap. 295).

IMMUNE-COMPLEX FORMATION

Clearance of antigen by immune-complex formation between antigen, complement, and antibody is a highly effective mechanism of host defense. However, depending on the level of immune complexes formed and their physicochemical properties, immune complexes may or may not result in host and foreign cell damage. After antigen exposure, certain types of soluble antigen-antibody complexes freely circulate and, if not cleared by the reticuloendothelial system, can be deposited in blood vessel walls and in other tissues such as renal glomeruli and cause *vasculitis* or *glomerulonephritis* syndromes (Chaps. 283 and 326). Deficiencies of early complement components are associated with inefficient clearance of immune complexes and immune complex mediated tissue damage in autoimmune syndromes, while deficiencies of the later complement components are associated with susceptibility to recurrent neisseria infections (Table 314-17).

Table 314–17. Complement Deficiencies and Associated Diseases

Component	Associated Diseases
Classic Pathway	
Clq, Clr, Cls, C4	Immune-complex syndromes,* pyogenic infections
C2	Immune-complex syndromes,* few with pyogenic infections
C1 Inhibitor	Rare immune-complex disease, few with pyogenic infections
C3 and Alternative Pathway C3	
C3	Immune-complex syndromes,* pyogenic infections
D	Pyogenic infections
Properdin	*Neisseria* infections
I	Pyogenic infections
H	Hemolytic uremic syndrome
Membrane Attack Complex	
C5, C6, C7, C8	Recurrent *Neisseria* infections, immune-complex disease
C9	Rare *Neisseria* infections

*Immune-complex syndromes include systemic lupus erythematosus (SLE) and SLE-like syndromes, glomerulonephritis, and vasculitis syndromes.

Source: After JA Schifferli, DK Peters: Lancet 88:957, 1983. Copyright 1983, with permission from Elsevier.

IMMEDIATE-TYPE HYPERSENSITIVITY

Helper T cells that drive antiallergen IgE responses are usually T_H2-type inducer T cells that secrete IL-4, IL-5, IL-6, and IL-10. Mast cells and basophils have high-affinity receptors for the Fc portion of IgE (FcRI), and cell-bound antiallergen IgE effectively "arms" basophils and mast cells. Mediator release is triggered by antigen (allergen) interaction with Fc receptor–bound IgE, the mediators released are responsible for the pathophysiologic changes of

allergic diseases (Table 314-12). Mediators released from mast cells and basophils can be divided into three broad functional types: (1) those that increase vascular permeability and contract smooth muscle (histamine, platelet-activating factor, SRS-A, BK-A), (2) those that are chemotactic for or activate other inflammatory cells (ECF-A, NCF, leukotriene B_4), and (3) those that modulate the release of other mediators (BK-A, platelet-activating factor) (Chap. 317).

CYTOTOXIC REACTIONS OF ANTIBODY

In this type of immunologic injury, complement-fixing (C1-binding) antibodies against normal or foreign cells or tissues (IgM, IgG1, IgG2, IgG3) bind complement via the classic pathway and initiate a sequence of events similar to that initiated by immune-complex deposition, resulting in cell lysis or tissue injury. Examples of antibody-mediated cytotoxic reactions include red cell lysis in *transfusion reactions*, *Goodpasture's syndrome* with anti–glomerular basement membrane antibody formation, and *pemphigus vulgaris* with antiepidermal antibodies inducing blistering skin disease.

CLASSIC DELAYED-TYPE HYPERSENSITIVITY REACTIONS

Inflammatory reactions initiated by mononuclear leukocytes and not by antibody alone have been termed *delayed-type hypersensitivity reactions*. The term *delayed* has been used to contrast a secondary cellular response that appears 48–72 h after antigen exposure with an *immediate* hypersensitivity response generally seen within 12 h of antigen challenge and initiated by basophil mediator release or preformed antibody. For example, in an individual previously infected with *M. tuberculosis* organisms, intradermal placement of tuberculin purified-protein derivative as a skin test challenge results in an indurated area of skin at 48–72 h, indicating previous exposure to tuberculosis.

The cellular events that result in classic delayed-type hypersensitivity responses are centered around T cells (predominantly, though not exclusively, IFN-γ, IL-2, and TNF-α-secreting T_H1-type helper T cells) and macrophages. Recently NK cells have been suggested to play a major role in the form of delayed hypersensitivity that occurs following skin contact with immunogens. First, local immune and inflammatory responses at the site of foreign antigen upregulate endothelial cell adhesion molecule expression, promoting the accumulation of lymphocytes at the tissue site. In the general schemes outlined in Figs. 314-2 and 314-3, antigen is processed by dendritic cells and presented to small numbers of CD4+ T cells expressing a TCR specific for the antigen. IL-12 produced by APCs induces T cells to produce IFN-γ (T_H1 response). Macrophages frequently undergo epithelioid cell transformation and fuse to form multinucleated giant cells in response to IFN-γ. This type of mononuclear cell infiltrate is termed *granulomatous inflammation*. Examples of diseases in which delayed-type hypersensitivity plays a major role are fungal infections (*histoplasmosis*; Chap. 199), mycobacterial infections (*tuberculosis*, *leprosy*; Chaps. 165 and 166), chlamydial infections (*lymphogranuloma venereum*; Chap. 176), helminth infections (*schistosomiasis*; Chap. 219), reactions to toxins (*berylliosis*; Chap. 256), and hypersensitivity reactions to organic dusts (*hypersensitivity pneumonitis*; Chap. 255). In addition, delayed-type hypersensitivity responses play important roles in tissue damage in autoimmune diseases such as *rheumatoid arthritis*, *temporal arteritis*, and granulomatosis with polyangiitis (Wegener's) (Chaps. 321 and 326).

Clinical Evaluation of Immune Function

Clinical assessment of immunity requires investigation of the four major components of the immune system that participate in host defense and in the pathogenesis of autoimmune diseases: (1) humoral immunity (B cells); (2) cell-mediated immunity (T cells, monocytes); (3) phagocytic cells of the reticuloendothelial system (macrophages), as well as polymorphonuclear leukocytes; and (4) complement. Clinical problems that require an evaluation of immunity include chronic infections, recurrent infections, unusual infecting agents, and certain autoimmune syndromes. The type of clinical syndrome under evaluation can provide information regarding possible immune defects (Chap. 316). Defects in cellular immunity generally result in viral, mycobacterial, and fungal infections. An extreme example of deficiency in cellular immunity is AIDS (Chap. 189). Antibody deficiencies result in recurrent bacterial infections, frequently with organisms such as *S. pneumoniae* and *Haemophilus influenzae* (Chap. 316). Disorders of phagocyte function are frequently manifested by recurrent skin infections, often due to *Staphylococcus aureus* (Chap. 60). Finally, deficiencies of early and late complement components are associated with autoimmune phenomena and recurrent *Neisseria* infections (Table 314-17). **For further discussion of useful initial screening tests of immune function, see Chap. 316.**

Immunotherapy

Many therapies for autoimmune and inflammatory diseases involve the use of nonspecific immune-modulating or immunosuppressive agents such as glucocorticoids or cytotoxic drugs. The goal of development of new treatments for immune-mediated diseases is to design ways to specifically interrupt pathologic immune responses, leaving nonpathologic immune responses intact. Novel ways to interrupt pathologic immune responses that are under investigation include the use of anti-inflammatory cytokines or specific cytokine inhibitors as anti-inflammatory agents, the use of monoclonal antibodies against T or B lymphocytes as therapeutic agents, the induction of anergy by administration of soluble CTLA-4 protein, the use of intravenous Ig for certain infections and immune complex –mediated diseases, the use of specific cytokines to reconstitute components of the immune system, and bone marrow transplantation to replace the pathogenic immune system with a more normal immune system (Chaps. 60, 316, and 189). In particular, the use of a monoclonal antibody to B cells (rituximab, anti-CD20 MAb) is approved in the United States for the treatment of non-Hodgkin's lymphoma (Chap. 110) and, in combination with methotrexate, for treatment of adult patients with severe rheumatoid arthritis resistant to TNF-α inhibitors (Chap. 321).

CYTOKINES AND CYTOKINE INHIBITORS

Recently, a humanized mouse anti-TNF-α monoclonal antibody (MAb) has been shown to be effective in both rheumatoid arthritis and ulcerative colitis. Use of anti-TNF-α antibody therapy has resulted in clinical improvement in patients with these diseases and has opened the way for targeting TNF-α to treat other severe forms of autoimmune and/or inflammatory disease. Blockage of TNF-α has been effective in *rheumatoid arthritis*, *psoriasis*, *Crohn's disease*, and *ankylosing spondylitis*. Anti-TNF-α MAb (infliximab) has been approved by the FDA for treatment of patients with rheumatoid arthritis.

Other cytokine inhibitors are recombinant soluble TNF-α receptor (R) fused to human Ig and Anakinra (soluble *IL-1 receptor antagonist*, or IL-1 ra). The treatment of autoinflammatory syndromes (Table 314-6) with recombinant IL-1 receptor antagonist can prevent symptoms in these syndromes, since the overproduction of IL-1β is a hallmark of these diseases. Soluble TNF-αR (etanercept) and IL-1 ra act to inhibit the activity of pathogenic cytokines in rheumatoid arthritis, i.e., TNF-α and IL-1, respectively. Similarly, anti-IL-6, IFN-β, and IL-11 act to inhibit pathogenic proinflammatory cytokines. Anti-IL-6 inhibits IL-6 activity, while IFN-β and IL-11 decrease IL-1 and TNF-α production.

Of particular note has been the successful use of IFN-γ in the treatment of the phagocytic cell defect in *chronic granulomatous disease* (Chap. 60).

MONOCLONAL ANTIBODIES TO T AND B CELLS

The OKT3 MAb against human T cells has been used for several years as a T cell–specific immunosuppressive agent that can substitute for horse anti-thymocyte globulin (ATG) in the treatment of solid organ transplant rejection. OKT3 produces fewer allergic reactions than ATG but does induce human anti-mouse Ig antibody—thus limiting its use. Anti-CD4 MAb therapy has been used in trials to treat patients with rheumatoid arthritis. While inducing profound immunosuppression, anti-CD4 MAb treatment also induces susceptibility to severe infections. Treatment of patients with a MAb against the T cell molecule CD40 ligand (CD154) is under investigation to induce tolerance to organ transplants, with promising results reported in animal studies. Monoclonal antibodies to the CD25 (IL-2α) receptor (Basiliximab) are being used for treatment of graft-versus-host disease in bone marrow transplantation, and anti-CD20 MAb (rituximab) is used to treat hematologic neoplasms, autoimmune diseases, and kidney transplant rejection. The anti-IgE monoclonal antibody (omalizumab) is used for blocking antigen-specific IgE that causes *hay fever* and *allergic rhinitis* (Chap. 317); however, side effects of anti-IgE include increased risk of anaphylaxis. Studies have shown that T_h17 cells, in addition to T_h1, are mediators of inflammation in Crohn's disease, and anti –IL-12/IL-23p40 antibody therapy has been studied as a treatment.

It is important to realize the potential risks for these immunosuppressive monoclonal antibodies. Natalizumab is a humanized IgG antibody against an α4 integrin that inhibits leukocyte migration into tissues, and has been approved for treatment of multiple sclerosis in the United States. Both it and anti-CD20 (rituximab) have been associated with the onset of progressive multifocal leukoencephalopathy (PML)—a serious and usually fatal CNS infection caused by JC polyoma virus. Efalizumab, a humanized IgG monoclonal antibody previously approved for treatment of plaque psoriasis, has now been taken off the market due to reactivation of JC virus leading to fatal PML. Thus, use of any currently approved immunosuppressant immunotherapies should be undertaken with caution and with careful monitoring of patients according to FDA guidelines.

TOLERANCE INDUCTION

Specific immunotherapy has moved into a new era with the introduction of soluble CTLA-4 protein into clinical trials. Use of this molecule to block T cell activation via TCR/CD28 ligation during organ or bone marrow transplantation has showed promising results in animals and in early human clinical trials. Specifically, treatment of bone marrow with CTLA-4 protein reduces rejection of the graft in HLA-mismatched bone marrow transplantation. In addition, promising results with soluble CTLA-4 have been reported in the downmodulation of autoimmune T cell responses in the treatment of psoriasis; and it is being studied for treatment of systemic lupus erythematosus (Chap. 319).

INTRAVENOUS IMMUNOGLOBULIN (IVIG)

IVIg has been used successfully to block reticuloendothelial cell function and immune complex clearance in various immune cytopenias such as immune thrombocytopenia (Chap. 115). In addition, IVIg is useful for prevention of tissue damage in certain inflammatory syndromes such as Kawasaki disease (Chap. 326) and as Ig replacement therapy for certain types of immunoglobulin deficiencies (Chap. 316). In addition, controlled clinical trials support the use of IVIg in selected patients with graft-versus-host disease, multiple sclerosis, myasthenia gravis, Guillain-Barré syndrome, and chronic demyelinating polyneuropathy.

STEM CELL TRANSPLANTATION

Hematopoietic stem cell transplantation (SCT) is now being comprehensively studied to treat several autoimmune diseases, including systemic lupus erythematosus, multiple sclerosis, and scleroderma. The goal of immune reconstitution in autoimmune disease syndromes is to replace a dysfunctional immune system with a normally reactive immune cell repertoire. Preliminary results in patients with scleroderma and lupus have showed encouraging results. Controlled clinical trials in these three diseases are now being launched in the United States and Europe to compare the toxicity and efficacy of conventional immunosuppression therapy with that of myeloablative autologous SCT.

Thus, a number of recent insights into immune system function have spawned a new field of interventional immunotherapy and have enhanced the prospect for development of specific and nontoxic therapies for immune and inflammatory diseases.

FURTHER READINGS

Andreakos ETH et al: Role of cytokines, in *Rheumatoid Arthritis*, EW St. Clair et al (eds). Philadelphia, Lippincott Williams & Wilkins, 2004, pp 134–149

Berger JR et al: Monoclonal antibodies and progressive multifocal leukoencephalopathy. MAbs 1:583, 2009[PMID: 20073129] [Full Text]

Blander JM et al: Toll-dependent selection of microbial antigens for presentation by dendritic cells. Nature 440:808, 2006[PMID: 16489357] [Full Text]

Carson KR et al: Monoclonal antibody-associated progressive multifocal leucoencephalopathy in patients treated with rituximab, natalizumab, and efalizumab: A review from the Research on Adverse Drug Events and Reports (RADAR) Project. Lancet Oncol 10:816, 2009[PMID: 19647202] [Full Text]

Davidson A, Diamond B: Autoimmune diseases. N Engl J Med 345:340, 2001[PMID: 11484692] [Full Text]

Diaz-Pena R et al: KIR genes and their role in spondyloarthropathies. Adv Exp Med Biol 649:286, 2009[PMID: 19731638] [Full Text]

Falschlehner C et al: Following TRAIL's path in the immune system. Immunology 127:145, 2009[PMID: 19476510] [Full Text]

Franchi L et al: The inflammasome: A caspase-1-activation platform that regulates immune responses. Nat Immunol 10:241, 2009[PMID: 19221555] [Full Text]

Hotchkiss RS et al: Cell death. N Engl J Med 361:1570, 2009[PMID: 19828534] [Full Text]

Iwasaki A, Medzhitov R: Regulation of adaptive immunity by the innate immune system. Science 327:291, 2010[PMID: 20075244] [Full Text]

Kelley WN et al (eds): *Textbook of Rheumatology* , 4th ed. Philadelphia, Saunders, 1993, chaps. 6, 7, 10, 13, 15, 16

Koretzky GA et al: SLP76 and SLP65: Complex regulation of signaling in lymphocytes and beyond. Nat Rev Immunol 6:67, 2006[PMID: 16493428] [Full Text]

Korman BD et al: Progressive multifocal leukoencephalopathy, efalizumab, and immunosuppression: A cautionary tale for dermatologists. Arch Dermatol 145:937, 2009[PMID: 19687432] [Full Text]

Kronenberg M et al: Innate-like recognition of microbes by invariant natural killer T cells. Curr Opin Immunol 21:391, 2009[PMID: 19646850] [Full Text]

Lanier L: NK cell recognition. Annu Rev Immunol 23:225, 2005[PMID: 15771571] [Full Text]

Lernmark A: Autoimmune diseases: Are markers ready for prediction? J Clin Invest 108:1091, 2001[PMID: 11602614] [Full Text]

Louten J et al: Development and function of Th17 cells in health and disease. J Allergy Clin Immunol 123:1004, 2009[PMID: 19410689] [Full Text]

Macdonald TT et al: Immunity, inflammation, and allergy in the gut. Science 307:1920, 2005[PMID: 15790845] [Full Text]

Martinon F et al: The inflammasomes: Guardians of the body. Annu Rev Immunol 27:229, 2009[PMID: 19302040] [Full Text]

Middendorp S et al: NKT cells in mucosal immunity. Mucosal Immunol 2(5):393, 2009[PMID: 19587641] [Full Text]

Morley BJ, Walport MJ: *The Complement Facts Books*. London, Academic Press, 2000, Chap. 2

Mullauer L et al: Mutations in apoptosis genes: A pathogenetic factor for human disease. Mutat Res 488:211, 2001[PMID: 11397650] [Full Text]

Paust S et al: Adaptive immune responses mediated by natural killer cells. Immunol Rev 235:286, 2010[PMID: 20536570] [Full Text]

Rajagopalan S et al: Understanding how combinations of HLA and KIR genes influence disease. J Exp Med 201:1025, 2005[PMID: 15809348] [Full Text]

Ramaswamy M et al: Harnessing programmed cell death as a therapeutic strategy in rheumatic diseases. Nat Rev Rheumatol 2011 [epub ahead of print]

Rinaudo C et al: Vaccinology in the Genome Era. J Clin Invest 119:2515, 2009[PMID: 19729849] [Full Text]

Rivera J et al: Molecular regulation of mast cell activation. J Allergy Clin Immunol 6:1214, 2006

Romagnani S: CD4 effector cells, in *Inflammation: Basic Principles and Clinical Correlates*, 3rd ed, J Gallin, R Snyderman (eds). Philadelphia, Lippincott Williams & Wilkins, 1999, pp 177

Schroder K et al: The NLRP3 inflammasome: a sensor for metabolic danger? Science 327:296, 2010[PMID: 20075245] [Full Text]

Steinman RM, Banchereau J: Taking dendritic cells into medicine. Nature 449:419, 2007[PMID: 17898760] [Full Text]

Ting Jpy et al: How the noninflammasome NLRs function in the innate immune system. Science 327:286, 2010[PMID: 20075243] [Full Text]

Ueno H et al: Targeting human dendritic cell subsets for improved vaccines. Semin Immunol 23:21, 2011[PMID: 21277223] [Full Text]

van Duin D et al: Triggering TLR signaling in vaccination. Trends Immunol 27:49, 2006

Weaver C et al: Interplay between the Th17 and Treg cell lineages: a co-evolutionary perspective. Nat Rev Immunol 9:883, 2009[PMID: 19935807] [Full Text]

Whelan B et al: mAbs in nonlupus autoimmune rheumatic disease. Curr Opin Hemat 16:280, 2009[PMID: 19444098] [Full Text]

Williams AP et al: Hanging in the balance. Mol Interv 5:226, 2005[PMID: 16123537] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE HLA COMPLEX AND ITS PRODUCTS

The human major histocompatibility complex (MHC), commonly called the human leukocyte antigen (HLA) complex, is a 4-megabase (Mb) region on chromosome 6 (6p21.3) that is densely packed with expressed genes. The best known of these genes are the HLA class I and class II genes, whose products are critical for immunologic specificity and transplantation histocompatibility, and they play a major role in susceptibility to a number of autoimmune diseases. Many other genes in the HLA region are also essential to the innate and antigen-specific functioning of the immune system. The HLA region shows extensive conservation with the MHC of other mammals in terms of genomic organization, gene sequence, and protein structure and function.

The *HLA class I genes* are located in a 2-Mb stretch of DNA at the telomeric end of the HLA region **(Fig. 315-1)**. The classic (MHC class Ia) HLA-A, -B, and -C loci, the products of which are integral participants in the immune response to intracellular infections, tumors, and allografts, are expressed in all nucleated cells and are highly polymorphic in the population. *Polymorphism* refers to a high degree of allelic variation within a genetic locus that leads to extensive variation between different individuals expressing different alleles. More than 650 alleles at HLA-A, 1000 at HLA-B, and 360 at HLA-C have been identified in different human populations, making this the most highly polymorphic segment known within the human genome. Each of the alleles at these loci encodes a *heavy chain* (also called an α *chain*) that associates noncovalently with the nonpolymorphic light chain β_2*-microglobulin*, encoded on chromosome 15.

Figure 315-1

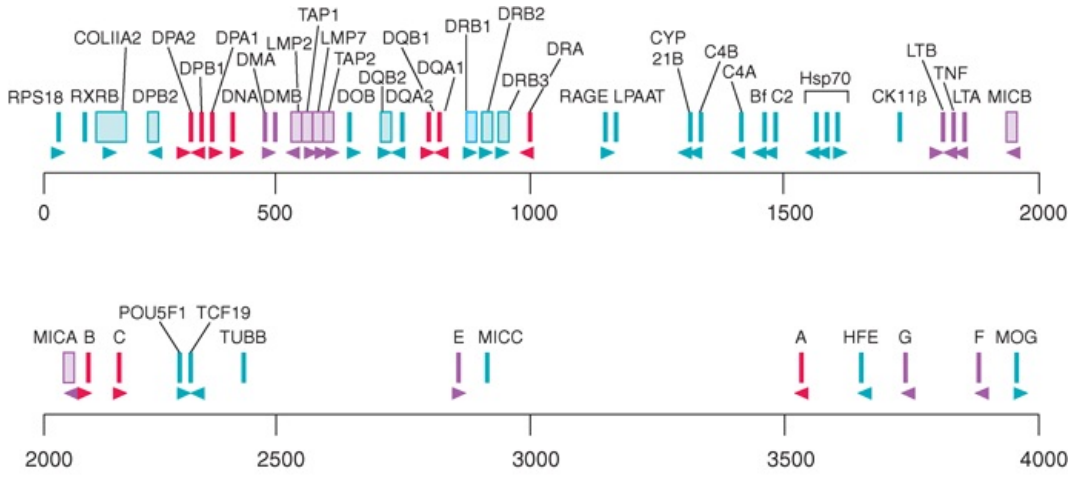

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Physical map of the HLA region, showing the class I and class II loci, other immunologically important loci, and a sampling of other genes mapped to this region. Gene orientation is indicated by arrowheads. Scale is in kilobase (kb). The approximate genetic distance from DP to A is 3.2 cM. This includes 0.8 cM between A and B (including 0.2 cM between C and B), 0.4–0.8 cM between B and DR-DQ, and 1.6–2.0 cM between DR-DQ and DP.

The nomenclature of HLAgenes and their products reflects the grafting of newer DNA sequence information on an older system based on serology. Among class I genes, alleles of the HLA-A, -B, and -C loci were originally identified in the 1950s, 1960s, and 1970s by alloantisera, derived primarily from multiparous women, who in the course of normal pregnancy produce antibodies against paternal antigens expressed on fetal cells. The serologic allotypes were designated by consecutive numbers (e.g., HLA-A1, HLA-B8). Currently, under World Health Organization (WHO) nomenclature, class I alleles are given a single designation that indicates locus, serologic specificity, and sequence-based subtype. For example, HLA-A*0201 indicates subtype 1 of the serologically defined allele HLA-A2. Subtypes that differ from each other at the nucleotide but not the amino acid sequence level are

designated by an extra numeral (e.g., HLA-B*07021 and HLA-B*07022 are two variants of the HLA-B702 subtype of HLA-B*07). The nomenclature of class II genes, discussed below, is made more complicated by the fact that both chains of a class II molecule are encoded by closely linked HLA-encoded loci, both of which may be polymorphic, and by the presence of differing numbers of isotypic DRB loci in different individuals. It has become clear that accurate HLA genotyping requires DNA sequence analysis, and the identification of alleles at the DNA sequence level has contributed greatly to the understanding of the role of HLA molecules as peptide-binding ligands, to the analysis of associations of HLA alleles with certain diseases, to the study of the population genetics of HLA, and to a clearer understanding of the contribution of HLA differences to allograft rejection and graft-versus-host disease. Current databases of HLA class I and class II sequences can be accessed by the Internet (e.g., from the IMGT/HLA Database, *http://www.ebi.ac.uk/imgt/hla*), and frequent updates of HLA gene lists are published in several journals.

The biologic significance of this MHC genetic diversity, resulting in extreme variation in the human population, is evident from the perspective of the structure of MHC molecules. As shown in **Fig. 315-2**, the MHC class I and class II genes encode MHC molecules that bind small peptides, and together this complex (pMHC; peptide-MHC) forms the ligand for recognition by T lymphocytes, through the antigen-specific T cell receptor (TCR). There is a direct link between the genetic variation and this structural interaction: The allelic changes in genetic sequence result in diversification of the peptide-binding capabilities of each MHC molecule and in differences for specific TCR binding. Thus, different pMHC complexes bind different antigens and are targets for recognition by different T cells.

Figure 315-2

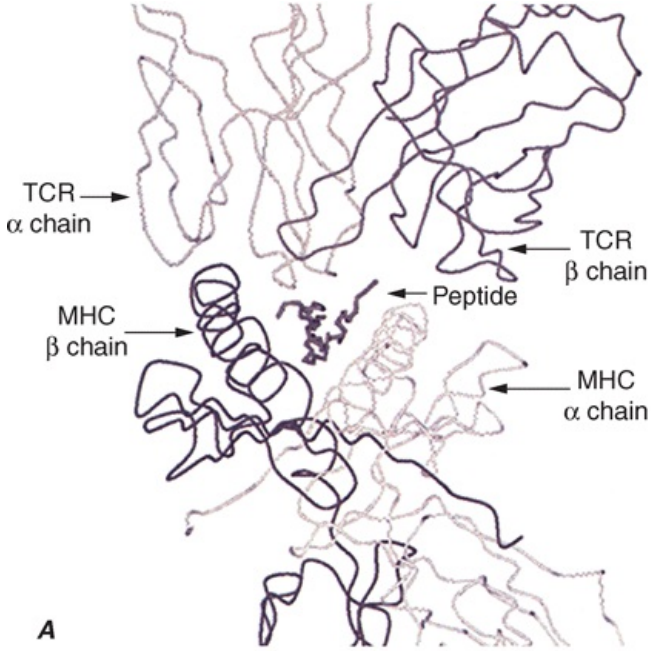

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

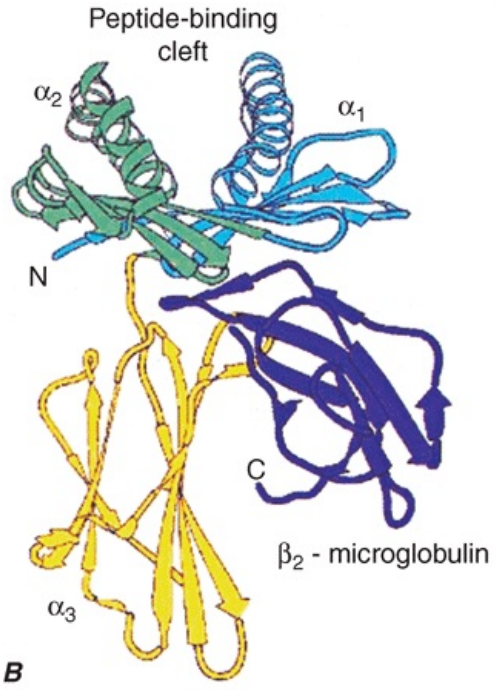

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

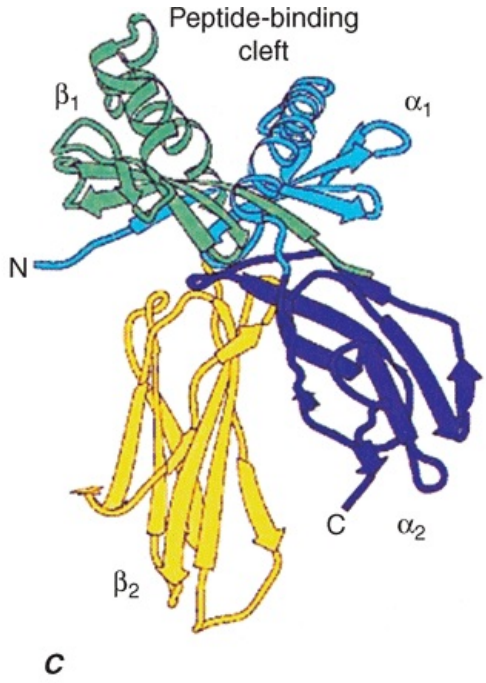

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A.* The trimolecular complex of TCR (top), MHC molecule (bottom)** and a bound peptide form the structural determinants of specific antigen recognition. Other panels (B.*** and ***C.***) show the domain structure of MHC class I (B) and class II (C) molecules. The α_1 and α_2 domains of class I and the α_1 and β_1 domains of class II form a β-sheet platform that forms the floor of the peptide-binding groove, and α helices that form the sides of the groove. The α_3 (*B*) and β_2 domains (*C*) project from the cell surface and form the contact sites for CD8 and CD4, respectively. *(Adapted from EL Reinhertz et al: Science 286:1913, 1999; and C Janeway et al: Immunobiology Bookshelf, 2nd ed, Garland Publishing, New York, 1997; with permission.)*

The class I MHC and class II MHC structures, shown in Fig. 315-2B, C, are structurally closely related; however, there are a few key differences. While both bind peptides and present them to T cells, the binding pockets have different shapes, which influence the types of immune responses that result (discussed below). In addition, there are structural contact sites for T cell molecules known as CD8 and CD4, expressed on the class I or class II membrane-proximal domains, respectively. This ensures that when peptide antigens are presented by class I molecules, the responding T cells are

predominantly of the CD8 class, and similarly, that T cells responding to class II pMHC complexes are predominantly CD4.

The nonclassic, or class Ib, MHC molecules, HLA-E, -F, and -G, are much less polymorphic than MHC Ia and appear to have distinct functions. The HLA-E molecule has a peptide repertoire displaying signal peptides cleaved from classic MHC class I molecules and is the major self-recognition target for the natural killer (NK) cell–inhibitory receptors NKG2A or NKG2C paired with CD94 (see below and Chap. 314). This appears to be a function of immune surveillance, as loss of MHC class I signal peptides serves as a surrogate marker for injured or infected cells, leading to release of the inhibitory signal and subsequent activation of NK cells. HLA-E can also bind and present peptides to CD8 T cells, albeit with a limited scope, as only three HLA-E alleles are known. HLA-G is expressed selectively in extravillous trophoblasts, the fetal cell population directly in contact with maternal tissues. It binds a wide array of peptides, is expressed in six different alternatively spliced forms, and provides inhibitory signals to both NK cells and T cells, presumably in the service of maintaining maternofetal tolerance; 14 HLA-G alleles have been identified. The protein product of HLA-F is found mainly intracellularly, and the function of this locus, which encodes four alleles, remains largely unknown.

Additional class I–like genes have been identified, some HLA-linked and some encoded on other chromosomes, that show only distant homology to the class Ia and Ib molecules but share the three-dimensional class I structure. Those on chromosome 6p21 include MIC-A and MIC-B, which are encoded centromeric to HLA-B, and HLA-HFE, located 3 to 4 cM (centi-Morgan) telomeric of HLA-F. MIC-A and MIC-B do not bind peptide but are expressed on gut and other epithelium in a stress-inducible manner and serve as activation signals for certain $\gamma\delta$ T cells, NK cells, CD8 T cells, and activated macrophages, acting through the activating NKG2D receptors. Sixty-seven MIC-A and 30 MIC-B alleles are known, and additional diversification comes from variable alanine repeat sequences in the transmembrane domain. Due to this structural diversity, MIC-A can be recognized as a foreign tissue target during organ transplantation, contributing to graft failure. HLA-HFE encodes the gene defective in hereditary hemochromatosis (Chap. 357). Among the non-HLA, class I–like genes, CD1 refers to a family of molecules that present glycolipids or other nonpeptide ligands to certain T cells, including T cells with NK activity; FcRn binds IgG within lysosomes and protects it from catabolism (Chap. 314); and Zn-α_2-glycoprotein 1 binds a nonpeptide ligand and promotes catabolism of triglycerides in adipose tissue. Like the HLA-A, -B, -C, -E, -F, and -G heavy chains, each of which forms a heterodimer with β_2-microglobulin (Fig. 315-2), the class I–like molecules, HLA-HFE, FcRn, and CD1 also bind to β_2-microglobulin, but MIC-A, MIC-B, and Zn-α_2-glycoprotein 1 do not.

The *HLA class II region* is also illustrated in Fig. 315-1. Multiple class II genes are arrayed within the centromeric 1 Mb of the HLA region, forming distinct haplotypes. A *haplotype* refers to an array of alleles at polymorphic loci along a chromosomal segment. Multiple class II genes are present on a single haplotype, clustered into three major subregions: HLA-DR, -DQ, and -DP. Each of these subregions contains at least one functional alpha (A) locus and one functional beta (B) locus. Together these encode proteins that form the α and β polypeptide chains of a mature class II HLA molecule. Thus, the DRA and DRB genes encode an HLA-DR molecule; products of the DQAI and DQBI genes form an HLA-DQ molecule; and the DPAI and DPBI genes encode an HLA-DP molecule. There are several DRB genes (DRB1, DRB2, DRB3, etc.), so that two expressed DR molecules are encoded on most haplotypes by combining the α-chain product of the DRA gene with separate β chains. More than 530 alleles have been identified at the HLA-DRB1 locus, with most of the variation occurring within limited segments encoding residues that interact with antigens. Detailed analysis of sequences and population distribution of these alleles strongly suggest that this diversity is actively selected by environmental pressures associated with pathogen diversity.

In the DQ region, both DQA1 and DQB1 are polymorphic, with 34 DQA1 alleles and 72 DQB1 alleles. The current nomenclature is largely analogous to that discussed above for class I, using the convention "locus * allele." Thus, for example, subtypes of the serologically defined specificity DR4, encoded by the DRB1 locus, are termed DRB1*0401, *0402, etc. In addition to allelic polymorphism, products of different DQA1 alleles can, with some limitations, pair with products of different DQB1 alleles through both *cis* and *trans* pairing to create combinatorial complexity and expand the number of expressed class II molecules. Because of the enormous allelic diversity in the general population, most individuals are heterozygous at all of the class I and class II loci. Thus, most individuals express six classic class I molecules (two each of HLA-A, -B, and -C) and around eight class II molecules—two DP, two DR (more in the case of haplotypes with additional functional DRB genes), and up to four DQ (two *cis* and two *trans*).

Other Genes in the MHC

In addition to the class I and class II genes themselves, there are numerous genes interspersed among the HLA loci that have interesting and important immunologic functions. Our current concept of the function of MHC genes now encompasses many of these additional genes, some of which are also highly polymorphic. Indeed, direct comparison of the complete DNA sequences for eight of the entire 4-Mb MHC regions from different haplotypes show >44,000 nucleotide variations, encoding an extremely high potential for biologic diversity, and at least 97 genes located in this region are known to

have coding region sequence variation. Specific examples include the TAP and LMP genes, as discussed in more detail below, which encode molecules that participate in intermediate steps in the HLA class I biosynthetic pathway. Another set of HLA genes, DMA and DMB, perform an analogous function for the class II pathway. These genes encode an intracellular molecule that facilitates the proper complexing of HLA class II molecules with antigen (see below). The *HLA class III region* is a name given to a cluster of genes between the class I and class II complexes, which includes genes for the two closely related cytokines tumor necrosis factor (TNF)-α and lymphotoxin (TNF-β); the complement components C2, C4, and Bf; heat shock protein (HSP)70; and the enzyme 21-hydroxylase.

The class I genesHLA-A, -B, and -C are expressed in all nucleated cells, although generally to a higher degree on leukocytes than on nonleukocytes. In contrast, the class II genes show a more restricted distribution: HLA-DR and HLA-DP genes are constitutively expressed on most cells of the myeloid cell lineage, whereas all three class II gene families (HLA-DR, -DQ, and -DP) are inducible by certain stimuli provided by inflammatory cytokines such as interferon γ. Within the lymphoid lineage, expression of these class II genes is constitutive on B cells and inducible on human T cells. Most endothelial and epithelial cells in the body, including the vascular endothelium and the intestinal epithelium, are also inducible for class II gene expression. Thus, while these somatic tissues normally express only class I and not class II genes, during times of local inflammation they are recruited by cytokine stimuli to express class II genes as well, thereby becoming active participants in ongoing immune responses. Class II expression is controlled largely at the transcriptional level through a conserved set of promoter elements that interact with a protein known as *CIITA*. Cytokine-mediated induction of CIITA is a principal method by which tissue-specific expression of HLA gene expression is controlled. Other HLA genes involved in the immune response such as TAP and LMP, are also susceptible to upregulation by signals such as interferon γ. Sequence data for the entire HLA region can be accessed on the Internet (e.g., *http://www.sanger.ac.uk/HGP/Chr6/MHC*).

Linkage Disequilibrium

In addition to extensive polymorphism at the class I and class II loci, another characteristic feature of the HLA complex is *linkage disequilibrium*. This is formally defined as a deviation from Hardy-Weinberg equilibrium for alleles at linked loci. This is reflected in the very low recombination rates between certain loci within the HLA complex. For example, recombination between DR and DQ loci is almost never observed in family studies, and characteristic haplotypes with particular arrays of DR and DQ alleles are found in every population. Similarly, the complement components C2, C4, and Bf are almost invariably inherited together, and the alleles at these loci are found in characteristic haplotypes. In contrast, there is a recombinational hotspot between DQ and DP, which are separated by 1–2 cM of genetic distance, despite their close physical proximity. Certain extended haplotypes encompassing the interval from DQ into the class I region are commonly found, the most notable being the haplotype DR3-B8-A1, which is found, in whole or in part, in 10–30% of northern European whites. It has been hypothesized that selective pressures may maintain linkage disequilibrium in HLA, but this remains to be determined. As discussed below under HLA and immunologic disease, one consequence of the phenomenon of linkage disequilibrium has been the resulting difficulty in assigning HLA-disease associations to a single allele at a single locus.

MHC STRUCTURE AND FUNCTION

Class I and class II molecules display a distinctive structural architecture, which contains specialized functional domains responsible for the unique genetic and immunologic properties of the HLA complex. The principal known function of both class I and class II HLA molecules is to bind antigenic peptides in order to present antigen to an appropriate T cell. The ability of a particular peptide to satisfactorily bind to an individual HLA molecule is a direct function of the molecular fit between the amino acid residues on the peptide with respect to the amino acid residues of the HLA molecule. The bound peptide forms a tertiary structure called the *MHC-peptide complex*, which communicates with T lymphocytes through binding to the TCR molecule. The first site of TCR-MHC-peptide interaction in the life of a T cell occurs in the thymus, where self-peptides are presented to developing thymocytes by MHC molecules expressed on thymic epithelium and hematopoietically derived antigen-presenting cells, which are primarily responsible for positive and negative selection, respectively (Chap. 314). Thus, the population of MHC–T cell complexes expressed in the thymus shapes the TCR repertoire. Mature T cells encounter MHC molecules in the periphery both in the maintenance of tolerance (Chap. 318) and in the initiation of immune responses. The MHC-peptide-TCR interaction is the central event in the initiation of most antigen-specific immune responses, since it is the structural determinant of the specificity. For potentially immunogenetic peptides, the ability of a given peptide to be generated and bound by an HLA molecule is a primary feature of whether or not an immune response to that peptide can be generated, and the repertoire of peptides that a particular individual's HLA molecules can bind exerts a major influence over the specificity of that individual's immune response.

When a TCR molecule binds to an HLA-peptide complex, it forms intermolecular contacts with both the antigenic peptide and with the HLA molecule itself. The outcome of this recognition event depends on the density and duration of the binding interaction, accounting for a dual specificity requirement for

activation of the T cell. That is, the TCR must be specific both for the antigenic peptide and for the HLA molecule. The polymorphic nature of the presenting molecules, and the influence that this exerts on the peptide repertoire of each molecule, results in the phenomenon of *MHC restriction* of the T cell specificity for a given peptide. The binding of CD8 or CD4 molecules to the class I or class II molecule, respectively, also contributes to the interaction between T cell and the HLA-peptide complex, by providing for the selective activation of the appropriate T cell.

Class I Structure

(Fig. 315-2*B*) As noted above, MHC class I molecules provide a cell-surface display of peptides derived from intracellular proteins, and they also provide the signal for self-recognition by NK cells. Surface-expressed class I molecules consist of an MHC-encoded 44-kD glycoprotein heavy chain, a non-MHC-encoded 12-kD light chain β_2-microglobulin, and an antigenic peptide, typically 8–11 amino acids in length and derived from intracellularly produced protein. The heavy chain displays a prominent peptide-binding groove. In HLA-A and -B molecules, the groove is ~3 nm in length by 1.2 nm in maximum width (30 Å x 12 Å), whereas it is apparently somewhat wider in HLA-C. Antigenic peptides are noncovalently bound in an extended conformation within the peptide-binding groove, with both N- and C-terminal ends anchored in pockets within the groove (A and F pockets, respectively) and, in many cases, with a prominent kink, or arch, approximately one-third of the way from the N-terminus that elevates the peptide main chain off the floor of the groove.

A remarkable property of peptide binding by MHC molecules is the ability to form highly stable complexes with a wide array of peptide sequences. This is accomplished by a combination of peptide sequence–independent and peptide sequence–dependent bonding. The former consists of hydrogen bond and van der Waals interactions between conserved residues in the peptide-binding groove and charged or polar atoms along the peptide backbone. The latter is dependent upon the six side pockets that are formed by the irregular surface produced by protrusion of amino acid side chains from within the binding groove. The side chains lining the pockets interact with some of the peptide side chains. The sequence polymorphism among different class I alleles and isotypes predominantly affects the residues that line these pockets, and the interactions of these residues with peptide residues constitute the sequence-dependent bonding that confers a particular sequence "motif" on the range of peptides that can bind any given MHC molecule.

Class I Biosynthesis

(Fig. 315-3*A*) The biosynthesis of the classic MHC class I molecules reflects their role in presenting endogenous peptides. The heavy chain is cotranslationally inserted into the membrane of the endoplasmic reticulum (ER), where it becomes glycosylated and associates sequentially with the chaperone proteins calnexin and ERp57. It then forms a complex with β_2-microglobulin, and this complex associates with the chaperone calreticulin and the MHC-encoded molecule tapasin, which physically links the class I complex to TAP, the MHC-encoded transporter associated with antigen processing. Meanwhile, peptides generated within the cytosol from intracellular proteins by the multisubunit, multicatalytic proteasome complex are actively transported into the ER by TAP, where they are trimmed by a peptidase known as *ERAAP* (ER aminopeptidase associated with antigen processing). At this point, peptides with appropriate sequence complementarity bind specific class I molecules to form complete, folded heavy chain–β_2-microglobulin–peptide trimer complexes. These are transported rapidly from the ER, through the *cis*- and *trans*-Golgi where the N-linked oligosaccharide is further processed, and thence to the cell surface.

Figure 315-3

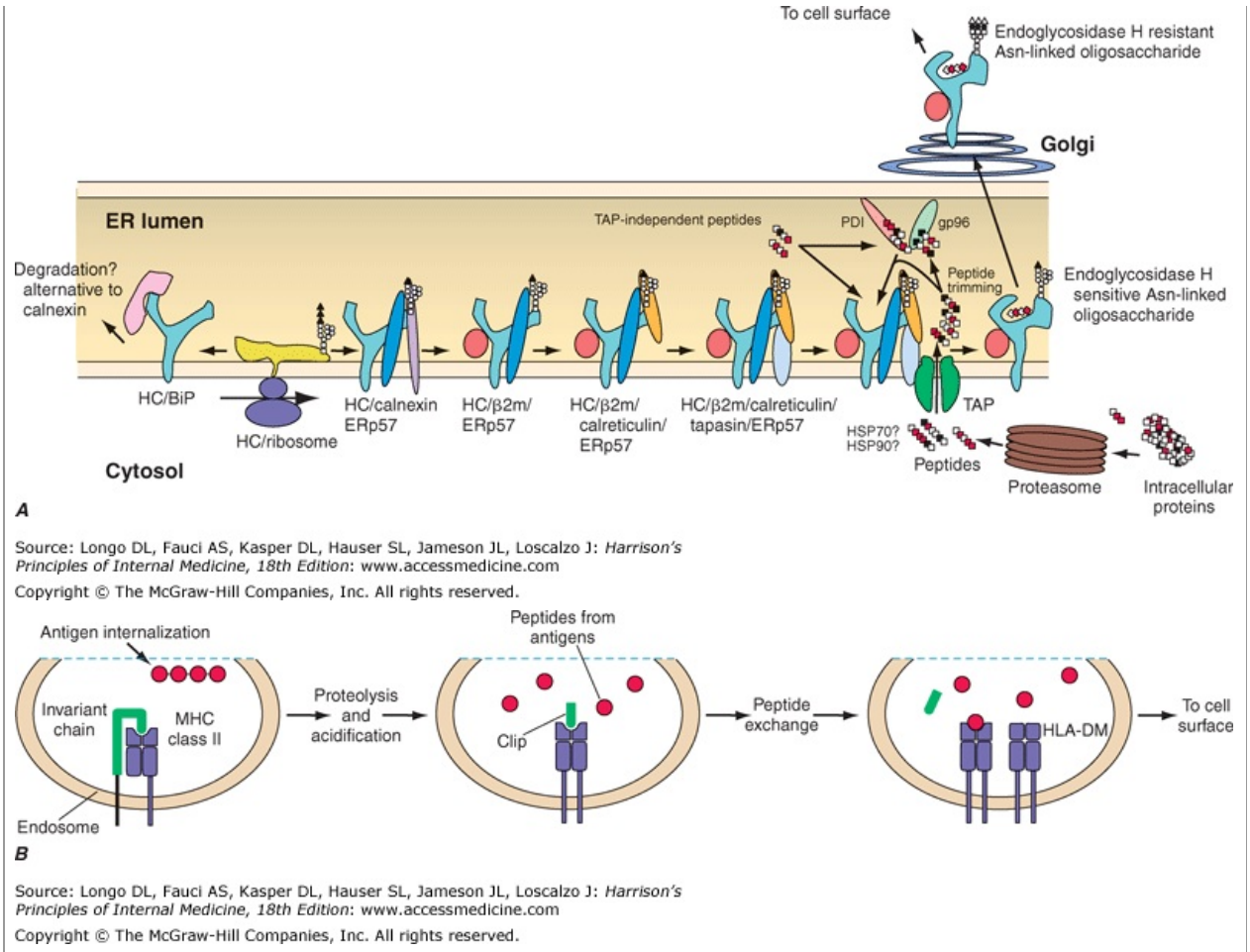

Biosynthesis of class I (*A*) and class II (*B*) molecules. *A.* Nascent heavy chain (HC) becomes associated with β2-microglobulin (β2m) and peptide through interactions with a series of chaperones. Peptides generated by the proteasome are transported into the endoplasmic reticulum (ER) by TAP. Peptides undergo N-terminal trimming in the ER and become associated with chaperones, including gp96 and PDI. Once peptide binds to HC-β2m, the HC-β2m-peptide trimeric complex exits the ER and is transported by the secretory pathway to the cell surface. In the Golgi, the N-linked oligosaccharide undergoes maturation, with addition of sialic acid residues. Molecules are not necessarily drawn to scale. ***B.*** Pathway of HLA class II molecule assembly and antigen processing. After transport through the Golgi and post-Golgi compartment, the class II–invariant chain complex moves to an acidic endosome, where the invariant chain is proteolytically cleaved into fragments and displaced by antigenic peptides, facilitated by interactions with the DMA-DMB chaperone protein. This class II molecule–peptide complex is then transported to the cell surface.

Most of the peptides transported by TAP are produced in the cytosol by proteolytic cleavage of intracellular proteins by the multisubunit, multicatalytic proteasome, and inhibitors of the proteasome dramatically reduce expression of class I–presented antigenic peptides. A thiol-dependent oxidoreductase Erp57, which mediates disulfide bond rearrangements, also appears to play an important role in folding the class I–peptide complex into a stable multicomponent molecule. The MHC-encoded proteasome subunits LMP2 and LMP7 may influence the spectrum of peptides produced but are not essential for proteasome function.

Class I Function

PEPTIDE ANTIGEN PRESENTATION

On any given cell, a class I molecule occurs in 100,000–200,000 copies and binds several hundred to several thousand distinct peptide species. The vast majority of these peptides are self-peptides to which the host immune system is tolerant by one or more of the mechanisms that maintain tolerance [e.g., clonal deletion in the thymus or clonal anergy or clonal ignorance in the periphery (Chaps. 314 and 318)]. However, class I molecules bearing foreign peptides expressed in a permissive immunologic context activate CD8 T cells, which, if naïve, will then differentiate into cytolytic T lymphocytes (CTLs). These T cells and their progeny, through their αβ TCRs, are then capable of Fas/CD95- and/or perforin-mediated cytotoxicity and/or cytokine secretion (Chap. 314) upon further encounter with the class I–peptide combination that originally activated it, and also with other combinations of class I molecule plus peptide that present a similar immunochemical stimulus to the TCR. As alluded to above, this phenomenon by which T cells recognize foreign antigens in the context of specific MHCalleles is termed *MHC restriction*, and the specific MHC molecule is termed the *restriction element*. The most common source of foreign peptides presented by class I molecules is viral infection, in the course of

which peptides from viral proteins enter the class I pathway. The generation of a strong CTL response that destroys virally infected cells represents an important antigen-specific defense against many viral infections (Chap. 314). In the case of some viral infections—hepatitis B, for example—CTL-induced target cell apoptosis is thought to be a more important mechanism of tissue damage than any direct cytopathic effect of the virus itself. The importance of the class I pathway in the defense against viral infection is underscored by the identification of a number of viral products that interfere with the normal class I biosynthetic pathway and thus block the immunogenetic expression of viral antigens.

Other examples of intracellularly generated peptides that can be presented by class I molecules in an immunogenic manner include peptides derived from nonviral intracellular infectious agents (e.g., *Listeria*, *Plasmodium*), tumor antigens, minor histocompatibility antigens, and certain autoantigens. There are also situations in which cell surface–expressed class I molecules are thought to acquire and present exogenously derived peptides.

HLA CLASS I RECEPTORS AND NK CELL RECOGNITION

(Chap. 314) NK cells, which play an important role in innate immune responses, are activated to cytotoxicity and cytokine secretion by contact with cells that lack MHC class I expression, and NK cell activation is inhibited by cells that express MHC class I. In humans, the recognition of class I molecules by NK cells is carried out by three classes of receptor families, the killer cell–inhibitory cell receptor (KIR) family, the leukocyte Ig-like receptor (LIR) family, and the CD94/NKG2 family. The KIR family, also called CD158, is encoded on chromosome 19q13.4. KIR gene nomenclature is based on the number of domains (2D or 3D) and the presence of long (L) or short (S) cytoplasmic domains. The KIR2DL1 and S1 molecules primarily recognize alleles of HLA-C, which possess a lysine at position 80 (HLA-Cw2, -4, -5 and -6), while the KIR2DL2/S2 and KIR2DL3/S3 families primarily recognize alleles of HLA-C with asparagine at this position (HLA-Cw1,-3, -7 and -8). The KIR3D L1 and S1 molecules predominantly recognize HLA-B alleles that fall into the HLA-Bw4 class determined by residues 77–83 in the α_1 domain of the heavy chain, while the KIR3DL2 molecule is an inhibitory receptor for HLA-A*03. One of the KIR products, KIR2DL4, is known to be an activating receptor for HLA-G. The most common KIR haplotype in whites contains one activating KIR and six inhibitory KIR genes, although there is a great deal of diversity in the population, with >100 different combinations. It appears that most individuals have at least one inhibitory KIR for a self-HLA class I molecule, providing a structural basis for NK cell target specificity, which helps prevent NK cells from attacking normal cells. The importance of KIR-HLA interactions to many immune responses is illustrated by studies associating KIR3DL1 or S1 with multiple sclerosis (Chapter 380), an autoimmune disease, but also with partial protection against HIV (Chap. 189); in both cases consistent with a role for HLA-KIR mediated NK activation.

The LIR gene family (CD85, also called ILT) is encoded centromeric of the KIR locus on 19q13.4, and it encodes a variety of inhibitory immunoglobulin-like receptors expressed on many lymphocyte and other hematopoietic lineages. Interaction of LIR-1 (ILT2) with NK or T cells inhibits activation and cytotoxicity, mediated by many different HLA class I molecules, including HLA-G. HLA-F also appears to interact with LIR molecules, although the functional context for this is not understood.

The third family of NK receptors for HLA is encoded in the NK complex on chromosome 12p12.3-13.1 and consists of CD94 and five NKG2 genes, A/B, C, E/H, D, and F. These molecules are C-type (calcium-binding) lectins, and most function as disulfide-bonded heterodimers between CD94 and one of the NKG2 glycoproteins. The principal ligand of CD94/NKG2A receptors is the HLA-E molecule, complexed to a peptide derived from the signal sequence of classic HLA class I molecules and HLA-G. Thus, analogous to the way in which KIR receptors recognize HLA-C, the NKG2 receptor monitors self–class I expression, albeit indirectly through peptide recognition in the context of HLA-E. NKG2C, -E, and -H appear to have similar specificities but act as activating receptors. NKG2D is expressed as a homodimer and functions as an activating receptor expressed on NK cells, $\gamma\delta$ TCR T cells, and activated CD8 T cells. When complexed with an adaptor called DAP10, NKG2D recognizes MIC-A and MIC-B molecules and activates the cytolytic response. NKG2D also binds a class of molecules known as *ULBP*, structurally related to class I molecules but not encoded in the MHC. **The function of NK cells in immune responses is discussed in Chap. 314**.

Class II Structure

(Fig. 315-2*C*) A specialized functional architecture similar to that of the class I molecules can be seen in the example of a class II molecule depicted in Fig. 315-2*C*, with an antigen-binding cleft arrayed above a supporting scaffold that extends the cleft toward the external cellular environment. However, in contrast to the HLA class I molecular structure, β_2-microglobulin is not associated with class II molecules. Rather, the class II molecule is a heterodimer, composed of a 29-kD α chain and a 34-kD β chain. The amino-terminal domains of each chain form the antigen-binding elements that, like the class I molecule, cradle a bound peptide in a groove bounded by extended α-helical loops, one encoded by the A (α chain) gene and one by the B (β chain) gene. Like the class I groove, the class II antigen-binding groove is punctuated by pockets that contact the side chains of amino acid residues of the bound peptide, but unlike the class I groove, it is open at both ends. Therefore, peptides bound by

class II molecules vary greatly in length, since both the N- and C-terminal ends of the peptides can extend through the open ends of this groove. Approximately 11 amino acids within the bound peptide form intimate contacts with the class II molecule itself, with backbone hydrogen bonds and specific side chain interactions combining to provide, respectively, stability and specificity to the binding **(Fig. 315-4)**.

Figure 315-4

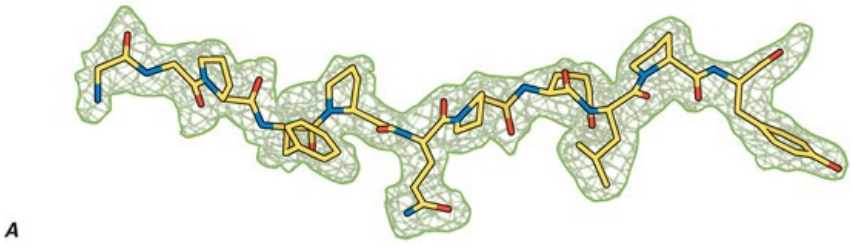

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

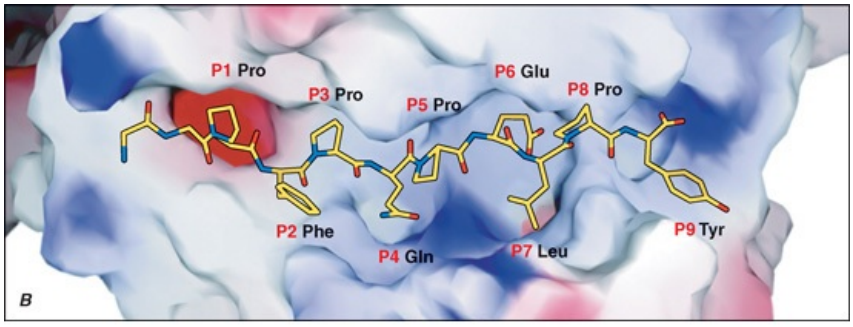

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

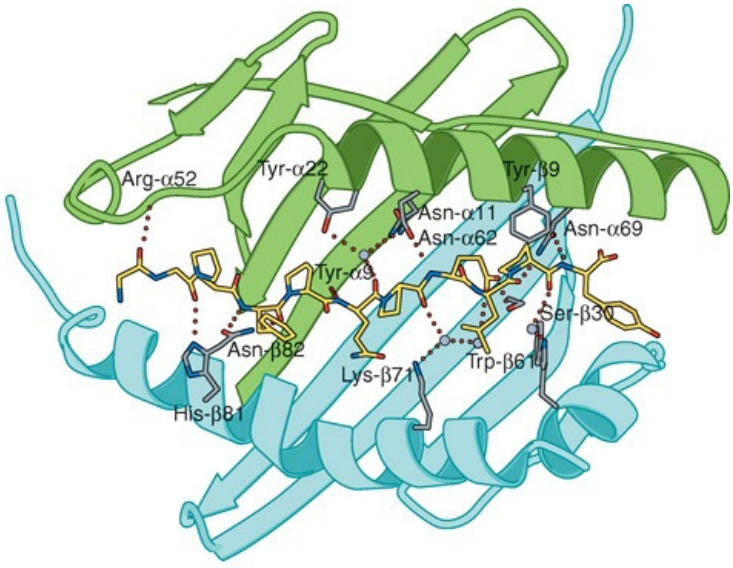

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Specific intermolecular interactions determine peptide binding to MHC class II molecules. A short peptide sequence derived from alpha-gliadin (***A.***) is accommodated within the MHC class II binding groove by specific interactions between peptide side chains (the P1–P9 residues illustrated in (***B.***) and corresponding pockets in the MHC class II structure. The latter are determined by the genetic polymorphisms of the MHC gene, in this case encoding an HLA-DQ2 molecule *(**C.**)*. This shows the extensive hydrogen bond and salt bridge network, which tightly constrains the pMHC complex and presents the complex of antigen and restriction element for CD4 T cell recognition. *(FromKim C et al: Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proc Natl Acad Sci USA 101:4175, 2004.)*

The genetic polymorphisms that distinguish different class II genes correspond to changes in the amino acid composition of the class II molecule, and these variable sites are clustered predominantly

around the pocket structures within the antigen-binding groove. As with class I, this is a critically important feature of the class II molecule, which explains how genetically different individuals have functionally different HLA molecules.

Biosynthesis and Function of Class II Molecules

(Fig. 315-3*B*) The intracellular assembly of class II molecules occurs within a specialized compartmentalized pathway that differs dramatically from the class I pathway described above. As illustrated in Fig. 315-3*B*, the class II molecule assembles in the ER in association with a chaperone molecule, known as the *invariant chain*. The invariant chain performs at least two roles. First, it binds to the class II molecule and blocks the peptide-binding groove, thus preventing antigenic peptides from binding. This role of the invariant chain appears to account for one of the important differences between class I and class II MHC pathways, since it can explain why class I molecules present endogenous peptides from proteins newly synthesized in the ER but class II molecules generally do not. Second, the invariant chain contains molecular localization signals that direct the class II molecule to traffic into post-Golgi compartments known as *endosomes*, which develop into specialized acidic compartments where proteases cleave the invariant chain, and antigenic peptides can now occupy the class II groove. The specificity and tissue distribution of these proteases appear to be an important way in which the immune system regulates access to the peptide-binding groove and T cells become exposed to specific self-antigens. Differences in protease expression in the thymus and in the periphery may in part determine which specific peptide sequences comprise the peripheral repertoire for T cell recognition. It is at this stage in the intracellular pathway, after cleavage of the invariant chain, that the MHC-encoded DM molecule catalytically facilitates the exchange of peptides within the class II groove to help optimize the specificity and stability of the MHC-peptide complex.

Once this MHC-peptide complex is deposited in the outer cell membrane it becomes the target for T cell recognition via a specific TCR expressed on lymphocytes. Because the endosome environment contains internalized proteins retrieved from the extracellular environment, the class II–peptide complex often contains bound antigens that were originally derived from extracellular proteins. In this way, the class II peptide–loading pathway provides a mechanism for immune surveillance of the extracellular space. This appears to be an important feature that permits the class II molecule to bind foreign peptides, distinct from the endogenous pathway of class I–mediated presentation.

Role of HLA in Transplantation

The development of modern clinical transplantation in the decades since the 1950s provided a major impetus for elucidation of the HLA system, as allograft survival is highest when donor and recipient are HLA-identical. Although many molecular events participate in transplantation rejection, allogeneic differences at class I and class II loci play a major role. Class I molecules can promote T cell responses in several different ways. In the cases of allografts in which the host and donor are mismatched at one or more class I loci, host T cells can be activated by classic *direct alloreactivity*, in which the antigen receptors on the host T cells react with the foreign class I molecule expressed on the allograft. In this situation, the response of any given TCR may be dominated by the allogeneic MHC molecule, the peptide bound to it, or some combination of the two. Another type of host antigraft T cell response involves the uptake and processing of donor MHC antigens by host antigen-presenting cells and the subsequent presentation of the resulting peptides by host MHC molecules. This mechanism is termed *indirect alloreactivity*.

In the case of class I molecules on allografts that are shared by the host and the donor, a host T cell response may still be triggered because of peptides that are presented by the class I molecules of the graft but not of the host. The most common basis for the existence of these endogenous antigen peptides, called *minor histocompatibility antigens* , is a genetic difference between donor and host at a non-MHC locus encoding the structural gene for the protein from which the peptide is derived. These loci are termed *minor histocompatibility loci* , and nonidentical individuals typically differ at many such loci. CD4 T cells react to analogous class II variation, both direct and indirect, and class II differences alone are sufficient to drive allograft rejection.

Association of HLA Alleles with Susceptibility to Disease

It has long been postulated that infectious agents provide the driving force for the allelic diversification seen in the HLA system. An important corollary of this hypothesis is that resistance to specific pathogens may differ between individuals, based on HLA genotype. Observations of specific HLA genes associated with resistance to malaria or dengue fever, persistence of hepatitis B, and to disease progression in HIV infection are consistent with this model. For example, failure to clear persistent hepatitis B or C viral infection may reflect the inability of particular HLA molecules to present viral antigens effectively to T cells. Similarly, both protective and susceptible HLA allelic associations have been described for human papilloma virus–associated cervical neoplasia, implicating the MHC as an influence in mediating viral clearance in this form of cancer.

Pathogen diversity is probably also the major selective pressure favoring HLA heterozygosity. The extraordinary scope of HLA allelic diversity increases the likelihood that most new pathogens will be recognized by some HLA molecules, helping to ensure immune fitness to the host. However, another

consequence of diversification is that some alleles may become capable of recognition of "innocent bystander" molecules, including drugs, environmental molecules, and tissue-derived self-antigens. In a few instances, single HLA alleles display a strong selectivity for binding of a particular agent that accounts for a genetically determined response: hypersensitivity to abacavir, an antiretroviral therapeutic, is directly linked to binding of abacavir in the antigen-binding pockets of HLA-B*5701, and chronic beryllium toxicity is linked to binding of beryllium by HLA-DP molecules with a specific glutamic acid polymorphic residue on the class II beta chain. Even in the case of more complex diseases, particular HLA alleles are strongly associated with certain inappropriate immune-mediated disease states, particularly for some common autoimmune disorders (Chap. 318). By comparing allele frequencies in patients with any particular disease and in control populations, >100 such associations have been identified, some of which are listed in **Table 315-1**. The strength of genetic association is reflected in the term *relative risk*, which is a statistical odds ratio representing the risk of disease for an individual carrying a particular genetic marker compared with the risk for individuals in that population without that marker. The nomenclature shown in Table 315-1 reflects both the HLA serotype (e.g., DR3, DR4) and the HLA genotype (e.g., DRB1*0301, DRB1*0401). It very likely the class I and class II alleles themselves are the true susceptibility alleles for most of these associations. However, because of the extremely strong linkage disequilibrium between the DR and DQ loci, in some cases it has been difficult to determine the specific locus or combination of class II loci involved. In some cases, the susceptibility gene may be one of the HLA-linked genes located near the class I or class II region, but not the HLA gene itself, and in other cases the susceptibility gene may be a non-HLA gene such as TNF-α, which is nearby. Indeed, since linkage disequilibrium of some haplotypes extends across large segments of the MHC region, it is quite possible that combinations of genes may account for the particular associations of HLA haplotypes with disease. For example, on some haplotypes associated with rheumatoid arthritis, both HLA-DRB1 alleles and a particular polymorphism associated with the TNF locus may be contributory to disease risk. Other candidates for similar epistatic effects include the IKBL gene and the MICA locus, potentially in combination with classic HLA class II risk alleles.

Table 315–1. Significant HLA Class I and Class II Associations with Disease

	Marker	Gene	Strength of Association
Spondyloarthropathies			
Ankylosing spondylitis	B27	B*2702, –04, –05	++++
Reiter's syndrome	B27		++++
Acute anterior uveitis	B27		+++
Reactive arthritis *(Yersinia, Salmonella, Shigella, Chlamydia)*	B27		+++
Psoriatic spondylitis	B27		+++
Collagen-Vascular Diseases			
Juvenile arthritis, pauciarticular	DR8 DR5		++ ++
Rheumatoid arthritis	DR4	DRB1*0401, –04, –05	+++
Sjögren's syndrome	DR3		++
Systemic lupus erythematosus			
White	DR3		+
Japanese	DR2		++
Autoimmune Gut and Skin			
Gluten-sensitive enteropathy (celiac disease)	DQ2	DQA1*0501 DQB1*0201	+++
Chronic active hepatitis	DR3		++
Dermatitis herpetiformis	DR3		+++
Psoriasis vulgaris	Cw6		++
Pemphigus vulgaris	DR4	DRB1*0402	+++
	DQ1	DQB1*0503	
Bullous pemphigoid variant	DQ7	DQB1*0301	+
Autoimmune Endocrine			
Type 1 diabetes mellitus	DQ8	DQB1*0302	+++
	DR4	DRB1*0401, –04	

	DR3		++
	DR2	DQB1 *0602	__a
Hyperthyroidism (Graves')	B8 DR3		+ +
Hyperthyroidism (Japanese)	B35		+
Adrenal insufficiency	DR3		++
Autoimmune Neurologic			
Myasthenia gravis	B8		+
	DR3		+
Multiple sclerosis	DR2	DRB1*1501	++
		DRB5*0101	
Other			
Behçet's disease	B51		++
Congenital adrenal hyperplasia	B47	21·OH (Cyp21B)	+++
Narcolepsy	DR2	DQB1*0602	++++
Goodpasture's syndrome (anti-GBM)	DR2		++
Abacavir hypersensitivity	B57	B*5701	++++

[a]Strong negative association; i.e., genetic association with protection from diabetes.

As might be predicted from the known function of the class I and class II gene products, almost all of the diseases associated with specific HLAalleles have an immunologic component to their pathogenesis. The recent development of soluble HLA-peptide recombinant molecules as biological probes of T cell function, often in multivalent complexes referred to as "MHC tetramers," represents an opportunity to use HLA genetic associations to develop biomarkers for detection of early disease progression. However, it should be stressed that even the strong HLA associations with disease (those associations with relative risk of ≳10) implicate normal, rather than defective, alleles. Most individuals who carry these susceptibility genes do not express the associated disease; in this way, the particular HLA gene is permissive for disease but requires other environmental (e.g., the presence of specific antigens) or genetic factors for full penetrance. In each case studied, even in diseases with very strong HLA associations, the concordance of disease in monozygotic twins is higher than in HLA-identical dizygotic twins or other sibling pairs, indicating that non-HLA genes contribute to susceptibility and can significantly modify the risk attributable to HLA.

Another group of diseases is genetically linked to HLA, not because of the immunologic function of HLA alleles but rather because they are caused by autosomal dominant or recessive abnormal alleles at loci that happen to reside in or near the HLA region. Examples of these are 21-hydroxylase deficiency (Chap. 342), hemochromatosis (Chap. 357), and spinocerebellar ataxia (Chap. 374).

Class I Associations with Disease

Although the associations of human disease with particular HLAalleles or haplotypes predominantly involve the class II region, there are also several prominent disease associations with class I alleles. These include the association of Behçet's disease (Chap. 327) with HLA-B51, psoriasis vulgaris (Chap. 52) with HLA-Cw6, and, most notably, the spondyloarthritides (Chap. 325) with HLA-B27. Twenty-five HLA-B locus alleles, designated HLA-B*2701–B*2725, encode the family of B27 class I molecules. All of the subtypes share a common B pocket in the peptide-binding groove—a deep, negatively charged pocket that shows a strong preference for binding the arginine side chain. In addition, B27 is among the most negatively charged of HLA class I heavy chains, and the overall preference is for positively charged peptides. HLA-B*2705 is the predominant subtype in whites and most other non-Asian populations, and this subtype is very highly associated with ankylosing spondylitis (AS) (Chap. 325), both in its idiopathic form and in association with chronic inflammatory bowel disease or psoriasis vulgaris. It is also associated with reactive arthritis (ReA) (Chap. 325), with other idiopathic forms of peripheral arthritis (undifferentiated spondyloarthropathy), and with recurrent acute anterior uveitis. B27 is found in 50–90% of individuals with these conditions, compared with a prevalence of ~7% in North American whites.

It can be concluded that the B27 molecule itself is involved in disease pathogenesis, based on strong evidence from clinical epidemiology and on the occurrence of a spondyloarthropathy-like disease in HLA-B27 transgenic rats. The association of B27 with these diseases may derive from the specificity of a particular peptide or family of peptides bound to B27 or through another mechanism that is independent of the peptide specificity of B27. In particular, HLA-B27 has been shown to form heavy chain homodimers, utilizing the cysteine residue at position 67 of the B57 α chain, in the absence of β_2-microglobulin. These homodimers are expressed on the surface of lymphocytes and monocytes from

patients with AS, and receptors including KIR3DL1, KIR3DL2, and ILT4 are capable of binding to them, promoting the activation and survival of cells expressing these receptors. Alternatively, this dimerization "misfolding" of B27 may initiate an intracellular stress signalling response, called the unfolded protein response (UPR), capable of modulating immune cell function. Whether these interactions contribute to disease susceptibility or pathogenesis is currently unknown.

Class II Disease Associations

As can be seen in Table 315-1, the majority of associations of HLA and disease are with class II alleles. Several diseases have complex HLA genetic associations.

CELIAC DISEASE

In the case of celiac disease (Chap. 294), it is probable that the HLA-DQ genes are the primary basis for the disease association. HLA-DQ genes present on both the celiac-associated DR3 and DR7 haplotypes include the *DQB1*0201*gene, and further detailed studies have documented a specific class II αβ dimer encoded by the *DQA1*0501* and *DQB1*0201* genes, which appears to account for most of the HLA genetic contribution to celiac disease susceptibility. This specific HLA association with celiac disease may have a straightforward explanation: peptides derived from the wheat gluten component gliaden are bound to the molecule encoded by *DQA1*0501* and *DQB1*0201* and presented to T cells. Gliaden-derived peptides that are implicated in this immune activation bind the DQ class II dimer best when the peptide contains a glutamine to glutamic acid substitution. It has been proposed that tissue transglutaminase, an enzyme present at increased levels in the intestinal cells of celiac patients, converts glutamine to glutamic acid in gliadin, creating peptides that are capable of being bound by the DQ2 molecule and presented to T cells.

PEMPHIGUS VULGARIS

In the case of pemphigus vulgaris (Chap. 54), there are two HLAgenes associated with disease, *DRB1*0402* and *DQB1*0503*. Peptides derived from desmoglien3, an epidermal autoantigen, bind to the DRB1*0402- and DQB1*0503-encoded HLA molecules, and this combination of specific peptide binding and disease-associated class II molecule is sufficient to stimulate desmoglien-specific T cells. A bullous pemphigoid clinical variant, not involving desmoglien recognition, has been found to be associated with HLA-DQB1*0301.

JUVENILE ARTHRITIS

Pauciarticular juvenile arthritis (Chap. 321) is an autoimmune disease associated with genes at the DRB1 locus and also with genes at the DPB1 locus. Patients with both DPB1*0201 and a DRB1 susceptibility allele (usually DRB1*08 or -*05) have a higher relative risk than expected from the additive effect of those genes alone. In juvenile patients with rheumatoid factor–positive polyarticular disease, heterozygotes carrying both DRB1*0401 and -*0404 have a relative risk > 100, reflecting an apparent synergy in individuals inheriting both of these susceptibility genes.

TYPE 1 DIABETES MELLITUS

Type 1 (autoimmune) diabetes mellitus (Chap. 344) is associated with MHCgenes on more than one haplotype. The presence of both the DR3 and DR4 haplotypes in one individual confers a twentyfold increased risk for type 1 diabetes; the strongest single association is with DQB1*0302, and all haplotypes that carry a *DQB1*0302*gene are associated with type 1 diabetes, whereas related haplotypes that carry a different DQB1 gene are not. However, the relative risk associated with inheritance of this gene can be modified, depending on other HLA genes present either on the same or a second haplotype. For example, the presence of a DR2-positive haplotype containing a *DQB1*0602* gene is associated with decreased risk. This gene, *DQB1*0602*, is considered "protective" for type 1 diabetes. Even some DRB1 genes that can occur on the same haplotype as DQB1*0302 may modulate risk, so that individuals with the DR4 haplotype that contains DRB1*0403 are less susceptible to type 1 diabetes than individuals with other DR4-DQB1*0302 haplotypes.

Although the presence of a DR3 haplotype in combination with the DR4-DQB1*0302 haplotype is a very high-risk combination for diabetes susceptibility, the specific gene on the DR3 haplotype that is responsible for this synergy is not yet identified. There are some characteristic structural features of the diabetes-associated DQ molecule encoded by DQB1*0302, particularly the capability for binding peptides that have negatively charged amino acids near their C-termini. This may indicate a role for specific antigenic peptides or T cell interactions in the immune response to islet-associated proteins.

HLA AND RHEUMATOID ARTHRITIS

The HLAgenes associated with rheumatoid arthritis (RA) (Chap. 321) encode a distinctive sequence of amino acids from codons 67–74 of the DRβ molecule: RA-associated class II molecules carry the sequence LeuLeuGluGlnArgArgAlaAla or LeuLeuGluGlnLysArgAlaAla in this region, while non-RA-associated genes carry one or more differences in this region. These residues form a portion of the molecule that lies in the middle of the α-helical portion of the DRB1-encoded class II molecule, termed the *shared epitope.*

The highest risk for susceptibility to RA comes in individuals who carry both a *DRB1*0401* and

*DRB1*0404*gene. These DR4-positive RA-associated alleles are most frequent among patients with more severe, erosive disease. Several mechanisms have been proposed that link the shared epitope to immune reactivity in RA. This portion of the class II molecule may allow preferential binding of an arthritogenic peptide, it may favor the expansion of a type of self-reactive T lymphocyte, or it may itself form part of the pMHC ligand recognized by TCR that initiates synovial tissue recognition.

Molecular Mechanisms for HLA-Disease Associations

As noted above, HLA molecules play a key role in the selection and establishment of the antigen-specific T cell repertoire and a major role in the subsequent activation of those T cells during the initiation of an immune response. Precise genetic polymorphisms characteristic of individual alleles dictate the specificity of these interactions and thereby instruct and guide antigen-specific immune events. These same genetically determined pathways are therefore implicated in disease pathogenesis when specific HLA genes are responsible for autoimmune disease susceptibility.

The fate of developing T cells within the thymus is determined by the affinity of interaction between T cell receptor and HLA molecules bearing self-peptides, and thus the particular HLA types of each individual control the precise specificity of the T cell repertoire (Chap. 314). The primary basis for HLA-associated disease susceptibility may well lie within this thymic maturation pathway. The positive selection of potentially autoreactive T cells, based on the presence of specific HLA susceptibility genes, may establish the threshold for disease risk in a particular individual.

At the time of onset of a subsequent immune response, the primary role of the HLA molecule is to bind peptide and present it to antigen-specific T cells. The HLA complex can therefore be viewed as encoding genetic determinants of precise immunologic activation events. Antigenic peptides that bind particular HLA molecules are capable of stimulating T cell immune responses; peptides that do not bind are not presented to T cells and are not immunogenic. This genetic control of the immune response is mediated by the polymorphic sites within the HLA antigen–binding groove that interact with the bound peptides. In autoimmune and immune-mediated diseases, it is likely that specific tissue antigens that are targets for pathogenic lymphocytes are complexed with the HLA molecules encoded by specific susceptibility alleles. In autoimmune diseases with an infectious etiology, it is likely that immune responses to peptides derived from the initiating pathogen are bound and presented by particular HLA molecules to activate T lymphocytes that play a triggering or contributory role in disease pathogenesis. The concept that early events in disease initiation are triggered by specific HLA-peptide complexes offers some prospects for therapeutic intervention, since it may be possible to design compounds that interfere with the formation or function of specific HLA-peptide–T cell receptor interactions.

When considering mechanisms of HLA associations with immune response and disease, it is well to remember that just as HLA genetics are complex, so are the mechanisms likely to be heterogeneous. Immune-mediated disease is a multistep process in which one of the HLA-associated functions is to establish a repertoire of potentially reactive T cells, while another HLA-associated function is to provide the essential peptide-binding specificity for T cell recognition. For diseases with multiple HLA genetic associations, it is possible that both of these interactions occur and synergize to advance an accelerated pathway of disease.

FURTHER READINGS

Caillat-Zucman S: Molecular mechanisms of HLA association with autoimmune diseases. Tissue Antigens 73:1, 2008[PMID: 19017300] [Full Text]

Carrington M et al: KIR-HLA intercourse in HIV disease. Trends Immunol 16:620, 2008[PMID: 18976921] [Full Text]

Horton R et al: Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project. Immunogenetics 60:1, 2008[PMID: 18193213] [Full Text]

Jones EY et al: MHC class II proteins and disease: A structural perspective. Nat Rev Immunol 6:271, 2006[PMID: 16557259] [Full Text]

Nepom GT: Major histocompatibility complex-directed susceptibility to rheumatoid arthritis. Adv Immunol 68:315, 1998[PMID: 9505093] [Full Text]

———: Tetramer analysis of human autoreactive CD4-positive T cells. Adv Immunol 88:51, 2005

Shiina T et al: The HLA genomic loci map: Expression, interaction, diversity and disease. J Hum Genet 54:15, 2009[PMID: 19158813] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRIMARY IMMUNE DEFICIENCY DISEASES: INTRODUCTION

Immunity is intrinsic to life and an important tool in the fight for survival against pathogenic microorganisms. The human immune system can be divided into two major components: the innate immune system and the adaptive immune system (Chap. 314). The innate immune system provides the rapid triggering of inflammatory responses based on the recognition (at the cell surface or within cells) of either molecules expressed by microorganisms or molecules that serve as "danger signals" released by cells under attack. These receptor/ligand interactions trigger signaling events that ultimately lead to inflammation. Virtually all cell lineages (not just immune cells) are involved in innate immune responses; however, myeloid cells (i.e., neutrophils and macrophages) play a major role because of their phagocytic capacity. The adaptive immune system operates by clonal recognition of antigens followed by a dramatic expansion of antigen-reactive cells and execution of an immune effector program. Most of the effector cells die off rapidly, whereas memory cells persist. Although both T and B lymphocytes recognize distinct chemical moieties and execute distinct adaptive immune responses, the latter is largely dependent on the former in generating long-lived humoral immunity. Adaptive responses utilize components of the innate immune system; for example, the antigen-presentation capabilities of dendritic cells help to determine the type of effector response. Not surprisingly, immune responses are controlled by a series of regulatory mechanisms.

Hundreds of gene products have been characterized as effectors or mediators of the immune system (Chap. 314). Whenever the expression or function of one of these products is genetically impaired (provided the function is nonredundant), a primary immunodeficiency (PID) occurs.

Primary immunodeficiencies are genetic diseases with primarily Mendelian inheritance. More than 200 conditions have now been described and deleterious mutations in approximately 150 genes have been identified. The overall prevalence of PIDs has been estimated in various countries at 5 per 100,000 individuals; however, given the difficulty in diagnosing these rare and complex diseases, this figure is probably an underestimate. Primary immunodeficiencies can involve all possible aspects of immune responses, from innate through adaptive, cell differentiation, and effector function and regulation. For the sake of clarity, PIDs should be classified according to (1) the arm of the immune system that is defective and (2) the mechanism of the defect (when known). **Table 316-1** classifies the most prevalent PIDs according to this manner of classification; however, one should bear in mind that the classification of PIDs sometimes involves arbitrary decisions because of overlap and, in some cases, lack of data.

Table 316–1. Classification of Primary Immune Deficiency Diseases

Deficiencies of the Innate Immune System	
• Phagocytic cells:	
- Impaired production: severe congenital neutropenia (SCN)	
- Asplenia	
- Impaired adhesion: leukocyte adhesion deficiency (LAD)	
- Impaired killing: chronic granulomatous disease (CGD)	
• Innate immunity receptors and signal transduction:	
- Defects in Toll-like receptor signaling	
- Mendelian susceptibility to mycobacterial disease	
• Complement deficiencies:	
- Classical, alternative, and lectin pathways	
- Lytic phase	
Deficiencies of the Adaptive Immune System	
• T lymphocytes:	
- Impaired development	Severe combined immune deficiencies (SCIDs) DiGeorge syndrome
- Impaired survival, migration, function	Severe combined immunodeficiencies Hyper-IgE syndrome (autosomal dominant)

	CD40 ligand deficiency Wiskott-Aldrich syndrome Ataxia-telangiectasia and other DNA repair deficiencies
• B lymphocytes:	
- Impaired development	XL and AR agammaglobulinemia
- Impaired function	Hyper-IgM syndrome Common variable immunodeficiency (CVID) IgA deficiency
Regulatory Defects	
• Innate immunity	Autoinflammatory syndromes (outside the scope of this chapter) Severe colitis
• Adaptive immunity	Hemophagocytic lymphohistiocytosis (HLH) Autoimmune lymphoproliferation syndrome (ALPS) Autoimmunity and inflammatory diseases (IPEX, APECED)

Abbreviations: APECED, autoimmune polyendocrinopathy candidiasis ectodermal dysplasia; AR, autosomal recessive; IPEX, immunodysregulation polyendocrinopathy enteropathy X-linked syndrome; XL, X-linked.

The consequences of PIDs vary widely as a function of the molecules that are defective. This concept translates into multiple levels of vulnerability to infection by pathogenic and opportunistic microorganisms, ranging from extremely broad [as in severe combined immunodeficiency (SCID)] to narrowly restricted to a single microorganism [as in Mendelian susceptibility to mycobacterial disease (MSMD)]. The locations of the sites of infection and the causal microorganisms involved will thus help physicians arrive at proper diagnoses. Primary immunodeficiencies can also lead to immunopathologic responses such as allergy (as in Wiskott-Aldrich syndrome), lymphoproliferation, and autoimmunity. A combination of recurrent infections, inflammation, and autoimmunity can be observed in a number of PIDs, thus creating obvious therapeutic challenges. Finally, some PIDs increase the risk of cancer, notably but not exclusively lymphocytic cancers, e.g., lymphoma.

DIAGNOSIS OF PRIMARY IMMUNODEFICIENCIES

The most frequent symptom prompting the diagnosis of a PID is the presence of recurrent or unusually severe infections. As mentioned above, recurrent allergic or autoimmune manifestations may also alert the physician to a possible diagnosis of PID. In such cases, a detailed account of the subject's personal and family medical history should be obtained. It is of the utmost importance to gather as much medical information as possible on relatives and up to several generations of ancestors. In addition to the obvious focus on primary symptoms, the clinical examination should evaluate the size of lymphoid organs and, when appropriate, look for the characteristic signs of a number of complex syndromes that may be associated with a PID.

The performance of laboratory tests should be guided to some extent by the clinical findings. Infections of the respiratory tract (bronchi, sinuses) mostly suggest a defective antibody response. In general, invasive bacterial infections can result from complement deficiencies, signaling defects of innate immune responses, asplenia, or defective antibody responses. Viral infections, recurrent *Candida* infections, and opportunistic infections are generally suggestive of impaired T cell immunity. Skin infections and deep-seated abscesses primarily reflect innate immune defects (such as chronic granulomatous disease); however, they may also appear in the autosomal dominant hyper-IgE syndrome. **Table 316-2** summarizes the laboratory tests that are most frequently used to diagnose a PID. More specific tests (notably genetic tests) are then used to make a definitive diagnosis.

Table 316–2. Test Most Frequently Used to Diagnose a Primary Immune Deficiency (PID)

Test	Information	PID Disease
• Blood cell counts and cell morphology	Neutrophil counts	↓Severe congenital neutropenia, ↑↑ LAD
	Lymphocyte counts*	T cell ID
	Eosinophilia	WAS, Hyper-IgE syndrome
	Howell-Jolly bodies	Asplenia
• Chest x-ray	Thymic shadow	SCID, DiGeorge syndrome
	Costochondral junctions	Adenosine deaminase

		deficiency
• Bone x-ray	Metaphyseal ends	Cartilage hair hypoplasia
• Immunoglobulin serum levels	IgG, IgA, IgM IgE	B cell ID Hyper-IgE syndrome, WAS, T cell ID
• Lymphocyte phenotype	T, B lymphocyte counts	T cell ID, agammaglobulinemia
• Dihydrorhodamine fluorescence (DHR) assay Nitroblue tetrazolium (NBT) assay	Reactive oxygen species production by PMN	Chronic granulomatous disease
• CH50, AP50	Classic and alternative complement pathways	Complement deficiencies
• Ultrasonography of the abdomen	Spleen size	Asplenia

*Normal counts vary with age. For example, the lymphocyte count is between 3000 and 9000/μL of blood below the age of 3 months and between 1500 and 2500/μL in adults.

Abbreviations: ID, immunodeficiency; LAD, leukocyte adhesion deficiency; PMNs, polymorphonuclear leukocytes; SCID, severe combined immunodeficiency; WAS, Wiskott-Aldrich syndrome.

The primary immunodeficiencies discussed below have been grouped together according to the affected cells and the mechanisms involved (Table 316-1, **Fig. 316-1**).

Figure 316-1

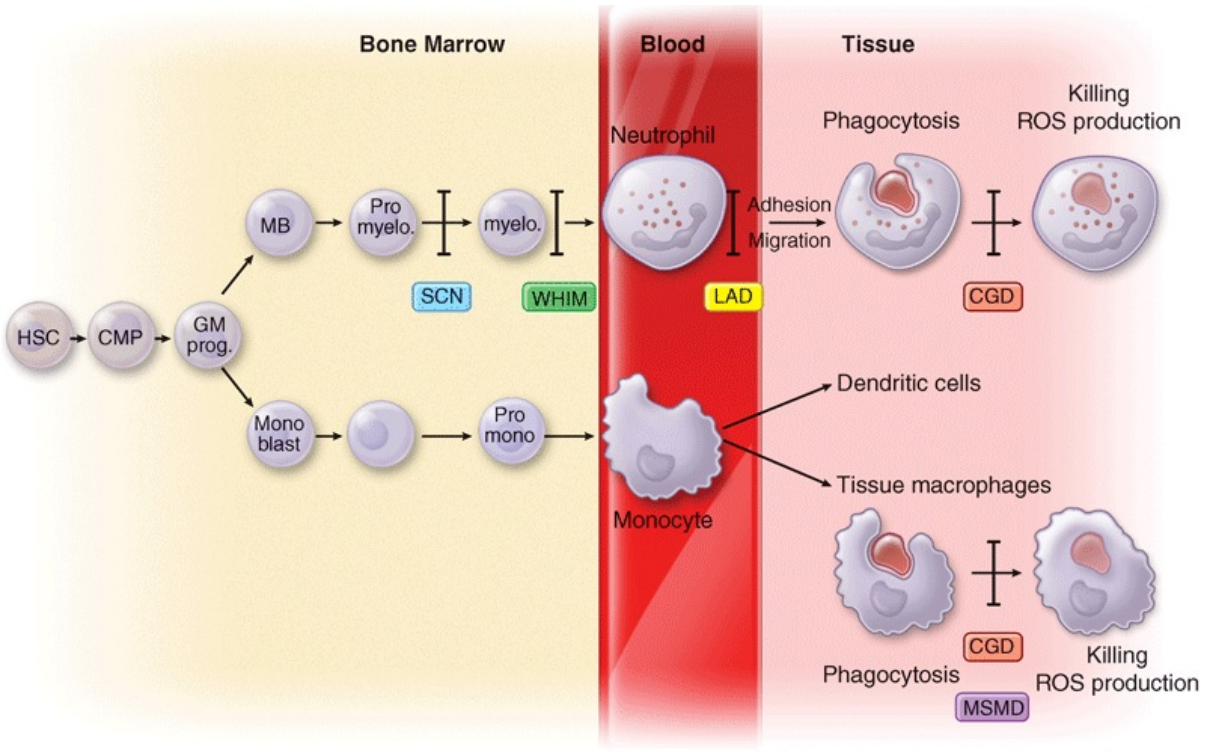

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Differentiation of phagocytic cells and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common myeloid progenitors (CMPs) and then granulocyte-monocyte progenitors (GM-prog.), which, in turn, differentiate into neutrophils (MB: myeloblasts; Promyelo: promyelocytes; myelo: myelocytes) or monocytes (monoblasts and promonocytes). Upon activation, neutrophils adhere to the vascular endothelium, transmigrate, and phagocytose the targets. Reactive oxygen species (ROS) are delivered to the microorganism-containing phagosomes. Macrophages in tissues kill using the same mechanism. Following activation by interferon γ (not shown here), macrophages can be armed to kill intracellular pathogens such as mycobacteria. For sake of simplicity, not all cell differentiation stages are shown. The abbreviations for PIDs are contained in boxes placed at corresponding stages of the pathway. SCN, severe congenital neutropenia; WHIM, warts, hypogammaglobulinemia, immunodeficiency myelokathexis; LAD, leukocyte adhesion deficiencies; CGD, chronic granulomatous diseases; MSMD, Mendelian susceptibility to mycobacterial disease.

PRIMARY IMMUNODEFICIENCIES OF THE INNATE IMMUNE SYSTEM

Primary immunodeficiencies of the innate immune system are relatively rare and account for approximately 10% of all PIDs.

Severe Congenital Neutropenia

Severe congenital neutropenia (SCN) consists of a group of inherited diseases that are characterized by severely impaired neutrophil counts [<500 polymorphonuclear leukocytes (PMN)/μL of blood]. The condition is usually manifested from birth. Severe congenital neutropenia may also be cyclic (with a 3-week periodicity), and other neutropenia syndromes can also be intermittent. Although the most frequent inheritance pattern for SCN is autosomal dominant, autosomal recessive and X-linked recessive conditions also exist. Bacterial infections at the interface between the body and the external milieu (e.g., the orifices, wounds, and the respiratory tract) are common manifestations. Bacterial infections can rapidly progress through soft tissue and are followed by dissemination in the bloodstream. Severe visceral fungal infections can also ensue. The absence of pus is a hallmark of this condition.

Diagnosis of SCN requires examination of the bone marrow. Most SCNs are associated with a block in granulopoiesis at the promyelocytic stage (Fig. 316-1). Severe congenital neutropenia has multiple etiologies, and to date mutations in 11 different genes have been identified. Most of these mutations result in isolated SCN, whereas others are syndromic (Chap. 60). The most frequent forms of SCN are caused by the premature cell death of granulocyte precursors, as observed in deficiencies of GFI1, HAX1, and elastase 2 (ELA2), with the latter accounting for 50% of SCN sufferers. Certain ELA2 mutations cause cyclic neutropenia syndrome. A gain-of-function mutation in the WASP gene (see the section on "Wiskott-Aldrich syndrome" below) causes X-linked SCN, which is also associated with monocytopenia.

As mentioned above, SCN exposes the patient to life-threatening, disseminated bacterial and fungal infections. Treatment requires careful hygiene measures, notably in infants. Later in life, special oral and dental care is essential, along with the prevention of bacterial infection by prophylactic administration of trimethoprim/sulfamethoxazole. Subcutaneous injection of the cytokinegranulocyte colony-stimulating factor (G-CSF) usually improves neutrophil development and thus prevents infection in most SCN diseases. However, there are two caveats: (1) a few cases of SCN with ELA2 mutation are refractory to G-CSF and may require curative treatment via allogeneic hematopoietic stem cell transplantation (HSCT) and (2) a subset of G-CSF-treated patients carrying ELA2 mutations are at a greater risk of developing acute myelogenous leukemia associated (in most cases) with somatic gain-of-function mutations of the G-CSF receptor gene.

Asplenia

Primary failure of the development of a spleen is an extremely rare disease that can be either syndromic (in Ivemark syndrome) or isolated with an autosomal dominant expression; in the latter case, the gene remains to be identified. Due to the absence of natural filtration of microbes in the blood, asplenia predisposes affected individuals to fulminant infections by encapsulated bacteria. Although most infections occur in the first years of life, cases may also arise in adulthood. The diagnosis is confirmed by abdominal ultrasonography and the detection of Howell-Jolly bodies in red blood cells. Effective prophylactic measures (twice-daily oral penicillin and appropriate vaccination programs) usually prevent fatal outcomes. The genetic causes of asplenia remain unknown.

Leukocyte Adhesion Deficiency (LAD)

Leukocyte adhesion deficiency consists of three autosomal recessive conditions (LAD I, II, and III) (Chap. 60). The most frequent condition (LAD I) is caused by mutations in the β2 integrin gene; following leukocyte activation, β2 integrins mediate adhesion to inflamed endothelium expressing cognate ligands. LAD III results from a defect in a regulatory protein (kindlin, also known as Fermt 3) involved in activating the ligand affinity of β2 integrins. The extremely rare LAD II condition is the end result of a defect in selectin-mediated leukocyte rolling that occurs prior to β2 integrin binding. There is a primary defect in fucose transporter such that oligosaccharide selectin ligands are missing in this syndromic condition.

Given that neutrophils are not able to reach infected tissues, LAD renders the individual susceptible to bacterial and fungal infections in a way that is similar to that of patients with SCN. LAD also causes impaired wound healing and delayed loss of the umbilical cord. A diagnosis can be suspected in cases of pus-free skin/tissue infections and massive hyperleukocytosis (>30,000/μL) in the blood (mostly granulocytes). Patients with LAD III also develop bleeding because the β2 integrin in platelets is not functional. Use of immunofluorescence and functional assays to detect β2 integrin can help form a diagnosis. Severe forms of LAD may require hematopoietic stem cell transplantation (HSCT), although gene therapy is also now being considered. Neutrophil-specific granule deficiency (a very rare condition caused by a mutation in the gene for transcription factor C/EBPα) results in a condition that is clinically similar to LAD.

Chronic Granulomatous Diseases

Chronic granulomatous diseases (CGDs) are characterized by impaired phagocytic killing of microorganisms by neutrophils and macrophages (Chap. 60). The incidence is approximately 1 per 200,000 live births. About 70% of cases are associated with X-linked recessive inheritance versus

autosomal inheritance in the remaining 30%. CGD causes deep-tissue bacterial and fungal abscesses in macrophage-rich organs such as the lymph nodes, liver, and lungs. Recurrent skin infections (such as folliculitis) are common and can prompt an early diagnosis of CGD. The infectious agents are typically catalase-positive bacteria (such as *Staphylococcus aureus* and *Serratia marcescens*) but also include *Burkholderia cepacia*, pathogenic mycobacteria (in certain regions of the world), and fungi (mainly filamentous molds, such as *Aspergillus*).

CGD is caused by defective production of reactive oxygen species (ROS) in the phagolysosome membrane following phagocytosis of microorganisms. It results from the lack of a component of NADPH oxidase (gp91phox or p22phox) or of the associated adapter/ activating proteins (p47phox, p67phox, or p40phox) that mediate the transport of electrons into the phagolysosome for creating ROS by interaction with O_2. Under normal circumstances, these ROS either directly kill engulfed microorganisms or enable the rise in pH needed to activate the phagosomal proteases that contribute to microbial killing. Diagnosis of CGD is based on assays of ROS production in neutrophils and monocytes (Table 316-2). As its name suggests, CGD is also a granulomatous disease. Macrophage-rich granulomas can often arise in the liver, spleen, and other organs. These are sterile granulomas that cause disease by obstruction (bladder, pylorus, etc.) or inflammation (colitis, restrictive lung disease).

The management of infections in patients with CGD can be a complex process. The treatment of bacterial infections is generally based on combination therapy with antibiotics that are able to penetrate into cells. The treatment of fungal infections requires aggressive, long-term use of antifungals. Inflammatory/granulomatous lesions are usually steroid-sensitive; however, glucocorticoids often contribute to the spread of infections. Hence, there is strong need for new therapeutic options in what is still a poorly understood disease.

The treatment of CGD mostly relies on preventing infections. It has been unambiguously demonstrated that prophylactic usage of trimethoprim/sulfamethoxazole is both well tolerated and highly effective in reducing the risk of bacterial infection. Daily administration of azole derivatives (notably intraconazole) also reduces the frequency of fungal complications. It has long been suggested that interferon-γ administration is helpful, although medical experts continue to disagree over this controversial issue. Most patients do reasonably well with prophylaxis and careful management. However, some patients develop severe and persistent fungal infections and/or chronic inflammatory complications that ultimately require HSCT. The latter is an established curative approach for CGD; however, the risk-versus-benefit ratio must be carefully assessed on a case-by-case basis. Gene therapy approaches are also being evaluated.

Mendelian Susceptibility to Mycobacterial Disease (MSMD)

This group of diseases is characterized by a defect in the IL-12 interferon (IFN)-γ axis (including IL-12p40, IL-12 receptor (R) β_1, interferon-γ R_1 and R_2, and STAT1 deficiencies), which ultimately leads to impaired IFN-γ-dependent macrophage activation. Both recessive and dominant inheritance modes have been observed. The hallmark of this PID is a specific and narrow vulnerability to tuberculous and nontuberculous mycobacteria. The most severe phenotype (as observed in complete IFN-γ receptor deficiency) is characterized by disseminated infection that can be fatal even when aggressive and appropriate antimycobacterial therapy is applied. In addition to mycobacterial infections, MSMD patients (and particularly those with an IL-12/IL-12 R deficiency) are prone to developing *Salmonella* infections. Although MSMDs are very rare, they should be considered in any patient with persistent mycobacterial infection. Treatment with interferon γ may efficiently bypass an IL-12/IL-12 R deficiency.

Toll-Like Receptor (TLR) Pathway Deficiencies

In a certain group of patients with early-onset, invasive *Streptococcus pneumoniae* infections or (less frequently) *Staphylococcus aureus* or other pyogenic infections, conventional screening for PIDs does not identify the cause of the defect in host defense. It has been established that these patients carry recessive mutations in genes that encode essential adaptor molecules (IRAK4 and MYD88) involved in the signaling pathways of the majority of known Toll-like receptors (TLRs) (Chap. 314). Remarkably, susceptibility to infection appears to decrease after the first few years of life—perhaps an indication that adaptive immunity (once triggered by an initial microbial challenge) is then able to prevent recurrent infections.

Certain TLRs (TLR-3, -7, -8, and -9) are involved in the recognition of RNA and DNA and usually become engaged during viral infections. Very specific susceptibility to herpes simplex encephalitis has been described in patients with a deficiency in Unc93b (a molecule associated with TLR-3, -7, -8, and -9 and probably required for correct subcellular localization) or TLR-3. The fact that no other TLR deficiencies have been found—despite extensive screening of patients with unexplained, recurrent infections—strongly suggests that these receptors are functionally redundant. Hypomorphic mutations in NEMO/IKK-γ (a member of the NF-κB complex, which is activated downstream of TLR receptors) lead to a complex, variable immunodeficiency and a number of associated features. Susceptibility to both invasive, pyogenic infections and mycobacteria may be observed in this particular setting.

Complement Deficiency

The complement system is composed of a complex cascade of plasma proteins (Chap. 314) that leads to

the deposition of C3b fragments on the surface of particles and the formation of immune complexes that can culminate in the activation of a lytic complex at the bacterial surface. C3 cleavage can be mediated via three pathways: the classic, alternate, and lectin pathways. C3b coats particles as part of the opsonization process that facilitates phagocytosis following binding to cognate receptors. A deficiency in any component of the classic pathway (C1q, C1r, C1s, C4, and C2) can predispose an individual to bacterial infections that are tissue-invasive or that occur in the respiratory tract. Likewise, a C3 deficiency or a deficiency in factor I (a protein that regulates C3 consumption, thus leading to a C3 deficiency due to its absence) also results in the same type of vulnerability to infection. It has recently been reported that a very rare deficiency in Ficolin-3 predisposes affected individuals to bacterial infections. Deficiencies in the alternative pathway (factors D and properdin) are associated with the occurrence of invasive *Neisseria* infections.

Lastly, deficiencies of any complement component involved in the lytic phase (C5, C6, C7, C8, and, to a lesser extent, C9) predispose affected individuals to systemic infection by *Neisseria*. This is explained by the critical role of complement in the lysis of the thick cell wall possessed by this class of bacteria.

Diagnosis of a complement deficiency relies primarily on testing the status of the classic and alternate pathway via functional assays, i.e., the CH50 and AP50 tests, respectively. When either pathway is profoundly impaired, determination of the status of the relevant components in that pathway enables a precise diagnosis. Appropriate vaccinations and daily administration of oral penicillin are efficient means of preventing recurrent infections. It is noteworthy that several complement deficiencies (in the classic pathway and the lytic phase) may also predispose affected individuals to autoimmune diseases (notably systemic lupus erythematosus; Chap. 319).

PRIMARY IMMUNODEFICIENCIES OF THE ADAPTIVE IMMUNE SYSTEM

T Lymphocyte Deficiencies (Table 316-1, Figs. 316-2 and 316-3)

Figure 316-2

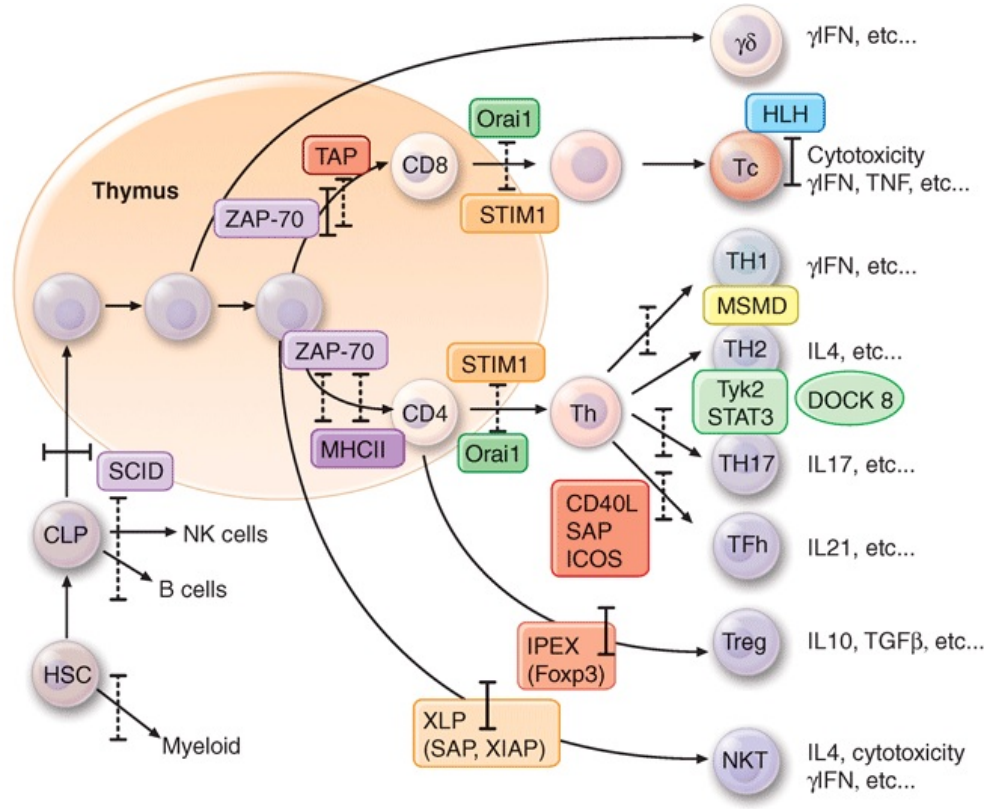

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

T cell differentiation, effector pathways, and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common lymphoid progenitors (CLPs), which, in turn, give rise to the T cell precursors that migrate to the thymus. The development of CD4+ and CD8+ T cells is shown. Known T cell effector pathways are indicated, i.e., γδ cells, cytotoxic T cells (Tc), TH1, TH2, TH17, TFh (follicular helper) CD4 effector T cells, regulatory T cells (Treg), and natural killer T cells (NKTs); abbreviations for PIDs are contained in boxes. Vertical bars indicate a complete deficiency; broken bars a partial deficiency. SCID, severe combined immunodeficiency; ZAP 70, zeta-associated protein deficiency, MHCII, major histocompatibility complex class II deficiency; TAP, TAP1 and 2 deficiencies; Orai1, Stim1 deficiencies; HLH, hematopoietic lymphohistiocytosis; MSMD, Mendelian susceptibility to mycobacterial disease; Tyk2, DOCK8, autosomal recessive form of hyper-

IgE syndrome; STAT3, autosomal dominant form of hyper-IgE syndrome; CD40L, ICOS, SAP deficiencies; IPEX, immunodysregulation polyendocrinopathy enteropathy X-linked syndrome; XLP, X-linked proliferative syndromes.

Figure 316-3

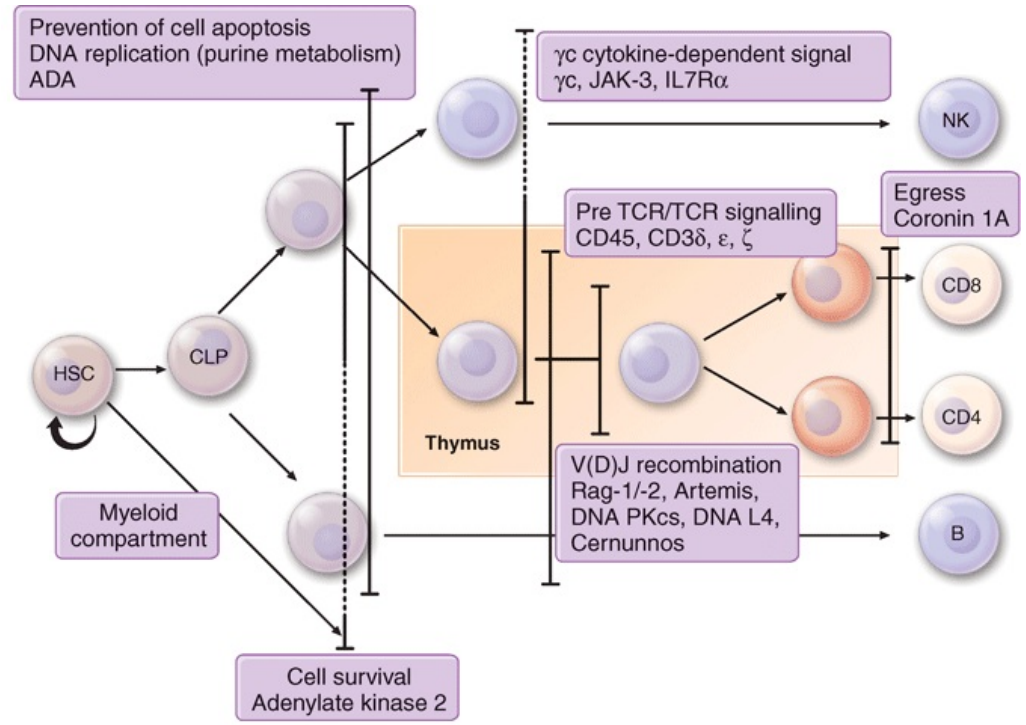

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

T cell differentiation and severe combined immunodeficiencies (SCIDs). The vertical bars indicate the six mechanisms currently known to lead to SCID. The names of deficient proteins are indicated in the boxes adjacent to the vertical bars. A broken line means that deficiency is partial or involves only some of the indicated immunodeficiencies. HSCs, hematopoietic stem cells; CLPs, common lymphoid progenitors; ADA, adenosine deaminase deficiency; NKs, natural killer cells; TCR, T cell receptor; DNAL4, DNA ligase 4.

Given the central role of T lymphocytes in adaptive immune responses (Chap. 314), PIDs involving T cells generally have severe pathologic consequences; this explains the poor overall prognosis and the need for early diagnosis and the early intervention with appropriate therapy. Several differentiation pathways of T cell effectors have been described, one or all of which may be affected by a given PID (Fig. 316-2). Follicular helper CD4+ T cells in germinal centers are required for T-dependent antibody production, including the generation of Ig class-switched, high-affinity antibodies. CD4+ T_{H1} cells provide cytokine-dependent (mostly interferon-γ-dependent) help to macrophages for intracellular killing of various microorganisms, including mycobacteria and *Salmonella*. CD4+ T_{H2} cells produce IL-4, IL-5, and IL-13 and thus recruit and activate eosinophils and other cells required to fight helminth infections. CD4+ T_{H17} cells produce IL-17 and IL-22 cytokines that recruit neutrophils to the skin and lungs to fight bacterial and fungal infections. Cytotoxic CD8+ T cells can kill infected cells, notably in the context of viral infections. In addition, certain T cell deficiencies predispose affected individuals to *Pneumocystis jiroveci* lung infections early in life and to chronic gut/biliary duct/liver infections by *Cryptosporidia* and related genera later on in life. Lastly, naturally occuring or induced regulatory T cells are essential for controlling inflammation (notably reactivity to commensal bacteria in the gut) and autoimmunity. The role of other T cell subsets with limited T cell receptor (TCR) diversity [such as γδTCR T cells or natural killer T (NKT) cells] in PIDs is less well known; however, these subsets can be defective in certain PIDs and this finding can sometimes contribute to the diagnosis (e.g., NKT cell deficiency in X-linked proliferative syndrome). T cell deficiencies account for approximately 20% of all cases of PID.

SEVERE COMBINED IMMUNODEFICIENCIES

Severe combined immunodeficiencies (SCIDs) constitute a group of rare PIDs characterized by a profound block in T cell development and thus the complete absence of these cells. The developmental block is always the consequence of an intrinsic deficiency. The incidence of SCID is estimated to be 1 in 50,000 to 100,000 live births. Given the severity of the T cell deficiency, clinical consequences occur early in life (usually within 3 to 6 months of birth). The most frequent clinical manifestations are recurrent

oral candidiasis, failure to thrive, and protracted diarrhea and/or acute interstitial pneumonitis caused by *Pneumocystis jiroveci* (although the latter can also be observed in the first year of life in children with B cell deficiencies). Severe viral infections or invasive bacterial infections can also occur. Patients may also experience complications related to infections caused by live vaccines (notably bacille Calmette-Guérin; BCG) that may lead not only to local and regional infection but also to disseminated infection manifested by fever, splenomegaly, and skin and lytic bone lesions. A scaly skin eruption can be observed in a context of maternal T cell engraftment (see below). A diagnosis of SCID can be suspected on the basis of the patient's clinical history and, possibly, a family history of deaths in very young children (suggestive of either X-linked or recessive inheritance). Lymphocytopenia is strongly suggestive of SCID in more than 90% of cases (Table 316-2). The absence of a thymic shadow on a chest x-ray can also be suggestive of SCID. An accurate diagnosis relies on precise determination of the number of circulating T, B, and NK lymphocytes and their subsets. T cell lymphopenia may be masked in some patients by the presence of maternal T cells (derived from maternal-fetal blood transfers) that cannot be eliminated. Although counts are usually low (<500/μL of blood), higher maternal T cell counts may, under some circumstances, initially mask the presence of SCID. Thus, screening for maternal cells by using adequate genetic markers should be performed whenever necessary. Inheritance pattern analysis and lymphocyte phenotyping can discriminate between various forms of SCID and provide guidance in the choice of accurate molecular diagnostic tests (see below). To date, six distinct causative mechanisms for SCID (Fig. 316-3) have been identified:

Severe Combined Immunodeficiency Caused by a Cytokine-Signaling Deficiency

The most frequent SCID phenotype (accounting for 40–50% of all cases) is the absence of both T and NK cells. This outcome results from a deficiency in either the common γ chain (γc) receptor (γc) that is shared by several cytokine receptors (the IL-2, -4, -7, -9, -15, and -21 receptors) or Jak-associated kinase (JAK) 3 that binds to the cytoplasmic portion of the γc chain receptor and induces signal transduction following cytokine binding. The former form of SCID (γc deficiency) has an X-linked inheritance mode, while the second is autosomal recessive. A lack of the IL-7Rα chain (which, together with γc, forms the IL-7 receptor) induces a selective T cell deficiency.

Purine Metabolism Deficiency

Ten to 20% of SCID patients exhibit a deficiency in adenosine deaminase (ADA), an enzyme of purine metabolism that deaminates adenosine (ado) and deoxyadenosine (dAdo). An ADA deficiency results in the accumulation of ado and dAdo metabolites that induce premature cell death of lymphocyte progenitors. The condition results in the absence of B and NK lymphocytes as well as T cells. The clinical expression of complete ADA deficiency typically occurs very early in life. Since ADA is a ubiquitous enzyme, its deficiency can also cause bone dysplasia with abnormal costochondral junctions and metaphyses (found in 50% of cases) and neurologic defects. The very rare purine nucleoside phosphorylase (PNP) deficiency causes a profound although incomplete T cell deficiency that is often associated with severe neurologic impairments.

Defective Rearrangements of T and B Cell Receptors

A series of SCID conditions are characterized by a selective deficiency in T and B lymphocytes with autosomal recessive inheritance. These conditions account for 20–30% of SCID cases and result from mutations in genes encoding proteins that mediate the recombination of V(D)J gene elements in T and B cell antigen receptor genes (required for the generation of diversity in antigen recognition). The main deficiencies involve RAG-1, RAG-2, DNA-dependent protein kinase, and Artemis. A less severe (albeit variable) immunologic phenotype can result from other deficiencies in the same pathway, i.e., DNA ligase 4 and Cernunnos deficiencies. Given that these latter factors are involved in DNA repair, these deficiencies also cause developmental defects.

Defective (Pre-)T Cell Receptor Signaling in the Thymus

A selective T cell defect can be caused by a series of rare deficiencies in molecules involved in signaling via the pre-TCR or the TCR. These include deficiencies in CD3 subunits associated with the (pre)TCR (i.e., CD3δ, ε, and ζ) and CD45.

Reticular Dysgenesis

Reticular dysgenesis is an extremely rare form of SCID that causes T and NK deficiencies with severe neutropenia and sensorineural deafness. It results from an adenylate kinase 2 deficiency.

Defective Egress of Lymphocytes

Defective egress of lymphocytes from the thymus has been found in a patient with very low T cell counts but a normal thymic shadow. This condition was found to result from a deficiency in coronin-1A.

Patients with SCID require appropriate care with aggressive anti-infective therapies, immunoglobulin replacement, and (when necessary) parenteral nutrition support. In most cases, curative treatment relies on HSCT. Today, HSCT provides a very high curative potential for SCID patients who are otherwise in reasonably good condition. In this regard, the feasibility of neonatal screening is now being evaluated. Gene therapy has been found to be successful for cases of X-linked SCID (γc deficiency) and SCID caused by an ADA deficiency, although toxicity has become an issue in the treatment of the former disease. Lastly, a third option for the treatment of ADA deficiency consists of enzyme substitution with a

pegylated enzyme.

THYMIC DEFECTS

A profound T cell defect can also result from faulty development of the thymus, as is most often observed in rare cases of DiGeorge syndrome—a relatively common condition leading to a constellation of developmental defects. In approximately 1% of such cases, the thymus is completely absent, leading to virtually no mature T cells. However, expansion of oligoclonal T cells can occur and is associated with skin lesions. Diagnosis (using immunofluorescence in situ hybridization) is based on the identification of a hemizygous deletion in the long arm of chromosome 22. To recover the capability for T cell differentiation, these cases require a thymic graft. CHARGE (coloboma of the eye, heart anomaly, choanal atresia, retardation, genital and ear anomalies) syndrome (CHD7 deficiency) is a less frequent cause of impaired thymus development. Lastly, the very rare "nude" defect is characterized by the absence of both hair and the thymus.

OMENN SYNDROME

Omenn syndrome consists of a subset of T cell deficiencies that present with a unique phenotype, including early-onset erythrodermia, alopecia, hepatosplenomegaly, and failure to thrive. These patients usually display T cell lymphocytosis, eosinophilia, and low B cell counts. It has been found that the T cells of these patients exhibit a low TCR heterogeneity. This peculiar syndrome is the consequence of hypomorphic mutations in genes usually associated with SCID, i.e., RAG-1, RAG-2, or (less frequently) Artemis or IL-7Rα. The impaired homeostasis of differentiating T cells thus causes this immune system –associated disease. These patients are very fragile, requiring simultaneous anti-infective therapy, nutritional support, and immunosuppression. HSCT provides a curative approach.

FUNCTIONAL T CELL DEFECTS (FIG. 316-2)

A subset of T cell PIDs with autosomal inheritance is characterized by partially preserved T cell differentiation but defective activation resulting in abnormal effector function. There are many causes of these defects, but all lead to susceptibility to viral and opportunistic infections, chronic diarrhea, and failure to thrive, with onset during childhood. Careful phenotyping and *in vitro* functional assays are required to identify these diseases, the best characterized of which are the following.

Zeta-Associated Protein 70 (ZAP70) Deficiency

Zeta-associated protein 70 is recruited to the TCR following antigen recognition. A ZAP70 deficiency leads typically to an almost complete absence of CD8+ T cells; CD4+ T cells are present but cannot be activated in vitro by TCR stimulation.

Calcium Signaling Defects

A small number of patients have been reported who exhibit a profound defect in in vitro T and B cell activation as a result of defective antigen receptor-mediated Ca^{2+} influx. This defect is caused by a mutation in the calcium channel gene (ORAI-) or its activator (STIM-1). It is noteworthy that these patients are also prone to autoimmune manifestations (blood cytopenias) and exhibit a nonprogressive muscle disease.

Human Leukocyteantigen (HLA) Class II Deficiency

Defective expression of HLA class II molecules is the hallmark of a group of four recessive genetic defects all of which affect molecules (RFX5, RFXAP, RFXANK, and CIITA) involved in the transactivation of the genes coding for HLA class II. As a result, low but variable CD4+ T cell counts are observed in addition to defective antigen-specific T and B cell responses. These patients are particularly susceptible to herpesvirus, adenovirus, and enterovirus infections and chronic gut/liver *Cryptosporidium* infections.

HLA Class I Deficiency

Defective expression of molecules involved in antigen presentation by HLA class I molecules (i.e., TAP-1, TAP-2, and Tapasin) leads to reduced CD8+ T cell counts, loss of HLA class I antigen expression, and a particular phenotype consisting of chronic obstructive pulmonary disease and severe vasculitis.

Other Defects

A variety of other T cell PIDs have been described, some of which are associated with a precise molecular defect [e.g., IL-2-inducible T cell kinase (ITK) deficiency]. These conditions are also characterized by profound vulnerability to infections, such as severe Epstein-Barr virus (EBV)–induced B cell proliferation and autoimmune disorders in ITK deficiency. Milder phenotypes are associated with CD8 and CD3γ deficiencies.

HSCT is indicated for most of these diseases, although the prognosis is worse than in SCID because many patients are chronically infected at the time of diagnosis. Fairly aggressive immunosuppression and myeloablation may be necessary to achieve engraftment of allogeneic stem cells.

T CELL PRIMARY IMMUNODEFICIENCIES WITH DNA REPAIR DEFECTS

This is a group of PIDs characterized by a combination of T and B cell defects of variable intensity, together with a number of nonimmunologic features resulting from DNA fragility. The autosomal recessive disorder *ataxia-telangiectasia* (AT) is the most frequently encountered condition in this group.

It has an incidence of 1:40,000 live births and causes B cell defects (low IgA, IgG2 deficiency, and low antibody production), which often require immunoglobulin replacement. Ataxia telangiectasia is associated with a progressive T cell immunodeficiency. As the name suggests, the hallmark features of AT are telangiectasia and cerebellar ataxia. The latter manifestations may not be detectable before the age of 3–4 years, so that AT should be considered in young children with IgA deficiency and recurrent and problematic infections. Diagnosis is based on a cytogenetic analysis showing excessive chromosomal rearrangements (mostly affecting chromosomes 7 and 14) in lymphocytes. Ataxia telangiectasia is caused by a mutation in the gene encoding the ATM protein—a kinase that plays an important role in the detection and repair of DNA lesions (or cell death if the lesions are too numerous) by triggering several different pathways. Overall, AT is a progressive disease that carries a very high risk of lymphoma, leukemia, and (during adulthood) carcinomas. A variant of AT ("AT-like disease") is caused by mutation in the MRE11 gene.

Nijmegen breakage syndrome (NBS) is a less common condition that also results from chromosome instability (with the same cytogenetic abnormalities as in AT). NBS is characterized by a severe T and B cell combined immune deficiency with autosomal recessive inheritance. Individuals with NBS exhibit microcephaly and a bird-like face, but have neither ataxia nor telangiectasia. The risk of malignancies is very high. Nijmegen breakage syndrome results from a deficiency in Nibrin (NBSI, a protein associated with MRE11 and Rad50 that is involved in checking DNA lesions) caused by hypomorphic mutations.

Severe forms of *dyskeratosis congenita* (also known as Hoyeraal-Hreidarsson syndrome) combine a progressive immunodeficiency that can also include an absence of B and NK lymphocytes, progressive bone marrow failure, microcephaly, in utero growth retardation, and gastrointestinal disease. The disease can be X-linked or, more rarely, autosomal recessive. It is caused by the mutation of genes encoding telomere maintenance proteins, including dyskerin (DKC1).

Finally, the *immunodeficiency with centromeric and facial anomalies* (ICF) is a complex syndrome of autosomal recessive inheritance that variably combines a mild T cell immune deficiency with a more severe B cell immune deficiency, coarse face, digestive disease, and mild mental retardation. A diagnostic feature is the detection by cytogenetic analysis of multiradial aspects in multiple chromosomes (most frequently 1, 9, and 16) corresponding to an abnormal DNA structure secondary to defective DNA methylation. It is the consequence of a deficiency in the DNA methyltransferase DNMT3B.

T CELL PRIMARY IMMUNODEFICIENCIES WITH HYPER-IGE

Several T cell PIDs are associated with elevated serum IgE levels (as in Omenn syndrome). A condition sometimes referred to as *autosomal recessive hyper-IgE syndrome* is notably characterized by recurrent bacterial infections in the skin and respiratory tract and severe skin and mucosal infections by pox viruses and human papillomaviruses, together with severe allergic manifestations. T and B lymphocyte counts are low. Mutations in the *DOCK8*gene have been found in a subset of these patients.

A very rare, related condition with autosomal recessive inheritance that causes a similar susceptibility to infection with various microbes (see above), including mycobacteria, reportedly results from a deficiency in Tyk-2, a JAK family kinase involved in the signaling of many different cytokine receptors.

AUTOSOMAL DOMINANT HYPER-IGE SYNDROME (HIES)

This unique condition, the *autosomal dominant hyper-IgE syndrome*, is usually diagnosed by the combination of recurrent skin and lung infections that can be complicated by pneumatoceles. Infections are caused by pyogenic bacteria and fungi. Several other manifestations characterize HIES, including facial dysmorphy, defective loss of primary teeth, hyperextensibility, scoliosis, and osteoporosis. Elevated serum IgE levels are typical of this syndrome. Recently, defective TH17 effector responses have been shown to account at least in part for the specific patterns of susceptibility to particular microbes. This condition is caused by a heterozygous (dominant) mutation in the gene encoding the transcription factor STAT3 that is required in a number of signaling pathways following binding of cytokine to cytokine receptors (such as that of IL-6 and the IL-6 receptor).

CARTILAGE HAIR HYPOPLASIA

The autosomal recessive *cartilage hair hypoplasia* (CHH) disease is characterized by short-limb dwarfism, metaphyseal dysostosis, and sparse hair, together with a combined T and B cell PID of extremely variable intensity (ranging from quasi-SCID to no clinically significant immune defects). The condition can predispose to erythroblastopenia, autoimmunity, and tumors. It is caused by mutations in the RMRP gene for a noncoding ribosome-associated RNA.

CD40 LIGAND AND CD40 DEFICIENCIES

Hyper-IgM syndrome (HIGM) is a well-known PID that is usually classified as a B cell immune deficiency (see **Fig. 316-4** and below). It results from defective immunoglobulin class switch recombination (CSR) in germinal centers and leads to profound deficiency in production of IgG, IgA, and IgE (although IgM production is maintained). Approximately half of HIGM sufferers are also prone to opportunistic infections, e.g., interstitial pneumonitis caused by *Pneumocystis jiroveci* (in young children), protracted diarrhea and cholangitis caused by *Cryptosporidium*, and infection of the brain with *Toxoplasma gondii*.

Figure 316-4

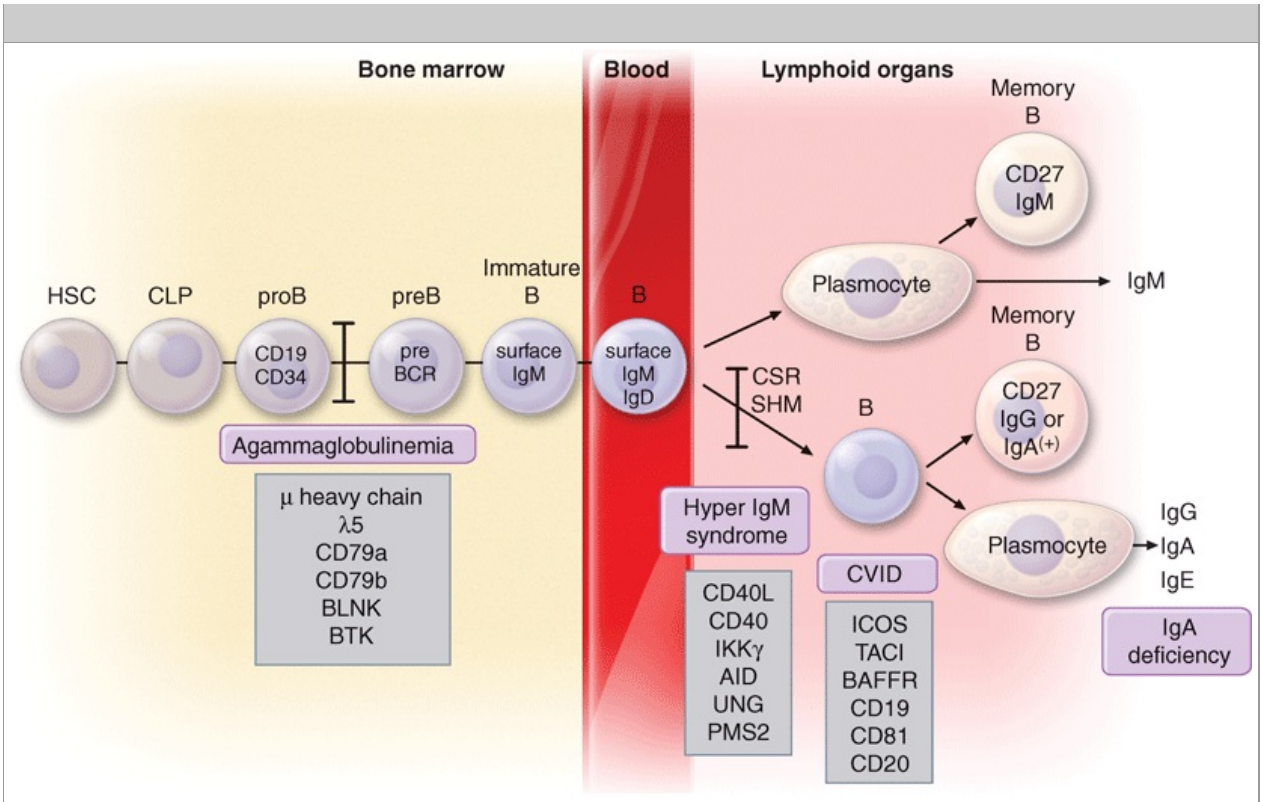

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B cell differentiation and related primary immunodeficiencies (PIDs). Hematopoietic stem cells (HSCs) differentiate into common lymphoid progenitors (CLPs), which give rise to pre-B cells. The B cell differentiation pathway goes through the pre–B cell stage (expression of the μ heavy chain and surrogate light chain), the immature B cell stage (expression of surface IgM), and the mature B cell stage (expression of surface IgM and IgD). The main phenotypic characteristics of these cells are indicated. In lymphoid organs, B cells can differentiate into plasma cells and produce IgM or undergo (in germinal centers) Ig class switch recombination (CSR) and somatic mutation of the variable region of V genes (SHM) that enable selection of high-affinity antibodies. These B cells produce antibodies of various isotypes and generate memory B cells. PIDs are indicated in the purple boxes. CVID: common variable immunodeficiency.

In the majority of cases, this condition has an X-linked inheritance and is caused by a deficiency in CD40 ligand (L). CD40L induces signaling events in B cells that are necessary for both CSR and adequate activation of other CD40-expressing cells that are involved in innate immune responses against the above-mentioned microorganisms. More rarely, the condition is caused by a deficiency in CD40 itself. The poorer prognosis of CD40L and CD40 deficiencies (relative to most other HIGM conditions) implies that (1) thorough investigations have to be performed in all cases of HIGM and (2) potentially curative HSCT should be discussed on a case-by-case basis for this group of patients.

WISKOTT-ALDRICH SYNDROME

Wiskott-Aldrich syndrome (WAS) is a complex, recessive, X-linked disease with an incidence of approximately 1 in 200,000 live births. It is caused by mutations in the WASP gene that affect not only T lymphocytes but also the other lymphocyte subsets, dendritic cells, and platelets. WAS is typically characterized by the following clinical manifestations: recurrent bacterial infections, eczema, and bleeding caused by thrombocytopenia. However, these manifestations are highly variable—mostly as a consequence of the many different WASP mutations that have been observed. Null mutations predispose affected individuals to invasive and bronchopulmonary infections, viral infections, severe eczema, and autoimmune manifestations. The latter include autoantibody-mediated blood cytopenia, glomerulonephritis, skin and visceral vasculitis (including brain vasculitis), erythema nodosum, and arthritis. Another possible consequence of WAS is lymphoma, which may be virally induced (e.g., by EBV or Kaposi's sarcoma–associated herpesvirus). Thrombocytopenia can be severe and compounded by the peripheral destruction of platelets associated with autoimmune disorders. Hypomorphic mutations usually lead to milder outcomes that are generally limited to thrombocytopenia. It is noteworthy that even patients with "isolated" X-linked thrombocytopenia can develop severe autoimmune disease or lymphoma later in life. The immunologic workup is not very informative; there can be a relative CD8+ T cell deficiency, frequently accompanied by low serum IgM levels and decreased antigen-specific antibody responses. A typical feature is reduced-sized platelets on a blood smear. Diagnosis is based on intracellular immunofluorescence analysis of WAS protein (WASp) expression in blood cells. WASp regulates the actin cytoskeleton and thus plays an important role in many lymphocyte functions, including cell adhesion and migration and the formation of synapses between antigen-presenting and target cells.

Predisposition to autoimmune disorders may (at least in part) be related to defective regulatory T cells. The treatment of WAS should match the severity of disease expression. Prophylactic antibiotics, immunoglobulin G (IgG) supplementation, and careful topical treatment of eczema are indicated. Although splenectomy improves platelet count in a majority of cases, this intervention is associated with a significant risk of infection (both pre- and post-HSCT). Allogeneic HSCT is curative, with fairly good results overall. Gene therapy trials are also under way.

A few other complex PIDs are worth mentioning. *Sp110 deficiency* causes a T cell PID with liver venoocclusive disease and hypogammaglobulinemia. *Chronic mucocutaneous candidiasis* (CMC) is probably a heterogeneous disease, considering the different inheritance patterns that have been observed. In some cases, chronic candidiasis is associated with late-onset bronchopulmonary infections, bronchiectasis, and brain aneurysms. Moderate forms of CMC are related to autoimmunity and AIRE deficiency (see below). In this setting, predisposition to candida infection is associated with the detection of autoantibodies to TH17 cytokines. Recently, innate immunodeficiencies (CARD9 and possibly Dectin-1) have been found in a few families with CMC.

B Lymphocyte Deficiencies (Table 316-1, Fig. 316-4)

Deficiencies that predominantly affect B lymphocytes are the most frequent PIDs and account for 60–70% of all cases. B lymphocytes make antibodies. Pentameric IgMs are found in the vascular compartment and are also secreted at mucosal surfaces. IgG antibodies diffuse freely into extravascular spaces, whereas IgA antibodies are produced and secreted predominantly from mucosa-associated lymphoid tissues. Although Ig isotypes have distinct effector functions, including Fc receptor–mediated and (indirectly) C_3 receptor–dependent phagocytosis of microorganisms, they share the ability to recognize and neutralize a given pathogen. Defective antibody production therefore allows the establishment of invasive, pyogenic bacterial infections as well as recurrent sinus and pulmonary infections (mostly caused by *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and, less frequently, by gram-negative bacteria). If left untreated, recurrent bronchial infections lead to bronchiectasis and, ultimately, cor pulmonale and death. Parasitic infections such as caused by *Giardia lambliasis* as well as bacterial infections caused by *Helicobacter and Campylobacter* of the gut are also observed. A complete lack of antibody production (namely agammaglobulinemia) can also predispose affected individuals to severe, chronic, disseminated enteroviral infections causing meningoencephalitis, hepatitis, and a dermatomyositis-like disease.

Even with the most profound of B cell deficiencies, infections rarely occur before the age of 6 months; this is because of transient protection provided by the transplacental passage of immunoglobulins during the last trimester of pregnancy. Conversely, a genetically nonimmunodeficient child born to a mother with hypogammaglobulinemia is, in the absence of maternal Ig substitution, usually prone to severe bacterial infections in utero and for several months after birth.

Diagnosis of B cell PIDs relies on the determination of serum Ig levels (Table 316-2). Determination of antibody production following immunization with tetanus toxoid vaccine or nonconjugated pneumococcal polysaccharide antigens can also help diagnose more subtle deficiencies. Another useful test is B cell phenotype determination in switched μ–S– CD27+ and nonswitched memory B cells (μ+S+ CD27+). In agammaglobulinemic patients, examination of bone marrow B cell precursors (Fig. 316-4) can help obtain a precise diagnosis and guide the choice of genetic tests.

AGAMMAGLOBULINEMIA

Agammaglobulinemia is characterized by a profound defect in B cell development (<1% of the normal B cell blood count). In most patients, very low residual Ig isotypes can be detected in the serum. In 85% of cases, agammaglobulinemia is caused by a mutation in the BTK gene that is located on the X-chromosome. The BTK gene product is a kinase that participates in (pre) B cell receptor signaling. When the kinase is defective, there is a block (albeit a leaky one) at the pre-B to B cell stage (Fig. 316-4). Detection of BTK by intracellular immunofluorescence of monocytes, and lack thereof in patients with X-linked agammaglobulinemia, is a useful diagnostic test. Not all of the mutations in BTK result in agammaglobulinemia, since some patients have a milder form of hypogammaglobulinemia and low but detectable B cell counts. These cases should not be confused with common variable immunodeficiency (CVID, see below). About 10% of agammaglobulinemia cases are caused by alterations in genes encoding elements of the pre-B cell receptor, i.e., the μ heavy chain, the λ5 surrogate light chain, Igα or Igβ, and the scaffold protein BLNK. In 5% of cases, the defect is unknown. It is noteworthy that agammaglobulinemia can be observed in patients with ICF syndrome, despite the presence of normal peripheral B cell counts. Lastly, agammaglobulinemia can be a manifestation of a myelodysplastic syndrome (associated or not with neutropenia). Treatment of agammaglobulinemic patients is based on immunoglobulin replacement (see below). Profound hypogammaglobulinemia is also observed in adults, in association with thymoma.

HYPER-IGM (HIGM) SYNDROMES

Hyper-IgM syndrome is a rare B cell PID characterized by defective Ig CSR. It results in very low serum levels of IgG and IgA and elevated or normal serum IgM levels. The clinical severity is similar to that seen in agammaglobulinemia, although chronic lung disease and sinusitis are less frequent and enteroviral infections are uncommon. As discussed above, a diagnosis of HIGM involves screening for an X-linked

CD40L deficiency and an autosomal recessive CD40 deficiency, which affect both B and T cells. In 50% of cases affecting only B cells, these isolated HIGM syndromes result from mutations in the gene encoding activation-induced deaminase, the protein that induces CSR in B cell germinal centers. These patients usually have enlarged lymphoid organs. In the other 50% of cases, the etiology is unknown (except for rare UNG and PMS2 deficiencies). Furthermore, IgM-mediated autoimmunity and lymphomas can occur in HIGM syndrome. It is noteworthy that HIGM can result from fetal rubella syndrome or can be a predominant immunologic feature of other PIDs, such as the immunodeficiency associated with ectodermic anhydrotic hypoplasia X-linked NEMO deficiency and the combined T and B cell PIDs caused by DNA repair defects such as ataxia-telangectasia and Cernunnos deficiency.

COMMON VARIABLE IMMUNODEFICIENCY (CVID)

CVID is an ill-defined condition characterized by low serum levels of one or more Ig isotypes. Its prevalence is estimated to be 1 in 20,000. The condition is recognized predominantly in adults, although clinical manifestations can occur earlier in life. Hypogammaglobulinemia is associated with at least partially defective antibody production in response to vaccine antigens. B lymphocyte counts are often normal but can be low. Besides infections, CVID patients may develop lymphoproliferation (splenomegaly), granulomatous lesions, colitis, antibody-mediated autoimmune disease, and lymphomas. A family history is found in 10% of cases. A clear-cut dominant inheritance pattern is found in some families, whereas recessive inheritance is observed more rarely. In most cases, no molecular cause can be identified. A small number of patients in Germany were found to carry mutations in the ICOS gene encoding a T cell-membrane protein that contributes to B cell activation and survival. In 10% of patients with CVID, monoallelic or biallelic mutations of the gene encoding TACI (a member of the TNF receptor family that is expressed on B cells) have been found. In fact, heterozygous TACI mutations correspond to a genetic susceptibility factor, since similar heterozygous mutations are found in 1% of controls. The BAFF receptor was found to be defective in a kindred with CVID, although not all individuals carrying the mutation have CVID.

A diagnosis of CVID should be made after excluding the presence of hypomorphic mutations associated with agammaglobulinemia or more subtle T cell defects; this is particularly the case in children. It is possible that many cases of CVID result from a constellation of factors, rather than a single genetic defect. Recently, rare cases of hypogammaglobulinemia were found to be associated with CD19 and CD81 deficiencies. These patients have B cells that can be identified by typing for other B cell markers.

SELECTIVE IG ISOTYPE DEFICIENCIES

IgA deficiency and CVID represent polar ends of a clinical spectrum due to the same underlying gene defect(s) in a large subset of these patients. IgA deficiency is the most common PID; it can be found in 1 in every 600 individuals. It is asymptomatic in most cases; however, individuals may present with increased numbers of acute and chronic respiratory infections that may lead to bronchiectasis. In addition, over their lifetime, these patients experience an increased susceptibility to drug allergies, atopic disorders, and autoimmune diseases. Symptomatic IgA deficiency is probably related to CVID, since it can be found in relatives of patients with CVID. Furthermore, IgA deficiency may progress to CVID. It is thus important to assess serum Ig levels in IgA-deficient patients (especially when infections occur frequently) in order to detect changes that should prompt the initiation of immunoglobulin replacement. Selective IgG2 (+G4) deficiency (which in some cases may be associated with IgA deficiency) can also result in recurrent sinopulmonary infections and should thus be specifically sought in this clinical setting. These conditions are ill-defined and a pathophysiologic explanation has not been found.

SELECTIVE ANTIBODY DEFICIENCY TO POLYSACCHARIDE ANTIGENS

Some patients with normal serum Ig levels are prone to *S. pneumoniae* and *H. influenzae* infections of the respiratory tract. Defective production of antibodies against polysaccharide antigens (such as those in the *S. pneumoniae* cell wall) can be observed and is probably causative. This condition may correspond to a defect in marginal zone B cells, a B cell subpopulation involved in T-independent antibody responses.

IMMUNOGLOBULIN REPLACEMENT

IgG antibodies have a half-life of 21–28 days. Thus, injection of plasma-derived polyclonal IgG containing a myriad of high-affinity antibodies can provide protection against disease-causing microorganisms in patients with defective IgG antibody production. This form of therapy should not be based on laboratory data alone (i.e., IgG and/or antibody deficiency) but should be guided by the occurrence or not of infections; otherwise, patients might be subjected to unjustified IgG infusions. Immunoglobulin replacement can be performed by IV or subcutaneous routes. In the former case, injections have to be repeated every 3–4 weeks, with a residual target level of 800 mg/mL in patients who had very low IgG levels prior to therapy. Subcutaneous injections are typically performed once a week, although the frequency can be adjusted on a case-by-case basis. A trough level of 800 mg/mL is desirable. Whatever the mode of administration, the main goal is to reduce the frequency of the respiratory tract infections and prevent chronic lung and sinus disease. The two routes appear to be equally safe and efficacious and so the choice should be left to the preference of the patient.

In patients with chronic lung disease, chest physical therapy with good pulmonary toilet and the cyclic use of antibiotics are also needed. Immunoglobulin replacement is well tolerated by most patients,

although the selection of the best-tolerated Ig preparation may be necessary in certain cases. Since IgG preparations contain a small proportion of IgAs, caution should be taken in patients with residual antibody production capacity and a complete IgA deficiency, as these subjects may develop anti-IgA antibodies that can trigger anaphylactic shock. These patients should be treated with IgA-free IgG preparations. Immunoglobulin replacement is a lifelong therapy; its rationale and procedures have to be fully understood and mastered by the patient and his or her family, in order to guarantee the strict observance required for efficacy.

PRIMARY IMMUNODEFICIENCIES AFFECTING REGULATORY PATHWAYS

(Table 316-1)

An increasing number of PIDs have been found to cause homeostatic dysregulation of the immune system, either alone or in association with increased vulnerability to infections. Defects of this type affecting the innate immune system and autoinflammatory syndromes will not be covered in this chapter. However, three specific entities (hemophagocytic lymphohistiocytosis, lymphoproliferation, and autoimmunity) will be described below.

Hemophagocytic Lymphohistiocytosis

Hemophagocytic lymphohistiocytosis (HLH) is characterized by an unremitting activation of CD8+ T lymphocytes and macrophages that leads to organ damage (notably in the liver, bone marrow, and central nervous system). This syndrome results from a broad set of inherited diseases, all of which impair T and NK lymphocyte cytotoxicity. The manifestations of HLH are often induced by a viral infection. EBV is the most frequent trigger. In severe forms of HLH, disease onset may start during the first year of life or even (in rare cases) at birth.

Diagnosis relies on the identification of the characteristic symptoms of HLH (fever, hepatosplenomegay, edema, neurologic diseases, blood cytopenia, increased liver enzymes, hypofibrinogenemia, high triglyceride levels, elevated markers of T cell activation, and hemophagocytic features in the bone marrow or cerebrospinal fluid). Functional assays of postactivation cytotoxic granule exocytosis (CD107 fluorescence at the cell membrane) can suggest genetically determined HLH. The conditions can be classified into three subsets:

1. Familial HLH with autosomal recessive inheritance, including perforin deficiency (30% of cases) that can be recognized by assessing intracellular perforin expression; Munc13-4 deficiency (30% of cases); syntaxin 11 deficiency (10% of cases); Munc18-2 deficiency (20% of cases); and a few residual cases that lack a known molecular defect.
2. HLH with partial albinism. Three conditions combine HLH and abnormal pigmentation, where hair examination can help in the diagnosis: Chediak-Higashi syndrome, Griscelli syndrome, and Hermansky Pudlak syndrome type II. Chediak-Higashi syndrome is also characterized by the presence of giant lysosomes within leukocytes (Chap. 60), in addition to a primary neurologic disorder with slow progression of symptoms over time.
3. X-linked proliferative syndrome (XLP) is characterized in most patients by the induction of HLH following EBV infection, while other patients develop progressive hypogammaglobulinemia similar to what is observed in CVID and/or certain lymphomas. XLP is caused by a mutation in the SH2DIA gene that encodes the adaptor protein SAP (associated with a SLAM family receptor). Several immunologic abnormalities have been described, including low 2B4-mediated NK cell cytotoxicity, impaired differentiation of NKT cells, defective antigen-induced T cell death, and defective T cell helper activity for B cells. A related disorder (XLP2) has recently been described. It is also X-linked and induces HLH (frequently after EBV infection), although the clinical manifestation may be less pronounced. The condition is associated with a deficiency of the antiapoptotic molecule XIAP. The pathophysiology of XLP2 and its relationship to XLP1 remain unclear.

HLH is a life-threatening complication. The treatment of this condition requires aggressive immunosuppression with either the cytotoxic agent VP-16 or anti-T cell antibodies. Once remission has been achieved, HSCT should be performed, since it provides the only curative form of therapy.

Autoimmune Lymphoproliferative Syndrome

Autoimmune lymphoproliferative syndrome (**ALPS**) is characterized by nonmalignant T and B lymphoproliferation causing splenomegaly and enlarged lymph nodes; 70% of patients also display autoimmune manifestations such as autoimmune cytopenias, Guillain-Barré syndrome, uveitis, and hepatitis (Chaps. 59 and 314). A hallmark of ALPS is the presence of CD4 -CD8- $TCR_{\alpha\beta}$+ T cells (2–50%) in the blood of affected individuals. Hypergammaglobulinemia involving IgG and IgA is also frequently observed. The syndrome is caused by a defect in Fas-mediated apoptosis of lymphocytes, which can thus accumulate and mediate autoimmunity. Furthermore, ALPS can lead to malignancies.

Most patients carry a heterozygous mutation in the gene encoding Fas that is characterized by dominant inheritance and variable penetrance, depending on the nature of the mutation. A rare and severe form of the disease with early onset can be observed in patients carrying a biallelic mutation of Fas, which profoundly impairs the protein's expression and/or function. Fas-ligand, caspase 10, caspase 8, and neuroblastoma RAS viral oncogene homologue (NRAS) mutations have also been reported in a few cases of ALPS. Many cases of ALPS have not been precisely delineated at the molecular level.

Treatment of ALPS is essentially based on the use of proapoptotic drugs, which need to be carefully administered in order to avoid toxicity.

Colitis, Autoimmunity, and Primary Immunodeficiencies

Several PIDs (most of which are T cell-related) can cause severe gut inflammation. The prototypic example is *immunodysregulation polyendocrinopathy enteropathy X-linked syndrome* (IPEX), characterized by a widespread inflammatory enteropathy, food intolerance, skin rashes, autoimmune cytopenias, and diabetes. The syndrome is caused by loss-of-function mutations in the gene encoding the transcription factor FOXP3, which is required for the acquisition of effector function by regulatory T cells. In most cases of IPEX, CD4+CD25+ regulatory T cells are absent from the blood. This condition has a poor prognosis and requires aggressive immunosuppression. The only possible curative approach is allogeneic HSCT. IPEX-like syndromes that lack a FOXP3 mutation have also been described. In some cases, a CD25 deficiency has been found. Defective CD25 expression also impairs regulatory cell expansion/function. This functional T cell deficiency means that CD25-deficient patients are also at increased risk of opportunistic infections. It is noteworthy that abnormalities in regulatory T cells have been described in other PID settings, such as in Omenn syndrome, STAT5b deficiency, STIM1 (Ca flux) deficiency, and WAS; these abnormalities may account (at least in part) for the occurrence of inflammation and autoimmunity. The autoimmune features observed in a small fraction of patients with DiGeorge syndrome may have the same cause. Recently, severe inflammatory gut disease has been described in patients with a deficiency in the IL-10 receptor.

A distinct autoimmune entity is observed in *autoimmune polyendocrinopathy candidiasis ectodermal dysplasia* (APECED) syndrome, which is characterized by autosomal recessive inheritance. It consists of multiple autoimmune manifestations that can affect solid organs in general and endocrine glands in particular. Mild, chronic *Candida* infection is often associated with this syndrome. The condition is due to mutations in the autoimmune regulator (AIRE) gene and results in impaired thymic expression of self-antigens by medullary epithelial cells and impaired negative selection of self-reactive T cells that leads to autoimmune manifestations.

CONCLUSION

The variety and complexity of the clinical manifestations of the many different PIDs strongly indicate that it is important to raise awareness of these diseases. Indeed, early diagnosis is essential for establishing an appropriate therapeutic regimen. Hence, patients with suspected PIDs must always be referred to experienced clinical centers that are able to perform appropriate molecular and genetic tests. A precise molecular diagnosis is not only necessary for initiating the most suitable treatment but it is also important for genetic counseling and prenatal diagnosis.

One pitfall that may hamper diagnosis is the high variability that is associated with many PIDs. Variable disease expression can result from the differing consequences of various mutations associated with a given condition, as exemplified by WAS and, to a lesser extent, X-linked agammaglobulinemia (XLA). There can also be effects of modifier genes (as also suspected in XLA) and environmental factors such as EBV infection that can be the main trigger of disease in X-linked lymphoproliferative (XLP) conditions. Furthermore, it has recently been established that somatic mutations in an affected gene can attenuate the phenotype of a number of T cell PIDs. This has been described for ADA deficiency, X-linked SCID, RAG deficiencies, NF-κB essential modulator (NEMO) deficiency, and, most frequently, in WAS. In contrast, somatic mutations can create disease states analogous to PID, as reported for ALPS. Lastly, cytokine-neutralizing autoantibodies can mimic a PID, as shown for interferon γ.

Many aspects of the pathophysiology of PIDs are still unknown, and the disease-causing gene mutations have not been identified in all cases (as illustrated by CVID and IgA deficiency). However, our medical understanding of PIDs has now reached the stage where scientifically based approaches to the diagnosis and treatment of these diseases can be implemented.

FURTHER READINGS

Boztug K, Klein C: Novel genetic etiologies of severe congenital neutropenia. Curr Opin Immunol 21:472, 2009[PMID: 19782549] [Full Text]

Browne SK, Holland SM: Anticytokine autoantibodies in infectious diseases: Pathogenesis and mechanisms. Lancet Infect Dis 10:875, 2010[PMID: 21109174] [Full Text]

Fischer A: Human primary immunodeficiency diseases. Immunity 27:835, 2007[PMID: 18093537] [Full Text]

Nelson KS, Lewis DB: Adult-onset presentations of genetic immunodeficiencies: Genes can throw slow curves. Curr Opin Infect Dis 23:359, 2010[PMID: 20712480] [Full Text]

Notarangelo LD, et al: Primary immunodeficiencies: 2009 update. J Allergy Clin Immunol 124:1161, 2009[PMID: 20004777] [Full Text]

Ochs HD et al: *Primary Immunodeficiency Diseases. A Molecular and Genetic Approach* . New York, Oxford University Press, 1999

Seger RA: Modern management of chronic granulomatous disease. Br J Haematol 140:255, 2008[PMID: 18217895] [Full Text]

Stiehm ER et al: *Immunologic Disorders in Infants and Children*. Philadelphia, Elsevier Saunders, 2004

Wood P et al: Recognition, clinical diagnosis and management of patients with primary antibody deficiencies: A systematic review. Clin Exp Immunol 149:410, 2007[PMID: 17565605] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRIMARY IMMUNODEFICIENCIES ASSOCIATED WITH(OR SECONDARY TO) OTHER DISEASES: INTRODUCTION

There are an increasing number of conditions in which a primary immunodeficiency (PID) has been described as one facet of a more complex disease setting. It is essential to consider associated diseases when a PID is identified as the primary manifestation and, conversely, not neglect the potentially harmful consequences of a PID that could be masked by other manifestations of a particular syndrome.

Below is a short description of these syndromes in which the PID is classified according to the arm of the immune system that is affected.

PRIMARY IMMUNODEFICIENCIES OF THE INNATE IMMUNE SYSTEM

a) Several severe congenital neutropenia (SCN) syndromes can be associated with malformations. The recently described SCN disease caused by glucose-6-phosphatase deficiency (G6PC3) can be associated with heart and urogenital malformations. The related glycogenesis Ib disease combines SCN with hypoglycemia and hepatosplenomegaly. Some *HAX-1* gene mutations lead to neurocognitive impairments as well as SCN. Barth syndrome combines SCN with cardiomyopathy. Lastly, Shwachman syndrome is a known autosomal recessive entity (caused by mutation of the *SBDS* gene) in which the defect in granulopoiesis can extend to the other hematopoietic lineages; short stature, bone metaphyseal dysplasia, and exocrine pancreatic insufficiency are known hallmarks of this condition.

b) Syndromic asplenia combines the risk of infection with heart defects and situs inversus.

c) Leukocyte adhesion deficiency (LAD) type II includes growth retardation and impairment of cognitive development.

d) A few patients with X-linked chronic granulomatous disease present with a contiguous gene deletion syndrome that can include the McLeod phenotype, which is characterized by anemia, acanthocytosis, and a severe risk of immune reaction against donor red cells because the patient's red cells do not express the Kell antigen. The McLeod phenotype also can result in a neurologic disease.

e) X-linked NF-kappa-B essential modulator (NEMO) deficiency provokes not only a variable set of deficiencies of both innate and adaptive immunity but also mild osteopetrosis, lymphedema, and, more frequently, anhydrotic ectodermal dysplasia, dysmorphic facies, and abnormal conical teeth. The last finding is often helpful in the diagnosis of that condition.

PRIMARY IMMUNODEFICIENCIES OF THE ADAPTIVE IMMUNE SYSTEM

a) T cell primary immunodeficiencies. Reticular dysgenesis, a rare severe combined immunodeficiency (SCID) characterized by T lymphopenia and agranulocytosis, can cause sensorineural deafness. Coronin A deficiency is another SCID variant that can be associated with behavioral disorders because the Coronin A gene is located in a genome area known to have been deleted in some patients with this disorder. The lack of enzymes of purine metabolism (adenosine deaminase and purine nucleoside phosphorylase) provokes not only profound T cell lymphocytopenia but also neurologic impairment, including dysautonomia and abnormalities of cognitive development of variable intensity, in many patients. The neurologic impairment can persist after hematopoietic stem cell transplantation (HSCT). Mild chondrodysplasia is a common finding in adenosine deaminase (ADA) deficiency and, indeed, can help the physician arrive at a final diagnosis.

b) **Primary thymic defects.** DiGeorge syndrome is a complex embryopathy that is caused by hemizygous interstitial deletion of chromosome 22, leading to multiple developmental defects, including conotruncal defects, hypoparathyroidism, and dysmorphic syndrome. Although a profound T cell immunodeficiency is rare in DiGeorge syndrome (~1% of cases), failure to recognize this feature is likely to have a fatal outcome. Similarly, some forms of the related CHARGE syndrome (mutation of the *CHD7* gene) also cause a profound T cell immunodeficiency.

c) **T cell primary immunodeficiencies related to calcium influx defects.** Very recently, rare T cell PIDs were found to be caused by defective store-operated entry of calcium ions into T and B lymphocytes after antigen stimulation. These defects (caused by ORA-1 and STIM-1 deficiencies) also lead to anhydrotic ectodermal dysplasia, abnormal teeth, and, above all, a nonprogressive muscle

disease characterized by excessive fatigue.

d) **DNA repair defects.** Several genetic defects impair DNA repair pathways. Many lead to combined T and B lymphocyte PIDs in a syndromal setting of varying complexity. The most common is ataxia telangiectasia (AT), an autosomal recessive disorder with an incidence of 1 in 40,000 live births; AT causes a B cell immunodeficiency (low IgA, IgG2 deficiency, and low antibody production) that often requires immunoglobulin replacement therapy.

AT is associated with a progressive T cell immunodeficiency. As the condition's name suggests, the hallmark features are telangiectasia and cerebellar ataxia.

These manifestations may not be detectable before age 3–4 years, and so AT should be considered in young children with IgA deficiency and problematic infections. Diagnosis is based on a cytogenetic analysis showing excessive chromosomal rearrangements (mostly affecting chromosomes 7and 14) in lymphocytes. AT is caused by mutation of the gene encoding the ATM protein, a kinase that plays a major role in the detection of DNA lesions and the organization of DNA repair (or cell death if the lesions are too numerous) by triggering several different pathways. Overall, AT is a progressive disease that carries a very high risk of lymphoma, leukemia, and (during adulthood) carcinomas. A variant of AT (AT-like disease) is caused by mutation of the *MRE11* gene.

Nijmegen breakage syndrome (NBS) is a less common condition that also results from chromosome instability (and the same cytogenetic abnormalities as in AT). It is characterized by a severe T and B cell combined immunodeficiency with autosomal recessive inheritance. Subjects with NBS exhibit microcephaly and a birdlike face but neither ataxia nor telangiectasia. The risk of malignancies is also very high. Nijmegen breakage syndrome results from a deficiency in Nibrin (NBS1, a protein associated with MRE11 and Rad50 that is involved in checking DNA lesions) caused by hypomorphic mutations.

Severe forms of dyskeratosis congenita (also known as Hoyeraal-Hreidersson syndrome) combine a progressive immunodeficiency that can include an absence of B and natural killer cell (NK) lymphocytes, progressive bone marrow failure, microcephaly, in utero growth retardation, and gut disease. The disease can be X-linked or, more rarely, autosomal recessive. It is caused by the mutation of genes encoding telomere maintenance proteins, including dyskerin (DKC1).

Bloom syndrome (helicase deficiency) combines a typical dysmorphic syndrome with growth retardation, skin lesions, and a mild immunodeficiency that also can be found in some patients with Fanconi's anemia.

Rare forms of combined T and B cell immunodeficiencies with autosomal recessive inheritance are associated in more complex syndromes with microcephaly, failure to grow, and a variable dysmorphic syndrome. These disorders are caused by mutation of the genes that encode DNA ligase 4 and Cernunnos (XLF), both of which are members of the nonhomologous end-joining DNA repair pathway.

Lastly, immunodeficiency, centromere instability, and facial anomalies (ICF) syndrome is a complex autosomal recessive syndrome that variably combines a mild T cell immunodeficiency and a more severe B cell immunodeficiency with a coarse face, intestinal disease, and mild mental retardation. A cytogenetic diagnostic feature is the presence of multiradial chromosomes (most frequently chromosomes 1, 9, and 16) caused by DNA defective methylation. The syndrome is a result of DNA methyltransferase DNMT3B deficiency.

e) **Growth hormone insensitivity syndrome (Laron dwarfism) with combined primary immunodeficiency.** Mutations in the *STAT5b* gene, which encodes a transcription factor involved in signaling downstream of the growth hormone receptor and the interleukin 2 (IL-2) receptor, lead to susceptibility to infection because of a partial, functional T cell immunodeficiency associated with autoimmune manifestations. The autoimmune manifestations probably result from defective generation/activation of regulatory T cells.

f) **Hyper-IgE syndrome (autosomal dominant form).** Hyper-IgE syndrome is a complex disorder that combines skin infections, inflammation, and susceptibility to bacterial and fungal infections of skin and lungs, often with pneumatoceles, with characteristic syndromic signs such as facial dysmorphy, defective loss of primary teeth, hyperextensibility, scoliosis, and osteoporosis. Elevated serum IgE levels are typical of hyper-IgE syndrome. The recently reported defects in TH17 effector responses account, at least in part, for the vulnerability to specific infections. This condition is caused by heterozygous (dominant) mutation of the gene encoding the transcription factor STAT3, which is required in a number of signaling pathways downstream of cytokine/cytokine receptor interactions (notably for IL-6).

g) **Primary immunodeficiencies with bone disease.** The autosomal recessive cartilage hair hypoplasia (CHH) disease is characterized by short-limb dwarfism, metaphyseal dysostosis, and sparse hair, together with a combined T and B cell PID of variable intensity, ranging from quasi-SCID to an absence of clinically significant immunodeficiency. The condition can predispose to erythroblastopenia, autoimmunity, and tumors. It is caused by mutations in the *RMRP* gene for a noncoding ribosome-associated RNA.

Schimke immunoosseous dysplasia is a rare autosomal recessive condition characterized by severe T and B cell immunodeficiency with spondyloepiphyseal dysplasia, growth retardation, and kidney and vascular diseases. It is the consequence of mutations in the *SMARCAL1* gene. The function of the gene product may be related to DNA repair.

h) **Venoocclusive disease with immunodeficiency (VODI syndrome)** is a rare autosomal recessive condition predominantly found in populations originating from Lebanon. It combines severe hepatic venoocclusive disease with usually mild T cell immunodeficiency and panhypogammaglobulinemia. It is caused by a deficiency in a nuclear protein, Sp110.

B CELL PRIMARY IMMUNODEFICIENCIES

Hypogammaglobulinemia can be associated with chromosomal defects such as trisomy 18 and Jacobsen syndrome (hemizygous deletion of part of the long arm of chromosome 11). A rare biallelic deficiency of the mismatch repair protein PMS2 leads to a partial deficiency in Ig class switch recombination in patients at a very high risk of cancer in general and colon carcinomas and lymphomas in particular. Transcobalamine deficiency disturbs vitamin B12 transport and therefore impairs hematopoiesis. Hypogammaglobulinemia is easily corrected by vitamin B_{12} administration and can be a characteristic of this very rare disorder.

PRIMARY IMMUNODEFICIENCIES AFFECTING REGULATORY PATHWAYS

Several inherited disorders that lead to hemophagocytic lymphohistiocytosis (HLH) also have features that are important in terms of both diagnosis and prognosis. Three of these disorders—Griscelli syndrome, Chédiak-Higashi syndrome, and the Hermansky-Pudlak type II syndrome—are characterized by partial albinism and silvery hair appearance that can facilitate diagnosis. Hermansky-Pudlak type II also can be a bleeding disorder if platelet aggregation is defective. Chediak-Higashi syndrome also is characterized by an early-onset progressive neurologic disorder with impaired cognitive development and motor and sensory deficiencies, culminating in a generalized encephalopathy. The encephalopathy is not prevented or arrested by allogeneic HSCT even when the HLH risk is controlled.

PRIMARY IMMUNODEFICIENCIES ASSOCIATED WITH OTHER CONDITIONS

A number of conditions can cause PIDs indirectly. For example, hypercatabolism in patients with Steinert's disease may cause hypogammaglobulinemia. Intestinal lymphangiectasia that includes both immunoglobulin and naive T cell loss and can expose the patient to a significant infectious risk. Urinary IgG loss may result from severe nephritic syndromes.

A number of drugs, including antimalarials, captopril, penicilla-mine, phenytoin, and sulfazaline, can induce predominantly IgA hypogammaglobulinemia in (probably predisposed) adults.

One also should also consider (1) diseases that are not thought to be PIDs but include the occurrence of recurrent infections and (2) genetic defects of the immune system that lead to other clinical manifestations. A very good example of the first group is cystic fibrosis (CF). Despite having a functionally normal immune system, patients with CF develop protracted bacterial respiratory tract infections, notably *Pseudomonas aeruginosa* colonization. This bacterium can incapacitate innate immune responses and cause unremitting inflammation that further facilitates infection. An example of the second group is primary alveolar proteinosis, which is caused by a defect in surfactant clearance by alveolar macrophages. The condition results from mutation of the gene encoding the granulocyte-macrophage colony-stimulating factor receptor α.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ALLERGIES, ANAPHYLAXIS, AND SYSTEMIC MASTOCYTOSIS: INTRODUCTION

The term *atopic allergy* implies a familial tendency to manifest such conditions as asthma, rhinitis, urticaria, and eczematous dermatitis (atopic dermatitis) alone or in combination, and in association with the presence of IgE. However, individuals without an atopic background may also develop hypersensitivity reactions, particularly urticaria and anaphylaxis, associated with the presence of IgE. Inasmuch as the mast cell is the key effector cell of the biologic response in allergic rhinitis, urticaria, anaphylaxis, and systemic mastocytosis, its developmental biology, activation pathway, product profile, and target tissues will be considered in the introduction to these clinical disorders.

The binding of IgE to human mast cells and basophils, a process termed *sensitization*, prepares these cells for subsequent antigen-specific activation. The sensitization of the high-affinity Fc receptor for IgE, designated FcεRI, also stabilizes the cellular expression of the receptor. FcεRI is composed of one α, one β, and two disulfide-linked γ chains, which together cross the plasma membrane seven times. The α chain is responsible for IgE binding, and the β and γ chains provide for signal transduction that follows the aggregation of the sensitized tetrameric receptors by polymeric antigen. Signal transduction is initiated through the action of an Src family–related tyrosine kinase, termed *Lyn*, that is constitutively associated with the β chain. Lyn transphosphorylates the canonical immunoreceptor tyrosine-based activation motifs (ITAMs) of the β and γ chains of the receptor, resulting in recruitment of more active Lyn to the β chain and of Syk tyrosine kinase. The phosphorylated tyrosines in the ITAMs function as binding sites for the tandem *src* homology two (SH2) domains within Syk. Syk activates not only phospholipase Cγ, which associates with the Linker of Activated T Cells at the plasma membrane, but also phosphatidylinositol 3-kinase to provide phosphatidylinositol-3,4,5-trisphosphate, which allows membrane targeting of the Tec family kinase Btk and its activation by Lyn. In addition, the Src family tyrosine kinase Fyn becomes activated after aggregation of IgE receptors and phosphorylates the adapter protein Gab2 that enhances activation of phosphatidylinositol 3-kinase. Indeed, this additional input is essential for mast cell activation, but it can be partially inhibited by Lyn, indicating that the extent of mast cell activation is in part regulated by the interplay between these Src family kinases. Activated phospholipase Cγ cleaves phospholipid membrane substrates to provide inositol-1,4,5-trisphosphate (IP_3) and 1,2-diacylglycerols (1,2-DAGs) so as to mobilize intracellular calcium and activate protein kinase C, respectively. The subsequent opening of calcium-regulated activated channels provides the sustained elevations of intracellular calcium required to recruit the mitogen-activated protein kinases, ERK, JNK, and p38 (serine/threonine kinases), which provide cascades to augment arachidonic acid release and to mediate nuclear translocation of transcription factors for various cytokines. The calcium ion–dependent activation of phospholipases cleaves membrane phospholipids to generate lysophospholipids, which, like 1,2-DAG, may facilitate the fusion of the secretory granule perigranular membrane with the cell membrane, a step that releases the membrane-free granules containing the preformed mediators of mast cell effects.

The secretory granule of the human mast cell has a crystalline structure, unlike mast cells of lower species. IgE-dependent cell activation results in solubilization and swelling of the granule contents within the first minute of receptor perturbation; this reaction is followed by the ordering of intermediate filaments about the swollen granule, movement of the granule toward the cell surface, and fusion of the perigranular membrane with that of other granules and with the plasmalemma to form extracellular channels for mediator release while maintaining cell viability.

In addition to exocytosis, aggregation of FcεRI initiates two other pathways for generation of bioactive products, namely, lipid mediators and cytokines. The biochemical steps involved in expression of such cytokines as tumor necrosis factor α (TNF-α), interleukin (IL) 1, IL-6, IL-4, IL-5, IL-13, granulocyte-macrophage colony-stimulating factor (GM-CSF), and others, including an array of chemokines, have not been specifically defined for mast cells. Inhibition studies of cytokine production (IL-1β, TNF-α, and IL-6) in mouse mast cells with cyclosporine or FK506 reveal binding to the ligand-specific immunophilin and attenuation of the calcium ion- and calmodulin-dependent serine/threonine phosphatase, calcineurin.

Lipid mediator generation **(Fig. 317-1)** involves translocation of calcium ion–dependent cytosolic phospholipase A_2 to the outer nuclear membrane, with subsequent release of arachidonic acid for metabolic processing by the distinct prostanoid and leukotriene pathways. The constitutive prostaglandin endoperoxide synthase-1 (PGHS-1/cyclooxygenase-1) and the de novo inducible PGHS-2 (cyclooxygenase-2) convert released arachidonic acid to the sequential intermediates, prostaglandins G_2 and H_2. The glutathione-dependent hematopoietic prostaglandin D_2 (PGD_2) synthase then converts PGH_2 to PGD_2, the predominant mast cell prostanoid. The PGD_2 receptors, DP_1 and DP_2, are distributed to smooth muscle as well as to T_H2 lymphocytes, eosinophils, and basophils implicated in allergic inflammation.

Figure 317-1

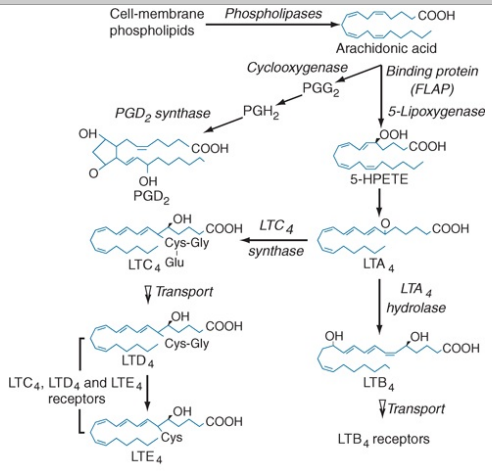

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathways for biosynthesis and release of membranederived lipid mediators from mast cells. In the 5-lipoxygenase pathway leukotriene A_4 (LTA_4) is the intermediate from which the terminal-pathway enzymes generate the distinct final products, leukotriene C_4 (LTC_4) and leukotriene B_4 (LTB_4), which leave the cell by separate saturable transport systems. Gamma glutamyl transpeptidase and a dipeptidase then cleave glutamic acid and glycine from LTC_4 to form LTD_4 and LTE_4, respectively. The major mast cell product of the cyclooxygenase system is PGD_2.

For the leukotriene biosynthetic pathway, the released arachidonic acid is metabolized by 5-lipoxygenase (5-LO) in the presence of an integral nuclear membrane protein, the 5-LO activating protein (FLAP). The calcium ion–dependent translocation of 5-LO to the nuclear membrane converts the arachidonic acid to the sequential intermediates, 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and leukotriene (LT) A_4. LTA_4 is conjugated with reduced glutathione by LTC_4 synthase, an integral nuclear membrane protein homologous to FLAP. Intracellular LTC_4 is released by a carrier-specific export step for extracellular metabolism to the additional cysteinyl leukotrienes, LTD_4 and LTE_4, by the sequential removal of glutamic acid and glycine. Alternatively, cytosolic LTA_4 hydrolase converts some LTA_4 to the dihydroxy leukotriene LTB_4, which also undergoes specific export. Two receptors for LTB_4, BLT_1 and BLT_2, mediate chemotaxis of human neutrophils. Two receptors for the cysteinyl leukotrienes, $CysLT_1$ and $CysLT_2$, are present on smooth muscle of the airways and the microvasculature and on hematopoietic cells such as macrophages, eosinophils, and mast cells. Whereas the $CysLT_1$ receptor has a preference for LTD_4 and is blocked by the receptor antagonists in clinical use, the $CysLT_2$ receptor is equally responsive to LTD_4 and LTC_4, is unaffected by these antagonists, and is a negative regulator of the function of the $CysLT_1$ receptor. The lysophospholipid formed during the release of arachidonic acid from 1-*O*-alkyl-2-acyl-*sn*-glyceryl-3-phosphorylcholine can be acetylated in the second position to form platelet-activating factor (PAF).

Unlike most other cells of bone marrow origin, mast cells leave the marrow and circulate as committed progenitors lacking their definitive secretory granules and characteristic FcεRI. These committed progenitors express the receptor, c-*kit*, for stem cell factor (SCF), and unlike other lineages, they retain and increase its expression with maturation. The SCF interaction with c-*kit* is an absolute requirement for the development of constitutive tissue mast cells residing in skin and

connective tissue sites and for the T_H2 cell-dependent comitogenesis providing mast cells to mucosal surfaces. Indeed, in clinical T cell deficiencies, mast cells are absent from the intestinal mucosa but are present in the submucosa. Based on the immunodetection of secretory granule neutral proteases, mast cells in the lung parenchyma and intestinal mucosa selectively express tryptase, and those in the intestinal and airway submucosa, skin, lymph nodes, and breast parenchyma express tryptase, chymase, and carboxypeptidase A (CPA). The secretory granules of mast cells selectively positive for tryptase exhibit closed scrolls with a periodicity suggestive of a crystalline structure by electron microscopy; whereas the secretory granules of mast cells with multiple proteases are scroll-poor, with an amorphous or lattice-like appearance.

Mast cells are distributed at cutaneous and mucosal surfaces and in submucosal tissues about venules and could influence the entry of foreign substances by their rapid response capability **(Fig. 317-2)**. Upon stimulus-specific activation and secretory granule exocytosis, histamine and acid hydrolases are solubilized, whereas the neutral proteases, which are cationic, remain largely bound to the anionic proteoglycans, heparin and chrondroitin sulfate E, with which they function as a complex. Histamine and the various lipid mediators (PGD_2, $LTC_4/D_4/E_4$, PAF) alter venular permeability, thereby allowing influx of plasma proteins such as complement and immunoglobulins, whereas LTB_4 mediates leukocyte-endothelial cell adhesion and subsequent directed migration (chemotaxis). The accumulation of leukocytes and plasma opsonins would facilitate defense of the microenvironment. The inflammatory response can also be detrimental, as in bronchial asthma, where the smooth-muscle constrictor activity of the cysteinyl leukotrienes is evident and much more potent than that of histamine.

Figure 317-2

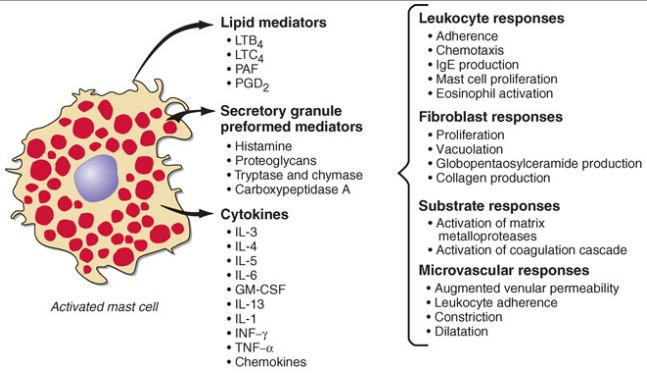

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bioactive mediators of three categories generated by IgE-dependent activation of murine mast cells can elicit common but sequential target cell effects leading to acute and sustained inflammatory responses. LT, leukotriene; PAF, platelet-activating factor; PGD_2, prostaglandin D_2; IL, interleukin; GM-CSF, granulocyte-macrophage colony-stimulating factor; INF, interferon; TNF, tumor necrosis factor.

The cellular component of the mast cell-mediated inflammatory response would be augmented and sustained by cytokines and chemokines of mast cell origin. IgE-dependent activation of human skin mast cells in situ elicits TNF-α production and release, which in turn induces endothelial cell responses favoring leukocyte adhesion. Similarly, activation of purified human lung mast cells or cord blood-derived cultured mast cells in vitro results in substantial production of proinflammatory (TNF-α) and immunomodulatory cytokines (IL-4, IL-5, IL-13) and chemokines. Bronchial biopsy specimens from patients with bronchial asthma reveal that mast cells are immunohistochemically positive for IL-4 and IL-5, but that the predominant localization of IL-4, IL-5, and GM-CSF is to T cells, defined as T_H2 by this profile. IL-4 modulates the T cell phenotype to the T_H2 subtype, determines the isotype switch to IgE (as does IL-13), and upregulates FcεRI-mediated expression of cytokines by mast cells.

An immediate and late cellular phase of allergic inflammation can be induced in the skin, nose, or lung of some allergic humans with local allergen challenge. In the immediate phase of a local challenge, there is pruritus and watery discharge from the nose, bronchospasm and mucus secretion in the lungs, and a wheal-and-flare response with pruritus in the skin. The reduced nasal patency, reduced pulmonary function, or evident erythema with swelling at the skin site in a late-phase response at 6-8 hours is associated with biopsy findings of infiltrating and activated T_H2 cells, eosinophils, basophils, and even some neutrophils. This allergic inflammation proceeding from early mast cell activation to late cellular infiltration has been used as an experimental surrogate of perennial rhinitis or bronchial asthma. However, in bronchial asthma, there is an intrinsic hyperreactivity of the airways independent of the associated inflammation.

Consideration of the mechanism of immediate-type hypersensitivity diseases in the human has focused largely on the IgE-dependent recognition of otherwise nontoxic substances. A region of chromosome 5 (5q23-31) contains genes implicated in the control of IgE levels including IL-4 and IL-13, as well as IL-3 and IL-9, which are involved in mucosal mast cell hyperplasia, and IL-5 and GM-CSF, which are central to eosinophil development and their enhanced tissue viability. Genes with linkage to the specific IgE response to particular allergens include those encoding the major histocompatibility complex (MHC) and certain chains of the T cell receptor (TCR-αδ). The complexity of atopy and the associated diseases includes susceptibility, severity, and therapeutic responses, each of which is among the separate variables modulated by both innate and adaptive immune stimuli.

The induction of allergic disease requires sensitization of a predisposed individual to a specific allergen. The greatest propensity for the development of atopic allergy occurs in childhood and early adolescence. The allergen is processed by antigen-presenting cells of the monocytic lineage located throughout the body at surfaces that contact the outside environment, such as the nose, lungs, eyes, skin, and intestine. These antigen-presenting cells present the epitope-bearing peptides via their MHC to T helper cells and their subsets. The T cell response depends both on cognate recognition and on the cytokine microenvironment provided by the antigen-presenting dendritic cells, with IL-4 directing a T_H2 subset, interferon (IFN) γ a T_H1 profile, and IL-6 with transforming growth factor β (TGFβ) a T_H17 subset. Allergens not only present antigenic epitopes via dendritic cells but also contain pattern recognition ligands that facilitate the immune response by direct initiation of cytokine generation from innate cell types such as basophils, mast cells, eosinophils, and others. The T_H2 response is associated with activation of specific B cells that can also present allergens or that transform into plasma cells for antibody production. Synthesis and release into the plasma of allergen-specific IgE results in sensitization of FcεR1-bearing cells such as mast cells and basophils, which become activated on exposure to the specific allergen. In certain diseases, including those associated with atopy, the monocyte and eosinophil populations can express a trimeric FcεR1, which lacks the β chain, and yet respond to its aggregation.

ANAPHYLAXIS

Definition

The life-threatening anaphylactic response of a sensitized human appears within minutes after systemic exposure to specific antigen and is manifested by respiratory distress due to laryngeal edema and/or intense bronchospasm, often followed by vascular collapse, or by shock without antecedent respiratory difficulty. Cutaneous manifestations exemplified by pruritus and urticaria with or without angioedema are characteristic of such systemic anaphylactic reactions. Gastrointestinal manifestations include nausea, vomiting, crampy abdominal pain, and diarrhea.

Predisposing Factors and Etiology

There is no convincing evidence that age, sex, race, or geographic location predisposes a human to anaphylaxis except through exposure to some immunogen. According to most studies, atopy does not predispose individuals to anaphylaxis from penicillin therapy or venom of a stinging insect but is a risk factor for allergens in food or latex.

The materials capable of eliciting the systemic anaphylactic reaction in humans include the following: heterologous proteins in the form of hormones (insulin, vasopressin, parathormone); enzymes (trypsin, chymotrypsin, penicillinase, streptokinase); pollen extracts (ragweed, grass, trees); nonpollen allergen extracts (dust mites, dander of cats, dogs, horses, and laboratory animals); food (peanuts, milk, eggs, seafood, nuts, grains, beans, gelatin in capsules); monoclonal antibodies; occupation-related products (latex rubber products); Hymenoptera venom (yellow jacket, yellow and baldfaced hornets, paper wasp, honey bee, imported fire ants); polysaccharides such as dextran and thiomersal as a vaccine preservative; drugs such as protamine; antibiotics (penicillins, cephalosporins, amphotericin B, nitrofurantoin, quinolones); chemotherapy agents (carboplatin, paclitaxel, doxorubicin); local anesthetics (procaine, lidocaine); muscle relaxants (suxamethonium, gallamine, pancuronium); vitamins (thiamine, folic acid); diagnostic agents (sodium dehydrocholate, sulfobromophthalein); and occupation-related chemicals (ethylene oxide). Drugs are considered to function as haptens that form immunogenic conjugates with host proteins. The conjugating hapten may be the parent compound, a nonenzymatically derived storage product, or a metabolite formed in the host.

Pathophysiology and Manifestations

Individuals differ in the time of appearance of symptoms and signs, but the hallmark of the anaphylactic reaction is the onset of some manifestation within seconds to minutes after introduction of the antigen, generally by injection or less commonly by ingestion. There may be upper or lower airway obstruction or both. Laryngeal edema may be experienced as a "lump" in the throat, hoarseness, or stridor, while bronchial obstruction is associated with a feeling of tightness in the chest and/or audible wheezing. Patients with bronchial asthma are predisposed to severe involvement of the lower airways. Flushing with diffuse erythema and a feeling of warmth may occur. A characteristic feature is the eruption of well-circumscribed, discrete cutaneous wheals with erythematous, raised, serpiginous borders and

blanched centers. These urticarial eruptions are intensely pruritic and may be localized or disseminated. They may coalesce to form giant hives, and they seldom persist beyond 48 h. A localized, nonpitting, deeper edematous cutaneous process, angioedema, may also be present. It may be asymptomatic or cause a burning or stinging sensation.

In fatal cases with clinical bronchial obstruction, the lungs show marked hyperinflation on gross and microscopic examination. The microscopic findings in the bronchi, however, are limited to luminal secretions, peribronchial congestion, submucosal edema, and eosinophilic infiltration, and the acute emphysema is attributed to intractable bronchospasm that subsides with death. The angioedema resulting in death by mechanical obstruction occurs in the epiglottis and larynx, but the process also is evident in the hypopharynx and to some extent in the trachea. On microscopic examination, there is wide separation of the collagen fibers and the glandular elements; vascular congestion and eosinophilic infiltration also are present. Patients dying of vascular collapse without antecedent hypoxia from respiratory insufficiency have visceral congestion with a presumptive loss of intravascular blood volume. The associated electrocardiographic abnormalities, with or without infarction, in some patients may reflect a primary cardiac event mediated by mast cells or may be secondary to a critical reduction in blood volume.

The angioedematous and urticarial manifestations of the anaphylactic syndrome have been attributed to the release of endogenous histamine. A role for the cysteinyl leukotrienes in causing marked bronchiolar constriction seems likely. Vascular collapse without respiratory distress in response to experimental challenge with the sting of a hymenopteran was associated with marked and prolonged elevations in blood histamine and intravascular coagulation and kinin generation. The finding that patients with systemic mastocytosis and episodic vascular collapse excrete large amounts of PGD_2 metabolites in addition to histamine suggests that PGD_2 is also of importance in the hypotensive anaphylactic reactions. The PAF level can be elevated in the serum of patients with anaphylaxis, and its concentration is correlated inversely with the constitutive level of the acetylhydrolase involved in its inactivation. The actions of the array of mast cell-derived mediators are likely additive or synergistic at the target tissues.

Diagnosis

The diagnosis of an anaphylactic reaction depends on a history revealing the onset of symptoms and signs within minutes after the responsible material is encountered. It is appropriate to rule out a complement-mediated immune complex reaction, an idiosyncratic response to a nonsteroidal anti-inflammatory drug (NSAID), or the direct effect of certain drugs or diagnostic agents on mast cells. Intravenous administration of a chemical mast cell-degranulating agent, including opiate derivatives and radiographic contrast media, may elicit generalized urticaria, angioedema, and a sensation of retrosternal oppression with or without clinically detectable bronchoconstriction or hypotension. Aspirin and other NSAIDs such as indomethacin, aminopyrine, and mefenamic acid may precipitate a life-threatening episode of obstruction of upper or lower airways, especially in patients with asthma, that is clinically indistinguishable from anaphylaxis but is not associated with the presence of specific IgE or elevation of blood tryptase. This syndrome, which is commonly associated with nasal polyposis, is due to inhibition of PGHS-1 with corresponding unregulated, amplified generation of the cysteinyl leukotrienes via the 5-LO/LTC_4 synthase pathway. In the transfusion anaphylactic reaction that occurs in patients with IgA deficiency, the responsible specificity resides in IgG or IgE anti-IgA; the mechanism of the reaction mediated by IgG anti-IgA is presumed to be complement activation with secondary mast cell participation.

The presence of specific IgE in the blood of patients with systemic anaphylaxis was demonstrated historically by passive transfer of the serum intradermally into a normal recipient, followed 24 h later by antigen challenge into the same site, with subsequent development of a wheal and flare (the *Prausnitz-Küstner reaction*). In current clinical practice, immunoassays using purified antigens can demonstrate the presence of specific IgE in the serum of patients with anaphylactic reactions, and intracutaneous skin testing may be performed after the patient has recovered to elicit a local wheal and flare in response to the putative antigen. Elevations of tryptase levels in serum implicate mast cell activation in a systemic reaction and are particularly informative for anaphylaxis with episodes of hypotension during general anesthesia or when there has been a fatal outcome. However, because of the short half-life of tryptase, elevated levels are best detected within 4 hours of a systemic reaction.

Treatment: Anaphylaxis

Early recognition of an anaphylactic reaction is mandatory, since death occurs within minutes to hours after the first symptoms. Mild symptoms such as pruritus and urticaria can be controlled by administration of 0.3 to 0.5 mL of 1:1000 (1 mg/mL) epinephrine SC or IM, with repeated doses as required at 5- to 20-min intervals for a severe reaction. If the antigenic material was injected into an extremity, the rate of absorption may be reduced by prompt application of a tourniquet proximal to the reaction site, administration of 0.2 mL of 1:1000 epinephrine into the site, and removal without compression of an insect stinger, if present. An IV infusion should be initiated to provide a route for administration of 2.5 mL epinephrine, diluted 1:10,000, at 5- to 10-min intervals, volume expanders such as normal saline, and vasopressor agents such as dopamine if intractable hypotension occurs. Replacement of intravascular volume due to postcapillary venular leakage may require several liters of saline. Epinephrine provides both a- and β-adrenergic effects, resulting in vasoconstriction, bronchial smooth-muscle relaxation, and attenuation of enhanced venular permeability. When epinephrine fails to control the anaphylactic reaction, hypoxia due to airway obstruction or related to a cardiac arrhythmia, or both, must be considered. Oxygen alone via a nasal catheter or with nebulized albuterol may be helpful, but either endotracheal intubation or a tracheostomy is mandatory for oxygen delivery if progressive hypoxia develops. Ancillary agents such as the antihistamine diphenhydramine, 50-100 mg IM or IV, and aminophylline, 0.25-0.5 g IV, are appropriate for urticaria-angioedema and bronchospasm, respectively. Intravenous glucocorticoids, 0.5-1 mg/kg of medrol, are not effective for the acute event but may alleviate later recurrence of bronchospasm, hypotension, or urticaria.

Prevention

Prevention of anaphylaxis must take into account the sensitivity of the recipient, the dose and character of the diagnostic or therapeutic agent, and the effect of the route of administration on the rate of absorption. Beta blockers are relatively contraindicated in persons at risk for anaphylactic reactions, especially those sensitive to Hymenoptera venom or those undergoing immunotherapy for respiratory system allergy. If there is a definite history of a past anaphylactic reaction to a medication, even though mild, it is advisable to select a structurally unrelated agent. A knowledge of cross-reactivity among agents is critical since, for example, cephalosporins have a cross-reactive ring structure with the penicillins. When skin testing, a prick or scratch skin test should precede an intradermal test, since the latter has a higher risk of causing anaphylaxis. These tests should be performed before the administration of certain materials that are likely to elicit anaphylactic reactions, such as allergenic extracts. Skin testing for antibiotics or chemotherapeutic agents should be performed only on patients with a positive clinical history consistent with an IgE-mediated reaction and in imminent need of the antibiotic in question; skin testing is of no value for non-IgE-mediated eruptions. With regard to penicillin, two-thirds of patients with a positive reaction history and positive skin tests to benzylpenicilloyl-polylysine (BPL) and/or the minor determinant mixture (MDM) of benzylpenicillin products experience allergic reactions with treatment, and these reactions are almost uniformly of the anaphylactic type in those patients with minor determinant reactivity. Even patients without a history of previous clinical reactions have a 2-6% incidence of positive skin tests to the two test materials, and about 3 per 1000 with a negative history experience anaphylaxis with therapy, with a mortality of about 1 per 100,000.

If an agent carrying a risk of an anaphylactic response is required because a non-cross-reactive alternative is not available, desensitization can be performed with most antibiotics and other classes of therapeutic agents by the IV, SC, or oral route. Typically, graded quantities of the drug are given by the selected route starting below the threshold dose for an adverse reaction and then doubling each dose until a therapeutic dosage is achieved. Due to the risk of systemic anaphylaxis during the course of desensitization, such a procedure should be performed under the supervision of a specialist and in a setting in which resuscitation equipment is at hand and an IV line is in place. Once a desensitized state is achieved, it is critical to continue administration of the therapeutic agent at regular intervals throughout the treatment period to prevent the reestablishment of a significant pool of sensitized cells.

A different form of protection involves the development of blocking antibody of the IgG class, which protects against Hymenoptera venom-induced anaphylaxis by interacting with antigen so that less reaches the sensitized tissue mast cells. The maximal risk for systemic anaphylactic reactions in persons with Hymenoptera sensitivity occurs in association with a currently positive skin test. Although there is only low-grade cross-reactivity between honey bee and yellow jacket venoms, there is a high degree of cross-reactivity between yellow jacket venom and the rest of the vespid venoms (yellow or baldfaced hornets and wasps). Prevention involves modification of outdoor activities to exclude bare feet, wearing perfumed toiletries, eating in areas attractive to insects, clipping hedges or grass, and hauling away trash or fallen fruit. As with each anaphylactic sensitivity, the individual should wear an informational bracelet and have immediate access to an unexpired autoinjectable epinephrine kit. The limitations of lifestyle and the psychological duress can be addressed by venom immunotherapy. Although it has been recommended that venom therapy be continued indefinitely or until the skin and specific serum IgE tests are unremarkable, there is evidence that 5 years of treatment induces a state of resistance to sting reactions that is independent of serum levels of specific IgG or IgE. For children with a systemic reaction limited to skin, the likelihood of progression to more serious respiratory or vascular manifestations is low, and thus immunotherapy is not recommended.

URTICARIA AND ANGIOEDEMA

Definition

Urticaria and angioedema may appear separately or together as cutaneous manifestations of localized nonpitting edema; a similar process may occur at mucosal surfaces of the upper respiratory or gastrointestinal tract. *Urticaria* involves only the superficial portion of the dermis, presenting as well-circumscribed wheals with erythematous raised serpiginous borders and blanched centers that may coalesce to become giant wheals. *Angioedema* is a well-demarcated localized edema involving the deeper layers of the skin, including the subcutaneous tissue. Recurrent episodes of urticaria and/or angioedema of less than 6 weeks' duration are considered acute, whereas attacks persisting beyond this period are designated chronic.

Predisposing Factors and Etiology

Urticaria and angioedema probably occur more frequently than usually described because of the evanescent, self-limited nature of such eruptions, which seldom require medical attention when limited to the skin. Although persons in any age group may experience acute or chronic urticaria and/or angioedema, these lesions increase in frequency after adolescence, with the highest incidence occurring in persons in the third decade of life; indeed, one survey of college students indicated that 15-20% had experienced a pruritic wheal reaction.

The classification of urticaria-angioedema presented in **Table 317-1** focuses on the different mechanisms for eliciting clinical disease and can be useful for differential diagnosis; nonetheless, most cases of chronic urticaria are idiopathic. Urticaria and/or angioedema occurring during the appropriate season in patients with seasonal respiratory allergy or as a result of exposure to animals or molds is attributed to inhalation or physical contact with pollens, animal dander, and mold spores, respectively. However, urticaria and angioedema secondary to inhalation are relatively uncommon compared to urticaria and angioedema elicited by ingestion of fresh fruits, shellfish, fish, milk products, chocolate, legumes including peanuts, and various drugs that may elicit not only the anaphylactic syndrome with prominent gastrointestinal complaints but also chronic urticaria.

Table 317-1 Classification of Urticaria and/or Angioedema

1. IgE-dependent

a. Specific antigen sensitivity (pollens, foods, drugs, fungi, molds, Hymenoptera venom, helminths)
b. Physical: dermographism, cold, solar
c. Autoimmune

2. Bradykinin-mediated

a. Hereditary angioedema: C1 inhibitor deficiency: null (type 1) and dysfunctional (type 2)
b. Acquired angioedema: C1 inhibitor deficiency: anti-idiotype and anti-C1 inhibitor
c. Angiotensin-converting enzyme inhibitors

3. Complement-mediated

a. Necrotizing vasculitis
b. Serum sickness
c. Reactions to blood products

4. Nonimmunologic

a. Direct mast cell-releasing agents (opiates, antibiotics, curare, D-tubocurarine, radiocontrast media)
b. Agents that alter arachidonic acid metabolism (aspirin and nonsteroidal anti-inflammatory agents, azo dyes, and benzoates)

5. Idiopathic

Additional etiologies include physical stimuli such as cold, heat, solar rays, exercise, and mechanical irritation. The physical urticarias can be distinguished by the precipitating event and other aspects of the clinical presentation. *Dermographism*, which occurs in 1-4% of the population, is defined by the appearance of a linear wheal at the site of a brisk stroke with a firm object or by any configuration appropriate to the eliciting event **(Fig. 317-3)**. Dermographism has a prevalence that peaks in the second to third decades. It is not influenced by an atopic diathesis and has a duration generally of <5 years. *Pressure urticaria*, which often accompanies chronic idiopathic urticaria, presents in response to a sustained stimulus such as a shoulder strap or belt, running (feet), or manual labor (hands). *Cholinergic urticaria* is distinctive in that the pruritic wheals are of small size (1-2 mm) and are surrounded by a large area of erythema; attacks are precipitated by fever, a hot bath or shower, or exercise and are presumptively attributed to a rise in core body temperature. *Exercise-related anaphylaxis* can be precipitated by exertion alone or can be dependent on prior food ingestion. The clinical presentation can be limited to flushing, erythema, and pruritic urticaria but may progress to angioedema of the face, oropharynx, larynx, or intestine or to vascular collapse; it is distinguished from cholinergic urticaria by presenting with wheals of conventional size and by not occurring with fever or a hot bath. *Cold urticaria* is local at body areas exposed to low ambient temperature or cold objects (ice cube) but can progress to vascular collapse with immersion in cold water (swimming). *Solar urticaria* is subdivided into six groups by the response to specific portions of the light spectrum. *Vibratory angioedema* may occur after years of occupational exposure or can be idiopathic; it may be accompanied by cholinergic urticaria. Other rare forms of physical allergy, always defined by stimulus-specific elicitation, include *local heat urticaria*, *aquagenic urticaria* from contact with water of any temperature (sometimes associated with polycythemia vera), and *contact urticaria* from direct interaction with some chemical substance.

Figure 317-3

Dermographism

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dermographic urticarial lesion induced by stroking the forearm lightly with the edge of a tongue blade. The photograph, taken after 2 minutes, demonstrates a prominent wheal and flare reaction in the shape of an *X. (Photograph provided by Allen P. Kaplan, MD, Medical University of South Carolina.)*

Angioedema without urticaria due to the generation of bradykinin occurs with C1 inhibitor (C1INH) deficiency that may be inborn as an autosomal dominant characteristic or may be acquired through the appearance of an autoantibody. The angiotensin-converting enzyme (ACE) inhibitors can provoke a similar clinical presentation in 0.1-0.5% of hypertensive patients due to attenuated degradation of bradykinin. The urticaria and angioedema associated with classic serum sickness or with hypocomplementemic cutaneous necrotizing angiitis are believed to be immune-complex diseases. The drug reactions to mast cell granule-releasing agents and to NSAIDs may be systemic, resembling anaphylaxis, or limited to cutaneous sites.

PATHOPHYSIOLOGY AND MANIFESTATIONS

Urticarial eruptions are distinctly pruritic, may involve any area of the body from the scalp to the soles of the feet, and appear in crops of 12- to 36-hour duration, with old lesions fading as new ones appear. Most of the physical urticarias (cold, cholinergic, dermatographism) are an exception, with individual lesions lasting less than 2 hours. The most common sites for urticaria are the extremities and face, with angioedema often being periorbital and in the lips. Although self-limited in duration, angioedema of the upper respiratory tract may be life-threatening due to laryngeal obstruction, while gastrointestinal involvement may present with abdominal colic, with or without nausea and vomiting, and may result in unnecessary surgical intervention. No residual discoloration occurs with either urticaria or angioedema unless there is an underlying process leading to superimposed extravasation of erythrocytes.

The pathology is characterized by edema of the superficial dermis in urticaria and of the subcutaneous tissue and deep dermis in angioedema. Collagen bundles in affected areas are widely separated, and the venules are sometimes dilated. Any perivenular infiltrate consists of lymphocytes, monocytes, eosinophils, and neutrophils that are present in varying combination and numbers.

Perhaps the best-studied example of IgE- and mast cell-mediated urticaria and angioedema is *cold urticaria*. Cryoglobulins or cold agglutinins may be recognized in up to 5% of these patients. Immersion of an extremity in an ice bath precipitates angioedema of the distal portion with urticaria at the air interface within minutes of the challenge. Histologic studies reveal marked mast cell degranulation with associated edema of the dermis and subcutaneous tissues. The histamine level in the plasma of venous effluent of the cold-challenged and angioedematous extremity is markedly increased, but no such increase appears in the plasma of effluent of the contralateral normal extremity. Elevated levels of histamine have been found in the plasma of venous effluent and in the fluid of suction blisters at experimentally induced lesional sites in patients with dermographism, pressure urticaria, vibratory angioedema, light urticaria, and heat urticaria. By ultrastructural analysis, the pattern of mast cell degranulation in cold urticaria resembles an IgE-mediated response with solubilization of granule contents, fusion of the perigranular and cell membranes, and discharge of granule contents, whereas in a dermographic lesion there is additional superimposed zonal (piecemeal) degranulation. Elevations of plasma histamine levels with biopsy-proven mast cell degranulation have also been demonstrated with generalized attacks of *cholinergic urticaria* and *exercise-related anaphylaxis* precipitated experimentally in subjects exercising on a treadmill while wearing a wet suit; however, only subjects with cholinergic urticaria have a concomitant decrease in pulmonary function.

Up to 40% of patients with chronic urticaria have an autoimmune cause for their disease including autoantibodies to IgE (5-10%) or, more commonly, to the a chain of FcεRI (35-45%). In these patients, autologous serum injected into their own skin can induce a wheal and flare reaction involving mast cell activation. The presence of these antibodies can also be recognized by their capacity to release histamine or induce activation markers such as CD63 or CD203 on basophils. An association with antibodies to microsomal peroxidase and/or thyroglobulin has been observed often with clinically significant Hashimoto's thyroiditis. In vitro studies reveal that these autoantibodies can mediate basophil degranulation with enhancement by serum as a source of the anaphylatoxic fragment, C5a.

Hereditary angioedema is an autosomal dominant disease due to a deficiency of C1INH (type 1) in about 85% of patients and to a dysfunctional protein (type 2) in

the remainder. In the acquired form of C1INH deficiency, there is excessive consumption due either to immune complexes formed between anti-idiotypic antibody and monoclonal IgG presented by B cell lymphomas or to an autoantibody directed to C1INH. C1INH blocks the catalytic function of activated factor XII (Hageman factor) and of kallikrein, as well as the C1r/C1s components of C1. During clinical attacks of angioedema, C1INH-deficient patients have elevated plasma levels of bradykinin, particularly in the venous effluent of an involved extremity, and reduced levels of prekallikrein and high-molecular-weight kininogen, from which bradykinin is cleaved. The parallel decline in the complement substrates C4 and C2 reflects the action of activated C1 during such attacks. Mice with targeted disruption of the gene for C1INH exhibit a chronic increase in vascular permeability. The pathobiology is aggravated by administration of an ACE inhibitor (captopril) and is attenuated by breeding the C1INH null strain to a bradykinin 2 receptor (Bk2R) null strain. As ACE is also described as kininase II, the use of blockers results in impaired bradykinin degradation and explains the angioedema that occurs idiosyncratically in hypertensive patients with a normal C1INH.

DIAGNOSIS

The rapid onset and self-limited nature of urticarial and angioedematous eruptions are distinguishing features. Additional characteristics are the occurrence of the urticarial crops in various stages of evolution and the asymmetric distribution of the angioedema. Urticaria and/or angioedema involving IgE-dependent mechanisms are often appreciated by historic considerations implicating specific allergens or physical stimuli, by seasonal incidence, and by exposure to certain environments. Direct reproduction of the lesion with physical stimuli is particularly valuable because it so often establishes the cause of the lesion. The diagnosis of an environmental allergen based on the clinical history can be confirmed by skin testing or assay for allergen-specific IgE in serum. IgE-mediated urticaria and/or angioedema may or may not be associated with an elevation of total IgE or with peripheral eosinophilia. Fever, leukocytosis, and an elevated sedimentation rate are absent.

The classification of urticarial and angioedematous states presented in Table 317-1 in terms of possible mechanisms necessarily includes some differential diagnostic points. Hypocomplementemia is not observed in IgE-mediated mast cell disease and may reflect either an acquired abnormality generally attributed to the formation of immune complexes or a genetic deficiency of C1INH. Chronic recurrent urticaria, generally in females, associated with arthralgias, an elevated sedimentation rate, and normo- or hypocomplementemia suggests an underlying cutaneous necrotizing angiitis. Vasculitic urticaria typically persists longer than 72 hours, whereas conventional urticaria often has a duration of 12-36 hours. Confirmation depends on a biopsy that reveals cellular infiltration, nuclear debris, and fibrinoid necrosis of the venules. The same pathobiologic process accounts for the urticaria in association with such diseases as systemic lupus erythematosus or viral hepatitis with or without associated arteritis. Serum sickness per se or a similar clinical entity due to drugs includes not only urticaria but also pyrexia, lymphadenopathy, myalgia, and arthralgia or arthritis. Urticarial reactions to blood products or intravenous administration of immunoglobulin are defined by the event and generally are not progressive unless the recipient is IgA-deficient in the former case or the reagent is aggregated in the latter.

The diagnosis of hereditary angioedema is suggested not only by family history but also by the lack of pruritus and of urticarial lesions, the prominence of recurrent gastrointestinal attacks of colic, and episodes of laryngeal edema. Laboratory diagnosis depends on demonstrating a deficiency of C1INH antigen (type 1) or a nonfunctional protein (type 2) by a catalytic inhibition assay. While levels of C1 are normal, its substrates, C4 and C2, are chronically depleted and fall further during attacks due to the activation of additional C1. Patients with the acquired forms of C1INH deficiency have the same clinical manifestations but differ in the lack of a familial element. Furthermore, their sera exhibit a reduction of C1 function and C1q protein as well as C1INH, C4, and C2. Inborn C1INH deficiency and ACE inhibitor-elicited angioedema are associated with elevated levels of bradykinin.

Urticaria and angioedema can be differentiated from contact sensitivity, a vesicular eruption that progresses to chronic thickening of the skin with continued allergenic exposure. They can also be differentiated from atopic dermatitis, a condition that may present as erythema, edema, papules, vesiculation, and oozing proceeding to a subacute and chronic stage in which vesiculation is less marked or absent and scaling, fissuring, and lichenification predominate in a distribution that characteristically involves the flexor surfaces. In cutaneous mastocytosis, the reddish brown macules and papules, characteristic of urticaria pigmentosa, urticate with pruritus upon trauma; and in systemic mastocytosis, without or with urticaria pigmentosa, there is episodic systemic flushing with or without urtication but no angioedema.

Treatment: Urticaria and Angioedema

Identification of the etiologic factor(s) and subsequent elimination provide the most satisfactory therapeutic program; this approach is feasible to varying degrees with IgE-mediated reactions to allergens or physical stimuli. For most forms of urticaria, H_1 antihistamines such as chlorpheniramine or diphenhydramine effectively attenuate both urtication and pruritus, but because of their side effects, nonsedating agents such as loratadine, desloratadine, and fexofenadine, or low-sedating agents such as cetirizine or levocetirizine generally are used first. Cyproheptadine in dosages beginning at 8 mg and ranging up to 32 mg daily and especially hydroxyzine in dosages beginning at 40 mg and ranging up to 200 mg daily have proven effective when H_1 antihistamines fail. The addition of an H2 antagonist such as cimetidine, ranitidine, or famotidine in conventional dosages may add benefit when H_1 antihistamines are inadequate. Doxepin, a dibenzoxepin tricyclic compound with both H_1 and H_2 receptor antagonist activity, is yet another alternative. A CysLT $_1$ receptor antagonist such as montelukast, 10 mg/d, or zafirlukast, 20 mg twice a day, can be important add-on therapy. Topical glucocorticoids are of no value, and systemic glucocorticoids are generally avoided in idiopathic, allergen-induced, or physical urticarias due to their long-term toxicity. Systemic glucocorticoids are useful in the management of patients with pressure urticaria, vasculitic urticaria (especially with eosinophil prominence), idiopathic angioedema with or without urticaria, or chronic urticaria that responds poorly to conventional treatment. With persistent vasculitic urticaria, hydroxychloroquine, dapsone, or colchicine may be added to the regimen after hydroxyzine and before or along with systemic glucocorticoids. Cyclosporine can be efficacious for patients with chronic idiopathic or chronic autoimmune urticaria that is severe and poorly responsive to other modalities and/or where a glucocorticoid requirement is excessive. For chronic urticaria induced by autoantibody activation of mast cells and basophils, monoclonal anti-IgE antibodies such as omalizumab may be considered.

The therapy of inborn C1INH deficiency has been simplified by the finding that attenuated androgens correct the biochemical defect and afford prophylactic protection; their efficacy is attributed to production by the normal gene of an amount of functional C1INH sufficient to control the spontaneous activation of C1. The antifibrinolytic agent ε-aminocaproic acid may be used for preoperative prophylaxis but is contraindicated in patients with thrombotic tendencies or ischemia due to arterial atherosclerosis. Infusion of isolated C1INH protein may be used for prophylaxis or treatment of an acute attack; a bradykinin 2 receptor antagonist and ecallantide, a kallikrein inhibitor, which are administered SC, are each being assessed for amelioration of attacks. For acquired C1INH deficiency, treatment of the underlying hematologicmalignancy is recommended.

SYSTEMIC MASTOCYTOSIS

Definition

Systemic mastocytosis is defined by a clonal expansion of mast cells that in most instances is indolent and nonneoplastic. The mast cell expansion is generally recognized only in bone marrow and in the normal peripheral distribution sites of the cells, such as skin, gastrointestinal mucosa, liver, and spleen. Mastocytosis occurs at any age and has a slight preponderance in males. The prevalence of systemic mastocytosis is not known, a familial occurrence is rare, and atopy is not increased.

Classification and Pathophysiology

A consensus classification for mastocytosis recognizes cutaneous mastocytosis with variants and four systemic forms **(Table 317-2)**. The form designated as *indolent systemic mastocytosis* (ISM) accounts for the majority of patients; it implies that there is no evidence of an associated hematologic disorder, liver disease, or lymphadenopathy and is not known to alter life expectancy. In *systemic mastocytosis associated with clonal hematologic non-mast cell lineage disease* (SM-AHNMD), the prognosis is determined by the nature of the associated disorder, which can range from dysmyelopoiesis to leukemia. In *aggressive systemic mastocytosis* (ASM), mast cell infiltration/proliferation in multiple organs such as liver, spleen, gut, CNS, and/or bone results in a poor prognosis; a subset of patients with this form has prominent eosinophilia with hepatosplenomegaly and lymphadenopathy. *Mast cell leukemia* is the rarest form of the disease and is invariably fatal at present; the peripheral blood contains circulating, metachromatically staining, atypical mast cells.

Table 317-2 Classification of Mastocytosis

Cutaneous mastocytosis (CM)
Urticaria pigmentosa (UP)/maculopapularcutaneous mastocytosis (MPCM)
Variants: plaque form, nodular form; telangiectasia macularis eruptiva perstans (TMEP); diffuse cutaneous mastocytosis (DCM)
Solitary mastocytoma of skin
Indolent systemic mastocytosis (ISM)
Systemic mastocytosis with an associated clonal hematologic non-mast cell lineage disease (SM-AHNMD)
Aggressive systemic mastocytosis (ASM)
Variant: lymphadenopathic mastocytosis with eosinophilia
Mast cell leukemia (MCL)
Mast cell sarcoma (MCS)

Source: Modified from SH Swerdlow et al (eds): *World Health Organization Classification of Tumors: Pathology and Genetics in Tumors of Hematopoietic and Lymphoid Tissues.* Lyon, IARC Press, 2008.

A point mutation of A to T at codon 816 of c- *kit* that causes an aspartic acid to valine substitution is found in multiple cell lineages in patients with mastocytosis, resulting in a somatic gain-in-function mutation. This substitution, as well as other rare mutations of c-*kit*, is characteristic of adults with SM-AHNMD but is also present in patients with ISM or cutaneous mastocytosis, as might be anticipated because mast cells are of bone marrow lineage. The prognosis for patients with cutaneous mastocytosis and for almost all with ISM is a normal life expectancy, while that for patients with SM-AHNMD is determined by a non-mast cell component. In infants and children with cutaneous manifestations, namely, urticaria pigmentosa or bullous lesions, visceral involvement is usually lacking, and resolution is

common because gain-in-function mutations are infrequent.

Clinical Manifestations

The clinical manifestations of systemic mastocytosis, distinct from a leukemic complication, are due to tissue occupancy by the mast cell mass, the tissue response to that mass, and the release of bioactive substances acting at both local and distal sites. The pharmacologically induced manifestations are pruritus, flushing, palpitations and vascular collapse, gastric distress, lower abdominal crampy pain, and recurrent headache. The increase in local cell burden is evidenced by the lesions of urticaria pigmentosa at skin sites and is a direct cause of bone pain and/or malabsorption. Mast cell-mediated fibrotic changes occur in liver, spleen, and bone marrow but not in gastrointestinal tissue or skin. Immunofluorescent analysis of bone marrow and skin lesions in ISM and of spleen, lymph node, and skin in ASM has revealed only one mast cell phenotype, namely, scroll-poor cells expressing tryptase, chymase, and CPA.

The cutaneous lesions of urticaria pigmentosa are reddish-brown macules or papules that respond to trauma with urtication and erythema (Darier's sign). The apparent incidence of these lesions is =90% in patients with ISM and <50% in those with SM-AHNMD or ASM. Approximately 1% of patients with ISM have skin lesions that appear as tan-brown macules with striking patchy erythema and associated telangiectasia (telangiectasia macularis eruptiva perstans). In the upper gastrointestinal tract, gastritis and peptic ulcer are significant problems. In the lower intestinal tract, the occurrence of diarrhea and abdominal pain is attributed to increased motility due to mast cell mediators; this problem can be aggravated by malabsorption, which can also cause secondary nutritional insufficiency and osteomalacia. The periportal fibrosis associated with mast cell infiltration and a prominence of eosinophils may lead to portal hypertension and ascites. In some patients, flushing and recurrent vascular collapse are markedly aggravated by an idiosyncratic response to a minimal dosage of NSAIDs. The neuropsychiatric disturbances are clinically most evident as impaired recent memory, decreased attention span, and "migraine-like" headaches. Patients may experience exacerbation of a specific clinical sign or symptom with alcohol ingestion, use of mast cell-interactive narcotics, or ingestion of NSAIDs.

Diagnosis

Although the diagnosis of mastocytosis is generally suspected on the basis of the clinical history and physical findings, and can be supported by laboratory procedures, it can be established only by a tissue diagnosis. By convention, the diagnosis of systemic mastocytosis depends heavily on bone marrow biopsy to meet the criteria of one major plus one minor or three minor findings **(Table 317-3)**. The bone marrow provides the major criterion by revealing aggregates of mast cells, often in paratrabecular and perivascular locations with lymphocytes and eosinophils, as well as the minor criteria of an abnormal mast cell morphology, an aberrant mast cell membrane immunophenotype, or a codon 816 mutation in any cell type. A serum total tryptase level and/or a 24-hour urine collection for measurement of histamine, histamine metabolites, or metabolites of PGD_2 are noninvasive approaches to consider before bone marrow biopsy. The a form of tryptase is elevated in more than one-half of patients with systemic mastocytosis and provides a minor criterion; the β form is increased in patients undergoing an anaphylactic reaction. Additional studies directed by the presentation include a bone scan or skeletal survey; contrast studies of the upper gastrointestinal tract with small-bowel follow-through, CT scan, or endoscopy; and a neuropsychiatric evaluation, including an electroencephalogram.

Table 317-3 Diagnostic Criteria for Systemic Mastocytosis*

Major: Multifocal dense infiltrates of mast cells in bone marrow or other extracutaneous tissues with confirmation by immunodetection of tryptase or metachromasia
Minor: Abnormal mast cell morphology with a spindle shape and/or multilobed or eccentric nucleus
Aberrant mast cell surface phenotype with expression of CD25 (IL-2 receptor) and CD2 in addition to C117 (c-*kit*)
Detection of codon 816 mutation in peripheral blood cells, bone marrow cells, or lesional tissue
Total serum tryptase (mostly alpha) greater than 20 ng/mL

*Diagnosis requires either the major and one minor or three minor criteria.

The differential diagnosis requires the exclusion of other flushing disorders. The 24-hour urine assessment of 5-hydroxy-indoleacetic acid and metanephrines should exclude a carcinoid tumor or a pheochromocytoma. Most patients with recurrent anaphylaxis, including the idiopathic group, present with angioedema and/or wheezing, which are not manifestations of systemic mastocytosis.

TREATMENT: SYSTEMIC MASTOCYTOSIS

The management of systemic mastocytosis uses a stepwise and symptom/sign-directed approach that includes an H_1 antihistamine for flushing and pruritus, an H_2 antihistamine or proton pump inhibitor for gastric acid hypersecretion, oral cromolyn sodium for diarrhea and abdominal pain, and aspirin for severe flushing with or without associated vascular collapse, despite use of H_1 and H_2 antihistamines, to block biosynthesis of PGD_2. Systemic glucocorticoids appear to alleviate the malabsorption. Headaches are generally managed with tricyclic antidepressants and other neurotransmitter-modifying agents. Ketotifen has been used to alleviate flushing in patients with gastric intolerance to NSAIDs and in patients with bone pain or intractable headaches. The efficacy of IFN-α in ASM is controversial, perhaps because of dosage limitations due to side effects. Treatment with hydroxyurea to reduce the mast cell lineage progenitors has a benefit in ASM. Chemotherapy is appropriate for the frank leukemias. Although c-*kit* is a receptor tyrosine kinase, the gain-in-function mutation of codon 816 is not susceptible to inhibition by imatinib mesylate.

ALLERGIC RHINITIS

Definition

Allergic rhinitis is characterized by sneezing; rhinorrhea; obstruction of the nasal passages; conjunctival, nasal, and pharyngeal itching; and lacrimation, all occurring in a temporal relationship to allergen exposure. Although commonly seasonal due to elicitation by airborne pollens, it can be perennial in an environment of chronic exposure. In North America, the incidence of allergic rhinitis is about 7%. The overall prevalence in North America is nearly 20%, with the peak prevalence of nearly 40% occurring in childhood and adolescence.

Predisposing Factors and Etiology

Allergic rhinitis generally occurs in atopic individuals, i.e., in persons with a family history of a similar or related symptom complex and a personal history of collateral allergy expressed as eczematous dermatitis, urticaria, and/or asthma **(Chap. 254)**. Up to 40% of patients with rhinitis manifest asthma, whereas ~70% of individuals with asthma experience rhinitis. Symptoms generally appear before the fourth decade of life and tend to diminish gradually with aging, although complete spontaneous remissions are uncommon. A relatively small number of weeds that depend on wind rather than insects for cross-pollination, as well as grasses and some trees, produce sufficient quantities of pollen suitable for wide distribution by air currents to elicit seasonal allergic rhinitis. The dates of pollination of these species generally vary little from year to year in a particular locale but may be quite different in another climate. In the temperate areas of North America, trees typically pollinate from March through May, grasses in June and early July, and ragweed from mid-August to early October. Molds, which are widespread in nature because they occur in soil or decaying organic matter, may propagate spores in a pattern that depends on climatic conditions. Perennial allergic rhinitis occurs in response to allergens that are present throughout the year, including desquamating epithelium in animal dander, cockroach-derived proteins, mold spores, or dust, which has mites such as *Dermatophagoides farinae* and *D. pteronyssinus*. Dust mites are scavengers of flecks of human skin and coat the digestate with mite-specific protein for excretion. In up to one-half of patients with perennial rhinitis, no clear-cut allergen can be demonstrated as causative. The ability of allergens to cause rhinitis rather than lower respiratory tract symptoms may be attributed to their large size, 10-100 μm, and retention within the nose.

Pathophysiology and Manifestations

Episodic rhinorrhea, sneezing, obstruction of the nasal passages with lacrimation, and pruritus of the conjunctiva, nasal mucosa, and oropharynx are the hallmarks of allergic rhinitis. The nasal mucosa is pale and boggy, the conjunctiva congested and edematous, and the pharynx generally unremarkable. Swelling of the turbinates and mucous membranes with obstruction of the sinus ostia and eustachian tubes precipitates secondary infections of the sinuses and middle ear, respectively. Nasal polyps, representing mucosal protrusions containing edema fluid with variable numbers of eosinophils, can increase obstructive symptoms and can concurrently arise within the nasopharynx or sinuses. Nasal polyps may occur independent of allergic rhinitis in patients with the aspirin-intolerant triad of rhinosinusitis and asthma and in patients with chronic staphylococcal colonization, which produces superantigens leading to an intense T_H2 inflammatory response.

The nose presents a large mucosal surface area through the folds of the turbinates and serves to adjust the temperature and moisture content of inhaled air and to filter out particulate materials >10 μm in size by impingement in a mucous blanket; ciliary action moves the entrapped particles toward the pharynx. Entrapment of pollen and digestion of the outer coat by mucosal enzymes such as lysozymes release protein allergens generally of 10,000-40,000 molecular weight. The initial interaction occurs between the allergen and intraepithelial mast cells and then proceeds to involve deeper perivenular mast cells, both of which are sensitized with specific IgE. During the symptomatic season when the mucosae are already swollen and hyperemic, there is enhanced adverse reactivity to the seasonal pollen. Biopsy specimens of nasal mucosa during seasonal rhinitis show submucosal edema with infiltration by eosinophils, along with some basophils and neutrophils.

The mucosal surface fluid contains IgA that is present because of its secretory piece and also IgE, which apparently arrives by diffusion from plasma cells in proximity to mucosal surfaces. IgE fixes to mucosal and submucosal mast cells, and the intensity of the clinical response to inhaled allergens is quantitatively related to the naturally occurring pollen dose. In sensitive individuals, the introduction of allergen into the nose is associated with sneezing, "stuffiness," and discharge, and the fluid contains histamine, PGD_2, and leukotrienes. Thus the mast cells of the nasal mucosa and submucosa generate and release mediators through IgE-dependent reactions that are capable of producing tissue edema and eosinophilic infiltration.

Diagnosis

The diagnosis of seasonal allergic rhinitis depends largely on an accurate history of occurrence coincident with the pollination of the offending weeds, grasses, or trees. The continuous character of perennial allergic rhinitis due to contamination of the home or place of work makes historic analysis difficult, but there may be a variability in symptoms that can be related to exposure to animal dander, dust mite and/or cockroach allergens, fungal spores, or work-related allergens such as latex. Patients with perennial rhinitis commonly develop the problem in adult life, and manifest nasal congestion and a postnasal discharge, often associated with thickening of the sinus membranes demonstrated by radiography. Perennial nonallergic rhinitis with eosinophilia syndrome (NARES) occurs in the middle decades of life and is characterized by nasal obstruction, anosmia, chronic sinusitis, and frequent aspirin intolerance. The term *vasomotor rhinitis* or *perennial nonallergic*

rhinitis designates a condition of enhanced reactivity of the nasopharynx in which a symptom complex resembling perennial allergic rhinitis occurs with nonspecific stimuli, including chemical odors, temperature and humidity variations, and position changes but occurs without tissue eosinophilia or an allergic etiology. Other entities to be excluded are structural abnormalities of the nasopharynx; exposure to irritants; gustatory rhinitis associated with cholinergic activation that occurs while eating or ingesting alcohol; hypothyroidism; upper respiratory tract infection; pregnancy with prominent nasal mucosal edema; prolonged topical use of a-adrenergic agents in the form of nose drops (rhinitis medicamentosa); and the use of certain therapeutic agents such as rauwolfia, β-adrenergic antagonists, estrogens, progesterone, ACE inhibitors, aspirin and other NSAIDS, and drugs for erectile dysfunction (phosphodiesterase-5 inhibitors).

The nasal secretions of allergic patients are rich in eosinophils, and a modest peripheral eosinophilia is a common feature. Local or systemic neutrophilia implies infection. Total serum IgE is frequently elevated, but the demonstration of immunologic specificity for IgE is critical to an etiologic diagnosis. A skin test by the intracutaneous route (puncture or prick) with the allergens of interest provides a rapid and reliable approach to identifying allergen-specific IgE that has sensitized cutaneous mast cells. A positive intracutaneous skin test with 1:10-1:20 weight/volume of extract has a high predictive value for the presence of allergy. An intradermal test with a 1:500-1:1000 dilution of 0.05 mL may follow if indicated by history when the intracutaneous test is negative, but while more sensitive, it is less reliable due to the reactivity of some asymptomatic individuals at the test dose. Skin testing by the intracutaneous route for food allergens can be supportive of the clinical history. A double-blind, placebo-controlled challenge may document a food allergy, but such a procedure does bear the risk of an anaphylactic reaction. An elimination diet is safer but is tedious and less definitive. Food allergy is uncommon as a cause of allergic rhinitis.

Newer methodology for detecting total IgE, including the development of enzyme-linked immunosorbent assays (ELISA) employing anti-IgE bound to either a solid-phase or a liquid-phase particle, provides rapid and cost-effective determinations. Measurements of specific anti-IgE in serum are obtained by its binding to an allergen and quantitation by subsequent uptake of labeled anti-IgE. As compared to the skin test, the assay of specific IgE in serum is less sensitive but has high specificity.

Prevention

Avoidance of exposure to the offending allergen is the most effective means of controlling allergic diseases; removal of pets from the home to avoid animal danders, utilization of air-filtration devices to minimize the concentrations of airborne pollens, elimination of cockroach-derived proteins by chemical destruction of the pest and careful food storage, travel to areas where the allergen is not being generated, and even a change of domicile to eliminate a mold spore problem may be necessary. Control of dust mites by allergen avoidance includes use of plastic-lined covers for mattresses, pillows, and comforters; using a filter-equipped vacuum cleaner; washing bedding and clothes at temperatures >54.5°C (above 130°F); and elimination of carpets and drapes.

TREATMENT: ALLERGIC RHINITIS

Although allergen avoidance is the most cost-effective means of managing allergic rhinitis, treatment with pharmacologic agents represents the standard approach to seasonal or perennial allergic rhinitis. Oral antihistamines of the H_1 class are effective for nasopharyngeal itching, sneezing, and watery rhinorrhea and for such ocular manifestations as itching, tearing, and erythema, but they are not efficacious for the nasal congestion. The older antihistamines are sedating, and they induce psychomotor impairment, including reduced eye-hand coordination and impaired automobile driving skills. Their anticholinergic (muscarinic) effects include visual disturbance, urinary retention, and constipation. Because the newer H_1 antihistamines such as fexofenadine, loratadine, desloradine, cetirizine, levocetirizine, olopatadine, bilastine, and azelastine are less lipophilic and more H_1 selective, their ability to cross the blood-brain barrier is reduced, and thus their sedating and anticholinergic side effects are minimized. These newer antihistamines do not differ appreciably in efficacy for relief of rhinitis and/or sneezing. Azelastine nasal spray may benefit individuals with nonallergic vasomotor rhinitis, but it has an adverse effect of dysgeusia (taste perversion) in some patients. Because antihistamines have little effect on congestion, α-adrenergic agents such as phenylephrine or oximetazoline are generally used topically to alleviate nasal congestion and obstruction. However, the duration of their efficacy is limited because of rebound rhinitis (i.e., 7- to 14-day use can lead to rhinitis medicamentosa) and such systemic responses as hypertension. Oral α-adrenergic agonist decongestants containing pseudoephedrine are standard for the management of nasal congestion, generally in combination with an antihistamine. While oral antihistamines typically reduce nasal and ocular symptoms by about one third, pseudoephedrine must be added to achieve a similar reduction in nasal congestion. These pseudoephedrine combination products can cause insomnia and are precluded from use in patients with narrow angle glaucoma, urinary retention, severe hypertension, marked coronary artery disease, or a first trimester pregnancy. The $CysLT_1$ blocker montelukast is approved for treatment of both seasonal and perennial rhinitis, and it reduces both nasal and ocular symptoms by about 20%. Cromolyn sodium, a nasal spray, is essentially without side effects and is used prophylactically on a continuous basis during the season. The clinical efficacy of cromolyn sodium used prophylactically is less than that of second-generation oral antihistamines and less than that of intranasal glucocorticoids. Intranasal high-potency glucocorticoids are the most potent drugs available for the relief of established rhinitis, seasonal or perennial, and are effective in relieving nasal congestion. They provide efficacy with substantially reduced side effects as compared with this same class of agent administered orally. Their most frequent side effect is local irritation, with *Candida* overgrowth being a rare occurrence. The currently available intranasal glucocorticoids-beclomethasone, flunisolide, triamcinolone, budesonide, fluticasone propionate, fluticasone furoate, ciclesonide, and mometasone furoate-are equally effective for nasal symptom relief, including nasal congestion; these agents all achieve up to 70% overall symptom relief with some variation in the time period for onset of benefit. Topical ipratropium is an anticholinergic agent effective in reducing rhinorrhea, including that in patients with perennial symptoms, and it can be additionally efficacious when combined with intranasal glucocorticoids. Local treatment with cromolyn sodium is effective in treating mild allergic conjunctivitis. Topical antihistamines such as olopatadine, azelastine, ketotifen, or epinastine administered to the eye provide rapid relief of itching and redness and are more effective than oral antihistamines.

Immunotherapy, often termed *hyposensitization*, consists of repeated subcutaneous injections of gradually increasing concentrations of the allergen(s) considered to be specifically responsible for the symptom complex. Controlled studies of ragweed, grass, dust mite, and cat dander allergens administered for treatment of allergic rhinitis have demonstrated at least partial relief of symptoms and signs. The duration of such immunotherapy is 3-5 years, with discontinuation being based on minimal symptoms over two consecutive seasons of exposure to the allergen. Clinical benefit appears related to the administration of a high dose of relevant allergen, advancing from weekly to monthly intervals. Patients should remain at the treatment site for at least 20 minutes after allergen administration so that any anaphylactic consequence can be managed. Local reactions with erythema and induration are not uncommon and may persist for 1-3 days. Immunotherapy is contraindicated in patients with significant cardiovascular disease or unstable asthma and should be conducted with particular caution in any patient requiring β-adrenergic blocking therapy because of the difficulty in managing an anaphylactic complication. The response to immunotherapy is derived from a complex of cellular and humoral effects that likely includes a modulation in T cell cytokine production. Immunotherapy should be reserved for clearly documented seasonal or perennial rhinitis that is clinically related to defined allergen exposure with confirmation by the presence of allergen-specific IgE. Systemic treatment with a monoclonal antibody to IgE (omalizumab) that blocks mast cell and basophil sensitization has efficacy for allergic rhinitis and can be used with immunotherapy to enhance safety and efficacy. However, current approval is only for treatment of patients with persistent allergic asthma not controlled by inhaled glucocorticoid therapy. A sequence for the management of allergic or perennial rhinitis based on an allergen-specific diagnosis and stepwise management as required for symptom control would include the following: (1) identification of the offending allergen(s) by history with confirmation of the presence of allergen-specific IgE by skin test and/or serum assay; (2) avoidance of the offending allergen; and (3) medical management in a stepwise fashion **(Fig. 317-4)**. Mild intermittent symptoms of allergic rhinitis are treated with oral antihistamines, oral $CysLT_1$ receptor antagonists, intranasal antihistamines, or intranasal cromolyn prophylaxis. Moderate to more severe allergic rhinitis is managed with intranasal glucocorticoids plus oral antihistamines, oral $CysLT_1$ receptor antagonists, or antihistamine-decongestant combinations. Persistent allergic rhinitis requiring the daily use of intranasal glucocorticoids with add-on interventions such as oral antihistamines, decongestant combinations, or topical ipratropium merits consideration of allergen-specific immunotherapy. Even a brief course of oral prednisone can be indicated for rapid relief of severe allergic rhinitis symptoms.

Figure 317-4

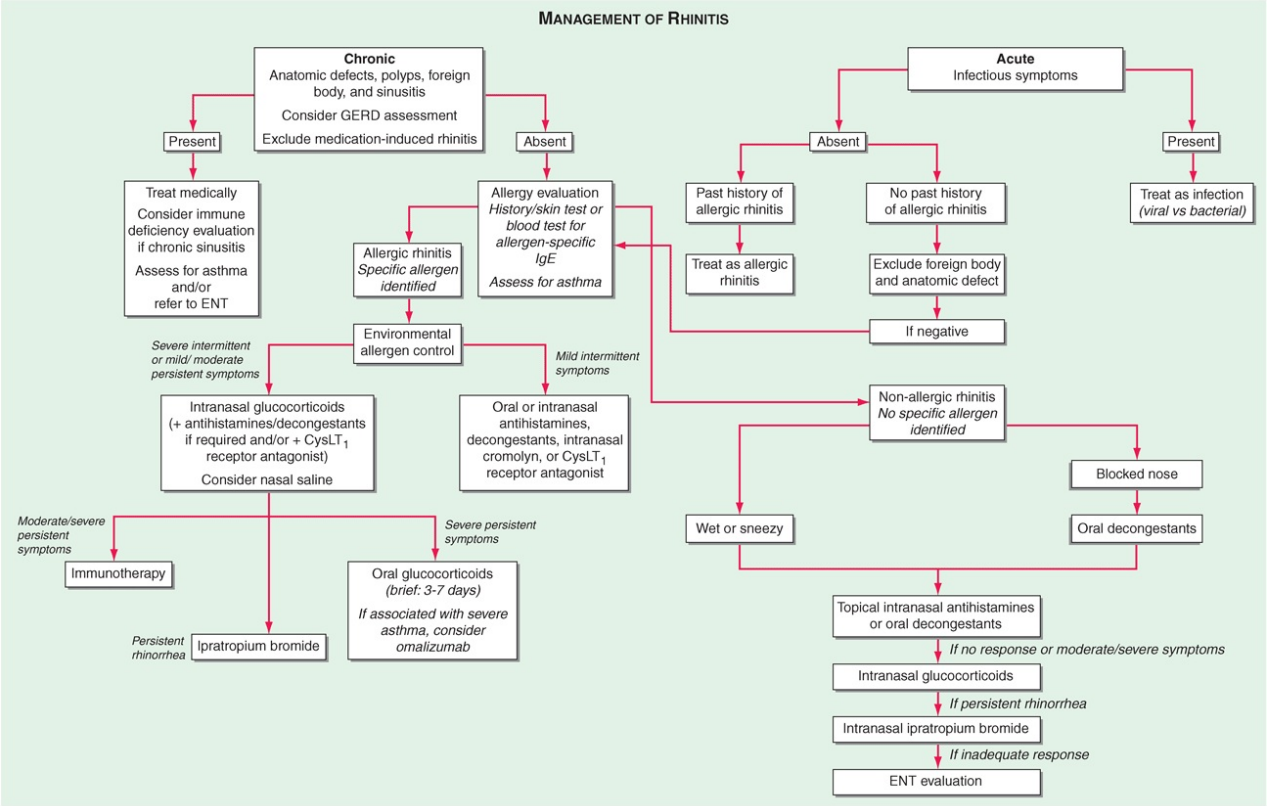

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the diagnosis and management of rhinitis. ENT, ear, nose, and throat; GERD, gastroesophageal reflux disease.

FURTHER READINGS

Austen KF, Kanaoka Y: Lipids, in *The Innate Immune Response to Infection*, SHE Kaufmann et al (eds). Washington, DC, ASM Press, 2004, pp 417-431

Benoist C, Mathis D: Mast cells in autoimmune disease. Nature 420:875, 2002[PMID: 12490961] [Full Text]

Galli SJ et al: Mast cells in the development of adaptive immune responses. Nat Immunol 6:135, 2005[PMID: 15662442] [Full Text]

Kaplan AP: Urticaria, in *Allergy and Allergic Diseases*, 2nd ed, AB Kay et al (eds). Oxford, Wiley-Blackwell, 2008, pp 1853-1875

Kushnir-Sukov NM et al: Mastocytosis, in *Allergy and Allergic Diseases*, 2nd ed, AB Kay et al (eds). Oxford, Wiley-Blackwell, 2008, pp 1878-1893

Valentine MD: Insect venom allergy, in *Immunologic Diseases*, 6th ed, KF Austen et al (eds). Philadelphia, Lippincott Williams & Wilkins, 2001

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

AUTOIMMUNITY AND AUTOIMMUNE DISEASES: INTRODUCTION

One of the central features of the immune system is the capacity to mount an inflammatory response to nonself while avoiding harm to self tissues. While recognition of self plays an important role in shaping the repertoires of immune receptors on both T and B cells, and in the clearance of apoptotic debris from tissues throughout the body, the development of potentially harmful immune responses to self-antigens is, in general, precluded. The essential feature of an autoimmune disease is that tissue injury is caused by the immunologic reaction of the organism against its own tissues. Autoimmunity, on the other hand, refers merely to the presence of antibodies or T lymphocytes that react with self-antigens and does not necessarily imply that the self-reactivity has pathogenic consequences. Autoimmunity is present in all individuals; however, autoimmune disease represents the end result of the breakdown of one or more of the basic mechanisms regulating immune tolerance.

Autoimmunity is seen in normal individuals and in higher frequency in normal older people. Polyreactive autoantibodies that recognize many host antigens are present throughout life. Expression of these antibodies may be increased following some inciting events. These are usually of the IgM heavy chain isotype and are encoded by nonmutated germline immunoglobulin variable region genes. When autoimmunity is induced by an inciting event, such as infection or tissue damage from trauma or ischemia, the autoreactivity is in general self-limited. Such autoimmunity may, however, be persistent, and then may or may not result in ensuing pathology. Even in the presence of organ pathology, it may be difficult to determine whether the damage is mediated by autoreactivity. Following an inciting event, the development of self-reactivity may be the consequence of an ongoing pathologic process, and be nonpathogenic, or may contribute to tissue inflammation and damage.

MECHANISMS OF AUTOIMMUNITY

Since Ehrlich first postulated the existence of mechanisms to prevent the generation of self-reactivity in 1900, ideas concerning the nature of this inhibition have developed in parallel with a progressive increase in understanding of the immune system. Burnet's clonal selection theory included the idea that interaction of lymphoid cells with their specific antigens during fetal or early postnatal life would lead to elimination of such "forbidden clones." This idea became untenable, however, when it was shown that autoimmune diseases could be induced in experimental animals by simple immunization procedures, that autoantigen-binding cells could be demonstrated easily in the circulation of normal individuals, and that self-limited autoimmune phenomena frequently developed following tissue damage from infection or trauma. These observations indicated that clones of cells capable of responding to autoantigens were present in the repertoire of antigen-reactive cells in normal adults and suggested that mechanisms in addition to clonal deletion were responsible for preventing their activation.

Currently, three general processes are thought to be involved in the maintenance of selective unresponsiveness to autoantigens **(Table 318-1)**: (1) sequestration of self-antigens, rendering them inaccessible to the immune system; (2) specific unresponsiveness (tolerance or anergy) of relevant T or B cells; and (3) limitation of potential reactivity by regulatory mechanisms.

Table 318-1 Mechanisms Preventing Autoimmunity

1. Sequestration of self-antigen
2. Generation and maintenance of tolerance
a. Central deletion of autoreactive lymphocytes b. Peripheral anergy of autoreactive lymphocytes
c. Receptor replacement in autoreactive lymphocytes
3. Regulatory mechanisms

Derangements of these normal processes may predispose to the development of autoimmunity **(Table 318-2)**. In general, these abnormal responses require an exogenous trigger such as bacterial or viral infection or cigarette smoking and require the presence of endogenous abnormalities in the cells of the immune system. Microbial superantigens, such as staphylococcal protein A and staphylococcal enterotoxins, are substances that can stimulate a broad range of T and B cells based upon specific

interactions with selected families of immune receptors, irrespective of their antigen specificity. If autoantigen-reactive T and/or B cells express these receptors, autoimmunity might develop. Alternatively, molecular mimicry or cross-reactivity between a microbial product and a self-antigen might lead to activation of autoreactive lymphocytes. One of the best examples of autoreactivity and autoimmune disease resulting from molecular mimicry is rheumatic fever, in which antibodies to the M protein of streptococci cross-react with myosin, laminin, and other matrix proteins as well as neuronal antigens. Deposition of these autoantibodies in the heart initiates an inflammatory response, whereas penetration of these antibodies into the brain can result in Sydenham's chorea. Molecular mimicry between microbial proteins and host tissues has been reported in type 1 diabetes mellitus, rheumatoid arthritis, and multiple sclerosis. It is presumed that infectious agents may be able to overcome self-tolerance because they possess molecules, such as bacterial endotoxin, RNA, or DNA, that have adjuvant-like effects on the immune system that increase the immunogenicity of the microbial antigens. The adjuvants activate dendritic cells through pattern recognition receptors and stimulate the activation of previously quiescent lymphocytes that recognize both microbial and self antigen.

Table 318-2 Mechanisms of Autoimmunity

I. Exogenous
A. Molecular mimicry
B. Superantigenic stimulation
C. Microbial adjuvanticity
II. Endogenous
A. Altered antigen presentation
1. Loss of immunologic privilege
2. Presentation of novel or cryptic epitopes (epitope spreading)
3. Alteration of self-antigen
4. Enhanced function of antigen-presenting cells
a. Co-stimulatory molecule expression
b. Cytokine production
B. Increased T cell help
1. Cytokine production
2. Co-stimulatory molecules
C. Increased B cell function
D. Apoptotic defects
E. Cytokine imbalance
F. Altered immunoregulation

Endogenous derangements of the immune system may also contribute to the loss of immunologic tolerance to self-antigens and the development of autoimmunity (Table 318-2). Some autoantigens reside in immunologically privileged sites, such as the brain or the anterior chamber of the eye. These sites are characterized by the inability of engrafted tissue to elicit immune responses. Immunologic privilege results from a number of events, including the limited entry of proteins from those sites into lymphatics, the local production of immunosuppressive cytokines such as transforming growth factor ß, and the local expression of molecules such as Fas ligand that can induce apoptosis of activated T cells. Lymphoid cells remain in a state of immunologic ignorance (neither activated nor anergized) to proteins expressed uniquely in immunologically privileged sites. If the privileged site is damaged by trauma or inflammation, or if T cells are activated elsewhere, proteins expressed at this site can become the targets of immunologic assault. Such an event may occur in multiple sclerosis and sympathetic ophthalmia, in which antigens uniquely expressed in the brain and eye, respectively, become the target of activated T cells.

Alterations in antigen presentation may also contribute to autoimmunity. Peptide determinants (*epitopes*) of a self antigen that are not routinely presented to lymphocytes may be recognized as a result of altered proteolytic processing of the molecule and the ensuing presentation of novel peptides (cryptic epitopes). When B cells rather than dendritic cells present self antigen, they may also present cryptic epitopes that can activate autoreactive T cells. These cryptic epitopes will not have previously been available to effect the silencing of autoreactive lymphocytes. Furthermore, once there is immunologic recognition of one protein component of a multimolecular complex, reactivity may be induced to other components of the complex following internalization and presentation of all molecules within the complex (epitope spreading). Finally, inflammation, drug exposure, or normal senescence may cause a primary chemical alteration in proteins, resulting in the generation of immune responses that cross-react with normal self-proteins. For example, the induction and/or release of protein arginine

deaminase enzymes results in the conversion of arginine residues to citrullines in a variety of proteins, thereby altering their capacity to induce immune responses. Production of anticitrullinated protein antibodies has been observed in rheumatoid arthritis, chronic lung disease, as well as normal smokers and may contribute to organ pathology. Alterations in the availability and presentation of autoantigens may be important components of immunoreactivity in certain models of organ-specific autoimmune diseases. In addition, these factors may be relevant in understanding the pathogenesis of various drug-induced autoimmune conditions. However, the diversity of autoreactivity manifest in non-organ-specific systemic autoimmune diseases suggests that these conditions might result from a more general activation of the immune system rather than from an alteration in individual self-antigens.

Many autoimmune diseases are characterized by the presence of antibodies that react with apoptotic material. Defects in the clearance of apoptotic material have been shown to elicit autoimmunity and autoimmune disease in a number of animal models. Moreover, defects in the clearance of apoptotic material have been found in subjects with systemic lupus erythematosus (SLE). Apoptotic debris not quickly cleared by the immune system can function as endogenous ligands for a number of pattern recognition receptors on dendritic cells. Under such circumstances, there is activation of dendritic cells, and an immune response to apoptotic debris can develop. In addition, the presence of extracellular apoptotic material within germinal centers of secondary lymphoid organs may facilitate the direct activation of autoimmune B cell clones or function to select autoimmune B cell clones during immune responses.

A number of experimental models have suggested that intense stimulation of T lymphocytes can produce nonspecific signals that bypass the need for antigen-specific helper T cells and lead to polyclonal B cell activation with the formation of multiple autoantibodies. For example, antinuclear, antierythrocyte, and antilymphocyte antibodies are produced during the chronic graft-versus-host reaction. In addition, true autoimmune diseases, including autoimmune hemolytic anemia and immune complex–mediated glomerulonephritis, can also be induced in this manner. While it is clear that such diffuse activation of helper T cell activity can cause autoimmunity, nonspecific stimulation of B lymphocytes can also lead to the production of autoantibodies. Thus, the administration of polyclonal B cell activators, such as bacterial endotoxin, to normal mice leads to the production of a number of autoantibodies, including those directed to DNA and IgG (rheumatoid factor). Moreover, excess BAFF can also cause T cell–independent B cell activation and heavy chain class switching and the development of autoimmunity. SLE, for example, can be induced in mice through exuberant dendritic cell activation, a redundancy of TLR7 on the y chromosome (BXSByaa mice) or through exposure to CpG, a ligand for TLR 9. The ensuing induction of inflammatory mediators can cause a switch from production of nonpathogenic IgM autoantibodies to pathogenic IgG autoantibodies in the absence of antigen-specific T cell help.

Aberrant selection of the B or T cell repertoire at the time of antigen receptor expression can also predispose to autoimmunity. For example, B cell immunodeficiency caused by an absence of the B cell receptor–associated kinase, Bruton's tyrosine kinase, leads to X-linked agammaglobulinemia. This syndrome is characterized by reduced B cell activation, but also by diminished negative selection of autoreactive B cells probably caused by high levels of BAFF, resulting in increased autoreactivity within a diminished B cell repertoire. Likewise, negative selection of autoreactive T cells in the thymus requires expression of the autoimmune regulator (AIRE) gene that enables the expression of tissue-specific proteins in thymic medullary epithelial cells. Peptides from these proteins are expressed in the context of major histocompatibility complex (MHC) molecules and mediate the elimination of autoreactive T cells. The absence of AIRE gene expression leads to a failure of negative selection of autoreactive cells, autoantibody production, and severe inflammatory destruction of multiple organs. Individuals deficient in AIRE gene expression develop autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).

Primary alterations in the activity of T and/or B cells, cytokine imbalances, or defective immunoregulatory circuits may also contribute to the emergence of autoimmunity. Diminished production of tumor necrosis factor (TNF) and interleukin (IL) 10 has been reported to be associated with the development of autoimmunity. Overproduction of type 1 interferon has also been associated with autoimmunity. Overexpression of co-stimulatory molecules on T cells similarly can lead to autoantibody production.

Autoimmunity may also result from an abnormality of immunoregulatory mechanisms. Observations made in both human autoimmune disease and animal models suggest that defects in the generation and expression of regulatory T cell activity may allow for the production of autoimmunity. It has recently been appreciated that the IPEX (immunodysregulation, polyendocrinopathy, enteropathy X-linked) syndrome results from the failure to express the FOXP3 gene, which encodes a molecule critical in the differentiation of regulatory T cells. Administration of normal regulatory T cells or factors derived from them can prevent the development of autoimmune disease in rodent models of autoimmunity. Abnormalities in the function of regulatory T cells have been noted in a number of human autoimmune diseases, although it remains uncertain whether these are causative or are secondary abnormalities owing to inflammation. Finally, recent data indicate that B cells may also exert regulatory function, largely through the production of the cytokine IL-10. Deficiency of IL-10-producing

regulatory B cells can prolong the course of an animal model of multiple sclerosis.

It should be apparent that no single mechanism can explain all the varied manifestations of autoimmunity. Furthermore, genetic evaluation has shown that a number of abnormalities often need to converge to induce an autoimmune disease. Additional factors that appear to be important determinants in the induction of autoimmunity include age, sex (many autoimmune diseases are far more common in women), genetic background, exposure to infectious agents, and environmental contacts. How all of these disparate factors affect the capacity to develop self-reactivity is currently being investigated intensively.

GENETIC CONSIDERATIONS

Evidence in humans that there are susceptibility genes for autoimmunity comes from family studies and especially from studies of twins. Studies in type 1 diabetes mellitus, rheumatoid arthritis, multiple sclerosis, and SLE have shown that approximately 15–30% of pairs of monozygotic twins show disease concordance, compared with <5% of dizygotic twins. The occurrence of different autoimmune diseases within the same family has suggested that certain susceptibility genes may predispose to a variety of autoimmune diseases. Genetic mapping has begun to identify chromosomal regions that predispose to specific autoimmune diseases. It is notable that some genes are associated with multiple autoimmune diseases, whereas others are more specifically associated with only one autoimmune condition. The gene encoding PTPN22 is associated with multiple autoimmune diseases. Its product is a phosphatase expressed by a variety of hematopoietic cells that downregulates antigen receptor–mediated stimulation of T and B cells. A gain-of-function polymorphism of this gene is associated with type 1 diabetes mellitus, rheumatoid arthritis, and SLE in some populations. The explanation of the association of this polymorphism with autoimmune disease is uncertain, but it is likely that it diminishes antigen receptor signaling during lymphocyte development permitting escape of autoreactive clones or decreased activation-induced apoptosis of autoantigen-reactive lymphocytes in the periphery. In recent years, genomewide association studies have demonstrated a variety of other genes that are involved in human autoimmune diseases. Most genes individually confer a relatively low risk for autoimmune diseases and are found in normal individuals. No gene has been identified that is essential for autoimmune diseases. In addition to this evidence from humans, certain inbred mouse strains reproducibly develop specific spontaneous or experimentally induced autoimmune diseases, whereas others do not. These findings have led to an extensive search for genes that determine susceptibility to autoimmune disease.

The strongest consistent association for susceptibility to autoimmune disease has been found with particular alleles of the MHC. It has been suggested that the association of MHC genotype with autoimmune disease relates to differences in the ability of different allelic variations of MHC molecules to present autoantigenic peptides to autoreactive T cells. An alternative hypothesis involves the role of MHC alleles in shaping the T cell receptor repertoire during T cell ontogeny in the thymus. Additionally, specific MHC gene products may themselves be the source of peptides that can be recognized by T cells. Cross-reactivity between such MHC peptides and peptides derived from proteins produced by common microbes may trigger autoimmunity by molecular mimicry. However, MHC genotype alone does not determine the development of autoimmunity. Identical twins are far more likely to develop the same autoimmune disease than MHC-identical nontwin siblings, suggesting that genetic factors other than the MHC also affect disease susceptibility. Recent studies of the genetics of type 1 diabetes mellitus, SLE, rheumatoid arthritis, and multiple sclerosis in humans and mice have shown that there are several independently segregating disease susceptibility loci in addition to the MHC. Genes that encode molecules of the innate immune response are also involved in autoimmunity. In humans, inherited homozygous deficiency of the early proteins of the classic pathway of complement (C1q, C4, or C2) as well as genes involved in the type 1 interferon pathway are very strongly associated with the development of SLE.

Immunopathogenic Mechanisms in Autoimmune Diseases

The mechanisms of tissue injury in autoimmune diseases can be divided into antibody-mediated and cell-mediated processes. Representative examples are listed in **Table 318-3**.

Table 318-3 Mechanisms of Tissue Damage in Autoimmune Disease

Effector	Mechanism	Target	Disease
Autoantibody	Blocking or inactivation	α Chain of the nicotinic acetylcholine receptor Phospholipid–β2-glycoprotein 1 complex Insulin receptor Intrinsic factor	Myasthenia gravis Antiphospholipid syndrome Insulin-resistant diabetes mellitus Pernicious anemia

	Stimulation	TSH receptor (LATS) Proteinase-3 (ANCA) Epidermal cadherin$_1$ Desmoglein 3	Graves' disease Granulomatosis with polyangiitis (Wegener's) Pemphigus vulgaris
	Complement activation Immune-complex formation Opsonization	α_3 Chain of collagen IV Double-stranded DNA Ig Platelet GpIIb:IIIa Rh antigens, I antigen	Goodpasture's syndrome Systemic lupus erythematosus Rheumatoid arthritis Autoimmune thrombocytopenic purpura Autoimmune hemolytic anemia
	Antibody-dependent cellular cytotoxicity	Thyroid peroxidase, thyroglobulin	Hashimoto's thyroiditis
T cells	Cytokine production	?	Rheumatoid arthritis, multiple sclerosis, type 1 diabetes mellitus
	Cellular cytotoxicity	?	Type 1 diabetes mellitus

The pathogenicity of autoantibodies can be mediated through several mechanisms, including opsonization of soluble factors or cells, activation of an inflammatory cascade via the complement system, and interference with the physiologic function of soluble molecules or cells.

In autoimmune thrombocytopenic purpura, opsonization of platelets targets them for elimination by phagocytes. Likewise, in autoimmune hemolytic anemia, binding of immunoglobulin to red cell membranes leads to phagocytosis and lysis of the opsonized cell. Goodpasture's syndrome, a disease characterized by lung hemorrhage and severe glomerulonephritis, represents an example of antibody binding leading to local activation of complement and neutrophil accumulation and activation. The autoantibody in this disease binds to the α_3 chain of type IV collagen in the basement membrane. In SLE, activation of the complement cascade at sites of immunoglobulin deposition in renal glomeruli is considered to be a major mechanism of renal damage. Moreover, the DNA- and RNA-containing immune complexes in SLE activate TLR 9 and 7, respectively, in dendritic cells and promote a proinflammatory, immunogenic milieu conducive to amplifying the autoimmune response.

Autoantibodies can also interfere with normal physiologic functions of cells or soluble factors. Autoantibodies against hormone receptors can lead to stimulation of cells or to inhibition of cell function through interference with receptor signaling. For example, long-acting thyroid stimulators, which are autoantibodies that bind to the receptor for thyroid-stimulating hormone (TSH), are present in Graves' disease and function as agonists, causing the thyroid to respond as if there were an excess of TSH. Alternatively, antibodies to the insulin receptor can cause insulin-resistant diabetes mellitus through receptor blockade. In myasthenia gravis, autoantibodies to the acetylcholine receptor can be detected in 85–90% of patients and are responsible for muscle weakness. The exact location of the antigenic epitope, the valence and affinity of the antibody, and perhaps other characteristics determine whether activation or blockade results from antibody binding.

Antiphospholipid antibodies are associated with thromboembolic events in primary and secondary antiphospholipid syndrome and have also been associated with fetal wastage. The major antibody is directed to the phospholipid–β_2-glycoprotein I complex and appears to exert a procoagulant effect. In pemphigus vulgaris, autoantibodies bind to a component of the epidermal cell desmosome, desmoglein 3, and play a role in the induction of the disease. They exert their pathologic effect by disrupting cell-cell junctions through stimulation of the production of epithelial proteases, leading to blister formation. Cytoplasmic antineutrophil cytoplasmic antibody (c-ANCA), found in granulomatosis with polyangiitis (Wegener's), is an antibody to an intracellular antigen, the 29-kDa serine protease (proteinase-3). In vitro experiments have shown that IgG anti-c-ANCA causes cellular activation and degranulation of primed neutrophils.

It is important to note that autoantibodies of a given specificity may cause disease only in genetically susceptible hosts, as has been shown in experimental models of myasthenia gravis, SLE, rheumatic fever and rheumatoid arthritis. It is also important to be aware that once organ damage is initiated, new inflammatory cascades are initiated that can sustain and amplify the autoimmune process. Finally, some autoantibodies seem to be markers for disease but have as yet no known pathogenic potential.

Autoimmune Diseases

Manifestations of autoimmunity are found in a large number of pathologic conditions. However, their presence does not necessarily imply that the pathologic process is an autoimmune disease. A number

of attempts to establish formal criteria for the diagnosis of autoimmune diseases have been made, but none is universally accepted. One set of criteria is shown in **Table 318-4**; however, this should be viewed merely as a guide in consideration of the problem.

Table 318-4 Human Autoimmune Disease: Presumptive Evidence for an Immunologic Pathogenesis

Major Criteria
1. Presence of autoantibodies or evidence of cellular reactivity to self
2. Documentation of relevant autoantibody or lymphocytic infiltrate in the pathologic lesion
3. Demonstration that relevant autoantibody or T cells can cause tissue pathology
a. Transplacental transmission
b. Adaptive transfer into animals
c. In vitro impact on cellular function Supportive Evidence
Supportive Evidence
1. Reasonable animal model
2. Beneficial effect from immunosuppressive agents
3. Association with other evidence of autoimmunity
4. No evidence of infection or other obvious cause

To classify a disease as autoimmune, it is necessary to demonstrate that the immune response to a self-antigen causes the observed pathology. Initially, the demonstration that antibodies against the affected tissue could be detected in the serum of patients suffering from various diseases was taken as evidence that these diseases had an autoimmune basis. However, such autoantibodies are also found when tissue damage is caused by trauma or infection, and the autoantibody is secondary to tissue damage. Thus, it is necessary to show that autoimmunity is pathogenic before classifying a disease as autoimmune.

If the autoantibodies are pathogenic, it may be possible to transfer disease to experimental animals by the administration of autoantibodies, with the subsequent development of pathology in the recipient similar to that seen in the patient from whom the antibodies were obtained. This has been shown, for example, in Graves' disease. Some autoimmune diseases can be transferred from mother to fetus and are observed in the newborn babies of diseased mothers. The symptoms of the disease in the newborn usually disappear as the levels of the maternal antibody decrease. An exception, however, is congenital heart block, in which damage to the developing conducting system of the heart follows in utero transfer of anti-Ro antibody from the mother to the fetus. This can result in a permanent developmental defect in the heart.

In most situations, the critical factors that determine when the development of autoimmunity results in autoimmune disease have not been delineated. The relationship of autoimmunity to the development of autoimmune disease may relate to the fine specificity of the antibodies or T cells or their specific effector capabilities. In many circumstances, a mechanistic understanding of the pathogenic potential of autoantibodies has not been established. In some autoimmune diseases, biased production of cytokines by helper T (T_H) cells may play a role in pathogenesis. In this regard, T cells can differentiate into specialized effector cells that predominantly produce interferon γ (T_H1), IL-4 (T_H2), IL-17 (T_H17) or provide help to B cells (T follicular helper, T_{FH}) **(Chap. 314)**. T_H1 cells facilitate macrophage activation and classic cell-mediated immunity, whereas T_H2 cells are thought to have regulatory functions and are involved in the resolution of normal immune responses and also the development of responses to a variety of parasites; T_H17 cells produce a number of inflammatory cytokines, including IL-17 and IL-22, and T_{FH} cells help B cells by constitutively producing IL-21. In a number of autoimmune diseases, such as rheumatoid arthritis, multiple sclerosis, type 1 diabetes mellitus, and Crohn's disease, there appears to be biased differentiation of T_H1 cells, with resultant organ damage. More recently, studies suggest accentuated differentiation of T_H17 cells associated with animal models of inflammatory arthritis and also rheumatoid arthritis, whereas increased differentiation of T_{FH} cells has been associated with animal models of SLE.

Organ-Specific versus Systemic Autoimmune Diseases

Autoimmune diseases form a spectrum, from those specifically affecting a single organ to systemic disorders with involvement of many organs **(Table 318-5)**. Hashimoto's autoimmune thyroiditis is an example of an organ-specific autoimmune disease **(Chap. 341)**. In this disorder, there is a specific lesion in the thyroid associated with infiltration of mononuclear cells and damage to follicular cells. Antibody to thyroid constituents can be demonstrated in nearly all cases. Other organ- or tissue-specific autoimmune disorders include pemphigus vulgaris, autoimmune hemolytic anemia, idiopathic

thrombocytopenic purpura, Goodpasture's syndrome, myasthenia gravis, and sympathetic ophthalmia. One important feature of some organ-specific autoimmune diseases is the tendency for overlap, such that an individual with one specific syndrome is more likely to develop a second syndrome. For example, there is a high incidence of pernicious anemia in individuals with autoimmune thyroiditis. More striking is the tendency for individuals with an organ-specific autoimmune disease to develop multiple other manifestations of autoimmunity without the development of associated organ pathology. Thus, as many as 50% of individuals with pernicious anemia have non-cross-reacting antibodies to thyroid constituents, whereas patients with myasthenia gravis may develop antinuclear antibodies, antithyroid antibodies, rheumatoid factor, antilymphocyte antibodies, and polyclonal hypergammaglobulinemia. Part of the explanation for this may relate to the genetic elements shared by individuals with these different diseases.

Table 318-5 Some Autoimmune Diseases

Organ Specific	
Graves' disease	Vitiligo
Hashimoto's thyroiditis	Autoimmune hemolytic anemia
Autoimmune polyglandular syndrome	Autoimmune thrombocytopenic purpura
Type 1 diabetes mellitus	Pernicious anemia
Insulin-resistant diabetes mellitus	Myasthenia gravis
Immune-mediated infertility	Multiple sclerosis
Autoimmune Addison's disease	Guillain-Barré syndrome
Pemphigus vulgaris	Stiff-man syndrome
Pemphigus foliaceus	Acute rheumatic fever
Dermatitis herpetiformis	Sympathetic ophthalmia
Autoimmune alopecia	Goodpasture's syndrome
Organ Nonspecific (Systemic)	
Systemic lupus erythematosus	Granulomatosis with polyangiitis (Wegener's)
Rheumatoid arthritis	Antiphospholipid syndrome
Systemic necrotizing vasculitis	Sjögren's syndrome

Systemic autoimmune diseases differ from organ-specific diseases in that pathologic lesions are found in multiple diverse organs and tissues. The hallmark of these conditions is the demonstration of associated relevant autoimmune manifestations that are likely to be etiologic in the organ pathology. SLE represents the prototype of these disorders because of its abundance of autoimmune manifestations.

SLE is a disease of protean manifestations that characteristically involves the kidneys, joints, skin, serosal surfaces, blood vessels, and central nervous system **(Chap. 319)**. The disease is associated with a vast array of autoantibodies whose production appears to be a part of a generalized hyperreactivity of the humoral immune system. Other features of SLE include generalized B cell hyperresponsiveness and polyclonal hypergammaglobulinemia. Current evidence suggests that both hypo- and hyperresponsiveness to antigen can lead to survival and activation of autoreactive B cells in SLE.

Treatment: Autoimmune Diseases

Treatment of autoimmune diseases can focus on either suppressing the induction of autoimmunity, restoring normal regulatory mechanisms, or inhibiting the effector mechanisms. To eliminate autoreactive cells, immunosuppressive or ablative therapies are most commonly used. In recent years, cytokine blockade has been demonstrated to be effective in preventing immune activation in some diseases. New therapies have also been developed to target lymphoid cells more specifically, either by blocking a co-stimulatory signal needed for T or B cell activation, by blocking the migratory capacity of lymphocytes, or by eliminating the effector T cells or B cells. The efficacy of these therapies is not yet demonstrated. Newer trials are testing the possibility of using autoantigen itself to induce tolerance. One major advance in inhibiting effector mechanisms has been the introduction of cytokine blockade, targeting TNF or IL-1, that appears to limit organ damage in some diseases. Biologicals that interface

with T cell activation (CTLA-4Ig) or delete B cells (anti-CD20 antibody) have also recently been approved for the treatment of rheumatoid arthritis. Therapies that prevent target organ damage or support target organ function remain an important therapeutic approach to autoimmune disease.

FURTHER READINGS

Annunziato F et al: Type 17 T helper cells—Origins, features and possible roles in rheumatic disease. Nat Rev 5:325, 2009[PMID: 19434074] [Full Text]

Baechler EC et al: Gene expression profiling in human autoimmunity. Immunol Rev 210:120, 2006[PMID: 16623768] [Full Text]

Drexler SK, Foxwell BM: The role of Toll-like receptors in chronic inflammation. Int J Biochem Cell Biol 42:506, 2010[PMID: 19837184] [Full Text]

Fujinami RS et al: Molecular mimicry, bystander activation, or viral persistence: Infections and autoimmune disease. Clin Microbiol Rev 19:80, 2006[PMID: 16418524] [Full Text]

Gauld SB et al: Silencing of autoreactive B cells by anergy: A fresh perspective. Curr Opin Immunol 18:292, 2006[PMID: 16616480] [Full Text]

Mackay IR: Clustering and commonalities among autoimmune diseases. J Autoimmun 33:170, 2009[PMID: 19837564] [Full Text]

Maier LM, Hafler DA: Autoimmunity risk alleles in costimulation pathways. Immunol Rev 229:322, 2009[PMID: 19426231] [Full Text]

Pisetsky DS: The role of innate immunity in the induction of autoimmunity. Autoimmun Rev 8:69, 2008[PMID: 18708168] [Full Text]

Podojil JR, Miller SD: Molecular mechanisms of T-cell receptor and costimulatory molecule ligation/blockade in autoimmune disease therapy. Immunol Rev 229:337, 2009[PMID: 19426232] [Full Text]

Rosen A, Casciola-Rosen L: Autoantigens in systemic autoimmunity: Critical partner in pathogenesis. J Intern Med 265:625, 2009[PMID: 19493056] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION AND PREVALENCE

Systemic lupus erythematosus (SLE) is an autoimmune disease in which organs and cells undergo damage initially mediated by tissue-binding autoantibodies and immune complexes. In most patients, autoantibodies are present for a few years before the first clinical symptom appears; clinical manifestations are heterogeneous. Ninety percent of patients at diagnosis are women of childbearing years; people of all genders, ages, and ethnic groups are susceptible. Prevalence of SLE in the United States is 10 to 400 per 100,000 depending on race and gender; highest prevalence is in black women and lowest is in white men.

PATHOGENESIS AND ETIOLOGY

The proposed pathogenic mechanisms of SLE are illustrated in **Fig. 319-1**. Interactions between susceptibility genes and environmental factors result in abnormal immune responses, which vary among different patients. Those responses may include (1) activation of innate immunity (dendritic cells, monocyte/macrophages) by CpG DNA, DNA in immune complexes, viral RNA, and RNA in RNA/protein self-antigens; (2) lowered activation thresholds and abnormal activation pathways in adaptive immunity cells (T and B lymphocytes); (3) ineffective regulatory CD4+ and CD8+ T cells; and (4) reduced clearance of immune complexes and of apoptotic cells. Self-antigens (nucleosomal DNA/protein; RNA/protein in Sm, Ro, and La; phospholipids) are available for recognition by the immune system in surface blebs of apoptotic cells; thus antigens, autoantibodies, and immune complexes persist for prolonged periods of time, allowing inflammation and disease to develop. Immune cell activation is accompanied by increased secretion of proinflammatory type 1 and 2 interferons (IFNs), tumor necrosis factor α (TNF-α), interleukin (IL)-17 and B cell–maturation/survival cytokines B lymphocyte stimulator (BLyS/BAFF), and IL-10. Upregulation of genes induced by interferons is a genetic "signature" in peripheral blood cells of SLE in approximately 50% of patients. Decreased production of other cytokines also contributes to SLE: Lupus T and natural killer (NK) cells fail to produce enough IL-2 and transforming growth factor β (TGF-β) to induce and sustain regulatory CD4+ and CD8+ T cells. The result of these abnormalities is sustained production of autoantibodies (referred to in **Fig. 319-1** and described in **Table 319-1)** and immune complexes; pathogenic subsets bind target tissues, with activation of complement, leading to release of cytokines, chemokines, vasoactive peptides, oxidants, and destructive enzymes. This is accompanied by influx into target tissues of T cells, monocyte/macrophages, and dendritic cells, as well as activation of resident macrophages and dendritic cells. In the setting of chronic inflammation, accumulation of growth factors and products of chronic oxidation contribute to irreversible tissue damage, including fibrosis/sclerosis, in glomeruli, arteries, brain, lungs, and other tissues.

Figure 319-1

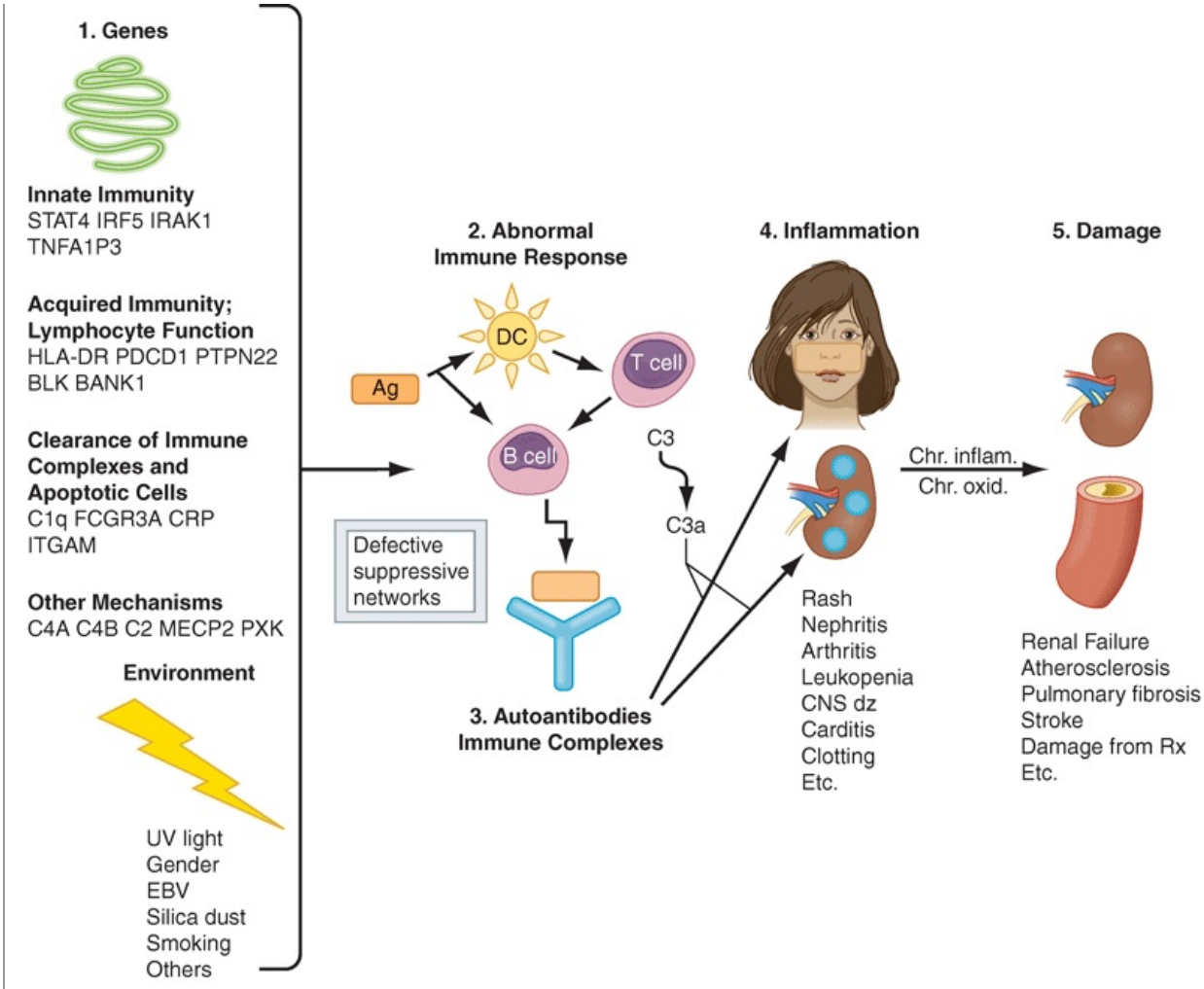

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenesis of SLE. Genes confirmed in more than one genome-wide association analysis in Northern European whites as increasing susceptibility to SLE or lupus nephritis are listed (reviewed in Moser KL et al, Recent insights into the genetic basis of SLE. Genes Immun 2009:10:373). Gene-environment interactions result in abnormal immune responses that generate pathogenic autoantibodies and immune complexes that deposit in tissue, activate complement, cause inflammation, and over time lead to irreversible organ damage. Ag, antigen; C1q, complement system; C3, complement component; CNS, central nervous system; DC, dendritic cell; EBV, Epstein-Barr virus; HLA, human leukocyte antigen; FcR, immunoglobulin Fc-binding receptor; IL, interleukin; MCP, monocyte chemotactic protein; PTPN, phosphotyrosine phosphatase; UV, ultraviolet.

Table 319-1 Autoantibodies in Systemic Lupus Erythematosus (SLE)

Antibody	Prevalence, %	Antigen Recognized	Clinical Utility
Antinuclear antibodies	98	Multiple nuclear	Best screening test; repeated negative tests make SLE unlikely
Anti-dsDNA	70	DNA (double-stranded)	High titers are SLE-specific and in some patients correlate with disease activity, nephritis, vasculitis
Anti-Sm	25	Protein complexed to 6 species of nuclear U1 RNA	Specific for SLE; no definite clinical correlations; most patients also have anti-RNP; more common in blacks and Asians than whites
Anti-RNP	40	Protein complexed to U1 RNA	Not specific for SLE; high titers associated with syndromes that have overlap features of several rheumatic syndromes including SLE; more common in blacks than whites
Anti-Ro (SS-A)	30	Protein complexed to	Not specific for SLE; associated with sicca syndrome, predisposes to subacutecutaneous lupus,

		hY RNA, primarily 60 kDa and 52 kDa	and to neonatal lupus with congenital heart block; associated with decreased risk for nephritis
Anti-La (SS-B)	10	47-kDa protein complexed to hY RNA	Usually associated with anti-Ro; associated with decreased risk for nephritis
Antihistone	70	Histones associated with DNA (in nucleosome, chromatin)	More frequent in drug-induced lupus than in SLE
Antiphospholipid	50	Phospholipids, β2 glycoprotein 1 cofactor, prothrombin	Three tests available—ELISAs for cardiolipin and β2G1, sensitive prothrombin time (DRVVT); predisposes to clotting, fetal loss, thrombocytopenia
Antierythrocyte	60	Erythrocyte membrane	Measured as direct Coombs' test; a small proportion develops overt hemolysis
Antiplatelet	30	Surface and altered cytoplasmic antigens on platelets	Associated with thrombocytopenia but sensitivity and specificity are not good; this is not a useful clinical test
Antineuronal (includes anti-glutamate receptor)	60	Neuronal and lymphocyte surface antigens	In some series a positive test in CSF correlates with active CNS lupus.
Antiribosomal P	20	Protein in ribosomes	In some series a positive test in serum correlates with depression or psychosis due to CNS lupus

Abbreviations: CNS, central nervous system; CSF, cerebrospinal fluid; DRVVT, dilute Russell viper venom time; ELISA, enzyme-linked immunosorbent assay.

SLE is a multigenic disease. Rare single-gene defects confer high hazard ratios (HR) for SLE (5–25), including homozygous deficiencies of early components of complement (C1q,r,s; C2; C4) and a mutation in TREX1 on the X chromosome. In most genetically susceptible individuals, normal alleles of multiple genes each contribute a small amount to abnormal immune/inflammation/tissue damage responses; if enough predisposing variations are present, disease results. Thirty to forty predisposing genes (examples listed in Fig 319-1) have been identified in recent genome-wide association studies in thousands of Northern European white patients and controls. They confer HR for SLE of 1.5–3. Such relatively weak gene polymorphisms that increase risk for SLE can be classified by their potential role in pathogenesis. Predisposing, antigen-presenting human leukocyte antigen (HLA)-molecules are most commonly found, in multiple ethnic groups (HLA DRB1 *0301 and *1501, as well as multiple genes across the 120-gene region). Other genetic factors in whites include innate immunity pathway gene polymorphisms, especially associated with interferon alpha (STAT4, IRF5, IRAK1. TNFAIP3, PTPN22), genes in lymphocyte signaling pathways (PTPN22, PDCD-1, Ox40L, BANK-1, LYN, BLK), genes that affect clearance of apoptotic cells or immune complexes (C1q, FCRG IIA and IIIA, CRP, ITGAM), and genes that influence neutrophil adherence (ITGAM), and endothelial cell function (TREX-1). Some polymorphisms influence clinical manifestations; such as single nucleotide polymorphisms (SNPs) of STAT 4 that associate with severe disease, anti-DNA, nephritis, and anti-phospholipid syndrome **(Chap. 320)**, and an allele of FCGRIIA encoding a receptor that binds immune complexes poorly and predisposes to nephritis. Some gene effects are in promoter regions (e.g., IL-10) and others are conferred by copy numbers (e.g., C4A). In addition to genome-encoded susceptibility and protective genes, the influence of certain micro (mi) RNAs on gene transcription, as well as posttranscriptional epigenetic modification of DNA, which is hypomethylated in SLE, also contribute to disease susceptibility.

Some genepolymorphisms contribute to several autoimmune diseases, such as STAT4 and CTLA4. All these gene polymorphisms/transcription/epigenetic combinations influence immune responses to the external and internal environment; when such responses are too high and/or too prolonged and/or inadequately regulated, autoimmune disease results.

Female sex is permissive for SLE with evidence for hormone effects, genes on the X chromosome, and epigenetic differences between genders playing a role. Females of many mammalian species make higher antibody responses than males. Women exposed to estrogen-containing oral contraceptives or hormone replacement have an increased risk of developing SLE (1.2–2-fold). Estradiol binds to

receptors on T and B lymphocytes, increasing activation and survival of those cells, thus favoring prolonged immune responses. Genes on the X chromosome that influence SLE, such as TREX-1, may play a role in gender predisposition—possibly because some genes on the second X in females are not silent. People with XXY karyotype (Klinefelter's syndrome) have a significantly increased risk for SLE.

Several environmental stimuli may influence SLE (Fig. 319-1). Exposure to ultraviolet light causes flares of SLE in approximately 70% of patients, possibly by increasing apoptosis in skin cells or by altering DNA and intracellular proteins to make them antigenic. It is likely that some infections induce a normal immune response that matures to contain some T and B cells that recognize self-antigens; such cells are not appropriately regulated, and autoantibody production occurs. Most SLE patients have autoantibodies for 3 years or more before the first symptoms of disease, suggesting that regulation controls the degree of autoimmunity for years before quantities and qualities of autoantibodies and pathogenic B and T cells cause clinical disease. Epstein-Barr virus (EBV) may be one infectious agent that can trigger SLE in susceptible individuals. Children and adults with SLE are more likely to be infected by EBV than age-, sex-, and ethnicity-matched controls. EBV contains amino acid sequences that mimic sequences on human spliceosomes (RNA/protein antigens) often recognized by autoantibodies in people with SLE. Current tobacco smoking increases risk for SLE [odds ratio (OR) 1.5]. Prolonged occupational exposure to silica (e.g., inhalation of soap powder dust) increases risk (OR 4.3) in black women. Thus, interplay between genetic susceptibility, environment, gender, and abnormal immune responses results in autoimmunity **(Chap. 318)**.

PATHOLOGY

In SLE, biopsies of affected skin show deposition of Ig at the dermal-epidermal junction (DEJ), injury to basal keratinocytes, and inflammation dominated by T lymphocytes in the DEJ and around blood vessels and dermal appendages. Clinically unaffected skin may also show Ig deposition at the DEJ.

In renal biopsies, the pattern and severity of injury are important in diagnosis and in selecting the best therapy. Many clinical studies of lupus nephritis have used the World Health Organization (WHO) classification of lupus nephritis. However, the International Society of Nephrology (ISN) and the Renal Pathology Society (RPS) have published a newer, similar classification **(Table 319-2)** that is replacing WHO standards. An advantage of the ISN/RPS classification is the addition of "a" for active and "c" for chronic changes, giving the physician information regarding the potential reversibility of disease. All the classification systems focus on glomerular disease, although the presence of tubular interstitial and vascular disease is important to clinical outcomes. In general, class III and IV disease, as well as class V accompanied by III or IV disease, should be treated with aggressive immunosuppression if possible, because there is a high risk for end-stage renal disease (ESRD) if patients are untreated or undertreated. Treatment for lupus nephritis is not recommended in patients with class I or II disease or with extensive irreversible changes. In children, a diagnosis of SLE can be established on the basis of renal histology without meeting additional diagnostic criteria **(Table 319-3)**.

Table 319-2 Classification of Lupus Nephritis (International Society of Nephrology and Renal Pathology Society)

Class I: Minimal Mesangial Lupus Nephritis

Normal glomeruli by light microscopy, but mesangial immune deposits by immunofluorescence.

Class II: Mesangial Proliferative Lupus Nephritis

Purely mesangial hypercellularity of any degree or mesangial matrix expansion by light microscopy, with mesangial immune deposits. A few isolated subepithelial or subendothelial deposits may be visible by immunofluorescence or electron microscopy, but not by light microscopy.

Class III: Focal Lupus Nephritis

Active or inactive focal, segmental or global endo- or extracapillary glomerulonephritis involving <50% of all glomeruli, typically with focal subendothelial immune deposits, with or without mesangial alterations.

Class III (A): Active lesions—focal proliferative lupus nephritis

Class III (A/C): Active and chronic lesions—focal proliferative and sclerosing lupus nephritis

Class III (C): Chronic inactive lesions with glomerular scars—focal sclerosing lupus nephritis

Class IV: Diffuse Lupus Nephritis

Active or inactive diffuse, segmental or global endo- or extracapillary glomerulonephritis involving ≥ 50% of all glomeruli, typically with diffuse subendothelial immune deposits, with or without mesangial alterations. This class is divided into diffuse segmental (IV-S) lupus nephritis when ≥50% of the involved glomeruli have segmental lesions, and diffuse global (IV-G) lupus nephritis when ≥50% of the involved glomeruli have global lesions. Segmental is defined as a glomerular lesion that involves less than one-half of the glomerular tuft. This class includes cases with diffuse wire loop deposits but with little or no glomerular proliferation.

Class IV-S (A): Active lesions—diffuse segmental proliferative lupus nephritis

Class IV-G (A): Active lesions—diffuse global proliferative lupus nephritis

Class IV-S (A/C): Active and chronic lesions—diffuse segmental proliferative and sclerosing lupus nephritis

Class IV-G (A/C): Active and chronic lesions—diffuse global proliferative and sclerosing lupus nephritis

Class IV-S (C): Chronic inactive lesions with scars—diffuse segmental sclerosing lupus nephritis

Class IV-G (C): Chronic inactive lesions with scars—diffuse global sclerosing lupus nephritis

Class V: Membranous Lupus Nephritis

Global or segmental subepithelial immune deposits or their morphologic sequelae by light microscopy and by immunofluorescence or electron microscopy, with or without mesangial alterations. Class V lupus nephritis may occur in combination with class III or IV, in which case both will be diagnosed. Class V lupus nephritis may show advanced sclerosis.

Class VI: Advanced Sclerotic Lupus Nephritis

≥90% of glomeruli globally sclerosed without residual activity.

Note: Indicate and grade (mild, moderate, severe) tubular atrophy, interstitial inflammation and fibrosis, severity of arteriosclerosis or other vascular lesions.

Source: JJ Weening et al: Kidney Int 65:521, 2004. Reprinted by permission from Macmillan Publishers Ltd., Copyright 2004.

Histologic abnormalities in blood vessels may also determine therapy. Patterns of vasculitis are not specific for SLE but may indicate active disease: leukocytoclastic vasculitis is most common **(Chap. 326)**.

Lymph node biopsies are usually performed to rule out infection or malignancies. In SLE, they show nonspecific diffuse chronic inflammation.

DIAGNOSIS

The diagnosis of SLE is based on characteristic clinical features and autoantibodies. Current criteria for classification are listed in Table 319-3, and an algorithm for diagnosis and initial therapy is shown in **Fig. 319-2**. The criteria are intended for confirming the diagnosis of SLE in patients included in studies; the author uses them in individual patients for estimating the probability that a disease is SLE. Any combination of ≥4 of 11 criteria, well documented at any time during an individual's history, makes it likely that the patient has SLE. (Specificity and sensitivity are ~95% and ~75%, respectively.) In many patients, criteria accrue over time. Antinuclear antibodies (ANA) are positive in >98% of patients during the course of disease; repeated negative tests suggest that the diagnosis is not SLE, unless other autoantibodies are present (Fig. 319-2). High-titer IgG antibodies to double-stranded DNA and antibodies to the Sm antigen are both specific for SLE and, therefore, favor the diagnosis in the presence of compatible clinical manifestations. The presence in an individual of multiple autoantibodies without clinical symptoms should not be considered diagnostic for SLE, although such persons are at increased risk.

Table 319-3 Diagnostic Criteria for Systemic Lupus Erythematosus

Malar rash	Fixed erythema, flat or raised, over the malar eminences
Discoid rash	Erythematous circular raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur
Photosensitivity	Exposure to ultraviolet light causes rash
Oral ulcers	Includes oral and nasopharyngeal ulcers, observed by physician
Arthritis	Nonerosive arthritis of two or more peripheral joints, with tenderness, swelling, or effusion
Serositis	Pleuritis or pericarditis documented by ECG or rub or evidence of effusion
Renal disorder	Proteinuria >0.5 g/d or √3+, or cellular casts
Neurologic disorder	Seizures or psychosis without other causes
Hematologic disorder	Hemolytic anemia or leukopenia (<4000/μL) or lymphopenia (<1500/μL) or thrombocytopenia(<100,000/μL) in the absence of offending drugs

Immunologic disorder	Anti-dsDNA, anti-Sm, and/or anti-phospholipid
Antinuclear antibodies	An abnormal titer of ANA by immunofluorescence or an equivalent assay at any point in time in the absence of drugs known to induce ANAs

If ≥4 of these criteria, well documented, are present at any time in a patient's history, the diagnosis is likely to be SLE. Specificity is ~95%; sensitivity is ~75%.

Abbreviations: ANA, antinuclear antibodies; dsDNA, double-strand DNA; ECG, electrocardiography.

Source: Criteria published by EM Tan et al: Arthritis Rheum 25:1271, 1982; update by MC Hochberg, Arthritis Rheum 40:1725, 1997.

Figure 319-2

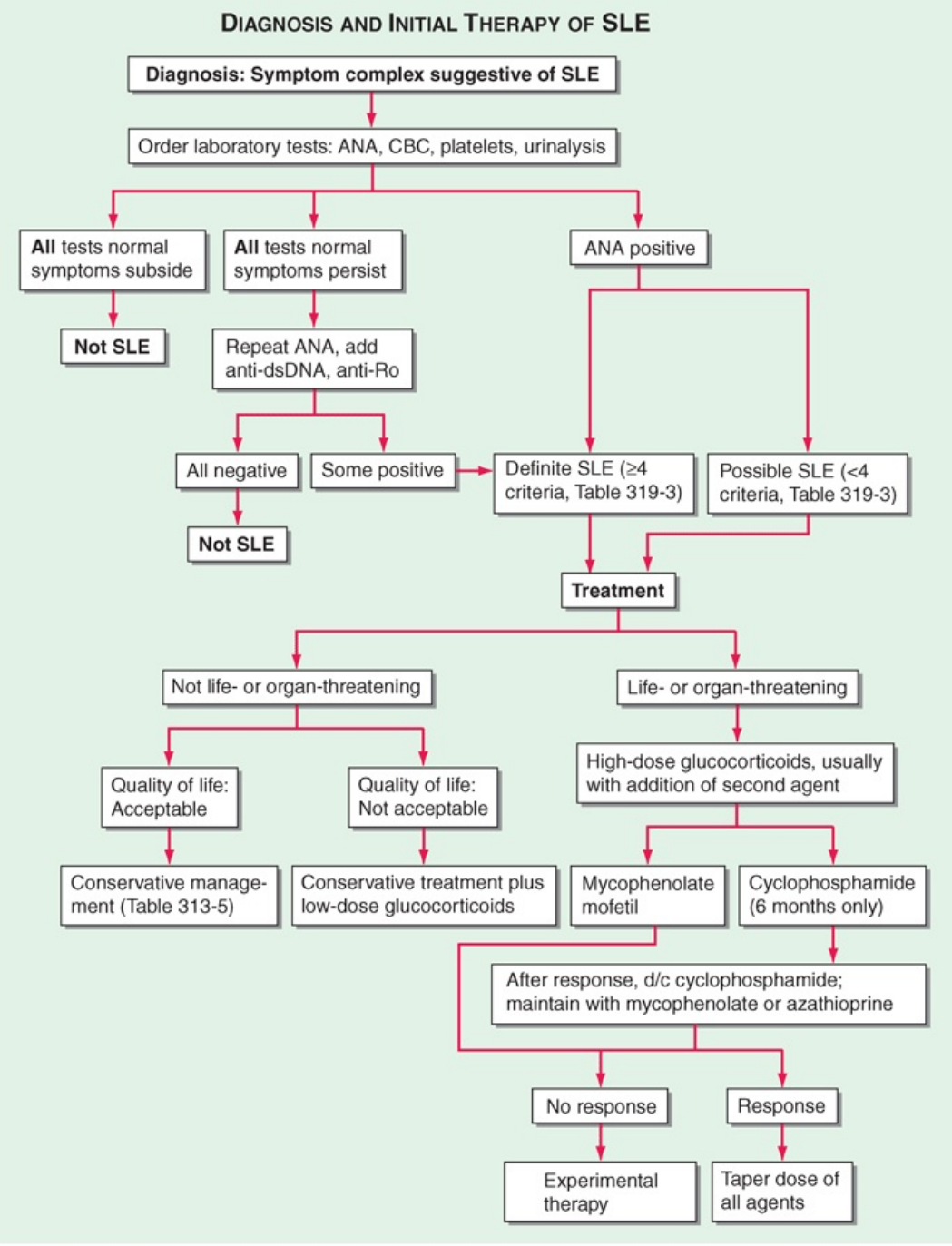

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for diagnosis and initial therapy of SLE. ANA, antinuclear antibodies; CBC, complete blood count.

INTERPRETATION OF CLINICAL MANIFESTATIONS

When a diagnosis of SLE is made, it is important to establish the severity and potential reversibility of the illness and to estimate the possible consequences of various therapeutic interventions. In the following sections, descriptions of some disease manifestations begin with relatively mild problems and progress to those more life-threatening.

OVERVIEW AND SYSTEMIC MANIFESTATIONS

At its onset, SLE may involve one or several organ systems; over time, additional manifestations may occur **(Tables 319-3 and 319-4)**. Most of the autoantibodies characteristic of each person are present at the time clinical manifestations appear (Tables 319-1 and 319-3). Severity of SLE varies from mild and intermittent to severe and fulminant. Most patients experience exacerbations interspersed with periods of relative quiescence; permanent complete remissions (absence of symptoms with no treatment) are rare. Systemic symptoms, particularly fatigue and myalgias/arthralgias, are present most of the time. Severe systemic illness requiring glucocorticoid therapy can occur with fever, prostration, weight loss, and anemia with or without other organ-targeted manifestations.

Table 319-4 Clinical Manifestations of SLE and Prevalence over the Entire Course of Disease*

Manifestation	Prevalence, %
Systemic: Fatigue, malaise, fever, anorexia, weight loss	95
Musculoskeletal	95
Arthralgias/myalgias	95
Nonerosive polyarthritis	60
Hand deformities	10
Myopathy/myositis	25/5
Ischemic necrosis of bone	15
Cutaneous	80
Photosensitivity	70
Malar rash	50
Oral ulcers	40
Alopecia	40
Discoid rash	20
Vasculitis rash	20
Other (e.g., urticaria, subacute cutaneous lupus)	15
Hematologic	85
Anemia (chronic disease)	70
Leukopenia (<4000/μL)	65
Lymphopenia (<1500/μL)	50
Thrombocytopenia (μ100,000/μL)	15
Lymphadenopathy	15
Splenomegaly	15
Hemolytic anemia	10
Neurologic	60
Cognitive disorder	50
Mood disorder	40

Headache	25
Seizures	20
Mono-, polyneuropathy	15
Stroke, TIA	10
Acute confusional state or movement disorder	2–5
Aseptic meningitis, myelopathy	<1
Cardiopulmonary	60
Pleurisy, pericarditis, effusions	30–50
Myocarditis, endocarditis	10
Lupus pneumonitis	10
Coronary artery disease	10
Interstitial fibrosis	5
Pulmonary hypertension, ARDS, hemorrhage	<5
Shrinking lung syndrome	<5
Renal	30–50
Proteinuria ≥500 mg/24 h, cellular casts	30–50
Nephrotic syndrome	25
End-stage renal disease	5–10
Gastrointestinal	40
Nonspecific (nausea, mild pain, diarrhea)	30
Abnormal liver enzymes	40
Vasculitis	5
Thrombosis	15
Venous	10
Arterial	5
Ocular	15
Sicca syndrome	15
Conjunctivitis, episcleritis	10
Vasculitis	5

*Numbers indicate percent of patients who have the manifestation at some time during the course of illness.

Abbreviations: ARDS, acute respiratory distress syndrome; TIA, transient ischemic attack.

Musculoskeletal Manifestations

Most people with SLE have intermittent polyarthritis, varying from mild to disabling, characterized by soft tissue swelling and tenderness in joints, most commonly in hands, wrists, and knees. Joint deformities (hands and feet) develop in only 10% of patients. Erosions on joint x-rays are rare; their presence suggests a non-lupus inflammatory arthropathy such as rheumatoid arthritis **(Chap. 321)**; some experts think that erosions can occur in SLE. If pain persists in a single joint, such as knee, shoulder, or hip, a diagnosis of ischemic necrosis of bone should be considered, particularly if there are no other manifestations of active SLE. The prevalence of ischemic necrosis of bone is increased in SLE, especially in patients treated with systemic glucocorticoids. Myositis with clinical muscle weakness, elevated creatine kinase levels, positive MRI scan, and muscle necrosis and inflammation on biopsy can occur, although most patients have myalgias without frank myositis. Glucocorticoid

therapies (commonly) and antimalarial therapies (rarely) can also cause muscle weakness; these adverse effects must be distinguished from active disease.

Cutaneous Manifestations

Lupus dermatitis can be classified as discoid lupus erythematosus (DLE), systemic rash, subacutecutaneous lupus erythematosus (SCLE), or "other." Discoid lesions are roughly circular with slightly raised, scaly hyperpigmented erythematous rims and depigmented, atrophic centers in which all dermal appendages are permanently destroyed. Lesions can be disfiguring, particularly on the face and scalp. Treatment consists primarily of topical or locally injected glucocorticoids and systemic antimalarials. Only 5% of people with DLE have SLE (although one-half have positive ANA); however, among individuals with SLE, as many as 20% have DLE. The most common SLE rash is a photosensitive, slightly raised erythema, occasionally scaly, on the face (particularly the cheeks and nose—the "butterfly" rash), ears, chin, V region of the neck and chest, upper back, and extensor surfaces of the arms. Worsening of this rash often accompanies flare of systemic disease. SCLE consists of scaly red patches similar to psoriasis, or circular flat red-rimmed lesions. Patients with these manifestations are exquisitely photosensitive; most have antibodies to Ro (SS-A). Other SLE rashes include recurring urticaria, lichen planus–like dermatitis, bullae, and panniculitis ("lupus profundus"). Rashes can be minor or severe; they may be the major disease manifestation. Small, painful ulcerations on the oral or nasal mucosa are common in SLE; the lesions resemble aphthous ulcers.

Renal Manifestations

Nephritis is usually the most serious manifestation of SLE, particularly since nephritis and infection are the leading causes of mortality in the first decade of disease. Since nephritis is asymptomatic in most lupus patients, urinalysis should be ordered in any person suspected of having SLE. The classification of lupus nephritis is primarily histologic (see "Pathology," above, and Table 319-2). Renal biopsy is useful in planning current and near-future therapies. Patients with dangerous proliferative forms of glomerular damage (ISN III and IV) usually have microscopic hematuria and proteinuria (>500 mg per 24 h); approximately one-half develop nephrotic syndrome, and most develop hypertension. If diffuse proliferative glomerulonephritis (DPGN) is untreated, virtually all patients develop ESRD within 2 years of diagnosis. Therefore, aggressive immunosuppression is indicated (usually systemic glucocorticoids plus a cytotoxic drug), unless 90% of glomeruli have irreversible damage **(Fig. 319-2, Table 319-5)**. Blacks are more likely to develop ESRD than are whites, even with the most current therapies. Overall in the United States, ~20% of individuals with lupus DPGN die or develop ESRD within 10 years of diagnosis. Such individuals require aggressive control of SLE and of the complications of renal disease and of therapy. A small proportion of SLE patients with proteinuria (usually nephrotic) have membranous glomerular changes without proliferation on renal biopsy. Their outcome is better than for those with DPGN. Lupus nephritis tends to be an ongoing disease, with flares requiring retreatment or intensification of treatment over many years. For most people with lupus nephritis, accelerated atherosclerosis becomes important after several years of disease; attention must be given to control of systemic inflammation, blood pressure, hyperlipidemia, and hyperglycemia.

Table 319-5 Medications for the Management of SLE

Medication	Dose Range	Drug Interactions	Serious or Common Adverse Effects
NSAIDs, salicylates (Ecotrin[a] and St. Joseph's aspirin[a] approved by FDA for use in SLE)	Doses toward upper limit of recommended range usually required	A2R/ACE inhibitors, glucocorticoids, fluconazole, methotrexate, thiazides	NSAIDs: Higher incidence of aseptic meningitis, transaminitis, decreased renal function, vasculitis of skin; entire class, especially COX-2-specific inhibitors, may increase risk for myocardial infarction Salicylates: ototoxicity, tinnitus Both: GI events and symptoms, allergic reactions, dermatitis, dizziness, acute renal failure, edema, hypertension
Topical glucocorticoids	Mid-potency for face; mid to high potency other areas	None known	Atrophy of skin, contact dermatitis, folliculitis, hypopigmentation, infection
Topical sunscreens	SPF 15 at least; 30+ preferred	None known	Contact dermatitis

Hydroxychloroquine[a] (quinacrine can be added or substituted)	200–400 mg qd (100 mg qd)	None known	Retinal damage, agranulocytosis, aplastic anemia, ataxia, cardiomyopathy, dizziness, myopathy, ototoxicity, peripheral neuropathy, pigmentation of skin, seizures, thrombocytopenia. Use in pregnancy may be acceptable. Pregnancy category DQuinacrine usually causes diffuse yellow skin coloration
DHEA (dehydroepiandrosterone)	200 mg qd	Unclear	Acne, menstrual irregularities, high serum levels of testosterone
Methotrexate (for dermatitis, arthritis)	10–25 mg once a week, PO or SC, with folic acid; decrease dose if CrCl <60 mL/min	Acitretin, leflunomide, NSAIDs and salicylates, penicillins, probenecid, sulfonamides, trimethoprim	Anemia, bone marrow suppression, leukopenia, thrombocytopenia, hepatotoxicity, nephrotoxicity, infections, neurotoxicity, pulmonary fibrosis, pneumonitis, severe dermatitis, seizures. Teratogenic. Pregnancy category X
Glucocorticoids, oral[a] (several specific brands are approved by FDA for use in SLE)	Prednisone, prednisolone: 0.5–1 mg/kg per day for severe SLE0.07–0.3 mg/kg per day or qod for milder disease	A2R/ACE antagonists, antiarrhythmics class III, β_2, cyclosporine, NSAIDs and salicylates, phenothiazines, phenytoins, quinolones, rifampin, risperidone, thiazides, sulfonylureas, warfarin	Infection, VZV infection, hypertension, hyperglycemia, hypokalemia, acne, allergic reactions, anxiety, aseptic necrosis of bone, cushingoid changes, CHF, fragile skin, insomnia, menstrual irregularities, mood swings, osteoporosis, psychosis
Methylprednisolone sodium succinate, IV[a] (FDA approved for lupus nephritis)	For severe disease, 1 g IV qd x 3 days	As for oral glucocorticoids	As for oral glucocorticoids (if used repeatedly); anaphylaxis
Cyclophosphamide[b,c]?IV	7–25 mg/kg q month x 6; consider mesna administration with dose	Allopurinol, bone marrow suppressants, colony-stimulating factors, doxorubicin, rituximab, succinylcholine, zidovudine	Infection, VZV infection, bone marrow suppression, leukopenia, anemia, thrombocytopenia, hemorrhagic cystitis (less with IV), carcinoma of the bladder, alopecia, nausea, diarrhea, malaise, malignancy, ovarian and testicular failure. Teratogenic. Pregnancy category D
Mycophenolate mofetil[b] or mycophenolic acid	MMF: 2–3 g/d PO; max 1 g bid if CrCl <25 mL/min.MPA: 360–1080 mg bid. Caution if CrCl <25 mL/min	Acyclovir, antacids, azathioprine, bile acid –binding resins, ganciclovir, iron, salts, probenecid, oral contraceptives	Infection, leukopenia, anemia, thrombocytopenia, lymphoma, lymphoproliferative disorders, malignancy, alopecia, cough, diarrhea, fever, GI symptoms, headache, hypertension, hypercholesterolemia, hypokalemia, insomnia, peripheral edema, transaminitis, tremor, rash. Teratogenic. Pregnancy category D
Azathioprine[b]	2–3 mg/kg per day PO; decrease frequency of dose if CrCl <50 mL/min	ACE inhibitors, allopurinol, bone marrow suppressants, interferons, mycophenolate mofetil, rituximab, warfarin, zidovudine	Infection, VZV infection, bone marrow suppression, leukopenia, anemia, thrombocytopenia, pancreatitis, hepatotoxicity, malignancy, alopecia, fever, flulike illness, GI symptoms. Use in pregnancy may be acceptable. Pregnancy category D
Belimumab	10 mg/kg i.v.	wk 0,2,4 then monthly	Infusion reactions Allergy Infections probable
Rituximab (for patients resistant to above	375 mg/M2 q wk x 4 or 1 g	IVIg	Infection (including PML), infusion reactions, headache, arrythmias,

therapies)	q 2 wks x 2[c]		allergic responses. Pregnancy category C
Belimumab	10 mg/kg i.v.	IV Ig	Infec

[a]Indicates medication is approved for use in SLE by the U.S. Food and Drug Administration.

[b]Indicates the medication has been used with glucocorticoids in the trials showing efficacy.

[c]See text for low dose regimen.

Abbreviations: A2R, angiotensin 2 receptor; ACE, angiotensin-converting enzyme; CHF, congestive heart failure; CrCl, creatinine clearance; FDA, U.S. Food and Drug Administration; GI, gastrointestinal; NSAIDs, nonsteroidal anti-inflammatory drugs; SPF, sun protection factor; VZV, varicella-zoster virus.

Nervous System Manifestations

There are many central nervous system (CNS) and peripheral nervous system manifestations of SLE; in some patients these are the major cause of morbidity and mortality. It is useful to approach this diagnostically by asking first whether the symptoms result from SLE or another condition (such as infection in immunosuppressed individuals). If symptoms are related to SLE, it should be determined whether they are caused by a diffuse process (requiring immunosuppression) or vascular occlusive disease (requiring anticoagulation). The most common manifestation of diffuse CNS lupus is cognitive dysfunction, including difficulties with memory and reasoning. Headaches are also common. When excruciating, they often indicate SLE flare; when milder, they are difficult to distinguish from migraine or tension headaches. Seizures of any type may be caused by lupus; treatment often requires both antiseizure and immunosuppressive therapies. Psychosis can be the dominant manifestation of SLE; it must be distinguished from glucocorticoid-induced psychosis. The latter usually occurs in the first weeks of glucocorticoid therapy, at daily doses of ≥40 mg of prednisone or equivalent; psychosis resolves over several days after glucocorticoids are decreased or stopped. Myelopathy is not rare and is often disabling; rapid immunosuppressive therapy starting with glucocorticoids is standard of care.

Vascular Occlusions

The prevalence of transient ischemic attacks, strokes, and myocardial infarctions is increased in patients with SLE. These vascular events are increased in, but not exclusive to, SLE patients with antibodies to phospholipids (aPL). Antiphospholipid antibodies are associated with hypercoagulability and acute thrombotic events, whereas chronic disease is associated with accelerated atherosclerosis **(Chap. 320)**. Ischemia in the brain can be caused by focal occlusion (either noninflammatory or associated with vasculitis) or by embolization from carotid artery plaque or from fibrinous vegetations of Libman-Sacks endocarditis. Appropriate tests for aPL (see below) and for sources of emboli should be ordered in such patients to estimate the need for, intensity of, and duration of anti-inflammatory and/or anticoagulant therapies. In SLE, myocardial infarctions are primarily manifestations of accelerated atherosclerosis. The increased risk for vascular events is seven- to tenfold overall, and higher in women <45 years old with SLE. Characteristics associated with increased risk for atherosclerosis include older age, hypertension, dyslipidemia, dysfunctional proinflammatory high-density lipoproteins, repeated high scores for disease activity, high cumulative or daily doses of glucocorticoids, and high levels of homocysteine. When it is most likely that an event results from clotting, long-term anticoagulation is the therapy of choice. Two processes can occur at once —vasculitis plus bland vascular occlusions—in which case it is appropriate to treat with anticoagulation plus immunosuppression. Statin therapies reduce levels of low-density lipoproteins (LDL) in SLE patients; reduction of cardiac events by statins has been shown in SLE patients with renal transplants but not in other SLE cohorts to date.

Pulmonary Manifestations

The most common pulmonary manifestation of SLE is pleuritis with or without pleural effusion. This manifestation, when mild, may respond to treatment with nonsteroidal anti-inflammatory drugs (NSAIDs); when more severe, patients require a brief course of glucocorticoid therapy. Pulmonary infiltrates also occur as a manifestation of active SLE and are difficult to distinguish from infection on imaging studies. Life-threatening pulmonary manifestations include interstitial inflammation leading to fibrosis, shrinking lung syndrome, and intra-alveolar hemorrhage; all of these probably require early aggressive immunosuppressive therapy as well as supportive care.

Cardiac Manifestations

Pericarditis is the most frequent cardiac manifestation; it usually responds to anti-inflammatory therapy and infrequently leads to tamponade. More serious cardiac manifestations are myocarditis and fibrinous endocarditis of Libman-Sacks. The endocardial involvement can lead to valvular insufficiencies, most commonly of the mitral or aortic valves, or to embolic events. It has not been proven that glucocorticoid or other immunosuppressive therapies lead to improvement of lupus myocarditis or endocarditis, but it is usual practice to administer a trial of high-dose steroids along with

appropriate supportive therapy for heart failure, arrhythmia, or embolic events. As discussed above, patients with SLE are at increased risk for myocardial infarction, usually due to accelerated atherosclerosis, which probably results from immune attack, chronic inflammation, and/or chronic oxidative damage to arteries.

Hematologic Manifestations

The most frequent hematologic manifestation of SLE is anemia, usually normochromic normocytic, reflecting chronic illness. Hemolysis can be rapid in onset and severe, requiring high-dose glucocorticoid therapy, which is effective in most patients. Leukopenia is also common and almost always consists of lymphopenia, not granulocytopenia; this rarely predisposes to infections and by itself usually does not require therapy. Thrombocytopenia may be a recurring problem. If platelet counts are >40,000/μL and abnormal bleeding is absent, therapy may not be required. High-dose glucocorticoid therapy (e.g., 1 mg/kg per day of prednisone or equivalent) is usually effective for the first few episodes of severe thrombocytopenia. Recurring or prolonged hemolytic anemia or thrombocytopenia, or disease requiring an unacceptably high dose of daily glucocorticoids, should be treated with an additional strategy (see "Treatment," below).

Gastrointestinal Manifestations

Nausea, sometimes with vomiting and diarrhea, can be manifestations of an SLE flare, as can diffuse abdominal pain probably caused by autoimmune peritonitis and/or intestinal vasculitis. Increases in serum aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are common when SLE is active. These manifestations usually improve promptly during systemic glucocorticoid therapy. Vasculitis involving the intestine may be life-threatening; perforations, ischemia, bleeding, and sepsis are frequent complications. Aggressive immunosuppressive therapy with high-dose glucocorticoids is recommended for short-term control; evidence of recurrence is an indication for additional therapies.

Ocular Manifestations

Sicca syndrome (Sjögren's syndrome; **Chap. 324)** and nonspecific conjunctivitis are common in SLE and rarely threaten vision. In contrast, retinal vasculitis and optic neuritis are serious manifestations: blindness can develop over days to weeks. Aggressive immunosuppression is recommended, although there are no controlled trials to prove effectiveness. Complications of glucocorticoid therapy include cataracts (common) and glaucoma.

LABORATORY TESTS

Laboratory tests serve (1) to establish or rule out the diagnosis; (2) to follow the course of disease, particularly to suggest that a flare is occurring or organ damage is developing; and (3) to identify adverse effects of therapies.

Tests for Autoantibodies (Tables 319-1 and 319-3)

Diagnostically, the most important autoantibodies to detect are ANA as the test is positive in >95% of patients, usually at the onset of symptoms. A few patients develop ANA within 1 year of symptom onset; repeated testing may thus be useful. ANA-negative lupus exists but is rare in adults and is usually associated with other autoantibodies (anti-Ro or anti-DNA). High-titer IgG antibodies to double-stranded DNA (dsDNA) (but not to single-stranded DNA) are specific for SLE. There is no international standardized test for ANA; variability between different service laboratories is high. Enzyme-linked immunosorbent assays (ELISA) and immunofluorescent reactions of sera with the dsDNA in the flagellate *Crithidia luciliae* have ~60% sensitivity for SLE; identification of high-avidity anti-dsDNA in the Farr assay is not as sensitive but may correlate better with risk for nephritis. Titers of anti-dsDNA vary over time. In some patients, increases in quantities of anti-dsDNA herald a flare, particularly of nephritis or vasculitis, especially when associated with declining levels of C3 or C4 complement. Antibodies to Sm are also specific for SLE and assist in diagnosis; anti-Sm antibodies do not usually correlate with disease activity or clinical manifestations. aPL are not specific for SLE, but their presence fulfills one classification criterion, and they identify patients at increased risk for venous or arterial clotting, thrombocytopenia, and fetal loss. There are two widely accepted tests that measure different antibodies (anticardiolipin and the lupus anticoagulant): (1) ELISA for anticardiolipin (internationally standardized with good reproducibility) and (2) a sensitive phospholipid-based activated prothrombin time such as the dilute Russell venom viper test. Some centers also recommend measurement of antibodies to β_2 glycoprotein 1, a serum protein cofactor that is the target of most antibodies to cardiolipin and some lupus anticoagulants. The higher the titers of IgG anticardiolipin (>40 IU is considered high), and the greater the number of different aPL that are detected, the greater is the risk for a clinical episode of clotting. Quantities of aPL may vary markedly over time; repeated testing is justified if clinical manifestations of the antiphospholipid antibody syndrome (APS) appear **(Chap. 320)**. To classify a patient as having APS, with or without SLE, by international criteria requires the presence of ≥1 clotting episode and/or repeated fetal losses plus at least two positive tests for aPL, at least 12 weeks apart; however, many patients with anti-phospholipid syndrome do not meet these stringent criteria, which are intended for inclusion of patients into studies.

An additional autoantibody test with predictive value (not used for diagnosis) detects anti-Ro, which indicates increased risk for neonatal lupus, sicca syndrome, and SCLE. Women with child-bearing potential and SLE should be screened for aPL and anti-Ro.

Standard Tests for Diagnosis

Screening tests for complete blood count, platelet count, and urinalysis may detect abnormalities that contribute to the diagnosis and influence management decisions.

Tests for Following Disease Course

It is useful to follow tests that indicate the status of organ involvement known to be present during SLE flares. These might include urinalysis for hematuria and proteinuria, hemoglobin levels, platelet counts, and serum levels of creatinine or albumin. There is great interest in identification of additional markers of disease activity. Candidates include levels of anti-DNA antibodies, several components of complement (C3 is most widely available), activated complement products (including those that bind to the C4d receptor on erythrocytes), IFN-inducible gene expression in peripheral blood cells, soluble IL-2 levels, and urinary levels of TNF-like weak inducer of apoptosis (TWEAK), neutrophil gelatinase-associated lipocalin (NGAL), or monocyte chemotactic protein 1 (MCP-1). None is uniformly agreed upon as a reliable indicator of flare or of response to therapeutic interventions. The physician should determine for each patient whether certain laboratory test changes predict flare. If so, altering therapy in response to these changes may be advisable (30 mg of prednisone daily for 2 weeks has been shown to prevent flares in patients with rising anti-DNA plus falling complement). In addition, given the increased prevalence of atherosclerosis in SLE, it is advisable to follow the recommendations of the National Cholesterol Education Program for testing and treatment, including scoring of SLE as an independent risk factor, similar to diabetes mellitus.

Treatment: Systemic Lupus Erythematosus

There is no cure for SLE, and complete sustained remissions are rare. Therefore, the physician should plan to induce improvement of acute flares and then maintain improvements with strategies that suppress symptoms to an acceptable level and prevent organ damage. Usually patients will endure some adverse effects of medications. Therapeutic choices depend on (1) whether disease manifestations are life-threatening or likely to cause organ damage, justifying aggressive therapies; (2) whether manifestations are potentially reversible; and (3) the best approaches to preventing complications of disease and its treatments. Therapies, doses, and adverse effects are listed in **Table 319-5**.

CONSERVATIVE THERAPIES FOR MANAGEMENT OF NON-LIFE-THREATENING DISEASE

Among patients with fatigue, pain, and autoantibodies of SLE, but without major organ involvement, management can be directed to suppression of symptoms. Analgesics and antimalarials are mainstays. NSAIDs are useful analgesics/anti-inflammatories, particularly for arthritis/arthralgias. However, two major issues currently indicate caution in using NSAIDs. First, SLE patients compared with the general population are at increased risk for NSAID-induced aseptic meningitis, elevated serum transaminases, hypertension, and renal dysfunction. Second, all NSAIDs, particularly those that inhibit cyclooxygenase-2 specifically, may increase risk for myocardial infarction. Acetaminophen to control pain may be a good strategy, but NSAIDs are more effective in some patients. The relative hazards of NSAIDs compared with low-dose glucocorticoid therapy have not been established. Antimalarials (hydroxychloroquine, chloroquine, and quinacrine) often reduce dermatitis, arthritis, and fatigue. A randomized, placebo-controlled, prospective trial has shown that withdrawal of hydroxychloroquine results in increased numbers of disease flares. Hydroxychloroquine reduces accrual of tissue damage over time. Because of potential retinal toxicity, patients receiving antimalarials should undergo ophthalmologic examinations annually. A placebo-controlled prospective trial suggests that administration of dehydroepiandrosterone may reduce disease activity. If quality of life is inadequate in spite of these conservative measures, treatment with low doses of systemic glucocorticoids may be necessary. Dermatitis should be managed with topical sunscreens, antimalarials, and topical glucocorticoids and/or tacrolimus. Since recent data show that mycophenolate mofetil, and belimumab (added to background therapies of glucocorticoids-plus-antimalarial-plus immunosuppressive) reduce disease activity in nonrenal manifestations of SLE, it is reasonable to consider these interventions in patients with persistent disease activity despite standard therapies. Azathioprine or methotrexate may also be considered for such patients (Table 319-5).

LIFE-THREATENING SLE: PROLIFERATIVE FORMS OF LUPUS NEPHRITIS

The mainstay of treatment for any inflammatory life-threatening or organ-threatening manifestations of SLE is systemic glucocorticoids (0.5–1 mg/kg per day PO or 1000 mg of methylprednisolone sodium succinate IV daily for 3 days followed by 0.5–1 mg/kg of daily prednisone or equivalent). Evidence that glucocorticoid therapy is lifesaving comes from retrospective studies from the predialysis era; survival is significantly better in people with DPGN treated with high-dose daily glucocorticoids (40–60 mg of prednisone daily for 4–6 months) versus lower doses. Currently, high doses are recommended for much shorter periods; recent trials of interventions for severe SLE employ 4–6 weeks of 0.5 to 1 mg/kg/day of prednisone or equivalent. Thereafter, doses are tapered as rapidly as the clinical

situation permits, usually to a maintenance dose varying from 5–10 mg of prednisone or equivalent per day or from 10–20 mg every other day. Most patients with an episode of severe lupus require many years of maintenance therapy with low-dose glucocorticoids, which can be increased to prevent or treat disease flares. Frequent attempts to gradually reduce the glucocorticoid requirement are recommended since virtually everyone develops important adverse effects **(Table 319-5)**. Prospective controlled trials in active lupus nephritis show that induction of improvement by administration of high doses of glucocorticoids (1000 mg of ethylprednisolone daily for 3 days) by IV routes compared with daily oral routes shortens the time to maximal improvement by a few weeks but ultimately improvements are similar. It has become standard practice to initiate therapy for active, potentially life-threatening SLE with high-dose IV glucocorticoid pulses, based on studies in lupus nephritis. This approach must be tempered by safety considerations, such as the presence of conditions adversely affected by glucocorticoids (infection, hyperglycemia, hypertension, osteoporosis, etc.).

Cytotoxic/immunosuppressive agents added to glucocorticoids are recommended to treat serious SLE. Almost all prospective controlled trials in SLE involving cytotoxic agents have been conducted in combination with glucocorticoids in patients with lupus nephritis. Therefore, the following recommendations apply to treatment of nephritis. Either cyclophosphamide (an alkylating agent) or mycophenolate mofetil (a relatively lymphocyte-specific inhibitor of inosine monophosphatase and therefore of purine synthesis) is an acceptable choice for induction of improvement in severely ill patients; azathioprine (a purine analogue and cycle-specific antimetabolite) is probably less effective but may be used if the other immunosuppressives are not tolerated or not available. In patients whose renal biopsies show ISN grade III or IV disease, early treatment with combinations of glucocorticoids and cyclophosphamide reduces progression to ESRD and improves survival; this difference can be seen after approximately 5 years of therapy. Shorter-term studies with glucocorticoids plus mycophenolate mofetil (prospective randomized trials of 6 months) show that this regimen is similar to cyclophosphamide in inducing improvement. Comparisons are complicated by effects of race, since higher proportions of blacks (and other non-Asian, non-white races) respond to mycophenolate than to cyclophosphamide, whereas similar proportions of whites and Asians respond to each drug. Regarding toxicity, diarrhea is more common with mycophenolate while herpetic infections, amenorrhea, and leukopenia are more common with cyclophosphamide; rates of severe infections and death are similar in some studies, although mycophenolate is less toxic than cyclophosphamide in others. Therapeutic responses to cyclophosphamide and mycophenolate begin 3–16 weeks after treatment is initiated, whereas glucocorticoid responses may begin within 24 h. For maintenance therapy, mycophenolate may be better than azathioprine in preventing flares and progression of lupus nephritis; either drug is acceptable and both are safer than cyclophosphamide. If cyclophosphamide is used for induction therapy, the recommended "National Institutes of Health (NIH)" dose (based on clinical trials at that institution) is 500–750 mg/m^2 intravenously, monthly for 6 months, followed by maintenance with daily oral mycophenolate or azathioprine. The incidence of ovarian failure, a common effect of cyclophosphamide therapy, can be reduced by treatment with a gonadotropin-releasing hormone agonist (e.g., Lupron 3.75 mg IM) prior to each monthly cyclophosphamide dose. Since cyclophosphamide has many adverse effects and is generally disliked by patients, alternative approaches using lower doses have been tested. European studies have shown that IV cyclophosphamide at doses of 500 mg every 2 weeks for six doses ("low dose") is as effective as the recommended higher dose given for a longer duration in the NIH regimen ("high dose"). All patients were maintained on azathioprine after the course of cyclophosphamide was completed. Ten-year follow-up has shown no differences in the high-dose and low-dose groups (death or ESRD in 9–20% in each group). The majority of the European patients were white; it is not clear that the data apply to U.S. populations. Patients with high serum creatinine levels [e.g., ≥265 μmol/L (≥3 mg/dL)] many months in duration and high chronicity scores on renal biopsy are not likely to respond to immunosuppression. In general, it may be better to induce improvement in a black or Hispanic patient with proliferative glomerulonephritis with mycophenolate (2–3 g daily) rather than with cyclophosphamide, with the option to switch if no evidence of response is detectable after 3–6 months of treatment. For whites and Asians, induction with either mycophenolate or cyclophosphamide is acceptable. Cyclophosphamide may be discontinued when it is clear that a patient is improving; the number of SLE flares is reduced by maintenance therapy with mycophenolate (1.5–2 g daily) or azathioprine (2 mg/kg/d). Both cyclophosphamide and mycophenolate are potentially teratogenic; patients should be off either medication for at least 3 months before attempting to conceive. If azathioprine is used either for induction or maintenance therapy, patients may be prescreened for homozygous deficiency of the TMPT enzyme (which is required to metabolize the 6-mercaptopurine product of azathioprine) since they are at higher risk for bone marrow suppression.

Good improvement occurs in ~80% of lupus nephritis patients receiving either cyclophosphamide or mycophenolate at 1–2 years of follow-up. However, at least 50% of these individuals have flares of nephritis over the next 5 years, and retreatment is required; such individuals are more likely to progress to ESRD. Long-term outcome of lupus nephritis to most interventions is better in whites than in blacks. Chlorambucil is an alkylating agent that can be substituted for cyclophosphamide; the risk of irreversible bone marrow suppression may be greater with this agent. Methotrexate (a folinic acid antagonist) may have a role in the treatment of arthritis and dermatitis, but probably not in nephritis or

other life-threatening disease. Small controlled trials (in Asia) of leflunomide, a relatively lymphocyte-specific pyrimidine antagonist licensed for use in rheumatoid arthritis, have suggested it can suppress disease activity in some SLE patients. Cyclosporine and tacrolimus, which inhibit production of IL-2 and T lymphocyte functions, are used by some clinicians particularly for membranous lupus nephritis. Since they have potential nephrotoxicity, but little bone marrow toxicity, the author uses them for periods of only a few months in patients with steroid-resistant cytopenias of SLE, or in steroid-resistant patients who have developed bone marrow suppression from standard cytotoxic agents.

Use of biologicals directed against B cells for active SLE is under intense study. Use of anti-CD20 (Rituximab), particularly in those patients with SLE who are resistant to the more standard combination therapies discussed above, is controversial. Several open trials have shown efficacy in a majority of such patients—both for nephritis and for extrarenal lupus. However, recent prospective placebo-controlled randomized trials did not show a difference between anti-CD20 and placebo when added to standard combination therapies. In contrast, recent trials of anti-BLyS (belimumab, directed against the ligand of the BLyS/BAFF receptor on B cells that promotes B cell survival and differentiation to plasmablasts) showed a small, but statistically significant, better suppression of disease activity in comparison to placebo, when added to standard combhnation therapies. The US FDA has approved belimumab for treatment of SLE: it has not been studied in active nephritis or central nervous system lupus.

It is important to note that there are few if any randomized, controlled, prospective studies of any agents in life-threatening SLE that do not include nephritis. Therefore, use of glucocorticoids plus byclophosphamide or mycophenolate in other life-threatening conditions is based on studies in nephritis.

SPECIAL CONDITIONS IN SLE THAT MAY REQUIRE ADDITIONAL OR DIFFERENT THERAPIES

Crescentic Lupus Nephritis

The presence of cellular or fibrotic crescents in glomeruli with proliferative glomerulonephritis (INS-IVG)] indicates a worse prognosis than in patients without this feature. There are few large prospective controlled trials showing efficacy of cyclophosphamide, mycophenolate, or cyclosporine in such cases. Most authorities currently recommend that cyclophosphamide in the NIH-recommended high dose or high doses of mycophenolate are the induction therapies of choice, in addition to glucocorticoids.

Membranous Lupus Nephritis

Most SLE patients with membranous (INS-V) nephritis also have proliferative changes and should be treated for proliferative disease; however, some have pure membranous changes. Treatment for this group is less well defined; recent prospective controlled trials suggest that alternate-day glucocorticoids plus cyclophosphamide or mycophenolate or cyclosporine are all effective in the majority of patients in reducing proteinuria; whether they preserve renal function over the lonf term is more controversial.

Pregnancy and Lupus

Fertility rates for men and women with SLE are probably normal. However, rate of fetal loss is increased (approximately two- to threefold) in women with SLE. Fetal demise is higher in mothers with high disease activity, antiphospholipid antibodies, and/or active nephritis. Suppression of disease activity can be achieved by administration of systemic glucocorticoids. A placental enzyme, 11-ß-dehydrogenase 2, deactivates glucocorticoids; it is more effective in deactivating prednisone and prednisolone than the fluorinated glucocorticoids dexamethasone and betamethasone. Glucocorticoids are listed by the FDA as pregnancy category A (no evidence of teratogenicity in human studies); cyclosporine, tacrolimus, and rituximab are listed as category C (may be teratogenic in animals but no good evidence in humans); azathioprine, hydroxychloroquine, mycophenolate mofetil, and cyclophosphamide are category D (there is evidence of teratogenicity in humans. but benefits might outweigh risks in certain situations); and methotrexate is category X (risks outweigh benefits). Therefore, active SLE in pregnant women should be controlled with prednisone/prednisolone at the lowest effective doses for the shortest time required. Adverse effects of prenatal glucocorticoid exposure (primarily betamethasone) on offspring may include low birth weight, developmental abnormalities in the CNS, and predilection toward adult metabolic syndrome. It is likely that each of these glucocorticoids and immunosuppressive medications get into breast milk, at least in low levels; patients should consider not breast-feeding if they need therapy for SLE. In SLE patients with aPL (on at least two occasions) and prior fetal losses, treatment with heparin (usually low-molecular-weight) plus low-dose aspirin has been shown in prospective controlled trials to increase significantly the proportion of live births; however, a recent prospective trial showed no differences in fetal outcomes in women taking aspirin compared to those on aspirin plus low-molecular-weight heparin. An additional potential problem for the fetus is the presence of antibodies to Ro, sometimes associated with neonatal lupus consisting of rash and congenital heart block. The latter can be life-threatening; therefore, the presence of anti-Ro requires vigilant monitoring of fetal heart rates with prompt intervention (delivery if possible) if distress occurs. To date, treatments of mother to reverse established heart block in the fetus, newborn, or infant (other than insertion of a pacemaker) have not

been successful. Women with SLE usually tolerate pregnancy without disease flares. However, a small proportion develops severe flares requiring aggressive glucocorticoid therapy or early delivery. Poor maternal outcomes are highest in women with active nephritis or irreversible organ damage in kidneys, brain, or heart.

Lupus and Antiphospholipid Antibody Syndrome (Chap. 20)

Patients with SLE who have venous or arterial clotting, and/or repeated fetal losses, and at least two positive tests for aPL have APS and should be managed with long-term anticoagulation. A target international normalized ratio (INR) of 2–2.5 is recommended for patients with one episode of venous clotting; an INR of 3–3.5 is recommended for patients with recurring clots or arterial clotting, particularly in the central nervous system. Recommendations are based on both retrospective and prospective studies of posttreatment clotting events and adverse effects from anticoagulation.

Microvascular Thrombotic Crisis (Thrombotic Thrombocytopenic Purpura, Hemolytic-Uremic Syndrome)

This syndrome of hemolysis, thrombocytopenia, and microvascular thrombosis in kidneys, brain, and other tissues carries a high mortality rate and occurs most commonly in young individuals with lupus nephritis. The most useful laboratory tests are identification of schistocytes on peripheral blood smears, elevated serum levels of lactate dehydrogenase, and antibodies to ADAMS13. Plasma exchange or extensive plasmapheresis is usually life-saving; most authorities recommend concomitant glucocorticoid therapy; there is no evidence that cytotoxic drugs are effective.

Lupus Dermatitis

Patients with any form of lupus dermatitis should minimize exposure to ultraviolet light, employing appropriate clothing and sunscreens with a sun protection factor of at least 15. Topical glucocorticoids and antimalarials (such as hydroxychloroquine) are effective in reducing lesion severity in most patients and are relatively safe. Systemic treatment with retinoic acid is a useful strategy in patients with inadequate improvement on topical glucocorticoids and antimalarials; adverse effects are potentially severe (particularly fetal abnormalities), and there are stringent reporting requirements for its use in the United States. Extensive, pruritic, bullous, or ulcerating dermatitides usually improve promptly after institution of systemic glucocorticoids; tapering may be accompanied by flare of lesions, thus necessitating use of a second medication such as hydroxychloroquine, retinoids, or cytotoxic medications such as methotrexate or azathioprine. In therapy-resistant lupus dermatitis there are reports of success with topical tacrolimus (caution must be exerted because of the possible increased risk for malignancies) or with systemic dapsone or thalidomide (the extreme danger of fetal deformities from thalidomide requires permission from and supervision by the supplier).

PREVENTIVE THERAPIES

Prevention of complications of SLE and its therapy include providing appropriate vaccinations (the administration of influenza and pneumococcal vaccines has been studied in patients with SLE; flare rates are similar to those receiving placebo) and suppressing recurrent urinary tract infections. In addition, strategies to prevent osteoporosis should be initiated in most patients likely to require long-term glucocorticoid therapy and/or with other predisposing factors. Control of hypertension and appropriate prevention strategies for atherosclerosis, including monitoring and treatment of dyslipidemias, management of hyperglycemia, and obesity, are recommended.

EXPERIMENTAL THERAPIES

Studies of highly targeted experimental therapies for SLE are in progress. They include targeting (1) activated B lymphocytes with anti-BLyS, or TACI-Ig; (2) inhibition of IFNα; (3) inhibition of B/T cell second signal co-activation with CTLA-Ig; and (4) inhibition of innate immune activation via TLR7 or TLR7 and 9, and induction of regulatory T cells with peptides from immunoglobulins or autoantigens. A few studies have employed vigorous untargeted immunosuppression with high-dose cyclophosphamide plus anti–T cell strategies, with rescue by transplantation of autologous hematopoietic stem cells for the treatment of severe and refractory SLE. One U.S. report showed an estimated mortality rate over 5 years of 15% and sustained remission in 50%. It is hoped that the next edition of this text will recommend more effective and less toxic approaches to treatment of SLE based on some of these strategies.

PATIENT OUTCOMES, PROGNOSIS, AND SURVIVAL

Survival in patients with SLE in the United States, Canada, Europe, and China is approximately 95% at 5 years, 90% at 10 years, and 78% at 20 years. In the United States, African Americans and Hispanic Americans with a mestizo heritage have a worse prognosis than whites, whereas Africans in Africa and Hispanic Americans with a Puerto Rican origin do not. The relative importance of gene mixtures and environmental differences accounting for ethnic differences is not known. Poor prognosis ($\sim$50% mortality in 10 years) in most series is associated with (at the time of diagnosis) high serum creatinine levels [>124 μmol/L (>1.4 mg/dL)], hypertension, nephrotic syndrome (24-h urine protein excretion >2.6 g), anemia [hemoglobin <124 g/L (<12.4 g/dL)], hypoalbuminemia, hypocomplementemia, aPL, male sex, and ethnicity (African American, Hispanic with mestizo heritage). Data regarding outcomes in SLE

patients with renal transplants show mixed results: some series have a twofold increase in graft rejection compared to patients with other causes of ESRD, whereas others show no differences. Overall patient survival is comparable (85% at 2 years). Lupus nephritis occurs in approximately 10% of transplanted kidneys. Disability in patients with SLE is common due primarily to chronic fatigue, arthritis, and pain, as well as renal disease. As many as 25% of patients may experience remissions, sometimes for a few years, but these are rarely permanent. The leading causes of death in the first decade of disease are systemic disease activity, renal failure, and infections; subsequently, thromboembolic events become increasingly frequent causes of mortality.

DRUG-INDUCED LUPUS

This is a syndrome of positive ANA associated with symptoms such as fever, malaise, arthritis or intense arthralgias/myalgias, serositis, and/or rash. The syndrome appears during therapy with certain medications and biologic agents, occurs predominantly in whites, has less female predilection than SLE, rarely involves kidneys or brain, is rarely associated with anti-dsDNA, is commonly associated with antibodies to histones, and usually resolves over several weeks after discontinuation of the offending medication. The list of substances that can induce lupus-like disease is long. Among the most frequent are the antiarrhythmics procainamide, disopyramide, and propafenone; the antihypertensive hydralazine; several angiotensin-converting enzyme inhibitors and beta blockers; the antithyroid propylthiouracil; the antipsychotics chlorpromazine and lithium; the anticonvulsants carbamazepine and phenytoin; the antibiotics isoniazid, minocycline, and macrodantin; the antirheumatic sulfasalazine; the diuretic hydrochlorothiazide; the antihyperlipidemics lovastatin and simvastatin; and interferons and TNF inhibitors. ANA usually appears before symptoms; however, many of the medications mentioned above induce ANA in patients who never develop symptoms of drug-induced lupus. It is appropriate to test for ANA at the first hint of relevant symptoms and to use test results to help decide whether to withdraw the suspect agent.

FURTHER READINGS

Falk RJ et al: Therapy of diffuse and proliferative lupus nephritis. UpToDate, October 2010

Helmick CG et al: Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part 1. Arthritis Rheum 58:15, 2008[PMID: 18163481] [Full Text]

Kasitanon N et al: Predictors of survival in systemic lupus erythematosus. Medicine (Baltimore) 85:147, 2006[PMID: 16721257] [Full Text]

Rahman A, Isenberg D: Systemic lupus erythematosus. New Engl J Med 358:929, 2008[PMID: 18305268] [Full Text]

Schur PH, Wallace DJ: Overview of the therapy and prognosis of systemic lupus erythematosus in adults. UpToDate, October 2010

Wallace DJ, Hahn BH (eds): *Dubois' Lupus Erythematosus*, 8th ed. Philadelphia, Lippincott, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION

Antiphospholipid antibody syndrome (APS) is an autoantibody-mediated acquired thrombophilia characterized by recurrent arterial or venous thrombosis and/or pregnancy morbidity in the presence of autoantibodies against phospholipid (PL)-binding plasma proteins, mainly a plasma apolipoprotein known as β2 glycoprotein I (β2GPI) and prothrombin **(Table 320-1)**. Another group of antibodies termed *lupus anticoagulant* (LA) prolong clotting times in vitro; this prolongation is not corrected by adding normal plasma to the detection system. APS may occur alone (primary), or in association with any other autoimmune disease (secondary). Catastrophic APS (CAPS) is defined as a rapidly progressive thromboembolic disease involving simultaneously three or more organs, organ systems, or tissues leading to corresponding functional defects.

Table 320-1 Classification and Nomenclature of Antiphospholipid Antibodies

• Antibodies against cardiolipin (aCL), a negatively charged phospholipid, detected by enzyme-linked immunosorbent assay (ELISA)
• Antibodies against β2GPI, (anti-β2GPI) detected by ELISA in the absence of PL.
• LA detected by clotting assays. LA constitutes a heterogeneous group of antibodies directed also against PL binding proteins, mainly β2GPI and prothrombin. LA antibodies induce elongation in vitro of the following clotting times:
Activated partial thromboplastin time (aPTT), kaolin clotting time (KCT), dilute Russel viper venom test (dRVVT)
• Antibodies against phospholipids/cholesterol complexes detected as biologic false-positive serologic test for syphilis (BFP-STS) and Venereal Disease Research Laboratory Test (VDRL)

EPIDEMIOLOGY

Anti-PL (aPL)-binding plasma protein antibodies occur in 1–5% of general population. Their prevalence increases with age; however, it is questionable whether they induce thrombotic events in elderly individuals. One-third of patients with systemic lupus erythematosus (SLE) **(Chap. 319)** possess these antibodies while their prevalence in other autoimmune connective tissue disorders such as systemic sclerosis (scleroderma), Sjögren's syndrome, dermatomyositis, rheumatoid arthritis, and early undifferentiated connective tissue disease, ranges from 6% to 15%. One-third of aPL positive individuals experience thrombotic events or pregnancy morbidity.

PATHOGENESIS

The trigger for the induction of antibodies to PL-binding proteins is not known. Preceding infections, however, have been proposed as the initiating event. These antibodies are pathogenic since anti-β2GPI/β2GPI complexes inactivate natural anticoagulants such as protein C, activate cells involved in the coagulation cascade to a prothrombotic phenotype, activate complement, and inhibit syncytium-trophoblast differentiation. Activated protein C (APC) binds the pro-coagulant factors Va and VIIIa and inactivates them. Anti-β2GPI/β2GPI complexes inhibit the APC activity in vivo by competing with the components of the APC/Va/VIIIa complexes for binding to a number of PL-binding sites, or by disrupting these complexes. Domain V of β2GPI can interact with apolipoprotein E receptor 2' (apoER2') and/or with the GPIbα subunit of the GPIb/IX/V receptor of platelets. Furthermore platelet factor 4 (PF4) tetramers dimerize β2GPI and the resulting complexes are recognized by anti-β2GPI antibodies, eventually activating the p38 mitogen-activated protein (p38 MAP) kinase phosphorylation, and leading to thromboxane B_2 (TXB_2) production in vitro. In fact, increased levels of 11-dehydro-TXB_2, have been found in the urine of patients with APS. Anti-β2GPI antibodies activate nuclear factor kappa B (NF-κB) in monocytes and endothelial cells by interacting with surface receptors not yet identified, leading to the secretion of pro-inflammatory cytokines, such as interleukins -1, -6, and -8; the expression of adhesion molecules such as intercellular adhesion molecule (ICAM-1), vascular cell adhesion molecule 1 (VCAM-1), and E-selectin; inhibition of cell-surface plasminogen activation; and the expressions of tissue factor, changing the phenotype of these cells to a prothrombotic form. As shown in mouse models, anti-β2GPI antibodies induce fetal injury through complement activation, since C4-deficient mice were protected from fetal injury.

CLINICAL MANIFESTATIONS AND LABORATORY FINDINGS

Clinical manifestations represent mainly a direct or indirect expression of venous or arterial thrombosis and/or pregnancy morbidity **(Table 320-2)**. Clinical features associated with venous thrombosis are superficial and deep vein thrombosis, cerebral venous thrombosis, signs and symptoms of intracranial hypertension, retinal vein thrombosis, pulmonary emboli, pulmonary arterial hypertension, and Budd-Chiari syndrome. Livedo reticularis consists of a mottled reticular vascular pattern that appears as a lace-like, purplish discoloration of the skin. It is probably caused by swelling of the venules owing to obstruction of capillaries by thrombi. This clinical manifestation correlates with vascular lesions such as those in the central nervous system as well as aseptic bone necrosis. Arterial thrombosis is manifested as migraines, cognitive dysfunction, transient ischemic attacks, stroke, myocardial infarction, arterial thrombosis of upper and lower extremities, ischemic leg ulcers, digital gangrene, avascular necrosis of bone, retinal artery occlusion leading to painless monocular loss of vision (amaurosis fugax), renal artery stenosis, and glomerular lesions, as well as infarcts of spleen, pancreas, and adrenals. Libman-Sacks endocarditis consists of very small vegetations, histologically characterized by organized platelet-fibrin microthrombi surrounded by growing fibroblasts and macrophages. Glomerular lesions are manifested with hypertension, mildly elevated serum creatinine levels, proteinuria, and mild hematuria. Histologically, these lesions are characterized in an acute phase by thrombotic microangiopathy involving glomerular capillaries, and in a chronic phase with fibrous intima hyperplasia, fibrous and/or fibrocellular occlusions of arterioles and focal cortical atrophy (Table 320-2). Premature atherosclerosis has been recognized as a rare feature of APS. Coombs-positive hemolytic anemia and thrombocytopenia are laboratory findings associated with APS. Discontinuation of therapy, major surgery, infection, and trauma may trigger CAPS.

Table 320-2 Clinical Features of Antiphospholipid Antibody Syndrome

Manifestation	%
Venous Thrombosis and Related Consequences	
Deep vein thrombosis	39
Livedo reticularis	24
Pulmonary embolism	14
Superficial thrombophlebitis	12
Thrombosis in various other sites	11
Arterial Thrombosis and Related Consequences	
Stroke	20
Cardiac valve thickening/dysfunction and/or Libman-Sacks vegetations	14
Transient ischemic attack	11
Myocardial ischemia (infarction or angina) and coronary bypass thrombosis	10
Leg ulcers and/or digital gangrene	9
Arterial thrombosis in the extremities	7
Retinal artery thrombosis/amaurosis fugax	7
Ischemia of visceral organs or avascular necrosis of bone	6
Multi-infarct dementia	3
Neurologic Manifestations of Uncertain Etiology	
Migraine	20
Epilepsy	7
Chorea	1
Cerebellar ataxia	1
Transverse myelopathy	0.5
Renal Manifestations Due to Various Reasons (Renal Artery/Renal Vein/Glomerular	3

Thrombosis, Fibrous Intima Hyperplasia)	
Osteoarticular Manifestations	
Arthralgia	39
Arthritis	27
Obstetric Manifestations (Referred to the Number of Pregnancies)	
Preeclampsia	10
Eclampsia	4
Fetal Manifestations (Referred to the Number of Pregnancies)	
Early fetal loss (<10 weeks)	35
Late fetal loss (≥10 weeks)	17
Premature birth among the live births	11
Hematologic Manifestations	
Thrombocytopenia	30
Autoimmune hemolytic anemia	10

Source: Adapted from R Cervera et al.

DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS

The diagnosis of APS should be seriously considered in cases of thrombosis, cerebral vascular accidents in individuals younger than 55 years of age, or pregnancy morbidity in the presence of livedo reticularis or thrombocytopenia. In these cases aPL antibodies should be measured. The presence of at least one clinical and one laboratory criterion ensures the diagnosis even in the presence of other causes of thrombophilia. Clinical criteria include: (1) vascular thrombosis defined as one or more clinical episodes of arterial, venous, or small vessel thrombosis in any tissue or organ; and (2) pregnancy morbidity, defined as (a) one or more unexplained deaths of a morphologically normal fetus at or beyond the tenth week of gestation, or (b) one or more premature births of a morphologically normal neonate before the thirty-fourth week of gestation because of eclampsia, severe preeclampsia, or placental insufficiency; or (c) three or more unexplained consecutive spontaneous abortions before the 10th week of gestation. Laboratory criteria include (1) LA, (2) anticardiolipin (aCL) and/or (3) anti-β2GPI antibodies, at intermediate or high titers on two occasions, 12 weeks apart.

Differential diagnosis is based on the exclusion of other inherited or acquired causes of thrombophilia **(Chap. 116)**, Coombs positive hemolytic anemia **(Chap. 106)**, and thrombocytopenia **(Chap. 115)**. Livedo reticularis with or without a painful ulceration on the lower extremities also may be a manifestation of disorders affecting (1) the vascular wall such as polyarteritis nodosa, SLE, cryoglobulinemia, and lymphomas; or (2) the vascular lumen, such as myeloproliferative disorders, atherosclerosis, hypercholesterolemia, or other causes of thrombophilia.

Treatment: Antiphospholipid Antibody Syndrome

After the first thrombotic event, APS patients should be placed on warfarin for life aiming to achieve an international normalized ratio (INR) ranging from 2.5 to 3.5, alone or in combination with 80 mg of aspirin daily. Pregnancy morbidity is prevented by a combination of heparin with aspirin 80 mg daily. Intravenous immunoglobulin (IVIg) 400 mg/kg qd for 5 days may also prevent abortions, while glucocorticoids are ineffective. Evidence-based treatment of patients with aPL in the absence of any clinical event is not available; however, aspirin 80 mg daily protects patients with SLE positive for aPL antibodies from developing thrombotic events.

Some patients with APS and patients with CAPS have recurrent thrombotic events despite appropriate anticoagulation. In these cases IVIg 400 mg/kg qd for 5 days or anti-CD20 monoclonal antibody 375 mg/m^2 per week for 4 weeks may be of benefit. Patients with CAPS, who are treated in the intensive care unit, are unable to receive warfarin; in this situation therapeutic doses of low-molecular-weight heparin should be administered. In cases of heparin- induced thrombocytopenia and thrombosis syndrome, inhibitors of phospholipid-bound activated factor X (FXa), such as fondaparinux 7.5 mg SC daily or rivaroxaban 10 mg PO daily are effective. The above drugs are administered by fixed doses and do not require close monitoring; their safety during the first trimester of pregnancy has not been clearly established.

FURTHER READINGS

Cervera R et al: Antiphospholipid syndrome: Clinical and immunologic manifestations and patterns of disease expression in a cohort of 1,000 patients. Arthritis Rheum 46:1019, 2002[PMID: 11953980] [Full Text]

Crowther MA et al: A comparison of two intensities of warfarin for the prevention of recurrent thrombosis in patients with the antiphospholipid antibody syndrome. N Engl J Med 349:1133, 2003[PMID: 13679527] [Full Text]

George D, Erkan D: Antiphospholipid syndrome. Prog Cardiovasc Dis 52:115, 2009[PMID: 19732604] [Full Text]

Miyakis S et al: International consensus statement on an update of the classification criteria for definite antiphospholipid syndrome (APS). J Thromb Haemost 295:4, 2006

Ruiz-Irastorza G et al: Antiphospholipid syndrome. Lancet 376:1498, 2010[PMID: 20822807] [Full Text]

Sikara MP et al: ß2 glycoprotein I (ß2GPI) binds platelet factor 4 (PF4): Implications for the pathogenesis of Antiphospholipid syndrome. Blood 115:713, 2010[PMID: 19805618] [Full Text]

Tektonidou M et al: Risk factors for thrombosis and primary thrombosis prevention in patients with systemic lupus erythematosus with or without antiphospholipid antibodies. Arthritis Rheum 61:29, 2009[PMID: 19116963] [Full Text]

Vlachoyiannopoulos PG et al: Antiphospholipid antibodies: Laboratory and pathogenetic aspects. Crit Rev Clin Lab Sci 271:44, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory disease of unknown etiology marked by a symmetric, peripheral polyarthritis. It is the most common form of chronic inflammatory arthritis and often results in joint damage and physical disability. Because it is a systemic disease, RA may result in a variety of extraarticular manifestations, including fatigue, subcutaneous nodules, lung involvement, pericarditis, peripheral neuropathy, vasculitis, and hematologic abnormalities.

Insights gained by a wealth of basic and clinical research over the past two decades have revolutionized the contemporary paradigms for the diagnosis and management of RA. Serum antibodies to cyclic citrullinated peptides (anti-CCPs) are now recognized to be a valuable biomarker of diagnostic and prognostic significance. Advances in ultrasound and magnetic resonance imaging have improved our ability to detect joint inflammation and destruction in RA. The science of RA has taken a major leap forward with the identification of new disease-related genes and further deciphering of the molecular pathways of disease pathogenesis. The relative importance of these different mechanisms has been highlighted by the observed benefits of the new class of highly targeted biologic therapies. Despite these gains, incomplete understanding of the initiating pathogenic pathways of RA remains a sizable barrier to its cure and prevention.

The last two decades have witnessed a remarkable improvement in the outcomes of RA. The historic descriptions of crippling arthritis are currently encountered much less frequently. Much of this progress can be traced to the expanded therapeutic armamentarium and the adoption of early treatment intervention. The shift in treatment strategy dictates a new mind-set for primary care practitioners—namely, one that demands early referral of patients with inflammatory arthritis to a rheumatologist for prompt diagnosis and initiation of therapy. Only then will patients achieve their best outcomes.

CLINICAL FEATURES

The incidence of RA increases between 25 and 55 years of age, after which it plateaus until the age of 75 and then decreases. The presenting symptoms of RA typically result from inflammation of the joints, tendons, and bursae. Patients often complain of early morning joint stiffness lasting more than 1 hour and easing with physical activity. The earliest involved joints are typically the small joints of the hands and feet. The initial pattern of joint involvement may be monoarticular, oligoarticular (≤4 joints), or polyarticular (>5 joints), usually in a symmetric distribution. Some patients with an inflammatory arthritis will present with too few affected joints and other characteristic features to be classified as having RA—so-called undifferentiated inflammatory arthritis. Those with an undifferentiated arthritis, who are most likely to be diagnosed later with RA, have a higher number of tender and swollen joints, test positive for serum rheumatoid factor (RF) or anti-CCP antibodies, and have higher scores for physical disability.

Once the disease process of RA is established, the wrists, metacarpophalangeal (MCP), and proximal interphalangeal (PIP) joints stand out as the most frequently involved joints **(Fig. 321-1)**. Distal interphalangeal (DIP) joint involvement may occur in RA, but it usually is a manifestation of coexistent osteoarthritis. Flexor tendon tenosynovitis is a frequent hallmark of RA and leads to decreased range of motion, reduced grip strength, and "trigger" fingers. Progressive destruction of the joints and soft tissues may lead to chronic, irreversible deformities. Ulnar deviation results from subluxation of the MCP joints, with subluxation of the proximal phalanx to the volar side of the hand. Hyperextension of the PIP joint with flexion of the DIP joint ("swan-neck deformity"), flexion of the PIP joint with hyperextension of the DIP joint ("boutonnière deformity"), and subluxation of the first MCP joint with hyperextension of the first interphalangeal (IP) joint ("Z-line deformity") also may result from damage to the tendons, joint capsule, and other soft tissues in these small joints. Inflammation about the ulnar styloid and tenosynovitis of the extensor carpi ulnaris may cause subluxation of the distal ulna, resulting in a "piano-key movement" of the ulnar styloid. While metatarsophalangeal joint (MTP) involvement is a feature of early disease in the feet, the ankle and midtarsal regions are usually affected later in the course of disease and often predispose to pes planovalgus ("flat feet"). Large joints, including the knees and shoulders, are often affected in established disease, although these joints may remain asymptomatic for many years after onset.

Figure 321-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Metacarpophalangeal and proximal interphalangeal joint swelling in rheumatoid arthritis. *(Courtesy of the American College of Rheumatology Image Bank.)*

Atlantoaxial involvement of the cervical spine is clinically noteworthy because of its potential to cause compressive myelopathy and neurologic dysfunction. Neurologic manifestations are rarely a presenting sign or symptom of atlantoaxial disease, but they may evolve over time with progressive instability of C1 on C2. The prevalence of atlantoaxial subluxation has been declining in recent years, and occurs now in less than 10% of patients. Unlike the spondyloarthritides **(Chap. 325)**, RA does not affect the thoracic and lumbar spine except in very unusual circumstances. Radiographic abnormalities of the temporomandibular joint occur commonly in patients with RA, but they are rarely associated with significant symptoms or functional impairment.

Extraarticular manifestations may develop during the clinical course of RA, even prior to the onset of arthritis **(Fig. 321-2)**. Patients most likely to develop extraarticular disease have a history of smoking, early onset of significant physical disability, and test positive for serum RF. Subcutaneous nodules, secondary Sjögren's syndrome, pulmonary nodules, and anemia are among the most frequently observed extraarticular manifestations. Recent studies have shown a decrease in the incidence and severity of at least some extraarticular manifestations, particularly Felty's syndrome and vasculitis.

Figure 321-2

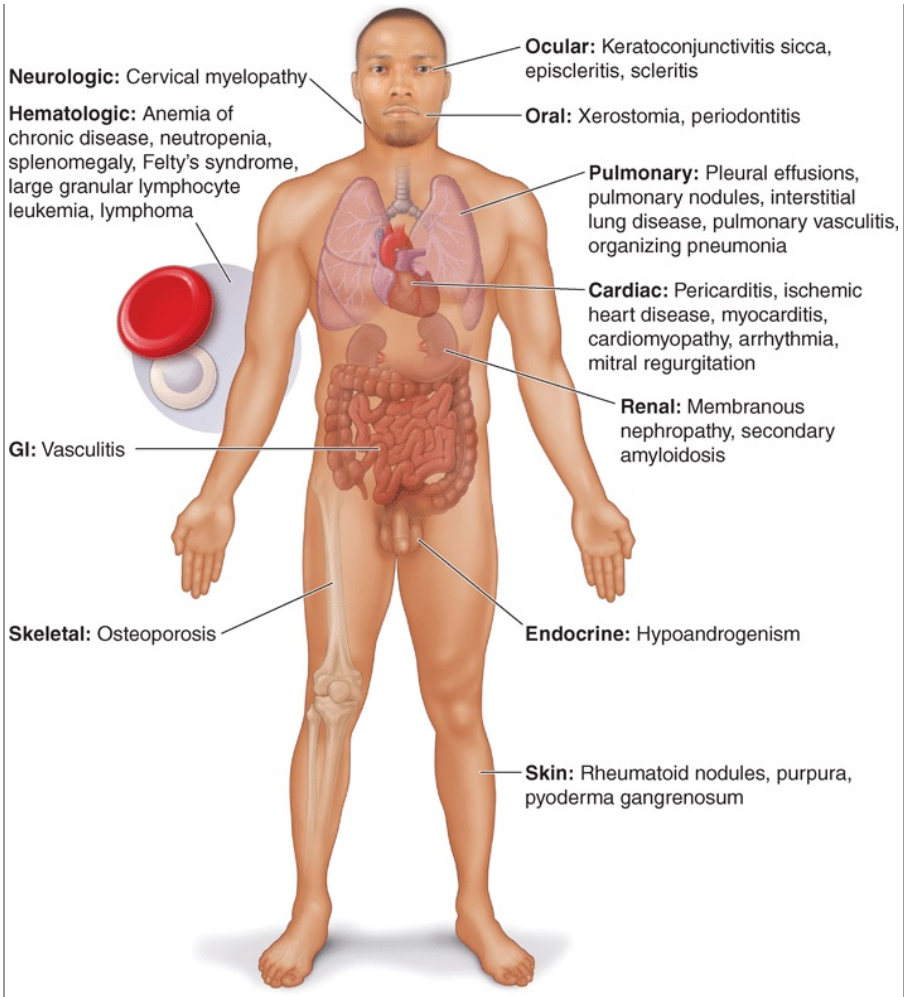

Extraarticular manifestations of rheumatoid arthritis.

The most common systemic and extraarticular features of RA are described in more detail in the sections below.

Constitutional

These signs and symptoms include weight loss, fever, fatigue, malaise, depression, and in the most severe cases, cachexia; they generally reflect a high degree of inflammation and may even precede the onset of joint symptoms. In general, the presence of a fever of >38.3°C (101°F) at any time during the clinical course should raise suspicion of systemic vasculitis (see below) or infection.

Nodules

Subcutaneous nodules occur in 30–40% of patients and more commonly in those with the highest levels of disease activity, the disease-related shared epitope (see below), a positive test for serum RF, and radiographic evidence of joint erosions. When palpated, the nodules are generally firm; nontender; and adherent to periosteum, tendons, or bursae; developing in areas of the skeleton subject to repeated trauma or irritation such as the forearm, sacral prominences, and the Achilles tendon. They may also occur in the lungs, pleura, pericardium, and peritoneum. Nodules are typically benign, although they can be associated with infection, ulceration, and gangrene.

Sjögren's Syndrome

Secondary Sjögren's syndrome **(Chap. 324)** is defined by the presence of either keratoconjunctivitis sicca (dry eyes) or xerostomia (dry mouth) in association with another connective tissue disease, such as RA. Approximately 10% of patients with RA have secondary Sjögren's syndrome.

Pulmonary

Pleural disease, the most common pulmonary manifestation of RA, may produce pleuritic chest pain and dyspnea, as well as a pleural friction rub and effusion. Pleural effusions tend to be exudative with increased numbers of monocytes and neutrophils. Interstitial lung disease (ILD) may also occur in patients with RA and is heralded by symptoms of dry cough and progressive shortness of breath. Diagnosis is readily made by high-resolution chest CT scan. Pulmonary function testing shows a restrictive pattern (e.g., reduced total lung capacity) with a reduced diffusing capacity for carbon monoxide (DLCO). The presence of ILD confers a poor prognosis. The prognosis is not quite as poor as that of idiopathic pulmonary fibrosis (e.g., usual interstitial pneumonitis) because ILD secondary to RA responds more favorably than idiopathic ILD to immunosuppressive therapy **(Chap. 261)**. Pulmonary nodules may be solitary or multiple. Caplan's syndrome is a rare subset of pulmonary nodulosis characterized by the development of nodules and pneumoconiosis following silica exposure. Other less common pulmonary findings include respiratory bronchiolitis and bronchiectasis.

Cardiac

The most frequent site of cardiac involvement in RA is the pericardium. However, clinical manifestations of pericarditis occur in less than 10% of patients with RA despite the fact that pericardial involvement may be detected in nearly one-half of the these

patients by echocardiogram or autopsy studies. Cardiomyopathy, another clinically important manifestation of RA, may result from necrotizing or granulomatous myocarditis, coronary artery disease, or diastolic dysfunction. This involvement too may be subclinical and only identified by echocardiography or cardiac MRI. Rarely, the heart muscle may contain rheumatoid nodules or be infiltrated with amyloid. Mitral regurgitation is the most common valvular abnormality in RA, occurring at a higher frequency than the general population.

Vasculitis

Rheumatoid vasculitis **(Chap. 326)** is seen most commonly in patients with long-standing disease, a positive test for serum RF, and hypocomplementemia; the overall incidence is quite rare, occurring in no more than 1% of cases. The cutaneous signs vary and include petechiae, purpura, digital infarcts, gangrene, livedo reticularis, and in severe cases large, painful lower extremity ulcerations. Vasculitic ulcers, which may be difficult to distinguish from those caused by venous insufficiency, may be treated successfully with immunosuppressive agents (requiring cytotoxic treatment in severe cases) as well as skin grafting. Sensorimotor polyneuropathies, such as mononeuritis multiplex, may occur in association with systemic rheumatoid vasculitis.

Hematologic

A normochromic, normocytic anemia often develops in patients with RA and is the most common hematologic abnormality. The degree of anemia parallels the degree of inflammation, correlating with the levels of serum C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR). Platelet counts may also be elevated in RA as an acute-phase reactant. Immune-mediated thrombocytopenia is rare in this disease.

Felty's syndrome is defined by the clinical triad of neutropenia, splenomegaly, and nodular RA and is seen in less than 1% of patients, although its incidence appears to be declining in the face of more aggressive treatment of the joint disease. It typically occurs in white patients in the late stages of severe RA. T cell large granular lymphocyte leukemia (T-LGL) may have a similar clinical presentation and often occurs in association with RA. T-LGL is characterized by a chronic, indolent clonal growth of LGL cells, leading to neutropenia and splenomegaly. As opposed to Felty's syndrome, T-LGL may develop early in the course of RA. Leukopenia apart from these disorders is uncommon and most often due to drug therapy.

Lymphoma

Large cohort studies have shown a two- to fourfold increased risk of lymphoma in RA patients compared with the general population. The most common histopathologic type of lymphoma is a diffuse large B-cell lymphoma. The risk of developing lymphoma increases if the patient has high levels of disease activity or Felty's syndrome.

Associated Conditions

In addition to extraarticular manifestations, several conditions associated with RA contribute to disease morbidity and mortality rates. They are worthy of mention because they affect chronic disease management.

CARDIOVASCULAR DISEASE

The most common cause of death in patients with RA is cardiovascular disease. The incidence of coronary artery disease and carotid atherosclerosis is higher in RA patients than in the general population even when controlling for traditional cardiac risk factors, such as hypertension, obesity, hypercholesterolemia, diabetes, and cigarette smoking. Furthermore, congestive heart failure (including both systolic and diastolic dysfunction) occurs at an approximately twofold higher rate in RA than in the general population. The presence of elevated serum inflammatory markers appears to confer an increased risk of cardiovascular disease in this population.

OSTEOPOROSIS

Osteoporosis is more common in patients with RA than an age- and sex-matched population, with prevalence rates of 20–30%. The inflammatory milieu of the joint probably spills over into the rest of the body and promotes generalized bone loss by activating osteoclasts. Chronic use of glucocorticoids and disability-related immobility also contributes to osteoporosis. Hip fractures are more likely to occur in patients with RA and are significant predictors of increased disability and mortality rate in this disease.

HYPOANDROGENISM

Men and postmenopausal women with RA have lower mean serum testosterone, luteinizing hormone (LH) and dehydroepiandrosterone (DHEA) levels than control populations. It has thus been hypothesized that hypoandrogenism may play a role in the pathogenesis of RA or arise as a consequence of the chronic inflammatory response. In fact, some studies suggest that higher testosterone levels offer some protection from RA in younger males. The idea that the low serum testosterone levels associated with RA arise from the chronic inflammatory state comes from observations that clinical improvement following successful treatment correlates with an increase in serum testosterone levels. From a clinical perspective, it is important to realize that patients receiving chronic glucocorticoid therapy may develop hypoandrogenism owing to inhibition of LH and follicle-stimulating hormone (FSH) secretion from the pituitary gland. Since low testosterone levels may lead to osteoporosis, men with serum testosterone levels below the physiologic range should be considered for androgen replacement therapy.

EPIDEMIOLOGY

RA affects approximately 0.5–1% of the adult population worldwide. There is evidence that the overall incidence of RA has been decreasing in recent decades, whereas the prevalence has remained the same because individuals with RA are living longer. The incidence and prevalence of RA varies based on geographic location, both globally and among certain ethnic groups within a country **(Fig. 321-3)**. For example, the Native American Yakima, Pima, and Chippewa tribes of North America have reported prevalence rates in some studies of nearly 7%. In contrast, many population studies from Africa and Asia show lower prevalence rates for RA in the range of 0.2–0.4%.

Figure 321-3

European ancestry:
HLA-DRB1:
*0401
*0404
*0301
***0101**

PTPN22: European
STAT 4: North American
TNFAIP3: North American
TRAFI/CF: North American
CTLA-4: European

Asian ancestry:
HLA-DRB1:
*0401 (East Asian)
*0405
*0901 (Japanese, Malaysian, Korean)

PADI4
CD244

Other:
CD40

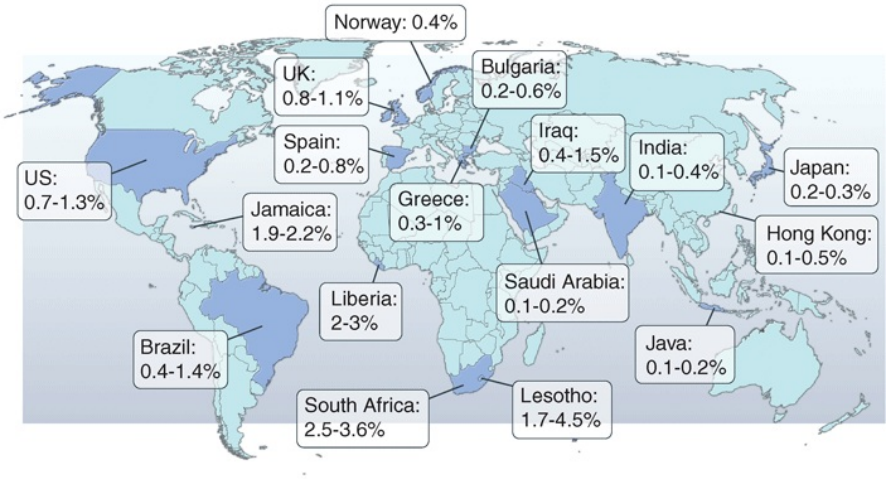

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Global prevalence rates of rheumatoid arthritis (RA) with genetic associations. Listed are the major genetic alleles associated with RA. While human leukocyte antigen (HLA)-DRB1 mutations are found globally, some alleles have been associated with RA in only certain ethnic groups.

Like many other autoimmune diseases, RA occurs more commonly in females than in males, with a 2–3:1 ratio. Interestingly, studies of RA from some of the Latin American and African countries show an even greater predominance of disease in females compared to males, with ratios of 6–8:1. Given this preponderance of females, various theories have been proposed to explain the possible role of estrogen in disease pathogenesis. Most of the theories center on the role of estrogens in enhancing the immune response. For example, some experimental studies have shown that estrogen can stimulate production of tumor necrosis factor α (TNF-α), a major cytokine in the pathogenesis of RA.

GENETIC CONSIDERATIONS

It has been recognized for over 30 years that genetic factors contribute to the occurrence of RA as well as to its severity. The likelihood that a first-degree relative of a patient will share the diagnosis of RA is 2–10 times greater than in the general population. There remains, however, some uncertainty in the extent to which genetics plays a role in the causative mechanisms of RA. While twin studies imply that genetic factors may explain up to 60% of the occurrence of RA, the more commonly stated estimate falls in the range of 10–25%. The estimate of genetic influence may vary across studies due to gene–environment interactions.

The alleles known to confer the greatest risk of RA are located within the major histocompatibility complex (MHC). It has been estimated that one-third of the genetic risk for RA resides within this locus. Most, but probably not all, of this risk is associated with allelic variation in the HLA-DRB1 gene, which encodes the MHC II β-chain molecule. The disease-associated HLA-DRB1 alleles share an amino acid sequence at positions 70–74 in the third hypervariable regions of the HLA-DR β-chain, termed the *shared epitope (SE)*. Carriership of the SE alleles is associated with production of anti-CCP antibodies and worse disease outcomes. Some of these HLA-DRB1 alleles bestow a high risk of disease (*0401), whereas others confer a more moderate risk (*0101, *0404, *1001, and *0901). In Greece, for example, where RA tends to be milder than in western European countries, RA susceptibility has been associated with the *0101 SE allele. By comparison, the *0401 or *0404 alleles are found in approximately 50–70% of Northern Europeans and are the predominant risk alleles in this group. The most common disease susceptibility SE alleles in Asians, namely the Japanese, Koreans, and Chinese, are *0405 and *0901. Lastly, disease susceptibility of Native American populations such as the Pima and Tlingit Indians, where the prevalence of RA can be as high as 7%, is associated with the SE allele *1042. The risk of RA conferred by these SE alleles is less in African and Hispanic Americans than in individuals of European ancestry.

Genome-wide association studies (GWAS) have made possible the identification of several non-MHC-related genes that contribute to RA susceptibility. GWAS are based on the detection of single-nucleotide polymorphisms (SNPs), which allow for examination of the genetic architecture of complex diseases such as RA. There are approximately 10 million common SNPs within a human genome consisting of 3 billion base pairs. As a rule, GWAS identify only common variants, namely, those with a frequency of more than 5% in the general population.

Overall, several themes have emerged from GWAS in RA. First, several of the non-MHC loci identified as risk alleles for RA have only a modest effect on risk; they also contribute to the risk for developing other autoimmune diseases, such as Type 1 diabetes mellitus, systemic lupus erythematosus, and multiple sclerosis. Second, most of the associations are described in patients with anti-CCP antibody-positive disease. Third, risk alleles vary among ethnic groups. Fourth, the risk loci mostly reside in genes encoding proteins involved in the regulation of the immune response, such as the signaling pathway for the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB). However, the risk alleles identified by GWAS only account at present for approximately 5% of the genetic risk, suggesting that rare variants or other classes of DNA variants, such as variants in copy number, may be yet found that significantly contribute to the overall risk model.

Among the best examples of the non-MHCgenes contributing to the risk of RA is the gene encoding protein tyrosine phosphatase non-receptor 22 (PTPN22). This gene varies in frequency among patients from different parts of Europe (e.g., 3–10%), but is absent in patients of East Asian ancestry. PTPN22 encodes lymphoid tyrosine phosphatase, a protein that regulates T and B cell function. Inheritance of the risk allele for PTPN22 produces a gain-of-function in the protein that is hypothesized to result in the abnormal selection of autoreactive T and B cells. In RA, carriage of the variant PTPN22 appears to be associated exclusively with anti-CCP-positive disease. The peptidyl arginine deiminase type IV (PADI4) gene is another risk allele that encodes an enzyme involved in the conversion of arginine to citrulline and is postulated to play a role in the development of antibodies to citrullinated antigens. A polymorphism in PADI4 has been only associated with RA in Asian

populations. Other SNPs associated with RA have been found in the genes for signal transducer and activator of transcription 4 (STAT4), CD244 (natural killer cell receptor 2B4), Fc receptor-like 3 (FCLR3), tumor necrosis factor (TNF) alpha-induced protein 3 (TNF-AIP3), and TNF receptor-associated factor 1 (TRAF1). These genes encode proteins with various roles in B and T cell signaling.

ENVIRONMENTAL FACTORS

In addition to genetic predisposition, a host of environmental factors have been implicated in the pathogenesis of RA. The most reproducible of these environmental links is cigarette smoking. Numerous cohort and case control studies have demonstrated that smoking confers a relative risk for developing RA of 1.5–3.5. A twin who smokes will have a significantly higher risk for RA than his or her monozygotic co-twin, theoretically with the same genetic risk, who does not smoke. Interestingly, the risk from smoking is almost exclusively related to RF- and anti-CCP antibody-positive disease.

Researchers began to aggressively seek an infectious etiology for RA after the discovery in 1931 that sera from patients with this disease could agglutinate strains of streptococci. Certain viruses such as Epstein-Barr virus (EBV) have garnered the most interest over the past 30 years given their ubiquity, ability to persist for many years in the host, and frequent association with arthritic complaints. For example, titers of IgG antibodies against EBV antigens in the peripheral blood and saliva are significantly higher in patients with RA than the general population. EBV DNA has also been found in synovial fluid and synovial cells of RA patients. Blood and synovial fluid analyses have also suggested a possible link with mycoplasma and parvovirus B19 infection. Since the evidence for these links is largely circumstantial, it has not been possible to directly implicate infection as a causative factor in RA.

PATHOLOGY

RA affects the synovial tissue and underlying cartilage and bone. The synovial membrane, which covers most articular surfaces, tendon sheaths, and bursae, normally is a thin layer of connective tissue. In joints, it faces the bone and cartilage, bridging the opposing bony surfaces and inserting at periosteal regions close to the articular cartilage. It consists primarily of two cell types—type A synoviocytes (macrophage-derived) and type B synoviocytes (fibroblast-derived). The synovial fibroblasts are the most abundant and produce the structural components of joints, including collagen, fibronectin, and laminin, as well as other extracellular constituents of the synovial matrix. The sublining layer consists of blood vessels and a sparse population of mononuclear cells within a loose network of connective tissue. Synovial fluid, an ultrafiltrate of blood, diffuses through the subsynovial lining tissue across the synovial membrane and into the joint cavity. Its main constituents are hyaluronan and lubricin. Hyaluronan is a glycosaminoglycan that contributes to the viscous nature of synovial fluid, which along with lubricin, lubricates the surface of the articular cartilage.

The pathologic hallmarks of RA are synovial inflammation and proliferation, focal bone erosions, and thinning of articular cartilage. Chronic inflammation leads to synovial lining hyperplasia and the formation of pannus, a thickened cellular membrane of granulation–reactive fibrovascular tissue that invades the underlying cartilage and bone. Cadherin-11, the chief organizing molecule of the synovial membrane, confers the invasive nature of the fibroblast-like synoviocytes, the prevailing cell type of the pannus. The inflammatory infiltrate is made up of no less than 6 cell types: T cells, B cells, plasma cells, dendritic cells, mast cells, and a few granulocytes. The T cells comprise 30–50% of the infiltrate, with the other cells accounting for the remainder. The topographical organization of these cells is complex and may vary among individuals with RA. Most often, the lymphocytes are diffusely organized among the tissue resident cells; however, in some cases, the B cells, T cells, and dendritic cells may form higher levels of organization, such as lymphoid follicles and germinal center–like structures. Growth factors secreted by synovial fibroblasts and macrophages promote the formation of new blood vessels in the synovial sublining that supply the increasing demands for oxygenation and nutrition required by the infiltrating leukocytes and expanding synovial tissue.

The structural damage to the mineralized cartilage and subchondral bone is mediated by the osteoclast. Osteoclasts are multinucleated giant cells that can be identified by their expression of CD68, tartrate-resistant acid phosphatase, cathepsin K, and the calcitonin receptor. They appear at the pannus-bone interface where they eventually form resorption lacunae. These lesions typically localize where the synovial membrane inserts into the periosteal surface at the edges of bones close to the rim of articular cartilage and at the attachment sites of ligaments and tendon sheaths. This process most likely explains why bone erosions usually develop at the radial sites of the MCP joints juxtaposed to the insertion sites of the tendons, collateral ligaments, and synovial membrane. Another form of bone loss is periarticular osteopenia that occurs in joints with active inflammation. It is associated with substantial thinning of the bony trabeculae along the metaphyses of bones, and likely results from inflammation of the bone marrow cavity. These lesions can be visualized on MRI scans, where they appear as signal alterations in the bone marrow adjacent to inflamed joints. Their signal characteristics show they are water-rich with a low fat content, and consistent with highly vascularized inflammatory tissue. These bone marrow lesions are often the forerunner of bone erosions.

The cortical bone layer that separates the bone marrow from the invading pannus is relatively thin and susceptible to penetration by the inflamed synovium. The bone marrow lesions seen on MRI scans are associated with an endosteal bone response characterized by the accumulation of osteoblasts and deposition of osteoid. Thus, in recent years, the concept of joint pathology in RA has been extended to include the bone marrow cavity. Finally, a third form of bone loss is generalized osteoporosis, which results in the thinning of trabecular bone throughout the body.

Articular cartilage is an avascular tissue comprised of a specialized matrix of collagens, proteoglycans, and other proteins. It is organized in four distinct regions (superficial, middle, deep, and calcified cartilage zones)—chondrocytes constitute the unique cellular component in these layers. Originally, cartilage was considered to be an inert tissue, but it is now known to be a highly responsive tissue that reacts to inflammatory mediators and mechanical factors, which in turn, alter the balance between cartilage anabolism and catabolism. In RA, the initial areas of cartilage degradation are juxtaposed to the synovial pannus. The cartilage matrix is characterized by a generalized loss of proteoglycan, most evident in the superficial zones adjacent to the synovial fluid. Degradation of cartilage may also take place in the perichondrocytic zone and in regions adjacent to the subchondral bone.

PATHOGENESIS

The pathogenic mechanisms of synovial inflammation are likely to result from a complex interplay of genetic, environmental, and immunologic factors that produces dysregulation of the immune system and a breakdown in self-tolerance **(Fig. 321-4)**. Precisely what triggers these initiating events and what genetic and environmental factors disrupt the immune system remain a mystery. However, a detailed molecular picture is emerging of the mechanisms underlying the chronic inflammatory response and the destruction of the articular cartilage and bone.

Figure 321-4

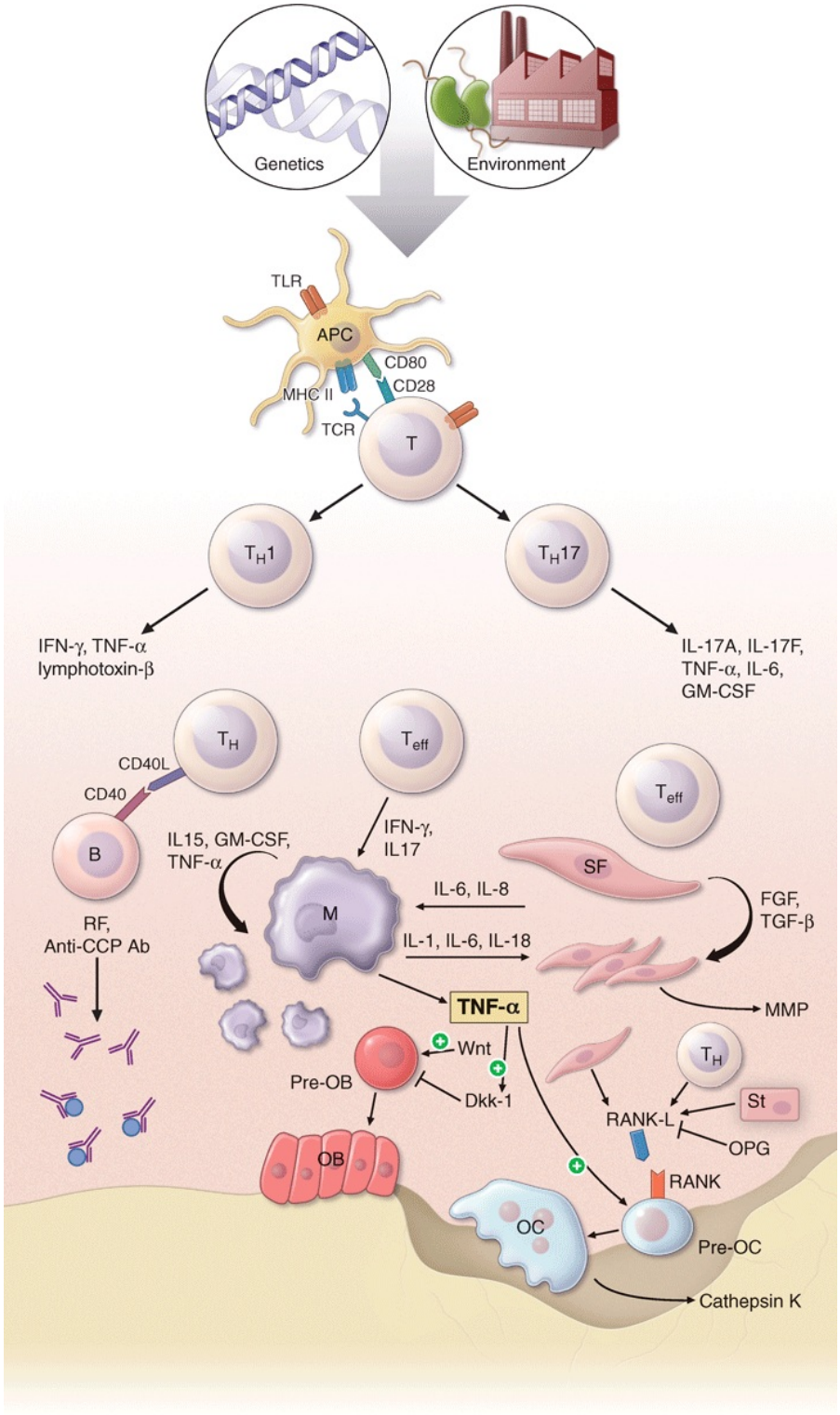

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathophysiologic mechanisms of inflammation and joint destruction. Genetic predisposition along with environmental factors may trigger the development of rheumatoid arthritis (RA), with subsequent synovial T cell activation. CD4+ T cells become activated by antigenpresenting cells (APCs) through interactions between the T cell receptor and class II major histocompatibility complex (MHC)-peptide antigen (signal 1) with co-stimulation through the CD28-CD80/86 pathway, as well as other pathways (signal 2). In theory, ligands binding Toll-like receptors (TLRs) may further stimulate activation of APCs inside the joint. Synovial CD4+ T cells differentiate into T_H1 and T_H17 cells, each with their distinctive cytokine profile. CD4+ T_H cells in turn activate B cells, some of which are destined to differentiate into autoantibody-producing plasma cells. Immune complexes, possibly comprised of rheumatoid factors (RFs) and anti–cyclic citrullinated peptides (CCP) antibodies, may form inside the joint, activating the complement pathway and amplifying inflammation. T effector cells stimulate synovial macrophages (M) and fibroblasts (SF) to secrete proinflammatory mediators, among which is tumor necrosis factor α (TNF-α). TNF-α upregulates adhesion molecules on endothelial cells, promoting leukocyte influx into the joint. It also stimulates the production of other inflammatory mediators, such as interleukin 1 (IL-1), IL-6, and granulocyte-macrophage colony-stimulating factor (GM-CSF). TNF-α has a critically important function in regulating the balance between bone destruction and formation. It upregulates the expression of dickkopf-1 (DKK-1), which can then internalize Wnt receptors on osteoblast precursors. Wnt is a soluble mediator that promotes osteoblastogenesis and bone formation. In RA, bone formation is inhibited through the Wnt pathway, presumably due to the action of elevated levels of DKK-1. In addition to inhibiting bone formation, TNF-α stimulates osteoclastogenesis. However, it is not sufficient by itself to induce the differentiation of osteoclast precursors (Pre-

stimulates osteoclastogenesis. However, it is not sufficient by itself to induce the differentiation of osteoclast precursors (Pre-OC) into activated osteoclasts capable of eroding bone. Osteoclast differentiation requires the presence of macrophage colony-stimulating factor (M-CSF) and receptor activator of nuclear factor κB (RANK) ligand, which binds to RANK on the surface of Pre-OC. Inside the joint, RANKL is mainly derived from stromal cells, synovial fibroblasts, and T cells. Osteoprotegerin (OPG) acts as a decoy receptor for RANKL, thereby inhibiting osteoclastogenesis and bone loss. FGF, fibroblast growth factor; IFN, interferon; TGF, transforming growth factor.

In RA, the earliest detectable preclinical stage is a breakdown in self-tolerance. This idea is supported by the finding that autoantibodies, such as RF and anti-CCP antibodies, may be found in sera from patients long before clinical disease. However, the antigenic targets of anti-CCP antibodies and RF are not restricted to the joint, and their role in disease pathogenesis remains speculative. Anti-CCP antibodies are directed against deiminated peptides, which result from posttranslational modification by the enzyme PADI4. They recognize citrulline-containing regions of several different matrix proteins, including filaggrin, keratin, fibrinogen, and vimentin. Other autoantibodies have been found in a minority of patients with RA, but they also occur in the setting of other types of arthritis. They bind to a diverse array of autoantigens, including type II collagen, human cartilage gp-39, aggrecan, calpastatin, BiP (immunoglobulin binding protein), and glucose-6-phosphate isomerase.

In theory, environmental stimulants may synergize with other factors to bring about inflammation in RA. People who smoke display higher citrullination of proteins in bronchoalveolar fluid than those who do not smoke. Thus, it has been speculated that long-term exposure to tobacco smoke might induce citrullination of cellular proteins in the lungs and enhance the expression of a neoepitope capable of inducing self-reactivity. Exposure to silicone dust and mineral oil, which has adjuvant effects, has also been linked to an increased risk for anti-CCP antibody-positive RA.

How might microbes or their products be involved in the initiating events of RA? The immune system is alerted to the presence of microbial infections through Toll-like receptors (TLRs). There are 10 TLRs in humans that recognize a variety of microbial products, including bacterial cell-surface lipopolysaccharides and heat- shock proteins (TLR4), lipoproteins (TLR2), double-strand RNA viruses (TLR3), and unmethylated CpG DNA from bacteria (TLR9). TLR2, -3, and -4 are abundantly expressed by synovial fibroblasts in early RA, and when bound by their ligands upregulate production of proinflammatory cytokines. Although such events could amplify inflammatory pathways in RA, a specific role for TLRs in disease pathogenesis has not been elucidated.

The pathogenesis of RA is built upon the concept that self-reactive T cells drive the chronic inflammatory response. In theory, self-reactive T cells might arise in RA from abnormal central (thymic) selection due to defects in DNA repair leading to an imbalance of T cell death and life, or defects in the cell signaling apparatus lowering the threshold for T cell activation. Similarly, abnormal selection of the T cell repertoire in the periphery might lead to a breakdown in T cell tolerance. The support for these theories comes mainly from studies of arthritis in mouse models. It has not been shown that patients with RA have abnormal thymic selection of T cells or defective apoptotic pathways regulating cell death. At least some antigen stimulation inside the joint seems likely, owing to the fact that T cells in the synovium express a cell-surface phenotype indicating prior antigen exposure and show evidence of clonal expansion. Of interest, peripheral blood T cells from patients with RA have been shown to display a fingerprint of premature aging that mostly affects inexperienced naïve T cells. In these studies, the most glaring findings have been the loss of telomeric sequences and a decrease in the thymic output of new T cells. While intriguing, it is not clear how generalized T cell abnormalities might provoke a systemic disease dominated by synovitis.

There is substantial evidence supporting a role for CD4+ T cells in the pathogenesis of RA. First, the co-receptor CD4 on the surface of T cells binds to invariant sites on MHC class II molecules, stabilizing the MHC-peptide–T cell receptor complex during T cell activation. Since the SE on MHC class II molecules is a risk factor for RA, it may be speculated that CD4+ T cell activation plays a role in the pathogenesis of this disease. Second, CD4+ memory T cells are enriched in the synovial tissue from patients with RA, and can be implicated through "guilt by association." Third, CD4+ T cells have been shown to be important in the initiation of arthritis in animal models. Fourth, T cell–directed therapies, such as cyclosporine and abatacept (CTLA4-Ig), have shown clinical efficacy in this disease. Taken together, these lines of evidence suggest that CD4+ T cells play an important role in orchestrating the chronic inflammatory response in RA. However, other cell types, such as CD8+ T cells, natural killer (NK) cells, and B cells are present in synovial tissue and may also influence pathogenic responses.

In the rheumatoid joint, by mechanisms of cell-cell contact and release of soluble mediators, activated T cells stimulate macrophages and fibroblast-like synoviocytes to generate proinflammatory mediators and proteases that drive the synovial inflammatory response and destroy the cartilage and bone. CD4+ T cells also provide help to B cells, which in turn, produce antibodies that may promote further inflammation in the joint. The previous T cell–centric model for the pathogenesis of RA was based on a T_H 1-driven paradigm, which came from studies indicating that CD4+ T helper (T_H) cells differentiated into T_H1 and T_H2 subsets, each with their distinctive cytokine profiles. T_H1 cells were found to mainly produce interferon γ (IFN-γ), lymphotoxin β, and TNF-α, while T_H2 cells predominately secreted interleukin (IL)-4, IL-5, IL-6, IL-10, and IL-13. The recent discovery of another subset of T_H cells, namely the T_H17 lineage, has revolutionized our concepts concerning the pathogenesis of RA. In humans, naïve T cells are induced to differentiate into T_H17 cells by exposure to transforming growth factor β (TGF-β), IL-1, IL-6, and IL-23. Upon activation, T_H17 cells secrete a variety of proinflammatory mediators such as IL-17, IL-21, IL-22, TNF-α, IL-26, IL-6, and granulocyte-macrophage colony-stimulating factor (GM-CSF). Substantial evidence now exists from both animal models and humans that IL-17 plays an important role not only in promoting joint inflammation but also in destroying cartilage and subchondral bone.

Activated B cells are also important players in the chronic inflammatory response. B cells give rise to plasma cells, which in turn, produce antibodies, including RF and anti-CCP antibodies. RFs may form large immune complexes inside the joint that contribute to the pathogenic process by fixing complement and promoting the release of proinflammatory chemokines and chemoattractants. In mouse models of arthritis, RF-containing immune complexes as well as anti-CCP-containing immune complexes synergize with other mechanisms to exacerbate the synovial inflammatory response.

RA is often considered to be a macrophage-driven disease because this cell type is the predominant source of proinflammatory cytokines inside the joint. Key proinflammatory cytokines released by synovial macrophages include TNF-α, IL-1, IL-6, IL-12, IL-15, IL-18, and IL-23. Synovial fibroblasts, the other major cell type in this microenvironment, produce the cytokines IL-1 and IL-6 as well as TNF-α. TNF-α is a pivotal cytokine in the pathobiology of synovial inflammation. It upregulates adhesion molecules on endothelial cells, promoting the influx of leukocytes into the synovial microenvironment. It also activates synovial fibroblasts, stimulates angiogenesis, promotes pain receptor sensitizing pathways, and drives osteoclastogenesis. Fibroblasts secrete matrix metalloproteinases (MMPs) as well as other proteases that are chiefly responsible for the breakdown of articular cartilage.

Osteoclast activation at the site of the pannus is closely tied to the presence of focal bone erosion. Receptor activator of nuclear factor κB ligand (RANKL) is expressed by stromal cells, synovial fibroblasts, and T cells. Upon binding to its receptor RANK on osteoclast progenitors, RANKL stimulates osteoclast differentiation and bone resorption. RANKL activity is regulated by osteoprotegerin (OPG), a decoy receptor of RANKL that blocks osteoclast formation. Monocytic cells in the synovium serve

as the precursors of osteoclasts and, when exposed to macrophage colony-stimulating factor (M-CSF) and RANKL, fuse to form polykaryons termed *preosteoclasts*. These precursor cells undergo further differentiation into osteoclasts with the characteristic ruffled membrane. Cytokines such as TNF-α, IL-1, IL-6, and IL-17 increase the expression of RANKL in the joint and thus promote osteoclastogenesis. Osteoclasts also secrete cathepsin K, which is a cysteineprotease that degrades the bone matrix by cleaving collagen.

Increased bone loss is only part of the story in RA, as decreased bone formation plays a crucial role in bone remodeling at sites of inflammation. Recent evidence shows that inflammation suppresses bone formation. The proinflammatory cytokineTNF-α plays a key role in actively suppressing bone formation by enhancing the expression of dickkopf-1 (DKK-1). DKK-1 is an important inhibitor of the Wnt pathway, which acts to promote osteoblast differentiation and bone formation. The Wnt system is a family of soluble glycoproteins that bind to cell-surface receptors known as frizzled (fz) and low-density lipoprotein (LDL) receptor–related proteins (LRPs) and promote cell growth. In animal models, increased levels of DKK-1 are associated with decreased bone formation, while inhibition of DKK-1 protects against structural damage in the joint. Wnt proteins also induce the formation of OPG and thereby shut down bone resorption, emphasizing their key role in tightly regulating the balance between bone resorption and formation.

DIAGNOSIS

The clinical diagnosis of RA is largely based on signs and symptoms of a chronic inflammatory arthritis, with laboratory and radiographic results providing important supplemental information. In 2010, a collaborative effort between the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) revised the 1987 ACR classification criteria for RA in an effort to improve early diagnosis with the goal of identifying patients who would benefit from early introduction of disease-modifying therapy **(Table 321-1)**. Application of the newly revised criteria yields a score of 0–10, with a score of ≥6 fulfilling the requirements for definite RA. The new classification criteria differ in several ways from the older criteria set. The new criteria include a positive test for serum anti-cyclic citrullinated peptide antibodies as an item, which carries greater specificity for the diagnosis of RA than a positive test for rheumatoid factor. The newer classification criteria also do not take into account if the patient has rheumatoid nodules or radiographic joint damage because these findings occur rarely in early RA. It is important to emphasize that the new 2010 ACR-EULAR criteria are "classification criteria" as opposed to "diagnostic criteria" and serve to distinguish patients at the onset of disease with a high likelihood of evolving into a chronic disease with persistent synovitis and joint damage. The presence of radiographic joint erosions or subcutaneous nodules may inform the diagnosis in the later stages of the disease.

Table 321-1 Classification Criteria for Rheumatoid Arthritis

		Score
Joint involvement	1 large joint (shoulder, elbow, hip, knee, ankle)	0
	2–10 large joints	1
	1–3 small joints (MCP, PIP, Thumb IP, MTP, wrists)	2
	4–10 small joints	3
	>10 joints (at least 1 small joint)	5
Serology	Negative RF and negative ACPA	0
	Low-positive RF or low-positive anti-CCP antibodies (≤3 times ULN)	2
	High-positive RF or high-positive anti-CCP antibodies (>3 times ULN)	3
Acute-phase reactants	Normal CRP and normal ESR	0
	Abnormal CRP or abnormal ESR	1
Duration of symptoms	<6 weeks	0
	≥6 weeks	1

Note: These criteria are aimed at classification of newly presenting patients who have at least 1 joint with definite clinical synovitis that is not better explained by another disease.

Abbreviations: CCP, cyclic citrullinated peptides; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; IP, interphalangeal joint; MCP, metacarpophalangeal joint; MTP, metatarsophalangeal joint; PIP, proximal interphalangeal joint; RF, rheumatoid factor; ULN, upper limit of normal.

Source: Neogi et al: Arthritis Rheum 62:2569, 2010.

Laboratory Features

IgM, IgG, and IgA isotypes of RF occur in sera from patients with RA, although the IgM isotype is the one most frequently measured by commercial laboratories. Serum IgM RF has been found in 75–80% of patients with RA; therefore, a negative result does not exclude the presence of this disease. It is also found in other connective tissue diseases, such as primary Sjögren's syndrome, systemic lupus erythematosus, and type II mixed essential cryoglobulinemia, as well as chronic infections such as subacute bacterial endocarditis and hepatitis B and C. Serum RF may also be detected in 1–5% of the healthy population.

The presence of serum anti-CCP antibodies has about the same sensitivity as serum RF for the diagnosis of RA. However, its diagnostic specificity approaches 95%, so a positive test for anti-CCP antibodies in the setting of an early inflammatory arthritis is useful for distinguishing RA from other forms of arthritis. There is some incremental value in testing for the presence of both RF and anti-CCP, as some patients with RA are positive for RF but negative for anti-CCP and visa versa. The presence of RF or anti-CCP antibodies also has prognostic significance, with anti-CCP antibodies showing the most value for predicting worse outcomes.

SYNOVIAL FLUID ANALYSIS

Typically, synovial fluid from patients with RA reflects an inflammatory state. Synovial fluid white blood cell (WBC) counts can vary widely, but generally range between 5000 and 50,000 WBC/μ^3 compared to <2000 WBC/μ for a non-inflammatory

condition such as osteoarthritis. In contrast to the synovial tissue, the overwhelming cell type in the synovial fluid is the neutrophil. Synovial fluid also contains RF and anti-CCP antibodies and immune complexes, as well as by-products of complement activation. Clinically, the analysis of synovial fluid is most useful for confirming an inflammatory arthritis (as opposed to osteoarthritis), while at the same time excluding infection or a crystal-induced arthritis such as gout or pseudogout **(Chap. 333)**.

Joint Imaging

Joint imaging is a valuable tool not only for diagnosing RA, but also for tracking progression of any joint damage. Plain x-ray is the most common imaging modality, but it is limited to visualization of the bony structures and inferences about the state of the articular cartilage based on the amount of narrowing of the joint space. MRI and ultrasound techniques offer the added value of detecting changes in the soft tissues such as synovitis, tenosynovitis, and effusions as well as greater sensitivity for identifying bony abnormalities. Plain radiographs are usually relied upon in clinical practice for the purpose of diagnosis and monitoring of affected joints. However, in selected cases, MRI and ultrasound can provide additional diagnostic information that may guide clinical decision-making.

PLAIN RADIOGRAPHY

Classically in RA, the initial radiographic finding is juxtaarticular osteopenia. Practically speaking, however, this finding is difficult to appreciate on plain films, and in particular, on the newer digitalized x-rays. Other findings on plain radiographs include soft tissue swelling, symmetric joint space loss, and subchondral erosions, most frequently in the wrists and hands (MCPs and PIPs) and the feet (MTPs). In the feet, the lateral aspect of the fifth MTP is often targeted first, but other MTP joints may be involved at the same time. X-ray imaging of advanced RA may reveal signs of severe destruction, including joint subluxation and collapse **(Fig. 321-5)**.

Figure 321-5

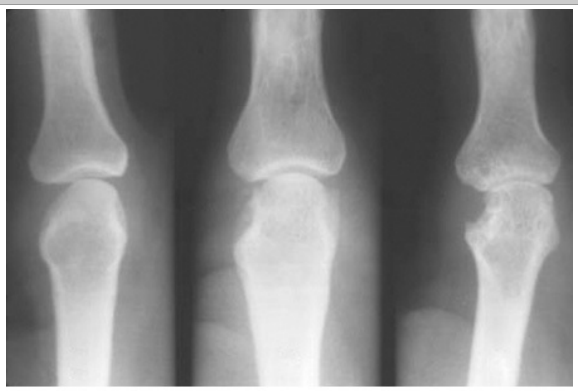

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

X-ray demonstrating progression of erosions on the proximal interphalangeal joint. *(Courtesy of the American College of Rheumatology.)*

MRI

MRI offers the greatest sensitivity for detecting synovitis and joint effusions, as well as early bone and bone marrow changes. These soft tissue abnormalities often occur before osseous changes are noted on x-ray. Presence of bone marrow edema has been recognized to be an early sign of inflammatory joint disease, and can predict the subsequent development of erosions on plain radiographs as well as MRI scans. Cost and availability of MRI are the main factors limiting its routine clinical use.

ULTRASOUND

Ultrasound, including power color Doppler, has the ability to detect more erosions than plain radiography, especially in easily accessible joints. Less clear, however, is the ability of ultrasound to reliably detect synovitis, including increased joint vascularity indicative of inflammation. The usefulness of ultrasound is dependent on the experience of the sonographer; however, it does offer the advantages of portability, lack of radiation, and low expense relative to MRI, factors that make it attractive as a clinical tool.

CLINICAL COURSE

The natural history of RA is complex and affected by a number of factors including age of onset, gender, genotype, phenotype (i.e., extraarticular manifestations or variants of RA), and comorbid conditions, which make for a truly heterogeneous disease. There is no simple way to predict the clinical course. It is important to realize that as many as 10% of patients with inflammatory arthritis fulfilling ACR classification criteria for RA will undergo a spontaneous remission within 6 months (particularly seronegative patients). However, the vast majority of patients will exhibit a pattern of persistent and progressive disease activity that waxes and wanes in intensity over time. A minority of patients will show intermittent and recurrent explosive attacks of inflammatory arthritis interspersed with periods of disease quiescence. Finally, an aggressive form of RA may occur in an unfortunate few with inexorable progression of severe erosive joint disease, although this highly destructive course is less common in the modern treatment era of biologics.

Disability, as measured by the Health Assessment Questionnaire (HAQ), shows gradual worsening of disability over time in the face of poorly controlled disease activity and disease progression. Disability may result from both a disease activity–related component that is potentially reversible with therapy and a joint damage–related component owing to the cumulative and largely irreversible effects of cartilage and bone breakdown. Early in the course of disease, the extent of joint inflammation is the primary determinant of disability, while in the later stages of disease, the amount of joint damage is the dominant contributing factor. Previous studies have shown that more than one-half of patients with RA are unable to work 10 years after the onset of their disease; however, increased employability and less work absenteeism has been reported recently with the use of newer therapies and earlier treatment intervention.

The overall mortality rate in RA is two times greater than the general population, with ischemic heart disease being the most common cause of death followed by infection. Median life expectancy is shortened by an average of 7 years for men and 3 years for women compared to control populations. Patients at higher risk for shortened survival are those with systemic

extraarticular involvement, low functional capacity, low socioeconomic status, low education, and chronic prednisone use.

Treatment: Rheumatoid Arthritis

The amount of clinical disease activity in patients with RA reflects the overall burden of inflammation and is the variable most influencing treatment decisions. Joint inflammation is the main driver of joint damage and is the most important cause of functional disability in the early stages of disease. Several composite indices have been developed to assess clinical disease activity. The ACR 20, 50, and 70 improvement criteria [which corresponds to a 20%, 50%, and 70% improvement in joint counts, physician/patient assessment of disease severity, pain scale, serum levels of acute-phase reactants (ESR or CRP), and a functional assessment of disability using a self-administered patient questionnaire] are a composite index with a dichotomous response variable. The ACR improvement criteria are commonly used in clinical trials as an endpoint for comparing the proportion of responders between treatment groups. In contrast, the Disease Activity Score (DAS), Simplified Disease Activity Index (SDAI), and the Clinical Disease Activity Index (CDAI) are continuous measures of disease activity. These scales are increasingly used in clinical practice for tracking disease status, and in particular, for documenting treatment response.

Several developments during the past two decades have changed the therapeutic landscape in RA. They include: (1) the emergence of methotrexate as the disease-modifying antirheumatic drug (DMARD) of first choice for the treatment of early RA; (2) the development of novel highly efficacious biologicals that can be used alone or in combination with methotrexate; and (3) the proven superiority of combination DMARD regimens over methotrexate alone. The medications used for the treatment of RA may be divided into broad categories: nonsteroidal anti-inflammatory drugs (NSAIDs); glucocorticoids, such as prednisone and methylprednisolone; conventional disease-modifying anti-rheumatic drugs (DMARDs); and biologic DMARDs **(Table 321-2)**. While disease for some patients with RA is managed adequately with a single DMARD, such as methotrexate, the situation entails in most cases the use of a combination DMARD regimen that may vary in its components over the treatment course depending on fluctuations in disease activity and emergence of drug-related toxicities and comorbidities.

Table 321-2 DMARDs Used for the Treatment of Rheumatoid Arthritis

Drug	Dosage	Serious Toxicities	Other Common Side Effects	Initial Evaluation	Monitoring
Hydroxychloroquine	200–400 mg/d orally (≤6.5 mg/kg)	Irreversible retinal damage Cardiotoxicity Blood dyscrasia	Nausea Diarrhea Headache Rash	Eye examination if > 40 years old or prior ocular disease	Funduscopic and visual field testing every 12 months
Sulfasalazine	Initial: 500 mg orally twice daily Maintenance: 1000–1500 mg twice daily	Granulocytopenia Hemolytic anemia (with G6PD deficiency)	Nausea Diarrhea Headache	CBC, LFTs G6PD level	CBC every 2–4 weeks for first 3 months, then every 3 months
Methotrexate	10–25 mg/week orally or SQ Folic acid 1 mg/d to reduce toxicities	Hepatotoxicity Myelosuppression Infection Interstitial pneumonitis Pregnancy category X	Nausea Diarrhea Stomatitis/mouth ulcers Alopecia Fatigue	CBC, LFTs Viral hepatitis panel* Chest x-ray	CBC, creatinine, LFTs every 2–3 months
Leflunomide	10–20 mg/d	Hepatotoxicity Myelosuppression Infection Pregnancy category X	Alopecia Diarrhea	CBC, LFTs Viral hepatitis panel*	CBC, creatinine, LFTs every 2–3 months
TNF-α Inhibitors	Infliximab: 3 mg/kg IV at weeks 0, 2, 6, then every 8 weeks. May increase dose up to 10 mg/kg every 4 weeks	↑ Risk bacterial, fungal infections Reactivation of latent TB ↑ Lymphoma risk (controversial) Drug-induced lupus Neurologic deficits	Infusion reaction ↑ LFTs	PPD skin test	LFTs periodically
	Etanercept: 50 mg SQ weekly, or 25 mg SQ biweekly	As above	Injection site reaction	PPD skin test	Monitor for injection site reactions
	Adalimumab: 40 mg SQ every other week	As above	Injection site reaction	PPD skin test	Monitor for injection site reactions
	Golimumab: 50 mg SQ monthly	As above	Injection site reaction	PPD skin test	Monitor for injection site reactions
	Certolizumab: 400 mg SQ weeks 0, 2, 4 then 200 mg every other week	As above	Injection site reaction	PPD skin test	Monitor for injection site

					reactions
Abatacept	Weight based: <60 kg: 500 mg 60 –100 kg: 750 mg >100 kg: 1000 mg IV dose at week 0, 2 and 4, and then every 4 weeks	↑ Risk bacterial, viral infections	Headache Nausea	PPD skin test	Monitor for infusion reactions
Anakinra	100 mg SQ daily	↑ Risk bacterial, viral infections Reactivation of latent TB Neutropenia	Injection site reaction Headache	PPD skin test CBC with differential	CBC every month for 3 months, then every 4 months for 1 year Monitor for injection site reactions
Rituximab	1000 mg IV x 2, day 0 and 14	↑ Risk bacterial, viral infections Infusion reaction	Rash Fever	CBC Viral hepatitis panel*	CBC at regular intervals
	May repeat course every 24 weeks or more Premedicate with methylprednisolone 100 mg to decrease infusion reaction	Cytopenia Hepatitis B reactivation			
Tocilizumab	4–8 mg/kg 8 mg/kg IV monthly	Risk of infection Infusion reaction LFT elevation Dyslipidemia Cytopenias		PPD skin test	CBC and LFTs at regular intervals

*Viral hepatitis panel: hepatitis B surface antigen, hepatitis C viral antibody.

Abbreviations: CBC, complete blood count; DMARDs, disease-modifying anti-rheumatic drugs; G6PD, glucose-6-phosphate dehydrogenase; IV, intravenous; LFTs, liver function tests; PPD, purified peptide derivative; SQ, subcutaneous; TB, tuberculosis.

NSAIDS

NSAIDs were formally viewed as the core of all other RA therapy, but they are now considered to be adjunctive therapy for management of symptoms uncontrolled by other measures. NSAIDs exhibit both analgesic and anti-inflammatory properties.

The anti-inflammatory effects of NSAIDs derive from their ability to nonselectively inhibit cyclooxygenase (COX)-1 and COX-2. Although the results of clinical trials suggest NSAIDs are roughly equivalent in their efficacy, experience suggests that some individuals may preferentially respond to a particular NSAID. Chronic use should be minimized due to the possibility of side effects, including gastritis and peptic ulcer disease as well as impairment of renal function.

GLUCOCORTICOIDS

Glucocorticoids may serve in several ways to control disease activity in RA. First, they may be administered in low-to-moderate doses to achieve rapid disease control before the onset of fully effective DMARD therapy, which often takes several weeks or even months. Second, a 1–2 week burst of glucocorticoids may be prescribed for the management of acute disease flares, with dose and duration guided by the severity of the exacerbation. Chronic administration of low doses (5–10 mg/d) of prednisone (or its equivalent) may also be warranted to control disease activity in patients with an inadequate response to DMARD therapy. Low-dose prednisone therapy has been shown in prospective studies to retard radiographic progression of joint disease; however, the benefits of this approach must be carefully weighed against the risks. Best practices minimize chronic use of low-dose prednisone therapy owing to the risk of osteoporosis and other long-term complications; however, the use of chronic prednisone therapy is unavoidable in many cases. Finally, if a patient exhibits one or a few actively inflamed joints, the clinician may consider intraarticular injection of an intermediate-acting glucocorticoid such as triamcinolone acetonide. This approach may allow for rapid control of inflammation in the setting of a limited number of affected joints. Caution must be exercised to appropriately exclude joint infection, as it often mimics an RA flare.

Osteoporosis ranks as an important long-term complication of chronic prednisone use. The ACR recommends primary prevention of glucocorticoid-induced osteoporosis with a bisphosphonate in any patient receiving 5 mg/d or more of prednisone for greater than 3 months. While prednisone use is known to increase the risk of peptic ulcer disease, especially with concomitant NSAID use, no evidence-based guidelines have been published regarding the use of gastrointestinal ulcer prophylaxis in this situation.

DMARDS

DMARDs are so named because of their ability to slow or prevent structural progression of RA. The conventional DMARDs include hydroxychloroquine, sulfasalazine, methotrexate, and leflunomide; they exhibit a delayed onset of action of approximately 6–12 weeks. Methotrexate is the DMARD of choice for the treatment of RA and is the anchor drug for most combination therapies. It was approved for the treatment of RA in 1986 and remains the benchmark for the efficacy and safety of new disease-modifying therapies. At the dosages used for the treatment of RA, methotrexate has been shown to stimulate adenosine release from cells, producing an anti-inflammatory effect. The clinical efficacy of leflunomide, an inhibitor of pyrimidine synthesis, appears similar to that of methotrexate; it has been shown in well-designed trials to be effective for the treatment of RA as monotherapy or in combination with methotrexate and other DMARDs.

Although similar to the other DMARDs in its slow onset of action, hydroxychloroquine has not been shown to delay radiographic progression of disease and thus is not considered to be a true DMARD. In clinical practice, hydroxychloroquine is generally used for treatment of early, mild disease or as adjunctive therapy in combination with other DMARDs. Sulfasalazine is utilized in a similar manner and has been shown in randomized, controlled trials to reduce radiographic progression of disease. Minocycline, gold salts, penicillamine, azathioprine, and cyclosporine have all been used for the treatment of RA with

varying degrees of success; however, they are used sparingly now due to their inconsistent clinical efficacy or unfavorable toxicity profile.

BIOLOGICALS

Biologic DMARDs have revolutionized the treatment of RA over the past decade (Table 321-2). They are protein therapeutics designed mostly to target cytokines and cell-surface molecules. The TNF inhibitors were the first biologicals approved for the treatment of RA. Anakinra, an IL-1 receptor antagonist, was approved shortly thereafter; however, its benefits have proved to be relatively modest compared with the other biologicals. Abatacept, rituximab, and tocilizumab are the newest members of this class.

Anti-TNF Agents

The development of TNF inhibitors was originally spurred by the experimental finding that TNF is a critical upstream mediator of joint inflammation. Currently, five agents that inhibit TNF-α are approved for the treatment of RA. There are three different anti-TNF monoclonal antibodies. Infliximab is a chimeric (part mouse and human) monoclonal antibody, while adalimumab and golimumab are humanized monoclonal antibodies. Certolizumab pegol is a pegylated Fc-free fragment of a humanized monoclonal antibody with binding specificity for TNF-α. Lastly, etanercept is a soluble fusion protein comprising the TNF receptor 2 in covalent linkage with the Fc portion of IgG1. All of the TNF inhibitors have been shown in randomized controlled clinical trials to reduce the signs and symptoms of RA, slow radiographic progression of joint damage, and improve physical function and quality of life. Anti-TNF drugs are typically used in combination with background methotrexate therapy. This combination regimen, which affords maximal benefit in many cases, is often the next step for treatment of patients with an inadequate response to methotrexate therapy. Etanercept, adalimumab, certolizumab pegol, and golimumab have also been approved for use as monotherapy.

Anti-TNF agents should be avoided in patients with active infection or a history of hypersensitivity to these agents. A major concern is the increased risk for infection, especially opportunistic fungal infection and reactivation of latent tuberculosis. For this reason, all patients are screened for latent tuberculosis according to national guidelines prior to starting anti-TNF therapy **(Chap. 165)**. In the United States, patients are skin tested using an intradermal injection of purified peptide derivative (PPD); individuals with skin reactions of more than 5 mm are presumed to have had previous exposure to TB and are evaluated for active disease and treated accordingly.

Anakinra

Anakinra, the recombinant form of the naturally occurring IL-1 receptor antagonist, has seen limited use for the treatment of RA owing to its modest clinical efficacy. However, anakinra has enjoyed a resurgence of late for the treatment of some rare syndromes dependent on IL-1 production, including neonatal-onset inflammatory disease, Muckle-Wells syndrome, and familial cold urticaria, systemic juvenile-onset inflammatory arthritis, and adult-onset Still's disease. Anakinra should not be combined with an anti-TNF drug due to the high rate of serious infections as observed with this regimen in a clinical trial.

Abatacept

Abatacept is a soluble fusion protein consisting of the extracellular domain of human cytotoxic T lymphocyte–associated antigen 4 (CTLA-4) linked to the modified portion of human IgG. It inhibits the co-stimulation of T cells by blocking CD28-CD80/86 interactions and may also inhibit the function of antigen-presenting cells by reverse signaling through CD80 and CD86. Abatacept has been shown in clinical trials to reduce disease activity, slow radiographic progression of damage, and improve functional disability. Most patients receive abatacept in combination with methotrexate or another DMARD such as leflunomide. Its onset of action is usually slower than that of the anti-TNF agents. Abatacept therapy has been associated with an increased risk of infection but is usually well tolerated otherwise.

Rituximab

Rituximab is a chimeric monoclonal antibody directed against CD20, a cell-surface molecule expressed by most mature B-lymphocytes. It works by depleting B cells, which in turn, leads to a reduction in the inflammatory response by unknown mechanisms. Rituximab has been approved for the treatment of refractory RA in combination with methotrexate and has been shown to be more effective for patients with seropositive than seronegative disease. Rituximab therapy has been associated with mild-to-moderate infusion reactions as well as an increased risk of infection. Notably, there have been isolated reports of a potentially lethal brain disorder, progressive multifocal leukoencephalopathy (PML), in association with rituximab therapy, although the absolute risk of this complication appears to be very low in patients with RA. Most of these cases have occurred on a background of previous or current exposure to other potent immunosuppressive drugs.

Tocilizumab

Tocilizumab is a humanized monoclonal antibody directed against the membrane and soluble forms of the IL-6 receptor. IL-6 is a proinflammatory cytokine implicated in the pathogenesis of RA, with detrimental effects on both joint inflammation and damage. IL-6 binding to its receptor activates intracellular signaling pathways that affect the acute-phase response, cytokine production, and osteoclast activation. Clinical trials have attested to the clinical efficacy of tocilizumab therapy for RA, both as monotherapy and in combination with methotrexate and other DMARDs. Tocilizumab has been associated with an increased risk of infection, neutropenia, and thrombocytopenia; however, the hematologic abnormalities appear to be reversible upon stopping the drug. In addition, this agent has been shown to increase LDL cholesterol; however, it is not known as yet if this effect on lipid levels increases the risk for development of atherosclerotic disease.

Approach to the Patient: Rheumatoid Arthritis

The original treatment pyramid for RA is now considered to be obsolete and has evolved into a new strategy that focuses on several goals: (1) early, aggressive therapy to prevent joint damage and disability; (2) frequent modification of therapy with utilization of combination therapy where appropriate; (3) individualization of therapy in an attempt to maximize response and minimize side effects; and (4) achieving, whenever possible, remission of clinical disease activity. A considerable amount of evidence supports this intensive treatment approach.

As mentioned earlier, methotrexate is the DMARD of first choice for initial treatment of moderate-to-severe RA. Failure to achieve adequate improvement with methotrexate therapy calls for consideration of an effective combination regimen. Effective combinations include: methotrexate, sulfasalazine, and hydroxychloroquine (triple therapy); methotrexate and leflunomide; and methotrexate plus a biological. The combination of methotrexate and an anti-TNF agent, for example, has been shown in randomized, controlled trials to be superior to methotrexate alone not only for reducing signs and symptoms of disease, but also for retarding the progression of structural joint damage. The caveat of these studies, however, is that the protection against structural damage afforded by combining an anti-TNF drug with methotrexate appears to be restricted to a subset of patients with a high risk for disease progression. This subset corresponds to approximately 25% of patients enrolled in clinical trials. The remaining patients do not show significant progression in joint damage over 12 months while receiving methotrexate alone. Predicting which patients will ultimately show radiologic joint damage is imprecise at best, although some factors such as an elevated serum level of acute-phase reactants, high burden of joint inflammation, and the presence of

erosive disease are associated with increased likelihood of developing structural injury.

Some patients may not respond to an anti-TNF drug or be intolerant of its side effects. Initial responders to an anti-TNF agent that later worsen may benefit from switching to another anti-TNF agent. Other biologicals, such as abatacept and rituximab, may also be considered for the treatment of patients whose disease is refractory to anti-TNF therapy. The addition of abatacept or rituximab to background methotrexate therapy has been shown in well-designed clinical trials to be effective for reducing the signs and symptoms of joint inflammation and slowing radiographic progression of disease. Early in the treatment course, abatacept may also be considered in lieu of an anti-TNF drug depending on the clinical circumstances (e.g., relative contraindication for the use of an anti-TNF agent).

A clinical state defined as low disease activity or remission is the optimal goal of therapy, although most patients never achieve remission despite every effort to achieve it. Composite indices, such as the Disease Activity Score -28 or DAS28, are useful for classifying states of low disease activity and remission; however, they are imperfect tools due to the limitations of the clinical joint examination in which low-grade synovitis may escape detection. Complete remission has been stringently defined as the total absence of all articular and extraarticular inflammation and immunologic activity related to RA. However, evidence for this state can be difficult to demonstrate in clinical practice. In an effort to standardize and simplify the definition of remission for clinical trials, the ACR and EULAR developed two provisional operational definitions of remission in RA **(Table 321-3)**. A patient may be considered in remission if he or she 1) meets all of the clinical and laboratory criteria listed in Table 321-3 or 2) has a composite Simplified Disease Activity Index (SDAI) score of <3.3. The SDAI is calculated by taking the sum of a tender joint and swollen joint count (using 28 joints), patient global assessment (0–10 scale), physician global assessment (0–10 scale), and C-reactive protein (in mg/dL). This definition of remission does not take into account the possibility of subclinical synovitis or that damage alone may produce a tender or swollen joint. Ignoring the semantics of these definitions, the aforementioned remission criteria are nonetheless useful for setting a level of disease control that will likely result in minimal or no progression of structural damage and disability.

Table 321-3 Acr/Eular Provisional Definition of Remission in Rheumatoid Arthritis

At any time point, patient must satisfy all of the following:
Tender joint count ≤ 1
Swollen joint count ≤ 1
C-reactive protein ≤ 1 mg/dL
Patient global assessment ≤ 1 (on a 0–10 scale)
OR
At any time point, patient must have a Simplified Disease Activity Index score of ≤ 3.3

Source: Adapted from Felson et al.

PHYSICAL THERAPY AND ASSISTIVE DEVICES

All patients should receive a prescription for exercise and physical activity. Dynamic strength training, community-based comprehensive physical therapy, and physical-activity coaching (emphasizing 30 minutes of moderately intensive activity most days a week) have all been shown to improve muscle strength and perceived health status. Foot orthotics for painful valgus deformity decreases foot pain and resulting disability and functional limitations. Judicious use of wrist splints can also decrease pain; however, their benefits may be offset by decreased dexterity and a variable effect on grip strength.

SURGERY

Surgical procedures may improve pain and disability in RA—most notably the hands, wrists, and feet, typically after the failure of medical therapy with varying degrees of reported long-term success. For large joints, such as the knee, hip, shoulder, or elbow, total joint arthroplasty is an option for advanced joint disease. A few surgical options exist for dealing with the smaller hand joints. Silicone implants are the most common prosthetic for MCP arthroplasty, and are generally implanted in patients with severe decreased arc of motion, marked flexion contractures, MCP joint pain with radiographic abnormalities and severe ulnar drift. Synovectomy and limited fusion are offered for the early rheumatoid wrist, but they are used much less frequently now compared to the past because of the availability of improved DMARD therapies. Arthrodesis and total wrist arthroplasty are reserved for patients with severe disease that have substantial pain and functional impairment. These two procedures appear to have equal efficacy in terms of pain control and patient satisfaction. Numerous surgical options exist for correction of hallux valgus in the forefoot, including arthrodesis and arthroplasty, as well as primarily arthrodesis for refractory hindfoot pain.

OTHER MANAGEMENT CONSIDERATIONS

Pregnancy

Up to 75% of female RA patients will note overall improvement in symptoms during pregnancy, but often will flare post-delivery. Flares during pregnancy are generally treated with low doses of prednisone; hydroxychloroquine and sulfasalazine are probably the safest DMARDs to use during pregnancy. Methotrexate and leflunomide therapy are contraindicated during pregnancy due to their teratogenicity in animals and humans. The experience with biologic agents has been insufficient to make specific recommendations for their use during pregnancy. Most rheumatologists avoid their use in this setting; however, exceptions are considered depending on the circumstances.

Elderly Patients

RA presents in up to one-third of patients after the age of 60; however, it has been recognized that older individuals receive less aggressive treatment due to concerns about increased risks of drug toxicity. Studies suggest that conventional DMARDs as well as biologic agents are equally effective and safe in younger and older patients. Due to comorbidities, many elderly patients have an increased risk of infection. Aging also leads to a gradual decline in renal function that may raise the risk for side effects from NSAIDs and some DMARDS, such as methotrexate. Renal function must be taken into consideration before prescribing methotrexate, which is mostly cleared by the kidneys. To reduce the risks of side effects, methotrexate doses may need to be adjusted downward for the drop in renal function that usually comes with the seventh and eighth decades of life. Methotrexate is usually not prescribed for patients with a serum creatinine greater than 2 mg/dl.

GLOBAL CHALLENGES

Developing countries are finding an increase in the incidence of noncommunicable, chronic diseases such as diabetes, cardiovascular disease, and RA in the face of ongoing poverty, rampant infectious disease, and poor access to modern health care facilities. In these areas, patients tend to have a greater delay in diagnosis and limited access to specialists, and thus

greater disease activity and disability at presentation. In addition, infection risk remains a significant issue for the treatment of RA in developing countries because of the immunosuppression associated with the use of glucocorticoids and most DMARDs. For example, in some developing countries, patients undergoing treatment for RA have a substantial increase in the incidence of tuberculosis, which demands the implementation of far more comprehensive screening practices and liberal use of isoniazid prophylaxis than in developed countries. The increased prevalence of hepatitis B and C, as well as human immunodeficiency virus (HIV), in these developing countries also poses challenges. Reactivation of viral hepatitis has been observed in association with some of the DMARDs, such as rituximab. Also, reduced access to antiretroviral therapy may limit the control of HIV infection and therefore the choice of DMARD therapies.

Despite these challenges, one should attempt to implement early treatment of RA in the developing countries with the resources at hand. Sulfasalazine and methotrexate are all reasonably accessible throughout the world where they can be utilized as both monotherapy and in combination with other drugs. The use of biologic agents is increasing in the developed countries as well as in other areas around the world, although their use is limited by high cost; national protocols restrict their use, and concerns remain about the risk for opportunistic infections.

SUMMARY

Improved understanding of the pathogenesis of RA and its treatment has dramatically revolutionized the management of this disease. The outcomes of patients with RA are vastly superior to those of the prebiologic modifier era; more patients than in years past are able to avoid significant disability and continue working, albeit with some job modifications in many cases. The need for early and aggressive treatment of RA as well as frequent follow-up visits for monitoring of drug therapy has implications for our health care system. Primary care physicians and rheumatologists must be prepared to work together as a team to reach the ambitious goals of best practice. In many settings, rheumatologists have reengineered their practice in a way that places high priority on consultations for any new patient with early inflammatory arthritis.

The therapeutic regimens for RA are becoming increasingly complex with the rapidly expanding therapeutic armamentarium. Patients receiving these therapies must be carefully monitored by both the primary care physician and the rheumatologist to minimize the risk of side effects and identify quickly any complications of chronic immunosuppression. Also, prevention and treatment of RA-associated conditions such as ischemic heart disease and osteoporosis will likely benefit from a team approach owing to the value of multidisciplinary care.

Research will continue to search for new therapies with superior efficacy and safety profiles and investigate treatment strategies that can bring the disease under control more rapidly and nearer to remission. However, prevention and cure of RA will likely require new breakthroughs in our understanding of disease pathogenesis. These insights may come from genetic studies illuminating critical pathways in the mechanisms of joint inflammation. Equally ambitious is the lofty goal of biomarker discovery that will open the door to personalized medicine for the care of patients with RA.

FURTHER READINGS

Felson DT et al: American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum 63:573, 2011[PMID: 21294106] [Full Text]

Goldring MB, Marcu KB: Cartilage homeostasis in health and rheumatic diseases. Arthritis Res Ther 11:224, 2009[PMID: 19519926] [Full Text]

Goldring SR, Goldring MB: Eating bone or adding it: The Wnt pathway decides. Nat Med 13:133, 2007[PMID: 17290270] [Full Text]

Goronzy JJ, Weyand CM: Developments in the scientific understanding of rheumatoid arthritis. Arthritis Res Ther 11:249, 2009[PMID: 19835638] [Full Text]

Gravallese EM et al: Identification of cell types responsible for bone resorption in rheumatoid arthritis and juvenile rheumatoid arthritis. Am J Pathol 152:943, 1998[PMID: 9546355] [Full Text]

Holers VM: Antibodies to citrullinated proteins: pathogenic and diagnostic significance. Curr Rheumatol Rep 9:396, 2007[PMID: 17915095] [Full Text]

Mcinnes IB, Schett G: Cytokines in the pathogenesis of rheumatoid arthritis. Nat Rev Immunol 7: 429, 2007[PMID: 17525752] [Full Text]

Michou L et al: Associations between genetic factors, tobacco smoking and autoantibodies in familial and sporadic rheumatoid arthritis. Ann Rheum Dis 67:466, 2008[PMID: 17660221] [Full Text]

Mody GM, Cardiel MH: Challenges in the management of rheumatoid arthritis in developing countries. Best Pract Res Clin Rheumatol 22:621, 2008[PMID: 18783741] [Full Text]

Myasoedova E et al: Epidemiology of rheumatiod arthritis: Rheumatoid arthritis and mortality. Curr Rheumatol Rep 12:379, 2010[PMID: 20645137] [Full Text]

Neogi T et al: 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 62:2569, 2010[PMID: 20872595] [Full Text]

Plenge RM: Rheumatoid arthritis genetics: 2009 update. Curr Rheumatol Rep 11:351, 2009[PMID: 19772830] [Full Text]

Scott DL, Kingsley GH: Tumor necrosis factor inhibitors for rheumatoid arthritis. N Engl J Med 355:704, 2006[PMID: 16914706] [Full Text]

———, Steer S. The course of established rheumatoid arthritis. Best Pract Res Clin Rheumatol 21:943, 2007

Smolen JS, et al: New therapies for treatment of rheumatoid arthritis. Lancet 370:1861, 2007[PMID: 17570481] [Full Text]

Turesson C et al: Extra-articular disease manifestations in rheumatoid arthritis: incidence trends and risk factors over 46 years. Ann Rheum Dis 62:722, 2003[PMID: 12860726] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACUTE RHEUMATIC FEVER: INTRODUCTION

Acute rheumatic fever (ARF) is a multisystem disease resulting from an autoimmune reaction to infection with group A streptococcus. Although many parts of the body may be affected, almost all of the manifestations resolve completely. The exception is cardiac valvular damage [rheumatic heart disease (RHD)], which may persist after the other features have disappeared.

GLOBAL CONSIDERATIONS

ARF and RHD are diseases of poverty. They were common in all countries until the early twentieth century, when their incidence began to decline in industrialized nations. This decline was largely attributable to improved living conditions—particularly less crowded housing and better hygiene—which resulted in reduced transmission of group A streptococci. The introduction of antibiotics and improved systems of medical care had a supplemental effect. Recurrent outbreaks of ARF began in the 1980s in the Rocky Mountain states of the United States, where elevated rates persist.

The virtual disappearance of ARF and reduction in the incidence of RHD in industrialized countries during the twentieth century unfortunately was not replicated in developing countries, where these diseases continue unabated. RHD is the most common cause of heart disease in children in developing countries and is a major cause of mortality and morbidity in adults as well. It has been estimated that between 15 and 19 million people worldwide are affected by RHD, with approximately one-quarter of a million deaths occurring each year. Some 95% of ARF cases and RHD deaths now occur in developing countries.

Although ARF and RHD are relatively common in all developing countries, they occur at particularly elevated rates in certain regions. These "hot spots" are sub-Saharan Africa, Pacific nations, Australasia, and the Indian subcontinent **(Fig. 322-1)**. Unfortunately, most developing countries do not currently have coordinated, register-based RHD control programs, which are proven to be cost-effective in reducing the burden of RHD. Enhancing awareness of RHD and mobilizing resources for its control in developing countries is an issue requiring international attention.

Figure 322-1

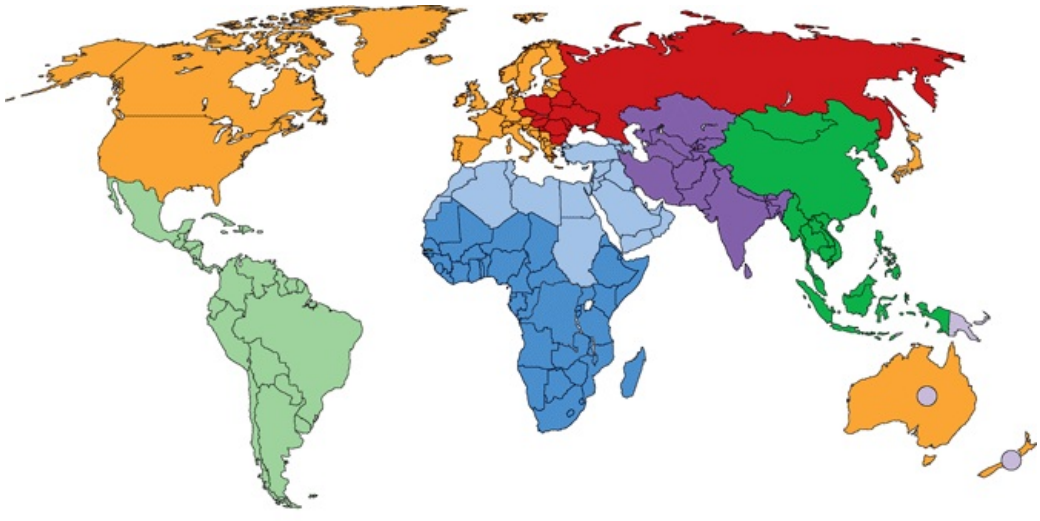

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prevalence of rheumatic heart disease in children aged 5–14 years. Circles within Australia and New Zealand represent indigenous populations, and also Pacific Islanders in New Zealand. *(From JR*

New Zealand represent indigenous populations, and also Pacific Islanders in New Zealand. (From A Carapetis et al: Lancet Infect Dis. Copyright 2005, with permission from Elsevier.)

EPIDEMIOLOGY

ARF is mainly a disease of children aged 5–14 years. Initial episodes become less common in older adolescents and young adults and are rare in persons aged >30 years. By contrast, recurrent episodes of ARF remain relatively common in adolescents and young adults. This pattern contrasts with the prevalence of RHD, which peaks between 25 and 40 years. There is no clear gender association for ARF, but RHD more commonly affects females, sometimes up to twice as frequently as males.

PATHOGENESIS

Organism Factors

Based on currently available evidence, ARF is exclusively caused by infection of the upper respiratory tract with group A streptococci **(see Chap. 136)**. Although classically, certain M-serotypes (particularly types 1, 3, 5, 6, 14, 18, 19, 24, 27, and 29) were associated with ARF, in high-incidence regions, it is now thought that any strain of group A streptococcus has the potential to cause ARF. Potential role of skin infection and of groups C and G streptococci are currently being investigated.

Host Factors

Approximately 3–6% of any population may be susceptible to ARF, and this proportion does not vary dramatically between populations. Findings of familial clustering of cases and concordance in monozygotic twins—particularly for chorea—confirm that susceptibility to ARF is an inherited characteristic. Particular human leukocyte antigen (HLA) class II alleles appear to be strongly associated with susceptibility. Associations have also been described with high levels of circulating mannose-binding lectin and polymorphisms of transforming growth factor β_1 gene and immunoglobulin genes. High-level expression of a particular alloantigen present on B cells, D8-17, has been found in patients with a history of ARF in many populations, with intermediate-level expression in first-degree family members, suggesting that this may be a marker of inherited susceptibility.

The Immune Response

When a susceptible host encounters a group A streptococcus, an autoimmune reaction results, which leads to damage to human tissues as a result of cross-reactivity between epitopes on the organism and the host **(Fig. 322-2)**. Cross-reactive epitopes are present in the streptococcal M protein and the *N*-acetylglucosamine of group A streptococcal carbohydrate and are immunologically similar to molecules in human myosin, tropomyosin, keratin, actin, laminin, vimentin, and *N*-acetylglucosamine. It is currently thought that the initial damage is due to cross-reactive antibodies attaching at the cardiac valve endothelium, allowing the entry of primed CD4+ T cells, leading to subsequent T cell-mediated inflammation.

Figure 322-2

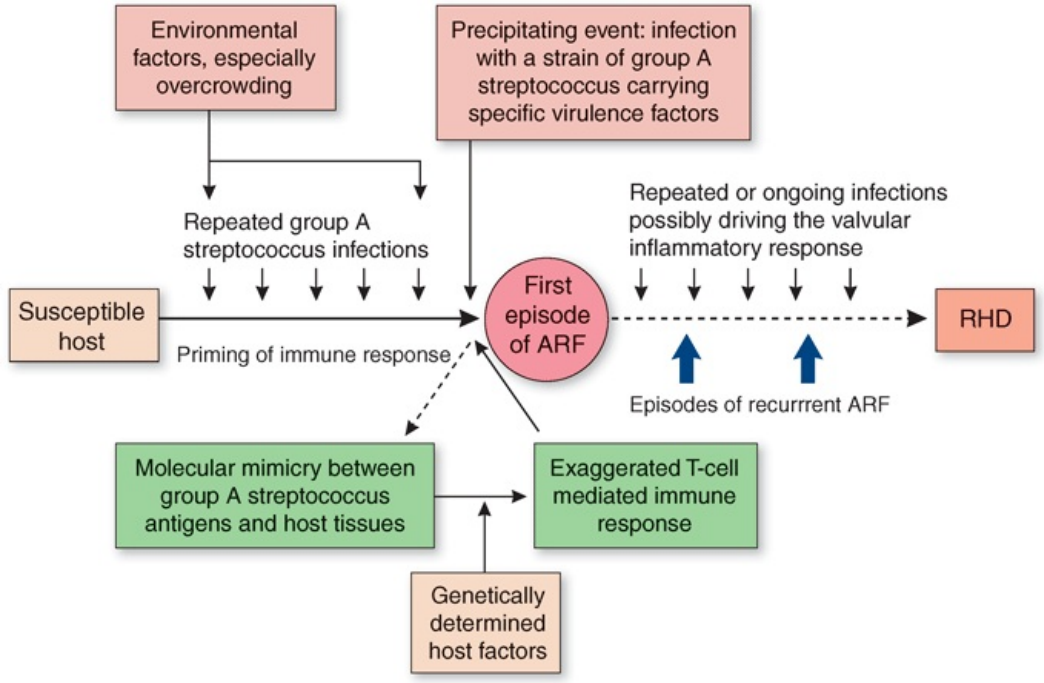

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathogenetic pathway for acute rheumatic fever and rheumatic heart disease.*(From JR Carapetis et al: Lancet 366:155, 2005. Copyright 2005, with permission from Elsevier.)*

CLINICAL FEATURES

There is a latent period of ~3 weeks (1–5 weeks) between the precipitating group A streptococcal infection and the appearance of the clinical features of ARF. The exceptions are chorea and indolent carditis, which may follow prolonged latent periods lasting up to 6 months. Although many patients report a prior sore throat, the preceding group A streptococcal infection is commonly subclinical; in these cases it can only be confirmed using streptococcal antibody testing. The most common clinical presentation of ARF is polyarthritis and fever. Polyarthritis is present in 60–75% of cases and carditis in 50–60%. The prevalence of chorea in ARF varies substantially between populations, ranging from <2% to 30%. Erythema marginatum and subcutaneous nodules are now rare, being found in <5% of cases.

Heart Involvement

Up to 60% of patients with ARF progress to RHD. The endocardium, pericardium, or myocardium may be affected. Valvular damage is the hallmark of rheumatic carditis. The mitral valve is almost always affected, sometimes together with the aortic valve; isolated aortic valve involvement is rare. Early valvular damage leads to regurgitation. Over ensuing years, usually as a result of recurrent episodes, leaflet thickening, scarring, calcification, and valvular stenosis may develop **(Fig. 322-3). See Videos 322-1A-D and 322-2A-B**. Therefore the characteristic manifestation of carditis in previously unaffected individuals is mitral regurgitation, sometimes accompanied by aortic regurgitation. Myocardial inflammation may affect electrical conduction pathways, leading to P-R interval prolongation (first-degree AV block or rarely higher-level block) and softening of the first heart sound.

Figure 322-3

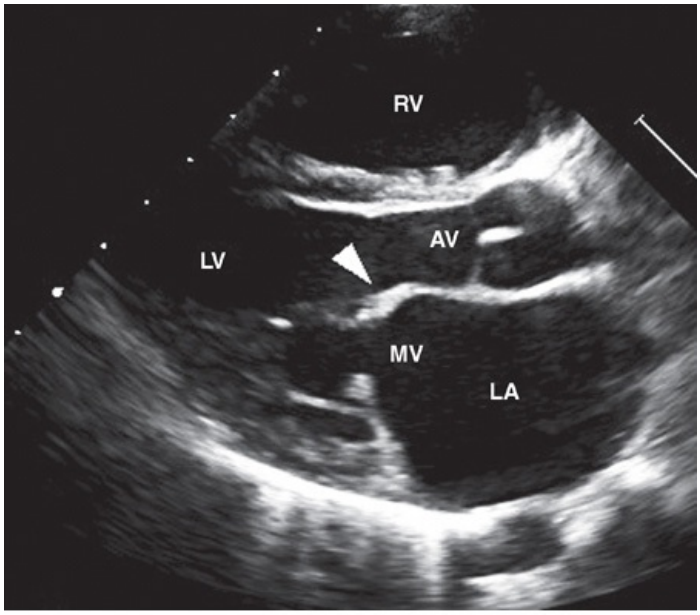

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transthoracic echocardiographic image from a 5-yearold boy with chronic rheumatic heart disease. This diastolic image demonstrates leaflet thickening, restriction of the anterior mitral valve leaflet tip, and doming of the body of the leaflet toward the interventricular septum. This appearance (marked by the arrowhead) is commonly described as a "hockey stick" or an "elbow" deformity. AV, aortic valve; LA, left atrium; LV, left ventricle; MV, mitral valve; RV, right ventricle. *(Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congential Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)*

Video 322-1A

Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical postero-laterally directed regurgitant jet of rheumatic carditis. ***A.*** Acute rheumatic carditis (apical 4-chamber view echocardiogram). (Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)

Video 322-1B

Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical postero-laterally directed regurgitant jet of rheumatic carditis. ***B.*** Acute rheumatic carditis (apical 4-chamber view colour-Doppler echocardiogram). (Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)

Video 322-1C

Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical postero-laterally directed regurgitant jet of rheumatic carditis. ***C.*** Acute rheumatic carditis (parasternal long-axis view echocardiogram). (Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)

Video 322-1D

Transthoracic echocardiographic images of a 9-year-old girl with first episode of acute rheumatic fever. Images demonstrate the typical echocardiographic findings of acute rheumatic carditis. The valve leaflets are relatively thin and highly mobile. The failure of coaptation of the mitral valve leaflets is the result of chordal elongation and annular dilatation. The mitral valve regurgitation is moderate with a typical postero-laterally directed regurgitant jet of rheumatic carditis. ***D.*** Acute rheumatic carditis (parasternal long-axis view colour-Doppler echocardiogram). (Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)

Video 322-2A

Transthoracic echocardiographic images are from a 5-year-old boy with chronic rheumatic heart disease with severe mitral valve regurgitation and moderate mitral valve stenosis. Images demonstrate the typical echocardiographic findings in advanced chronic rheumatic heart disease. Both the anterior and posterior mitral valve leaflets are markedly thickened. During diastole, the motion of the anterior mitral valve leaflet tip is restricted with doming of the body of the leaflet towards the interventricular septum. This appearance is commonly described as a "hockey stick" or an "elbow" deformity. ***A.*** Chronic rheumatic heart disease (parasternal long-axis view). (Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)

Video 322-2B

Transthoracic echocardiographic images are from a 5-year-old boy with chronic rheumatic heart disease with severe mitral valve regurgitation and moderate mitral valve stenosis. Images demonstrate the typical echocardiographic findings in advanced chronic rheumatic heart disease. Both the anterior and posterior mitral valve leaflets are markedly thickened. During diastole, the motion of the anterior mitral valve leaflet tip is restricted with doming of the body of the leaflet towards the interventricular septum. This appearance is commonly described as a "hockey stick" or an "elbow" deformity. ***B.*** Chronic rheumatic heart disease (apical 2-chamber view echocardiogram). (Courtesy of Dr. Bo Remenyi, Department of Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand.)

Joint Involvement

To qualify as a major manifestation, joint involvement in ARF must be arthritic, i.e., objective evidence of inflammation, with hot, swollen, red and/or tender joints, and involvement of more than one joint (i.e., polyarthritis). The typical arthritis is migratory, moving from one joint to another over a period of hours. ARF almost always affects the large joints—most commonly the knees, ankles, hips, and elbows—and is asymmetric. The pain is severe and usually disabling until anti-inflammatory medication is commenced.

Less severe joint involvement is also relatively common but qualifies only as a minor manifestation. Arthralgia without objective joint inflammation usually affects large joints in the same migratory pattern as polyarthritis. In some populations, aseptic monoarthritis may be a presenting feature of ARF. This may occur because of early commencement of anti-inflammatory medication before the typical migratory pattern is established.

The joint manifestations of ARF are highly responsive to salicylates and other nonsteroidal anti-inflammatory drugs (NSAIDs). Indeed, joint involvement that persists more than 1 or 2 days after starting salicylates is unlikely to be due to ARF. Conversely, if salicylates are commenced early in the illness, before fever and migratory polyarthritis have become manifest, it may be difficult to make a diagnosis of ARF. For this reason, salicylates and other NSAIDs should be withheld—and pain managed with acetaminophen or codeine—until the diagnosis is confirmed.

Chorea

Sydenham's chorea commonly occurs in the absence of other manifestations, follows a prolonged latent period after group A streptococcal infection, and is found mainly in females. The choreiform movements affect particularly the head (causing characteristic darting movements of the tongue) and the upper limbs **(Chap. 371)**. They may be generalized or restricted to one side of the body (hemi-chorea). The chorea varies in severity. In mild cases it may be evident only on careful examination, while in the most severe cases the affected individuals are unable to perform activities of daily living and are at risk of injuring themselves. Chorea eventually resolves completely, usually within 6 weeks.

Skin Manifestations

The classic rash of ARF is *erythema marginatum***(Chap. 17)**, which begins as pink macules that clear centrally, leaving a serpiginous, spreading edge. The rash is evanescent, appearing and disappearing before the examiner's eyes. It occurs usually on the trunk, sometimes on the limbs, but almost never on the face.

Subcutaneous nodules occur as painless, small (0.5–2 cm), mobile lumps beneath the skin overlying bony prominences, particularly of the hands, feet, elbows, occiput, and occasionally the vertebrae. They are a delayed manifestation, appearing 2–3 weeks after the onset of disease, last for just a few days up to 3 weeks, and are commonly associated with carditis.

Other Features

Fever occurs in most cases of ARF, although rarely in cases of pure chorea. Although high-grade fever (≳39°C) is the rule, lower grade temperature elevations are not uncommon. Elevated acute-phase reactants are also present in most cases. C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) are often dramatically elevated. Occasionally the peripheral leukocyte count is mildly elevated.

Evidence of a Preceding Group a Streptococcal Infection

With the exception of chorea and low-grade carditis, both of which may become manifest many months later, evidence of a preceding group A streptococcal infection is essential in making the diagnosis of ARF. As most cases do not have a positive throat swab culture or rapid antigen test, serologic evidence is usually needed. The most common serologic tests are the anti-streptolysin O (ASO) and anti-DNase B (ADB) titers. Where possible, age-specific reference ranges should be determined in a local population of healthy people without a recent group A streptococcal infection.

Other Post-Streptococcal Syndromes that May Be Confused with Rheumatic Fever

Post-streptococcal reactive arthritis (PSRA) is differentiated from ARF on the basis of: (1) small-joint involvement that is often symmetric; (2) a short latent period following streptococcal infection (usually <1 week); (3) occasional causation by nongroup A β-hemolytic streptococcal infection; (4) slower responsiveness to salicylates; and (5) the absence of other features of ARF, particularly carditis.

Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infection (PANDAS) is a term that links a range of tic disorders and obsessive-compulsive symptoms with group A streptococcal infections. People with PANDAS are said not to be at risk of carditis, unlike patients with Sydenham's chorea. The diagnoses of PANDAS and PSRA should rarely be made in populations with a high incidence of ARF.

Confirming the Diagnosis

Because there is no definitive test, the diagnosis of ARF relies on the presence of a combination of typical clinical features together with evidence of the precipitating group A streptococcal infection, and the exclusion of other diagnoses. This uncertainty led Dr. T. Duckett Jones in 1944 to develop a set of criteria (subsequently known as the *Jones criteria*) to aid in the diagnosis. An expert panel convened by the World Health Organization (WHO) clarified the use of the Jones criteria in ARF recurrences **(Table 322-1)**. Because each revision of the Jones criteria since 1944 has reduced sensitivity and increased specificity, in response to the decline in incidence of ARF in high-income countries, there is now concern that they may be too insensitive for countries where ARF incidence remains high. As a result, some countries (e.g., Australia and New Zealand) have developed their own, more sensitive, diagnostic criteria for ARF in their populations (links available at the *RHDnet* websitewww.worldheart.org/rhd).

Table 322-1 2002–2003 World Health Organization Criteria for the Diagnosis of Rheumatic Fever and Rheumatic Heart Disease (Based on the 1992 Revised Jones Criteria)

Diagnostic Categories	Criteria
Primary episode of rheumatic fever[a]	Two major or one major and two minor manifestations plus evidence of preceding group A streptococcal infection
Recurrent attack of rheumatic fever in a patient without established rheumatic heart disease	Two major or one major and two minor manifestations plus evidence of preceding group A streptococcal infection
Recurrent attack of rheumatic fever in a patient with established rheumatic heart disease[b]	Two minor manifestations plus evidence of preceding group A streptococcal infection[c]
Rheumatic chorea Insidious onset rheumatic carditis[b]	Other major manifestations or evidence of group A streptococcal infection not required
Chronic valve lesions of rheumatic heart disease (patients presenting for the first time with pure mitral stenosis or mixed mitral valve disease and/or aortic valve disease)[d]	Do not require any other criteria to be diagnosed as having rheumatic heart disease
Major manifestations	Carditis Polyarthritis Chorea Erythema marginatum Subcutaneous nodules
Minor manifestations	Clinical: fever, polyarthralgia Laboratory: elevated erythrocyte sedimentation rate or leukocyte count[e]

	Electrocardiogram: prolonged P-R interval
Supporting evidence of a preceding streptococcal infection within the last 45 days	Elevated or rising anti-streptolysin O or other streptococcal antibody, *or* A positive throat culture, *or* Rapid antigen test for group A streptococcus, *or* Recent scarlet fever[e]

[a]Patients may present with polyarthritis (or with only polyarthralgia or monoarthritis) and with several (three or more) other minor manifestations, together with evidence of recent group A streptococcal infection. Some of these cases may later turn out to be rheumatic fever. It is prudent to consider them as cases of "probable rheumatic fever" (once other diagnoses are excluded) and advise regular secondary prophylaxis. Such patients require close follow-up and regular examination of the heart. This cautious approach is particularly suitable for patients in vulnerable age groups in high incidence settings.

[b]Infective endocarditis should be excluded.

[c]Some patients with recurrent attacks may not fulfill these criteria.

[d]Congenital heart disease should be excluded.

[e]1992 Revised Jones criteria do not include elevated leukocyte count as a laboratory minor manifestation (but do include elevated C-reactive protein), and do not include recent scarlet fever as supporting evidence of a recent streptococcal infection.

Source: Reprinted with permission from WHO Expert Consultation on Rheumatic Fever and Rheumatic Heart Disease (2001: Geneva, Switzerland): *Rheumatic Fever and Rheumatic Heart Disease: Report of a WHO Expert Consultation* (WHO Tech Rep Ser, 923). Geneva, World Health Organization, 2004.

Treatment: Acute Rheumatic Fever

Patients with possible ARF should be followed closely to ensure that the diagnosis is confirmed, treatment of heart failure and other symptoms is undertaken, and preventive measures including commencement of secondary prophylaxis, inclusion on an ARF registry, and health education are commenced. Echocardiography should be performed on all possible cases to aid in making the diagnosis and to determine the severity at baseline of any carditis. Other tests that should be performed are listed in **Table 322-2**.

Table 322-2 Recommended Tests in Cases of Possible Acute Rheumatic Fever

Recommended for all cases
White blood cell count
Erythrocyte sedimentation rate
C-reactive protein
Blood cultures if febrile
Electrocardiogram (repeat in 2 weeks and 2 months if prolonged P-R interval or other rhythm abnormality)
Chest x-ray if clinical or echocardiographic evidence of carditis
Echocardiogram (consider repeating after 1 month if negative)
Throat swab (preferably before giving antibiotics)–culture for group A streptococcus
Anti-streptococcal serology: both anti-streptolysin O and anti-DNase B titres, if available (repeat 10–14 days later if 1st test not confirmatory)
Tests for alternative diagnoses, depending on clinical features
Repeated blood cultures if possible endocarditis
Joint aspirate (microscopy and culture) for possible septic arthritis
Copper, ceruloplasmin, anti-nuclear antibody, drug screen for choreiform movements
Serology and auto-immune markers for arboviral, auto-immune or reactive arthritis

Source: Reprinted with permission from National Heart Foundation of Australia.

There is no treatment for ARF that has been proven to alter the likelihood of developing, or the severity

of, RHD. With the exception of treatment of heart failure, which may be life-saving in cases of severe carditis, the treatment of ARF is symptomatic.

ANTIBIOTICS

All patients with ARF should receive antibiotics sufficient to treat the precipitating group A streptococcal infection **(Chap. 136)**. Penicillin is the drug of choice and can be given orally [as phenoxymethyl penicillin, 500 mg (250 mg for children ≤27 kg) PO twice daily, or amoxicillin 50 mg/kg (max 1 g) daily, for 10 days] or as a single dose of 1.2 million units (600,000 units for children ≤27 kg) IM benzathine penicillin G.

SALICYLATES AND NSAIDS

These may be used for the treatment of arthritis, arthralgia, and fever, once the diagnosis is confirmed. They are of no proven value in the treatment of carditis or chorea. Aspirin is the drug of choice. An initial dose of 80–100 mg/kg per day in children (4–8 g/d in adults) in 4–5 divided doses is often needed for the first few days up to 2 weeks. A lower dose should be used if symptoms of salicylate toxicity emerge, such as nausea, vomiting, or tinnitus. When the acute symptoms are substantially resolved, the dose can be reduced to 60–70 mg/kg per day for a further 2–4 weeks. Fever, joint manifestations, and elevated acute-phase reactants sometimes recur up to 3 weeks after the medication is discontinued. This does not indicate a recurrence and can be managed by recommencing salicylates for a brief period. Although less well studied, naproxen at a dose of 10–20 mg/kg per day has been reported to lead to good symptomatic response.

CONGESTIVE HEART FAILURE

Glucocorticoids

The use of glucocorticoids in ARF remains controversial. Two meta-analyses have failed to demonstrate a benefit of glucocorticoids compared to placebo or salicylates in improving the short- or longer term outcome of carditis. However, the studies included in these meta-analyses all took place >40 years ago and did not use medications in common usage today. Many clinicians treat cases of severe carditis (causing heart failure) with glucocorticoids in the belief that they may reduce the acute inflammation and result in more rapid resolution of failure. However, the potential benefits of this treatment should be balanced against the possible adverse effects, including gastrointestinal bleeding and fluid retention. If used, prednisone or prednisolone are recommended at doses of 1–2 mg/kg per day (maximum, 80 mg). Glucocorticoids are often only required for a few days or up to a maximum of 3 weeks.

MANAGEMENT OF HEART FAILURE

See Chap. 234.

BED REST

Traditional recommendations for long-term bed rest, once the cornerstone of management, are no longer widely practiced. Instead, bed rest should be prescribed as needed while arthritis and arthralgia are present, and for patients with heart failure. Once symptoms are well controlled, gradual mobilization can commence as tolerated.

CHOREA

Medications to control the abnormal movements do not alter the duration or outcome of chorea. Milder cases can usually be managed by providing a calm environment. In patients with severe chorea, carbamazepine or sodium valproate are preferred to haloperidol. A response may not be seen for 1–2 weeks, and a successful response may only be to reduce rather than resolve the abnormal movements. Medication should be continued for 1–2 weeks after symptoms subside.

INTRAVENOUS IMMUNOGLOBULIN (IVIG)

Small studies have suggested that IVIg may lead to more rapid resolution of chorea but has shown no benefit on the short- or long-term outcome of carditis in ARF without chorea. In the absence of better data, IVIg is *not* recommended except in cases of severe chorea refractory to other treatments.

PROGNOSIS

Untreated, ARF lasts on average 12 weeks. With treatment, patients are usually discharged from hospital within 1–2 weeks. Inflammatory markers should be monitored every 1–2 weeks until they have normalized (usually within 4–6 weeks), and an echocardiogram should be performed after 1 month to determine if there has been progression of carditis. Cases with more severe carditis need close clinical and echocardiographic monitoring in the longer term.

Once the acute episode has resolved, the priority in management is to ensure long-term clinical follow-up and adherence to a regimen of secondary prophylaxis. Patients should be entered onto the local ARF registry (if present) and contact made with primary care practitioners to ensure a plan for follow-up and administration of secondary prophylaxis before the patient is discharged. Patients and their families should also be educated about their disease, emphasizing the importance of adherence to

secondary prophylaxis. If carditis is present, they should also be informed of the need for antibiotic prophylaxis against endocarditis for dental and surgical procedures.

PREVENTION

Primary Prevention

Ideally, primary prevention would entail elimination of the major risk factors for streptococcal infection, particularly overcrowded housing. This is difficult to achieve in most places where ARF is common.

Therefore, the mainstay of primary prevention for ARF remains primary prophylaxis (i.e., the timely and complete treatment of group A streptococcal sore throat with antibiotics). If commenced within 9 days of sore throat onset, a course of penicillin (as outlined above for treatment of ARF) will prevent almost all cases of ARF that would otherwise have developed. This important strategy relies on individuals presenting for medical care when they have a sore throat, the availability of trained health and microbiology staff along with the materials and infrastructure to take throat swabs, and a reliable supply of penicillin. Unfortunately, many of these elements are not available in developing countries.

Secondary Prevention

The mainstay of controlling ARF and RHD is secondary prevention. Because patients with ARF are at dramatically higher risk than the general population of developing a further episode of ARF after a group A streptococcal infection, they should receive long-term penicillin prophylaxis to prevent recurrences. The best antibiotic for secondary prophylaxis is benzathine penicillin G (1.2 million units, or 600,000 units if ≤27 kg) delivered every 4 weeks. It can be given every 3 weeks, or even every 2 weeks, to persons considered to be at particularly high risk, although in settings where good compliance with 4-weekly dosing can be achieved, more frequent dosing is rarely needed. Oral penicillin V (250 mg) can be given twice-daily instead but is somewhat less effective than benzathine penicillin G. Penicillin allergic patients can receive erythromycin (250 mg) twice daily.

The duration of secondary prophylaxis is determined by many factors, in particular the duration since the last episode of ARF (recurrences become less likely with increasing time), age (recurrences are less likely with increasing age), and the severity of RHD (if severe, it may be prudent to avoid even a very small risk of recurrence because of the potentially serious consequences) **(Table 322-3)**. Secondary prophylaxis is best delivered as part of a coordinated RHD control program, based around a registry of patients. Registries improve the ability to follow patients and identify those who default from prophylaxis and institute strategies to improve adherence.

Table 322-3 American Heart Association Recommendations for Duration of Secondary Prophylaxis*

Category of Patient	Duration of Prophylaxis
Rheumatic fever without carditis	For 5 years after the last attack or 21 years of age (whichever is longer)
Rheumatic fever with carditis but no residual valvular disease	For 10 years after the last attack, or 21 years of age (whichever is longer)
Rheumatic fever with persistent valvular disease, evident clinically or on echocardiography	For 10 years after the last attack, or 40 years of age (whichever is longer). Sometimes lifelong prophylaxis.

*These are only recommendations and must be modified by individual circumstances as warranted. Note that other organizations have slightly different recommendations (see www.worldheart.org/rhd for links).

Source: Adapted from AHA Scientific Statement Prevention of Rheumatic Fever and Diagnosis and Treatment of Acute Streptococcal Pharyngitis. Circulation 119:1541, 2009.

FURTHER READINGS

Carapetis JR et al: Acute rheumatic fever. Lancet 366:155, 2005 [PMID: 16005340] [Full Text]

——— et al: The global burden of group A streptococcal diseases. Lancet Infect Dis 5:685, 2005

Gerber MA et al: Prevention of rheumatic fever and diagnosis and treatment of acute Streptococcal pharyngitis: A scientific statement from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee of the Council on Cardiovascular Disease in the Young, the Interdisciplinary Council on Functional Genomics and Translational Biology, and the Interdisciplinary Council on Quality of Care and Outcomes Research: Endorsed by the American Academy of Pediatrics. Circulation 119:1541, 2009 [PMID: 19246689] [Full Text]

Lennon D et al: School-based prevention of acute rheumatic fever: A group randomized trial in New Zealand. Pediatr Infect Dis J 28:787, 2009 [PMID: 19710585] [Full Text]

National Heart Foundation of Australia: *Diagnosis and Management of Acute Rheumatic Fever and Rheumatic Heart Disease in Australia: Complete Evidence-Based Review and Guideline.* Melbourne, National Heart Foundation of Australia, 2006

Paar JA et al: Prevalence of rheumatic heart disease in children and young adults in Nicaragua. Am J Cardiol 105:1809, 2010 [PMID: 20538135] [Full Text]

Special Writing Group of the Committee on Rheumatic Fever, Endocarditis and Kawasaki Disease of the American Heart Association: Guidelines for the diagnosis of acute rheumatic fever: Jones criteria, 1992 update. JAMA 268:2069, 1992

Steer A et al: Group A streptococcal vaccines: facts versus fantasy. Curr Opin Infect Dis 22:544, 2009 [PMID: 19797947] [Full Text]

Steer AC, Carapetis JR: Acute rheumatic fever and rheumatic heart disease in indigenous populations. Pediatr Clin North Am 56:1401, 2009 [PMID: 19962028] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION

Systemic sclerosis (SSc) is a connective tissue disorder of unknown etiology, heterogeneous clinical manifestations, and chronic and often progressive course. The diffuse cutaneous form of SSc (dcSSc) is characterized by thickening of the skin (scleroderma) and distinctive involvement of multiple internal organs, most notably the lungs, gastrointestinal tract, heart, and kidneys. The early stage of the disease is associated with prominent inflammatory features. Over time, patients develop functional and structural alterations in multiple vascular beds and progressive visceral organ dysfunction due to fibrosis. Although the presence of thickened skin (scleroderma) distinguishes SSc from other connective tissue diseases, scleroderma-like skin induration can occur in localized forms of scleroderma and other disorders **(Table 323-1)**. Patients can be classified into two principal subsets defined largely by the pattern of skin involvement, as well as clinical and laboratory manifestations **(Table 323-2)**. Diffuse cutaneous SSc is associated with progressive skin induration, starting in the fingers and ascending from distal to proximal extremities, the face, and the trunk. These patients are at risk for early pulmonary fibrosis and acute renal involvement. Patients with limited cutaneous SSc (lcSSc) generally have long-standing Raynaud's phenomenon before other manifestations of SSc appear. Skin involvement in lcSSc is slowly progressive and remains limited to the fingers (sclerodactyly), distal extremities, and face, but the trunk is not affected. A subset of patients with lcSSc have prominent calcinosis cutis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia, a constellation termed *CREST syndrome.* However, these features may also be seen in patients with dcSSc. Visceral organ involvement in lcSSc tends to show insidious progression. Although the long-term prognosis of lcSSc is better than that of dcSSc, pulmonary arterial hypertension (PAH), interstitial lung disease, hypothyroidism, and primary biliary cirrhosis may develop in the late stage of lcSSc. In some patients, Raynaud's phenomenon and other typical features of SSc occur in the absence of detectable skin thickening. This syndrome has been termed *SSc sine scleroderma.*

Table 323-1 Conditions Associated with Scleroderma-Like Induration

Systemic sclerosis (SSc)
Limited cutaneous SSc Diffuse cutaneous SSc
Localized scleroderma Guttate morphea, diffuse morphea Linear scleroderma, coup de sabre, hemifacial atrophy
Pansclerotic morphea
Overlap syndromes Mixed connective tissue disease SSc/polymyositis
Stiff Skin Syndrome
Undifferentiated connective tissue disease Scleredema and diabetic scleredema
Scleromyxedema (papular mucinosis)
Nephrogenic systemic fibrosis (nephrogenic fibrosing dermatopathy)
Chronic graft-versus-host disease Diffuse fasciitis with eosinophilia (Shulman disease, eosinophilic fasciitis)
Eosinophilia-myalgia syndrome

Chemically induced scleroderma-like conditions
Vinyl chloride–induced disease Pentazocine-induced skin fibrosis
Paraneoplastic

Table 323-2 Subsets of Systemic Sclerosis (Ssc): Limited Cutaneous Ssc versus Diffuse Cutaneous Ssc

Features	Limited Cutaneous SSc	Diffuse Cutaneous SSc
Skin involvement	Indolent onset. Limited to fingers, distal to elbows, face; slow progression	Rapid onset. Diffuse: fingers, extremities, face, trunk; rapid progression
Raynaud's phenomenon	Precedes skin involvement; associated with critical ischemia	Onset coincident with skin involvement, may be mild
Musculoskeletal	Early arthralgia, fatigue	Severe arthralgia, carpal tunnel syndrome, tendon friction rubs
Pulmonary fibrosis	Occasional, moderate	Frequent, early and severe
Pulmonary arterial hypertension	Frequent, late, may be isolated	May occur, often in association with pulmonary fibrosis
Scleroderma renal crisis	Very rare	Occurs in 15%; early
Calcinosis cutis	Frequent, prominent	May occur, mild
Characteristic autoantibodies	Anticentromere	Antitopoisomerase I (Scl-70), anti-RNA polymerase III

EPIDEMIOLOGY

SSc is an acquired sporadic disease with a worldwide distribution and affecting all races. In the United States, the incidence is estimated at 9–19 cases per million per year. The only community-based survey of SSc yielded a prevalence of 286 cases per million population. There are an estimated 100,000 cases in the United States, although this number may be significantly higher if patients who do not meet strict classification criteria are also included. Studies from England, Australia, and Japan showed rates of SSc that were lower than in the United States. Age, gender, and ethnicity are important factors determining disease susceptibility. Like other connective tissue diseases, SSc shows a female predominance that is most pronounced in the childbearing years and declines after menopause. While SSc can present at any age, the most common age of onset for both limited and diffuse cutaneous forms is in the range of 30–50 years. The incidence is higher in blacks than whites, and disease onset occurs at an earlier age. Furthermore, blacks are more likely to have the diffuse cutaneous form of SSc associated with interstitial lung involvement and a worse prognosis.

GENETIC CONSIDERATIONS

SSc shows a non-Mendelian pattern of inheritance. Monozygotic twins have a relatively low concordance rate for SSc (4.7%), although concordance for antinuclear antibodies is significantly greater. A genetic contribution to disease susceptibility is indicated by the fact that 1.6% of SSc patients have a first-degree relative with SSc, a prevalence rate substantially higher than in the general population. The risk of other autoimmune diseases, including systemic lupus erythematosus (SLE) **(Chap. 319)** and rheumatoid arthritis **(Chap. 321)**, is also increased. Among Choctaw Indians from Oklahoma, SSc prevalence as high as 4690 per million has been reported. Genetic investigations in SSc have focused on candidate gene polymorphisms. Small studies have shown associations with the genes encoding angiotensin-converting enzyme (ACE); endothelin-1 and nitric oxide synthase; B cell markers (CD19); chemokines (monocyte chemoattractant protein-1) and chemokine receptors; interferon signaling mediators STAT4 and IRF5; migration inhibitory factor; cytokines [interleukin 1α (IL-1α, IL-4, and tumor necrosis factor α (TNF-α)]; growth factors and their receptors [connective tissue growth factor (CTGF) and transforming growth factor β (TGF-β)]; and extracellular matrix proteins [fibronectin, fibrillin, and secreted protein acidic-rich in cysteine (SPARC)]. To date, these genetic studies indicate that as in other complex diseases, multiple genetic loci are involved in SSc, and their

individual contributions to disease susceptibility are modest. Genome-wise association studies to identify additional genetic susceptibility loci in SSc are currently underway.

ENVIRONMENTAL AND OCCUPATIONAL RISK FACTORS

Patients with SSc have increased serum antibodies to human cytomegalovirus (hCMV), and antitopoisomerase-I (Scl-70) autoantibodies recognize antigenic epitopes present on the hCMV-derived proteins, suggesting molecular mimicry as a possible mechanistic link between hCMV infection and SSc. Evidence of human parvovirus B19 infection in SSc patients has also been presented; however, the etiologic role of viruses remains unproven. Reports of geographic clustering of SSc cases suggesting shared environmental exposures have not been substantiated by careful investigation. An epidemic of a novel syndrome with features suggestive of SSc occurred in Spain in the 1980s. The outbreak, termed *toxic oil syndrome* and affecting over 20,000 individuals, was linked to contaminated rapeseed oils used for cooking. A similar epidemic outbreak, termed *eosinophilia-myalgia syndrome* (EMS), occurred a decade later in the United States. Affected individuals presented with marked eosinophilia and severe myalgia, followed by the development of scleroderma-like chronic skin lesions. The EMS epidemic was linked to the consumption of imported batches of L-tryptophan used as dietary supplements. While both of these apparently novel toxic-epidemic syndromes were characterized by scleroderma-like chronic skin changes and variable visceral organ involvement they were associated with clinical, pathologic, and laboratory features that clearly distinguished them from SSc. The incidence of SSc is increased among miners exposed to silica. Other occupational exposures tentatively linked with SSc include polyvinyl chloride, epoxy resins, and aromatic hydrocarbons including toluene and trichloroethylene. Drugs implicated in SSc-like illnesses include bleomycin, pentazocine and cocaine, and appetite suppressants linked with pulmonary hypertension. As yet unknown inhaled factors may play a role in the development of SSc-associated interstitial lung disease. Case reports and series describing SSc in women with silicone breast implants had raised concern regarding a possible causal role of silicone in SSc. However, large-scale epidemiologic investigations found no evidence of increased risk of SSc.

PATHOGENESIS

A comprehensive view of the pathogenesis of SSc must incorporate the three cardinal features of the disease: (1) vasculopathy, (2) cellular and humoral auto immunity, and (3) progressive visceral and vascular fibrosis in multiple organs **(Fig. 323-1)**. Autoimmunity and altered vascular reactivity may be the earliest manifestations of SSc. Complex interplay between these processes is thought to initiate and then amplify the fibrotic process.

Figure 323-1

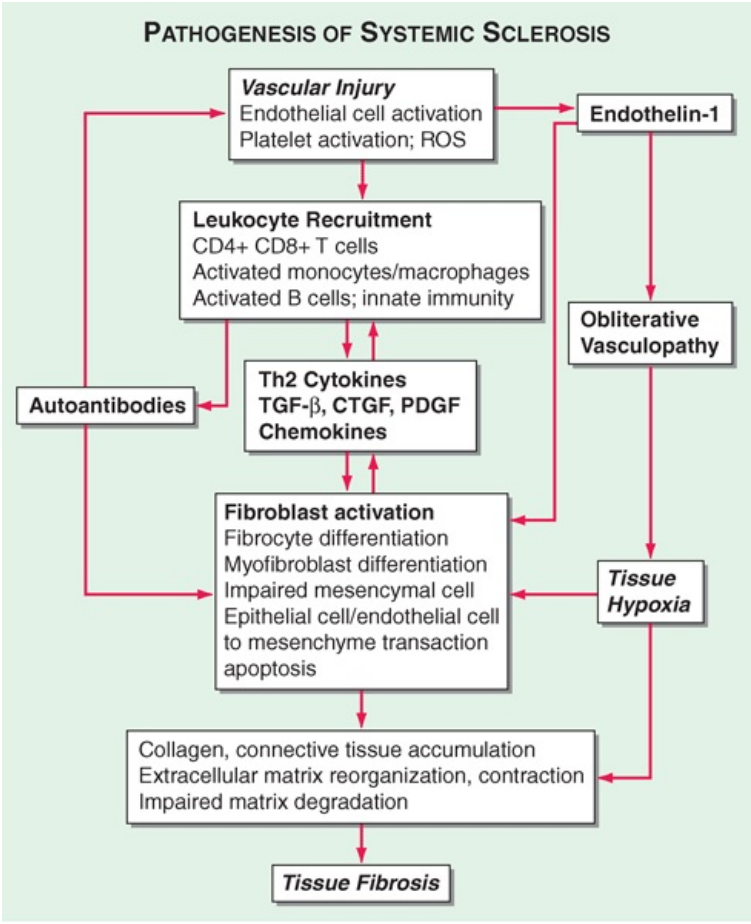

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Initial vascular injury in a genetically susceptible individual leads to functional and structural vascular alterations, inflammation, and autoimmunity. The inflammatory and immune responses initiate and sustain fibroblast activation and differentiation, resulting in pathologic fibrogenesis and irreversible tissue damage. Vascular damage results in tissue ischemia that further contributes to progressive fibrosis and atrophy. CTGF, connective tissue growth factor; PDGF, platelet-derived growth factor; TGF-β, transforming growth factor-β.

Animal Models of Disease

There is no single animal model of SSc that reproduces the three cardinal processes that underlie the pathogenesis, but some models recapitulate selected disease characteristics. The tight-skin mouse (Tsk1) is a naturally occurring fibrosis model characterized by spontaneous skin thickening. The mutation responsible for the phenotype, a duplication in the fibrillin-1 gene, gives rise to an abnormally large fibrillin-1 protein that contributes to defective extracellular matrix assembly and aberrant activation of TGF-β. Mutations in the fibrillin-1 gene are associated with Marfan's disease and the stiff skin syndrome but have not been described in patients with SSc. Fibrosis in the skin and lungs can be induced in mice by bleomycin injections or by transplantation of human leukocyte antigen (HLA)-mismatched bone marrow or spleen cells. Increasingly, manipulation of mice via mutagenesis or targeted genetic modification such as knock-out or transgenesis are utilized to create new disease models and for dissecting the roles of individual molecules in the underlying processes. For example, genetic targeting of Smad3, an intracellular TGF-β signal transducer, or of the nuclear receptor peroxisome proliferator-activated receptor (PPAR) gamma, yielded mice that were resistant or hypersensitive to bleomycin-induced scleroderma. These mouse models become increasingly useful for preclinical testing of novel treatments.

Microangiopathy

Vascular involvement in SSc is extensive, involves multiple vascular beds, and has important clinical consequences. Raynaud's phenomenon, an early manifestation, is characterized by an altered blood-flow response to cold challenge. This initially reversible functional vascular abnormality is associated with alterations in the autonomic and peripheral nervous systems, with impaired production of neuropeptides such as calcitoningene-related peptide from sensory afferent nerves and heightened sensitivity of α_2-adrenergic receptors on vascular smooth-muscle cells. While isolated Raynaud's phenomenon is common, relatively benign, and nonprogressive, SSc-associated Raynaud's phenomenon is frequently complicated by irreversible structural and functional changes. Viruses,

superoxide radicals, vascular cytotoxic factors, and immune responses such as complement and circulating autoantibodies to endothelial cells may each contribute to endothelial cell injury in early SSc. Endothelial injury results in dysregulated production of endothelium-derived vasodilatory (nitric oxide and prostacyclin) and vasoconstricting (endothelin-1) substances, as well as increased expression of intercellular adhesion molecule 1 (ICAM-1) and other surface adhesion molecules. Microvessels show enhanced permeability and transendothelial leukocyte diapedesis, activation of coagulation and fibrinolytic cascades, and platelet aggregation. Smooth muscle cell–like myointimal cells proliferate, the basement membrane is thickened and reduplicated, and fibrosis of the adventitial layers develops. The vasculopathic process affects capillaries, as well as arterioles, and even large vessels in many organs, resulting in reduced blood flow, tissue ischemia, and generation of profibrotic factors. Progressive luminal occlusion due to intimal and medial hypertrophy, combined with persistent endothelial cell damage and adventitial fibrosis, establish a vicious cycle culminating in the striking absence of blood vessels seen on angiograms of the hands and kidneys in late-stage disease. Damaged endothelium promotes platelet aggregation with release of serotonin and platelet alpha granules including thromboxane, a potent vasoconstrictor, and of platelet-derived growth factor (PDGF). Vascular compromise is aggravated by defective fibrinolysis. Oxidative stress due to ischemia-reperfusion is associated with generation of reactive oxygen species (ROS) that further damage the endothelium through peroxidation of membrane lipids. Paradoxically, the process of revascularization that normally reestablishes blood flow to ischemic tissue is defective in SSc despite elevated levels of vascular endothelial growth factor (VEGF) and other angiogenic factors. The number of bone marrow–derived CD34+ CD133+ endothelial progenitor cells is markedly reduced in the circulation, and their differentiation in vitro into mature endothelial cells is impaired. Thus, widespread capillary malformation and loss, obliterative vasculopathy of small and medium-sized arteries, and failure to repair damaged vessels are hallmarks of SSc.

Inflammation and Cellular Immunity

In the early stages of SSc, activated T cells and monocytes/macrophages accumulate in lesional skin, lungs, and other affected organs. Infiltrating T cells express CD45 and HLA-DR activation markers and display restricted T cell receptor signatures indicative of oligoclonal expansion in response to (unknown) antigen. Circulating CD4+ T cells have elevated levels of chemokine receptors and α_1 integrin adhesion molecules, accounting for their enhanced ability to bind to endothelium and to fibroblasts. Endothelial cells express ICAM-1 and other adhesion molecules that facilitate leukocyte diapedesis. Activated macrophages and T cells show a T_H2-polarized immune response and secrete IL-4 and IL-13. T_H2 cytokines induce the production of TGF-β and promote collagen synthesis and other profibrotic responses, whereas the T_H1 cytokineinterferon γ (IFN-γ) inhibits collagen synthesis and blocks cytokine-mediated fibroblast activation. Because TGF-β stimulates its own synthesis, as well as that of CTGF (also termed CCN2) and other cytokines, TGF-β establishes an autocrine/paracrine stimulatory loop that sustains activation of fibroblasts and other effector cells **(Chaps. 314 and 318)**. Regulatory T cells (Tregs) are essential for maintaining normal immune tolerance. While the frequency of Tregs in the peripheral blood is elevated in SSc, their immunosuppressive function is defective.

HUMORAL AUTOIMMUNITY

Antinuclear antibodies occur in virtually all patients with SSc. In addition, a number of mutually exclusive autoantibodies that are highly specific for SSc have been described. These antibodies show strong association with specific disease phenotypes and genetically determined HLA haplotypes **(Table 323-3)**. Autoantibody levels correlate with disease severity, and titers fluctuate with disease activity. While some SSc-specific autoantibodies are antinuclear and directed against intracellular proteins such as topoisomerase-I, and the RNA polymerases, others are directed against cell-surface antigens or secreted proteins. Functional autoantibodies have well-established clinical utility as diagnostic and prognostic markers in SSc, although their pathogenetic role in the disease manifestations remains uncertain. Autoantibodies to fibroblasts, endothelial cells, PDGF cell-surface receptors, fibrillin-1, and matrix metalloproteinase enzymes have been described in SSc. The direct pathogenetic role of these self antibodies in SSc remains to be firmly established.

Table 323-3 Autoantibodies and Associated Features in Systemic Sclerosis (Ssc)

Target Antigen	SSc Subset	Characteristic Clinical Association
Topoisomerase-I	dcSSc	Tendon friction rubs, ILD, cardiac involvement, scleroderma renal crisis
Centromere proteins	lcSSc	Digital ischemia, calcinosis, isolated PAH; renal crisis rare
RNA polymerase III	dcSSc	Extensive skin, tendon friction rubs, renal crisis

U3-RNP	dcSSc	PAH, ILD, scleroderma renal crisis, myositis
Th/T0	lcSSc	ILD, PAH
PM/Scl	lcSSc	Calcinosis, myositis
U1-RNP	MCTD	PAH

Abbreviations: dcSSc, diffuse cutaneous SSc; ILD, interstitial lung disease; lcSSc, limited cutaneous SSc; MCTD, mixed connective tissue disease; PAH, pulmonary arterial hypertension.

A variety of mechanisms have been proposed for the occurrence of autoantibodies in SSc. Proteolytic cleavage, increased expression, or altered subcellular localization of certain cellular proteins in SSc could lead to their recognition as neoepitopes by the immune system. For example, cytotoxic T cells release the protease granzyme B that cleaves peptides, and generates neoepitopes that can break immune tolerance. Recent studies implicate B cells in both the autoimmune and fibrotic responses in SSc. In addition to their well-recognized role in antibody production, B cells can also present antigen, produce IL-6 and TGF-β, and modulate T cell and dendritic cell function. In SSc, B cells show elevated CD19 expression, and reduced numbers of memory B cells and early plasma cells. Gene expression profiling of lesional skin has identified mRNA expression signatures characteristic of B cell activation.

Fibrosis

Fibrosis affecting multiple organs distinguishes SSc from other connective tissue diseases. Fibrosis characteristically follows, and is thought to be a consequence of, autoimmunity and vascular damage. The process, characterized by progressive replacement of normal tissue architecture with dense connective tissue, accounts for substantial morbidity and mortality. Fibroblasts are mesenchymal cells responsible for maintaining the functional and structural integrity of connective tissue. When activated by TGF-β and related factors, fibroblasts proliferate, migrate, secrete collagens and extracellular matrix, growth factors, and cytokines, and transdifferentiate into myofibroblasts. Under normal conditions, these responses allow fibroblasts to repair tissue damage. The rapid and self-limited physiologic repair program becomes sustained and amplified in pathologic fibrosis, resulting in the irreversible accumulation of scar tissue.

In addition to connective tissue–resident fibroblasts, and transformation of epithelial cells into fibroblasts, circulating mesenchymal progenitor cells of bone marrow origin might also contribute to fibrosis. The factors that regulate the development of mesenchymal progenitor cells in the bone marrow, their trafficking from the circulation into lesional tissue, and in situ into matrix-producing fibroblasts, are unknown. Epithelial and endothelial cells and fibroblasts can differentiate into smooth-muscle–like myofibroblasts characterized by prominent cytoskeleal structures containing alpha smooth muscle actin. While myofibroblasts can be transiently detected during normal wound healing, they persist in tissue during pathologic fibrogenesis, possibly due to resistance to apoptosis. Myofibroblasts contribute to scar formation via production of collagen and TGF-β, and contraction of the surrounding extracellular matrix.

Explanted fibroblasts display an abnormally activated phenotype in culture. Compared to normal fibroblasts, SSc fibroblasts have variably increased rates of collagengene transcription and display smooth-muscle actin stress fibers. Furthermore, they show enhanced secretion of extracellular matrix molecules, cytokines, and growth factors; expression of chemokine receptors and cell surface adhesion molecules; resistance to apoptosis; spontaneously generate ROS; and autocrine TGF-β signaling. The abnormal "scleroderma phenotype" of these cells persists during their serial passage in vitro. Factors contributing to the autonomously activated phenotype include autocrine TGF-β stimulatory loops, hypoxia, deregulated microRNA expressions and other epigenetic modifications, and altered cell-matrix interaction. Global transcriptome analyses show differential expression of many extracellular matrix genes, including collagens, fibronectin, and fibrillins in SSc fibroblasts. A majority of the abnormally expressed genes could be linked to TGF-β responses, but other fibrogenic signaling pathways involving CTGF, endothelin-1m hypoxia, PDGF and Wnts are also operative in SSc.

PATHOLOGY

The distinguishing pathologic hallmark of SSc is the combination of widespread capillary loss and obliterative vasculopathy of small arteries and arterioles, together with fibrosis in the skin and internal organs. In early disease, perivascular cellular infiltrates composed of CD4+ and CD8+ T lymphocytes, monocytes/macrophages, plasma cells, mast cells, and occasionally B cells may be detected in multiple organs prior to the appearance of fibrosis. The vascular lesion is characterized by intimal proliferation in the small and medium-sized arteries, resulting in luminal narrowing. Obliterative vasculopathy as a late finding is prominent in the heart, lungs, kidneys, and intestinal tract. Fibrosis is found in the skin, lungs, gastrointestinal tract, heart, tendon sheath, perifascicular tissue surrounding skeletal muscle, and some endocrine organs. In these tissues, accumulation of connective tissue composed of endothelin-1m collagens, fibronectin, proteoglycans, and other structural macromolecules progressively disrupts normal architecture, resulting in functional impairment of affected organs.

Skin

In the skin, fibrosis causes dermal expansion and obliteration of the hair follicles, sweat glands, and other appendages **(Fig. 323-2*A*)**. Collagen fiber accumulation is most prominent in the reticular dermis, and the fibrotic process invades the subjacent adipose layer with entrapment of fat cells. The epidermis is atrophic, and the rete pegs are effaced.

Figure 323-2

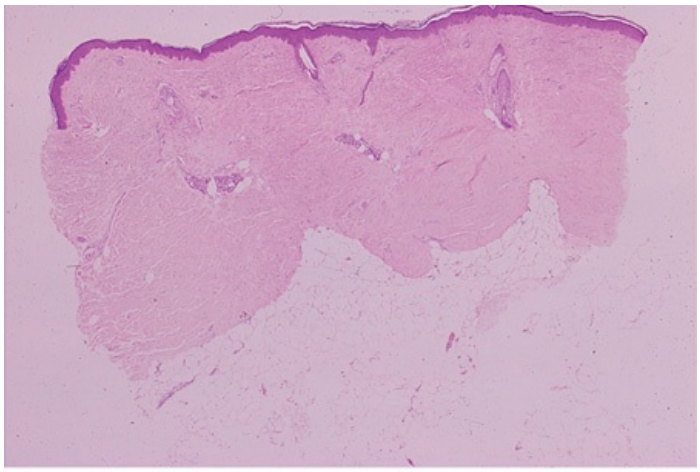

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

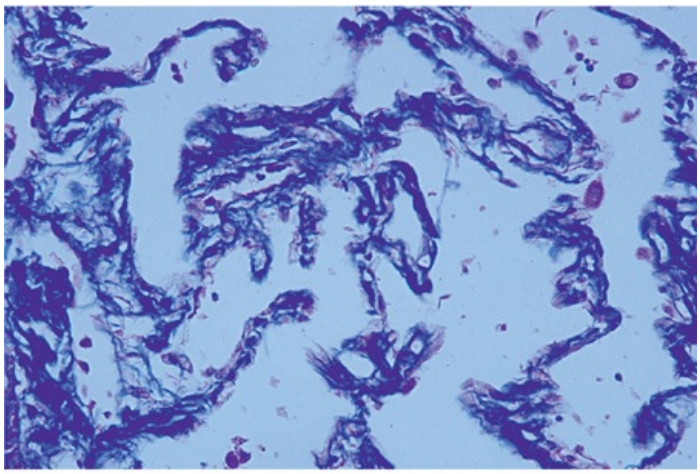

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

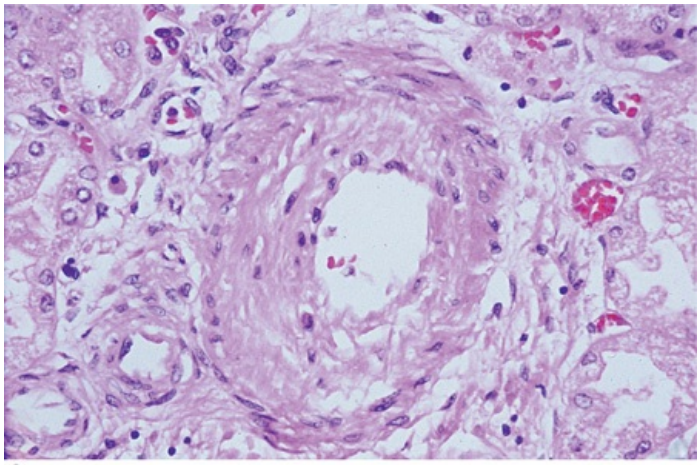

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathologic findings in systemic sclerosis (SSc).*A.* Dermal sclerosis. The skin is thickened due to marked expansion of the dermis. Thick bundles of densely packed collagen replace skin appendages. ***B.*** Early interstitial lung disease. Diffuse fibrosis of the alveolar septae and a chronic inflammatory cell infiltrate. Trichrome stain.***C.*** Pulmonary arterial obliterative vasculopathy. Striking intimal hyperplasia and narrowing of the lumen of a small pulmonary artery, with minimal interstitial fibrosis, in a patient with limited cutaneous SSc.

Lungs

Patchy infiltration of the alveolar walls with T lymphocytes, macrophages, and eosinophils occurs in early disease. With progression, interstitial fibrosis and vascular damage dominate the pathologic picture, often coexisting within the same lesions in patients with dcSSc. Pulmonary fibrosis is characterized by expansion of the alveolar interstitium, with accumulation of collagen and other connective tissue proteins. The most common histologic pattern in SSc is fibrotic nonspecific interstitial pneumonia **(Fig. 323-2***B*). Progressive thickening of the alveolar septae results in obliteration of the airspaces and honeycombing, as well as loss of pulmonary blood vessels. This process impairs gas exchange and contributes to worsening of pulmonary hypertension. Intimal thickening of the pulmonary arteries, best seen with elastin stain, underlies pulmonary hypertension **(Fig. 323-2***C*) and, at autopsy, is often associated with multiple pulmonary emboli and evidence of myocardial fibrosis.

Gastrointestinal Tract

Pathologic changes can be found at any level from the mouth to the rectum. The lower esophagus is frequently involved, with prominent atrophy of the muscular layers; striated muscle in the upper third of the esophagus is generally spared. Characteristic vascular lesions are often present. Replacement of the normal intestinal tract architecture results in diminished peristaltic activity, with gastroesophageal reflux, dysmotility, and small-bowel obstruction. Chronic reflux is associated with esophageal inflammation, ulcerations, and stricture formation and may lead to Barrett's metaplasia.

Kidneys

In the kidneys, lesions in the interlobular and arcuate arteries predominate, whereas glomerulonephritis is rare. Chronic renal ischemia is associated with shrunken glomeruli. Patients with scleroderma renal crisis show dramatic changes in small renal arteries with reduplication of elastic lamina, marked intimal proliferation, and narrowing of the lumen, often accompanied by thrombosis and microangiopathic hemolysis.

Heart

The heart is frequently affected, with prominent involvement of the myocardium and pericardium. The characteristic arteriolar lesions are concentric intimal hypertrophy and luminal narrowing, accompanied by contraction band necrosis reflecting ischemia-reperfusion injury, and patchy myocardial fibrosis that may also affect the conduction system. Despite the prominent role of ischemia in SSc, the frequency of atherosclerotic coronary artery disease is comparable to the general population.

Other Organs

Synovitis may be found in early SSc; however, with progression of the disease, the synovium becomes fibrotic. Fibrosis of tendon sheaths and fascia produces palpable and sometimes audible tendon friction rubs. Inflammation, and, in later stages, atrophy and fibrosis of the muscles are common findings. Fibrosis of the thyroid gland and of the minor salivary glands may be seen.

CLINICAL FEATURES

Overview

Virtually every organ is affected in SSc **(Table 323-4)**. While there is a great deal of variability in the clinical manifestations from one patient to the next, patients can be classified into one of two major subsets based on the pattern of skin involvement (Table 323-2). Moreover, while dcSSc is associated with prominent and early internal organ involvement, lcSSc presents with long-standing Raynaud's phenomenon, indolent skin, limited internal organ involvement, and a better prognosis. While patient stratification into diffuse and limited cutaneous subsets is useful, disease expression is far more complex, and several distinct phenotypes exist within each subset. For example, 10–15% of patients with lcSSc develop severe pulmonary arterial hypertension without significant interstitial lung disease (ILD). Other patients have systemic features of SSc without appreciable skin involvement (SSc sine scleroderma). Unique clinical phenotypes of SSc associate with specific autoantibodies (Table 323-3). Patients with "overlap" have typical SSc features coexisting with clinical and laboratory evidence of another autoimmune disease such as polymyositis, Sjögren's syndrome, polyarthritis, autoimmune liver disease, or SLE.

Table 323-4 Internal Organ Involvement: Limited Cutaneous and Diffuse Cutaneous Forms of Systemic Sclerosis

Features	Limited Cutaneous SSc (%)	Diffuse Cutaneous SSc (%)
Skin involvement	90*	100
Raynaud's phenomenon	99	98
Esophageal involvement	90	80
Pulmonary fibrosis	35	65
Pulmonary arterial hypertension	15	15
Myopathy	11	23
Cardiac involvement	9	12
Scleroderma renal crisis	2	15

*10% of lcSSc patients have SSc sine scleroderma.

The term *scleroderma* refers to localized scleroderma and is used to describe a group of localized skin disorders that primarily affect children (Table 323-1). In contrast to SSc, localized scleroderma is rarely associated with Raynaud's phenomenon or internal organ involvement. Morphea presents as solitary or multiple circular patches of thickened skin and, less commonly, widespread induration (generalized or pansclerotic morphea); the fingers are spared. Linear scleroderma—streaks of thickened skin, typically in one or both lower extremities—may affect the subcutaneous tissues with fibrosis and atrophy of supporting structures, muscle, and bone. In children, the growth of affected long bones can be retarded. When linear scleroderma lesions cross joints, significant contractures can develop.

Initial Clinical Presentation

The initial presentation is quite different in the diffuse and the limited cutaneous forms of the disease. In patients with dcSSc, the interval between Raynaud's phenomenon and appearance of other manifestations is generally brief (weeks to months). Soft tissue swelling and intense pruritus are signs of the early inflammatory "edematous" phase of dcSSc. The fingers, hands, distal limbs, and face are usually affected first. Diffuse hyperpigmentation and carpal tunnel syndrome can occur. Arthralgias, muscle weakness and decreased joint mobility are common. During the ensuing weeks to months, the inflammatory edematous phase evolves into the "fibrotic" phase, with skin induration that is associated with loss of body hair, reduced production of skin oils, and a decline in sweating capacity. The subcutaneous tissue becomes affected, with fat atrophy and fibrosis of underlying fascia, muscle, and other soft tissue structures. Progressive flexion contractures of the fingers ensue. The wrists, elbows, shoulders, hip girdles, knees, and ankles become stiff due to fibrosis of the supporting joint structures. While advancing skin involvement is the most visible manifestation of early dcSSc, important internal organ involvement develops during this stage. The initial 4 years from disease onset is the period of rapidly evolving systemic involvement and greatest risk for pulmonary and renal damage. If organ failure does not occur during this period, the systemic process may stabilize.

Compared to dcSSc, the course of lcSSc is generally more indolent. The period between the onset of Raynaud's phenomenon and manifestations such as gastroesophageal reflux, telangiectasia, or calcinosis can be several years. Raynaud's phenomenon tends to be more severe than in dcSSc, and can be associated with critical ischemia, ulcerations, and autoamputation of the fingers. On the other

hand, significant renal involvement and pulmonary fibrosis are uncommon in lcSSc patients. Cardiac involvement and isolated pulmonary arterial hypertension develop in 10–15%. Overlap of SSc with the sicca complex, polyarthritis, cutaneous vasculitis, and biliary cirrhosis is seen primarily in the lcSSc subset.

ORGAN INVOLVEMENT

Raynaud's Phenomenon

Raynaud's phenomenon is an episodic vasoconstriction in the fingers and toes that occurs in virtually every patient with SSc. Vasoconstriction may also affect the tip of the nose and earlobes. Attacks are triggered by exposure to cold, a decrease in temperature, emotional stress, and vibration. Typical attacks start with pallor, followed by cyanosis of variable duration. Eventually erythema develops spontaneously or with rewarming of the digit. The progression of the three color phases reflects the underlying pathogenic mechanisms of vasoconstriction, ischemia, and reperfusion.

As much as 3–5% of the general population has Raynaud's phenomenon, and it is more frequent in women. In the absence of associated signs or symptoms of an underlying condition, Raynaud's phenomenon is classified as primary, and represents an exaggerated physiologic response to cold. Secondary Raynaud's phenomenon can occur as a complication of SSc and other connective tissue diseases, hematologic and endocrine conditions, and occupational disorders, and with the use of drugs such as the beta blocker atenolol and anticancer drugs such as cisplatin and bleomycin. Distinguishing primary versus secondary Raynaud's phenomenon can present a diagnostic challenge. The diagnosis of primary Raynaud's phenomenon is supported by the following: absence of an underlying cause on history and physical examination; a family history of Raynaud's phenomenon; absence of digital tissue necrosis, ulceration, or gangrene; and a negative test for antinuclear antibodies. Secondary Raynaud's phenomenon tends to develop at an older age (>30 years), is clinically more severe (episodes more frequent, prolonged, and painful), and is frequently associated with ischemic lesions and infarction in the digits **(Fig. 323-3)**. The cutaneous capillaries at the nail bed can be viewed under a drop of grade B immersion oil using a low-power stereoscopic microscope. Nailfold capillaroscopy can be helpful in the evaluation of Raynaud's phenomenon; patients with primary Raynaud's phenomenon have normal capillaries that appear as regularly spaced parallel vascular loops, whereas in SSc and other connective tissue diseases, nailfold capillaries are distorted with widened and irregular loops, dilated lumen, and areas of vascular "dropout." In SSc, abnormal vascular reactivity may involve multiple vascular beds, and cold-induced Raynaud's-like episodic vasospasm has been documented in the pulmonary, renal, gastrointestinal, and coronary circulations.

Figure 323-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Digital necrosis. Sharply demarcated necrosis of the fingertip in a patient with limited cutaneous systemic sclerosis (SSc) associated with severe Raynaud's phenomenon.

Skin Features

While early-stage SSc is associated with edematous skin changes, skin thickening is the hallmark that distinguishes SSc from other connective tissue diseases. The distribution of skin thickening is invariably symmetric and bilateral. It typically starts in the fingers, and then characteristically advances from distal to proximal extremities in an ascending fashion. The involved skin is firm, coarse, and thickened, and the extremities and trunk may be darkly pigmented. In some patients, diffuse tanning in the absence of sun exposure is a very early manifestation of skin involvement. In dark-skinned patients, vitiligo-like hypopigmentation may occur. Because pigment loss spares the perifollicular areas, the skin may have a "salt-and-pepper" appearance, most prominently on the scalp, upper back, and chest. Dermal sclerosis due to collagen accumulation causes obliteration of hair follicles, sweat glands, and eccrine and sebaceous glands, resulting in hair loss, decreased sweating, and dry skin. Transverse creases on the dorsum of the fingers disappear **(Fig. 323-4)**. Fixed flexion contractures of the fingers cause reduced hand mobility and lead to muscle atrophy. Skin thickening in combination with fibrosis of the subjacent tendons accounts for contractures of the wrists, elbows, and knees. Thick ridges at the neck due to firm adherence of skin to the underlying platysma muscle interfere with neck extension. The face assumes a characteristic "mauskopf" appearance with taut and shiny skin, loss of wrinkles, and occasionally an expressionless facies due to reduced mobility of the eyelids, cheeks, and mouth. Thinning of the lips with accentuation of the central incisor teeth and fine wrinkles (radial furrowing) around the mouth complete the picture. Reduced oral aperture (microstomia) interferes with eating and oral hygiene. The nose assumes a pinched, beak-like appearance.

Figure 323-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sclerodactyly. Note skin induration on the fingers, and fixed flexion contractures at the proximal interphalangeal joints in a patient with limited cutaneous systemic sclerosis (SSc).

In established SSc, the skin is firmly bound to the subcutaneous fat (tethering) and undergoes thinning and atrophy. Telangiectasia are dilated skin capillaries 2–20 mm in diameter frequently seen in lcSSc. These lesions, reminiscent of hereditary hemorrhagic telangiectasia, are prominent on the face, hands, lips, and oral mucosa **(Fig. 323-5)**. Breakdown of atrophic skin leads to chronic ulcerations at the extensor surfaces of the proximal interphalangeal joints, the volar pads of the fingertips, and bony prominences such as the elbows and malleoli. Ulcers are painful and may become secondarily infected, resulting in osteomyelitis. Healing of ischemic fingertip ulcerations leaves characteristic fixed digital "pits." Loss of soft tissue at the fingertips due to ischemia is frequent and may be associated with striking resorption of the terminal phalanges (acro-osteolysis) **(Fig. 323-6)**.

Figure 323-5

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cutaneous vascular changes.*A.* Capillary changes at the nailfold in a patient with limited cutaneous systemic sclerosis (lcSSc). ***B.*** Telangiectasia on the face.

Figure 323-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acro-osteolysis. Note dissolution of terminal phalanges in a patient with long-standing limited cutaneous systemic sclerosis (lcSSc) and Raynaud's phenomenon.

Calcium deposits occur in the skin and soft tissues. Calcinosis cutis is most common in patients with lcSSc who are positive for anticentromere antibodies. The deposits, varying in size from tiny punctate lesions to large conglomerate masses, are composed of calcium hydroxyapatite crystals and can be readily visualized on plain x-rays. Frequent locations include the finger pads, palms, extensor surfaces of the forearms, and the olecranon and prepatellar bursae **(Fig. 323-7)**. Paraspinal calcifications may cause neurologic complications. Calcific deposits appears as persistent firm, nontender subcutaneous lumps. They may occasionally ulcerate through the overlying skin, producing drainage of chalky white material, pain, and local inflammation.

Figure 323-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Calcinosis cutis. Note large calcific deposit breaking through the skin in a patient with limited cutaneous systemic sclerosis (lcSSc).

Pulmonary Features

Pulmonary involvement can be documented in most patients with SSc and is now the leading cause of death. There are two main types of significant pulmonary involvement: ILD and PAH. Many patients with SSc develop some degree of both complications. Less frequent pulmonary manifestations of SSc include aspiration pneumonitis complicating gastroesophageal reflux, pulmonary hemorrhage due to endobronchial telangiectasia, obliterative bronchiolitis, pleural reactions, restrictive ventilatory defect due to chest wall fibrosis, spontaneous pneumothorax, and drug-induced lung toxicity. The incidence of lung cancer, particularly bronchioloalveolar carcinoma, may be increased.

Pulmonary involvement can remain asymptomatic until it is advanced. The most frequent presenting respiratory symptoms—exertional dyspnea, fatigue, and reduced exercise tolerance—are often subtle and slowly progressive. A chronic dry cough may be present. Physical examination may reveal "Velcro"

crackles at the lung bases. Pulmonary function testing (PFT) is a sensitive method for detecting early pulmonary involvement. The most common abnormalities are reductions in forced vital capacity (FVC) or single breath diffusing capacity of the lung for carbon monoxide (DLCO). A reduction in DLCO that is significantly out of proportion to the reduction in FVC suggests pulmonary vascular disease, but may also be due to anemia. With exercise, patients show a decrease in PO_2.

INTERSTITIAL LUNG DISEASE (ILD)

Some evidence of ILD can be found in up to 90% of patients with SSc at autopsy and 85% by thin-section high-resolution computed tomography (HRCT). ILD and pulmonary fibrosis cause restrictive pulmonary function defect with impaired gas exchange, characterized on PFT by decreased FVC and DLCO but unaffected flow rates. Clinically significant ILD develops in 16–43% of patients with SSc; the frequency varies depending on the detection method used. Risk factors include male gender, African American race, diffuse skin involvement, severe gastroesophageal reflux, and the presence of topoisomerase-I autoantibodies, as well as a low FVC or DLCO at initial presentation. In patients who develop significant ILD, the most rapid progression in lung disease occurs early in the course of the disease (within the first 3 years), when the FVC can decline by 30% per year.

Chest radiography is useful for ruling out infection and other causes of pulmonary involvement, but compared to HRCT it is relatively insensitive for detection of early ILD. HRCT may show subpleural reticular linear opacities, predominantly in the lower lobes, even in asymptomatic patients **(Fig. 323-8)**. Additional findings include mediastinal lymphadenopathy, nodules, traction bronchiectasis and in some cases, honeycomb cystic changes. Ground-glass opacification, alone or in combination with a reticular pattern, is seen in 50% of patients. Ground-glass opacification on HRCT is indicative of fine fibrosis, and does not identify alveolitis or predict rapid progression. The extent of lung disease on HRCT is a predictor of mortality in SSc. Bronchoalveolar lavage (BAL) can demonstrate inflammation in the lower respiratory tract and may be useful for ruling out infection., While an elevated proportion of neutrophils (> 2%) and/or eosinophils (> 3%) in the BAL fluid is correlated with more extensive lung disease on HRCT, and is associated with more rapid decline in FVC and reduced survival, BAL is not useful for identifying reversible alveolitis. Lung biopsy is indicated only in patients with atypical findings on chest radiographs and should be thoracoscopically guided. The histologic pattern on lung biopsy may be helpful in predicting the risk of progression of ILD. The most common pattern in SSc, nonspecific interstitial pneumnia, carries a better prognosis than usual interstitial pneumnia. Recent studies suggest that measurement of serum factors such as KL-6, a glycoprotein found in type II pneumocytes and alveolar macrophages, may have utility as biomarkers for the detection and serial monitoring of ILD in patients with SSc.

Figure 323-8

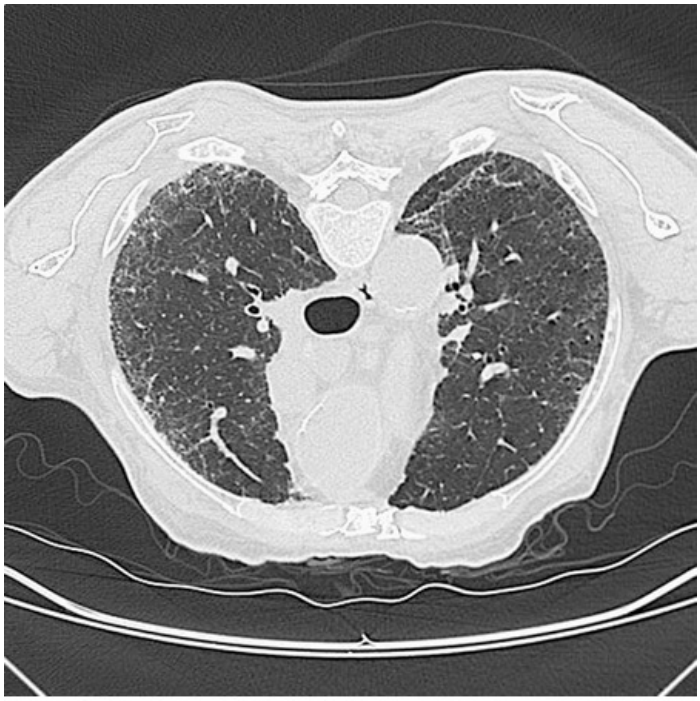

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

High-resolution CT scan of the lungs: interstitial lung disease. Note bilateral reticulonodular opacifications in a peripheral distribution in the lower lobes of the lungs in a patient with diffuse cutaneous systemic sclerosis (dcSSc).

PULMONARY ARTERIAL HYPERTENSION (PAH)

PAH, defined as a mean pulmonary arterial pressure >25mmHg with a pulmonary capillary wedge pressure <15 mmHg, is a major complication of SSc. Approximately 15% of SSc patients have PAH that can occur in association with ILD or as an isolated pulmonary abnormality. The natural history of SSc-associated PAH is variable, but in many patients it follows a downhill course with development of right heart failure and significant mortality. Risk factors for PAH include limited cutaneous disease with anticentromere antibodies, late age at disease onset, severe Raynaud's phenomenon, and the presence of antibodies to U1-RNP, U3-RNP (fibrillarin), and B23.

Patients with early PAH are generally asymptomatic. The initial symptom is typically exertional dyspnea and reduced exercise capacity. With progression, angina, exertional near-syncope, and symptoms and signs of right-sided heart failure appear. Physical examination shows tachypnea, a prominent pulmonic S_2 heart sound, palpable right ventricular heave, elevated jugular venous pressure, and dependent edema. Doppler echocardiography provides a noninvasive method for estimating the pulmonary arterial pressure is widely used to screen and for pulmonary hypertension. Echocardiographic estimates of pulmonary arterial systolic pressures exceeding 40 mmHg at rest suggest PAH. Pulmonary function testing may show a reduced DLCO in isolation or combined with a restrictive pattern. Because echocardiography can result in over- or underestimation of pulmonary arterial pressures in SSc, right heart catheterization is always required to confirm the presence of PAH and accurately assess its severity, including the degree of right heart dysfunction. Serum levels of brain natriuretic peptide (BNP) and N-terminal pro-BNP correlate with the presence and severity of PAH in SSc, as well as survival. Therefore, BNP measurements can be useful in screening SSc patients and in monitoring the response to treatment. The prognosis of PAH is determined by the degree of pulmonary arterial pressure elevation.

Gastrointestinal Involvement

The gastrointestinal tract is affected in up to 90% of SSc patients with both limited and diffuse cutaneous forms of the disease. The pathologic features of atrophy of smooth muscle, intact mucosa, and obliterative small-vessel vasculopathy are similar throughout the length of the gastrointestinal tract.

UPPER GASTROINTESTINAL TRACT INVOLVEMENT

Oropharyngeal manifestations due to a combination of xerostomia, reduced oral aperture, periodontal disease and resorption of the mandibular condyles are frequent. The frenulum of the tongue may be shortened. Symptoms of gastroesophageal reflux disease (GERD) develop early. Most patients have heartburn, regurgitation, and dysphagia. A combination of reduced lower esophageal sphincter pressure resulting in gastroesophageal reflux, impaired esophageal clearance of refluxed gastric contents due to diminished motility in the distal two-thirds of the esophagus, and delayed gastric emptying accounts for GERD. Chest CT scan characteristically shows a dilated esophagus with intraluminal air. Severe erosive esophagitis may be found on endoscopy in patients with minimal symptoms. Endoscopy may be necessary to rule out opportunistic infections with candida, herpes virus, and cytomegalovirus. Esophageal strictures and Barrett's esophagus may complicate chronic GERD. Because Barrett's esophagus is associated with an increased risk of adenocarcinoma, SSc patients with this lesion need to undergo periodic endoscopy and biopsy. Extraesophageal manifestations of GERD such as hoarseness and chronic cough may occur. Aspiration pneumonitis also may occur, aggravating underlying ILD.

Gastroparesis with early satiety, abdominal distention, and aggravated reflux symptoms is common. The presence and severity of gastroparesis can be assessed by radionuclide gastric emptying studies. Gastric antral vascular ectasia (GAVE) in the antrum may occur. These subepithelial lesions, reflecting the diffuse small-vessel vasculopathy of SSc, are described as "watermelon stomach" due to their endoscopic appearance. Patients with GAVE can have recurrent episodes of gastrointestinal bleeding, resulting in chronic unexplained anemia. Manometric testing shows abnormalities in the upper small intestines of most patients with SSc.

LOWER GASTROINTESTINAL TRACT INVOLVEMENT

Impaired intestinal motility may result in malabsorption and chronic diarrhea secondary to bacterial overgrowth. Fat and protein malabsorption and B_{12} and vitamin D deficiency ensue, sometimes culminating in severe malnutrition. Disturbed intestinal motor function can also cause intestinal pseudoobstruction, with symptoms of nausea and abdominal distension that are indistinguishable from those of delayed gastric emptying. Patients present with recurrent episodes of acute abdominal pain, nausea, and vomiting. Radiographic studies show acute intestinal obstruction, and the major diagnostic challenge is to differentiate pseudoobstruction, which responds to supportive care and intravenous nutritional supplementation, from mechanical obstruction. Colonic involvement may cause severe constipation, fecal incontinence, gastrointestinal bleeding from telangiectasia, and rectal prolapse. In late-stage SSc, wide-mouth sacculations or diverticula occur in the colon, occasionally causing perforation and bleeding. An occasional radiologic finding is pneumatosis cystoides intestinalis

due to air trapping in the bowel wall that may rarely rupture and cause benign pneumoperitoneum. Although the liver is rarely affected, primary biliary cirrhosis may coexist with SSc.

Renal Involvement: Scleroderma Renal Crisis

Scleroderma renal crisis, the most dreaded complication of SSc, occurs in 10–15% of patients, and almost always within 4 years of the onset of the disease. Prior to the advent of ACE inhibitors, short-term survival in scleroderma renal crisis was <10%. The pathogenesis involves obliterative vasculopathy and luminal narrowing of the renal arcuate and interlobular arteries. Progressive reduction in renal blood flow, aggravated by vasospasm, leads to juxtaglomerular hyperplasia, increased renin secretion, and activation of angiotensin, with further renal vasoconstriction resulting in a vicious cycle that culminates in malignant hypertension. Risk factors for scleroderma renal crisis include African American race, male gender, diffuse cutaneous SSc with extensive and progressive skin involvement, and autoantibodies to RNA polymerases I and III. Palpable tendon friction rubs, pericardial effusion, new unexplained anemia, and thrombocytopenia may be harbingers of impending scleroderma renal crisis. High-risk patients with early SSc should be counseled to check their blood pressure daily. Patients with lcSSc infrequently develop scleroderma renal crisis. Because there is an association between glucocorticoid use and the onset of scleroderma renal crisis, prednisone should be used in high-risk SSc patients only when absolutely required, and at low doses (<10 mg/d).

Patients characteristically present with accelerated hypertension and progressive renal insufficiency. However, in approximately 10% of patients, blood pressure remains normal. Normotensive renal crisis is generally associated with a poor outcome. Headache, blurred vision, and chest pain may accompany elevation of blood pressure. Urinalysis typically shows mild proteinuria, granular casts, and microscopic hematuria; thrombocytopenia and microangiopathic hemolysis with fragmented red blood cells can be seen. Progressive oliguric renal failure over several days generally follows. In some cases, scleroderma renal crisis is misdiagnosed as thrombotic thrombocytopenic purpura. The value of kidney biopsy in this setting is uncertain. Oliguria or a creatinine >3 mg/dL at presentation predicts poor outcome, with permanent hemodialysis and high mortality. Prompt aggressive intervention with short-acting ACE inhibitors to achieve adequate blood pressure control before the onset of renal failure results in improved prognosis. In contrast, there is no evidence to support the practice of "prophylactic" use of ACE inhibitors in normotensive SSc patients. Rarely, crescentic glomerulonephritis occurs in the setting of SSc.

Cardiac Involvement

Although cardiac involvement is often clinically silent, it is frequently detected when sensitive diagnostic tools are used. Cardiac disease occurs more frequently in patients with dcSSc than in those with lcSSc, and generally develops within 3 years of the onset of skin thickening. Clinically evident cardiac involvement in SSc is a poor prognostic factor. The endocardium, myocardium, and pericardium may be affected separately or together. Manifestations include pericardial effusions, atrial and ventricular tachycardias, conduction abnormalities, valvular regurgitation, hypertrophy, and heart failure. Systemic and pulmonary hypertension and lung and renal involvement may also impact on the heart. Despite the presence of widespread obliterative vasculopathy, the frequency of clinical or pathologic epicardial coronary artery disease in SSc is not increased. While conventional echocardiography has low sensitivity for detecting SSc preclinical heart involvement, newer modalities such as tissue Doppler echocardiography (TDE) and cardiac magnetic resonance imaging (MRI) reveal a high prevalence of abnormal myocardial function. Thallium perfusion studies document abnormal cardiac perfusion in a majority of patients. The serum level of N-terminal pro-brain natriuretic peptide (NT-pro-BNP), a ventricular hormone, is a sensitive and specific diagnostic marker for increased pulmonary artery pressure in SSc, but may also have utility as a marker of primary cardiac involvement. Myocarditis can occur in association with inflammatory polymyositis, and can be diagnosed using cardiac MRI. Pericardial effusions occur in over 15% of patients and, rarely, may cause tamponade.

Musculoskeletal Complications

Carpal tunnel syndrome occurs frequently and may be a presenting manifestation of SSc. Generalized arthralgia and stiffness are prominent in early disease. Joint mobility is progressively impaired, especially in patients with dcSSc. Most commonly affected are the hands. Contractures develop at the proximal interphalangeal joints and wrists. In patients with dcSSc, large joint contractures can be accompanied by tendon friction rubs characterized by leathery crepitation that can be heard or palpated upon passive movement. Tendon rubs are due to extensive fibrosis and adhesion of the tendon sheaths and fascial planes at the affected joint. Movement at the elbows, shoulders, and knees is frequently reduced. True joint inflammation is uncommon; however, occasional patients develop erosive polyarthritis in the hands. Muscle weakness is common, and may indicate deconditioning, disuse atrophy, and malnutrition. Less commonly, inflammatory myositis indistinguishable from idiopathic polymyositis may occur. A chronic noninflammatory myopathy characterized by atrophy and fibrosis in the absence of elevated muscle enzyme levels can be seen in late-stage SSc. Bone resorption occurs most commonly in the terminal phalanges, where it causes loss of the distal tufts (acro-osteolysis) (Fig. 323-5). Resorption of the mandibular condyles can lead to bite difficulties.

Osteolysis can also affect the ribs and distal clavicles.

Other Disease Manifestations

Many SSc patients develop dry eyes and dry mouth (sicca complex). Biopsy of the minor salivary glands shows fibrosis rather than focal lymphocytic infiltration characteristic of primary Sjögren's syndrome **(Chap. 324)**. Hypothyroidism is common and generally due to fibrosis of the thyroid gland. Whereas the central nervous system is generally spared in SSc, sensory trigeminal neuropathy due to fibrosis or vasculopathy can occur, presenting with gradual onset of pain and numbness. Pregnancy in women with SSc is associated with an increased rate of adverse fetal outcomes. Furthermore, cardiopulmonary involvement may worsen during pregnancy, and new onset of scleroderma renal crisis has been described. Erectile dysfunction is frequent in men with SSc and may be the initial disease manifestation. Inability to attain or maintain penile erection is due to vascular insufficiency and fibrosis. The risk of certain malignancies is increased in SSc. Some studies have indicated that cancers of the lung, tongue, and breast occur more frequently in patients with SSc. Barrett's metaplasia is associated with increased risk for adenocarcinoma of the esophagus.

Laboratory Features

A mild normocytic or microcytic anemia due to chronic inflammation is frequent in patients with SSc. Iron deficiency anemia may indicate gastrointestinal bleeding caused by GAVE or chronic esophagitis. Macrocytic anemia, indicating a maturation disorder, may be caused by folate and vitamin B12 deficiency due to small bowel bacterial overgrowth and malabsorption, or by drugs such as methotrexate or alkylating agents. Microangiopathic hemolytic anemia, caused by mechanical trauma and fragmentation of red blood cells during their passage through microvessels coated with fibrin or platelet thrombi, is a hallmark of scleroderma renal crisis. Thrombocytopenia and leukopenia generally indicate drug toxicity. In contrast to other connective tissue diseases, the erythrocyte sedimentation rate (ESR) is generally normal; an elevation may signal coexisting myositis or malignancy.

Antinuclear autoantibodies are present in almost all patients with SSc and can be detected at disease onset. Autoantibodies against topoisomerase-I (Scl-70) and centromere are specific for SSc and are mutually exclusive. Topoisomerase-I antibodies are detected in 31% of patients with dcSSc, but in only 13% of patients with lcSSc; conversely, anticentromere antibodies are detected in 38% of patients with lcSSc, but in only 2% of patients with dcSSc. Anticentromere antibodies are commonly associated with lcSSc and PAH, and only rarely with cardiac or renal involvement or significant ILD. Topoisomerase-I-positive patients have reduced survival compared to those without this antibody; whereas anticentromere antibody-positive patients have improved survival compared to those without this antibody. Nucleolar immunofluorescence pattern on serologic testing reflects antibodies to U3-RNP (fibrillarin), Th/To, or PM/Scl, whereas a speckled immunofluorescence pattern indicates antibodies to RNA polymerase III. Although antibodies to ß2GPI occur in antiphospholipid antibody syndrome and are not specific for SSc, their presence in SSc is associated with an increased risk of ischemic lesions in the fingers. No direct pathogenic role has been firmly established for any of the SSc-associated autoantibodies; however, antibody titers can correlate with disease severity and fluctuate with disease activity.

Diagnosis

The diagnosis of SSc is made primarily on clinical grounds and is generally straightforward in patients with established disease. The presence of skin induration, with a characteristic symmetric distribution pattern associated with typical visceral organ manifestations, establishes the diagnosis with a high degree of certainty. While the conditions listed in Table 323-1 can be associated with skin induration, the distribution pattern of skin lesions, together with the absence of Raynaud's phenomenon or typical visceral organ manifestations or SSc-specific autoantibodies, differentiates these conditions from SSc. Occasionally, full-thickness biopsy of the skin is required for establishing the diagnosis of scleredema, scleromyxedema, or nephrogenic systemic fibrosis. In lcSSc a history of antecedent Raynaud's phenomenon and gastroesophageal reflux symptoms, coupled with the presence of sclerodactyly and capillary changes on nailfold capillaroscopy, often in combinations with telangiectasia and calcinosis cutis, helps to establish the diagnosis. The finding of digital tip pitting scars and radiologic evidence of pulmonary fibrosis in the lower lobes are particularly helpful diagnostically. Primary Raynaud's phenomenon is a common benign condition that must be differentiated from early or limited SSc. Nailfold microscopy is particularly helpful in this situation, because in primary Raynaud's phenomenon the nailfold capillaries are normal, whereas in SSc capillary abnormalities, as well as serum autoantibodies can be detected even before other disease manifestations.

Establishing the diagnosis of SSc in early disease may be a challenge. In dcSSc, initial symptoms are often nonspecific and relate to inflammation. Patients complain of fatigue, swelling, aching, and stiffness, and Raynaud's phenomenon may initially be absent. Physical examination may reveal diffuse upper extremity edema and puffy fingers. Patients at this stage are sometimes diagnosed as early rheumatoid arthritis, systemic lupus erythematosus, myositis, or, most commonly, undifferentiated connective tissue disease. Within weeks to months, Raynaud's phenomenon and characteristic clinical features appear accompanied by advancing induration of the skin. The presence of antinuclear and SSc-specific autoantibodies provides a high degree of diagnostic specificity. Raynaud's phenomenon

with fingertip ulcerations or other evidence of digital ischemia, coupled with telangiectasia, distal esophageal dysmotility, unexplained ILD or PAH, or accelerated hypertension with renal failure in the absence of skin induration, suggests the diagnosis of SSc sine scleroderma. These patients may have anticentromere antibodies.

Treatment: Systemic Sclerosis

OVERVIEW

To date, no therapy has been shown to significantly alter the natural history of SSc. In contrast, multiple interventions are highly effective in alleviating the symptoms and in slowing the progression of the cumulative organ damage. A significant reduction in disease-related mortality has been noted during the past 25 years. Because of the marked heterogeneity in clinical presentations, patients need careful investigation at baseline, and evaluation and treatment approaches must be individually tailored according to each patient's unique needs. Optimal management incorporates the following principles: prompt and accurate diagnosis; classification and risk stratification based on clinical and laboratory evaluation; early recognition of organ-based complications and assessment of their extent, severity, and likelihood of deterioration; regular monitoring for progression, disease activity, and response to therapy; and continuing patient education. In order to minimize irreversible organ damage, the management of life-threatening complications must be proactive, with regular screening and initiation of appropriate intervention at the earliest possible opportunity. In light of the complex, multisystemic nature of the disease, an integrated team-based approach is optimal. Most patients are treated with combinations of drugs that act upon different aspects of the disease. Patients should become familiar with the spectrum of potential complications, have an understanding of the therapeutic options and natural history of the disease, and be empowered to partner with their physicians. This typically requires a long-term relationship between patient and physician, with ongoing counseling and encouragement.

DISEASE-MODIFYING THERAPY: IMMUNOSUPPRESSIVE AGENTS

Immunosuppressive agents effective in other connective tissue diseases have generally shown modest or no benefit in the treatment of SSc. Glucocorticoids may be useful for alleviating stiffness and aching in early-stage dcSSc, but do not influence the progression of skin or internal organ involvement. Furthermore, their use in high doses is associated with an increased risk of scleroderma renal crisis. Therefore, glucocorticoids should be avoided if possible; when absolutely necessary, they should be given at the lowest dose possible and for brief periods only. The use of cyclophosphamide has been extensively studied in light of its efficacy in the treatment of vasculitis **(Chap. 326)**, systemic lupus erythematosus **(Chap. 319)**, and other autoimmune diseases **(Chap. 318)**.

Cyclophosphamide has been evaluated in the treatment of SSc in retrospective and prospective controlled clinical trials. Both oral and intermittent intravenous cyclophosphamide were shown to reduce the progression of ILD in SSc patients with early symptomatic disease, with stabilization, and, rarely, modest improvement of pulmonary function and HRCT after 1 year of treatment. Improvement in respiratory symptoms and the extent of skin induration was also noted. The beneficial effect of cyclophosphamide on lung function wanes upon discontinuation of therapy. The benefits of cyclophosphamide need to be balanced against its potential toxicity, including bone marrow suppression, opportunistic infections, hemorrhagic cystitis and bladder cancer, premature ovarian failure, and late secondary malignancies.

In small clinical trials in SSc, methotrexate treatment was associated with a modest improvement in skin scores. Mycophenolate mofetil treatment was associated with improved skin induration in uncontrolled studies and was generally well tolerated. The use of immunomodulatory agents such as cyclosporine, azathioprine, rituximab, extracorporeal photophoresis, imatinib, thalidomide, or rapamycin for the treatment of SSc is currently not well supported by the literature. Immune ablation using high-dose chemotherapy with or without irradiation, followed by autologous stem cell reconstitution, is undergoing evaluation in randomized clinical trials in SSc. In light of its potential morbidity and mortality, as well as cost, autologous stem cell transplantation in SSc is still considered experimental.

Anti-Fibrotic Therapy

Because widespread tissue fibrosis causes progressive organ damage in dcSSc, drugs that interfere with the fibrotic process represent a rational approach to therapy. D-Penicillamine has been extensively used as an antifibrotic agent. Retrospective studies in SSc indicated that D-penicillamine stabilized and improved skin induration, prevented new internal organ involvement, and improved survival. However, a randomized controlled clinical trial in early active SSc found no difference in the extent of skin involvement between patients treated with standard-dose (750 mg/d) or very low-dose (125 mg every other day) D-penicillamine. Minocycline, recombinant relaxin, interferon (INF)-γ, and inhibitors of tumor necrosis factor have failed to show meaningful clinical benefit in SSc.

Vascular Therapy

The goal of vascular therapy is to control Raynaud's phenomenon, prevent the development and enhance the healing of ischemic complications, and slow the progression of obliterative vasculopathy. Patients with Raynaud's phenomenon should dress warmly, minimize cold exposure or stress, and

avoid drugs that could precipitate or exacerbate vasospastic episodes. Some patients may respond to biofeedback therapy. Calcium channel blockers such as nifedipine or diltiazem are commonly used but show only moderate benefit, and their use is often limited by side effects (palpitations, dependent edema, light-headedness). While ACE inhibitors do not reduce the frequency or severity of episodes, angiotensin II receptor blockers such as losartan are effective and generally well tolerated. Some patients with Raynaud's phenomenon may require α_1-adrenergic receptor blockers (e.g., prazosin), 5-phosphodiesterase inhibitors (e.g., sildenafil), serotonin reuptake inhibitors (e.g., fluoxetine), topical nitroglycerine, and intravenous prostaglandins. Low-dose aspirin and dipyridamole prevent platelet aggregation and may have a role as adjunctive agents. In patients with ischemic finger ulcerations, the endothelin-1 receptor antagonist bosentan reduces the development of new ulcers. Digital sympathectomy and local injections of botulinum type A (botox) into the digits are options in some patients with severe Raynaud's phenomenon associated with ischemia. Empirical long-term therapy with statins and antioxidants may delay the progression of vascular damage and obliteration. The use of calcium channel blockers has been associated with improved cardiac perfusion and cardiac function in SSc patients with cardiac involvement.

TREATMENT OF GASTROINTESTINAL COMPLICATIONS

Because gastroesophageal reflux is very common, all patients with SSc should be treated for this complication. Significant reflux may occur in the absence of symptoms. Patients should be instructed to elevate the head of the bed and eat frequent small meals. Proton pump inhibitors reduce acid reflux and may need to be given in relatively high doses. Recurrent gastrointestinal bleeding from vascular ectasia in the gastric antrum (watermelon stomach) is amenable to treatment with laser photocoagulation. Bacterial overgrowth due to small-bowel dysmotility causes abdominal bloating and diarrhea and may lead to malabsorption and severe malnutrition. Treatment with short courses of rotating broad-spectrum antibiotics such as metronidazole, erythromycin, and tetracycline can eradicate bacterial overgrowth. Parenteral hyperalimentation is indicated if malnutrition develops. Chronic hypomotility of the small bowel may respond to octreotide.

TREATMENT OF PULMONARY ARTERIAL HYPERTENSION (PAH)

Because PAH is asymptomatic until it is advanced, patients with SSc should be screened for the presence of PAH on a regular basis. When PAH is symptomatic, treatment should be started with an oral endothelin-1 receptor antagonist or a phosphodiesterase inhibitor such as sildenafil. Patients may also require diuretics, oral anticoagulation, and digoxin when appropriate. If hypoxemia is documented, supplemental oxygen should be given by nasal cannula in order to avoid hypoxia-induced secondary pulmonary vasoconstriction. Inhibitors of phosphodiesterase type 5 (e.g., sildenafil) have been shown to have short-term efficacy in PAH and may be used in combination with bosentan. Prostacyclin analogues such as epoprostenol or treprostinil can be administered intravenously or by continuous subcutaneous infusion, or frequent inhalations via nebulizer. Lung transplantation remains an option for patients with SSc-associated PAH who fail medical therapy.

TREATMENT OF RENAL CRISIS

Scleroderma renal crisis is a medical emergency because the outcome is largely determined by the extent of renal damage present at the time that aggressive therapy is initiated. Prompt recognition of impending or early scleroderma renal crisis is therefore essential, and efforts should be made to avoid its occurrence. High risk patients with early SSc and extensive and progressive skin involvement should be instructed to monitor their blood pressure daily and report significant alterations immediately. Potentially nephrotoxic drugs should be avoided, and glucocorticoids used only when absolutely necessary, and at low doses. When scleroderma renal crisis occurs, treatment should be started promptly with short-acting ACE inhibitors, with the goal of achieving rapid normalization of the blood pressure. Kidney biopsy is rarely useful in this setting. Up to two-thirds of patients require dialysis. However, substantial renal recovery can occur following renal crisis, and up to one-half of the patients may be able to discontinue dialysis. Kidney transplantation is appropriate for patients who are unable to discontinue dialysis after 2 years. Survival of SSc patients with with renal transplantation is comparable to that in other connective tissue diseases, and recurrence of renal crisis is rare.

SKIN CARE

Because skin involvement in SSc is never life-threatening, and it stabilizes, and may even regress spontaneously over time, the overall management of the disease should not be dictated by its cutaneous manifestations. The inflammatory symptoms of early skin involvement can be effectively controlled with systemic antihistamines and cautious and short-term use of low-dose glucocorticoids (<5 mg/d of prednisone). Retrospective studies have shown that D-penicillamine reduced the extent and progression of skin induration; however, these benefits could not be substantiated in a controlled prospective trial. Cyclophosphamide and methotrexate have also been shown to have modest effects on skin induration. Because induration is associated with dryness, patients should use hydrophilic ointments and bath oils. Regular skin massage is helpful. Telangiectasia may present a cosmetic problem, especially when they occur on the face. Treatment with pulsed dye laser may have short-term benefit. Fingertip ulcerations should be protected by occlusive dressing to promote healing and prevent infection. Infected skin ulcers are treated with topical antibiotics. Surgical debridement may be

indicated. No therapy has been shown to be effective in preventing the formation of calcific soft tissue deposits or in promoting their dissolution.

COURSE

The natural history of SSc is highly variable and difficult to predict, especially in early stages when the specific disease subset—diffuse or limited cutaneous form—is not clear. Patient with dcSSc have a more rapidly progressive disease and worse prognosis than those with lcSSc.

In dcSSc, early inflammatory symptoms such as fatigue, edema, arthralgia, and pruritus tend to subside 2–4 years after the onset of disease, and the extent of skin thickening reaches a plateau after which it generally shows slow regression. It is during the early edematous stage, generally lasting <3 years, that visceral organ involvement develops and progresses. While existing visceral organ involvement, such as pulmonary fibrosis, may continue to progress, new organ involvement is rare after the skin involvement reaches its peak. Scleroderma renal crisis almost invariably occurs within the first 4 years of disease. In dcSSc patients with late-stage disease (>6 years), the skin is usually soft and atrophic. Skin regression characteristically occurs in an order that is the reverse of initial involvement, with softening on the trunks followed by proximal and then distal extremities. Sclerodactyly and finger contractures generally persist. Cutaneous telangiectasia and calcinosis are common, making it difficult to differentiate late-stage dcSSc from lcSSc. Relapse or recurrence of skin thickening after the peak of skin involvement has been reached is rare. Patients with lcSSc follow a clinical course that is markedly different than that of dcSSc. In this subset of SSc, Raynaud's phenomenon typically precedes other disease manifestations by years or even decades. Visceral organ complications such as PAH and ILD generally develop late and tend to be slowly progressive.

PROGNOSIS

SSc confers a substantial increase in the risk of premature death, with age- and gender-adjusted mortality rates that are fivefold to eightfold higher compared to the general population. In one population-based study of SSc patients with all forms of the disease, the median survival was 11 years. In patients with dcSSc, 5- and 10-year survivals are 70% and 55%, respectively, whereas in patients with lcSSc, 5- and 10-year survivals are 90% and 75%, respectively. The prognosis of SSc correlates with the extent of skin involvement, which itself is a surrogate for visceral organ involvement. Major causes of death are PAH, pulmonary fibrosis, gastrointestinal involvement, and cardiac disease. Scleroderma renal crisis is associated with a 30% 3-year mortality. Lung cancer and excess cardiovascular deaths also contribute to increased mortality. Markers of worse prognosis include male gender, African American race, older age of disease onset, extensive skin thickening with truncal involvement, evidence of significant or progressive visceral organ involvement, and the presence of anti-topoisomerase-I and anti-RNA polymerase III antibodies. Additional predictors of increased mortality at initial evaluation include an elevated ESR, anemia, and proteinuria. In one study, SSc patients with extensive skin involvement, lung vital capacity <55% predicted, significant gastrointestinal involvement (pseudoobstruction or malabsorption), evidence of cardiac involvement (arrhythmias or congestive heart failure), or scleroderma renal crisis had a cumulative 9-year survival <40%. The severity of PAH is itself strongly associated with mortality, and SSc patients who had a mean pulmonary arterial pressure ≥45 mmHg had a 33% 3-year survival. The advent of ACE inhibitor therapy for scleroderma renal crisis had a dramatic impact on survival, increasing from a <10% 1-year survival in the pre-ACE inhibitor era to a >70% 3-year survival at the present time.

MIXED CONNECTIVE TISSUE DISEASE

Patients who have lcSSc coexisting with features of SLE, polymyositis, and rheumatoid arthritis may have mixed connective tissue disease (MCTD). This overlap syndrome is generally associated with the presence of high titers of autoantibodies to U1-RNP. The characteristic initial presentation is Raynaud's phenomenon associated with puffy fingers and myalgia. Gradually, lcSSc features of sclerodactyly, calcinosis, and cutaneous telangiectasia develop. Skin rashes suggestive of systemic lupus erythematosus (malar rash, photosensitivity) or of dermatomyositis (heliotrope rash on the eyelids, erythematous rash on the knuckles) occur. Arthralgia is common, and some patients develop erosive polyarthritis. Pulmonary fibrosis and isolated or secondary PAH may develop. Other manifestations include esophageal dysmotility, pericarditis, Sjögren's syndrome, and renal disease, especially membranous glomerulonephritis. Laboratory evaluation indicates features of inflammation with elevated ESR and hypergammaglobulinemia. While anti-U1RNP antibodies are detected in the serum in high titers, SSc-specific autoantibodies are not found. In contrast to SSc, patients with MCTD often show a good response to treatment with glucocorticoids, and the long-term prognosis is better than that of SSc. Whether MCTD is a truly distinct entity or is, rather, a subset of SLE or SSc remains controversial.

EOSINOPHILIC FASCIITIS

Eosinophilic fasciitis is a rare idiopathic disorder associated with induration of the skin that generally develops rapidly. Adults are primarily affected. The skin has a coarse cobblestone "peau d'orange"

appearance. In contrast to SSc, internal organ involvement is rare, and Raynaud's phenomenon and SSc-associated autoantibodies are absent. Furthermore, skin involvement spares the fingers. Full-thickness excisional biopsy of the lesional skin reveals fibrosis of the subcutaneous fascia, and is generally required for diagnosis. Inflammation and eosinophil infiltration in the fascia are variably present. In the acute phase of the illness, peripheral blood eosinophilia may be prominent. MRI appears to be a sensitive tool for the diagnosis of eosinophilic fasciitis. In some patients, eosinophilic fasciitis occurs in association with, or preceding, myelodysplastic syndromes or multiple myeloma. Treatment with glucocorticoids leads to prompt resolution of the eosinophilia. In contrast, skin changes generally show slow and variable improvement. The prognosis of patients with eosinophilic fasciitis is good.

FURTHER READINGS

Feghali-Bostwick C et al: Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. Arthritis Rheum 48:1956, 2003[PMID: 12847690] [Full Text]

Kuwana M et al: Defective vasculogenesis in systemic sclerosis. Lancet 364:603, 2004[PMID: 15313361] [Full Text]

Mayes MD et al: Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 48:2246, 2003[PMID: 12905479] [Full Text]

Milano A et al: Molecular subsets in the gene expression signatures of scleroderma skin. PLoS ONE 3:e2696, 2008; erratum: PLoS ONE 3:doi: 10.1371/annotation/05bed72c-c6f6-4685-a732-02c78e5f66c2, 2008

Scussel-Lonzetti L et al: Predicting mortality in systemic sclerosis: Analysis of a cohort of 309 French Canadian patients with emphasis on features at diagnosis as predictive factors for survival. Medicine (Baltimore) 81:154, 2002[PMID: 11889414] [Full Text]

Tashkin DP et al: Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 354:2655, 2006[PMID: 16790698] [Full Text]

———: Effects of 1-year treatment with cyclophosphamide on outcomes at 2 years in scleroderma lung disease. Am J Respir Crit Care Med 176:1026, 2007

Varga J, Abraham D: Systemic sclerosis: A prototypic multisystem fibrotic disorder. J Clin Invest 117:557, 2007[PMID: 17332883] [Full Text]

Whitfield ML et al: Systemic and cell type-specific gene expression patterns in SSc skin. Proc Natl Acad Sci U S A 100:12319, 2003[PMID: 14530402] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE SPONDYLOARTHRITIDES: INTRODUCTION

The spondyloarthritides are a group of overlapping disorders that share certain clinical features and genetic associations. These disorders include ankylosing spondylitis, reactive arthritis, psoriatic arthritis and spondylitis, enteropathic arthritis and spondylitis, juvenile-onset spondyloarthritis (SpA), and undifferentiated SpA. The similarities in clinical manifestations and genetic predisposition suggest that these disorders share pathogenic mechanisms.

ANKYLOSING SPONDYLITIS

Ankylosing spondylitis (AS) is an inflammatory disorder of unknown cause that primarily affects the axial skeleton; peripheral joints and extraarticular structures are also frequently involved. The disease usually begins in the second or third decade; male to female prevalence is between 2:1 and 3:1. The term *axial spondyloarthritis*, coming into common use, includes early or mild forms that do not meet classical criteria for AS.

Epidemiology

AS shows a striking correlation with the histocompatibility antigen HLA-B27 and occurs worldwide roughly in proportion to the prevalence of B27 **(Chap. 315)**. In North American whites, the prevalence of B27 is 7%, whereas it is 90% in patients with AS, independent of disease severity.

In population surveys, AS is present in 1–6% of adults inheriting B27, whereas the prevalence is 10–30% among B27+ adult first-degree relatives of AS probands. Concordance rate in identical twins is about 65%. Susceptibility to AS is determined largely by genetic factors, with B27 comprising up to one-half of the genetic component. Other HLA-linked genes may also contribute to susceptibility to AS. Genome-wide single-nucleotide polymorphism (SNP) analysis has identified additional susceptibility alleles in the genes encoding ERAP1 (chromosome 5q15) and IL-23R (chromosome 1p31.3). The genes encoding TNFSF15, TNFSF1A, STAT3, ANTXR2, and IL1R2, and at least six other chromosomal regions have also been implicated.

Pathology

The sites of axial inflammation in AS are inaccessible to routine biopsy and are rarely approached surgically. Knowledge of the axial histopathology is therefore based mostly on advanced cases. Sacroiliitis is often the earliest manifestations of AS. Synovitis, pannus, myxoid marrow, subchondral granulation tissue and marrow edema, enthesitis, and chondroid differentiation are found. Macrophages, T cells, and osteoclasts are prevalent. Eventually the eroded joint margins are gradually replaced by fibrocartilage regeneration and then by ossification. The joint may become totally obliterated.

In the spine, there is inflammatory granulation tissue at the junction of annulus fibrosis and vertebral bone. The outer annular fibers are eroded and eventually replaced by bone, forming the beginning of a syndesmophyte, which then grows by continued endochondral ossification, ultimately bridging the adjacent vertebral bodies. Ascending progression of this process leads to the "bamboo spine." Other lesions in the spine include diffuse osteoporosis, erosion of vertebral bodies at the disk margin, "squaring" or "barreling" of vertebrae, and inflammation and destruction of the disk-bone border. Inflammatory arthritis of the apophyseal joints is common, with erosion of cartilage by pannus, often followed by bony ankylosis. Bone mineral density is diminished in the spine and proximal femur early in the course of the disease.

Peripheral synovitis in AS shows marked vascularity, lining layer hyperplasia, lymphoid infiltration, and pannus formation. Central cartilaginous erosions caused by proliferation of subchondral granulation tissue are common.

Inflammation in the fibrocartilaginous enthesis, the region where a tendon, ligament, or joint capsule attaches to bone, is a characteristic lesion in AS and other SpA, both at axial and peripheral sites. Enthesitis is associated with prominent edema of the adjacent bone marrow and is often characterized by erosive lesions that eventually undergo ossification.

Pathogenesis

The pathogenesis of AS is thought to be immune-mediated, but there is no direct evidence for autoimmunity. There is uncertainty regarding the primary site of disease initiation. A unifying concept is

that the AS disease process begins at sites where articular cartilage, ligaments, and other structures attach to bone. The dramatic response of the disease to therapeutic blockade of tumor necrosis factor α (TNF-α) indicates that this cytokine plays a central role in the immunopathogenesis of AS. There is recent evidence that TH17 T cells and their cytokines may also play an important role.

The inflamed sacroiliac joint is infiltrated with CD4+ and CD8+ T cells and macrophages and shows high levels of TNF-α, particularly early in the disease. Abundant transforming growth factor β (TGF-β) has been found in more advanced lesions. Peripheral synovitis in AS and the other spondyloarthritides is characterized by neutrophils, macrophages expressing CD68 and CD163, CD4+ and CD8+ T cells, and B cells. There is prominent staining for intercellular adhesion molecule 1 (ICAM-1), vascular cell adhesion molecule 1 (VCAM-1), matrix metalloproteinase 3 (MMP-3), and myeloid-related proteins 8 and 14 (MRP-8 and MRP-14). Unlike rheumatoid arthritis (RA) synovium, citrullinated proteins and cartilage gp39 peptide–major histocompatibility complexes (MHCs) are absent. No specific event or exogenous agent that triggers the onset of disease has been identified, although overlapping features with reactive arthritis and inflammatory bowel disease (IBD) suggest that enteric bacteria may play a role. Triggering of innate immunity by microdamage at entheseal sites has recently been emphasized. Strong evidence that B27 plays a direct role is provided by genetic epidemiology studies and by the finding that rats transgenic for B27 spontaneously develop dramatic arthritis and spondylitis. However, the role of B27 remains unresolved. Since B27 rats lacking CD8+ T cells still develop arthritis and spondylitis, classical peptide antigen presentation to CD8+ T cells is probably not the primary disease mechanism. However, the association of AS with ERAP1, which strongly influences the MHC class I peptide repertoire, is only found in B27+ patients and suggests that peptide binding to B27 is nonetheless important. The B27 heavy chain has an unusual tendency to misfold, a process that may be proinflammatory. Genetic and functional studies in humans have suggested a role for natural killer (NK) cells in AS, possibly through interaction with B27. Defective dendritic cell function is a consistent feature of SpA-prone B27 rats not yet well investigated in patients.

Clinical Manifestations

The symptoms of the disease are usually first noticed in late adolescence or early adulthood; the median age in Western countries is 23. In 5% of patients, symptoms begin after age 40. The initial symptom is usually dull pain, insidious in onset, felt deep in the lower lumbar or gluteal region, accompanied by low-back morning stiffness of up to a few hours' duration that improves with activity and returns following inactivity. Within a few months, the pain has usually become persistent and bilateral. Nocturnal exacerbation of pain often forces the patient to rise and move around.

In some patients, bony tenderness (presumably reflecting enthesitis or osteitis) may accompany back pain or stiffness, while in others it may be the predominant complaint. Common sites include the costosternal junctions, spinous processes, iliac crests, greater trochanters, ischial tuberosities, tibial tubercles, and heels. Arthritis in the hips and shoulders ("root" joints) occurs in 25–35% of patients. Severe isolated hip arthritis or bony chest pain may be the presenting complaint. Arthritis of peripheral joints other than the hips and shoulders, usually asymmetric, occurs in up to 30% of patients. Neck pain and stiffness from involvement of the cervical spine are usually relatively late manifestations but are occasionally dominant symptoms. Rare patients, particularly in the older age group, present with predominantly constitutional symptoms.

AS often has a juvenile onset in developing countries. Peripheral arthritis and enthesitis usually predominate, with axial symptoms supervening in late adolescence.

Initially, physical findings mirror the inflammatory process. The most specific findings involve loss of spinal mobility, with limitation of anterior and lateral flexion and extension of the lumbar spine and of chest expansion. Limitation of motion is usually out of proportion to the degree of bony ankylosis, reflecting muscle spasm secondary to pain and inflammation. Pain in the sacroiliac joints may be elicited either with direct pressure or with stress on the joints. In addition, there is commonly tenderness upon palpation at the sites of symptomatic bony tenderness and paraspinous muscle spasm.

The modified Schober test is a useful measure of lumbar spine flexion. The patient stands erect, with heels together, and marks are made on the spine at the lumbosacral junction (identified by a horizontal line between the posterosuperior iliac spines) and 10 cm above. The patient then bends forward maximally with knees fully extended, and the distance between the two marks is measured. This distance increases by ≥5 cm in the case of normal mobility and by <4 cm in the case of decreased mobility. Chest expansion is measured as the difference between maximal inspiration and maximal forced expiration in the fourth intercostal space in males or just below the breasts in females, with the patient's hands resting on or just behind the head. Normal chest expansion is ≥5 cm.

Limitation or pain with motion of the hips or shoulders is usually present if these joints are involved. It should be emphasized that early in the course of mild cases, symptoms may be subtle and nonspecific, and the physical examination may be completely normal.

The course of the disease is extremely variable, ranging from the individual with mild stiffness and normal radiographs to the patient with a totally fused spine and severe bilateral hip arthritis,

accompanied by severe peripheral arthritis and extraarticular manifestations. Pain tends to be persistent early in the disease and then becomes intermittent, with alternating exacerbations and quiescent periods. In a typical severe untreated case with progression of the spondylitis to syndesmophyte formation, the patient's posture undergoes characteristic changes, with obliterated lumbar lordosis, buttock atrophy, and accentuated thoracic kyphosis. There may be a forward stoop of the neck or flexion contractures at the hips, compensated by flexion at the knees. Disease progression can be estimated clinically from loss of height, limitation of chest expansion and spinal flexion, and occiput-to-wall distance. Occasional individuals are encountered with advanced deformities who report having never had significant symptoms.

There is little consensus regarding the factors that predict disease progression and functional outcome. In some but not all studies, onset of AS in adolescence and early hip involvement correlate with a worse prognosis. AS in women tends to progress less frequently to total spinal ankylosis, although there is some evidence for an increased prevalence of isolated cervical ankylosis and peripheral arthritis in women. In industrialized countries, peripheral arthritis (distal to hips and shoulders) occurs in less than one-half of patients with AS, usually as a late manifestation, whereas in developing countries, the prevalence is much higher, with onset typically early in the disease course. Pregnancy has no consistent effect on AS, with symptoms improving, remaining the same, or deteriorating in about one-third of pregnant patients, respectively. Smoking correlates with adverse outcome.

The most serious complication of the spinal disease is spinal fracture, which can occur with even minor trauma to the rigid, osteoporotic spine. The lower cervical spine is most commonly involved. These fractures are often displaced and cause spinal cord injury. A recent survey suggested a >10% lifetime risk of fracture. Occasionally, fracture through a diskovertebral junction and adjacent neural arch, termed *pseudoarthrosis*, most common in the thoracolumbar spine, can be an unrecognized source of persistent localized pain and/or neurologic dysfunction. Wedging of thoracic vertebrae is common and correlates with accentuated kyphosis.

The most common extraarticular manifestation is acute anterior uveitis, which occurs in 40% of patients and can antedate the spondylitis. Attacks are typically unilateral, causing pain, photophobia, and increased lacrimation. These tend to recur, often in the opposite eye. Cataracts and secondary glaucoma are not uncommon sequelae. Up to 60% of patients have inflammation in the colon or ileum. This is usually asymptomatic, but frank IBD occurs in 5–10% of patients with AS (see "Enteropathic Arthritis," below). About 10% of patients meeting criteria for AS have psoriasis (see "Psoriatic Arthritis," below). Aortic insufficiency, sometimes leading to congestive heart failure, occurs in a few percent of patients, occasionally early in the course of the spinal disease but usually after prolonged disease. Third-degree heart block may occur alone or together with aortic insufficiency. Subclinical pulmonary lesions and cardiac dysfunction may be relatively common. Cauda equina syndrome and upper pulmonary lobe fibrosis are rare late complications. Retroperitoneal fibrosis is a rare associated condition. Prostatitis has been reported to have an increased prevalence. Amyloidosis is rare **(Chap. 112)**.

Several validated measures of disease activity and functional outcome are in widespread use in the study and management of AS, particularly the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), a measure of disease activity; the Bath Ankylosing Spondylitis Functional Index (BASFI), a measure of limitation in activities of daily living; and several measures of radiographic changes. Despite persistence of the disease, most patients remain gainfully employed. Some but not all studies of survival in AS have suggested that AS shortens life span, compared with the general population. Mortality attributable to AS is largely the result of spinal trauma, aortic insufficiency, respiratory failure, amyloid nephropathy, or complications of therapy such as upper gastrointestinal hemorrhage. The impact of anti-TNF therapy on outcome and mortality is not yet known, but there is evidence for significantly improved work productivity.

Laboratory Findings

No laboratory test is diagnostic of AS. In most ethnic groups, HLA-B27 is present in 80–90% of patients. Erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) are often, but not always, elevated. Mild anemia may be present. Patients with severe disease may show an elevated alkaline phosphatase level. Elevated serum IgA levels are common. Rheumatoid factor, anti-cyclic citrullinated peptide (CCP), and antinuclear antibodies (ANAs) are largely absent unless caused by a coexistent disease, although ANAs may appear with anti-TNF therapy. Synovial fluid from peripheral joints in AS is nonspecifically inflammatory. In cases with restriction of chest wall motion, decreased vital capacity and increased functional residual capacity are common, but airflow is normal and ventilatory function is usually well maintained.

Radiographic Findings

Radiographically demonstrable sacroiliitis is eventually present in AS. The earliest changes by standard radiography are blurring of the cortical margins of the subchondral bone, followed by erosions and sclerosis. Progression of the erosions leads to "pseudowidening" of the joint space; as fibrous and then bony ankylosis supervene, the joints may become obliterated. The changes and

progression of the lesions are usually symmetric.

In the lumbar spine, progression of the disease leads to straightening, caused by loss of lordosis, and reactive sclerosis, caused by osteitis of the anterior corners of the vertebral bodies with subsequent erosion, leading to "squaring" or even "barreling" of one or more vertebral bodies. Progressive ossification leads to eventual formation of marginal syndesmophytes, visible on plain films as bony bridges connecting successive vertebral bodies anteriorly and laterally.

In many cases, years may elapse before unequivocal sacroiliac abnormalities are evident on plain radiographs, and consequently magnetic resonance imaging (MRI) is being increasingly used in diagnosing AS. Active sacroiliitis is best visualized by dynamic MRI with fat saturation, either T2-weighed turbo spin-echo sequence or short tau inversion recovery (STIR) with high resolution, or T1-weighted images with contrast enhancement. These techniques are much more sensitive than conventional radiography for identifying early intraarticular inflammation, cartilage changes, and underlying bone marrow edema in sacroiliitis **(Fig. 325-1)**. They are also highly sensitive for evaluation of acute and chronic spinal changes **(Fig. 325-2)**.

Figure 325-1

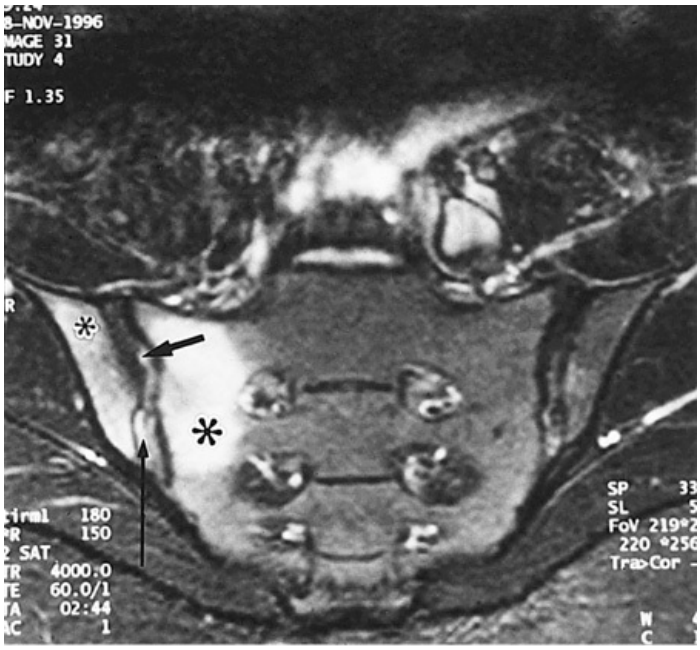

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Early sacroiliitis in a patient with AS, indicated by prominent edema in the juxtaarticular bone marrow (asterisks), synovium and joint capsule (thin arrow), and interosseous ligaments (thick arrow) on a STIR (short tau inversion recovery) magnetic resonance image. (*From M. Bollow et al. Zeitschrift für Rheumatologie 58:61, 1999. Reproduced with permission.)*

Figure 325-2

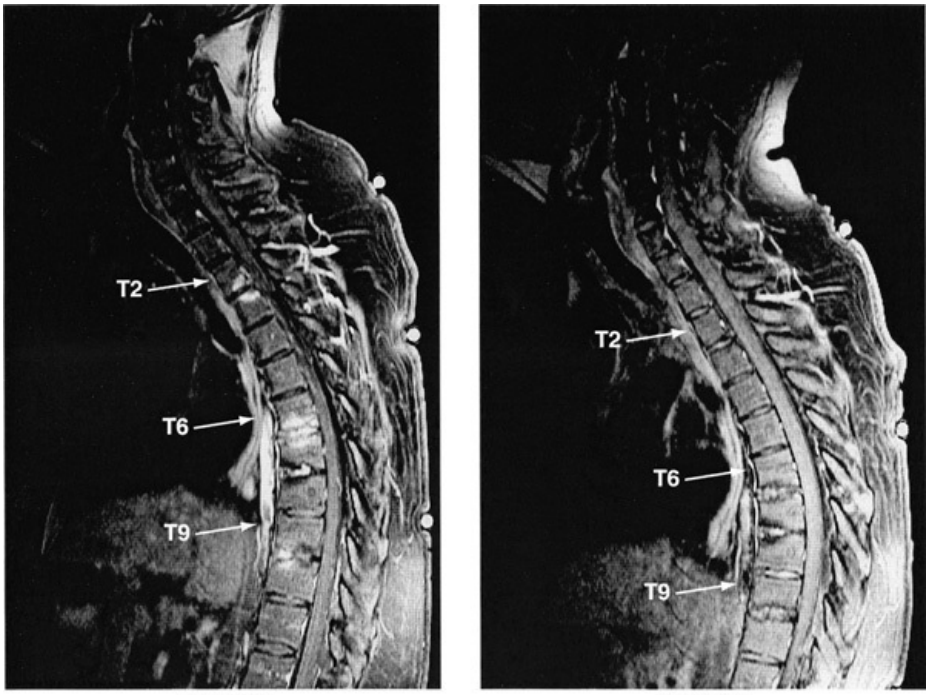

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spinal inflammation (spondylodiskitis) in a patient with AS and its dramatic response to treatment with infliximab. Gadolinium-enhanced T1-weighted magnetic resonance images, with fat saturation, at baseline and after 24 weeks of infliximab therapy. (*From J Braun et al.*)

Reduced bone mineral density can be detected by dual-energy x-ray absorptiometry of the femoral neck and the lumbar spine. By using a lateral projection of the L3 vertebral body, falsely elevated readings related to spinal ossification can be avoided.

Diagnosis

It is important to establish the diagnosis of early AS before the development of irreversible deformity. This goal presents a challenge for several reasons: (1) Back pain is very common, but AS is much less common; (2) an early presumptive diagnosis often relies on clinical grounds requiring considerable expertise; and (3) young individuals with early AS are often reluctant to seek medical care. The widely used modified New York criteria (1984) are based on the presence of definite radiographic sacroiliitis and are too insensitive in early or mild cases. In 2009, new criteria for axial SpA were proposed by the Assessment of Spondyloarthritis International Society (ASAS) **(Table 325-1)**. They are applicable to individuals with ≥3 months of back pain with age of onset <45 years old. Active inflammation of the sacroiliac (SI) joints as determined by dynamic MRI is considered equivalent to the older criterion of definite radiographic sacroiliitis (see below).

Table 325-1 Asas Criteria for Classification of Axial Spondyloarthritis (to Be Applied for Patients with Back Pain ≥3 Months and Age of Onset <45 Years)[a]

Sacroiliitis on imaging plus ≥1 SpA feature	**or HLA-B27 plus ≥2 other SpA features**
Sacroiliitis on imaging	SpA features
• Active (acute) inflammation on MRI highly suggestive of SpA-associated sacroiliitis[b] and/or • Definite radiographic sacroiliitis according to modified New York criteria[c]	• Inflammatory back pain[d] • Arthritis[e] • Enthesitis (heel)[f] • Anterior uveitis[g] • Dactylitis[e] • Psoriasis[e] • Crohn's disease or ulcerative colitis[e]

	• Good response to NSAIDs[h] • Family history of SpA[i] • HLA-B27 • Elevated CRP[j]

[a]Sensitivity 83%, specificity 84%. The imaging arm (sacroiliitis) alone has a sensitivity of 66% and a specificity of 97%.

[b]Bone marrow edema and/or osteitis on short tau inversion recovery (STIR) or gadolinium-enhanced T1 image.

[c]Bilateral grade =2 or unilateral grade 3 or 4.

[d] See text for criteria.

[e]Past or present, diagnosed by a physician.

[f] Past or present pain or tenderness on examination at calcaneus insertion of Achilles tendon or plantar fascia.

[g] Past or present, confirmed by an ophthalmologist.

[h] Substantial relief of back pain at 24–48 h after a full dose of NSAID.

[i] First- or second-degree relatives with ankylosing spondylitis (AS), psoriasis, uveitis, reactive arthritis (ReA), or inflammatory bowel disease (IBD).

[j] After exclusion of other causes of elevated CRP.

Abbreviations: ASAS, Assessment of Spondyloarthritis International Society; CRP, C-reactive protein; NSAIDs, nonsteroidal anti-inflammatorydrugs; SpA, spondyloarthritis.

Source: From M Rudwaleit et al: Ann Rheum Dis 68:777, 2009. Copyright 2009, with permission from BMJ Publishing Group Ltd.

AS must be differentiated from numerous other causes of low-back pain, some far more common than AS. To qualify as the criterion for inflammatory back pain of axial SpA (Table 325-1), the chronic (≥3 months) back pain should have four or more of these characteristic features: (1) age of onset below 40 years old, (2) insidious onset, (3) improvement with exercise and (4) no improvement with rest, and (5) pain at night with improvement upon getting up. Other common features of inflammatory back pain include morning stiffness >30 min, awakening from back pain during only the second half of the night, and alternating buttock pain. In clinical decision-making, all of these features are additive. The most common causes of back pain other than AS are primarily mechanical or degenerative rather than primarily inflammatory and tend not to show clustering of these features.

Less-common metabolic, infectious, and malignant causes of back pain must also be differentiated from AS, including infectious spondylitis, spondylodiskitis, and sacroiliitis, and primary or metastatic tumor. Ochronosis can produce a phenotype that is clinically and radiographically similar to AS. Calcification and ossification of paraspinous ligaments occur in *diffuse idiopathic skeletal hyperostosis* (DISH), which occurs in the middle-aged and elderly and is usually not symptomatic. Ligamentous calcification gives the appearance of "flowing wax" on the anterior bodies of the vertebrae. Intervertebral disk spaces are preserved, and sacroiliac and apophyseal joints appear normal, helping to differentiate DISH from spondylosis and from AS, respectively.

Treatment: Ankylosing Spondylitis

All management of AS should include an exercise program designed to maintain posture and range of motion. Nonsteroidal anti-inflammatory drugs (NSAIDs) are the first line of pharmacologic therapy for AS. These agents reduce pain and tenderness and increase mobility in many patients with AS. There is evidence that daily NSAID therapy slows radiographic progression. However, many patients with AS have continued symptoms and develop deformity despite NSAID therapy. Beginning in 2000, dramatic responses to anti-TNF-α therapy were reported in patients with AS and other spondyloarthritides. Patients with AS treated with either infliximab (chimeric human/mouse anti-TNF-α monoclonal antibody), etanercept (soluble p75 TNF-α receptor–IgG fusion protein), or adalimumab or golimumab (human anti-TNF-α monoclonal antibodies) have shown rapid, profound, and sustained reductions in all clinical and laboratory measures of disease activity. Patients with long-standing disease and even some with complete spinal ankylosis have shown significant improvement in both objective and subjective indicators of disease activity and function, including morning stiffness, pain, spinal mobility, peripheral joint swelling, CRP, and ESR. MRI studies indicate substantial resolution of bone marrow edema, enthesitis, and joint effusions in the sacroiliac joints, spine, and peripheral joints (Fig. 325-2). Similar results have been obtained in large randomized controlled trials of all four agents and many

open-label studies. About one-half of the patients achieve a ≥50% reduction in the BASDAI. The response tends to be stable over time, and partial or full remissions are common. Increased bone mineral density is found as early as 24 weeks after onset of therapy. There is evidence that anti-TNF therapy does not prevent syndesmophyte formation, although the practical clinical significance of this is not yet clear. A mechanism for this has been proposed based on the observation that TNF-α inhibits new bone formation by upregulating DKK-1, a negative regulator of the wingless (Wnt) signaling pathway that promotes osteoblast activity. Serum DKK-1 levels are inappropriately low in AS patients and are also suppressed by anti-TNF therapy.

The dosages of anti-TNF agents used in AS patients have usually been similar to those in RA. Infliximab is given intravenously, 3–5 mg/kg body weight, and then repeated 2 weeks later, again 6 weeks later, and then at 8-week intervals. Etanercept is given by subcutaneous injection, 50 mg once weekly. Adalimumab is given by subcutaneous injection, 40 mg biweekly. Golimumab is given by subcutaneous injection, 50 or 100 mg every 4 weeks.

Although these potent immunosuppressive agents have so far been relatively safe, seven types of side effects are not rare: (1) serious infections, including disseminated tuberculosis; (2) hematologic disorders, such as pancytopenia; (3) demyelinating disorders; (4) exacerbation of congestive heart failure; (5) systemic lupus erythematosus–related autoantibodies and clinical features; (6) hypersensitivity infusion or injection site reactions; and (7) severe liver disease. No increased incidence of malignancy has been observed in AS patients treated for over 5 years.

Because of the expense, potentially serious side effects, and unknown long-term effects of these agents, their use should be restricted to patients with a definite diagnosis and active disease (BASDAI ≥4 out of 10 and expert opinion) that is inadequately responsive to therapy with at least two different NSAIDs. Before initiation of anti-TNF therapy, all patients should be tested for tuberculin (TB) reactivity, and reactors (≥5 mm) should be treated with anti-TB agents. Contraindications include active infection or high risk of infection; malignancy or premalignancy; and history of systemic lupus erythematosus, multiple sclerosis, or related autoimmunity. Pregnancy and breast-feeding are relative contraindications. Continuation beyond 12 weeks of therapy requires either a 50% reduction in BASDAI or absolute reduction of ≥2 out of 10, and favorable expert opinion. Sulfasalazine, in doses of 2–3 g/d, has been shown to be of modest benefit, primarily for peripheral arthritis. A therapeutic trial of this agent should precede any use of anti-TNF agents in patients with predominantly peripheral arthritis. Methotrexate, although widely used, has not been shown to be of benefit in AS, nor has any therapeutic role for gold or oral glucocorticoids been documented. Potential benefit in AS has been reported for thalidomide, 200 mg/d, perhaps acting through inhibition of TNF-α.

The most common indication for surgery in patients with AS is severe hip joint arthritis, the pain and stiffness of which are usually dramatically relieved by total hip arthroplasty. Rare patients may benefit from surgical correction of extreme flexion deformities of the spine or of atlantoaxial subluxation.

Attacks of uveitis are usually managed effectively with local glucocorticoid administration in conjunction with mydriatic agents, although systemic glucocorticoids, immunosuppressive drugs, or anti-TNF therapy may be required. TNF inhibitors reduce the frequency of attacks of uveitis in patients with AS, although cases of new or recurrent uveitis after use of a TNF inhibitor have been observed, especially with etanercept.

Coexistent cardiac disease may require pacemaker implantation and/or aortic valve replacement. Management of axial osteoporosis is at present similar to that used for primary osteoporosis, since data specific for AS are not available.

REACTIVE ARTHRITIS

Reactive arthritis (ReA) refers to acute nonpurulent arthritis complicating an infection elsewhere in the body. In recent years, the term has been used primarily to refer to SpA following enteric or urogenital infections.

Other forms of reactive and infection-related arthritis not associated with B27 and showing a spectrum of clinical features different from SpA, such as Lyme disease and rheumatic fever, are discussed in Chaps. 173 and 322.

Historic Background

The association of acute arthritis with episodes of diarrhea or urethritis has been recognized for centuries. A large number of cases during World Wars I and II focused attention on the triad of arthritis, urethritis, and conjunctivitis, often with additional mucocutaneous lesions, which became widely known by eponyms that are now of historic interest only.

The identification of bacterial species capable of triggering the clinical syndrome and the finding that many patients possess the B27 antigen led to the unifying concept of ReA as a clinical syndrome triggered by specific etiologic agents in a genetically susceptible host. A similar spectrum of clinical manifestations can be triggered by enteric infection with any of several *Shigella*, *Salmonella*, *Yersinia*, and *Campylobacter* species; by genital infection with *Chlamydia trachomatis*; and by other agents as

well. The triad of arthritis, urethritis, and conjunctivitis represents a small part of the spectrum of the clinical manifestations of ReA. For the purposes of this chapter, the use of the term *ReA* will be restricted to those cases of SpA in which there is at least presumptive evidence for a related antecedent infection. Patients with clinical features of ReA who lack evidence of an antecedent infection will be considered to have *undifferentiated spondyloarthritis*, discussed below.

Epidemiology

Following the first reports of association of ReA with HLA-B27, in most hospital-based series in which *Shigella*, *Yersinia*, or *Chlamydia* were the triggering infectious agents, 60–85% of patients were found to be B27-positive, with a lower prevalence in ReA triggered by *Salmonella* and *Campylobacter*. In more recent community-based or common-source epidemic studies, the prevalence of B27 in ReA has often been below 50%, and in some instances not elevated at all. The most common age range is 18–40 years, but ReA can occur in children over 5 years of age and in older adults.

The gender ratio in ReA following enteric infection is nearly 1:1, whereas venereally acquired ReA occurs mainly in men. The overall prevalence and incidence of ReA are difficult to assess because of the variable prevalence of triggering infections and genetic susceptibility factors in different populations. In Scandinavia, an annual incidence of 10–28:100,000 has been reported. The spondylo-arthritides were formerly almost unknown in sub-Saharan Africa. However, ReA and other peripheral SpA have now become the most common rheumatic diseases in Africans in the wake of the AIDSepidemic, without association to B27, which is very rare in these populations. SpA in Africans with HIV infection usually occurs in individuals with stage I disease (as classified by the World Health Organization). It is often the first manifestation of infection and often remits with disease progression. In contrast, Western white patients with HIV and SpA are usually B27-positive, and the arthritis flares as AIDS advances.

Pathology

Synovial histology is similar to that of other SpA. Enthesitis shows increased vascularity and macrophage infiltration of fibrocartilage. Microscopic histopathologic evidence of inflammation has occasionally been noted in the colon and ileum of patients with post-venereal ReA, but much less commonly than in postenteric ReA. The skin lesions of keratoderma blenorrhagica, associated mainly with venereally acquired ReA, are histologically indistinguishable from psoriatic lesions.

Etiology and Pathogenesis

Of the four *Shigella* species *S. sonnei*, *S. boydii*, *S. flexneri*, and *S. dysenteriae*, *S. flexneri* has most often been implicated in cases of ReA, both sporadic and epidemic. *S. sonnei* and *S. dysenteriae* trigger some cases of ReA.

Other bacteria identified definitively as triggers of ReA include several *Salmonella* spp., *Yersinia enterocolitica*, *Y. pseudotuberculosis, Campylobacter jejuni* , and *Chlamydia trachomatis*. There is also evidence implicating *Clostridium difficile*, *Campylobacter coli* , certain toxigenic *E. coli*, and possibly *Ureaplasma urealyticum* and *Mycoplasma genitalium*. Respiratory infection with *Chlamydia pneumoniae* has also been implicated. There are also numerous isolated reports of acute arthritis preceded by other bacterial, viral, or parasitic infections, and even following intravesicular bacillus Calmette-Guérin (BCG) treatment for bladder cancer.

It has not been determined whether ReA occurs by the same pathogenic mechanism following infection with each of these microorganisms, nor has the mechanism been elucidated in the case of any one of the known bacterial triggers. Most, if not all, of the organisms well established to be triggers produce lipopolysaccharide (LPS) and share a capacity to attack mucosal surfaces, to invade host cells, and to survive intracellularly. Antigens from *Chlamydia*, *Yersinia*, *Salmonella*, and *Shigella* have been shown to be present in the synovium and/or synovial fluid leukocytes of patients with ReA for long periods following the acute attack. In ReA triggered by *Y. enterocolitica*, bacterial LPS and heat-shock protein antigens have been found in peripheral blood cells years after the triggering infection. *Yersinia* DNA and *C. trachomatis* DNA and RNA have been detected in synovial tissue from ReA patients, suggesting the presence of viable organisms despite uniform failure to culture the organism from these specimens. The specificity of these findings is unclear, however, since chromosomal bacterial DNA has also been found in synovium in other rheumatic diseases, and 16S rRNA from a very wide variety of bacteria has been found in ReA synovium. In several older studies, synovial T cells that specifically responded to antigens of the inciting organism were reported and characterized as predominantly CD4+ with a T_H2 or T regulatory phenotype. More recent work has documented high levels of IL-17 in ReA synovial fluid, but the source has not been identified. HLA-B27 seems to be associated with more severe and chronic forms of ReA, but its pathogenic role remains to be determined. HLA-B27 significantly prolongs the intracellular survival of *Y. enterocolitica* and *S. enteritidis* in human and mouse cell lines. Prolonged intracellular bacterial survival, promoted by B27, other factors, or both, may permit trafficking of infected leukocytes from the site of primary infection to joints, where an innate and/or adaptive immune response to persistent bacterial antigens may then promote arthritis.

Clinical Features

The clinical manifestations of ReA constitute a spectrum that ranges from an isolated, transient monarthritis or enthesitis to severe multisystem disease. A careful history will usually elicit evidence of an antecedent infection 1–4 weeks before onset of symptoms of the reactive disease. However, in a sizable minority, no clinical or laboratory evidence of an antecedent infection can be found. In cases of presumed venereally acquired reactive disease, there is often a history of a recent new sexual partner, even without laboratory evidence of infection.

Constitutional symptoms are common, including fatigue, malaise, fever, and weight loss. The musculoskeletal symptoms are usually acute in onset. Arthritis is usually asymmetric and additive, with involvement of new joints occurring over a few days to 1–2 weeks. The joints of the lower extremities, especially the knee, ankle, and subtalar, metatarsophalangeal, and toe interphalangeal joints, are most commonly involved, but the wrist and fingers can be involved as well. The arthritis is usually quite painful, and tense joint effusions are not uncommon, especially in the knee. Patients often cannot walk without support. Dactylitis, or "sausage digit," a diffuse swelling of a solitary finger or toe, is a distinctive feature of ReA and other peripheral spondyloarthritides but can be seen in polyarticular gout and sarcoidosis. Tendinitis and fasciitis are particularly characteristic lesions, producing pain at multiple insertion sites (entheses), especially the Achilles insertion, the plantar fascia, and sites along the axial skeleton. Spinal and low-back pain are quite common and may be caused by insertional inflammation, muscle spasm, acute sacroiliitis, or, presumably, arthritis in intervertebral joints.

Urogenital lesions may occur throughout the course of the disease. In males, urethritis may be marked or relatively asymptomatic and may be either an accompaniment of the triggering infection or a result of the reactive phase of the disease. Prostatitis is also common. Similarly, in females, cervicitis or salpingitis may be caused either by the infectious trigger or by the sterile reactive process.

Ocular disease is common, ranging from transient, asymptomatic conjunctivitis to an aggressive anterior uveitis that occasionally proves refractory to treatment and may result in blindness.

Mucocutaneous lesions are frequent. Oral ulcers tend to be superficial, transient, and often asymptomatic. The characteristic skin lesions, *keratoderma blenorrhagica*, consist of vesicles that become hyperkeratotic, ultimately forming a crust before disappearing. They are most common on the palms and soles but may occur elsewhere as well. In patients with HIV infection, these lesions are often extremely severe and extensive, sometimes dominating the clinical picture **(Chap. 189)**. Lesions may occur on the glans penis, termed *circinate balanitis*; these consist of vesicles that quickly rupture to form painless superficial erosions, which in circumcised individuals can form crusts similar to those of keratoderma blenorrhagica. Nail changes are common and consist of onycholysis, distal yellowish discoloration, and/or heaped-up hyperkeratosis.

Less-frequent or rare manifestations of ReA include cardiac conduction defects, aortic insufficiency, central or peripheral nervous system lesions, and pleuropulmonary infiltrates.

Arthritis typically persists 3–5 months, but courses up to 1 year can occur. Chronic joint symptoms persist in about 15% of patients and in up to 60% in hospital-based series. Recurrences of the acute syndrome are also common. Work disability or forced change in occupation are common in those with persistent joint symptoms. Chronic heel pain is often particularly distressing. Low-back pain, sacroiliitis, and frank AS are also common sequelae. In most studies, HLA-B27–positive patients have shown a worse outcome than B27-negative patients. Patients with *Yersinia*- or *Salmonella*- induced arthritis have less chronic disease than those whose initial episode follows epidemic shigellosis.

Laboratory and Radiographic Findings

The ESR and acute-phase reactants are usually elevated during the acute phase of the disease. Mild anemia may be present. Synovial fluid is nonspecifically inflammatory. In most ethnic groups, about one-half of the patients are B27-positive. The triggering infection usually does not persist at the site of primary mucosal infection through the time of onset of the reactive disease, but it may be possible to culture the organism, e.g., in the case of *Yersinia*- or *Chlamydia*-induced disease. Serologic evidence of a recent infection may be present, such as a marked elevation of antibodies to *Yersinia*, *Salmonella*, or *Chlamydia*. Polymerase chain reaction (PCR) for chlamydial DNA in first-voided urine specimens is said to have high sensitivity.

In early or mild disease, radiographic changes may be absent or confined to juxtaarticular osteoporosis. With long-standing persistent disease, marginal erosions and loss of joint space can be seen in affected joints. Periostitis with reactive new bone formation is characteristic, as in all the SpA. Spurs at the insertion of the plantar fascia are common.

Sacroiliitis and spondylitis may be seen as late sequelae. Sacroiliitis is more commonly asymmetric than in AS, and spondylitis, rather than ascending symmetrically, can begin anywhere along the lumbar spine. The syndesmophytes may be asymmetric, coarse and nonmarginal, arising from the middle of a vertebral body, a pattern less commonly seen in primary AS. Progression to spinal fusion is uncommon.

Diagnosis

ReA is a clinical diagnosis with no definitively diagnostic laboratory test or radiographic finding. The diagnosis should be entertained in any patient with an acute inflammatory, asymmetric, additive arthritis or tendinitis. The evaluation should include questioning regarding possible triggering events such as an episode of diarrhea or dysuria. On physical examination, attention must be paid to the distribution of the joint and tendon involvement and to possible sites of extraarticular involvement, such as the eyes, mucous membranes, skin, nails, and genitalia. Synovial fluid analysis may be helpful in excluding septic or crystal-induced arthritis. Culture, serology, or molecular methods may help to identify a triggering infection.

Although typing for B27 has low negative predictive value in ReA, it may have prognostic significance in terms of severity, chronicity, and the propensity for spondylitis and uveitis. Furthermore, if positive, it can be helpful diagnostically in atypical cases. HIV testing is often indicated and may be necessary in order to select appropriate therapy.

It is important to differentiate ReA from disseminated gonococcal disease **(Chap. 144)**, both of which can be venereally acquired and associated with urethritis. Unlike ReA, gonococcal arthritis and tenosynovitis tend to involve both upper and lower extremities equally, to lack back symptoms, and to be associated with characteristic vesicular skin lesions. A positive gonococcal culture from the urethra or cervix does not exclude a diagnosis of ReA; however, culturing gonococci from blood, skin lesion, or synovium establishes the diagnosis of disseminated gonococcal disease. PCR assay for *N. gonorrhoeae* and *C. trachomatis* may be helpful. Occasionally, only a therapeutic trial of antibiotics can distinguish the two.

ReA shares many features in common with psoriatic arthropathy. However, psoriatic arthritis is usually gradual in onset; the arthritis tends to affect primarily the upper extremities; there is less associated periarthritis; and there are usually no associated mouth ulcers, urethritis, or bowel symptoms.

Treatment Reactive Arthritis

Most patients with ReA benefit to some degree from high-dose NSAIDs, although acute symptoms are rarely completely ameliorated, and some patients fail to respond at all. Indomethacin, 75–150 mg/d in divided doses, is the initial treatment of choice, but other NSAIDs may be tried.

Prompt, appropriate antibiotic treatment of acute chlamydial urethritis or enteric infection may prevent the emergence of ReA. However, several controlled trials have failed to demonstrate any benefit for antibiotic therapy that is initiated after onset of arthritis. One long-term follow-up study suggested that although antibiotic therapy had no effect on the acute episode of ReA, it helped prevent subsequent chronic SpA. Another such study failed to demonstrate any long-term benefit. A promising recent double-blind placebo-controlled study showed that a majority of patients with chronic ReA due to *Chlamydia* benefited significantly from a 6-month course of rifampin 300 mg daily plus azithromycin 500 mg daily for 5 days then twice weekly, or 6 months of rifampin 300 mg daily plus doxycycline 100 mg twice daily.

Multicenter trials have suggested that sulfasalazine, up to 3 g/d in divided doses, may be beneficial to patients with persistent ReA.[1] Patients with persistent disease may respond to azathioprine, 1–2 mg/kg per day, or to methotrexate, up to 20 mg per week. Although no controlled trials of anti-TNF-α in ReA have been reported, anecdotal evidence supports the use of these agents in severe chronic cases, although lack of response has also been observed.[1]

Tendinitis and other enthesitic lesions may benefit from intralesional glucocorticoids. Uveitis may require aggressive treatment to prevent serious sequelae (see above). Skin lesions ordinarily require only symptomatic treatment. In patients with HIV infection and ReA, many of whom have severe skin lesions, the skin lesions in particular respond to antiretroviral therapy. Cardiac complications are managed conventionally; management of neurologic complications is symptomatic.

Comprehensive management includes counseling of patients in the avoidance of sexually transmitted disease and exposure to enteropathogens, as well as appropriate use of physical therapy, vocational counseling, and continued surveillance for long-term complications such as ankylosing spondylitis.

PSORIATIC ARTHRITIS

Psoriatic arthritis (PsA) refers to an inflammatory arthritis that characteristically occurs in individuals with psoriasis.

Historic Background

The association between arthritis and psoriasis was noted in the nineteenth century. In the 1960s, on the basis of epidemiologic and clinical studies, it became clear that unlike RA the arthritis associated with psoriasis was usually seronegative, often involved the distal interphalangeal (DIP) joints of the fingers and the spine and sacroiliac joints, had distinctive radiographic features, and showed considerable familial aggregation. In the 1970s, PsA was included in the broader category of the spondyloarthritides because of features similar to those of AS and ReA.

Epidemiology

Estimates of the prevalence of PsA among individuals with psoriasis range from 5 to 30%. In white populations, psoriasis is estimated to have a prevalence of 1–3%. Psoriasis and PsA are less common in other races in the absence of HIV infection, and the prevalence of PsA in individuals with psoriasis may be less common. First-degree relatives of PsA patients have an elevated risk for psoriasis, for PsA itself, and for other forms of SpA. Of patients with psoriasis, up to 30% have an affected first-degree relative. In monozygotic twins, the reported concordance for psoriasis varies from 35 to 72%, and for PsA from 10 to 30%. A variety of HLA associations have been found. The HLA-Cw6 gene is directly associated with psoriasis, particularly familial juvenile-onset (type I) psoriasis. HLA-B27 is associated with psoriatic spondylitis (see below). HLA-DR7, -DQ3, and -B57 are associated with PsA because of linkage disequilibrium with Cw6. Other associations with PsA include HLA-B13, -B37, -B38, -B39, and DR4. A recent genome-wide scan found association of both psoriasis and PsA with a polymorphism at the HCP5 locus closely linked to HLA-B, and also to IL-23R, IL-12B (chromosome 5q31), and several other chromosomal regions.

Pathology

The inflamed synovium in PsA resembles that of RA, although with somewhat less hyperplasia and cellularity than in RA, and somewhat greater vascularity. Some studies have indicated a higher tendency to synovial fibrosis in PsA. Unlike RA, PsA shows prominent enthesitis, with histology similar to that of the other spondyloarthritides.

Pathogenesis

PsA is almost certainly immune-mediated and probably shares pathogenic mechanisms with psoriasis. PsA synovium shows infiltration with T cells, B cells, macrophages, and NK receptor–expressing cells, with upregulation of leukocyte homing receptors. Clonally expanded CD8+ T cells are frequent in PsA. Plasmacytoid dendritic cells are thought to play a key role in psoriasis, and there is some evidence for their participation in psoriatic arthritis. There is abundant synovial overexpression of proinflammatory cytokines. Interleukin 2, interferon-γ, TNF-α, and IL-1β, -6, -8, -10, -12, -13, and -15 are found in PsA synovium or synovial fluid. T_H17 derived cytokines are likely to be important in PsA, given their prominence in psoriasis and in other spondyloarthritides, the genetic association with genes in the IL-12/IL-23 axis, and the therapeutic response to an antibody to the shared IL-12/23 p40 subunit (see below). Consistent with the extensive bone lesions in PsA, patients with PsA have been found to have a marked increase in osteoclastic precursors in peripheral blood and upregulation of receptor activator of nuclear factor κβ ligand (RANKL) in the synovial lining layer.

Clinical Features

In 60–70% of cases, psoriasis precedes joint disease. In 15–20% of cases, the two manifestations appear within 1 year of each other. In about 15–20% of cases, the arthritis precedes the onset of psoriasis and can present a diagnostic challenge. The frequency in men and women is almost equal, although the frequency of disease patterns differs somewhat in the two sexes. The disease can begin in childhood or late in life but typically begins in the fourth or fifth decade, at an average age of 37 years.

The spectrum of arthropathy associated with psoriasis is quite broad. Many classification schemes have been proposed. In the original scheme of Wright and Moll, five patterns are described: (1) arthritis of the DIP joints; (2) asymmetric oligoarthritis; (3) symmetric polyarthritis similar to RA; (4) axial involvement (spine and sacroiliac joints); and (5) arthritis mutilans, a highly destructive form of disease. These patterns are not fixed, and the pattern that persists chronically often differs from that of the initial presentation. A simpler scheme in recent use contains three patterns: oligoarthritis, polyarthritis, and axial arthritis.

Nail changes in the fingers or toes occur in 90% of patients with PsA, compared with 40% of psoriatic patients without arthritis, and pustular psoriasis is said to be associated with more severe arthritis. Several articular features distinguish PsA from other joint disorders. Dactylitis occurs in >30%; enthesitis and tenosynovitis are also common and are probably present in most patients, although often not appreciated on physical examination. Shortening of digits because of underlying osteolysis is particularly characteristic of PsA **(Fig. 325-3)**, and there is a much greater tendency than in RA for both fibrous and bony ankylosis of small joints. Rapid ankylosis of one or more proximal interphalangeal (PIP) joints early in the course of disease is not uncommon. Back and neck pain and stiffness are also common in PsA.

Figure 325-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Characteristic lesions of psoriatic arthritis. Inflammation is prominent in the DIP joints (left 5th, 4th, 2nd; right 2nd, 3rd, and 5th) and PIP joints (left 2nd, right 2nd, 4th, and 5th). There is dactylitis in the left 2nd finger and thumb, with pronounced telescoping of the left 2nd finger. Nail dystrophy (hyperkeratosis and onycholysis) affects each of the fingers except the left 3rd finger, the only finger without arthritis. (*Courtesy of Donald Raddatz, MD; with permission.*)

Arthropathy confined to the DIP joints predominates in about 15% of cases. Accompanying nail changes in the affected digits are almost always present. These joints are also often affected in the other patterns of PsA. Approximately 30% of patients have asymmetric oligoarthritis. This pattern commonly involves a knee or another large joint with a few small joints in the fingers or toes, often with dactylitis. Symmetric polyarthritis occurs in about 40% of PsA patients at presentation. It may be indistinguishable from RA in terms of the joints involved, but other features characteristic of PsA are usually also present. In general, peripheral joints in PsA tend to be somewhat less tender than in RA, although signs of inflammation are usually present. Almost any peripheral joint can be involved. Axial arthropathy without peripheral involvement is found in about 5% of PsA patients. It may be indistinguishable from idiopathic AS, although more neck involvement and less thoracolumbar spinal involvement is characteristic, and nail changes are not found in idiopathic AS. A small percentage of PsA patients have arthritis mutilans, in which there can be widespread shortening of digits ("telescoping"), sometimes coexisting with ankylosis and contractures in other digits.

Six patterns of nail involvement are identified: pitting, horizontal ridging, onycholysis, yellowish discoloration of the nail margins, dystrophic hyperkeratosis, and combinations of these findings. Other extraarticular manifestations of the spondyloarthritides are common. Eye involvement, either conjunctivitis or uveitis, is reported in 7–33% of PsA patients. Unlike the uveitis associated with AS, the uveitis in PsA is more often bilateral, chronic, and/or posterior. Aortic valve insufficiency has been found in <4% of patients, usually after long-standing disease.

Widely varying estimates of clinical outcome have been reported in PsA. At its worst, severe PsA with arthritis mutilans is at least as crippling and ultimately fatal as severe RA. Unlike RA, however, many patients with PsA experience temporary remissions. Overall, erosive disease develops in the majority of patients, progressive disease with deformity and disability is common, and in some large published series, mortality was found to be significantly increased compared with the general population.

The psoriasis and associated arthropathy seen with HIV infection both tend to be severe and can occur in populations with very little psoriasis in noninfected individuals. Severe enthesopathy, dactylitis, and rapidly progressive joint destruction are seen, but axial involvement is very rare. This condition is prevented by or responds well to antiretroviral therapy.

Laboratory and Radiographic Findings

There are no laboratory tests diagnostic of PsA. ESR and CRP are often elevated. A small percentage of patients may have low titers of rheumatoid factor or antinuclear antibodies. About 10% of patients have anti-CCP antibodies. Uric acid may be elevated in the presence of extensive psoriasis. HLA-B27 is found in 50–70% of patients with axial disease, but ≤20% in patients with only peripheral joint involvement.

The peripheral and axial arthropathies in PsA show a number of radiographic features that distinguish them from RA and AS, respectively. Characteristics of peripheral PsA include DIP involvement, including the classic "pencil-in-cup" deformity; marginal erosions with adjacent bony proliferation ("whiskering"); small-joint ankylosis; osteolysis of phalangeal and metacarpal bone, with telescoping of digits; and periostitis and proliferative new bone at sites of enthesitis. Characteristics of axial PsA include asymmetric sacroiliitis; compared with idiopathic AS, less zygapophyseal joint arthritis, fewer

and less symmetric and delicate syndesmophytes; fluffy hyperperiostosis on anterior vertebral bodies; severe cervical spine involvement, with a tendency to atlantoaxial subluxation but relative sparing of the thoracolumbar spine; and paravertebral ossification. Ultrasound and MRI both readily demonstrate enthesitis and tendon sheath effusions that can be difficult to assess on physical examination. A recent MRI study of 68 PsA patients found sacroiliitis in 35%, unrelated to B27 but correlated with restricted spinal movement.

Diagnosis

Classification criteria for PsA were published in 2006 [Classification of Psoriatic Arthritis (CASPAR) criteria] that have been widely accepted **(Table 325-2)**. The sensitivity and specificity of these criteria exceed 90%, and they are useful for early diagnosis. The criteria are based on the history, presence of psoriasis, characteristic peripheral or spinal joint symptoms, signs, and imaging. Diagnosis can be challenging when the arthritis precedes psoriasis, the psoriasis is undiagnosed or obscure, or the joint involvement closely resembles another form of arthritis. A high index of suspicion is needed in any patient with an undiagnosed inflammatory arthropathy. The history should include inquiry about psoriasis in the patient and family members. Patients should be asked to disrobe for the physical examination, and psoriasiform lesions should be sought in the scalp, ears, umbilicus, and gluteal folds in addition to more accessible sites; the finger and toe nails should also be carefully examined. Axial symptoms or signs, dactylitis, enthesitis, ankylosis, the pattern of joint involvement, and characteristic radiographic changes can be helpful clues. The differential diagnosis includes all other forms of arthritis, which can occur coincidentally in individuals with psoriasis. The differential diagnosis of isolated DIP involvement is short. Osteoarthritis (Heberden's nodes) is usually not inflammatory; gout involving more than one DIP joint often involves other sites and may be accompanied by tophi; the very rare entity multicentric reticulohistiocytosis involves other joints and has characteristic small pearly periungual skin nodules; and the uncommon entity inflammatory osteoarthritis, like the others, lacks the nail changes of PsA. Radiography can be helpful in all of these cases and in distinguishing between psoriatic spondylitis and idiopathic AS. A history of trauma to an affected joint preceding the onset of arthritis is said to occur more frequently in PsA than in other types of arthritis, perhaps reflecting the Koebner phenomenon in which psoriatic skin lesions can arise at sites of the skin trauma.

Table 325-2 The Caspar (C Lassification Criteria for Psoriatic Arthritis) Criteria[a]

To meet the CASPAR criteria, a patient must have inflammatory articular disease (joint, spine, or entheseal) with ≥3 points from any of the following five categories:

1. Evidence of current psoriasis,[b, c] a personal history of psoriasis, or a family history of psoriasis[d]
2. Typical psoriatic nail dystrophy[e] observed on current physical examination
3. A negative test result for rheumatoid factor
4. Either current dactylitisf or a history of dactylitis recorded by a rheumatologist
5. Radiographic evidence of juxtaarticular new bone formation[g] in the hand or foot

[a] Specificity of 99% and sensitivity of 91%.

[b] Current psoriasis is assigned 2 points; all other features are assigned 1 point.

[c] Psoriatic skin or scalp disease present at the time of examination, as judged by a rheumatologist or dermatologist.

[d] History of psoriasis in a first- or second-degree relative.

[e] Onycholysis, pitting, or hyperkeratosis.

[f] Swelling of an entire digit.

[g] Ill-defined ossification near joint margins, excluding osteophyte formation.

Source: From W Taylor et al.

Treatment: Psoriatic Arthritis

Ideally, coordinated therapy is directed at both the skin and joints in PsA. As described above for AS, use of the anti-TNF-α agents has revolutionized the treatment of PsA. Prompt and dramatic resolution of both arthritis and skin lesions has been observed in large, randomized controlled trials of etanercept, infliximab, adalimumab, and golimumab. Many of the responding patients had long-standing disease that was resistant to all previous therapy, as well as extensive skin disease. The clinical response is more dramatic than in RA, and delay of disease progression has been demonstrated radiographically. Paradoxically, rare cases have been reported of exacerbation or de

novo appearance of psoriasis precipitated by anti-TNF therapy for a variety of conditions. In some cases, the therapy can nevertheless be continued.

The anti-T cell biologic agent alefacept, in combination with methotrexate, has shown benefit in both psoriatic arthritis and psoriasis. Ustekinumab, a monoclonal antibody to the shared IL-23/IL-12p40 subunit, has shown promise in treating both psoriasis and PsA in early clinical trials.

Other treatment for PsA has been based on drugs that have efficacy in RA and/or in psoriasis. Although methotrexate in doses of 15–25 mg/week and sulfasalazine (usually given in doses of 2–3 g/d) have each been found to have clinical efficacy in controlled trials, neither effectively halts progression of erosive joint disease. Other agents with efficacy in psoriasis reported to benefit PsA are cyclosporine, retinoic acid derivatives, and psoralens plus ultraviolet A light (PUVA). There is controversy regarding the efficacy in PsA of gold and antimalarials, which have been widely used in RA. The pyrimidine synthetase inhibitor leflunomide has been shown in a randomized controlled trial to be beneficial in both psoriasis and psoriatic arthritis.

All of these treatments require careful monitoring. Immunosuppressive therapy may be used cautiously in HIV-associated PsA if the HIV infection is well controlled.

In one large prospective series, 7% of patients with PsA required musculoskeletal surgery beginning at a mean of 13 years' disease duration. Indications for surgery are similar to those in RA, although there is an impression that outcomes in PsA may be less satisfactory.

UNDIFFERENTIATED AND JUVENILE-ONSET SPONDYLOARTHRITIS

Many patients, usually young adults, present with some features of one or more of the spondyloarthritides discussed above. Until recently, these patients were said to have undifferentiated spondyloarthritis, or simply spondyloarthritis, as defined by the 1991 European Spondyloarthropathy Study Group criteria **(Table 325-3)**. For example, a patient may present with inflammatory synovitis of one knee, Achilles tendinitis, and dactylitis of one digit. Some of these patients may have ReA in which the triggering infection remains clinically silent. In some other cases, the patient subsequently develops IBD or psoriasis or the process eventually meets criteria for AS. This diagnosis of undifferentiated SpA was also commonly applied to patients with inflammatory back pain, who did meet modified New York criteria for AS. Most of these would now be classified under the new category of axial spondyloarthritis (Table 325-1).

Table 325-3 European Spondyloarthropathy Study Group (Essg) Criteria for Spondyloarthritis[a]

Inflammatory Back Pain[b]	***or***	**Synovitis**
		• **Asymmetric *or*** • **Predominantly in lower extremities**
	and	
One or more of the following:		
• Family history of SpA[b]		
• Psoriasis[b]		
• Crohn's disease or ulcerative colitis[c]		
• Nongonococcal urethritis, cervicitis, or acute diarrhea within 1 month before arthritis		
• Alternating buttock pain[d]		
• ?Enthesitis[b]		
• ?Radiographic sacroiliitis[b]		

[a] Sensitivity >85%, specificity >85%.

[b] See definition in Table 325-1.

[c] Past or present, diagnosed by a physician and confirmed by endoscopy or radiography.

[d] Past or present pain alternating between the right and left gluteal regions.

SpA, spondyloarthritis.

Source: From M Dougados et al; J Sieper J et al. Copyright 2009, with permission from BMJ Publishing Group Ltd.

Approximately one-half of the patients with undifferentiated SpA are HLA-B27-positive, and thus the absence of B27 is not useful in establishing or excluding the diagnosis. In familial cases, which are much more frequently B27-positive, there is often eventual progression to classical AS.

In juvenile-onset SpA, which begins between ages 7 and 16, most commonly in boys (60–80%), an asymmetric, predominantly lower-extremity oligoarthritis and enthesitis without extraarticular features is the typical mode of presentation. The prevalence of B27 in this condition, which has been termed the *seronegative enthesopathy and arthropathy (SEA) syndrome*, is approximately 80%. Many, but not all, of these patients go on to develop AS in late adolescence or adulthood.

Management of undifferentiated SpA is similar to that of the other spondyloarthritides. Response to anti-TNF-α therapy has been documented, and this therapy is indicated in severe, persistent cases not responsive to other treatment. One 2004 publication reported significant benefit in patients with long-standing undifferentiated spondyloarthropathy treated for 9 months with doxycycline and rifampin. These data await confirmation.

Current pediatric textbooks and journals should be consulted for information on management of juvenile-onset SpA.

ENTEROPATHIC ARTHRITIS

Historic Background

A relationship between arthritis and IBD was observed in the 1930s. The relationship was further defined by the epidemiologic studies in the 1950s and 1960s and included in the concept of the spondyloarthritides in the 1970s.

Epidemiology

Both of the common forms of IBD, ulcerative colitis (UC) and Crohn's disease (CD) **(Chap. 295)**, are associated with SpA. UC and CD both have an estimated prevalence of 0.05–0.1%, and the incidence of each is thought to have increased in recent decades. AS and peripheral arthritis are both associated with UC and with CD. Wide variations have been reported in the estimated frequencies of these associations. In recent series, AS was diagnosed in 1–10%, and peripheral arthritis in 10–50% of patients with IBD. Inflammatory back pain and enthesopathy are common, and many patients have sacroiliitis on imaging studies.

The prevalence of UC or CD in patients with AS is thought to be 5–10%. However, investigation of unselected SpA patients by ileocolonoscopy has revealed that from one-third to two-thirds of patients with AS have subclinical intestinal inflammation that is evident either macroscopically or histologically. These lesions have also been found in patients with undifferentiated SpA or ReA (both enterically and urogenitally acquired).

Both UC and CD have a tendency to familial aggregation, more so for CD. HLA associations have been weak and inconsistent. HLA-B27 is found in up to 70% of patients with IBD and AS, but in ≤15% of patients with IBD and peripheral arthritis or IBD alone. Three alleles of the *NOD2/CARD15* gene on chromosome 16 have been found in approximately one-half of patients with CD. These alleles are not associated with the spondyloarthritides per se. However, they are found significantly more often in (1) CD patients with sacroiliitis than in those without sacroiliitis, and (2) SpA patients with chronic inflammatory gut lesions than in those with normal gut histology. These associations are independent of HLA-B27.

Genome studies have shown that CD and UC have some susceptibility genes in common and some specific to each condition. Among these, IL-23R (highly associated with CD and to a lesser degree with UC) is shared with AS and psoriasis. TNFSF15, associated with CD, has also been found linked to SpA.

Pathology

Available data for IBD-associated peripheral arthritis suggest a synovial histology similar to other spondyloarthritides. Association with arthropathy does not affect the gut histology of UC or CD **(Chap. 295)**. The subclinical inflammatory lesions in the colon and distal ileum associated with SpA have been classified as either acute or chronic. The former resemble acute bacterial enteritis, with largely intact architecture and neutrophilic infiltration in the lamina propria. The latter resemble the lesions of CD, with distortion of villi and crypts, aphthoid ulceration, and mononuclear cell infiltration in the lamina propria.

Pathogenesis

Both IBD and SpA are immune-mediated, but the specific pathogenic mechanisms are poorly understood, and the connection between the two is obscure. The shared genetics could reflect either

shared pathogenetic mechanisms, close genetic linkage of separate susceptibility alleles, or both. IBD is a common phenotype in a number of rodent lines with transgenic overexpression or targeted deletion of genes involved in immune processes. Arthritis is an accompanying prominent feature in two of these IBD models, B27 transgenic rats and mice with constitutive overexpression of TNF-α, and immune dysregulation is prominent in both. Several lines of evidence indicate trafficking of leukocytes between the gut and the joint. Mucosal leukocytes from IBD patients have been shown to bind avidly to synovial vasculature through several different adhesion molecules. Macrophages expressing CD163 are prominent in the inflammatory lesions of both gut and synovium in the spondyloarthritides.

Clinical Features

AS associated with IBD is clinically indistinguishable from idiopathic AS. It runs a course independent of the bowel disease, and in many patients it precedes the onset of IBD, sometimes by many years. Peripheral arthritis not infrequently begins before onset of overt bowel disease. The spectrum of peripheral arthritis includes acute self-limited attacks of oligoarthritis that often coincide with relapses of IBD, and more chronic and symmetric polyarticular arthritis that runs a course independent of IBD activity. The patterns of joint involvement are similar in UC and CD. In general, erosions and deformities are infrequent in IBD-associated peripheral arthritis, and joint surgery is infrequently required. Isolated destructive hip arthritis is a rare complication of CD, apparently distinct from osteonecrosis and septic arthritis. Dactylitis and enthesopathy are occasionally found. In addition to the ~20% of IBD patients with SpA, a comparable percentage have arthralgias or fibromyalgia symptoms.

Other extraintestinal manifestations of IBD are seen in addition to arthropathy, including uveitis, pyoderma gangrenosum, erythema nodosum, and finger clubbing, all somewhat more commonly in CD than UC. The uveitis shares the features described above for PsA-associated uveitis.

Laboratory and Radiographic Findings

Laboratory findings reflect the inflammatory and metabolic manifestations of IBD. Joint fluid is usually at least mildly inflammatory. Of patients with AS and IBD, 30–70% carry the HLA-B27 gene, compared with >90% of patients with AS alone and 50–70% of those with AS and psoriasis. Hence, definite or probable AS in a B27-negative individual in the absence of psoriasis should prompt a search for occult IBD. Radiographic changes in the axial skeleton are the same as in uncomplicated AS. Erosions are uncommon in peripheral arthritis but may occur, particularly in the metatarsophalangeal joints. Isolated destructive hip disease has been described.

Diagnosis

Diarrhea and arthritis are both common conditions that can coexist for a variety of reasons. When etiopathogenically related, reactive arthritis and IBD-associated arthritis are the most common causes. Rare causes include celiac disease, blind loop syndromes, and Whipple's disease. In most cases, diagnosis depends upon investigation of the bowel disease.

Treatment: Enteropathic Arthritis

Treatment of CD has been improved by therapy with anti-TNF agents. Infliximab and adalimumab are effective for induction and maintenance of clinical remission in CD, and infliximab has been shown to be effective in fistulizing CD. IBD-associated arthritis also responds to these agents. Other treatment for IBD, including sulfasalazine and related drugs, systemic glucocorticoids, and immunosuppressive drugs, are also usually of benefit for associated peripheral arthritis. NSAIDs are generally helpful and well tolerated, but they can precipitate flares of IBD. Rare cases of IBD, usually UC, have apparently been precipitated by anti-TNF therapy, usually etanercept, given for any of several rheumatic diseases.

SAPHO SYNDROME

The syndrome of synovitis, acne, pustulosis, hyperostosis, and osteitis (SAPHO) is characterized by a variety of skin and musculoskeletal manifestations. Dermatologic manifestations include palmoplantar pustulosis, acne conglobata, acne fulminans, and hidradenitis suppurativa. The main musculoskeletal findings are sternoclavicular and spinal hyperostosis, chronic recurrent foci of sterile osteomyelitis, and axial or peripheral arthritis. Cases with one or a few manifestations are probably the rule. The ESR is usually elevated, sometimes dramatically. In some cases, bacteria, most often *Propionibacterium acnes*, have been cultured from bone biopsy specimens and occasionally other sites. Inflammatory bowel disease was coexistent in 8% of patients in one large series. B27 is not associated. Either bone scan or CT scan is helpful diagnostically. A recent MRI report described a characteristic vertebral body corner cortical erosions in 12 out of 12 patients. High-dose NSAIDs may provide relief from bone pain. A number of uncontrolled series and case reports describe successful therapy with pamidronate or other bisphosphonates. Response to anti-TNF-α therapy has also been observed, although in a few cases this has been associated with a flare of skin manifestations. Successful prolonged antibiotic therapy has also been reported.

WHIPPLE'S DISEASE

Whipple's disease **(Chap. 294)** is a rare chronic bacterial infection, mostly of middle-aged white men, caused by *Tropheryma whipplei*. At least 75% of affected individuals develop an oligo- or polyarthritis. The joint manifestations usually precede other symptoms of the disease by 5 years or more; they are thus particularly important because appropriate antibiotic therapy is curative, whereas the untreated disease is fatal. Large and small peripheral joints and sacroiliac joints may be involved. The arthritis is abrupt in onset, migratory, usually lasts hours to a few days, and then resolves completely. Chronic polyarthritis can occur but is not typical. Eventually prolonged diarrhea, malabsorption, and weight loss occur. Other manifestations of systemic disease include fever, edema, serositis, endocarditis, pneumonia, hypotension, lymphadenopathy, hyperpigmentation, subcutaneous nodules, clubbing, and uveitis. Central nervous system involvement eventually develops in 80% of untreated patients, with cognitive changes, headache, diplopia, and papilledema, and may be detectable on MRI. Oculomasticatory and oculofacial-skeletal myorhythmia with supranuclear vertical gaze palsy are said to be pathognomonic. Laboratory abnormalities include anemia and changes from malabsorption. Synovial fluid is usually inflammatory. Radiography rarely shows joint erosions but may show sacroiliitis. Abdominal CT may reveal lymphadenopathy. Foamy macrophages containing periodic acid–Schiff (PAS)-staining bacterial remnants can be seen in biopsies of small intestine, synovium, lymph node, and other tissues.

The complete genome sequence of *T. whipplei* was published in 2003. Diagnosis is facilitated by PCR amplification of sequences of the 16S ribosomal gene or other genes of *T. whipplei* in biopsied tissue. In the future, this may be supplanted or complemented by serologic tests. The organism is ubiquitous in the environment and is found in some healthy individuals, so the mere presence of DNA does not establish a diagnosis. The syndrome responds to therapy with penicillin (or ceftriaxone) and streptomycin for 2 weeks followed by trimethoprim-sulfamethoxazole for 1–2 years, but other antibiotic regimens may be preferable, and infectious disease consultation is strongly advised. Monitoring for central nervous system relapse is critical. Recently, nonclassical infections with *T. whipplei* have been described, including endocarditis.

FURTHER READINGS

Australo-Anglo-American Spondyloarthritis Consortium (TASC) et al: Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet 42:123, 2010

Benjamin M et al: Microdamage and altered vascularity at the enthesis-bone interface provides an anatomic explanation for bone involvement in the HLA-B27-associated spondylarthritides and allied disorders. Arthritis Rheum 56:224, 2007[PMID: 17195226] [Full Text]

Braun J et al: Persistent clinical efficacy and safety of anti-tumour necrosis factor alpha therapy with infliximab in patients with ankylosing spondylitis over 5 years: Evidence for different types of response. Ann Rheum Dis 67:340, 2008[PMID: 17967831] [Full Text]

Brown MA: Genetics and the pathogenesis of ankylosing spondylitis. Curr Opin Rheumatol 21:318, 2009[PMID: 19496308] [Full Text]

Burton PR et al: Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 39:1329, 2007[PMID: 17952073] [Full Text]

Carter J et al: Combination antibiotics as a treatment for chronic *Chlamydia*-induced reactive arthritis. Arthritis Rheum 62:1298, 2010[PMID: 20155838] [Full Text]

Chandran V et al: Sensitivity of the classification of psoriatic arthritis criteria in early psoriatic arthritis. Arthritis Rheum 57:1560, 2007[PMID: 18050230] [Full Text]

Davis JC, Jr et al: Efficacy and safety of up to 192 weeks of etanercept therapy in patients with ankylosing spondylitis. Ann Rheum Dis 67:346, 2008[PMID: 17967833] [Full Text]

Diarra D et al: Dickkopf-1 is a master regulator of joint remodeling. Nat Med 13:156, 2007[PMID: 17237793] [Full Text]

Dougados M et al: The European Spondyloarthropathy Study Group preliminary criteria for the classification of spondyloarthropathy. Arthritis Rheum 34:1218, 1991[PMID: 1930310] [Full Text]

Fitzgerald O, Winchester R: Psoriatic arthritis: From pathogenesis to therapy. Arthritis Res Ther 11:214, 2009[PMID: 19232079] [Full Text]

Inman rd et al: Efficacy and safety of golimumab in patients with ankylosing spondylitis: Results of a randomized, double-blind, placebo-controlled, phase III trial. Arthritis Rheum 58:3402, 2008[PMID: 18975305] [Full Text]

López-larrea C, Diaz-Peña R (eds): Molecular mechanisms of spondyloarthropathies, in *Advances in Experimental Medicine and Biology*, vol. 649. Austin, TX, Springer Science+ Business Media and Landes Bioscience, 2009

Magrey M, Khan MA: New insights into synovitis, acne, pustulosis, hyperostosis, and osteitis (SAPHO) syndrome. Curr Rheumatol Rep 11:329, 2009[PMID: 19772827] [Full Text]

Maksymowych WP et al: Inflammatory lesions of the spine on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis: evidence of a relationship between inflammation and new bone formation. Arthritis Rheum 60:93, 2009[PMID: 19116919] [Full Text]

Rudwaleit M et al: The early disease stage in axial spondylarthritis: Results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum 60:717, 2009[PMID: 19248087] [Full Text]

———: The development of Assessment of Spondyloarthritis International Society classification criteria for axial spondyloarthritis (pt. I): Classification of paper patients by expert opinion including uncertainty appraisal. Ann Rheum Dis 68:770, 2009

———: The development of Assessment of Spondyloarthritis International Society classification criteria for axial spondyloarthritis (pt. II): Validation and final selection. Ann Rheum Dis 68:777, 2009

Schneider T et al: Whipple's disease: New aspects of pathogenesis and treatment. Lancet Infect Dis 8:179, 2008[PMID: 18291339] [Full Text]

Sieper J et al: The Assessment of Spondyloarthritis International Society (ASAS) handbook: A guide to assess spondyloarthritis. Ann Rheum Dis 68:ii1, 2009

Singh R et al: Th1/Th17 cytokine profiles in patients with reactive arthritis/undifferentiated spondyloarthropathy. J Rheumatol 34:2285, 2007[PMID: 17937463] [Full Text]

Taylor W et al: Classification criteria for psoriatic arthritis. Development of new criteria from a large international study. Arthritis Rheum, 54:2665, 2006[PMID: 16871531] [Full Text]

Zinovieva E et al: Comprehensive linkage and association analyses identify haplotype, near to the TNFSF15 gene, significantly associated with spondyloarthritis. PLoS Genet 5:e1000528, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION

Vasculitis is a clinicopathologic process characterized by inflammation of and damage to blood vessels. The vessel lumen is usually compromised, and this is associated with ischemia of the tissues supplied by the involved vessel. A broad and heterogeneous group of syndromes may result from this process, since any type, size, and location of blood vessel may be involved. Vasculitis and its consequences may be the primary or sole manifestation of a disease; alternatively, vasculitis may be a secondary component of another primary disease. Vasculitis may be confined to a single organ, such as the skin, or it may simultaneously involve several organ systems.

CLASSIFICATION

A major feature of the vasculitic syndromes as a group is the fact that there is a great deal of heterogeneity at the same time as there is considerable overlap among them. This heterogeneity and overlap in addition to a lack of understanding of the pathogenesis of these syndromes have been major impediments to the development of a coherent classification system for these diseases. **Table 326-1** lists the major vasculitis syndromes. The distinguishing and overlapping features of these syndromes are discussed below.

Table 326-1 Vasculitis Syndromes

Primary Vasculitis Syndromes	Secondary Vasculitis Syndromes
Granulomatosis with polyangiitis (Wegener's)	Drug-induced vasculitis
Churg-Strauss syndrome	Serum sickness
Polyarteritis nodosa	Vasculitis associated with other primary diseases
Microscopic polyangiitis	Infection
Giant cell arteritis	Malignancy
Takayasu's arteritis	Rheumatic disease
Henoch-Schönlein purpura	
Idiopathiccutaneousvasculitis	
Cryoglobulinemic vasculitis	
Behçet's syndrome	
Isolated vasculitis of the central nervous system	
Cogan's syndrome	
Kawasaki disease	

PATHOPHYSIOLOGY AND PATHOGENESIS

Generally, most of the vasculitic syndromes are assumed to be mediated at least in part by immunopathogenic mechanisms that occur in response to certain antigenic stimuli **(Table 326-2)**. However, evidence supporting this hypothesis is for the most part indirect and may reflect epiphenomena as opposed to true causality. Furthermore, it is unknown why some individuals might develop vasculitis in response to certain antigenic stimuli, whereas others do not. It is likely that a number of factors are involved in the ultimate expression of a vasculitic syndrome. These include the genetic predisposition, environmental exposures, and the regulatory mechanisms associated with immune response to certain antigens.

Table 326-2 Potential Mechanisms of Vessel Damage in Vasculitis Syndromes

Pathogenic immune complex formation and/or deposition
Henoch-Schönlein purpura
Vasculitis associated with collagen vascular diseases

Serum sickness and cutaneousvasculitis syndromes
Hepatitis C–associated cryoglobulinemic vasculitis
Polyarteritis nodosa–like vasculitis associated with hepatitis B
Production of antineutrophilic cytoplasmic antibodies
Granulomatosis with polyangiitis (Wegener's)
Churg-Strauss syndrome
Microscopic polyangiitis
Pathogenic T lymphocyte responses and granuloma formation
Giant cell arteritis
Takayasu's arteritis
Granulomatosis with polyangiitis (Wegener's)
Churg-Strauss syndrome

Source: Adapted from Sneller and Fauci.

Pathogenic Immune-Complex Formation

Vasculitis is generally considered within the broader category of *immune-complex diseases* that include serum sickness and certain of the connective tissue diseases, of which systemic lupus erythematosus **(Chap. 319)** is the prototype. Although deposition of immune complexes in vessel walls is the most widely accepted pathogenic mechanism of vasculitis, the causal role of immune complexes has not been clearly established in most of the vasculitic syndromes. Circulating immune complexes need not result in deposition of the complexes in blood vessels with ensuing vasculitis, and many patients with active vasculitis do not have demonstrable circulating or deposited immune complexes. The actual antigen contained in the immune complex has only rarely been identified in vasculitic syndromes. In this regard, hepatitis B antigen has been identified in both the circulating and deposited immune complexes in a subset of patients who have features of a systemic vasculitis, most notably in polyarteritis nodosa (PAN; see "Polyarteritis Nodosa"). Cryoglobulinemic vasculitis is strongly associated with hepatitis C virus infection; hepatitis C virions and hepatitis C virus antigen-antibody complexes have been identified in the cryoprecipitates of these patients (see "Cryoglobulinemic Vasculitis").

The mechanisms of tissue damage in immune complex–mediated vasculitis resemble those described for serum sickness. In this model, antigen-antibody complexes are formed in antigen excess and are deposited in vessel walls whose permeability has been increased by vasoactive amines such as histamine, bradykinin, and leukotrienes released from platelets or from mast cells as a result of IgE-triggered mechanisms. The deposition of complexes results in activation of complement components, particularly C5a, which is strongly chemotactic for neutrophils. These cells then infiltrate the vessel wall, phagocytose the immune complexes, and release their intracytoplasmic enzymes, which damage the vessel wall. As the process becomes subacute or chronic, mononuclear cells infiltrate the vessel wall. The common denominator of the resulting syndrome is compromise of the vessel lumen with ischemic changes in the tissues supplied by the involved vessel. Several variables may explain why only certain types of immune complexes cause vasculitis and why only certain vessels are affected in individual patients. These include the ability of the reticuloendothelial system to clear circulating complexes from the blood, the size and physicochemical properties of immune complexes, the relative degree of turbulence of blood flow, the intravascular hydrostatic pressure in different vessels, and the preexisting integrity of the vessel endothelium.

ANTINEUTROPHIL CYTOPLASMIC ANTIBODIES (ANCA)

ANCA are antibodies directed against certain proteins in the cytoplasmic granules of neutrophils and monocytes. These autoantibodies are present in a high percentage of patients with active granulomatosis with polyangiitis (Wegener's) and microscopic polyangiitis, and in a lower percentage of patients with Churg-Strauss syndrome. Because these diseases share the presence of ANCA and small-vessel vasculitis, some investigators have come to refer to them collectively as "ANCA-associated vasculitis." However, as these diseases possess unique clinical phenotypes in which ANCA may be absent, it remains our opinion that granulomatosis with polyangiitis (Wegener's), microscopic polyangiitis, and Churg-Strauss syndrome should continue to be viewed as separate entities.

There are two major categories of ANCA based on different targets for the antibodies. The terminology of *cytoplasmic ANCA* (cANCA) refers to the diffuse, granular cytoplasmic staining pattern observed by immunofluorescence microscopy when serum antibodies bind to indicator neutrophils. Proteinase-3, a 29-kDa neutral serine proteinase present in neutrophil azurophilic granules, is the major cANCA antigen. More than 90% of patients with typical active granulomatosis with polyangiitis (Wegener's)

have detectable antibodies to proteinase-3 (see below). The terminology of *perinuclear ANCA* (pANCA) refers to the more localized perinuclear or nuclear staining pattern of the indicator neutrophils. The major target for pANCA is the enzyme myeloperoxidase; other targets that can produce a pANCA pattern of staining include elastase, cathepsin G, lactoferrin, lysozyme, and bactericidal/permeability-increasing protein. However, only antibodies to myeloperoxidase have been convincingly associated with vasculitis. Antimyeloperoxidase antibodies have been reported to occur in variable percentages of patients with microscopic polyangiitis, Churg-Strauss syndrome, crescentic glomerulonephritis, and granulomatosis with polyangiitis (Wegener's) (see below). A pANCA pattern of staining that is not due to antimyeloperoxidase antibodies has been associated with nonvasculitic entities such as rheumatic and nonrheumatic autoimmune diseases, inflammatory bowel disease, certain drugs, and infections such as endocarditis and bacterial airway infections in patients with cystic fibrosis.

It is unclear why patients with these vasculitis syndromes develop antibodies to myeloperoxidase or proteinase-3, whereas such antibodies are rare in other inflammatory diseases and autoimmune diseases. It is also unclear what role these antibodies play in disease pathogenesis. There are a number of in vitro observations that suggest possible mechanisms whereby these antibodies can contribute to the pathogenesis of the vasculitis syndromes. Proteinase-3 and myeloperoxidase reside in the azurophilic granules and lysosomes of resting neutrophils and monocytes, where they are apparently inaccessible to serum antibodies. However, when neutrophils or monocytes are primed by tumor necrosis factor α (TNF-α) or interleukin 1 (IL-1), proteinase-3 and myeloperoxidase translocate to the cell membrane, where they can interact with extracellular ANCA. The neutrophils then degranulate and produce reactive oxygen species that can cause tissue damage. Furthermore, ANCA-activated neutrophils can adhere to and kill endothelial cells in vitro. Activation of neutrophils and monocytes by ANCA also induces the release of proinflammatory cytokines such as IL-1 and IL-8. Recent adoptive transfer experiments in genetically engineered mice provide further evidence for a direct pathogenic role of ANCA in vivo. In contradiction, however, a number of clinical and laboratory observations argue against a primary pathogenic role for ANCA. Patients may have active granulomatosis with polyangiitis (Wegener's) in the absence of ANCA; the absolute height of the antibody titers does not correlate well with disease activity; and patients with granulomatosis with polyangiitis (Wegener's) in remission may continue to have high antiproteinase-3 (cANCA) titers for years (see below). Thus, the role of these autoantibodies in the pathogenesis of systemic vasculitis remains unclear.

Pathogenic T Lymphocyte Responses and Granuloma Formation

In addition to the classic immune complex–mediated mechanisms of vasculitis as well as ANCA, other immunopathogenic mechanisms may be involved in damage to vessels. The most prominent of these are delayed hypersensitivity and cell-mediated immune injury as reflected in the histopathologic feature of granulomatous vasculitis. However, immune complexes themselves may induce granulomatous responses. Vascular endothelial cells can express HLA class II molecules following activation by cytokines such as interferon (IFN) γ. This allows these cells to participate in immunologic reactions such as interaction with CD4+ T lymphocytes in a manner similar to antigen-presenting macrophages. Endothelial cells can secrete IL-1, which may activate T lymphocytes and initiate or propagate in situ immunologic processes within the blood vessel. In addition, IL-1 and TNF-α are potent inducers of endothelial-leukocyte adhesion molecule 1 (ELAM-1) and vascular cell adhesion molecule 1 (VCAM-1), which may enhance the adhesion of leukocytes to endothelial cells in the blood vessel wall. Other mechanisms such as direct cellular cytotoxicity, antibody directed against vessel components, or antibody-dependent cellular cytotoxicity have been suggested in certain types of vessel damage. However, there is no convincing evidence to support their causal contribution to the pathogenesis of any of the recognized vasculitic syndromes.

Approach to the Patient: General Principles of Diagnosis

The diagnosis of vasculitis is often considered in any patient with an unexplained systemic illness. However, there are certain clinical abnormalities that when present alone or in combination should suggest a diagnosis of vasculitis. These include palpable purpura, pulmonary infiltrates and microscopic hematuria, chronic inflammatory sinusitis, mononeuritis multiplex, unexplained ischemic events, and glomerulonephritis with evidence of multisystem disease. A number of nonvasculitic diseases may also produce some or all of these abnormalities. Thus, the first step in the workup of a patient with suspected vasculitis is to exclude other diseases that produce clinical manifestations that can mimic vasculitis **(Table 326-3)**. It is particularly important to exclude infectious diseases with features that overlap those of vasculitis, especially if the patient's clinical condition is deteriorating rapidly and empirical immunosuppressive treatment is being contemplated.

Table 326-3 Conditions that Can Mimic Vasculitis

Infectious diseases
Bacterial endocarditis
Disseminated gonococcal infection

- Pulmonary histoplasmosis
- Coccidioidomycosis
- Syphilis
- Lyme disease
- Rocky Mountain spotted fever
- Whipple's disease

Coagulopathies/thrombotic microangiopathies
- Antiphospholipid antibody syndrome
- Thrombotic thrombocytopenic purpura

Neoplasms
- Atrial myxoma
- Lymphoma
- Carcinomatosis

Drug toxicity
- Cocaine
- Amphetamines
- Ergot alkaloids
- Methysergide
- Arsenic

Sarcoidosis

Atheroembolic disease

Antiglomerular basement membrane disease (Goodpasture's syndrome)

Amyloidosis

Migraine

Once diseases that mimic vasculitis have been excluded, the workup should follow a series of progressive steps that establish the diagnosis of vasculitis and determine, where possible, the category of the vasculitis syndrome **(Fig. 326-1)**. This approach is of considerable importance since several of the vasculitis syndromes require aggressive therapy with glucocorticoids and cytotoxic agents, while other syndromes usually resolve spontaneously and require symptomatic treatment only. The definitive diagnosis of vasculitis is made upon biopsy of involved tissue. The yield of "blind" biopsies of organs with no subjective or objective evidence of involvement is very low and should be avoided. When syndromes such as PAN, Takayasu's arteritis, or isolated central nervous system (CNS) vasculitis are suspected, arteriogram of organs with suspected involvement should be performed. However, arteriograms should not be performed routinely when patients present with localized cutaneous vasculitis with no clinical indication of visceral involvement.

Figure 326-1

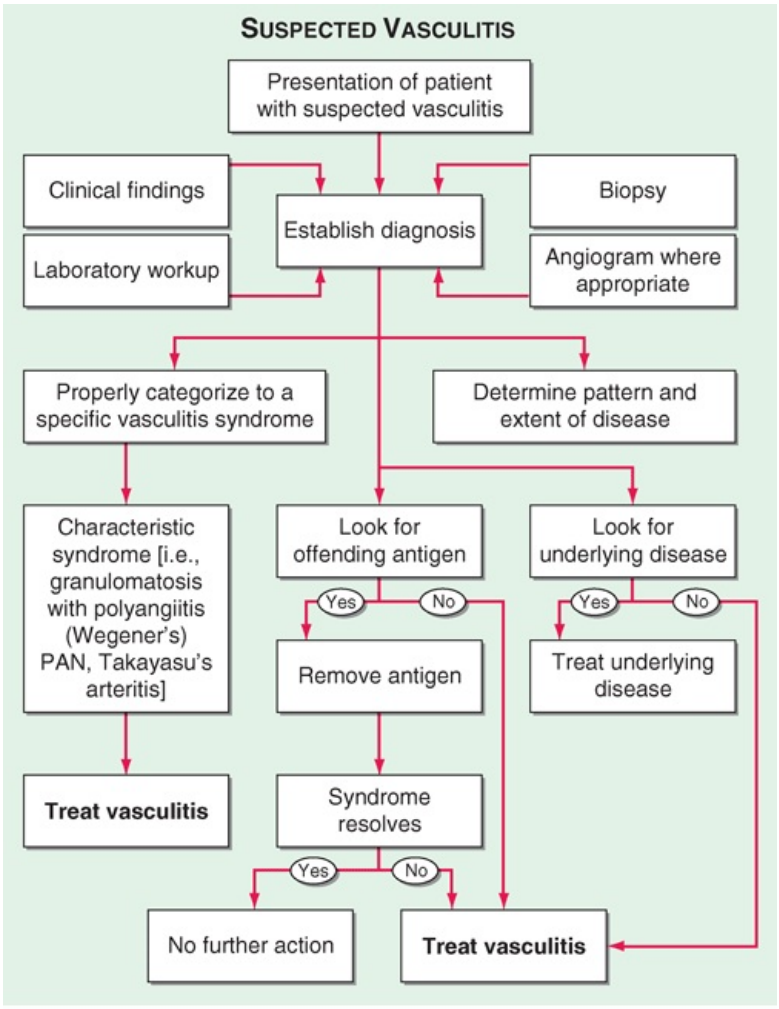

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the approach to a patient with suspected diagnosis of vasculitis. PAN, polyarteritis nodosa.

GENERAL PRINCIPLES OF TREATMENT

Once a diagnosis of vasculitis has been established, a decision regarding therapeutic strategy must be made **(Fig. 326-1)**. If an offending antigen that precipitates the vasculitis is recognized, the antigen should be removed where possible. If the vasculitis is associated with an underlying disease such as an infection, neoplasm, or connective tissue disease, the underlying disease should be treated. If the syndrome represents a primary vasculitic disease, treatment should be initiated according to the category of the vasculitis syndrome. Specific therapeutic regimens are discussed below for the individual vasculitis syndromes; however, certain general principles regarding therapy should be considered. Decisions regarding treatment should be based upon the use of regimens for which there has been published literature supporting efficacy for that particular vasculitic disease. Since the potential toxic side effects of certain therapeutic regimens may be substantial, the risk-versus-benefit ratio of any therapeutic approach should be weighed carefully. On the one hand, glucocorticoids and/or cytotoxic therapy should be instituted immediately in diseases where irreversible organ system dysfunction and high morbidity and mortality rates have been clearly established. Granulomatosis with polyangiitis (Wegener's) is the prototype of a severe systemic vasculitis requiring such a therapeutic approach (see below). On the other hand, when feasible, aggressive therapy should be avoided for vasculitic manifestations that rarely result in irreversible organ system dysfunction and that usually do not respond to such therapy. For example, idiopathiccutaneous vasculitis usually resolves with symptomatic treatment, and prolonged courses of glucocorticoids uncommonly result in clinical benefit. Cytotoxic agents have not proved to be beneficial in idiopathic cutaneous vasculitis, and their toxic side effects generally outweigh any potential beneficial effects. Glucocorticoids should be initiated in those systemic vasculitides that cannot be specifically categorized or for which there is no established standard therapy; cytotoxic therapy should be added in these diseases only if an adequate response does not result or if remission can only be achieved and maintained with an unacceptably toxic regimen of glucocorticoids. When remission is achieved, one should continually attempt to taper glucocorticoids and discontinue when possible. When using cytotoxic regimens, one should base the choice of agent upon the available therapeutic data supporting efficacy in that disease, the site and

severity of organ involvement, and the toxicity profile of the drug.

Physicians should be thoroughly aware of the toxic side effects of therapeutic agents employed that can include both acute and long-term complications **(Table 326-4)**. Morbidity and mortality can occur as a result of treatment and strategies to monitor for and prevent toxicity represent an essential part of patient care. Glucocorticoids are an important part of treatment for most vasculitides but are associated with substantial toxicities. Monitoring and prevention of glucocorticoid-induced bone loss is important in all patients. With the use of daily cyclophosphamide, strategies are particularly important and are directed toward minimization of bladder toxicity and prevention of leukopenia. Instructing the patient to take cyclophosphamide all at once in the morning with a large amount of fluid throughout the day in order to maintain a dilute urine can reduce the risk of bladder injury. Bladder cancer can occur several years after discontinuation of cyclophosphamide therapy; therefore, monitoring for bladder cancer should continue indefinitely in patients who have received cyclophosphamide. Bone marrow suppression is an important toxicity of cyclophosphamide and can be observed during glucocorticoid tapering or over time, even after periods of stable measurements. Monitoring of the complete blood count every 1–2 weeks for as long as the patient receives cyclophosphamide can effectively prevent cytopenias. Maintaining the white blood count (WBC) at >3000/μL and the neutrophil count >1500/μL is essential to reduce the risk of life-threatening infections.

Table 326-4 Major Toxic Side Effects of Drugs Commonly Used in the Treatment of Systemic Vasculitis

Glucocorticoids	
Osteoporosis	Growth suppression in children
Cataracts	Hypertension
Glaucoma	Avascular necrosis of bone
Diabetes mellitus	Myopathy
Electrolyte abnormalities	Alterations in mood
Metabolic abnormalities	Psychosis
Suppression of inflammatory and immune responses leading to opportunistic infections	Pseudotumor cerebri Peptic ulcer diathesis Pancreatitis
Cushingoid features	
Cyclophosphamide	
Bone marrow suppression	Hypogammaglobulinemia
Cystitis	Pulmonary fibrosis
Bladder carcinoma	Myelodysplasia
Gonadal suppression	Oncogenesis
Gastrointestinal intolerance	Teratogenicity
	Opportunistic infections
Methotrexate	
Gastrointestinal intolerance	Pneumonitis
Stomatitis	Teratogenicity
Bone marrow suppression	Opportunistic infections
Hepatotoxicity (may lead to fibrosis or cirrhosis)	
Azathioprine	
Gastrointestinal intolerance	Opportunistic infections
Bone marrow suppression	Hypersensitivity
Hepatotoxicity	

Methotrexate and azathioprine are also associated with bone marrow suppression, and complete blood counts should be obtained every 1–2 weeks for the first 1–2 months after their initiation and once a month thereafter. To lessen toxicity, methotrexate is often given together with folic acid, 1 mg daily, or folinic acid, 5–10 mg once a week 24 h following methotrexate. Prior to initiation of azathioprine, thiopurine methyltransferase (TPMT), an enzyme involved in the metabolism of azathioprine, should be assayed because inadequate levels may result in severe cytopenia.

Infection represents a significant toxicity for all vasculitis patients treated with immunosuppressive therapy. Infections with *Pneumocystis jiroveci* and certain fungi can be seen even in the face of WBCs that are within normal limits, particularly in patients receiving glucocorticoids. All vasculitis patients who are receiving daily glucocorticoids in combination with a cytotoxic drug should receive trimethoprim-sulfamethoxazole (TMP-SMX) or another prophylactic therapy to prevent *P. jiroveci* infection.

Finally, it should be emphasized that each patient is unique and requires individual decision-making. The above outline should serve as a framework to guide therapeutic approaches; however, flexibility should be practiced in order to provide maximal therapeutic efficacy with minimal toxic side effects in each patient.

GRANULOMATOSIS WITH POLYANGIITIS (WEGENER'S)

Definition

Granulomatosis with polyangiitis (Wegener's) is a distinct clinicopathologic entity characterized by granulomatous vasculitis of the upper and lower respiratory tracts together with glomerulonephritis. In addition, variable degrees of disseminated vasculitis involving both small arteries and veins may occur.

Incidence and Prevalence

Granulomatosis with polyangiitis (Wegener's) is an uncommon disease with an estimated prevalence of 3 per 100,000. It is extremely rare in blacks compared with whites; the male-to-female ratio is 1:1. The disease can be seen at any age; ~15% of patients are <19 years of age, but only rarely does the disease occur before adolescence; the mean age of onset is ~40 years.

Pathology and Pathogenesis

The histopathologic hallmarks of granulomatosis with polyangiitis (Wegener's) are necrotizing vasculitis of small arteries and veins together with granuloma formation, which may be either intravascular or extravascular **(Fig. 326-2)**. Lung involvement typically appears as multiple, bilateral, nodular cavitary infiltrates **(Fig. 326-3)**, which on biopsy almost invariably reveal the typical necrotizing granulomatous vasculitis. Upper airway lesions, particularly those in the sinuses and nasopharynx, typically reveal inflammation, necrosis, and granuloma formation, with or without vasculitis.

Figure 326-2

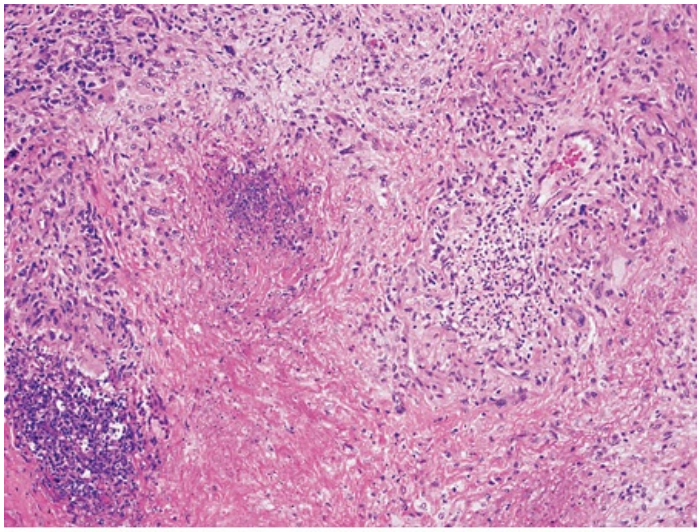

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lung histology in granulomatosis with polyangiitis (Wegener's). This area of geographic necrosis has a serpiginous border of histiocytes and giant cells surrounding a central necrotic zone. Vasculitis is also present with neutrophils and lymphocytes infiltrating the wall of a small arteriole *(upper right). (Courtesy of William D. Travis, MD; with permission.)*

Figure 326-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomography scan of a patient with granulomatosis with polyangiitis (Wegener's). The patient developed multiple, bilateral, and cavitary infiltrates.

In its earliest form, renal involvement is characterized by a focal and segmental glomerulitis that may evolve into a rapidly progressive crescentic glomerulonephritis. Granuloma formation is only rarely seen on renal biopsy. In contrast to other forms of glomerulonephritis, evidence of immune complex deposition is not found in the renal lesion of granulomatosis with polyangiitis (Wegener's). In addition to the classic triad of disease of the upper and lower respiratory tracts and kidney, virtually any organ can be involved with vasculitis, granuloma, or both.

The immunopathogenesis of this disease is unclear, although the involvement of upper airways and lungs with granulomatous vasculitis suggests an aberrant cell-mediated immune response to an exogenous or even endogenous antigen that enters through or resides in the upper airway. Chronic nasal carriage of *Staphylococcus aureus* has been reported to be associated with a higher relapse rate of granulomatosis with polyangiitis (Wegener's); however, there is no evidence for a role of this organism in the pathogenesis of the disease.

Peripheral blood mononuclear cells obtained from patients with granulomatosis with polyangiitis (Wegener's) manifest increased secretion of IFN-γ but not of IL-4, IL-5, or IL-10 compared to normal controls. In addition, TNF-α production from peripheral blood mononuclear cells and CD4+ T cells is elevated. Furthermore, monocytes from patients with granulomatosis with polyangiitis (Wegener's) produce increased amounts of IL-12. These findings indicate an unbalanced T_H1-type T cell cytokine pattern in this disease that may have pathogenic and perhaps ultimately therapeutic implications.

A high percentage of patients with granulomatosis with polyangiitis (Wegener's) develop ANCA, and these autoantibodies may play a role in the pathogenesis of this disease (see above).

Clinical and Laboratory Manifestations

Involvement of the upper airways occurs in 95% of patients with granulomatosis with polyangiitis (Wegener's). Patients often present with severe upper respiratory tract findings such as paranasal sinus pain and drainage and purulent or bloody nasal discharge, with or without nasal mucosal ulceration **(Table 326-5)**. Nasal septal perforation may follow, leading to saddle nose deformity. Serous otitis media may occur as a result of eustachian tube blockage. Subglottic tracheal stenosis resulting from active disease or scarring occurs in ~16% of patients and may result in severe airway obstruction.

Table 326-5 Granulomatosis with Polyangiitis (Wegener's): Frequency of Clinical Manifestations in 158 Patients Studied at the National Institutes of Health

Manifestation	Percent at Disease Onset	Percent Throughout Course of Disease
Kidney	18	77
Glomerulonephritis		
Ear/nose/throat	73	92
Sinusitis	51	85

Nasal disease	36	68
Otitis media	25	44
Hearing loss	14	42
Subglottic stenosis	1	16
Ear pain	9	14
Oral lesions	3	10
Lung	45	85
Pulmonary infiltrates	25	66
Pulmonary nodules	24	58
Hemoptysis	12	30
Pleuritis	10	28
Eyes	5	18
Conjunctivitis	1	18
Dacryocystitis	6	16
Scleritis	2	15
Proptosis	3	11
Eye pain	0	8
Visual loss	0	4
Retinal lesions	0	1
Corneal lesions	0	2
Iritis		
Other[a]	32	67
Arthralgias/arthritis	23	50
Fever	19	46
Cough	13	46
Skin abnormalities	15	35
Weight loss (>10% body weight)	1	15
Peripheral neuropathy	1	8
Central nervous system disease	2	6
Pericarditis	1	3
Hyperthyroidism		

[a]Fewer than 1% had parotid, pulmonary artery, breast, or lower genitourinary (urethra, cervix, vagina, testicular) involvement.

Source: Hoffman et al.

Pulmonary involvement may be manifested as asymptomatic infiltrates or may be clinically expressed as cough, hemoptysis, dyspnea, and chest discomfort. It is present in 85–90% of patients. Endobronchial disease, either in its active form or as a result of fibrous scarring, may lead to obstruction with atelectasis.

Eye involvement (52% of patients) may range from a mild conjunctivitis to dacryocystitis, episcleritis, scleritis, granulomatous sclerouveitis, ciliary vessel vasculitis, and retroorbital mass lesions leading to proptosis.

Skin lesions (46% of patients) appear as papules, vesicles, palpable purpura, ulcers, or subcutaneous nodules; biopsy reveals vasculitis, granuloma, or both. Cardiac involvement (8% of patients) manifests as pericarditis, coronary vasculitis, or, rarely, cardiomyopathy. Nervous system manifestations (23% of patients) include cranial neuritis, mononeuritis multiplex, or, rarely, cerebral vasculitis and/or granuloma.

Renal disease (77% of patients) generally dominates the clinical picture and, if left untreated, accounts directly or indirectly for most of the mortality rate in this disease. Although it may smolder in some cases as a mild glomerulitis with proteinuria, hematuria, and red blood cell casts, it is clear that once

clinically detectable renal functional impairment occurs, rapidly progressive renal failure usually ensues unless appropriate treatment is instituted.

While the disease is active, most patients have nonspecific symptoms and signs such as malaise, weakness, arthralgias, anorexia, and weight loss. Fever may indicate activity of the underlying disease but more often reflects secondary infection, usually of the upper airway.

Characteristic laboratory findings include a markedly elevated erythrocyte sedimentation rate (ESR), mild anemia and leukocytosis, mild hypergammaglobulinemia (particularly of the IgA class), and mildly elevated rheumatoid factor. Thrombocytosis may be seen as an acute-phase reactant. Approximately 90% of patients with active granulomatosis with polyangiitis (Wegener's) have a positive antiproteinase-3 ANCA. However, in the absence of active disease, the sensitivity drops to ~60–70%. A small percentage of patients with granulomatosis with polyangiitis (Wegener's) may have antimyeloperoxidase rather than antiproteinase-3 antibodies, and up to 20% may lack ANCA.

Patients with granulomatosis with polyangiitis (Wegener's) have been found to have an increased incidence of venous thrombotic events. Although routine anticoagulation for all patients is not recommended, a heightened awareness for any clinical features suggestive of deep venous thrombosis or pulmonary emboli is warranted.

Diagnosis

The diagnosis of granulomatosis with polyangiitis (Wegener's) is made by the demonstration of necrotizing granulomatous vasculitis on tissue biopsy in a patient with compatible clinical features. Pulmonary tissue offers the highest diagnostic yield, almost invariably revealing granulomatous vasculitis. Biopsy of upper airway tissue usually reveals granulomatous inflammation with necrosis but may not show vasculitis. Renal biopsy can confirm the presence of pauci-immune glomerulonephritis.

The specificity of a positive antiproteinase-3 ANCA for granulomatosis with polyangiitis (Wegener's) is very high, especially if active glomerulonephritis is present. However, the presence of ANCA should be adjunctive and, with rare exceptions, should not substitute for a tissue diagnosis. False-positive ANCA titers have been reported in certain infectious and neoplastic diseases.

In its typical presentation, the clinicopathologic complex of granulomatosis with polyangiitis (Wegener's) usually provides ready differentiation from other disorders. However, if all the typical features are not present at once, it needs to be differentiated from the other vasculitides, antiglomerular basement membrane disease (Goodpasture's syndrome) **(Chap. 283)**, relapsing polychondritis **(Chap. 328)**, tumors of the upper airway or lung, and infectious diseases such as histoplasmosis **(Chap. 199)**, mucocutaneous leishmaniasis **(Chap. 212)**, and rhinoscleroma **(Chap. 31)** as well as noninfectious granulomatous diseases.

Of particular note is the differentiation from *midline granuloma* and *upper airway neoplasms* , which are part of the spectrum of *midline destructive diseases*. These diseases lead to extreme tissue destruction and mutilation localized to the midline upper airway structures including the sinuses; erosion through the skin of the face commonly occurs, a feature that is extremely rare in granulomatosis with polyangiitis (Wegener's). Although blood vessels may be involved in the intense inflammatory reaction and necrosis, primary vasculitis is not seen. Midline granuloma is part of the spectrum of *angiocentric immunoproliferative lesions* that are considered to represent a spectrum of postthymic T cell proliferative lesions and should be treated as such **(Chap. 110)**. The term *idiopathic* has been applied to midline granuloma when extensive diagnostic workup including multiple biopsies has failed to reveal anything other than inflammation and necrosis. Under these circumstances, it is possible that the tumor cells were masked by the intensive inflammatory response. Such cases have responded to local irradiation with 50 Gy (5000 rad). Upper airway lesions should never be irradiated in granulomatosis with polyangiitis (Wegener's). Cocaine-induced tissue injury can be another important mimic of granulomatosis with polyangiitis (Wegener's) in patients who present with isolated midline destructive disease. ANCA that target human neutrophil elastase can be found in patients with cocaine-induced midline destructive lesions and can confound the differentiation from granulomatosis with polyangiitis (Wegener's).

Granulomatosis with polyangiitis (Wegener's) must also be differentiated from *lymphomatoid granulomatosis*, which is an Epstein-Barr virus–positive B cell proliferation that is associated with an exuberant T cell reaction. Lymphomatoid granulomatosis is characterized by lung, skin, CNS, and kidney involvement in which atypical lymphocytoid and plasmacytoid cells infiltrate nonlymphoid tissue in an angioinvasive manner. In this regard, it clearly differs from granulomatosis with polyangiitis (Wegener's) in that it is not an inflammatory vasculitis in the classic sense but an infiltration of vessels with atypical mononuclear cells; granuloma may be present in involved tissues. Up to 50% of patients may develop a true malignant lymphoma.

Treatment: Granulomatosis with Polyangiitis (Wegener's)

Prior to the introduction of effective therapy, granulomatosis with polyangiitis (Wegener's) was universally fatal within a few months of diagnosis. Glucocorticoids alone led to some symptomatic improvement, with little effect on the ultimate course of the disease. The development of treatment with

cyclophophamide dramatically changed patient outcome such that marked improvement was seen in >90% of patients, complete remission in 75% of patients, and 5-year patient survival was seen in over 80%.

Despite the ability to successfully induce remission, 50–70% of remissions are later associated with one or more relapses. The determination of relapse should be based on objective evidence of disease activity, taking care to rule out other features that may have a similar appearance such as infection, medication toxicity, or chronic disease sequelae. The ANCA titer can be misleading and should not be used to assess disease activity. Many patients who achieve remission continue to have elevated titers for years. Results from a large prospective study found that increases in ANCA were not associated with relapse and that only 43% relapsed within 1 year of an increase in ANCA levels. Thus, a rise in ANCA by itself is not a harbinger of immediate disease relapse and should not lead to reinstitution or increase in immunosuppressive therapy.

Reinduction of remission after relapse is almost always achieved; however, a high percentage of patients ultimately have some degree of damage from irreversible features of their disease, such as varying degrees of renal insufficiency, hearing loss, tracheal stenosis, saddle nose deformity, and chronically impaired sinus function. Patients who developed irreversible renal failure but who achieved subsequent remission have undergone successful renal transplantation.

Because long-term cyclophosphamide is associated with substantial toxicity, approaches have been developed that seek to minimize the duration of exposure to cyclophosphamide while still taking advantage of its efficacy for severe disease. Treatment of granulomatosis with polyangiitis (Wegener's) is currently viewed as having two phases: *induction*, where active disease is put into remission, followed by *maintenance*. The decision regarding which agents to use for induction and maintenance is based upon disease severity together with individual patient factors that include contraindication, relapse history, and comorbidities.

CYCLOPHOSPHAMIDE INDUCTION FOR SEVERE DISEASE?

For patients with severe disease, daily cyclophosphamide combined with glucocorticoids has been repeatedly proved to effectively induce remission and prolong survival. At the initiation of therapy, glucocorticoids are usually given as prednisone, 1 mg/kg per day for the first month, followed by gradual tapering on an alternate-day or daily schedule with discontinuation after ~6–9 months. Cyclophosphamide is given in doses of 2 mg/kg per day orally, but as it is renally eliminated, dosage reduction should be considered in patients with renal insufficiency. Some reports have indicated therapeutic success with less frequent and severe toxic side effects using IV cyclophosphamide. In a recent randomized trial, IV cyclophosphamide 15 mg/kg, three infusions given every 2 weeks, then every 3 weeks thereafter, was compared to cyclophosphamide 2 mg/kg daily given for 3 months followed by 1.5 mg/kg daily. Although IV cyclophosphamide was found to have a comparable rate of remission with a lower cumulative cyclophosphamide dose and occurrence of leukopenia, the use of a consolidation phase and an insufficient frequency of blood count monitoring may have negatively influenced the results in those who received daily cyclophosphamide. Of note in this study was that relapse occurred in 19% of those who received IV cyclophosphamide as compared to 9% who received daily oral administration. We continue to strongly favor daily cyclophosphamide with utilization of blood count monitoring every 1–2 weeks (as discussed above) and limiting the duration of induction exposure to 3–6 months.

In patients with imminently life-threatening disease, such as rapidly progressive glomerulonephritis or pulmonary hemorrhage requiring mechanical ventilation, a regimen of daily cyclophosphamide and glucocorticoids is the treatment of choice to induce remission. Adjunctive plasmapheresis was found to further improve renal recovery in a study of patients with rapidly progressive glomerulonephritis who had a creatinine of greater than 5.8 mg/dL.

REMISSION MAINTENANCE AFTER CYCLOPHOSPHAMIDE?

After 3–6 months of induction treatment, cyclophosphamide should be stopped and switched to another agent for remission maintenance. The agents with which there has been the greatest published experience are methotrexate and azathioprine. Methotrexate is administered orally or subcutaneously starting at a dosage of 0.3 mg/kg as a single weekly dose, not to exceed 15 mg/week. If the treatment is well tolerated after 1–2 weeks, the dosage should be increased by 2.5 mg weekly up to a dosage of 20–25 mg/week and maintained at that level. Azathioprine, 2 mg/kg per day, has also proved effective in maintaining remission following induction with daily cyclophosphamide. In a randomized trial comparing methotrexate to azathioprine for remission maintenance, comparable rates of toxicity and relapse were seen. Therefore, the choice of agent is often based on toxicity profile, as methotrexate cannot be given to patients with renal insufficiency or chronic liver disease, as well as on other individual patient factors. In patients who are unable to receive methotrexate or azathioprine or who have relapsed through such treatment, mycophenolate mofetil, 1000 mg twice a day, may also sustain remission following cyclophosphamide induction.

The optimal duration of maintenance therapy is uncertain. In the absence of toxicity, maintenance therapy is usually given for a minimum of 2 years past remission, after which time consideration can be

given for tapering over a 6–12 month period until discontinuation. Some patients with significant organ damage or a history of relapse may benefit from longer-term continuation of a maintenance agent.

METHOTREXATE INDUCTION FOR NONSEVERE DISEASE?

For selected patients whose disease is not immediately life threatening or in those patients who have experienced significant cyclophosphamide toxicity, methotrexate together with glucocorticoids given at the dosages described above may be considered as an alternative for induction therapy, which is then continued for maintenance.

RITUXIMAB INDUCTION FOR SEVERE DISEASE?

Rituximab is a chimeric monoclonal antibody directed against CD20 present on normal and malignant B lymphocytes that is FDA approved for the treatment of non-Hodgkin's lymphoma, chronic lymphocytic lymphoma, and rheumatoid arthritis. In two recent randomized trials that enrolled ANCA positive patients with severe active granulomatosis with polyangiitis (Wegener's) or microscopic polyangiitis, rituximab 375 mg/m^2 once a week for 4 weeks in combination with glucocorticoids was found to be as effective as cyclophosphamide with glucocorticoids for inducing disease remission. In the trial which also enrolled patients with relapsing disease, rituximab was found to be statistically superior to cyclophosphamide.

While the data supports that rituximab is effective for remission induction severe active granulomatosis with polyangiitis (Wegener's) or microscopic polyangiitis, there remain a number of ongoing questions regarding rituximab that must be considered in weighing its use in the individual patient. These include that there are no long-term data regarding relapse risk or long-term safety with rituximab, it is unclear how often rituximab needs to be given, and as all patients in the randomized trials were ANCA positive, its efficacy in ANCA negative patients is unknown. It is also uncertain whether the use of other maintenance agents after rituximab would provide any additional benefit in prolonging remission or increase toxicity as these were not used in either randomized trial.

While rituximab does not have the bladder toxicity or infertility concerns, as can occur with cyclophosphamide, in both of the randomized trials, the rate of adverse events was similar in the rituximab and cyclophosphamide arms. Serious side effects of rituximab include infusion reactions, severe mucocutaneous reactions, and rare reports of progressive multifocal leukoencephalopathy. As rituximab can bring about reactivation of hepatitis B, all patients should undergo hepatitis screening prior to treatment with rituximab.

OTHER BIOLOGIC THERAPIES?

Etanercept, a dimeric fusion protein containing the 75-kDa TNF receptor bound to human IgG1, was not found to sustain remission when used adjunctively to standard therapy and should not be used in the treatment of granulomatosis with polyangiitis (Wegener's).

TRIMETHOPRIM-SULFAMETHOXAZOLE?

Although certain reports have indicated that TMP-SMX may be of benefit in the treatment of granulomatosis with polyangiitis (Wegener's) isolated to the sinonasal tissues, it should never be used alone to treat active granulomatosis with polyangiitis (Wegener's) outside of the upper airway such as in patients with renal or pulmonary disease. In a study examining the effect of TMP-SMX on relapse, decreased relapses were shown only with regard to upper airway disease, and no differences in major organ relapses were observed.

ORGAN-SPECIFIC TREATMENT

Not all manifestations of granulomatosis with polyangiitis (Wegener's) require or respond to cytotoxic therapy. In managing non-major organ disease, such as that isolated to the sinus, joints, or skin, the risks of treatment should be carefully weighed against the benefits. Treatment with cyclophosphamide is rarely if ever justified for the treatment of isolated sinus disease in granulomatosis with polyangiitis (Wegener's). Although patients with non-major organ disease may be effectively treated without cytotoxic therapy, these individuals must be monitored closely for the development of disease activity affecting the lungs, kidneys, or other major organs. Subglottic tracheal stenosis and endobronchial stenosis are examples of disease manifestations that do not typically respond to systemic immunosuppressive treatment.

MICROSCOPIC POLYANGIITIS

Definition

The term *microscopic polyarteritis* was introduced into the literature by Davson in 1948 in recognition of the presence of glomerulonephritis in patients with PAN. In 1992, the Chapel Hill Consensus Conference on the Nomenclature of Systemic Vasculitis adopted the term *microscopic polyangiitis* to connote a necrotizing vasculitis with few or no immune complexes affecting small vessels (capillaries, venules, or arterioles). Glomerulonephritis is very common in microscopic polyangiitis, and pulmonary capillaritis often occurs. The absence of granulomatous inflammation in microscopic polyangiitis is said to differentiate it from granulomatosis with polyangiitis (Wegener's).

Incidence and Prevalence

The incidence of microscopic polyangiitis has not yet been reliably established due to its previous inclusion as part of PAN. The mean age of onset is ~57 years of age, and males are slightly more frequently affected than females.

Pathology and Pathogenesis

The vasculitis seen in microscopic polyangiitis has a predilection to involve capillaries and venules in addition to small and medium-sized arteries. Immunohistochemical staining reveals a paucity of immunoglobulin deposition in the vascular lesion of microscopic polyangiitis, suggesting that immune-complex formation does not play a role in the pathogenesis of this syndrome. The renal lesion seen in microscopic polyangiitis is identical to that of granulomatosis with polyangiitis (Wegener's). Like granulomatosis with polyangiitis (Wegener's), microscopic polyangiitis is highly associated with the presence of ANCA, which may play a role in pathogenesis of this syndrome (see above).

Clinical and Laboratory Manifestations

Because of its predilection to involve the small vessels, microscopic polyangiitis and granulomatosis with polyangiitis (Wegener's) share similar clinical features. Disease onset may be gradual, with initial symptoms of fever, weight loss, and musculoskeletal pain; however, it is often acute. Glomerulonephritis occurs in at least 79% of patients and can be rapidly progressive, leading to renal failure. Hemoptysis may be the first symptom of alveolar hemorrhage, which occurs in 12% of patients. Other manifestations include mononeuritis multiplex and gastrointestinal tract and cutaneousvasculitis. Upper airway disease and pulmonary nodules are not typically found in microscopic polyangiitis and, if present, suggest granulomatosis with polyangiitis (Wegener's).

Features of inflammation may be seen, including an elevated ESR, anemia, leukocytosis, and thrombocytosis. ANCA are present in 75% of patients with microscopic polyangiitis, with antimyeloperoxidase antibodies being the predominant ANCA associated with this disease.

Diagnosis

The diagnosis is based on histologic evidence of vasculitis or pauci-immune glomerulonephritis in a patient with compatible clinical features of multisystem disease. Although microscopic polyangiitis is strongly ANCA-associated, no studies have as yet established the sensitivity and specificity of ANCA in this disease.

Treatment: Microscopic Polyangiitis

The 5-year survival rate for patients with treated microscopic polyangiitis is 74%, with disease-related mortality occurring from alveolar hemorrhage or gastrointestinal, cardiac, or renal disease. Studies on treatment have come from trials that have included patients with granulomatosis with polyangiitis (Wegener's) or microscopic polyangiitis. Currently, the treatment approach for microscopic polyangiitis is the same as is used for granulomatosis with polyangiitis (Wegener's) [see "Granulomatosis With Polyangiitis (Wegener's)"] for a detailed description of this therapeutic regimen), and patients with immediately life-threatening disease should be treated with the combination of prednisone and daily cyclophosphamide. Recent studies with rituximab also included ANCA positive patients with microscopic polyangiitis. Disease relapse has been observed in at least 34% of patients. Treatment for such relapses would be similar to that used at the time of initial presentation and based upon site and severity of disease.

CHURG-STRAUSS SYNDROME

Definition

Churg-Strauss syndrome, also referred to as *allergic angiitis and granulomatosis*, was described in 1951 by Churg and Strauss and is characterized by asthma, peripheral and tissue eosinophilia, extravascular granuloma formation, and vasculitis of multiple organ systems.

Incidence and Prevalence

Churg-Strauss syndrome is an uncommon disease with an estimated annual incidence of 1–3 per million. The disease can occur at any age with the possible exception of infants. The mean age of onset is 48 years, with a female-to-male ratio of 1.2:1.

Pathology and Pathogenesis

The necrotizing vasculitis of Churg-Strauss syndrome involves small and medium-sized muscular arteries, capillaries, veins, and venules. A characteristic histopathologic feature of Churg-Strauss syndrome is granulomatous reactions that may be present in the tissues or even within the walls of the vessels themselves. These are usually associated with infiltration of the tissues with eosinophils. This process can occur in any organ in the body; lung involvement is predominant, with skin, cardiovascular system, kidney, peripheral nervous system, and gastrointestinal tract also commonly involved. Although the precise pathogenesis of this disease is uncertain, its strong association with asthma and its clinicopathologic manifestations, including eosinophilia, granuloma, and vasculitis, point to aberrant

immunologic phenomena.

Clinical and Laboratory Manifestations

Patients with Churg-Strauss syndrome often exhibit nonspecific manifestations such as fever, malaise, anorexia, and weight loss, which are characteristic of a multisystem disease. The pulmonary findings in Churg-Strauss syndrome clearly dominate the clinical picture with severe asthmatic attacks and the presence of pulmonary infiltrates. Mononeuritis multiplex is the second most common manifestation and occurs in up to 72% of patients. Allergic rhinitis and sinusitis develop in up to 61% of patients and are often observed early in the course of disease. Clinically recognizable heart disease occurs in ~14% of patients and is an important cause of mortality. Skin lesions occur in ~51% of patients and include purpura in addition to cutaneous and subcutaneous nodules. The renal disease in Churg-Strauss syndrome is less common and generally less severe than that of granulomatosis with polyangiitis (Wegener's) and microscopic polyangiitis.

The characteristic laboratory finding in virtually all patients with Churg-Strauss syndrome is a striking eosinophilia, which reaches levels >1000 cells/μL in >80% of patients. Evidence of inflammation as evidenced by elevated ESR, fibrinogen, or α_2-globulins can be found in 81% of patients. The other laboratory findings reflect the organ systems involved. Approximately 48% of patients with Churg-Strauss syndrome have circulating ANCA that is usually antimyeloperoxidase.

Diagnosis

Although the diagnosis of Churg-Strauss syndrome is optimally made by biopsy in a patient with the characteristic clinical manifestations (see above), histologic confirmation can be challenging as the pathognomonic features often do not occur simultaneously. In order to be diagnosed with Churg-Strauss syndrome, a patient should have evidence of asthma, peripheral blood eosinophilia, and clinical features consistent with vasculitis.

Treatment: Churg-Strauss Syndrome

The prognosis of untreated Churg-Strauss syndrome is poor, with a reported 5-year survival of 25%. With treatment, prognosis is favorable, with one study finding a 78-month actuarial survival rate of 72%. Myocardial involvement is the most frequent cause of death and is responsible for 39% of patient mortality. Glucocorticoids alone appear to be effective in many patients. Dosage tapering is often limited by asthma, and many patients require low-dose prednisone for persistent asthma many years after clinical recovery from vasculitis. In glucocorticoid failure or in patients who present with fulminant multisystem disease, the treatment of choice is a combined regimen of daily cyclophosphamide and prednisone (see "Granulomatosis With Polyangiitis (Wegener's)" for a detailed description of this therapeutic regimen).

POLYARTERITIS NODOSA

Definition

PAN, also referred to as *classic PAN*, was described in 1866 by Kussmaul and Maier. It is a multisystem, necrotizing vasculitis of small and medium-sized muscular arteries in which involvement of the renal and visceral arteries is characteristic. PAN does not involve pulmonary arteries, although bronchial vessels may be involved; granulomas, significant eosinophilia, and an allergic diathesis are not observed.

Incidence and Prevalence

It is difficult to establish an accurate incidence of PAN because previous reports have included PAN and microscopic polyangiitis as well as other related vasculitides. PAN, as currently defined, is felt to be a very uncommon disease.

Pathology and Pathogenesis

The vascular lesion in PAN is a necrotizing inflammation of small and medium-sized muscular arteries. The lesions are segmental and tend to involve bifurcations and branchings of arteries. They may spread circumferentially to involve adjacent veins. However, involvement of venules is not seen in PAN and, if present, suggests microscopic polyangiitis (see below). In the acute stages of disease, polymorphonuclear neutrophils infiltrate all layers of the vessel wall and perivascular areas, which results in intimal proliferation and degeneration of the vessel wall. Mononuclear cells infiltrate the area as the lesions progress to the subacute and chronic stages. Fibrinoid necrosis of the vessels ensues with compromise of the lumen, thrombosis, infarction of the tissues supplied by the involved vessel, and, in some cases, hemorrhage. As the lesions heal, there is collagen deposition, which may lead to further occlusion of the vessel lumen. Aneurysmal dilations up to 1 cm in size along the involved arteries are characteristic of PAN. Granulomas and substantial eosinophilia with eosinophilic tissue infiltrations are not characteristically found and suggest Churg-Strauss syndrome (see above).

Multiple organ systems are involved, and the clinicopathologic findings reflect the degree and location of vessel involvement and the resulting ischemic changes. As mentioned above, pulmonary arteries are not involved in PAN, and bronchial artery involvement is uncommon. The pathology in the kidney in

classic PAN is that of arteritis without glomerulonephritis. In patients with significant hypertension, typical pathologic features of glomerulosclerosis may be seen. In addition, pathologic sequelae of hypertension may be found elsewhere in the body.

The presence of a PAN-like vasculitis in patients with hepatitis B together with the isolation of circulating immune complexes composed of hepatitis B antigen and immunoglobulin, and the demonstration by immunofluorescence of hepatitis B antigen, IgM, and complement in the blood vessel walls, strongly suggest the role of immunologic phenomena in the pathogenesis of this disease. Hairy cell leukemia can be associated with PAN; the pathogenic mechanisms of this association are unclear.

Clinical and Laboratory Manifestations

Nonspecific signs and symptoms are the hallmarks of PAN. Fever, weight loss, and malaise are present in over one-half of cases. Patients usually present with vague symptoms such as weakness, malaise, headache, abdominal pain, and myalgias that can rapidly progress to a fulminant illness. Specific complaints related to the vascular involvement within a particular organ system may also dominate the presenting clinical picture as well as the entire course of the illness **(Table 326-6)**. In PAN, renal involvement most commonly manifests as hypertension, renal insufficiency, or hemorrhage due to microaneurysms.

Table 326-6 Clinical Manifestations Related to Organ System Involvement in Classic Polyarteritis Nodosa

Organ System	Percent Incidence	Clinical Manifestations
Renal	60	Renal failure, hypertension
Musculoskeletal	64	Arthritis, arthralgia, myalgia
Peripheral nervous system	51	Peripheral neuropathy, mononeuritis multiplex
Gastrointestinal tract	44	Abdominal pain, nausea and vomiting, bleeding, bowel infarction and perforation, cholecystitis, hepatic infarction, pancreatic infarction
Skin	43	Rash, purpura, nodules, cutaneous infarcts, livedo reticularis, Raynaud's phenomenon
Cardiac	36	Congestive heart failure, myocar-dial infarction, pericarditis
Genitourinary	25	Testicular, ovarian, or epididymal pain
Central nervous system	23	Cerebral vascular accident, altered mental status, seizure

Source: From TR Cupps, AS Fauci: *The Vasculitides*. Philadelphia, Saunders, 1981.

There are no diagnostic serologic tests for PAN. In >75% of patients, the leukocyte count is elevated with a predominance of neutrophils. Eosinophilia is seen only rarely and, when present at high levels, suggests the diagnosis of Churg-Strauss syndrome. The anemia of chronic disease may be seen, and an elevated ESR is almost always present. Other common laboratory findings reflect the particular organ involved. Hypergammaglobulinemia may be present, and all patients should be screened for hepatitis B. Antibodies against myeloperoxidase or proteinase-3 (ANCA) are rarely found in patients with PAN.

Diagnosis

The diagnosis of PAN is based on the demonstration of characteristic findings of vasculitis on biopsy material of involved organs. In the absence of easily accessible tissue for biopsy, the arteriographic demonstration of involved vessels, particularly in the form of aneurysms of small and medium-sized arteries in the renal, hepatic, and visceral vasculature, is sufficient to make the diagnosis. Aneurysms of vessels are not pathognomonic of PAN; furthermore, aneurysms need not always be present, and arteriographic findings may be limited to stenotic segments and obliteration of vessels. Biopsy of symptomatic organs such as nodular skin lesions, painful testes, and nerve/muscle provides the highest diagnostic yields.

Treatment: Polyarteritis Nodosa

The prognosis of untreated PAN is extremely poor, with a reported 5-year survival rate between 10 and 20%. Death usually results from gastrointestinal complications, particularly bowel infarcts and perforation, and cardiovascular causes. Intractable hypertension often compounds dysfunction in other organ systems, such as the kidneys, heart, and CNS, leading to additional late morbidity and mortality

in PAN. With the introduction of treatment, survival rate has increased substantially. Favorable therapeutic results have been reported in PAN with the combination of prednisone and cyclophosphamide (see "Granulomatosis With Polyangiitis (Wegener's)" for a detailed description of this therapeutic regimen). In less severe cases of PAN, glucocorticoids alone have resulted in disease remission. In patients with hepatitis B who have a PAN-like vasculitis, antiviral therapy represents an important part of therapy and has been used in combination with glucocorticoids and plasma exchange. Careful attention to the treatment of hypertension can lessen the acute and late morbidity and mortality rates associated with renal, cardiac, and CNS complications of PAN. Following successful treatment, relapse of PAN has been estimated to occur in 10–20% of patients.

GIANT CELL ARTERITIS AND POLYMYALGIA RHEUMATICA

Definition

Giant cell arteritis, also referred to as *cranial arteritis* or *temporal arteritis*, is an inflammation of medium- and large-sized arteries. It characteristically involves one or more branches of the carotid artery, particularly the temporal artery. However, it is a systemic disease that can involve arteries in multiple locations, particularly the aorta and its main branches.

Giant cell arteritis is closely associated with *polymyalgia rheumatica*, which is characterized by stiffness, aching, and pain in the muscles of the neck, shoulders, lower back, hips, and thighs. Most commonly, polymyalgia rheumatica occurs in isolation, but it may be seen in 40–50% of patients with giant cell arteritis. In addition, ~10–20% of patients who initially present with features of isolated polymyalgia rheumatica later go on to develop giant cell arteritis. This strong clinical association together with data from pathophysiologic studies has increasingly supported that giant cell arteritis and polymyalgia rheumatica represent differing clinical spectrums of a single disease process.

Incidence and Prevalence

Giant cell arteritis occurs almost exclusively in individuals >50 years. It is more common in women than in men and is rare in blacks. The incidence of giant cell arteritis varies widely in different studies and in different geographic regions. A high incidence has been found in Scandinavia and in regions of the United States with large Scandinavian populations, compared to a lower incidence in southern Europe. The annual incidence rates in individuals ≥50 years range from 6.9 to 32.8 per 100,000 population. Familial aggregation has been reported, as has an association with HLA-DR4. In addition, genetic linkage studies have demonstrated an association of giant cell arteritis with alleles at the HLA-DRB1 locus, particularly HLA-DRB1* 04 variants. In Olmsted County, Minnesota, the annual incidence of polymyalgia rheumatica in individuals ≥50 years is 58.7 per 100,000 population.

Pathology and Pathogenesis

Although the temporal artery is most frequently involved in giant cell arteritis, patients often have a systemic vasculitis of multiple medium- and large-sized arteries, which may go undetected. Histopathologically, the disease is a panarteritis with inflammatory mononuclear cell infiltrates within the vessel wall with frequent giant cell formation. There is proliferation of the intima and fragmentation of the internal elastic lamina. Pathophysiologic findings in organs result from the ischemia related to the involved vessels.

Experimental data support that giant cell arteritis is an antigen-driven disease in which activated T lymphocytes, macrophages, and dendritic cells play a critical role in the disease pathogenesis. Sequence analysis of the T cell receptor of tissue-infiltrating T cells in lesions of giant cell arteritis indicates restricted clonal expansion, suggesting the presence of an antigen residing in the arterial wall. Giant cell arteritis is believed to be initiated in the adventitia where CD4+ T cells enter through the vasa vasorum, become activated, and orchestrate macrophage differentiation. T cells recruited to vasculitic lesions in patients with giant cell arteritis produce predominantly IL-2 and IFN-γ, and the latter has been suggested to be involved in the progression to overt arteritis.

Clinical and Laboratory Manifestations

Giant cell arteritis is most commonly characterized clinically by the complex of fever, anemia, high ESR, and headaches in a patient over the age of 50 years. Other phenotypic manifestations include features of systemic inflammation including malaise, fatigue, anorexia, weight loss, sweats, arthralgias, polymyalgia rheumatica, or large-vessel disease.

In patients with involvement of the cranial arteries, headache is the predominant symptom and may be associated with a tender, thickened, or nodular artery, which may pulsate early in the disease but may become occluded later. Scalp pain and claudication of the jaw and tongue may occur. A well-recognized and dreaded complication of giant cell arteritis, particularly in untreated patients, is ischemic optic neuropathy, which may lead to serious visual symptoms, even sudden blindness in some patients. However, most patients have complaints relating to the head or eyes before visual loss. Attention to such symptoms with institution of appropriate therapy (see below) will usually avoid this complication. Other cranial ischemic complications include strokes, scalp or tongue infarction.

Up to one-third of patients can have large-vessel disease that can be the primary presentation of giant cell arteritis or can emerge at a later point in patients who have had previous cranial arteritis features or polymyalgia rheumatica. Manifestations of large-vessel disease can include subclavian artery stenosis that can present as arm claudication or aortic aneurysms involving the thoracic and to a lesser degree the abdominal aorta, which carry risks of rupture or dissection.

Characteristic laboratory findings in addition to the elevated ESR include a normochromic or slightly hypochromic anemia. Liver function abnormalities are common, particularly increased alkaline phosphatase levels. Increased levels of IgG and complement have been reported. Levels of enzymes indicative of muscle damage such as serum creatine kinase are not elevated.

Diagnosis

The diagnosis of giant cell arteritis and its associated clinicopathologic syndrome can often be suggested clinically by the demonstration of the complex of fever, anemia, and high ESR with or without symptoms of polymyalgia rheumatica in a patient >50 years. The diagnosis is confirmed by biopsy of the temporal artery. Since involvement of the vessel may be segmental, positive yield is increased by obtaining a biopsy segment of 3–5 cm together with serial sectioning of biopsy specimens. Ultrasonography of the temporal artery has been reported to be helpful in diagnosis. A temporal artery biopsy should be obtained as quickly as possible in the setting of ocular signs and symptoms, and under these circumstances therapy should not be delayed pending a biopsy. In this regard, it has been reported that temporal artery biopsies may show vasculitis even after ~14 days of glucocorticoid therapy. A dramatic clinical response to a trial of glucocorticoid therapy can further support the diagnosis.

Large vessel disease may be suggested by symptoms and findings on physical examination such as diminished pulses or bruits. It is confirmed by vascular imaging, most commonly through magnetic resonance or computed tomography.

Isolated polymyalgia rheumatica is a clinical diagnosis made by the presence of typical symptoms of stiffness, aching, and pain in the muscles of the hip and shoulder girdle, an increased ESR, the absence of clinical features suggestive of giant cell arteritis, and a prompt therapeutic response to low-dose prednisone.

Treatment: Giant Cell Arteritis and Polymyalgia Rheumatica

Acute disease-related mortality directly from giant cell arteritis is very uncommon with fatalities occurring from cerebrovascular events or myocardial infarction. However, patients are at risk of late mortality from aortic aneurysm rupture or dissection as patients with giant cell arteritis are 18 times more likely to develop thoracic aortic aneurysms than the general population.

The goals of treatment in giant cell arteritis are to reduce symptoms and, most importantly, to prevent visual loss. The treatment approach for cranial and large-vessel disease in giant cell arteritis is currently the same. Giant cell arteritis and its associated symptoms are exquisitely sensitive to glucocorticoid therapy. Treatment should begin with prednisone, 40–60 mg/d for ~1 month, followed by a gradual tapering. When ocular signs and symptoms occur, consideration should be given for the use of methylprednisolone 1000 mg daily for 3 days to protect remaining vision. Although the optimal duration of glucocorticoid therapy has not been established, most series have found that patients require treatment for ≥2 years. Symptom recurrence during prednisone tapering develops in 60–85% of patients with giant cell arteritis, requiring a dosage increase. The ESR can serve as a useful indicator of inflammatory disease activity in monitoring and tapering therapy and can be used to judge the pace of the tapering schedule. However, minor increases in the ESR can occur as glucocorticoids are being tapered and do not necessarily reflect an exacerbation of arteritis, particularly if the patient remains symptom-free. Under these circumstances, the tapering should continue with caution. Glucocorticoid toxicity occurs in 35–65% of patients and represents an important cause of patient morbidity. Aspirin 81 mg daily has been found to reduce the occurrence of cranial ischemic complications in giant cell arteritis and should be given in addition to glucocorticoids in patients who do not have contraindications. The use of weekly methotrexate as a glucocorticoid-sparing agent has been examined in two randomized placebo-controlled trials that reached conflicting conclusions. Infliximab, a monoclonal antibody to TNF, was studied in a randomized trial and was not found to provide benefit.

Patients with isolated polymyalgia rheumatica respond promptly to prednisone, which can be started at a lower dose of 10–20 mg/d. Similar to giant cell arteritis, the ESR can serve as a useful indicator in monitoring and prednisone reduction. Recurrent polymyalgia symptoms develop in the majority of patients during prednisone tapering. One study of weekly methotrexate found that the use of this drug reduced the prednisone dose on average by only 1 mg and did not decrease prednisone-related side effects. A randomized trial in polymyalgia rheumatica did not find infliximab to lessen relapse or glucocorticoid requirements.

DEFINITION

Takayasu's arteritis is an inflammatory and stenotic disease of medium- and large-sized arteries characterized by a strong predilection for the aortic arch and its branches. For this reason, it is often

referred to as the *aortic arch syndrome*.

Incidence and Prevalence

Takayasu's arteritis is an uncommon disease with an estimated annual incidence rate of 1.2–2.6 cases per million. It is most prevalent in adolescent girls and young women. Although it is more common in Asia, it is neither racially nor geographically restricted.

Pathology and Pathogenesis

The disease involves medium- and large-sized arteries, with a strong predilection for the aortic arch and its branches; the pulmonary artery may also be involved. The most commonly affected arteries seen by arteriography are listed in **Table 326-7**. The involvement of the major branches of the aorta is much more marked at their origin than distally. The disease is a panarteritis with inflammatory mononuclear cell infiltrates and occasionally giant cells. There are marked intimal proliferation and fibrosis, scarring and vascularization of the media, and disruption and degeneration of the elastic lamina. Narrowing of the lumen occurs with or without thrombosis. The vasa vasorum are frequently involved. Pathologic changes in various organs reflect the compromise of blood flow through the involved vessels.

Table 326-7 Frequency of Arteriographic Abnormalities and Potential Clinical Manifestations of Arterial Involvement in Takayasu's Arteritis

Artery	Percent of Arteriographic Abnormalities	Potential Clinical Manifestations
Subclavian	93	Arm claudication, Raynaud's phenomenon
Common carotid	58	Visual changes, syncope, transient ischemic attacks, stroke
Abdominal aorta[a]	47	Abdominal pain, nausea, vomiting
Renal	38	Hypertension, renal failure
Aortic arch or root	35	Aortic insufficiency, congestive heart failure
Vertebral	35	Visual changes, dizziness
Coeliac axis[a]	18	Abdominal pain, nausea, vomiting
Superior mesenteric[a]	18	Abdominal pain, nausea, vomiting
Iliac	17	Leg claudication
Pulmonary	10–40	Atypical chest pain, dyspnea
Coronary	<10	Chest pain, myocardial infarction

[a]Arteriographic lesions at these locations are usually asymptomatic but may potentially cause these symptoms.

Source: Kerr et al.

Immunopathogenic mechanisms, the precise nature of which is uncertain, are suspected in this disease. As with several of the vasculitis syndromes, circulating immune complexes have been demonstrated, but their pathogenic significance is unclear.

Clinical and Laboratory Manifestations

Takayasu's arteritis is a systemic disease with generalized as well as vascular symptoms. The generalized symptoms include malaise, fever, night sweats, arthralgias, anorexia, and weight loss, which may occur months before vessel involvement is apparent. These symptoms may merge into those related to vascular compromise and organ ischemia. Pulses are commonly absent in the involved vessels, particularly the subclavian artery. The frequency of arteriographic abnormalities and the potentially associated clinical manifestations are listed in Table 326-7. Hypertension occurs in 32–93% of patients and contributes to renal, cardiac, and cerebral injury.

Characteristic laboratory findings include an elevated ESR, mild anemia, and elevated immunoglobulin levels.

Diagnosis

The diagnosis of Takayasu's arteritis should be suspected strongly in a young woman who develops a decrease or absence of peripheral pulses, discrepancies in blood pressure, and arterial bruits. The diagnosis is confirmed by the characteristic pattern on arteriography, which includes irregular vessel walls, stenosis, poststenotic dilation, aneurysm formation, occlusion, and evidence of increased collateral circulation. Complete aortic arteriography by catheter-directed dye arteriography or magnetic resonance arteriography should be obtained in order to fully delineate the distribution and degree of arterial disease. Histopathologic demonstration of inflamed vessels adds confirmatory data; however, tissue is rarely readily available for examination.

Treatment: Takayasu's Arteritis

The long-term outcome of patients with Takayasu's arteritis has varied widely between studies. Although two North American reports found overall survival to be >94%, the 5-year mortality rate from other studies has ranged from 0 to 35%. Disease-related mortality most often occurs from congestive heart failure, cerebrovascular events, myocardial infarction, aneurysm rupture, or renal failure. Even in the absence of life-threatening disease, Takayasu's arteritis can be associated with significant morbidity. The course of the disease is variable, and although spontaneous remissions may occur, Takayasu's arteritis is most often chronic and relapsing. Although glucocorticoid therapy in doses of 40–60 mg prednisone per day alleviates symptoms, there are no convincing studies that indicate that they increase survival. The combination of glucocorticoid therapy for acute signs and symptoms and an aggressive surgical and/or arterioplastic approach to stenosed vessels has markedly improved outcome and decreased morbidity by lessening the risk of stroke, correcting hypertension due to renal artery stenosis, and improving blood flow to ischemic viscera and limbs. Unless it is urgently required, surgical correction of stenosed arteries should be undertaken only when the vascular inflammatory process is well controlled with medical therapy. In individuals who are refractory to or unable to taper glucocorticoids, methotrexate in doses up to 25 mg per week has yielded encouraging results. Preliminary results with anti-TNF therapies have been encouraging, but will require further study through randomized trials to determine efficacy.

HENOCH-SCHÖNLEIN PURPURA

Definition

Henoch-Schönlein purpura, also referred to as *anaphylactoid purpura*, is a small-vessel vasculitis characterized by palpable purpura (most commonly distributed over the buttocks and lower extremities), arthralgias, gastrointestinal signs and symptoms, and glomerulonephritis.

Incidence and Prevalence

Henoch-Schönlein purpura is usually seen in children; most patients range in age from 4 to 7 years; however, the disease may also be seen in infants and adults. It is not a rare disease; in one series it accounted for between 5 and 24 admissions per year at a pediatric hospital. The male-to-female ratio is 1.5:1. A seasonal variation with a peak incidence in spring has been noted.

Pathology and Pathogenesis

The presumptive pathogenic mechanism for Henoch-Schönlein purpura is immune-complex deposition. A number of inciting antigens have been suggested including upper respiratory tract infections, various drugs, foods, insect bites, and immunizations. IgA is the antibody class most often seen in the immune complexes and has been demonstrated in the renal biopsies of these patients.

Clinical and Laboratory Manifestations

In pediatric patients, palpable purpura is seen in virtually all patients; most patients develop polyarthralgias in the absence of frank arthritis. Gastrointestinal involvement, which is seen in almost 70% of pediatric patients, is characterized by colicky abdominal pain usually associated with nausea, vomiting, diarrhea, or constipation and is frequently accompanied by the passage of blood and mucus per rectum; bowel intussusception may occur. Renal involvement occurs in 10–50% of patients and is usually characterized by mild glomerulonephritis leading to proteinuria and microscopic hematuria, with red blood cell casts in the majority of patients **(Chap. 283)**; it usually resolves spontaneously without therapy. Rarely, a progressive glomerulonephritis will develop. In adults, presenting symptoms are most frequently related to the skin and joints, while initial complaints related to the gut are less common. Although certain studies have found that renal disease is more frequent and more severe in adults, this has not been a consistent finding. However, the course of renal disease in adults may be more insidious and thus requires close follow-up. Myocardial involvement can occur in adults but is rare in children.

Laboratory studies generally show a mild leukocytosis, a normal platelet count, and occasionally eosinophilia. Serum complement components are normal, and IgA levels are elevated in about one-half of patients.

Diagnosis

The diagnosis of Henoch-Schönlein purpura is based on clinical signs and symptoms. Skin biopsy specimen can be useful in confirming leukocytoclastic vasculitis with IgA and C3 deposition by immunofluorescence. Renal biopsy is rarely needed for diagnosis but may provide prognostic information in some patients.

Treatment: Henoch-Schönlein Purpura

The prognosis of Henoch-Schönlein purpura is excellent. Mortality is exceedingly rare, and 1–5% of children progress to end-stage renal disease. Most patients recover completely, and some do not require therapy. Treatment is similar for adults and children. When glucocorticoid therapy is required, prednisone, in doses of 1 mg/kg per day and tapered according to clinical response, has been shown to be useful in decreasing tissue edema, arthralgias, and abdominal discomfort; however, it has not proved beneficial in the treatment of skin or renal disease and does not appear to shorten the duration of active disease or lessen the chance of recurrence. Patients with rapidly progressive glomerulonephritis have been anecdotally reported to benefit from intensive plasma exchange combined with cytotoxic drugs. Disease recurrences have been reported in 10–40% of patients.

IDIOPATHIC CUTANEOUS VASCULITIS

Definition

The term *cutaneousvasculitis* is defined broadly as inflammation of the blood vessels of the dermis. Due to its heterogeneity, cutaneous vasculitis has been described by a variety of terms including *hypersensitivity vasculitis* and *cutaneous leukocytoclastic angiitis*. However, cutaneous vasculitis is not one specific disease but a manifestation that can be seen in a variety of settings. In >70% of cases, cutaneous vasculitis occurs either as part of a primary systemic vasculitis or as a secondary vasculitis related to an inciting agent or an underlying disease (see "Secondary Vasculitis," below). In the remaining 30% of cases, cutaneous vasculitis occurs idiopathically.

Incidence and Prevalence

Cutaneousvasculitis represents the most commonly encountered vasculitis in clinical practice. The exact incidence of idiopathic cutaneous vasculitis has not been determined due to the predilection for cutaneous vasculitis to be associated with an underlying process and the variability of its clinical course.

Pathology and Pathogenesis

The typical histopathologic feature of cutaneousvasculitis is the presence of vasculitis of small vessels. Postcapillary venules are the most commonly involved vessels; capillaries and arterioles may be involved less frequently. This vasculitis is characterized by a *leukocytoclasis*, a term that refers to the nuclear debris remaining from the neutrophils that have infiltrated in and around the vessels during the acute stages. In the subacute or chronic stages, mononuclear cells predominate; in certain subgroups, eosinophilic infiltration is seen. Erythrocytes often extravasate from the involved vessels, leading to palpable purpura.

Clinical and Laboratory Manifestations

The hallmark of idiopathiccutaneousvasculitis is the predominance of skin involvement. Skin lesions may appear typically as palpable purpura; however, other cutaneous manifestations of the vasculitis may occur, including macules, papules, vesicles, bullae, subcutaneous nodules, ulcers, and recurrent or chronic urticaria. The skin lesions may be pruritic or even quite painful, with a burning or stinging sensation. Lesions most commonly occur in the lower extremities in ambulatory patients or in the sacral area in bedridden patients due to the effects of hydrostatic forces on the postcapillary venules. Edema may accompany certain lesions, and hyperpigmentation often occurs in areas of recurrent or chronic lesions.

There are no specific laboratory tests diagnostic of idiopathiccutaneousvasculitis. A mild leukocytosis with or without eosinophilia is characteristic, as is an elevated ESR. Laboratory studies should be aimed toward ruling out features to suggest an underlying disease or a systemic vasculitis.

Diagnosis

The diagnosis of cutaneousvasculitis is made by the demonstration of vasculitis on biopsy. An important diagnostic principle in patients with cutaneous vasculitis is to search for an etiology of the vasculitis—be it an exogenous agent, such as a drug or an infection, or an endogenous condition, such as an underlying disease (Fig. 326-1). In addition, a careful physical and laboratory examination should be performed to rule out the possibility of systemic vasculitis. This should start with the least invasive diagnostic approach and proceed to the more invasive only if clinically indicated.

Treatment: Idiopathic Cutaneous Vasculitis

When an antigenic stimulus is recognized as the precipitating factor in the cutaneousvasculitis, it should be removed; if this is a microbe, appropriate antimicrobial therapy should be instituted. If the vasculitis is associated with another underlying disease, treatment of the latter often results in

resolution of the former. In situations where disease is apparently self-limited, no therapy, except possibly symptomatic therapy, is indicated. When cutaneous vasculitis persists and when there is no evidence of an inciting agent, an associated disease, or an underlying systemic vasculitis, the decision to treat should be based on weighing the balance between the degree of symptoms and the risk of treatment. Some cases of idiopathic cutaneous vasculitis resolve spontaneously, while others remit and relapse. In those patients with persistent vasculitis, a variety of therapeutic regimens have been tried with variable results. In general, the treatment of idiopathic cutaneous vasculitis has not been satisfactory. Fortunately, since the disease is generally limited to the skin, this lack of consistent response to therapy usually does not lead to a life-threatening situation. Agents with which there have been anecdotal reports of success include dapsone, colchicine, hydroxychloroquine, and nonsteroidal anti-inflammatory agents. Glucocorticoids are often used in the treatment of idiopathic cutaneous vasculitis. Therapy is usually instituted as prednisone, 1 mg/kg per day, with rapid tapering where possible, either directly to discontinuation or by conversion to an alternate-day regimen followed by ultimate discontinuation. In cases that prove refractory to glucocorticoids, a trial of a cytotoxic agent may be indicated. Patients with chronic vasculitis isolated to cutaneous venules rarely respond dramatically to any therapeutic regimen, and cytotoxic agents should be used only as a last resort in these patients. Methotrexate and azathioprine have been used in such situations in anecdotal reports. Although cyclophosphamide is the most effective therapy for the systemic vasculitides, it should almost never be used for idiopathic cutaneous vasculitis because of the potential toxicity.

CRYOGLOBULINEMIC VASCULITIS

Definition

Cryoglobulins are cold-precipitable monoclonal or polyclonal immunoglobulins. Cryoglobulinemia may be associated with a systemic vasculitis characterized by palpable purpura, arthralgias, weakness, neuropathy, and glomerulonephritis. Although this can be observed in association with a variety of underlying disorders including multiple myeloma, lymphoproliferative disorders, connective tissue diseases, infection, and liver disease, in many instances it appeared to be idiopathic. Because of the apparent absence of an underlying disease and the presence of cryoprecipitate containing oligoclonal/polyclonal immunoglobulins, this entity was referred to as *essential mixed cryoglobulinemia*. Since the discovery of hepatitis C, it has been established that the vast majority of patients who were considered to have essential mixed cryoglobulinemia have cryoglobulinemic vasculitis related to hepatitis C infection.

Incidence and Prevalence

The incidence of cryoglobulinemic vasculitis has not been established. It has been estimated, however, that 5% of patients with chronic hepatitis C will develop cryoglobulinemic vasculitis.

Pathology and Pathogenesis

Skin biopsies in cryoglobulinemic vasculitis reveal an inflammatory infiltrate surrounding and involving blood vessel walls, with fibrinoid necrosis, endothelial cell hyperplasia, and hemorrhage. Deposition of immunoglobulin and complement is common. Abnormalities of uninvolved skin including basement membrane alterations and deposits in vessel walls may be found. Membranoproliferative glomerulonephritis is responsible for 80% of all renal lesions in cryoglobulinemic vasculitis.

The association between hepatitis C and cryoglobulinemic vasculitis has been supported by the high frequency of documented hepatitis C infection, the presence of hepatitis C RNA and anti–hepatitis C antibodies in serum cryoprecipitates, evidence of hepatitis C antigens in vasculitic skin lesions, and the effectiveness of antiviral therapy (see below). Current evidence suggests that in the majority of cases, cryoglobulinemic vasculitis occurs when an aberrant immune response to hepatitis C infection leads to the formation of immune complexes consisting of hepatitis C antigens, polyclonal hepatitis C –specific IgG, and monoclonal IgM rheumatoid factor. The deposition of these immune complexes in blood vessel walls triggers an inflammatory cascade that results in cryoglobulinemic vasculitis.

Clinical and Laboratory Manifestations

The most common clinical manifestations of cryoglobulinemic vasculitis are cutaneous vasculitis, arthritis, peripheral neuropathy, and glomerulonephritis. Renal disease develops in 10–30% of patients. Life-threatening rapidly progressive glomerulonephritis or vasculitis of the CNS, gastrointestinal tract, or heart occurs infrequently.

The presence of circulating cryoprecipitates is the fundamental finding in cryoglobulinemic vasculitis. Rheumatoid factor is almost always found and may be a useful clue to the disease when cryoglobulins are not detected. Hypocomplementemia occurs in 90% of patients. An elevated ESR and anemia occur frequently. Evidence for hepatitis C infection must be sought in all patients by testing for hepatitis C antibodies and hepatitis C RNA.

Treatment: Cryoglobulinemic Vasculitis

Acute mortality directly from cryoglobulinemic vasculitis is uncommon, but the presence of glomerulonephritis is a poor prognostic sign for overall outcome. In such patients, 15% progress to

end-stage renal disease, with 40% later experiencing fatal cardiovascular disease, infection, or liver failure. As indicated above, the majority of cases are associated with hepatitis C infection. In such patients, treatment with IFN-α and ribavirin **(Chap. 304)** can prove beneficial. Clinical improvement with antiviral therapy is dependent on the virologic response. Patients who clear hepatitis C from the blood have objective improvement in their vasculitis along with significant reductions in levels of circulating cryoglobulins, IgM, and rheumatoid factor. However, substantial portions of patients with hepatitis C do not have a sustained virologic response to such therapy, and the vasculitis typically relapses with the return of viremia. While transient improvement can be observed with glucocorticoids, a complete response is seen in only 7% of patients. Plasmapheresis and cytotoxic agents have been used in anecdotal reports. These observations have not been confirmed, and such therapies carry significant risks.

BEHÇET'S SYNDROME

Behçet's syndrome is a clinicopathologic entity characterized by recurrent episodes of oral and genital ulcers, iritis, and cutaneous lesions. The underlying pathologic process is a leukocytoclastic venulitis, although vessels of any size and in any organ can be involved. **This disorder is described in detail in Chap. 327.**

ISOLATED VASCULITIS OF THE CENTRAL NERVOUS SYSTEM

Isolated vasculitis of the CNS , which is also called *primary angiitis of the CNS* (PACNS), is an uncommon clinicopathologic entity characterized by vasculitis restricted to the vessels of the CNS without other apparent systemic vasculitis. The inflammatory process is usually composed of mononuclear cell infiltrates with or without granuloma formation.

Patients may present with headaches, altered mental function, and focal neurologic defects. Systemic symptoms are generally absent. Devastating neurologic abnormalities may occur depending on the extent of vessel involvement. The diagnosis can be suggested by abnormal MRI of the brain, an abnormal lumbar puncture, and/or demonstration of characteristic vessel abnormalities on arteriography **(Fig. 326-4)**, but it is confirmed by biopsy of the brain parenchyma and leptomeninges. In the absence of a brain biopsy, care should be taken not to misinterpret as true primary vasculitis arteriographic abnormalities that might actually be related to another cause. An important entity in the differential diagnosis is reversible cerebral vasoconstrictive syndrome, which typically presents with "thunderclap" headache and is associated with arteriographic abnormalities that mimic PACNS that are reversible. Other diagnostic considerations include infection, atherosclerosis, emboli, connective tissue disease, sarcoidosis, malignancy, and drug-associated causes. The prognosis of granulomatous PACNS is poor; however, some reports indicate that glucocorticoid therapy, alone or together with cyclophosphamide administered as described above, has induced clinical remissions.

Figure 326-4

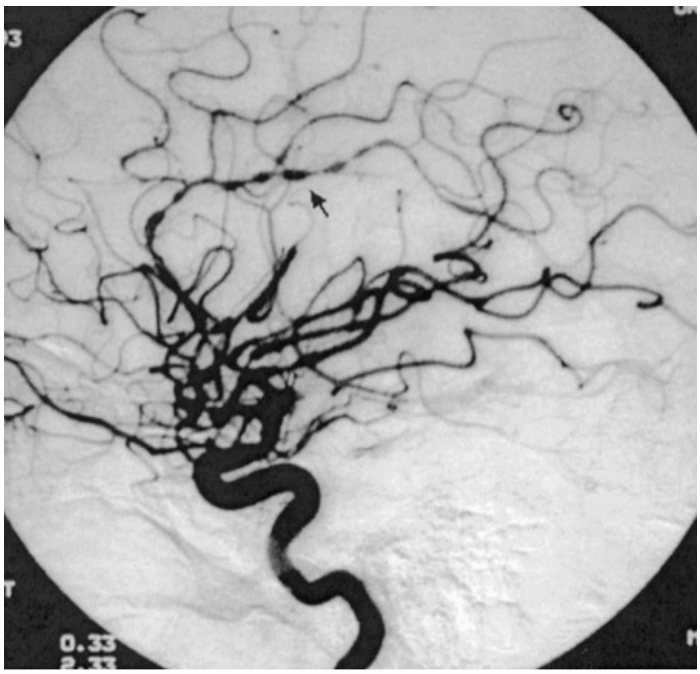

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cerebral arteriogram from a 32-year-old male with central nervous system vasculitis.

Dramatic beading (*arrow*) typical of vasculitis is seen.

COGAN'S SYNDROME

Cogan's syndrome is characterized by interstitial keratitis together with vestibuloauditory symptoms. It may be associated with a systemic vasculitis, particularly aortitis with involvement of the aortic valve. Glucocorticoids are the mainstay of treatment. Initiation of treatment as early as possible after the onset of hearing loss improves the likelihood of a favorable outcome.

KAWASAKI DISEASE

Kawasaki disease, also referred to as *mucocutaneous lymph node syndrome*, is an acute, febrile, multisystem disease of children. Some 80% of cases occur prior to the age of 5, with the peak incidence occurring at ≤2 years. It is characterized by nonsuppurative cervical adenitis and changes in the skin and mucous membranes such as edema; congested conjunctivae; erythema of the oral cavity, lips, and palms; and desquamation of the skin of the fingertips. Although the disease is generally benign and self-limited, it is associated with coronary artery aneurysms in ~25% of cases, with an overall case-fatality rate of 0.5–2.8%. These complications usually occur between the third and fourth weeks of illness during the convalescent stage. Vasculitis of the coronary arteries is seen in almost all the fatal cases that have been autopsied. There is typical intimal proliferation and infiltration of the vessel wall with mononuclear cells. Beadlike aneurysms and thromboses may be seen along the artery. Other manifestations include pericarditis, myocarditis, myocardial ischemia and infarction, and cardiomegaly.

Apart from the up to 2.8% of patients who develop fatal complications, the prognosis of this disease for uneventful recovery is excellent. High-dose IV γ globulin (2 g/kg as a single infusion over 10 h) together with aspirin (100 mg/kg per day for 14 days followed by 3–5 mg/kg per day for several weeks) have been shown to be effective in reducing the prevalence of coronary artery abnormalities when administered early in the course of the disease. Surgery may be necessary for Kawasaki disease patients that have giant coronary artery aneurysms or other coronary complications. Surgical treatment most commonly includes thromboendarterectomy, thrombus clearing, aneurysmal reconstruction, and coronary artery bypass grafting.

POLYANGIITIS OVERLAP SYNDROMES

Some patients with systemic vasculitis manifest clinicopathologic characteristics that do not fit precisely into any specific disease but have overlapping features of different vasculitides. Active systemic vasculitis in such settings has the same potential for causing irreversible organ system damage as when it occurs in one of the defined syndromes listed in Table 326-1. The diagnostic and therapeutic considerations as well as the prognosis for these patients depend on the sites and severity of active vasculitis. Patients with vasculitis that could potentially cause irreversible damage to a major organ system should be treated as described under "Granulomatosis With Polyangiitis (Wegener's)."

SECONDARY VASCULITIS

Drug-Induced Vasculitis

Vasculitis associated with drug reactions usually presents as palpable purpura that may be generalized or limited to the lower extremities or other dependent areas; however, urticarial lesions, ulcers, and hemorrhagic blisters may also occur **(Chap. 55)**. Signs and symptoms may be limited to the skin, although systemic manifestations such as fever, malaise, and polyarthralgias may occur. Although the skin is the predominant organ involved, systemic vasculitis may result from drug reactions. Drugs that have been implicated in vasculitis include allopurinol, thiazides, gold, sulfonamides, phenytoin, and penicillin **(Chap. 55)**.

An increasing number of drugs have been reported to cause vasculitis associated with antimyeloperoxidase ANCA. Of these, the best evidence of causality exists for hydralazine and propylthiouracil. The clinical manifestations in ANCA-positive drug-induced vasculitis can range from cutaneous lesions to glomerulonephritis and pulmonary hemorrhage. Outside of drug discontinuation, treatment should be based on the severity of the vasculitis. Patients with immediately life-threatening small-vessel vasculitis should initially be treated with glucocorticoids and cyclophosphamide as described for granulomatosis with polyangiitis (Wegener's). Following clinical improvement, consideration may be given for tapering such agents along a more rapid schedule.

Serum Sickness and Serum Sickness–Like Reactions

These reactions are characterized by the occurrence of fever, urticaria, polyarthralgias, and lymphadenopathy 7–10 days after primary exposure and 2–4 days after secondary exposure to a heterologous protein (classic serum sickness) or a nonprotein drug such as penicillin or sulfa (serum sickness–like reaction). Most of the manifestations are not due to a vasculitis; however, occasional patients will have typical cutaneous venulitis that may progress rarely to a systemic vasculitis.

Vasculitis Associated with Other Underlying Diseases

Certain *infections* may directly trigger an inflammatory vasculitic process. For example, rickettsias can invade and proliferate in the endothelial cells of small blood vessels causing a vasculitis **(Chap. 174)**. In addition, the inflammatory response around blood vessels associated with certain systemic fungal diseases such as histoplasmosis **(Chap. 199)** may mimic a primary vasculitic process. A leukocytoclastic vasculitis predominantly involving the skin with occasional involvement of other organ systems may be a minor component of many other infections. These include *subacute bacterial endocarditis*, *Epstein-Barr virus infection*, *HIV infection*, as well as a number of other infections.

Vasculitis can be associated with certain *malignancies*, particularly lymphoid or reticuloendothelial neoplasms. Leukocytoclastic venulitis confined to the skin is the most common finding; however, widespread systemic vasculitis may occur. Of particular note is the association of *hairy cell leukemia***(Chap. 110)** with PAN.

A number of *connective tissue diseases* have vasculitis as a secondary manifestation of the underlying primary process. Foremost among these are *systemic lupus erythematosus***(Chap. 319)**, *rheumatoid arthritis***(Chap. 321)**, *inflammatory myositis***(Chap. 388)**, *relapsing polychondritis***(Chap. 328)**, and *Sjögren's syndrome***(Chap. 324)**. The most common form of vasculitis in these conditions is the small-vessel venulitis isolated to the skin. However, certain patients may develop a fulminant systemic necrotizing vasculitis.

Secondary vasculitis has also been observed in association with *ulcerative colitis*, *congenital deficiencies of various complement components*, *retroperitoneal fibrosis*, *primary biliary cirrhosis*, α_1*-antitrypsin deficiency*, and *intestinal bypass surgery*.

FURTHER READINGS

de Groot K et al: Randomized trial of cyclophosphamide versus methotrexate for induction of remission in early systemic antineutrophil cytoplasmic antibody-associated vasculitis. Arthritis Rheum 52:2461, 2005

Finkielman JD et al: Antiproteinase 3 antineutrophil cytoplasmic antibodies and disease activity in Wegener granulomatosis. Ann Intern Med 147:611, 2007[PMID: 17975183] [Full Text]

Guillevin L et al: Microscopic polyangiitis. Clinical and laboratory findings in eighty-five patients. Arthritis Rheum 42:421, 1999[PMID: 10088763] [Full Text]

Hoffman GS, Specks U: Antineutrophil cytoplasmic antibodies. Arthritis Rheum 41:1521, 1998[PMID: 9751084] [Full Text]

——— et al: Wegener's granulomatosis: An analysis of 158 patients. Ann Intern Med 116:488, 1992

Jayne D et al: A randomized trial of maintenance therapy for vasculitis associated with antineutrophil cytoplasmic autoantibodies. N Engl J Med 349:36, 2003[PMID: 12840090] [Full Text]

Jennette JC et al: Nomenclature of systemic vasculitis. Proposal of an international consensus conference. Arthritis Rheum 37:187, 1994[PMID: 8129773] [Full Text]

Kerr G et al: Takayasu arteritis. Ann Intern Med 120:919, 1994[PMID: 7909656] [Full Text]

Langford CA et al: Use of a cyclophosphamide-induction methotrexate-maintenance regimen for the treatment of Wegener's granulomatosis: Extended follow-up and rate of relapse. Am J Med 114:463, 2003[PMID: 12727579] [Full Text]

Ludviksson BR et al: Active Wegener's granulomatosis is associated with HLA-DR+ CD4+ T cells exhibiting an unbalanced Th-1 type T cell cytokine pattern: Reversal with IL-10. J Immunol 160:3602, 1998[PMID: 9531324] [Full Text]

Sneller MC, Fauci AS: Pathogenesis of vasculitis syndromes. Med Clin North Am 81:221, 1997[PMID: 9012762] [Full Text]

Stone JH et al: Rituximab versus cyclophosphamide for ANCA-associated vasculitis. N Engl J Med 363:221, 2010[PMID: 20647199] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF THE VASCULITIC SYNDROMES: INTRODUCTION

Diagnosis of the vasculitic syndromes is usually based upon characteristic histologic or arteriographic findings in a patient who has clinically compatible features. The images provided in this atlas highlight some of the characteristic histologic and radiographic findings that may be seen in the vasculitic diseases. These images demonstrate the importance that tissue histology may have in securing the diagnosis of vasculitis, the utility of diagnostic imaging in the vasculitic diseases, and the improvements in the care of vasculitis patients that have resulted from radiologic innovations.

Tissue biopsies represent vital information in many patients with a suspected vasculitic syndrome, not only in confirming the presence of vasculitis and other characteristic histologic features, but also in ruling out other diseases that can have similar clinical presentations. The determination of where biopsies should be performed is based upon the presence of clinical disease in an affected organ, the likelihood of a positive diagnostic yield from data contained in the published literature, and the risk of performing a biopsy in an affected site. Common sites where biopsies may be performed include the lung, kidney, and skin. Other sites such as sural nerve, brain, testicle, and gastrointestinal tissues may also demonstrate features of vasculitis and be appropriate locations for biopsy when clinically affected.

Surgical biopsies of radiographically abnormal pulmonary parenchyma, have a diagnostic yield of 90% in patients with granulomatosis with polyangiitis (Wegener's), and play an important role in ruling out infection or malignancy. The yield of lung biopsies is highly associated with amount of tissue that can be obtained, and transbronchial biopsies, while less invasive, have a yield of only 7%. Lung biopsies also play an important role in microscopic polyangiitis, Churg-Strauss syndrome, and in any vasculitic disease where an immunosuppressed patient has pulmonary disease that is suspected to be an infection.

Kidney biopsy findings of a focal, segmental, crescentic, necrotizing glomerulonephritis with few to no immune complexes (pauci-immune glomerulonephritis) are characteristic in patients with granulomatosis with polyangiitis (Wegener's), microscopic polyangiitis, or Churg-Strauss syndrome, who have active renal disease. These findings not only distinguish these entities from other causes of glomerulonephritis, they can confirm the presence of active glomerulonephritis that requires treatment. Because of this, renal biopsies can also be helpful to guide management decisions in these diseases when an established patient has worsening renal function and an inactive or equivocal urine sediment. Cryoglobulinemic vasculitis and Henoch-Schönlein purpura are other vasculitides where renal involvement may occur and where biopsy may be important in diagnosis or prognosis.

Biopsies of the skin are commonly performed and are well tolerated. As not all purpuric or ulcerative lesions are due to vasculitis, skin biopsy plays an important role to confirm the presence of vasculitis as the cause of the manifestation. Cutaneous vasculitis represents the most common vasculitic feature that affects people and can be seen in a broad spectrum of settings including infections, medications, malignancies, and connective tissue diseases. Because of this, for systemic vasculitides that will require aggressive immunosuppressive treatment, a skin biopsy may not represent sufficient evidence to secure the diagnosis.

Diagnostic imaging represents a critical assessment tool in patients who are known or suspected to have a systemic vasculitic disease. Imaging contributes unique information about the patient that, when taken together with the history, physical examination, and laboratory determinations, can guide the differential diagnosis and the subsequent assessment or treatment plan. A diverse range of imaging techniques are utilized in the assessment of vasculitis including plain radiography, ultrasonography, CT, MRI, positron emission tomography, and catheter-directed dye arteriography. These procedures have specific utilities that can allow differing perspectives on the spectrum and severity of vasculitis.

For vasculitic diseases that involve the large- or medium-sized blood vessels, arteriography provides information regarding blood vessel stenoses or aneurysms that can support the diagnosis. Catheter-directed dye arteriography offers the most precise information regarding the vessel lumen but carries risks related to dye exposure and the invasive nature of the procedure. Advancements in MR and CT arteriography have brought about noninvasive options to view the lumen and vessel wall, thus enhancing the ability to perform serial studies for patient monitoring.

Although vasculitis involving the small blood vessels cannot be directly visualized, diagnostic imaging plays an essential role in detecting tissue injury that occurs as result of blood vessel and tissue

inflammation. In granulomatosis with polyangiitis (Wegener's), 80% of patients may have pulmonary involvement during their disease course. Chest imaging should be obtained whenever active disease is suspected, as up to one-third of patients with radiographic abnormalities are asymptomatic. Pulmonary imaging is also important to detect complications of vasculitis therapy such as opportunistic pneumonias as well as medication-related pneumonitis.

Figure e40-1

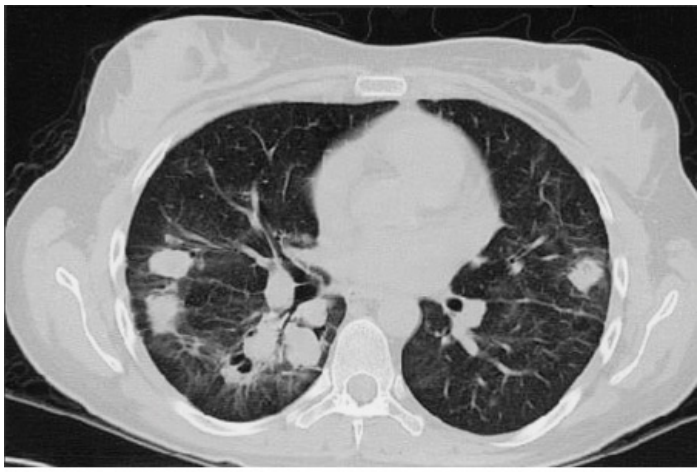

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bilateral nodular infiltrates seen on computed tomography of the chest in a 40-year-old woman with granulomatosis with polyangiitis (Wegener's).

Figure e40-2

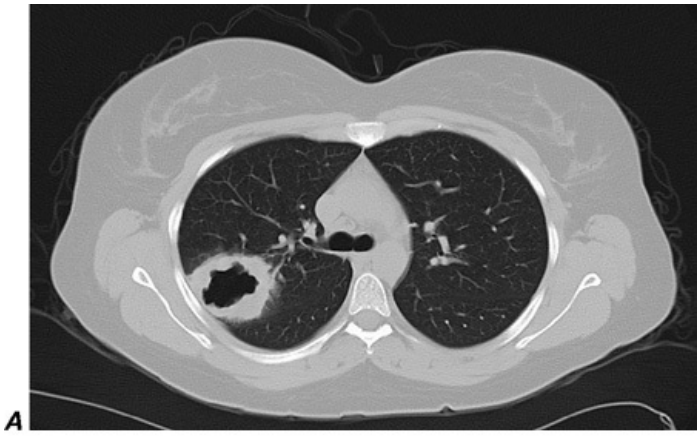

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomography of the chest in two patients with granulomatosis with polyangiitis (Wegener's) demonstrating (***A***) single and (***B***) multiple cavitary lung lesions.

Figure e40-3

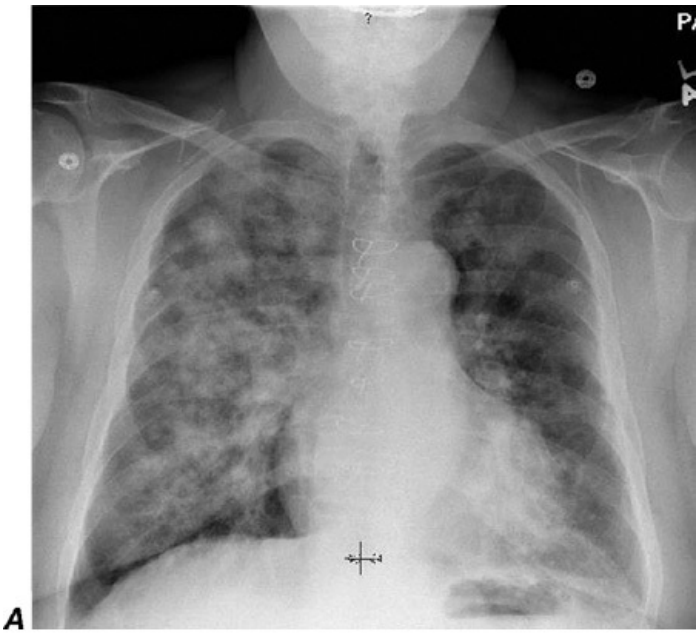

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bilateral ground-glass infiltrates due to alveolar hemorrhage from pulmonary capillaritis as seen in the same patient by (*A*) chest radiograph and (*B*) computed tomography. This manifestation can occur in granulomatosis with polyangiitis (Wegener's) or microscopic polyangiitis.

Figure e40-4

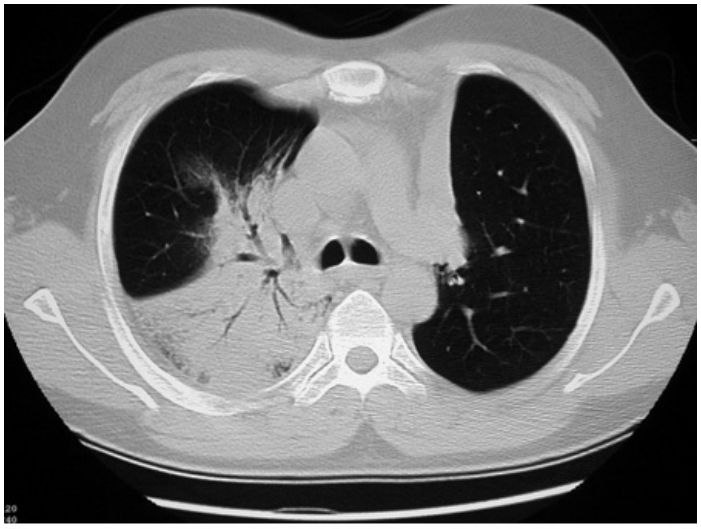

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomography of the chest demonstrating a dense infiltrate with air bronchograms involving a segment of the right upper lobe due to bacterial pneumonia in an immunosuppressed patient with granulomatosis with polyangiitis (Wegener's). Collapse of the left upper lobe secondary to endobronchial stenosis from granulomatosis with polyangiitis (Wegener's) also is seen on this image.

Figure e40-5

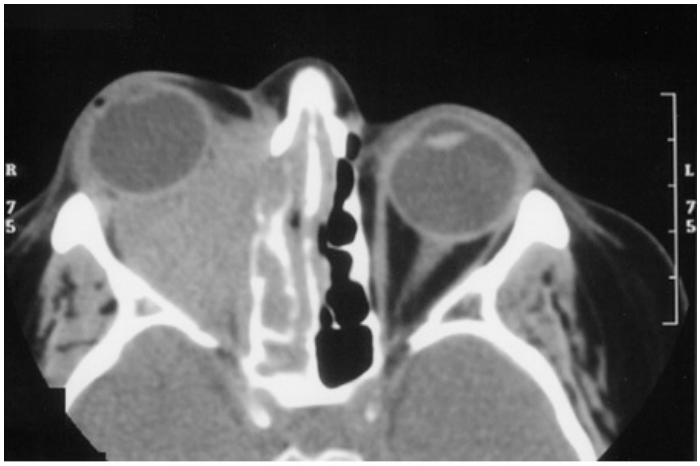

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomography of the orbits in a patient with granulomatosis with polyangiitis (Wegener's), who presented with right-eye proptosis. The image demonstrates inflammatory tissue extending from the ethmoid sinus through the lamina papyracea and filling the orbital space.

Figure e40-6

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

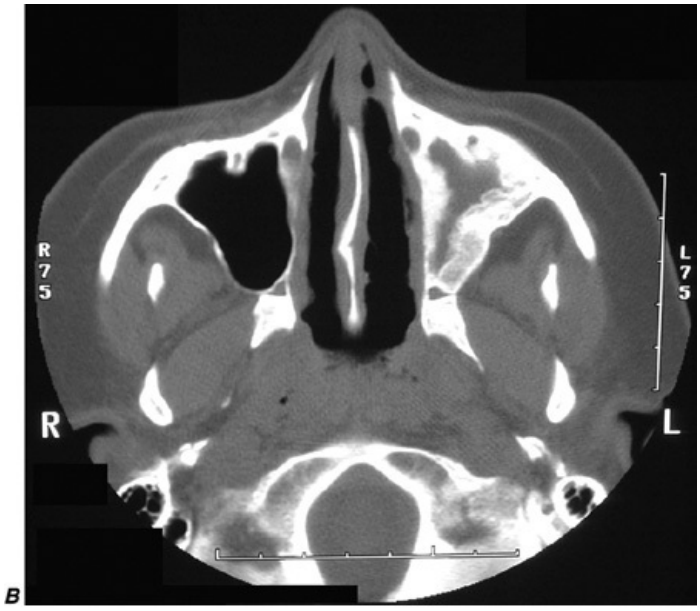

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomography of the sinuses in two patients with granulomatosis with polyangiitis (Wegener's) (***A***) Mucosal thickening of the bilateral maxillary sinuses and a perforation of the nasal septum. (***B***) Osteitis with obliteration of the left maxillary sinus in a patient with long-standing sinus disease.

Figure e40-7

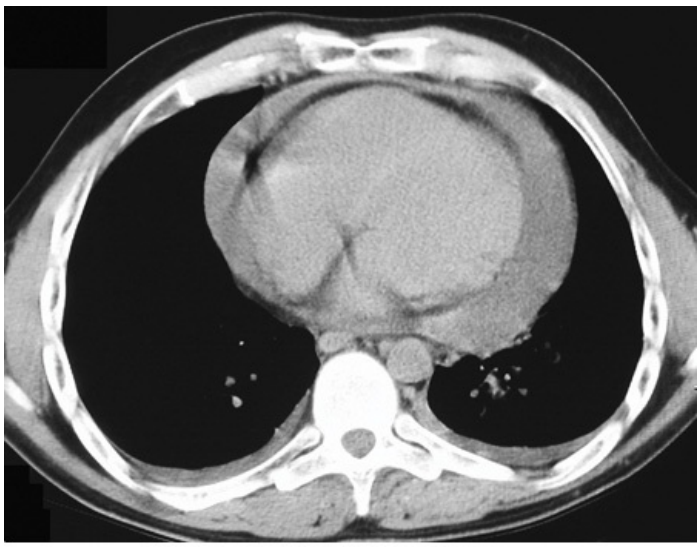

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Computed tomography of the chest demonstrating a large pericardial effusion in a patient with Churg-Strauss syndrome. Cardiac involvement is an important cause of morbidity and mortality in Churg-Strauss syndrome and can include myocarditis, endocarditis, and pericarditis.

Figure e40-8

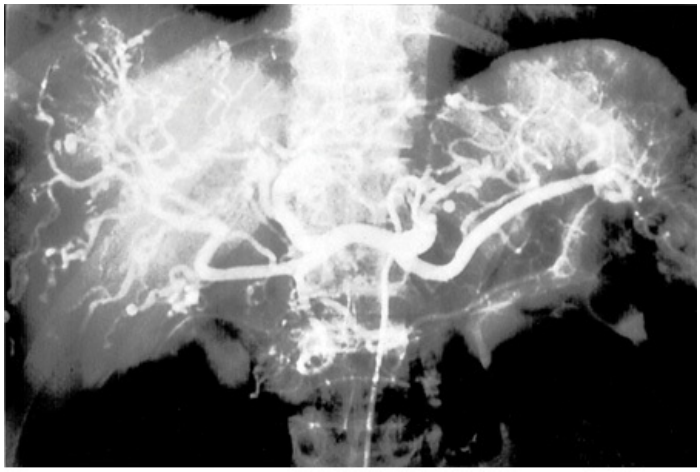

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arteriogram of a 40-year-old man with polyarteritis nodosa demonstrating microaneurysms in the hepatic circulation.

Figure e40-9

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cerebral arteriogram demonstrating beading along branches of the internal carotid artery in a patient with isolated central nervous system vasculitis.

Figure e40-10

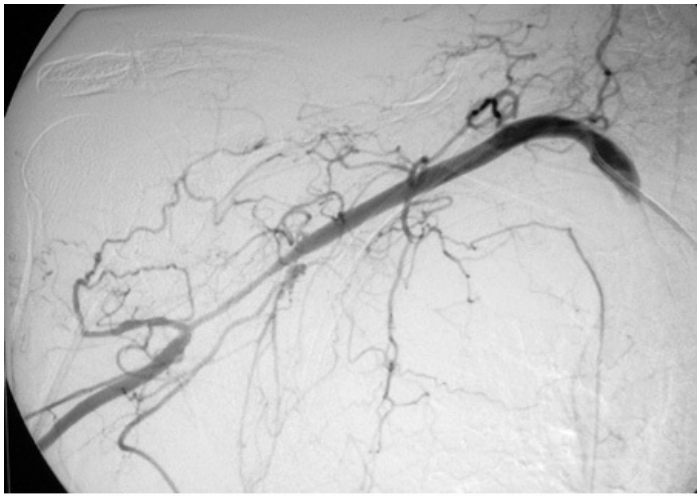

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Upper-extremity arteriogram demonstrating a long stenotic lesion of the axillary artery in a 75-year-old female with giant cell arteritis.

Figure e40-11

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Magnetic resonance imaging demonstrating extensive aneurysmal disease of the thoracic aorta in an 80-year-old female. The patient had been diagnosed with biopsy-proven giant cell arteritis 10 years prior to presenting with this aneurysm.

Figure e40-12

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arteriogram of the aortic arch demonstrating complete occlusion of the left common carotid artery just after its origin from the aorta. This 20-year-old female presented with syncope and was subsequently diagnosed with Takayasu's arteritis.

Figure e40-13

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arteriogram demonstrating stenosis of the abdominal aorta in a 25-year-old female with Takayasu's arteritis.

Figure e40-14

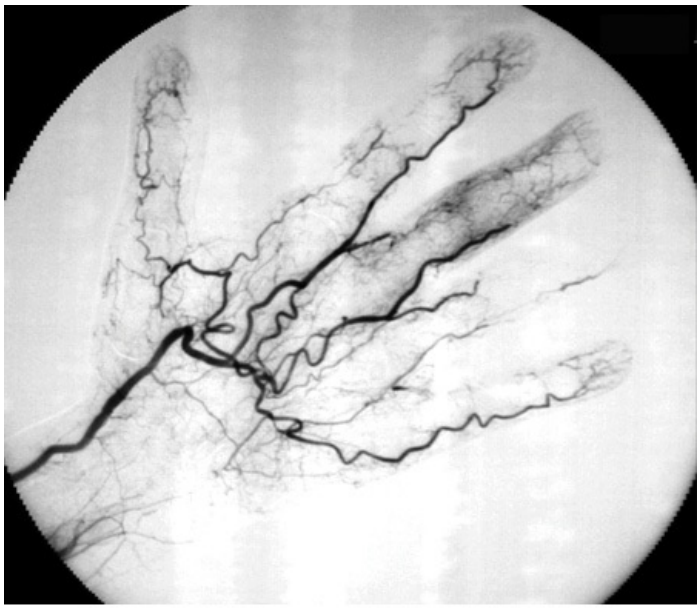

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arteriogram of the hand demonstrating arterial skip lesions and vessel cutoffs in a patient with cryoglobulinemia due to multiple myeloma.

Figure e40-15

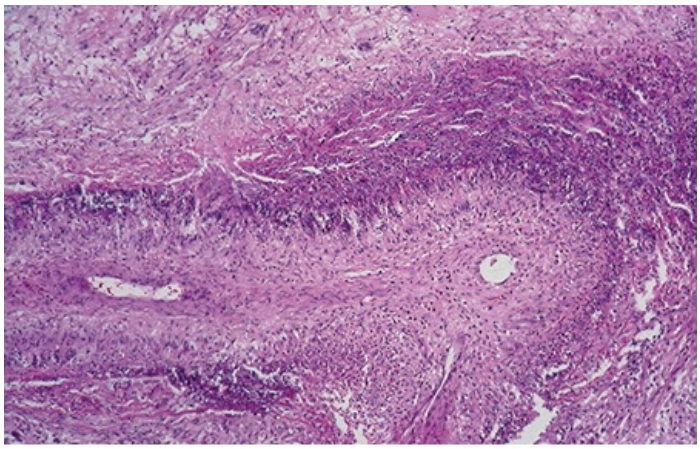

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lung histology in granulomatosis with polyangiitis (Wegener's). This lung biopsy shows areas of geographic necrosis with a border of histiocytes and giant cells. There is also vasculitis with neutrophils, lymphocytes, and giant cells infiltrating the wall of an artery.

Figure e40-16

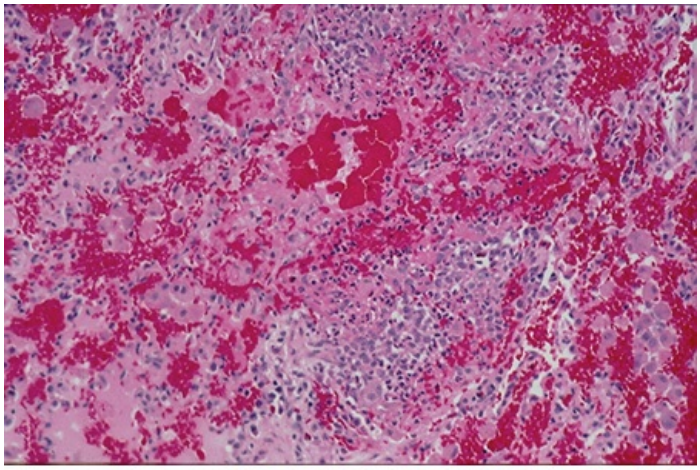

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lung histology in microscopic polyangiitis. This lung biopsy demonstrates hemorrhage in the alveolar spaces due to capillaritis in a patient with microscopic polyangiitis. Similar findings can also be seen in granulomatosis with polyangiitis (Wegener's) and less commonly in Churg-Strauss syndrome.

Figure e40-17

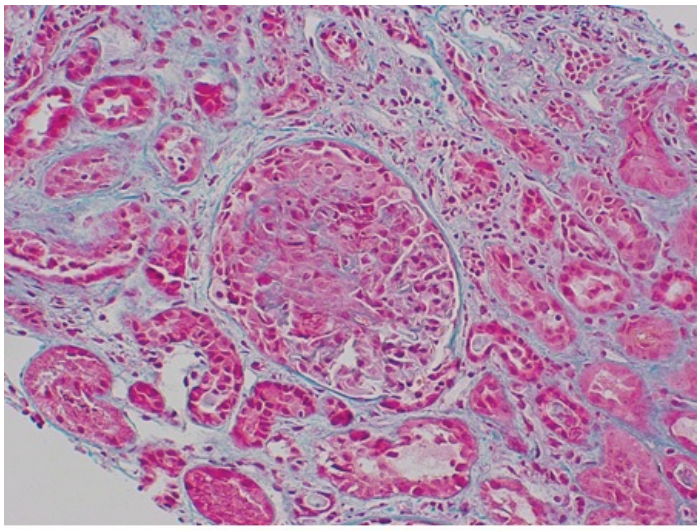

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Kidney biopsy in granulomatosis with polyangiitis (Wegener's). This renal biopsy shows a crescentic and necrotizing glomerulonephritis. These findings were focal and segmental with normal and scarred glomeruli also being found in the biopsy. By immunofluorescence and electron microscopy, no immune deposits were present, indicative of a pauci-immune glomerulonephritis. Similar findings can also be seen in microscopic polyangiitis and Churg-Strauss syndrome.

Figure e40-18

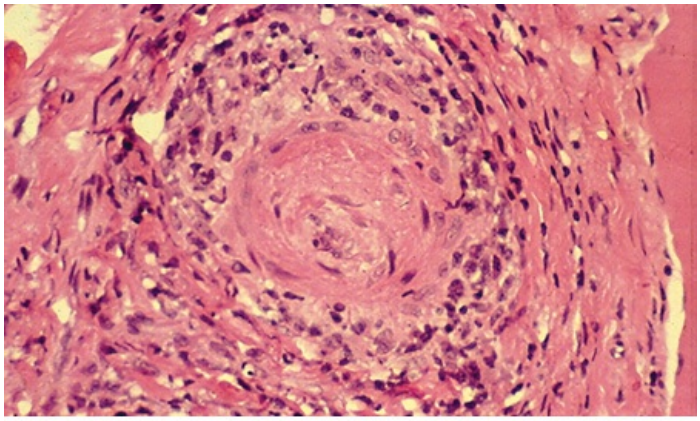

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sural nerve biopsy in polyarteritis nodosa. This sural nerve biopsy was performed in a patient with polyarteritis nodosa, who had presented with a mononeuritis multiplex. Neutrophils are seen infiltrating all layers of this medium-sized vessel, which resulted in vessel occlusion and nerve infraction.

Figure e40-19

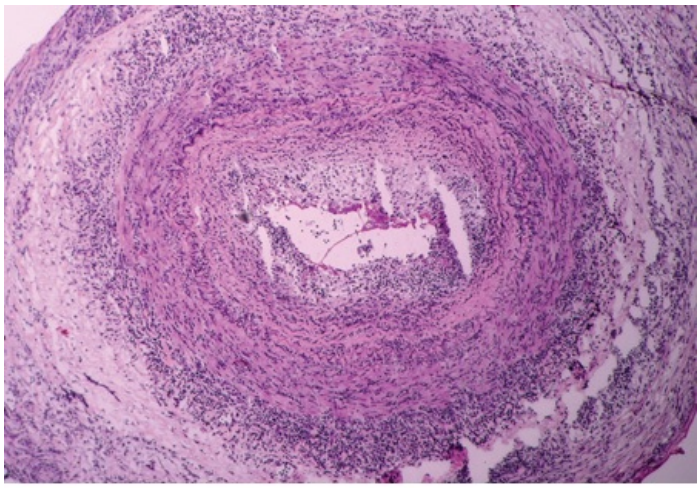

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Temporal artery biopsy in giant cell arteritis. This temporal artery biopsy demonstrates a panmural infiltration of mononuclear cells and lymphocytes that are particularly seen in the media and adventitia. Scattered giant cells are also present.

Figure e40-20

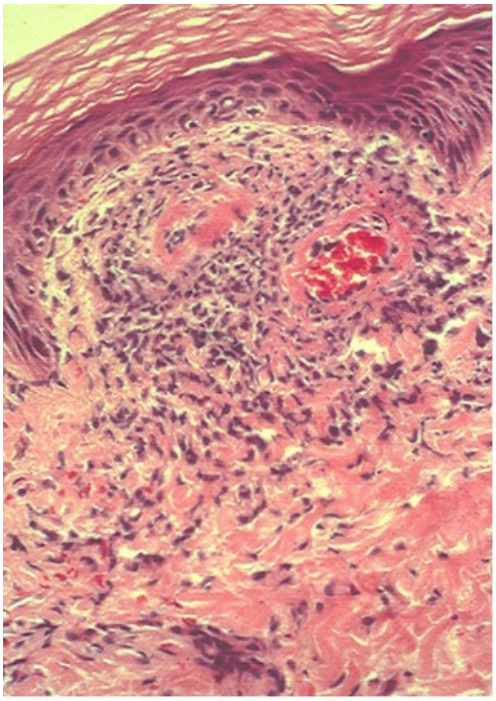

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cutaneousvasculitis. This skin biopsy reveals two arterioles beneath the dermis with a neutrophilic inflammatory infiltrate in and around the vessel wall with leukocytoclasis (nuclear debris). While such features are diagnostic of vasculitis, they can be seen in a variety of settings and are not specific for any single disease.

Figure e40-21

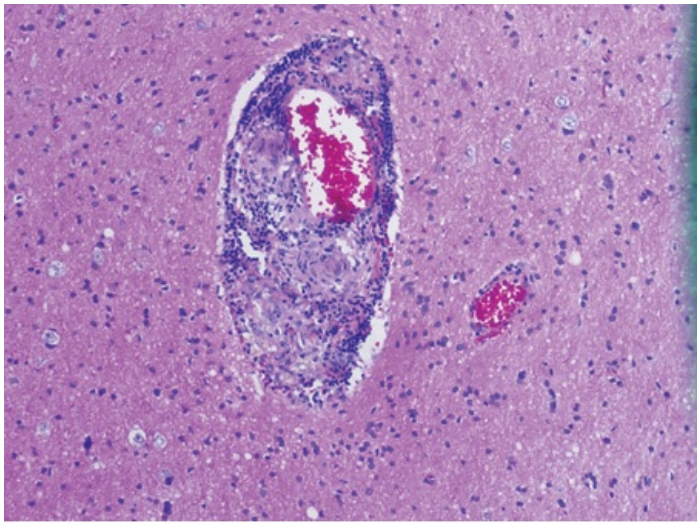

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Granulomatous primary angiitis of the central nervous system. This brain biopsy demonstrates a medium-sized artery with granulomatous inflammation present within the vessel wall indicative of a granulomatous vasculitis. This patient presented with progressive headache, clinical and radiographic features of a stroke, and had arteriographic features consistent with vasculitis. As no evidence of vasculitis could be found outside of the brain, this was consistent with granulomatous primary angiitis of the central nervous system (PACNS).

Copyright © The McGraw-Hill Companies. All rights reserved.
The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

RELAPSING POLYCHONDRITIS: INTRODUCTION

Relapsing polychondritis is an uncommon disorder of unknown cause characterized by inflammation of cartilage predominantly affecting the ears, nose, and laryngotracheobronchial tree. Other manifestations include scleritis, neurosensory hearing loss, polyarthritis, cardiac abnormalities, skin lesions, and glomerulonephritis. Relapsing polychondritis has been estimated to have an incidence of 3.5 per million population per year. The peak age of onset is between the ages of 40–50 years, but relapsing polychondritis may affect children and the elderly. It is found in all races, and both sexes are equally affected. No familial tendency is apparent. A significantly higher frequency of HLA-DR4 has been found in patients with relapsing polychondritis than in healthy individuals. A predominant subtype allele(s) of HLA-DR4 was not found. Approximately 30% of patients with relapsing polychondritis will have another rheumatologic disorder, the most frequent being systemic vasculitis, followed by rheumatoid arthritis, systemic lupus erythematosus (SLE), Sjögren's syndrome, or the spondyloarthritides. Nonrheumatic disorders associated with relapsing polychondritis include inflammatory bowel disease, primary biliary cirrhosis, and myelodysplastic syndrome **(Table 328-1)**. In most cases, these disorders antedate the appearance of relapsing polychondritis, usually by months or years.

Table 328-1 Disorders Associated with Relapsing Polychondritis[a]

Systemic vasculitis
Rheumatoid arthritis
Systemic lupus erythematosus
Sjögren's syndrome
Spondyloarthritides Behçet's disease
Inflammatory bowel disease
Primary biliary cirrhosis Myelodysplastic syndrome

[a]Systemic vasculitis is the most common association followed by rheumatoid arthritis, systemic lupus erythematosus, and Sjögren's syndrome.

Source: Modified from Michet.

PATHOLOGY AND PATHOPHYSIOLOGY

The earliest abnormality of hyaline and elastic cartilage noted histologically is a focal or diffuse loss of basophilic staining indicating depletion of proteoglycan from the cartilage matrix. Inflammatory infiltrates are found adjacent to involved cartilage and consist predominantly of mononuclear cells and occasional plasma cells. In acute disease, polymorphonuclear white cells may also be present. Destruction of cartilage begins at the outer edges and advances centrally. There is lacunar breakdown and loss of chondrocytes. Degenerating cartilage is replaced by granulation tissue and later by fibrosis and focal areas of calcification. Small loci of cartilage regeneration may be present. Immunofluorescence studies have shown immunoglobulins and complement at sites of involvement. Extracellular granular material observed in the degenerating cartilage matrix by electron microscopy has been interpreted to be enzymes, immunoglobulins, or proteoglycans.

Immunologic mechanisms play a role in the pathogenesis of relapsing polychondritis. The accumulating data strongly suggest that both humoral and cell-mediated immunity play an important role in the pathogenesis of relapsing polychondritis. Immunoglobulin and complement deposits are found at sites of inflammation. In addition, antibodies to type II collagen and to matrilin-1 and immune complexes are detected in the sera of some patients. The possibility that an immune response to type

II collagen may be important in the pathogenesis is supported experimentally by the occurrence of auricular chondritis in rats immunized with type II collagen. Antibodies to type II collagen are found in the sera of these animals, and immune deposits are detected at sites of ear inflammation. Humoral immune responses to type IX and type XI collagen, matrilin-1, and cartilage oligomeric matrix protein have been demonstrated in some patients. In a study, rats immunized with matrilin-1 were found to develop severe inspiratory stridor and swelling of the nasal septum. The rats had severe inflammation with erosions of the involved cartilage, which was characterized by increased numbers of CD4+ and CD8+ T cells in the lesions. The cartilage of the joints and ear pinna was not involved. All had IgG antibodies to matrilin-1. Matrilin-1 is a noncollagenous protein present in the extracellular matrix in cartilage. It is present in high concentrations in the trachea and is also present in the nasal septum but not in articular cartilage. A subsequent study demonstrated serum anti-matrilin-1 antibodies in approximately 13% of patients with relapsing polychondritis; approximately 70% of these patients had respiratory symptoms. Cell-mediated immunity may also be operative in causing tissue injury, since lymphocyte transformation can be demonstrated when lymphocytes of patients are exposed to cartilage extracts. T cells specific for type II collagen have been found in some patients, and CD4+ T cells have been observed at sites of cartilage inflammation.

CLINICAL MANIFESTATIONS

The onset of relapsing polychondritis is frequently abrupt with the appearance of one or two sites of cartilaginous inflammation. The pattern of cartilaginous involvement and the frequency of episodes vary widely among patients. Non-cartilaginous presentations may also occur. Systemic inflammatory features such as fever, fatigue, and weight loss occur and may precede the clinical signs of relapsing polychondritis by several weeks. Relapsing polychondritis may go unrecognized for several months or even years in patients who only initially manifest intermittent joint pain and/or swelling, or who have unexplained eye inflammation, hearing loss, valvular heart disease, or pulmonary symptoms.

Auricular chondritis is the most frequent presenting manifestation of relapsing polychondritis occurring in 40% of patients and eventually affecting about 85% of patients **(Table 328-2)**. One or both ears are involved, either sequentially or simultaneously. Patients experience the sudden onset of pain, tenderness, and swelling of the cartilaginous portion of the ear **(Fig. 328-1)**. This typically involves the pinna of the ears, sparing the earlobes because they do not contain cartilage. The overlying skin has a beefy red or violaceous color. Prolonged or recurrent episodes lead to cartilage destruction and result in a flabby or droopy ear. Swelling may close off the eustachian tube or the external auditory meatus, either of which can impair hearing. Inflammation of the internal auditory artery or its cochlear branch produces hearing loss, vertigo, ataxia, nausea, and vomiting. Vertigo is almost always accompanied by hearing loss.

Table 328-2 Clinical Manifestations of Relapsing Polychondritis

Clinical Feature	Presenting	Cumulative
	Frequency, %	
Auricular chondritis	43	89
Arthritis	32	72
Nasal chondritis	21	61
Ocular inflammation	18	59
Laryngotracheal symptoms	23	55
Reduced hearing	7	40
Saddle nose deformity	11	25
Cutaneous	4	25
Laryngotracheal stricture	15	23
Vasculitis	2	14
Elevated creatinine	7	13
Aortic or mitral regurgitation	0	12

Source: Modified from Kent et al.

Figure 328-1

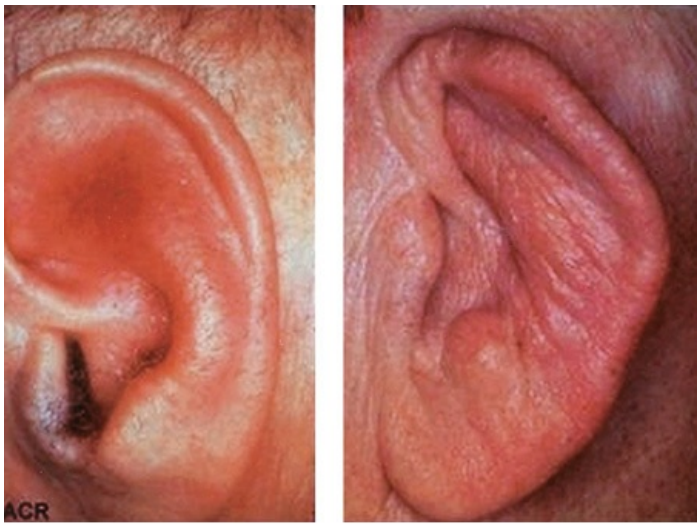

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left. The pinna is erythematous, swollen, and tender. Not shown is the ear lobule that is spared as there is no underlying cartilage. ***Right***. The pinna is thickened and deformed. The destruction of the underlying cartilage results in a floppy ear. *(Reprinted from the Clinical Slide Collection on the Rheumatic Diseases, ©1991, 1995, 1997, 1998, 1999. Used by permission of the American College of Rheumatology.)*

Approximately 61% of patients will develop nasal involvement, with 21% having this at the time of presentation. Patients may experience nasal stuffiness, rhinorrhea, and epistaxis. The bridge of the nose and surrounding tissue becomes red, swollen, and tender and may collapse, producing a saddlenose deformity **(Fig. 328-2)**. In some patients, nasal deformity develops insidiously without overt inflammation. Saddlenose is observed more frequently in younger patients, especially in women.

Figure 328-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Saddlenose results from destruction and collapse of the nasal cartilage. (*Reprinted from the Clinical Slide Collection on the Rheumatic Diseases, ©1991, 1995, 1997, 1998, 1999. Used by permission of the American College of Rheumatology.*)

Joint involvement is the presenting manifestation in relapsing polychondritis in approximately one-third of patients and may be present for several months before other features appear. Eventually, more than one-half of the patients will have arthralgias or arthritis. The arthritis is usually asymmetric and oligo- or polyarticular, and it involves both large and small peripheral joints. An episode of arthritis lasts from a few days to several weeks and resolves spontaneously without joint erosion or deformity. Attacks of arthritis may not be temporally related to other manifestations of relapsing polychondritis. Joint fluid has been reported to be noninflammatory. In addition to peripheral joints, inflammation may involve the costochondral, sternomanubrial, and sternoclavicular cartilages. Destruction of these cartilages may result in a pectus excavatum deformity or even a flail anterior chest wall.

Eye manifestations occur in more than one-half of patients and include conjunctivitis, episcleritis, scleritis, iritis, uveitis, and keratitis. Ocular inflammation can be severe and visually threatening. Other manifestations include eyelid and periorbital edema, proptosis, optic neuritis, extraocular muscle palsies, retinal vasculitis, and renal vein occlusion.

Laryngotracheobronchial involvement occurs in ~50% of patients and is among the most serious manifestations of relapsing polychondritis. Symptoms include hoarseness, a nonproductive cough, and tenderness over the larynx and proximal trachea. Mucosal edema, strictures, and/or collapse of laryngeal or tracheal cartilage may cause stridor and life-threatening airway obstruction necessitating tracheostomy. Involvement can extend into the lower airways resulting in tracheobronchomalacia. Collapse of cartilage in bronchi leads to pneumonia and, when extensive, to respiratory insufficiency.

Cardiac valvular regurgitation occurs in about 5–10% of patients and is due to progressive dilation of the valvular ring or to destruction of the valve cusps. Aortic regurgitation occurs in about 7% of patients with the mitral and other heart valves being affected less often. Other cardiac manifestations include pericarditis, myocarditis, coronary vasculitis, and conduction abnormalities. Aneurysms of the proximal, thoracic, or abdominal aorta may occur even in the absence of active chondritis and occasionally rupture.

Renal disease occurs in about 10% of patients. The most common renal lesions include mesangial expansion or segmental necrotizing glomerulonephritis, which have been reported to have small amounts of electron-dense deposits in the mesangium where there is also faint deposition of C3 and/or IgG or IgM. Tubulointerstitial disease and IgA nephropathy have also been reported.

Approximately 25% of patients have skin lesions, which can include purpura, erythema nodosum, erythema multiforme, angioedema/urticaria, livedo reticularis, and panniculitis.

Features of vasculitis are seen in up to 25% of patients and can affect any size vessel. Large vessel vasculitis may present with aortic aneurysms and medium vessel disease may affect the coronary, hepatic, mesenteric, or renal arteries or vessel supplying nerves. Skin vessel disease and involvement of the postcapillary venules can also occur. A variety of primary vasculitides have also been reported to occur in association with relapsing polychondritis **(Chap. 326)**. One specific overlap is the "MAGIC" syndrome (*m*outh *a*nd *g*enital ulcers with *i*nflamed *c*artilage) in which patients present with features of both relapsing polychondritis and Behçet's disease **(Chap. 327)**.

LABORATORY FINDINGS AND DIAGNOSTIC IMAGING

There are no laboratory features that are diagnostic for relapsing polychondritis. Mild leukocytosis and normocytic, normochromic anemia are often present. Eosinophilia is observed in 10% of patients. The erythrocyte sedimentation rate and C-reactive protein are usually elevated. Rheumatoid factor and antinuclear antibody tests are occasionally positive in low titers and complement levels are normal. Antibodies to type II collagen are present in fewer than one-half of the patients and are not specific. Circulating immune complexes may be detected, especially in patients with early active disease. Elevated levels of γ globulin may be present. Antineutrophil cytoplasmic antibodies (ANCA), either cytoplasmic (cANCA) or perinuclear (pANCA), are found in some patients with active disease. However, on target antigen specific testing, there are only occasional reports of positive myeloperoxidase-ANCA and proteinase 3-ANCA are very rarely found in relapsing polychondritis.

The upper and lower airways can be evaluated by imaging techniques such as computed tomography and magnetic resonance imaging. Bronchoscopy provides direct visualization of the airways but can be a high-risk procedure in patients with airway compromise. Pulmonary function testing with flow-volume loops can show inspiratory and/or expiratory obstruction. Imaging can also be useful to detect extracartilaginous disease. The chest film may show widening of the ascending or descending aorta due to an aneurysm, and cardiomegaly when aortic insufficiency is present. MRI can assess aortic aneurysmal dilatation. Electrocardiography and echocardiography can be useful in further evaluating for cardiac features of disease.

DIAGNOSIS

Diagnosis is based on recognition of the typical clinical features. Biopsies of the involved cartilage from the ear, nose, or respiratory tract will confirm the diagnosis but are only necessary when clinical features are not typical. Diagnostic criteria were suggested in 1976 by McAdam et al and modified by

Damiani and Levine in 1979. These criteria continue to be generally used in clinical practice. McAdam et al proposed the following: (1) recurrent chondritis of both auricles; (2) nonerosive inflammatory arthritis; (3) chondritis of nasal cartilage; (4) inflammation of ocular structures, including conjunctivitis, keratitis, scleritis/episcleritis, and/or uveitis; (5) chondritis of the laryngeal and/or tracheal cartilages; and (6) cochlear and/or vestibular damage manifested by neurosensory hearing loss, tinnitus, and/or vertigo. The diagnosis is certain when three or more of these features are present along with a positive biopsy from the ear, nasal, or respiratory cartilage. Damiani and Levine later suggested that the diagnosis could be made when one or more of the above features and a positive biopsy were present, when two or more separate sites of cartilage inflammation were present that responded to glucocorticoids or dapsone, or when three or more of the above features were present.

The differential diagnosis of relapsing polychondritis is centered around its sites of clinical involvement. Patients with granulomatosis with polyangiitis (Wegener's) may have a saddlenose and tracheal involvement but can be distinguished by the primary inflammation occurring in the mucosa at these sites, the absence of auricular involvement, and the presence of pulmonary parenchymal disease. Patients with Cogan's syndrome have interstitial keratitis and vestibular and auditory abnormalities, but this syndrome does not involve the respiratory tract or ears. Reactive arthritis may initially resemble relapsing polychondritis because of oligoarticular arthritis and eye involvement, but it is distinguished in time by the appearance of urethritis and typical mucocutaneous lesions and the absence of nose or ear cartilage involvement. Rheumatoid arthritis may initially suggest relapsing polychondritis because of arthritis and eye inflammation. The arthritis in rheumatoid arthritis, however, is erosive and symmetric. In addition, rheumatoid factor titers are usually high compared with those in relapsing polychondritis and anti-cyclic citrullinated peptide is usually not seen. Bacterial infection of the pinna may be mistaken for relapsing polychondritis but differs by usually involving only one ear, including the earlobe. Auricular cartilage may also be damaged by trauma or frostbite.

Treatment: Relapsing Polychondritis

In patients with active chondritis, prednisone, 40–60 mg/d, is often effective in suppressing disease activity; it is tapered gradually once disease is controlled. In some patients, prednisone can be stopped, while in others low doses in the range of 5–10 mg/d are required for continued suppression of disease. Dapsone 50–100 mg/d has been effective for cartilage inflammation and joint features in some patients. Other immunosuppressive drugs such as cyclophosphamide, methotrexate, azathioprine, or cyclosporine should be reserved for patients who have severe organ-threatening disease, fail to respond to prednisone, or who require high doses for control of disease activity. Patients with significant ocular inflammation often require intraocular glucocorticoids as well as high doses of prednisone. There are a small number of reports on the use of tumor necrosis factor antagonists, which are too few in number to assess efficacy. A small retrospective series of nine patients did not find anti-CD20 (rituximab) to provide benefit although this experience remains too small to draw firm conclusions. Heart valve replacement or repair of an aortic aneurysm may be necessary. When obstruction is severe, tracheostomy is required. Stents may be necessary in patients with tracheobronchial collapse.

PATIENT OUTCOME, PROGNOSIS, AND SURVIVAL

The course of relapsing polychondritis is highly variable, with inflammatory episodes lasting from a few days to several weeks and then subsiding spontaneously. Attacks may recur at intervals varying from weeks to months. In other patients, the disease has a chronic, smoldering course. In a few patients, the disease may be limited to one or two episodes of cartilage inflammation. In one study, the 5-year estimated survival rate was 74% and the 10-year survival rate 55%. In contrast to earlier series, only about one-half of the deaths could be attributed to relapsing polychondritis or complications of treatment. Pulmonary complications accounted for only 10% of all fatalities. In general, patients with more widespread disease have a worse prognosis.

ACKNOWLEDGMENT

This chapter represents a revised version of the text authored by Dr. Bruce C. Gilliland that appeared in previous editions of Harrison's Principles of Internal Medicine. Dr. Gilliland passed away on February 17, 2007, and had been a contributor to Harrison's since the 11th edition.

FURTHER READINGS

Ernst A et al: Relapsing polychondritis and airway involvement. Chest 135:1024, 2009[PMID: 19017885] [Full Text]

Kent PD et al: Relapsing polychondritis. Curr Opin Rheumatol 16:56, 2004[PMID: 14673390] [Full Text]

Letko E et al: Relapsing polychondritis: A clinical review. Semin Arthritis Rheum 31:384, 2002[PMID: 12077711] [Full Text]

Michet CJ et al: Relapsing polychondritis. Survival and predictive role of early disease manifestations. Ann Intern Med 104:74, 1986[PMID: 3484422] [Full Text]

Staats BA et al: Relapsing polychondritis. Semin Respir Crit Care Med 23:145, 2002[PMID: 16088607] [Full Text]

Trentham DE, Le CH: Relapsing polychondritis. Ann Intern Med 129:114, 1998[PMID: 9669970] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEFINITION

Sarcoidosis is an inflammatory disease characterized by the presence of noncaseating granulomas. The disease is often multisystem and requires the presence of involvement in two or more organs for a specific diagnosis. The finding of granulomas is not specific for sarcoidosis, and other conditions known to cause granulomas must be ruled out. These conditions include mycobacterial and fungal infections, malignancy, and environmental agents such as beryllium. While sarcoidosis can affect virtually every organ of the body, the lung is most commonly affected. Other organs commonly affected are the liver, skin, and eye. The clinical outcome of sarcoidosis varies, with remission occurring in over one-half of the patients within a few years of diagnosis; however, the remaining patients may develop a chronic disease that lasts for decades.

ETIOLOGY

Despite multiple investigations, the cause of sarcoidosis remains unknown. Currently, the most likely etiology is an infectious or noninfectious environmental agent that triggers an inflammatory response in a genetically susceptible host. Among the possible infectious agents, careful studies have shown a much higher incidence of *Propionibacter acnes* in the lymph nodes of sarcoidosis patients compared to controls. An animal model has shown that *P. acnes* can induce a granulomatous response in mice similar to sarcoidosis. Others have demonstrated the presence of a mycobacterial protein [*Mycobacterium tuberculosis* catalase-peroxidase (mKatG)] in the granulomas of some sarcoidosis patients. This protein is very resistant to degradation and may represent the persistent antigen in sarcoidosis. Immune response to this and other mycobacterial proteins has been documented by another laboratory. These studies suggest that a mycobacterium similar to *M. tuberculosis* could be responsible for sarcoidosis. The mechanism exposure/infection with such agents has been the focus of other studies. Environmental exposures to insecticides and mold have been associated with an increased risk for disease. In addition, health care workers appear to have an increased risk. Also, sarcoidosis in a donor organ has occurred after transplantation into a sarcoidosis patient. Some authors have suggested that sarcoidosis is not due to a single agent but represents a particular host response to multiple agents. Some studies have been able to correlate the environmental exposures to genetic markers. These studies have supported the hypothesis that a genetically susceptible host is a key factor in the disease.

INCIDENCE AND PREVALENCE

Sarcoidosis is seen worldwide, with the highest prevalence reported in the Nordic population. In the United States, the disease has been reported more commonly in African Americans than whites, with the ratio of African Americans to whites ranging from 3:1 to 17:0. Women appear to be slightly more susceptible than men. The lower estimate is from a large health maintenance organization in Detroit. The earlier American studies finding the higher incidence in African Americans may have been influenced by the fact that African Americans seem to develop more extensive and chronic pulmonary disease. Since most sarcoidosis clinics are run by pulmonologists, a selection bias may have occurred. Worldwide, the prevalence of the disease varies from 20–60 per 100,000 for many groups such as Japanese, Italians, and American whites. Higher rate occurs in Ireland and Nordic countries. In one closely observed community in Sweden, the lifetime risk for developing sarcoidosis was 3%.

Sarcoidosis often occurs in young, otherwise healthy adults. It is uncommon to diagnose the disease in someone under age 18. However, it has become clear that a second peak in incidence develops around age 60. In a study of >700 newly diagnosed sarcoidosis patients in the United States, one-half of the patients were ≥40 years at the time of diagnosis.

Although most cases of sarcoidosis are sporadic, a familial form of the disease exists. At least 5% of patients with sarcoidosis will have a family member with sarcoidosis. Sarcoidosis patients who are Irish or African American seem to have a two to three times higher rate of familial disease.

PATHOPHYSIOLOGY AND IMMUNOPATHOGENESIS

The granuloma is the pathologic hallmark of sarcoidosis. A distinct feature of sarcoidosis is the local accumulation of inflammatory cells. Extensive studies in the lung using bronchoalveolar lavage (BAL) have demonstrated that the initial inflammatory response is an influx of T helper cells. In addition, there is an accumulation of activated monocytes. **Figure 329-1** is a proposed model for sarcoidosis. Using

the HLA-CD4 complex, antigen-presenting cells present an unknown antigen to the helper T cell. Studies have clarified that specific HLA haplotypes such as HLA-DRB1*1101 are associated with an increased risk for developing sarcoidosis. In addition, different HLA haplotypes are associated with different clinical outcomes.

Figure 329-1

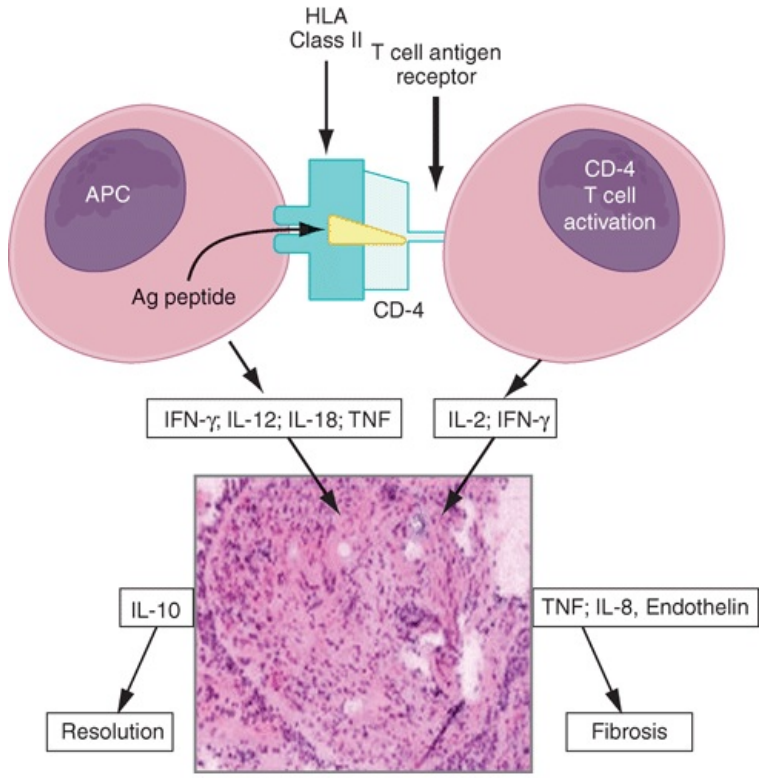

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of initial events of sarcoidosis. The antigen-presenting cell and helper T cell complex leads to the release of multiple cytokines. This forms a granuloma. Over time, the granuloma may resolve or lead to chronic disease, including fibrosis. APC, antigen-presenting cell; HLA, human leukocyte antigen; IFN, interferon; IL, interleukin; TNF, tumor necrosis factor.

The macrophage/helper T cell cluster leads to activation with the increased release of several cytokines. These include interleukin (IL)-2 released from the T cell and interferon ϒ and tumor necrosis factor (TNF) released by the macrophage. The T cell is a necessary part of the initial inflammatory response. In advanced, untreated HIV infection, patients who lack helper T cells rarely develop sarcoidosis. In contrast, several reports confirm that sarcoidosis becomes unmasked as HIV-infected individuals receive antiretroviral therapy, with subsequent restoration of their immune system. In contrast, treatment of established pulmonary sarcoidosis with cyclosporine, a drug that downregulates helper T cell responses, seems to have little impact on sarcoidosis.

The granulomatous response of sarcoidosis can resolve with or without therapy. However, in at least 20% of patients with sarcoidosis, a chronic form of the disease develops. This persistent form of the disease is associated with the secretion of high levels of IL-8. Also, studies have reported that in patients with this chronic form of disease excessive amounts of TNF are released in the areas of inflammation.

It is sometimes difficult to determine early on the ultimate clinical outcome of sarcoidosis. One form of the disease, *Löfgren's syndrome*, consists of erythema nodosum, hilar adenopathy on chest roentgenogram, and uveitis. Löfgren's syndrome is associated with a good prognosis, with >90% of patients experiencing disease resolution within 2 years. A recently proposed expansion of the term Lofgren's syndrome includes periarticular arthritis without erythema nodosum. Recent studies have demonstrated that the HLA-DRB1*03 was found in two-thirds of Scandinavian patients with Löfgren's syndrome. More than 95% of those patients who were HLA-DRB1*03 positive had resolution of their disease within 2 years, while nearly one-half of the remaining patients had disease for more than 2 years. It remains to be determined whether these observations can be applied to a non-Scandinavian population.

CLINICAL MANIFESTATIONS

The presentation of sarcoidosis ranges from patients who are asymptomatic to those with organ failure. It is unclear how often sarcoidosis is asymptomatic. In countries where routine chest roentgenogram screening is performed, 20–30% of pulmonary cases are detected in asymptomatic individuals. The inability to screen for other asymptomatic forms of the disease would suggest that as many as one-third of sarcoidosis patients are asymptomatic.

Respiratory complaints including cough and dyspnea are the most common presenting symptoms. In many cases, the patient presents with a 2–4 week history of these symptoms. Unfortunately, due to the nonspecific nature of pulmonary symptoms, the patient may see physicians for up to a year before a diagnosis is confirmed. For these patients, the diagnosis of sarcoidosis is usually only suggested when a chest roentgenogram is performed.

Symptoms related to cutaneous and ocular disease are the next two most common complaints. Skin lesions are often nonspecific. However, since these lesions are readily observed, the patient and treating physician are often led to a diagnosis. In contrast to patients with pulmonary disease, patients with cutaneous lesions are more likely to be diagnosed within 6 months of symptoms.

Nonspecific constitutional symptoms include fatigue, fever, night sweats, and weight loss. Fatigue is perhaps the most common constitutional symptom that affects these patients. Given its insidious nature, patients are usually not aware of the association with their sarcoidosis until their disease resolves.

The overall incidence of sarcoidosis at the time of diagnosis and eventual common organ involvement are summarized in **Table 329-1**. Over time, skin, eye, and neurologic involvement seem more apparent. In the United States, the frequency of specific organ involvement appears to be affected by age, race, and gender. For example, eye disease is more common among African Americans. Under the age of 40, it occurs more frequently in women. However, in those diagnosed over the age of 40, eye disease is more common in men.

Table 329-1 Frequency of Common Organ Involvement and Lifetime Risk[a]

	Presentation, %[b]	Follow-up, %[c]
Lung	95	94
Skin	24	43
Eye	12	29
Extrathoracic lymph node	15	16
Liver	12	14
Spleen	7	8
Neurologic	5	16
Cardiac	2	3

[a]Patients could have more than one organ involved.

[b]From ACCESS study of 736 patients evaluated. within 6 months of diagnosis.

[c]From follow-up of 1024 sarcoidosis patients seen at the University of Cincinnati Interstitial Lung Disease and Sarcoidosis Clinic from 2002–2006.

Lung

Lung involvement occurs in >90% of sarcoidosis patients. The most commonly used method for detecting lung disease is still the chest roentgenogram. **Figure 329-2** illustrates the chest roentgenogram from a sarcoidosis patient with bilateral hilar adenopathy. Although the CT scan has changed the diagnostic approach to interstitial lung disease, the CT scan is not usually considered a monitoring tool for patients with sarcoidosis. **Figure 329-3** demonstrates some of the characteristic CT features, including peribronchial thickening and reticular nodular changes, which are predominantly subpleural. The peribronchial thickening seen on CT scan seems to explain the high yield of granulomas from bronchial biopsies performed for diagnosis.

Figure 329-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Posterior-anterior chest roentgenogram demonstrating bilateral hilar adenopathy, stage 1 disease.

Figure 329-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

High-resolution CT scan of chest demonstrating patchy reticular nodularity, including areas of confluence.

While the CT scan is more sensitive, the standard scoring system described by Scadding in 1961 for chest roentgenograms remains the preferred method of characterizing the chest involvement. Stage 1 is hilar adenopathy alone (Fig. 329-2), often with right paratracheal involvement. Stage 2 is a combination of adenopathy plus infiltrates, whereas stage 3 reveals infiltrates alone. Stage 4 consists of fibrosis. Usually the infiltrates in sarcoidosis are predominantly an upper lobe process. Only in a few noninfectious diseases is an upper lobe predominance noted. In addition to sarcoidosis, the differential diagnosis of upper lobe disease includes hypersensitivity pneumonitis, silicosis, and Langerhans cell histiocytosis. For infectious diseases, tuberculosis and *Pneumocystis* pneumonia can often present as upper lobe diseases.

Lung volumes, mechanics, and diffusion all are useful in evaluating interstitial lung diseases such as sarcoidosis. The diffusion of carbon monoxide (DLCO) is the most sensitive test for an interstitial lung disease. Reduced lung volumes are a reflection of the restrictive lung disease seen in sarcoidosis.

However, a third of the patients presenting with sarcoidosis still have lung volumes within the normal range, despite abnormal chest roentgenograms and dyspnea.

Approximately one-half of sarcoidosis patients present with obstructive disease, reflected by a reduced ratio of forced vital capacity expired in one second (FEV_1/FVC). Cough is a very common symptom.

Airway involvement causing varying degrees of obstruction underlies the cough in most sarcoidosis patients. Airway hyperreactivity as determined by methacholine challenge will be positive in some of these patients. A few patients with cough will respond to traditional bronchodilators as the only form of treatment. In some cases, high-dose inhaled glucocorticoids alone are useful.

Pulmonary arterial hypertension is reported in at least 5% of sarcoidosis patients. Either direct vascular involvement or the consequence of fibrotic changes in the lung can lead to pulmonary arterial hypertension. In sarcoidosis patients with end-stage fibrosis awaiting lung transplant, 70% will have pulmonary arterial hypertension. This is a much higher incidence than that reported for other fibrotic lung diseases. In less advanced, but still symptomatic, patients pulmonary arterial hypertension has been noted in up to 50% of the cases. Because sarcoidosis-associated pulmonary arterial hypertension may respond to therapy, evaluation for this should be considered in persistently symptomatic patients.

Skin

Skin involvement is eventually identified in over a third of patients with sarcoidosis. The classic cutaneous lesions include erythema nodosum, maculopapular lesions, hyper- and hypopigmentation, keloid formation, and subcutaneous nodules. A specific complex of involvement of the bridge of the nose, the area beneath the eyes, and the cheeks is referred to as *lupus pernio* **(Fig. 329-4)** and is diagnostic for a chronic form of sarcoidosis.

Figure 329-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic inflammatory lesions around nose, eyes, and cheeks, referred to as lupus pernio.

In contrast, erythema nodosum is a transient rash that can be seen in association with hilar adenopathy and uveitis (Löfgren's syndrome). Erythema nodosum is more common in women and in certain self-described demographic groups including whites and Puerto Ricans. In the United States, the other manifestations of skin sarcoidosis, especially lupus pernio, are more common in African Americans than whites.

The maculopapular lesions from sarcoidosis are the most common chronic form of the disease **(Fig. 329-5)**. These are often overlooked by the patient and physician, since they are chronic and not painful. Initially, these lesions are usually purplish papules and are often indurated. They can become confluent and infiltrate large areas of the skin. With treatment, the color and induration may fade. Because these lesions are caused by noncaseating granulomas, the diagnosis of sarcoidosis can be readily made by a skin biopsy.

Figure 329-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Maculopapular lesions on the trunk of a sarcoidosis patient.

Eye

The frequency of ocular manifestations for sarcoidosis varies depending on race. In Japan, >70% of sarcoidosis patients develop ocular disease, while in the United States only 30% have eye disease, with problems more common in African Americans than whites. Although the most common manifestation is an anterior uveitis, over a quarter of patients will have inflammation at the posterior of the eye, including retinitis and pars planitis. While symptoms such as photophobia, blurred vision, and increased tearing can occur, some asymptomatic patients still have active inflammation. Initially asymptomatic patients with ocular sarcoidosis can eventually develop blindness. Therefore, it is recommended that all patients with sarcoidosis receive a dedicated ophthalmologic examination. Sicca is seen in over one-half of the chronic sarcoidosis patients. Dry eyes appear to be a reflection of prior lacrimal gland disease. Although the patient may no longer have active inflammation, the dry eyes may require natural tears or other lubricants.

Liver

Using biopsies to detect granulomatous disease, liver involvement can be identified in over one-half of sarcoidosis patients. However, using liver function studies, only 20–30% of patients will have evidence of liver involvement. The most common abnormality of liver function is an elevation of the alkaline phosphatase level, consistent with an obstructive pattern. In addition, elevated transaminase levels can occur. An elevated bilirubin level is a marker for more advanced liver disease. Overall, only 5% of sarcoidosis patients have sufficient symptoms from their liver disease to require specific therapy. Although symptoms can be due to hepatomegaly, more frequently symptoms result from extensive intrahepatic cholestasis leading to portal hypertension. In this case, ascites and esophageal varices can occur. It is rare that a sarcoidosis patient will require a liver transplant, because even the patient with cirrhosis due to sarcoidosis can respond to systemic therapy. On a cautionary note, patients with both sarcoidosis and hepatitis C should avoid therapy with interferon α because of its association with the development or worsening of granulomatous disease.

Bone Marrow and Spleen

One or more bone marrow manifestations can be identified in many sarcoidosis patients. The most common hematologic problem is lymphopenia, which is a reflection of sequestration of the lymphocytes into the areas of inflammation. Anemia occurs in 20% of patients and leukopenia is less common. Bone marrow examination will reveal granulomas in about a third of patients. Splenomegaly can be detected in 5–10% of patients, but splenic biopsy reveals granulomas in 60% of patients. The CT scan can be relatively specific for sarcoidosis involvement of the spleen **(Fig. 329-6)**. Both bone marrow and spleen involvement are more common in African Americans than whites. These manifestations alone are rarely an indication for therapy. On occasion, splenectomy may be indicated for massive symptomatic splenomegaly or profound pancytopenia.

Figure 329-6

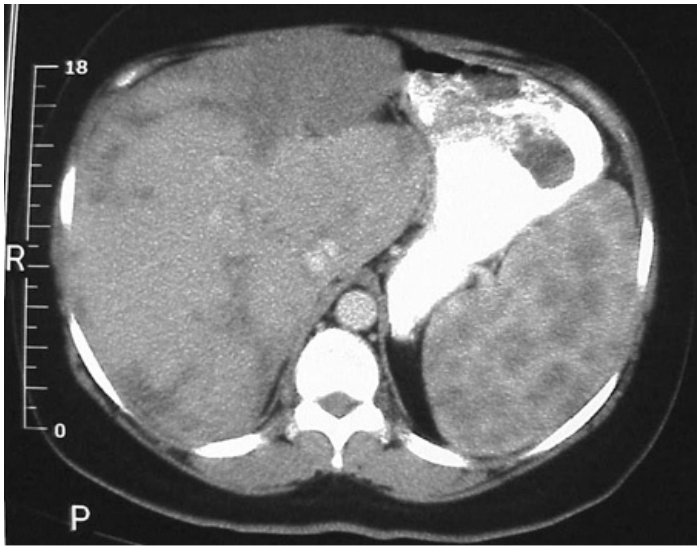

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of the abdomen after oral and intravenous contrast. The stomach is compressed by the enlarged spleen. Within the spleen, areas of hypo- and hyperdensity are identified.

Calcium Metabolism

Hypercalcemia and/or hypercalciuria occurs in about 10% of sarcoidosis patients. It is more common in whites than African Americans and in men. The mechanism of abnormal calcium metabolism is increased production of 1,25-dihydroxyvitamin D by the granuloma itself. The 1,25-dihydroxyvitamin D causes increased intestinal absorption of calcium, leading to hypercalcemia with a suppressed parathyroid hormone (PTH) level **(Chap. 353)**. Increased exogenous vitamin D from diet or sunlight exposure may exacerbate this problem. Serum calcium should be determined as part of the initial evaluation of all sarcoidosis patients, and a repeat determination may be useful during the summer months with increased sun exposure. In patients with a history of renal calculi, a 24-h urine calcium measurement should be obtained. If a sarcoidosis patient with a history of renal calculi is to be placed on calcium supplements, a follow-up 24-h urine calcium level should be measured.

Renal Disease

Direct kidney involvement occurs in <5% of sarcoidosis patients. It is associated with granulomas in the kidney itself and can lead to nephritis. However, hypercalcemia is the most likely cause of sarcoidosis-associated renal disease. In 1–2% of sarcoidosis patients, acute renal failure has been encountered as a result of hyper-calcemia. Treatment of the hypercalcemia with glucocorticoids and other therapies often improves, but does not totally resolve, the renal dysfunction.

Nervous System

Neurologic disease is reported in 5–10% of sarcoidosis patients and appears to be of equal frequency across all ethnic groups. Any part of the central or peripheral nervous system can be affected. The presence of granulomatous inflammation is often visible on MRI studies. The MRI with gadolinium enhancement may demonstrate space-occupying lesions, but the MRI can be negative due to small lesions or the effect of systemic therapy in reducing the inflammation. The cerebral spinal fluid (CSF) findings include lymphocytic meningitis with a mild increase in protein. The CSF glucose is usually normal but can be low. Certain areas of the nervous system are more commonly affected in neurosarcoidosis. These include cranial nerve involvement, basilar meningitis, myelopathy, and anterior hypothalamic disease with associated diabetes insipidus **(Chap. 340)**. Seizures and cognitive changes also occur. Of the cranial nerves, seventh nerve paralysis can be transient and mistaken for Bell's palsy (idiopathic seventh nerve paralysis). Since this form of neurosarcoidosis often resolves within weeks and may not recur, it may have occurred prior to a definitive diagnosis of sarcoidosis. Optic neuritis is another cranial nerve manifestation of sarcoidosis. This manifestation is more chronic and usually requires long-term systemic therapy. It can be associated with both anterior and posterior uveitis. Differentiating between neurosarcoidosis and multiple sclerosis can be difficult at times. Optic neuritis can occur in both diseases. In some patients with sarcoidosis, multiple enhancing white matter abnormalities may be detected by MRI, suggesting multiple sclerosis. In such cases, the presence of meningeal enhancement or hypothalamic involvement suggests neurosarcoidosis, as does evidence of extraneurologic disease such as pulmonary or skin involvement, which also suggests sarcoidosis. Since the response of neurosarcoidosis to glucocorticoids and cytotoxic therapy is different from that of multiple sclerosis, differentiating between these disease entities is important.

Cardiac

The presence of cardiac involvement is influenced by race. Over a quarter of Japanese sarcoidosis patients develop cardiac disease, whereas only 5% of sarcoidosis patients in the United States and Europe develop cardiac disease. However, there is no apparent difference between whites and African Americans. Cardiac disease usually presents as either congestive heart failure or cardiac arrhythmias. Both manifestations result from infiltration of the heart muscle by granulomas. Diffuse granulomatous involvement of the heart muscle can lead to ejection fractions below 10%. Even in this situation, improvement in the ejection fraction can occur with systemic therapy. Arrhythmias can also occur with diffuse infiltration or with more patchy cardiac involvement. If the atrioventricular (AV) node is infiltrated, heart block can occur. This can be detected by routine electrocardiography. Ventricular arrhythmias and sudden death due to ventricular tachycardia are common causes of death. Arrhythmias are best detected using 24-h ambulatory monitoring. Because ventricular arrhythmias are usually multifocal due to patchy multiple granulomas in the heart, ablation therapy is not useful. Patients with significant ventricular arrhythmias should be considered for an implanted defibrillator, which appears to have reduced the rate of death in cardiac sarcoidosis. While systemic therapy can be useful in treating the arrhythmias, patients may still have malignant arrhythmias up to 6 months after starting successful treatment, and the risk for recurrent arrhythmias occurs whenever medications are tapered.

Musculoskeletal System

Direct granulomatous bone and muscle involvement as documented by x-ray, MRI **(Fig. 329-7)**, gallium scan, or biopsy can be seen in about 10% of sarcoidosis patients. However, a larger percentage of sarcoidosis patients complain of myalgias and arthralgias. These complaints are similar to those reported by patients with other inflammatory diseases, including chronic infections such as mononucleosis. Fatigue associated with sarcoidosis may be overwhelming for many patients. Recent studies have demonstrated a link between fatigue and small peripheral nerve fiber disease in sarcoidosis.

Figure 329-7

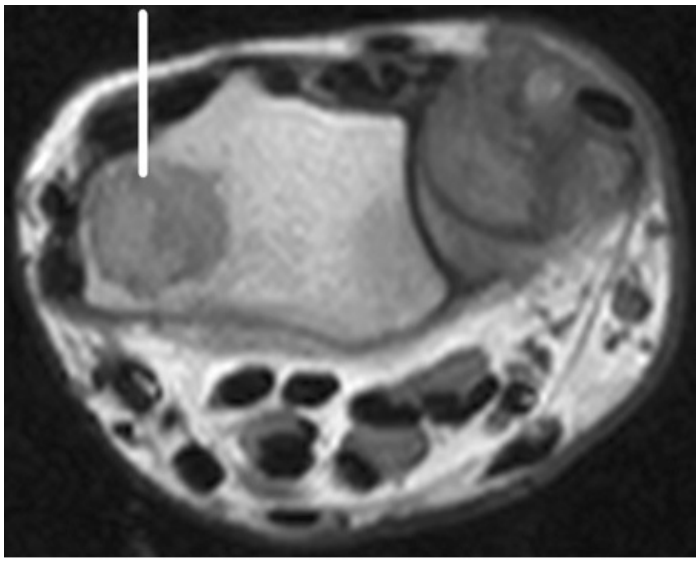

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI of wrist demonstrating large cyst in a sarcoidosis patient (*line*).

Other Organ Involvement

Although sarcoidosis can affect any organ of the body, rarely does it involve the breast, testes, ovary, or stomach. Because of the rarity of involvement, a mass in one of these areas requires a biopsy to rule out other diseases including cancer. For example, in a study of breast problems in female sarcoidosis patients, a breast lesion was more likely to be granulomas from sarcoidosis than from breast cancer. However, findings on the physical examination or mammogram cannot reliably differentiate between these lesions. More importantly, as women with sarcoidosis age, breast cancer becomes more common. Therefore, it is recommended that routine screening including mammography be performed along with other imaging studies (ultrasound, MRI) or biopsy as clinically indicated.

Complications

Sarcoidosis is usually a self-limited, non-life-threatening disease. However, organ-threatening disease can occur. These complications can include blindness, paraplegia, or renal failure. Death from sarcoidosis occurs in about 5% of patients seen in sarcoidosis referral clinics. The usual causes of death related to sarcoidosis are from lung, cardiac, neurologic, or liver involvement. In respiratory

failure, an elevation of the right atrial pressure is a poor prognostic finding. Lung complications can also include infections such as mycetoma, which can subsequently lead to massive bleeding. In addition, the use of immunosuppressive agents can increase the incidence of serious infections.

LABORATORY FINDINGS

The chest roentgenogram remains the most commonly used tool to assess lung involvement in sarcoidosis. As noted above, the chest roentgenogram classifies involvement into four stages, with stages 1 and 2 having hilar and paratracheal adenopathy. The CT scan has been used increasingly in evaluating interstitial lung disease. In sarcoidosis, the presence of adenopathy and a nodular infiltrate is not specific for sarcoidosis. Adenopathy up to 2 cm can be seen in other inflammatory lung diseases such as idiopathic pulmonary fibrosis. However, adenopathy >2 cm in the short axis supports the diagnosis of sarcoidosis over other interstitial lung diseases.

The positive emission tomography (PET) scan has increasingly replaced gallium 67 scanning to identify areas of sarcoidosis in the chest and other parts of the body. Both tests can be used to identify potential areas for biopsy. Cardiac PET scanning has also proved useful in assessing cardiac sarcoidosis. A positive PET scan may be due to the granulomas from sarcoidosis and not to disseminated malignancy.

Serum levels of angiotensin-converting enzyme (ACE) can be helpful in the diagnosis of sarcoidosis. However, the test has somewhat low sensitivity and specificity. Elevated levels of ACE are reported in 60% of patients with acute disease and only 20% of patients with chronic disease. Although there are several causes for mild elevation of ACE, including diabetes, elevations of >50% of the upper limit of normal are seen in only a few conditions including sarcoidosis, leprosy, Gaucher's disease, hyperthyroidism, and disseminated granulomatous infections such as miliary tuberculosis. There is an insertion/deletion (I/D) polymorphism of the *ACE* gene on what is felt to be in the noncritical part of the gene. There is a phenotypic difference for ACE levels, with II polymorphism having the lowest and DD polymorphism the highest levels of ACE for both sarcoidosis patients and healthy controls. There is no clear-cut association between ACE phenotype and clinical manifestation of disease. Because the ACE level is determined by a biologic assay, the concurrent use of an ACE inhibitor such as lisinopril will lead to a very low ACE level.

DIAGNOSIS

The diagnosis of sarcoidosis requires both compatible clinical features and pathologic findings. Since the cause of sarcoidosis remains elusive, the diagnosis cannot be made with 100% certainty. Nevertheless, the diagnosis can be made with reasonable certainty based on history and physical features along with laboratory and pathologic findings.

Patients are usually evaluated for possible sarcoidosis based on two scenarios **(Fig. 329-8)**. In the first scenario, a patient may undergo a biopsy revealing a noncaseating granuloma in either a pulmonary or an extrapulmonary organ. If the clinical presentation is consistent with sarcoidosis and there is no alternative cause for the granulomas identified, then the patient is felt to have sarcoidosis.

Figure 329-8

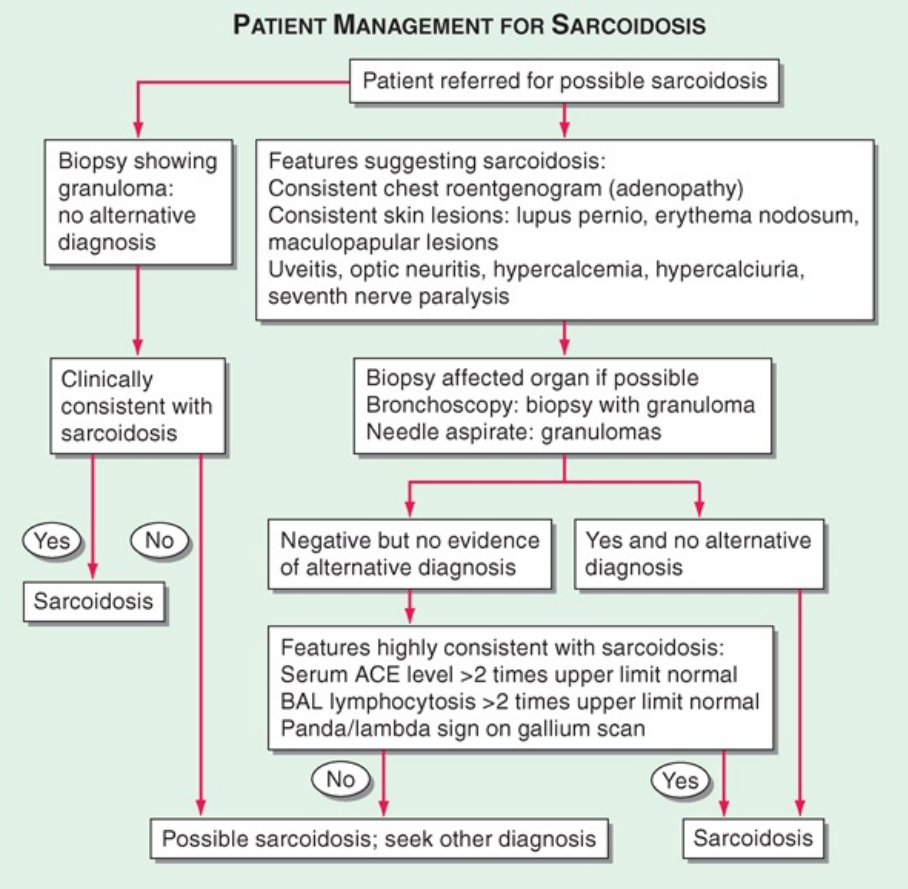

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Proposed approach to management of patient with possible sarcoidosis. Presence of one or more of these features supports the diagnosis of sarcoidosis: uveitis, optic neuritis, hypercalcemia, hypercalciuria, seventh cranial nerve paralysis, diabetes insipidus.

In the second scenario, signs or symptoms suggesting sarcoidosis such as the presence of bilateral adenopathy may be present in an otherwise asymptomatic patient or a patient with uveitis or a rash consistent with sarcoidosis. At this point, a diagnostic procedure should be performed. For the patient with a compatible skin lesion, a skin biopsy should be considered. Other biopsies to consider could include liver, extrathoracic lymph node, or muscle. In some cases, a biopsy of the affected organ may not be easy to perform (such as a brain or spinal cord lesion). In other cases, such as an endomyocardial biopsy, the likelihood of a positive biopsy is low. Because of the high rate of pulmonary involvement in these cases, the lung may be easier to approach by bronchoscopy. During the bronchoscopy, a transbronchial biopsy, bronchial biopsy, or transbronchial needle aspirate of an enlarged mediastinal lymph node can be performed. The endobronchial ultrasonography-guided transbronchial needle aspirate may be particularly useful in the patient with stage 1 disease (i.e., adenopathy without infiltrates).

If the biopsy reveals granulomas, an alternative diagnosis such as infection or malignancy must be excluded. Bronchoscopic washings can be sent for cultures for fungi and tuberculosis. For the pathologist, the more tissue that is provided, the more comfortable is the diagnosis of sarcoidosis. A needle aspirate may be adequate in an otherwise classic case of sarcoidosis, but may be insufficient in a patient in whom lymphoma or fungal infection is a likely alternative diagnosis. Since granulomas can be seen on the edge of a lymphoma, the presence of a few granulomas from a needle aspirate may not be sufficient to clarify the diagnosis. Mediastinoscopy remains the procedure of choice to confirm the presence or absence of lymphoma in the mediastinum. Alternatively, for most patients, evidence of extrathoracic disease (e.g., eye involvement) may further support the diagnosis of sarcoidosis.

For patients with negative pathology, positive supportive tests may increase the likelihood of the diagnosis of sarcoidosis. These tests include an elevated ACE level, which can also be elevated in other granulomatous diseases but not in malignancy. A positive gallium scan can support the diagnosis if increased activity is noted in the parotids and lacrimal glands (*panda sign*) or in the right paratracheal and left hilar area (*lambda sign*). A BAL is often performed during the bronchoscopy. An increase in the percentage of lymphocytes supports the diagnosis of sarcoidosis. The use of the lymphocyte markers CD4 and CD8 can be used to determine the CD4/CD8 ratio of these increased

lymphocytes in the BAL fluid. A ratio of >3.5 is strongly supportive of sarcoidosis but is less sensitive than an increase in lymphocytes alone. Although in general, an increase in BAL lymphocytes is supportive of the diagnosis, other conditions must be considered.

These supportive tests when combined with commonly associated clinical features of the disease, which are not diagnostic of sarcoidosis, can enhance the diagnostic probability. These nondiagnostic features include uveitis, renal stones, hypercalcemia, seventh cranial nerve paralysis, or erythema nodosum.

The *Kviem-Siltzbach procedure* is a specific diagnostic test for sarcoidosis. An intradermal injection of specially prepared tissue derived from the spleen of a known sarcoidosis patient is biopsied 4–6 weeks after injection. If noncaseating granulomas are seen, this is highly specific for the diagnosis of sarcoidosis. Unfortunately, there is no commercially available Kviem-Siltzbach reagent, and some locally prepared batches have lower specificity. Thus, this test is of historic interest and is rarely used in current clinical practice.

Because the diagnosis of sarcoidosis can never be certain, over time other features may arise that lead to an alternative diagnosis. Conversely, evidence for new organ involvement may eventually confirm the diagnosis of sarcoidosis.

PROGNOSIS

The risk of death or loss of organ function remains low in sarcoidosis. Poor outcomes usually occur in patients who present with advanced disease in whom treatment seems to have little impact. In these cases, irreversible fibrotic changes have frequently occurred.

For the majority of patients, initial presentation occurs during the granulomatous phase of the disease as depicted in Fig. 329-1. It is clear that many patients resolve their disease within 2–5 years. These patients are felt to have acute, self-limiting sarcoidosis. However, there is a form of the disease that does not resolve within the first 2–5 years. These chronic patients can be identified at presentation by certain risk factors at presentation such as fibrosis on chest roentgenogram, presence of lupus pernio, bone cysts, cardiac or neurologic disease (except isolated seventh nerve paralysis), and presence of renal calculi due to hypercalciuria. Recent studies also indicate that patients who require glucocorticoids for any manifestation of their disease in the first 6 months of presentation have a >50% chance of having chronic disease. In contrast, <10% of patients who require no systemic therapy in the first 6 months will require chronic therapy.

Treatment: Sarcoidosis

The indications for therapy should be based on symptoms. The patient with elevated liver function tests or an abnormal chest roentgenogram probably does not benefit from treatment. However, these patients should be monitored for evidence of progressive, symptomatic disease.

One approach to therapy is summarized in **Figs. 329-9 and 329-10**. We have divided the approach into treating acute versus chronic disease. For acute disease, no therapy remains a viable option for patients with no or mild symptoms. For symptoms confined to only one organ, topical therapy is preferable. For multiorgan disease or disease too extensive for topical therapy, an approach to systemic therapy is outlined. Glucocorticoids remain the drugs of choice for this disease. However, the decision to continue to treat with glucocorticoids or to add steroid-sparing agents depends on the tolerability, duration, and dosage of glucocorticoids. **Table 329-2** summarizes the dosage and monitoring of several commonly used drugs. According to the available trials, evidence-based recommendations are made. Most of these recommendations are for pulmonary disease because most of the trials were performed only in pulmonary disease. Treatment recommendations for extrapulmonary disease are usually similar with a few modifications. For example, the dosage of glucocorticoids is usually higher for neurosarcoidosis and lower for cutaneous disease. There was some suggestion that higher doses would be beneficial for cardiac sarcoidosis, but one study found that initial doses >40 mg/d prednisone were associated with a worse outcome because of toxicity.

Figure 329-9

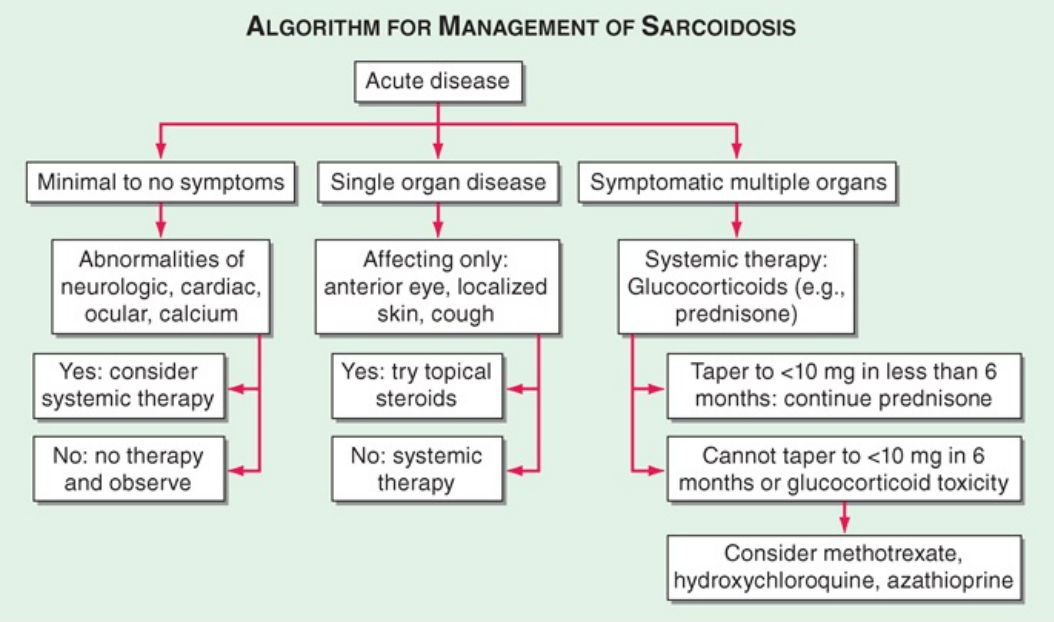

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The management of acute sarcoidosis is based on level of symptoms and extent of organ involvement. In patients with mild symptoms, no therapy may be needed unless specified manifestations are noted.

Figure 329-10

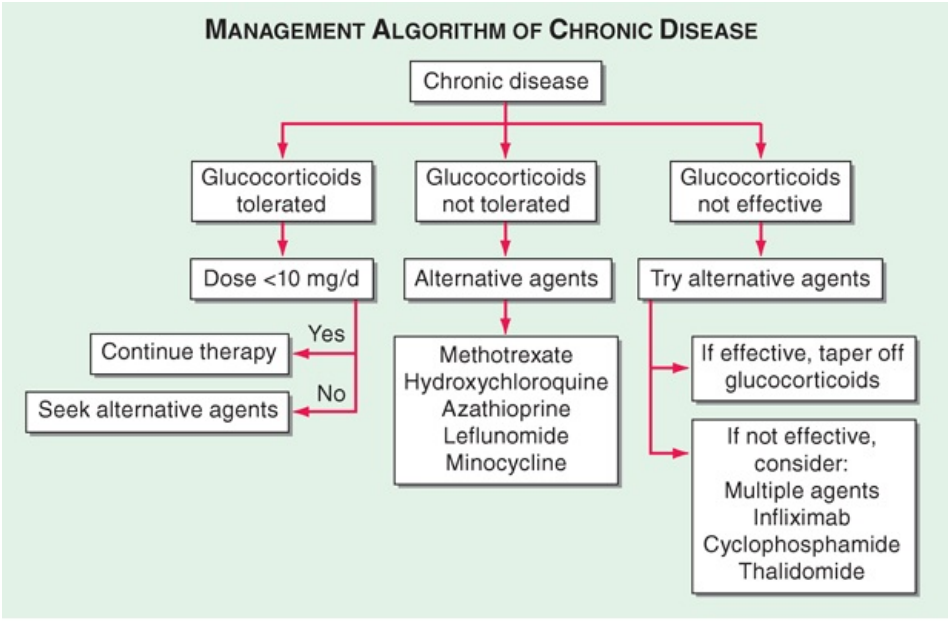

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to chronic disease is based on whether glucocorticoid therapy is tolerated or not.

Table 329-2 Commonly Used Drugs to Treat Sarcoidosis

Drug	Initial Dose	Maintenance Dose	Monitoring	Toxicity	Support Therapy[a]	Support Monitoring[a]
Prednisone	20–40 mg qd	Taper to 5–10 mg	Glucose, blood pressure, bone	Diabetes, osteoporosis	A: Acute pulmonary D:	

			bone density		Extrapulmonary	
Hydroxychloroquine	200 –400 mg qd	400 mg qd	Eye exam q6–12 mo	Ocular	B: Some forms of disease	D: Routine eye exam
Methotrexate	10 mg qw	2.5–15 mg qw	CBC, renal, hepatic q2mo	Hematologic, nausea, hepatic, pulmonary	B: Steroid sparing C: Some forms chronic disease	D: Routine hematologic, renal, and hepatic monitoring
Azathioprine	50–150 mg qd	50–200 mg qd	CBC, renal q2mo	Hematologic, nausea	C: Some forms chronic disease	D: Routine hematologic monitoring
Infliximab	3–5 mg/kg q2wk for 2 doses	3–10 mg/kg q4–8 wk	Initial PPD	Infections, allergic reaction, carcinogen	A: Chronic pulmonary disease	B: Caution in patients with latent tuberculosis or advanced congestive heart failure

[a]Grade A: supported by at least two double-blind randomized control trials; grade B: supported by prospective cohort studies; grade C: supported primarily by two or more retrospective studies; grade D: only one retrospective study or based on experience in other diseases.

Abbreviations: CBC, complete blood count; PPD, purified protein derivative test for tuberculosis.

Source: Adapted from Baughman and Selroos.

While most patients receive glucocorticoids as their initial systemic therapy, toxicity associated with prolonged therapy often leads to steroid-sparing alternatives. The antimalarial drugs such as hydroxychloroquine are more effective for skin than pulmonary disease. Minocycline may also be useful for cutaneoussarcoidosis. For pulmonary and other extra-pulmonary disease, cytotoxic agents are often employed. These include methotrexate, azathioprine, chlorambucil, and cyclophosphamide. The most widely studied cytotoxic agent has been methotrexate. This agent works in approximately two-thirds of sarcoidosis patients, regardless of the disease manifestation. As noted in Table 329-2, specific guidelines for monitoring therapy have been recommended. Cytokine modulators such as thalidomide and pentoxifylline have also been used in a limited number of cases.

The anti-TNF agents have recently been studied in sarcoidosis, with prospective randomized trials of both etanercept and infliximab completed. Etanercept has a limited role as a steroid-sparing agent. Conversely, infliximab significantly improved lung function when given to patients with chronic disease already on glucocorticoids and cytotoxic agents. The difference in response for these two agents is similar to that observed in Crohn's disease, where infliximab is effective and etanercept is not. In addition, there is a higher risk for reactivation of tuberculosis with infliximab compared to etanercept. The differential response rate could be explained by differences in mechanism of action since etanercept is a TNF receptor antagonist and infliximab is a monoclonal antibody against TNF. In contrast to etanercept, infliximab also binds to TNF on the surface of some cells that are releasing TNF and this can lead to cell lysis. This effect has been documented in Crohn's disease. There is currently limited information about the dose and effectiveness of adalimumab, another anti-TNF antibody, versus infliximab in sarcoidosis. The role of the newer therapeutic agents for sarcoidosis is still evolving. However, these targeted therapies confirm that TNF may be an important target, especially in the treatment of chronic disease. However, these agents are not a panacea, since sarcoidosis-like disease has occurred in patients treated with anti-TNF agents for non-sarcoidosis indications.

FURTHER READINGS

Baughman RP, Selroos O: Evidence-based approach to treatment of sarcoidosis, in PG Gibson et al (eds): *Evidence-Based Respiratory Medicine*. Oxford, BMJ Books Blackwell, 2005, pp 491–508

de Kleijn WP et al: Fatigue in sarcoidosis: a systematic review. Curr Opin Pulm Med 15:499, 2009

Drake WP et al: Cellular recognition of Mycobacterium tuberculosis ESAT-6 and KatG peptides in systemic sarcoidosis. Infect Immun 75:527, 2007[PMID: 17088357] [Full Text]

Grunewald J, Eklund A: Löfgren's syndrome: Human leukocyte antigen strongly influences the disease course. Am J Respir Crit Care Med 179:307, 2009[PMID: 18996998] [Full Text]

Hunninghake GW et al: ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and other Granulomatous Disorders. Sarcoidosis Vasc Diffuse Lung Dis 16:149, 1999[PMID: 10560120] [Full Text]

Iannuzzi MC et al: Sarcoidosis. N Engl J Med 357:2153, 2007[PMID: 18032765] [Full Text]

Judson MA et al: Efficacy of infliximab in extrapulmonary sarcoidosis: Results from a randomised trial. Eur Respir J 31:1189, 2008[PMID: 18256069] [Full Text]

Newman LS et al: A case control etiologic study of sarcoidosis: Environmental and occupational risk factors. Am J Respir Crit Care Med 170:1324, 2004[PMID: 15347561] [Full Text]

Rossman MD et al: HLA-DRB1*1101: A significant risk factor for sarcoidosis in blacks and whites. Am J Hum Genet 73:720, 2003[PMID: 14508706] [Full Text]

Ziegenhagen MW et al: Exaggerated TNF-alpha release of alveolar macrophages in corticosteroid resistant sarcoidosis. Sarcoidosis Vasc Diffuse Lung Dis 19:185, 2002[PMID: 12405487] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

FAMILIAL MEDITERRANEAN FEVER AND OTHER HEREDITARY RECURRENT FEVERS: INTRODUCTION

Familial Mediterranean fever (FMF) is the prototype of a group of inherited diseases **(Table 330-1)** that are characterized by recurrent episodes of fever with serosal, synovial, or cutaneous inflammation and, in some individuals, the eventual development of systemic AA amyloidosis **(Chap. 112)**. Because of the relative infrequency of high-titer autoantibodies or antigen-specific T cells, the term *autoinflammatory* has been proposed to describe these disorders, rather than autoimmune. The innate immune system, with its myeloid effector cells and germline receptors for pathogen-associated molecular patterns and endogenous danger signals, plays a predominant role in the pathogenesis of the autoinflammatory diseases.

Table 330-1 The Hereditary Recurrent Fever Syndromes

	FMF	TRAPS	HIDS	MWS	FCAS	NOMID
Ethnicity	Jewish, Arab, Turkish, Armenian, Italian	Any ethnic group	Predominantly Dutch, northern European	Any ethnic group	Any ethnic group	Any ethnic group
Inheritance	Recessive[a]	Dominant	Recessive	Dominant	Dominant	Usually de novo mutations
Gene/chromosome	*MEFV*/16p13.3	*TNFRSF1A*/12p13	*MVK*/12q24	*NLRP3*/1q44	*NLRP3*/1q44	*NLRP3*/1q44
Protein	Pyrin	p55 TNF receptor	Mevalonate kinase	Cryopyrin	Cryopyrin	Cryopyrin
Attack length	1–3 days	Often > 7 days	3–7 days	1–2 days	Minutes–3 days	Continuous, with flares
Serosa	Pleurisy, peritonitis; asymptomatic pericardial effusions	Pleurisy, peritonitis, pericarditis	Abd pain, but seldom peritonitis; pleurisy, pericarditis uncommon	Abd pain common; pleurisy, pericarditis rare	Rare	Rare
Skin	Erysipeloid erythema	Centrifugally migrating erythema	Diffuse maculopapular rash; oral ulcers	Diffuse urticaria-like rash	Cold-induced urticaria-like rash	Diffuse urticaria-like rash
Joints	Acute monoarthritis; chronic hip arthritis (rare)	Acute monoarthritis, arthralgia	Arthralgia, oligoarthritis	Arthralgia, large joint oligoarthritis	Polyarthralgia	Epiphyseal, patellar overgrowth, clubbing
Muscle	Exercise-induced myalgia common; protracted febrile myalgia rare	Migratory myalgia	Uncommon	Myalgia common	Sometimes myalgia	Sometimes myalgia
Eyes, ears	Uncommon	Periorbital edema, conjunctivitis, rarely uveitis	Uncommon	Conjunctivitis, episcleritis, optic disc edema; sensorineural hearing loss	Conjunctivitis	Conjunctivitis, uveitis, optic disc edema, blindness, sensorineural hearing loss
CNS	Aseptic meningitis rare	Headache	Headache	Headache	Headache	Aseptic meningitis, seizures
Amyloidosis	Most common in M694V homozygotes	~15% of cases	Uncommon	~25% of cases	Uncommon	Late complication
Treatment	Oral colchicine prophylaxis	Glucocorticoids, etanercept, anakinra (IL-1 receptor	NSAIDs for fever; IL-1β and TNF inhibitors	Anakinra, rilonacept, canakinumab	Anakinra, rilonacept, canakinumab	Anakinra

	antagonist)	investigational			

[a]A substantial percentage of patients with clinical FMF have only a single demonstrable *MEFV* mutation on DNA sequencing.

Abbreviations: FCAS, familial cold autoinflammatory syndrome; FMF, familial Mediterranean fever; HIDS, hyperimmunoglobulinemia D with periodic fever syndrome; IL, interleukin; MWS, Muckle-Wells syndrome; NOMID, neonatal-onset multisystem inflammatory disease; NSAIDs, nonsteroidal anti-inflammatories; TNF, tumor necrosis factor; TRAPS, TNF receptor-associated periodic syndrome.

BACKGROUND AND PATHOPHYSIOLOGY

FMF was first recognized among Armenians, Arabs, Turks, and non-Ashkenazi (primarily North African and Iraqi) Jews. With the advent of genetic testing, FMF has been documented with increasing frequency among Ashkenazi Jews, Italians, and other Mediterranean populations, and occasional cases have been confirmed even in the absence of known Mediterranean ancestry. FMF is recessively inherited, but, particularly in countries where families are small, a positive family history can only be elicited in ~50% of cases. DNA testing demonstrates carrier frequencies as high as 1:3 among affected populations, suggesting a heterozygote advantage.

The FMF gene encodes a 781-amino acid, ~95 kDa protein denoted *pyrin* (or *marenostrin*) that is expressed in granulocytes, eosinophils, monocytes, dendritic cells, and synovial and peritoneal fibroblasts. The N-terminal 92 amino acids of pyrin define a motif, the PYRIN domain, that is similar in structure to death domains, death effector domains, and caspase recruitment domains. PYRIN domains mediate homotypic protein-protein interactions and have been found in several other proteins, including cryopyrin, which is mutated in three other recurrent fever syndromes. Through a number of mechanisms, including the interaction of the PYRIN domain with an intermediary adaptor protein, pyrin regulates caspase-1 [interleukin (IL) 1β-converting enzyme], and thereby IL-1β secretion. Mice bearing FMF-associated pyrin mutations exhibit inflammation and excessive IL-1 production.

ACUTE ATTACKS

Febrile episodes in FMF may begin even in early infancy; 90% of patients have had their first attack by age 20. Typical FMF episodes generally last 24–72 hours, with arthritic attacks tending to last somewhat longer. In some patients the episodes occur with great regularity, but more often the frequency of attacks varies over time, ranging from as often as once every few days to remissions lasting several years. Attacks are often unpredictable, although some patients relate them to physical exertion, emotional stress, or menses; pregnancy may be associated with remission.

If measured, fever is nearly always present throughout FMF attacks. Severe hyperpyrexia and even febrile seizures may be seen in infants, and fever is sometimes the only manifestation of FMF in young children.

Over 90% of FMF patients experience abdominal attacks at some time. Episodes range in severity from dull, aching pain and distention with mild tenderness on direct palpation to severe generalized pain with absent bowel sounds, rigidity, rebound tenderness, and air-fluid levels on upright radiographs. CT scanning may demonstrate a small amount of fluid in the abdominal cavity. If such patients undergo exploratory laparotomy, a sterile, neutrophil-rich peritoneal exudate is present, sometimes with adhesions from previous episodes. Ascites is rare.

Pleural attacks are usually manifested by unilateral, sharp, stabbing chest pain. Radiographs may show atelectasis and sometimes an effusion. If performed, thoracentesis demonstrates an exudative fluid rich in neutrophils. After repeated attacks, pleural thickening may develop.

FMF arthritis is most frequent among individuals homozygous for the M694V mutation, which is especially common in the non-Ashkenazi Jewish population. Acute arthritis in FMF is usually monoarticular, affecting the knee, ankle, or hip, although other patterns can be seen, particularly in children. Large sterile effusions rich in neutrophils are frequent, without commensurate erythema or warmth. Even after repeated arthritic attacks, radiographic changes are rare. Before the advent of colchicine prophylaxis, chronic arthritis of the knee or hip were seen in ~5% of FMF patients with arthritis. Chronic sacroiliitis can occur in FMF irrespective of the HLA-B27 antigen, even in the face of colchicine therapy. In the United States, FMF patients are much more likely to have arthralgia than arthritis.

The most characteristic cutaneous manifestation of FMF is erysipelas-like erythema, a raised erythematous rash that most commonly occurs on the dorsum of the foot, ankle, or lower leg alone or in combination with abdominal pain, pleurisy, or arthritis. Biopsy demonstrates perivascular infiltrates of granulocytes and monocytes. This rash is seen most often in M694V homozygotes and is relatively rare in the United States.

Exercise-induced (nonfebrile) myalgia is common in FMF, and a small percentage of patients develop a protracted febrile myalgia that can last several weeks. Symptomatic pericardial disease is rare, although some patients have small pericardial effusions as an incidental echocardiographic finding. Unilateral acute scrotal inflammation may occur in prepubertal boys. Aseptic meningitis has been reported in FMF, but the causal connection is controversial. Vasculitis, including Henoch-Schönlein purpura and polyarteritis nodosum **(Chap. 326)** may be seen at increased frequency in FMF.

Laboratory features of FMF attacks are consistent with acute inflammation and include an elevated erythrocyte sedimentation rate, leukocytosis, thrombocytosis (in children), and elevations in C-reactive protein, fibrinogen, haptoglobin, and serum immunoglobulins. Transient albuminuria and hematuria may also be seen.

AMYLOIDOSIS

Before the advent of colchicine prophylaxis, systemic amyloidosis was a common complication of FMF. It is caused by deposition of a fragment of serum amyloid A, an acute-phase reactant, in the kidneys, adrenals, intestine, spleen, lung, and testes **(Chap. 112)**. Amyloidosis should be suspected in patients who have proteinuria between attacks; renal or rectal biopsy are used most often to establish the diagnosis. Risk factors include the M694V homozygousgenotype, positive family history (independent of FMF mutational status), the SAA 1 genotype, male gender, noncompliance with colchicine therapy, and having grown up in the Middle East.

Diagnosis

For typical cases, physicians experienced with FMF can often make the diagnosis on clinical grounds alone. Clinical

criteria sets for FMF have been shown to have high sensitivity and specificity in parts of the world where the pretest probability of FMF is high. Genetic testing can provide a useful adjunct in ambiguous cases or for physicians not experienced in FMF. Most of the more severe disease-associated FMF mutations are in exon 10 of the gene, with a smaller group of milder variants in exon 2. An updated list of mutations for FMF and other hereditary recurrent fevers can be found online at http://fmf.igh.cnrs.fr/infevers/.

Genetic testing has permitted a broadening of the clinical spectrum and geographic distribution of FMF and may be of prognostic value. Most studies indicate that M694V homozygotes have an earlier age of onset and a higher frequency of arthritis, rash, and amyloidosis. In contrast, the E148Q variant is usually associated with milder disease. E148Q is sometimes found in *cis* with exon 10 mutations, which complicates the interpretation of genetic test results. Only ~70% of patients with clinically typical FMF have two identifiable mutations in *trans*, suggesting either that current screening methods do not detect all of the relevant mutations or that one mutation may be sufficient to cause disease under some circumstances. In these cases clinical judgment is very important, and sometimes a therapeutic trial of colchicine may help to confirm the diagnosis. Genetic testing of unaffected individuals is usually inadvisable, because of the possibility of nonpenetrance and the potential impact of a positive test on future insurability.

If a patient is seen during his or her first attack, the differential diagnosis may be broad, although delimited by the specific organ involvement. After several attacks the differential diagnosis may include the other hereditary recurrent fever syndromes (Table 330-1); the syndrome of periodic fever with aphthous ulcers, pharyngitis, and cervical adenopathy (PFAPA); systemic-onset juvenile rheumatoid arthritis or adult Still's disease; porphyria; hereditary angioedema; inflammatory bowel disease; and, in women, gynecologic disorders.

Treatment: Familial Mediterranean Fever

The treatment of choice for FMF is daily oral colchicine, which decreases the frequency and intensity of attacks and prevents the development of amyloidosis in compliant patients. Intermittent dosing at the onset of attacks is not as effective as daily prophylaxis and is of unproven value in preventing amyloidosis. The usual adult dose of colchicine is 1.2–1.8 mg/d, which causes substantial reduction in symptoms in two-thirds of patients and some improvement in >90%. Children may require lower doses, although not proportionately to body weight.

Common side effects of colchicine include bloating, abdominal cramps, lactose intolerance, and diarrhea. They can be minimized by starting at a low dose and gradually advancing as tolerated, splitting the dose, use of simethecone for flatulence, and avoidance of dairy products. If taken by either parent at the time of conception, colchicine may cause a small increase in the risk of trisomy 21 (Down syndrome). In elderly patients with renal insufficiency, colchicine can cause a myoneuropathy characterized by proximal muscle weakness and elevation of the creatine kinase. Cyclosporine inhibits hepatic excretion of colchicine by its effects on the MDR-1 transport system, sometimes leading to colchicine toxicity in patients who have undergone renal transplantation for amyloidosis. Intravenous colchicine should generally not be administered to patients already taking oral colchicine, because severe, sometimes fatal, toxicity can occur in this setting.

There are no established alternatives for the small number of patients who do not respond to colchicine or cannot tolerate therapeutic dosages, although the IL-1 receptor antagonist and inhibitors of tumor necrosis factor (TNF) are investigational. Bone marrow transplantation has been suggested for refractory FMF, but the risk-benefit ratio is currently regarded as unacceptable.

OTHER HEREDITARY RECURRENT FEVERS

Within 5 years of the discovery of the FMF gene, three additional genes causing five other hereditary recurrent fever syndromes were identified, catalyzing a paradigm shift in diagnosis and treatment of these disorders.

Tnf Receptor–Associated Periodic Syndrome (Traps)

TRAPS is caused by dominantly inherited mutations in the extracellular domains of the 55-kDa TNF receptor (TNFRSF1A, p55). Although originally described in a large Irish family (and hence the name *familial Hibernian fever*), TRAPS has a broad ethnic distribution. TRAPS episodes often begin in childhood. The duration of attacks ranges from 1–2 days to as long as several weeks, and in severe cases symptoms may be nearly continuous. In addition to peritoneal, pleural, and synovial attacks similar to FMF, TRAPS patients frequently have ocular inflammation (most often conjunctivitis and/or periorbital edema), and a distinctive migratory myalgia with overlying painful erythema may be present. TRAPS patients generally respond better to glucocorticoids than to prophylactic colchicine. About 15% develop amyloidosis. The diagnosis of TRAPS is based on the demonstration of *TNFRSF1A* mutations in the presence of characteristic symptoms. Leukocytes from patients with certain TRAPS mutations exhibit a defect in TNF receptor-shedding, possibly impairing normal homeostasis. However, a more complex picture is emerging, with a number of functional abnormalities, some of which are ligand-independent, contributing to the autoinflammatory phenotype. Etanercept, a TNF inhibitor, ameliorates TRAPS attacks, although its effect on amyloidosis is unproven. Perhaps because of the ligand-independent signaling abnormalities in TRAPS, IL-1 inhibition has been beneficial in some patients.

Hyperimmunoglobulinemia D with Periodic Fever Syndrome (Hids)

HIDS is a recessively inherited recurrent fever syndrome found primarily in individuals of northern European ancestry. It is caused by mutations in mevalonate kinase (*MVK*), encoding an enzyme involved in the synthesis of cholesterol and nonsterol isoprenoids. Attacks usually begin in infancy, and last 3–5 days. Clinically distinctive features include painful cervical adenopathy, a diffuse maculopapular rash sometimes affecting the palms and soles, and aphthous ulcers; pleurisy is rare, as is amyloidosis. Although originally defined by the persistent elevation of serum IgD, disease activity is not related to IgD levels, and some patients with FMF or TRAPS may have modestly increased serum IgD. Moreover, occasional patients with *MVK* mutations and recurrent fever have normal IgD levels. All patients with mutations have markedly elevated urinary mevalonate levels during their febrile attacks, although the inflammatory manifestations are likely to be due to a deficiency of isoprenoids rather than an excess of mevalonate. There is currently no established treatment for HIDS, although intermittent or continuous IL-1 inhibition are investigational.

The Cryopyrinopathies, or Cryopyrin-Associated Periodic Syndromes (Caps)

Three hereditary febrile syndromes, familial cold autoinflammatory syndrome (FCAS), Muckle-Wells syndrome (MWS), and neonatal-onset multisystem inflammatory disease (NOMID), are all caused by mutations in *NLRP3* (formerly

known as *CIAS1)*, the gene encoding cryopyrin (or NLRP3), and represent a clinical spectrum of disease. FCAS patients develop chills, fever, headache, arthralgia, conjunctivitis, and an urticaria-like rash in response to generalized cold exposure. In MWS, an urticarial rash is noted, but it is not usually induced by cold; MWS patients also develop fevers, abdominal pain, limb pain, arthritis, conjunctivitis, and, over time, sensorineural hearing loss. NOMID is the most severe of the three disorders, with chronic aseptic meningitis, a characteristic arthropathy, and rash. Like the FMF protein, pyrin, cryopyrin has an N-terminal PYRIN domain. Cryopyrin regulates IL-1β production through the formation of a macromolecular complex termed the *inflammasome*. Peripheral blood leukocytes from patients with FCAS, MWS, and NOMID release increased amounts of IL-1β upon in vitro stimulation, relative to healthy controls. Macrophages from cryopyrin-deficient mice exhibit decreased IL-1β production in response to certain gram-positive bacteria, bacterial RNA, and monosodium urate crystals. Patients with all three cryopyrino-pathies show a dramatic response to injections of IL-1 inhibitors. Increased IL-1 signaling is also a feature of the recently-described deficiency in the interleukin-1 receptor antagonist (DIRA), a recessively inherited disorder similarly responsive to anakinra treatment. In contrast to the cryopyrinopathies, DIRA presents with pustular skin lesions and multifocal sterile osteomyelitis, and fever is often not a prominent feature.

FURTHER READINGS

Aksentijevich I et al: An autoinflammatory disease with deficiency of the interleukin-1-receptor antagonist. N Engl J Med 360:2426, 2009[PMID: 19494218] [Full Text]

Dinarello CA: Interleukin-1α and the autoinflammatory diseases. N Engl J Med 360:2467, 2009[PMID: 19494224] [Full Text]

Goldbach-Manksy R et al: Neonatal-onset multisystem inflammatory disease responsive to interleukin-1β inhibition. N Engl J Med 355:581, 2006

Hoffman HM et al: Prevention of cold-associated acute inflammation in familial cold autoinflammatory syndrome by interleukin-1 receptor antagonist. Lancet 364:1779, 2004[PMID: 15541451] [Full Text]

Hull KM et al: The TNF receptor–associated periodic syndrome (TRAPS). Emerging concepts of an autoinflammatory disorder. Medicine (Baltimore) 81:349, 2002[PMID: 12352631] [Full Text]

Kastner DL et al: Autoinflammatory disease reloaded: a clinical perspective. Cell 140:784, 2010[PMID: 20303869] [Full Text]

Kuijk LM et al: HMG-CoA reductase inhibition induces IL-1β release through Rac1/PI3K/PKB-dependent caspase-1 activation. Blood 112:3563, 2008[PMID: 18684863] [Full Text]

Lachmann HJ et al: In vivo regulation of interleukin 1β in patients with cryopyrin-associated periodic syndromes. J Exp Med 206:1029, 2009[PMID: 19364880] [Full Text]

Masters SL et al: *Horror autoinflammaticus:* The molecular pathophysiology of autoinflammatory disease. Annu Rev Immunol 27:621, 2009[PMID: 19302049] [Full Text]

Schroder K, Tschopp J: The inflammasomes. Cell 140: 140:821, 2010[PMID: 20303873] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO ARTICULAR AND MUSCULOSKELETAL DISORDERS: INTRODUCTION

Musculoskeletal complaints account for >315 million outpatient visits per year and nearly 20% of all outpatient visits in the United States. The Centers for Disease Control and Prevention estimate that 22% (46 million) of the U.S. population has physician-diagnosed arthritis and 19 million have significant functional limitation. While many patients will have self-limited conditions requiring minimal evaluation and only symptomatic therapy and reassurance, specific musculoskeletal presentations or their persistence may herald a more serious condition that requires further evaluation or laboratory testing to establish a diagnosis. The goal of the musculoskeletal evaluation is to formulate a differential diagnosis that leads to an accurate diagnosis and timely therapy, while avoiding excessive diagnostic testing and unnecessary treatment **(Table 331-1)**. There are several urgent conditions that must be diagnosed promptly to avoid significant morbid or mortal sequelae. These "red flag" diagnoses include septic arthritis, acute crystal-induced arthritis (e.g., gout), and fracture. Each may be suspected by its acute onset and monarticular or focal musculoskeletal pain (see below).

Table 331-1 Evaluation of Patients with Musculoskeletal Complaints

Goals
Accurate diagnosis
Timely provision of therapy
Avoidance of unnecessary diagnostic testing
Approach
Anatomic localization of complaint (articular vs. nonarticular)
Determination of the nature of the pathologic process(inflammatory vs. noninflammatory)
Determination of the extent of involvement (monarticular,polyarticular, focal, widespread)
Determination of chronology (acute vs. chronic)
Consider the most common disorders first
Formulation of a differential diagnosisentry

Individuals with musculoskeletal complaints should be evaluated with a thorough history, a comprehensive physical and musculo-skeletal examination, and, if appropriate, laboratory testing. The initial encounter should determine whether the musculoskeletal complaint signals a red flag condition (septic arthritis, gout, or fracture) or not. The evaluation should proceed to ascertain if the complaint is (1) *articular* or *nonarticular* in origin, (2) *inflammatory* or *noninflammatory* in nature, (3) *acute* or *chronic* in duration, and (4) *localized (monarticular)* or *widespread* (*polyarticular*) in distribution.

With such an approach and an understanding of the pathophysiologic processes, the musculoskeletal complaint or presentation can be characterized (e.g., acute inflammatory monarthritis or a chronic noninflammatory, nonarticular widespread pain) to narrow the diagnostic possibilities. A diagnosis can be made in the vast majority of individuals. However, some patients will not fit immediately into an established diagnostic category. Many musculoskeletal disorders resemble each other at the outset, and some may take weeks or months to evolve into a readily recognizable diagnostic entity. This consideration should temper the desire to establish a definitive diagnosis at the first encounter.

ARTICULAR VERSUS NONARTICULAR

The musculoskeletal evaluation must discriminate the anatomic origin(s) of the patient's complaint. For example, ankle pain can result from a variety of pathologic conditions involving disparate anatomic structures, including gonococcal arthritis, calcaneal fracture, Achilles tendinitis, plantar fasciitis, cellulitis, and peripheral or entrapment neuropathy. Distinguishing between articular and nonarticular conditions requires a careful and detailed examination. Articular structures include the synovium, synovial fluid, articular cartilage, intraarticular ligaments, joint capsule, and juxtaarticular bone. Nonarticular (or periarticular) structures, such as supportive extraarticular ligaments, tendons, bursae, muscle, fascia, bone, nerve, and overlying skin, may be involved in the pathologic process. Although musculoskeletal complaints are often ascribed to the joints, nonarticular disorders more frequently underlie such complaints. Distinguishing between these potential sources of pain may be challenging

to the unskilled examiner. Articular disorders may be characterized by deep or diffuse pain, pain or limited range of motion on active and passive movement, and swelling (caused by synovial proliferation, effusion, or bony enlargement), crepitation, instability, "locking," or deformity. By contrast, nonarticular disorders tend to be painful on active, but not passive (or assisted), range of motion. Periarticular conditions often demonstrate point or focal tenderness in regions adjacent to articular structures, and have physical findings remote from the joint capsule. Moreover, nonarticular disorders seldom demonstrate swelling, crepitus, instability, or deformity of the joint itself.

INFLAMMATORY VERSUS NONINFLAMMATORY DISORDERS

In the course of a musculoskeletal evaluation, the examiner should determine the nature of the underlying pathologic process and whether inflammatory or noninflammatory findings exist. Inflammatory disorders may be infectious (infection with *Neisseria gonorrhoea* or *Mycobacterium tuberculosis*), crystal-induced (gout, pseudogout), immune-related [rheumatoid arthritis (RA), systemic lupus erythematosus (SLE)], reactive (rheumatic fever, reactive arthritis), or idiopathic. Inflammatory disorders may be identified by any of the four cardinal signs of inflammation (erythema, warmth, pain, or swelling), systemic symptoms (fatigue, fever, rash, weight loss), or laboratory evidence of inflammation [elevated erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP), thrombocytosis, anemia of chronic disease, or hypoalbuminemia]. Articular stiffness commonly accompanies chronic musculoskeletal disorders and can extend beyond the joint. However, the severity and duration of stiffness may be diagnostically important. Morning stiffness related to inflammatory disorders (such as RA or polymyalgia rheumatica) is precipitated by prolonged rest, is described as severe, lasts for hours, and may improve with activity or anti-inflammatory medications. By contrast, intermittent stiffness (also known as gel phenomenon), associated with noninflammatory conditions [such as osteoarthritis (OA)], is precipitated by brief periods of rest, usually lasts less than 60 minutes, and is exacerbated by activity. Fatigue may accompany inflammation (as seen in RA and polymyalgia rheumatica) but may also be a consequence of fibromyalgia (a noninflammatory disorder), anemia, cardiac failure, endocrinopathy, poor nutrition, chronic pain, poor sleep, or depression. Noninflammatory disorders may be related to trauma (rotator cuff tear), repetitive use (bursitis, tendinitis), degeneration or ineffective repair (OA), neoplasm (pigmented villonodular synovitis), or pain amplification (fibromyalgia). Noninflammatory disorders are often characterized by pain without synovial swelling or warmth, absence of inflammatory or systemic features, daytime gel phenomena rather than morning stiffness, and normal (for age) or negative laboratory investigations.

Identification of the nature of the underlying process and the site of the complaint will enable the examiner to characterize the musculoskeletal presentation (e.g., acute inflammatory monarthritis, chronic noninflammatory, nonarticular widespread pain), narrow the diagnostic considerations, and assess the need for immediate diagnostic or therapeutic intervention or for continued observation. Figure 331-1 presents an algorithmic approach to the evaluation of patients with musculoskeletal complaints. This approach is remarkably effective and relies on clinical and historic features, rather than laboratory testing, to diagnose many common rheumatic disorders.

Figure 331-1

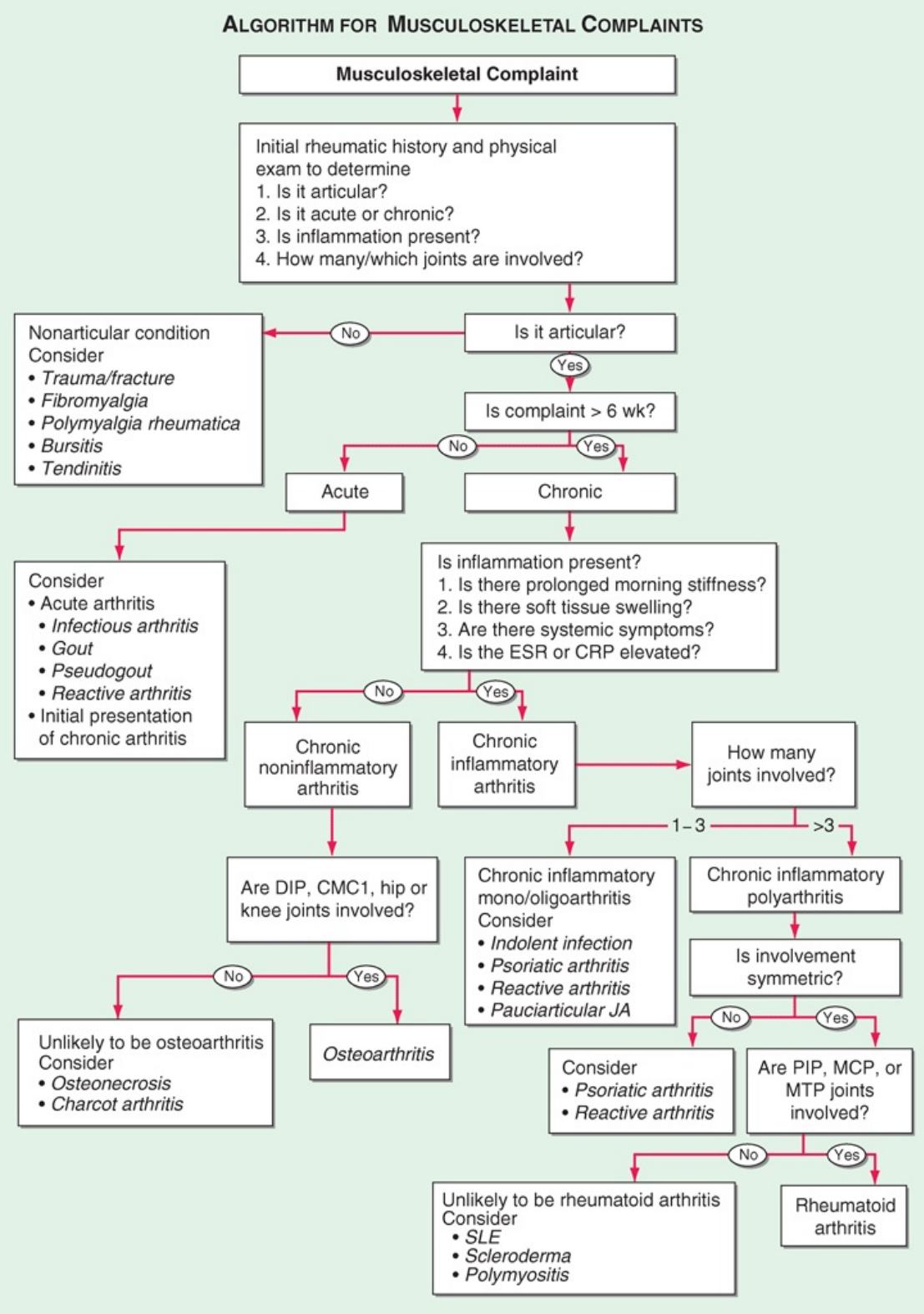

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the diagnosis of musculoskeletal complaints. An approach to formulating a differential diagnosis (shown in italics). CMC, carpometacarpal; CRP, C-reactive protein; DIP, distal interphalangeal; ESR, erythrocyte sedimentation rate; JA, juvenile arthritis; MCP, metacarpophalangeal; MTP, metatarsophalangeal; PIP, proximal interphalangeal; PMR, polymyalgia rheumatica; SLE, systemic lupus erythematosus.

The algorithmic approach may be unnecessary in patients with the most commonly encountered ailments; as these can also be considered based on frequency and characteristic presentations. The most prevalent causes of musculoskeletal complaints are shown in **Fig. 331-2**. As trauma, fracture, overuse syndromes, and fibromyalgia are among the most common causes of presentation, these

should be considered during the initial encounter. If these possibilities are excluded, other frequently occurring disorders should be considered according to the patient's age. Hence, those younger than 60 years are commonly affected by repetitive use/strain disorders, gout (men only), RA, spondyloarthritis, and uncommonly, infectious arthritis. Patients over age 60 years are frequently affected by OA, crystal (gout and pseudogout) arthritis, polymyalgia rheumatica, osteoporotic fracture, and uncommonly, septic arthritis. These conditions are between 10 and 100 times more prevalent than other serious autoimmune conditions, such as systemic lupus erythematosus, scleroderma, polymyositis, and vasculitis.

Figure 331-2

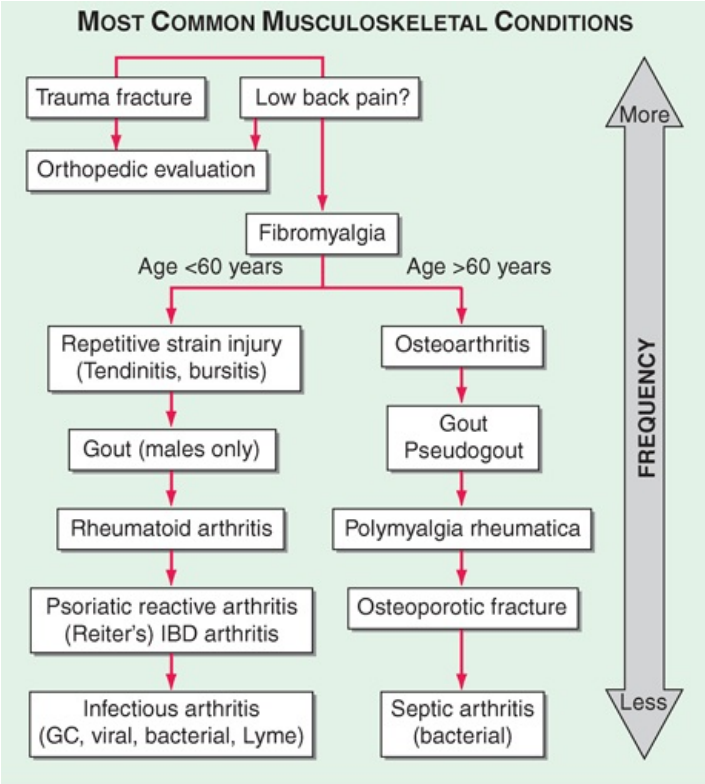

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for consideration of the most common musculoskeletal conditions. GC, gonococcal; IBD, inflammatory bowel disease.

Clinical History

Additional historic features may reveal important clues to the diagnosis. Aspects of the patient profile, complaint chronology, extent of joint involvement, and precipitating factors can provide important information. Certain diagnoses are more frequent in different *age* groups (Fig. 331-2). SLE and reactive arthritis occur more frequently in the young, whereas fibromyalgia and RA are frequent in middle age, and OA and polymyalgia rheumatica are more prevalent among the elderly. Diagnostic clustering is also evident when *sex* and *race* are considered. Gout and the spondyloarthropathies (e.g., ankylosing spondylitis) are more common in men, whereas RA, fibromyalgia, and lupus are more frequent in women. *Racial predilections* may be evident. Thus, polymyalgia rheumatica, giant cell arteritis, and Wegener's granulomatosis commonly affect whites, whereas sarcoidosis and SLE more commonly affect African Americans. *Familial aggregation* may be seen in disorders such as ankylosing spondylitis, gout, and Heberden's nodes of OA.

The chronology of the complaint is an important diagnostic feature and can be divided into the *onset*, *evolution*, and *duration*. The onset of disorders such as septic arthritis or gout tends to be abrupt, whereas OA, RA, and fibromyalgia may have more indolent presentations. The patients' complaints may evolve differently and be classified as chronic (OA), intermittent (crystal or Lyme arthritis), migratory (rheumatic fever, gonococcal or viral arthritis), or additive (RA, psoriatic arthritis). Musculoskeletal disorders are typically classified as acute or chronic based upon a symptom duration that is either less than or greater than 6 weeks, respectively. Acute arthropathies tend to be infectious, crystal-induced, or reactive. Chronic conditions include noninflammatory or immunologic arthritides (e.g., OA, RA) and nonarticular disorders (e.g., fibromyalgia).

The *extent* or *distribution* of articular involvement is often informative. Articular disorders are classified based on the number of joints involved, as either *monarticular* (one joint), *oligoarticular* or *pauciarticular* (two or three joints), or *polyarticular* (four or more joints). Although crystal and infectious arthritis are often mono- or oligoarticular, OA and RA are polyarticular disorders. Nonarticular disorders may be classified as either focal or widespread. Complaints secondary to tendinitis or carpal tunnel syndrome are typically focal, whereas weakness and myalgia, caused by polymyositis or fibromyalgia, are more diffuse in their presentation. Joint involvement in RA tends to be symmetric, whereas the spondyloarthropathies and gout are often asymmetric and oligoarticular. The upper extremities are frequently involved in RA and OA, whereas lower extremity arthritis is characteristic of reactive arthritis and gout at their onset. Involvement of the axial skeleton is common in OA and ankylosing spondylitis but is infrequent in RA, with the notable exception of the cervical spine.

The clinical history should also identify *precipitating events*, such as trauma (osteonecrosis, meniscal tear), drug administration **(Table 331-2)**, or antecedent or intercurrent illnesses (rheumatic fever, reactive arthritis, hepatitis), that may have contributed to the patient's complaint. Certain comorbidities may predispose to musculoskeletal consequences. This is especially so for diabetes mellitus (carpal tunnel syndrome), renal insufficiency (gout), psoriasis (psoriatic arthritis), myeloma (low back pain), cancer (myositis), and osteoporosis (fracture) or when using certain drugs such as glucocorticoids (osteonecrosis, septic arthritis) and diuretics or chemotherapy (gout) (Table 331-2).

Table 331-2 Drug-Induced Musculoskeletal Conditions

Arthralgias
Quinidine, cimetidine, quinolones, chronic acyclovir, interferon, IL-2, nicardipine, vaccines, rifabutin, aromatase and HIVprotease inhibitors
Myalgias/myopathy
Glucocorticoids, penicillamine, hydroxychloroquine, AZT, lovastatin, simvastatin, pravastatin, clofibrate, interferon, IL-2, alcohol, cocaine, taxol, docetaxel, colchicine, quinolones, cyclosporine, protease inhibitors
Tendon rupture/tendinitis
Quinolones, glucocorticoids, isotretinoin
Gout
Diuretics, aspirin, cytotoxics, cyclosporine, alcohol, moonshine, ethambutol
Drug-induced lupus
Hydralazine, procainamide, quinidine, phenytoin, carbamazepine, methyldopa, isoniazid, chlorpromazine, lithium, penicillamine, tetracyclines, TNF inhibitors, ACE inhibitors, ticlopidine
Osteonecrosis
Glucocorticoids, alcohol, radiation, bisphosphonates
Osteopenia
Glucocorticoids, chronic heparin, phenytoin, methotrexate
Scleroderma
Vinyl chloride, bleomycin, pentazocine, organic solvents, carbidopa, tryptophan, rapeseed oil
Vasculitis
Allopurinol, amphetamines, cocaine, thiazides, penicillamine, propylthiouracil, montelukast, TNF inhibitors, hepatitis B vaccine, trimethoprim/sulfamethoxazole

Abbreviations: ACE, angiotensin-converting enzyme; IL-2, interleukin 2; TNF, tumor necrosis factor.

Lastly, a thorough *rheumatic review of systems* may disclose useful diagnostic information. A variety of musculoskeletal disorders may be associated with systemic features such as fever (SLE, infection), rash (SLE, psoriatic arthritis), nail abnormalities (psoriatic or reactive arthritis), myalgias (fibromyalgia, statin- or drug-induced myopathy), or weakness (polymyositis, neuropathy). In addition, some conditions are associated with involvement of other organ systems including the eyes (Behçet's disease, sarcoidosis, spondyloarthritis), gastrointestinal tract (scleroderma, inflammatory bowel disease), genitourinary tract (reactive arthritis, gonococcemia), or the nervous system (Lyme disease, vasculitis).

RHEUMATOLOGIC EVALUATION OF THE ELDERLY

The incidence of rheumatic diseases rises with age, such that 58% of those >65 years will have joint complaints. Musculoskeletal disorders in elderly patients are often not diagnosed because the signs and symptoms may be insidious, overlooked, or overshadowed by comorbidities. These difficulties are compounded by the diminished reliability of laboratory testing in the elderly, who often manifest nonpathologic abnormal results. For example, the ESR may be misleadingly elevated, and low-titer

positive tests for rheumatoid factor and antinuclear antibodies (ANAs) may be seen in up to 15% of elderly patients. Although nearly all rheumatic disorders afflict the elderly, certain diseases and drug-induced disorders (Table 331-2) are more common in this age group. The elderly should be approached in the same manner as other patients with musculoskeletal complaints, but with an emphasis on identifying the potential rheumatic consequences of medical comorbidities and therapies. OA, osteoporosis, gout, pseudogout, polymyalgia rheumatica, vasculitis, and drug-induced disorders are all more common in the elderly than in other individuals. The physical examination should identify the nature of the musculoskeletal complaint as well as coexisting diseases that may influence diagnosis and choice of treatment.

RHEUMATOLOGIC EVALUATION OF THE HOSPITALIZED PATIENT

Inpatient and outpatient evaluations and diagnostic considerations may differ, owing to greater symptom severity, more acute presentations, and greater interplay of comorbidities with the hospitalized patient. Patients with rheumatic disorders tend to be admitted for one of several reasons: (1) acute onset of inflammatory arthritis; (2) undiagnosed systemic or febrile illness; (3) musculoskeletal trauma; or (4) exacerbation or deterioration of an existing autoimmune disorder (e.g., SLE); or (5) new medical comorbidities (e.g., thrombotic event, lymphoma, infection) arising in patients with articular or connective tissue disorders. Notably, in the United States, rheumatic patients are seldom if ever admitted because of widespread pain, serologic abnormalities, or for the initiation of new therapies, although this is routinely done in other parts of the world.

Acute monarticular inflammatory arthritis may be a "red flag condition" (e.g., septic arthritis, gout, pseudogout) that will require arthrocentesis. However, new-onset polyarticular inflammatory arthritis will have a wider differential diagnosis (e.g., RA, hepatitis-related arthritis, serum sickness, drug-induced lupus, polyarticular septic arthritis) and may require targeted laboratory investigations rather than synovial fluid analysis. Patients with febrile, multisystem disorders will require exclusion of infectious or neoplastic etiologies and an evaluation driven by dominant symptoms with the greatest specificity. Conditions worthy of consideration may include vasculitis (giant cell arteritis in the elderly or polyarteritis nodosa in younger patients), adult-onset Still's disease, SLE, anti-phospholipid antibody syndrome, and sarcoidosis. As misdiagnosis of connective tissue disorders is common, patients who present with a reported preexisting rheumatic condition (e.g., SLE, RA, ankylosing spondylitis) should have their diagnosis confirmed by careful history, physicial and musculoskeletal examination, and detailed review of their medical records. It is important to note that when rheumatic disease patients are admitted to the hospital, it is usually for medical problems unrelated to their autoimmune disease, but rather because of either a comorbid condition or complication of drug therapy. Patients with chronic inflammatory disorders (e.g., RA, SLE, psoriasis, etc.) have an augmented risk of infection, cardiovascular events, and neoplasia.

Certain conditions, such as acute gout, can be precipitated in hospitalized patients by surgery, dehydration, or other events and should be considered when hospitalized patients are evaluated for the acute onset of a musculoskeletal condition. It is also common for positive results obtained from overly aggressive and unfocused laboratory testing to generate the need for a full rheumatologic evaluation.

PHYSICAL EXAMINATION

The goal of the physical examination is to ascertain the structures involved, the nature of the underlying pathology, the functional consequences of the process, and the presence of systemic or extraarticular manifestations. A knowledge of topographic anatomy is necessary to identify the primary site(s) of involvement and differentiate articular from nonarticular disorders. The musculoskeletal examination depends largely on careful inspection, palpation, and a variety of specific physical maneuvers to elicit diagnostic signs **(Table 331-3)**. Although most articulations of the appendicular skeleton can be examined in this manner, adequate inspection and palpation are not possible for many axial (e.g., zygapophyseal) and inaccessible (e.g., sacroiliac or hip) joints. For such joints, there is a greater reliance upon specific maneuvers and imaging for assessment.

Table 331-3 Glossary of Musculoskeletal Terms

Crepitus
A palpable (less commonly audible) vibratory or crackling sensation elicited with joint motion; fine joint crepitus is common and often insignificant in large joints; coarse joint crepitus indicates advanced cartilaginous and degenerative changes (as in osteoarthritis)
Subluxation
Alteration of joint alignment such that articulating surfaces incompletely approximate each other
Dislocation
Abnormal displacement of articulating surfaces such that the surfaces are not in contact
Range of motion

For diarthrodial joints, the arc of measurable movement through which the joint moves in a single plane
Contracture
Loss of full movement resulting from a fixed resistance caused either by tonic spasm of muscle (reversible) or by fibrosis of periarticular structures (permanent)
Deformity
Abnormal shape or size of a structure; may result from bony hypertrophy, malalignment of articulating structures, or damage to periarticular supportive structures
Enthesitis
Inflammation of the entheses (tendinous or ligamentous insertions on bone)
Epicondylitis
Infection or inflammation involving an epicondyle

Examination of involved and uninvolved joints will determine whether *pain*, *warmth*, *erythema*, or *swelling* is present. The locale and level of pain elicited by palpation or movement should be quantified. One example would be to count the number of tender joints on palpation of 28 easily examined joints [proximal interphalangeals (PIPs), metacarpophalangeals (MCPs), wrists, elbows, shoulders, and knees] (with a range of 0–28). Similarly, the number of swollen joints (0–28) can be counted and recorded. Careful examination should distinguish between true articular swelling (caused by synovial effusion or synovial proliferation) and nonarticular (or periarticular) involvement, which usually extends beyond the normal joint margins. Synovial effusion can be distinguished from synovial hypertrophy or bony hypertrophy by palpation or specific maneuvers. For example, small to moderate knee effusions may be identified by the "bulge sign" or "ballottement of the patellae." Bursal effusions (e.g., effusions of the olecranon or prepatellar bursa) are often focal, periarticular, overlie bony prominences, and are fluctuant with sharply defined borders. Joint *stability* can be assessed by palpation and by the application of manual stress. *Subluxation* or *dislocation*, which may be secondary to traumatic, mechanical, or inflammatory causes, can be assessed by inspection and palpation. Joint *swelling or volume* can be assessed by palpation. Distention of the articular capsule usually causes pain and evident swelling. The patient will attempt to minimize the pain by maintaining the joint in the position of least intraarticular pressure and greatest volume, usually partial flexion. For this reason, inflammatory effusions may give rise to flexion contractures. Clinically, this may be detected as fluctuant or "squishy" swelling, with grapelike compressibility. Inflammation may result in fixed flexion deformities, or diminished range of motion—especially on extension, when joint volumes are decreased. Active and passive *range of motion* should be assessed in all planes, with contralateral comparison. Serial evaluations of the joints should record the number of tender and swollen joints and loss of a normal range of motion, using a goniometer to quantify the arc of movement. Each joint should be passively manipulated through its full range of motion (including, as appropriate, flexion, extension, rotation, abduction, adduction, lateral bending, inversion, eversion, supination, pronation, medial/lateral deviation, plantar- or dorsiflexion). Limitation of motion is frequently caused by effusion, pain, deformity, or contracture. If passive motion exceeds active motion, a periarticular process (e.g., tendinitis, tendon rupture, or myopathy) should be considered. *Contractures* may reflect antecedent synovial inflammation or trauma. Minor joint *crepitus* is common during joint palpation and maneuvers, but may indicate significant cartilage degeneration as it becomes coarser (e.g., OA). Joint *deformity* usually indicates a long-standing or aggressive pathologic process. Deformities may result from ligamentous destruction, soft tissue contracture, bony enlargement, ankylosis, erosive disease, or subluxation. Examination of the musculature will document strength, atrophy, pain, or spasm. Appendicular muscle weakness should be characterized as proximal or distal. Muscle strength should be assessed by observing the patient's performance (e.g., walking, rising from a chair, grasping, writing). Strength may also be graded on a 5-point scale: 0 for no movement; 1 for trace movement or twitch; 2 for movement with gravity eliminated; 3 for movement against gravity only; 4 for movement against gravity and resistance; and 5 for normal strength. The examiner should assess for often-overlooked nonarticular or periarticular involvement, especially when articular complaints are not supported by objective findings referable to the joint capsule. The identification of soft tissue/nonarticular pain will prevent unwarranted and often expensive additional evaluations. Specific maneuvers may reveal common nonarticular abnormalities, such as a carpal tunnel syndrome (which can be identified by Tinel's or Phalen's sign). Other examples of soft tissue abnormalities include olecranon bursitis, epicondylitis (e.g., tennis elbow), enthesitis (e.g., Achilles tendinitis), and tender trigger points associated with fibromyalgia.

APPROACH TO REGIONAL RHEUMATIC COMPLAINTS

Although all patients should be evaluated in a logical and thorough manner, many cases with focal musculoskeletal complaints are caused by commonly encountered disorders that exhibit a predictable pattern of onset, evolution, and localization; they can often be diagnosed immediately on the basis of limited historic information and selected maneuvers or tests. Although nearly every joint could be

approached in this manner, the evaluation of four common involved anatomic regions—the hand, shoulder, hip, and knee—are reviewed here.

Hand Pain

Focal or unilateral hand pain may result from trauma, overuse, infection, or a reactive or crystal-induced arthritis. By contrast, bilateral hand complaints commonly suggest a degenerative (e.g., OA), systemic, or inflammatory/immune (e.g., RA) etiology. The distribution or pattern of joint involvement is highly suggestive of certain disorders **(Fig. 331-3)**. Thus, OA (or degenerative arthritis) may manifest as distal interphalangeal (DIP) and PIP joint pain with bony hypertrophy sufficient to produce Heberden's and Bouchard's nodes, respectively. Pain, with or without bony swelling, involving the base of the thumb (first carpometacarpal joint) is also highly suggestive of OA. By contrast, RA tends to involve the PIP, MCP, intercarpal, and carpometacarpal joints (wrist) with pain, prolonged stiffness, and palpable synovial tissue hypertrophy. Psoriatic arthritis may mimic the pattern of joint involvement seen in OA (DIP and PIP joints), but can be distinguished by the presence of inflammatory signs (erythema, warmth, synovial swelling), with or without carpal involvement, nail pitting, or onycholysis. Hemochromatosis should be considered when degenerative changes (bony hypertrophy) are seen at the second and third MCP joints with associated chondrocalcinosis or episodic, inflammatory wrist arthritis.

Figure 331-3

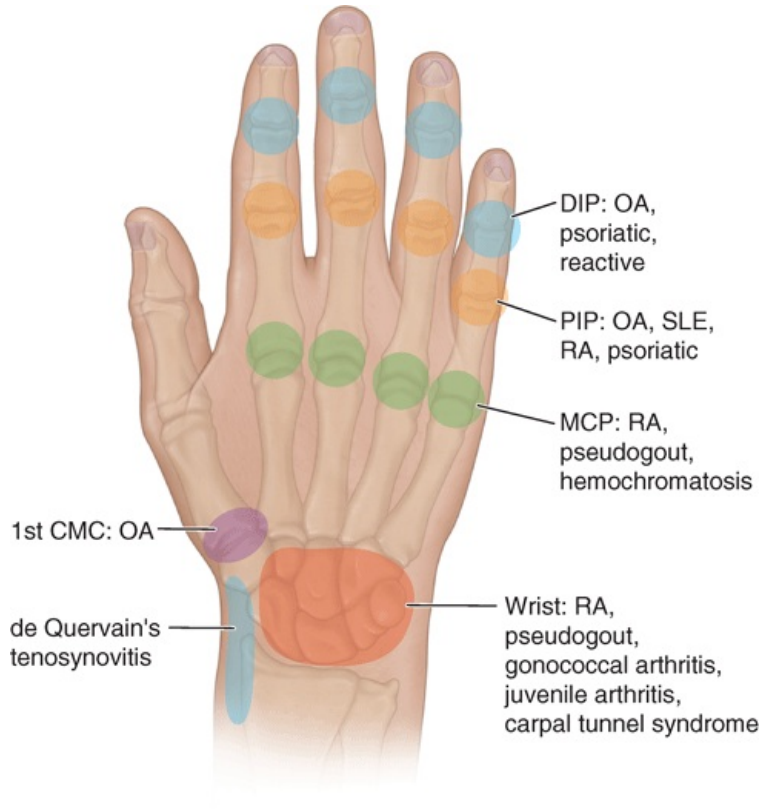

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sites of hand or wrist involvement and their potential disease associations. CMC, carpometacarpal; DIP, distal interphalangeal; MCP, metacarpophalangeal; OA, osteoarthritis; PIP, proximal interphalangeal; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus. *(From Cush et al, with permission.)*

Soft tissue swelling over the dorsum of the hand and wrist may suggest an inflammatory extensor tendon tenosynovitis possibly caused by gonococcal infection, gout, or inflammatory arthritis (e.g., RA). Tenosynovitis is suggested by localized warmth, swelling, or pitting edema and may be confirmed when the soft tissue swelling tracks with tendon movement, such as flexion and extension of fingers or when pain is induced while stretching the extensor tendon sheaths (flexing the digits distal to the MCP joints and maintaining the wrist in a fixed, neutral position).

Focal wrist pain localized to the radial aspect may be caused by de Quervain's tenosynovitis resulting from inflammation of the tendon sheath(s) involving the abductor pollicis longus or extensor pollicis brevis (Fig. 331-3). This commonly results from overuse or follows pregnancy and may be diagnosed with Finkelstein's test. A positive result is present when radial wrist pain is induced after the thumb is

flexed and placed inside a clenched fist and the patient actively deviates the hand downward with ulnar deviation at the wrist. Carpal tunnel syndrome is another common disorder of the upper extremity and results from compression of the median nerve within the carpal tunnel. Manifestations include pain in the wrist that may radiate with paresthesia to the thumb, second and third fingers, and radial half of the fourth finger and, at times, atrophy of thenar musculature. Carpal tunnel syndrome is commonly associated with pregnancy, edema, trauma, OA, inflammatory arthritis, and infiltrative disorders (e.g., amyloidosis). The diagnosis may be suggested by a positive Tinel's or Phalen's sign. With each test, paresthesia in a median nerve distribution is induced or increased by either "thumping" the volar aspect of the wrist (Tinel's sign) or pressing the extensor surfaces of both flexed wrists against each other (Phalen's sign). The variable sensitivity of these tests may require nerve conduction velocity testing to confirm a suspected diagnosis.

Shoulder Pain

During the evaluation of shoulder disorders, the examiner should carefully note any history of trauma, fibromyalgia, infection, inflammatory disease, occupational hazards, or previous cervical disease. In addition, the patient should be questioned as to the activities or movement(s) that elicit shoulder pain. While arthritis is suggested by pain on movement in all planes, pain with specific active motion suggests a periarticular (nonarticular) process. Shoulder pain may originate in the glenohumeral or acromioclavicular joints, subacromial (subdeltoid) bursa, periarticular soft tissues (e.g., fibromyalgia, rotator cuff tear/tendinitis), or cervical spine (**Fig. 331-4**). Shoulder pain is referred frequently from the cervical spine but may also be referred from intrathoracic lesions (e.g., a Pancoast tumor) or from gall bladder, hepatic, or diaphragmatic disease. Fibromyalgia should be suspected when glenohumeral pain is accompanied by diffuse periarticular (i.e., subacromial, bicipital) pain and tender points (i.e., trapezius or supraspinatus). The shoulder should be put through its full range of motion both actively and passively (with examiner assistance): forward flexion, extension, abduction, adduction, and internal and external rotation. Manual inspection of the periarticular structures will often provide important diagnostic information. Glenohumeral involvement is best detected by placing the thumb over the glenohumeral joint and applying pressure anteriorly while internally and externally rotating the humeral head. The examiner should apply direct manual pressure over the subacromial bursa that lies lateral to and immediately beneath the acromion (Fig. 331-4). Subacromial bursitis is a frequent cause of shoulder pain. Anterior to the subacromial bursa, the bicipital tendon traverses the bicipital groove. This tendon is best identified by palpating it in its groove as the patient rotates the humerus internally and externally. Direct pressure over the tendon may reveal pain indicative of bicipital tendinitis. Palpation of the acromioclavicular joint may disclose local pain, bony hypertrophy, or, uncommonly, synovial swelling. Whereas OA and RA commonly affect the acromioclavicular joint, OA seldom involves the glenohumeral joint, unless there is a traumatic or occupational cause. The glenohumeral joint is best palpated anteriorly by placing the thumb over the humeral head (just medial and inferior to the coracoid process) and having the patient rotate the humerus internally and externally. Pain localized to this region is indicative of glenohumeral pathology. Synovial effusion or tissue is seldom palpable but, if present, may suggest infection, RA, or an acute tear of the rotator cuff.

Figure 331-4

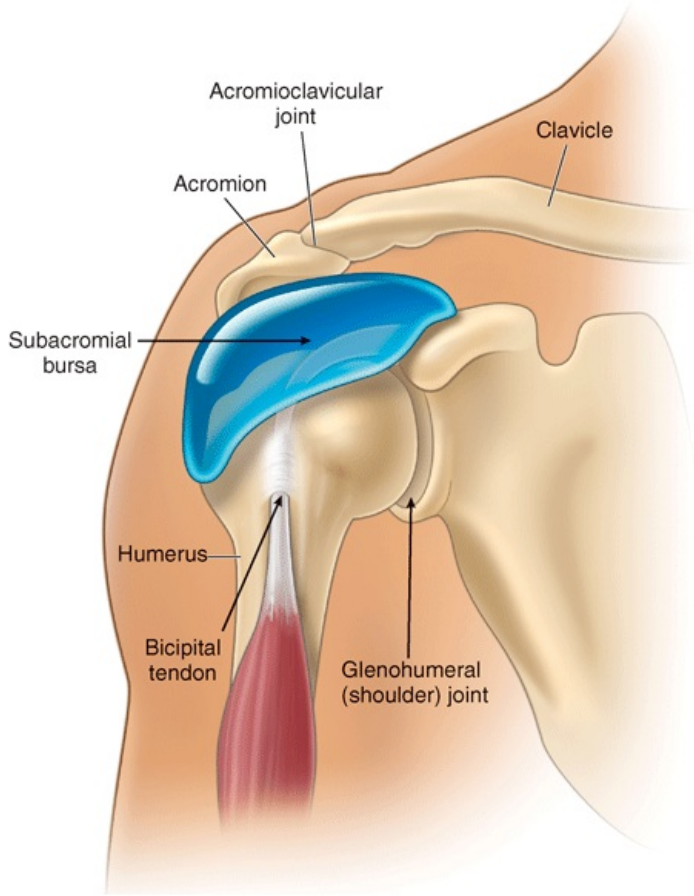

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Origins of shoulder pain. The schematic diagram of the shoulder indicates with arrows the most common causes and locations of shoulder pain.

Rotator cuff tendinitis or tear is a very common cause of shoulder pain. The rotator cuff is formed by the tendons of the supraspinatus, infraspinatus, teres minor, and subscapularis muscles. Rotator cuff tendinitis is suggested by pain on active abduction (but not passive abduction), pain over the lateral deltoid muscle, night pain, and evidence of the impingement sign. This maneuver is performed by the examiner raising the patient's arm into forced flexion while stabilizing and preventing rotation of the scapula. A positive sign is present if pain develops before 180° of forward flexion. A complete tear of the rotator cuff is more common in the elderly and often results from trauma; it may manifest in the same manner as tendinitis but is less common. The diagnosis is also suggested by the drop arm test in which the patient is unable to maintain his or her arm outstretched once it is passively abducted. If the patient is unable to hold the arm up once 90° of abduction is reached, the test is positive. Tendinitis or tear of the rotator cuff can be confirmed by magnetic resonance imaging (MRI) or ultrasound.

Knee Pain

Knee pain may result from intraarticular (OA, RA) or periarticular (anserine bursitis, collateral ligament strain) processes or be referred from hip pathology. A careful history should delineate the chronology of the knee complaint and whether there are predisposing conditions, trauma, or medications that might underlie the complaint. For example, patellofemoral disease (e.g., OA) may cause anterior knee pain that worsens with climbing stairs. Observation of the patient's gait is also important. The knee should be carefully inspected in the upright (weight-bearing) and prone positions for swelling, erythema, malalignment, visible trauma (contusion, laceration), or muscle wasting. The most common form of malalignment in the knee is *genu varum* (bowlegs) or *genu valgum* (knock-knees). Bony swelling of the knee joint commonly results from hypertrophic osseous changes seen with disorders such as OA and neuropathic arthropathy. Swelling caused by hypertrophy of the synovium or synovial effusion may manifest as a fluctuant, ballotable, or soft tissue enlargement in the suprapatellar pouch (suprapatellar reflection of the synovial cavity) or regions lateral and medial to the patella. Synovial effusions may also be detected by balloting the patella downward toward the femoral groove or by eliciting a "bulge sign." With the knee extended the examiner should manually compress, or "milk," synovial fluid down from the suprapatellar pouch and lateral to the patellae. The application of manual pressure lateral to the patella may cause an observable shift in synovial fluid (bulge) to the medial aspect. The examiner should note that this maneuver is only effective in detecting small to moderate effusions (<100 mL). Inflammatory disorders such as RA, gout, pseudogout, and reactive arthritis may

involve the knee joint and produce significant pain, stiffness, swelling, or warmth. A popliteal or *Baker's cyst* is best palpated with the knee partially flexed and is best viewed posteriorly with the patient standing and knees fully extended to visualize isolated or unilateral popliteal swelling or fullness.

Anserine bursitis is an often missed periarticular cause of knee pain in adults. The pes anserine bursa underlies the insertion of the conjoined tendons (sartorius, gracilis, semitendinosis) on the anteromedial proximal tibia and may be painful following trauma, overuse, or inflammation. It is often tender in patients with fibromyalgia, obesity, and knee osteoarthritis. Other forms of bursitis may also present as knee pain. The prepatellar bursa is superficial and is located over the inferior portion of the patella. The infrapatellar bursa is deeper and lies beneath the patellar ligament before its insertion on the tibial tubercle.

Internal derangement of the knee may result from trauma or degenerative processes. Damage to the meniscal cartilage (medial or lateral) frequently presents as chronic or intermittent knee pain. Such an injury should be suspected when there is a history of trauma, athletic activity, or chronic knee arthritis, and when the patient relates symptoms of "locking," clicking, or "giving way" of the joint. With the knee flexed 90° and the patient's foot on the table, pain elicited during palpation over the joint line or when the knee is stressed laterally or medially may suggest a meniscal tear. A positive McMurray test may also indicate a meniscal tear. To perform this test, the knee is first flexed at 90°, and the leg is then extended while the lower extremity is simultaneously torqued medially or laterally. A painful click during inward rotation may indicate a lateral meniscus tear, and pain during outward rotation may indicate a tear in the medial meniscus. Lastly, damage to the cruciate ligaments should be suspected with acute onset of pain, possibly with swelling, a history of trauma, or a synovial fluid aspirate that is grossly bloody. Examination of the cruciate ligaments is best accomplished by eliciting a drawer sign. With the patient recumbent, the knee should be partially flexed and the foot stabilized on the examining surface. The examiner should manually attempt to displace the tibia anteriorly or posteriorly with respect to the femur. If anterior movement is detected, then anterior cruciate ligament damage is likely. Conversely, significant posterior movement may indicate posterior cruciate damage. Contralateral comparison will assist the examiner in detecting significant anterior or posterior movement.

Hip Pain

The hip is best evaluated by observing the patient's gait and assessing range of motion. The vast majority of patients reporting "hip pain" localize their pain unilaterally to the posterior gluteal musculature **(Fig. 331-5)**. Such pain tends to radiate down the posterolateral aspect of the thigh and may or may not be associated with complaints of low back pain. This presentation frequently results from degenerative arthritis of the lumbosacral spine or disks and commonly follows a dermatomal distribution with involvement of nerve roots between L4 and S1. Sciatica is caused by impingement of the L4, L5, or S1 nerve (i.e., from a herniated disk) and manifests as unilateralneuropathic pain extending from the gluteal region down the posterolateral leg to the foot. Some individuals instead localize their "hip pain" laterally to the area overlying the trochanteric bursa. Because of the depth of this bursa, swelling and warmth are usually absent. Diagnosis of trochanteric bursitis can be confirmed by inducing point tenderness over the trochanteric bursa. Gluteal and trochanteric pain may also indicate underlying fibromyalgia. Range of movement may be limited by pain. Pain in the hip joint is less common and tends to be located anteriorly, over the inguinal ligament; it may radiate medially to the groin. Uncommonly, ilio-psoas bursitis may mimic true hip joint pain. Diagnosis of iliopsoas bursitis may be suggested by a history of trauma or inflammatory arthritis. Pain associated with iliopsoas bursitis is localized to the groin or anterior thigh and tends to worsen with hyperextension of the hip; many patients prefer to flex and externally rotate the hip to reduce the pain from a distended bursa.

Figure 331-5

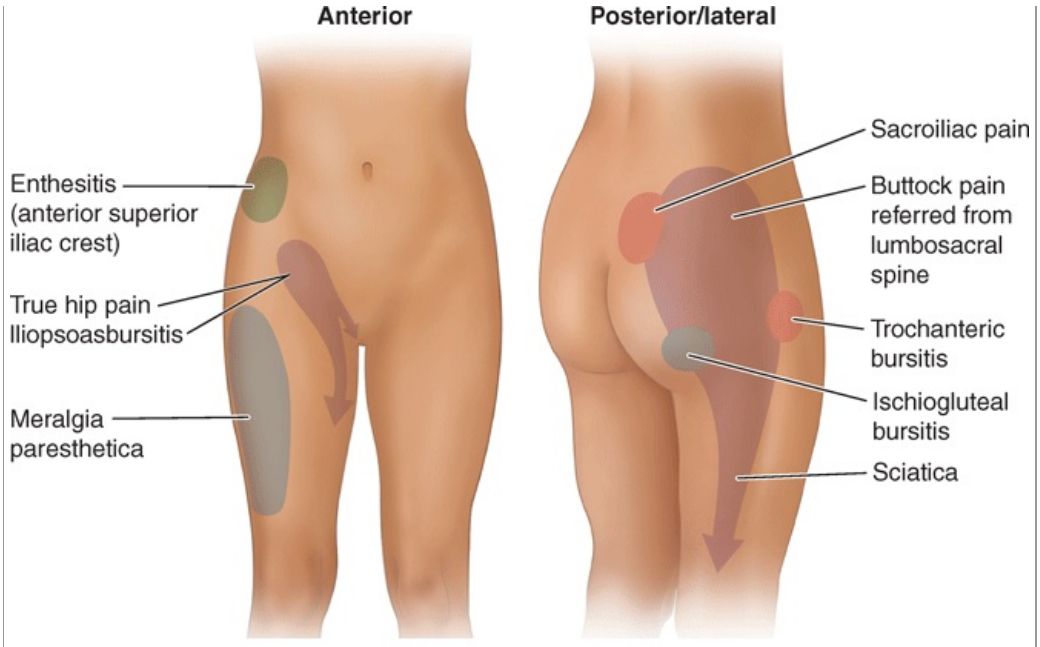

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Origins of hip pain and dysesthesias. *(From Cush et al, with permission.)*

LABORATORY INVESTIGATIONS

The vast majority of musculoskeletal disorders can be easily diagnosed by a complete history and physical examination. An additional objective of the initial encounter is to determine whether additional investigations or immediate therapy is required. A number of features indicate the need for additional evaluation. Monarticular conditions require additional evaluation, as do traumatic or inflammatory conditions and conditions accompanied by neurologic changes or systemic manifestations of serious disease. Finally, individuals with chronic symptoms (>6 weeks), especially when there has been a lack of response to symptomatic measures, are candidates for additional evaluation. The extent and nature of the additional investigation should be dictated by the clinical features and suspected pathologic process. Laboratory tests should be used to confirm a specific clinical diagnosis and not be used to screen or evaluate patients with vague rheumatic complaints. Indiscriminate use of broad batteries of diagnostic tests and radiographic procedures is rarely a useful or cost-effective means to establish a diagnosis.

Besides a complete blood count, including a white blood cell (WBC) and differential count, the routine evaluation should include a determination of an acute-phase reactant such as the ESR or CRP, which can be useful in discriminating inflammatory from noninflammatory disorders. Both are inexpensive, easily obtained, and may be elevated with infection, inflammation, autoimmune disorders, neoplasia, pregnancy, renal insufficiency, advanced age, and hyperlipidemia. Extreme elevation of the acute-phase reactants (CRP, ESR) is seldom seen without evidence of serious illness (e.g., sepsis, pleuropericarditis, polymyalgia rheumatica, giant cell arteritis, adult Still's disease).

Serum uric acid determinations are useful in the diagnosis of gout and in monitoring the response to urate-lowering therapy. Uric acid, the end product of purine metabolism, is primarily excreted in the urine. Serum values range from 238 to 516 μmol/L (4.0–8.6 mg/dL) in men; the lower values [178–351 μmol/L (3.0–5.9 mg/dL)] seen in women are caused by the uricosuric effects of estrogen. Urinary uric acid levels are normally <750 mg per 24 h. Although hyperuricemia [especially levels >535 μmol/L (9 mg/dL)] is associated with an increased incidence of gout and nephrolithiasis, levels may not correlate with the severity of articular disease. Uric acid levels (and the risk of gout) may be increased by inborn errors of metabolism (Lesch-Nyhan syndrome), disease states (renal insufficiency, myeloproliferative disease, psoriasis), or drugs (alcohol, cytotoxic therapy, thiazides). Although nearly all patients with gout will demonstrate hyperuricemia at some time during their illness, up to 5% of patients with an acute gouty attack will have normal serum uric acid levels, presumably from acute inflammation augmented excretion of uric acid. Monitoring serum uric acid may be useful in assessing the response to hypouricemic therapy or chemotherapy as the goal of therapy is to lower serum urate below 6 mg/dL.

Serologic tests for rheumatoid factor (RF), cyclic citrullinated peptide (CCP) antibodies, antinuclear antibodies (ANA), complement levels, Lyme and antineutrophil cytoplasmic antibodies (ANCA), or antistreptolysin O (ASO) titer should be carried out only when there is clinical evidence to suggest an associated diagnosis, as these have poor predictive value when used for screening, especially when

the pretest probability is low. Although 4–5% of a healthy population will have positive tests for RF and ANAs, only 1% and <0.4% of the population will have RA or SLE, respectively. IgM RF (autoantibodies against the Fc portion of IgG) is found in 80% of patients with RA and may also be seen in low titers in patients with chronic infections (*tuberculosis*, leprosy, hepatitis); other autoimmune diseases (SLE, Sjögren's syndrome); and chronic pulmonary, hepatic, or renal diseases. When considering RA, both serum RF and anti-CCP antibodies should be obtained as these are complementary. Both are comparably sensitive but CCP antibodies are more specific than RF. In RA, the presence of anti-CCP and rheumatoid factor antibodies may indicate a greater risk for more severe, erosive polyarthritis. ANAs are found in nearly all patients with SLE and may also be seen in patients with other autoimmune diseases (polymyositis, scleroderma, antiphospholipid syndrome, Sjogren's syndrome), drug-induced lupus (resulting from hydralazine, procainamide, quinidine, tetracyclines, tumor necrosis factor inhibitors), chronic liver or renal disorders, and advanced age. Positive ANAs are found in 5% of adults and in up to 14% of elderly or chronically ill individuals. The ANA test is very sensitive but poorly specific for lupus, as <5% of all positive results will be caused by lupus alone. The interpretation of a positive ANA test may depend on the magnitude of the titer and the pattern observed by immunofluorescence microscopy **(Table 331-4)**. Diffuse and speckled patterns are least specific, whereas a peripheral, or rim, pattern [related to autoantibodies against double-strand (native) DNA] is highly specific and suggestive of lupus. Centromeric patterns are seen in patients with limited scleroderma [calcinosis, Raynaud's phenomenon, esophageal involvement, sclerodactyly, telangiectasia (CREST) syndrome] or primary biliary sclerosis, and nucleolar patterns may be seen in patients with diffuse systemic sclerosis or inflammatory myositis.

Table 331-4 Antinuclear Antibody (Ana) Patterns and Clinical Associations

ANA Pattern	Antigen Identified	Clinical Correlate
Diffuse	Deoxyribonucl eoprotein Histones	Nonspecific Drug-induced lupus, lupus
Peripheral(rim)	ds-DNA	50% of SLE (specific)
Speckled	U1-RNP	≤90% of MCTD
	Sm	30% of SLE (specific)
	Ro (SS-A)	Sjögrens 60%, SCLE, neonatal lupus, ANA(-) lupus
	La (SS-B)	50% of Sjögrens, 15% lupus
	Scl-70	40% of diffuse scleroderma
	PM-1	Polymyositis (PM), dermatomyositis
	Jo-1	PM w/pneumonitis + arthritis
Nucleolar	RNA polymerase I, others	40% of PSS
Centromere	Kinetochore	75% CREST (limited scleroderma)

Aspiration and analysis of synovial fluid are always indicated in acute monarthritis or when an infectious or crystal-induced arthropathy is suspected. Synovial fluid may distinguish between noninflammatory and inflammatory processes by analysis of the appearance, viscosity, and cell count. Tests for synovial fluid glucose, protein, lactate dehydrogenase, lactic acid, or autoantibodies are not recommended as they have no diagnostic value. Normal synovial fluid is clear or a pale straw color and is viscous, primarily because of the high levels of hyaluronate. Noninflammatory syno-vial fluid is clear, viscous, and amber-colored, with a white blood cell count of <2000/μL and a predominance of mononuclear cells. The viscosity of synovial fluid is assessed by expressing fluid from the syringe one drop at a time. Normally, there is a stringing effect, with a long tail behind each synovial drop. Effusions caused by OA or trauma will have normal viscosity. Inflammatory fluid is turbid and yellow, with an increased white cell count (2000–50,000/μL) and a polymorphonuclear leukocyte predominance. Inflammatory fluid has reduced viscosity, diminished hyaluronate, and little or no tail following each drop of synovial fluid. Such effusions are found in RA, gout, and other inflammatory arthritides. Septic fluid is opaque and purulent, with a WBC count usually >50,000/μL, a predominance of polymorphonuclear leukocytes (>75%), and low viscosity. Such effusions are typical of septic arthritis but may occur with RA or gout. In addition, hemorrhagic synovial fluid may be seen with trauma, hemarthrosis, or neuropathic arthritis. An algorithm for synovial fluid aspiration and analysis is shown in **Fig. 331-6**. Synovial fluid should be analyzed immediately for appearance, viscosity, and cell count. Monosodium urate crystals (observed in gout) are seen by polarized microscopy and are long, needle-shaped, negatively birefringent, and usually intracellular. In chondrocalcinosis and pseudogout, calcium pyrophosphate dihydrate crystals are usually short, rhomboid-shaped, and positively birefringent. Whenever infection is suspected, synovial fluid should be Gram-stained and cultured appropriately. If gonococcal arthritis is suspected, immediate plating of the fluid on appropriate culture medium is indicated. Synovial fluid from patients with chronic monarthritis should also be cultured for *M. tuberculosis* and fungi. Last, it should be noted that crystal-induced and septic arthritis occasionally occur together in the same joint.

Figure 331-6

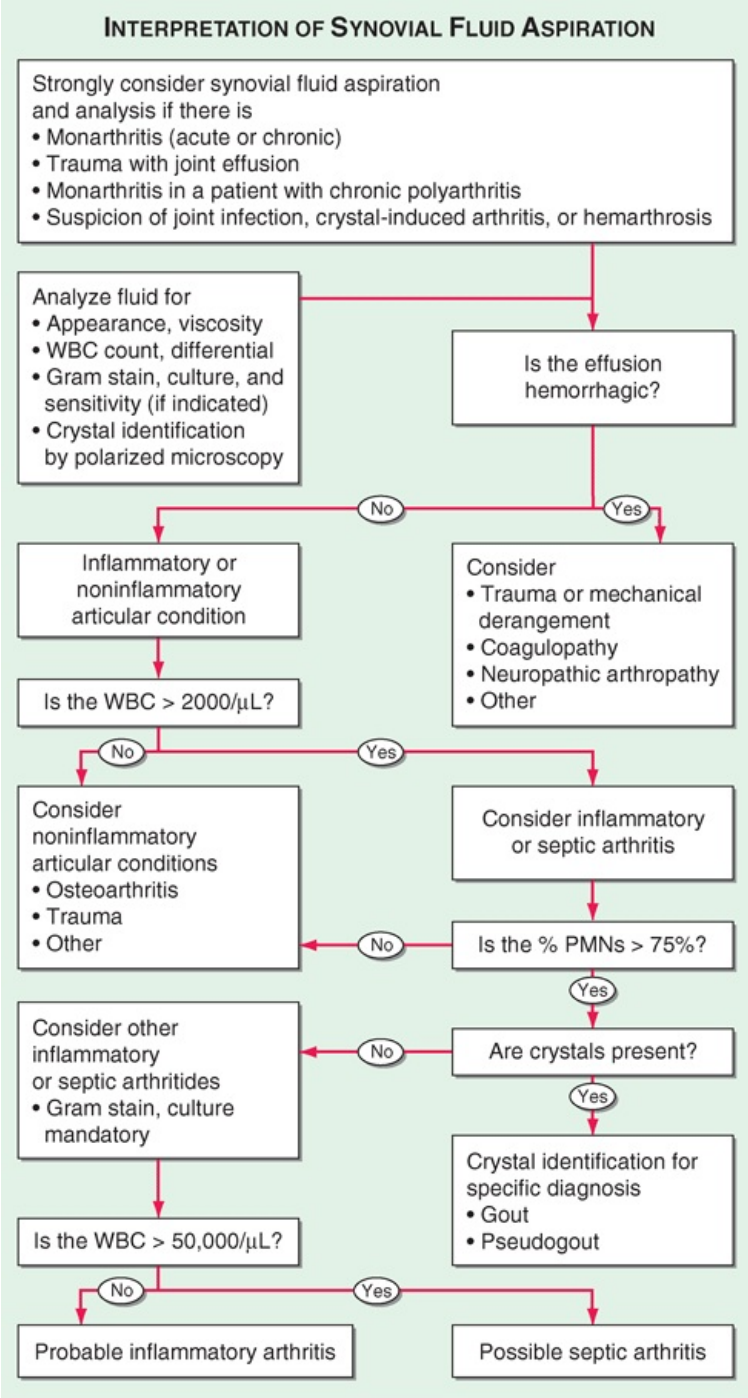

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithmic approach to the use and interpretation of synovial fluid aspiration and analysis. PMNs, polymorphonuclear (leukocytes); WBC, white blood cell (count).

DIAGNOSTIC IMAGING IN JOINT DISEASES

Conventional radiography has been a valuable tool in the diagnosis and staging of articular disorders. Plain x-rays are most appropriate when there is a history of trauma, suspected chronic infection, progressive disability, or monarticular involvement; when therapeutic alterations are considered; or when a baseline assessment is desired for what appears to be a chronic process. However, in acute inflammatory arthritis, early radiography is rarely helpful in establishing a diagnosis and may only reveal soft tissue swelling or juxtaarticular demineralization. As the disease progresses, calcification (of soft tissues, cartilage, or bone), joint space narrowing, erosions, bony ankylosis, new bone formation (sclerosis, osteophytes, or periostitis), or subchondral cysts may develop and suggest specific clinical entities. Consultation with a radiologist will help define the optimal imaging modality, technique, or

positioning and prevent the need for further studies.

Additional imaging techniques may possess greater diagnostic sensitivity and facilitate early diagnosis in a limited number of articular disorders and in selected circumstances and are indicated when conventional radiography is inadequate or nondiagnostic **(Table 331-5)**. *Ultrasonography* is useful in the detection of soft tissue abnormalities, such as tenosynovitis, that cannot be fully appreciated by clinical examination. Owing to low cost, portability, and wider use, ultrasound use has grown and is the preferred method for the evaluation of synovial (Baker's) cysts, rotator cuff tears, tendinitis and tendon injury, and suspected early synovitis. Its utility is enhanced by operator experience. *Radionuclide scintigraphy* provides useful information regarding the metabolic status of bone and, along with radiography, is well suited for total-body assessment of the extent and distribution of skeletal involvement. Radionuclide imaging is a very sensitive, but poorly specific, means of detecting inflammatory or metabolic alterations in bone or periarticular soft tissue structures. The limited tissue contrast resolution of scintigraphy may obscure the distinction between a bony or periarticular process and may necessitate the additional use of MRI. Scintigraphy, using ^{99m}Tc, ^{67}Ga, or ^{111}In-labeled WBCs has been applied to a variety of articular disorders with variable success (Table 331-5). Although [^{99m}Tc] pertechnate or diphosphate scintigraphy **(Fig. 331-7)** may be useful in identifying osseous infection, neoplasia, inflammation, increased blood flow, bone remodeling, heterotopic bone formation, or avascular necrosis, MRI is preferred in most instances. The poor specificity of ^{99m}Tc scanning has largely limited its use to surveys for bone metastases and Paget's disease of bone. Gallium scanning utilizes ^{67}Ga, which binds serum and cellular transferrin and lactoferrin, and is preferentially taken up by neutrophils, macrophages, bacteria, and tumor tissue (e.g., lymphoma). As such, it is primarily used in the identification of occult infection or malignancy. Scanning with ^{111}In-labeled WBCs has been used to detect osteomyelitis and infectious or inflammatory arthritis. Nevertheless, the use of ^{111}In-labeled WBC or ^{67}Ga scanning has largely been replaced by MRI, except when there is a suspicion of prosthetic joint infections.

Table 331-5 Diagnostic Imaging Techniques for Musculoskeletal Disorders

Method	Imaging Time, h	Cost[a]	Current Indications
Ultrasound[b]	≤1	++	Synovial cysts Rotator cuff tears Tendon injury
Radionuclide scintigraphy			
^{99m}Tc	1–4	++	Metastatic bone survey Evaluation of Paget's disease Acute and chronic osteomyelitis
^{111}In-WBC	24	+++	Acute infection Prosthetic infection Acute osteomyelitis
^{67}Ga	24–48	++++	Acute and chronic infection Acute osteomyelitis
Computed tomography	≤1	+++	Herniated intervertebral disk Sacroiliitis Spinal stenosis Spinal trauma Osteoid osteoma Stress fracture
Magnetic resonance imaging	1/2–2	++++	Avascular necrosis Osteomyelitis Intraarticular derangement and soft tissue injury Derangements of axial skeleton and spinal cord Herniated intervertebral disk Pigmented villonodular synovitis Inflammatory and metabolic muscle pathology

[a]Relative cost for imaging study. [b]Results depend on operator.

Figure 331-7

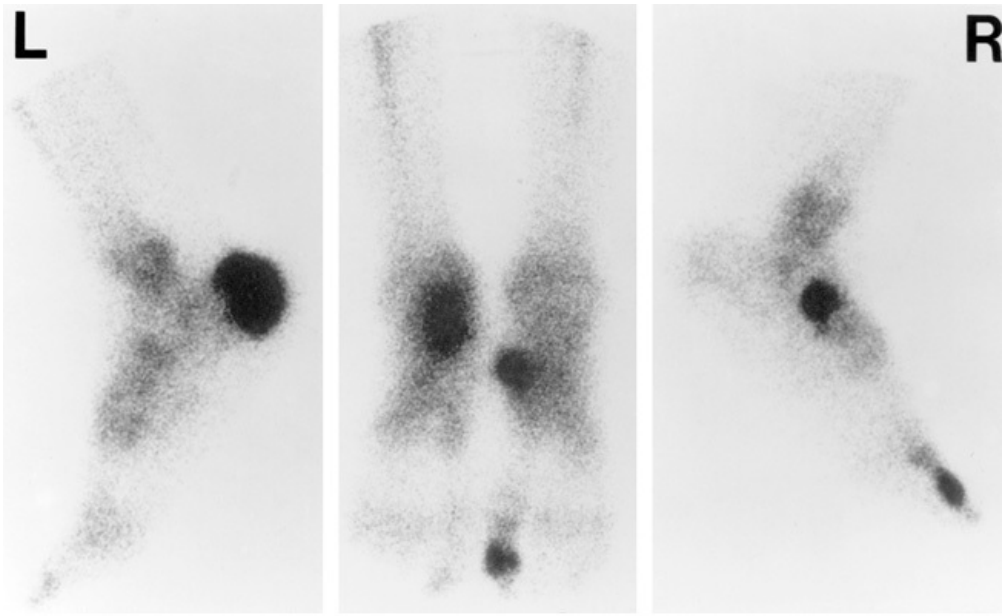

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

[^{99m}Tc]Diphosphonate scintigraphy of the feet of a 33-year-old African-American male with reactive arthritis, manifested by sacroiliitis, urethritis, uveitis, asymmetric oligoarthritis, and enthesitis. This bone scan demonstrates increased uptake indicative of enthesitis involving the insertions of the left Achilles tendon, plantar aponeurosis, and right tibialis posterior tendon as well as arthritis of the right first interphalangeal joint.

CT provides detailed visualization of the axial skeleton. Articulations previously considered difficult to visualize by radiography (e.g., zygapophyseal, sacroiliac, sternoclavicular, hip joints) can be effectively evaluated using CT. CT has been demonstrated to be useful in the diagnosis of low back pain syndromes (e.g., spinal stenosis vs. herniated disk), sacroiliitis, osteoid osteoma, and stress fractures. Helical or spiral CT (with or without contrast angiography) is a novel technique that is rapid, cost-effective, and sensitive in diagnosing pulmonary embolism or obscure fractures, often in the setting of initially equivocal findings. High-resolution CT can be advocated in the evaluation of suspected or established infiltrative lung disease (e.g., scleroderma or rheumatoid lung). The recent use of hybrid [positron emission tomography (PET)/CT or single-photon emission CT (SPECT)] scans in metastatic evaluations have incorporated CT to provide better anatomic localization of scintigraphic abnormalities.

MRI has significantly advanced the ability to image musculoskeletal structures. MRI has the advantages of providing multiplanar images with fine anatomic detail and contrast resolution **(Fig. 331-8)** that allows for the superior ability to visualize bone marrow and soft tissue periarticular structures. Although more costly with a longer procedural time than CT, the MRI has become the preferred technique when evaluating complex musculoskeletal disorders.

Figure 331-8

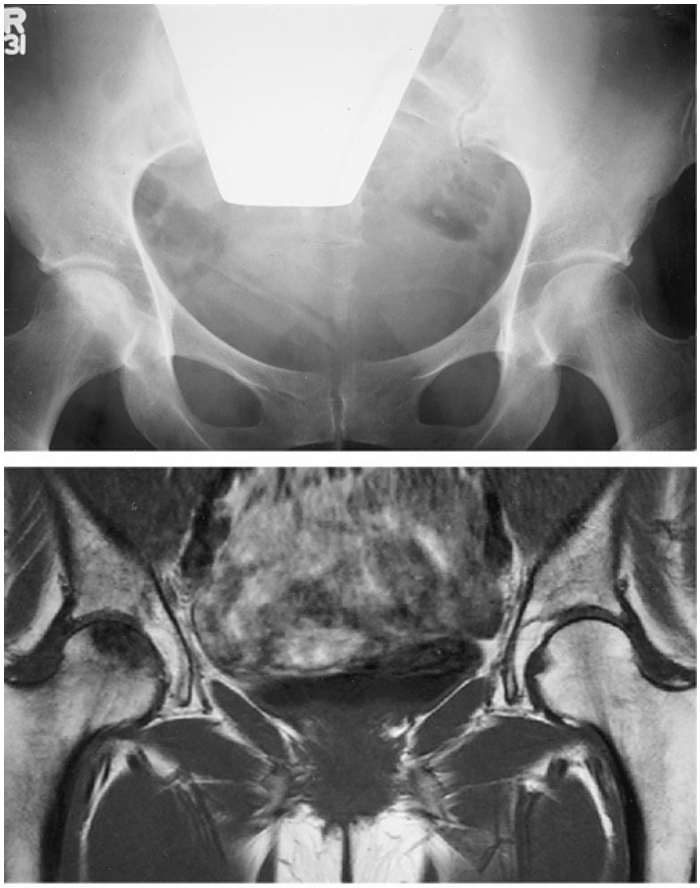

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Superior sensitivity of MRI in the diagnosis of osteonecrosis of the femoral head. A 45-year-old woman receiving highdose glucocorticoids developed right hip pain. Conventional x-rays (*top*) demonstrated only mild sclerosis of the right femoral head. T1-weighted MRI (*bottom*) demonstrated low-density signal in the right femoral head, diagnostic of osteonecrosis.

MRI can image fascia, vessels, nerve, muscle, cartilage, ligaments, tendons, pannus, synovial effusions, and bone marrow. Visualization of particular structures can be enhanced by altering the pulse sequence to produce either T1- or T2-weighted spin echo, gradient echo, or inversion recovery [including short tau inversion recovery (STIR)] images. Because of its sensitivity to changes in marrow fat, MRI is a sensitive but nonspecific means of detecting osteonecrosis, osteomyelitis, and marrow inflammation indicating overlying synovitis or osteitis (Fig. 331-8). Because of its enhanced soft tissue resolution, MRI is more sensitive than arthrography or CT in the diagnosis of soft tissue injuries (e.g., meniscal and rotator cuff tears); intraarticular derangements; marrow abnormalities (osteonecrosis, myeloma); and spinal cord or nerve root damage or synovitis.

FURTHER READINGS

American College of Rheumatology Ad Hoc Committee on Clinical Guidelines: Guidelines for the initial evaluation of the adult patient with acute musculoskeletal symptoms. Arthritis Rheum 39:1, 1996

Avouac J et al: Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: A systematic literature review. Ann Rheum Dis 65:845, 2006[PMID: 16606649] [Full Text]

Cush JJ et al: Evaluation of musculoskeletal complaints, in *Rheumatology: Diagnosis and Therapeutics,* 2nd ed, JJ Cush et al (eds). Philadelphia, Lippincott Williams & Wilkins, 2005, pp 3–20

Hootman JM, Helmick CG: Projections of US prevalence of arthritis and associated activity limitations. Arthritis Rheum 54:226, 2006[PMID: 16385518] [Full Text]

Kavanaugh A: The utility of immunologic laboratory tests in patients with rheumatic diseases. Arthritis Rheum 44:2221, 2001[PMID: 11665961] [Full Text]

Ory PA: Radiography in the assessment of musculoskeletal conditions. Best Pract Res Clin Rheumatol 17:495, 2003[PMID: 12787514] [Full Text]

Rudwaleit M et al: How to diagnose axial spondyloarthritis early. Ann Rheum Dis 63:535, 2004[PMID: 15082484] [Full Text]

Shmerling RH et al: Synovial fluid tests: What should be ordered? JAMA 264:1009, 1990[PMID: 2198352] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OSTEOARTHRITIS: INTRODUCTION

Osteoarthritis (OA) is the most common type of arthritis. Its high prevalence, especially in the elderly, and the high rate of disability related to disease make it a leading cause of disability in the elderly. Because of the aging of Western populations and because obesity, a major risk factor, is increasing in prevalence, the occurrence of osteoarthritis is on the rise. In the United States, osteoarthritis prevalence will increase by 66–100% by 2020.

OA affects certain joints, yet spares others **(Fig. 332-1)**. Commonly affected joints include the cervical and lumbosacral spine, hip, knee, and first metatarsal phalangeal joint (MTP). In the hands, the distal and proximal interphalangeal joints and the base of the thumb are often affected. Usually spared are the wrist, elbow, and ankle. Our joints were designed, in an evolutionary sense, for brachiating apes, animals that still walked on four limbs. We thus develop OA in joints that were ill designed for human tasks such as pincer grip (OA in the thumb base) and walking upright (OA in knees and hips) Some joints, like the ankles, may be spared because their articular cartilage may be uniquely resistant to loading stresses.

Figure 332-1

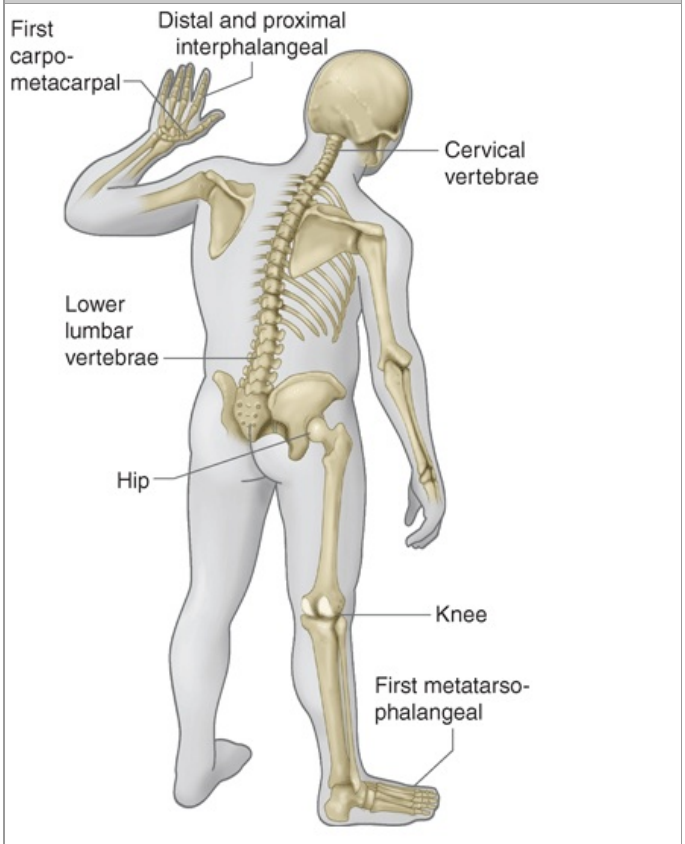

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Joints affected by osteoarthritis.

OA can be diagnosed based on structural abnormalities or on the symptoms these abnormalities evoke. According to cadaveric studies, by elderly years, structural changes of OA are nearly universal. These include cartilage loss (seen as joint space loss on x-rays) and osteophytes. Many persons with x-ray evidence of OA have no joint symptoms and, while the prevalence of structural abnormalities is of interest in understanding disease pathogenesis, what matters more from a clinical perspective is the prevalence of symptomatic OA. Symptoms, usually joint pain, determine disability, visits to clinicians,

and disease costs.

Symptomatic OA of the knee (pain on most days of a recent month in a knee plus x-ray evidence of OA in that knee) occurs in ~12% of persons age ≤60 in the United States and 6% of all adults age ≤30. Symptomatic hip OA is roughly one-third as common as disease in the knee. While radiographically evident hand OA and the appearance of bony enlargement in affected hand joints **(Fig. 332-2)** are extremely common in older persons, most cases are often not symptomatic. Even so, symptomatic hand OA occurs in ~10% of elderly individuals and often produces measurable limitation in function.

Figure 332-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Severe osteoarthritis of the hands affecting the distal interphalangeal joints (Heberden's nodes) and the proximal interphalangeal joints (Bouchard's nodes). There is no clear bony enlargement of the other common site in the hands, the thumb base.

The prevalence of OA rises strikingly with age. Regardless of how it is defined, OA is uncommon in adults under age 40 and highly prevalent in those over age 60. It is also a disease that, at least in middle-aged and elderly persons, is much more common in women than in men, and sex differences in prevalence increase with age.

X-ray evidence of OA is common in the lower back and neck, but back pain and neck pain have not been tied to findings of OA on x-ray. Thus, back pain and neck pain are treated separately (Chap. 15).

DEFINITION

OA is joint failure, a disease in which all structures of the joint have undergone pathologic change, often in concert. The pathologic sine qua non of disease is hyaline articular cartilage loss, present in a focal and, initially, nonuniform manner. This is accompanied by increasing thickness and sclerosis of the subchondral bony plate, by outgrowth of osteophytes at the joint margin, by stretching of the articular capsule, by mild synovitis in many affected joints, and by weakness of muscles bridging the joint. In knees, meniscal degeneration is part of the disease. There are numerous pathways that lead to joint failure, but the initial step is often joint injury in the setting of a failure of protective mechanisms.

JOINT PROTECTIVE MECHANISMS AND THEIR FAILURE

Joint protectors include: joint capsule and ligaments, muscle, sensory afferents, and underlying bone. Joint capsule and ligaments serve as joint protectors by providing a limit to excursion, thereby fixing the range of joint motion.

Synovial fluid reduces friction between articulating cartilage surfaces, thereby serving as a major protector against friction-induced cartilage wear. This lubrication function depends on the molecule *lubricin*, a mucinous glycoprotein secreted by synovial fibroblasts whose concentration diminishes after joint injury and in the face of synovial inflammation.

The ligaments, along with overlying skin and tendons, contain mechanoreceptor sensory afferent nerves. These mechanoreceptors fire at different frequencies throughout a joint's range of motion, providing feedback by way of the spinal cord to muscles and tendons. As a consequence, these muscles and tendons can assume the right tension at appropriate points in joint excursion to act as optimal joint protectors, anticipating joint loading.

Muscles and tendons that bridge the joint are key joint protectors. Their contractions at the appropriate time in joint movement provide the appropriate power and acceleration for the limb to accomplish its tasks. Focal stress across the joint is minimized by muscle contraction that decelerates the joint before impact and assures that when joint impact arrives, it is distributed broadly across the joint surface.

Failure of these joint protectors increases the risk of joint injury and OA. For example, in animals, OA develops rapidly when a sensory nerve to the joint is sectioned and joint injury induced. Similarly, in humans, Charcot arthropathy, a severe and rapidly progressive OA, develops when minor joint injury occurs in the presence of posterior column peripheral neuropathy. Another example of joint protector failure is rupture of ligaments, a well-known cause of the early development of OA.

Cartilage and Its Role in Joint Failure

In addition to being a primary target tissue for disease, cartilage also functions as a joint protector. A thin rim of tissue at the ends of two opposing bones, cartilage is lubricated by synovial fluid to provide an almost frictionless surface across which these two bones move. The compressible stiffness of cartilage compared to bone provides the joint with impact-absorbing capacity.

Since the earliest changes of OA may occur in cartilage and abnormalities there can accelerate disease development, understanding the structure and physiology of cartilage is critical to an appreciation of disease pathogenesis. The two major macromolecules in cartilage are type 2 collagen, which provides cartilage its tensile strength, and aggrecan, a proteoglycan macromolecule linked with hyaluronic acid, which consists of highly negatively charged glycosaminoglycans. In normal cartilage, type 2 collagen is woven tightly, constraining the aggrecan molecules in the interstices between collagen strands, forcing these highly negatively charged molecules into close proximity with one another. The aggrecan molecule, through electrostatic repulsion of its negative charges, gives cartilage its compressive stiffness. Chondrocytes, the cells within this avascular tissue, synthesize all elements of the matrix. In addition, they produce enzymes that break down the matrix and cytokines and growth factors, which in turn provide autocrine/paracrine feedback that modulates synthesis of matrix molecules **(Fig. 332-3)**. Cartilage matrix synthesis and catabolism are in a dynamic equilibrium influenced by the cytokine and growth factor environment. Mechanical and osmotic stress on chondrocytes induces these cells to alter gene expression and increase production of inflammatory cytokines and matrix-degrading enzymes. While chondrocytes synthesize numerous enzymes, especially matrix metalloproteinases (MMP), only a few of these enzymes are critical in regulating cartilage breakdown. Type 2 cartilage is degraded primarily by MMP-13 (collagenase 3), with other collagenases playing a minor role. Aggrecan degradation is a consequence, in part, of activation of two aggrecanases (ADAMTS-4 and ADAMTS-5) and perhaps of MMPs. Both collagenase and aggrecanases act primarily in the territorial matrix surrounding chondrocytes; however, as the osteoarthritic process develops, their activities and effects spread throughout the matrix, especially in the superficial layers of cartilage.

Figure 332-3

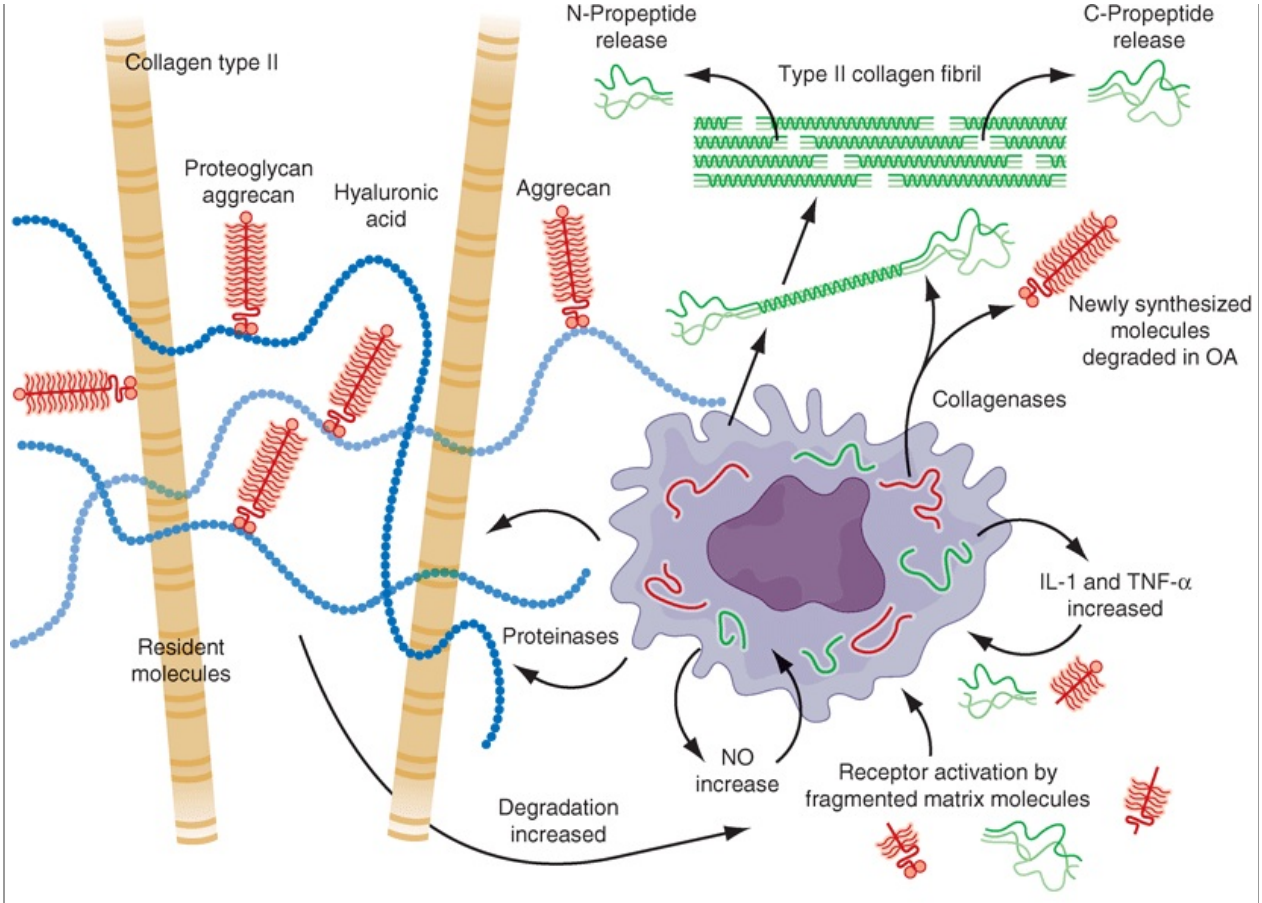

The chondrocyte and its products, type II collagen, aggrecan, and enzymes, which degrade these structures along with molecules stimulating chondrocytes. IL, interleukin; NO, nitric oxide; OA, osteoarthritis; TNF, tumor necrosis factor. (*From AR Poole et al: Ann Rheum Dis 61(S):ii78, 2002*.)

The synovium and chondrocytes synthesize numerous growth factors and cytokines. Chief among them is interleukin (IL) 1, which exerts transcriptional effects on chondrocytes, stimulating production of proteinases and suppressing cartilage matrix synthesis. In animal models of OA, IL-1 blockade prevents cartilage loss. Tumor necrosis factor (TNF) α may play a similar role to that of IL-1. These cytokines also induce chondrocytes to synthesize prostaglandin E2, nitric oxide, and bone morphogenic protein 2 (BMP-2), which together have complex effects on matrix synthesis and degradation. Nitric oxide inhibits aggrecan synthesis and enhances proteinase activity, whereas BMP-2 stimulates anabolic activity. At early stages in the matrix response to injury and in the healthy response to loading, the net effect of cytokine stimulation may be matrix synthesis but, ultimately, excess IL-1 triggers matrix degradation. Enzymes in the matrix are held in check by activation inhibitors, including tissue inhibitor of metalloproteinase (TIMP). Growth factors are also part of this complex network, with insulin-like growth factor type 1 and transforming growth factor β playing prominent roles in stimulating anabolism by chondrocytes.

Whereas healthy cartilage is metabolically sluggish, with slow matrix turnover and synthesis and degradation in balance, cartilage in early OA or after an injury is highly metabolically active. In the latter situation, stimulated chondrocytes synthesize enzymes and new matrix molecules, with those enzymes becoming activated in the matrix, causing release of degraded aggrecan and type 2 collagen into cartilage and into the synovial fluid. OA cartilage is characterized by gradual depletion of aggrecan, an unfurling of the tightly woven collagen matrix, and loss of type 2 collagen. With these changes comes increasing vulnerability of cartilage, which loses its compressive stiffness.

RISK FACTORS

Joint vulnerability and joint loading are the two major factors contributing to the development of OA. On the one hand, a vulnerable joint whose protectors are dysfunctional can develop OA with minimal levels of loading, perhaps even levels encountered during every- day activities. On the other hand, in a young joint with competent protectors, a major acute injury or long-term overloading is necessary to precipitate disease. Risk factors for OA can be understood in terms of their effect either on joint vulnerability or on loading **(Fig. 332-4)**.

Figure 332-4

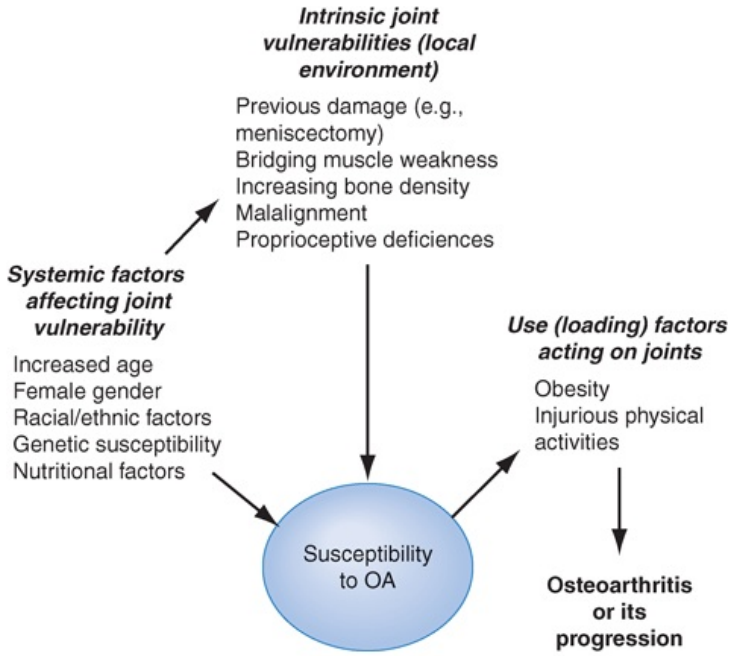

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Risk factors for osteoarthritis either contribute to the susceptibility of the joint (systemic factors or factors in the local joint environment) or increase risk by the load they put on the joint. Usually a combination of loading and susceptibility factors is required to cause disease or its progression.

Systemic Risk Factors

Age is the most potent risk factor for OA. Radiographic evidence of OA is rare in individuals under age 40; however, in some joints, such as the hands, OA occurs in >50% of persons over age 70. Aging increases joint vulnerability through several mechanisms. Whereas dynamic loading of joints stimulates cartilage matrix synthesis by chondrocytes in young cartilage, aged cartilage is less responsive to these stimuli. Indeed, because of the poor responsiveness of older cartilage to such stimulation, cartilage transplant operations are far more challenging in older than in younger persons. Partly because of this failure to synthesize matrix with loading, cartilage thins with age, and thinner cartilage experiences higher shear stress at basal layers and is at greater risk of cartilage damage. Also, joint protectors fail more often with age. Muscles that bridge the joint become weaker with age and also respond less quickly to oncoming impulses. Sensory nerve input slows with age, retarding the feedback loop of mechanoreceptors to muscles and tendons related to their tension and position. Ligaments stretch with age, making them less able to absorb impulses. These factors work in concert to increase the vulnerability of older joints to OA.

Older women are at high risk of OA in all joints, a risk that emerges as women reach their sixth decade. While hormone loss with menopause may contribute to this risk, there is little understanding of the unique vulnerability of older women vs. men to OA.

Heritability and Genetics

OA is a highly heritable disease, but its heritability varies by joint. Fifty percent of the hand and hip OA in the community is attributable to inheritance, i.e., to disease present in other members of the family. However, the heritable proportion of knee OA is at most 30%, with some studies suggesting no heritability at all. Whereas many people with OA have disease in multiple joints, this "generalized OA" phenotype is rarely inherited and is more often a consequence of aging.

Emerging evidence has identified genetic mutations that confer a high risk of OA, one of which is a polymorphism within the growth differentiation factor 5 gene. *This polymorphism diminishes the quantity of GDF5, which normally has anabolic effects on the synthesis of cartilage matrix*.

Global Considerations

Hip OA is rare in China and in immigrants from China to the United States. However, OA in the knees is at least as common, if not more so, in Chinese than in whites from the United States, and knee OA represents a major cause of disability in China, especially in rural areas. Anatomic differences between Chinese and white hips may account for much of the difference in hip OA prevalence, with white hips having a higher prevalence of anatomic predispositions to the development of OA. Persons from Africa, but not African Americans, may also have a very low rate of hip OA.

Risk Factors in the Joint Environment

Some risk factors increase vulnerability of the joint through local effects on the joint environment. With changes in joint anatomy, for example, load across the joint is no longer distributed evenly across the joint surface, but rather shows an increase in focal stress. In the hip, three uncommon developmental abnormalities occurring in utero or childhood, congenital dysplasia, Legg-Perthes disease, and slipped capital femoral epiphysis, leave a child with distortions of hip joint anatomy that often lead to OA later in life. Girls are predominantly affected by acetabular dysplasia, a mild form of congenital dislocation, whereas the other abnormalities more often affect boys. Depending on the severity of the anatomic abnormalities, hip OA occurs either in young adulthood (severe abnormalities) or middle age (mild abnormalities).

Major injuries to a joint also can produce anatomic abnormalities that leave the joint susceptible to OA. For example, a fracture through the joint surface often causes OA in joints in which the disease is otherwise rare such as the ankle and the wrist. Avascular necrosis can lead to collapse of dead bone at the articular surface, producing anatomic irregularities and subsequent OA.

Tears of ligamentous and fibrocartilaginous structures that protect the joints, such as the anterior cruciate ligament and the meniscus in the knee and the labrum in the hip, increase joint susceptibility and can lead to premature OA. Meniscal tears increase with age and when chronic are often asymptomatic but lead to adjacent cartilage damage and accelerated osteoarthritis. Even injuries that do not produce diagnosed joint injuries may increase risk of OA, perhaps because the structural injury was not detected at the time. For example, in the Framingham study subjects, men with a history of major knee injury, but no surgery, had a 3.5-fold increased risk for subsequent knee OA.

Another source of anatomic abnormality is malalignment across the joint **(Fig. 332-5)**. This factor has been best studied in the knee, which is the fulcrum of the longest lever arm in the body. Varus (bowlegged) knees with OA are at exceedingly high risk of cartilage loss in the medial or inner compartment of the knee, whereas valgus (knock-kneed) malalignment predisposes to rapid cartilage loss in the lateral compartment. Malalignment causes this effect by decreasing contact area during loading, increasing stress on a focal area of cartilage, which then breaks down. There is evidence that malalignment in the knee not only causes cartilage loss but leads to underlying bone damage, producing bone marrow lesions seen on MRI. Malalignment in the knee often produces such a substantial increase in focal stress within the knee (as evidenced by its destructive effects on subchondral bone) that severely malaligned knees may be destined to progress regardless of the status of other risk factors.

Figure 332-5

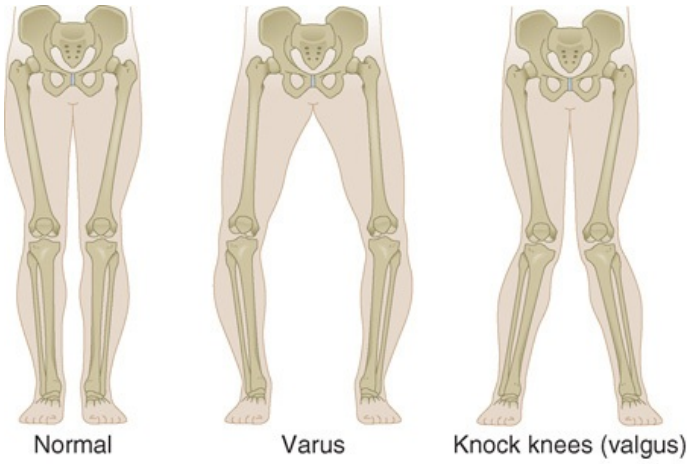

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The two types of limb malalignment in the frontal plane: varus, in which the stress is placed across the medial compartment of the knee joint, and valgus, which places excess stress across the lateral compartment of the knee.

Weakness in the quadriceps muscles bridging the knee increases the risk of the development of painful OA in the knee.

Patients with knee OA have impaired proprioception across their knees, and this may predispose them to further disease progression. The role of bone in serving as a shock absorber for impact load is not well understood, but persons with increased bone density are at high risk of OA, suggesting that the resistance of bone to impact during joint use may play a role in disease development.

Loading Factors

OBESITY

Three to six times body weight is transmitted across the knee during single-leg stance. Any increase in

weight may be multiplied by this factor to reveal the excess force across the knee in overweight persons during walking. Obesity is a well-recognized and potent risk factor for the development of knee OA and, less so, for hip OA. Obesity precedes the development of disease and is not just a consequence of the inactivity present in those with disease. It is a stronger risk factor for disease in women than in men, and in women, the relationship of weight to the risk of disease is linear, so that with each increase in weight, there is a commensurate increase in risk. Weight loss in women lowers the risk of developing symptomatic disease. Not only is obesity a risk factor for OA in weight-bearing joints, but obese persons have more severe symptoms from the disease.

Obesity's effect on the development and progression of disease is mediated mostly through the increased loading in weight-bearing joints that occurs in overweight persons. However, a modest association of obesity with an increased risk of hand OA suggests that there may be a systemic metabolic factor circulating in obese persons that affects disease risk also.

REPEATED USE OF JOINT

There are two categories of repetitive joint use, occupational use and leisure time physical activities. Workers performing repetitive tasks as part of their occupations for many years are at high risk of developing OA in the joints they use repeatedly. For example, farmers are at high risk for hip OA, and miners have high rates of OA in knees and spine, Even within a textile mill, women whose jobs required fine pincer grip [increasing the stress across the interphalangeal (IP) joints] had much more distal IP (DIP) joint OA than women whose jobs required repeated power grip, a motion that does not stress the DIP joints. Workers whose jobs require regular knee bending or lifting or carrying heavy loads have a high rate of knee OA. One reason why workers may get disease is that during long days at work, their muscles may gradually become exhausted, no longer serving as effective joint protectors.

While exercise is a major element of the treatment of OA, certain types of exercise may paradoxically increase the risk of disease. While recreational runners are not at increased risk of knee OA, studies suggest that they have a modest increased risk of disease in the hip. However, persons who have already sustained major knee injuries are at increased risk of progressive knee OA as a consequence of running. Compared to nonrunners, elite runners (professional runners and those on Olympic teams) have high risks of both knee and hip OA. Given the widespread recommendation to adopt a healthier, more exercise-filled lifestyle; longitudinal epidemiologic studies of exercise contain cautionary notes. For example, women with increased levels of physical activity, either as teenagers or at age 50, had a higher risk of developing symptomatic hip disease later in life than women who were sedentary. Other athletic activities that pose high risks of joint injury, such as football, may thereby predispose to OA.

PATHOLOGY

The pathology of OA provides evidence of the involvement of many joint structures in disease. Cartilage initially shows surface fibrillation and irregularity. As disease progresses, focal erosions develop there, and these eventually extend down to the subjacent bone. With further progression, cartilage erosion down to bone expands to involve a larger proportion of the joint surface, even though OA remains a focal disease with nonuniform loss of cartilage **(Fig. 332-6)**.

Figure 332-6

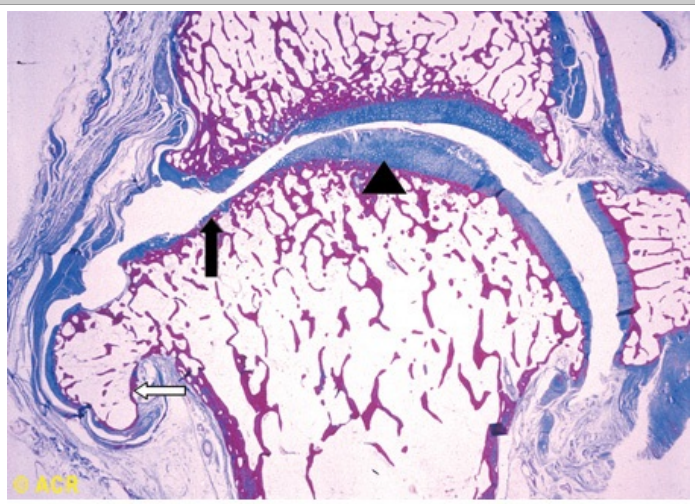

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathologic changes of osteoarthritis in a toe joint. Note the nonuniform loss of cartilage (*arrowhead* vs. *solid arrow*), the increased thickness of the subchondral bone envelope (*solid arrow*), and the osteophyte (*open arrow*). (*From the American College of Rheumatology slide collection.*)

After an injury to cartilage, chondrocytes undergo mitosis and clustering. While the metabolic activity of these chondrocyte clusters is high, the net effect of this activity is to promote proteoglycan depletion in the matrix surrounding the chondrocytes. This is because the catabolic is greater than the synthetic activity. As disease develops, collagen matrix becomes damaged, the negative charges of proteoglycans get exposed, and cartilage swells from ionic attraction to water molecules. Because in damaged cartilage proteoglycans are no longer forced into close proximity, cartilage does not bounce back after loading as it did when healthy, and cartilage becomes vulnerable to further injury. Chondrocytes at the basal level of cartilage undergo apoptosis.

With loss of cartilage come alterations in subchondral bone. Stimulated by growth factors and cytokines, osteoclasts and osteoblasts in the subchondral bony plate, just underneath cartilage, become activated. Bone formation produces a thickening and stiffness of the subchondral plate that occurs even before cartilage ulcerates. Trauma to bone during joint loading may be the primary factor driving this bone response, with healing from injury (including microcracks) producing stiffness. Small areas of osteonecrosis usually exist in joints with advanced disease. Bone death may also be caused by bone trauma with shearing of microvasculature, leading to a cutoff of vascular supply to some bone areas.

At the margin of the joint, near areas of cartilage loss, osteophytes form. These start as outgrowths of new cartilage and, with neurovascular invasion from the bone, this cartilage ossifies. Osteophytes are an important radiographic hallmark of OA. In malaligned joints, osteophytes grow larger on the side of the joint subject to most loading stress (e.g., in varus knees, osteophytes grow larger on the medial side).

The synovium produces lubricating fluids that minimize shear stress during motion. In healthy joints, the synovium consists of a single discontinuous layer filled with fat and containing two types of cells, macrophages and fibroblasts, but in OA, it can sometimes become edematous and inflamed. There is a migration of macrophages from the periphery into the tissue, and cells lining the synovium proliferate. Enzymes secreted by the synovium digest cartilage matrix that has been sheared from the surface of the cartilage.

Additional pathologic changes occur in the capsule, which stretches, becomes edematous, and can become fibrotic.

The pathology of OA is not identical across joints. In hand joints with severe OA, for example, there are often cartilage erosions in the center of the joint probably produced by bony pressure from the opposite side of the joint. In hand OA, pathology has also been noted in ligament site insertions, which may help propagate disease.

Basic calcium phosphate and calcium pyrophosphate dihydrate crystals are present microscopically in most joints with end-stage OA. Their role in osteoarthritic cartilage is unclear, but their release from cartilage into the joint space and joint fluid likely triggers synovial inflammation, which can, in turn, produce release of enzymes and trigger nociceptive stimulation.

SOURCES OF PAIN

Because cartilage is aneural, cartilage loss in a joint is not accompanied by pain. Thus, pain in OA likely arises from structures outside the cartilage. Innervated structures in the joint include the synovium, ligaments, joint capsule, muscles, and subchondral bone. Most of these are not visualized by the x-ray, and the severity of x-ray changes in OA correlates poorly with pain severity.

Based on MRI studies in osteoarthritic knees comparing those with and without pain and on studies mapping tenderness in unanesthetized joints, likely sources of pain include synovial inflammation, joint effusions, and bone marrow edema. Modest synovitis develops in many but not all osteoarthritic joints. Some diseased joints have no synovitis, whereas others have synovial inflammation that approaches the severity of joints with rheumatoid arthritis (Chap. 321). The presence of synovitis on MRI is correlated with the presence and severity of knee pain. Capsular stretching from fluid in the joint stimulates nociceptive fibers there, inducing pain. Increased focal loading as part of the disease not only damages cartilage but probably also injures the underlying bone. As a consequence, bone marrow edema appears on the MRI; histologically, this edema signals the presence of microcracks and scar, which are the consequences of trauma. These lesions may stimulate bone nociceptive fibers. Also, hemostatic pressure within bone rises in OA, and the increased pressure itself may stimulate nociceptive fibers, causing pain. Lastly, osteophytes themselves may be a source of pain. When osteophytes grow, neurovascular innervation penetrates through the base of the bone into the cartilage and into the developing osteophyte.

Pain may arise from outside the joint also, including bursae near the joints. Common sources of pain near the knee are anserine bursitis and iliotibial band syndrome.

CLINICAL FEATURES

Joint pain from OA is activity-related. Pain comes on either during or just after joint use and then gradually resolves. Examples include knee or hip pain with going up or down stairs, pain in weight-bearing joints when walking, and, for hand OA, pain when cooking. Early in disease, pain is episodic, triggered often by a day or two of overactive use of a diseased joint, such as a person with knee OA

taking a long run and noticing a few days of pain thereafter. As disease progresses, the pain becomes continuous and even begins to be bothersome at night. Stiffness of the affected joint may be prominent, but morning stiffness is usually brief (<30 min).

In knees, buckling may occur, in part, due to weakness of muscles crossing the joint. Mechanical symptoms, such as buckling, catching, or locking, could also signify internal derangement, such as meniscal tears, and need to be evaluated. In the knee, pain with activities requiring knee flexion, such as stair climbing and arising from a chair, often emanates from the patellofemoral compartment of the knee, which does not actively articulate until the knee is bent ~35°.

OA is the most common cause of chronic knee pain in persons over age 45, but the differential diagnosis is long. Inflammatory arthritis is likely if there is prominent morning stiffness and many other joints are affected. Bursitis occurs commonly around knees and hips. A physical examination should focus on whether tenderness is over the joint line (at the junction of the two bones around which the joint is articulating) or is outside of it. Anserine bursitis, medial and distal to the knee, is an extremely common cause of chronic knee pain that may respond to a glucocorticoid injection. Prominent nocturnal pain in the absence of end-stage OA merits a distinct workup. For hip pain, OA can be detected by loss of internal rotation on passive movement, and pain isolated to an area lateral to the hip joint usually reflects the presence of trochanteric bursitis.

No blood tests are routinely indicated for workup of patients with OA unless symptoms and signs suggest inflammatory arthritis. Examination of the synovial fluid is often more helpful diagnostically than an x-ray. If the synovial fluid white count is >1000 per μL, inflammatory arthritis or gout or pseudogout are likely, the latter two being also identified by the presence of crystals.

X-rays are indicated to evaluate chronic hand pain and hip pain thought to be due to OA, as the diagnosis is often unclear without confirming radiographs. For knee pain, x-rays should be obtained if symptoms or signs are not typical of OA or if knee pain persists after inauguration of effective treatment. In OA, radiographic findings **(Fig. 332-7)** correlate poorly with the presence and severity of pain. Further, radiographs may be normal in early disease as they are insensitive to cartilage loss and other early findings.

Figure 332-7

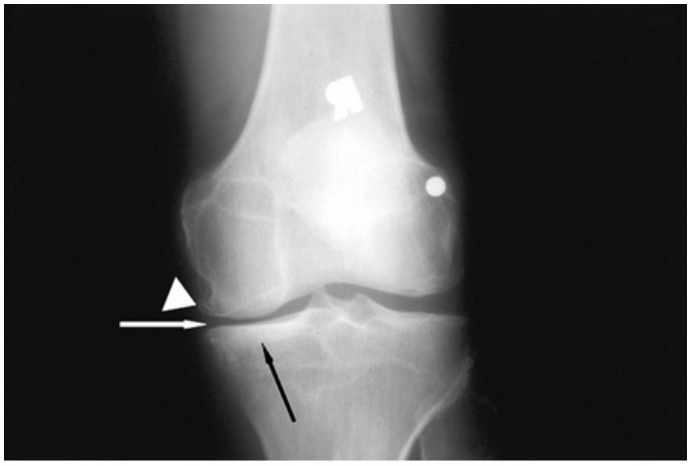

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

X-ray of knee with medial osteoarthritis. Note the narrowed joint space on medial side of the joint only (*white arrow*), the sclerosis of the bone in the medial compartment providing evidence of cortical thickening (*black arrow*), and the osteophytes in the medial femur (*white wedge*).

While MRI may reveal the extent of pathology in an osteoarthritic joint, it is not indicated as part of the diagnostic workup. Findings such as meniscal tears in cartilage and bone lesions occur in most patients with OA in the knee, but almost never warrant a change in therapy.

Treatment: Osteoarthritis

The goals of the treatment of OA are to alleviate pain and minimize loss of physical function. To the extent that pain and loss of function are consequences of inflammation, of weakness across the joint, and of laxity and instability, the treatment of OA involves addressing each of these impairments. Comprehensive therapy consists of a multimodality approach including nonpharmacologic and pharmacologic elements.

Patients with mild and intermittent symptoms may need only reassurance or nonpharmacologic treatments. Patients with ongoing, disabling pain are likely to need both nonpharmaco- and pharmacotherapy.

Treatments for knee OA have been more completely evaluated than those for hip and hand OA or for disease in other joints. Thus, while the principles of treatment are identical for OA in all joints, we shall focus below on the treatment of knee OA, noting specific recommendations for disease in other joints, especially when they differ from those for disease in the knee.

NONPHARMACOTHERAPY

Since OA is a mechanically driven disease, the mainstay of treatment involves altering loading across the painful joint and improving the function of joint protectors, so they can better distribute load across the joint. Ways of lessening focal load across the joint include

1. (1) avoiding activities that overload the joint, as evidenced by their causing pain;
2. (2) improving the strength and conditioning of muscles that bridge the joint, so as to optimize their function; and
3. (3) unloading the joint, either by redistributing load within the joint with a brace or a splint or by unloading the joint during weight bearing with a cane or a crutch.

The simplest effective treatment for many patients is to avoid activities that precipitate pain. For example, for the middle-aged patient whose long-distance running brings on symptoms of knee OA, a less demanding form of weight-bearing activity may alleviate all symptoms. For an older person whose daily constitutionals up and down hills bring on knee pain, routing the constitutional away from hills might eliminate symptoms.

Each pound of weight increases the loading across the knee three- to sixfold. Weight loss may have a commensurate multiplier effect, unloading both knees and hips. Thus, weight loss, especially if substantial, may lessen symptoms of knee and hip OA.

In hand joints affected by OA, splinting, by limiting motion, often minimizes pain for patients with involvement either in the base of the thumb or in the DIP or proximal IP joints. With an appropriate splint, function can often be preserved. Weight-bearing joints such as knees and hips can be unloaded by using a cane in the hand opposite to the affected joint for partial weight bearing. A physical therapist can help teach the patient how to use the cane optimally, including ensuring that its height is optimal for unloading. Crutches or walkers can serve a similar beneficial function.

Exercise

Osteoarthritic pain in knees or hips during weight bearing results in lack of activity and poor mobility and, because OA is so common, the inactivity that results represents a public health concern, increasing the risk of cardiovascular disease and of obesity. Aerobic capacity is poor in most elders with symptomatic knee OA, worse than others of the same age.

The development of weakness in muscles that bridge osteoarthritic joints is multifactorial in etiology. First, there is a decline in strength with age. Second, with limited mobility comes disuse muscle atrophy. Third, patients with painful knee or hip OA alter their gait so as to lessen loading across the affected joint, and this further diminishes muscle use. Fourth, "arthrogenous inhibition" may occur, whereby contraction of muscles bridging the joint is inhibited by a nerve afferent feedback loop emanating in a swollen and stretched joint capsule; this prevents maximal attainment of voluntary maximal strength. Since adequate muscle strength and conditioning are critical to joint protection, weakness in a muscle that bridges a diseased joint makes the joint more susceptible to further damage and pain. The degree of weakness correlates strongly with the severity of joint pain and the degree of physical limitation. One of the cardinal elements of the treatment of OA is to improve the functioning of muscles surrounding the joint.

For knee and hip OA, trials have shown that exercise lessens pain and improves physical function. Most effective exercise regimens consist of aerobic and/or resistance training, the latter of which focuses on strengthening muscles across the joint. Exercises are likely to be effective, especially if they train muscles for the activities a person performs daily. Some exercises may actually increase pain in the joint; these should be avoided, and the regimen needs to be individualized to optimize effectiveness and minimize discomfort. Range-of-motion exercises, which do not strengthen muscles, and isometric exercises that strengthen muscles, but not through range of motion, are unlikely to be effective by themselves. Isokinetic and isotonic strengthening (strengthening that occurs when a person flexes or extends the knees against resistance) have been shown consistently to be efficacious. Low-impact exercises, including water aerobics and water resistance training, are often better tolerated by patients than exercises involving impact loading, such as running or treadmill exercises. A patient should be referred to an exercise class or to a therapist who can create an individualized regimen, and then an individualized home-based regimen can be crafted.

There is no strong evidence that patients with hand OA benefit from therapeutic exercise, although for any patient with OA, individualized exercise programs should be tried. Adherence to exercise over the long term is the major challenge to an exercise prescription. In trials involving patients with knee OA, who are interested in exercise treatment, a third to over a half of patients stopped exercising by 6 months. Less than 50% continued regular exercise at 1 year. The strongest predictor of continued exercise in a patient is a previous personal history of successful exercise. Physicians should reinforce

the exercise prescription at each clinic visit, help the patient recognize barriers to ongoing exercise, and identify convenient times for exercise to be done routinely. The combination of exercise with calorie restriction is especially effective in lessening pain.

One clinical trial has suggested that, among those with very early OA, participating in a strengthening and multimodality exercise program led to improvement in cartilage biochemistry, as evidenced by MRI imaging. There is little other evidence, however, that strengthening or other exercise has an effect on joint structure.

Correction of Malalignment

Malalignment in the frontal plane (varus-valgus) markedly increases the stress across the joint, which can lead to progression of disease and to pain and disability (Fig. 332-5). Correcting malalignment, either surgically or with bracing, may relieve pain in persons whose knees are maligned. Malalignment develops over years as a consequence of gradual anatomic alterations of the joint and bone, and correcting it is often very challenging. One way is with a fitted brace, which takes an often varus osteoarthritic knee and straightens it by putting valgus stress across the knee. Unfortunately, many patients are unwilling to wear a realigning knee brace, plus in patients with obese legs, braces may slip with usage and lose their realigning effect. They are indicated for willing patients who can learn to put them on correctly and on whom they do not slip.

Other ways of correcting malalignment across the knee include the use of orthotics in footwear. Unfortunately, while they may have modest effects on knee alignment, trials have heretofore not demonstrated efficacy of a lateral wedge orthotic vs. placebo wedges.

Pain from the patellofemoral compartment of the knee can be caused by tilting of the patella or patellar malalignment with the patella riding laterally (or less often, medially) in the femoral trochlear groove. Using a brace to realign the patella, or tape to pull the patella back into the trochlear sulcus or reduce its tilt, has been shown, when compared to placebo taping in clinical trials, to lessen patellofemoral pain. However, patients may find it difficult to apply tape, and skin irritation from the tape is common. Commercial patellar braces may be a solution, but they have not been tested.

While their effect on malalignment is questionable, neoprene sleeves pulled to cover the knee lessen pain and are easy to use and popular among patients. The explanation for their therapeutic effect on pain is unclear.

In patients with knee OA, acupuncture produces modest pain relief compared to placebo needles and may be an adjunctive treatment.

PHARMACOTHERAPY

While nonpharmacologic approaches to therapy constitute its mainstay, pharmacotherapy serves an important adjunctive role in OA treatment. Available drugs are administered using oral, topical, and intraarticular routes.

Acetaminophen, Nonsteroidal Anti-Inflammatory Drugs (NSAIDs), and COX-2 Inhibitors

Acetaminophen (paracetamol) is the initial analgesic of choice for patients with OA in knee, hip, or hands. For some patients, it is adequate to control symptoms, in which case more toxic drugs such as NSAIDs can be avoided. Doses up to 1 g 4 times daily can be used **(Table 332-1)**.

Table 332-1 Pharmacologic Treatment for Osteoarthritis

Treatment	Dosage	Comments
Acetaminophen	Up to 1 g qid	Prolongs half-life of warfarin
Oral NSAIDs and COX-2 inhibitors[a]		Take with food. Increased risk of myocardial infarction and stroke for some NSAIDs and especially COX-2 inhibitors. High rates of gastrointestinal side effects, including ulcers and bleeding, occur. Patients at high risk for gastrointestinal side effects should also take either a proton pump inhibitor or misoprostol.b There is an increased gastrointestinal side effects or bleeding when taken with acetylsalicylic acid. Can also cause edema and renal insufficiency.
Naproxen	375–500 mg bid	
Salsalate	1500 mg bid	
Ibuprofen	600–800 mg 3–4 times a day	
Topical NSAIDs		Rub onto joint. Few systemic side effects. Skin irritation common.
Diclofenac Na 1% gel	4gm qid (for knees)	
Opiates	Various	Common side effects include dizziness, sedation, nausea or vomiting, dry mouth, constipation, urinary retention, and pruritis. Respiratory and central

		nervous system depression can occur.
Capsaicin	0.025 –0.075% cream 3–4 times a day	Can irritate mucous membranes.
Intraarticular injections		
Steroids		
Hyaluronans	Varies from 3–5 weekly injections depending on preparation	Mild to moderate pain at injection site. Controversy exists re: efficacy.

[a]COX-2, cyclooxygenase 2; NSAIDs, nonsteroidal anti-inflammatory drugs.

[b]Patients at high risk include those with previous gastrointestinal events, persons ≥60 years, and persons taking glucocorticoids. Trials have shown the efficacy of proton pump inhibitors and misoprostol in the prevention of ulcers and bleeding. Misoprostol is associated with a high rate of diarrhea and cramping; therefore, proton pump inhibitors are more widely used to reduce NSAID-related gastrointestinal symptoms.

Source: Adapted from Felson 2006.

NSAIDs are the most popular drugs to treat osteoarthritic pain. They can be administered either topically or orally. In clinical trials, oral NSAIDs produce ~30% greater improvement in pain than high-dose acetaminophen. Occasional patients treated with NSAIDs experience dramatic pain relief, whereas others experience little improvement. Initially, NSAIDs should be administered topically or taken orally on an "as needed" basis because side effects are less frequent with low intermittent doses, which may be highly efficacious. If occasional medication use is insufficiently effective, then daily treatment may be indicated, with an anti-inflammatory dose selected (Table 332-1). Patients should be reminded to take low-dose aspirin and ibuprofen at different times to eliminate a drug interaction.

NSAIDs taken orally have substantial and frequent side effects, the most common of which is upper gastrointestinal toxicity, including dyspepsia, nausea, bloating, gastrointestinal bleeding, and ulcer disease. Some 30–40% of patients experience upper gastrointestinal (GI) side effects so severe as to require discontinuation of medication. To minimize the risk of nonsteroidal-related GI side effects, patients should not take two NSAIDs, and should take medications after food; if risk is high, patients should take a gastroprotective agent, such as a proton pump inhibitor. Certain oral agents are safer to the stomach than others including nonacetylated salicylates and nabumetone. Major NSAID-related GI side effects can occur in patients who do not complain of upper GI symptoms. In one study of patients hospitalized for GI bleeding, 81% had no premonitory symptoms.

Because of the increased rates of cardiovascular events associated with cyclooxygenase 2 (COX-2) inhibitors and with some conventional NSAIDs such as diclofenac, many of these drugs are not appropriate long-term treatment choices for older persons with osteoarthritis, especially those at high risk of heart disease or stroke. The American Heart Association has identified rofecoxib and all other COX-2 inhibitors as putting patients at high risk, although low doses of celecoxib, such as 200 mg/d, may not be associated with an elevation of risk. The only conventional NSAID that appears safe from a cardiovascular perspective is naproxen, but it does have gastrointestinal toxicity.

There are other common side effects of NSAIDs, including the tendency to develop edema, because of prostaglandin inhibition of afferent blood supply to glomeruli in the kidneys and, for similar reasons, a predilection toward reversible renal insufficiency. Blood pressure may increase modestly in some NSAID-treated patients.

With the approval by the U.S. Food and Drug Administration of topical diclofenac and the availability of these agents in Europe, clinicians have a choice of administration modality for anti-inflammatory drugs. NSAIDs can be placed into a gel or topical solution with another chemical modality that enhances penetration of the skin barrier. When absorbed through the skin, plasma concentrations are an order of magnitude lower than with the same amount of drug administered orally or parenterally. However, when these drugs are administered topically in proximity to a superficial joint, (knees, hands, but not hips), the drug can be found in joint tissues such as the synovium and cartilage. Trial results have varied but generally have found that topical NSAIDs are slightly less efficacious than oral agents, but have far fewer gastrointestinal and systemic side effects. Unfortunately, topical NSAIDs often cause local skin irritation where the medication is applied, inducing redness, burning or itching in up to 40 percent of patients (see Table 332-1).

Intraarticular Injections: Glucocorticoids and Hyaluronic Acid

Since synovial inflammation is likely to be a major cause of pain in patients with OA, local anti-inflammatory treatments administered intraarticularly may be effective in ameliorating pain, at least

temporarily. Glucocorticoid injections provide such efficacy, but work better than placebo injections for only 1 or 2 weeks. This may be because the disease remains mechanically driven and, when a person begins to use the joint, the loading factors that induce pain return. Glucocorticoid injections are useful to get patients over acute flares of pain and may be especially indicated if the patient has coexistent OA and crystal deposition disease, especially from calcium pyrophosphate dihydrate crystals (Chap. 333). There is no evidence that repeated glucocorticoid injections into the joint are dangerous.

Hyaluronic acid injections can be given for treatment of symptoms in knee and hip OA, but there is controversy as to whether they have efficacy vs. placebo (Table 332-1).

Optimal therapy for OA is often achieved by trial and error, with each patient having idiosyncratic responses to specific treatments. When medical therapies have failed and the patient has an unacceptable reduction in their quality of life and ongoing pain and disability, then at least for knee and hip OA, total joint arthroplasty is indicated.

SURGERY

For knee OA, several operations are available. Among the most popular surgeries, at least in the United States, is arthroscopic debridement and lavage. Randomized trials evaluating this operation have showed that its efficacy is no greater than that of sham surgery or no treatment for relief of pain or disability. Even mechanical symptoms such as buckling, which are extremely common in patients with knee OA, do not respond to arthroscopic debridement. Arthroscopic meniscectomy is indicated for acute meniscal tears in which symptoms such as locking and acute pain are clearly related temporally to a knee injury that produced the tear.

For patients with knee OA isolated to the medial compartment, operations to realign the knee to lessen medial loading can relieve pain. These include a high tibial osteotomy, in which the tibia is broken just below the tibial plateau and realigned so as to load the lateral, nondiseased compartment, or a unicompartmental replacement with realignment. Each surgery may provide the patient with years of pain relief before they require a total knee replacement.

Ultimately, when the patient with knee or hip OA has failed medical treatment modalities and remains in pain, with limitations of physical function that compromise the quality of life, the patient should be referred for total knee or hip arthroplasty. These are highly efficacious operations that relieve pain and improve function in the vast majority of patients. Currently failure rates are ~1% per year, although these rates are higher in obese patients. The chance of surgical success is greater in centers where at least 25 such operations are performed yearly or with surgeons who perform multiple operations annually. The timing of knee or hip replacement is critical. If the patient suffers for many years until their functional status has declined substantially, with considerable muscle weakness, postoperative functional status may not improve to a level achieved by others who underwent operation earlier in their disease course.

Cartilage Regeneration

Chondrocyte transplantation has not been found to be efficacious in OA, perhaps because OA includes pathology of joint mechanics, which is not corrected by chondrocyte transplants. Similarly, abrasion arthroplasty (chondroplasty) has not been well studied for efficacy in OA, but it produces fibrocartilage in place of damaged hyaline cartilage. Both of these surgical attempts to regenerate and reconstitute articular cartilage may be more likely to be efficacious early in disease when joint malalignment and many of the other noncartilage abnormalities that characterize OA have not yet developed.

FURTHER READINGS

Abramson SB, Attur M: Developments in the scientific understanding of osteoarthritis. Arthritis Res Ther 11:227, 2009[PMID: 19519925] [Full Text]

Bennell K, Hinman R: Exercise as a treatment for osteoarthritis. Curr Opin Rheumatol 17:634, 2005[PMID: 16093845] [Full Text]

Felson DT: Osteoarthritis of the knee. N Engl J Med 354:841, 2006[PMID: 16495396] [Full Text]

: Developments in the clinical understanding of osteoarthritis. Arthritis Res Ther 11:203, 2009

Krohn K: Footwear alterations and bracing as treatments for knee osteoarthritis. Curr Opin Rheumatol 17:653, 2005[PMID: 16093848] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GOUT AND OTHERCRYSTAL-ASSOCIATED ARTHROPATHIES: INTRODUCTION

The use of polarizing light microscopy during synovial fluid analysis in 1961 by McCarty and Hollander and the subsequent application of other crystallographic techniques, such as electron microscopy, energy-dispersive elemental analysis, and x-ray diffraction, have allowed investigators to identify the roles of different microcrystals, including monosodium urate (MSU), calcium pyrophosphate dihydrate (CPPD), calcium apatite (apatite), and calcium oxalate (CaOx), in inducing acute or chronic arthritis or periarthritis. The clinical events that result from deposition of MSU, CPPD, apatite, and CaOx have many similarities but also have important differences. Before the use of crystallographic techniques in rheumatology, much of what was considered to be gouty arthritis in fact was not. Because of often similar clinical presentations, the need to perform synovial fluid analysis to distinguish the type of crystal involved must be emphasized. Polarized light microscopy alone can identify most typical crystals; apatite, however, is an exception. Aspiration and analysis of effusions are also important to assess the possibility of infection. Apart from the identification of specific microcrystalline materials or organisms, synovial fluid characteristics in crystal-associated diseases are nonspecific, and synovial fluid can be inflammatory or noninflammatory. A list of possible musculoskeletal manifestations of crystal-associated arthritis is shown in **Table 333-1**.

Table 333-1 Musculoskeletal Manifestations of Crystal-Induced Arthritis

Acute mono- or polyarthritis	Destructive arthropathies
Bursitis	Pseudo-rheumatoid arthritis
Tendinitis	Pseudo-ankylosing spondylitis
Enthesitis	Spinal stenosis
Tophaceous deposits	Crowned dens syndrome
Peculiar type of osteoarthritis	Carpal tunnel syndrome
Synovial osteochondromatosis	Tendon rupture

GOUT

Gout is a metabolic disease that most often affects middle-aged to elderly men and postmenopausal women. It results from an increased body pool of urate with hyperuricemia. It typically is characterized by episodic acute and chronic arthritis caused by deposition of MSU crystals in joints and connective tissue tophi and the risk for deposition in kidney interstitium or uric acid nephrolithiasis (Chap. 359).

Acute and Chronic Arthritis

Acute arthritis is the most common early clinical manifestation of gout. Usually, only one joint is affected initially, but polyarticular acute gout can occur in subsequent episodes. The metatarsophalangeal joint of the first toe often is involved, but tarsal joints, ankles, and knees also are affected commonly. Especially in elderly patients or in advanced disease, finger joints may be involved. Inflamed Heberden's or Bouchard's nodes may be a first manifestation of gouty arthritis. The first episode of acute gouty arthritis frequently begins at night with dramatic joint pain and swelling. Joints rapidly become warm, red, and tender, with a clinical appearance that often mimics that of cellulitis. Early attacks tend to subside spontaneously within 3–10 days, and most patients have intervals of varying length with no residual symptoms until the next episode. Several events may precipitate acute gouty arthritis: dietary excess, trauma, surgery, excessive ethanol ingestion, hypouricemic therapy, and serious medical illnesses such as myocardial infarction and stroke.

After many acute mono- or oligoarticular attacks, a proportion of gouty patients may present with a chronic nonsymmetric synovitis, causing potential confusion with rheumatoid arthritis (Chap. 321). Less commonly, chronic gouty arthritis will be the only manifestation, and, more rarely, the disease will manifest only as periarticular tophaceous deposits in the absence of synovitis. Women represent only 5–20% of all patients with gout. Premenopausal gout is rare; it is seen mostly in individuals with a strong family history of gout. Kindreds of precocious gout in young females caused by decreased renal urate clearance and renal insufficiency have been described. Most women with gouty arthritis are postmenopausal and elderly, have osteoarthritis and arterial hypertension that cause mild renal insufficiency, and usually are receiving diuretics.

LABORATORY DIAGNOSIS

Even if the clinical appearance strongly suggests gout, the presumptive diagnosis ideally should be confirmed by needle aspiration of acutely or chronically involved joints or tophaceous deposits. Acute septic arthritis, several of the other crystalline-associated arthropathies, palindromic rheumatism, and psoriatic arthritis may present with similar clinical features. During acute gouty attacks, needle-shaped MSU crystals typically are seen both intracellularly and extracellularly **(Fig. 333-1)**. With compensated polarized light these crystals are brightly birefringent with negative elongation. Synovial fluid leukocyte counts are elevated from 2000 to 60,000/μL. Effusions appear cloudy due to the increased numbers of leukocytes. Large amounts of crystals occasionally produce a thick pasty or chalky joint fluid. Bacterial infection can coexist with urate crystals in synovial fluid; if there is any suspicion of septic arthritis, joint fluid must be cultured.

Figure 333-1

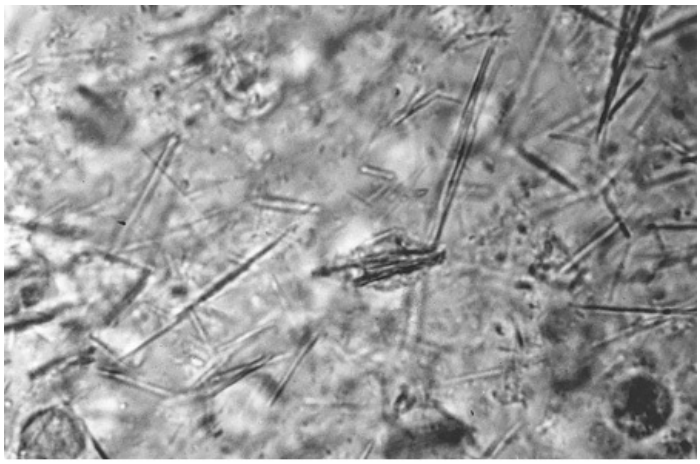

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Extracellular and intracellular monosodium urate crystals, as seen in a fresh preparation of synovial fluid, illustrate needle- and rod-shaped crystals. These crystals are strongly negative birefringent crystals under compensated polarized light microscopy; 400x.

MSU crystals also can often be demonstrated in the first metatarsophalangeal joint and in knees not acutely involved with gout. Arthrocentesis of these joints is a useful technique to establish the diagnosis of gout between attacks.

Serum uric acid levels can be normal or low at the time of an acute attack, as inflammatory cytokines can be uricosuric and effective initiation of hypouricemic therapy can precipitate attacks. This limits the value of serum uric acid determinations for the diagnosis of gout. Nevertheless, serum urate levels are almost always elevated at some time and are important to use to follow the course of hypouricemic therapy. A 24-h urine collection for uric acid can, in some cases, be useful in assessing the risk of stones, elucidating overproduction or underexcretion of uric acid, and deciding whether it may be appropriate to use a uricosuric therapy (Chap. 359). Excretion of >800 mg of uric acid per 24 h on a regular diet suggests that causes of overproduction of purine should be considered. Urinalysis, serum creatinine, hemoglobin, white blood cell (WBC) count, liver function tests, and serum lipids should be obtained because of possible pathologic sequelae of gout and other associated diseases requiring treatment and as baselines because of possible adverse effects of gout treatment.

RADIOGRAPHIC FEATURES

Early in the disease radiographic studies may only confirm clinically evident swelling. Cystic changes, well-defined erosions with sclerotic margins (often with overhanging bony edges), and soft tissue masses are characteristic features of advanced chronic tophaceous gout. Ultrasound, CT and MRI are being studied and are likely to become more sensitive for early changes.

Treatment: Gout

ACUTE GOUTY ARTHRITIS

The mainstay of treatment during an acute attack is the administration of anti-inflammatory drugs such as nonsteroidal anti-inflammatory drugs (NSAIDs), colchicine, or glucocorticoids. NSAIDs are used most often in individuals without complicating comorbid conditions. Both colchicine and NSAIDs may be poorly tolerated and dangerous in the elderly and in the presence of renal insufficiency and gastrointestinal disorders. This was repeated later. Ice pack applications and rest of the involved joints can be helpful. Colchicine given orally is a traditional and effective treatment if used early in an attack. One useful regimen is one 0.6-mg tablet given every 8 h with subsequent tapering. This is generally

better tolerated than the formerly advised hourly regimen. The drug must be stopped promptly at the first sign of loose stools, and symptomatic treatment must be given for the diarrhea. Intravenous colchicine has been taken off the market. NSAIDs given in full anti-inflammatory doses are effective in ~90% of patients, and the resolution of signs and symptoms usually occurs in 5–8 days. The most effective drugs are any of those with a short half-life and include indomethacin, 25–50 mg tid; naproxen, 500 mg bid; ibuprofen, 800 mg tid; and diclofenac, 50 mg tid. Glucocorticoids given IM or orally, for example, prednisone, 30–50 mg/d as the initial dose and gradually tapered with the resolution of the attack, can be effective in polyarticular gout. For a single joint or a few involved joints intraarticular triamcinolone acetonide, 20–40 mg, or methylprednisolone, 25–50 mg, have been effective and well tolerated. Based on recent evidence on the essential role of the inflammasome and interleukin 1 β (IL-1β) in acute gout, anakinra has been used and other inhibitors of IL-1β are under investigation.

HYPOURICEMIC THERAPY

Ultimate control of gout requires correction of the basic underlying defect: the hyperuricemia. Attempts to normalize serum uric acid to <300–360 μmol/L (5.0–6.0 mg/dL) to prevent recurrent gouty attacks and eliminate tophaceous deposits entail a commitment to long-term hypouricemic regimens and medications that generally are required for life. Hypouricemic therapy should be considered when, as in most patients, the hyperuricemia cannot be corrected by simple means (control of body weight, low-purine diet, increase in liquid intake, limitation of ethanol use, decreased use of fructose-containing foods and beverages, and avoidance of diuretics). The decision to initiate hypouricemic therapy usually is made taking into consideration the number of acute attacks (urate lowering may be cost-effective after two attacks), serum uric acid levels [progression is more rapid in patients with serum uric acid >535 μmol/L (>9.0 mg/dL)], the patient's willingness to commit to lifelong therapy, or the presence of uric acid stones. Urate-lowering therapy should be initiated in any patient who already has tophi or chronic gouty arthritis. Uricosuric agents such as probenecid can be used in patients with good renal function who underexcrete uric acid, with <600 mg in a 24-h urine sample. Urine volume must be maintained by ingestion of 1500 mL of water every day. Probenecid can be started at a dose of 250 mg twice daily and increased gradually as needed up to 3 g per day to maintain a serum uric acid level <360 μmol/L (6 mg/dL). Probenecid is generally not effective in patients with serum creatinine levels >177 μmol/L (2 mg/dL). These patients may require allopurinol or benzbromarone (not available in the United States). Benzbromarone is another uricosuric drug that is more effective in patients with renal failure. Some agents used to treat common comorbidities, including losartan, fenofibrate, and amlodipine, have some mild uricosuric effects.

The xanthine oxidase inhibitor allopurinol is by far the most commonly used hypouricemic agent and is the best drug to lower serum urate in overproducers, urate stone formers, and patients with renal disease. It can be given in a single morning dose, 100–300 mg initially and increasing up to 800 mg if needed. In patients with chronic renal disease, the initial allopurinol dose should be lower and adjusted depending on the serum creatinine concentration; for example, with a creatinine clearance of 10 mL/min, one generally would use 100 mg every other day. Doses can be increased gradually to reach the target urate level of 6 mg/dL; however, more studies are needed to provide exact guidance. Toxicity of allopurinol has been recognized increasingly in patients who use thiazide diuretics and patients allergic to penicillin and ampicillin. The most serious side effects include life-threatening toxic epidermal necrolysis, systemic vasculitis, bone marrow suppression, granulomatous hepatitis, and renal failure. Patients with mild cutaneous reactions to allopurinol can reconsider the use of a uricosuric agent, undergo an attempt at desensitization to allopurinol, or take febuxostat, a new, chemically unrelated specific xanthine oxidase inhibitor. Febuxostat is approved at 40 or 80 mg once a day and does not require dose adjustment in mild to moderate renal disease. Patients can also pay increased attention to diet and should be aware of new alternative agents (see below). Urate-lowering drugs are generally not initiated during acute attacks but after the patient is stable and low-dose colchicine has been initiated to decrease the risk of the flares that often occur with urate lowering. Colchicine anti-inflammatory prophylaxis in doses of 0.6 mg one to two times daily should be given along with the hypouricemic therapy until the patient is normouricemic and without gouty attacks for 6 months or as long as tophi are present. Colchicine should not be used in dialysis patients and is given in lower doses in patients with renal disease or with P gylcoprotein or CYP3A4 inhibitors such as clarithromycin that can increase toxicity or colchicine. Pegloticase is a new urate-lowering biologic agent that can be effective in patients allergic to or failing xanthine oxidase inhibitors. New uricosurics are undergoing investigation.

CPPD DEPOSITION DISEASE

Pathogenesis

The deposition of CPPD crystals in articular tissues is most common in the elderly, occurring in 10–15% of persons age 65–75 years and 30–50% of those >85 years. In most cases, this process is asymptomatic and the cause of CPPD deposition is uncertain. Because >80% of patients are >60 years and 70% have preexisting joint damage from other conditions, it is likely that biochemical changes in aging or diseased cartilage favors crystal nucleation. In patients with CPPD arthritis there

is increased production of inorganic pyrophosphate and decreased levels of pyrophosphatases in cartilage extracts. Mutations in the *ANKH* gene, as described in both familial and sporadic cases, can increase elaboration and extracellular transport of pyrophosphate. The increase in pyrophosphate production appears to be related to enhanced activity of ATP pyrophosphohydrolase and 5'-nucleotidase, which catalyze the reaction of ATP to adenosine and pyrophosphate. This pyrophosphate could combine with calcium to form CPPD crystals in matrix vesicles or on collagen fibers. There are decreased levels of cartilage glycosaminoglycans that normally inhibit and regulate crystal nucleation. High activities of transglutaminase enzymes also may contribute to the deposition of CPPD crystals.

Release of CPPD crystals into the joint space is followed by the phagocytosis of those crystals by monocyte-macrophages and neutrophils, which respond by releasing chemotactic and inflammatory substances and, as with MSU crystals, activate the inflammasome.

A minority of patients with CPPD arthropathy have metabolic abnormalities or hereditary CPPD disease **(Table 333-2)**. These associations suggest that a variety of different metabolic products may enhance CPPD deposition either by directly altering cartilage or by inhibiting inorganic pyrophosphatases. Included among these conditions are hyperparathyroidism, hemochromatosis, hypophosphatasia, hypomagnesemia, and possibly myxedema. The presence of CPPD arthritis in individuals <50 years old should lead to consideration of these metabolic disorders (Table 333-2) and inherited forms of disease, including those identified in a variety of ethnic groups. Genomic DNA studies performed on different kindreds have shown a possible location of genetic defects on chromosome 8q or on chromosome 5p in a region that expresses the gene of the membrane pyrophosphate channel (*ANKH* gene). As noted above, mutations described in the *ANKH* gene in kindreds with CPPD arthritis can increase extracellular pyrophosphate and induce CPPD crystal formation. Investigation of younger patients with CPPD deposition should include inquiry for evidence of familial aggregation and evaluation of serum calcium, phosphorus, alkaline phosphatase, magnesium, serum iron, and transferrin.

Table 333-2 Conditions Associated with Calcium Pyrophosphate Dihydrate Disease

Aging
Disease-associated
Primary hyperparathyroidism
Hemochromatosis
Hypophosphatasia
Hypomagnesemia
Chronic gout
Postmeniscectomy
Gitelman's syndrome
Epiphyseal dysplasias

Clinical Manifestations

CPPD arthropathy may be asymptomatic, acute, subacute, or chronic or may cause acute synovitis superimposed on chronically involved joints. Acute CPPD arthritis originally was termed *pseudogout* by McCarty and co-workers because of its striking similarity to gout. Other clinical manifestations of CPPD deposition include (1) induction or enhancement of peculiar forms of osteoarthritis, (2) induction of severe destructive disease that may radiographically mimic neuropathic arthritis, (3) production of symmetric synovitis that is clinically similar to rheumatoid arthritis and sometimes seen in familial forms with early onset, (4) intervertebral disk and ligament calcification with restriction of spine mobility that mimics ankylosing spondylitis (also seen in hereditary forms), (5) spinal stenosis (most commonly seen in the elderly), and (6) rarely periarticular tophus-like nodules.

The knee is the joint most frequently affected in CPPD arthropathy. Other sites include the wrist, shoulder, ankle, elbow, and hands. The temporomandibular joint and ligamentum flavum of the spinal canal may be involved. Clinical and radiographic evidence indicates that CPPD deposition is polyarticular in at least two-thirds of patients. When the clinical picture resembles that of slowly progressive osteoarthritis, diagnosis may be difficult. Joint distribution may provide important clues suggesting CPPD disease. For example, primary osteoarthritis less often involves a metacarpophalangeal, wrist, elbow, shoulder, or ankle joints. If radiographs reveal punctate and/or linear radiodense deposits in fibrocartilaginous joint menisci or articular hyaline cartilage (*chondrocalcinosis*), the diagnostic likelihood of CPPD disease is further increased. *Definitive diagnosis* requires demonstration of typical rhomboid or rodlike crystals in synovial fluid or articular tissue **(Fig. 333-2)**. In the absence of joint effusion or indications to obtain a synovial biopsy,

chondrocalcinosis is presumptive of CPPD deposition. One exception is chondrocalcinosis due to CaOx in some patients with chronic renal failure.

Figure 333-2

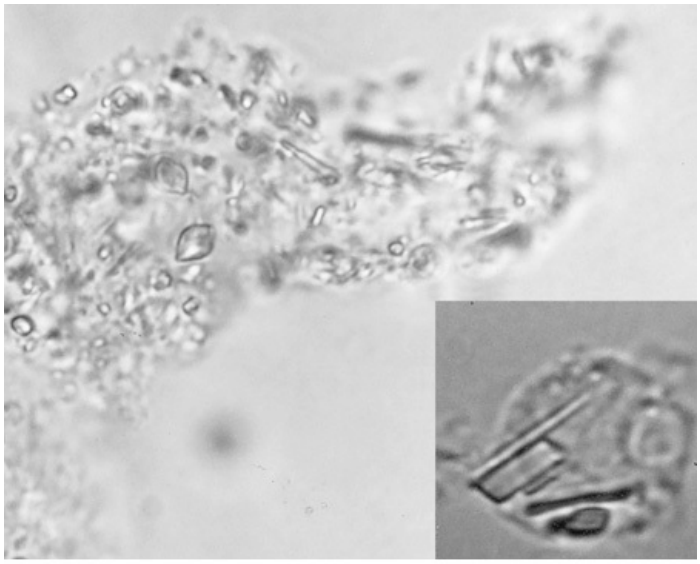

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Intracellular and extracellular calcium pyrophosphate dihydrate crystals, as seen in a fresh preparation of synovial fluid, illustrate rectangular, rod-shaped, and rhomboid crystals that are weakly positive birefringent crystals (compensated polarized light microscopy; 400x).

Acute attacks of CPPD arthritis may be precipitated by trauma. Rapid diminution of serum calcium concentration, as may occur in severe medical illness or after surgery (especially parathyroidectomy), can also lead to pseudogout attacks.

In as many as 50% of cases, episodes of CPPD-induced inflammation are associated with low-grade fever and, on occasion, temperatures as high as 40°C (104°F). Whether or not radiographic proof of chondrocalcinosis is evident in the involved joint(s), synovial fluid analysis with microbial cultures is essential to rule out the possibility of infection. In fact, infection in a joint with any microcrystalline deposition process can lead to crystal shedding and subsequent synovitis from both crystals and microorganisms. Synovial fluid in acute CPPD disease has inflammatory characteristics. The leukocyte count can range from several thousand cells to 100,000 cells/μL, with the mean being about 24,000 cells/μL and the predominant cell being the neutrophil. Polarized light microscopy usually reveals rhomboid, square, or rodlike crystals with weak positive birefringence inside tissue fragments and fibrin clots and in neutrophils (Fig. 333-2). CPPD crystals may coexist with MSU and apatite in some cases.

Treatment: Cppd Deposition Disease

Untreated acute attacks may last a few days to as long as a month. Treatment by joint aspiration and NSAIDs or by intraarticular glucocorticoid injection may result in return to prior status in ≤10 days. For patients with frequent recurrent attacks of pseudogout, daily prophylactic treatment with low doses of colchicine may be helpful in decreasing the frequency of the attacks. Severe polyarticular attacks usually require short courses of glucocorticoids or, as recently reported, an IL-1β antagonist, anakinra. Unfortunately, there is no effective way to remove CPPD deposits from cartilage and synovium. Uncontrolled studies suggest that the administration of antimalarial agents or even methotrexate may be helpful in controlling persistent synovitis. Patients with progressive destructive large-joint arthropathy may require joint replacement.

CALCIUM APATITE DEPOSITION DISEASE

Pathogenesis

Apatite is the primary mineral of normal bone and teeth. Abnormal accumulation of basic calcium phosphates, largely carbonate substituted apatite, can occur in areas of tissue damage (dystrophic calcification), hypercalcemic or hyperparathyroid states (metastatic calcification), and certain conditions of unknown cause **(Table 333-3)**. In chronic renal failure, hyperphosphatemia can contribute to extensive apatite deposition both in and around joints. Familial aggregation is rarely seen; no association with *ANKH* mutations has been described thus far. Apatite crystals are deposited primarily on matrix vessels. Incompletely understood alterations in matrix proteoglycans, phosphatases, hormones, and cytokines probably can influence crystal formation.

Table 333-3 Conditions Associated with Apatite Deposition Disease

Aging
Osteoarthritis
Hemorrhagic shoulder effusions in the elderly (Milwaukee shoulder)
Destructive arthropathy
Tendinitis, bursitis
Tumoral calcinosis (sporadic cases)
Disease-associated
Hyperparathyroidism
Milk-alkali syndrome
Renal failure/long-term dialysis
Connective tissue diseases (e.g., systemic sclerosis, idiopathicmyositis, SLE)
Heterotopic calcification after neurologic catastrophes (e.g., stroke, spinal cord injury)
Heredity
Bursitis, arthritis
Tumoral calcinosis
Fibrodysplasia ossificans progressiva

Abbreviation: SLE, systemic lupus erythematosus.

Apatite aggregates are commonly present in synovial fluid in an extremely destructive chronic arthropathy of the elderly that occurs most often in the shoulders (Milwaukee shoulder) and in a similar process in hips, knees, and erosive osteoarthritis of fingers. Joint destruction is associated with damage to cartilage and supporting structures, leading to instability and deformity. Progression tends to be indolent, and synovial fluid leukocyte counts are usually <2000/μL. Symptoms range from minimal to severe pain and disability that may lead to joint replacement surgery. Whether severely affected patients merely represent an extreme synovial tissue response to the apatite crystals that are so common in osteoarthritis is uncertain. Synovial lining cell or fibroblast cultures exposed to apatite (or CPPD) crystals can undergo mitosis and markedly increase the release of prostaglandin E2 and cytokines and also collagenases and neutral proteases, underscoring the destructive potential of abnormally stimulated synovial lining cells.

Clinical Manifestations

Periarticular or articular deposits may occur and may be associated with acute reversible inflammation and/or chronic damage to the joint capsule, tendons, bursa, or articular surfaces. The most common sites of apatite deposition include bursae and tendons in and/or around the knees, shoulders, hips, and fingers. Clinical manifestations include asymptomatic radiographic abnormalities, acute synovitis, bursitis, tendinitis, and chronic destructive arthropathy. Although the true incidence of apatite arthritis is not known, 30–50% of patients with osteoarthritis have apatite microcrystals in their synovial fluid. Such crystals frequently can be identified in clinically stable osteoarthritic joints, but they are more likely to come to attention in persons experiencing acute or subacute worsening of joint pain and swelling. The synovial fluid leukocyte count in apatite arthritis is usually low (<2000/μL) despite dramatic symptoms, with predominance of mononuclear cells.

Diagnosis

Intra- and/or periarticular calcifications with or without erosive, destructive, or hypertrophic changes may be seen on radiographs **(Fig. 333-3)**. They should be distinguished from the linear calcifications typical of CPPD deposition disease.

Figure 333-3

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

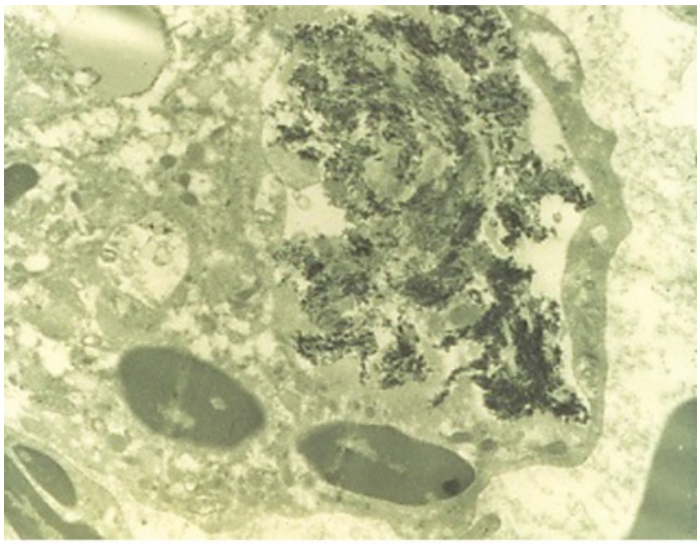

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. **Radiograph showing calcification** due to apatite crystals surrounding an eroded joint. ***B.*** An electron micrograph demonstrates dark needle-shaped apatite crystals within a vacuole of a synovial fluid mononuclear cell (30,000x).

Definitive diagnosis of apatite arthropathy, also called basic calcium phosphate disease, depends on identification of crystals from synovial fluid or tissue (Fig. 333-3). Individual crystals are very small and can be seen only by electron microscopy. Clumps of crystals may appear as 1- to 20-μm shiny intra- or extracellular nonbirefringent globules or aggregates that stain purplish with Wright's stain and bright red with alizarin red S. Tetracycline binding is under investigation as a labeling alternative. Absolute identification depends on electron microscopy with energy-dispersive elemental analysis, x-ray diffraction, infrared spectroscopy, or Raman microspectroscopy, but they usually are not required in clinical diagnosis.

Treatment: Calcium Apatite Deposition Disease

Treatment of apatite arthritis or periarthritis is nonspecific. Acute attacks of bursitis or synovitis may be self-limiting, resolving in days to several weeks. Aspiration of effusions and the use of either NSAIDs or oral colchicine for 2 weeks or intra- or periarticular injection of a depot glucocorticoid appear to shorten the duration and intensity of symptoms. Local injection of disodium ethylenediaminetetraacetic acid (EDTA) was effective in one study of calcific tendinitis at the shoulder. Periarticular apatite deposits

may be resorbed with resolution of attacks. Agents to lower serum phosphate levels may lead to resorption of deposits in renal failure patients receiving hemodialysis. In patients with underlying severe destructive articular changes, response to medical therapy is usually less rewarding.

CAOX DEPOSITION DISEASE

Pathogenesis

Primary oxalosis is a rare hereditary metabolic disorder (Chap. 364). Enhanced production of oxalic acid may result from at least two different enzyme defects, leading to hyperoxalemia and deposition of calcium oxalate crystals in tissues. Nephrocalcinosis, renal failure, and death usually occur before age 20. Acute and/or chronic CaOx arthritis and periarthritis may complicate primary oxalosis during later years of illness.

Secondary oxalosis is more common than the primary disorder. It is one of the many metabolic abnormalities that complicate end-stage renal disease. In chronic renal disease, calcium oxalate deposits have long been recognized in visceral organs, blood vessels, bones, and cartilage and are now known to be one of the causes of arthritis in chronic renal failure. Thus far, reported patients have been dependent on long-term hemodialysis or peritoneal dialysis (Chap. 281), and many had received ascorbic acid supplements. Ascorbic acid is metabolized to oxalate, which is inadequately cleared in uremia and by dialysis. Such supplements usually are avoided in dialysis programs because of the risk of enhancing hyperoxalosis and its sequelae.

Clinical Manifestations and Diagnosis

CaOx aggregates can be found in bone, articular cartilage, synovium, and periarticular tissues. From these sites, crystals may be shed, causing acute synovitis. Persistent aggregates of CaOx can, like apatite and CPPD, stimulate synovial cell proliferation and enzyme release, resulting in progressive articular destruction. Deposits have been documented in fingers, wrists, elbows, knees, ankles, and feet.

Clinical features of acute CaOx arthritis may not be distinguishable from those due to sodium urate, CPPD, or apatite. Radiographs may reveal chondrocalcinosis or soft tissue calcifications. CaOx-induced synovial effusions are usually noninflammatory, with <2000 leukocytes/μL, or mildly inflammatory. Neutrophils or mononuclear cells can predominate. CaOx crystals have a variable shape and variable birefringence to polarized light. The most easily recognized forms are bipyramidal, have strong birefringence **(Fig. 333-4)**, and stain with alizarin red S.

Figure 333-4

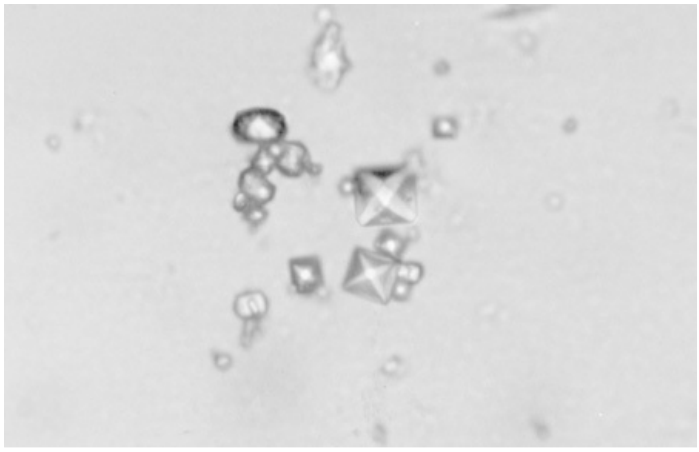

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bipyramidal and small polymorphic calcium oxalate crystals from synovial fluid are a classic finding in CaOx arthropathy (ordinary light microscopy; 400x).

Treatment: Calcium Oxalate Deposition Disease

Treatment of CaOx arthropathy with NSAIDs, colchicine, intraarticular glucocorticoids, and/or an increased frequency of dialysis has produced only slight improvement. In primary oxalosis, liver transplantation has induced a significant reduction in crystal deposits (Chap. 364).

ACKNOWLEDGMENT

This chapter has been revised for the previous edition and this edition from an original version written by Antonio Reginato, MD, in earlier editions of Harrison's Principles of Internal Medicine .

FURTHER READINGS

Maldonado I et al: Oxalate crystal deposition disease. Curr Rheumatol Rep 4:257, 2002[PMID: 12010612] [Full Text]

Molloy ES, McCarthy GM: Basic calcium phosphate crystals: Pathways to joint destruction. Curr Opinion Rheumatol 18:187, 2006[PMID: 16462527] [Full Text]

Reiders MK et al: A randomised controlled trial on the efficacy and tolerability with dose escalation of allopurinol 300–600 mg/day versus benzbromarone 100–200 mg/day in patients with gout. Ann Rheum Dis 68:892, 2009

Richette P et al: An update on the epidemiology of calcium pyrophosphate dihydrate crystal deposition disease. Rheumatology (Oxford) 48:711, 2009[PMID: 19398486] [Full Text]

Thouverey C et al: Inorganic pyrophosphate as a regulator of hydroxyapatite or calcium pyrophosphate dihydrate mineral deposition by matrix vesicles. Osteoarthritis Cartilage 17:64, 2009[PMID: 18603452] [Full Text]

Wortmann RL et al (eds): *Crystal-Induced Arthropathies: Gout, Pseudogout and Apatite-Associated Syndromes*. New York, Taylor & Francis, 2006

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INFECTIOUS ARTHRITIS: INTRODUCTION

Although *Staphylococcus aureus*, *Neisseria gonorrhoeae*, and other bacteria are the most common causes of infectious arthritis, various mycobacteria, spirochetes, fungi, and viruses also infect joints **(Table 334-1)**. Since acute bacterial infection can destroy articular cartilage rapidly, all inflamed joints must be evaluated without delay to exclude noninfectious processes and determine appropriate antimicrobial therapy and drainage procedures. For more detailed information on infectious arthritis caused by specific organisms, the reader is referred to the chapters on those organisms.

Table 334-1 Differential Diagnosis of Arthritis Syndromes

Acute Monarticular Arthritis	Chronic Monarticular Arthritis	Polyarticular Arthritis
Staphylococcus aureus	*Mycobacterium tuberculosis*	*Neisseria meningitidis*
Streptococcus pneumoniae	Nontuberculous mycobacteria	*N. gonorrhoeae*
β-Hemolytic streptococci	*Borrelia burgdorferi*	Nongonococcal bacterial arthritis
Gram-negative bacilli	*Treponema pallidum*	Bacterial endocarditis
Neisseria gonorrhoeae	*Candida* species	*Candida* species
Candida species	*Sporothrix schenckii*	Poncet's disease (tuberculous rheumatism)
Crystal-induced arthritis	*Coccidioides immitis*	Hepatitis B virus
Fracture	*Blastomyces dermatitidis*	Parvovirus B19
Hemarthrosis	*Aspergillus species*	HIV
Foreign body	*Cryptococcus neoformans*	Human T-lymphotropic virus type I
Osteoarthritis	*Nocardia* species	Rubella virus
Ischemic necrosis	*Brucella* species	Arthropod-borne viruses
Monarticular rheumatoid arthritis	Legg-Calvé-Perthes disease	Sickle cell disease flare
	Osteoarthritis	Reactive arthritis
		Serum sickness
		Acute rheumatic fever
		Inflammatory bowel disease
		Systemic lupus erythematosus
		Rheumatoid arthritis/Still's disease
		Other vasculitides
		Sarcoidosis

Acute bacterial infection typically involves a single joint or a few joints. Subacute or chronic monarthritis or oligoarthritis suggests mycobacterial or fungal infection; episodic inflammation is seen in syphilis, Lyme disease, and the reactive arthritis that follows enteric infections and chlamydial urethritis. Acute polyarticular inflammation occurs as an immunologic reaction during the course of endocarditis, rheumatic fever, disseminated neisserial infection, and acute hepatitis B. Bacteria and viruses occasionally infect multiple joints, the former most commonly in persons with rheumatoid arthritis.

Approach to the Patient: Infectious Arthritis

Aspiration of synovial fluid an essential element in the evaluation of potentially infected joints can be performed without difficulty in most cases by the insertion of a large-bore needle into the site of maximal fluctuance or tenderness or by the route of easiest access. Ultrasonography or fluoroscopy may be used to guide aspiration of difficult-to-localize effusions of the hip and, occasionally, the shoulder and other joints. Normal synovial fluid contains <180 cells (predominantly mononuclear cells) per microliter. Synovial cell counts averaging 100,000/μL (range, 25,000–250,000/μL), with >90%

neutrophils, are characteristic of acute bacterial infections. Crystal-induced, rheumatoid, and other noninfectious inflammatory arthritides usually are associated with <30,000–50,000 cells/μL; cell counts of 10,000–30,000/μL, with 50–70% neutrophils and the remainder lymphocytes, are common in mycobacterial and fungal infections. Definitive diagnosis of an infectious process relies on identification of the pathogen in stained smears of synovial fluid, isolation of the pathogen from cultures of synovial fluid and blood, or detection of microbial nucleic acids and proteins by nucleic acid amplification (NAA)–based assays and immunologic techniques.

ACUTE BACTERIAL ARTHRITIS

PATHOGENESIS

Bacteria enter the joint from the bloodstream; from a contiguous site of infection in bone or soft tissue; or by direct inoculation during surgery, injection, animal or human bite, or trauma. In hematogenous infection, bacteria escape from synovial capillaries, which have no limiting basement membrane, and within hours provoke neutrophilic infiltration of the synovium. Neutrophils and bacteria enter the joint space; later, bacteria adhere to articular cartilage. Degradation of cartilage begins within 48 h as a result of increased intraarticular pressure, release of proteases and cytokines from chondrocytes and synovial macrophages, and invasion of the cartilage by bacteria and inflammatory cells. Histologic studies reveal bacteria lining the synovium and cartilage as well as abscessesextending into the synovium, cartilage, and'in severe cases'subchondral bone. Synovial proliferation results in the formation of a pannus over the cartilage, and thrombosis of inflamed synovial vessels develops. Bacterial factors that appear important in the pathogenesis of infective arthritis include various surface-associated adhesins in *S. aureus* that permit adherence to cartilage and endotoxins that promote chondrocyte-mediated breakdown of cartilage.

MICROBIOLOGY

The hematogenous route of infection is the most common route in all age groups, and nearly every bacterial pathogen is capable of causing septic arthritis. In infants, group B streptococci, gram-negative enteric bacilli, and *S. aureus* are the most common pathogens. Since the advent of the *Haemophilus influenzae* vaccine, the predominant causes among children <5 years of age have been *S. aureus*, *Streptococcus pyogenes* (group A *Streptococcus*), and (in some centers) *Kingella kingae*. Among young adults and adolescents, *N. gonorrhoeae* is the most commonly implicated organism. *S. aureus* accounts for most nongonococcal isolates in adults of all ages; gram-negative bacilli, pneumococci, and β-hemolytic streptococci'particularly groups A and B but also groups C, G, and F'are involved in up to one-third of cases in older adults, especially those with underlying comorbid illnesses.

Infections after surgical procedures or penetrating injuries are due most often to *S. aureus* and occasionally to other gram-positive bacteria or gram-negative bacilli. Infections with coagulase-negative staphylococci are unusual except after the implantation of prosthetic joints or arthroscopy. Anaerobic organisms, often in association with aerobic or facultative bacteria, are found after human bites and when decubitus ulcers or intraabdominal abscesses spread into adjacent joints. Polymicrobial infections complicate traumatic injuries with extensive contamination. Bites and scratches from cats and other animals may introduce *Pasteurella multocida* into joints, and bites from humans may introduce *Eikenella corrodens* or other components of the oral flora.

NONGONOCOCCAL BACTERIAL ARTHRITIS

Epidemiology

Although hematogenous infections with virulent organisms such as *S. aureus*, *H. influenzae*, and pyogenic streptococci occur in healthy persons, there is an underlying host predisposition in many cases of septic arthritis. Patients with rheumatoid arthritis have the highest incidence of infective arthritis (most often secondary to *S. aureus*) because of chronically inflamed joints; glucocorticoid therapy; and frequent breakdown of rheumatoid nodules, vasculitic ulcers, and skin overlying deformed joints. Diabetes mellitus, glucocorticoid therapy, hemodialysis, and malignancy all carry an increased risk of infection with *S. aureus* and gram-negative bacilli. Tumor necrosis factor inhibitors (etanercept and infliximab), which increasingly are used for the treatment of rheumatoid arthritis, predispose to mycobacterial infections and possibly to other pyogenic bacterial infections and could be associated with septic arthritis in this population. Pneumococcal infections complicate alcoholism, deficiencies of humoral immunity, and hemoglobinopathies. Pneumococci, *Salmonella* species, and *H. influenzae* cause septic arthritis in persons infected with HIV. Persons with primary immunoglobulin deficiency are at risk for mycoplasmal arthritis, which results in permanent joint damage if tetracycline and replacement therapy with IV immunoglobulin are not administered promptly. IV drug users acquire staphylococcal and streptococcal infections from their own flora and acquire pseudomonal and other gram-negative infections from drugs and injection paraphernalia.

Clinical Manifestations

Some 90% of patients present with involvement of a single joint'most commonly the knee; less frequently the hip; and still less often the shoulder, wrist, or elbow. Small joints of the hands and feet

are more likely to be affected after direct inoculation or a bite. Among IV drug users, infections of the spine, sacroiliac joints, and sternoclavicular joints **(Fig. 334-1)** are more common than infections of the appendicular skeleton. Polyarticular infection is most common among patients with rheumatoid arthritis and may resemble a flare of the underlying disease.

Figure 334-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute septic arthritis of the sternoclavicular joint. A man in his forties with a history of cirrhosis presented with a new onset of fever and lower neck pain. He had no history of IV drug use or previous catheter placement. Jaundice and a painful swollen area over his left sternoclavicular joint were evident on physical examination. Cultures of blood drawn at admission grew group B *Streptococcus*. The patient recovered after treatment with IV penicillin. *(Courtesy of Francisco M. Marty, MD, Brigham and Women's Hospital, Boston; with permission.)*

The usual presentation consists of moderate to severe pain that is uniform around the joint, effusion, muscle spasm, and decreased range of motion. Fever in the range of 38.3°–38.9°C (101°–102°F) and sometimes higher is common but may not be present, especially in persons with rheumatoid arthritis, renal or hepatic insufficiency, or conditions requiring immunosuppressive therapy. The inflamed, swollen joint is usually evident on examination except in the case of a deeply situated joint such as the hip, shoulder, or sacroiliac joint. Cellulitis, bursitis, and acute osteomyelitis, which may produce a similar clinical picture, should be distinguished from septic arthritis by their greater range of motion and less than circumferential swelling. A focus of extraarticular infection such as a boil or pneumonia should be sought. Peripheral-blood leukocytosis with a left shift and elevation of the erythrocyte sedimentation rate or C-reactive protein level are common.

Plain radiographs show evidence of soft tissue swelling, joint-space widening, and displacement of tissue planes by the distended capsule. Narrowing of the joint space and bony erosions indicate advanced infection and a poor prognosis. Ultrasound is useful for detecting effusions in the hip, and CT or MRI can demonstrate infections of the sacroiliac joint, the sternoclavicular joint, and the spine very well.

Laboratory Findings

Specimens of peripheral blood and synovial fluid should be obtained before antibiotics are administered. Blood cultures are positive in up to 50–70% of *S. aureus* infections but are less frequently positive in infections due to other organisms. The synovial fluid is turbid, serosanguineous, or frankly purulent. Gram-stained smears confirm the presence of large numbers of neutrophils. Levels of total protein and lactate dehydrogenase in synovial fluid are elevated, and the glucose level is depressed; however, these findings are not specific for infection, and measurement of these levels is not necessary for diagnosis. The synovial fluid should be examined for crystals, because gout and pseudogout can resemble septic arthritis clinically, and infection and crystal-induced disease occasionally occur together. Organisms are seen on synovial fluid smears in nearly three-quarters of infections with *S. aureus* and streptococci and in 30–50% of infections due to gram-negative and other bacteria. Cultures of synovial fluid are positive in >90% of cases. Inoculation of synovial fluid into bottles containing liquid media for blood cultures increases the yield of a culture, especially if the pathogen is a fastidious organism or the patient is taking an antibiotic. Although not yet widely available, NAA-based assays for bacterial DNA will be useful for the diagnosis of partially treated or culture-negative bacterial arthritis.

Treatment: Nongonococcal Bacterial Arthritis

Prompt administration of systemic antibiotics and drainage of the involved joint can prevent destruction of cartilage, postinfectious degenerative arthritis, joint instability, or deformity. Once samples of blood and synovial fluid have been obtained for culture, empirical antibiotics should be given that are directed against the bacteria visualized on smears or the pathogens that are likely in light of the patient's age and risk factors. Initial therapy should consist of IV administration of bactericidal agents; direct instillation of antibiotics into the joint is not necessary to achieve adequate levels in synovial fluid and tissue. An IV third-generation cephalosporin such as cefotaxime (1 g every 8 h) or ceftriaxone (1–2 g every 24 h) provides adequate empirical coverage for most community-acquired infections in adults when smears show no organisms. IV vancomycin (1 g every 12 h) is used if there are gram-positive cocci on the smear. If methicillin-resistant *S. aureus* is an unlikely pathogen (e.g., when it is not widespread in the community), either oxacillin or nafcillin (2 g every 4 h) should be given. In addition, an aminoglycoside or third-generation cephalosporin should be given to IV drug users or other patients in whom *Pseudomonas aeruginosa* may be the responsible agent.

Definitive therapy is based on the identity and antibiotic susceptibility of the bacteria isolated in culture. Infections due to staphylococci are treated with oxacillin, nafcillin, or vancomycin for 4 weeks. Pneumococcal and streptococcal infections due to penicillin-susceptible organisms respond to 2 weeks of therapy with penicillin G (2 million units IV every 4 h); infections caused by *H. influenzae* and by strains of *Streptococcus pneumoniae* that are resistant to penicillin are treated with cefotaxime or ceftriaxone for 2 weeks. Most enteric gram-negative infections can be cured in 3–4 weeks by a second- or third-generation cephalosporin given IV or by a fluoroquinolone such as levofloxacin (500 mg IV or PO every 24 h). *P. aeruginosa* infection should be treated for at least 2 weeks with a combination regimen of an aminoglycoside plus either an extended-spectrum penicillin such as mezlocillin (3 g IV every 4 h) or an antipseudomonal cephalosporin such as ceftazidime (1 g IV every 8 h). If tolerated, this regimen is continued for an additional 2 weeks; alternatively, a fluoroquinolone such as ciprofloxacin (750 mg PO twice daily) is given by itself or with the penicillin or cephalosporin in place of the aminoglycoside.

Timely drainage of pus and necrotic debris from the infected joint is required for a favorable outcome. Needle aspiration of readily accessible joints such as the knee may be adequate if loculations or particulate matter in the joint does not prevent its thorough decompression. Arthroscopic drainage and lavage may be employed initially or within several days if repeated needle aspiration fails to relieve symptoms, decrease the volume of the effusion and the synovial white cell count, and clear bacteria from smears and cultures. In some cases, arthrotomy is necessary to remove loculations and debride infected synovium, cartilage, or bone. Septic arthritis of the hip is best managed with arthrotomy, particularly in young children, in whom infection threatens the viability of the femoral head. Septic joints do not require immobilization except for pain control before symptoms are alleviated by treatment. Weight bearing should be avoided until signs of inflammation have subsided, but frequent passive motion of the joint is indicated to maintain full mobility. Although addition of glucocorticoids to antibiotic treatment improves the outcome of *S. aureus* arthritis in experimental animals, no clinical trials have evaluated this approach in humans.

GONOCOCCAL ARTHRITIS

Epidemiology

Although its incidence has declined in recent years, gonococcal arthritis (Chap. 144) has accounted for up to 70% of episodes of infectious arthritis in persons <40 years of age in the United States. Arthritis due to *N. gonorrhoeae* is a consequence of bacteremia arising from gonococcal infection or, more frequently, from asymptomatic gonococcal mucosal colonization of the urethra, cervix, or pharynx. Women are at greatest risk during menses and during pregnancy and overall are two to three times more likely than men to develop disseminated gonococcal infection (DGI) and arthritis. Persons with complement deficiencies, especially of the terminal components, are prone to recurrent episodes of gonococcemia. Strains of gonococci that are most likely to cause DGI include those which produce transparent colonies in culture, have the type IA outer-membrane protein, or are of the AUH-auxotroph type.

Clinical Manifestations and Laboratory Findings

The most common manifestation of DGI is a syndrome of fever, chills, rash, and articular symptoms. Small numbers of papules that progress to hemorrhagic pustules develop on the trunk and the extensor surfaces of the distal extremities. Migratory arthritis and tenosynovitis of the knees, hands, wrists, feet, and ankles are prominent. The cutaneous lesions and articular findings are believed to be the consequence of an immune reaction to circulating gonococci and immune-complex deposition in tissues. Thus, cultures of synovial fluid are consistently negative, and blood cultures are positive in <45% of patients. Synovial fluid may be difficult to obtain from inflamed joints and usually contains only 10,000–20,000 leukocytes/μL.

True gonococcal septic arthritis is less common than the DGI syndrome and always follows DGI, which is unrecognized in one-third of patients. A single joint such as the hip, knee, ankle, or wrist is usually involved. Synovial fluid, which contains >50,000 leukocytes/μL, can be obtained with ease; the

gonococcus is only occasionally evident in gram-stained smears, and cultures of synovial fluid are positive in <40% of cases. Blood cultures are almost always negative.

Because it is difficult to isolate gonococci from synovial fluid and blood, specimens for culture should be obtained from potentially infected mucosal sites. Cultures and gram-stained smears of skin lesions are occasionally positive. All specimens for culture should be plated onto Thayer-Martin agar directly or in special transport media at the bedside and transferred promptly to the microbiology laboratory in an atmosphere of 5% CO_2, as generated in a candle jar. NAA-based assays are extremely sensitive in detecting gonococcal DNA in synovial fluid. A dramatic alleviation of symptoms within 12–24 h after the initiation of appropriate antibiotic therapy supports a clinical diagnosis of the DGI syndrome if cultures are negative.

Treatment: Gonococcal Arthritis

Initial treatment consists of ceftriaxone (1 g IV or IM every 24 h) to cover possible penicillin-resistant organisms. Once local and systemic signs are clearly resolving and if the sensitivity of the isolate permits, the 7-day course of therapy can be completed with an oral agent such as ciprofloxacin (500 mg twice daily). If penicillin-susceptible organisms are isolated, amoxicillin (500 mg three times daily) may be used. Suppurative arthritis usually responds to needle aspiration of involved joints and 7–14 days of antibiotic treatment. Arthroscopic lavage or arthrotomy is rarely required. Patients with DGI should be treated for *Chlamydia trachomatis* infection unless this infection is ruled out by appropriate testing.

It is noteworthy that arthritis symptoms similar to those seen in DGI occur in meningococcemia. A dermatitis-arthritis syndrome, purulent monarthritis, and reactive polyarthritis have been described. All respond to treatment with IV penicillin.

SPIROCHETAL ARTHRITIS

LYME DISEASE

Lyme disease (Chap. 173) due to infection with the spirochete *Borrelia burgdorferi* causes arthritis in up to 70% of persons who are not treated. Intermittent arthralgias and myalgias'but not arthritis'occur within days or weeks of inoculation of the spirochete by the *Ixodes* tick. Later, there are three patterns of joint disease: (1) Fifty percent of untreated persons experience intermittent episodes of monarthritis or oligoarthritis involving the knee and/or other large joints. The symptoms wax and wane without treatment over months, and each year 10–20% of patients report loss of joint symptoms. (2) Twenty percent of untreated persons develop a pattern of waxing and waning arthralgias. (3) Ten percent of untreated patients develop chronic inflammatory synovitis that results in erosive lesions and destruction of the joint. Serologic tests for IgG antibodies to *B. burgdorferi* are positive in >90% of persons with Lyme arthritis, and an NAA-based assay detects *Borrelia* DNA in 85%.

Treatment: Lyme Arthritis

Lyme arthritis generally responds well to therapy. A regimen of oral doxycycline (100 mg twice daily for 30 days), oral amoxicillin (500 mg four times daily for 30 days), or parenteral ceftriaxone (2 g/d for 2–4 weeks) is recommended. Patients who do not respond to a total of 2 months of oral therapy or 1 month of parenteral therapy are unlikely to benefit from additional antibiotic therapy and are treated with anti-inflammatory agents or synovectomy. Failure of therapy is associated with host features such as the human leukocyte antigen DR4 (HLA-DR4) genotype, persistent reactivity to OspA (outer-surface protein A), and the presence of hLFA-1 (human leukocyte function–associated antigen 1), which cross-reacts with OspA.

SYPHILITIC ARTHRITIS

Articular manifestations occur in different stages of syphilis (Chap. 169). In early congenital syphilis, periarticular swelling and immobilization of the involved limbs (Parrot's pseudoparalysis) complicate osteochondritis of long bones. Clutton's joint, a late manifestation of congenital syphilis that typically develops between ages 8 and 15 years, is caused by chronic painless synovitis with effusions of large joints, particularly the knees and elbows. Secondary syphilis may be associated with arthralgias, with symmetric arthritis of the knees and ankles and occasionally of the shoulders and wrists, and with sacroiliitis. The arthritis follows a subacute to chronic course with a mixed mononuclear and neutrophilic synovial-fluid pleocytosis (typical cell counts, 5000–15,000/μL). Immunologic mechanisms may contribute to the arthritis, and symptoms usually improve rapidly with penicillin therapy. In tertiary syphilis, Charcot's joint results from sensory loss due to tabes dorsalis. Penicillin is not helpful in this setting.

MYCOBACTERIAL ARTHRITIS

Tuberculous arthritis (Chap. 165) accounts for ~1% of all cases of tuberculosis and 10% of extrapulmonary cases. The most common presentation is chronic granulomatous monarthritis. An unusual syndrome, Poncet's disease, is a reactive symmetric form of polyarthritis that affects persons with visceral or disseminated tuberculosis. No mycobacteria are found in the joints, and symptoms

resolve with antituberculous therapy.

Unlike tuberculous osteomyelitis (Chap. 126), which typically involves the thoracic and lumbar spine (50% of cases), tuberculous arthritis primarily involves the large weight-bearing joints, in particular the hips, knees, and ankles, and only occasionally involves smaller non-weight-bearing joints. Progressive monarticular swelling and pain develop over months or years, and systemic symptoms are seen in only half of all cases. Tuberculous arthritis occurs as part of a disseminated primary infection or through late reactivation, often in persons with HIV infection or other immunocompromised hosts. Coexistent active pulmonary tuberculosis is unusual.

Aspiration of the involved joint yields fluid with an average cell count of 20,000/μL, with ~50% neutrophils. Acid-fast staining of the fluid yields positive results in fewer than one-third of cases, and cultures are positive in 80%. Culture of synovial tissue taken at biopsy is positive in ~90% of cases and shows granulomatous inflammation in most. NAA methods can shorten the time to diagnosis to 1 or 2 days. Radiographs reveal peripheral erosions at the points of synovial attachment, periarticular osteopenia, and eventually joint-space narrowing. Therapy for tuberculous arthritis is the same as that for tuberculous pulmonary disease, requiring the administration of multiple agents for 6–9 months. Therapy is more prolonged in immunosuppressed individuals such as those infected with HIV.

Various atypical mycobacteria (Chap. 167) found in water and soil may cause chronic indolent arthritis. Such disease results from trauma and direct inoculation associated with farming, gardening, or aquatic activities. Smaller joints, such as the digits, wrists, and knees, are usually involved. Involvement of tendon sheaths and bursae is typical. The mycobacterial species involved include *Mycobacterium marinum*, *M. avium-intracellulare*, *M. terrae*, *M. kansasii*, *M. fortuitum*, and *M. chelonae*. In persons who have HIV infection or are receiving immunosuppressive therapy, hematogenous spread to the joints has been reported for *M. kansasii*, *M. avium-intracellulare*, and *M. haemophilum*. Diagnosis usually requires biopsy and culture, and therapy is based on antimicrobial susceptibility patterns.

FUNGAL ARTHRITIS

Fungi are an unusual cause of chronic monarticular arthritis. Granulomatous articular infection with the endemic dimorphic fungi *Coccidioides immitis*, *Blastomyces dermatitidis*, and (less commonly) *Histoplasma capsulatum* **(Fig. 334-2)** results from hematogenous seeding or direct extension from bony lesions in persons with disseminated disease. Joint involvement is an unusual complication of sporotrichosis (infection with *Sporothrix schenckii*) among gardeners and other persons who work with soil or sphagnum moss. Articular sporotrichosis is six times more common among men than among women, and alcoholics and other debilitated hosts are at risk for polyarticular infection.

Figure 334-2

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

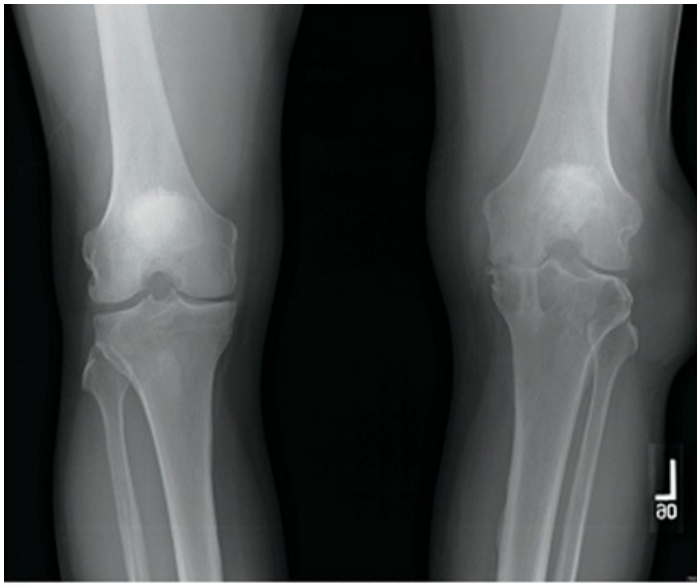

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

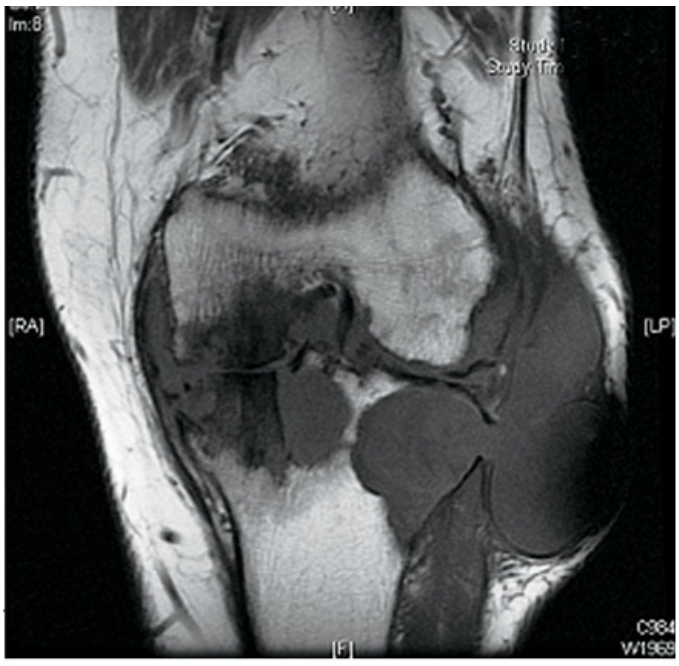

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chronic arthritis caused by *Histoplasma capsulatum* in the left knee. ***A.*** A man in his sixties from El Salvador presented with a history of progressive knee pain and difficulty walking for several years. He had undergone arthroscopy for a meniscal tear 7 years before presentation (without relief) and had received several intraarticular glucocorticoid injections. The patient developed significant deformity of the knee over time, including a large effusion in the lateral aspect. ***B.*** An x-ray of the knee showed multiple abnormalities, including severe medial femorotibial joint-space narrowing, several large subchondral cysts within the tibia and the patellofemoral compartment, a large suprapatellar joint effusion, and a large soft tissue mass projecting laterally over the knee. ***C.*** MRI further defined these abnormalities and demonstrated the cystic nature of the lateral knee abnormality. Synovial biopsies demonstrated chronic inflammation with giant cells, and cultures grew *H. capsulatum* after 3 weeks of incubation. All clinical cystic lesions and the effusion resolved after 1 year of treatment with itraconazole. The patient underwent a left total knee replacement for definitive treatment. *(Courtesy of Francisco M. Marty, MD, Brigham and Women's Hospital, Boston; with permission.)*

Candida infection involving a single joint'usually the knee, hip, or shoulder'results from surgical procedures, intraarticular injections, or (among critically ill patients with debilitating illnesses such as diabetes mellitus or hepatic or renal insufficiency and patients receiving immunosuppressive therapy)

hematogenous spread. *Candida* infections in IV drug users typically involve the spine, sacroiliac joints, or other fibrocartilaginous joints. Unusual cases of arthritis due to *Aspergillus* species, *Cryptococcus neoformans*, *Pseudallescheria boydii*, and the dematiaceous fungi also have resulted from direct inoculation or disseminated hematogenous infection in immunocompromised persons.

The synovial fluid in fungal arthritis usually contains 10,000–40,000 cells/μL, with ~70% neutrophils. Stained specimens and cultures of synovial tissue often confirm the diagnosis of fungal arthritis when studies of synovial fluid give negative results. Treatment consists of drainage and lavage of the joint and systemic administration of an antifungal agent directed at a specific pathogen. The doses and duration of therapy are the same as for disseminated disease (see Part 8, Section 16). Intraarticular instillation of amphotericin B has been used in addition to IV therapy.

VIRAL ARTHRITIS

Viruses produce arthritis by infecting synovial tissue during systemic infection or by provoking an immunologic reaction that involves joints. As many as 50% of women report persistent arthralgias, and 10% report frank arthritis within 3 days of the rash that follows natural infection with rubella virus and within 2–6 weeks after receipt of live-virus vaccine. Episodes of symmetric inflammation of fingers, wrists, and knees uncommonly recur for >1 year, but a syndrome of chronic fatigue, low-grade fever, headaches, and myalgias can persist for months or years. IV immunoglobulin has been helpful in selected cases. Self-limited monarticular or migratory polyarthritis may develop within 2 weeks of the parotitis of mumps; this sequela is more common among men than among women. Approximately 10% of children and 60% of women develop arthritis after infection with parvovirus B19. In adults, arthropathy sometimes occurs without fever or rash. Pain and stiffness, with less prominent swelling (primarily of the hands but also of the knees, wrists, and ankles), usually resolve within weeks, although a small proportion of patients develop chronic arthropathy.

About 2 weeks before the onset of jaundice, up to 10% of persons with acute hepatitis B develop an immune complex–mediated, serum sickness–like reaction with maculopapular rash, urticaria, fever, and arthralgias. Less common developments include symmetric arthritis involving the hands, wrists, elbows, or ankles and morning stiffness that resembles a flare of rheumatoid arthritis. Symptoms resolve at the time jaundice develops. Many persons with chronic hepatitis C infection report persistent arthralgia or arthritis, both in the presence and in the absence of cryoglobulinemia.

Painful arthritis involving larger joints often accompanies the fever and rash of several arthropod-borne viral infections, including those caused by chikungunya, O'nyong-nyong, Ross River, Mayaro, and Barmah Forest viruses (Chap. 196).Symmetric arthritis involving the hands and wrists may occur during the convalescent phase of infection with lymphocytic choriomeningitis virus. Patients infected with an enterovirus frequently report arthralgias, and echovirus has been isolated from patients with acute polyarthritis.

Several arthritis syndromes are associated with HIV infection. Reactive arthritis (Reiter's syndrome) with painful lower-extremity oligoarthritis often follows an episode of urethritis in HIV-infected persons. HIV-associated reactive arthritis appears to be extremely common among persons with the HLA-B27 haplotype, but sacroiliac joint disease is unusual and is seen mostly in the absence of HLA-B27. Up to one-third of HIV-infected persons with psoriasis develop psoriatic arthritis. Painless monarthropathy and persistent symmetric polyarthropathy occasionally complicate HIV infection. Chronic persistent oligoarthritis of the shoulders, wrists, hands, and knees occurs in women infected with human T cell lymphotropic virus type I. Synovial thickening, destruction of articular cartilage, and leukemic-appearing atypical lymphocytes in synovial fluid are characteristic, but progression to T cell leukemia is unusual.

PARASITIC ARTHRITIS

Arthritis due to parasitic infection is rare. The guinea worm *Dracunculus medinensis* may cause destructive joint lesions in the lower extremities as migrating gravid female worms invade joints or cause ulcers in adjacent soft tissues that become secondarily infected. Hydatid cysts infect bones in 1–2% of cases of infection with *Echinococcus granulosus*. The expanding destructive cystic lesions may spread to and destroy adjacent joints, particularly the hip and pelvis. In rare cases, chronic synovitis has been associated with the presence of schistosomal eggs in synovial biopsies. Monarticular arthritis in children with lymphatic filariasis appears to respond to therapy with diethylcarbamazine even in the absence of microfilariae in synovial fluid. Reactive arthritis has been attributed to hookworm, *Strongyloides*, *Cryptosporidium*, and *Giardia* infection in case reports, but confirmation is required.

POSTINFECTIOUS OR REACTIVE ARTHRITIS

Reactive polyarthritis develops several weeks after ~1% of cases of nongonococcal urethritis and 2% of enteric infections, particularly those due to *Yersinia enterocolitica*, *Shigella flexneri*, *Campylobacter jejuni* , and *Salmonella* species. Only a minority of these patients have the other findings of classic reactive arthritis, including urethritis, conjunctivitis, uveitis, oral ulcers, and rash.

Studies have identified microbial DNA or antigen in synovial fluid or blood, but the pathogenesis of this condition is poorly understood.

Reactive arthritis is most common among young men (except after *Yersinia* infection) and has been linked to the HLA-B27 locus as a potential genetic predisposing factor. Patients report painful, asymmetric oligoarthritis that affects mainly the knees, ankles, and feet. Low-back pain is common, and radiographic evidence of sacroiliitis is found in patients with long-standing disease. Most patients recover within 6 months, but prolonged recurrent disease is more common in cases that follow chlamydial urethritis. Anti-inflammatory agents help relieve symptoms, but the role of prolonged antibiotic therapy in eliminating microbial antigen from the synovium is controversial.

Migratory polyarthritis and fever constitute the usual presentation of acute rheumatic fever in adults (Chap. 322). This presentation is distinct from that of poststreptococcal reactive arthritis, which also follows infections with group A *Streptococcus* but is not migratory, lasts beyond the typical 3-week maximum of acute rheumatic fever, and responds poorly to aspirin.

INFECTIONS IN PROSTHETIC JOINTS

Infection complicates 1–4% of total joint replacements. The majority of infections are acquired intraoperatively or immediately postoperatively as a result of wound breakdown or infection; less commonly, these joint infections develop later after joint replacement and are the result of hematogenous spread or direct inoculation. The presentation may be acute, with fever, pain, and local signs of inflammation, especially in infections due to *S. aureus*, pyogenic streptococci, and enteric bacilli. Alternatively, infection may persist for months or years without causing constitutional symptoms when less virulent organisms, such as coagulase-negative staphylococci or diphtheroids, are involved. Such indolent infections usually are acquired during joint implantation and are discovered during evaluation of chronic unexplained pain or after a radiograph shows loosening of the prosthesis; the erythrocyte sedimentation rate and C-reactive protein level are usually elevated in such cases.

The diagnosis is best made by needle aspiration of the joint; accidental introduction of organisms during aspiration must be avoided meticulously. Synovial fluid pleocytosis with a predominance of polymorphonuclear leukocytes is highly suggestive of infection, since other inflammatory processes uncommonly affect prosthetic joints. Culture and Gram's stain usually yield the responsible pathogen. Sonication of explanted prosthetic material can improve the yield of culture, presumably by breaking up bacterial biofilms on the surfaces of prostheses. Use of special media for unusual pathogens such as fungi, atypical mycobacteria, and *Mycoplasma* may be necessary if routine and anaerobic cultures are negative.

Treatment: Prosthetic Joint Infections

Treatment includes surgery and high doses of parenteral antibiotics, which are given for 4–6 weeks because bone is usually involved. In most cases, the prosthesis must be replaced to cure the infection. Implantation of a new prosthesis is best delayed for several weeks or months because relapses of infection occur most commonly within this time frame. In some cases, reimplantation is not possible, and the patient must manage without a joint, with a fused joint, or even with amputation. Cure of infection without removal of the prosthesis is occasionally possible in cases that are due to streptococci or pneumococci and that lack radiologic evidence of loosening of the prosthesis. In these cases, antibiotic therapy must be initiated within several days of the onset of infection, and the joint should be drained vigorously by open arthrotomy or arthroscopically. In selected patients who prefer to avoid the high morbidity rate associated with joint removal and reimplantation, suppression of the infection with antibiotics may be a reasonable goal. A high cure rate with retention of the prosthesis has been reported when the combination of oral rifampin and ciprofloxacin is given for 3–6 months to persons with staphylococcal prosthetic joint infection of short duration. This approach, which is based on the ability of rifampin to kill organisms adherent to foreign material and in the stationary growth phase, requires confirmation in prospective trials.

PREVENTION

To avoid the disastrous consequences of infection, candidates for joint replacement should be selected with care. Rates of infection are particularly high among patients with rheumatoid arthritis, persons who have undergone previous surgery on the joint, and persons with medical conditions requiring immunosuppressive therapy. Perioperative antibiotic prophylaxis, usually with cefazolin, and measures to decrease intraoperative contamination, such as laminar flow, have lowered the rates of perioperative infection to <1% in many centers. After implantation, measures should be taken to prevent or rapidly treat extraarticular infections that might give rise to hematogenous spread to the prosthesis. The effectiveness of prophylactic antibiotics for the prevention of hematogenous infection after dental procedures has not been demonstrated; in fact, viridans streptococci and other components of the oral flora are extremely unusual causes of prosthetic joint infection. Accordingly, the American Dental Association and the American Academy of Orthopaedic Surgeons do not recommend antibiotic prophylaxis for most dental patients with total joint replacements. They do, however, recommend prophylaxis for patients who may be at high risk of hematogenous infection, including those with inflammatory arthropathies, immunosuppression, type 1 diabetes mellitus, joint replacement

within the preceding 2 years, previous prosthetic joint infection, malnourishment, or hemophilia. The recommended regimen is amoxicillin (2 g PO) 1 h before dental procedures associated with a high incidence of bacteremia. Clindamycin (600 mg PO) is suggested for patients allergic to penicillin.

ACKNOWLEDGMENTS

The contributions of James H. Maguire and the late Scott J. Thaler to this chapter in earlier editions are gratefully acknowledged.

FURTHER READINGS

Bardin T: Gonococcal arthritis. Best Pract Res Clin Rheumatol 17:201, 2003[PMID: 12787521] [Full Text]

Franssila R, Hedman K: Infection and musculoskeletal conditions: Viral causes of arthritis. Best Pract Res Clin Rheumatol 20:1139, 2006[PMID: 17127201] [Full Text]

Harrington JT: Mycobacterial and fungal arthritis. Curr Opin Rheumatol 10:335, 1998[PMID: 9725095] [Full Text]

Mathews CJ et al: Bacterial septic arthritis in adults. Lancet 375:846, 2010[PMID: 20206778] [Full Text]

Medina Rodriguez F: Rheumatic manifestations of human immunodeficiency virus infection. Rheum Dis Clin North Am 29:145, 2003

Meehan AM et al: Outcome of penicillin-susceptible streptococcal prosthetic joint infection treated with debridement and retention of the prosthesis. Clin Infect Dis 36:845, 2003[PMID: 12652384] [Full Text]

Moyad TF et al: Evaluation and management of the infected total hip and knee. Orthopedics 31:589, 2008

Shirtliff ME, Mader JT: Acute septic arthritis. Clin Microbiol Rev 15:527, 2002[PMID: 12364368] [Full Text]

Stengel D et al: Systematic review and meta-analysis of antibiotic therapy for bone and joint infections. Lancet Infect Dis 1:175, 2001[PMID: 11871494] [Full Text]

Tarkowski A: Infection and musculoskeletal conditions: Infectious arthritis. Best Pract Res Clin Rheumatol 20:1029, 2006[PMID: 17127195] [Full Text]

Trampuz A et al: Sonication of removed hip and knee prostheses for diagnosis of infection. N Engl J Med 357:654, 2007[PMID: 17699815] [Full Text]

Zimmerli W et al: Prosthetic-joint infections. N Engl J Med 351:145, 2004

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

FIBROMYALGIA: INTRODUCTION

Definition

Fibromyalgia (FM) is characterized by chronic widespread musculoskeletal pain and tenderness. Although it is defined primarily as a pain syndrome, FM patients also commonly complain of associated neuropsychological symptoms of fatigue, unrefreshing sleep, cognitive dysfunction, anxiety, and depression. Patients with FM have an increased prevalence of other syndromes associated with pain and fatigue, including chronic fatigue syndrome (Chap. 389), temporomandibular disorder, chronic headaches, irritable bowel syndrome, interstitial cystitis/painful bladder syndrome, and other pelvic pain syndromes. Available evidence implicates the central nervous system as key to maintaining pain and other core symptoms of FM and related conditions. The presence of FM is associated with substantial negative consequences for physical and social functioning.

Epidemiology

FM is far more common in women than in men, with a ratio of about 9:1. In population-based studies worldwide, there is general agreement that the prevalence rate is approximately 2–3%, with rates of closer to 5–10% in primary care practices. The prevalence data are similar across socioeconomic classes. Cultural factors may play a role in determining whether patients with FM symptoms seek medical attention; however, even in cultures in which secondary gain is not expected to play a significant role, the prevalence of FM remains in this range.

Clinical Manifestations

PAIN AND TENDERNESS

The most common presenting complaint of a patient with FM is "pain all over." Patients with FM have pain that is typically above and below the waist on both sides of the body and involves the axial skeleton (neck, back, or chest). The pain attributable to FM is poorly localized, difficult to ignore, severe in its intensity, and associated with a reduced functional capacity. Pain should have been present most of the day on most days for at least 3 months.

The clinical pain of FM is associated with increased evoked pain sensitivity. In clinical practice, this is determined by a tender point examination in which the examiner uses the thumbnail to exert pressure of approximately 4 kg/m^2, or the pressure leading to blanching of the tip of the thumbnail, on well-defined musculotendinous sites (**Fig. 335-1**). American College of Rheumatology classification criteria previously required that 11 of 18 sites be perceived as painful for a diagnosis of FM. In practice, tenderness is a continuous variable, and strict application of a categorical threshold for diagnosis specifics is no longer necessary. Increased pain sensitivity can be demonstrated not only for the mechanical pressure-induced pain used in the clinic but also for nonmuscular mechanical pressure, heat, cold, and other sensory stimuli; this reinforces the idea that the pathogenic mechanisms of FM are not related to specific musculoskeletal pathology but to altered pain processing. New criteria eliminate tender points and focus on clinical symptoms of widespread pain and neuropsychological symptoms.

Figure 335-1

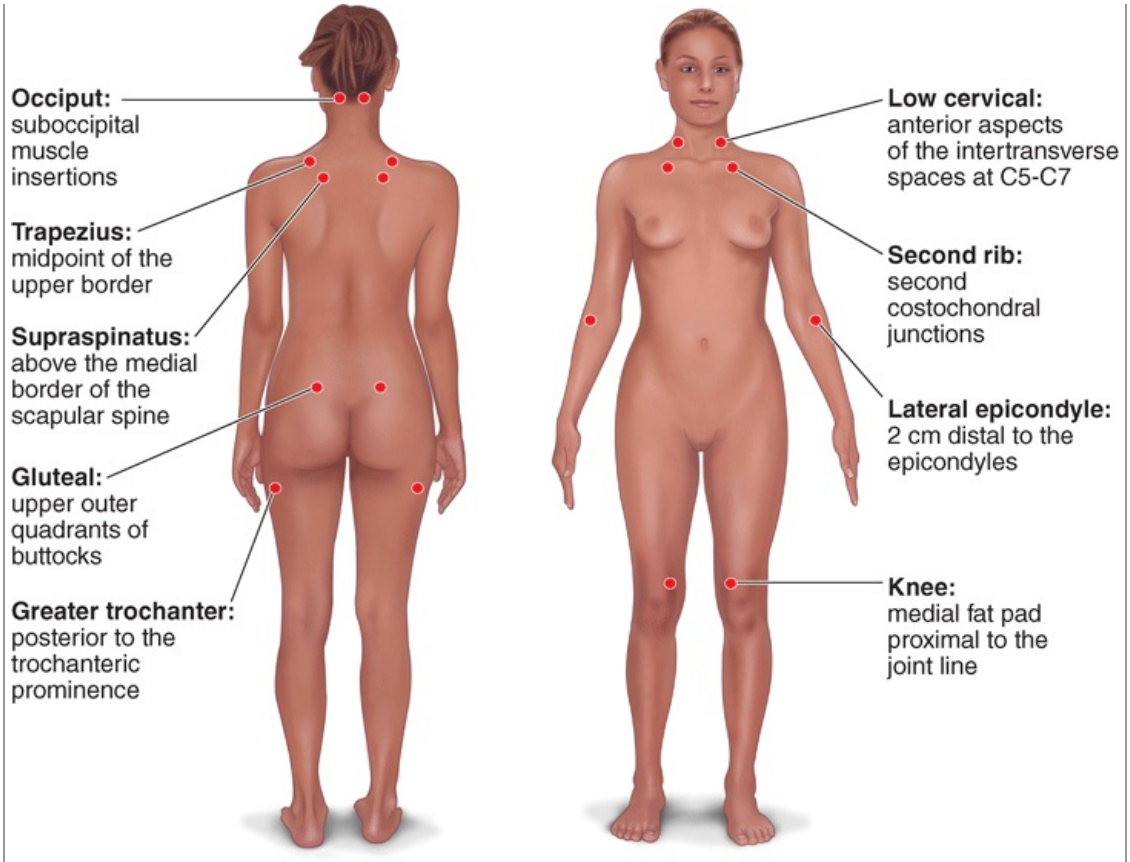

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tender point assessment in patients with fibromyalgia.

Patients with FM often have peripheral pain generators that are thought to serve as triggers for the more widespread pain attributed to central nervous system factors. Potential pain generators such as arthritis, bursitis, tendinitis, neuropathies, and other inflammatory or degenerative conditions should be identified by history and physical examination. More subtle pain generators may include joint hypermobility and scoliosis. Patients also may have chronic myalgias triggered by infectious, metabolic, or psychiatric conditions that can serve as triggers for the development of FM. These conditions are often in the differential diagnosis of patients with FM, and a major challenge is to distinguish the ongoing activity of a triggering condition from FM as a consequence of a comorbid condition that should itself be treated.

NEUROPSYCHOLOGICAL SYMPTOMS

In addition to widespread pain, FM patients typically complain of fatigue, stiffness, sleep disturbance, cognitive dysfunction, anxiety, and depression. These symptoms are present to varying degrees in most FM patients but are not present in every patient or at all times. Such symptoms may, however, have an equal or even greater impact on function and quality of life. Fatigue is highly prevalent in patients under primary care who ultimately are diagnosed with FM. Pain, stiffness, and fatigue often are worsened by exercise or unaccustomed activity (postexertional malaise). The sleep complaints include difficulty falling asleep, difficulty staying asleep, and early-morning awakening. Regardless of the specific complaint, patients awake feeling unrefreshed. Patients with FM may meet criteria for restless legs syndrome and sleep-disordered breathing; frank sleep apnea can also be present. Cognitive complaints are characterized as slowness in processing, difficulties with attention or concentration, problems with word retrieval, and short-term memory loss. Studies have demonstrated altered cognitive function in these domains in patients with FM, though speed of processing is age-appropriate. Symptoms of anxiety and depression are common, and the lifetime prevalence of mood disorders in patients with FM approaches 80%. Although depression is neither necessary nor sufficient for the diagnosis of FM, it is important to screen for major depressive disorders by querying for depressed mood and anhedonia. Analysis of genetic factors that are likely to predispose to FM reveals shared neurobiologic pathways with mood disorders, providing the basis for comorbidity.

OVERLAPPING SYNDROMES

Because FM can overlap in presentation with other chronic pain conditions, review of systems often reveals headaches, facial/jaw pain, regional myofascial pain particularly involving the neck or back, and arthritis. Visceral pain complaints involving the gastrointestinal tract, bladder, and pelvic or

perineal region are also often present. Patients may or may not meet defined criteria for specific syndromes. It is important for patients to understand that there may be shared pathways that mediate symptoms and that using treatment strategies effective for one condition may help with global symptom management.

COMORBID CONDITIONS

FM is often comorbid with chronic musculoskeletal, infectious, metabolic, or psychiatric conditions. Whereas FM is present in only 2–5% of the general population, it occurs in 20% or more of patients with degenerative or inflammatory rheumatic disorders, probably because these conditions serve as peripheral pain generators to alter central pain-processing pathways. Similarly, chronic infectious, metabolic, or psychiatric diseases associated with musculoskeletal pain can mimic FM and/or serve as a trigger for the development of FM. It is particularly important for clinicians to be sensitive to pain management of these comorbid conditions so that when FM emerges, as characterized by pain outside the boundaries of what could reasonably be explained by the triggering condition, development of neuropsychological symptoms, or tenderness on physical examination, treatment of central pain processes will be undertaken rather than continuing to focus on treating peripheral or inflammatory causes of pain.

PSYCHOSOCIAL CONSIDERATIONS

Symptoms of FM often have their onset and are exacerbated during periods of high levels of real or perceived stress. This may reflect an interaction between central stress physiology, vigilance or anxiety, and central pain-processing pathways. Understanding current psychosocial stressors will aid in patient management as many factors that exacerbate symptoms cannot be addressed by using pharmacologic approaches. Furthermore, there is a high prevalence of exposure to previous interpersonal and other forms of violence in patients with FM and related conditions. If posttraumatic stress disorder is an issue, the clinician should be aware of it and consider treatment options.

FUNCTIONAL IMPAIRMENT

It is crucial to evaluate the impact of FM symptoms on function and role fulfillment. In defining the success of a management strategy, improved function is a key measure. Functional assessment should include physical, mental, and social domains. Understanding where role functioning falls short will assist in establishing treatment goals.

Differential Diagnosis

Because musculoskeletal pain is such a common complaint, the differential diagnosis of FM is broad. **Table 335-1** lists some of the more common conditions that should be considered. Patients with inflammatory causes for widespread pain should be identifiable on the basis of specific history, physical findings, and laboratory or radiographic tests.

Table 335-1 Common Conditions in the Differential Diagnosis of Fibromyalgia

Inflammatory
Polymyalgia rheumatica
Inflammatory arthritis: rheumatoid arthritis, spondyloarthritides
Connective tissues diseases: systemic lupus erythematosus, Sjögren's syndrome
Infectious
Hepatitis C
Human immunodeficiency virus (HIV)
Lyme disease
Parvovirus B19
Epstein-Barr virus
Noninflammatory
Degenerative joint/spine/disk disease
Myofascial pain syndromes
Bursitis, tendinitis, repetitive strain injuries
Endocrine
Hypo- or hyperthyroidism
Hyperparathyroidism
Neurologic diseases
Multiple sclerosis

Neuropathic pain syndromes
Psychiatric disease
Major depressive disorder
Drugs
Statins
Aromatase inhibitors

Laboratory or Radiographic Testing

Routine laboratory and radiographic tests are normal in patients with FM, and so diagnostic testing is focused on excluding other diagnoses and evaluating for pain generators or comorbid conditions (**Table 335-2**). Most patients with a new complaint of chronic widespread pain should be assessed for the most common entities in the differential diagnosis. Radiographic testing should be used sparingly and only for diagnosis of inflammatory arthritis. After the patient has been evaluated thoroughly, repeat testing is discouraged unless the symptom complex changes. Particularly to be discouraged is advanced imaging (MRI) of the spine unless there are features suggesting inflammatory spine disease or neurologic symptoms.

Table 335-2 Laboratory and Radiographic Testing in Patients with Fibromyalgia Symptoms

Routine
Erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP)
Complete blood count (CBC)
Complete metabolic panel
Thyroid-stimulating hormone (TSH)
Guided by history and physical examination
Antinuclear antibody (ANA)
Anti-SSA (anti-Sjögren's syndrome A) and anti-SSB
Rheumatoid factor and anticyclic citrullinated peptide (anti-CCP)
Creatine phosphokinase (CPK)
Viral and bacterial serologies
Spine and joint radiographs

Genetics and Physiology

As in most complex diseases, it is likely that a number of genes contribute to vulnerability to the development of FM. To date, these genes appear to be in pathways controlling pain sensitivity and stress response. Some of the genetic underpinnings of FM are shared across other chronic pain conditions. For example, catechol-O-methyltransferase, which controls the synaptic levels of norepinephrine and dopamine, has been associated with pain sensitivity in the general population and certain polymorphisms or haplotypes have been associated with FM, chronic fatigue syndrome, and temporomandibular disorder. Polymorphisms of the β-adrenergic receptor and dopamine receptor are also associated with FM and other chronic pain conditions. Genes associated with metabolism, transport, and receptors of serotonin and other monoamines have also been implicated in FM and overlapping conditions. Taken together, the pathways in which polymorphisms have been identified in FM patients further implicate central factors as mediating the physiology that leads to FM clinical manifestations.

Psychophysical testing of patients with FM has demonstrated altered sensory afferent pain processing and impaired descending noxious inhibitory control leading to hyperalgesia and allodynia. Functional MRI and other research imaging procedures clearly demonstrate activation of the brain regions involved in the experience of pain in response to stimuli that are innocuous in study participants without FM. Pain perception in FM patients is influenced by the emotional and cognitive dimensions, such as catastrophizing and perceptions of control, providing a solid basis for recommendations for cognitive and behavioral treatment strategies.

Approach to the Patient: Fibromyalgia

FM occurs commonly and has an extraordinary impact on functioning and health-related quality of life; however, symptoms and impact can be managed effectively by physicians and other health professionals. Developing a partnership with patients with a goal of understanding and implementing a treatment strategy and choosing appropriate nonpharmacologic and pharmacologic treatments are

essential for improving the outcome of FM.

Treatment: Fibromyalgia

NONPHARMACOLOGIC TREATMENT

Patients with chronic pain, fatigue, and other neuropsychological symptoms require a framework for understanding the symptoms that have such an important impact on their function and quality of life. Providing explanation of the genetics, triggers, and physiology of FM can be an important adjunct in relieving the associated anxiety as well as reducing the overall cost of health care resources. In addition, patients must be educated regarding the expectations for treatment. The physician should focus on improved function and quality of life rather than elimination of pain. Illness behaviors should be discouraged, and behaviors that focus on improved function strongly encouraged.

Treatment strategies should include physical conditioning, with encouragement to begin at low levels of aerobic exercise with slow but consistent advancement. Patients who have been physically inactive or who report postexertional malaise may do best in supervised or water-based programs to start. Treatments that incorporate improved physical function with relaxation, such as yoga and Tai Chi, may also be helpful. Strength training may be recommended after a patient has reached his or her aerobic goals. Exercise programs are helpful for reductions in tenderness and for enhanced self-efficacy. Cognitive-behavioral strategies to improve sleep hygiene and reduce illness behaviors can also be helpful in management.

PHARMACOLOGIC APPROACHES

It is essential for the clinician to treat any comorbid triggering condition and clearly delineate for the patient the treatment goals for each medication. For example, glucocorticoids or nonsteroidal anti-inflammatory drugs may be useful for management of inflammatory triggers but are not effective for FM-related symptoms. At present, the treatment approaches that have proved most successful in FM patients target afferent or descending pain pathways. **Table 335-3** outlines the drugs with demonstrated effectiveness. It should be emphasized strongly that opioid analgesics are to be avoided in patients with FM. These agents have no demonstrated efficacy in FM and are associated with opioid-induced hyperalgesia that can worsen both symptoms and function. Utilization of single agents to treat multiple symptom domains is strongly encouraged. For example, if a patient's symptom complex is dominated by pain and sleep disturbance, using an agent that exerts both analgesic and sleep-promoting effects is desirable. These agents include sedating antidepressants such as amitriptyline or alpha-2-delta ligands such as gabapentin and pregabalin. For patients with pain associated with fatigue, anxiety, or depression, drugs that have both analgesic and antidepressant/anxiolytic effects, such as duloxetine or milnacipran, may be the best first choice.

Table 335-3 Pharmacologic Agents Effective for Treatment of Fibromyalgia

Antidepressants: balanced serotonin:norepinephrine reuptake inhibition
Amitryptiline
Duloxetinea
Milnaciprana
Anticonvulsants: ligands of the alpha-2-delta subunit of voltage-gated calcium channels
Gabapentin
Pregabalina

[a] Approved for fibromyalgia by the U.S. Food and Drug Administration.

FURTHER READINGS

Annemans L et al: Health economic consequences related to the diagnosis of fibromyalgia syndrome. Arthritis Rheum 58:895, 2008[PMID: 18311794] [Full Text]

Arnold LM: New therapies in fibromyalgia. Arthritis Res Ther 8:212, 2006[PMID: 16762044] [Full Text]

Boomershine CS, Crofford LJ: A symptom-based approach to pharmacologic management of fibromyalgia. Nat Rev Rheumatol 5:191, 2009[PMID: 19337283] [Full Text]

Busch AJ et al: Exercise for fibromyalgia: A systematic review. J Rheumatol 35:1130, 2008[PMID: 18464301] [Full Text]

Wolfe F et al: The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res 62:600, 2010[PMID: 20461783] [Full Text]

Williams DA, Clauw DJ: Understanding fibromyalgia: Lessons from the broader pain research community. J Pain 10:777, 2009[PMID: 19638325] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ARTHRITIS ASSOCIATED WITH SYSTEMIC DISEASE

Arthropathy of Acromegaly

Acromegaly is the result of excessive production of growth hormone by an adenoma in the anterior pituitary gland (Chap. 339). The excessive secretion of growth hormone along with insulin-like growth factor I stimulates proliferation of cartilage, periarticular connective tissue, and bone, resulting in several musculoskeletal problems, including osteoarthritis, back pain, muscle weakness, and carpal tunnel syndrome.

Osteoarthritis is a common feature, most often affecting the knees, shoulders, hips, and hands. Single or multiple joints may be affected. Hypertrophy of cartilage initially produces radiographic widening of the joint space. The newly synthesized cartilage is abnormally susceptible to fissuring, ulceration, and destruction. Ligamental laxity of joints further contributes to the development of osteoarthritis. Cartilage degrades, the joint space narrows, and subchondral sclerosis and osteophytes develop. Joint examination reveals crepitus and laxity. Joint fluid is noninflammatory. Calcium pyrophosphate dihydrate crystals are found in the cartilage in some cases of acromegalyarthropathy and, when shed into the joint, these can elicit attacks of pseudogout. Chondrocalcinosis may be observed on radiographs. Back pain is extremely common, perhaps as a result of spine hypermobility. Spine radiographs show normal or widened intervertebral disk spaces, hypertrophic anterior osteophytes, and ligamental calcification. These changes are similar to those observed in patients with diffuse idiopathic skeletal hyperostosis. Dorsal kyphosis in conjunction with elongation of the ribs contributes to the development of the barrel chest seen in acromegalic patients. The hands and feet become enlarged, owing to soft tissue proliferation. The fingers are thickened and have spadelike distal tufts. One-third of patients have a thickened heel pad. Approximately 25% of patients have Raynaud's phenomenon. Carpal tunnel syndrome occurs in about half of patients. The median nerve is compressed by excess connective tissue in the carpal tunnel. Patients with acromegaly could develop proximal muscle weakness, which is thought to be caused by the effect of growth hormone on muscle. Serum muscle enzyme levels and electromyography are normal. Muscle biopsy specimens show muscle fibers of varying size but with no inflammation.

Arthropathy of Hemochromatosis

Hemochromatosis is a disorder of iron storage. Excessive amounts of iron are absorbed from the intestine, leading to iron deposition in parenchymal cells, which results in impairment of organ function (Chap. 357). Symptoms of hemochromatosis usually begin between the ages of 40 and 60 but can occur earlier. Arthropathy, which occurs in 20–40% of patients, usually begins after the age of 50 and may be the first clinical feature of hemochromatosis. The arthropathy is an osteoarthritis-like disorder affecting the small joints of the hands, followed later by larger joints such as knees, ankles, shoulders, and hips. The second and third metacarpophalangeal joints of both hands are often the first and prominent joints affected; this may provide an important clue to the possibility of hemochromatosis because these joints are not predominantly affected by "routine" osteoarthritis. Patients experience some morning stiffness and pain with use of involved joints. The affected joints are enlarged and mildly tender. Radiographs show narrowing of the joint space, subchondral sclerosis, subchondral cysts, and juxtaarticular proliferation of bone with frequent hooklike osteophytes. The synovial fluid is noninflammatory. The synovium shows mild to moderate proliferation of iron containing lining cells, fibrosis, and some mononuclear cell infiltration. In approximately half of patients, there is evidence of calcium pyrophosphate deposition disease (CPPD), and some patients experience episodes of acute pseudogout.

Iron may damage the articular cartilage in several ways. Iron catalyzes superoxide-dependent lipid peroxidation, which may play a role in joint damage. In animal models, ferric iron has been shown to interfere with collagen formation and increase the release of lysosomal enzymes from cells in the synovial membrane. Iron inhibits synovial tissue pyrophosphatase in vitro and, therefore, may inhibit pyrophosphatase in vivo, resulting in chondrocalcinosis.

Treatment: Arthropathy of Hemochromatosis

The treatment of hemochromatosis is repeated phlebotomy. Unfortunately, this treatment has little effect on established arthritis, which, along with chondrocalcinosis, may progress. Symptomatic treatment of the arthritis consists of administration of acetaminophen and nonsteroidal anti-inflammatory drugs (NSAIDs), as tolerated. Acute pseudogout attacks are treated with high doses of an

NSAID or a short course of glucocorticoids. Hip or knee total joint replacement has been successful in advanced disease.

Hemophilic Arthropathy

Hemophilia is a sex-linked recessive genetic disorder characterized by the absence or deficiency of factor VIII (hemophilia A, or classic hemophilia) or factor IX (hemophilia B, or Christmas disease) (Chap. 116). Hemophilia A constitutes 85% of cases. Spontaneous hemarthrosis is a common problem with both types of hemophilia and can lead to a deforming arthritis. The frequency and severity of hemarthrosis are related to the degree of clotting factor deficiency. Hemarthrosis is not common in other disorders of coagulation such as von Willebrand disease, factor V deficiency, warfarin therapy, or thrombocytopenia.

Hemarthrosis occurs after one year of age, when the child begins to walk and run. In order of frequency, the joints most commonly affected are the knees, ankles, elbows, shoulders, and hips. Small joints of the hands and feet are occasionally involved.

In the initial stage of arthropathy, hemarthrosis produces a warm, tensely swollen, and painful joint. The patient holds the affected joint in flexion and guards against any movement. Blood in the joint remains liquid because of the absence of intrinsic clotting factors and the absence of tissue thromboplastin in the synovium. The synovial blood is resorbed over a period of a week or longer, depending on the size of the hemarthrosis. Joint function usually returns to normal or baseline in about two weeks. Low-grade temperature elevation may accompany hemarthrosis, but a fever >101 warrants concern for infection.

Recurrent hemarthrosis may result in chronic arthritis. The involved joints remain swollen, and flexion deformities develop. Joint motion may be restricted and function severely limited. Restricted joint motion, or laxity with subluxation, are features of end-stage disease.

Bleeding into muscle and soft tissue also causes musculoskeletal dysfunction. When bleeding into the iliopsoas muscle occurs, the hip is held in flexion because of the pain, resulting in a hip flexion contracture. Rotation of the hip is preserved, which distinguishes this problem from hemarthrosis or other causes of hip synovitis. Expansion of the hematoma may place pressure on the femoral nerve, resulting in a femoral neuropathy. Hemorrhage into a closed compartment space, such as the calf or volar compartment in the forearm, can result in muscle necrosis, neuropathy, and flexion deformities of the ankles, wrists, and fingers. When bleeding involves periosteum or bone, a painful pseudotumor forms. These occur distal to the elbows or knees in children and improve with treatment of the hemophilia. Surgical removal is indicated if the pseudotumor continues to enlarge. In adults, pseudotumors occur in the femur and pelvis and are usually refractory to treatment. When bleeding occurs in muscle, cysts may develop within the muscle. Needle aspiration of a cyst is contraindicated because it can induce further bleeding; however, if they become secondarily infected, drainage may be necessary (after factor repletion).

Septic arthritis rarely occurs in hemophilia and is difficult to distinguish from acute hemarthrosis on physical examination. If there is serious suspicion of an infected joint, the joint should be aspirated immediately, the fluid cultured, and the patient started on antibiotics that provide broad coverage, including *staphylococcus*, until the results of the culture return. Clotting-factor deficiency should be corrected before arthrocentesis to minimize the risk of traumatic bleeding.

Radiographs of joints reflect the stage of disease. In early stages, there is only capsule distention; later, juxtaarticular osteopenia, marginal erosions, and subchondral cysts develop. Late in the disease, the joint space is narrowed and there is bony overgrowth similar to osteoarthritis.

Treatment: Hemarthrosis

The treatment of musculoskeletal bleeding is initiated with the immediate infusion of factor VIII or IX at the first sign of joint or muscle hemorrhage. Patients who have developed factor inhibitors are at greater risk for joint damage and may benefit from receiving recombinant activated factor VII or activated prothrombin complex concentrate. The joint should be rested in a position of forced extension, as tolerated, to avoid contracture. Analgesia should be provided; ideally, the nonselective NSAIDs, which can diminish platelet function, should be avoided if possible. Selective cyclooxygenase-2 inhibitors do not interfere with platelet function and may be preferable based upon a demonstrated increased risk of upper GI bleeding and theoretical risk of increased articular bleeding with nonselective NSAIDs. Synovectomy, open or arthroscopic, may be attempted in patients with chronic symptomatic synovial proliferation and recurrent hemarthrosis, although hypertrophied synovium is very vascular and subject to bleeding. Both types of synovectomy reduce the number of hemarthroses. Open surgical synovectomy, however, is associated with some loss of range of motion. Both require aggressive prophylaxis against bleeding. Radiosynovectomy with either yttrium 90 silicate or phosphorus 31 colloid has been effective and may be attempted when surgical synovectomy is not practical. Total joint replacement is indicated for severe joint destruction and incapacitating pain.

Arthropathies Associated with Hemoglobinopathies

SICKLE CELL DISEASE

Sickle cell disease (Chap. 104) is associated with several musculoskeletal abnormalities **(Table 336-1)**. Children under the age of five may develop diffuse swelling, tenderness, and warmth of the hands and feet lasting from one to three weeks. The condition, referred to as *sickle cell dactylitis* or *hand-foot syndrome*, has also been observed in sickle cell thalassemia. Dactylitis is believed to result from infarction of the bone marrow and cortical bone leading to periostitis and soft tissue swelling. Radiographs show periosteal elevation, subperiosteal new bone formation, and areas of radiolucency and increased density involving the metacarpals, metatarsals, and proximal phalanges. These bone changes disappear after several months. The syndrome leaves little or no residual damage. Because hematopoiesis ceases in the small bones of hands and feet with age, the syndrome is rarely seen after age five.

Table 336-1 Musculoskeletal Abnormalities in Sickle Cell Disease

Sickle Cell Dactylitis	Avascular Necrosis
Joint effusions in sickle cell crises	Bone changes secondary to marrow hyperplasia
Osteomyelitis Infarction of bone	Septic arthritis Gouty arthritis
Infarction of bone marrow	

Sickle cell crisis is associated with periarticular pain and occasionally with joint effusions. The joint and periarticular area are warm and tender. Knees and elbows are most often affected, but other joints can be involved. Joint effusions are usually noninflammatory. Acute synovial infarction can cause a sterile effusion with high synovial fluid neutrophil counts. Synovial biopsies have shown mild lining cell proliferation and microvascular thrombosis with infarctions. Scintigraphic studies have shown decreased marrow uptake adjacent to the involved joint. The treatment is that for sickle cell crisis (Chap. 104).

Patients with sickle cell disease seem predisposed to osteomyelitis, which commonly involves the long tubular bones (Chap. 126) and *Salmonella* is a particularly frequent cause (Chap. 153). Radiographs of the involved site show periosteal elevation initially, followed by disruption of the cortex. Treatment of the infection results in healing of the bone lesion. Sickle cell disease is also more frequently associated with bone infarction resulting from vasoocclusion secondary to the sickling of red cells. Bone infarction also occurs in hemoglobin sickle cell disease and sickle cell thalassemia (Chap. 104). The bone pain in sickle cell crisis is due to bone and bone marrow infarction. In children, infarction of the epiphyseal growth plate interferes with normal growth of the affected extremity. Radiographically, infarction of the bone cortex results in periosteal elevation and irregular thickening of the bone cortex. Infarction in the bone marrow leads to lysis, fibrosis, and new bone formation. Clinical distinction between osteomyelitis and bone infarctions can be difficult; imaging can be helpful.

Avascular necrosis of the head of the femur occurs in ~5% of patients. It also occurs in the humeral head and less commonly in the distal femur, tibial condyles, distal radius, vertebral bodies, and other juxtaarticular sites. Irregularity of the femoral head and other articular surfaces often results in degenerative joint disease. Radiograph of the affected joint may show patchy radiolucency and density followed by flattening of the bone. MRI is a sensitive technique for detecting early avascular necrosis as well as bone infarction elsewhere. Total hip replacement and placement of prostheses in other joints may improve function and relieve the joint pain in these patients.

Septic arthritis is occasionally encountered in sickle cell disease (Chap. 334). Multiple joints may be infected. Joint infection may result from bacteremia due to splenic dysfunction or fromcontiguous osteomyelitis. The more common microorganisms include *Staphylococcus aureus*, *Streptococcus*, and *Salmonella*. Salmonella does not cause septic arthritis as frequently as osteomyelitis. Acute gouty arthritis is uncommon in sickle cell disease, even though 40% of patients are hyperuricemic. However, it may occur in patients generally not expected to get gout (younger, including female patients). Hyperuricemia is due to overproduction of uric acid secondary to increased red cell turnover as well as suboptimal renal excretion. Attacks may be polyarticular, and arthrocentesis should be performed as the diagnostic test to distinguish infection from gout or synovial infarction.

The bone marrow hyperplasia in sickle cell disease results in widening of the medullary cavities, thinning of the cortices, and coarse trabeculations and central cupping of the vertebral bodies. These changes are also seen to a lesser degree in hemoglobin sickle cell disease and sickle cell thalassemia. In normal individuals, red marrow is located mostly in the axial skeletal, but in sickle cell disease, red marrow is found in the bones of the extremities and even in the tarsal and carpal bones. Vertebral compression may lead to dorsal kyphosis, and softening of the bone in the acetabulum may result in protrusio acetabuli.

THALASSEMIA

β Thalassemia is a congenital disorder of hemoglobin synthesis characterized by impaired production of β chains (Chap. 104). Bone and joint abnormalities occur in β thalassemia, being most common in

the major and intermedia groups. In one study, ~50% of patients with β thalassemia had evidence of symmetric ankle arthropathy, characterized by a dull aching pain aggravated by weight bearing. The onset was most often in the second or third decade of life. The degree of ankle pain in these patients varied. Some patients experienced self-limited ankle pain, which occurred only after strenuous physical activity and lasted several days to weeks. Other patients had chronic ankle pain, which became worse with walking. Symptoms eventually abated in a few patients. Compression of the ankle, calcaneus, or forefoot was painful in some patients. Synovial fluid from two patients was noninflammatory. Radiographs of ankle showed osteopenia, widened medullary spaces, thin cortices, and coarse trabeculations. These findings are largely the result of bone marrow expansion. The joint space was preserved. Specimens of bone from three patients revealed osteomalacia, osteopenia, and microfractures. Increased osteoblasts as well as increased foci of bone resorption were present on the bone surface. Iron staining was found in the bone trabeculae, in osteoid, and in the cement line. Synovium showed hyperplasia of lining cells, which contained deposits of hemosiderin. This arthropathy was considered to be related to the underlying bone pathology. The role of iron overload or abnormal bone metabolism in the pathogenesis of this arthropathy is not known. The arthropathy was treated with analgesics and splints. Patients were also transfused to decrease hematopoiesis and bone marrow expansion.

Patients with β-thalassemia major and intermedia also have involvement of other joints, including the knees, hips, and shoulders. Acquired hemochromatosis with arthropathy has been described in a patient with thalassemia. Gouty arthritis and septic arthritis can occur. Avascular necrosis is not a feature of thalassemia because there is no sickling of red cells leading to thrombosis and infarction.

β-Thalassemia minor (trait) is also associated with joint manifestations. Chronic seronegative oligoarthritis affecting predominantly ankles, wrists, and elbows has been described. These patients had mild persistent synovitis without large effusions. Joint erosions were not seen. Recurrent episodes of an acute asymmetric arthritis have also been reported; episodes last less than a week and may affect knees, ankles, shoulders, elbows, wrists, and metacarpal phalangeal joints. The mechanism for this arthropathy is unknown. Treatment with nonsteroidal drugs was not particularly effective.

Musculoskeletal Disorders Associated with Hyperlipidemia (See Also Chap. 356)

Musculoskeletal or cutaneous manifestations may be the first clinical indication of a specific hereditary disorder of lipoprotein metabolism. Patients with familial hypercholesterolemia (previously referred to as *type II hyperlipoproteinemia*) may have recurrent migratory polyarthritis involving knees and other large peripheral joints and, to a lesser degree, peripheral small joints. Pain ranges from moderate to incapacitating. The involved joints can be warm, erythematous, swollen, and tender. Arthritis usually has a sudden onset, lasts from a few days to two weeks, and does not cause joint damage. Episodes may suggest acute gout attacks. Several attacks occur per year. Synovial fluid from involved joints is not inflammatory and contains few white cells and no crystals. Joint involvement may actually represent inflammatory periarthritis or peritendinitis and not true arthritis. The recurrent, transient nature of the arthritis may suggest rheumatic fever, especially because patients with hyperlipoproteinemia may have an elevated erythrocyte sedimentation rate and elevated antistreptolysin O titers, because the latter are quite common. Attacks of tendinitis, including the large Achilles and patellar tendons, may come on gradually and last only a few days or be acute as described above. Patients may be asymptomatic between attacks. Achilles tendinitis and other joint manifestations often precede the appearance of xanthomas and may be the first clinical indication of hyperlipoproteinemia. Attacks of tendinitis may occur following treatment with a lipid-lowering drug. Patients, over time, may develop tendinous xanthomas in the Achilles, patellar, and extensor tendons of the hands and feet. Xanthomas have also been reported in the peroneal tendon, the plantar aponeurosis, and the periosteum overlying the distal tibia. These xanthomas are located within tendon fibers. Tuberous xanthomas are soft subcutaneous masses located over the extensor surfaces of the elbows, knees, and hands, as well as on the buttocks. They appear in childhood in homozygous patients and, after the age of 30, in heterozygous patients. Patients with elevated plasma levels of very low density lipoprotein (VLDL) and triglyceride (previously referred to as *type IV hyperlipoproteinemia*)may also have a mild inflammatory arthritis affecting large and small peripheral joints, usually in an asymmetric pattern, with only a few joints involved at a time. The onset of arthritis is usually in middle age. Arthritis may be persistent or recurrent, with episodes lasting a few days to weeks. Joint pain is severe in some patients. Patients may experience morning stiffness. Joint tenderness and periarticular hyperesthesia may also be present, as may synovial thickening. Joint fluid is usually noninflammatory and without crystals but may have increased white blood cell counts with predominantly mononuclear cells. Radiographs may show juxtaarticular osteopenia and cystic lesions. Large bone cysts have been noted in a few patients. Xanthoma and bone cysts are also observed in other lipoprotein disorders. The pathogenesis of arthritis in patients with familial hypercholesterolemia or with elevated levels of VLDL and triglyceride is not well understood. NSAIDs or analgesics usually provide adequate relief of symptoms when used on an as-needed basis.

Clinical improvement may occur in patients as they are treated with lipid-lowering agents; however, patients treated with an HMG-CoA reductase inhibitor may experience myalgias, and a few patients may develop a myopathy, myositis, or even rhabdomyolysis. Patients who develop myositis while on

statin therapy may be susceptible to this side effect due to an underlying muscle disorder and should be reevaluated after discontinuation of the drug. Myositis has also been reported with the use of niacin (Chap. 388) but is less common than myalgias.

Musculoskeletal syndromes have not clearly been associated with the more common mixed hyperlipidemias seen in general practice.

OTHER ARTHRITIDES

Neuropathic Joint Disease

Neuropathic joint disease (Charcot's joint) is a progressive destructive arthritis associated with loss of pain sensation, proprioception, or both. Normal muscular reflexes that modulate joint movement are decreased. Without these protective mechanisms, joints are subjected to repeated trauma, resulting in progressive cartilage and bone damage. Neuropathic arthropathy was first described by Jean-Martin Charcot in 1868 in patients with tabes dorsalis. The term *Charcot joint* is commonly used interchangeably with *neuropathic joint*. Today, diabetes mellitus is the most frequent cause of neuropathic joint disease **(Fig. 336-1)**. A variety of other disorders are associated with neuropathic arthritis including leprosy, yaws, syringomyelia, meningomyelocele, congenital indifference to pain, peroneal muscular atrophy (Charcot-Marie-Tooth disease), and amyloidosis. An arthritis resembling neuropathic joint disease has been reported in patients who have received frequent intraarticular glucocorticoid injections into a weight-bearing joint, but this is a rare complication. The distribution of joint involvement depends on the underlying neurologic disorder **(Table 336-2)**. In tabes dorsalis, knees, hips, and ankles are most commonly affected; in syringomyelia, the glenohumeral joint, elbow, and wrist; and in diabetes mellitus, the tarsal and tarsometatarsal joints are most commonly affected.

Figure 336-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Charcot arthropathy associated with diabetes mellitus. Lateral foot radiograph demonstrating complete loss of the arch due to bony fragmentation and dislocation in the midfoot. (*Courtesy of Andrew Neckers, MD and Jean Schils, MD; with permission.)*

Table 336-2 Disorders Associated with Neuropathic Joint Disease

Diabetes mellitus	Amyloidosis
Tabes dorsalis	Leprosy
Meningomyelocele	Congenital indifference to pain
Syringomyelia	Peroneal muscular atrophy

PATHOLOGY AND PATHOPHYSIOLOGY

The pathologic changes in the neuropathic joint are similar to those found in the severe osteoarthritic joint. There is fragmentation and eventual loss of articular cartilage with eburnation of the underlying bone. Osteophytes are found at the joint margins. With more advanced disease, erosions are present on the joint surface. Fractures, devitalized bone, intraarticular loose bodies, and microscopic fragments of cartilage and bone may be present.

At least two underlying mechanisms are believed to be involved in the pathogenesis of neuropathic arthritis. An abnormal autonomic nervous system is thought to be responsible for the dysregulated blood flow to the joint with subsequent resorption of bone. Loss of bone, particularly in the diabetic

foot, may be the initial finding. With the loss of deep pain, proprioception, and protective neuromuscular reflexes, the joint is subjected to repeated microtrauma, resulting in ligamental tears and bone fractures. The mechanism of injury that occurs following frequent intraarticular glucocorticoid injections is thought to be due to the analgesic effect of glucocorticoids leading to overuse of an already damaged joint, which results in accelerated cartilage damage, although steroid-induced cartilage damage be more common in some animal species than in humans. It is not understood why only a few patients with neuropathy develop clinically evident neuropathic arthritis.

CLINICAL MANIFESTATIONS

Neuropathic joint disease usually begins in a single joint and then becomes apparent in other joints, depending on the underlying neurologic disorder. The involved joint progressively becomes enlarged due to bony overgrowth and synovial effusion. Loose bodies may be palpated in the joint cavity. Joint instability, subluxation, and crepitus occur as the disease progresses. Neuropathic joints may develop rapidly, and a totally disorganized joint with multiple bony fragments may evolve in a patient within weeks or months. The amount of pain experienced by the patient is less than would be anticipated based on the degree of joint damage. Patients may experience sudden joint pain from intraarticular fractures of osteophytes or condyles.

Neuropathic arthritis is encountered most often in patients with diabetes mellitus, with the incidence estimated in the range of 0.5%. The usual age of onset is ≤50 years, following several years of diabetes, but exceptions occur. The tarsal and tarsometatarsal joints are most often affected, followed by the metatarsophalangeal and talotibial joints. The knees and spine are occasionally involved. Patients often attribute the onset of foot pain to antecedent trauma such as twisting their foot. Neuropathic changes may develop rapidly following a foot fracture or dislocation. Swelling of the foot and ankle are often present. Downward collapse of the tarsal bones leads to convexity of the sole, referred to as a "rocker foot." Large osteophytes may protrude from the top of the foot. Calluses frequently form over the metatarsal heads and may lead to infected ulcers and osteomyelitis. The value of protective inserts and orthotics, as well as regular foot examination, cannot be overstated. Radiographs may show resorption and tapering of the distal metatarsal bones. The term *Lisfranc fracture-dislocation* is sometimes used to describe the destructive changes at the tarsometatarsal joints.

DIAGNOSIS

The diagnosis of neuropathic arthritis is based on the clinical features and characteristic radiographic findings in a patient with an underlying sensory neuropathy. The differential diagnosis of neuropathic arthritis depends upon the severity of the process and includes osteomyelitis, osteonecrosis, advanced osteoarthritis, stress fractures, and CPPD. Radiographs in neuropathic arthritis initially show changes of osteoarthritis with joint space narrowing, subchondral bone sclerosis, osteophytes, and joint effusions followed later by marked destructive and hypertrophic changes. The radiographic findings of neuropathic arthritis may be difficult to differentiate from those of osteomyelitis, especially in the diabetic foot. The joint margins in a neuropathic joint tend to be distinct, while in osteomyelitis, they are blurred. Imaging studies may be helpful, but cultures of tissue from the joint are often required to exclude osteomyelitis. MRI and bone scans using indium 111–labeled white blood cells or indium 111 –labeled immunoglobulin G, which will show an increased uptake in osteomyelitis but not in a neuropathic joint may be useful. A technetium bone scan will not distinguish osteomyelitis from neuropathic arthritis, as increased uptake is observed in both. The joint fluid in neuropathic arthritis is noninflammatory; may be xanthochromic or even bloody; and may contain fragments of synovium, cartilage, and bone. The finding of calcium pyrophosphate dihydrate crystals supports the diagnosis of crystal-associated arthropathy. In the absence of such crystals, an increased number of leukocytes may indicate osteomyelitis.

TREATMENT: NEUROPATHIC JOINT DISEASE

The primary focus of treatment is to stabilize the joint. Treatment of the underlying disorder, even if successful, does not usually affect established joint disease. Braces and splints are helpful. Their use requires close surveillance, because patients may be unable to appreciate pressure from a poorly adjusted brace. In the diabetic patient, early recognition and treatment of a Charcot's foot by prohibiting weight bearing of the foot for at least eight weeks may possibly prevent severe disease from developing. Fusion of an unstable joint may improve function and reduce pain, but nonunion is frequent, especially when immobilization of the joint is inadequate.

Hypertrophic Osteoarthropathy and Clubbing

Hypertrophic osteoarthropathy (HOA) is characterized by clubbing of digits and, in more advanced stages, by periosteal new bone formation and synovial effusions. HOA may be primary or familial and begin in childhood. Secondary HOA is associated with intrathoracic malignancies, suppurative and some hypoxemic lung diseases, congenital heart disease, and a variety of other disorders. Clubbing is almost always a feature of HOA but can occur as an isolated manifestation **(Fig. 336-2)**. The presence of clubbing in isolation may be congenital or represent either an early stage or one element in the spectrum of HOA. The presence of isolated acquired clubbing has the same clinical significance as

clubbing associated with periostitis.

Figure 336-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clubbing of fingers. *(Reprinted from the Clinical Slide Collection on the Rheumatic Diseases, Copyright 1991, 1995. Used by permission of the American College of Rheumatology.)*

PATHOLOGY AND PATHOPHYSIOLOGY OF ACQUIRED HOA

In HOA, the bone changes in the distal extremities begin as periostitis followed by new bone formation. At this stage, a radiolucent area may be observed between the new periosteal bone and subjacent cortex. As the process progresses, multiple layers of new bone are deposited, which become contiguous with the cortex and result in cortical thickening. The outer portion of bone is laminated in appearance, with an irregular surface. Initially, the process of periosteal new bone formation involves the proximal and distal diaphyses of the tibia, fibula, radius, and ulna and, less frequently, the femur, humerus, metacarpals, metatarsals, and phalanges. Occasionally, scapulae, clavicles, ribs, and pelvic bones are also affected. The adjacent interosseous membranes may become ossified. The distribution of the bone manifestations is usually bilateral and symmetric. The soft tissue overlying the distal third of the arms and legs may be thickened. Proliferation of connective tissue occurs in the nail bed and volar pad of digits, giving the distal phalanges a clubbed appearance. Small blood vessels in the clubbed digits are dilated and have thickened walls. In addition, the number of arteriovenous anastomoses is increased.

Several theories have been suggested for the pathogenesis of HOA, but many have been disproved or have not explained the development in all clinical disorders associated with HOA. Previously proposed neurogenic and humoral theories are no longer considered likely explanations for HOA. Recent studies have suggested a role for platelets in the development of HOA. It has been observed that megakaryocytes and large platelet particles, present in venous circulation, were fragmented in their passage through normal lung. In patients with cyanotic congenital heart disease and in other disorders associated with right-to-left shunts, these large platelet particles bypass the lung and reach the distal extremities, where they can interact with endothelial cells. Platelet-endothelial activation in the distal portion of extremities may result in the release of platelet-derived growth factor (PDGF) and other factors leading to the proliferation of connective tissue and periosteum. Stimulation of fibroblasts by PDGF and transforming growth factor β results in cell growth and collagen synthesis. Elevated plasma levels of von Willebrand factor antigen have been found in patients with both primary and secondary forms of HOA, indicating endothelial activation or damage. Abnormalities of collagen synthesis have been demonstrated in the involved skin of patients with primary HOA. Other factors are undoubtedly involved in the pathogenesis of HOA, and further studies are needed to better understand this disorder.

CLINICAL MANIFESTATIONS

Primary or familial HOA, also referred to as *pachydermoperiostitis* or *Touraine-Solente-Golé syndrome*, usually begins insidiously at puberty. In a smaller number of patients, the onset is in the first year of life. The disorder is inherited as an autosomal dominant trait with variable expression and is nine times more common in boys than in girls. Approximately one-third of patients have a family history of primary HOA.

Primary HOA is characterized by clubbing, periostitis, and unusual skin features. A small number of patients with this syndrome do not express clubbing. The skin changes and periostitis are prominent features of this syndrome. The skin becomes thickened and coarse. Deep nasolabial folds develop, and the forehead may become furrowed. Patients may have heavy-appearing eyelids and ptosis. The skin is often greasy, and there may be excessive sweating of the hands and feet. Patients may also

experience acnevulgaris, seborrhea, and folliculitis. In a few patients, the skin over the scalp becomes very thick and corrugated, a feature that has been descriptively termed *cutis verticis gyrata*. The distal extremities, particularly the legs, become thickened owing to proliferation of new bone and soft tissue; when the process is extensive, the distal lower extremities resemble those of an elephant. The periostitis is usually not painful, because it may be in secondary HOA. Clubbing of the fingers may be extensive, producing large, bulbous deformities and clumsiness. Clubbing also affects the toes. Patients may experience articular and periarticular pain, especially in the ankles and knees, and joint motion may be mildly restricted owing to periarticular bone overgrowth. Noninflammatory effusions occur in the wrists, knees, and ankles. Synovial hypertrophy is not found. Associated abnormalities observed in patients with primary HOA include hypertrophic gastropathy, bone marrow failure, female escutcheon, gynecomastia, and cranial suture defects. In patients with primary HOA, the symptoms disappear when adulthood is reached.

HOA secondary to an underlying disease occurs more frequently than primary HOA. It accompanies a variety of disorders and may precede clinical features of the associated disorder by months. Clubbing is more frequent than the full syndrome of HOA in patients with associated illnesses. Because clubbing evolves over months and is usually asymptomatic, it is often recognized first by the physician and not the patient. Patients may experience a burning sensation in their fingertips. Clubbing is characterized by widening of the fingertips, enlargement of the distal volar pad, convexity of the nail contour, and the loss of the normal 15° angle between the proximal nail and cuticle. The thickness of the digit at the base of the nail is greater than the thickness at the distal interphalangeal joint. An objective measurement of finger clubbing can be made by determining the diameter at the base of the nail and at the distal interphalangeal joint of all 10 digits. Clubbing is present when the sum of the individual digit ratios is >10. At the bedside, clubbing can be appreciated by having the patient place the dorsal surface of the distal phalanges of the fourth fingers together with the nails of the fourth fingers opposing each other. Normally, an open area is visible between the bases of the opposing fingernails; when clubbing is present, this open space is no longer visible. The base of the nail feels spongy when compressed, and the nail can be easily rocked on its bed. When clubbing is advanced, the finger may have a drumstick appearance, and the distal interphalangeal joint can be hyperextended. Periosteal involvement in the distal extremities may produce a burning or deep-seated aching pain. The pain can be quite incapacitating and is aggravated by dependency and relieved by elevation of the affected limbs. Pressure applied over the distal forearms and legs or gentle percussion of the distal long bones like the tibia may be quite painful.

Patients may experience joint pain, most often in the ankles, wrists, and knees. Joint effusions may be present; usually, they are small and noninflammatory. The small joints of the hands are rarely affected. Severe joint or long bone pain may be the presenting symptom of an underlying lung malignancy and may precede the appearance of clubbing. In addition, the progression of HOA tends to be more rapid when associated with malignancies, most notably bronchogenic carcinoma. Noninflammatory but variably painful knee effusions may occur prior to the appearance of clubbing and symptoms of distal periostitis. Unlike primary HOA, excessive sweating and oiliness of the skin and thickening of the facial skin are uncommon in secondary HOA.

HOA occurs in 5–10% of patients with intrathoracic malignancies, the most common being bronchogenic carcinoma and pleural tumors **(Table 336-3)**. Lung metastases infrequently cause HOA. HOA is also seen in patients with intrathoracic infections, including lung abscesses, empyema, bronchiectasis, and chronic obstructive lung disease, but uncommonly in pulmonary tuberculosis. HOA may also accompany chronic interstitial pneumonitis, sarcoidosis, and cystic fibrosis. In the latter, clubbing is more common than the full syndrome of HOA. Other causes of clubbing include congenital heart disease with right-to-left shunts, bacterial endocarditis, Crohn's disease, ulcerative colitis, sprue, and neoplasms of the esophagus, liver, and small and large bowel. In patients with congenital heart disease with right-to-left shunts, clubbing alone occurs more often than the full syndrome of HOA.

Table 336-3 Disorders Associated with Hypertrophic Osteoarthropathy

Pulmonary	Cardiovascular
Bronchogenic carcinoma and other neoplasms	Cyanotic congenital heart disease
Bronchogenic carcinoma and other neoplasms	Subacute bacterial endocarditis
Lung abscesses, empyema, bronchiectasisChronic interstitial pneumonitisCystic fibrosis	Infected arterial grafts[a]
Chronic obstructive lung disease	Aortic aneurysm[b]
Sarcoidosis	Aneurysm of major extremity artery[a]
Gastrointestinal	Patent ductus arteriosus[b]
Inflammatory bowel disease	Arteriovenous fistula of major
Sprue	

Neoplasms: esophagus, liver, bowel	Arteriovenous fistula of major extremity vessel[a] Thyroid (thyroid acropachy) Hyperthyroidism (Graves' disease)

[a]Unilateral involvement.

[b]Bilateral lower extremity involvement.

Unilateral clubbing has been found in association with aneurysms of major extremity arteries, with infected arterial grafts, and with arteriovenous fistulas of brachial vessels. Clubbing of the toes but not fingers has been associated with an infected abdominal aortic aneurysm and patent ductus arteriosus. Clubbing of a single digit may follow trauma and has been reported in tophaceous gout and sarcoidosis. While clubbing occurs more commonly than the full syndrome in most diseases, periostitis in the absence of clubbing has been observed in the affected limb of patients with infected arterial grafts.

Hyperthyroidism (Graves' disease), treated or untreated, is occasionally associated with clubbing and periostitis of the bones of the hands and feet. This condition is referred to as *thyroid acropachy*. Periostitis may be asymptomatic and occurs in the midshaft and diaphyseal portion of the metacarpal and phalangeal bones. Significant hand joint pain may occur; this may respond to successful therapy of the thyroid dysfunction. The long bones of the extremities are seldom affected. Elevated levels of long-acting thyroid stimulator are found in the serum of these patients.

LABORATORY FINDINGS

The laboratory abnormalities reflect the underlying disorder. The synovial fluid of involved joints has <500 white cells per microliter, and the cells are predominantly mononuclear. Radiographs show a faint radiolucent line beneath the new periosteal bone along the shaft of long bones at their distal end. These changes are observed most frequently at the ankles, wrists, and knees. The ends of the distal phalanges may show osseous resorption. Radionuclide studies show pericortical linear uptake along the cortical margins of long bones that may be present before any radiographic changes.

TREATMENT HYPERTROPHIC OSTEOARTHROPATHY

The treatment of HOA is to identify the associated disorder and treat it appropriately. The symptoms and signs of HOA may disappear completely with removal or effective chemotherapy of a tumor or with antibiotic therapy and drainage of a chronic pulmonary infection. Vagotomy or percutaneous block of the vagus nerve leads to symptomatic relief in some patients. NSAIDs or analgesics may help control symptoms of HOA.

Reflex Sympathetic Dystrophy Syndrome

The reflex sympathetic dystrophy syndrome is now referred to as *complex regional pain syndrome, type 1*, by the new Classification of the International Association for the Study of Pain. It is characterized by pain and swelling, usually of a distal extremity, accompanied by vasomotor instability, trophic skin changes, and the rapid development of bony demineralization. Reflex sympathetic dystrophy syndrome, including its treatment, is covered in greater detail in Chap. 376.

Tietze Syndrome and Costochondritis

Tietze syndrome is manifested by painful swelling of one or more costochondral articulations. The age of onset is usually before 40, and both sexes are affected equally. In most patients, only one joint is involved, usually the second or third costochondral joint. The onset of anterior chest pain may be sudden or gradual. The pain may radiate to the arms or shoulders and is aggravated by sneezing, coughing, deep inspirations, or twisting motions of the chest. The term *costochondritis* is often used interchangeably with *Tietze syndrome*, but some workers restrict the former term to pain of the costochondral articulations without swelling. Costochondritis is observed in patients over age 40; tends to affect the third, fourth, and fifth costochondral joints; and occurs more often in women. Both syndromes may mimic cardiac or upper abdominal causes of pain. Rheumatoid arthritis, ankylosing spondylitis, and reactive arthritis (Reiter's syndrome) may involve costochondral joints but are distinguished easily by their other clinical features. Other skeletal causes of anterior chest wall pain are xiphoidalgia and the slipping rib syndrome, which usually involves the tenth rib. Malignancies such as breast cancer, prostate cancer, plasma cell cytoma, and sarcoma can invade the ribs, thoracic spine, or chest wall and produce symptoms suggesting Tietze syndrome. Patients with osteomalacia may have significant rib pain, with or without documented micro fractures. These conditions should be distinguishable by radiographs, bone scanning, vitamin D measurement, or biopsy. Analgesics, anti-inflammatory drugs, and local glucocorticoid injections usually relieve symptoms of costochondritis/Tietze syndrome. Care should be taken to avoid overdiagnosing these syndromes in patients with acute chest pain syndromes; many patients will be tender to overly vigorous palpation of the costochondral joints.

MYOFASCIAL PAIN SYNDROME

Myofascial pain syndrome is characterized by multiple areas of localized musculoskeletal pain and tenderness in association with tender points. The pain is deep and aching and may be accompanied by a burning sensation. Myofascial pain may be regional and follow trauma, overuse, or prolonged static contraction of a muscle or muscle group, which may occur when reading or writing at a desk or working at a computer. In addition, this syndrome may be associated with underlying osteoarthritis of the neck or low back. Pain may be referred from tender points to defined areas distant from the area of original tenderness. Palpation of the tender point reproduces or accentuates the pain. The tender points are usually located in the center of a muscle belly, but they can occur at other sites such as costosternal junctions, the xiphoid process, ligamentous and tendinous insertions, fascia, and fatty areas. Tender point sites in muscle have been described as feeling indurated and taut, and palpation may cause the muscle to twitch. These findings, however, have been shown not to be unique for myofascial pain syndrome, because in a controlled study, they were also present in some "normal" subjects. Myofascial pain most often involves the posterior neck, low back, shoulders, and chest. Chronic pain in the muscles of the posterior neck may involve referral of pain from a tender point in the erector neck muscle or upper trapezius to the head, leading to persistent headaches, which may last for days. Tender points in the paraspinal muscles of the low back may refer pain to the buttock. Pain may be referred down the leg from a tender point in the gluteus medius and can mimic sciatica. A tender point in the infraspinatus muscle may produce local and referred pain over the lateral deltoid and down the outside of the arm into the hand. Injection of a local anesthetic such as 1% lidocaine into the tender point site often results in at least transient pain relief. Another useful technique is first to spray from the tender point toward the area of referred pain with an agent such as ethyl chloride and then to stretch the muscle. This maneuver may need to be repeated several times. Massage and application of ultrasound to the affected area also may be beneficial. Patients should be instructed in methods to prevent muscle stresses related to work and recreation. Posture and resting positions are important in preventing muscle tension. The prognosis in most patients is good. In some patients, regionally localized myofascial pain syndrome may seem to evolve into more generalized fibromyalgia (Chap. 335). Abnormal or nonrestorative sleep is a common accompaniment in these patients and may need to be specifically addressed.

Neoplasias and Arthritis

Primary tumors and tumor-like disorders of synovium are uncommon but should be considered in the differential diagnosis of monarticular joint disease. In addition, metastases to bone and primary bone tumors adjacent to a joint may produce joint symptoms. **For further discussion, see Chap. 98.**

Pigmented villonodular synovitis (PVNS) is characterized by the slowly progressive, exuberant, benign proliferation of synovial tissue, usually involving a single joint. The most common age of onset is in the third decade, and women are affected slightly more often than men. The cause of this disorder is unknown.

The synovium has a brownish color and numerous large, finger-like villi that fuse to form pedunculated nodules. There is marked hyperplasia of synovial cells in the stroma of the villi. Hemosiderin granules and lipids are found in the cytoplasm of macrophages and in the interstitial tissue. Multinucleated giant cells may be present. The proliferative synovium grows into the subsynovial tissue and invades adjacent cartilage and bone.

The clinical picture of pigmented villonodular synovitis is characterized by the insidious onset of persistent swelling and pain in affected joints, most commonly the knee. Other joints affected include the hips, ankles, calcaneocuboid joints, elbows, and small joints of the fingers or toes. The disease may also involve the common flexor sheath of the hands or fingers. Less commonly, tendon sheaths in the wrist, ankle, or foot may be involved. Symptoms of pain, a catching sensation, or stiffness may initially be mild and intermittent and may be present for years before the patient seeks medical attention. Radiographs may show joint space narrowing, erosions, and subchondral cysts. The joint fluid contains blood and is dark red or almost black in color. Lipid-containing macrophages may be present in the fluid. The joint fluid may be clear if hemorrhage has not occurred. Some patients have polyarticular involvement.

The treatment of pigmented villonodular synovitis is complete synovectomy. With incomplete synovectomy, the villonodular synovitis recurs, and the rate of tissue growth may be faster than it was originally. Irradiation of the involved joint has been successful in some patients.

Synovial chondromatosis is a disorder characterized by multiple focal metaplastic growths of normal-appearing cartilage in the synovium or tendon sheath. Segments of cartilage break loose and continue to grow as loose bodies. When calcification and ossification of loose bodies occur, the disorder is referred to as *synovial osteochondromatosis*. The disorder is usually monarticular and affects young to middle-aged individuals. The knee is most often involved, followed by hip, elbow, and shoulder. Symptoms are pain, swelling, and decreased motion of the joint. Radiographs may show several rounded calcifications within the joint cavity. Treatment is synovectomy; however, as in PVNS, the tumor may recur.

Synovial sarcoma is a malignantneoplasm often found near a large joint of both upper and lower extremities, being more common in the lower extremity. It seldom arises within the joint itself. Synovial sarcomas constitute 10% of soft tissue sarcomas. The tumor is believed to arise from primitive mesenchymal tissue that differentiates into epithelial cells and/or spindle cells. Small foci of calcification may be present in the tumor mass. It occurs most often in young adults and is more common in men. The tumor presents as a slowly growing deep-seated mass near a joint, without much pain. The area of the knee is the most common site, followed by the foot, ankle, elbow, and shoulder. Other primary sites include the buttocks, abdominal wall, retroperitoneum, and mediastinum. The tumor spreads along tissue planes. The most common site of visceral metastasis is lung. The diagnosis is made by biopsy. Treatment is wide resection of the tumor, including adjacent muscle and regional lymph nodes, followed by chemotherapy and radiation therapy. Amputation of the involved distal extremity may be required. Chemotherapy may be beneficial in some patients with metastatic disease. Isolated pulmonary metastasis can be surgically removed. The five-year survival rate with treatment is variable depending on the staging of the tumor, ranging from approximately 25% to 60% or higher. Synovial sarcomas tend to recur locally and metastasize to regional lymph nodes, lungs, and skeleton.

In addition to the rare direct metastases of solid cell tumors to the highly vascular synovium, neoplasia arising from nonarticular organ sites can affect joints in other ways. Acute leukemias in children can mimic juvenile inflammatory arthritis with severe joint pain and fever. In adults, chronic and acute myeloid leukemia can rarely infiltrate the synovium. The rarely occurring hairy cell leukemia has a peculiar tendency to cause episodic inflammatory oligoarthritis and tenosynovitis; these episodes are dramatic and mimic acute gout attacks. They respond to potent anti-inflammatory therapy with glucocorticoids; with remission of the leukemia, they may abate. Carcinomas can be associated with several paraneoplastic articular syndromes, including hypertrophic pulmonary osteoarthropathy (discussed above). Acute palmar fasciitis with polyarthritis is a well-described, but rare association with certain cancers, mainly adenocarcinomas. Clinically, this is fairly abrupt in onset with pain in the MCP and PIP joints of the hands with rapidly evolving contractures of the fingers due to thickening of the palmar (flexor) tendons. A similar syndrome can be seen in diabetics. Paraneoplastic arthritis has been described and may occur in several patterns: asymmetric predominantly lower extremity joints and symmetric, polyarthritis with hand joint involvement. Tumors were often found after the onset of the arthritis, although many patients had a preceding period of malaise or weight loss. The onset is often acute, and patients tend to be older males. These features should raise the specter of an underlying malignancy (or viral infection such as hepatitis C) as the cause of the arthritis. In one series, the symptoms resolved with successful therapy of the malignancy but did not recur with relapse of the malignancy. Dermatomyositis is also well described as a paraneoplastic syndrome and may have joint pain and arthritis as components of the syndrome. Malignancy associated arthritis may be responsive to NSAID therapy and to treatment of the primary neoplasm.

ACKNOWLEDGMENT

This chapter represents a revised version of the chapter authored by Dr. Bruce C. Gilliland that was in the previous editions of Harrison's*. Dr. Gilliland passed away on February 17, 2007, and had been a contributor to* Harrison's Principles of Internal Medicine *since the 11th edition.*

FURTHER READINGS

Altman RD, Tenenbaum J: Hypertrophic osteoarthropathy, in *Kelley's Textbook of Rheumatology* , 7th ed, ED Harris et al (eds). Philadelphia, Saunders, 2005, pp 1748–1753

Cronin ME: Rheumatic aspects of endocrinopathies, in *Arthritis and Allied Conditions,* 15th ed, WJ Koopman (ed). Philadelphia, Lippincott Williams & Wilkins, 2005, pp 2559–2576

Heck LW JR: Arthritis associated with hematologic disorders, storage diseases, disorders of lipid metabolism, and dysproteinemias, in *Arthritis and Allied Conditions* , 15th ed, WJ Koopman (ed). Philadelphia, Lippincott Williams & Wilkins, 2005, pp 1969–1990

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PERIARTICULAR DISORDERS OF THE EXTREMITIES: INTRODUCTION

A number of periarticular disorders have become increasingly common over the past two to three decades, due in part to greater participation in recreational sports by individuals of a wide range of ages. Periarticular disorders most commonly affect the knee or shoulder. With the exception of bursitis, hip pain is most often articular or is being referred from disease affecting another structure (Chap 331). This chapter discusses some of the more common periarticular disorders.

Bursitis

Bursitis is inflammation of a bursa, which is a thin-walled sac lined with synovial tissue. The function of the bursa is to facilitate movement of tendons and muscles over bony prominences. Excessive frictional forces from overuse, trauma, systemic disease (e.g., rheumatoid arthritis, gout), or infection may cause bursitis. *Subacromial bursitis* (subdeltoid bursitis) is the most common form of bursitis. The subacromial bursa, which is contiguous with the subdeltoid bursa, is located between the undersurface of the acromion and the humeral head and is covered by the deltoid muscle. Bursitis is caused by repetitive overhead motion and often accompanies rotator cuff tendinitis. Another frequently encountered form is *trochanteric bursitis*, which involves the bursa around the insertion of the gluteus medius onto the greater trochanter of the femur. Patients experience pain over the lateral aspect of the hip and upper thigh and have tenderness over the posterior aspect of the greater trochanter. External rotation and resisted abduction of the hip elicit pain. *Olecranon bursitis* occurs over the posterior elbow, and when the area is acutely inflamed, infection or gout should be excluded by aspirating the bursa and performing a Gram stain and culture on the fluid as well as examining the fluid for urate crystals. *Achilles bursitis* involves the bursa located above the insertion of the tendon to the calcaneus and results from overuse and wearing tight shoes. *Retrocalcaneal bursitis* involves the bursa that is located between the calcaneus and posterior surface of the Achilles tendon. The pain is experienced at the back of the heel, and swelling appears on the medial and/or lateral side of the tendon. It occurs in association with spondyloarthropathies, rheumatoid arthritis, gout, or trauma. *Ischial bursitis* (weaver's bottom) affects the bursa separating the gluteus medius from the ischial tuberosity and develops from prolonged sitting and pivoting on hard surfaces. *Iliopsoas bursitis* affects the bursa that lies between the iliopsoas muscle and hip joint and is lateral to the femoral vessels. Pain is experienced over this area and is made worse by hip extension and flexion. *Anserine bursitis* is an inflammation of the sartorius bursa located over the medial side of the tibia just below the knee and under the conjoint tendon and is manifested by pain on climbing stairs. Tenderness is present over the insertion of the conjoint tendon of the sartorius, gracilis, and semitendinosus. *Prepatellar bursitis* (housemaid's knee) occurs in the bursa situated between the patella and overlying skin and is caused by kneeling on hard surfaces. Gout or infection may also occur at this site. Treatment of bursitis consists of prevention of the aggravating situation, rest of the involved part, administration of a nonsteroidal anti-inflammatory drug (NSAID) where appropriate for an individual patient, or local glucocorticoid injection.

Rotator Cuff Tendinitis and Impingement Syndrome

Tendinitis of the rotator cuff is the major cause of a painful shoulder and is currently thought to be caused by inflammation of the tendon(s). The rotator cuff consists of the tendons of the supraspinatus, infraspinatus, subscapularis, and teres minor muscles, and inserts on the humeral tuberosities. Of the tendons forming the rotator cuff, the supraspinatus tendon is the most often affected, probably because of its repeated impingement (*impingement syndrome*) between the humeral head and the undersurface of the anterior third of the acromion and coracoacromial ligament above as well as the reduction in its blood supply that occurs with abduction of the arm **(Fig. 337-1)**. The tendon of the infraspinatus and that of the long head of the biceps are less commonly involved. The process begins with edema and hemorrhage of the rotator cuff, which evolves to fibrotic thickening and eventually to rotator cuff degeneration with tendon tears and bone spurs. Subacromial bursitis also accompanies this syndrome. Symptoms usually appear after injury or overuse, especially with activities involving elevation of the arm with some degree of forward flexion. Impingement syndrome occurs in persons participating in baseball, tennis, swimming, or occupations that require repeated elevation of the arm. Those over age 40 are particularly susceptible. Patients complain of a dull aching in the shoulder, which may interfere with sleep. Severe pain is experienced when the arm is actively abducted into an overhead position. The arc between 60° and 120° is especially painful. Tenderness is present over the lateral aspect of the humeral head just below the acromion. NSAIDs, local glucocorticoid injection, and

physical therapy may relieve symptoms. Surgical decompression of the subacromial space may be necessary in patients refractory to conservative treatment.

Figure 337-1

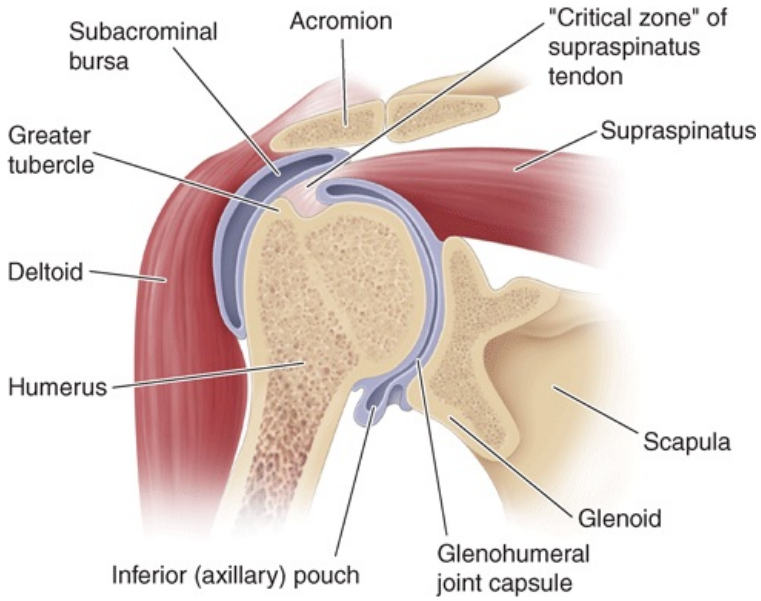

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronal section of the shoulder illustrating the relationships of the glenohumeral joint, the joint capsule, the subacromial bursa, and the rotator cuff (supraspinatus tendon). *[From F Kozin, in Arthritis and Allied Conditions, 13th ed, WJ Koopman (ed). Baltimore, Williams & Wilkins, 1997, with permission.]*

Patients may tear the supraspinatus tendon acutely by falling on an outstretched arm or lifting a heavy object. Symptoms are pain along with weakness of abduction and external rotation of the shoulder. Atrophy of the supraspinatus muscles develops. The diagnosis is established by arthrogram, ultrasound, or MRI. Surgical repair may be necessary in patients who fail to respond to conservative measures. In patients with moderate-to-severe tears and functional loss, surgery is indicated.

Calcific Tendinitis

This condition is characterized by deposition of calcium salts, primarily hydroxyapatite, within a tendon. The exact mechanism of calcification is not known but may be initiated by ischemia or degeneration of the tendon. The supraspinatus tendon is most often affected because it is frequently impinged on and has a reduced blood supply when the arm is abducted. The condition usually develops after age 40. Calcification within the tendon may evoke acute inflammation, producing sudden and severe pain in the shoulder. However, it may be asymptomatic or not related to the patient's symptoms.

Bicipital Tendinitis and Rupture

Bicipital tendinitis, or tenosynovitis, is produced by friction on the tendon of the long head of the biceps as it passes through the bicipital groove. When the inflammation is acute, patients experience anterior shoulder pain that radiates down the biceps into the forearm. Abduction and external rotation of the arm are painful and limited. The bicipital groove is very tender to palpation. Pain may be elicited along the course of the tendon by resisting supination of the forearm with the elbow at 90° (Yergason's supination sign). Acute rupture of the tendon may occur with vigorous exercise of the arm and is often painful. In a young patient, it should be repaired surgically. Rupture of the tendon in an older person may be associated with little or no pain and is recognized by the presence of persistent swelling of the biceps ("Popeye" muscle) produced by the retraction of the long head of the biceps. Surgery is usually not necessary in this setting.

De Quervain's Tenosynovitis

In this condition, inflammation involves the abductor pollicis longus and the extensor pollicis brevis as these tendons pass through a fibrous sheath at the radial styloid process. The usual cause is repetitive twisting of the wrist. It may occur in pregnancy, and it also occurs in mothers who hold their babies with the thumb outstretched. Patients experience pain on grasping with their thumb, such as with pinching. Swelling and tenderness are often present over the radial styloid process. The Finkelstein sign is positive, which is elicited by having the patient place the thumb in the palm and close the fingers over it. The wrist is then ulnarly deviated, resulting in pain over the involved tendon

sheath in the area of the radial styloid. Treatment consists initially of splinting the wrist and an NSAID. When severe or refractory to conservative treatment, glucocorticoid injections can be very effective.

Patellar Tendinitis (Jumper's Knee)

Tendinitis involves the patellar tendon at its attachment to the lower pole of the patella. Patients may experience pain when jumping during basketball or volleyball, going up stairs, or doing deep knee squats. Tenderness is noted on examination over the lower pole of the patella. Treatment consists of rest, icing, and NSAIDs, followed by strengthening and increasing flexibility.

Iliotibial Band Syndrome

The iliotibial band is a thick connective tissue that runs from the ilium to the fibula. Patients with iliotibial band syndrome most commonly present with aching or burning pain at the site where the band courses over the lateral femoral condyle of the knee; pain may also radiate up the thigh, toward the hip. Predisposing factors for iliotibial band syndrome include a varus alignment of the knee, excessive running distance, poorly fitted shoes, or continuous running on uneven terrain. Treatment consists of rest, NSAIDs, physical therapy, and addressing risk factors such as shoes and running surface. Glucocorticoid injection into the area of tenderness can provide relief, but running must be avoided for at least two weeks after the injection. Surgical release of the iliotibial band has been helpful in rare patients for whom conservative treatment has failed.

Adhesive Capsulitis

Often referred to as "frozen shoulder," adhesive capsulitis is characterized by pain and restricted movement of the shoulder, usually in the absence of intrinsic shoulder disease. Adhesive capsulitis may follow bursitis or tendinitis of the shoulder or be associated with systemic disorders such as chronic pulmonary disease, myocardial infarction, and diabetes mellitus. Prolonged immobility of the arm contributes to the development of adhesive capsulitis. Pathologically, the capsule of the shoulder is thickened, and a mild chronic inflammatory infiltrate and fibrosis may be present.

Adhesive capsulitis occurs more commonly in women after age 50. Pain and stiffness usually develop gradually but progress rapidly in some patients. Night pain is often present in the affected shoulder and pain may interfere with sleep. The shoulder is tender to palpation, and both active and passive movement are restricted. Radiographs of the shoulder show osteopenia. The diagnosis is typically made by physical examination but can be confirmed if necessary by arthrography, in that only a limited amount of contrast material, usually <15 mL, can be injected under pressure into the shoulder joint.

In most patients, the condition improves spontaneously 1–3 years after onset. While pain usually improves, many patients are left with some limitation of shoulder motion. Early mobilization of the arm following an injury to the shoulder may prevent the development of this disease. Physical therapy provides the foundation of treatment for adhesive capsulitis. Local injections of glucocorticoids and NSAIDs may also provide relief of symptoms. Slow but forceful injection of contrast material into the joint may lyse adhesions and stretch the capsule, resulting in improvement of shoulder motion. Manipulation under anesthesia may be helpful in some patients.

Lateral Epicondylitis (Tennis Elbow)

Lateral epicondylitis, or tennis elbow, is a painful condition involving the soft tissue over the lateral aspect of the elbow. The pain originates at or near the site of attachment of the common extensors to the lateral epicondyle and may radiate into the forearm and dorsum of the wrist. The pain usually appears after work or recreational activities involving repeated motions of wrist extension and supination against resistance. Most patients with this disorder injure themselves in activities other than tennis, such as pulling weeds, carrying suitcases or briefcases, or using a screwdriver. The injury in tennis usually occurs when hitting a backhand with the elbow flexed. Shaking hands and opening doors can reproduce the pain. Striking the lateral elbow against a solid object may also induce pain.

The treatment is usually rest along with administration of an NSAID. Ultrasound, icing, and friction massage may also help relieve pain. When pain is severe, the elbow is placed in a sling or splinted at 90° of flexion. When the pain is acute and well localized, injection of a glucocorticoid using a small-gauge needle may be effective. Following injection, the patient should be advised to rest the arm for at least one month and avoid activities that would aggravate the elbow. Once symptoms have subsided, the patient should begin rehabilitation to strengthen and increase flexibility of the extensor muscles before resuming physical activity involving the arm. A forearm band placed 2.5–5.0 cm (1–2 in.) below the elbow may help to reduce tension on the extensor muscles at their attachment to the lateral epicondyle. The patient should be advised to restrict activities requiring forcible extension and supination of the wrist. Improvement may take several months. The patient may continue to experience mild pain but, with care, can usually avoid the return of debilitating pain. Occasionally, surgical release of the extensor aponeurosis may be necessary.

Medial Epicondylitis

Medial epicondylitis is an overuse syndrome resulting in pain over the medial side of the elbow with radiation into the forearm. The cause of this syndrome is considered to be repetitive resisted motions

of wrist flexion and pronation, which lead to microtears and granulation tissue at the origin of the pronator teres and forearm flexors, particularly the flexor carpi radialis. This overuse syndrome is usually seen in patients >35 years and is much less common than lateral epicondylitis. It occurs most often in work-related repetitive activities but also occurs with recreational activities such as swinging a golf club (golfer's elbow) or throwing a baseball. On physical examination, there is tenderness just distal to the medial epicondyle over the origin of the forearm flexors. Pain can be reproduced by resisting wrist flexion and pronation with the elbow extended. Radiographs are usually normal. The differential diagnosis of patients with medial elbow symptoms include tears of the pronator teres, acute medial collateral ligament tear, and medial collateral ligament instability. Ulnar neuritis has been found in 25–50% of patients with medial epicondylitis and is associated with tenderness over the ulnar nerve at the elbow as well as hypesthesia and paresthesia on the ulnar side of the hand.

The initial treatment of medial epicondylitis is conservative, involving rest, NSAIDs, friction massage, ultrasound, and icing. Some patients may require splinting. Injections of glucocorticoids at the painful site may also be effective. Patients should be instructed to rest for at least one month. Also, patients should start physical therapy once the pain has subsided. In patients with chronic debilitating medial epicondylitis that remains unresponsive after at least a year of treatment, surgical release of the flexor muscle at its origin may be necessary and is often successful.

Plantar Fasciitis

Plantar fasciitis is a common cause of foot pain in adults, with the peak incidence occurring in people between the ages of 40 and 60 years. It is also seen more frequently in a younger population consisting of runners, aerobic exercise dancers, and ballet dancers. The pain originates at or near the site of the plantar fascia attachment to the medial tuberosity of the calcaneus. Several factors that increase the risk of developing plantar fasciitis include obesity, pes planus (flat foot or absence of the foot arch when standing), pes cavus (high-arched foot), limited dorsiflexion of the ankle, prolonged standing, walking on hard surfaces, and faulty shoes. In runners, excessive running and a change to a harder running surface may precipitate plantar fasciitis.

The diagnosis of plantar fasciitis can usually be made on the basis of history and physical examination alone. Patients experience severe pain with the first steps on arising in the morning or following inactivity during the day. The pain usually lessens with weight-bearing activity during the day, only to worsen with continued activity. Pain is made worse on walking barefoot or up stairs. On examination, maximal tenderness is elicited on palpation over the inferior heel corresponding to the site of attachment of the plantar fascia.

Imaging studies may be indicated when the diagnosis is not clear. Plain radiographs may show heel spurs, which are of little diagnostic significance. Ultrasonography in plantar fasciitis can demonstrate thickening of the fascia and diffuse hypoechogenicity, indicating edema at the attachment of the plantar fascia to the calcaneus. MRI is a sensitive method for detecting plantar fasciitis, but it is usually not required for establishing the diagnosis.

The differential diagnosis of inferior heel pain includes calcaneal stress fractures, the spondyloarthritides, rheumatoid arthritis, gout, neoplastic or infiltrative bone processes, and nerve compression/entrapment syndromes.

Resolution of symptoms occurs within 12 months in more than 80% of patients with plantar fasciitis. The patient is advised to reduce or discontinue activities that can exacerbate plantar fasciitis. Initial treatment consists of ice, heat, massage, and stretching. Stretching of the plantar fascia and calf muscles are commonly employed and can be beneficial. Orthotics provide medial arch support and can be effective. Foot strapping or taping are commonly performed, and some patients may benefit by wearing a night splint designed to keep the ankle in a neutral position. A short course of NSAIDs can be given to patients when the benefits outweigh the risks. Local glucocorticoid injections have also been shown to be efficacious but may carry an increased risk for plantar fascia rupture. Plantar fasciotomy is reserved for those patients who have failed to improve after at least 6–12 months of conservative treatment.

FURTHER READINGS

Buchbinder R: Plantar fasciitis. N Engl J Med 350:2159, 2004[PMID: 15152061] [Full Text]

Husni EM, Donohue JP: Painful shoulder and reflex sympathetic dystrophy syndrome, in *Arthritis and Allied Conditions*, 15th ed, WJ Koopman (ed). Philadelphia, Lippincott Williams & Wilkins, 2005, pp 2133–2152

Martin SD, Thornhill TS: Shoulder pain, in *Kelley's Textbook of Rheumatology* , 7th ed, ED Harris et al (eds). Philadelphia, Saunders, 2005, pp 557–587

Matsen FA 3rd. Clinical practice. Rotator-cuff failure. N Engl J Med 358:2138, 2008[PMID: 18480206] [Full Text]

Neer CS II: Impingement lesions. Clin Orthop 173:70, 1983[PMID: 6825348] [Full Text]

Sheridan MA, Hannafin JA. Upper extremity: emphasis on frozen shoulder. Orthop Clin North Am 37:531, 2006[PMID: 17141009] [Full Text]

† Deceased. A contributor to HPIM since the 11th edition, Dr. Gilliland passed away on February 17, 2007.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRINCIPLES OF ENDOCRINOLOGY: INTRODUCTION

The management of endocrine disorders requires a broad understanding of intermediary metabolism, reproductive physiology, bone metabolism, and growth. Accordingly, the practice of endocrinology is intimately linked to a conceptual framework for understanding hormone secretion, hormone action, and principles of feedback control. The endocrine system is evaluated primarily by measuring hormone concentrations, arming the clinician with valuable diagnostic information. Most disorders of the endocrine system are amenable to effective treatment once the correct diagnosis is determined. Endocrine deficiency disorders are treated with physiologic hormone replacement; hormone excess conditions, which usually are due to benign glandular adenomas, are managed by removing tumors surgically or reducing hormone levels medically.

SCOPE OF ENDOCRINOLOGY

The specialty of endocrinology encompasses the study of glands and the hormones they produce. The term *endocrine* was coined by Starling to contrast the actions of hormones secreted internally (*endocrine*) with those secreted externally (*exocrine*) or into a lumen, such as the gastrointestinal tract. The term *hormone*, derived from a Greek phrase meaning "to set in motion," aptly describes the dynamic actions of hormones as they elicit cellular responses and regulate physiologic processes through feedback mechanisms.

Unlike many other specialties in medicine, it is not possible to define endocrinology strictly along anatomic lines. The classic endocrine glands—pituitary, thyroid, parathyroid, pancreatic islets, adrenals, and gonads—communicate broadly with other organs through the nervous system, hormones, cytokines, and growth factors. In addition to its traditional synaptic functions, the brain produces a vast array of peptide hormones, and this has led to the discipline of neuroendocrinology. Through the production of hypothalamic releasing factors, the central nervous system (CNS) exerts a major regulatory influence over pituitary hormone secretion (Chap. 339). The peripheral nervous system stimulates the adrenal medulla. The immune and endocrine systems are also intimately intertwined. The adrenal hormone cortisol is a powerful immunosuppressant. Cytokines and interleukins (ILs) have profound effects on the functions of the pituitary, adrenal, thyroid, and gonads. Common endocrine diseases such as autoimmune thyroid disease and Type 1 diabetes mellitus are caused by dysregulation of immune surveillance and tolerance. Less common diseases such as polyglandular failure, Addison's disease, and lymphocytic hypophysitis also have an immunologic basis.

The interdigitation of endocrinology with physiologic processes in other specialties sometimes blurs the role of hormones. For example, hormones play an important role in maintenance of blood pressure, intravascular volume, and peripheral resistance in the cardiovascular system. Vasoactive substances such as catecholamines, angiotensin II, endothelin, and nitric oxide are involved in dynamic changes of vascular tone in addition to their multiple roles in other tissues. The heart is the principal source of atrial natriuretic peptide, which acts in classic endocrine fashion to induce natriuresis at a distant target organ (the kidney). Erythropoietin, a traditional circulating hormone, is made in the kidney and stimulates erythropoiesis in bone marrow (Chap. 57). The kidney is also integrally involved in the renin-angiotensin axis (Chap. 342) and is a primary target of several hormones, including parathyroid hormone (PTH), mineralocorticoids, and vasopressin. The gastrointestinal tract produces a surprising number of peptide hormones, such as cholecystokinin, ghrelin, gastrin, secretin, and vasoactive intestinal peptide, among many others. Adipose tissue produces leptin, which acts centrally to control appetite. Carcinoid and islet tumors can secrete excessive amounts of these hormones, leading to specific clinical syndromes (Chap. 350). Many of these gastrointestinal hormones are also produced in the CNS, where their functions are poorly understood. As hormones such as inhibin, ghrelin, and leptin are discovered, they become integrated into the science and practice of medicine on the basis of their functional roles rather than their tissues of origin.

Characterization of hormone receptors frequently reveals unexpected relationships to factors in nonendocrine disciplines. The growth hormone (GH) and leptin receptors, for example, are members of the cytokine receptor family. The G protein–coupled receptors (GPCRs), which mediate the actions of many peptide hormones, are used in numerous physiologic processes, including vision, smell, and neurotransmission.

NATURE OF HORMONES

Hormones can be divided into five major classes: (1) *amino acid derivatives* such as dopamine, catecholamine, and thyroid hormone; (2) *small neuropeptides* such as gonadotropin-releasing hormone (GnRH), thyrotropin-releasing hormone (TRH), somatostatin, and vasopressin; (3) *large proteins* such as insulin, luteinizing hormone (LH), and PTH produced by classic endocrine glands; (4) *steroid hormones* such as cortisol and estrogen that are synthesized from cholesterol-based precursors; and (5) *vitamin derivatives* such as retinoids (vitamin A) and vitamin D. A variety of *peptide growth factors*, most of which act locally, share actions with hormones. As a rule, amino acid derivatives and peptide hormones interact with cell-surface membrane receptors. Steroids, thyroid hormones, vitamin D, and retinoids are lipid-soluble and interact with intracellular nuclear receptors.

Hormone and Receptor Families

Many hormones and receptors can be grouped into families, reflecting their structural similarities (Table 338-1). The evolution of these families generates diverse but highly selective pathways of hormone action. Recognition of these relationships allows extrapolation of information gleaned from one hormone or receptor to other family members.

Table 338-1 Membrane Receptor Families and Signaling Pathways

Receptors	Effectors	Signaling Pathways
G Protein–Coupled Seven-Transmembrane (GPCR)		
β-Adrenergic LH, FSH, TSH	$G_{S}\alpha$, adenylate cyclase	Stimulation of cyclic AMP production, protein kinase A
Glucagon PTH, PTHrP ACTH, MSH GHRH, CRH	Ca^{2+} channels	Calmodulin, Ca^{2+}-dependent kinases
α-Adrenergic Somatostatin	$G_{i}\alpha$	Inhibition of cyclic AMP production Activation of K^{+}, Ca^{2+} channels
TRH, GnRH	G_q, G_{11}	Phospholipase C, diacylglycerol, IP_3, protein kinase C, voltage-dependent Ca^{2+} channels
Receptor Tyrosine Kinase		
Insulin, IGF-I	Tyrosine kinases, IRS	MAP kinases, PI 3-kinase; AKT, also known as protein kinase B, PKB
EGF, NGF	Tyrosine kinases, ras	Raf, MAP kinases, RSK
Cytokine Receptor–Linked Kinase		
GH, PRL	JAK, tyrosine kinases	STAT, MAP kinase, PI 3-kinase, IRS-1
Serine Kinase		
Activin, TGF-β, MIS	Serine kinase	Smads

Abbreviations: IP_3, inositol triphosphate; IRS, insulin receptor substrates; MAP, mitogen-activated protein; MSH, melanocyte-stimulating hormone; NGF, nerve growth factor; PI, phosphatidylinositol; RSK, ribosomal S6 kinase; TGF-β, transforming growth factor β. For all other abbreviations, see text.

The glycoprotein hormone family, consisting of thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), LH, and human chorionic gonadotropin (hCG), illustrates many features of related hormones. The glycoprotein hormones are heterodimers that have the α subunit in common; the β subunits are distinct and confer specific biologic actions. The overall three-dimensional architecture of the β subunits is similar, reflecting the locations of conserved disulfide bonds that restrain protein conformation. The cloning of the β-subunit genes from multiple species suggests that this family arose from a common ancestral gene, probably by gene duplication and subsequent divergence to evolve new biologic functions.

As the hormone families enlarge and diverge, their receptors must co-evolve if new biologic functions are to be derived. Related GPCRs, for example, have evolved for each of the glycoprotein hormones. These receptors are structurally similar, and each is coupled to the $G_{S}\alpha$ signaling pathway. However, there is minimal overlap of hormone binding. For example, TSH binds with high specificity to the TSH receptor but interacts minimally with the LH or the FSH receptor. Nonetheless, there can be subtle physiologic consequences of hormone cross-reactivity with other receptors. Very high levels of hCG

during pregnancy stimulate the TSH receptor and increase thyroid hormone levels, resulting in a compensatory decrease in TSH.

Insulin and insulin-like growth factor I (IGF-I) and IGF-II have structural similarities that are most apparent when precursor forms of the proteins are compared. In contrast to the high degree of specificity seen with the glycoprotein hormones, there is moderate cross-talk among the members of the insulin/IGF family. High concentrations of an IGF-II precursor produced by certain tumors (e.g., sarcomas) can cause hypoglycemia, partly because of binding to insulin and IGF-I receptors (Chap. 100). High concentrations of insulin also bind to the IGF-I receptor, perhaps accounting for some of the clinical manifestations seen in severe insulin resistance.

Another important example of receptor cross-talk is seen with PTH and parathyroid hormone–related peptide (PTHrP) (Chap. 353). PTH is produced by the parathyroid glands, whereas PTHrP is expressed at high levels during development and by a variety of tumors (Chap. 100). These hormones have amino acid sequence similarity, particularly in their amino-terminal regions. Both hormones bind to a single PTH receptor that is expressed in bone and kidney. Hypercalcemia and hypophosphatemia therefore may result from excessive production of either hormone, making it difficult to distinguish hyperparathyroidism from hypercalcemia of malignancy solely on the basis of serum chemistries. However, sensitive and specific assays for PTH and PTHrP now allow these disorders to be distinguished more readily.

Based on their specificities for DNA binding sites, the nuclear receptor family can be subdivided into type 1 receptors (GR, MR, AR, ER, PR) that bind steroids and type 2 receptors (TR, VDR, RAR, PPAR) that bind thyroid hormone, vitamin D, retinoic acid, or lipid derivatives. Certain functional domains in nuclear receptors, such as the zinc finger DNA-binding domains, are highly conserved. However, selective amino acid differences within this domain confer DNA sequence specificity. The hormone-binding domains are more variable, providing great diversity in the array of small molecules that bind to different nuclear receptors. With few exceptions, hormone binding is highly specific for a single type of nuclear receptor. One exception involves the glucocorticoid and mineralocorticoid receptors. Because the mineralocorticoid receptor also binds glucocorticoids with high affinity, an enzyme (11ß-hydroxysteroid dehydrogenase) in renal tubular cells inactivates glucocorticoids, allowing selective responses to mineralocorticoids such as aldosterone. However, when very high glucocorticoid concentrations occur, as in Cushing's syndrome, the glucocorticoid degradation pathway becomes saturated, allowing excessive cortisol levels to exert mineralocorticoid effects (sodium retention, potassium wasting). This phenomenon is particularly pronounced in ectopic adrenocorticotropic hormone (ACTH) syndromes (Chap. 342). Another example of relaxed nuclear receptor specificity involves the estrogen receptor, which can bind an array of compounds, some of which have little apparent structural similarity to the high-affinity ligand estradiol. This feature of the estrogen receptor makes it susceptible to activation by "environmental estrogens" such as resveratrol, octylphenol, and many other aromatic hydrocarbons. However, this lack of specificity provides an opportunity to synthesize a remarkable series of clinically useful antagonists (e.g., tamoxifen) and selective estrogen response modulators (SERMs) such as raloxifene. These compounds generate distinct conformations that alter receptor interactions with components of the transcription machinery (see below), thereby conferring their unique actions.

Hormone Synthesis and Processing

The synthesis of peptide hormones and their receptors occurs through a classic pathway of gene expression: transcription → mRNA → protein → posttranslational protein processing → intracellular sorting, followed by membrane integration or secretion (Chap. 61).

Many hormones are embedded within larger precursor polypeptides that are proteolytically processed to yield the biologically active hormone. Examples include proopiomelanocortin (POMC) → ACTH; proglucagon → glucagon; proinsulin → insulin; and pro-PTH → PTH, among others. In many cases, such as POMC and proglucagon, these precursors generate multiple biologically active peptides. It is provocative that hormone precursors are typically inactive, presumably adding an additional level of regulatory control. Prohormone conversion occurs not only for peptide hormones but also for certain steroids (testosterone → dihydrotestosterone) and thyroid hormone ($T_4 \rightarrow T_3$).

Hormone precursor processing is intimately linked to intracellular sorting pathways that transport proteins to appropriate vesicles and enzymes, resulting in specific cleavage steps, followed by protein folding and translocation to secretory vesicles. Hormones destined for secretion are translocated across the endoplasmic reticulum under the guidance of an amino-terminal signal sequence that subsequently is cleaved. Cell-surface receptors are inserted into the membrane via short segments of hydrophobic amino acids that remain embedded within the lipid bilayer. During translocation through the Golgi and endoplasmic reticulum, hormones and receptors are also subject to a variety of posttranslational modifications, such as glycosylation and phosphorylation, which can alter protein conformation, modify circulating half-life, and alter biologic activity.

Synthesis of most steroid hormones is based on modifications of the precursor, cholesterol. Multiple regulated enzymatic steps are required for the synthesis of testosterone (Chap. 346), estradiol (Chap.

347), cortisol (Chap. 342), and vitamin D (Chap. 352). This large number of synthetic steps predisposes to multiple genetic and acquired disorders of steroidogenesis.

Although endocrine genes contain regulatory DNA elements similar to those found in many other genes, their exquisite control by other hormones also necessitates the presence of specific hormone response elements. For example, the TSH genes are repressed directly by thyroid hormones acting through the thyroid hormone receptor (TR), a member of the nuclear receptor family. Steroidogenic enzymegene expression requires specific transcription factors, such as steroidogenic factor-1 (SF-1), acting in conjunction with signals transmitted by trophic hormones (e.g., ACTH or LH). For some hormones, substantial regulation occurs at the level of translational efficiency. Insulin biosynthesis, although it requires ongoing gene transcription, is regulated primarily at the translational level in response to elevated levels of glucose or amino acids.

Hormone Secretion, Transport, and Degradation

The circulating level of a hormone is determined by its rate of secretion and its circulating half-life. After protein processing, peptide hormones (GnRH, insulin, GH) are stored in secretory granules. As these granules mature, they are poised beneath the plasma membrane for imminent release into the circulation. In most instances, the stimulus for hormone secretion is a releasing factor or neural signal that induces rapid changes in intracellular calcium concentrations, leading to secretory granule fusion with the plasma membrane and release of its contents into the extracellular environment and bloodstream. Steroid hormones, in contrast, diffuse into the circulation as they are synthesized. Thus, their secretory rates are closely aligned with rates of synthesis. For example, ACTH and LH induce steroidogenesis by stimulating the activity of *st*eroidogenic *a*cute *r*egulatory (StAR) protein (transports cholesterol into the mitochondrion) along with other rate-limiting steps (e.g., cholesterol side-chain cleavage enzyme, CYP11A1) in the steroidogenic pathway.

Hormone transport and degradation dictate the rapidity with which a hormonal signal decays. Some hormonal signals are evanescent (e.g., somatostatin), whereas others are longer-lived (e.g., TSH). Because somatostatin exerts effects in virtually every tissue, a short half-life allows its concentrations and actions to be controlled locally. Structural modifications that impair somatostatin degradation have been useful for generating long-acting therapeutic analogues such as octreotide (Chap. 339). In contrast, the actions of TSH are highly specific for the thyroid gland. Its prolonged half-life accounts for relatively constant serum levels even though TSH is secreted in discrete pulses.

An understanding of circulating hormonehalf-life is important for achieving physiologic hormone replacement, as the frequency of dosing and the time required to reach steady state are intimately linked to rates of hormone decay. T4, for example, has a circulating half-life of 7 days. Consequently, >1 month is required to reach a new steady state, and single daily doses are sufficient to achieve constant hormone levels. T_3, in contrast, has a half-life of 1 day. Its administration is associated with more dynamic serum levels, and it must be administered two to three times per day. Similarly, synthetic glucocorticoids vary widely in their half-lives; those with longer half-lives (e.g., dexamethasone) are associated with greater suppression of the hypothalamic-pituitary-adrenal (HPA) axis. Most protein hormones [e.g., ACTH, GH, prolactin (PRL), PTH, LH] have relatively short half-lives (<20 min), leading to sharp peaks of secretion and decay. The only accurate way to profile the pulse frequency and amplitude of these hormones is to measure levels in frequently sampled blood (every 10 min or less) over long durations (8–24 h). Because this is not practical in a clinical setting, an alternative strategy is to pool three to four samples drawn at about 30-min intervals or interpret the results in the context of a relatively wide normal range. Rapid hormone decay is useful in certain clinical settings. For example, the short half-life of PTH allows the use of intraoperative PTH determinations to confirm successful removal of an adenoma. This is particularly valuable diagnostically when there is a possibility of multicentric disease or parathyroid hyperplasia, as occurs with multiple endocrine neoplasia (MEN) or renal insufficiency.

Many hormones circulate in association with serum-binding proteins. Examples include (1) T4 and T_3 binding to thyroxine- binding globulin (TBG), albumin, and thyroxine-binding prealbumin (TBPA); (2) cortisol binding to cortisol-binding globulin (CBG); (3) androgen and estrogen binding to sex hormone –binding globulin (SHBG) [also called testosterone-binding globulin (TeBG)]; (4) IGF-I and -II binding to multiple IGF-binding proteins (IGFBPs); (5) GH interactions with GH-binding protein (GHBP), a circulating fragment of the GH receptor extracellular domain; and (6) activin binding to follistatin. These interactions provide a hormonal reservoir, prevent otherwise rapid degradation of unbound hormones, restrict hormone access to certain sites (e.g., IGFBPs), and modulate the unbound, or "free," hormone concentrations. Although a variety of binding protein abnormalities have been identified, most have few clinical consequences aside from creating diagnostic problems. For example, TBG deficiency can reduce total thyroid hormone levels greatly, but the free concentrations of T_4 and T_3 remain normal. Liver disease and certain medications can also influence binding protein levels (e.g., estrogen increases TBG) or cause displacement of hormones from binding proteins (e.g., salsalate displaces T_4 from TBG). In general, only unbound hormone is available to interact with receptors and thus elicit a biologic response. Short-term perturbations in binding proteins change the free hormone concentration, which in turn induces compensatory adaptations through feedback loops. SHBG

changes in women are an exception to this self-correcting mechanism. When SHBG decreases because of insulin resistance or androgen excess, the unbound testosterone concentration is increased, potentially leading to hirsutism (Chap. 49). The increased unbound testosterone level does not result in an adequate compensatory feedback correction because estrogen, not testosterone, is the primary regulator of the reproductive axis.

An additional exception to the unbound hormone hypothesis involves megalin, a member of the low-density lipoprotein (LDL) receptor family that serves as an endocytotic receptor for carrier-bound vitamins A and D and SHBG-bound androgens and estrogens. After internalization, the carrier proteins are degraded in lysosomes and release their bound ligands within the cells. Membrane transporters have also been identified for thyroid hormones.

Hormone degradation can be an important mechanism for regulating concentrations locally. As noted above, 11β-hydroxysteroid dehydrogenase inactivates glucocorticoids in renal tubular cells, preventing actions through the mineralocorticoid receptor. Thyroid hormone deiodinases convert T4 to T_3 and can inactivate T_3. During development, degradation of retinoic acid by Cyp26b1 prevents primordial germ cells in the male from entering meiosis, as occurs in the female ovary.

HORMONE ACTION THROUGH RECEPTORS

Receptors for hormones are divided into two major classes: membrane and nuclear. *Membrane receptors* primarily bind peptide hormones and catecholamines. *Nuclear receptors* bind small molecules that can diffuse across the cell membrane, such as steroids and vitamin D. Certain general principles apply to hormone-receptor interactions regardless of the class of receptor. Hormones bind to receptors with specificity and an affinity that generally coincides with the dynamic range of circulating hormone concentrations. Low concentrations of free hormone (usually 10^{12} to 10^{9} *M*) rapidly associate and dissociate from receptors in a bimolecular reaction such that the occupancy of the receptor at any given moment is a function of hormone concentration and the receptor's affinity for the hormone. Receptor numbers vary greatly in different target tissues, providing one of the major determinants of specific cellular responses to circulating hormones. For example, ACTH receptors are located almost exclusively in the adrenal cortex, and FSH receptors are found predominantly in the gonads. In contrast, insulin and TRs are widely distributed, reflecting the need for metabolic responses in all tissues.

Membrane Receptors

Membrane receptors for hormones can be divided into several major groups: (1) seven transmembrane GPCRs, (2) tyrosine kinase receptors, (3) cytokine receptors, and (4) serine kinase receptors (Fig. 338-1). The *seven transmembrane GPCR family* binds a remarkable array of hormones, including large proteins (e.g., LH, PTH), small peptides (e.g., TRH, somatostatin), catecholamines (epinephrine, dopamine), and even minerals (e.g., calcium). The extracellular domains of GPCRs vary widely in size and are the major binding site for large hormones. The transmembrane- spanning regions are composed of hydrophobic α-helical domains that traverse the lipid bilayer. Like some channels, these domains are thought to circularize and form a hydrophobic pocket into which certain small ligands fit. Hormone binding induces conformational changes in these domains, transducing structural changes to the intracellular domain, which is a docking site for G proteins.

Figure 338-1

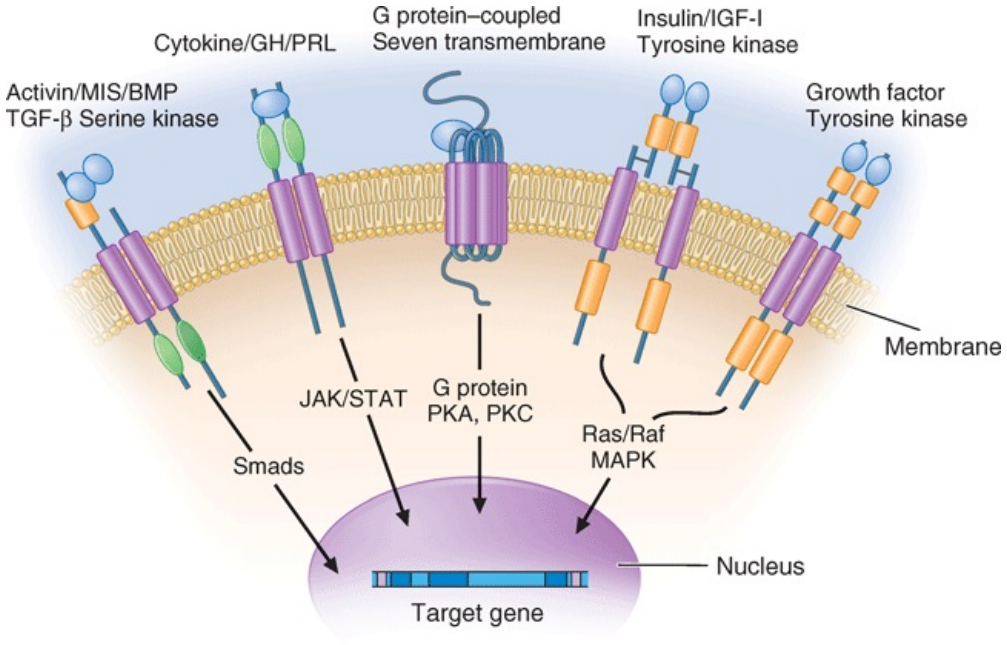

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Membrane receptor signaling. MAPK, mitogen-activated protein kinase; PKA, -C, protein kinase A, C; TGF, transforming growth factor. For other abbreviations, see text.

The large family of *G proteins*, so named because they bind guanine nucleotides [guanosine triphosphate (GTP), guanosine diphosphate (GDP)], provides great diversity for coupling receptors to different signaling pathways. G proteins form a heterotrimeric complex that is composed of various α and βγ subunits. The α subunit contains the guanine nucleotide–binding site and hydrolyzes GTP → GDP. The βγ subunits are tightly associated and modulate the activity of the α subunit as well as mediating their own effector signaling pathways. G protein activity is regulated by a cycle that involves GTP hydrolysis and dynamic interactions between the α and αβ subunits. Hormone binding to the receptor induces GDP dissociation, allowing Gα to bind GTP and dissociate from the αβ complex. Under these conditions, the Gα subunit is activated and mediates signal transduction through various enzymes, such as adenylate cyclase and phospholipase C. GTP hydrolysis to GDP allows reassociation with the αβ subunits and restores the inactive state. As described below, a variety of endocrinopathies result from G protein mutations or from mutations in receptors that modify their interactions with G proteins. G proteins interact with other cellular proteins, including kinases, channels, G protein–coupled receptor kinases (GRKs), and arrestins, that mediate signaling as well as receptor desensitization and recycling.

The *tyrosine kinase receptors* transduce signals for insulin and a variety of growth factors, such as IGF-I, epidermal growth factor (EGF), nerve growth factor, platelet-derived growth factor, and fibroblast growth factor. The cysteine-rich extracellular ligand-binding domains contain growth factor binding sites. After ligand binding, this class of receptors undergoes autophosphorylation, inducing interactions with intracellular adaptor proteins such as Shc and insulin receptor substrates. In the case of the insulin receptor, multiple kinases are activated, including the Raf-Ras-MAPK and the Akt/protein kinase B pathways. The tyrosine kinase receptors play a prominent role in cell growth and differentiation as well as in intermediary metabolism.

The GH and PRL receptors belong to the *cytokine receptor* family. Analogous to the tyrosine kinase receptors, ligand binding induces receptor interaction with intracellular kinases—the Janus kinases (JAKs), which phosphorylate members of the signal transduction and activators of transcription (STAT) family—as well as with other signaling pathways (Ras, PI3-K, MAPK). The activated STAT proteins translocate to the nucleus and stimulate expression of target genes.

The *serine kinase receptors* mediate the actions of activins, transforming growth factor β, müllerian-inhibiting substance (MIS, also known as anti-müllerian hormone, AMH), and bone morphogenic proteins (BMPs). This family of receptors (consisting of type I and II subunits) signals through proteins termed *smads* (fusion of terms for *Caenorhabditis elegans* sma + mammalian mad). Like the STAT proteins, the smads serve a dual role of transducing the receptor signal and acting as transcription factors. The pleomorphic actions of these growth factors dictate that they act primarily in a local (paracrine or autocrine) manner. Binding proteins such as follistatin (which binds activin and other members of this family) function to inactivate the growth factors and restrict their distribution.

Nuclear Receptors

The family of nuclear receptors has grown to nearly 100 members, many of which are still classified as orphan receptors because their ligands, if they exist, have not been identified (Fig. 338-2). Otherwise, most nuclear receptors are classified on the basis of the nature of their ligands. Though all nuclear receptors ultimately act to increase or decrease gene transcription, some (e.g., glucocorticoid receptor) reside primarily in the cytoplasm, whereas others (e.g., thyroid hormone receptor) are always located in the nucleus. After ligand binding, the cytoplasmically localized receptors translocate to the nucleus. There is growing evidence that certain nuclear receptors (e.g., glucocorticoid, estrogen) can also act at the membrane or in the cytoplasm to activate or repress signal transduction pathways, providing a mechanism for cross-talk between membrane and nuclear receptors.

Figure 338-2

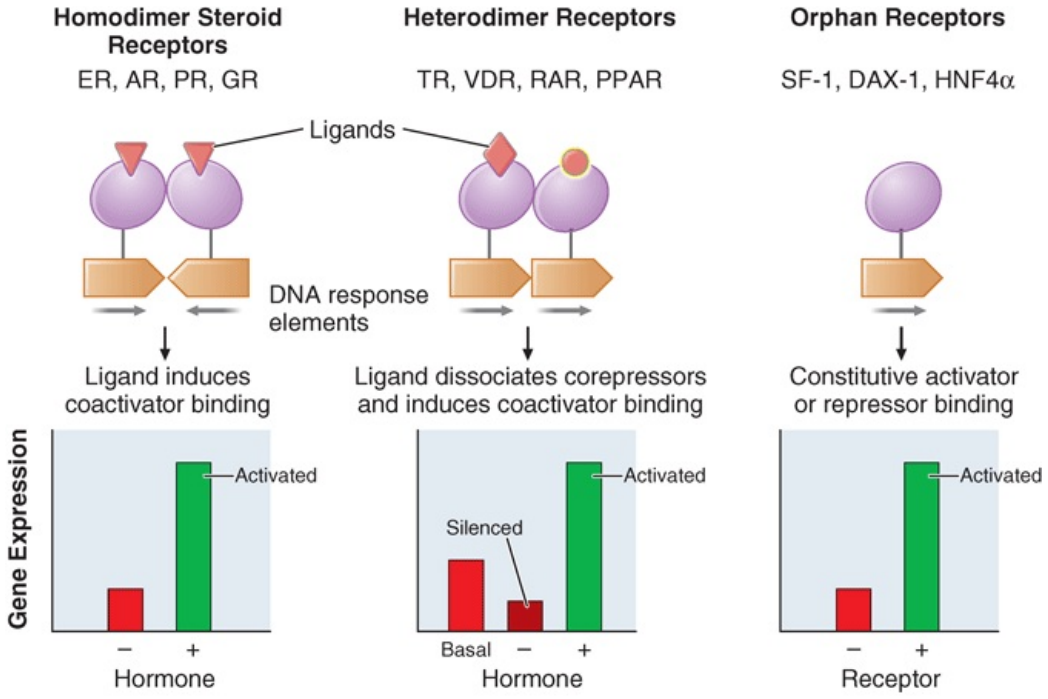

Nuclear receptor signaling. ER, estrogen receptor; AR, androgen receptor; PR, progesterone receptor; GR, glucocorticoid receptor; TR, thyroid hormone receptor; VDR, vitamin D receptor; RAR, retinoic acid receptor; PPAR, peroxisome proliferator activated receptor; SF-1, steroidogenic factor-1; DAX, *d*osage-sensitive sex-reversal, *a*drenal hypoplasia congenita, *X*-chromosome; HNF4α, hepatic nuclear factor 4α.

The structures of nuclear receptors have been studied extensively, including by x-ray crystallography. The DNA binding domain, consisting of two zinc fingers, contacts specific DNA recognition sequences in target genes. Most nuclear receptors bind to DNA as dimers. Consequently, each monomer recognizes an individual DNA motif, referred to as a "half-site." The steroid receptors, including the glucocorticoid, estrogen, progesterone, and androgen receptors, bind to DNA as homodimers. Consistent with this twofold symmetry, their DNA recognition half-sites are palindromic. The thyroid, retinoid, peroxisome proliferator activated, and vitamin D receptors bind to DNA preferentially as heterodimers in combination with retinoid X receptors (RXRs). Their DNA half-sites are arranged as direct repeats.

The carboxy-terminal hormone-binding domain mediates transcriptional control. For type II receptors such as thyroid hormone receptor (TR) and retinoic acid receptor (RAR), co-repressor proteins bind to the receptor in the absence of ligand and silence gene transcription. Hormone binding induces conformational changes, triggering the release of co-repressors and inducing the recruitment of coactivators that stimulate transcription. Thus, these receptors are capable of mediating dramatic changes in the level of gene activity. Certain disease states are associated with defective regulation of these events. For example, mutations in the TR prevent co-repressor dissociation, resulting in a dominant form of hormone resistance (Chap. 341). In promyelocytic leukemia, fusion of RARα to other nuclear proteins causes aberrant gene silencing and prevents normal cellular differentiation. Treatment with retinoic acid reverses this repression and allows cellular differentiation and apoptosis to occur. Most type 1 steroid receptors interact weakly with co-repressors, but ligand binding still induces interactions with an array of coactivators. X-ray crystallography shows that various SERMs induce distinct estrogen receptor conformations. The tissue-specific responses caused by these agents in breast, bone, and uterus appear to reflect distinct interactions with coactivators. The receptor-

coactivator complex stimulates gene transcription by several pathways, including (1) recruitment of enzymes (histone acetyl transferases) that modify chromatin structure, (2) interactions with additional transcription factors on the target gene, and (3) direct interactions with components of the general transcription apparatus to enhance the rate of RNA polymerase II–mediated transcription. Studies of nuclear receptor-mediated transcription show that these are dynamic events that involve relatively rapid (e.g., 30–60 min) cycling of transcription complexes on any specific target gene.

FUNCTIONS OF HORMONES

The functions of individual hormones are described in detail in subsequent chapters. Nevertheless, it is useful to illustrate how most biologic responses require integration of several different hormone pathways. The physiologic functions of hormones can be divided into three general areas: (1) growth and differentiation, (2) maintenance of homeostasis, and (3) reproduction.

Growth

Multiple hormones and nutritional factors mediate the complex phenomenon of growth (Chap. 339). Short stature may be caused by GH deficiency, hypothyroidism, Cushing's syndrome, precocious puberty, malnutrition, chronic illness, or genetic abnormalities that affect the epiphyseal growth plates (e.g., *FGFR3* and *SHOX* mutations). Many factors (GH, IGF-I, thyroid hormones) stimulate growth, whereas others (sex steroids) lead to epiphyseal closure. Understanding these hormonal interactions is important in the diagnosis and management of growth disorders. For example, delaying exposure to high levels of sex steroids may enhance the efficacy of GH treatment.

Maintenance of Homeostasis

Though virtually all hormones affect homeostasis, the most important among them are the following:

1. Thyroid hormone—controls about 25% of basal metabolism in most tissues
2. Cortisol—exerts a permissive action for many hormones in addition to its own direct effects
3. PTH—regulates calcium and phosphorus levels
4. Vasopressin—regulates serum osmolality by controlling renal free-water clearance
5. Mineralocorticoids—control vascular volume and serum electrolyte (Na^+, K^+) concentrations
6. Insulin—maintains euglycemia in the fed and fasted states

The defense against hypoglycemia is an impressive example of integrated hormone action (Chap. 345). In response to the fasted state and falling blood glucose, insulin secretion is suppressed, resulting in decreased glucose uptake and enhanced glycogenolysis, lipolysis, proteolysis, and gluconeogenesis to mobilize fuel sources. If hypoglycemia develops (usually from insulin administration or sulfonylureas), an orchestrated counterregulatory response occurs—glucagon and epinephrine rapidly stimulate glycogenolysis and gluconeogenesis, whereas GH and cortisol act over several hours to raise glucose levels and antagonize insulin action.

Although free-water clearance is controlled primarily by vasopressin, cortisol and thyroid hormone are also important for facilitating renal tubular responses to vasopressin (Chap. 340). PTH and vitamin D function in an interdependent manner to control calcium metabolism (Chap. 352). PTH stimulates renal synthesis of 1,25-dihydroxyvitamin D, which increases calcium absorption in the gastrointestinal tract and enhances PTH action in bone. Increased calcium, along with vitamin D, feeds back to suppress PTH, thus maintaining calcium balance.

Depending on the severity of a specific stress and whether it is acute or chronic, multiple endocrine and cytokine pathways are activated to mount an appropriate physiologic response. In severe acute stress such as trauma or shock, the sympathetic nervous system is activated and catecholamines are released, leading to increased cardiac output and a primed musculoskeletal system. Catecholamines also increase mean blood pressure and stimulate glucose production. Multiple stress-induced pathways converge on the hypothalamus, stimulating several hormones, including vasopressin and corticotropin-releasing hormone (CRH). These hormones, in addition to cytokines (tumor necrosis factor α, IL-2, IL-6), increase ACTH and GH production. ACTH stimulates the adrenal gland, increasing cortisol, which in turn helps sustain blood pressure and dampen the inflammatory response. Increased vasopressin acts to conserve free water.

Reproduction

The stages of reproduction include (1) sex determination during fetal development (Chap. 349); (2) sexual maturation during puberty (Chaps. 346 and 347); (3) conception, pregnancy, lactation, and child rearing (Chap. 347); and (4) cessation of reproductive capability at menopause (Chap. 348). Each of these stages involves an orchestrated interplay of multiple hormones, a phenomenon well illustrated by the dynamic hormonal changes that occur during each 28-day menstrual cycle. In the early follicular phase, pulsatile secretion of LH and FSH stimulates the progressive maturation of the ovarian follicle. This results in gradually increasing estrogen and progesterone levels, leading to enhanced pituitary

sensitivity to GnRH, which, when combined with accelerated GnRH secretion, triggers the LH surge and rupture of the mature follicle. Inhibin, a protein produced by the granulosa cells, enhances follicular growth and feeds back to the pituitary to selectively suppress FSH without affecting LH. Growth factors such as EGF and IGF-I modulate follicular responsiveness to gonadotropins. Vascular endothelial growth factor and prostaglandins play a role in follicle vascularization and rupture.

During pregnancy, the increased production of prolactin, in combination with placentally derived steroids (e.g., estrogen and progesterone), prepares the breast for lactation. Estrogens induce the production of progesterone receptors, allowing for increased responsiveness to progesterone. In addition to these and other hormones involved in lactation, the nervous system and oxytocin mediate the suckling response and milk release.

HORMONAL FEEDBACK REGULATORY SYSTEMS

Feedback control, both negative and positive, is a fundamental feature of endocrine systems. Each of the major hypothalamic-pituitary-hormone axes is governed by negative feedback, a process that maintains hormone levels within a relatively narrow range (Chap. 339). Examples of hypothalamic-pituitary negative feedback include (1) thyroid hormones on the TRH-TSH axis, (2) cortisol on the CRH-ACTH axis, (3) gonadal steroids on the GnRH-LH/FSH axis, and (4) IGF-I on the growth hormone–releasing hormone (GHRH)-GH axis (Fig. 338-3). These regulatory loops include both positive (e.g., TRH, TSH) and negative (e.g., T4, T_3) components, allowing for exquisite control of hormone levels. As an example, a small reduction of thyroid hormone triggers a rapid increase of TRH and TSH secretion, resulting in thyroid gland stimulation and increased thyroid hormone production. When thyroid hormone reaches a normal level, it feeds back to suppress TRH and TSH, and a new steady state is attained. Feedback regulation also occurs for endocrine systems that do not involve the pituitary gland, such as calcium feedback on PTH, glucose inhibition of insulin secretion, and leptin feedback on the hypothalamus. An understanding of feedback regulation provides important insights into endocrine testing paradigms (see below).

Figure 338-3

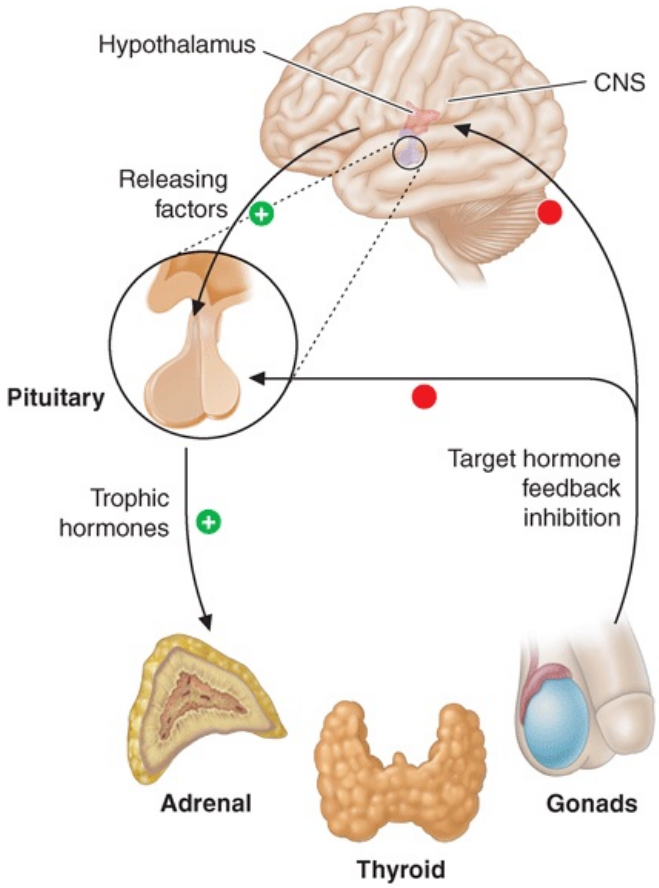

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Feedback regulation of endocrine axes. CNS, central nervous system.

Positive feedback control also occurs but is not well understood. The primary example is estrogen-mediated stimulation of the midcycle LH surge. Though chronic low levels of estrogen are inhibitory,

gradually rising estrogen levels stimulate LH secretion. This effect, which is illustrative of an endocrine rhythm (see below), involves activation of the hypothalamic GnRH pulse generator. In addition, estrogen-primed gonadotropes are extraordinarily sensitive to GnRH, leading to amplification of LH release.

Paracrine and Autocrine Control

The previously mentioned examples of feedback control involve classic endocrine pathways in which hormones are released by one gland and act on a distant target gland. However, local regulatory systems, often involving growth factors, are increasingly recognized. *Paracrine regulation* refers to factors released by one cell that act on an adjacent cell in the same tissue. For example, somatostatin secretion by pancreatic islet δ cells inhibits insulin secretion from nearby β cells. *Autocrine regulation* describes the action of a factor on the same cell from which it is produced. IGF-I acts on many cells that produce it, including chondrocytes, breast epithelium, and gonadal cells. Unlike endocrine actions, paracrine and autocrine control are difficult to document because local growth factor concentrations cannot be measured readily.

Anatomic relationships of glandular systems also greatly influence hormonal exposure: the physical organization of islet cells enhances their intercellular communication; the portal vasculature of the hypothalamic-pituitary system exposes the pituitary to high concentrations of hypothalamic releasing factors; testicular seminiferous tubules gain exposure to high testosterone levels produced by the interdigitated Leydig cells; the pancreas receives nutrient information and local exposure to peptide hormones (incretins) from the gastrointestinal tract; and the liver is the proximal target of insulin action because of portal drainage from the pancreas.

Hormonal Rhythms

The feedback regulatory systems described above are superimposed on hormonal rhythms that are used for adaptation to the environment. Seasonal changes, the daily occurrence of the light-dark cycle, sleep, meals, and stress are examples of the many environmental events that affect hormonal rhythms. The *menstrual cycle* is repeated on average every 28 days, reflecting the time required to follicular maturation and ovulation (Chap. 347). Essentially all pituitary hormone rhythms are entrained to sleep and to the *circadian cycle*, generating reproducible patterns that are repeated approximately every 24 h. The HPA axis, for example, exhibits characteristic peaks of ACTH and cortisol production in the early morning, with a nadir during the night. Recognition of these rhythms is important for endocrine testing and treatment. Patients with Cushing's syndrome characteristically exhibit increased midnight cortisol levels compared with normal individuals (Chap. 342). In contrast, morning cortisol levels are similar in these groups, as cortisol is normally high at this time of day in normal individuals. The HPA axis is more susceptible to suppression by glucocorticoids administered at night as they blunt the early-morning rise of ACTH. Understanding these rhythms allows glucocorticoid replacement that mimics diurnal production by administering larger doses in the morning than in the afternoon. Disrupted sleep rhythms can alter hormonal regulation. For example, sleep deprivation causes mild insulin resistance, food craving, and hypertension, which are reversible, at least in the short term.

Other endocrine rhythms occur on a more rapid time scale. Many peptide hormones are secreted in discrete bursts every few hours. LH and FSH secretion are exquisitely sensitive to GnRH pulse frequency. Intermittent pulses of GnRH are required to maintain pituitary sensitivity, whereas continuous exposure to GnRH causes pituitary gonadotrope desensitization. This feature of the hypothalamic-pituitary-gonadotrope axis forms the basis for using long-acting GnRH agonists to treat central precocious puberty or to decrease testosterone levels in the management of prostate cancer. It is important to be aware of the pulsatile nature of hormone secretion and the rhythmic patterns of hormone production in relating serum hormone measurements to normal values. For some hormones, integrated markers have been developed to circumvent hormonal fluctuations. Examples include 24-h urine collections for cortisol, IGF-I as a biologic marker of GH action, and HbA1c as an index of long-term (weeks to months) blood glucose control.

Often, one must interpret endocrine data only in the context of other hormones. For example, PTH levels typically are assessed in combination with serum calcium concentrations. A high serum calcium level in association with elevated PTH is suggestive of hyperparathyroidism, whereas a suppressed PTH in this situation is more likely to be caused by hypercalcemia of malignancy or other causes of hypercalcemia. Similarly, TSH should be elevated when T4 and T_3 concentrations are low, reflecting reduced feedback inhibition. When this is not the case, it is important to consider secondary hypothyroidism, which is caused by a defect at the level of the pituitary.

PATHOLOGIC MECHANISMS OF ENDOCRINE DISEASE

Endocrine diseases can be divided into three major types of conditions: (1) hormone excess, (2) hormone deficiency, and (3) hormone resistance (Table 338-2).

Table 338-2 Causes of Endocrine Dysfunction

Type of Endocrine Disorder	Examples
Hyperfunction	
Neoplastic	
Benign	Pituitary adenomas, hyperparathyroidism, autonomous thyroid or adrenal nodules, pheochromocytoma
Malignant	Adrenal cancer, medullary thyroid cancer, carcinoid
Ectopic	Ectopic ACTH, SIADH secretion
Multiple endocrine neoplasia	MEN 1, MEN 2
Autoimmune	Graves' disease
Iatrogenic	Cushing's syndrome, hypoglycemia
Infectious/inflammatory	Subacute thyroiditis
Activating receptor mutations	LH, TSH, Ca^{2+} and PTH receptors, $G_{S}\alpha$
Hypofunction	
Autoimmune	Hashimoto's thyroiditis, Type 1 diabetes mellitus, Addison's disease, polyglandular failure
Iatrogenic	Radiation-induced hypopituitarism, hypothyroidism, surgical
Infectious/inflammatory	Adrenal insufficiency, hypothalamic sarcoidosis
Hormone mutations	GH, LHβ, FSHβ, vasopressin
Enzyme defects	21-Hydroxylase deficiency
Developmental defects	Kallmann syndrome, Turner syndrome, transcription factors
Nutritional/vitamin deficiency	Vitamin D deficiency, iodine deficiency
Hemorrhage/infarction	Sheehan's syndrome, adrenal insufficiency
Hormone resistance	
Receptor mutations	
Membrane	GH, vasopressin, LH, FSH, ACTH, GnRH, GHRH, PTH, leptin, Ca^{2+}
Nuclear	AR, TR, VDR, ER, GR, PPARγ
Signaling pathway mutations	Albright's hereditary osteodystrophy
Postreceptor	Type 2 diabetes mellitus, leptin resistance

Abbreviations: AR, androgen receptor; ER, estrogen receptor; GR, glucocorticoid receptor; PPAR, peroxisome proliferator activated receptor; SIADH, syndrome of inappropriate antidiuretic hormone; TR, thyroid hormone receptor; VDR, vitamin D receptor. For all other abbreviations, see text.

Causes of Hormone Excess

Syndromes of hormone excess can be caused by neoplastic growth of endocrine cells, autoimmune disorders, and excess hormone administration. Benign endocrine tumors, including parathyroid, pituitary, and adrenal adenomas, often retain the capacity to produce hormones, perhaps reflecting the fact that they are relatively well differentiated. Many endocrine tumors exhibit subtle defects in their "set points" for feedback regulation. For example, in Cushing's disease, impaired feedback inhibition of ACTH secretion is associated with autonomous function. However, the tumor cells are not completely resistant to feedback, as evidenced by ACTH suppression by higher doses of dexamethasone (e.g., high-dose dexamethasone test) (Chap. 342). Similar set point defects are also typical of parathyroid adenomas and autonomously functioning thyroid nodules.

The molecular basis of some endocrine tumors, such as the MEN syndromes (MEN 1, 2A, 2B), have provided important insights into tumorigenesis (Chap. 351). MEN 1 is characterized primarily by the triad of parathyroid, pancreatic islet, and pituitary tumors. MEN 2 predisposes to medullary thyroid carcinoma, pheochromocytoma, and hyperparathyroidism. The *MEN1* gene, located on chromosome 11q13, encodes a putative tumor-suppressor gene, menin. Analogous to the paradigm first described for retinoblastoma, the affected individual inherits a mutant copy of the *MEN1* gene, and tumorigenesis ensues after a somatic "second hit" leads to loss of function of the normal *MEN1* gene (through deletion or point mutations).

In contrast to inactivation of a tumor-suppressor gene, as occurs in MEN 1 and most other inherited cancer syndromes, MEN 2 is caused by activating mutations in a single allele. In this case, activating

mutations of the *RET* protooncogene, which encodes a receptor tyrosine kinase, leads to thyroid C cell hyperplasia in childhood before the development of medullary thyroid carcinoma. Elucidation of this pathogenic mechanism has allowed early genetic screening for *RET* mutations in individuals at risk for MEN 2, permitting identification of those who may benefit from prophylactic thyroidectomy and biochemical screening for pheochromocytoma and hyperparathyroidism.

Mutations that activate hormone receptor signaling have been identified in several GPCRs. For example, activating mutations of the LH receptor cause a dominantly transmitted form of male-limited precocious puberty, reflecting premature stimulation of testosterone synthesis in Leydig cells (Chap. 346). Activating mutations in these GPCRs are located predominantly in the transmembrane domains and induce receptor coupling to $G_{S}\alpha$ even in the absence of hormone. Consequently, adenylate cyclase is activated, and cyclic adenosine monophosphate (AMP) levels increase in a manner that mimics hormone action. A similar phenomenon results from activating mutations in $G_{S}\alpha$. When these mutations occur early in development, they cause McCune-Albright syndrome. When they occur only in somatotropes, the activating $G_{S}\alpha$ mutations cause GH-secreting tumors and acromegaly (Chap. 339).

In autoimmune Graves' disease, antibody interactions with the TSH receptor mimic TSH action, leading to hormone overproduction (Chap. 341). Analogous to the effects of activating mutations of the TSH receptor, these stimulating autoantibodies induce conformational changes that release the receptor from a constrained state, thereby triggering receptor coupling to G proteins.

Causes of Hormone Deficiency

Most examples of hormone deficiency states can be attributed to glandular destruction caused by autoimmunity, surgery, infection, inflammation, infarction, hemorrhage, or tumor infiltration (Table 338-2). Autoimmune damage to the thyroid gland (Hashimoto's thyroiditis) and pancreatic islet β cells (Type 1 diabetes mellitus) is a prevalent cause of endocrine disease. Mutations in a number of hormones, hormone receptors, transcription factors, enzymes, and channels can also lead to hormone deficiencies.

Hormone Resistance

Most severe hormone resistance syndromes are due to inherited defects in membrane receptors, nuclear receptors, or the pathways that transduce receptor signals. These disorders are characterized by defective hormone action despite the presence of increased hormone levels. In complete androgen resistance, for example, mutations in the androgen receptor result in a female phenotypic appearance in genetic (XY) males, even though LH and testosterone levels are increased (Chap. 349). In addition to these relatively rare genetic disorders, more common acquired forms of functional hormone resistance include insulin resistance in Type 2 diabetes mellitus, leptin resistance in obesity, and GH resistance in catabolic states. The pathogenesis of functional resistance involves receptor downregulation and postreceptor desensitization of signaling pathways; functional forms of resistance are generally reversible.

Approach to the Patient: Endocrine Disease

Because most glands are relatively inaccessible, the examination usually focuses on the manifestations of hormone excess or deficiency as well as direct examination of palpable glands, such as the thyroid and gonads. For these reasons, it is important to evaluate patients in the context of their presenting symptoms, review of systems, family and social history, and exposure to medications that may affect the endocrine system. Astute clinical skills are required to detect subtle symptoms and signs suggestive of underlying endocrine disease. For example, a patient with Cushing's syndrome may manifest specific findings, such as central fat redistribution, striae, and proximal muscle weakness, in addition to features seen commonly in the general population, such as obesity, plethora, hypertension, and glucose intolerance. Similarly, the insidious onset of hypothyroidism—with mental slowing, fatigue, dry skin, and other features—can be difficult to distinguish from similar, nonspecific findings in the general population. Clinical judgment that is based on knowledge of disease prevalence and pathophysiology is required to decide when to embark on more extensive evaluation of these disorders. Laboratory testing plays an essential role in endocrinology by allowing quantitative assessment of hormone levels and dynamics. Radiologic imaging tests such as CT scan, MRI, thyroid scan, and ultrasound are also used for the diagnosis of endocrine disorders. However, these tests generally are employed only after a hormonal abnormality has been established by biochemical testing.

HORMONE MEASUREMENTS AND ENDOCRINE TESTING

Immunoassays are the most important diagnostic tool in endocrinology, as they allow sensitive, specific, and quantitative determination of steady-state and dynamic changes in hormone concentrations. Immunoassays use antibodies to detect specific hormones. For many peptide hormones, these measurements are now configured to use two different antibodies to increase binding affinity and specificity. There are many variations of these assays; a common format involves using one antibody to capture the antigen (hormone) onto an immobilized surface and a second antibody, coupled to a chemiluminescent [immunochemiluminescent assay (ICMA)] or radioactive

immunoradiometric assay (IRMA)] signal, to detect the antigen. These assays are sensitive enough to detect plasma hormone concentrations in the picomolar to nanomolar range, and they can readily distinguish structurally related proteins, such as PTH from PTHrP. A variety of other techniques are used to measure specific hormones, including mass spectroscopy, various forms of chromatography, and enzymatic methods; bioassays are now rarely used.

Most hormone measurements are based on plasma or serum samples. However, urinary hormone determinations remain useful for the evaluation of some conditions. Urinary collections over 24 h provide an integrated assessment of the production of a hormone or metabolite, many of which vary during the day. It is important to assure complete collections of 24-h urine samples; simultaneous measurement of creatinine provides an internal control for the adequacy of collection and can be used to normalize some hormone measurements. A 24-h urine free cortisol measurement largely reflects the amount of unbound cortisol, thus providing a reasonable index of biologically available hormone. Other commonly used urine determinations include 17-hydroxycorticosteroids, 17-ketosteroids, vanillylmandelic acid, metanephrine, catecholamines, 5-hydroxyindoleacetic acid, and calcium.

The value of quantitative hormone measurements lies in their correct interpretation in a clinical context. The normal range for most hormones is relatively broad, often varying by a factor of two- to tenfold. The normal ranges for many hormones are sex- and age-specific. Thus, using the correct normative database is an essential part of interpreting hormone tests. The pulsatile nature of hormones and factors that can affect their secretion, such as sleep, meals, and medications, must also be considered. Cortisol values increase fivefold between midnight and dawn; reproductive hormone levels vary dramatically during the female menstrual cycle.

For many endocrine systems, much information can be gained from basal hormone testing, particularly when different components of an endocrine axis are assessed simultaneously. For example, low testosterone and elevated LH levels suggest a primary gonadal problem, whereas a hypothalamic-pituitary disorder is likely if both LH and testosterone are low. Because TSH is a sensitive indicator of thyroid function, it is generally recommended as a first-line test for thyroid disorders. An elevated TSH level is almost always the result of primary hypothyroidism, whereas a low TSH is most often caused by thyrotoxicosis. These predictions can be confirmed by determining the free thyroxine level. Elevated calcium and PTH levels suggest hyperparathyroidism, whereas PTH is suppressed in hypercalcemia caused by malignancy or granulomatous diseases. A suppressed ACTH in the setting of hypercortisolemia, or increased urine free cortisol, is seen with hyperfunctioning adrenal adenomas.

It is not uncommon, however, for baseline hormone levels associated with pathologic endocrine conditions to overlap with the normal range. In this circumstance, dynamic testing is useful to separate the two groups further. There are a multitude of dynamic endocrine tests, but all are based on principles of feedback regulation, and most responses can be remembered on the basis of the pathways that govern endocrine axes. *Suppression tests* are used in the setting of suspected endocrine hyperfunction. An example is the dexamethasone suppression test used to evaluate Cushing's syndrome (Chaps. 339 and 342). *Stimulation tests* generally are used to assess endocrine hypofunction. The ACTH stimulation test, for example, is used to assess the adrenal gland response in patients with suspected adrenal insufficiency. Other stimulation tests use hypothalamic-releasing factors such as CRH and GHRH to evaluate pituitary hormone reserve (Chap. 339). Insulin-induced hypoglycemia also evokes pituitary ACTH and GH responses. Stimulation tests based on reduction or inhibition of endogenous hormones are now used infrequently. Examples include metyrapone inhibition of cortisol synthesis and clomiphene inhibition of estrogen feedback.

SCREENING AND ASSESSMENT OF COMMON ENDOCRINE DISORDERS

Many endocrine disorders are prevalent in the adult population (Table 338-3) and can be diagnosed and managed by general internists, family practitioners, or other primary health care providers. The high prevalence and clinical impact of certain endocrine diseases justifies vigilance for features of these disorders during routine physical examinations; laboratory screening is indicated in selected high-risk populations.

Table 338-3 Examples of Prevalent Endocrine and Metabolic Disorders in the Adult

Disorder	Approx. Prevalence in Adults[a]	Screening/Testing Recommendations[b]	Chapter
Obesity	31% BMI ≥30 65% BMI ≥25	Calculate BMI Measure waist circumference Exclude secondary causes Consider comorbid complications	78
Type 2 diabetes mellitus	>7%	Beginning at age 45, screen every 3 years, or earlier in high-risk groups: Fasting plasma glucose (FPG) >126 mg/dL Random plasma glucose >200 mg/dL An elevated HbA1c Consider comorbid complications	344
Hyperlipidemia	20–25%	Cholesterol screening at least every 5 years; more often	356

		in high-risk groups Lipoprotein analysis (LDL, HDL) for increased cholesterol, CAD, diabetes Consider secondary causes	
Hypothyroidism	5–10%, women 0.5–2%, men	TSH; confirm with free T4 Screen women after age 35 and every 5 years thereafter	341
Graves' disease	1–3%, women 0.1%, men	TSH, free T4	341
Thyroid nodules and neoplasia	2–5% palpable >25% by ultrasound	Physical examination of thyroid Fine-needle aspiration biopsy	341
Osteoporosis	5–10%, women 2–5%, men	Bone mineral density measurements in women >65 years or in postmenopausal women or men at risk Exclude secondary causes	354
Hyperparathyroidism	0.1–0.5%, women > men	Serum calcium PTH, if calcium is elevated Assess comorbid conditions	353
Infertility	10%, couples	Investigate both members of couple Semen analysis in male Assess ovulatory cycles in female Specific tests as indicated	346, 347
Polycystic ovarian syndrome	5–10%, women	Free testosterone, DHEAS Consider comorbid conditions	347
Hirsutism	5–10%	Free testosterone, DHEAS Exclude secondary causes Additional tests as indicated	49
Menopause	Median age, 51	FSH	348
Hyperprolactinemia	15% in women with amenorrhea or galactorrhea	PRL level MRI, if not medication-related	339
Erectile dysfunction	20–30%	Careful history, PRL, testosterone Consider secondary causes (e.g., diabetes)	48
Gynecomastia	15%	Often, no tests are indicated Consider Klinefelter syndrome Consider medications, hypogonadism, liver disease	346
Klinefelter syndrome	0.2%, men	Karyotype Testosterone	349
Vitamin D deficiency	40–50%	Measure serum 25-OH vitamin D Consider secondary causes	352
Turner syndrome	0.03%, women	Karyotype Consider comorbid conditions	349

[a] The prevalence of most disorders varies among ethnic groups and with aging. Data based primarily on U.S. population.

[b] See individual chapters for additional information on evaluation and treatment. Early testing is indicated in patients with signs and symptoms of disease and in those at increased risk.

Abbreviations: BMI, body mass index; CAD, coronary artery disease; DHEAS, dehydroepiandrosterone; HDL, high-density lipoprotein; LDL, low-density lipoprotein. For other abbreviations, see text.

FURTHER READINGS

Golden SH et al: Clinical review: Prevalence and incidence of endocrine and metabolic disorders in the United States: A comprehensive review. J Clin Endocrinol Metab 94:1853, 2009[PMID: 19494161] [Full Text]

Jameson JL, Degroot LJ (eds): *Endocrinology*, 6th ed. Philadelphia, Elsevier, 2010

Kleinau G, Krause G: Thyrotropin and homologous glycoprotein hormone receptors: Structural and functional aspects of extracellular signaling mechanisms. Endocr Rev 30:133, 2009[PMID: 19176466] [Full Text]

Marx SJ, Simonds WF: Hereditary hormone excess: Genes, molecular pathways, and syndromes. Endocr Rev 26:615, 2005[PMID: 15632315] [Full Text]

Veldhuis JD et al: Motivations and methods for analyzing pulsatile hormone secretion. Endocr Rev 29:823, 2008[PMID: 18940916] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE ANTERIOR PITUITARY AND HYPOTHALAMUS: INTRODUCTION

The anterior pituitary often is referred to as the "master gland" because, together with the hypothalamus, it orchestrates the complex regulatory functions of many other endocrine glands. The anterior pituitary gland produces six major hormones: (1) prolactin (PRL), (2) growth hormone (GH), (3) adrenocorticotropic hormone (ACTH), (4) luteinizing hormone (LH), (5) follicle-stimulating hormone (FSH), and (6) thyroid-stimulating hormone (TSH) (Table 339-1). Pituitary hormones are secreted in a pulsatile manner, reflecting stimulation by an array of specific hypothalamic releasing factors. Each of these pituitary hormones elicits specific responses in peripheral target tissues. The hormonal products of those peripheral glands, in turn, exert feedback control at the level of the hypothalamus and pituitary to modulate pituitary function (Fig. 339-1). Pituitary tumors cause characteristic hormone-excess syndromes. Hormone deficiency may be inherited or acquired. Fortunately, there are efficacious treatments for the various pituitary hormone-excess and -deficiency syndromes. Nonetheless, these diagnoses are often elusive; this emphasizes the importance of recognizing subtle clinical manifestations and performing the correct laboratory diagnostic tests. **For discussion of disorders of the posterior pituitary, or neurohypophysis, see Chap. 340.**

Table 339-1 Anterior Pituitary Hormone Expression and Regulation

Cell	Corticotrope	Somatotrope	Lactotrope	Thyrotrope	Gonadotrope
Tissue-specific transcription factor	T-Pit	Prop-1, Pit-1	Prop-1, Pit-1	Prop-1, Pit-1, TEF	SF-1, DAX-1
Fetal appearance	6 weeks	8 weeks	12 weeks	12 weeks	12 weeks
Hormone	POMC	GH	PRL	TSH	FSH LH
Protein	Polypeptide	Polypeptide	Polypeptide	Glycoprotein α, β subunits	Glycoprotein α, β subunits
Amino acids	266 (ACTH 1–39)	191	199	211	210 204
Stimulators	CRH, AVP, gp-130 cytokines	GHRH, ghrelin	Estrogen, TRH, VIP	TRH	GnRH, activins, estrogen
Inhibitors	Glucocorticoids	Somatostatin, IGF-I	Dopamine	T_3, T4, dopamine, somatostatin, glucocorticoids	Sex steroids, inhibin
Target gland	Adrenal	Liver, other tissues	Breast, other tissues	Thyroid	Ovary, testis
Trophic effect	Steroid production	IGF-I production, growth induction, insulin antagonism	Milk production	T4 synthesis and secretion	Sex steroid production, follicle growth, germ cell maturation
Normal range	ACTH, 4–22 pg/L	<0.5 μg/L*	M < 15; F <20 μg/L	0.1–5 mU/L	M, 5–20 IU/L, F (basal), 5–20 IU/L

*Hormone secretion integrated over 24 h.

Abbreviations: M, male; F, female. For other abbreviations, see text.

Source: Adapted from I Shimon, S Melmed, in S Melmed, P Conn (eds): *Endocrinology: Basic and Clinical Principles*. Totowa, NJ, Humana, 2005.

ANATOMY AND DEVELOPMENT

Anatomy

The pituitary gland weighs ~600 mg and is located within the sella turcica ventral to the diaphragma sella; it consists of anatomically and functionally distinct anterior and posterior lobes. The bony sella is contiguous to vascular and neurologic structures, including the cavernous sinuses, cranial nerves, and optic chiasm. Thus, expanding intrasellar pathologic processes may have significant central mass effects in addition to their endocrinologic impact.

Figure 339-1

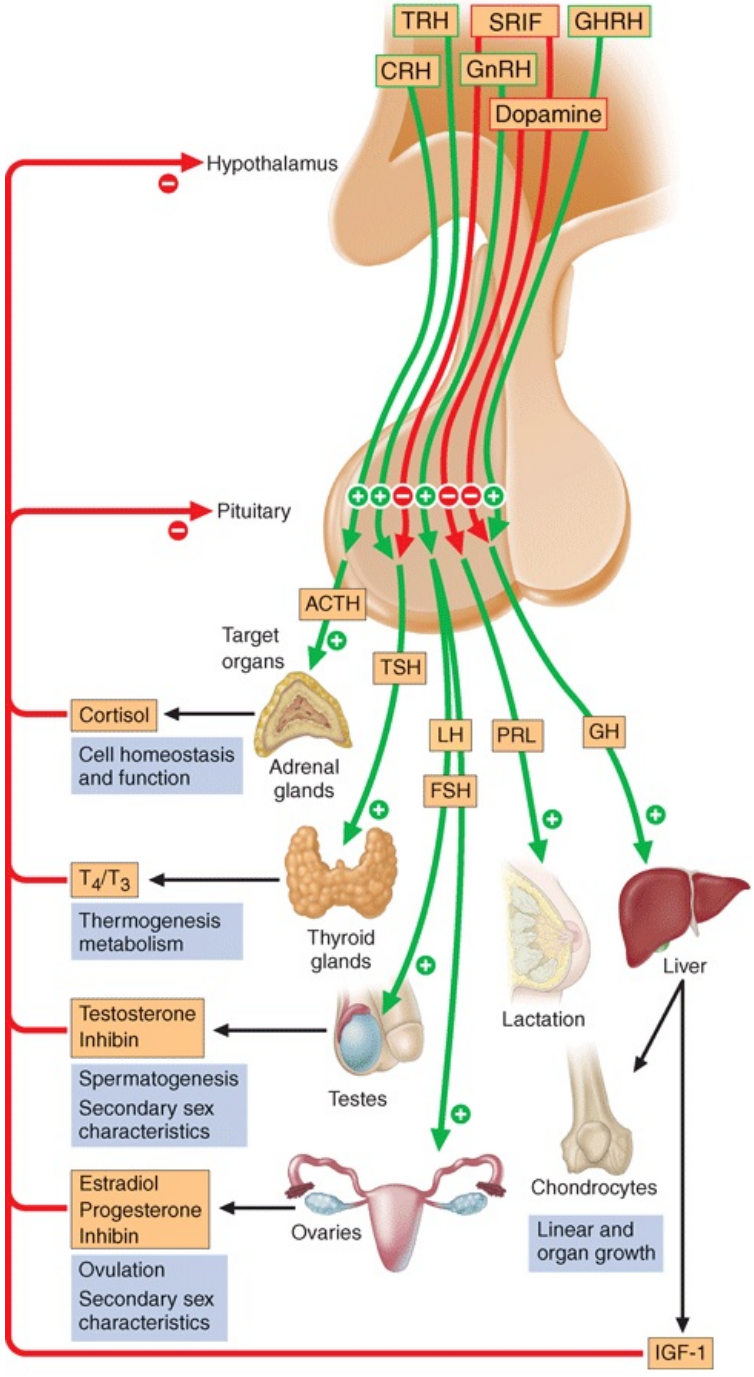

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of pituitary axes. Hypothalamic hormones regulate anterior pituitary trophic hormones that in turn determine target gland secretion. Peripheral hormones feed back to regulate hypothalamic and pituitary hormones. For abbreviations, see text.

Hypothalamic neural cells synthesize specific releasing and inhibiting hormones that are secreted directly into the portal vessels of the pituitary stalk. Blood supply of the pituitary gland comes from the superior and inferior hypophyseal arteries (Fig. 339-2). The hypothalamic-pituitary portal plexus provides the major blood source for the anterior pituitary, allowing reliable transmission of

hypothalamic peptide pulses without significant systemic dilution; consequently, pituitary cells are exposed to releasing or inhibiting factors and in turn release their hormones as discrete pulses (Fig. 339-3).

Figure 339-2

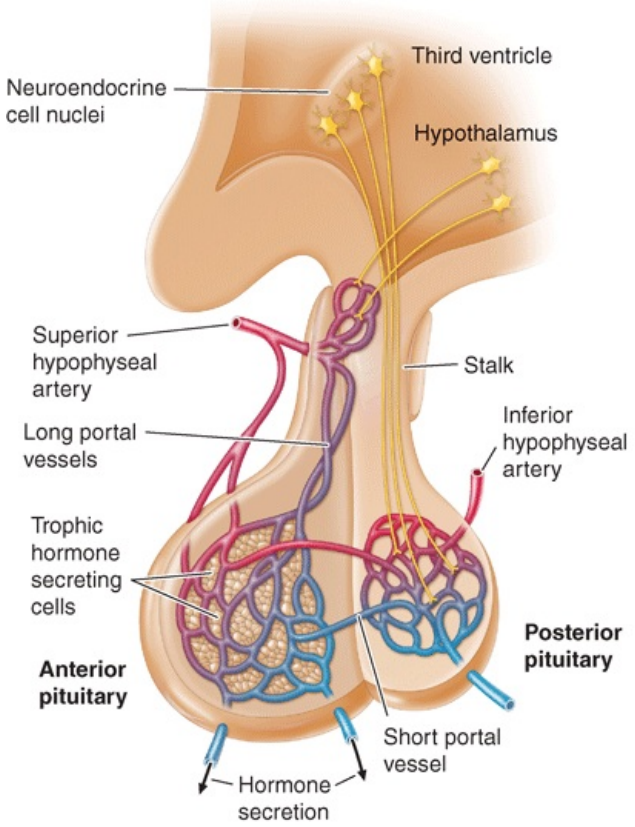

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of hypothalamic-pituitary vasculature. The hypothalamic nuclei produce hormones that traverse the portal system and impinge on anterior pituitary cells to regulate pituitary hormone secretion. Posterior pituitary hormones are derived from direct neural extensions.

Figure 339-3

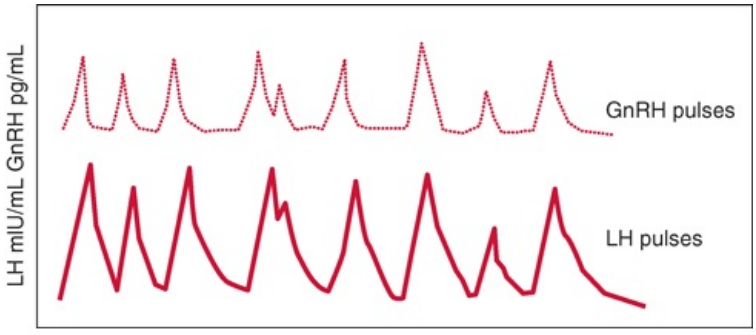

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypothalamic gonadotropin-releasing hormone (GnRH) pulses induce secretory pulses of luteinizing hormone (LH).

The posterior pituitary is supplied by the inferior hypophyseal arteries. In contrast to the anterior pituitary, the posterior lobe is directly innervated by hypothalamic neurons (supraopticohypophyseal and tuberohypophyseal nerve tracts) via the pituitary stalk (Chap. 340). Thus, posterior pituitary production of vasopressin [antidiuretic hormone (ADH)] and oxytocin is particularly sensitive to neuronal damage by lesions that affect the pituitary stalk or hypothalamus.

Pituitary Development

The embryonic differentiation and maturation of anterior pituitary cells have been elucidated in considerable detail. Pituitary development from Rathke's pouch involves a complex interplay of lineage-specific transcription factors expressed in pluripotent precursor cells and gradients of locally produced growth factors (Table 339-1). The transcription factor Prop-1 induces pituitary development of Pit-1-specific lineages as well as gonadotropes. The transcription factor Pit-1 determines cell-specific expression of GH, PRL, and TSH in somatotropes, lactotropes, and thyrotropes. Expression of high levels of estrogen receptors in cells that contain Pit-1 favors PRL expression, whereas thyrotrope embryonic factor (TEF) induces TSH expression. Pit-1 binds to GH, PRL, and TSH gene regulatory elements as well as to recognition sites on its own promoter, providing a mechanism for maintaining specific pituitary phenotypic stability. Gonadotrope cell development is further defined by the cell-specific expression of the nuclear receptors steroidogenic factor (SF-1) and dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX-1). Development of corticotrope cells, which express the proopiomelanocortin (POMC) gene, requires the T-Pit transcription factor. Abnormalities of pituitary development caused by mutations of Pit-1, Prop-1, SF-1, DAX-1, and T-Pit result in a series of rare, selective or combined pituitary hormone deficits.

HYPOTHALAMIC AND ANTERIOR PITUITARY INSUFFICIENCY

Hypopituitarism results from impaired production of one or more of the anterior pituitary trophic hormones. Reduced pituitary function can result from inherited disorders; more commonly, hypopituitarism is acquired and reflects the compressive mass effects of tumors or the consequences of inflammation or vascular damage. These processes also may impair synthesis or secretion of hypothalamic hormones, with resultant pituitary failure (Table 339-2).

Table 339-2 Etiology of Hypopituitarism*

Development/structural
Transcription factor defect
Pituitary dysplasia/aplasia
Congenital CNS mass, encephalocele
Primary empty sella
Congenital hypothalamic disorders (septo-optic dysplasia,
Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,
Kallmann syndrome)
Traumatic
Surgical resection
Radiation damage
Head injuries
Neoplastic
Pituitary adenoma
Parasellar mass (germinoma, ependymoma, glioma)
Rathke's cyst
Craniopharyngioma
Hypothalamic hamartoma, gangliocytoma
Pituitary metastases (breast, lung, colon carcinoma)
Lymphoma and leukemia
Meningioma
Infiltrative/inflammatory
Lymphocytic hypophysitis
Hemochromatosis
Sarcoidosis
Histiocytosis X
Granulomatous hypophysitis
Vascular
Pituitary apoplexy
Pregnancy-related (infarction with diabetes; postpartum necrosis)
Sickle cell disease
Arteritis

Infections
Fungal (histoplasmosis)
Parasitic (toxoplasmosis)
Tuberculosis
Pneumocystis carinii

*Trophic hormone failure associated with pituitary compression or destruction usually occurs sequentially: GH > FSH > LH > TSH > ACTH. During childhood, growth retardation is often the presenting feature, and in adults, hypogonadism is the earliest symptom.

Developmental and Genetic Causes of Hypopituitarism

PITUITARY DYSPLASIA

Pituitary dysplasia may result in aplastic, hypoplastic, or ectopic pituitary gland development. Because pituitary development follows midline cell migration from the nasopharyngeal Rathke's pouch, midline craniofacial disorders may be associated with pituitary dysplasia. Acquired pituitary failure in the newborn also can be caused by birth trauma, including cranial hemorrhage, asphyxia, and breech delivery.

Septo-Optic Dysplasia

Hypothalamic dysfunction and hypopituitarism may result from dysgenesis of the septum pellucidum or corpus callosum. Affected children have mutations in the *HESX1* gene, which is involved in early development of the ventral prosencephalon. These children exhibit variable combinations of cleft palate, syndactyly, ear deformities, hypertelorism, optic atrophy, micropenis, and anosmia. Pituitary dysfunction leads to diabetes insipidus, GH deficiency and short stature, and, occasionally, TSH deficiency.

TISSUE-SPECIFIC FACTOR MUTATIONS

Several pituitary cell–specific transcription factors, such as Pit-1 and Prop-1, are critical for determining the development and committed function of differentiated anterior pituitary cell lineages. Autosomal dominant or recessive Pit-1 mutations cause combined GH, PRL, and TSH deficiencies. These patients usually present with growth failure and varying degrees of hypothyroidism. The pituitary may appear hypoplastic on MRI.

Prop-1 is expressed early in pituitary development and appears to be required for Pit-1 function. Familial and sporadic *PROP1* mutations result in combined GH, PRL, TSH, and gonadotropin deficiency. Over 80% of these patients have growth retardation; by adulthood, all are deficient in TSH and gonadotropins, and a small minority later develop ACTH deficiency. Because of gonadotropin deficiency, these individuals do not enter puberty spontaneously. In some cases, the pituitary gland is enlarged. *TPIT* mutations result in ACTH deficiency associated with hypocortisolism.

DEVELOPMENTAL HYPOTHALAMIC DYSFUNCTION

Kallmann Syndrome

Kallmann syndrome results from defective hypothalamic gonadotropin-releasing hormone (GnRH) synthesis and is associated with anosmia or hyposmia due to olfactory bulb agenesis or hypoplasia (Chap. 346). The syndrome also may be associated with color blindness, optic atrophy, nerve deafness, cleft palate, renal abnormalities, cryptorchidism, and neurologic abnormalities such as mirror movements. Defects in the X-linked *KAL* gene impair embryonic migration of GnRH neurons from the hypothalamic olfactory placode to the hypothalamus. Genetic abnormalities, in addition to *KAL* mutations, also can cause isolated GnRH deficiency. Autosomal recessive (i.e., *GPR54, KISS1*) and dominant (i.e., *FGFR1*) modes of transmission have been described, and there is a growing list of genes associated with GnRH deficiency (*GNRH1, PROK2, PROKR2, CH7, PCSK1, FGF8, TAC3, TACR3*). GnRH deficiency prevents progression through puberty. Males present with delayed puberty and pronounced hypogonadal features, including micropenis, probably the result of low testosterone levels during infancy. Females present with primary amenorrhea and failure of secondary sexual development.

Kallmann syndrome and other causes of congenital GnRH deficiency are characterized by low LH and FSH levels and low concentrations of sex steroids (testosterone or estradiol). In sporadic cases of isolated gonadotropin deficiency, the diagnosis is often one of exclusion after other causes of hypothalamic-pituitary dysfunction have been eliminated. Repetitive GnRH administration restores normal pituitary gonadotropin responses, pointing to a hypothalamic defect.

Long-term treatment of men with human chorionic gonadotropin (hCG) or testosterone restores pubertal development and secondary sex characteristics; women can be treated with cyclic estrogen and progestin. Fertility also may be restored by the administration of gonadotropins or by using a portable infusion pump to deliver subcutaneous, pulsatile GnRH.

Bardet-Biedl Syndrome

This is a rare genetically heterogeneous disorder characterized by mental retardation, renal abnormalities, obesity, and hexadactyly, brachydactyly, or syndactyly. Central diabetes insipidus may or may not be associated. GnRH deficiency occurs in 75% of males and half of affected females. Retinal degeneration begins in early childhood, and most patients are blind by age 30. Numerous subtypes of Bardet-Biedl syndrome (BBS) have been identified, with genetic linkage to at least nine different loci. Several of the loci encode genes involved in basal body cilia function, and this may account for the diverse clinical manifestations.

Leptin and Leptin Receptor Mutations

Deficiencies of leptin or its receptor cause a broad spectrum of hypothalamic abnormalities, including hyperphagia, obesity, and central hypogonadism (Chap. 77). Decreased GnRH production in these patients results in attenuated pituitary FSH and LH synthesis and release.

PRADER-WILLI SYNDROME

This is a contiguous gene syndrome that results from deletion of the paternal copies of the imprinted *SNRPN* gene, the *NECDIN* gene, and possibly other genes on chromosome 15q. Prader-Willi syndrome is associated with hypogonadotropic hypogonadism, hyperphagia-obesity, chronic muscle hypotonia, mental retardation, and adult-onset diabetes mellitus (Chap. 62). Multiple somatic defects also involve the skull, eyes, ears, hands, and feet. Diminished hypothalamic oxytocin- and vasopressin-producing nuclei have been reported. Deficient GnRH synthesis is suggested by the observation that chronic GnRH treatment restores pituitary LH and FSH release.

Acquired Hypopituitarism

Hypopituitarism may be caused by accidental or neurosurgical trauma; vascular events such as apoplexy; pituitary or hypothalamic neoplasms, craniopharyngioma, lymphoma, or metastatic tumors; inflammatory disease such as lymphocytic hypophysitis; infiltrative disorders such as sarcoidosis, hemochromatosis (Chap. 357), and tuberculosis; or irradiation.

Increasing evidence suggests that patients with brain injury, including sports trauma, subarachnoid hemorrhage, and irradiation, have transient hypopituitarism and require intermittent long-term endocrine follow-up, as permanent hypothalamic or pituitary dysfunction will develop in 25–40% of these patients.

HYPOTHALAMIC INFILTRATION DISORDERS

These disorders—including sarcoidosis, histiocytosis X, amyloidosis, and hemochromatosis —frequently involve both hypothalamic and pituitary neuronal and neurochemical tracts. Consequently, diabetes insipidus occurs in half of patients with these disorders. Growth retardation is seen if attenuated GH secretion occurs before pubertal epiphyseal closure. Hypogonadotropic hypogonadism and hyperprolactinemia are also common.

INFLAMMATORY LESIONS

Pituitary damage and subsequent dysfunction can be seen with chronic infections such as tuberculosis, with opportunistic fungal infections associated with AIDS, and in tertiary syphilis. Other inflammatory processes, such as granulomas and sarcoidosis, may mimic the features of a pituitary adenoma. These lesions may cause extensive hypothalamic and pituitary damage, leading to trophic hormone deficiencies.

CRANIAL IRRADIATION

Cranial irradiation may result in long-term hypothalamic and pituitary dysfunction, especially in children and adolescents, as they are more susceptible to damage after whole-brain or head and neck therapeutic irradiation. The development of hormonal abnormalities correlates strongly with irradiation dosage and the time interval after completion of radiotherapy. Up to two-thirds of patients ultimately develop hormone insufficiency after a median dose of 50 Gy (5000 rad) directed at the skull base. The development of hypopituitarism occurs over 5–15 years and usually reflects hypothalamic damage rather than primary destruction of pituitary cells. Although the pattern of hormone loss is variable, GH deficiency is most common, followed by gonadotropin and ACTH deficiency. When deficiency of one or more hormones is documented, the possibility of diminished reserve of other hormones is likely. Accordingly, anterior pituitary function should be continually evaluated over the long term in previously irradiated patients, and replacement therapy instituted when appropriate (see below).

LYMPHOCYTIC HYPOPHYSITIS

This occurs most often in postpartum women; it usually presents with hyperprolactinemia and MRI evidence of a prominent pituitary mass that often resembles an adenoma, with mildly elevated PRL levels. Pituitary failure caused by diffuse lymphocytic infiltration may be transient or permanent but requires immediate evaluation and treatment. Rarely, isolated pituitary hormone deficiencies have been described, suggesting a selective autoimmune process targeted to specific cell types. Most patients manifest symptoms of progressive mass effects with headache and visual disturbance. The erythrocyte sedimentation rate often is elevated. As the MRI image may be indistinguishable from that of a pituitary adenoma, hypophysitis should be considered in a postpartum woman with a newly

diagnosed pituitary mass before an unnecessary surgical intervention is undertaken. The inflammatory process often resolves after several months of glucocorticoid treatment, and pituitary function may be restored, depending on the extent of damage.

PITUITARY APOPLEXY

Acute intrapituitary hemorrhagic vascular events can cause substantial damage to the pituitary and surrounding sellar structures. Pituitary apoplexy may occur spontaneously in a preexisting adenoma; postpartum (Sheehan's syndrome); or in association with diabetes, hypertension, sickle cell anemia, or acute shock. The hyperplastic enlargement of the pituitary, which occurs normally during pregnancy, increases the risk for hemorrhage and infarction. Apoplexy is an endocrine emergency that may result in severe hypoglycemia, hypotension and shock, central nervous system (CNS) hemorrhage, and death. Acute symptoms may include severe headache with signs of meningeal irritation, bilateral visual changes, ophthalmoplegia, and, in severe cases, cardiovascular collapse and loss of consciousness. Pituitary CT or MRI may reveal signs of intratumoral or sellar hemorrhage, with deviation of the pituitary stalk and compression of pituitary tissue.

Patients with no evident visual loss or impaired consciousness can be observed and managed conservatively with high-dose glucocorticoids. Those with significant or progressive visual loss or loss of consciousness require urgent surgical decompression. Visual recovery after sellar surgery is inversely correlated with the length of time after the acute event. Therefore, severe ophthalmoplegia or visual deficits are indications for early surgery. Hypopituitarism is very common after apoplexy.

EMPTY SELLA

A partial or apparently totally empty sella is often an incidental MRI finding. These patients usually have normal pituitary function, implying that the surrounding rim of pituitary tissue is fully functional. Hypopituitarism, however, may develop insidiously. Pituitary masses also may undergo clinically silent infarction and involution with development of a partial or totally empty sella by cerebrospinal fluid (CSF) filling the dural herniation. Rarely, small but functional pituitary adenomas may arise within the rim of pituitary tissue, and they are not always visible on MRI.

Presentation and Diagnosis

The clinical manifestations of hypopituitarism depend on which hormones are lost and the extent of the hormone deficiency. GH deficiency causes growth disorders in children and leads to abnormal body composition in adults (see below). Gonadotropin deficiency causes menstrual disorders and infertility in women and decreased sexual function, infertility, and loss of secondary sexual characteristics in men. TSH and ACTH deficiency usually develop later in the course of pituitary failure. TSH deficiency causes growth retardation in children and features of hypothyroidism in children and adults. The secondary form of adrenal insufficiency caused by ACTH deficiency leads to hypocortisolism with relative preservation of mineralocorticoid production. PRL deficiency causes failure of lactation. When lesions involve the posterior pituitary, polyuria and polydipsia reflect loss of vasopressin secretion. Epidemiologic studies have documented an increased mortality rate in patients with long-standing pituitary damage, primarily from increased cardiovascular and cerebrovascular disease. Previous head or neck irradiation is also a determinant of increased mortality rates in patients with hypopituitarism.

Laboratory Investigation

Biochemical diagnosis of pituitary insufficiency is made by demonstrating low levels of trophic hormones in the setting of low levels of target hormones. For example, low free thyroxine in the setting of a low or inappropriately normal TSH level suggests secondary hypothyroidism. Similarly, a low testosterone level without elevation of gonadotropins suggests hypogonadotropic hypogonadism. Provocative tests may be required to assess pituitary reserve (Table 339-3). GH responses to insulin-induced hypoglycemia, arginine, l-dopa, growth hormone–releasing hormone (GHRH), or growth hormone–releasing peptides (GHRPs) can be used to assess GH reserve. Corticotropin-releasing hormone (CRH) administration induces ACTH release, and administration of synthetic ACTH (cosyntropin) evokes adrenal cortisol release as an indirect indicator of pituitary ACTH reserve (Chap. 342). ACTH reserve is most reliably assessed by measuring ACTH and cortisol levels during insulin-induced hypoglycemia. However, this test should be performed cautiously in patients with suspected adrenal insufficiency because of enhanced susceptibility to hypoglycemia and hypotension. Administering insulin to induce hypoglycemia is contraindicated in patients with active coronary artery disease or seizure disorders.

Table 339-3 Tests of Pituitary Sufficiency

Hormone	Test	Blood Samples	Interpretation
Growth hormone	Insulin tolerance test: Regular insulin (0.05 –0.15 U/kg IV)	–30, 0, 30, 60, 120 min for glucose and GH	Glucose < 40 mg/dL; GH should be >3 μg/L

	GHRH test: 1 μg/kg IV	0, 15, 30, 45, 60, 120 min for GH	Normal response is GH >3 μg/L	
	L-Arginine test: 30 g IV over 30 min	0, 30, 60, 120 min for GH	Normal response is GH >3 μg/L	
	L-Dopa test: 500 mg PO	0, 30, 60, 120 min for GH	Normal response is GH >3 μg/L	
	Prolactin	TRH test: 200–500 μg IV	0, 20, and 60 min for TSH and PRL	Normal prolactin is >2 μg/L and increase >200% of baseline
ACTH	Insulin tolerance test: regular insulin (0.05–0.15 U/kg IV)	–30, 0, 30, 60, 90 min for glucose and cortisol	Glucose <40 mg/dL Cortisol should increase by >7 μg/dL or to >20 μg/dL	
	CRH test: 1 μg/kg ovine CRH IV at 8 A.M.	0, 15, 30, 60, 90, 120 min for ACTH and cortisol	Basal ACTH increases 2- to 4-fold and peaks at 20–100 pg/mL Cortisol levels >20–25 μg/dL	
	Metyrapone test: Metyrapone (30 mg/kg) at midnight	Plasma 11-deoxycortisol and cortisol at 8 A.M.; ACTH can also be measured	Plasma cortisol should be <4 μg/dL to assure an adequate response Normal response is 11-deoxycortisol >7.5 μg/dL or ACTH >75 pg/mL	
	Standard ACTH stimulation test: ACTH 1-24 (cosyntropin), 0.25 mg IM or IV	0, 30, 60 min for cortisol and aldosterone	Normal response is cortisol >21 μg/dL and aldosterone response of >4 ng/dL above baseline	
	Low-dose ACTH test: ACTH 1-24 (cosyntropin), 1 μg IV	0, 30, 60 min for cortisol	Cortisol should be >21 μg/dL	
	3-day ACTH stimulation test consists of 0.25 mg ACTH 1-24 given IV over 8 h each day		Cortisol >21 μg/dL	
TSH	Basal thyroid function tests: T_4, T_3, TSH	Basal measurements	Low free thyroid hormone levels in the setting of TSH levels that are not appropriately increased indicate pituitary insufficiency	
	TRH test: 200–500 μg IV	0, 20, 60 min for TSH and PRL[a]	TSH should increase by >5 mU/L unless thyroid hormone levels are increased	
LH, FSH	LH, FSH, testosterone, estrogen	Basal measurements	Basal LH and FSH should be increased in postmenopausal women Low testosterone levels in the setting of low LH and FSH indicate pituitary insufficiency	
	GnRH test: GnRH (100 μg) IV	0, 30, 60 min for LH and FSH	In most adults, LH should increase by 10 IU/L and FSH by 2 IU/L Normal responses are variable	
Multiple hormones	Combined anterior pituitary test: GHRH (1 μg/kg), CRH (1 μg/kg), GnRH (100 μg), TRH (200 μg) are given IV	–30, 0, 15, 30, 60, 90, 120 min for GH, ACTH, cortisol, LH, FSH, and TSH	Combined or individual releasing hormone responses must be elevated in the context of basal target gland hormone values and may not be uniformly diagnostic (see text)	

[a]Evoked PRL response indicates lactotrope integrity. ***Note:*** For abbreviations, see text.

Treatment: Hypopituitarism

Hormone replacement therapy, including glucocorticoids, thyroid hormone, sex steroids, growth hormone, and vasopressin, is usually safe and free of complications. Treatment regimens that mimic physiologic hormone production allow for maintenance of satisfactory clinical homeostasis. Effective dosage schedules are outlined in Table 339-4. Patients in need of glucocorticoid replacement require careful dose adjustments during stressful events such as acute illness, dental procedures, trauma, and acute hospitalization.

Table 339-4 Hormone Replacement Therapy for Adult Hypopituitarism*

Trophic Hormone Deficit	Hormone Replacement
ACTH	Hydrocortisone (10–20 mg A.M.; 5–10 mg P.M.) Cortisone acetate (25 mg A.M.; 12.5 mg P.M.) Prednisone (5 mg A.M.)
TSH	L-Thyroxine (0.075–0.15 mg daily)
FSH/LH	Males Testosterone enanthate (200 mg IM every 2 weeks) Testosterone skin patch (5 mg/d) Females Conjugated estrogen (0.65–1.25 mg qd for 25 days) Progesterone (5–10 mg qd) on days 16–25 Estradiol skin patch (0.5 mg, every other day) For fertility: Menopausal gonadotropins, human chorionic gonadotropins
GH	Adults: Somatotropin (0.1–1.25 mg SC qd) Children: Somatotropin [0.02–0.05 (mg/kg per day)]
Vasopressin	Intranasal desmopressin (5–20 μg twice daily) Oral 300–600 μg qd

*All doses shown should be individualized for specific patients and should be reassessed during stress, surgery, or pregnancy. Male and female fertility requirements should be managed as discussed in Chap. 44.

Note: For abbreviations, see text.

HYPOTHALAMIC, PITUITARY, AND OTHER SELLAR MASSES

Pituitary Tumors

Pituitary adenomas are the most common cause of pituitary hormone hypersecretion and hyposecretion syndromes in adults. They account for ~15% of all intracranial neoplasms and have been identified with a population prevalence of ~80/100,000. At autopsy, up to one-quarter of all pituitary glands harbor an unsuspected microadenoma (<10 mm diameter). Similarly, pituitary imaging detects small clinically inapparent pituitary lesions in at least 10% of individuals.

PATHOGENESIS

Pituitary adenomas are benign neoplasms that arise from one of the five anterior pituitary cell types. The clinical and biochemical phenotypes of pituitary adenomas depend on the cell type from which they are derived. Thus, tumors arising from lactotrope (PRL), somatotrope (GH), corticotrope (ACTH), thyrotrope (TSH), or gonadotrope (LH, FSH) cells hypersecrete their respective hormones (Table 339-5). Plurihormonal tumors that express combinations of GH, PRL, TSH, ACTH, and the glycoprotein hormone α or β subunit may be diagnosed by careful immunocytochemistry or may manifest as clinical syndromes that combine features of these hormonal hypersecretory syndromes. Morphologically, these tumors may arise from a single polysecreting cell type or include cells with mixed function within the same tumor.

Table 339-5 Classification of Pituitary Adenomas*

Adenoma Cell Origin	Hormone Product	Clinical Syndrome
Lactotrope	PRL	Hypogonadism, galactorrhea
Gonadotrope	FSH, LH, subunits	Silent or hypogonadism

Somatotrope	GH	Acromegaly/gigantism
Corticotrope	ACTH	Cushing's disease
Mixed growth hormone and prolactin cell	GH, PRL	Acromegaly, hypogonadism, galactorrhea
Other plurihormonal cell	Any	Mixed
Acidophil stem cell	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Mammosomatotrope	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Thyrotrope	TSH	Thyrotoxicosis
Null cell	None	Pituitary failure
Oncocytoma	None	Pituitary failure

*Hormone-secreting tumors are listed in decreasing order of frequency. All tumors may cause local pressure effects, including visual disturbances, cranial nerve palsy, and headache. ***Note:*** For abbreviations, see text. ***Source:*** Adapted from S Melmed, in JL Jameson (ed): *Principles of Molecular Medicine,* Totowa, NJ, Humana Press, 1998.

Hormonally active tumors are characterized by autonomous hormone secretion with diminished feedback responsiveness to physiologic inhibitory pathways. Hormone production does not always correlate with tumor size. Small hormone-secreting adenomas may cause significant clinical perturbations, whereas larger adenomas that produce less hormone may be clinically silent and remain undiagnosed (if no central compressive effects occur). About one-third of all adenomas are clinically nonfunctioning and produce no distinct clinical hypersecretory syndrome. Most of them arise from gonadotrope cells and may secrete small amounts of α- and β-glycoprotein hormone subunits or, very rarely, intact circulating gonadotropins. True pituitary carcinomas with documented extracranial metastases are exceedingly rare.

Almost all pituitary adenomas are monoclonal in origin, implying the acquisition of one or more somatic mutations that confer a selective growth advantage. Consistent with their clonal origin, complete surgical resection of small pituitary adenomas usually cures hormone hypersecretion. Nevertheless, hypothalamic hormones such as GHRH and CRH also enhance mitotic activity of their respective pituitary target cells in addition to their role in pituitary hormone regulation. Thus, patients who harbor rare abdominal or chest tumors that elaborate ectopic GHRH or CRH may present with somatotrope or corticotrope hyperplasia with GH or ACTH hypersecretion.

Several etiologic genetic events have been implicated in the development of pituitary tumors. The pathogenesis of sporadic forms of acromegaly has been particularly informative as a model of tumorigenesis. GHRH, after binding to its G protein–coupled somatotrope receptor, utilizes cyclic AMP (adenosine monophosphate) as a second messenger to stimulate GH secretion and somatotrope proliferation. A subset (~35%) of GH-secreting pituitary tumors contain sporadic mutations in $Gs\alpha$ (Arg 201 → Cys or His; Gln 227 → Arg). These mutations attenuate intrinsic GTPase activity, resulting in constitutive elevation of cyclic AMP, Pit-1 induction, and activation of cyclic AMP response element binding protein (CREB), thereby promoting somatotrope cell proliferation and GH secretion.

Characteristic loss of heterozygosity (LOH) in various chromosomes has been documented in large or invasive macroadenomas, suggesting the presence of putative tumor suppressor genes at these loci. LOH of chromosome regions on 11q13, 13, and 9 is present in up to 20% of sporadic pituitary tumors, including GH-, PRL-, and ACTH-producing adenomas and some nonfunctioning tumors.

Compelling evidence also favors growth factor promotion of pituitary tumor proliferation. Basic fibroblast growth factor (bFGF) is abundant in the pituitary and has been shown to stimulate pituitary cell mitogenesis. Other factors involved in initiation and promotion of pituitary tumors include loss of negative-feedback inhibition (as seen with primary hypothyroidism or hypogonadism) and estrogen-mediated or paracrine angiogenesis. Growth characteristics and neoplastic behavior also may be influenced by several activated oncogenes, including *RAS* and pituitary tumor transforming gene (*PTTG*), or inactivation of growth suppressor genes, including *MEG3*.

GENETIC SYNDROMES ASSOCIATED WITH PITUITARY TUMORS

Several familial syndromes are associated with pituitary tumors, and the genetic mechanisms for some of them have been unraveled (Table 339-6).

Table 339-6 Familial Pituitary Tumor Syndromes

	Gene Mutated	Clinical Features
Multiple endocrine neoplasia 1 (MEN 1)	*MEN1* (11q13)	Hyperparathyroidism Pancreatic neuroendocrine tumors

		Foregut carcinoids Adrenal adenomas Skin lesions Pituitary adenomas (40%)
Multiple endocrine neoplasia 4 (MEN 4)	*CDKNIB (12p13)*	Hyperparathyroidsm Pituitary adenomas Other tumors
Carney complex	*PRKAR1A 17q23-24*	Pituitary hyperplasia and adenomas (10%) Atrial myxomas Schwannomas Adrenal hyperplasia Lentigines
Familial pituitary adenomas	*AIP (11q13.3)*	Acromegaly/gigantism (15%)

Multiple endocrine neoplasia (MEN) 1 is an autosomal dominant syndrome characterized primarily by a genetic predisposition to parathyroid, pancreatic islet, and pituitary adenomas (Chap. 351). MEN1 is caused by inactivating germ-line mutations in *MENIN*, a constitutively expressed tumor-suppressor gene located on chromosome 11q13. Loss of heterozygosity, or a somatic mutation of the remaining normal *MENIN* allele, leads to tumorigenesis. About half of affected patients develop prolactinomas; acromegaly and Cushing's syndrome are less commonly encountered.

Carney syndrome is characterized by spotty skin pigmentation, myxomas, and endocrine tumors, including testicular, adrenal, and pituitary adenomas. Acromegaly occurs in about 20% of these patients. A subset of patients have mutations in the R1α regulatory subunit of protein kinase A (*PRKAR1A*).

McCune-Albright syndrome consists of polyostotic fibrous dysplasia, pigmented skin patches, and a variety of endocrine disorders, including acromegaly, adrenal adenomas, and autonomous ovarian function (Chap. 347). Hormonal hypersecretion results from constitutive cyclic AMP production caused by inactivation of the GTPase activity of Gsα. The Gsα mutations occur postzygotically, leading to a mosaic pattern of mutant expression.

Familial acromegaly is a rare disorder in which family members may manifest either acromegaly or gigantism. The disorder is associated with LOH at a chromosome 11q13 locus distinct from that of *MENIN*. A subset of families with a predisposition for familial pituitary tumors, especially acromegaly, have been found to harbor inactivating mutations in the AIP gene, which encodes the aryl hydrocarbon receptor interacting protein.

Other Sellar Masses

Craniopharyngiomas are benign, suprasellar cystic masses that present with headaches, visual field deficits, and variable degrees of hypopituitarism. They are derived from Rathke's pouch and arise near the pituitary stalk, commonly extending into the suprasellar cistern. Craniopharyngiomas are often large, cystic, and locally invasive. Many are partially calcified, exhibiting a characteristic appearance on skull x-ray and CT images. More than half of all patients present before age 20, usually with signs of increased intracranial pressure, including headache, vomiting, papilledema, and hydrocephalus. Associated symptoms include visual field abnormalities, personality changes and cognitive deterioration, cranial nerve damage, sleep difficulties, and weight gain. Hypopituitarism can be documented in about 90%, and diabetes insipidus occurs in about 10% of patients. About half of affected children present with growth retardation. MRI is generally superior to CT for evaluating cystic structure and tissue components of craniopharyngiomas. CT is useful to define calcifications and evaluate invasion into surrounding bony structures and sinuses.

Treatment usually involves transcranial or transsphenoidal surgical resection followed by postoperative radiation of residual tumor. Surgery alone is curative in less than half of patients because of recurrences due to adherence to vital structures or because of small tumor deposits in the hypothalamus or brain parenchyma. The goal of surgery is to remove as much tumor as possible without risking complications associated with efforts to remove firmly adherent or inaccessible tissue. In the absence of radiotherapy, about 75% of craniopharyngiomas recur, and 10-year survival is less than 50%. In patients with incomplete resection, radiotherapy improves 10-year survival to 70–90% but is associated with increased risk of secondary malignancies. Most patients require lifelong pituitary hormone replacement.

Developmental failure of Rathke's pouch obliteration may lead to *Rathke's cysts*, which are small (<5

mm) cysts entrapped by squamous epithelium and are found in about 20% of individuals at autopsy. Although Rathke's cleft cysts do not usually grow and are often diagnosed incidentally, about a third present in adulthood with compressive symptoms, diabetes insipidus, and hyperprolactinemia due to stalk compression. Rarely, hydrocephalus develops. The diagnosis is suggested preoperatively by visualizing the cyst wall on MRI, which distinguishes these lesions from craniopharyngiomas. Cyst contents range from CSF-like fluid to mucoid material. *Arachnoid cysts* are rare and generate an MRI image that is isointense with cerebrospinal fluid.

Sella chordomas usually present with bony clival erosion, local invasiveness, and, on occasion, calcification. Normal pituitary tissue may be visible on MRI, distinguishing chordomas from aggressive pituitary adenomas. Mucinous material may be obtained by fine-needle aspiration.

Meningiomas arising in the sellar region may be difficult to distinguish from nonfunctioning pituitary adenomas. Meningiomas typically enhance on MRI and may show evidence of calcification or bony erosion. Meningiomas may cause compressive symptoms.

Histiocytosis X includes a variety of syndromes associated with foci of eosinophilic granulomas. Diabetes insipidus, exophthalmos, and punched-out lytic bone lesions (*Hand-Schüller-Christian disease*) are associated with granulomatous lesions visible on MRI, as well as a characteristic axillary skin rash. Rarely, the pituitary stalk may be involved.

Pituitary metastases occur in ~3% of cancer patients. Bloodborne metastatic deposits are found almost exclusively in the posterior pituitary. Accordingly, diabetes insipidus can be a presenting feature of lung, gastrointestinal, breast, and other pituitary metastases. About half of pituitary metastases originate from breast cancer; about 25% of patients with metastatic breast cancer have such deposits. Rarely, pituitary stalk involvement results in anterior pituitary insufficiency. The MRI diagnosis of a metastatic lesion may be difficult to distinguish from an aggressive pituitary adenoma; the diagnosis may require histologic examination of excised tumor tissue. Primary or metastatic lymphoma, leukemias, and plasmacytomas also occur within the sella.

Hypothalamic hamartomas and *gangliocytomas* may arise from astrocytes, oligodendrocytes, and neurons with varying degrees of differentiation. These tumors may overexpress hypothalamic neuropeptides, including GnRH, GHRH, and CRH. With GnRH-producing tumors, children present with precocious puberty, psychomotor delay, and laughing-associated seizures. Medical treatment of GnRH-producing hamartomas with long-acting GnRH analogues effectively suppresses gonadotropin secretion and controls premature pubertal development. Rarely, hamartomas also are associated with craniofacial abnormalities; imperforate anus; cardiac, renal, and lung disorders; and pituitary failure as features of *Pallister-Hall syndrome*, which is caused by mutations in the carboxy terminus of the *GLI3* gene. Hypothalamic hamartomas are often contiguous with the pituitary, and preoperative MRI diagnosis may not be possible. Histologic evidence of hypothalamic neurons in tissue resected at transsphenoidal surgery may be the first indication of a primary hypothalamic lesion.

Hypothalamic gliomas and *optic gliomas* occur mainly in childhood and usually present with visual loss. Adults have more aggressive tumors; about a third are associated with neurofibromatosis.

Brain germ-cell tumors may arise within the sellar region. They include *dysgerminomas*, which frequently are associated with diabetes insipidus and visual loss. They rarely metastasize. *Germinomas, embryonal carcinomas, teratomas*, and *choriocarcinomas* may arise in the parasellar region and produce hCG. These germ-cell tumors present with precocious puberty, diabetes insipidus, visual field defects, and thirst disorders. Many patients are GH-deficient with short stature.

Metabolic Effects of Hypothalamic Lesions

Lesions involving the anterior and preoptic hypothalamic regions cause paradoxical vasoconstriction, tachycardia, and hyperthermia. Acute hyperthermia usually is due to a hemorrhagic insult, but poikilothermia may also occur. Central disorders of thermoregulation result from posterior hypothalamic damage. The *periodic hypothermia syndrome* is characterized by episodic attacks of rectal temperatures <30°C (86°F), sweating, vasodilation, vomiting, and bradycardia (Chap. 19). Damage to the ventromedial hypothalamic nuclei by craniopharyngiomas, hypothalamic trauma, or inflammatory disorders may be associated with *hyperphagia* and *obesity*. This region appears to contain an energy-satiety center where melanocortin receptors are influenced by leptin, insulin, POMC products, and gastrointestinal peptides (Chap. 77). Polydipsia and hypodipsia are associated with damage to central osmoreceptors located in preoptic nuclei (Chap. 340). Slow-growing hypothalamic lesions can cause increased somnolence and disturbed sleep cycles as well as obesity, hypothermia, and emotional outbursts. Lesions of the central hypothalamus may stimulate sympathetic neurons, leading to elevated serum catecholamine and cortisol levels. These patients are predisposed to cardiac arrhythmias, hypertension, and gastric erosions.

Evaluation

LOCAL MASS EFFECTS

Clinical manifestations of sellar lesions vary, depending on the anatomic location of the mass and the

direction of its extension (Table 339-7). The dorsal sellar diaphragm presents the least resistance to soft tissue expansion from the sella; consequently, pituitary adenomas frequently extend in a suprasellar direction. Bony invasion may occur as well.

Table 339-7 Features of Sellar Mass Lesions*

Impacted Structure	Clinical Impact
Pituitary	Hypogonadism Hypothyroidism Growth failure and adult hyposomatotropism Hypoadrenalism
Optic chiasm	Loss of red perception Bitemporal hemianopia Superior or bitemporal field defect Scotoma Blindness
Hypothalamus	Temperature dysregulation Appetite and thirst disorders Obesity Diabetes insipidus Sleep disorders Behavioral dysfunction Autonomic dysfunction
Cavernous sinus	Opthalmoplegia with or without ptosis or diplopia Facial numbness
Frontal lobe	Personality disorder Anosmia
Brain	Headache Hydrocephalus Psychosis Dementia Laughing seizures

*As the intrasellar mass expands, it first compresses intrasellar pituitary tissue, then usually invades dorsally through the dura to lift the optic chiasm or laterally to the cavernous sinuses. Bony erosion is rare, as is direct brain compression. Microadenomas may present with headache.

Headaches are common features of small intrasellar tumors, even with no demonstrable suprasellar extension. Because of the confined nature of the pituitary, small changes in intrasellar pressure stretch the dural plate; however, headache severity correlates poorly with adenoma size or extension.

Suprasellar extension can lead to visual loss by several mechanisms, the most common being compression of the optic chiasm, but rarely, direct invasion of the optic nerves or obstruction of CSF flow leading to secondary visual disturbances also occurs. Pituitary stalk compression by a hormonally active or inactive intrasellar mass may compress the portal vessels, disrupting pituitary access to hypothalamic hormones and dopamine; this results in early hyperprolactinemia and later concurrent loss of other pituitary hormones. This "stalk section" phenomenon may also be caused by trauma, whiplash injury with posterior clinoid stalk compression, or skull base fractures. Lateral mass invasion may impinge on the cavernous sinus and compress its neural contents, leading to cranial nerve III, IV, and VI palsies as well as effects on the ophthalmic and maxillary branches of the fifth cranial nerve (Chap. 376). Patients may present with diplopia, ptosis, ophthalmoplegia, and decreased facial sensation, depending on the extent of neural damage. Extension into the sphenoid sinus indicates that the pituitary mass has eroded through the sellar floor. Aggressive tumors rarely invade the palate roof and cause nasopharyngeal obstruction, infection, and CSF leakage. Temporal and frontal lobe involvement may rarely lead to uncinate seizures, personality disorders, and anosmia. Direct hypothalamic encroachment by an invasive pituitary mass may cause important metabolic sequelae, including precocious puberty or hypogonadism, diabetes insipidus, sleep disturbances, dysthermia,

and appetite disorders.

MRI

Sagittal and coronal T1-weighted MRI imaging before and after administration of gadolinium allows precise visualization of the pituitary gland with clear delineation of the hypothalamus, pituitary stalk, pituitary tissue and surrounding suprasellar cisterns, cavernous sinuses, sphenoid sinus, and optic chiasm. Pituitary gland height ranges from 6 mm in children to 8 mm in adults; during pregnancy and puberty, the height may reach 10–12 mm. The upper aspect of the adult pituitary is flat or slightly concave, but in adolescent and pregnant individuals, this surface may be convex, reflecting physiologic pituitary enlargement. The stalk should be midline and vertical. CT scan is reserved to define the extent of bony erosion or the presence of calcification.

Anterior pituitary gland soft tissue consistency is slightly heterogeneous on MRI, and signal intensity resembles that of brain matter on T1-weighted imaging (Fig. 339-4). Adenoma density is usually lower than that of surrounding normal tissue on T1-weighted imaging, and the signal intensity increases with T2-weighted images. The high phospholipid content of the posterior pituitary results in a "pituitary bright spot."

Figure 339-4

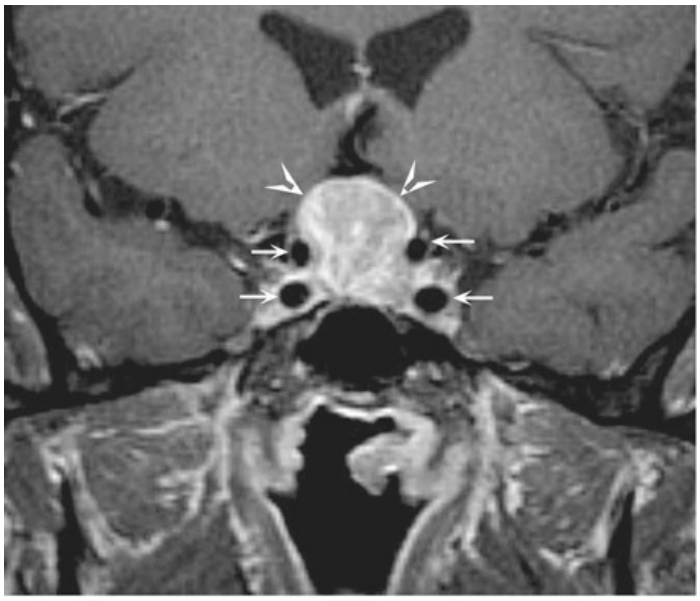

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pituitary adenoma. Coronal T1-weighted postcontrast MR image shows a homogeneously enhancing mass (*arrowheads*) in the sella turcica and suprasellar region compatible with a pituitary adenoma; the small arrows outline the carotid arteries.

Sellar masses are encountered commonly as incidental findings on MRI, and most of them are pituitary adenomas (incidentalomas). In the absence of hormone hypersecretion, these small intrasellar lesions can be monitored safely with MRI, which is performed annually and then less often if there is no evidence of further growth. Resection should be considered for incidentally discovered macroadenomas, as about one-third become invasive or cause local pressure effects. If hormone hypersecretion is evident, specific therapies are indicated. When larger masses (>1 cm) are encountered, they should also be distinguished from nonadenomatous lesions. Meningiomas often are associated with bony hyperostosis; craniopharyngiomas may be calcified and are usually hypodense, whereas gliomas are hyperdense on T2-weighted images.

OPHTHALMOLOGIC EVALUATION

Because optic tracts may be contiguous to an expanding pituitary mass, reproducible visual field assessment using perimetry techniques should be performed on all patients with sellar mass lesions that abut the optic chiasm (Chap. 28). Bitemporal hemianopia or superior bitemporal defects are classically observed, reflecting the location of these tracts within the inferior and posterior part of the chiasm. Homonymous cuts reflect postchiasmal lesions, and monocular field cuts prechiasmal lesions. Loss of red perception is an early sign of optic tract pressure. Early diagnosis reduces the risk of blindness, scotomas, or other visual disturbances.

LABORATORY INVESTIGATION

The presenting clinical features of functional pituitary adenomas (e.g., acromegaly, prolactinomas, or

Cushing's syndrome) should guide the laboratory studies (Table 339-8). However, for a sellar mass with no obvious clinical features of hormone excess, laboratory studies are geared toward determining the nature of the tumor and assessing the possible presence of hypopituitarism. When a pituitary adenoma is suspected based on MRI, initial hormonal evaluation usually includes (1) basal PRL; (2) insulin-like growth factor (IGF) I; (3) 24-h urinary free cortisol (UFC) and/or overnight oral dexamethasone (1 mg) suppression test; (4) α subunit, FSH, and LH; and (5) thyroid function tests. Additional hormonal evaluation may be indicated based on the results of these tests. Pending more detailed assessment of hypopituitarism, a menstrual history, measurement of testosterone and 8 A.M. cortisol levels, and thyroid function tests usually identify patients with pituitary hormone deficiencies that require hormone replacement before further testing or surgery.

Table 339-8 Screening Tests for Functional Pituitary Adenomas

	Test	Comments
Acromegaly	Serum IGF-I	Interpret IGF-I relative to age- and sex-matched controls
	Oral glucose tolerance test with GH obtained at 0, 30, and 60 min	Normal subjects should suppress growth hormone to <1 μg/L
Prolactinoma	Serum PRL	Exclude medications
		MRI of the sella should be ordered if prolactin is elevated
Cushing's disease	24-h urinary free cortisol	Ensure urine collection is total and accurate
	Dexamethasone (1 mg) at 11 P.M. and fasting plasma cortisol measured at 8 A.M.	Normal subjects suppress to <5 μg/dL
	ACTH assay	Distinguishes adrenal adenoma (ACTH suppressed) from ectopic ACTH or Cushing's disease (ACTH normal or elevated)

Note: For abbreviations, see text.

HISTOLOGIC EVALUATION

Immunohistochemical staining of pituitary tumor specimens obtained at transsphenoidal surgery confirms clinical and laboratory studies and provides a histologic diagnosis when hormone studies are equivocal and in cases of clinically nonfunctioning tumors. Occasionally, ultrastructural assessment by electron microscopy is required for diagnosis.

Treatment: Hypothalamic, Pituitary, and Other Sellar Masses

OVERVIEW

Successful management of sellar masses requires accurate diagnosis as well as selection of optimal therapeutic modalities. Most pituitary tumors are benign and slow-growing. Clinical features result from local mass effects and hormonal hypo- or hypersecretion syndromes caused directly by the adenoma or occurring as a consequence of treatment. Thus, lifelong management and follow-up are necessary for these patients.

MRI with gadolinium enhancement for pituitary visualization, new advances in transsphenoidal surgery and in stereotactic radiotherapy (including gamma-knife radiotherapy), and novel therapeutic agents have improved pituitary tumor management. The goals of pituitary tumor treatment include normalization of excess pituitary secretion, amelioration of symptoms and signs of hormonal hypersecretion syndromes, and shrinkage or ablation of large tumor masses with relief of adjacent structure compression. Residual anterior pituitary function should be preserved during treatment and sometimes can be restored by removing the tumor mass. Ideally, adenoma recurrence should be prevented.

TRANSSPHENOIDAL SURGERY

Transsphenoidal rather than transfrontal resection is the desired surgical approach for pituitary tumors, except for the rare invasive suprasellar mass surrounding the frontal or middle fossa or the optic nerves or invading posteriorly behind the clivus. Intraoperative microscopy facilitates visual distinction between adenomatous and normal pituitary tissue as well as microdissection of small tumors that may not be visible by MRI (Fig. 339-5). Transsphenoidal surgery also avoids the cranial invasion and manipulation of brain tissue required by subfrontal surgical approaches. Endoscopic techniques with three-dimensional intraoperative localization have also improved visualization and access to tumor tissue.

Figure 339-5

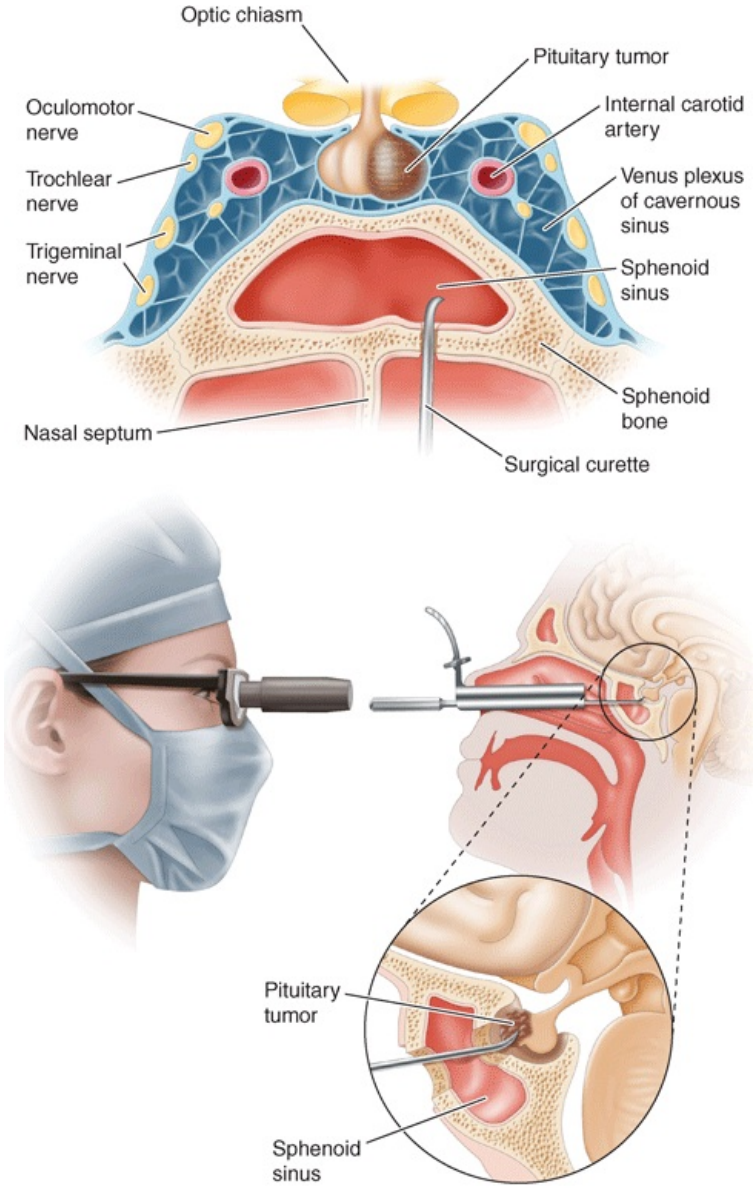

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transsphenoidal resection of pituitary mass via the endonasal approach. *(Adapted from R Fahlbusch: Endocrinol Metab Clin 21:669, 1992.)*

In addition to correction of hormonal hypersecretion, pituitary surgery is indicated for mass lesions that impinge on surrounding structures. Surgical decompression and resection are required for an expanding pituitary mass accompanied by persistent headache, progressive visual field defects, cranial nerve palsies, hydrocephalus, and, occasionally, intrapituitary hemorrhage and apoplexy. Transsphenoidal surgery sometimes is used for pituitary tissue biopsy to establish a histologic diagnosis.

Whenever possible, the pituitary mass lesion should be selectively excised; normal pituitary tissue should be manipulated or resected only when critical for effective mass dissection. Nonselective hemihypophysectomy or total hypophysectomy may be indicated if no hypersecreting mass lesion is clearly discernible, multifocal lesions are present, or the remaining nontumorous pituitary tissue is obviously necrotic. This strategy, however, increases the likelihood of hypopituitarism and the need for lifelong hormone replacement.

Preoperative mass effects, including visual field defects and compromised pituitary function, may be reversed by surgery, particularly when the deficits are not long-standing. For large and invasive tumors, it is necessary to determine the optimal balance between maximal tumor resection and preservation of anterior pituitary function, especially for preserving growth and reproductive function in younger patients. Similarly, tumor invasion outside the sella is rarely amenable to surgical cure; the surgeon must judge the risk-versus-benefit ratio of extensive tumor resection.

Side Effects

Tumor size, the degree of invasiveness, and experience of the surgeon largely determine the incidence of surgical complications. Operative mortality rate is about 1%. Transient diabetes insipidus and hypopituitarism occur in up to 20% of patients. Permanent diabetes insipidus, cranial nerve damage, nasal septal perforation, or visual disturbances may be encountered in up to 10% of patients. CSF leaks occur in 4% of patients. Less common complications include carotid artery injury, loss of vision, hypothalamic damage, and meningitis. Permanent side effects are rare after surgery for microadenomas.

RADIATION

Radiation is used either as a primary therapy for pituitary or parasellar masses or, more commonly, as an adjunct to surgery or medical therapy. Focused megavoltage irradiation is achieved by precise MRI localization, using a high-voltage linear accelerator and accurate isocentric rotational arcing. A major determinant of accurate irradiation is reproduction of the patient's head position during multiple visits and maintenance of absolute head immobility. A total of <50 Gy (5000 rad) is given as 180-cGy (180-rad) fractions divided over about 6 weeks. Stereotactic radiosurgery delivers a large single high-energy dose from a cobalt 60 source (gamma knife), linear accelerator, or cyclotron. Long-term effects of gamma-knife surgery are unclear but appear to be similar to those encountered with conventional radiation.

The role of radiation therapy in pituitary tumor management depends on multiple factors, including the nature of the tumor, the age of the patient, and the availability of surgical and radiation expertise. Because of its relatively slow onset of action, radiation therapy is usually reserved for postsurgical management. As an adjuvant to surgery, radiation is used to treat residual tumor and in an attempt to prevent regrowth. Irradiation offers the only means for potentially ablating significant postoperative residual nonfunctioning tumor tissue. In contrast, PRL- and GH-secreting tumor tissues are amenable to medical therapy.

Side Effects

In the short term, radiation may cause transient nausea and weakness. Alopecia and loss of taste and smell may be more long-lasting. Failure of pituitary hormone synthesis is common in patients who have undergone head and neck or pituitary-directed irradiation. More than 50% of patients develop loss of GH, ACTH, TSH, and/or gonadotropin secretion within 10 years, usually due to hypothalamic damage. Lifelong follow-up with testing of anterior pituitary hormone reserve is therefore required after radiation treatment. Optic nerve damage with impaired vision due to optic neuritis is reported in about 2% of patients who undergo pituitary irradiation. Cranial nerve damage is uncommon now that radiation doses are ≤2 Gy (200 rad) at any one treatment session and the maximum dose is <50 Gy (5000 rad). The use of stereotactic radiotherapy may reduce damage to adjacent structures. Radiotherapy for pituitary tumors has been associated with adverse mortality rates, mainly from cerebrovascular disease. The cumulative risk of developing a secondary tumor after conventional radiation is 1.3% after 10 years and 1.9% after 20 years.

MEDICAL

Medical therapy for pituitary tumors is highly specific and depends on tumor type. For prolactinomas, dopamine agonists are the treatment of choice. For acromegaly, somatostatin analogues and GH receptor antagonists are indicated. For TSH-secreting tumors, somatostatin analogues and occasionally dopamine agonists are indicated. ACTH-secreting tumors and nonfunctioning tumors are generally not responsive to medications and require surgery and/or irradiation.

PROLACTIN

Synthesis

PRL consists of 198 amino acids and has a molecular mass of 21,500 kDa; it is weakly homologous to GH and human placental lactogen (hPL), reflecting the duplication and divergence of a common GH-PRL-hPL precursor gene. PRL is synthesized in lactotropes, which constitute about 20% of anterior pituitary cells. Lactotropes and somatotropes are derived from a common precursor cell that may give rise to a tumor that secretes both PRL and GH. Marked lactotrope cell hyperplasia develops during pregnancy and the first few months of lactation. These transient functional changes in the lactotrope population are induced by estrogen.

Secretion

Normal adult serum PRL levels are about 10–25 μg/L in women and 10–20 μg/L in men. PRL secretion is pulsatile, with the highest secretory peaks occurring during rapid eye movement sleep. Peak serum PRL levels (up to 30 μg/L) occur between 4:00 and 6:00 a.m. The circulating half-life of PRL is about 50 min.

PRL is unique among the pituitary hormones in that the predominant central control mechanism is inhibitory, reflecting dopamine-mediated suppression of PRL release. This regulatory pathway accounts for the spontaneous PRL hypersecretion that occurs with pituitary stalk section, often a consequence of compressive mass lesions at the skull base. Pituitary dopamine type 2 (D_2) receptors

mediate inhibition of PRL synthesis and secretion. Targeted disruption (gene knockout) of the murine D_2 receptor in mice results in hyperprolactinemia and lactotrope proliferation. As discussed below, dopamine agonists play a central role in the management of hyperprolactinemic disorders.

Thyrotropin-releasing hormone (TRH) (pyro Glu-His-Pro-NH_2) is a hypothalamic tripeptide that elicits prolactin release within 15–30 min after intravenous injection. The physiologic relevance of TRH for PRL regulation is unclear, and it appears primarily to regulate TSH (Chap. 341). *Vasoactive intestinal peptide* (VIP) also induces PRL release, whereas glucocorticoids and thyroid hormone weakly suppress PRL secretion.

Serum PRL levels rise transiently after exercise, meals, sexual intercourse, minor surgical procedures, general anesthesia, chest wall injury, acute myocardial infarction, and other forms of acute stress. PRL levels increase markedly (about tenfold) during pregnancy and decline rapidly within 2 weeks of parturition. If breast-feeding is initiated, basal PRL levels remain elevated; suckling stimulates reflex increases in PRL levels that last for about 30–45 min. Breast suckling activates neural afferent pathways in the hypothalamus that induce PRL release. With time, suckling-induced responses diminish and interfeeding PRL levels return to normal.

Action

The PRL receptor is a member of the type I cytokine receptor family that also includes GH and interleukin (IL) 6 receptors. Ligand binding induces receptor dimerization and intracellular signaling by Janus kinase (JAK), which stimulates translocation of the signal transduction and activators of transcription (STAT) family to activate target genes. In the breast, the lobuloalveolar epithelium proliferates in response to PRL, placental lactogens, estrogen, progesterone, and local paracrine growth factors, including IGF-I.

PRL acts to induce and maintain lactation, decrease reproductive function, and suppress sexual drive. These functions are geared toward ensuring that maternal lactation is sustained and not interrupted by pregnancy. PRL inhibits reproductive function by suppressing hypothalamic GnRH and pituitary gonadotropin secretion and by impairing gonadal steroidogenesis in both women and men. In the ovary, PRL blocks folliculogenesis and inhibits granulosa cell aromatase activity, leading to hypoestrogenism and anovulation. PRL also has a luteolytic effect, generating a shortened, or inadequate, luteal phase of the menstrual cycle. In men, attenuated LH secretion leads to low testosterone levels and decreased spermatogenesis. These hormonal changes decrease libido and reduce fertility in patients with hyperprolactinemia.

Hyperprolactinemia

ETIOLOGY

Hyperprolactinemia is the most common pituitary hormone hypersecretion syndrome in both men and women. PRL-secreting pituitary adenomas (prolactinomas) are the most common cause of PRL levels >200 μg/L (see below). Less pronounced PRL elevation can also be seen with microprolactinomas but is more commonly caused by drugs, pituitary stalk compression, hypothyroidism, or renal failure (Table 339-9).

Table 339-9 Etiology of Hyperprolactinemia

I. Physiologic hypersecretion
Pregnancy
Lactation
Chest wall stimulation
Sleep
Stress
II. Hypothalamic–pituitary stalk damage
Tumors
Craniopharyngioma
Suprasellar pituitary mass
Meningioma
Dysgerminoma
Metastases
Empty sella
Lymphocytic hypophysitis
Adenoma with stalk compression

- Granulomas
- Rathke's cyst
- Irradiation
- Trauma
  - Pituitary stalk section
  - Suprasellar surgery

III. Pituitary hypersecretion

- Prolactinoma
- Acromegaly

IV. Systemic disorders

- Chronic renal failure
- Hypothyroidism
- Cirrhosis
- Pseudocyesis
- Epileptic seizures

V. Drug-induced hypersecretion

- Dopamine receptor blockers
  - Atypical antipsychotics: risperidone
  - Phenothiazines: chlorpromazine, perphenazine
  - Butyrophenones: haloperidol
  - Thioxanthenes
  - Metoclopramide
- Dopamine synthesis inhibitors
  - α-Methyldopa
- Catecholamine depletors
  - Reserpine
- Opiates
- H_2 antagonists
  - Cimetidine, ranitidine
- Imipramines
  - Amitriptyline, amoxapine
- Serotonin reuptake inhibitors
  - Fluoxetine
- Calcium channel blockers
  - Verapamil
  - Estrogens
  - TRH

Pregnancy and lactation are the important physiologic causes of hyperprolactinemia. Sleep-associated hyperprolactinemia reverts to normal within an hour of awakening. Nipple stimulation and sexual orgasm also may increase PRL. Chest wall stimulation or trauma (including chest surgery and herpes zoster) invoke the reflex suckling arc with resultant hyperprolactinemia. Chronic renal failure elevates PRL by decreasing peripheral clearance. Primary hypothyroidism is associated with mild hyperprolactinemia, probably because of compensatory TRH secretion.

Lesions of the hypothalamic-pituitary region that disrupt hypothalamic dopamine synthesis, portal vessel delivery, or lactotrope responses are associated with hyperprolactinemia. Thus, hypothalamic tumors, cysts, infiltrative disorders, and radiation-induced damage cause elevated PRL levels, usually in the range of 30–100 μg/L. Plurihormonal adenomas (including GH and ACTH tumors) may hypersecrete PRL directly. Pituitary masses, including clinically nonfunctioning pituitary tumors, may compress the pituitary stalk to cause hyperprolactinemia.

Drug-induced inhibition or disruption of dopaminergic receptor function is a common cause of

hyperprolactinemia (Table 339-9). Thus, antipsychotics and antidepressants are a relatively common cause of mild hyperprolactinemia. Most patients receiving risperidone have elevated prolactin levels, sometimes exceeding 200 ug/L. Methyldopa inhibits dopamine synthesis and verapamil blocks dopamine release, also leading to hyperprolactinemia. Hormonal agents that induce PRL include estrogens and TRH.

PRESENTATION AND DIAGNOSIS

Amenorrhea, galactorrhea, and infertility are the hallmarks of hyperprolactinemia in women. If hyperprolactinemia develops before menarche, primary amenorrhea results. More commonly, hyperprolactinemia develops later in life and leads to oligomenorrhea and ultimately to amenorrhea. If hyperprolactinemia is sustained, vertebral bone mineral density can be reduced compared with age-matched controls, particularly when it is associated with pronounced hypoestrogenemia. Galactorrhea is present in up to 80% of hyperprolactinemic women. Although usually bilateral and spontaneous, it may be unilateral or expressed only manually. Patients also may complain of decreased libido, weight gain, and mild hirsutism.

In men with hyperprolactinemia, diminished libido, infertility, and visual loss (from optic nerve compression) are the usual presenting symptoms. Gonadotropin suppression leads to reduced testosterone, impotence, and oligospermia. True galactorrhea is uncommon in men with hyperprolactinemia. If the disorder is long-standing, secondary effects of hypogonadism are evident, including osteopenia, reduced muscle mass, and decreased beard growth.

The diagnosis of idiopathichyperprolactinemia is made by exclusion of known causes of hyperprolactinemia in the setting of a normal pituitary MRI. Some of these patients may harbor small microadenomas below visible MRI sensitivity (~2 mm).

Galactorrhea

Galactorrhea, the inappropriate discharge of milk-containing fluid from the breast, is considered abnormal if it persists longer than 6 months after childbirth or discontinuation of breast-feeding. Postpartum galactorrhea associated with amenorrhea is a self-limiting disorder usually associated with moderately elevated PRL levels. Galactorrhea may occur spontaneously, or it may be elicited by nipple pressure. In both men and women, galactorrhea may vary in color and consistency (transparent, milky, or bloody) and arise either unilaterally or bilaterally. Mammography or ultrasound is indicated for bloody discharges (particularly from a single nipple), which may be caused by breast cancer. Galactorrhea is commonly associated with hyperprolactinemia caused by any of the conditions listed in Table 339-9. Acromegaly is associated with galactorrhea in about one-third of patients. Treatment of galactorrhea usually involves managing the underlying disorder (e.g., replacing T4 for hypothyroidism, discontinuing a medication, treating prolactinoma).

LABORATORY INVESTIGATION

Basal, fasting morning PRL levels (normally <20 μg/L) should be measured to assess hypersecretion. Both false-positive and false-negative results may be encountered. In patients with markedly elevated PRL levels (>1000 μg/L), reported results may be falsely lowered because of assay artifacts; sample dilution is required to measure these high values accurately. Falsely elevated values may be caused by aggregated forms of circulating PRL, which are usually biologically inactive (macroprolactinemia). Hypothyroidism should be excluded by measuring TSH and T4 levels.

TREATMENT: HYPERPROLACTINEMIA

Treatment of hyperprolactinemia depends on the cause of elevated PRL levels. Regardless of the etiology, however, treatment should be aimed at normalizing PRL levels to alleviate suppressive effects on gonadal function, halt galactorrhea, and preserve bone mineral density. Dopamine agonists are effective for most causes of hyperprolactinemia (see the treatment section for prolactinoma, below) regardless of the underlying cause.

If the patient is taking a medication known to cause hyperprolactinemia, the drug should be withdrawn, if possible. For psychiatric patients who require neuroleptic agents, supervised dose titration or the addition of a dopamine agonist can help restore normoprolactinemia and alleviate reproductive symptoms. However, dopamine agonists sometimes worsen the underlying psychiatric condition, especially at high doses. Hyperprolactinemia usually resolves after adequate thyroid hormone replacement in hypothyroid patients or after renal transplantation in patients undergoing dialysis. Resection of hypothalamic or sellar mass lesions can reverse hyperprolactinemia caused by stalk compression and reduced dopamine tone. Granulomatous infiltrates occasionally respond to glucocorticoid administration. In patients with irreversible hypothalamic damage, no treatment may be warranted. In up to 30% of patients with hyperprolactinemia—usually without a visible pituitary microadenoma—the condition may resolve spontaneously.

Prolactinoma

ETIOLOGY AND PREVALENCE

Tumors arising from lactotrope cells account for about half of all functioning pituitary tumors, with a

population prevalence of ~10/100,000 in men and ~30/100,000 in women. Mixed tumors that secrete combinations of GH and PRL, ACTH and PRL, and rarely TSH and PRL are also seen. These plurihormonal tumors are usually recognized by immunohistochemistry, sometimes without apparent clinical manifestations from the production of additional hormones. Microadenomas are classified as <1 cm in diameter and usually do not invade the parasellar region. Macroadenomas are >1 cm in diameter and may be locally invasive and impinge on adjacent structures. The female:male ratio for microprolactinomas is 20:1, whereas the sex ratio is near 1:1 for macroadenomas. Tumor size generally correlates directly with PRL concentrations; values >250 μg/L usually are associated with macroadenomas. Men tend to present with larger tumors than women, possibly because the features of male hypogonadism are less readily evident. PRL levels remain stable in most patients, reflecting the slow growth of these tumors. About 5% of microadenomas progress in the long term to macroadenomas.

PRESENTATION AND DIAGNOSIS

Women usually present with amenorrhea, infertility, and galactorrhea. If the tumor extends outside the sella, visual field defects or other mass effects may be seen. Men often present with impotence, loss of libido, infertility, or signs of central CNS compression, including headaches and visual defects. Assuming that physiologic and medication-induced causes of hyperprolactinemia are excluded (Table 339-9), the diagnosis of prolactinoma is likely with a PRL level >200 μg/L. PRL levels <100 μg/L may be caused by microadenomas, other sellar lesions that decrease dopamine inhibition, or nonneoplastic causes of hyperprolactinemia. For this reason, an MRI should be performed in all patients with hyperprolactinemia. It is important to remember that hyperprolactinemia caused secondarily by the mass effects of nonlactotrope lesions is also corrected by treatment with dopamine agonists despite failure to shrink the underlying mass. Consequently, PRL suppression by dopamine agonists does not necessarily indicate that the underlying lesion is a prolactinoma.

TREATMENT: PROLACTINOMA

As microadenomas rarely progress to become macroadenomas, no treatment may be needed if fertility is not desired. Estrogen replacement is indicated to prevent bone loss and other consequences of hypoestrogenemia and does not appear to increase the risk of tumor enlargement; these patients should be monitored by regular serial PRL and MRI measurements.

For symptomatic microadenomas, therapeutic goals include control of hyperprolactinemia, reduction of tumor size, restoration of menses and fertility, and resolution of galactorrhea. Dopamine agonist doses should be titrated to achieve maximal PRL suppression and restoration of reproductive function (Fig. 339-6). A normalized PRL level does not ensure reduced tumor size. However, tumor shrinkage usually is not seen in those who do not respond with lowered PRL levels. For macroadenomas, formal visual field testing should be performed before initiating dopamine agonists. MRI and visual fields should be assessed at 6- to 12-month intervals until the mass shrinks and annually thereafter until maximum size reduction has occurred.

Figure 339-6

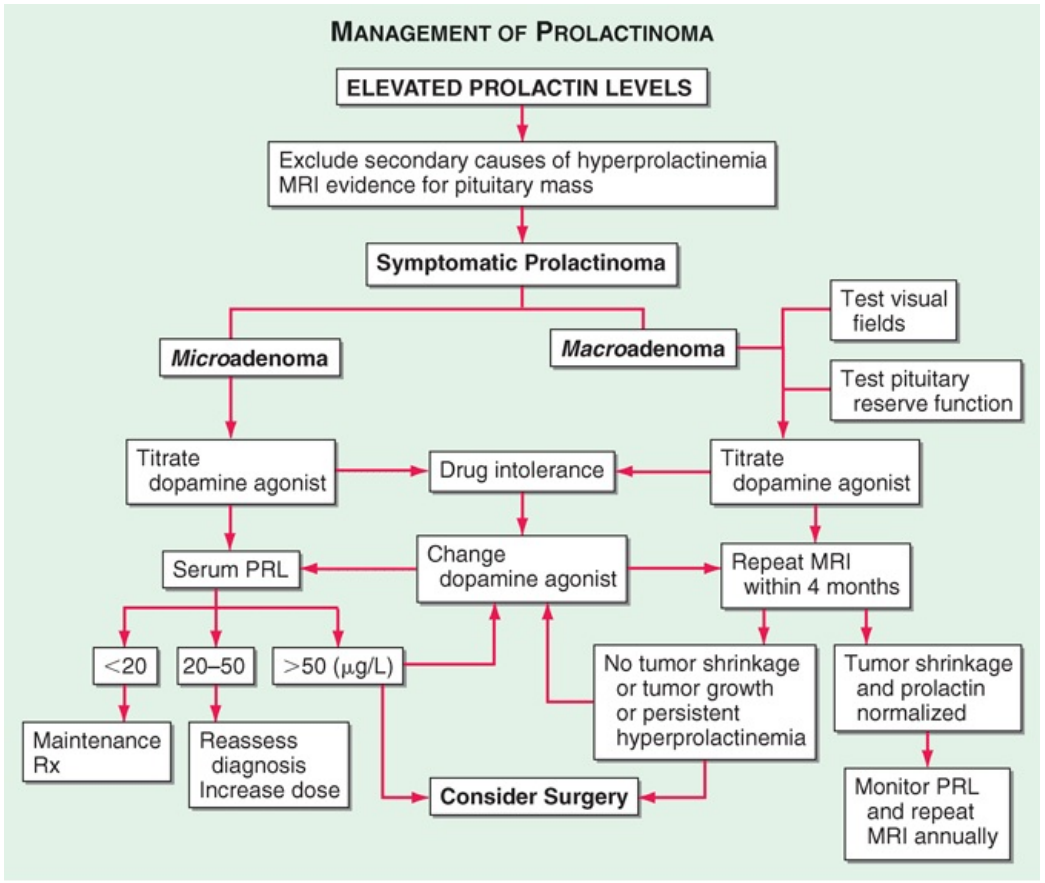

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of prolactinoma. MRI, magnetic resonance imaging; PRL, prolactin.

Medical

Oral dopamine agonists (cabergoline and bromocriptine) are the mainstay of therapy for patients with micro- or macroprolactinomas. Dopamine agonists suppress PRL secretion and synthesis as well as lactotrope cell proliferation. In patients with microadenomas who have achieved normoprolactinemia and significant reduction of tumor mass, the dopamine agonist may be withdrawn after 2 years. These patients should be monitored carefully for evidence of prolactinoma recurrence. About 20% of patients (especially males) are resistant to dopaminergic treatment; these adenomas may exhibit decreased D_2 dopamine receptor numbers or a postreceptor defect. D_2 receptor gene mutations in the pituitary have not been reported.

Cabergoline

An ergoline derivative, cabergoline is a long-acting dopamine agonist with high D_2 receptor affinity. The drug effectively suppresses PRL for >14 days after a single oral dose and induces prolactinoma shrinkage in most patients. Cabergoline (0.5 to 1.0 mg twice weekly) achieves normoprolactinemia and resumption of normal gonadal function in ~80% of patients with microadenomas; galactorrhea improves or resolves in 90% of patients. Cabergoline normalizes PRL and shrinks ~70% of macroprolactinomas. Mass effect symptoms, including headaches and visual disorders, usually improve dramatically within days after cabergoline initiation; improvement of sexual function requires several weeks of treatment but may occur before complete normalization of prolactin levels. After initial control of PRL levels has been achieved, cabergoline should be reduced to the lowest effective maintenance dose. In ~5% of treated patients harboring a microadenoma, hyperprolactinemia may resolve and not recur when dopamine agonists are discontinued after long-term treatment. Cabergoline also may be effective in patients resistant to bromocriptine. Adverse effects and drug intolerance are encountered less commonly than with bromocriptine.

Bromocriptine

The ergot alkaloid bromocriptine mesylate is a dopamine receptor agonist that suppresses prolactin secretion. Because it is short-acting, the drug is preferred when pregnancy is desired. In microadenomas bromocriptine rapidly lowers serum prolactin levels to normal in up to 70% of patients, decreases tumor size, and restores gonadal function. In patients with macroadenomas, prolactin levels are also normalized in 70% of patients and tumor mass shrinkage (≥50%) is achieved in most patients.

Therapy is initiated by administering a low bromocriptine dose (0.625–1.25 mg) at bedtime with a snack, followed by gradually increasing the dose. Most patients are controlled with a daily dose of ≤7.5 mg (2.5 mg tid).

Side Effects

Side effects of dopamine agonists include constipation, nasal stuffiness, dry mouth, nightmares, insomnia, and vertigo; decreasing the dose usually alleviates these problems. Nausea, vomiting, and postural hypotension with faintness may occur in ~25% of patients after the initial dose. These symptoms may persist in some patients. In general, fewer side effects are reported with cabergoline. For the approximately 15% of patients who are intolerant of oral bromocriptine, cabergoline may be better tolerated. Intravaginal administration of bromocriptine is often efficacious in patients with intractable gastrointestinal side effects. Auditory hallucinations, delusions, and mood swings have been reported in up to 5% of patients and may be due to the dopamine agonist properties or to the lysergic acid derivative of the compounds. Rare reports of leukopenia, thrombocytopenia, pleural fibrosis, cardiac arrhythmias, and hepatitis have been described. Patients with Parkinson's disease who receive at least 3 mg of cabergoline daily have been reported to be at risk for development of cardiac valve regurgitation. Studies analyzing over 500 prolactinoma patients receiving recommended doses of cabergoline (up to 2 mg weekly) have shown no evidence for an increased incidence of valvular disorders. Nevertheless, as no controlled prospective studies are available, it is prudent to perform echocardiograms before initiating standard-dose cabergoline therapy.

Surgery

Indications for surgical adenoma debulking include dopamine resistance or intolerance and the presence of an invasive macroadenoma with compromised vision that fails to improve after drug treatment. Initial PRL normalization is achieved in about 70% of microprolactinomas after surgical resection, but only 30% of macroadenomas can be resected successfully. Follow-up studies have shown that hyperprolactinemia recurs in up to 20% of patients within the first year after surgery; long-term recurrence rates exceed 50% for macroadenomas. Radiotherapy for prolactinomas is reserved for patients with aggressive tumors that do not respond to maximally tolerated dopamine agonists and/or surgery.

Pregnancy

The pituitary increases in size during pregnancy, reflecting the stimulatory effects of estrogen and perhaps other growth factors on pituitary vascularity and lactotrope cell hyperplasia. About 5% of microadenomas significantly increase in size, but 15–30% of macroadenomas grow during pregnancy. Bromocriptine has been used for more than 30 years to restore fertility in women with hyperprolactinemia, without evidence of teratogenic effects. Nonetheless, most authorities recommend strategies to minimize fetal exposure to the drug. For women taking bromocriptine who desire pregnancy, mechanical contraception should be used through three regular menstrual cycles to allow for conception timing. When pregnancy is confirmed, bromocriptine should be discontinued and PRL levels followed serially, especially if headaches or visual symptoms occur. For women harboring macroadenomas, regular visual field testing is recommended, and the drug should be reinstituted if tumor growth is apparent. Although pituitary MRI may be safe during pregnancy, this procedure should be reserved for symptomatic patients with severe headache and/or visual field defects. Surgical decompression may be indicated if vision is threatened. Although comprehensive data support the efficacy and relative safety of bromocriptine-facilitated fertility, patients should be advised of potential unknown deleterious effects and the risk of tumor growth during pregnancy. As cabergoline is long-acting with a high D_2-receptor affinity, it is not recommended for use in women when fertility is desired.

GROWTH HORMONE

Synthesis

GH is the most abundant anterior pituitary hormone, and GH-secreting somatotrope cells constitute up to 50% of the total anterior pituitary cell population. Mammosomatotrope cells, which coexpress PRL with GH, can be identified by using double immunostaining techniques. Somatotrope development and GH transcription are determined by expression of the cell-specific Pit-1 nuclear transcription factor. Five distinct genes encode GH and related proteins. The pituitary GH gene (*hGH-N*) produces two alternatively spliced products that give rise to 22-kDa GH (191 amino acids) and a less abundant 20-kDa GH molecule with similar biologic activity. Placental syncytiotrophoblast cells express a GH variant (*hGH-V*) gene; the related hormone human chorionic somatotropin (HCS) is expressed by distinct members of the gene cluster.

Secretion

GH secretion is controlled by complex hypothalamic and peripheral factors. *GHRH* is a 44-amino-acid hypothalamic peptide that stimulates GH synthesis and release. Ghrelin, an octanoylated gastric-derived peptide, and synthetic agonists of the *GHS-R* induce GHRH and also directly stimulate GH release. *Somatostatin* [somatotropin-release inhibiting factor (SRIF)] is synthesized in the medial preoptic area of the hypothalamus and inhibits GH secretion. GHRH is secreted in discrete spikes that

elicit GH pulses, whereas SRIF sets basal GH secretory tone. SRIF also is expressed in many extrahypothalamic tissues, including the CNS, gastrointestinal tract, and pancreas, where it also acts to inhibit islet hormone secretion. *IGF-I*, the peripheral target hormone for GH, feeds back to inhibit GH; estrogen induces GH, whereas chronic glucocorticoid excess suppresses GH release.

Surface receptors on the somatotrope regulate GH synthesis and secretion. The GHRH receptor is a G protein–coupled receptor (GPCR) that signals through the intracellular cyclic AMP pathway to stimulate somatotrope cell proliferation as well as GH production. Inactivating mutations of the GHRH receptor cause profound dwarfism (see below). A distinct surface receptor for ghrelin, the gastric-derived GH secretagogue, is expressed in the hypothalamus and pituitary. Somatostatin binds to five distinct receptor subtypes (SSTR1 to SSTR5); SSTR2 and SSTR5 subtypes preferentially suppress GH (and TSH) secretion.

GH secretion is pulsatile, with highest peak levels occurring at night, generally correlating with sleep onset. GH secretory rates decline markedly with age so that hormone levels in middle age are about 15% of pubertal levels. These changes are paralleled by an age-related decline in lean muscle mass. GH secretion is also reduced in obese individuals, though IGF-I levels may not be suppressed, suggesting a change in the setpoint for feedback control. Elevated GH levels occur within an hour of deep sleep onset as well as after exercise, physical stress, and trauma and during sepsis. Integrated 24-h GH secretion is higher in women and is also enhanced by estrogen replacement. Using standard assays, random GH measurements are undetectable in ~50% of daytime samples obtained from healthy subjects and are also undetectable in most obese and elderly subjects. Thus, single random GH measurements do not distinguish patients with adult GH deficiency from normal persons.

GH secretion is profoundly influenced by nutritional factors. Using newer ultrasensitive GH assays with a sensitivity of 0.002 μg/L, a glucose load suppresses GH to <0.7 μg/L in women and to <0.07 μg/L in men. Increased GH pulse frequency and peak amplitudes occur with chronic malnutrition or prolonged fasting. GH is stimulated by intravenous l-arginine, dopamine, and apomorphine (a dopamine receptor agonist), as well as by α-adrenergic pathways. β-Adrenergic blockage induces basal GH and enhances GHRH- and insulin-evoked GH release.

Action

The pattern of GH secretion may affect tissue responses. The higher GH pulsatility observed in men compared with the relatively continuous GH secretion in women may be an important biologic determinant of linear growth patterns and liver enzyme induction.

The 70-kDa peripheral GH receptor protein has structural homology with the cytokine/hematopoietic superfamily. A fragment of the receptor extracellular domain generates a soluble GH binding protein (GHBP) that interacts with GH in the circulation. The liver and cartilage contain the greatest number of GH receptors. GH binding to preformed receptor dimers is followed by internal rotation and subsequent signaling through the JAK/STAT pathway. Activated STAT proteins translocate to the nucleus, where they modulate expression of GH-regulated target genes. GH analogues that bind to the receptor but are incapable of mediating receptor signaling are potent antagonists of GH action. A GH receptor antagonist (pegvisomant) is approved for treatment of acromegaly.

GH induces protein synthesis and nitrogen retention and impairs glucose tolerance by antagonizing insulin action. GH also stimulates lipolysis, leading to increased circulating fatty acid levels, reduced omental fat mass, and enhanced lean body mass. GH promotes sodium, potassium, and water retention and elevates serum levels of inorganic phosphate. Linear bone growth occurs as a result of complex hormonal and growth factor actions, including those of IGF-I. GH stimulates epiphyseal prechondrocyte differentiation. These precursor cells produce IGF-I locally, and their proliferation is also responsive to the growth factor.

Insulin-Like Growth Factors

Although GH exerts direct effects in target tissues, many of its physiologic effects are mediated indirectly through IGF-I, a potent growth and differentiation factor. The liver is the major source of circulating IGF-I. In peripheral tissues, IGF-I exerts local paracrine actions that appear to be both dependent on and independent of GH. Thus, GH administration induces circulating IGF-I as well as stimulating local IGF-I production in multiple tissues.

Both IGF-I and IGF-II are bound to high-affinity circulating IGF-binding proteins (IGFBPs) that regulate IGF bioactivity. Levels of IGFBP3 are GH-dependent, and it serves as the major carrier protein for circulating IGF-I. GH deficiency and malnutrition usually are associated with low IGFBP3 levels. IGFBP1 and IGFBP2 regulate local tissue IGF action but do not bind appreciable amounts of circulating IGF-I.

Serum IGF-I concentrations are profoundly affected by physiologic factors. Levels increase during puberty, peak at 16 years, and subsequently decline by >80% during the aging process. IGF-I concentrations are higher in women than in men. Because GH is the major determinant of hepatic IGF-I synthesis, abnormalities of GH synthesis or action (e.g., pituitary failure, GHRH receptor defect, GH receptor defect) reduce IGF-I levels. Hypocaloric states are associated with GH resistance; IGF-I levels

are therefore low with cachexia, malnutrition, and sepsis. In acromegaly, IGF-I levels are invariably high and reflect a log-linear relationship with GH concentrations.

IGF-I PHYSIOLOGY

IGF-I has been approved for use in patients with GH-resistance syndromes. Injected IGF-I (100 μg/kg) induces hypoglycemia, and lower doses improve insulin sensitivity in patients with severe insulin resistance and diabetes. In cachectic subjects, IGF-I infusion (12 μg/kg per hour) enhances nitrogen retention and lowers cholesterol levels. Longer-term subcutaneous IGF-I injections enhance protein synthesis and are anabolic. Although bone formation markers are induced, bone turnover also may be stimulated by IGF-I.

IGF-I side effects are dose-dependent, and overdose may result in hypoglycemia, hypotension, fluid retention, temporomandibular jaw pain, and increased intracranial pressure, all of which are reversible. Avascular femoral head necrosis has been reported. Chronic excess IGF-I administration presumably would result in features of acromegaly.

Disorders of Growth and Development

SKELETAL MATURATION AND SOMATIC GROWTH

The growth plate is dependent on a variety of hormonal stimuli, including GH, IGF-I, sex steroids, thyroid hormones, paracrine growth factors, and cytokines. The growth-promoting process also requires caloric energy, amino acids, vitamins, and trace metals and consumes about 10% of normal energy production. Malnutrition impairs chondrocyte activity and reduces circulating IGF-I and IGFBP3 levels.

Linear bone growth rates are very high in infancy and are pituitary-dependent. Mean growth velocity is ~6 cm/year in later childhood and usually is maintained within a given range on a standardized percentile chart. Peak growth rates occur during midpuberty when bone age is 12 (girls) or 13 (boys). Secondary sexual development is associated with elevated sex steroids that cause progressive epiphyseal growth plate closure. *Bone age* is delayed in patients with all forms of true GH deficiency or GH receptor defects that result in attenuated GH action.

Short stature may occur as a result of constitutive intrinsic growth defects or because of acquired extrinsic factors that impair growth. In general, delayed bone age in a child with short stature is suggestive of a hormonal or systemic disorder, whereas normal bone age in a short child is more likely to be caused by a genetic cartilage dysplasia or growth plate disorder (Chap. 363).

GH DEFICIENCY IN CHILDREN

GH Deficiency

Isolated GH deficiency is characterized by short stature, micropenis, increased fat, high-pitched voice, and a propensity to hypoglycemia due to relatively unopposed insulin action. Familial modes of inheritance are seen in one-third of these individuals and may be autosomal dominant, recessive, or X-linked. About 10% of children with GH deficiency have mutations in the *GH-N* gene, including gene deletions and a wide range of point mutations. Mutations in transcription factors Pit-1 and Prop-1, which control somatotrope development result in GH deficiency in combination with other pituitary hormone deficiencies, which may become manifest only in adulthood. The diagnosis of *idiopathic GH deficiency* (IGHD) should be made only after known molecular defects have been rigorously excluded.

GHRH Receptor Mutations

Recessive mutations of the GHRH receptor gene in subjects with severe proportionate dwarfism are associated with low basal GH levels that cannot be stimulated by exogenous GHRH, GHRP, or insulin-induced hypoglycemia, as well as anterior pituitary hypoplasia The syndrome exemplifies the importance of the GHRH receptor for somatotrope cell proliferation and hormonal responsiveness.

Growth Hormone Insensitivity

This is caused by defects of GH receptor structure or signaling. Homozygous or heterozygous mutations of the GH receptor are associated with partial or complete GH insensitivity and growth failure (*Laron syndrome*). The diagnosis is based on normal or high GH levels, with decreased circulating GHBP, and low IGF-I levels. Very rarely, defective IGF-I, IGF-I receptor, or IGF-I signaling defects are also encountered. *STAT5B* mutations result in immunodeficiency with abrogated GH signaling, leading to short stature with normal or elevated GH levels and low IGF-I levels.

Nutritional Short Stature

Caloric deprivation and malnutrition, uncontrolled diabetes, and chronic renal failure represent secondary causes of abrogated GH receptor function. These conditions also stimulate production of proinflammatory cytokines, which act to exacerbate the block of GH-mediated signal transduction. Children with these conditions typically exhibit features of acquired short stature with normal or elevated GH, and low IGF-I levels. Circulating GH receptor antibodies may rarely cause peripheral GH insensitivity.

Psychosocial Short Stature

Emotional and social deprivation lead to growth retardation accompanied by delayed speech, discordant hyperphagia, and an attenuated response to administered GH. A nurturing environment restores growth rates.

PRESENTATION AND DIAGNOSIS

Short stature is commonly encountered in clinical practice, and the decision to evaluate these children requires clinical judgment in association with auxologic data and family history. Short stature should be evaluated comprehensively if a patient's height is >3 standard deviations (SD) below the mean for age or if the growth rate has decelerated. Skeletal maturation is best evaluated by measuring a radiologic bone age, which is based mainly on the degree of wrist bone growth plate fusion. Final height can be predicted using standardized scales (Bayley-Pinneau or Tanner-Whitehouse) or estimated by adding 6.5 cm (boys) or subtracting 6.5 cm (girls) from the midparental height.

LABORATORY INVESTIGATION

Because GH secretion is pulsatile, GH deficiency is best assessed by examining the response to provocative stimuli, including exercise, insulin-induced hypoglycemia, and other pharmacologic tests that normally increase GH to >7 μg/L in children. Random GH measurements do not distinguish normal children from those with true GH deficiency. Adequate adrenal and thyroid hormone replacement should be assured before testing. Age- and sex-matched IGF-I levels are not sufficiently sensitive or specific to make the diagnosis but can be useful to confirm GH deficiency. Pituitary MRI may reveal pituitary mass lesions or structural defects. Molecular analyses for known mutations should be undertaken when the cause of short stature remains cryptic, or when additional clinical features suggest a gentic cause.

TREATMENT: DISORDERS OF GROWTH AND DEVELOPMENT

Replacement therapy with recombinant GH (0.02–0.05 mg/kg per day subcutaneously) restores growth velocity in GH-deficient children to ~10 cm/year. If pituitary insufficiency is documented, other associated hormone deficits should be corrected—especially adrenal steroids. GH treatment is also moderately effective for accelerating growth rates in children with Turner syndrome and chronic renal failure.

In patients with GH insensitivity and growth retardation due to mutations of the GH receptor, treatment with IGF-I bypasses the dysfunctional GH receptor.

Adult GH Deficiency (Aghd)

This disorder usually is caused by hypothalamic or pituitary somatotrope damage. Acquired pituitary hormone deficiency follows a typical pattern in which loss of adequate GH reserve foreshadows subsequent hormone deficits. The sequential order of hormone loss is usually GH → FSH/LH → TSH → ACTH.

PRESENTATION AND DIAGNOSIS

The clinical features of AGHD include changes in body composition, lipid metabolism, and quality of life and cardiovascular dysfunction (Table 339-10). Body composition changes are common and include reduced lean body mass, increased fat mass with selective deposition of intraabdominal visceral fat, and increased waist-to-hip ratio. Hyperlipidemia, left ventricular dysfunction, hypertension, and increased plasma fibrinogen levels also may be present. Bone mineral content is reduced, with resultant increased fracture rates. Patients may experience social isolation, depression, and difficulty maintaining gainful employment. Adult hypopituitarism is associated with a threefold increase in cardiovascular mortality rates in comparison to age- and sex-matched controls, and this may be due to GH deficiency, as patients in these studies were replaced with other deficient pituitary hormones.

Table 339-10 Features of Adult Growth Hormone Deficiency

Clinical
Impaired quality of life
Decreased energy and drive
Poor concentration
Low self-esteem
Social isolation
Body composition changes
Increased body fat mass
Central fat deposition
Increased waist-hip ratio

Decreased lean body mass
Reduced exercise capacity
Reduced maximum O_2 uptake
Impaired cardiac function
Reduced muscle mass
Cardiovascular risk factors
Impaired cardiac structure and function
Abnormal lipid profile
Decreased fibrinolytic activity
Atherosclerosis
Omental obesity
Imaging
Pituitary: mass or structural damage
Bone: reduced bone mineral density
Abdomen: excess omental adiposity
Laboratory
Evoked GH <3 ng/mL
IGF-I and IGFBP3 low or normal
Increased LDL cholesterol
Concomitant gonadotropin,
TSH, and/or ACTH reserve deficits may be present

Abbreviation: LDL, low-density lipoprotein. For other abbreviations, see text.

LABORATORY INVESTIGATION

AGHD is rare, and in light of the nonspecific nature of associated clinical symptoms, patients appropriate for testing should be selected carefully on the basis of well-defined criteria. With few exceptions, testing should be restricted to patients with the following predisposing factors: (1) pituitary surgery, (2) pituitary or hypothalamic tumor or granulomas, (3) history of cranial irradiation, (4) radiologic evidence of a pituitary lesion, (5) childhood requirement for GH replacement therapy, and rarely (6) unexplained low age- and sex-matched IGF-I levels. The transition of a GH-deficient adolescent to adulthood requires retesting to document subsequent adult GH deficiency. Up to 20% of patients previously treated for childhood-onset GH deficiency are found to be GH-sufficient on repeat testing as adults.

A significant proportion (~25%) of truly GH-deficient adults have low-normal IGF-I levels. Thus, as in the evaluation of GH deficiency in children, valid age- and sex-matched IGF-I measurements provide a useful index of therapeutic responses but are not sufficiently sensitive for diagnostic purposes. The most validated test to distinguish pituitary-sufficient patients from those with AGHD is insulin-induced (0.05–0.1 U/kg) hypoglycemia. After glucose reduction to ~40 mg/dL, most individuals experience neuroglycopenic symptoms (Chap. 345), and peak GH release occurs at 60 min and remains elevated for up to 2 h. About 90% of healthy adults exhibit GH responses >5 μg/L; AGHD is defined by a peak GH response to hypoglycemia of <3 μg/L. Although insulin-induced hypoglycemia is safe when performed under appropriate supervision, it is contraindicated in patients with diabetes, ischemic heart disease, cerebrovascular disease, or epilepsy and in elderly patients. Alternative stimulatory tests include intravenous arginine (30 g), GHRH (1 μg/kg), GHRP-6 (90 μg) and in glucagon (1 mg). Combinations of these tests may evoke GH secretion in subjects who are not responsive to a single test.

TREATMENT: ADULT GH DEFICIENCY

Once the diagnosis of AGHD is unequivocally established, replacement of GH may be indicated. Contraindications to therapy include the presence of an active neoplasm, intracranial hypertension, and uncontrolled diabetes and retinopathy. The starting dose of 0.1–0.2 mg/d should be titrated (up to a maximum of 1.25 mg/d) to maintain IGF-I levels in the mid-normal range for age- and sex-matched controls (Fig. 339-7). Women require higher doses than men, and elderly patients require less GH. Long-term GH maintenance sustains normal IGF-I levels and is associated with persistent body composition changes (e.g., enhanced lean body mass and lower body fat). High-density lipoprotein cholesterol increases, but total cholesterol and insulin levels do not change significantly. Lumbar spine bone mineral density increases, but this response is gradual (>1 year). Many patients note significant

improvement in quality of life when evaluated by standardized questionnaires. The effect of GH replacement on mortality rates in GH-deficient patients is currently the subject of long-term prospective investigation.

Figure 339-7

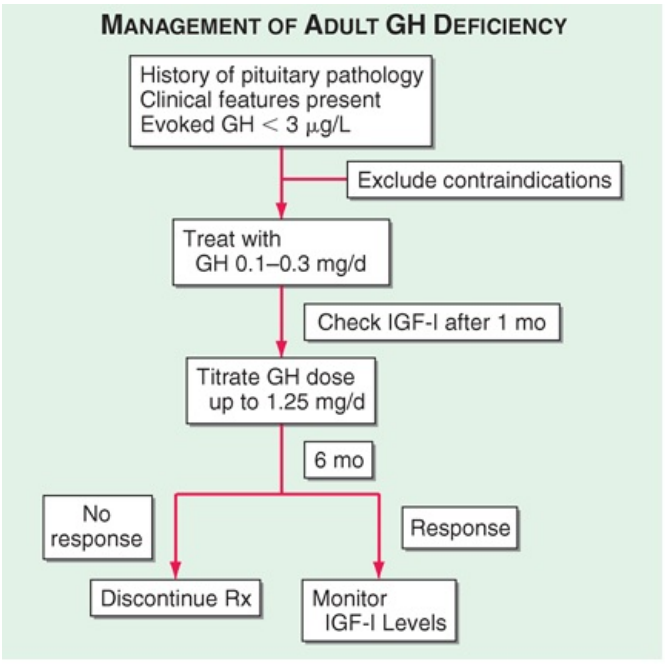

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of adult growth hormone (GH) deficiency. IGF, insulin-like growth factor.

About 30% of patients exhibit reversible dose-related fluid retention, joint pain, and carpal tunnel syndrome, and up to 40% exhibit myalgias and paresthesia. Patients receiving insulin require careful monitoring for dosing adjustments, as GH is a potent counterregulatory hormone for insulin action. Patients with Type 2 diabetes mellitus initially develop further insulin resistance. However, glycemic control improves with the sustained loss of abdominal fat associated with long-term GH replacement. Headache, increased intracranial pressure, hypertension, and tinnitus occur rarely. Pituitary tumor regrowth and progression of skin lesions or other tumors are being assessed in long-term surveillance programs. To date, development of these potential side effects does not appear significant.

Acromegaly

ETIOLOGY

GH hypersecretion is usually the result of a somatotrope adenoma but may rarely be caused by extrapituitary lesions (Table 339-11). In addition to more common GH-secreting somatotrope adenomas, mixed mammosomatotrope tumors and acidophilic stem-cell adenomas secrete both GH and PRL. In patients with acidophilic stem-cell adenomas, features of hyperprolactinemia (hypogonadism and galactorrhea) predominate over the less clinically evident signs of acromegaly. Occasionally, mixed plurihormonal tumors are encountered that also secrete ACTH, the glycoprotein hormone α subunit, or TSH in addition to GH. Patients with partially empty sellas may present with GH hypersecretion due to a small GH-secreting adenoma within the compressed rim of pituitary tissue; some of these may reflect the spontaneous necrosis of tumors that were previously larger. GH-secreting tumors rarely arise from ectopic pituitary tissue remnants in the nasopharynx or midline sinuses.

Table 339-11 Causes of Acromegaly

	Prevalence, %
Excess Growth Hormone Secretion	
Pituitary	98
Densely or sparsely granulated GH cell adenoma	60
Mixed GH cell and PRL cell adenoma	25
Mammosomatrope cell adenoma	10

	Plurihormonal adenoma
	GH cell carcinoma or metastases
	Multiple endocrine neoplasia 1 (GH cell adenoma)
	McCune-Albright syndrome
	Ectopic sphenoid or parapharyngeal sinus pituitary adenoma
Extrapituitary tumor	
Pancreatic islet cell tumor	<1
Lymphoma	
Excess Growth Hormone–Releasing Hormone Secretion	
Central	<1
Hypothalamic hamartoma, choristoma, ganglioneuroma	<1
Peripheral	<1
Bronchial carcinoid, pancreatic islet cell tumor, small cell lung cancer, adrenal adenoma, medullary thyroid carcinoma, pheochromocytoma	

Source: Adapted from S Melmed: N Engl J Med 322:966, 1990.

Abbreviations: GH, growth hormone; PRL, prolactin.

There are case reports of ectopic GH secretion by tumors of pancreatic, ovarian, lung, or hematopoietic origin. Rarely, excess GHRH production may cause acromegaly because of chronic stimulation of somatotropes. These patients present with classic features of acromegaly, elevated GH levels, pituitary enlargement on MRI, and pathologic characteristics of pituitary hyperplasia. The most common cause of GHRH-mediated acromegaly is a chest or abdominal carcinoid tumor. Although these tumors usually express positive GHRH immunoreactivity, clinical features of acromegaly are evident in only a minority of patients with carcinoid disease. Excessive GHRH also may be elaborated by hypothalamic tumors, usually choristomas or neuromas.

PRESENTATION AND DIAGNOSIS

Protean manifestations of GH and IGF-I hypersecretion are indolent and often are not clinically diagnosed for 10 years or more. Acral bony overgrowth results in frontal bossing, increased hand and foot size, mandibular enlargement with prognathism, and widened space between the lower incisor teeth. In children and adolescents, initiation of GH hypersecretion before epiphyseal long bone closure is associated with development of pituitary gigantism (Fig. 339-8). Soft tissue swelling results in increased heel pad thickness, increased shoe or glove size, ring tightening, characteristic coarse facial features, and a large fleshy nose. Other commonly encountered clinical features include hyperhidrosis, a deep and hollow-sounding voice, oily skin, arthropathy, kyphosis, carpal tunnel syndrome, proximal muscle weakness and fatigue, acanthosis nigricans, and skin tags. Generalized visceromegaly occurs, including cardiomegaly, macroglossia, and thyroid gland enlargement.

Figure 339-8

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Features of acromegaly/gigantism. A 22-year-old man with gigantism due to excess growth

Features of acromegaly/gigantism. A 22-year-old man with gigantism due to excess growth hormone is shown to the left of his identical twin. The increased height and prognathism ***(A)*** and enlarged hand ***(B)*** and foot ***(C)*** of the affected twin are apparent. Their clinical features began to diverge at the age of approximately 13 years. *(Reproduced from R Gagel, IE McCutcheon: N Engl J Med 324:524, 1999; with permission.)*

The most significant clinical impact of GH excess occurs with respect to the cardiovascular system. Coronary heart disease, cardiomyopathy with arrhythmias, left ventricular hypertrophy, decreased diastolic function, and hypertension ultimately occur in most patients if untreated. Upper airway obstruction with sleep apnea occurs in more than 60% of patients and is associated with both soft tissue laryngeal airway obstruction and central sleep dysfunction. Diabetes mellitus develops in 25% of patients with acromegaly, and most patients are intolerant of a glucose load (as GH counteracts the action of insulin). Acromegaly is associated with an increased risk of colon polyps and mortality from colonic malignancy; polyps are diagnosed in up to one-third of patients. Overall mortality is increased about threefold and is due primarily to cardiovascular and cerebrovascular disorders and respiratory disease. Unless GH levels are controlled, survival is reduced by an average of 10 years compared with an age-matched control population.

LABORATORY INVESTIGATION

Age- and sex-matched serum IGF-I levels are elevated in acromegaly. Consequently, an IGF-I level provides a useful laboratory screening measure when clinical features raise the possibility of acromegaly. Due to the pulsatility of GH secretion, measurement of a single random GH level is not useful for the diagnosis or exclusion of acromegaly and does not correlate with disease severity. The diagnosis of acromegaly is confirmed by demonstrating the failure of GH suppression to <0.4 μg/L within 1–2 h of an oral glucose load (75 g). When newer ultrasensitive GH assays are used, normal nadir GH levels are even lower (<0.05 μg/L). About 20% of patients exhibit a paradoxical GH rise after glucose. PRL should be measured, as it is elevated in ~25% of patients with acromegaly. Thyroid function, gonadotropins, and sex steroids may be attenuated because of tumor mass effects. Because most patients will undergo surgery with glucocorticoid coverage, tests of ACTH reserve in asymptomatic patients are more efficiently deferred until after surgery.

TREATMENT: ACROMEGALY

The goal of treatment is to control GH and IGF-I hypersecretion, ablate or arrest tumor growth, ameliorate comorbidities, restore mortality rates to normal, and preserve pituitary function.

Surgical resection of GH-secreting adenomas is the initial treatment for most patients (Fig. 339-9). Somatostatin analogues are used as adjuvant treatment for preoperative shrinkage of large invasive macroadenomas, immediate relief of debilitating symptoms, and reduction of GH hypersecretion; in frail patients experiencing morbidity; and in patients who decline surgery or, when surgery fails, to achieve biochemical control. Irradiation or repeat surgery may be required for patients who cannot tolerate or do not respond to adjunctive medical therapy. The high rate of late hypopituitarism and the slow rate (5–15 years) of biochemical response are the main disadvantages of radiotherapy. Irradiation is also relatively ineffective in normalizing IGF-I levels. Stereotactic ablation of GH-secreting adenomas by gamma-knife radiotherapy is promising, but initial reports suggest that long-term results and side effects are similar to those observed with conventional radiation. Somatostatin analogues may be required while awaiting the full benefits of radiotherapy. Systemic sequelae of acromegaly, including cardiovascular disease, diabetes, and arthritis, should be managed aggressively. Mandibular surgical repair may be indicated.

Figure 339-9

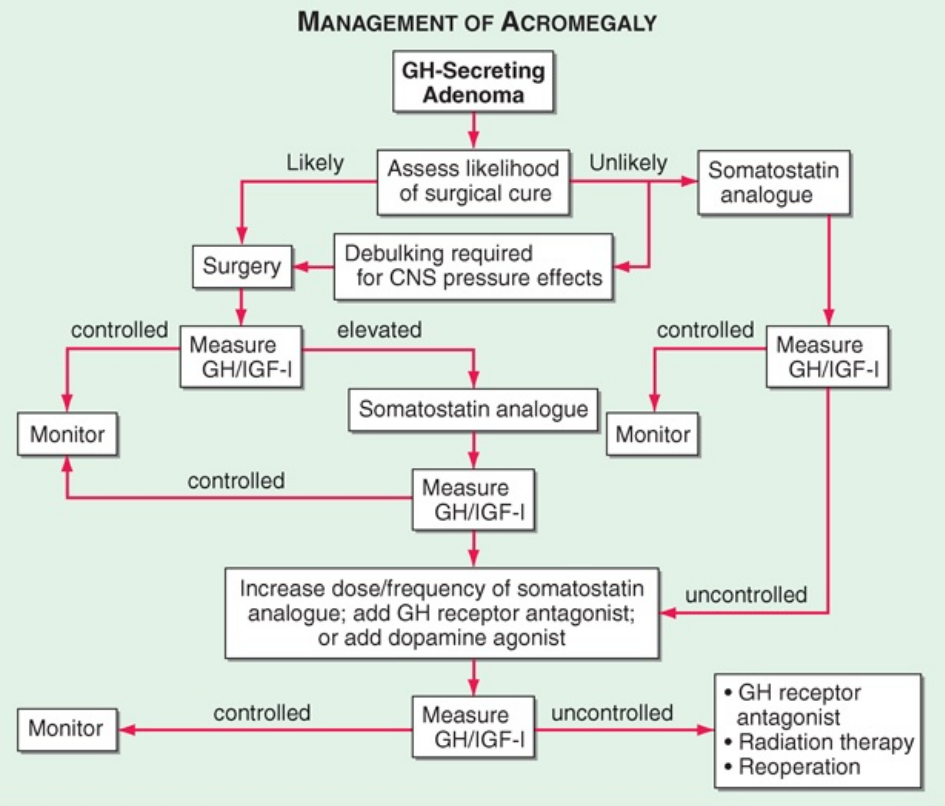

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of acromegaly. GH, growth hormone; CNS, central nervous system; IGF, insulin-like growth factor. *(Adapted from S Melmed et al: J Clin Endocrinol Metab 94:1509–1517, 2009; © The Endocrine Society.)*

Surgery

Transsphenoidal surgical resection by an experienced surgeon is the preferred primary treatment for both microadenomas (cure rate ~70%) and macroadenomas (<50% cured). Soft tissue swelling improves immediately after tumor resection. GH levels return to normal within an hour, and IGF-I levels are normalized within 3–4 days. In ~10% of patients, acromegaly may recur several years after apparently successful surgery; hypopituitarism develops in up to 15% of patients after surgery.

Somatostatin Analogues

Somatostatin analogues exert their therapeutic effects through SSTR2 and SSTR5 receptors, both of which invariably are expressed by GH-secreting tumors. Octreotide acetate is an eight-amino-acid synthetic somatostatin analogue. In contrast to native somatostatin, the analogue is relatively resistant to plasma degradation. It has a 2-h serum half-life and possesses fortyfold greater potency than native somatostatin to suppress GH. Octreotide is administered by subcutaneous injection, beginning with 50 μg tid; the dose can be increased gradually up to 1500 μg/d. Fewer than 10% of patients do not respond to the analogue. Octreotide suppresses integrated GH levels and normalizes IGF-I levels in ~ 75% of treated patients.

The long-acting somatostatin depot formulations, octreotide and lanreotide, are the preferred medical treatment for patients with acromegaly. *Sandostatin-LAR* is a sustained-release, long-acting formulation of octreotide incorporated into microspheres that sustain drug levels for several weeks after intramuscular injection. GH suppression occurs for as long as 6 weeks after a 30-mg intramuscular injection; long-term monthly treatment sustains GH and IGF-I suppression and also reduces pituitary tumor size in ~50% of patients. *Lanreotide* autogel, a slow-release depot somatostatin preparation, is a cyclic somatostatin octapeptide analogue that suppresses GH and IGF-I hypersecretion after a 60-mg subcutaneous injection. Long-term monthly administration controls GH hypersecretion in two-thirds of treated patients and improves patient compliance because of the long interval required between drug injections. Rapid relief of headache and soft tissue swelling occurs in ~ 75% of patients within days to weeks of somatostatin analogue initiation. Most patients report symptomatic improvement, including amelioration of headache, perspiration, obstructive apnea, and cardiac failure.

Side Effects

Somatostatin analogues are well tolerated in most patients. Adverse effects are short-lived and mostly

relate to drug-induced suppression of gastrointestinal motility and secretion. Nausea, abdominal discomfort, fat malabsorption, diarrhea, and flatulence occur in one-third of patients, and these symptoms usually remit within 2 weeks. Octreotide suppresses postprandial gallbladder contractility and delays gallbladder emptying; up to 30% of patients develop long-term echogenic sludge or asymptomatic cholesterol gallstones. Other side effects include mild glucose intolerance due to transient insulin suppression, asymptomatic bradycardia, hypothyroxinemia, and local injection site discomfort.

GH Receptor Antagonist

Pegvisomant antagonizes endogenous GH action by blocking peripheral GH binding to its receptor. Consequently, serum IGF-I levels are suppressed, reducing the deleterious effects of excess endogenous GH. Pegvisomant is administered by daily subcutaneous injection (10–20 mg) and normalizes IGF-I in >90% of patients. GH levels, however, remain elevated as the drug does not have antitumor actions. Side effects include reversible liver enzyme elevation, lipodystrophy, and injection site pain. Tumor size should be monitored by MRI.

Combined treatment with monthly somatostatin analogues and weekly or biweekly pegvisomant injections has been used effectively in resistant patients.

Dopamine Agonists

Bromocriptine and cabergoline may modestly suppress GH secretion in some patients. High doses of bromocriptine (≥20 mg/d) or cabergoline (0.5 mg/d) are usually required to achieve modest GH therapeutic efficacy. Combined treatment with octreotide and cabergoline may induce additive biochemical control compared with either drug alone.

Radiation

External radiation therapy or high-energy stereotactic techniques are used as adjuvant therapy for acromegaly. An advantage of radiation is that patient compliance with long-term treatment is not required. Tumor mass is reduced, and GH levels are attenuated over time. However, 50% of patients require at least 8 years for GH levels to be suppressed to <5 μg/L; this level of GH reduction is achieved in about 90% of patients after 18 years but represents suboptimal GH suppression. Patients may require interim medical therapy for several years before attaining maximal radiation benefits. Most patients also experience hypothalamic-pituitary damage, leading to gonadotropin, ACTH, and/or TSH deficiency within 10 years of therapy.

In summary, surgery is the preferred primary treatment for GH-secreting microadenomas (Fig. 339-9). The high frequency of GH hypersecretion after macroadenoma resection usually necessitates adjuvant or primary medical therapy for these larger tumors. Patients unable to receive or respond to unimodal medical treatment may benefit from combined treatments or can be offered radiation.

ADRENOCORTICOTROPIC HORMONE (SEE ALSO CHAP. 342)

Synthesis

ACTH-secreting corticotrope cells constitute about 20% of the pituitary cell population. ACTH (39 amino acids) is derived from the POMC precursor protein (266 amino acids) that also generates several other peptides, including β-lipotropin, β-endorphin, met-enkephalin, α-melanocyte-stimulating hormone (α-MSH), and corticotropin-like intermediate lobe protein (CLIP). The POMC gene is potently suppressed by glucocorticoids and induced by CRH, arginine vasopressin (AVP), and proinflammatory cytokines, including IL-6, as well as leukemia inhibitory factor.

CRH, a 41-amino-acid hypothalamic peptide synthesized in the paraventricular nucleus as well as in higher brain centers, is the predominant stimulator of ACTH synthesis and release. The CRH receptor is a GPCR that is expressed on the corticotrope and signals to induce POMC transcription.

Secretion

ACTH secretion is pulsatile and exhibits a characteristic circadian rhythm, peaking at 6 a.m. and reaching a nadir about midnight. Adrenal glucocorticoid secretion, which is driven by ACTH, follows a parallel diurnal pattern. ACTH circadian rhythmicity is determined by variations in secretory pulse amplitude rather than changes in pulse frequency. Superimposed on this endogenous rhythm, ACTH levels are increased by physical and psychological stress, exercise, acute illness, and insulin-induced hypoglycemia.

Loss of cortisol feedback inhibition, as occurs in primary adrenal failure, results in extremely high ACTH levels. Glucocorticoid-mediated negative regulation of the hypothalamic-pituitary-adrenal (HPA) axis occurs as a consequence of both hypothalamic CRH suppression and direct attenuation of pituitary POMC gene expression and ACTH release.

Acute inflammatory or septic insults activate the HPA axis through the integrated actions of proinflammatory cytokines, bacterial toxins, and neural signals. The overlapping cascade of ACTH-inducing cytokines [tumor necrosis factor (TNF); IL-1, -2, and -6; and leukemia inhibitory factor] activates hypothalamic CRH and AVP secretion, pituitary POMC gene expression, and local pituitary

paracrine cytokine networks. The resulting cortisol elevation restrains the inflammatory response and enables host protection. Concomitantly, cytokine-mediated central glucocorticoid receptor resistance impairs glucocorticoid suppression of the HPA. Thus, the neuroendocrine stress response reflects the net result of highly integrated hypothalamic, intrapituitary, and peripheral hormone and cytokine signals.

Action

The major function of the HPA axis is to maintain metabolic homeostasis and mediate the neuroendocrine stress response. ACTH induces adrenocortical steroidogenesis by sustaining adrenal cell proliferation and function. The receptor for ACTH, designated *melanocortin-2 receptor*, is a GPCR that induces steroidogenesis by stimulating a cascade of steroidogenic enzymes (Chap. 342).

ACTH Deficiency

PRESENTATION AND DIAGNOSIS

Secondary adrenal insufficiency occurs as a result of pituitary ACTH deficiency. It is characterized by fatigue, weakness, anorexia, nausea, vomiting, and, occasionally, hypoglycemia. In contrast to primary adrenal failure, hypocortisolism associated with pituitary failure usually is not accompanied by hyperpigmentation or mineralocorticoid deficiency.

ACTH deficiency is commonly due to glucocorticoid withdrawal after treatment-associated suppression of the HPA axis. Isolated ACTH deficiency may occur after surgical resection of an ACTH-secreting pituitary adenoma that has suppressed the HPA axis; this phenomenon is suggestive of a surgical cure. The mass effects of other pituitary adenomas or sellar lesions may lead to ACTH deficiency, but usually in combination with other pituitary hormone deficiencies. Partial ACTH deficiency may be unmasked in the presence of an acute medical or surgical illness, when clinically significant hypocortisolism reflects diminished ACTH reserve. Rarely, *TPIT* or POMC mutations result in primary ACTH deficiency.

LABORATORY DIAGNOSIS

Inappropriately low ACTH levels in the setting of low cortisol levels are characteristic of diminished ACTH reserve. Low basal serum cortisol levels are associated with blunted cortisol responses to ACTH stimulation and impaired cortisol response to insulin-induced hypoglycemia, or testing with metyrapone or CRH. **For a description of provocative ACTH tests, see Chap. 342.**

TREATMENT: ACTH DEFICIENCY

Glucocorticoid replacement therapy improves most features of ACTH deficiency. The total daily dose of hydrocortisone replacement preferably should not exceed 25 mg daily, divided into two or three doses. Prednisone (5 mg each morning) is longer-acting and has fewer mineralocorticoid effects than hydrocortisone. Some authorities advocate lower maintenance doses in an effort to avoid cushingoid side effects. Doses should be increased severalfold during periods of acute illness or stress.

Cushing's Syndrome (ACTH-Producing Adenoma)

(See also Chap. 342)

ETIOLOGY AND PREVALENCE

Pituitary corticotrope adenomas account for 70% of patients with endogenous causes of Cushing's syndrome. However, it should be emphasized that iatrogenic hypercortisolism is the most common cause of cushingoid features. Ectopic tumor ACTH production, cortisol-producing adrenal adenomas, adrenal carcinoma, and adrenal hyperplasia account for the other causes; rarely, ectopic tumor CRH production is encountered.

ACTH-producing adenomas account for about 10–15% of all pituitary tumors. Because the clinical features of Cushing's syndrome often lead to early diagnosis, most ACTH-producing pituitary tumors are relatively small microadenomas. However, macroadenomas also are seen while some ACTH-expressing adenomas are clinically silent. Cushing's disease is 5–10 times more common in women than in men. These pituitary adenomas exhibit unrestrained ACTH secretion, with resultant hypercortisolemia. However, they retain partial suppressibility in the presence of high doses of administered glucocorticoids, providing the basis for dynamic testing to distinguish pituitary from nonpituitary causes of Cushing's syndrome.

PRESENTATION AND DIAGNOSIS

The diagnosis of Cushing's syndrome presents two great challenges: (1) to distinguish patients with pathologic cortisol excess from those with physiologic or other disturbances of cortisol production and (2) to determine the etiology of cortisol excess.

Typical features of chronic cortisol excess include thin skin, central obesity, hypertension, plethoric moon facies, purple striae and easy bruisability, glucose intolerance or diabetes mellitus, gonadal dysfunction, osteoporosis, proximal muscle weakness, signs of hyperandrogenism (acne, hirsutism), and psychological disturbances (depression, mania, and psychoses) (Table 339-12). Hematopoietic

features of hypercortisolism include leukocytosis, lymphopenia, and eosinopenia. Immune suppression includes delayed hypersensitivity. These protean yet commonly encountered manifestations of hypercortisolism make it challenging to decide which patients mandate formal laboratory evaluation. Certain features make pathologic causes of hypercortisolism more likely; they include characteristic central redistribution of fat, thin skin with striae and bruising, and proximal muscle weakness. In children and in young females, early osteoporosis may be particularly prominent. The primary cause of death is cardiovascular disease, but infections and risk of suicide are also increased.

Table 339-12 Clinical Features of Cushing's Syndrome (All Ages)

Symptoms/Signs	Frequency, %
Obesity or weight gain (>115% ideal body weight)	80
Thin skin	80
Moon facies	75
Hypertension	75
Purple skin striae	65
Hirsutism	65
Menstrual disorders (usually amenorrhea)	60
Plethora	60
Abnormal glucose tolerance	55
Impotence	55
Proximal muscle weakness	50
Truncal obesity	50
Acne	45
Bruising	45
Mental changes	45
Osteoporosis	40
Edema of lower extremities	30
Hyperpigmentation	20
Hypokalemic alkalosis	15
Diabetes mellitus	15

Source: Adapted from MA Magiokou et al, in ME Wierman (ed): *Diseases of the Pituitary.* Totowa, NJ, Humana, 1997.

Rapid development of features of hypercortisolism associated with skin hyperpigmentation and severe myopathy suggests an ectopic source of ACTH. Hypertension, hypokalemic alkalosis, glucose intolerance, and edema are also more pronounced in these patients. Serum potassium levels <3.3 mmol/L are evident in ~70% of patients with ectopic ACTH secretion but are seen in <10% of patients with pituitary-dependent Cushing's syndrome.

LABORATORY INVESTIGATION

The diagnosis of Cushing's syndrome is based on laboratory documentation of endogenous hypercortisolism. Measurement of 24-h urine free cortisol (UFC) is a precise and cost-effective screening test. Alternatively, the failure to suppress plasma cortisol after an overnight 1-mg dexamethasone suppression test can be used to identify patients with hypercortisolism. As nadir levels of cortisol occur at night, elevated midnight samples of cortisol are suggestive of Cushing's syndrome. Basal plasma ACTH levels often distinguish patients with ACTH-independent (adrenal or exogenous glucocorticoid) from those with ACTH-dependent (pituitary, ectopic ACTH) Cushing's syndrome. Mean basal ACTH levels are about eightfold higher in patients with ectopic ACTH secretion than in those with pituitary ACTH-secreting adenomas. However, extensive overlap of ACTH levels in these two disorders precludes using ACTH measurements to make the distinction. Instead, dynamic testing based on differential sensitivity to glucocorticoid feedback or ACTH stimulation in response to CRH or cortisol reduction is used to distinguish ectopic from pituitary sources of excess ACTH (Table 339-13). Very rarely, circulating CRH levels are elevated, reflecting ectopic tumor-derived secretion of CRH and often ACTH. **For further discussion of dynamic testing for Cushing's syndrome, see Chap. 342**.

Table 339-13 Differential Diagnosis of ACTH-Dependent Cushing's Syndrome*

	ACTH-Secreting Pituitary Tumor	Ectopic ACTH Secretion
Etiology	Pituitary corticotrope adenoma	Bronchial, abdominal carcinoid

	Plurihormonal adenoma	Small cell lung cancer Thymoma
Sex	F > M	M > F
Clinical features	Slow onset	Rapid onset Pigmentation Severe myopathy
Serum potassium <3.3 μg/L	<10%	75%
24-h urinary free cortisol (UFC)	High	High
Basal ACTH level	Inappropriately high	Very high
Dexamethasone suppression		
1 mg overnight		
Low dose (0.5 mg q6h)	Cortisol >5 μg/dL	Cortisol >5 μg/dL
High dose (2 mg q6h) UFC > 80% suppressed	Cortisol <5 μg/dL Microadenomas: 90% Macroadenomas: 50%	Cortisol >5 μg/dL 10%
Inferior petrosal sinus sampling (IPSS)		
Basal		
IPSS: peripheral	>2	<2
CRH-induced IPSS: peripheral	>3	<3

*ACTH-independent causes of Cushing's syndrome are diagnosed by suppressed ACTH levels and an adrenal mass in the setting of hypercortisolism. Iatrogenic Cushing's syndrome is excluded by history.

Abbreviations: ACTH, adrenocorticotropic hormone; CRH, corticotropin-releasing hormone; F, female; M, male.

Most ACTH-secreting pituitary tumors are <5 mm in diameter, and about half are undetectable by sensitive MRI. The high prevalence of incidental pituitary microadenomas diminishes the ability to distinguish ACTH-secreting pituitary tumors accurately from nonsecreting incidentalomas.

INFERIOR PETROSAL VENOUS SAMPLING

Because pituitary MRI with gadolinium enhancement is insufficiently sensitive to detect small (<2 mm) pituitary ACTH-secreting adenomas, bilateral inferior petrosal sinus ACTH sampling before and after CRH administration may be required to distinguish these lesions from ectopic ACTH-secreting tumors that may have similar clinical and biochemical characteristics. Simultaneous assessment of ACTH in each inferior petrosal vein and in the peripheral circulation provides a strategy for confirming and localizing pituitary ACTH production. Sampling is performed at baseline and 2, 5, and 10 min after intravenous bovine CRH (1 μg/kg) injection. An increased ratio (>2) of inferior petrosal:peripheral vein ACTH confirms pituitary Cushing's syndrome. After CRH injection, peak petrosal:peripheral ACTH ratios ≥3 confirm the presence of a pituitary ACTH-secreting tumor. The sensitivity of this test is >95%, with very rare false-positive results. False-negative results may be encountered in patients with aberrant venous drainage. Petrosal sinus catheterizations are technically difficult, and about 0.05% of patients develop neurovascular complications. The procedure should not be performed in patients with hypertension or in the presence of a well-visualized pituitary adenoma on MRI.

TREATMENT: CUSHING'S SYNDROME

Selective transsphenoidal resection is the treatment of choice for Cushing's disease (Fig. 339-10). The remission rate for this procedure is ~80% for microadenomas but <50% for macroadenomas. After successful tumor resection, most patients experience a postoperative period of symptomatic ACTH deficiency that may last up to 12 months. This usually requires low-dose cortisol replacement, as patients experience both steroid withdrawal symptoms and have a suppressed HPA axis. Biochemical recurrence occurs in approximately 5% of patients in whom surgery was initially successful.

Figure 339-10

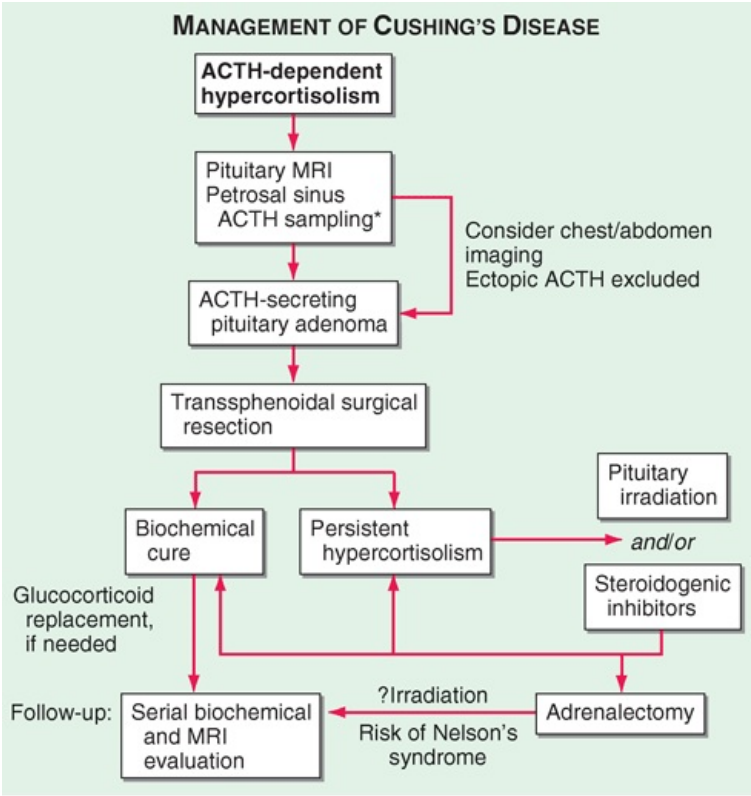

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of Cushing's syndrome. ACTH, adrenocorticotropin hormone; MRI, magnetic resonance imaging. *, Not usually required.

When initial surgery is unsuccessful, repeat surgery is sometimes indicated, particularly when a pituitary source for ACTH is well documented. In older patients, in whom issues of growth and fertility are less important, hemi- or total hypophysectomy may be necessary if a discrete pituitary adenoma is not recognized. Pituitary irradiation may be used after unsuccessful surgery, but it cures only about 15% of patients. Because the effects of radiation are slow and only partially effective in adults, steroidogenic inhibitors are used in combination with pituitary irradiation to block adrenal effects of persistently high ACTH levels.

Ketoconazole, an imidazole derivative antimycotic agent, inhibits several P450 enzymes and effectively lowers cortisol in most patients with Cushing's disease when administered twice daily (600–1200 mg/d). Elevated hepatic transaminases, gynecomastia, impotence, gastrointestinal upset, and edema are common side effects. *Metyrapone* (2–4 g/d) inhibits 11β-hydroxylase activity and normalizes plasma cortisol in up to 75% of patients. Side effects include nausea and vomiting, rash, and exacerbation of acne or hirsutism. *Mitotane* (*o,p'*-DDD; 3–6 g/d orally in four divided doses) suppresses cortisol hypersecretion by inhibiting 11β-hydroxylase and cholesterol side-chain cleavage enzymes and by destroying adrenocortical cells. Side effects of mitotane include gastrointestinal symptoms, dizziness, gynecomastia, hyperlipidemia, skin rash, and hepatic enzyme elevation. It also may lead to hypoaldosteronism. Other agents include aminoglutethimide (250 mg tid), trilostane (200–1000 mg/d), cyproheptadine (24 mg/d), and IV etomidate (0.3 mg/kg per hour). Glucocorticoid insufficiency is a potential side effect of agents used to block steroidogenesis.

The use of steroidogenic inhibitors has decreased the need for bilateral adrenalectomy. Removal of both adrenal glands corrects hypercortisolism but may be associated with significant morbidity rates and necessitates permanent glucocorticoid and mineralocorticoid replacement. Adrenalectomy in the setting of residual corticotrope adenoma tissue predisposes to the development of Nelson's syndrome, a disorder characterized by rapid pituitary tumor enlargement and increased pigmentation secondary to high ACTH levels. Radiation therapy may be indicated to prevent the development of Nelson's syndrome after adrenalectomy.

GONADOTROPINS: FSH AND LH

Synthesis and Secretion

Gonadotrope cells constitute about 10% of anterior pituitary cells and produce two gonadotropins—LH and FSH. Like TSH and hCG, LH and FSH are glycoprotein hormones that consist of α and β subunits. The α subunit is common to these glycoprotein hormones; specificity is conferred by the β subunits,

which are expressed by separate genes.

Gonadotropin synthesis and release are dynamically regulated. This is particularly true in women, in whom rapidly fluctuating gonadal steroid levels vary throughout the menstrual cycle. Hypothalamic GnRH, a 10-amino-acid peptide, regulates the synthesis and secretion of both LH and FSH. GnRH is secreted in discrete pulses every 60–120 min, and the pulses in turn elicit LH and FSH pulses (Fig. 339-3). The pulsatile mode of GnRH input is essential to its action; pulses prime gonadotrope responsiveness, whereas continuous GnRH exposure induces desensitization. Based on this phenomenon, long-acting GnRH agonists are used to suppress gonadotropin levels in children with precocious puberty and in men with prostate cancer (Chap. 95) and are used in some ovulation-induction protocols to reduce levels of endogenous gonadotropins (Chap. 347). Estrogens act at both the hypothalamus and the pituitary to modulate gonadotropin secretion. Chronic estrogen exposure is inhibitory, whereas rising estrogen levels, as occur during the preovulatory surge, exert positive feedback to increase gonadotropin pulse frequency and amplitude. Progesterone slows GnRH pulse frequency but enhances gonadotropin responses to GnRH. Testosterone feedback in men also occurs at the hypothalamic and pituitary levels and is mediated in part by its conversion to estrogens.

Although GnRH is the main regulator of LH and FSH secretion, FSH synthesis is also under separate control by the gonadal peptides inhibin and activin, which are members of the transforming growth factor β (TGF-β) family. Inhibin selectively suppresses FSH, whereas activin stimulates FSH synthesis (Chap. 347).

Action

The gonadotropin hormones interact with their respective GPCRs expressed in the ovary and testis, evoking germ-cell development and maturation and steroid hormone biosynthesis. In women, FSH regulates ovarian follicle development and stimulates ovarian estrogen production. LH mediates ovulation and maintenance of the corpus luteum. In men, LH induces Leydig cell testosterone synthesis and secretion and FSH stimulates seminiferous tubule development and regulates spermatogenesis.

Gonadotropin Deficiency

Hypogonadism is the most common presenting feature of adult hypopituitarism even when other pituitary hormones are also deficient. It is often a harbinger of hypothalamic or pituitary lesions that impair GnRH production or delivery through the pituitary stalk. As noted above, hypogonadotropic hypogonadism is a common presenting feature of hyperprolactinemia.

A variety of inherited and acquired disorders are associated with isolated hypogonadotropic hypogonadism (IHH) (Chap. 346). Hypothalamic defects associated with GnRH deficiency include two X-linked disorders, Kallmann syndrome (see above) and mutations in the DAX1 gene, as well as dominant mutations in FGFR1. Mutations in *GPR54*, kisspeptin, the GnRH receptor, and the LH β or FSH β subunit genes are additional causes of selective gonadotropin deficiency. Acquired forms of GnRH deficiency leading to hypogonadotropism are seen in association with anorexia nervosa (Chap. 79), stress, starvation, and extreme exercise but also may be idiopathic. Hypogonadotropic hypogonadism in these disorders is reversed by removal of the stressful stimulus or by caloric replenishment.

PRESENTATION AND DIAGNOSIS

In premenopausal women, hypogonadotropic hypogonadism presents as diminished ovarian function leading to oligomenorrhea or amenorrhea, infertility, decreased vaginal secretions, decreased libido, and breast atrophy. In hypogonadal adult men, secondary testicular failure is associated with decreased libido and potency, infertility, decreased muscle mass with weakness, reduced beard and body hair growth, soft testes, and characteristic fine facial wrinkles. Osteoporosis occurs in both untreated hypogonadal women and men.

LABORATORY INVESTIGATION

Central hypogonadism is associated with low or inappropriately normal serum gonadotropin levels in the setting of low sex hormone concentrations (testosterone in men, estradiol in women). Because gonadotropin secretion is pulsatile, valid assessments may require repeated measurements or the use of pooled serum samples. Men have reduced sperm counts.

Intravenous GnRH (100 μg) stimulates gonadotropes to secrete LH (which peaks within 30 min) and FSH (which plateaus during the ensuing 60 min). Normal responses vary according to menstrual cycle stage, age, and sex of the patient. Generally, LH levels increase about threefold, whereas FSH responses are less pronounced. In the setting of gonadotropin deficiency, a normal gonadotropin response to GnRH indicates intact pituitary gonadotrope function and suggests a hypothalamic abnormality. An absent response, however, cannot reliably distinguish pituitary from hypothalamic causes of hypogonadism. For this reason, GnRH testing usually adds little to the information gained from baseline evaluation of the hypothalamic-pituitary-gonadotrope axis except in cases of isolated GnRH deficiency (e.g., Kallmann syndrome).

MRI examination of the sellar region and assessment of other pituitary functions usually are indicated

in patients with documented central hypogonadism.

TREATMENT: GONADOTROPIN DEFI CIENCY

In males, testosterone replacement is necessary to achieve and maintain normal growth and development of the external genitalia, secondary sex characteristics, male sexual behavior, and androgenic anabolic effects, including maintenance of muscle function and bone mass. Testosterone may be administered by intramuscular injections every 1–4 weeks or by using skin patches that are replaced daily (Chap. 346). Testosterone gels are also available. Gonadotropin injections [hCG or human menopausal gonadotropin (hMG)] over 12–18 months are used to restore fertility. Pulsatile GnRH therapy (25–150 ng/kg every 2 h), administered by a subcutaneous infusion pump, is also effective for treatment of hypothalamic hypogonadism when fertility is desired.

In premenopausal women, cyclical replacement of estrogen and progesterone maintains secondary sexual characteristics and integrity of genitourinary tract mucosa and prevents premature osteoporosis (Chap. 347). Gonadotropin therapy is used for ovulation induction. Follicular growth and maturation are initiated using hMG or recombinant FSH; hCG or human luteinizing hormone (hLH) is subsequently injected to induce ovulation. As in men, pulsatile GnRH therapy can be used to treat hypothalamic causes of gonadotropin deficiency.

Nonfunctioning and Gonadotropin-Producing Pituitary Adenomas

ETIOLOGY AND PREVALENCE

Nonfunctioning pituitary adenomas include those that secrete little or no pituitary hormones as well as tumors that produce too little hormone to result in recognizable clinical features. They are the most common type of pituitary adenoma and are usually macroadenomas at the time of diagnosis because clinical features are not apparent until tumor mass effects occur. Based on immunohistochemistry, most clinically nonfunctioning adenomas can be shown to originate from gonadotrope cells. These tumors typically produce small amounts of intact gonadotropins (usually FSH) as well as uncombined α, LH β, and FSH β subunits. Tumor secretion may lead to elevated α and FSH β subunits and, rarely, to increased LH β subunit levels. Some adenomas express α subunits without FSH or LH. TRH administration often induces an atypical increase of tumor-derived gonadotropins or subunits.

PRESENTATION AND DIAGNOSIS

Clinically nonfunctioning tumors often present with optic chiasm pressure and other symptoms of local expansion or may be incidentally discovered on an MRI performed for another indication (incidentaloma). Rarely, menstrual disturbances or ovarian hyperstimulation occur in women with large tumors that produce FSH and LH. More commonly, adenoma compression of the pituitary stalk or surrounding pituitary tissue leads to attenuated LH and features of hypogonadism. PRL levels are usually slightly increased, also because of stalk compression. It is important to distinguish this circumstance from true prolactinomas, as nonfunctioning tumors do not shrink in response to treatment with dopamine agonists.

LABORATORY INVESTIGATION

The goal of laboratory testing in clinically nonfunctioning tumors is to classify the type of the tumor, identify hormonal markers of tumor activity, and detect possible hypopituitarism. Free α subunit levels may be elevated in 10–15% of patients with nonfunctioning tumors. In female patients, peri- or postmenopausal basal FSH concentrations are difficult to distinguish from tumor-derived FSH elevation. Premenopausal women have cycling FSH levels, also preventing clear-cut diagnostic distinction from tumor-derived FSH. In men, gonadotropin-secreting tumors may be diagnosed because of slightly increased gonadotropins (FSH > LH) in the setting of a pituitary mass. Testosterone levels are usually low despite the normal or increased LH level, perhaps reflecting reduced LH bioactivity or the loss of normal LH pulsatility. Because this pattern of hormone test results is also seen in primary gonadal failure and, to some extent, with aging (Chap. 346), the finding of increased gonadotropins alone is insufficient for the diagnosis of a gonadotropin-secreting tumor. In the majority of patients with gonadotrope adenomas, TRH administration stimulates LH β subunit secretion; this response is not seen in normal individuals. GnRH testing, however, is not helpful for making the diagnosis. For nonfunctioning and gonadotropin-secreting tumors, the diagnosis usually rests on immunohistochemical analyses of surgically resected tumor tissue, as the mass effects of these tumors usually necessitate resection.

Although acromegaly or Cushing's syndrome usually presents with unique clinical features, clinically inapparent (silent) somatotrope or corticotrope adenomas may only be diagnosed by immunostaining of resected tumor tissue. If PRL levels are <100 μg/L in a patient harboring a pituitary mass, a nonfunctioning adenoma causing pituitary stalk compression should be considered.

Treatment: Nonfunctioning and Gonadotropin-Producing Pituitary Adenomas

Asymptomatic small nonfunctioning microadenomas adenomas with no threat to vision may be followed with regular MRI and visual field testing without immediate intervention. However, for macroadenomas, transsphenoidal surgery is indicated to reduce tumor size and relieve mass effects (Fig. 339-11). Although it is not usually possible to remove all adenoma tissue surgically, vision improves in 70% of

patients with preoperative visual field defects. Preexisting hypopituitarism that results from tumor mass effects may improve or resolve completely. Beginning about 6 months postoperatively, MRI scans should be performed yearly to detect tumor regrowth. Within 5–6 years after successful surgical resection, ~15% of nonfunctioning tumors recur. When substantial tumor remains after transsphenoidal surgery, adjuvant radiotherapy may be indicated to prevent tumor regrowth. Radiotherapy may be deferred if no postoperative residual mass is evident.

Figure 339-11

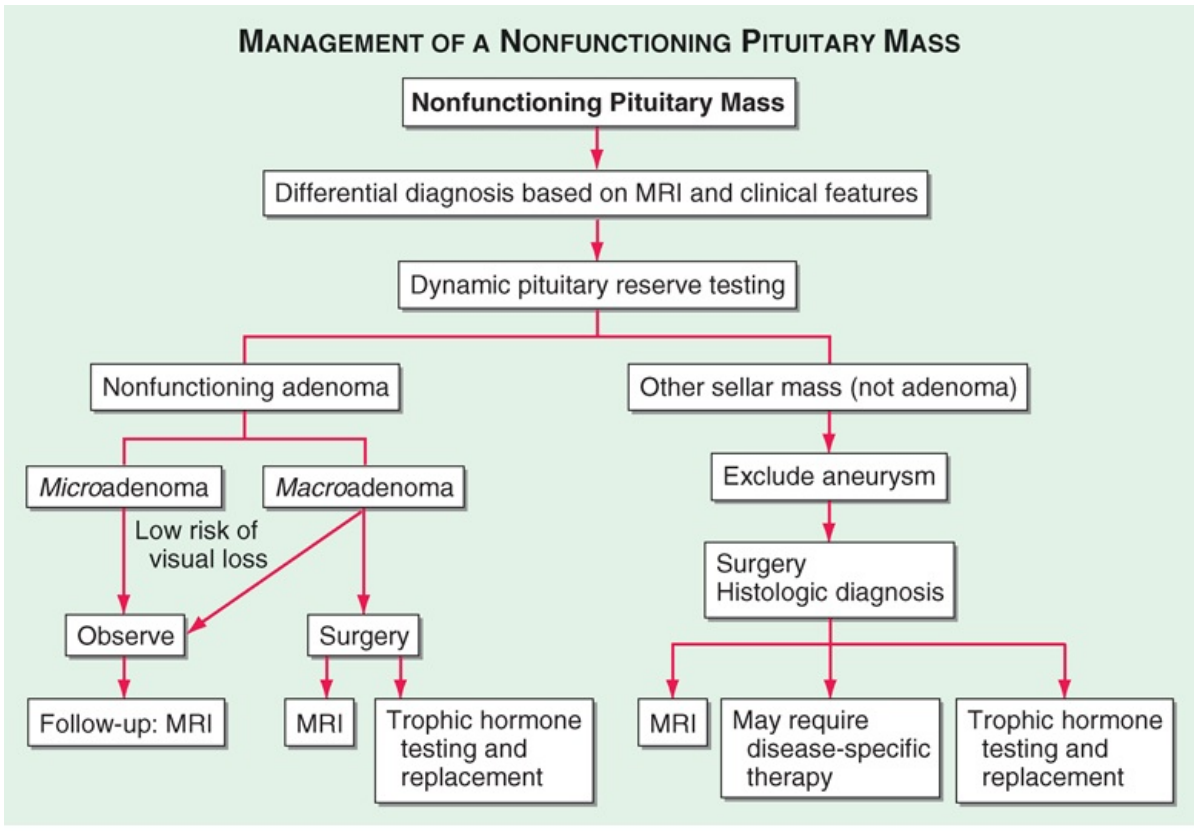

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of a nonfunctioning pituitary mass.

Nonfunctioning pituitary tumors respond poorly to dopamine agonist treatment and somatostatin analogues are largely ineffective for shrinking these tumors. The selective GnRH antagonist Nal-Glu GnRH suppresses FSH hypersecretion but has no effect on adenoma size.

THYROID-STIMULATING HORMONE

Synthesis and Secretion

TSH-secreting thyrotrope cells constitute 5% of the anterior pituitary cell population. TSH is structurally related to LH and FSH. It shares a common α subunit with these hormones but contains a specific TSH β subunit. TRH is a hypothalamic tripeptide (pyroglutamyl histidylprolinamide) that acts through a GPCR to stimulate TSH synthesis and secretion; it also stimulates the lactotrope cell to secrete PRL. TSH secretion is stimulated by TRH, whereas thyroid hormones, dopamine, somatostatin, and glucocorticoids suppress TSH by overriding TRH induction.

Thyrotrope growth and TSH secretion are both induced when negative feedback inhibition by thyroid hormones is removed. Thus, thyroid damage (including surgical thyroidectomy), radiation-induced hypothyroidism, chronic thyroiditis, and prolonged goitrogen exposure are associated with increased TSH. Long-standing untreated hypothyroidism can lead to thyrotrope hyperplasia and pituitary enlargement, which may be evident on MRI.

Action

TSH is secreted in pulses, though the excursions are modest in comparison to other pituitary hormones because of the low amplitude of the pulses and the relatively long half-life of TSH. Consequently, single determinations of TSH suffice to assess its circulating levels. TSH binds to a GPCR on thyroid follicular cells to stimulate thyroid hormone synthesis and release (Chap. 341).

TSH Deficiency

Features of central hypothyroidism due to TSH deficiency mimic those seen with primary hypothyroidism but are generally less severe. Pituitary hypothyroidism is characterized by low basal TSH levels in the setting of low free thyroid hormone. In contrast, patients with hypothyroidism of hypothalamic origin (presumably due to a lack of endogenous TRH) may exhibit normal or even slightly elevated TSH levels. The TSH produced in this circumstance appears to have reduced biologic activity because of altered glycosylation.

TRH (200 μg) injected intravenously causes a two- to threefold increase in TSH (and PRL) levels within 30 min. Although TRH testing can be used to assess TSH reserve, abnormalities of the thyroid axis usually can be detected based on basal free T4 and TSH levels, and TRH testing is rarely indicated.

Thyroid-replacement therapy should be initiated after adequate adrenal function has been established. Dose adjustment is based on thyroid hormone levels and clinical parameters rather than the TSH level.

TSH-Secreting Adenomas

TSH-producing macroadenomas are rare but are often large and locally invasive when they occur. Patients usually present with thyroid goiter and hyperthyroidism, reflecting overproduction of TSH. Diagnosis is based on demonstrating elevated serum free T4 levels, inappropriately normal or high TSH secretion, and MRI evidence of a pituitary adenoma.

It is important to exclude other causes of inappropriate TSH secretion, such as resistance to thyroid hormone, an autosomal dominant disorder caused by mutations in the thyroid hormone β receptor (Chap. 341). The presence of a pituitary mass and elevated α subunit levels are suggestive of a TSH-secreting tumor. Dysalbuminemic hyperthyroxinemia syndromes, caused by mutations in serum thyroid hormone binding proteins, are also characterized by elevated thyroid hormone levels, but with normal rather than suppressed TSH levels. Moreover, free thyroid hormone levels are normal in these disorders, most of which are familial.

TREATMENT: TSH-SECRETING ADENOMAS

The initial therapeutic approach is to remove or debulk the tumor mass surgically, usually using a transsphenoidal approach. Total resection is not often achieved as most of these adenomas are large and locally invasive. Normal circulating thyroid hormone levels are achieved in about two-thirds of patients after surgery. Thyroid ablation or antithyroid drugs (methimazole and propylthiouracil) can be used to reduce thyroid hormone levels. Somatostatin analogue treatment effectively normalizes TSH and α subunit hypersecretion, shrinks the tumor mass in 50% of patients, and improves visual fields in 75% of patients; euthyroidism is restored in most patients. Because somatostatin analogues markedly suppress TSH, biochemical hypothyroidism often requires concomitant thyroid hormone replacement, which may also further control tumor growth.

Diabetes Insipidus

See Chap. 340 for diagnosis and treatment of diabetes insipidus.

FURTHER READINGS

Castinetti F et al: Long-term results of stereotactic radiosurgery in secretory pituitary adenomas. J Clin Endocrinol Metab 94:3400, 2009[PMID: 19509108] [Full Text]

Elston MS et al: Familial pituitary tumor syndromes. Nat Rev Endocrinol 5:453, 2009[PMID: 19564887] [Full Text]

Fernandez A et al: Prevalence of pituitary adenomas: A community-based, cross-sectional study in Banbury (Oxfordshire, UK). Clin Endocrinol (Oxf) 72:377, 2010[PMID: 19650784] [Full Text]

Ilias I et al: Cushing's syndrome due to ectopic corticotropin secretion: Twenty years' experience at the National Institutes of Health. J Clin Endocrinol Metab 90:4955, 2005[PMID: 15914534] [Full Text]

Kharlip J et al: Recurrence of hyperprolactinemia after withdrawal of long-term cabergoline therapy. J Clin Endocrinol Metab 94:2428, 2009[PMID: 19336508] [Full Text]

Liu H et al: Systematic review: The safety and efficacy of growth hormone in the healthy elderly. Ann Intern Med 146:104, 2007[PMID: 17227934] [Full Text]

Melmed S: Acromegaly. N Engl J Med 355:2558, 2006[PMID: 17167139] [Full Text]

——— et al: Diagnosis and treatment of hyperprolactinemia: An Endorcine Society Clinical Practice Guideline. J Clin Endocrinol Metab 96:273, 2011

Molitch ME: Evaluation and treatment of adult growth hormone deficiency: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 91:1621, 2006[PMID: 16636129] [Full Text]

Nieman LK et al: The diagnosis of Cushing's syndrome: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 93:1526, 2008[PMID: 18334580] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE NEUROHYPOPHYSIS: INTRODUCTION

The neurohypophysis, or posterior pituitary, is formed by axons that originate in large cell bodies in the supraoptic and paraventricular nuclei of the hypothalamus. It produces two hormones: (1) arginine vasopressin (AVP), also known as antidiuretic hormone, and (2) oxytocin. AVP acts on the renal tubules to reduce water loss by concentrating the urine. Oxytocin stimulates postpartum milk letdown in response to suckling. AVP deficiency causes diabetes insipidus (DI), which is characterized by the production of large amounts of dilute urine. Excessive or inappropriate AVP production predisposes to hyponatremia if water intake is not reduced in parallel with urine output.

VASOPRESSIN

Synthesis and Secretion

AVP is a nonapeptide composed of a six-member disulfide ring and a tripeptide tail (Fig. 340-1). It is synthesized via a polypeptide precursor that includes AVP, neurophysin, and copeptin, all encoded by a single gene on chromosome 20. After preliminary processing and folding, the precursor is packaged in neurosecretory vesicles, where it is transported down the axon, further processed to AVP, and stored in neurosecretory vesicles until the hormone and other components are released by exocytosis into peripheral blood.

Figure 340-1

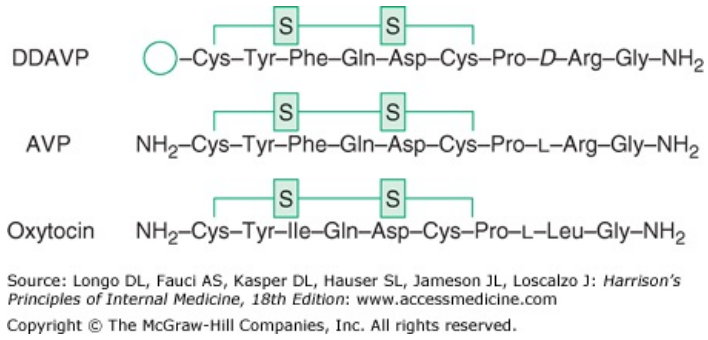

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary structures of arginine vasopressin (AVP), oxytocin, and desmopressin.

AVP secretion is regulated primarily by the "effective" osmotic pressure of body fluids. This control is mediated by specialized hypothalamic cells known as *osmoreceptors*, which are extremely sensitive to small changes in the plasma concentration of sodium and certain other solutes but normally are insensitive to other solutes such as urea and glucose. The osmoreceptors appear to include inhibitory as well as stimulatory components that function in concert to form a threshold, or set point, control system. Below this threshold, plasma AVP is suppressed to levels that permit the development of a maximum water diuresis. Above it, plasma AVP rises steeply in direct proportion to plasma osmolarity, quickly reaching levels sufficient to effect a maximum antidiuresis. The absolute levels of plasma osmolarity/sodium at which minimally and maximally effective levels of plasma AVP occur, vary appreciably from person to person, apparently owing to genetic influences on the set and sensitivity of the system. However, the average threshold, or set point, for AVP release corresponds to a plasma osmolarity or sodium of about 280 mosmol/L or 135 meq/L, respectively; levels only 2–4% higher normally result in maximum antidiuresis.

Though it is relatively stable in a healthy adult, the set point of the osmoregulatory system can be lowered by pregnancy, the menstrual cycle, estrogen, and relatively large, acute reductions in blood pressure or volume. Those reductions are mediated largely by neuronal afferents that originate in transmural pressure receptors of the heart and large arteries and project via the vagus and glossopharyngeal nerves to the brainstem, from which postsynaptic projections ascend to the hypothalamus. These pathways maintain a tonic inhibitory tone that decreases when blood volume or pressure falls by >10–20%. This baroregulatory system is probably of minor importance in the physiology of AVP secretion because the hemodynamic changes required to effect it usually do not occur during normal activities. However, the baroregulatory system undoubtedly plays an important

role in AVP secretion in patients with large, acute disturbances of hemodynamic function.

AVP secretion also can be stimulated by nausea, acute hypoglycemia, glucocorticoid deficiency, smoking, and, possibly, hyperangiotensinemia. The emetic stimuli are extremely potent since they typically elicit immediate, 50- to 100-fold increases in plasma AVP even when the nausea is transient and is not associated with vomiting or other symptoms. They appear to act via the emetic center in the medulla and can be blocked completely by treatment with antiemetics such as fluphenazine. There is no evidence that pain or other noxious stresses have any effect on AVP unless they elicit a vasovagal reaction with its associated nausea and hypotension.

Action

The most important, if not the only, physiologic action of AVP is to reduce water excretion by promoting concentration of urine. This antidiuretic effect is achieved by increasing the hydroosmotic permeability of cells that line the distal tubule and medullary collecting ducts of the kidney (Fig. 340-2). In the absence of AVP, these cells are impermeable to water and reabsorb little, if any, of the relatively large volume of dilute filtrate that enters from the proximal nephron. This results in the excretion of very large volumes (as much as 0.2 mL/kg per min) of maximally dilute urine (specific gravity and osmolarity ~ 1.000 and 50 mosmol/L, respectively), a condition known as *water diuresis*. In the presence of AVP, these cells become selectively permeable to water, allowing the water to diffuse back down the osmotic gradient created by the hypertonic renal medulla. As a result, the dilute fluid passing through the tubules is concentrated and the rate of urine flow decreases. The magnitude of this effect varies in direct proportion to the plasma AVP concentration and, at maximum levels, approximates a urine flow rate as low as 0.35 mL/min and a urine osmolarity as high as 1200 mosmol/L. This action is mediated via binding to G protein–coupled V_2 receptors on the serosal surface of the cell, activation of adenyl cyclase, and insertion into the luminal surface of water channels composed of a protein known as *aquaporin 2* (AQP2). The V_2 receptors and aquaporin 2 are encoded by genes on chromosomes Xq28 and 12q13, respectively.

Figure 340-2

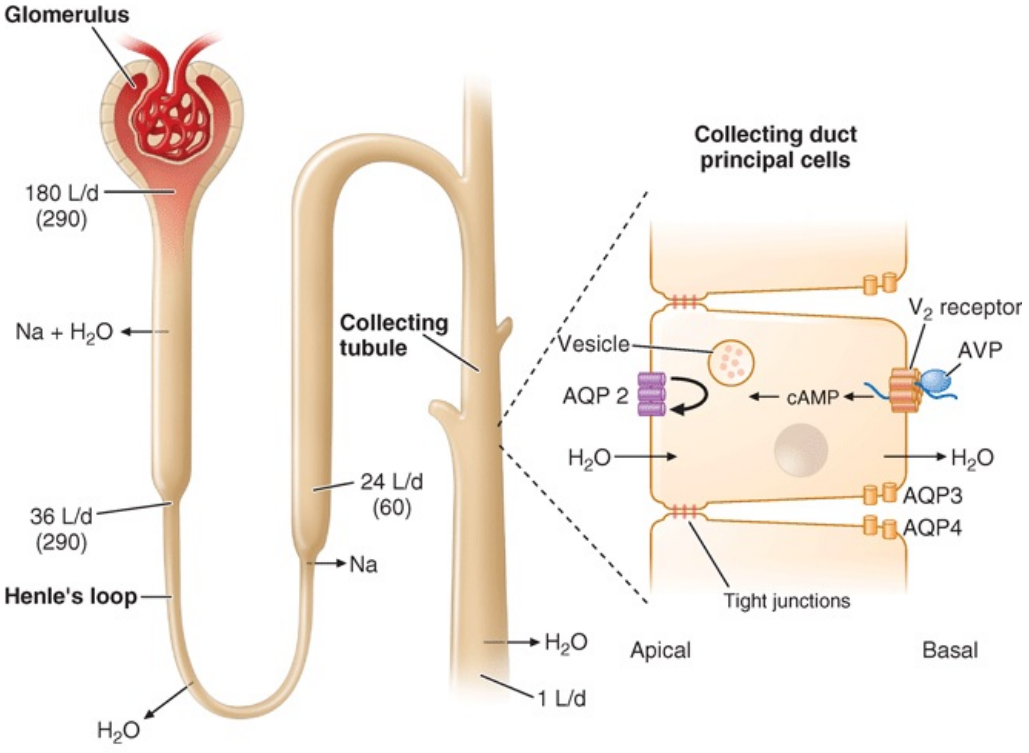

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Antidiuretic effect of arginine vasopressin (AVP) in the regulation of urine volume. In a typical 70-kg adult, the kidney filters ~180 L/d of plasma. Of this, ~144 L (80%) is reabsorbed isosmotically in the proximal tubule and another 8 L (4–5%) is reabsorbed without solute in the descending limb of Henle's loop. The remainder is diluted to an osmolarity of ~60 mmol/kg by selective reabsorption of sodium and chloride in the ascending limb. In the absence of AVP, the urine issuing from the loop passes largely unmodified through the distal tubules and collecting ducts, resulting in a maximum water diuresis. In the presence of AVP, solute-free water is reabsorbed osmotically through the principal cells of the collecting ducts, resulting in the excretion of a much smaller volume of concentrated urine. This antidiuretic effect is mediated via a G protein–coupled V_2 receptor that

increases intracellular cyclic AMP, thereby inducing translocation of aquaporin 2 (AQP 2) water channels into the apical membrane. The resultant increase in permeability permits an influx of water that diffuses out of the cell through AQP 3 and AQP 4 water channels on the basal-lateral surface. The net rate of flux across the cell is determined by the number of AQP 2 water channels in the apical membrane and the strength of the osmotic gradient between tubular fluid and the renal medulla. Tight junctions on the lateral surface of the cells serve to prevent unregulated water flow.

At high concentrations, AVP also causes contraction of smooth muscle in blood vessels and in the gastrointestinal tract, induces glycogenolysis in the liver, and potentiates adrenocorticotropic hormone (ACTH) release by corticotropin-releasing factor. These effects are mediated by V_{1a} or V_{1b} receptors that are coupled to phospholipase C. Their role, if any, in human physiology/pathophysiology is uncertain.

Metabolism

AVP distributes rapidly into a space roughly equal to the extracellular fluid volume. It is cleared irreversibly with a $t_{1/2}$ of 10–30 minutes. Most AVP clearance is due to degradation in the liver and kidneys. During pregnancy, the metabolic clearance of AVP is increased three- to fourfold due to placental production of an N-terminal peptidase.

THIRST

Because AVP cannot reduce water loss below a certain minimum level obligated by urinary solute load and evaporation from skin and lungs, a mechanism for ensuring adequate intake is essential for preventing dehydration. This vital function is performed by the thirst mechanism. Like AVP, thirst is regulated primarily by an osmostat that is situated in the anteromedial hypothalamus and is able to detect very small changes in the plasma concentration of sodium and certain other effective solutes. The thirst osmostat appears to be "set" about 5% higher than the AVP osmostat. This arrangement ensures that thirst, polydipsia, and dilution of body fluids do not occur until plasma osmolarity/sodium start to exceed the defensive capacity of the antidiuretic mechanism.

OXYTOCIN

Oxytocin is also a nonapeptide, and it differs from AVP only at positions 3 and 8 (Fig. 340-1). However, it has relatively little antidiuretic effect and seems to act mainly on mammary ducts to facilitate milk letdown during nursing. It also may help initiate or facilitate labor by stimulating contraction of uterine smooth muscle, but it is not clear if this action is physiologic or necessary for normal delivery.

DEFICIENCIES OF VASOPRESSIN SECRETION AND ACTION

Diabetes Insipidus

CLINICAL CHARACTERISTICS

Decreased secretion or action of AVP usually manifests as diabetes insipidus, a syndrome characterized by the production of abnormally large volumes of dilute urine. The 24-hour urine volume is >50 mL/kg body weight, and the osmolarity is <300 mosmol/L. The polyuria produces symptoms of urinary frequency, enuresis, and/or nocturia, which may disturb sleep and cause mild daytime fatigue or somnolence. It also results in a slight rise in plasma osmolarity that stimulates thirst and a commensurate increase in fluid intake (polydipsia). Overt clinical signs of dehydration are uncommon unless fluid intake is impaired.

ETIOLOGY

Deficient secretion of AVP can be primary or secondary. The primary form usually results from agenesis or irreversible destruction of the neurohypophysis and is referred to variously as *neurohypophyseal DI*, *pituitary DI*, or *central DI*. It can be caused by a variety of congenital, acquired, or genetic disorders, but in about one-half of all adult patients it is idiopathic (Table 340-1). The surgically induced forms of pituitary DI usually appear within 24 hours and then go through a 2- to 3-week interim period of inappropriate antidiuresis, after which they may or may not recur. The genetic form usually is transmitted in an autosomal dominant mode and is caused by diverse mutations in the coding region of the AVP–neurophysin II (or *AVP-NPII*) gene. All the mutations alter one or more amino acids known to be critical for correct folding of the prohormone, thus interfering with its processing and trafficking through the endoplasmic reticulum. The AVP deficiency and DI develop gradually several months to years after birth, progressing from partial to severe and permanent DI. They appear to result from accumulation of misfolded mutant precursor followed by selective degeneration of AVP-producing magnocellular neurons. An autosomal recessive form due to an inactivating mutation in the AVP portion of the gene, an X-linked recessive form due to an unidentified gene on Xq28, and an autosomal recessive form due to mutations of the *WFS 1* gene responsible for Wolfram's syndrome [diabetes insipidus, diabetes mellitus, optic atrophy, and neural deafness (DIDMOAD)] have also been described. A primary deficiency of plasma AVP also can result from increased metabolism by an N-terminal aminopeptidase produced by the placenta. It is referred to as *gestational DI* since the signs and symptoms manifest during pregnancy and usually remit several weeks after delivery.

Table 340-1 Causes of Diabetes Insipidus
Pituitary diabetes insipidus
Acquired
Head trauma (closed and penetrating) including pituitary surgery
Neoplasms
Primary
Craniopharyngioma
Pituitary adenoma (suprasellar)
Dysgerminoma
Meningioma
Metastatic (lung, breast)
Hematologic (lymphoma, leukemia)
Granulomas
Sarcoidosis
Histiocytosis
Xanthoma disseminatum
Infectious
Chronic meningitis
Viral encephalitis
Toxoplasmosis
Inflammatory
Lymphocytic infundibuloneurohypophysitis
Granulomatosis with polyangiitis (Wegener's)
Lupus erythematosus
Scleroderma
Chemical toxins
Tetrodotoxin
Snake venom
Vascular
Sheehan's syndrome
Aneurysm (internal carotid)
Aortocoronary bypass
Hypoxic encephalopathy
Pregnancy (vasopressinase)
Idiopathic
Congenital malformations
Septo-optic dysplasia
Midline craniofacial defects
Holoprosencephaly
Hypogenesis, ectopia of pituitary
Genetic
Autosomal dominant (*AVP-neurophysin gene*)
Autosomal recessive (*AVP-neurophysin gene*)
Autosomal recessive-Wolfram-(4p – WFS 1 gene)
X-linked recessive (Xq28)
Deletion chromosome 7q
Nephrogenic diabetes insipidus
Acquired
Drugs
Lithium
Demeclocycline
Methoxyflurane

Amphotericin B
Aminoglycosides
Cisplatin
Rifampin
Foscarnet
Metabolic
Hypercalcemia, hypercalciuria
Hypokalemia
Obstruction (ureter or urethra)
Vascular
Sickle cell disease and trait
Ischemia (acute tubular necrosis)
Granulomas
Sarcoidosis
Neoplasms
Sarcoma
Infiltration
Amyloidosis
Pregnancy
Idiopathic
Genetic
X-linked recessive (*AVP receptor-2 gene*)
Autosomal recessive (AQP2 gene)
Autosomal dominant (AQP2 gene)
Primary polydipsia
Acquired
Psychogenic
Schizophrenia
Obsessive compulsive disorder
Dipsogenic (abnormal thirst)
Granulomas
Sarcoidosis
Infectious
Tuberculous meningitis
Head trauma (closed and penetrating)
Demyelination
Multiple sclerosis
Drugs
Lithium
Carbamazepine
Idiopathic
Iatrogenic

Secondary deficiencies of AVP result from inhibition of secretion by excessive intake of fluids. They are referred to as *primary polydipsia* and can be divided into three subcategories. One of them, *dipsogenic DI*, is characterized by inappropriate thirst caused by a reduction in the set of the osmoregulatory mechanism. It sometimes occurs in association with multifocal diseases of the brain such as neurosarcoid, tuberculous meningitis, and multiple sclerosis but is often idiopathic. The second subtype, *psychogenic polydipsia*, is not associated with thirst, and the polydipsia seems to be a feature of psychosis or obsessive compulsive disorder. The third subtype, *iatrogenic polydipsia*, results from recommendations to increase fluid intake for its presumed health benefits.

Primary deficiencies in the antidiuretic action of AVP result in *nephrogenic DI* (Table 340-1). They can be genetic, acquired, or drug induced. The genetic form usually is transmitted in a semirecessive X-linked manner and is caused by mutations in the coding region of the V_2 receptor gene that impair trafficking and/or ligand binding of the mutant receptor. Autosomal recessive or dominant forms are

caused by AQP2 gene mutations that result in complete or partial defects in trafficking and function of the water channels in distal and collecting tubules of the kidney.

Secondary deficiencies in the antidiuretic response to AVP result from polyuria per se. They are caused by washout of the medullary concentration gradient and/or suppression of aquaporin function. They usually resolve 24–48 hours after the polyuria is corrected but can complicate interpretation of some acute tests used for differential diagnosis.

PATHOPHYSIOLOGY

When the secretion or action of AVP falls below 80–85% of normal, urine concentration ceases and the rate of urine output rises to symptomatic levels. If the defect is due to pituitary, gestational, or nephrogenic DI, the polyuria results in a small (1–2%) decrease in body water and a commensurate increase in plasma osmolarity and sodium concentration that stimulate thirst and a compensatory increase in water intake. As a result, *hypernatremia and other overt physical or laboratory signs of dehydration do not develop unless the patient also has a defect in thirstor fails to drink for some other reason*.

The severity of the antidiuretic defect varies markedly from patient to patient. In some, the deficiencies in AVP secretion or action are so severe that even an intense stimulus such as nausea or severe dehydration does not raise plasma AVP enough to concentrate the urine. In others, the deficiency is incomplete, and a modest stimulus such as a few hours of fluid deprivation, smoking, or a vasovagal reaction increases plasma AVP sufficiently to raise urine osmolarity as high as 800 mosmol/L. The maximum achieved is usually less than normal, but that is the case largely because maximal concentrating capacity is temporarily impaired by chronic polyuria.

In primary polydipsia, the pathogenesis of the polydipsia and polyuria is the reverse of that in pituitary, nephrogenic, and gestational DI. Thus, the excessive intake of fluids slightly increases body water, thereby reducing plasma osmolarity, AVP secretion, and urinary concentration. The latter results in a compensatory increase in urinary free-water excretion that varies in direct proportion to intake. Therefore, hyponatremia or clinically appreciable overhydration is uncommon unless the polydipsia is very severe or the compensatory water diuresis is impaired by a drug or disease that stimulates or mimics endogenous AVP.

In the dipsogenic form of primary polydipsia, fluid intake is excessive because the osmotic threshold for thirst appears to be reset to the left, often well below that for AVP release. When deprived of fluids or subjected to another acute osmotic or nonosmotic stimulus, these individuals invariably increase plasma AVP normally, but the resultant increase in urine concentration is usually subnormal because the individuals' renal capacities to concentrate the urine also are blunted temporarily by chronic polyuria. Thus, the maximum level of urine osmolarity achieved is usually indistinguishable from that in patients with partial pituitary, partial gestational, or partial nephrogenic DI. Patients with psychogenic or iatrogenic polydipsia respond similarly to fluid restriction but do not complain of thirst and usually offer other explanations for their high fluid intake.

DIFFERENTIAL DIAGNOSIS

When symptoms of urinary frequency, enuresis, nocturia, and/or persistent thirst are present, the possibility of DI should be evaluated after excluding glucosuria by collecting a 24-hour urine on ad libitum fluid intake. If the volume exceeds 50 mL/kg per day (3500 mL in a 70-kg male) and the osmolarity is >300 mosmol/L, DI is confirmed and the patient should be evaluated further to determine the type.

In differentiating among the various types of DI, the history alone may be sufficient if it reveals a likely antecedent such as pituitary surgery. Usually, however, that type of indicator is absent, ambiguous, or misleading and other approaches are needed. Except in the rare patient with hypertonic dehydration under basal conditions, differentiation should begin with a *fluid deprivation test*. It can be performed on an outpatient basis if the necessary staff and facilities are available. To minimize patient discomfort, avoid excessive dehydration, and maximize the information obtained, the test should be started in the morning and continued with hourly monitoring of body weight, plasma osmolarity and/or sodium concentration, urine volume, and urine osmolarity until either of two endpoints is reached. If fluid deprivation does not result in urine concentration (osmolarity >300 mosmol/L, specific gravity >1.010) before body weight decreases by 5% or plasma osmolarity/sodium rise above the upper limit of normal, the patient has severe pituitary or severe nephrogenic DI. These disorders usually can be distinguished by administering desmopressin (0.03 μg/kg SC or IV) and repeating the measurement of urine osmolarity 1–2 hours later. An increase of >50% indicates severe pituitary DI, whereas a smaller or absent response is strongly suggestive of nephrogenic DI.

Conversely, if fluid deprivation results in concentration of the urine, severe defects in AVP secretion and action are excluded and the question becomes whether the patient has partial pituitary DI, partial nephrogenic DI, or primary polydipsia. The maximum levels of urine osmolarity achieved before and after desmopressin injection are of no help in this regard because the values in the three groups vary widely and overlap owing to impairment of renal concentrating capacity caused by polyuria per se. Therefore, another approach is needed to differentiate among them. The easiest and least expensive

method is to measure plasma AVP before and during the fluid deprivation test and analyze the results in relation to the concurrent plasma and urine osmolarity (Fig. 340-3). This approach invariably differentiates partial nephrogenic DI from partial pituitary DI and primary polydipsia. It also differentiates partial pituitary DI from primary polydipsia if plasma osmolarity and/or sodium are clearly above the normal range when the hormone is measured. However, the requisite level of hypertonic dehydration may be difficult to produce by fluid deprivation alone when urine concentration occurs. Therefore, it is usually necessary to continue the fluid deprivation and infuse hypertonic (3%) saline at a rate of 0.1 mL/kg per min until plasma osmolarity/sodium measured every 20 to 30 minutes reach or slightly exceed the upper limit of normal. At that point the plasma osmolarity/sodium, which is usually reached in 30 to 90 minutes, the measurement of plasma AVP should be repeated and the result related to plasma osmolarity/sodium is as before.

Figure 340-3

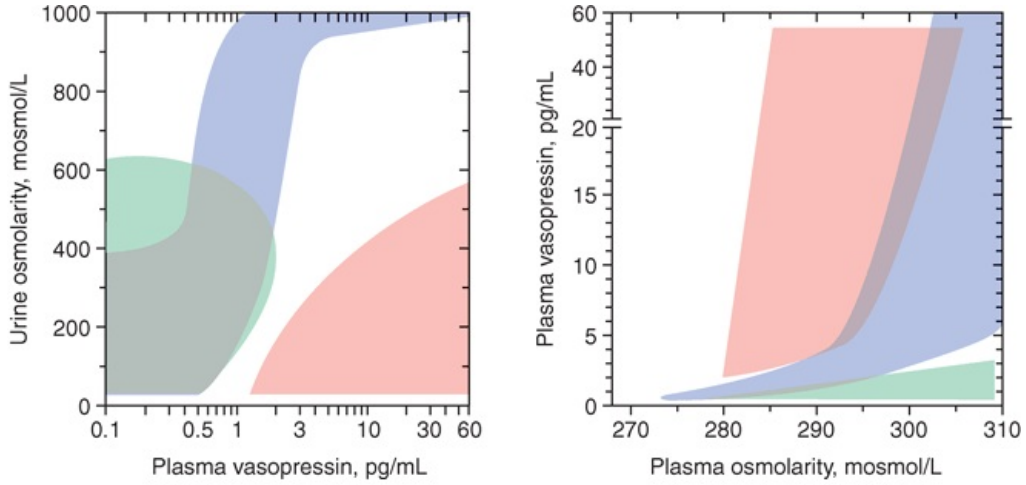

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship of plasma AVP to urine osmolarity (*left*) and plasma osmolarity (*right*) before and during fluid deprivation–hypertonic saline infusion test in patients who are normal or have primary polydipsia (blue zones), pituitary diabetes insipidus (green zones), or nephrogenic diabetes insipidus (pink zones).

An alternative method of differential diagnosis is MRI of the pituitary and hypothalamus. In most healthy adults and children, the posterior pituitary emits a hyperintense signal in T1-weighted midsagittal images. This "bright spot" is almost always present in patients with primary polydipsia but is invariably absent or abnormally small in patients with pituitary DI. It is usually also small or absent in nephrogenic DI presumably because of high secretion and turnover of AVP. Thus, a normal bright spot virtually excludes pituitary DI, argues against nephrogenic DI, and strongly suggests primary polydipsia. Lack of the bright spot is less helpful, however, because it is absent not only in pituitary and nephrogenic DI but also in some healthy adults and patients with empty sella who do not have DI or AVP deficiency.

The other way to distinguish among the three basic types of DI is a closely monitored trial of desmopressin therapy.

TREATMENT: DIABETES INSIPIDUS

The signs and symptoms of uncomplicated pituitary DI can be eliminated completely by treatment with desmopressin (DDAVP), a synthetic analogue of AVP (Fig. 340-1). DDAVP acts selectively at V_2 receptors to increase urine concentration and decrease urine flow in a dose-dependent manner (Fig. 340-4). It is also more resistant to degradation than is AVP and has a three- to fourfold longer duration of action. Desmopressin can be given by IV or SC injection, nasal inhalation, or oral tablet. The doses required to control pituitary DI completely vary widely, depending on the patient and the route of administration. However, they usually range from 1–2 μg qd or bid by injection, 10–20 μg bid or tid by nasal spray, or 100–400 μg bid or tid orally. The onset of action is rapid, ranging from as little as 15 minutes after injection to 60 minutes after oral administration. When given in doses sufficient to normalize urinary osmolarity and flow completely, desmopressin produces a slight (1–3%) increase in total body water and a commensurate decrease in plasma osmolarity and sodium concentration that rapidly eliminates thirst and polydipsia. Consequently, water balance is maintained and hyponatremia does not develop unless the osmoregulation of thirst is also impaired or fluid intake is excessive for another reason such as a misconception about the need to prevent dehydration. Fortunately, thirst abnormalities are rare in pituitary DI, and other motivations to drink excessively usually can be eliminated by patient education. Therefore, desmopressin usually can be given safely in doses

sufficient to normalize urine output completely, thereby avoiding the inconvenience and discomfort of intermittent escape otherwise needed to prevent water intoxication.

Figure 340-4

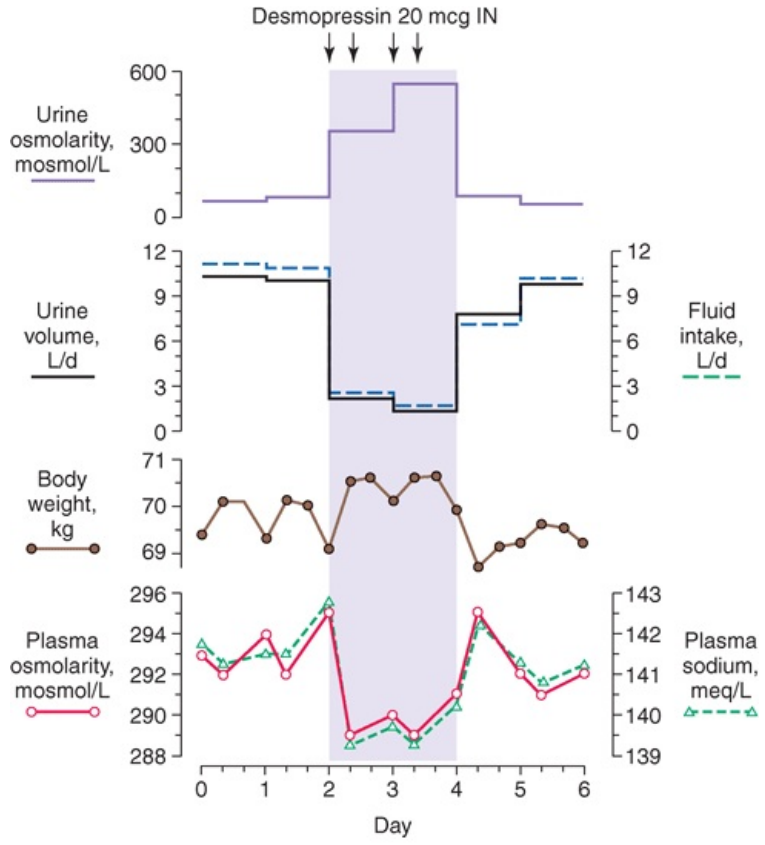

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effect of desmopressin therapy on water balance in a patient with uncomplicated pituitary diabetes insipidus. Note that treatment rapidly reduces thirst and fluid intake as well as urine output to normal, with only a slight increase in body water (weight) and a decrease in plasma osmolarity/sodium. *[From P Felig, L Frohman (eds): Endocrinology and Metabolism, 4th ed. New York, McGraw-Hill, 2001, with permission.]*

Primary polydipsia cannot be treated safely with desmopressin or any other antidiuretic drug because eliminating the polyuria does not eliminate the urge to drink. Therefore, it produces hyponatremia and/or other signs of water intoxication, usually within 24 to 48 hours if urine output is normalized completely. Patient education may eliminate iatrogenic polydipsia, but it is largely ineffective in psychogenic or dipsogenic DI. In these patients, the only help currently available is to try to prevent water intoxication by warning about the use of drugs that can impair urinary free-water excretion directly or indirectly.

The polyuria and polydipsia of nephrogenic DI are not affected by treatment with standard doses of desmopressin. If resistance is partial, it may be overcome by tenfold higher doses, but this treatment is too expensive and inconvenient to be useful chronically. However, treatment with conventional doses of a thiazide diuretic and/or amiloride in conjunction with a low-sodium diet and a prostaglandin synthesis inhibitor (e.g., indomethacin) usually reduces the polyuria and polydipsia by 30–70% and may eliminate them completely in some patients. Side effects such as hypokalemia and gastric irritation can be minimized by the use of amiloride or potassium supplements and by taking medications with meals.

Adipsic Hypernatremia

CLINICAL CHARACTERISTICS

A defect in the thirst mechanism results in adipsic hypernatremia, a syndrome characterized by chronic or recurrent hypertonic dehydration. The hypernatremia varies widely in severity and usually is associated with signs of hypovolemia such as tachycardia, postural hypotension, azotemia, hyperuricemia, and hypokalemia. Muscle weakness, pain, rhabdomyolysis, hyperglycemia, hyperlipidemia, and acute renal failure may also occur. DI usually does not exist at presentation but may develop during rehydration.

ETIOLOGY

Deficient thirst is usually due to hypogenesis or destruction of the osmoreceptors in the anterior hypothalamus. Because of their proximity, the osmoreceptors that regulate AVP secretion also are usually impaired. These defects can result from various congenital malformations of midline brain structures or may be acquired due to diseases such as occlusions of the anterior communicating artery, primary or metastatic tumors in the hypothalamus, head trauma, surgery, granulomatous diseases such as sarcoidosis and histiocytosis, AIDS, and cytomegalovirus encephalitis.

PATHOPHYSIOLOGY

A deficiency of thirst results in a failure to drink enough water to replenish obligatory renal and extrarenal losses, causing hypertonic dehydration. In most patients, the response of AVP to osmotic stimulation is also deficient (Fig. 340-5). If the deficiency is partial, it may not be clinically apparent at first because the hypertonicity and hypovolemia are severe enough to stimulate the release of AVP in the small amounts necessary to concentrate the urine. However, when the hypertonicity and hypovolemia are reduced, plasma AVP falls and polyuria develop, often before the dehydration is corrected fully. Patients with a complete lack of osmoregulation do not develop DI at any level of hydration because they cannot osmotically suppress or stimulate AVP secretion. Therefore, a hyponatremic syndrome indistinguishable from inappropriate antidiuresis may develop if rehydration is excessive. In most patients, the neurohypophysis and the AVP response to hemodynamic or emetic stimuli are normal. In a few, however, the neurohypophysis is also destroyed, resulting in a combination of chronic pituitary DI and hypodipsia that is particularly difficult to manage.

Figure 340-5

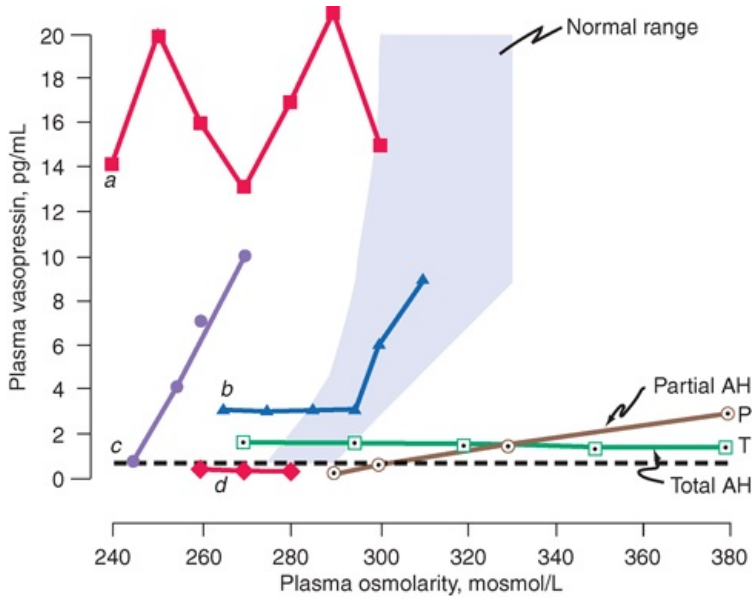

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Heterogeneity of osmoregulatory dysfunction in adipsic hypernatremia (AH) and the syndrome of inappropriate antidiuresis (SIAD). Each line depicts schematically the relationship of plasma arginine vasopressin (AVP) to plasma osmolarity during water loading and/or infusion of 3% saline in a patient with either AH (open symbols) or SIAD (closed symbols). The shaded area indicates the normal range of the relationship. The horizontal broken line indicates the plasma AVP level below which the hormone is undetectable and urinary concentration usually does not occur. Lines P and T represent patients with a selective deficiency in the osmoregulation of thirst and AVP that is either partial () or total (). In the latter, plasma AVP does not change in response to increases or decreases in plasma osmolarity but remains within a range sufficient to concentrate the urine even if overhydration produces hypotonic hyponatremia. In contrast, if the osmoregulatory deficiency is partial (), rehydration of the patient suppresses plasma AVP to levels that result in urinary dilution and polyuria before plasma osmolarity and sodium are reduced to normal. Lines *a*–*d* represent different defects in the osmoregulation of plasma AVP observed in patients with SIAD. In *a* (), plasma AVP is markedly elevated and fluctuates widely without relation to changes in plasma osmolarity, indicating complete loss of osmoregulation. In *b* (), plasma AVP remains fixed at a slightly elevated level until plasma osmolarity reaches the normal range, at which point it begins to rise appropriately, indicating a selective defect in the inhibitory component of the osmoregulatory mechanism. In *c* (), plasma AVP rises in close correlation with plasma osmolarity before the latter reaches the normal range, indicating downward resetting of the osmostat. In *d* (), plasma AVP appears to be osmoregulated normally, suggesting that the inappropriate antidiuresis is caused by some other abnormality.

DIFFERENTIAL DIAGNOSIS

Adipsic hypernatremia usually can be distinguished clinically from other causes of inadequate fluid intake (e.g., coma, paralysis, restraints, absence of fresh water) that can also result in hypertonic dehydration. Previous episodes and/or denial of thirst and failure to drink spontaneously when the patient is conscious, unrestrained, and hypernatremic are virtually diagnostic of adipsia. The hypernatremia caused by excessive oral or intravenous intake of sodium can also be distinguished by the history and/or physical examination and laboratory signs of volume expansion rather than contraction.

TREATMENT: ADIPSIC HYPERNATREMIA

Adipsic hypernatremia should be treated by administering water orally if the patient is alert and cooperative or by using hypotonic fluids (0.45% saline or 5% dextrose and water) via IV if the patient is not. The amount of free water in liters required to correct the deficit (ΔFW) can be estimated from body weight in kg (BW) and the serum sodium concentration in mmol/L (S_{Na}) by the formula $\Delta FW = 0.5BW \times [(S_{Na} - 140)/140]$. If serum glucose ($S_{Glu}$) is elevated, the measured S_{Na} should be corrected (S_{Na}^*) by the formula $S_{Na}^* = S_{Na} + [(S_{Glu} - 90)/36]$. This amount plus an allowance for continuing insensible and urinary losses should be given over a 24- to 48-hour period. Close monitoring of serum sodium as well as fluid intake and urinary output is essential because, depending on the extent of osmoreceptor deficiency (Fig. 340-5), some patients will develop AVP-deficient DI, requiring desmopressin therapy to complete rehydration; others will develop hyponatremia and a syndrome of inappropriate antidiuresis (SIAD)-like picture if overhydrated. If hyperglycemia and/or hypokalemia are present, insulin and/or potassium supplements should be given with the expectation that both can be discontinued soon after rehydration is complete. Plasma urea/creatinine should be monitored closely for signs of acute renal failure.

Once the patient has been rehydrated, an MRI of the brain and tests of anterior pituitary function should be performed to look for the cause and collateral defects in other hypothalamic functions. A long-term management plan to prevent or minimize recurrence of the fluid and electrolyte imbalance also should be developed. This should include a practical method that can be used to regulate fluid intake in accordance with day-to-day variations in water balance. The most effective way to do this is to prescribe desmopressin to control DI if it is present and teach the patient how to adjust daily fluid intake in accordance with day-to-day changes in body weight or serum sodium as determined by home monitoring analyzers. Prescribing a constant fluid intake is ineffective and potentially dangerous because it does not take into account the large, uncontrolled variations in insensible loss that inevitably result from changes in ambient temperature and physical activity.

EXCESS VASOPRESSIN SECRETION AND ACTION

Hyponatremia

(see also Chap. 45)

CLINICAL CHARACTERISTICS

Excessive secretion or action of AVP results in the production of decreased volumes of more highly concentrated urine. If not accompanied by a commensurate reduction in fluid intake or an increase in insensible loss, the reduction in urine output results in excess water retention with expansion and dilution of all body fluids. In some patients, excessive intake results from inappropriate thirst. If the hyponatremia develops gradually or has been present for more than a few days, it may be largely asymptomatic. However, if it develops acutely, it usually is accompanied by symptoms and signs of water intoxication that may include mild headache, confusion, anorexia, nausea, vomiting, coma, and convulsions. Severe hyponatremia may be lethal. Other clinical signs and symptoms vary greatly, depending on the pathogenesis of the defect in antidiuretic function.

ETIOLOGY

Hyponatremia and impaired urinary dilution can be caused by either a primary or a secondary defect in the regulation of AVP secretion or action. The primary forms are generally referred to as the syndrome of inappropriate antidiuresis. They have many different causes, including ectopic production of AVP by lung cancer or other neoplasms; eutopic release by various diseases or drugs; and exogenous administration of AVP, desmopressin, or large doses of oxytocin (Table 340-2). The ectopic forms result from abnormal expression of the AVP-NPII gene by primary or metastatic malignancies. The eutopic forms occur most often in patients with acute infections or strokes but have also been associated with many other neurologic diseases and injuries. In this case, the SIAD is usually self-limited and remits spontaneously within 2–3 weeks, but about 10% of cases are chronic. The mechanisms by which these diseases disrupt osmoregulation are not known. The defect in osmoregulation can take any of four distinct forms (Fig. 340-5). In one of the most common (reset osmostat), AVP secretion remains fully responsive to changes in plasma osmolarity/sodium but the threshold, or set point, of the osmoregulatory system is abnormally low. These patients differ from those with the other types of osmoregulatory defect in that they are able to maximally suppress plasma AVP and dilute their urine if their fluid intake is high enough to reduce their plasma osmolarity/sodium

to the new set point. Another, smaller subgroup (~10% of the total) has inappropriate antidiuresis without a demonstrable defect in the osmoregulation of plasma AVP (Fig. 340-5). In some of them, all young boys, the inappropriate antidiuresis has been traced to a constitutively activating mutation of the V_2 receptor gene. This unusual variant may be referred to as familial nephrogenic SIAD to distinguish it from other possible causes of the syndrome.

Table 340-2 Causes of Syndrome of Inappropriate Antidiuresis (Siad)

Neoplasms
Carcinomas
Lung
Duodenum
Pancreas
Ovary
Bladder, ureter
Other neoplasms
Thymoma
Mesothelioma
Bronchial adenoma
Carcinoid
Gangliocytoma
Ewing's sarcoma
Head trauma (closed and penetrating)
Infections
Pneumonia, bacterial or viral
Abscess, lung or brain
Cavitation (aspergillosis)
Tuberculosis, lung or brain
Meningitis, bacterial or viral
Encephalitis
AIDS
Vascular
Cerebrovascular occlusions, hemorrhage
Cavernous sinus thrombosis
Genetic
X-linked recessive
(V_2 receptor gene)
Neurologic
Guillain-Barré syndrome
Multiple sclerosis
Delirium tremens
Amyotrophic lateral sclerosis
Hydrocephalus
Psychosis
Peripheral neuropathy
Congenital malformations
Agenesis corpus callosum
Cleft lip/palate
Other midline defects
Metabolic
Acute intermittent porphyria
Pulmonary
Asthma
Pneumothorax

Positive-pressure respiration
Drugs
Vasopressin or desmopressin
Chlorpropamide
Oxytocin, high dose
Vincristine
Carbamazepine
Nicotine
Phenothiazines
Cyclophosphamide
Tricyclic antidepressants
Monoamine oxidase inhibitors
Serotonin reuptake inhibitors

The secondary forms of osmotically inappropriate antidiuresis also have multiple causes. They usually are subdivided into three types, depending on the nature of the abnormal stimulus and the state of extracellular fluid volume. Type I occurs in sodium-retaining, edema-forming states such as congestive heart failure, cirrhosis, and nephrosis. It is associated with markedly excessive retention of water and sodium that is thought to be stimulated by a large reduction in "effective" blood volume caused by low cardiac output and/or redistribution of plasma from the intravascular space to the interstitial space. Type II occurs in sodium-depleted states such as severe gastroenteritis, diuretic abuse, and mineralocorticoid deficiency. It is due to stimulation of AVP by a large reduction in blood volume and/or pressure. In both types, the increased AVP secretion appears to be due to downward resetting of the osmostat. Type III is due to nonosmotic, nonhemodynamic AVP stimuli such as nausea or isolated glucocorticoid deficiency that produce a form of euvolemic hyponatremia similar to SIAD (Table 340-3). They are differentiated because the cause of excess AVP secretion in type III can be corrected quickly and completely by treatments (antiemetics or glucocorticoids) that are not useful in SIAD.

Table 340-3 Differential Diagnosis of Hyponatremia Based on Clinical Assessment of Extracellular Fluid Volume (Ecfv)

Clinical Findings	Type I, Hypervolemic	Type II, Hypovolemic	Type III, Euvolemic	SIAD Euvolemic
History				
CHF, cirrhosis, or nephrosis	Yes	No	No	No
Salt and water loss	No	Yes	No	No
ACTH–cortisol deficiency and/or nausea and vomiting	No	No	Yes	No
Physical examination				
Generalized edema, ascites	Yes	No	No	No
Postural hypotension	Maybe	Maybe	Maybe[a]	No
Laboratory				
BUN, creatinine	High-normal	High-normal	Low-normal	Low-normal
Uric acid	High-normal	High-normal	Low-normal	Low-normal
Serum potassium	Low-normal	Low-normal[b]	Normal[c]	Normal
Serum albumin	Low-normal	High-normal	Normal	Normal
Serum cortisol	Normal-high	Normal-high[d]	Low[e]	Normal
Plasma renin activity	High	High	Low[f]	Low
Urinary sodium (meq unit of time)[g]	Low	Low[h]	High[i]	High[i]

[a] Postural hypotension may occur in secondary (ACTH-dependent) adrenal insufficiency even though extracellular fluid volume and aldosterone are usually normal.

[b] Serum potassium may be high if hypovolemia is due to aldosterone deficiency.

[c] Serum potassium may be low if vomiting causes alkalosis.

[d] Serum cortisol is low if hypovolemia is due to primary adrenal insufficiency (Addison's disease).

[e] Serum cortisol will be normal or high if the cause is nausea and vomiting rather than secondary (ACTH-dependent) adrenal insufficiency.

[f] Plasma renin activity may be high if the cause is secondary (ACTH) adrenal insufficiency.

[g] Urinary sodium should be expressed as the rate of excretion rather than the concentration. In a hyponatremic adult, an excretion rate >25 meq/d (or 25 μeq/mg of creatinine) could be considered high.

[h] The rate of urinary sodium excretion may be high if the hypovolemia is due to diuretic abuse, primary adrenal insufficiency, or other causes of renal sodium wasting.

[i] The rate of urinary sodium excretion may be low if intake is curtailed by symptoms or treatment.

Abbreviations: ACTH, adrenocorticotropic hormone; BUN, blood urea nitrogen; CHF, congestive heart failure; SIAD, syndrome of inappropriate antidiuresis.

PATHOPHYSIOLOGY

When osmotic suppression of antidiuresis is impaired for any reason, retention of water and dilution of body fluids occur only if intake exceeds the rate of obligatory and insensible and urinary losses. The excess water intake sometimes is due to an associated defect in the osmoregulation of thirst (dipsogenic) but also can be psychogenic or iatrogenic, including IV administration of hypotonic fluids.

In SIAD, the excessive retention of water expands extracellular and intracellular volume, increases glomerular filtration and atrial natriuretic hormone, suppresses plasma renin activity, and increases urinary sodium excretion. This natriuresis reduces total body sodium, and this serves to counteract the extracellular hypervolemia but aggravates the hyponatremia. The osmotically driven increase in intracellular volume results in swelling of brain cells and increases intracranial pressure; this is probably responsible for the symptoms of acute water intoxication. Within a few days, this swelling may be counteracted by inactivation or elimination of intracellular solutes, resulting in the remission of symptoms even though the hyponatremia persists. The pathophysiology of type III (euvolemic) hyponatremia is probably similar to that of SIAD.

In type I (hypervolemic) or type II (hypovolemic) hyponatremia, the antidiuretic effect of hemodynamically induced AVP release is enhanced by decreased distal delivery of glomerular filtrate that results from increased reabsorption of sodium in proximal nephrons. If the marked reduction in urine output is not associated with a commensurate reduction in water intake or an increase in insensible loss, body fluids are expanded and diluted, resulting in hyponatremia. Unlike SIAD, however, glomerular filtration is reduced and plasma renin activity and aldosterone are elevated due to either effective hypovolemia (type I) or absolute hypovolemia (type II). Thus, urinary sodium excretion is low (unless sodium reabsorption is impaired by a diuretic), and the hyponatremia is usually accompanied by hypokalemia, azotemia, and hyperuricemia. The sodium retention is an appropriate compensatory response to severe volume and sodium depletion in type II but is inappropriate and deleterious in type I since body sodium and extracellular volume are already markedly increased, as evidenced by the presence of generalized edema.

DIFFERENTIAL DIAGNOSIS

SIAD is a diagnosis of exclusion that usually can be made from the history, physical examination, and basic laboratory data. The possibility that hyponatremia is due to an osmotically driven shift of water from the intracellular space to the extracellular space can be excluded if plasma glucose is not high enough to account for the hyponatremia [serum sodium decreases ~1 meq/L for each rise in glucose of 2 mmol/L (36 mg/dL)] and/or plasma osmolarity is reduced in proportion to sodium (each decrease in serum sodium of 1 meq/L should reduce plasma osmolarity by ~2 mosmol/L). The type of hypotonic hyponatremia can then be determined by standard clinical indicators of the extracellular fluid volume (Table 340-3). If these findings are ambiguous or contradictory, measuring the rate of urinary sodium excretion or plasma renin activity may be helpful *provided* that the hyponatremia is not in the recovery phase or due to a primary defect in renal conservation of sodium, diuretic abuse, or hyporeninemic hypoaldosteronism. The latter may be suspected if serum potassium is elevated instead of low as it usually is in types I and II hyponatremia. Measurements of plasma AVP are currently of no value in differentiating among the three types of hyponatremia since the abnormalities are similar. In patients who fulfill the clinical criteria for type III (euvolemic) hyponatremia, morning plasma cortisol should also be measured to exclude secondary adrenal insufficiency. If it is normal and there is no history of nausea/vomiting, the diagnosis of SIAD is confirmed and a careful search for occult lung cancer or other common causes of the syndrome (Table 340-2) should be undertaken. If an activating mutation of the V_2 receptor gene is suspected, plasma AVP should be measured while the hyponatremia and antidiuresis are present. If it is undetectable, DNA should be collected for analysis of the V_2 receptor gene.

TREATMENT: HYPONATREMIA

The management of hyponatremia differs depending not only on the type but also on the severity and duration of symptoms. In a patient with SIAD and few symptoms, the objective is to reduce body water

gradually by restricting total fluid intake to less than the sum of urinary and insensible losses. Because the water derived from food (300–700 mL/d) usually approximates basal insensible losses in adults, total discretionary intake (all liquids) should be at least 500 mL less than urinary output. If achievable, this usually reduces body water and increases serum sodium by about 1–2% per day. If the symptoms or signs of water intoxication are more severe, the hyponatremia can be corrected more rapidly by supplementing the fluid restriction with IV infusion of hypertonic (3%) saline. This treatment also has the advantage of correcting the sodium deficiency that is partly responsible for the hyponatremia in SIAD and produces a solute diuresis that serves to remove some of the excess water. However, if plasma sodium is raised too rapidly or too much and the hyponatremia has been present for >24–48 hours, it also has the potential to produce central pontine myelinolysis, an acute, potentially fatal neurologic syndrome characterized by quadriparesis, ataxia, and abnormal extraocular movements. The risk of this complication can be minimized by observing several precautions: 3% saline should be infused at a rate ≤0.05 mL/kg body weight per min; the effect should be monitored continuously by STAT measurements of serum sodium at least once every 2 hours; and the infusion should be stopped as soon as serum sodium increases by 12 mmol/L or to 130 mmol/L, whichever comes first. Urinary output should be monitored continuously since SIAD can remit spontaneously at any time, resulting in an acute water diuresis that greatly accelerates the rate of rise in serum sodium produced by fluid restriction and 3% saline infusion.

In chronic SIAD, the hyponatremia can be corrected by treatment with demeclocycline, 150–300 mg PO tid or qid, or fludrocortisone, 0.05–0.2 mg PO bid. The effect of the demeclocycline manifests in 7–14 days and is due to production of a reversible form of nephrogenic DI. Potential side effects include phototoxicity and azotemia. The effect of fludrocortisone also requires 1–2 weeks and is partly due to increased retention of sodium and possibly inhibition of thirst. It also increases urinary potassium excretion, which may require replacement through dietary adjustments or supplements and may induce hypertension, occasionally necessitating discontinuation of the treatment.

Nonpeptide AVP antagonists that block the antidiuretic effect of AVP have been used experimentally to treat SIAD. They produce a dose-dependent increase in urinary free-water excretion, that, if combined with a modest restriction of fluid intake, reduces body water and corrects the hyponatremia. The antagonists appear to have no adverse side effects, but, like hypertonic saline, they probably carry the risk of inducing osmotic demyelinization if the hyponatremia is corrected too rapidly. One of them, a combined V_2/V_{1a} antagonist (Conivaptan), has been approved for short-term in-hospital IV treatment of SIAD and the hyponatremia of congestive heart failure. It is a substrate and inhibitor of cytochrome P450 and should not be used in conjunction with other drugs metabolized by these pathways. Other V_2 receptor antagonists are currently in phase III trials.

In type I hyponatremia, fluid restriction is also appropriate and somewhat effective if it can be maintained. However, infusion of hypertonic saline is contraindicated because it further increases total body sodium and edema and may precipitate cardiovascular decompensation. Preliminary studies with antagonists of V_2 receptors indicate that they are almost as effective in type I hyponatremia as they are in SIAD.

In type II hyponatremia, the defect in AVP secretion and water balance usually can be corrected easily and quickly by stopping the loss of sodium and water and/or replacing the deficits by mouth or IV infusion of normal or hypertonic saline. As with the treatment of other forms of hyponatremia, care must be taken to ensure that plasma sodium does not increase too rapidly. Fluid restriction and administration of AVP antagonists are contraindicated in type II as they would only aggravate the underlying volume depletion and could result in hemodynamic collapse.

In euvolemic hyponatremia due to protracted nausea and vomiting or isolated glucocorticoid deficiency (type III), all abnormalities can be corrected quickly and completely by giving an antiemetic or stress doses of hydrocortisone. As with other treatments, care must be taken to ensure that serum sodium does not rise too quickly or too far.

Global Perspectives

The incidence, clinical characteristics, etiology, pathophysiology, differential diagnosis, and treatments of fluid and electrolyte disorders in tropical and nonindustrialized countries differ in some respects from those in the United States and other industrialized parts of the world. Hyponatremia, for example, appears to be more common and is more likely to be due to infectious diseases such as cholera, shigellosis, and other diarrheal disorders. In these circumstances, hyponatremia is probably due to gastrointestinal losses of salt and water (hypovolemia type II), but other abnormalities, including undefined infectious toxins, also may contribute. The causes of DI are similar worldwide except that malaria and venoms from snake or insect bites are much more common.

FURTHER READINGS

Bichet D: Vasopressin receptor mutations in nephrogenic diabetes insipidus. Semin Nephrol 28:245, 2008[PMID: 18519085] [Full Text]

Biswas M, Davies JS: Hyponatremia in clinical practice. Postgrad Med J 83:373, 2007[PMID: 17551067] [Full Text]

Christensen JH, Rittig S: Familial neurohypophyseal diabetes insipidus—an update. Semin Nephrol 26:209, 2006[PMID: 16713494] [Full Text]

Feldman BJ et al: Nephrogenic syndrome of inappropriate antidiuresis. N Engl J Med 352:1884, 2005[PMID: 15872203] [Full Text]

Kristof RA et al: Incidence, clinical manifestations and course of water and electrolytes metabolism disturbances following transsphenoidal pituitary adenoma surgery: A prospective observational study. J Neurosurg 111:555, 2009[PMID: 19199508] [Full Text]

Liamis G et al: A review of drug induced hyponatremia. Am J Kidney Dis 52:144, 2008[PMID: 18468754] [Full Text]

Loonen AJM et al: Aquaporin 2 mutations in nephrogenic diabetes insipidus. Semin Nephrol 28:252, 2008[PMID: 18519086] [Full Text]

Robertson GL: Antidiuretic hormone: Normal and disordered function. Endocrinol Metab Clin North Am 30:671, 2001[PMID: 11571936] [Full Text]

Sitprija V: Altered fluid, electrolyte and mineral status in tropical disease with an emphasis on malaria and leptospirosis. Nature Clin Pract Nephrol 42:91, 208

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE THYROID GLAND: INTRODUCTION

The thyroid gland produces two related hormones, thyroxine (T4) and triiodothyronine (T_3) **(Fig. 341-1)**. Acting through thyroid hormone receptors α and β, these hormones play a critical role in cell differentiation during development and help maintain thermogenic and metabolic homeostasis in the adult. Autoimmune disorders of the thyroid gland can stimulate overproduction of thyroid hormones (*thyrotoxicosis*) or cause glandular destruction and hormone deficiency (*hypothyroidism*). In addition, benign nodules and various forms of thyroid cancer are relatively common and amenable to detection by physical examination.

Figure 341-1 Structures of Thyroid Hormones.

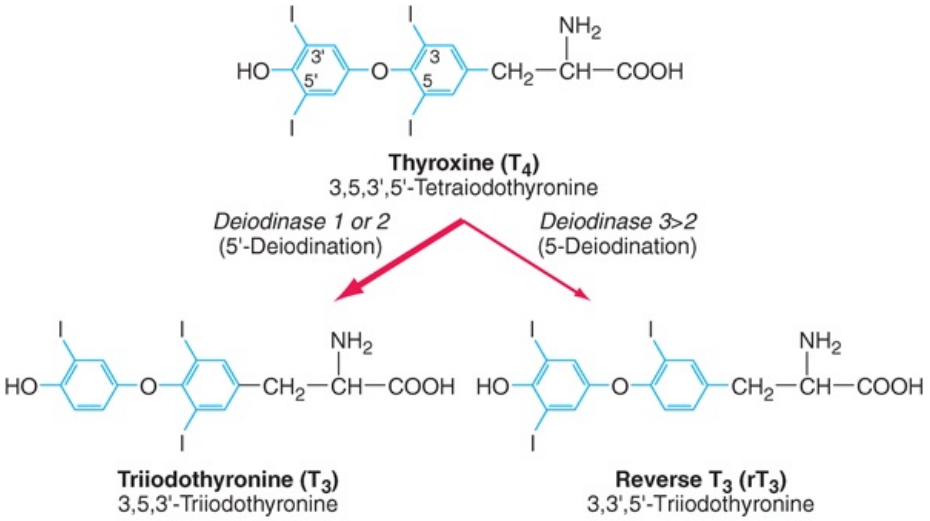

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Thyroxine (T4) contains four iodine atoms. Deiodination leads to production of the potent hormone triiodothyronine (T_3), or the inactive hormone reverse T_3.

ANATOMY AND DEVELOPMENT

The thyroid (Greek *thyreos*, shield, plus *eidos*, form) consists of two lobes connected by an isthmus. It is located anterior to the trachea between the cricoid cartilage and the suprasternal notch. The normal thyroid is 12–20 g in size, highly vascular, and soft in consistency. Four parathyroid glands, which produce parathyroid hormone (Chap. 353), are located posterior to each pole of the thyroid. The recurrent laryngeal nerves traverse the lateral borders of the thyroid gland and must be identified during thyroid surgery to avoid injury and vocal cord paralysis.

The thyroid gland develops from the floor of the primitive pharynx during the third week of gestation. The developing gland migrates along the thyroglossal duct to reach its final location in the neck. This feature accounts for the rare ectopic location of thyroid tissue at the base of the tongue (lingual thyroid) as well as the occurrence of thyroglossal duct cysts along this developmental tract. Thyroid hormone synthesis normally begins at about 11 weeks' gestation.

Neural crest derivatives from the ultimobranchial body give rise to thyroid medullary C cells that produce calcitonin, a calcium-lowering hormone. The C cells are interspersed throughout the thyroid gland, although their density is greatest in the juncture of the upper one-third and lower two-thirds of the gland. Calcitonin plays a minimal role in calcium homeostasis in humans but the C-cells are important because of their involvement in medullary thyroid cancer.

Thyroid gland development is orchestrated by the coordinated expression of several developmental transcription factors. Thyroid transcription factor (TTF)-1, TTF-2, and paired homeobox-8 (PAX-8) are expressed selectively, but not exclusively, in the thyroid gland. In combination, they dictate thyroid cell

development and the induction of thyroid-specific genes such as thyroglobulin (Tg), thyroid peroxidase (TPO), the sodium iodide symporter (Na^+/I, NIS), and the thyroid-stimulating hormone receptor (TSH-R). Mutations in these developmental transcription factors or their downstream target genes are rare causes of thyroid agenesis or dyshormonogenesis, though the causes of most forms of congenital hypothyroidism remain unknown **(Table 341-1)**. Because congenital hypothyroidism occurs in approximately 1 in 4000 newborns, neonatal screening is now performed in most industrialized countries (see below). Transplacental passage of maternal thyroid hormone occurs before the fetal thyroid gland begins to function and provides partial hormone support to a fetus with congenital hypothyroidism. Early thyroid hormone replacement in newborns with congenital hypothyroidism prevents potentially severe developmental abnormalities.

Table 341-1 Genetic Causes of Congenital Hypothyroidism

Defective Gene Protein	Inheritance	Consequences
PROP-1	Autosomal recessive	Combined pituitary hormone deficiencies with preservation of adrenocorticotropic hormone
PIT-1	Autosomal recessive Autosomal dominant	Combined deficiencies of growth hormone, prolactin, thyroid-stimulating hormone (TSH)
TSHß	Autosomal recessive	TSH deficiency
TTF-1 (TITF-1)	Autosomal dominant	Variable thyroid hypoplasia, choreoathetosis, pulmonary problems
TTF-2 (FOXE-1)	Autosomal recessive	Thyroid agenesis, choanal atresia, spiky hair
PAX-8	Autosomal dominant	Thyroid dysgenesis
TSH-receptor	Autosomal recessive	Resistance to TSH
$G_{S\alpha}$ (Albright hereditary osteodystrophy)	Autosomal dominant	Resistance to TSH
Na^+/I^- symporter	Autosomal recessive	Inability to transport iodide
THOX2	Autosomal dominant	Organification defect
Thyroid peroxidase	Autosomal recessive	Defective organification of iodide
Thyroglobulin	Autosomal recessive	Defective synthesis of thyroid hormone
Pendrin	Autosomal recessive	Pendred syndrome: sensorineural deafness and partial organification defect in thyroid
Dehalogenase 1	Autosomal recessive	Loss of iodide reutilization

The thyroid gland consists of numerous spherical follicles composed of thyroid follicular cells that surround secreted colloid, a proteinaceous fluid containing large amounts of thyroglobulin, the protein precursor of thyroid hormones **(Fig. 341-2)**. The thyroid follicular cells are polarized—the basolateral surface is apposed to the bloodstream and an apical surface faces the follicular lumen. Increased demand for thyroid hormone is regulated by thyroid-stimulating hormone (TSH), which binds to its receptor on the basolateral surface of the follicular cells, leading to Tg reabsorption from the follicular lumen, proteolysis within the cytoplasm, yielding thyroid hormones for secretion into the bloodstream.

Figure 341-2 Regulation of Thyroid Hormone Synthesis.

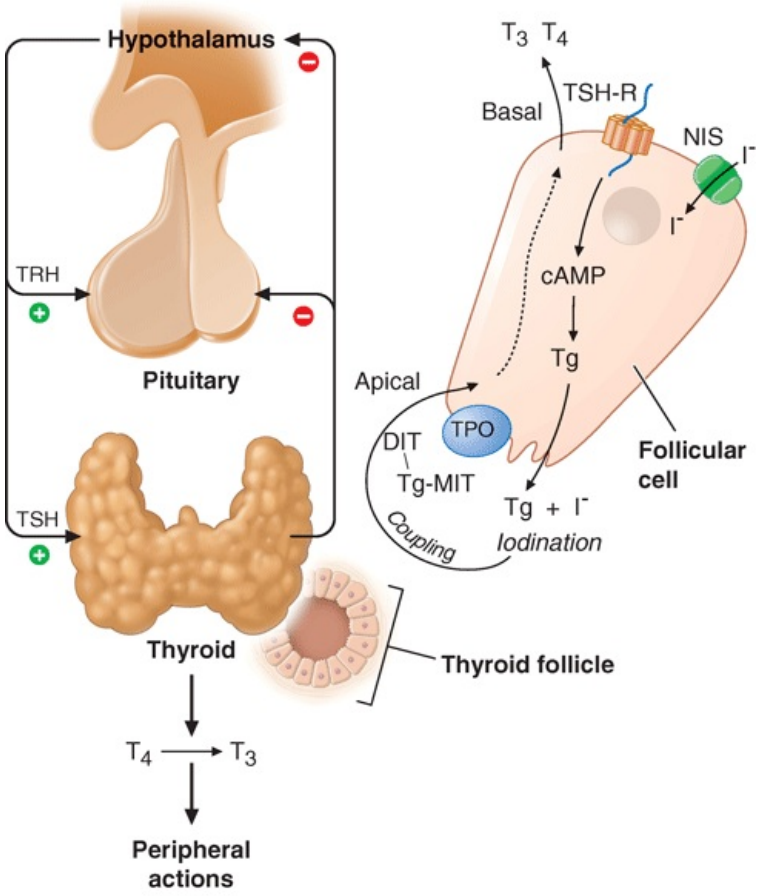

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Left. Thyroid hormones T4 and T_3 feed back to inhibit hypothalamic production of thyrotropin-releasing hormone (TRH) and pituitary production of thyroid-stimulating hormone (TSH). TSH stimulates thyroid gland production of T_4 and T_3. ***Right.*** Thyroid follicles are formed by thyroid epithelial cells surrounding proteinaceous colloid, which contains thyroglobulin. Follicular cells, which are polarized, synthesize thyroglobulin and carry out thyroid hormone biosynthesis (see text for details). TSH-R, thyroid-stimulating hormone receptor; Tg, thyroglobulin; NIS, sodium iodide symporter; TPO, thyroid peroxidase; DIT, diiodotyrosine; MIT, monoiodotyrosine.

REGULATION OF THE THYROID AXIS

TSH, secreted by the thyrotrope cells of the anterior pituitary, plays a pivotal role in control of the thyroid axis and serves as the most useful physiologic marker of thyroid hormone action. TSH is a 31-kDa hormone composed of α and β subunits; the α subunit is common to the other glycoprotein hormones [luteinizing hormone, follicle-stimulating hormone, human chorionic gonadotropin (hCG)], whereas the TSH β subunit is unique to TSH. The extent and nature of carbohydrate modification are modulated by thyrotropin-releasing hormone (TRH) stimulation and influence the biologic activity of the hormone.

The thyroid axis is a classic example of an endocrine feedback loop. Hypothalamic TRH stimulates pituitary production of TSH, which, in turn, stimulates thyroid hormone synthesis and secretion. Thyroid hormones, acting predominantly through thyroid hormone receptor β2 (TRβ2), feed back to inhibit TRH and TSH production (Fig. 341-2). The "set-point" in this axis is established by TSH. TRH is the major positive regulator of TSH synthesis and secretion. Peak TSH secretion occurs ~15 min after administration of exogenous TRH. Dopamine, glucocorticoids, and somatostatin suppress TSH but are not of major physiologic importance except when these agents are administered in pharmacologic doses. Reduced levels of thyroid hormone increase basal TSH production and enhance TRH-mediated stimulation of TSH. High thyroid hormone levels rapidly and directly suppress TSH gene expression secretion and inhibit TRH stimulation of TSH, indicating that thyroid hormones are the dominant regulator of TSH production. Like other pituitary hormones, TSH is released in a pulsatile manner and exhibits a diurnal rhythm; its highest levels occur at night. However, these TSH excursions are modest in comparison to those of other pituitary hormones, in part, because TSH has a relatively long plasma half-life (50 minutes). Consequently, single measurements of TSH are adequate for assessing its circulating level. TSH is measured using immunoradiometric assays that are highly sensitive and specific. These assays readily distinguish between normal and suppressed TSH values; thus, TSH can be used for the diagnosis of hyperthyroidism (low TSH) as well as hypothyroidism (high

TSH).

THYROID HORMONE SYNTHESIS, METABOLISM, AND ACTION

Thyroid Hormone Synthesis

Thyroid hormones are derived from Tg, a large iodinated glycoprotein. After secretion into the thyroid follicle, Tg is iodinated on tyrosine residues that are subsequently coupled via an ether linkage. Reuptake of Tg into the thyroid follicular cell allows proteolysis and the release of newly synthesized T4 and T_3.

IODINE METABOLISM AND TRANSPORT

Iodide uptake is a critical first step in thyroid hormone synthesis. Ingested iodine is bound to serum proteins, particularly albumin. Unbound iodine is excreted in the urine. The thyroid gland extracts iodine from the circulation in a highly efficient manner. For example, 10–25% of radioactive tracer (e.g., ^{123}I) is taken up by the normal thyroid gland over 24 hours; this value can rise to 70–90% in Graves' disease. Iodide uptake is mediated by NIS, which is expressed at the basolateral membrane of thyroid follicular cells. NIS is most highly expressed in the thyroid gland, but low levels are present in the salivary glands, lactating breast, and placenta. The iodide transport mechanism is highly regulated, allowing adaptation to variations in dietary supply. Low iodine levels increase the amount of NIS and stimulate uptake, whereas high iodine levels suppress NIS expression and uptake. The selective expression of NIS in the thyroid allows isotopic scanning, treatment of hyperthyroidism, and ablation of thyroid cancer with radioisotopes of iodine, without significant effects on other organs. Mutation of the *NIS* gene is a rare cause of congenital hypothyroidism, underscoring its importance in thyroid hormone synthesis. Another iodine transporter, pendrin, is located on the apical surface of thyroid cells and mediates iodine efflux into the lumen. Mutation of the *pendrin* gene causes *Pendred syndrome*, a disorder characterized by defective organification of iodine, goiter, and sensorineural deafness.

Iodine deficiency is prevalent in many mountainous regions and in central Africa, central South America, and northern Asia **(Fig. 341-3)**. Europe remains mildly iodine deficient and health surveys indicate that iodine intake has been falling in the United States and Australia. The World Health Organization (WHO) estimates that about 2 billion people are iodine-deficient, based on urinary excretion data. In areas of relative iodine deficiency, there is an increased prevalence of goiter and, when deficiency is severe, hypothyroidism and cretinism. *Cretinism* is characterized by mental and growth retardation and occurs when children who live in iodine-deficient regions are not treated with iodine or thyroid hormone to restore normal thyroid hormone levels during early life. These children are often born to mothers with iodine deficiency, and it is likely that maternal thyroid hormone deficiency worsens the condition. Concomitant selenium deficiency may also contribute to the neurologic manifestations of cretinism. Iodine supplementation of salt, bread, and other food substances has markedly reduced the prevalence of cretinism. Unfortunately, however, iodine deficiency remains the most common cause of preventable mental deficiency, often because of societal resistance to food additives or the cost of supplementation. In addition to overt cretinism, mild iodine deficiency can lead to subtle reduction of IQ. Oversupply of iodine, through supplements or foods enriched in iodine (e.g., shellfish, kelp), is associated with an increased incidence of autoimmune thyroid disease. The recommended average daily intake of iodine is 150–250 μg/d for adults, 90–120 μg/d for children, and 250 μg/d for pregnant and lactating women. Urinary iodine is >10 μg/dL in iodine-sufficient populations.

Figure 341-3 Worldwide Iodine Nutrition.

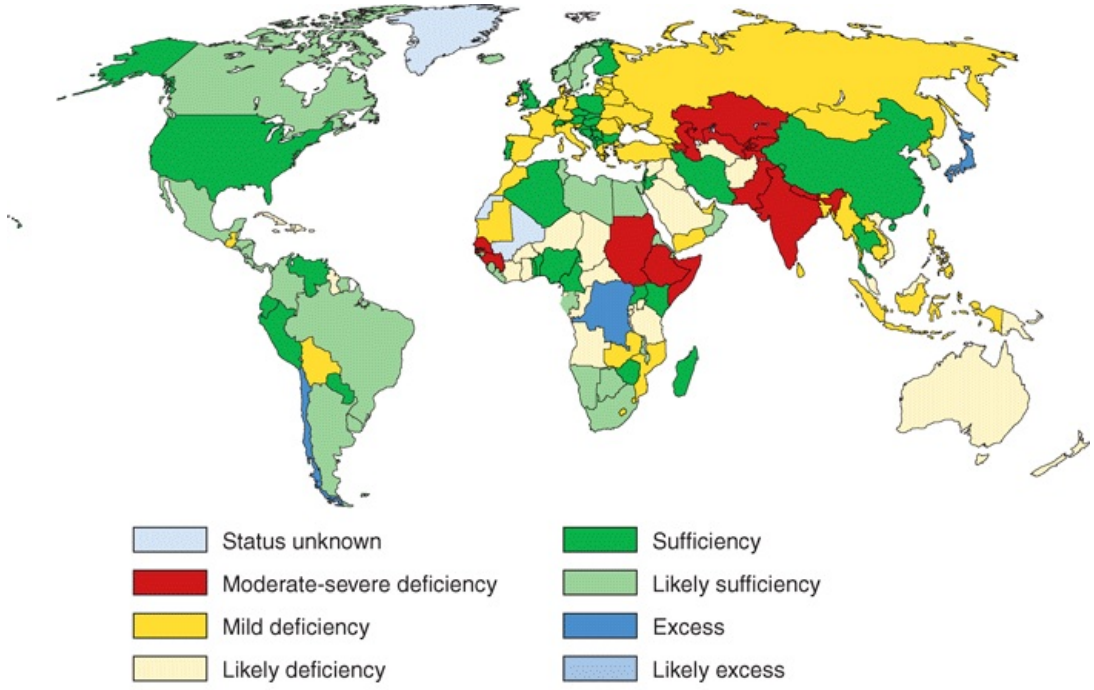

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Data are from the WHO and the International Council for the Control of Iodine Deficiency Disorders (*http://indorgs.virginia.edu/iccidd/mi/cidds.html*).

ORGANIFICATION, COUPLING, STORAGE, RELEASE

After iodide enters the thyroid, it is trapped and transported to the apical membrane of thyroid follicular cells, where it is oxidized in an organification reaction that involves TPO and hydrogen peroxide. The reactive iodine atom is added to selected tyrosyl residues within Tg, a large (660 kDa) dimeric protein that consists of 2769 amino acids. The iodotyrosines in Tg are then coupled via an ether linkage in a reaction that is also catalyzed by TPO. Either T4 or T_3 can be produced by this reaction, depending on the number of iodine atoms present in the iodotyrosines. After coupling, Tg is taken back into the thyroid cell, where it is processed in lysosomes to release T_4 and T_3. Uncoupled mono- and diiodotyrosines (MIT, DIT) are deiodinated by the enzyme dehalogenase, thereby recycling any iodide that is not converted into thyroid hormones.

Disorders of thyroid hormone synthesis are rare causes of congenital hypothyroidism. The vast majority of these disorders are due to recessive mutations in TPO or Tg, but defects have also been identified in the TSH-R, NIS, pendrin, hydrogen peroxide generation, and dehalogenase. Because of the biosynthetic defect, the gland is incapable of synthesizing adequate amounts of hormone, leading to increased TSH and a large goiter.

TSH ACTION

TSH regulates thyroid gland function through the TSH-R, a seven-transmembrane G protein–coupled receptor (GPCR). The TSH-R is coupled to the α subunit of stimulatory G protein ($G_{s}\alpha$), which activates adenylyl cyclase, leading to increased production of cyclic AMP. TSH also stimulates phosphatidylinositol turnover by activating phospholipase C. The functional role of the TSH-R is exemplified by the consequences of naturally occurring mutations. Recessive loss-of-function mutations cause thyroid hypoplasia and congenital hypothyroidism. Dominant gain-of-function mutations cause sporadic or familial hyperthyroidism that is characterized by goiter, thyroid cell hyperplasia, and autonomous function. Most of these activating mutations occur in the transmembrane domain of the receptor. They mimic the conformational changes induced by TSH binding or the interactions of thyroid-stimulating immunoglobulins (TSI) in Graves' disease. Activating TSH-R mutations also occur as somatic events, leading to clonal selection and expansion of the affected thyroid follicular cell and autonomously functioning thyroid nodules (see below).

OTHER FACTORS THAT INFLUENCE HORMONE SYNTHESIS AND RELEASE

Although TSH is the dominant hormonal regulator of thyroid gland growth and function, a variety of growth factors, most produced locally in the thyroid gland, also influence thyroid hormone synthesis. These include insulin-like growth factor I (IGF-1), epidermal growth factor, transforming growth factor ß (TGF-ß), endothelins, and various cytokines. The quantitative roles of these factors are not well understood, but they are important in selected disease states. In acromegaly, for example, increased levels of growth hormone and IGF-1 are associated with goiter and predisposition to multinodular

goiter (MNG). Certain cytokines and interleukins (ILs) produced in association with autoimmune thyroid disease induce thyroid growth, whereas others lead to apoptosis. Iodine deficiency increases thyroid blood flow and upregulates the NIS, stimulating more efficient iodine uptake. Excess iodide transiently inhibits thyroid iodide organification, a phenomenon known as the *Wolff-Chaikoff effect* . In individuals with a normal thyroid, the gland escapes from this inhibitory effect and iodide organification resumes; the suppressive action of high iodide may persist, however, in patients with underlying autoimmune thyroid disease.

Thyroid Hormone Transport and Metabolism

SERUM BINDING PROTEINS

T4 is secreted from the thyroid gland in about twentyfold excess over T_3**(Table 341-2)**. Both hormones are bound to plasma proteins, including thyroxine-binding globulin (TBG), transthyretin (TTR, formerly known as thyroxine-binding prealbumin, or TBPA), and albumin. The plasma-binding proteins increase the pool of circulating hormone, delay hormone clearance, and may modulate hormone delivery to selected tissue sites. The concentration of TBG is relatively low (1–2 mg/dL), but because of its high affinity for thyroid hormones (T_4 > T_3), it carries about 80% of the bound hormones. Albumin has relatively low affinity for thyroid hormones but has a high plasma concentration (~3.5 g/dL), and it binds up to 10% of T_4 and 30% of T_3. TTR carries about 10% of T_4 but little T_3.

Table 341-2 Characteristics of Circulating T_4 and T_3

Hormone Property	T_4	T_3
Serum concentrations		
Total hormone	8 μg/dL	0.14 μg/dL
Fraction of total hormone in the free form	0.02%	0.3%
Free (unbound) hormone	21×10^{-12} M	6×10^{-12} M
Serum half-life	7 d	0.75 d
Fraction directly from the thyroid	100%	20%
Production rate, including peripheral conversion	90 μg/d	32 μg/d
Intracellular hormone fraction	~20%	~70%
Relative metabolic potency	0.3	1
Receptor binding	10^{-10} M	10^{-11} M

When the effects of the various binding proteins are combined, approximately 99.98% of T4 and 99.7% of T_3 are protein-bound. Because T_3 is less tightly bound than T_4, the fraction of unbound T_3 is greater than unbound T_4, but there is less unbound T_3 in the circulation because it is produced in smaller amounts and cleared more rapidly than T_4. The unbound or "free" concentrations of the hormones are ~$2 \times 10^{11} M$ for T_4 and ~$6 \times 10^{12} M$ for T_3, which roughly correspond to the thyroid hormone receptor binding constants for these hormones (see below). The unbound hormone is thought to be biologically available to tissues. Nonetheless, the homeostatic mechanisms that regulate the thyroid axis are directed toward maintenance of normal concentrations of unbound hormones.

ABNORMALITIES OF THYROID HORMONE BINDING PROTEINS

A number of inherited and acquired abnormalities affect thyroid hormone binding proteins. X-linked TBG deficiency is associated with very low levels of total T4 and T_3. However, because unbound hormone levels are normal, patients are euthyroid and TSH levels are normal. It is important to recognize this disorder to avoid efforts to normalize total T_4 levels, as this leads to thyrotoxicosis and is futile because of rapid hormone clearance in the absence of TBG. TBG levels are elevated by estrogen, which increases sialylation and delays TBG clearance. Consequently, in women who are pregnant or taking estrogen-containing contraceptives, elevated TBG increases total T_4 and T_3 levels; however, unbound T_4 and T_3 levels are normal. These features are part of the explanation for why women with hypothyroidism require increased amounts of l-thyroxine replacement as TBG levels are increased by pregnancy or estrogen treatment. Mutations in TBG, TTR, and albumin may increase the binding affinity for T_4 and/or T_3 and cause disorders known as *euthyroid hyperthyroxinemia* or *familial dysalbuminemic hyperthyroxinemia* (FDH) **(Table 341-3)**. These disorders result in increased total T_4 and/or T_3, but unbound hormone levels are normal. The familial nature of the disorders, and the fact that TSH levels are normal rather than suppressed, should suggest this diagnosis. Unbound

hormone levels (ideally measured by dialysis) are normal in FDH. The diagnosis can be confirmed by using tests that measure the affinities of radiolabeled hormone binding to specific transport proteins or by performing DNA sequence analyses of the abnormal transport protein genes.

Table 341-3 Conditions Associated with Euthyroid Hyperthyroxinemia

Disorder	Cause	Transmission	Characteristics
Familial dysalbuminemic hyperthyroxinemia (FDH)	Albumin mutations, usually R218H	AD	Increased T4 Normal unbound T4 Rarely increased T_3
TBG			
Familial excess	Increased TBG production	XL	Increased total T4, T_3 Normal unbound T4, T_3
Acquired excess	Medications (estrogen), pregnancy, cirrhosis, hepatitis	Acquired	Increased total T4, T_3 Normal unbound T4, T_3
Transthyretin[a]			
Excess	Islet tumors	Acquired	Usually normal T4, T_3
Mutations	Increased affinity for T4 or T_3	AD	Increased total T4, T_3 Normal unbound T4, T_3
Medications: propranolol, ipodate, iopanoic acid, amiodarone	Decreased $T_4 \rightarrow T_3$ conversion	Acquired	Increased T4 Decreased T_3 Normal or increased TSH
Sick-euthyroid syndrome	Acute illness, especially psychiatric disorders	Acquired	Transiently increased unbound T4 Decreased TSH T4 and T_3 may also be decreased (see text)
Resistance to thyroid hormone (RTH)	Thyroid hormone receptor β mutations	AD	Increased unbound T4, T_3 Normal or increased TSH Some patients clinically thyrotoxic

[a]Also known as thyroxine-binding prealbumin, TBPA.

Abbreviations: AD, autosomal dominant; TBG, thyroxine-binding globulin; TSH, thyroid-stimulating hormone; XL, X-linked.

Certain medications, such as salicylates and salsalate, can displace thyroid hormones from circulating binding proteins. Although these drugs transiently perturb the thyroid axis by increasing free thyroid hormone levels, TSH is suppressed until a new steady state is reached, thereby restoring euthyroidism. Circulating factors associated with acute illness may also displace thyroid hormone from

binding proteins (see "Sick Euthyroid Syndrome," below).

DEIODINASES

T4 may be thought of as a precursor for the more potent T_3. T_4 is converted to T_3 by the deiodinase enzymes (Fig. 341-1). Type I deiodinase, which is located primarily in thyroid, liver, and kidneys, has a relatively low affinity for T_4. Type II deiodinase has a higher affinity for T_4 and is found primarily in the pituitary gland, brain, brown fat, and thyroid gland. Expression of type II deiodinase allows it to regulate T_3 concentrations locally, a property that may be important in the context of levothyroxine (T_4) replacement. Type II deiodinase is also regulated by thyroid hormone; hypothyroidism induces the enzyme, resulting in enhanced $T_4 \rightarrow T_3$ conversion in tissues such as brain and pituitary. $T_4 \rightarrow T_3$ conversion is impaired by fasting, systemic illness or acute trauma, oral contrast agents, and a variety of medications (e.g., propylthiouracil, propranolol, amiodarone, glucocorticoids). Type III deiodinase inactivates T_4 and T_3 and is the most important source of reverse T_3 (rT_3). Massive hemangiomas that express type III deiodinase are a rare cause of hypothyroidism in infants.

Thyroid Hormone Action

THYROID HORMONE TRANSPORT

Circulating thyroid hormones enter cells by passive diffusion and via specific transporters such as the monocarboxylate 8 (MCT8) transporter. Mutations in the *MCT8* gene have been identified in patients with X-linked psychomotor retardation and thyroid function abnormalities (low T4, high T_3, and high TSH). After entering cells, thyroid hormones act primarily through nuclear receptors, although they also have nongenomic actions through stimulating plasma membrane and mitochondrial enzymatic responses.

NUCLEAR THYROID HORMONE RECEPTORS

Thyroid hormones bind with high affinity to nuclear *thyroid hormone receptors* (TRs) α and β. Both TRα and TRβ are expressed in most tissues, but their relative expression levels vary among organs; TRα is particularly abundant in brain, kidneys, gonads, muscle, and heart, whereas TRβ expression is relatively high in the pituitary and liver. Both receptors are variably spliced to form unique isoforms. The TRβ2 isoform, which has a unique amino terminus, is selectively expressed in the hypothalamus and pituitary, where it plays a role in feedback control of the thyroid axis (see above). The TRα2 isoform contains a unique carboxy terminus that precludes thyroid hormone binding; it may function to block the action of other TR isoforms.

The TRs contain a central DNA-binding domain and a C-terminal ligand-binding domain. They bind to specific DNA sequences, termed *thyroid response elements* (TREs), in the promoter regions of target genes **(Fig. 341-4)**. The receptors bind as homodimers or, more commonly, as heterodimers with retinoic acid X receptors (RXRs) (Chap. 338). The activated receptor can either stimulate gene transcription (e.g., myosin heavy chain α) or inhibit transcription (e.g., TSH β-subunit gene), depending on the nature of the regulatory elements in the target gene.

Figure 341-4 Mechanism of Thyroid Hormone Receptor Action.

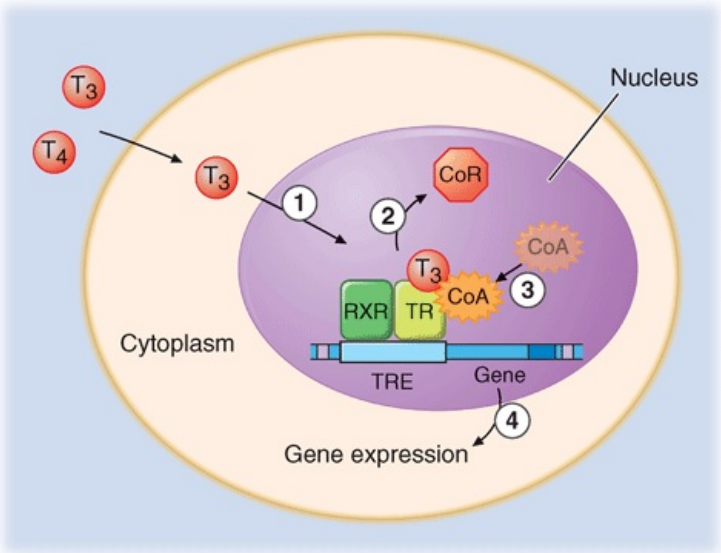

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The thyroid hormone receptor (TR) and retinoid X receptor (RXR) form heterodimers that bind specifically to thyroid hormone response elements (TRE) in the promoter regions of target genes. In the absence of hormone, TR binds co-repressor (CoR) proteins that silence gene expression. The numbers refer to a series of ordered reactions that occur in response to thyroid hormone: (1) T4 or T_3 enters the nucleus; (2) T_3 binding dissociates CoR from TR; (3) Coactivators (CoA) are recruited to the T_3-bound receptor; (4) gene expression is altered.

Thyroid hormones (T_3 and T4) bind with similar affinities to TRα and TRβ. However, structural differences in the ligand binding domains provide the potential for developing receptor-selective agonists or antagonists. T_3 is bound with 10–15 times greater affinity than T_4, which explains its increased hormonal potency. Though T_4 is produced in excess of T_3, receptors are occupied mainly by T_3, reflecting $T_4 \rightarrow T_3$ conversion by peripheral tissues, greater T_3 bioavailability in the plasma, and receptors' greater affinity for T_3. After binding to TRs, thyroid hormone induces conformational changes in the receptors that modify its interactions with accessory transcription factors. Importantly, in the absence of thyroid hormone binding, the aporeceptors bind to co-repressor proteins that inhibit gene transcription. Hormone binding dissociates the co-repressors and allows the recruitment of coactivators that enhance transcription. The discovery of TR interactions with corepressors explains the fact that TR silences gene expression in the absence of hormone binding. Consequently, hormone deficiency has a profound effect on gene expression because it causes gene repression as well as loss of hormone-induced stimulation. This concept has been corroborated by the finding that targeted deletion of the TR genes in mice has a less-pronounced phenotypic effect than hormone deficiency.

THYROID HORMONE RESISTANCE

Resistance to thyroid hormone (RTH) is an autosomal dominant disorder characterized by elevated thyroid hormone levels and inappropriately normal or elevated TSH. Individuals with RTH do not, in general, exhibit signs and symptoms that are typical of hypothyroidism because hormone resistance is partial and is compensated by increased levels of thyroid hormone. The clinical features of RTH can include goiter, attention deficit disorder, mild reduction in IQ, delayed skeletal maturation, tachycardia, and impaired metabolic responses to thyroid hormone.

RTH is caused by mutations in the TRβ receptor gene. These mutations, located in restricted regions of the ligand-binding domain, cause loss of receptor function. However, because the mutant receptors retain the capacity to dimerize with RXRs, bind to DNA, and recruit co-repressor proteins, they function as antagonists of the remaining normal TRβ and TRα receptors. This property, referred to as "dominant negative" activity, explains the autosomal dominant mode of transmission. The diagnosis is suspected when unbound thyroid hormone levels are increased without suppression of TSH. Similar hormonal abnormalities are found in other affected family members, although the TRβ mutation arises de novo in about 20% of patients. DNA sequence analysis of the TRβ gene provides a definitive diagnosis. RTH must be distinguished from other causes of euthyroid hyperthyroxinemia (e.g., FDH) and inappropriate secretion of TSH by TSH-secreting pituitary adenomas (Chap. 339). In most patients, no treatment is indicated; the importance of making the diagnosis is to avoid inappropriate treatment of mistaken hyperthyroidism and to provide genetic counseling.

Physical Examination

In addition to the examination of the thyroid itself, the physical examination should include a search for signs of abnormal thyroid function and the extrathyroidal features of ophthalmopathy and dermopathy (see below). Examination of the neck begins by inspecting the seated patient from the front and side and noting any surgical scars, obvious masses, or distended veins. The thyroid can be palpated with both hands from behind or while facing the patient, using the thumbs to palpate each lobe. It is best to use a combination of these methods, especially when nodules are small. The patient's neck should be slightly flexed to relax the neck muscles. After locating the cricoid cartilage, the isthmus can be identified and followed laterally to locate either lobe (normally, the right lobe is slightly larger than the left). By asking the patient to swallow sips of water, thyroid consistency can be better appreciated as the gland moves beneath the examiner's fingers.

Features to be noted include thyroid size, consistency, nodularity, and any tenderness or fixation. An estimate of thyroid size (normally 12–20 g) should be made, and a drawing is often the best way to record findings. However, ultrasound is the method of choice when it is important to determine thyroid size accurately. The size, location, and consistency of any nodules should also be defined. A bruit over the gland indicates increased vascularity, as occurs in hyperthyroidism. If the lower borders of the thyroid lobes are not clearly felt, a goiter may be retrosternal. Large retrosternal goiters can cause venous distention over the neck and difficulty breathing, especially when the arms are raised (Pemberton's sign). With any central mass above the thyroid, the tongue should be extended, as thyroglossal cysts then move upward. The thyroid examination is not complete without assessment for lymphadenopathy in the supraclavicular and cervical regions of the neck.

Laboratory Evaluation

MEASUREMENT OF THYROID HORMONES

The enhanced sensitivity and specificity of *TSH assays* have greatly improved laboratory assessment of thyroid function. Because TSH levels change dynamically in response to alterations of T4 and T_3, a logical approach to thyroid testing is to first determine whether TSH is suppressed, normal, or elevated. With rare exceptions (see below), a normal TSH level excludes a primary abnormality of thyroid function. This strategy depends on the use of immunochemiluminometric assays (ICMAs) for TSH that are sensitive enough to discriminate between the lower limit of the reference range and the suppressed values that occur with thyrotoxicosis. Extremely sensitive (fourth-generation) assays can detect TSH levels ≤0.004 mU/L, but, for practical purposes, assays sensitive to ≤0.1 mU/L are sufficient. The widespread availability of the TSH ICMA has rendered the TRH stimulation test obsolete, because the failure of TSH to rise after an intravenous bolus of 200–400 μg TRH has the same implications as a suppressed basal TSH measured by ICMA.

The finding of an abnormal TSH level must be followed by measurements of circulating thyroid hormone levels to confirm the diagnosis of hyperthyroidism (suppressed TSH) or hypothyroidism (elevated TSH). Radioimmunoassays are widely available for serum *total* T_4 and *total* T_3. T4 and T_3 are highly protein-bound, and numerous factors (illness, medications, genetic factors) can influence protein binding. It is useful, therefore, to measure the free, or unbound, hormone levels, which correspond to the biologically available hormone pool. Two direct methods are used to measure *unbound thyroid hormones*: (1) unbound thyroid hormone competition with radiolabeled T_4 (or an analogue) for binding to a solid-phase antibody, and (2) physical separation of the unbound hormone fraction by ultracentrifugation or equilibrium dialysis. Though early unbound hormone immunoassays suffered from artifacts, newer assays correlate well with the results of the more technically demanding and expensive physical separation methods. An indirect method to estimate unbound thyroid hormone levels is to calculate the free T_3 or free T_4 index from the total T_4 or T_3 concentration and the *thyroid hormone binding ratio* (THBR). The latter is derived from the *T_3-resin uptake test*, which determines the distribution of radiolabeled T_3 between an absorbent resin and the unoccupied thyroid hormone binding proteins in the sample. The binding of the labeled T_3 to the resin is increased when there is reduced unoccupied protein binding sites (e.g., TBG deficiency) or increased total thyroid hormone in the sample; it is decreased under the opposite circumstances. The product of THBR and total T_3 or T_4 provides the *free T_3 or T_4 index*. In effect, the index corrects for anomalous total hormone values caused by abnormalities in hormone-protein binding.

Total thyroid hormone levels are *elevated* when TBG is increased due to estrogens (pregnancy, oral contraceptives, hormone therapy, tamoxifen), and *decreased* when TBG binding is reduced (androgens, nephrotic syndrome). Genetic disorders and acute illness can also cause abnormalities in thyroid hormone binding proteins, and various drugs [phenytoin, carbamazepine, salicylates, and nonsteroidal anti-inflammatory drugs (NSAIDs)] can interfere with thyroid hormone binding. Because unbound thyroid hormone levels are normal and the patient is euthyroid in all of these circumstances, assays that measure unbound hormone are preferable to those for total thyroid hormones.

For most purposes, the unbound T4 level is sufficient to confirm thyrotoxicosis, but 2–5% of patients have only an elevated T_3 level (T_3 toxicosis). Thus, unbound T_3 levels should be measured in patients with a suppressed TSH but normal unbound T_4 levels.

There are several clinical conditions in which the use of TSH as a screening test may be misleading, particularly without simultaneous unbound T4 determinations. Any severe nonthyroidal illness can cause abnormal TSH levels (see below). Although hypothyroidism is the most common cause of an elevated TSH level, rare causes include a TSH-secreting pituitary tumor (Chap. 339), thyroid hormone resistance, and assay artifact. Conversely, a suppressed TSH level, particularly <0.1 mU/L, usually indicates thyrotoxicosis but may also be seen during the first trimester of pregnancy (due to hCG secretion), after treatment of hyperthyroidism (because TSH can remain suppressed for several months), and in response to certain medications (e.g., high doses of glucocorticoids or dopamine). Importantly, secondary hypothyroidism, caused by hypothalamic-pituitary disease, is associated with a variable (low to high-normal) TSH level, which is inappropriate for the low T_4 level. Thus, *TSH should not be used as an isolated laboratory test to assess thyroid function in patients with suspected or known pituitary disease*.

Tests for the end-organ effects of thyroid hormone excess or depletion, such as estimation of basal metabolic rate, tendon reflex relaxation rates, or serum cholesterol, are not useful as clinical determinants of thyroid function.

TESTS TO DETERMINE THE ETIOLOGY OF THYROID DYSFUNCTION

Autoimmune thyroid disease is detected most easily by measuring circulating antibodies against TPO and Tg. As antibodies to Tg alone are uncommon, it is reasonable to measure only TPO antibodies. About 5–15% of euthyroid women and up to 2% of euthyroid men have thyroid antibodies; such individuals are at increased risk of developing thyroid dysfunction. Almost all patients with autoimmune hypothyroidism, and up to 80% of those with Graves' disease, have TPO antibodies, usually at high

levels.

TSI are antibodies that stimulate the TSH-R in Graves' disease. They can be measured in bioassays or indirectly in assays for TSH-binding inhibiting immunoglobulins (TBII) that detect antibody binding to the receptor. The main use of these assays is to predict neonatal thyrotoxicosis caused by high maternal levels of TSI in the last trimester of pregnancy.

Serum Tg levels are increased in all types of thyrotoxicosis except *thyrotoxicosis factitia* caused by self-administration of thyroid hormone. Tg levels are particularly increased in thyroiditis, reflecting thyroid tissue destruction and release of Tg. The main role for Tg measurement, however, is in the follow-up of thyroid cancer patients. After total thyroidectomy and radioablation, Tg levels should be undetectable; in the absence of anti-Tg antibodies, measurable levels indicate incomplete ablation or recurrent cancer.

RADIOIODINE UPTAKE AND THYROID SCANNING

The thyroid gland selectively transports radioisotopes of iodine (^{123}I, ^{125}I, ^{131}I) and ^{99m}Tc pertechnetate, allowing thyroid imaging and quantitation of radioactive tracer fractional uptake.

Nuclear imaging of Graves' disease is characterized by an enlarged gland and increased tracer uptake that is distributed homogeneously. Toxic adenomas appear as focal areas of increased uptake, with suppressed tracer uptake in the remainder of the gland. In toxic MNG, the gland is enlarged—often with distorted architecture—and there are multiple areas of relatively increased or decreased tracer uptake. Subacute thyroiditis is associated with very low uptake because of follicular cell damage and TSH suppression. Thyrotoxicosis factitia is also associated with low uptake.

Although the use of fine-needle aspiration (FNA) biopsy has diminished the use of thyroid scans in the evaluation of solitary thyroid nodules, the functional features of thyroid nodules have some prognostic significance. So-called cold nodules, which have diminished tracer uptake, are usually benign. However, these nodules are more likely to be malignant (~5–10%) than so-called hot nodules, which are almost never malignant.

Thyroid scanning is also used in the follow-up of thyroid cancer. After thyroidectomy and ablation using ^{131}I, there is diminished radioiodine uptake in the thyroid bed, allowing the detection of metastatic thyroid cancer deposits that retain the ability to transport iodine. Whole-body scans using 111–185 MBq (3–5 mCi) ^{131}I are typically performed after thyroid hormone withdrawal to raise the TSH level or after the administration of recombinant human TSH.

THYROID ULTRASOUND

Ultrasonography is used increasingly to assist in the diagnosis of nodular thyroid disease, a reflection of the limitations of the physical examination and improvements in ultrasound technology. Using 10-MHz instruments, spatial resolution and image quality are excellent, allowing the detection of nodules and cysts >3 mm. In addition to detecting thyroid nodules, ultrasound is useful for monitoring nodule size and for the aspiration of nodules or cystic lesions. Ultrasound-guided FNA biopsy of thyroid lesions lowers the rate of inadequate sampling. Ultrasonography is also used in the evaluation of recurrent thyroid cancer, including possible spread to cervical lymph nodes.

HYPOTHYROIDISM

Iodine deficiency remains the most common cause of hypothyroidism worldwide. In areas of iodine sufficiency, autoimmune disease (Hashimoto's thyroiditis) and iatrogenic causes (treatment of hyperthyroidism) are most common **(Table 341-4)**.

Table 341-4 Causes of Hypothyroidism

Primary
Autoimmune hypothyroidism: Hashimoto's thyroiditis, atrophic thyroiditis
Iatrogenic: ^{131}I treatment, subtotal or total thyroidectomy, external irradiation of neck for lymphoma or cancer
Drugs: iodine excess (including iodine-containing contrast media and amiodarone), lithium, antithyroid drugs, *p*-aminosalicylic acid, interferon-α and other cytokines, aminoglutethimide, sunitinib
Congenital hypothyroidism: absent or ectopic thyroid gland, dyshormonogenesis, TSH-R mutation
Iodine deficiency
Infiltrative disorders: amyloidosis, sarcoidosis, hemochromatosis, scleroderma, cystinosis, Riedel's thyroiditis
Overexpression of type 3 deoiodinase in infantile hemangioma
Transient

Silent thyroiditis, including postpartum thyroiditis
Subacute thyroiditis
Withdrawal of thyroxine treatment in individuals with an intact thyroid
After ^{131}I treatment or subtotal thyroidectomy for Graves' disease
Secondary
Hypopituitarism: tumors, pituitary surgery or irradiation, infiltrative disorders, Sheehan's syndrome, trauma, genetic forms of combined pituitary hormone deficiencies
Isolated TSH deficiency or inactivity
Bexarotene treatment
Hypothalamic disease: tumors, trauma, infiltrative disorders, idiopathic

Abbreviations: TSH, thyroid-stimulating hormone; TSH-R, TSH receptor.

Congenital Hypothyroidism

PREVALENCE

Hypothyroidism occurs in about 1 in 4000 newborns. It may be transient, especially if the mother has TSH-R blocking antibodies or has received antithyroid drugs, but permanent hypothyroidism occurs in the majority. Neonatal hypothyroidism is due to thyroid gland dysgenesis in 80–85%, to inborn errors of thyroid hormone synthesis in 10–15%, and is TSH-R antibody-mediated in 5% of affected newborns. The developmental abnormalities are twice as common in girls. Mutations that cause congenital hypothyroidism are being increasingly identified, but the vast majority remain idiopathic(Table 341-1).

CLINICAL MANIFESTATIONS

The majority of infants appear normal at birth, and <10% are diagnosed based on clinical features, which include prolonged jaundice, feeding problems, hypotonia, enlarged tongue, delayed bone maturation, and umbilical hernia. Importantly, permanent neurologic damage results if treatment is delayed. Typical features of adult hypothyroidism may also be present **(Table 341-5)**. Other congenital malformations, especially cardiac, are four times more common in congenital hypothyroidism.

Table 341-5 Signs and Symptoms of Hypothyroidism (Descending Order of Frequency)

Symptoms	Signs
Tiredness, weakness	Dry coarse skin; cool peripheral extremities
Dry skin	Puffy face, hands, and feet (myxedema)
Feeling cold	Diffuse alopecia
Hair loss	Bradycardia
Difficulty concentrating and poor memory	Peripheral edema
Constipation	Delayed tendon reflex relaxation
Weight gain with poor appetite	Carpal tunnel syndrome
Dyspnea	Serous cavity effusions
Hoarse voice	
Menorrhagia (later oligomenorrhea or amenorrhea)	
Paresthesia	
Impaired hearing	

DIAGNOSIS AND TREATMENT

Because of the severe neurologic consequences of untreated congenital hypothyroidism, neonatal screening programs have been established. These are generally based on measurement of TSH or T4 levels in heel-prick blood specimens. When the diagnosis is confirmed, T_4 is instituted at a dose of 10–15 μg/kg per day, and the dose is adjusted by close monitoring of TSH levels. T_4 requirements are relatively great during the first year of life, and a high circulating T_4 level is usually needed to normalize TSH. Early treatment with T_4 results in normal IQ levels, but subtle neurodevelopmental abnormalities may occur in those with the most severe hypothyroidism at diagnosis or when treatment is delayed or suboptimal.

Autoimmune Hypothyroidism

CLASSIFICATION

Autoimmune hypothyroidism may be associated with a goiter (Hashimoto's, or *goitrous thyroiditis*) or, at the later stages of the disease, minimal residual thyroid tissue (*atrophic thyroiditis*). Because the autoimmune process gradually reduces thyroid function, there is a phase of compensation when normal thyroid hormone levels are maintained by a rise in TSH. Though some patients may have minor symptoms, this state is called *subclinical hypothyroidism*. Later, unbound T4 levels fall and TSH levels rise further; symptoms become more readily apparent at this stage (usually TSH >10 mIU/L), which is referred to as *clinical hypothyroidism* or *overt hypothyroidism*.

PREVALENCE

The mean annual incidence rate of autoimmune hypothyroidism is up to 4 per 1000 women and 1 per 1000 men. It is more common in certain populations, such as the Japanese, probably because of genetic factors and chronic exposure to a high-iodine diet. The mean age at diagnosis is 60 years, and the prevalence of overt hypothyroidism increases with age. Subclinical hypothyroidism is found in 6–8% of women (10% over the age of 60) and 3% of men. The annual risk of developing clinical hypothyroidism is about 4% when subclinical hypothyroidism is associated with positive TPO antibodies.

PATHOGENESIS

In Hashimoto's thyroiditis, there is a marked lymphocytic infiltration of the thyroid with germinal center formation, atrophy of the thyroid follicles accompanied by oxyphil metaplasia, absence of colloid, and mild to moderate fibrosis. In atrophic thyroiditis, the fibrosis is much more extensive, lymphocyte infiltration is less pronounced, and thyroid follicles are almost completely absent. Atrophic thyroiditis likely represents the end stage of Hashimoto's thyroiditis rather than a distinct disorder.

As with most autoimmune disorders, susceptibility to autoimmune hypothyroidism is determined by a combination of genetic and environmental factors, and the risk of either autoimmune hypothyroidism or Graves' disease is increased among siblings. HLA-DR polymorphisms are the best documented genetic risk factors for autoimmune hypothyroidism, especially HLA-DR3, -DR4, and -DR5 in Caucasians. A weak association also exists between polymorphisms in *CTLA-4*, a T cell–regulatory gene, and autoimmune hypothyroidism. Both of these genetic associations are shared by other autoimmune diseases, which may explain the relationship between autoimmune hypothyroidism and other autoimmune diseases, especially type 1 diabetes mellitus, Addison's disease, pernicious anemia, and vitiligo (Chap. 351). HLA-DR and *CTLA*-4 polymorphisms account for approximately half of the genetic susceptibility to autoimmune hypothyroidism. Other contributory loci remain to be identified. A gene on chromosome 21 may be responsible for the association between autoimmune hypothyroidism and Down syndrome. The female preponderance of thyroid autoimmunity is most likely due to sex steroid effects on the immune response, but an X chromosome–related genetic factor is also possible and may account for the high frequency of autoimmune hypothyroidism in Turner's syndrome. Environmental susceptibility factors are poorly defined at present. A high iodine intake may increase the risk of autoimmune hypothyroidism by immunologic effects or direct thyroid toxicity. There is no convincing evidence for a role of infection except for the congenital rubella syndrome, in which there is a high frequency of autoimmune hypothyroidism. Viral thyroiditis does not induce subsequent autoimmune thyroid disease.

The thyroid lymphocytic infiltrate in autoimmune hypothyroidism is composed of activated CD4+ and CD8+ T cells as well as B cells. Thyroid cell destruction is primarily mediated by the CD8+ cytotoxic T cells, which destroy their targets by either perforin-induced cell necrosis or granzyme B–induced apoptosis. In addition, local T cell production of cytokines, such as tumor necrosis factor (TNF), IL-1, and interferon γ (IFN-γ), may render thyroid cells more susceptible to apoptosis mediated by death receptors, such as Fas, which are activated by their respective ligands on T cells. These cytokines also impair thyroid cell function directly and induce the expression of other proinflammatory molecules by the thyroid cells themselves, such as cytokines, HLA class I and class II molecules, adhesion molecules, CD40, and nitric oxide. Administration of high concentrations of cytokines for therapeutic purposes (especially IFN-α) is associated with increased autoimmune thyroid disease, possibly through mechanisms similar to those in sporadic disease.

Antibodies to TPO and Tg are clinically useful markers of thyroid autoimmunity, but any pathogenic effect is restricted to a secondary role in amplifying an ongoing autoimmune response. TPO antibodies fix complement, and complement membrane-attack complexes are present in the thyroid in autoimmune hypothyroidism. However, transplacental passage of Tg or TPO antibodies has no effect on the fetal thyroid, which suggests that T cell–mediated injury is required to initiate autoimmune damage to the thyroid.

Up to 20% of patients with autoimmune hypothyroidism have antibodies against the TSH-R, which, in contrast to TSI, do not stimulate the receptor but prevent the binding of TSH. These TSH-R-blocking antibodies, therefore, cause hypothyroidism and, especially in Asian patients, thyroid atrophy. Their transplacental passage may induce transient neonatal hypothyroidism. Rarely, patients have a mixture of TSI and TSH-R-blocking antibodies, and thyroid function can oscillate between hyperthyroidism and

hypothyroidism as one or the other antibody becomes dominant. Predicting the course of disease in such individuals is difficult, and they require close monitoring of thyroid function. Bioassays can be used to document that TSH-R-blocking antibodies reduce the cyclic AMP–inducing effect of TSH on cultured TSH-R-expressing cells, but these assays are difficult to perform. TBII assays that measure the binding of antibodies to the receptor by competition with radiolabeled TSH do not distinguish between TSI- and TSH-R-blocking antibodies, but a positive result in a patient with spontaneous hypothyroidism is strong evidence for the presence of blocking antibodies. The use of these assays does not generally alter clinical management, although it may be useful to confirm the cause of transient neonatal hypothyroidism.

CLINICAL MANIFESTATIONS

The main clinical features of hypothyroidism are summarized in Table 341-5. The onset is usually insidious, and the patient may become aware of symptoms only when euthyroidism is restored. Patients with Hashimoto's thyroiditis may present because of goiter rather than symptoms of hypothyroidism. The goiter may not be large, but it is usually irregular and firm in consistency. It is often possible to palpate a pyramidal lobe, normally a vestigial remnant of the thyroglossal duct. Rarely is uncomplicated Hashimoto's thyroiditis associated with pain.

Patients with atrophic thyroiditis or the late stage of Hashimoto's thyroiditis present with symptoms and signs of hypothyroidism. The skin is dry, and there is decreased sweating, thinning of the epidermis, and hyperkeratosis of the stratum corneum. Increased dermal glycosaminoglycan content traps water, giving rise to skin thickening without pitting (*myxedema*). Typical features include a puffy face with edematous eyelids and nonpitting pretibial edema **(Fig. 341-5)**. There is pallor, often with a yellow tinge to the skin due to carotene accumulation. Nail growth is retarded, and hair is dry, brittle, difficult to manage, and falls out easily. In addition to diffuse alopecia, there is thinning of the outer third of the eyebrows, although this is not a specific sign of hypothyroidism.

Figure 341-5 Facial Appearance in Hypothyroidism.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Note puffy eyes and thickened skin.

Other common features include constipation and weight gain (despite a poor appetite). In contrast to popular perception, the weight gain is usually modest and due mainly to fluid retention in the myxedematous tissues. Libido is decreased in both sexes, and there may be oligomenorrhea or amenorrhea in long-standing disease, but menorrhagia is also common. Fertility is reduced, and the incidence of miscarriage is increased. Prolactin levels are often modestly increased (Chap. 339) and may contribute to alterations in libido and fertility and cause galactorrhea.

Myocardial contractility and pulse rate are reduced, leading to a reduced stroke volume and bradycardia. Increased peripheral resistance may be accompanied by hypertension, particularly diastolic. Blood flow is diverted from the skin, producing cool extremities. Pericardial effusions occur in

up to 30% of patients but rarely compromise cardiac function. Though alterations in myosin heavy chain isoform expression have been documented, cardiomyopathy is unusual. Fluid may also accumulate in other serous cavities and in the middle ear, giving rise to conductive deafness. Pulmonary function is generally normal, but dyspnea may be caused by pleural effusion, impaired respiratory muscle function, diminished ventilatory drive, or sleep apnea.

Carpal tunnel and other entrapment syndromes are common, as is impairment of muscle function with stiffness, cramps, and pain. On examination, there may be slow relaxation of tendon reflexes and pseudomyotonia. Memory and concentration are impaired. Experimentally, PET scans examining glucose metabolism in hypothyroid subjects show lower regional activity in the amygdala, hippocampus, and perigenual anterior cingulated cortex, among other regions, and this activity corrects after thyroxine replacement. Rare neurologic problems include reversible cerebellar ataxia, dementia, psychosis, and myxedema coma. *Hashimoto's encephalopathy* has been defined as a steroid-responsive syndrome associated with TPO antibodies, myoclonus, and slow-wave activity on electroencephalography, but the relationship with thyroid autoimmunity or hypothyroidism is not established. The hoarse voice and occasionally clumsy speech of hypothyroidism reflect fluid accumulation in the vocal cords and tongue.

The features described above are the consequence of thyroid hormone deficiency. However, autoimmune hypothyroidism may be associated with signs or symptoms of other autoimmune diseases, particularly vitiligo, pernicious anemia, Addison's disease, alopecia areata, and type 1 diabetes mellitus. Less-common associations include celiac disease, dermatitis herpetiformis, chronic active hepatitis, rheumatoid arthritis, systemic lupus erythematosus (SLE), myasthenia gravis, and Sjögren's syndrome. Thyroid-associated ophthalmopathy, which usually occurs in Graves' disease (see below), occurs in about 5% of patients with autoimmune hypothyroidism.

Autoimmune hypothyroidism is uncommon in children and usually presents with slow growth and delayed facial maturation. The appearance of permanent teeth is also delayed. Myopathy, with muscle swelling, is more common in children than in adults. In most cases, puberty is delayed, but precocious puberty sometimes occurs. There may be intellectual impairment if the onset is before 3 years and the hormone deficiency is severe.

LABORATORY EVALUATION

A summary of the investigations used to determine the existence and cause of hypothyroidism is provided in **Fig. 341-6**. A normal TSH level excludes primary (but not secondary) hypothyroidism. If the TSH is elevated, an unbound T4 level is needed to confirm the presence of clinical hypothyroidism, but T_4 is inferior to TSH when used as a screening test, because it will not detect subclinical hypothyroidism. Circulating unbound T_3 levels are normal in about 25% of patients, reflecting adaptive deiodinase responses to hypothyroidism. T_3 measurements are, therefore, not indicated.

Figure 341-6 Evaluation of Hypothyroidism.

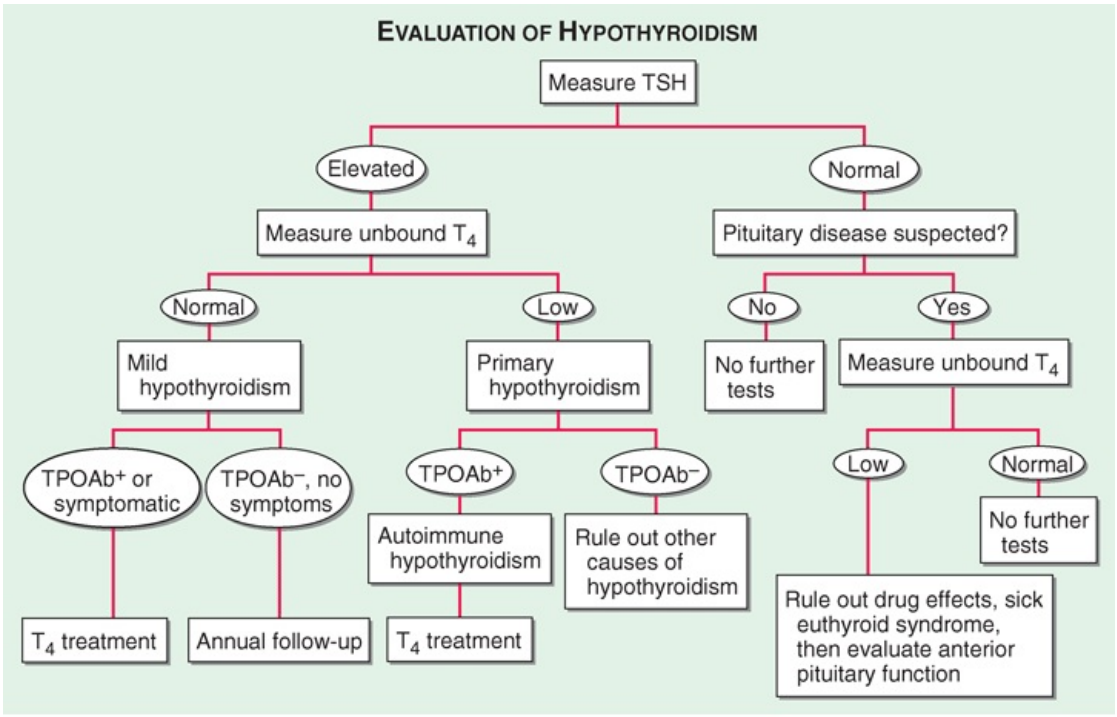

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

$TPOAb^+$, thyroid peroxidase antibodies present; $TPOAb^-$, thyroid peroxidase antibodies not present; TSH, thyroid-stimulating hormone.

Once clinical or subclinical hypothyroidism is confirmed, the etiology is usually easily established by demonstrating the presence of TPO antibodies, which are present in >90% of patients with autoimmune hypothyroidism. TBII can be found in 10–20% of patients, but these determinations are not needed routinely. If there is any doubt about the cause of a goiter associated with hypothyroidism, FNA biopsy can be used to confirm the presence of autoimmune thyroiditis. Other abnormal laboratory findings in hypothyroidism may include increased creatine phosphokinase, elevated cholesterol and triglycerides, and anemia (usually normocytic or macrocytic). Except when accompanied by iron deficiency, the anemia and other abnormalities gradually resolve with thyroxine replacement.

DIFFERENTIAL DIAGNOSIS

An asymmetric goiter in Hashimoto's thyroiditis may be confused with a multinodular goiter or thyroid carcinoma, in which thyroid antibodies may also be present. Ultrasound can be used to show the presence of a solitary lesion or a multinodular goiter rather than the heterogeneous thyroid enlargement typical of Hashimoto's thyroiditis. FNA biopsy is useful in the investigation of focal nodules. Other causes of hypothyroidism are discussed below and in Table 341-14, but rarely cause diagnostic confusion.

Other Causes of Hypothyroidism

Iatrogenic hypothyroidism is a common cause of hypothyroidism and can often be detected by screening before symptoms develop. In the first 3–4 months after radioiodine treatment, transient hypothyroidism may occur due to reversible radiation damage. Low-dose thyroxine treatment can be withdrawn if recovery occurs. Because TSH levels are suppressed by hyperthyroidism, unbound T4 levels are a better measure of thyroid function than TSH in the months following radioiodine treatment. Mild hypothyroidism after subtotal thyroidectomy may also resolve after several months, as the gland remnant is stimulated by increased TSH levels.

Iodine deficiency is responsible for endemic goiter and cretinism but is an uncommon cause of adult hypothyroidism unless the iodine intake is very low or there are complicating factors, such as the consumption of thiocyanates in cassava or selenium deficiency. Though hypothyroidism due to iodine deficiency can be treated with thyroxine, public health measures to improve iodine intake should be advocated to eliminate this problem. Iodized salt or bread or a single bolus of oral or intramuscular iodized oil have all been used successfully.

Paradoxically, chronic iodine excess can also induce goiter and hypothyroidism. The intracellular events that account for this effect are unclear, but individuals with autoimmune thyroiditis are especially susceptible. Iodine excess is responsible for the hypothyroidism that occurs in up to 13% of patients treated with amiodarone (see below). Other drugs, particularly lithium, may also cause hypothyroidism. Transient hypothyroidism caused by thyroiditis is discussed below.

Secondary hypothyroidism is usually diagnosed in the context of other anterior pituitary hormone deficiencies; isolated TSH deficiency is very rare (Chap. 339). TSH levels may be low, normal, or even slightly increased in secondary hypothyroidism; the latter is due to secretion of immunoactive but bioinactive forms of TSH. The diagnosis is confirmed by detecting a low unbound T4 level. The goal of treatment is to maintain T_4 levels in the upper half of the reference range, because TSH levels cannot be used to monitor therapy.

Treatment: Hypothyroidism

CLINICAL HYPOTHYROIDISM

If there is no residual thyroid function, the daily replacement dose of levothyroxine is usually 1.6 μg/kg body weight (typically 100–150 μg). In many patients, however, lower doses suffice until residual thyroid tissue is destroyed. In patients who develop hypothyroidism after the treatment of Graves' disease, there is often underlying autonomous function, necessitating lower replacement doses (typically 75–125 μg/d).

Adult patients under 60 without evidence of heart disease may be started on 50–100 μg levothyroxine (T4) daily. The dose is adjusted on the basis of TSH levels, with the goal of treatment being a normal TSH, ideally in the lower half of the reference range. TSH responses are gradual and should be measured about two months after instituting treatment or after any subsequent change in levothyroxine dosage. The clinical effects of levothyroxine replacement are slow to appear. Patients may not experience full relief from symptoms until 3–6 months after normal TSH levels are restored. Adjustment of levothyroxine dosage is made in 12.5- or 25-μg increments if the TSH is high; decrements of the same magnitude should be made if the TSH is suppressed. Patients with a suppressed TSH of any cause, including T_4 overtreatment, have an increased risk of atrial fibrillation and reduced bone density.

Although dessicated animal thyroid preparations (thyroid extract USP) are available, they are not recommended because the ratio of T_3 to T4 is nonphysiologic. The use of levothyroxine combined with liothyronine (triiodothyronine, T_3) has been investigated, but benefit has not been confirmed in prospective studies. There is no place for liothyronine alone as long-term replacement, because the short half-life necessitates three or four daily doses and is associated with fluctuating T_3 levels.

Once full replacement is achieved and TSH levels are stable, follow-up measurement of TSH is recommended at annual intervals and may be extended to every 2–3 years if a normal TSH is maintained over several years. It is important to ensure ongoing adherence, however, as patients do not feel any symptomatic difference after missing a few doses of levothyroxine, and this sometimes leads to self-discontinuation.

In patients of normal body weight who are taking ≳200 μg of levothyroxine per day, an elevated TSH level is often a sign of poor adherence to treatment. This is also the likely explanation for fluctuating TSH levels, despite a constant levothyroxine dosage. Such patients often have normal or high unbound T4 levels, despite an elevated TSH, because they remember to take medication for a few days before testing; this is sufficient to normalize T_4, but not TSH levels. It is important to consider variable adherence, because this pattern of thyroid function tests is otherwise suggestive of disorders associated with inappropriate TSH secretion (Table 341-3). Because T_4 has a long half-life (7 days), patients who miss a dose can be advised to take two doses of the skipped tablets at once. Other causes of increased levothyroxine requirements must be excluded, particularly malabsorption (e.g., celiac disease, small-bowel surgery), estrogen therapy, and drugs that interfere with T_4 absorption or clearance such as cholestyramine, ferrous sulfate, calcium supplements, lovastatin, aluminum hydroxide, rifampicin, amiodarone, carbamazepine, and phenytoin.

SUBCLINICAL HYPOTHYROIDISM

By definition, subclinical hypothyroidism refers to biochemical evidence of thyroid hormone deficiency in patients who have few or no apparent clinical features of hypothyroidism. There are no universally accepted recommendations for the management of subclinical hypothyroidism, but the most recently published guidelines do not recommend routine treatment when TSH levels are below 10 mU/L. It is important to confirm that any elevation of TSH is sustained over a 3-month period before treatment is given. As long as excessive treatment is avoided, there is no risk in correcting a slightly increased TSH. Moreover, there is a risk that patients will progress to overt hypothyroidism, particularly when the TSH level is elevated and TPO antibodies are present. Treatment is administered by starting with a low dose of levothyroxine (25–50 μg/d) with the goal of normalizing TSH. If thyroxine is not given, thyroid function should be evaluated annually.

SPECIAL TREATMENT CONSIDERATIONS

Rarely, levothyroxine replacement is associated with pseudotumor cerebri in children. Presentation appears to be idiosyncratic and occurs months after treatment has begun. Women with a history or high risk of hypothyroidism should ensure that they are euthyroid prior to conception and during early pregnancy as maternal hypothyroidism may adversely affect fetal neural development and cause

preterm delivery. The presence of thyroid autoantibodies alone, in a euthyroid patient, is also associated with preterm delivery, and outcome may be improved by levothyroxine treatment. Thyroid function should be evaluated immediately after pregnancy is confirmed and at the beginning of the second and third trimesters. The dose of levothyroxine may need to be increased by ≳50% during pregnancy and returned to previous levels after delivery. Elderly patients may require 20% less thyroxine than younger patients. In the elderly, especially patients with known coronary artery disease, the starting dose of levothyroxine is 12.5–25 μg/d with similar increments every 2–3 months until TSH is normalized. In some patients, it may be impossible to achieve full replacement despite optimal antianginal treatment. *Emergency surgery* is generally safe in patients with untreated hypothyroidism, although routine surgery in a hypothyroid patient should be deferred until euthyroidism is achieved.

Myxedema coma still has a high mortality rate, despite intensive treatment. Clinical manifestations include reduced level of consciousness, sometimes associated with seizures, as well as the other features of hypothyroidism (Table 341-5). Hypothermia can reach 23°C (74°F). There may be a history of treated hypothyroidism with poor compliance, or the patient may be previously undiagnosed. Myxedema coma almost always occurs in the elderly and is usually precipitated by factors that impair respiration, such as drugs (especially sedatives, anesthetics, antidepressants), pneumonia, congestive heart failure, myocardial infarction, gastrointestinal bleeding, or cerebrovascular accidents. Sepsis should also be suspected. Exposure to cold may also be a risk factor. Hypoventilation, leading to hypoxia and hypercapnia, plays a major role in pathogenesis; hypoglycemia and dilutional hyponatremia also contribute to the development of myxedema coma.

Levothyroxine can initially be administered as a single IV bolus of 500 μg, which serves as a loading dose. Although further levothyroxine is not strictly necessary for several days, it is usually continued at a dose of 50–100 μg/d. If suitable IV preparation is not available, the same initial dose of levothyroxine can be given by nasogastric tube (though absorption may be impaired in myxedema). An alternative is to give liothyronine (T_3) intravenously or via nasogastric tube, in doses ranging from 10 to 25 μg every 8–12 h. This treatment has been advocated because T4 → T_3 conversion is impaired in myxedema coma. However, excess liothyronine has the potential to provoke arrhythmias. Another option is to combine levothyroxine (200 μg) and liothyronine (25 μg) as a single, initial IV bolus followed by daily treatment with levothyroxine (50–100 μg/d) and liothyronine (10 μg every 8 h).

Supportive therapy should be provided to correct any associated metabolic disturbances. External warming is indicated only if the temperature is <30°C, as it can result in cardiovascular collapse (Chap. 19). Space blankets should be used to prevent further heat loss. Parenteral hydrocortisone (50 mg every 6 h) should be administered, because there is impaired adrenal reserve in profound hypothyroidism. Any precipitating factors should be treated, including the early use of broad-spectrum antibiotics, pending the exclusion of infection. Ventilatory support with regular blood gas analysis is usually needed during the first 48 hours. Hypertonic saline or IV glucose may be needed if there is severe hyponatremia or hypoglycemia; hypotonic IV fluids should be avoided because they may exacerbate water retention secondary to reduced renal perfusion and inappropriate vasopressin secretion. The metabolism of most medications is impaired, and sedatives should be avoided if possible or used in reduced doses. Medication blood levels should be monitored, when available, to guide dosage.

THYROTOXICOSIS

Thyrotoxicosis is defined as the state of thyroid hormone excess and is not synonymous with *hyperthyroidism*, which is the result of excessive thyroid function. However, the major etiologies of thyrotoxicosis are hyperthyroidism caused by Graves' disease, toxic MNG, and toxic adenomas. Other causes are listed in **Table 341-6**.

Table 341-6 Causes of Thyrotoxicosis

Primary hyperthyroidism
Graves' disease Toxic multinodular goiter
Toxic adenoma
Functioning thyroid carcinoma metastases
Activating mutation of the TSH receptor
Activating mutation of $G_{s\alpha}$ (McCune-Albright syndrome)
Struma ovarii
Drugs: iodine excess (Jod-Basedow phenomenon)
Thyrotoxicosis without hyperthyroidism
Subacute thyroiditis

Silent thyroiditis
Other causes of thyroid destruction: amiodarone, radiation, infarction of adenoma
Ingestion of excess thyroid hormone (thyrotoxicosis factitia) or thyroid tissue
Secondary hyperthyroidism TSH-secreting pituitary adenoma
Thyroid hormone resistance syndrome: occasional patients may have features of thyrotoxicosis
Chorionic gonadotropin-secreting tumors[a]
Gestational thyrotoxicosis[a]

[a]Circulating TSH levels are low in these forms of secondary hyperthyroidism.

Abbreviations: TSH, thyroid-stimulating hormone.

Graves' Disease

EPIDEMIOLOGY

Graves' disease accounts for 60–80% of thyrotoxicosis. The prevalence varies among populations, reflecting genetic factors and iodine intake (high iodine intake is associated with an increased prevalence of Graves' disease). Graves' disease occurs in up to 2% of women but is one-tenth as frequent in men. The disorder rarely begins before adolescence and typically occurs between 20 and 50 years of age; it also occurs in the elderly.

PATHOGENESIS

As in autoimmune hypothyroidism, a combination of environmental and genetic factors, including polymorphisms in HLA-DR, *CTLA-4*, *CD25*, *PTPN22* (a T cell regulatory gene) and *TSH-R*, contribute to Graves' disease susceptibility. The concordance for Graves' disease in monozygotic twins is 20–30%, compared to <5% in dizygotic twins. Indirect evidence suggests that stress is an important environmental factor, presumably operating through neuroendocrine effects on the immune system. Smoking is a minor risk factor for Graves' disease and a major risk factor for the development of ophthalmopathy. Sudden increases in iodine intake may precipitate Graves' disease, and there is a threefold increase in the occurrence of Graves' disease in the postpartum period. Graves' disease may occur during the immune reconstitution phase after highly active antiretroviral therapy (HAART) or alemtuzumab treatment.

The hyperthyroidism of Graves' disease is caused by TSI that are synthesized in the thyroid gland as well as in bone marrow and lymph nodes. Such antibodies can be detected by bioassays or by using the more widely available TBII assays. The presence of TBII in a patient with thyrotoxicosis implies the existence of TSI, and these assays are useful in monitoring pregnant Graves' patients in whom high levels of TSI can cross the placenta and cause neonatal thyrotoxicosis. Other thyroid autoimmune responses, similar to those in autoimmune hypothyroidism (see above), occur concurrently in patients with Graves' disease. In particular, TPO antibodies occur in up to 80% of cases and serve as a readily measurable marker of autoimmunity. Because the coexisting thyroiditis can also affect thyroid function, there is no direct correlation between the level of TSI and thyroid hormone levels in Graves' disease. In the long term, spontaneous autoimmune hypothyroidism may develop in up to 15% of patients with Graves' disease.

Cytokines appear to play a major role in thyroid-associated ophthalmopathy. There is infiltration of the extraocular muscles by activated T cells; the release of cytokines such as IFN-γ, TNF, and IL-1 results in fibroblast activation and increased synthesis of glycosaminoglycans that trap water, thereby leading to characteristic muscle swelling. Late in the disease, there is irreversible fibrosis of the muscles. Orbital fibroblasts may be particularly sensitive to cytokines, perhaps explaining the anatomic localization of the immune response. Though the pathogenesis of thyroid-associated ophthalmopathy remains unclear, there is mounting evidence that the TSH-R may be a shared autoantigen that is expressed in the orbit; this would explain the close association with autoimmune thyroid disease. Increased fat is an additional cause of retrobulbar tissue expansion. The increase in intraorbital pressure can lead to proptosis, diplopia, and optic neuropathy

CLINICAL MANIFESTATIONS

Signs and symptoms include features that are common to any cause of thyrotoxicosis **(Table 341-7)** as well as those specific for Graves' disease. The clinical presentation depends on the severity of thyrotoxicosis, the duration of disease, individual susceptibility to excess thyroid hormone, and the patient's age. In the elderly, features of thyrotoxicosis may be subtle or masked, and patients may present mainly with fatigue and weight loss, a condition known as *apathetic thyrotoxicosis*.

Table 341-7 Signs and Symptoms of Thyrotoxicosis (Descending Order of Frequency)

Symptoms	Signs[a]
Hyperactivity, irritability, dysphoria	Tachycardia; atrial fibrillation in the elderly
Heat intolerance and sweating	Tremor
Palpitations	Goiter
Fatigue and weakness	Warm, moist skin
Weight loss with increased appetite	Muscle weakness, proximal myopathy
Diarrhea	Lid retraction or lag
Polyuria	Gynecomastia
Oligomenorrhea, loss of libido	

[a]Excludes the signs of ophthalmopathy and dermopathy specific for Graves' disease.

Thyrotoxicosis may cause unexplained weight loss, despite an enhanced appetite, due to the increased metabolic rate. Weight gain occurs in 5% of patients, however, because of increased food intake. Other prominent features include hyperactivity, nervousness, and irritability, ultimately leading to a sense of easy fatigability in some patients. Insomnia and impaired concentration are common; apathetic thyrotoxicosis may be mistaken for depression in the elderly. Fine tremor is a frequent finding, best elicited by having patients stretch out their fingers while feeling the fingertips with the palm. Common neurologic manifestations include hyperreflexia, muscle wasting, and proximal myopathy without fasciculation. Chorea is rare. Thyrotoxicosis is sometimes associated with a form of hypokalemic periodic paralysis; this disorder is particularly common in Asian males with thyrotoxicosis, but it occurs in other ethnic groups as well.

The most common cardiovascular manifestation is sinus tachycardia, often associated with palpitations, occasionally caused by supraventricular tachycardia. The high cardiac output produces a bounding pulse, widened pulse pressure, and an aortic systolic murmur and can lead to worsening of angina or heart failure in the elderly or those with preexisting heart disease. Atrial fibrillation is more common in patients >50 years of age. Treatment of the thyrotoxic state alone converts atrial fibrillation to normal sinus rhythm in about half of patients, suggesting the existence of an underlying cardiac problem in the remainder.

The skin is usually warm and moist, and the patient may complain of sweating and heat intolerance, particularly during warm weather. Palmar erythema, onycholysis, and, less commonly, pruritus, urticaria, and diffuse hyperpigmentation may be evident. Hair texture may become fine, and a diffuse alopecia occurs in up to 40% of patients, persisting for months after restoration of euthyroidism. Gastrointestinal transit time is decreased, leading to increased stool frequency, often with diarrhea and occasionally mild steatorrhea. Women frequently experience oligomenorrhea or amenorrhea; in men, there may be impaired sexual function and, rarely, gynecomastia. The direct effect of thyroid hormones on bone resorption leads to osteopenia in long-standing thyrotoxicosis; mild hypercalcemia occurs in up to 20% of patients, but hypercalciuria is more common. There is a small increase in fracture rate in patients with a previous history of thyrotoxicosis.

In Graves' disease, the thyroid is usually diffusely enlarged to two to three times its normal size. The consistency is firm, but less so than in MNG. There may be a thrill or bruit due to the increased vascularity of the gland and the hyperdynamic circulation.

Lid retraction, causing a staring appearance, can occur in any form of thyrotoxicosis and is the result of sympathetic overactivity. However, Graves' disease is associated with specific eye signs that comprise *Graves' ophthalmopathy* **(Fig. 341-7*A*)**. This condition is also called *thyroid-associated ophthalmopathy*, as it occurs in the absence of Graves' disease in 10% of patients. Most of these individuals have autoimmune hypothyroidism or thyroid antibodies. The onset of Graves' ophthalmopathy occurs within the year before or after the diagnosis of thyrotoxicosis in 75% of patients but can sometimes precede or follow thyrotoxicosis by several years, accounting for some cases of euthyroid ophthalmopathy.

Figure 341-7 Features of Graves' Disease.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Ophthalmopathy in Graves' disease; lid retraction, periorbital edema, conjunctival injection, and proptosis are marked. ***B.*** Thyroid dermopathy over the lateral aspects of the shins. ***C.*** Thyroid acropachy.

Some patients with Graves' disease have little clinical evidence of ophthalmopathy. However, the enlarged extraocular muscles typical of the disease, and other subtle features, can be detected in almost all patients when investigated by ultrasound or CT imaging of the orbits. Unilateral signs are found in up to 10% of patients. The earliest manifestations of ophthalmopathy are usually a sensation of grittiness, eye discomfort, and excess tearing. About one-third of patients have proptosis, best detected by visualization of the sclera between the lower border of the iris and the lower eyelid, with the eyes in the primary position. Proptosis can be measured using an exophthalmometer. In severe cases, proptosis may cause corneal exposure and damage, especially if the lids fail to close during sleep. Periorbital edema, scleral injection, and chemosis are also frequent. In 5–10% of patients, the muscle swelling is so severe that diplopia results, typically, but not exclusively, when the patient looks up and laterally. The most serious manifestation is compression of the optic nerve at the apex of the orbit, leading to papilledema; peripheral field defects; and, if left untreated, permanent loss of vision.

Many scoring systems have been used to gauge the extent and activity of the orbital changes in Graves' disease. The "NO SPECS" scheme is an acronym derived from the following eye changes:

0 = **N**o signs or symptoms

1 = **O**nly signs (lid retraction or lag), no symptoms

2 = **S**oft-tissue involvement (periorbital edema)

3 = **P**roptosis (>22 mm)

4 = **E**xtraocular-muscle involvement (diplopia)

5 = **C**orneal involvement

6 = **S**ight loss

Although useful as a mnemonic, the NO SPECS scheme is inadequate to describe the eye disease fully, and patients do not necessarily progress from one class to another. When Graves' eye disease is active and severe, referral to an ophthalmologist is indicated and objective measurements are needed, such as lid-fissure width; corneal staining with fluorescein; and evaluation of extraocular muscle function (e.g., Hess chart), intraocular pressure and visual fields, acuity, and color vision.

Thyroid dermopathy occurs in <5% of patients with Graves' disease **(Fig. 341-7*B*)**, almost always in the presence of moderate or severe ophthalmopathy. Although most frequent over the anterior and lateral aspects of the lower leg (hence the term *pretibial myxedema*), skin changes can occur at other sites, particularly after trauma. The typical lesion is a noninflamed, indurated plaque with a deep pink or purple color and an "orange skin" appearance. Nodular involvement can occur, and the condition can rarely extend over the whole lower leg and foot, mimicking elephantiasis. *Thyroid acropachy* refers to a form of clubbing found in <1% of patients with Graves' disease **(Fig. 341-7*C*)**. It is so strongly associated with thyroid dermopathy that an alternative cause of clubbing should be sought in a Graves' patient without coincident skin and orbital involvement.

LABORATORY EVALUATION

Investigations used to determine the existence and cause of thyrotoxicosis are summarized in **Fig. 341-8**. In Graves' disease, the TSH level is suppressed and total and unbound thyroid hormone levels are increased. In 2–5% of patients (and more in areas of borderline iodine intake), only T_3 is increased (T_3 toxicosis). The converse state of T4 toxicosis, with elevated total and unbound T_4 and

normal T_3 levels, is occasionally seen when hyperthyroidism is induced by excess iodine, providing surplus substrate for thyroid hormone synthesis. Measurement of TPO antibodies or TBII may be useful if the diagnosis is unclear clinically but is not needed routinely. Associated abnormalities that may cause diagnostic confusion in thyrotoxicosis include elevation of bilirubin, liver enzymes, and ferritin. Microcytic anemia and thrombocytopenia may occur.

Figure 341-8 Evaluation of Thyrotoxicosis.

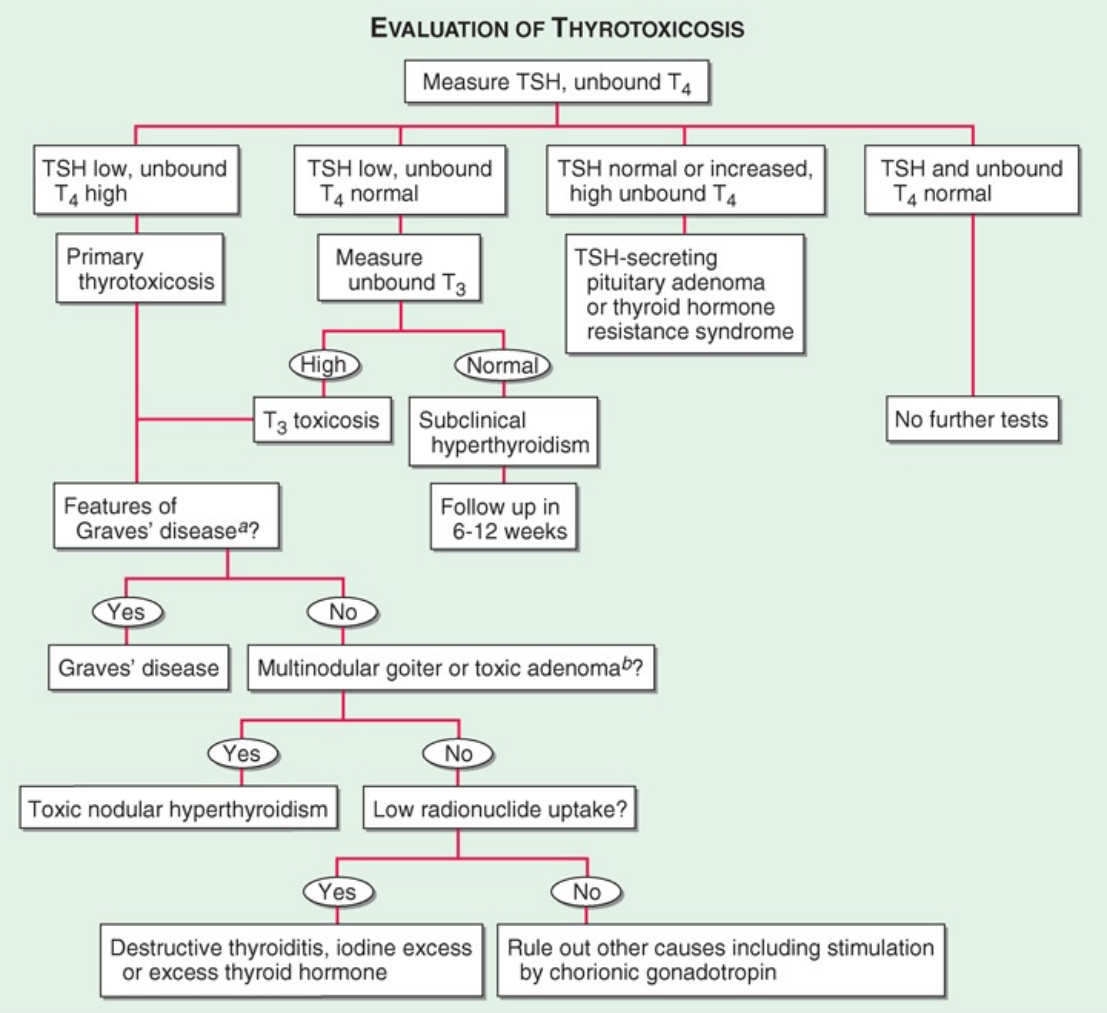

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

[a]Diffuse goiter, positive TPO antibodies, ophthalmopathy, dermopathy; [b]can be confirmed by radionuclide scan. TSH, thyroid-stimulating hormone.

DIFFERENTIAL DIAGNOSIS

Diagnosis of Graves' disease is straightforward in a patient with biochemically confirmed thyrotoxicosis, diffuse goiter on palpation, ophthalmopathy, and often a personal or family history of autoimmune disorders. For patients with thyrotoxicosis who lack these features, the most reliable diagnostic method is to measure TBII or TSI. An alternative is to undertake a radionuclide (^{99m}Tc, ^{123}I, or ^{131}I) scan of the thyroid, which will distinguish the diffuse, high uptake of Graves' disease from nodular thyroid disease, destructive thyroiditis, ectopic thyroid tissue, and factitious thyrotoxicosis. In secondary hyperthyroidism due to a TSH-secreting pituitary tumor, there is also a diffuse goiter. The presence of a nonsuppressed TSH level and the finding of a pituitary tumor on CT or MRI scan readily identify such patients.

Clinical features of thyrotoxicosis can mimic certain aspects of other disorders, including panic attacks, mania, pheochromocytoma, and weight loss associated with malignancy. The diagnosis of thyrotoxicosis can be easily excluded if the TSH and unbound T4 and T_3 levels are normal. A normal TSH also excludes Graves' disease as a cause of diffuse goiter.

CLINICAL COURSE

Clinical features generally worsen without treatment; mortality was 10–30% before the introduction of satisfactory therapy. Some patients with mild Graves' disease experience spontaneous relapses and remissions. Rarely, there may be fluctuation between hypo- and hyperthyroidism due to changes in the functional activity of TSH-R antibodies. About 15% of patients who enter remission after treatment develop hypothyroidism 10–15 years later as a result of the destructive autoimmune process.

The clinical course of ophthalmopathy does not follow that of the thyroid disease. Ophthalmopathy typically worsens over the initial 3–6 months, followed by a plateau phase over the next 12–18 months, with spontaneous improvement, particularly in the soft tissue changes. However, the course is more fulminant in up to 5% of patients, requiring intervention in the acute phase if there is optic nerve compression or corneal ulceration. Diplopia may appear late in the disease due to fibrosis of the extraocular muscles. Some studies suggest that radioiodine treatment for hyperthyroidism worsens the eye disease in a small proportion of patients (especially smokers). Antithyroid drugs or surgery have no adverse effects on the clinical course of ophthalmopathy. Thyroid dermopathy, when it occurs, usually appears 1–2 years after the development of Graves' hyperthyroidism; it may improve spontaneously.

TREATMENT: GRAVES' DISEASE

The *hyperthyroidism* of Graves' disease is treated by reducing thyroid hormone synthesis, using antithyroid drugs, or reducing the amount of thyroid tissue with radioiodine (^{131}I) treatment or by thyroidectomy. Antithyroid drugs are the predominant therapy in many centers in Europe and Japan, whereas radioiodine is more often the first line of treatment in North America. These differences reflect the fact that no single approach is optimal and that patients may require multiple treatments to achieve remission.

The main *antithyroid drugs* are the thionamides, such as propylthiouracil, carbimazole, and the active metabolite of the latter, methimazole. All inhibit the function of TPO, reducing oxidation and organification of iodide. These drugs also reduce thyroid antibody levels by mechanisms that remain unclear, and they appear to enhance rates of remission. Propylthiouracil inhibits deiodination of T4 → T_3. However, this effect is of minor benefit, except in the most severe thyrotoxicosis, and is offset by the much shorter half-life of this drug (90 min) compared to methimazole (6 h).

There are many variations of antithyroid drug regimens. The initial dose of carbimazole or methimazole is usually 10–20 mg every 8 or 12 h, but once-daily dosing is possible after euthyroidism is restored. Propylthiouracil is given at a dose of 100–200 mg every 6–8 h, and divided doses are usually given throughout the course. Lower doses of each drug may suffice in areas of low iodine intake. The starting dose of antithyroid drugs can be gradually reduced (titration regimen) as thyrotoxicosis improves. Alternatively, high doses may be given combined with levothyroxine supplementation (block-replace regimen) to avoid drug-induced hypothyroidism. Initial reports suggesting superior remission rates with the block-replace regimen have not been reproduced in several other trials. The titration regimen is often preferred to minimize the dose of antithyroid drug and provide an index of treatment response.

Thyroid function tests and clinical manifestations are reviewed 3–4 weeks after starting treatment, and the dose is titrated based on unbound T4 levels. Most patients do not achieve euthyroidism until 6–8 weeks after treatment is initiated. TSH levels often remain suppressed for several months and therefore do not provide a sensitive index of treatment response. The usual daily maintenance doses of antithyroid drugs in the titration regimen are 2.5–10 mg of carbimazole or methimazole and 50–100 mg of propylthiouracil. In the block-replace regimen, the initial dose of antithyroid drug is held constant, and the dose of levothyroxine is adjusted to maintain normal unbound T_4 levels. When TSH suppression is alleviated, TSH levels can also be used to monitor therapy.

Maximum remission rates (up to 30–50% in some populations) are achieved by 18–24 months for the titration regimen and by 6 months for the block-replace regimen. For unclear reasons, remission rates appear to vary in different geographic regions. Patients with severe hyperthyroidism and large goiters are most likely to relapse when treatment stops, but outcomes are difficult to predict. All patients should be followed closely for relapse during the first year after treatment and at least annually thereafter.

The common side effects of antithyroid drugs are rash, urticaria, fever, and arthralgia (1–5% of patients). These may resolve spontaneously or after substituting an alternative antithyroid drug. Rare but major side effects include hepatitis; an SLE-like syndrome; and, most important, agranulocytosis (<1%). It is essential that antithyroid drugs are stopped and not restarted if a patient develops major side effects. Written instructions should be provided regarding the symptoms of possible agranulocytosis (e.g., sore throat, fever, mouth ulcers) and the need to stop treatment pending a complete blood count to confirm that agranulocytosis is not present. Management of agranulocytosis is described in Chap. 107. It is not useful to monitor blood counts prospectively, because the onset of agranulocytosis is idiosyncratic and abrupt.

Propranolol (20–40 mg every 6 h) or longer-acting beta blockers such as atenolol, may be helpful to control adrenergic symptoms, especially in the early stages before antithyroid drugs take effect. Beta

blockers are also useful in patients with thyrotoxic periodic paralysis, pending correction of thyrotoxicosis. The need for anticoagulation with coumadin should be considered in all patients with atrial fibrillation. If digoxin is used, increased doses are often needed in the thyrotoxic state.

Radioiodine causes progressive destruction of thyroid cells and can be used as initial treatment or for relapses after a trial of antithyroid drugs. There is a small risk of thyrotoxic crisis (see below) after radioiodine, which can be minimized by pretreatment with antithyroid drugs for at least a month before treatment. Antecedent treatment with antithyroid drugs should be considered for all elderly patients or for those with cardiac problems to deplete thyroid hormone stores before administration of radioiodine. Carbimazole or methimazole must be stopped at least 2 days before radioiodine administration to achieve optimum iodine uptake. Propylthiouracil has a prolonged radioprotective effect and should be stopped several weeks before radioiodine is given, or a larger dose of radioiodine will be necessary.

Efforts to calculate an optimal dose of radioiodine that achieves euthyroidism without a high incidence of relapse or progression to hypothyroidism have not been successful. Some patients inevitably relapse after a single dose because the biologic effects of radiation vary between individuals, and hypothyroidism cannot be uniformly avoided even using accurate dosimetry. A practical strategy is to give a fixed dose based on clinical features, such as the severity of thyrotoxicosis, the size of the goiter (increases the dose needed), and the level of radioiodine uptake (decreases the dose needed). ^{131}I dosage generally ranges between 185 MBq (5 mCi) to 555 MBq (15 mCi). Incomplete treatment or early relapse is more common in males and in patients <40 years of age. Many authorities favor an approach aimed at thyroid ablation (as opposed to euthyroidism), given that levothyroxine replacement is straightforward and most patients ultimately progress to hypothyroidism over 5–10 years, frequently with some delay in the diagnosis of hypothyroidism.

Certain radiation safety precautions are necessary in the first few days after radioiodine treatment, but the exact guidelines vary depending on local protocols. In general, patients need to avoid close, prolonged contact with children and pregnant women for several days because of possible transmission of residual isotope and excessive exposure to radiation emanating from the gland. Rarely, there may be mild pain due to radiation thyroiditis 1–2 weeks after treatment. Hyperthyroidism can persist for 2–3 months before radioiodine takes full effect. For this reason, β-adrenergic blockers or antithyroid drugs can be used to control symptoms during this interval. Persistent hyperthyroidism can be treated with a second dose of radioiodine, usually 6 months after the first dose. The risk of hypothyroidism after radioiodine depends on the dosage but is at least 10–20% in the first year and 5% per year thereafter. Patients should be informed of this possibility before treatment and require close follow-up during the first year and annual thyroid function testing.

Pregnancy and breast-feeding are absolute contraindications to radioiodine treatment, but patients can conceive safely 6 months after treatment. The presence of severe ophthalmopathy requires caution, and some authorities advocate the use of prednisone, 40 mg/d, at the time of radioiodine treatment, tapered over 2–3 months to prevent exacerbation of ophthalmopathy. The overall risk of cancer after radioiodine treatment in adults is not increased. Although many physicians avoid radioiodine in children and adolescents because of the theoretical risks of malignancy, emerging evidence suggests that radioiodine can be used safely in older children.

Subtotal or near-total thyroidectomy is an option for patients who relapse after antithyroid drugs and they prefer this treatment to radioiodine. Some experts recommend surgery in young individuals, particularly when the goiter is very large. Careful control of thyrotoxicosis with antithyroid drugs, followed by potassium iodide (3 drops SSKI orally tid), is needed prior to surgery to avoid thyrotoxic crisis and to reduce the vascularity of the gland. The major complications of surgery—bleeding, laryngeal edema, hypoparathyroidism, and damage to the recurrent laryngeal nerves—are unusual when the procedure is performed by highly experienced surgeons. Recurrence rates in the best series are <2%, but the rate of hypothyroidism is only slightly less than that following radioiodine treatment.

The titration regimen of antithyroid drugs should be used to manage Graves' disease in *pregnancy*, as blocking doses of these drugs produce fetal hypothyroidism. Propylthiouracil is usually used because of relatively low transplacental transfer and its ability to block $T4 \rightarrow T_3$ conversion. Also, carbimazole and methimazole have been associated with rare cases of fetal *aplasia cutis* and other defects, such as choanal atresia. The lowest effective dose of propylthiouracil should be given, and it is often possible to stop treatment in the last trimester because TSI tend to decline in pregnancy. Nonetheless, the transplacental transfer of these antibodies rarely causes *fetal* or *neonatal thyrotoxicosis*. Poor intrauterine growth, a fetal heart rate of >160 beats/min, and high levels of maternal TSI in the last trimester may herald this complication. Antithyroid drugs given to the mother can be used to treat the fetus and may be needed for 1–3 months after delivery, until the maternal antibodies disappear from the baby's circulation. The postpartum period is a time of major risk for relapse of Graves' disease. Breast-feeding is safe with low doses of antithyroid drugs. Graves' disease in *children* is usually managed with antithyroid drugs, often given as a prolonged course of the titration regimen. Surgery or radioiodine may be indicated for severe disease.

Thyrotoxic crisis, or *thyroid storm*, is rare and presents as a life-threatening exacerbation of

hyperthyroidism, accompanied by fever, delirium, seizures, coma, vomiting, diarrhea, and jaundice. The mortality rate due to cardiac failure, arrhythmia, or hyperthermia is as high as 30%, even with treatment. Thyrotoxic crisis is usually precipitated by acute illness (e.g., stroke, infection, trauma, diabetic ketoacidosis), surgery (especially on the thyroid), or radioiodine treatment of a patient with partially treated or untreated hyperthyroidism. Management requires intensive monitoring and supportive care, identification and treatment of the precipitating cause, and measures that reduce thyroid hormone synthesis. Large doses of propylthiouracil (600 mg loading dose and 200–300 mg every 6 h) should be given orally or by nasogastric tube or per rectum; the drug's inhibitory action on $T4 \rightarrow T_3$ conversion makes it the antithyroid drug of choice. One hour after the first dose of propylthiouracil, stable iodide is given to block thyroid hormone synthesis via the Wolff-Chaikoff effect (the delay allows the antithyroid drug to prevent the excess iodine from being incorporated into new hormone). A saturated solution of potassium iodide (5 drops SSKI every 6 h), or ipodate or iopanoic acid (500 mg per 12 h), may be given orally. (Sodium iodide, 0.25 g IV every 6 h, is an alternative but is not generally available.) Propranolol should also be given to reduce tachycardia and other adrenergic manifestations (40–60 mg PO every 4 h; or 2 mg IV every 4 h). Although other β-adrenergic blockers can be used, high doses of propranolol decrease $T_4 \rightarrow T_3$ conversion, and the doses can be easily adjusted. Caution is needed to avoid acute negative inotropic effects, but controlling the heart rate is important, as some patients develop a form of high-output heart failure. Additional therapeutic measures include glucocorticoids (e.g., dexamethasone, 2 mg every 6 h), antibiotics if infection is present, cooling, oxygen, and intravenous fluids.

Ophthalmopathy requires no active treatment when it is mild or moderate, because there is usually spontaneous improvement. General measures include meticulous control of thyroid hormone levels, cessation of smoking, and an explanation of the natural history of ophthalmopathy. Discomfort can be relieved with artificial tears (e.g., 1% methylcellulose), eye ointment, and the use of dark glasses with side frames. Periorbital edema may respond to a more upright sleeping position or a diuretic. Corneal exposure during sleep can be avoided by using patches or taping the eyelids shut. Minor degrees of diplopia improve with prisms fitted to spectacles. Severe ophthalmopathy, with optic nerve involvement or chemosis resulting in corneal damage, is an emergency requiring joint management with an ophthalmologist. Short-term benefit can be gained in about two-thirds of patients by the use of high-dose glucocorticoids (e.g., prednisone, 40–80 mg daily), sometimes combined with cyclosporine. Glucocorticoid doses are tapered by 5 mg every 2 weeks, but the taper often results in reemergence of congestive symptoms. Pulse therapy with IV methylprednisolone (e.g., 500–1000 mg of methylprednisolone in 250 mL of saline infused over 2 h daily for 1 week) followed by an oral regimen is also used. When glucocorticoids are ineffective, orbital decompression can be achieved by removing bone from any wall of the orbit, thereby allowing displacement of fat and swollen extraocular muscles. The transantral route is used most often, because it requires no external incision. Proptosis recedes an average of 5 mm, but there may be residual or even worsened diplopia. Once the eye disease has stabilized, surgery may be indicated for relief of diplopia and correction of the appearance. External beam radiotherapy of the orbits has been used for many years, but the efficacy of this therapy remains unclear, and it is best reserved for those who have failed or are not candidates for glucocorticoid therapy.

Thyroid dermopathy does not usually require treatment, but it can cause cosmetic problems or interfere with the fit of shoes. Surgical removal is not indicated. If necessary, treatment consists of topical, high-potency glucocorticoid ointment under an occlusive dressing. Octreotide may be beneficial in some cases.

Other Causes of Thyrotoxicosis

Destructive thyroiditis (subacute or silent thyroiditis) typically presents with a short thyrotoxic phase due to the release of preformed thyroid hormones and catabolism of Tg (see "Subacute Thyroiditis," below). True hyperthyroidism is absent, as demonstrated by a low radionuclide uptake. Circulating Tg levels are usually increased. Other causes of thyrotoxicosis with low or absent thyroid radionuclide uptake include *thyrotoxicosis factitia*, iodine excess, and, rarely, ectopic thyroid tissue, particularly teratomas of the ovary (*struma ovarii*); and functional metastatic follicular carcinoma. Whole-body radionuclide studies can demonstrate ectopic thyroid tissue, and thyrotoxicosis factitia can be distinguished from destructive thyroiditis by the clinical features and low levels of Tg. Amiodarone treatment is associated with thyrotoxicosis in up to 10% of patients, particularly in areas of low iodine intake (see below).

TSH-secreting pituitary adenoma is a rare cause of thyrotoxicosis. It can be identified by the presence of an inappropriately normal or increased TSH level in a patient with hyperthyroidism, diffuse goiter, and elevated T4 and T_3 levels (Chap. 339). Elevated levels of the α-subunit of TSH, released by the TSH-secreting adenoma, support this diagnosis, which can be confirmed by demonstrating the pituitary tumor on MRI or CT scan. A combination of transsphenoidal surgery, sella irradiation, and octreotide may be required to normalize TSH, because many of these tumors are large and locally invasive at the time of diagnosis. Radioiodine or antithyroid drugs can be used to control thyrotoxicosis.

Thyrotoxicosis caused by *toxic MNG* and *hyperfunctioning solitary nodules* is discussed below.

THYROIDITIS

A clinically useful classification of thyroiditis is based on the onset and duration of disease **(Table 341-8)**.

Table 341-8 Causes of Thyroiditis

Acute
Bacterial infection: especially *Staphylococcus*, *Streptococcus*, and *Enterobacter*
Fungal infection: *Aspergillus*, *Candida*, *Coccidioides*, *Histoplasma*, and *Pneumocystis*
Radiation thyroiditis after ^{131}I treatment Amiodarone (may also be subacute or chronic)
Subacute
Viral (or granulomatous) thyroiditis
Silent thyroiditis (including postpartum thyroiditis)
Mycobacterial infection
Chronic
Autoimmunity: focal thyroiditis, Hashimoto's thyroiditis, atrophic thyroiditis Riedel's thyroiditis
Parasitic thyroiditis: echinococcosis, strongyloidiasis, cysticercosis
Traumatic: after palpation

Acute Thyroiditis

Acute thyroiditis is rare and due to suppurative infection of the thyroid. In children and young adults, the most common cause is the presence of a piriform sinus, a remnant of the fourth branchial pouch that connects the oropharynx with the thyroid. Such sinuses are predominantly left-sided. A long-standing goiter and degeneration in a thyroid malignancy are risk factors in the elderly. The patient presents with thyroid pain, often referred to the throat or ears, and a small, tender goiter that may be asymmetric. Fever, dysphagia, and erythema over the thyroid are common, as are systemic symptoms of a febrile illness and lymphadenopathy.

The differential diagnosis of *thyroid pain* includes subacute or, rarely, chronic thyroiditis; hemorrhage into a cyst; malignancy including lymphoma; and, rarely, amiodarone-induced thyroiditis or amyloidosis. However, the abrupt presentation and clinical features of acute thyroiditis rarely cause confusion. The erythrocyte sedimentation rate (ESR) and white cell count are usually increased, but thyroid function is normal. FNA biopsy shows infiltration by polymorphonuclear leukocytes; culture of the sample can identify the organism. Caution is needed in immunocompromised patients as fungal, mycobacterial, or *Pneumocystis* thyroiditis can occur in this setting. Antibiotic treatment is guided initially by Gram stain and, subsequently, by cultures of the FNA biopsy. Surgery may be needed to drain an abscess, which can be localized by CT scan or ultrasound. Tracheal obstruction, septicemia, retropharyngeal abscess, mediastinitis, and jugular venous thrombosis may complicate acute thyroiditis but are uncommon with prompt use of antibiotics.

Subacute Thyroiditis

This is also termed *de Quervain's thyroiditis*, *granulomatous thyroiditis*, or *viral thyroiditis*. Many viruses have been implicated, including mumps, coxsackie, influenza, adenoviruses, and echoviruses, but attempts to identify the virus in an individual patient are often unsuccessful and do not influence management. The diagnosis of subacute thyroiditis is often overlooked because the symptoms can mimic pharyngitis. The peak incidence occurs at 30–50 years, and women are affected three times more frequently than men.

PATHOPHYSIOLOGY

The thyroid shows a characteristic patchy inflammatory infiltrate with disruption of the thyroid follicles and multinucleated giant cells within some follicles. The follicular changes progress to granulomas accompanied by fibrosis. Finally, the thyroid returns to normal, usually several months after onset. During the initial phase of follicular destruction, there is release of Tg and thyroid hormones, leading to increased circulating T4 and T_3 and suppression of TSH **(Fig. 341-9)**. During this destructive phase, radioactive iodine uptake is low or undetectable. After several weeks, the thyroid is depleted of stored thyroid hormone and a phase of hypothyroidism typically occurs, with low unbound T_4 (and sometimes T_3) and moderately increased TSH levels. Radioactive iodine uptake returns to normal or is even

increased as a result of the rise in TSH. Finally, thyroid hormone and TSH levels return to normal as the disease subsides.

Figure 341-9 Clinical Course of Subacute Thyroiditis.

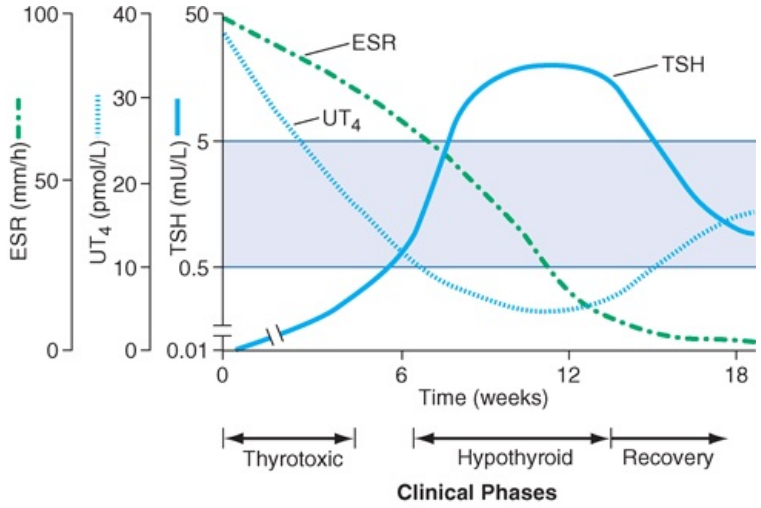

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The release of thyroid hormones is initially associated with a thyrotoxic phase and suppressed thyroid-stimulating hormone (TSH). A hypothyroid phase then ensues, with low T4 and TSH levels that are initially low but gradually increase. During the recovery phase, increased TSH levels combined with resolution of thyroid follicular injury leads to normalization of thyroid function, often several months after the beginning of the illness. ESR, erythrocyte sedimentation rate; FT_4, free or unbound T_4.

CLINICAL MANIFESTATIONS

The patient usually presents with a painful and enlarged thyroid, sometimes accompanied by fever. There may be features of thyrotoxicosis or hypothyroidism, depending on the phase of the illness. Malaise and symptoms of an upper respiratory tract infection may precede the thyroid-related features by several weeks. In other patients, the onset is acute, severe, and without obvious antecedent. The patient typically complains of a sore throat, and examination reveals a small goiter that is exquisitely tender. Pain is often referred to the jaw or ear. Complete resolution is the usual outcome, but permanent hypothyroidism can occur, particularly in those with coincidental thyroid autoimmunity. A prolonged course over many months, with one or more relapses, occurs in a small percentage of patients.

LABORATORY EVALUATION

As depicted in **(Fig. 341-9)**, thyroid function tests characteristically evolve through three distinct phases over about 6 months: (1) thyrotoxic phase, (2) hypothyroid phase, and (3) recovery phase. In the thyrotoxic phase, T4 and T_3 levels are increased, reflecting their discharge from the damaged thyroid cells, and TSH is suppressed. The T_4/T_3 ratio is greater than in Graves' disease or thyroid autonomy, in which T_3 is often disproportionately increased. The diagnosis is confirmed by a high ESR and low radioiodine uptake. The white blood cell count may be increased, and thyroid antibodies are negative. If the diagnosis is in doubt, FNA biopsy may be useful, particularly to distinguish unilateral involvement from bleeding into a cyst or neoplasm.

TREATMENT: SUBACUTE THYROIDITIS

Relatively large doses of aspirin (e.g., 600 mg every 4–6 h) or NSAIDs are sufficient to control symptoms in many cases. If this treatment is inadequate, or if the patient has marked local or systemic symptoms, glucocorticoids should be given. The usual starting dose is 40–60 mg prednisone, depending on severity. The dose is gradually tapered over 6–8 weeks, in response to improvement in symptoms and the ESR. If a relapse occurs during glucocorticoid withdrawal, treatment should be started again and withdrawn more gradually. In these patients, it is useful to wait until the radioactive iodine uptake normalizes before stopping treatment. Thyroid function should be monitored every 2–4 weeks using TSH and unbound T4 levels. Symptoms of thyrotoxicosis improve spontaneously but may be ameliorated by β-adrenergic blockers; antithyroid drugs play no role in treatment of the thyrotoxic phase. Levothyroxine replacement may be needed if the hypothyroid phase is prolonged, but doses should be low enough (50 to 100 μg daily) to allow TSH-mediated recovery.

Silent Thyroiditis

Painless thyroiditis, or *"silent" thyroiditis*, occurs in patients with underlying autoimmune thyroid disease. It has a clinical course similar to that of subacute thyroiditis, except that there is little or no thyroid tenderness. The condition occurs in up to 5% of women 3–6 months after pregnancy and is then termed *postpartum thyroiditis*. Typically, patients have a brief phase of thyrotoxicosis lasting 2–4 weeks, followed by hypothyroidism for 4–12 weeks, and then resolution; often, however, only one phase is apparent. The condition is associated with the presence of TPO antibodies antepartum, and it is three times more common in women with type 1 diabetes mellitus. As in subacute thyroiditis, the radioactive iodine uptake is initially suppressed. In addition to the painless goiter, silent thyroiditis can be distinguished from subacute thyroiditis by a normal ESR and the presence of TPO antibodies. Glucocorticoid treatment is not indicated for silent thyroiditis. Severe thyrotoxic symptoms can be managed with a brief course of propranolol, 20–40 mg three or four times daily. Thyroxine replacement may be needed for the hypothyroid phase but should be withdrawn after 6–9 months, as recovery is the rule. Annual follow-up thereafter is recommended, because a proportion of these individuals develop permanent hypothyroidism. The condition may recur in subsequent pregnancies.

Drug-Induced Thyroiditis

Patients receiving cytokines such as IFN-α or IL-2 may develop painless thyroiditis. IFN-α, which is used to treat chronic hepatitis B or C and hematologic and skin malignancies, causes thyroid dysfunction in up to 5% of treated patients. It has been associated with painless thyroiditis, hypothyroidism, and Graves' disease, and is most common in women with TPO antibodies prior to treatment. For discussion of amiodarone, see "Amiodarone Effects on Thyroid Function," below.

Chronic Thyroiditis

Focal thyroiditis is present in 20–40% of euthyroid autopsy cases and is associated with serologic evidence of autoimmunity, particularly the presence of TPO antibodies. These antibodies are 4–10 times more common in otherwise healthy women than men. The most common clinically apparent cause of chronic thyroiditis is *Hashimoto's thyroiditis*, an autoimmune disorder that often presents as a firm or hard goiter of variable size (see above). *Riedel's thyroiditis* is a rare disorder that typically occurs in middle-aged women. It presents with an insidious, painless goiter with local symptoms due to compression of the esophagus, trachea, neck veins, or recurrent laryngeal nerves. Dense fibrosis disrupts normal gland architecture and can extend outside the thyroid capsule. Despite these extensive histologic changes, thyroid dysfunction is uncommon. The goiter is hard, nontender, often asymmetric, and fixed, leading to suspicion of a malignancy. Diagnosis requires open biopsy as FNA biopsy is usually inadequate. Treatment is directed to surgical relief of compressive symptoms. Tamoxifen may also be beneficial. There is an association between Riedel's thyroiditis and idiopathic fibrosis at other sites (retroperitoneum, mediastinum, biliary tree, lung, and orbit).

SICK EUTHYROID SYNDROME

Any acute, severe illness can cause abnormalities of circulating TSH or thyroid hormone levels in the absence of underlying thyroid disease, making these measurements potentially misleading. The major cause of these hormonal changes is the release of cytokines such as IL-6. Unless a thyroid disorder is strongly suspected, the routine testing of thyroid function should be avoided in acutely ill patients.

The most common hormone pattern in sick euthyroid syndrome (SES) is a decrease in total and unbound T_3 levels (low T_3 syndrome) with normal levels of T4 and TSH. The magnitude of the fall in T_3 correlates with the severity of the illness. T_4 conversion to T_3 via peripheral deiodination is impaired, leading to increased reverse T_3 (rT_3). Despite this effect, decreased clearance rather than increased production is the major basis for increased rT_3. Also, T_4 is alternately metabolized to the hormonally inactive T_3 sulfate. It is generally assumed that this low T_3 state is adaptive, because it can be induced in normal individuals by fasting. Teleologically, the fall in T_3 may limit catabolism in starved or ill patients.

Very sick patients may exhibit a dramatic fall in total T4 and T_3 levels (low T_4 syndrome). This state has a poor prognosis. A key factor in the fall in T_4 levels is altered binding to TBG. T_4 assays usually demonstrate a normal unbound T_4 level in such patients, depending on the assay method used.

Fluctuation in TSH levels also creates challenges in the interpretation of thyroid function in sick patients. TSH levels may range from <0.1 to >20 mIU/L; these alterations reverse after recovery, confirming the absence of underlying thyroid disease. A rise in cortisol or administration of glucocorticoids may provide a partial explanation for decreased TSH levels. The exact mechanisms underlying the subnormal TSH seen in 10% of sick patients and the increased TSH seen in 5% remain unclear but may be mediated by cytokines including IL-12 and IL-18.

Any severe illness can induce changes in thyroid hormone levels, but certain disorders exhibit a distinctive pattern of abnormalities. Acute liver disease is associated with an initial rise in total (but not unbound) T_3 and T4 levels, due to TBG release; these levels become subnormal with progression to liver failure. A transient increase in total and unbound T_4 levels, usually with a normal T_3 level, is seen

in 5–30% of acutely ill psychiatric patients. TSH values may be transiently low, normal, or high in these patients. In the early stage of HIV infection, T_3 and T_4 levels rise, even if there is weight loss. T_3 levels fall with progression to AIDS, but TSH usually remains normal. Renal disease is often accompanied by low T_3 concentrations, but with normal rather than increased rT_3 levels, due to an unknown factor that increases uptake of rT_3 into the liver.

The diagnosis of SES is challenging. Historic information may be limited, and patients often have multiple metabolic derangements. Useful features to consider include previous history of thyroid disease and thyroid function tests, evaluation of the severity and time course of the patient's acute illness, documentation of medications that may affect thyroid function or thyroid hormone levels, and measurements of rT_3 together with unbound thyroid hormones and TSH. The diagnosis of SES is frequently presumptive, given the clinical context and pattern of laboratory values; only resolution of the test results with clinical recovery can clearly establish this disorder. Treatment of SES with thyroid hormone (T4 and/or T_3) is controversial, but most authorities recommend monitoring the patient's thyroid function tests during recovery, without administering thyroid hormone, unless there is historic or clinical evidence suggestive of hypothyroidism. Sufficiently large randomized controlled trials using thyroid hormone are unlikely to resolve this therapeutic controversy in the near future, because clinical presentations and outcomes are highly variable.

AMIODARONE EFFECTS ON THYROID FUNCTION

Amiodarone is a commonly used type III antiarrhythmic agent (Chap. 233). It is structurally related to thyroid hormone and contains 39% iodine by weight. Thus, typical doses of amiodarone (200 mg/d) are associated with very high iodine intake, leading to greater than fortyfold increases in plasma and urinary iodine levels. Moreover, because amiodarone is stored in adipose tissue, high iodine levels persist for >6 months after discontinuation of the drug. Amiodarone inhibits deiodinase activity, and its metabolites function as weak antagonists of thyroid hormone action. Amiodarone has the following effects on thyroid function: (1) acute, transient suppression of thyroid function; (2) hypothyroidism in patients susceptible to the inhibitory effects of a high iodine load; and (3) thyrotoxicosis that may be caused by either a Jod-Basedow effect from the iodine load, in the setting of MNG or incipient Graves' disease, or a thyroiditis-like condition.

The initiation of amiodarone treatment is associated with a transient decrease of T4 levels, reflecting the inhibitory effect of iodine on T_4 release. Soon thereafter, most individuals escape from iodide-dependent suppression of the thyroid (Wolff-Chaikoff effect), and the inhibitory effects on deiodinase activity and thyroid hormone receptor action become predominant. These events lead to the following pattern of thyroid function tests: increased T_4, decreased T_3, increased rT_3, and a transient TSH increase (up to 20 mIU/L). TSH levels normalize or are slightly suppressed within 1–3 months.

The incidence of hypothyroidism from amiodarone varies geographically, apparently correlating with iodine intake. Hypothyroidism occurs in up to 13% of amiodarone-treated patients in iodine-replete countries, such as the United States, but is less common (<6% incidence) in areas of lower iodine intake, such as Italy or Spain. The pathogenesis appears to involve an inability of the thyroid gland to escape from the Wolff-Chaikoff effect in autoimmune thyroiditis. Consequently, amiodarone-associated hypothyroidism is more common in women and individuals with positive TPO antibodies. It is usually unnecessary to discontinue amiodarone for this side effect, because levothyroxine can be used to normalize thyroid function. TSH levels should be monitored, because T4 levels are often increased for the reasons described above.

The management of amiodarone-induced thyrotoxicosis (AIT) is complicated by the fact that there are different causes of thyrotoxicosis and because the increased thyroid hormone levels exacerbate underlying arrhythmias and coronary artery disease. Amiodarone treatment causes thyrotoxicosis in 10% of patients living in areas of low iodine intake and in 2% of patients in regions of high iodine intake. There are two major forms of AIT, although some patients have features of both. Type 1 AIT is associated with an underlying thyroid abnormality (preclinical Graves' disease or nodular goiter). Thyroid hormone synthesis becomes excessive as a result of increased iodine exposure (Jod-Basedow phenomenon). Type 2 AIT occurs in individuals with no intrinsic thyroid abnormalities and is the result of drug-induced lysosomal activation leading to destructive thyroiditis with histiocyte accumulation in the thyroid; the incidence rises as cumulative amiodarone dosage increases. Mild forms of type 2 AIT can resolve spontaneously or can occasionally lead to hypothyroidism. Color-flow doppler thyroid scanning shows increased vascularity in type 1 AIT but decreased vascularity in type 2 AIT. Thyroid scintiscans are difficult to interpret in this setting because the high endogenous iodine levels diminish tracer uptake. However, the presence of normal or rarely increased uptake favors type 1 AIT.

In AIT, the drug should be stopped, if possible, although this is often impractical because of the underlying cardiac disorder. Discontinuation of amiodarone will not have an acute effect because of its storage and prolonged half-life. High doses of antithyroid drugs can be used in type 1 AIT but are often ineffective. In type 2 AIT, oral contrast agents, such as sodium ipodate (500 mg/d) or sodium

tyropanoate (500 mg, 1–2 doses/d), rapidly reduce T4 and T_3 levels, decrease $T_4 \rightarrow T_3$ conversion, and may block tissue uptake of thyroid hormones. Potassium perchlorate, 200 mg every 6 h, has been used to reduce thyroidal iodide content. Perchlorate treatment has been associated with agranulocytosis, though the risk appears relatively low with short-term use. Glucocorticoids, as administered for subacute thyroiditis, have modest benefit in type 2 AIT. Lithium blocks thyroid hormone release and can also provide some benefit. Near-total thyroidectomy rapidly decreases thyroid hormone levels and may be the most effective long-term solution if the patient can undergo the procedure safely.

THYROID FUNCTION IN PREGNANCY

Five factors alter thyroid function in pregnancy: (1) the transient increase in hCG during the first trimester, which stimulates the TSH-R; (2) the estrogen-induced rise in TBG during the first trimester, which is sustained during pregnancy; (3) alterations in the immune system, leading to the onset, exacerbation, or amelioration of an underlying autoimmune thyroid disease (see above); (4) increased thyroid hormone metabolism by the placenta; and (5) increased urinary iodide excretion, which can cause impaired thyroid hormone production in areas of marginal iodine sufficiency. Women with a precarious iodine intake (<50 μg/d) are most at risk of developing a goiter during pregnancy, and iodine supplementation should be considered to prevent maternal and fetal hypothyroidism and the development of neonatal goiter.

The rise in circulating hCG levels during the first trimester is accompanied by a reciprocal fall in TSH that persists into the middle of pregnancy. This appears to reflect weak binding of hCG, which is present at very high levels, to the TSH-R. Rare individuals have been described with variant TSH-R sequences that enhance hCG binding and TSH-R activation. Human chorionic gonadotropin-induced changes in thyroid function can result in transient gestational hyperthyroidism and/or *hyperemesis gravidarum*, a condition characterized by severe nausea and vomiting and risk of volume depletion. Antithyroid drugs are rarely needed, and parenteral fluid replacement usually suffices until the condition resolves.

Maternal hypothyroidism occurs in 2–3% of women of child-bearing age and is associated with increased risk of developmental delay in the offspring. Consequently, TSH screening for hypothyroidism is indicated in early pregnancy and should be considered in women who are planning pregnancy, particularly if they have a goiter or strong family history of autoimmune thyroid disease. Thyroid hormone requirements are increased by 25–50 μg/d during pregnancy.

GOITER AND NODULAR THYROID DISEASE

Goiter refers to an enlarged thyroid gland. Biosynthetic defects, iodine deficiency, autoimmune disease, and nodular diseases can each lead to goiter, though by different mechanisms. Biosynthetic defects and iodine deficiency are associated with reduced efficiency of thyroid hormone synthesis, leading to increased TSH, which stimulates thyroid growth as a compensatory mechanism to overcome the block in hormone synthesis. Graves' disease and Hashimoto's thyroiditis are also associated with goiter. In Graves' disease, the goiter results mainly from the TSH-R–mediated effects of TSI. The goitrous form of Hashimoto's thyroiditis occurs because of acquired defects in hormone synthesis, leading to elevated levels of TSH and its consequent growth effects. Lymphocytic infiltration and immune system–induced growth factors also contribute to thyroid enlargement in Hashimoto's thyroiditis. Nodular disease is characterized by the disordered growth of thyroid cells, often combined with the gradual development of fibrosis. Because the management of goiter depends on the etiology, the detection of thyroid enlargement on physical examination should prompt further evaluation to identify its cause.

Nodular thyroid disease is common, occurring in about 3–7% of adults when assessed by physical examination. Using more sensitive techniques, such as ultrasound, it is present in >25% of adults. Thyroid nodules may be solitary or multiple, and they may be functional or nonfunctional.

Diffuse Nontoxic (Simple) Goiter

ETIOLOGY AND PATHOGENESIS

When diffuse enlargement of the thyroid occurs in the absence of nodules and hyperthyroidism, it is referred to as a *diffuse nontoxic goiter*. This is sometimes called *simple goiter*, because of the absence of nodules, or *colloid goiter*, because of the presence of uniform follicles that are filled with colloid. Worldwide, diffuse goiter is most commonly caused by iodine deficiency and is termed *endemic goiter* when it affects >5% of the population. In nonendemic regions, *sporadic goiter* occurs, and the cause is usually unknown. Thyroid enlargement in teenagers is sometimes referred to as *juvenile goiter*. In general, goiter is more common in women than men, probably because of the greater prevalence of underlying autoimmune disease and the increased iodine demands associated with pregnancy.

In *iodine-deficient areas*, thyroid enlargement reflects a compensatory effort to trap iodide and produce sufficient hormone under conditions in which hormone synthesis is relatively inefficient.

Somewhat surprisingly, TSH levels are usually normal or only slightly increased, suggesting increased sensitivity to TSH or activation of other pathways that lead to thyroid growth. Iodide appears to have direct actions on thyroid vasculature and may indirectly affect growth through vasoactive substances such as endothelins and nitric oxide. Endemic goiter is also caused by exposure to environmental *goitrogens* such as cassava root, which contains a thiocyanate; vegetables of the Cruciferae family (known as cruciferous vegetables) (e.g., brussels sprout, cabbage, and cauliflower); and milk from regions where goitrogens are present in grass. Though relatively rare, inherited defects in thyroid hormone synthesis lead to a diffuse nontoxic goiter. Abnormalities at each step in hormone synthesis, including iodide transport (NIS), Tg synthesis, organification and coupling (TPO), and the regeneration of iodide (dehalogenase), have been described.

Clinical Manifestations and Diagnosis

If thyroid function is preserved, most goiters are asymptomatic. Spontaneous hemorrhage into a cyst or nodule may cause the sudden onset of localized pain and swelling. Examination of a diffuse goiter reveals a symmetrically enlarged, nontender, generally soft gland without palpable nodules. Goiter is defined, somewhat arbitrarily, as a lateral lobe with a volume greater than the thumb of the individual being examined. If the thyroid is markedly enlarged, it can cause tracheal or esophageal compression. These features are unusual, however, in the absence of nodular disease and fibrosis. *Substernal goiter* may obstruct the thoracic inlet. *Pemberton's sign* refers to symptoms of faintness with evidence of facial congestion and external jugular venous obstruction when the arms are raised above the head, a maneuver that draws the thyroid into the thoracic inlet. Respiratory flow measurements and CT or MRI should be used to evaluate substernal goiter in patients with obstructive signs or symptoms.

Thyroid function tests should be performed in all patients with goiter to exclude thyrotoxicosis or hypothyroidism. It is not unusual, particularly in iodine deficiency, to find a low total T4, with normal T_3 and TSH, reflecting enhanced $T_4 \rightarrow T_3$ conversion. A low TSH with a normal free T_3 and free T_4, particularly in older patients, suggests the possibility of thyroid autonomy or undiagnosed Graves' disease, and is termed *subclinical thyrotoxicosis*. The benefit of treatment (typically with radioiodine) in subclinical thyrotoxicosis, versus follow-up and implementing treatment if free T_3 or free T_4 levels become abnormal, is unclear, but treatment is increasingly recommended in the elderly to reduce the risk of atrial fibrillation and bone loss. TPO antibodies may be useful to identify patients at increased risk of autoimmune thyroid disease. Low urinary iodine levels (<10 μg/dL) support a diagnosis of iodine deficiency. Thyroid scanning is not generally necessary but will reveal increased uptake in iodine deficiency and most cases of dyshormonogenesis. Ultrasound is not generally indicated in the evaluation of diffuse goiter unless a nodule is palpable on physical examination.

TREATMENT: DIFFUSE NONTOXIC (SIMPLE) GOITER

Iodine or thyroid hormone replacement induces variable regression of goiter in iodine deficiency, depending on how long it has been present and the degree of fibrosis that has developed. Because of the possibility of underlying thyroid autonomy, caution should be exercised when instituting suppressive thyroxine therapy in patients with goiter, particularly if the baseline TSH is in the low to normal range. In younger patients, the dose of levothyroxine can be started at 100 mcg/d and adjusted to suppress the TSH into the low to normal, but detectable, range. Treatment of elderly patients should be initiated at 50 mcg/d. The efficacy of suppressive treatment is greater in younger patients and for those with soft goiters. Significant regression is usually seen within 3–6 months of treatment; after this time, it is unlikely to occur. In older patients and in those with some degree of nodular disease or fibrosis, fewer than one-third demonstrate significant shrinkage of the goiter. Surgery is rarely indicated for diffuse goiter. Exceptions include documented evidence of tracheal compression or obstruction of the thoracic outlet, which are more likely to be associated with substernal multinodular goiters (see below). Subtotal or near-total thyroidectomy for these or cosmetic reasons should be performed by an experienced surgeon to minimize complication rates. Surgery should be followed by replacement with levothyroxine, with the aim of keeping the TSH level at the lower end of the reference range to prevent regrowth of the goiter. Radioiodine reduces goiter size by about 50% in the majority of patients over 6–12 months. It is rarely associated with transient acute swelling of the thyroid, which is usually inconsequential unless there is severe tracheal narrowing. If they are not treated with levothyroxine, patients should be followed after radioiodine treatment for the possible development of hypothyroidism.

Nontoxic Multinodular Goiter

ETIOLOGY AND PATHOGENESIS

Depending on the population studied, MNG occurs in up to 12% of adults. MNG is more common in women than men and increases in prevalence with age. It is more common in iodine-deficient regions but also occurs in regions of iodine sufficiency, reflecting multiple genetic, autoimmune, and environmental influences on the pathogenesis.

There is typically wide variation in nodule size. Histology reveals a spectrum of morphologies ranging from hypercellular regions to cystic areas filled with colloid. Fibrosis is often extensive, and areas of

hemorrhage or lymphocytic infiltration may be seen. Using molecular techniques, most nodules within a MNG are polyclonal in origin, suggesting a hyperplastic response to locally produced growth factors and cytokines. TSH, which is usually not elevated, may play a permissive or contributory role. Monoclonal lesions also occur within a MNG, reflecting mutations in genes that confer a selective growth advantage to the progenitor cell.

CLINICAL MANIFESTATIONS

Most patients with nontoxic MNG are asymptomatic and euthyroid. MNG typically develops over many years and is detected on routine physical examination or when an individual notices an enlargement in the neck. If the goiter is large enough, it can ultimately lead to compressive symptoms including difficulty swallowing, respiratory distress (tracheal compression), or plethora (venous congestion); but these symptoms are uncommon. Symptomatic MNGs are usually extraordinarily large and/or develop fibrotic areas that cause compression. Sudden pain in a MNG is usually caused by hemorrhage into a nodule but should raise the possibility of invasive malignancy. Hoarseness, reflecting laryngeal nerve involvement, also suggests malignancy.

DIAGNOSIS

On examination, thyroid architecture is distorted, and multiple nodules of varying size can be appreciated. Because many nodules are deeply embedded in thyroid tissue or reside in posterior or substernal locations, it is not possible to palpate all nodules. A TSH level should be measured to exclude subclinical hyper- or hypothyroidism, but thyroid function is usually normal. Tracheal deviation is common, but compression must usually exceed 70% of the tracheal diameter before there is significant airway compromise. Pulmonary function testing can be used to assess the functional effects of compression and to detect tracheomalacia, which characteristically causes inspiratory stridor. CT or MRI can be used to evaluate the anatomy of the goiter and the extent of substernal extension, which is often much greater than is apparent on physical examination. A barium swallow may reveal the extent of esophageal compression. The risk of malignancy in MNG is similar to that in solitary nodules. Ultrasonography can be used to identify which nodules should be biopsied, including large, dominant nodules or those with sonographic characteristics suggestive of malignancy (e.g., microcalcifications, hypoechogenicity, increased vascularity).

TREATMENT: NONTOXIC MULTINODULAR GOITER

Most nontoxic MNGs can be managed conservatively. T4 suppression is rarely effective for reducing goiter size and introduces the risk of subclinical or overt thyrotoxicosis, particularly if there is underlying autonomy or if it develops during treatment. If levothyroxine is used, it should be started at low doses (50 μg) and advanced gradually while monitoring the TSH level to avoid excessive suppression. Contrast agents and other iodine-containing substances should be avoided because of the risk of inducing the *Jod-Basedow effect*, characterized by enhanced thyroid hormone production by autonomous nodules. Radioiodine is used with increasing frequency because it can decrease goiter size and may selectively ablate regions of autonomy. Dosage of ^{131}I depends on the size of the goiter and radioiodine uptake but is usually about 3.7 MBq (0.1 mCi) per gram of tissue, corrected for uptake [typical dose 370–1070 MBq (10 to 29 mCi)]. Repeat treatment may be needed and effectiveness may be increased by concurrent administration of recombinant TSH. It is possible to achieve a 40–50% reduction in goiter size in most patients. Earlier concerns about radiation-induced thyroid swelling and tracheal compression have diminished; studies have shown this complication to be rare. When acute compression occurs, glucocorticoid treatment or surgery may be needed. Radiation-induced hypothyroidism is less common than after treatment for Graves' disease. However, posttreatment autoimmune thyrotoxicosis may occur in up to 5% of patients treated for nontoxic MNG. Surgery remains highly effective but is not without risk, particularly in older patients with underlying cardiopulmonary disease.

Toxic Multinodular Goiter

The pathogenesis of toxic MNG appears to be similar to that of nontoxic MNG; the major difference is the presence of functional autonomy in toxic MNG. The molecular basis for autonomy in toxic MNG remains unknown. As in nontoxic goiters, many nodules are polyclonal, while others are monoclonal and vary in their clonal origins. Genetic abnormalities known to confer functional autonomy, such as activating TSH-R or $G_{S\alpha}$ mutations (see below), are not usually found in the autonomous regions of toxic MNG goiter.

In addition to features of goiter, the clinical presentation of toxic MNG includes subclinical hyperthyroidism or mild thyrotoxicosis. The patient is usually elderly and may present with atrial fibrillation or palpitations, tachycardia, nervousness, tremor, or weight loss. Recent exposure to iodine, from contrast dyes or other sources, may precipitate or exacerbate thyrotoxicosis. The TSH level is low. The T4 level may be normal or minimally increased; T_3 is often elevated to a greater degree than T_4. Thyroid scan shows heterogeneous uptake with multiple regions of increased and decreased uptake; 24-hour uptake of radioiodine may not be increased.

TREATMENT: TOXIC MULTINODULAR GOITER

The management of toxic MNG is challenging. Antithyroid drugs, often in combination with beta blockers, can normalize thyroid function and address clinical features of thyrotoxicosis. This treatment, however, often stimulates the growth of the goiter, and in contrast to Graves' disease, spontaneous remission does not occur. Radioiodine can be used to treat areas of autonomy as well as to decrease the mass of the goiter. Usually, however, some degree of autonomy remains, presumably because multiple autonomous regions emerge as soon as others are treated. Nonetheless, a trial of radioiodine should be considered before subjecting patients, many of whom are elderly, to surgery. Surgery provides definitive treatment of underlying thyrotoxicosis as well as goiter. Patients should be rendered euthyroid using an antithyroid drug before operation.

Hyperfunctioning Solitary Nodule

A solitary, autonomously functioning thyroid nodule is referred to as *toxic adenoma*. The pathogenesis of this disorder has been unraveled by demonstrating the functional effects of mutations that stimulate the TSH-R signaling pathway. Most patients with solitary hyperfunctioning nodules have acquired somatic, activating mutations in the TSH-R **(Fig. 341-10)**. These mutations, located primarily in the receptor transmembrane domain, induce constitutive receptor coupling to $G_{S\alpha}$, increasing cyclic AMP levels and leading to enhanced thyroid follicular cell proliferation and function. Less commonly, somatic mutations are identified in $G_{S\alpha}$. These mutations, which are similar to those seen in McCune-Albright syndrome (Chap. 347) or in a subset of somatotrope adenomas (Chap. 339), impair GTP hydrolysis, also causing constitutive activation of the cyclic AMP signaling pathway. In most series, activating mutations in either the TSH-R or the $G_{S\alpha}$ subunit genes are identified in >90% of patients with solitary hyperfunctioning nodules.

Figure 341-10 Activating Mutations of the TSH-R.

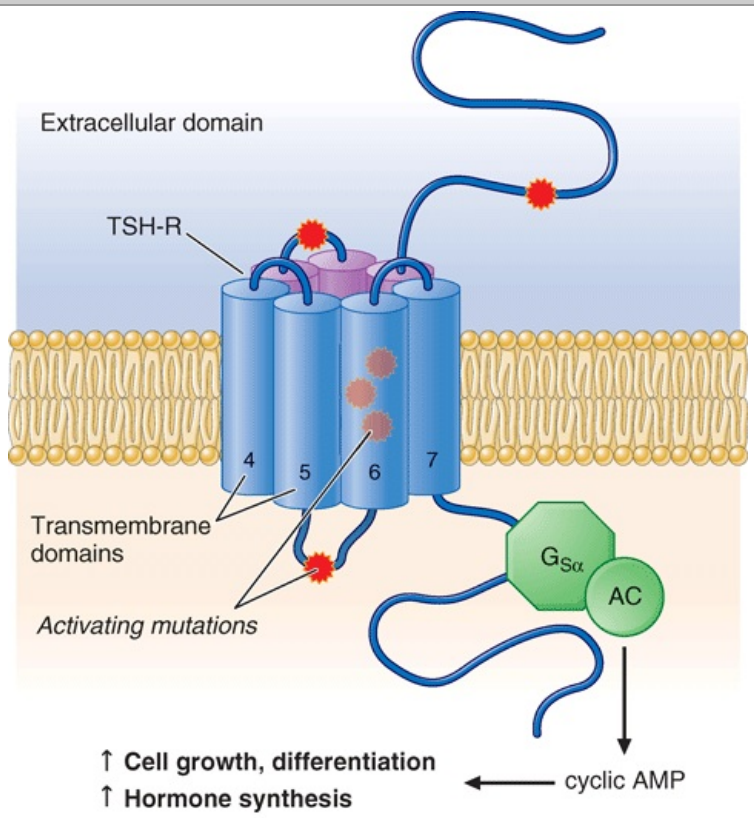

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mutations (*) that activate the thyroid-stimulating hormone receptor (TSH-R) reside mainly in transmembrane 5 and intracellular loop 3, though mutations have occurred in a variety of different locations. The effect of these mutations is to induce conformational changes that mimic TSH binding, thereby leading to coupling to stimulatory G protein ($G_{S\alpha}$) and activation of adenylate cyclase (AC), an enzyme that generates cyclic AMP.

Thyrotoxicosis is usually mild. The disorder is suggested by the presence of the thyroid nodule, which is generally large enough to be palpable, and by the absence of clinical features suggestive of Graves' disease or other causes of thyrotoxicosis. A thyroid scan provides a definitive diagnostic test, demonstrating focal uptake in the hyperfunctioning nodule and diminished uptake in the remainder of the gland, as activity of the normal thyroid is suppressed.

TREATMENT: HYPERFUNCTIONING SOLITARY NODULE

Radioiodine ablation is usually the treatment of choice. Because normal thyroid function is suppressed, ^{131}I is concentrated in the hyperfunctioning nodule with minimal uptake and damage to normal thyroid tissue. Relatively large radioiodine doses [e.g., 370–1110 MBq (10–29.9 mCi) ^{131}I] have been shown to correct thyrotoxicosis in about 75% of patients within 3 months. Hypothyroidism occurs in <10% of those patients over the next 5 years. Surgical resection is also effective and is usually limited to enucleation of the adenoma or lobectomy, thereby preserving thyroid function and minimizing risk of hypoparathyroidism or damage to the recurrent laryngeal nerves. Medical therapy using antithyroid drugs and beta blockers can normalize thyroid function but is not an optimal long-term treatment. Using ultrasound guidance, repeated ethanol injections, or percutaneous radiofrequency thermal ablation have been used successfully in some centers to ablate hyperfunctioning nodules, and these techniques have also been used to reduce the size of nonfunctioning thyroid nodules.

BENIGN NEOPLASMS

The various types of benign thyroid nodules are listed in **Table 341-9**. These lesions are common (5–10% adults), particularly when assessed by sensitive techniques such as ultrasound. The risk of malignancy is very low for *macrofollicular adenomas* and *normofollicular adenomas*. *Microfollicular, trabecular, and Hürthle cell variants* raise greater concern, and the histology is more difficult to interpret. About one-third of palpable nodules are *thyroid cysts*. These may be recognized by their ultrasound appearance or based on aspiration of large amounts of pink or straw-colored fluid (colloid). Many are mixed cystic/solid lesions, in which case it is desirable to aspirate cellular components under ultrasound or harvest cells after cytospin of cyst fluid. Cysts frequently recur, even after repeated aspiration, and may require surgical excision if they are large or if the cytology is suspicious. Sclerosis has been used with variable success but is often painful and may be complicated by infiltration of the sclerosing agent.

Table 341-9 Classification of Thyroid Neoplasms

Benign	
Follicular epithelial cell adenomas	
Macrofollicular (colloid)	
Normofollicular (simple)	
Microfollicular (fetal) Trabecular (embryonal)	
Hürthle cell variant (oncocytic)	
Malignant	**Approximate Prevalence, %**
Follicular epithelial cell	
Well-differentiated carcinomas	
Papillary carcinomas	80–90
Pure papillary	
Follicular variant Diffuse sclerosing variant	
Tall cell, columnar cell variants	
Follicular carcinomas	5–10
Minimally invasive	
Widely invasive	
Hürthle cell carcinoma (oncocytic)	
Insular carcinoma	
Undifferentiated (anaplastic) carcinomas C cell (calcitonin-producing)	
Medullary thyroid cancer	<10
Sporadic	
Familial	
MEN 2	
Other malignancies	
Lymphomas	1–2
Sarcomas Metastases	

Others	

Note: MEN, multiple endocrine neoplasia.

The treatment approach for benign nodules is similar to that for MNG. TSH suppression with levothyroxine decreases the size of about 30% of nodules and may prevent further growth. If a nodule has not decreased in size after 6–12 months of suppressive therapy, treatment should be discontinued because little benefit is likely to accrue from long-term treatment; the risk of iatrogenic subclinical thyrotoxicosis should also be considered.

THYROID CANCER

Thyroid carcinoma is the most common malignancy of the endocrine system. Malignant tumors derived from the follicular epithelium are classified according to histologic features. Differentiated tumors, such as papillary thyroid cancer (PTC) or follicular thyroid cancer (FTC), are often curable, and the prognosis is good for patients identified with early-stage disease. In contrast, anaplastic thyroid cancer (ATC) is aggressive, responds poorly to treatment, and is associated with a bleak prognosis.

The incidence of thyroid cancer (~9/100,000 per year) increases with age, plateauing after about age 50 **(Fig. 341-11)**. Age is also an important prognostic factor—thyroid cancer at a young age (<20) or in older persons (>45) is associated with a worse prognosis. Thyroid cancer is twice as common in women as men, but male gender is associated with a worse prognosis. Additional important risk factors include a history of childhood head or neck irradiation, large nodule size (≥4 cm), evidence for local tumor fixation or invasion into lymph nodes, and the presence of metastases **(Table 341-10)**.

Figure 341-11 Age-Associated Incidence (—♦—) and Mortality (—•—) Rates for Invasive Thyroid Cancer.

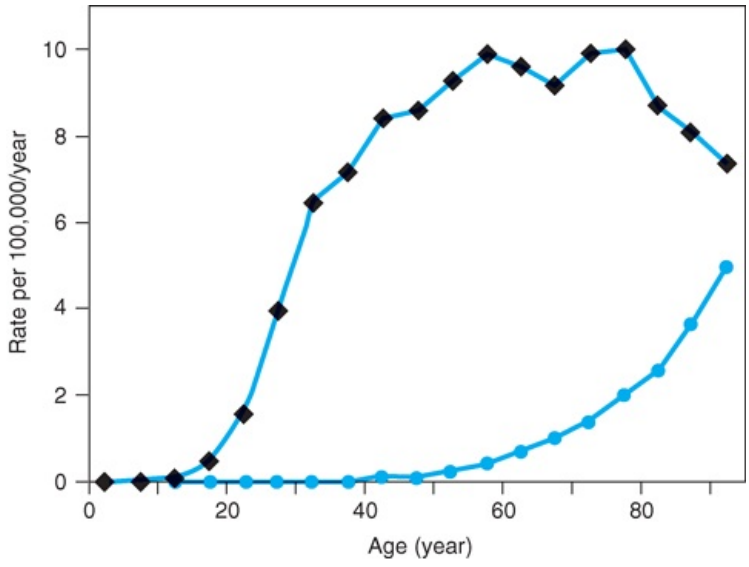

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

[Adapted from LAG Ries et al (eds): SEER Cancer Statistics Review, 1973–1996, Bethesda, National Cancer Institute, 1999.]

Table 341-10 Risk Factors for Thyroid Carcinoma in Patients with Thyroid Nodule

History of head and neck irradiation Age <20 or >45 years
Bilateral disease
Increased nodule size (>4 cm)
New or enlarging neck mass
Male gender
Family history of thyroid cancer or MEN 2
Vocal cord paralysis, hoarse voice
Nodule fixed to adjacent structures Extrathyroidal extension

Suspected lymph node involvement	
Iodine deficiency (follicular cancer)	

Abbreviation: MEN, multiple endocrine neoplasia.

Several unique features of thyroid cancer facilitate its management: (1) thyroid nodules are readily palpable, allowing early detection and biopsy by FNA; (2) iodine radioisotopes can be used to diagnose (^{123}I) and treat (^{131}I) differentiated thyroid cancer, reflecting the unique uptake of this anion by the thyroid gland; and (3) serum markers allow the detection of residual or recurrent disease, including the use of Tg levels for PTC and FTC and calcitonin for medullary thyroid cancer (MTC).

Classification

Thyroid neoplasms can arise in each of the cell types that populate the gland, including thyroid follicular cells, calcitonin-producing C cells, lymphocytes, and stromal and vascular elements, as well as metastases from other sites (Table 341-9). The American Joint Committee on Cancer (AJCC) has designated a staging system using the TNM classification **(Table 341-11)**. Several other classification and staging systems are also widely used, some of which place greater emphasis on histologic features or risk factors such as age or gender.

Table 341-11 Thyroid Cancer Classification[a]

Papillary or follicular thyroid cancers		
	<45 years	>45 years
Stage I	Any T, any N, M0	T1, N0, M0
Stage II	Any T, any N, M1	T2 or T3, N0, M0
Stage III	—	T4, N0, M0 Any T, N1, M0
Stage IV	—	Any T, any N, M1
Anaplastic thyroid cancer		
Stage IV	All cases are stage IV	
Medullary thyroid cancer		
Stage I	T1, N0, M0	
Stage II	T2–T4, N0, M0	
Stage III	Any T, N1, M0	
Stage IV	Any T, any N, M1	

[a]Criteria include: T, the size and extent of the primary tumor (T1 ≤ 1 cm; 1 cm < T2 ≤ 4 cm; T3 > 4 cm; T4 direct invasion through the thyroid capsule); N, the absence (N0) or presence (N1) of regional node involvement; M, the absence (M0) or presence (M1) of metastases.

Source: American Joint Committee on Cancer staging system for thyroid cancers using the TNM classification.

Pathogenesis and Genetic Basis

RADIATION

Early studies of the pathogenesis of thyroid cancer focused on the role of external radiation, which predisposes to chromosomal breaks, leading to genetic rearrangements and loss of tumor-suppressor genes. External radiation of the mediastinum, face, head, and neck region was administered in the past to treat an array of conditions, including acne and enlargement of the thymus, tonsils, and adenoids. Radiation exposure increases the risk of benign and malignant thyroid nodules, is associated with multicentric cancers, and shifts the incidence of thyroid cancer to an earlier age group. Radiation from nuclear fallout also increases the risk of thyroid cancer. Children seem more predisposed to the effects of radiation than adults. Of note, radiation derived from ^{131}I therapy appears to contribute minimal increased risk of thyroid cancer.

TSH AND GROWTH FACTORS

Many differentiated thyroid cancers express TSH receptors and, therefore, remain responsive to TSH. This observation provides the rationale for T4 suppression of TSH in patients with thyroid cancer. Residual expression of TSH receptors also allows TSH-stimulated uptake of ^{131}I therapy (see below).

ONCOGENES AND TUMOR-SUPPRESSOR GENES

Thyroid cancers are monoclonal in origin, consistent with the idea that they originate as a consequence of mutations that confer a growth advantage to a single cell. In addition to increased rates of proliferation, some thyroid cancers exhibit impaired apoptosis and features that enhance

invasion, angiogenesis, and metastasis. Thyroid neoplasms have been analyzed for a variety of genetic alterations, but without clear evidence of an ordered acquisition of somatic mutations as they progress from the benign to the malignant state. On the other hand, certain mutations are relatively specific for thyroid neoplasia, some of which correlate with histologic classification **(Table 341-12)**.

Table 341-12 Genetic Alterations in Thyroid Neoplasia

Gene/Protein	Type of Gene	Chromosomal Location	Genetic Abnormality	Tumor
TSH receptor	GPCR receptor	14q31	Point mutations	Toxic adenoma, differentiated carcinomas
$G_{S\alpha}$	G protein	20q13.2	Point mutations	Toxic adenoma, differentiated carcinomas
RET/PTC	Receptor tyrosine kinase	10q11.2	Rearrangements PTC1: (inv(10)q11.2q21)PTC2: (t(10;17)(q11.2;q23))PTC3: ELE1/TK	PTC
RET	Receptor tyrosine kinase	10q11.2	Point mutations	MEN 2, medullary thyroid cancer
BRAF	MEK kinase	7q24	Point mutations, rearrangements	PTC, ATC
TRK	Receptor tyrosine kinase	1q23-24	Rearrangements	Multinodular goiter, papillary thyroid cancer
RAS	Signal transducing p21	Hras 11p15.5Kras 12p12.1; Nras 1p13.2	Point mutations	Differentiated thyroid carcinoma, adenomas
p53	Tumor suppressor, cell cycle control, apoptosis	17p13	Point mutations Deletion, insertion	Anaplastic cancer
APC	Tumor suppressor, adenomatous polyposis coli gene	5q21-q22	Point mutations	Anaplastic cancer, also associated with familial polyposis coli
p16 (MTS1, CDKN2A)	Tumor suppressor, cell cycle control	9p21	Deletions	Differentiated carcinomas
p21/WAF	Tumor suppressor, cell cycle control	6p21.2	Overexpression	Anaplastic cancer
MET	Receptor tyrosine kinase	7q31	Overexpression	Follicular thyroid cancer
c-MYC	Receptor tyrosine kinase	8q24.12.-13	Overexpression	Differentiated carcinoma
PTEN	Phosphatase	10q23	Point mutations	PTC in Cowden's syndrome (multiple hamartomas, breast tumors, gastrointestinal polyps, thyroid tumors)
CTNNB1	β-Catenin	3p22	Point mutations	Anaplastic cancer
Loss of heterozygosity (LOH)	Tumor suppressors	3p; 11q13, other loci	Deletions	Differentiated thyroid carcinomas, anaplastic cancer
PAX8-PPARγ 1	Transcription factor Nuclear receptor fusion	t(2;3)(q13;p25)	Translocation	Follicular adenoma or carcinoma

Abbreviations: TSH, thyroid-stimulating hormone; $G_{S\alpha}$, G-protein stimulating α-subunit; RET, rearranged during transfection proto-oncogene; PTC, papillary thyroid cancer; TRK, tyrosine kinase receptor; RAS, rat sarcoma proto-oncogene; p53, p53 tumor suppressor gene; MET, met proto-oncogene (hepatocyte growth factor receptor); c-MYC, cellular homologue of myelocytomatosis virus

proto-oncogene; PTEN, phosphatase and tensin homologue; APC, adenomatous polyposis coli; MTS, multiple tumor suppressor; CDKN2A, cyclin-dependent kinase inhibitor 2A; P21, p21 tumor suppressor; WAF, wild-type p53 activated fragment; GPCR, G protein–coupled receptor; ELE1/TK, RET-activating gene ele1/tyrosine kinase; MEN 2, multiple endocrine neoplasia-2; PAX8, paired domain transcription factor; PPARγ1, peroxisome-proliferator activated receptor γ1; BRAF, v-raf homologue, B1; MEK, mitogen extracellular signal-regulated kinase.

Source: Adapted with permission from P Kopp, JL Jameson, in JL Jameson (ed): Principles of Molecular Medicine. Totowa, NJ, Humana Press, 1998.

As described above, activating mutations of the TSH-R and the $G_{S\alpha}$ subunit are associated with autonomously functioning nodules. Though these mutations induce thyroid cell growth, this type of nodule is almost always benign.

Activation of the RET-RAS-BRAF signaling pathway is seen in most PTCs, though the types of mutations are heterogeneous. A variety of rearrangements involving the *RET* gene on chromosome 10 brings this receptor tyrosine kinase under the control of other promoters, leading to receptor overexpression. *RET* rearrangements occur in 20–40% of PTCs in different series and were observed with increased frequency in tumors developing after the Chernobyl radiation accident. Rearrangements in PTC have also been observed for another tyrosine kinase gene, *TRK1*, which is located on chromosome 1. To date, the identification of PTC with *RET* or *TRK1* rearrangements has not proven useful for predicting prognosis or treatment responses. *BRAF* mutations appear to be the most common genetic alteration in PTC. These mutations activate the kinase, which stimulates the mitogen-activated protein MAP kinase (MAPK) cascade. *RAS* mutations, which also stimulate the MAPK cascade, are found in about 20–30% of thyroid neoplasms, including both PTC and FTC. Of note, simultaneous *RET*, *BRAF*, and *RAS* mutations do not occur in the same tumor, suggesting that activation of the MAPK cascade is critical for tumor development, independent of the step that initiates the cascade.

RAS mutations also occur in FTCs. In addition, a rearrangement of the thyroid developmental transcription factor PAX8 with the nuclear receptor PPARγ is identified in a significant fraction of FTCs. Loss of heterozygosity of 3p or 11q, consistent with deletions of tumor-suppressor genes, is also common in FTCs.

Most of the mutations seen in differentiated thyroid cancers have also been detected in ATCs. BRAF mutations are seen in up to 50% of ATCs. Mutations in CTNNB1, which encodes β-catenin, occur in about two-thirds of ATCs, but not in PTC or FTC. Mutations of the tumor suppressor p53 also play an important role in the development of ATC. Because p53 plays a role in cell cycle surveillance, DNA repair, and apoptosis, its loss may contribute to the rapid acquisition of genetic instability as well as poor treatment responses (Chap. 84) (Table 341-12).

The role of molecular diagnostics in the clinical management of thyroid cancer is under investigation. In principle, analyses of specific mutations might aid in classification, prognosis, or choice of treatment. However, there is no clear evidence to date that this information alters clinical decision making.

MTC, when associated with multiple endocrine neoplasia (MEN) type 2, harbors an inherited mutation of the *RET* gene. Unlike the rearrangements of *RET* seen in PTC, the mutations in MEN2 are point mutations that induce constitutive activity of the tyrosine kinase (Chap. 351). MTC is preceded by hyperplasia of the C cells, raising the likelihood that as-yet-unidentified "second hits" lead to cellular transformation. A subset of sporadic MTC contain somatic mutations that activate *RET*.

Well-Differentiated Thyroid Cancer

PAPILLARY

PTC is the most common type of thyroid cancer, accounting for 70–90% of well-differentiated thyroid malignancies. Microscopic PTC is present in up to 25% of thyroid glands at autopsy, but most of these lesions are very small (several millimeters) and are not clinically significant. Characteristic cytologic features of PTC help make the diagnosis by FNA or after surgical resection; these include psammoma bodies, cleaved nuclei with an "orphan-Annie" appearance caused by large nucleoli, and the formation of papillary structures.

PTC tends to be multifocal and to invade locally within the thyroid gland as well as through the thyroid capsule and into adjacent structures in the neck. It has a propensity to spread via the lymphatic system but can metastasize hematogenously as well, particularly to bone and lung. Because of the relatively slow growth of the tumor, a significant burden of pulmonary metastases may accumulate, sometimes with remarkably few symptoms. The prognostic implication of lymph node spread is debated. Lymph node involvement by thyroid cancer can be well tolerated but appears to increase the risk of recurrence and mortality, particularly in older patients. The staging of PTC by the TNM system is outlined in Table 341-11. Most papillary cancers are identified in the early stages (>80% stages I or II) and have an excellent prognosis, with survival curves similar to expected survival **(Fig. 341-12*A*)**. Mortality is markedly increased in stage IV disease (distant metastases), but this group comprises only about 1% of patients. The treatment of PTC is described below.

Figure 341-12 Survival Rates in Patients with Differentiated Thyroid Cancer.

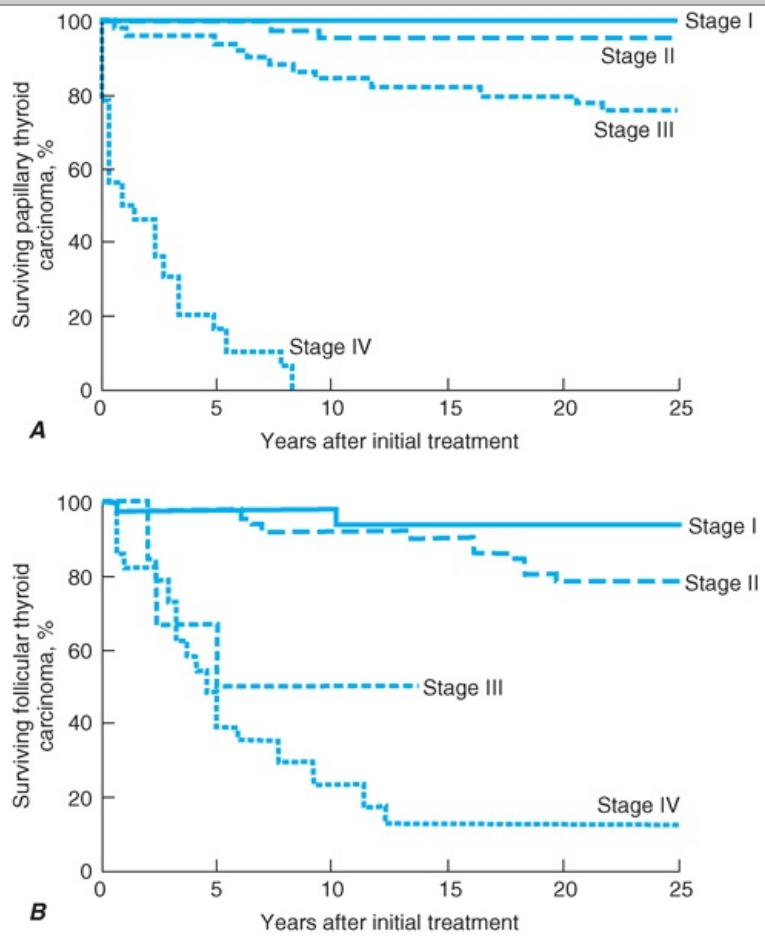

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Papillary cancer, cohort of 1851 patients. I, 1107 (60%); II, 408 (22%); III, 312 (17%); IV, 24 (1%); *n* = 1185. ***B.*** Follicular cancer, cohort of 153 patients. I, 42 (27%); II, 82 (54%); III, 6 (4%); IV, 23 (15%); *n* = 153.

[Adapted from PR Larsen et al: William's Textbook of Endocrinology, 9th ed, JD Wilson et al (eds). Philadelphia, Saunders, 1998, pp 389–575, with permission.]

FOLLICULAR

The incidence of FTC varies widely in different parts of the world; it is more common in iodine-deficient regions. FTC is difficult to diagnose by FNA because the distinction between benign and malignant follicular neoplasms rests largely on evidence of invasion into vessels, nerves, or adjacent structures. FTC tends to spread by hematogenous routes leading to bone, lung, and central nervous system metastases. Mortality rates associated with FTC are less favorable than for PTC, in part because a larger proportion of patients present with stage IV disease (Fig. 341-12*B*). Poor prognostic features include distant metastases, age >50 years, primary tumor size >4 cm, Hürthle cell histology, and the presence of marked vascular invasion.

TREATMENT: WELL-DIFFERENTIATED THYROID CANCER

Surgery

All well-differentiated thyroid cancers should be surgically excised. In addition to removing the primary lesion, surgery allows accurate histologic diagnosis and staging, and multicentric disease is commonly found in the contralateral thyroid lobe. Lymph node spread can also be assessed at the time of surgery, and involved nodes can be removed. Recommendations about the extent of surgery vary for stage I disease, as survival rates are similar for lobectomy and near-total thyroidectomy. Lobectomy is associated with a lower incidence of hypoparathyroidism and injury to the recurrent laryngeal nerves. However, it is not possible to monitor Tg levels or to perform whole-body ^{131}I scans in the presence of the residual lobe. Moreover, if final staging or subsequent follow-up indicates the need for radioiodine scanning or treatment, repeat surgery is necessary to remove the remaining thyroid tissue. Therefore, near-total thyroidectomy is preferable in almost all patients; complication rates are acceptably low if the surgeon is highly experienced in the procedure. Postsurgical radioablation of the remnant thyroid

tissue is increasingly being used because it may destroy remaining or multifocal thyroid carcinoma, and it facilitates the use of Tg determinations and radioiodine scanning for long-term follow-up by eliminating residual normal or neoplastic tissue.

TSH SUPPRESSION THERAPY

As most tumors are still TSH-responsive, levothyroxine suppression of TSH is a mainstay of thyroid cancer treatment. Though TSH suppression clearly provides therapeutic benefit, there are no prospective studies that identify the optimal level of TSH suppression. A reasonable goal is to suppress TSH as much as possible without subjecting the patient to unnecessary side effects from excess thyroid hormone, such as atrial fibrillation, osteopenia, anxiety, and other manifestations of thyrotoxicosis. For patients at low risk of recurrence, TSH should be suppressed into the low but detectable range (0.1–0.5 mIU/L). For patients at high risk of recurrence or with known metastatic disease, complete TSH suppression is indicated if there are no strong contraindications to mild thyrotoxicosis. In this instance, unbound T4 must also be monitored to avoid excessive treatment.

RADIOIODINE TREATMENT

Well-differentiated thyroid cancer still incorporates radioiodine, though less efficiently than normal thyroid follicular cells. Radioiodine uptake is determined primarily by expression of the NIS and is stimulated by TSH, requiring expression of the TSH-R. The retention time for radioactivity is influenced by the extent to which the tumor retains differentiated functions such as iodide trapping and organification. After near-total thyroidectomy, substantial thyroid tissue often remains, particularly in the thyroid bed and surrounding the parathyroid glands. Consequently, ^{131}I ablation is necessary to eliminate remaining normal thyroid tissue and to treat residual tumor cells.

Indications

The use of therapeutic doses of radioiodine remains an area of controversy in thyroid cancer management. However, postoperative thyroid ablation and radioiodine treatment of known residual PTC or FTC clearly reduces recurrence rates but has a smaller impact on mortality, particularly in patients at relatively low risk. This low-risk group includes most patients with stage 1 PTC with primary tumors <1.5 cm in size. For patients with larger papillary tumors, spread to the adjacent lymph nodes, FTC, or evidence of metastases, thyroid ablation and radioiodine treatment are generally indicated.

^{131}I Thyroid Ablation and Treatment

As noted above, the decision to use ^{131}I for thyroid ablation should be coordinated with the surgical approach, as radioablation is much more effective when there is minimal remaining normal thyroid tissue. A typical strategy is to treat the patient for several weeks postoperatively with liothyronine (25 μg bid or tid), followed by thyroid hormone withdrawal. Ideally, the TSH level should increase to >50 mU/L over 3–4 weeks. The level to which TSH rises is dictated largely by the amount of normal thyroid tissue remaining postoperatively. Recombinant human TSH (rhTSH) has also been used to enhance ^{131}I uptake for postsurgical ablation. It appears to be at least as effective as thyroid hormone withdrawal and should be particularly useful as residual thyroid tissue prevents an adequate endogenous TSH rise.

A pretreatment scanning dose of ^{131}I [usually 111–185 MBq (3–5 mCi)] can reveal the amount of residual tissue and provides guidance about the dose needed to accomplish ablation. However, because of concerns about radioactive "stunning" that impairs subsequent treatment, there is a trend to avoid pretreatment scanning and to proceed directly to ablation, unless there is suspicion that the amount of residual tissue will alter therapy. A maximum outpatient ^{131}I dose is 1110 MBq (29.9 mCi) in the United States, though ablation is often more complete using greater doses [1850–3700 MBq (50–100 mCi)]. Patients should be placed on a low-iodine diet (<50 μg/d urinary iodine) to increase radioiodine uptake. In patients with known residual cancer, the larger doses ensure thyroid ablation and may destroy remaining tumor cells. A whole-body scan following the high-dose radioiodine treatment is useful to identify possible metastatic disease.

Follow-Up Whole-Body Thyroid Scanning and Thyroglobulin Determinations

An initial whole-body scan should be performed about 6 months after thyroid ablation. The strategy for follow-up management of thyroid cancer has been altered by the availability of rhTSH to stimulate ^{131}I uptake and by the improved sensitivity of Tg assays to detect residual or recurrent disease. A scheme for using either rhTSH or thyroid hormone withdrawal for thyroid scanning is summarized in **Fig. 341-13**. After thyroid ablation, rhTSH can be used in follow-up to stimulate Tg and ^{131}I uptake without subjecting patients to thyroid hormone withdrawal and its associated symptoms of hypothyroidism as well as the risk of tumor growth after prolonged TSH stimulation. Alternatively, in patients who are likely to require ^{131}I treatment, the traditional approach of thyroid hormone withdrawal can be used to increase TSH. This involves switching patients from levothyroxine (T4) to the more rapidly cleared hormone liothyronine (T_3), thereby allowing TSH to increase more quickly. Because TSH stimulates

Tg levels, Tg measurements should be obtained after administration of rhTSH or when TSH levels have risen after thyroid hormone withdrawal.

In low-risk patients who have no clinical evidence of residual disease after ablation and a basal Tg <1 ng/ml, increasing evidence supports the use of rhTSH-stimulated Tg levels one year after ablation, without the need for radioiodine scanning. If stimulated Tg levels are low (<2 ng/ml) and, ideally, undetectable, these patients can be managed with suppressive therapy and measurements of unstimulated Tg every 6–12 months. The absence of Tg antibodies should be confirmed in these patients. On the other hand, patients with residual disease on whole-body scanning or those with elevated Tg levels require additional ^{131}I therapy. In addition, most authorities advocate radioiodine treatment for scan-negative, Tg-positive (Tg >5–10 ng/mL) patients, as many derive therapeutic benefit from a large dose of ^{131}I.

Figure 341-13 Use of Recombinant Human Thyroid-Stimulating Hormone (TSH) in the Follow-Up of Patients with Thyroid Cancer.

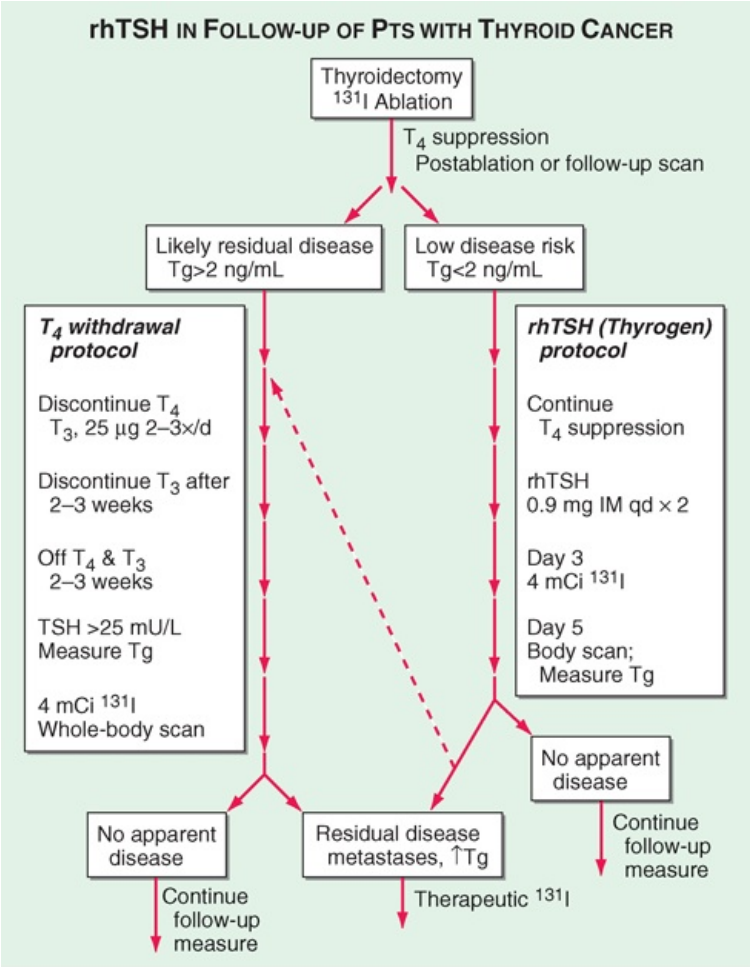

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tg, thyroglobulin; rhTSH, recombinant human TSH.

In addition to radioiodine, external beam radiotherapy is also used to treat specific metastatic lesions, particularly when they cause bone pain or threaten neurologic injury (e.g., vertebral metastases).

New Potential Therapies

Kinase inhibitors are being explored as a means to target pathways known to be active in thyroid cancer, including the Ras, BRAF, EGFR, VEGFR, and angiogenesis pathways. Partial responses have been seen in small trials using motesaniv, sorafenib, and other agents, but the efficacy of these agents awaits larger studies.

Anaplastic and Other Forms of Thyroid Cancer

ANAPLASTIC THYROID CANCER

As noted above, ATC is a poorly differentiated and aggressive cancer. The prognosis is poor, and most

patients die within 6 months of diagnosis. Because of the undifferentiated state of these tumors, the uptake of radioiodine is usually negligible, but it can be used therapeutically if there is residual uptake. Chemotherapy has been attempted with multiple agents, including anthracyclines and paclitaxel, but it is usually ineffective. External beam radiation therapy can be attempted and continued if tumors are responsive.

THYROID LYMPHOMA

Lymphoma in the thyroid gland often arises in the background of Hashimoto's thyroiditis. A rapidly expanding thyroid mass suggests the possibility of this diagnosis. Diffuse large-cell lymphoma is the most common type in the thyroid. Biopsies reveal sheets of lymphoid cells that can be difficult to distinguish from small-cell lung cancer or ATC. These tumors are often highly sensitive to external radiation. Surgical resection should be avoided as initial therapy because it may spread disease that is otherwise localized to the thyroid. If staging indicates disease outside of the thyroid, treatment should follow guidelines used for other forms of lymphoma (Chap. 110).

Medullary Thyroid Carcinoma

MTC can be sporadic or familial and accounts for about 5% of thyroid cancers. There are three familial forms of MTC: MEN 2A, MEN 2B, and familial MTC without other features of MEN (Chap. 351). In general, MTC is more aggressive in MEN 2B than in MEN 2A, and familial MTC is more aggressive than sporadic MTC. Elevated serum calcitonin provides a marker of residual or recurrent disease. It is reasonable to test all patients with MTC for *RET* mutations, as genetic counseling and testing of family members can be offered to those individuals who test positive for mutations.

The management of MTC is primarily surgical. Unlike tumors derived from thyroid follicular cells, these tumors do not take up radioiodine. External radiation treatment and chemotherapy may provide palliation in patients with advanced disease (Chap. 351).

APPROACH TO THE PATIENT

A Thyroid Nodule

Palpable thyroid nodules are found in about 5% of adults, but the prevalence varies considerably worldwide. Given this high prevalence rate, practitioners commonly identify thyroid nodules. The main goal of this evaluation is to identify, in a cost-effective manner, the small subgroup of individuals with malignant lesions.

Nodules are more common in iodine-deficient areas, in women, and with aging. Most palpable nodules are >1 cm in diameter, but the ability to feel a nodule is influenced by its location within the gland (superficial versus deeply embedded), the anatomy of the patient's neck, and the experience of the examiner. More sensitive methods of detection, such as CT, thyroid ultrasound, and pathologic studies, reveal thyroid nodules in >20% of glands. The presence of these thyroid incidentalomas has led to much debate about how to detect nodules and which nodules to investigate further. Most authorities still rely on physical examination to detect thyroid nodules, reserving ultrasound for monitoring nodule size or as an aid in thyroid biopsy.

An approach to the evaluation of a solitary nodule is outlined in **Fig. 341-14**. Most patients with thyroid nodules have normal thyroid function tests. Nonetheless, thyroid function should be assessed by measuring a TSH level, which may be suppressed by one or more autonomously functioning nodules. If the TSH is suppressed, a radionuclide scan is indicated to determine if the identified nodule is "hot," as lesions with increased uptake are almost never malignant and FNA is unnecessary. Otherwise, FNA biopsy, ideally performed with ultrasound guidance, should be the first step in the evaluation of a thyroid nodule. FNA has good sensitivity and specificity when performed by physicians familiar with the procedure and when the results are interpreted by experienced cytopathologists. The technique is particularly useful for detecting PTC. The distinction of benign and malignant follicular lesions is often not possible using cytology alone.

Figure 341-14 Approach to the Patient with a Thyroid Nodule.

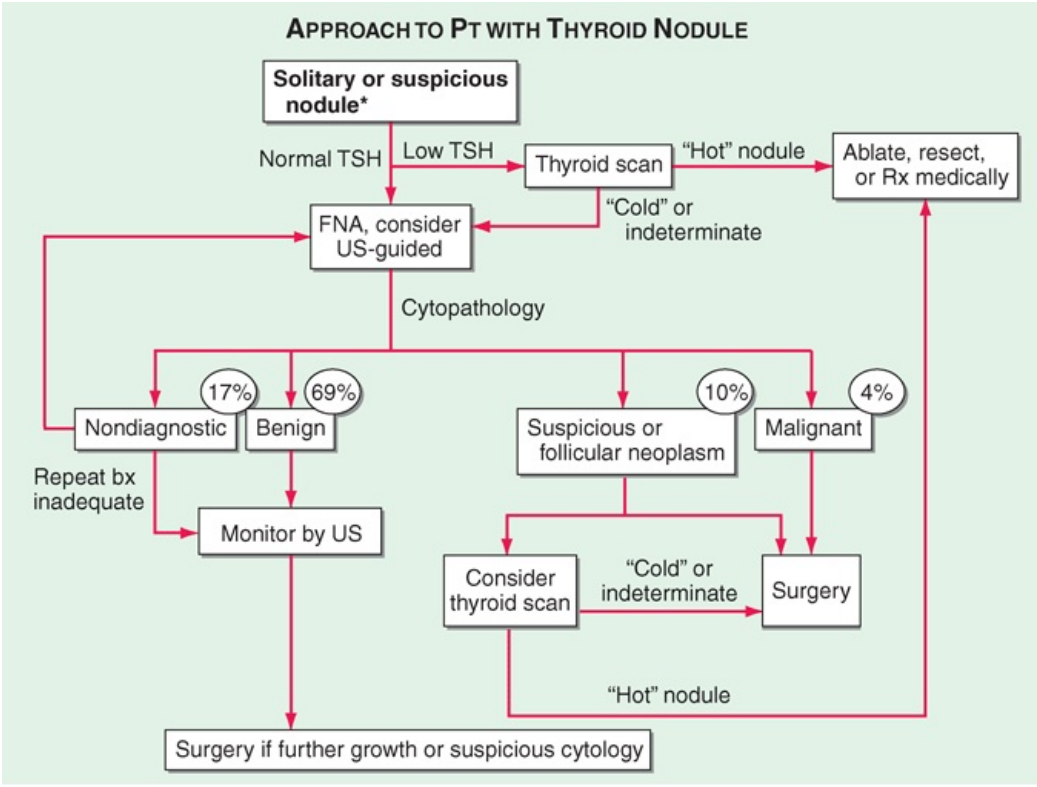

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

See text and references for details. *About one-third of nodules are cystic or mixed solid-cystic. US, ultrasound; TSH, thyroid-stimulating hormone; FNA, fine-needle aspiration.

In several large studies, FNA biopsies yielded the following findings: 70% benign, 10% malignant or suspicious for malignancy, and 20% nondiagnostic or yielding insufficient material for diagnosis. Characteristic features of malignancy mandate surgery. A diagnosis of follicular neoplasm also warrants surgery, as benign and malignant lesions cannot be distinguished based on cytopathology or frozen section. The management of patients with benign lesions is more variable. Many authorities advocate TSH suppression, whereas others monitor nodule size without suppression. With either approach, thyroid nodule size should be monitored, ideally using ultrasound. Repeat FNA is indicated if a nodule enlarges, and a second biopsy should be performed within 2–5 years to confirm the benign status of the nodule.

Nondiagnostic biopsies occur for many reasons, including a fibrotic reaction with relatively few cells available for aspiration, a cystic lesion in which cellular components reside along the cyst margin, or a nodule that may be too small for accurate aspiration. For these reasons, ultrasound-guided FNA is indicated when the FNA is repeated. Ultrasound characteristics are also useful for deciding which nodules to biopsy when multiple nodules are present. Sonographic characteristics suggestive of malignancy include microcalcifications, increased vascularity, and hypoechogenicity within the nodule.

The evaluation of a thyroid nodule is stressful for most patients. They are concerned about the possibility of thyroid cancer, whether verbalized or not. It is constructive, therefore, to review the diagnostic approach and to reassure patients when no malignancy is found. When a suspicious lesion or thyroid cancer is identified, the generally favorable prognosis and available treatment options can be reassuring.

FURTHER READINGS

Abalovich M et al: Management of thyroid dysfunction during pregnancy and postpartum: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 92 (Suppl): S1, 2007

Bahn RS: Graves' ophthalmopathy. N Engl J Med 362:726, 2010[PMID: 20181974] [Full Text]

Bauer M et al: Brain glucose metabolism in hypothyroidism: a positron emission tomography study before and after thyroid hormone replacement therapy. J Clin Endocrinol Metab 94:2922, 2009[PMID: 19435829] [Full Text]

Biondi B, Cooper DS: The clinical significance of subclinical thyroid disease. Endocr Rev 29:76, 2008[PMID: 17991805] [Full Text]

———: Benefits of thyrotropin suppression versus the risks of adverse effects in differentiated thyroid cancer. Thyroid 20:135, 2010

Brent GA: Clinical practice. Graves' disease. N Engl J Med 358:2594, 2008[PMID: 18550875] [Full Text]

DeGroot LJ et al: Thyroid gland, in *Endocrinology*, 6th ed, JL Jameson, LJ DeGroot (eds). Philadelphia, Elsevier Saunders, 2009

Fagin JA, Mitsiades N: Molecular pathology of thyroid cancer: diagnostic and clinical implications. Best Pract Res Clin Endocrinol Metab. 22: 955, 2008[PMID: 19041825] [Full Text]

Hegedüs L: Treatment of Graves' hyperthyroidism: Evidence-based and emerging modalities. Endocrinol Metab Clin North Am. 38:355, 2009

Sherman SI et al: Motesanib diphosphate in progressive differentiated thyroid cancer. N Engl J Med 359:31, 2008[PMID: 18596272] [Full Text]

Stagnaro-Green A: Maternal thyroid disease and preterm delivery. J Clin Endocrinol Metab 94: 21, 2009[PMID: 18984665] [Full Text]

Zeitlin AA et al: Genetic developments in autoimmune thyroid disease: an evolutionary process. Clin Endocrinol 68:671, 2008[PMID: 18081880] [Full Text]

Zimmermann MB: Iodine deficiency. Endocr Rev 30:376, 2009[PMID: 19460960] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE ADRENAL CORTEX: INTRODUCTION

The adrenal cortex produces three classes of corticosteroid hormones: glucocorticoids (e.g., cortisol), mineralocorticoids (e.g., aldosterone), and adrenal androgen precursors (e.g., dehydroepiandrosterone, DHEA) **(Fig. 342-1)**. Glucocorticoids and mineralocorticoids act through specific nuclear receptors, regulating aspects of the physiologic stress response as well as blood pressure and electrolyte homeostasis. Adrenal androgen precursors are converted in the gonads and peripheral target cells to sex steroids that act via nuclear androgen and estrogen receptors.

Figure 342-1

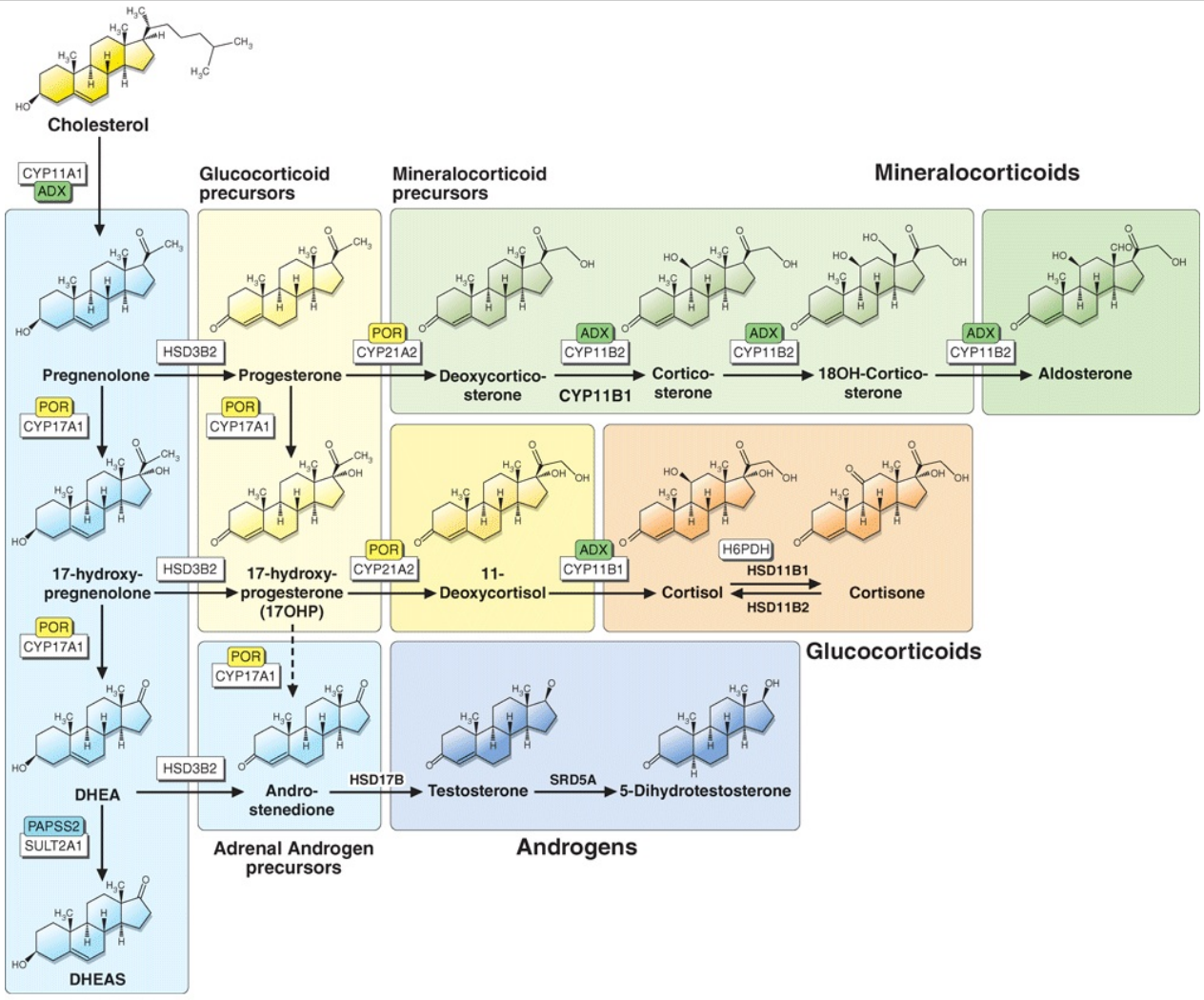

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adrenal steroidogenesis.

CYP11A1, side chain cleavage enzyme; CYP17A1, 17α-hydroxylase/17,20 lyase; POR, P450 oxidoreductase; ADX, adrenodoxin; HSD3B2, 3β-hydroxysteroid dehydrogenase type 2; CYP21A2, 21-hydroxylase; CYP11B1, 11β-hydroxylase; CYP11B2, aldosterone synthase; HSD11B1, 11β-hydroxysteroid dehydrogenase type 1; HSD11B2, 11β-hydroxysteroid dehydrogenase type 2; H6PDH, hexose-6-phosphate dehydrogenase; HSD17B, 17β-hydroxysteroid dehydrogenase; SRD5A, 5 α-reductase; SULT2A1, DHEA sulfotransferase; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulfate; PAPSS2, PAPS synthase type 2.

Disorders of the adrenal cortex are characterized by deficiency or excess of one or several of the three major corticosteroid classes. Hormone deficiency can be caused by inherited glandular or enzymatic disorders or by destruction of the pituitary or adrenal gland by autoimmune disorders, infection, infarction, or by iatrogenic events such as surgery or hormonal suppression. Hormone excess is usually the result of neoplasia, leading to increased production of adrenocorticotropic hormone (ACTH) by the pituitary or neuroendocrine cells (ectopic ACTH), or increased production of glucocorticoids or mineralocorticoids by adrenal nodules. Adrenal nodules are increasingly identified incidentally during abdominal imaging performed for other reasons.

Adrenal Anatomy and Development

The normal adrenal glands weigh 6–11 g each. They are located above the kidneys and have their own blood supply. Arterial blood flows initially to the subcapsular region and then meanders from the outer cortical zona glomerulosa through the

intermediate zona fasciculata to the inner zona reticularis and eventually to the adrenal medulla. The right suprarenal vein drains directly into the vena cava while the left suprarenal vein drains into the left renal vein.

During early embryonic development, the adrenals originate from the urogenital ridge and then separate from gonads and kidneys about the 6th week of gestation. Concordant with the time of sexual differentiation (seventh to ninth week of gestation, see Chap. 349), the adrenal cortex starts to produce cortisol and the adrenal sex steroid precursor DHEA. The orphan nuclear receptors SF1 (steroidogenic factor 1) and DAX1 (dosage-sensitive sex reversal gene 1), among others, play a crucial role during this period of development, as they regulate a multitude of adrenal genes involved in steroidogenesis.

Regulatory Control of Steroidogenesis

Production of glucocorticoids and adrenal androgens is under the control of the hypothalamic-pituitary-adrenal (HPA) axis, whereas mineralocorticoids are regulated by the renin-angiotensin-aldosterone (RAA) system.

Glucocorticoid synthesis is under inhibitory feedback control by the hypothalamus and the pituitary **(Fig. 342-2)**. Hypothalamic release of corticotropin-releasing hormone (CRH) occurs in response to endogenous or exogenous stress. CRH stimulates the cleavage of the 241–amino acid polypeptide pro opiomelanocortin (POMC) by pituitary-specific prohormone convertase, yielding adrenocorticotropic hormone (ACTH). ACTH is released by the corticotrope cells of the anterior pituitary and acts as the pivotal regulator of cortisol synthesis, with additional short-term effects on mineralocorticoid and adrenal androgen synthesis. The release of CRH, and subsequently ACTH, occurs in a pulsatile fashion that follows a circadian rhythm under the control of the hypothalamus, specifically its suprachiasmatic nucleus (SCN), with additional regulation by a complex network of cell-specific clock genes. Reflecting the pattern of ACTH secretion, adrenal cortisol secretion exhibits a distinct circadian rhythm, with peak levels in the morning and low levels in the evening **(Fig. 342-3)**.

Figure 342-2

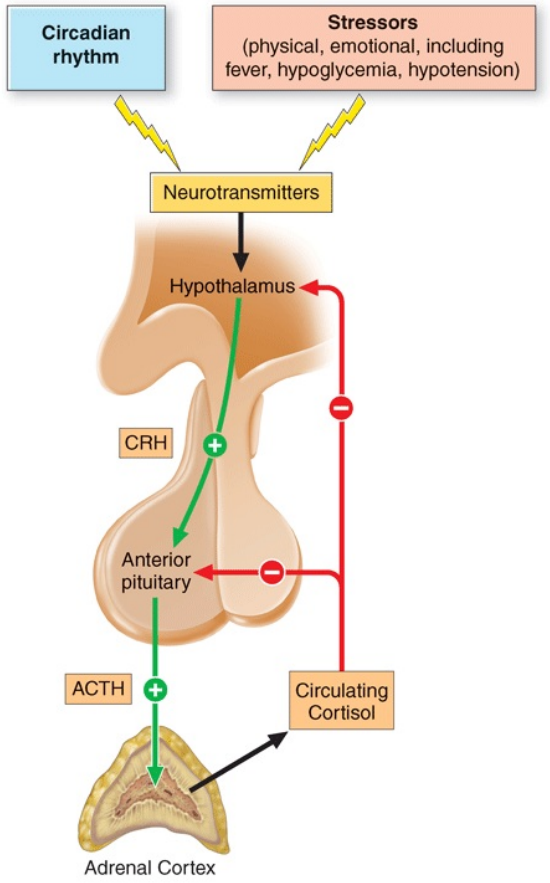

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Regulation of the hypothalamic-pituitary-adrenal (HPA) axis. CRH, corticotropin-releasing hormone; ACTH, adrenocorticotropic hormone.

Figure 342-3

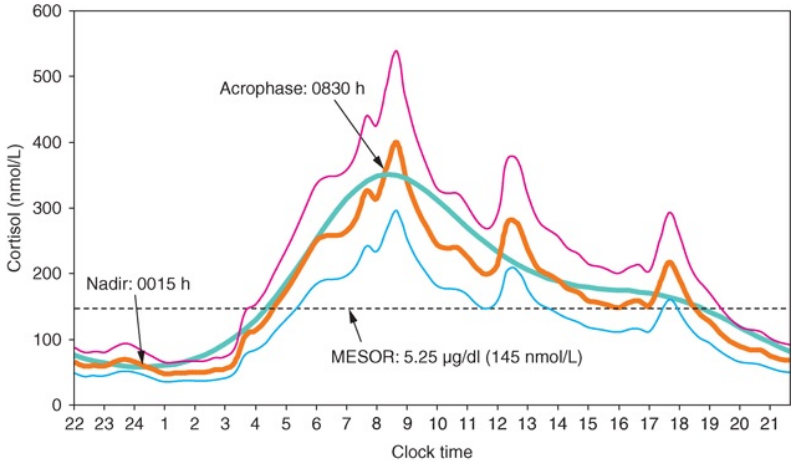

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Physiologic cortisol circadian rhythm.

Circulating cortisol concentrations drop under the rhythm-adjusted mean (MESOR) in the early evening hours, with nadir levels around midnight and a rise in the early morning hours; peak levels are observed ~8:30 A.M. (acrophase). *(Modified after Debono M et al: Modified-release hydrocortisone to provide circadian cortisol profiles. J Clin Endocrinol Metab 94:1548, 2009.)*

Figure 342-4

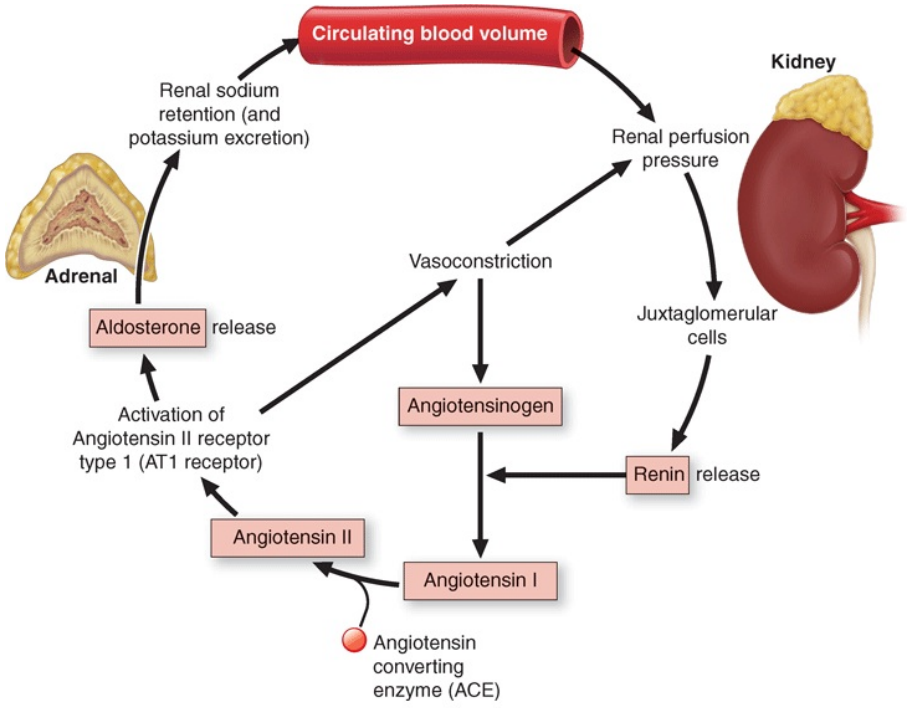

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Regulation of the renin-angiotensin-aldosterone (RAA) system.

Diagnostic tests assessing the HPA axis make use of the fact that it is regulated by negative feedback. Glucocorticoid excess is diagnosed by employing a dexamethasone suppression test. Dexamethasone, a potent glucocorticoid, suppresses CRH/ACTH and, therefore, endogenous cortisol. Various versions of the dexamethasone suppression test are described in detail in Chap. 339. If cortisol production is autonomous (e.g., adrenal nodule), ACTH is already suppressed and dexamethasone has little additional effect. If cortisol production is driven by an ACTH-producing pituitary adenoma, dexamethasone suppression is ineffective at low doses but usually induces suppression at high doses. If cortisol production is driven by an ectopic source of ACTH, the tumors are usually resistant to dexamethasone suppression. Thus, the dexamethasone suppression test is useful to establish the diagnosis of Cushing's syndrome and to assist with the differential diagnosis of cortisol excess.

Conversely, to assess glucocorticoid deficiency, ACTH stimulation of cortisol production is used. The ACTH peptide contains 39 amino acids but the first 24 are sufficient to elicit a physiologic response. The standard ACTH stimulation test involves administration of cosyntropin (ACTH 1-24), 0.25 mg IM or IV, and collection of blood samples at 0, 30, and 60 minutes for cortisol. A normal response is defined as a cortisol level >20 µg/dL or an increment of >10 µg/dL over baseline. A low-dose (1 µg cosyntropin IV) version of this test has been advocated to avoid overstimulation of the adrenal gland. Alternatively, an insulin tolerance test (ITT) can be used to assess adrenal insufficiency. It involves injection of insulin to induce hypoglycemia, which represents a strong stress signal that triggers hypothalamic CRH release and activation of the entire HPA axis. The ITT involves administration of regular insulin 0.1 U/kg IV (dose should be lower if hypopituitarism is likely) and collection of blood samples at

0, 30, 60, and 120 minutes for glucose, cortisol, and growth hormone (GH), if also assessing the GH axis. Oral or IV glucose is administered after the patient has achieved symptomatic hypoglycemia (usually glucose <40 mg/dL). A normal response is defined as a cortisol >20 μg/dL and GH >5.1 μg/L. The ITT requires careful clinical monitoring and sequential measurements of glucose. It is contraindicated in patients with coronary disease, cerebrovascular disease, or seizure disorders, which has made the short cosyntropin test the commonly accepted first-line test.

Mineralocorticoid production is controlled by the RAA regulatory cycle, which is initiated by the release of renin from the juxtaglomerular cells in the kidney, resulting in cleavage of angiotensinogen to angiotensin I in the liver **(Fig. 342-4)**. Angiotensin-converting enzyme (ACE) cleaves angiotensin I to angiotensin II, which binds and activates the angiotensin II receptor type 1 (AT1 receptor), resulting in increased aldosterone production and vasoconstriction. Aldosterone enhances sodium retention and potassium excretion, and increases the arterial perfusion pressure, which in turn regulates renin release. Because mineralocorticoid synthesis is primarily under the control of the RAA system, hypothalamic-pituitary damage does not significantly impact the capacity of the adrenal to synthesize aldosterone.

Similar to the HPA axis, the assessment of the RAA system can be used for diagnostic purposes. If mineralocorticoid excess is present, there is a counter-regulatory downregulation of plasma renin (see below for testing). Conversely, in mineralocorticoid deficiency, plasma renin is markedly increased. Physiologically, oral or IV sodium loading results in suppression of aldosterone, a response that is attenuated or absent in patients with autonomous mineralocorticoid excess.

Steroid Hormone Synthesis, Metabolism, and Action

ACTH stimulation is required for the initiation of steroidogenesis. The ACTH receptor MC2R (melanocortin 2 receptor) interacts with the MC2R-accessory protein MRAP, and the complex is transported to the adrenocortical cell membrane, where it binds to ACTH **(Fig. 342-5)**. ACTH stimulation generates cyclic AMP (cAMP), which upregulates the protein kinase A (PKA) signaling pathway. PKA activation impacts steroidogenesis in three distinct ways: (1) increases the import of cholesterol esters; (2) increases the activity of hormone- sensitive lipase, which cleaves cholesterol esters to cholesterol for import into the mitochondrion; and (3) increases the availability and phosphorylation of CREB (cAMP response element binding), a transcription factor that enhances transcription of CYP11A1 and other enzymes required for glucocorticoid synthesis.

Figure 342-5

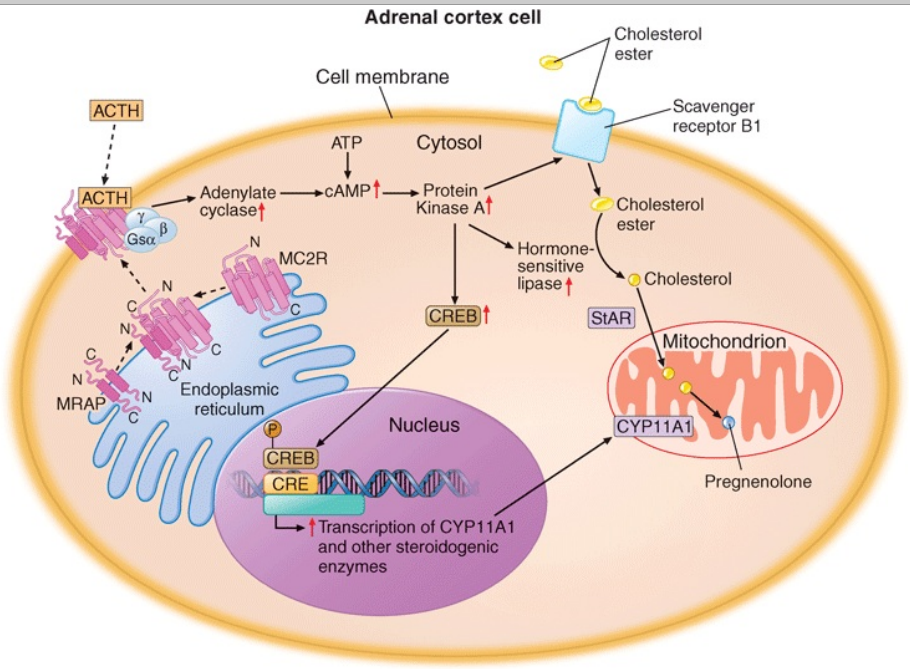

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

ACTH effects on adrenal steroidogenesis. ACTH, adrenocorticotropic hormone; ATP, adenosine triphosphate; CRE, cAMP response element; CREB, cAMP response element binding; MRAP, MC2R-accessory protein; StAR, steroidogenic acute regulatory [protein].

Adrenal steroidogenesis occurs in a zone-specific fashion, with mineralocorticoid synthesis occurring in the outer zona glomerulosa, glucocorticoid synthesis in the zona fasciculata, and adrenal androgen synthesis in the inner zona reticularis (Fig. 342-1). All steroidogenic pathways require cholesterol import into the mitochondrion, a process initiated by the action of the steroidogenic acute regulatory (StAR) protein, which shuttles cholesterol from the outer to the inner mitochondrial membrane. The majority of steroidogenic enzymes are cytochrome P450 (CYP) enzymes, which are either located in the mitochondrion (side chain cleavage enzyme, CYP11A1; 11β-hydroxylase, CYP11B1; aldosterone synthase, CYP11B2) or in the endoplasmic reticulum membrane (17α-hydroxylase, CYP17A1; 21-hydroxylase, CYP21A2; aromatase, CYP19A1). These enzymes require electron donation via specific redox cofactor enzymes, P450 oxidoreductase (POR), and adrenodoxin/adrenodoxin reductase (ADX/ADR) for the microsomal and mitochondrial CYP enzymes, respectively. In addition, the short-chain dehydrogenase 3β-hydroxysteroid dehydrogenase type 2 (3β-HSD2), also termed Δ4,Δ5 isomerase, plays a major role in adrenal steroidogenesis.

The cholesterol side chain cleavage enzyme CYP11A1 generates pregnenolone. Glucocorticoid synthesis requires conversion of pregnenolone to progesterone by 3β-HSD2, followed by conversion to 17-hydroxyprogesterone by CYP17A1, further hydroxylation at carbon 21 by 21-hydroxylase, and eventually, 11β-hydroxylation by CYP11B1 to generate active cortisol (Fig. 342-1). Mineralocorticoid synthesis also requires progesterone, which is first converted to deoxycorticosterone by CYP21A2 and then converted via corticosterone and 18-hydroxycorticosterone to aldosterone in three steps catalyzed by CYP11B2. For adrenal androgen synthesis, pregnenolone undergoes conversion by CYP17A1, which uniquely catalyzes two enzymatic reactions. Via its 17α-hydroxylase activity, CYP17A1 converts pregnenolone to 17-hydroxypregnenolone, followed by generation of the universal sex steroid precursor DHEA via CYP17A1 17,20 lyase activity. The majority of DHEA is secreted by the adrenal

in the form of its sulfate ester, DHEAS, generated by DHEA sulfotransferase (SULT2A1).

Following its release from the adrenal, cortisol circulates in the bloodstream mainly bound to cortisol-binding globulin (CBG) and to a lesser extent to albumin, with only a minor fraction circulating as free, unbound hormone. Free cortisol is thought to enter cells directly, not requiring active transport. In addition, in a multitude of peripheral target tissues of glucocorticoid action, including adipose, liver, muscle, and brain, cortisol is generated from inactive cortisone within the cell by the enzyme 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1) **(Fig. 342-6)**. Thereby, 11β-HSD1 functions as a tissue-specific prereceptor regulator of glucocorticoid action. For the conversion of inactive cortisone to active cortisol, 11β-HSD1 requires nicotinamide adenine dinucleotide phosphate [NADPH (reduced form)], which is provided by the enzyme hexose-6-phosphate dehydrogenase (H6PDH). Like the catalytic domain of 11β-HSD1, H6PDH is located in the lumen of the endoplasmic reticulum, and converts glucose-6-phosphate (G6P) to 6-phosphogluconate (6PGL), thereby regenerating NADP+ to NADPH, which drives the activation of cortisol from cortisone by 11β-HSD1.

Figure 342-6

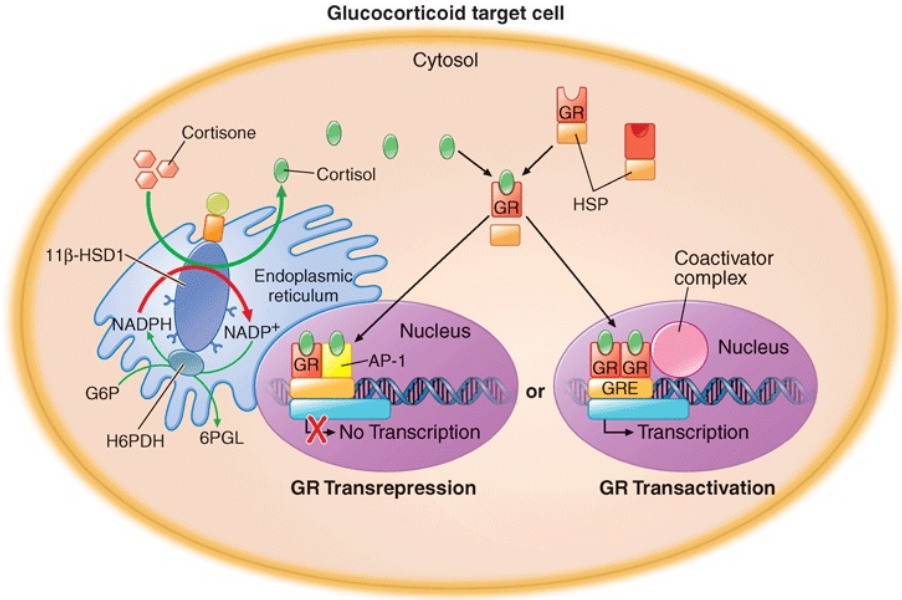

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prereceptor activation of cortisol and glucocorticoid receptor (GR) action. GRE, glucocorticoid response elements; HSP, heat shock proteins; NADPH, nicotinamide adenine dinucleotide phosphate (reduced form).

Figure 342-7

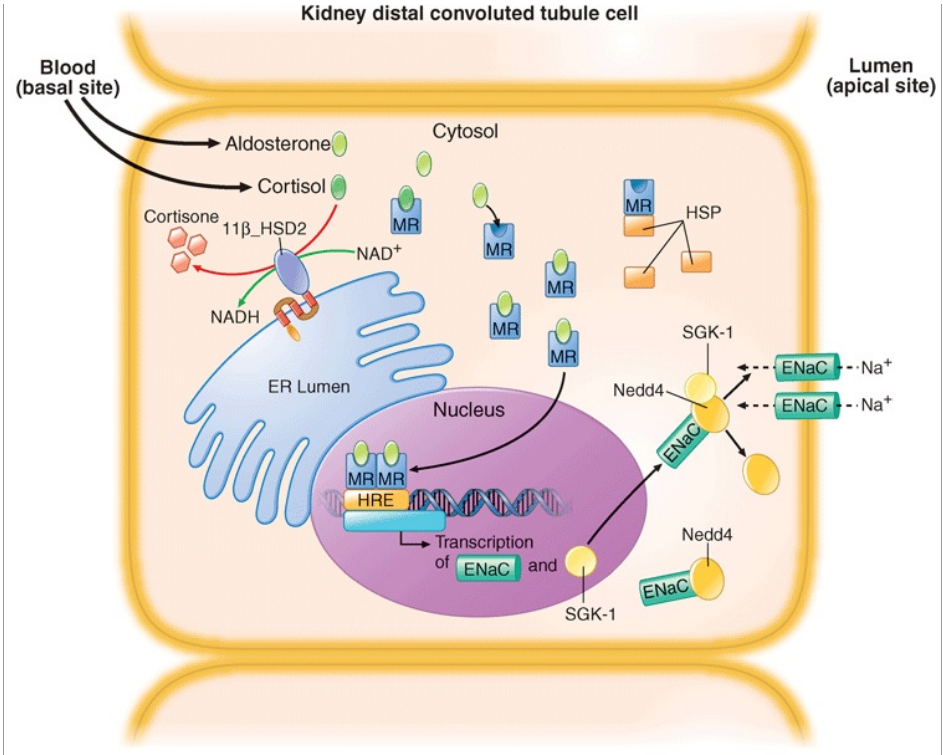

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prereceptor inactivation of cortisol and mineralocorticoid receptor action.

In the cytosol of target cells, cortisol binds and activates the glucocorticoid receptor (GR), which results in dissociation of heat shock proteins (HSP) from the receptor and subsequent dimerization (Fig. 342-6). Cortisol-bound GR dimers translocate to the nucleus and activate glucocorticoid response elements (GRE) in the DNA sequence, thereby enhancing transcription of glucocorticoid-regulated genes (GR transactivation). However, cortisol-bound GR can also form heterodimers with transcription factors such as AP-1 or NF-κB, resulting in transrepression of proinflammatory genes, a mechanism of major importance for the anti-inflammatory action of glucocorticoids. It is important to note that corticosterone also exerts glucocorticoid activity, albeit much weaker than cortisol itself. However, in rodents corticosterone is the major glucocorticoid and in patients with 17-hydroxylase deficiency, lack of cortisol can be compensated for by higher concentrations of corticosterone that accumulates as a consequence of the enzymatic block.

Cortisol is inactivated to cortisone by the microsomal enzyme 11β-hydroxysteroid dehydrogenase type 2 (11β-HSD2) **(Fig. 342-7)**, mainly in the kidney, but also in the colon, salivary glands, and other target tissues. Cortisol and aldosterone bind the mineralocorticoid receptor (MR) with equal affinity; however, cortisol circulates in the bloodstream at about a thousandfold higher concentration. Thus, only rapid inactivation of cortisol to cortisone by 11β-HSD2 prevents MR activation by excess cortisol, thereby acting as a tissue-specific modulator of the MR pathway. In addition to cortisol and aldosterone, deoxycorticosterone (DOC) (Fig. 342-1) also exerts mineralocorticoid activity. DOC accumulation due to 11β-hydroxylase deficiency or due to tumor-related excess production can result in mineralocorticoid excess.

Analogous to cortisol action via the GR, aldosterone (or cortisol) binding to the MR dissociates the HSP–receptor complex, allowing homodimerization of the MR, and translocation of the hormone-bound MR dimer to the nucleus (Fig. 342-7). The activated MR enhances transcription of the epithelial sodium channel (ENaC) and serum glucocorticoid-inducible kinase 1 (SGK-1). In the cytosol, interaction of ENaC with Nedd4 prevents cell surface expression of ENaC. However, SGK-1 phosphorylates serine residues within the Nedd4 protein, reduces the interaction between Nedd4 and ENaC and consequently enhances the trafficking of ENaC to the cell surface, where it mediates sodium retention.

Cushing's Syndrome

(See also Chap. 339) Cushing's syndrome reflects a constellation of clinical features that result from chronic exposure to excess glucocorticoids of any etiology. The disorder can be ACTH-dependent (e.g., pituitary corticotrope adenoma, ectopic secretion of ACTH by nonpituitary tumor) or ACTH-independent (e.g., adrenocortical adenoma, adrenocortical carcinoma, nodular adrenal hyperplasia), as well as iatrogenic (e.g., administration of exogenous glucocorticoids to treat various inflammatory conditions). The term *Cushing's disease* refers specifically to Cushing's syndrome caused by a pituitary corticotrope adenoma.

EPIDEMIOLOGY

Cushing's syndrome is generally considered a rare disease. It occurs with an incidence of 1–2 per 100,000 population per year. However, it is debated whether mild cortisol excess may be more prevalent among patients with several features of Cushing's such as centripetal obesity, type 2 diabetes, and osteoporotic vertebral fractures, recognizing that these are relatively nonspecific and common in the population.

In the overwhelming majority of patients, Cushing's syndrome is caused by an ACTH-producing corticotrope adenoma of the pituitary **(Table 342-1)**, as initially described by Harvey Cushing in 1912. Cushing's disease more frequently affects women, with the exception of prepubertal cases, where it is more common in boys. By contrast, ectopic ACTH syndrome is more frequently identified in men. Only 10% of patients with Cushing's syndrome have a primary, adrenal cause of their disease (e.g., autonomous cortisol excess independent of ACTH), and most of these patients are women. Overall, the medical use of glucocorticoids for immunosuppression, or for the treatment of inflammatory disorders, is the most common cause of Cushing's syndrome.

Table 342-1 Causes of Cushing's Syndrome

Causes of Cushing's Syndrome	Female:Male Ratio	%
ACTH-Dependent Cushing's		**90**
Cushing's disease (= ACTH-producing pituitary adenoma)	4:1	75
Ectopic ACTH syndrome (due to ACTH secretion by bronchial or pancreatic carcinoid tumors, small cell lung cancer, medullary thyroid carcinoma, pheochromocytoma and others)	1:1	15
ACTH-Independent Cushing's	4:1	**10**
Adrenocortical adenoma		5-10
Adrenocortical carcinoma		1%
Rare causes: PPNAD, primary pigmented nodular adrenal disease; AIMAH, ACTH-independent massive adrenal hyperplasia; McCune-Albright syndrome		<1%

Abbreviations: ACTH, adrenocorticotropic hormone; AIMAH, ACTH-independent macronodular hyperplasia; PPNAD, primary pigmented nodular adrenal disease.

ETIOLOGY

In at least 90% of patients with Cushing's disease, ACTH excess is caused by a corticotrope pituitary microadenoma, often only a few millimeters in diameter. Pituitary macroadenomas (i.e. tumors >1 cm in size), are found in only 5–10% of patients. Pituitary corticotrope adenomas usually occur sporadically, but very rarely can be found in the context of multiple endocrine neoplasia type 1 (MEN1) (Chap. 351).

Ectopic ACTH production is predominantly caused by occult carcinoid tumors, most frequently in the lung, but also in thymus or pancreas. Because of their small size, these tumors are often difficult to locate. Advanced small cell lung cancer can cause ectopic ACTH production. In rare cases, ectopic ACTH production has been found to originate from medullary thyroid carcinoma or pheochromocytoma, the latter co-secreting catecholamines and ACTH.

The majority of patients with ACTH-independent cortisol excess harbor a cortisol-producing adrenal adenoma. Adrenocortical carcinomas may also cause ACTH-independent disease and are often large, with excess production of several corticosteroid classes. A rare but notable cause of adrenal cortisol excess is ACTH-independent macronodular hyperplasia (AIMAH), generally characterized by ectopic expression of receptors not usually found in the adrenal, including receptors for luteinizing hormone, vasopressin, serotonin, interleukin-1, or gastric inhibitory peptide (GIP), the cause of food-dependent Cushing's. Activation of these receptors results in upregulation of PKA signaling, as physiologically occurs with ACTH, with a subsequent increase in cortisol production. Mutations in a regulatory subunit of PKA (PRKAR1A) are found in patients with primary pigmented nodular adrenal disease (PPNAD) as part of *Carney's complex*, an autosomal dominant multiple neoplasia condition associated with cardiac myxomas, hyperlentiginosis, Sertoli's cell tumors, and PPNAD. PPNAD can present as micronodular or macronodular hyperplasia, or both. Another rare cause of ACTH-independent Cushing's is *McCune-Albright syndrome*, also associated with polyostotic fibrous dysplasia, unilateral café-au-lait spots, and precocious puberty. McCune-Albright syndrome is caused by activating mutations in GNAS-1 (guanine nucleotide binding protein alpha stimulating activity polypeptide 1), and such mutations have also been found in bilateral macronodular hyperplasia without other McCune-Albright features (Table 342-1; see also Chap. 355).

CLINICAL MANIFESTATIONS

Glucocorticoids affect almost all cells of the body and thus signs of cortisol excess impact multiple physiologic systems **(Table 342-2)**, with upregulation of gluconeogenesis, lipolysis, and protein catabolism causing the most prominent features. In addition, excess glucocorticoid secretion overcomes the ability of 11ß-HSD2 to rapidly inactivate cortisol to cortisone in the kidney, thereby exerting mineralocorticoid actions, manifest as diastolic hypertension, hypokalemia, and edema. Excess glucocorticoids also interfere with central regulatory systems, leading to suppression of gonadotropins with subsequent hypogonadism and amenorrhea, and suppression of the hypothalamic-pituitary-thyroid axis, resulting in decreased TSH (thyroid-stimulating hormone) secretion.

Table 342-2 Signs and Symptoms of Cushing's Syndrome

Body Compartment/System	Signs and Symptoms
Body fat	Weight gain, central obesity, rounded face, fat pad on back of neck ("buffalo hump")
Skin	Facial plethora, thin and brittle skin, easy bruising, broad and purple stretch marks, acne, hirsutism
Bone	Osteopenia, osteoporosis (vertebral fractures), decreased linear growth in children
Muscle	Weakness, proximal myopathy (prominent atrophy of gluteal and upper leg muscles)
Cardiovascular system	Hypertension, hypokalemia, edema, atherosclerosis
Metabolism	Glucose intolerance/diabetes, dyslipidemia
Reproductive system	Decreased libido, in women amenorrhea (due to cortisol-mediated inhibition of gonadotropin release)
Central nervous system	Irritability, emotional lability, depression, sometimes cognitive defects, in severe cases, paranoid psychosis
Blood and immune system	Increased susceptibility to infections, increased white blood cell count, eosinopenia, hypercoagulation with increased risk of deep vein thrombosis and pulmonary embolism

The majority of clinical signs and symptoms observed in Cushing's syndrome are relatively nonspecific and include features such as obesity, diabetes, diastolic hypertension, hirsutism, and depression that are commonly found in patients who do not have Cushing's. Therefore, careful clinical assessment is an important aspect of evaluating suspected cases. A diagnosis of Cushing's should be considered when several clinical features are found in the same patient, in particular when more specific features are found. These include fragility of the skin, with easy bruising and broad (>1 cm), purplish striae **(Fig. 342-8)**, and signs of proximal myopathy, which becomes most obvious when trying to stand up from a chair without the use of hands or when climbing stairs. Clinical manifestations of Cushing's do not differ substantially among the different causes of Cushing's. In ectopic ACTH syndrome, hyperpigmentation of the knuckles, scars, or skin areas exposed to increased friction can be observed (Fig. 342-8), and is caused by stimulatory effects of excess ACTH and other POMC cleavage products on melanocyte pigment production. Furthermore, patients with ectopic ACTH syndrome, and some with adrenocortical carcinoma as the cause of Cushing's, may have a more brisk onset and rapid progression of clinical signs and symptoms.

Figure 342-8

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical features of Cushing's syndrome. ***A.*** Note central obesity and broad, purple stretch marks (***B***. close-up). ***C.*** Note thin and brittle skin in an elderly patient with Cushing's. ***D.*** Hyperpigmentation of the knuckles in a patient with ectopic ACTH excess.

Patients with Cushing's syndrome can be acutely endangered by deep vein thrombosis, with subsequent pulmonary embolism due to a hypercoagulable state associated with Cushing's. The majority of patients also experience psychiatric symptoms, mostly in the form of anxiety or depression, but acute paranoid or depressive psychosis may also occur. Even after cure, long-term health may be affected by an increased risk of cardiovascular disease and osteoporosis with vertebral fractures, depending on the duration and degree of exposure to significant cortisol excess.

DIAGNOSIS

The most important first step in the management of patients with suspected Cushing's syndrome is to establish the correct diagnosis. Most mistakes in clinical management, leading to unnecessary imaging or surgery, are made because the diagnostic protocol is not followed **(Fig. 342-9)**. This protocol requires establishing the diagnosis of Cushing's beyond doubt prior to employing any tests used for the differential diagnosis of the condition. In principle, after excluding exogenous glucocorticoid use as the cause of clinical signs and symptoms, suspected cases should be tested if there are multiple and progressive features of Cushing's, particularly features with a potentially higher discriminatory value. Exclusion of Cushing's is also indicated in patients with incidentally discovered adrenal masses.

Figure 342-9

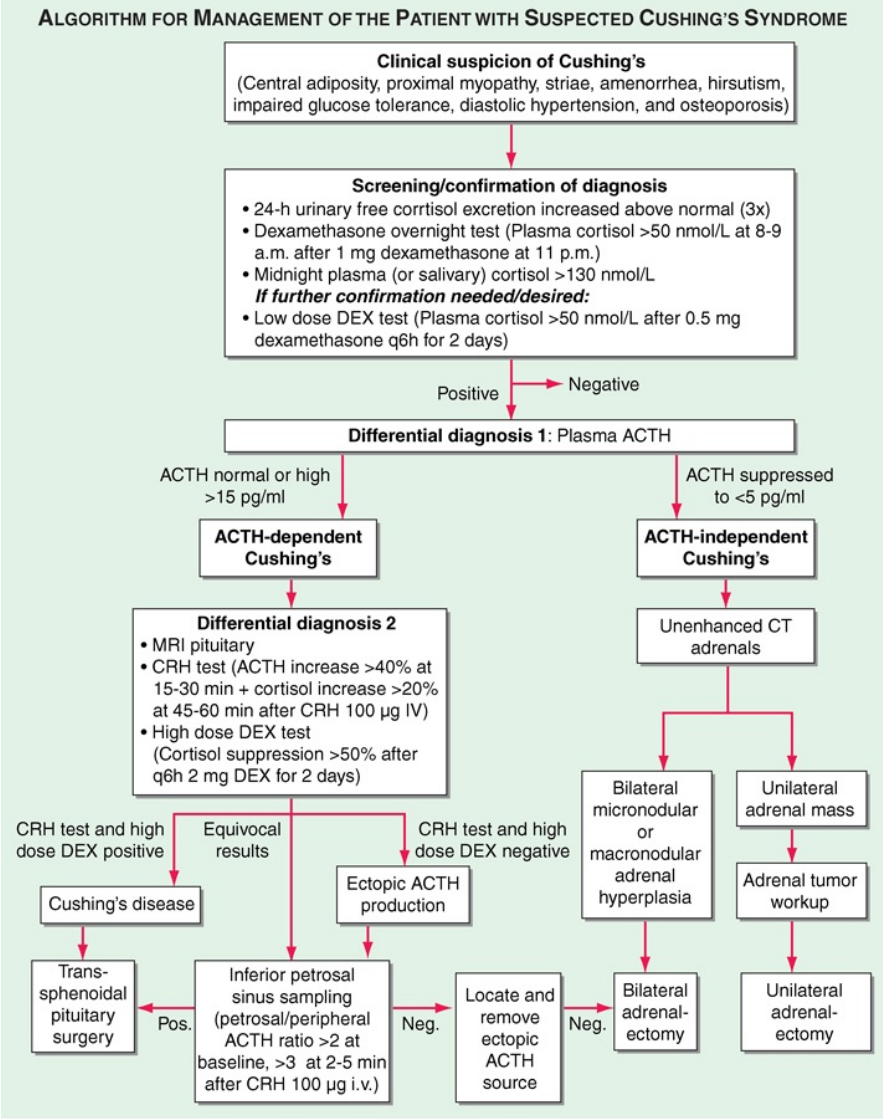

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of the patient with suspected Cushing's syndrome. CHR, corticotropin-releasing hormone; DEX, dexamethasone.

A diagnosis of Cushing's can be considered as established if the results of several tests are consistently suggestive of Cushing's. These tests may include increased 24-hour urinary free cortisol excretion in three separate collections, failure to appropriately suppress morning cortisol after overnight exposure to dexamethasone, and evidence of loss of diurnal cortisol secretion with high levels at midnight, the time of the physiologically lowest secretion (Fig. 342-9). Factors potentially affecting the outcome of these diagnostic tests have to be excluded such as incomplete 24-hour urine collection or rapid inactivation of dexamethasone due to concurrent intake of CYP3A4-inducing drugs (e.g., antiepileptics, rifampicin). Concurrent intake of oral contraceptives that raise CBG and thus total cortisol can cause failure to suppress after dexamethasone. If in doubt, testing should be repeated after 4–6 weeks off estrogens. Patients with pseudo-Cushing states, i.e., alcohol-related, and those with cyclic Cushing's may require further testing to safely confirm or exclude the diagnosis of Cushing's. In addition, the biochemical assays employed can affect the test results, with specificity representing a common problem with antibody-based assays for the measurement of urinary free cortisol. These assays have been greatly improved by the introduction of highly specific tandem mass spectrometry.

DIFFERENTIAL DIAGNOSIS

The evaluation of patients with confirmed Cushing's should be carried out by an endocrinologist and begins with the differential diagnosis of ACTH-dependent and ACTH-independent cortisol excess (Fig. 342-9). Generally, plasma ACTH levels are suppressed in cases of autonomous adrenal cortisol excess, as a consequence of enhanced negative feedback to the hypothalamus and pituitary. By contrast, patients with ACTH-dependent Cushing's have normal or increased plasma ACTH, with very high levels being found in some patients with ectopic ACTH syndrome. Importantly, imaging should only be used after it is established whether the cortisol excess is ACTH-dependent or ACTH-independent, as nodules in the pituitary or the adrenal are a common finding in the general population. In patients with confirmed ACTH-independent excess, adrenal imaging is indicated **(Fig. 342-10)**, preferably using an unenhanced CT scan. This allows assessment of adrenal morphology and determination of tumor density in Hounsfield Units (HU), which helps to distinguish between benign and malignant adrenal lesions.

Figure 342-10

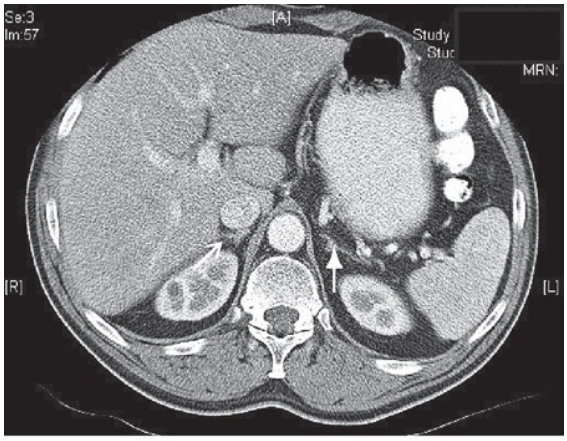

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

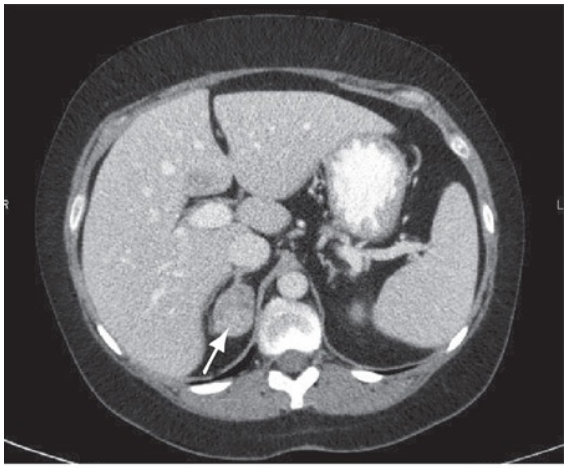

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

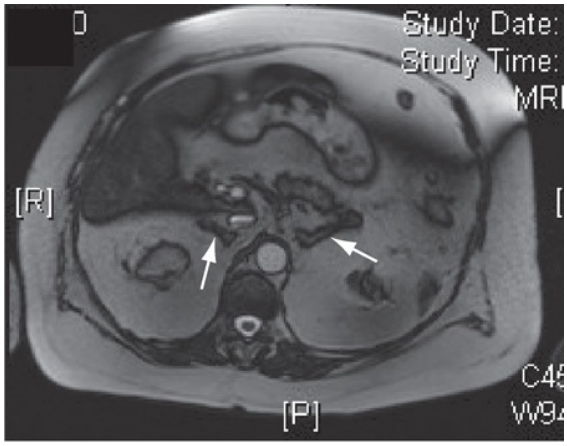

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

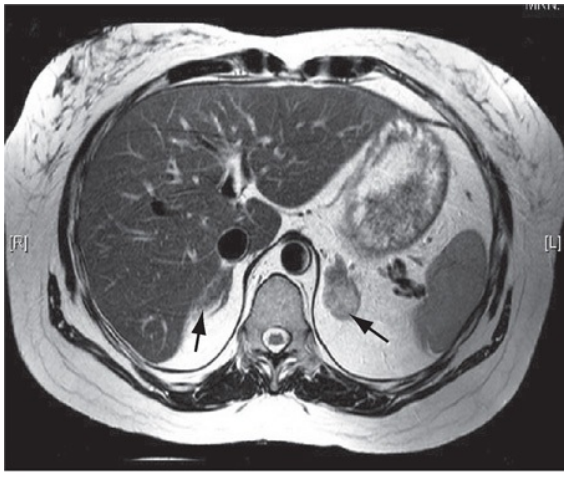

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Adrenal imaging in Cushing's syndrome.*A.* Adrenal CT showing normal bilateral adrenal morphology (arrows). ***B.*** MRI showing bilateral adrenal hyperplasia due to excess ACTH stimulation in Cushing's disease. ***C.*** CT scan depicting a right adrenocortical adenoma (arrow) causing Cushing's syndrome. ***D.*** MRI showing bilateral macronodular hyperplasia causing Cushing's syndrome.

For ACTH-dependent cortisol excess (Chap. 339), an MRI of the pituitary is the investigation of choice, but it may not show an abnormality in up to 40% of cases because small tumors are below the sensitivity of detection. Characteristically, pituitary corticotrope adenomas fail to enhance following gadolinium administration on T1-weighted MRI images. In all cases of confirmed ACTH-dependent Cushing's, further tests are required for the differential diagnosis of pituitary Cushing's disease and ectopic ACTH syndrome. These tests exploit the fact that most pituitary corticotrope adenomas still display regulatory features, including residual ACTH suppression by high-dose glucocorticoids and CRH responsiveness. In contrast, ectopic sources of ACTH are typically resistant to dexamethasone suppression and unresponsive to CRH (Fig. 342-9). However, it should be noted that a small minority of ectopic ACTH-producing tumors exhibit dynamic responses similar to pituitary corticotrope tumors. If the two tests show discordant results, or if there is any other reason for doubt, the differential diagnosis can be further clarified by performing bilateral inferior petrosal sinus sampling (IPSS) with concurrent blood sampling for ACTH in the right and left inferior petrosal sinus and a peripheral vein. An increased central/peripheral plasma ACTH ratio >2 at baseline and >3 after CRH injection is indicative of Cushing's disease (Fig. 342-9), with very high sensitivity and specificity. Of note, the results of the IPSS cannot be reliably used for lateralization (i.e. prediction of the location of the tumor within the pituitary), because there is broad interindividual variability in the venous drainage of the pituitary region. Importantly, no cortisol-lowering agents should be used prior to IPSS.

If the differential diagnostic testing indicates ectopic ACTH syndrome, then further imaging should include high-resolution, fine-cut CT scanning of the chest and abdomen for scrutiny of the lung, thymus, and pancreas. If no lesions are identified, an MRI of the chest can be considered as carcinoid tumors usually show high signal intensity on T2-weighted images. Furthermore, octreotide scintigraphy can be helpful in some cases as ectopic ACTH-producing tumors often express somatostatin receptors. Depending on the suspected cause, patients with ectopic ACTH syndrome should also undergo blood sampling for fasting gut hormones, chromogranin A, calcitonin, and biochemical exclusion of pheochromocytoma.

TREATMENT: CUSHING'S SYNDROME

Overt Cushing's is associated with a poor prognosis if left untreated. In ACTH-independent disease, treatment consists of surgical removal of the adrenal tumor. For smaller tumors, a minimally invasive approach can be employed, whereas for larger tumors and those suspected of malignancy, an open approach is preferred.

In Cushing's disease, the treatment of choice is selective removal of the pituitary corticotrope tumor, usually via a transsphenoidal approach. This results in an initial cure rate of 70–80% when performed by a highly experienced surgeon. However, even after initial remission following surgery, long-term follow-up is important as late relapse occurs in a significant number of patients. If pituitary disease recurs, there are several options, including second surgery, radiotherapy, stereotactic radiosurgery, and bilateral adrenalectomy. These options need to be applied in a highly individualized fashion.

In some with very severe, overt Cushing's (e.g., difficult to control hypokalemic hypertension or acute psychosis), it may be necessary to introduce medical therapy to rapidly control the cortisol excess during the period leading up to surgery. Similarly, patients with metastasized, glucocorticoid-producing carcinomas may require long-term antiglucocorticoid drug treatment. In case of ectopic ACTH syndrome, in which the tumor cannot be located, one must carefully weigh whether drug treatment or bilateral adrenalectomy is the most appropriate choice, with the latter facilitating immediate cure but requiring life-long corticosteroid replacement. In this instance, it is paramount to ensure regular imaging follow-up for identification of the ectopic ACTH source.

Oral agents with established efficacy in Cushing's syndrome are metyrapone and ketoconazole. Metyrapone inhibits cortisol synthesis at the level of 11β-hydroxylase (Fig. 342-1), whereas the antimycotic drug ketoconazole inhibits the early steps of steroidogenesis. Typical starting doses are 500 mg/tid for metyrapone (maximum dose, 6 g) and 200 mg/tid for ketoconazole (maximum dose, 1200 mg). Mitotane, a derivative of the insecticide o,p'DDD, is an adrenolytic agent that is also effective for reducing cortisol. Because of its side effect profile, it is most commonly used in the context of adrenocortical carcinoma, but low-dose treatment (500–1000 mg per day) has also been used in benign Cushing's. In severe cases of cortisol excess, etomidate can be used to lower cortisol. It is administered by continuous IV infusion in low, nonanesthetic doses.

After the successful removal of an ACTH- or cortisol-producing tumor, the HPA axis will remain suppressed. Thus, hydrocortisone replacement needs to be initiated at the time of surgery and slowly tapered following recovery, to allow physiologic adaptation to normal cortisol levels. Depending on degree and duration of cortisol excess, the HPA axis may require many months or even years to resume normal function.

Mineralocorticoid Excess

EPIDEMIOLOGY

Following the first description of a patient with an aldosterone-producing adrenal adenoma (*Conn's syndrome*), mineralocorticoid

excess was thought to represent a rare cause of hypertension. However, in studies systematically screening all patients with hypertension, a much higher prevalence is now recognized, ranging from 5 to 12%. The prevalence is higher when patients are preselected for hypokalemic hypertension.

ETIOLOGY

The most common cause of mineralocorticoid excess is primary hyperaldosteronism, reflecting excess production of aldosterone by the adrenal zona glomerulosa. Bilateral micronodular hyperplasia is somewhat more common than unilateral adrenal adenomas **(Table 342-3)**. Bilateral adrenal hyperplasia is usually micronodular but can also contain larger nodules that might be mistaken for a unilateral adenoma. In rare instances, primary hyperaldosteronism is caused by an adrenocortical carcinoma. Carcinomas should be considered in younger patients and in those with larger tumors, as benign aldosterone-producing adenomas usually measure <1 cm in diameter.

Table 342-3 Causes of Mineralocorticoid Excess

Causes of Mineralocorticoid Excess	Mechanism	%
Primary Hyperaldosteronism		
Adrenal (Conn's) adenoma	Autonomous aldosterone excess	40
Bilateral (micronodular) adrenal hyperplasia	Autonomous aldosterone excess	60
Glucocorticoid-remediable hyperaldosteronism (dexamethasone-suppressible hyperaldosteronism)	Crossover between the *CYP11B1* and *CYP11B2* genes results in ACTH-driven aldosterone production	<1
Other Causes (Rare)		<1
Syndrome of apparent mineralocorticoid excess (AME)	Mutations in *HSD11B2* result in lack of renal activation of cortisol to cortisone, leading to excess activation of the MR by cortisol	
Cushing's syndrome	Cortisol excess overcomes the capacity of HSD11B2 to inactivate cortisol to cortisone, consequently flooding the MR	
Glucocorticoid resistance	Upregulation of cortisol production due to GR mutations results in flooding of the MR by cortisol	
Adrenocortical carcinoma	Autonomous aldosterone and/or DOC excess	
Congenital adrenal hyperplasia	Accumulation of DOC due to mutations in *CYP11B1* or *CYP17A1*	
Progesterone-induced hypertension	Progesterone acts as an abnormal ligand due to mutations in the *MR* gene	
Liddle's syndrome	Mutant ENaC β or γ subunits resulting in reduced degradation of ENaC keeping the membrane channel in open conformation for longer, enhancing mineralocorticoid action	

Abbreviations: DOC, deoxycorticosterone; ENaC, epithelial sodium channel; GR, glucocorticoid receptor; MR, mineralocorticoid receptor.

A rare cause of aldosterone excess is glucocorticoid-remediable aldosteronism (GRA), which is caused by a chimeric gene resulting from cross-over of promoter sequences between the CYP11B1 and CYP11B2 genes that are involved in glucocorticoid and mineralocorticoid synthesis, respectively (Fig. 342-1). This rearrangement brings CYP11B2 under the control of ACTH receptor signaling; consequently, aldosterone production is regulated by ACTH rather than by renin. The family history can be helpful as there may be evidence for dominant transmission of hypertension. Recognition of the disorder is important because it can be associated with early-onset hypertension and strokes. In addition, glucocorticoid suppression can reduce aldosterone production.

Other rare causes of mineralocorticoid excess are listed in Table 342-3. An important cause is excess binding and activation of the mineralocorticoid receptor by a steroid other than aldosterone. Cortisol acts as a potent mineralocorticoid if it escapes efficient inactivation to cortisone by 11β-HSD2 in the kidney (Fig. 342-7). This can be caused by inactivating mutations in the *HSD11B2* gene resulting in the syndrome of apparent mineralocorticoid excess (AME) that characteristically manifests with severe hypokalemic hypertension in childhood. However, milder mutations may cause normokalemic hypertension manifesting in adulthood (Type II AME). Inhibition of 11β-HSD2 by excess licorice ingestion also results in hypokalemic hypertension, as does overwhelming of 11β-HSD2 conversion capacity by cortisol excess in Cushing's syndrome. Desoxycorticosterone (DOC) also binds and activates the mineralocorticoid receptor and can cause hypertension if its circulating concentrations are increased. This can arise through autonomous DOC secretion by an adrenocortical carcinoma, but also when DOC accumulates as a consequence of an adrenal enzymatic block, as seen in congenital adrenal hyperplasia due to CYP11B1 (11β-hydroxylase) or CYP17A1 (17α-hdyroxylase) deficiency (Fig. 342-1). Progesterone can cause hypokalemic hypertension in rare individuals who harbor a mineralocorticoid receptor mutation that enhances binding and activation by progesterone; physiologically, progesterone normally exerts antimineralocorticoid activity. Finally, excess mineralocorticoid activity can be caused by mutations in the β or γ subunits of the ENaC, disrupting its interaction with Nedd4 (Fig. 342-7), and thereby decreasing receptor internalization and degradation. The constitutively active ENAC drives hypokalemic hypertension, resulting in an autosomal dominant disorder termed *Liddle's syndrome*.

CLINICAL MANIFESTATIONS

Excess activation of the mineralocorticoid receptor leads to potassium depletion and increased sodium retention, with the latter causing an expansion of extracellular and plasma volume. Increased ENaC activity also results in hydrogen depletion that can cause metabolic alkalosis. Aldosterone also has direct effects on the vascular system, where it increases cardiac remodeling and decreases compliance. Aldosterone excess may cause direct damage to the myocardium and the kidney glomeruli, in addition to secondary damage due to systemic hypertension.

The clinical hallmark of mineralocorticoid excess is hypokalemic hypertension; serum sodium tends to be normal due to the concurrent fluid retention, which in some cases can lead to peripheral edema. Hypokalemia can be exacerbated by thiazide drug treatment, which leads to increased delivery of sodium to the distal renal tubule, thereby driving potassium excretion. Severe hypokalemia can be associated with muscle weakness, overt proximal myopathy, or even hypokalemic paralysis. Severe alkalosis contributes to muscle cramps and, in severe cases, can cause tetany.

DIAGNOSIS

Diagnostic screening for mineralocorticoid excess is not currently recommended for all patients with hypertension, but should be restricted to those who exhibit hypertension associated with drug resistance, hypokalemia, an adrenal mass, or hypertension before the age of 40 years **(Fig. 342-11)**. The accepted screening test is concurrent measurement of plasma renin and aldosterone with subsequent calculation of the aldosterone-renin ratio (ARR) (Fig. 342-11); serum potassium needs to be normalized prior to testing. Stopping antihypertensive medication can be cumbersome, particularly in patients with severe

hypertension. Thus, for practical purposes, in the first instance the patient can remain on the usual antihypertensive medications, with the exception that mineralocorticoid receptor antagonists need to be ceased at least 4 weeks prior to ARR measurement. The remaining antihypertensive drugs usually do not affect the outcome of ARR testing, except that β-blocker treatment can cause false-positive results and ACE/AT1R inhibitors can cause false-negative results in milder cases **(Table 342-4)**.

Figure 342-11

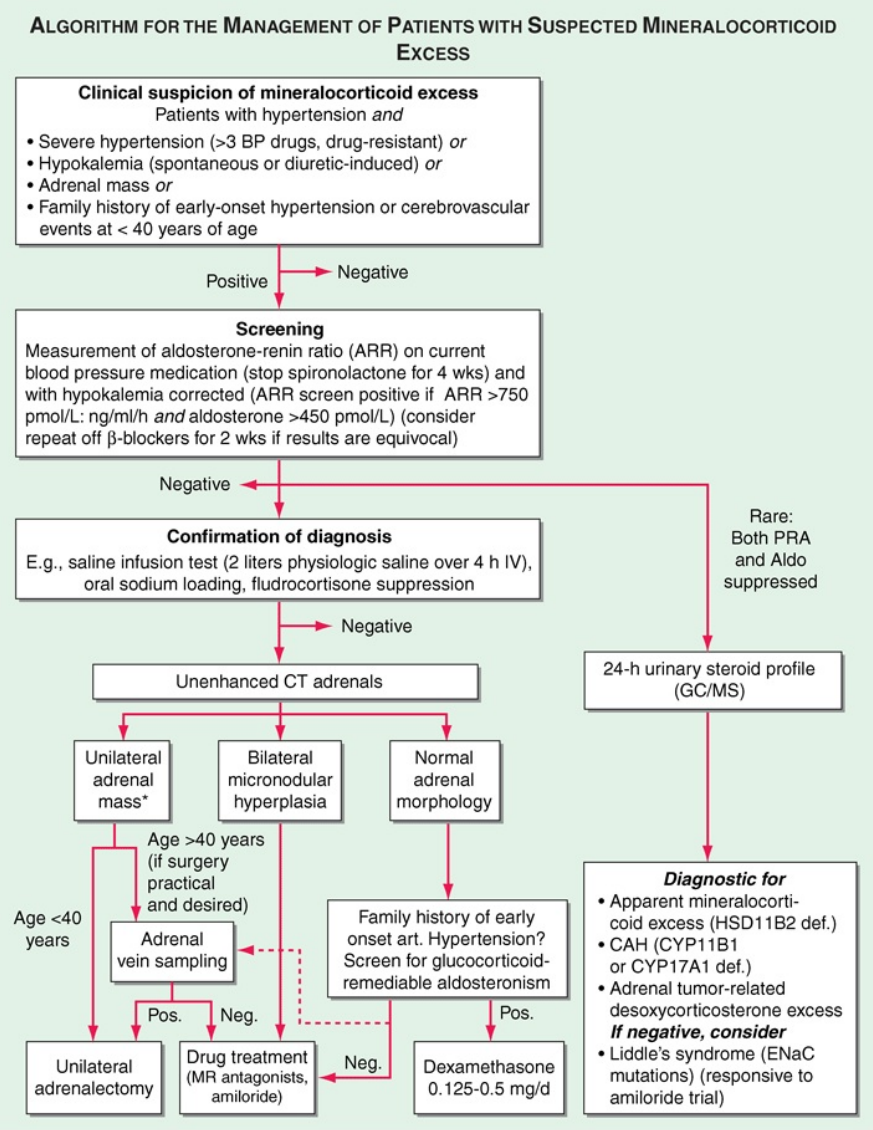

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of patients with suspected mineralocorticoid excess. Perform adrenal tumor workup (see Fig. 342-12). GC/MS, gaschromatography/mass spectometry.

Table 342-4 Effects of Antihypertensive Drugs on the Aldosterone-Renin-Ratio (Arr)

Drug	Effect on Renin	Effect on Aldosterone	Net Effect on ARR
β-Blockers	↓	↑	↑
α_1-Blockers	→	→	→
α_2-Sympathomimetics	→	→	→
ACE inhibitors	↑	↓	↓
AT1R blockers	↑	↓	↓
Calcium antagonists	→	→	→
Diuretics	(↑)	(↑)	→/(↓)

ARR screening is positive if the ratio is lower than 750 pmol/L: ng/mL per hour, with a concurrently high normal or increased aldosterone (Fig. 342-11). If one relies on the ARR only, the likelihood of a false-positive ARR becomes greater when renin levels are very low. The characteristics of the biochemical assays are also important. Some labs measure plasma renin activity whereas others measure plasma renin concentrations. Antibody-based assays for the measurement of serum aldosterone lack the reliability of tandem mass spectrometry assays but these are not yet ubiquitously available.

Diagnostic confirmation of mineralocorticoid excess in a patient with positive ARR screening result should be undertaken by an endocrinologist as the tests lack optimized validation. The most straightforward is the saline infusion test, which involves the IV administration of 2 L of physiologic saline over a 4-hour period. Failure of aldosterone to suppress below 140 pmol/L (5 ng/dL) is indicative of autonomous mineralocorticoid excess. Alternative tests are the oral sodium loading test (300 mmol NaCl/d for 3 days) or the fludrocortisone suppression test (0.1 mg q6h with 30 mmol NaCl q8h for 4 days); the latter can be difficult because of the risk of profound hypokalemia and increased hypertension. In patients with overt hypokalemic hypertension, strongly positive ARR, and concurrently increased aldosterone levels, confirmatory testing is usually not necessary.

DIFFERENTIAL DIAGNOSIS AND TREATMENT

After the diagnosis of hyperaldosteronism is established, the next step is to use adrenal imaging to further assess the cause. Fine-cut CT scanning of the adrenal region is the method of choice as it provides excellent visualization of adrenal morphology. CT will readily identify larger tumors suspicious of malignancy but may miss lesions smaller than 5 mm. The differentiation between bilateral micronodular hyperplasia and a unilateral adenoma is only required if a surgical approach is feasible and desired. Consequently, selective adrenal vein sampling (AVS) should only be carried out in surgical candidates with either no obvious lesion on CT or evidence of a unilateral lesion in patients older than 40 years, as the latter patients have a high likelihood of harboring a coincidental, endocrine inactive adrenal adenoma (Fig. 342-11). AVS is used to compare aldosterone levels in the inferior vena cava and between the right and left adrenal veins. AVS requires concurrent measurement of cortisol to document correct placement of the catheter in the adrenal veins and should demonstrate a cortisol gradient >3 between the vena cava and each adrenal vein. Lateralization is confirmed by an aldosterone/cortisol ratio that is at least twofold higher on one side than the other. AVS is a complex procedure that requires a highly skilled interventional radiologist. Even then, the right adrenal vein can be difficult to cannulate correctly, which invalidates the procedure. There is also no agreement as to whether the two adrenal veins should be cannulated simultaneously or successively and whether ACTH stimulation enhances the diagnostic value of AVS.

Patients younger than 40 years with confirmed mineralocorticoid excess and a unilateral lesion can go straight to surgery, which is also indicated in patients with confirmed lateralization documented by a valid AVS procedure. Laparoscopic adrenalectomy is the preferred approach. Patients who are not surgical candidates, or with evidence of bilateral hyperplasia based on CT or AVS, should be treated medically (Fig. 342-11). Medical treatment, which can also be considered prior to surgery to avoid postsurgical hypoaldosteronism, consists primarily of the mineralocorticoid receptor antagonist spironolactone. It can be started at 12.5–50 mg bid and titrated up to a maximum of 400 mg/d to control blood pressure and normalize potassium. Side effects include menstrual irregularity, decreased libido, and gynecomastia. The more selective MR antagonist eplerenone can also be used. Doses start at 25 mg bid and it can be titrated up to 200 mg/d. Another useful drug is the sodium channel blocker amiloride (5–10 mg/bid).

In patients with normal adrenal morphology and family history of early-onset, severe hypertension, a diagnosis of GRA should be considered and can be evaluated using genetic testing. Treatment of GRA consists of administering dexamethasone, using the lowest dose possible to control blood pressure. Some patients also require additional MR antagonist treatment.

The diagnosis of nonaldosterone-related mineralocorticoid excess is based on documentation of suppressed renin and suppressed aldosterone in the presence of hypokalemic hypertension. This testing is best carried out by employing urinary steroid metabolite profiling by gas chromatography/mass spectrometry (GC/MS). An increased free cortisol over free cortisone ratio is suggestive of AME and can be treated with dexamethasone. Steroid profiling by GC/MS also detects the steroids associated with CYP11B1 and CYP17A1 deficiency or the irregular steroid secretion pattern in a DOC-producing adrenocortical carcinoma (Fig. 342-11). If the GC/MS profile is normal, then Liddle's syndrome should be considered. It is very sensitive to amiloride treatment but will not respond to MR antagonist treatment, as the defect is due to a constitutively active ENaC.

Approach to the Patient: Incidentally Discovered Adrenal Mass

EPIDEMIOLOGY

Incidentally discovered adrenal masses, commonly termed adrenal "incidentalomas," are common, with a prevalence of at least 2% in the general population as documented in CT and autopsy series. The prevalence increases with age, with 1% of 40-year-olds and 7% of 70-year-olds harboring an adrenal mass.

ETIOLOGY

Most solitary adrenal tumors are monoclonal neoplasms. Several genetic syndromes, including MEN-1 (*MEN1*), MEN-2 (*RET*), Carney's complex (*PRKAR1A*), and McCune-Albright (*GNAS1*), can have adrenal tumors as one of their features. Somatic mutations in MEN1, GNAS1, and PRKAR1A have been identified in a small proportion of sporadic adrenocortical adenomas. Aberrant expression of membrane receptors (gastric inhibitory peptide, β-adrenergic, luteinizing hormone, vasopressin V1 and interleukin-I receptors) have been identified in some sporadic cases of macronodular adrenocortical hyperplasia.

The majority of adrenal nodules are endocrine inactive adrenocortical adenomas. However, larger series suggest that up to 25% of adrenal nodules are hormonally active, due to a cortisol- or aldosterone-producing adrenocortical adenoma or a pheochromocytoma associated with catecholamine excess **(Table 342-5)**. Adrenocortical carcinoma is rare but it is the cause of an adrenal mass in 5% of patients. However, the most common cause of a malignant adrenal mass is metastasis originating from another solid tissue tumor (Table 342-5).

Table 342-5 Classification of Unilateral Adrenal Masses

Benign	**Approximate Prevalence (%)**
Adrenocortical adenoma	
Endocrine inactive	60–85
Cortisol-producing	5–10
Aldosterone-producing	2–5
Pheochromocytoma	5–10
Adrenal myelolipoma	<1
Adrenal ganglioneuroma	<0.1
Adrenal hemangioma	<0.1
Adrenal cyst	<1
Adrenal hematoma/hemorrhagic infarction	<1
Indeterminate	
Adrenocortical oncocytoma	<1
Malignant	
Adrenocortical carcinoma	2–5

Malignant pheochromocytoma	<1
Adrenal neuroblastoma	<0.1
Lymphomas (incl. primary adrenal lymphoma)	<1
Metastases (most frequent: breast, lung)	15

Note: Bilateral adrenal enlargement/masses may be caused by congenital adrenal hyperplasia, bilateral macronodular hyperplasia, bilateral hemorrhage (due to antiphospholipid syndrome or sepsis-associated Waterhouse-Friderichsen syndrome), granuloma, amyloidosis, infiltrative disease including tuberculosis.

DIFFERENTIAL DIAGNOSIS AND TREATMENT

Patients with an adrenal mass >1 cm require a diagnostic evaluation. Two key questions need to be addressed: (1) Does the tumor autonomously secrete hormones that could have a detrimental effect on health?, and (2) Is the adrenal mass benign or malignant?

Hormone secretion by an adrenal mass occurs along a continuum, with a gradual increase in clinical manifestations in parallel with hormone levels. Exclusion of catecholamine excess from a pheochromocytoma arising from the adrenal medulla is a mandatory part of the diagnostic workup **(Fig. 342-12)**. Furthermore, autonomous cortisol and aldosterone secretion resulting in Cushing's syndrome or primary hyperaldosteronism, respectively, require exclusion. Adrenal incidentalomas are associated with lower levels of autonomous cortisol secretion, and patients may lack overt clinical features of Cushing's syndrome. Nonetheless, they may exhibit one or more components of the metabolic syndrome (e.g., obesity, type 2 diabetes, or hypertension). There is ongoing debate about the optimal treatment for these patients with mild or subclinical Cushing's syndrome. Overproduction of adrenal androgen precursors, DHEA and its sulfate, is rare and is most frequently seen in the context of adrenocortical carcinoma, as are increased levels of steroid precursors such as 17-hydroxyprogesterone.

Figure 342-12

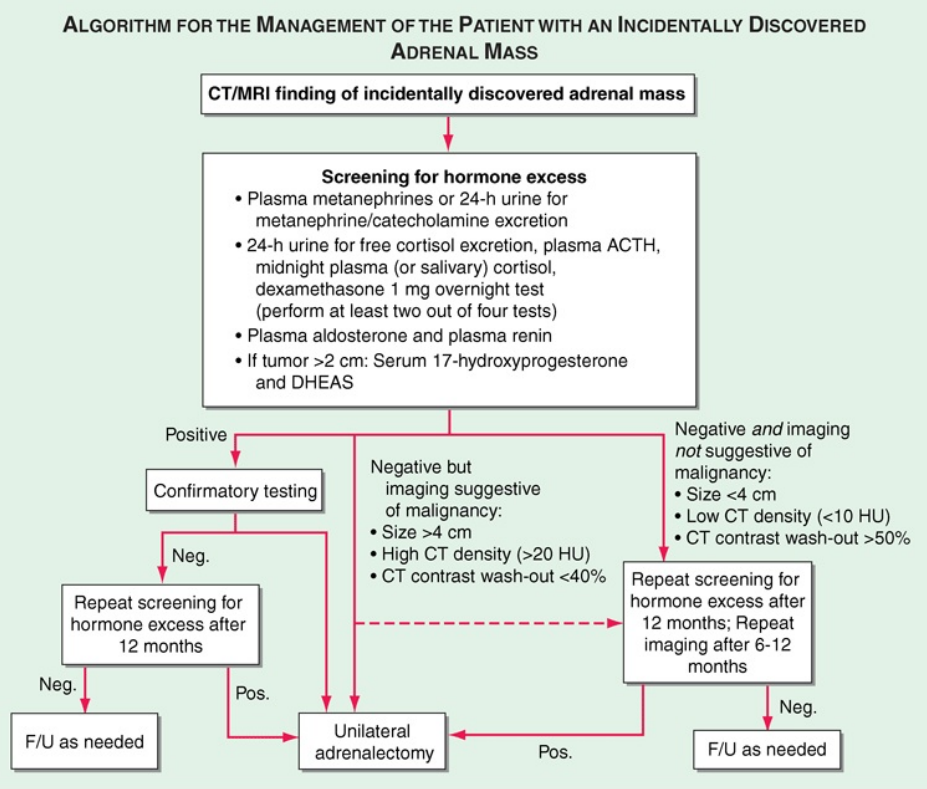

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of the patient with an incidentally discovered adrenal mass. F/U, follow-up.

For the differentiation of benign from malignant adrenal masses, imaging is relatively sensitive though specificity is suboptimal. CT is the procedure of choice for imaging the adrenal glands (Fig. 342-12). The risk of adrenocortical carcinoma, pheochromocytoma, and benign adrenal myelolipoma increases with the diameter of the adrenal mass. However, size alone is of poor predictive value, with only 80% sensitivity and 60% specificity for the differentiation of benign from malignant masses when using a 4-cm cut-off. Metastases are found with similar frequency in adrenal masses of all sizes. Tumor density on unenhanced CT is of additional diagnostic value, with most adrenocortical adenomas being lipid rich and thus presenting with low attenuation values [i.e., densities of <10 Hounsfield Units (HU)]. By contrast, adrenocortical carcinomas, but also pheochromocytomas, usually have high attenuation values (i.e. densities >20 HU on precontrast scans). Generally, benign lesions are rounded and homogenous whereas most malignant lesions appear lobulated and inhomogeneous. Pheochromocytoma and adrenomyelolipoma may also exhibit lobulated and inhomogeneous features. Additional information can be obtained from CT by assessment of contrast wash-out after 15 minutes, which is >50% in benign lesions but <40% in malignant lesions, which usually have a more extensive vascularization. MRI also allows for the visualization of the adrenal glands with somewhat lower resolution than CT. However, as it does not involve exposure to ionizing radiation, it is preferred in children, young adults, and during pregnancy. MRI has a valuable role in the characterization of indeterminate adrenal lesions using chemical shift analysis, with malignant tumors rarely showing loss of signal on opposed-phase MRI.

Fine-needle aspiration (FNA) or CT-guided biopsy of an adrenal mass is almost never indicated. FNA of a pheochromocytoma can cause a life-threatening hypertensive crisis. FNA of an adrenocortical carcinoma violates the tumor capsule. FNA should only be considered in a patient with a history of nonadrenal malignancy and a newly detected adrenal mass. FNA should be

carried out only after careful exclusion of pheochromocytoma and if the outcome will influence therapeutic management. It is important to recognize that in 25% of patients with a previous history of nonadrenal malignancy, a newly detected mass on CT is not a metastasis.

Adrenal masses associated with confirmed hormone excess or suspected malignancy are usually treated surgically (Fig. 342-12) or, if adrenalectomy is not feasible or desired, with medication. Preoperative exclusion of glucocorticoid excess is particularly important for the prediction of postoperative suppression of the contralateral adrenal gland, which requires glucocorticoid replacement before surgery. If the initial decision is for observation, imaging and biochemical testing should be repeated about a year after the first assessment. However, this may be performed earlier in patients with borderline imaging or hormonal findings. There is no agreement with regard to the required long-term follow-up beyond 1 year and in patients with normal biochemistry and no evidence of increased tumor size at follow-up.

Adrenocortical Carcinoma

Adrenocortical carcinoma (ACC) is a rare malignancy with an annual incidence of 1–2 per million population. ACC is generally considered a highly malignant tumor; however, it presents with broad interindividual variability with regard to biologic characteristics and clinical behavior. Somatic mutations in the tumor suppressor gene *TP53* are found in 25% of apparently sporadic ACC. Germline *TP53* mutations are the cause of the Li-Fraumeni syndrome associated with multiple solid organ cancers including ACC and are found in 25% of pediatric ACC cases; the *TP53* mutation R337H is found in almost all pediatric ACC in Brazil. Other genetic changes identified in ACC include alterations in the Wnt/ß-catenin pathway and in the insulin-like growth factor 2 (IGF2) cluster; IGF2 overexpression is found in 90% of ACC.

Patients with large adrenal tumors suspicious of malignancy should be managed by a multidisciplinary specialist team, including an endocrinologist, an oncologist, a surgeon, a radiologist, and a histopathologist. FNA is not indicated in suspected ACC: first, cytology and also histopathology of a core biopsy cannot differentiate between benign and malignant primary adrenal masses; second, FNA violates the tumor capsule and may even cause needle canal metastasis. Even when the entire tumor specimen is available, the histopathologic differentiation between benign and malignant lesions is a diagnostic challenge. The most common histopathologic classification is the Weiss score, taking into account high nuclear grade; mitotic rate (>5/HPF); atypical mitosis; <25% clear cells; diffuse architecture; and presence of necrosis, venous invasion, and invasion of sinusoidal structures and tumor capsule. The presence of three or more elements suggests ACC.

Although 60–70% of ACCs are biochemically found to overproduce hormones, this is not clinically apparent in many patients due to the relatively inefficient steroid production by the adrenocortical cancer cells. Excess production of glucocorticoids and adrenal androgen precursors are most common. Mixed excess production of several corticosteroid classes by an adrenal tumor is generally indicative of malignancy.

Tumor staging at diagnosis **(Table 342-6)** has important prognostic implications and requires scanning of the chest and abdomen for local organ invasion, lymphadenopathy, and metastases. Intravenous contrast medium is necessary for maximum sensitivity for hepatic metastases. An adrenal origin may be difficult to determine on standard axial CT imaging if the tumors are large and invasive, but CT reconstructions or MRI are more informative **(Fig. 342-13)** using multiple planes and different sequences. Vascular and adjacent organ invasion is diagnostic of malignancy. 18-Fluoro-2-deoxy-d-glucosepositron emission tomography (18-FDG PET) is highly sensitive for the detection of malignancy and can be used to detect small metastases or local recurrence that may not be obvious on CT (Fig. 342-13). However, FDG PET is not specific and therefore cannot be used for differentiating benign from malignant adrenal lesions. Metastasis in ACC most frequently occurs to liver and lung.

Table 342-6 Classification System for Staging of Adrenocortical Carcinoma

Stage	ENSAT Stage	TNM Definitions
I	**T1,N0, M0**	T1, tumor ≤5 cm N0, no positive lymph node M0, no distant metastases
II	**T2,N0,M0**	T2, tumor >5 cm N0, no positive lymph node M0, no distant metastases
III	**T1–T2,N1,M0** **T3–T4,N0–N1,M0**	N1, positive lymph node(s) M0, no distant metastases T3, tumor infiltration into surrounding tissue T4, tumor invasion into adjacent organs ***or*** venous tumor thrombus in vena cava or renal vein
IV	**T1–T4,N0–N1,M1**	M1, presence of distant metastases

Abbreviation: ENSAT, European Network for the Study of Adrenal Tumors.

Figure 342-13

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

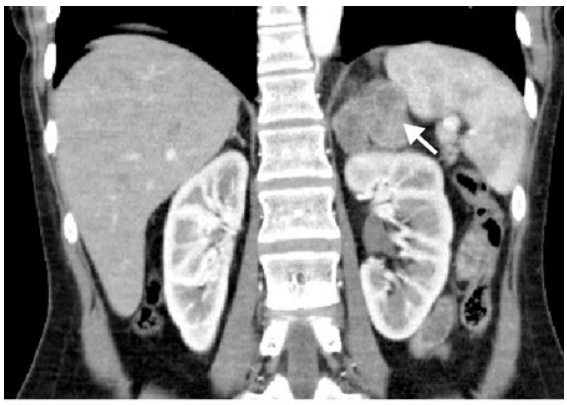

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

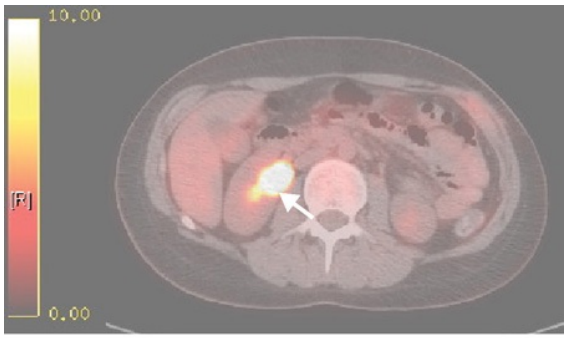

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Imaging in adrenocortical carcinoma. MRI scan with ***A.*** frontal and ***B.*** lateral views of a left adrenocortical carcinoma that was detected incidentally. CT scan with ***C.*** coronal and ***D.*** transverse views depicting a right-sided adrenocortical carcinoma. Note the irregular border and inhomogeneous structure. CT scan ***E.*** and PET-CT ***F.*** visualizing a peritoneal metastasis of an adrenocortical carcinoma in close proximity to the left kidney (arrow).

ACC carries a poor prognosis and cure can be achieved only by complete surgical removal. Capsule violation during primary surgery, metastasis at diagnosis, and primary treatment in a nonspecialist center are major determinants of poor survival. If the primary tumor invades adjacent organs, en bloc removal of kidney and spleen should be considered to reduce the risk of recurrence. Surgery can also be considered in a patient with metastases if there is severe tumor-related hormone excess. This indication needs to be carefully weighed against surgical risk, including thromboembolic complications, and the resulting delay in the introduction of other therapeutic options. Patients with confirmed ACC and successful removal of the primary tumor should receive adjuvant treatment with mitotane (o,p'DDD), particularly in patients with a high risk of recurrence as determined by tumor size >8 cm, histopathologic signs of vascular invasion, capsule invasion or violation, and a Ki67 proliferation index ≥10%. Mitotane is usually started at 500 mg qid, with doses increased by 1000 mg/d every 1–2 weeks as tolerated. The maximum tolerated dose is usually 8–10 g/m^2 per day. Adjuvant mitotane should be continued for at least 2 years, if the patient can tolerate side effects. Regular monitoring of plasma mitotane levels is mandatory (therapeutic range 14–20 mg/L; neurotoxic complications more frequent >20 mg/L), as is concurrent replacement with hydrocortisone. The latter should be given at higher doses than usually employed in adrenal insufficiency (e.g., 20 mg tid), as mitotane increases glucocorticoid inactivation due to the induction of hepatic CYP3A4 activity. It also increases circulating cortisol-binding globulin, thereby decreasing the available free cortisol fraction. Single metastases can be addressed surgically or with radiofrequency ablation as appropriate. If the tumor recurs or progresses during mitotane treatment, chemotherapy should be considered (e.g., cisplatin, etoposide, doxorubicin plus continuing mitotane, the so-called Berrutti regimen); painful bone metastasis responds to irradiation. Overall survival in ACC is still poor, with 5-year survival rates of 30–40%.

Adrenal Insufficiency

EPIDEMIOLOGY

The prevalence of well-documented, permanent adrenal insufficiency is 5 in 10,000 in the general population. Hypothalamic-pituitary origin of disease is most frequent, with a prevalence of 3 in 10,000, whereas primary adrenal insufficiency has a prevalence of 2 in 10,000. Approximately one-half of the latter cases are acquired, mostly caused by autoimmune destruction of the adrenal glands; the other one-half are genetic, most commonly caused by distinct enzymatic blocks in adrenal steroidogenesis affecting glucocorticoid synthesis (i.e. congenital adrenal hyperplasia.)

Adrenal insufficiency arising from suppression of the HPA axis as a consequence of exogenous glucocorticoid treatment is much more common, occurring in 0.5–2% of the population in developed countries.

ETIOLOGY

Primary adrenal insufficiency is most commonly caused by autoimmune adrenalitis. Isolated autoimmune adrenalitis accounts for 30–40%, whereas 60–70% develop adrenal insufficiency as part of autoimmune polyglandular syndromes (APS) (Chap. 351) **(Table 342-7)**. APS1, also termed APECED (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy), is the underlying cause in 10% of patients affected by APS. APS1 is transmitted in an autosomal recessive manner and is caused by mutations in the autoimmune regulator gene *AIRE*. Associated autoimmune conditions overlap with those seen in APS2, but may also include total alopecia, primary hypoparathyroidism, and, in rare cases, lymphoma. APS1 patients invariably develop chronic mucocutaneous candidiasis, usually manifest in childhood, and preceding adrenal insufficiency by years or decades. The much more prevalent APS2 is of polygenic inheritance, with confirmed associations with the *HLA-DR3* gene region in the major histocompatibility complex and distinct gene regions involved in immune regulation (*CTLA-4*, *PTPN22*, *CLEC16A*). Coincident autoimmune disease most frequently includes thyroid autoimmune disease, vitiligo, and premature ovarian failure. Less commonly, additional features may include type 1 diabetes and pernicious anemia caused by vitamin B12 deficiency.

Table 342-7 Causes of Primary Adrenal Insufficiency

Diagnosis	Gene	Associated Features
Autoimmune polyglandular syndrome 1 (APS1)	*AIRE*	Hypoparathyroidism, chronic mucocutaneous candidiasis, other autoimmune disorders, rarely lymphomas
Autoimmune polyglandular syndrome 2 (APS2)	Associations with HLA-DR3, CTLA-4	Hypothyroidism, hyperthyroidism, premature ovarian failure, vitiligo, type 1 diabetes mellitus, pernicious anemia
Isolated autoimmune adrenalitis	Associations with HLA-DR3, CTLA-4	
Congenital adrenal hyperplasia (CAH)	*CYP21A2, CYP11B1, CYP17A1, HSD3B2, POR*	See Table 342-10 (see also Chap. 349)
Congenital lipoid adrenal hyperplasia (CLAH)	*StAR, CYP11A1*	46,XY DSD, gonadal failure (see also Chap. 349)
Adrenal hypoplasia congenita (AHC)	*NR0B1* (DAX-1), *NR5A1* (SF-1)	46,XY DSD, gonadal failure (see also Chap. 349)
Adrenoleukodystrophy (ALD), adrenomyeloneuropathy (AMN)	*X-ALD*	Demyelination of central nervous system (ALD) or spinal cord and peripheral nerves (AMN)
Familial glucocorticoid deficiency - FGD1 - FGD2 - FGD3 Triple A syndrome	*MC2R* *MRAP* ? *AAAS*	ACTH insensitivity syndromes due to mutations in the ACTH receptor MC2R and its accessory protein MRAP tall stature Alacrima, achalasia, neurologic impairment
Smith-Lemli-Opitz-Syndrome	*SLOS*	Cholesterol synthesis disorder associated with mental retardation, craniofacial malformations, growth failure
Kearns-Sayre syndrome	Mitochondrial DNA deletions	Progressive external ophthalmoplegia, pigmentary retinal degeneration, cardiac conduction defects, gonadal failure, hypoparathyroidism, type 1 diabetes
IMAGe syndrome	?	Intrauterine growth retardation, metaphyseal dysplasia, genital anomalies
Adrenal infections		Tuberculosis, HIV, CMV, cryptococcosis, histoplasmosis, coccidioidomycosis
Adrenal infiltration		Metastases, lymphomas, sarcoidosis, amyloidosis, hemochromatosis
Adrenal hemorrhage		Meningococcal sepsis (Waterhouse-Friderichsen syndrome), primary antiphospholipid syndrome
Drug-induced		Mitotane, aminoglutethimide, arbiraterone, trilostane, etomidate, ketoconazole, suramin, RU486
Bilateral adrenalectomy		E.g., in the management of Cushing's or after bilateral nephrectomy

Abbreviations: CMV, cytomegalovirus; DSD, disordered sex development.

X-linked adrenoleukodystrophy has an incidence of 1:20,000 males and is caused by mutations in the *X-ALD* gene encoding the peroxisomal membrane transporter protein ABCD1; its disruption results in accumulation of very long chain (>24 carbon atoms) fatty acids. Approximately 50% of cases manifest in early childhood with rapidly progressive white matter disease (cerebral ALD); 35% present during adolescence or in early adulthood with neurologic features indicative of myelin and peripheral nervous system involvement (adrenomyeloneuropathy, AMN). In the remaining 15%, adrenal insufficiency is the sole manifestation of disease. Of note, distinct mutations manifest with variable penetrance within affected families.

Rarer causes of adrenal insufficiency involve destruction of the adrenal glands as a consequence of infection, hemorrhage, or infiltration (Table 342-7); tuberculous adrenalitis is still a frequent cause of disease in developing countries. Adrenal metastases rarely cause adrenal insufficiency, and this occurs only with bilateral, bulky metastases.

Inborn causes of primary adrenal insufficiency other than congenital adrenal hyperplasia are rare, causing less than 1% of cases. However, their elucidation provides important insights into adrenal gland development and physiology. Mutations causing

primary adrenal insufficiency (Table 342-7) include factors regulating adrenal development and steroidogenesis (DAX-1, SF-1), cholesterol synthesis, import and cleavage (DHCR7, StAR, CYP11A1), and elements of the adrenal ACTH response pathway (MC2R, MRAP) (Fig. 342-5).

Secondary adrenal insufficiency is the consequence of dysfunction of the hypothalamic-pituitary component of the HPA axis **(Table 342-8)**. Excluding iatrogenic suppression, the overwhelming majority of cases are caused by pituitary or hypothalamic tumors, or their treatment by surgery or irradiation (Chap. 339). Rarer causes include pituitary apoplexy, either as a consequence of an infarcted pituitary adenoma or transient reduction in the blood supply of the pituitary during surgery or after rapid blood loss associated with parturition, also termed Sheehan's syndrome. Isolated ACTH deficiency is rarely caused by autoimmune disease or pituitary infiltration (Table 342-8). Mutations in the ACTH precursor POMC or in factors regulating pituitary development are genetic causes of ACTH deficiency (Table 342-8).

Table 342-8 Causes of Secondary Adrenal Insufficiency

Diagnosis	Gene	Associated Features
Pituitary tumors (endocrine active and inactive adenomas, very rare: carcinoma)		Depending on tumor size and location: visual field impairment (bilateral hemianopia), hyperprolactinemia, secondary hypothyroidism, hypogonadism, growth hormone deficiency
Other mass lesions affecting the hypothalamic-pituitary region		Craniopharyngioma, meningioma, ependymoma, metastases
Pituitary irradiation		Radiotherapy administered for pituitary tumors, brain tumors, or craniospinal irradiation in leukemia
Autoimmune hypophysitis		Often associated with pregnancy; may present with panhypopituitarism or isolated ACTH deficiency; can be associated with autoimmune thyroid disease, more rarely with vitiligo, premature ovarian failure, type 1 diabetes, pernicious anemia
Pituitary apoplexy/hemorrhage		Hemorrhagic infarction of large pituitary adenomas or pituitary infarction consequent to traumatic major blood loss (e.g., surgery or pregnancy: Sheehan's syndrome)
Pituitary infiltration		Tuberculosis, actinomycosis, sarcoidosis, histiocytosis X, granulomatosis with polyangiitis (Wegener's), metastases
Drug-induced		Chronic glucocorticoid excess (endogenous or exogenous)
Congenital isolated ACTH deficiency	*TBX19* (Tpit)	
Combined pituitary hormone deficiency (CPHD)	*PROP-1* *HESX1* *LHX3* *LHX4* *SOX3*	Progressive development of CPHD in the order GH, PRL, TSH, LH/FSH, ACTH CPHD and septo-optic dysplasia CPHD and limited neck rotation, sensorineural deafness CPHD and cerebellar abnormalities CPHD and variable mental retardation
Proopiomelanocortin (POMC) deficiency	*POMC*	Early-onset obesity, red hair pigmentation

Abbreviations: ACTH, adrenocorticotropic hormone; GH, growth hormone; LH/FSH, luteinizing hormone/follicle-stimulating hormone; PRL, prolactin; TSH, thyroid-stimulating hormone.

CLINICAL MANIFESTATIONS

In principle, the clinical features of primary adrenal insufficiency are characterized by the loss of both glucocorticoid and mineralocorticoid secretion **(Table 342-9)**. In secondary adrenal insufficiency, only glucocorticoid deficiency is present, as the adrenal itself is intact and thus still amenable to regulation by the RAA system. Adrenal androgen secretion is disrupted in both primary and secondary adrenal insufficiency (Table 342-9). Hypothalamic-pituitary disease can lead to additional clinical manifestations due to involvement of other endocrine axes (thyroid, gonads, growth hormone, prolactin) or visual impairment with bitemporal hemianopia caused by chiasmal compression. It is important to recognize that iatrogenic adrenal insufficiency caused by exogenous glucocorticoid suppression of the HPA axis may result in all symptoms associated with glucocorticoid deficiency (Table 342-9), if exogenous glucocorticoids are stopped abruptly. However, patients will appear clinically Cushingoid as a result of the preceding overexposure to glucocorticoids.

Table 342-9 Signs and Symptoms of Adrenal Insufficiency

Signs and Symptoms Caused by Glucocorticoid Deficiency
Fatigue, lack of energy
Weight loss, anorexia
Myalgia, joint pain
Fever
Anemia, lymphocytosis, eosinophilia
Slightly increased TSH (due to loss of feedback inhibition of TSH release)
Hypoglycemia (more frequent in children)
Low blood pressure, postural hypotension
Hyponatremia (due to loss of feedback inhibition of AVP release)
Signs and Symptoms Caused by Mineralocorticoid Deficiency (Primary AI Only)
Abdominal pain, nausea, vomiting
Dizziness, postural hypotension
Salt craving
Low blood pressure, postural hypotension
Increased serum creatinine (due to volume depletion)
Hyponatremia

Hyperkalemia
Signs and Symptoms Caused by Adrenal Androgen Deficiency
Lack of energy
Dry and itchy skin (in women)
Loss of libido (in women)
Loss of axillary and pubic hair (in women)
Other Signs and Symptoms
Hyperpigmentation (primary AI only) [due to excess of pro-opiomelanocortin (POMC)–derived peptides]
Alabaster-colored pale skin (secondary AI only) (due to deficiency of POMC-derived peptides)

Chronic adrenal insufficiency manifests with relatively nonspecific signs and symptoms such as fatigue and loss of energy, often resulting in delayed or missed diagnoses (e.g., as depression or anorexia). A distinguishing feature of primary adrenal insufficiency is hyperpigmentation, which is caused by excess ACTH stimulation of melanocytes. Hyperpigmentation is most pronounced in skin areas exposed to increased friction or shear stress and is increased by sunlight **(Fig. 342-14)**. Conversely, in secondary adrenal insufficiency, the skin has an alabaster-like paleness due to lack of ACTH secretion.

Figure 342-14

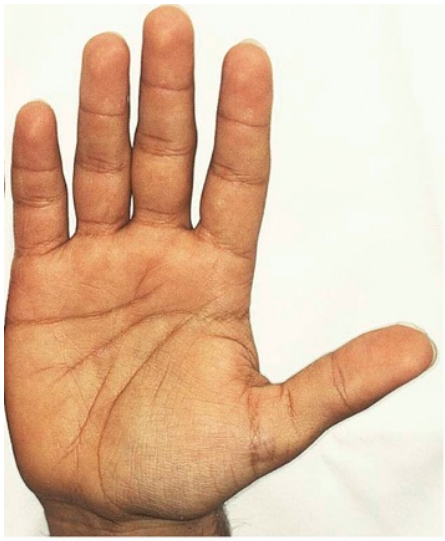

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

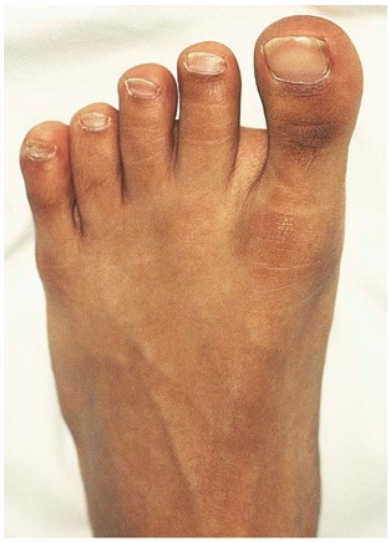

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical features of Addison's disease. Note the hyperpigmentation in areas of increased friction including (***A***) palmar creases, (***B***) dorsal foot, (***C***) nipples and axillary region, and (***D***) patchy hyperpigmentation of the oral mucosa.

Hyponatremia is a characteristic biochemical feature in primary adrenal insufficiency and is found in 80% of patients at presentation. Hyperkalemia is present in 40% of patients at initial diagnosis. Hyponatremia is primarily caused by mineralocorticoid deficiency but can also occur in secondary adrenal insufficiency due to diminished inhibition of ADH by cortisol, resulting in mild syndrome of inappropriate secretion of antidiuretic hormone (SIADH). Glucocorticoid deficiency also results in slightly increased TSH concentrations that normalize within days to weeks after initiation of glucocorticoid replacement.

Acute adrenal insufficiency usually occurs after a prolonged period of nonspecific complaints and is more frequently observed in patients with primary adrenal insufficiency, due to the loss of both glucocorticoid and mineralocorticoid secretion. Postural hypotension may progress to hypovolemic shock. Adrenal insufficiency may mimic features of acute abdomen with abdominal tenderness, nausea, vomiting, and fever. In some cases, the primary presentation may resemble neurologic disease, with decreased responsiveness, progressing to stupor and coma. An adrenal crisis can be triggered by an intercurrent illness, surgical or other stress, or increased glucocorticoid inactivation (e.g., hyperthyroidism).

DIAGNOSIS

The diagnosis of adrenal insufficiency is established by the short cosyntropin test, a safe and reliable tool with excellent predictive diagnostic value **(Fig. 342-15)**. The cut-off for failure is usually defined at cortisol levels of <500–550 nmol/L (18–20 μg/dL) sampled 30–60 minutes after ACTH stimulation; the exact cut-off is dependent on the locally available assay. During the early phase of HPA disruption (e.g., within 4 weeks of pituitary insufficiency), patients may still respond to exogenous ACTH stimulation. In this circumstance, the insulin tolerance test is an alternative choice but is more invasive and should be carried out only under a specialist's supervision (see above). Induction of hypoglycemia is contraindicated in individuals with diabetes mellitus, cardiovascular disease, or history of seizures. Random serum cortisol measurements are of limited diagnostic value, as baseline cortisol levels may be coincidentally low due to the physiologic diurnal rhythm of cortisol secretion (Fig. 342-3). Similarly, many patients with secondary adrenal insufficiency have relatively normal baseline cortisol levels but fail to mount an appropriate cortisol response to ACTH, which can only be revealed by stimulation testing. Importantly, tests to establish the diagnosis of adrenal insufficiency should never delay treatment. Thus, in a patient with suspected adrenal crisis, it is reasonable to draw baseline cortisol levels, provide replacement therapy, and defer formal stimulation testing until a later time.

Figure 342-15

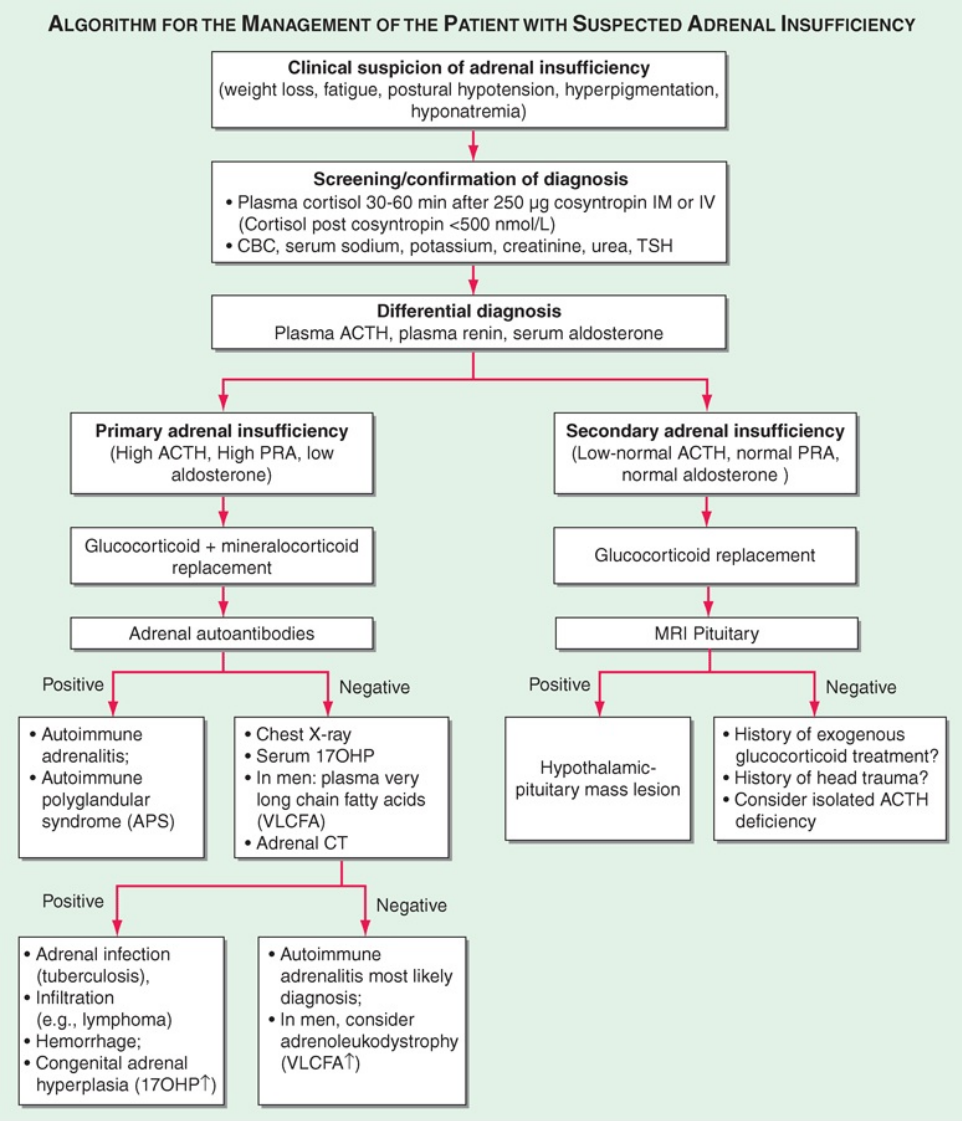

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Management of the patient with suspected adrenal insufficiency. PRA, plasma renin activity.

Once adrenal insufficiency is confirmed, measurement of plasma ACTH is the next step, with increased or inappropriately low levels defining primary and secondary origin of disease, respectively (Fig. 342-15). In primary adrenal insufficiency, increased plasma renin will confirm the presence of mineralocorticoid deficiency. At initial presentation, patients with primary adrenal insufficiency should undergo screening for steroid autoantibodies as a marker of autoimmune adrenalitis. If these tests are negative, adrenal imaging by CT is indicated to investigate possible hemorrhage, infiltration, or masses. In male patients with negative autoantibodies in the plasma, very long chain fatty acids should be measured to exclude X-ALD. Patients with inappropriately low ACTH, in the presence of confirmed cortisol deficiency, should undergo hypothalamic-pituitary imaging by MRI. Features suggestive of preceding pituitary apoplexy such as sudden-onset severe headache, or history of previous head trauma, should be carefully explored, particularly in patients with no obvious MRI lesion.

TREATMENT: ACUTE ADRENAL INSUFFICIENCY

Acute adrenal insufficiency requires immediate initiation of rehydration, usually carried out by saline infusion at initial rates of 1 L/h with continuous cardiac monitoring. Glucocorticoid replacement should be initiated by bolus injection of 100 mg hydrocortisone, followed by the administration of 100–200 mg hydrocortisone over 24 h, either by continuous infusion or provided by several IV or IM injections. Mineralocorticoid replacement can be initiated once the daily hydrocortisone dose has been reduced to <50 mg because at higher doses hydrocortisone provides sufficient stimulation of mineralocorticoid receptors.

Glucocorticoid replacement for the treatment of chronic adrenal insufficiency should be administered at a dose that replaces the physiologic daily cortisol production, which is usually achieved by the oral administration of 15–25 mg hydrocortisone in two to three divided doses. Pregnancy may require an increase in hydrocortisone dose by 50% during the last trimester. In all patients, at least one-half of the daily dose should be administered in the morning. Currently available glucocorticoid preparations fail to mimic the physiologic cortisol secretion rhythm (Fig. 342-3). Long-acting glucocorticoids such as prednisolone or dexamethasone are not preferred as they result in increased glucocorticoid exposure due to extended glucocorticoid receptor activation at times of physiologically low cortisol secretion. There are no well-established dose equivalencies, but as a guide, equipotency can be assumed for 1 mg hydrocortisone, 1.6 mg cortisone acetate, 0.2 mg prednisolone, 0.25 mg prednisone, and 0.025 mg dexamethasone.

Monitoring of glucocorticoid replacement is mainly based on the history and examination for signs and symptoms suggestive of glucocorticoid over- or under-replacement, including assessment of body weight and blood pressure. Plasma ACTH, 24-hour urinary free cortisol, or serum cortisol day curves reflect whether hydrocortisone has been taken or not, but do not convey reliable information about replacement quality. In patients with isolated primary adrenal insufficiency, monitoring should include screening for autoimmune thyroid disease, and female patients should be made aware of the possibility of premature ovarian

failure. Supraphysiologic glucocorticoid treatment with doses equivalent to 30 mg hydrocortisone or more will affect bone metabolism, and these patients should undergo regular bone mineral density evaluation. All patients with adrenal insufficiency need to be instructed about the requirement for stress-related glucocorticoid dose adjustments. These generally consist of doubling the routine oral glucocorticoid dose in the case of intercurrent illness with fever and bedrest and the need for IV hydrocortisone injection at a daily dose of 100 mg in cases of prolonged vomiting, surgery, or trauma. Patients living or traveling in regions with delayed access to acute health care should carry a hydrocortisone self-injection emergency kit, in addition to their usual steroid emergency cards and bracelets.

Mineralocorticoid replacement in primary adrenal insufficiency should be initiated at a dose of 100–150 μg fludrocortisone. The adequacy of treatment can be evaluated by measuring blood pressure, sitting and standing, to detect a postural drop indicative of hypovolemia. In addition, serum sodium, potassium, and plasma renin should be measured regularly. Renin levels should be kept in the upper normal reference range. Changes in glucocorticoid dose may also impact on mineralocorticoid replacement as cortisol also binds the mineralocorticoid receptor; 40 mg hydrocortisone is equivalent to 100 μg fludrocortisone. In patients living or traveling in areas with hot or tropical weather conditions, the fludrocortisone dose should be increased by 50–100 μg during the summer. Mineralocorticoid dose may also need to be adjusted during pregnancy, due to the antimineralocorticoid activity of progesterone, but this is less often required than hydrocortisone dose adjustment. Plasma renin cannot serve as a monitoring tool during pregnancy, as renin rises physiologically during gestation.

Adrenal androgen replacement is an option in patients with lack of energy, despite optimized glucocorticoid and mineralocorticoid replacement. It may also be indicated in women with features of androgen deficiency, including loss of libido. Adrenal androgen replacement can be achieved by once-daily administration of 25–50 mg DHEA. Treatment is monitored by measurement of DHEAS, androstenedione, testosterone, and SHBG 24 hours after the last DHEA dose.

Congenital Adrenal Hyperplasia

(See also Chap. 349) Congenital adrenal hyperplasia (CAH) is caused by mutations in genes encoding steroidogenic enzymes involved in glucocorticoid synthesis (CYP21A2, CYP17A1, HSD3B2, CYP11B1) or in the cofactor enzyme P450 oxidoreductase that serves as an electron donor to CYP21A2 and CYP17A1 (Fig. 342-1). Invariably, patients affected by CAH exhibit glucocorticoid deficiency. Depending on the exact step of enzymatic block, they may also have excess production mineralocorticoids or deficient production of sex steroids **(Table 342-10)**. The diagnosis of CAH is readily established by measurement of the steroids accumulating before the distinct enzymatic block, either in serum or in urine, preferably by the use of mass spectrometry–based assays (Table 342-10).

Table 342-10 Variants of Congenital Adrenal Hyperplasia

Variant	Gene	Impact on Steroid Synthesis	Diagnostic Marker Steroids in Serum (and Urine)
21-Hydroxylase deficiency (21OHD)	*CYP21A2*	Glucocorticoid deficiency, mineralocorticoid deficiency, adrenal androgen excess	17-Hydroxyprogesterone, 21-deoxycortisol (pregnanetriol, 17-hydroxypregnanolone, pregnanetriolone)
11β-Hydroxylase deficiency (11OHD)	*CYP11B1*	Glucocorticoid deficiency, mineralocorticoid excess, adrenal androgen excess	11-Deoxycortisol, 11-deoxycorticosterone (tetrahydro-11-deoxycortisol, tetrahydro-11-deoxycorticosterone)
17α-Hydroxylase deficiency (17OHD)	*CYP17A1*	(Glucocorticoid deficiency), mineralocorticoid excess, androgen deficiency	11-Deoxycorticosterone, corticosterone, pregnenolone, progesterone (tetrahydro-11-deoxycorticosterone, tetrahydrocorticosterone, pregnenediol, pregnanediol)
3β-Hydroxysteroid dehydrogenase deficiency (3bHSDD)	*HSD3B2*	Glucocorticoid deficiency, (mineralocorticoid deficiency), adrenal androgen excess	17-Hydroxypregnanolone (pregnanetriol)
P450 oxidoreductase deficiency (ORD)	*POR*	Glucocorticoid deficiency, (mineralocorticoid excess), androgen deficiency, skeletal malformations	Pregnenolone, progesterone, 17-hydroxyprogesterone (pregnanediol, pregnanetriol)

Mutations in CYP21A2 are the most prevalent cause of CAH, responsible for 90–95% of cases. 21-Hydroxylase deficiency disrupts glucocorticoid and mineralocorticoid synthesis (Fig. 342-1), resulting in diminished negative feedback via the HPA axis. This leads to increased pituitary ACTH release, which drives increased synthesis of adrenal androgen precursors and subsequent androgen excess. The degree of impairment of glucocorticoid and mineralocorticoid secretion depends on the severity of mutations. Major loss-of-function mutations result in combined glucocorticoid and mineralocorticoid deficiency (classic CAH, neonatal presentation), whereas less severe mutations affect glucocorticoid synthesis only (simple virilizing CAH, neonatal or early childhood presentation). The mildest mutations result in the least severe clinical phenotype, nonclassical CAH, usually presenting during adolescence and early adulthood and with preserved glucocorticoid production.

Androgen excess is present in all patients and manifests with broad phenotypic variability, ranging from severe virilization of the external genitalia in neonatal girls (e.g., 46,XX DSD) to hirsutism and oligomenorrhea resembling a polycystic ovary syndrome phenotype in young women with nonclassic CAH. In countries without neonatal screening for CAH, boys with classic CAH usually present with life-threatening adrenal crisis in the first few weeks of life (salt-wasting crisis); a simple-virilizing genotype manifests with precocious pseudo-puberty and advanced bone age in early childhood, whereas men with nonclassic CAH are usually detected only through family screening.

Glucocorticoid treatment is more complex than for other causes of primary adrenal insufficiency. It not only needs to replace missing glucocorticoids but also aims to suppress the increased ACTH drive and subsequent androgen excess. Current treatment is hampered by the lack of glucocorticoid preparations that mimic the diurnal cortisol secretion profile, resulting in a prolonged period of ACTH stimulation and subsequent androgen production during the early morning hours. In childhood, optimization of growth and pubertal development are important goals of glucocorticoid treatment, in addition to prevention of adrenal crisis and treatment of 46,XX DSD. In adults, the focus shifts to preserving fertility and preventing side effects of glucocorticoid overtreatment, namely, the metabolic syndrome and osteoporosis. Fertility can be compromised in women due to oligo/amenorrhea with chronic anovulation as a consequence of androgen excess. Men may develop so-called testicular adrenal rest tumors (Fig. 342-16). These consist of hyperplastic cells with adrenocortical characteristics located in the rete testis and should not be confused with testicular tumors. Testicular adrenal rest tumors can compromise sperm production and induce fibrosis that may be irreversible.

Figure 342-16

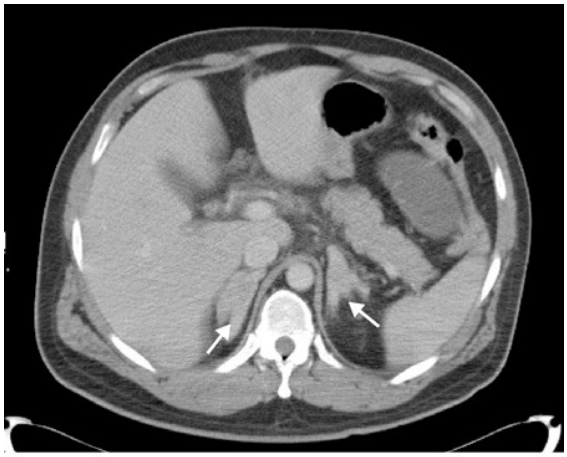

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

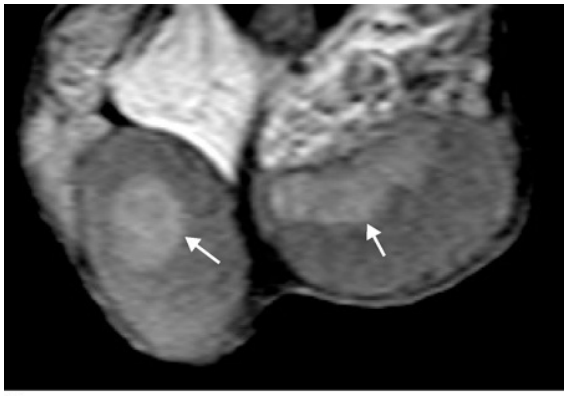

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Imaging in congenital adrenal hyperplasia (CAH). Adrenal CT scans showing homogenous bilateral hyperplasia in a young patient with classic CAH ***A*** and macronodular bilateral hyperplasia ***B*** in a middle-aged classic CAH patient with longstanding poor disease control. MRI scan with T1-weighted ***C*** and T2-weighted ***D*** images showing bilateral testicular adrenal rest tumors (arrows) in a young patient with salt-wasting congenital adrenal hyperplasia. (Courtesy of N. Reisch.)

TREATMENT: CONGENITAL ADRENAL HYPERPLASIA

Hydrocortisone is a good treatment option for the prevention of adrenal crisis but longer-acting prednisolone may be needed to control androgen excess. In children, hydrocortisone is given in divided doses at 1–1.5 times the normal cortisol production rate (about 10–13 mg/m^2 per day). In adults, intermediate-acting glucocorticoids (e.g., prednisone) may be given, using the lowest dose necessary to suppress excess androgen production. For achieving fertility, dexamethasone treatment may be required, but should be only given for the shortest possible time period to limit metabolic side effects. Biochemical monitoring should include androstenedione and testosterone, aiming for the normal sex-specific reference range. 17-Hydroxyprogesterone (17OHP) is a useful marker of overtreatment, indicated by 17OHP levels within the normal range of healthy controls. Glucocorticoid overtreatment may suppress the hypothalamic-pituitary-gonadal axis. Thus, treatment needs to be carefully titrated against clinical features of disease control. Stress dose glucocorticoids should be given at double or triple the daily dose for surgery, acute illness, or severe trauma. Poorly controlled CAH can result in adrenocortical hyperplasia, which gave the disease its name, and may present as macronodular hyperplasia subsequent to long-standing ACTH excess **(Fig. 342-15)**. The nodular areas can develop autonomous adrenal androgen production, and may be unresponsive to glucocorticoid treatment.

Mineralocorticoid requirements change during life and are higher in children, explained by relative mineralocorticoid resistance that diminishes with ongoing maturation of the kidney. Children with CAH usually receive mineralocorticoid and salt replacement. However, young adults with CAH should undergo reassessment of their mineralocorticoid reserve. Plasma renin should be regularly monitored and kept within the upper half of the normal reference range.

FURTHER READINGS

Arlt W et al: Health status of adults with congenital adrenal hyperplasia: A cohort study of 203 patients. J Clin Endocrinol Metab 95:5110, 2010[PMID: 20719839] [Full Text]

Biller BM et al: Treatment of adrenocorticotropin-dependent Cushing's syndrome: A consensus statement. J Clin Endocrinol Metab 93:2454, 2008[PMID: 18413427] [Full Text]

Debono M et al: Novel strategies for hydrocortisone replacement. Best Pract Res Clin Endocrinol Metab 23:221, 2009[PMID: 19500765] [Full Text]

Fassnacht M, Allolio B: Clinical management of adrenocortical carcinoma. Best Pract Res Clin Endocrinol Metab 23:273, 2009[PMID: 19500769] [Full Text]

Funder JW et al: Case detection, diagnosis, and treatment of patients with primary aldosteronism: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 93:3266, 2008[PMID: 18552288] [Full Text]

Krone N, Arlt W: Genetics of congenital adrenal hyperplasia. Best Pract Res Clin Endocrinol Metab 23:181, 2009[PMID: 19500762] [Full Text]

Nieman LK et al: The diagnosis of Cushing's syndrome: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 93:1526, 2008[PMID: 18334580] [Full Text]

Speiser PW et al: Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 95:4133, 2010[PMID: 20823466] [Full Text]

Young WF Jr: Clinical practice. The incidentally discovered adrenal mass. N Engl J Med 356:601, 2007[PMID: 17287480] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PHEOCHROMOCYTOMA: INTRODUCTION

Pheochromocytomas and paragangliomas are catecholamine- producing tumors derived from the sympathetic or parasympathetic nervous system. These tumors may arise sporadically or be inherited as features of multiple endocrine neoplasia type 2 or several other pheochromocytoma-associated syndromes. The diagnosis of pheochromocytomas provides a potentially correctable cause of hypertension, and their removal can prevent hypertensive crises that can be lethal. The clinical presentation is variable, ranging from an adrenal incidentaloma to a patient in hypertensive crisis with associated cerebrovascular or cardiac complications.

Epidemiology

Pheochromocytoma is estimated to occur in 2–8 of 1 million persons per year, and about 0.1% of hypertensive patients harbor a pheochromocytoma. Autopsy series reveal prevalence of 0.2%. The mean age at diagnosis is about 40 years, although the tumors can occur from early childhood until late in life. The "rule of tens" for pheochromocytomas states that about 10% are bilateral, 10% are extraadrenal, and 10% are malignant. However, these percentages are higher in the inherited syndromes.

Etiology and Pathogenesis

Pheochromocytomas and paragangliomas are well-vascularized tumors that arise from cells derived from the sympathetic (e.g., adrenal medulla) or parasympathetic (e.g., carotid body, glomus vagale) paraganglia (Fig. 343-1). The name *pheochromocytoma* reflects the black-colored staining caused by chromaffin oxidation of catecholamines. Although a variety of terms have been used to describe these tumors, most clinicians use the term *pheochromocytoma* to describe symptomatic catecholamine-producing tumors, including those in extraadrenal retroperitoneal, pelvic, and thoracic sites. The term *paraganglioma* is used to describe catecholamine-producing tumors in the head and neck. These tumors may secrete little or no catecholamines.

Figure 343-1

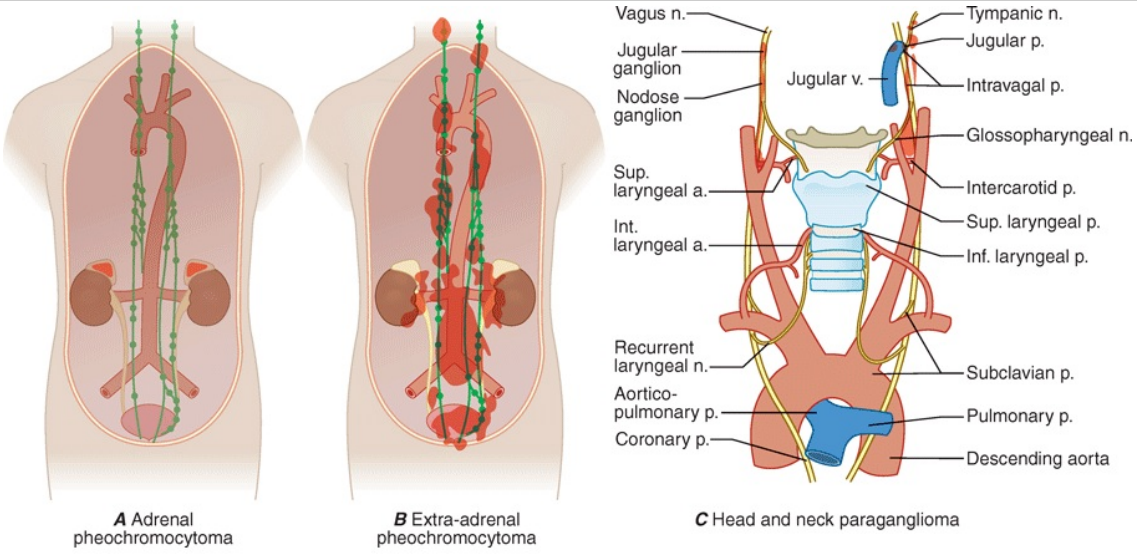

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The paraganglial system and topographic sites (in red) of pheochromocytomas and paragangliomas. *[Parts **A**,**B**, from WM Manger, RW Gifford: Clinical and experimental pheochromocytoma. Cambridge, Blackwell Science, 1996; Part C, from GG Glenner, PM Grimley: Tumors of the Extra-adrenal Paraganglion System (Including Chemoreceptors), Atlas of Tumor Pathology, 2nd Series, Fascicle 9. Washington, DC, AFIP, 1974.]*

The etiology of sporadic pheochromocytomas and paragangliomas is unknown. However, about 25% of patients have an inherited condition, including germ-line mutations in the *RET, VHL, NF1, SDHB, SDHC, SDHD,* or *SDHAF2* genes. Biallelic gene inactivation has been demonstrated for the *VHL, NF1*, and *SDH* genes, whereas *RET* mutations activate the receptor tyrosine kinase activity. SDH is an enzyme of the Krebs cycle and the mitochondrial respiratory chain. The VHL protein is a component of a ubiquitin E3 ligase. *VHL* mutations reduce protein degradation, resulting in upregulation of components involved in cell cycle progression, glucose metabolism, and oxygen sensing.

Clinical Features

The clinical presentation is so variable that pheochromocytoma has been termed "the great masquerader" (Table 343-1). Among the presenting symptoms, episodes of palpitations, headaches, and profuse sweating are typical and constitute a classic triad. The presence of all three symptoms in association with hypertension makes pheochromocytoma a likely diagnosis. However, a pheochromocytoma can be asymptomatic for years, and some tumors grow to a considerable size before patients note symptoms.

Table 343-1 Clinical Features Associated with Pheochromocytoma
Headaches

Sweating attacks
Palpitations and tachycardia Hypertension, sustained or paroxysmal
Anxiety and panic attacks
Pallor
Nausea
Abdominal pain
Weakness
Weight loss Paradoxical response to antihypertensive drugs
Polyuria and polydipsia
Constipation
Orthostatic hypotension
Dilated cardiomyopathy
Erythrocytosis
Elevated blood sugar Hypercalcemia

The dominant sign is hypertension. Classically, patients have episodic hypertension, but sustained hypertension is also common. Catecholamine crises can lead to heart failure, pulmonary edema, arrhythmias, and intracranial hemorrhage. During episodes of hormone release, which can occur at very divergent intervals, patients are anxious and pale, and they experience tachycardia and palpitations. These paroxysms generally last less than an hour and may be precipitated by surgery, positional changes, exercise, pregnancy, urination (particularly bladder pheochromocytomas), and various medications (e.g., tricyclic antidepressants, opiates, metoclopramide).

Diagnosis

The diagnosis is based on documentation of catecholamine excess by biochemical testing and localization of the tumor by imaging. Both are of equal importance, although measurement of catecholamines is traditionally the first step.

BIOCHEMICAL TESTING

Pheochromocytomas and paragangliomas synthesize and store catecholamines, which include norepinephrine (noradrenaline), epinephrine (adrenaline), and dopamine. Elevated plasma and urinary levels of catecholamines and the methylated metabolites, metanephrines, are the cornerstone for the diagnosis. The hormonal activity of tumors fluctuates, resulting in considerable variation in serial catecholamine measurements. Thus, there is some value in obtaining tests during or soon after a symptomatic crisis. However, most tumors continuously leak O-methylated metabolites, which are detected by measurements of metanephrines.

Catecholamines and metanephrines can be measured by using different methods (e.g., high-performance liquid chromatography, enzyme-linked immunosorbent assay, and liquid chromatography/mass spectrometry). In a clinical context suspicious for pheochromocytoma, when values are increased three times the upper limit of normal, a pheochromocytoma is highly likely regardless of the assay used. However, as summarized in Table 343-2, the sensitivity and specificity of available biochemical tests vary greatly, and these differences are important in assessing patients with borderline elevations of different compounds. Urinary tests for vanillylmandelic acid (VMA), metanephrines (total or fractionated), and catecholamines are widely available and are used commonly for initial testing. Among these tests, the fractionated metanephrines and catecholamines are the most sensitive. Plasma tests are more convenient and include measurements of catecholamines and metanephrines. Measurements of plasma metanephrine are the most sensitive and are less susceptible to false-positive elevations from stress, including venipuncture. Although the incidence of false-positive test results has been reduced by the introduction of newer assays, physiologic stress responses and medications that increase catecholamines still can confound testing. Because the tumors are relatively rare, borderline elevations are likely to be false positives. In this circumstance, it is important to exclude diet or drug exposure (withdrawal of levodopa, sympathomimetics, diuretics, tricyclic antidepressants, alpha and beta blockers) that might cause false positives and then repeat testing or perform a clonidine suppression test (measurement of plasma metanephrines 3 h after oral administration of 300 μg of clonidine). Other pharmacologic tests, such as the phentolamine test and the glucagon provocation test, are of relatively low sensitivity and are not recommended.

Table 343-2 Biochemical and Imaging Methods Used for Pheochromocytoma and Paraganglioma Diagnosis

Diagnostic Method	Sensitivity	Specificity
24-h urinary tests		
Vanillylmandelic acid	++	++++
Catecholamines	+++	+++
Fractionated metanephrines	++++	++
Total metanephrines	+++	++++
Plasma tests		
Catecholamines	+++	++
Free metanephrines	++++	+++
CT	++++	+++
MRI	++++	+++
MIBG scintigraphy	+++	++++
Somatostatin receptor scintigraphy*	++	++
Dopa (dopamine) PET	+++	++++

*Particularly high in head and neck paragangliomas.

Abbreviations: MIBG, metaiodobenzylguanidine; PET, positron emission tomography.

DIAGNOSTIC IMAGING

A variety of methods have been used to localize pheochromocytomas and paragangliomas (Table 343-2). CT and MRI are similar in sensitivity. CT should be performed with contrast. T2-weighted MRI with gadolinium contrast is optimal for detecting pheochromocytomas and is somewhat better than CT for imaging extra-adrenal pheochromocytomas and paragangliomas. About 5% of adrenal incidentalomas, which usually are detected by CT or MRI, prove to be pheochromocytomas after endocrinologic evaluation.

Tumors also can be localized by using radioactive tracers, including 131I- or 123I-metaiodobenzylguanidine (MIBG), 111In-somatostatin analogues, or 18F-dopa (or dopamine) positron emission tomography (PET). Because these agents exhibit selective uptake in paragangliomas, nuclear imaging is particularly useful in the hereditary syndromes.

DIFFERENTIAL DIAGNOSIS

When one is entertaining the possibility of a pheochromocytoma, other disorders to consider include essential hypertension, anxiety attacks, use of cocaine or amphetamines, mastocytosis or carcinoid syndrome (usually lacking hypertension), intracranial lesions, clonidine withdrawal, autonomic epilepsy, and factitious crises (usually from sympathomimetic amines). When an asymptomatic adrenal mass is identified, likely diagnoses other than pheochromocytoma include a nonfunctioning adrenal adenoma, aldosteronoma, and cortisol-producing adenoma (Cushing's syndrome).

Treatment: Pheochromocytoma

Complete tumor removal is the ultimate therapeutic goal. Preoperative patient preparation is essential for safe surgery. α-Adrenergic blockers (phenoxybenzamine) should be initiated at relatively low doses (e.g., 5–10 mg orally three times per day) and increased as tolerated every few days. Because patients are volume-constricted, liberal salt intake and hydration are necessary to avoid orthostasis. Adequate alpha blockade generally requires 7 days, with a typical final dose of 20–30 mg phenoxybenzamine three times per day. Oral prazosin or intravenous phentolamine can be used to manage paroxysms while awaiting adequate alpha blockade. Before surgery, blood pressure should be consistently below 160/90 mmHg, with moderate orthostasis. Beta blockers (e.g., 10 mg propranolol three to four times per day) can be added after starting alpha blockers and increased as needed if tachycardia persists. Other antihypertensives, such as calcium channel blockers or angiotensin-converting enzyme inhibitors, have been used when blood pressure is difficult to control with phenoxybenzamine alone.

Surgery should be performed by teams of anesthesiologists and surgeons with experience in the management of pheochromocytomas. Blood pressure can be labile during surgery, particularly at the onset of intubation or when the tumor is manipulated. Nitroprusside infusion is useful for intraoperative hypertensive crises, and hypotension usually responds to volume infusion. Although laparotomy was the traditional surgical approach, endoscopic surgery, using either a transperitoneal or a retroperitoneal approach, is associated with fewer complications, a faster recovery, and optimal cosmetic results. Atraumatic endoscopic surgery has become the method of choice. It may be possible to preserve the normal adrenal cortex, particularly in hereditary disorders in which bilateral pheochromocytomas are more likely. Extra-adrenal abdominal as well as most thoracic pheochromocytomas also can be removed endoscopically. Postoperatively, catecholamine normalization should be documented. An adrenocorticotropic hormone test should be used to exclude cortisol deficiency when bilateral adrenal cortex–sparing surgery is performed.

Malignant Pheochromocytoma

About 5–10% of pheochromocytomas and paragangliomas are malignant. The diagnosis of malignant pheochromocytoma is problematic. Typical histologic criteria of cellular atypia, presence of mitoses, and invasion of vessels or adjacent tissues do not reliably identify which tumors have the capacity to metastasize. Thus, the term *malignant pheochromocytoma* generally is restricted to tumors with distant metastases, most commonly found in lungs, bone, or liver, suggesting a vascular pathway of spread. Because hereditary syndromes are associated with multifocal tumor sites, these features should be anticipated in patients with germ-line mutations of *RET, VHL, SDHD, or SDHB*. However, distant metastases also occur in these syndromes, especially in carriers of *SDHB* mutations.

Treatment of malignant pheochromocytoma or paraganglioma is challenging. Options include tumor mass reduction, alpha blockers for symptoms, chemotherapy, and nuclear medicine radiotherapy. Averbuch's chemotherapy protocol includes dacarbazine (600 mg/m^2 days 1 and 2), cyclophosphamide (750 mg/m^2 day 1), and vincristine (1.4 mg/m^2 day 1), repeated every 21 days for three to six cycles. Palliation (stable disease to shrinkage) is achieved in about one-half of patients. Other chemotherapeutic protocols remain in the experimental stage. An alternative is 1^{31}I-MIBG treatment using 200-mCi doses at monthly intervals over three to six cycles. The prognosis of metastatic pheochromocytoma or paraganglioma is variable, with a 5-year survival of 30–60%.

Pheochromocytoma in Pregnancy

Pheochromocytomas occasionally are diagnosed in pregnancy. Endoscopic removal, preferably in the forth to sixth month of gestation, is possible and can be followed by uneventful childbirth. Regular screening in families with inherited pheochromocytomas provides an opportunity to identify and remove asymptomatic tumors in women of reproductive age.

Pheochromocytoma-Associated Syndromes

About 25–33% of patients with a pheochromocytoma or paraganglioma have an inherited syndrome. The mean age at diagnosis is about 15 years lower in patients with inherited syndromes compared with patients with sporadic tumors.

Neurofibromatosis type 1 (NF 1) was the first described pheochromocytoma-associated syndrome (Chap. 379). The *NF1* gene functions as a tumor suppressor by regulating the Ras signaling cascade. Classic features of neurofibromatosis include multiple neurofibromas, café au lait spots, axillary freckling of the skin, and Lisch nodules of the iris (Fig. 343-2). Pheochromocytomas occur in only about 1% of these patients and are located predominantly in the adrenals. Malignant pheochromocytoma is not uncommon.

Figure 343-2

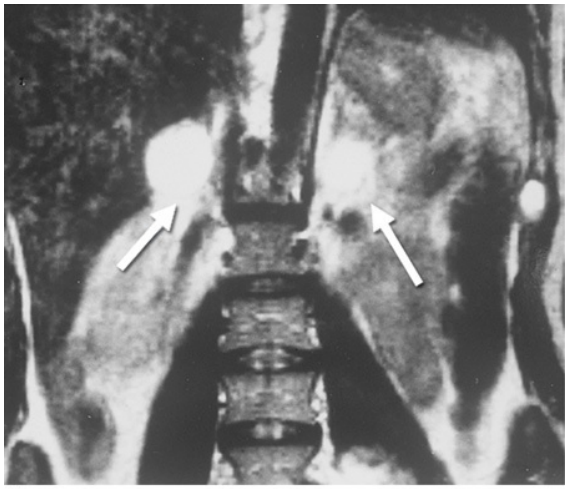

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

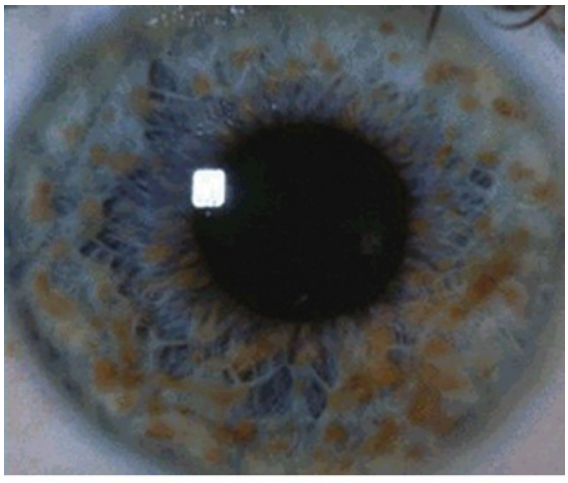

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurofibromatosis. A. MRI of bilateral adrenal pheochromocytoma. **B**. Cutaneous neurofibromas. **C**. Lisch nodules of the iris. **D**. Axillary freckling. (*Part A from HPH Neumann et al: Keio J Med 54:15, 2005; with permission.*)

The best-known pheochromocytoma-associated syndrome is the autosomal dominant disorder *multiple endocrine neoplasia type 2A and type 2B (MEN 2A, MEN 2B)* (Chap. 351). Both types of MEN 2 are caused by mutations in *RET* (*RE*arranged during *T*ransfection), which encodes a tyrosine kinase. The locations of RET mutations correlate with the severity of disease and the type of MEN 2 (Chap. 351). MEN 2A is characterized by medullary thyroid carcinoma (MTC), pheochromocytoma, and hyperparathyroidism; MEN 2B also includes MTC and pheochromocytoma, as well as multiple mucosal neuromas, marfanoid habitus, and other developmental disorders, though it typically lacks hyperparathyroidism. MTC is seen in virtually all patients with MEN 2, but pheochromocytoma occurs in only about 50% of these patients. Nearly all pheochromocytomas are benign and located in the adrenals, often bilateral (Fig. 343-3). Pheochromocytoma may be symptomatic before MTC. Prophylactic thyroidectomy is being performed in many carriers of *RET* mutations; pheochromocytomas should be excluded before any surgery in these patients.

Figure 343-3

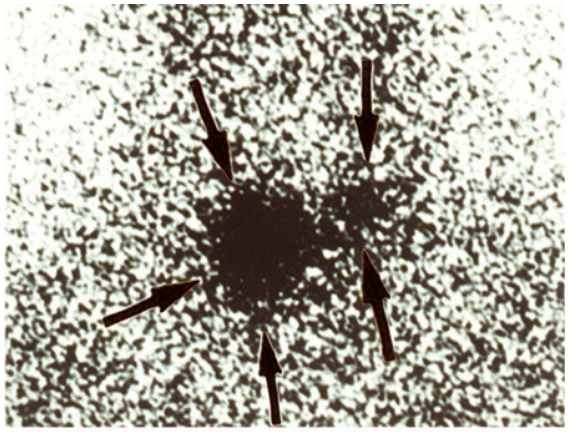

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

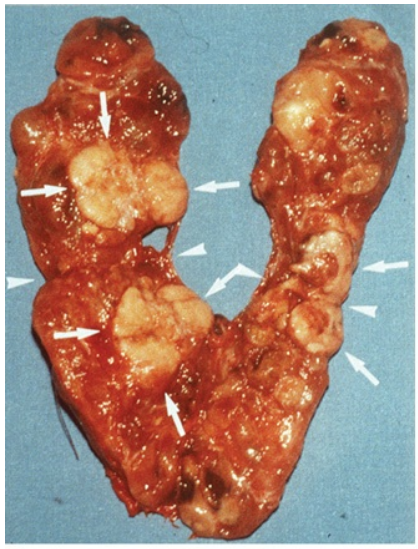

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

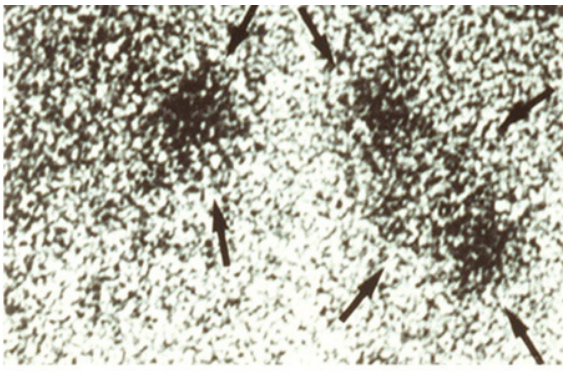

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

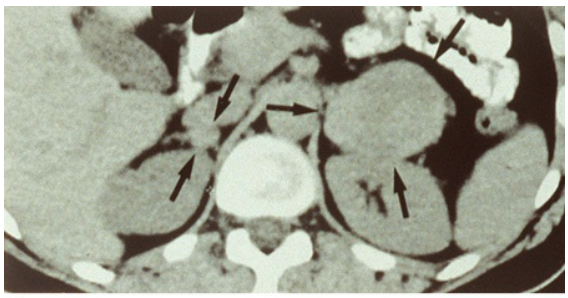

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

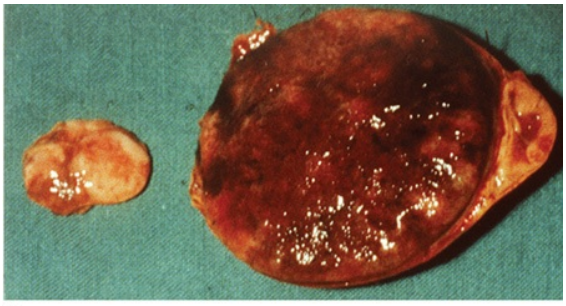

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple endocrine neoplasia type 2. Multifocal medullary thyroid carcinoma shown by (***A***) MIBG scintigraphy and (***B***) operative specimen Arrows demonstrate the tumors; arrowheads show the tissue bridge of the cut specimen. Bilateral adrenal pheochromocytoma shown by (***C***) MIBG scintigraphy, (***D***) CT imaging, and (***E***) operative specimens. (*From HPH Neumann et*

al: Keio J Med 54:15, 2005; with permission.)

Von Hippel-Lindau syndrome (VHL) is an autosomal dominant disorder that predisposes to retinal and cerebellar hemangioblastomas, which also occur in the brainstem and spinal cord (Fig. 343-4). Other important features of VHL are clear cell renal carcinomas, pancreatic islet cell tumors, endolymphatic sac tumors (ELSTs) of the inner ear, cystadenomas of the epididymis and broad ligament, and multiple pancreatic or renal cysts.

Figure 343-4

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

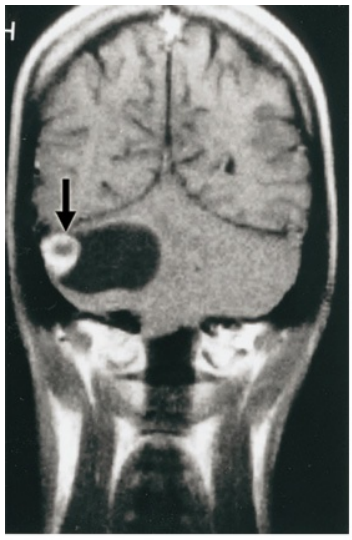

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

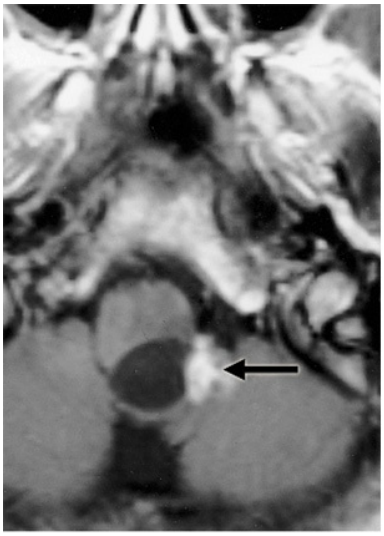

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

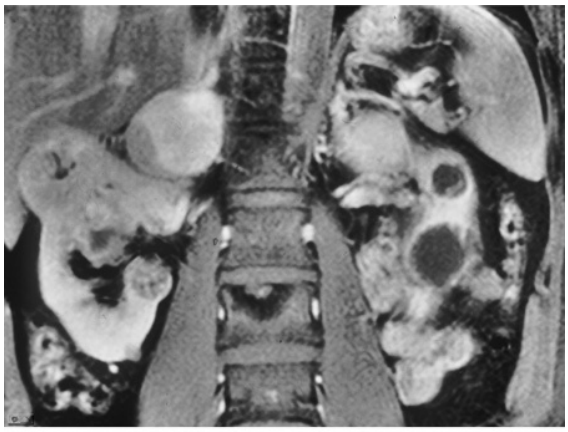

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

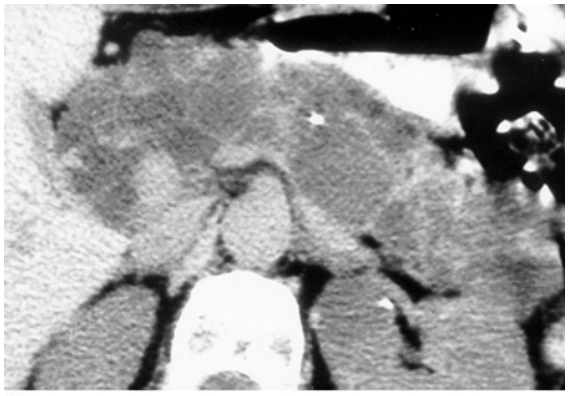

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Von Hippel-Lindau disease. Retinal angioma (*A*); hemangioblastomas of cerebellum are shown by MRI in (***B***) brainstem; (***C*** and ***D***) spinal cord; (***E***) bilateral pheochromocytomas and bilateral renal clear cell carcinomas; and (***F***) multiple pancreatic cysts. *[Parts A and D from HPH Neumann et al: Adv Nephrol Necker Hosp 27:361, 1997. Copyright Elsevier. Part B from SH Morgan, J-P Grunfeld (eds): Inherited Disorders of the Kidney, Oxford, UK, Oxford University Press, 1998. Part F from HPH Neumann et al: Contrib Nephrol 136:193, 2001. Copyright S. Karger AG, Basel.]*

The *VHL* gene encodes an E3 ubiquitin ligase that regulates expression of hypoxia-inducible factor-1 (HIF-1), among other genes. Loss of *VHL* is associated with increased expression of vascular endothelial growth factor (VEGF) that induces angiogenesis. Although the *VHL* gene can be inactivated by all types of mutations, patients with pheochromocytoma predominantly have missense mutations. About 20–30% of patients with VHL have pheochromocytomas, but in some families the incidence can reach 90%. The recognition of pheochromocytoma as a VHL-associated feature provides an opportunity to diagnose retinal, central nervous system, renal, and pancreatic tumors at a stage when effective treatment may still be possible

The *paraganglioma syndromes (PGL)* have been classified by genetic analyses of families with head and neck paragangliomas. The susceptibility genes encode subunits of the enzyme succinate dehydrogenase (SDH), a component in the Krebs cycle and the mitochondrial electron transport chain. SDH is formed by four subunits (A–D). Mutations of *SDHB* (PGL4), *SDHC* (PGL3), *SDHD* (PGL1), and *SDHAF2* (PGL2) predispose to the paraganglioma syndromes. Mutations of SDHA do not predispose to paraganglioma tumors but instead cause Leigh's disease, a form of encephalopathy. The transmission of the disease in carriers of *SDHB, SDHC*, and *SDHAF2* germ-line mutations is autosomal dominant. In contrast, in SDHD families, only the progeny of affected fathers develop tumors if they inherit the mutation. In a small number of patients with familial pheochromocytoma, a mutation has not been identified. PGL1 is most common, followed by PGL4; PGL2 and PGL3 are rare. Adrenal, extra-adrenal abdominal, and thoracic pheochromocytomas that are components of PGL1 and PGL4, are rare in PGL3, but absent in PGL2 (Fig. 343-5). About one-third of the patients with PGL4 develop metastases.

Figure 343-5

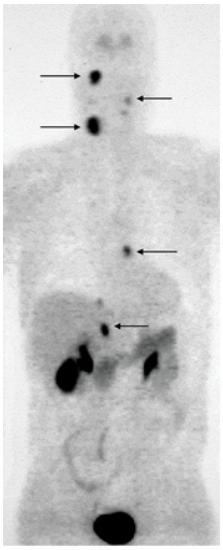

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

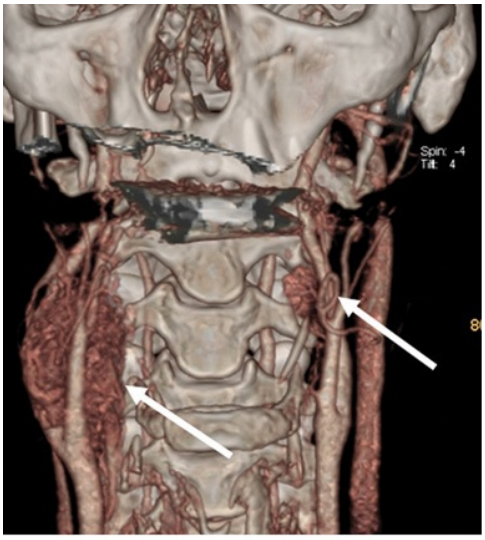

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Paraganglioma syndrome. PGL1, a patient with incomplete resection of a left carotid body tumor and the *SDHD* W5X mutation. ***A***. ^{18}F-dopa positron emission tomography demonstrating tumor uptake in the right jugular glomus, the right carotid body, the left carotid body, the left coronary glomus, and the right adrenal gland. Note the physiologic accumulation of the radiopharmaceutical agent in the kidneys, liver, gallbladder, renal pelvis, and urinary bladder. ***B*** and ***C***. CT angiography with three-dimensional reconstruction. Arrows point to the paraganglial tumors. (*From S Hoegerle et al: Eur J Nucl Med Mol Imaging 30:689, 2003; with permission.*)

Familial pheochromocytoma (FP) has been attributed to hereditary, exclusively adrenal tumors in patients with germline mutations in the TMEM127 gene.

Guidelines for Genetic Screening in Patients with Pheochromocytoma or Paraganglioma

In addition to family history, general features suggesting an inherited syndrome include young age, multifocal tumors, extra-adrenal tumors, and malignant tumors (Fig. 343-6). Because of the relatively high prevalence of familial syndromes among patients who present with pheochromocytoma or paraganglioma, it is useful to identify germline mutations even in patients without a known family history. A first step is to search for clinical features of inherited syndromes and to perform an in-depth, multigenerational family history. Each of these syndromes exhibits autosomal dominant transmission with variable penetrance, but a proband with a mother affected by paraganglial tumors is not predisposed to PLG1 (*SDHD* mutation carrier). Cutaneous neurofibromas, café au lait spots, and axillary freckling suggest neurofibromatosis. Germ-line mutations in *NF1* have not been reported in patients with sporadic pheochromocytomas. Thus, *NF1* testing does not have to be performed in the absence of other clinical features of neurofibromatosis. A personal or family history of medullary thyroid cancer or elevation of serum calcitonin strongly suggest MEN 2 and should prompt testing for *RET* mutations. A history of visual impairment, or tumors of the cerebellum, kidney, brainstem, or spinal cord, suggests the possibility of VHL. A personal and/or family history for head and neck paraganglioma suggests PGL1 or PGL4.

Figure 343-6

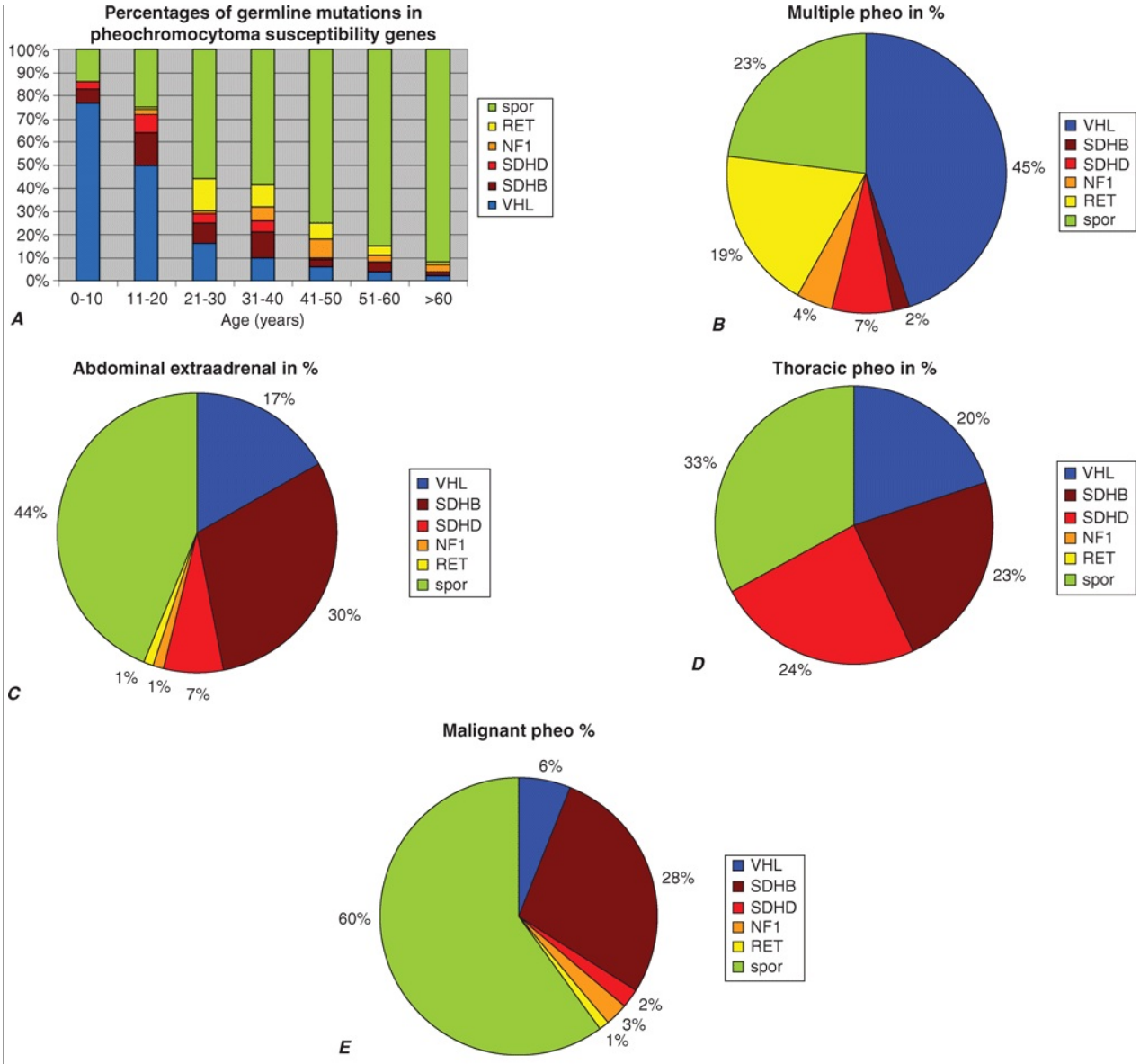

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mutation distribution in the*RET, VHL, NF1, SDHB,and SDHD genes. A*. Correlation with age. The bars depict the frequency of sporadic or various inherited forms of pheochromocytoma in different age groups. The inherited disorders are much more common among younger individuals presenting with pheochromocytoma. Germ-line mutations according to (***B***) multiple, (***C***) extraadrenal retroperitoneal, (***D***) thoracic, and (***E***) malignant pheochromocytomas (*Data from the Freiburg International Pheochromocytoma and Paraganglioma Registry in 2009.*)

A single adrenal pheochromocytoma in a patient with an otherwise unremarkable history may still be associated with mutations of *VHL, RET, SDHB*, or *SDHD* (in decreasing order of frequency). Two-thirds of extra-adrenal tumors are associated with one of these syndromes, and multifocal tumors occur with decreasing frequency in carriers of *RET, SDHD, VHL*, and *SDHB* mutations. About 30% of head and neck paragangliomas are associated with germ-line mutations of one of the SDH subunit genes (particularly *SDHD*) and are rare in carriers of *VHL* and *RET* mutations.

Once the underlying syndrome is diagnosed, the benefit of genetic testing can be extended to relatives. For this purpose, it is necessary to identify the germ-line mutation in the proband and, after genetic counseling, perform DNA sequence analyses of the responsible gene in relatives to determine whether they are affected (Chap. 63). Other family members may benefit from biochemical screening for paraganglial tumors in individuals who carry a germ-line mutation.

FURTHER READINGS

Erlic Z et al: Clinical predictors and algorithm for the genetic diagnosis of pheochromocytoma patients. Clin Cancer Res 15:6378, 2009[PMID: 19825962] [Full Text]

Lenders JW et al: Phaeochromocytoma. Lancet 366:665, 2005[PMID: 16112304] [Full Text]

Neumann HP et al: Germ-line mutations in nonsyndromic pheochromocytoma. N Engl J Med 346:1459, 2002[PMID: 12000816] [Full Text]

———: Evidence of MEN-2 in the original description of classic pheochromocytoma. N Engl J Med 357:1311, 2007

Pacak K et al: Pheochromocytoma: Recommendations for clinical practice from the First International Symposium. Nat Clin Pract Endocrinol Metab 3:92, 2007[PMID: 17237836] [Full Text]

Tischler AS: Pheochromocytoma and extra-adrenal paraganglioma: Arch Pathol Lab Med 132:1272, 2008[PMID: 18684026] [Full Text]

Yao L et al: Spectrum and prevalence of FP/TMEM127 gene mutations in pheochromocytomas and paragangliomas. JAMA 304:2611, 2010[PMID: 21156949] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DIABETES MELLITUS: INTRODUCTION

Diabetes mellitus (DM) refers to a group of common metabolic disorders that share the phenotype of hyperglycemia. Several distinct types of DM are caused by a complex interaction of genetics and environmental factors. Depending on the etiology of the DM, factors contributing to hyperglycemia include reduced insulin secretion, decreased glucose utilization, and increased glucose production. The metabolic dysregulation associated with DM causes secondary pathophysiologic changes in multiple organ systems that impose a tremendous burden on the individual with diabetes and on the health care system. In the United States, DM is the leading cause of end-stage renal disease (ESRD), nontraumatic lower extremity amputations, and adult blindness. It also predisposes to cardiovascular diseases. With an increasing incidence worldwide, DM will be a leading cause of morbidity and mortality for the foreseeable future.

CLASSIFICATION

DM is classified on the basis of the pathogenic process that leads to hyperglycemia, as opposed to earlier criteria such as age of onset or type of therapy **(Fig. 344-1)**. The two broad categories of DM are designated type 1 and type 2 **(Table 344-1)**. Both types of diabetes are preceded by a phase of abnormal glucose homeostasis as the pathogenic processes progress. Type 1 DM is the result of complete or near-total insulin deficiency. Type 2 DM is a heterogeneous group of disorders characterized by variable degrees of insulin resistance, impaired insulin secretion, and increased glucose production. Distinct genetic and metabolic defects in insulin action and/or secretion give rise to the common phenotype of hyperglycemia in type 2 DM and have important potential therapeutic implications now that pharmacologic agents are available to target specific metabolic derangements. Type 2 DM is preceded by a period of abnormal glucose homeostasis classified as impaired fasting glucose (IFG) or impaired glucose tolerance (IGT).

Figure 344-1

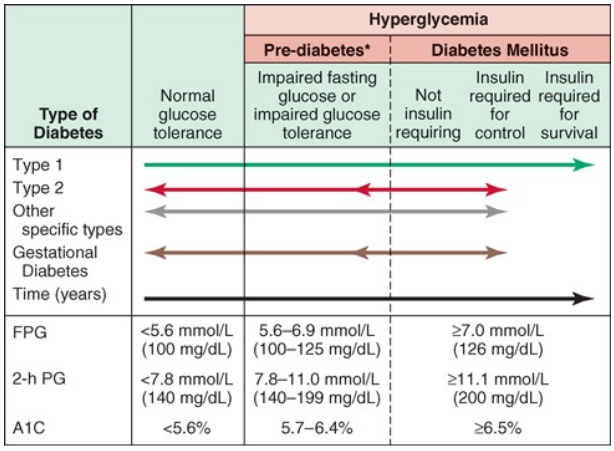

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spectrum of glucose homeostasis and diabetes mellitus (DM). The spectrum from normal glucose tolerance to diabetes in type 1 DM, type 2 DM, other specific types of diabetes, and gestational DM is shown from left to right. In most types of DM, the individual traverses from normal glucose tolerance to impaired glucose tolerance to overt diabetes (these should be viewed not as abrupt categories but as a spectrum). Arrows indicate that changes in glucose tolerance may be bidirectional in some types of diabetes. For example, individuals with type 2 DM may return to the impaired glucose tolerance category with weight loss; in gestational DM, diabetes may revert to impaired glucose tolerance or even normal glucose tolerance after delivery. The fasting plasma glucose (FPG), the 2-h plasma glucose (PG) after a glucose challenge, and the A1C for the different categories of glucose tolerance are shown at the lower part of the figure. These values do not apply to the diagnosis of gestational DM. The World Health Organization uses an FPG of 110–125 mg/dL for the prediabetes category. Some types of DM may or may not require insulin for survival. *Some use the term "increased risk for diabetes" (ADA) or "intermediate hyperglycemia" (WHO) rather than "prediabetes." *(Adapted from the American Diabetes Association, 2007.)*

Table 344-1 Etiologic Classification of Diabetes Mellitus

I. Type 1 diabetes (beta cell destruction, usually leading to absolute insulin deficiency)
A. Immune-mediated
B. Idiopathic
II. Type 2 diabetes (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly insulin secretory defect with insulin resistance)
III. Other specific types of diabetes
A. Genetic defects of beta cell function characterized by mutations in:
1. Hepatocyte nuclear transcription factor (HNF) 4α (MODY 1)
2. Glucokinase (MODY 2)
3. HNF-1α (MODY 3)
4. Insulin promoter factor-1 (IPF-1; MODY 4)

5. HNF-1β (MODY 5)
6. NeuroD1 (MODY 6)
7. Mitochondrial DNA
8. Subunits of ATP-sensitive potassium channel
9. Proinsulin or insulin

B. Genetic defects in insulin action
1. Type A insulin resistance
2. Leprechaunism
3. Rabson-Mendenhall syndrome
4. Lipodystrophy syndromes

C. Diseases of the exocrine pancreas—pancreatitis, pancreatectomy, neoplasia, cystic fibrosis, hemochromatosis, fibrocalculous pancreatopathy, mutations in carboxyl ester lipase

D. Endocrinopathies—acromegaly, Cushing's syndrome, glucagonoma, pheochromocytoma, hyperthyroidism, somatostatinoma, aldosteronoma

E. Drug- or chemical-induced—glucocorticoids, vacor (a rodenticide), pentamidine, nicotinic acid, diazoxide, β-adrenergic agonists, thiazides, hydantoins, asparaginase, α-interferon, protease inhibitors, antipsychotics (atypicals and others), epinephrine

F. Infections—congenital rubella, cytomegalovirus, coxsackievirus

G. Uncommon forms of immune-mediated diabetes— "stiff-person" syndrome, anti-insulin receptor antibodies

H. Other genetic syndromes sometimes associated with diabetes— Wolfram's syndrome, Down's syndrome, Klinefelter's syndrome, Turner's syndrome, Friedreich's ataxia, Huntington's chorea, Laurence-Moon-Biedl syndrome, myotonic dystrophy, porphyria, Prader-Willi syndrome

IV. Gestational diabetes mellitus (GDM)

Abbreviation: MODY, maturity-onset diabetes of the young.

Source: Adapted from American Diabetes Association, 2011.

Two features of the current classification of DM diverge from previous classifications. First, the terms *insulin-dependent diabetes mellitus* (IDDM) and *non-insulin-dependent diabetes mellitus* (NIDDM) are obsolete. Since many individuals with type 2 DM eventually require insulin treatment for control of glycemia, the use of the term NIDDM generated considerable confusion. A second difference is that age is not a criterion in the classification system. Although type 1 DM most commonly develops before the age of 30, an autoimmune beta cell destructive process can develop at any age. It is estimated that between 5 and 10% of individuals who develop DM after age 30 years have type 1 DM. Although type 2 DM more typically develops with increasing age, it is now being diagnosed more frequently in children and young adults, particularly in obese adolescents.

Other Types of DM

Other etiologies for DM include specific genetic defects in insulin secretion or action, metabolic abnormalities that impair insulin secretion, mitochondrial abnormalities, and a host of conditions that impair glucose tolerance (Table 344-1). *Maturity-onset diabetes of the young* (MODY) is a subtype of DM characterized by autosomal dominant inheritance, early onset of hyperglycemia (usually <25 years), and impairment in insulin secretion (discussed below). Mutations in the insulin receptor cause a group of rare disorders characterized by severe insulin resistance.

DM can result from pancreatic exocrine disease when the majority of pancreatic islets are destroyed. Cystic fibrosis-related DM is an important consideration in this patient population. Hormones that antagonize insulin action can also lead to DM. Thus, DM is often a feature of endocrinopathies such as acromegaly and Cushing's disease. Viral infections have been implicated in pancreatic islet destruction but are an extremely rare cause of DM. A form of acute onset of type 1 diabetes, termed *fulminant diabetes*, has been noted in Japan and may be related to viral infection of islets.

Gestational Diabetes Mellitus (GDM)

Glucose intolerance developing during pregnancy is classified as gestational diabetes. Insulin resistance is related to the metabolic changes of late pregnancy, and the increased insulin requirements may lead to IGT or diabetes. GDM occurs in ~7% (range 2–10%) of pregnancies in the United States; most women revert to normal glucose tolerance postpartum but have a substantial risk (35–60%) of developing DM in the next 10–20 years. The International Diabetes and Pregnancy Study Groups now recommends that diabetes diagnosed at the initial prenatal visit should be classified as "overt" diabetes rather than gestational diabetes.

EPIDEMIOLOGY

The worldwide prevalence of DM has risen dramatically over the past two decades, from an estimated 30 million cases in 1985 to 285 million in 2010. Based on current trends, the International Diabetes Federation projects that 438 million individuals will have diabetes by the year 2030 **(Fig. 344-2)**. Although the prevalence of both type 1 and type 2 DM is increasing worldwide, the prevalence of type 2 DM is rising much more rapidly, presumably because of increasing obesity, reduced activity levels as countries become more industrialized, and the aging of the population. In 2010, the prevalence of diabetes ranged from 11.6 to 30.9% in the 10 countries with the highest prevalence (Naurua, United Arab Emigrates, Saudi Arabia, Mauritius, Bahrain, Reunion, Kuwait, Oman, Tonga, Malaysia—in descending prevalence; Fig. 344-2). In the most recent estimate for the United States (2010), the Centers for Disease Control and Prevention (CDC) estimated that 25.8 million persons, or 8.3% of the population, had diabetes (~27% of the individuals with diabetes were undiagnosed). Approximately 1.6 million individuals (>20 years) were newly diagnosed with diabetes in 2010. DM increases with aging. In 2010, the prevalence of DM in the United Sates was estimated to be 0.2% in individuals aged <20 years and 11.3% in individuals aged >20 years. In individuals aged >65 years, the prevalence of DM was 26.9%. The prevalence is similar in men and women throughout most age ranges (11.8% and 10.8%, respectively, in individuals aged >20 years). Worldwide estimates project that in 2030 the greatest number of individuals with diabetes will be aged 45–64 years.

Figure 344-2

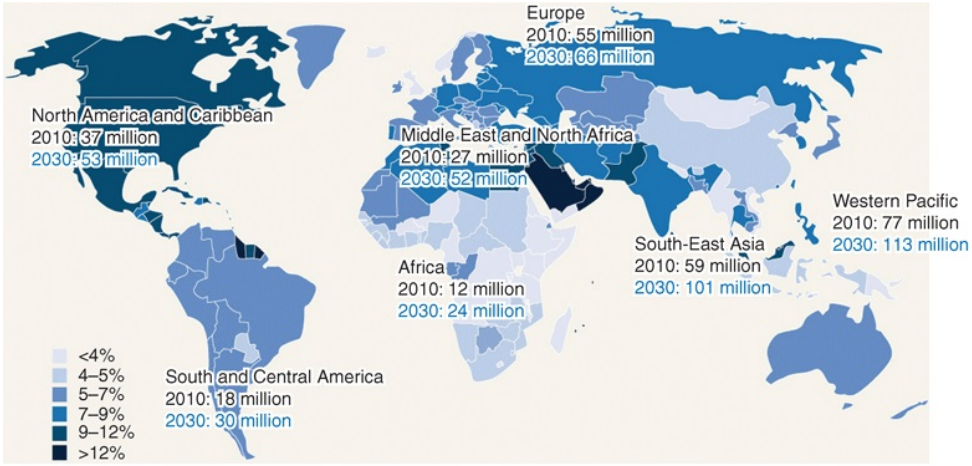

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Worldwide prevalence of diabetes mellitus. Comparative prevalence (%) of estimates of diabetes (20–79 years), 2010. *(Used with permission from IDF Diabetes Atlas, the International Diabetes Federation, 2009.)*

There is considerable geographic variation in the incidence of both type 1 and type 2 DM. Scandinavia has the highest incidence of type 1 DM (e.g., in Finland, the incidence is 57.4/100,000 per year). The Pacific Rim has a much lower rate of type 1 DM (in Japan and China, the incidence is 0.6–2.4/100,000 per year); Northern Europe and the United States have an intermediate rate (8–20/100,000 per year). Much of the increased risk of type 1 DM is believed to reflect the frequency of high-risk human leukocyte antigen (HLA) alleles among ethnic groups in different geographic locations. The prevalence of type 2 DM and its harbinger, IGT, is highest in certain Pacific islands and the Middle East and intermediate in countries such as India and the United States. This variability is likely due to genetic, behavioral, and environmental factors. DM prevalence also varies among different ethnic populations within a given country. For example, the CDC estimated that the age-adjusted prevalence of DM in the United States (age > 20 years; 2007–2009) was 7.1% in non-Hispanic whites, 7.5% in Asian Americans, 11.8% in Hispanics, and 12.6% in non-Hispanic blacks. Comparable statistics for individuals belonging to American Indian, Alaska Native, or Pacific-Islander ethnic groups are not available, but the prevalence likely exceeds the rate in non-Hispanic whites. The onset of type 2 DM occurs, on average, at an earlier age in ethnic groups other than non-Hispanic whites. In Asia, the prevalence of diabetes is increasing rapidly and the diabetes phenotype appears to be different from that in the United States and Europe —onset at a lower BMI and younger age, greater visceral adiposity, and reduced insulin secretory capacity.

Diabetes is a major cause of mortality, but several studies indicate that diabetes is likely underreported as a cause of death. In the United States, diabetes was listed as the seventh leading cause of death in 2007; a recent estimate suggested that diabetes was the fifth leading cause of death worldwide and was responsible for almost 4 million deaths in 2010 (6.8% of deaths were attributed to diabetes worldwide).

DIAGNOSIS

Glucose tolerance is classified into three broad categories: normal glucose homeostasis, diabetes mellitus, and impaired glucose homeostasis. Glucose tolerance can be assessed using the fasting plasma glucose (FPG), the response to oral glucose challenge, or the hemoglobin A1C (A1C). An FPG <5.6 mmol/L (100 mg/dL), a plasma glucose <140 mg/dL (11.1 mmol/L) following an oral glucose challenge, and an A1C <5.6% are considered to define normal glucose tolerance. The International Expert Committee with members appointed by the American Diabetes Association, the European Association for the Study of Diabetes, and the International Diabetes Federation has issued diagnostic criteria for DM **(Table 344-2)** based on the following premises: (1) the FPG, the response to an oral glucose challenge (OGTT—oral glucose tolerance test), and A1C differ among individuals, and (2) DM is defined as the level of glycemia at which diabetes-specific complications occur rather than on deviations from a population-based mean. For example, the prevalence of retinopathy in Native Americans (Pima Indian population) begins to increase at an FPG >6.4 mmol/L (116 mg/dL) **(Fig. 344-3)**.

An FPG ≥7.0 mmol/L (126 mg/dL), a glucose >11.1 mmol/L (200 mg/dL) 2 h after an oral glucose challenge, or an A1C ≥6.5% warrant the diagnosis of DM (Table 344-2). A random plasma glucose concentration ≥11.1 mmol/L (200 mg/dL) accompanied by classic symptoms of DM (polyuria, polydipsia, weight loss) also is sufficient for the diagnosis of DM (Table 344-2).

Table 344-2 Criteria for the Diagnosis of Diabetes Mellitus

• Symptoms of diabetes plus random blood glucose concentration ≥11.1 mmol/L (200 mg/dL)[a] *or*
• Fasting plasma glucose ≥7.0 mmol/L (126 mg/dL)[b] *or*
• A1C > 6.5%[c] *or*
• Two-hour plasma glucose ≥11.1 mmol/L (200 mg/dL) during an oral glucose tolerance test[d]

[a]Random is defined as without regard to time since the last meal. [b]Fasting is defined as no caloric intake for at least 8 h. [c]The test should be performed in laboratory certified according to A1C standards of the Diabetes Control and Complications Trial. [d]The test should be performed using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water, not recommended for routine clinical use. ***Note:*** In the absence of unequivocal hyperglycemia and acute metabolic decompensation, these criteria should be confirmed by repeat testing on a different day. ***Source:*** American Diabetes Association, 2011.

Figure 344-3

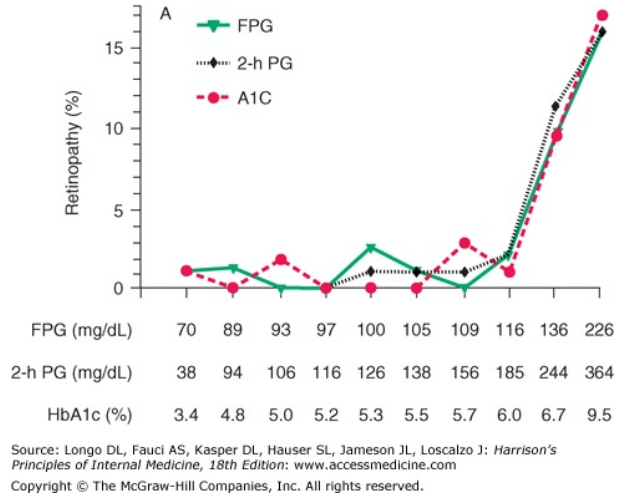

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship of diabetes-specific complication and glucose tolerance. This figure shows the incidence of retinopathy in Pima Indians as a function of the fasting plasma glucose (FPG), the 2-h plasma glucose after a 75-g oral glucose challenge (2-h PG), or the A1C. Note that the incidence of retinopathy greatly increases at a fasting plasma glucose >116 mg/dL, or a 2-h plasma glucose of 185 mg/dL, or an A1C > 6.5%. (Blood glucose values are shown in mg/dL; to convert to mmol/L, divide value by 18.) *(Copyright 2002, American Diabetes Association. From Diabetes Care 25(Suppl 1): S5–S20, 2002.)*

Abnormal glucose homeostasis (Fig. 344-1) is defined as (1) FPG = 5.6–6.9 mmol/L (100–125 mg/dL), which is defined as IFG (note that the World Health Organization uses an FPG of 6.1–6.9 mmol/L (110–125 mg/dL); (2) plasma glucose levels between 7.8 and 11 mmol/L (140 and 199 mg/dL) following an oral glucose challenge, which is termed impaired glucose tolerance (IGT); or (3) A1C of 5.7–6.4%. An A1C of 5.7–6.4%, IFG, and IGT do not identify the same individuals, but individuals in all three groups are at greater risk of progressing to type 2 diabetes and have an increased risk of cardiovascular disease. Some use the term "prediabetes," "increased risk of diabetes" (ADA), or "intermediate hyperglycemia" (WHO) for this category. The current criteria for the diagnosis of DM emphasize that the A1C or the FPG as the most reliable and convenient tests for identifying DM in asymptomatic individuals. Oral glucose tolerance testing, although still a valid means for diagnosing DM, is not often used in routine clinical care.

The diagnosis of DM has profound implications for an individual from both a medical and a financial standpoint. Thus, abnormalities on screening tests for diabetes should be repeated before making a definitive diagnosis of DM, unless acute metabolic derangements or a markedly elevated plasma glucose are present (Table 344-2). These criteria also allow for the diagnosis of DM to be withdrawn in situations when the glucose intolerance reverts to normal.

Screening

Widespread use of the FPG or the A1C as a screening test for type 2 DM is recommended because (1) a large number of individuals who meet the current criteria for DM are asymptomatic and unaware that they have the disorder, (2) epidemiologic studies suggest that type 2 DM may be present for up to a decade before diagnosis, (3) some individuals with type 2 DM have one or more diabetes-specific complications at the time of their diagnosis, and (4) treatment of type 2 DM may favorably alter the natural history of DM. The ADA recommends screening all individuals >45 years every 3 years and screening individuals at an earlier age if they are overweight [body mass index (BMI) >25 kg/m^2] and have one additional risk factor for diabetes **(Table 344-3)**. In contrast to type 2 DM, a long asymptomatic period of hyperglycemia is rare prior to the diagnosis of type 1 DM. A number of immunologic markers for type 1 DM are becoming available (discussed below), but their routine use is discouraged pending the identification of clinically beneficial interventions for individuals at high risk for developing type 1 DM.

Table 344-3 Risk Factors for Type 2 Diabetes Mellitus
Family history of diabetes (i.e., parent or sibling with type 2 diabetes)
Obesity (BMI ≥25 kg/m^2)
Physical inactivity
Race/ethnicity (e.g., African American, Latino, Native American, Asian American, Pacific Islander)
Previously identified with IFG, IGT, or an A1C of 5.7–6.4%
History of GDM or delivery of baby >4 kg (9 lb)
Hypertension (blood pressure ≥140/90 mmHg)
HDL cholesterol level <35 mg/dL (0.90 mmol/L) and/or a triglyceride level >250 mg/dL (2.82 mmol/L)
Polycystic ovary syndrome or acanthosis nigricans
History of cardiovascular disease

Abbreviations: BMI, body mass index; GDM, gestational diabetes mellitus; HDL, high-density lipoprotein; IFG, impaired fasting glucose; IGT, impaired glucose tolerance. ***Source:*** Adapted from American Diabetes Association, 2011.

INSULIN BIOSYNTHESIS, SECRETION, AND ACTION

Biosynthesis

Insulin is produced in the beta cells of the pancreatic islets. It is initially synthesized as a single-chain 86-amino-acid precursor polypeptide, preproinsulin. Subsequent proteolytic processing removes the amino-terminal signal peptide, giving rise to proinsulin. Proinsulin is structurally related to insulin-like growth factors I and II, which bind weakly to the insulin receptor. Cleavage of an internal 31-residue fragment from proinsulin generates the C peptide and the A (21 amino acids) and B (30 amino acids) chains of insulin, which are connected by disulfide bonds. The mature insulin molecule and C peptide are stored together and co-secreted from secretory granules in the beta cells. Because C peptide is cleared more slowly than insulin, it is a useful marker of insulin secretion and allows discrimination of endogenous and exogenous sources of insulin in the evaluation of hypoglycemia (Chaps. 345 and 350). Pancreatic beta cells co-secrete islet amyloid polypeptide (IAPP) or amylin, a 37-amino-acid peptide, along with insulin. The role of IAPP in normal physiology is incompletely defined, but it is the major component of

the amyloid fibrils found in the islets of patients with type 2 diabetes, and an analogue is sometimes used in treating type 1 and type 2 DM. Human insulin is produced by recombinant DNA technology; structural alterations at one or more amino acid residues modify its physical and pharmacologic characteristics (see below).

Secretion

Glucose is the key regulator of insulin secretion by the pancreatic beta cell, although amino acids, ketones, various nutrients, gastrointestinal peptides, and neurotransmitters also influence insulin secretion. Glucose levels >3.9 mmol/L (70 mg/dL) stimulate insulin synthesis, primarily by enhancing protein translation and processing. Glucose stimulation of insulin secretion begins with its transport into the beta cell by a facilitative glucose transporter **(Fig. 344-4)**. Glucose phosphorylation by glucokinase is the rate-limiting step that controls glucose-regulated insulin secretion. Further metabolism of glucose-6-phosphate via glycolysis generates ATP, which inhibits the activity of an ATP-sensitive K^+ channel. This channel consists of two separate proteins: one is the binding site for certain oral hypoglycemics (e.g., sulfonyl-ureas, meglitinides); the other is an inwardly rectifying K^+ channel protein (Kir6.2). Inhibition of this K^+ channel induces beta cell membrane depolarization, which opens voltage-dependent calcium channels (leading to an influx of calcium), and stimulates insulin secretion. Insulin secretory profiles reveal a pulsatile pattern of hormone release, with small secretory bursts occurring about every 10 min, superimposed upon greater amplitude oscillations of about 80–150 min. Incretins are released from neuroendocrine cells of the gastrointestinal tract following food ingestion and amplify glucose-stimulated insulin secretion and suppress glucagon secretion. Glucagon-like peptide 1 (GLP-1), the most potent incretin, is released from L cells in the small intestine and stimulates insulin secretion only when the blood glucose is above the fasting level. Incretin analogues, are used to enhance endogenous insulin secretion (see below).

Figure 344-4

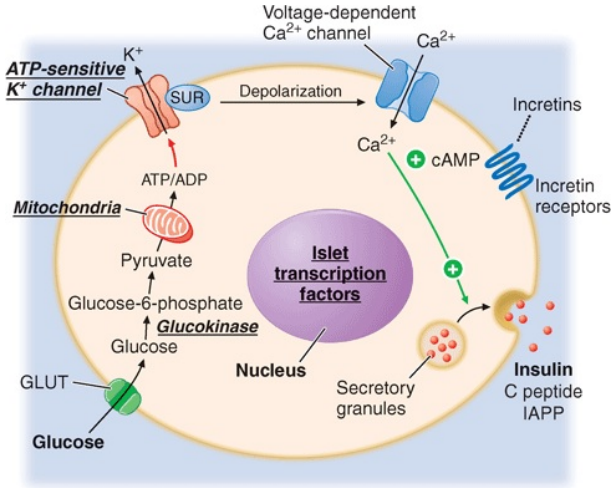

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanisms of glucose-stimulated insulin secretion and abnormalities in diabetes. Glucose and other nutrients regulate insulin secretion by the pancreatic beta cell. Glucose is transported by a glucose transporter (GLUT1 in humans, GLUT2 in rodents); subsequent glucose metabolism by the beta cell alters ion channel activity, leading to insulin secretion. The SUR receptor is the binding site for some drugs that act as insulin secretagogues. Mutations in the events or proteins underlined are a cause of maturity-onset diabetes of the young (MODY) or other forms of diabetes. SUR, sulfonylurea receptor; ATP, adenosine triphosphate; ADP, adenosine diphosphate, cAMP, cyclic adenosine monophosphate. IAPP, islet amyloid polypeptide or amylin.

Action

Once insulin is secreted into the portal venous system, ~50% is removed and degraded by the liver. Unextracted insulin enters the systemic circulation where it binds to receptors in target sites. Insulin binding to its receptor stimulates intrinsic tyrosine kinase activity, leading to receptor autophosphorylation and the recruitment of intracellular signaling molecules, such as insulin receptor substrates (IRS) **(Fig. 344-5)**. IRS and other adaptor proteins initiate a complex cascade of phosphorylation and dephosphorylation reactions, resulting in the widespread metabolic and mitogenic effects of insulin. As an example, activation of the phosphatidylinositol-3′-kinase (PI-3-kinase) pathway stimulates translocation of a facilitative glucose transporter (e.g., GLUT4) to the cell surface, an event that is crucial for glucose uptake by skeletal muscle and fat. Activation of other insulin receptor signaling pathways induces glycogen synthesis, protein synthesis, lipogenesis, and regulation of various genes in insulin-responsive cells.

Figure 344-5

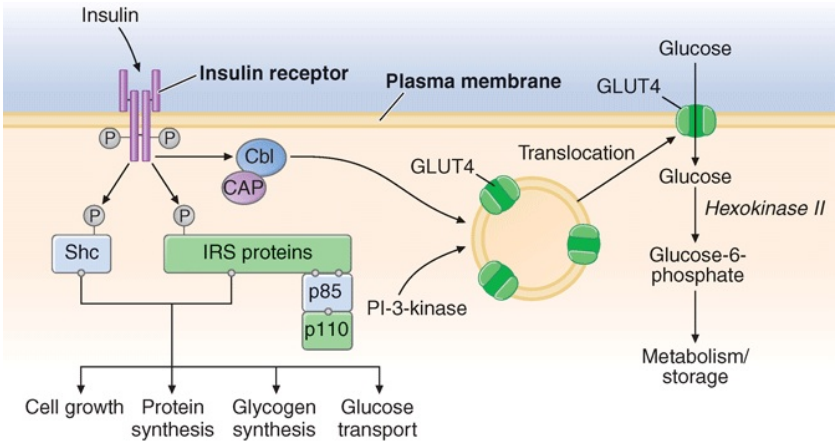

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Insulin signal transduction pathway in skeletal muscle. The insulin receptor has intrinsic tyrosine kinase activity and interacts with insulin receptor substrates (IRS and Shc) proteins. A number of "docking" proteins bind to these cellular proteins and initiate the metabolic actions of insulin [GrB-2, SOS, SHP-2, p110, and phosphatidylinositol-3′-kinase (PI-3-kinase)]. Insulin increases glucose transport through PI-3-kinase and the Cbl pathway, which promotes the translocation of intracellular vesicles containing GLUT4 glucose transporter to the plasma membrane.

Glucose homeostasis reflects a balance between hepatic glucose production and peripheral glucose uptake and utilization. Insulin is the most important regulator of this metabolic equilibrium, but neural input, metabolic signals, and other hormones (e.g., glucagon) result in integrated control of glucose supply and utilization (Chap. 345; see Fig. 345-1). In the fasting state, low insulin levels increase glucose production by promoting hepatic gluconeogenesis and glycogenolysis and reduce glucose uptake in insulin-sensitive tissues (skeletal muscle and fat), thereby promoting mobilization of stored precursors such as amino acids and free fatty acids (lipolysis). Glucagon, secreted by pancreatic alpha cells when blood glucose or insulin levels are low, stimulates glycogenolysis and gluconeogenesis by the liver and renal medulla. Postprandially, the glucose load elicits a rise in insulin and fall in glucagon, leading to a reversal of these processes. Insulin, an anabolic hormone, promotes the storage of carbohydrate and fat and protein synthesis. The major portion of postprandial glucose is utilized by skeletal muscle, an effect of insulin-stimulated glucose uptake. Other tissues, most notably the brain, utilize glucose in an insulin-independent fashion.

PATHOGENESIS

Type 1 DM

Type 1 DM is the result of interactions of genetic, environmental, and immunologic factors that ultimately lead to the destruction of the pancreatic beta cells and insulin deficiency. Type 1 DM results from autoimmune beta cell destruction, and most, but not all, individuals have evidence of islet-directed autoimmunity. Some individuals who have the clinical phenotype of type 1 DM lack immunologic markers indicative of an autoimmune process involving the beta cells and the genetic markers of type 1 diabetes. These individuals are thought to develop insulin deficiency by unknown, nonimmune mechanisms and are ketosis prone; many are African American or Asian in heritage. The temporal development of type 1 DM is shown schematically as a function of beta cell mass in **Fig. 344-6**. Individuals with a genetic susceptibility have normal beta cell mass at birth but begin to lose beta cells secondary to autoimmune destruction that occurs over months to years. This autoimmune process is thought to be triggered by an infectious or environmental stimulus and to be sustained by a beta cell–specific molecule. In the majority, immunologic markers appear after the triggering event but before diabetes becomes clinically overt. Beta cell mass then begins to decrease, and insulin secretion progressively declines, although normal glucose tolerance is maintained. The rate of decline in beta cell mass varies widely among individuals, with some patients progressing rapidly to clinical diabetes and others evolving more slowly. Features of diabetes do not become evident until a majority of beta cells are destroyed (70–80%). At this point, residual functional beta cells exist but are insufficient in number to maintain glucose tolerance. The events that trigger the transition from glucose intolerance to frank diabetes are often associated with increased insulin requirements, as might occur during infections or puberty. After the initial clinical presentation of type 1 DM, a "honeymoon" phase may ensue during which time glycemic control is achieved with modest doses of insulin or, rarely, insulin is not needed. However, this fleeting phase of endogenous insulin production from residual beta cells disappears as the autoimmune process destroys remaining beta cells, and the individual becomes insulin deficient. Some individuals with long-standing type 1 diabetes produce a small amount of insulin (as reflected by C-peptide production) and some individuals have insulin-positive cells in the pancreas at autopsy.

Figure 344-6

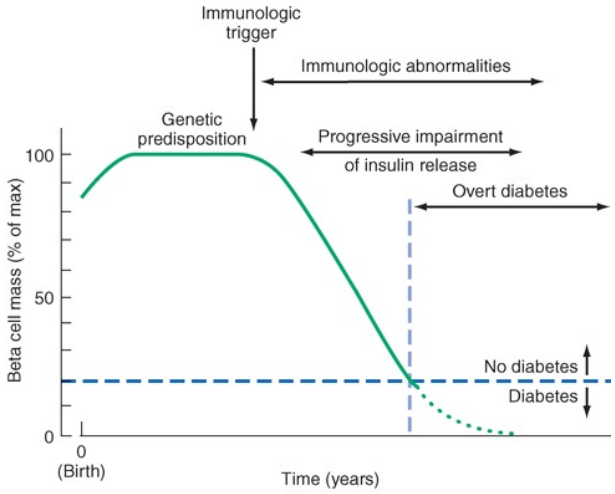

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Temporal model for development of type 1 diabetes. Individuals with a genetic predisposition are exposed to an immunologic trigger that initiates an autoimmune process, resulting in a gradual decline in beta cell mass. The downward slope of the beta cell mass varies among individuals and may not be continuous. This progressive impairment in insulin release results in diabetes when ~80% of the beta cell mass is destroyed. A "honeymoon" phase may be seen in the first 1 or 2 years after the onset of diabetes and is associated with reduced insulin requirements. *[Adapted from Medical Management of Type 1 Diabetes, 3rd ed, JS Skyler (ed). American Diabetes Association, Alexandria, VA, 1998.]*

GENETIC CONSIDERATIONS

Susceptibility to type 1 DM involves multiple genes. The concordance of type 1 DM in identical twins ranges between 40 and 60%, indicating that additional modifying factors are likely involved in determining whether diabetes develops. The major susceptibility gene for type 1 DM is located in the HLA region on chromosome 6. Polymorphisms in the HLA complex account for 40–50% of the genetic risk of developing type 1 DM. This region contains genes that encode the class II major histocompatibility complex (MHC) molecules, which present antigen to helper T cells and thus are involved in initiating the immune response (Chap. 315). The ability of class II MHC molecules to present antigen is dependent on the amino acid composition of their antigen-binding sites. Amino acid substitutions may influence the specificity of the immune response by altering the binding affinity of different antigens for class II molecules.

Most individuals with type 1 DM have the HLA DR3 and/or DR4 haplotype. Refinements in genotyping of HLA loci have shown that the haplotypes DQA1*0301, DQB1*0302, and DQB1*0201 are most strongly associated with type 1 DM. These haplotypes are present in 40% of children with type 1 DM as compared to 2% of the normal U.S. population. However, most individuals with predisposing haplotypes do not develop diabetes.

In addition to MHC class II associations, genome association studies have identified at least 20 different genetic loci that contribute susceptibility to type 1 DM (polymorphisms in the promoter region of the insulin gene, the CTLA-4 gene, interleukin-2 receptor, *CTLA4*, and PTPN22, etc.). Genes that confer protection against the development of the disease also exist. The haplotype DQA1*0102, DQB1*0602 is extremely rare in individuals with type 1 DM (<1%) and appears to provide protection from type 1 DM.

Although the risk of developing type 1 DM is increased tenfold in relatives of individuals with the disease, the risk is relatively low: 3–4% if the parent has type 1 diabetes and 5–15% in a sibling (depending on which HLA haplotypes are shared). Hence, most individuals with type 1 DM do not have a first-degree relative with this disorder.

PATHOPHYSIOLOGY

Although other islet cell types [alpha cells (glucagon-producing), delta cells (somatostatin-producing), or PP cells (pancreatic polypeptide-producing)] are functionally and embryologically similar to beta cells and express most of the same proteins as beta cells, they are spared from the autoimmune destruction. Pathologically, the pancreatic islets are infiltrated with lymphocytes (a process termed *insulitis*). After all beta cells are destroyed, the inflammatory process abates, the islets become atrophic, and most immunologic markers disappear. Studies of the autoimmune process in humans and in animal models of type 1 DM (NOD mouse and BB rat) have identified the following abnormalities in the humoral and cellular arms of the immune system: (1) islet cell autoantibodies; (2) activated lymphocytes in the islets, peripancreatic lymph nodes, and systemic circulation; (3) T lymphocytes that proliferate when stimulated with islet proteins; and (4) release of cytokines within the insulitis. Beta cells seem to be particularly susceptible to the toxic effect of some cytokines [tumor necrosis factor α (TNF-α), interferon γ, and interleukin 1 (IL-1)]. The precise mechanisms of beta cell death are not known but may involve formation of nitric oxide metabolites, apoptosis, and direct CD8+ T cell cytotoxicity. The islet destruction is mediated by T lymphocytes rather than islet autoantibodies, as these antibodies do not generally react with the cell surface of islet cells and are not capable of transferring DM to animals. Suppression of the autoimmune process at the time of diagnosis of diabetes slows the decline in beta cell destruction, but the safety of such interventions is unknown.

Pancreatic islet molecules targeted by the autoimmune process include insulin, glutamic acid decarboxylase (GAD, the biosynthetic enzyme for the neurotransmitter GABA), ICA-512/IA-2 (homology with tyrosine phosphatases), and a beta cell –specific zinc transporter (ZnT-8). Most of the autoantigens are not beta cell–specific, which raises the question of how the beta cells are selectively destroyed. Current theories favor initiation of an autoimmune process directed at one beta cell molecule, which then spreads to other islet molecules as the immune process destroys beta cells and creates a series of secondary autoantigens. The beta cells of individuals who develop type 1 DM do not differ from beta cells of normal individuals, since islets transplanted from a genetically identical twin are destroyed by a recurrence of the autoimmune process of type 1 DM.

IMMUNOLOGIC MARKERS

Islet cell autoantibodies (ICAs) are a composite of several different antibodies directed at pancreatic islet molecules such as GAD, insulin, IA-2/ICA-512, and ZnT-8, and serve as a marker of the autoimmune process of type 1 DM. Assays for autoantibodies to GAD-65 are commercially available. Testing for ICAs can be useful in classifying the type of DM as type 1 and in identifying nondiabetic individuals at risk for developing type 1 DM. ICAs are present in the majority of individuals (>85%) diagnosed with new-onset type 1 DM, in a significant minority of individuals with newly diagnosed type 2 DM (5–10%), and

occasionally in individuals with GDM (<5%). ICAs are present in 3–4% of first-degree relatives of individuals with type 1 DM. In combination with impaired insulin secretion after IV glucose tolerance testing, they predict a >50% risk of developing type 1 DM within 5 years. At present, the measurement of ICAs in nondiabetic individuals is a research tool because no treatments have been approved to prevent the occurrence or progression to type 1 DM. Clinical trials are testing interventions to slow the autoimmune beta cell destruction.

ENVIRONMENTAL FACTORS

Numerous environmental events have been proposed to trigger the autoimmune process in genetically susceptible individuals; however, none have been conclusively linked to diabetes. Identification of an environmental trigger has been difficult because the event may precede the onset of DM by several years (Fig. 344-6). Putative environmental triggers include viruses (coxsackie, rubella, enteroviruses most prominently), bovine milk proteins, and nitrosourea compounds.

PREVENTION OF TYPE 1 DM

A number of interventions have successfully delayed or prevented diabetes in animal models. Some interventions have targeted the immune system directly (immunosuppression, selective T cell subset deletion, induction of immunologic tolerance to islet proteins), whereas others have prevented islet cell death by blocking cytotoxic cytokines or increasing islet resistance to the destructive process. Though results in animal models are promising, these interventions have not been successful in preventing type 1 DM in humans. The Diabetes Prevention Trial—type 1 concluded that administering insulin (IV or PO) to individuals at high risk for developing type 1 DM did not prevent type 1 DM.

In patients with new-onset type 1 diabetes, treatment with anti-CD3 monoclonal antibodies, a GADvaccine, and anti-B lymphocyte monoclonal antibody have been shown to slow the decline in C-peptide levels. This is an area of active clinical investigation.

Type 2 DM

Insulin resistance and abnormal insulin secretion are central to the development of type 2 DM. Although the primary defect is controversial, most studies support the view that insulin resistance precedes an insulin secretory defect but that diabetes develops only when insulin secretion becomes inadequate. Type 2 DM likely encompasses a range of disorders with common phenotype of hyperglycemia. Most of our current understanding (and the discussion below) of the pathophysiology and genetics is based on studies of individuals of European descent. It is becoming increasing apparent that DM in other ethnic groups (Asian, African, and Latin American) has a different, but yet undefined, pathophysiology. In these groups, DM that is ketosisprone (often obese) or ketosis-resistant (often lean) is commonly seen.

Genetic Considerations

Type 2 DM has a strong genetic component. The concordance of type 2 DM in identical twins is between 70 and 90%. Individuals with a parent with type 2 DM have an increased risk of diabetes; if both parents have type 2 DM, the risk approaches 40%. Insulin resistance, as demonstrated by reduced glucose utilization in skeletal muscle, is present in many nondiabetic, first-degree relatives of individuals with type 2 DM. The disease is polygenic and multifactorial, since in addition to genetic susceptibility, environmental factors (such as obesity, nutrition, and physical activity) modulate the phenotype. The genes that predispose to type 2 DM are incompletely identified, but recent genome-wide association studies have identified a large number of genes that convey a relatively small risk for type 2 DM (>20 genes, each with a relative risk of 1.06–1.5). Most prominent is a variant of the transcription factor 7–like 2 gene that has been associated with type 2 diabetes in several populations and with impaired glucose tolerance in one population at high risk for diabetes. Genetic polymorphisms associated with type 2 diabetes have also been found in the genes encoding the peroxisome proliferators–activated receptor-γ, inward rectifying potassium channel, zinc transporter, IRS, and calpain 10. The mechanisms by which these genetic loci increase the susceptibility to type 2 diabetes are not clear, but most are predicted to alter islet function or development or insulin secretion. While the genetic susceptibility to type 2 diabetes is under active investigation (estimation that <10% of genetic risk is determined by loci identified thus far), it is currently not possible to use a combination of known genetic loci to predict type 2 diabetes.

PATHOPHYSIOLOGY

Type 2 DM is characterized by impaired insulin secretion, insulin resistance, excessive hepatic glucose production, and abnormal fat metabolism. Obesity, particularly visceral or central (as evidenced by the hip-waist ratio), is very common in type 2 DM (80% or more are obese). In the early stages of the disorder, glucose tolerance remains near-normal, despite insulin resistance, because the pancreatic beta cells compensate by increasing insulin output **(Fig. 344-7)**. As insulin resistance and compensatory hyperinsulinemia progress, the pancreatic islets in certain individuals are unable to sustain the hyperinsulinemic state. IGT, characterized by elevations in postprandial glucose, then develops. A further decline in insulin secretion and an increase in hepatic glucose production lead to overt diabetes with fasting hyperglycemia. Ultimately, beta cell failure ensues.

Figure 344-7

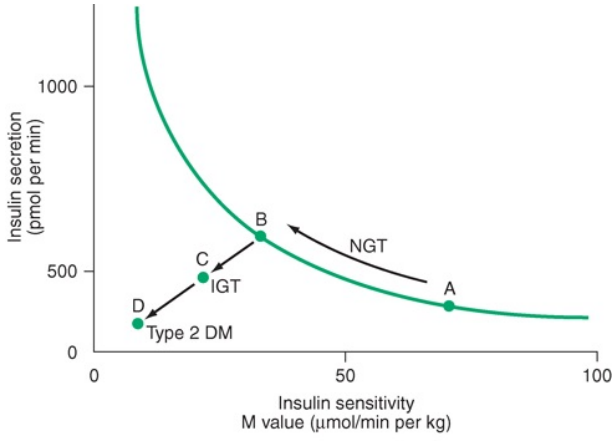

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Metabolic changes during the development of type 2 diabetes mellitus (DM). Insulin secretion and insulin sensitivity are related, and as an individual becomes more insulin resistant (by moving from point A to point B), insulin secretion increases. A failure to compensate by increasing the insulin secretion results initially in impaired glucose tolerance (IGT; point C) and

failure to compensate by increasing the insulin secretion results in impaired glucose tolerance (point C) and ultimately in type 2 DM (point D). *(Adapted from SE Kahn: J Clin Endocrinol Metab 86:4047, 2001; RN Bergman, M Ader: Trends Endocrinol Metab 11:351, 2000.)*

METABOLIC ABNORMALITIES

Abnormal Muscle and Fat Metabolism

Insulin resistance, the decreased ability of insulin to act effectively on target tissues (especially muscle, liver, and fat), is a prominent feature of type 2 DM and results from a combination of genetic susceptibility and obesity. Insulin resistance is relative, however, since supranormal levels of circulating insulin will normalize the plasma glucose. Insulin dose-response curves exhibit a rightward shift, indicating reduced sensitivity, and a reduced maximal response, indicating an overall decrease in maximum glucose utilization (30–60% lower than in normal individuals). Insulin resistance impairs glucose utilization by insulin-sensitive tissues and increases hepatic glucose output; both effects contribute to the hyperglycemia. Increased hepatic glucose output predominantly accounts for increased FPG levels, whereas decreased peripheral glucose usage results in postprandial hyperglycemia. In skeletal muscle, there is a greater impairment in nonoxidative glucose usage (glycogen formation) than in oxidative glucose metabolism through glycolysis. Glucose metabolism in insulin-independent tissues is not altered in type 2 DM.

The precise molecular mechanism leading to insulin resistance in type 2 DM has not been elucidated. Insulin receptor levels and tyrosine kinase activity in skeletal muscle are reduced, but these alterations are most likely secondary to hyperinsulinemia and are not a primary defect. Therefore, "postreceptor" defects in insulin-regulated phosphorylation/dephosphorylation appear to play the predominant role in insulin resistance (Fig. 344-5). For example, a PI-3-kinase signaling defect might reduce translocation of GLUT4 to the plasma membrane. Other abnormalities include the accumulation of lipid within skeletal myocytes, which may impair mitochondrial oxidative phosphorylation and reduce insulin-stimulated mitochondrial ATP production. Impaired fatty acid oxidation and lipid accumulation within skeletal myocytes also may generate reactive oxygen species such as lipid peroxides. Of note, not all insulin signal transduction pathways are resistant to the effects of insulin (e.g., those controlling cell growth and differentiation using the mitogenic-activated protein kinase pathway). Consequently, hyperinsulinemia may increase the insulin action through these pathways, potentially accelerating diabetes-related conditions such as atherosclerosis.

The obesity accompanying type 2 DM, particularly in a central or visceral location, is thought to be part of the pathogenic process. The increased adipocyte mass leads to increased levels of circulating free fatty acids and other fat cell products (Chap. 77). For example, adipocytes secrete a number of biologic products (nonesterified free fatty acids, retinol-binding protein 4, leptin, TNF-α, resistin, and adiponectin). In addition to regulating body weight, appetite, and energy expenditure, adipokines also modulate insulin sensitivity. The increased production of free fatty acids and some adipokines may cause insulin resistance in skeletal muscle and liver. For example, free fatty acids impair glucose utilization in skeletal muscle, promote glucose production by the liver, and impair beta cell function. In contrast, the production by adipocytes of adiponectin, an insulin-sensitizing peptide, is reduced in obesity, and this may contribute to hepatic insulin resistance. Adipocyte products and adipokines also produce an inflammatory state and may explain why markers of inflammation such as IL-6 and C-reactive protein are often elevated in type 2 DM. In addition, inflammatory cells have been found infiltrating adipose tissue. Inhibition of inflammatory signaling pathways such as the nuclear factor κB (NF-κB) pathway appears to reduce insulin resistance and improve hyper-glycemia in animal models.

Impaired Insulin Secretion

Insulin secretion and sensitivity are interrelated (Fig. 344-7). In type 2 DM, insulin secretion initially increases in response to insulin resistance to maintain normal glucose tolerance. Initially, the insulin secretory defect is mild and selectively involves glucose-stimulated insulin secretion. The response to other nonglucose secretagogues, such as arginine, is preserved. Abnormalities in proinsulin processing is reflected by increased secretion of proinsulin in type 2 diabetes. Eventually, the insulin secretory defect progresses to a state of inadequate insulin secretion.

The reason(s) for the decline in insulin secretory capacity in type 2 DM is unclear. The assumption is that a second genetic defect—superimposed upon insulin resistance—leads to beta cell failure. Beta cell mass is decreased by approximately 50% in individuals with long-standing type 2 diabetes. Islet amyloid polypeptide or amylin is co-secreted by the beta cell and forms the amyloid fibrillar deposit found in the islets of individuals with long-standing type 2 DM. Whether such islet amyloid deposits are a primary or secondary event is not known. The metabolic environment of diabetes may also negatively impact islet function. For example, chronic hyperglycemia paradoxically impairs islet function ("glucose toxicity") and leads to a worsening of hyperglycemia. Improvement in glycemic control is often associated with improved islet function. In addition, elevation of free fatty acid levels ("lipotoxicity") and dietary fat may also worsen islet function.

Increased Hepatic Glucose and Lipid Production

In type 2 DM, insulin resistance in the liver reflects the failure of hyperinsulinemia to suppress gluconeogenesis, which results in fasting hyperglycemia and decreased glycogen storage by the liver in the postprandial state. Increased hepatic glucose production occurs early in the course of diabetes, though likely after the onset of insulin secretory abnormalities and insulin resistance in skeletal muscle. As a result of insulin resistance in adipose tissue, lipolysis and free fatty acid flux from adipocytes are increased, leading to increased lipid [very low density lipoprotein (VLDL) and triglyceride] synthesis in hepatocytes. This lipid storage or steatosis in the liver may lead to nonalcoholic fatty liver disease (Chap. 309) and abnormal liver function tests. This is also responsible for the dyslipidemia found in type 2 DM [elevated triglycerides, reduced high-density lipoprotein (HDL), and increased small dense low-density lipoprotein (LDL) particles].

INSULIN RESISTANCE SYNDROMES

The insulin resistance condition comprises a spectrum of disorders, with hyperglycemia representing one of the most readily diagnosed features. The *metabolic syndrome*, the *insulin resistance syndrome*, or *syndrome X* are terms used to describe a constellation of metabolic derangements that includes insulin resistance, hypertension, dyslipidemia (decreased HDL and elevated triglycerides), central or visceral obesity, type 2 diabetes or IGT/IFG, and accelerated cardiovascular disease. This syndrome is discussed in Chap. 242.

A number of relatively rare forms of severe insulin resistance include features of type 2 DM or IGT (Table 344-1). Mutations in the insulin receptor that interfere with binding or signal transduction are a rare cause of insulin resistance. Acanthosis nigricans and signs of hyperandrogenism (hirsutism, acne, and oligomenorrhea in women) are also common physical features. Two distinct syndromes of severe insulin resistance have been described in adults: (1) type A, which affects young women and is characterized by severe hyperinsulinemia, obesity, and features of hyperandrogenism; and (2) type B, which affects middle-aged women and is characterized by severe hyperinsulinemia, features of hyperandrogenism, and autoimmune disorders. Individuals with the type A insulin resistance syndrome have an undefined defect in the insulin-signaling pathway; individuals with the type B insulin resistance syndrome have autoantibodies directed at the insulin receptor. These receptor autoantibodies may block insulin binding or may stimulate the insulin receptor, leading to intermittent hypoglycemia.

Polycystic ovary syndrome (PCOS) is a common disorder that affects premenopausal women and is characterized by chronic anovulation and hyperandrogenism (Chap. 347). Insulin resistance is seen in a significant subset of women with PCOS, and the disorder substantially increases the risk for type 2 DM, independent of the effects of obesity.

PREVENTION

Type 2 DM is preceded by a period of IGT or IFG, and a number of lifestyle modifications and pharmacologic agents prevent or delay the onset of DM. The Diabetes Prevention Program (DPP) demonstrated that intensive changes in lifestyle (diet and exercise for 30 min/d five times/week) in individuals with IGT prevented or delayed the development of type 2 DM by 58% compared to placebo. This effect was seen in individuals regardless of age, sex, or ethnic group. In the same study, metformin prevented or delayed diabetes by 31% compared to placebo. The lifestyle intervention group lost 5–7% of their body weight during the 3 years of the study. Studies in Finnish and Chinese populations noted similar efficacy of diet and exercise in preventing or delaying type 2 DM; α-glucosidase inhibitors, metformin, thiazolidinediones, and orlistat prevent or delay type 2 DM but are not approved for this purpose. Individuals with a strong family history of type 2 DM and individuals with IFG or IGT should be strongly encouraged to maintain a normal BMI and engage in regular physical activity. Pharmacologic therapy for individuals with prediabetes is currently controversial because its cost-effectiveness and safety profile are not known. The ADA has suggested that metformin be considered in individuals with both IFG and IGT who are at very high risk for progression to diabetes (age <60 years, BMI ≥35 kg/m^2, family history of diabetes in first-degree relative, elevated triglycerides, reduced HDL, hypertension, or A1C >6.0%). Individuals with IFG, IGT, or an A1C of 5.7–6.4% should be monitored annually to determine if diagnostic criteria for diabetes are present.

GENETICALLY DEFINED, MONOGENIC

Forms of Diabetes Mellitus

Several monogenic forms of DM have been identified. Six different variants of MODY, caused by mutations in genes encoding islet-enriched transcription factors or glucokinase (Fig. 344-4; Table 344-1), are transmitted as autosomal dominant disorders. MODY 1, MODY 3, and MODY 5 are caused by mutations in the hepatocyte nuclear transcription factor (HNF) 4α, HNF-1α, and HNF-1β, respectively. As their names imply, these transcription factors are expressed in the liver but also in other tissues, including the pancreatic islets and kidney. These factors most likely affect islet development or the expression of genes important in glucose-stimulated insulin secretion or the maintenance of beta cell mass. For example, individuals with an HNF-1α mutation (MODY 3) have a progressive decline in glycemic control but may respond to sulfonylureas. In fact, some of these patients were initially thought to have type 1 DM but were later shown to respond to a sulfonylurea, and insulin was discontinued. Individuals with a HNF-1β mutation have progressive impairment of insulin secretion, hepatic insulin resistance, and require insulin treatment (minimal response to sulfonylureas). These individuals often have other abnormalities such as renal cysts, mild pancreatic exocrine insufficiency, and abnormal liver function tests. Individuals with MODY 2, the result of mutations in the glucokinase gene, have mild-to-moderate, stable hyperglycemia that does not respond to oral hypoglycemic agents. Glucokinase catalyzes the formation of glucose-6-phosphate from glucose, a reaction that is important for glucose sensing by the beta cells and for glucose utilization by the liver. As a result of glucokinase mutations, higher glucose levels are required to elicit insulin secretory responses, thus altering the set point for insulin secretion. MODY 4 is a rare variant caused by mutations in the insulin promoter factor (IPF) 1, which is a transcription factor that regulates pancreatic development and insulin gene transcription. Homozygous inactivating mutations cause pancreatic agenesis, whereas heterozygous mutations may result in DM. Studies of populations with type 2 DM suggest that mutations in MODY-associated genes are an uncommon (<5%) cause of type 2 DM.

Transient or permanent neonatal diabetes (onset <6 months of age) occurs. Permanent neonatal diabetes may be caused by several genetic mutations and usually requires treatment with insulin. Mutations in the ATP-sensitive potassium channel subunits (Kir6.2 and ABCC8) and the insulin gene (interfere with proinsulin folding and processing) (Fig. 344-4) are the major causes of permanent neonatal diabetes. Although these activating mutations in the ATP-sensitive potassium channel subunits impair glucose-stimulated insulin secretion, these individuals may respond to sulfonylureas and be treated with these agents. These mutations are associated with a spectrum of neurologic dysfunction. Homozygous glucokinase mutations cause a severe form of neonatal diabetes.

ACUTE COMPLICATIONS OF DM

Diabetic ketoacidosis (DKA) and hyperglycemic hyperosmolar state (HHS) are acute complications of diabetes. DKA was formerly considered a hallmark of type 1 DM, but also occurs in individuals who lack immunologic features of type 1 DM and who can sometimes subsequently be treated with oral glucose-lowering agents (these obese individuals with type 2 DM are often of Hispanic or African-American descent. The initial management of DKA is similar. HHS is primarily seen in individuals with type 2 DM. Both disorders are associated with absolute or relative insulin deficiency, volume depletion, and acid-base abnormalities. DKA and HHS exist along a continuum of hyperglycemia, with or without ketosis. The metabolic similarities and differences in DKA and HHS are highlighted in **Table 344-4**. Both disorders are associated with potentially serious complications if not promptly diagnosed and treated.

Table 344-4 Laboratory Values in Diabetic Ketoacidosis (DKA) and Hyperglycemic Hyperosmolar State (HHS) (Representative Ranges at Presentation)

	DKA	HHS
Glucose,[a] mmol/L (mg/dL)	13.9–33.3 (250–600)	33.3–66.6 (600–1200)
Sodium, meq/L	125–135	135–145
Potassium[a,b]	Normal to ↑	Normal
Magnesium[a]	Normal	Normal
Chloride[a]	Normal	Normal
Phosphate[a,b]	Normal	Normal
Creatinine	Slightly ↑	Moderately ↑
Osmolality (mOsm/mL)	300–320	330–380
Plasma ketones[a]	++++	+/–
Serum bicarbonate,[a] meq/L	<15 meq/L	Normal to slightly ↓
Arterial pH	6.8–7.3	>7.3

Arterial PCO_2,[a] mmHg	20–30	Normal
Anion gap[a] $[Na – (Cl + HCO_3)]$	↑	Normal to slightly ↑

[a]Large changes occur during treatment of DKA.

[b]Although plasma levels may be normal or high at presentation, total-body stores are usually depleted.

Diabetic Ketoacidosis

CLINICAL FEATURES

The symptoms and physical signs of DKA are listed in **Table 344-5** and usually develop over 24 h. DKA may be the initial symptom complex that leads to a diagnosis of type 1 DM, but more frequently it occurs in individuals with established diabetes. Nausea and vomiting are often prominent, and their presence in an individual with diabetes warrants laboratory evaluation for DKA. Abdominal pain may be severe and can resemble acute pancreatitis or ruptured viscus. Hyperglycemia leads to glucosuria, volume depletion, and tachycardia. Hypotension can occur because of volume depletion in combination with peripheral vasodilatation. Kussmaul respirations and a fruity odor on the patient's breath (secondary to metabolic acidosis and increased acetone) are classic signs of the disorder. Lethargy and central nervous system depression may evolve into coma with severe DKA but should also prompt evaluation for other reasons for altered mental status (infection, hypoxemia, etc.). Cerebral edema, an extremely serious complication of DKA, is seen most frequently in children. Signs of infection, which may precipitate DKA, should be sought on physical examination, even in the absence of fever. Tissue ischemia (heart, brain) can also be a precipitating factor. Omission of insulin because of an eating disorder may sometimes precipitate DKA.

Table 344-5 Manifestations of Diabetic Ketoacidosis

Symptoms
Nausea/vomiting
Thirst/polyuria
Abdominal pain
Shortness of breath
Precipitating events
Inadequate insulin administration
Infection (pneumonia/UTI/gastroenteritis/sepsis)
Infarction (cerebral, coronary, mesenteric, peripheral)
Drugs (cocaine)
Pregnancy
Physical Findings
Tachycardia
Dehydration/hypotension
Tachypnea/Kussmaul respirations/respiratory distress
Abdominal tenderness (may resemble acute pancreatitis or surgical abdomen)
Lethargy/obtundation/cerebral edema/possibly coma

PATHOPHYSIOLOGY

DKA results from relative or absolute insulin deficiency combined with counterregulatory hormone excess (glucagon, catecholamines, cortisol, and growth hormone). Both insulin deficiency and glucagon excess, in particular, are necessary for DKA to develop. The decreased ratio of insulin to glucagon promotes gluconeogenesis, glycogenolysis, and ketone body formation in the liver, as well as increases in substrate delivery from fat and muscle (free fatty acids, amino acids) to the liver. Markers of inflammation (cytokines, C-reactive protein) are elevated in both DKA and HHS

The combination of insulin deficiency and hyperglycemia reduces the hepatic level of fructose-2,6-bisphosphate, which alters the activity of phosphofructokinase and fructose-1,6-bisphosphatase. Glucagon excess decreases the activity of pyruvate kinase, whereas insulin deficiency increases the activity of phosphoenolpyruvate carboxykinase. These changes shift the handling of pyruvate toward glucose synthesis and away from glycolysis. The increased levels of glucagon and catecholamines in the face of low insulin levels promote glycogenolysis. Insulin deficiency also reduces levels of the GLUT4 glucose transporter, which impairs glucose uptake into skeletal muscle and fat and reduces intracellular glucose metabolism (Fig. 344-5).

Ketosis results from a marked increase in free fatty acid release from adipocytes, with a resulting shift toward ketone body synthesis in the liver. Reduced insulin levels, in combination with elevations in catecholamines and growth hormone, increase lipolysis and the release of free fatty acids. Normally, these free fatty acids are converted to triglycerides or VLDL in the liver. However, in DKA, hyperglucagonemia alters hepatic metabolism to favor ketone body formation, through activation of the enzyme carnitine palmitoyltransferase I. This enzyme is crucial for regulating fatty acid transport into the mitochondria, where beta oxidation and conversion to ketone bodies occur. At physiologic pH, ketone bodies exist as ketoacids, which are neutralized by bicarbonate. As bicarbonate stores are depleted, metabolic acidosis ensues. Increased lactic acid production also contributes to the acidosis. The increased free fatty acids increase triglyceride and VLDL production. VLDL clearance is also reduced because the activity of insulin-sensitive lipoprotein lipase in muscle and fat is decreased. Hypertriglyceridemia may be severe enough to cause pancreatitis.

DKA is often precipitated by increased insulin requirements, as occurs during a concurrent illness (Table 344-5). Failure to augment insulin therapy often compounds the problem. Complete omission or inadequate administration of insulin by the patient or health care team (in a hospitalized patient with type 1 DM) may precipitate DKA. Patients using insulin infusion devices with short-acting insulin may develop DKA, since even a brief interruption in insulin delivery (e.g., mechanical malfunction) quickly leads to insulin deficiency.

LABORATORY ABNORMALITIES AND DIAGNOSIS

The timely diagnosis of DKA is crucial and allows for prompt initiation of therapy. DKA is characterized by hyperglycemia, ketosis, and metabolic acidosis (increased anion gap) along with a number of secondary metabolic derangements (Table 344-4). Occasionally, the serum glucose is only minimally elevated. Serum bicarbonate is frequently <10 mmol/L, and arterial pH ranges between 6.8 and 7.3, depending on the severity of the acidosis. Despite a total-body potassium deficit, the serum potassium at presentation may be mildly elevated, secondary to the acidosis. Total-body stores of sodium, chloride, phosphorus, and magnesium are reduced in DKA but are not accurately reflected by their levels in the serum because of dehydration and hyperglycemia. Elevated blood urea nitrogen (BUN) and serum creatinine levels reflect intravascular volume depletion. Interference from acetoacetate may falsely elevate the serum creatinine measurement. Leukocytosis, hypertriglyceridemia, and hyperlipoproteinemia are commonly found as well. Hyperamylasemia may suggest a diagnosis of pancreatitis, especially when accompanied by abdominal pain. However, in DKA the amylase is usually of salivary origin and thus is not diagnostic of pancreatitis. Serum lipase should be obtained if pancreatitis is suspected.

The measured serum sodium is reduced as a consequence of the hyperglycemia [1.6-mmol/L (1.6 meq) reduction in serum sodium for each 5.6-mmol/L (100 mg/dL) rise in the serum glucose]. A normal serum sodium in the setting of DKA indicates a more profound water deficit. In "conventional" units, the calculated serum osmolality [2 x (serum sodium + serum potassium) + plasma glucose (mg/dL)/18 + BUN/2.8] is mildly to moderately elevated, though to a lesser degree than that found in HHS (see below).

In DKA, the ketone body, β-hydroxybutyrate, is synthesized at a threefold greater rate than acetoacetate; however, acetoacetate is preferentially detected by a commonly used ketosis detection reagent (nitroprusside). Serum ketones are present at significant levels (usually positive at serum dilution of ≥1:8). The nitroprusside tablet, or stick, is often used to detect urine ketones; certain medications such as captopril or penicillamine may cause false-positive reactions. Serum or plasma assays for β-hydroxybutyrate are preferred since they more accurately reflect the true ketone body level.

The metabolic derangements of DKA exist along a spectrum, beginning with mild acidosis with moderate hyperglycemia evolving into more severe findings. The degree of acidosis and hyperglycemia do not necessarily correlate closely since a variety of factors determine the level of hyperglycemia (oral intake, urinary glucose loss). Ketonemia is a consistent finding in DKA and distinguishes it from simple hyperglycemia. The differential diagnosis of DKA includes starvation ketosis, alcoholic ketoacidosis (bicarbonate usually >15 meq/L) and other forms of increased anion-gap acidosis (Chap. 47).

TREATMENT: DIABETIC KETOACIDOSIS

The management of DKA is outlined in **Table 344-6**. After initiating IV fluid replacement and insulin therapy, the agent or event that precipitated the episode of DKA should be sought and aggressively treated. If the patient is vomiting or has altered mental status, a nasogastric tube should be inserted to prevent aspiration of gastric contents. Central to successful treatment of DKA is careful monitoring and frequent reassessment to ensure that the patient and the metabolic derangements are improving. A comprehensive flow sheet should record chronologic changes in vital signs, fluid intake and output, and laboratory values as a function of insulin administered.

Table 344-6 Management of Diabetic Ketoacidosis

1. Confirm diagnosis (↑plasma glucose, positive serum ketones, metabolic acidosis).
2. Admit to hospital; intensive-care setting may be necessary for frequent monitoring or if pH <7.00 or unconscious. 3. Assess: Serum electrolytes (K^+, Na^+, Mg^{2+}, Cl^-, bicarbonate, phosphate) Acid-base status—pH, HCO_3^-, PCO_2, β-hydroxybutyrate Renal function (creatinine, urine output)
4. Replace fluids: 2–3 L of 0.9% saline over first 1–3 h ? (15–20 mL/kg per hour); subsequently, 0.45% saline at 250–500 mL/h; change to 5% glucose and 0.45% saline at 150–250 mL/h when plasma glucose reaches 200 mg/dL (11.2 mmol/L).
5. Administer short-acting insulin: IV (0.1 units/kg), then 0.1 units/kg per hour by continuous IV infusion; increase two- to threefold if no response by 2–4 h. If the initial serum potassium is <3.3 mmol/L (3.3 meq/L), do not administer insulin until the potassium is corrected. If the initial serum potassium is >5.2 mmol/L (5.2 meq/L), do not supplement K^+ until the potassium is corrected.
6. Assess patient: What precipitated the episode (noncompliance, infection, trauma, infarction, cocaine)? Initiate appropriate workup for precipitating event (cultures, CXR, ECG).
7. Measure capillary glucose every 1–2 h; measure electrolytes (especially K^+, bicarbonate, phosphate) and anion gap every 4 h for first 24 h.
8. Monitor blood pressure, pulse, respirations, mental status, fluid intake and output every 1–4 h.
9. Replace K^+: 10 meq/h when plasma K^+ < 5.0–5.2 meq/L (or 20–30 meq/L of infusion fluid), ECG normal, urine flow and normal creatinine documented; administer 40–80 meq/h when plasma K^+ < 3.5 meq/L or if bicarbonate is given. See text about bicarbonate or phosphate supplementation.
10. Continue above until patient is stable, glucose goal is 8.3–13.9 mmol/L (150–250 mg/dL), and acidosis is resolved. Insulin infusion may be decreased to 0.05–0.1 units/kg per hour.
11. Administer long-acting insulin as soon as patient is eating. Allow for overlap in insulin infusion and SC insulin injection.

Abbreviations: CXR, chest x-ray; ECG, electrocardiogram. ***Source:*** Adapted from M Sperling, in *Therapy for Diabetes Mellitus and Related Disorders*, American Diabetes Association, Alexandria, VA, 1998; and AE Kitabchi et al: Diabetes Care 32:1335, 2009.

After the initial bolus of normal saline, replacement of the sodium and free water deficit is carried out over the next 24 h (fluid deficit is often 3–5 L). When hemodynamic stability and adequate urine output are achieved, IV fluids should be switched to 0.45% saline depending on the calculated volume deficit. The change to 0.45% saline helps to reduce the trend toward hyperchloremia later in the course of DKA. Alternatively, initial use of lactated Ringer's IV solution may reduce the hyperchloremia that commonly occurs with normal saline.

A bolus of IV (0.1 units/kg) short-acting insulin should be administered immediately (Table 344-6), and subsequent treatment should provide continuous and adequate levels of circulating insulin. IV administration is preferred (0.1 units/kg of regular insulin per hour), because it ensures rapid distribution and allows adjustment of the infusion rate as the patient responds to therapy. In

mild episodes of DKA, short-acting insulin analogues can be used SC. IV insulin should be continued until the acidosis resolves and the patient is metabolically stable. As the acidosis and insulin resistance associated with DKA resolve, the insulin infusion rate can be decreased (to 0.05–0.1 units/kg per hour). Long-acting insulin, in combination with SC short-acting insulin, should be administered as soon as the patient resumes eating, as this facilitates transition to an outpatient insulin regimen and reduces length of hospital stay. It is crucial to continue the insulin infusion until adequate insulin levels are achieved by administering long-acting insulin by the SC route. Even relatively brief periods of inadequate insulin administration in this transition phase may result in DKA relapse.

Hyperglycemia usually improves at a rate of 4.2–5.6 mmol/L (75–100 mg/dL) per hour as a result of insulin-mediated glucose disposal, reduced hepatic glucose release, and rehydration. The latter reduces catecholamines, increases urinary glucose loss, and expands the intravascular volume. The decline in the plasma glucose within the first 1–2 h may be more rapid and is mostly related to volume expansion. When the plasma glucose reaches 11.2 mmol/L (200 mg/dL), glucose should be added to the 0.45% saline infusion to maintain the plasma glucose in the 8.3–13.9 mmol/L (150–250 mg/dL) range, and the insulin infusion should be continued. Ketoacidosis begins to resolve as insulin reduces lipolysis, increases peripheral ketone body use, suppresses hepatic ketone body formation, and promotes bicarbonate regeneration. However, the acidosis and ketosis resolve more slowly than hyperglycemia. As ketoacidosis improves, β-hydroxybutyrate is converted to acetoacetate. Ketone body levels may appear to increase if measured by laboratory assays that use the nitroprusside reaction, which only detects acetoacetate and acetone. The improvement in acidosis and anion gap, a result of bicarbonate regeneration and decline in ketone bodies, is reflected by a rise in the serum bicarbonate level and the arterial pH. Depending on the rise of serum chloride, the anion gap (but not bicarbonate) will normalize. A hyperchloremic acidosis [serum bicarbonate of 15–18 mmol/L (15–18 meq/L)] often follows successful treatment and gradually resolves as the kidneys regenerate bicarbonate and excrete chloride.

Potassium stores are depleted in DKA [estimated deficit 3–5 mmol/kg (3–5 meq/kg)]. During treatment with insulin and fluids, various factors contribute to the development of hypokalemia. These include insulin-mediated potassium transport into cells, resolution of the acidosis (which also promotes potassium entry into cells), and urinary loss of potassium salts of organic acids. Thus, potassium repletion should commence as soon as adequate urine output and a normal serum potassium are documented. If the initial serum potassium level is elevated, then potassium repletion should be delayed until the potassium falls into the normal range. Inclusion of 20–40 meq of potassium in each liter of IV fluid is reasonable, but additional potassium supplements may also be required. To reduce the amount of chloride administered, potassium phosphate or acetate can be substituted for the chloride salt. The goal is to maintain the serum potassium at >3.5 mmol/L (3.5 meq/L).

Despite a bicarbonate deficit, bicarbonate replacement is not usually necessary. In fact, theoretical arguments suggest that bicarbonate administration and rapid reversal of acidosis may impair cardiac function, reduce tissue oxygenation, and promote hypokalemia. The results of most clinical trials do not support the routine use of bicarbonate replacement, and one study in children found that bicarbonate use was associated with an increased risk of cerebral edema. However, in the presence of severe acidosis (arterial pH <6.9), the ADA advises bicarbonate [50 mmol/L (meq/L) of sodium bicarbonate in 200 mL of sterile water with 10 meq/L KCl per hour for 2 h until the pH is >7.0]. Hypophosphatemia may result from increased glucose usage, but randomized clinical trials have not demonstrated that phosphate replacement is beneficial in DKA. If the serum phosphate < 0.32 mmol/L (1 mg/dL), then phosphate supplement should be considered and the serum calcium monitored. Hypomagnesemia may develop during DKA therapy and may also require supplementation.

With appropriate therapy, the mortality rate of DKA is low (<1%) and is related more to the underlying or precipitating event, such as infection or myocardial infarction. Venous thrombosis, upper gastrointestinal bleeding, and acute respiratory distress syndrome occasionally complicate DKA. The major nonmetabolic complication of DKA therapy is cerebral edema, which most often develops in children as DKA is resolving. The etiology of and optimal therapy for cerebral edema are not well established, but overreplacement of free water should be avoided.

Following treatment, the physician and patient should review the sequence of events that led to DKA to prevent future recurrences. Foremost is patient education about the symptoms of DKA, its precipitating factors, and the management of diabetes during a concurrent illness. During illness or when oral intake is compromised, patients should (1) frequently measure the capillary blood glucose; (2) measure urinary ketones when the serum glucose > 16.5 mmol/L (300 mg/dL); (3) drink fluids to maintain hydration; (4) continue or increase insulin; and (5) seek medical attention if dehydration, persistent vomiting, or uncontrolled hyperglycemia develop. Using these strategies, early DKA can be prevented or detected and treated appropriately on an outpatient basis.

Hyperglycemic Hyperosmolar State

CLINICAL FEATURES

The prototypical patient with HHS is an elderly individual with type 2 DM, with a several-week history of polyuria, weight loss, and diminished oral intake that culminates in mental confusion, lethargy, or coma. The physical examination reflects profound dehydration and hyperosmolality and reveals hypotension, tachycardia, and altered mental status. Notably absent are symptoms of nausea, vomiting, and abdominal pain and the Kussmaul respirations characteristic of DKA. HHS is often precipitated by a serious, concurrent illness such as myocardial infarction or stroke. Sepsis, pneumonia, and other serious infections are frequent precipitants and should be sought. In addition, a debilitating condition (prior stroke or dementia) or social situation that compromises water intake usually contributes to the development of the disorder.

PATHOPHYSIOLOGY

Relative insulin deficiency and inadequate fluid intake are the underlying causes of HHS. Insulin deficiency increases hepatic glucose production (through glycogenolysis and gluconeogenesis) and impairs glucose utilization in skeletal muscle (see above discussion of DKA). Hyperglycemia induces an osmotic diuresis that leads to intravascular volume depletion, which is exacerbated by inadequate fluid replacement. The absence of ketosis in HHS is not understood. Presumably, the insulin deficiency is only relative and less severe than in DKA. Lower levels of counterregulatory hormones and free fatty acids have been found in HHS than in DKA in some studies. It is also possible that the liver is less capable of ketone body synthesis or that the insulin/glucagon ratio does not favor ketogenesis.

LABORATORY ABNORMALITIES AND DIAGNOSIS

The laboratory features in HHS are summarized in Table 344-4. Most notable are the marked hyperglycemia [plasma glucose may be >55.5 mmol/L (1000 mg/dL)], hyperosmolality (>350 mosmol/L), and prerenal azotemia. The measured serum sodium may be normal or slightly low despite the marked hyperglycemia. The corrected serum sodium is usually increased [add 1.6 meq to measured sodium for each 5.6-mmol/L (100 mg/dL) rise in the serum glucose]. In contrast to DKA, acidosis and ketonemia are absent or mild. A small anion-gap metabolic acidosis may be present secondary to increased lactic acid. Moderate ketonuria, if present, is secondary to starvation.

TREATMENT: HYPERGLYCEMIC HYPEROSMOLAR STATE

Volume depletion and hyperglycemia are prominent features of both HHS and DKA. Consequently, therapy of these disorders shares several elements (Table 344-6). In both disorders, careful monitoring of the patient's fluid status, laboratory values, and insulin infusion rate is crucial. Underlying or precipitating problems should be aggressively sought and treated. In HHS, fluid

losses and dehydration are usually more pronounced than in DKA due to the longer duration of the illness. The patient with HHS is usually older, more likely to have mental status changes, and more likely to have a life-threatening precipitating event with accompanying comorbidities. Even with proper treatment, HHS has a substantially higher mortality rate than DKA (up to 15% in some clinical series).

Fluid replacement should initially stabilize the hemodynamic status of the patient (1–3 L of 0.9% normal saline over the first 2–3 h). Because the fluid deficit in HHS is accumulated over a period of days to weeks, the rapidity of reversal of the hyperosmolar state must balance the need for free water repletion with the risk that too rapid a reversal may worsen neurologic function. If the serum sodium > 150 mmol/L (150 meq/L), 0.45% saline should be used. After hemodynamic stability is achieved, the IV fluid administration is directed at reversing the free water deficit using hypotonic fluids (0.45% saline initially, then 5% dextrose in water, D5W). The calculated free water deficit (which averages 9–10 L) should be reversed over the next 1–2 days (infusion rates of 200–300 mL/h of hypotonic solution). Potassium repletion is usually necessary and should be dictated by repeated measurements of the serum potassium. In patients taking diuretics, the potassium deficit can be quite large and may be accompanied by magnesium deficiency. Hypophosphatemia may occur during therapy and can be improved by using KPO_4 and beginning nutrition.

As in DKA, rehydration and volume expansion lower the plasma glucose initially, but insulin is also required. A reasonable regimen for HHS begins with an IV insulin bolus of 0.1 units/kg followed by IV insulin at a constant infusion rate of 0.1 units/kg per hour. If the serum glucose does not fall, increase the insulin infusion rate by twofold. As in DKA, glucose should be added to IV fluid when the plasma glucose falls to 13.9–16.7 mmol/L (250–300 mg/dL), and the insulin infusion rate should be decreased to 0.05–0.1 units/kg per hour. The insulin infusion should be continued until the patient has resumed eating and can be transferred to a SC insulin regimen. The patient should be discharged from the hospital on insulin, though some patients can later switch to oral glucose-lowering agents.

CHRONIC COMPLICATIONS OF DM

The chronic complications of DM affect many organ systems and are responsible for the majority of morbidity and mortality associated with the disease. Chronic complications can be divided into vascular and nonvascular complications **(Table 344-7)**. The vascular complications of DM are further subdivided into microvascular (retinopathy, neuropathy, nephropathy) and macrovascular complications [coronary heart disease (CHD), peripheral arterial disease (PAD), cerebrovascular disease]. Nonvascular complications include problems such as gastroparesis, infections, and skin changes. Long-standing diabetes may be associated with hearing loss. Whether type 2 DM in elderly individuals is associated with impaired mental function is not clear.

Table 344-7 Chronic Complications of Diabetes Mellitus

Microvascular
Eye disease
Retinopathy (nonproliferative/proliferative)
Macular edema
Neuropathy
Sensory and motor (mono- and polyneuropathy)
Autonomic
Nephropathy
Macrovascular
Coronary heart disease
Peripheral arterial disease
Cerebrovascular disease
Other
Gastrointestinal (gastroparesis, diarrhea)
Genitourinary (uropathy/sexual dysfunction)
Dermatologic
Infectious
Cataracts
Glaucoma
Periodontal disease
Hearing loss

The risk of chronic complications increases as a function of the duration and degree of hyperglycemia; they usually do not become apparent until the second decade of hyperglycemia. Since type 2 DM often has a long asymptomatic period of hyperglycemia, many individuals with type 2 DM have complications at the time of diagnosis.

The microvascular complications of both type 1 and type 2 DM result from chronic hyperglycemia. Large, randomized clinical trials of individuals with type 1 or type 2 DM have conclusively demonstrated that a reduction in chronic hyperglycemia prevents or delays retinopathy, neuropathy, and nephropathy. Other incompletely defined factors may modulate the development of complications. For example, despite long-standing DM, some individuals never develop nephropathy or retinopathy. Many of these patients have glycemic control that is indistinguishable from those who develop microvascular complications, suggesting that there is a genetic susceptibility for developing particular complications.

Evidence implicating a causative role for chronic hyperglycemia in the development of macrovascular complications is less conclusive. However, coronary heart disease events and mortality rate are two to four times greater in patients with type 2 DM. These events correlate with fasting and postprandial plasma glucose levels as well as with the A1C. Other factors (dyslipidemia and hypertension) also play important roles in macrovascular complications.

Mechanisms of Complications

Although chronic hyperglycemia is an important etiologic factor leading to complications of DM, the mechanism(s) by which it leads to such diverse cellular and organ dysfunction is unknown. At least four prominent theories, which are not mutually exclusive, have been proposed to explain how hyperglycemia might lead to the chronic complications of DM. An emerging hypothesis is that hyperglycemia leads to epigenetic changes in the affected cells.

One theory is that increased intracellular glucose leads to the formation of advanced glycosylation end products (AGEs), which

bind to a cell surface receptor, via the nonenzymatic glycosylation of intra- and extracellular proteins. Nonenzymatic glycosylation results from the interaction of glucose with amino groups on proteins. AGEs have been shown to cross-link proteins (e.g., collagen, extracellular matrix proteins), accelerate atherosclerosis, promote glomerular dysfunction, reduce nitric oxide synthesis, induce endothelial dysfunction, and alter extracellular matrix composition and structure. The serum level of AGEs correlates with the level of glycemia, and these products accumulate as the glomerular filtration rate (GFR) declines.

A second theory is based on the observation that hyperglycemia increases glucose metabolism via the sorbitol pathway. Intracellular glucose is predominantly metabolized by phosphorylation and subsequent glycolysis, but when increased, some glucose is converted to sorbitol by the enzyme aldose reductase. Increased sorbitol concentration alters redox potential, increases cellular osmolality, generates reactive oxygen species, and likely leads to other types of cellular dysfunction. However, testing of this theory in humans, using aldose reductase inhibitors, has not demonstrated significant beneficial effects on clinical endpoints of retinopathy, neuropathy, or nephropathy.

A third hypothesis proposes that hyperglycemia increases the formation of diacylglycerol leading to activation of protein kinase C (PKC). Among other actions, PKC alters the transcription of genes for fibronectin, type IV collagen, contractile proteins, and extracellular matrix proteins in endothelial cells and neurons. Inhibitors of PKC are being studied in clinical trials.

A fourth theory proposes that hyperglycemia increases the flux through the hexosamine pathway, which generates fructose-6-phosphate, a substrate for O-linked glycosylation and proteoglycan production. The hexosamine pathway may alter function by glycosylation of proteins such as endothelial nitric oxide synthase or by changes in gene expression of transforming growth factor β (TGF-β) or plasminogen activator inhibitor-1 (PAI-1).

Growth factors appear to play an important role in some DM-related complications, and their production is increased by most of these proposed pathways. Vascular endothelial growth factor A (VEGF-A) is increased locally in diabetic proliferative retinopathy and decreases after laser photocoagulation. TGF-β is increased in diabetic nephropathy and stimulates basement membrane production of collagen and fibronectin by mesangial cells. Other growth factors, such as platelet-derived growth factor, epidermal growth factor, insulin-like growth factor I, growth hormone, basic fibroblast growth factor, and even insulin, have been suggested to play a role in DM-related complications. A possible unifying mechanism is that hyperglycemia leads to increased production of reactive oxygen species or superoxide in the mitochondria; these compounds may activate all four of the pathways described above. Although hyperglycemia serves as the initial trigger for complications of diabetes, it is still unknown whether the same pathophysiologic processes are operative in all complications or whether some pathways predominate in certain organs.

Glycemic Control and Complications

The Diabetes Control and Complications Trial (DCCT) provided definitive proof that reduction in chronic hyperglycemia can prevent many of the early complications of type 1 DM. This large multicenter clinical trial randomized more than 1400 individuals with type 1 DM to either intensive or conventional diabetes management and prospectively evaluated the development of retinopathy, nephropathy, and neuropathy. Individuals in the intensive diabetes management group received multiple administrations of insulin each day along with extensive educational, psychological, and medical support. Individuals in the conventional diabetes management group received twice-daily insulin injections and quarterly nutritional, educational, and clinical evaluation. The goal in the former group was normoglycemia; the goal in the latter group was prevention of symptoms of diabetes. Individuals in the intensive diabetes management group achieved a substantially lower hemoglobin A_{1C} (7.3%) than individuals in the conventional diabetes management group (9.1%).

The DCCT demonstrated that improvement of glycemic control reduced nonproliferative and proliferative retinopathy (47% reduction), microalbuminuria (39% reduction), clinical nephropathy (54% reduction), and neuropathy (60% reduction). Improved glycemic control also slowed the progression of early diabetic complications. There was a nonsignificant trend in reduction of macrovascular events during the trial (most individuals were young and had a low risk of cardiovascular disease). The results of the DCCT predicted that individuals in the intensive diabetes management group would gain 7.7 additional years of vision, 5.8 additional years free from ESRD, and 5.6 years free from lower extremity amputations. If all complications of DM were combined, individuals in the intensive diabetes management group would experience 15.3 more years of life without significant microvascular or neurologic complications of DM, compared to individuals who received standard therapy. This translates into an additional 5.1 years of life expectancy for individuals in the intensive diabetes management group. The long-term prognosis for type 1 diabetes continues to improve as shown by 30-year incidence data in the intensively treated group from the DCCT of retinopathy (21%), nephropathy (9%), and cardiovascular disease (9%). During this follow-up, fewer than 1% of the cohort had become blind, lost a limb to amputation, or required dialysis. The benefit of the improved glycemic control during the DCCT persisted even after the study concluded and glycemic control worsened. For example, individuals in the intensive diabetes management group for a mean of 6.5 years had a 42–57% reduction in cardiovascular events [nonfatal myocardial infarction (MI), stroke, or death from a cardiovascular event] at a mean follow-up of 17 years, even though their subsequent glycemic control was the same as those in the conventional diabetes management group from years 6.5–17 (discussed below).

The benefits of an improvement in glycemic control occurred over the entire range of A_{1C} values **(Fig. 344-8)**, suggesting that at any A_{1C} level, an improvement in glycemic control is beneficial. The goal of therapy is to achieve an A_{1C} level as close to normal as possible, without subjecting the patient to excessive risk of hypoglycemia.

Figure 344-8

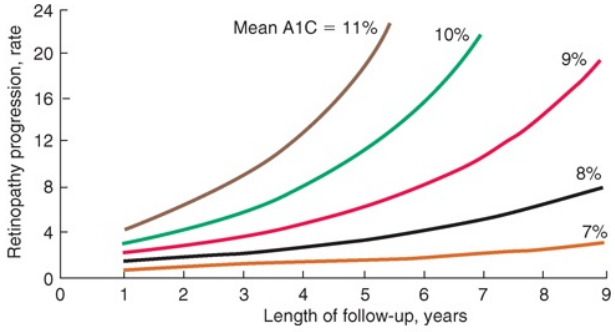

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship of glycemic control and diabetes duration to diabetic retinopathy. The progression of retinopathy in individuals in the Diabetes Control and Complications Trial is graphed as a function of the length of follow-up with different

curves for different A_{1C} values. *(Adapted from The Diabetes Control and Complications Trial Research Group: Diabetes 44:968, 1995.)*

The United Kingdom Prospective Diabetes Study (UKPDS) studied the course of >5000 individuals with type 2 DM for >10 years. This study utilized multiple treatment regimens and monitored the effect of intensive glycemic control and risk factor treatment on the development of diabetic complications. Newly diagnosed individuals with type 2 DM were randomized to (1) intensive management using various combinations of insulin, a sulfonylurea, or metformin or (2) conventional therapy using dietary modification and pharmacotherapy with the goal of symptom prevention. In addition, individuals were randomly assigned to different antihypertensive regimens. Individuals in the intensive treatment arm achieved an A_{1C} of 7%, compared to a 7.9% A_{1C} in the standard treatment group. The UKPDS demonstrated that each percentage point reduction in A_{1C} was associated with a 35% reduction in microvascular complications. As in the DCCT, there was a continuous relationship between glycemic control and development of complications. Improved glycemic control did not conclusively reduce (nor worsen) cardiovascular mortality rate during the period of the trial, but was associated with improvement with lipoprotein risk profiles, such as reduced triglycerides and increased HDL.

One of the major findings of the UKPDS was that strict blood pressure control significantly reduced both macro- and microvascular complications. In fact, the beneficial effects of blood pressure control were greater than the beneficial effects of glycemic control. Lowering blood pressure to moderate goals (144/82 mmHg) reduced the risk of DM-related death, stroke, microvascular end- points, retinopathy, and heart failure (risk reductions between 32 and 56%).

Similar reductions in the risks of retinopathy and nephropathy were also seen in a small trial of lean Japanese individuals with type 2 DM randomized to either intensive glycemic control or standard therapy with insulin (Kumamoto study). These results demonstrate the effectiveness of improved glycemic control in individuals of different ethnicity and, presumably, a different etiology of DM (i.e., phenotypically different from those in the DCCT and UKPDS).

The findings of the DCCT, UKPDS, and Kumamoto study strongly support the idea that chronic hyperglycemia plays a causative role in the pathogenesis of diabetic microvascular complications. These landmark studies prove the value of metabolic control and emphasize the importance of (1) intensive glycemic control in all forms of DM and (2) early diagnosis and strict blood pressure control in type 2 DM. Optimal targets for glycemic control and blood pressure are not entirely clear (see below).

Ophthalmologic Complications of Diabetes Mellitus

DM is the leading cause of blindness between the ages of 20 and 74 in the United States. The gravity of this problem is highlighted by the finding that individuals with DM are 25 times more likely to become legally blind than individuals without DM. Blindness is primarily the result of progressive diabetic retinopathy and clinically significant macular edema. Diabetic retinopathy is classified into two stages: nonproliferative and proliferative. Nonproliferative diabetic retinopathy usually appears late in the first decade or early in the second decade of the disease and is marked by retinal vascular microaneurysms, blot hemorrhages, and cotton-wool spots **(Fig. 344-9)**. Mild nonproliferative retinopathy progresses to more extensive disease, characterized by changes in venous vessel caliber, intraretinal microvascular abnormalities, and more numerous microaneurysms and hemorrhages. The pathophysiologic mechanisms invoked in nonproliferative retinopathy include loss of retinal pericytes, increased retinal vascular permeability, alterations in retinal blood flow, and abnormal retinal microvasculature, all of which lead to retinal ischemia.

Figure 344-9

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diabetic retinopathy results in scattered hemorrhages, yellow exudates, and neovascularization. This patient has neovascular vessels proliferating from the optic disc, requiring urgent panretinal laser photocoagulation.

The appearance of neovascularization in response to retinal hypoxemia is the hallmark of proliferative diabetic retinopathy (Fig. 344-9). These newly formed vessels appear near the optic nerve and/or macula and rupture easily, leading to vitreous hemorrhage, fibrosis, and ultimately retinal detachment. Not all individuals with nonproliferative retinopathy develop proliferative retinopathy, but the more severe the nonproliferative disease, the greater the chance of evolution to proliferative retinopathy within 5 years. This creates an important opportunity for early detection and treatment of diabetic retinopathy. Clinically significant macular edema can occur when only nonproliferative retinopathy is present. Fluorescein angiography is useful to detect macular edema, which is associated with a 25% chance of moderate visual loss over the next 3 years.

Duration of DM and degree of glycemic control are the best predictors of the development of retinopathy; hypertension is also a risk factor. Nonproliferative retinopathy is found in many individuals who have had DM for >20 years (25% incidence with 5 years, and 80% incidence with 15 years of type 1 DM). Although there is genetic susceptibility for retinopathy, it confers less influence than either the duration of DM or the degree of glycemic control.

TREATMENT: DIABETIC RETINOPATHY

The most effective therapy for diabetic retinopathy is prevention. Intensive glycemic and blood pressure control will delay the development or slow the progression of retinopathy in individuals with either type 1 or type 2 DM. Paradoxically, during the first 6–12 months of improved glycemic control, established diabetic retinopathy may transiently worsen. Fortunately, this progression is temporary, and in the long term, improved glycemic control is associated with less diabetic retinopathy. Individuals with known retinopathy are candidates for prophylactic photocoagulation when initiating intensive therapy. Once advanced

retinopathy is present, improved glycemic control imparts less benefit, though adequate ophthalmologic care can prevent most blindness.

Regular, comprehensive eye examinations are essential for all individuals with DM. Most diabetic eye disease can be successfully treated if detected early. Routine, nondilated eye examinations by the primary care provider or diabetes specialist are inadequate to detect diabetic eye disease, which requires an ophthalmologist for optimal care of these disorders. Laser photocoagulation is very successful in preserving vision. Proliferative retinopathy is usually treated with panretinal laser photocoagulation, whereas macular edema is treated with focal laser photocoagulation. Although exercise has not been conclusively shown to worsen proliferative diabetic retinopathy, most ophthalmologists advise individuals with advanced diabetic eye disease to limit physical activities associated with repeated Valsalva maneuvers. Aspirin therapy (650 mg/d) does not appear to influence the natural history of diabetic retinopathy.

Renal Complications of Diabetes Mellitus

Diabetic nephropathy is the leading cause of ESRD in the United States and a leading cause of DM-related morbidity and mortality. Both microalbuminuria and macroalbuminuria in individuals with DM are associated with increased risk of cardiovascular disease. Individuals with diabetic nephropathy commonly have diabetic retinopathy.

Like other microvascular complications, the pathogenesis of diabetic nephropathy is related to chronic hyperglycemia. The mechanisms by which chronic hyperglycemia leads to ESRD, though incompletely defined, involve the effects of soluble factors (growth factors, angiotensin II, endothelin, AGEs), hemodynamic alterations in the renal microcirculation (glomerular hyperfiltration or hyperperfusion, increased glomerular capillary pressure), and structural changes in the glomerulus (increased extracellular matrix, basement membrane thickening, mesangial expansion, fibrosis). Some of these effects may be mediated through angiotensin II receptors. Smoking accelerates the decline in renal function. Because only 20–40% of patients with diabetes develop diabetic nephropathy, additional susceptibility factors remain unidentified. One known risk factor is a family history of diabetic nephropathy.

The natural history of diabetic nephropathy is characterized by a fairly predictable sequence of events that was initially defined for individuals with type 1 DM but appears to be similar in type 2 DM **(Fig. 344-10)**. Glomerular hyperperfusion and renal hypertrophy occur in the first years after the onset of DM and are associated with an increase of the GFR. During the first 5 years of DM, thickening of the glomerular basement membrane, glomerular hypertrophy, and mesangial volume expansion occur as the GFR returns to normal. After 5–10 years of type 1 DM, ~40% of individuals begin to excrete small amounts of albumin in the urine. Microalbuminuria is defined as 30–299 mg/d in a 24-h collection or 30–299 μg/mg creatinine in a spot collection (preferred method). Although the appearance of microalbuminuria in type 1 DM is an important risk factor for progression to macroalbuminuria (>300 mg/d or > 300 μg/mg creatinine), only ~50% of individuals progress to macroalbuminuria over the next 10 years. In some individuals with type 1 diabetes and microalbuminuria of short duration, the microalbuminuria regresses. Microalbuminuria is a risk factor for cardiovascular disease. Once macroalbuminuria is present, there is a steady decline in GFR, and ~50% of individuals reach ESRD in 7–10 years. Once macroalbuminuria develops, blood pressure rises slightly and the pathologic changes are likely irreversible.

Figure 344-10

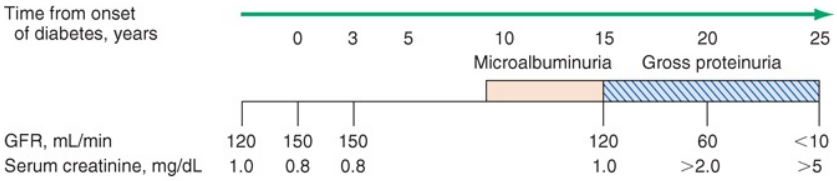

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Time course of development of diabetic nephropathy. The relationship of time from onset of diabetes, the glomerular filtration rate (GFR), and the serum creatinine are shown. *(Adapted from RA DeFranzo, in Therapy for Diabetes Mellitus and Related Disorders, 3rd ed. American Diabetes Association, Alexandria, VA, 1998.)*

The nephropathy that develops in type 2 DM differs from that of type 1 DM in the following respects: (1) microalbuminuria or macroalbuminuria may be present when type 2 DM is diagnosed, reflecting its long asymptomatic period; (2) hypertension more commonly accompanies microalbuminuria or macroalbuminuria in type 2 DM; and (3) microalbuminuria may be less predictive of diabetic nephropathy and progression to macroalbuminuria in type 2 DM. Finally, it should be noted that albuminuria in type 2 DM may be secondary to factors unrelated to DM, such as hypertension, congestive heart failure (CHF), prostate disease, or infection. Diabetic nephropathy and ESRD secondary to DM develop more commonly in African Americans, Native Americans, and Hispanic individuals than in Caucasians with type 2 DM.

Type IV renal tubular acidosis (hyporeninemic hypoaldosteronism) may occur in type 1 or 2 DM. These individuals develop a propensity to hyperkalemia, which may be exacerbated by medications [especially angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs)]. Patients with DM are predisposed to radiocontrast-induced nephrotoxicity. Risk factors for radiocontrast-induced nephrotoxicity are preexisting nephropathy and volume depletion. Individuals with DM undergoing radiographic procedures with contrast dye should be well hydrated before and after dye exposure, and the serum creatinine should be monitored for 24–48 h following the procedure.

TREATMENT: DIABETIC NEPHROPATHY

The optimal therapy for diabetic nephropathy is prevention by control of glycemia. As part of comprehensive diabetes care, microalbuminuria should be detected at an early stage when effective therapies can be instituted. The recommended strategy for detecting microalbuminuria is outlined in **Fig. 344-11**. Since some individuals with type 1 or type 2 DM have a decline in GFR in the absence of micro- or macroalbuminuria, annual measurement of the serum creatinine to estimate GFR should also be performed. Interventions effective in slowing progression from microalbuminuria to macroalbuminuria include (1) normalization of glycemia, (2) strict blood pressure control, and (3) administration of ACE inhibitors or ARBs. Dyslipidemia should also be treated.

Figure 344-11

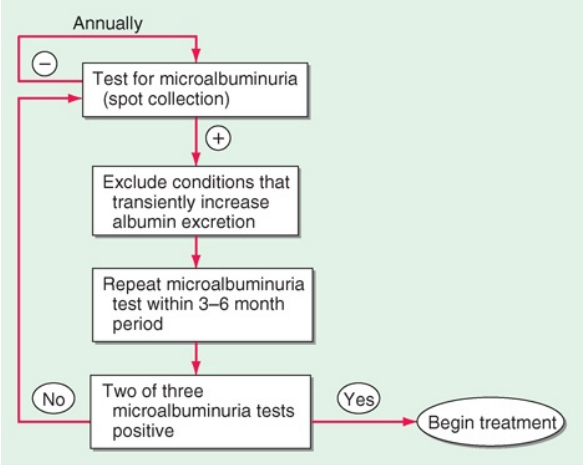

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Screening for microalbuminuria should be performed in patients with type 1 diabetes for ≥5 years, in patients with type 2 diabetes, and during pregnancy. Non-diabetes-related conditions that might increase microalbuminuria are urinary tract infection, hematuria, heart failure, febrile illness, severe hyperglycemia, severe hypertension, and vigorous exercise. *(Adapted from RA DeFronzo, in Therapy for Diabetes Mellitus and Related Disorders, 3rd ed. American Diabetes Association, Alexandria, VA, 1998.)*

Improved glycemic control reduces the rate at which microalbuminuria appears and progresses in type 1 and type 2 DM. However, once macroalbuminuria exists, it is unclear whether improved glycemic control will slow progression of renal disease. During the later phase of declining renal function, insulin requirements may fall as the kidney is a site of insulin degradation. Furthermore, many glucose-lowering medications (sulfonylureas and metformin) are contraindicated in advanced renal insufficiency.

Many individuals with type 1 or type 2 DM develop hypertension. Numerous studies in both type 1 and type 2 DM demonstrate the effectiveness of strict blood pressure control in reducing albumin excretion and slowing the decline in renal function. Blood pressure should be maintained at <130/80 mmHg in diabetic individuals.

Either ACE inhibitors or ARBs should be used to reduce the progression from microalbuminuria to macroalbuminuria and the associated decline in GFR that accompanies macroalbumin-uria in individuals with type 1 or type 2 DM (see "Hypertension," below). Although direct comparisons of ACE inhibitors and ARBs are lacking, most experts believe that the two classes of drugs are equivalent in the patient with diabetes. ARBs can be used as an alternative in patients who develop ACE inhibitor –associated cough or angioedema. After 2–3 months of therapy in patients with microalbuminuria, the drug dose is increased until either the microalbuminuria disappears or the maximum dose is reached. If use of either ACE inhibitors or ARBs is not possible or the blood pressure is not controlled, then calcium channel blockers (non-dihydropyridine class), beta blockers, or diuretics should be used. However, their efficacy in slowing the fall in the GFR is not proven. Blood pressure control with any agent is extremely important, but a drug-specific benefit in diabetic nephropathy, independent of blood pressure control, has been shown only for ACE inhibitors and ARBs in patients with DM.

The ADA suggests modest restriction of protein intake in diabetic individuals with microalbuminuria (0.8–1.0 g/kg per day) or macroalbuminuria (<0.8 g/kg per day).

Nephrology consultation should be considered when the estimated GFR <60 mL/min per 1.743 m^2. Once macroalbuminuria ensues, the likelihood of ESRD is very high. As compared to nondiabetic individuals, hemodialysis in patients with DM is associated with more frequent complications, such as hypotension (due to autonomic neuropathy or loss of reflex tachycardia), more difficult vascular access, and accelerated progression of retinopathy. Survival after the onset of ESRD is shorter in the diabetic population compared to nondiabetics with similar clinical features. Atherosclerosis is the leading cause of death in diabetic individuals on dialysis, and hyperlipidemia should be treated aggressively. Renal transplantation from a living related donor is the preferred therapy but requires chronic immunosuppression. Combined pancreas-kidney transplant offers the promise of normoglycemia and freedom from dialysis.

Neuropathy and Diabetes Mellitus

Diabetic neuropathy occurs in ~50% of individuals with long-standing type 1 and type 2 DM. It may manifest as polyneuropathy, mononeuropathy, and/or autonomic neuropathy. As with other complications of DM, the development of neuropathy correlates with the duration of diabetes and glycemic control. Additional risk factors are BMI (the greater the BMI, the greater the risk of neurop-athy) and smoking. The presence of cardiovascular disease, elevated triglycerides, and hypertension is also associated with diabetic peripheral neuropathy. Both myelinated and unmyelinated nerve fibers are lost. Because the clinical features of diabetic neuropathy are similar to those of other neuropathies, the diagnosis of diabetic neuropathy should be made only after other possible etiologies are excluded (Chap. 384).

POLYNEUROPATHY/MONONEUROPATHY

The most common form of diabetic neuropathy is distal symmetric polyneuropathy. It most frequently presents with distal sensory loss, but up to 50% of patients do not have symptoms of neuropathy. Hyperesthesia, paresthesia, and dysesthesia also may occur. Any combination of these symptoms may develop as neuropathy progresses. Symptoms may include a sensation of numbness, tingling, sharpness, or burning that begins in the feet and spreads proximally. Neuropathic pain develops in some of these individuals, occasionally preceded by improvement in their glycemic control. Pain typically involves the lower extremities, is usually present at rest, and worsens at night. Both an acute (lasting <12 months) and a chronic form of painful diabetic neuropathy have been described. As diabetic neuropathy progresses, the pain subsides and eventually disappears, but a sensory deficit in the lower extremities persists. Physical examination reveals sensory loss, loss of ankle reflexes, and abnormal position sense.

Diabetic polyradiculopathy is a syndrome characterized by severe disabling pain in the distribution of one or more nerve roots. It may be accompanied by motor weakness. Intercostal or truncal radiculopathy causes pain over the thorax or abdomen. Involvement of the lumbar plexus or femoral nerve may cause severe pain in the thigh or hip and may be associated with muscle weakness in the hip flexors or extensors (diabetic amyotrophy). Fortunately, diabetic polyradiculopathies are usually self-limited and resolve over 6–12 months.

Mononeuropathy (dysfunction of isolated cranial or peripheral nerves) is less common than polyneuropathy in DM and presents with pain and motor weakness in the distribution of a single nerve. A vascular etiology has been suggested, but the pathogenesis is unknown. Involvement of the third cranial nerve is most common and is heralded by diplopia. Physical examination reveals ptosis and ophthalmoplegia with normal pupillary constriction to light. Sometimes other cranial nerves IV, VI, or VII (Bell's palsy) are affected. Peripheral mononeuropathies or simultaneous involvement of more than one nerve (mononeuropathy multiplex) may also occur.

AUTONOMIC NEUROPATHY

Individuals with long-standing type 1 or 2 DM may develop signs of autonomic dysfunction involving the cholinergic, noradrenergic, and peptidergic (peptides such as pancreatic polypeptide, substance P, etc.) systems. DM-related autonomic neuropathy can involve multiple systems, including the cardiovascular, gastrointestinal, genitourinary, sudomotor, and metabolic systems. Autonomic neuropathies affecting the cardiovascular system cause a resting tachycardia and orthostatic hypotension. Reports of sudden death have also been attributed to autonomic neuropathy. Gastroparesis and bladder-emptying abnormalities are often caused by the autonomic neuropathy seen in DM (discussed below). Hyperhidrosis of the upper extremities and anhidrosis of the lower extremities result from sympathetic nervous system dysfunction. Anhidrosis of the feet can promote dry skin with cracking, which increases the risk of foot ulcers. Autonomic neuropathy may reduce counterregulatory hormone release (especially catecholamines), leading to an inability to sense hypoglycemia appropriately (hypoglycemia unawareness; Chap. 345), thereby subjecting the patient to the risk of severe hypoglycemia and complicating efforts to improve glycemic control.

TREATMENT: DIABETIC NEUROPATHY

Treatment of diabetic neuropathy is less than satisfactory. Improved glycemic control should be aggressively pursued and will improve nerve conduction velocity, but symptoms of diabetic neuropathy may not necessarily improve. Efforts to improve glycemic control may be confounded by autonomic neuropathy and hypoglycemia unawareness. Risk factors for neuropathy such as hypertension and hypertriglyceridemia should be treated. Avoidance of neurotoxins (alcohol) and smoking, supplementation with vitamins for possible deficiencies (B_{12}, folate; Chap. 74), and symptomatic treatment are the mainstays of therapy. Loss of sensation in the foot places the patient at risk for ulceration and its sequelae; consequently, prevention of such problems is of paramount importance. Patients with symptoms or signs of neuropathy (see "Physical Examination," below) should check their feet daily and take precautions (footwear) aimed at preventing calluses or ulcerations. If foot deformities are present, a podiatrist should be involved.

Chronic, painful diabetic neuropathy is difficult to treat but may respond to antidepressants (tricyclic antidepressants such as amitriptyline, desipramine, nortriptyline, imipramine or selective serotoninnorepinephrine reuptake inhibitors such as duloxetine) or anticonvulsants (gabapentin, pregabalin, carbamazepine, lamotrigine). Two agents, duloxetine and pregabalin, have been approved by the U.S. Food and Drug Administration (FDA) for pain associated with diabetic neuropathy. However, pending further study, most recommend beginning with other agents such as a tricyclic antidepressant and switching if there is no response or if side effects develop. Referral to a pain management center may be necessary. Since the pain of acute diabetic neuropathy may resolve over time, medications may be discontinued as progressive neuronal damage from DM occurs.

Therapy of orthostatic hypotension secondary to autonomic neuropathy is also challenging. A variety of agents have limited success (fludrocortisone, midodrine, clonidine, octreotide, and yohimbine) but each has significant side effects. Nonpharmacologic maneuvers (adequate salt intake, avoidance of dehydration and diuretics, and lower extremity support hose) may offer some benefit.

Gastrointestinal/Genitourinary Dysfunction

Long-standing type 1 and 2 DM may affect the motility and function of gastrointestinal (GI) and genitourinary systems. The most prominent GI symptoms are delayed gastric emptying (gastroparesis) and altered small- and large-bowel motility (constipation or diarrhea). Gastroparesis may present with symptoms of anorexia, nausea, vomiting, early satiety, and abdominal bloating. Microvascular complications (retinopathy and neuropathy) are usually present. Nuclear medicine scintigraphy after ingestion of a radiolabeled meal may document delayed gastric emptying, but may not correlate well with the patient's symptoms. Noninvasive "breath tests" following ingestion of a radiolabeled meal are under development. Though parasympathetic dysfunction secondary to chronic hyperglycemia is important in the development of gastroparesis, hyperglycemia itself also impairs gastric emptying. Nocturnal diarrhea, alternating with constipation, is a feature of DM-related GI autonomic neuropathy. In type 1 DM, these symptoms should also prompt evaluation for celiac sprue because of its increased frequency. Esophageal dysfunction in long-standing DM may occur but is usually asymptomatic.

Diabetic autonomic neuropathy may lead to genitourinary dysfunction including cystopathy, erectile dysfunction, and female sexual dysfunction (reduced sexual desire, dyspareunia, reduced vaginal lubrication). Symptoms of diabetic cystopathy begin with an inability to sense a full bladder and a failure to void completely. As bladder contractility worsens, bladder capacity and the postvoid residual increase, leading to symptoms of urinary hesitancy, decreased voiding frequency, incontinence, and recurrent urinary tract infections. Diagnostic evaluation includes cystometry and urodynamic studies.

Erectile dysfunction and retrograde ejaculation are very common in DM and may be one of the earliest signs of diabetic neuropathy (Chap. 48). Erectile dysfunction, which increases in frequency with the age of the patient and the duration of diabetes, may occur in the absence of other signs of diabetic autonomic neuropathy.

TREATMENT: GASTROINTESTINAL/GENITOURINARY DYSFUNCTION

Current treatments for these complications of DM are inadequate. Improved glycemic control should be a primary goal, as some aspects (neuropathy, gastric function) may improve. Smaller, more frequent meals that are easier to digest (liquid) and low in fat and fiber may minimize symptoms of gastroparesis. Agents with some efficacy include dopamine antagonists metoclopramide, 5–10 mg, and domperidone, 10–20 mg, before each meal. Erythromycin interacts with the motilin receptor and may promote gastric emptying. Diabetic diarrhea in the absence of bacterial overgrowth is treated symptomatically with loperamide and may respond to octreotide (50–75 μg three times daily, SC). Treatment of bacterial overgrowth with antibiotics is sometimes useful (Chap. 294).

Diabetic cystopathy should be treated with timed voiding or self-catheterization, possibly with the addition of bethanechol. Drugs that inhibit type 5 phosphodiesterase are effective for erectile dysfunction, but their efficacy in individuals with DM is slightly lower than in the nondiabetic population (Chap. 48). Sexual dysfunction in women may be improved with use of vaginal lubricants, treatment of vaginal infections, and systemic or local estrogen replacement.

Cardiovascular Morbidity and Mortality

Cardiovascular disease is increased in individuals with type 1 or type 2 DM. The Framingham Heart Study revealed a marked increase in PAD, CHF, CHD, MI, and sudden death (risk increase from one- to fivefold) in DM. The American Heart Association has designated DM as a "CHD risk equivalent." Type 2 diabetes patients without a prior MI have a similar risk for coronary artery–related events as nondiabetic individuals who have had a prior MI. Because of the extremely high prevalence of underlying cardiovascular disease in individuals with diabetes (especially in type 2 DM), evidence of atherosclerotic vascular disease (e.g., cardiac stress test) should be sought in an individual with diabetes who has symptoms suggestive of cardiac ischemia or

peripheral or carotid arterial disease. The screening of asymptomatic individuals with diabetes for CHD is controversial, and recent studies have not shown a clinical benefit. The absence of chest pain ("silent ischemia") is common in individuals with diabetes, and a thorough cardiac evaluation should be considered in individuals undergoing major surgical procedures. The prognosis for individuals with diabetes who have CHD or MI is worse than for nondiabetics. CHD is more likely to involve multiple vessels in individuals with DM.

The increase in cardiovascular morbidity and mortality rates appears to relate to the synergism of hyperglycemia with other cardiovascular risk factors. For example, after controlling for all known cardiovascular risk factors, type 2 DM increases the cardiovascular death rate twofold in men and fourfold in women. Risk factors for macrovascular disease in diabetic individuals include dyslipidemia, hypertension, obesity, reduced physical activity, and cigarette smoking. Additional risk factors more prevalent in the diabetic population include microalbuminuria, macroalbuminuria, an elevation of serum creatinine, and abnormal platelet function. Insulin resistance, as reflected by elevated serum insulin levels, is associated with an increased risk of cardiovascular complications in individuals with and without DM. Individuals with insulin resistance and type 2 DM have elevated levels of plasminogen activator inhibitors (especially PAI-1) and fibrinogen, which enhances the coagulation process and impairs fibrinolysis, thus favoring the development of thrombosis. Diabetes is also associated with endothelial, vascular smooth-muscle, and platelet dysfunction.

Improved glycemic control started soon after the diagnosis of diabetes reduces cardiovascular complications in DM, but the glycemic goal for individuals with long-standing diabetes remains unclear. In both the DCCT (type 1 diabetes) and the UKPDS (type 2 diabetes), cardiovascular events were not reduced by intensive treatment during the trial but were reduced at follow-up 10–17 years later (this effect has been termed *legacy effect* or *metabolic memory*). During the DCCT, an improvement in the lipid profile of individuals in the intensive group (lower total and LDL cholesterol, lower triglycerides) during intensive diabetes management was noted. Trials to examine whether very aggressive glycemic targets (A_{1C} near 6%) reduce cardiovascular events in type 2 diabetes did not show a survival benefit of reducing the A_{1C} below 7% (and in one trial, the outcome was worse). Current recommendations do not suggest more aggressive glucose lowering in this patient population. The possibility of atherogenic potential of insulin is suggested by the data in nondiabetic individuals showing higher serum insulin levels (indicative of insulin resistance) in association with greater risk of cardiovascular morbidity and mortality. However, treatment with insulin and the sulfonylureas did not appear to increase the risk of cardiovascular disease in individuals with type 2 DM, refuting prior claims about the atherogenic potential of these agents.

In addition to CHD, cerebrovascular disease is increased in individuals with DM (threefold increase in stroke). Individuals with DM have an increased incidence of CHF. The etiology of this abnormality is probably multifactorial and includes factors such as myocardial ischemia from atherosclerosis, hypertension, and myocardial cell dysfunction secondary to chronic hyperglycemia.

TREATMENT: CARDIOVASCULAR DISEASE

In general, the treatment of coronary disease is not different in the diabetic individual (Chap. 243). Revascularization procedures for CHD, including percutaneous coronary interventions (PCI) and coronary artery bypass grafting (CABG), may be less efficacious in the diabetic individual. Initial success rates of PCI in diabetic individuals are similar to those in the nondiabetic population, but diabetic patients have higher rates of restenosis and lower long-term patency and survival rates in older studies. More recently, the use of drug-eluting stents and a GPIIb/IIIa platelet inhibitor has improved the outcomes in diabetic patients, and whether there is a difference in efficacy of PCI in diabetic individuals is not clear. Although CABG may be preferred over PCI in diabetic individuals with multivessel coronary artery disease or recent Q-wave MI, PCI is preferred in patients with single-vessel coronary artery disease or two-vessel disease (no involvement of left anterior descending).

The ADA has emphasized the importance of glycemic control and aggressive cardiovascular risk modification in all individuals with DM (see below). Past trepidation about using beta blockers in individuals who have diabetes should not prevent use of these agents since they clearly benefit diabetic patients after MI. ACE inhibitors (or ARBs) may also be particularly beneficial and should be considered in individuals with type 2 DM and other risk factors (smoking, dyslipidemia, history of cardiovascular disease, microalbuminuria). Patients with atypical chest pain or an abnormal resting ECG should be considered for screening for CHD.

Antiplatelet therapy reduces cardiovascular events in individuals with DM who have CHD. Current recommendations by the ADA include the use of aspirin for secondary prevention of coronary events and the consideration of aspirin use in diabetic individuals with an increased cardiovascular risk (based on risk stratification using risk factors such as hypertension, smoking, family history, albuminuria, or dyslipidemia). Data demonstrating efficacy of aspirin in primary prevention of coronary events in individuals with DM and a low risk for CHD are lacking. The aspirin dose (75–162 mg) is the same as that in nondiabetic individuals. Aspirin therapy does not have detrimental effects on renal function or hypertension, nor does it influence the course of diabetic retinopathy.

CARDIOVASCULAR RISK FACTORS

Dyslipidemia

Individuals with DM may have several forms of dyslipidemia (Chap. 356). Because of the additive cardiovascular risk of hyperglycemia and hyperlipidemia, lipid abnormalities should be assessed aggressively and treated as part of comprehensive diabetes care. The most common pattern of dyslipidemia is hypertriglyceridemia and reduced HDL cholesterol levels. DM itself does not increase levels of LDL, but the small dense LDL particles found in type 2 DM are more atherogenic because they are more easily glycated and susceptible to oxidation.

Almost all treatment studies of diabetic dyslipidemia have been performed in individuals with type 2 DM because of the greater frequency of dyslipidemia in this form of diabetes. Interventional studies have shown that the beneficial effects of LDL reduction are similar in the diabetic and nondiabetic populations. Large prospective trials of primary and secondary intervention for CHD have included some individuals with type 2 DM, and subset analyses have consistently found that reductions in LDL reduce cardiovascular events and morbidity in individuals with DM. No prospective studies have addressed similar questions in individuals with type 1 DM. Since the frequency of cardiovascular disease is low in children and young adults with diabetes, assessment of CV risk should be incorporated into the guidelines discussed below.

Based on the guidelines provided by the ADA and the American Heart Association, priorities in the treatment of dyslipidemia are as follows: (1) lower the LDL cholesterol, (2) raise the HDL cholesterol, and (3) decrease the triglycerides. A treatment strategy depends on the pattern of lipoprotein abnormalities. Initial therapy for all forms of dyslipidemia should include dietary changes, as well as the same lifestyle modifications recommended in the nondiabetic population (smoking cessation, blood pressure control, weight loss, increased physical activity). The dietary recommendations for individuals with DM are similar to those advocated by the National Cholesterol Education Program (Chap. 356) and include increased monounsaturated fat and carbohydrates and reduced saturated fats and cholesterol. Though viewed as important, the response to dietary alterations is often modest (<10% reduction in the LDL). Improvement in glycemic control will lower triglycerides and have a modest beneficial effect by raising HDL. HMG-CoA reductase inhibitors are the agents of choice for lowering the LDL. According to guidelines of the ADA and the American Heart Association, the target lipid values in diabetic individuals (age >40 years) without cardiovascular disease should be as follows: LDL < 2.6 mmol/L (100 mg/dL); HDL >1 mmol/L (40 mg/dL) in men and >1.3 mmol/L (50 mg/dL) in women; and triglycerides <1.7 mmol/L (150 mg/dL). In patients >40 years, the ADA recommends addition

of a statin, regardless of the LDL level in patients with CHD and those without CHD, but who have CHD risk factors.

If the patient is known to have CHD, the ADA recommends an LDL goal of <1.8 mmol/L (70 mg/dL) as an "option" [in keeping with evidence that such a goal is beneficial in nondiabetic individuals with CHD (Chap. 356)]. Older studies with fibrates indicated efficacy, but recent trials have not shown a benefit of this class of agents. Combination therapy with an HMG-CoA reductase inhibitor and a fibrate or another lipid-lowering agent (ezetimibe, niacin) may be considered to reach LDL goals, but statin/fibrate combinations increase the possibility of side effects such as myositis. Nicotinic acid effectively raises HDL and can be used in patients with diabetes, but high doses (>2 g/d) may worsen glycemic control and increase insulin resistance. Bile acid–binding resins should not be used if hypertriglyceridemia is present.

Hypertension

Hypertension can accelerate other complications of DM, particularly cardiovascular disease and nephropathy. In targeting a goal of BP <130/80 mmHg, therapy should first emphasize life-style modifications such as weight loss, exercise, stress management, and sodium restriction. Realizing that more than one agent is usually required to reach the blood pressure goal, the ADA recommends that all patients with diabetes and hypertension be treated with an ACE inhibitor or an ARB. Subsequently, agents that reduce cardiovascular risk (beta blockers, thiazide diuretics, and calcium channel blockers) should be incorporated into the regimen. While ACE inhibitors and ARBs are likely equivalent in most patients with diabetes and renal disease, the ADA notes (1) in patients with type 1 diabetes, hypertension, and micro- or macroalbuminuria, an ACE inhibitor slowed progression of nephropathy; (2) an ACE inhibitor or an ARB slowed the progression to macroalbuminuria in patients with type 2 diabetes, hypertension, and microalbumin-uria; and (3) ARB slowed the decline in GFR in patients with type 2 diabetes, hypertension, macroalbuminuria, and renal insufficiency. Additional points of emphasis include the following:

1. ACE inhibitors are either glucose- and lipid-neutral or glucose- and lipid-beneficial and thus positively impact the cardiovascular risk profile. Calcium channel blockers, central adrenergic antagonists, and vasodilators are lipid- and glucose-neutral.
2. Beta blockers and thiazide diuretics can increase insulin resistance and negatively impact the lipid profile; beta blockers may slightly increase the risk of developing type 2 DM. Beta blockers are safe in patients with diabetes and reduce cardiovascular events.
3. Sympathetic inhibitors and α-adrenergic blockers may worsen orthostatic hypotension in the diabetic individual with autonomic neuropathy.
4. Equivalent reduction in blood pressure by different classes of agents may not translate into equivalent protection from cardiovascular and renal endpoints. Thiazides, beta blockers, ACE inhibitors, and ARBs positively impact cardiovascular endpoints (MI or stroke).
5. Serum potassium and renal function should be monitored.

Because of the high prevalence of atherosclerotic disease in individuals with type 2 DM, the possibility of renovascular hypertension should be considered when the blood pressure is not readily controlled.

Lower Extremity Complications

DM is the leading cause of nontraumatic lower extremity amputation in the United States. Foot ulcers and infections are also a major source of morbidity in individuals with DM. The reasons for the increased incidence of these disorders in DM involve the interaction of several pathogenic factors: neuropathy, abnormal foot biomechanics, PAD, and poor wound healing. The peripheral sensory neuropathy interferes with normal protective mechanisms and allows the patient to sustain major or repeated minor trauma to the foot, often without knowledge of the injury. Disordered proprioception causes abnormal weight bearing while walking and subsequent formation of callus or ulceration. Motor and sensory neuropathy lead to abnormal foot muscle mechanics and to structural changes in the foot (hammertoe, claw toe deformity, prominent metatarsal heads, Charcot joint). Autonomic neuropathy results in anhidrosis and altered superficial blood flow in the foot, which promote drying of the skin and fissure formation. PAD and poor wound healing impede resolution of minor breaks in the skin, allowing them to enlarge and to become infected.

Approximately 15% of individuals with type 2 DM develop a foot ulcer (great toe or MTP areas are most common), and a significant subset will ultimately undergo amputation (14–24% risk with that ulcer or subsequent ulceration). Risk factors for foot ulcers or amputation include: male sex, diabetes >10 years' duration, peripheral neuropathy, abnormal structure of foot (bony abnormalities, callus, thickened nails), peripheral arterial disease, smoking, history of previous ulcer or amputation, and poor glycemic control. Large calluses are often precursors to or overlie ulcerations.

TREATMENT: LOWER EXTREMITY COMPLICATIONS

The optimal therapy for foot ulcers and amputations is prevention through identification of high-risk patients, education of the patient, and institution of measures to prevent ulceration. High-risk patients should be identified during the routine foot examination performed on all patients with DM (see "Ongoing Aspects of Comprehensive Diabetes Care," below). Patient education should emphasize (1) careful selection of footwear, (2) daily inspection of the feet to detect early signs of poor-fitting footwear or minor trauma, (3) daily foot hygiene to keep the skin clean and moist, (4) avoidance of self-treatment of foot abnormalities and high-risk behavior (e.g., walking barefoot), and (5) prompt consultation with a health care provider if an abnormality arises. Patients at high risk for ulceration or amputation may benefit from evaluation by a foot care specialist. Interventions directed at risk factor modification include orthotic shoes and devices, callus management, nail care, and prophylactic measures to reduce increased skin pressure from abnormal bony architecture. Attention to other risk factors for vascular disease (smoking, dyslipidemia, hypertension) and improved glycemic control are also important.

Despite preventive measures, foot ulceration and infection are common and represent a serious problem. Due to the multifactorial pathogenesis of lower extremity ulcers, management of these lesions is multidisciplinary and often demands expertise in orthopedics, vascular surgery, endocrinology, podiatry, and infectious diseases. The plantar surface of the foot is the most common site of ulceration. Ulcers may be primarily neuropathic (no accompanying infection) or may have surrounding cellulitis or osteomyelitis. Cellulitis without ulceration is also frequent and should be treated with antibiotics that provide broad-spectrum coverage, including anaerobes (see below).

An infected ulcer is a clinical diagnosis, since superficial culture of any ulceration will likely find multiple possible bacterial species. The infection surrounding the foot ulcer is often the result of multiple organisms (gram-positive and -negative organisms and anaerobes), and gas gangrene may develop in the absence of clostridial infection. Cultures taken from the surface of the ulcer are not helpful; a culture from the debrided ulcer base or from purulent drainage or aspiration of the wound is the most helpful. Wound depth should be determined by inspection and probing with a blunt-tipped sterile instrument. Plain radiographs of the foot should be performed to assess the possibility of osteomyelitis in chronic ulcers that have not responded to therapy. Nuclear medicine bone scans may be helpful, but overlying subcutaneous infection is often difficult to distinguish from osteomyelitis. Indium-labeled white cell studies are more useful in determining if the infection involves bony structures or only soft tissue, but they are technically demanding. MRI of the foot may be the most specific modality, although distinguishing bony destruction due to osteomyelitis from destruction secondary to Charcot arthropathy is difficult. If surgical debridement is

necessary, bone biopsy and culture may provide the answer.

Osteomyelitis is best treated by a combination of prolonged antibiotics (IV, then oral) and possibly debridement of infected bone. The possible contribution of vascular insufficiency should be considered in all patients. Noninvasive blood-flow studies are often unreliable in DM, and angiography may be required, recognizing the risk of contrast-induced nephrotoxicity. Peripheral arterial bypass procedures are often effective in promoting wound healing and in decreasing the need for amputation of the ischemic limb.

A growing number of possible treatments for diabetic foot ulcers exist, but they have yet to demonstrate clear efficacy in prospective, controlled trials. A consensus statement from the ADA identified six interventions with demonstrated efficacy in diabetic foot wounds: (1) off-loading, (2) debridement, (3) wound dressings, (4) appropriate use of antibiotics, (5) revascularization, and (6) limited amputation. Off-loading is the complete avoidance of weight bearing on the ulcer, which removes the mechanical trauma that retards wound healing. Bed rest and a variety of orthotic devices or contact casting limit weight bearing on wounds or pressure points. Surgical debridement is important and effective, but clear efficacy of other modalities for wound cleaning (enzymes, soaking, whirlpools) is lacking. Dressings such as hydrocolloid dressings promote wound healing by creating a moist environment and protecting the wound. Antiseptic agents should be avoided. Topical antibiotics are of limited value. Referral for physical therapy, orthotic evaluation, and rehabilitation should occur once the infection is controlled.

Mild or non-limb-threatening infections can be treated with oral antibiotics (cephalosporin, clindamycin, amoxicillin/clavulanate, and fluoroquinolones), surgical debridement of necrotic tissue, local wound care (avoidance of weight bearing over the ulcer), and close surveillance for progression of infection. More severe ulcers may require IV antibiotics as well as bed rest and local wound care. Urgent surgical debridement may be required. Strict control of glycemia should be a goal (see below). IV antibiotics should provide broad-spectrum coverage directed toward *Staphylococcus aureus*, streptococci, gram-negative aerobes, and anaerobic bacteria. Initial antimicrobial regimens include ertapenem, piperacillin/tazobactam, cefotetan, ampicillin/sulbactam, linezolid, or the combination of clindamycin and a fluoroquinolone. Severe infections, or infections that do not improve after 48 h of antibiotic therapy, require expansion of antimicrobial therapy to treat methicillin-resistant *S. aureus* (e.g., vancomycin) and *Pseudomonas aeruginosa*. If the infection surrounding the ulcer is not improving with IV antibiotics, reassessment of antibiotic coverage and reconsideration of the need for surgical debridement or revascularization are indicated. With clinical improvement, oral antibiotics and local wound care can be continued on an outpatient basis with close follow-up.

New information about wound biology has led to a number of new technologies (e.g., living skin equivalents and growth factors) that may prove useful, especially in neuropathic ulcers. Hyperbaric oxygen has been used, but rigorous proof of efficacy is lacking. Negative wound pressure has been shown to accelerate wound healing of plantar wounds.

Infections

Individuals with DM have a greater frequency and severity of infection. The reasons for this include incompletely defined abnormalities in cell-mediated immunity and phagocyte function associated with hyperglycemia, as well as diminished vascularization. Hyperglycemia aids the colonization and growth of a variety of organisms (*Candida* and other fungal species). Many common infections are more frequent and severe in the diabetic population, whereas several rare infections are seen almost exclusively in the diabetic population. Examples of this latter category include rhinocerebral mucormycosis, emphysematous infections of the gall bladder and urinary tract, and "malignant" or invasive otitis externa. Invasive otitis externa is usually secondary to *P. aeruginosa* infection in the soft tissue surrounding the external auditory canal, usually begins with pain and discharge, and may rapidly progress to osteomyelitis and meningitis. These infections should be sought, in particular, in patients presenting with HHS.

Pneumonia, urinary tract infections, and skin and soft tissue infections are all more common in the diabetic population. In general, the organisms that cause pulmonary infections are similar to those found in the nondiabetic population; however, gram-negative organisms, *S. aureus*, and *Mycobacterium tuberculosis* are more frequent pathogens. Urinary tract infections (either lower tract or pyelonephritis) are the result of common bacterial agents such as *Escherichia coli*, though several yeast species (*Candida* and *Torulopsis glabrata*) are commonly observed. Complications of urinary tract infections include emphysematous pyelonephritis and emphysematous cystitis. Bacteriuria occurs frequently in individuals with diabetic cystopathy. Susceptibility to furunculosis, superficial candidal infections, and vulvovaginitis are increased. Poor glycemic control is a common denominator in individuals with these infections. Diabetic individuals have an increased rate of colonization of *S. aureus* in the skinfolds and nares. Diabetic patients also have a greater risk of postoperative wound infections. Strict glycemic control reduces postoperative infections in diabetic individuals undergoing CABG and should be the goal in all diabetic patients with an infection.

Dermatologic Manifestations

The most common skin manifestations of DM are protracted wound healing and skin ulcerations. Diabetic dermopathy, sometimes termed *pigmented pretibial papules*, or "diabetic skin spots," begins as an erythematous area and evolves into an area of circular hyperpigmentation. These lesions result from minor mechanical trauma in the pretibial region and are more common in elderly men with DM. Bullous diseases, such as bullosa diabeticorum (shallow ulcerations or erosions in the pretibial region), are also seen. *Necrobiosis lipoidica diabeticorum* is a rare disorder of DM that predominantly affects young women with type 1 DM, neuropathy, and retinopathy. It usually begins in the pretibial region as an erythematous plaque or papules that gradually enlarge, darken, and develop irregular margins, with atrophic centers and central ulceration. They may be painful. Vitiligo occurs at increased frequency in individuals with type 1 diabetes. *Acanthosis nigricans* (hyperpigmented velvety plaques seen on the neck, axilla, or extensor surfaces) is sometimes a feature of severe insulin resistance and accompanying diabetes. Generalized or localized *granuloma annulare* (erythematous plaques on the extremities or trunk) and *scleredema* (areas of skin thickening on the back or neck at the site of previous superficial infections) are more common in the diabetic population. *Lipoatrophy* and *lipohypertrophy* can occur at insulin injection sites but are now unusual with the use of human insulin. Xerosis and pruritus are common and are relieved by skin moisturizers.

Approach to the Patient: Diabetes Mellitus

DM and its complications produce a wide range of symptoms and signs; those secondary to acute hyperglycemia may occur at any stage of the disease, whereas those related to chronic complications begin to appear during the second decade of hyperglycemia. Individuals with previously undetected type 2 DM may present with chronic complications of DM at the time of diagnosis. The history and physical examination should assess for symptoms or signs of acute hyperglycemia and should screen for the chronic complications and conditions associated with DM.

HISTORY

A complete medical history should be obtained with special emphasis on DM-relevant aspects such as weight, family history of DM and its complications, risk factors for cardiovascular disease, exercise, smoking, and ethanol use. Symptoms of hyperglycemia include polyuria, polydipsia, weight loss, fatigue, weakness, blurry vision, frequent superficial infections (vaginitis, fungal skin infections), and slow healing of skin lesions after minor trauma. Metabolic derangements relate mostly to hyperglycemia (osmotic diuresis) and to the catabolic state of the patient (urinary loss of glucose and calories, muscle breakdown due to protein degradation and decreased protein synthesis). Blurred vision results from changes in the water

content of the lens and resolves as the hyperglycemia is controlled.

In a patient with established DM, the initial assessment should also include special emphasis on prior diabetes care, including the type of therapy, prior A1C levels, self-monitoring blood glucose results, frequency of hypoglycemia, presence of DM-specific complications, and assessment of the patient's knowledge about diabetes, exercise, and nutrition. The chronic complications may afflict several organ systems, and an individual patient may exhibit some, all, or none of the symptoms related to the complications of DM (see above). In addition, the presence of DM-related comorbidities should be sought (cardiovascular disease, hypertension, dyslipidemia).

PHYSICAL EXAMINATION

In addition to a complete physical examination, special attention should be given to DM-relevant aspects such as weight or BMI, retinal examination, orthostatic blood pressure, foot examination, peripheral pulses, and insulin injection sites. Blood pressure >130/80 mmHg is considered hypertension in individuals with diabetes. Careful examination of the lower extremities should seek evidence of peripheral arterial disease (pedal pulses), peripheral neuropathy, calluses, superficial fungal infections, nail disease, ankle reflexes, and foot deformities (such as hammertoes or claw toes and Charcot foot) in order to identify sites of potential skin ulceration. Vibratory sensation (128-MHz tuning fork at the base of the great toe), the ability to sense touch with a monofilament (5.07, 10-g monofilament), pinprick sensation, testing for ankle reflexes, and vibration perception threshold (using a biothesiometer) are used to detect moderately advanced diabetic neuropathy. Since periodontal disease is more frequent in DM, the teeth and gums should also be examined.

CLASSIFICATION OF DM IN AN INDIVIDUAL PATIENT

The etiology of diabetes in an individual with new-onset disease can usually be assigned on the basis of clinical criteria. Individuals with type 1 DM tend to have the following characteristics: (1) onset of disease prior to age 30 years; (2) lean body habitus; (3) requirement of insulin as the initial therapy; (4) propensity to develop ketoacidosis; and (5) an increased risk of other autoimmune disorders such as autoimmune thyroid disease, adrenal insufficiency, pernicious anemia, celiac disease, and vitiligo. In contrast, individuals with type 2 DM often exhibit the following features: (1) develop diabetes after the age of 30 years; (2) are usually obese (80% are obese, but elderly individuals may be lean); (3) may not require insulin therapy initially; and (4) may have associated conditions such as insulin resistance, hypertension, cardiovascular disease, dyslipidemia, or PCOS. In type 2 DM, insulin resistance is often associated with abdominal obesity (as opposed to hip and thigh obesity) and hypertriglyceridemia. Although most individuals diagnosed with type 2 DM are older, the age of diagnosis is declining, and there is a marked increase among overweight children and adolescents. Some individuals with phenotypic type 2 DM present with DKA but lack autoimmune markers and may be later treated with oral glucose-lowering agents rather than insulin (this clinical picture is sometimes referred to as *ketosis-prone type 2 DM*). On the other hand, some individuals (5–10%) with the phenotypic appearance of type 2 DM do not have absolute insulin deficiency but have autoimmune markers (ICA, GAD autoantibodies) suggestive of type 1 DM (termed *latent autoimmune diabetes of the adult*). Such individuals are more likely to be <50 years of age, have a normal BMI, and have a personal or family history of other autoimmune disease. They are much more likely to require insulin treatment within 5 years. Monogenic forms of diabetes (discussed above) should be considered in those with diabetes onset <30 years of age, an autosomal pattern of diabetes inheritance, and the lack of nearly complete insulin deficiency. Despite recent advances in the understanding of the pathogenesis of diabetes, it remains difficult to categorize some patients unequivocally. Individuals who deviate from the clinical profile of type 1 and type 2 DM, or who have other associated defects such as deafness, pancreatic exocrine disease, and other endocrine disorders, should be classified accordingly (Table 344-1).

LABORATORY ASSESSMENT

The laboratory assessment should first determine whether the patient meets the diagnostic criteria for DM (Table 344-2) and then assess the degree of glycemic control (A1C, discussed below). In addition to the standard laboratory evaluation, the patient should be screened for DM-associated conditions (e.g., microalbuminuria, dyslipid-emia, thyroid dysfunction). Individuals at high risk for cardiovascular disease should be screened for asymptomatic CHD by appropriate cardiac stress testing, when indicated.

The classification of the type of DM may be facilitated by laboratory assessments. Serum insulin or C-peptide measurements do not always distinguish type 1 from type 2 DM, but a low C-peptide level confirms a patient's need for insulin. Many individuals with new-onset type 1 DM retain some C-peptide production. Measurement of islet cell antibodies at the time of diabetes onset may be useful if the type of DM is not clear based on the characteristics described above.

LONG-TERM TREATMENT

Overall Principles

The goals of therapy for type 1 or type 2 DM are to (1) eliminate symptoms related to hyperglycemia, (2) reduce or eliminate the long-term microvascular and macrovascular complications of DM, and (3) allow the patient to achieve as normal a lifestyle as possible. To reach these goals, the physician should identify a target level of glycemic control for each patient, provide the patient with the educational and pharmacologic resources necessary to reach this level, and monitor/treat DM-related complications. Symptoms of diabetes usually resolve when the plasma glucose is <11.1 mmol/L (200 mg/dL), and thus most DM treatment focuses on achieving the second and third goals. The treatment goals for patients with diabetes are summarized in **Table 344-8**.

Table 344-8 Treatment Goals for Adults with Diabetes[a]

Index	Goal
Glycemic control[b]	
A1C	<7.0%[c]
Preprandial capillary plasma glucose	3.9–7.2 mmol/L (70–130 mg/dL)
Peak postprandial capillary plasma glucose[d]	<10.0 <1.7 mmol/L (<180 mg/dL)
Blood pressure	<130/80
Lipids[e]	
Low-density lipoprotein	<2.6 mmol/L (100 mg/dL)
High-density lipoprotein	>1 mmol/L (40 mg/dL) in men >1.3 mmol/L (50 mg/dL) in women
Triglycerides	<1.7 mmol/L (150 mg/dL)

[a]As recommended by the ADA; goals should be individualized for each patient (see text). Goals may be different for certain patient populations. [b]A1C is primary goal. [c]Normal range for A1C: 4.0–6.0% (DCCT-based assay). [d]One–two hours after beginning of a meal. [e]In decreasing order of priority. ***Source:*** Adapted from American Diabetes Association, 2011.

The care of an individual with either type 1 or type 2 DM requires a multidisciplinary team. Central to the success of this team are the patient's participation, input, and enthusiasm, all of which are essential for optimal diabetes management. Members of the health care team include the primary care provider and/or the endocrinologist or diabetologist, a certified diabetes educator, and a nutritionist. In addition, when the complications of DM arise, subspecialists (including neurologists, nephrologists, vascular surgeons, cardiologists, ophthalmologists, and podiatrists) with experience in DM-related complications are essential.

A number of names are sometimes applied to different approaches to diabetes care, such as intensive insulin therapy, intensive glycemic control, and "tight control." The current chapter, and other sources, use the term *comprehensive diabetes care* to emphasize the fact that optimal diabetes therapy involves more than plasma glucose management. Though glycemic control is central to optimal diabetes therapy, comprehensive diabetes care of both type 1 and type 2 DM should also detect and manage DM-specific complications and modify risk factors for DM-associated diseases. In addition to the physical aspects of DM, social, family, financial, cultural, and employment-related issues may impact diabetes care. The International Diabetes Federation (IDF), recognizing that resources available for diabetes care varies widely throughout the world, has issued guidelines for standard care (a well-developed service base and with health care funding systems consuming a significant part of their national wealth), minimal care (health care settings with very limited resources), and comprehensive care (health care settings with considerable resources). This chapter provides guidance for this comprehensive level of diabetes care.

Patient Education About DM, Nutrition, and Exercise

The patient with type 1 or type 2 DM should receive education about nutrition, exercise, care of diabetes during illness, and medications to lower the plasma glucose. Along with improved compliance, patient education allows individuals with DM to assume greater responsibility for their care. Patient education should be viewed as a continuing process with regular visits for reinforcement; it should not be a process that is completed after one or two visits to a nurse educator or nutritionist. The ADA refers to education about the individualized management plan for the patient as diabetes self-management education (DSME). More frequent contact between the patient and the diabetes management team (electronic, telephone, etc.) improves glycemic control.

DIABETES EDUCATION

The diabetes educator is a health care professional (nurse, dietician, or pharmacist) with specialized patient education skills who is certified in diabetes education (e.g., American Association of Diabetes Educators). Education topics important for optimal diabetes care include self-monitoring of blood glucose; urine ketone monitoring (type 1 DM); insulin administration; guidelines for diabetes management during illnesses; prevention and management of hypoglycemia (Chap. 345); foot and skin care; diabetes management before, during, and after exercise; and risk factor–modifying activities.

NUTRITION

Medical nutrition therapy (MNT) is a term used by the ADA to describe the optimal coordination of caloric intake with other aspects of diabetes therapy (insulin, exercise, weight loss). Primary prevention measures of MNT are directed at preventing or delaying the onset of type 2 DM in high-risk individuals (obese or with prediabetes) by promoting weight reduction. Medical treatment of obesity is a rapidly evolving area and is discussed in Chap. 78. Secondary prevention measures of MNT are directed at preventing or delaying diabetes-related complications in diabetic individuals by improving glycemic control. Tertiary prevention measures of MNT are directed at managing diabetes-related complications (cardiovascular disease, nephropathy) in diabetic individuals. For example, in individuals with diabetes and chronic kidney disease, protein intake should be limited to 0.8 g/kg of body weight per day. MNT in patients with diabetes and cardiovascular disease should incorporate dietary principles used in nondiabetic patients with cardiovascular disease. While the recommendations for all three types of MNT overlap, this chapter emphasizes secondary prevention measures of MNT. Pharmacologic approaches that facilitate weight loss and bariatric surgery should be considered in selected patients (Chap. 78).

In general, the components of optimal MNT are similar for individuals with type 1 or type 2 DM and similar to those for the general population (fruits, vegetables, fiber-containing foods, and low fat; **Table 344-9**). MNT education is an important component of comprehensive diabetes care and should be reinforced by regular patient education. Historically, nutrition education imposed restrictive, complicated regimens on the patient. Current practices have greatly changed, though many patients and health care providers still view the diabetic diet as monolithic and static. For example, MNT now includes foods with sucrose and seeks to modify other risk factors such as hyperlipidemia and hypertension rather than focusing exclusively on weight loss in individuals with type 2 DM. The *glycemic index* is an estimate of the postprandial rise in the blood glucose when a certain amount of that food is consumed. Consumption of foods with a low glycemic index appears to reduce postprandial glucose excursions and improve glycemic control. Reduced calorie and nonnutritive sweeteners are useful. Currently, evidence does not support supplementation of the diet with vitamins, antioxidants (vitamin C and E), or micronutrients (chromium) in patients with diabetes.

Table 344-9 Nutritional Recommendations for Adults with Diabetes[a]

Weight loss diet (in prediabetes and type 2 DM)
• Hypocaloric diet that is low-fat or low-carbohydrate
Fat in diet
• Minimal *trans* fat consumption
Carbohydrate in diet
• Monitor carbohydrate intake in regards to calories
• Sucrose-containing foods may be consumed with adjustments in insulin dose
• Amount of carbohydrate determined by estimating grams of carbohydrate in diet for (type 1 DM)
• Glycemic index reflects how consumption of a particular food affects the blood glucose
Protein in diet: as part of an optimal diet
Other components
• Nonnutrient sweeteners
• Routine supplements of vitamins, antioxidants, or trace elements not advised

[a]See text for differences for patients with type 1 or type 2 diabetes. As for the general population, a healthy diet includes fruits,

vegetables, and fiber-containing foods.

Source: Adapted from American Diabetes Association, 2011.

The goal of MNT in the individual with type 1 DM is to coordinate and match the caloric intake, both temporally and quantitatively, with the appropriate amount of insulin. MNT in type 1 DM and self-monitoring of blood glucose must be integrated to define the optimal insulin regimen. The ADA encourages patients and providers to utilize carbohydrate counting or exchange systems to estimate the nutrient content of a meal or snack. Based on the patient's estimate of the carbohydrate content of a meal, an insulin-to-carbohydrate ratio determines the bolus insulin dose for a meal or snack. MNT must be flexible enough to allow for exercise, and the insulin regimen must allow for deviations in caloric intake. An important component of MNT in type 1 DM is to minimize the weight gain often associated with intensive diabetes management.

The goals of MNT in type 2 DM should focus on weight loss and address the greatly increased prevalence of cardiovascular risk factors (hypertension, dyslipidemia, obesity) and disease in this population. The majority of these individuals are obese, and weight loss is strongly encouraged and should remain an important goal. Hypocaloric diets and modest weight loss (5–7%) often result in rapid and dramatic glucose lowering in individuals with new-onset type 2 DM. Nevertheless, numerous studies document that long-term weight loss is uncommon. MNT for type 2 DM should emphasize modest caloric reduction (low-carbohydrate or low-fat), reduced fat intake, and increased physical activity. Increased consumption of soluble, dietary fiber may improve glycemic control in individuals with type 2 DM. Weight loss and exercise improve insulin resistance.

EXERCISE

Exercise has multiple positive benefits including cardiovascular risk reduction, reduced blood pressure, maintenance of muscle mass, reduction in body fat, and weight loss. For individuals with type 1 or type 2 DM, exercise is also useful for lowering plasma glucose (during and following exercise) and increasing insulin sensitivity. In patients with diabetes, the ADA recommends 150 min/week (distributed over at least 3 days) of moderate aerobic physical activity. The exercise regimen should also include resistance training.

Despite its benefits, exercise presents challenges for individuals with DM because they lack the normal glucoregulatory mechanisms (normally, insulin falls and glucagon rises during exercise). Skeletal muscle is a major site for metabolic fuel consumption in the resting state, and the increased muscle activity during vigorous, aerobic exercise greatly increases fuel requirements. Individuals with type 1 DM are prone to either hyperglycemia or hypoglycemia during exercise, depending on the pre-exercise plasma glucose, the circulating insulin level, and the level of exercise-induced catecholamines. If the insulin level is too low, the rise in catecholamines may increase the plasma glucose excessively, promote ketone body formation, and possibly lead to ketoacidosis. Conversely, if the circulating insulin level is excessive, this relative hyperinsulinemia may reduce hepatic glucose production (decreased glycogenolysis, decreased gluconeogenesis) and increase glucose entry into muscle, leading to hypoglycemia.

To avoid exercise-related hyper- or hypoglycemia, individuals with type 1 DM should (1) monitor blood glucose before, during, and after exercise; (2) delay exercise if blood glucose is >14 mmol/L (250 mg/dL) and ketones are present; (3) if the blood glucose is <5.6 mmol/L (100 mg/dL), ingest carbohydrate before exercising; (3) monitor glucose during exercise and ingest carbohydrate to prevent hypoglycemia; (4) decrease insulin doses (based on previous experience) before exercise and inject insulin into a non-exercising area; and (5) learn individual glucose responses to different types of exercise and increase food intake for up to 24 h after exercise, depending on intensity and duration of exercise. In individuals with type 2 DM, exercise-related hypoglycemia is less common but can occur in individuals taking either insulin or insulin secretagogues.

Because asymptomatic cardiovascular disease appears at a younger age in both type 1 and type 2 DM, formal exercise tolerance testing may be warranted in diabetic individuals with any of the following: age >35 years, diabetes duration >15 years (type 1 DM) or >10 years (type 2 DM), microvascular complications of DM (retinopathy, microalbuminuria, or nephropathy), PAD, other risk factors of CHD, or autonomic neuropathy. Untreated proliferative retinopathy is a relative contraindication to vigorous exercise, as this may lead to vitreous hemorrhage or retinal detachment.

Monitoring the Level of Glycemic Control

Optimal monitoring of glycemic control involves plasma glucose measurements by the patient and an assessment of long-term control by the physician (measurement of hemoglobin A1C and review of the patient's self-measurements of plasma glucose). These measurements are complementary: the patient's measurements provide a picture of short-term glycemic control, whereas the A1C reflects average glycemic control over the previous 2–3 months.

SELF-MONITORING OF BLOOD GLUCOSE

Self-monitoring of blood glucose (SMBG) is the standard of care in diabetes management and allows the patient to monitor his or her blood glucose at any time. In SMBG, a small drop of blood and an easily detectable enzymatic reaction allow measurement of the capillary plasma glucose. Many glucose monitors can rapidly and accurately measure glucose (calibrated to provide plasma glucose value even though blood glucose is measured) in small amounts of blood (3–10 μL) obtained from the fingertip; alternative testing sites (e.g., forearm) are less reliable, especially when the blood glucose is changing rapidly (postprandially). A large number of blood glucose monitors are available, and the certified diabetes educator is critical in helping the patient select the optimal device and learn to use it properly. By combining glucose measurements with diet history, medication changes, and exercise history, the diabetes management team and patient can improve the treatment program.

The frequency of SMBG measurements must be individualized and adapted to address the goals of diabetes care. Individuals with type 1 DM or individuals with type 2 DM taking multiple insulin injections each day should routinely measure their plasma glucose three or more times per day to estimate and select mealtime boluses of short-acting insulin and to modify long-acting insulin doses. Most individuals with type 2 DM require less frequent monitoring, though the optimal frequency of SMBG has not been clearly defined. Individuals with type 2 DM who are taking insulin should utilize SMBG more frequently than those on oral agents. Individuals with type 2 DM who are on oral medications should utilize SMBG as a means of assessing the efficacy of their medication and the impact of diet. Since plasma glucose levels fluctuate less in these individuals, one to two SMBG measurements per day (or fewer in patients who are on oral agents or are diet-controlled) may be sufficient. Most measurements in individuals with type 1 or type 2 DM should be performed prior to a meal and supplemented with postprandial measurements to assist in reaching postprandial glucose targets (Table 344-8).

Devices for continuous blood glucose monitoring (CGM) have been approved by the FDA, and others are in various stages of development. These devices do not replace the need for traditional glucose measurements. This rapidly evolving technology requires substantial expertise on the part of the diabetes management team and the patient. Current continuous glucose-monitoring systems measure the glucose in interstitial fluid, which is in equilibrium with the blood glucose. These devices provide useful short-term information about the patterns of glucose changes as well as an enhanced ability to detect hypoglycemic episodes. Alarms notify the patient if the blood glucose falls into the hypoglycemic range. Clinical experience with these devices is rapidly growing, and they are most useful in individuals with hypoglycemia unawareness, frequent hypoglycemia, or those who have not achieved glycemic targets despite major efforts. The utility of CGM in the ICU setting remains to be determined.

Ketones are an indicator of early diabetic ketoacidosis and should be measured in individuals with type 1 DM when the plasma

glucose is consistently >16.7 mmol/L (300 mg/dL) during a concurrent illness or with symptoms such as nausea, vomiting, or abdominal pain. Blood measurement of β-hydroxybutyrate is preferred over urine testing with nitroprusside-based assays that measure only acetoacetate and acetone.

ASSESSMENT OF LONG-TERM GLYCEMIC CONTROL

Measurement of glycated hemoglobin is the standard method for assessing long-term glycemic control. When plasma glucose is consistently elevated, there is an increase in nonenzymatic glycation of hemoglobin; this alteration reflects the glycemic history over the previous 2–3 months, since erythrocytes have an average life span of 120 days (glycemic level in the preceding month contributes about 50% to the A1C value). There are numerous laboratory methods for measuring the various forms of glycated hemoglobin, and these have significant interassay variations; assays that are calibrated against the DCCT A1C assay are essential. Depending on the assay methodology, hemoglobinopathies, anemias, reticulocytosis, transfusions, and uremia may interfere with the A1C result. Measurement of A1C at the "point of care" allows for more rapid feedback and may therefore assist in adjustment of therapy.

A_{1C} should be measured in all individuals with DM during their initial evaluation and as part of their comprehensive diabetes care. As the primary predictor of long-term complications of DM, the A_{1C} should mirror, to a certain extent, the short-term measurements of SMBG. These two measurements are complementary in that recent intercurrent illnesses may impact the SMBG measurements but not the A_{1C}. Likewise, postprandial and nocturnal hyperglycemia may not be detected by the SMBG of fasting and preprandial capillary plasma glucose but will be reflected in the A_{1C}. In standardized assays, the A_{1C} approximates the following mean plasma glucose values: an A_{1C} of 6% = 7.0 mmol/L (126 mg/dL), 7% = 8.6 mmol/L (154 mg/dL), 8% = 10.2 mmol/L (183 mg/dL), 9% = 11.8 mmol/L (212 mg/dL), 10% = 13.4 mmol/L (240 mg/dL), 11% = 14.9 mmol/L (269 mg/dL), and 12% = 16.5 mmol/L (298 mg/dL). In patients achieving their glycemic goal, the ADA recommends measurement of the A_{1C} at least twice per year. More frequent testing (every 3 months) is warranted when glycemic control is inadequate or when therapy has changed. The degree of glycation of other proteins, such as albumin, can be used as an alternative indicator of glycemic control when the A_{1C} is inaccurate (hemolytic anemia, hemoglobinopathies). The fructosamine assay (measuring glycated albumin) reflects the glycemic status over the prior 2 weeks. Alternative assays of glycemic control should not be routinely used since studies demonstrating that it accurately predicts the complications of DM are lacking.

Treatment: Type 1 and Type 2 Diabetes Mellitus

ESTABLISHMENT OF TARGET LEVEL OF GLYCEMIC CONTROL

Because the complications of DM are related to glycemic control, normoglycemia or near-normoglycemia is the desired, but often elusive, goal for most patients. However, normalization of the plasma glucose for long periods of time is extremely difficult, as demonstrated by the DCCT. Regardless of the level of hyperglycemia, improvement in glycemic control will lower the risk of diabetes complications (Fig. 344-8).

The target for glycemic control (as reflected by the A_{1C}) must be individualized, and the goals of therapy should be developed in consultation with the patient after considering a number of medical, social, and lifestyle issues. Some important factors to consider include the patient's age, ability to understand and implement a complex treatment regimen, presence and severity of complications of diabetes, ability to recognize hypoglycemic symptoms, presence of other medical conditions or treatments that might alter the response to therapy, lifestyle and occupation (e.g., possible consequences of experiencing hypoglycemia on the job), and level of support available from family and friends.

The ADA suggests that the glycemic goal is to achieve an A_{1C} as close to normal as possible without significant hypoglycemia. In general, the target A_{1C} should be <7% (Table 344-8) with a more stringent target for some patients. A higher A_{1C} goal may be appropriate for the very young or old or in individuals with limited life span or comorbid conditions. The major consideration is the frequency and severity of hypoglycemia, since this becomes more common with a more stringent A_{1C} goal. More stringent glycemic control (A1C of 6% or less) is not beneficial, and may be deterimental, in type 2 DM and a high risk of CV disease.

TYPE 1 DIABETES MELLITUS

General Aspects

The ADA recommendations for fasting and bedtime glycemic goals and A_{1C} targets are summarized in Table 344-8. The goal is to design and implement insulin regimens that mimic physiologic insulin secretion. Because individuals with type 1 DM partially or completely lack endogenous insulin production, administration of basal insulin is essential for regulating glycogen breakdown, gluconeogenesis, lipolysis, and ketogenesis. Likewise, insulin replacement for meals should be appropriate for the carbohydrate intake and promote normal glucose utilization and storage.

Intensive Management

Intensive diabetes management has the goal of achieving euglycemia or near-normal glycemia. This approach requires multiple resources, including thorough and continuing patient education, comprehensive recording of plasma glucose measurements and nutrition intake by the patient, and a variable insulin regimen that matches glucose intake and insulin dose. Insulin regimens usually include multiple-component insulin regimens, multiple daily injections (MDI), or insulin infusion devices (each discussed below).

The benefits of intensive diabetes management and improved glycemic control include a reduction in the microvascular complications of DM and a reduction in the macrovascular complications of DM, which persists after a period of near-normoglycemia. From a psychological standpoint, the patient experiences greater control over his or her diabetes and often notes an improved sense of well-being, greater flexibility in the timing and content of meals, and the capability to alter insulin dosing with exercise. In addition, intensive diabetes management prior to and during pregnancy reduces the risk of fetal malformations and morbidity. Intensive diabetes management is strongly encouraged in newly diagnosed patients with type 1 DM because it may prolong the period of C-peptide production, which may result in better glycemic control and a reduced risk of serious hypoglycemia.

Although intensive management confers impressive benefits, it is also accompanied by significant personal and financial costs and is therefore not appropriate for all individuals.

Insulin Preparations

Current insulin preparations are generated by recombinant DNA technology and consist of the amino acid sequence of human insulin or variations thereof. In the United States, most insulin is formulated as U-100 (100 units/mL). Regular insulin formulated as U-500 (500 units/mL) is available and sometimes useful in patients with severe insulin resistance. Human insulin has been formulated with distinctive pharmacokinetics or genetically modified to more closely mimic physiologic insulin secretion. Insulins can be classified as short-acting or long-acting **(Table 344-10)**. For example, one short-acting insulin formulation, insulin lispro, is an insulin analogue in which the 28th and 29th amino acids (lysine and proline) on the insulin B chain have been reversed by recombinant DNA technology. Insulin aspart and insulin glulisine are other genetically modified insulin analogues with properties similar to lispro. These insulin analogues have full biologic activity but less tendency for self-aggregation, resulting in more rapid

absorption and onset of action and a shorter duration of action. These characteristics are particularly advantageous for allowing entrainment of insulin injection and action to rising plasma glucose levels following meals. The shorter duration of action also appears to be associated with a decreased number of hypoglycemic episodes, primarily because the decay of insulin action corresponds to the decline in plasma glucose after a meal. Thus, insulin aspart, lispro, or glulisine is preferred over regular insulin for prandial coverage. Insulin glargine is a long-acting biosynthetic human insulin that differs from normal insulin in that asparagine is replaced by glycine at amino acid 21, and two arginine residues are added to the C terminus of the B chain. Compared to NPH insulin, the onset of insulin glargine action is later, the duration of action is longer (~24 h), and there is a less pronounced peak. A lower incidence of hypoglycemia, especially at night, has been reported with insulin glargine when compared to NPH insulin. The possible association between glargine and increased cancer risk is being investigated (FDA review underway) and is controversial. Insulin detemir has a fatty acid side chain that prolongs its action by slowing absorption and catabolism. Twice-daily injections of glargine of detemir are sometimes required to provide 24-h coverage. Regular and NPH insulin have the native insulin amino acid sequence.

Table 344-10 Properties of Insulin Preparations

	Time of Action		
Preparation	**Onset, h**	**Peak, h**	**Effective Duration, h**
Short-acting			
Aspart	<0.25	0.5–1.5	3–4
Glulisine	<0.25	0.5–1.5	3–4
Lispro	<0.25	0.5–1.5	3–4
Regular	0.5–1.0	2–3	4–6
Long-acting			
Detemir	1–4	—[a]	Up to 24
Glargine	1–4	—[a]	Up to 24
NPH	1–4	6–10	10–16
Insulin combinations			
75/25–75% protamine lispro, 25% lispro	<0.25	1.5 h	Up to 10–16
70/30–70% protamine aspart, 30% aspart	<0.25	1.5 h	Up to 10–16
50/50–50% protamine lispro, 50% lispro	<0.25	1.5 h	Up to 10–16
70/30–70% NPH, 30% regular	0.5–1	Dual[b]	10–16

[a]Glargine and detemir have minimal peak activity.

[b]Dual: two peaks—one at 2–3 h and the second one several hours later.

Source: Adapted from JS Skyler, Therapy for Diabetes Mellitus and Related Disorders, American Diabetes Association, Alexandria, VA, 2004.

Basal insulin requirements are provided by long-acting (NPH insulin, insulin glargine, or insulin detemir) insulin formulations. These are usually prescribed with short-acting insulin in an attempt to mimic physiologic insulin release with meals. Although mixing of NPH and short-acting insulin formulations is common practice, this mixing may alter the insulin absorption profile (especially the short-acting insulins). For example, lispro absorption is delayed by mixing with NPH. The alteration in insulin absorption when the patient mixes different insulin formulations should not discourage mixing insulins. However, the following guidelines should be followed: (1) Mix the different insulin formulations in the syringe immediately before injection (inject within 2 min after mixing); (2) do not store insulin as a mixture; (3) follow the same routine in terms of insulin mixing and administration to standardize the physiologic response to injected insulin; and (4) do not mix insulin glargine or detemir with other insulins. The miscibility of human regular and NPH insulin allows for the production of combination insulins that contain 70% NPH and 30% regular (70/30), or equal mixtures of NPH and regular (50/50). Other combination insulin formulations are insulin aspart (70/30) and insulin lispro (75/25 and 50/50). By including the insulin analogue mixed with protamine, these combinations have a short-acting and long-acting profile (Table 344-10). While more convenient for the patient (only two injections/day), combination insulin formulations do not allow independent adjustment of short-acting and long-acting activity. Several insulin formulations are available as insulin "pens," which may be more convenient for some patients. Insulin delivery by inhalation is no longer available but remains under investigation.

Insulin Regimens

Representations of the various insulin regimens that may be utilized in type 1 DM are illustrated in **Fig. 344-12**. Although the insulin profiles are depicted as "smooth," symmetric curves, there is considerable patient-to-patient variation in the peak and duration. In all regimens, long-acting insulins (NPH, glargine, or detemir) supply basal insulin, whereas regular, insulin aspart, glulisine, or lispro insulin provides prandial insulin. Short-acting insulin analogues should be injected just before (<20 min) or just after a meal; regular insulin is given 30–45 min prior to a meal.

Figure 344-12

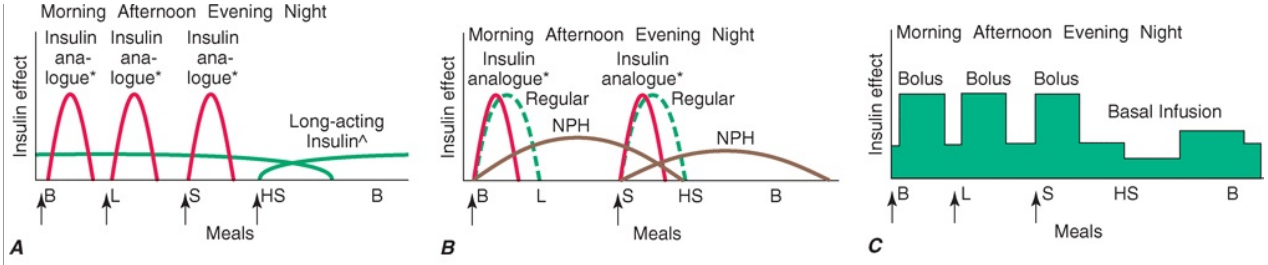

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Representative insulin regimens for the treatment of diabetes. For each panel, the *y*-axis shows the amount of insulin effect and the *x*-axis shows the time of day. B, breakfast; L, lunch; S, supper; HS, bedtime; CSII, continuous subcutaneous insulin infusion. *Lispro, glulisine, or insulin aspart can be used. The time of insulin injection is shown with a vertical arrow. The type of insulin is noted above each insulin curve. ***A***. A multiple-component insulin regimen consisting of long-acting insulin A glargine or detemir may be required each day) to provide basal insulin coverage and three shots of glulisine, lispro, or insulin aspart to provide glycemic coverage for each meal. ***B***. The injection of two shots of long-acting insulin (NPH) and short-acting insulin [glulisine, lispro, insulin aspart (solid red line), or regular (green dashed line)]. Only one formulation of short-acting insulin is used. ***C***. Insulin administration by insulin infusion device is shown with the basal insulin and a bolus injection at each meal. The basal insulin rate is decreased during the evening and increased slightly prior to the patient awakening in the morning. Glulisine, lispro, or insulin aspart is used in the insulin pump. *[Adapted from H Lebovitz (ed): Therapy for Diabetes Mellitus. American Diabetes Association, Alexandria, VA, 2004.]*

A shortcoming of current insulin regimens is that injected insulin immediately enters the systemic circulation, whereas endogenous insulin is secreted into the portal venous system. Thus, exogenous insulin administration exposes the liver to subphysiologic insulin levels. No insulin regimen reproduces the precise insulin secretory pattern of the pancreatic islet. However, the most physiologic regimens entail more frequent insulin injections, greater reliance on short-acting insulin, and more frequent capillary plasma glucose measurements. In general, individuals with type 1 DM require 0.5–1 U/kg per day of insulin divided into multiple doses, with ~50% of the insulin given as basal insulin.

Multiple-component insulin regimens refer to the combination of basal insulin and bolus insulin (preprandial short-acting insulin). The timing and dose of short-acting, preprandial insulin are altered to accommodate the SMBG results, anticipated food intake, and physical activity. Such regimens offer the patient with type 1 diabetes more flexibility in terms of lifestyle and the best chance for achieving near normoglycemia. One such regimen, shown in Fig. 344-12*B*, consists of basal insulin with glargine or detemir and preprandial lispro, glulisine, or insulin aspart. The insulin aspart, glulisine, or lispro dose is based on individualized algorithms that integrate the preprandial glucose and the anticipated carbohydrate intake. To determine the meal component of the preprandial insulin dose, the patient uses an insulin-to- carbohydrate ratio (a common ratio for type 1 DM is 1–1.5 units/10 g of carbohydrate, but this must be determined for each individual). To this insulin dose is added the supplemental or correcting insulin based on the preprandial blood glucose [one formula uses 1 unit of insulin for every 2.7 mmol/L (50 mg/dL) over the preprandial glucose target; another formula uses (body weight in kg) x (blood glucose – desired glucose in mg/dL)/1700]. An alternative multiple-component insulin regimen consists of bedtime NPH insulin, a small dose of NPH insulin at breakfast (20 –30% of bedtime dose), and preprandial short-acting insulin. Other variations of this regimen are in use but have the disadvantage that NPH has a significant peak, making hypoglycemia more common. Frequent SMBG (more than three times per day) is absolutely essential for these types of insulin regimens.

One commonly used regimen consists of twice-daily injections of NPH mixed with a short-acting insulin before the morning and evening meals (Fig. 344-12*A*). Such regimens usually prescribe two-thirds of the total daily insulin dose in the morning (with about two-thirds given as long-acting insulin and one-third as short-acting) and one-third before the evening meal (with approximately one-half given as long-acting insulin and one-half as short-acting). The drawback to such a regimen is that it enforces a rigid schedule on the patient, in terms of daily activity and the content and timing of meals. Although it is simple and effective at avoiding severe hyperglycemia, it does not generate near-normal glycemic control in individuals with type 1 DM. Moreover, if the patient's meal pattern or content varies or if physical activity is increased, hyperglycemia or hypoglycemia may result. Moving the long-acting insulin from before the evening meal to bedtime may avoid nocturnal hypoglycemia and provide more insulin as glucose levels rise in the early morning (so-called dawn phenomenon). The insulin dose in such regimens should be adjusted based on SMBG results with the following general assumptions: (1) the fasting glucose is primarily determined by the prior evening long-acting insulin; (2) the pre-lunch glucose is a function of the morning short-acting insulin; (3) the pre-supper glucose is a function of the morning long-acting insulin; and (4) the bedtime glucose is a function of the pre-supper, short-acting insulin. This is not an optimal regimen for the patient with type 1 DM, but is sometimes used for patients with type 2 diabetes.

Continuous SC insulin infusion (CSII) is a very effective insulin regimen for the patient with type 1 diabetes (Fig. 344-12*C*). To the basal insulin infusion, a preprandial insulin ("bolus") is delivered by the insulin infusion device based on instructions from the patient, who uses an individualized algorithm incorporating the preprandial plasma glucose and anticipated carbohydrate intake (see above). These sophisticated insulin infusion devices can accurately deliver small doses of insulin (microliters per hour) and have several advantages: (1) multiple basal infusion rates can be programmed to accommodate nocturnal versus daytime basal insulin requirement, (2) basal infusion rates can be altered during periods of exercise, (3) different waveforms of insulin infusion with meal-related bolus allow better matching of insulin depending on meal composition, and (4) programmed algorithms consider prior insulin administration and blood glucose values in calculating the insulin dose. These devices require instruction by a health professional with considerable experience with insulin infusion devices and very frequent patient interactions with the diabetes management team. Insulin infusion devices present unique challenges, such as infection at the infusion site, unexplained hyperglycemia because the infusion set becomes obstructed, or diabetic ketoacidosis if the pump becomes disconnected. Since most physicians use lispro, glulisine, or insulin aspart in CSII, the extremely short half-life of these insulins quickly lead to insulin deficiency if the delivery system is interrupted. Essential to the safe use of infusion devices is thorough patient education about pump function and frequent SMBG. Efforts to create a closed-loop system in which data from continuous glucose measurement regulates the insulin infusion rate continue.

Other Agents that Improve Glucose Control

The role of amylin, a 37-amino-acid peptide co-secreted with insulin from pancreatic beta cells, in normal glucose homeostasis is uncertain. However, based on the rationale that patients who are insulin deficient are also amylin deficient, an analogue of amylin (pramlintide) was created and found to reduce postprandial glycemic excursions in type 1 and type 2 diabetic patients taking insulin. Pramlintide injected just before a meal slows gastric emptying and suppresses glucagon but does not alter insulin levels. Pramlintide is approved for insulin-treated patients with type 1 and type 2 DM. Addition of pramlintide produces a modest reduction in the A1C and seems to dampen meal-related glucose excursions. In type 1 diabetes, pramlintide is started as a 15-μg

SC injection before each meal and titrated up to a maximum of 30–60 μg as tolerated. In type 2 DM, pramlintide is started as a 60-μg SC injection before each meal and may be titrated up to a maximum of 120 μg. The major side effects are nausea and vomiting, and dose escalations should be slow to limit these side effects. Because pramlintide slows gastric emptying, it may influence absorption of other medications and should not be used in combination with other drugs that slow GI motility. The short-acting insulin given before the meal should initially be reduced to avoid hypoglycemia and then titrated as the effects of the pramlintide become evident. α-Glucosidase inhibitors are another type of agent that is sometimes used in conjunction in patients with type 1 DM.

TYPE 2 DIABETES MELLITUS

General Aspects

The goals of therapy for type 2 DM are similar to those in type 1. While glycemic control tends to dominate the management of type 1 DM, the care of individuals with type 2 DM must also include attention to the treatment of conditions associated with type 2 DM (obesity, hypertension, dyslipidemia, cardiovascular disease) and detection/management of DM-related complications **(Fig. 344-13)**. DM-specific complications may be present in up to 20–50% of individuals with newly diagnosed type 2 DM. Reduction in cardiovascular risk is of paramount importance as this is the leading cause of mortality in these individuals. Efforts to achieve blood pressure and lipid goals (Table 344-8) should begin in concert with glucose-lowering interventions.

Figure 344-13

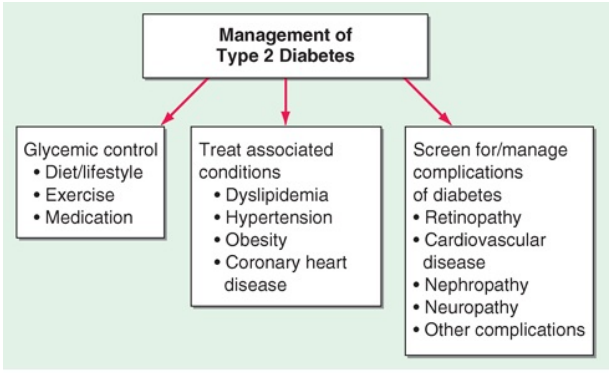

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Essential elements in comprehensive diabetes care of type 2 diabetes.

Type 2 diabetes management should begin with MNT (discussed above). An exercise regimen to increase insulin sensitivity and promote weight loss should also be instituted. Pharmacologic approaches to the management of type 2 DM include oral glucose-lowering agents, insulin, and other agents that improve glucose control; most physicians and patients prefer oral glucose-lowering agents as the initial choice (discussed below) after review of various medications. Any therapy that improves glycemic control reduces "glucose toxicity" to the islet cells and improves endogenous insulin secretion. However, type 2 DM is a progressive disorder and ultimately requires multiple therapeutic agents and often insulin.

Glucose-Lowering Agents

Advances in the therapy of type 2 DM have generated oral glucose-lowering agents that target different pathophysiologic processes in type 2 DM. Based on their mechanisms of action, glucose-lowering agents are subdivided into agents that increase insulin secretion, reduce glucose production, increase insulin sensitivity, and enhance GLP-1 action **(Table 344-11)**. Glucose-lowering agents other than insulin (with the exception of amylin analogue and α-glucosidase inhibitors) are ineffective in type 1 DM and should not be used for glucose management of severely ill individuals with type 2 DM. Insulin is sometimes the initial glucose-lowering agent in type 2 diabetes.

Table 344-11 Agents Used for Treatment of Type 1 and Type 2 Diabetes

	Mechanism of Action	Examples	A1C Reduction (%)[a]	Agent-Specific Advantages	Agent-Specific Disadvantages	Contraindications
Oral						
Biguanides[b]	↓ Hepatic glucose production	Metformin	1–2	Weight neutral, Do not cause hypoglycemia, inexpensive	Diarrhea, nausea, lactic acidosis	Serum creatinine >1.5 mg/dL (men) >1.4 mg/dL (women), CHF, radiographic contrast studies, seriously ill patients, acidosis
α-Glucosidase inhibitors[b]	↓ GI glucose absorption	Acarbose, Miglitol	0.5–0.8	Reduce postprandial glycemia	GI flatulence, liver function tests	Renal/liver disease
Dipeptidyl peptidase IV inhibitors[b]	Prolong endogenous GLP-1 action	Saxagliptin, Sitagliptin, Vildagliptin	0.5–0.8	Do not cause hypoglycemia		Reduce dose with renal disease
Insulin secretagogues: Sulfonylureas[b]	↑ Insulin secretion	See text and Table 344-12	1–2	Inexpensive	Hypoglycemia, weight gain	Renal/liver disease
Insulin secretagogues:	↑ Insulin secretion	See text and Table 344-12	1–2	Short onset of action, lower	Hypoglycemia	Renal/liver disease

Non-sulfonylureas[b]				postprandial glucose		
Thiazolidinediones[b]	↓ Insulin resistance, ↑ glucose utilization	Rosiglitazone, Pioglitazone	0.5–1.4	Lower insulin requirements	Peripheral edema, CHF, weight gain, fractures, macular edema; rosiglitazone may increase cardiovascular risk	CHF, liver disease; see text about rosiglitazone
Bile acid sequestrants	Bind bile acids; mechanism of glucose lowering not known	Colesevelam	0.5		Constipation, dyspepsia, abdominal pain, nausea, ↑ tri-glycerides, interfere with absorption of other drugs, intestinal obstruction	Elevated plasma triglycerides
Parenteral						
Insulin[b,c]	↑ Glucose utilization, ↓ Hepatic glucose production, and other anabolic actions	See text and Table 344-10	Not limited	Known safety profile	Injection, weight gain, hypoglycemia	
GLP-1 receptor agonists[b]	↑ Insulin, ↓ glucagon, slow gastric emptying, satiety	Exenatide, liraglutide	0.5–1.0	Weight loss, do not cause hypoglycemia	Injection, nausea, ↑ risk of hypoglycemia with insulin secretagogues, pancreatitis, renal failure	Renal disease, agents that also slow GI motility; see text
Amylin agonists[b,c]	Slow gastric emptying, ↓ glucagon	Pramlintide	0.25–0.5	Reduce postprandial glycemia; weight loss	Injection, nausea, ↑ risk of hypoglycemia with insulin	Agents that also slow GI motility
Medical nutrition therapy and physical activity[b,c]	↓ Insulin resistance, ↑ insulin secretion	Low-calorie, low-fat diet, exercise	1–3	Other health benefits	Compliance difficult, long-term success low	

[a]A1C reduction (absolute) depends partly on starting A1C.

[b]Used for treatment of type 2 diabetes.

[c]Used in conjunction with insulin for treatment of type 1 diabetes.

Biguanides

Metformin, representative of this class of agents, reduces hepatic glucose production and improves peripheral glucose utilization slightly (Table 344-11). Metformin activates AMP-dependent protein kinase and enters cells through organic cation transporters (polymorphisms of these may influence the response to metformin). Metformin reduces fasting plasma glucose and insulin levels, improves the lipid profile, and promotes modest weight loss. The initial starting dose of 500 mg once or twice a day can be increased to 1000 mg bid. An extended-release form is available and may have fewer gastrointestinal side effects (diarrhea, anorexia, nausea, metallic taste). Because of its relatively slow onset of action and gastrointestinal symptoms with higher doses, the dose should be escalated every 2–3 weeks based on SMBG measurements. Metformin is effective as monotherapy and can be used in combination with other oral agents or with insulin. The major toxicity of metformin, lactic acidosis is very rare and can be prevented by careful patient selection. Vitamin B12 levels are ~30% lower during metformin treatment. Metformin should not be used in patients with renal insufficiency [GFR < 60 mL/min], any form of acidosis, CHF, liver disease, or severe hypoxemia. Metformin should be discontinued in patients who are seriously ill, in patients who can take nothing orally, and in those receiving radiographic contrast material. Insulin should be used until metformin can be restarted.

Insulin Secretagogues—Agents that Affect the ATP-Sensitive K^+ Channel

Insulin secretagogues stimulate insulin secretion by interacting with the ATP-sensitive potassium channel on the beta cell (Fig. 344-4). These drugs are most effective in individuals with type 2 DM of relatively recent onset (<5 years), who have residual endogenous insulin production. First-generation sulfonylureas (chlorpropamide, tolazamide, tolbutamide; not shown in **Table 344-12**) have a longer half-life, a greater incidence of hypoglycemia, more frequent drug interactions, and are now rarely used. Second-generation sulfonylureas have a more rapid onset of action and better coverage of the postprandial glucose rise, but the shorter half-life of some agents may require more than once-a-day dosing (Table 344-12). Sulfonylureas reduce both fasting and postprandial glucose and should be initiated at low doses and increased at 1- to 2-week intervals based on SMBG. In general, sulfonylureas increase insulin acutely and thus should be taken shortly before a meal; with chronic therapy, though, the insulin release is more sustained. Glimepiride and glipizide can be given in a single daily dose and are preferred over glyburide. Repaglinide and nateglinide are not sulfonylureas but also interact with the ATP-sensitive potassium channel. Because of their short half-life, these agents are given with each meal or immediately before to reduce meal-related glucose excursions.

Table 344-12 Properties of Insulin Secretagogues

Class/Generic Name	Daily Dosage, mg	Duration of Action, h
Sulfonylureas		
Glimepiride	1–8	24
Glipizide	5–40	12–18
Glipizide (extended release)	5–20	24
Glyburide	1.25–20	12–24
Glyburide (micronized)	0.75–12	12–24
Nonsulfonylureas (Meglititinides)		
Repaglinide	0.5–16	2–6

Nateglinide	180–360	2–4
GLP-1 agonist		
Exenatide	0.01–0.02	4–6
Liraglutide	0.6–1.8	12–24
Dipeptidyl Peptidase-4 Inhibitors		
Saxagliptin	2.5–5	12–16
Sitagliptin	100	12–16
Vildagliptin	50–100	12–24

Abbreviation: GLP-1, glucagon-like peptide 1.

These insulin secretagogues are generally well tolerated. These agents, especially the longer acting ones, have the potential to cause profound and persistent hypoglycemia, especially in elderly individuals. Hypoglycemia is usually related to delayed meals, increased physical activity, alcohol intake, or renal insufficiency. Individuals who ingest an overdose of some agents develop prolonged and serious hypoglycemia and should be monitored closely in the hospital (Chap. 345). Most sulfonylureas are metabolized in the liver to compounds (some of which are active) that are cleared by the kidney. Thus, their use in individuals with significant hepatic or renal dysfunction is not advisable. Weight gain, a common side effect of sulfonylurea therapy, results from the increased insulin levels and improvement in glycemic control. Some sulfonylureas have significant drug interactions with alcohol and some medications including warfarin, aspirin, ketoconazole, α-glucosidase inhibitors, and fluconazole. A related isoform of ATP-sensitive potassium channels is present in the myocardium and the brain. All of these agents except glyburide have a low affinity for this isoform. Despite concerns that this agent might affect the myocardial response to ischemia and observational studies suggesting that sulfonylureas increase cardiovascular risk, studies have not shown an increased cardiac mortality with glyburide.

Insulin Secretagogues—Agents that Enhance Glp-1 Receptor Signaling

"Incretins" amplify glucose-stimulated insulin secretion (Fig. 344-4). Agents that either act as a GLP-1 agonist or enhance endogenous GLP-1 activity are approved for the treatment of type 2 diabetes (Table 344-12). Agents in this class do not cause hypoglycemia because of the glucose-dependent nature of incretin-stimulated insulin secretion (unless there is concomitant use of an agent that can lead to hypoglycemia—sulfonylureas, etc.). Exenatide, a synthetic version of a peptide initially identified in the saliva of the Gila monster (exendin-4), is an analogue of GLP-1. Unlike native GLP-1, which has a half-life of <5 min, differences in the exenatide amino acid sequence render it resistant to the enzyme that degrades GLP-1 (dipeptidyl peptidase IV, or DPP-IV). Thus, exenatide has prolonged GLP-1-like action and binds to GLP-1 receptors found in islets, the gastrointestinal tract, and the brain. Liraglutide, another GLP-1 receptor agonist, is almost identical to native GLP-1 except for an amino acid substitution and addition of a fatty acyl group (coupled with a γ-glutamic acid spacer) that promote binding to albumin and plasma proteins and prolong its half-life. GLP-1 receptor agonists increase glucose-stimulated insulin secretion, suppress glucagon, and slow gastric emptying. These agents do not promote weight gain; in fact, most patients experience modest weight loss and appetite suppression. Treatment with these agents should start at a low dose to minimize initial side effects (nausea being the limiting one). Exenatide is approved for monotherapy and for use as combination therapy with metformin, sulfonylureas, and thiazolidinediones. Some patients taking insulin secretagogues may require a reduction in those agents to prevent hypoglycemia. GLP-1 receptor agonists should not be used in patients taking insulin. The major side effects are nausea, vomiting, and diarrhea; pancreatitis and reduced renal function have been reported in surveillance data with exenatide. Liraglutide carries a black box warning from the FDA because of an increased risk of thyroid C-cell tumors in rodents and is contraindicated in individuals with medullary carcinoma of the thyroid and multiple endocrine neoplasia. Because GLP-1 receptor agonists slow gastric emptying, they may influence the absorption of other drugs. Whether GLP-1 receptor agonists enhance beta cell survival, promote beta cell proliferation, or alter the natural history of type 2 DM is not known. Other GLP-1 receptor agonists and formulations are under development.

DPP-IV inhibitors inhibit degradation of native GLP-1 and thus enhance the incretin effect. DPP-IV, which is widely expressed on the cell surface of endothelial cells and some lymphocytes, degrades a wide range of peptides (not GLP-1 specific). DPP-IV inhibitors promote insulin secretion in the absence of hypoglycemia or weight gain, and appear to have a preferential effect on postprandial blood glucose. DPP-IV inhibitors are used either alone or in combination with other oral agents in adults with type 2 DM. Reduced doses should be given to patients with renal insufficiency. These agents have relatively few side effects.

α-Glucosidase Inhibitors

α-Glucosidase inhibitors (acarbose and miglitol) reduce postprandial hyperglycemia by delaying glucose absorption; they do not affect glucose utilization or insulin secretion (Table 344-11). Postprandial hyperglycemia, secondary to impaired hepatic and peripheral glucose disposal, contributes significantly to the hyperglycemic state in type 2 DM. These drugs, taken just before each meal, reduce glucose absorption by inhibiting the enzyme that cleaves oligosaccharides into simple sugars in the intestinal lumen. Therapy should be initiated at a low dose (25 mg of acarbose or miglitol) with the evening meal and may be increased to a maximal dose over weeks to months (50–100 mg for acarbose or 50 mg for miglitol with each meal). The major side effects (diarrhea, flatulence, abdominal distention) are related to increased delivery of oligosaccharides to the large bowel and can be reduced somewhat by gradual upward dose titration. α-Glucosidase inhibitors may increase levels of sulfonylureas and increase the incidence of hypoglycemia. Simultaneous treatment with bile acid resins and antacids should be avoided. These agents should not be used in individuals with inflammatory bowel disease, gastroparesis, or a serum creatinine >177 μmol/L (2 mg/dL). This class of agents is not as potent as other oral agents in lowering the A1C but is unique because it reduces the postprandial glucose rise even in individuals with type 1 DM. If hypoglycemia from other diabetes treatments occurs while taking these agents, the patient should consume glucose since the degradation and absorption of complex carbohydrates will be retarded.

Thiazolidinediones

Thiazolidinediones reduce insulin resistance by binding to the PPAR-γ (peroxisome proliferator-activated receptor γ) nuclear receptor (which forms a heterodimer with the retinoid X receptor). The PPAR-γ receptor is found at highest levels in adipocytes but is expressed at lower levels in many other tissues. Agonists of this receptor regulate a large number of genes, promote adipocyte differentiation, reduce hepatic fat accumulation, and promote fatty acid storage (Table 344-11). Thiazolidinediones promote a redistribution of fat from central to peripheral locations. Circulating insulin levels decrease with use of the thiazolidinediones, indicating a reduction in insulin resistance. Although direct comparisons are not available, the two currently available thiazolidinediones appear to have similar efficacy; the therapeutic range for pioglitazone is 15–45 mg/d in a single daily dose, and for rosiglitazone the total daily dose is 2–8 mg/d administered either once daily or twice daily in divided doses. The prototype of this class of drugs, troglitazone, was withdrawn from the U.S. market after reports of hepatotoxicity and an association with an idiosyncratic liver reaction that sometimes led to hepatic failure. Although rosiglitazone and pioglitazone do not appear to induce the liver abnormalities seen with troglitazone, the FDA recommends measurement of liver function tests prior to initiating therapy with a thiazolidinedione and at regular intervals (every 2 months for the first year and then periodically).

Rosiglitazone raises LDL, HDL, and triglycerides slightly. Pioglitazone raises HDL to a greater degree and LDL a lesser degree

but lowers triglycerides. The clinical significance of the lipid changes with these agents is not known and may be difficult to ascertain since most patients with type 2 diabetes are also treated with a statin.

Thiazolidinediones are associated with weight gain (2–3 kg), a small reduction in the hematocrit, and a mild increase in plasma volume. Peripheral edema and CHF are more common in individuals treated with these agents. These agents are contraindicated in patients with liver disease or CHF (class III or IV). The FDA has issued an alert that rare patients taking these agents may experience a worsening of diabetic macular edema. An increased risk of fractures has been noted in women taking these agents. Thiazolidinediones have been shown to induce ovulation in premenopausal women with PCOS. Women should be warned about the risk of pregnancy, since the safety of thiazolidinediones in pregnancy is not established.

Recent concerns about increased cardiovascular risk associated with rosiglitazone have led to considerable restrictions on its use. The FDA issued a "black box" warning for rosiglitazone in 2007. Under recent FDA guidelines (September 2010), new prescriptions for rosiglitazone in the United States must be written under a "risk evaluation and mitigation strategy" and only for patients with diabetes that cannot be controlled by other medications. The European Medicines Agency (2010) advised removal of rosiglitazone and formulations containing rosiglitazone from the European market. As a result of thesse rulings, rosiglitazone will be available in the United States on a limited basis, but will not be available in Europe.

Other Therapies for Type 2 Diabetes

Bile acid–binding resins. Bile acid metabolism is abnormal type 2 diabetes. The bile acid–binding resin colesevelam has been approved for the treatment of type 2 diabetes (already approved for treatment of hypercholesterolemia). Emerging evidence indicates that bile acids, by signaling through nuclear receptors, may have a role in metabolism. Since bile acid–binding resins are minimally absorbed into the systemic circulation, how bile acid–binding resins lower blood glucose is not known. Colesevelam (available as a powder for oral solution and as 625-mg tablets) is prescribed as 3–6 tablets prior to meals. The most common side effects are gastrointestinal (constipation, abdominal pain, and nausea). Bile acid–binding resins can increase plasma triglycerides and should be used cautiously in patients with a tendency to hypertriglyceridemia. The role of this class of drugs in the treatment of type 2 diabetes is not yet defined.

Bromocriptine. A formulation of the dopamine receptor agonist bromocriptine (Cycloset), has been approved by the FDA for the treatment of type 2 diabetes (approved in 2009). However, this formulation is not available in the United States, and its role in the treatment of type 2 diabetes is uncertain.

Insulin Therapy in Type 2 DM

Insulin should be considered as the initial therapy in type 2 DM, particularly in lean individuals or those with severe weight loss, in individuals with underlying renal or hepatic disease that precludes oral glucose-lowering agents, or in individuals who are hospitalized or acutely ill. Insulin therapy is ultimately required by a substantial number of individuals with type 2 DM because of the progressive nature of the disorder and the relative insulin deficiency that develops in patients with long-standing diabetes. Both physician and patient reluctance often delay the initiation of insulin therapy, but glucose control and patient well-being are improved by insulin therapy in patients who have not reached the glycemic target.

Because endogenous insulin secretion continues and is capable of providing some coverage of mealtime caloric intake, insulin is usually initiated in a single dose of long-acting insulin (0.3–0.4 U/kg per day), given either before breakfast and in the evening (NPH) or just before bedtime (NPH, glargine, detemir). Since fasting hyperglycemia and increased hepatic glucose production are prominent features of type 2 DM, bedtime insulin is more effective in clinical trials than a single dose of morning insulin. Glargine given at bedtime has less nocturnal hypoglycemia than NPH insulin. Some physicians prefer a relatively low, fixed starting dose of long-acting insulin (5–15 units) to reduce the chance of hypoglycemia in the initial treatment period. The insulin dose may then be adjusted in 10% increments as dictated by SMBG results. Both morning and bedtime long-acting insulin may be used in combination with oral glucose-lowering agents. Initially, basal insulin may be sufficient, but more often prandial insulin coverage with multiple insulin injections is needed as diabetes progresses (see insulin regimens used for type 1 diabetes). Other insulin formulations that have a combination of short-acting and long-acting insulin (Table 344-10) are sometimes used in patients with type 2 DM because of convenience but do not allow independent adjustment of short-acting and long-acting insulin dose and often do not achieve the same degree of glycemic control as basal/bolus regimens. In selected patients with type 2 DM (usually insulin deficient as defined by C-peptide level), insulin infusion devices may be considered.

Choice of Initial Glucose-Lowering Agent

The level of hyper-glycemia should influence the initial choice of therapy. Assuming maximal benefit of MNT and increased physical activity has been realized, patients with mild to moderate hyperglycemia [FPG <11.1–13.9 mmol/L (200–250 mg/dL)] often respond well to a single, oral glucose-lowering agent. Patients with more severe hyperglycemia [FPG >13.9 mmol/L (250 mg/dL)] may respond partially but are unlikely to achieve normoglycemia with oral monotherapy. A stepwise approach that starts with a single agent and adds a second agent to achieve the glycemic target can be used (see "Combination Therapy," below). Insulin can be used as initial therapy in individuals with severe hyperglycemia [FPG >13.9–16.7 mmol/L (250–300 mg/dL)] or in those who are symptomatic from the hyperglycemia. This approach is based on the rationale that more rapid glycemic control will reduce "glucose toxicity" to the islet cells, improve endogenous insulin secretion, and possibly allow oral glucose-lowering agents to be more effective. If this occurs, the insulin may be discontinued.

Insulin secretagogues, biguanides, α-glucosidase inhibitors, thiazolidinediones, GLP-1 receptor agonists, DPP-IV inhibitors, and insulin are approved for monotherapy of type 2 DM. Although each class of oral glucose-lowering agents has unique advantages and disadvantages, certain generalizations apply: (1) insulin secretagogues, biguanides, GLP-1 receptor agonists, and thiazolidinediones improve glycemic control to a similar degree (1–2% reduction in A1C) and are more effective than α-glucosidase inhibitors and DPP-IV inhibitors; (2) assuming a similar degree of glycemic improvement, no clinical advantage to one class of drugs has been demonstrated, and any therapy that improves glycemic control is likely beneficial; (3) insulin secretagogues, GLP-1 receptor agonists, DPP-IV inhibitors, and α-glucosidase inhibitors begin to lower the plasma glucose immediately, whereas the glucose-lowering effects of the biguanides and thiazolidinediones are delayed by several weeks; (4) not all agents are effective in all individuals with type 2 DM (primary failure); (5) biguanides, α-glucosidase inhibitors, GLP-1 receptor agonists, DPP-IV inhibitors, and thiazolidinediones do not directly cause hypoglycemia; and (6) most individuals will eventually require treatment with more than one class of oral glucose-lowering agents or insulin, reflecting the progressive nature of type 2 DM; (7) durability of glycemic control is slightly less for glyburide compared to metformin or rosiglitazone.

Considerable clinical experience exists with metformin and sulfonylureas because they have been available for several decades. It is assumed that the α-glucosidase inhibitors, GLP-1 agonists, DPP-IV inhibitors, and thiazolidinediones, which are newer classes of oral glucose-lowering drugs, will reduce DM-related complications by improving glycemic control, although long-term data are not yet available. The thiazolidine-diones are theoretically attractive because they target a fundamental abnormality in type 2 DM, namely insulin resistance. However, all of these agents are currently more costly than metformin and sulfonylureas.

A reasonable treatment algorithm for initial therapy uses metformin as initial therapy because of its efficacy, known side-effect profile, and relatively low cost **(Fig. 344-14)**. Metformin has the advantage that it promotes mild weight loss, lowers insulin levels, and improves the lipid profile slightly. Based on SMBG results and the A_{1C}, the dose of metformin should be increased until the glycemic target is achieved or maximum dose is reached.

Figure 344-14

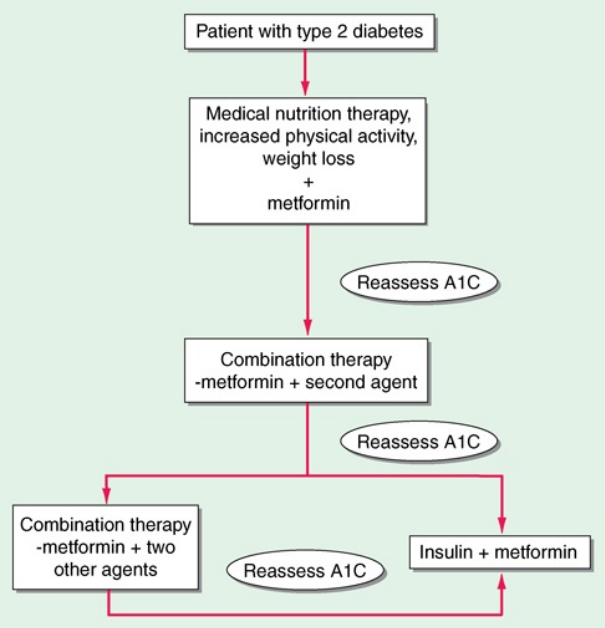

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Glycemic management of type 2 diabetes. See text for discussion of treatment of severe hyperglycemia or symptomatic hyperglycemia. Agents that can be combined with metformin include insulin secre-tagogues, thiazolidinediones, α-glucosidase inhibitors, DPP-IV inhibitors, and GLP-1 receptor agonists. A_{1C}, hemoglobin A_{1C}.

Combination Therapy with Glucose-Lowering Agents

A number of combinations of therapeutic agents are successful in type 2 DM, and the dosing of agents in combination is the same as when the agents are used alone. Because mechanisms of action of the first and second agents used are different, the effect on glycemic control is usually additive. Several fixed dose combinations of oral agents are available, but evidence that they are superior to titration of single agent to a maximum dose and then addition of a second agent is lacking. If adequate control is not achieved with the combination of two agents (based on reassessment of the A_{1C} every 3 months), a third oral agent or basal insulin should be added (Fig. 344-14).

Treatment with insulin becomes necessary as type 2 DM enters the phase of relative insulin deficiency (as seen in long-standing DM) and is signaled by inadequate glycemic control with one or two oral glucose-lowering agents. Insulin alone or in combination should be used in patients who fail to reach the glycemic target. For example, a single dose of long-acting insulin at bedtime is effective in combination with metformin. As endogenous insulin production falls further, multiple injections of long-acting and short-acting insulin regimens are necessary to control postprandial glucose excursions. These insulin regimens are identical to the long-acting and short-acting combination regimens discussed above for type 1 DM. Since the hyper-glycemia of type 2 DM tends to be more "stable," these regimens can be increased in 10% increments every 2–3 days using the FBG results. The daily insulin dose required can become quite large (1–2 units/kg per day) as endogenous insulin production falls and insulin resistance persists. Individuals who require >1 unit/kg per day of long-acting insulin should be considered for combination therapy with metformin or a thiazolidinedione. The addition of metformin or a thiazolidinedione can reduce insulin requirements in some individuals with type 2 DM, while maintaining or even improving glycemic control. Insulin plus a thiazolidinedione promotes weight gain and is associated with peripheral edema. Addition of a thiazolidinedione to a patient's insulin regimen may necessitate a reduction in the insulin dose to avoid hypoglycemia.

EMERGING THERAPIES

Whole pancreas transplantation (performed concomitantly with a renal transplant) may normalize glucose tolerance and is an important therapeutic option in type 1 DM with ESRD, though it requires substantial expertise and is associated with the side effects of immunosuppression. Pancreatic islet transplantation has been plagued by limitations in pancreatic islet supply and graft survival and remains an area of clinical investigation. Many individuals with long-standing type 1 DM still produce very small amounts of insulin or have insulin-positive cells within the pancreas. This suggests that beta cells may slowly regenerate but are quickly destroyed by the autoimmune process. Thus, efforts to suppress the autoimmune process and to stimulate beta cell regeneration are being tested both at the time of diagnosis and in years after the diagnosis of type 1 DM. Closed-loop pumps that infuse the appropriate amount of insulin in response to changing glucose levels are potentially feasible now that continuous glucose-monitoring technology has been developed. New therapies under development for type 2 diabetes include an inhibitor of the sodium-glucose co- transporter in the kidney, activators of glucokinase, an inhibitor of 11 β-hydroxysteroid dehydrogenase-1, and salsalate.

Bariatric surgery for markedly obese individuals with type 2 diabetes has shown considerable promise—sometimes with resolution of the diabetes or major reductions in the needed dose of glucose-lowering therapies (Chap. 78). The ADA clinical guidelines state that bariatric surgery should be considered in individuals with DM and a BMI >35 kg/m^2.

COMPLICATIONS OF THERAPY FOR DIABETES MELLITUS

As with any therapy, the benefits of efforts directed toward glycemic control must be balanced against the risks of treatment. Side effects of intensive treatment include an increased frequency of serious hypoglycemia, weight gain, increased economic costs, and greater demands on the patient. In the DCCT, quality of life was very similar in the intensive and standard therapy groups. The most serious complication of therapy for DM is hypoglycemia, and its treatment with oral glucose or glucagon injection is discussed in Chap. 345. Severe, recurrent hypoglycemia warrants examination of treatment regimen and glycemic goal for the individual patient. Weight gain occurs with most (insulin, insulin secretagogues, thiazolidine-diones) but not all (metformin, α-glucosidase inhibitors, GLP-1 receptor agonists, DPP-IV inhibitors) therapies that improve glycemic control. The

weight gain is partially due to the anabolic effects of insulin and the reduction in glucosuria. In the DCCT, individuals with the greatest weight gain exhibited increases in LDL cholesterol and triglycerides as well as increases in blood pressure (both systolic and diastolic) similar to those seen in individuals with type 2 DM and insulin resistance. These effects could increase the risk of cardiovascular disease. As discussed previously, transient worsening of diabetic retinopathy or neuropathy sometimes accompanies improved glycemic control. As a result of recent controversies about the optimal glycemic goal and concerns about safety, the FDA now requires information about the cardiovascular safety profile as part of its new evaluation for treatment of type 2 diabetes.

ONGOING ASPECTS OF COMPREHENSIVE DIABETES CARE

The morbidity and mortality rates of DM-related complications can be greatly reduced by timely and consistent surveillance procedures **(Table 344-13)**. These screening procedures are indicated for all individuals with DM, but many individuals with diabetes do not receive comprehensive diabetes care. Screening for dyslipidemia and hypertension should be performed annually. In addition to routine health maintenance, individuals with diabetes should also receive the pneumococcal and tetanus vaccines (at recommended intervals) and the influenza vaccine (annually). As discussed above, aspirin therapy should be considered in many patients with diabetes (primary prevention in type 1 or type 2 DM men >50 years or women >60 years with one risk factor CV disease), but its role in primary prevention in low-risk individuals is uncertain and not recommended.

Table 344-13 Guidelines for Ongoing Medical Care for Patients with Diabetes

• Self-monitoring of blood glucose (individualized frequency)
• A1C testing (2–4 times/year)
• Patient education in diabetes management (annual)
• Medical nutrition therapy and education (annual)
• Eye examination (annual)
• Foot examination (1–2 times/year by physician; daily by patient)
• Screening for diabetic nephropathy (annual; see Fig. 344-11)
• Blood pressure measurement (quarterly)
• Lipid profile and serum creatinine (estimate GFR) (annual)
• Influenza/pneumococcal immunizations
• Consider antiplatelet therapy (see text)

Abbreviation: A1C, hemoglobin A1C.

An annual comprehensive eye examination should be performed by a qualified optometrist or ophthalmologist. If abnormalities are detected, further evaluation and treatment require an ophthalmologist skilled in diabetes-related eye disease. Because many individuals with type 2 DM have had asymptomatic diabetes for several years before diagnosis, the ADA recommends the following ophthalmologic examination schedule: (1) individuals with type 1 DM should have an initial eye examination within 5 years of diagnosis, (2) individuals with type 2 DM should have an initial eye examination at the time of diabetes diagnosis, (3) women with DM who are pregnant or contemplating pregnancy should have an eye examination prior to conception and during the first trimester, and (4) if eye exam is normal, repeat examination in 2–3 years may be appropriate.

An annual foot examination should (1) assess blood flow, sensation (monofilament testing, pin prick, or tuning fork), ankle reflexes, and nail care; (2) look for the presence of foot deformities such as hammer or claw toes and Charcot foot; and (3) identify sites of potential ulceration. The ADA recommends annual screening for distal symmetric neuropathy beginning with the initial diagnosis of diabetes and annual screening for autonomic neuropathy 5 years after diagnosis of type 1 DM and at the time of diagnosis of type 2 DM. This includes testing for loss of protective sensation (LOPS) using monofilament testing plus one of the following tests: vibration, pinprick, ankle reflexes, or vibration perception threshold (using a biothesiometer). If the monofilament test or one of the other tests is abnormal, the patient is diagnosed with LOPS and counseled accordingly. Calluses and nail deformities should be treated by a podiatrist; the patient should be discouraged from self-care of even minor foot problems but should be strongly encouraged to check his or her feet daily for any early lesions. Providers should consider screening for asymptomatic peripheral arterial disease using ankle-brachial index testing in high-risk individuals.

An annual microalbuminuria measurement (albumin-to- creatinine ratio in spot urine) is advised in individuals with type 1 or type 2 DM (Fig. 344-10). The urine protein measurement in a routine urinalysis does not detect these low levels of albumin excretion (microalbuminuria). Screening for microalbuminuria should commence 5 years after the onset of type 1 DM and at the time of diagnosis of type 2 DM. Regardless of protein excretion results, the GFR should be estimated using the serum creatinine in all patients on an annual basis.

SPECIAL CONSIDERATIONS IN DIABETES MELLITUS

Psychosocial Aspects

Because the individual with DM can face challenges that affect many aspects of daily life, psychosocial assessment and treatment are a critical part of providing comprehensive diabetes care. The individual with DM must accept that he or she may develop complications related to DM. Even with considerable effort, normoglycemia can be an elusive goal, and solutions to worsening glycemic control may not be easily identifiable. The patient should view him- or herself as an essential member of the diabetes care team and not as someone who is cared for by the diabetes management team. Emotional stress may provoke a change in behavior so that individuals no longer adhere to a dietary, exercise, or therapeutic regimen. This can lead to the appearance of either hyper- or hypoglycemia. Eating disorders, including binge eating disorders, bulimia, and anorexia nervosa, appear to occur more frequently in individuals with type 1 or type 2 DM (Chap. 79).

Management in the Hospitalized Patient

Virtually all medical and surgical subspecialties are involved in the care of hospitalized patients with diabetes. Hyperglycemia, whether in a patient with known diabetes or in someone without known diabetes, appears to be a predictor of poor outcome in hospitalized patients. General anesthesia, surgery, infection, or concurrent illness raises the levels of counterregulatory hormones (cortisol, growth hormone, catecholamines, and glucagon) and cytokines that may lead to transient insulin resistance and hyperglycemia. These factors increase insulin requirements by increasing glucose production and impairing glucose utilization and thus may worsen glycemic control. The concurrent illness or surgical procedure may lead to variable insulin absorption and also prevent the patient with DM from eating normally and, thus, may promote hypoglycemia. Glycemic control should be assessed on admission using the A_{1C}. Electrolytes, renal function, and intravascular volume status should be assessed as well. The high prevalence of cardiovascular disease in individuals with DM (especially in type 2 DM) may require preoperative cardiovascular evaluation.

The goals of diabetes management during hospitalization are near-normoglycemia, avoidance of hypoglycemia, and transition back to the outpatient diabetes treatment regimen. Glycemic control appears to improve the clinical outcomes in a variety of settings, but optimal glycemic goals for the hospitalized patient are incompletely defined. In a number of cross-sectional studies of patients with diabetes, a greater degree of hyperglycemia was associated with worse cardiac, neurologic, and infectious outcomes. In some studies, patients who do not have preexisting diabetes but who develop modest blood glucose elevations during their hospitalization appear to benefit from achieving near-normoglycemia using insulin treatment. However, a large randomized clinical trial (NICE-SUGAR) of individuals in the intensive care unit (most of whom were receiving mechanical ventilation) found an increased mortality rate and a greater number of episodes of severe hypoglycemia with very strict glycemic control [target BG of 4.5–6 mmol/L or 81–108 mg/dL] compared to individuals with a more moderate glycemic goal (mean blood glucose of 8 mmol/L or 144 mg/dL). Currently, most data suggest that very strict blood glucose control in acutely ill patients likely worsens outcomes and increases the frequency of hypoglycemia. The ADA suggests these glycemic goals for hospitalized patients: (1) in critically ill patients: glucose of 7.8–10.0 mmol/L or 140–180 mg/dL; (2) in non–critically ill patients: pre-meal glucose <7.8 mmol/L (140 mg/dL) and at other times BG <10 mmol/L (180 mg/dL).

The physician caring for an individual with diabetes in the perioperative period, during times of infection or serious physical illness, or simply when the patient is fasting for a diagnostic procedure must monitor the plasma glucose vigilantly, adjust the diabetes treatment regimen, and provide glucose infusion as needed. Hypoglycemia is frequent in hospitalized patients, and many of these episodes are avoidable. Measures to reduce or prevent hypoglycemia include frequent glucose monitoring and anticipating potential modifications of insulin/glucose administration because of changes in the clinical situation or treatment (tapering of glucocorticoids, etc.) or interruption of enteral or parenteral infusions or PO intake.

Depending on the severity of the patient's illness and the hospital setting, the physician can use either an insulin infusion or SC insulin. Insulin infusions are preferred in the ICU or in a clinically unstable setting. The absorption of SC insulin may be variable in such situations. Insulin infusions can also effectively control plasma glucose in the perioperative period and when the patient is unable to take anything by mouth. Regular insulin is preferred over insulin analogues for IV insulin infusion since it is less expensive and equally effective. The physician must consider carefully the clinical setting in which an insulin infusion will be utilized, including whether adequate ancillary personnel are available to monitor the plasma glucose frequently and whether they can adjust the insulin infusion rate to maintain the plasma glucose within the optimal range. Insulin-infusion algorithms should integrate the insulin sensitivity of the patient, frequent blood glucose monitoring, and the trend of changes in the blood glucose to determine the insulin-infusion rate. Insulin-infusion algorithms jointly developed and implemented by nursing and physician staff are advised. Because of the short half-life of IV regular insulin, it is necessary to administer long-acting insulin prior to discontinuation of the insulin infusion (2–4 h) to avoid a period of insulin deficiency.

In patients who are not critically ill or not in the ICU, basal or "scheduled" insulin is provided by SC, long-acting insulin supplemented by prandial and/or "corrective" insulin using a short-acting insulin (insulin analogues preferred). The use of "sliding scale," short-acting insulin alone, where no insulin is given unless the blood glucose is elevated, is inadequate for in-patient glucose management and should not be used. The short-acting, preprandial insulin dose should include coverage for food consumption (based on anticipated carbohydrate intake) plus a corrective or supplemental insulin based on the patient's insulin sensitivity and the blood glucose. For example, if the patient is thin (and likely insulin-sensitive), a corrective insulin supplement might be 1 unit for each 2.7 mmol/L (50 mg/dL) over the glucose target. If the patient is obese and insulin-resistant, then the insulin supplement might be 2 units for each 2.7 mmol/L (50 mg/dL) over the glucose target. It is critical to individualize the regimen and adjust the basal or "scheduled" insulin dose frequently, based on the corrective insulin required. A "consistent carbohydrate diabetes meal plan" for hospitalized patients provides a predictable amount of carbohydrate for a particular meal each day (but not necessarily the same amount for breakfast, lunch, and supper). The hospital diet should be determined by a nutritionist; terms such as *ADA diet* or *low-sugar diet* are no longer used.

Individuals with type 1 DM who are undergoing general anesthesia and surgery, or who are seriously ill, should receive continuous insulin, either through an IV insulin infusion or by SC administration of a reduced dose of long-acting insulin. Short-acting insulin alone is insufficient. Prolongation of a surgical procedure or delay in the recovery room is not uncommon and may result in periods of insulin deficiency leading to DKA. Insulin infusion is the preferred method for managing patients with type 1 DM in the perioperative period or when serious concurrent illness is present (0.5–1.0 units/h of regular insulin). If the diagnostic or surgical procedure is brief and performed under local or regional anesthesia, a reduced dose of SC, long-acting insulin may suffice (30–50% reduction, with short-acting insulin withheld or reduced). This approach facilitates the transition back to long-acting insulin after the procedure. Glucose may be infused to prevent hypoglycemia. The blood glucose should be monitored frequently during the illness or in the perioperative period.

Individuals with type 2 DM can be managed with either an insulin infusion or SC long-acting insulin (25–50% reduction depending on clinical setting) plus preprandial, short-acting insulin. Oral glucose-lowering agents should be discontinued upon admission and are not useful in regulating the plasma glucose in clinical situations where the insulin requirements and glucose intake are changing rapidly. Moreover, these oral agents may be dangerous if the patient is fasting (e.g., hypoglycemia with sulfonylureas). Metformin should be withheld when radiographic contrast media will be given or if severe CHF, acidosis, or declining renal function is present.

TOTAL PARENTERAL NUTRITION

(See also Chap. 76.) Total parenteral nutrition (TPN) greatly increases insulin requirements. In addition, individuals not previously known to have DM may become hyperglycemic during TPN and require insulin treatment. IV insulin infusion is the preferred treatment for hyperglycemia, and rapid titration to the required insulin dose is done most efficiently using a separate insulin infusion. After the total insulin dose has been determined, insulin may be added directly to the TPN solution or, preferably, given as a separate infusion. Often, individuals receiving either TPN or enteral nutrition receive their caloric loads continuously and not at "meal times"; consequently, SC insulin regimens must be adjusted.

GLUCOCORTICOIDS

Glucocorticoids increase insulin resistance, decrease glucose utilization, increase hepatic glucose production, and impair insulin secretion. These changes lead to a worsening of glycemic control in individuals with DM and may precipitate diabetes in other individuals ("steroid-induced diabetes"). The effects of glucocorticoids on glucose homeostasis are dose-related, usually reversible, and most pronounced in the postprandial period. If the FPG is near the normal range, oral diabetes agents (e.g., sulfonylureas, metformin) may be sufficient to reduce hyperglycemia. If the FPG >11.1 mmol/L (200 mg/dL), oral agents are usually not efficacious and insulin therapy is required. Short-acting insulin may be required to supplement long-acting insulin in order to control postprandial glucose excursions.

REPRODUCTIVE ISSUES

Reproductive capacity in either men or women with DM appears to be normal. Menstrual cycles may be associated with alterations in glycemic control in women with DM. Pregnancy is associated with marked insulin resistance; the increased insulin requirements often precipitate DM and lead to the diagnosis of GDM. Glucose, which at high levels is a teratogen to the developing fetus, readily crosses the placenta, but insulin does not. Thus, hyperglycemia from the maternal circulation may stimulate insulin secretion in the fetus. The anabolic and growth effects of insulin may result in macrosomia. GDM complicates ~ 7% of pregnancies in the United States. The incidence of GDM is greatly increased in certain ethnic groups, including African

Americans and Latinas, consistent with a similar increased risk of type 2 DM. Current recommendations advise screening for glucose intolerance between weeks 24 and 28 of pregnancy in women with increased risk for GDM (≥25 years; obesity; family history of DM; member of an ethnic group such as Latina, Native American, Asian American, African American, or Pacific Islander). Therapy for GDM is similar to that for individuals with pregnancy-associated diabetes and involves MNT and insulin, if hyperglycemia persists. Oral glucose-lowering agents are not approved for use during pregnancy, but studies using metformin or glyburide have shown efficacy and have not found toxicity. However, most physicians use insulin to treat GDM. With current practices, the morbidity and mortality rates of the mother with GDM and the fetus are not different from those in the nondiabetic population. Individuals who develop GDM are at marked increased risk for developing type 2 DM in the future and should be screened periodically for DM. Most individuals with GDM revert to normal glucose tolerance after delivery, but some will continue to have overt diabetes or impairment of glucose tolerance after delivery. In addition, children of women with GDM appear to be at risk for obesity and glucose intolerance and have an increased risk of diabetes beginning in the later stages of adolescence.

Pregnancy in individuals with known DM requires meticulous planning and adherence to strict treatment regimens. Intensive diabetes management and normalization of the A1C are essential for individuals with existing DM who are planning pregnancy. The most crucial period of glycemic control is soon after fertilization. The risk of fetal malformations is increased 4–10 times in individuals with uncontrolled DM at the time of conception, and normal plasma glucose during the preconception period and throughout the periods of organ development in the fetus should be the goal.

Lipodystrophic DM

Lipodystrophy, or the loss of subcutaneous fat tissue, may be generalized in certain genetic conditions such as leprechaunism. Generalized lipodystrophy is associated with severe insulin resistance and is often accompanied by acanthosis nigricans and dyslipidemia. Localized lipodystrophy associated with insulin injections has been reduced considerably by the use of human insulin.

PROTEASE INHIBITORS AND LIPODYSTROPHY

Protease inhibitors used in the treatment of HIV disease (Chap. 189) have been associated with a centripetal accumulation of fat (visceral and abdominal area), accumulation of fat in the dorsocervical region, loss of extremity fat, decreased insulin sensitivity (elevations of the fasting insulin level and reduced glucose tolerance on IV glucose tolerance testing), and dyslipidemia. Although many aspects of the physical appearance of these individuals resemble Cushing's syndrome, increased cortisol levels do not account for this appearance. The possibility remains that this is related to HIV infection by some undefined mechanism, since some features of the syndrome were observed before the introduction of protease inhibitors. Therapy for HIV-related lipodystrophy is not well established.

FURTHER READINGS

American Diabetes Association: Standards of medical care in diabetes. Diabetes Care 34:S11, 2011

Chan JC et al: Diabetes in Asia: Epidemiology, risk factors, and pathophysiology. JAMA 301:2129, 2009[PMID: 19470990] [Full Text]

Grant RW et al: Genetic architecture of type 2 diabetes: Recent progress and clinical implications. Diabetes Care 32:1107, 2009[PMID: 19460916] [Full Text]

Holman RR et al: 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 359:1577, 2008[PMID: 18784090] [Full Text]

Kahn SE et al: Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med 355:2427, 2006[PMID: 17145742] [Full Text]

Kitabchi AE et al: Hyperglycemic crises in adult patients with diabetes. Diabetes Care 32:1335, 2009[PMID: 19564476] [Full Text]

Nathan DM et al: Modern-day clinical course of type 1 diabetes mellitus after 30 years' duration: The diabetes control and complications trial/epidemiology of diabetes interventions and complications and Pittsburgh epidemiology of diabetes complications experience (1983–2005). Arch Intern Med 169:1307, 2009[PMID: 19636033] [Full Text]

Vaxillaire M, Froguel P: Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. Endocr Rev 29:254, 2008[PMID: 18436708] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HYPOGLYCEMIA: INTRODUCTION

Hypoglycemia is most commonly caused by drugs used to treat diabetes mellitus or by exposure to other drugs, including alcohol. However, a number of other disorders, including critical organ failure, sepsis and inanition, hormone deficiencies, non–beta-cell tumors, insulinoma, and prior gastric surgery, may cause hypoglycemia **(Table 345-1)**. Hypoglycemia is most convincingly documented by *Whipple's triad:* (1) symptoms consistent with hypoglycemia, (2) a low plasma glucose concentration measured with a precise method (not a glucose monitor), and (3) relief of those symptoms after the plasma glucose level is raised. The lower limit of the fasting plasma glucose concentration is normally approximately 70 mg/dL (3.9 mmol/L), but substantially lower venous glucose levels occur normally, late after a meal. Glucose levels <55 mg/dL (3.0 mmol/L) with symptoms that are relieved promptly after the glucose level is raised document hypoglycemia. Hypoglycemia can cause serious morbidity; if severe and prolonged, it can be fatal. It should be considered in any patient with episodes of confusion, an altered level of consciousness, or a seizure.

Table 345-1 Causes of Hypoglycemia in Adults

Ill or medicated individual
1. Drugs
Insulin or insulin secretagogue
Alcohol
Others
2. Critical illness
Hepatic, renal or cardiac failure
Sepsis
Inanition
3. Hormone deficiency
Cortisol
Glucagon and epinephrine (in insulin-deficient diabetes)
4. Non–islet cell tumor
Seemingly well individual
5. Endogenous hyperinsulinism
Insulinoma
Functional beta-cell disorders (nesidioblastosis)
Noninsulinoma pancreatogenous hypoglycemia
Post–gastric bypass hypoglycemia
Insulin autoimmune hypoglycemia
Antibody to insulin
Antibody to insulin receptor
Insulin secretagogue
Other
6. Accidental, surreptitious or malicious hypoglycemia

Source: From PE Cryer et al: J Clin Endocrinol Metab 94:709, 2009. © The Endocrine Society, 2009.

SYSTEMIC GLUCOSE BALANCE AND GLUCOSE COUNTERREGULATION

Glucose is an obligate metabolic fuel for the brain under physiologic conditions. The brain cannot synthesize glucose or store more than a few minutes' supply as glycogen and therefore requires a continuous supply of glucose from the arterial circulation. As the arterial plasma glucose concentration falls below the physiologic range, blood-to-brain glucose transport becomes insufficient to support brain energy metabolism and function. However, redundant glucose counterregulatory mechanisms

normally prevent or rapidly correct hypoglycemia.

Plasma glucose concentrations are normally maintained within a relatively narrow range, roughly 70–110 mg/dL (3.9–6.1 mmol/L) in the fasting state with transient higher excursions after a meal, despite wide variations in exogenous glucose delivery from meals and in endogenous glucose utilization by, for example, exercising muscle. Between meals and during fasting, plasma glucose levels are maintained by endogenous glucose production, hepatic glycogenolysis, and hepatic (and renal) gluconeogenesis **(Fig. 345-1)**. Although hepatic glycogen stores are usually sufficient to maintain plasma glucose levels for approximately 8 h, this time period can be shorter if glucose demand is increased by exercise or if glycogen stores are depleted by illness or starvation.

Figure 345-1

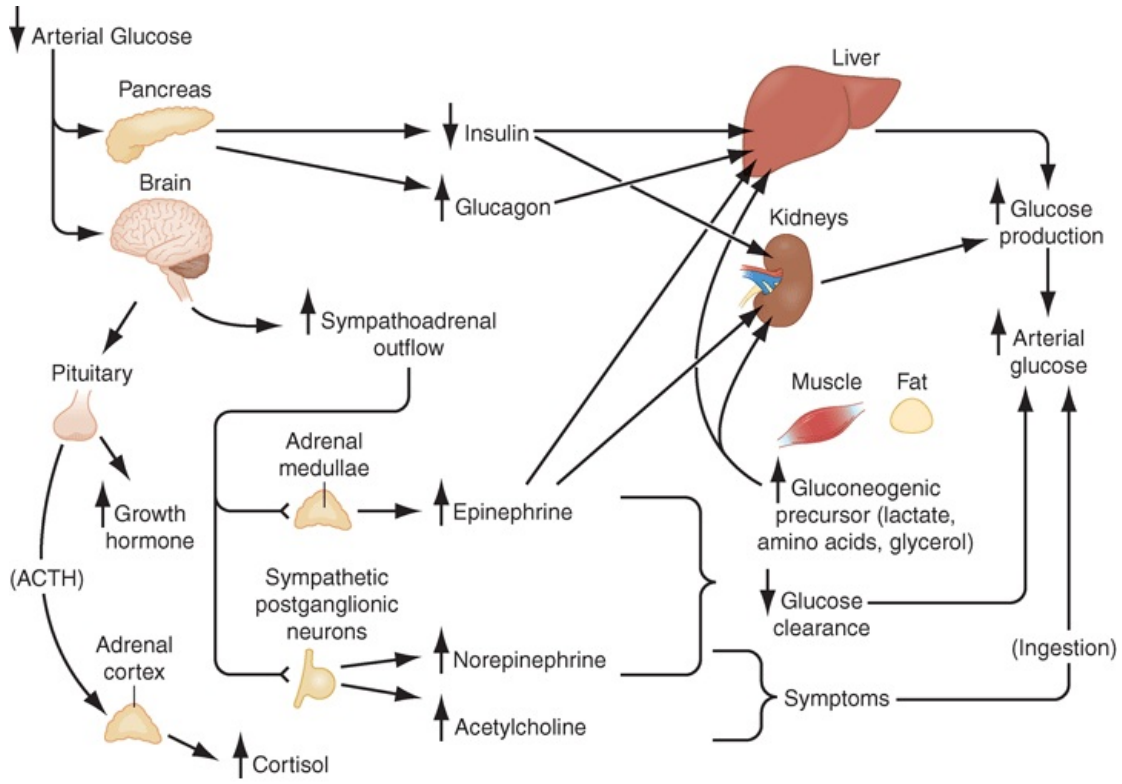

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Physiology of glucose counterregulation—the mechanisms that normally prevent or rapidly correct hypoglycemia. In insulin-deficient diabetes, the key counterregulatory responses—suppression of insulin and increase of glucagon—are lost, and the stimulation of sympathoadrenal outflow is attenuated.

Gluconeogenesis normally requires low insulin levels and the presence of anti-insulin (counterregulatory) hormones, together with a coordinated supply of precursors from muscle and adipose tissue to the liver (and kidneys). Muscle provides lactate, pyruvate, alanine, glutamine, and other amino acids. Triglycerides in adipose tissue are broken down into fatty acids and glycerol, which is a gluconeogenic precursor. Fatty acids provide an alternative oxidative fuel to tissues other than the brain (which requires glucose).

Systemic glucose balance—maintenance of the normal plasma glucose concentration—is accomplished by a network of hormones, neural signals, and substrate effects that regulate endogenous glucose production and glucose utilization by tissues other than the brain (Chap. 344). Among the regulatory factors, insulin plays a dominant role (**Table 345-2**; Fig. 345-1). As plasma glucose levels decline within the physiologic range in the fasting state, pancreatic beta-cell insulin secretion decreases, thereby increasing hepatic glycogenolysis and hepatic (and renal) gluconeogenesis. Low insulin levels also reduce glucose utilization in peripheral tissues, inducing lipolysis and proteolysis, thereby releasing gluconeogenic precursors. Thus, a decrease in insulin secretion is the first defense against hypoglycemia.

Table 345-2 Physiologic Responses to Decreasing Plasma Glucose Concentrations

Response	Glycemic Threshold, mmol/L	Physiologic Effects	Role in the Prevention or Correction of Hypoglycemia (Glucose Counterregulation)

	(mg/dL)		
↓ Insulin	4.4–4.7 (80–85)	↑ R_a (↓ R_d)	Primary glucose regulatory factor/first defense against hypoglycemia
↑ Glucagon	3.6–3.9 (65–70)	↑ R_a	Primary glucose counterregulatory factor/second defense against hypoglycemia
↑ Epinephrine	3.6–3.9 (65–70)	↑ R_a, ↓ R_c	Third defense against hypoglycemia, critical when glucagon is deficient
↑ Cortisol and growth hormone	3.6–3.9 (65–70)	↑ R_a, ↓ R_c	Involved in defense against prolonged hypoglycemia, not critical
Symptoms	2.8–3.1 (50–55)	Recognition of hypoglycemia	Prompt behavioral defense against hypoglycemia (food ingestion)
↓ Cognition	<2.8 (<50)	—	(Compromises behavioral defense against hypoglycemia)

Note: R_a, rate of glucose appearance, glucose production by the liver and kidneys; R_c, rate of glucose clearance, glucose utilization relative to the ambient plasma glucose concentration; R_d, rate of glucose disappearance, glucose utilization by the brain (which is unaltered by the glucoregulatory hormones) and by insulin-sensitive tissues such as skeletal muscle (which is regulated by insulin, epinephrine, cortisol, and growth hormone). ***Source:*** From PE Cryer: Hypoglycemia, in *Williams Textbook of Endocrinology,* 11th ed, HM Kronenberg et al (eds). Philadelphia, Saunders, 2008.

As plasma glucose levels decline just below the physiologic range, glucose counterregulatory (plasma glucose–raising) hormones are released (Table 345-2; Fig. 345-1). Among these, pancreatic α-cell glucagon, which stimulates hepatic glycogenolysis, plays a primary role. Glucagon is the second defense against hypoglycemia. Adrenomedullary epinephrine, which stimulates hepatic glycogenolysis and gluconeogenesis (and renal gluconeogenesis), is not normally critical. However, it becomes critical when glucagon is deficient. Epinephrine is the third defense against hypoglycemia. When hypoglycemia is prolonged beyond ~4 hours, cortisol and growth hormone also support glucose production and limit glucose utilization, albeit only ~20% of that of epinephrine.

As plasma glucose levels fall to lower levels, symptoms prompt the behavioral defense against hypoglycemia, including the ingestion of food (Table 345-2; Fig. 345-1).

The normal glycemic thresholds for these responses to decreasing plasma glucose concentrations are shown in Table 345-2. However, these thresholds are dynamic. They shift to higher-than-normal glucose levels in people with poorly controlled diabetes who can experience symptoms of hypoglycemia when their glucose levels decline into the normal range. On the other hand, they shift to lower-than-normal glucose levels in people with recurrent hypoglycemia, e.g., those with aggressively treated diabetes or an insulinoma. Such patients have symptoms at glucose levels lower than those that cause symptoms in healthy individuals.

Clinical Manifestations

Neuroglycopenic symptoms of hypoglycemia are the direct result of central nervous system (CNS) glucose deprivation. They include behavioral changes, confusion, fatigue, seizure, loss of consciousness, and, if hypoglycemia is severe and prolonged, death. Neurogenic (or autonomic) symptoms of hypoglycemia are the result of the perception of physiologic changes caused by the CNS-mediated sympathoadrenal discharge triggered by hypoglycemia. They include adrenergic symptoms (mediated largely by norepinephrine released from sympathetic postganglionic neurons but perhaps also by epinephrine released from the adrenal medullae) such as palpitations, tremor, and anxiety. They also include cholinergic symptoms (mediated by acetylcholine released from sympathetic postganglionic neurons) such as sweating, hunger, and paresthesias. Clearly, these are nonspecific symptoms. Their attribution to hypoglycemia requires a corresponding low plasma glucose concentration and their resolution after the glucose level is raised (Whipple's triad).

Common signs of hypoglycemia include diaphoresis and pallor. Heart rate and systolic blood pressure are typically increased but may not be raised in an individual who has experienced repeated, recent episodes of hypoglycemia. Neuroglycopenic manifestations are often observable. Transient focal neurologic deficits occur occasionally. Permanent neurologic deficits are rare.

Etiology and Pathophysiology

Hypoglycemia is most commonly a result of the treatment of diabetes. This topic is therefore addressed before considering other causes of hypoglycemia.

Hypoglycemia in Diabetes

IMPACT AND FREQUENCY

Hypoglycemia is the limiting factor in the glycemic management of diabetes. First, it causes recurrent morbidity in most people with type 1 diabetes (T1DM) and many with advanced type 2 diabetes

(T2DM) and is sometimes fatal. Second, it precludes maintenance of euglycemia over a lifetime of diabetes and thus full realization of the well-established vascular benefits of glycemic control. Third, it causes a vicious cycle of recurrent hypoglycemia by producing hypoglycemia-associated autonomic failure—the clinical syndromes of defective glucose counterregulation and of hypoglycemia unawareness.

Hypoglycemia is a fact of life for people with T1DM. They suffer an average of two episodes of symptomatic hypoglycemia per week and at least one episode of severe, at least temporarily disabling, hypoglycemia each year. An estimated 6–10% of people with T1DM die as a result of hypoglycemia. Hypoglycemia occurs less frequently in T2DM. However, its prevalence in insulin-requiring T2DM is greater than commonly appreciated, Recent studies investigating insulin pump or multiple injection therapies have revealed a hypoglycemia prevalence approaching 70%. Metformin, thiazolidinediones, α-glucosidase inhibitors, glucagon-like peptide-1 (GLP-1) receptor agonists, and dipeptidyl peptidase-IV (DPP-IV) inhibitors should not cause hypoglycemia. However, they increase the risk when combined with an insulin secretagogue, such as one of the sulfonylureas or glinides, or with insulin. Notably, the frequency of hypoglycemia approaches that in T1DM as persons with T2DM develop absolute insulin deficiency and require more complex treatment with insulin.

CONVENTIONAL RISK FACTORS

The conventional risk factors for hypoglycemia in diabetes are based on the premise that relative or absolute insulin excess is the sole determinant of risk. Relative or absolute insulin excess occurs when (1) insulin (or insulin secretagogue) doses are excessive, ill-timed, or of the wrong type; (2) the influx of exogenous glucose is reduced (e.g., during an overnight fast or following missed meals or snacks); (3) insulin-independent glucose utilization is increased (e.g., during exercise); (4) sensitivity to insulin is increased (e.g., with improved glycemic control, in the middle of the night, late after exercise, or with increased fitness or weight loss); (5) endogenous glucose production is reduced (e.g., following alcohol ingestion); and (6) insulin clearance is reduced (e.g., in renal failure). However, these conventional risk factors alone explain a minority of episodes; other factors are typically involved.

HYPOGLYCEMIA-ASSOCIATED AUTONOMIC FAILURE

While marked insulin excess alone can cause hypoglycemia, iatrogenic hypoglycemia in diabetes is typically the result of the interplay of relative or absolute therapeutic insulin excess and compromised physiologic and behavioral defenses against falling plasma glucose concentrations (Table 345-2; **Fig. 345-2**). Defective glucose counterregulation compromises physiologic defense (particularly decrements in insulin and increments in glucagon and epinephrine), and hypoglycemia unawareness compromises behavioral defense (ingestion of carbohydrate).

Figure 345-2

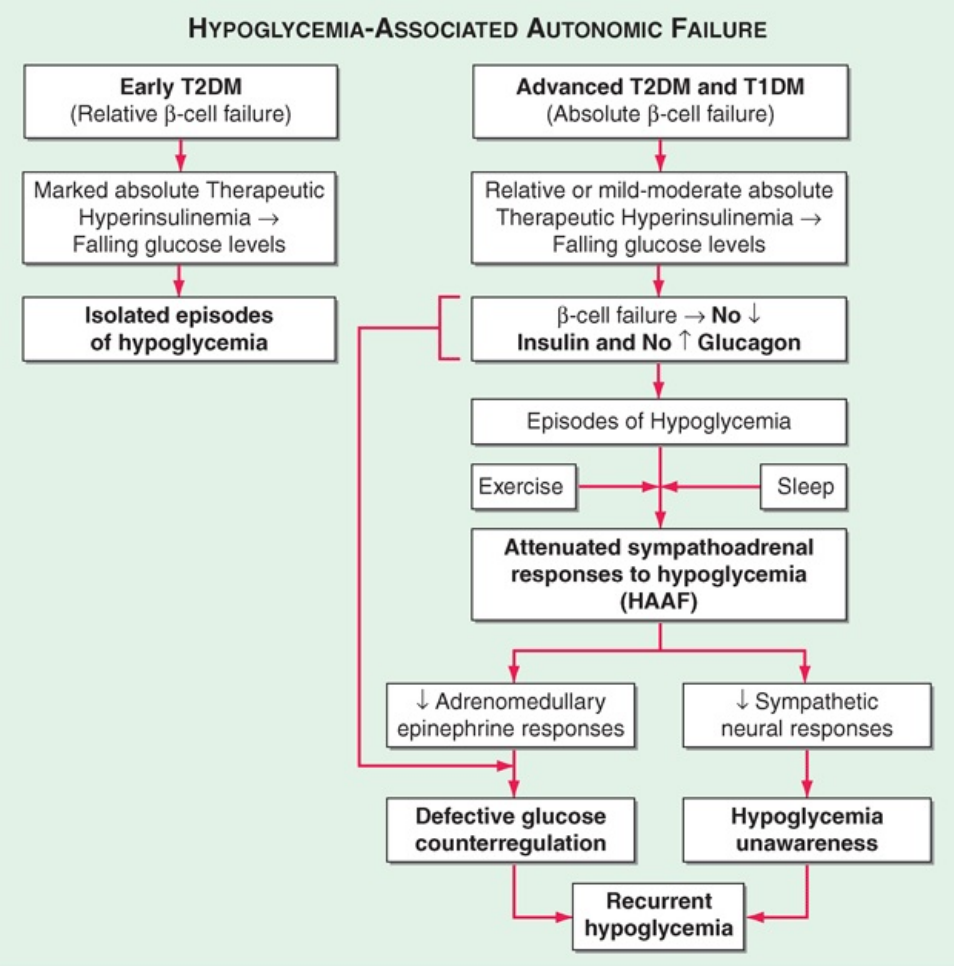

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypoglycemia-associated autonomic failure in insulin-deficient diabetes.

Defective Glucose Counterregulation

In the setting of absolute endogenous insulin deficiency, insulin levels do not decrease as plasma glucose levels fall; the first defense against hypoglycemia is lost. Furthermore, probably because the decrement in intraislet insulin is normally a signal to stimulate glucagon secretion, glucagon levels do not increase as plasma glucose levels fall further; the second defense against hypoglycemia is lost. Finally, the increase in epinephrine levels, the third defense against hypoglycemia, in response to a given level of hypoglycemia is typically attenuated. The glycemic threshold for the sympathoadrenal (adrenomedullary epinephrine and sympathetic neural norepinephrine) response is shifted to lower plasma glucose concentrations. That is typically the result of recent antecedent iatrogenic hypoglycemia. In the setting of absent decrements in insulin and of absent increments in glucagon, the attenuated increment in epinephrine causes the clinical syndrome of defective glucose counterregulation. Affected patients are at twenty-five-fold or greater increased risk of severe iatrogenic hypoglycemia during aggressive glycemic therapy of their diabetes compared with those with normal epinephrine responses. This functional, and reversible, disorder is distinct from classical diabetic autonomic neuropathy, a structural and irreversible disorder.

Hypoglycemia Unawareness

The attenuated sympathoadrenal response (largely the reduced sympathetic neural response) to hypoglycemia causes the clinical syndrome of hypoglycemia unawareness, i.e., loss of the warning adrenergic and cholinergic symptoms that previously allowed the patient to recognize developing hypoglycemia and therefore abort the episode by ingesting carbohydrates. Affected patients are at a sixfold increased risk of severe iatrogenic hypoglycemia during aggressive glycemic therapy of their diabetes.

Hypoglycemia-Associated Autonomic Failure

The concept of hypoglycemia-associated autonomic failure (HAAF) in diabetes posits that recent antecedent iatrogenic hypoglycemia (or sleep or prior exercise) causes both defective glucose counterregulation (by reducing the epinephrine response to a given level of subsequent hypoglycemia in the setting of absent insulin and glucagon responses) and hypoglycemia unawareness (by reducing

the sympathoadrenal response to a given level of subsequent hypoglycemia). These impaired responses create a vicious cycle of recurrent iatrogenic hypoglycemia (Fig. 345-2). Hypoglycemia unawareness, and to some extent the reduced epinephrine component of defective glucose counterregulation, is reversible by as little as 2–3 weeks of scrupulous avoidance of hypoglycemia in most affected patients.

Based on this pathophysiology, additional risk factors for hypoglycemia in diabetes include (1) absolute insulin deficiency that indicates that insulin levels will not decrease and glucagon levels will not increase as plasma glucose levels fall; (2) a history of severe hypoglycemia or of hypoglycemia unawareness, implying recent antecedent hypoglycemia, as well as prior exercise or sleep, that indicates that the sympathoadrenal response will be attenuated; and (3) lower HbA_{1C} levels or lower glycemic goals that, all other factors being equal, increase the probability of recent antecedent hypoglycemia.

HYPOGLYCEMIA RISK FACTOR REDUCTION

Several recent multicenter randomized controlled trials investigating the potential benefits of tight glucose control in either inpatient or outpatient settings have reported a high prevalence of severe hypoglycemia. In the NICE-SUGAR study, attempts to control in-hospital plasma glucose values toward physiologic levels resulted in increased mortality. The ADVANCE, ACCORD, and VADT studies also reported a significant incidence of severe hypoglycemia in T2DM individuals. Somewhat surprisingly, all three studies found little or no benefit of intensive glucose control to reduce macrovascular events in T2DM. In fact, the ACCORD study was ended early due to increased mortality in the intensive glucose control arm. Whether iatrogenic hypoglycemia was the cause of the increased mortality is not known. In the light of the above findings, new recommendations and paradigms have been formulated. Whereas there is little debate regarding the need to reduce hyperglycemia in hospitalized patients, the glycemic maintenance goals have been modified to fall between 140 and 180 mg/dL. Thus, the benefits from insulin therapy and reduced hyperglycemia can be obtained while reducing the prevalence of hypoglycemia.

Similarly, evidence exists that intensive glucose control can reduce the prevalence of microvascular disease in both T1DM and T2DM. These benefits need to be weighed against the increased prevalence of hypoglycemia. The level of glucose control (HBA_{1C}) should be evaluated for each patient. Multicenter trials have demonstrated that patients with recently diagnosed T1DM or T2DM can have better glycemic control with less hypoglycemia. In addition, there is still long-term benefit in reducing HBA_{1C} from higher to lower levels, albeit still above recommended levels. Perhaps a reasonable therapeutic goal is the lowest HBA_{1C} that does not cause severe hypoglycemia and preserves awareness of hypoglycemia.

Pancreatic transplantation (both whole organ and islet cells) has been used as a treatment option for recurrent severe hypoglycemia. Generally, rates of hypoglycemia are reduced following transplantation. This appears to be due to more physiologic insulin and glucagon responses during hypoglycemia following whole organ transplant and resolution of insulin modulation following islet cell transplantation.

The use of continuous glucose monitors appears promising as a method of reducing hypoglycemia while improving HBA_{1C}. Other interventions to stimulate counterregulatory responses such as terbutaline, fluoxetine, thiazolidinediones or fructose remain experimental and have not been subjected to large-scale clinical trials.

Thus, intensive glycemic therapy (Chap. 344) needs to be applied in combination with patient education and empowerment; frequent self-monitoring of blood glucose; flexible insulin (and other drug) regimens including the use of insulin analogues (both short and longer acting); individualized glycemic goals; ongoing professional guidance and support; and consideration of both the conventional risk factors and those indicative of compromised glucose counterregulation. Given a history of hypoglycemia unawareness, a 2- to 3-week period of scrupulous avoidance of hypoglycemia is indicated.

Hypoglycemia Without Diabetes

There are many causes of hypoglycemia (Table 345-1). Because hypoglycemia is common in insulin- or insulin secretagogue–treated diabetes, it is often reasonable to assume that a clinically suspicious episode was the result of hypoglycemia. On the other hand, because hypoglycemia is rare in the absence of relevant drug–treated diabetes, it is reasonable to conclude that a hypoglycemic disorder is present only in patients in whom Whipple's triad can be demonstrated.

Particularly in patients who are ill or medicated, the initial diagnostic considerations should focus should on medications, and then critical illnesses, hormone deficiency, or nonislet cell tumor hypoglycemia. In the absence of any of these and in a seemingly well individual, the focus should shift to the possibilities of endogenous hyperinsulinism or accidental, surreptitious, or even malicious hypoglycemia.

DRUGS

Insulin and insulin secretagogues suppress glucose production and stimulate glucose utilization. Ethanol blocks gluconeogenesis but not glycogenolysis. Thus, alcohol-induced hypoglycemia typically occurs after a several-day ethanol binge during which the person eats little food, thereby causing glycogen depletion. Ethanol is usually measurable in blood at the time of presentation, but its levels correlate poorly with plasma glucose concentrations. Because gluconeogenesis becomes the predominant route of glucose production during prolonged hypoglycemia, alcohol can contribute to the progression of hypoglycemia in patients with insulin-treated diabetes.

A large number of other drugs have been associated with hypoglycemia. These include commonly used drugs such as angiotensin-converting enzyme inhibitors and angiotensin receptor antagonists, ß-adrenergic receptor antagonists, quinolone antibiotics, indomethacin, quinine, and sulfonamides among many others.

CRITICAL ILLNESS

Among hospitalized patients, serious illnesses such as renal, hepatic, or cardiac failure, sepsis, and inanition are second only to drugs as causes of hypoglycemia.

Rapid and extensive hepatic destruction (e.g., toxic hepatitis) causes fasting hypoglycemia because the liver is the major site of endogenous glucose production. The mechanism of hypoglycemia in patients with cardiac failure is unknown. It may involve hepatic congestion and hypoxia. Although the kidneys are a source of glucose production, hypoglycemia in patients with renal failure is also caused by the reduced clearance of insulin and reduced mobilization of gluconeogenic precursors in renal failure.

Sepsis is a relatively common cause of hypoglycemia. Increased glucose utilization is induced by cytokine production in macrophage-rich tissues such as the liver, spleen, and lung. Hypoglycemia develops if glucose production fails to keep pace. Cytokine-induced inhibition of gluconeogenesis in the setting of nutritional glycogen depletion, in combination with hepatic and renal hypoperfusion, may also contribute to hypoglycemia.

Hypoglycemia can be seen with starvation, perhaps because of loss of whole-body fat stores and subsequent depletion of gluconeogenic precursors (e.g., amino acids), necessitating increased glucose utilization.

HORMONE DEFICIENCIES

Neither cortisol nor growth hormone is critical to the prevention of hypoglycemia, at least in adults. Nonetheless, hypoglycemia can occur with prolonged fasting in patients with primary adrenocortical failure (Addison's disease) or hypopituitarism. Anorexia and weight loss are typical features of chronic cortisol deficiency and likely result in glycogen depletion. Cortisol deficiency is associated with impaired gluconeogenesis and low levels of gluconeogenic precursors, suggesting that substrate-limited gluconeogenesis, in the setting of glycogen depletion, is the cause of hypoglycemia. Growth hormone deficiency can cause hypoglycemia in young children. In addition to extended fasting, high rates of glucose utilization (e.g., during exercise or in pregnancy) or low rates of glucose production (e.g., following alcohol ingestion) can precipitate hypoglycemia in adults with previously unrecognized hypopituitarism.

Hypoglycemia is not a feature of the epinephrine-deficient state that results from bilateral adrenalectomy, when glucocorticoid replacement is adequate, nor does it occur during pharmacologic adrenergic blockade when other glucoregulatory systems are intact. Combined deficiencies of glucagon and epinephrine play a key role in the pathogenesis of iatrogenic hypoglycemia in people with insulin-deficient diabetes as discussed earlier. Otherwise, deficiencies of these hormones are not usually considered in the differential diagnosis of a hypoglycemic disorder.

NON–BETA-CELL TUMORS

Fasting hypoglycemia, often termed non–islet cell tumor hypoglycemia, occurs occasionally in patients with large mesenchymal or epithelial tumors (e.g., hepatomas, adrenocortical carcinomas, carcinoids). The glucose kinetic patterns resemble those of hyperinsulinism (see below), but insulin secretion is suppressed appropriately during hypoglycemia. In most instances, hypoglycemia is due to overproduction of an incompletely processed form of insulin-like growth factor II ("big IGF-II") that does not complex normally with circulating binding proteins and thus more readily gains access to target tissues. The tumors are usually apparent clinically, plasma IGF-II to IGF-I ratios are high and free IGF-II levels [and levels of pro-IGF-II (E1-21)] are elevated. Curative surgery is seldom possible, but reduction of tumor bulk may ameliorate hypoglycemia. Therapy with a glucocorticoid, growth hormone, or both has also been reported to alleviate hypoglycemia. Hypoglycemia attributed to ectopic IGF-I production has been reported but is rare.

ENDOGENOUS HYPERINSULINISM

Hypoglycemia due to endogenous hyperinsulinism can be caused by (1) a primary beta-cell disorder, typically a beta-cell tumor (insulinoma), sometimes multiple insulinomas, or a functional beta-cell

disorder with beta-cell hypertrophy or hyperplasia; (2) an antibody to insulin or to the insulin receptor; (3) a beta-cell secretagogue such as a sulfonylurea; or (4) perhaps ectopic insulin secretion among other very rare mechanisms. None of these causes is common.

The fundamental pathophysiologic feature of endogenous hyperinsulinism caused by a primary beta-cell disorder or an insulin secretagogue is the failure of insulin secretion to fall to very low levels during hypoglycemia. This is assessed by measuring plasma insulin, C-peptide (the connecting peptide that is cleaved from proinsulin to produce insulin), proinsulin, and glucose concentrations during hypoglycemia. Insulin, C-peptide, and proinsulin levels need not be high relative to normal, euglycemic values; they are inappropriately high in the setting of a low plasma glucose concentration. Critical diagnostic findings are a plasma insulin concentration ≥3 μU/mL (≥18 pmol/L), a plasma C-peptide concentration ≥0.6 ng/mL (≥0.2 nmol/L) and a plasma proinsulin concentration ≥5.0 pmol/L when the plasma glucose concentration is <55 mg/dL (<3 mmol/L) with symptoms of hypoglycemia. A low plasma β-hydroxybutyrate concentration (≥2.7 mmol/L) and an increment in plasma glucose >25 mg/dL (1.4 mmol/L) following intravenous glucagon (1 mg) indicate increased insulin (or insulin-like growth factor) actions.

The diagnostic strategy is to obtain measurements of plasma glucose, insulin, C-peptide, proinsulin and β-hydroxybutyrate concentrations—and to screen for circulating oral hypoglycemic agents—during an episode of hypoglycemia and to assess symptoms during the episode and seek their resolution following correction of hypoglycemia by intravenous injection of glucagon (i.e., to document Whipple's triad). This approach is straightforward if the patient is hypoglycemic during evaluation. Since endogenous hyperinsulinemic disorders usually, but not invariably, cause fasting hypoglycemia, a diagnostic episode may develop after a relatively short outpatient fast. Serial sampling during an up to 72-h inpatient diagnostic fast or following a mixed meal is more problematic. An alternative is to give the patient a detailed list of the required measurements and ask them to present to an emergency room, with the list, during a symptomatic episode. Obviously, a normal plasma glucose concentration during a symptomatic episode indicates that the symptoms are not the result of hypoglycemia.

An insulinoma, an insulin-secreting pancreatic islet beta-cell tumor, is the prototypical cause of endogenous hyperinsulinism and, therefore, should be sought in patients with the clinical syndrome. However, insulinoma is not the only cause of endogenous hyperinsulinism. Some patients with fasting endogenous hyperinsulinemic hypoglycemia have diffuse islet involvement with beta-cell hypertrophy and sometimes hyperplasia. This pattern is commonly referred to as nesidioblastosis, although the finding of beta-cells budding from ducts is not invariably present. Other patients have a similar islet pattern but postprandial hypoglycemia, a disorder termed noninsulinoma pancreatogenous hypoglycemia. Post–gastric bypass postprandial hypoglycemia also involves diffuse islet involvement and endogenous hyperinsulinism. It most often follows Roux en Y gastric bypass. Some have suggested that exaggerated GLP-1 responses to meals may cause hyperinsulinemia and hypoglycemia, but the pathogenesis has not been clearly established. If medical treatments such as an α-glucosidase inhibitor or octreotide fail, partial pancreatectomy may be required. Autoimmune hypoglycemias include those caused by an antibody to insulin, which gradually disassociates, leading to late postprandial hypoglycemia. Alternatively, an insulin receptor antibody can function as an agonist. The presence of an insulin secretagogue, such as a sulfonylurea or a glinide, results in a clinical and biochemical pattern similar to that of an insulinoma but can be distinguished by the presence of the circulating secretagogue. Finally, very rare phenomena include ectopic insulin secretion, a gain of function insulin receptor mutation, and exercise-induced hyperinsulinemia.

Insulinomas are uncommon—the yearly incidence is estimated to be 1 in 250,000—but because more than 90% are benign, they are a treatable cause of potentially fatal hypoglycemia. The median age at presentation is 50 years in sporadic cases, but it usually presents in the third decade when it is a component of multiple endocrine neoplasia type 1 (Chap. 351). More than 99% of insulinomas are within the substance of the pancreas and they are usually small (90% <2.0 cm). Therefore, they come to clinical attention because of hypoglycemia rather than mass effects. CT or MRI detects approximately 70–80% of insulinomas. These methods detect metastases in the roughly 10% of patients with a malignant insulinoma. Transabdominal ultrasound will often identify insulinomas, and endoscopic ultrasound has a sensitivity of about 90%. Somatostatin receptor scintigraphy is thought to detect insulinomas in about half of patients. Selective pancreatic arterial calcium injections, with the endpoint of a sharp increase in hepatic venous insulin levels, regionalizes insulinomas with high sensitivity, but this invasive procedure is seldom necessary except to confirm endogenous hyperinsulinism in the diffuse islet disorders. Intraoperative pancreatic ultrasonography almost invariably localizes insulinomas that are not readily palpable by the surgeon. Surgical resection of a solitary insulinoma is generally curative. Diazoxide, which inhibits insulin secretion, or the somatostatin analogue octreotide can be used to treat hypoglycemia in patients with unresectable tumors; everolimus, an mTOR (mammalian target of rapamycin) inhibitor, is promising.

Accidental, Surreptitious or Malicious Hypoglycemia

Accidental ingestion of an insulin secretagogue (e.g., the result of a pharmacy or other medical error), or administration of insulin, can occur. Factitious hypoglycemia, caused by surreptitious or even

malicious administration of insulin or an insulin secretagogue, shares many clinical and laboratory features with insulinoma. It is most common among health care workers, patients with diabetes or their relatives, and people with a history of other factitious illnesses. However, it should be considered in all patients being evaluated for hypoglycemia of obscure cause. Ingestion of an insulin secretagogue causes hypoglycemia with increased C-peptide levels whereas exogenous insulin causes hypoglycemia with low C-peptide levels, reflecting suppression of insulin secretion.

Analytical error in the measurement of plasma glucose concentrations is rare. On the other hand, glucose monitors used to guide treatment of diabetes are not quantitative instruments, particularly at low glucose levels, and these should not be used for the definitive diagnosis of hypoglycemia. Even with a quantitative method, low measured glucose concentrations can be artifactual, e.g., the result of continued glucose metabolism by the formed elements of the blood ex vivo, particularly in the presence of leukocytosis, erythrocytosis, or thrombocytosis, or if separation of the serum from the formed elements is delayed (pseudohypoglycemia).

Approach to the Patient: Hypoglycemia

In addition to recognition and documentation of hypoglycemia, and often urgent treatment, diagnosis of the hypoglycemic mechanism is critical for choosing a treatment that prevents, or at least minimizes, recurrent hypoglycemia.

RECOGNITION AND DOCUMENTATION

Hypoglycemia is suspected in patients with typical symptoms; in the presence of confusion, an altered level of consciousness, or a seizure; or in a clinical setting in which hypoglycemia is known to occur. Blood should be drawn, whenever possible, before the administration of glucose to allow documentation of a low plasma glucose concentration. Convincing documentation of hypoglycemia requires the fulfillment of Whipple's triad. Thus, the ideal time to measure the plasma glucose level is during a symptomatic episode. A normal glucose level excludes hypoglycemia as the cause of the symptoms. A low glucose level confirms that hypoglycemia is the cause of the symptoms, provided the latter resolve after the glucose level is raised. When the cause of the hypoglycemic episode is obscure, additional measurements, while the glucose level is low and before treatment, should include plasma insulin, C-peptide, proinsulin and β-hydroxybutyrate levels, as well as screening for circulating oral hypoglycemic agents, and symptoms should be assessed during and after the plasma glucose concentration is raised.

When the history suggests prior hypoglycemia, and a potential mechanism is not apparent, the diagnostic strategy is to measure these values, and assess for Whipple's triad, during and after an episode of hypoglycemia. On the other hand, while it cannot be ignored, a distinctly low plasma glucose concentration measured in a patient without corresponding symptoms raises the possibility of an artifact (pseudohypoglycemia).

DIAGNOSIS OF THE HYPOGLYCEMIC MECHANISM

In a patient with documented hypoglycemia, a plausible hypoglycemic mechanism can often be deduced from the history, physical examination, and available laboratory data (Table 345-1). Drugs, particularly those used to treat diabetes or alcohol, should be the first consideration, even in the absence of known use of a relevant drug, given the possibility of surreptitious, accidental, or malicious drug administration. Other considerations include evidence of a relevant critical illness, less commonly hormone deficiencies, and rarely a non–beta-cell tumor that can be pursued diagnostically. Absent one of these mechanisms, in an otherwise seemingly well individual, one should consider endogenous hyperinsulinism and proceed with measurements and assessment of symptoms during spontaneous hypoglycemia or under conditions that might elicit hypoglycemia.

URGENT TREATMENT

Oral treatment with glucose tablets or glucose-containing fluids, candy, or food is appropriate if the patient is able and willing to take these. A reasonable initial dose is 20 g of glucose. If the patient is unable or unwilling, because of neuroglycopenia, to take carbohydrates orally, parenteral therapy is necessary. Intravenous glucose (25 g) should be given and followed by a glucose infusion guided by serial plasma glucose measurements. If intravenous therapy is not practical, subcutaneous or intramuscular glucagon (1.0 mg in adults) can be used, particularly in patients with T1DM. Because it acts by stimulating glycogenolysis, glucagon is ineffective in glycogen-depleted individuals (e.g., those with alcohol-induced hypoglycemia). It also stimulates insulin secretion and is therefore less useful in T2DM. These treatments raise plasma glucose concentrations only transiently, and patients should therefore be urged to eat as soon as is practical to replete glycogen stores.

PREVENTION OF RECURRENT HYPOGLYCEMIA

Prevention of recurrent hypoglycemia requires an understanding of the hypoglycemic mechanism. Offending drugs can be discontinued or their doses reduced. Hypoglycemia caused by a sulfonylurea can persist for hours, or even days. Underlying critical illnesses can often be treated. Cortisol and growth hormone can be replaced if they are deficient. Surgical, radiotherapeutic, or chemotherapeutic reduction of a nonislet cell tumor can alleviate hypoglycemia even if the tumor cannot be cured;

glucocorticoid or growth hormone administration also may reduce hypoglycemic episodes in such patients. Surgical resection of an insulinoma is curative; medical therapy with diazoxide or octreotide can be used if resection is not possible and in patients with a nontumor beta-cell disorder. Partial pancreatectomy may be necessary in the latter patients. The treatment of autoimmune hypoglycemia (e.g., with a glucocorticoid or immunosuppressive drugs) is problematic, but the disorders are sometimes self-limited. Failing these treatments, frequent feedings and avoidance of fasting may be required. Administration of uncooked cornstarch at bedtime or even an overnight intragastric infusion of glucose may be necessary in some patients.

FURTHER READINGS

The Advance Collaborative Group: Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 358:2560, 2008

Cryer PE: *Hypoglycemia in Diabetes. Pathophysiology, Prevalence and Prevention*. Alexandria, VA, American Diabetes Association, 2009

——— et al: Evaluation and management of adult hypoglycemic disorders: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 94:709, 2009

———: Hypoglycemia, in Williams Textbook of Endocrinology, 12th ed, S Melmed et al (eds). Philadelphia, Saunders, in press

Davis SN: Diabetes: Hypoglycemia—A new approach to an old problem. Nat Rev Endocrinol 5:243, 2009[PMID: 19444253] [Full Text]

——— et al: Effects of intensive therapy and antecedent hypoglycemia on counterregulatory responses to hypoglycemia in type 2 diabetes. Diabetes 58:701, 2009

Murad MH et al: Drug-induced hypoglycemia: A systematic review. J Clin Endocrinol Metab 94:741, 2009[PMID: 19088166] [Full Text]

The Nice-Sugar Study Investigators: Intensive versus conventional glucose control in critically ill patients. N Engl J Med 360:1283, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE TESTES AND MALE REPRODUCTIVE SYSTEM: INTRODUCTION

The male reproductive system regulates sex differentiation, virilization, and the hormonal changes that accompany puberty, ultimately leading to spermatogenesis and fertility. Under the control of the pituitary hormones—luteinizing hormone (LH) and follicle-stimulating hormone (FSH)—the Leydig cells of the testes produce testosterone and germ cells are nurtured by Sertoli cells to divide, differentiate, and mature into sperm. During embryonic development, testosterone and dihydrotestosterone (DHT) induce the wolffian duct and virilization of the external genitalia. During puberty, testosterone promotes somatic growth and the development of secondary sex characteristics. In adults, testosterone is necessary for spermatogenesis, stimulation of libido, normal sexual function, and maintenance of muscle and bone mass. This chapter focuses on the physiology of the testes and disorders associated with decreased androgen production, which may be caused by gonadotropin deficiency or primary testis dysfunction. A variety of testosterone formulations now allow more physiologic androgen replacement. Infertility occurs in ~5% of men and is increasingly amenable to treatment by hormone replacement or by using sperm transfer techniques. **For further discussion of sexual dysfunction, disorders of the prostate, and testicular cancer, see Chaps. 48, 95, and 96, respectively**.

DEVELOPMENT AND STRUCTURE OF THE TESTIS

The fetal testis develops from the undifferentiated gonad after expression of a genetic cascade that is initiated by the SRY (sex-related gene on the Y chromosome) (Chap. 349). SRY induces differentiation of Sertoli cells, which surround germ cells and, together with peritubular myoid cells, form testis cords that later develop into seminiferous tubules. Fetal Leydig cells and endothelial cells migrate into the gonad from the adjacent mesonephros but also may arise from interstitial cells that reside between testis cords. Leydig cells produce testosterone, which supports the growth and differentiation of wolffian duct structures that develop into the epididymis, vas deferens, and seminal vesicles. Testosterone is also converted to DHT (see below), which induces formation of the prostate and the external male genitalia, including the penis, urethra, and scrotum. Testicular descent through the inguinal canal is controlled in part by Leydig cell production of insulin-like factor 3 (INSL3), which acts via a receptor termed LGR8 (leucine-rich repeat-containing G protein–coupled receptor 8, also known as GREAT: *G* protein–coupled *re*ceptor *a*ffecting *t*estis descent). Sertoli cells produce müllerian inhibiting substance (MIS), which causes regression of the müllerian structures, including the fallopian tube, uterus, and upper segment of the vagina.

NORMAL MALE PUBERTAL DEVELOPMENT

Although *puberty* commonly refers to the maturation of the reproductive axis and the development of secondary sex characteristics, it involves a coordinated response of multiple hormonal systems that include the adrenal gland and the growth hormone (GH) axis **(Fig. 346-1)**. The development of secondary sex characteristics is initiated by *adrenarche*, which usually occurs between 6 and 8 years of age when the adrenal gland begins to produce greater amounts of androgens from the zona reticularis, the principal site of dehydroepiandrosterone (DHEA) production. The sex maturation process is greatly accelerated by the activation of the hypothalamic-pituitary axis and the production of gonadotropin-releasing hormone (GnRH). The GnRH pulse generator in the hypothalamus is active during fetal life and early infancy but is restrained until the early stages of puberty by a neuroendocrine brake imposed by the inhibitory actions of glutamate, γ-amino butyric acid (GABA), and neuropeptide Y. Although the pathways that initiate reactivation of the GnRH pulse generator at the onset of puberty have been elusive, mounting evidence supports involvement of GPR54, a G protein–coupled receptor that binds an endogenous ligand, kisspeptin. Individuals with mutations of GPR54 fail to enter puberty, and experiments in primates demonstrate that infusion of kisspeptin is sufficient to induce premature puberty. Kisspeptin signaling plays an important role in mediating the feedback action of sex steroids on gonadotropin secretion and in regulating the tempo of sexual maturation at puberty. Leptin, a hormone produced by adipose cells, plays a permissive role in the resurgence of GnRH secretion at the onset of puberty, as leptin-deficient individuals also fail to enter puberty (Chap. 77).

Figure 346-1

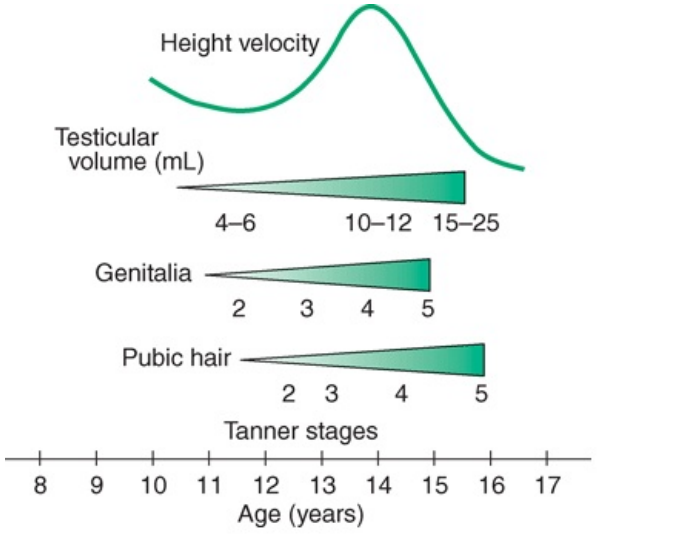

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pubertal events in males. Sexual maturity ratings for genitalia and pubic hair and divided into five stages. *(From WA Marshall, JM Tanner: Arch Dis Child 45:13, 1970 .)*

The early stages of puberty are characterized by nocturnal surges of LH and FSH. Growth of the testes is usually the first sign of puberty, reflecting an increase in seminiferous tubule volume. Increasing levels of testosterone deepen the voice and increase muscle growth. Conversion of testosterone to DHT leads to growth of the external genitalia and pubic hair. DHT also stimulates prostate and facial hair growth and initiates recession of the temporal hairline. The growth spurt occurs at a testicular volume of about 10–12 mL. GH increases early in puberty and is stimulated in part by the rise in gonadal steroids. GH increases the level of insulin-like growth factor I (IGF-I), which enhances linear bone growth. The prolonged pubertal exposure to gonadal steroids (mainly estradiol) ultimately causes epiphyseal closure and limits further bone growth.

REGULATION OF TESTICULAR FUNCTION

Regulation of the Hypothalamic-Pituitary-Testis Axis in Adult Men

Hypothalamic GnRH regulates the production of the pituitary gonadotropins LH and FSH **(Fig. 346-2)**. GnRH is released in discrete pulses approximately every 2 hours, resulting in corresponding pulses of LH and FSH. These dynamic hormone pulses account in part for the wide variations in LH and testosterone even within the same individual. LH acts primarily on the Leydig cell to stimulate testosterone synthesis. The regulatory control of androgen synthesis is mediated by testosterone and estrogen feedback on both the hypothalamus and the pituitary. FSH acts on the Sertoli cell to regulate spermatogenesis and the production of Sertoli products such as inhibin B, which acts to suppress pituitary FSH selectively. Despite these somewhat distinct Leydig and Sertoli cell–regulated pathways, testis function is integrated at several levels: GnRH regulates both gonadotropins; spermatogenesis requires high levels of testosterone; numerous paracrine interactions between Leydig and Sertoli cells are necessary for normal testis function.

Figure 346-2

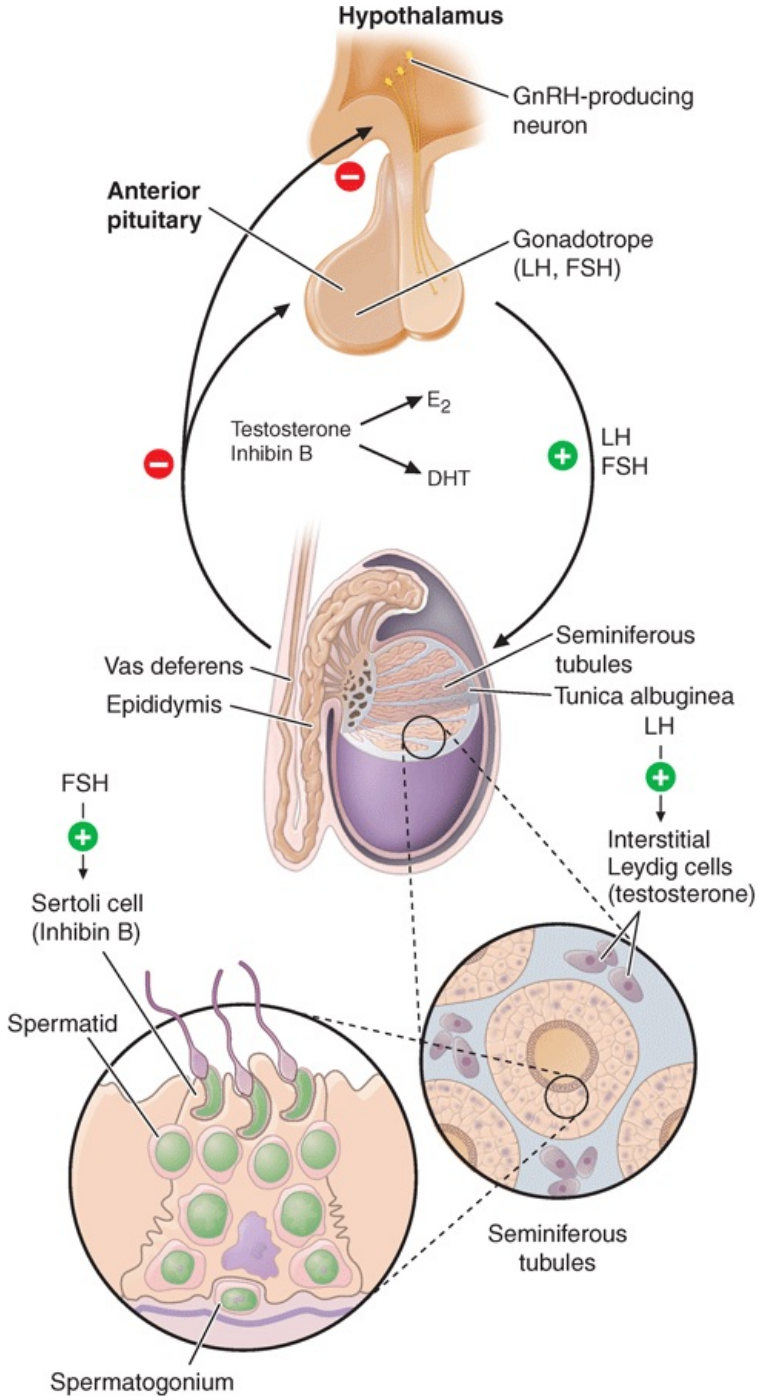

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Human pituitary gonadotropin axis, structure of testis, seminiferous tubule. E_2, 17β–estradiol; DHT, dihydrotestosterone.

The Leydig Cell: Androgen Synthesis

LH binds to its seven transmembrane, G protein–coupled receptor to activate the cyclic AMP pathway. Stimulation of the LH receptor induces *st*eroid *a*cute *r*egulatory (StAR) protein, along with several steroidogenic enzymes involved in androgen synthesis. LH receptor mutations cause Leydig cell hypoplasia or agenesis, underscoring the importance of this pathway for Leydig cell development and function. The rate-limiting process in testosterone synthesis is the delivery of cholesterol by the StAR protein to the inner mitochondrial membrane. Peripheral benzodiazepine receptor, a mitochondrial cholesterol-binding protein, is also an acute regulator of Leydig cell steroidogenesis. The five major enzymatic steps involved in testosterone synthesis are summarized in **Fig. 346-3**. After cholesterol transport into the mitochondrion, the formation of pregnenolone by CYP11A1 (side chain cleavage enzyme) is a limiting enzymatic step. The 17α-hydroxylase and the 17,20-lyase reactions are catalyzed by a single enzyme, CYP17; posttranslational modification (phosphorylation) of this enzyme and the presence of specific enzyme cofactors confer 17,20-lyase activity selectively in the testis and zona

reticularis of the adrenal gland. Testosterone can be converted to the more potent DHT by 5α-reductase, or it can be aromatized to estradiol by CYP19 (aromatase). Two isoforms of steroid 5α-reductase, SRD5A1 and SRD5A2, have been described; all known individuals with clinical features of 5α-reductase deficiency have mutations in SRD5A2, the predominant form in the prostate and the skin.

Figure 346-3

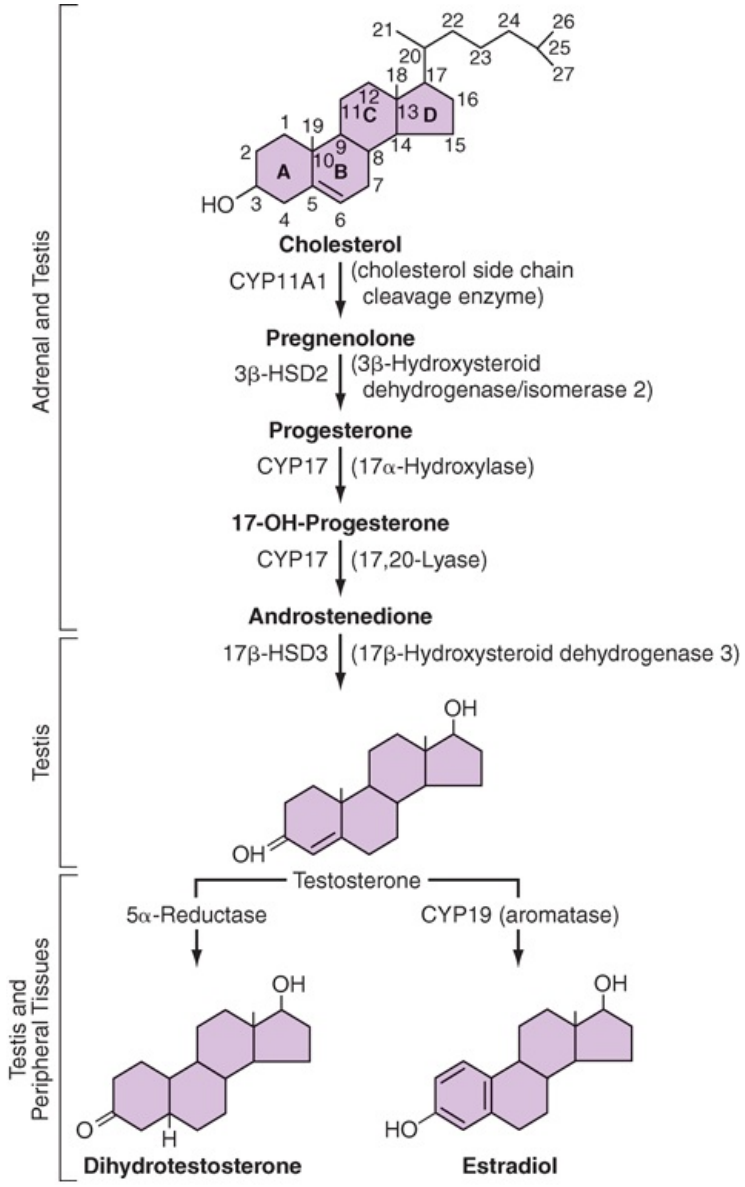

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The biochemical pathway in the conversion of 27-carbon sterol cholesterol to androgens and estrogens.

TESTOSTERONE TRANSPORT AND METABOLISM

In males, 95% of circulating testosterone is derived from testicular production (3–10 mg/d). Direct secretion of testosterone by the adrenal and peripheral conversion of androstenedione to testosterone collectively account for another 0.5 mg/d of testosterone. Only a small amount of DHT (70 μg/d) is secreted directly by the testis; most circulating DHT is derived from peripheral conversion of testosterone. Most of the daily production of estradiol (approximately 45 μg/d) in men is derived from aromatase-mediated peripheral conversion of testosterone and androstenedione.

Circulating testosterone is bound to two plasma proteins: sex hormone–binding globulin (SHBG) and albumin **(Fig. 346-4)**. SHBG binds testosterone with much greater affinity than albumin. Only 0.5–3% of testosterone is unbound. According to the "free hormone" hypothesis, only the unbound fraction is biologically active; however, albumin-bound hormone dissociates readily in the capillaries and may be bioavailable. SHBG concentrations are decreased by androgens, obesity, diabetes mellitus, insulin, and nephrotic syndrome. Conversely, estrogen administration, hyperthyroidism, many chronic

inflammatory illnesses, and aging are associated with high SHBG concentrations.

Figure 346-4

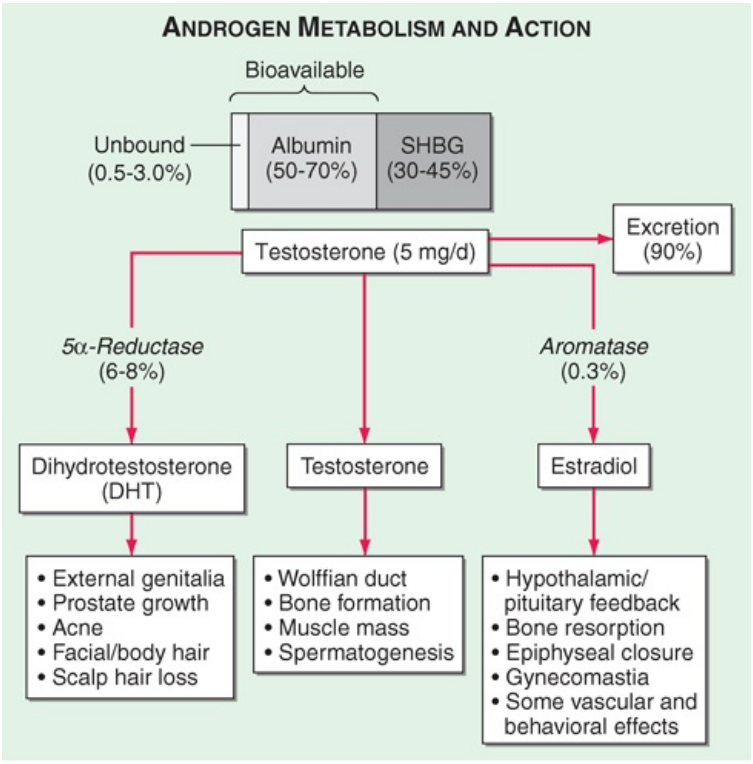

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Androgen metabolism and actions. SHBG, sex hormone–binding globulin.

Testosterone is metabolized predominantly in the liver, although some degradation occurs in peripheral tissues, particularly the prostate and the skin. In the liver, testosterone is converted by a series of enzymatic steps that involve 5α- and 5β-reductases, 3α- and 3β-hydroxysteroid dehydrogenases, and 17β-hydroxysteroid dehydrogenase into androsterone, etiocholanolone, DHT, and 3-α-androstanediol. These compounds undergo glucuronidation or sulfation before being excreted by the kidneys.

MECHANISM OF ANDROGEN ACTION

The androgen receptor (AR) is structurally related to the nuclear receptors for estrogen, glucocorticoids, and progesterone (Chap. 338). The AR is encoded by a gene on the long arm of the X chromosome and has a molecular mass of about 110 kDa. A polymorphic region in the amino terminus of the receptor, which contains a variable number of glutamine repeats, modifies the transcriptional activity of the receptor. The AR protein is distributed in both the cytoplasm and the nucleus. Ligand binding to the AR induces conformational changes that allow the recruitment and assembly of tissue-specific cofactors and causes it to translocate into the nucleus, where it binds to DNA or other transcription factors already bound to DNA. Thus, the AR is a ligand-regulated transcription factor. Some androgen effects may be mediated by nongenomic AR signal transduction pathways. Testosterone binds to AR with one-half the affinity of DHT. The DHT-AR complex also has greater thermostability and a slower dissociation rate than the testosterone-AR complex. However, the molecular basis for selective testosterone versus DHT actions has not been explained completely.

The Seminiferous Tubules: Spermatogenesis

The seminiferous tubules are convoluted, closed loops with both ends emptying into the rete testis, a network of progressively larger efferent ducts that ultimately form the epididymis (Fig. 346-2). The seminiferous tubules total about 600 m in length and account for about two-thirds of testis volume. The walls of the tubules are formed by polarized Sertoli cells that are apposed to peritubular myoid cells. Tight junctions between Sertoli cells create a blood-testis barrier. Germ cells constitute the majority of the seminiferous epithelium (~60%) and are intimately embedded within the cytoplasmic extensions of the Sertoli cells, which function as "nurse cells." Germ cells progress through characteristic stages of mitotic and meiotic divisions. A pool of type A spermatogonia serve as stem cells capable of self-renewal. Primary spermatocytes are derived from type B spermatogonia and undergo meiosis before progressing to spermatids that undergo spermiogenesis (a differentiation process involving chromatin condensation, acquisition of an acrosome, elongation of cytoplasm, and formation of a tail) and are

released from Sertoli cells as mature spermatozoa. The complete differentiation process into mature sperm requires 74 days. Peristaltic-type action by peritubular myoid cells transports sperm into the efferent ducts. The spermatozoa spend an additional 21 days in the epididymis, where they undergo further maturation and capacitation. The normal adult testes produce >100 million sperm per day.

Naturally occurring mutations in the *FSHb* gene and the FSH receptor confirm an important, but not essential, role for this pathway in spermatogenesis. Females with these mutations are hypogonadal and infertile because ovarian follicles do not mature; males exhibit variable degrees of reduced spermatogenesis, presumably because of impaired Sertoli cell function. Because Sertoli cells produce inhibin B, an inhibitor of FSH, seminiferous tubule damage (e.g., by radiation) causes a selective increase of FSH. Testosterone reaches very high concentrations locally in the testis and is essential for spermatogenesis. The cooperative actions of FSH and testosterone are important in the progression of meiosis and spermiation. FSH and testosterone regulate germ cell survival via intrinsic and extrinsic apoptotic mechanisms. FSH also may play an important role in supporting spermatogonia. Gonadotropin-regulated testicular RNA helicase (GRTH/DDX25), a testis-specific gonadotropin/androgen-regulated RNA helicase, is present in germ cells and Leydig cells and may participate in paracrine regulation of germ cell development.

A number of knockout mouse models exhibit impaired germ cell development or spermatogenesis, presaging possible mutations associated with male infertility. The human Y chromosome contains a small pseudoautosomal region that can recombine with homologous regions of the X chromosome. Most of the Y chromosome does not recombine with the X chromosome and is referred to as the male-specific region of the Y (MSY). The MSY contains 156 transcription units that encode for 26 proteins, including nine families of Y-specific multicopy genes; many of these Y-specific genes are also testis-specific and necessary for spermatogenesis. Microdeletions of several Y chromosome azoospermia factor (*AZF*) genes (e.g., RNA-binding motif, *RBM*; deleted in azoospermia, *DAZ*) are associated with oligospermia or azoospermia.

Treatment: Male Factor Infertility

Treatment options for male factor infertility have expanded greatly in recent years. Secondary hypogonadism is highly amenable to treatment with gonadotropins (see below). Assisted reproductive technologies such as in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI) have provided new opportunities for patients with male factor infertility primary testicular failure and disorders of sperm transport. Choice of initial treatment options depends on sperm concentration and motility. Expectant management should be attempted initially in men with mild male factor infertility (sperm count 15–20 x 10^6/mL and normal motility). Moderate male factor infertility (10–15 x 10^6/mL and 20–40% motility) should begin with intrauterine insemination alone or in combination with treatment of the female partner with clomiphene or gonadotropins, but it may require IVF with or without ICSI. For men with a severe defect (sperm count <10 x 10^6/mL, 10% motility), IVF with ICSI or donor sperm should be used.

CLINICAL AND LABORATORY EVALUATION OF MALE REPRODUCTIVE FUNCTION

History and Physical Examination

The history should focus on developmental stages such as puberty and growth spurts, as well as androgen-dependent events such as early-morning erections, frequency and intensity of sexual thoughts, and frequency of masturbation or intercourse. Although libido and the overall frequency of sexual acts are decreased in androgen-deficient men, young hypogonadal men may achieve erections in response to visual erotic stimuli. Men with acquired androgen deficiency often report decreased energy and increased irritability.

The physical examination should focus on secondary sex characteristics such as hair growth, gynecomastia, testicular volume, prostate, and height and body proportions. *Eunuchoid proportions* are defined as an arm span >2 cm greater than height and suggest that androgen deficiency occurred before epiphyseal fusion. Hair growth on the face, axilla, chest, and pubic regions is androgen-dependent; however, changes may not be noticeable unless androgen deficiency is severe and prolonged. Ethnicity also influences the intensity of hair growth (Chap. 49). Testicular volume is best assessed by using a Prader orchidometer. Testes range from 3.5 to 5.5 cm in length, which corresponds to a volume of 12–25 mL. Advanced age does not influence testicular size, although the consistency becomes less firm. Asian men generally have smaller testes than Western Europeans, independent of differences in body size. Because of its possible role in infertility, the presence of varicocele should be sought by palpation while the patient is standing; it is more common on the left side. Patients with Klinefelter's syndrome have markedly reduced testicular volumes (1–2 mL). In congenital hypogonadotropic hypogonadism, testicular volumes provide a good index for the degree of gonadotropin deficiency and the likelihood of response to therapy.

Gonadotropin and Inhibin Measurements

LH and FSH are measured by using two-site immunoradiometric, immunofluorometric, or

chemiluminescent assays, which have very low cross-reactivity with other pituitary glycoprotein hormones and human chorionic gonadotropin (hCG) and have sufficient sensitivity to measure the low levels present in patients with hypogonadotropic hypogonadism. In men with a low testosterone level, an LH level can distinguish primary (high LH) from secondary (low or inappropriately normal LH) hypogonadism. An elevated LH level indicates a primary defect at the testicular level, whereas a low or inappropriately normal LH level suggests a defect at the hypothalamic-pituitary level. LH pulses occur about every 1–3 hours in normal men. Thus, gonadotropin levels fluctuate, and samples should be pooled or repeated when results are equivocal. FSH is less pulsatile than LH because it has a longer half-life. Selective increase in FSH suggests damage to the seminiferous tubules. Inhibin B, a Sertoli cell product that suppresses FSH, is reduced with seminiferous tubule damage. Inhibin B is a dimer with α-β_B subunits and is measured by two-site immunoassays.

GNRH STIMULATION TESTING

The GnRH test is performed by measuring LH and FSH concentrations at baseline and at 30 and 60 minutes after intravenous administration of 100 μg of GnRH. A minimally acceptable response is a twofold LH increase and a 50% FSH increase. In the prepubertal period or with severe GnRH deficiency, the gonadotrope may not respond to a single bolus of GnRH because it has not been primed by endogenous hypothalamic GnRH. With the advent of highly sensitive gonadotropin assays, this test is rarely used in practice.

Testosterone Assays

TOTAL TESTOSTERONE

Total testosterone includes both unbound and protein-bound testosterone and is measured by radioimmunoassays, immunometric assays, or liquid chromatography tandem mass spectrometry (LC-MS/MS). LC-MS/MS involves extraction of serum by organic solvents, separation of testosterone from other steroids by high-performance liquid chromatography and mass spectrometry, and quantitation of unique testosterone fragments by mass spectrometry. LC-MS/MS provides accurate and sensitive measurements of testosterone levels even in the low range and is emerging as the method of choice for testosterone measurement. A single random sample provides a good approximation of the average testosterone concentration with the realization that testosterone levels fluctuate in response to pulsatile LH. Testosterone is generally lower in the late afternoon and is reduced by acute illness. The testosterone concentration in healthy young men ranges from 300 to 1000 ng/dL in most laboratories, although these reference ranges are not derived from population-based random samples. Alterations in SHBG levels due to aging, obesity, diabetes mellitus, hyperthyroidism, some types of medications, or chronic illness or on a congenital basis can affect total testosterone levels. Heritable factors contribute substantially to the population-level variation in testosterone levels. Genomewide association studies have revealed polymorphisms in the SHBG gene as important contributors to variation in testosterone levels.

MEASUREMENT OF UNBOUND TESTOSTERONE LEVELS

Most circulating testosterone is bound to SHBG and to albumin; only 0.5–3% of circulating testosterone is unbound, or "free." The unbound testosterone concentration can be measured by equilibrium dialysis or calculated from total testosterone, SHBG, and albumin concentrations by using published mass-action equations. Tracer analogue methods are relatively inexpensive and convenient but are inaccurate. *Bioavailable testosterone* refers to unbound testosterone plus testosterone that is loosely bound to albumin; it can be measured by the ammonium sulfate precipitation method.

HCG STIMULATION TEST

The hCG stimulation test is performed by administering a single injection of 1500–4000 IU of hCG intramuscularly and measuring testosterone levels at baseline and 24, 48, 72, and 120 hours after hCG injection. An alternative regimen involves three injections of 1500 units of hCG on successive days and measurement of testosterone levels 24 hours after the last dose. An acceptable response to hCG is a doubling of the testosterone concentration in adult men. In prepubertal boys, an increase in testosterone to >150 ng/dL indicates the presence of testicular tissue. No response may indicate an absence of testicular tissue or marked impairment of Leydig cell function. Measurement of MIS, a Sertoli cell product, is also used to detect the presence of testes in prepubertal boys with cryptorchidism.

Semen Analysis

Semen analysis is the most important step in the evaluation of male infertility. Samples are collected by masturbation after a period of abstinence of 2–3 days. Semen volumes and sperm concentrations vary considerably among fertile men, and several samples may be needed before it is possible to conclude that the results are abnormal. Analysis should be performed within an hour of collection. The normal ejaculate volume is 2–6 mL and contains sperm counts >20 million/mL, with a motility of >50% and >15% normal morphology. Some men with low sperm counts are nevertheless fertile. A variety of tests for sperm function can be performed in specialized laboratories, but they add relatively little to the treatment options.

Testicular Biopsy

Testicular biopsy is useful in some patients with oligospermia or azoospermia as an aid in diagnosis and an indication for the feasibility of treatment. Using local anesthesia, fine-needle aspiration biopsy is performed to aspirate tissue for histology. Alternatively, open biopsies can be performed under local or general anesthesia when more tissue is required. A normal biopsy in an azoospermic man with a normal FSH level suggests obstruction of the vas deferens, which may be correctable surgically. Biopsies are also used to harvest sperm for ICSI and to classify disorders such as hypospermatogenesis (all stages present but in reduced numbers), germ cell arrest (usually at primary spermatocyte stage), and Sertoli cell–only syndrome (absent germ cells) or hyalinization (sclerosis with absent cellular elements).

DISORDERS OF SEXUAL DIFFERENTIATION

See Chap. 349.

DISORDERS OF PUBERTY

The onset and tempo of puberty vary greatly in the general population and are affected by genetic and environmental factors. Although some of the variance in the timing of puberty is explained by heritable factors, the genes involved, and their relative contributions to the timing of puberty are not known.

Precocious Puberty

Puberty in boys before age 9 is considered precocious. *Isosexual precocity* refers to premature sexual development consistent with phenotypic sex and includes features such as the development of facial hair and phallic growth. Isosexual precocity is divided into gonadotropin-dependent and gonadotropin-independent causes of androgen excess **(Table 346-1)**. *Heterosexual precocity* refers to the premature development of estrogenic features in boys such as breast development.

Table 346-1 Causes of Precocious or Delayed Puberty in Boys

I. Precocious puberty

- A. Gonadotropin-dependent
  - 1. Idiopathic
  - 2. Hypothalamic hamartoma or other lesions
  - 3. CNS tumor or inflammatory state
- B. Gonadotropin-independent
  - 1. Congenital adrenal hyperplasia
  - 2. hCG-secreting tumor
  - 3. McCune-Albright syndrome
  - 4. Activating LH receptor mutation
  - 5. Exogenous androgens

II. Delayed puberty

- A. Constitutional delay of growth and puberty
- B. Systemic disorders
  - 1. Chronic disease
  - 2. Malnutrition
  - 3. Anorexia nervosa
- C. CNS tumors and their treatment (radiotherapy and surgery)
- D. Hypothalamic-pituitary causes of pubertal failure (low gonadotropins)
  - 1. Congenital disorders (Table 346-2)
    - a. Hypothalamic syndromes (e.g., Prader-Willi)
    - b. Idiopathic hypogonadotropic hypogonadism
    - c. Congenital
    - d. PROP1 mutations and other mutations affecting pituitary development/function
  - 2. Acquired disorders
    - a. Pituitary tumors
    - b. Hyperprolactinemia

E. Gonadal causes of pubertal failure (elevated gonadotropins)
1. Klinefelter's syndrome
2. Bilateral undescended testes
3. Orchitis
4. Chemotherapy or radiotherapy
5. Anorchia

F. Androgen insensitivity

Abbreviations: CNS, central nervous system; GnRH, gonadotropin-releasing hormone; hCG, human chronic gonadotropin; LH, luteinizing hormone.

Table 346-2 Causes of Congenital Hypogonadotropic Hypogonadism

Gene	Locus	Inheritance	Associated Features
KAL1	Xp22	X-linked	Anosmia, renal agenesis, synkinesia, cleft lip/palate, oculomotor/visuospatial defects, gut malrotations
NELF	9q34.3	AR	Anosmia, hypogonadotropic hypogonadism
FGFR1	8p11-p12	AD	Anosmia, cleft lip/palate, synkinesia, syndactyly
PROK2	20p13	AR	Anosmia, hypogonadotropic hypogonadism
PROKR2	20p12.3	AR	Anosmia
LEP	7q31	AR	Obesity
LEPR	1p31	AR	Obesity
PC1	5q15-21	AR	Obesity, diabetes mellitus, ACTH deficiency
HESX1	3p21	AR	Septo-optic dysplasia, CPHD
		AD	Isolated GH insufficiency
LHX3	9q34	AR	CPHD (ACTH spared), cervical spine rigidity
PROP1	5q35	AR	CPHD (ACTH usually spared)
GPR54	19p13	AR	None
GnRHR	4q21	AR	None
FSHβ	11p13	AR	↑ LH
LHβ	19q13	AR	↑ FSH
SF1 (NR5A1)	9p33	AD/AR	Primary adrenal failure, XY sex reversal
DAX1 (NR0B1)	Xp21	X-linked	Primary adrenal failure, impaired spermatogenesis
TAC3R	4q25	AR	None
TAC3	12q13-q21	AR	None

Abbreviations: ACTH, adrenocorticotropic hormone; AD, autosomal dominant; AR, autosomal recessive; CPHD, combined pituitary hormone deficiency; KAL1, Interval-1 gene; NELF, nasal embryonic LHRH factor; FGFR1, fibroblast growth factor receptor 1; PROKR2, prokineticin receptor 2; LEP, leptin; LEPR, leptin receptor; PC1, prohormone convertase 1; HESX1, homeobox gene expressed in embryonic stem cells 1; LHX3, LIM homeobox gene 3; PROP1, Prophet of Pit 1; GPR54, G protein–coupled receptor 54; GnRHR, gonadotropin-releasing hormone receptor; FSHβ, follicle-stimulating hormone β subunit; LHβ, luteinizing hormone β subunit; SF1, steroidogenic factor 1; DAX1, *d*osage-sensitive sex-reversal, *a*drenal hypoplasia congenita, *X*-chromosome.

GONADOTROPIN-DEPENDENT PRECOCIOUS PUBERTY

This disorder, called *central precocious puberty* (CPP), is less common in boys than in girls. It is caused by premature activation of the GnRH pulse generator, sometimes because of central nervous system (CNS) lesions such as hypothalamic hamartomas, but it is often idiopathic. CPP is characterized by gonadotropin levels that are inappropriately elevated for age. Because pituitary priming has occurred, GnRH elicits LH and FSH responses typical of those seen in puberty or in adults. MRI should be performed to exclude a mass, structural defect, infection, or inflammatory process.

GONADOTROPIN-INDEPENDENT PRECOCIOUS PUBERTY

In these disorders, androgens from the testis or the adrenal are increased but gonadotropins are low. This group of disorders includes hCG-secreting tumors; congenital adrenal hyperplasia; sex steroid–producing tumors of the testis, adrenal, and ovary; accidental or deliberate exogenous sex steroid administration; hypothyroidism; and activating mutations of the LH receptor or $G_{S\alpha}$ subunit.

Familial Male-Limited Precocious Puberty

Also called *testotoxicosis*, familial male-limited precocious puberty is an autosomal dominant disorder caused by activating mutations in the LH receptor, leading to constitutive stimulation of the cyclic AMP pathway and testosterone production. Clinical features include premature androgenization in boys, growth acceleration in early childhood, and advanced bone age followed by premature epiphyseal fusion. Testosterone is elevated, and LH is suppressed. Treatment options include inhibitors of testosterone synthesis (e.g., ketoconazole), androgen receptor antagonists (e.g., flutamide and bicalutamide), and aromatase inhibitors (e.g., anastrazole).

Mccune-Albright Syndrome

This is a sporadic disorder caused by somatic (postzygotic) activating mutations in the $G_{S\alpha}$ subunit that links G protein–coupled receptors to intracellular signaling pathways (Chap. 355). The mutations impair the guanosine triphosphatase activity of the $G_{S\alpha}$ protein, leading to constitutive activation of adenylyl cyclase. Like activating LH receptor mutations, this stimulates testosterone production and causes gonadotropin-independent precocious puberty. In addition to sexual precocity, affected individuals may have autonomy in the adrenals, pituitary, and thyroid glands. Café au lait spots are characteristic skin lesions that reflect the onset of the somatic mutations in melanocytes during embryonic development. Polyostotic fibrous dysplasia is caused by activation of the parathyroid hormone receptor pathway in bone. Treatment is similar to that in patients with activating LH receptor mutations. Bisphosphonates have been used to treat bone lesions.

Congenital Adrenal Hyperplasia

Boys with congenital adrenal hyperplasia (CAH) who are not well controlled with glucocorticoid suppression of adrenocorticotropic hormone (ACTH) can develop premature virilization because of excessive androgen production by the adrenal gland (Chaps. 342 and 349). LH is low, and the testes are small. Adrenal rests may develop within the testis of poorly controlled patients with CAH because of chronic ACTH stimulation; adrenal rests do not require surgical removal and regress with effective glucocorticoid therapy. Some children with CAH may develop gonadotropin-dependent precocious puberty with early maturation of the hypothalamic- pituitary-gonadal axis, elevated gonadotropins, and testicular growth.

HETEROSEXUAL SEXUAL PRECOCITY

Breast enlargement in prepubertal boys can result from familial aromatase excess, estrogen-producing tumors in the adrenal gland, Sertoli cell tumors in the testis, marijuana smoking, or exogenous estrogens or androgens. Occasionally, germ cell tumors that secrete hCG can be associated with breast enlargement due to excessive stimulation of estrogen production (see "Gynecomastia," below).

APPROACH TO THE PATIENT: PRECOCIOUS PUBERTY

After verification of precocious development, serum LH and FSH levels should be measured to determine whether gonadotropins are increased in relation to chronologic age (gonadotropin-dependent) or whether sex steroid secretion is occurring independent of LH and FSH (gonadotropin-independent). In children with gonadotropin-dependent precocious puberty, CNS lesions should be excluded by history, neurologic examination, and MRI scan of the head. If organic causes are not found, one is left with the diagnosis of idiopathic central precocity. Patients with high testosterone but suppressed LH concentrations have gonadotropin-independent sexual precocity; in these patients, DHEA sulfate (DHEAS) and 17α-hydroxyprogesterone should be measured. High levels of testosterone and 17α-hydroxyprogesterone suggest the possibility of CAH due to 21α-hydroxylase or 11β-hydroxylase deficiency. If testosterone and DHEAS are elevated, adrenal tumors should be excluded by obtaining a CT scan of the adrenal glands. Patients with elevated testosterone but without increased 17α-hydroxyprogesterone or DHEAS should undergo careful evaluation of the testis by palpation and ultrasound to exclude a Leydig cell neoplasm. Activating mutations of the LH receptor should be considered in children with gonadotropin-independent precocious puberty in whom CAH, androgen abuse, and adrenal and testicular neoplasms have been excluded.

PRECOCIOUS PUBERTY

In patients with a known cause (e.g., a CNS lesion or a testicular tumor), therapy should be directed toward the underlying disorder. In patients with idiopathic CPP, long-acting GnRH analogues can be used to suppress gonadotropins and decrease testosterone, halt early pubertal development, delay accelerated bone maturation, and prevent early epiphyseal closure, thus increasing final height and mitigating the psychosocial consequences of early pubertal development. The treatment is most effective for increasing final adult height if it is initiated before age 6. Puberty resumes after

discontinuation of the GnRH analogue. Counseling is an important aspect of the overall treatment strategy.

In children with gonadotropin-independent precocious puberty, inhibitors of steroidogenesis such as ketoconazole, and AR antagonists have been used empirically. Long-term treatment with spironolactone (a weak androgen antagonist) and ketoconazole has been reported to normalize growth rate and bone maturation and to improve predicted height in small, nonrandomized trials in boys with familial male-limited precocious puberty. Aromatase inhibitors such as testolactone and letrozole have been used as an adjunct to antiandrogen and GnRH analogue therapy for children with familial male-limited precocious puberty, congenital adrenal hyperplasia, and McCune-Albright syndrome.

Delayed Puberty

Puberty is delayed in boys if it has not ensued by age 14, an age that is 2–2.5 standard deviations above the mean for healthy children. Delayed puberty is more common in boys than in girls. There are four main categories of delayed puberty: (1) constitutional delay of growth and puberty (~60% of cases), (2) functional hypogonadotropic hypogonadism caused by systemic illness or malnutrition (~20% of cases), (3) hypogonadotropic hypogonadism caused by genetic or acquired defects in the hypothalamic-pituitary region (~10% of cases), and (4) hypergonadotropic hypogonadism secondary to primary gonadal failure (~15% of cases) (Table 346-1). Functional hypogonadotropic hypogonadism is more common in girls than in boys. Permanent causes of hypogonadotropic or hypergonadotropic hypogonadism are identified in <25% of boys with delayed puberty.

APPROACH TO THE PATIENT: DELAYED PUBERTY

Any history of systemic illness, eating disorders, excessive exercise, social and psychological problems, and abnormal patterns of linear growth during childhood should be verified. Boys with pubertal delay may have accompanying emotional and physical immaturity relative to their peers, which can be a source of anxiety. Physical examination should focus on height; arm span; weight; visual fields; and secondary sex characteristics, including hair growth, testicular volume, phallic size, and scrotal reddening and thinning. Testicular size >2.5 cm generally indicates that the child has entered puberty.

The main diagnostic challenge is to distinguish those with constitutional delay, who will progress through puberty at a later age, from those with an underlying pathologic process. Constitutional delay should be suspected when there is a family history and when there are delayed bone age and short stature. Pituitary priming by pulsatile GnRH is required before LH and FSH are synthesized and secreted normally. Thus, blunted responses to exogenous GnRH can be seen in patients with constitutional delay, GnRH deficiency, or pituitary disorders (see "GnRH Stimulation Testing," above). In contrast, low-normal basal gonadotropin levels or a normal response to exogenous GnRH is consistent with an early stage of puberty, which often is heralded by nocturnal GnRH secretion. Thus, constitutional delay is a diagnosis of exclusion that requires ongoing evaluation until the onset of puberty and the growth spurt.

TREATMENT: DELAYED PUBERTY

If therapy is considered appropriate, it can begin with 25–50 mg testosterone enanthate or testosterone cypionate every 2 weeks or by using a 2.5-mg testosterone patch or 25-mg testosterone gel. Because aromatization of testosterone to estrogen is obligatory for mediating androgen effects on epiphyseal fusion, concomitant treatment with aromatase inhibitors may allow attainment of greater final adult height. Testosterone treatment should be interrupted after 6 months to determine whether endogenous LH and FSH secretion has ensued. Other causes of delayed puberty should be considered when there are associated clinical features or when boys do not enter puberty spontaneously after a year of observation or treatment.

Reassurance without hormonal treatment is appropriate for many individuals with presumed constitutional delay of puberty. However, the impact of delayed growth and pubertal progression on a child's social relationships and school performance should be weighed. Also, boys with constitutional delay of puberty are less likely to achieve their full genetic height potential and have reduced total body bone mass as adults, mainly due to narrow limb bones and vertebrae as a result of impaired periosteal expansion during puberty. Administration of androgen therapy to boys with constitutional delay does not affect final height, and when administered with an aromatase inhibitor, it may improve final height.

DISORDERS OF THE MALE REPRODUCTIVE AXIS DURING ADULTHOOD

Hypogonadotropic Hypogonadism

Because LH and FSH are trophic hormones for the testes, impaired secretion of these pituitary gonadotropins results in secondary hypogonadism, which is characterized by low testosterone in the setting of low LH and FSH. Those with the most severe deficiency have complete absence of pubertal development, sexual infantilism, and, in some cases, hypospadias and undescended testes. Patients with partial gonadotropin deficiency have delayed or arrested sex development. The 24-hour LH secretory profiles are heterogeneous in patients with hypogonadotropic hypogonadism, reflecting

variable abnormalities of LH pulse frequency or amplitude. In severe cases, basal LH is low and there are no LH pulses. A smaller subset of patients has low-amplitude LH pulses or markedly reduced pulse frequency. Occasionally, only sleep-entrained LH pulses occur, reminiscent of the pattern seen in the early stages of puberty. Hypogonadotropic hypogonadism can be classified into congenital and acquired disorders. Congenital disorders most commonly involve GnRH deficiency, which leads to gonadotropin deficiency. Acquired disorders are more common than congenital disorders and may result from a variety of sellar mass lesions or infiltrative diseases of the hypothalamus or pituitary.

CONGENITAL DISORDERS ASSOCIATED WITH GONADOTROPIN DEFICIENCY

Most cases of congenital hypogonadotropic hypogonadism are idiopathic despite extensive endocrine testing and imaging studies of the sellar region. Among known causes, familial hypogonadotropic hypogonadism can be transmitted as an X-linked (20%), autosomal recessive (30%), or autosomal dominant (50%) trait. Some individuals with idiopathic hypogonadotropic hypogonadism (IHH) have sporadic mutations in the same genes that cause inherited forms of the disorder **(Table 346-2)**. *Kallmann's syndrome* is an X-linked disorder caused by mutations in the *KAL1* gene, which encodes anosmin, a protein that mediates the migration of neural progenitors of the olfactory bulb and GnRH-producing neurons. These individuals have GnRH deficiency and variable combinations of anosmia or hyposmia, renal defects, and neurologic abnormalities, including mirror movements. Gonadotropin secretion and fertility can be restored by administration of pulsatile GnRH or gonadotropin replacement. Mutations in the *FGFR1* gene cause an autosomal dominant form of hypogonadotropic hypogonadism that clinically resembles Kallmann's syndrome; mutations in its putative ligand, the*FGF8* gene product, have also been associated with IHH. Prokineticin 2 (PROK2) also encodes a protein involved in migration and development of olfactory and GnRH neurons. Recessive mutations in PROK2 or in its receptor, PROKR2, have been associated with both anosmic and normosmic forms of hypogonadotropic hypogonadism. X-linked hypogonadotropic hypogonadism also occurs in *adrenal hypoplasia congenita*, a disorder caused by mutations in the *DAX1* gene, which encodes a nuclear receptor in the adrenal gland and reproductive axis. Adrenal hypoplasia congenita is characterized by absent development of the adult zone of the adrenal cortex, leading to neonatal adrenal insufficiency. Puberty usually does not occur or is arrested, reflecting variable degrees of gonadotropin deficiency. Although sexual differentiation is normal, some patients have testicular dysgenesis and impaired spermatogenesis despite gonadotropin replacement. Less commonly, adrenal hypoplasia congenita, sex reversal, and hypogonadotropic hypogonadism can be caused by mutations of steroidogenic factor 1 (SF1). *GnRH receptor mutations* , the most common identifiable cause of normosmic IHH, account for ~40% of autosomal recessive and 10% of sporadic cases of hypogonadotropic hypogonadism. These patients have decreased LH response to exogenous GnRH. Some receptor mutations alter GnRH binding affinity, allowing apparently normal responses to pharmacologic doses of exogenous GnRH, whereas other mutations may alter signal transduction downstream of hormone binding. Mutations of the *GnRH1* gene also have been reported in patients with hypogonadotropic hypogonadism, although they are rare. G protein–coupled receptor GPR54 and its cognate receptor, kisspeptin, are important regulators of sexual maturation. Recessive mutations in GPR54 cause gonadotropin deficiency without anosmia. Patients retain responsiveness to exogenous GnRH, suggesting an abnormality in the neural pathways that control GnRH release. The genes encoding neurokinin B (TAC3), which is involved in preferential activation of GnRH release in early development, and its receptor (TAC3R) have been implicated in some families with normosmic IHH. Mutations in more than one gene (digenicity) may contribute to clinical heterogeneity in IHH patients. Rarely, recessive mutations in the *LHb* or *FSHb* gene have been described in patients with selective deficiencies of these gonadotropins. In approximately 10% of men with IHH, reversal of gonadotropin deficiency may occur in adult life. Also, a small fraction of men with IHH may present with androgen deficiency and infertility in adult life after having gone through apparently normal pubertal development.

A number of homeodomain transcription factors are involved in the development and differentiation of the specialized hormone-producing cells within the pituitary gland (Table 346-2). Patients with mutations of *PROP1* have combined pituitary hormone deficiency that includes GH, prolactin (PRL), thyroid-stimulating hormone (TSH), LH, and FSH but not ACTH. *LHX3* mutations cause combined pituitary hormone deficiency in association with cervical spine rigidity. *HESX1* mutations cause septo-optic dysplasia and combined pituitary hormone deficiency.

Prader-Willi syndrome is characterized by obesity, hypotonic musculature, mental retardation, hypogonadism, short stature, and small hands and feet. Prader-Willi syndrome is a genomic imprinting disorder caused by deletions of the proximal portion of paternally derived chromosome 15q, uniparental disomy of the maternal alleles, or mutations of the genes/loci involved in imprinting (Chap. 62). *Laurence-Moon syndrome* is an autosomal recessive disorder characterized by obesity, hypogonadism, mental retardation, polydactyly, and retinitis pigmentosa. Recessive mutations of leptin, or its receptor, cause severe obesity and pubertal arrest, apparently because of hypothalamic GnRH deficiency (Chap. 77).

ACQUIRED HYPOGONADOTROPIC DISORDERS

Severe Illness, Stress, Malnutrition, and Exercise

These factors may cause reversible gonadotropin deficiency. Although gonadotropin deficiency and reproductive dysfunction are well documented in these conditions in women, men exhibit similar but less pronounced responses. Unlike women, most male runners and other endurance athletes have normal gonadotropin and sex steroid levels despite low body fat and frequent intensive exercise. Testosterone levels fall at the onset of illness and recover during recuperation. The magnitude of gonadotropin suppression generally correlates with the severity of illness. Although hypogonadotropic hypogonadism is the most common cause of androgen deficiency in patients with acute illness, some have elevated levels of LH and FSH, which suggest primary gonadal dysfunction. The pathophysiology of reproductive dysfunction during acute illness is unknown but probably involves a combination of cytokine and/or glucocorticoid effects. There is a high frequency of low testosterone levels in patients with chronic illnesses such as HIV infection, end-stage renal disease, chronic obstructive lung disease, and many types of cancer and in patients receiving glucocorticoids. About 20% of HIV-infected men with low testosterone levels have elevated LH and FSH levels; these patients presumably have primary testicular dysfunction. The remaining 80% have either normal or low LH and FSH levels; these men have a central hypothalamic-pituitary defect or a dual defect involving both the testis and the hypothalamic-pituitary centers. Muscle wasting is common in chronic diseases associated with hypogonadism, which also leads to debility, poor quality of life, and adverse outcome of disease. There is great interest in exploring strategies that can reverse androgen deficiency or attenuate the sarcopenia associated with chronic illness.

Men using opioids for relief of cancer or noncancerous pain or because of addiction often have suppressed testosterone and LH levels; the degree of suppression is dose-related and particularly severe with long-acting opioids such as methadone. Opioids suppress GnRH secretion and alter the sensitivity to feedback inhibition by gonadal steroids. Men who are heavy users of marijuana have decreased testosterone secretion and sperm production. The mechanism of marijuana-induced hypogonadism is decreased GnRH secretion. Gynecomastia observed in marijuana users can also be caused by plant estrogens in crude preparations. Androgen deprivation therapy in men with prostate cancer has been associated with increased risk of bone fractures, diabetes mellitus, cardiovascular events, fatigue, sexual dysfunction, and poor quality of life.

Obesity

In men with mild to moderate obesity, SHBG levels decrease in proportion to the degree of obesity, resulting in lower total testosterone levels. However, free testosterone levels usually remain within the normal range. The decrease in SHBG levels is caused by increased circulating insulin, which inhibits SHBG production. Estradiol levels are higher in obese men than in healthy, nonobese controls because of aromatization of testosterone to estradiol in adipose tissue. Weight loss is associated with reversal of these abnormalities, including an increase in total and free testosterone levels and a decrease in estradiol levels. A subset of massively obese men may have a defect in the hypothalamic-pituitary axis as suggested by low free testosterone in the absence of elevated gonadotropins. Weight gain in adult men can accelerate the rate of age-related decline in testosterone levels.

Hyperprolactinemia

(See also Chap. 339.) Elevated PRL levels are associated with hypogonadotropic hypogonadism. PRL inhibits hypothalamic GnRH secretion either directly or through modulation of tuberoinfundibular dopaminergic pathways. A PRL-secreting tumor also may destroy the surrounding gonadotropes by invasion or compression of the pituitary stalk. Treatment with dopamine agonists reverses gonadotropin deficiency, although there may be a delay relative to PRL suppression.

Sellar Mass Lesions

Neoplastic and nonneoplastic lesions in the hypothalamus or pituitary can affect gonadotrope function directly or indirectly. In adults, pituitary adenomas constitute the largest category of space-occupying lesions affecting gonadotropin and other pituitary hormone production. Pituitary adenomas that extend into the suprasellar region can impair GnRH secretion and mildly increase PRL secretion (usually <50 μg/L) because of impaired tonic inhibition by dopaminergic pathways. These tumors should be distinguished from prolactinomas, which typically secrete higher PRL levels. The presence of diabetes insipidus suggests the possibility of a craniopharyngioma, infiltrative disorder, or other hypothalamic lesions (Chap. 340).

Hemochromatosis

(See also Chap. 357.) Both the pituitary and the testis can be affected by excessive iron deposition. However, the pituitary defect is the predominant lesion in most patients with hemochromatosis and hypogonadism. The diagnosis of hemochromatosis is suggested by the association of characteristic skin discoloration, hepatic enlargement or dysfunction, diabetes mellitus, arthritis, cardiac conduction defects, and hypogonadism.

Primary Testicular Causes of Hypogonadism

Common causes of primary testicular dysfunction include Klinefelter's syndrome, uncorrected

cryptorchidism, cancer chemotherapy, radiation to the testes, trauma, torsion, infectious orchitis, HIV infection, anorchia syndrome, and myotonic dystrophy. Primary testicular disorders may be associated with impaired spermatogenesis, decreased androgen production, or both. **See Chap. 349 for disorders of testis development, androgen synthesis, and androgen action**.

KLINEFELTER'S SYNDROME

(See also Chap. 349.) Klinefelter's syndrome is the most common chromosomal disorder associated with testicular dysfunction and male infertility. It occurs in about 1 in 1000 live-born males. Azoospermia is the rule in men with Klinefelter's syndrome who have the 47,XXY karyotype; however, men with mosaicism may have germ cells, especially at a younger age. The clinical phenotype of Klinefelter's syndrome can be heterogeneous, possibly because of mosaicism, polymorphisms in androgen receptor gene, variable testosterone levels, or other genetic factors. Testicular histology shows hyalinization of seminiferous tubules and absence of spermatogenesis. Although their function is impaired, the number of Leydig cells appears to increase. Testosterone is decreased and estradiol is increased, leading to clinical features of undervirilization and gynecomastia. Men with Klinefelter's syndrome are at increased risk of systemic lupus erythematosus, breast cancer, non-Hodgkin's lymphoma, and lung cancer and reduced risk of prostate cancer. Periodic mammography for breast cancer surveillance is recommended for men with Klinefelter's syndrome. Fertility has been achieved by intracytoplasmic injection of sperm retrieved surgically from testicular biopsies of men with Klinefelter's syndrome, including some men with nonmosaic forms of Klinefelter's syndrome.

CRYPTORCHIDISM

Cryptorchidism occurs when there is incomplete descent of the testis from the abdominal cavity into the scrotum. About 3% of full-term and 30% of premature male infants have at least one undescended testis at birth, but descent is usually complete by the first few weeks of life. The incidence of cryptorchidism is <1% by 9 months of age. Androgens regulate both the transabdominal and inguinoscrotal descent of the testes through degeneration of the craniosuspensory ligament and a shortening of the gubernacula, respectively. Mutations in INSL3 and its receptor, which regulate the transabdominal portion of testicular descent, have been found in some patients with cryptorchidism.

Cryptorchidism is associated with increased risk of malignancy and infertility. Unilateral cryptorchidism, even when corrected before puberty, is associated with decreased sperm count, possibly reflecting unrecognized damage to the fully descended testis or other genetic factors. Epidemiologic, clinical, and molecular evidence supports the idea that cryptorchidism, hypospadias, impaired spermatogenesis, and testicular cancer may be causally related to common genetic and environment perturbations and are components of the testicular dysgenesis syndrome.

ACQUIRED TESTICULAR DEFECTS

Viral orchitis may be caused by the mumps virus, echovirus, lymphocytic choriomeningitis virus, and group B arboviruses. Orchitis occurs in as many as one-fourth of adult men with mumps; the orchitis is unilateral in about two-thirds and bilateral in the remainder. Orchitis usually develops a few days after the onset of parotitis but may precede it. The testis may return to normal size and function or undergo atrophy. Semen analysis returns to normal for three-fourths of men with unilateral involvement but for only one-third of men with bilateral orchitis. *Trauma*, including testicular torsion, also can cause secondary atrophy of the testes. The exposed position of the testes in the scrotum renders them susceptible to both thermal and physical trauma, particularly in men with hazardous occupations.

The testes are sensitive to *radiation damage*. Doses >200 mGy (20 rad) are associated with increased FSH and LH levels and damage to the spermatogonia. After ~800 mGy (80 rad), oligospermia or azoospermia develops, and higher doses may obliterate the germinal epithelium. Permanent androgen deficiency in adult men is uncommon after therapeutic radiation; however, most boys given direct testicular radiation therapy for acute lymphoblastic leukemia have permanently low testosterone levels. Sperm banking should be considered before patients undergo radiation treatment or chemotherapy.

Drugs interfere with testicular function by several mechanisms, including inhibition of testosterone synthesis (e.g., ketoconazole), blockade of androgen action (e.g., spironolactone), increased estrogen (e.g., marijuana), or direct inhibition of spermatogenesis (e.g., chemotherapy).

Combination chemotherapy for acute leukemia, Hodgkin's disease, and testicular and other cancers may impair Leydig cell function and cause infertility. The degree of gonadal dysfunction depends on the type of chemotherapeutic agent and the dose and duration of therapy. Because of high response rates and the young age of these men, infertility and androgen deficiency have emerged as important long-term complications of cancer chemotherapy. Cyclophosphamide and combination regimens containing procarbazine are particularly toxic to germ cells. Thus, 90% of men with Hodgkin's lymphoma receiving MOPP (mechlorethamine, oncovin, procarbazine, prednisone) therapy develop azoospermia or extreme oligozoospermia; newer regimens that do not include procarbazine, such as ABVD (adriamycin, bleomycin, vinblastine, dacarbazine), are less toxic to germ cells.

Alcohol, when consumed in excess for prolonged periods, decreases testosterone, independent of liver disease or malnutrition. Elevated estradiol and decreased testosterone levels may occur in men

taking digitalis.

The occupational and recreational history should be evaluated carefully in all men with infertility because of the toxic effects of many *chemical agents* on spermatogenesis. Known environmental hazards include microwaves and ultrasound and chemicals such as nematocide dibromochloropropane, cadmium, phthalates, and lead. In some populations, sperm density is said to have declined by as much as 40% in the last 50 years. Environmental estrogens or antiandrogens may be partly responsible.

Testicular failure also occurs as a part of *polyglandular autoimmune insufficiency* (Chap. 351). Sperm antibodies can cause isolated male infertility. In some instances, these antibodies are secondary phenomena resulting from duct obstruction or vasectomy. Granulomatous diseases can affect the testes, and testicular atrophy occurs in 10–20% of men with lepromatous leprosy because of direct tissue invasion by the mycobacteria. The tubules are involved initially, followed by endarteritis and destruction of Leydig cells.

Systemic disease can cause primary testis dysfunction in addition to suppressing gonadotropin production. In cirrhosis, a combined testicular and pituitary abnormality leads to decreased testosterone production independent of the direct toxic effects of ethanol. Impaired hepatic extraction of adrenal androstenedione leads to extraglandular conversion to estrone and estradiol, which partially suppresses LH. Testicular atrophy and gynecomastia are present in approximately one-half of men with cirrhosis. In chronic renal failure, androgen synthesis and sperm production decrease despite elevated gonadotropins. The elevated LH level is due to reduced clearance, but it does not restore normal testosterone production. About one-fourth of men with renal failure have hyperprolactinemia. Improvement in testosterone production with hemodialysis is incomplete, but successful renal transplantation may return testicular function to normal. Testicular atrophy is present in one-third of men with sickle cell anemia. The defect may be at either the testicular or the hypothalamic-pituitary level. Sperm density can decrease temporarily after acute febrile illness in the absence of a change in testosterone production. Infertility in men with celiac disease is associated with a hormonal pattern typical of androgen resistance, namely, elevated testosterone and LH levels.

Neurologic diseases associated with altered testicular function include myotonic dystrophy, spinobulbar muscular atrophy, and paraplegia. In myotonic dystrophy, small testes may be associated with impairment of both spermatogenesis and Leydig cell function. Spinobulbar muscular atrophy is caused by an expansion of the glutamine repeat sequences in the amino-terminal region of the AR; this expansion impairs function of the AR, but it is unclear how the alteration is related to the neurologic manifestations. Men with spinobulbar muscular atrophy often have undervirilization and infertility as a late manifestation. Spinal cord lesions that cause paraplegia can lead to a temporary decrease in testosterone levels and may cause persistent defects in spermatogenesis; some patients retain the capacity for penile erection and ejaculation.

Androgen Insensitivity Syndromes

Mutations in the AR cause resistance to the action of testosterone and DHT. These X-linked mutations are associated with variable degrees of defective male phenotypic development and undervirilization (Chap. 349). Although not technically hormone-insensitivity syndromes, two genetic disorders impair testosterone conversion to active sex steroids. Mutations in the *SRD5A2* gene, which encodes 5α-reductase type 2, prevent the conversion of testosterone to DHT, which is necessary for the normal development of the male external genitalia. Mutations in the *CYP19* gene, which encodes aromatase, prevent testosterone conversion to estradiol. Males with *CYP19* mutations have delayed epiphyseal fusion, tall stature, eunuchoid proportions, and osteoporosis, consistent with evidence from an estrogen receptor–deficient individual that these testosterone actions are mediated indirectly via estrogen.

GYNECOMASTIA

Gynecomastia refers to enlargement of the male breast. It is caused by excess estrogen action and is usually the result of an increased estrogen/androgen ratio. True gynecomastia is associated with glandular breast tissue that is >4 cm in diameter and often tender. Glandular tissue enlargement should be distinguished from excess adipose tissue: glandular tissue is firmer and contains fibrous-like cords. Gynecomastia occurs as a normal physiologic phenomenon in the newborn (due to transplacental transfer of maternal and placental estrogens), during puberty (high ratio of estrogen to androgen in early stages of puberty), and with aging (increased fat tissue and increased aromatase activity), but it also can result from pathologic conditions associated with androgen deficiency or estrogen excess. The prevalence of gynecomastia increases with age and body mass index (BMI), probably because of increased aromatase activity in adipose tissue. Medications that alter androgen metabolism or action may also cause gynecomastia. The relative risk of breast cancer is increased in men with gynecomastia, although the absolute risk is relatively small.

Pathologic Gynecomastia

Any cause of *androgen deficiency* can lead to gynecomastia, reflecting an increased

estrogen/androgen ratio, as estrogen synthesis still occurs by aromatization of residual adrenal and gonadal androgens. Gynecomastia is a characteristic feature of Klinefelter's syndrome (Chap. 349). *Androgen insensitivity* disorders also cause gynecomastia. *Excess estrogen production* may be caused by tumors, including Sertoli cell tumors in isolation or in association with Peutz-Jegher syndrome or Carney complex. Tumors that produce hCG, including some testicular tumors, stimulate Leydig cell estrogen synthesis. *Increased conversion of androgens to estrogens* can be a result of increased availability of substrate (androstenedione) for extraglandular estrogen formation (CAH, hyperthyroidism, and most feminizing adrenal tumors) or diminished catabolism of androstenedione (liver disease) so that androgen precursors are shunted to aromatase in peripheral sites. Obesity is associated with increased aromatization of androgen precursors to estrogens. Extraglandular aromatase activity can also be increased in tumors of the liver or adrenal gland or rarely as an inherited disorder. Several families with *increased peripheral aromatase activity* inherited as an autosomal dominant or an X-linked disorder have been described. In some families with this disorder, an inversion in chromosome 15q21.2-3 causes the *CYP19* gene to be activated by the regulatory elements of contiguous genes, resulting in excessive estrogen production in the fat and other extragonadal tissues. *Drugs* can cause gynecomastia by acting directly as estrogenic substances (e.g., oral contraceptives, phytoestrogens, digitalis), inhibiting androgen synthesis (e.g., ketoconazole), or action (e.g., spironolactone).

Because up to two-thirds of pubertal boys and one-half of hospitalized men have palpable glandular tissue that is benign, detailed investigation or intervention is not indicated in all men presenting with gynecomastia **(Fig. 346-5)**. In addition to the extent of gynecomastia, recent onset, rapid growth, tender tissue, and occurrence in a lean subject should prompt more extensive evaluation. This should include a careful drug history, measurement and examination of the testes, assessment of virilization, evaluation of liver function, and hormonal measurements, including testosterone, estradiol, and androstenedione, LH, and hCG. A karyotype should be obtained in men with very small testes to exclude Klinefelter's syndrome. In spite of extensive evaluation, the etiology is established in fewer than one-half of patients.

Figure 346-5

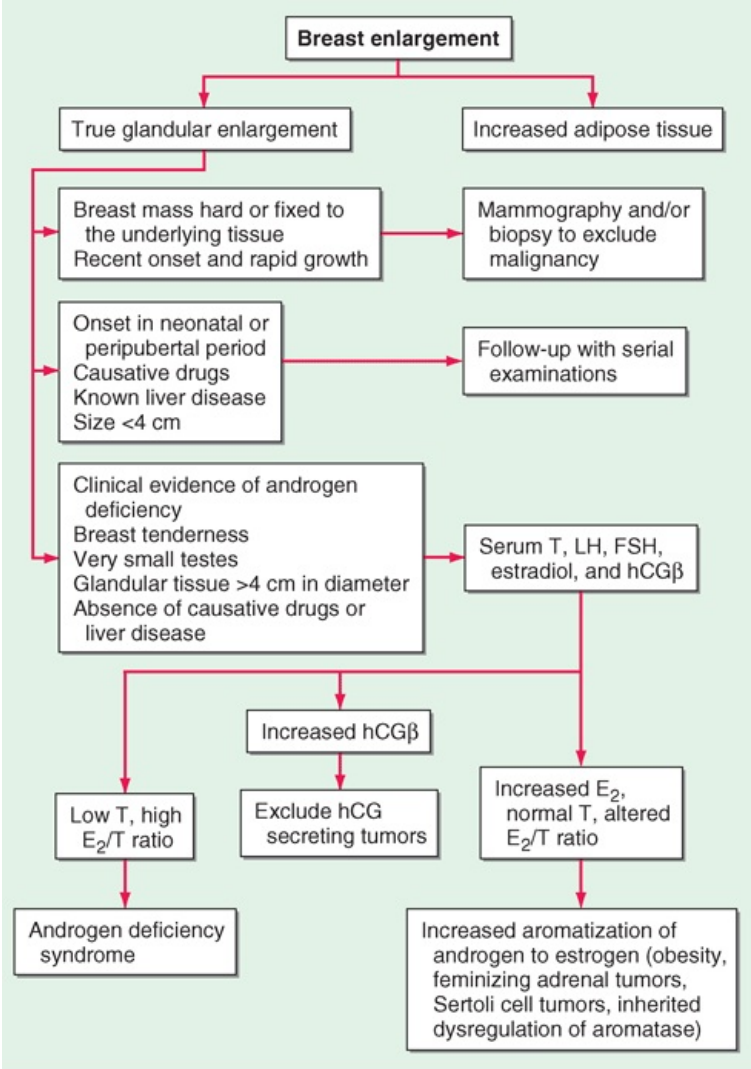

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evaluation of gynecomastia. T, testosterone; LH, luteinizing hormone; FSH, follicle-stimulating hormones; hCGβ, human chorionic gonadotropin β; E_2, 17β-estradiol.

Treatment: Gynecomastia

When the primary cause can be identified and corrected, breast enlargement usually subsides over several months. However, if gynecomastia is of long duration, surgery is the most effective therapy. Indications for surgery include severe psychological and/or cosmetic problems, continued growth or tenderness, and suspected malignancy. In patients who have painful gynecomastia and in whom surgery cannot be performed, treatment with antiestrogens such as tamoxifen (20 mg/d) can reduce pain and breast tissue size in over one-half of the patients. The estrogen receptor antagonists tamoxifen and raloxifen have been reported in small trials to reduce breast size in men with pubertal gynecomastia, although complete regression of breast enlargement is unusual with the use of estrogen receptor antagonists. Aromatase inhibitors can be effective in the early proliferative phase of the disorder. However, in a randomized trial in men with established gynecomastia, anastrozole was no more effective than placebo in reducing breast size.

AGING-RELATED CHANGES IN MALE REPRODUCTIVE FUNCTION

A number of cross-sectional and longitudinal studies (e.g., the Baltimore Longitudinal Study of Aging, the Massachusetts Male Aging Study, and the European Male Aging Study) have established that testosterone concentrations decrease with advancing age. This age-related decline starts in the third decade of life and progresses slowly; the rate of decline in testosterone concentrations is greater in obese men, men with chronic illness, and those taking medications than in healthy older men. Because SHBG concentrations are higher in older men than in younger men, free or bioavailable testosterone concentrations decline with aging to a greater extent than do total testosterone concentrations. The age-related decline in testosterone is due to defects at all levels of the hypothalamic-pituitary-testicular axis: pulsatile GnRH secretion is attenuated, LH response to GnRH is reduced, and testicular response to LH is impaired. However, the gradual rise of LH with aging suggests that testis

dysfunction is the main cause of declining androgen levels. The term *andropause* has been used to denote age-related decline in testosterone concentrations; this term is a misnomer because there is no discrete time when testosterone concentrations decline abruptly.

In epidemiologic surveys, low total and bioavailable testosterone concentrations have been associated with decreased appendicular skeletal muscle mass and strength, decreased self-reported physical function, higher visceral fat mass, insulin resistance, and increased risk of coronary artery disease and mortality. An analysis of signs and symptoms in older men in the European Male Aging Study revealed a syndromic association of sexual and physical symptoms with testosterone levels <320 ng/dL in community-dwelling older men. In systematic reviews of randomized controlled trials, testosterone therapy in healthy older men with low or low-normal testosterone levels was associated with greater increments in lean body mass, grip strength, and self-reported physical function than was the case with placebo. Testosterone therapy also induced greater improvement in vertebral but not femoral bone mineral density. Testosterone therapy in older men with sexual dysfunction and unequivocally low testosterone levels improves libido, but testosterone effects on erectile function and response to selective phosphodiesterase inhibitors have been inconsistent. Testosterone therapy has not been shown to improve depression scores, fracture risk, cognitive function, or clinical outcomes in older men. Furthermore, neither the long-term risks nor the clinical benefits of testosterone therapy in older men have been demonstrated in adequately powered trials. Although there is no evidence that testosterone causes prostate cancer, there is concern that testosterone therapy might induce the growth of subclinical prostate cancers or exacerbate cardiovascular disease. One randomized testosterone trial in older men with mobility limitation and a high burden of chronic conditions such as diabetes, heart disease, hypertension, and hyperlipidemia reported a greater number of cardiovascular events in men randomized to the testosterone arm of the study than in those randomized to the placebo arm. Population screening of all older men for low testosterone levels is not recommended, and testing should be restricted to men who have symptoms or physical features attributable to androgen deficiency. Testosterone therapy is not recommended for all older men with low testosterone levels. In older men with significant symptoms of androgen deficiency who have testosterone levels <200 ng/dL, testosterone therapy may be considered on an individualized basis and should be instituted after careful discussion of the risks and benefits (see "Testosterone Replacement," below).

Testicular morphology, semen production, and fertility are maintained up to a very old age in men. Although concern has been expressed about age-related increases in germ cell mutations and impairment of DNA repair mechanisms, the frequency of chromosomal aneuploidy or structural abnormalities does not increase in the sperm of older men. However, the incidence of autosomal dominant diseases such as achondroplasia, polyposis coli, Marfan's syndrome, and Apert's syndrome increases in the offspring of men who are advanced in age, consistent with transmission of sporadic missense mutations.

Approach to the Patient: Androgen Deficiency

Hypogonadism often is characterized by decreased sex drive, reduced frequency of sexual intercourse or inability to maintain erections, reduced beard growth, loss of muscle mass, decreased testicular size, and gynecomastia. Less than 10% of patients with erectile dysfunction alone have testosterone deficiency. Thus, it is useful to look for a constellation of symptoms and signs suggestive of androgen deficiency. Except when extreme, these clinical features may be difficult to distinguish from changes that occur with normal aging. Moreover, androgen deficiency may develop gradually. Although population studies such as the Massachusetts Male Aging Study and the Baltimore Longitudinal Study of Aging have reported a high prevalence of low testosterone levels in middle-aged and older men, the age-related decline in testosterone should be distinguished from classic hypogonadism due to diseases of the testes, the pituitary, or the hypothalamus.

When symptoms or clinical features suggest possible androgen deficiency, the laboratory evaluation is initiated by the measurement of total testosterone, preferably in the morning, using a reliable assay such as LC-MS/MS (**Fig. 346-6**). A consistently low total testosterone level <300 ng/dL measured by a reliable assay in association with symptoms provides evidence of testosterone deficiency. An early-morning testosterone level >350 ng/dL makes the diagnosis of androgen deficiency unlikely. In men with testosterone levels between 200 and 350 ng/dL, the total testosterone level should be repeated and a free testosterone level should be measured. In older men and in patients with other clinical states that are associated with alterations in SHBG levels, a direct measurement of free testosterone level by equilibrium dialysis can be useful in unmasking testosterone deficiency.

Figure 346-6

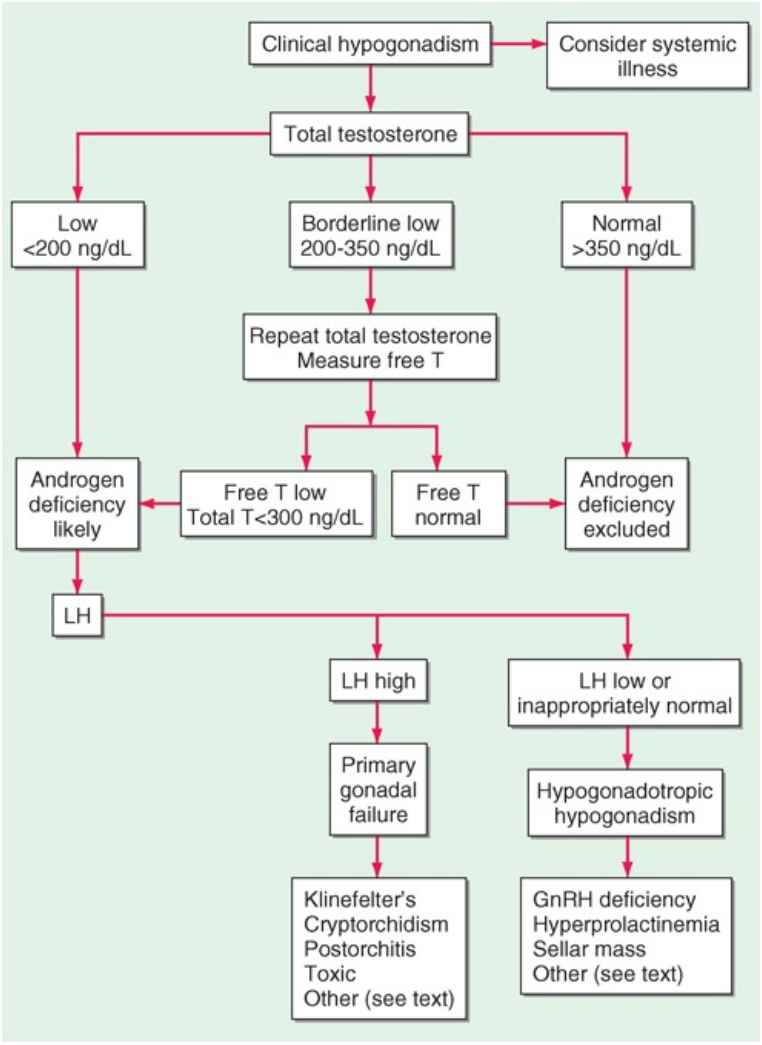

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evaluation of hypogonadism. GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; T, testosterone.

When androgen deficiency has been confirmed by low testosterone concentrations, LH should be measured to classify the patient as having primary (high LH) or secondary (low or inappropriately normal LH) hypogonadism. An elevated LH level indicates that the defect is at the testicular level. Common causes of primary testicular failure include Klinefelter's syndrome, HIV infection, uncorrected cryptorchidism, cancer chemotherapeutic agents, radiation, surgical orchiectomy, and prior infectious orchitis. Unless causes of primary testicular failure are known, a karyotype should be performed in men with low testosterone and elevated LH to exclude Klinefelter's syndrome. Men who have low testosterone but "inappropriately normal" or low LH levels have secondary hypogonadism; their defect resides at the hypothalamic-pituitary level. Common causes of acquired secondary hypogonadism include space-occupying lesions of the sella, hyperprolactinemia, chronic illness, hemochromatosis, excessive exercise, and substance abuse. Measurement of PRL and an MRI scan of the hypothalamic-pituitary region can help exclude the presence of a space-occupying lesion. Patients in whom known causes of hypogonadotropic hypogonadism have been excluded are classified as having IHH. It is not unusual for congenital causes of hypogonadotropic hypogonadism such as Kallmann's syndrome to be diagnosed in young adults.

Treatment: Age-Related Reproductive Dysfunction

Gonadotropins

Gonadotropin therapy is used to establish or restore fertility in patients with gonadotropin deficiency of any cause. Several gonadotropin preparations are available. Human menopausal gonadotropin (hMG; purified from the urine of postmenopausal women) contains 75 IU FSH and 75 IU LH per vial. hCG (purified from the urine of pregnant women) has little FSH activity and resembles LH in its ability to stimulate testosterone production by Leydig cells. Recombinant LH is now available. Because of the expense of hMG, treatment usually is begun with hCG alone, and hMG is added later to promote the FSH-dependent stages of spermatid development. Recombinant human FSH (hFSH) is now available

and is indistinguishable from purified urinary hFSH in its biologic activity and pharmacokinetics in vitro and in vivo, although the mature β subunit of recombinant hFSH has seven fewer amino acids. Recombinant hFSH is available in ampules containing 75 IU (~7.5 μg FSH), which accounts for >99% of protein content. Once spermatogenesis is restored using combined FSH and LH therapy, hCG alone is often sufficient to maintain spermatogenesis.

Although a variety of treatment regimens are used, 1500–2000 IU of hCG or recombinant human LH (rhLH) administered intramuscularly three times weekly is a reasonable starting dose. Testosterone levels should be measured 6–8 weeks later and 48–72 hours after the hCG or rhLH injection; the hCG/rhLH dose should be adjusted to achieve testosterone levels in the mid-normal range. Sperm counts should be monitored on a monthly basis. It may take several months for spermatogenesis to be restored; therefore, it is important to forewarn patients about the potential length and expense of the treatment and to provide conservative estimates of success rates. If testosterone levels are in the mid-normal range but the sperm concentrations are low after 6 months of therapy with hCG alone, FSH should be added. This can be done by using hMG, highly purified urinary hFSH, or recombinant hFSH. The selection of the FSH dose is empirical. A common practice is to start with the addition of 75 IU FSH three times a week in conjunction with the hCG/rhLH injections. If sperm densities are still low after 3 months of combined treatment, the FSH dose should be increased to 150 IU. Occasionally, it may take ≥18–24 months for spermatogenesis to be restored.

The two best predictors of success using gonadotropin therapy in hypogonadotropic men are testicular volume at presentation and time of onset. In general, men with testicular volumes >8 mL have better response rates than those who have testicular volumes <4 mL. Patients who became hypogonadotropic after puberty experience higher success rates than those who have never undergone pubertal changes. Spermatogenesis usually can be reinitiated by hCG alone, with high rates of success for men with postpubertal onset of hypogonadotropism. The presence of a primary testicular abnormality such as cryptorchidism will attenuate testicular response to gonadotropin therapy. Prior androgen therapy does not preclude subsequent response to gonadotropin therapy, although some studies suggest that it may attenuate response to subsequent gonadotropin therapy.

Testosterone Replacement

Androgen therapy is indicated to restore testosterone levels to normal to correct features of androgen deficiency. Testosterone replacement improves libido and overall sexual activity and increases energy, lean muscle mass, and bone density. The benefits of testosterone replacement therapy have been proved only in men who have documented androgen deficiency, as demonstrated by testosterone levels that are well below the lower limit of normal (<250 ng/dL).

Testosterone is available in a variety of formulations with distinct pharmacokinetics **(Table 346-3)**. Testosterone serves as a prohormone and is converted to 17β-estradiol by aromatase and to 5α-dihydrotestosterone by 5α-reductase. Therefore, in evaluating testosterone formulations, it is important to consider whether the formulation being used can achieve physiologic estradiol and DHT concentrations in addition to normal testosterone concentrations. Although testosterone concentrations at the lower end of the normal male range can restore sexual function, it is not clear whether low-normal testosterone levels can maintain bone mineral density and muscle mass. The current recommendation is to restore testosterone levels to the mid-normal range.

Table 346-3 Clinical Pharmacology of Some Testosterone Formulations

Formulation	Regimen	Pharmacokinetic profile	DHT and E2	Advantages	Disadvantages
Testosterone enanthate or cypionate	150–200 mg IM q 2 weeks or 75-100 mg/week	After a single IM injection, serum T levels rise into the supraphysiologic range, then decline gradually into the hypogonadal range by the end of the dosing interval	DHT and E_2 levels rise in proportion to the increase in T levels; T:DHT and T:E_2 ratios do not change	Corrects symptoms of androgen deficiency; relatively inexpensive if self-administered; flexibility of dosing	Requires IM injection; peaks and valleys in serum T levels
1% Testosterone gel	Available in sachets, tubes, and pumps 5–10 g T gel containing 50–100 mg T qid	Restores serum T and E_2 levels to physiologic male range	Serum DHT levels are higher and T:DHT ratios are lower in hypogonadal men treated with the T gel than in healthy	Corrects symptoms of androgen deficiency; provides flexibility of dosing, ease of application; good skin tolerability	Potential of transfer to a female partner or child by direct skin-to-skin contact; skin irritation in a small proportion of treated men; moderately high DHT levels

			eugonadal men		
Transdermal testosterone patch	1 or 2 patches, designed to nominally deliver 5–10 mg T over 24 h applied qid on nonpressure areas	Restores serum T, DHT, and E_2 levels to physiologic male range	T:DHT and T:E_2 levels are in physiologic male range	Ease of application; corrects symptoms of androgen deficiency	Serum T levels in some androgen-deficient men may be in the low-normal range; these men may need application of 2 patches daily; skin irritation at the application site occurs frequently in many patients
Buccal, bioadhesive, T tablets	30-mg controlled-release, bioadhesive tablets bid	Absorbed from buccal mucosa	Normalizes serum T and DHT levels in hypogonadal men	Corrects symptoms of androgen deficiency in healthy, hypogonadal men	Gum-related adverse events in 16% of treated men.
Testosterone pellets	3–6 pellets implanted SC; dose and regimen vary with formulation	Serum T peaks at 1 month and then is sustained in normal range for 3–6 months, depending on formulation	T:DHT and T:E_2 ratios do not change	Corrects symptoms of androgen deficiency	Requires surgical incision for insertions; pellets may extrude spontaneously
17-α-methyl testosterone	This 17-α-alkylated compound should *not* be used because of potential for liver toxicity.	Orally active			Clinical responses are variable; potential for liver toxicity; should *not* be used for treatment of androgen deficiency
Oral testosterone undecanoate*	40–80 mg PO bid or tid with meals	When administered in oleic acid, T undecanoate is absorbed through lymphatics, bypassing portal system; considerable variability in the same individual on different days and among individuals	High DHT to T ratio	Convenience of oral administration	Not approved in the United States; variable clinical responses, variable serum T levels, high DHT:T ratio
Injectable long-acting testosterone undecanoate in oil*	European regimen 1000 mg IM, followed by 1000 mg at 6 weeks and 1000 mg q 10–14 weeks	When administered at a dose of 750–1000 mg IM, serum T levels are maintained in the normal range in a majority of treated men	DHT and E_2 levels rise in proportion to increase in T levels; T:DHT and T:E_2 ratios do not change	Corrects symptoms of androgen deficiency; requires infrequent administration.	Requires IM injection of a large volume (4 mL); cough reported immediately after injection in a very small number of men
Testosterone-inadhesive matrix patch*	2 x 60 cm^2 patches delivering approximately 4.8 mg of T/d	Restores serum T, DHT, and E_2 to physiologic range	T:DHT and T:E_2 are in physiologic range.	Lasts 2 d	Some skin irritation

*These formulations are not approved for clinical use in the United States but are available outside the United States in many countries. Physicians in countries where these formulations are available should follow the approved drug regimens.

Abbreviatons: DHT, dihydrotestosterone; E_2, estradiol; T, testosterone.

ORAL DERIVATIVES OF TESTOSTERONE

Testosterone is well-absorbed after oral administration but is quickly degraded during the first pass through the liver. Therefore, it is difficult to achieve sustained blood levels of testosterone after oral administration of crystalline testosterone. 17α-Alkylated derivatives of testosterone (e.g., 17α-methyl testosterone, oxandrolone, fluoxymesterone) are relatively resistant to hepatic degradation and can be administered orally; however, because of the potential for hepatotoxicity, including cholestatic jaundice, peliosis, and hepatoma, these formulations should not be used for testosterone replacement. Hereditary angioedema due to C1 esterase deficiency is the only exception to this general recommendation; in this condition, oral 17α-alkylated androgens are useful because they stimulate hepatic synthesis of the C1 esterase inhibitor.

INJECTABLE FORMS OF TESTOSTERONE

The esterification of testosterone at the 17β-hydroxy position makes the molecule hydrophobic and extends its duration of action. The slow release of testosterone ester from an oily depot in the muscle accounts for its extended duration of action. The longer the side chain, the greater the hydrophobicity of the ester, and the longer the duration of action. Thus, testosterone enanthate, cypionate, and undecanoate with longer side chains have longer durations of action than does testosterone propionate. Within 24 hours after intramuscular administration of 200 mg testosterone enanthate or cypionate, testosterone levels rise into the high-normal or supraphysiologic range and then gradually decline into the hypogonadal range over the next 2 weeks. A bimonthly regimen of testosterone enanthate or cypionate therefore results in peaks and troughs in testosterone levels that are accompanied by changes in a patient's mood, sexual desire, and energy level. The kinetics of testosterone enanthate and cypionate are similar. Estradiol and DHT levels are normal if testosterone replacement is physiologic.

TRANSDERMAL TESTOSTERONE PATCH

Nongenital testosterone patches, when applied in an appropriate dose, can normalize testosterone, DHT, and estradiol levels 4–12 hours after application. Sexual function and well-being are restored in androgen-deficient men treated with the nongenital patch. One 5-mg patch may not be sufficient to increase testosterone into the mid-normal male range in all hypogonadal men; some patients may need two 5-mg patches daily to achieve the targeted testosterone concentrations. The use of testosterone patches may be associated with skin irritation in some individuals.

TESTOSTERONE GEL

Two testosterone gels, Androgel and Testim, when applied topically to the skin in 5-, 7.5-, and 10-g doses, can maintain total and free testosterone concentrations in the mid- to high-normal range in hypogonadal men. The current recommendations are to begin with a 50-mg dose and adjust the dose on the basis of testosterone levels. The advantages of the testosterone gel include ease of application and flexibility of dosing. A major concern is the potential for inadvertent transfer of the gel to a sexual partner or to children who may come in close contact with the patient. The ratio of DHT to testosterone concentrations is higher in men treated with the testosterone gel than in healthy men. Also, there is considerable intra- and interindividual variation in serum testosterone levels in men treated with the transdermal gel.

BUCCAL ADHESIVE TESTOSTERONE

A buccal testosterone tablet that adheres to the buccal mucosa and releases testosterone as it is slowly dissolved has been approved. After twice-daily application of 30-mg tablets, serum testosterone levels are maintained within the normal male range in a majority of treated hypogonadal men. The adverse effects include buccal ulceration and gum problems in a few subjects. The effects of food and brushing on absorption have not been studied in detail.

Implants of crystalline testosterone can be inserted in the subcutaneous tissue by means of a trocar through a small skin incision. Testosterone is released by surface erosion of the implant and absorbed into the systemic circulation. Two to six 200-mg implants can maintain testosterone in the mid- to high-normal range for up to 6 months. Potential drawbacks include incising the skin for insertion and removal and spontaneous extrusions and fibrosis at the site of the implant.

TESTOSTERONE FORMULATIONS NOT AVAILABLE IN THE UNITED STATES

Testosterone undecanoate, when administered orally in oleic acid, is absorbed preferentially through the lymphatics into the systemic circulation and is spared first-pass degradation in the liver. Doses of 40–80 mg orally, two or three times daily, are typically used. However, the clinical responses are variable and suboptimal. Ratios of DHT to testosterone are higher in hypogonadal men treated with oral testosterone undecanoate compared with eugonadal men.

After initial priming, long-acting testosterone undecanoate in oil, when administered intramuscularly every 12 weeks, maintains serum testosterone, estradiol, and DHT in the normal male range and corrects symptoms of androgen deficiency in a majority of treated men. However, the large injection volume (4 mL) is a relative drawback.

NOVEL ANDROGEN FORMULATIONS

A number of androgen formulations with better pharmacokinetics or more selective activity profiles are under development. Two long-acting esters, testosterone buciclate and testosterone undecanoate, when injected intramuscularly, can maintain circulating testosterone concentrations in the male range for 7–12 weeks. Initial clinical trials have demonstrated the feasibility of administering testosterone by the sublingual or buccal route. 7α-Methyl-19-nortestosterone is an androgen that cannot be 5α-reduced; therefore, compared to testosterone, it has relatively greater agonist activity in muscle and gonadotropin suppression but lesser activity on the prostate.

Selective androgen receptor modulators (SARMs) are a class of androgen receptor ligands that bind the androgen receptor and display tissue-selective actions. A number of nonsteroidal SARMs that act as full agonists on the muscle and bone and spare the prostate to varying degrees have advanced to phase I and II human trials. Nonsteroidal SARMs do not serve as substrates for either the steroid 5α-reductase or the CYP19 aromatase. SARM binding to AR induces specific conformational changes in the AR protein, which then modulates protein-protein interactions between AR and its coregulators, resulting in tissue-specific regulation of gene expression.

PHARMACOLOGIC USES OF ANDROGENS

Androgens and SARMs are being evaluated as anabolic therapies for functional limitations associated with aging and chronic illness. Testosterone supplementation increases skeletal muscle mass, maximal voluntary strength, and muscle power in healthy men, hypogonadal men, older men with low testosterone levels, HIV-infected men with weight loss, and men receiving glucocorticoids. These anabolic effects of testosterone are related to testosterone dose and circulating concentrations. Systematic reviews have confirmed that testosterone therapy in HIV-infected men with weight loss promotes improvements in body weight, lean body mass, muscle strength, and depression indices, leading to the recommendation that testosterone be considered as an adjunctive therapy in HIV-infected men who are experiencing unexplained weight loss and have low testosterone levels. Similarly, in glucocorticoid-treated men, testosterone therapy should be considered to maintain muscle mass and strength and vertebral bone mineral density. It is not known whether testosterone therapy in older men with functional limitations is safe and effective in improving physical function and health-related quality of life and reducing disability. Concerns about potential adverse effects of testosterone on prostate and cardiovascular event rates have encouraged the development of selective androgen receptor modulators that are preferentially anabolic and spare the prostate.

Testosterone administration induces hypertrophy of both types 1 and 2 fibers and increases satellite cell (muscle progenitor cells) and myonuclear numbers. Androgens promote the differentiation of mesenchymal, multipotent progenitor cells into the myogenic lineage and inhibit their differentiation into the adipogenic lineage. Testosterone may have additional effects on satellite cell replication and muscle protein synthesis that may contribute to an increase in skeletal muscle mass.

Other indications for androgen therapy are in selected patients with anemia due to bone marrow failure (an indication largely supplanted by erythropoietin) and for hereditary angioedema.

MALE HORMONAL CONTRACEPTION BASED ON COMBINED ADMINISTRATION OF TESTOSTERONE AND GONADOTROPIN INHIBITORS

Supraphysiologic doses of testosterone (200 mg testosterone enanthate weekly) suppress LH and FSH secretion and induce azoospermia in 50% of white men and >95% of Asian men. The World Health Organization (WHO)-supported multicenter efficacy trials have demonstrated that suppression of spermatogenesis to azoospermia or severe oligozoospermia (<3 million/mL) by administration of testosterone enanthate to men results in effective contraception. Because of concern about long-term adverse effects of supraphysiologic testosterone doses, regimens that combine other gonadotropin inhibitors such as GnRH antagonists and progestins, with replacement doses of testosterone are being investigated. Oral etonogestrel daily in combination with intramuscular testosterone decanoate every 4–6 weeks induced azoospermia or severe oligozoospermia (sperm density <1 million/mL) in 99% of treated men over a 1-year period. This regimen was associated with weight gain, deceased testicular volume, and decreased plasma high- density lipoprotein (HDL) cholesterol, and its long-term safety has not been demonstrated. Selective androgen receptor modulators that are more potent inhibitors of gonadotropins than testosterone and spare the prostate hold promise for their contraceptive potential.

RECOMMENDED REGIMENS FOR ANDROGEN REPLACEMENT

Testosterone esters are administered typically at doses of 75–100 mg intramuscularly every week or 150–200 mg every 2 weeks. One or two 5-mg nongenital testosterone patches can be applied daily over the skin of the back, thigh, or upper arm away from pressure areas. Testosterone gel typically is applied over a covered area of skin at a dose of 5–10 g daily; patients should wash their hands after gel application. Bioadhesive buccal testosterone tablets at a dose of 30 mg typically are applied twice daily on the buccal mucosa.

ESTABLISHING EFFICACY OF TESTOSTERONE REPLACEMENT THERAPY

Because a clinically useful marker of androgen action is not available, restoration of testosterone levels into the mid-normal range remains the goal of therapy. Measurements of LH and FSH are not

useful in assessing the adequacy of testosterone replacement. Testosterone should be measured 3 months after initiating therapy to assess adequacy of therapy. There is substantial interindividual variability in serum testosterone levels, presumably due to genetic differences in testosterone clearance. In patients who are treated with testosterone enanthate or cypionate, testosterone levels should be 350–600 ng/dL 1 week after the injection. If testosterone levels are outside this range, adjustments should be made either in the dose or in the interval between injections. In men on transdermal patch or gel or buccal testosterone therapy, testosterone levels should be in the mid-normal range (500–700 ng/dL) 4–12 hours after application. If testosterone levels are outside this range, the dose should be adjusted.

Restoration of sexual function, secondary sex characteristics, energy, and well-being and maintenance of muscle and bone health are important objectives of testosterone replacement therapy. The patient should be asked about sexual desire and activity, the presence of early-morning erections, and the ability to achieve and maintain erections adequate for sexual intercourse. Some hypogonadal men continue to complain about sexual dysfunction even after testosterone replacement has been instituted; these patients may benefit from counseling. The hair growth in response to androgen replacement is variable and depends on ethnicity. Hypogonadal men with prepubertal onset of androgen deficiency who begin testosterone therapy in their late twenties or thirties may find it difficult to adjust to their newly found sexuality and may benefit from counseling. If the patient has a sexual partner, the partner should be included in counseling because of the dramatic physical and sexual changes that occur with androgen treatment.

CONTRAINDICATIONS FOR ANDROGEN ADMINISTRATION

Testosterone administration is contraindicated in men with a history of prostate or breast cancer **(Table 346-4)**. Testosterone therapy should not be administered without further urologic evaluation to men with a palpable prostate nodule or induration or prostate-specific antigen >4 ng/mL or >3 ng/mL in men at high risk for prostate cancer such as blacks or men with first-degree relatives with prostate cancer or with severe lower urinary tract symptoms (American Urological Association lower urinary tract symptom score >19). Testosterone replacement should not be administered to men with baseline hematocrit ≥50%, severe untreated obstructive sleep apnea, uncontrolled or poorly controlled congestive heart failure, or recent myocardial infarction or unstable angina.

Table 346-4 Conditions in Which Testosterone Administration Is Associated with a Risk of Adverse Outcome

Conditions in Which Testosterone Administration is Associated with Very High Risk of Serious Adverse Outcomes:
Metastatic prostate cancer
Breast cancer
Conditions in Which Testosterone Administration is Associated with Moderate to High Risk of Adverse Outcomes:
Undiagnosed prostate nodule or induration
PSA >4 ng/mL (>3 ng/mL in individuals at high risk for prostate cancer, such as blacks and men with first-degree relatives who have prostate cancer)
Erythrocytosis (hematocrit >50%)
Severe lower urinary tract symptoms associated with benign prostatic hypertrophy as indicated by the American Urological Association/International prostate symptom score >19
Uncontrolled or poorly controlled congestive heart failure

Abbreviation: PSA, prostate-specific antigen.

Source: Reproduced from the Endocrine Society Guideline for Testosterone Therapy of Androgen Deficiency Syndromes in Men (Bhasin et al).

MONITORING POTENTIAL ADVERSE EXPERIENCES

The clinical effectiveness and safety of testosterone replacement therapy should be assessed 3–6 months after initiating testosterone therapy and annually thereafter **(Table 346-5)**. Potential adverse effects include acne, oiliness of skin, erythrocytosis, breast tenderness and enlargement, leg edema, induction and exacerbation of obstructive sleep apnea, and increased risk of detection of prostate disease. In addition, there may be formulation-specific adverse effects such as skin irritation with transdermal patches, risk of gel transfer to a sexual partner with testosterone gels, buccal ulceration and gum problems with buccal testosterone, and pain and mood fluctuation with injectable testosterone esters.

Table 346-5 Monitoring Men Receiving Testosterone Therapy

1. Evaluate patient 3–6 months after treatment initiation and then annually to assess whether symptoms have responded to treatment and whether patient is experiencing any adverse effects.

2. Monitor testosterone level 3–6 months after initiation of testosterone therapy:

- Therapy should aim to raise serum testosterone level into mid-normal range.
- Injectable testosterone enanthate or cypionate: Measure serum testosterone level midway between injections. If testosterone is >700 ng/dL (24.5 nmol/L) or <400 ng/dL (14.1 nmol/L), adjust dose or frequency.
- Transdermal patches: Assess testosterone level 3–12 h after application of the patch; adjust dose to achieve testosterone level in mid-normal range.
- Buccal testosterone bioadhesive tablet: Assess level immediately before or after application of fresh system.
- Transdermal gels: Assess testosterone level any time after patient has been on treatment for at least 1 week; adjust dose to achieve serum testosterone level in the mid-normal range.
- Testosterone pellets: Measure testosterone levels at the end of dosing interval. Adjust number of pellets and/or dosing interval to achieve serum testosterone levels in normal range.
- Oral testosterone undecanoate*: Monitor serum testosterone level 3 to 5 h after ingestion.
- Injectable testosterone undecanoate: Measure serum testosterone level just before each subsequent injection and adjust dosing interval to maintain serum testosterone in mid-normal range.

3. Check hematocrit at baseline at 3–6 months and then annually. If hematocrit is >54%, stop therapy until hematocrit decreases to a safe level; evaluate patient for hypoxia and sleep apnea; reinitiate therapy with a reduced dose.

4. Measure bone mineral density of lumbar spine and/or femoral neck after 1–2 years of testosterone therapy in hypogonadal men with osteoporosis or low trauma fracture, consistent with regional standard of care.

5. In men 40 years of age or older with baseline PSA >0.6 ng/mL, perform digital rectal examination and check PSA level before initiating treatment at 3–6 months and then in accordance with guidelines for prostate cancer screening depending on age and race of patient.

6. Obtain urologic consultation if there is:

- An increase in serum PSA concentration >1.4 ng/mL within any 12-month period of testosterone treatment.
- A PSA velocity >0.4 ng/mL per year using PSA level after 6 months of testosterone administration as reference (applicable only if PSA data are available for a period exceeding 2 years).
- Detection of a prostatic abnormality on digital rectal examination.
- An AUA/IPSS prostate symptom score >19.

7. Evaluate formulation-specific adverse effects at each visit:

- Buccal testosterone tablets: Inquire about alterations in taste and examine gums and oral mucosa for irritation.
- Injectable testosterone esters (enanthate, cypionate, and undecanoate): Ask about fluctuations in mood or libido and, rarely, cough after injections.
- Testosterone patches: Look for skin reaction at application site.
- Testosterone gels: Advise patients to cover application sites with a shirt and wash skin with soap and water before having skin-to-skin contact because testosterone gels leave a testosterone residue on skin that can be transferred to a woman or child who comes in close contact. Serum testosterone levels are maintained when application site is washed 4–6 h after application of testosterone gel.
- Testosterone pellets: Look for signs of infection, fibrosis, or pellet extrusion.

*Not approved for clinical use in the United States. ***Abbreviations:*** AUA/IPSS, American Urological Association International Prostate Symptom Score; PSA, prostate-specific antigen.

Source: Reproduced with permission from the Endocrine Society Guideline for Testosterone Therapy of Androgen Deficiency Syndromes in Men (Bhasin et al).

Hemoglobin Levels

Administration of testosterone to androgen-deficient men typically is associated with a 3–5% increase in hemoglobin levels due to suppression of hepcidin and increased iron availability for erythropoiesis. The magnitude of hemoglobin increase during testosterone therapy is greater in older men than younger men and in men who have sleep apnea, a significant smoking history, or chronic obstructive lung disease. The frequency of erythrocytosis is higher in hypogonadal men treated with injectable testosterone esters than in those treated with transdermal formulations, presumably due to the higher testosterone dose delivered by the typical regimens of testosterone esters. Erythrocytosis is the most

common adverse event reported in testosterone trials in middle-aged and older men and also the most common cause of treatment discontinuation in these trials. If hematocrit rises above 54%, testosterone therapy should be stopped until hematocrit has fallen to <50%. After evaluation of the patient for hypoxia and sleep apnea, testosterone therapy may be reinitiated at a lower dose.

Prostate and Serum PSA Levels

Testosterone replacement therapy increases prostate volume to the size seen in age-matched controls but does not increase prostate volume beyond that expected for age. There is no evidence that testosterone therapy causes prostate cancer. However, androgen administration can exacerbate preexisting metastatic prostate cancer. Many older men harbor microscopic foci of cancer in their prostates. It is not known whether long-term testosterone administration will induce these microscopic foci to grow into clinically significant cancers.

Prostate-specific antigen (PSA) levels are lower in testosterone-deficient men and are restored to normal after testosterone replacement. There is considerable test-retest variability in PSA measurements. Increments in PSA levels after testosterone supplementation in androgen-deficient men are generally <0.5 ng/mL, and increments >1 ng/mL over a 3–6-month period are unusual. The 90% confidence interval for the change in PSA values in men with benign prostatic hypertrophy measured 3–6 months apart is 1.4 ng/mL. Therefore, the Endocrine Society expert panel suggests that an increase in PSA >1.4 ng/mL in any one year after starting testosterone therapy, if confirmed, should lead to urologic evaluation. PSA velocity criterion can be used for patients who have sequential PSA measurements for >2 years; a change of >0.40 ng/mL per year merits closer urologic follow-up.

Cardiovascular Risk

In epidemiologic studies, testosterone concentrations are negatively related to the risk of diabetes mellitus, heart disease, and all-cause and cardiovascular mortality. A recent testosterone trial in older men with mobility limitation was stopped early because of the higher rates of cardiovascular events in the testosterone arm than in the placebo arm of the trial. Meta-analyses of testosterone trials have found no statistically significant increase in cardiovascular event rates in men receiving testosterone therapy, although nonsignificant increases have been noted. Inferences about adverse events from previous trials included in these meta-analyses were limited by poor ascertainment, small numbers of events, and small numbers of participants. Adequately powered prospective studies are needed to determine the effect of testosterone replacement on cardiovascular risk.

ANDROGEN ABUSE BY ATHLETES AND RECREATIONAL BODYBUILDERS

The illicit use of androgenic-anabolic steroids (AAS) to enhance athletic performance first surfaced in the 1950s among power lifters and spread rapidly to other sports, professional as well as high school athletes, and recreational bodybuilders. In the early 1980s, the use of AAS spread beyond the athletic community into the general population, and now as many as 2 million Americans, most of them men, probably have used these compounds. The most commonly used androgenic steroids include testosterone esters, nandrolone, stanozolol, methandienone, and methenolol. Athletes generally use increasing doses of multiple steroids in a practice known as stacking.

The adverse effects of long-term AAS abuse are poorly understood. Most of the information about the adverse effects of AAS has emerged from case reports, uncontrolled studies, or clinical trials that used replacement doses of testosterone. The adverse event data from clinical trials using physiologic replacement doses of testosterone have been extrapolated unjustifiably to AAS users who may administer 10–100 times the replacement doses of testosterone over many years and to support the claim that AAS use is safe. A substantial fraction of androgenic steroid users also use other drugs that are perceived to be muscle-building or performance-enhancing such as growth hormone; IGF-I; insulin; stimulants such as amphetamine, clenbuterol, cocaine, ephedrine, and thyroxine; and drugs perceived to reduce adverse effects such as hCG, aromatase inhibitors, and estrogen antagonists. The men who abuse androgenic steroids are more likely to engage in other high-risk behaviors than are nonusers. The adverse events associated with AAS use may be due to AAS themselves, concomitant use of other drugs, high-risk behaviors, and host characteristics that may render these individuals more susceptible to AAS use or other high-risk behaviors.

The high rates of mortality and morbidity observed in AAS users are alarming. One Finnish study reported 4.6 times the risk of death among elite power lifters than among age-matched men from the general population. The causes of death among power lifters included suicides, myocardial infarction, hepatic coma, and non-Hodgkin's lymphoma. A retrospective review of patient records in Sweden also reported higher standardized mortality ratios for AAS users than for nonusers.

Numerous reports of cardiac death among young AAS users raise concerns about the adverse cardiovascular effects of AAS. High doses of AAS may induce proatherogenic dyslipidemia, increase thrombosis risk via effects on clotting factors and platelets, and induce vasospasm through their effects on vascular nitric oxide. The finding of androgen receptors on myocardial cells suggests that AAS may be directly toxic to myocardial cells.

Replacement doses of testosterone, when administered parenterally, are associated with only a small decrease in HDL cholesterol and little or no effect on total cholesterol, low-density lipoprotein (LDL)

cholesterol, and triglyceride levels. In contrast, supraphysiologic doses of testosterone and orally administered, 17α-alylated, nonaromatizable AAS are associated with marked reductions in HDL cholesterol and increases in LDL cholesterol.

Long-term AAS use suppresses LH and FSH secretion and inhibits endogenous testosterone production and spermatogenesis. Men who have used AAS for more than a few months experience suppression of the hypothalamic-pituitary- testicular (HPT) axis after stopping AAS that may be associated with sexual dysfunction, infertility, and depression; in some AAS users, gonadotropin suppression may last more than a year. The dysphoria caused by androgen withdrawal may cause some men to revert to using AAS, leading to continued use and AAS dependence. As many as 30% of AAS users develop a syndrome of AAS dependence that is characterized by long-term AAS use despite adverse medical and psychiatric effects.

Unsafe injection practices, high-risk behaviors, and increased rates of incarceration put AAS users at increased risk of HIV and hepatitis B and C. In one survey, nearly 1 in 10 gay men had injected AAS or other substances, and AAS users were more likely to report high-risk unprotected anal sex than were other men.

Some AAS users develop manic symptoms during AAS exposure (sometimes associated with violence) and major depression (sometimes associated with suicidality) during AAS withdrawal. Users also may engage in other forms of illicit drug use, which may be potentiated or exacerbated by AAS.

Elevated liver enzymes, cholestatic jaundice, hepatic neoplasms, and peliosis hepatis have been reported with oral 17α-alkylated AAS. AAS use may cause muscle hypertrophy without compensatory adaptations in tendons, ligaments, and joints, thus increasing the risk of tendon and joint injuries. AAS use is associated with acne and baldness, as well as increased body hair.

Accredited laboratories use gas chromatography–mass spectrometry or liquid chromatography–mass spectrometry to detect anabolic steroid abuse. In recent years, the availability of high-resolution mass spectrometry and tandem mass spectrometry has improved the sensitivity of detecting androgen abuse. Illicit testosterone use generally is detected by the application of the measurement of the ratio of urinary testosterone to epitestosterone and further confirmed by the use of the ^{13}C:^{12}C ratio in testosterone by the use of isotope ratio combustion mass spectrometry. Exogenous testosterone administration increases urinary testosterone glucuronide excretion and consequently the ratio of testosterone to epitestosterone. Ratios above 4 suggest exogenous testosterone use but can also reflect genetic variation. Synthetic testosterone has a lower ^{13}C:^{12}C ratio than endogenously produced testosterone, and these differences in the ^{13}C:^{12}C ratio can be detected by isotope ratio combustion mass spectrometry, which is used to confirm exogenous testosterone use in individuals with a high ratio of testosterone to epitestosterone.

FURTHER READINGS

Basaria S et al: Adverse events associated with testosterone administration. N Engl J Med 363:109, 2010[PMID: 20592293] [Full Text]

Bhasin S: An approach to infertile men. J Clin Endocrinol Metab 92:1995, 2007[PMID: 17554051] [Full Text]

——— et al: Testosterone therapy in men with androgen deficiency syndromes: An endocrine society clinical practice guideline. J Clin Endocrinol Metab 95:2536, 2010

Bolona ER et al: Testosterone use in men with sexual dysfunction: A systematic review and meta-analysis of randomized placebo-controlled trials. Mayo Clin Proc 82:20, 2007[PMID: 17285782] [Full Text]

Crowley WF Jr et al: New genes controlling human reproduction and how you find them. Trans Am Clin Climatol Assoc 119:29, 2008[PMID: 18596868] [Full Text]

Ferlin A et al: Molecular and clinical characterizations of Y chromosome microdeletions in infertile men: A 10-year experience in Italy. J Clin Endocrinol Metab 92:762, 2007[PMID: 17213277] [Full Text]

Fernández-Balsells MM et al: Clinical review: 1. Adverse effects of testosterone therapy in adult men: A systematic review and meta-analysis. J Clin Endocrinol Metab 95:2560, 2010[PMID: 21713498] [Full Text]

Kanayama G et al: Long-term psychiatric and medical consequences of anabolic-androgenic steroid abuse: A looming public health concern? Drug Alcohol Depend 98:1, 2008[PMID: 18599224] [Full Text]

Menke A et al: Sex steroid hormone concentrations and risk of death in US men. Am J Epidemiol 171:583, 2010[PMID: 20083549] [Full Text]

Sedlmeyer IL, Palmert MR: Delayed puberty: Analysis of a large case series from an academic center. J Clin Endocrinol Metab 87:1613, 2002[PMID: 11932291] [Full Text]

Srinivas-Shankar U et al: Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: A randomized, double-blind, placebo-controlled study. J Clin Endocrinol Metab 95:639, 2010[PMID: 20061435] [Full Text]

Wu F et al: Identification of late-onset hypogonadism in middle-aged and elderly men. N Engl J Med 363:123, 2010[PMID: 20554979] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE FEMALE REPRODUCTIVE SYSTEM, INFERTILITY, AND CONTRACEPTION: INTRODUCITON

The female reproductive system regulates the hormonal changes responsible for puberty and adult reproductive function. Normal reproductive function in women requires the dynamic integration of hormonal signals from the hypothalamus, pituitary, and ovary, resulting in repetitive cycles of follicle development, ovulation, and preparation of the endometrial lining of the uterus for implantation should conception occur.

For further discussion of related topics, see the following chapters: hyperandrogenic disorders (Chap. 49), menstrual cycle disorders (Chap. 50), gynecologic malignancies (Chap. 97), sexually transmitted diseases (Chap. 130), male hormonal contraception (Chap. 346), menopause (Chap. 348), and sexual differentiation (Chap. 349).

DEVELOPMENT OF THE OVARY AND EARLY FOLLICULAR GROWTH

The ovary orchestrates the development and release of a mature oocyte and also elaborates hormones (e.g., estrogen, progesterone, inhibin, relaxin) that are critical for pubertal development and preparation of the uterus for conception, implantation, and the early stages of pregnancy. To achieve these functions in repeated monthly cycles, the ovary undergoes some of the most dynamic changes of any organ in the body. Primordial germ cells can be identified by the third week of gestation and their migration to the genital ridge is complete by 6 weeks' gestation. Germ cells persist within the genital ridge, are then referred to as *oogonia*, and are essential for induction of ovarian development. Although one X chromosome undergoes X inactivation in somatic cells, it is reactivated in oogonia and genes on both X chromosomes are required for normal ovarian development. A streak ovary containing only stromal cells is found in patients with 45,X Turner's syndrome (Chap. 349).

The germ cell population expands, and starting at ~8 weeks' gestation, oogonia begin to enter prophase of the first meiotic division and become primary oocytes. This allows the oocyte to be surrounded by a single layer of flattened granulosa cells to form a primordial follicle **(Fig. 347-1)**. Granulosa cells are derived from mesonephric cells that invade the ovary early in its development, pushing the germ cells to the periphery. Although recent studies have reopened the debate, the weight of evidence strongly supports the concept that the ovary contains a nonrenewable pool of germ cells. Through the combined processes of mitosis, meiosis, and atresia, the population of oogonia reaches its maximum of 6–7 million by 20 weeks' gestation, after which there is a progressive loss of both oogonia and primordial follicles through the process of atresia. At birth, oogonia are no longer present in the ovary, and only 1–2 million germ cells remain in the form of primordial follicles **(Fig. 347-2)**. The oocyte persists in prophase of the first meiotic division until just before ovulation, when meiosis resumes.

Figure 347-1

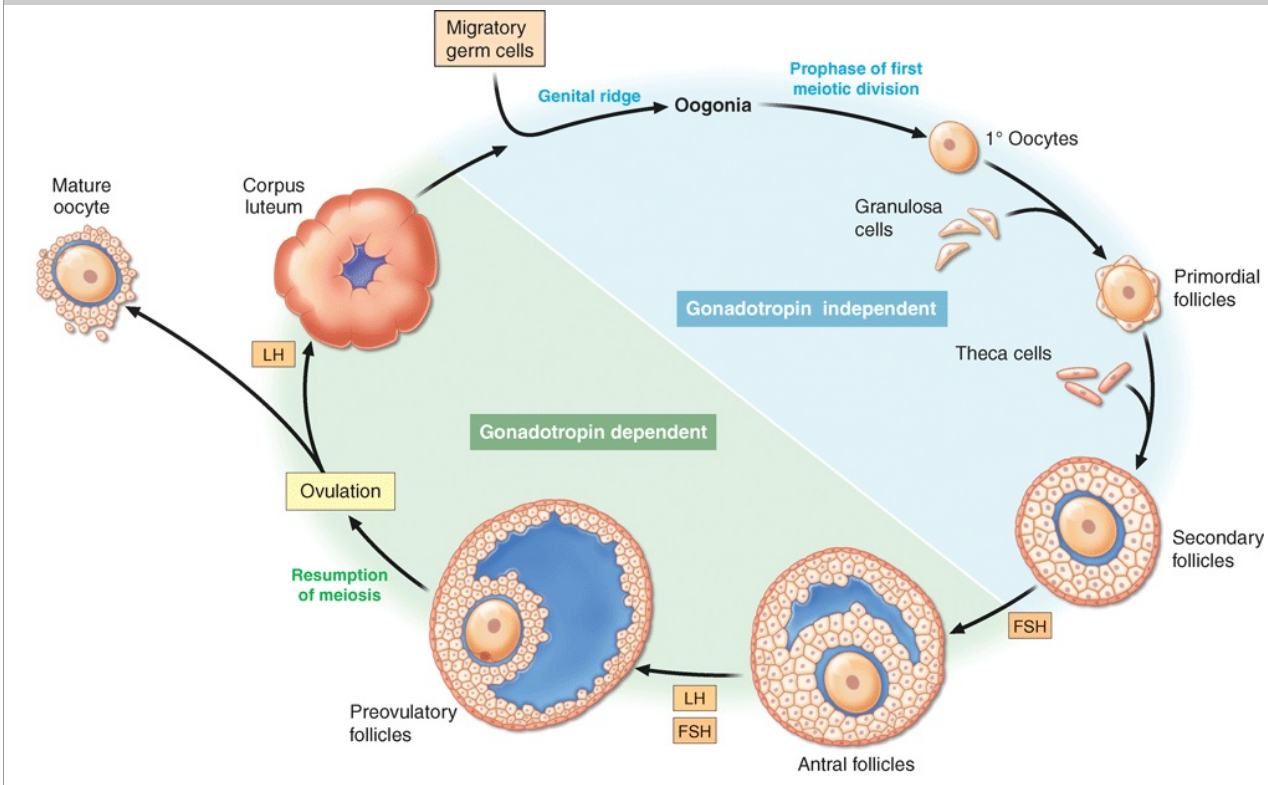

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Stages of ovarian development from the arrival of the migratory germ cells at the genital ridge through gonadotropin-independent and gonadotropin-dependent phases that ultimately result in ovulation of a mature oocyte. FSH, follicle-stimulating hormone; LH, luteinizing hormone.

Figure 347-2

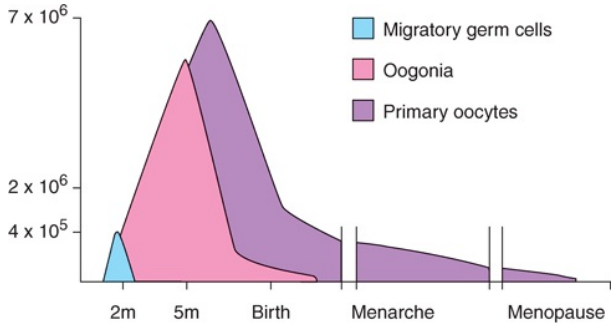

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ovarian germ cell number is maximal at mid-gestation, then decreases precipitously.

The quiescent primordial follicles are recruited to further growth and differentiation through a highly regulated process that limits the size of the developing cohort to ensure that folliculogenesis can continue throughout the reproductive life span. This initial recruitment of primordial follicles to form primary follicles (Fig. 347-1) is characterized by growth of the oocyte and the transition from squamous to cuboidal granulosa cells. The theca interna cells that surround the developing follicle begin to form as the primary follicle grows. Acquisition of a zona pellucida by the oocyte and the presence of several layers of surrounding cuboidal granulosa cells mark the development of secondary follicles. It is at this stage that granulosa cells develop follicle-stimulating hormone (FSH), estradiol, and androgen receptors and communicate with one another through the development of gap junctions.

Bidirectional signaling between the germ cells and the somatic cells in the ovary are a necessary component underlying the maturation of the oocyte and the capacity for hormone secretion. For example, the oocyte-derived factor in the germline α(FIGα) is required for initial follicle formation. Anti-müllerian hormone [AMH, also known as mullerian inhibiting substance (MIS)] and activins derived from somatic cells induce the development of primary follicles from primordial follicles. Oocyte-derived growth differentiation factor 9 (GDF-9) is required for migration of pre-theca cells to the outer surface of the developing follicle. GDF-9 is also required for formation of secondary follicles, as are granulosa cell–derived KIT ligand (KITL) and the forkhead transcription factor (FOXL2). All of these genes are potential candidates for premature ovarian failure in women, and mutations in the human *FOXL2* gene have been shown to cause the syndrome of blepharophimosis/ ptosis/epicanthus inversus, which is associated with ovarian failure.

DEVELOPMENT OF A MATURE FOLLICLE

The early stages of follicle growth are primarily driven by intraovarian factors and may take up to a year from the time of initial recruitment. Maturation to the state required for ovulation, including the resumption of meiosis in the oocyte, requires the combined stimulus of FSH and luteinizing hormone (LH) (Fig. 347-1) and can be accomplished within weeks. This phase of recruitment of secondary follicles from the resting follicle pool requires the direct action of FSH. Accumulation of follicular fluid between the layers of granulosa cells creates an antrum that divides the granulosa cells into two functionally distinct groups: mural cells that line the follicle wall and cumulus cells that surround the oocyte **(Fig. 347-3)**. Recent evidence suggests that, in addition to its role in normal development of the mullerian system, the WNT signaling pathway is required for normal antral follicle development and may also play a role in ovarian steroidogenesis. A single dominant follicle emerges from the growing follicle pool within the first 5–7 days after the onset of menses, and the majority of follicles fall off their growth trajectory and become atretic. Autocrine actions of activin and bone morphogenic protein 6 (BMP-6), derived from the granulosa cells, and paracrine actions of GDF-9, BMP-15, BMP-6, and Gpr149, derived from the oocyte, are involved in granulosa cell proliferation and modulation of FSH responsiveness. Differential exposure to these factors may explain why one follicle is selected for continued growth to the preovulatory stage. The dominant follicle can be distinguished by its size, evidence of granulosa cell proliferation, large number of FSH receptors, high aromatase activity, and elevated concentrations of estradiol and inhibin A in follicular fluid.

Figure 347-3

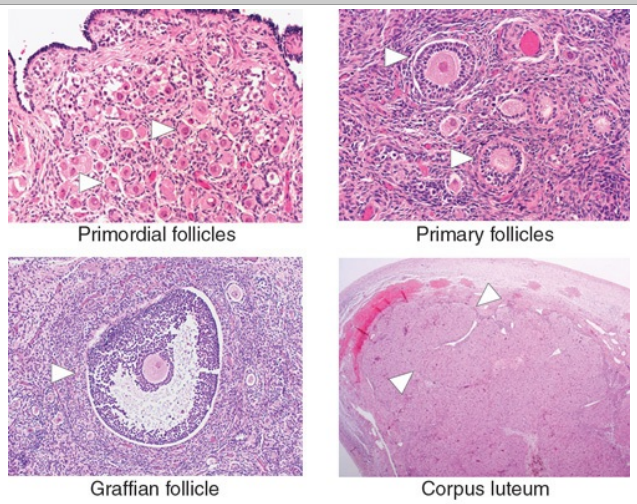

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Development of ovarian follicles. The Graffian follicle is also known as a tertiary or preovulatory follicle. *(Courtesy of JH*

Eichhorn and D Roberts, Massachusetts General Hospital; with permission.)

The dominant follicle undergoes rapid expansion during the 5–6 days prior to ovulation, reflecting granulosa cell proliferation and accumulation of follicular fluid. FSH induces LH receptors on the granulosa cells, and the preovulatory, or Graffian, follicle moves to the outer ovarian surface in preparation for ovulation. The LH surge triggers the resumption of meiosis, the suppression of granulosa cell proliferation, and the induction of cyclooxygenase 2 (COX-2), prostaglandins, and the progesterone receptor, each of which is required for ovulation. EGF-like factors are thought to mediate these follicular responses to LH. Ovulation also involves production of extracellular matrix leading to expansion of the cumulus cell population that surrounds the oocyte and the controlled expulsion of the egg and follicular fluid. Both progesterone and prostaglandins (induced by the ovulatory stimulus) are essential for this process. After ovulation, luteinization is induced by LH in conjunction with the acquisition of a rich vascular network in response to vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (basic FGF). Traditional regulators of central reproductive control, gonadotropin-releasing hormone (GnRH) and its receptor (GnRHR), are also produced in the ovary and may be involved in corpus luteum function.

REGULATION OF OVARIAN FUNCTION

Hypothalamic and Pituitary Secretion

GnRH neurons develop from epithelial cells outside the central nervous system and migrate, initially alongside the olfactory neurons, to the medial basal hypothalamus. Studies in GnRH-deficient patients who fail to undergo puberty have provided insights into genes that control the ontogeny and function of GnRH neurons **(Fig. 347-4)**. *KAL1*, *FGF8/FGFR1*, *PROK2/PROKR2*, *NELF* and *CDH7* (Chap. 346) have been implicated in the migration of GnRH neurons to the hypothalamus. Approximately 7000 GnRH neurons, scattered throughout the medial basal hypothalamus, establish contacts with capillaries of the pituitary portal system in the median eminence. GnRH is secreted into the pituitary portal system in discrete pulses to stimulate synthesis and secretion of LH and FSH from pituitary gonadotropes, which comprise ~10% of cells in the pituitary (Chap. 339). Functional connections of GnRH neurons with the portal system are established by the end of the first trimester, coinciding with the production of pituitary gonadotropins. Thus, like the ovary, the hypothalamic and pituitary components of the reproductive system are present before birth. However, the high levels of estradiol and progesterone produced by the placenta suppress hypothalamic-pituitary stimulation of ovarian hormonal secretion in the fetus.

Figure 347-4

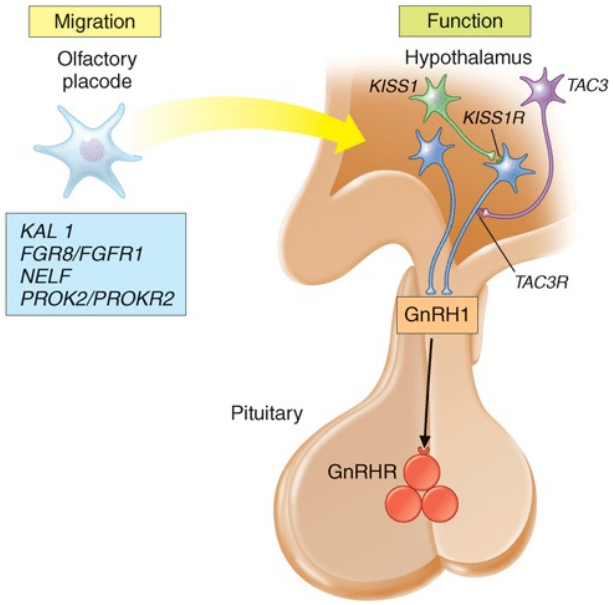

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Establishment of a functional GnRH system requires the participation of a number of genes that are essential for development and migration of GnRH neurons from the olfactory placode to the hypothalamus in addition to genes involved in the functional control of GnRH secretion and action.

After birth and the loss of placenta-derived steroids, gonadotropin levels rise. FSH levels are much higher in girls than in boys. This rise in FSH results in ovarian activation (evident on ultrasound) and increased inhibin B and estradiol levels. Studies that have identified mutations in *TAC3*, which encodes neurokinin B, and its receptor, *TAC3R*, in patients with GnRH deficiency indicate that both are involved in control of GnRH secretion and may be particularly important at this early stage of development. By 12–20 months of age, the reproductive axis is again suppressed, and a period of relative quiescence persists until puberty **(Fig. 347-5)**. At the onset of puberty, pulsatile GnRH secretion induces pituitary gonadotropin production. In the early stages of puberty, LH and FSH secretion are apparent only during sleep, but as puberty develops, pulsatile gonadotropin secretion occurs throughout the day and night.

Figure 347-5

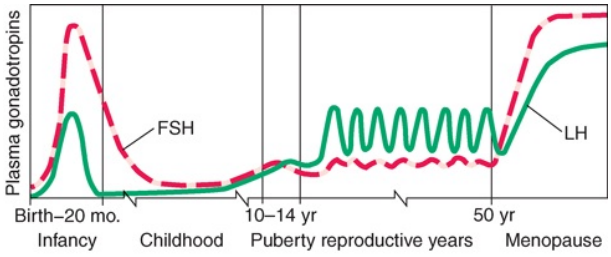

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Follicle-stimulating hormone (FSH) and luteinizing hormone (LH) are increased during the neonatal years but go through a period of childhood quiescence before increasing again during puberty. Gonadotropin levels are cyclic during the reproductive years and increase dramatically with the loss of negative feedback that accompanies menopause.

The mechanisms responsible for the childhood quiescence and pubertal reactivation of the reproductive axis remain incompletely understood. GnRH neurons in the hypothalamus respond to both excitatory and inhibitory factors. Increased sensitivity to the inhibitory influence of gonadal steroids has long been implicated in the inhibition of GnRH secretion during childhood but has not been definitively established in the human. Metabolic signals such as adipocyte-derived leptin, play a permissive role in reproductive function (Chap. 77). Studies of patients with isolated GnRH deficiency reveal that mutations in the G protein–coupled receptor 54 (*GPR54*) gene (now known as *KISS1R*) preclude the onset of puberty. The ligand for this receptor, metastin, is derived from the parent peptide, kisspeptin-1 (*KISS1*), and is a powerful stimulant for GnRH release. A potential role for kisspeptin in the onset of puberty has been suggested by upregulation of *KISS1* and *KISS1R* transcripts in the hypothalamus at the time of puberty. The KISS1/KISS1R system may also be involved in estrogen feedback regulation of GnRH secretion.

Ovarian Steroids

Ovarian steroid-producing cells do not store hormones but produce them in response to LH and FSH during the normal menstrual cycle. The sequence of steps and the enzymes involved in the synthesis of steroid hormones are similar in the ovary, adrenal, and testis. However, the enzymes required to catalyze specific steps are compartmentalized and may not be abundant or even present in all cell types. Within the developing ovarian follicle, estrogen synthesis from cholesterol requires close integration between theca and granulosa cells—sometimes called the *two-cell model for steroidogenesis* **(Fig. 347-6)**. FSH receptors are confined to the granulosa cells, whereas LH receptors are restricted to the theca cells until the late stages of follicular development, when they are also found on granulosa cells. The theca cells surrounding the follicle are highly vascularized and use cholesterol, derived primarily from circulating lipoproteins, as the starting point for the synthesis of androstenedione and testosterone under the control of LH. Androstenedione and testosterone are transferred across the basal lamina to the granulosa cells, which receive no direct blood supply. The mural granulosa cells are particularly rich in aromatase and, under the control of FSH, produce estradiol, the primary steroid secreted from the follicular phase ovary and the most potent estrogen. Theca cell–produced androstenedione and, to a lesser extent, testosterone are also secreted into peripheral blood, where they can be converted to dihydrotestosterone in skin and to estrogens in adipose tissue. The hilar interstitial cells of the ovary are functionally similar to Leydig cells and are also capable of secreting androgens. Although stromal cells proliferate in response to androgens [as in polycystic ovarian syndrome (PCOS)], they do not secrete androgens.

Figure 347-6

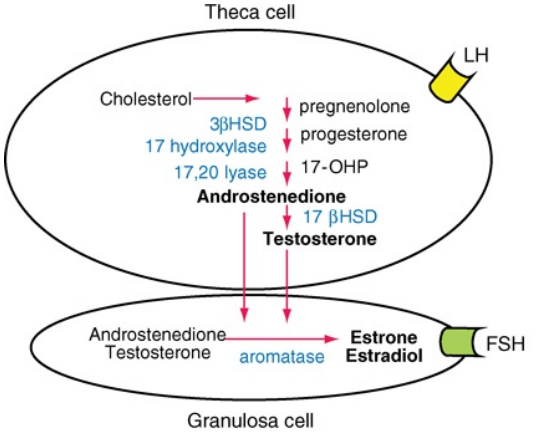

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Estrogen production in the ovary requires the cooperative function of the theca and granulosa cells under the control of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). HSD, hydroxysteroid dehydrogenase; OHP, hydroxyprogesterone.

Development of the rich capillary network following rupture of the follicle at the time of ovulation makes it possible for large molecules such as low-density lipoprotein (LDL) to reach the luteinized granulosa and theca lutein cells. As in the follicle, both cell types are required for steroidogenesis in the corpus luteum. The luteinized granulosa cells are the main source of progesterone production whereas the theca lutein cells produce 17-hydroxyprogesterone, a substrate for aromatization to estradiol by the luteinized granulosa cells. LH is critical for normal structure and function of the corpus luteum. Because LH and human chorionic gonadotropin (hCG) bind to a common receptor, the role of LH in support of the corpus luteum can be replaced by hCG in the first 10 weeks after conception, and hCG is commonly used for luteal phase support in the treatment of infertility.

STEROID HORMONE ACTIONS

Both estrogen and progesterone play critical roles in the expression of secondary sexual characteristics in women (Chap. 338). Estrogen promotes development of the ductule system in the breast, whereas progesterone is responsible for glandular

development. In the reproductive tract, estrogens create a receptive environment for fertilization and support pregnancy and parturition through carefully coordinated changes in the endometrium, thickening of the vaginal mucosa, thinning of the cervical mucus, and uterine growth and contractions. Progesterone induces secretory activity in the estrogen-primed endometrium, increases the viscosity of cervical mucus, and inhibits uterine contractions. Both gonadal steroids play critical roles in the negative and positive feedback controls of gonadotropin secretion. Progesterone also increases basal body temperature and has therefore been used clinically as a marker of ovulation.

The vast majority of circulating estrogens and androgens are carried in the blood bound to carrier proteins, which restrain their free diffusion into cells and prolong their clearance, serving as a reservoir. High-affinity binding proteins include sex hormone –binding globulin (SHBG), which binds androgens with somewhat greater affinity than estrogens, and corticosteroid-binding globulin (CBG), which also binds progesterone. Modulations in binding protein levels by insulin, androgens, and estrogens contribute to high bioavailable testosterone levels in PCOS and to high circulating estrogen and progesterone levels during pregnancy.

Estrogens act primarily through binding to the nuclear receptors, estrogen receptor (ER) α and β. Transcriptional coactivators and co-repressors modulate ER action (Chap. 338). Both ER subtypes are present in the hypothalamus, pituitary, ovary, and reproductive tract. Although ERα and -β exhibit some functional redundancy, there is also a high degree of specificity, particularly in cell type expression. For example, ERα functions in the ovarian theca cells, whereas ERβ is critical for granulosa cell function. There is also evidence for membrane-initiated signaling by estrogen. Similar signaling mechanisms pertain for progesterone with evidence of transcriptional regulation through progesterone receptor (PR) A and B protein isoforms, as well as rapid membrane signaling.

Ovarian Peptides

Inhibin was initially isolated from gonadal fluids based on its ability to selectively inhibit FSH secretion from pituitary cells. Inhibin is a heterodimer composed of an α-subunit and a βA- or βB-subunit to form inhibin A or inhibin B, both of which are secreted from the ovary. Activin is a homodimer of inhibin β subunits with the capacity to stimulate the synthesis and secretion of FSH. Inhibins and activins are members of the transforming growth factor β (TGF-β) superfamily of growth and differentiation factors. During the purification of inhibin, follistatin, an unrelated monomeric protein that inhibits FSH secretion, was discovered. Within the pituitary, follistatin inhibits FSH secretion indirectly through binding and neutralizing activin.

Inhibin B is secreted from the granulosa cells of small antral follicles, whereas inhibin A is present in both granulosa and theca cells and is secreted by dominant follicles. Inhibin A is also present in luteinized granulosa cells and is a major secretory product of the corpus luteum. Inhibin B is constitutively secreted by granulosa cells and increases in serum in conjunction with cycle recruitment to the pool of actively growing follicles under the control of FSH. Inhibin B has been used clinically as a marker of ovarian reserve. Inhibin B is an important inhibitor of FSH, independent of estradiol, during the menstrual cycle. Although activin is also secreted from the ovary, the excess of follistatin in serum, combined with its nearly irreversible binding of activin, make it unlikely that ovarian activin plays an endocrine role in FSH regulation. However, there is evidence that activin plays an autocrine/paracrine role in the ovary, in addition to its intrapituitary role in modulation of FSH production.

AMH (also known as MIS) is important in ovarian biology in addition to the function from which it derived it name (i.e., promotion of the degeneration of the müllerian system during embryogenesis in the male). AMH is produced by granulosa cells and, like inhibin B, is a marker of ovarian reserve. AMH may also inhibit the recruitment of primordial follicles into the follicle pool and appears to increase the effect of FSH on aromatase expression.

Relaxin, which is produced by the theca lutein cells of the corpus luteum, is thought to play a role in decidualization of the endometrium and suppression of myometrial contractile activity, both of which are essential for the early establishment of pregnancy.

HORMONAL INTEGRATION OF THE NORMAL MENSTRUAL CYCLE

The sequence of changes responsible for mature reproductive function is coordinated through a series of negative and positive feedback loops that alter pulsatile GnRH secretion, the pituitary response to GnRH, and the relative secretion of LH and FSH from the gonadotrope. The frequency and amplitude of pulsatile GnRH secretion differentially modulate the synthesis and secretion of LH and FSH, with slow frequencies favoring FSH synthesis and increased amplitudes favoring LH synthesis. Activin is produced in both pituitary gonadotropes and folliculostellate cells and stimulates the synthesis and secretion of FSH. Inhibins function as potent antagonists of activins through sequestration of the activin receptors. Although inhibin is expressed in the pituitary, gonadal inhibin is the principal source of feedback inhibition of FSH.

For the majority of the cycle, the reproductive system functions in a classic endocrine negative feedback mode. Estradiol and progesterone inhibit GnRH secretion, and the inhibins act at the pituitary to selectively inhibit FSH synthesis and secretion **(Fig. 347-7)**. This negative feedback control of FSH is critical for development of the single mature oocyte that characterizes normal reproductive function in women. In addition to these negative feedback controls, the menstrual cycle is uniquely dependent on estrogen-induced positive feedback to produce an LH surge that is essential for ovulation of a mature follicle. The neural signaling pathways that distinguish estrogen negative versus positive feedback are incompletely understood.

Figure 347-7

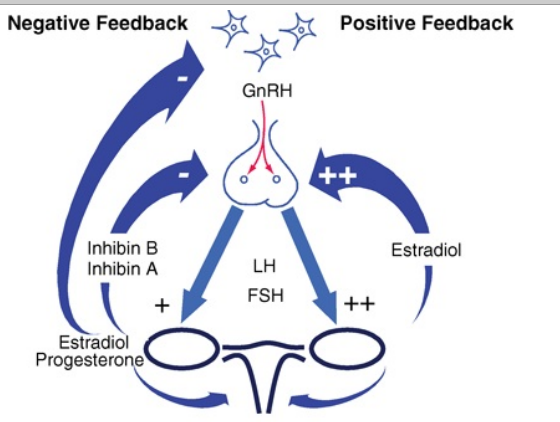

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The reproductive system in women is critically dependent on both negative feedback of gonadal steroids and inhibin to modulate follicle-stimulating hormone (FSH) secretion and on estrogen positive feedback to generate the preovulatory luteinizing hormone (LH) surge. GnRH, gonadotropin-releasing hormone.

The Follicular Phase

The follicular phase is characterized by recruitment of a cohort of secondary follicles and the ultimate selection of a dominant preovulatory follicle **(Fig. 347-8)**. The follicular phase begins, by convention, on the first day of menses. However, follicle recruitment is initiated by the rise in FSH that begins in the late luteal phase in conjunction with the loss of negative feedback of gonadal steroids and likely inhibin A. The fact that a 20–30% increase in FSH is adequate for follicular recruitment speaks to the marked sensitivity of the resting follicle pool to FSH. The resultant granulosa cell proliferation is responsible for increasing early follicular phase levels of inhibin B. Inhibin B in conjunction with rising levels of estradiol, and probably inhibin A, restrain FSH secretion during this critical period such that only a single follicle matures in the vast majority of cycles. The increased risk of multiple gestation associated with the increased levels of FSH characteristic of advanced maternal age, or with exogenous gonadotropin administration in the treatment of infertility, attest to the importance of the negative feedback regulation of FSH. With further growth of the dominant follicle, estradiol and inhibin A increase exponentially and the follicle acquires LH receptors. Increasing levels of estradiol are responsible for proliferative changes in the endometrium. The exponential rise in estradiol results in positive feedback on the pituitary, leading to the generation of an LH surge (and a smaller FSH surge), thereby triggering ovulation and luteinization of the granulosa cells.

Figure 347-8

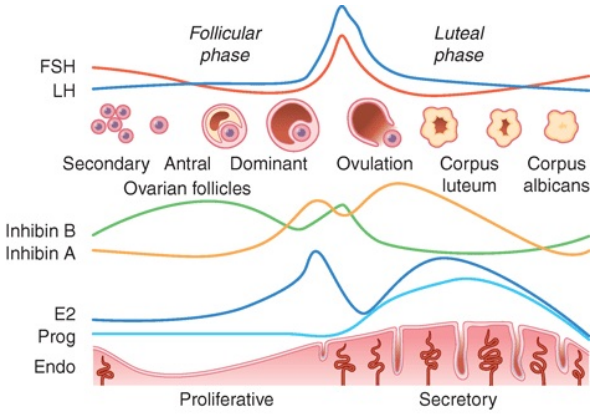

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship between gonadotropins, follicle development, gonadal secretion, and endometrial changes during the normal menstrual cycle. FSH, follicle-stimulating hormone; LH, luteinizing hormone; E_2, estradiol; Prog, progesterone; Endo, endometrium.

The Luteal Phase

The luteal phase begins with the formation of the corpus luteum from the ruptured follicle. Progesterone and inhibin A are produced from the luteinized granulosa cells, which continue to aromatize theca-derived androgen precursors, producing estradiol. The combined actions of estrogen and progesterone are responsible for the secretory changes in the endometrium that are necessary for implantation. The corpus luteum is supported by LH but has a finite life span because of diminished sensitivity to LH. The demise of the corpus luteum results in a progressive decline in hormonal support of the endometrium. Inflammation or local hypoxia and ischemia result in vascular changes in the endometrium leading to the release of cytokines, cell death, and shedding of the endometrium.

If conception occurs, hCG produced by the trophoblast binds to LH receptors on the corpus luteum, maintaining steroid hormone production and preventing involution of the corpus luteum. The corpus luteum is essential for the hormonal maintenance of the endometrium during the first 6–10 weeks of pregnancy until this function is taken over by the placenta.

CLINICAL ASSESSMENT OF OVARIAN FUNCTION

Menstrual bleeding should become regular within 2 to 4 years of menarche, although anovulatory and irregular cycles are common before that. For the remainder of adult reproductive life, the cycle length counted from the first day of menses to the first day of subsequent menses, is ~28 days, with a range of 25–35 days. However, cycle-to-cycle variability for an individual woman is ±2 days. Luteal phase length is relatively constant between 12 and 14 days in normal cycles; thus, the major variability in cycle length is due to variations in the follicular phase. The duration of menstrual bleeding in ovulatory cycles varies between 4 and 6 days. There is a gradual shortening of cycle length with age such that women over the age of 35 have cycles that are shorter than during their younger reproductive years. Anovulatory cycles increase as women approach the menopause, and bleeding patterns may be erratic.

Women who report regular monthly bleeding with cycles that do not vary by >4 days generally have ovulatory cycles, but several other clinical signs can be used to assess the likelihood of ovulation. Some women experience *mittelschmerz*, described as midcycle pelvic discomfort that is thought to be caused by the rapid expansion of the dominant follicle at the time of ovulation. A constellation of premenstrual moliminal symptoms such as bloating, breast tenderness, and food cravings often occur several days before menses in ovulatory cycles, but their absence cannot be used as evidence of anovulation. Methods that can be used to determine whether ovulation is likely to include a serum progesterone level >5 ng/mL ~7 days before expected menses, an increase in basal body temperature of 0.24°C (>0.5°F) in the second half of the cycle due to the thermoregulatory effect of progesterone, or the detection of the urinary LH surge using ovulation predictor kits. Because ovulation occurs ~36 hours after the LH surge, urinary LH can be helpful in timing intercourse to coincide with ovulation.

Ultrasound can be used to detect the growth of the fluid-filled antrum of the developing follicle and to assess endometrial proliferation in response to increasing estradiol levels in the follicular phase, as well as the characteristic echogenicity of the secretory endometrium of the luteal phase.

PUBERTY

Normal Pubertal Development in Girls

The first menstrual period (*menarche*) occurs relatively late in the series of developmental milestones that characterize normal pubertal development **(Table 347-1)**. Menarche is preceded by the appearance of pubic and then axillary hair as a result of maturation of the zona reticularis in the adrenal gland and increased adrenal androgen secretion, particularly dehydroepiandrosterone (DHEA). The triggers for adrenarche remain unknown but may involve increases in body mass index as well as in utero and neonatal factors. Menarche is also preceded by breast development (*thelarche*), which is exquisitely sensitive to the very low levels of estrogen that result from peripheral conversion of adrenal androgens and the low levels of estrogen secreted from the ovary early in pubertal maturation. Breast development precedes the appearance of pubic and axillary hair in ~60% of girls. The interval between the onset of breast development and menarche is ~2 years. There has been a gradual decline in the age of menarche over the past century, attributed in large part to improvement in nutrition, and there is a relationship between adiposity and earlier sexual maturation in girls. In the United States, menarche occurs at an average age of 12.5 years (Table 347-1). Much of the variation in the timing of puberty is due to genetic factors, with heritability estimates of 50–80%. Both adrenarche and breast development occur ~1 year earlier in black compared with white girls, although the timing of menarche differs by only 6 months between these ethnic groups.

Table 347-1 Mean Age (Years) of Pubertal Milestones in Girls

	Onset of Breast/Pubic Hair Development	Age of Peak Height Velocity	Menarche	Final Breast/Pubic Hair Development	Adult Height
White	10.2	11.9	12.6	14.3	17.1
Black	?9.6	11.5	12	13.6	16.5

Source: From FM Biro et al: J Pediatr 148:234, 2006.

Other important hormonal changes also occur in conjunction with puberty. Growth hormone (GH) levels increase early in puberty, stimulated in part by the pubertal increases in estrogen secretion. GH increases insulin-like growth factor-I (IGF-I), which enhances linear growth. The growth spurt is generally less pronounced in girls than in boys, with a peak growth velocity of ~7 cm/year. Linear growth is ultimately limited by closure of epiphyses in the long bones as a result of prolonged exposure to estrogen. Puberty is also associated with mild insulin resistance.

Disorders of Puberty

The differential diagnosis of precocious and delayed puberty is similar in boys (Chap. 346) and girls. However, there are differences in the timing of normal puberty and differences in the relative frequency of specific disorders in girls compared with boys.

PRECOCIOUS PUBERTY

Traditionally, precocious puberty has been defined as the development of secondary sexual characteristics before the age of 8 in girls based on data from Marshall and Tanner in British girls studied in the 1960s. More recent studies led to recommendations that girls be evaluated for precocious puberty if breast development or pubic hair are present at <7 years of age for white girls or <6 years for black girls.

Precocious puberty is most often centrally mediated **(Table 347-2)**, resulting from early activation of the hypothalamic-pituitary-ovarian axis. It is characterized by pulsatile LH secretion and an enhanced LH and FSH response to exogenous GnRH (two- to threefold stimulation) **(Table 347-3)**. True precocity is marked by advancement in bone age of >2 SD, a recent history of growth acceleration, and progression of secondary sexual characteristics. In girls, centrally mediated precocious puberty is idiopathic in ~85% of cases; however, neurogenic causes must also be considered. GnRH agonists that induce pituitary desensitization are the mainstay of treatment to prevent premature epiphyseal closure and preserve adult height, as well as to manage psychosocial repercussions of precocious puberty.

Table 347-2 Differential Diagnosis of Precocious Puberty

Central (GnRH Dependent)	Peripheral (GnRH Independent)
Idiopathic	Congenital adrenal hyperplasia
CNS tumors	Estrogen-producing tumors
Hamartomas	Adrenal tumors
Astrocytomas	Ovarian tumors
Adenomyomas	Gonadotropin/hCG-producing tumors
Gliomas Germinomas	Exogenous exposure to estrogen or androgen
CNS infection	McCune-Albright syndrome
Head trauma	Aromatase excess syndrome
Iatrogenic	
Radiation Chemotherapy	
Surgical	
CNS malformation	
Arachnoid or suprasellar cysts	
Septo-optic dysplasia Hydrocephalus	

Abbreviations: CNS, central nervous system; GnRH, gonadotropin-releasing hormone; hCG, human chorionic gonadotropin.

Table 347-3 Evaluation of Precocious and Delayed Puberty

	Precocious	Delayed
Initial Screening Tests		
History and physical	x	x
Assessment of growth velocity	x	x

Bone age	x	x
LH, FSH	x	x
Estradiol, testosterone	x	x
DHEAS	x	x
17-Hydroxyprogesterone	x	
TSH, T4	x	x
Complete blood count		x
Sedimentation rate, C-reactive protein		x
Electrolytes, renal function		x
Liver enzymes		x
IGF-I, IGFBP-3		x
Urinalysis		x
Secondary Tests		
Pelvic ultrasound	x	x
Cranial MRI	x	x
β-hCG	x	
GnRH/agonist stimulation test	x	x
ACTH stimulation test	x	
Inflammatory bowel disease panel	x	x
Celiac disease panel		x
Prolactin		x
Karyotype		x

Abbreviations: LH, luteinizing hormone; FSH, follicle-stimulating hormone; DHEAS, dehydroepiandrosterone sulfate; TSH, thyroid-stimulating hormone; T4, thyroxine; IGF-I, insulin-like growth factor-I; IGFBP-3, IGF-binding protein 3; hCG, human chorionic gonadotropin; ACTH, adrenocorticotropic hormone.

Peripherally mediated precocious puberty does not involve activation of the hypothalamic-pituitary-ovarian axis and is characterized by suppressed gonadotropins in the presence of elevated estradiol. Management of peripheral precocious puberty involves treating the underlying disorder (Table 347-2) and limiting the effects of gonadal steroids using aromatase inhibitors, inhibitors of steroidogenesis, and estrogen receptor blockers. It is important to be aware that central precocious puberty can also develop in girls whose precocity was initially peripherally mediated, as in McCune-Albright syndrome and congenital adrenal hyperplasia.

Incomplete and intermittent forms of precocious puberty may also occur. For example, premature breast development may occur in girls before the age of 2 years, with no further progression and without significant advancement in bone age, androgen production, or compromised height. Premature adrenarche can also occur in the absence of progressive pubertal development, but it must be distinguished from late-onset congenital adrenal hyperplasia and androgen-secreting tumors, in which case it may be termed *heterosexual precocity*. Premature adrenarche may be associated with obesity, hyperinsulinemia, and the subsequent predisposition to PCOS.

DELAYED PUBERTY

Delayed puberty **(Table 347-4)** is defined as the absence of secondary sexual characteristics by age 13 in girls. The diagnostic considerations are very similar to those for primary amenorrhea (Chap. 50). Between 25 and 40% of delayed puberty in girls is of ovarian origin, with Turner's syndrome accounting for the majority of such patients. Functional hypogonadotropic hypogonadism encompasses diverse etiologies such as systemic illnesses, including celiac disease and chronic renal disease, and endocrinopathies such as diabetes and hypothyroidism. In addition, girls appear to be particularly susceptible to the adverse effects of abnormalities in energy balance that result from exercise, dieting, and/or eating disorders. Together these reversible conditions account for ~25% of delayed puberty in girls. Congenital hypogonadotropic hypogonadism in girls or boys can be caused by mutations in several different genes or combinations of genes (Fig. 347-4, Chap. 346, Table 346-2). Family studies suggest that genes identified in association with absent puberty may also cause delayed puberty, and recent reports have further suggested that a genetic susceptibility to environmental stresses such as diet and exercise may account for at least some cases of functional hypothalamic amenorrhea. Although neuroanatomic causes of delayed puberty are considerably less common in girls than in boys, it is always important to rule these out in the setting of hypogonadotropic hypogonadism.

Table 347-4 Differential Diagnosis of Delayed Puberty

Hypergonadotropic
Ovarian
Turner's syndrome
Gonadal dysgenesis
Chemotherapy/radiation therapy
Galactosemia
Autoimmune oophoritis
Congenital lipoid hyperplasia
Steroidogenic enzyme abnormalities
17α-hydroxylase deficiency
Aromatase deficiency
Gonadotropin/receptor mutations
FSHβ, LHR, FSHR
Androgen resistance syndrome
Hypogonadotropic
Genetic
Hypothalamic syndromes

Leptin/leptin receptor
HESX1 (septo-optic dysplasia)
PC1 (prohormone convertase)
IHH and Kallmann syndrome
KAL1, FGF8, FGFR1, NELF, PROK2, PROKR2
KISS1, KISS1R, TAC3, TAC3R, GnRH1, GnRHR
Abnormalities of pituitary development/function
PROP1
CNS tumors/infiltrative disorders
Craniopharyngioma
Astrocytoma, germinoma, glioma
Prolactinomas, other pituitary tumors
Histiocytosis X
Chemotherapy/radiation
Functional
Chronic diseases
Malnutrition
Excessive exercise
Eating disorders

Abbreviations: CNS, central nervous system; FGF8, fibroblast growth factor 8; FGFR1, fibroblast growth factor 1 receptor; FSH β, follicle-stimulating hormone β chain; FSHR, FSH receptor; GnRHR, gonadotropin-releasing hormone receptor; HESX1, homeobox, embryonic stem cell expressed 1; IHH, idiopathic hypogonadotropic hypogonadism; KAL, Kallmann; KISS1, kisspeptin 1; KISSR1, KISS1 receptor; LHR, luteinizing hormone receptor; NELF, nasal embryonic LHRH factor; PROK2, prokineticin 2; PROKR2 prokineticin receptor 2; PROP1, prophet of Pit1, paired-like homeodomain transcription factor.

INFERTILITY

Definition and Prevalence

Infertility is defined as the inability to conceive after 12 months of unprotected sexual intercourse. In a study of 5574 English and American women who ultimately conceived, pregnancy occurred in 50% within 3 months, 72% within 6 months, and 85% within 12 months. These findings are consistent with predictions based on *fecundability*, the probability of achieving pregnancy in one menstrual cycle (approximately 20–25% in healthy young couples). Assuming a fecundability of 0.25, 98% of couples should conceive within 13 months. Based on this definition, the National Survey of Family Growth reports a 14% rate of infertility in the United States in married women aged 15–44. The infertility rate has remained relatively stable over the past 30 years, although the proportion of couples without children has risen, reflecting a trend to delay childbearing. This trend has important implications because of an age-related decrease in fecundability, which begins at age 35 and decreases markedly after age 40. It is estimated that ~8% of women in the United States have received medical assistance for infertility; of these, 74% received counseling, ~60% underwent infertility testing of the female and/or male partner, and ~46% used ovulation-inducing medication.

Causes of Infertility

The spectrum of infertility ranges from reduced conception rates or the need for medical intervention to irreversible causes of infertility. Infertility can be attributed primarily to male factors in 25%, female factors in 58%, and is unexplained in about 17% of couples **(Fig. 347-9)**. Not uncommonly, both male and female factors contribute to infertility.

Figure 347-9

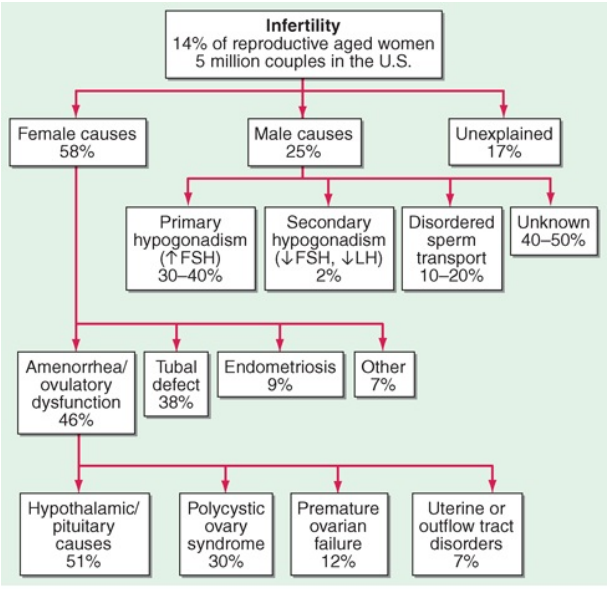

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Causes of infertility. FSH, follicle-stimulating hormone; LH, luteinizing hormone.

Approach to the Patient: Infertility

INITIAL EVALUATION

In all couples presenting with infertility, the initial evaluation includes discussion of the appropriate timing of intercourse and discussion of modifiable risk factors such as smoking, alcohol, caffeine, and obesity. The range of required investigations should be reviewed as well as a brief description of infertility treatment options, including adoption. Initial investigations are focused on determining whether the primary cause of the infertility is male, female, or both. These investigations include a semen analysis in the male, confirmation of ovulation in the female, and, in the majority of situations, documentation of tubal patency in the female. In some cases, after an extensive workup excluding all male and female factors, a specific cause cannot be identified and infertility may ultimately be classified as unexplained.

PSYCHOLOGICAL ASPECTS OF INFERTILITY

Infertility is invariably associated with psychological stress related not only to the diagnostic and therapeutic procedures themselves but also to repeated cycles of hope and loss associated with each new procedure or cycle of treatment that does not result in the birth of a child. These feelings are often combined with a sense of isolation from friends and family. Counseling and stress-management techniques should be introduced early in the evaluation of infertility. Importantly, infertility and its treatment do not appear to be associated with long-term psychological sequelae.

FEMALE CAUSES

Abnormalities in menstrual function constitute the most common cause of female infertility. These disorders, which include ovulatory dysfunction and abnormalities of the uterus or outflow tract, may present as amenorrhea or as irregular or short menstrual cycles. A careful history and physical examination and a limited number of laboratory tests will help to determine whether the abnormality is: (1) hypothalamic or pituitary (low FSH, LH, and estradiol with or without an increase in prolactin), (2) PCOS (irregular cycles and hyperandrogenism in the absence of other causes of androgen excess), (3) ovarian (low estradiol with increased FSH), or (4) uterine or outflow tract abnormality. The frequency of these diagnoses depends on whether the amenorrhea is primary or occurs after normal puberty and menarche (see Fig. 50-2).

The approach to further evaluation of these disorders is described in detail in Chap. 50.

Ovulatory Dysfunction

In women with a history of regular menstrual cycles, *evidence of ovulation* should be sought as described above. An endometrial biopsy to exclude luteal phase insufficiency is no longer considered a usual part of the infertility workup. Even in the presence of ovulatory cycles, evaluation of *ovarian reserve* is recommended for women aged >35 years. Measurement of FSH on day 3 of the cycle (an FSH level <10 IU/mL on cycle day 3 predicts adequate ovarian oocyte reserve) or in response to clomiphene (blocks estrogen negative feedback on FSH), antral follicle count and serum levels of inhibin B and AMH have all been used for this purpose.

Tubal Disease

Tubal dysfunction may result from pelvic inflammatory disease (PID), appendicitis, endometriosis, pelvic adhesions, tubal surgery, previous use of an intrauterine device (IUD), and a previous ectopic pregnancy. However, a cause is not identified in up to 50% of patients with documented tubal factor infertility. Because of the high prevalence of tubal disease, evaluation of tubal patency by hysterosalpingogram (HSG) or laparoscopy should occur early in the majority of couples with infertility. Subclinical infections with *Chlamydia trachomatis* may be an underdiagnosed cause of tubal infertility and requires the treatment of both partners.

Endometriosis

Endometriosis is defined as the presence of endometrial glands or stroma outside the endometrial cavity and uterine musculature. Its presence is suggested by a history of dyspareunia (painful intercourse), worsening dysmenorrhea that often begins before menses, or by a thickened rectovaginal septum or deviation of the cervix on pelvic examination. The pathogenesis of the infertility associated with endometriosis is unclear but may involve effects on fertilization, normal function of the endometrium, as well as adhesions. Endometriosis is often clinically silent, however, and can only be excluded definitively by laparoscopy.

MALE CAUSES

(See also Chap. 346) Known causes of male infertility include primary testicular disease, disorders of sperm transport, and hypothalamic-pituitary disease resulting in secondary hypogonadism. However, the etiology is not ascertained in up to one-half of men with suspected male factor infertility. The key initial diagnostic test is a *semen analysis*. Testosterone levels should be measured if the sperm count is low on repeated examination or if there is clinical evidence of hypogonadism. Gonadotropin levels will help to determine a gonadal versus a central cause of hypogonadism.

Treatment: Infertility

In addition to addressing the negative impact of smoking on fertility and pregnancy outcome, counseling about nutrition and weight is a critical part of infertility and pregnancy management. Both low and increased body mass index (BMI) are associated with infertility in women and with increased morbidity during pregnancy. Obesity has also been associated with infertility in men. The treatment of infertility should be tailored to the problems unique to each couple. In many situations, including unexplained infertility, mild-to-moderate endometriosis, and/or borderline semen parameters, a stepwise approach to infertility is optimal, beginning with low-risk interventions and moving to more invasive, higher risk interventions only if necessary. After determination of all infertility factors and their correction, if possible, this approach might include, in increasing order of complexity: (1) expectant management, (2) clomiphene citrate (see below) with or without intrauterine insemination (IUI), (3) gonadotropins with or without IUI, and (4) in vitro fertilization (IVF). The time used to complete the evaluation, correction, and expectant management can be longer in women aged <30 years, but this process should be advanced rapidly in women aged >35 years. In some situations, expectant management will not be appropriate.

OVULATORY DYSFUNCTION

Treatment of ovulatory dysfunction should first be directed at identification of the etiology of the disorder to allow specific management when possible. Dopamine agonists, for example, may be indicated in patients with hyperprolactinemia (Chap. 339); lifestyle modification may be successful in women with obesity, low body weight, or a history of intensive exercise (Chap. 79).

Medications used for ovulation induction include clomiphene citrate, gonadotropins, and pulsatile GnRH. *Clomiphene citrate* is a nonsteroidal estrogen antagonist that increases FSH and LH levels by blocking estrogen negative feedback at the hypothalamus. The efficacy of clomiphene for ovulation induction is highly dependent on patient selection. It induces ovulation in ~60% of women with PCOS and is the initial treatment of choice. It may be combined with agents that modify insulin levels such as metformin. Clomiphene citrate is less successful in patients with hypogonadotropic hypogonadism. *Aromatase inhibitors* have also been investigated for the treatment of infertility. Initial studies are promising, but these medications have not been approved for this indication.

Gonadotropins are highly effective for ovulation induction in women with hypogonadotropic hypogonadism and PCOS and are used to induce the development of multiple follicles in unexplained infertility and in older reproductive-age women.

Disadvantages include a significant risk of multiple gestation and the risk of ovarian hyperstimulation. Careful monitoring and a conservative approach to ovarian stimulation reduce these risks. Currently available gonadotropins include urinary preparations of LH and FSH, highly purified FSH, and recombinant FSH. Though FSH is the key component, there are growing data that the addition of some LH (or hCG) may improve results and this is particularly important in hypogonadotropic patients.

None of these methods are effective in women with premature ovarian failure in whom donor oocyte or adoption are the methods of choice.

TUBAL DISEASE

If hysterosalpingography suggests a tubal or uterine cavity abnormality, or if a patient is aged ≥35 at the time of initial evaluation, laparoscopy with tubal lavage is recommended, often with a hysteroscopy. Although tubal reconstruction may be attempted if tubal disease is identified, it is generally being replaced by the use of IVF. These patients are at increased risk of developing an ectopic pregnancy.

ENDOMETRIOSIS

Though 60% of women with minimal or mild endometriosis may conceive within 1 year without treatment, laparoscopic resection or ablation appears to improve conception rates. Medical management of advanced stages of endometriosis is widely used for symptom control but has not been shown to enhance fertility. In moderate-to-severe endometriosis, conservative surgery is associated with pregnancy rates of 50 and 39%, respectively, compared with rates of 25 and 5% with expectant management alone. In some patients, IVF may be the treatment of choice.

MALE FACTOR INFERTILITY

The treatment options for male factor infertility have expanded greatly in recent years (Chap. 346). Secondary hypogonadism is highly amenable to treatment with pulsatile GnRH or gonadotropins. In vitro techniques have provided new opportunities for patients with primary testicular failure and disorders of sperm transport. Choice of initial treatment options depends on sperm concentration and motility. Expectant management should be attempted initially in men with mild male factor infertility (sperm count of 15 to 20 x 10^6/mL and normal motility). Moderate male factor infertility (10 to 15 x 10^6/mL and 20–40% motility) should begin with IUI alone or in combination with treatment of the female partner with clomiphene or gonadotropins, but it may require IVF with or without intracytoplasmic sperm injection (ICSI). For men with a severe defect (sperm count of <10 x 10^6/mL, 10% motility), IVF with ICSI or donor sperm should be used. If ICSI is performed because of azoospermia due to congenital bilateral absence of the vas deferens, genetic testing and counseling should be provided because of the risk of cystic fibrosis.

ASSISTED REPRODUCTIVE TECHNOLOGIES

The development of assisted reproductive technologies (ARTs) has dramatically altered the treatment of male and female infertility. IVF is indicated for patients with many causes of infertility that have not been successfully managed with more conservative approaches. IVF or ICSI is often the treatment of choice in couples with a significant male factor or tubal disease, whereas IVF using donor oocytes is used in patients with premature ovarian failure and in women of advanced reproductive age. Success rates depend on the age of the woman and the cause of the infertility. The number of cycles canceled increases from approximately 10% in women aged <35 years to 24% in women aged >40 while the birth rate decreases from ~39% of cycles in which embryos were transferred in women aged <35 to 24% in women aged 38-40 and only 10% in women aged >40. Though often effective, IVF is expensive and requires careful monitoring of ovulation induction and invasive techniques, including the aspiration of multiple follicles. IVF is associated with a significant risk of multiple gestation, particularly in women aged <35, in whom it can be as high as 40%.

CONTRACEPTION

Only 15% of couples in the United States report having unprotected sexual intercourse in the past 3 months. However, despite the wide availability and widespread use of a variety of effective methods of contraception, approximately one-half of all births in the United States are the result of unintended pregnancy. Teenage pregnancies continue to represent a serious public health problem in the United States, with >1 million unintended pregnancies each year—a significantly greater incidence than in other industrialized nations.

Of the contraceptive methods available **(Table 347-5)**, a reversible form of contraception is used by >50% of couples, while sterilization (male or female) has been employed as a permanent form of contraception by over one-third of couples. Pregnancy termination is relatively safe when directed by health care professionals, but is rarely the option of choice.

Table 347-5 Effectiveness of Different Forms of Contraception

Method of Contraception	Theoretical[a] Effectiveness, %	Actual[a] Effectiveness, %	Percent Continuing Use at 1 Year[b]	Contraceptive Methods Used by U.S. Women[c]
Barrier methods				
Condoms	98	88	63	18
Diaphragm	94	82	58	2
Cervical cap	94	82	50	<1
Spermicides	97	79	43	1
Sterilization				
Male	99.9	99.9	100	9
Female	99.8	99.6	100	27
Intrauterine device				1
Copper T380	99	97	78	
Progestasert	98	97	81	
Mirena	99.9	99.8		
Hormonal contraceptives	99.7	92	72	31
Combination pill				
Progestin only pill				
Transdermal patch				
Vaginal ring				

Long-acting progestins				
Depo-Provera	99.7	99.7	70	9

[a]Adapted from J Trussel et al: Obstet Gynecol 76:558, 1990.

[b]Adapted from Contraceptive Technology Update. Contraceptive Technology, Feb. 1996, Vol 17, No 1, pp 13–24.

[c]Adapted from LJ Piccinino and WD Mosher: Fam Plan Perspective 30:4, 1998.

No single contraceptive method is ideal, although all are safer than carrying a pregnancy to term. The effectiveness of a given method of contraception does not just depend on the efficacy of the method itself. Discrepancies between theoretical and actual effectiveness emphasize the importance of patient education and compliance when considering various forms of contraception (Table 347-5). Knowledge of the advantages and disadvantages of each contraceptive is essential for counseling an individual about the methods that are safest and most consistent with his or her lifestyle. The World Health Organization (WHO) has extensive family planning resources for the physician and patient that can be accessed online. Similar resources for determining medical eligibility are available through the Centers for Disease Control and Prevention (CDC). Considerations for contraceptive use in obese patients and after bariatric surgery are discussed below.

Barrier Methods

Barrier contraceptives (such as condoms, diaphragms, and cervical caps) and spermicides are easily available, reversible, and have fewer side effects than hormonal methods. However, their effectiveness is highly dependent on adherence and proper use (Table 347-5). A major advantage of barrier contraceptives is the protection provided against sexually transmitted infections (STIs) (Chap. 130). Consistent use is associated with a decreased risk of HIV, gonorrhea, nongonococcal urethritis, and genital herpes, probably due in part to the concomitant use of spermicides. Natural membrane condoms may be less effective than latex condoms, and petroleum-based lubricants can degrade condoms and decrease their efficacy for preventing HIV infection. A highly effective female condom, which also provides protection against STIs, was approved in 1994 but has not achieved widespread use.

Sterilization

Sterilization is the method of birth control most frequently chosen by fertile men and multiparous women >30 (Table 347-5). Sterilization refers to a procedure that prevents fertilization by surgical interruption of the fallopian tubes in women or the vas deferens in men. Although tubal ligation and vasectomy are potentially reversible, these procedures should be considered permanent and should not be undertaken without patient counseling.

Several methods of *tubal ligation* have been developed, all of which are highly effective with a 10-year cumulative pregnancy rate of 1.85 per 100 women. However, when pregnancy does occur, the risk of ectopic pregnancy may be as high as 30%. The success rate of tubal reanastomosis depends on the method used, but even after successful reversal, the risk of ectopic pregnancy remains high. In addition to prevention of pregnancy, tubal ligation reduces the risk of ovarian cancer, possibly by limiting the upward migration of potential carcinogens.

Vasectomy is a highly effective outpatient surgical procedure that has little risk. The development of azoospermia may be delayed for 2–6 months, and other forms of contraception must be used until two sperm-free ejaculations provide proof of sterility. Reanastomosis may restore fertility in 30–50% of men, but the success rate declines with time after vasectomy and may be influenced by factors such as the development of antisperm antibodies.

Intrauterine Devices

IUDs inhibit pregnancy primarily through a spermicidal effect caused by a sterile inflammatory reaction induced by the presence of a foreign body in the uterine cavity (copper IUDs) or by the release of progestins (Progestasert, Mirena). IUDs provide a high level of efficacy in the absence of systemic metabolic effects, and ongoing motivation is not required to ensure efficacy once the device has been placed. However, only 1% of women in the United States use this method compared to a utilization rate of 15–30% in much of Europe and Canada, despite evidence that the newer devices are not associated with increased rates of pelvic infection and infertility, as occurred with earlier devices. An IUD should not be used in women at high risk for development of STI or in women at high risk for bacterial endocarditis. The IUD may not be effective in women with uterine leiomyomas because they alter the size or shape of the uterine cavity. IUD use is associated with increased menstrual blood flow, although this is less pronounced with the progestin-releasing IUD, which is associated with a more frequent occurrence of spotting or amenorrhea.

Hormonal Methods

ORAL CONTRACEPTIVE PILLS

Because of their ease of use and efficacy, oral contraceptive pills are the most widely used form of hormonal contraception. They act by suppressing ovulation, changing cervical mucus, and altering the endometrium. The current formulations are made from synthetic estrogens and progestins. The estrogen component of the pill consists of ethinyl estradiol or mestranol, which is metabolized to ethinyl estradiol. Multiple synthetic progestins are used. Norethindrone and its derivatives are used in many formulations. Low-dose norgestimate and the more recently developed progestins (desogestrel, gestodene, drospirenone) have a less androgenic profile; levonorgestrel appears to be the most androgenic of the progestins and should be avoided in patients with hyperandrogenic symptoms. The three major formulations of oral contraceptives are (1) fixed-dose estrogen-progestin combination, (2) phasic estrogen-progestin combination, and (3) progestin only. Each of these formulations is administered daily for 3 weeks followed by a week of no medication during which menstrual bleeding generally occurs. Two extended oral contraceptives are approved for use in the United States; Seasonale is a 3-month preparation with 84 days of active drug and 7 days of placebo, and Lybrel is a continuous preparation containing 90 μg of levonorgestrel and 20 μg of ethinyl estradiol. Current doses of ethinyl estradiol range from 20 to 50 μg. However, indications for the 50-μg dose are rare, and the majority of formulations contain 35 μg of ethinyl estradiol. The reduced estrogen and progestin content in the second- and third-generation pills has decreased both side effects and risks associated with oral contraceptive use **(Table 347-6)**. At the currently used doses, patients must be cautioned not to miss pills due to the potential for ovulation. Side effects, including breakthrough bleeding, amenorrhea, breast tenderness, and weight gain, often respond to a change in formulation.

Table 347-6 Oral Contraceptives: Contraindications and Disease Risk

Contraindications
Absolute
Previous thromboembolic event or stroke
History of an estrogen-dependent tumor
Active liver disease

Pregnancy
Undiagnosed abnormal uterine bleeding
Hypertriglyceridemia
Women aged >35 years who smoke heavily
Relative
Hypertension
Women receiving anticonvulsant drug therapy
Women following bariatric surgery (malapsorptive procedure)
Disease Risks
Increased
Coronary heart disease—increased in smokers >35; no relation to progestin type
Hypertension—relative risk 1.8 (current users) and 1.2 (previous users)
Venous thrombosis—relative risk ~4; may be higher with third-generation progestin, drosperinone, and patch; compounded by obesity (tenfold increased risk compared with nonobese, no OCP); markedly increased with factor V Leiden or prothrombin-gene mutations (see Chap. 116)
Stroke—slight increase; unclear relation to migraine headache
Cerebral vein thrombosis—relative risk ~13–15; synergistic with prothrombin-gene mutation
Cervical cancer—relative risk 2–4 Breast cancer—may increase risk in carriers of BRCA1 and possible BRCA2
Decreased
Ovarian cancer—50% reduction in risk
Endometrial cancer—40% reduction in risk

Abbreviation: OCP, oral contraceptive pill.

The microdose progestin-only minipill is less effective as a contraceptive, having a pregnancy rate of 2–7 per 100 women-years. However, it may be appropriate for women with cardiovascular disease or for women who cannot tolerate synthetic estrogens.

NEW METHODS

A *weekly contraceptive patch* (Ortho Evra) is available and has similar efficacy to oral contraceptives but may be associated with less breakthrough bleeding. Approximately 2% of patches fail to adhere, and a similar percentage of women have skin reactions. Efficacy is lower in women weighing >90 kg. The amount of estrogen delivered may be comparable to that of a 40-μg ethinyl estradiol oral contraceptive, raising the possibility of increased risk of venous thromboembolism, which must be balanced against potential benefits for women not able to successfully use other methods. A *monthly contraceptive estrogen/progestin injection* (Lunelle) is highly effective, with a first-year failure rate of <0.2%, but it may be less effective in obese women. Its use is associated with bleeding irregularities that diminish over time. Fertility returns rapidly after discontinuation. A *monthly vaginal ring* (NuvaRing) that is intended to be left in place during intercourse is also available for contraceptive use. It is highly effective, with a 12-month failure rate of 0.7%. Ovulation returns within the first recovery cycle after discontinuation.

LONG-TERM CONTRACEPTIVES

Long-term progestin administration acts primarily by inhibiting ovulation and causing changes in the endometrium and cervical mucus that result in decreased implantation and sperm transport. Depot Medroxyprogesterone acetate (Depo Provera, DMPA), the only injectable form available in the United States, is effective for 3 months, but return of fertility after discontinuation may be delayed for up to 12–18 months. DMPA is now available for both SC and IM injection. Irregular bleeding, amenorrhea, and weight gain are the most common side effects. This form of contraception may be particularly good for women in whom an estrogen-containing contraceptive is contraindicated (e.g., migraine exacerbation, sickle-cell anemia, fibroids).

Postcoital Contraception

Postcoital contraceptive methods prevent implantation or cause regression of the corpus luteum and are highly efficacious if used appropriately. Unprotected intercourse without regard to the time of the month carries an 8% incidence of pregnancy, an incidence that can be reduced to 2% by the use of emergency contraceptives within 72 hours of unprotected intercourse. A notice published in 1997 by the U.S. Food and Drug Administration (FDA) indicated that certain oral contraceptive pills could be used within 72 hours of unprotected intercourse [Ovral (2 tablets, 12 hours apart) and Lo/Ovral (4 tablets, 12 hours apart)]. Preven (50 mg ethinyl estradiol and 0.25 mg levonorgestrel) and Plan B or Next Choice (0.75 mg levonorgestrel) are now approved for postcoital contraception and are available over the counter for women aged >17 years. Levonorgestrel is more effective and is associated with fewer side effects than the combination estrogen-progestin regimens. Ulipristal acetate is a progesterone antagonist that has been developed for emergency contraception. It is available in Europe and was approved by the FDA for prescription use in 2010. This medication is effective for up to 5 days after unprotected intercourse. Mifepristone is also a progesterone antagonist that is available for medical termination of intrauterine pregnancy but is not approved for emergency contraception in the United States. Insertion of a copper IUD within 5 days after unprotected intercourse is also a highly effective method.

Impact of Obesity on Contraceptive Choice

Approximately one-third of adults in the United States are obese. While obesity is associated with some reduction in fertility, the vast majority of obese women can conceive. The risk of pregnancy-associated complications is higher in obese women. Intrauterine contraception may be more effective than oral or transdermal methods for obese women. The WHO guidelines provide no restrictions (class 1) for the use of intrauterine contraception, DMPA, and progestin-only pills for obese women (BMI ≥ 30) in the absence of coexistent medical problems whereas methods that include estrogen (pill, patch, ring) are considered class 2 (advantages generally outweigh theoretical or proven risks) due to the increased risk of thromboembolic disease. There are no restrictions to the use of any contraceptive methods following restrictive bariatric surgery procedures, but both combined and progestin-only pills are relatively less effective following procedures associated with malabsorption.

FURTHER READINGS

Balasubramanian R et al: Human GnRH deficiency: A unique disease model to unravel the ontogeny of GnRH neurons. Neuroendocrinol 92:81, 2010[PMID: 20606386] [Full Text]

Boivin J et al: International estimates of infertility prevalence and treatment-seeking: Potential need and demand for infertility medical care. Hum Reprod 22:1506, 2007[PMID: 17376819] [Full Text]

Division of Reproductive Health et al: U S. medical eligibility criteria for contraceptive use, 2010: Adapted from the World Health

Organization medical eligibility criteria for contraceptive use, 4th edition. MMWR Recomm Rep 59:1, 2010

Edson MA et al: The mammalian ovary from genesis to revelation. Endocr Rev 30:624, 2010

Gajdos ZK et al: Genetic determinants of pubertal timing in the general population. Mol Cell Endocrinol 324:21, 2010[PMID: 20144687] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE MENOPAUSE TRANSITION AND POSTMENOPAUSAL HORMONE THERAPY: INTRODUCTION

Menopause is the permanent cessation of menstruation due to loss of ovarian follicular function. It is diagnosed retrospectively after 12 months of amenorrhea. The average age at menopause is 51 years among U.S. women. *Perimenopause* refers to the time period preceding menopause, when fertility wanes and menstrual cycle irregularity increases, until the first year after cessation of menses. The onset of perimenopause precedes the final menses by 2 to 8 years, with a mean duration of four years. Smoking accelerates the menopausal transition by 2 years.

Although the peri- and postmenopausal transitions share many symptoms, the physiology and clinical management differ. Low-dose oral contraceptives have become a therapeutic mainstay in perimenopause, whereas postmenopausal hormone therapy (HT) has been a common method of symptom alleviation after menstruation ceases.

PERIMENOPAUSE

Physiology

Ovarian mass and fertility decline sharply after age 35 and even more precipitously during perimenopause; depletion of primary follicles, a process that begins before birth, occurs steadily until menopause (Chap. 347). In perimenopause, intermenstrual intervals shorten significantly (typically by 3 days) due to an accelerated follicular phase. Follicle-stimulating hormone (FSH) levels rise due to altered folliculogenesis and reduced inhibin secretion. In contrast to the consistently high FSH and low estradiol levels seen in menopause, perimenopause is characterized by "irregularly irregular" hormone levels. The propensity for anovulatory cycles can produce a hyperestrogenic, hypoprogestagenic environment that may account for the increased incidence of endometrial hyperplasia or carcinoma, uterine polyps, and leiomyoma observed among women of perimenopausal age. Mean serum levels of selected ovarian and pituitary hormones during the menopausal transition are shown in **Fig. 348-1**. With transition into menopause, estradiol levels fall markedly, whereas estrone levels are relatively preserved, reflecting peripheral aromatization of adrenal and ovarian androgens. FSH levels increase more than those of luteinizing hormone (LH), presumably because of the loss of inhibin, as well as estrogen feedback.

Figure 348-1

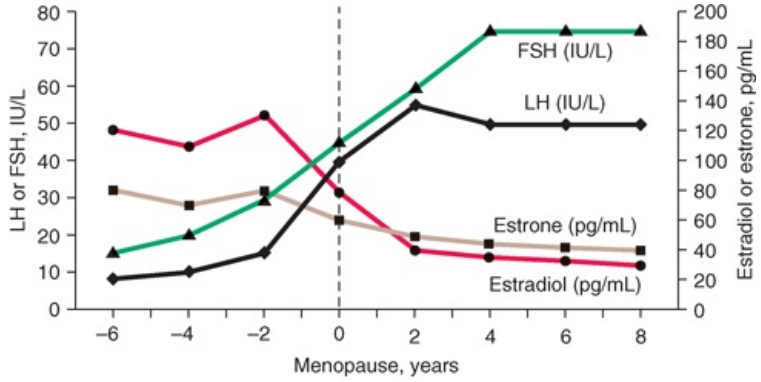

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mean serum levels of ovarian and pituitary hormones during the menopausal transition.

FSH, follicle-stimulating hormone; LH, luteinizing hormone. *(From JL Shifren, I Schiff: J Womens Health Gend Based Med 9 Suppl 1:S3, 2000, with permission.)*

Diagnostic Tests

Because of their extreme intraindividual variability, FSH and estradiol levels are imperfect diagnostic

indicators of perimenopause in menstruating women. However, a low FSH in the early follicular phase (days 2 through 5) of the menstrual cycle is inconsistent with a diagnosis of perimenopause. FSH measurement can also aid in assessing fertility; levels of <20 mIU/mL, 20 to <30 mIU/mL, and ≥30 mIU/mL measured on day 3 of the cycle indicate a good, fair, and poor likelihood of achieving pregnancy, respectively.

Symptoms

Determining whether symptoms that develop in midlife are due to ovarian senescence or to other age-related changes is difficult. There is strong evidence that the menopausal transition can cause hot flashes, night sweats, irregular bleeding, and vaginal dryness, and moderate evidence that it can cause sleep disturbances in some women. There is inconclusive or insufficient evidence that ovarian aging is a major cause of mood swings, depression, impaired memory or concentration, somatic symptoms, urinary incontinence, or sexual dysfunction. In one U.S. study, nearly 60% of women reported hot flashes in the 2 years before their final menses. Symptom intensity, duration, frequency, and effects on quality of life are highly variable.

Treatment: Perimenopause

For women with irregular or heavy menses or hormone-related symptoms that impair quality of life, low-dose combined oral contraceptives are a staple of therapy. Static doses of estrogen and progestin (e.g., 20 μg of ethinyl estradiol and 1 mg of norethindrone acetate daily for 21 days each month) can eliminate vasomotor symptoms and restore regular cyclicity. Oral contraceptives provide other benefits, including protection against ovarian and endometrial cancers and increased bone density, although it is not clear whether use during perimenopause decreases fracture risk later in life. Moreover, the contraceptive benefit is important, given that the unintentional pregnancy rate among women in their forties rivals that of adolescents. Contraindications to oral contraceptive use include cigarette smoking, liver disease, a history of thromboembolism or cardiovascular disease, breast cancer, or unexplained vaginal bleeding. Progestin-only formulations (e.g., 0.35 mg norethindrone daily) or medroxyprogesterone (Depo-Provera) injections (e.g., 150 mg IM every 3 months) may provide an alternative for the treatment of perimenopausal menorrhagia in women who smoke or have cardiovascular risk factors. Although progestins neither regularize cycles nor reduce the number of bleeding days, they reduce the volume of menstrual flow.

Nonhormonal strategies to reduce menstrual flow include use of nonsteroidal anti-inflammatory agents such as mefenamic acid (initial dose of 500 mg at start of menses, then 250 mg qid for 2–3 days) or, when medical approaches fail, endometrial ablation. It should be noted that menorrhagia requires an evaluation to rule out uterine disorders. Transvaginal ultrasound with saline enhancement is useful for detecting leiomyomata or polyps, and endometrial aspiration can identify hyperplastic changes.

TRANSITION TO MENOPAUSE

For sexually active women using contraceptive hormones to alleviate perimenopausal symptoms, the question of when and if to switch to HT must be individualized. Doses of estrogen and progestogen (either synthetic progestins or natural forms of progesterone) in HT are lower than those in oral contraceptives and have not been documented to prevent pregnancy. Although a 1-year absence of spontaneous menses reliably indicates ovulation cessation, it is not possible to assess the natural menstrual pattern while a woman is taking an oral contraceptive. Women willing to switch to a barrier method of contraception should do so; if menses occur spontaneously, oral contraceptive use can be resumed. The average age of final menses among relatives can serve as a guide for when to initiate this process, which can be repeated yearly until menopause has occurred.

MENOPAUSE AND POSTMENOPAUSAL HORMONE

Therapy

One of the most complex health care decisions facing women is whether to use postmenopausal HT. Once prescribed primarily to relieve vasomotor symptoms, HT has been promoted as a strategy to forestall various disorders that accelerate after menopause, including osteoporosis and cardiovascular disease. In 2000, nearly 40% of postmenopausal women age 50–74 in the United States had used HT. This widespread use occurred despite the paucity of conclusive data, until recently, on the health consequences of such therapy. Although many women rely on their health care providers for a definitive answer to the question of whether to use postmenopausal hormones, balancing the benefits and risks for an individual patient is challenging.

Although observational studies suggest that HT prevents cardiovascular and other chronic diseases, the apparent benefits may result at least in part from differences between women who opt to take postmenopausal hormones and women who do not. Those choosing HT tend to be healthier, have greater access to medical care, are more compliant with prescribed treatments, and maintain a more health-promoting lifestyle. Randomized trials, which eliminate these confounding factors, have not consistently confirmed the benefits found in observational studies. Indeed, the largest HT trial to date, the Women's Health Initiative (WHI), which examined more than 27,000 postmenopausal women age 50–79 (mean age, 63) for an average of 5–7 years, was stopped early because of an overall

unfavorable risk-benefit ratio in the estrogen-progestin arm and an excess risk of stroke that was not offset by a reduced risk of coronary heart disease (CHD) in the estrogen-only arm.

The following summary offers a decision-making guide based on a synthesis of currently available evidence. Prevention of cardiovascular disease is eliminated from the equation due to lack of evidence for such benefits in recent randomized clinical trials.

Benefits and Risks of Postmenopausal Hormone Therapy (Table 348-1)

Table 348-1 Benefits and Risks of Postmenopausal Hormone Therapy (Ht) in Primary Prevention Settings[a]

		Benefit or Risk		
		Relative		Absolute
Outcome	**Effect**	**Observational Studies**	**WHI[b], Except Where Noted**	**WHI[b], Except Where Noted**
Definite Benefits				
Symptoms of menopause	Definite improvement	↓ 70–80% decreased risk	↓65–90% decreased risk[c]	
Osteoporosis	Definite increase in bone mineral density and decrease in fracture risk	↓ 20–50% decreased risk for fracture	E+P: ↓ 33% decreased risk for hip fracture	E+P: 50 fewer hip fractures (110 vs 160) per 100,000 woman-years
			E: ↓ 39% decreased risk for hip fracture	E: 60 fewer hip fractures (110 vs 170) per 100,000 woman-years
Definite Risks				
Endometrial cancer	Definite increase in risk with estrogen alone; no increase in risk with estrogen-progestin	E+P: No increase in risk E: ↑ >300% increased risk (1–5 years); >600% increased risk (≥5 years)	E+P: No increase in risk E: Not applicable	E+P: No difference in risk E: 46 excess cases per 100,000 woman-years with unopposed estrogen (observational studies)[d]
Venous thromboembolism	Definite increase in risk	↑ 110% increased risk	E+P: ↑ 106% increased risk	E+P: 180 excess cases (350 vs 170) per 100,000 woman-years
			E: ↑ 32% increased risk	E: 80 excess cases (300 vs 220) per 100,000 woman-years
Breast cancer	Increase in risk with long-term use (≥5 years) of estrogen-progestin	E+P: ↑ 63% increased risk (≥5 years) E: ↑ 20% increased risk (≥5 years)	E+P: ↑ 24% increased risk E: No increase in risk	10–30 excess cases per 10,000 women using HT for 5 years; 30–90 excess cases per 10,000 women after 10 years use; 50–200 excess cases per 10,000 women after 15 years' use (estimate derived from observational data and WHI E+P findings)
Gallbladder disease	Definite increase in risk	↑ 110% increased risk	E+P: ↑ 67% increased risk	E+P: 180 excess cases (460 vs 280) per 100,000 woman-years
			E: ↑ 93% increased risk	E: 310 excess cases (650 vs 340) per 100,000 woman-years
Probable or Uncertain Risks and Benefits				
Coronary heart disease	Probable increase in risk among older	E+P: ↓36% decreased risk	E+P: ↑ 24%	E+P: 60 excess cases (390 vs 330) per 100,000 woman-

	women and women many years past menopause; possible decrease in risk or no effect in younger or recent menopausal women	E: ↓ 45% decreased risk	increased risk E: No increase or decrease in risk	years E: No difference in risk
Stroke	Probable increase in risk	↑ 12% increased risk	E+P: ↑ 31% increased risk	E+P: 70 excess cases (310 vs 240) per 100,000 woman-years
			E: ↑ 39% increased risk	E: 120 excess cases (440 vs 320) per 100,000 woman-years
Ovarian cancer	Probable increase in risk with long-term use (≥5 years)	E+P: No effect (<4 years use)	E+P: ↑ 58% increased risk[e]	E+P: 10 excess cases (40 vs 30) per 100,000 woman-years[e]
		E: ↑ 80% increased risk (≥10 years)	E: Not yet available	
Colorectal cancer	Probable decrease in risk with estrogen-progestin	↓ 34% decreased risk	E+P: ↓ 37% decreased risk	E+P: 70 fewer cases (90 vs 160) per 100,000 woman-years
			E: No increase or decrease in risk	E: No difference in risk
Diabetes mellitus	Probable decrease in risk	↓ 20% decreased risk	E+P: ↓ 21% decreased risk	E+P: 150 fewer cases (610 vs 760) per 100,000 woman-years
			E: ↓ 12% decreased risk[e]	E: 140 fewer cases (1160 vs 1300) per 100,000 woman-years[e]
Cognitive dysfunction	Unproven decrease in risk (inconsistent data from observational studies and randomized trials)	↓ 34% decreased risk	↑ 76% increased risk for dementia at age ≥65	120–230 excess cases of dementia per 100,000 woman-years

[a]E, estrogen alone; E+P, estrogen plus progestin. Most studies have assessed conjugated equine estrogen alone or in combination with medroxyprogesterone acetate.

[b]WHI, Women's Health Initiative. The estrogen–plus–progestin arm of the WHI assessed 5.6 years of conjugated equine estrogen (0.625 mg/d) plus medroxyprogesterone acetate (2.5 mg/d) versus placebo. The estrogen–alone arm of the WHI assessed 7.1 years of conjugated equine estrogen (0.625 mg/d) versus placebo.

[c]Data are from other randomized trials. The WHI was not designed to assess effect of HT on menopausal symptoms.

[d]JE Manson, KA Martin: N Engl J Med 345:34, 2001.

[e]Not statistically significant.

DEFINITE BENEFITS

Symptoms of Menopause

Compelling evidence, including data from randomized clinical trials, indicates that estrogen therapy is highly effective for controlling vasomotor and genitourinary symptoms. Alternative approaches, including the use of antidepressants (such as venlafaxine, 75–150 mg/d), gabapentin (300–900 mg/d), clonidine (0.1–0.2 mg/d), or vitamin E (400–800 IU/d), or the consumption of soy-based products or other phytoestrogens, may also alleviate vasomotor symptoms, although they are less effective than HT. For genitourinary symptoms, the efficacy of vaginal estrogen is similar to that of oral or transdermal estrogen.

Osteoporosis(See Also Chap. 354)

Bone Density

By reducing bone turnover and resorption rates, estrogen slows the aging-related bone loss experienced by most postmenopausal women. More than 50 randomized trials have demonstrated that postmenopausal estrogen therapy, with or without a progestogen, rapidly increases bone mineral density at the spine by 4–6% and at the hip by 2–3%, and maintains those increases during treatment.

Fractures

Data from observational studies indicate a 50–80% lower risk of vertebral fracture and a 25–30% lower risk of hip, wrist, and other peripheral fractures among current estrogen users; addition of a progestogen does not appear to modify this benefit. In the WHI, 5–7 years of either combined estrogen-progestin or estrogen-only therapy was associated with a 30–40% reduction in hip fracture and 20–30% fewer total fractures among a population unselected for osteoporosis. Bisphosphonates (such as alendronate, 10 mg/d or 70 mg once per week; risedronate, 5 mg/d or 35 mg once per week; or ibandronate, 2.5 mg/d or 150 mg once per month or 3 mg every 3 months IV) and raloxifene (60 mg/d), a selective estrogen receptor modulator (SERM), have been shown in randomized trials to increase bone mass density and decrease fracture rates. A recently available option for treatment of osteoporosis is parathyroid hormone (teriparatide, 20 μg/d SC). These agents, unlike estrogen, do not appear to have adverse effects on the endometrium or breast. Increased physical activity and adequate calcium (1000–1200 mg/d through diet or supplements in two to three divided doses) and vitamin D (600–1000 IU/d) intakes may also reduce the risk of osteoporosis-related fractures. (Blood levels of 25-hydroxyvitamin D of ≥75 nmol/L are optimal for bone-density maintenance and fracture prevention.) The Fracture Risk Assessment (FRAX®) score, an algorithm that combines an individual's bone-density score with age and other risk factors to predict her 10-year risk of hip and major osteoporotic fracture, may be of use in guiding decisions about pharmacologic treatment (see http://www.shef.ac.uk/FRAX/index.htm).

DEFINITE RISKS

Endometrial Cancer (with Estrogen Alone)

A combined analysis of 30 observational studies found a tripling of endometrial cancer risk among short-term (1–5 years) users of unopposed estrogen and a nearly tenfold increased risk among users for 10 or more years. These findings are supported by results from the randomized Postmenopausal Estrogen/Progestin Interventions (PEPI) trial, in which 24% of women assigned to unopposed estrogen for 3 years developed atypical endometrial hyperplasia, a premalignant lesion, compared with only 1% of women assigned to placebo. Use of a progestogen, which opposes the effects of estrogen on the endometrium, eliminates these risks.

Venous Thromboembolism

A meta-analysis of 12 studies—8 case-control, 1 cohort, and 3 randomized trials—found that current estrogen use was associated with a doubling of venous thromboembolism risk in postmenopausal women. Relative risks of thromboembolic events were even greater (2.7–5.1) in the three trials included in the meta-analysis. Results from the WHI indicate a twofold increase in risk of venous and pulmonary thromboembolism associated with estrogen-progestin and a one-third increase in this risk with estrogen-only therapy. Transdermal estrogen, taken alone or with certain progestogens (micronized progesterone or pregnane derivatives), appears to be a safer alternative with respect to thrombotic risk.

Breast Cancer (with Estrogen-Progestin)

An increased risk of breast cancer has been found among current or recent estrogen users in observational studies; this risk is directly related to duration of use. In a meta-analysis of 51 case-control and cohort studies, short-term use (<5 years) of postmenopausal HT did not appreciably elevate breast cancer incidence, whereas long-term use (≥5 years) was associated with a 35% increase in risk. In contrast to findings for endometrial cancer, combined estrogen-progestin regimens appear to increase breast cancer risk more than estrogen alone. Data from randomized trials also indicate that estrogen-progestin raises breast cancer risk. In the WHI, women assigned to receive combination hormones for an average of 5.6 years were 24% more likely to develop breast cancer than women assigned to placebo, but 7.1 years of estrogen-only therapy did not increase risk. Indeed, the WHI showed a trend toward a reduction in breast cancer risk with estrogen alone, although it is unclear whether this finding would pertain to formulations of estrogen other than conjugated equine estrogens or to treatment durations longer than 7 years. In the Heart and Estrogen/progestin Replacement Study (HERS), 4 years of combination therapy was associated with a 27% increase in breast cancer risk. Although the latter finding was not statistically significant, the totality of evidence strongly implicates estrogen-progestin therapy in breast carcinogenesis.

Gallbladder Disease

Large observational studies report a two- to threefold increased risk of gallstones or cholecystectomy

among postmenopausal women taking oral estrogen. In the WHI, women randomized to estrogen-progestin or estrogen alone had a 67% and 93% greater risk, respectively, of undergoing cholecystectomy than those assigned to placebo. Increased risks were also observed in HERS. Transdermal HT might be a safer alternative, but further research is needed.

PROBABLE OR UNCERTAIN RISKS AND BENEFITS

Coronary Heart Disease/Stroke

Until recently, HT had been enthusiastically recommended as a possible cardioprotective agent. In the past 3 decades, multiple observational studies suggested, in the aggregate, that estrogen use leads to a 35–50% reduction in CHD incidence among postmenopausal women. The biologic plausibility of such an association is supported by data from randomized trials demonstrating that exogenous estrogen lowers plasma low-density lipoprotein (LDL) cholesterol and raises high-density lipoprotein (HDL) cholesterol levels by 10–15%. Administration of estrogen also favorably affects lipoprotein(a) levels, LDL oxidation, endothelial vascular function, fibrinogen, and plasminogen activator inhibitor-1. However, estrogen therapy also has unfavorable effects on other biomarkers of cardiovascular risk: it boosts triglyceride levels; promotes coagulation via factor VII, prothrombin fragments 1 and 2, and fibrinopeptide A elevations; and raises levels of the inflammatory marker C-reactive protein.

Randomized trials of estrogen or combined estrogen-progestin in women with preexisting cardiovascular disease (CVD) have not confirmed the benefits reported in observational studies. In HERS, a secondary prevention trial designed to test the efficacy and safety of estrogen-progestin therapy on clinical cardiovascular outcomes, the 4-year incidence of coronary mortality and nonfatal myocardial infarction was similar in the active treatment and placebo groups, and a 50% increase in risk of coronary events was noted during the first year of the study among participants assigned to the active treatment group. Although it is possible that progestin may mitigate estrogen's benefits, the Estrogen Replacement and Atherosclerosis (ERA) trial indicated that angiographically determined progression of coronary atherosclerosis was unaffected by either opposed or unopposed estrogen treatment. Moreover, the Papworth Hormone Replacement Therapy Atherosclerosis Study, a trial of transdermal estradiol with and without norethindrone; the Women's Estrogen for Stroke Trial (WEST), a trial of oral 17β-estradiol; and the EStrogen in the Prevention of ReInfarction Trial (ESPRIT), a trial of oral estradiol valerate, found no cardiovascular benefits of the regimens studied. Thus, in clinical trials, HT has not proved effective for the secondary prevention of CVD in postmenopausal women.

Primary prevention trials also suggest an early increase in cardiovascular risk and absence of cardioprotection with postmenopausal HT. Results from the WHI suggest a deleterious cardiovascular effect of HT. Women assigned to 5.6 years of estrogen-progestin therapy were 24% more likely to develop CHD and 31% more likely to suffer a stroke than those assigned to placebo. In the estrogen-only arm of the WHI, a similar increase in stroke and no effect on CHD were observed.

However, a closer look at available data suggests that timing of initiation of HT may critically influence the association between such therapy and CHD. Estrogen may slow early stages of atherosclerosis but have adverse effects on advanced atherosclerotic lesions. It has been hypothesized that the prothrombotic and proinflammatory effects of estrogen manifest themselves predominantly among women with subclinical lesions who initiate HT well after the menopausal transition, whereas women with less arterial damage who start HT early in menopause may derive cardiovascular benefit because they have not yet developed advanced lesions. Nonhuman primate data support this concept. Conjugated estrogens had no effect on the extent of coronary artery plaque in cynomolgus monkeys assigned to estrogen alone or combined with progestin starting 2 years (approximately 6 human years) after oophorectomy and well after the establishment of atherosclerosis. However, administration of exogenous hormones immediately after oophorectomy, during the early stages of atherosclerosis, reduced the extent of plaque by 70%.

Lending further credence to this hypothesis are results of subgroup analyses of observational and clinical trial data. For example, among women who entered the WHI trial with a better cholesterol profile, estrogen with or without progestin led to a 40% lower risk for incident CHD. Among women who entered with a worse cholesterol profile, therapy resulted in a 73% higher risk (p for interaction=0.02). Moreover, although there was no association between estrogen-only therapy and CHD in the WHI trial cohort as a whole, such therapy was associated with a CHD risk reduction of 37% among participants age 50–59. By contrast, a risk reduction of only 8% was observed among those age 60–69, and a risk increase of 11% was found among those age 70–79. Due to the relatively small number of cases of myocardial infarction or coronary death (the primary definition of CHD in the WHI), especially in the younger women, these intra- and inter-age group differences were not statistically significant. However, when the definition of CHD was widened to include coronary bypass surgery or percutaneous coronary interventions, estrogen-only therapy was associated with a significant 45% reduction in CHD among women in the youngest age group. Moreover, estrogen was associated with lower levels of coronary artery calcified plaque.

Although age did not have a similar effect in the estrogen-progestin arm of the WHI, CHD risks steadily increased with years since menopause. Estrogen-progestin was associated with an 11% risk reduction for women less than 10 years beyond menopause but was associated with a 22% increase in risk for

women 10–19 years from menopause, and a 71% increase in risk for women 20 years or more from menopause (only the latter was statistically significant). In the large observational Nurses' Health Study, women who chose to start HT within 4 years of menopause experienced a lower risk of CHD than did nonusers, whereas those who began therapy 10 or more years after menopause appeared to receive little coronary benefit. Because observational studies include a high proportion of women who begin HT within 3–4 years of menopause and clinical trials include a high proportion of women 12 or more years past menopause, these findings help to reconcile some of the apparent discrepancies between the two types of studies.

Whether or not age at initiation of HT influences stroke risk is not well understood. In the WHI and the Nurses' Health Study, HT was associated with an excess risk of stroke in all age groups. Further research is needed on age, time since menopause, and other clinical characteristics as well as on biomarkers that predict increases or decreases in cardiovascular risk associated with exogenous HT. Furthermore, it remains uncertain whether different doses, formulations, or routes of administration of HT will produce different cardiovascular effects.

Colorectal Cancer

Observational studies have suggested that HT reduces risks of colon and rectal cancer, although the estimated magnitudes of the relative benefits ranged from 8 to 34% in various meta-analyses. In the WHI, the sole trial to examine the issue, estrogen-progestin was associated with a significant 44% reduction in colorectal cancer over a 5.6-year period, although no benefit was seen with 7 years of estrogen-only therapy.

Cognitive Decline and Dementia

A meta-analysis of ten case-control and two cohort studies suggested that postmenopausal HT is associated with a 34% decreased risk of dementia. Subsequent randomized trials, including the WHI, however, have failed to demonstrate any benefit of estrogen or estrogen-progestin therapy on the progression of mild to moderate Alzheimer's disease and/or have indicated a potential adverse effect of HT on the incidence of dementia, at least in women age 65 and older. Determining whether timing of initiation of HT influences cognitive outcomes will require further study.

Ovarian Cancer and Other Disorders

On the basis of limited observational and randomized data, it has been hypothesized that HT increases the risk of ovarian cancer and reduces the risk of type 2 diabetes mellitus. Results from the WHI support these hypotheses. The WHI also found that estrogen-progestin use was associated with increased lung cancer mortality.

Changes in Health Status after Discontinuation of Hormone Therapy

In the WHI cohort as a whole, the elevated risks for CHD, stroke, and venous thromboembolism associated with active use of estrogen-progestin disappeared within 2.4 years after discontinuation of therapy, as did the benefits, including amelioration of hot flashes and protection against osteoporotic fractures and colorectal cancer. A slightly elevated risk for breast cancer persisted, and a suggestion of higher risks for lung cancer, total cancer, and total mortality emerged. Postintervention results stratified by age or time since menopause onset are not yet available.

Approach to the Patient: Postmenopausal Hormone Therapy

The rational use of postmenopausal HT requires balancing the potential benefits and risks. **Figure 348-2** provides one approach to decision making. The clinician should first determine whether the patient has moderate to severe menopausal symptoms, the only indication for initiating systemic HT (urogenital symptoms in the absence of vasomotor symptoms may be treated with vaginal estrogen). The benefits and risks of such therapy should then be reviewed with the patient, giving more emphasis to absolute than to relative measures of effect, and pointing out uncertainties in clinical knowledge where relevant. Because chronic disease rates generally increase with age, absolute risks tend to be greater in older women, even when relative risks remain similar. Potential side effects—especially vaginal bleeding that may result from use of combined estrogen-progestogen formulations recommended for women with an intact uterus—should be noted. The patient's own preference regarding therapy should be elicited and factored into the decision. Contraindications to HT should be assessed routinely and include unexplained vaginal bleeding, active liver disease, venous thromboembolism, history of endometrial cancer (except stage 1 without deep invasion) or breast cancer, and history of CHD, stroke, transient ischemic attack, or diabetes. Relative contraindications include hypertriglyceridemia (>400 mg/dl) and active gallbladder disease; in such cases, transdermal estrogen may be an option. Primary prevention of heart disease should not be viewed as an expected benefit of HT, and an increase in stroke and a small early increase in coronary artery disease risk should be considered. Nevertheless, such therapy may be appropriate if the noncoronary benefits of treatment clearly outweigh risks. A woman who suffers an acute coronary event or stroke while on HT should stop therapy immediately.

Figure 348-2

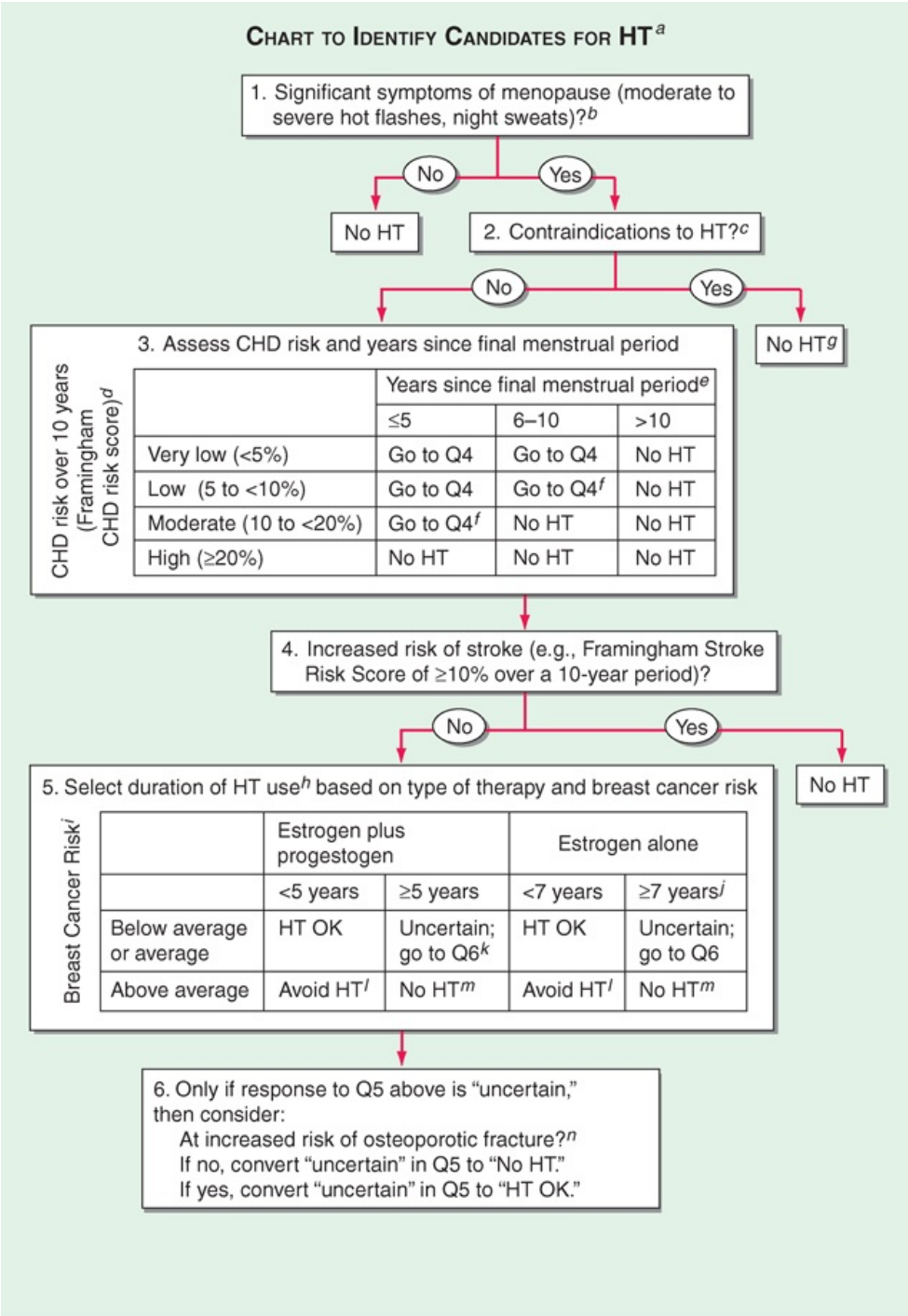

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chart for identifying appropriate candidates for postmenopausal hormone therapy (HT). CHD, coronary heart disease.

a Reassess each step at least once every 6–12 months (assuming patient's continued preference for HT).

b Women who have vaginal dryness without moderate to severe vasomotor symptoms may be candidates for vaginal estrogen.

c Traditional contraindications: unexplained vaginal bleeding; active liver disease; history of venous thromboembolism due to pregnancy, oral contraceptive use, or unknown etiology; blood clotting disorder; history of breast or endometrial cancer; history of CHD, stroke, transient ischemic attack, or diabetes. For other contraindications, including high triglycerides (>400 mg/dL); active gallbladder disease; and history of venous thromboembolism due to past immobility, surgery, or bone fracture; oral HT should be avoided but transdermal HT may be an option (see *f* below).

d 10-year risk of CHD, based on Framingham Coronary Heart Disease Risk Score (Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults: JAMA 285:2486, 2001), as

Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 106:3143, 2002[PMID: 12485966]), as modified by JE Manson, with SS Bassuk: *Hot Flashes, Hormones & Your Health* . New York, McGraw-Hill, 2007.

[e] Women >10 years past menopause are not good candidates for starting (first use of) HT.

[f] Avoid oral HT. Transdermal HT may be an option because it has a less adverse effect on clotting factors, triglyceride levels, and inflammation factors than oral HT.

[g] Consider selective serotonin or serotonin-norepinephrine reuptake inhibitor, gabapentin, clonidine, soy, or alternative.

[h] HT should be continued only if moderate to severe menopausal symptoms persist. The recommended cutpoints for duration are based on results of the Women's Health Initiative estrogen-progestin and estrogen-alone trials, which lasted 5.6 and 7.1 years, respectively. For longer durations of HT use, balance of benefits and risks is not known.

[i] Above-average risk of breast cancer: one or more first-degree relatives with breast cancer; susceptibility genes such as *BRCA1* or *BRCA2*; or a personal history of breast biopsy demonstrating atypia.

[j] Women with premature surgical menopause may take HT until average age at menopause (age 51 in the United States) and then follow flowchart for subsequent decision making.

[k] If progestogen is taken daily, avoid extending duration. If progestogen is cyclical or infrequent, avoid extending duration more than 1–2 years.

[l] If menopausal symptoms are severe, estrogen plus progestin can be taken for 2–3 years maximum and estrogen alone for 4–5 years maximum.

[m] If at high risk of osteoporotic fracture (see Q6), consider bisphosphonate, raloxifene, or alternative.

[n] Increased risk of osteoporotic fracture: documented osteopenia, personal or family history of nontraumatic fracture, current smoking, or weight <125 lbs.

Source: Adapted from JE Manson with SS Bassuk: *Hot Flashes, Hormones & Your Health* . New York, McGraw-Hill, 2007.

Short-term use (<5 years) of HT is appropriate for relief of menopausal symptoms among women without contraindications to such use. However, such therapy should be avoided among women with an elevated baseline risk of future cardiovascular events. Women who have contraindications, or are opposed to HT, may derive benefit from the use of certain antidepressants (including venlafaxine, fluoxetine, or paroxetine), gabapentin, clonidine, soy, or black cohosh, and, for genitourinary symptoms, intravaginal estrogen creams or devices.

Long-term use (≥5 years) of HT, especially estrogen-progestogen, is more problematic because a heightened risk of breast cancer must be factored into the decision. Reasonable candidates for such use include a small percentage of postmenopausal women and comprise those who have persistent severe vasomotor symptoms along with an increased risk of osteoporosis (e.g., those with osteopenia, a personal or family history of nontraumatic fracture, or a weight below 125 lbs), who also have no personal or family history of breast cancer in a first-degree relative or other contraindications, and who have a strong personal preference for therapy. Poor candidates are women with elevated cardiovascular risk, those at increased risk of breast cancer (e.g., women who have a first-degree relative with breast cancer, susceptibility genes such as *BRCA1* or *BRCA2*, or a personal history of cellular atypia detected by breast biopsy), and those at low risk of osteoporosis. Even in reasonable candidates, strategies to minimize dose and duration of use should be employed. For example, women using HT to relieve intense vasomotor symptoms in early postmenopause should consider discontinuing therapy before 5 years, resuming it only if such symptoms persist. Because of the role of progestogens in increasing breast cancer risk, regimens that employ cyclic rather than continuous progestogen exposure should be considered if treatment is extended. For prevention of osteoporosis, alternative therapies such as bisphosphonates or SERMs should be considered. Research on androgen-containing preparations has been limited, particularly in terms of long-term safety. Additional research on the effects of these agents on CVD, glucose tolerance, and breast cancer will be of particular interest.

In addition to HT, control of symptoms and prevention of chronic disease can be accomplished by lifestyle choices, including smoking abstention, adequate physical activity, and a healthy diet. An expanding array of pharmacologic options (e.g., bisphosphonates, SERMs, and other agents for osteoporosis, and cholesterol-lowering or antihypertensive agents for CVD should also reduce the widespread reliance on hormone use. However, short-term HT may still benefit some women.

FURTHER READINGS

CANONICO M et al: Hormone replacement therapy and risk of venous thromboembolism in postmenopausal women: systematic review and meta-analysis. BMJ 336:1227, 2008[PMID: 18495631] [Full Text]

CHLEBOWSKI RT et al: Oestrogen plus progestin and lung cancer in postmenopausal women (Women's Health Initiative trial): a post-hoc analysis of a randomised controlled trial. Lancet 374:1243, 2009[PMID: 19767090] [Full Text]

HEISS G et al: Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin. JAMA 299:1036, 2008[PMID: 18319414] [Full Text]

HSIA J et al: Conjugated equine estrogens and coronary heart disease. Arch Intern Med 166:357, 2006[PMID: 16476878] [Full Text]

MANSON JE, WITH BASSUK SS: *Hot Flashes, Hormones & Your Health* . New York: McGraw-Hill, 2007

———, et al: Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med 349:523, 2003

ROSSOUW JE et al: Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. JAMA 297:1465, 2007[PMID: 17405972] [Full Text]

SANTEN RJ et al: Postmenopausal hormone therapy: an Endocrine Society scientific statement. J Clin Endocrinol Metab 95(Suppl 1): S1, 2010

STEERING COMMITTEE FOR THE WOMEN'S HEALTH INITIATIVE: Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: Principal results from the Women's Health Initiative randomized controlled trial. JAMA 291:1701, 2004

UTIAN WH et al.: Estrogen and progestogen use in postmenopausal women: 2010 position statement of The North American Menopause Society. Menopause 17:242, 2010

Writing Group for the Women's Health Initiative Investigators: Risks and benefits of estrogen plus progestin in healthy postmenopausal women. Principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321, 2002

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF SEX DEVELOPMENT: INTRODUCTION

Sex development begins in utero but continues into young adulthood with the achievement of sexual maturity and reproductive capability. The major determinants of sex development can be divided into three major components: chromosomal sex, gonadal sex (sex determination), and phenotypic sex (sex differentiation) (Fig. 349-1). Abnormalities at each of these stages can result in disorders of sex development (DSDs) (Table 349-1). A child born with ambiguous genitalia requires urgent assessment, as some causes such as congenital adrenal hyperplasia (CAH), can be associated with life-threatening adrenal crises. Early gender assignment and clear communication with parents about the diagnosis and treatment options are essential. The involvement of an experienced multidisciplinary team is crucial for counseling, medical management, and surgical evaluation/intervention (if needed). Subtler forms of gonadal dysfunction [e.g., Klinefelter's syndrome (KS), Turner's syndrome (TS)] often are diagnosed later in life by internists. Because these conditions are associated with a variety of psychological, reproductive, and metabolic consequences, an open dialogue must be established between the patient and health care providers to ensure continuity and attention to these issues.

Figure 349-1

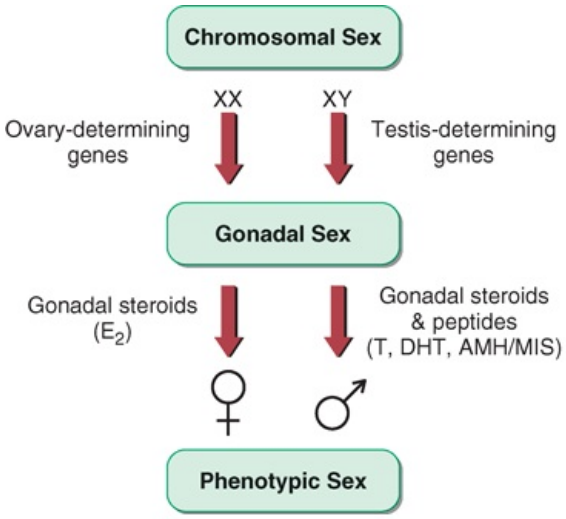

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sex development can be divided into three major components: chromosomal sex, gonadal sex, and phenotypic sex. T, testosterone; DHT, dihydrotestosterone; MIS, müllerian-inhibiting substance also known as anti-müllerian hormone, AMH.

Table 349-1 Classification of Disorders of Sex Development (Dsds)

Sex Chromosome DSD	46,XY DSD	46,XX DSD
47,XXY (Klinefelter's syndrome and variants) 45,X (Turner's syndrome and variants) 45,X/46,XY mosaicism (mixed gonadal dysgenesis) 46,XX/46,XY (chimerism/mosaicism)	**Disorders of gonadal (testis) development** Complete or partial gonadal dysgenesis (e.g., *SRY, SOX9, SF1, WT1, DHH)* Impaired fetal Leydig cell function (e.g., *SF1/NR5A1, CXorf6/MAMLD1*) Ovotesticular DSD Testis regression	**Disorders of gonadal (ovary) development** Gonadal dysgenesis Ovotesticular DSD Testicular DSD (e.g., *SRY*+, dup *SOX9, RSPO1*)
	Disorders in androgen synthesis or action (see Table 349-3)	**Androgen excess (see Table 349-4)**

	Disorders of androgen biosynthesis LH receptor (*LHCGR*) mutations Smith-Lemli-Opitz syndrome Steroidogenic acute regulatory (*STAR*) protein Cholesterol side-chain cleavage (*CYP11A1*) 3β-Hydroxysteroid dehydrogenase II (*HSD3b2*) 17α-Hydroxylase/17,20-lyase (*CYP17A1*) P450 oxidoreductase (*POR*) 17β-Hydroxysteroid dehydrogenase III (*HSD17b3*) 5α-Reductase II (*SRD5A2*) Disorders of androgen action Androgen Insensitivity syndrome Drugs and environmental modulators	Fetal 3β-Hydroxysteroid dehydrogenase II (*HSD3b2?*) 21-Hydroxylase (*CYP21A2?*) P450 oxidoreductase (*POR?*) 11β-Hydroxylase (*CYP11B1?*) Glucocorticoid receptor mutations Fetoplacental Aromatase deficiency (*CYP19*) Oxidoreductase deficiency (*POR*) Maternal Maternal virilizing tumors (e.g., luteomas) Androgenic drugs
	Other Syndromic associations of male genital development Persistent müllerian duct syndrome Vanishing testis syndrome Isolated hypospadias Congenital hypogonadotropic hypogonadism Cryptorchidism Environmental influences	**Other** Syndromic associations (e.g., cloacal anomalies) Müllerian agenesis/hypoplasia (e.g., MRKH) Uterine abnormalities (e.g., MODY5) Vaginal atresia (e.g., McKusick-Kaufman) Labial adhesions

Source: Modified from IA Hughes. Arch Dis Child 91:554, 2006.

SEX DEVELOPMENT

Chromosomal sex describes the X and/or Y chromosome complement (46,XY male; 46,XX female) that is established at the time of fertilization. The presence of a normal Y chromosome determines that testis development will occur even in the presence of multiple X chromosomes (e.g., 47,XXY or 48,XXXY). The loss of an X chromosome impairs gonad development (45,X or 45,X/46,XY mosaicism). Fetuses with no X chromosome (45,Y) are not viable.

Gonadal sex refers to the assignment of gonadal tissue as testis or ovary. The embryonic gonad is bipotential and can develop (from ~42 days gestation) into either a testis or an ovary, depending on which genes are expressed (Fig. 349-2). Testis development is initiated by expression of the Y chromosome gene *SRY* (sex- determining region on the Y chromosome) that encodes an HMG box transcription factor. *SRY* is expressed transiently in cells destined to become Sertoli cells and serves as a pivotal switch to establish the testis lineage. Mutation of *SRY* prevents testis development in chromosomal 46,XY males, whereas translocation of *SRY* in 46,XX females is sufficient to induce testis development and a male phenotype. Other genes are necessary to continue testis development. *SOX9* (*SRY*-related HMG-box gene 9) is upregulated by *SRY* in the developing male gonad but is suppressed in the female gonad. Transgenic expression of *SOX9* is sufficient to initiate testis formation in mice, and mutations that disrupt *SOX9* impair testis development. *WT1* (Wilms' tumor–related gene 1) acts early in the genetic pathway and regulates the transcription of several genes, including *SFI* (officially called *NRSA1*), *DAX1* (NR0B1), and *AMH* (encoding *MIS*, müllerian-inhibiting substance). *SF1* encodes steroidogenic factor 1, a nuclear receptor that functions in cooperation with other transcription factors to regulate a large array of adrenal and gonadal genes, including *SOX9* and many genes involved in steroidogenesis. Heterozygous *SF1* mutations account for ~10% of XY patients with gonadal dysgenesis and impaired androgenization, indicating the sensitivity of the testis to *SF1* gene dosage. The early expression pattern of *SF1* in the gonad parallels that of another orphan nuclear

receptor, *DAX1* (dosage sensitive sex-reversal, adrenal hypoplasia congenita on the X chromosome, gene 1). Duplication of *DAX1* impairs testis development, whereas deletions or mutations of *DAX1* lead to disordered formation of testis cords, again revealing the exquisite sensitivity of the male sex-determining pathway to gene dosage effects. In addition to the genes mentioned above, studies of human and murine mutations indicate that at least 15 other genes are also involved in gonadal differentiation, gonadal development, and final positioning of the gonad (Fig. 349-2). These genes encode an array of signaling molecules and paracrine growth factors in addition to transcription factors.

Figure 349-2

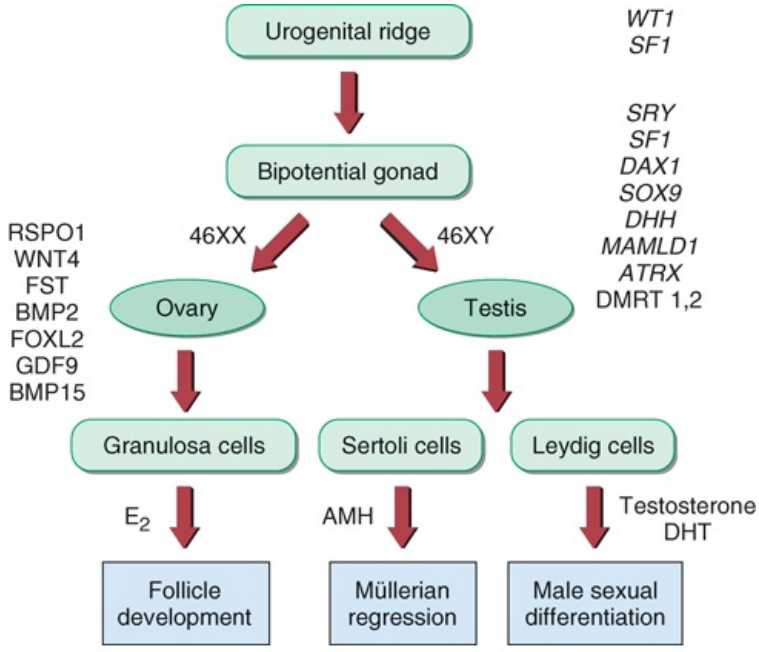

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The genetic regulation of gonadal development. *WT1*, Wilms' tumor–related gene 1; *SF1*, steroidogenic factor 1 (also known as NR5A1); *SRY*, sex-determining region on the Y chromosome; *SOX9*, SRY-related HMG-box gene 9; *DHH*, desert hedgehog; *ATRX*, (α-thalassemia, mental retardation on the X); *DAX1*, dosage sensitive sex-reversal, adrenal hypoplasia congenita on the X chromosome, gene 1; DMRT 1,2, doublesex MAB3-related transcription factor 1,2; WNT4, wingless-type MMTV integration site 4; FST, follistatin; BMP2 and 15, bone morphogenic factors 2 and 15; FOXL2, forkhead transcription factor L2; GDF9, growth differentiation factor 9; AMH, anti-müllerian hormone (müllerian-inhibiting substance); DHT, dihydrotestosterone; RSPO1, R-spondin 1; MAMLD1, mastermind-like domain containing 1.

Although ovarian development once was considered a "default" process, it is now clear that specific genes are expressed during the earliest stages of ovary development. Some of these factors may repress testis development (e.g., WNT4, R-spondin-1) (Fig. 349-2). Once the ovary has formed, additional genes are required for normal follicular development [e.g., follicle-stimulating hormone (FSH) receptor, *GDF9*]. Steroidogenesis in the ovary requires the development of follicles that contain granulosa cells and theca cells surrounding the oocytes (Chap. 347). Thus, there is relatively limited ovarian steroidogenesis until gonadotropins are produced at puberty.

Germ cells also develop in a sex dimorphic manner. In the developing ovary, primordial germ cells (PGCs) proliferate and enter meiosis, whereas they proliferate and then undergo mitotic arrest in the developing testis. PGC entry into meiosis is initiated by retinoic acid that activates *STRA8* (*stimulated by retinoic acid 8*) and other genes involved in meiosis. The developing testis produces high levels of CYP26B1, an enzyme that degrades retinoic acid, preventing PGC entry into meiosis. Approximately 7 million germ cells are present in the fetal ovary in the second trimester, and 1 million remain at birth. Only 400 are ovulated during a woman's reproductive life span (Chap. 347).

Phenotypic sex refers to the structures of the external and internal genitalia and secondary sex characteristics. The male phenotype requires the secretion of anti-müllerian hormone (AMH, also known as müllerian-inhibiting substance, MIS) from Sertoli cells and testosterone from testicular Leydig cells. AMH is a member of the transforming growth factor (TGF) β family and acts through specific receptors to cause regression of the müllerian structures (from 60–80 days' gestation). At ~60–140 days' gestation, testosterone supports the development of wolffian structures, including the

epididymides, vasa deferentia, and seminal vesicles. Testosterone is the precursor for dihydrotestosterone (DHT), a potent androgen that promotes development of the external genitalia, including the penis and scrotum (65–100 days, and thereafter) (Fig. 349-3). The urogenital sinus develops into the prostate and prostatic urethra in the male and into the urethra and lower portion of the vagina in the female. The genital tubercle becomes the glans penis in the male and the clitoris in the female. The urogenital swellings form the scrotum or the labia majora, and the urethral folds fuse to form the shaft of the penis and the male urethra or the labia minora. In the female, wolffian ducts regress and the müllerian ducts form the fallopian tubes, uterus, and upper segment of the vagina. A female phenotype will develop in the absence of the gonad, but estrogen is needed for maturation of the uterus and breast at puberty.

Figure 349-3

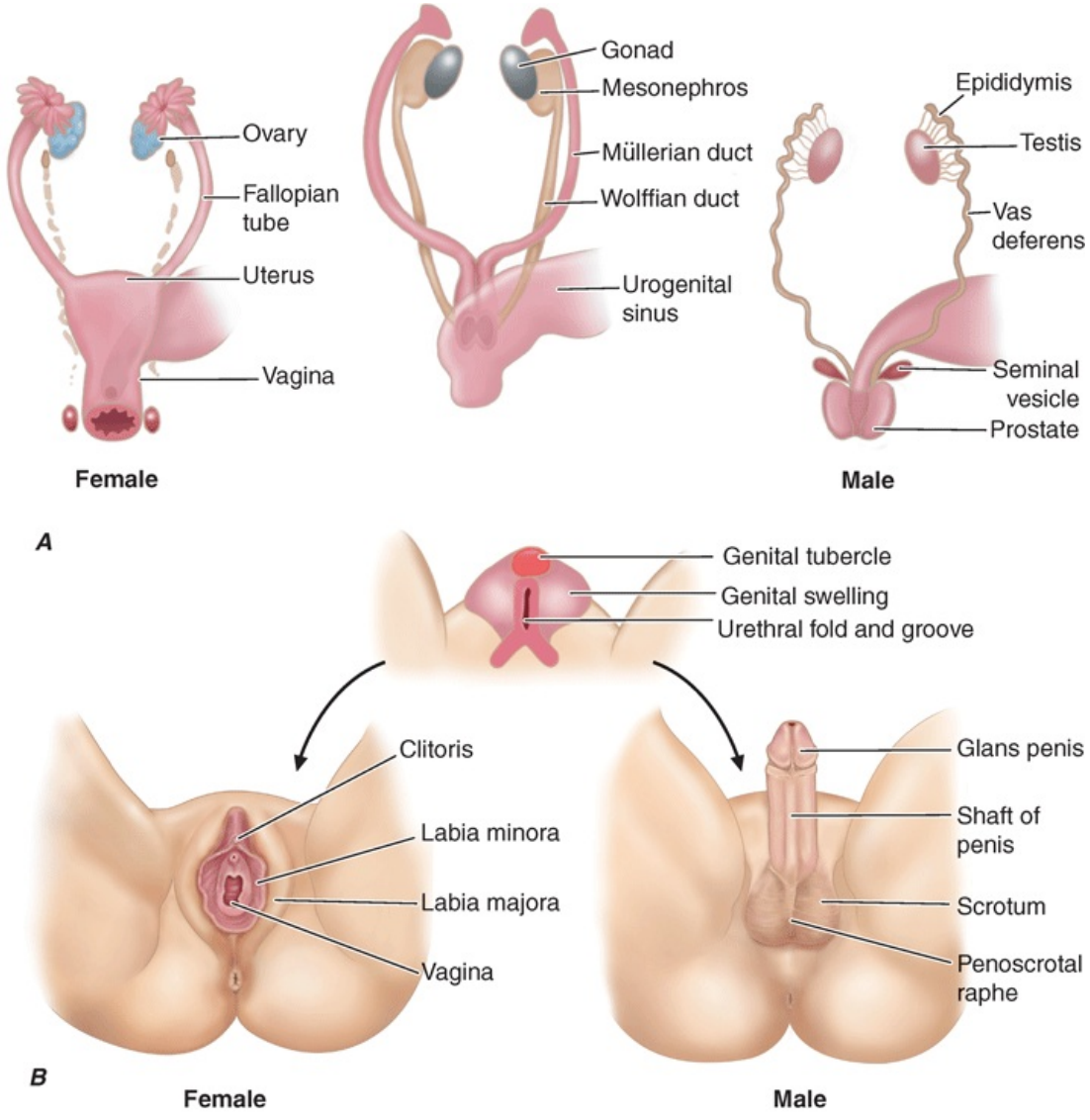

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sex development.*A.* Internal urogenital tract. ***B.*** External genitalia. *[After E Braunwald et al (eds): Harrison's Principles of Internal Medicine, 15th ed. New York, McGraw-Hill, 2001.]*

DISORDERS OF CHROMOSOMAL SEX

Variations in sex chromosome number and structure can present as disorders of sex development (e.g., 45,X/46,XY). KS (47,XXY) and TS (45,X) do not usually present with genital ambiguity but are associated with gonadal dysfunction (Table 349-2).

Table 349-2 Clinical Features of Chromosomal Disorders of Sex Development (DSD)

			Genitalia

Disorder	Common Chromosomal Complement	Gonad	External	Internal	Breast Development
Klinefelter's syndrome	47,XXY or 46,XY/47,XXY	Hyalinized testes	Male	Male	Gynecomastia
	Clinical Features				
	Small testes, azoospermia, decreased facial and axillary hair, decreased libido, tall stature and increased leg length, decreased penile length, increased risk of breast tumors, thromboembolic disease, learning difficulties, obesity, diabetes mellitus, varicose veins				
Turner's syndrome	45,X or 45,X/46,XX	Streak gonad or immature ovary	Female	Hypoplastic female	Immature female
	Clinical Features				
	Infancy: lymphedema, web neck, shield chest, low-set hairline, cardiac defects and coarctation of the aorta, urinary tract malformations and horseshoe kidney				
	Childhood: short stature, cubitus valgus, short neck, short 4th metacarpals, hypoplastic nails, micrognathia, scoliosis, otitis media and sensorineural hearing loss, ptosis and amblyopia, multiple nevi and keloid formation, autoimmune thyroid disease, visuospatial learning difficulties				
	Adulthood: pubertal failure and primary amenorrhea, hypertension, obesity, dyslipidemia, impaired glucose tolerance and insulin resistance, autoimmune thyroid disease, cardiovascular disease, aortic root dilation, osteoporosis, inflammatory bowel disease, chronic hepatic dysfunction, increased risk of colon cancer, hearing loss				
Mixed gonadal dysgenesis	45,X/46,XY	Testis or streak gonad	Variable	Variable	Usually male
	Clinical Features				
	Short stature, increased risk of gonadal tumors, some Turner's syndrome features				
Ovotesticular DSD (true hermaphroditism)	46,XX/46,XY	Testis and ovary or ovotestis	Variable	Variable	Gynecomastia
	Clinical Features				
	Possible increased risk of gonadal tumors				

Klinefelter's Syndrome (47,XXY)

PATHOPHYSIOLOGY

The classic form of KS (47,XXY) occurs after meiotic nondisjunction of the sex chromosomes during gametogenesis (40% during spermatogenesis, 60% during oogenesis) (Chap. 62). Mosaic forms of KS (46,XY/47,XXY) are thought to result from chromosomal mitotic nondisjunction within the zygote and occur in at least 10% of individuals with this condition. Other chromosomal variants of KS (e.g., 48,XXYY, 48,XXXY) have been reported but are less common.

CLINICAL FEATURES

KS is characterized by small testes, infertility, gynecomastia, "eunuchoid" proportions, and incomplete virilization in phenotypic males. It has an incidence of at least 1 in 1000 men, but approximately 75% of cases are not diagnosed. In severe cases, individuals present prepubertally with small testes or with impaired androgenization and gynecomastia at the time of puberty. Developmental delay and learning disabilities may be a feature. Later in life, eunuchoid features or infertility lead to the diagnosis. Testes are small and firm [median length 2.5 cm (4 mL volume); almost always <3.5 cm (12 mL)] and typically seem inappropriately small for the degree of androgenization. Biopsies are not usually necessary but reveal seminiferous tubule hyalinization and azoospermia. Other clinical features of KS are listed in Table 349-2. Plasma concentrations of FSH and luteinizing hormone (LH) are increased in most patients with 47,XXY (90% and 80%, respectively), and plasma testosterone is decreased (50–75%), reflecting primary gonadal failure. Estradiol is often increased because of chronic Leydig cell stimulation by LH and because of aromatization of androstenedione by adipose tissue; the increased ratio of estradiol/testosterone results in gynecomastia. Patients with mosaic forms of KS have less severe clinical features, have larger testes, and sometimes achieve spontaneous fertility.

Treatment: Klinefelter's Syndrome

Gynecomastia should be treated by surgical reduction if it causes concern (Chap. 346). Androgen supplementation improves virilization, libido, energy, hypofibrinolysis, and bone mineralization in underandrogenized men but may occasionally worsen gynecomastia (Chap. 346). Fertility has been achieved by using in vitro fertilization in men with oligospermia or with intracytoplasmic sperm injection (ICSI) after retrieval of spermatozoa by testicular sperm extraction techniques. In specialized centers,

successful spermatozoa retrieval using this technique is possible in >50% of men with nonmosaic KS. After ICSI and embryo transfer, successful pregnancies can be achieved in ~50% of these cases. The risk of transmission of this chromosomal abnormality needs to be considered, and preimplantation screening may be desired, although this outcome is much less common than originally predicted.

Turner's Syndrome (Gonadal Dysgenesis; 45,X)

PATHOPHYSIOLOGY

Approximately one-half of individuals with Turner's syndrome have a 45,X karyotype, about 20% have 45,*X*/46,*XX* mosaicism, and the remainder have structural abnormalities of the X chromosome such as X fragments, isochromosomes, or rings. The clinical features of TS result from haploinsufficiency of multiple X chromosomal genes (e.g., short stature homeobox, *SHOX*). However, imprinted genes also may be affected when the inherited X has different parental origins.

CLINICAL FEATURES

TS is characterized by bilateral streak gonads, primary amenorrhea, short stature, and multiple congenital anomalies in phenotypic females. It affects ~1 in 2500 women and is diagnosed at different ages depending on the dominant clinical features (Table 349-2). Prenatally, a diagnosis of TS usually is made incidentally after chorionic villus sampling or amniocentesis for unrelated reasons such as advanced maternal age. Prenatal ultrasound findings include increased nuchal translucency. The postnatal diagnosis of TS should be considered in female neonates or infants with lymphedema, nuchal folds, low hairline, or left-sided cardiac defects and in girls with unexplained growth failure or pubertal delay. Although limited spontaneous pubertal development occurs in up to 30% of girls with TS (10%, 45,X; 30–40%, 45,X/46,XX) and ~2% reach menarche, the vast majority of women with TS develop complete ovarian failure. This diagnosis should be considered, therefore, in all women who present with primary or secondary amenorrhea and elevated gonadotropin levels.

Treatment: Turner's Syndrome

The management of girls and women with TS requires a multidisciplinary approach because of the number of potentially involved organ systems. Detailed cardiac and renal evaluation should be performed at the time of diagnosis. Individuals with congenital heart defects (CHDs) (30%) (bicuspid aortic valve, 30–50%; coarctation of the aorta, 30%; aortic root dilation, 5%) require long-term follow-up by an experienced cardiologist, antibiotic prophylaxis for dental or surgical procedures, and serial imaging of aortic root dimensions, as progressive aortic root dilation is associated with increased risk of aortic dissection. Individuals found to have congenital renal and urinary tract malformations (30%) are at risk for urinary tract infections, hypertension, and nephrocalcinosis. Hypertension can occur independently of cardiac and renal malformations and should be monitored and treated as in other patients with essential hypertension. Clitoral enlargement or other evidence of virilization suggests the presence of covert, translocated Y chromosomal material and is associated with increased risk of gonadoblastoma, apparently as a consequence of Y chromosomal genes distinct from *SRY*. Regular assessment of thyroid function, weight, dentition, hearing, speech, vision, and educational issues should be performed during childhood. Otitis media and middle-ear disease are prevalent in childhood (50–85%), and sensorineural hearing loss becomes progressively common with age (70–90%). Autoimmune hypothyroidism (15–30%) can occur in childhood but has a mean age of onset in the third decade. Counseling about long-term growth and fertility issues should be provided. Patient support groups are active throughout the world and can play an invaluable role.

The treatment of short stature in children with TS remains a challenge, as untreated final height rarely exceeds 150 cm in nonmosaic 45,X TS. High-dose recombinant growth hormone stimulates growth rate in children with TS and may be used alone or in combination with low doses of the nonaromatizable anabolic steroid oxandrolone (up to 0.05 mg/kg per d) in an older child (>9 years). However, final height increments are often modest (5–10 cm), and individualization of treatment response to regimens may be beneficial. Girls with evidence of gonadal failure require estrogen replacement to induce breast and uterine development, support growth, and maintain bone mineralization. Most physicians now choose to initiate low-dose estrogen therapy (one-tenth to one-eighth of the adult replacement dose) to induce puberty at an age-appropriate time (~12 years). Doses of estrogen are increased gradually to allow feminization over a 2–4 year period. Progestins are added later to regulate withdrawal bleeds, and some women with TS have achieved successful pregnancy after ovum donation and in vitro fertilization. Long-term follow-up of women with TS involves careful surveillance of sex hormone replacement and reproductive function, bone mineralization, cardiac function and aortic root dimensions, blood pressure, weight and glucose tolerance, hepatic and lipid profiles, thyroid function, and hearing. This service is provided by a dedicated TS clinic in some centers.

Mixed Gonadal Dysgenesis (45,X/46,Xy)

Mixed gonadal dysgenesis typically results from 45,*X*/46,XY mosaicism. The phenotype of patients with this condition varies considerably. Although some patients have a predominantly female phenotype with somatic features of TS, streak gonads, and müllerian structures, most 45,*X*/46,*XY* individuals have a male phenotype and testes, and the diagnosis is made incidentally after

amniocentesis or during investigation of infertility. In practice, most children referred for assessment have ambiguous genitalia and variable somatic features. A female sex-of-rearing is often assigned (60%) if uterine structures are present, gonads are intraabdominal, and phallic development is poor. In such situations, gonadectomy usually is undertaken to prevent further androgen secretion and prevent development of gonadoblastoma (up to 25%). Individuals raised as males require reconstructive surgery for hypospadias and removal of dysgenetic gonads if the gonads cannot be brought down into the scrotum. Scrotal testes can be preserved but require regular examination for tumor development. Biopsy for carcinoma in situ is recommended in adolescence, and testosterone supplementation may be required to support androgenization in puberty. Height potential is usually attenuated.

Ovotesticular Dsd

Ovotesticular DSD (formerly called *true hermaphroditism*) occurs when both an ovary and a testis—or when an ovotestis—are found in one individual. For unclear reasons, gonadal asymmetry most often occurs with a testis on the right and an ovary on the left. Most individuals with this diagnosis have a 46,XX karyotype, especially in sub-Saharan Africa. A 46,XX/46,XY chimeric karyotype is less common and has a variable phenotype.

DISORDERS OF GONADAL AND PHENOTYPIC SEX

The clinical features of patients with disorders of gonadal and phenotypic sex are divided into the underandrogenization of 46,XY males (46,XY DSD) and the excess androgenization of 46,XX females (46,XX DSD) (Table 349-1). These disorders cover a spectrum of phenotypes ranging from "46,XY phenotypic females" or "46,XX males" to individuals with ambiguous genitalia.

46,Xy Dsd (Underandrogenized Males)

Underandrogenization of the 46,XY fetus (formerly called *male pseudohermaphroditism*) reflects defects in androgen production or action. It can result from disorders of testis development, defects of androgen synthesis, or resistance to testosterone and DHT (Table 349-1).

DISORDERS OF TESTIS DEVELOPMENT

Testicular Dysgenesis

Patients with *pure* (or *complete*) *gonadal dysgenesis (Swyer syndrome)* have streak gonads, müllerian structures (due to insufficient AMH/MIS secretion), and a complete absence of androgenization. Serum AMH/MIS is low, and testosterone response to human chorionic gonadotropin (hCG) stimulation is impaired. Patients with *partial gonadal dysgenesis (dysgenetic testes)* may produce enough MIS to regress the uterus and, sometimes, sufficient testosterone for partial androgenization. Gonadal dysgenesis can result from mutations or deletions of testis-promoting genes (*WT1*, *SF1*, *SRY*, *SOX9*, *DHH*, *ATRX*, *ARX*, *DMRT*) or duplication of chromosomal loci containing "antitestis" genes (e.g., *WNT4/RSPO1*, *DAX1*) (Table 349-3). Among these, deletions or mutations of *SRY* and heterozygous mutations of *SF1(NR5A1)* appear to be most common but still account collectively for <25% of cases. Associated clinical features may be present, reflecting additional functional roles for these genes. For example, renal dysfunction occurs in patients with specific *WT1* mutations (Denys-Drash and Fraser's syndromes), primary adrenal failure occurs in some patients with *SF1* mutations, and severe cartilage abnormalities (campomelic dysplasia) are the predominant clinical feature of *SOX9* mutations. A family history of DSD or premature ovarian insufficiency is important (e.g., *SF1/NR5A1*). Intraabdominal dysgenetic testes should be removed to prevent malignancy, and estrogens can be used to induce secondary sex characteristics in 46,XY individuals raised as females. *Absent (vanishing) testis syndrome* (*bilateral anorchia*) reflects regression of the testis during development. The etiology is unknown, but the absence of müllerian structures indicates adequate secretion of AMH in utero. Early testicular regression causes impaired androgenization in utero, and in most cases, androgenization of the external genitalia is either normal or slightly impaired (e.g., small penis, hypospadias). These individuals can be offered testicular prostheses and should receive androgen replacement in adolescence.

Table 349-3 Selected Genetic Causes of Underandrogenization of Karyotypic Males (46,Xy Dsd)

Gene	Inheritance	Gonad	Uterus	External Genitalia	Associated Features
Disorders of Testis Development					
WT1	AD	Dysgenetic testis	+/–	Female or ambiguous	Wilms' tumor, renal abnormalities, gonadal tumors (WAGR, Denys-Drash and Fraser's syndromes)
CBX2	AD	Ovary	+	Female	
SF1	AR/AD	Dysgenetic testis/Leydig's	+/–	Female or ambiguous	Primary adrenal failure; primary ovarian insufficiency in female

		dysfunction			(46,XX) relatives
SRY	Y	Dysgenetic testis or ovotestis	+/–	Female or ambiguous	
SOX9	AD	Dysgenetic testis or ovotestis	+/–	Female or ambiguous	Campomelic dysplasia
DHH	AR	Dysgenetic testis	+	Female	Minifascicular neuropathy
ATRX	X	Dysgenetic testis	–	Female or ambiguous	α Thalassemia, developmental delay
ARX	X	Dysgenetic testis	–	Male or ambiguous	Developmental delay; X-linked lissencephaly
MAMLD1	X	Dysgenetic testis/Leydig's dysfunction	–	Hypospadias	
DAX1	dupXp21	Dysgenetic testis	+/–	Female or ambiguous	
WNT4/RSPO1	dup1p35	Dysgenetic testis	+	Ambiguous	
Disorders of Androgen Synthesis					
LHR	AR	Testis	–	Female, ambiguous or micropenis	Leydig's cell hypoplasia
DHCR7	AR	Testis	–	Variable	Smith-Lemli-Opitz syndrome: coarse facies, second-third toe syndactyly, failure to thrive, developmental delay, cardiac and visceral abnormalities
StAR	AR	Testis	–	Female or ambiguous	Congenital lipoid adrenal hyperplasia (primary adrenal failure)
CYP11A1	AR	Testis	–	Ambiguous	Primary adrenal failure
HSD3b2	AR	Testis	–	Ambiguous	CAH, primary adrenal failure ± salt loss, partial androgenization due to ↑ DHEA
CYP17	AR	Testis	–	Female or ambiguous	CAH, hypertension due to ↑ corticosterone and 11-deoxycorticosterone, except in isolated 17,20 lyase deficiency
POR	AR	Testis	–	Ambiguous or male	Mixed features of 21-hydroxylase deficiency and 17α-hydroxylase/17,20 lyase deficiency, sometimes associated with Antley-Bixler craniosynostosis
HSD17b3	AR	Testis	–	Female or ambiguous	Partial androgenization at puberty, ↑ androstenedione: testosterone ratio
SRD5A2	AR	Testis	–	Ambiguous or micropenis	Partial androgenization at puberty, ↑ testosterone: dihydrotestosterone ratio
Disorders of Androgen Action					
Androgen receptor	X	Testis	–	Female, ambiguous, micropenis or normal male	Phenotypic spectrum from complete androgen insensitivity syndrome (female external genitalia) and partial androgen insensitivity (ambiguous) to normal male genitalia and infertility

Abbreviations: AR, autosomal recessive; AD, autosomal dominant; CAH, congenital adrenal hyperplasia; *WT1,* Wilms' tumor–related gene 1; WAGR, Wilms' tumor, aniridia, genitourinary anomalies, and mental retardation; *SF1,* steroidogenic factor 1; *SRY,* sex-related gene on the Y chromosome; *SOX9*,*SRY*-related HMG-box gene 9; *DHH*, desert hedgehog; *ATRX*, (α-thalassemia,

mental retardation on the X); *ARX*, aristaless related homeobox, X-linked; *DAX1*, dosage sensitive sex-reversal, adrenal hypoplasia congenita on the X chromosome, gene 1; *WNT4*, wingless-type mouse mammary tumor virus integration site, 4; *LHR*, LH receptor; *DHCR7*, sterol 7 δ reductase; *StAR*, steroidogenic acute regulatory protein; *CYP11A1*, P450 cholesterol side-chain cleavage; *HSD3b2*, 3β-hydroxysteroid dehydrogenase type 2; *CYP17*, 17α-hydroxylase and 17,20-lyase; *POR*, P450 oxidoreductase; *HSD17b3,* 17β-hydroxysteroid dehydrogenase type 3; *SRD5A2*, 5α-reductase type 2.

DISORDERS OF ANDROGEN SYNTHESIS

Defects in the pathway that regulates androgen synthesis (Fig. 349-4) cause underandrogenization of the male fetus (Table 349-1). Müllerian regression is unaffected because Sertoli cell function is preserved.

Figure 349-4

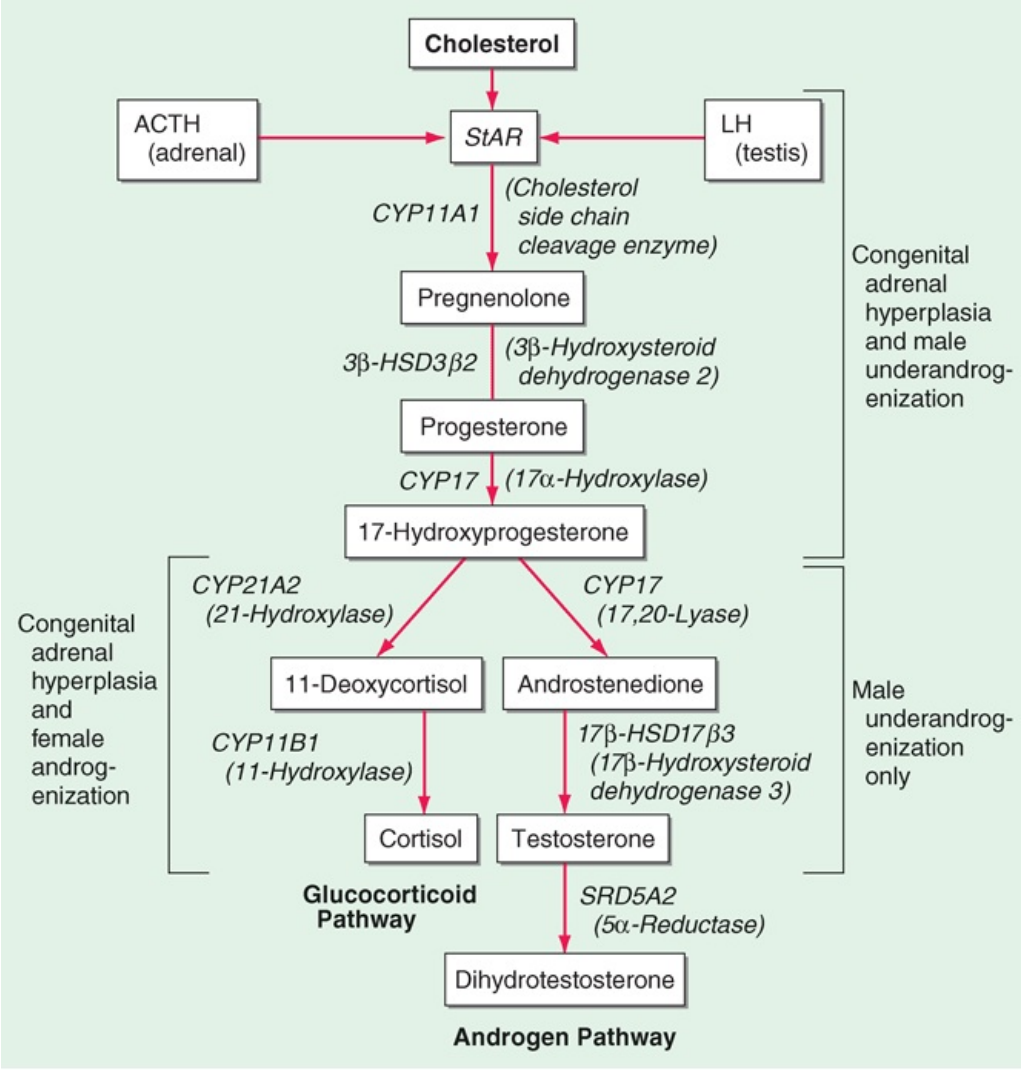

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Simplified overview of glucocorticoid and androgen synthesis pathways. Defects in *CYP21A2* and *CYP11B1* shunt steroid precursors into the androgen pathway and cause androgenization of 46,XX females. Testosterone is synthesized in the testicular Leydig's cells and converted to dihydrotestosterone peripherally. Defects in enzymes involved in androgen synthesis result in underandrogenization of 46,XY males. *StAR*, steroidogenic acute regulatory protein. *[After E Braunwald et al (eds): Harrison's Principles of Internal Medicine, 15th ed. New York, McGraw-Hill, 2001.]*

LH Receptor

Mutations in the LH receptor (LHCGR) cause Leydig's cell hypoplasia and androgen deficiency. Defects of LH receptor synthesis or function preclude hCG stimulation of Leydig's cells in utero, as well as LH stimulation of Leydig's cells late in gestation and during the neonatal period. As a result, testosterone and DHT synthesis are insufficient for normal androgenization of the internal and external genitalia, causing a spectrum of phenotypes that range from complete underandrogenization to

micropenis, depending on the severity of the mutation.

Steroidogenic Enzyme Pathways

Mutations in *steroidogenic acute regulatory protein (StAR)* and *CYP11A1* affect both adrenal and gonadal steroidogenesis (Chap. 342). Affected individuals (46,XY) usually have severe early-onset salt-losing adrenal failure and a female phenotype, although later-onset milder variants have been reported. Defects in *3β-hydroxysteroid dehydrogenase type 2* (*HSD3b2*) also cause adrenal insufficiency in severe cases, but the accumulation of dehydroepiandrosterone (DHEA) has a mild androgenizing effect, resulting in ambiguous genitalia or hypospadias. Patients with CAH due to *17α-hydroxylase (CYP17)deficiency* have variable underandrogenization and develop hypertension and hypokalemia due to the potent salt-retaining effects of corticosterone and 11-deoxycorticosterone. Patients with complete loss of 17α-hydroxylase function often present as phenotypic females who fail to enter puberty and are found to have inguinal testes and hypertension in adolescence. Some mutations in *CYP17* selectively impair 17,20 lyase activity without altering 17α-hydroxylase activity, leading to underandrogenization without mineralocorticoid excess and hypertension. Mutations in *P450 oxidoreductase (POR)* affect multiple steroidogenic enzymes, leading to impaired androgenization and a biochemical pattern of apparent combined 21-hydroxylase and 17α-hydroxylase deficiency, sometimes with skeletal abnormalities (Antley-Bixler craniosynostosis). Defects in *17*b*-hydroxysteroid dehydrogenase type 3(HSD17b3)* and *5*a*-reductase type 2 (SRD5A2)* interfere with the synthesis of testosterone and DHT, respectively. These conditions are characterized by minimal or absent androgenization in utero, but some phallic development can occur during adolescence due to the action of other enzyme isoforms. Individuals with *5α-reductase type 2* deficiency have normal wolffian structures and usually do not develop breast tissue. At puberty, the increase in testosterone induces muscle mass and other virilizing features despite DHT deficiency. Some individuals change gender from female to male at puberty. Thus, the management of this disorder is challenging. DHT cream can improve prepubertal phallic growth in patients raised as male. Gonadectomy before adolescence and estrogen replacement at puberty can be considered in individuals raised as females.

DISORDERS OF ANDROGEN ACTION

Androgen Insensitivity Syndrome

Mutations in the androgen receptor (AR) cause resistance to androgen (testosterone, DHT) action or the *androgen insensitivity syndrome* (AIS). AIS is a spectrum of disorders that affects at least 1 in 100,000 46,XY individuals. Because the androgen receptor is X-linked, only 46,XY offspring are affected if the mother is a carrier of a mutation. XY individuals with *complete AIS (formerly called testicular feminization syndrome)* have a female phenotype, normal breast development (due to aromatization of testosterone), a short vagina but no uterus (because MIS production is normal), scanty pubic and axillary hair, and a female psychosexual orientation. Gonadotropins and testosterone levels can be low, normal, or elevated, depending on the degree of androgen resistance and the contribution of estradiol to feedback inhibition of the hypothalamic-pituitary-gonadal axis. AMH/MIS levels in childhood are normal or high. Most patients present with inguinal hernias (containing testes) in childhood or with primary amenorrhea in adulthood. Gonadectomy sometimes is performed, as there is a low risk of malignancy, and estrogen replacement is prescribed. Alternatively, the gonads can be left in situ until breast development is complete. The use of graded dilators in adolescence is usually sufficient to dilate the vagina and permit sexual intercourse.

Partial AIS (Reifenstein's syndrome) results from less severe AR mutations. Patients often present in infancy with perineoscrotal hypospadias and small undescended testes and with gynecomastia at the time of puberty. Those individuals raised as males require hypospadias repair in childhood and breast reduction in adolescence. Supplemental testosterone rarely enhances androgenization significantly, as endogenous testosterone is already increased. More severely underandrogenized patients present with clitoral enlargement and labial fusion and may be raised as females. The surgical and psychosexual management of these patients is complex and requires active involvement of the parents and the patient during the appropriate stages of development. *Azoospermia* and male-factor infertility also have been described in association with mild loss-of-function mutations in the androgen receptor.

Other Disorders Affecting 46, Xy Males

Persistent müllerian duct syndrome is the presence of a uterus in an otherwise normal male. This condition can result from mutations in AMH or its receptor (AMHR2). The uterus may be removed, but damage to the vasa deferentia must be avoided. *Isolated hypospadias* occurs in ~1 in 200 males and is treated by surgical repair. Most cases are idiopathic, although evidence of penoscrotal hypospadias, poor phallic development, and/or bilateral cryptorchidism require investigation for an underlying disorder of sex development (e.g., partial gonadal dysgenesis, mild defect in testosterone action, or even severe forms of 46,XX CAH). Unilateral undescended testes (cryptorchidism) affects more than 3% of boys at birth. Orchidopexy should be considered if the testis has not descended by 6 to 9 months of age. Bilateral cryptorchidism occurs less frequently and should raise suspicion of gonadotropin deficiency or DSD. A small subset of patients with cryptorchidism may have mutations in the insulin-like 3 (*INSL3*) gene or its receptor LGR8 (also known as *GREAT*), which mediates normal

testicular descent. Ascending testis is being recognized increasingly as a distinct condition for which management is currently unclear. *Syndromic associations* and *intrauterine growth retardation* also occur relatively frequently in association with impaired testicular function or target tissue responsiveness, but the underlying etiology of many of these conditions is unknown.

46,XX Dsd (Androgenized Females)

Inappropriate androgenization of females (formerly called *female pseudohermaphroditism*) occurs when the gonad (ovary) contains androgen-secreting testicular material or after increased androgen exposure, which is usually adrenal in origin (Table 349-1).

46,XX TESTICULAR/OVOTESTICULAR DSD

Testicular tissue can develop in 46,XX testicular DSD (46,XX males) after translocation of *SRY* or duplication of *SOX9* or defects in *RSPO1* (Table 349-4).

Table 349-4 Selected Genetic Causes of Androgenization of Karyotypic Females (46,XX Dsd)

Gene	Inheritance	Gonad	Uterus	External Genitalia	Associated Features
Testicular/Ovotesticular DSD					
SRY	Translocation	Testis or ovotestis	–	Male or ambiguous	
SOX9	dup17q24	Unknown	–	Male or ambiguous	
RSPO1	AR	Testis or ovotestis	±	Male or ambiguous	Palmar plantar hyperkeratosis, squamous cell skin carcinoma
WNT4	AR	Testis or ovotestis	–	Male or ambiguous	SERKAL syndrome (renal dysgenesis, adrenal and lung hypoplasia)
Increased Androgen Synthesis					
HSD3b2	AR	Ovary	+	Clitoromegaly	CAH, primary adrenal failure, mild androgenization due to ↑ DHEA
CYP21A2	AR	Ovary	+	Ambiguous	CAH, phenotypic spectrum from severe salt-losing forms associated with adrenal failure to simple virilizing forms with compensated adrenal function, ↑ 17-hydroxyprogesterone
POR	AR	Ovary	+	Ambiguous or female	Mixed features of 21-hydroxylase deficiency and 17α -hydroxylase/17,20 lyase deficiency, sometimes associated with Antley-Bixler craniosynostosis
CYP11B1	AR	Ovary	+	Ambiguous	CAH, hypertension due to ↑ 11-deoxycortisol and 11-deoxycorticosterone
CYP19	AR	Ovary	+	Ambiguous	Maternal virilization during pregnancy, absent breast development at puberty
Glucocorticoid receptor	AR	Ovary	+	Ambiguous	↑ ACTH, 17-hydroxyprogesterone and cortisol; failure of dexamethasone suppression

Abbreviations: DSD, disorders of sex development; AR, autosomal recessive; *SRY*, sex-related gene on the Y chromosome; *SOX9*, *SRY*-related HMG-box gene 9; CAH, congenital adrenal hyperplasia; *HSD3b2*, 3β-hydroxysteroid dehydrogenase type 2; *CYP21A2*, 21-hydroxylase; POR, P450 oxidoreductase; *CYP11B1*, 11β-hydroxylase; *CYP19*, aromatase; ACTH, adrenocorticotropin; *RSPO1*, R-spondin 1.

INCREASED ANDROGEN EXPOSURE

21-Hydroxylase Deficiency (Congenital Adrenal Hyperplasia)

The *classic form* of 21-hydroxylase deficiency (21-OHD) is the most common cause of CAH (Chap. 342). It has an incidence between 1 in 10,000 and 1 in 15,000 and is the most common cause of androgenization in chromosomal 46,XX females (Table 349-4). Affected individuals are homozygous or compound heterozygous for severe mutations in the enzyme 21-hydroxylase (*CYP21A2*). This mutation causes a block in adrenal glucocorticoid and mineralocorticoid synthesis, increasing 17-hydroxyprogesterone and shunting steroid precursors into the androgen synthesis pathway (Fig. 349-4). Glucocorticoid insufficiency causes a compensatory elevation of adrenocorticotropin (ACTH), resulting in adrenal hyperplasia and additional synthesis of steroid precursors proximal to the

enzymatic block. Increased androgen synthesis in utero causes androgenization of the female fetus in the first trimester. Ambiguous genitalia are seen at birth, with varying degrees of clitoral enlargement and labial fusion. Excess androgen production causes gonadotropin-independent precocious puberty in males with 21-OHD.

The *salt-wasting* form of 21-OHD results from severe combined glucocorticoid and mineralocorticoid deficiency. A salt-wasting crisis usually manifests between 7 and 21 days of life and is a potentially life-threatening event that requires urgent fluid resuscitation and steroid treatment. Thus, a diagnosis of 21-OHD should be considered in any baby with ambiguous genitalia with bilateral nonpalpable gonads. Males (46,XY) with 21-OHD have no genital abnormalities at birth but are equally susceptible to adrenal insufficiency and salt-losing crises.

Females with the *classic simple virilizing* form of 21-OHD also present with genital ambiguity. They have impaired cortisol biosynthesis but do not develop salt loss. Patients with *nonclassic 21-OHD* produce normal amounts of cortisol and aldosterone, but at the expense of producing excess androgens. Hirsutism (60%), oligomenorrhea (50%), and acne (30%) are the most common presenting features. This is one of the most common recessive disorders in humans, with an incidence as high as 1 in 100 to 500 in many populations and 1 in 27 in Ashkenazi Jews of Eastern European origin.

Biochemical features of acute salt-wasting 21-OHD are hyponatremia, hyperkalemia, hypoglycemia, low cortisol and aldosterone, and elevated 17-hydroxyprogesterone, ACTH, and plasma renin activity. Presymptomatic diagnosis of classic 21-OHD is now made by neonatal screening tests for increased 17-hydroxyprogesterone in many centers. In most cases, 17-hydroxyprogesterone is markedly increased. In adults, ACTH stimulation (0.25 mg cosyntropin IV) with assays for 17-hydroxyprogesterone at 0 and 30 min can be useful for detecting nonclassic 21-OHD and heterozygotes (Chap. 342).

Treatment: Congenital Adrenal Hyperplasia

Acute salt-wasting crises require fluid resuscitation, IV hydrocortisone, and correction of hypoglycemia. Once the patient is stabilized, glucocorticoids must be given to correct the cortisol insufficiency and suppress ACTH stimulation, thereby preventing further virilization, rapid skeletal maturation, and the development of polycystic ovaries. Typically, hydrocortisone (10–15 mg/m^2 per day in three divided doses) is used in childhood with a goal of partially suppressing 17-hydroxyprogesterone (100–<1000 ng/dL). The aim of treatment is to use the lowest glucocorticoid dose that adequately suppresses adrenal androgen production without causing signs of glucocorticoid excess such as impaired growth and obesity. Salt-wasting conditions are treated with mineralocorticoid replacement. Infants usually need salt supplements up to the first year of life. Plasma renin activity and electrolytes are used to monitor mineralocorticoid replacement. Some patients with simple virilizing 21-OHD also benefit from mineralocorticoid supplements. Newer therapeutic approaches such as antiandrogens and aromatase inhibitors (to block premature epiphyseal closure) are under evaluation. Parents and patients should be aware of the need for increased doses of steroids during sickness, and patients should carry medic alert systems.

Older adolescents and adults often are treated with prednisolone or with dexamethasone at night to provide more complete ACTH suppression. Steroid doses should be adjusted to individual requirements as overtreatment results in weight gain and hypertension and can affect bone turnover. Androstenedione and testosterone may be useful measurements of long-term control, with less fluctuation than 17-hydroxyprogesterone. Mineralocorticoid requirements often decrease in adulthood, and doses should be reduced to avoid hypertension. In very severe cases, adrenalectomy has been advocated but incurs the risks of surgery and total adrenal insufficiency.

Girls with significant genital androgenization due to classic 21-OHD usually undergo vaginal reconstruction and clitoral reduction (maintaining the glans and nerve supply), but the optimal timing of these procedures is debated, as is the need for the individual to be able to consent. There is a higher threshold for undertaking clitoral surgery in some centers as long-term sensation and ability to achieve orgasm can be affected, but the long-term results of newer techniques are not yet known. If surgery is performed in infancy, surgical revision or regular vaginal dilatation may be needed in adolescence or adulthood, and long-term psychological support and psychosexual counseling may be appropriate. Women with 21-OHD frequently develop polycystic ovaries and have reduced fertility, especially when control is poor. Fecundity is achieved in up to 90% of women, but ovulation induction (or even adrenalectomy) is frequently required. Dexamethasone should be avoided in pregnancy. Men with poorly controlled 21-OHD may develop testicular adrenal rests and are at risk for reduced fertility. Prenatal treatment of 21-OHD by the administration of dexamethasone to mothers is still under evaluation. However, treatment of the mother and child must be started ideally before 6 to 7 weeks; long-term effects of prenatal dexamethasone exposure on fetal development are still under evaluation.

The treatment of other forms of CAH includes mineralocorticoid and glucocorticoid replacement for salt-losing conditions (e.g., *StAR*, *CYP11A1*, *HSD3b2*), suppression of ACTH drive with glucocorticoids in disorders associated with hypertension (e.g., *CYP17*, *CYP11B1*), and appropriate sex-hormone replacement in adolescence and adulthood, when necessary.

Other Causes

Increased androgen synthesis can also occur in CAH due to defects in *POR*, *11*b-*hydroxylase(CYP11B1),* and *3*b-*hydroxysteroid dehydrogenase type 2 (HSD3b2)* and with mutations in the genes encoding *aromatase (CYP19)* and the glucocorticoid receptor. Increased androgen exposure in utero can occur with maternal virilizing tumors and with ingestion of androgenic compounds.

Other Disorders Affecting 46,XX Females

Congenital absence of the vagina occurs in association with *müllerian agenesis* or *hypoplasia* as part of the Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome (rarely caused by *WNT4* mutations). This diagnosis should be considered in otherwise phenotypically normal females with primary amenorrhea. Associated features include renal (agenesis) and cervical spinal abnormalities.

Global Considerations

The approach to a child or adolescent with ambiguous genitalia or another DSD requires cultural sensitivity, as the concepts of sex and gender vary widely. Rare genetic DSDs can occur more frequently in specific populations (e.g., *5*β*-reductase type 2* in the Dominican Republic). Different forms of CAH also show ethnic and geographic variability. In many countries, appropriate biochemical tests may not be readily available, and access to appropriate forms of surgery or treatment may be limited.

FURTHER READINGS

Bojesen A, Gravholt CH: Klinefelter syndrome in clinical practice. Nat Clin Pract Urol 4:192, 2007[PMID: 17415352] [Full Text]

Bondy CA for the Turner Syndrome Study Group: Care of girls and women with Turner syndrome: A guideline of the Turner Syndrome Study Group. J Clin Endocrinol Metab 92:10, 2007[PMID: 17047017] [Full Text]

Brain CE et al: Holistic management of DSD. Best Pract Res Clin Endocrinol Metab 24:335, 2010[PMID: 20541156] [Full Text]

Cools M et al: Germ cell tumors in the intersex gonad: Old paths, new directions, moving frontiers. Endocrinol Rev 27:468, 2006[PMID: 16735607] [Full Text]

Lee PA et al: Consensus statement of management of intersex disorders. Pediatrics 118:e488, 2006

Mendonca BB et al: 46,XY disorders of sex development (DSD). Clin Endocrinol 70:173, 2009[PMID: 18811725] [Full Text]

Speiser PW et al: Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 95:4133, 2010[PMID: 20823466] [Full Text]

Wilhelm D et al: Sex determination and gonadal development in mammals. Physiol Rev 87:1, 2007[PMID: 17237341] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GENERAL FEATURES OF GASTROINTESTINAL (GI) NEUROENDOCRINE TUMORS

Gastrointestinal neuroendocrine tumors (NETs) are tumors derived from the diffuse neuroendocrine system of the GI tract; that system is composed of amine- and acid-producing cells with different hormonal profiles, depending on the site of origin. The tumors historically are divided into carcinoid tumors and pancreatic endocrine tumors (PETs), although recent pathologic classifications have proposed that they all be classified as gastrointestinal NETs. In this chapter the term *carcinoid tumor* is retained because it is widely used. These tumors originally were classified as APUDomas (for *a*mine *p*recursor *u*ptake and *d*ecarboxylation), as were pheochromocytomas, melanomas, and medullary thyroid carcinomas, because they share certain cytochemical features as well as various pathologic, biologic, and molecular features (Table 350-1). It was originally proposed that APUDomas had a similar embryonic origin from neural crest cells, but it is now known the peptide-secreting cells are not of neuroectodermal origin. Nevertheless, the concept of APUDomas is useful because the tumors from the cells have important similarities as well as some differences (Table 350-1). In this section, the areas of similarity between PETs and carcinoids will be discussed together and areas in which there are important differences will be discussed separately.

Table 350-1 General Characteristics of Gastrointestinal Neuroendocrine Tumors [Carcinoids, Pancreatic Endocrine Tumors (Pets)]

A. Share general neuroendocrine cell markers (identification used for diagnosis)
1. Chromogranins (A, B, C) are acidic monomeric soluble proteins found in the large secretory granules. Chromogranin A is the most widely used.
2. Neuron-specific enolase (NSE) is the γ-γ dimer of the enzyme enolase and is a cytosolic marker of neuroendocrine differentiation.
3. Synaptophysin is an integral membrane glycoprotein of 38,000 molecular weight found in small vesicles of neurons and neuroendocrine tumors.
B. Pathologic similarities
1. All are APUDomas showing amine precursor uptake and decarboxylation.
2. Ultrastructurally they have dense-core secretory granules (>80 nm).
3. Histologically, generally appear similar with few mitoses and uniform nuclei.
4. Frequently synthesize multiple peptides/amines, which can be detected immunocytochemically but may not be secreted.
5. Presence or absence of clinical syndrome or type cannot be predicted by immunocytochemical studies.
6. Histologic classifications increasingly predictive of biologic behavior. Only invasion or metastases establish malignancy
C. Similarities of biologic behavior
1. Generally slow growing, but a proportion are aggressive.
2. Secrete biologically active peptides/amines, which can cause clinical symptoms.
3. Generally have high densities of somatostatin receptors, which are used for both localization and treatment.
D. Similarities/differences in molecular abnormalities
1. Similarities
a. Uncommon—alterations in common oncogenes (ras, jun, fos, etc).
b. Uncommon—alterations in common tumor-suppressor genes (p53, retinoblastoma).
c. Alterations at MEN 1 locus (11q13) and $p16^{INK4a}$ (9p21) occur in a proportion (10–45%).
d. Methylation of various genes occurs in 40–87% (ras-associated domain family I, p14, p16, O^6 methyl guanosine methyltransferases, retinoic acid receptor β)
2. Differences
a. PETs—loss of 1p (21%), 3p (8–47%), 3q (8–41%), 11q (21–62%), 6q (18–68%). Gains at 17q (10–55%), 7q (16–68%), 4q (33%).
b. Carcinoids—loss of 18q (38–67%) >18p (33–43%) > 9p, 16q21(21–23%). Gains at 17q, 19p (57%), 4q (33%), 14q (20%).

Abbreviation: MEN 1, multiple endocrine neoplasia type 1.

CLASSIFICATION/PATHOLOGY/TUMOR BIOLOGY OF NETS

NETs generally are composed of monotonous sheets of small round cells with uniform nuclei, and mitoses are uncommon. They can be identified tentatively on routine histology; however, these tumors are now recognized principally by their histologic staining patterns due to shared cellular proteins. Historically, silver staining was used, and tumors were classified as showing an argentaffin reaction if they took up and reduced silver or as being argyrophilic if they did not reduce it. More recently immunocytochemical localization of chromogranins (A, B, C), neuron-specific enolase, and

synaptophysin, which are all neuroendocrine cell markers, is used (Table 350-1). Chromogranin A is currently the most widely used.

Ultrastructurally, these tumors possess electron-dense neurosecretory granules and frequently contain small clear vesicles that correspond to synaptic vesicles of neurons. NETs synthesize numerous peptides, growth factors, and bioactive amines that may be ectopically secreted, giving rise to a specific clinical syndrome (Table 350-2). The diagnosis of the specific syndrome requires the clinical features of the disease (Table 350-2) and cannot be made from the immunocytochemistry results alone. The presence or absence of a specific clinical syndrome also cannot be predicted from the immunocytochemistry alone (Table 350-1). Furthermore, pathologists cannot distinguish between benign and malignant NETs unless metastases or invasion is present.

Table 350-2 Gastrointestinal Neuroendocrine Tumor Syndrome

Name	Biologically Active Peptide(s) Secreted	Incidence (New Cases/10^6 Population/Year)	Tumor Location	Malignant, %	Associated with MEN 1, %	Main Symptoms/Signs
I. Established Specific Functional Syndrome						
A. Carcinoid tumor						
Carcinoid syndrome	Serotonin, possibly tachykinins, motilin, prostaglandins	0.5–2	Midgut (75–87%) Foregut (2–33%) Hindgut (1–8%) Unknown (2–15%)	95–100	Rare	Diarrhea (32–84%) Flushing (63–75%) Pain (10–34%) Asthma (4–18%) Heart disease (11–41%)
B. Pancreatic endocrine tumor						
Zollinger-Ellison syndrome	Gastrin	0.5–1.5	Duodenum (70%) Pancreas (25%) Other sites (5%)	60–90	20–25	Pain (79–100%) Diarrhea (30–75%) Esophageal symptoms (31–56%)
Insulinoma	Insulin	1–2	Pancreas (>99%)	<10	4–5	Hypoglycemic symptoms (100%)
VIPoma (Verner-Morrison syndrome, pancreatic cholera, WDHA)	Vasoactive intestinal peptide	0.05–0.2	Pancreas (90%, adult) Other (10%, neural,adrenal, periganglionic)	40–70	6	Diarrhea (90–100%) Hypokalemia (80–100%) Dehydration (83%)
Glucagonoma	Glucagon	0.01–0.1	Pancreas (100%)	50–80	1–20	Rash (67–90%) Glucose intolerance (38–87%) Weight loss (66–96%)
Somatostatinoma	Somatostatin	Rare	Pancreas (55%) Duodenum/jejunum (44%)	>70	45	Diabetes mellitus (63–90%) Cholelithiases (65–90%) Diarrhea (35–90%)
GRFoma	Growth hormone–releasing hormone	Unknown	Pancreas (30%) Lung (54%) Jejunum (7%) Other (13%)	>60	16	Acromegaly (100%)
ACTHoma	ACTH	Rare	Pancreas (4–16% all ectopic Cushing's)	>95	Rare	Cushing's syndrome (100%)
PET causing carcinoid syndrome	Serotonin, tachykinins	Rare (43 cases)	Pancreas (<1% all carcinoids)	60–88	Rare	Same as carcinoid syndrome above
PET causing hypercalcemia	PTHrP Others unknown	Rare	Pancreas (rare cause of hypercalcemia)	84	Rare	Abdominal pain due to hepatic metastases
II. Possible Specific Functional Syndrome						
PET secreting calcitonin	Calcitonin	Rare	Pancreas (rare cause of hypercalcitonemia)	>80	16	Diarrhea (50%)
PET secreting renin	Renin	Rare	Pancreas	Unknown	No	Hypertension
PET secreting luteinizing hormone	Luteinizing hormone	Rare	Pancreas	Unknown	No	Anovulation, virilization (female); reduced libido (male)

PET secreting erythropoietin	Erythropoietin	Rare	Pancreas	100	No	Polycythemia
PET secreting IF-II	Insulin-like growth growth factor II	Rare	Pancreas	Unknown	No	Hypoglycemia
III. No Functional Syndrome						
PPoma/nonfunctional	None	1–2	Pancreas (100%)	>60	18–44	Weight loss (30–90%) Abdominal mass (10–30%) Pain (30–95%)

Abbreviations: ACTH, adrenocorticotropic hormone; GRFoma, growth hormone-releasing factor secreting pancreatic endocrine tumor; IF-II, insulin-like growth factor 2; MEN, multiple endocrine neoplasia; PET, pancreatic endocrine tumor; PPoma, tumor secreting pancreatic polypeptide; PTHrP, parathyroid hormone–related peptide; VIPoma, tumor secreting vasoactive intestinal peptide; WDHA, *w*atery *d*iarrhea, *h*ypokalemia, and *a*chlorhydria syndrome.

Carcinoid tumors frequently are classified according to their anatomic area of origin (i.e., foregut, midgut, hindgut) because tumors with similar areas of origin share functional manifestations, histochemistry, and secretory products (Table 350-3). Foregut tumors generally have a low serotonin (5-HT) content; are argentaffin-negative but argyrophilic; occasionally secrete adrenocorticotropic hormone (ACTH) or 5-hydroxytryptophan (5-HTP), causing an atypical carcinoid syndrome (Fig. 350-1); are often multihormonal; and may metastasize to bone. They uncommonly produce a clinical syndrome due to the secreted products. Midgut carcinoids are argentaffin-positive, have a high serotonin content, most frequently cause the typical carcinoid syndrome when they metastasize (Table 350-3, Fig. 350-1), release serotonin and tachykinins (substance P, neuropeptide K, substance K), rarely secrete 5-HTP or ACTH, and less commonly metastasize to bone. Hindgut carcinoids (rectum, transverse and descending colon) are argentaffin-negative, are often argyrophilic, rarely contain serotonin or cause the carcinoid syndrome (Fig. 350-1, Table 350-3), rarely secrete 5-HTP or ACTH, contain numerous peptides, and may metastasize to bone.

Table 350-3 Carcinoid Tumor Location, Frequency of Metastases, and Association with the Carcinoid Syndrome

	Location (% of Total)	Incidence of Metastases	Incidence of Carcinoid Syndrome
Foregut			
Esophagus	<0.1	—	—
Stomach	4.6	10	9.5
Duodenum	2.0	—	3.4
Pancreas	0.7	71.9	20
Gallbladder	0.3	17.8	5
Bronchus, lung, trachea	27.9	5.7	13
Midgut			
Jejunum	1.8	{58.4	9
Ileum	14.9		9
Meckel's diverticulum	0.5	—	13
Appendix	4.8	38.8	<1
Colon	8.6	51	5
Liver	0.4	32.	—
Ovary	1.0	2 32	50
Testis	<0.1	—	50
Hindgut			
Rectum	13.6	3.9	—

Source: Location is from the PAN-SEER data (1973–1999), and incidence of metastases from the SEER data (1992–1999), reported by IM Modlin et al: Cancer 97:934, 2003. Incidence of carcinoid syndrome is from 4349 cases studied from 1950–1971, reported by JD Godwin: Cancer 36:560, 1975.

Pancreatic endocrine tumors can be classified into nine well-established specific functional syndromes (Table 350-2), five possible specific functional syndromes (PETs secreting calcitonin, renin, luteinizing hormone, erythropoietin, or insulin-like growth factor II) (Table 350-2), and nonfunctional PETs (pancreatic polypeptide-secreting tumors; PPomas). Other functional hormonal syndromes due to nonpancreatic tumors (usually intraabdominal in location) have been described only rarely and are not included in Table 350-2. They include secretion of glucagon-like peptide-2 (GLP-2) that causes intestinal villus hypertrophy (enteroglucagonomas), secretion of GLP-1 that causes hypoglycemia and delayed transit, and intestinal and ovarian tumors secreting peptide tyrosine tyrosine (PYY) that result in altered motility and constipation. Each of the functional syndromes listed in Table 350-2 is associated with symptoms due to the specific hormone released. In contrast, nonfunctional PETs release no products that cause a specific clinical syndrome. "Nonfunctional" is a misnomer in the strict sense because those tumors frequently ectopically secrete a number of peptides [pancreatic polypeptide (PP), chromogranin A, ghrelin, neurotensin, ∝ subunits of human chorionic gonadotropin, neuron-specific enolase]; however, they cause no specific clinical syndrome. The symptoms caused by nonfunctional PETs are entirely due to the tumor per se.

Carcinoid tumors can occur in almost any GI tissue (Table 350-3); however, at present most (70%) have their origin in one of three sites: bronchus, jejunoileum, or colon/rectum. In the past, carcinoid tumors most frequently were reported in the appendix (i.e., 40%); however, the bronchus/lung, rectum, and small intestine are now the most common sites. Overall,

the GI tract is the most common site for these tumors, accounting for 64%, with the respiratory tract a distant second at 28%. Both race and sex can affect the frequency as well as the distribution of carcinoid tumors. African Americans have a high incidence of carcinoids, and rectal carcinoids are the most common. Females have a lower incidence of small-intestinal and pancreatic carcinoids.

The term *pancreatic endocrine tumor*, although widely used and therefore retained here, is also a misnomer, strictly speaking, because these tumors can occur either almost entirely in the pancreas (insulinomas, glucagonomas, nonfunctional PETs, PETs causing hypercalcemia) or at both pancreatic and extrapancreatic sites [gastrinomas, VIPomas (vasoactive intestinal peptide), somatostatinomas, GRFomas (growth hormone–releasing factor)]. PETs are also called islet cell tumors; however, the use of this term is discouraged because it is not established that they originate from the islets and many can occur at extrapancreatic sites.

A number of new classification systems have been proposed for both carcinoids and PETs. In the World Health Organization (WHO) classification it has been proposed that these tumors all be classified as GI neuroendocrine tumors (including carcinoids and PETs), which divides them into three general categories: (1a) well-differentiated NETs, (1b) well-differentiated neuroendocrine carcinomas that have low-grade malignancy, and (2) poorly differentiated neuroendocrine carcinomas that are usually small cell neuroendocrine carcinomas of high-grade malignancy. The term *carcinoid* is synonymous with *well-differentiated NETs* (1a). This classification is further divided on the basis of tumor location and biology. In addition, for the first time a standard TNM (tumor, node, metastasis) classification and grading system has been proposed for GI neuroendocrine tumors. The new WHO classification and the TNM classification and grading system were proposed to facilitate the comparison and evaluation of clinical, pathologic, and prognostic features and results of treatment in GI NETs from different studies. These classification systems may provide important prognostic information that can guide treatment (Table 350-4).

Table 350-4 Prognostic Factors in Neuroendocrine Tumors

I. Both carcinoid tumors and PETs
Presence of liver metastases ($p < .001$)
Extent of liver metastases ($p < .001$)
Presence of lymph node metastases ($p < .001$)
Depth of invasion ($p < .001$)
Rapid rate of tumor growth
Elevated serum alkaline phosphatase levels ($p = .003$)
Primary tumor site ($p < .001$)
Primary tumor size ($p < .005$)
Various histologic features
Tumor differentiation ($p < .001$)
High growth indices (high K_{i-67} index, PCNA expression)
High mitotic counts ($p < .001$)
Necrosis present
Presence of cytokeratin 19 ($p < .02$)
Vascular or perineural invasion
Vessel density (low microvessel density, increased lymphatic density)
High CD10 metalloproteinase expression (in series with all grades of NETs)
Flow cytometric features (i.e., aneuploidy)
High VEGF expression (in low-grade or well-differentiated NETs only)
WHO, TNM, and grading classification
Presence of a pancreatic NET rather than GI NET associated with poorer prognosis ($p = .0001$)
Older age ($p < .01$)
II. Carcinoid tumors
Presence of carcinoid syndrome
Laboratory results [urinary 5-HIAA levels ($p < .01$), plasma neuropeptide K ($p < .05$), serum chromogranin A ($p < .01$)]
Presence of a second malignancy
Male sex ($p < .001$)
Mode of discovery (incidental > symptomatic)
Molecular findings [TGF-α expression ($p < .05$), chr 16q LOH or gain chr 4p ($p < .05$)]
WHO, TNM, and grading classification
Molecular findings [gain in chr 14, loss of 3p13 (ileal carcinoid), upregulation of Hoxc6]
III. PETs
Ha-ras oncogene or p53 overexpression
Female gender
MEN 1 syndrome absent
Presence of nonfunctional tumor (some studies, not all)
WHO, TNM, and grading classification
Laboratory findings (increased chromogranin A in some studies; gastrinomas—increased gastrin level)
Molecular findings [increased HER2/neu expression ($p = .032$), chr 1q, 3p, 3q, or 6q LOH ($p = .0004$), EGF receptor overexpression ($p = .034$), gains in chr 7q, 17q, 17p, 20q; alterations in the VHL gene (deletion, methylation)

The exact incidence of carcinoid tumors or PETs varies according to whether only symptomatic tumors or all tumors are considered. The incidence of clinically significant carcinoids is 7–13 cases/million population per year, whereas any malignant carcinoids at autopsy are reported in 21–84 cases/million population per year. The incidence of GI NETs is approximately 25–50 cases per million in the United States, which makes them less common than adenocarcinomas of the GI tract. However, their incidence has increased sixfold in the last 30 years. Clinically significant PETs have a prevalence of 10 cases/million population, with insulinomas, gastrinomas, and nonfunctional PETs having an incidence of 0.5–2 cases/million population per year (Table 350-2). VIPomas are two to eight times less common, glucagonomas are 17 to 30 times less common, and somatostatinomas are the least common. In autopsy studies 0.5–1.5% of all cases have a PET; however, in less than 1 in 1000 cases was a functional tumor thought to occur.

Both carcinoid tumors and PETs commonly show malignant behavior (Tables 350-2 and 350-3). With PETs, except for insulinomas in which <10% are malignant, 50–100% in different series are malignant. With carcinoid tumors the percentage showing malignant behavior varies in different locations. For the three most common sites of occurrence, the incidence of metastases varies greatly from jejunoileum (58%) > lung/bronchus (6%) > rectum (4%) (Table 350-3). With both carcinoid tumors and PETs, a number of factors, summarized in Table 350-4, are important prognostic factors in determining survival and the aggressiveness of the tumor. Patients with PETs (excluding insulinomas) generally have a poorer prognosis than do patients with GI NETs (carcinoids). The presence of liver metastases is the single most important prognostic factor in single and multivariate analyses for both carcinoid tumors and PETs. Particularly important in the development of liver metastases is the size of the primary tumor. For example, with small-intestinal carcinoids, which are the most common cause of the carcinoid syndrome due to metastatic disease in the liver (Table 350-2), metastases occur in 15–25% if the tumor diameter is <1 cm, 58–80% if it is 1–2 cm in diameter, and >75% if it is >2 cm in diameter. Similar data exist for gastrinomas and other PETs in which the size of the primary tumor is an independent predictor of the development of liver metastases. The presence of lymph node metastases; the depth of invasion; the rapid rate of growth; various histologic features [differentiation, mitotic rates, growth indices, vessel density, vascular endothelial growth factor (VEGF), and CD10 metalloproteinase expression]; necrosis; presence of cytokeratin; elevated serum alkaline phosphatase levels; older age; advanced stages in WHO, TNM, or grading classification systems; and flow cytometric results such as the presence of aneuploidy are all important prognostic factors for the development of metastatic disease (Table 350-4). For patients with carcinoid tumors, additional associations with a worse prognosis include the development of the carcinoid syndrome (especially the development of carcinoid heart disease), male sex, the presence of a symptomatic tumor or greater increases in a number of tumor markers [5-hydroxyindolacetic acid (5-HIAA), neuropeptide K, chromogranin A], and the presence of various molecular features. With PETs or gastrinomas, which have been the best studied PET long-term, a worse prognosis is associated with female sex, overexpression of the Ha-ras oncogene or p53, the absence of multiple endocrine neoplasia type 1 (MEN 1), higher levels of various tumor markers (i.e., chromogranin A, gastrin), and various molecular features (Table 350-4).

A number of diseases due to various genetic disorders are associated with an increased incidence of neuroendocrine tumors (Table 350-5). Each one is caused by a loss of a possible tumor-suppressor gene. The most important is MEN 1, which is an autosomal dominant disorder due to a defect in a 10-exon gene on 11q13, which encodes for a 610-amino-acid nuclear protein, menin (Chap. 351). Patients with MEN 1 develop hyperparathyroidism due to parathyroid hyperplasia in 95–100% of cases, PETs in 80–100%, pituitary adenomas in 54–80%, adrenal adenomas in 27–36%, bronchial carcinoids in 8%, thymic carcinoids in 8%, gastric carcinoids in 13–30% of patients with Zollinger-Ellison syndrome, skin tumors [angiofibromas (88%), collagenomas (72%)], central nervous system (CNS) tumors [meningiomas (<8%)], and smooth-muscle tumors [leiomyomas, leiomyosarcomas (1–7%)]. Among patients with MEN 1, 80–100% develop nonfunctional PETs (most are microscopic with 0–13% large/symptomatic), functional PETs occur in 20–80% in different series with a mean of 54% developing Zollinger-Ellison syndrome, 18% insulinomas, 3% glucagonomas, 3% VIPomas, and <1% GRFomas or somatostatinomas. MEN 1 is present in 20–25% of all patients with Zollinger-Ellison syndrome, 4% of patients with insulinomas, and a low percentage (<5%) of patients with the other PETs.

Table 350-5 Genetic Syndromes Associated with an Increased Incidence of Neuroendocrine Tumors (Nets) [Carcinoids or Pancreatic Endocrine Tumors (Pets)]

Syndrome	Location of Gene Mutation and Gene Product	NETs Seen/Frequency
Multiple endocrine neoplasia type 1 (MEN 1)	11q13 (encodes 610-amino-acid protein, menin)	80–100% develop PETs (microscopic), 20–80% (clinical): (nonfunctional > gastrinoma > insulinoma) Carcinoids: gastric (13–30%), bronchial/thymic (8%)
von Hippel–Lindau disease	3q25 (encodes 213-amino-acid protein)	12–17% develop PETs (almost always nonfunctional)
von Recklinghausen's disease [neurofibromatosis 1 (NF-1)]	17q11.2 (encodes 2485-amino-acid protein, neurofibromin)	0–10% develop PETs, primarily duodenal somatostatinomas (usually nonfunctional) Rarely insulinoma, gastrinoma
Tuberous sclerosis	9q34 (TSCI) encodes 1164-amino-acid protein, hamartin) 16p13 (TSC2) (encodes 1807-amino-acid protein, tuberin)	Uncommonly develop PETs [nonfunctional and functional (insulinoma, gastrinoma)]

Three phacomatoses associated with neuroendocrine tumors are von Hippel–Lindau disease (VHL), von Recklinghausen's disease [neurofibromatosis type 1 (NF-1)], and tuberous sclerosis (Bourneville's disease) (Table 350-5). VHL is an autosomal dominant disorder due to defects on chromosome 3p25, which encodes for a 213-amino-acid protein that interacts with the elongin family of proteins as a transcriptional regulator (Chaps. 284, 343, 351, 379). In addition to cerebellar hemangioblastomas, renal cancer, and pheochromocytomas, 10–17% develop a PET. Most are nonfunctional, although insulinomas and VIPomas have been reported. Patients with NF-1 (von Recklinghausen's disease) have defects in a gene on chromosome 17q11.2 that encodes for a 2845-amino-acid protein, neurofibromin, which functions in normal cells as a suppressor of the ras signaling cascade (Chap. 379). Up to 10% of these patients develop an upper GI carcinoid tumor, characteristically in the periampullary region (54%). Many are classified as somatostatinomas because they contain somatostatin immunocytochemically; however, they uncommonly secrete somatostatin and rarely produce a clinical somatostatinoma syndrome. NF-1 has rarely been associated with insulinomas and Zollinger-Ellison syndrome. NF-1 accounts for 48% of all duodenal somatostastinomas and 23% of all ampullary carcinoid tumors. Tuberous sclerosis

is caused by mutations that alter either the 1164-amino-acid protein hamartin (TSC1) or the 1807-amino-acid protein tuberin (TSC2) (Chap. 379). Both hamartin and tuberin interact in a pathway related to phosphatidylinositol 3-kinases and mTor signaling cascades. A few cases including nonfunctional and functional PETs (insulinomas and gastrinomas) have been reported in these patients (Table 350-5).

In contrast to most common nonendocrine tumors, such as carcinoma of the breast, colon, lung, or stomach, in neither PETs nor carcinoid tumors have alterations in common oncogenes (*ras, myc, fos, src, jun*) or common tumor-suppressor genes (p53, retinoblastoma susceptibility gene) been found to be generally important in their molecular pathogenesis (Table 350-1). Alterations that may be important in their pathogenesis include changes in the *MEN 1* gene, p16/MTS1 tumor-suppressor gene, and *DPC 4/Smad 4 gene*; amplification of the HER-2/*neu* protooncogene; alterations in transcription factors [Hoxc6 (GI carcinoids)], growth factors, and their receptor expression; methylation of a number of genes that probably results in their inactivation; and deletions of unknown tumor-suppressor genes as well as gains in other unknown genes (Table 350-1). Comparative genomic hybridization, genome-wide allelotyping studies, and genome-wide single-nucleotide polymorphism analyses have shown that chromosomal losses and gains are common in PETs and carcinoids, but they differ between these two NETs and some have prognostic significance (Table 350-4). Mutations in the *MEN 1* gene are probably particularly important. There is loss of heterozygosity at the MEN 1 locus on chromosome 11q13 in 93% of sporadic PETs (i.e., in patients without MEN 1) and in 26–75% of sporadic carcinoid tumors. Mutations in the *MEN 1* gene are reported in 31–34% of sporadic gastrinomas. The presence of a number of these molecular alterations (PET or carcinoid) correlates with tumor growth, tumor size, and disease extent or invasiveness and may have prognostic significance.

CARCINOID TUMORS AND CARCINOID SYNDROME

Characteristics of the Most Common GI Carcinoid Tumors

APPENDICEAL CARCINOIDS

Appendiceal carcinoids occur in 1 in every 200–300 appendectomies, usually in the appendiceal tip. Most (i.e., >90%) are <1 cm in diameter without metastases in older studies, but more recently 2–35% have had metastases (Table 350-3). In the SEER data of 1570 appendiceal carcinoids, 62% were localized, 27% had regional metastases, and 8% had distant metastases. Approximately 50% between 1 and 2 cm metastasized to lymph nodes. Their percentage of the total number of carcinoids decreased from 43.9% (1950–1969) to 2.4% (1992–1999).

SMALL-INTESTINAL CARCINOIDS

Small-intestinal carcinoids account for approximately one-third of all small-bowel tumors in various surgical series. These are frequently multiple; 70–80% are present in the ileum, and 70% within 6 cm (24 in.) of the ileocecal valve. Forty percent are <1 cm in diameter, 32% are 1–2 cm, and 29% are >2 cm. Between 35 and 70% are associated with metastases (Table 350-3). They characteristically cause a marked fibrotic reaction, which can lead to intestinal obstruction. Distant metastases occur to liver in 36–60%, to bone in 3%, and to lung in 4%. As discussed previously, tumor size is an important variable in the frequency of metastases. However, even a proportion of small carcinoid tumors of the small intestine (<1 cm) have metastases in 15–25% of cases, whereas the proportion increases to 58–100% for tumors 1–2 cm in diameter. Carcinoids also occur in the duodenum, with 31% having metastases. No duodenal tumor <1 cm in two series metastasized, whereas 33% of those >2 cm had metastases. Small-intestinal carcinoids are the most common cause (60–87%) of the carcinoid syndrome and are discussed in a later section (Table 350-6).

Table 350-6 Clinical Characteristics in Patients with Carcinoid Syndrome

	At Presentation	During Course of Disease
Symptoms/signs		
Diarrhea	32–73%	68–84%
Flushing	23–65%	63–74%
Pain	10%	34%
Asthma/wheezing	4–8%	3–18%
Pellagra	2%	5%
None	12%	22%
Carcinoid heart disease present	11%	14–41%
Demographics		
Male	46–59%	46–61%
Age		
Mean	57 yrs	52–54 yrs
Range	25–79 yrs	9–91 yrs
Tumor location		
Foregut	5–9%	2–33%
Midgut	78–87%	60–87%
Hindgut	1–5%	1–8%
Unknown	2–11%	2–15%

RECTAL CARCINOIDS

Rectal carcinoids represent 1–2% of all rectal tumors. They are found in approximately 1 in every 2500 proctoscopies. Nearly all occur between 4 and 13 cm above the dentate line. Most are small, with 66–80% being <1 cm in diameter, and rarely metastasize (5%). Tumors between 1 and 2 cm can metastasize in 5–30%, and those >2 cm, which are uncommon, in >70%.

BRONCHIAL CARCINOIDS

Bronchial carcinoids account for 1–2% of primary lung tumors. The frequency of bronchial carcinoids has increased more than fivefold over the last 30 years. A number of different classifications of bronchial carcinoid tumors have been

proposed. In some studies, lung NETs are classified into four categories: typical carcinoid [also called bronchial carcinoid tumor, Kulchitsky cell carcinoma I (KCC-I)], atypical carcinoid [also called well-differentiated neuroendocrine carcinoma (KC-II)], intermediate small cell neuroendocrine carcinoma, and small cell neuroendocarcinoma (KC-III). Another proposed classification includes three categories of lung NETs: benign or low-grade malignant (typical carcinoid), low-grade malignant (atypical carcinoid), and high-grade malignant (poorly differentiated carcinoma of the large cell or small cell type). The WHO classification includes four general categories: typical carcinoid, atypical carcinoid, large cell neuroendocrine carcinoma, and small cell carcinoma. These different categories of lung NETs have different prognoses, varying from excellent for typical carcinoid to poor for small cell neuroendocrine carcinomas. The occurrence of large cell and small cell lung carcinoids, but not typical or atypical lung carcinoids, is related to tobacco use.

GASTRIC CARCINOIDS

Gastric carcinoids account for 3 of every 1000 gastric neoplasms. Three different subtypes of gastric carcinoids are proposed to occur. Each originates from gastric enterochromaffin-like cells (ECL cells), one of the six types of gastric neuroendocrine cells, in the gastric mucosa. Two subtypes are associated with hypergastrinemic states, either chronic atrophic gastritis (type I) (80% of all gastric carcinoids) or Zollinger-Ellison syndrome, which is almost always a part of the MEN 1 syndrome (type II) (6% of all cases). These tumors generally pursue a benign course, with type 1 uncommonly (<10%) associated with metastases, whereas type II tumors are slightly more aggressive with 10–30% percentage associated with metastases. They are usually multiple and small and infiltrate only to the submucosa. The third subtype of gastric carcinoid (type III) (sporadic) occurs without hypergastrinemia (14–25% of all gastric carcinoids) and has an aggressive course, with 54–66% developing metastases. Sporadic carcinoids are usually single, large tumors; 50% have atypical histology, and they can be a cause of the carcinoid syndrome. Gastric carcinoids as a percentage of all carcinoids are increasing in frequency [1.96% (1969–1971), 3.6% (1973–1991), 5.8% (1991–1999)].

Carcinoid Tumors Without the Carcinoid Syndrome

The age of patients at diagnosis ranges from 10 to 93 years, with a mean age of 63 years for the small intestine and 66 years for the rectum. The presentation is diverse and is related to the site of origin and the extent of malignant spread. In the appendix, carcinoid tumors usually are found incidentally during surgery for suspected appendicitis. Small-intestinal carcinoids in the jejunoileum present with periodic abdominal pain (51%), intestinal obstruction with ileus/invagination (31%), an abdominal tumor (17%), or GI bleeding (11%). Because of the vagueness of the symptoms, the diagnosis usually is delayed approximately 2 years from onset of the symptoms, with a range up to 20 years. Duodenal, gastric, and rectal carcinoids are most frequently found by chance at endoscopy. The most common symptoms of rectal carcinoids are melena/bleeding (39%), constipation (17%), and diarrhea (12%). Bronchial carcinoids frequently are discovered as a lesion on a chest radiograph, and 31% of the patients are asymptomatic. Thymic carcinoids present as anterior mediastinal masses, usually on chest radiograph or CT scan. Ovarian and testicular carcinoids usually present as masses discovered on physical examination or ultrasound. Metastatic carcinoid tumor in the liver frequently presents as hepatomegaly in a patient who may have minimal symptoms and nearly normal liver function test results.

Carcinoid Tumors with Systemic Symptoms Due to Secreted Products

Carcinoid tumors immunocytochemically can contain numerous GI peptides: gastrin, insulin, somatostatin, motilin, neurotensin, tachykinins (substance K, substance P, neuropeptide K), glucagon, gastrin-releasing peptide, vasoactive intestinal peptide (VIP), pancreatic polypeptide (PP), ghrelin, other biologically active peptides (ACTH, calcitonin, growth hormone), prostaglandins, and bioactive amines (serotonin). These substances may or may not be released in sufficient amounts to cause symptoms. In various studies of patients with carcinoid tumors, elevated serum levels of PP were found in 43%, motilin in 14%, gastrin in 15%, and VIP in 6%. Foregut carcinoids are more likely to produce various GI peptides than are midgut carcinoids. Ectopic ACTH production causing Cushing's syndrome is seen increasingly with foregut carcinoids (respiratory tract primarily) and in some series has been the most common cause of the ectopic ACTH syndrome, accounting for 64% of all cases. Acromegaly due to growth hormone–releasing factor release occurs with foregut carcinoids, as does the somatostatinoma syndrome, but rarely occurs with duodenal carcinoids. The most common systemic syndrome with carcinoid tumors is the carcinoid syndrome, which is discussed in detail in the next section.

Carcinoid Syndrome

CLINICAL FEATURES

The cardinal features from a number of series at presentation as well as during the disease course are shown in Table 350-6. Flushing and diarrhea are the two most common symptoms, occurring in up to 73% initially and in up to 89% during the course of the disease. The characteristic flush is of sudden onset; it is a deep red or violaceous erythema of the upper body, especially the neck and face, often associated with a feeling of warmth and occasionally associated with pruritus, lacrimation, diarrhea, or facial edema. Flushes may be precipitated by stress; alcohol; exercise; certain foods, such as cheese; or certain agents, such as catecholamines, pentagastrin, and serotonin reuptake inhibitors. Flushing episodes may be brief, lasting 2 to 5 min, especially initially, or may last hours, especially later in the disease course. Flushing usually is associated with metastastic midgut carcinoids but can also occur with foregut carcinoids. With bronchial carcinoids the flushes frequently are prolonged for hours to days, reddish in color, and associated with salivation, lacrimation, diaphoresis, diarrhea, and hypotension. The flush associated with gastric carcinoids can also be reddish in color, but with a patchy distribution over the face and neck, although the classic flush seen with midgut carcinoids can also be seen with gastric carcinoids. It may be provoked by food and have accompanying pruritus.

Diarrhea is present in 32–73% initially and 68–84% at some time in the disease course. Diarrhea usually occurs with flushing (85% of cases). The diarrhea usually is described as watery, with 60% of patients having <1 L/d of diarrhea. Steatorrhea is present in 67%, and in 46% it is greater than 15 g/d (normal <7 g). Abdominal pain may be present with the diarrhea or independently in 10–34% of cases.

Cardiac manifestations occur in 11–20% initially of patients with carcinoid syndrome and in 17–56% (mean 40%) at some time in the disease course. The cardiac disease is due to the formation of fibrotic plaques (composed of smooth-muscle cells, myofibroblasts, and elastic tissue) involving the endocardium, primarily on the right side, although lesions on the left side also occur occasionally, especially if a patent foramen ovale exists. The dense fibrous deposits are most commonly on the ventricular aspect of the tricuspid valve and less commonly on the pulmonary valve cusps. They can result in constriction of the valves, and pulmonic stenosis is usually predominant, whereas the tricuspid valve is often fixed open, resulting in regurgitation predominating. Overall, in patients with carcinoid heart disease, 97% have tricuspid insufficiency, 59% tricuspid stenosis, 50% pulmonary insufficiency, 25% pulmonary stenosis, and 11% (0–25%) left-side lesions. Up to 80% of patients with cardiac lesions develop heart failure. Lesions on the left side are much less extensive, occur in 30%

at autopsy, and most frequently affect the mitral valve.

Other clinical manifestations include wheezing or asthma-like symptoms (8–18%) and pellagra-like skin lesions (2–25%). A variety of noncardiac problems due to increased fibrous tissue have been reported, including retroperitoneal fibrosis causing urethral obstruction, Peyronie's disease of the penis, intraabdominal fibrosis, and occlusion of the mesenteric arteries or veins.

PATHOBIOLOGY

Carcinoid syndrome occurred in 8% of 8876 patients with carcinoid tumors, with a rate of 1.4–18.4% in different studies. It occurs only when sufficient concentrations of products secreted by the tumor reach the systemic circulation. In 91% of cases this occurs after distant metastases to the liver. Rarely, primary gut carcinoids with nodal metastases with extensive retroperitoneal invasion, pancreatic carcinoids with retroperitoneal lymph nodes, or carcinoids of the lung or ovary with direct access to the systemic circulation can cause the carcinoid syndrome without hepatic metastases. All carcinoid tumors do not have the same propensity to metastasize and cause the carcinoid syndrome (Table 350-3). Midgut carcinoids account for 60–67% of cases of carcinoid syndrome, foregut tumors for 2–33%, hindgut for 1–8%, and an unknown primary location for 2–15%.

One of the main secretory products of carcinoid tumors involved in the carcinoid syndrome is serotonin [5-hydroxytryptamine (5-HT)] (Fig. 350-1), which is synthesized from tryptophan. Up to 50% of dietary tryptophan can be used in this synthetic pathway by tumor cells, and this can result in inadequate supplies for conversion to niacin; hence, some patients (2.5%) develop pellagra-like lesions. Serotonin has numerous biologic effects, including stimulating intestinal secretion with inhibition of absorption, stimulating increases in intestinal motility, and stimulating fibrogenesis. In various studies 56–88% of all carcinoid tumors were associated with serotonin overproduction; however, 12–26% of the patients did not have the carcinoid syndrome. In one study platelet serotonin was elevated in 96% of patients with midgut carcinoids, 43% with foregut tumors, and 0% with hindgut tumors. In 90–100% of patients with the carcinoid syndrome there is evidence of serotonin overproduction. Serotonin is thought to be predominantly responsible for the diarrhea because of its effects on gut motility and intestinal secretion, primarily through 5-HT_3 and, to a lesser degree, 5-HT_4 receptors. Serotonin receptor antagonists (especially 5-HT_3 antagonists) relieve the diarrhea in many, but not all, patients. Additional studies suggest that prostaglandin E2 (PGE2) and tachykinins may be important mediators of the diarrhea in some patients. In one study, plasma tachykinin levels correlated with symptoms of both flushing and diarrhea. Serotonin does not appear to be involved in the flushing because serotonin receptor antagonists do not relieve flushing. In patients with gastric carcinoids the characteristic red, patchy pruritic flush probably is due to histamine release because H_1 and H_2 receptor antagonists can prevent it. Numerous studies have shown that tachykinins are stored in carcinoid tumors and released during flushing. However, some studies have demonstrated that octreotide can relieve the flushing induced by pentagastrin in these patients without altering the stimulated increase in plasma substance P, suggesting that other mediators must be involved in the flushing. A correlation between plasma tachykinin levels, but not substance P levels, and flushing has been reported. Both histamine and serotonin may be responsible for the wheezing as well as the fibrotic reactions involving the heart, causing Peyronie's disease and intraabdominal fibrosis. The exact mechanism of the heart disease has remained unclear, although increasing evidence supports a central role for serotonin. The valvular heart disease caused by the appetite-suppressant drug dexfenfluramine is histologically indistinguishable from that observed in carcinoid disease. Furthermore, ergot-containing dopamine receptor agonists used for Parkinson's disease (pergolide, cabergoline) cause valvular heart disease that closely resembles that seen in the carcinoid syndrome. Metabolites of fenfluramine, as well as the dopamine receptor agonists, have high affinity for serotonin receptor subtype 5-HT_{2B} receptors, whose activation is known to cause fibroblast mitogenesis. Serotonin receptor subtypes 5-$HT_{1B,1D,2A,2B}$ normally are expressed in human heart valve interstitial cells. High levels of 5-HT_{2B} receptors are known to occur in heart valves and occur in cardiac fibroblasts and cardiomyocytes. Studies of cultured interstitial cells from human cardiac valves have demonstrated that these valvulopathic drugs induce mitogenesis by activating 5-HT_{2B} receptors and stimulating upregulation of transforming growth factor β and collagen biosynthesis. These observations support the conclusion that serotonin overproduction by carcinoid tumors is important in mediating the valvular changes, possibly by activating 5-HT_{2B} receptors in the endocardium. Both the magnitude of serotonin overproduction and prior chemotherapy are important predictors of progression of the heart disease. Atrial natriuretic peptide (ANP) overproduction also has been reported in patients with cardiac disease, but its role in the pathogenesis is unknown. However, high plasma levels of ANP have a worse prognosis. Plasma connective tissue growth factor levels are elevated in many fibrotic conditions; elevated levels occur in patients with carcinoid heart disease and correlate with the presence of right-ventricular dysfunction and the extent of valvular regurgitation in patients with carcinoid tumors.

Figure 350-1

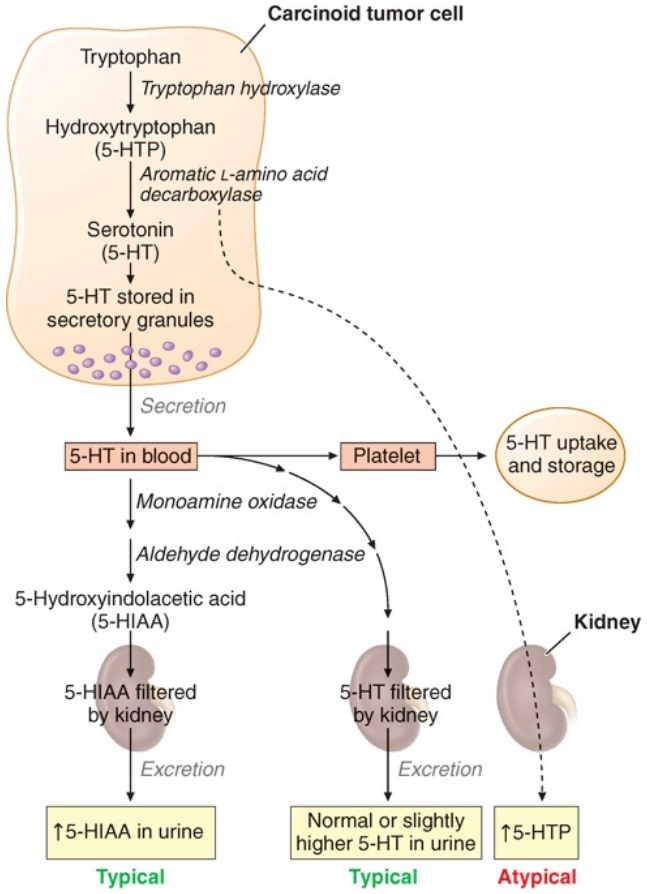

Synthesis, secretion, and metabolism of serotonin (5-HT) in patients with typical and atypical carcinoid syndromes. 5-HIAA, 5-hydroxyindolacetic acid.

Patients may develop either a typical or, rarely, an atypical carcinoid syndrome. In patients with the typical form, which characteristically is caused by a midgut carcinoid tumor, the conversion of tryptophan to 5-HTP is the rate-limiting step (Fig. 350-1). Once 5-HTP is formed, it is rapidly converted to 5-HT and stored in secretory granules of the tumor or in platelets. A small amount remains in plasma and is converted to 5-HIAA, which appears in large amounts in the urine. These patients have an expanded serotonin pool size, increased blood and platelet serotonin, and increased urinary 5-hydroxyindolacetic acid (5-HIAA). Some carcinoid tumors cause an atypical carcinoid syndrome that is thought to be due to a deficiency in the enzyme dopa decarboxylase; thus, 5-HTP cannot be converted to 5-HT (serotonin), and 5-HTP is secreted into the bloodstream (Fig. 350-1). In these patients, plasma serotonin levels are normal but urinary levels may be increased because some 5-HTP is converted to 5-HT in the kidney. Characteristically, urinary 5-HTP and 5-HT are increased, but urinary 5-HIAA levels are only slightly elevated. Foregut carcinoids are the most likely to cause an atypical carcinoid syndrome.

One of the most immediate life-threatening complications of the carcinoid syndrome is the development of a carcinoid crisis. This is more common in patients who have intense symptoms or have greatly increased urinary 5-HIAA levels (i.e., >200 mg/d). The crises may occur spontaneously or be provoked by stress, anesthesia, chemotherapy, or a biopsy. Patients develop intense flushing, diarrhea, abdominal pain, cardiac abnormalities including tachycardia, hypertension, or hypotension. If not adequately treated, this can be a terminal event.

Diagnosis of the Carcinoid Syndrome and Carcinoid Tumors

The diagnosis of carcinoid syndrome relies on measurement of urinary or plasma serotonin or its metabolites in the urine. The measurement of 5-HIAA is used most frequently. False-positive elevations may occur if the patient is eating serotonin-rich foods such as bananas, pineapples, walnuts, pecans, avocados, or hickory nuts or is taking certain medications (cough syrup containing guaifenesin, acetaminophen, salicylates, serotonin reuptake inhibitors, or l-dopa). The normal range in daily urinary 5-HIAA excretion is 2–8 mg/d. Serotonin overproduction was noted in 92% of patients with carcinoid syndrome in one study, and in another study, 5-HIAA had 73% sensitivity and 100% specificity for carcinoid syndrome.

Most physicians use only the urinary 5-HIAA excretion rate; however, plasma and platelet serotonin levels, if available, may provide additional information. Platelet serotonin levels are more sensitive than urinary 5-HIAA but are not generally available. Because patients with foregut carcinoids may produce an atypical carcinoid syndrome, if this syndrome is suspected and the urinary 5-HIAA is minimally elevated or normal, other urinary metabolites of tryptophan, such as 5-HTP and 5-HT, should be measured (Fig. 350-1).

Flushing occurs in a number of other diseases, including systemic mastocytosis, chronic myeloid leukemia with increased histamine release, menopause, reactions to alcohol or glutamate, side effects of chlorpropamide, calcium channel blockers, and nicotinic acid. None of these conditions cause increased urinary 5-HIAA.

The diagnosis of carcinoid tumor can be suggested by the carcinoid syndrome, recurrent abdominal symptoms in a healthy-appearing individual, or the discovery of hepatomegaly or hepatic metastases associated with minimal symptoms. Ileal carcinoids, which make up 25% of all clinically detected carcinoids, should be suspected in patients with bowel obstruction, abdominal pain, flushing, or diarrhea.

Serum chromogranin A levels are elevated in 56–100% of patients with carcinoid tumors, and the level correlates with tumor bulk. Serum chromogranin A levels are not specific for carcinoid tumors because they are also elevated in patients with PETs and other neuroendocrine tumors. Plasma neuron-specific enolase levels are also used as a marker of carcinoid tumors but are less sensitive than chromogranin A, being increased in only 17–47% of patients.

Treatment: Carcinoid Syndrome and Nonmetastatic Carcinoid Tumors

CARCINOID SYNDROME

Treatment includes avoiding conditions that precipitate flushing, dietary supplementation with nicotinamide, treatment of heart failure with diuretics, treatment of wheezing with oral bronchodilators, and control of the diarrhea with antidiarrheal agents such as loperamide and diphenoxylate. If patients still have symptoms, serotonin receptor antagonists or somatostatin analogues (Fig. 350-2) are the drugs of choice.

Figure 350-2

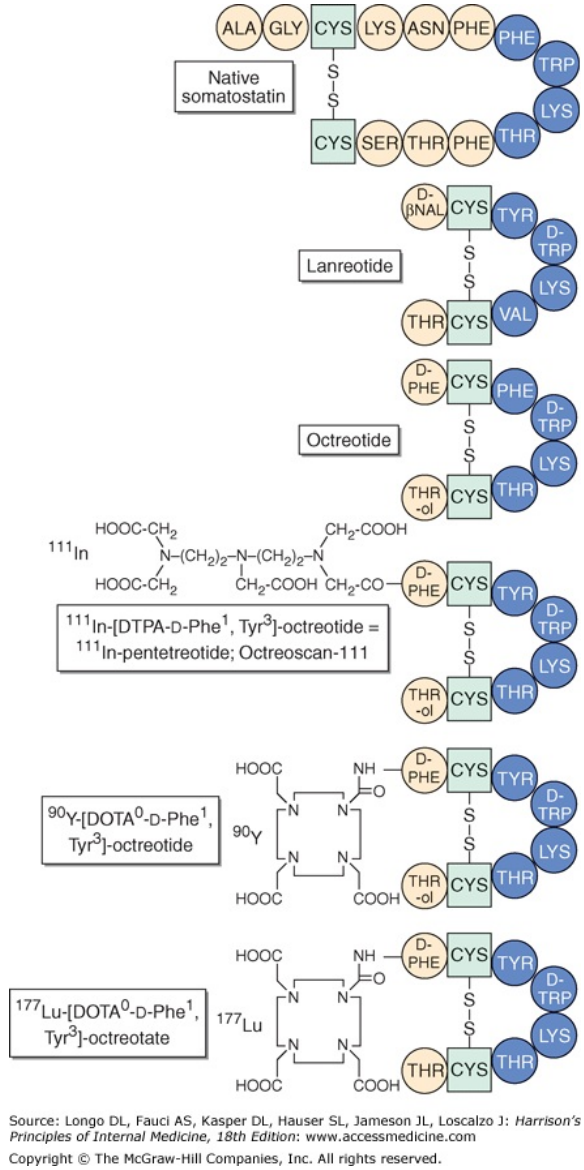

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Structure of somatostatin and synthetic analogues used for diagnostic or therapeutic indications.

There are 14 subclasses of serotonin receptors, and antagonists for many are not available. The 5-HT$_1$ and 5-HT$_2$ receptor antagonists methylsergide, cyproheptadine, and ketanserin have all been used to control the diarrhea but usually do not decrease flushing. The use of methylsergide is limited because it can cause or enhance retroperitoneal fibrosis. Ketanserin diminishes diarrhea in 30–100% of patients. 5-HT$_3$ receptor antagonists (ondansetron, tropisetron, alosetron) can control diarrhea and nausea in up to 100% of patients and occasionally ameliorate the flushing. A combination of histamine H$_1$ and H$_2$ receptor antagonists (i.e., diphenhydramine and cimetidine or ranitidine) may control flushing in patients with foregut carcinoids.

Synthetic analogues of somatostatin (octreotide, lanreotide) are now the most widely used agents to control the symptoms of patients with carcinoid syndrome (Fig. 350-2). These drugs are effective at relieving symptoms and decreasing urinary 5-HIAA levels in patients with this syndrome. Octreotide-LAR and lanreotide-SR/autogel (Somatuline) control symptoms in 74% and 68%, respectively, of patients with carcinoid syndrome and show a biochemical response in 51% and 39%. Patients with mild to moderate symptoms usually are treated initially with octreotide 100 μg SC every 8 h and begun on long-acting monthly depot forms (octreotide-LAR or lanreotide-autogel). Forty percent of patients escape control after a median time of 4 months, and the depot dosage may have to be increased as well as supplemented with the shorter-acting formulation, SC octreotide.

Carcinoid heart disease is associated with a decreased mean survival (3.8 years), and therefore it should be sought for and carefully assessed in all patients with carcinoid syndrome. Transthoracic echocardiography remains a key element in establishing the diagnosis of carcinoid heart disease and determining the extent and type of cardiac abnormalities. Treatment with diuretics and somatostatin analogues can reduce the negative hemodynamic effects and secondary heart failure. It remains unclear whether long-term treatment with these drugs will decrease the progression of carcinoid heart disease. Balloon valvuloplasty for stenotic valves or cardiac valve surgery may be required.

In patients with carcinoid crises, somatostatin analogues are effective at both treating the condition and preventing their development during known precipitating events such as surgery, anesthesia, chemotherapy, and stress. It is recommended that octreotide 150–250 μg SC every 6 to 8 h be used 24–48 h before anesthesia and then continued throughout the procedure.

Currently, sustained-release preparations of both octreotide [octreotide-LAR (long-acting release), 10, 20, 30 mg] and lanreotide [lanreotide-PR (prolonged release, lanreotide-autogel), 60, 90, 120 mg] are available and widely used because their use greatly facilitates long-term treatment. Octreotide-LAR (30 mg/month) gives a plasma level ≥1 ng/mL for 25 days, whereas this requires three to six injections a day of the non-sustained-release form. Lanreotide autogel (Somatuline) is given every 4–6 weeks.

Short-term side effects occur in up to one-half of patients. Pain at the injection site and side effects related to the GI tract (59% discomfort, 15% nausea, diarrhea) are the most common. They are usually short-lived and do not interrupt treatment. Important long-term side effects include gallstone formation, steatorrhea, and deterioration in glucose tolerance. The overall incidence of gallstones/biliary sludge in one study was 52%, with 7% having symptomatic disease that required surgical treatment.

Interferon α is reported to be effective in controlling symptoms of the carcinoid syndrome either alone or combined with hepatic artery embolization. With interferon α alone the response rate is 42%, and with interferon α with hepatic artery embolization, diarrhea was controlled for 1 year in 43% and flushing was controlled in 86%.

Hepatic artery embolization alone or with chemotherapy (chemoembolization) has been used to control the symptoms of carcinoid syndrome. Embolization alone is reported to control symptoms in up to 76% of patients, and chemoembolization (5-fluorouracil, doxorubicin, cisplatin, mitomycin) in 60–75% of patients. Hepatic artery embolization can have major side effects, including nausea, vomiting, pain, and fever. In two studies 5–7% of patients died from complications of hepatic artery occlusion.

Other drugs have been used successfully in small numbers of patients to control the symptoms of carcinoid syndrome. Parachlorophenylanine can inhibit tryptophan hydroxylase and therefore the conversion of tryptophan to 5-HTP. However, its severe side effects, including psychiatric disturbances, make it intolerable for long-term use. α-Methyldopa inhibits the conversion of 5-HTP to 5-HT, but its effects are only partial.

Peptide radioreceptor therapy (using radiotherapy with radiolabeled somatostatin analogues), the use of radiolabeled microspheres, and other methods for treatment of advanced metastatic disease may facilitate control of the carcinoid syndrome and are discussed in a later section dealing with treatment of advanced disease.

CARCINOID TUMORS (NONMETASTATIC)

Surgery is the only potentially curative therapy. Because with most carcinoids the probability of metastatic disease increases with increasing size, the extent of surgical resection is determined accordingly. With appendiceal carcinoids <1 cm, simple appendectomy was curative in 103 patients followed for up to 35 years. With rectal carcinoids <1 cm, local resection is curative. With small-intestinal carcinoids <1 cm, there is not complete agreement. Because 15–69% of small-intestinal carcinoids this size have metastases in different studies, some recommend a wide resection with en bloc resection of the adjacent lymph-bearing mesentery. If the carcinoid tumor is >2 cm for rectal, appendiceal, or small-intestinal carcinomas, a full cancer operation should be done. This includes a right hemicolectomy for appendiceal carcinoid, an abdominoperineal resection or low anterior resection for rectal carcinoids, and an en bloc resection of adjacent lymph nodes for small-intestinal carcinoids. For carcinoids 1–2 cm in diameter for appendiceal tumors, a simple appendectomy is proposed by some, whereas others favor a formal right hemicolectomy. For rectal carcinoids 1–2 cm, it is recommended that a wide local full-thickness excision be performed.

With type I or II gastric carcinoids, which are usually <1 cm, endoscopic removal is recommended. In type I or II gastric carcinoids, if the tumor is >2 cm or if there is local invasion, some recommend total gastrectomy, whereas others recommend antrectomy in type I to reduce the hypergastrinemia, which led to regression of the carcinoids in a number of studies. For types I and II gastric carcinoids of 1–2 cm, there is no agreement, with some recommending endoscopic treatment followed by chronic somatostatin treatment and careful follow-up and others recommending surgical treatment. With type III gastric carcinoids >2 cm, excision and regional lymph node clearance are recommended. Most tumors <1 cm are treated endoscopically.

Resection of isolated or limited hepatic metastases may be beneficial and will be discussed in a later section on treatment of advanced disease.

PANCREATIC ENDOCRINE TUMORS

Functional PETs usually present clinically with symptoms due to the hormone-excess state. Only late in the course of the disease does the tumor per se cause prominent symptoms such as abdominal pain. In contrast, all the symptoms due to nonfunctional PETs are due to the tumor per se. The overall result of this is that some functional PETs may present with severe symptoms with a small or undetectable primary tumor, whereas nonfunctional tumors usually present late in the disease course with large tumors, which are frequently metastatic. The mean delay between onset of continuous symptoms and diagnosis of a functional PET syndrome is 4–7 years. Therefore, the diagnoses frequently are missed for extended periods.

Treatment: Pancreatic Endocrine Tumor

Treatment of PETs requires two different strategies. First, treatment must be directed at the hormone-excess state such as the gastric acid hypersecretion in gastrinomas or the hypoglycemia in insulinomas. Ectopic hormone secretion usually causes the presenting symptoms and can cause life-threatening complications. Second, with all the tumors except insulinomas, >50% are malignant (Table 350-2); therefore, treatment must also be directed against the tumor per se. Because in many patients these tumors are not surgically curable due to the presence of advanced disease at diagnosis, surgical resection for cure, which addresses both treatment aspects, is often not possible.

Gastrinoma (Zollinger-Ellison Syndrome) (Zes)

A gastrinoma is a neuroendocrine tumor that secretes gastrin; the resultant hypergastrinemia causes gastric acid hypersecretion (Zollinger-Ellison syndrome). The chronic hypergastrinemia results in marked gastric acid hypersecretion and growth of the gastric mucosa with increased numbers of parietal cells and proliferation of gastric ECL cells. The gastric acid hypersecretion characteristically causes peptic ulcer disease, often refractory and severe, as well as diarrhea. The most common presenting symptoms are abdominal pain (70–100%), diarrhea (37–73%), and gastroesophageal reflux disease (GERD) (30–35%); 10–20% have diarrhea only. Although peptic ulcers may occur in unusual locations, most patients have a typical duodenal ulcer. Important observations that should suggest this diagnosis include peptic ulcer disease (PUD); with diarrhea; PUD in an unusual location or with multiple ulcers; PUD refractory to treatment or persistent; PUD associated with prominent gastric folds; PUD associated with findings suggestive of MEN 1 (endocrinopathy, family history of ulcer or endocrinopathy, nephrolithiases); and PUD without *Helicobacter pylori* present. *H. pylori* is present in >90% of idiopathic peptic ulcers but is present in <50% of patients with gastrinomas. Chronic unexplained diarrhea also should suggest gastrinoma.

Approximately 20–25% of patients with ZES have MEN 1, and in most cases hyperparathyroidism is present before the gastrinoma. These patients are treated differently from those without MEN 1; therefore, MEN 1 should be sought in all patients by family history and by measuring plasma ionized calcium and prolactin levels and plasma hormone levels (parathormone, growth hormone).

Most gastrinomas (50–70%) are present in the duodenum, followed by the pancreas (20–40%) and other intraabdominal sites (mesentery, lymph nodes, biliary tract, liver, stomach, ovary). Rarely, the tumor may involve extraabdominal sites. In MEN 1 the gastrinomas are also usually in the duodenum (70–90%), followed by the pancreas (10–30%), and are almost always multiple. About 60–90% of gastrinomas are malignant (Table 350-2) with metastatic spread to lymph nodes and liver. Distant metastases to bone occur in 12–30% of patients with liver metastases.

DIAGNOSIS

The diagnosis of ZES requires the demonstration of inappropriate fasting hypergastrinemia, usually by demonstrating hypergastrinemia occurring with an increased basal gastric acid output (BAO) (hyperchlorhydria). More than 98% of patients with gastrinomas have fasting hypergastrinemia, although in 40–60% the level may be elevated less than tenfold. Therefore, when the diagnosis is suspected, a fasting gastrin should be determined first. It is important to remember that potent gastric acid suppressant drugs such as proton pump inhibitors (omeprazole, esomeprazole, pantoprazole, lansoprazole, rabeprazole) can suppress acid secretion sufficiently to cause hypergastrinemia; because of their prolonged duration of action, these drugs have to be discontinued for a week before the gastrin determination. Withdrawal of proton pump inhibitors (PPIs) should be performed carefully and is best done in consultation with GI units with experience in this area. The widespread use of PPIs can confound the diagnosis of ZES by raising a false-positive diagnosis by causing hypergastrinemia in a patient being treated with idiopathic peptic disease (without ZES) and lead to a false-negative diagnosis because at routine doses used to treat patients with idiopathic peptic disease, PPIs control symptoms in most ZES patients and thus mask the diagnosis. If ZES is suspected and the gastrin level is elevated, it is important to show that it is increased when gastric pH is ≤2.0 because physiologically hypergastrinemia secondary to achlorhydria (atrophic gastritis, pernicious anemia) is one of the most common causes of hypergastrinemia. Nearly all gastrinoma patients have a fasting pH ≤2 when off antisecretory drugs. If the fasting gastrin is >1000 pg/mL (increased tenfold) and the pH is ≤2.0, which occurs in 40–60% of patients with gastrinoma, the diagnosis of ZES is established after the possibility of retained antrum syndrome has been ruled out by history. In patients with hypergastrinemia with fasting gastrins <1000 pg/mL and gastric pH ≤2.0, other conditions, such as *H. pylori* infections, antral G-cell hyperplasia/hyperfunction, gastric outlet obstruction, and, rarely, renal failure, can masquerade as ZES. To establish the diagnosis in this group, a determination of BAO and a secretin provocative test should be done. In patients with ZES without previous gastric acid–reducing surgery, the BAO is usually (>90%) elevated (i.e., >15 meq/h). The secretin provocative test is usually positive, with the criterion of a >120-pg/mL increase over the basal level having the highest sensitivity (94%) and specificity (100%).

Treatment: Gastrinomas

Gastric acid hypersecretion in patients with gastrinomas can be controlled in almost every case by oral gastric antisecretory drugs. Because of their long duration of action and potency, which allows dosing once or twice a day, the PPIs (H^+,K^+-ATPase inhibitors) are the drugs of choice. Histamine H_2-receptor antagonists are also effective, although more frequent dosing (q 4–8 h) and high doses are required. In patients with MEN 1 with hyperparathyroidism, correction of the hyperparathyroidism increases the sensitivity to gastric antisecretory drugs and decreases the basal acid output. Long-term treatment with PPIs (>15 years) has proved to be safe and effective, without development of tachyphylaxis. Although patients with ZES, especially those with MEN 1, more frequently develop gastric carcinoids, no data suggest that the long-term use of PPIs increases this risk in these patients. With long-term PPI use in ZES patients, vitamin B12 deficiency can develop; thus, vitamin B_{12} levels should be assessed during follow-up.

With the increased ability to control acid hypersecretion, more than 50% of patients who are not cured (>60% of patients) will die from tumor-related causes. At presentation, careful imaging studies are essential to localize the extent of the tumor. A third of patients present with hepatic metastases, and in <15% of those patients the disease is limited, so that surgical resection may be possible. Surgical short-term cure is possible in 60% of all patients without MEN 1 or liver metastases (40% of all patients) and in 30% of patients long-term. In patients with MEN 1, long-term surgical cure is rare because the tumors are multiple, frequently with lymph node metastases. Therefore, all patients with gastrinomas without MEN 1 or a medical condition that limits life expectancy should undergo surgery by a surgeon experienced in the treatment of these disorders.

Insulinomas

An insulinoma is an endocrine tumor of the pancreas that is thought to be derived from beta cells that ectopically secrete insulin, which results in hypoglycemia. The average age of occurrence is 40–50 years old. The most common clinical symptoms are due to the effect of the hypoglycemia on the CNS (neuroglycemic symptoms) and include confusion, headache, disorientation, visual difficulties, irrational behavior, and even coma. Also, most patients have symptoms due to excess catecholamine release secondary to the hypoglycemia, including sweating, tremor, and palpitations. Characteristically, these attacks are associated with fasting.

Insulinomas are generally small (>90% <2 cm) and usually not multiple (90%); only 5–15% are malignant, and they almost invariably occur only in the pancreas, distributed equally in the pancreatic head, body, and tail.

Insulinomas should be suspected in all patients with hypoglycemia, especially when there is a history suggesting that attacks are provoked by fasting, or with a family history of MEN 1. Insulin is synthesized as proinsulin, which consists of a

21-amino-acid α chain and a 30-amino-acid β chain connected by a 33-amino-acid connecting peptide (C peptide). In insulinomas, in addition to elevated plasma insulin levels, elevated plasma proinsulin levels are found, and C-peptide levels can be elevated.

DIAGNOSIS

The diagnosis of insulinoma requires the demonstration of an elevated plasma insulin level at the time of hypoglycemia. A number of other conditions may cause fasting hypoglycemia, such as the inadvertent or surreptitious use of insulin or oral hypoglycemic agents, severe liver disease, alcoholism, poor nutrition, and other extrapancreatic tumors. Furthermore, postprandial hypoglycemia can be caused by a number of conditions that confuse the diagnosis of insulinoma. Particularly important here is the increased occurrence of hypoglycemia after gastric bypass surgery for obesity, which is now widely performed. The most reliable test to diagnose insulinoma is a fast up to 72 h with serum glucose, C-peptide, proinsulin, and insulin measurements every 4–8 h. If at any point the patient becomes symptomatic or glucose levels are persistently below <2.2 mmol/L (40 mg/dL), the test should be terminated and repeat samples for the above studies should be obtained before glucose is given. Some 70–80% of patients will develop hypoglycemia during the first 24 h, and 98% by 48 h. In nonobese normal subjects, serum insulin levels should decrease to <43 pmol/L (<6 μU/mL) when blood glucose decreases to <2.2 mmol/L (<40 mg/dL) and the ratio of insulin to glucose is <0.3 (in mg/dL). In addition to having an insulin level >6 μU/mL when blood glucose is <40 mg/dL, some investigators also require an elevated C-peptide and serum proinsulin level, an insulin/glucose ratio >0.3, and a decreased plasma β-hydroxybutyrate level for the diagnosis of insulinomas. Surreptitious use of insulin or hypoglycemic agents may be difficult to distinguish from insulinomas. The combination of proinsulin levels (normal in exogenous insulin/hypoglycemic agent users), C-peptide levels (low in exogenous insulin users), antibodies to insulin (positive in exogenous insulin users), and measurement of sulfonylurea levels in serum or plasma will allow the correct diagnosis to be made. The diagnosis of insulinoma has been complicated by the introduction of specific insulin assays that do not also interact with proinsulin, as do many of the older radioimmunoassays (RIAs), and therefore give lower plasma insulin levels. The increased use of these specific insulin assays has resulted in increased numbers of patients with insulinomas having lower plasma insulin values than the 6 μU/mL levels proposed to be characteristic of insulinomas by RIA. In these patients the assessment of proinsulin and C-peptide levels at the time of hypoglycemia is particularly helpful for establishing the correct diagnosis. An elevated proinsulin level when the fasting glucose level is <45 mg/dL is sensitive and specific.

Treatment: Insulinomas

Only 5–15% of insulinomas are malignant; therefore, after appropriate imaging (see below), surgery should be performed. In different studies, 75–100% of patients are cured by surgery. Before surgery, the hypoglycemia can be controlled by frequent small meals and the use of diazoxide (150–800 mg/d). Diazoxide is a benzothiadiazide whose hyperglycemic effect is attributed to inhibition of insulin release. Its side effects are sodium retention and GI symptoms such as nausea. Approximately 50–60% of patients respond to diazoxide. Other agents effective in some patients to control the hypoglycemia include verapamil and diphenylhydantoin. Long-acting somatostatin analogues such as octreotide and lanreotide are acutely effective in 40% of patients. However, octreotide must be used with care because it inhibits growth hormone secretion and can alter plasma glucagon levels; therefore, in some patients it can worsen the hypoglycemia.

For the 5–15% of patients with malignant insulinomas, these drugs or somatostatin analogues are used initially. In a small number of patients with insulinomas, some with malignant tumors, mammalian target of rapamycin (mTor) inhibitors (everolimus, rapamycin) are reported to control the hypoglycemia. If they are not effective, various antitumor treatments such as hepatic arterial embolization, chemoembolization, chemotherapy, and peptide receptor radiotherapy have been used (see below).

Insulinomas, which are usually benign (>90%) and intrapancreatic in location, are increasingly resected using a laparoscopic approach, which has lower morbidity rates. This approach requires that the insulinoma be localized on preoperative imaging studies.

Glucagonomas

A glucagonoma is an endocrine tumor of the pancreas that secretes excessive amounts of glucagon, which causes a distinct syndrome characterized by dermatitis, glucose intolerance or diabetes, and weight loss. Glucagonomas principally occur between 45 and 70 years of age. The tumor is clinically heralded by a characteristic dermatitis (migratory necrolytic erythema) (67–90%), accompanied by glucose intolerance (40–90%), weight loss (66–96%), anemia (33–85%), diarrhea (15–29%), and thromboembolism (11–24%). The characteristic rash usually starts as an annular erythema at intertriginous and periorificial sites, especially in the groin or buttock. It subsequently becomes raised, and bullae form; when the bullae rupture, eroded areas form. The lesions can wax and wane. The development of a similar rash in patients receiving glucagon therapy suggests that the rash is a direct effect of the hyperglucagonemia. A characteristic laboratory finding is hypoaminoacidemia, which occurs in 26–100% of patients.

Glucagonomas are generally large tumors at diagnosis (5–10 cm). Some 50–80% occur in the pancreatic tail. From 50 to 82% have evidence of metastatic spread at presentation, usually to the liver. Glucagonomas are rarely extrapancreatic and usually occur singly.

DIAGNOSIS

The diagnosis is confirmed by demonstrating an increased plasma glucagon level. Characteristically, plasma glucagon levels exceed 1000 pg/mL (normal is <150 pg/mL) in 90%; 7% are between 500 and 1000 pg/mL, and 3% are <500 pg/mL. A trend toward lower levels at diagnosis has been noted in the last decade. A plasma glucagon level >1000 pg/mL is considered diagnostic of glucagonoma. Other diseases causing increased plasma glucagon levels include renal insufficiency, acute pancreatitis, hypercorticism, hepatic insufficiency, severe stress, and prolonged fasting or familial hyperglucagonemia, as well as danazol treatment. With the exception of cirrhosis, these disorders do not increase plasma glucagon >500 pg/mL.

Necrolytic migratory erythema is not pathognomonic for glucagonoma and occurs in myeloproliferative disorders, hepatitis B infection, malnutrition, short-bowel syndrome, inflammatory bowel disease, and malabsorption disorders.

Treatment: Glucagonomas

In 50–80% of patients, hepatic metastases are present, and so curative surgical resection is not possible. Surgical debulking in patients with advanced disease or other antitumor treatments may be beneficial (see below). Long-acting somatostatin analogues such as octreotide and lanreotide improve the skin rash in 75% of patients and may improve the weight loss, pain, and diarrhea but usually do not improve the glucose intolerance.

Somatostatinoma Syndrome

The somatostatinoma syndrome is due to an NET that secretes excessive amounts of somatostatin, which causes a distinct syndrome characterized by diabetes mellitus, gallbladder disease, diarrhea, and steatorrhea. There is no general distinction in the literature between a tumor that contains somatostatin-like immunoreactivity (somatostatinoma) and does (11–45%) or does not (55–90%) produce a clinical syndrome (somatostatinoma syndrome) by secreting somatostatin. In a review of 173 cases of somatostatinomas, only 11% were associated with the somatostatinoma syndrome. The mean age is 51 years. Somatostatinomas occur primarily in the pancreas and small intestine, and the frequency of the symptoms and occurrence of the somatostatinoma syndrome differ in each. Each of the usual symptoms is more common in pancreatic than in intestinal somatostatinomas: diabetes mellitus (95% vs. 21%), gallbladder disease (94% vs. 43%), diarrhea (92% vs. 38%), steatorrhea (83% vs. 12%), hypochlorhydria (86% vs. 12%), and weight loss (90% vs. 69%). The somatostatinoma syndrome occurs in 30–90% of pancreatic and 0–5% of small-intestinal somatostatinomas. In various series 43% of all duodenal NETs contain somatostatin; however, the somatostatinoma syndrome is rarely present (<2%). Somatostatinomas occur in the pancreas in 56–74% of cases, with the primary location being the pancreatic head. The tumors are usually solitary (90%) and large (mean size 4.5 cm). Liver metastases are common, being present in 69–84% of patients. Somatostatinomas are rare in patients with MEN 1, occurring in only 0.65%.

Somatostatin is a tetradecapeptide that is widely distributed in the CNS and GI tract, where it functions as a neurotransmitter or has paracrine and autocrine actions. It is a potent inhibitor of many processes, including release of almost all hormones, acid secretion, intestinal and pancreatic secretion, and intestinal absorption. Most of the clinical manifestations are directly related to these inhibitory actions.

DIAGNOSIS

In most cases somatostatinomas have been found by accident either at the time of cholecystectomy or during endoscopy. The presence of psammoma bodies in a duodenal tumor should particularly raise suspicion. Duodenal somatostatin-containing tumors are increasingly associated with von Recklinghausen's disease. Most of these tumors (>98%) do not cause the somatostatinoma syndrome. The diagnosis of the somatostatinoma syndrome requires the demonstration of elevated plasma somatostatin levels.

Treatment: Somatostatinomas

Pancreatic tumors are frequently (70–92%) metastatic at presentation, whereas 30–69% of small-intestinal somatostatinomas have metastases. Surgery is the treatment of choice for those without widespread hepatic metastases. Symptoms in patients with the somatostatinoma syndrome are also improved by octreotide treatment.

Vipomas

VIPomas are endocrine tumors that secrete excessive amounts of vasoactive intestinal peptide, which causes a distinct syndrome characterized by large-volume diarrhea, hypokalemia, and dehydration. This syndrome also is called Verner-Morrison syndrome, pancreatic cholera, and WDHA syndrome for *w*atery *d*iarrhea, *h*ypokalemia, and *a*chlorhydria, which some patients develop. The mean age of patients with this syndrome is 49 years; however, it can occur in children, and when it does, it is usually caused by a ganglioneuroma or ganglioneuroblastoma.

The principal symptoms are large-volume diarrhea (100%) severe enough to cause hypokalemia (80–100%), dehydration (83%), hypochlorhydria (54–76%), and flushing (20%). The diarrhea is secretory in nature, persisting during fasting, and is almost always >1 L/d and in 70% is >3 L/d. In a number of studies, the diarrhea was intermittent initially in up to half the patients. Most patients do not have accompanying steatorrhea (16%), and the increased stool volume is due to increased excretion of sodium and potassium, which, with the anions, accounts for the osmolality of the stool. Patients frequently have hyperglycemia (25–50%) and hypercalcemia (25–50%).

VIP is a 28-amino-acid peptide that is an important neurotransmitter, ubiquitously present in the CNS and GI tract. Its known actions include stimulation of small-intestinal chloride secretion as well as effects on smooth muscle contractility, inhibition of acid secretion, and vasodilatory effects, which explain most features of the clinical syndrome.

In adults 80–90% of VIPomas are pancreatic in location, with the rest due to VIP-secreting pheochromocytomas, intestinal carcinoids, and rarely ganglioneuromas. These tumors are usually solitary, 50–75% are in the pancreatic tail, and 37–68% have hepatic metastases at diagnosis. In children <10 years old, the syndrome is usually due to ganglioneuromas or ganglioblastomas and is less often malignant (10%).

DIAGNOSIS

The diagnosis requires the demonstration of an elevated plasma VIP level and the presence of large-volume diarrhea. A stool volume <700 mL/d is proposed to exclude the diagnosis of VIPoma. When the patient fasts, a number of diseases can be excluded that can cause marked diarrhea. Other diseases that can produce a secretory large-volume diarrhea include gastrinomas, chronic laxative abuse, carcinoid syndrome, systemic mastocytosis, rarely medullary thyroid cancer, diabetic diarrhea, sprue, and AIDS. Among these conditions, only VIPomas caused a marked increase in plasma VIP. Chronic surreptitious use of laxatives/diuretics can be particularly difficult to detect clinically. Hence, in a patient with unexplained chronic diarrhea, screens for laxatives should be performed; they will detect many, but not all, laxative abusers.

Treatment: Vasoactive Intestinal Peptidomas

The most important initial treatment in these patients is to correct their dehydration, hypokalemia, and electrolyte losses with fluid and electrolyte replacement. These patients may require 5 L/d of fluid and >350 meq/d of potassium. Because 37–68% of adults with VIPomas have metastatic disease in the liver at presentation, a significant number of patients cannot be cured surgically. In these patients long-acting somatostatin analogues such as octreotide and lanreotide are the drugs of choice.

Octreotide/lanreotide will control the diarrhea short- and long-term in 75–100% of patients. In nonresponsive patients the combination of glucocorticoids and octreotide/lanreotide has proved helpful in a small number of patients. Other drugs reported to be helpful in small numbers of patients include prednisone (60–100 mg/d), clonidine, indomethacin, phenothiazines, loperamide, lidamidine, lithium, propranolol, and metoclopramide. Treatment of advanced disease with embolization, chemoembolization, chemotherapy, radiotherapy, radiofrequency ablation, and peptide receptor radiotherapy may be helpful (see below).

Nonfunctional Pancreatic Endocrine Tumors (NF-Pets)

NF-PETs are endocrine tumors that originate in the pancreas and secrete no products, or their products do not cause a specific clinical syndrome. The symptoms are due entirely to the tumor per se. NF-PETs secrete chromogranin A (90–100%), chromogranin B (90–100%), PP (58%), α-HCG (human chorionic gonadotropin) (40%), and β-HCG (20%).

Because the symptoms are due to the tumor mass, patients with NF-PETs usually present late in the disease course with invasive tumors and hepatic metastases (64–92%) and the tumors are usually large (72% >5 cm). NF-PETs are usually solitary except in patients with MEN 1, in which case they are multiple. They occur primarily in the pancreatic head. Even though these tumors do not cause a functional syndrome, immunocytochemical studies show that they synthesize numerous peptides and cannot be distinguished from functional tumors by immunocytochemistry. In MEN 1, 80–100% of patients have microscopic NF-PETs, but they become large or symptomatic in only a minority (0–13%) of cases. In VHL 12–17% develop NF-PETs, and in 4% they are ≥3 cm in diameter.

The most common symptoms are abdominal pain (30–80%), jaundice (20–35%), and weight loss, fatigue, or bleeding; 10–30% are found incidentally. The average time from the beginning of symptoms to diagnosis is 5 years.

DIAGNOSIS

The diagnosis is established by histologic confirmation in a patient without either the clinical symptoms or the elevated plasma hormone levels of one of the established syndromes. The principal difficulty in diagnosis is to distinguish an NF-PET from a nonendocrine pancreatic tumor, which is more common. Even though chromogranin A levels are elevated in almost every patient, this is not specific for this disease as it can be found in functional PETs, carcinoids, and other neuroendocrine disorders. Plasma pancreatic polypeptide is increased in 22–71% of patients and should strongly suggest the diagnosis in a patient with a pancreatic mass because it is usually normal in patients with pancreatic adenocarcinomas. Elevated plasma PP is not diagnostic of this tumor because it is elevated in a number of other conditions, such as chronic renal failure, old age, inflammatory conditions, and diabetes. A positive somatostatin receptor scan in a patient with a pancreatic mass should suggest the presence of PET/NF-PET rather than a nonendocrine tumor.

Treatment: Nonfunctional Pancreatic Endocrine Tumors

Overall survival in patients with sporadic NF-PET is 30–63% at 5 years, with a median survival of 6 years. Unfortunately, surgical curative resection can be considered only in a minority of these patients because 64–92% present with metastatic disease. Treatment needs to be directed against the tumor per se using the various modalities discussed below for advanced disease. The treatment of NF-PETs in either MEN 1 patients or patients with VHL is controversial. Most recommend surgical resection for any tumor >2–3 cm in diameter; however, there is no consensus on smaller NF-PETs, with most recommending careful surveillance of these patients.

Grfomas

GRFomas are endocrine tumors that secrete excessive amounts of growth hormone–releasing factor (GRF) that cause acromegaly. GRF is a 44-amino-acid peptide, and 25–44% of PETs have GRF immunoreactivity, although it is uncommonly secreted. GRFomas are lung tumors in 47–54% of cases, PETs in 29–30%, and small-intestinal carcinoids in 8–10%; up to 12% occur at other sites. Patients have a mean age of 38 years, and the symptoms usually are due to either acromegaly or the tumor per se. The acromegaly caused by GRFomas is indistinguishable from classic acromegaly. The pancreatic tumors are usually large (>6 cm), and liver metastases are present in 39%. They should be suspected in any patient with acromegaly and an abdominal tumor, a patient with MEN 1 with acromegaly, or a patient without a pituitary adenoma with acromegaly or associated with hyperprolactinemia, which occurs in 70% of GRFomas. GRFomas are an uncommon cause of acromegaly. GRFomas occur in <1% of MEN 1 patients. The diagnosis is established by performing plasma assays for GRF and growth hormone. Most GRFomas have a plasma GRF level >300 pg/mL (normal <5 pg/mL men, <10 pg/mL women). Patients with GRFomas also have increased plasma levels of insulin-like growth factor type I (IGF-I) levels similar to those in classic acromegaly. Surgery is the treatment of choice if diffuse metastases are not present. Long-acting somatostatin analogues such as octreotide and lanreotide are the agents of choice, with 75–100% of patients responding.

Other Rare Pancreatic Endocrine Tumor Syndromes

Cushing's syndrome (ACTHoma) due to a PET occurs in 4–16% of all ectopic Cushing's syndrome cases. It occurs in 5% of cases of sporadic gastrinomas, almost invariably in patients with hepatic metastases, and is an independent poor prognostic factor. Paraneoplastic hypercalcemia due to PETs releasing parathyroid hormone–related peptide (PTHrP), a PTH-like material, or unknown factor, is rarely reported. The tumors are usually large, and liver metastases are usually present. Most (88%) appear to be due to release of PTHrP. PETs occasionally can cause the carcinoid syndrome. PETs secreting calcitonin have been proposed as a specific clinical syndrome. One-half of the patients have diarrhea, which disappears with resection of the tumor. The proposal that this could be a discrete syndrome is supported by the finding that 25–42% of patients with medullary thyroid cancer with hypercalcitonemia develop diarrhea, probably secondary to a motility disorder. This is classified in Table 350-2 as a possible specific disorder because so few cases have been described. Similarly classified with only a few cases described are a renin-producing PET in a patient presenting with hypertension; PETs secreting luteinizing hormone, resulting in masculinization or decreased libido; a PET-secreting erythropoietin resulting in polycythemia; and PETs secreting insulin-like growth factor II causing hypoglycemia (Table 350-2). Ghrelin is a 28-amino-acid peptide with a number of metabolic functions. Even though it is detectable immunohistochemically in most PETs, no specific syndrome is associated with release of ghrelin by the PET.

TUMOR LOCALIZATION

Localization of the primary tumor and knowledge of the extent of the disease are essential to the proper management of all carcinoids and PETs. Without proper localization studies it is not possible to determine whether the patient is a candidate for curative resection or cytoreductive surgery or requires antitumor treatment or to predict the patient's prognosis reliably.

Numerous tumor localization methods are used in both types of NETs, including conventional imaging studies (computed tomographic scanning, magnetic resonance imaging, transabdominal ultrasound, selective angiography), somatostatin receptor scintigraphy (SRS), and positron emission tomographic scanning. In PETs, endoscopic ultrasound (EUS) and functional localization by measuring venous hormonal gradients are also reported to be useful. Bronchial carcinoids are usually detected by standard chest radiography and assessed by CT. Rectal, duodenal, colonic, and gastric carcinoids are usually detected by GI endoscopy.

PETs, as well as carcinoid tumors, frequently overexpress high-affinity somatostatin receptors in both their primary tumors and their metastases. Of the five types of somatostatin receptors ($sst_{1–5}$), radiolabeled octreotide binds with high affinity to sst_2 and sst_5, has a lower affinity for sst_3, and has a very low affinity for sst_1 and sst_4. Between 90 and 100% of carcinoid tumors and PETs possess sst_2, and many also have the other four sst subtypes. Interaction with these receptors can be used to localize NETs by using [^{111}In-DTPA-d-Phe^1]octreotide and radionuclide scanning (SRS) as well as for

treatment of the hormone-excess state with octreotide or lanreotide, as discussed earlier. Because of its sensitivity and ability to localize tumor throughout the body, SRS is the initial imaging modality of choice for localizing both the primary and metastatic NETs. SRS localizes tumor in 73–89% of patients with carcinoids and in 56–100% of patients with PETs, except insulinomas. Insulinomas are usually small and have low densities of sst receptors, resulting in SRS being positive in only 12–50% of patients with insulinomas. Figure 350-3 shows an example of the increased sensitivity of SRS in a patient with a carcinoid tumor. The CT scan showed a single liver metastasis, whereas the SRS demonstrated three metastases in the liver in multiple locations. Occasional false-positive responses with SRS can occur (12% in one study) because numerous other normal tissues as well as diseases can have high densities of sst receptors, including granulomas (sarcoid, tuberculosis, etc.), thyroid diseases (goiter, thyroiditis), and activated lymphocytes (lymphomas, wound infections). For PETs in the pancreas, EUS is highly sensitive, localizing 77–100% of insulinomas, which occur almost exclusively within the pancreas. Endoscopic ultrasound is less sensitive for extrapancreatic tumors. It is increasingly used in patients with MEN 1 and to a lesser extent VHL to detect small PETs not seen with other modalities or for serial PET assessments to determine size changes or rapid growth in patients in whom surgery is deferred. EUS with cytologic evaluation also is used frequently to distinguish an NF-PET from a pancreatic adenocarcinoma or another nonendocrine pancreatic tumor.

Figure 350-3

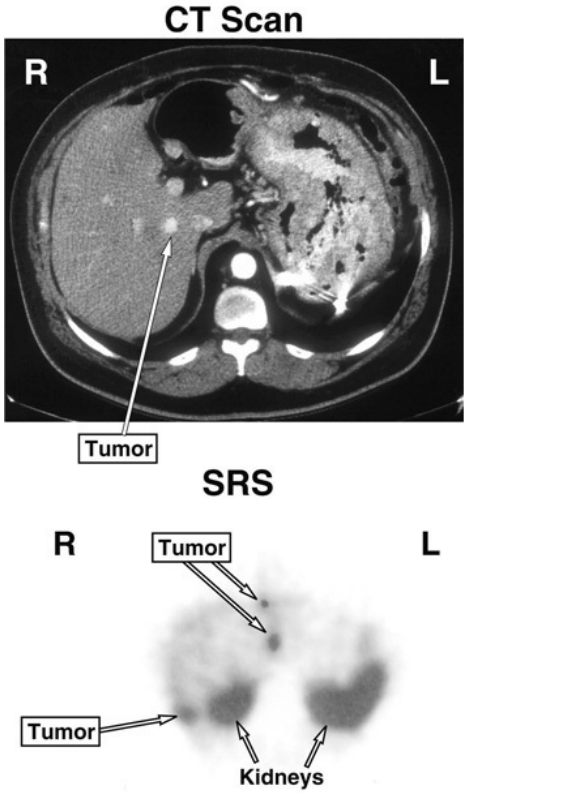

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ability of CT scanning (*top*) or somatostatin receptor scintigraphy (SRS) (*bottom*) to localize metastatic carcinoid in the liver.

Insulinomas overexpress receptors for GLP-1; a radiolabeled GLP-1 analogue can detect occult insulinomas not localized by other imaging modalities. Functional localization by measuring hormonal gradients is now uncommonly used with gastrinomas (after intraarterial secretin injections) but is still frequently used in insulinoma patients in whom other imaging studies are negative (assessing hepatic vein insulin concentrations post-intraarterial calcium injections). The intraarterial calcium test may also allow differentiation of the cause of the hypoglycemia and indicate whether it is due to an insulinoma or a nesidioblastosis. The latter entity is becoming increasingly important because hypoglycemia after gastric bypass surgery for obesity is increasing in frequency, and it is primarily due to nesidioblastosis, although it can occasionally be due to an insulinoma.

If liver metastases are identified by SRS, to plan the proper treatment either a CT or an MRI is recommended to assess the size and exact location of the metastases because SRS does not provide information on tumor size. Functional localization measuring hormone gradients after intraarterial calcium injections in insulinomas (insulin) or gastrin gradients after secretin injections in gastrinoma is a sensitive method, being positive in 80–100% of patients. However, this method provides only regional localization and therefore is reserved for cases in which the other imaging modalities are negative.

Two newer imaging modalities (positron emission tomography and use of hybrid scanners such as CT and SRS) may have increased sensitivity. Positron emission tomographic scanning with ^{18}F-fluoro-DOPA in patients with carcinoids or with ^{11}C-5-HTP or ^{68}gallium-labeled somatostatin analogues in patients with PETs or carcinoids has greater sensitivity than conventional imaging studies or SRS and probably will be used increasingly in the future. Positron emission tomographic scanning for GI NETs is not currently approved in the United States.

Treatment: Advanced Disease (Diffuse Metastatic Disease)

The single most important prognostic factor for survival is the presence of liver metastases (Fig. 350-4). For patients with foregut carcinoids without hepatic metastases, the 5-year survival in one study was 95%, and with distant metastases, it

was 20% (Fig. 350-4, bottom). With gastrinomas the 5-year survival without liver metastases is 98%; with limited metastases in one hepatic lobe, it is 78%; and with diffuse metastases, 16% (Fig. 350-4, top). In a large study of 156 patients (67 PETs, rest carcinoids) the overall 5-year survival rate was 77%; it was 96% without liver metastases, 73% with liver metastases, and 50% with distant disease. Therefore, treatment for advanced metastatic disease is an important challenge. A number of different modalities are reported to be effective, including cytoreductive surgery [surgically or by radio frequency ablation (RFA)], treatment with chemotherapy, somatostatin analogues, interferon α, hepatic embolization alone or with chemotherapy (chemoembolization), radiotherapy with radiolabeled beads/microspheres, peptide radio-receptor therapy, and liver transplantation.

Figure 350-4

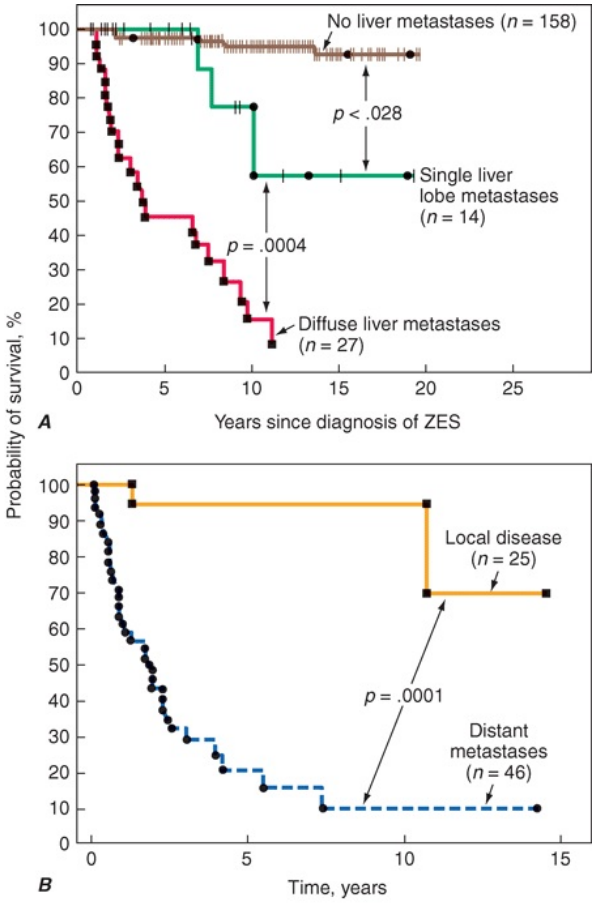

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effect of the presence and extent of liver metastases on survival in patients with gastrinomas (***A***) or carcinoid tumors (***B***). ZES, Zollinger-Ellison syndrome. *(Top panel is drawn from data from 199 patients with gastrinomas modified from F Yu et al: J Clin Oncol 17:615, 1999. Bottom panel is drawn from data from 71 patients with foregut carcinoid tumors from EW McDermott et al: Br J Surg 81:1007, 1994.)*

Specific Antitumor Treatments

Cytoreductive surgery, unfortunately, is possible in only 9–22% of patients who present with limited hepatic metastases. Although no randomized studies have proved that it extends life, results from a number of studies suggest that it probably increases survival; therefore, it is recommended, if possible. Radio frequency thermal ablation can be applied to GI NET liver metastases if they are limited in number (usually <5) and size (usually <3.5 cm in diameter). Response rates are >80%, the morbidity rate is low, and this procedure may be particularly helpful in patients with functional PETs that are difficult to control medically.

Chemotherapy for metastatic carcinoid tumors has generally been disappointing, with response rates of 0–40% with various two- and three-drug combinations. Chemotherapy for PETs has been more successful, with tumor shrinkage reported in 30–70% of patients. The current regimen of choice is streptozotocin and doxorubicin. In poorly differentiated PETs, chemotherapy with cisplatin, etoposide, or their derivatives is the recommended treatment, with response rates of 40–70%; however, responses are generally short-lived. Some newer combinations of chemotherapeutic agents show promise in small numbers of patients, including temozolomide (TMZ) alone, especially in PETs, which frequently have O^6-methylguanine DNA methyltransferase deficiency, which increases their TMZ sensitivity (34% response rate), and TMZ plus capecitabine (response rate 59–71%, retrospective studies).

Long-acting somatostatin analogues such as octreotide, lanreotide, and interferon α rarely decrease tumor size (i.e., 0–17%); however, these drugs have tumoristatic effects, stopping additional growth in 26–95% of patients with NETs. A randomized, double-blind study in patients with metastatic midgut carcinoids demonstrated a marked lengthening of time to progression (14.3 vs. 6 months, $p = .000072$) from the use of octreotide-LAR. This improvement was seen in patients with limited liver involvement. Whether this change will result in extended survival has not been proved. Somatostatin analogues can induce apoptosis in carcinoid tumors, and interferon α can decrease Bcl-2 protein expression, which probably contributes to its antiproliferative effects.

Hepatic embolization and chemoembolization (with dacarbazine, cisplatin, doxorubicin, 5-fluorouracil, or streptozotocin) have been reported to decrease tumor bulk and help control the symptoms of the hormone-excess state. These modalities generally are reserved for liver-directed therapy in cases in which treatment with somatostatin analogues, interferon (carcinoids), or chemotherapy (PETs) fails. Embolization, when combined with treatment with octreotide and interferon α, significantly reduces tumor progression (p = .008) compared with treatment with embolization and octreotide alone in patients with advanced midgut carcinoids.

Radiotherapy with radiolabeled somatostatin analogues that are internalized by the tumors is being investigated. Three different radionuclides are being used. High doses of [^{111}In-DTPA-d-Phe^{1}]octreotide, which emits γ-rays, internal conversion, and Auger electrons; yttrium-90, which emits high-energy β-particles coupled by a DOTA chelating group to octreotide or octreotate; and ^{177}lutetium-coupled analogues, which emit both, are all in clinical studies. ^{111}Indium-, ^{90}yttrium-, and ^{177}lutetium-labeled compounds caused tumor stabilization in 41–81%, 44–88%, and 23–40%, respectively, and a decrease in tumor size in 8–30%, 6–37%, and 38%, respectively, of patients with advanced metastatic NETs. Use of ^{177}Lu-labeled analogues to treat 504 patients with malignant NETs produced a reduction of tumor size of >50% in 30% of patients (2% complete) and tumor stabilization in 51%. An effect on survival has not been established. These results suggest that this novel therapy may be helpful, especially in patients with widespread metastatic disease.

Selective internal radiation therapy (SIRT) using ^{90}yttrium glass or resin microspheres is being evaluated in patients with unresectable NET liver metastases. The treatment requires careful evaluation for vascular shunting before treatment and generally is reserved for patients without extrahepatic metastatic disease and with adequate hepatic reserve. The ^{90}Y-microspheres are delivered to the liver by intraarterial injection from percutaneous placed catheters. In four studies involving metastatic NETs, the response rate varied from 50–61% (partial or complete), tumor stabilization occurred in 22–41%, and overall survival varied from 25–70 months. In the largest study (148 patients), no radiation-induced liver failure occurred and the most common side effect was fatigue (6.5%).

The use of liver transplantation has been abandoned for treatment of most metastatic tumors to the liver. However, for metastatic NETs, it is still a consideration. In a review of 103 cases of malignant NETs (48 PETs, 43 carcinoids) the 2- and 5-year survival rates were 60% and 47%, respectively. However, recurrence-free survival was low (<24%). For younger patients with metastatic NETs limited to the liver, liver transplantation may be justified.

Newer approaches show some promise in the treatment of advanced GI NETs. They include the use of growth factor inhibitors or inhibitors of their receptors (using tyrosine kinase inhibitors, monoclonal antibodies), inhibitors of mTor signaling (everolimus, temsirolimus), angiogenesis inhibitors, and VEGF or vascular endothelial growth factor receptor (VEGFR) tyrosine kinase inhibitors. A number of these agents, particularly sunitinib (tyrosine kinase inhibitor), various mTor inhibitors, and bevacizumub (monoclonal antibody against VEGF), show impressive activity. Additional value may result from selected combinations of agents.

FURTHER READINGS

Boudreaux JP et al: The NANETS consensus guideline for the diagnosis and management of neuroendocrine tumors: Well-differentiated neuroendocrine tumors of the jejunum, ileum, appendix, and cecum. Pancreas 2010 39:753, 2010[PMID: 20664473] [Full Text]

Capurso G et al: Molecular target therapy for gastroenteropancreatic endocrine tumours: Biological rationale and clinical perspectives. Crit Rev Oncol Hematol 72:110, 2009[PMID: 19249226] [Full Text]

Clark OH et al: NCCN Clinical Practice Guidelines in Oncology: Neuroendocrine tumors. J Natl Compr Canc Netw 7:712, 2009[PMID: 19635226] [Full Text]

Jensen RT et al: Inherited pancreatic endocrine tumor syndromes: Advances in molecular pathogenesis, diagnosis, management and controversies. Cancer 113:1807, 2008[PMID: 18798544] [Full Text]

Kulke MH et al: NANETS treatment guidelines: Well-differentiated neuroendocrine tumors of the stomach and pancreas. Pancreas 39:735, 2010[PMID: 20664472] [Full Text]

Metz DC et al: Gastrointestinal neuroendocrine tumors: Pancreatic endocrine tumors. Gastroenterology 135:1469, 2008[PMID: 18703061] [Full Text]

Modlin IM et al: Review article: Somatostatin analogues in the treatment of gastroenteropancreatic neuroendocrine (carcinoid) tumours. Aliment Pharmacol Ther 31:169, 2010[PMID: 19845567] [Full Text]

Pinchot SN et al: Carcinoid tumors. Oncologist 13:1255, 2008[PMID: 19091780] [Full Text]

Poncet G et al: Recent trends in the treatment of well-differentiated endocrine carcinoma of the small bowel. World J Gastroenterol 16:1696, 2010[PMID: 20380000] [Full Text]

Rinke A et al: Placebo-controlled, double-blind, prospective, randomized study on the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: A report from the PROMID Study Group. J Clin Oncol 27:4656, 2009[PMID: 19704057] [Full Text]

Van Essen M et al: Peptide-receptor radionuclide therapy for endocrine tumors. Nat Rev Endocrinol 5:382, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NEOPLASTIC DISORDERS AFFECTING MULTIPLE ENDOCRINE ORGANS

Multiple endocrine neoplasia syndrome is defined as a disorder with neoplasms in two or more different hormonal tissues in several members of a family. Several distinct genetic disorders predispose to endocrine gland neoplasia and cause hormone excess syndromes **(Table 351-1)**. DNA-based genetic testing is available for these disorders, but effective management requires an understanding of endocrine neoplasia and the range of clinical features that may be manifested in an individual patient.

Table 351-1 Disease Associations in the Multiple Endocrine Neoplasia (Men) Syndromes

MEN 1	MEN 2	Mixed Syndromes
Parathyroid hyperplasia or adenoma	**MEN 2A**	**Von Hippel–Lindau syndrome**
Islet cell hyperplasia, adenoma, or carcinoma	MTC	Pheochromocytoma
Pituitary hyperplasia or adenoma	Pheochromocytoma	Islet cell tumor
Other, less common manifestations: foregut carcinoid, pheochromocytoma, subcutaneous or visceral lipomas	Parathyroid hyperplasia or adenoma	Renal cell carcinoma
	MEN 2A with cutaneous lichen amyloidosis	Hemangioblastoma of central nervous system
	MEN 2A with Hirschsprung disease	Retinal angiomas
	Familial MTC	**Neurofibromatosis with features of MEN 1 or 2**
	MEN 2B	**Carney complex**
	MTC	Myxomas of heart, skin, and breast
	Pheochromocytoma	Spotty cutaneous pigmentation
	Mucosal and gastrointestinal neuromas	Testicular, adrenal, and GH-producing pituitary tumors
	Marfanoid features	Peripheral nerve schwannomas
		Familial growth hormone or prolactin-producing pituitary tumors

Abbreviations: GH, growth hormone; MTC, medullary thyroid carcinoma.

Multiple Endocrine Neoplasia (Men) Type 1

MEN 1, or Wermer's syndrome, is inherited as an autosomal dominant trait. This syndrome is characterized by neoplasia of the parathyroid glands, enteropancreatic tumors, anterior pituitary adenomas, and other neuroendocrine tumors with variable penetrance (Table 351-1). Although rare, MEN 1 is the most common multiple endocrine neoplasia syndrome, with an estimated prevalence of 2–20 per 100,000 in the general population. It is caused by inactivating mutations of the tumor-suppressor gene *MEN1* located at chromosome 11q13. The *MEN1* gene codes for a nuclear protein called Menin. Menin interacts with JunD, suppressing JunD-dependent transcriptional activation. It is unclear how this accounts for Menin growth regulatory activity, since JunD is associated with inhibition of cell growth. Each child born to an affected parent has a 50% probability of inheriting the gene. The variable penetrance of the several neoplastic components can make the differential diagnosis and treatment challenging.

CLINICAL MANIFESTATIONS

Primary hyperparathyroidism is the most common manifestation of MEN 1, with an estimated penetrance of 95–100%. Hypercalcemia may develop during the teenage years, and most individuals

are affected by age 40 **(Fig. 351-1)**. Hyperparathyroidism is the earliest manifestation of the syndrome in most MEN 1 patients. The neoplastic changes in hyperparathyroidism provide a specific example of one of the cardinal features of endocrine tumors in MEN 1: multicentricity. The neoplastic changes inevitably affect multiple parathyroid glands, making surgical cure difficult. Screening for hyperparathyroidism involves measurement of either an albumin-adjusted or an ionized serum calcium level. The diagnosis is established by demonstrating elevated levels of serum calcium and intact parathyroid hormone. Manifestations of hyperparathyroidism in MEN 1 do not differ substantially from those in sporadic hyperparathyroidism and include calcium-containing kidney stones, kidney failure, nephrocalcinosis, bone abnormalities (i.e., osteoporosis, osteitis fibrosa cystica), and gastrointestinal and musculoskeletal complaints. Management is challenging because of early onset, significant recurrence rates, and the multiplicity of parathyroid gland involvement. Differentiation of hyperparathyroidism of MEN 1 from other forms of familial primary hyperparathyroidism usually is based on family history, histologic features of resected parathyroid tissue, the presence of a *MEN1* mutation, and, sometimes, long-term observation to determine whether other manifestations of MEN 1 develop. Parathyroid hyperplasia is the most common cause of hyperparathyroidism in MEN 1, although single and multiple adenomas have been described. Hyperplasia of one or more parathyroid glands is common in younger patients; adenomas usually are found in older patients or those with long-standing disease.

Figure 351-1

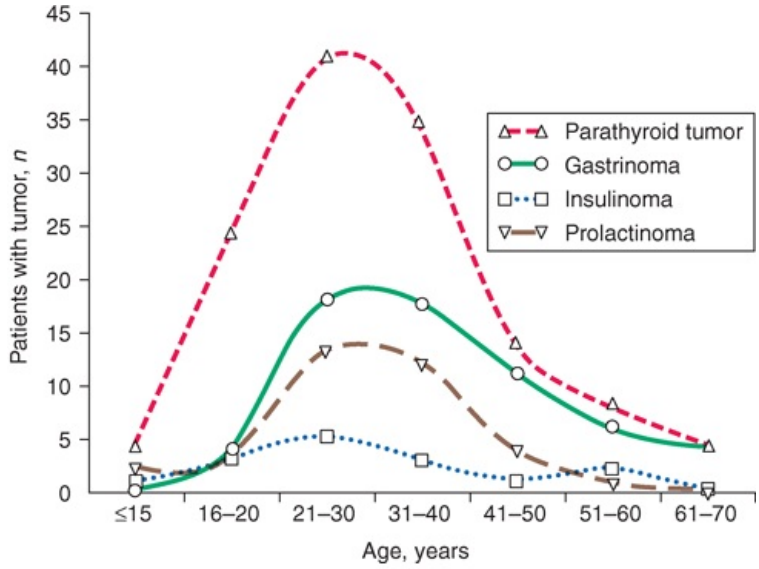

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Age at onset of endocrine tumor expression in multiple endocrine neoplasia type 1 (MEN 1). Data derived from retrospective analysis for each endocrine organ hyperfunction in 130 cases of MEN 1. Age at onset is the age at first symptom or, with tumors not causing symptoms, age at the time of the first abnormal finding on a screening test. The rate of diagnosis of hyperparathyroidism increased sharply between ages 16 and 20 years. *(Reprinted with permission from S Marx et al: Ann Intern Med 129:484, 1998.)*

Enteropancreatic tumors are the second most common manifestation of MEN 1, with an estimated penetrance of 50%. They tend to occur in parallel with hyperparathyroidism (Fig. 351-1); 30% are malignant. Most of these tumors secrete peptide hormones that cause specific clinical syndromes. Those syndromes, however, may have an insidious onset and a slow progression, making their diagnosis difficult and in many cases delayed. Some enteropancreatic tumors do not secrete hormones. Those "silent" tumors usually are found during radiographic screening. Metastasis, most commonly to the liver, occurs in about a third of patients.

Gastrinomas are the most common enteropancreatic tumors observed in MEN 1 patients and result in the Zollinger-Ellison syndrome (ZES). ZES is caused by excessive gastrin production and occurs in more than one-half of MEN 1 patients with small carcinoid- like tumors in the duodenal wall or, less often, by pancreatic islet cell tumors. There may be more than one gastrin-producing tumor, making localization difficult. The robust acid production may cause esophagitis, duodenal ulcers throughout the duodenum, ulcers involving the proximal jejunum, and diarrhea. The ulcer diathesis is commonly refractory to conservative therapy such as antacids. The diagnosis is made by finding increased gastric acid secretion, elevated basal gastrin levels in the serum [generally >115 pmol/L (200 pg/mL)], and an exaggerated response of serum gastrin to either secretin or calcium. Other causes of elevated serum gastrin levels, such as achlorhydria, treatment with H_2 receptor antagonists or proton pump

inhibitors, retained gastric antrum, small-bowel resection, gastric outlet obstruction, and hypercalcemia, should be excluded (Fig. 351-1). High-resolution, early-phase CT scanning, abdominal MRI with contrast, octreotide scan, and/or endoscopic ultrasound are the best preoperative techniques for identification of the primary and metastatic gastrinoma; intraoperative ultrasonography is the most sensitive method for detection of small tumors. Approximately one-fourth of all cases of ZES occur in the context of MEN 1.

Insulinomas are the second most common enteropancreatic tumors in patients who have MEN 1. Unlike gastrinomas, most insulinomas originate in the pancreas bed, becoming the most common pancreatic tumor in MEN 1. Hypoglycemia caused by insulinomas is observed in about one-third of MEN 1 patients with pancreatic islet cell tumors (Fig. 351-1). The tumors may be benign or malignant (25%). The diagnosis can be suggested by documenting hypoglycemia during a short fast with simultaneous inappropriate elevation of serum insulin and C-peptide levels. More commonly, it is necessary to subject the patient to a supervised 12- to 72-h fast to provoke hypoglycemia (Chap. 345). Large insulinomas may be identified by CT or MRI scanning; small tumors not detected by conventional radiographic techniques may be localized by endoscopic ultrasound or selective arteriographic injection of calcium into each of the arteries that supply the pancreas and sampling of the hepatic vein for insulin to determine the anatomic region containing the tumor. Intraoperative ultrasonography is used frequently to localize these tumors. The trend toward earlier diagnosis of, hence, smaller tumors has reduced the usefulness of octreotide scanning, which is positive in a minority of these patients.

Glucagonoma, which is seen occasionally in MEN 1, causes a syndrome of hyperglycemia, skin rash (necrolytic migratory erythema), anorexia, glossitis, anemia, depression, diarrhea, and venous thrombosis. In about half of these patients the plasma glucagon level is high, leading to its designation as the *glucagonoma syndrome*, although elevation of the plasma glucagon level in MEN 1 patients is not necessarily associated with these symptoms. Some patients with this syndrome also have elevated plasma ghrelin levels. The glucagonoma syndrome may represent a complex interaction between glucagon and ghrelin overproduction and the nutritional status of the patient.

The *Verner-Morrison,* or *watery diarrhea, syndrome* consists of watery diarrhea, hypokalemia, hypochlorhydria, and metabolic acidosis. The diarrhea can be voluminous and almost always is found in association with an islet cell tumor in the context of MEN 1, prompting use of the term *pancreatic cholera*. However, when not associated with MEN 1, the syndrome outside of MEN 1 is not restricted to pancreatic islet tumors and has been observed with carcinoids or other tumors. This syndrome is believed to be due to overproduction of vasoactive intestinal peptide (VIP), although plasma VIP levels may not be elevated. Hypercalcemia may be induced by the effects of VIP to stimulate osteoclastic bone resorption as well as by hyperparathyroidism. Other disorders that should be considered in the differential diagnosis of chronic diarrhea include infectious or parasitic diseases, inflammatory bowel disease, sprue, and other endocrine causes such as ZES, carcinoid, and medullary thyroid carcinoma.

The pancreatic neoplasms differ from the other components of MEN 1 in that approximately one-third of the tumors display malignant features, including hepatic metastases. The pancreatic neoplasms also can be used to highlight another characteristic of MEN 1: the specific impact of a hormone produced by one component of MEN 1 on another neoplastic component of this syndrome. Specific examples include the effects of either corticotropin-releasing hormone (CRH) or growth hormone –releasing hormone (GHRH) production by an islet cell tumor to cause a syndrome of excess adrenocorticotropin (ACTH) (Cushing's disease) or GH (acromegaly) production by the pituitary gland. These secondary interactions add complexity to the diagnosis and management of these tumor syndromes. Pancreatic islet cell tumors are diagnosed by identification of a characteristic clinical syndrome, hormonal assays with or without provocative stimuli, or radiographic techniques. One approach involves annual screening of individuals at risk with measurement of basal and meal-stimulated levels of pancreatic polypeptide to identify the tumors as early as possible; the rationale for this screening strategy is the concept that surgical removal of islet cell tumors at an early stage will be curative. Other approaches to screening include measurement of serum gastrin and pancreatic polypeptide levels every 2–3 years, with the rationale that pancreatic neoplasms will be detected at a later stage but can be managed medically, if possible, or by surgery. High-resolution, early-phase CT scanning or endoscopic ultrasound provides the best preoperative technique for identification of these tumors; intraoperative ultrasonography is the most sensitive method for detection of small tumors. Although fluorodeoxyglucose–positron emission tomography (FDG-PET) scanning detects ~50% of pancreatic islet cell tumors, most of these tumors are large; as most of these tumors can be identified by CT or ultrasound, the lack of sensitivity for small tumors makes FDG-PET scanning unhelpful for early diagnosis.

Pituitary tumors occur in 20–30% of patients with MEN 1 and tend to be multicentric. These tumors can exhibit aggressive behavior and local invasiveness that makes them difficult to resect (Chap. 339). Prolactinomas are the most common (Fig. 351-1) and are diagnosed by finding serum prolactin levels >200 μg/L with or without a pituitary mass evident on MRI. Values <200 μg/L may be due to a prolactin-secreting neoplasm or to compression of the pituitary stalk by a different type of pituitary tumor. Acromegaly due to excessive GH production is the second most common syndrome caused by pituitary

tumors in MEN 1 and can rarely be due to production of GHRH by an islet cell tumor (see above). The possibility of hereditary growth hormone– or prolactin-secreting tumors (discussed below in "Other Genetic Endocrine Tumor Syndromes") should be considered in the differential diagnosis. Cushing's disease can be caused by ACTH-producing pituitary tumors or by ectopic production of ACTH or CRH by other components of the MEN 1 syndrome, including islet cell or carcinoid tumors or adrenal adenomas. Diagnosis of pituitary Cushing's disease is generally best accomplished by a high-dose dexamethasone suppression test or by petrosal venous sinus sampling for ACTH after IV injection of CRH. Differentiation of a primary pituitary tumor from an ectopic CRH-producing tumor may be difficult because the pituitary is the source of ACTH in both disorders; documentation of CRH production by a pancreatic islet or carcinoid tumor may be the only method of proving ectopic CRH production.

Adrenal cortical tumors are found in almost one-half of gene carriers but are rarely functional; malignancy in cortical adenomas is uncommon. Rare cases of pheochromocytoma have been described in the context of MEN 1. Due to their rarity, screening for these tumors is indicated only when there are suggestive symptoms.

Carcinoid tumors in MEN 1 are of the foregut type and are derived from thymus, lung, stomach, or duodenum; they may metastasize or be locally invasive. These tumors usually produce serotonin, calcitonin, or CRH; the typical carcinoid syndrome with flushing, diarrhea, and bronchospasm is rare (Chap. 350). Mediastinal carcinoid tumors (an upper mediastinal mass) are more common in men; bronchial carcinoid tumors are more common in women. Carcinoid tumors are a late manifestation of MEN 1; some reports have emphasized the importance of routine chest CT screening for mediastinal carcinoid tumors because of their high rate of malignant transformation and aggressive behavior.

Unusual Manifestations of MEN 1

Subcutaneous or visceral lipomas and cutaneous leiomyomas also may be present but rarely undergo malignant transformation. Skin angiofibromas or collagenomas are seen in most patients with MEN 1 when carefully sought.

GENETIC CONSIDERATIONS

MEN1 gene mutations are found in >90% of families with the syndrome **(Fig. 351-2)**. Genetic testing can be performed in individuals at risk for the development of MEN 1 and is commercially available in the United States and Europe. The major value of genetic testing in a kindred with an identifiable mutation is the assignment or exclusion of gene carrier status. In those identified as carrying the mutant gene, routine screening for individual manifestations of MEN 1 should be performed as outlined above. Those with negative genetic test results in a kindred with a known germ-line mutation can be excluded from further screening for MEN 1. A significant percentage of sporadic parathyroid, islet cell, and carcinoid tumors also have loss or mutation of *MEN1*. There is no correlation between a particular germ-line mutation and a clinical phenotype. It is presumed that these mutations are somatic and occur in a single cell, leading to subsequent transformation.

Figure 351-2

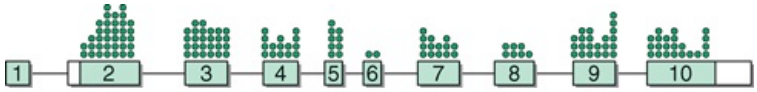

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic depiction of the *MEN1* gene and the distribution of mutations. The shaded areas show coding sequence. The closed circles show the relative distribution of mutations, mostly inactivating, in each exon. Mutation data are derived from the Human Gene Mutation Database, from which more detailed information can be obtained at *www.uwcm.ac.uk/uwcm/mg/hgmd0.html*. *(From M Krawczak, DN Cooper: Trends Genet 13:1321, 1998.)*

TREATMENT: MULTIPLE ENDOCRINE NEOPLASIA TYPE 1

Almost everyone who inherits a mutant *MEN1* gene develops at least one clinical manifestation of the syndrome. Most develop hyperparathyroidism, 80% develop pancreatic islet cell tumors, and more than half develop pituitary tumors. For most of these tumors, initial surgery is not curative and patients frequently require multiple surgical procedures and surgery on two or more endocrine glands during a lifetime. For this reason, it is essential to establish clear goals for management of these patients rather than to recommend surgery casually each time a tumor is discovered. Ranges for acceptable management are discussed below.

Hyperparathyroidism

Individuals with serum calcium levels >3.0 mmol/L (12 mg/dL), evidence of calcium nephrolithiasis or renal dysfunction, neuropathic or muscular symptoms, or bone involvement (including osteopenia) and individuals <50 years of age should undergo parathyroid exploration. There is less agreement about

the necessity for parathyroid exploration in individuals who do not meet these criteria, and observation may be appropriate in MEN 1 patients with asymptomatic hyperparathyroidism.

When parathyroid surgery is indicated in MEN 1, there are two approaches. In the first, all parathyroid tissue is identified and removed at the time of primary operation, and parathyroid tissue is implanted in the nondominant forearm. Thymectomy also should be performed because of the potential for later development of malignantcarcinoid tumors. If reoperation for hyperparathyroidism is necessary at a later date, transplanted parathyroid tissue can be resected from the forearm with titration of tissue removal to lower the intact parathyroid hormone (PTH) to <50% of basal.

Another approach is to remove 3–3.5 parathyroid glands from the neck (leaving ~50 mg of parathyroid tissue), carefully marking the location of residual tissue so that the remaining tissue can be located easily during subsequent surgery. If this approach is used, intraoperative PTH measurements should be utilized to monitor adequacy of removal of parathyroid tissue with a goal of reducing postoperative serum intact PTH to ≥50% of basal values.

The use of high-resolution CT scanning (1 mm) and imaging during three phases of contrast flow has substantially improved the ability to identify aberrantly located parathyroid tissue. As this issue arises with some frequency in the context of parathyroid disease in MEN 1, this technique should be utilized to locate parathyroid tissue before reoperation for a failed exploration, and it may be useful before the initial operation.

Pancreatic Islet Cell Tumors

(See Chap. 350 for discussion of pancreatic islet cell tumors not associated with MEN 1.) Two features of pancreatic islet cell tumors in MEN 1 complicate management. First, the tumors are multicentric, are malignant about a third of the time, and cause death in 10–20% of patients. Second, performance of a total pancreatectomy to prevent malignancy causes diabetes mellitus, a disease with significant long-term complications that include neuropathy, retinopathy, and nephropathy. These features make it difficult to formulate clear-cut guidelines, but some general concepts appear to be valid. (1) Islet cell tumors producing insulin, glucagon, VIP, GHRH, or CRH should be resected because medical therapy for the hormonal effects of these tumors are generally ineffective. (2) Gastrin-producing islet cell tumors that cause ZES are frequently multicentric. Recent experience suggests that a high percentage of ZES in MEN 1 is caused by duodenal wall carcinoid tumors and that resection of these tumors improves the cure rate. Treatment with H_2 receptor antagonists (cimetidine or ranitidine) or proton pump inhibitors (omeprazole, lansoprazole, esomeprazole, etc.) provides an alternative, and some think preferable, therapy to surgery for control of ulcer disease in patients with multicentric tumors or hepatic metastases. (3) In families in which there is a high incidence of malignant islet cell tumors that cause death, total pancreatectomy at an early age may be considered to prevent malignancy, although it should be noted that this surgical intervention does not prevent the development of neuroendocrine tumors outside the pancreato-duodenal region.

Management of metastatic islet cell carcinoma is unsatisfactory. Hormonal abnormalities sometimes can be controlled. For example, ZES can be treated with H_2 receptor antagonists or proton pump inhibitors; the somatostatin analogues octreotide and lanreotide are useful in the management of carcinoid, glucagonoma, and the watery diarrhea syndrome. Bilateral adrenalectomy may be required for ectopic ACTH syndrome if medical therapy is ineffective (Chap. 342). Islet cell carcinomas frequently metastasize to the liver but may grow slowly. Hepatic artery embolization, radiofrequency ablation, or chemotherapy (5-fluorouracil, streptozocin, chlorozotocin, doxorubicin, or dacarbazine) may reduce tumor streptozotocin mass, control symptoms of hormone excess, and prolong life; however, these treatments are never curative. There is evolving evidence that everolimus, an inhibitor of mTor (mammalian target of rapamycin) causes regression of tumor size; 2 of 13 islet cell carcinomas and 2 of 12 carcinoid tumors had a >30% reduction in size and >60% had stable disease.

Pituitary Tumors

Treatment of prolactinomas with dopamine agonists (bromocriptine, cabergoline, or quinagolide) usually returns the serum prolactin level to normal and prevents further tumor growth (Chap. 339). Surgical resection of a prolactinoma is rarely curative but may relieve mass effects. Transsphenoidal resection is appropriate for neoplasms that secrete ACTH, GH, or the α subunit of the pituitary glycoprotein hormones. Octreotide reduces tumor mass in one-third of GH-secreting tumors and reduces GH and insulin-like growth factor I levels in >75% of patients. Pegvisomant, a GH antagonist, rapidly lowers insulin-like growth factor levels in patients with acromegaly (Chap. 339). Radiation therapy may be useful for large or recurrent tumors.

Improvements in the management of MEN 1, particularly the earlier recognition of islet cell and pituitary tumors, have improved outcomes in these patients. As a result, other neoplastic manifestations that develop later in the course of this disorder, such as carcinoid syndrome, are now seen with increased frequency.

Multiple Endocrine Neoplasia Type 2

CLINICAL MANIFESTATIONS

Medullary thyroid carcinoma (MTC) and pheochromocytoma are associated in two major syndromes: MEN type 2A and MEN type 2B (Table 351-1). MEN 2A is the combination of MTC, hyperparathyroidism, and pheochromocytoma. Three subvariants of MEN 2A are familial medullary thyroid carcinoma (FMTC), MEN 2A with cutaneous lichen amyloidosis, and MEN 2A with Hirschsprung disease. MEN 2B is the combination of MTC, pheochromocytoma, mucosal neuromas, intestinal ganglioneuromatosis, and marfanoid features.

Multiple Endocrine Neoplasia Type 2a

MTC is the most common manifestation. This tumor usually develops in childhood, beginning as hyperplasia of the calcitonin-producing cells (C cells) of the thyroid. MTC typically is located at the junction of the upper one-third and lower two-thirds of each lobe of the thyroid, reflecting the high density of C cells in this location; tumors >1 cm in size frequently are associated with local or distant metastases.

Pheochromocytoma occurs in ~50% of patients with MEN 2A and causes palpitations, nervousness, headaches, and sometimes sweating (Chap. 343). About half of the tumors are bilateral, and >50% of patients who have had unilateral adrenalectomy develop a pheochromocytoma in the contralateral gland within a decade. A second feature of these tumors is a disproportionate increase in the secretion of epinephrine relative to norepinephrine. This characteristic differentiates the MEN 2 pheochromocytomas from sporadic pheochromocytoma and those associated with von Hippel–Lindau (VHL) syndrome, hereditary paraganglioma, or neurofibromatosis. Capsular invasion is common, but metastasis is uncommon. Finally, the pheochromocytomas almost always are found in the adrenal gland, differentiating the pheochromocytomas in MEN 2 from the extraadrenal tumors more commonly found in hereditary paraganglioma syndromes.

Hyperparathyroidism occurs in 15–20% of patients, with the peak incidence in the third or fourth decade. The manifestations of hyperparathyroidism do not differ from those in other forms of primary hyperparathyroidism (Chap. 353). Diagnosis is established by finding hypercalcemia, hypophosphatemia, hypercalciuria, and an inappropriately high serum level of intact PTH. Multiglandular parathyroid hyperplasia is the most common histologic finding, although with long-standing disease adenomatous changes may be superimposed on hyperplasia.

The most common subvariant of MEN 2A is familial MTC, an autosomal dominant syndrome in which MTC is the only manifestation (Table 351-1). The clinical diagnosis of FMTC is established by the identification of MTC in multiple generations without a pheochromocytoma. Since the penetrance of pheochromocytoma is 50% in MEN 2A, it is possible that MEN 2A could masquerade as FMTC in small kindreds. It is important to consider this possibility carefully before classifying a kindred as having FMTC; failure to do so could lead to death or serious morbidity from pheochromocytoma in an affected kindred member. The difficulty of differentiating MEN 2A and FMTC is discussed further below.

Multiple Endocrine Neoplasia Type 2b

The association of MTC, pheochromocytoma, mucosal neuromas, and a marfanoid habitus is designated MEN 2B. MTC in MEN 2B develops earlier and is more aggressive than in MEN 2A. Metastatic disease has been described before 1 year of age, and death may occur in the second or third decade of life. However, the prognosis is not invariably bad even in patients with metastatic disease, as evidenced by a number of multigenerational families with this disease.

Pheochromocytoma occurs in more than half of MEN 2B patients and does not differ from that in MEN 2A. Hypercalcemia is rare in MEN 2B, and there are no well-documented examples of hyperparathyroidism.

The mucosal neuromas and marfanoid body habitus are the most distinctive features and are recognizable in childhood. Neuromas are present on the tip of the tongue, under the eyelids, and throughout the gastrointestinal tract and are true neuromas, distinct from neurofibromas. The most common presentation in children relates to gastrointestinal symptomatology, including intermittent colic, pseudoobstruction, and diarrhea.

GENETIC CONSIDERATIONS

Mutations of the *RET* protooncogene have been identified in most patients with MEN 2 **(Fig. 351-3)**. *RET* encodes a tyrosine kinase receptor that in combination with a co-receptor, GFRα, normally is activated by glial cell–derived neurotrophic factor (GDNF) or other members of this transforming growth factor ß–like family of peptides, including artemin, persephin, and neurturin. In the C cell there is evidence that persephin normally activates the *RET*/GFRα-4 receptor complex and is partially responsible for migration of the C cells into the thyroid gland, whereas in the developing neuronal system of the gastrointestinal tract, GDNF activates the *RET*/GFRα-1 complex. *RET* mutations induce constitutive activity of the receptor, explaining the autosomal dominant transmission of the disorder.

Figure 351-3

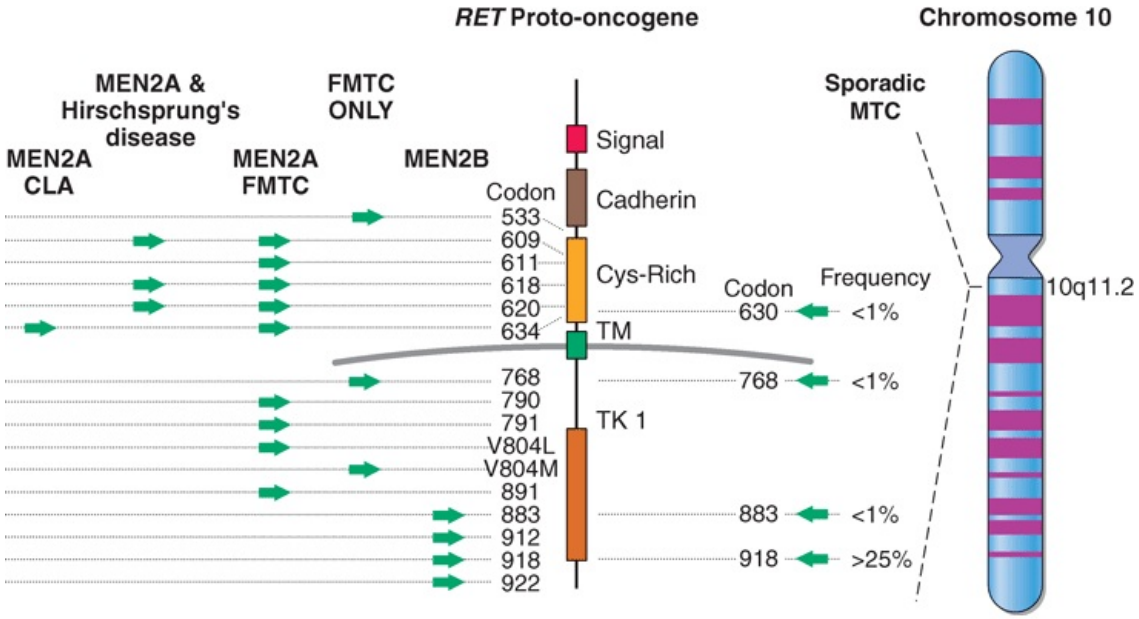

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of the *RET* protooncogene showing mutations found in MEN type 2 and sporadic medullary thyroid carcinoma (MTC). The *RET* protooncogene is located on the proximal arm of chromosome 10q (10q11.2). Activating mutations of two functional domains of RET tyrosine kinase receptor have been identified. The first affects a cysteine-rich (Cys-Rich) region in the extracellular portion of the receptor. Each germ-line mutation changes a cysteine at codons 609, 611, 618, 620, or 634 to another amino acid. The second region is the intracellular tyrosine kinase (TK) domain. Codon 634 mutations account for ~80% of all germ-line mutations. Mutations of codons 630, 768, 883, and 918 have been identified as somatic (non-germ-line) mutations that occur in a single parafollicular or C cell within the thyroid gland in sporadic MTC. A codon 918 mutation is the most common somatic mutation. MEN2, multiple endocrine neoplasia type 2; CLA, cutaneous lichen amyloidosis; FMTC, familial medullary thyroid carcinoma; Signal, the signal peptide; Cadherin, a cadherin-like region in the extracellular domain; TM, transmembrane domain; TK, tyrosine kinase domain.

Naturally occurring mutations localize to two regions of the *RET* tyrosine kinase receptor. The first is a cysteine-rich extracellular domain; point mutations in the coding sequence for one of six cysteines (codons 609, 611, 618, 620, 630, and 634) cause amino acid substitutions that induce receptor dimerization and activation in the absence of its ligand. Codon 634 mutations occur in 80% of MEN 2A kindreds and are most commonly associated with classic MEN 2A features (Figs. 351-3 and 351-2); an arginine substitution at this codon accounts for half of all MEN 2A mutations. All reported families with MEN 2A and cutaneous lichen amyloidosis have a codon 634 mutation. Mutations of codon 609, 611, 618, or 620 occur in 10–15% of MEN 2A kindreds and are more commonly associated with FMTC (Fig. 351-3). Mutations in codons 609, 618, and 620 also have been identified in a variant of MEN 2A that includes Hirschsprung disease (Fig. 351-3). The second region of the *RET* tyrosine kinase that is mutated in MEN 2 is in the substrate recognition pocket at codon 918 (Fig. 351-3). This activating mutation is present in ~95% of patients with MEN 2B and accounts for 5% of all *RET* protooncogene mutations in MEN 2. Mutations of codon 883 and 922 also have been identified in a few patients with MEN 2B.

Uncommon mutations (<5% of the total) include those of codons 533 (exon 8), 666, 768, 777, 790, 791, 804, 891, and 912. Mutations associated with only FMTC include codons 533, 768, and 912. With greater experience, mutations that once were associated with FMTC only (666, 791, V804L, V804M, and 891) have been found in MEN 2A as there have been occasional descriptions of pheochromocytoma. At present it is reasonable to conclude that only kindreds with codon 533, 768, or 912 mutations are consistently associated with FMTC; in kindreds with all other *RET* mutations, pheochromocytoma is a possibility. The recognition that germ-line mutations occur in at least 6% of patients with apparently sporadic MTC has led to the firm recommendation that all patients with MTC should be screened for these mutations. The effort to screen patients with sporadic MTC, combined with the fact that new kindreds with classic MEN 2A are being recognized less frequently, has led to a shift in the mutation frequencies. These findings mirror results in other malignancies in which germ-line mutations of cancer-causing genes contribute to a greater percentage of apparently sporadic cancer than was considered previously. The recognition of new *RET* mutations suggests that more will be identified in the future.

Somatic mutations (found only in the tumor and not transmitted in the germ line) of the *RET* protooncogene have been identified in sporadic MTC; 25–60% of sporadic tumors have codon 918

mutations, and somatic mutations in codons 630, 768, and 804 have been identified (Fig. 351-3).

Treatment: Multiple Endocrine Neoplasia Type 2

SCREENING FOR MULTIPLE ENDOCRINE NEOPLASIA TYPE 2

Death from MTC can be prevented by early thyroidectomy. The identification of *RET* protooncogene mutations and the application of DNA-based molecular diagnostic techniques to identify these mutations have simplified the screening process. During the initial evaluation of a kindred, a *RET* protooncogene analysis should be performed on an individual with proven MEN 2A. Establishment of the specific germ-line mutation facilitates the subsequent analysis of other family members. Each family member at risk should be tested twice for the presence of the specific mutation; the second analysis should be performed on a new DNA sample and, ideally, in a second laboratory to exclude sample mix-up or technical error (see *www.genetests.org* for an up-to-date list of laboratory testing sites). Both false-positive and false-negative analyses have been described. A false-negative test result is of the greatest concern because calcitonin testing is now rarely performed as a diagnostic backup study; if there is a genetic test error, a child may present in the second or third decade with metastatic MTC. Individuals in a kindred with a known mutation who have two normal analyses can be excluded from further screening.

There is a consensus that children with codon 883, 918, and 922 mutations, those associated with MEN 2B, should have a total thyroidectomy and central lymph node dissection (level VI) performed during the first months of life or soon after identification of the syndrome. If local metastasis is discovered, a more extensive lymph node dissection (levels II to V) is generally indicated. In children with codon 611, 618, 620, 630, 634, and 891 mutations, thyroidectomy should be performed before age 6 years because of reports of local metastatic disease in children this age. Finally, there are kindreds with codon 609, 768, 790, 791, 804, and 912 mutations in which the phenotype of MTC appears to be less aggressive. A clinician caring for children with one of these mutations faces a dilemma. In many kindreds there has never been a death from MTC caused by one of these mutations. However, in other kindreds there are examples of metastatic disease occurring early in life. For example, metastatic disease before age 6 years has been described with codon 609 and 804 mutations and before age 14 years in a patient with a codon 912 mutation. In kindreds with these mutations, two management approaches have been suggested: (1) perform a total thyroidectomy with or without central node dissection at some arbitrary age (perhaps 6–10 years of age) or (2) continue annual or biannual calcitonin provocative testing with performance of total thyroidectomy with or without central neck dissection when the test becomes abnormal. The pentagastrin test involves measurement of serum calcitonin basally and 2, 5, 10, and 15 min after a bolus injection of 5 μg pentagastrin per kilogram of body weight. Patients should be warned before pentagastrin injection of epigastric tightness, nausea, warmth, and tingling of extremities and reassured that the symptoms will last ~2 min. If pentagastrin is unavailable, an alternative is a short calcium infusion, performed by obtaining a baseline serum calcitonin and then infusing 150 mg calcium salt IV over 10 min with measurement of serum calcitonin at 5, 10, 15, 30 min after initiation of the infusion.

The *RET* protooncogene analysis should be performed in patients with suspected MEN 2B to detect codon 883, 918, and 922 mutations, especially in newborn children in whom the diagnosis is suspected but the clinical phenotype is not fully developed. Other family members at risk for MEN 2B also should be tested because the mucosal neuromas can be subtle. Most MEN 2B mutations represent de novo mutations derived from the paternal allele. In the rare families with proven germ-line transmission of MTC but no identifiable *RET* protooncogene mutation (sequencing of the entire *RET* gene should be performed), annual pentagastrin or calcium testing should be performed on members at risk.

Annual screening for pheochromocytoma in patients with germ-line *RET* mutations should be performed by measuring basal plasma or 24-h urine catecholamines and metanephrines. The goal is to identify a pheochromocytoma before it causes significant symptoms or is likely to cause sudden death, an event most commonly associated with large tumors. Although there are kindreds with FMTC and specific *RET* mutations in which no pheochromocytomas have been identified (Fig. 351-3), clinical experience is insufficient to exclude pheochromocytoma screening in these individuals. Radiographic studies such as MRI or CT scans generally are reserved for individuals with abnormal screening tests or symptoms suggestive of pheochromocytoma (Chap. 343). Women should be tested during pregnancy because undetected pheochromocytoma can cause maternal death during childbirth.

Measurement of serum calcium and parathyroid hormone levels every 2–3 years provides an adequate screen for hyperparathyroidism, except in families in which hyperparathyroidism is a prominent component, in which measurements should be made annually.

MEDULLARY THYROID CARCINOMA

Hereditary MTC is a multicentric disorder. Total thyroidectomy with a central lymph node dissection should be performed in children who carry the mutant gene. Incomplete thyroidectomy leaves the possibility of later transformation of residual C cells. The goal of early therapy is cure, and a strategy that does not accomplish this goal is shortsighted. Long-term follow-up studies indicate an excellent

outcome, with ~90% of children free of disease 15–20 years after surgery. In contrast, 15–25% of patients in whom the diagnosis is made on the basis of a palpable thyroid nodule die from the disease within 15–20 years.

In adults with MTC >1 cm in size, metastases to regional lymph nodes are common (>75%). Total thyroidectomy with central lymph node dissection and selective dissection of other regional chains provides the best chance for cure. In patients with extensive local metastatic disease in the neck, external radiation may prevent local recurrence or reduce tumor mass but is not curative. Chemotherapy with combinations of adriamycin, vincristine, cyclophosphamide, and dacarbazine may provide palliation. Clinical trials with small compounds (tyrosine kinase inhibitors) that interact with the ATP-binding pocket of the RET, vascular endothelial receptor, and type 2 and epidermal growth factor receptors and prevent phosphorylation have shown promise for treatment of hereditary and sporadic MTC. A phase I trial of vandetanib has shown that 45% of patients have a 30% or greater reduction of tumor size and prolongation of progression-free survival by at least 11 months. Similar phase II results have been observed for XL184, sunitinib, tipifarnib, and sorafenib, and phase II trials of E7080 and pazopanib are under way. It seems likely that one or more of these compounds will be approved for treatment of metastatic MTC within the next few years.

PHEOCHROMOCYTOMA

The long-term goal for management of pheochromocytoma is to prevent death and cardiovascular complications. Improvements in radiographic imaging of the adrenals make direct examination of the apparently normal contralateral gland during surgery less important, and the rapid evolution of laparoscopic abdominal or retroperitoneal surgery has simplified management of early pheochromocytoma. The major question is whether to remove both adrenal glands or remove only the affected adrenal at the time of primary surgery. Issues to be considered in making this decision include the possibility of malignancy (<15 reported cases), the high probability of developing pheochromocytoma in the apparently unaffected gland over an 8- to 10-year period, and the risks of adrenal insufficiency caused by removal of both glands (at least two deaths related to adrenal insufficiency have occurred in MEN 2 patients). Most clinicians recommend removing only the affected gland. If both adrenals are removed, glucocorticoid and mineralocorticoid replacement is mandatory. An alternative approach is to perform a cortical-sparing adrenalectomy, removing the pheochromocytoma and adrenal medulla and leaving the adrenal cortex behind. This approach is usually successful and eliminates the necessity for steroid hormone replacement in most patients, although the pheochromocytoma recurs in a small percentage.

HYPERPARATHYROIDISM

Hyperparathyroidism has been managed by one of two approaches. Removal of 3.5 glands with maintenance of the remaining half gland in the neck is the usual procedure. In families in which hyperparathyroidism is a prominent manifestation (almost always associated with a codon 634 *RET* mutation) and recurrence is common, total parathyroidectomy with transplantation of parathyroid tissue into the nondominant forearm is preferred. This approach is discussed above in the context of hyperparathyroidism associated with MEN 1.

Other Genetic Endocrine Tumor Syndromes

A number of mixed syndromes exist in which the neoplastic associations differ from those in MEN 1 or 2 (Table 351-1).

The cause of VHL syndrome—the association of central nervous system tumors, renal cell carcinoma, pheochromocytoma, and islet cell neoplasms—is a mutation in the *VHL* tumor-suppressor gene. Germ-line-inactivating mutations of the *VHL* gene cause tumor formation when there is additional loss or somatic mutation of the normal *VHL* allele in brain, kidney, pancreatic islet, or adrenal medullary cells. Missense mutations been identified in >40% of VHL families with pheochromocytoma, suggesting that families with this type of mutation should be surveyed routinely for pheochromocytoma. A point that may be useful in differentiating VHL from MEN 1 (overlapping features include islet cell tumor and rare pheochromocytoma) or MEN 2 (overlapping feature is pheochromocytoma) is that hyperparathyroidism rarely occurs in VHL.

The molecular defect in type 1 neurofibromatosis inactivates neurofibromin, a cell membrane–associated protein that normally activates a GTPase. Inactivation of this protein impairs GTPase and causes continuous activation of p21 Ras and its downstream tyrosine kinase pathway. Endocrine tumors also form in less common neoplastic genetic syndromes. These include Cowden disease, Carney complex, familial growth hormone and prolactin tumors, and familial carcinoid syndrome. Carney complex includes myxomas of the heart, skin, and breast; peripheral nerve schwannomas; spotty skin pigmentation; and testicular, adrenal, and GH-secreting pituitary tumors. Linkage analysis has identified two loci: chromosome 2p in half of the families and 17q in the others. The 17q gene has been identified as the regulatory subunit (type IA) of protein kinase A (*PRKA1A*). Familial growth hormone– or prolactin-producing neoplasms without other manifestations of MEN 1 are caused by germ-line-inactivating mutation of the aryl hydrocarbon receptor interacting protein (AIP). It is transmitted in an autosomal dominant manner. Other types of endocrine tumors have not, to date,

been associated with AIP mutations.

IMMUNOLOGIC SYNDROMES AFFECTING MULTIPLE ENDOCRINE ORGANS

When immune dysfunction affects two or more endocrine glands and other nonendocrine immune disorders are present, the *polyglandular autoimmune* (PGA) *syndromes* should be considered. The PGA syndromes are classified as two main types: the type I syndrome starts in childhood and is characterized by mucocutaneous candidiasis, hypoparathyroidism, and adrenal insufficiency; the type II, or Schmidt, syndrome is more likely to present in adults and most commonly includes adrenal insufficiency, thyroiditis, or type 1 diabetes mellitus. Some authors have attempted to subdivide PGA II on the basis of association with some autoimmune disorders but not others (i.e., type II and type III). The type III syndrome is heterogeneous and may consist of autoimmune thyroid disease along with a variety of other autoimmune endocrine disorders **(Table 351-2)**. However, little information is gained by making this subdivision in terms of understanding pathogenesis or prevention of future endocrine complications in individual patients or in the affected families.

Table 351-2 Features of Polyglandular Autoimmune (Pga) Syndromes

PGA I	PGA II
Epidemiology	
Autosomal recessive	Polygenic inheritance
Mutations in *APECED* gene	HLA-DR3 and HLA-DR4 associated
Childhood onset	Adult onset
Equal male:female ratio	Female predominance
Disease Associations	
Mucocutaneous candidiasis	Adrenal insufficiency
Hypoparathyroidism	Hypothyroidism
Adrenal insufficiency	Graves' disease
Hypogonadism	Type 1 diabetes
Alopecia	Hypogonadism
Hypothyroidism	Hypophysitis
Dental enamel hypoplasia	Myasthenia gravis
Malabsorption	Vitiligo
Chronic active hepatitis	Alopecia
Vitiligo	Pernicious anemia
Pernicious anemia	Celiac disease

Abbreviation: APECED, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy.

Polyglandular Autoimmune Syndrome Type I

PGA type I usually is recognized in the first decade of life and requires two of three components for diagnosis: mucocutaneous candidiasis, hypoparathyroidism, and adrenal insufficiency. Mucocutaneous candidiasis and hypoparathyroidism present with similar high frequency (100% and 79–96%, respectively). Adrenal insufficiency is observed in 60–72% of patients. Mineralocorticoids and glucocorticoids may be lost simultaneously or sequentially. PGA type 1 also is called *autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy* (APECED). Other endocrine defects can include gonadal failure (60% female, 14% male), hypothyroidism (5%), and destruction of the beta cells of the pancreatic islets and development of insulin-dependent (type 1) diabetes mellitus (14% lifetime risk). Additional features include hypoplasia of the dental enamel, nail dystrophy, tympanic membrane sclerosis, vitiligo, keratopathy, and gastric parietal cell dysfunction resulting in pernicious anemia (13%). Some patients develop autoimmune hepatitis (12%), malabsorption (variably attributed to intestinal lymphangiectasia, bacterial overgrowth, or hypoparathyroidism), asplenism, achalasia, and cholelithiasis (Table 351-2). At the outset, only one organ may be involved, but the number increases with time so that patients eventually manifest two to five components of the syndrome.

Most patients initially present with oral candidiasis in childhood; it is poorly responsive to treatment (Chap. 353) and relapses frequently. Chronic hypoparathyroidism usually occurs before adrenal insufficiency develops. More than 60% of postpubertal women develop premature hypogonadism. The endocrine components, including adrenal insufficiency and hypoparathyroidism, may not develop until the fourth decade, making continued surveillance necessary.

Type I PGA syndrome is not associated with a particular HLA type and usually is inherited as an autosomal recessive trait. It may occur sporadically. The responsible gene, designated as either *APECED* or *AIRE*, encodes a transcription factor that is expressed in thymus and lymph nodes; a

variety of different mutations have been reported. The mechanism by which these mutations lead to the diverse manifestations of type I PGA is unknown.

Polyglandular Autoimmune Syndrome Type II

PGA type II is characterized by two or more of the endocrinopathies listed in Table 351-2. Most often these endocrinopathies include primary adrenal insufficiency, Graves' disease or autoimmune hypothyroidism, type 1 diabetes mellitus, and primary hypogonadism. Because adrenal insufficiency is relatively rare, it is used frequently to define the presence of the syndrome. Among patients with adrenal insufficiency, type 1 diabetes mellitus coexists in 52% and autoimmune thyroid disease occurs in 69%. However, many patients with antimicrosomal and antithyroglobulin antibodies never develop abnormalities of thyroid function. Thus, increased antibody titers alone are poor predictors of future disease. Other associated conditions include hypophysitis, celiac disease (2–3%), atrophic gastritis, and pernicious anemia (13%). Vitiligo, which is caused by antibodies against the melanocyte, and alopecia are less common than in the type I syndrome. Mucocutaneous candidiasis does not occur. A few patients develop a late-onset, usually transient hypoparathyroidism caused by antibodies that compete with PTH for binding to the PTH receptor. Up to 25% of patients with myasthenia gravis and an even higher percentage who have myasthenia and a thymoma have PGA type II.

The type II syndrome is familial in nature, often transmitted as an autosomal dominant trait with incomplete penetrance. As in many of the individual autoimmune endocrinopathies, certain HL-DR3 and -DR4 alleles increase disease susceptibility; several different genes probably contribute to the expression of this syndrome.

A variety of autoantibodies are seen in PGA type II, including antibodies directed against (1) thyroid antigens such as thyroid peroxidase, thyroglobulin, and the thyroid-stimulating hormone (TSH) receptor; (2) adrenal side chain cleavage enzyme, steroid 21-hydroxylase, or ACTH receptor; and (3) pancreatic islet glutamic acid decarboxylase or the insulin receptor, among others. The roles of cytokines such as interferon and cell-mediated immunity are unclear.

Diagnosis

The clinical manifestations of adrenal insufficiency often develop slowly, may be difficult to detect, and can be fatal if not diagnosed and treated appropriately. Thus, prospective screening should be performed routinely in all patients and family members at risk for PGA types I and II. The most effective screening test for adrenal disease is a cosyntropin stimulation test (Chap. 342). A fasting blood glucose level can be obtained to screen for hyperglycemia. Additional screening tests should include measurements of TSH, luteinizing hormone, follicle-stimulating hormone, and, in men, testosterone levels. In families with suspected type I PGA syndrome, calcium and phosphorus levels should be measured. These screening studies should be performed every 1–2 years up to about age 50 in families with PGA type II syndrome and until about age 40 in patients with type I syndrome. Screening measurements of autoantibodies against potentially affected endocrine organs are of uncertain prognostic value. The differential diagnosis of PGA syndrome should include the DiGeorge syndrome (hypoparathyroidism due to glandular agenesis and mucocutaneous candidiasis), Kearns-Sayre syndrome (hypoparathyroidism, primary hypogonadism, type 1 diabetes mellitus, and panhypopituitarism), Wolfram's syndrome (congenital diabetes insipidus and diabetes mellitus), IPEX syndrome (*i*mmunodysregulation, *p*olyendocrinopathy, and *e*nteropathy, *X*-linked), and congenital rubella (type 1 diabetes mellitus and hypothyroidism).

TREATMENT: POLYGLANDULAR AUTOIMMUNE SYNDROME

With the exception of Graves' disease, the management of each of the endocrine components of the disease involves hormone replacement and is covered in detail in the chapters on adrenal, thyroid, gonadal, and parathyroid disease (Chaps. 341, 342, 346, 347, and 353). Some aspects of therapy merit special emphasis. Primary hypothyroidism can mask adrenal insufficiency by prolonging the half-life of cortisol; consequently, administration of thyroid hormone to a patient with unsuspected adrenal insufficiency can precipitate adrenal crisis. Thus, all patients with hypothyroidism in the context of PGA syndrome should be screened for adrenal disease and, if it is present, treated with glucocorticoids before or concurrently with thyroid hormone therapy. Hypoglycemia or decreasing insulin requirements in a patient with diabetes mellitus type 1 may be the earliest symptom of adrenal insufficiency. Consequently, such patients should be screened for adrenal disease. Treatment of mucocutaneous candidiasis with ketoconazole may induce adrenal insufficiency. This drug also may elevate liver enzymes, making the diagnosis of autoimmune hepatitis more difficult. Hypocalcemia in PGA type II is more commonly due to malabsorption associated with celiac disease than to hypoparathyroidism.

Other Autoimmune Endocrine Syndromes

INSULIN RESISTANCE CAUSED BY ANTIBODIES

Rare insulin-resistance syndromes occur in patients who develop antibodies that block the interaction of insulin with its receptor. Conversely, other classes of anti-insulin receptor antibodies can activate the receptor and can cause hypoglycemia; this disorder should be considered in the differential diagnosis of fasting hypoglycemia (Chap. 345).

Patients with insulin receptor antibodies and acanthosis nigricans are often middle-aged women who acquire insulin resistance in association with other autoimmune disorders, such as systemic lupus erythematosus and Sjögren's syndrome. Vitiligo, alopecia, Raynaud's phenomenon, and arthritis also may be seen. Other autoimmune endocrine disorders, including thyrotoxicosis, hypothyroidism, and hypogonadism, occur rarely. Acanthosis nigricans, a velvety, hyperpigmented, thickened skin lesion, is prominent on the dorsum of the neck and other skinfold areas in the axillae or groin and often heralds the diagnosis in these patients. However, acanthosis nigricans also occurs in patients with obesity or polycystic ovarian syndrome, in which insulin resistance appears to be due to a postreceptor defect; thus, acanthosis nigricans itself is not diagnostic of the immunologic form of insulin resistance.

Some patients with acanthosis nigricans have mild glucose intolerance, with a compensatory increase in insulin secretion that is detected only when insulin levels are measured. Others have severe diabetes mellitus that requires massive doses of insulin (several thousand units per day) to lower the blood glucose levels. The nature of the antibodies determines the manifestations; though insulin resistance is more common, fasting hypoglycemia can result from insulinomimetic antibodies.

Insulin-resistant diabetes mellitus associated with anti-insulin antibodies occurs in patients with ataxia-telangiectasia. This is an autosomal recessive disorder caused by mutations in *ATM*, a gene involved in cellular responses to ionizing radiation and oxidative damage. This disorder is characterized by ataxia, telangiectasia, immune abnormalities, and an increased incidence of malignancies.

AUTOIMMUNE INSULIN SYNDROME WITH HYPOGLYCEMIA

This disorder typically occurs in patients with other autoimmune disorders and is caused by polyclonal autoantibodies that bind to endogenously synthesized insulin. If the insulin dissociates from the antibodies several hours or more after a meal, hypoglycemia can result. Most cases of the syndrome have been described from Japan, and there may be a genetic component. In plasma cell dyscrasias such as multiple myeloma, the plasma cells may produce monoclonal antibodies against insulin and cause hypoglycemia by a similar mechanism.

ANTITHYROXINE ANTIBODIES AND HYPOTHYROIDISM

Circulating autoantibodies against thyroid hormones in patients with both immune thyroid disease and plasma cell dyscrasias such as Waldenström's macroglobulinemia can bind thyroid hormones, decrease their biologic activity, and result in primary hypothyroidism. In other patients the antibodies simply interfere with thyroid hormone immunoassays and cause false elevations or decreases in measured hormone levels.

CROW-FUKASE SYNDROME

The features of this syndrome are highlighted by an acronym that emphasizes its important features: *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M*-proteins, and *s*kin changes (POEMS). The most important feature is a severe, progressive sensorimotor polyneuropathy associated with a plasma cell dyscrasia. Localized collections of plasma cells (plasmacytomas) can cause sclerotic bone lesions and produce monoclonal IgG or IgA proteins. Endocrine manifestations in men or women include hyperprolactinemia, diabetes mellitus type 2, primary hypothyroidism, and adrenal insufficiency. Additional findings include ovarian failure and amenorrhea in women and testicular failure, impotence, and gynecomastia in men. Skin changes include hyperpigmentation, thickening of the dermis, hirsutism, and hyperhidrosis. Hepatomegaly and lymphadenopathy occur in about two-thirds of patients, and splenomegaly is seen in about one-third. Other manifestations include increased cerebrospinal fluid pressure with papilledema, peripheral edema, ascites, pleural effusions, glomerulonephritis, and fever. Median survival may be >10 years, though shorter in patients with extravascular volume overload or clubbing.

The systemic nature of the disorder may cause confusion with other connective tissue diseases. The endocrine manifestations suggest an autoimmune basis of the disorder, but circulating antibodies against endocrine cells have not been demonstrated. Increased serum and tissue levels of interleukin 6, interleukin 1β, vascular endothelial growth factor, matrix metalloproteins, and tumor necrosis factor α are present, but the pathophysiologic basis for the POEMS syndrome is uncertain. Therapy directed against the plasma cell dyscrasia such as local radiation of bony lesions, chemotherapy, thalidomide, plasmapheresis, bone marrow or stem cell transplantation, and treatment with all-*trans* retinoic acid may improve the endocrine manifestations.

FURTHER READINGS

Burgess J: How should the patient with multiple endocrine neoplasia type 1 (MEN 1) be followed? Clin Endocrinol (Oxf) 72:13, 2010[PMID: 19552677] [Full Text]

Falchetti A et al: Multiple endocrine neoplasia type 1 (MEN 1): Not only inherited endocrine tumors. Genet Med 11:825, 2009[PMID: 19904212] [Full Text]

Gracanin A et al: Tissue selectivity in multiple endocrine neoplasia type 1-associated tumorigenesis. Cancer Res 69:6371, 2009[PMID: 19654304] [Full Text]

Jensen RT et al: Inherited pancreatic endocrine tumor syndromes: Advances in molecular pathogenesis, diagnosis, management, and controversies. Cancer 113:1807, 2008[PMID: 18798544] [Full Text]

Jimenez C et al: Management of medullary thyroid carcinoma. Endocrinol Metab Clin North Am 37:481, 2008[PMID: 18502338] [Full Text]

Kahaly GJ: Polyglandular autoimmune syndromes. Eur J Endocrinol 161:11, 2009[PMID: 19411300] [Full Text]

Mathis D, Benoist C: Aire. Annu Rev Immunol 27:287, 2009[PMID: 19302042] [Full Text]

Vierimma O et al: Pituitary adenoma predisposition caused by germline mutations in the AIP gene. Science 312:1228, 2006

Ye L et al: The evolving field of tyrosine kinase inhibitors in the treatment of endocrine tumors. Endocrine Rev 31:578, 2010[PMID: 20605972] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BONE STRUCTURE AND METABOLISM

Bone is a dynamic tissue that is remodeled constantly throughout life. The arrangement of compact and cancellous bone provides strength and density suitable for both mobility and protection. In addition, bone provides a reservoir for calcium, magnesium, phosphorus, sodium, and other ions necessary for homeostatic functions. Bone also hosts and regulates hematopoiesis by providing niches for hematopoietic cell proliferation and differentiation. The skeleton is highly vascular and receives about 10% of the cardiac output. Remodeling of bone is accomplished by two distinct cell types: osteoblasts produce bone matrix, and osteoclasts resorb the matrix.

The extracellular components of bone consist of a solid mineral phase in close association with an organic matrix, of which 90–95% is type I collagen (Chap. 363). The noncollagenous portion of the organic matrix is heterogeneous and contains serum proteins such as albumin as well as many locally produced proteins, whose functions are incompletely understood. Those proteins include cell attachment/signaling proteins such as thrombospondin, osteopontin, and fibronectin; calcium-binding proteins such as matrix gla protein and osteocalcin; and proteoglycans such as biglycan and decorin. Some of the proteins organize collagen fibrils; others influence mineralization and binding of the mineral phase to the matrix.

The mineral phase is made up of calcium and phosphate and is best characterized as a poorly crystalline hydroxyapatite. The mineral phase of bone is deposited initially in intimate relation to the collagen fibrils and is found in specific locations in the "holes" between the collagen fibrils. This architectural arrangement of mineral and matrix results in a two-phase material well suited to withstand mechanical stresses. The organization of collagen influences the amount and type of mineral phase formed in bone. Although the primary structures of type I collagen in skin and bone tissues are similar, there are differences in posttranslational modifications and distribution of intermolecular cross-links. The holes in the packing structure of the collagen are larger in mineralized collagen of bone and dentin than in unmineralized collagens such as those in tendon. Single amino acid substitutions in the helical portion of either the α1 (*COL1A1*) or α2 (*COL1A2*) chains of type I collagen disrupt the organization of bone in osteogenesis imperfecta. The severe skeletal fragility associated with this group of disorders highlights the importance of the fibrillar matrix in the structure of bone (Chap. 363).

Osteoblasts synthesize and secrete the organic matrix. They are derived from cells of mesenchymal origin (Fig. 352-1*A*). Active osteoblasts are found on the surface of newly forming bone. As an osteoblast secretes matrix, which then is mineralized, the cell becomes an *osteocyte*, still connected with its blood supply through a series of canaliculi. Osteocytes account for the vast majority of the cells in bone. They are thought to be the mechanosensors in bone that communicate signals to surface osteoblasts and their progenitors through the canalicular network and thereby serve as master regulators of bone formation and resorption. Remarkably, osteocytes also secrete fibroblast growth factor 23 (FGF23), a major regulator of phosphate metabolism (see below). Mineralization of the matrix, both in trabecular bone and in osteones of compact cortical bone (*Haversian systems*), begins soon after the matrix is secreted (primary mineralization) but is not completed for several weeks or even longer (secondary mineralization). Although this mineralization takes advantage of the high concentrations of calcium and phosphate, already near saturation in serum, mineralization is a carefully regulated process that is dependent on the activity of osteoblast-derived alkaline phosphatase, which probably works by hydrolyzing inhibitors of mineralization.

Figure 352-1

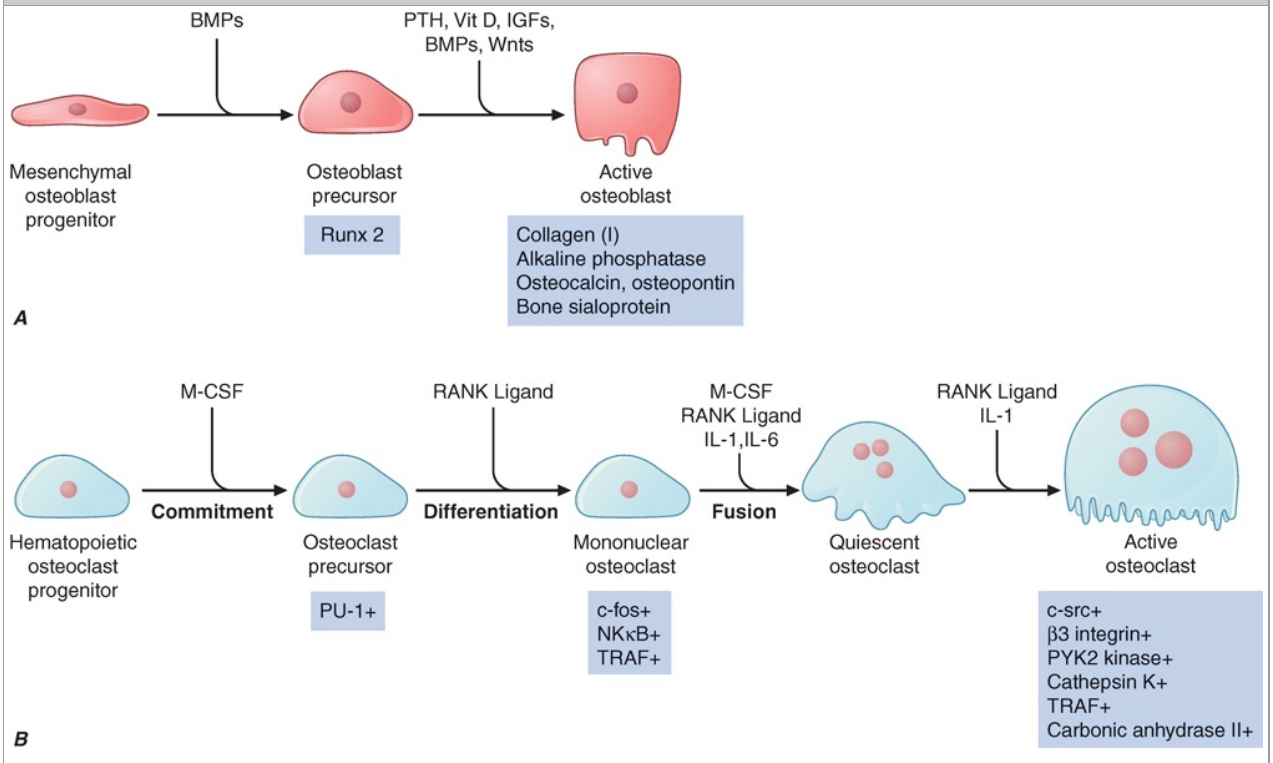

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathways regulating development of*A.*osteoblasts and*B.*osteoclasts. Hormones, cytokines, and growth factors that control cell proliferation and differentiation are shown above the arrows. Transcription factors and other markers specific for various stages of development are depicted below the arrows. BMPs, bone morphogenic proteins; wnts, wingless-type mouse mammary tumor virus integration site; PTH, parathyroid hormone; Vit D, vitamin D; IGFs, insulin-like growth factors; Runx2, Runt-related transcription factor 2; M-CSF, macrophage colony-stimulating factor; PU-1, a monocyte- and B lymphocyte –specific ets family transcription factor; NFκB, nuclear factor κB; TRAF, tumor necrosis factor receptor–associated factors; RANK ligand, receptor activator of NFκB ligand; IL-1, interleukin 1; IL-6, interleukin 6. *(Modified from T Suda et al: Endocr Rev 20:345, 1999, with permission.)*

Genetic studies in humans and mice have identified several key genes that control osteoblast development. *Runx2* is a transcription factor expressed specifically in chondrocyte (cartilage cells) and osteoblast progenitors as well as in hypertrophic chondrocytes and mature osteoblasts. *Runx2* regulates the expression of several important osteoblast proteins, including osterix (another transcription factor needed for osteoblast maturation), osteopontin, bone sialoprotein, type I collagen, osteocalcin, and receptor-activator of NFκB (RANK) ligand. *Runx2* expression is regulated in part by bone morphogenic proteins (BMPs). *Runx2*-deficient mice are devoid of osteoblasts, whereas mice with a deletion of only one allele (*Runx2* +/−) exhibit a delay in formation of the clavicles and some cranial bones. The latter abnormalities are similar to those in the human disorder *cleidocranial dysplasia*, which is also caused by heterozygous inactivating mutations in *Runx2*.

The paracrine signaling molecule, Indian hedgehog (Ihh), also plays a critical role in osteoblast development, as evidenced by Ihh-deficient mice that lack osteoblasts in bone formed on a cartilage mold (endochondral ossification). Signals originating from members of the wnt (wingless-type mouse mammary tumor virus integration site) family of paracrine factors are also important for osteoblast proliferation and differentiation. Numerous other growth-regulatory factors affect osteoblast function, including the three closely related transforming growth factor βs, fibroblast growth factors (FGFs) 2 and 18, platelet-derived growth factor, and insulin-like growth factors (IGFs) I and II. Hormones such as parathyroid hormone (PTH) and 1,25-dihydroxyvitamin D [1,25$(OH)_2$D] activate receptors expressed by osteoblasts to assure mineral homeostasis and influence a variety of bone cell functions.

Resorption of bone is carried out mainly by *osteoclasts*, multinucleated cells that are formed by fusion of cells derived from the common precursor of macrophages and osteoclasts. Multiple factors that regulate osteoclast development have been identified (Fig. 352-1*B*). Factors produced by osteoblasts or marrow stromal cells allow osteoblasts to control osteoclast development and activity. Macrophage colony-stimulating factor (M-CSF) plays a critical role during several steps in the pathway and ultimately leads to fusion of osteoclast progenitor cells to form multinucleated, active osteoclasts. RANK ligand, a member of the tumor necrosis factor (TNF) family, is expressed on the surface of osteoblast progenitors and stromal fibroblasts. In a process involving cell-cell interactions, RANK ligand binds to the RANK receptor on osteoclast progenitors, stimulating osteoclast differentiation and activation. Alternatively, a soluble decoy receptor, referred to as osteoprotegerin, can bind RANK ligand and inhibit osteoclast differentiation. Several growth factors and cytokines (including interleukins 1, 6, and 11; TNF; and interferon γ) modulate osteoclast differentiation and function. Most hormones that influence osteoclast function do not target these cells directly but instead influence M-CSF and RANK ligand signaling by osteoblasts. Both PTH and 1,25$(OH)_2$D increase osteoclast number and activity, whereas estrogen decreases osteoclast number and activity by this indirect mechanism. Calcitonin, in contrast, binds to its receptor on the basal surface of osteoclasts and directly inhibits osteoclast function.

Osteoclast-mediated resorption of bone takes place in scalloped spaces (*Howship's lacunae*) where the osteoclasts are attached through a specific $\alpha_V\beta_3$ integrin to components of the bone matrix such as osteopontin. The osteoclast forms a tight seal to the underlying matrix and secretes protons, chloride, and proteinases into a confined space that has been likened to an extracellular lysosome. The active osteoclast surface forms a ruffled border that contains a specialized proton-pump ATPase, which secretes acid and solubilizes the mineral phase. Carbonic anhydrase (type II isoenzyme) within the osteoclast generates the needed protons. The bone matrix is resorbed in the acid environment adjacent to the ruffled border by proteases that act at low pH such as cathepsin K.

In the embryo and the growing child, bone develops by remodeling and replacing previously calcified cartilage (endochondral bone formation) or is formed without a cartilage matrix (intramembranous bone formation). During endochondral bone formation, chondrocytes proliferate, secrete and mineralize a matrix, enlarge (hypertrophy), and then die, enlarging bone and providing the matrix and factors that stimulate endochondral bone formation. This program is regulated by both local factors such as IGF-I and -II, Ihh, parathyroid hormone–related peptide (PTHrP), and FGFs and by systemic hormones such as growth hormone, glucocorticoids, and estrogen.

New bone, whether formed in infants or in adults during repair, has a relatively high ratio of cells to matrix and is characterized by coarse fiber bundles of collagen that are interlaced and randomly dispersed (woven bone). In adults, the more mature bone is organized with fiber bundles regularly arranged in parallel or concentric sheets (lamellar bone). In long bones, deposition of lamellar bone in a concentric arrangement around blood vessels forms the Haversian systems. Growth in length of bones is dependent on proliferation of cartilage cells and the endochondral sequence at the growth plate. Growth in width and thickness is accomplished by formation of bone at the periosteal surface and by resorption at the endosteal surface, with the rate of formation exceeding that of resorption. In adults, after the growth plates close, growth in length and endochondral bone formation cease except for some activity in the cartilage cells beneath the articular surface. Even in adults, however, remodeling of bone (within Haversian systems as well as along the surfaces of trabecular bone) continues throughout life. In adults, ~4% of the surface of trabecular bone (such as iliac crest) is involved in active resorption, whereas 10–15% of trabecular surfaces are covered with osteoid, unmineralized new bone formed by osteoblasts. Radioisotope studies indicate that as much as 18% of the total skeletal calcium is deposited and removed each year. Thus, bone is an active metabolizing tissue that requires an intact blood supply. The cycle of bone resorption and formation is a highly orchestrated process carried out by the basic multicellular unit, which is composed of a group of osteoclasts and osteoblasts (Fig. 352-2).

Figure 352-2

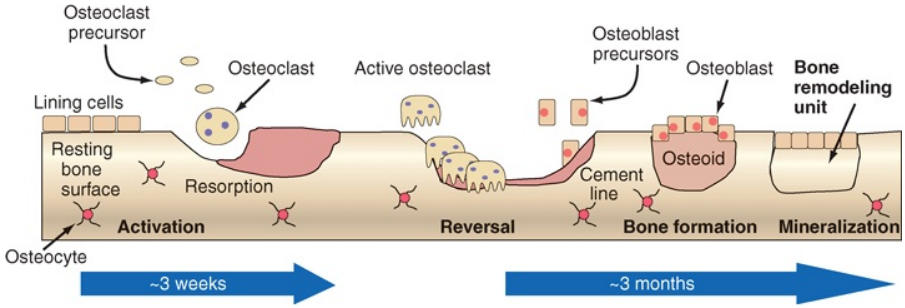

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of bone remodeling. The cycle of bone remodeling is carried out by the basic multicellular unit (BMU), which consists of a group of osteoclasts and osteoblasts. In cortical bone, the BMUs tunnel through the tissue, whereas in cancellous bone, they move across the trabecular surface. The process of bone remodeling is initiated by contraction of the lining cells and the recruitment of osteoclast precursors. These precursors fuse to form multinucleated, active osteoclasts that mediate bone resorption. Osteoclasts adhere to bone and subsequently remove it by acidification and proteolytic digestion. As the BMU advances, osteoclasts leave the resorption site and osteoblasts move in to cover the excavated area and begin the process of new bone formation by secreting osteoid, which eventually is mineralized into new bone. After osteoid mineralization, osteoblasts flatten and form a layer of lining cells over new bone.

The response of bone to fractures, infection, and interruption of blood supply and to expanding lesions is relatively limited. Dead bone must be resorbed, and new bone must be formed, a process carried out in association with growth of new blood vessels into the involved area. In injuries that disrupt the organization of the tissue such as a fracture in which apposition of fragments is poor or when motion exists at the fracture site, the progenitor stromal cells recapitulate the endochondral bone formation of early development and form cartilage that is replaced by bone and, variably, fibrous tissue. When there is good apposition with fixation and little motion at the fracture site, repair occurs predominantly by formation of new bone without other mediating tissue.

Remodeling of bone occurs along lines of force generated by mechanical stress. The signals from these mechanical stresses are sensed by osteocytes, which transmit signals to osteoclasts and osteoblasts or their precursors. One such signal is sclerostin, an inhibitor of wnt signaling. Mechanical forces suppress sclerostin production and thus increase bone formation by osteoblasts. Expanding lesions in bone such as tumors induce resorption at the surface in contact with the tumor by producing ligands such as PTHrP that stimulate osteoclast differentiation and function. Even in a disorder as architecturally disruptive as Paget's disease, remodeling is dictated by mechanical forces. Thus, bone plasticity reflects the interaction of cells with each other and with the environment.

Measurement of the products of osteoblast and osteoclast activity can assist in the diagnosis and management of bone diseases. Osteoblast activity can be assessed by measuring serum bone-specific alkaline phosphatase. Similarly, osteocalcin, a protein secreted from osteoblasts, is made virtually only by osteoblasts. Osteoclast activity can be assessed by measurement of products of collagen degradation. Collagen molecules are covalently linked to each other in the extracellular matrix through the formation of hydroxypyridinium cross-links (Chap. 363). After digestion by osteoclasts, these cross-linked peptides can be measured both in urine and in blood.

CALCIUM METABOLISM

Over 99% of the 1–2 kg of calcium present normally in the adult human body resides in the skeleton, where it provides mechanical stability and serves as a reservoir sometimes needed to maintain extracellular fluid (ECF) calcium concentration (Fig. 352-3). Skeletal calcium accretion first becomes significant during the third trimester of fetal life, accelerates throughout childhood and adolescence, reaches a peak in early adulthood, and gradually declines thereafter at rates that rarely exceed 1–2% per year. These slow changes in total skeletal calcium content contrast with relatively high daily rates of closely matched fluxes of calcium into and out of bone (~250–500 mg each), a process mediated by coupled osteoblastic and osteoclastic activity. Another 0.5–1% of skeletal calcium is freely exchangeable (e.g., in chemical equilibrium) with that in the ECF.

Figure 352-3

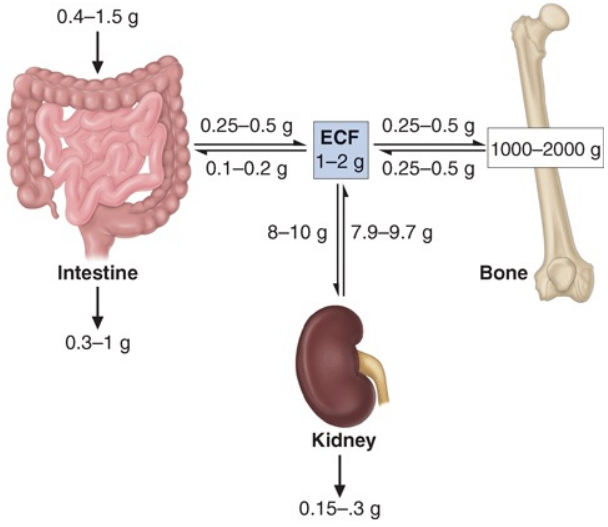

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Calcium homeostasis. Schematic illustration of calcium content of extracellular fluid (ECF) and bone as well as of diet and

Calcium Homeostasis. Schematic illustration of calcium content of extracellular fluid (ECF) and bone as well as of diet and feces; magnitude of calcium flux per day as calculated by various methods is shown at sites of transport in intestine, kidney, and bone. Ranges of values shown are approximate and were chosen to illustrate certain points discussed in the text. In conditions of calcium balance, rates of calcium release from and uptake into bone are equal.

The concentration of ionized calcium in the ECF must be maintained within a narrow range because of the critical role it plays in a wide array of cellular functions, especially those involved in neuromuscular activity, secretion, and signal transduction. Intracellular cytosolic free calcium levels are ~100 nmol/L and are 10,000-fold lower than ionized calcium concentration in the blood and ECF (1.1–1.3 mmol/L). Cytosolic calcium does not play the structural role played by extracellular calcium; instead, it serves a signaling function. The steep chemical gradient of calcium from outside to inside the cell promotes rapid calcium influx through various membrane calcium channels that can be activated by hormones, metabolites, or neurotransmitters, swiftly changing cellular function. In blood, total calcium concentration is normally 2.2–2.6 m*M* (8.5–10.5 mg/dL), of which ~50% is ionized. The remainder is bound ionically to negatively charged proteins (predominantly albumin and immunoglobulins) or loosely complexed with phosphate, citrate, sulfate, or other anions. Alterations in serum protein concentrations directly affect the total blood calcium concentration even if the ionized calcium concentration remains normal. An algorithm to correct for protein changes adjusts the total serum calcium (in mg/dL) upward by 0.8 times the deficit in serum albumin (g/dL) or by 0.5 times the deficit in serum immunoglobulin (in g/dL). Such corrections provide only rough approximations of actual free calcium concentrations, however, and may be misleading, particularly during acute illness. Acidosis also alters ionized calcium by reducing its association with proteins. The best practice is to measure blood ionized calcium directly by a method that employs calcium-selective electrodes in acute settings during which calcium abnormalities might occur.

Control of the ionized calcium concentration in the ECF ordinarily is accomplished by adjusting the rates of calcium movement across intestinal and renal epithelia. These adjustments are mediated mainly via changes in blood levels of the hormones PTH and $1,25(OH)_2D$. Blood ionized calcium directly suppresses PTH secretion by activating parathyroid calcium-sensing receptors (CaSRs). Also, ionized calcium indirectly affects PTH secretion via effects on $1,25(OH)_2D$ production. This active vitamin D metabolite inhibits PTH production by an incompletely understood mechanism of negative feedback (Chap. 353).

Normal dietary calcium intake in the United States varies widely, ranging from 10–37 mmol/d (400–1500 mg/d). Many individuals, in an effort to prevent osteoporosis, routinely supplement this with oral calcium salts to a total intake of 37–50 mmol/d (1500–2000 mg/d). Intestinal absorption of ingested calcium involves both active (transcellular) and passive (paracellular) mechanisms. Passive calcium absorption is nonsaturable and approximates 5% of daily calcium intake, whereas active absorption involves apical calcium entry via specific ion channels (TRPV5 and TRPV6), whose expression is controlled principally by $1,25(OH)_2D$, and normally ranges from 20 to 70%. Active calcium transport occurs mainly in the proximal small bowel (duodenum and proximal jejunum), although some active calcium absorption occurs in most segments of the small intestine. Optimal rates of calcium absorption require gastric acid. This is especially true for weakly dissociable calcium supplements such as calcium carbonate. In fact, large boluses of calcium carbonate are poorly absorbed because of their neutralizing effect on gastric acid. In achlorhydric subjects and for those taking drugs that inhibit gastric acid secretion, supplements should be taken with meals to optimize their absorption. Use of calcium citrate may be preferable in these circumstances. Calcium absorption may also be blunted in disease states such as pancreatic or biliary insufficiency, in which ingested calcium remains bound to unabsorbed fatty acids or other food constituents. At high levels of calcium intake, synthesis of $1,25(OH)_2D$ is reduced; this decreases the rate of active intestinal calcium absorption. The opposite occurs with dietary calcium restriction. Some calcium, ~2.5–5 mmol/d (100–200 mg/d), is excreted as an obligate component of intestinal secretions and is not regulated by calciotropic hormones.

The feedback-controlled hormonal regulation of intestinal absorptive efficiency results in a relatively constant daily net calcium absorption of ~5–7.5 mmol/d (200–400 mg/d) despite large changes in daily dietary calcium intake. This daily load of absorbed calcium is excreted by the kidneys in a manner that is also tightly regulated by the concentration of ionized calcium in the blood. Approximately 8–10 g/d of calcium is filtered by the glomeruli, of which only 2–3% appears in the urine. Most filtered calcium (65%) is reabsorbed in the proximal tubules via a passive, paracellular route that is coupled to concomitant NaCl reabsorption and not specifically regulated. The cortical thick ascending limb of Henle's loop (cTAL) reabsorbs roughly another 20% of filtered calcium, also via a paracellular mechanism. Calcium reabsorption in the cTAL requires a tight-junctional protein called paracellin-1 and is inhibited by increased blood concentrations of calcium or magnesium, acting via the CaSR, which is highly expressed on basolateral membranes in this nephron segment. Operation of the renal CaSR provides a mechanism, independent of those engaged directly by PTH or $1,25(OH)_2D$, by which serum ionized calcium can control renal calcium reabsorption. Finally, ~10% of filtered calcium is reabsorbed in the distal convoluted tubules (DCTs) by a transcellular mechanism. Calcium enters the luminal surface of the cell through specific apical calcium channels (TRPV5), whose number is regulated. It then moves across the cell in association with a specific calcium-binding protein (calbindin-D28k) that buffers cytosolic calcium concentrations from the large mass of transported calcium. Ca^{2+}-ATPases and Na^+/Ca^{2+} exchangers actively extrude calcium across the basolateral surface and thereby maintain the transcellular calcium gradient. All these processes are stimulated directly or indirectly by PTH. The DCT is also the site of action of thiazide diuretics, which lower urinary calcium excretion by inducing sodium depletion and thereby augmenting proximal calcium reabsorption. Conversely, dietary sodium loads, or increased distal sodium delivery caused by loop diuretics or saline infusion, induce calciuresis.

The homeostatic mechanisms that normally maintain a constant serum ionized calcium concentration may fail at extremes of calcium intake or when the hormonal systems or organs involved are compromised. Thus, even with maximal activity of the vitamin D–dependent intestinal active transport system, sustained calcium intakes <5 mmol/d (<200 mg/d) cannot provide enough net calcium absorption to replace obligate losses via the intestine, the kidney, sweat, and other secretions. In this case, increased blood levels of PTH and $1,25(OH)_2D$ activate osteoclastic bone resorption to obtain needed calcium from bone, which leads to progressive bone loss and negative calcium balance. Increased PTH and $1,25(OH)_2D$ also enhance renal calcium reabsorption, and $1,25(OH)_2D$ enhances calcium absorption in the gut. At very high calcium intakes [>100 mmol/d (>4 g/d)], passive intestinal absorption continues to deliver calcium into the ECF despite maximally downregulated intestinal active transport and renal tubular calcium reabsorption. This can cause severe hypercalciuria, nephrocalcinosis, progressive renal failure, and hypercalcemia (e.g., "milk-alkali syndrome"). Deficiency or excess of PTH or vitamin D, intestinal disease, and renal failure represent other commonly encountered challenges to normal calcium homeostasis (Chap. 353).

PHOSPHORUS METABOLISM

Although 85% of the ~600 g of body phosphorus is present in bone mineral, phosphorus is also a major intracellular constituent both as the free anion(s) and as a component of numerous organophosphate compounds, including structural proteins, enzymes, transcription factors, carbohydrate and lipid intermediates, high-energy stores [ATP (adenosine triphosphate), creatine phosphate], and nucleic acids. Unlike calcium, phosphorus exists intracellularly at concentrations close to those present in ECF (e.g., 1–2 mmol/L). In cells and in the ECF, phosphorus exists in several forms, predominantly as $H_2PO_4^-$ or $NaHPO_4^-$, with perhaps 10% as HPO_4^{2-}. This mixture of anions will be referred to here as "phosphate." In serum, about 12% of phosphorus is bound to proteins. Concentrations of phosphates in blood and ECF generally are expressed in terms of

elemental phosphorus, with the normal range in adults being 0.75–1.45 mmol/L (2.5–4.5 mg/dL). Because the volume of the intracellular fluid compartment is twice that of the ECF, measurements of ECF phosphate may not accurately reflect phosphate availability within cells that follows even modest shifts of phosphate from one compartment to the other.

Phosphate is widely available in foods and is absorbed efficiently (65%) by the small intestine even in the absence of vitamin D. However, phosphate absorptive efficiency may be enhanced (to 85–90%) via active transport mechanisms that are stimulated by $1,25(OH)_2D$. These mechanisms involve activation of Na^+/PO_4^{2-} co-transporters that move phosphate into intestinal cells against an unfavorable electrochemical gradient. Daily net intestinal phosphate absorption varies widely with the composition of the diet but is generally in the range of 500–1000 mg/d. Phosphate absorption can be inhibited by large doses of calcium salts or by sevelamer hydrochloride (Renagel), strategies commonly used to control levels of serum phosphate in renal failure. Aluminum hydroxide antacids also reduce phosphate absorption but are used less commonly because of the potential for aluminum toxicity. Low serum phosphate stimulates renal proximal tubular synthesis of $1,25(OH)_2D$, perhaps by suppressing blood levels of FGF23 (see below).

Serum phosphate levels vary by as much as 50% on a normal day. This reflects the effect of food intake but also an underlying circadian rhythm that produces a nadir between 7 and 10 a.m. Carbohydrate administration, especially as IV dextrose solutions in fasting subjects, can decrease serum phosphate by >0.7 mmol/L (2 mg/dL) due to rapid uptake into and utilization by cells. A similar response is observed in the treatment of diabetic ketoacidosis and during metabolic or respiratory alkalosis. Because of this wide variation in serum phosphate, it is best to perform measurements in the basal, fasting state.

Control of serum phosphate is determined mainly by the rate of renal tubular reabsorption of the filtered load, which is ~4–6 g/d. Because intestinal phosphate absorption is highly efficient, urinary excretion is not constant but varies directly with dietary intake. The fractional excretion of phosphate (ratio of phosphate to creatinine clearance) is generally in the range of 10–15%. The proximal tubule is the principal site at which renal phosphate reabsorption is regulated. This is accomplished by changes in the levels of apical expression and activity of specific Na^+/PO_4^{2-} co-transporters (NaPi-2 and NaPi-2c) in the proximal tubule. Levels of these transporters at the apical surface of these cells are reduced rapidly by PTH, the major known hormonal regulator of renal phosphate excretion. FGF23 can impair phosphate reabsorption dramatically by a similar mechnism. Activating *FGF23* mutations cause the rare disorder autosomal dominant hypophosphatemic rickets. In contrast to PTH, FGF23 also leads to reduced synthesis of $1,25(OH)_2D$, which may worsen the resulting hypophosphatemia by lowering intestinal phosphate absorption. Renal reabsorption of phosphate is responsive to changes in dietary intake such that experimental dietary phosphate restriction leads to a dramatic lowering of urinary phosphate within hours, preceding any decline in serum phosphate (e.g., filtered load). This physiologic renal adaptation to changes in dietary phosphate availability occurs independently of PTH and may be mediated in part by changes in levels of serum FGF23. Findings in *FGF23*-knockout mice suggest that FGF23 normally acts to lower blood phosphate and $1,25(OH)_2D$ levels. In turn, elevation of blood phosphate increases blood levels of FGF23.

Renal phosphate reabsorption is impaired by hypocalcemia, hypomagnesemia, and severe hypophosphatemia. Phosphate clearance is enhanced by ECF volume expansion and impaired by dehydration. Phosphate retention is an important pathophysiologic feature of renal insufficiency (Chap. 280).

Hypophosphatemia

CAUSES

Hypophosphatemia can occur by one or more of three primary mechanisms: (1) inadequate intestinal phosphate absorption, (2) excessive renal phosphate excretion, and (3) rapid redistribution of phosphate from the ECF into bone or soft tissue (Table 352-1). Because phosphate is so abundant in foods, inadequate intestinal absorption is almost never observed now that aluminum hydroxide antacids, which bind phosphate in the gut, are no longer used commonly. Fasting or starvation, however, may result in depletion of body phosphate and predispose to subsequent hypophosphatemia during refeeding, especially if this is accomplished with IV glucose alone.

Table 352-1 Causes of Hypophosphatemia

I. Reduced renal tubular phosphate reabsorption
A. PTH/PTHrP-dependent
1. Primary hyperparathyroidism
2. Secondary hyperparathyroidism
a. Vitamin D deficiency/resistance
b. Calcium starvation/malabsorption
c. Bartter's syndrome
d. Autosomal recessive renal hypercalciuria with hypomagnesemia
3. PTHrP-dependent hypercalcemia of malignancy
4. Familial hypocalciuric hypercalcemia
B. PTH/PTHrP-independent
1. Excess FGF23 or other "phosphatonins"
a. X-linked hypophosphatemic rickets (XLHR)
b. Autosomal recessive hypophosphatemia (ARHP)
c. Autosomal dominant hypophosphatemic rickets (ADHR)
d. Tumor-induced osteomalacia syndrome (TIO)
e. McCune-Albright syndrome (fibrous dysplasia)
f. Epidermal nevus syndrome
2. Intrinsic renal disease
a. Fanconi's syndrome(s)
b. Cystinosis
c. Wilson's disease
d. NaPi-2a or NaPi-2c mutations
3. Other systemic disorders
a. Poorly controlled diabetes mellitus

b. Alcoholism
c. Hyperaldosteronism
d. Hypomagnesemia
e. Amyloidosis
f. Hemolytic-uremic syndrome
g. Renal transplantation or partial liver resection
h. Rewarming or induced hyperthermia
4. Drugs or toxins
a. Ethanol
b. Acetazolamide, other diuretics
c. High-dose estrogens or glucocorticoids
d. Heavy metals (lead, cadmium)
e. Toluene, *N*-methyl formamide
f. Cisplatin, ifosfamide, foscarnet, rapamycin
II. Impaired intestinal phosphate absorption
Aluminum-containing antacids
Sevalamer
III. Shifts of extracellular phosphate into cells
A. Intravenous glucose
B. Insulin therapy for prolonged hyperglycemia or diabetic ketoacidosis
C. Catecholamines (epinephrine, dopamine, albuterol)
D. Acute respiratory alkalosis
E. Gram-negative sepsis, toxic shock syndrome
F. Recovery from starvation or acidosis
G. Rapid cellular proliferation
1. Leukemic blast crisis
2. Intensive erythropoietin, other growth factor therapy
IV. Accelerated net bone formation
A. After parathyroidectomy
B. Treatment of vitamin D deficiency, Paget's disease
C. Osteoblastic metastases

Chronic hypophosphatemia usually signifies a persistent renal tubular phosphate-wasting disorder. Excessive activation of PTH/PTHrP receptors in the proximal tubule as a result of primary or secondary hyperparathyroidism or because of the PTHrP-mediated hypercalcemia syndrome in malignancy (Chap. 353) is among the more common causes of renal hypophosphatemia, especially because of the high prevalence of vitamin D deficiency in older Americans. Familial hypocalciuric hypercalcemia and Jansen's chondrodystrophy are rare examples of genetic disorders in this category (Chap. 353).

Several genetic and acquired diseases cause PTH/PTHrP-independent tubular phosphate wasting with associated rickets and osteomalacia. All these diseases manifest severe hypophosphatemia; renal phosphate wasting, sometimes accompanied by aminoaciduria; low blood levels of $1,25(OH)_2D$; low-normal serum levels of calcium; and evidence of impaired cartilage or bone mineralization. Analysis of these diseases has led to the discovery of the hormone FGF23, which is an important physiologic regulator of phosphate metabolism. FGF23 decreases phosphate reabsorption in the proximal tubule and also suppresses the 1α-hydroxylase responsible for synthesis of $1,25(OH)_2D$. FGF23 is synthesized by cells of the osteoblast lineage, primarily osteocytes. High-phosphate diets increase FGF23 levels, and low-phosphate diets decrease them. Autosomal dominant hypophosphatemic rickets (ADHR) was the first disease linked to abnormalities in FGF23. ADHR results from activating mutations in the gene that encodes FGF23. The most common inherited cause of hypophosphatemia is X-linked hypophosphatemic rickets (XLHR), which results from inactivating mutations in an endopeptidase termed *PHEX* (*p*hosphate-regulating gene with *h*omologies to *e*ndopeptidases on the *X* chromosome) that is expressed most abundantly on the surface of osteocytes and mature osteoblasts. Patients with XLH usually have high FGF23 levels, and ablation of the FGF23 gene reverses the hypophosphatemia found in the mouse version of XLH. How inactivation of PHEX leads to increased levels of FGF23 has not been determined. A third hypophosphatemic disorder, tumor-induced osteomalacia (TIO), is an acquired disorder in which tumors, usually of mesenchymal origin and generally histologically benign, secrete molecules that induce renal phosphate wasting. The hypophosphatemic syndrome resolves completely within hours to days after successful resection of the responsible tumor. Such tumors express large amounts of FGF23 mRNA, and patients with TIO usually exhibit elevations of FGF23 in their blood.

Dent's disease is an X-linked recessive disorder caused by inactivating mutations in CLCN5, a chloride transporter expressed in endosomes of the proximal tubule; features include hypercalciuria, hypophosphatemia, and recurrent kidney stones. Renal phosphate wasting is common among poorly controlled diabetic patients and alcoholics, who therefore are at risk for iatrogenic hypophosphatemia when treated with insulin or IV glucose, respectively. Diuretics and certain other drugs and toxins can cause defective renal tubular phosphate reabsorption (Table 352-1).

In hospitalized patients, hypophosphatemia is often attributable to massive redistribution of phosphate from the ECF into cells. Insulin therapy for diabetic ketoacidosis is a paradigm for this phenomenon, in which the severity of the hypophosphatemia is related to the extent of antecedent depletion of phosphate and other electrolytes (Chap. 344). The hypophosphatemia is usually greatest at a point many hours after initiation of insulin therapy and is difficult to predict from baseline measurements of serum phosphate at the time of presentation, when prerenal azotemia can obscure significant phosphate depletion. Other factors that may contribute to such acute redistributive hypophosphatemia include antecedent starvation or malnutrition, administration of IV glucose without other nutrients, elevated blood catecholamines (endogenous or exogenous), respiratory alkalosis, and recovery from metabolic acidosis.

Hypophosphatemia also can occur transiently (over weeks to months) during the phase of accelerated net bone formation that follows parathyroidectomy for severe primary hyperparathyroidism or during treatment of vitamin D deficiency or lytic Paget's disease. This is usually most prominent in patients who preoperatively have evidence of high bone turnover (e.g., high serum levels of alkaline phosphatase). Osteoblastic metastases can also lead to this syndrome.

CLINICAL AND LABORATORY FINDINGS

The clinical manifestations of severe hypophosphatemia reflect a generalized defect in cellular energy metabolism because of ATP depletion, a shift from oxidative phosphorylation toward glycolysis, and associated tissue or organ dysfunction. Acute, severe hypophosphatemia occurs mainly or exclusively in hospitalized patients with underlying serious medical or surgical illness and preexisting phosphate depletion due to excessive urinary losses, severe malabsorption, or malnutrition. Chronic hypophosphatemia tends to be less severe, with a clinical presentation dominated by musculoskeletal complaints such as bone pain, osteomalacia, pseudofractures, and proximal muscle weakness or, in children, rickets and short stature.

Neuromuscular manifestations of severe hypophosphatemia are variable but may include muscle weakness, lethargy, confusion, disorientation, hallucinations, dysarthria, dysphagia, oculomotor palsies, anisocoria, nystagmus, ataxia, cerebellar tremor, ballismus, hyporeflexia, impaired sphincter control, distal sensory deficits, paresthesia, hyperesthesia, generalized or Guillain-Barré–like ascending paralysis, seizures, coma, and even death. Serious sequelae such as paralysis, confusion, and seizures are likely only at phosphate concentrations <0.25 mmol/L (<0.8 mg/dL). Rhabdomyolysis may develop during rapidly progressive hypophosphatemia. The diagnosis of hypophosphatemia-induced rhabdomyolysis may be overlooked, as up to 30% of patients with acute hypophosphatemia (<0.7 m*M*) have creatine phosphokinase elevations that peak one to two days after the nadir in serum phosphate, when the release of phosphate from injured myocytes may have led to a near normalization of circulating levels of phosphate.

Respiratory failure and cardiac dysfunction, which are reversible with phosphate treatment, may occur at serum phosphate levels of 0.5–0.8 mmol/L (1.5–2.5 mg/dL). Renal tubular defects, including tubular acidosis, glycosuria, and impaired reabsorption of sodium and calcium, may occur. Hematologic abnormalities correlate with reductions in intracellular ATP and 2,3-diphosphoglycerate and may include erythrocyte microspherocytosis and hemolysis; impaired oxyhemoglobin dissociation; defective leukocyte chemotaxis, phagocytosis, and bacterial killing; and platelet dysfunction with spontaneous gastrointestinal hemorrhage.

TREATMENT: HYPOPHOSPHATEMIA

Severe hypophosphatemia [<0.75 mmol/L (<2 mg/dL)], particularly in the setting of underlying phosphate depletion, constitutes a dangerous electrolyte abnormality that should be corrected promptly. Unfortunately, the cumulative deficit in body phosphate cannot be predicted easily from knowledge of the circulating level of phosphate, and therapy must be approached empirically. The threshold for IV phosphate therapy and the dose administered should reflect consideration of renal function, the likely severity and duration of the underlying phosphate depletion, and the presence and severity of symptoms consistent with those of hypophosphatemia. In adults, phosphate may be safely administered IV as neutral mixtures of sodium and potassium phosphate salts at initial doses of 0.2–0.8 mmol/kg of elemental phosphorus over 6 hours (e.g., 10–50 mmol over 6 hours), with doses >20 mmol/6 hours reserved for those who have serum levels <0.5 mmol/L (1.5 mg/dL) and normal renal function. A suggested approach is presented in Table 352-2. Serum levels of phosphate and calcium must be monitored closely (every 6–12 hours) throughout treatment. It is necessary to avoid a serum calcium-phosphorus product >50 to reduce the risk of heterotopic calcification. Hypocalcemia, if present, should be corrected before administering IV phosphate. Less severe hypophosphatemia, in the range of 0.5–0.8 mmol/L (1.5–2.5 mg/dL), usually can be treated with oral phosphate in divided doses of 750–2000 mg/d as elemental phosphorus; higher doses can cause bloating and diarrhea.

Table 352-2 Intravenous Therapy for Hypophosphatemia

Consider			
Likely severity of underlying phosphate depletion			
Concurrent parenteral glucose administration			
Presence of neuromuscular, cardiopulmonary, or hematologic complications of hypophosphatemia			
Renal function [reduce dose by 50% if serum creatinine >220 μmol/L (>2.5 mg/dL)]			
Serum calcium level (correct hypocalcemia first; reduce dose by 50% in hypercalcemia)			
Guidelines			
Serum Phosphorus, mM (mg/dL)	**Rate of Infusion, mmol/h**	**Duration, h**	**Total Administered, mmol**
<0.8 (<2.5)	2	6	12
<0.5 (<1.5)	4	6	24
<0.3 (<1)	8	6	48

Rates shown are calculated for a 70-kg person; levels of serum calcium and phosphorus must be measured every 6 to 12 h during therapy; infusions can be repeated to achieve stable serum phosphorus levels >0.8 mmol/L (>2.5 mg/dL); most formulations available in the United States provide 3 mmol/mL of sodium or potassium phosphate.

Management of chronic hypophosphatemia requires knowledge of the cause(s) of the disorder. Hypophosphatemia related to the secondary hyperparathyroidism of vitamin D deficiency usually responds to treatment with vitamin D and calcium alone. XLHR, ADHR, TIO, and related renal tubular disorders usually are managed with divided oral doses of phosphate, often with calcium and $1,25(OH)_2D$ supplements to bypass the block in renal $1,25(OH)_2D$ synthesis and prevent secondary hyperparathyroidism caused by suppression of ECF calcium levels. Thiazide diuretics may be used to prevent nephrocalcinosis in patients who are managed this way. Complete normalization of hypophosphatemia is generally not possible in these conditions. Optimal therapy for TIO is extirpation of the responsible tumor, which may be localized by radiographic skeletal survey or bone scan (many are located in bone) or by radionuclide scanning using sestamibi or labeled octreotide. Successful treatment of TIO-induced hypophosphatemia with octreotide has been reported in a small number of patients.

Hyperphosphatemia

CAUSES

When the filtered load of phosphate and glomerular filtration rate (GFR) are normal, control of serum phosphate levels is achieved by adjusting the rate at which phosphate is reabsorbed by the proximal tubular NaPi-2 co-transporters. The principal hormonal regulators of NaPi-2 activity are PTH and FGF23. Hyperphosphatemia, defined in adults as a fasting serum phosphate concentration >1.8 mmol/L (5.5 mg/dL), usually results from impaired glomerular filtration, hypoparathyroidism, excessive delivery of phosphate into the ECF (from bone, gut, or parenteral phosphate therapy), or a combination of these factors (Table 352-3). The upper limit of normal serum phosphate concentrations is higher in children and neonates [2.4 mmol/L (7 mg/dL)]. It is useful to distinguish hyperphosphatemia caused by impaired renal phosphate excretion from that which results from excessive delivery of phosphate into the ECF (Table 352-3).

Table 352-3 Causes of Hyperphosphatemia

I. Impaired renal phosphate excretion
A. Renal insufficiency
B. Hypoparathyroidism
1. Developmental
2. Autoimmune
3. After neck surgery or radiation
4. Activating mutations of the calcium-sensing receptor
C. Parathyroid suppression
1. Parathyroid-independent hypercalcemia
a. Vitamin D or vitamin Aintoxication
b. Sarcoidosis, other granulomatous diseases
c. Immobilization, osteolytic metastases
d. Milk-alkali syndrome
2. Severe hypermagnesemia or hypomagnesemia
D. Pseudohypoparathyroidism
E. Acromegaly
F. Tumoral calcinosis
G. Heparin therapy
II. Massive extracellular fluid phosphate loads
A. Rapid administration of exogenous phosphate (intravenous, oral, rectal)
B. Extensive cellular injury or necrosis
1. Crush injuries
2. Rhabdomyolysis
3. Hyperthermia
4. Fulminant hepatitis
5. Cytotoxic therapy
6. Severe hemolytic anemia
C. Transcellular phosphate shifts
1. Metabolic acidosis
2. Respiratory acidosis

In chronic renal insufficiency, reduced GFR leads to phosphate retention. Hyperphosphatemia in turn further impairs renal synthesis of $1,25(OH)_2D$ and stimulates PTH secretion and hypertrophy both directly and indirectly (by lowering blood ionized calcium levels). Thus, hyperphosphatemia is a major cause of the secondary hyperparathyroidism of renal failure and must be addressed early in the course of the disease (Chaps. 280 and 353).

Hypoparathyroidism leads to hyperphosphatemia via increased expression of NaPi-2 co-transporters in the proximal tubule. Hypoparathyroidism, or parathyroid suppression, has multiple potential causes, including autoimmune disease; developmental, surgical, or radiation-induced absence of functional parathyroid tissue; vitamin D intoxication or other causes of PTH-independent hypercalcemia; cellular PTH resistance (pseudohypoparathyroidism or hypomagnesemia); infiltrative disorders such as Wilson's disease and hemochromatosis; and impaired PTH secretion caused by hypermagnesemia, severe hypomagnesemia, or activating mutations in the CaSR. Hypocalcemia may also contribute directly to impaired phosphate clearance, as calcium infusion can induce hyperphosphaturia in hypoparathyroid subjects. Increased tubular phosphate reabsorption also occurs in acromegaly, during heparin administration, and in tumoral calcinosis. Tumoral calcinosis is caused by a rare group of genetic disorders in which the *FGF23* gene is inactivated directly or FGF23 is processed in a way that leads to low levels of active FGF23 in the bloodstream. A similar syndrome results from FGF23 resistance due to inactivating mutations of the FGF23 co-receptor Klotho. These abnormalities cause elevated serum $1,25(OH)_2D$, parathyroid suppression, increased intestinal calcium absorption, and focal hyperostosis with large, lobulated periarticular heterotopic ossifications (especially at shoulders or hips) and are accompanied by hyperphosphatemia. In some forms of tumoral calcinosis serum phosphorus levels are normal.

When large amounts of phosphate are delivered rapidly into the ECF, hyperphosphatemia can occur despite normal renal function. Examples include overzealous IV phosphate therapy, oral or rectal administration of large amounts of phosphate-containing laxatives or enemas (especially in children), extensive soft tissue injury or necrosis (crush injuries, rhabdomyolysis, hyperthermia, fulminant hepatitis, cytotoxic chemotherapy), extensive hemolytic anemia, and transcellular phosphate shifts induced by severe metabolic or respiratory acidosis.

CLINICAL FINDINGS

The clinical consequences of acute, severe hyperphosphatemia are due mainly to the formation of widespread calcium phosphate precipitates and resulting hypocalcemia. Thus, tetany, seizures, accelerated nephrocalcinosis (with renal failure, hyperkalemia, hyperuricemia, and metabolic acidosis), and pulmonary or cardiac calcifications (including development of acute heart block) may occur. The severity of these complications relates to the elevation of serum phosphate levels, which can reach concentrations as high as 7 mmol/L (20 mg/dL) in instances of massive soft tissue injury or tumor lysis syndrome.

TREATMENT: HYPERPHOSPHATEMIA

Therapeutic options for management of severe hyperphosphatemia are limited. Volume expansion may enhance renal phosphate clearance. Aluminum hydroxide antacids or sevalamer may be helpful in chelating and limiting absorption of offending phosphate salts present in the intestine. Hemodialysis is the most effective therapeutic strategy and should be considered early in the course of severe hyperphosphatemia, especially in the setting of renal failure and symptomatic hypocalcemia.

MAGNESIUM METABOLISM

Magnesium is the major intracellular divalent cation. Normal concentrations of extracellular magnesium and calcium are crucial for normal neuromuscular activity. Intracellular magnesium forms a key complex with ATP and is an important cofactor for a wide range of enzymes, transporters, and nucleic acids required for normal cellular function, replication, and energy metabolism. The concentration of magnesium in serum is closely regulated within the range of 0.7–1 mmol/L (1.5–2 meq/L; 1.7–2.4 mg/dL), of

which 30% is protein-bound and another 15% is loosely complexed to phosphate and other anions. One-half of the 25 g (1000 mmol) of total body magnesium is located in bone, only one-half of which is insoluble in the mineral phase. Almost all extraskeletal magnesium is present within cells, where the total concentration is 5 m*M*, 95% of which is bound to proteins and other macromolecules. Because only 1% of body magnesium resides in the ECF, measurements of serum magnesium levels may not accurately reflect the level of total body magnesium stores.

Dietary magnesium content normally ranges from 6 to 15 mmol/d (140–360 mg/d), of which 30–40% is absorbed, mainly in the jejunum and ileum. Intestinal magnesium absorptive efficiency is stimulated by $1,25(OH)_2D$ and can reach 70% during magnesium deprivation. Urinary magnesium excretion normally matches net intestinal absorption and is ~4 mmol/d (100 mg/d). Regulation of serum magnesium concentrations is achieved mainly by control of renal magnesium reabsorption. Only 20% of filtered magnesium is reabsorbed in the proximal tubule, whereas 60% is reclaimed in the cTAL and another 5–10% in the DCT. Magnesium reabsorption in the cTAL occurs via a paracellular route that requires both a lumen-positive potential, created by NaCl reabsorption, and tight-junction proteins encoded by members of the Claudin gene family. Magnesium reabsorption in the cTAL is increased by PTH but inhibited by hypercalcemia or hypermagnesemia, both of which activate the CaSR in this nephron segment.

Hypomagnesemia

CAUSES

Hypomagnesemia usually signifies substantial depletion of body magnesium stores (0.5–1 mmol/kg). Hypomagnesemia can result from intestinal malabsorption; protracted vomiting, diarrhea, or intestinal drainage; defective renal tubular magnesium reabsorption; or rapid shifts of magnesium from the ECF into cells, bone, or third spaces (Table 352-4). Dietary magnesium deficiency is unlikely except possibly in the setting of alcoholism. A rare genetic disorder that causes selective intestinal magnesium malabsorption has been described (primary infantile hypomagnesemia). Another rare inherited disorder (hypomagnesemia with secondary hypocalcemia) is caused by mutations in the gene encoding TRPM6, a protein that, along with TRPM7, forms a channel important for both intestinal and distal-tubular renal magnesium transport. Malabsorptive states, often compounded by vitamin D deficiency, can critically limit magnesium absorption and produce hypomagnesemia despite the compensatory effects of secondary hyperparathyroidism and of hypocalcemia and hypomagnesemia to enhance cTAL magnesium reabsorption. Diarrhea or surgical drainage fluid may contain ≥5 mmol/L of magnesium.

Table 352-4 Causes of Hypomagnesemia

I. Impaired intestinal absorption
A. Hypomagnesemia with secondary hypocalcemia (TRPM6 mutations)
B. Malabsorption syndromes
C. Vitamin D deficiency
II. Increased intestinal losses
A. Protracted vomiting/diarrhea
B. Intestinal drainage, fistulas
III. Impaired renal tubular reabsorption
A. Genetic magnesium-wasting syndromes
1. Gitelman's syndrome
2. Bartter's syndrome
3. Claudin 16 or 19 mutations
4. Na+,K+-ATPase γ-subunit mutations (FXYD2)
5. Autosomal dominant, with low bone mass
B. Acquired renal disease
1. Tubulointerstitial disease
2. Postobstruction, ATN (diuretic phase)
3. Renal transplantation
C. Drugs and toxins
1. Ethanol
2. Diuretics (loop, -thiazide, osmotic)
3. Cisplatin
4. Pentamidine, foscarnet
5. Cyclosporine
6. Aminoglycosides, amphotericin B
7. Cetuximab
D. Other
1. Extracellular fluid volume expansion
2. Hyperaldosteronism
3. SIADH
4. Diabetes mellitus
5. Hypercalcemia
6. Phosphate depletion
7. Metabolic acidosis
8. Hyperthyroidism
IV. Rapid shifts from extracellular fluid
A. Intracellular redistribution
1. Recovery from diabetic ketoacidosis
2. Refeeding syndrome
3. Correction of respiratory acidosis
4. Catecholamines

B. Accelerated bone formation
1. Post-parathyroidectomy
2. Treatment of vitamin D deficiency
3. Osteoblastic -metastases
C. Other
1. Pancreatitis, burns, excessive sweating
2. Pregnancy (third trimester) and lactation

Abbreviations: ATN, acute tubular necrosis; SIADH, syndrome of inappropriate antidiuretic hormone.

Several genetic magnesium-wasting syndromes have been described, including inactivating mutations of genes encoding the DCT NaCl co-transporter (Gitelman's syndrome), proteins required for cTAL Na-K-2Cl transport (Bartter's syndrome), paracellin-1 (autosomal recessive renal hypomagnesemia with hypercalciuria), a DCT Na^+,K^+-ATPase γ-subunit (autosomal dominant renal hypomagnesemia with hypocalciuria), and a mitochondrial DNA gene encoding a mitochondrial tRNA. ECF expansion, hypercalcemia, and severe phosphate depletion may impair magnesium reabsorption, as can various forms of renal injury, including those caused by drugs such as cisplatin, cyclosporine, aminoglycosides, and pentamidine as well as the EGF receptor inhibitory antibody cetuximab (Table 352-4). A rising blood concentration of ethanol directly impairs tubular magnesium reabsorption, and persistent glycosuria with osmotic diuresis leads to magnesium wasting and probably contributes to the high frequency of hypomagnesemia in poorly controlled diabetic patients. Magnesium depletion is aggravated by metabolic acidosis, which causes intracellular losses as well.

Hypomagnesemia due to rapid shifts of magnesium from ECF into the intracellular compartment can occur during recovery from diabetic ketoacidosis, starvation, or respiratory acidosis. Less acute shifts may be seen during rapid bone formation after parathyroidectomy, with treatment of vitamin D deficiency, or with osteoblastic metastases. Large amounts of magnesium may be lost with acute pancreatitis, extensive burns, and protracted and severe sweating and during pregnancy and lactation.

CLINICAL AND LABORATORY FINDINGS

Hypomagnesemia may cause generalized alterations in neuromuscular function, including tetany, tremor, seizures, muscle weakness, ataxia, nystagmus, vertigo, apathy, depression, irritability, delirium, and psychosis. Patients are usually asymptomatic when serum magnesium concentrations are >0.5 mmol/L (1 meq/L; 1.2 mg/dL), although the severity of symptoms may not correlate with serum magnesium levels. Cardiac arrhythmias may occur, including sinus tachycardia, other supraventricular tachycardias, and ventricular arrhythmias. Electrocardiographic abnormalities may include prolonged PR or QT intervals, T-wave flattening or inversion, and ST straightening. Sensitivity to digitalis toxicity may be enhanced.

Other electrolyte abnormalities often seen with hypomagnesemia, including hypocalcemia (with hypocalciuria) and hypokalemia, may not be easily corrected unless magnesium is administered as well. The hypocalcemia may be a result of concurrent vitamin D deficiency, although hypomagnesemia can cause impaired synthesis of $1,25(OH)_2D$, cellular resistance to PTH, and, at very low serum magnesium [<0.4 mmol/L (<0.8 meq/L; <1 mg/dL)], a defect in PTH secretion; these abnormalities are reversible with therapy.

TREATMENT: HYPOMAGNESEMIA

Mild, asymptomatic hypomagnesemia may be treated with oral magnesium salts [$MgCl_2$, MgO, $Mg(OH)_2$] in divided doses totaling 20–30 mmol/d (40–60 meq/d). Diarrhea may occur with larger doses. More severe hypomagnesemia should be treated parenterally, preferably with IV $MgCl_2$, which can be administered safely as a continuous infusion of 50 mmol/d (100 meq Mg^{2+}/d) if renal function is normal. If GFR is reduced, the infusion rate should be lowered by 50–75%. Use of IM MgSO4 is discouraged; the injections are painful and provide relatively little magnesium (2 mL of 50% $MgSO_4$ supplies only 4 mmol). $MgSO_4$ may be given IV instead of $MgCl_2$, although the sulfate anions may bind calcium in serum and urine and aggravate hypocalcemia. Serum magnesium should be monitored at intervals of 12–24 hours during therapy, which may continue for several days because of impaired renal conservation of magnesium (only 50–70% of the daily IV magnesium dose is retained) and delayed repletion of intracellular deficits, which may be as high as 1–1.5 mmol/kg (2–3 meq/kg).

It is important to consider the need for calcium, potassium, and phosphate supplementation in patients with hypomagnesemia. Vitamin D deficiency frequently coexists and should be treated with oral or parenteral vitamin D or 25(OH)D [but not with $1,25(OH)_2D$, which may impair tubular magnesium reabsorption, possibly via PTH suppression]. In severely hypomagnesemic patients with concomitant hypocalcemia and hypophosphatemia, administration of IV magnesium alone may worsen hypophosphatemia, provoking neuromuscular symptoms or rhabdomyolysis, due to rapid stimulation of PTH secretion. This is avoided by administering both calcium and magnesium.

Hypermagnesemia

CAUSES

Hypermagnesemia is rarely seen in the absence of renal insufficiency, as normal kidneys can excrete large amounts (250 mmol/d) of magnesium. Mild hypermagnesemia due to excessive reabsorption in the cTAL occurs with calcium-sensing receptor mutations in familial hypocalciuric hypercalcemia and has been described in some patients with adrenal insufficiency, hypothyroidism, or hypothermia. Massive exogenous magnesium exposures, usually via the gastrointestinal tract, can overwhelm renal excretory capacity and cause life-threatening hypermagnesemia (Table 352-5). A notable example of this is prolonged retention of even normal amounts of magnesium-containing cathartics in patients with intestinal ileus, obstruction, or perforation. Extensive soft tissue injury or necrosis can also deliver large amounts of magnesium into the ECF in patients who have suffered trauma, shock, sepsis, cardiac arrest, or severe burns.

Table 352-5 Causes of Hypermagnesemia
I. Excessive magnesium intake
A. Cathartics, urologic irrigants
B. Parenteral magnesium administration
II. Rapid mobilization from soft tissues
A. Trauma, shock, sepsis
B. Cardiac arrest
C. Burns
III. Impaired magnesium excretion

A. Renal failure
B. Familial hypocalciuric hypercalcemia
IV. Other
A. Adrenal insufficiency
B. Hypothyroidism
C. Hypothermia
V. Excessive magnesium intake
A. Cathartics, urologic irrigants
B. Parenteral magnesium administration
VI. Rapid mobilization from soft tissues
A. Trauma, shock, sepsis
B. Cardiac arrest
C. Burns
VII. Impaired magnesium excretion
A. Renal failure
B. Familial hypocalciuric hypercalcemia
VIII. Other
A. Adrenal insufficiency
B. Hypothyroidism
C. Hypothermia

CLINICAL AND LABORATORY FINDINGS

The most prominent clinical manifestations of hypermagnesemia are vasodilation and neuromuscular blockade, which may appear at serum magnesium concentrations >2 mmol/L (>4 meq/L; >4.8 mg/dL). Hypotension that is refractory to vasopressors or volume expansion may be an early sign. Nausea, lethargy, and weakness may progress to respiratory failure, paralysis, and coma, with hypoactive tendon reflexes, at serum magnesium levels >4 mmol/L. Other findings may include gastrointestinal hypomotility or ileus; facial flushing; pupillary dilation; paradoxical bradycardia; prolongation of PR, QRS, and QT intervals; heart block; and, at serum magnesium levels approaching 10 mmol/L, asystole.

Hypermagnesemia, acting via the CaSR, causes hypocalcemia and hypercalciuria due to both parathyroid suppression and impaired cTAL calcium reabsorption.

TREATMENT: HYPERMAGNESEMIA

Successful treatment of hypermagnesemia generally involves identifying and interrupting the source of magnesium and employing measures to increase magnesium clearance from the ECF. Use of magnesium-free cathartics or enemas may be helpful in clearing ingested magnesium from the gastrointestinal tract. Vigorous IV hydration should be attempted, if appropriate. Hemodialysis is effective and may be required in patients with significant renal insufficiency. Calcium, administered IV in doses of 100–200 mg over 1–2 hours, has been reported to provide temporary improvement in signs and symptoms of hypermagnesemia.

VITAMIN D

Synthesis and Metabolism

1,25-dihydroxyvitamin D [1,25$(OH)_2$D] is the major steroid hormone involved in mineral ion homeostasis regulation. Vitamin D and its metabolites are hormones and hormone precursors rather than vitamins, since in the proper biologic setting, they can be synthesized endogenously (Fig. 352-4). In response to ultraviolet radiation of the skin, a photochemical cleavage results in the formation of vitamin D from 7-dehydrocholesterol. Cutaneous production of vitamin D is decreased by melanin and high solar protection factor sunblocks, which effectively impair skin penetration by ultraviolet light. The increased use of sunblocks in North America and Western Europe and a reduction in the magnitude of solar exposure of the general population over the last several decades has led to an increased reliance on dietary sources of vitamin D. In the United States and Canada, these sources largely consist of fortified cereals and dairy products, in addition to fish oils and egg yolks. Vitamin D from plant sources is in the form of vitamin D2, whereas that from animal sources is vitamin D3. These two forms have equivalent biologic activity and are activated equally well by the vitamin D hydroxylases in humans. Vitamin D enters the circulation, whether absorbed from the intestine or synthesized cutaneously, bound to vitamin D–binding protein, an α-globulin synthesized in the liver. Vitamin D is subsequently 25-hydroxylated in the liver by cytochrome P450–like enzymes in the mitochondria and microsomes. The activity of this hydroxylase is not tightly regulated, and the resultant metabolite, 25-hydroxyvitamin D [25(OH)D], is the major circulating and storage form of vitamin D. Approximately 88% of 25(OH)D circulates bound to the vitamin D–binding protein, 0.03% is free, and the rest circulates bound to albumin. The half-life of 25(OH)D is approximately two to three weeks; however, it is shortened dramatically when vitamin D–binding protein levels are reduced, as can occur with increased urinary losses in the nephrotic syndrome.

Figure 352-4

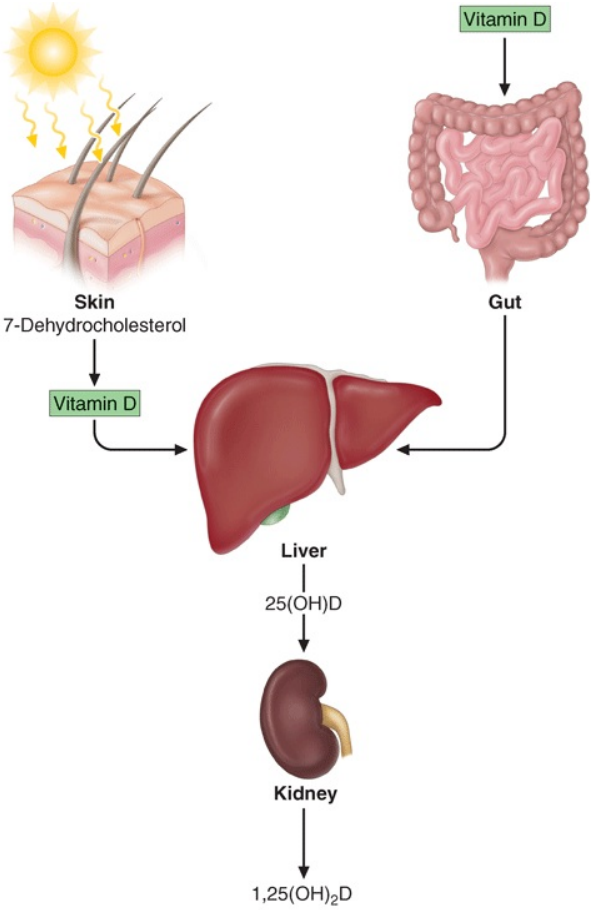

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Vitamin D synthesis and activation. Vitamin D is synthesized in the skin in response to ultraviolet radiation and also is absorbed from the diet. It is then transported to the liver, where it undergoes 25-hydroxylation. This metabolite is the major circulating form of vitamin D. The final step in hormone activation, 1α-hydroxylation, occurs in the kidney.

The second hydroxylation, required for the formation of the mature hormone, occurs in the kidney (Fig. 352-5). The 25-hydroxyvitamin D-1α-hydroxylase is a tightly regulated cytochrome P450–like mixed-function oxidase expressed in the proximal convoluted tubule cells of the kidney. PTH and hypophosphatemia are the major inducers of this microsomal enzyme, whereas calcium, FGF23, and the enzyme's product, $1,25(OH)_2D$, repress it. The 25-hydroxyvitamin D-1α-hydroxylase is also present in epidermal keratinocytes, but keratinocyte production of $1,25(OH)_2D$ is not thought to contribute to circulating levels of this hormone. In addition to being present in the trophoblastic layer of the placenta, the 1α-hydroxylase is produced by macrophages associated with granulomas and lymphomas. In these latter pathologic states, the activity of the enzyme is induced by interferon γ and TNF-α but is not regulated by calcium or $1,25(OH)_2D$; therefore, hypercalcemia, associated with elevated levels of $1,25(OH)_2D$, may be observed. Treatment of sarcoidosis-associated hypercalcemia with glucocorticoids, ketoconazole, or chloroquine reduces $1,25(OH)_2D$ production and effectively lowers serum calcium. In contrast, chloroquine has not been shown to lower the elevated serum $1,25(OH)_2D$ levels in patients with lymphoma.

Figure 352-5

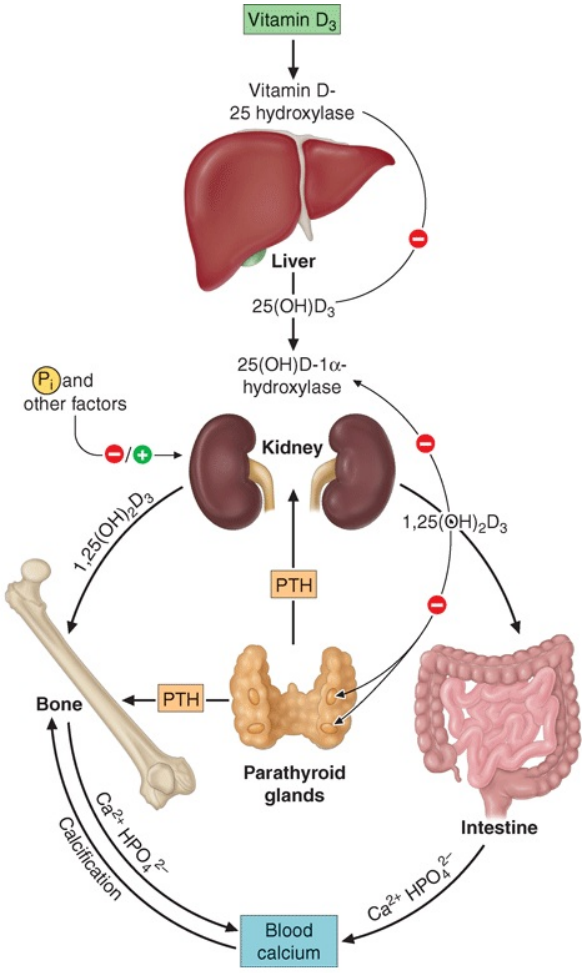

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of the hormonal control loop for vitamin D metabolism and function. A reduction in the serum calcium below ~2.2 mmol/L (8.8 mg/dL) prompts a proportional increase in the secretion of parathyroid hormone (PTH) and so mobilizes additional calcium from the bone. PTH promotes the synthesis of $1,25(OH)_2D$ in the kidney, which in turn stimulates the mobilization of calcium from bone and intestine and regulates the synthesis of PTH by negative feedback.

The major pathway for inactivation of vitamin D metabolites is an additional hydroxylation step by the vitamin D 24-hydroxylase, an enzyme that is expressed in most tissues. $1,25(OH)_2D$ is the major inducer of this enzyme; therefore, this hormone promotes its own inactivation, thereby limiting its biologic effects. Polar metabolites of $1,25(OH)_2D$ are secreted into the bile and reabsorbed via the enterohepatic circulation. Impairment of this recirculation, which is seen with diseases of the terminal ileum, leads to accelerated losses of vitamin D metabolites.

Actions of $1,25(Oh)_2D$

$1,25(OH)_2D$ mediates its biologic effects by binding to a member of the nuclear receptor superfamily, the vitamin D receptor (VDR). This receptor belongs to the subfamily that includes the thyroid hormone receptors, the retinoid receptors, and the peroxisome proliferator–activated receptors; however, in contrast to the other members of this subfamily, only one VDR isoform has been isolated. The VDR binds to target DNA sequences as a heterodimer with the retinoid X receptor, recruiting a series of coactivators that modify chromatin and approximate the VDR to the basal transcriptional apparatus, resulting in the induction of target gene expression. The mechanism of transcriptional repression by the VDR varies with different target genes but has been shown to involve either interference with the action of activating transcription factors or the recruitment of novel proteins to the VDR complex, resulting in transcriptional repression.

The affinity of the VDR for $1,25(OH)_2D$ is approximately three orders of magnitude higher than that for other vitamin D metabolites. In normal physiologic circumstances, these other metabolites are not thought to stimulate receptor-dependent actions. However, in states of vitamin D toxicity, the markedly elevated levels of 25(OH)D may lead to hypercalcemia by interacting directly with the VDR and by displacing $1,25(OH)_2D$ from vitamin D–binding protein, resulting in increased bioavailability of the active hormone.

The VDR is expressed in a wide range of cells and tissues. The molecular actions of $1,25(OH)_2D$ have been studied most extensively in tissues involved in the regulation of mineral ion homeostasis. This hormone is a major inducer of calbindin 9K, a calcium-binding protein expressed in the intestine, which is thought to play an important role in the active transport of calcium across the enterocyte. The two major calcium transporters expressed by intestinal epithelia, TRPV5 and TRPV6 (transient receptor potential vanilloid), are also vitamin D responsive. By inducing the expression of these and other genes in the small intestine, $1,25(OH)_2D$ increases the efficiency of intestinal calcium absorption, and it also has been shown to have several important actions in the skeleton. The VDR is expressed in osteoblasts and regulates the expression of several genes in this cell. These genes include the bone matrix proteins osteocalcin and osteopontin, which are upregulated by $1,25(OH)_2D$, in addition to type I collagen, which is transcriptionally repressed by $1,25(OH)_2D$. Both $1,25(OH)_2D$ and PTH induce the expression of RANK ligand, which promotes osteoclast differentiation and increases osteoclast activity, by binding to RANK on osteoclast progenitors and mature osteoclasts. This is the mechanism by which $1,25(OH)_2D$ induces bone resorption. However, the skeletal features associated with VDR-knockout mice (rickets, osteomalacia) are largely corrected by increasing calcium and

phosphorus intake, underscoring the importance of vitamin D action in the gut.

The VDR is expressed in the parathyroid gland, and $1,25(OH)_2D$ has been shown to have antiproliferative effects on parathyroid cells and to suppress the transcription of the parathyroid hormonegene. These effects of $1,25(OH)_2D$ on the parathyroid gland are an important part of the rationale for current therapies directed at preventing and treating hyperparathyroidism associated with renal insufficiency.

The VDR is also expressed in tissues and organs that do not play a role in mineral ion homeostasis. Notable in this respect is the observation that $1,25(OH)_2D$ has an antiproliferative effect on several cell types, including keratinocytes, breast cancer cells, and prostate cancer cells. The effects of $1,25(OH)_2D$ and the VDR on keratinocytes are particularly intriguing. Alopecia is seen in humans and mice with mutant VDRs but is not a feature of vitamin D deficiency; thus, the effects of the VDR on the hair follicle are ligand-independent.

Vitamin D Deficiency

The mounting concern about the relationship between solar exposure and the development of skin cancer has led to increased reliance on dietary sources of vitamin D. Although the prevalence of vitamin D deficiency varies, the third National Health and Nutrition Examination Survey (NHANES III) revealed that vitamin D deficiency is prevalent throughout the United States. The clinical syndrome of vitamin D deficiency can be a result of deficient production of vitamin D in the skin, lack of dietary intake, accelerated losses of vitamin D, impaired vitamin D activation, or resistance to the biologic effects of $1,25(OH)_2D$ (Table 352-6).

The elderly and nursing home residents are particularly at risk for vitamin D deficiency, since both the efficiency of vitamin D synthesis in the skin and the absorption of vitamin D from the intestine decline with age. Similarly, intestinal malabsorption of dietary fats leads to vitamin D deficiency. This is further exacerbated in the presence of terminal ileal disease, which results in impaired enterohepatic circulation of vitamin D metabolites. In addition to intestinal diseases, accelerated inactivation of vitamin D metabolites can be seen with drugs that induce hepatic cytochrome P450 mixed-function oxidases such as barbiturates, phenytoin, and rifampin. Impaired 25-hydroxylation, associated with severe liver disease or isoniazid, is an uncommon cause of vitamin D deficiency. A mutation in the gene responsible for 25- hydroxylation has been identified in one kindred. Impaired 1α-hydroxylation is prevalent in the population with profound renal dysfunction due to an increase in circulating FGF23 levels and a decrease in functional renal mass. Thus, therapeutic interventions should be considered in patients whose creatinine clearance is <0.5 mL/s (30 mL/min). Mutations in the renal 1α-hydroxylase are the basis for the genetic disorder, pseudovitamin D –deficiency rickets. This autosomal recessive disorder presents with the syndrome of vitamin D deficiency in the first year of life. Patients present with growth retardation, rickets, and hypocalcemic seizures. Serum $1,25(OH)_2D$ levels are low despite normal 25(OH)D levels and elevated PTH levels. Treatment with vitamin D metabolites that do not require 1α-hydroxylation results in disease remission, although lifelong therapy is required. A second autosomal recessive disorder, hereditary vitamin D –resistant rickets, a consequence of vitamin D receptor mutations, is a greater therapeutic challenge. These patients present in a similar fashion during the first year of life, but alopecia often accompanies the disorder, demonstrating a functional role of the VDR in postnatal hair regeneration. Serum levels of $1,25(OH)_2D$ are dramatically elevated in these individuals both because of increased production due to stimulation of 1α-hydroxylase activity as a consequence of secondary hyperparathyroidism and because of impaired inactivation, since induction of the 24-hydroxylase by $1,25(OH)_2D$ requires an intact VDR. Since the receptor mutation results in hormone resistance, daily calcium and phosphorus infusions may be required to bypass the defect in intestinal mineral ion absorption.

Table 352-6 Causes of Impaired Vitamin D Action

Vitamin D deficiency
Impaired cutaneous production
Dietary absence
Malabsorption
Accelerated loss of vitamin D
Increased metabolism (barbiturates, phenytoin, rifampin)
Impaired enterohepatic circulation
Nephrotic syndrome
Impaired 25-hydroxylation
Liver disease, isoniazid
Impaired 1α-hydroxylation
Hypoparathyroidism
Renal failure
Ketoconazole
1α-hydroxylase mutation
Oncogenic osteomalacia
X-linked hypophosphatemic rickets
Target organ resistance
Vitamin D receptor mutation
Phenytoin

Regardless of the cause, the clinical manifestations of vitamin D deficiency are largely a consequence of impaired intestinal calcium absorption. Mild to moderate vitamin D deficiency is asymptomatic, whereas long-standing vitamin D deficiency results in hypocalcemia accompanied by secondary hyperparathyroidism, impaired mineralization of the skeleton (osteopenia on x-ray or decreased bone mineral density), and proximal myopathy. Vitamin D deficiency also has been shown to be associated with an increase in overall mortality rates, including cardiovascular causes. In the absence of an intercurrent illness, the hypocalcemia associated with long-standing vitamin D deficiency rarely presents with acute symptoms of hypocalcemia such as numbness, tingling, and seizures. However, the concurrent development of hypomagnesemia, which impairs parathyroid function, or the administration of potent bisphosphonates, which impair bone resorption, can lead to acute symptomatic hypocalcemia in vitamin D–deficient individuals.

RICKETS AND OSTEOMALACIA

In children, before epiphyseal fusion, vitamin D deficiency results in growth retardation associated with an expansion of the growth plate known as *rickets*. Three layers of chondrocytes are present in the normal growth plate: the reserve zone, the

proliferating zone, and the hypertrophic zone. Rickets associated with impaired vitamin D action is characterized by expansion of the hypertrophic chondrocyte layer. The proliferation and differentiation of the chondrocytes in the rachitic growth plate are normal, and the expansion of the growth plate is a consequence of impaired apoptosis of the late hypertrophic chondrocytes, an event that precedes replacement of these cells by osteoblasts during endochondral bone formation. Investigations in murine models demonstrate that hypophosphatemia, which in vitamin D deficiency is a consequence of secondary hyperparathyroidism, is a key etiologic factor in the development of the rachitic growth plate.

The hypocalcemia and hypophosphatemia that accompany vitamin D deficiency result in impaired mineralization of bone matrix proteins, a condition known as *osteomalacia*. Osteomalacia is also a feature of long-standing hypophosphatemia, which may be a consequence of renal phosphate wasting or chronic use of etidronate or phosphate-binding antacids. This hypomineralized matrix is biomechanically inferior to normal bone; as a result, patients with vitamin D deficiency are prone to bowing of weight-bearing extremities and skeletal fractures. Vitamin D and calcium supplementation have been shown to decrease the incidence of hip fracture among ambulatory nursing home residents in France, suggesting that undermineralization of bone contributes significantly to morbidity in the elderly. Proximal myopathy is a striking feature of severe vitamin D deficiency both in children and in adults. Rapid resolution of the myopathy is observed upon vitamin D treatment.

Though vitamin D deficiency is the most common cause of rickets and osteomalacia, many disorders lead to inadequate mineralization of the growth plate and bone. Calcium deficiency without vitamin D deficiency, the disorders of vitamin D metabolism previously discussed, and hypophosphatemia can all lead to inefficient mineralization. Even in the presence of normal calcium and phosphate levels, chronic acidosis and drugs such as bisphosphonates can lead to osteomalacia. The inorganic calcium/phosphate mineral phase of bone cannot form at low pH, and bisphosphonates bind to and prevent mineral crystal growth. Since alkaline phosphatase is necessary for normal mineral deposition, probably because the enzyme can hydrolyze inhibitors of mineralization such as inorganic pyrophosphate, genetic inactivation of the alkaline phosphatase gene (hereditary hypophosphatasia) also can lead to osteomalacia in the setting of normal calcium and phosphate levels.

DIAGNOSIS OF VITAMIN D DEFICIENCY, RICKETS, AND OSTEOMALACIA

The most specific screening test for vitamin D deficiency in otherwise healthy individuals is a serum 25(OH)D level. Although the normal ranges vary, levels of 25(OH)D <37 nmol/L (<15 ng/mL) are associated with increasing PTH levels and lower bone density; optimal vitamin D levels are >80 nmol/L (>32 ng/mL). Vitamin D deficiency leads to impaired intestinal absorption of calcium, resulting in decreased serum total and ionized calcium values. This hypocalcemia results in secondary hyperparathyroidism, a homeostatic response that initially maintains serum calcium levels at the expense of the skeleton. Due to the PTH-induced increase in bone turnover, alkaline phosphatase levels are often increased. In addition to increasing bone resorption, PTH decreases urinary calcium excretion while promoting phosphaturia. This results in hypophosphatemia, which exacerbates the mineralization defect in the skeleton. With prolonged vitamin D deficiency resulting in osteomalacia, calcium stores in the skeleton become relatively inaccessible, since osteoclasts cannot resorb unmineralized osteoid, and frank hypocalcemia ensues. Since PTH is a major stimulus for the renal 25(OH)D 1α-hydroxylase, there is increased synthesis of the active hormone, $1,25(OH)_2D$. Paradoxically, levels of this hormone are often normal in severe vitamin D deficiency. Therefore, measurements of $1,25(OH)_2D$ are not accurate reflections of vitamin D stores and should not be used to diagnose vitamin D deficiency in patients with normal renal function.

Radiologic features of vitamin D deficiency in children include a widened, expanded growth plate that is characteristic of rickets. These findings not only are apparent in the long bones but also are present at the costochondral junction, where the expansion of the growth plate leads to swellings known as the "rachitic rosary." Impairment of intramembranous bone mineralization leads to delayed fusion of the calvarial sutures and a decrease in the radi-opacity of cortical bone in the long bones. If vitamin D deficiency occurs after epiphyseal fusion, the main radiologic finding is a decrease in cortical thickness and relative radiolucency of the skeleton. A specific radiologic feature of osteomalacia, whether associated with phosphate wasting or vitamin D deficiency, is pseudofractures, or Looser's zones. These are radiolucent lines that occur where large arteries are in contact with the underlying skeletal elements; it is thought that the arterial pulsations lead to the radiolucencies. As a result, these pseudofractures are usually a few millimeters wide, are several centimeters long, and are seen particularly in the scapula, the pelvis, and the femoral neck.

TREATMENT: VITAMIN D DEFICIENCY

Daily intake of a multivitamin (400 IU) is often insufficient to prevent vitamin D deficiency. Based on the observation that 800 IU of vitamin D, with calcium supplementation, decreases the risk of hip fractures in elderly women, this higher dose is thought to be an appropriate daily intake for prevention of vitamin D deficiency in adults. The safety margin for vitamin D is large, and vitamin D toxicity usually is observed only in patients taking doses in the range of 40,000 IU daily. Treatment of vitamin D deficiency should be directed at the underlying disorder, if possible, and also should be tailored to the severity of the condition. Vitamin D should always be repleted in conjunction with calcium supplementation since most of the consequences of vitamin D deficiency are a result of impaired mineral ion homeostasis. In patients in whom 1α-hydroxylation is impaired, metabolites that do not require this activation step are the treatment of choice. They include $1,25(OH)_2D3$ [calcitriol (Rocaltrol), 0.25–0.5 μg/d] and 1α-hydroxyvitamin D_2 (Hectorol, 2.5–5 μg/d). If the pathway required for activation of vitamin D is intact, severe vitamin D deficiency can be treated with pharmacologic repletion initially (50,000 IU weekly for 3–12 weeks), followed by maintenance therapy (800 IU daily). Pharmacologic doses may be required for maintenance therapy in patients who are taking medications such as barbiturates or phenytoin, that accelerate metabolism of, or cause resistance to $1,25(OH)_2D$. Calcium supplementation should include 1.5–2 g/d of elemental calcium. Normocalcemia is usually observed within one week of the institution of therapy, although increases in PTH and alkaline phosphatase levels may persist for three to six months. The most efficacious methods to monitor treatment and resolution of vitamin D deficiency are serum and urinary calcium measurements. In patients who are vitamin D replete and are taking adequate calcium supplementation, the 24-hour urinary calcium excretion should be in the range of 100–250 mg/ 24 hours. Lower levels suggest problems with adherence to the treatment regimen or with absorption of calcium or vitamin D supplements. Levels >250 mg/24 hours predispose to nephrolithiasis and should lead to a reduction in vitamin D dosage and/or calcium supplementation.

FURTHER READINGS

Bouillon R et al: Vitamin D and human health: Lessons from the vitamin D receptor null mice. Endocr Rev 29:726, 2008[PMID: 18694980] [Full Text]

DeLuca HF: Overview of general physiologic features and functions of vitamin D. Am J Clin Nutr 80:1689S, 2004

Heaney RP: Bone health. Am J Clin Nutr 85:300S, 2007

Holick MF: Vitamin D deficiency. N Engl J Med 357:266, 2007[PMID: 17634462] [Full Text]

Karsenty G: The complexities of skeletal biology. Nature 423:316, 2003[PMID: 12748648] [Full Text]

Kobayashi T et al: Minireview: Transcriptional regulation in development of bone. Endocrinology 146:1012, 2005[PMID: 15604211] [Full Text]

Razzaque MS: The FGF23-Klotho axis: Endocrine regulation of phosphate homeostasis. Nat Rev Endocrinol 5:611, 2009.[PMID: 19844248] [Full Text]

Teitelbaum SL, Ross SP: Genetic regulation of osteoclast development and function. Nat Rev Genet 4:638, 2003[PMID: 12897775] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE PARATHYROID GLAND AND CALCIUM HOMEOSTASIS: INTRODUCTION

The four parathyroid glands are located posterior to the thyroid gland. They produce parathyroid hormone (PTH), which is the primary regulator of calcium physiology. PTH acts directly on bone, where it induces calcium resorption, and on the kidney, where it enhances calcium reabsorption and synthesis of 1,25-dihydroxyvitamin D [1,25$(OH)_2$D], a hormone that increases gastrointestinal calcium absorption. Serum PTH levels are tightly regulated by a negative feedback loop. Calcium, acting through the calcium-sensing receptor, and vitamin D, acting through its nuclear receptor, reduce PTH release and synthesis. Additional evidence indicates that fibroblast growth factor 23 (FGF23), a phosphaturic hormone, can suppress PTH secretion. Understanding the hormonal pathways that regulate calcium levels and bone metabolism is essential for effective diagnosis and management of a wide array of hyper- and hypocalcemic disorders.

Hyperparathyroidism (HPT), characterized by excess production of PTH, is a common cause of hypercalcemia and is usually the result of autonomously functioning adenomas or hyperplasia. Surgery for this disorder is highly effective and has been shown to reverse some of the deleterious effects of long-standing PTH excess on bone density. Humoral hypercalcemia of malignancy is also common and is usually due to the overproduction of parathyroid hormone–related peptide (PTHrP) by cancer cells. The similarities in the biochemical characteristics of hyperparathyroidism and humoral hypercalcemia of malignancy, first noted by Albright in 1941, are now known to reflect the actions of PTH and PTHrP through the same G protein–coupled PTH/PTHrP receptor.

The genetic basis of multiple endocrine neoplasia (MEN) types 1 and 2, familial hypocalciuric hypercalcemia (FHH), different forms of pseudohypoparathyroidism (PHP), Jansen's syndrome, disorders of vitamin D synthesis and action, and the molecular events associated with parathyroid gland neoplasia have provided new insights into the regulation of calcium homeostasis. PTH and *possibly some* of its analogues are promising therapeutic agents for the treatment of postmenopausal or senile osteoporosis, and calcimimetic agents, which activate the calcium-sensing receptor, have provided new approaches for PTH suppression.

PARATHYROID HORMONE

Physiology

The primary function of PTH is to maintain the extracellular fluid (ECF) calcium concentration within a narrow normal range. The hormone acts directly on bone and kidney and indirectly on the intestine through its effects on synthesis of 1,25$(OH)_2$D to increase serum calcium concentrations; in turn, PTH production is closely regulated by the concentration of serum ionized calcium. This feedback system is the critical homeostatic mechanism for maintenance of ECF calcium. Any tendency toward hypocalcemia, as might be induced by calcium-deficient diets, is counteracted by an increased secretion of PTH. This in turn (1) increases the rate of dissolution of bone mineral, thereby increasing the flow of calcium from bone into blood; (2) reduces the renal clearance of calcium, returning more of the calcium filtered at the glomerulus into ECF; and (3) increases the efficiency of calcium absorption in the intestine by stimulating the production of 1,25$(OH)_2$D. Immediate control of blood calcium is due to PTH effects on bone and, to a lesser extent, on renal calcium clearance. Maintenance of steady-state calcium balance, on the other hand, probably results from the effects of 1,25$(OH)_2$D on calcium absorption (Chap. 352). The renal actions of the hormone are exerted at multiple sites and include inhibition of phosphate transport (proximal tubule), augmentation of calcium reabsorption (distal tubule), and stimulation of the renal 25(OH)D-1α-hydroxylase. As much as 12 mmol (500 mg) calcium is transferred between the ECF and bone each day (a large amount in relation to the total ECF calcium pool), and PTH has a major effect on this transfer. The homeostatic role of the hormone can preserve calcium concentration in blood at the cost of bone demineralization.

PTH has multiple actions on bone, some direct and some indirect. PTH-mediated changes in bone calcium release can be seen within minutes. The chronic effects of PTH are to increase the number of bone cells, both osteoblasts and osteoclasts, and to increase the remodeling of bone; these effects are apparent within hours after the hormone is given and persist for hours after PTH is withdrawn. Continuous exposure to elevated PTH (as in hyperparathyroidism or long-term infusions in animals) leads to increased osteoclast- mediated bone resorption. However, the intermittent administration of PTH, elevating hormone levels for 1–2 hours each day, leads to a net stimulation of bone formation rather than bone breakdown. Striking increases, especially in trabecular bone in the spine and hip, have been reported with the use of PTH in combination with estrogen. PTH(1-34) as monotherapy caused a highly significant reduction in fracture incidence in a worldwide placebo-controlled trial.

Osteoblasts (or stromal cell precursors), which have PTH/PTHrP receptors, are crucial to this bone-forming effect of PTH; osteoclasts, which mediate bone breakdown, lack such receptors. PTH-mediated stimulation of osteoclasts is indirect, acting in part, through cytokines released from osteoblasts to activate osteoclasts; in experimental studies of bone resorption in vitro, osteoblasts must be present for PTH to activate osteoclasts to resorb bone (Chap. 352).

Structure

PTH is an 84-amino-acid single-chain peptide. The amino-terminal portion, PTH(1–34), is highly conserved and is critical for the biologic actions of the molecule. Modified synthetic fragments of the amino-terminal sequence as small as PTH(1–11) are sufficient to activate the PTH/PTHrP receptor (see below). The carboxyl-terminal regions of the full-length PTH(1–84) molecule also can bind to a separate binding protein/receptor (cPTH-R), but this receptor has been incompletely characterized. Fragments shortened at the amino terminus possibly by binding to cPTH can inhibit some of the biologic actions of full-length PTH(1–84) and of PTH(1–34).

Biosynthesis, Secretion, and Metabolism

SYNTHESIS

Parathyroid cells have multiple methods of adapting to increased needs for PTH production. Most rapid (within minutes) is secretion of preformed hormone in response to hypocalcemia. Second, within hours, PTH mRNA expression is induced by sustained hypocalcemia. Finally, protracted challenge leads within days to cellular replication to increase gland mass.

PTH is initially synthesized as a larger molecule (preproparathyroid hormone, consisting of 115 amino acids). After a first cleavage step to remove the "pre" sequence of 25 amino acid residues, a second cleavage step removes the "pro" sequence of 6 amino acid residues before secretion of the mature peptide comprising 84 residues. In one kindred with hypoparathyroidism, a mutation in the preprotein region of the gene interferes with hormone transport and secretion.

Transcriptional suppression of the PTH gene by calcium is nearly maximal at physiologic calcium concentrations. Hypocalcemia increases transcriptional activity within hours. $1,25(OH)_2D3$ strongly suppresses PTH gene transcription. In patients with renal failure, IV administration of supraphysiologic levels of $1,25(OH)_2D_3$ or analogues of this active metabolite can dramatically suppress PTH overproduction, which is sometimes difficult to control due to severe secondary HPT. Regulation of proteolytic destruction of preformed hormone (posttranslational regulation of hormone production) is an important mechanism for mediating rapid (minutes) changes in hormone availability. High calcium increases and low calcium inhibit the proteolytic destruction of hormone stores.

REGULATION OF PTH SECRETION

PTH secretion increases steeply to a maximum value of about five times the basal rate of secretion as calcium concentration falls from normal to the range of 1.9–2 mmol/L (7.5–8 mg/dL) (measured as total calcium). The ionized fraction of blood calcium is the important determinant of hormone secretion. Severe intracellular magnesium deficiency impairs PTH secretion (see below).

ECF calcium controls PTH secretion by interaction with a calcium sensor, a G protein–coupled receptor (GPCR) for which Ca^{2+} ions act as the primary ligand (see below). This receptor is a member of a distinctive subgroup of the GPCR superfamily that is characterized by a large extracellular domain suitable for "clamping" the small-molecule ligand. Stimulation of the receptor by high calcium levels suppresses PTH secretion. The receptor is present in parathyroid glands and the calcitonin-secreting cells of the thyroid (C cells), as well as in other sites such as brain and kidney. Genetic evidence has revealed a key biologic role for the calcium-sensing receptor in parathyroid gland responsiveness to calcium and in renal calcium clearance. Heterozygous point mutations associated with loss-of-function cause the syndrome of FHH, in which the blood calcium abnormality resembles that observed in hyperparathyroidism but with hypocalciuria. On the other hand, heterozygous gain-of-function mutations cause a form of hypocalcemia resembling hypoparathyroidism (see below).

METABOLISM

The secreted form of PTH is indistinguishable by immunologic criteria and by molecular size from the 84-amino-acid peptide [PTH(1–84)] extracted from glands. However, much of the immunoreactive material found in the circulation is smaller than the extracted or secreted hormone. The principal circulating fragments of immunoreactive hormone lack a portion of the critical amino-terminal sequence required for biologic activity and, hence, are biologically inactive fragments (so-called middle and carboxyl-terminal fragments). Much of the proteolysis of hormone occurs in the liver and kidney. Peripheral metabolism of PTH does not appear to be regulated by physiologic states (high versus low calcium, etc.); hence, peripheral metabolism of hormone, although responsible for rapid clearance of secreted hormone, appears to be a high-capacity, metabolically invariant catabolic process.

The rate of clearance of the secreted 84-amino-acid peptide from blood is more rapid than the rate of clearance of the biologically inactive fragment(s) corresponding to the middle and carboxyl-terminal regions of PTH. Consequently, the interpretation of results obtained with earlier PTH radioimmunoassays is influenced by the nature of the peptide fragments detected by the antibodies.

Although the problems inherent in PTH measurements have been largely circumvented by use of double-antibody immunometric assays, it is now known that some of these assays detect, besides the intact molecule, large amino-terminally truncated forms of PTH, which are present in normal and uremic individuals in addition to PTH(1–84). The concentration of these fragments relative to that of intact PTH(1–84) is higher with induced hypercalcemia than in eucalcemic or hypocalcemic conditions and is higher in patients with renal failure. These fragments have limited portions of the amino-terminal portion of the hormone removed; PTH(7–84) has been identified as a major component of these amino-terminally truncated fragments. Growing evidence suggests that the PTH(7–84) (and probably related amino-terminally truncated fragments) can act, through yet undefined mechanisms, as an inhibitor of PTH action and may be of clinical significance, particularly in renal failure. In this group of patients, efforts to prevent secondary HPT by a variety of measures (vitamin D analogues, higher calcium intake, higher dialysate calcium, and phosphate-lowering strategies) may have led to oversuppression of biologically active, intact PTH since some amino-terminally truncated PTH fragments such as PTH(7–84), react in many immunometric PTH assays (now termed second-generation assays; see below under "Diagnosis"). Excessive parathyroid gland suppression due to overly aggressive treatment with vitamin D analogues and calcium-containing phosphate binders or inaccurate PTH measurements can lead to adynamic bone disease in renal failure (see below). Adynamic bone disease has been associated in children with further impaired growth and increased bone fracture rates in adults, and can furthermore lead to significant hypercalcemia. The measurement of PTH with newer third-generation immunometric assays, which use detection antibodies directed against extreme amino-terminal PTH epitopes and thus detect only full-length PTH(1–84), may provide some advantage to prevent bone disease in chronic kidney disease.

PARATHYROID HORMONE–RELATED PROTEIN (PTHRP)

PTHrP is responsible for most instances of hypercalcemia of malignancy (Chap. 100), a syndrome that resembles HPT but without elevated PTH levels. Most cell types produce PTHrP, including brain, pancreas, heart, lung, mammary tissue, placenta, endothelial cells, and smooth muscle. In fetal animals, PTHrP directs transplacental calcium transfer, and high concentrations of PTHrP are produced in mammary tissue and secreted into milk, but the biologic significance of the very high concentrations of this hormone in breast milk is unknown. PTHrP also plays an essential role in endochondral bone formation and in branching morphogenesis of the breast, and possibly in uterine contraction and other biologic functions.

PTH and PTHrP, although distinctive products of different genes, exhibit considerable functional and structural homology **(Fig. 353-1)** and have evolved from a shared ancestral gene. The structure of the gene for human PTHrP, however, is more complex than that of PTH, containing multiple additional exons, which can undergo alternate splicing patterns during formation of the mature mRNA. Protein products of 141, 139, and 173 amino acids are produced, and other molecular forms may result from tissue-specific degradation at accessible internal cleavage sites. The biologic roles of these various molecular species and the nature of the circulating forms of PTHrP are unclear. It is uncertain whether PTHrP circulates at any significant level in adults. As a paracrine factor, PTHrP may be produced, act, and be destroyed locally within tissues. In adults, PTHrP appears to have little influence on calcium homeostasis, except in disease states, when large tumors, especially of the squamous cell type as well as renal cell carcinomas, lead to massive overproduction of the hormone and hypercalcemia.

Figure 353-1

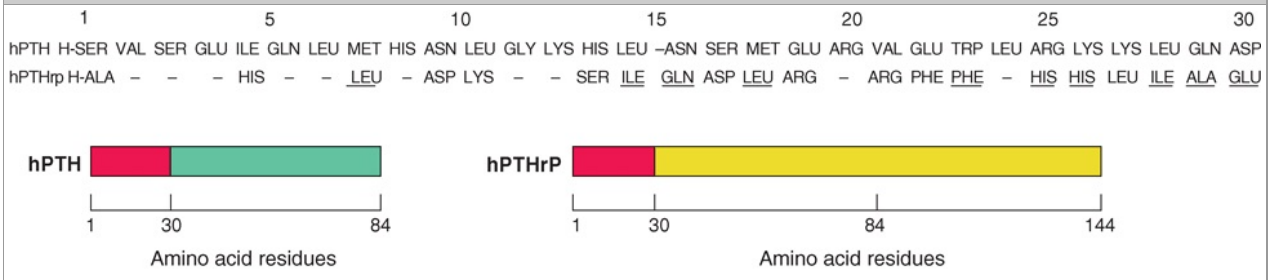

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram to illustrate similarities and differences in structure of human parathyroid hormone (PTH) and human PTH-related peptide (PTHrP). Close structural (and functional) homology exists between the first 30 amino acids of hPTH and hPTHrP. The PTHrP sequence may be ≥144 amino acid residues in length. PTH is only 84 residues long; after residue 30, there is little structural homology between the two. Dashed lines in the PTHrP sequence indicate identity; underlined residues, although different from those of PTH, still represent conservative changes (charge or polarity preserved). Ten amino acids are identical, and a total of 20 of 30 are homologues.

PTH and Pthrp Hormone Action

Both PTH and PTHrP bind to and activate the PTH/PTHrP receptor. The PTH/PTHrP receptor (also known as the PTH-1 receptor, PTH1R) belongs to a subfamily of GPCRs that includes the receptors for calcitonin, glucagon, secretin, vasoactive intestinal peptide, and other peptides. Although both ligands activate the PTH1R, the two peptides induce distinct responses in the receptor, which explains how a single receptor without isoforms can serve two biologic roles. The extracellular regions of the receptor are involved in hormone binding, and the intracellular domains, after hormone activation, bind G protein subunits to transduce hormone signaling into cellular responses through the stimulation of second messenger formation. A second receptor that binds PTH, originally termed the *PTH-2 receptor* (PTH2R), is primarily expressed in brain, pancreas, and testis. Different mammalian PTH1Rs respond equivalently to PTH and PTHrP, whereas only the human PTH2R responds efficiently to PTH (but not to PTHrP), PTH2Rs from other species show little or no stimulation of second-messenger formation in response to PTH or PTHrP. The endogenous ligand of the PTH2R was shown to be a hypothalamic peptide referred to as *tubular infundibular peptide of 39 residues* , TIP39, that is distantly related to PTH. The PTH1R and the PTH2R can be traced backward in evolutionary time to fish; in fact, the zebrafish genome contains, in addition to the PTH1R and the PTH2R orthologs, a third receptor, the PTH3R, that is more closely related to the fish PTH1R than the fish PTH2R. The evolutionary conservation of structure and function suggests important biologic roles for these receptors, even in fish, which lack discrete parathyroid glands but produce two molecules that are closely related to mammalian PTH.

Studies using the cloned PTH1R confirm that it can be coupled to more than one G protein and second-messenger pathway, apparently explaining the multiplicity of pathways stimulated by PTH. Stimulation of protein kinases (A and C) and calcium transport channels is associated with a variety of hormone-specific tissue responses. These responses include inhibition of phosphate and bicarbonate transport, stimulation of calcium transport, and activation of renal 1α-hydroxylase in the kidney. The responses in bone include effects on collagen synthesis, alkaline phosphatase, ornithine decarboxylase, citrate decarboxylase, and glucose-6-phosphate dehydrogenase activities, phospholipid synthesis, as well as calcium and phosphate transport. Ultimately, these biochemical events lead to an integrated hormonal response in bone turnover and calcium homeostasis. PTH also activates Na^+/Ca^{2+} exchangers at renal distal tubular sites and stimulates translocation of preformed calcium transport channels, moving them from the interior to the apical surface to increase tubular uptake of calcium. PTH-dependent stimulation of phosphate excretion (reducing reabsorption—the opposite effect from actions on calcium in the kidney) involves the down regulation of two sodium-dependent phosphate co-transporters, NPT2a and NPT2c, and expression at the apical membrane, thereby reducing phosphate reabsorption. Similar mechanisms may be involved in other renal tubular transporters that are influenced by PTH.

PTHrP exerts important developmental influences on fetal bone development and in adult physiology. A homozygous ablation of the gene encoding PTHrP (or disruption of the gene encoding the PTH/PTHrP receptor) in mice causes a lethal phenotype in which animals are born with pronounced acceleration of chondrocyte maturation that resembles a lethal form of chondrodysplasia in humans **(Fig. 353-2)**. Mice that are heterozygous for ablation of the PTHrP gene display reduced mineral density consistent with osteoporosis. Experiments with these mouse models point to a hitherto unappreciated role of PTHrP as a paracrine/autocrine factor that modulates bone metabolism in adults as well as during bone development.

Figure 353-2

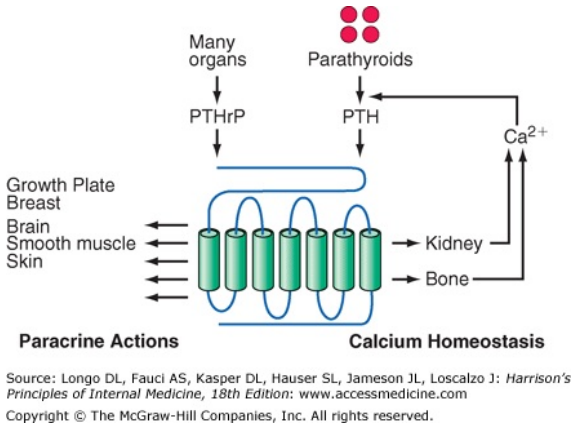

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Dual role for the actions of the PTH/PTHrP receptor (PTH1R). Parathyroid hormone (PTH; endocrine-calcium homeostasis) and PTH-related peptide (PTHrP; paracrine–multiple tissue actions including growth plate cartilage in developing bone) use the single receptor for their disparate functions mediated by the amino-terminal 30 residues of either peptide. Other regions of both ligands interact with other receptors (not shown).

CALCITONIN

(See also Chap. 351) Calcitonin is a hypocalcemic peptide hormone that in several mammalian species acts as an antagonist to PTH. Calcitonin seems to be of limited physiologic significance in humans, at least with regard to calcium homeostasis. It is of medical significance because of its role as a tumor marker in sporadic and hereditary cases of medullary carcinoma and its medical use as an adjunctive treatment in severe hypercalcemia and in Paget's disease of bone.

The hypocalcemic activity of calcitonin is accounted for primarily by inhibition of osteoclast-mediated bone resorption and secondarily by stimulation of renal calcium clearance. These effects are mediated by receptors on osteoclasts and renal tubular cells. Calcitonin exerts additional effects through receptors present in the brain, the gastrointestinal tract, and the immune system. The hormone, for example, exerts analgesic effects directly on cells in the hypothalamus and related structures, possibly by interacting with receptors for related peptide hormones such as calcitonin gene–related peptide (CGRP) or amylin. The latter ligands have specific high-affinity receptors and can also bind to and trigger calcitonin receptors. The calcitonin receptor shares considerable structural similarity with the PTH1R.

The thyroid is the major source of the hormone, and the cells involved in calcitonin synthesis arise from neural crest tissue. During embryogenesis, these cells migrate into the ultimobranchial body, derived from the last branchial pouch. In submammalian vertebrates, the ultimobranchial body constitutes a discrete organ, anatomically separate from the thyroid gland; in mammals, the ultimobranchial gland fuses with and is incorporated into the thyroid gland.

The naturally occurring calcitonins consist of a peptide chain of 32 amino acids. There is considerable sequence variability among species. Calcitonin from salmon, which is used therapeutically, is 10–100 times more potent than mammalian forms in lowering serum calcium.

There are two calcitoningenes, α and β; the transcriptional control of these genes is complex. Two different mRNA molecules are transcribed from the α gene; one is translated into the precursor for calcitonin, and the other message is translated into an alternative product, CGRP. CGRP is synthesized wherever the calcitonin mRNA is expressed (e.g., in medullary carcinoma of the thyroid). The β, or CGRP-2, gene is transcribed into the mRNA for CGRP in the central nervous system (CNS); this gene does not produce calcitonin, however. CGRP has cardiovascular actions and may serve as a neurotransmitter or play a developmental role in the CNS.

The circulating level of calcitonin in humans is lower than that in many other species. In humans, even extreme variations in calcitonin production do not change calcium and phosphate metabolism; no definite effects are attributable to calcitonin deficiency (totally thyroidectomized patients receiving only replacement thyroxine) or excess (patients with medullary carcinoma of the thyroid, a calcitonin-secreting tumor) (Chap. 351). Calcitonin has been a useful pharmacologic agent to suppress bone resorption in Paget's disease (Chap. 355) and osteoporosis (Chap. 354) and in the treatment of hypercalcemia of malignancy (see below). However, the physiologic role, if any, of calcitonin in humans is uncertain. On the other hand, a knockout of the calcitonin gene in mice leads to reduced bone mineral density, suggesting that its biologic role in mammals is still not fully understood.

HYPERCALCEMIA

(See also Chap. 46) Hypercalcemia can be a manifestation of a serious illness such as malignancy or can be detected coincidentally by laboratory testing in a patient with no obvious illness. The number of patients recognized with asymptomatic hypercalcemia, usually hyperparathyroidism, increased in the late twentieth century.

Whenever hypercalcemia is confirmed, a definitive diagnosis must be established. Although hyperparathyroidism, a frequent cause of asymptomatic hypercalcemia, is a chronic disorder in which manifestations, if any, may be expressed only after months or years, hypercalcemia can also be the earliest manifestation of malignancy, the second most common cause of hypercalcemia in the adult. The causes of hypercalcemia are numerous **(Table 353-1)**, but hyperparathyroidism and cancer account for 90% of all cases.

Table 353-1 Classification of Causes of Hypercalcemia

I. Parathyroid-related

A. Primary hyperparathyroidism Adenoma(s) Multiple endocrine neoplasia Carcinoma
B. Lithium therapy
C. Familial hypocalciuric hypercalcemia

II. Malignancy-related

A. Solid tumor with metastases (breast)
B. Solid tumor with humoral mediation of hypercalcemia (lung, kidney)
C. Hematologicmalignancies (multiple myeloma, lymphoma, leukemia)

III. Vitamin D—related

A. Vitamin D intoxication
B. ↑ 1,25(OH)2D; sarcoidosis and other granulomatous diseases
C. Idiopathic hypercalcemia of infancy

IV. Associated with high bone turnover

A. Hyperthyroidism
B. Immobilization
C. Thiazides
D. Vitamin Aintoxication

V. Associated with renal failure

A. Severe secondary hyperparathyroidism
B. Aluminum intoxication
C. Milk-alkali syndrome

Before undertaking a diagnostic workup, it is essential to be sure that true hypercalcemia, not a false-positive laboratory test, is present. A false-positive diagnosis of hypercalcemia is usually the result of inadvertent hemoconcentration during blood collection or elevation in serum proteins such as albumin. Hypercalcemia is a chronic problem, and it is cost-effective to obtain several serum calcium measurements; these tests need not be in the fasting state.

Clinical features are helpful in differential diagnosis. Hypercalcemia in an adult who is asymptomatic is usually due to primary hyperparathyroidism. In malignancy-associated hypercalcemia, the disease is usually not occult; rather, symptoms of malignancy bring the patient to the physician, and hypercalcemia is discovered during the evaluation. In such patients, the interval between detection of hypercalcemia and death, especially without vigorous treatment, is often <6 months. Accordingly, if an asymptomatic individual has had hypercalcemia or some manifestation of hypercalcemia such as kidney stones for >1 or 2 years, it is unlikely that malignancy is the cause. Nevertheless, differentiating primary hyperparathyroidism from *occult* malignancy can occasionally be difficult, and careful evaluation is required, particularly when the duration of the hypercalcemia is unknown. Hypercalcemia not due to hyperparathyroidism or malignancy can result from excessive vitamin D action, high bone turnover from any of several causes, or from renal failure (Table 353-1). Dietary history and a history of ingestion of vitamins or drugs are often helpful in diagnosing some of the less frequent causes. Immunometric PTH assays serve as the principal laboratory test in establishing the diagnosis.

Hypercalcemia from any cause can result in fatigue, depression, mental confusion, anorexia, nausea, vomiting, constipation, reversible renal tubular defects, increased urination, a short QT interval in the electrocardiogram, and, in some patients, cardiac arrhythmias. There is a variable relation from one patient to the next between the severity of hypercalcemia and the symptoms. Generally, symptoms are more common at calcium levels >2.9–3 mmol/L (11.5–12 mg/dL), but some patients, even at this level, are asymptomatic. When the calcium level is >3.2 mmol/L (13 mg/dL), calcification in kidneys, skin, vessels, lungs, heart, and stomach occurs and renal insufficiency may develop, particularly if blood phosphate levels are normal or elevated due to impaired renal function. Severe hypercalcemia, usually defined as ≥3.7–4.5 mmol/L (15–18 mg/dL), can be a medical emergency; coma and cardiac arrest can occur.

Acute management of the hypercalcemia is usually successful. The type of treatment is based on the severity of the hypercalcemia and the nature of associated symptoms, as outlined below.

Primary Hyperparathyroidism

NATURAL HISTORY AND INCIDENCE

Primary hyperparathyroidism is a generalized disorder of calcium, phosphate, and bone metabolism due to an increased secretion of PTH. The elevation of circulating hormone usually leads to hypercalcemia and hypophosphatemia. There is great variation in the manifestations. Patients may present with multiple signs and symptoms, including recurrent nephrolithiasis, peptic ulcers, mental changes, and, less frequently, extensive bone resorption. However, with greater awareness of the disease and wider use of multiphasic screening tests, including measurements of blood calcium, the diagnosis is frequently made in patients who have no symptoms and minimal, if any, signs of the disease other than hypercalcemia and elevated levels of PTH. The manifestations may be subtle, and the disease may have a benign course for many years or a lifetime. This milder form of the disease is usually termed *asymptomatic HPT*. Rarely, hyperparathyroidism develops or worsens abruptly and causes severe complications such as marked dehydration and coma, so-called hypercalcemic parathyroid crisis.

The annual incidence of the disease is calculated to be as high as 0.2% in patients >60, with an estimated prevalence, including undiscovered asymptomatic patients, of ≥1%; some reports suggest the incidence may be declining. If confirmed, these changing estimates may reflect less frequent routine testing of serum calcium in recent years, earlier overestimates in incidence, or unknown factors. The disease has a peak incidence between the third and fifth decades but occurs in young children and in the elderly.

ETIOLOGY

Parathyroid tumors are most often encountered as isolated adenomas without other endocrinopathy. They may also arise in hereditary syndromes such as MEN syndromes. Parathyroid tumors may also arise as secondary to underlying disease (excessive stimulation in secondary hyperparathyroidism, especially chronic renal failure), or after other forms of excessive stimulation such as lithium therapy. These etiologies are discussed below.

Solitary Adenomas

A single abnormal gland is the cause in ~80% of patients; the abnormality in the gland is usually a benign neoplasm or adenoma and rarely a parathyroid carcinoma. Some surgeons and pathologists report that the enlargement of multiple

glands is common; double adenomas are reported. In ~15% of patients, all glands are hyperfunctioning; *chief cell parathyroid hyperplasia* is usually hereditary and frequently associated with other endocrine abnormalities.

Hereditary Syndromes and Multiple Parathyroid Tumors

Hereditary hyperparathyroidism can occur without other endocrine abnormalities but is usually part of a *multiple endocrine neoplasia* syndrome (Chap. 351). MEN 1 (Wermer's syndrome) consists of hyperparathyroidism and tumors of the pituitary and pancreas, often associated with gastric hypersecretion and peptic ulcer disease (Zollinger-Ellison syndrome). MEN 2A is characterized by pheochromocytoma and medullary carcinoma of the thyroid, as well as hyperparathyroidism; MEN 2B has additional associated features such as multiple neuromas but usually lacks hyperparathyroidism. Each of these MEN syndromes is transmitted in an apparent autosomal dominant manner, although, as noted below, the genetic basis of MEN 1 involves biallelic loss of a tumor suppressor.

The *hyperparathyroidism jaw tumor* (HPT-JT) syndrome occurs in families with parathyroid tumors (sometimes carcinomas) in association with benign jaw tumors. Some kindreds exhibit hereditary hyperparathyroidism without other endocrinopathies. This disorder is often termed *nonsyndromic familial isolated hyperparathyroidism* (FIHP). There is speculation that these families may be examples of variable expression of the other syndromes such as MEN 1, MEN 2, or the HPT-JT syndrome, but they may also have distinctive, still unidentified genetic causes.

PATHOLOGY

Adenomas are most often located in the inferior parathyroid glands, but in 6–10% of patients, parathyroid adenomas may be located in the thymus, the thyroid, the pericardium, or behind the esophagus. Adenomas are usually 0.5–5 g in size but may be as large as 10–20 g (normal glands weigh 25 mg on average). Chief cells are predominant in both hyperplasia and adenoma. With chief cell hyperplasia, the enlargement may be so asymmetric that some involved glands appear grossly normal. If generalized hyperplasia is present, however, histologic examination reveals a uniform pattern of chief cells and disappearance of fat even in the absence of an increase in gland weight. Thus, microscopic examination of biopsy specimens of several glands is essential to interpret findings at surgery.

Parathyroid carcinoma is often not aggressive. Long-term survival without recurrence is common if at initial surgery the entire gland is removed without rupture of the capsule. Recurrent parathyroid carcinoma is usually slow-growing with local spread in the neck, and surgical correction of recurrent disease may be feasible. Occasionally, however, parathyroid carcinoma is more aggressive, with distant metastases (lung, liver, and bone) found at the time of initial operation. It may be difficult to appreciate initially that a primary tumor is carcinoma; increased numbers of mitotic figures and increased fibrosis of the gland stroma may precede invasion. The diagnosis of carcinoma is often made in retrospect. Hyperparathyroidism from a parathyroid carcinoma may be indistinguishable from other forms of primary hyperparathyroidism but is usually more severe clinically. A potential clue to the diagnosis is offered by the degree of calcium elevation. Calcium values of 3.5–3.7 mmol/L (14–15 mg/dL) are frequent with carcinoma and may alert the surgeon to remove the abnormal gland with care to avoid capsular rupture. Recent findings concerning the genetic basis of parathyroid carcinoma (distinct from that of benign adenomas) indicate the need, in these kindreds, for family screening (see below).

GENETIC DEFECTS ASSOCIATED WITH HYPERPARATHYROIDISM

As in many other types of neoplasia, two fundamental types of genetic defects have been identified in parathyroid gland tumors: (1) overactivity of protooncogenes and (2) loss of function of tumor-suppressor genes. The former, by definition, can lead to uncontrolled cellular growth and function by activation (gain-of-function mutation) of a single allele of the responsible gene, whereas the latter requires loss of function of both allelic copies. Biallelic loss of function of a tumor-suppressor gene is usually characterized by a germ-line defect (all cells) and an additional somatic deletion/ mutation in the tumor **(Fig. 353-3)**.

Figure 353-3

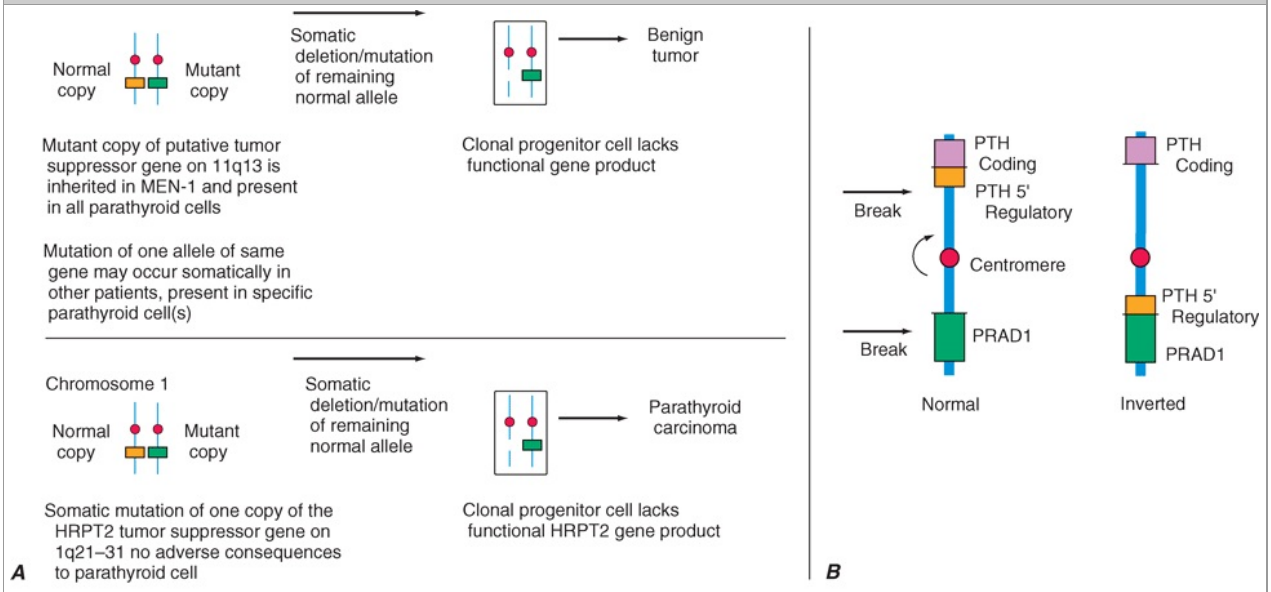

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

***A*.Schematic diagram indicating molecular events** in tumor susceptibility. The patient with the hereditary abnormality (multiple endocrine neoplasia, or MEN) is envisioned as having one defective gene inherited from the affected parent on chromosome 11, but one copy of the normal gene is present from the other parent. In the monoclonal tumor (benign tumor), a somatic event, here partial chromosomal deletion, removes the remaining normal gene from a cell. In nonhereditary tumors, two successive somatic mutations must occur, a process that takes a longer time. By either pathway, the cell, deprived of growth-regulating influence from this gene, has unregulated growth and becomes a tumor. A

pathway, the cell deprived of growth regulating influence from this gene, has unregulated growth and becomes a tumor. A different genetic locus also involving loss of a tumor-suppressor gene termed HRPT2 is involved in the pathogenesis of parathyroid carcinoma. *(From A Arnold: J Clin Endocrine Metab 77:1108, 1993. Copyright 1993, The Endocrine Society.)* ***B.*** **Schematic illustration of the mechanism and consequences of gene rearrangement** and overexpression of the PRAD 1 protooncogene (pericentromeric inversion of chromosome 11) in parathyroid adenomas. The excessive expression of PRAD 1 (a cell cycle control protein, cyclin D_1) by the highly active PTH gene promoter in the parathyroid cell contributes to excess cellular proliferation. *[From J Habener et al, in L DeGroot, JL Jameson (eds): Endocrinology, 4th ed. Philadelphia, Saunders, 2001; with permission.]*

Mutations in the *MEN1* gene locus, encoding the protein MENIN, on chromosome 11q13 are responsible for causing MEN 1; the normal allele of this gene fits the definition of a tumor-suppressor gene. Inheritance of one mutated allele in this hereditary syndrome, followed by loss of the other allele via somatic cell mutation, leads to monoclonal expansion and tumor development. Also, in ~15–20% of sporadic parathyroid adenomas, both alleles of the *MEN1* locus on chromosome 11 are somatically deleted, implying that the same defect responsible for MEN 1 can also cause the sporadic disease (Fig. 353-3*A*). Consistent with the Knudson hypothesis for two-step neoplasia in certain inherited cancer syndromes (Chap. 83), the earlier onset of hyperparathyroidism in the hereditary syndromes reflects the need for only one mutational event to trigger the monoclonal outgrowth. In sporadic adenomas, typically occurring later in life, two different somatic events must occur before the *MEN1* gene is silenced.

Other presumptive anti-oncogenes involved in hyperparathyroidism include a still unidentified gene mapped to chromosome 1p seen in 40% of sporadic parathyroid adenomas and a gene mapped to chromosome Xp11 in patients with secondary hyperparathyroidism and renal failure, who progressed to "tertiary" hyperparathyroidism, now known to reflect monoclonal outgrowths within previously hyperplastic glands.

A more complex pattern, still incompletely resolved, arises with genetic defects and carcinoma of the parathyroids. This appears to be due to biallelic loss of a functioning copy of a gene, *HRPT2* (or CDC73), originally identified as the cause of the HPT-JT syndrome. Several inactivating mutations have been identified in *HRPT2* (located on chromosome 1q21-31), which encodes a 531-amino-acid protein called parafibromin. The responsible genetic mutations in *HRPT2* appear to be necessary, but not sufficient, for parathyroid cancer.

In general, the detection of additional genetic defects in these parathyroid tumor–related syndromes and the variations seen in phenotypic expression/penetrance indicate the multiplicity of the genetic factors responsible. Nonetheless, the ability to detect the presence of the major genetic contributors has greatly aided a more informed management of family members of patients identified in the hereditary syndromes such as MEN 1, MEN 2, and HPT-JT.

An important contribution from studies on the genetic origin of parathyroid carcinoma has been the realization that the mutations involve a different pathway than that involved with the benign gland enlargements. Unlike the pathogenesis of genetic alterations seen in colon cancer, where lesions evolve from benign adenomas to malignant disease by progressive genetic changes, the alterations commonly seen in most parathyroid cancers (*HRPT2* mutations) are infrequently seen in sporadic parathyroid adenomas.

Abnormalities at the *Rb* gene were the first to be noted in parathyroid cancer. The *Rb* gene, a tumor-suppressor gene located on chromosome 13q14, was initially associated with retinoblastoma but has since been implicated in other neoplasias, including parathyroid carcinoma. Early studies implicated allelic deletions of the *Rb* gene in many parathyroid carcinomas and decreased or absent expression of the Rb protein. However, because there are often large deletions in chromosome 13 that include many genes in addition to the *Rb* locus (with similar findings in some pituitary carcinomas), it remains possible that other tumor-suppressor genes on chromosome 13 may be playing a role in parathyroid carcinoma.

Study of the parathyroid cancers found in some patients with the HPT-JT syndrome has led to identification of a much larger role for mutations in the *HRPT2* gene in most parathyroid carcinomas, including those that arise sporadically, without apparent association with the HPT-JT syndrome. Mutations in the coding region have been identified in 75–80% of all parathyroid cancers analyzed, leading to the conclusion that, with addition of presumed mutations in the noncoding regions, this genetic defect may be seen in essentially all parathyroid carcinomas. Of special importance was the discovery that, in some sporadic parathyroid cancers, germ-line mutations have been found; this, in turn, has led to careful investigation of the families of these patients and a new clinical indication for genetic testing in this setting.

Hypercalcemia occurring in family members (who are also found to have the germ-line mutations) can lead to the finding, at parathyroid surgery, of premalignant parathyroid tumors.

Overall, it seems there are multiple factors in parathyroid cancer, in addition to the *HRPT2* and *Rb* gene, although the *HRPT2* gene mutation is the most invariant abnormality. *RET* encodes a tyrosine kinase type receptor; specific inherited germ-line mutations lead to a constitutive activation of the receptor, thereby explaining the autosomal dominant mode of transmission and the relatively early onset of neoplasia. In the MEN 2 syndrome, the *RET* protooncogene may be responsible for the earliest disorder detected, the polyclonal disorder (C cell hyperplasia, which then is transformed into a clonal outgrowth—a medullary carcinoma with the participation of other, still uncharacterized genetic defects).

In some parathyroid adenomas, activation of a protooncogene has been identified (Fig. 353-3*B*). A reciprocal translocation involving chromosome 11 has been identified that juxtaposes the *PTH* gene promoter upstream of a gene product termed *PRAD-1*, encoding a cyclin D protein that plays a key role in normal cell division. This translocation plus other mechanisms that cause an equivalent overexpression of cyclin D1 are found in 20–40% of parathyroid adenomas.

Mouse models have confirmed the role of several of the major identified genetic defects in parathyroid disease and the MEN syndromes. Loss of the *MEN1* gene locus or overexpression of the *PRAD-1* protooncogene or the mutated *RET* protooncogene have been analyzed by genetic manipulation in mice, with the expected onset of parathyroid tumors or medullary carcinoma, respectively.

SIGNS AND SYMPTOMS

One-half or more of patients with hyperparathyroidism are asymptomatic. In series in which patients are followed without operation, as many as 80% are classified as without symptoms. Manifestations of hyperparathyroidism involve primarily the kidneys and the skeletal system. Kidney involvement, due either to deposition of calcium in the renal parenchyma or to recurrent nephrolithiasis, was present in 60–70% of patients prior to 1970. With earlier detection, renal complications occur in <20% of patients in many large series. Renal stones are usually composed of either calcium oxalate or calcium phosphate. In occasional patients, repeated episodes of nephrolithiasis or the formation of large calculi may lead to urinary tract obstruction, infection, and loss of renal function. Nephrocalcinosis may also cause decreased renal function and phosphate retention.

The distinctive bone manifestation of hyperparathyroidism is *osteitis fibrosa cystica*, which occurred in 10–25% of patients in series reported 50 years ago. Histologically, the pathognomonic features are an increase in the giant multinucleated

osteoclasts in scalloped areas on the surface of the bone (Howship's lacunae) and a replacement of the normal cellular and marrow elements by fibrous tissue. X-ray changes include resorption of the phalangeal tufts and replacement of the usually sharp cortical outline of the bone in the digits by an irregular outline (subperiosteal resorption). In recent years, osteitis fibrosa cystica is very rare in primary hyperparathyroidism, probably due to the earlier detection of the disease.

Dual-energy x-ray absorptiometry (DEXA) of the spine provides reproducible quantitative estimates (within a few percent) of spinal bone density. Similarly, bone density in the extremities can be quantified by densitometry of the hip or of the distal radius at a site chosen to be primarily cortical. CT is a very sensitive technique for estimating spinal bone density, but reproducibility of standard CT is no better than 5%. Newer CT techniques (spiral, "extreme" CT) are more reproducible but are currently available in a limited number of medical centers. Cortical bone density is reduced while cancellous bone density, especially in the spine, is relatively preserved. In symptomatic patients, dysfunctions of the CNS, peripheral nerve and muscle, gastrointestinal tract, and joints also occur. It has been reported that severe neuropsychiatric manifestations may be reversed by parathyroidectomy. When present in symptomatic patients, neuromuscular manifestations may include proximal muscle weakness, easy fatigability, and atrophy of muscles and may be so striking as to suggest a primary neuromuscular disorder. The distinguishing feature is the complete regression of neuromuscular disease after surgical correction of the hyperparathyroidism.

Gastrointestinal manifestations are sometimes subtle and include vague abdominal complaints and disorders of the stomach and pancreas. Again, cause and effect are unclear. In MEN 1 patients with hyperparathyroidism, duodenal ulcer may be the result of associated pancreatic tumors that secrete excessive quantities of gastrin (Zollinger-Ellison syndrome). Pancreatitis has been reported in association with hyperparathyroidism, but the incidence and the mechanism are not established.

Much attention has been paid in recent years to the manifestations of and optimum management strategies for asymptomatic hyperparathyroidism. This is now the most prevalent form of the disease. Asympomatatic primary hyperparathyroidism is defined as biochemically confirmed hyperparathyroidism (elevated or inappropriately normal PTH levels despite hypercalcemia) with the absence of signs and symptoms typically associated with more severe hyperparathyroidism such as features of renal or bone disease.

Three conferences on the topic have been held in the United States over the past two decades, with the most recent in 2008. The published proceedings include discussion of more subtle manifestations of disease, its natural history (without parathyroidectomy), and guidelines both for indications for surgery and medical monitoring in nonoperated patients.

Issues of concern include the potential for cardiovascular deterioration, the presence of subtle neuropsychiatric symptoms, and the longer-term status of skeletal integrity in patients not treated surgically. The current consensus is that medical monitoring rather than surgical correction of hyperparathyroidism may be justified in certain patients. The current recommendation is that patients who show mild disease, as defined by specific criteria **(Table 353-2)**, can be safely followed under management guidelines **(Table 353-3)**. There is, however, growing uncertainty about subtle disease manifestations and whether surgery is therefore indicated in most patients. Among the issues is the evidence of eventual (>8 years) deterioration in bone mineral density after a decade of relative stability. There is concern that this late-onset deterioration in bone density in nonoperated patients could contribute significantly to the well-known age-dependent fracture risk (osteoporosis). One study reported significant and sustained improvements in bone mineral density after successful parathyroidectomy, again raising the issue regarding benefits of surgery. Other randomized studies, however, did not report major gains post-surgery.

Table 353-2 Guidelines for Surgery in Asymptomatic Primary Hyperparathyroidisma

Parameter	Guideline
Serum calcium (above normal)	>1 mg/dL
24-h urinary Ca	No indication
Creatinine clearance (calculated)	If <60 mL/min
Bone density	*T* score <–2.5 at Any of 3 sitesb
Age	<50

[a]Bilezikian et al.

[b]Spine, distal radius, hip.

Table 353-3 Guidelines for Monitoring in Asymptomatic Primary Hyperparathyroidism[a]

Parameter	Guideline
Serum calcium	Annually
24-h urinary calcium	Not recommended
Creatinine clearance	Not recommended
Serum creatinine[b]	Annually
Bone density	Annually (3 sites)[a]

[a]Updates guidelines (Bilezikian et al).

[b]Creatinine clearance calculated by Cockcroft-Gault equation or MDRD equation.

Cardiovascular disease including left ventricular hypertrophy, cardiac functional defects, and endothelial dysfunction have been reported as reversible in European patients with more severe symptomatic disease after surgery, leading to numerous studies of these cardiovascular features in those with milder disease. There are reports of endothelial dysfunction in patients with mild asymptomatic hyperparathyroidism, but more observation is needed the expert panels concluded, especially whether there is reversibility with surgery.

A topic of considerable interest and some debate is assessment of neuropsychiatric status and health-related quality of life

(QOL) status in hyperparathyroid patients both before surgery and in response to parathyroidectomy. Several observational studies suggest considerable improvements in symptom score after surgery. Randomized studies of surgery versus observation, however, have yielded inconclusive results, especially regarding benefits of surgery. Most studies report that hyperparathyroidism is associated with increased neuropsychiatric symptoms, so the issue remains a significant factor in decisions regarding the impact of surgery in this disease.

Diagnosis

The diagnosis is typically made by detecting an elevated immunoreactive PTH level in a patient with asymptomatic hypercalcemia (see "Differential Diagnosis: Special Tests," below). Serum phosphate is usually low but may be normal, especially if renal failure has developed.

Several modifications in PTH assays have been introduced in efforts to improve their utility in light of information about metabolism of PTH (as discussed above). First-generation assays were based on displacement of radiolabeled PTH from antibodies that reacted with PTH (often also PTH fragments). Double-antibody or immunometric assays (one antibody that is usually directed against the carboxyl-terminal portion of intact PTH to capture the hormone and a second radio- or enzyme-labeled antibody that is usually directed against the amino-terminal portion of intact PTH) greatly improved the diagnostic discrimination of the tests by eliminating interference from circulating biologically inactive fragments, detected by the original first-generation assays. Double-antibody assays are now referred to as second-generation. Such PTH assays have in some centers and testing laboratories been replaced by third-generation assays after it was discovered that large PTH fragments, devoid of only the extreme amino-terminal portion of the PTH molecule, are also present in blood and are detected, incorrectly as intact PTH. These amino-terminally truncated PTH fragments were prevented from registering in the newer third-generation assays by use of a detection antibody directed against the extreme amino-terminal epitope. These assays may be useful for clinical research studies as in management of chronic renal disease, but the consensus is that either second- or third-generation assays are useful in the diagnosis of primary hyperparathyroidism and for the diagnosis of high-turnover bone disease in chronic kidney disease.

Many tests based on renal responses to excess PTH (renal calcium and phosphate clearance; blood phosphate, chloride, magnesium; urinary or nephrogenous cyclic AMP) were used in earlier decades. These tests have low specificity for hyperparathyroidism and are therefore not cost-effective; they have been replaced by PTH immunometric assays combined with simultaneous blood calcium measurements **(Fig. 353-4)**.

Figure 353-4

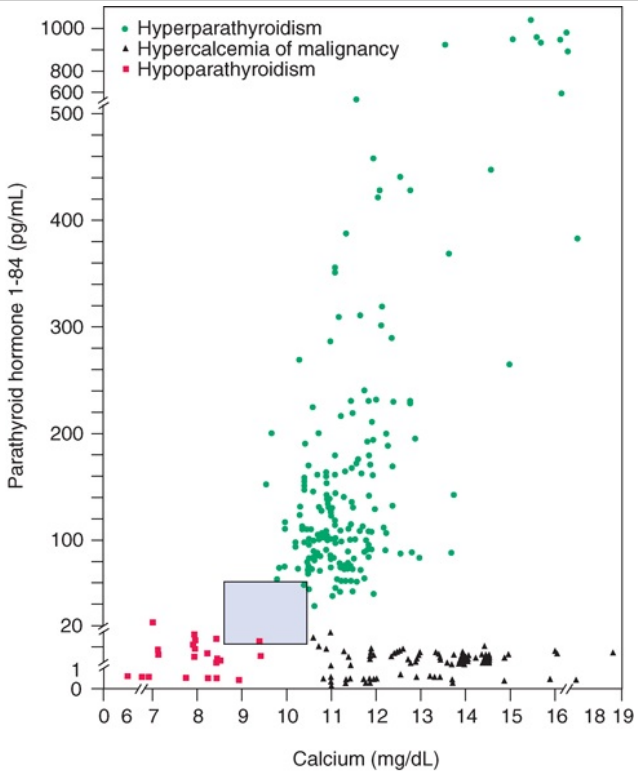

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Levels of immunoreactive parathyroid hormone (PTH) detected in patients with primary hyperparathyroidism, hypercalcemia of malignancy, and hypoparathyroidism. Boxed area represents the upper and normal limits of blood calcium and/or immunoreactive PTH. [*From SR Nussbaum, JT Potts, Jr, in L DeGroot, JL Jameson (eds): Endocrinology, 4th ed. Philadelphia, Saunders, 2001; with permission.*]

TREATMENT: HYPERPARATHYROIDISM

Surgical excision of the abnormal parathyroid tissue is the definitive therapy for this disease. As noted above, medical surveillance without operation for patients with mild, asymptomatic disease is, however, still preferred by some physicians and patients, particularly when the patients are more elderly. Evidence favoring surgery, if medically feasible, is growing because of concerns about skeletal, cardiovascular, and neuropsychiatric disease, even in mild hyperparathyroidism.

Two surgical approaches are generally practiced. The conventional parathyroidectomy procedure was neck exploration with general anesthesia; this procedure is being replaced in many centers, whenever feasible, by an outpatient procedure with local anesthesia, termed minimally invasive parathyroidectomy.

Parathyroid exploration is challenging and should be undertaken by an experienced surgeon. Certain features help in predicting the pathology (e.g., multiple abnormal glands in familial cases). However, some critical decisions regarding management can be made only during the operation.

With conventional surgery, one approach is still based on the view that typically only one gland (the adenoma) is abnormal. If an enlarged gland is found, a normal gland should be sought. In this view, if a biopsy of a normal-sized second gland confirms its histologic (and presumed functional) normality, no further exploration, biopsy, or excision is needed. At the other extreme is the minority viewpoint that all four glands be sought and that most of the total parathyroid tissue mass be removed. The concern with the former approach is that the recurrence rate of hyperparathyroidism may be high if a second abnormal gland is missed; the latter approach could involve unnecessary surgery and an unacceptable rate of hypoparathyroidism. When normal glands are found in association with one enlarged gland, excision of the single adenoma usually leads to cure or at least years free of symptoms. Long-term follow-up studies to establish true rates of recurrence are limited.

Recently, there has been growing experience with new surgical strategies that feature a minimally invasive approach guided by improved preoperative localization and intraoperative monitoring by PTH assays. Preoperative ^{99m}Tc sestamibi scans with single-photon emission CT (SPECT) are used to predict the location of an abnormal gland and intraoperative sampling of PTH before and at 5-minute intervals after removal of a suspected adenoma to confirm a rapid fall (>50%) to normal levels of PTH. In several centers, a combination of preoperative sestamibi imaging, cervical block anesthesia, minimal surgical incision, and intraoperative PTH measurements has allowed successful outpatient surgical management with a clear-cut cost benefit compared to general anesthesia and more extensive neck surgery. The use of these minimally invasive approaches requires clinical judgment to select patients unlikely to have multiple gland disease (e.g., MEN or secondary hyperparathyroidism). The growing acceptance of the technique and its relative ease for the patient has lowered the threshold for surgery.

Usually the severity of the hypercalcemia provides a preoperative clue to parathyroid carcinoma. In such cases, when neck exploration is undertaken, the tissue should be widely excised; care is taken to avoid rupture of the capsule to prevent local seeding of tumor cells.

Multiple-gland hyperplasia, as predicted in familial cases, poses more difficult questions of surgical management. Once a diagnosis of hyperplasia is established, all the glands must be identified. Two schemes have been proposed for surgical management. One is to totally remove three glands with partial excision of the fourth gland; care is taken to leave a good blood supply for the remaining gland. Other surgeons advocate total parathyroidectomy with immediate transplantation of a portion of a removed, minced parathyroid gland into the muscles of the forearm, with the view that surgical excision is easier from the ectopic site in the arm if there is recurrent hyperfunction.

In a minority of cases, if no abnormal parathyroid glands are found in the neck, the issue of further exploration must be decided. There are documented cases of five or six parathyroid glands and of unusual locations for adenomas such as in the mediastinum.

When a second parathyroid exploration is indicated, the minimally invasive techniques for preoperative localization such as ultrasound, CT scan, and isotope scanning are combined with venous sampling and/or selective digital arteriography in one of the centers specializing in these procedures. Intraoperative monitoring of PTH levels by rapid PTH immunoassays may be useful in guiding the surgery. At one center, long-term cures have been achieved with selective embolization or injection of large amounts of contrast material into the end-arterial circulation feeding the parathyroid tumor.

A decline in serum calcium occurs within 24 hours after successful surgery; usually blood calcium falls to low-normal values for 3–5 days until the remaining parathyroid tissue resumes full hormone secretion. Acute postoperative hypocalcemia is likely only if severe bone mineral deficits are present or if injury to all the normal parathyroid glands occurs during surgery. In general, there are few problems encountered in patients with uncomplicated disease such as a single adenoma (the clear majority), who do not have symptomatic bone disease nor a large deficit in bone mineral, who are vitamin D and magnesium sufficient, and who have good renal and gastrointestinal function. The extent of postoperative hypocalcemia varies with the surgical approach. If all glands are biopsied, hypocalcemia may be transiently symptomatic and more prolonged. Hypocalcemia is more likely to be symptomatic after second parathyroid explorations, particularly when normal parathyroid tissue was removed at the initial operation and when the manipulation and/or biopsy of the remaining normal glands are more extensive in the search for the missing adenoma.

Patients with hyperparathyroidism have efficient intestinal calcium absorption due to the increased levels of $1,25(OH)_2D$ stimulated by PTH excess. Once hypocalcemia signifies successful surgery, patients can be put on a high-calcium intake or be given oral calcium supplements. Despite mild hypocalcemia, most patients do not require parenteral therapy. If the serum calcium falls to <2 mmol/L (8 mg/dL), *and if the phosphate level rises simultaneously* , the possibility that surgery has caused hypoparathyroidism must be considered. With unexpected hypocalcemia, coexistent hypomagnesemia should be considered, as it interferes with PTH secretion and causes functional hypoparathyroidism (Chap. 352).

Signs of hypocalcemia include symptoms such as muscle twitching, a general sense of anxiety, and positive Chvostek's and Trousseau's signs coupled with serum calcium consistently <2 mmol/L (8 mg/dL). Parenteral calcium replacement at a low level should be instituted when hypocalcemia is symptomatic. The rate and duration of IV therapy are determined by the severity of the symptoms and the response of the serum calcium to treatment. An infusion of 0.5–2 mg/kg per hour or 30–100 mL/h of a 1-mg/mL solution usually suffices to relieve symptoms. Usually, parenteral therapy is required for only a few days. If symptoms worsen or if parenteral calcium is needed for >2–3 days, therapy with a vitamin D analogue and/or oral calcium (2–4 g/d) should be started (see below). It is cost-effective to use calcitriol (doses of 0.5–1 μg/d) because of the rapidity of onset of effect and prompt cessation of action when stopped, in comparison to other forms of vitamin D. A rise in blood calcium after several months of vitamin D replacement may indicate restoration of parathyroid function to normal. It is also appropriate to monitor serum PTH serially to estimate gland function in such patients.

If magnesium deficiency was present, it can complicate the postoperative course since magnesium deficiency impairs the secretion of PTH. Hypomagnesemia should be corrected whenever detected. Magnesium replacement can be effective orally (e.g., $MgCl_2$, $MgOH_2$), but parenteral repletion is usual to ensure postoperative recovery, if magnesium deficiency is suspected due to low blood magnesium levels. Because the depressant effect of magnesium on central and peripheral nerve functions does not occur at levels <2 mmol/L (normal range 0.8–1.2 mmol/L), parenteral replacement can be given rapidly. A cumulative dose as great as 0.5–1 mmol/kg of body weight can be administered if severe hypomagnesemia is present; often, however, total doses of 20–40 mmol are sufficient.

Medical Management

The guidelines for recommending surgical intervention, if feasible (Table 353-2), as well as for monitoring patients with

asymptomatic hyperparathyroidism who elect not to undergo parathyroidectomy (Table 353-3), reflect the changes over time since the first conference on the topic in 1990. Medical monitoring rather than corrective surgery is still acceptable, but it is clear that surgical intervention is the more frequently recommended option for the reasons noted above. Tightened guidelines favoring surgery include lowering the recommended level of serum calcium elevation, more careful attention to skeletal integrity through reference to peak skeletal mass at baseline (T scores) rather than age-adjusted bone density (*Z* scores), as well as the presence of any fragility fracture. The other changes noted in the two guidelines (Tables 353-2 and 353-3) reflect accumulated experience and practical consideration, such as a difficulty in quantity of urine collections. Despite the usefulness of the guidelines, the importance of individual patient and physician judgment and preference are clear in all recommendations.

When surgery is not selected, or not medically feasible, there is interest in the potential value of specific medical therapies. There is no long-term experience regarding specific clinical outcomes such as fracture prevention, but it has been established that bisphosphonates increase bone mineral density significantly without changing serum calcium (as does estrogen, but the latter is not favored because of reported adverse effects in other organ systems). Calcimimetics that lower PTH secretion lower calcium but do not affect bone mass density (BMD).

Other Parathyroid-Related Causes of Hypercalcemia

LITHIUM THERAPY

Lithium, used in the management of bipolar depression and other psychiatric disorders, causes hypercalcemia in ~10% of treated patients. The hypercalcemia is dependent on continued lithium treatment, remitting and recurring when lithium is stopped and restarted. The parathyroid adenomas reported in some hypercalcemic patients with lithium therapy may reflect the presence of an independently occurring parathyroid tumor; a permanent effect of lithium on parathyroid gland growth need not be implicated as most patients have complete reversal of hypercalcemia when lithium is stopped. However, long-standing stimulation of parathyroid cell replication by lithium may predispose to development of adenomas (as is documented in secondary hyperparathyroidism and renal failure).

At the levels achieved in blood in treated patients, lithium can be shown in vitro to shift the PTH secretion curve to the right in response to calcium; i.e., higher calcium levels are required to lower PTH secretion, probably acting at the calcium sensor (see below). This effect can cause elevated PTH levels and consequent hypercalcemia in otherwise normal individuals. Fortunately, there are usually alternative medications for the underlying psychiatric illness. Parathyroid surgery should not be recommended unless hypercalcemia and elevated PTH levels persist after lithium is discontinued.

Genetic Disorders Causing Hyperparathyroid-Like Syndromes

FAMILIAL HYPOCALCIURIC HYPERCALCEMIA

FHH (also called *familial benign hypercalcemia*) is inherited as an autosomal dominant trait. Affected individuals are discovered because of asymptomatic hypercalcemia. FHH, which is caused by an inactivating mutation in a single allele of the calcium sensing receptor (see below), involves inappropriately normal or even increased secretion of PTH, whereas another hypercalcemic disorder, namely the exceedingly rare Jansen's disease, is caused by a constitutively active PTH/PTHrP receptor in target tissues. Neither FHH nor Jansen's disease, however, is a growth disorder of the parathyroids.

The pathophysiology of FHH is now understood. The primary defect is abnormal sensing of the blood calcium by the parathyroid gland and renal tubule, causing inappropriate secretion of PTH and excessive renal reabsorption of calcium. The calcium-sensing receptor is a member of the third family of GPCRs (type C, or III). The receptor responds to increased ECF calcium concentration by suppressing PTH secretion through second-messenger signaling, thereby providing negative-feedback regulation of PTH secretion. Many different inactivating mutations in the calcium-sensing receptor have been identified in patients with FHH. These mutations lower the capacity of the sensor to bind calcium, and the mutant receptors function as though blood calcium levels were low; excessive secretion of PTH occurs from an otherwise normal gland. Approximately two-thirds of patients with FHH have mutations within the protein-coding region of the gene. The remaining one-third of kindreds may have mutations in the gene promoter or may involve still unknown mechanisms in other regions of the genome identified through mapping studies (e.g., chromosome 19p and 19q).

Even before elucidation of the pathophysiology of FHH, abundant clinical evidence served to separate the disorder from primary hyperparathyroidism; these clinical features are still useful in differential diagnosis. Patients with primary hyperparathyroidism have <99% renal calcium reabsorption, whereas most patients with FHH have >99% reabsorption. The hypercalcemia in FHH is often detectable in affected members of the kindreds in the first decade of life, whereas hypercalcemia rarely occurs in patients with primary hyperparathyroidism or the MEN syndromes who are aged <10 years. PTH may be elevated in FHH, but the values are usually normal or lower for the same degree of calcium elevation that is observed in patients with primary hyperparathyroidism. Parathyroid surgery performed in a few patients with FHH before the nature of the syndrome was understood led to permanent hypoparathyroidism; nevertheless, hypocalciuria persisted, establishing that hypocalciuria is not PTH-dependent (now known to be due to the abnormal calcium-sensing receptor in the kidney).

Few clinical signs or symptoms are present in patients with FHH, and other endocrine abnormalities are not present. Most patients are detected as a result of family screening after hypercalcemia is detected in a proband. In those patients inadvertently operated upon, the parathyroids appeared normal or moderately hyperplastic. Parathyroid surgery is not appropriate, nor, in view of the lack of symptoms, does medical treatment seem needed to lower the calcium. One striking exception to the rule against parathyroid surgery in this syndrome is the occurrence, usually in consanguineous marriages (due to the rarity of the gene mutation), of a homozygous or compound heterozygote state, resulting in severe impairment of calcium-sensing receptor function. In this condition, neonatal severe hypercalcemia, total parathyroidectomy is mandatory. Rare but well-documented cases of acquired hypocalciuric hypercalcemia are reported due to antibodies against the calcium-sensing receptor. They appear to be a complication of an underlying autoimmune disorder and respond to therapies directed against the underlying disorder.

JANSEN'S DISEASE

Activating mutations in the PTH/PTHrP receptor (PTH1R) have been identified as the cause of this rare autosomal dominant syndrome. Because the mutations lead to constitutive receptor function, one abnormal copy of the mutant receptor is sufficient to cause the disease, thereby accounting for its dominant mode of transmission. The disorder leads to short-limbed dwarfism due to abnormal regulation of chondrocyte maturation in the growth plates of the bone that are formed through an endochondral process. In adult life, there are numerous abnormalities in bone, including multiple cystic resorptive areas resembling those seen in severe hyperparathyroidism. Hypercalcemia and hypophosphatemia with undetectable or low PTH levels are typically seen. The pathogenesis of the growth plate abnormalities in Jansen's disease

has been confirmed by transgenic experiments in which targeted expression of the mutant PTH/PTHrP receptor to the proliferating chondrocyte layer of growth plate emulated several features of the human disorder. **Figure 353-5** illustrates some of these genetic mutations in the parathyroid gland or PTH target cells that affect Ca2 metabolism.

Figure 353-5

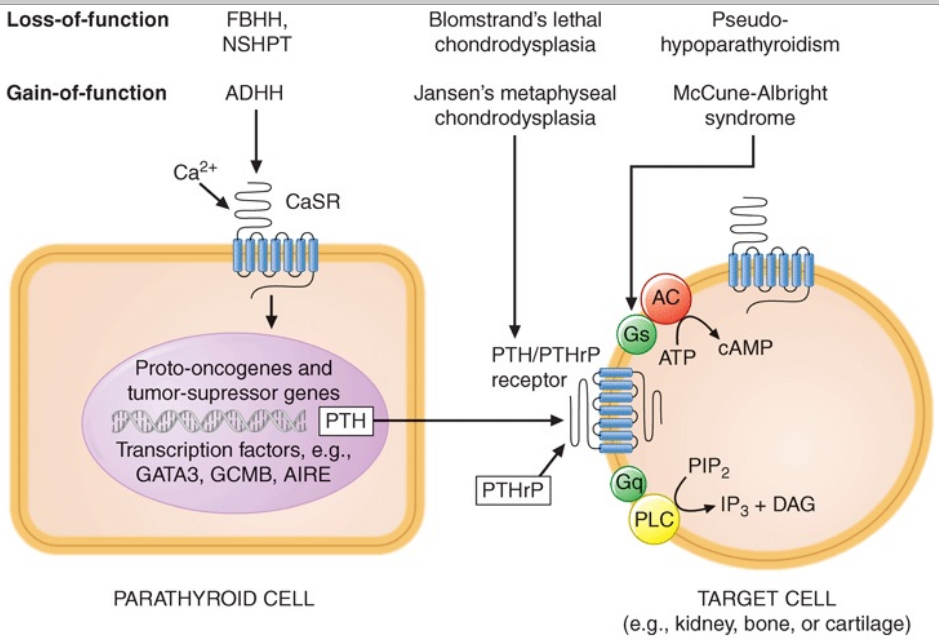

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Illustration of some genetic mutations that alter calcium metabolism by effects on the parathyroid cell or target cells of PTH action. Alterations in PTH production by the parathyroid cell can be caused by changes in the response to extracellular fluid calcium (Ca2+) that are detected by the calcium-sensing receptor (CaSR). Furthermore, PTH (or PTHrP) can show altered efficacy in target cells such as in proximal tubular cells, by altered function of its receptor (PTH/PTHrP receptor) or the signal transduction proteins, G proteins such as Gsα that is linked to adenylate cyclase (AC), the enzyme responsible for producing cyclic AMP (cAMP) [also illustrated is Gq, which activates an alternate pathway of receptor signal transmission involving the generation of inositol triphosphate (IP3) or diacylglycerol (DAG)]. Heterozygous loss-of-function mutations in the CaSR cause familial benign hypocalciuric hypercalcemia (FBHH) and homozygous mutations (both alleles mutated) and severe neonatal hyperparathyroidism (NSHPT); heterozygous gain-of-functions causes autosomal dominant hypercalciuric hypocalcemia (ADHH). Other defects in parathyroid cell function that occur at the level of gene regulation (oncogenes or tumor suppressor genes) or transcription factors are discussed in the text. Blomstrand's lethal chondrodysplasia is due to homozygous or compound heterozygous loss-of-function mutations in the PTH/PTHrP receptor, a neonatally lethal disorder, while pseudohypoparathyroidism involves inactivation at the level of the G proteins, specifically mutations that eliminate or reduce Gsα activity in the kidney (see text for details). Jansen's metaphyseal chondrodysplasia and McCune-Albright syndrome represent gain-of-function mutations in the PTH/PTHrP receptor and Gsα protein, respectively.

Malignancy-Related Hypercalcemia

CLINICAL SYNDROMES AND MECHANISMS OF HYPERCALCEMIA

Hypercalcemia due to malignancy is common (occurring in as many as 20% of cancer patients, especially with certain types of tumor such as lung carcinoma), often severe and difficult to manage, and, on rare occasions, difficult to distinguish from primary hyperparathyroidism. Although malignancy is often clinically obvious or readily detectable by medical history, hypercalcemia can occasionally be due to an occult tumor. Previously, hypercalcemia associated with malignancy was thought to be due to local invasion and destruction of bone by tumor cells; many cases are now known to result from the elaboration by the malignant cells of humoral mediators of hypercalcemia. PTHrP is the responsible humoral agent in most solid tumors that cause hypercalcemia.

The histologic character of the tumor is more important than the extent of skeletal metastases in predicting hypercalcemia. Small cell carcinoma (oat cell) and adenocarcinoma of the lung, though the most common lung tumors associated with skeletal metastases, rarely cause hypercalcemia. By contrast, many patients with squamous cell carcinoma of the lung develop hypercalcemia. Histologic studies of bone in patients with squamous cell or epidermoid carcinoma of the lung, in sites invaded by tumor as well as areas remote from tumor invasion, reveal increased bone resorption.

Two main mechanisms of hypercalcemia are operative in cancer hypercalcemia. Many solid tumors associated with hypercalcemia, particularly squamous cell and renal tumors, produce and secrete PTHrP that causes increased bone resorption and mediate the hypercalcemia through systemic actions on the skeleton. Alternatively, direct bone marrow invasion occurs with hematologicmalignancies such as leukemia, lymphoma, and multiple myeloma. Lymphokines and cytokines (including PTHrP) produced by cells involved in the marrow response to the tumors promote resorption of bone through local destruction. Several hormones, hormone analogues, cytokines, and growth factors have been implicated as the result of clinical assays, in vitro tests, or chemical isolation. The etiologic factor produced by activated normal lymphocytes and by myeloma and lymphoma cells, originally termed *osteoclast activation factor* , now appears to represent the biologic action of several different cytokines, probably interleukin 1 and lymphotoxin or tumor necrosis factor (TNF). In some lymphomas, there is a third mechanism, caused by an increased blood level of $1,25(OH)_2D$, produced by the abnormal lymphocytes.

In the more common mechanism, usually termed *humoral hypercalcemia of malignancy* , solid tumors (cancers of the lung and kidney, in particular), in which bone metastases are absent, minimal, or not detectable clinically, secrete PTHrP measurable by immunoassay. Secretion by the tumors of the PTH-like factor, PTHrP, activates the PTH1R, resulting in a

pathophysiology closely resembling hyperparathyroidism. The clinical picture resembles primary hyperparathyroidism (hypophosphatemia accompanies hypercalcemia), and elimination or regression of the primary tumor leads to disappearance of the hypercalcemia.

As in hyperparathyroidism, patients with the humoral hypercalcemia of malignancy have elevated urinary nephrogenous cyclic AMP excretion, hypophosphatemia, and increased urinary phosphate clearance. However, in humoral hypercalcemia of malignancy, immunoreactive PTH is undetectable or suppressed, making the differential diagnosis easier. Other features of the disorder differ from those of true hyperparathyroidism. Although the biologic actions of PTH and PTHrP are exerted through the same receptor, subtle differences in receptor activation by the two ligands must account for some of the discordance in pathophysiology, when an excess of one or the other peptide occurs. Other cytokines elaborated by the malignancy may contribute to the variations from hyperparathyroidism in these patients as well. Patients with humoral hypercalcemia of malignancy may have low to normal levels of $1,25(OH)_2D$ instead of elevated levels as in true hyperparathyroidism. In some patients with the humoral hypercalcemia of malignancy, osteoclastic resorption is unaccompanied by an osteoblastic or bone-forming response, implying inhibition of the normal coupling of bone formation and resorption.

Several different assays (single- or double-antibody, different epitopes) have been developed to detect PTHrP. Most data indicate that circulating PTHrP levels are undetectable (or low) in normal individuals except perhaps in pregnancy (high in human milk) and elevated in most cancer patients with the humoral syndrome. The etiologic mechanisms in cancer hypercalcemia may be multiple in the same patient. For example, in breast carcinoma (metastatic to bone) and in a distinctive type of T cell lymphoma/leukemia initiated by human T cell lymphotropic virus I, hypercalcemia is caused by direct local lysis of bone as well as by a humoral mechanism involving excess production of PTHrP. Hyperparathyroidism has been reported to coexist with the humoral cancer syndrome and, rarely, ectopic hyperparathyroidism due to tumor elaboration of true PTH is reported.

DIAGNOSTIC ISSUES

Levels of PTH measured by the double-antibody technique are undetectable or extremely low in tumor hypercalcemia, as would be expected with the mediation of the hypercalcemia by a factor other than PTH (the hypercalcemia suppresses the normal parathyroid glands). In a patient with minimal symptoms referred for hypercalcemia, low or undetectable PTH levels would focus attention on a possible occult malignancy (except for very rare cases of ectopic hyperparathyroidism).

Ordinarily, the diagnosis of cancer hypercalcemia is not difficult because tumor symptoms are prominent when hypercalcemia is detected. Indeed, hypercalcemia may be noted incidentally during the workup of a patient with known or suspected malignancy. Clinical suspicion that malignancy is the cause of the hypercalcemia is heightened when there are other signs or symptoms of a paraneoplastic process such as weight loss, fatigue, muscle weakness, or unexplained skin rash, or when symptoms specific for a particular tumor are present. Squamous cell tumors are most frequently associated with hypercalcemia, particularly tumors of the lung, kidney, head and neck, and urogenital tract. Radiologic examinations can focus on these areas when clinical evidence is unclear. Bone scans with technetium-labeled bisphosphonate are useful for detection of osteolytic metastases; the sensitivity is high, but specificity is low; results must be confirmed by conventional x-rays to be certain that areas of increased uptake are due to osteolytic metastases per se. Bone marrow biopsies are helpful in patients with anemia or abnormal peripheral blood smears.

TREATMENT: MALIGNANCY-RELATED HYPERCALCEMIA

Treatment of the hypercalcemia of malignancy is first directed to control of tumor; reduction of tumor mass usually corrects hypercalcemia. If a patient has severe hypercalcemia yet has a good chance for effective tumor therapy, treatment of the hypercalcemia should be vigorous while awaiting the results of definitive therapy. If hypercalcemia occurs in the late stages of a tumor that is resistant to antitumor therapy, the treatment of the hypercalcemia should be judicious as high calcium levels can have a mild sedating effect. Standard therapies for hypercalcemia (discussed below) are applicable to patients with malignancy.

Vitamin D–Related Hypercalcemia

Hypercalcemia caused by vitamin D can be due to excessive ingestion or abnormal metabolism of the vitamin. Abnormal metabolism of the vitamin is usually acquired in association with a widespread granulomatous disorder. Vitamin D metabolism is carefully regulated, particularly the activity of renal 1α-hydroxylase, the enzyme responsible for the production of $1,25(OH)_2D$ (Chap. 352). The regulation of 1α-hydroxylase and the normal feedback suppression by $1,25(OH)_2D$ seem to work less well in infants than in adults and to operate poorly, if at all, in sites other than the renal tubule; these phenomena may explain the occurrence of hypercalcemia secondary to excessive $1,25(OH)_2D3$ production in infants with Williams' syndrome (see below) and in adults with sarcoidosis or lymphoma.

VITAMIN D INTOXICATION

Chronic ingestion of 40–100 times the normal physiologic requirement of vitamin D (amounts >40,000–100,000 U/d) is usually required to produce significant hypercalcemia in normal individuals. The stated upper limit of safe dietary intake is 2000 U/d (50 μg/d) in adults because of concerns about potential toxic effects of cumulative supraphysiologic doses. These recommendations are now regarded as too restrictive, since some estimates are that in elderly individuals in northern latitudes, 2000 U/daily or more may be necessary to avoid vitamin D insufficiency.

Hypercalcemia in vitamin D intoxication is due to an excessive biologic action of the vitamin, perhaps the consequence of increased levels of 25(OH)D rather than merely increased levels of the active metabolite $1,25(OH)_2D$ (the latter may not be elevated in vitamin D intoxication). 25(OH)D has definite, if low, biologic activity in the intestine and bone. The production of 25(OH)D is less tightly regulated than is the production of $1,25(OH)_2D$. Hence concentrations of 25(OH)D are elevated severalfold in patients with excess vitamin D intake.

The diagnosis is substantiated by documenting elevated levels of 25(OH)D >100 mg/mL. Hypercalcemia is usually controlled by restriction of dietary calcium intake and appropriate attention to hydration. These measures, plus discontinuation of vitamin D, usually lead to resolution of hypercalcemia. However, vitamin D stores in fat may be substantial, and vitamin D intoxication may persist for weeks after vitamin D ingestion is terminated. Such patients are responsive to glucocorticoids, which in doses of 100 mg/d of hydrocortisone or its equivalent usually return serum calcium levels to normal over several days; severe intoxication may require intensive therapy.

SARCOIDOSIS AND OTHER GRANULOMATOUS DISEASES

In patients with sarcoidosis and other granulomatous diseases, such as tuberculosis and fungal infections, excess $1,25(OH)_2D$ is synthesized in macrophages or other cells in the granulomas. Indeed, increased $1,25(OH)_2D$ levels have

been reported in anephric patients with sarcoidosis and hypercalcemia. Macrophages obtained from granulomatous tissue convert 25(OH)D to 1,25$(OH)_2$D at an increased rate. There is a positive correlation in patients with sarcoidosis between 25(OH)D levels (reflecting vitamin D intake) and the circulating concentrations of 1,25$(OH)_2$D, whereas normally there is no increase in 1,25$(OH)_2$D with increasing 25(OH)D levels due to multiple feedback controls on renal 1α-hydroxylase (Chap. 352). The usual regulation of active metabolite production by calcium and phosphate or by PTH does not operate in these patients. Clearance of 1,25$(OH)_2$D from blood may be decreased in sarcoidosis as well. PTH levels are usually low and 1,25$(OH)_2$D levels are elevated, but primary hyperparathyroidism and sarcoidosis may coexist in some patients.

Management of the hypercalcemia can often be accomplished by avoiding excessive sunlight exposure and limiting vitamin D and calcium intake. Presumably, however, the abnormal sensitivity to vitamin D and abnormal regulation of 1,25$(OH)_2$D synthesis will persist as long as the disease is active. Alternatively, glucocorticoids in the equivalent of 100 mg/d of hydrocortisone or equivalent doses of glucorticoids may help control hypercalcemia. Glucocorticoids appear to act by blocking excessive production of 1,25$(OH)_2$D, as well as the response to it in target organs.

IDIOPATHIC HYPERCALCEMIA OF INFANCY

This rare disorder, usually referred to as *Williams' syndrome*, is an autosomal dominant disorder characterized by multiple congenital development defects, including supravalvular aortic stenosis, mental retardation, and an elfin facies, in association with hypercalcemia due to abnormal sensitivity to vitamin D. The hypercalcemia associated with the syndrome was first recognized in England after fortification of milk with vitamin D. The cardiac and developmental abnormalities were independently described, but the connection between these defects and hypercalcemia were not described until later. Levels of 1,25$(OH)_2$D can be elevated, ranging from 46 to 120 nmol/L (150–500 pg/mL). The mechanism of the abnormal sensitivity to vitamin D and of the increased circulating levels of 1,25$(OH)_2$D is still unclear. Studies suggest that genetic mutations involving microdeletions at the elastin locus and perhaps other genes on chromosome 7 may play a role in the pathogenesis.

Hypercalcemia Associated with High Bone Turnover

HYPERTHYROIDISM

As many as 20% of hyperthyroid patients have high-normal or mildly elevated serum calcium concentrations; hypercalciuria is even more common. The hypercalcemia is due to increased bone turnover, with bone resorption exceeding bone formation. Severe calcium elevations are not typical, and the presence of such suggests a concomitant disease such as hyperparathyroidism. Usually, the diagnosis is obvious, but signs of hyperthyroidism may occasionally be occult, particularly in the elderly (Chap. 341). Hypercalcemia is managed by treatment of the hyperthyroidism. Reports that thyroid-stimulating hormone (TSH) itself normally has a bone-protective effect suggest that suppressed TSH levels also play a role in hypercalcemia.

IMMOBILIZATION

Immobilization is a rare cause of hypercalcemia in adults in the absence of an associated disease but may cause hypercalcemia in children and adolescents, particularly after spinal cord injury and paraplegia or quadriplegia. With resumption of ambulation, the hypercalcemia in children usually returns to normal.

The mechanism appears to involve a disproportion between bone formation and bone resorption; the former decreased and the latter increased. Hypercalciuria and increased mobilization of skeletal calcium can develop in normal volunteers subjected to extensive bed rest, although hypercalcemia is unusual. Immobilization of an adult with a disease associated with high bone turnover, however, such as Paget's disease, may cause hypercalcemia.

THIAZIDES

Administration of benzothiadiazines (thiazides) can cause hypercalcemia in patients with high rates of bone turnover such as patients with hypoparathyroidism treated with high doses of vitamin D. Traditionally, thiazides are associated with aggravation of hypercalcemia in primary hyperparathyroidism, but this effect can be seen in other high-bone-turnover states as well. The mechanism of thiazide action is complex. Chronic thiazide administration leads to reduction in urinary calcium; the hypocalciuric effect appears to reflect the enhancement of proximal tubular resorption of sodium and calcium in response to sodium depletion. Some of this renal effect is due to augmentation of PTH action and is more pronounced in individuals with intact PTH secretion. However, thiazides cause hypocalciuria in hypoparathyroid patients on high-dose vitamin D and oral calcium replacement if sodium intake is restricted. This finding is the rationale for the use of thiazides as an adjunct to therapy in hypoparathyroid patients, as discussed below. Thiazide administration to normal individuals causes a transient increase in blood calcium (usually within the high-normal range) that reverts to preexisting levels after a week or more of continued administration. If hormonal function and calcium and bone metabolism are normal, homeostatic controls are reset to counteract the calcium-elevating effect of the thiazides. In the presence of hyperparathyroidism or increased bone turnover from another cause, homeostatic mechanisms are ineffective. The abnormal effects of the thiazide on calcium metabolism disappear within days of cessation of the drug.

VITAMIN A INTOXICATION

Vitamin Aintoxication is a rare cause of hypercalcemia and is most commonly a side effect of dietary faddism (Chap. 74). Calcium levels can be elevated into the 3–3.5-mmol/L (12–14 mg/dL) range after the ingestion of 50,000–100,000 units of vitamin A daily (10–20 times the minimum daily requirement). Typical features of severe hypercalcemia include fatigue, anorexia, and, in some, severe muscle and bone pain. Excess vitamin A intake is presumed to increase bone resorption.

The diagnosis can be established by history and by measurement of vitamin A levels in serum. Occasionally, skeletal x-rays reveal periosteal calcifications, particularly in the hands. Withdrawal of the vitamin is usually associated with prompt disappearance of the hypercalcemia and reversal of the skeletal changes. As in vitamin Dintoxication, administration of 100 mg/d hydrocortisone or its equivalent leads to a rapid return of the serum calcium to normal.

Hypercalcemia Associated with Renal Failure

SEVERE SECONDARY HYPERPARATHYROIDISM

The pathogenesis of secondary hyperparathyroidism in chronic kidney disease is incompletely understood. Resistance to the normal level of PTH is a major factor contributing to the development of hypocalcemia, which, in turn, is a stimulus to parathyroid gland enlargement. However, recent findings have indicated that an increase of FGF23 production by osteocytes (and possibly osteoblasts) in bone occurs well before an elevation in PTH is detected. FGF23 is a potent inhibitor of the renal 1-alpha hydroxylase, and the FGF23-dependent reduction in 1,25(OH)2 vitamin D may be an important stimulus for the development of secondary hyperparathyroidism.

Secondary hyperparathyroidism occurs not only in patients with renal failure but also in those with osteomalacia due to multiple causes (Chap. 352), including deficiency of vitamin D action and PHP (deficient response to PTH at the level of the receptor). For both disorders, hypocalcemia seems to be the common denominator in initiating the development of secondary hyperparathyroidism. Primary (1°) and secondary (2°) hyperparathyroidism can be distinguished conceptually by the autonomous growth of the parathyroid glands in primary hyperparathyroidism (presumably irreversible) and the adaptive response of the parathyroids in secondary hyperparathyroidism (typically reversible). In fact, reversal over weeks from an abnormal pattern of secretion, presumably accompanied by involution of parathyroid gland mass to normal, occurs in patients with osteomalacia who have been treated effectively with calcium and vitamin D. However, it is now recognized that a true clonal outgrowth (irreversible) can arise in long-standing, inadequately treated chronic renal failure [e.g., tertiary (3°) hyperparathyroidism; see below].

Patients with secondary hyperparathyroidism may develop bone pain, ectopic calcification, and pruritus. The bone disease seen in patients with secondary hyperparathyroidism and chronic kidney disease is termed *renal osteodystrophy* and affects primarily bone turnover. However, osteomalacia is frequently encountered as well and may be related to the circulating levels of FGF23.

Two other skeletal disorders have been frequently associated in the past with CKD patients treated by long-term dialysis who received aluminum-containing phosphate binders. Aluminum deposition in bone (see below) leads to an osteomalacia-like picture. The other entity is a low-bone-turnover state termed "aplastic" or "adynamic" bone disease; PTH levels are lower than typically observed in CKD patients with secondary hyperparathyroidism. It is believed that the condition is caused, at least in part, by excessive PTH suppression, which may be even greater than previously appreciated in light of evidence that some of the immunoreactive PTH detected by most commercially available PTH assays is not the full-length biologically active molecule (as discussed above) but may consist of amino-terminally truncated fragments that do not activate the PTH1R.

TREATMENT: SECONDARY HYPERPARATHYROIDISM

Medical therapy to reverse secondary hyperparathyroidism in CKD includes reduction of excessive blood phosphate by restriction of dietary phosphate, the use of nonabsorbable antacids, and careful, selective addition of calcitriol (0.25–2 μg/d) or related analogues. Calcium carbonate became preferred over aluminum-containing antacids to prevent aluminum-induced bone disease. However, synthetic gels that also bind phosphate (such as sevelamer; Chap. 280) are now widely used, with the advantage of avoiding not only aluminum retention but excess calcium elevation. Intravenous calcitriol (or related analogues), administered as several pulses each week, helps control secondary hyperparathyroidism. Aggressive but carefully administered medical therapy can often, but not always, reverse hyperparathyroidism and its symptoms and manifestations.

Occasional patients develop severe manifestations of secondary hyperparathyroidism, including hypercalcemia, pruritus, extraskeletal calcifications, and painful bones, despite aggressive medical efforts to suppress the hyperparathyroidism. PTH hypersecretion no longer responsive to medical therapy, a state of severe hyperparathyroidism in patients with renal failure that requires surgery, has been referred to as *tertiary hyperparathyroidism*. Parathyroid surgery is necessary to control this condition. Based on genetic evidence from examination of tumor samples in these patients, the emergence of autonomous parathyroid function is due to a monoclonal outgrowth of one or more previously hyperplastic parathyroid glands. The adaptive response has become an independent contributor to disease; this finding seems to emphasize the importance of optimal medical management to reduce the proliferative response of the parathyroid cells that enables the irreversible genetic change.

ALUMINUM INTOXICATION

Aluminum intoxication (and often hypercalcemia as a complication of medical treatment) in the past occurred in patients on chronic dialysis; manifestations included acute dementia and unresponsive and severe osteomalacia. Bone pain, multiple nonhealing fractures, particularly of the ribs and pelvis, and a proximal myopathy occur. Hypercalcemia develops when these patients are treated with vitamin D or calcitriol because of impaired skeletal responsiveness. Aluminum is present at the site of osteoid mineralization, osteoblastic activity is minimal, and calcium incorporation into the skeleton is impaired. The disorder is now rare because of the avoidance of aluminum-containing antacids or aluminum excess in the dialysis regimen (Chap. 360).

MILK-ALKALI SYNDROME

The milk-alkali syndrome is due to excessive ingestion of calcium and absorbable antacids such as milk or calcium carbonate. It is much less frequent since proton-pump inhibitors and other treatments became available for peptic ulcer disease. For a time, the increased use of calcium carbonate in the management of secondary hyperparathyroidism led to reappearance of the syndrome. Several clinical presentations—acute, subacute, and chronic—have been described, all of which feature hypercalcemia, alkalosis, and renal failure. The chronic form of the disease, termed *Burnett's syndrome*, is associated with irreversible renal damage. The acute syndromes reverse if the excess calcium and absorbable alkali are stopped.

Individual susceptibility is important in the pathogenesis, as some patients are treated with calcium carbonate and alkali regimens without developing the syndrome. One variable is the fractional calcium absorption as a function of calcium intake. Some individuals absorb a high fraction of calcium, even with intakes ≥2 g of elemental calcium per day, instead of reducing calcium absorption with high intake, as occurs in most normal individuals. Resultant mild hypercalcemia after meals in such patients is postulated to contribute to the generation of alkalosis. Development of hypercalcemia causes increased sodium excretion and some depletion of total-body water. These phenomena and perhaps some suppression of endogenous PTH secretion due to mild hypercalcemia lead to increased bicarbonate resorption and to alkalosis in the face of continued calcium carbonate ingestion. Alkalosis per se selectively enhances calcium resorption in the distal nephron, thus aggravating the hypercalcemia. The cycle of mild hypercalcemia → bicarbonate retention → alkalosis → renal calcium retention → severe hypercalcemia perpetuates and aggravates hypercalcemia and alkalosis as long as calcium and absorbable alkali are ingested.

Differential Diagnosis: Special Tests

Differential diagnosis of hypercalcemia is best achieved by using clinical criteria, but immunometric assays to measure PTH is especially useful in distinguishing among major causes **(Fig. 353-6)**. The clinical features that deserve emphasis are the presence or absence of symptoms or signs of disease and evidence of chronicity. If one discounts fatigue or depression, >90% of patients with primary hyperparathyroidism have *asymptomatic hypercalcemia*; symptoms of malignancy are usually present in cancer-associated hypercalcemia. Disorders other than hyperparathyroidism and malignancy cause <10% of cases of hypercalcemia, and some of the nonparathyroid causes are associated with clear-cut manifestations such as renal failure.

Figure 353-6

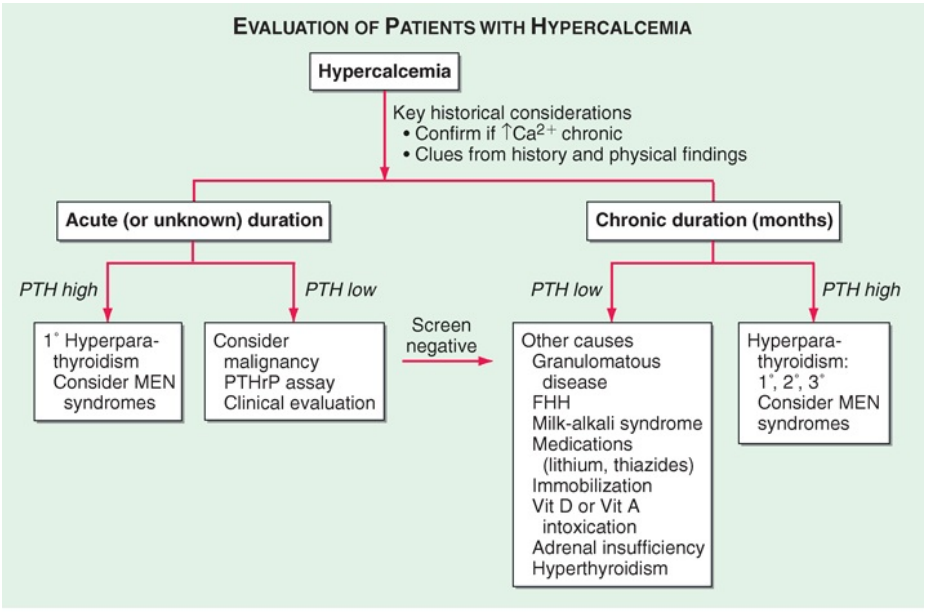

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the evaluation of patients with hypercalcemia. See text for details. FHH, familial hypocalciuric hypercalcemia; MEN, multiple endocrine neoplasia; PTH, parathyroid hormone; PTHrP, parathyroid hormone–related peptide.

Hyperparathyroidism is the likely diagnosis in patients with *chronic hypercalcemia*. If hypercalcemia has been manifest for >1 year, malignancy can usually be excluded as the cause. A striking feature of malignancy-associated hypercalcemia is the rapidity of the course, whereby signs and symptoms of the underlying malignancy are evident within months of the detection of hypercalcemia. Although clinical considerations are helpful in arriving at the correct diagnosis of the cause of hypercalcemia, appropriate laboratory testing is essential for definitive diagnosis. The immunoassay for PTH usually separates hyperparathyroidism from all other causes of hypercalcemia. There are very rare reports of ectopic production of excess PTH by nonparathyroid tumors. Patients with hyperparathyroidism have elevated PTH levels despite hypercalcemia, whereas patients with malignancy and the other causes of hypercalcemia (except for disorders mediated by PTH such as lithium-induced hypercalcemia) have levels of hormone below normal or undetectable. Assays based on the double-antibody method for PTH exhibit very high sensitivity (especially if serum calcium is simultaneously evaluated) and specificity for the diagnosis of primary hyperparathyroidism (Fig. 353-4).

In summary, PTH values are elevated in >90% of parathyroid-related causes of hypercalcemia, undetectable or low in malignancy-related hypercalcemia, and undetectable or normal in vitamin D–related and high-bone-turnover causes of hypercalcemia. In view of the specificity of the PTH immunoassay and the high frequency of hyperparathyroidism in hypercalcemic patients, it is cost-effective to measure the PTH level in all hypercalcemic patients unless malignancy or a specific nonparathyroid disease is obvious. False-positive PTH assay results are rare. Immunoassays for PTHrP are helpful in diagnosing certain types of malignancy-associated hypercalcemia. Although FHH is parathyroid-related, the disease should be managed distinctively from hyperparathyroidism. Clinical features and the low urinary calcium excretion can help make the distinction. Because the incidence of malignancy and hyperparathyroidism both increase with age, they can coexist as two independent causes of hypercalcemia.

$1,25(OH)_2D$ levels are elevated in many (but not all) patients with primary hyperparathyroidism. In other disorders associated with hypercalcemia, concentrations of $1,25(OH)_2D$ are low or, at the most, normal. However, this test is of low specificity and is not cost-effective, as not all patients with hyperparathyroidism have elevated $1,25(OH)_2D$ levels and not all nonparathyroid hypercalcemic patients have suppressed $1,25(OH)_2D$. Measurement of $1,25(OH)_2D$ is, however, critically valuable in establishing the cause of hypercalcemia in sarcoidosis and certain lymphomas.

A useful general approach is outlined in Fig. 353-6. If the patient is *asymptomatic* and there is evidence of *chronicity* to the hypercalcemia, hyperparathyroidism is almost certainly the cause. If PTH levels (usually measured at least twice) are elevated, the clinical impression is confirmed and little additional evaluation is necessary. If there is only a short history or no data as to the duration of the hypercalcemia, *occult malignancy* must be considered; if the PTH levels are not elevated, then a thorough workup must be undertaken for malignancy, including chest x-ray, CT of chest and abdomen, and bone scan. Immunoassays for PTHrP may be especially useful in such situations. Attention should also be paid to clues for underlying hematologic disorders such as anemia, increased plasma globulin, and abnormal serum immunoelectrophoresis; bone scans can be negative in some patients with metastases such as in multiple myeloma. Finally, if a patient with chronic hypercalcemia is asymptomatic and malignancy therefore seems unlikely on clinical grounds, but PTH values are not elevated, it is useful to search for other chronic causes of hypercalcemia such as occult sarcoidosis. A careful history of dietary supplements and drug use may suggest intoxication with vitamin D or vitamin A or the use of thiazides.

TREATMENT: HYPERCALCEMIC STATES

The approach to medical treatment of hypercalcemia varies with its severity **(Table 353-4)**. Mild hypercalcemia, <3 mmol/L (12 mg/dL), can be managed by hydration. More severe hypercalcemia [levels of 3.2–3.7 mmol/L (13–15 mg/dL)] must be managed aggressively; above that level, hypercalcemia can be life-threatening and requires emergency measures. By using a combination of approaches in severe hypercalcemia, the serum calcium concentration can be decreased by 0.7–2.2 mmol/L (3–9 mg/dL) within 24–48 hours in most patients, enough to relieve acute symptoms, prevent death from hypercalcemic crisis, and permit diagnostic evaluation. Therapy can then be directed at the underlying disorder—the

second priority.

Table 353-4 Therapies for Severe Hypercalcemia

Treatment	Onset of Action	Duration of Action	Advantages	Disadvantages
Most Useful Therapies				
Hydration with saline	Hours	During infusion	Rehydration invariably needed	Volume overload
Forced diuresis; saline plus loop diuretic	Hours	During treatment	Rapid action	Volume overload, cardiac decompensation, intensive monitoring, electrolyte disturbance, inconvenience
Bisphosphonates				
Pamidronate	1–2 days	10–14 days to weeks	High potency; intermediate onset of action	Fever in 20%, hypophosphatemia, hypocalcemia, hypomagnesemia, rarely jaw necrosis
Zolendronate	1–2 days	>3 weeks	Same as for pamidronate (may last longer)	Same as pamidronate above
Calcitonin	Hours	1–2 days	Rapid onset of action; useful as adjunct in severe hypercalcemia	Rapid tachyphylaxis
Special Use Therapies				
Phosphate Oral	24 h	During use	Chronic management (with hypophosphatemia); low toxicity if P < 4 mg/dL	Limited use except as adjuvant or chronic therapy
Glucocorticoids	Days	Days, weeks	Oral therapy, antitumor agent	Active only in certain malignancies, vitamin D excess and sarcoidosis; glucocorticoid side effects
Dialysis	Hours	During use and 24–48 h afterward	Useful in renal failure; onset of effect in hours; can immediately reverse life-threatening hypercalcemia	Complex procedure, reserved for extreme or special circumstances

Hypercalcemia develops because of excessive skeletal calcium release, increased intestinal calcium absorption, or inadequate renal calcium excretion. Understanding the particular pathogenesis helps guide therapy. For example, hypercalcemia in patients with malignancy is primarily due to excessive skeletal calcium release and is, therefore, minimally improved by restriction of dietary calcium. On the other hand, patients with vitamin D hypersensitivity or vitamin Dintoxication have excessive intestinal calcium absorption, and restriction of dietary calcium is beneficial. Decreased renal function or ECF depletion decreases urinary calcium excretion. In such situations, rehydration may rapidly reduce or reverse the hypercalcemia, even though increased bone resorption persists. As outlined below, the more severe the hypercalcemia, the greater the number of combined therapies that should be used. Rapid acting (hours) approaches —rehydration, forced diuresis, and calcitonin—can be used with the most effective antiresorptive agents such as bisphosphonates (since severe hypercalcemia usually involves excessive bone resorption).

Hydration, Increased Salt Intake, Mild and Forced Diuresis

The first principle of treatment is to restore normal hydration. Many hypercalcemic patients are dehydrated because of vomiting, inanition, and/or hypercalcemia-induced defects in urinary concentrating ability. The resultant drop in glomerular filtration rate is accompanied by an additional decrease in renal tubular sodium and calcium clearance. Restoring a normal ECF volume corrects these abnormalities and increases urine calcium excretion by 2.5–7.5 mmol/d (100–300 mg/d). Increasing urinary sodium excretion to 400–500 mmol/d increases urinary calcium excretion even further than simple rehydration. After rehydration has been achieved, saline can be administered or furosemide or ethacrynic acid can be given twice daily to depress the tubular reabsorptive mechanism for calcium (care must be taken to prevent dehydration). The combined use of these therapies can increase urinary calcium excretion to ≥12.5 mmol/d (500 mg/d) in most hypercalcemic patients. Since this is a substantial percentage of the exchangeable calcium pool, the serum calcium concentration usually falls 0.25–0.75 mmol/L (1–3 mg/dL) within 24 hours. Precautions should be taken to prevent potassium and magnesium depletion; calcium-containing renal calculi are a potential complication.

Under life-threatening circumstances, the preceding approach can be pursued more aggressively, but the availability of effective agents to block bone resorption (such as bisphosphonates) has reduced the need for extreme diuresis regimens (**Table 353-5**). Depletion of potassium and magnesium is inevitable unless replacements are given; pulmonary edema can be precipitated. The potential complications can be reduced by careful monitoring of central venous pressure and plasma or urine electrolytes; catheterization of the bladder may be necessary. Dialysis treatment may be needed when renal function is compromised.

Table 353-5 Functional Classification of Hypocalcemia (Excluding Neonatal Conditions)

PTH Absent	
Hereditary hypoparathyroidism	Hypomagnesemia
Acquired hypoparathyroidism	
PTH Ineffective	
Chronic renal failure	Active vitamin D ineffective
Active vitamin D lacking	Intestinal malabsorption
↓ Dietary intake or sunlight	Vitamin D–dependent rickets type II
Defective metabolism:	
Anticonvulsant therapy	Pseudohypoparathyroidism
Vitamin D–dependent rickets type I	

PTH Overwhelmed	
Severe, acute hyperphosphatemia Tumor lysis	Osteitis fibrosa after parathyroidectomy
Acute renal failure	
Rhabdomyolysis	

Abbreviation: PTH, parathyroid hormone.

Bisphosphonates

The bisphosphonates are analogues of pyrophosphate, with high affinity for bone, especially in areas of increased bone turnover, where they are powerful inhibitors of bone resorption. These bone-seeking compounds are stable in vivo because phosphatase enzymes cannot hydrolyze the central carbon-phosphorus-carbon bond. The bisphosphonates are concentrated in areas of high bone turnover and are taken up by and inhibit osteoclast action; the mechanism of action is complex. The bisphosphonate molecules that contain amino groups in the side chain structure (see below) interfere with prenylation of proteins and can lead to cellular apoptosis. The highly active nonamino group–containing bisphosphonates are also metabolized to cytotoxic products.

The initial bisphosphonate widely used in clinical practice, etidronate, was effective but had several disadvantages, including the capacity to inhibit bone formation as well as blocking resorption. Subsequently, a number of second- or third-generation compounds have become the mainstays of antiresorptive therapy for treatment of hypercalcemia and osteoporosis. The newer bisphosphonates have a highly favorable ratio of blocking resorption versus inhibiting bone formation; they inhibit osteoclast-mediated skeletal resorption yet do not cause mineralization defects at ordinary doses. Though the bisphosphonates have similar structures, the routes of administration, efficacy, toxicity, and side effects vary. The potency of the compounds for inhibition of bone resorption varies more than 10,000-fold, increasing in the order of etidronate, tiludronate, pamidronate, alendronate, risedronate, and zolendronate. The IV use of pamidronate and zolendronate is approved for the treatment of hypercalcemia; between 30 and 90 mg pamidronate, given as a single IV dose over a few hours, returns serum calcium to normal within 24–48 hours with an effect that lasts for weeks in 80–100% of patients. Zolendronate given in doses of 4 or 8 mg/5-minute infusion has a more rapid and more sustained effect than pamidronate in direct comparison.

These drugs are used extensively in cancer patients. Absolute survival improvements are noted with pamidronate and zolendronate in multiple myeloma, for example. However, though still rare, there are increasing reports of jaw necrosis, especially after dental surgery, mainly in cancer patients treated with multiple doses of the more potent bisphosphonates.

Calcitonin

Calcitonin acts within a few hours of its administration, principally through receptors on osteoclasts, to block bone resorption. Calcitonin, after 24 hours of use, is no longer effective in lowering calcium. Tachyphylaxis, a known phenomenon with this drug, seems to explain the results, since the drug is often effective in the first 24 hours of use. Therefore, in life-threatening hypercalcemia, calcitonin can be used effectively within the first 24 hours in combination with rehydration and saline diuresis while waiting for more sustained effects from a simultaneously administered bisphosphonate such as pamidronate. Usual doses of calcitonin are 2–8 U/kg of body weight IV, SC, or IM every 6–12 hours.

Other Therapies

Plicamycin (formerly mithramycin), which inhibits bone resorption, has been a useful therapeutic agent but is now seldom used because of its toxicity and the effectiveness of bisphosphonates. Plicamycin must be given IV, either as a bolus or by slow infusion; the usual dose is 25 μg/kg body weight. *Gallium nitrate* exerts a hypocalcemic action by inhibiting bone resorption and altering the structure of bone crystals. It is not often used now because of superior alternatives.

Glucocorticoids have utility, especially in hypercalcemia complicating certain malignancies. They increase urinary calcium excretion and decrease intestinal calcium absorption when given in pharmacologic doses, but they also cause negative skeletal calcium balance. In normal individuals and in patients with primary hyperparathyroidism, glucocorticoids neither increase nor decrease the serum calcium concentration. In patients with hypercalcemia due to certain osteolytic malignancies, however, glucocorticoids may be effective as a result of antitumor effects. The malignancies in which hypercalcemia responds to glucocorticoids include multiple myeloma, leukemia, Hodgkin's disease, other lymphomas, and carcinoma of the breast, at least early in the course of the disease. Glucocorticoids are also effective in treating hypercalcemia due to vitamin D intoxication and sarcoidosis. Glucocorticoids are also useful in the rare form of hypercalcemia, now recognized in certain autoimmune disorders in which inactivating antibodies against the receptor imitate FHH. Elevated PTH and calcium levels are effectively lowered by the glucocorticoids. In all the preceding situations, the hypocalcemic effect develops over several days, and the usual glucocorticoid dosage is 40–100 mg prednisone (or its equivalent) daily in four divided doses. The side effects of chronic glucocorticoid therapy may be acceptable in some circumstances.

Dialysis is often the treatment of choice for severe hypercalcemia complicated by renal failure, which is difficult to manage medically. Peritoneal dialysis with calcium-free dialysis fluid can remove 5–12.5 mmol (200–500 mg) of calcium in 24–48 hours and lower the serum calcium concentration by 0.7–3 mmol/L (3–12 mg/dL). Large quantities of phosphate are lost during dialysis, and serum inorganic phosphate concentration usually falls, potentially aggravating hypercalcemia. Therefore, the serum inorganic phosphate concentration should be measured after dialysis, and phosphate supplements should be added to the diet or to dialysis fluids if necessary.

Phosphate therapy, PO or IV, has a limited role in certain circumstances (Chap. 352). Correcting hypophosphatemia lowers the serum calcium concentration by several mechanisms, including bone/calcium exchange. The usual oral treatment is 1–1.5 g phosphorus per day for several days, given in divided doses. It is generally believed, but not established, that toxicity does not occur if therapy is limited to restoring serum inorganic phosphate concentrations to normal.

Raising the serum inorganic phosphate concentration above normal decreases serum calcium levels, sometimes strikingly. Intravenous phosphate is one of the most dramatically effective treatments available for severe hypercalcemia but is toxic and even dangerous (fatal hypocalcemia). For these reasons, it is used rarely and only in severely hypercalcemic patients with cardiac or renal failure where dialysis, the preferable alternative, is not feasible or is unavailable.

Summary

The various therapies for hypercalcemia are listed in Table 353-4. The choice depends on the underlying disease, the severity of the hypercalcemia, the serum inorganic phosphate level, and the renal, hepatic, and bone marrow function. Mild

hypercalcemia [≤3 mmol/L (12 mg/dL)] can usually be managed by hydration. Severe hypercalcemia [≥3.7 mmol/L (15 mg/dL)] requires rapid correction. Calcitonin should be given for its rapid, albeit short-lived, blockade of bone resorption, and IV pamidronate or zolendronate should be administered, although its onset of action is delayed for 1–2 days. In addition, for the first 24–48 hours, aggressive sodium-calciumdiuresis with IV saline should be given and, following rehydration, large doses of furosemide or ethacrynic acid, but only if appropriate monitoring is available and cardiac and renal function are adequate. Otherwise, dialysis may be necessary. Intermediate degrees of hypercalcemia between 3 and 3.7 mmol/L (12 and 15 mg/dL) should be approached with vigorous hydration and then the most appropriate selection for the patient of the combinations used with severe hypercalcemia.

HYPOCALCEMIA

(See also Chap. 46)

Pathophysiology of Hypocalcemia: Classification Based on Mechanism

Chronic hypocalcemia is less common than hypercalcemia; causes include chronic renal failure, hereditary and acquired hypoparathyroidism, vitamin D deficiency, PHP, and hypomagnesemia.

Acute rather than chronic hypocalcemia is seen in critically ill patients or as a consequence of certain medications and often does not require specific treatment. Transient hypocalcemia is seen with severe sepsis, burns, acute renal failure, and extensive transfusions with citrated blood. Although as many as one-half of patients in an intensive care setting are reported to have calcium concentrations <2.1 mmol/L (8.5 mg/dL), most do not have a reduction in ionized calcium. Patients with severe sepsis may have a decrease in ionized calcium (true hypocalcemia), but in other severely ill individuals, hypoalbuminemia is the primary cause of the reduced total calcium concentration. Alkalosis increases calcium binding to proteins, and in this setting direct measurements of ionized calcium should be made.

Medications such as protamine, heparin, and glucagon may cause transient hypocalcemia. These forms of hypocalcemia are usually not associated with tetany and resolve with improvement in the overall medical condition. The hypocalcemia after repeated transfusions of citrated blood usually resolves quickly.

Patients with *acute pancreatitis* have hypocalcemia that persists during the acute inflammation and varies in degree with the severity of the pancreatitis. The cause of hypocalcemia remains unclear. PTH values are reported to be low, normal, or elevated, and both resistance to PTH and impaired PTH secretion have been postulated. Occasionally, a chronic low total calcium and low ionized calcium concentration are detected in an elderly patient without obvious cause and with a paucity of symptoms; the pathogenesis is unclear.

Chronic hypocalcemia, however, is usually symptomatic and requires treatment. Neuromuscular and neurologic manifestations of chronic hypocalcemia include muscle spasms, carpopedal spasm, facial grimacing, and, in extreme cases, laryngeal spasm and convulsions. Respiratory arrest may occur. Increased intracranial pressure occurs in some patients with long-standing hypocalcemia, often in association with papilledema. Mental changes include irritability, depression, and psychosis. The QT interval on the electrocardiogram is prolonged, in contrast to its shortening with hypercalcemia. Arrhythmias occur, and digitalis effectiveness may be reduced. Intestinal cramps and chronic malabsorption may occur. Chvostek's or Trousseau's sign can be used to confirm latent tetany.

The classification of hypocalcemia shown in Table 353-5 is based on an organizationally useful premise that PTH is responsible for minute-to-minute regulation of plasma calcium concentration and, therefore, that the occurrence of hypocalcemia must mean a failure of the homeostatic action of PTH. Failure of the PTH response can occur if there is hereditary or acquired parathyroid gland failure, if PTH is ineffective in target organs, or if the action of the hormone is overwhelmed by the loss of calcium from the ECF at a rate faster than it can be replaced.

PTH Absent

Whether hereditary or acquired, hypoparathyroidism has a number of common components. Symptoms of untreated hypocalcemia are shared by both types of hypoparathyroidism, although the onset of hereditary hypoparathyroidism is more gradual and is often associated with other developmental defects. Basal ganglia calcification and extrapyramidal syndromes are more common and earlier in onset in hereditary hypoparathyroidism. In earlier decades, acquired hypoparathyroidism secondary to surgery in the neck was more common than hereditary hypoparathyroidism, but the frequency of surgically induced parathyroid failure has diminished as a result of improved surgical techniques that spare the parathyroid glands and increased use of nonsurgical therapy for hyperthyroidism. PHP, an example of ineffective PTH action rather than a failure of parathyroid gland production, may share several features with hypoparathyroidism, including extraosseous calcification and extrapyramidal manifestations such as choreoathetotic movements and dystonia.

Papilledema and raised intracranial pressure may occur in both hereditary and acquired hypoparathyroidism, as do chronic changes in fingernails and hair and lenticular cataracts, the latter usually reversible with treatment of hypocalcemia. Certain skin manifestations, including alopecia and candidiasis, are characteristic of hereditary hypoparathyroidism associated with autoimmune polyglandular failure (Chap. 351).

Hypocalcemia associated with hypomagnesemia is associated with both deficient PTH release and impaired responsiveness to the hormone. Patients with hypocalcemia secondary to hypomagnesemia have absent or low levels of circulating PTH, indicative of diminished hormone release despite a maximum physiologic stimulus by hypocalcemia. Plasma PTH levels return to normal with correction of the hypomagnesemia. Thus hypoparathyroidism with low levels of PTH in blood can be due to hereditary gland failure, acquired gland failure, or acute but reversible gland dysfunction (hypomagnesemia).

GENETIC ABNORMALITIES AND HEREDITARY HYPOPARATHYROIDISM

Hereditary hypoparathyroidism can occur as an isolated entity without other endocrine or dermatologic manifestations. More typically, it occurs in association with other abnormalities such as defective development of the thymus or failure of other endocrine organs such as the adrenal, thyroid, or ovary (Chap. 351). Hereditary hypoparathyroidism is often manifest within the first decade but may appear later.

Genetic defects associated with hypoparathyroidism serve to illuminate the complexity of organ development, hormonal biosynthesis and secretion, and tissue-specific patterns of endocrine effector function **(Fig. 353-5)**. Often, hypoparathyroidism is isolated, signifying a highly specific functional disturbance. When hypoparathyroidism is associated with other developmental or organ defects, treatment of the hypocalcemia can still be effective.

A rare form of hypoparathyroidism associated with defective development of both the thymus and the parathyroid glands is termed the *DiGeorge syndrome*, or the *velocardiofacial syndrome*. Congenital cardiovascular, facial, and other developmental defects are present, and patients may die in early childhood with severe infections, hypocalcemia and

seizures, or cardiovascular complications. Patients can survive into adulthood, and milder, incomplete forms occur. Most cases are sporadic, but an autosomal dominant form involving microdeletions of chromosome 22q11.2 has been described. Smaller deletions in chromosome 22 are seen in incomplete forms of the DiGeorge syndrome, appearing in childhood or adolescence, that are manifest primarily by parathyroid gland failure. The chromosome 22 defect is now termed *DSG1*; more recently, a defect in chromosome 10p is also recognized—now called *DSG2*. The phenotypes seem similar. Studies on the chromosome 22 defect have pinpointed a transcription factor, TBX1. Deletions of the orthologous mouse gene show a phenotype similar to the human syndrome.

Another autosomal dominant developmental defect, featuring hypoparathyroidism, deafness, and renal dysplasia (HDR) has been studied at a genetic level. Cytogenic abnormalities in some, but not all kindred, point to translocation defects on chromosome 10, as in DiGeorge syndrome. However, the lack of immunodeficiency and heart defects distinguishes the two syndromes. Mouse models, as well as deletional analysis in some HDR patients, has pointed to transcription factor GATA3, which is important in embryonic development and is expressed in developing kidney, ear structures, and the parathyroids.

Another pair of linked developmental disorders involving the parathyroids is recognized. *Kenney-Caffey syndrome* features hypoparathyroidism, short stature, osteosclerosis, and thick cortical bones. A defect seen in Middle Eastern patients, particularly in Saudi Arabia, termed *Sanjad-Sakati syndrome*, also exhibits growth failure and other dysmorphic features. This syndrome, which is clearly autosomal recessive, involves a gene on chromosome 1q42-q43. Both syndromes apparently involve a chaperone protein, called *TBCE*, relevant to tubulin function.

Hypoparathyroidism can occur in association with a complex hereditary autoimmune syndrome involving failure of the adrenals, the ovaries, the immune system, and the parathyroids in association with recurrent mucocutaneous candidiasis, alopecia, vitiligo, and pernicious anemia (Chap. 351). The responsible gene on chromosome 21q22.3 has been identified. The protein product, which resembles a transcription factor, has been termed the *autoimmune regulator*, or AIRE. A stop codon mutation occurs in many Finnish families with the disorder, commonly referred to as *polyglandular autoimmune type 1 deficiency*.

Hypoparathyroidism is seen in two disorders associated with mitochondrial dysfunction and myopathy, one termed the *Kearns-Sayre syndrome* (KSS), with ophthalmoplegia and pigmentary retinopathy, and the other termed the *MELAS syndrome*, *m*itochondrial *e*ncephalopathy, *l*actic *a*cidosis, and *s*troke-like episodes. Mutations or deletions in mitochondrial genes have been identified.

Several forms of hypoparathyroidism, each rare in frequency, are seen as isolated defects; the genetic mechanisms are varied. The inheritance includes autosomal dominant, autosomal recessive, and X-linked modes. Three separate autosomal defects involving the parathyroid gene have been recognized: one is dominant and the other two are recessive. The dominant form has a point mutation in the signal sequence, a critical region involved in intracellular transport of the hormone precursor. An Arg for Cys mutation interferes with processing of the precursor and is believed to trigger an apoptotic cellular response, hence acting as a dominant negative. The other two forms are recessive. One point mutation also blocks cleavage of the PTH precursor but requires both alleles to cause hypoparathyroidism. The third involves a single-nucleotide base change that results in an exon splicing defect; the lost exon contains the promoter—hence, the gene is silenced. An X-linked recessive form of hypoparathyroidism has been described in males and the defect has been localized to chromosome Xq26-q27, perhaps involving the *SOX3* gene.

Abnormalities in the calcium-sensing receptor (CaSR) are detected in three distinctive hypocalcemic disorders. All are rare but more than 10 different gain-of-function mutations have been found in one form of hypocalcemia termed *autosomal dominant hypocalcemic hypercalciuria (ADHH)*. The receptor senses the ambient calcium level as excessive and suppresses PTH secretion, leading to hypocalcemia. The hypocalcemia is aggravated by constitutive receptor activity in the renal tubule causing excretion of inappropriate amounts of calcium. Recognition of the syndrome is important because efforts to treat the hypocalcemia with vitamin D analogues and increased oral calcium exacerbate the already excessive urinary calcium excretion (several grams or more per 24 hours), leading to irreversible renal damage from stones and ectopic calcification.

Other causes of isolated hypoparathyroidism include homozygous, inactivating mutations in the parathyroid-specific transcription factor GCMB, which lead to an autosomal recessive form of the disease, or heterozygous point mutations in GCMB, which have a dominant negative effect on the wild-type protein and thus lead to an autosomal dominant form of hypoparathyroidism.

The *Bartter syndrome* is a group of disorders associated with disturbances in electrolyte and acid/base balance, sometimes with nephrocalcinosis and other features. Several types of ion channels or transporters are involved. Curiously, *Bartter syndrome type V* has the electrolyte and pH disturbances seen in the other syndromes but appears to be due to a gain of function in the CaSR. The defect may be more severe than in ADHH and explains the additional features seen beyond hypocalcemia and hypercalciuria.

As with autoimmune disorders that block the CaSR (discussed above under hypercalcemic conditions), there are autoantibodies that at least transiently activate the CaSR, leading to suppressed PTH secretion and hypocalcemia. This disorder, which may wax and wane, could be classified as an acquired form of hypoparathyroidism but is listed here with other disorders involving the CaSR.

ACQUIRED HYPOPARATHYROIDISM

Acquired chronic hypoparathyroidism is usually the result of inadvertent surgical removal of all the parathyroid glands; in some instances, not all the tissue is removed, but the remainder undergoes vascular supply compromise secondary to fibrotic changes in the neck after surgery. In the past, the most frequent cause of acquired hypoparathyroidism was surgery for hyperthyroidism. Hypoparathyroidism now usually occurs after surgery for hyperparathyroidism when the surgeon, facing the dilemma of removing too little tissue and thus not curing the hyperparathyroidism, removes too much. Parathyroid function may not be totally absent in all patients with postoperative hypoparathyroidism.

Even rarer causes of acquired chronic hypoparathyroidism include radiation-induced damage subsequent to radioiodine therapy of hyperthyroidism and glandular damage in patients with hemochromatosis or hemosiderosis after repeated blood transfusions. Infection may involve one or more of the parathyroids but usually does not cause hypoparathyroidism because all four glands are rarely involved.

Transient hypoparathyroidism is frequent following surgery for hyperparathyroidism. After a variable period of hypoparathyroidism, normal parathyroid function may return due to hyperplasia or recovery of remaining tissue. Occasionally, recovery occurs months after surgery.

Treatment: Acquired and Hereditary Hypoparathyroidism

Treatment involves replacement with vitamin D or 1,25(OH)$_2$D3 (calcitriol) combined with a high oral calcium intake. In most patients, blood calcium and phosphate levels are satisfactorily regulated, but some patients show resistance and a brittleness, with a tendency to alternate between hypocalcemia and hypercalcemia. For many patients, vitamin D in doses of 40,000–120,000 U/d (1–3 mg/d) combined with ≥1 g elemental calcium is satisfactory. The wide dosage range reflects the variation encountered from patient to patient; precise regulation of each patient is required. Compared to typical daily requirements in euparathyroid patients of 200 U/d (or in older patients as high as 800 U/d), the high dose of vitamin D (as much as 100-fold higher) reflects the reduced conversion of vitamin D to 1,25(OH)$_2$D. Many physicians now use 0.5–1 μg of calcitriol in management of such patients, especially if they are difficult to control. Because of its storage in fat, when vitamin D is withdrawn, weeks are required for the disappearance of the biologic effects, compared with a few days for calcitriol, which has a rapid turnover.

Oral calcium and vitamin D restore the overall calcium-phosphate balance but do not reverse the lowered urinary calcium reabsorption typical of hypoparathyroidism. Therefore, care must be taken to avoid excessive urinary calcium excretion after vitamin D and calcium replacement therapy; otherwise, kidney stones can develop. Thiazide diuretics lower urine calcium by as much as 100 mg/d in hypoparathyroid patients on vitamin D, provided they are maintained on a low-sodium diet. Use of thiazides seems to be of benefit in mitigating hypercalciuria and easing the daily management of these patients.

There are now trials of parenterally administered PTH [either PTH(1–34) or PTH(1–84)] in patients with hypoparathyroidism providing greater ease of maintaining serum calcium and reducing urinary calcium excretion (desirable to protect any renal damage). However, PTH therapy is not approved as of yet.

HYPOMAGNESEMIA

Severe hypomagnesemia (<0.4 mmol/L; <0.8 meq/L) is associated with hypocalcemia (Chap. 352). Restoration of the total-body magnesium deficit leads to rapid reversal of hypocalcemia. There are at least two causes of the hypocalcemia —impaired PTH secretion and reduced responsiveness to PTH. **For further discussion of causes and treatment of hypomagnesemia, see**Chap. 352.

The effects of magnesium on PTH secretion are similar to those of calcium; hypermagnesemia suppresses and hypomagnesemia stimulates PTH secretion. The effects of magnesium on PTH secretion are normally of little significance, however, because the calcium effects dominate. Greater change in magnesium than in calcium is needed to influence hormone secretion. Nonetheless, hypomagnesemia might be expected to increase hormone secretion. It is therefore surprising to find that severe hypomagnesemia is associated with blunted secretion of PTH. The explanation for the paradox is that severe, chronic hypomagnesemia leads to intracellular magnesium deficiency, which interferes with secretion and peripheral responses to PTH. The mechanism of the cellular abnormalities caused by hypomagnesemia is unknown, although effects on adenylate cyclase (for which magnesium is a cofactor) have been proposed.

PTH levels are undetectable or inappropriately low in severe hypomagnesemia despite the stimulus of severe hypocalcemia, and acute repletion of magnesium leads to a rapid increase in PTH level. Serum phosphate levels are often not elevated, in contrast to the situation with acquired or idiopathic hypoparathyroidism, probably because phosphate deficiency is often seen in hypomagnesmia (see Chap. 331).

Diminished peripheral responsiveness to PTH also occurs in some patients, as documented by subnormal response in urinary phosphorus and urinary cyclic AMP excretion after administration of exogenous PTH to patients who are hypocalcemic and hypomagnesemic. Both blunted PTH secretion and lack of renal response to administered PTH can occur in the same patient. When acute magnesium repletion is undertaken, the restoration of PTH levels to normal or supranormal may precede restoration of normal serum calcium by several days.

Treatment: Hypomagnesemia

Repletion of magnesium cures the condition. Repletion should be parenteral. Attention must be given to restoring the intracellular deficit, which may be considerable. After IV magnesium administration, serum magnesium may return transiently to the normal range, but unless replacement therapy is adequate, serum magnesium will again fall. If the cause of the hypomagnesemia is renal magnesium wasting, magnesium may have to be given long-term to prevent recurrence (Chap. 352).

PTH Ineffective

PTH is ineffective when the PTH/PTHrP receptor–guanyl nucleotide–binding protein complex is defective (PHP, discussed below), when PTH action to promote calcium absorption from the diet is impaired because of vitamin D deficiency or because vitamin D is ineffective (defects in vitamin D receptor or vitamin D synthesis), or in chronic renal failure in which the calcium-elevating action of PTH is impaired.

Typically, hypophosphatemia is more severe than hypocalcemia in vitamin D deficiency states because of the increased secretion of PTH, which, although only partly effective in elevating blood calcium, is capable of promoting phosphaturia.

PHP, on the other hand, has a pathophysiology different from the other disorders of ineffective PTH action. PHP resembles hypoparathyroidism (in which PTH synthesis is deficient) and is manifested by hypocalcemia and hyperphosphatemia, yet elevated PTH levels. The cause of the disorder is defective PTH-dependent activation of guanyl nucleotide–binding proteins, resulting in failure of PTH to increase intracellular cyclic AMP (see below).

CHRONIC RENAL FAILURE

Improved medical management of chronic kidney disease now allows many patients to survive for decades and hence time enough to develop features of renal osteodystrophy, which must be controlled to avoid additional morbidity. Impaired production of 1,25(OH)$_2$D is now thought to be the principal factor that causes calcium deficiency, secondary hyperparathyroidism, and bone disease; hyperphosphatemia typically occurs only in the later stages of CKD. Low levels of 1,25(OH)$_2$D due to increased FGF23 production in bone are critical in the development of hypocalcemia. The uremic state also causes impairment of intestinal absorption by mechanisms other than defects in vitamin D metabolism. Nonetheless, treatment with supraphysiologic amounts of vitamin D or calcitriol corrects the impaired calcium absorption. Since increased FGF23 levels are seen even in early stages of renal failure in some patients, and have been reported to correlate with increased mortality, there is current interest in methods (lowering phosphate absorption) to lower FGF23 levels and concern as to whether vitamin D supplementation (known physiologically to increase FGF23) increases FGF23 in CKD.

Hyperphosphatemia in renal failure lowers blood calcium levels by several mechanisms, including extraosseous deposition of calcium and phosphate, impairment of the bone-resorbing action of PTH, and reduction in 1,25(OH)$_2$D production by remaining renal tissue.

Treatment: Chronic Renal Failure

Therapy of chronic renal failure (Chap. 280) involves appropriate management of patients prior to dialysis and adjustment of regimens once dialysis is initiated. Attention should be paid to restriction of phosphate in the diet; avoidance of aluminum-containing phosphate-binding antacids to prevent the problem of aluminum intoxication; provision of an adequate calcium intake by mouth, usually 1–2 g/d; and supplementation with 0.25–1 μg/d calcitriol. Each patient must be monitored closely. The aim of therapy is to restore normal calcium balance to prevent osteomalacia and severe secondary hyperparathyroidism (it is usually recommended to maintain PTH levels between 100 and 300 pg/mL) and, in light of evidence of genetic changes and monoclonal outgrowths of parathyroid glands in CKD patients, to prevent secondary from becoming autonomous hyperparathyroidism. Reduction of hyperphosphatemia and restoration of normal intestinal calcium absorption by calcitriol can improve blood calcium levels and reduce the manifestations of secondary hyperparathyroidism. Since adynamic bone disease can occur in association with low PTH levels, it is important to avoid excessive suppression of the parathyroid glands while recognizing the beneficial effects of controlling the secondary hyperparathyroidism. These patients should probably be closely monitored with PTH assays that detect only the full-length PTH(1–84) to ensure that biologically active PTH and not inactive, inhibitory PTH fragments are measured. Use of phosphate-binding agents such as sevelamer are approved for payment only in end-stage renal disease (ESRD).

VITAMIN D DEFICIENCY DUE TO INADEQUATE DIET AND/OR SUNLIGHT

Vitamin D deficiency due to inadequate intake of dairy products enriched with vitamin D, lack of vitamin supplementation, and reduced sunlight exposure in the elderly, particularly during winter in northern latitudes, is more common in the United States than previously recognized. Biopsies of bone in elderly patients with hip fracture (documenting osteomalacia) and abnormal levels of vitamin D metabolites, PTH, calcium, and phosphate indicate that vitamin D deficiency may occur in as many as 25% of elderly patients, particularly in northern latitudes in the United States. Concentrations of 25(OH)D are low or low-normal in these patients. Quantitative histomorphometry of bone biopsy specimens reveals widened osteoid seams consistent with osteomalacia (Chap. 352). PTH hypersecretion compensates for the tendency for the blood calcium to fall but also induces renal phosphate wasting and results in osteomalacia.

Treatment involves adequate replacement with vitamin D and calcium until the deficiencies are corrected. Severe hypocalcemia rarely occurs in moderately severe vitamin D deficiency of the elderly, but vitamin D deficiency must be considered in the differential diagnosis of mild hypocalcemia.

Mild hypocalcemia, secondary hyperparathyroidism, severe hypophosphatemia, and a variety of nutritional deficiencies occur with gastrointestinal diseases. Hepatocellular dysfunction can lead to reduction in 25(OH)D levels, as in portal or biliary cirrhosis of the liver, and malabsorption of vitamin D and its metabolites, including $1,25(OH)_2D$, may occur in a variety of bowel diseases, hereditary or acquired. Hypocalcemia itself can lead to steatorrhea, due to deficient production of pancreatic enzymes and bile salts. Depending on the disorder, vitamin D or its metabolites can be given parenterally, guaranteeing adequate blood levels of active metabolites.

DEFECTIVE VITAMIN D METABOLISM

Anticonvulsant Therapy

Anticonvulsant therapy with any of several agents induces acquired vitamin D deficiency by increasing the conversion of vitamin D to inactive compounds and/or causing resistance to its action. The more marginal the vitamin D intake in the diet, the more likely that anticonvulsant therapy will lead to abnormal mineral and bone metabolism.

Vitamin D–Dependent Rickets Type I

Rickets can be due to *resistance to the action* of vitamin D as well as to vitamin D deficiency. Vitamin D–dependent rickets type I, previously termed *pseudo-vitamin D–resistant rickets*, differs from true vitamin D–resistant rickets (vitamin D –dependent rickets type II, see below) in that it is less severe and the biochemical and radiographic abnormalities can be reversed with appropriate doses of the vitamin's active metabolite, $1,25(OH)_2D3$. Physiologic amounts of calcitriol cure the disease (Chap. 352). This finding fits with the pathophysiology of the disorder, which is autosomal recessive, and is now known to be caused by mutations in the gene encoding 25(OH)D-1α-hydroxylase. Both alleles are inactivated in all patients, and compound heterozygotes, harboring distinct mutations, are common.

Clinical features include hypocalcemia, often with tetany or convulsions, hypophosphatemia, secondary hyperparathyroidism, and osteomalacia, often associated with skeletal deformities and increased alkaline phosphatase. Treatment involves physiologic replacement doses of $1,25(OH)_2D3$ (Chap. 352).

Vitamin D–Dependent Rickets Type II

Vitamin D–dependent rickets type II results from end-organ resistance to the active metabolite $1,25(OH)_2D3$. The clinical features resemble those of the type I disorder and include hypocalcemia, hypophosphatemia, secondary hyperparathyroidism, and rickets but also partial or total alopecia. Plasma levels of $1,25(OH)_2D$ are elevated, in keeping with the refractoriness of the end organs. This disorder is caused by mutations in the gene encoding the vitamin D receptor; treatment is difficult and requires regular, usually nocturnal calcium infusions (Chap. 352).

PSEUDOHYPOPARATHYROIDISM

PHP refers to a group of distinct inherited disorders. Patients are characterized by symptoms and signs of hypocalcemia in association with distinctive skeletal and developmental defects. The hypocalcemia is due to a deficient response to PTH, which is probably restricted to the proximal renal tubules. Hyperplasia of the parathyroids, a response to hormone-resistant hypocalcemia, causes elevation of PTH levels. Studies, both clinical and basic, have clarified some aspects of these disorders, including the variable clinical spectrum, the pathophysiology, the genetic defects, and their mode of inheritance.

A working classification of the various forms of PHP is given in **Table 353-6**. The classification scheme is based on the signs of ineffective PTH action (low calcium and high phosphate), urinary cyclic AMP response to exogenous PTH, the presence or absence of *Albright's hereditary osteodystrophy* (AHO), and assays to measure the concentration of the $G_s\alpha$ subunit of the adenylate cyclase enzyme. Using these criteria, there are four types: PHP types Ia and Ib; pseudopseudohypoparathyroidism (PPHP), and the related disorder POH, and PHP-II.

Table 353-6 Classification of Pseudohypoparathyroidism (Php) and Pseudopseudohypoparathyroidism (Pphp)

Type	Hypocalcemia, Hyperphosphatemia	Response of Urinary cAMP to PTH	Serum PTH	$Gs\alpha$ Subunit Deficiency	AHO	Resistance to Hormones in Addition to PTH
PHP-	Yes	↓	↑	Yes	Yes	Yes

Ia						
PHP-Ib	Yes	↓	↑	No	No	Yes (in some patients)
PHP-II	Yes	Normal	↑	No	No	No
PPHP	No	Normal	Normal	Yes	Yes	No

Abbreviations: ↓, decreased; ↑, increased; AHO, Albright's hereditary osteodystrophy; PTH, parathyroid hormone.

Php-Ia and Php-Ib

Individuals with PHP-I, the most common of the disorders, show a deficient urinary cyclic AMP response to administration of exogenous PTH. Patients with PHP-I are divided into type Ia and type Ib. Patients with PHP-Ia show evidence for AHO and reduced amounts of $G_S\alpha$ protein/activity in readily accessible tissues, such as erythrocytes, lymphocytes, and fibroblasts. Patients with PHP-Ib typically lack evidence for AHO and they have normal $G_S\alpha$ activity. PHP-Ic, sometimes listed as a third form of PHP, is really a variant of PHP-Ia, since the mutant $G_S\alpha$ shows normal activity in certain in vitro assays.

Most patients who have PHP-Ia reveal characteristic features of AHO, consisting of short stature, round face, skeletal anomalies (brachydactyly), and/or heterotopic calcification. Patients have low calcium and high phosphate levels, as with true hypoparathyroidism. PTH levels, however, are elevated, reflecting resistance to hormone action.

Amorphous deposits of calcium and phosphate are found in the basal ganglia in about one-half of patients. The defects in metacarpal and metatarsal bones are sometimes accompanied by short phalanges as well, possibly reflecting premature closing of the epiphyses. The typical findings are short fourth and fifth metacarpals and metatarsals. The defects are usually bilateral. Exostoses and radius curvus are frequent. Impairments in olfaction and taste and unusual dermatoglyphic abnormalities have been reported.

INHERITANCE AND GENETIC PATTERNS

Multiple defects at the *GNAS* locus have now been identified in PHP-Ia, PHP-Ib, and PPHP patients. This gene, which is located on chromosome 20q13.3, encodes the α-subunit of the stimulatory G-protein ($G_S\alpha$), among other products (see below). Mutations include abnormalities in splice junctions associated with deficient mRNA production, point mutations, insertions, and/or deletion that all result in a protein with defective function, i.e. a 50% reduction in $G_S\alpha$ activity in erythrocytes or other cells.

Detailed analyses of disease transmission in affected kindreds have clarified many features of PHP-Ia, PPHP, and PHP-Ib **(Fig. 353-7)**. The former two entities, traced through multiple kindreds, have an inheritance pattern consistent with gene imprinting—only females, not males, can transmit the full disease with hypocalcemia—and PHP-Ia and PPHP do not coexist in the same generation. The phenomenon of gene imprinting, involving methylation of gene loci, independent of any mutation, involves selective inactivation of either the maternal or the paternal allele (Chap. 61). In the case of the $G_S\alpha$ transcript, it is paternally imprinted (silenced) in the renal cortex (where the disease manifestation is expressed), so that the disease PHP-Ia can never be inherited from the father carrying the defective allele but only from a mother whose allelic product is critical for the PTH-dependent function in the proximal tubules of the kidney. In the renal cortex, it is postulated that only the maternal allele is normally active (independent of any mutation), such that lack of activity from a defective paternal allele is not of consequence. This explains the occurrence in PHP-Ia of hypocalcemia, hyperphosphatemia, and resistance to PTH and often to other hormones that mediate their actions through a G protein–coupled receptor in tissues where imprinting also occurs. Strong additional evidence favoring this overall hypothesis comes from gene knockout studies in the mouse (ablating exon 2 of the gene responsible for $G_S\alpha$ synthesis). Mice inheriting the mutant allele from the female had undetectable $G_S\alpha$ protein in renal cortex and were hypocalcemic and resistant to renal actions of PTH. Offspring inheriting the mutant allele from the male showed no evidence of PTH resistance or hypocalcemia.

Figure 353-7

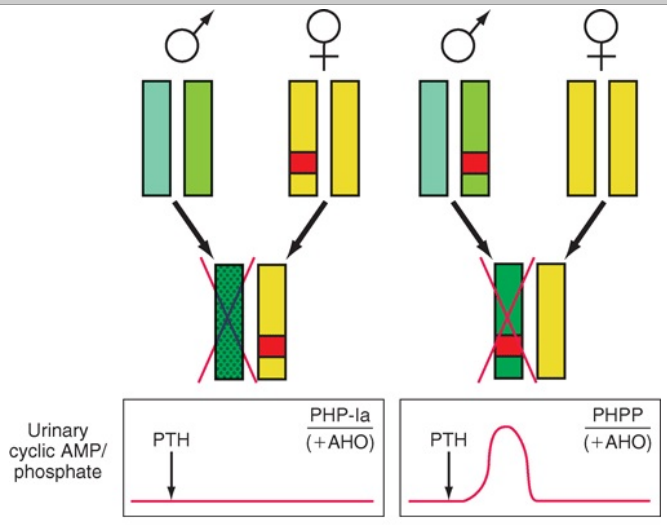

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Paternal imprinting of renal parathyroid hormone (PTH) resistance (*GNAS-1* gene for $G_S\alpha$ subunit) in pseudohypoparathyroidism (PHPIa). An impaired excretion of urinary cyclic AMP and phosphate is observed in patients with PHP. In the renal cortex, there is selective silencing of the paternal $G_S\alpha$ gene mRNA. The disease becomes manifest only in patients who inherit the defective gene from an obligate female carrier (*left*). If the genetic defect is inherited from an obligate male gene carrier, there is no biochemical abnormality; administration of PTH causes an appropriate increase

an obligate male gene carrier, there is no biochemical abnormality. Administration of PTH causes an appropriate increase in the urinary cyclic AMP and phosphate concentration [pseudo-PHP (PPHP); *right*]. Both patterns of inheritance lead to Albright's hereditary osteodystrophy (AHO), perhaps because of haplotype insufficiency—i.e., both copies of $G_s\alpha$ must be active in the fetus for normal bone development.

Imprinting is tissue selective. Paternal $G_s\alpha$ expression is not silenced in most tissues. It seems likely, therefore, that the AHO phenotype recognized in PPHP as well as PHP-Ia reflects $G_s\alpha$ haploinsufficiency during embryonic or postnatal development.

The complex mechanisms that control the *GNAS* gene contribute to challenges involved in unraveling the pathogenesis of these disorders, especially that of PHP-Ib. Much intensive work with families in which multiple members are affected by PHP-Ib, as well as studies of the complex regulation of the *GNAS* gene locus, have now shown that PHP-Ib is caused by microdeletions within or up-stream of the *GNAS* locus, which are associated with a loss of DNA methylation at one or several loci of the maternal allele (Table 353-6). These abnormalities in methylation silence the expression of the gene. This leads in the renal cortex—where $G_s\alpha$ appears to be expressed exclusively from the maternal allele—to PTH resistance. In other tissues, $G_s\alpha$ expression is not imprinted; hence erythrocytes show normal levels of $G_s\alpha$.

PHP-Ib, lacking the AHO phenotype in most instances, shares with PHP-Ia the hypocalcemia and hyperphosphatemia caused by PTH resistance, and thus the blunted urinary cyclic AMP response to administered PTH, a standard test to assess the presence or absence of hormone resistance (Table 353-6). Furthermore, these endocrine abnormalities become apparent only if the disease-causing mutation is inherited maternally. Bone responsiveness may be excessive rather than blunted in PHP-Ib (and in PHP-Ia) patients, based on case reports that have emphasized an osteitis fibrosa–like pattern in several PHP-Ib patients.

PHP-II refers to patients with hypocalcemia and hyperphosphatemia who have a normal urinary cyclic AMP, but an impaired urinary phosphaturic response to PTH. These patients are assumed to have a defect in the response to PTH at a locus distal to cyclic AMP production. It remains unclear why the PTH resistance in some patients, labeled as PHP-II, can be treated with vitamin D supplements.

The diagnosis of these hormone-resistant states can usually be made without difficulty when there is a positive family history for features of AHO, in association with the signs and symptoms of hypocalcemia. In both categories—PHP-Ia and PHP-Ib—serum PTH levels are elevated, particularly when patients are hypocalcemic. However, patients with PHP-Ib or PHP-II usually do not have phenotypic abnormalities, only hypocalcemia with high PTH levels, as evidence for hormone resistance. In PHP-Ib, the response of urinary cyclic AMP to the administration of exogenous PTH is blunted. The diagnosis of PHP-II is more complex, in that cyclic AMP responses in urine are, by definition, normal. Vitamin D deficiency must be excluded before the diagnosis of PHP-II can be entertained.

Treatment: Pseudohypoparathyroidism

Treatment of PHP is similar to that of hypoparathyroidism, except that calcium and vitamin D doses are usually lower. Patients with PHP show no PTH resistance in the distal tubules—hence, urinary calcium clearance is not affected and they are at less risk of developing nephrocalcinosis than in patients with true hypoparathyroidism. Variability in response makes it necessary to establish the optimal regimen for each patient, based on maintaining the appropriate blood calcium level and urinary calcium excretion.

PTH Overwhelmed

Occasionally, loss of calcium from the ECF is so severe that PTH cannot compensate. Such situations include acute pancreatitis and severe, acute hyperphosphatemia, often in association with renal failure, conditions in which there is rapid efflux of calcium from the ECF. Severe hypocalcemia can occur quickly; PTH rises in response to hypocalcemia but does not return blood calcium to normal.

SEVERE, ACUTE HYPERPHOSPHATEMIA

Severe hyperphosphatemia is associated with extensive tissue damage or cell destruction (Chap. 352). The combination of increased release of phosphate from muscle and impaired ability to excrete phosphorus because of renal failure causes moderate to severe hyperphosphatemia, the latter causing calcium loss from the blood and mild to moderate hypocalcemia. Hypocalcemia is usually reversed with tissue repair and restoration of renal function as phosphorus and creatinine values return to normal. There may even be a mild hypercalcemic period in the oliguric phase of renal function recovery. This sequence, severe hypocalcemia followed by mild hypercalcemia, reflects widespread deposition of calcium in muscle and subsequent redistribution of some of the calcium to the ECF after phosphate levels return to normal.

Other causes of hyperphosphatemia include hypothermia, massive hepatic failure, and hematologicmalignancies, either because of high cell turnover of malignancy or because of cell destruction by chemotherapy.

Treatment: Severe, Acute Hyperphosphatemia

Treatment is directed toward lowering of blood phosphate by the administration of phosphate-binding antacids or dialysis, often needed for the management of renal failure. Although calcium replacement may be necessary if hypocalcemia is severe and symptomatic, calcium administration during the hyperphosphatemic period tends to increase extraosseous calcium deposition and aggravate tissue damage. The levels of $1,25(OH)_2D$ may be low during the hyperphosphatemic phase and return to normal during the oliguric phase of recovery.

OSTEITIS FIBROSA AFTER PARATHYROIDECTOMY

Severe hypocalcemia after parathyroid surgery is rare now that osteitis fibrosa cystica is an infrequent manifestation of hyperparathyroidism. When osteitis fibrosa cystica is severe, however, bone mineral deficits can be large. After parathyroidectomy, hypocalcemia can persist for days if calcium replacement is inadequate. Treatment may require parenteral administration of calcium; addition of calcitriol and oral calcium supplementation is sometimes needed for weeks to a month or two until bone defects are filled (which, of course, is of therapeutic benefit in the skeleton), making it possible to discontinue parenteral calcium and/or reduce the amount.

Differential Diagnosis of Hypocalcemia

Care must be taken to ensure that true hypocalcemia is present; in addition, acute transient hypocalcemia can be a manifestation of a variety of severe, acute illnesses, as discussed above. *Chronic hypocalcemia*, however, can usually be ascribed to a few disorders associated with absent or ineffective PTH. Important clinical criteria include the duration of the illness, signs or symptoms of associated disorders, and the presence of features that suggest a hereditary abnormality. A nutritional history can be helpful in recognizing a low intake of vitamin D and calcium in the elderly, and a history of excessive alcohol intake may suggest magnesium deficiency.

Hypoparathyroidism and PHP are typically lifelong illnesses, usually (but not always) appearing by adolescence; hence, a recent onset of hypocalcemia in an adult is more likely due to nutritional deficiencies, renal failure, or intestinal disorders that result in deficient or ineffective vitamin D. Neck surgery, even long past, however, can be associated with a delayed onset of postoperative hypoparathyroidism. A history of seizure disorder raises the issue of anticonvulsive medication. Developmental defects may point to the diagnosis of PHP. Rickets and a variety of neuromuscular syndromes and deformities may indicate ineffective vitamin D action, either due to defects in vitamin D metabolism or to vitamin D deficiency.

A pattern of *low calcium with high phosphorus* in the absence of renal failure or massive tissue destruction almost invariably means hypoparathyroidism or PHP. A *low calcium and low phosphorus* pattern points to absent or ineffective vitamin D, thereby impairing the action of PTH on calcium metabolism (but not phosphate clearance). The relative ineffectiveness of PTH in calcium homeostasis in vitamin D deficiency, anticonvulsant therapy, gastrointestinal disorders, and hereditary defects in vitamin D metabolism leads to secondary hyperparathyroidism as a compensation. The excess PTH on renal tubule phosphate transport accounts for renal phosphate wasting and hypophosphatemia.

Exceptions to these patterns may occur. Most forms of hypomagnesemia are due to long-standing nutritional deficiency as seen in chronic alcoholics. Despite the fact that the hypocalcemia is principally due to an acute absence of PTH, phosphate levels are usually low, rather than elevated, as in hypoparathyroidism. Chronic renal failure is often associated with hypocalcemia and hyperphosphatemia, despite secondary hyperparathyroidism.

Diagnosis is usually established by application of the PTH immunoassay, tests for vitamin D metabolites, and measurements of the urinary cyclic AMP response to exogenous PTH. In hereditary and acquired hypoparathyroidism and in severe hypomagnesemia, PTH is either undetectable or in the normal range **(Fig. 353-4)**. This finding in a hypocalcemic patient is supportive of hypoparathyroidism, as distinct from ineffective PTH action, in which even mild hypocalcemia is associated with elevated PTH levels. Hence a failure to detect elevated PTH levels establishes the diagnosis of hypoparathyroidism; elevated levels suggest the presence of secondary hyperparathyroidism, as found in many of the situations in which the hormone is ineffective due to associated abnormalities in vitamin D action. Assays for 25(OH)D can be helpful. Low or low-normal 25(OH)D indicates vitamin D deficiency due to lack of sunlight, inadequate vitamin D intake, or intestinal malabsorption. Recognition that mild hypocalcemia, rickets, and hypophosphatemia are due to anticonvulsant therapy is made by history.

Treatment: Hypocalcemic States

The management of hypoparathyroidism, PHP, chronic renal failure, and hereditary defects in vitamin D metabolism involves the use of vitamin D or vitamin D metabolites and calcium supplementation. Vitamin D itself is the least expensive form of vitamin D replacement and is frequently used in the management of uncomplicated hypoparathyroidism and some disorders associated with ineffective vitamin D action. When vitamin D is used prophylactically, as in the elderly or in those with chronic anticonvulsant therapy, there is a wider margin of safety than with the more potent metabolites. However, most of the conditions in which vitamin D is administered chronically for hypocalcemia require amounts 50–100 times the daily replacement dose because the formation of $1,25(OH)_2D$ is deficient. In such situations, vitamin D is no safer than the active metabolite because intoxication can occur with high-dose therapy (because of storage in fat). Calcitriol is more rapid in onset of action and also has a short biologic half-life.

Vitamin D [at least 1000 U/d (2–3 μg/d) (higher levels required in older persons)] or calcitriol (0.25–1 μg/d) is required to prevent rickets in normal individuals. In contrast, 40,000–120,000 U (1–3 mg) of vitamin D2 or D3 is typically required in hypoparathyroidism. The dose of calcitriol is unchanged in hypoparathyroidism, since the defect is in hydroxylation by the 25(OH)D-1α-hydroxylase. Calcitriol is also used in disorders of 25(OH)D-1α-hydroxylase; vitamin D receptor defects are much more difficult to treat.

Patients with hypoparathyroidism should be given 2–3 g elemental calcium PO each day. The two agents, vitamin D or calcitriol and oral calcium, can be varied independently. Urinary calcium excretion needs to be monitored carefully. If hypocalcemia alternates with episodes of hypercalcemia in high-brittleness patients with hypoparathyroidism, administration of calcitriol and use of thiazides, as discussed above, may make management easier.

FURTHER READINGS

Bastepe M, Jüppner H: Pseudohypoparathyroidism, Albright's hereditary osteodystrophy, and progressive osseous heteroplasia: Disorders caused by inactivating GNAS mutations, in *Endocrinology*, 6th ed, in Endocrinology, JL Jameson, LJ DeGroot (eds), Philadelphia, W.B. Saunders Company, 2011

Bilezikian JP et al: Guidelines for the management of asymptomatic primary hyperparathyroidism: Summary statement from the third international workshop. J Clin Endocrinol Metab 94:335, 2009[PMID: 19193908] [Full Text]

Thakker RV et al: Genetic disorders of calcium homeostasis caused by abnormal regulation of parathyroid hormone secretion or responsiveness, in *Endocrinology*, 6th ed, in Endocrinology, JL Jameson, LJ DeGroot (eds), Philadelphia, W.B. Saunders Company, 2011

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

OSTEOPOROSIS: INTRODUCTION

Osteoporosis, a condition characterized by decreased bone strength, is prevalent among postmenopausal women but also occurs in men and women with underlying conditions or major risk factors associated with bone demineralization. Its chief clinical manifestations are vertebral and hip fractures, although fractures can occur at any skeletal site. Osteoporosis affects >10 million individuals in the United States, but only a small proportion are diagnosed and treated.

DEFINITION

Osteoporosis is defined as a reduction in the strength of bone that leads to an increased risk of fractures. Loss of bone tissue is associated with deterioration in skeletal microarchitecture. The World Health Organization (WHO) operationally defines osteoporosis as a bone density that falls 2.5 standard deviations (SD) below the mean for young healthy adults of the same sex —also referred to as a *T-score* of –2.5. Postmenopausal women who fall at the lower end of the young normal range (a T-score <–1.0) are defined as having low bone density and are also at increased risk of osteoporosis. More than 50% of fractures among postmenopausal women, including hip fractures, occur in this group with low bone density.

EPIDEMIOLOGY

In the United States, as many as 8 million women and 2 million men have osteoporosis (T-score <–2.5), and an additional 18 million individuals have bone mass levels that put them at increased risk of developing osteoporosis (e.g., bone mass T-score < –1.0). Osteoporosis occurs more frequently with increasing age as bone tissue is lost progressively. In women, the loss of ovarian function at menopause (typically about age 50) precipitates rapid bone loss so that most women meet the diagnostic criterion for osteoporosis by age 70–80.

The epidemiology of fractures follows the trend for loss of bone density. Fractures of the distal radius increase in frequency before age 50 and plateau by age 60, with only a modest age-related increase thereafter. In contrast, incidence rates for hip fractures double every 5 years after age 70 **(Fig. 354-1)**. This distinct epidemiology may be related to the way people fall as they age, with fewer falls on an outstretched hand and more falls directly on the hip. At least 1.5 million fractures occur each year in the United States as a consequence of osteoporosis. As the population continues to age, the total number of fractures will continue to escalate.

Figure 354-1

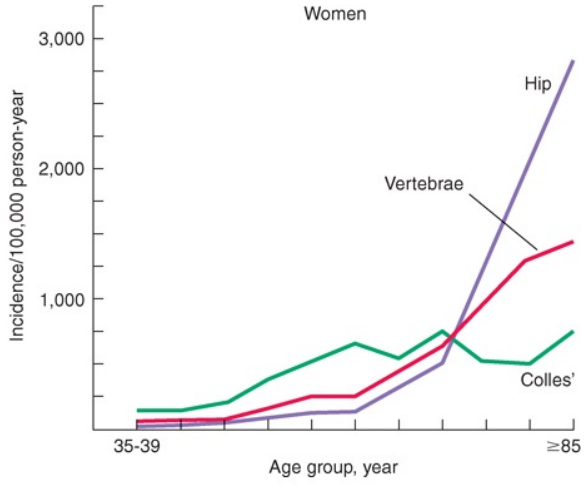

Epidemiology of vertebral, hip, and Colles' fractures with age. *(Adapted from C Cooper, LJ Melton III: Trends Endocrinol Metab 3:224, 1992; with permission.)*

About 300,000 hip fractures occur each year in the United States, most of which require hospital admission and surgical intervention. The probability that a 50-year-old white individual will have a hip fracture during his or her lifetime is 14% for women and 5% for men; the risk for blacks is lower (about one-half those rates). Hip fractures are associated with a high incidence of deep vein thrombosis and pulmonary embolism (20–50%) and a mortality rate between 5 and 20% during the year after surgery.

There are about 700,000 vertebral crush fractures per year in the United States. Only a fraction of them are recognized clinically, since many are relatively asymptomatic and are identified incidentally during radiography for other purposes **(Fig. 354-2)**. Vertebral fractures rarely require hospitalization but are associated with long-term morbidity and a slight increase in mortality rates, primarily related to pulmonary disease. Multiple vertebral fractures lead to height loss (often of several inches), kyphosis, and secondary pain and discomfort related to altered biomechanics of the back. Thoracic fractures can be associated with restrictive lung disease, whereas lumbar fractures are associated with abdominal symptoms that include distention, early satiety, and constipation.

Figure 354-2

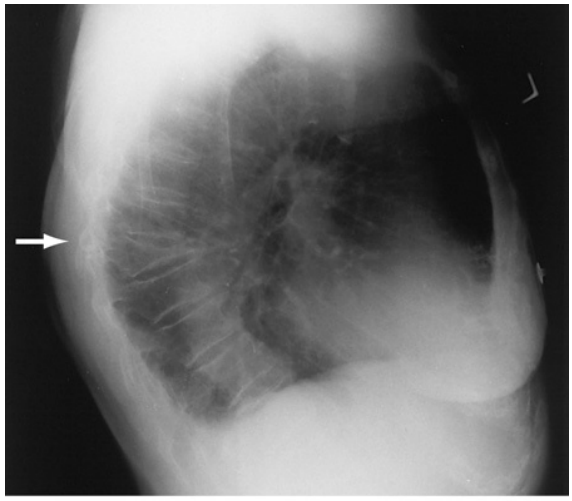

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lateral spine x-ray showing severe osteopenia and a severe wedge-type deformity (severe anterior compression).

Approximately 250,000 wrist fractures occur in the United States each year. Fractures of other bones (estimated to be ~300,000 per year) also occur with osteoporosis; this is not surprising in light of the fact that bone loss is a systemic phenomenon. Fractures of the pelvis and proximal humerus clearly are associated with osteoporosis. Although some fractures result from major trauma, the threshold for fracture is reduced for an osteoporotic bone **(Fig. 354-3)**. In addition to bone density, there are a number of risk factors for fracture; the common ones are summarized in **Table 354-1**. Age, prior fractures, a family history of osteoporosis-related fractures, low body weight, cigarette consumption, and excessive alcohol use are all independent predictors of fracture. Chronic diseases with inflammatory components that increase skeletal remodeling such as rheumatoid arthritis, increase the risk of osteoporosis, as do diseases associated with malabsorption. Chronic diseases that increase the risk of falling or frailty, including dementia, Parkinson's disease, and multiple sclerosis, also increase fracture risk.

Figure 354-3

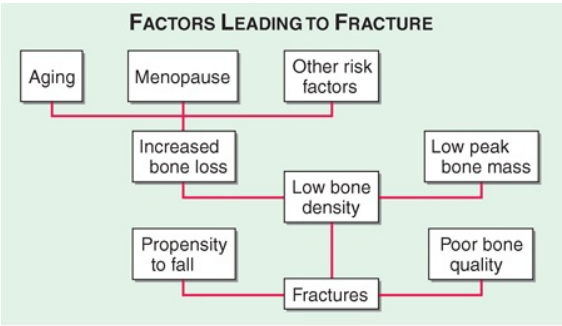

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Factors leading to osteoporotic fractures.

Table 354-1 Risk Factors for Osteoporosis Fracture

Nonmodifiable	Estrogen deficiency
Personal history of fracture as an adult	Early menopause (<45 years) or bilateral ovariectomy
History of fracture in first-degree relative	Prolonged premenstrual amenorrhea (>1 year)
Female sex	Low calcium intake
Advanced age	Alcoholism
White race	Impaired eyesight despite adequate correction
Dementia	
Potentially modifiable	Recurrent falls
Current cigarette smoking	Inadequate physical activity
Low body weight [<58 kg (127 lb)]	Poor health/frailty

In the United States and Europe, osteoporosis-related fractures are more common among women than men, presumably due to a lower peak bone mass as well as postmenopausal bone loss in women. However, this sex difference in bone density and age-related increase in hip fractures is not as apparent in some other cultures, possibly due to genetics, physical activity level, or diet.

Fractures are themselves risk factors for future fractures (Table 354-1). Vertebral fractures increase the risk of other vertebral fractures as well as fractures of the peripheral skeleton such as the hip and wrist. Wrist fractures also increase the risk of vertebral and hip fractures. Consequently, among individuals over age 50, any fracture should be considered as potentially related to osteoporosis regardless of the circumstances of the fracture. Osteoporotic bone is more likely to fracture than is

normal bone at any level of trauma, and a fracture in a person over 50 should trigger evaluation for osteoporosis.

PATHOPHYSIOLOGY

Bone Remodeling

Osteoporosis results from bone loss due to age-related changes in bone remodeling as well as extrinsic and intrinsic factors that exaggerate this process. These changes may be superimposed on a low peak bone mass. Consequently, understanding the bone remodeling process is fundamental to understanding the pathophysiology of osteoporosis (Chap. 352). During growth, the skeleton increases in size by linear growth and by apposition of new bone tissue on the outer surfaces of the cortex **(Fig. 354-4)**. The latter process is called *modeling*, a process that also allows the long bones to adapt in shape to the stresses placed on them. Increased sex hormone production at puberty is required for skeletal maturation, which reaches maximum mass and density in early adulthood. It is around puberty that the sexual dimorphism in skeletal size becomes obvious, although true bone density remains similar between the sexes. Nutrition and lifestyle also play an important role in growth, though genetic factors primarily determine peak skeletal mass and density. Numerous genes control skeletal growth, peak bone mass, and body size, as well as skeletal structure and density. Heritability estimates of 50–80% for bone density and size have been derived on the basis of twin studies. Though peak bone mass is often lower among individuals with a family history of osteoporosis, association studies of candidate genes [vitamin D receptor; type I collagen, the estrogen receptor (ER), and interleukin 6 (IL-6); and insulin-like growth factor I (IGF-I)] and bone mass, bone turnover, and fracture prevalence have been inconsistent. Linkage studies suggest that a genetic locus on chromosome 11 is associated with high bone mass. Families with high bone mass and without much apparent age-related bone loss have been shown to have a point mutation in LRP5, a low-density lipoprotein receptor–related protein. The role of this gene in the general population is not clear, although a nonfunctional mutation results in osteoporosis-pseudoglioma syndrome, and LRP5 signaling appears to be important in controlling bone formation.

Figure 354-4

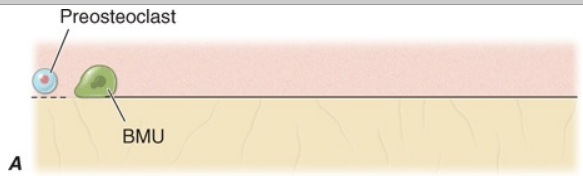

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

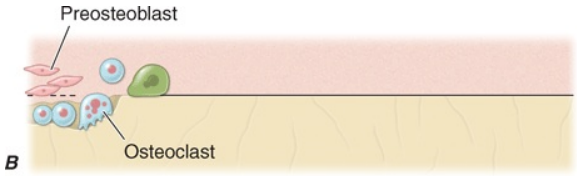

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

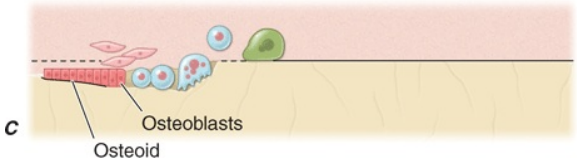

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

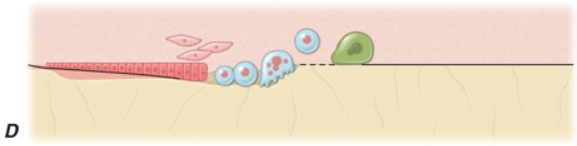

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

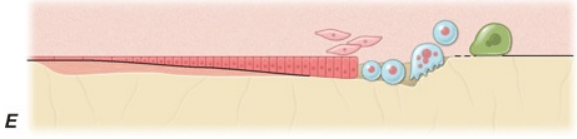

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mechanism of bone remodeling. The basic molecular unit (BMU) moves along the trabecular surface at a rate of about 10 μm/d. The figure depicts remodeling over ~120 days. ***A.*** Origination of BMU-lining cells contracts to expose collagen and attract preosteoclasts. ***B.*** Osteoclasts fuse into multinucleated cells that resorb a cavity. Mononuclear cells continue resorption, and preosteoblasts are stimulated to proliferate. ***C.*** Osteoblasts align at bottom of cavity and start forming osteoid (black). ***D.*** Osteoblasts continue formation and mineralization. Previous osteoid starts to mineralize (horizontal lines). ***E.*** Osteoblasts begin to flatten. ***F.*** Osteoblasts turn into lining cells; bone remodeling at initial surface (left of drawing) is now complete, but BMU is still advancing (to the right). *[Adapted from SM Ott, in JP Bilezikian et al (eds): Principles of Bone Biology, vol. 18. San Diego, Academic Press, 1996, pp 231–241.]*

In adults, bone remodeling, not modeling, is the principal metabolic skeletal process. Bone remodeling has two primary functions: (1) to repair microdamage within the skeleton to maintain skeletal strength and (2) to supply calcium from the skeleton to maintain serum calcium. Remodeling may be activated by microdamage to bone as a result of excessive or accumulated stress. Acute demands for calcium involve osteoclast-mediated resorption as well as calcium transport by osteocytes. Chronic demands for calcium result in secondary hyperparathyroidism, increased bone remodeling, and overall loss of bone tissue.

Bone remodeling also is regulated by several circulating hormones, including estrogens, androgens, vitamin D, and parathyroid hormone (PTH), as well as locally produced growth factors such as IGF-I and immunoreactive growth hormone II (IGH-II), transforming growth factor β (TGF-β), parathyroid hormone–related peptide (PTHrP), interleukins (ILs), prostaglandins, and members of the tumor necrosis factor (TNF) superfamily. These factors primarily modulate the rate at which new remodeling sites are activated, a process that results initially in bone resorption by osteoclasts, followed by a period of repair during which new bone tissue is synthesized by osteoblasts. The cytokine responsible for communication between the osteoblasts, other marrow cells, and osteoclasts has been identified as RANK ligand (RANKL) [receptor activator of nuclear factor-kappa-B (NFκB); RANKL]. RANKL, a member of the TNF family, is secreted by osteoblasts and certain cells of the immune system (Chap. 352). The osteoclast receptor for this protein is referred to as *RANK*. Activation of RANK by RANKL is a final common path in osteoclast development and activation. A humoral decoy for RANKL, also secreted by osteoblasts, is referred to as *osteoprotegerin***(Fig. 354-5)**. Modulation of osteoclast recruitment and activity appears to be related to the interplay among these three factors. Additional influences include nutrition (particularly calcium intake) and physical activity level.

Figure 354-5

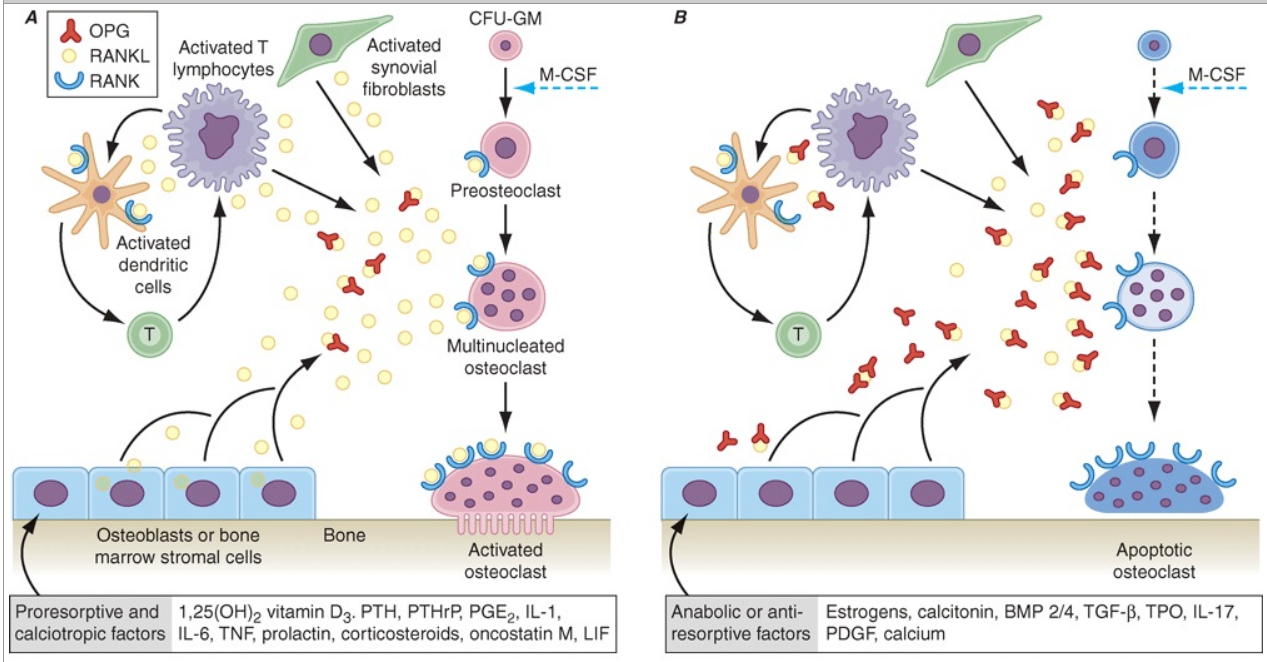

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hormonal control of bone resorption.*A.* Proresorptive and calciotropic factors. ***B.*** Anabolic and antiosteoclastic factors. RANKL expression is induced in osteoblasts, activated T cells, synovial fibroblasts, and bone marrow stromal cells. It binds to membrane-bound receptor RANK to promote osteoclast differentiation, activation, and survival. Conversely, osteoprotegerin (OPG) expression is induced by factors that block bone catabolism and promote anabolic effects. OPG binds and neutralizes RANKL, leading to a block in osteoclastogenesis and decreased survival of preexisting osteoclasts. CFU-GM, colony-forming units, granulocyte macrophage; M-CSF, macrophage colony-stimulating factor; RANKL, receptor activator of nuclear factor NFκB; PTH, parathyroid hormone; PGE2, prostaglandin E2 ; TNF, tumor necrosis factor; LIF, leukemia inhibitory factor; TPO, thrombospondin; PDGF, platelet-derived growth factor; OPG-L, osteoprotegerin-ligand; IL, interleukin; TGF-β, transforming growth factor β. *(From WJ Boyle et al: Nature 423: 337, 2003.)*

In young adults resorbed bone is replaced by an equal amount of new bone tissue. Thus, the mass of the skeleton remains constant after peak bone mass is achieved in adulthood. After age 30–45, however, the resorption and formation processes become imbalanced, and resorption exceeds formation. This imbalance may begin at different ages and varies at different skeletal sites; it becomes exaggerated in women after menopause. Excessive bone loss can be due to an increase in osteoclastic activity and/or a decrease in osteoblastic activity. In addition, an increase in remodeling activation frequency, and

thus the number of remodeling sites, can magnify the small imbalance seen at each remodeling unit. Increased recruitment of bone remodeling sites produces a reversible reduction in bone tissue but also can result in permanent loss of tissue and disrupted skeletal architecture. In trabecular bone, if the osteoclasts penetrate trabeculae, they leave no template for new bone formation to occur, and, consequently, rapid bone loss ensues and cancellous connectivity becomes impaired. A higher number of remodeling sites increases the likelihood of this event. In cortical bone, increased activation of remodeling creates more porous bone. The effect of this increased porosity on cortical bone strength may be modest if the overall diameter of the bone is not changed. However, decreased apposition of new bone on the periosteal surface coupled with increased endocortical resorption of bone decreases the biomechanical strength of long bones. Even a slight exaggeration in normal bone loss increases the risk of osteoporosis-related fractures because of the architectural changes that occur.

Calcium Nutrition

Peak bone mass may be impaired by inadequate calcium intake during growth among other nutritional factors (calories, protein, and other minerals), leading to increased risk of osteoporosis later in life. During the adult phase of life, insufficient calcium intake contributes to relative secondary hyperparathyroidism and an increase in the rate of bone remodeling to maintain normal serum calcium levels. PTH stimulates the hydroxylation of vitamin D in the kidney, leading to increased levels of 1,25-dihydroxyvitamin D [1,25$(OH)_2$D] and enhanced gastrointestinal calcium absorption. PTH also reduces renal calcium loss. Although these are all appropriate compensatory homeostatic responses for adjusting calcium economy, the long-term effects are detrimental to the skeleton because the increased remodeling rates and the ongoing imbalance between resorption and formation at remodeling sites combine to accelerate loss of bone tissue.

Total daily calcium intakes <400 mg are detrimental to the skeleton, and intakes in the range of 600–800 mg, which is about the average intake among adults in the United States, are also probably suboptimal. The recommended daily required intake of 1000–1200 mg for adults accommodates population heterogeneity in controlling calcium balance (Chap. 73).

Vitamin D

(See also Chap. 352) Severe vitamin D deficiency causes rickets in children and osteomalacia in adults. However, there is accumulating evidence that vitamin D insufficiency may be more prevalent than previously thought, particularly among individuals at increased risk such as the elderly; those living in northern latitudes; and individuals with poor nutrition, malabsorption, or chronic liver or renal disease. Dark-skinned individuals are also at high risk of vitamin D deficiency. An expert consensus panel has suggested that the accepted levels for serum 25-hydroxyvitamin D [25(OH)D] have been set too low and that optimal targets for serum 25(OH)D are >75 nmol/L (30 ng/mL). To achieve this level for most adults requires an intake of 800–1000 units/d, particularly in individuals who avoid sunlight or routinely use ultraviolet-blocking lotions. Vitamin D insufficiency leads to compensatory secondary hyperparathyroidism and is an important risk factor for osteoporosis and fractures. Some studies have shown that >50% of inpatients on a general medical service exhibit biochemical features of vitamin D deficiency, including increased levels of PTH and alkaline phosphatase and lower levels of ionized calcium. In women living in northern latitudes, vitamin D levels decline during the winter months. This is associated with seasonal bone loss, reflecting increased bone turnover. Even among healthy ambulatory individuals, mild vitamin D deficiency is increasing in prevalence. Treatment with vitamin D can return levels to normal [>75 μmol/L (30 ng/mL)] and prevent the associated increase in bone remodeling, bone loss, and fractures. Reduced fracture rates also have been documented among individuals in northern latitudes who have greater vitamin D intake and have higher 25(OH)D levels (see below). Vitamin D adequacy also may affect risk and/or severity of other diseases, including cancers (colorectal, prostate, and breast), autoimmune diseases, and diabetes.

Estrogen Status

Estrogen deficiency probably causes bone loss by two distinct but interrelated mechanisms: (1) activation of new bone remodeling sites and (2) exaggeration of the imbalance between bone formation and resorption. The change in activation frequency causes a transient bone loss until a new steady state between resorption and formation is achieved. The remodeling imbalance, however, results in a permanent decrement in mass. In addition, the very presence of more remodeling sites in the skeleton increases the probability that trabeculae will be penetrated, eliminating the template on which new bone can be formed and accelerating the loss of bony tissue.

The most common estrogen-deficient state is the cessation of ovarian function at the time of menopause, which occurs on average at age 51 (Chap. 348). Thus, with current life expectancy, an average woman will spend about 30 years without an ovarian supply of estrogen. The mechanism by which estrogen deficiency causes bone loss is summarized in Fig. 354-5. Marrow cells (macrophages, monocytes, osteoclast precursors, mast cells) as well as bone cells (osteoblasts, osteocytes, osteoclasts) express ERs α and β. Loss of estrogen increases production of RANKL and may reduce production of osteoprotegerin, increasing osteoclast recruitment. Estrogen also may play an important role in determining the life span of bone cells by controlling the rate of apoptosis. Thus, in situations of estrogen deprivation, the life span of osteoblasts may be decreased, whereas the longevity and activity of osteoclasts are increased.

Since remodeling is initiated at the surface of bone, it follows that trabecular bone—which has a considerably larger surface area (80% of the total) than cortical bone—will be affected preferentially by estrogen deficiency. Fractures occur earliest at sites where trabecular bone contributes most to bone strength; consequently, vertebral fractures are the most common early consequence of estrogen deficiency.

Physical Activity

Inactivity such as prolonged bed rest or paralysis, results in significant bone loss. Concordantly, athletes have higher bone mass than does the general population. These changes in skeletal mass are most marked when the stimulus begins during growth and before the age of puberty. Adults are less capable than children of increasing bone mass after restoration of physical activity. Epidemiologic data support the beneficial effects on the skeleton of chronic high levels of physical activity. Fracture risk is lower in rural communities and in countries where physical activity is maintained into old age. However, when exercise is initiated during adult life, the effects of moderate exercise on the skeleton are modest, with a bone mass increase of 1–2% in short-term studies of <2 years' duration. It is argued that more active individuals are less likely to fall and are more capable of protecting themselves upon falling, thereby reducing fracture risk.

Chronic Disease

Various genetic and acquired diseases are associated with an increase in the risk of osteoporosis **(Table 354-2)**. Mechanisms that contribute to bone loss are unique for each disease and typically result from multiple factors, including nutrition, reduced physical activity levels, and factors that affect rates of bone remodeling. In most, but not all, circumstances the primary diagnosis is made before osteoporosis presents clinically.

Table 354-2 Diseases Associated with an Increased Risk of Generalized Osteoporosis in Adults
Hypogonadal states

- Turner's syndrome
- Klinefelter's syndrome
- Anorexia nervosa
- Hypothalamic amenorrhea
- Hyperprolactinemia
- Other primary or secondary hypogonadal states

Endocrine disorders

- Cushing's syndrome
- Hyperparathyroidism
- Thyrotoxicosis
- Type 1 diabetes mellitus
- Acromegaly
- Adrenal insufficiency

Nutritional and gastrointestinal disorders

- Malnutrition
- Parenteral nutrition
- Malabsorption syndromes
- Gastrectomy
- Severe liver disease, especially biliary cirrhosis
- Pernicious anemia

Rheumatologic disorders

- Rheumatoid arthritis
- Ankylosing spondylitis

Hematologic disorders/malignancy

- Multiple myeloma
- Lymphoma and leukemia
- Malignancy-associated parathyroid hormone (PTHrP) production
- Mastocytosis
- Hemophilia
- Thalassemia

Selected inherited disorders

- Osteogenesis imperfecta
- Marfan's syndrome
- Hemochromatosis
- Hypophosphatasia
- Glycogen storage diseases
- Homocystinuria
- Ehlers-Danlos syndrome
- Porphyria
- Menkes' syndrome
- Epidermolysis bullosa

Other disorders

- Immobilization
- Chronic obstructive pulmonary disease
- Pregnancy and lactation
- Scoliosis
- Multiple sclerosis
- Sarcoidosis
- Amyloidosis

Medications

A large number of medications used in clinical practice have potentially detrimental effects on the skeleton **(Table 354-3)**. *Glucocorticoids* are the most common cause of medication-induced osteoporosis. It is often not possible to determine the extent to which osteoporosis is related to glucocorticoid or to other factors, as treatment is superimposed on the effects of the primary disease, which in itself may be associated with bone loss (e.g., rheumatoid arthritis). Excessive doses of thyroid hormone can accelerate bone remodeling and result in bone loss.

Table 354-3 Drugs Associated with an Increased Risk of Generalized Osteoporosis in Adults

Glucocorticoids	Excessive thyroxine
Cyclosporine	Aluminum
Cytotoxic drugs	Gonadotropin-releasing hormone agonists
Anticonvulsants	Heparin

Excessive alcohol	Lithium
Aromatase inhibitors	

Other medications have less detrimental effects on the skeleton than pharmacologic doses of glucocorticoids. *Anticonvulsants* are thought to increase the risk of osteoporosis, although many affected individuals have concomitant insufficiency of 1,25(OH)2D, as some anticonvulsants induce the cytochrome P450 system and vitamin D metabolism. Patients undergoing transplantation are at high risk for rapid bone loss and fracture not only from glucocorticoids but also from treatment with other *immunosuppressants* such as cyclosporine and tacrolimus (FK506). In addition these patients often have underlying metabolic abnormalities such as hepatic or renal failure, that predispose to bone loss.

Aromatase inhibitors, which potently block the aromatase enzyme that converts androgens and other adrenal precursors to estrogen, reduce circulating postmenopausal estrogen levels dramatically. These agents, which are used in various stages for breast cancer treatment, also have been shown to have a detrimental effect on bone density and risk of fracture.

Cigarette Consumption

The use of cigarettes over a long period has detrimental effects on bone mass. These effects may be mediated directly by toxic effects on osteoblasts or indirectly by modifying estrogen metabolism. On average, cigarette smokers reach menopause 1–2 years earlier than the general population. Cigarette smoking also produces secondary effects that can modulate skeletal status, including intercurrent respiratory and other illnesses, frailty, decreased exercise, poor nutrition, and the need for additional medications (e.g., glucocorticoids for lung disease).

MEASUREMENT OF BONE MASS

Several noninvasive techniques are available for estimating skeletal mass or density. They include dual-energy x-ray absorptiometry (DXA), single-energy x-ray absorptiometry (SXA), quantitative CT, and ultrasound (US).

DXA is a highly accurate x-ray technique that has become the standard for measuring bone density in most centers. Though it can be used for measurement in any skeletal site, clinical determinations usually are made of the lumbar spine and hip. Portable DXA machines have been developed that measure the heel (calcaneus), forearm (radius and ulna), or finger (phalanges). DXA also can be used to measure body composition. In the DXA technique, two x-ray energies are used to estimate the area of mineralized tissue, and the mineral content is divided by the area, which partially corrects for body size. However, this correction is only partial since DXA is a two-dimensional scanning technique and cannot estimate the depth or posteroanterior length of the bone. Thus, small people tend to have lower than average bone mineral density (BMD). Bone spurs, which are common in osteoarthritis, tend to falsely increase bone density of the spine and are a particular problem in measuring the spine in older individuals. Because DXA instrumentation is provided by several different manufacturers, the output varies in absolute terms. Consequently, it has become standard practice to relate the results to "normal" values by using T-scores, which compare individual results to those in a young population that is matched for race and sex. Z-scores compare individual results to those of an age-matched population that also is matched for race and sex. Thus, a 60-year-old woman with a Z-score of –1 (1 SD below mean for age) has a T-score of –2.5 (2.5 SD below mean for a young control group) **(Fig. 354-6)**. A T-score below –2.5 in the lumbar spine, femoral neck, or total hip is taken as a diagnosis of osteoporosis.

Figure 354-6

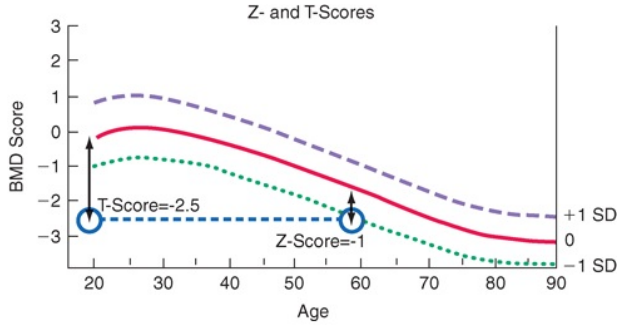

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship between Z-scores and T-scores in a 60-year-old woman. BMD, bone mineral density; SD, standard deviation.

CT is used primarily to measure the spine and, more recently, the hip. Peripheral CT is used to measure bone in the forearm or tibia. The results obtained from CT are different from all others currently available since this technique is three-dimensional and can provide a true density (mass of bone tissue per unit volume). CT also can specifically analyze trabecular bone and cortical bone content and volume separately. However, CT remains expensive, involves greater radiation exposure, and is less reproducible than DXA. A new technique employing high-resolution CT scanning called *Xtreme CT* also can provide information on skeletal architecture, including cancellous connectivity.

US is used to measure bone mass by calculating the attenuation of the signal as it passes through bone or the speed with which it traverses the bone. It is unclear whether US assesses properties of bone other than mass (e.g., quality), but this is a potential advantage of the technique. Because of its relatively low cost and mobility, US is amenable for use as a screening procedure.

All these techniques for measuring BMD have been approved by the U.S. Food and Drug Administration (FDA) on the basis of their capacity to predict fracture risk. The hip is the preferred site of measurement in most individuals, since it predicts the risk of hip fracture, the most important consequence of osteoporosis, better than any other bone density measurement site. When hip measurements are performed by DXA, the spine can be measured at the same time. In younger individuals such as perimenopausal or early postmenopausal women, spine measurements may be the most sensitive indicator of bone loss.

When to Measure Bone Mass

Clinical guidelines have been developed for the use of bone densitometry in clinical practice. The original National Osteoporosis Foundation guidelines recommend bone mass measurements in postmenopausal women, assuming they have one or more risk factors for osteoporosis in addition to age, sex, and estrogen deficiency. The guidelines further recommend that bone mass measurement be considered in *all* women by age 65, a position ratified by the U.S. Preventive Health Services

Task Force. Criteria approved for Medicare reimbursement of BMD are summarized in **Table 354-4**.

Table 354-4 FDA-Approved Indications for Bmd Tests*
Estrogen-deficient women at clinical risk of osteoporosis
Vertebral abnormalities on x-ray suggestive of osteoporosis (osteopenia, vertebral fracture)
Glucocorticoid treatment equivalent to ≥7.5 mg of prednisone or duration of therapy >3 months
Primary hyperparathyroidism
Monitoring response to an FDA-approved medication for osteoporosis
Repeat BMD evaluations at >23-month intervals or more frequently if medically justified

*Criteria adapted from the 1998 Bone Mass Measurement Act.

Abbreviations: BMD, bone mineral density; FDA, U.S. Food and Drug Administration.

When to Treat Based on Bone Mass Results

Most guidelines suggest that patients be considered for treatment when BMD is >2.5 SD below the mean value for young adults (T-score ≤–2.5), a level consistent with the diagnosis of osteoporosis. Treatment also should also be considered in postmenopausal women with fracture risk factors even if BMD is not in the osteoporosis range. Risk factors (age, prior fracture, family history of hip fracture, low body weight, cigarette consumption, excessive alcohol use, steroid use, and rheumatoid arthritis) can be combined with BMD to assess the likelihood of a fracture over a 5- or 10-year period. Treatment thresholds depend on cost-effectiveness analyses but probably is ~1% per year of risk in the United States.

TREATMENT OSTEOPOROSIS

The perimenopausal transition is a good opportunity to initiate a discussion about risk factors for osteoporosis and consideration of indications for a BMD test. A careful history and physical examination should be performed to identify risk factors for osteoporosis. A low Z-score increases the suspicion of a secondary disease. Height loss >2.5–3.8 cm (>1–1.5 in.) is an indication for radiography or vertebral fracture assessment by DXA to rule out asymptomatic vertebral fractures, as is the presence of significant kyphosis or back pain, particularly if it began after menopause. For patients who present with fractures, it is important to ensure that the fractures are not caused by an underlying malignancy. Usually this is clear on routine radiography, but on occasion, CT, MRI, or radionuclide scans may be necessary.

Routine Laboratory Evaluation

There is no established algorithm for the evaluation of women who present with osteoporosis. A general evaluation that includes complete blood count, serum and 24-h urine calcium, and renal and hepatic function tests is useful for identifying selected secondary causes of low bone mass, particularly for women with fractures or very low Z-scores. An elevated serum calcium level suggests hyperparathyroidism or malignancy, whereas a reduced serum calcium level may reflect malnutrition and osteomalacia. In the presence of hypercalcemia, a serum PTH level differentiates between hyperparathyroidism (PTH↑) and malignancy (PTH↓), and a high PTHrP level can help document the presence of humoral hypercalcemia of malignancy (Chap. 353). A low urine calcium (<50 mg/24 h) suggests osteomalacia, malnutrition, or malabsorption; a high urine calcium (>300 mg/24 h) is indicative of hypercalciuria and must be investigated further. Hypercalciuria occurs primarily in three situations: (1) a renal calcium leak, which is more common in males with osteoporosis; (2) absorptive hypercalciuria, which can be idiopathic or associated with increased $1,25(OH)_2D$ in granulomatous disease; or (3) hematologicmalignancies or conditions associated with excessive bone turnover such as Paget's disease, hyperparathyroidism, and hyperthyroidism.

Individuals who have osteoporosis-related fractures or bone density in the osteoporotic range should have a measurement of serum 25(OH)D level, since the intake of vitamin D required to achieve a target level >32 ng/mL is highly variable. Vitamin D levels should be optimized in all individuals being treated for osteoporosis. Hyperthyroidism should be evaluated by measuring thyroid-stimulating hormone (TSH).

When there is clinical suspicion of Cushing's syndrome, urinary free cortisol levels or a fasting serum cortisol should be measured after overnight dexamethasone. When bowel disease, malabsorption, or malnutrition is suspected, serum albumin, cholesterol, and a complete blood count should be checked. Asymptomatic malabsorption may be heralded by anemia (macrocytic—vitamin B_{12} or folate deficiency; microcytic—iron deficiency) or low serum cholesterol or urinary calcium levels. If these or other features suggest malabsorption, further evaluation is required. Asymptomatic celiac disease with selective malabsorption is being found with increasing frequency; the diagnosis can be made by testing for antigliadin, antiendomysial, or transglutaminase antibodies but may require endoscopic biopsy. A trial of a gluten-free diet can be confirmatory (Chap. 294). When osteoporosis is found associated with symptoms of rash, multiple allergies, diarrhea, or flushing, mastocytosis should be excluded by using 24-h urine histamine collection or serum tryptase.

Myeloma can masquerade as generalized osteoporosis, although it more commonly presents with bone pain and characteristic "punched-out" lesions on radiography. Serum and urine electrophoresis and evaluation for light chains in urine are required to exclude this diagnosis. A bone marrow biopsy may be required to rule out myeloma (in patients with equivocal electrophoretic results) and also can be used to exclude mastocytosis, leukemia, and other marrow infiltrative disorders such as Gaucher's disease.

Bone Biopsy

Tetracycline labeling of the skeleton allows determination of the rate of remodeling as well as evaluation for other metabolic bone diseases. The current use of BMD tests, in combination with hormonal evaluation and biochemical markers of bone remodeling, has largely replaced the clinical use of bone biopsy, although it remains an important tool in clinical research.

Biochemical Markers

Several biochemical tests are available that provide an index of the overall rate of bone remodeling **(Table 354-5)**. Biochemical markers usually are characterized as those related primarily to *bone formation* or *bone resorption*. These tests measure the overall state of bone remodeling at a single point in time. Clinical use of these tests has been hampered by biologic variability (in part related to circadian rhythm) as well as analytic variability, although the latter is improving.

Table 354-5 Biochemical Markers of Bone Metabolism in Clinical Use
Bone formation
Serum bone-specific alkaline phosphatase
Serum osteocalcin

Serum propeptide of type I procollagen
Bone resorption
Urine and serum cross-linked N-telopeptide
Urine and serum cross-linked C-telopeptide
Urine total free deoxypyridinoline

For the most part, remodeling markers do not predict rates of bone loss well enough for this information to be used clinically. However, markers of bone resorption may help in the prediction of fracture risk, independently of bone density, particularly in older individuals. In women ≥65 years, when bone density results are greater than the usual treatment thresholds noted above, a high level of bone resorption should prompt consideration of treatment. The primary use of biochemical markers is for monitoring the response to treatment. With the introduction of antiresorptive therapeutic agents, bone remodeling declines rapidly, with the fall in resorption occurring earlier than the fall in formation. Inhibition of bone resorption is maximal within 3–6 months. Thus, measurement of bone resorption before initiating therapy and 4–6 months after starting therapy provides an earlier estimate of patient response than does bone densitometry. A decline in resorptive markers can be ascertained after treatment with bisphosphonates or estrogen; this effect is less marked after treatment with either raloxifene or intranasal calcitonin. A biochemical marker response to therapy is particularly useful for asymptomatic patients and may help ensure long-term adherence to treatment. Bone turnover markers are also useful in monitoring the effects of 1-34hPTH, or teriparatide, which rapidly increases bone formation and later bone resorption.

Treatment: Osteoporosis

MANAGEMENT OF OSTEOPOROTIC FRACTURES

Treatment of a patient with osteoporosis frequently involves management of acute fractures as well as treatment of the underlying disease. Hip fractures almost always require surgical repair if the patient is to become ambulatory again. Depending on the location and severity of the fracture, condition of the neighboring joint, and general status of the patient, procedures may include open reduction and internal fixation with pins and plates, hemiarthroplasties, and total arthroplasties. These surgical procedures are followed by intense rehabilitation in an attempt to return patients to their prefracture functional level. Long bone fractures often require either external or internal fixation. Other fractures (e.g., vertebral, rib, and pelvic fractures) usually are managed with supportive care, requiring no specific orthopedic treatment.

Only ~25–30% of vertebral compression fractures present with sudden-onset back pain. For acutely symptomatic fractures, treatment with analgesics is required, including nonsteroidal anti-inflammatory agents and/or acetaminophen, sometimes with the addition of a narcotic agent (codeine or oxycodone). A few small, randomized clinical trials suggest that calcitonin may reduce pain related to acute vertebral compression fracture. A recently developed technique involves percutaneous injection of artificial cement (polymethylmethacrylate) into the vertebral body (vertebroplasty or kyphoplasty); this offers significant immediate pain relief in the majority of patients. Long-term effects are unknown, and conclusions are based on observational studies in patients with severe persistent back pain from acute or subacute vertebral fractures. There have been no long-term randomized controlled trials of either vertebroplasty or kyphoplasty to date. Short periods of bed rest may be helpful for pain management, but in general, early mobilization is recommended as it helps prevent further bone loss associated with immobilization. Occasionally, use of a soft elastic-style brace may facilitate earlier mobilization. Muscle spasms often occur with acute compression fractures and can be treated with muscle relaxants and heat treatments.

Severe pain usually resolves within 6–10 weeks. Chronic pain is probably not bony in origin; instead, it is related to abnormal strain on muscles, ligaments, and tendons and to secondary facet-joint arthritis associated with alterations in thoracic and/or abdominal shape. Chronic pain is difficult to treat effectively and may require analgesics, sometimes including narcotic analgesics. Frequent intermittent rest in a supine or semireclining position is often required to allow the soft tissues, which are under tension, to relax. Back-strengthening exercises (paraspinal) may be beneficial. Heat treatments help relax muscles and reduce the muscular component of discomfort. Various physical modalities, such as US and transcutaneous nerve stimulation, may be beneficial in some patients. Pain also occurs in the neck region, not as a result of compression fractures (which almost never occur in the cervical spine as a result of osteoporosis) but because of chronic strain associated with trying to elevate the head in a person with a severe thoracic kyphosis.

Multiple vertebral fractures often are associated with psychological symptoms; this is not always appreciated. The changes in body configuration and back pain can lead to marked loss of self-image and a secondary depression. Altered balance, precipitated by the kyphosis and the anterior movement of the body's center of gravity, leads to a fear of falling, a consequent tendency to remain indoors, and the onset of social isolation. These symptoms sometimes can be alleviated by family support and/or psychotherapy. Medication may be necessary when depressive features are present.

MANAGEMENT OF THE UNDERLYING DISEASE

Risk Factor Reduction

Assessment of fracture risk can be estimated by using FRAX calculators that are available online (*http://www.shef.ac.uk/FRAX/tool.jsp?locationValue=9*) **(Fig. 354-7)**. Patients should be thoroughly educated to reduce the impact of modifiable risk factors associated with bone loss and falling. Medications should be reviewed to ensure that all are necessary. Glucocorticoid medication, if present, should be evaluated to determine that it is truly indicated and is being given in doses that are as low as possible. For those on thyroid hormone replacement, TSH testing should be performed to determine that an excessive dose is not being used, as thyrotoxicosis can be associated with increased bone loss. In patients who smoke, efforts should be made to facilitate smoking cessation. Reducing risk factors for falling also include alcohol abuse treatment and a review of the medical regimen for any drugs that might be associated with orthostatic hypotension and/or sedation, including hypnotics and anxiolytics. If nocturia occurs, the frequency should be reduced, if possible (e.g., by decreasing or modifying diuretic use), as arising in the middle of sleep is a common precipitant of a fall. Patients should be instructed about environmental safety with regard to eliminating exposed wires, curtain strings, slippery rugs, and mobile tables. Avoiding stocking feet on wood floors, checking carpet condition (particularly on stairs), and providing good light in paths to bathrooms and outside the home are important preventive measures. Treatment for impaired vision is recommended, particularly a problem with depth perception, which is specifically associated with increased falling risk. Elderly patients with neurologic impairment (e.g., stroke, Parkinson's disease, Alzheimer's disease) are particularly at risk of falling and require specialized supervision and care.

Figure 354-7

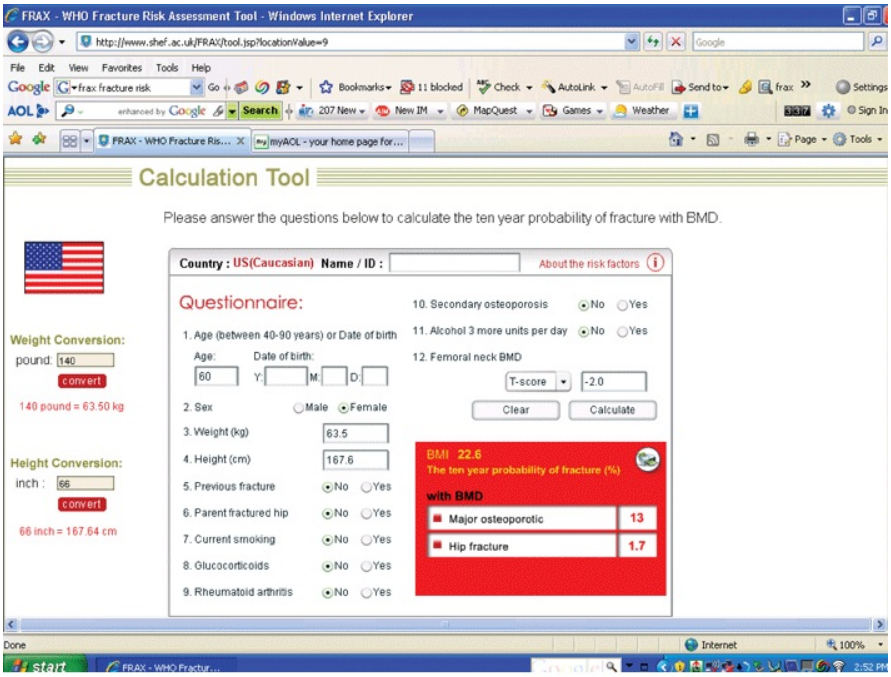

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

FRAX calculation tool. When the answers to the indicated questions are filled in, the calculator can be used to assess the 10-year probability of fracture. The calculator (available online at *http://www.shef.ac.uk/FRAX/tool.jsp?locationValue=9*) also can risk adjust for various ethnic groups.

Nutritional Recommendations

Calcium

A large body of data indicates that optimal calcium intake reduces bone loss and suppresses bone turnover. Recommended intakes from an Institute of Medicine report are shown in **Table 354-6**. The National Health and Nutritional Evaluation Studies (NHANES) have consistently documented that average calcium intakes fall considerably short of these recommendations. The preferred source of calcium is dairy products and other foods, but many patients require calcium supplementation. Food sources of calcium are dairy products (milk, yogurt, and cheese) and fortified foods such as certain cereals, waffles, snacks, juices, and crackers. Some of these fortified foods contain as much calcium per serving as milk.

Table 354-6 Adequate Calcium Intake

Life Stage Group	Estimated Adequate Daily Calcium Intake, mg/d
Young children (1–3 years)	500
Older children (4–8 years)	800
Adolescents and young adults (9–18 years)	1300
Men and women (19–50 years)	1000
Men and women (51 and older)	1200

Note: Pregnancy and lactation needs are the same as for nonpregnant women (e.g., 1300 mg/d for adolescents/young adults and 1000 mg/d for ≥19 years).

Source: Adapted from the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes. Food and Nutrition Board. Institute of Medicine. Washington, DC, 1997, National Academy Press.

If a calcium supplement is required, it should be taken in doses ≤600 mg at a time, as the calcium absorption fraction decreases at higher doses. Calcium supplements should be calculated on the basis of the elemental calcium content of the supplement, not the weight of the calcium salt **(Table 354-7)**. Calcium supplements containing carbonate are best taken with food since they require acid for solubility. Calcium citrate supplements can be taken at any time. To confirm bioavailability, calcium supplements can be placed in distilled vinegar. They should dissolve within 30 min.

Table 354-7 Elemental Calcium Content of Various Oral Calcium Preparations

Calcium Preparation	Elemental Calcium Content
Calcium citrate	60 mg/300 mg
Calcium lactate	80 mg/600 mg
Calcium gluconate	40 mg/500 mg
Calcium carbonate	400 mg/g
Calcium carbonate + 5 μg vitamin D2 (OsCal 250)	250 mg/tablet
Calcium carbonate (Tums 500)	500 mg/tablet

Source: Adapted from SM Krane, MF Holick, Chap. 355, in *Harrison's Principles of Internal Medicine* , 14th ed. New York, McGraw-Hill, 1998.

Several controlled clinical trials of calcium plus vitamin D have confirmed reductions in clinical fractures, including fractures of the hip (~20–30% risk reduction). All recent studies of pharmacologic agents have been conducted in the context of calcium replacement (± vitamin D). Thus, it is standard practice to ensure an adequate calcium and vitamin D intake in patients with

osteoporosis whether they are receiving additional pharmacologic therapy or not. A systematic review confirmed a greater BMD response to antiresorptive therapy when calcium intake was adequate.

Although side effects from supplemental calcium are minimal (eructation and constipation mostly with carbonate salts), individuals with a history of kidney stones should have a 24-h urine calcium determination before starting increased calcium to avoid significant hypercalciuria.

Vitamin D

Vitamin D is synthesized in skin under the influence of heat and ultraviolet light (Chap. 352). However, large segments of the population do not obtain sufficient vitamin D to maintain what is now considered an adequate supply [serum 25(OH)D consistently >75 μmol/L (30 ng/mL)]. Since vitamin D supplementation at doses that would achieve these serum levels is safe and inexpensive, the Institute of Medicine recommends daily intakes of 200 IU for adults <50 years of age, 400 IU for those 50 –70 years, and 600 IU for those >70 years. Multivitamin tablets usually contain 400 IU, and many calcium supplements also contain vitamin D. Some data suggest that higher doses (≥1000 IU) may be required in the elderly and chronically ill.

Other Nutrients

Other nutrients such as salt, high animal protein intakes, and caffeine may have modest effects on calcium excretion or absorption. Adequate vitamin K status is required for optimal carboxylation of osteocalcin. States in which vitamin K nutrition or metabolism is impaired, such as with long-term warfarin therapy, have been associated with reduced bone mass. Research concerning cola intake is controversial but suggests a possible link to reduced bone mass through factors that are independent of caffeine.

Magnesium is abundant in foods, and magnesium deficiency is quite rare in the absence of a serious chronic disease. Magnesium supplementation may be warranted in patients with inflammatory bowel disease, celiac disease, chemotherapy, severe diarrhea, malnutrition, or alcoholism. Dietary phytoestrogens, which are derived primarily from soy products and legumes [e.g., garbanzo beans (chickpeas) and lentils], exert some estrogenic activity but are insufficiently potent to justify their use in place of a pharmacologic agent in the treatment of osteoporosis.

Patients with hip fractures are often frail and relatively malnourished. Some data suggest an improved outcome in such patients when they are provided calorie and protein supplementation. Excessive protein intake can increase renal calcium excretion, but this can be corrected by an adequate calcium intake.

Exercise

Exercise in young individuals increases the likelihood that they will attain the maximal genetically determined peak bone mass. Meta-analyses of studies performed in postmenopausal women indicate that weight-bearing exercise prevents bone loss but does not appear to result in substantial gain of bone mass. This beneficial effect wanes if exercise is discontinued. Most of the studies are short term, and a more substantial effect on bone mass is likely if exercise is continued over a long period. Exercise also has beneficial effects on neuromuscular function, and it improves coordination, balance, and strength, thereby reducing the risk of falling. A walking program is a practical way to start. Other activities such as dancing, racquet sports, cross-country skiing, and use of gym equipment, are also recommended, depending on the patient's personal preference and general condition. Even women who cannot walk benefit from swimming or water exercises, not so much for the effects on bone, which are quite minimal, but because of effects on muscle. Exercise habits should be consistent, optimally at least three times a week.

PHARMACOLOGIC THERAPIES

Until fairly recently, estrogen treatment, either by itself or in concert with a progestin, was the primary therapeutic agent for prevention or treatment of osteoporosis. However, a number of new drugs have appeared, and more are expected in the near future. Some are agents that specifically treat osteoporosis (bisphosphonates, calcitonin, PTH); others such as selective estrogen response modulators (SERMs), have broader effects. The availability of these drugs allows therapy to be tailored to the needs of an individual patient.

Estrogens

A large body of clinical trial data indicates that various types of estrogens (conjugated equine estrogens, estradiol, estrone, esterified estrogens, ethinyl estradiol, and mestranol) reduce bone turnover, prevent bone loss, and induce small increases in bone mass of the spine, hip, and total body. The effects of estrogen are seen in women with natural or surgical menopause and in late postmenopausal women with or without established osteoporosis. Estrogens are efficacious when administered orally or transdermally. For both oral and transdermal routes of administration, combined estrogen/progestin preparations are now available in many countries, obviating the problem of taking two tablets or using a patch and oral progestin. One large study, referred to as PEPI (Postmenopausal Estrogen/Progestin Intervention Trial), indicated that C-21 progestins alone do not augment the effect of standard estrogen doses on bone mass.

Dose of Estrogen

For oral estrogens, the standard recommended doses have been 0.3 mg/d for esterified estrogens, 0.625 mg/d for conjugated equine estrogens, and 5 μg/d for ethinyl estradiol. For transdermal estrogen, the commonly used dose supplies 50 μg estradiol per day, but a lower dose may be appropriate for some individuals. Dose-response data for conjugated equine estrogens indicate that lower doses (0.3 and 0.45 mg/d) are effective. Doses even lower have been associated with bone mass protection.

Fracture Data

Epidemiologic databases indicate that women who take estrogen replacement have a 50% reduction, on average, of osteoporotic fractures, including hip fractures. The beneficial effect of estrogen is greatest among those who start replacement early and continue the treatment; the benefit declines after discontinuation to the extent that there is no residual protective effect against fracture by 10 years after discontinuation. The first clinical trial evaluating fractures as secondary outcomes, the Heart and Estrogen-Progestin Replacement Study (HERS) trial, showed no effect of hormone therapy on hip or other clinical fractures in women with established coronary artery disease. These data made the results of the Women's Health Initiative (WHI) exceedingly important (Chap. 348). The estrogen-progestin arm of the WHI in >16,000 postmenopausal healthy women indicated that hormone therapy reduces the risk of hip and clinical spine fracture by 34% and that of all clinical fractures by 24%.

A few smaller clinical trials have evaluated spine fracture occurrence as an outcome with estrogen therapy. They have consistently shown that estrogen treatment reduces the incidence of vertebral compression fracture.

The WHI has provided a vast amount of data on the multisystemic effects of hormone therapy. Although earlier observational studies suggested that estrogen replacement might reduce heart disease, the WHI showed that combined estrogen-progestin treatment increased risk of fatal and nonfatal myocardial infarction by ~29%, confirming data from the HERS study. Other important relative risks included a 40% increase in stroke, a 100% increase in venous thromboembolic disease, and a 26% increase in risk of breast cancer. Subsequent analyses have confirmed the increased risk of stroke and shown a twofold increase in dementia. Benefits other than the fracture reductions noted above included a 37% reduction in the risk of colon

cancer. These relative risks have to be interpreted in light of absolute risk **(Fig. 354-8)**. For example, out of 10,000 women treated with estrogen-progestin for 1 year, there will be 8 excess heart attacks, 8 excess breast cancers, 18 excess venous thromboembolic events, 5 fewer hip fractures, 44 fewer clinical fractures, and 6 fewer colorectal cancers. These numbers must be multiplied by years of hormone treatment. There was no effect of hormone treatment on the risk of uterine cancer or total mortality.

Figure 354-8

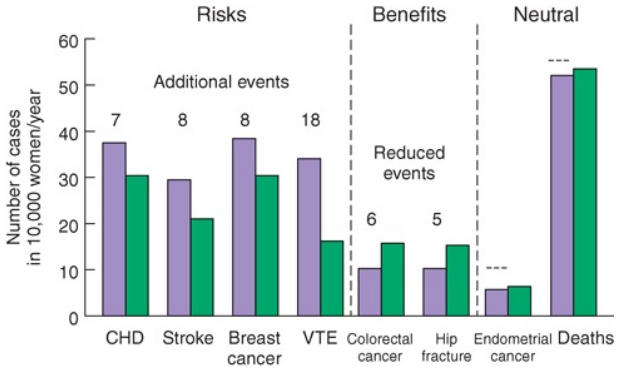

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effects of hormone therapy on event rates: green, placebo; purple, estrogen and progestin. CHD, coronary heart disease; VTE, venous thromboembolic events. (*Adapted from Women's Health Initiative. WHI HRT Update. Available at http://www.nhlbi.nih.gov/health/women/upd2002.htm.*)

It is important to note that these WHI findings apply specifically to hormone treatment in the form of conjugated equine estrogen plus medroxyprogesterone acetate. The relative benefits and risks of unopposed estrogen in women who had hysterectomies vary somewhat. They still show benefits against fracture occurrence and increased risk of venous thrombosis and stroke, similar in magnitude to the risks for combined hormone therapy. In contrast, though, the estrogen-only arm of WHI indicated no increased risk of heart attack or breast cancer. The data suggest that at least some of the detrimental effects of combined therapy are related to the progestin component.

Mode of Action

Two subtypes of ERs, α and β, have been identified in bone and other tissues. Cells of monocyte lineage express both ERα and ERβ, as do osteoblasts. Estrogen-mediated effects vary with the receptor type. Using ER knockout mouse models, elimination of ERα produces a modest reduction in bone mass, whereas mutation of ERβ has less of an effect on bone. A male patient with a homozygous mutation of ERα had markedly decreased bone density as well as abnormalities in epiphyseal closure, confirming the important role of ERα in bone biology. The mechanism of estrogen action in bone is an area of active investigation (Fig. 354-5). Although data are conflicting, estrogens may inhibit osteoclasts directly. However, the majority of estrogen (and androgen) effects on bone resorption are mediated indirectly through paracrine factors produced by osteoblasts. These actions include (1) increasing IGF-I and TGF-β and (2) suppressing IL-1 (α and β), IL-6, TNF-α, and osteocalcin synthesis. The indirect estrogen actions primarily decrease bone resorption.

Progestins

In women with a uterus, daily progestin or cyclical progestins at least 12 days per month are prescribed in combination with estrogens to reduce the risk of uterine cancer. Medroxyprogesterone acetate and norethindrone acetate blunt the high-density lipoprotein response to estrogen, but micronized progesterone does not. Neither medroxyprogesterone acetate nor micronized progesterone appears to have an independent effect on bone; at lower doses of estrogen, norethindrone acetate may have an additive benefit. On breast tissue, progestins may increase the risk of breast cancer.

Serms

Two SERMs are used currently in postmenopausal women: raloxifene, which is approved for the prevention and treatment of osteoporosis, and tamoxifen, which is approved for the prevention and treatment of breast cancer.

Tamoxifen reduces bone turnover and bone loss in postmenopausal women compared with placebo groups. These findings support the concept that tamoxifen acts as an estrogenic agent in bone. There are limited data on the effect of tamoxifen on fracture risk, but the Breast Cancer Prevention study indicated a possible reduction in clinical vertebral, hip, and Colles' fractures. The major benefit of tamoxifen is on breast cancer occurrence. The breast cancer prevention trial indicated that tamoxifen administration over 4–5 years reduced the incidence of new invasive and noninvasive breast cancer by ~45% in women at increased risk of breast cancer. The incidence of ER-positive breast cancers was reduced by 65%. Tamoxifen increases the risk of uterine cancer in postmenopausal women, limiting its use for breast cancer prevention in women at low or moderate risk.

Raloxifene (60 mg/d) has effects on bone turnover and bone mass that are very similar to those of tamoxifen, indicating that this agent is also estrogenic on the skeleton. The effect of raloxifene on bone density (+1.4–2.8% versus placebo in the spine, hip, and total body) is somewhat less than that seen with standard doses of estrogens. Raloxifene reduces the occurrence of vertebral fracture by 30–50%, depending on the population; however, there are no data confirming that raloxifene can reduce the risk of nonvertebral fractures over 8 years of observation.

Raloxifene, like tamoxifen and estrogen, has effects in other organ systems. The most beneficial effect appears to be a reduction in invasive breast cancer (mainly decreased ER-positive) occurrence of ~65% in women who take raloxifene compared to placebo. In a head-to-head study raloxifene was as effective as tamoxifen in preventing breast cancer in high-risk women, but in a separate study it had no effect on heart disease in women with increased risk for this outcome. In contrast to tamoxifen, raloxifene is not associated with an increase in the risk of uterine cancer or benign uterine disease. Raloxifene increases the occurrence of hot flashes but reduces serum total and low-density lipoprotein cholesterol, lipoprotein(a), and fibrinogen.

Mode of Action of Serms

All SERMs bind to the ER, but each agent produces a unique receptor-drug conformation. As a result, specific coactivator or co-repressor proteins are bound to the receptor (Chap. 338), resulting in differential effects on gene transcription that vary

depending on other transcription factors present in the cell. Another aspect of selectivity is the affinity of each SERM for the different ERα and ERβ subtypes, which are expressed differentially in various tissues. These tissue-selective effects of SERMs offer the possibility of tailoring estrogen therapy to best meet the needs and risk factor profile of an individual patient.

Bisphosphonates

Alendronate, risedronate, and ibandronate are approved for the prevention and treatment of postmenopausal osteoporosis. Risedronate and alendronate are approved for the treatment of steroid-induced osteoporosis, and risedronate also is approved for prevention of steroid-induced osteoporosis. Both alendronate and risedronate are approved for treatment of osteoporosis in men.

Alendronate has been shown to decrease bone turnover and increase bone mass in the spine by up to 8% versus placebo and by 6% versus placebo in the hip. Multiple trials have evaluated its effect on fracture occurrence. The Fracture Intervention Trial provided evidence in >2000 women with prevalent vertebral fractures that daily alendronate treatment (5 mg/d for 2 years and 10 mg/d for 9 months afterward) reduces vertebral fracture risk by about 50%, multiple vertebral fractures by up to 90%, and hip fractures by up to 50%. Several subsequent trials have confirmed these findings **(Figs. 354-9 and 354-10)**. For example, in a study of >1900 women with low bone mass treated with alendronate (10 mg/d) versus placebo, the incidence of all nonvertebral fractures was reduced by ~47% after only 1 year.

Figure 354-9

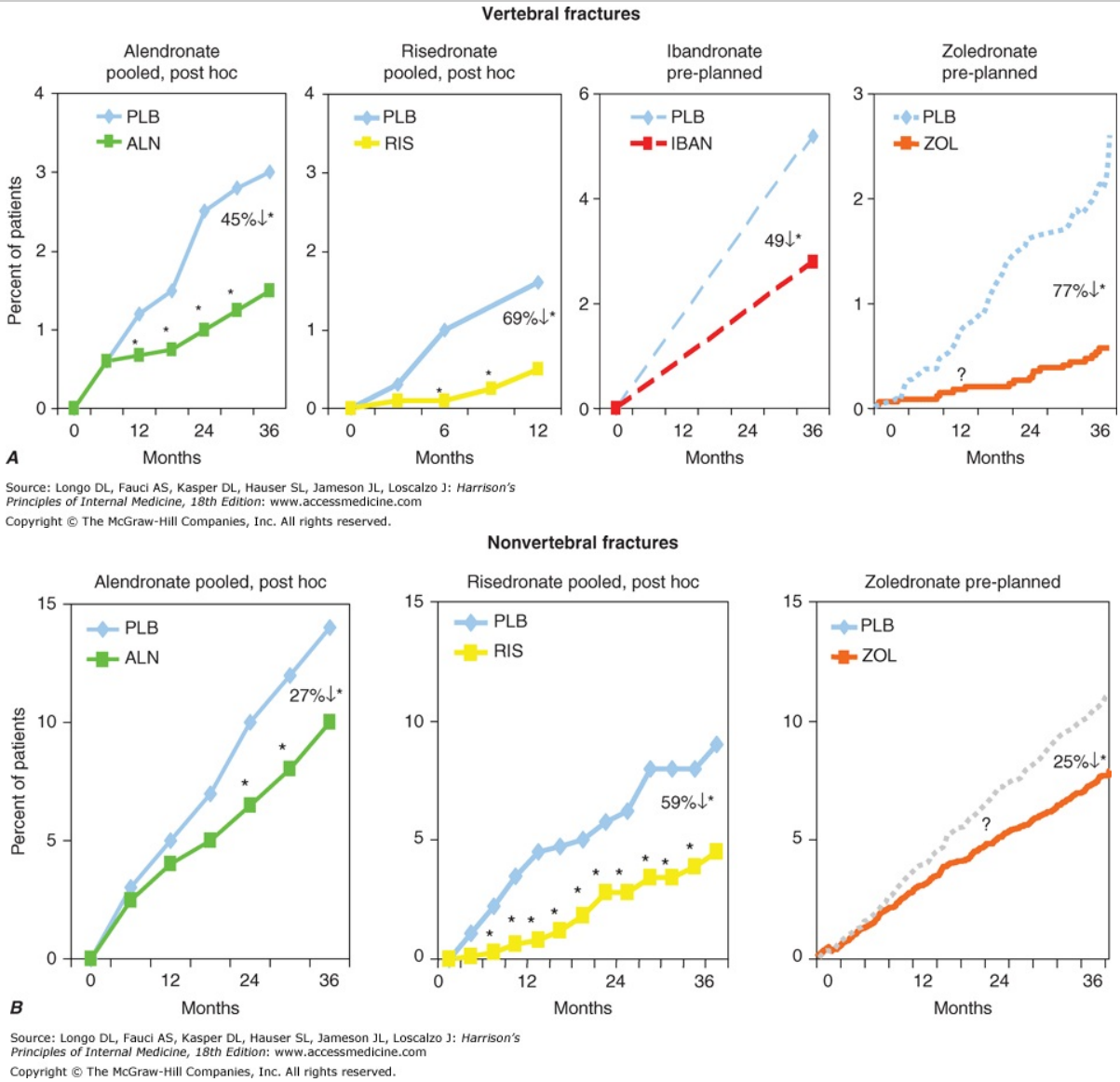

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

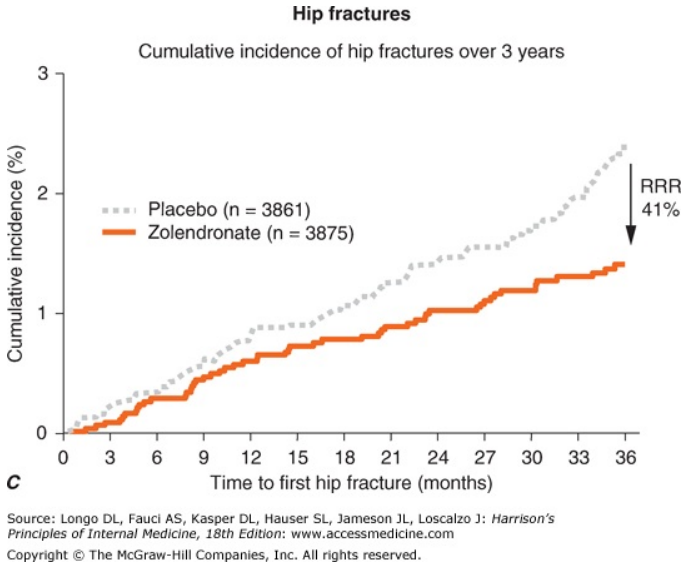

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effects of various bisphosphonates on clinical vertebral fractures A., nonvertebral fractures B., and hip fractures C. Plb, placebo; RRR, relative risk reduction. (*After DM Black et al: J Clin Endocrinol Metab 85:4118, 2000; C Roux et al: Curr Med Res Opin 4:433, 2004; CH Chesnut et al: J Bone Miner Res 19: 1241, 2004; DM Black et al: N Engl J Med 356:1809, 2007;JT Harrington et al: Calcif Tissue Int 74:129, 2003.*)

Figure 354-10

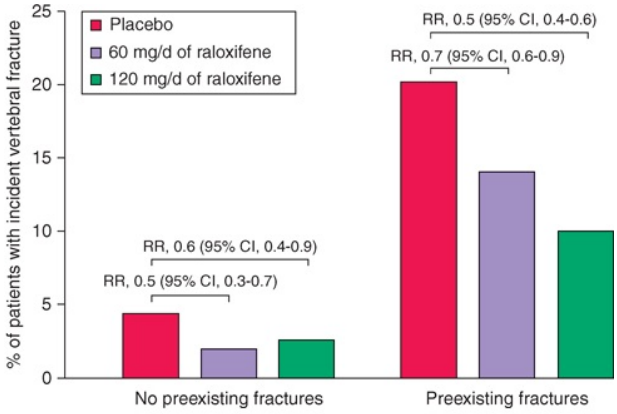

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effects of two doses of raloxifene on incident vertebral fractures in the MORE trial. *(After B Ettinger et al: JAMA:282:637, 1999.)*

Trials comparing once-weekly alendronate, 70 mg, with daily 10-mg dosing have shown equivalence with regard to bone mass and bone turnover responses. Consequently, once-weekly therapy generally is preferred because of the low incidence of gastrointestinal side effects and ease of administration. Alendronate should be given with a full glass of water before breakfast, as bisphosphonates are poorly absorbed. Because of the potential for esophageal irritation, alendronate is contraindicated in patients who have stricture or inadequate emptying of the esophagus. It is recommended that patients remain upright for at least 30 min after taking the medication to avoid esophageal irritation. Cases of esophagitis, esophageal ulcer, and esophageal stricture have been described, but the incidence appears to be low. In clinical trials, overall gastrointestinal symptomatology was no different with alendronate than with placebo. Alendronate is also available in a preparation that contains vitamin D.

Risedronate also reduces bone turnover and increases bone mass. Controlled clinical trials have demonstrated 40–50% reduction in vertebral fracture risk over 3 years, accompanied by a 40% reduction in clinical nonspine fractures. The only clinical trial specifically designed to evaluate hip fracture outcome (HIP) indicated that risedronate reduced hip fracture risk in women in their seventies with confirmed osteoporosis by 40%. In contrast, risedronate was not effective at reducing hip fracture occurrence in older women (80+ years) without proven osteoporosis. Studies have shown that 35 mg of risedronate administered once weekly is therapeutically equivalent to 5 mg/d. Patients should take risedronate with a full glass of plain water to facilitate delivery to the stomach and should not lie down for 30 min after taking the drug. The incidence of gastrointestinal side effects in trials with risedronate was similar to that of placebo.

Etidronate was the first bisphosphonate to be approved, initially for use in Paget's disease and hypercalcemia. This agent has also been used in osteoporosis trials of smaller magnitude than those performed for alendronate and risedronate but is not approved by the FDA for treatment of osteoporosis. Etidronate probably has some efficacy against vertebral fracture when given as an intermittent cyclical regimen (2 weeks on, 2.5 months off). Its effectiveness against nonvertebral fractures has not been studied.

Ibandronate is the third amino-bisphosphonate approved in the United States. Ibandronate (2.5 mg/d) has been shown in clinical trials to reduce vertebral fracture risk by ~40% but with no overall effect on nonvertebral fractures. In a post hoc analysis of subjects with a femoral neck T-score of –3 or below, ibandronate reduced the risk of nonvertebral fractures by ~60%. In clinical trials, ibandronate doses of 150 mg/month PO or 3 mg every 3 months IV had greater effects on turnover and bone mass than did 2.5 mg/d. Patients should take oral ibandronate in the same way as other bisphosphonates, but with 1 h elapsing before other food or drink (other than plain water).

Zoledronic acid is a potent bisphosphonate with unique administration regimens (once yearly IV). Although it has not been approved for use in osteoporosis, the data suggest that it is highly effective in fracture risk reduction. In a study of >7000 women followed for 3 years, zoledronic acid (5 mg as a single IV infusion annually) reduced the risk of vertebral fractures by 70%, nonvertebral fractures by 25%, and hip fractures by 40%. These results were associated with less height loss and disability. In the treated population, there was an increased risk of atrial fibrillation (2%) and arthralgia and a 15% risk of fever in comparison to placebo.

Mode of Action

Bisphosphonates are structurally related to pyrophosphates, compounds that are incorporated into bone matrix. Bisphosphonates specifically impair osteoclast function and reduce osteoclast number, in part by inducing apoptosis. Recent evidence suggests that the nitrogen-containing bisphosphonates also inhibit protein prenylation, one of the end products in the mevalonic acid pathway, by inhibiting the enzyme farnesyl pyrophosphate synthase. This effect disrupts intracellular protein trafficking and ultimately may lead to apoptosis. Some bisphosphonates have very long retention in the skeleton and may exert long-term effects. The consequences of this, if any, are unknown. A phenomenon that has been called *osteonecrosis of the jaw* (ONJ) has been described, mostly in patients with cancer who are given high doses of zoledronic acid or pamidronate. A few cases have been described in patients with osteoporosis treated with oral bisphosphonates. The background incidence of ONJ in this population is not known, and thus the attributable risk for bisphosphonates is not clear, although it appears to be relatively low.

Calcitonin

Calcitonin is a polypeptide hormone produced by the thyroid gland (Chap. 353). Its physiologic role is unclear as no skeletal disease has been described in association with calcitonin deficiency or excess. Calcitonin preparations are approved by the FDA for Paget's disease, hypercalcemia, and osteoporosis in women >5 years past menopause.

Injectable calcitonin produces small increments in bone mass of the lumbar spine. However, difficulty of administration and frequent reactions, including nausea and facial flushing, make general use limited. A nasal spray containing calcitonin (200 IU/d) is available for treatment of osteoporosis in postmenopausal women. One study suggests that nasal calcitonin produces small increments in bone mass and a small reduction in new vertebral fractures in calcitonin-treated patients versus those on calcium alone. There has been no proven effectiveness against nonvertebral fractures. An oral preparation of calcitonin recently was approved for use in osteoporosis.

Calcitonin is not indicated for prevention of osteoporosis and is not sufficiently potent to prevent bone loss in early postmenopausal women. Calcitonin might have an analgesic effect on bone pain, both in the subcutaneous and possibly in the nasal form.

Mode of Action

Calcitonin suppresses osteoclast activity by direct action on the osteoclast calcitonin receptor. Osteoclasts exposed to calcitonin cannot maintain their active ruffled border, which normally maintains close contact with underlying bone.

Denosumab

A novel agent that was given twice yearly by SC administration in a randomized controlled trial in postmenopausal women with osteoporosis has been shown to increase BMD in the spine, hip, and forearm and reduce vertebral, hip, and nonvertebral fractures over a 3-year period by 70, 40, and 20%, respectively **(Fig. 354-11)**. Other clinical trials indicate ability to increase bone mass in postmenopausal women with low bone mass (above osteoporosis range) and in postmenopausal women with breast cancer treated with hormonal agents. Furthermore, a study of men with prostate cancer treated with gonadotropin-releasing hormone (GnRH) agonist therapy indicated the ability of denosumab to improve bone mass and reduce vertebral fracture occurrence. Denosumab was approved by the FDA in 2010 for the treatment of postmenopausal women who have a high risk for osteoporotic fractures, including those with a history of fracture or multiple risk factors for fracture, and those who have failed or are intolerant to other osteoporosis therapy.

Figure 354-11

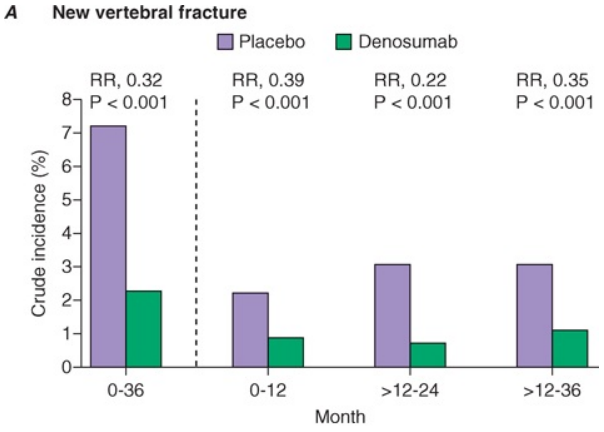

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

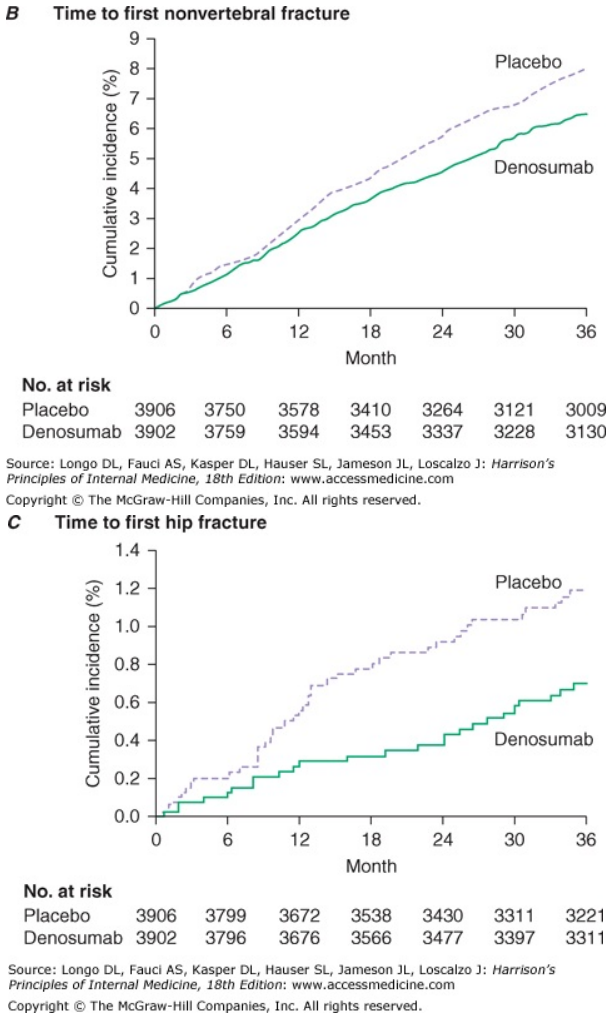

Effects of denosumab on new vertebral fractures ***A.*** and times to nonvertebral and hip fracture B. and C. *(After SR Cummings et al: N Engl J Med:361:756, 2009.)*

Mode of Action

Denosumab is a fully human monoclonal antibody to RANKL, the final common effector of osteoclast formation, activity, and survival. Denosumab binds to RANKL, inhibiting its ability to initiate formation of mature osteoclasts from osteoclast precursors and to bring mature osteoclasts to the bone surface and initiate bone resorption. Denosumab also plays a role in reducing the survival of the osteoclast. Through these actions on the osteoclast, denosumab induces potent antiresorptive action, as assessed biochemically and histomorphometrically, and may contribute to the occurrence of ONJ. Serious adverse reactions include hypocalcemia, infections, and dermatologic reactions such as dermatitis, rashes, and eczema.

Parathyroid Hormone

Endogenous PTH is an 84-amino-acid peptide that is largely responsible for calcium homeostasis (Chap. 353). Although chronic elevation of PTH, as occurs in hyperparathyroidism, is associated with bone loss (particularly cortical bone), PTH also can exert anabolic effects on bone. Consistent with this, some observational studies have indicated that mild elevations in PTH are associated with maintenance of trabecular bone mass. On the basis of these findings, several clinical trials have been performed using an exogenous PTH analogue (1-34hPTH; teriparatide) that has been approved for the treatment of established osteoporosis in both men and women. The first randomized controlled trial in postmenopausal women showed that PTH, when superimposed on ongoing estrogen therapy, produced substantial increments in bone mass (13% over a 3-year period compared with estrogen alone) and reduced the risk of vertebral compression deformity. In the pivotal study (median, 19 months' duration), 20 μg PTH(1–34) daily by SC injection reduced vertebral fractures by 65% and nonvertebral fractures by 45% **(Fig. 354-12)**. Treatment is administered as a single daily injection given for a maximum of 2 years. Teriparatide produces increases in bone mass and mediates architectural improvements in skeletal structure. These effects are lower when patients have been exposed previously to bisphosphonates, possibly in proportion to the potency of the antiresorptive effect. When 1–34hPTH is being considered for treatment-naive patients, it is best administered as monotherapy and followed by an antiresorptive agent such as a bisphosphonate.

Figure 354-12

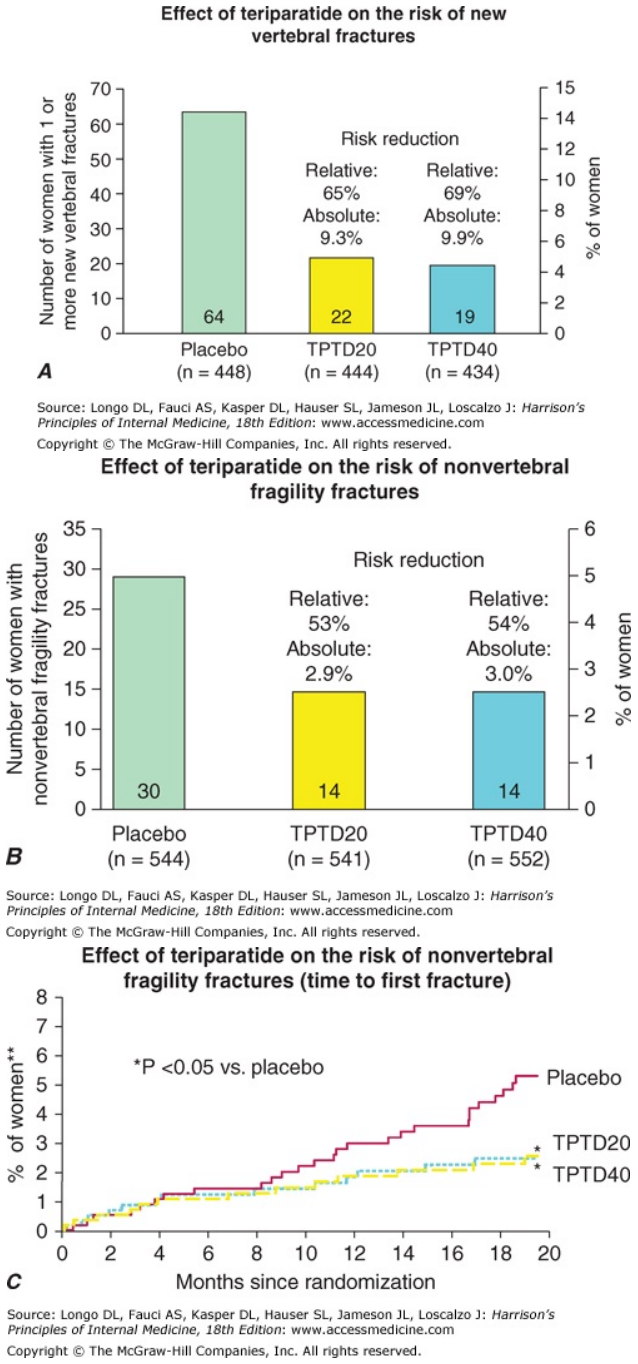

Effects of teriparatide on new vertebral fractures A. and nonvertebral fragility fractures B. and C. *(After RM Neer et al: N Engl J Med 344:1434, 2001.)*

Side effects of teriparatide are generally mild and can include muscle pain, weakness, dizziness, headache, and nausea. Rodents given prolonged treatment with PTH in relatively high doses developed osteogenic sarcomas. One case of osteosarcoma has been described in a patient treated with teriparatide. At present this seems to equate to the background incidence of osteosarcoma in this population.

PTH use may be limited by its mode of administration; alternative modes of delivery are being investigated. The optimal frequency of administration also remains to be established, and it is possible that PTH might be effective when used intermittently. Cost also may be a limiting factor.

Mode of Action

Exogenously administered PTH appears to have direct actions on osteoblast activity, with biochemical and histomorphometric evidence of de novo bone formation early in response to PTH, before activation of bone resorption. Subsequently, PTH activates bone remodeling but still appears to favor bone formation over bone resorption. PTH stimulates IGF-I and collagen production and appears to increase osteoblast number by stimulating replication, enhancing osteoblast recruitment, and inhibiting apoptosis. Unlike all other treatments, PTH produces a true increase in bone tissue and an apparent restoration of bone microarchitecture **(Fig. 354-13)**.

Figure 354-13

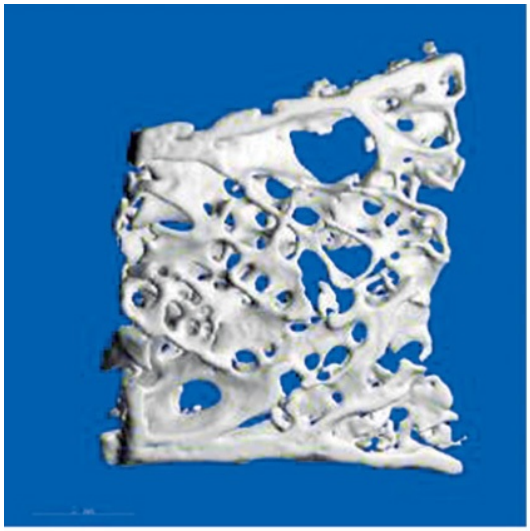

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

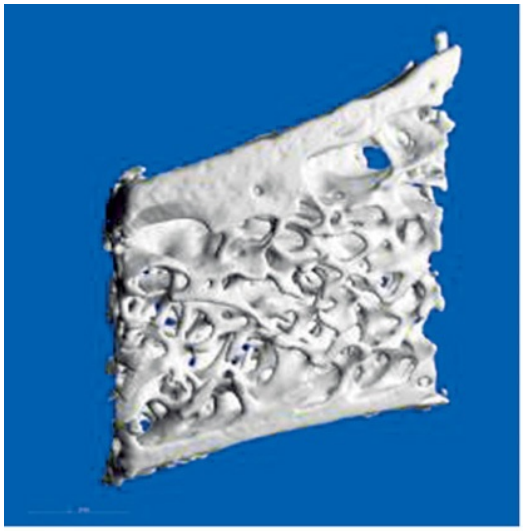

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Effect of parathyroid hormone (PTH) treatment on bone microarchitecture. Paired biopsy specimens from a 64-year-old woman before ***A.*** and after ***B.*** treatment with PTH. *(From DW Dempster et al: J Bone Miner Res 16:1846, 2001.*)

Fluoride

Fluoride has been available for many years and is a potent stimulator of osteoprogenitor cells when studied in vitro. It has been used in multiple osteoporosis studies with conflicting results, in part because of the use of varying doses and preparations. Despite increments in bone mass of up to 10%, there are no consistent effects of fluoride on vertebral or nonvertebral fracture; the latter may actually increase when high doses of fluoride are used. Fluoride remains an experimental agent despite its long history and multiple studies.

Strontium Ranelate Strontium ranelate is approved in several European countries for the treatment of osteoporosis. It increases bone mass throughout the skeleton; in clinical trials, the drug reduced the risk of vertebral fractures by 37% and that of nonvertebral fractures by 14%. It appears to be modestly antiresorptive while at the same time not causing as much of a decrease in bone formation (measured biochemically). Strontium is incorporated into hydroxyapatite, replacing calcium, a feature that might explain some of its fracture benefits. Small increased risks of venous thrombosis, seizures, and abnormal cognition have been seen and require further study.

Other Potential Anabolic Agents

Several small studies of growth hormone (GH), alone or in combination with other agents, have not shown consistent or substantial positive effects on skeletal mass. Many of these studies have been relatively short term, and the effects of GH, growth hormone–releasing hormone, and the IGFs are still under investigation. Anabolic steroids, mostly derivatives of testosterone, act primarily as antiresorptive agents to reduce bone turnover but also may stimulate osteoblastic activity. Effects on bone mass remain unclear but appear weak in general, and use is limited by masculinizing side effects. Several recent observational studies suggest that the statin drugs, which currently are used to treat hypercholesterolemia, may be associated with increased bone mass and reduced fractures, but conclusions from clinical trials are mixed.

NONPHARMACOLOGIC APPROACHES

Protective pads worn around the outer thigh, which cover the trochanteric region of the hip, can prevent hip fractures in elderly residents in nursing homes. The use of hip protectors is limited largely by issues of compliance and comfort, but new devices are being developed that may circumvent these problems and provide adjunctive treatments.

Kyphoplasty and *vertebroplasty* are also useful nonpharmacologic approaches for the treatment of painful vertebral fractures. However, no long-term data are available.

TREATMENT MONITORING

There are currently no well-accepted guidelines for monitoring treatment of osteoporosis. Because most osteoporosis treatments produce small or moderate bone mass increments on average, it is reasonable to consider BMD as a monitoring tool. Changes must exceed ~4% in the spine and 6% in the hip to be considered significant in any individual. The hip is the preferred site due to larger surface area and greater reproducibility. Medication-induced increments may require several years to produce changes of this magnitude (if they do at all). Consequently, it can be argued that BMD should be repeated at intervals >2 years. Only significant BMD reductions should prompt a change in medical regimen, as it is expected that many individuals will not show responses greater than the detection limits of the current measurement techniques.

Biochemical markers of bone turnover may prove useful for treatment monitoring, but little hard evidence currently supports this concept; it remains unclear which endpoint is most useful. If bone turnover markers are used, a determination should be made before therapy is started and repeated ≥4 months after therapy is initiated. In general, a change in bone turnover markers must be 30–40% lower than the baseline to be significant because of the biologic and technical variability in these tests. A positive change in biochemical markers and/or bone density can be useful to help patients adhere to treatment regimens.

GLUCOCORTICOID-INDUCED OSTEOPOROSIS

Osteoporotic fractures are a well-characterized consequence of the hypercortisolism associated with Cushing's syndrome. However, the therapeutic use of glucocorticoids is by far the most common form of glucocorticoid-induced osteoporosis. Glucocorticoids are used widely in the treatment of a variety of disorders, including chronic lung disorders, rheumatoid arthritis, and other connective tissue diseases, inflammatory bowel disease, and after transplantation. Osteoporosis and related fractures are serious side effects of chronic glucocorticoid therapy. Because the effects of glucocorticoids on the skeleton are often superimposed on the consequences of aging and menopause, it is not surprising that women and the elderly are most frequently affected. The skeletal response to steroids is remarkably heterogeneous, however, and even young, growing individuals treated with glucocorticoids can present with fractures.

The risk of fractures depends on the dose and duration of glucocorticoid therapy, although recent data suggest that there may be no completely safe dose. Bone loss is more rapid during the early months of treatment, and trabecular bone is affected more severely than cortical bone. As a result, fractures have been shown to increase within 3 months of steroid treatment. There is an increase in fracture risk in both the axial skeleton and the appendicular skeleton, including risk of hip fracture. Bone loss can occur with any route of steroid administration, including high-dose inhaled glucocorticoids and intraarticular injections. Alternate-day delivery does not appear to ameliorate the skeletal effects of glucocorticoids.

Pathophysiology

Glucocorticoids increase bone loss by multiple mechanisms, including (1) inhibition of osteoblast function and an increase in osteoblast apoptosis, resulting in impaired synthesis of new bone; (2) stimulation of bone resorption, probably as a secondary effect; (3) impairment of the absorption of calcium across the intestine, probably by a vitamin D–independent effect; (4) increase of urinary calcium loss and perhaps induction of some degree of secondary hyperparathyroidism; (5) reduction of adrenal androgens and suppression of ovarian and testicular secretion of estrogens and androgens; and (6) induction of glucocorticoid myopathy, which may exacerbate effects on skeletal and calcium homeostasis as well as increase the risk of falls.

Evaluation of the Patient

Because of the prevalence of glucocorticoid-induced bone loss, it is important to evaluate the status of the skeleton in all patients starting or already receiving long-term glucocorticoid therapy. Modifiable risk factors should be identified, including those for falls. Examination should include testing of height and muscle strength. Laboratory evaluation should include an assessment of 24-h urinary calcium. All patients on long-term (>3 months) glucocorticoids should have measurement of bone mass at both the spine and the hip using DXA. If only one skeletal site can be measured, it is best to assess the spine in individuals <60 years and the hip in those >60 years.

Prevention

Bone loss caused by glucocorticoids can be prevented, and the risk of fractures significantly reduced. Strategies must include using the lowest dose of glucocorticoid for disease management. Topical and inhaled routes of administration are preferred, where appropriate. Risk factor reduction is important, including smoking cessation, limitation of alcohol consumption, and participation in weight-bearing exercise, when appropriate. All patients should receive an adequate calcium and vitamin D intake from the diet or from supplements.

TREATMENT GLUCOCORTICOID-INDUCED OSTEOPOROSIS

Only bisphosphonates have been demonstrated in large clinical trials to reduce the risk of fractures in patients being treated with glucocorticoids. Risedronate prevents bone loss and reduces vertebral fracture risk by ~70%. Similar beneficial effects are observed in studies of alendronate. Controlled trials of hormone therapy have shown bone-sparing effects, and calcitonin also has some protective effect in the spine. Thiazides reduce urine calcium loss, but their role in prevention of fractures is unclear. PTH has been studied in a small group of women with glucocorticoid-induced osteoporosis, among whom bone mass increased substantially, and teriparatide is being investigated in a larger multicenter trial.

FURTHER READINGS

Bischoff-Ferrari HA et al: Benefit-risk assessment of vitamin D supplementation. Osteoporos Int 21:1121, 2010 [PMID: 19957164] [Full Text]

Bolognese MA: SERMS and SERMS with estrogens for postmenopausal osteoporosis. Rev Endocr Metab Disord 11:253, 2010 [PMID: 20446043] [Full Text]

Cosman F: Parathyroid hormone treatment for osteoporosis. Curr Opin Endocrinol Diabetes Obes 15:495, 2008 [PMID: 18971677] [Full Text]

Cummings SR et al: FREEDOM Trial: Denosumab for prevention of fractures in postmenopausal women with osteoporosis. N Engl J Med 361:756, 2009 [PMID: 19671655] [Full Text]

Ettinger B et al: Updated fracture incidence rates for the US version of FRAX. Osteoporos Int 2010 21:25, 2009 [PMID: 19705048] [Full Text]

Khosla S: Update on estrogens and the skeleton. J Clin Endocrinol Metab 95:3569, 2010 [PMID: 20685883] [Full Text]

Lyles KW et al: Zolendroic acid and clinical fractures and mortality after hip fracture. N Engl J Med 357:1799, 2007 [PMID: 17878149] [Full Text]

Tucker KL: Osteoporosis prevention and nutrition. Current Curr Osteoporos Rep 7:111, 2009 [PMID: 19968914] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PAGET'S DISEASE OF BONE

Paget's disease is a localized bone-remodeling disorder that affects widespread, noncontiguous areas of the skeleton. The pathologic process is initiated by overactive osteoclastic bone resorption followed by a compensatory increase in osteoblastic new bone formation, resulting in a structurally disorganized mosaic of woven and lamellar bone. Pagetic bone is expanded, less compact, and more vascular; thus, it is more susceptible to deformities and fractures. Although most patients are asymptomatic, symptoms resulting directly from bony involvement (bone pain, secondary arthritis, fractures) or secondarily from the expansion of bone causing compression of surrounding neural tissue are not uncommon.

EPIDEMIOLOGY

There is a marked geographic variation in the frequency of Paget's disease, with high prevalence in Western Europe (Great Britain, France, and Germany, but not Switzerland or Scandinavia) and among those who have immigrated to Australia, New Zealand, South Africa, and North and South America. The disease is rare in native populations of the Americas, Africa, Asia, and the Middle East; when it does occur, the affected subjects usually have evidence of European ancestry, supporting the migration theory. For unclear reasons, the prevalence and severity of Paget's disease are decreasing, and the age of diagnosis is increasing.

The prevalence is greater in males and increases with age. Autopsy series reveal Paget's disease in about 3% of those over age 40. Prevalence of positive skeletal radiographs in patients over age 55 is 2.5% for men and 1.6% for women. Elevated alkaline phosphatase (ALP) levels in asymptomatic patients have an age-adjusted incidence of 12.7 and 7 per 100,000 person-years in men and women, respectively.

ETIOLOGY

The etiology of Paget's disease of bone remains unknown, but evidence supports both genetic and viral etiologies. A positive family history is found in 15–25% of patients and, when present, raises the prevalence of the disease seven- to tenfold among first-degree relatives.

A clear genetic basis has been established for several rare familial bone disorders that clinically and radiographically resemble Paget's disease but have more severe presentation and earlier onset. A homozygous deletion of the *TNFRSF11B* gene, which encodes osteoprotegrin **(Fig. 355-1)**, causes *juvenile Paget's disease,* also known as *familial idiopathic hypophosphatasia*, a disorder characterized by uncontrolled osteoclastic differentiation and resorption. Familial patterns of disease in several large kindred are consistent with an autosomal-dominant pattern of inheritance with variable penetrance. *Familial expansile osteolysis, expansile skeletal hyperphosphatasia, and early-onset Paget's disease* are associated with mutations in *TNFRSF11A* gene, which encodes RANK (receptor activator of nuclear factor-κB), a member of the tumor-necrosis factor superfamily critical for osteoclast differentiation (Fig. 355-1). Finally, mutations in the gene for valosin-containing protein cause a rare syndrome with autosomal-dominant inheritance and variable penetrance known as *inclusion body myopathy with Paget's disease and frontotemporal dementia (IBMPFD).* The role of genetic factors is less clear in the more common form of late-onset Paget's disease. Although a few families with mutations in the gene encoding RANK have been reported, the most common mutations identified in familial and sporadic cases of Paget's disease have been in the *SQSTM1* gene (sequestasome-1 or p62 protein) in the C-terminal ubiquitin-binding domain. The p62 protein is involved in NF–κB signaling and regulates osteoclastic differentiation. The phenotypic variability in patients with *SQSTM1* mutations suggests that additional factors, such as other genetic influences or viral infection, may influence clinical expression of the disease.

Figure 355-1

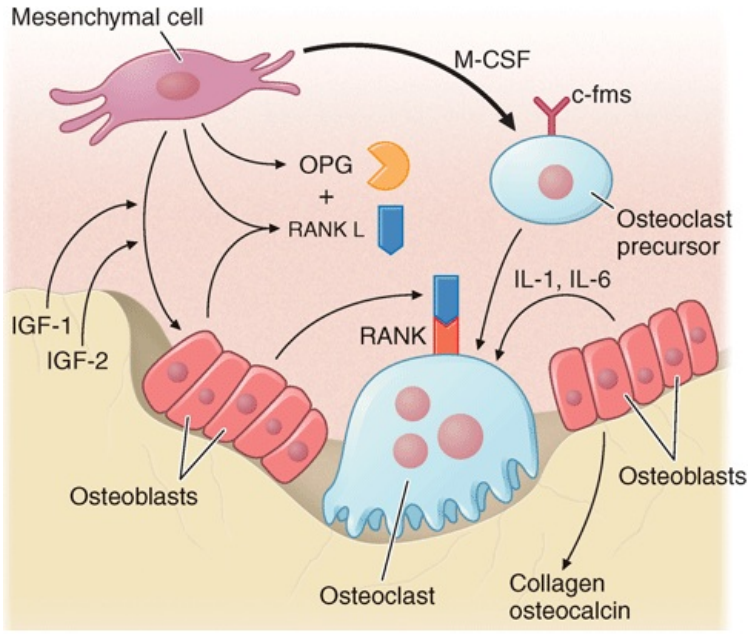

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram illustrating factors that promote differentiation and function of osteoclasts and osteoblasts and the role of the RANK pathway. Stromal bone marrow (mesenchymal) cells and differentiated osteoblasts produce multiple growth factors and cytokines, including macrophage colony-stimulating factor (M-CSF), to modulate osteoclastogenesis. RANKL (receptor activator of NFκB ligand) is produced by osteoblast progenitors and mature osteoblasts and can bind to a soluble decoy receptor known as OPG (osteoprotegerin) to inhibit RANKL action. Alternatively, a cell-cell interaction between osteoblast and osteoclast progenitors allows RANKL to bind to its membrane-bound receptor, RANK, thereby stimulating osteoclast differentiation and function. RANK binds intracellular proteins called TRAFs (tumor necrosis factor receptor–associated factors) that mediate receptor signaling through transcription factors such as NFκB. M-CSF binds to its receptor, c-fms, which is the cellular homologue of the *fms* oncogene. See text for the potential role of these pathways in disorders of osteoclast function such as Paget's disease and osteopetrosis.

Several lines of evidence suggest that a viral infection may contribute to the clinical manifestations of Paget disease, including (1) the presence of cytoplasmic and nuclear inclusions resembling paramyxoviruses (measles and respiratory syncytial virus) in pagetic osteoclasts and (2) viral mRNA in precursor and mature osteoclasts. The viral etiology is further supported by conversion of osteoclast precursors to pagetic-like osteoclasts by vectors containing the measles virus nucleocapsid or matrix genes. However, the viral etiology has been questioned by the inability to culture a live virus from pagetic bone and by failure to clone the full-length viral genes from material obtained from patients with Paget's disease.

PATHOPHYSIOLOGY

The principal abnormality in Paget's disease is the increased number and activity of osteoclasts. Pagetic osteoclasts are large, increased 10- to 100-fold in number and have a greater number of nuclei (as many as 100 compared to 3–5 nuclei in the normal osteoclast). The overactive osteoclasts may create a sevenfold increase in resorptive surfaces and an erosion rate of 9 μg/day (normal is 1 μg/day). Several causes for the increased number and activity of pagetic osteoclasts have been identified: (1) osteoclastic precursors are hypersensitive to $1,25(OH)_2D3$; (2) osteoclasts are hyperresponsive to RANK ligand (RANKL), the osteoclast stimulatory factor that mediates the effects of most osteotropic factors on osteoclast formation; (3) marrow stromal cells from pagetic lesions have increased RANKL expression; (4) osteoclast precursor recruitment is increased by interleukin (IL) 6, which is increased in the blood of patients with active Paget disease and is overexpressed in pagetic osteoclasts; (5) expression of the proto-oncogene c-*fos*, which increases osteoclastic activity, is increased; and (6) the antiapoptotic oncogene *Bcl*-2 in pagetic bone is overexpressed. Numerous osteoblasts are recruited to active resorption sites and produce large amounts of new bone matrix. As a result, bone turnover is high, and bone mass is normal or increased, not reduced, unless there is concomitant deficiency of calcium and/or vitamin D.

The characteristic feature of Paget's disease is increased bone resorption accompanied by accelerated bone formation. An initial osteolytic phase involves prominent bone resorption and marked hypervascularization. Radiographically, this manifests as an advancing lytic wedge, or "blade of grass" lesion. The second phase is a period of very active bone formation and resorption that replaces normal lamellar bone with haphazard (woven) bone. Fibrous connective tissue may replace normal bone marrow. In the final sclerotic phase, bone resorption declines progressively and leads to a hard,

dense, less vascular pagetic or mosaic bone, which represents the so-called burned-out phase of Paget's disease. All three phases may be present at the same time at different skeletal sites.

CLINICAL MANIFESTATIONS

Diagnosis is often made in asymptomatic patients because they have elevated ALP levels on routine blood chemistry testing or an abnormality on a skeletal radiograph obtained for another indication. The skeletal sites most commonly involved are the pelvis, vertebral bodies, skull, femur, and tibia. Familial cases with an early presentation often have numerous active sites of skeletal involvement.

The most common presenting symptom is pain, which may result from increased bony vascularity, expanding lytic lesions, fractures, bowing, or other deformities. Bowing of the femur or tibia causes gait abnormalities and abnormal mechanical stresses with secondary osteoarthritis of the hip or knee joints. Long bone bowing also causes extremity pain by stretching the muscles attached to the bone softened by the pagetic process. Back pain results from enlarged pagetic vertebrae, vertebral compression fractures, spinal stenosis, degenerative changes of the joints, and altered body mechanics with kyphosis and forward tilt of the upper back. Rarely, spinal cord compression may result from bone enlargement or from the vascular steal syndrome. Skull involvement may cause headaches, symmetric or asymmetric enlargement of the parietal or frontal bones (frontal bossing), and increased head size. Cranial expansion may narrow cranial foramens and cause neurologic complications including hearing loss from cochlear nerve damage from temporal bone involvement, cranial nerve palsies, and softening of the base of the skull (*platybasia*) with the risk of brainstem compression. Pagetic involvement of the facial bones may cause facial deformity; loss of teeth and other dental conditions; and, rarely, airway compression.

Fractures are serious complications of Paget's disease and usually occur in long bones at areas of active or advancing lytic lesions. Common fracture sites are the femoral shaft and subtrochanteric regions. Neoplasms arising from pagetic bone are rare. The incidence of sarcoma appears to be decreasing, possibly because of earlier, more effective treatment with potent antiresorptive agents. The majority of tumors are osteosarcomas, which usually present with new pain in a long-standing pagetic lesion. Osteoclast-rich benign giant cell tumors may arise in areas adjacent to pagetic bone, and they respond to glucocorticoid therapy.

Cardiovascular complications may occur in patients with involvement of large (15–35%) portions of the skeleton and a high degree of disease activity (ALP four times above normal). The extensive arteriovenous shunting and marked increases in blood flow through the vascular pagetic bone lead to a high-output state and cardiac enlargement. However, high-output heart failure is relatively rare and usually develops in patients with concomitant cardiac pathology. In addition, calcific aortic stenosis and diffuse vascular calcifications have been associated with Paget's disease.

DIAGNOSIS

The diagnosis may be suggested on clinical examination by the presence of an enlarged skull with frontal bossing, bowing of an extremity, or short stature with simian posturing. An extremity with an area of warmth and tenderness to palpation may suggest an underlying pagetic lesion. Other findings include bony deformity of the pelvis, skull, spine, and extremities; arthritic involvement of the joints adjacent to lesions; and leg-length discrepancy resulting from deformities of the long bones.

Paget's disease is usually diagnosed from radiologic and biochemical abnormalities. Radiographic findings typical of Paget's disease include enlargement or expansion of an entire bone or area of a long bone, cortical thickening, coarsening of trabecular markings, and typical lytic and sclerotic changes. Skull radiographs **(Fig. 355-2)** reveal regions of "cotton wool," or osteoporosis circumscripta, thickening of diploic areas, and enlargement and sclerosis of a portion or all of one or more skull bones. Vertebral cortical thickening of the superior and inferior end plates creates a "picture frame" vertebra. Diffuse radiodense enlargement of a vertebra is referred to as "ivory vertebra." Pelvic radiographs may demonstrate disruption or fusion of the sacroiliac joints; porotic and radiodense lesions of the ilium with whorls of coarse trabeculation; thickened and sclerotic iliopectineal line (brim sign); and softening with protrusio acetabuli, with axial migration of the hips and functional flexion contracture. Radiographs of long bones reveal bowing deformity and typical pagetic changes of cortical thickening and expansion and areas of lucency and sclerosis **(Fig. 355-3)**. Radionuclide ^{99m}Tc bone scans are less specific but are more sensitive than standard radiographs for identifying sites of active skeletal lesions. Although CT and MRI studies are not necessary in most cases, CT may be useful for the assessment of possible fracture and MRI is necessary to assess the possibility of sarcoma, giant cell tumor, or metastatic disease in pagetic bone. Definitive diagnosis of malignancy often requires bone biopsy.

Figure 355-2

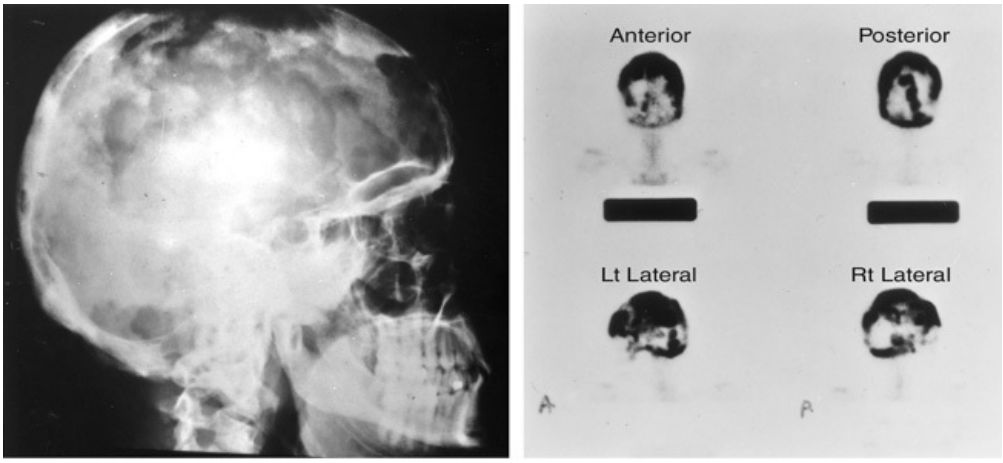

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A 48-year-old woman with Paget's disease of the skull. ***Left.*** Lateral radiograph showing areas of both bone resorption and sclerosis. ***Right.*** ^{99m}Tc HDP bone scan with anterior, posterior, and lateral views of the skull showing diffuse isotope uptake by the frontal, parietal, occipital, and petrous bones.

Figure 355-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiograph of a 73-year-old man with Paget's disease of the right proximal femur. Note the coarsening of the trabecular pattern with marked cortical thickening and narrowing of the joint space consistent with osteoarthritis secondary to pagetic deformity of the right femur.

Biochemical evaluation is useful in the diagnosis and management of Paget disease. The marked increase in bone turnover can be monitored using biochemical markers of bone formation and resorption. The parallel rise in markers of bone formation and resorption confirms the coupling of bone formation and resorption in Paget's disease. The degree of bone-marker elevation reflects the extent and severity of the disease. Patients with the highest elevation of ALP (10 times the upper limit of normal) typically have involvement of the skull and at least one other skeletal site. Lower values suggest less extensive involvement or a quiescent phase of the disease. For most patients, serum total ALP remains the test of choice both for diagnosis and assessing response to therapy. Occasionally, a symptomatic patient with evidence of progression at a single site may have a normal total ALP level but increased bone-specific ALP. For unclear reasons, serum osteocalcin, another marker of bone formation, is not always elevated and is not recommended for use in diagnosis or management of Paget's disease. Bone-resorption markers (serum or urine N-telopeptide or C-telopeptide measured in the blood or urine) are also elevated in active Paget's disease and decrease more rapidly in response to therapy than does ALP.

Serum calcium and phosphate levels are normal in Paget's disease. Immobilization of a patient with active Paget's disease may rarely cause hypercalcemia and hypercalciuria and increase the risk for nephrolithiasis. However, the discovery of hypercalcemia, even in the presence of immobilization, should prompt a search for another cause of hypercalcemia. In contrast, hypocalcemia or mild secondary hyperparathyroidism may develop in Paget patients with very active bone formation and insufficient dietary calcium intake, particularly during bisphosphonate therapy when bone resorption is rapidly suppressed and active bone formation continues. Therefore, adequate calcium and vitamin D intake should be instituted prior to administration of bisphosphonates.

TREATMENT: PAGET'S DISEASE OF BONE

The development of effective and potent pharmacologic agents **(Table 355-1)** has changed the treatment philosophy from treating only symptomatic patients to treating asymptomatic patients who are at risk for complications. Pharmacologic therapy is indicated in the following circumstances: to control symptoms caused by metabolically active Paget's disease such as bone pain, fracture, headache, pain from pagetic radiculopathy or arthropathy, or neurologic complications; to decrease local blood flow and minimize operative blood loss in patients who need surgery at an active pagetic site; to reduce hypercalciuria that may occur during immobilization; and to decrease the risk of complications when disease activity is high (elevated ALP) and when the site of involvement involves weight-bearing bones, areas adjacent to major joints, vertebral bodies, and the skull. Whether or not early therapy prevents late complications remains to be determined. A recent randomized study of over 1200 patients from the UK showed no difference in bone pain, fracture rates, quality of life, and hearing loss between patients who received pharmacologic therapy to control symptoms (bone pain) and those receiving bisphosphonates to normalize serum ALP. However, the most potent agent (zoledronic acid) was not used and the duration of observation (mean of 3 years with a range of 2 to 5 years) may not be long enough to assess the impact of treatment on long-term outcomes. It seems likely that the restoration of normal bone architecture following suppression of pagetic activity will prevent further deformities and complications.

Table 355-1 Pharmacologic Agents Approved for Treatment of Paget's Disease

Name (Brand)	Dose and Mode of Delivery	Normalization of ALP
Zoledronate (Zometa)	5 mg IV over 15 min	90% of patients at 6 mo
Pamidronate (Aredia)	30 mg IV/d over 4 h on 3 days	~50% of patients
Risedronate (Actonel)	30 mg PO/d for 2 mo	73% of patients
Alendronate (Fosamax)	40 mg PO/d for 6 mo	63% of patients
Tiludronate (Skelid)	800 mg PO daily for 3 mo	35% of patients
Etidronate (Didronel)	200–400 mg PO/d x 6 mo	15% of patients
Calcitonin (Miacalcin)	100 U SC daily for 6–18 mo (may reduce to 50 U 3 times per wk)	(Reduction of ALP by to 50%)

Agents approved for treatment of Paget's disease suppress the very high rates of bone resorption and secondarily decrease the high rates of bone formation **(Table 355-1)**. As a result of decreasing bone turnover, pagetic structural patterns, including areas of poorly mineralized woven bone, are replaced by more normal cancellous or lamellar bone. Reduced bone turnover can be documented by a decline in serum ALP and urine or serum resorption markers (N-telopeptide, C-telopeptide).

The first clinically useful agent, etidronate, is now rarely used because the doses required to suppress bone resorption may impair mineralization, necessitating that the drug be given for a maximum of 6 months followed by a 6-month drug-free period. The second-generation oral bisphosphonates—tiludronate, alendronate, and risedronate—are more potent than etidronate in controlling bone turnover and, thus, induce a longer remission at a lower dose. The lower doses reduce the risks of impaired mineralization and osteomalacia. Oral bisphosphonates should be taken first thing in the morning on an empty stomach, followed by maintenance of upright posture with no food, drink, or other medications for 30–60 minutes. The efficacy of different agents, based on their ability to normalize or decrease ALP levels, is summarized in **Table 355-1**, although the response rates are not comparable because they are obtained from different studies.

Intravenous bisphosphonates approved for Paget's disease include pamidronate and zoledronic acid. Although the recommended dose for pamidronate is 30 mg dissolved in 500 mL of normal saline or dextrose IV over 4 h on three consecutive days, a more commonly used simpler regimen is a single infusion of 60–90 mg in patients with mild elevation of serum ALP and multiple 90-mg infusions in

those with higher levels of ALP. In many patients, particularly those who have severe disease or need rapid normalization of bone turnover (neurologic symptoms, severe bone pain due to a lytic lesion, risk of an impending fracture, or pretreatment prior to elective surgery in an area of active disease), treatment with zoledronic acid is the first choice. It normalizes ALP in about 90% of patients by 6 months, and the therapeutic effect persists for at least 6 more months in most patients. About 10–20% of patients experience a flulike syndrome after the first infusion, which can be partly ameliorated by pretreatment with acetaminophen or NSAIDs. In patients with high bone turnover, vitamin D (400–800 IU daily) and calcium (500 mg three times daily) should be provided to prevent hypocalcemia and secondary hyperparathyroidism. Remission following treatment with IV bisphosphonates, particularly zoledronic acid, may persist for well over 1 year. Bisphosphonates should not be used in patients with renal insufficiency (glomerular filtration rate <35ml/min).

The subcutaneous injectable form of salmon calcitonin is approved for the treatment of Paget's disease. Intranasal calcitonin spray is approved for osteoporosis at a dose of 200 U/d; however, the efficacy of this dose in Paget's disease has not been thoroughly studied. The usual starting dose of injectable calcitonin (100 U/d) reduces ALP by 50% and may relieve skeletal symptoms. The dose may be reduced to 50 U/d three times weekly after an initial favorable response to 100 U daily; however, the lower dose may require long-term use to sustain efficacy. The common side effects of calcitonin therapy are nausea and facial flushing. Secondary resistance after prolonged use may be due to either the formation of anticalcitonin antibodies or downregulation of osteoclastic cell–surface calcitonin receptors. The lower potency and injectable mode of delivery make this agent a less-attractive treatment option that should be reserved for patients who either do not tolerate bisphosphonates or have a contraindication to their use.

SCLEROSING BONE DISORDERS

Osteopetrosis

Osteopetrosis refers to a group of disorders caused by severe impairment of osteoclast-mediated bone resorption. Other terms that are often used include marble bone disease, which captures the solid x-ray appearance of the involved skeleton, and Albers-Schonberg disease, which refers to the milder, adult form of osteopetrosis also known as autosomal-dominant osteopetrosis type II. The major types of osteopetrosis include malignant (severe, infantile, autosomal recessive) osteopetrosis and benign (adult, autosomal dominant) osteopetrosis types I and II. A rare autosomal recessive intermediate form has a more benign prognosis. Autosomal recessive carbonic anhydrase (CA) II deficiency produces osteopetrosis of intermediate severity associated with renal tubular acidosis and cerebral calcification.

ETIOLOGY AND GENETICS

Naturally occurring and gene-knockout animal models with phenotypes similar to those of the human disorders have been used to explore the genetic basis of osteopetrosis. The primary defect in osteopetrosis is the loss of osteoclastic bone resorption and preservation of normal osteoblastic bone formation. Osteoprotegerin (OPG) is a soluble decoy receptor that binds osteoblast-derived RANK ligand, which mediates osteoclast differentiation and activation (Fig. 355-1). Transgenic mice that overexpress OPG develop osteopetrosis, presumably by blocking RANK ligand. Mice deficient in RANK lack osteoclasts and develop severe osteopetrosis.

Recessive mutations of CA II prevent osteoclasts from generating an acid environment in the clear zone between its ruffled border and the adjacent mineral surface. Absence of CA II, therefore, impairs osteoclastic bone resorption. Other forms of human disease have less clear genetic defects. About one-half of the patients with malignant infantile osteopetrosis have a mutation in the *TCIRG1* gene encoding the osteoclast-specific subunit of the vacuolar proton pump, which mediates the acidification of the interface between bone mineral and the osteoclast ruffled border. Mutations in the *CICN7* chloride channel gene cause autosomal-dominant osteopetrosis type II.

CLINICAL PRESENTATION

The incidence of autosomal-recessive severe (malignant) osteopetrosis ranges from 1 in 200,000 to 1 in 500,000 live births. As bone and cartilage fail to undergo modeling, paralysis of one or more cranial nerves may occur due to narrowing of the cranial foramens. Failure of skeletal modeling also results in inadequate marrow space, leading to extramedullary hematopoiesis with hypersplenism and pancytopenia. Hypocalcemia due to lack of osteoclastic bone resorption may occur in infants and young children. The untreated infantile disease is fatal, often before age five.

Adult (benign) osteopetrosis is an autosomal-dominant disease that is usually diagnosed by the discovery of typical skeletal changes in young adults who undergo radiologic evaluation of a fracture. The prevalence is 1 in 100,000 to 1 in 500,000 adults. The course is not always benign, because fractures may be accompanied by loss of vision, deafness, psychomotor delay, mandibular osteomyelitis, and other complications usually associated with the juvenile form. In some kindred, nonpenetrance results in skip generations, while in other families, severely affected children are born into families with benign disease. The milder form of the disease does not usually require treatment.

RADIOGRAPHY

Typically, there are generalized symmetric increases in bone mass with thickening of both cortical and trabecular bone. Diaphyses and metaphyses are broadened, and alternating sclerotic and lucent bands may be seen in the iliac crests, at the ends of long bones, and in vertebral bodies. The cranium is usually thickened, particularly at the base of the skull, and the paranasal and mastoid sinuses are underpneumatized.

LABORATORY FINDINGS

The only significant laboratory findings are elevated serum levels of osteoclast-derived tartrate-resistant acid phosphatase (TRAP) and the brain isoenzyme of creatine kinase. Serum calcium may be low in severe disease, and parathyroid hormone and 1,25-dihydroxyvitamin D levels may be elevated in response to hypocalcemia.

Treatment: Osteopetrosis

Allogeneic HLA-identical bone marrow transplantation has been successful in some children. Following transplantation, the marrow contains progenitor cells and normally functioning osteoclasts. A cure is most likely when children are transplanted before age four. Marrow transplantation from nonidentical HLA-matched donors has a much higher failure rate. Limited studies in small numbers of patients have suggested variable benefits following treatment with interferon γ-1β, 1,25- dihydroxyvitamin D (which stimulates osteoclasts directly), methylprednisolone, and a low-calcium/high-phosphate diet.

Surgical intervention is indicated to decompress optic or auditory nerve compression. Orthopedic management is required for the surgical treatment of fractures and their complications including malunion and postfracture deformity.

Pyknodysostosis

This is an autosomal-recessive form of osteosclerosis that is believed to have affected the French impressionist painter Henri de Toulouse-Lautrec. The molecular basis involves mutations in the gene that encodes cathepsin K, a lysosomal metalloproteinase highly expressed in osteoclasts and important for bone-matrix degradation. Osteoclasts are present but do not function normally. Pyknodysostosis is a form of short-limb dwarfism that presents with frequent fractures but usually a normal life span. Clinical features include short stature; kyphoscoliosis and deformities of the chest; high arched palate; proptosis; blue sclerae; dysmorphic features including small face and chin, frontooccipital prominence, pointed beaked nose, large cranium, and obtuse mandibular angle; and small, square hands with hypoplastic nails. Radiographs demonstrate a generalized increase in bone density, but in contrast to osteopetrosis, the long bones are normally shaped. Separated cranial sutures, including the persistent patency of the anterior fontanel, are characteristic of the disorder. There may also be hypoplasia of the sinuses, mandible, distal clavicles, and terminal phalanges. Persistence of deciduous teeth and sclerosis of the calvarium and base of the skull are also common. Histologic evaluation shows normal cortical bone architecture with decreased osteoblastic and osteoclastic activities. Serum chemistries are normal, and, unlike osteopetrosis, there is no anemia. There is no known treatment for this condition, and there are no reports of attempted bone marrow transplant.

Progressive Diaphyseal Dysplasia

Also known as *Camurati-Engelmann disease*, progressive diaphyseal dysplasia is an autosomal-dominant disorder that is characterized radiographically by diaphyseal hyperostosis and a symmetric thickening and increased diameter of the endosteal and periosteal surfaces of the diaphyses of the long bones, particularly the femur and tibia, and, less often, the fibula, radius, and ulna. The genetic defect responsible for the disease has been localized to the area of chromosome 19q13.2 encoding tumor growth factor (TGF)-β1. The mutation promotes activation of TGF-β1. The clinical severity is variable. The most common presenting symptoms are pain and tenderness of the involved areas, fatigue, muscle wasting, and gait disturbance. The weakness may be mistaken for muscular dystrophy. Characteristic body habitus includes thin limbs with little muscle mass yet prominent and palpable bones and, when the skull is involved, large head with prominent forehead and proptosis. Patients may also display signs of cranial nerve palsies, hydrocephalus, central hypogonadism, and Raynaud's phenomenon. Radiographically, patchy progressive endosteal and periosteal new bone formation is observed along the diaphyses of the long bones. Bone scintigraphy shows increased radiotracer uptake in involved areas.

Treatment with low-dose glucocorticoids relieves bone pain and may reverse the abnormal bone formation. Intermittent bisphosphonate therapy has produced clinical improvement in a limited number of patients.

Hyperostosis Corticalis Generalisata

This is also known as *van Buchem's disease*; it is an autosomal-recessive disorder characterized by endosteal hyperostosis in which osteosclerosis involves the skull, mandible, clavicles, and ribs. The major manifestations are due to narrowed cranial foramens with neural compressions that may result in

optic atrophy, facial paralysis, and deafness. Adults may have an enlarged mandible. Serum ALP levels may be elevated, which reflect the uncoupled bone remodeling with high osteoblastic formation rates and low osteoclastic resorption. As a result, there is increased accumulation of normal bone. Endosteal hyperostosis with syndactyly, known as *sclerosteosis*, is a more severe form. The genetic defects for both sclerosteosis and van Buchem's disease have been assigned to the same region of the chromosome 17q12-q21. It is possible that both conditions may have deactivating mutations in the *BEER* (bone-expressed equilibrium regulator) gene.

Melorheostosis

Melorheostosis (Greek, "flowing hyperostosis") may occur sporadically or follow a pattern consistent with an autosomal-recessive disorder. The major manifestation is progressive linear hyperostosis in one or more bones of one limb, usually a lower extremity. The name comes from the radiographic appearance of the involved bone, which resembles melted wax that has dripped down a candle. Symptoms appear during childhood as pain or stiffness in the area of sclerotic bone. There may be associated ectopic soft tissue masses, composed of cartilage or osseous tissue, and skin changes overlying the involved bone, consisting of scleroderma-like areas and hypertrichosis. The disease does not progress in adults, but pain and stiffness may persist. Laboratory tests are unremarkable. No specific etiology is known. There is no specific treatment. Surgical interventions to correct contractures are often unsuccessful.

Osteopoikilosis

The literal translation of osteopoikilosis is "spotted bones"; it is a benign autosomal-dominant condition in which numerous small, variably shaped (usually round or oval) foci of bony sclerosis are seen in the epiphyses and adjacent metaphyses. The lesions may involve any bone except the skull, ribs, and vertebrae. They may be misidentified as metastatic lesions. The main differentiating points are that bony lesions of osteopoikilosis are stable over time and do not accumulate radionucleotide on bone scanning. In some kindred, osteopoikilosis is associated with connective tissue nevi known as *dermatofibrosis lenticularis disseminata*, also known as *Buschke-Ollendorff syndrome*. Histologic inspection reveals thickened but otherwise normal trabeculae and islands of normal cortical bone. No treatment is indicated.

Hepatitis C–Associated Osteosclerosis

Hepatitis C–associated osteosclerosis (HCAO) is a rare acquired diffuse osteosclerosis in adults with prior hepatitis C infection. After a latent period of several years, patients develop diffuse appendicular bone pain and a generalized increase in bone mass with elevated serum ALP. Bone biopsy and histomorphometry reveal increased rates of bone formation, decreased bone resorption with a marked decrease in osteoclasts, and dense lamellar bone. One patient had increased serum OPG levels, and bone biopsy showed large numbers of osteoblasts positive for OPG and reduced osteoclast number. Empirical therapy includes pain control, and there may be beneficial response to either calcitonin or bisphosphonate.

DISORDERS ASSOCIATED WITH DEFECTIVE MINERALIZATION

Hypophosphatasia

This is a rare inherited disorder that presents as rickets in infants and children or osteomalacia in adults with paradoxically low serum levels of ALP. The frequency of the severe neonatal and infantile forms is about 1 in 100,000 live births in Canada, where the disease is most common because of its high prevalence among Mennonites and Hutterites. It is rare in African Americans. The severity of the disease is remarkably variable, ranging from intrauterine death associated with profound skeletal hypomineralization at one extreme to premature tooth loss as the only manifestation in some adults. Severe cases are inherited in an autosomal-recessive manner, but the genetic patterns are less clear for the milder forms. The disease is caused by a deficiency of tissue nonspecific (bone/liver/kidney) ALP (TNSALP), which, although ubiquitous, results only in bone abnormalities. Protein levels and functions of the other ALP isozymes (germ cell, intestinal, placental) are normal. Defective ALP permits accumulation of its major naturally occurring substrates including phosphoethanolamine (PEA), inorganic pyrophosphate (PPi), and pyridoxal 5′-phosphate (PLP). The accumulation of PPi interferes with mineralization through its action as a potent inhibitor of hydroxyapatite crystal growth.

Perinatal hypophosphatasia becomes manifest during pregnancy and is often complicated by polyhydramnios and intrauterine death. The infantile form becomes clinically apparent before the age of 6 months with failure to thrive, rachitic deformities, functional craniosynostosis despite widely open fontanels (which are actually hypomineralized areas of the calvarium), raised intracranial pressure, and flail chest and predisposition to pneumonia. Hypercalcemia and hypercalciuria are common. This form has a mortality rate of about 50%. Prognosis seems to improve for the children who survive infancy. Childhood hypophosphatasia has variable clinical presentation. Premature loss of deciduous teeth (before age five) is the hallmark of the disease. Rickets causes delayed walking with waddling gait, short stature, and dolichocephalic skull with frontal bossing. The disease often improves during puberty but may recur in adult life. Adult hypophosphatasia presents during middle age with painful,

poorly healing metatarsal stress fractures or thigh pain due to femoral pseudofractures.

Laboratory investigation reveals low ALP levels and normal or elevated levels of serum calcium and phosphorus despite clinical and radiologic evidence of rickets or osteomalacia. Serum parathyroid hormone, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D levels are normal. The elevation of PLP is specific for the disease and may even be present in asymptomatic parents of severely affected children. Because vitamin B6 increases PLP levels, vitamin B_6 supplements should be discontinued one week before testing.

There is no established medical therapy. In contrast to other forms of rickets and osteomalacia, calcium and vitamin D supplementation should be avoided because they may aggravate hypercalcemia and hypercalciuria. A low-calcium diet, glucocorticoids, and calcitonin have been used in a small number of patients with variable responses. Because fracture healing is poor, placement of intramedullary rods is best for acute fracture repair and for prophylactic prevention of fractures.

Axial Osteomalacia

This is a rare disorder characterized by defective skeletal mineralization despite normal serum calcium and phosphate levels. Clinically, the disorder presents in middle-aged or elderly men with chronic axial skeletal discomfort. Cervical spine pain may also be present. Radiographic findings are mainly osteosclerosis due to coarsened trabecular patterns typical of osteomalacia. Spine, pelvis, and ribs are most commonly affected. Histologic changes show defective mineralization and flat, inactive osteoblasts. The primary defect appears to be an acquired defect in osteoblast function. The course is benign, and there is no established treatment. Calcium and vitamin D therapies are not effective.

Fibrogenesis Imperfecta Ossium

This is a rare condition of unknown etiology. It presents in both sexes; in middle age or later; with progressive, intractable skeletal pain and fractures; worsening immobilization; and a debilitating course. Radiographic evaluation reveals generalized osteomalacia, osteopenia, and occasional pseudofractures. Histologic features include a tangled pattern of collagen fibrils with abundant osteoblasts and osteoclasts. There is no effective treatment. Spontaneous remission has been reported in a small number of patients. Calcium and vitamin D have not been beneficial.

FIBROUS DYSPLASIA AND MCCUNE-ALBRIGHT SYNDROME

Fibrous dysplasia is a sporadic disorder characterized by the presence of one (monostotic) or more (polyostotic) expanding fibrous skeletal lesions composed of bone-forming mesenchyme. The association of the polyostotic form with café-au-lait spots and hyperfunction of an endocrine system such as pseudo-precocious puberty of ovarian origin is known as *McCune-Albright syndrome* (MAS). A spectrum of the phenotypes is caused by activating mutations in the *GNAS1* gene, which encodes the α subunit of the stimulatory G protein ($G_S\alpha$). As the postzygotic mutations occur at different stages of early development, the extent and type of tissue affected are variable and explain the mosaic pattern of skin and bone changes. GTP binding activates the $G_S\alpha$ regulatory protein and mutations in regions of $G_S\alpha$ that selectively inhibit GTPase activity, which results in constitutive stimulation of the cyclic AMP–protein kinase A signal transduction pathway. Such mutations of the $G_S\alpha$ protein–coupled receptor may cause autonomous function in bone (parathyroid hormone receptor); skin (melanocyte-stimulating hormone receptor); and various endocrine glands including ovary (follicle-stimulating hormone receptor), thyroid (thyroid-stimulating hormone receptor), adrenal (adrenocorticotropic hormone receptor), and pituitary (growth hormone–releasing hormone receptor). The skeletal lesions are composed largely of mesenchymal cells that do not differentiate into osteoblasts, resulting in the formation of imperfect bone. In some areas of bone, fibroblast-like cells develop features of osteoblasts in that they produce extracellular matrix that organizes into woven bone. Calcification may occur in some areas. In other areas, cells have features of chondrocytes and produce cartilage-like extracellular matrix.

CLINICAL PRESENTATION

Fibrous dysplasia occurs with equal frequency in both sexes, whereas MAS with precocious puberty is more common (10:1) in girls. The monostotic form is the most common and is usually diagnosed in patients between 20 and 30 years of age without associated skin lesions. The polyostotic form typically manifests in children <10 years old and may progress with age. Early-onset disease is generally more severe. Lesions may become quiescent in puberty and progress during pregnancy or with estrogen therapy. In polyostotic fibrous dysplasia, the lesions most commonly involve the maxilla and other craniofacial bones, ribs, and metaphyseal or diaphyseal portions of the proximal femur or tibia. Expanding bone lesions may cause pain, deformity, fractures, and nerve entrapment. Sarcomatous degeneration involving the facial bones or femur is infrequent (<1%). The risk of malignant transformation is increased by radiation, which has proven to be ineffective treatment. In rare patients with widespread lesions, renal phosphate wasting and hypophosphatemia may cause rickets or osteomalacia. Hypophosphatemia may be due to production of a phosphaturic factor by the abnormal fibrous tissue.

MAS patients may have café-au-lait spots, which are flat, hyperpigmented skin lesions that have rough borders ("coast of Maine") in contrast to the café-au-lait lesions of neurofibromatosis that have smooth borders ("coast of California"). The most common endocrinopathy is isosexual pseudoprecocious puberty in girls. Other less common endocrine disorders include thyrotoxicosis, Cushing's syndrome, acromegaly, hyperparathyroidism, hyperprolactinemia, and pseudoprecocious puberty in boys.

RADIOGRAPHIC FINDINGS

In long bones, the fibrous dysplastic lesions are typically well-defined, radiolucent areas with thin cortices and a ground-glass appearance. Lesions may be lobulated with trabeculated areas of radiolucency **(Fig. 355-4)**. Involvement of facial bones usually presents as radiodense lesions, which may create a leonine appearance (leontiasis osea). Expansile cranial lesions may narrow foramens and cause optic lesions, reduce hearing, and create other manifestations of cranial nerve compression.

Figure 355-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Radiograph of a 16-year-old male with fibrous dysplasia of the right proximal femur. Note the multiple cystic lesions, including the large lucent lesion in the proximal midshaft with scalloping of the interior surface. The femoral neck contains two lucent cystic lesions.

LABORATORY RESULTS

Serum ALP is occasionally elevated but calcium, parathyroid hormone, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D levels are normal. Patients with extensive polyostotic lesions may have hypophosphatemia, hyperphosphaturia, and osteomalacia. The hypophosphatemia and phosphaturia are directly related to the levels of fibroblast growth factor 23 (FGF23). Biochemical markers of bone turnover may be elevated.

Treatment: Fibrous Dysplasia and Mccune-Albright Syndrome

Spontaneous healing of the lesions does not occur, and there is no established effective treatment. Improvement in bone pain and partial or complete resolution of radiographic lesions have been reported after IV bisphosphonate therapy. Surgical stabilization is used to prevent pathologic fracture or destruction of a major joint space and to relieve nerve root or cranial nerve compression or sinus obstruction.

OTHER DYSPLASIAS OF BONE AND CARTILAGE

Pachydermoperiostosis

Pachydermoperiostosis, or hypertrophic osteoarthropathy (primary or idiopathic), is an autosomal-dominant disorder characterized by periosteal new bone formation that involves the distal extremities. The lesions present as clubbing of the digits and hyperhydrosis and thickening of the skin, primarily of

the face and forehead. The changes usually appear during adolescence, progress over the next decade, and then become quiescent. During the active phase, progressive enlargement of the hands and feet produces a pawlike appearance, which may be mistaken for acromegaly. Arthralgias, pseudogout, and limited mobility may also occur. The disorder must be differentiated from secondary hypertrophic osteopathy that develops during the course of serious pulmonary disorders. The two conditions can be differentiated by standard radiography of the digits in which secondary pachydermoperiostosis has exuberant periosteal new bone formation and a smooth and undulating surface. In contrast, primary hypertrophic osteopathy has an irregular periosteal surface.

There are no diagnostic blood or urine tests. Synovial fluid does not have an inflammatory profile. There is no specific therapy, although a limited experience with colchicine suggests some benefit in controlling the arthralgias.

Osteochondrodysplasias

These include several hundred heritable disorders of connective tissue. These primary abnormalities of cartilage manifest as disturbances in cartilage and bone growth. Selected growth-plate chondrodysplasias are described here. **For discussion of chondrodysplasias, see Chap. 363.**

ACHONDRODYSPLASIA

This is a relatively common form of short-limb dwarfism that occurs in 1 in 15,000 to 1 in 40,000 live births. The disease is caused by a mutation of the fibroblast growth factor receptor 3 (*FGFR3*) gene that results in a gain-of-function state. Most cases are sporadic mutations. However, when the disorder appears in families, the inheritance pattern is consistent with an autosomal-dominant disorder. The primary defect is abnormal chondrocyte proliferation at the growth plate that causes development of short, but proportionately thick, long bones. Other regions of the long bones may be relatively unaffected. The disorder is manifest by the presence of short limbs (particularly the proximal portions), normal trunk, large head, saddle nose, and an exaggerated lumbar lordosis. Severe spinal deformity may lead to cord compression. The homozygous disorder is more serious than the sporadic form and may cause neonatal death. Pseudoachondroplasia clinically resembles achondrodysplasia but has no skull abnormalities.

ENCHONDROMATOSIS

This is also called *dyschondroplasia* or *Ollier's disease*; it is also a disorder of the growth plate in which the primary cartilage is not resorbed. Cartilage ossification proceeds normally, but it is not resorbed normally, leading to cartilage accumulation. The changes are most marked at the ends of long bones, where the highest growth rates occur. Chondrosarcoma develops infrequently. The association of enchondromatosis and cavernous hemangiomas of the skin and soft tissues is known as *Maffucci syndrome*. Both Ollier's disease and Maffucci syndrome are associated with various malignancies, including granulosa cell tumor of the ovary and cerebral glioma.

MULTIPLE EXOSTOSES

This is also called *diaphyseal aclasis* or *osteochondromatosis*; it is a genetic disorder that follows an autosomal-dominant pattern of inheritance. In this condition, areas of growth plates become displaced, presumably by growing through a defect in the perichondrium. The lesion begins with vascular invasion of the growth-plate cartilage, resulting in a characteristic radiographic finding of a mass that is in direct communication with the marrow cavity of the parent bone. The underlying cortex is resorbed. The disease is caused by inactivating mutations of the *EXT1* and *EXT2* genes, whose products normally regulate processing of chondrocyte cytoskeletal proteins. The products of the *EXT* gene likely function as tumor suppressors, with the loss-of-function mutation resulting in abnormal proliferation of growth-plate cartilage. Solitary or multiple lesions are located in the metaphyses of long bones. Although usually asymptomatic, the lesions may interfere with joint or tendon function or compress peripheral nerves. The lesions stop growing when growth ceases but may recur during pregnancy. There is a small risk for malignant transformation into chondrosarcoma.

EXTRASKELETAL (ECTOPIC) CALCIFICATION AND OSSIFICATION

Deposition of calcium phosphate crystals (*calcification*) or formation of true bone (*ossification*) in nonosseous soft tissue may occur by one of three mechanisms: (1) metastatic calcification due to a supranormal calcium &multi; phosphate concentration product in extracellular fluid; (2) dystrophic calcification due to mineral deposition into metabolically impaired or dead tissue despite normal serum levels of calcium and phosphate; and (3) ectopic ossification, or true bone formation. Disorders that may cause extraskeletal calcification or ossification are listed in **Table 355-2**.

Table 355-2 Diseases and Conditions Associated with Ectopic Calcification and Ossification

Metastatic calcification
Hypercalcemic states
Primary hyperparathyroidism

Sarcoidosis Vitamin D intoxication Milk-alkali syndrome Renal failure Hyperphosphatemia Tumoral calcinosis Secondary hyperparathyroidism Pseudohypoparathyroidism Renal failure Hemodialysis Cell lysis following chemotherapy Therapy with vitamin D and phosphate
Dystrophic calcification Inflammatory disorders Scleroderma Dermatomyositis Systemic lupus erythematosus Trauma-induced
Ectopic ossification Myositis ossificans Postsurgery Burns Neurologic injury Other trauma Fibrodysplasia ossificans progressiva

Metastatic Calcification

Soft tissue calcification may complicate diseases associated with significant hypercalcemia, hyperphosphatemia, or both. In addition, vitamin D and phosphate treatments or calcium administration in the presence of mild hyperphosphatemia, such as during hemodialysis, may induce ectopic calcification. Calcium phosphate precipitation may complicate any disorder when the serum calcium &multi; phosphate concentration product is >75. The initial calcium phosphate deposition is in the form of small, poorly organized crystals, which subsequently organize into hydroxyapatite crystals. Calcifications that occur in hypercalcemic states with normal or low phosphate have a predilection for kidney, lungs, and gastric mucosa. Hyperphosphatemia with normal or low serum calcium may promote soft tissue calcification with predilection for the kidney and arteries. The disturbances of calcium and phosphate in renal failure and hemodialysis are common causes of soft tissue (metastatic) calcification.

Tumoral Calcinosis

This is a rare genetic disorder characterized by masses of metastatic calcifications in soft tissues around major joints, most often shoulders, hips, and ankles. Tumoral calcinosis differs from other disorders in that the periarticular masses contain hydroxyapatite crystals or amorphous calcium phosphate complexes, while in fibrodysplasia ossificans progressiva (below), true bone is formed in soft tissues. About one-third of tumoral calcinosis cases are familial, with both autosomal-recessive and autosomal-dominant modes of inheritance reported. The disease is also associated with a variably expressed abnormality of dentition marked by short bulbous roots, pulp calcification, and radicular dentin deposited in swirls. The primary defect responsible for the metastatic calcification appears to be hyperphosphatemia resulting from the increased capacity of the renal tubule to reabsorb filtered phosphate. Spontaneous soft tissue calcification is related to the elevated serum phosphate, which, along with normal serum calcium, exceeds the concentration product of 75.

All of the North American patients reported have been African American. The disease usually presents in childhood and continues throughout the patient's life. The calcific masses are typically painless and grow at variable rates, sometimes becoming large and bulky. The masses are often located near major joints but remain extracapsular. Joint range of motion is not usually restricted unless the tumors are very large. Complications include compression of neural structures and ulceration of the overlying skin

with drainage of chalky fluid and risk of secondary infection. Small deposits not detected by standard radiographs may be detected by ^{99m}Tc bone scanning. The most common laboratory findings are hyperphosphatemia and elevated serum 1,25-dihydroxyvitamin D levels. Serum calcium, parathyroid hormone, and ALP levels are usually normal. Renal function is also usually normal. Urine calcium and phosphate excretions are low, and calcium and phosphate balances are positive.

An acquired form of the disease may occur with other causes of hyperphosphatemia, such as secondary hyperparathyroidism associated with hemodialysis, hypoparathyroidism, pseudohypoparathyroidism, and massive cell lysis following chemotherapy for leukemia. Tissue trauma from joint movement may contribute to the periarticular calcifications. Metastatic calcifications are also seen in conditions associated with hypercalcemia, such as in sarcoidosis, vitamin D intoxication, milk-alkali syndrome, and primary hyperparathyroidism. In these conditions, however, mineral deposits are more likely to occur in proton-transporting organs such as kidney, lungs, and gastric mucosa in which an alkaline milieu is generated by the proton pumps.

TREATMENT: TUMORAL CALCINOSIS

Therapeutic successes have been achieved with surgical removal of subcutaneous calcified masses, which tend not to recur if all calcification is removed from the site. Reduction of serum phosphate by chronic phosphorus restriction may be accomplished using low dietary phosphorus intake alone or in combination with oral phosphate binders. The addition of the phosphaturic agent acetazolamide may be useful. Limited experience using the phosphaturic action of calcitonin deserves further testing.

Dystrophic Calcification

Posttraumatic calcification may occur with normal serum calcium and phosphate levels and normal ion-solubility product. The deposited mineral is either in the form of amorphous calcium phosphate or hydroxyapatite crystals. Soft tissue calcification complicating connective tissue disorders such as scleroderma, dermatomyositis, and systemic lupus erythematosus may involve localized areas of the skin or deeper subcutaneous tissue and is referred to as *calcinosis circumscripta*. Mineral deposition at sites of deeper tissue injury including periarticular sites is called *calcinosis universalis*.

Ectopic Ossification

True extraskeletal bone formation that begins in areas of fasciitis following surgery, trauma, burns, or neurologic injury is referred to as *myositis ossificans*. The bone formed is organized as lamellar or trabecular, with normal osteoblasts and osteoclasts conducting active remodeling. Well-developed haversian systems and marrow elements may be present. A second cause of ectopic bone formation occurs in an inherited disorder, *fibrodysplasia ossificans progressiva*.

Fibrodysplasia Ossificans Progressiva

This is also called *myositis ossificans progressiva*; it is a rare autosomal-dominant disorder characterized by congenital deformities of the hands and feet and episodic soft tissue swellings that ossify. Ectopic bone formation occurs in fascia, tendons, ligaments, and connective tissue within voluntary muscles. Tender, rubbery induration, sometimes precipitated by trauma, develops in the soft tissue and gradually calcifies. Eventually, heterotopic bone forms at these sites of soft tissue trauma. Morbidity results from heterotopic bone interfering with normal movement and function of muscle and other soft tissues. Mortality is usually related to restrictive lung disease caused by an inability of the chest to expand. Laboratory tests are unremarkable.

There is no effective medical therapy. Bisphosphonates, glucocorticoids, and a low-calcium diet have largely been ineffective in halting progression of the ossification. Surgical removal of ectopic bone is not recommended, because the trauma of surgery may precipitate formation of new areas of heterotopic bone. Dental complications including frozen jaw may occur following injection of local anesthetics. Thus, CT imaging of the mandible should be undertaken to detect early sites of soft tissue ossification before they are appreciated by standard radiography.

FURTHER READINGS

Albagha OM et al: Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. Nat Genet 42:520, 2010[PMID: 20436471] [Full Text]

Avramidis A et al: Scintigraphic, biochemical, and clinical response to zoledronic acid treatment in patients with Paget's disease of bone. J Bone Miner Metab 26:635, 2008[PMID: 18979164] [Full Text]

Goode A, Layfield R: Recent advances in understanding the molecular basis of Paget's disease of bone. J Clin Pathol 63:199, 2010[PMID: 19858527] [Full Text]

Hiruma Y et al: A SQSTM1/p62 mutation linked to Paget's disease increases the osteoclastogenic potential of the bone microenvironment. Hum Mol Genet 17:3708, 2008[PMID: 18765443] [Full Text]

Hosking D et al: Long-term control of bone turnover in Paget's disease with zoledronic acid and risedronate. J Bone Miner Res 22:142, 2007[PMID: 17032148] [Full Text]

Langston AL et al: Randomized trial of intensive bisphosphonate treatment versus symptomatic management in Paget's disease of bone. J Bone Miner Res 25:20, 2010[PMID: 19580457] [Full Text]

Ralston SH et al: Pathogenesis and management of Paget's disease of bone. Lancet 372:155, 2008[PMID: 18620951] [Full Text]

Reid IR et al: Comparison of a single infusion of zoledronic acid with risedronate for Paget's disease. N Engl J Med 353:898,2005[PMID: 16135834] [Full Text]

Roodman GE et al: Paget's disease of bone. J Clin Invest 115: 200, 2005[PMID: 15690073] [Full Text]

Whyte MP: Paget's disease of bone. N Engl J Med 355:593, 2006[PMID: 16899779] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF LIPOPROTEIN METABOLISM: INTRODUCTION

Lipoproteins are complexes of lipids and proteins that are essential for the transport of cholesterol, triglycerides, and fat-soluble vitamins. Previously, lipoprotein disorders were the purview of specialized lipidologists, but the demonstration that lipid-lowering therapy significantly reduces the clinical complications of atherosclerotic cardiovascular disease (ASCVD) has brought the diagnosis and treatment of these disorders into the domain of the internist. The number of individuals who are candidates for lipid-lowering therapy has continued to increase. The development of safe, effective, and well-tolerated pharmacologic agents has greatly expanded the therapeutic armamentarium available to the physician to treat disorders of lipid metabolism. Therefore, the appropriate diagnosis and management of lipoprotein disorders is of critical importance in the practice of medicine. This chapter will review normal lipoprotein physiology, the pathophysiology of primary (inherited) disorders of lipoprotein metabolism, the diseases and environmental factors that cause secondary disorders of lipoprotein metabolism, and the practical approaches to their diagnosis and management.

LIPOPROTEIN METABOLISM

Lipoprotein Classification and Composition

Lipoproteins are large macromolecular complexes that transport hydrophobic lipids (primarily triglycerides, cholesterol, and fat-soluble vitamins) through body fluids (plasma, interstitial fluid, and lymph) to and from tissues. Lipoproteins play an essential role in the absorption of dietary cholesterol, long-chain fatty acids, and fat-soluble vitamins; the transport of triglycerides, cholesterol, and fat-soluble vitamins from the liver to peripheral tissues; and the transport of cholesterol from peripheral tissues to the liver.

Lipoproteins contain a core of hydrophobic lipids (triglycerides and cholesteryl esters) surrounded by hydrophilic lipids (phospholipids, unesterified cholesterol) and proteins that interact with body fluids. The plasma lipoproteins are divided into five major classes based on their relative density **(Fig. 356-1** and **Table 356-1)**: chylomicrons, very low-density lipoproteins (VLDLs), intermediate-density lipoproteins (IDLs), low-density lipoproteins (LDLs), and high-density lipoproteins (HDLs). Each lipoprotein class comprises a family of particles that vary slightly in density, size and protein composition. The density of a lipoprotein is determined by the amount of lipid per particle. HDL is the smallest and most dense lipoprotein, whereas chylomicrons and VLDLs are the largest and least dense lipoprotein particles. Most plasma triglyceride is transported in chylomicrons or VLDLs, and most plasma cholesterol is carried as cholesteryl esters in LDLs and HDLs.

Figure 356-1

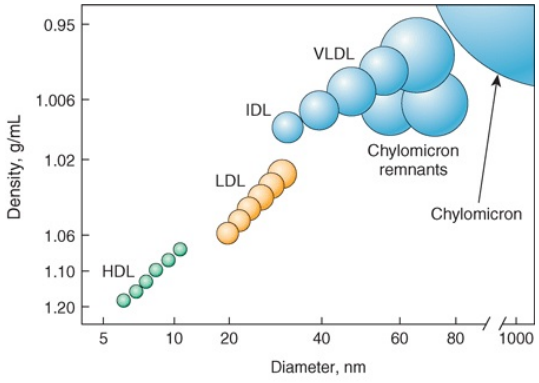

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The density and size-distribution of the major classes of lipoprotein particles. Lipoproteins are classified by density and size, which are inversely related. VLDL, very low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein.

Table 356-1 Major Lipoprotein Classes

Lipoprotein	Density, g/mL[a]	Size, nm[b]	Electrophoretic Mobility[c]	Apolipoproteins		
				Major	Other	Other Constituents
Chylomicrons	0.930	75–1200	Origin	ApoB-48	A-I, A-IV, C-I, C-II, C-III, E	Retinyl esters
Chylomicron remnants	0.930–1.006	30–80	Slow pre-β	ApoB-48	A-I, A-IV, C-I, C-II, C-III, E	Retinyl esters
VLDL	0.930–1.006	30–80	Pre-β	ApoB-100	A-I, A-II, A-V, C-I, C-II, C-III, E	Vitamin E
IDL	1.006–1.019	25–35	Slow pre-β	ApoB-100	C-I, C-II, C-III, E	Vitamin E
LDL	1.019–1.063	18–25	β	ApoB-100		Vitamin E
HDL	1.063–1.210	5–12	α	ApoA-I	A-II, A-IV, A-V, C-III, E	LCAT, CETP paroxonase
Lp(a)	1.050–1.120	25	Pre-β	ApoB-100	Apo(a)	

[a]The density of the particle is determined by ultracentrifugation.

[b]The size of the particle is measured using gel electrophoresis.

[c]The electrophoretic mobility of the particle on agarose gel electrophores reflects the size and surface charge of the particle, with being the position of LDL and a being the position of HDL.

Note: All of the lipoprotein classes contain phospholipids, esterified and unesterified cholesterol, and triglycerides to varying degrees.

Abbreviations: CETP, cholesteryl ester transfer protein; HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LCAT, lecithin-cholesterol acyltransferase; LDL, low-density lipoprotein; Lp(a), lipoprotein A; VLDL, very low-density lipoprotein.

The proteins associated with lipoproteins, called *apolipoproteins* **(Table 356-2)**, are required for the assembly, structure, and function of lipoproteins. Apolipoproteins activate enzymes important in lipoprotein metabolism and act as ligands for cell surface receptors. ApoA-I, which is synthesized in the liver and intestine, is found on virtually all HDL particles. ApoA-II is the second most abundant HDL apolipoprotein and is on approximately two-thirds of the HDL particles. ApoB is the major structural protein of chylomicrons, VLDLs, IDLs, and LDLs; one molecule of apoB, either apoB-48 (chylomicron) or apoB-100 (VLDL, IDL or LDL), is present on each lipoprotein particle. The human liver synthesizes apoB-100, and the intestine makes apoB-48, which is derived from the same gene by mRNA editing. ApoE is present in multiple copies on chylomicrons, VLDL, and IDL, and it plays a critical role in the metabolism and clearance of triglyceride-rich particles. Three apolipoproteins of the C-series (apoC-I, apoC-II, and apoC-III) also participate in the metabolism of triglyceride-rich lipoproteins. ApoB is the only major apolipoprotein that does not transfer between lipoprotein particles. Some of the minor apolipoproteins are listed in **Table 356-2**.

Table 356-2 Major Apolipoproteins

Apolipoprotein	Primary Source	Lipoprotein Association	Function
ApoA-I	Intestine, liver	HDL, chylomicrons	Structural protein for HDL Activates LCAT
ApoA-II	Liver	HDL, chylomicrons	Structural protein for HDL
ApoA-IV	Intestine	HDL, chylomicrons	Unknown
ApoA-V	Liver	VLDL, chylomicrons	Promotes LPL-mediated triglyceride lipolysis
Apo(a)	Liver	Lp(a)	Unknown
ApoB-48	Intestine	Chylomicrons	Structural protein for chylomicrons
ApoB-100	Liver	VLDL, IDL, LDL, Lp(a)	Structural protein for VLDL, LDL, IDL, Lp(a)
			Ligand for binding to LDL receptor
ApoC-I	Liver	Chylomicrons, VLDL, HDL	Unknown
ApoC-II	Liver	Chylomicrons, VLDL, HDL	Cofactor for LPL
ApoC-III	Liver	Chylomicrons, VLDL, HDL	Inhibits lipoprotein binding to receptors
ApoE	Liver	Chylomicron remnants, IDL, HDL	Ligand for binding to LDL receptor
ApoH	Liver	Chylomicrons, VLDL, LDL, HDL	B2 glycoprotein I
ApoJ	Liver	HDL	Unknown
ApoL	Unknown	HDL	Unknown
ApoM	Liver	HDL	Unknown

Abbreviations: HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LCAT, lecithin-cholesterol acyltransferase; LDL, low-density lipoprotein; Lp(a), lipoprotein A; LPL, lipoprotein lipase; VLDL, very low- density lipoprotein.

Transport of Dietary Lipids (Exogenous Pathway)

The exogenous pathway of lipoprotein metabolism permits efficient transport of dietary lipids **(Fig. 356-2)**. Dietary triglycerides are hydrolyzed by lipases within the intestinal lumen and emulsified with bile acids to form micelles. Dietary cholesterol, fatty acids, and fat-soluble vitamins are absorbed in the proximal small intestine. Cholesterol and retinol are esterified (by the addition of a fatty acid) in the enterocyte to form cholesteryl esters and retinyl esters, respectively. Longer-chain fatty acids (>12 carbons) are incorporated into triglycerides and packaged with apoB-48, cholesteryl esters, retinyl esters, phospholipids, and cholesterol to form chylomicrons. Nascent chylomicrons are secreted into the intestinal lymph and delivered via the thoracic duct directly to the systemic circulation, where they are extensively processed by peripheral tissues before reaching the liver. The particles encounter lipoprotein lipase (LPL), which is anchored to a glycosylphosphatidylinositol-anchored protein, GPIHBP1, that is attached to the endothelial surfaces of capillaries in adipose tissue, heart, and skeletal muscle **(Fig. 356-2)**. The triglycerides of chylomicrons are hydrolyzed by LPL, and free fatty acids are released. ApoC-II, which is transferred to circulating chylomicrons from HDL, acts as a required cofactor for LPL in this reaction. The released free fatty acids are taken up by adjacent myocytes or adipocytes and either oxidized to generate energy or reesterified and stored as triglyceride. Some of the released free fatty acids bind albumin before entering cells and are transported to other tissues, especially the liver. The chylomicron particle progressively shrinks in size as the hydrophobic core is hydrolyzed and the hydrophilic lipids (cholesterol and phospholipids) and apolipoproteins on the particle surface are transferred to HDL, creating chylomicron remnants. Chylomicron remnants are rapidly removed from the circulation by the liver through a process that requires apoE as a ligand for receptors in the liver. Consequently, few, if any, chylomicrons or chylomicron remnants are present in the blood after a 12-hour fast, except in patients with disorders of chylomicron metabolism.

Figure 356-2

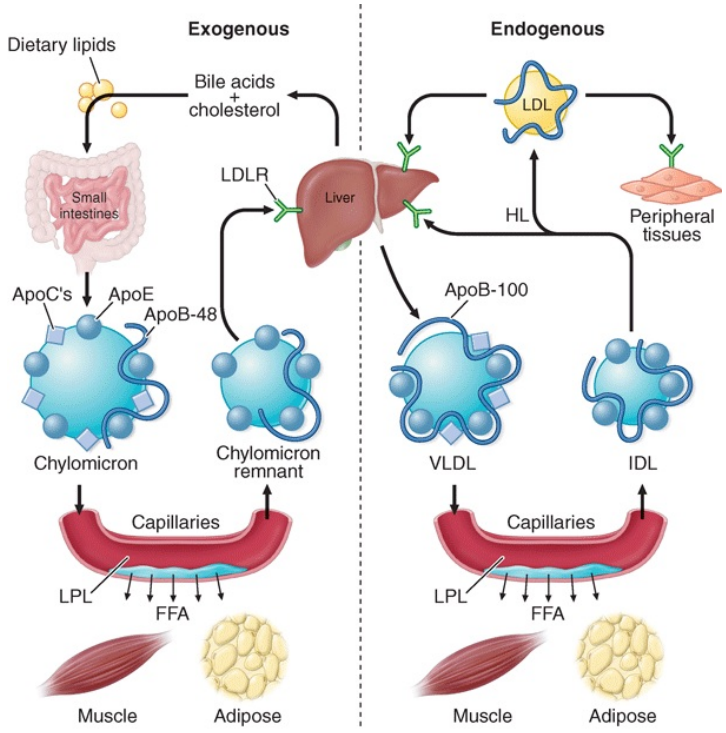

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The exogenous and endogenous lipoprotein metabolic pathways. The exogenous pathway transports dietary lipids to the periphery and the liver. The endogenous pathway transports hepatic lipids to the periphery. LPL, lipoprotein lipase; FFA, free fatty acid; VLDL, very low density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; LDLR, low-density lipoprotein receptor; HL, hepatic lipase.

Transport of Hepatic Lipids (Endogenous Pathway)

The endogenous pathway of lipoprotein metabolism refers to the secretion of apoB-containing lipoproteins from the liver and the metabolism of these triglyceride-rich particles in peripheral tissues **(Fig. 356-2)**. VLDL particles resemble chylomicrons in protein composition but contain apoB-100 rather than apoB-48 and have a higher ratio of cholesterol to triglyceride (~1 mg of cholesterol for every 5 mg of triglyceride). The triglycerides of VLDL are derived predominantly from the esterification of long-chain fatty acids in the liver. The packaging of hepatic triglycerides with the other major components of the nascent VLDL particle (apoB-100, cholesteryl esters, phospholipids, and vitamin E) requires the action of the enzyme microsomal triglyceride transfer protein (MTP). After secretion into the plasma, VLDL acquires multiple copies of apoE and apolipoproteins of the C series by transfer from HDL. As with chylomicrons, the triglycerides of VLDL are hydrolyzed by LPL, especially in muscle, heart, and adipose tissue. After the VLDL remnants dissociate from LPL, they are referred to as IDLs, which contain roughly similar amounts of cholesterol and triglyceride. The liver removes approximately 40–60% of IDL by LDL receptor–mediated endocytosis via binding to apoE. The remainder of IDL is remodeled by hepatic lipase (HL) to form LDL. During this process, most of the triglyceride in the particle hydrolyzed, and all apolipoproteins except apoB-100 are transferred to other lipoproteins. The cholesterol in LDL accounts for more than one-half of the plasma cholesterol in most individuals. Approximately 70% of circulating LDL is cleared by LDL receptor–mediated endocytosis in the liver. *Lipoprotein(a)* [Lp(a)] is a lipoprotein similar to LDL in lipid and protein composition, but it contains an additional protein called *apolipoprotein(a)* [apo(a)]. Apo(a) is synthesized in the liver and attached to apoB-100 by a disulfide linkage. The major site of clearance of Lp(a) is the liver, but the uptake pathway is not known.

HDL Metabolism and Reverse Cholesterol Transport

All nucleated cells synthesize cholesterol, but only hepatocytes and enterocytes can effectively excrete cholesterol from the body, into either the bile or the gut lumen. In the liver, cholesterol is secreted into the bile, either directly or after conversion to bile acids. Cholesterol in peripheral cells is transported from the plasma membranes of peripheral cells to the liver and intestine by a process termed "reverse cholesterol transport" that is facilitated by HDL **(Fig. 356-3)**.

Figure 356-3

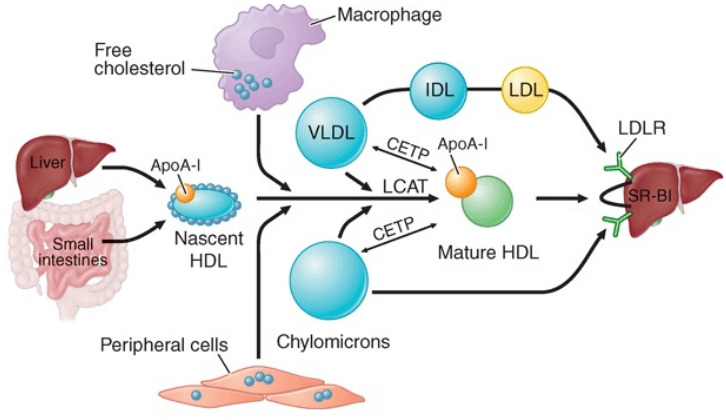

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

HDL metabolism and reverse cholesterol transport. This pathway transports excess cholesterol from the periphery back to the liver for excretion in the bile. The liver and the intestine produce nascent HDLs. Free cholesterol is acquired from macrophages and other peripheral cells and esterified by LCAT, forming mature HDLs. HDL cholesterol can be selectively taken up by the liver via SR-BI (scavenger receptor class BI). Alternatively, HDL cholesteryl ester can be transferred by CETP from HDLs to VLDLs and chylomicrons, which can then be taken up by the liver. LCAT, lecithin-cholesterol acyltransferase; CETP, cholesteryl ester transfer protein; VLDL, very low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein; LDLR, low-density lipoprotein receptor.

Nascent HDL particles are synthesized by the intestine and the liver. Newly secreted apoA-I rapidly acquires phospholipids and unesterified cholesterol from its site of synthesis (intestine or liver) via efflux promoted by the membrane protein ATP-binding cassette protein A1 (ABCA1). This process results in the formation of discoidal HDL particles, which then recruit additional unesterified cholesterol from the periphery. Within the HDL particle, the cholesterol is esterified by lecithin-cholesterol acyltransferase (LCAT), a plasma enzyme associated with HDL, and the more hydrophobic cholesteryl ester moves to the core of the HDL particle. As HDL acquires more cholesteryl ester it becomes spherical, and additional apolipoproteins and lipids are transferred to the particles from the surfaces of chylomicrons and VLDLs during lipolysis.

HDL cholesterol is transported to hepatocytes by both an indirect and a direct pathway. HDL cholesteryl esters can be transferred to apoB-containing lipoproteins in exchange for triglyceride by the cholesteryl ester transfer protein (CETP). The cholesteryl esters are then removed from the circulation by LDL receptor–mediated endocytosis. HDL cholesterol can also be taken up directly by hepatocytes via the scavenger receptor class B1 (SR-B1), a cell surface receptor that mediates the selective transfer of lipids to cells.

HDL particles undergo extensive remodeling within the plasma compartment by a variety of lipid transfer proteins and lipases. The phospholipid transfer protein (PLTP) has the net effect of transferring phospholipids from other lipoproteins to HDL or among different classes of HDL particles. After CETP- and PLTP-mediated lipid exchange, the triglyceride-enriched HDL becomes a much better substrate for HL, which hydrolyzes the triglycerides and phospholipids to generate smaller HDL particles. A related enzyme called *endothelial lipase* hydrolyzes HDL phospholipids, generating smaller HDL particles that are catabolized faster. Remodeling of HDL influences the metabolism, function, and plasma concentrations of HDL.

DISORDERS OF LIPOPROTEIN METABOLISM

Fredrickson and Levy classified hyperlipoproteinemias according to the type of lipoprotein particles that accumulate in the blood (Type I to Type V) **(Table 356-3)**. A classification scheme based on the molecular etiology and pathophysiology of the lipoprotein disorders complements this system and forms the basis for this chapter. The identification and characterization of genes responsible for the genetic forms of hyperlipidemia have provided important molecular insights into the critical roles of structural apolipoproteins, enzymes, and receptors in lipid metabolism **(Table 356-4)**.

Table 356-3 Fredrickson Classification of Hyperlipoproteinemias

Phenotype	I	IIa	IIb	III	IV	V
Lipoprotein, elevated	Chylomicrons	LDL	LDL and VLDL	Chylomicron and VLDL remnants	VLDL	Chylomicrons and VLDL
Triglycerides	↑↑↑	N	↑	↑↑	↑↑	↑↑↑
Cholesterol (total)	↑	↑↑↑	↑↑	↑↑	N/↑	↑↑
LDL-cholesterol	↓	↑↑↑	↑↑	↓	↓	↓
HDL-cholesterol	↓↓↓	N/↓	↓	N	↓↓	↓↓↓
Plasma appearance	Lactescent	Clear	Clear	Turbid	Turbid	Lactescent
Xanthomas	Eruptive	Tendon, tuberous	None	Palmar, tuberoeruptive	None	Eruptive
Pancreatitis	+++	0	0	0	0	+++
Coronary atherosclerosis	0	+++	+++	+++	+/–	+/–
Peripheral atherosclerosis	0	+	+	++	+/–	+/–
Molecular defects	LPL and ApoC-II	LDL receptor, ApoB-100, PCSK9, LDLRAP, ABCG5 and ABCG8		ApoE	ApoA-V	ApoA-V and GPIHBP1
Genetic nomenclature	FCS	FH, FDB, ADH, ARH, sitosterolemia	FCHL	FDBL	FHTG	FHTG

Abbreviations: ADH, autosomal dominant hypercholesterolemia; Apo, apolipoprotein; ARH, autosomal recessive hypercholesterolemia; FCHL, familial combined hyperlipidemia; FCS, familial chylomicronemia syndrome; FDB, familial defective ApoB; FDBL, familial dysbetalipoproteinemia; FH, familial hypercholesterolemia; FHTG, familial hypertriglyceridemia; LPL, lipoprotein lipase; LDLRAP, LDL receptor associated protein; GPIHBP1, glycosylphosphatidylinositol-anchored high density lipoprotein binding protein1; N, normal

Table 356-4 Primary Hyperlipoproteinemias Caused by Known Single Gene Mutations

Genetic Disorder	Protein (Gene) Defect	Lipoproteins Elevated	Clinical Findings	Genetic Transmission	Estimated Incidence
Lipoprotein lipase deficiency	LPL (LPL)	Chylomicrons	Eruptive xanthomas, hepatosplenomegaly, pancreatitis	AR	1/1,000,000
Familial apolipoprotein C-II deficiency	ApoC-II (APOC2)	Chylomicrons	Eruptive xanthomas, hepatosplenomegaly, pancreatitis	AR	<1/1,000,000
ApoA-V deficiency	ApoA-V (APOA5)	Chylomicrons, VLDL	Eruptive xanthomas, hepatosplenomegaly, pancreatitis	AD	<1/1,000,000
GPIHBP1 deficiency	GDIHBP1	Chylomicrons	Eruptive xanthomas, pancreatitis	AD	<1/1,000,000
Familial hepatic lipase deficiency	Hepatic lipase (LIPC)	VLDL remnants	Pancreatitis, CHD	AR	<1/1,000,000
Familial dysbetalipoproteinemia	ApoE (APOE)	Chylomicron and VLDL remnants	Palmar and tuberoeruptive xanthomas, CHD, PVD	AR AD	1/10,000
Familial hypercholesterolemia	LDL receptor (LDLR)	LDL	Tendon xanthomas, CHD	AD	1/500
Familial defective apoB-100	ApoB-100 (APOB)	LDL	Tendon xanthomas, CHD	AD	<1/1000
Autosomal dominant hypercholesterolemia	PCSK9 (PCSK9)	LDL	Tendon xanthomas, CHD	AD	<1/1,000,000

Autosomal recessive hypercholesterolemia	LDLRAP	LDL	Tendon xanthomas, CHD	AR	<1/1,000,000
Sitosterolemia	ABCG5 or ABCG8	LDL	Tendon xanthomas, CHD	AR	<1/1,000,000

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ARH, autosomal recessive hypercholesterolemia; CHD, coronary heart disease; LDL, low-density lipoprotein; LPL, lipoprotein lipase; PVD, peripheral vascular disease; VLDL, very-low density lipoprotein.

Primary Disorders of Elevated Apob-Containing Lipoproteins

A variety of genetic conditions are associated with the accumulation in plasma of specific classes of lipoprotein particles. In general, these can be divided into those causing elevated LDL-cholesterol (LDL-C) with normal triglycerides and those causing elevated triglycerides **(Table 356-4)**.

LIPID DISORDERS ASSOCIATED WITH ELEVATED LDL-C AND NORMAL TRIGLYCERIDES

Familial Hypercholesterolemia (Fh)

FH is an autosomal codominant disorder characterized by elevated plasma levels of LDL-C with normal triglycerides, tendon xanthomas, and premature coronary atherosclerosis. FH is caused by a large number (>1000) mutations in the LDL receptor gene. It has a higher incidence in certain founder populations, such as Afrikaners, Christian Lebanese, and French Canadians. The elevated levels of LDL-C in FH are due to an increase in the production of LDL from IDL (since a portion of IDL is normally cleared by LDL receptor–mediated endocytosis) and a delayed removal of LDL from the blood. Individuals with two mutated LDL receptor alleles (FH homozygotes) have much higher LDL-C levels than those with one mutant allele (FH heterozygotes).

Homozygous FH occurs in approximately 1 in 1 million persons worldwide. Patients with homozygous FH can be classified into one of two groups based on the amount of LDL receptor activity measured in their skin fibroblasts: those patients with <2% of normal LDL receptor activity (receptor negative) and those patients with 2–25% of normal LDL receptor activity (receptor defective). Most patients with homozygous FH present in childhood with cutaneous xanthomas on the hands, wrists, elbows, knees, heels, or buttocks. Total cholesterol levels are usually >500 mg/dL and can be higher than 1000 mg/dL. The devastating complication of homozygous FH is accelerated atherosclerosis, which can result in disability and death in childhood. Atherosclerosis often develops first in the aortic root, where it can cause aortic valvular or supravalvular stenosis, and typically extends into the coronary ostia, which become stenotic. Children with homozygous FH often develop symptomatic coronary atherosclerosis before puberty; symptoms can be atypical, and sudden death is not uncommon. Untreated, receptor-negative patients with homozygous FH rarely survive beyond the second decade; patients with receptor-defective LDL receptor defects have a better prognosis but almost invariably develop clinically apparent atherosclerotic vascular disease by age 30, and often much sooner. Carotid and femoral disease develops later in life and is usually not clinically significant.

A careful family history should be taken, and plasma lipid levels should be measured in the parents and other first-degree relatives of patients with homozygous FH. The disease has >90% penetrance so both parents of FH homozygotes usually have hypercholesterolemia. The diagnosis of homozygous FH can be confirmed by obtaining a skin biopsy and measuring LDL receptor activity in cultured skin fibroblasts, or by quantifying the number of LDL receptors on the surfaces of lymphocytes using cell sorting technology. Molecular assays are also available to define the mutations in the LDL receptor by DNA sequencing. In selected populations where particular mutations predominate (e.g., Africaners and French Canadians), the common mutations can be screened for directly. Alternatively, the entire coding region needs to be sequenced for mutation detection because a large number of different LDL receptor mutations can cause disease. Ten to 15% of LDL receptor mutations are large deletions or insertions, which may be missed by routine DNA sequencing.

Combination therapy with an HMG-CoA reductase inhibitor and a second drug (cholesterol absorption inhibitor or bile acid sequestrant) sometimes reduces plasma LDL-C in those FH homozygotes who have residual LDL receptor activity, but patients with homozygous FH invariably require additional lipid-lowering therapy. Since the liver is quantitatively the most important tissue for removing circulating LDLs via the LDL receptor, liver transplantation is effective in decreasing plasma LDL-C levels in this disorder. Liver transplantation, however, is associated with substantial risks, including the requirement for long-term immunosuppression. The current treatment of choice for homozygous FH is LDL apheresis (a process by which the LDL particles are selectively removed from the circulation), which can promote regression of xanthomas and may slow the progression of atherosclerosis. Initiation of LDL apheresis should generally be delayed until approximately 5 years of age, except when evidence of atherosclerotic vascular disease is present.

Heterozygous FH is caused by the inheritance of one mutant LDL receptor allele and occurs in approximately 1 in 500 persons worldwide, making it one of the most common single-gene disorders. It is characterized by elevated plasma levels of LDL-C (usually 200–400 mg/dL) and normal levels of triglyceride. Patients with heterozygous FH have hypercholesterolemia from birth, and disease recognition is usually based on detection of hypercholesterolemia on routine screening, the appearance of tendon xanthomas, or the development of symptomatic ASCVD. Since the disease is codominant in inheritance, one parent and ~50% of the patient's siblings usually also have hypercholesterolemia. The family history is frequently positive for premature ASCVD on one side of the family. Corneal arcus is common, and tendon xanthomas involving the dorsum of the hands, elbows, knees, and especially the Achilles tendons are present in ~75% of patients. The age of onset of ASCVD is highly variable and depends in part on the molecular defect in the LDL receptor gene and also on coexisting cardiac risk factors. FH heterozygotes with elevated plasma levels of Lp(a) appear to be at greater risk for cardiovascular complications. Untreated men with heterozygous FH have an ~50% chance of having a myocardial infarction before age 60 years. Although the age of onset of atherosclerotic heart disease is later in women with FH, coronary heart disease (CHD) is significantly more common in women with FH than in the general female population.

No definitive diagnostic test for heterozygous FH is available. Although FH heterozygotes tend to have reduced levels of LDL receptor function in skin fibroblasts, significant overlap with the LDL receptor activity levels in normal fibroblasts exists. Molecular assays are now available to identify mutations in the LDL receptor gene by DNA sequencing, but the clinical utility of pinpointing the mutation has not been demonstrated. The clinical diagnosis is usually not problematic, but it is critical that hypothyroidism, nephrotic syndrome, and obstructive liver disease be excluded before initiating therapy.

FH patients should be aggressively treated to lower plasma levels of LDL-C. Initiation of a low-cholesterol, low-fat diet is recommended, but heterozygous FH patients require lipid-lowering drug therapy. Statins are effective in heterozygous FH, but combination drug therapy with the addition of a cholesterol absorption inhibitor and/or bile acid sequestrant is frequently required, and the addition of nicotinic acid is sometimes needed. Heterozygous FH patients who cannot be adequately controlled on combination drug therapy are candidates for LDL apheresis.

FAMILIAL DEFECTIVE APOB-100 (FDB)

FDB is a dominantly inherited disorder that clinically resembles heterozygous FH. The disease is rare in most populations except individuals of German descent, where the frequency can be as high as 1 in 1000. FDB is characterized by elevated plasma LDL-C levels with normal triglycerides, tendon xanthomas, and an increased incidence of premature ASCVD. FDB is caused by mutations in the LDL receptor–binding domain of apoB-100, most commonly due to a substitution of glutamine for arginine at position 3500. As a consequence of the mutation in apoB-100, LDL binds the LDL receptor with reduced affinity, and LDL is removed from the circulation at a reduced rate. Patients with FDB cannot be clinically distinguished from patients with heterozygous FH, although patients with FDB tend to have lower plasma levels of LDL-C than FH heterozygotes. The apoB-100 gene mutation can be detected directly, but genetic diagnosis is not currently encouraged since the recommended management of FDB and heterozygous FH is identical.

Autosomal Dominant Hypercholesterolemia Due to Mutations in Pcsk9 (ADH-Pcsk9 or Adh3)

ADH-PCSK9 is a rare autosomal dominant disorder caused by gain-of-function mutations in proprotein convertase subtilisin/kexin type 9

(PCSK9). PCSK9 is a secreted protein that binds to the LDL receptor, resulting in its degradation. Normally, after LDL binds to the receptor it is internalized along with the receptor. In the low pH of the endosome, LDL dissociates from the receptor and returns to the cell surface. The LDL is delivered to the lysosome. When PCSK9 binds the receptor, the complex is internalized and the receptor is redirected to the lysosome rather than to the cell surface. The missense mutations in PCSK9 that cause hypercholesterolemia enhance the activity of PCSK9. As a consequence, the number of hepatic LDL receptors is reduced. Patients with ADH-PCSK9 are indistinguishable clinically from patients with FH. Interestingly, loss-of-function mutations in PCSK9 cause low LDL-C levels (see below).

Autosomal Recessive Hypercholesterolemia (Arh)

ARH is a rare disorder (except in Sardinia, Italy) due to mutations in a protein (ARH, also called LDLR adaptor protein, LDLRAP) involved in LDL receptor–mediated endocytosis in the liver. In the absence of LDLRAP, LDL binds to the LDL receptor but the lipoprotein-receptor complex fails to be internalized. ARH, like homozygous FH, is characterized by hypercholesterolemia, tendon xanthomas, and premature coronary artery disease (CAD). The levels of plasma LDL-C tend to be intermediate between the levels present in FH homozygotes and FH heterozygotes, and CAD is not usually symptomatic until at least the third decade. LDL receptor function in cultured fibroblasts is normal or only modestly reduced in ARH, whereas LDL receptor function in lymphocytes and the liver is negligible. Unlike FH homozygotes, the hyperlipidemia responds partially to treatment with HMG-CoA reductase inhibitors, but these patients usually require LDL apheresis to lower plasma LDL-C to recommended levels.

Sitosterolemia

Sitosterolemia is another rare autosomal recessive disease that can result in severe hypercholesterolemia, tendon xanthomas, and premature ASCVD. Sitosterolemia is caused by mutations in either of two members of the ATP-binding cassette (ABC) half transporter family, ABCG5 and ABCG8. These genes are expressed in enterocytes and hepatocytes. The proteins heterodimerize to form a functional complex that pumps plant sterols such as sitosterol and campesterol, and animal sterols, predominantly cholesterol, into the gut lumen and into the bile. In normal individuals, <5% of dietary plant sterols are absorbed by the proximal small intestine and delivered to the liver. Absorbed plant sterols are preferentially secreted into the bile and are maintained at very low levels. In sitosterolemia, the intestinal absorption of sterols is increased and biliary excretion of the sterols is reduced, resulting in increased plasma and tissue levels of both plant sterols and cholesterol.

Incorporation of plant sterols into cell membranes results in misshapen red blood cells and megathrombocytes that are visible on blood smear. Episodes of hemolysis are a distinctive clinical feature of this disease compared to other genetic forms of hypercholesterolemia.

Sitosterolemia is diagnosed by demonstrating an increase in the plasma level of sitosterol using gas chromatography. The hypercholesterolemia is unusually responsive to reductions in dietary cholesterol content and should be suspected in individuals who have a >40% reduction in plasma cholesterol level on a low-cholesterol diet. The hypercholesterolemia does not respond to HMG-CoA reductase inhibitors, whereas bile acid sequestrants and cholesterol-absorption inhibitors such as ezetimibe, are effective in reducing plasma sterol levels in these patients.

Polygenic Hypercholesterolemia

This condition is characterized by hypercholesterolemia due to elevated LDL-C with a normal plasma level of triglyceride in the absence of secondary causes of hypercholesterolemia. Plasma LDL-C levels are generally not as elevated as they are in FH and FDB. Family studies are useful to differentiate polygenic hypercholesterolemia from the single-gene disorders described above; one-half of the first-degree relatives of patients with FH and FDB are hypercholesterolemic, whereas <10% of first-degree relatives of patients with polygenic hypercholesterolemia have hypercholesterolemia. Treatment of polygenic hypercholesterolemia is identical to that of other forms of hypercholesterolemia.

ELEVATED PLASMA LEVELS OF LIPOPROTEIN(A)

Unlike the other major classes of lipoproteins, that have a normal distribution in the population, plasma levels of Lp(a) have a highly skewed distribution with levels varying over a 1000-fold range. Levels are strongly influenced by genetic factors, with individuals of African and South Asian descent having higher levels than those of European descent. Although it has been well documented that elevated levels of Lp(a) are associated with an increase in ASCVD, lowering plasma levels of Lp(a) has not been demonstrated to reduce cardiovascular risk.

LIPID DISORDERS ASSOCIATED WITH ELEVATED TRIGLYCERIDES

Familial Chylomicronemia Syndrome (Type I Hyperlipoproteinemia; Lipoprotein Lipase and Apoc-II Deficiency)

As noted above, LPL is required for the hydrolysis of triglycerides in chylomicrons and VLDLs, and apoC-II is a cofactor for LPL **(Fig. 356-2)**. Genetic deficiency or inactivity of either protein results in impaired lipolysis and profound elevations in plasma chylomicrons. These patients can also have elevated plasma levels of VLDL, but chylomicronemia predominates. The fasting plasma is turbid, and if left at 4°C (39.2°F) for a few hours, the chylomicrons float to the top and form a creamy supernatant. In these disorders, called *familial chylomicronemia syndromes*, fasting triglyceride levels are almost invariably >1000 mg/dL. Fasting cholesterol levels are also elevated but to a lesser degree.

LPL deficiency has autosomal recessive inheritance and has a frequency of approximately 1 in 1 million in the population. *ApoC-II deficiency* is also recessive in inheritance pattern and is even less common than LPL deficiency. Multiple different mutations in the LPL and apoC-II genes cause these diseases. Obligate LPL heterozygotes have normal or mild-to-moderate elevations in plasma triglyceride levels, whereas individuals heterozygous for mutation in apoC-II do not have hypertriglyceridemia.

Both LPL and apoC-II deficiency usually present in childhood with recurrent episodes of severe abdominal pain due to acute pancreatitis. On funduscopic examination, the retinal blood vessels are opalescent (lipemia retinalis). Eruptive xanthomas, which are small, yellowish-white papules, often appear in clusters on the back, buttocks, and extensor surfaces of the arms and legs. These typically painless skin lesions may become puritic. Hepatosplenomegaly results from the uptake of circulating chylomicrons by reticuloendothelial cells in the liver and spleen. For unknown reasons, some patients with persistent and pronounced chylomicronemia never develop pancreatitis, eruptive xanthomas, or hepatosplenomegaly. Premature CHD is not generally a feature of familial chylomicronemia syndromes.

The diagnoses of LPL and apoC-II deficiency are established enzymatically in specialized laboratories by assaying triglyceride lipolytic activity in postheparin plasma. Blood is sampled after an IV heparin injection to release the endothelial-bound LPL. LPL activity is profoundly reduced in both LPL and apoC-II deficiency; in patients with apoC-II deficiency, it normalizes after the addition of normal plasma (providing a source of apoC-II). Molecular sequencing of the genes can be used to confirm the diagnosis.

The major therapeutic intervention in familial chylomicronemia syndromes is dietary fat restriction (to as little as 15 g/d) with fat-soluble vitamin supplementation. Consultation with a registered dietician familiar with this disorder is essential. Caloric supplementation with medium-chain triglycerides, which are absorbed directly into the portal circulation, can be useful but may be associated with hepatic fibrosis if used for prolonged periods. If dietary fat restriction alone is not successful in resolving the chylomicronemia, fish oils have been effective in some patients. In patients with apoC-II deficiency, apoC-II can be provided by infusing fresh-frozen plasma to resolve the chylomicronemia in the acute setting. Management of patients with familial chylomicronemia syndrome is particularly challenging during pregnancy when VLDL production is increased and may require plasmapheresis to remove the circulating chylomicrons.

Apoa-V Deficiency

Another apolipoprotein, ApoA-V, circulates at much lower concentrations than the other major apolipoproteins. Individuals harboring mutations in both ApoA-V alleles can present as adults with chylomicronemia. The exact mechanism of action of ApoA-V is not known, but it appears to be required for the association of VLDL and chylomicrons with LPL.

Gpihbp1 Deficiency

After LPL is synthesized in adipocytes, myocytes or other cells, it is transported across the vascular endothelium and is attached to a protein on the endothelial surface of capillaries called GPIHBP1. Homozygosity for mutations that interfere with GPIHBP1 synthesis or folding cause severe hypertriglyceridemia. The frequency of chylomicronemia due to mutations in GHIHBP1 has not been established but appears to be very rare.

Hepatic Lipase Deficiency

HL is a member of the same gene family as LPL and hydrolyzes triglycerides and phospholipids in remnant lipoproteins and HDLs. HL deficiency is a very rare autosomal recessive disorder characterized by elevated plasma levels of cholesterol and triglycerides (mixed hyperlipidemia) due to the accumulation of circulating lipoprotein remnants and either a normal or elevated plasma level of HDL-C. The diagnosis is confirmed by measuring HL activity in postheparin plasma. Due to the small number of patients with HL deficiency, the association of this genetic defect with ASCVD is not clearly known, but lipid-lowering therapy is recommended.

Familial Dysbetalipoproteinemia (Type III Hyperlipoproteinemia)

Like HL deficiency, familial dysbetalipoproteinemia (FDBL) (also known as *type III hyperlipoproteinemia* or *familial broad β disease*) is characterized by a mixed hyperlipidemia due to the accumulation of remnant lipoprotein particles. ApoE is present in multiple copies on chylomicron and VLDL remnants and mediates their removal via hepatic lipoprotein receptors **(Fig. 356-2)**. FDBL is due to genetic variations in apoE that interfere with its ability to bind lipoprotein receptors. The *APOE* gene is polymorphic in sequence, resulting in the expression of three common isoforms: apoE3, which is the most common; and apoE2 and apoE4, which both differ from apoE3 by a single amino acid. Although associated with slightly higher LDL-C levels and increased CHD risk, the apoE4 allele is not associated with FDBL. Patients with apoE4 have an increased incidence of late-onset Alzheimer's disease. ApoE2 has a lower affinity for the LDL receptor; therefore, chylomicron and VLDL remnants containing apoE2 are removed from plasma at a slower rate. Individuals who are homozygous for the E2 allele (the E2/E2 genotype) comprise the most common subset of patients with FDBL.

Approximately 0.5% of the general population are apoE2/E2 homozygotes, but only a small minority of these individuals develop FDBL. In most cases, an additional, identifiable factor precipitates the development of hyperlipoproteinemia. The most common precipitating factors are a high-fat diet, diabetes mellitus, obesity, hypothyroidism, renal disease, HIV infection, estrogen deficiency, alcohol use, or certain drugs. Other mutations in apoE can cause a dominant form of FDBL where the hyperlipidemia is fully manifest in the heterozygous state, but these mutations are rare.

Patients with FDBL usually present in adulthood with incidental hyperlipidemia, xanthomas, premature coronary disease, or peripheral vascular disease. The disease seldom presents in women before menopause. Two distinctive types of xanthomas, tuberoeruptive and palmar, are seen in FDBL patients. Tuberoeruptive xanthomas begin as clusters of small papules on the elbows, knees, or buttocks and can grow to the size of small grapes. Palmar xanthomas (alternatively called *xanthomata striata palmaris*) are orange-yellow discolorations of the creases in the palms and wrists. In FDBL, in contrast to other disorders of elevated triglycerides, the plasma levels of cholesterol and triglyceride are often elevated to a similar degree and the level of HDL-C is usually normal rather than being low.

The traditional approaches to diagnosis of this disorder are lipoprotein electrophoresis (broad β band) or ultracentrifugation (ratio of VLDL-C to total plasma triglyceride >0.30). Protein methods (apoE phenotyping) or DNA-based methods (apoE genotyping) can be performed to confirm homozygosity for apoE2. However, absence of the apoE2/E2 genotype does not rule out the diagnosis of FDBL, since other mutations in apoE can cause this condition.

Since FDBL is associated with increased risk of premature ASCVD, it should be treated aggressively. Subjects with FDBL tend to have more peripheral vascular disease than is typically seen in FH. Other metabolic conditions that can worsen the hyperlipidemia (see above) should be aggressively treated. Patients with FDBL are typically very diet-responsive and can respond favorably to weight reduction and to low-cholesterol, low-fat diets. Alcohol intake should be curtailed. HMG-CoA reductase inhibitors, fibrates, and niacin are all generally effective in the treatment of FDBL, and sometimes combination drug therapy is required.

Familial Hypertriglyceridemia (Fhtg)

FHTG is a relatively common (~1 in 500) autosomal dominant disorder of unknown etiology characterized by moderately elevated plasma triglycerides accompanied by more modest elevations in cholesterol. Since the major class of lipoproteins elevated in this disorder is VLDL, patients with this disorder are often referred to as having *Type IV hyperlipoproteinemia* (Fredrickson classification, Table 356-3). The elevated plasma levels of VLDL are due to increased production of VLDL, impaired catabolism of VLDL, or a combination of these mechanisms. Some patients with FHTG have a more severe form of hyperlipidemia in which both VLDLs and chylomicrons are elevated (*Type V hyperlipidemia*), since these two classes of lipoproteins compete for the same lipolytic pathway. Increased intake of simple carbohydrates, obesity, insulin resistance, alcohol use, and estrogen treatment, all of which increase VLDL synthesis, can exacerbate this syndrome. FHTG appears not to be associated with increased risk of ASCVD in many families.

The diagnosis of FHTG is suggested by the triad of elevated levels of plasma triglycerides (250–1000 mg/dL), normal or only mildly increased cholesterol levels (<250 mg/dL), and reduced plasma levels of HDL-C. Plasma LDL-C levels are generally not increased and are often reduced due to defective metabolism of the triglyceride-rich particles, The identification of other first-degree relatives with hypertriglyceridemia is useful in making the diagnosis. FDBL and familial combined hyperlipidemia (FCHL) should also be ruled out since these two conditions are associated with a significantly increased risk of ASCVD. The plasma apoB levels are lower and the ratio of plasma triglyceride to cholesterol is higher in FHTG than in either FDBL or FCHL.

It is important to consider and rule out secondary causes of the hypertriglyceridemia **(Table 356-5)** before making the diagnosis of FHTG. Lipid-lowering drug therapy can frequently be avoided with appropriate dietary and lifestyle changes. Patients with plasma triglyceride levels >500 mg/dL after a trial of diet and exercise should be considered for drug therapy to avoid the development of chylomicronemia and pancreatitis. Fibrate drugs or fish oils (omega 3 fatty acids) are reasonable first-line approaches for FHTG, and niacin can also be considered in this condition. For more moderate elevations in triglyceride levels (250-500 mg/dL), statins are effective at lowering triglyceride levels.

Table 356-5 Secondary Forms of Hyperlipidemia

LDL		HDL					
Elevated	**Reduced**	**Elevated**	**Reduced**	**VLDL Elevated**	**IDL Elevated**	**ChylomicronsElevated**	**Lp(a)Elevated**
Hypothyroidism	Severe liver disease	Alcohol	Smoking	Obesity	Multiple myeloma	Autoimmune disease	Renal insufficiency
Nephrotic syndrome	Malabsorption	Exercise	DM type 2	DM type 2			
	Malnutrition	Exposure to chlorinated hydrocarbons	Obesity	Glycogen storage disease	Monoclonal gammopathy	DM type 2	Inflammation
Cholestasis	Gaucher's disease		Malnutrition				Menopause
Acute intermittent porphyria	Chronic infectious disease	Drugs: estrogen	Gaucher's disease	Hepatitis Alcohol	Autoimmune disease		Orchidectomy
Anorexia nervosa	Hyperthyroidism Drugs: niacin		Drugs: anabolic steroids	Renal failure Sepsis	Hypothyroidism		Hypothyroidism Acromegaly

Hepatoma Drugs: thiazides, cyclosporin, tegretol	toxicity		steroids, beta blockers	Stress Cushing's syndrome Pregnancy Acromegaly Lipodystrophy Drugs: estrogen, beta blockers, glucocorticoids, bile acid binding resins, retinoic acid		Nephrosis Drugs: growth hormone, isotretinoin

Abbreviations: DM, diabetes mellitus; HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; Lp(a), lipoprotein A; VLDL, very low-density lipoprotein.

Familial Combined Hyperlipidemia (Fchl)

FCHL is generally characterized by moderate elevations in plasma levels of triglycerides (VLDL) and cholesterol (LDL) and reduced plasma levels of HDL-C. Approximately 20% of patients who develop CHD under age 60 have FCHL. The disease appears to be autosomal dominant with incomplete penetrance and affected family members typically have one of three possible phenotypes: (1) elevated plasma levels of LDL-C, (2) elevated plasma levels of triglycerides due to elevation in VLDL, or (3) elevated plasma levels of both LDL-C and triglyceride. A classic feature of FCHL is that the lipoprotein profile can switch among these three phenotypes in the same individual over time and may depend on factors such as diet, exercise, and weight. FCHL can manifest in childhood but is usually not fully expressed until adulthood. A cluster of other metabolic risk factors are often found in association with this hyperlipidemia, including obesity, glucose intolerance, insulin resistance, and hypertension (the so-called metabolic syndrome, Chap. 242). These patients do not develop xanthomas.

Patients with FCHL almost always have significantly elevated plasma levels of apoB. The levels of apoB are disproportionately high relative to the plasma LDL-C concentration, indicating the presence of small, dense LDL particles, which are characteristic of this syndrome. *Hyperapobetalipoproteinemia*, which has been used to describe the state of elevated plasma levels of apoB with normal plasma LDL-C levels, is probably a form of FCHL. Individuals with FCHL generally share the same metabolic defect, which is overproduction of VLDL by the liver. The molecular etiology of FCHL remains poorly understood, and it is likely that defects in several different genes can cause the phenotype of FCHL.

The presence of a mixed dyslipidemia (plasma triglyceride levels between 200 and 800 mg/dL and total cholesterol levels between 200 and 400 mg/dL, usually with HDL-C levels <40 mg/dL in men and <50 mg/dL in women) and a family history of hyperlipidemia and/or premature CHD strongly suggests the diagnosis of FCHL.

Individuals with FCHL should be treated aggressively due to significantly increased risk of premature CHD. Decreased dietary intake of saturated fat and simple carbohydrates, aerobic exercise, and weight loss can all have beneficial effects on the lipid profile. Patients with diabetes should be aggressively treated to maintain good glucose control. Most patients with FCHL require lipid-lowering drug therapy to reduce lipoprotein levels to the recommended range and reduce the high risk of ASCVD. Statins are effective in this condition, but many patients will require a second drug (cholesterol absorption inhibitor, niacin, fibrate, or fish oils) for optimal control of lipoprotein levels.

Inherited Causes of Low Levels of Apob-Containing Lipoproteins

FAMILIAL HYPOBETALIPOPROTEINEMIA (FHB)

Low plasma levels of LDL-C (the "β-lipoprotein") with a genetic or inherited basis are referred to generically as *familial hypobetalipoproteinemia*. Traditionally this term has been used to refer to the condition of low total cholesterol and LDL-C due to mutations in apoB, which represents the most common inherited form of hypocholesterolemia. Most of the mutations causing FHB interfere with the production of apoB, resulting in reduced secretion and/or accelerated catabolism of the protein. Individuals heterozygous for these mutations usually have LDL-C levels <80 mg/dL and may enjoy protection from ASCVD, though this has not been rigorously demonstrated. Some heterozygotes have elevated levels of hepatic triglycerides.

Mutations in both apoB alleles cause homozygous FHB, a disorder resembling abetalipoproteinemia (see below), although the neurologic findings tend to be less severe. Patients with homozygous hypobetalipoproteinemia can be distinguished from individuals with abetalipoproteinemia by measuring the levels of LDL-C in the parents, which are low in hypobetalipoproteinemia and normal in abetalipoproteinemia.

PCSK9 DEFICIENCY

A phenocopy of FHB results from loss-of-function mutations in PCSK9. As reviewed previously, PCSK9 normally promotes the degradation of the LDL receptor. Mutations that interfere with the synthesis of PCSK9, which are more common in individuals of African descent, result in increased LDL receptor activity and ~40% reduction in plasma level of LDL-C. A sequence variation of higher frequency (R46L) is found predominantly in individuals of European descent and is associated with a 15% reduction in LDL-C. Individuals with inactivating mutations are protected from developing CHD relative to those without these sequence variations, presumably due to having lower plasma cholesterol levels since birth.

ABETALIPOPROTEINEMIA

The synthesis and secretion of apoB-containing lipoproteins in the enterocytes of the proximal small bowel and in the hepatocytes of the liver involve a complex series of events that coordinate the coupling of various lipids with apoB-48 and apoB-100, respectively. Abetalipoproteinemia is a rare autosomal recessive disease caused by loss-of-function mutations in the gene encoding microsomal triglyceride transfer protein (MTP), a protein that transfers lipids to nascent chylomicrons and VLDLs in the intestine and liver, respectively. Plasma levels of cholesterol and triglyceride are extremely low in this disorder, and chylomicrons, VLDLs, LDLs, and apoB are undetectable in plasma. The parents of patients with abetalipoproteinemia (obligate heterozygotes) have normal plasma lipid and apoB levels. Abetalipoproteinemia usually presents in early childhood with diarrhea and failure to thrive due to fat malabsorption. The initial neurologic manifestations are loss of deep-tendon reflexes, followed by decreased distal lower extremity vibratory and proprioceptive sense, dysmetria, ataxia, and the development of a spastic gait, often by the third or fourth decade. Patients with abetalipoproteinemia also develop a progressive pigmented retinopathy presenting with decreased night and color vision, followed by reductions in daytime visual acuity and ultimately progressing to near-blindness. The presence of spinocerebellar degeneration and pigmented retinopathy in this disease has resulted in some patients with abetalipoproteinemia being misdiagnosed as having Friedreich's ataxia.

Most clinical manifestations of abetalipoproteinemia result from defects in the absorption and transport of fat-soluble vitamins. Vitamin E and retinyl esters are normally transported from enterocytes to the liver by chylomicrons, and vitamin E is dependent on VLDL for transport out of the liver and into the circulation. As a consequence of the inability of these patients to secrete apoB-containing particles, patients with abetalipoproteinemia are markedly deficient in vitamin E and are also mildly to moderately deficient in vitamins A and K. Patients with abetalipoproteinemia should be referred to specialized centers for confirmation of the diagnosis and appropriate therapy. Treatment consists of a low-fat, high-caloric, vitamin-enriched diet accompanied by large supplemental doses of vitamin E. It is

imperative that treatment be initiated as soon as possible to help forestall development of neurologic sequelae, which can progress even with appropriate therapy. New therapies for this serious disease are needed.

Genetic Disorders of HDL Metabolism

Mutations in genes encoding proteins that play critical roles in HDL synthesis and catabolism can result in both reductions and elevations in plasma levels of HDL-C. Unlike the genetic forms of hypercholesterolemia, which are invariably associated with premature coronary atherosclerosis, genetic forms of hypoalphalipoproteinemia (low HDL-C) are not always associated with accelerated atherosclerosis.

Inherited Causes of Low Levels of HDL-C

GENE DELETIONS IN THE APOAV-AI-CIII-AIV LOCUS AND CODING MUTATIONS IN APOA-I

Complete genetic deficiency of apoA-I due to deletion of the apoA-I gene results in the virtual absence of HDL from the plasma. The genes encoding apoA-I, apoC-III, apoA-IV, and apoA-V are clustered together on chromosome 11, and some patients with no apoA-I have genomic deletions that include other genes in the cluster. ApoA-I is required for LCAT activity. In the absence of LCAT, free cholesterol levels increase in both HDL and in tissues. The free cholesterol can form deposits in the cornea and in the skin, resulting in corneal opacities and planar xanthomas. Premature CHD is a common feature of apoA-I deficiency, especially when additional genes in the complex are also deleted.

Missense and nonsense mutations in the apoA-I gene have been identified in some patients with low plasma levels of HDL-C (usually 15–30 mg/dL), but these are very rare causes of low HDL-C levels. Patients heterozygous for an Arg173Cys substitution in APOAI (so-called apoA-I$_{Milano}$) have very low plasma levels of HDL due to impaired LCAT activation and rapid catabolism of the mutant apolipoprotein and yet have no increased risk of premature CHD. Most other individuals with low plasma HDL-C levels due to missense mutations in apoA-I do not appear to have premature CHD. A few selected missense mutations in apoA-I and apoA-II promote the formation of amyloid fibrils causing systemic amyloidosis.

TANGIER DISEASE (ABCA1 DEFICIENCY)

Tangier disease is a very rare autosomal codominant form of extremely low plasma HDL-C caused by mutations in the gene encoding ABCA1, a cellular transporter that facilitates efflux of unesterified cholesterol and phospholipids from cells to apoA-I **(Fig. 356-3)**. ABCA1 in the liver and intestine rapidly lipidates the apoA-I secreted from these tissues. In the absence of ABCA1, the nascent, poorly lipidated apoA-I is immediately cleared from the circulation. Thus, patients with Tangier disease have extremely low circulating plasma levels of HDL-C (<5 mg/dL) and apoA-I (<5 mg/dL). Cholesterol accumulates in the reticuloendothelial system of these patients, resulting in hepatosplenomegaly and pathognomonic enlarged, grayish yellow or orange tonsils. An intermittent peripheral neuropathy (mononeuritis multiplex) or a sphingomyelia-like neurologic disorder can also be seen in this disorder. Tangier disease is probably associated with some increased risk of premature atherosclerotic disease, although the association is not as robust as might be anticipated, given the very low levels of HDL-C and apoA-I in these patients. Patients with Tangier disease also have low plasma levels of LDL-C, which may attenuate the atherosclerotic risk. Obligate heterozygotes for ABCA1 mutations have moderately reduced plasma HDL-C levels (15–30 mg/dL) but their risk of premature CHD remains uncertain. ABCA1 mutations appear to be the cause of low HDL-C in a minority of individuals.

LCAT DEFICIENCY

This very rare autosomal recessive disorder is caused by mutations in LCAT, an enzyme synthesized in the liver and secreted into the plasma, where it circulates associated with lipoproteins **(Fig. 356-3)**. As reviewed previously, the enzyme is activated by apoA-I and mediates the esterification of cholesterol to form cholesteryl esters. Consequently, in LCAT deficiency the proportion of free cholesterol in circulating lipoproteins is greatly increased (from ~25% to >70% of total plasma cholesterol). Lack of normal cholesterol esterification impairs formation of mature HDL particles, resulting in the rapid catabolism of circulating apoA-I. Two genetic forms of LCAT deficiency have been described in humans: complete deficiency (also called *classic LCAT deficiency*) and partial deficiency (also called *fish-eye disease*). Progressive corneal opacification due to the deposition of free cholesterol in the cornea, very low plasma levels of HDL-C (usually <10 mg/dL), and variable hypertriglyceridemia are characteristic of both disorders. In partial LCAT deficiency, there are no other known clinical sequelae. In contrast, patients with complete LCAT deficiency have hemolytic anemia and progressive renal insufficiency that eventually leads to end-stage renal disease (ESRD). Remarkably, despite the extremely low plasma levels of HDL-C and apoA-I, premature ASCVD is not a consistent feature of either LCAT deficiency or fish eye disease. The diagnosis can be confirmed in a specialized laboratory by assaying plasma LCAT activity or by sequencing the LCAT gene.

PRIMARY HYPOALPHALIPOPROTEINEMIA

Low plasma levels of HDL-C (the "alpha lipoprotein") is referred to as *hypoalphalipoproteinemia*. Primary hypoalphalipoproteinemia is defined as a plasma HDL-C level below the tenth percentile in the setting of relatively normal cholesterol and triglyceride levels, no apparent secondary causes of low plasma HDL-C, and no clinical signs of LCAT deficiency or Tangier disease. This syndrome is often referred to as *isolated low HDL* . A family history of low HDL-C facilitates the diagnosis of an inherited condition, which usually follows an autosomal dominant pattern. The metabolic etiology of this disease appears to be primarily accelerated catabolism of HDL and its apolipoproteins. Some of these patients may have ABCA1 mutations and therefore technically have heterozygous Tangier disease. Several kindreds with primary hypoalphalipoproteinemia have been described in association with an increased incidence of premature CHD, although this is not an invariant association. Association of hypoalphalipoproteinemia with premature CHD may depend on the specific nature of the gene defect or the underlying metabolic defect responsible for the low plasma HDL-C level.

Inherited Causes of High Levels of HDL-C

CETP DEFICIENCY

Loss-of-function mutations in both alleles of the gene encoding CETP cause substantially elevated HDL-C levels (usually >150 mg/dL). As noted above, CETP facilitates the transfer of cholesteryl esters from HDL to apoB-containing lipoproteins **(Fig. 356-3)**. The absence of this transfer results in an increase in the cholesteryl ester content of HDL and a reduction in plasma levels of LDL-C. The large, cholesterol-rich HDL particles circulating in these patients are cleared at a reduced rate. CETP deficiency was first diagnosed in Japanese persons and is rare outside of Japan. The relationship of CETP deficiency to ASCVD remains unresolved. Heterozygotes for CETP deficiency have only modestly elevated HDL-C levels. Based on the phenotype of high HDL-C in CETP deficiency, pharmacologic inhibition of CETP is under development as a new therapeutic approach to both raise HDL-C levels and lower LDL-C levels, but whether it will reduce risk of ASCVD remains to be determined.

FAMILIAL HYPERALPHALIPOPROTEINEMIA

The condition of high plasma levels of HDL-C is referred to as *hyperalphalipoproteinemia* and is defined as a plasma HDL-C level above the ninetieth percentile. This trait runs in families, and outside of Japan it is unlikely to be due to CETP deficiency. Most, but not all, persons with this condition appear to have a reduced risk of CHD and increased longevity. Recent evidence is consistent with mutations in endothelial lipase contributing to this phenotype in some cases.

Secondary Disorders of Lipoprotein Metabolism

Significant changes in plasma levels of lipoproteins are seen in a variety of diseases. It is crucial that secondary causes of dyslipidemias (Table 356-5) are considered prior to initiation of lipid-lowering therapy.

OBESITY

(See also Chaps. 77 and 78) Obesity is frequently accompanied by dyslipidemia. The increase in adipocyte mass and accompanying decreased insulin sensitivity associated with obesity has multiple effects on lipid metabolism. More free fatty acids are delivered from the expanded adipose tissue to the liver, where they are reesterified in hepatocytes to form triglycerides, which are packaged into

VLDLs for secretion into the circulation. The increased insulin levels promote fatty acid synthesis in the liver. Increased dietary intake of simple carbohydrates also drives hepatic production of VLDLs, resulting in elevations in VLDL and/or LDL in some obese subjects. Plasma levels of HDL-C tend to be low in obesity, due in part to reduced lipolysis. Weight loss is often associated with reductions in plasma levels of circulating apoB-containing lipoproteins and increases in the plasma levels of HDL-C.

DIABETES MELLITUS

(See also Chap. 344) Patients with type I diabetes mellitus generally do not have hyperlipidemia if they remain under good glycemic control. Diabetic ketoacidosis is frequently accompanied by hypertriglyceridemia due to an increased hepatic influx of free fatty acids from adipose tissue. Patients with type II diabetes mellitus are usually dyslipidemic, even when under relatively good glycemic control. The high levels of insulin and insulin resistance associated with type II diabetes has multiple effects on fat metabolism: (1) a decrease in LPL activity resulting in reduced catabolism of chylomicrons and VLDLs, (2) an increase in the release of free fatty acid from the adipose tissue, (3) an increase in fatty acid synthesis in the liver, and (4) an increase in hepatic VLDL production. Patients with type II diabetes mellitus have several lipid abnormalities, including elevated plasma triglycerides (due to increased VLDL and lipoprotein remnants), elevated levels of dense LDL, and decreased plasma levels of HDL-C. In some diabetic patients, especially those with a genetic defect in lipid metabolism, the triglycerides can be extremely elevated, resulting in the development of pancreatitis. Elevated plasma LDL-C levels usually are not a feature of diabetes mellitus and suggest the presence of an underlying lipoprotein abnormality or may indicate the development of diabetic nephropathy.

Lipodystrophy is associated with profound insulin resistance and elevated plasma levels of VLDL and chylomicrons that can be especially difficult to control. Those with congenital generalized lipodystrophy have absence of subcutaneous fat associated with muscle hypertrophy and hepatic steatosis; some of these patients have been treated successfully with leptin. Partial lipodystropy can present with dyslipidemia and the diagnosis should be entertained in patients with variations in body fat distribution, particularly increased truncal fat accompanied by reduced fat in the buttocks and extremities.

THYROID DISEASE

(See also Chap. 341) Hypothyroidism is associated with elevated plasma LDL-C levels due primarily to a reduction in hepatic LDL receptor function and delayed clearance of LDL. Conversely, plasma levels of LDL-C are often reduced in the hyperthyroid patient. Hypothyroid patients also frequently have increased levels of circulating IDL, and some patients with hypothyroidism also have mild hypertriglyceridemia. Because hypothyroidism is often subtle and therefore easily overlooked, all patients presenting with elevated plasma levels of LDL-C, IDL or triglycerides should be screened for hypothyroidism. Thyroid replacement therapy usually ameliorates the hypercholesterolemia; if not, the patient probably has a primary lipoprotein disorder and may require lipid-lowering drug therapy.

RENAL DISORDERS

(See also Chap. 280) Nephrotic syndrome is often associated with pronounced hyperlipoproteinemia, which is usually mixed but can manifest as hypercholesterolemia or hypertriglyceridemia. The hyperlipidemia of nephrotic syndrome appears to be due to a combination of increased hepatic production and decreased clearance of VLDLs, with increased LDL production. Effective treatment of the underlying renal disease normalizes the lipid profile, but most patients with chronic nephrotic syndrome require lipid-lowering drug therapy.

ESRD is often associated with mild hypertriglyceridemia (<300 mg/dL) due to the accumulation of VLDLs and remnant lipoproteins in the circulation. Triglyceride lipolysis and remnant clearance are both reduced in patients with renal failure. Because the risk of ASCVD is increased in ESRD subjects with hyperlipidemia, they should probably be aggressively treated with lipid-lowering agents, even though there is inadequate data at present to indicate that this population benefits from LDL-lowering therapy.

Patients with renal transplants usually have increased lipid levels due to the effect of the drugs required for immunosuppression (cyclosporine and glucocorticoids) and present a difficult management problem since HMG-CoA reductase inhibitors must be used cautiously in these patients.

LIVER DISORDERS

(See also Chap. 301) Because the liver is the principal site of formation and clearance of lipoproteins, it is not surprising that liver diseases can affect plasma lipid levels in a variety of ways. Hepatitis due to infection, drugs, or alcohol is often associated with increased VLDL synthesis and mild to moderate hypertriglyceridemia. Severe hepatitis and liver failure are associated with dramatic reductions in plasma cholesterol and triglycerides due to reduced lipoprotein biosynthetic capacity. Cholestasis is associated with hypercholesterolemia, which can be very severe. A major pathway by which cholesterol is excreted from the body is via secretion into bile, either directly or after conversion to bile acids, and cholestasis blocks this critical excretory pathway. In cholestasis, free cholesterol, coupled with phospholipids, is secreted into the plasma as a constituent of a lamellar particle called *LP-X*. The particles can deposit in skinfolds, producing lesions resembling those seen in patients with FDBL (xanthomata strata palmaris). Planar and eruptive xanthomas can also be seen in patients with cholestasis.

ALCOHOL

Regular alcohol consumption has a variable effect on plasma lipid levels. The most common effect of alcohol is to increase plasma triglyceride levels. Alcohol consumption stimulates hepatic secretion of VLDL, possibly by inhibiting the hepatic oxidation of free fatty acids, which then promote hepatic triglyceride synthesis and VLDL secretion. The usual lipoprotein pattern seen with alcohol consumption is Type IV (increased VLDLs), but persons with an underlying primary lipid disorder may develop severe hypertriglyceridemia (Type V) if they drink alcohol. Regular alcohol use also raises plasma levels of HDL-C.

ESTROGEN

Estrogen administration is associated with increased VLDL and HDL synthesis, resulting in elevated plasma levels of both triglycerides and HDL-C. This lipoprotein pattern is distinctive since the levels of plasma triglyceride and HDL-C are typically inversely related. Plasma triglyceride levels should be monitored when birth control pills or postmenopausal estrogen therapy is initiated to ensure that the increase in VLDL production does not lead to severe hypertriglyceridemia. Use of low-dose preparations of estrogen or the estrogen patch can minimize the effect of exogenous estrogen on lipids.

LYSOSOMAL STORAGE DISEASES

(See also Chap. 361) Cholesteryl ester storage disease (due to deficiency in lysosomal acid lipase) and glycogen storage diseases such as von Gierke's disease (caused by mutations in glucose-6-phosphatase) are rare causes of secondary hyperlipidemias.

CUSHING'S SYNDROME

(See also Chap. 342) Glucocorticoid excess is associated with increased VLDL synthesis and hypertriglyceridemia. Patients with Cushing's syndrome can also have mild elevations in plasma levels of LDL-C.

DRUGS

Many drugs have an impact on lipid metabolism and can result in significant alterations in the lipoprotein profile (Table 356-5).

Screening

(See also Chaps. 225 and 242) Guidelines for the screening and management of lipid disorders have been provided by an expert Adult Treatment Panel (ATP) convened by the National Cholesterol Education Program (NCEP) of the National Heart, Lung, and Blood Institute. The NCEP ATPIII guidelines published in 2001 recommend that all adults older than age 20 years should have plasma levels of cholesterol, triglyceride, LDL-C, and HDL-C measured after a 12-hour overnight fast. In most clinical laboratories, the total cholesterol and triglycerides in the plasma are measured enzymatically, and then the cholesterol in the supernatant is measured after precipitation of apoB-containing lipoproteins to determine the HDL-C. The LDL-C is estimated using the following equation:

LDL-C = total cholesterol – (triglycerides/5) – HDL-C.

(The VLDL-C is estimated by dividing the plasma triglyceride by 5, reflecting the ratio of cholesterol to triglyceride in VLDL particles.) This formula is reasonably accurate if test results are obtained on fasting plasma and if the triglyceride level does not exceed ~200 mg/dL; by convention it cannot be used if the triglyceride level is >400 mg/dL. The accurate determination of LDL-C levels in patients with triglyceride levels >200 mg/dL requires application of ultracentrifugation techniques or other direct assays for LDL-C. If the triglyceride level is >200 mg/dL, the guidelines recommend that the "non-HDL-C" be calculated by simple subtraction of HDL-C from the total cholesterol and that this be considered a secondary target of therapy. Further evaluation and treatment is based primarily on the plasma LDL-C and non-HDL-C levels as well as assessment of overall cardiovascular risk.

Diagnosis

The critical first step in managing a lipid disorder is to determine the class or classes of lipoproteins that are increased or decreased in the patient. The Fredrickson classification scheme for hyperlipoproteinemias (Table 356-3), though less commonly used now than in the past, can be helpful in this regard. Once the hyperlipidemia is accurately classified, efforts should be directed to rule out any possible secondary causes of the hyperlipidemia (Table 356-5). Although many patients with hyperlipidemia have a primary or genetic cause of their lipid disorder, secondary factors frequently contribute to the hyperlipidemia. A fasting glucose should be obtained in the initial workup of all subjects with an elevated triglyceride level. Nephrotic syndrome and chronic renal insufficiency should be excluded by obtaining urine protein and serum creatinine. Liver function tests should be performed to rule out hepatitis and cholestasis. Hypothyroidism should be ruled out by measuring serum TSH. Patients with hyperlipidemia, especially hypertriglyceridemia, who drink alcohol should be encouraged to decrease their intake. Sedentary lifestyle, obesity, and smoking are all associated with low HDL-C levels, and patients should be counseled about these issues.

Once secondary causes for the elevated lipoprotein levels have been ruled out, attempts should be made to diagnose the primary lipid disorder since the underlying etiology has a significant effect on the risk of developing CHD, on the response to drug therapy, and on the management of other family members. Often, determining the correct diagnosis requires a detailed family medical history and, in some cases, lipid analyses in family members.

If the fasting plasma triglyceride level is >1000 mg/dL, the patient almost always has chylomicronemia and either has Type I or Type V hyperlipoproteinemia (Table 356-3). The plasma triglyceride to cholesterol ratio helps distinguish between these two possibilities and is higher in Type I than Type V hyperlipoproteinemia. If the patient has Type I hyperlipoproteinemia, a postheparin lipolytic assay should be performed to determine if the patient has LPL or apoC-II deficiency. Type V is a much more frequent form of chylomicronemia in the adult patient. Often treatment of secondary factors contributing to the hyperlipidemia (diet, obesity, glucose intolerance, alcohol ingestion, estrogen therapy) will change a Type V into a Type IV pattern, reducing the risk of developing acute pancreatitis.

If the levels of LDL-C are very high (greater than a ninety-fifth percentile), it is likely the patient has a genetic form of hyperlipidemia. The presence of severe hypercholesterolemia, tendon xanthomas, and an autosomal dominant pattern of inheritance are consistent with the diagnosis of either FH, FDB, or ADH-PCSK9. At the present time, there is no compelling reason to perform molecular studies to further refine the molecular diagnosis, since the treatment of FH and FDB is identical. Recessive forms of severe hypercholesterolemia are rare and if the patient with severe hypercholesterolemia has parents with normal cholesterol levels, sitosterolemia should be considered; a clue to the diagnosis of sitosterolemia is the greater than expected response of the hypercholesterolemia to reductions in dietary cholesterol content or to treatment with either a cholesterol absorption inhibitor (ezetimibe) or to bile acid resins. Patients with more moderate hypercholesterolemia that does not segregate in families as a monogenic trait are likely to have polygenic hypercholesterolemia.

The most common error in the diagnosis and treatment of lipid disorders is in patients with a mixed hyperlipidemia without chylomicronemia. Elevations in the plasma levels of both cholesterol and triglycerides are seen in patients with increased plasma levels of IDL (Type III) and of LDL and VLDL (Type IIB) and in patients with increased levels of VLDL (Type IV). The ratio of triglyceride to cholesterol is higher in Type IV than the other two disorders. The plasma levels of apoB are highest in Type IIB. A beta quantification to determine the VLDL-C/triglyceride ratio in plasma (see discussion of FDBL) or a direct measurement of the plasma LDL-C should be performed at least once prior to initiation of lipid-lowering therapy to determine if the hyperlipidemia is due to the accumulation of remnants or to an increase in both LDL and VLDL.

Treatment: Lipoprotein Disorders

CLINICAL EVIDENCE THAT TREATMENT OF DYSLIPIDEMIA REDUCES RISK OF CHD

Observational Data

Multiple epidemiologic studies have demonstrated a strong relationship between plasma levels of LDL-C and CHD. A direct connection between plasma cholesterol levels and the atherosclerotic process was made in humans when aortic fatty streaks in young persons were shown to be strongly correlated with serum cholesterol levels. The elucidation of homozygous familial hypercholesterolemia was proof that high plasma levels of LDL-C alone are sufficient to cause CAD. Moreover, PCSK9 deficiency proves that having a lifelong reduction in plasma level of LDL-C is associated with a marked reduction in cardiovascular risk.

Clinical Trials: LDL-C Reduction

Early clinical trials of cholesterol (mostly LDL-C) reduction utilized niacin, bile acid sequestrants, and even the surgical approach of partial ileal bypass to reduce serum cholesterol levels. Although most of these early studies found a small but significant reduction in cardiac events, no decrease in total mortality was seen. The discovery of more potent and well-tolerated cholesterol-lowering agents, namely HMG-CoA reductase inhibitors (statins), ushered in a series of large cholesterol reduction trials that unequivocally established the benefit of cholesterol reduction. The first of these studies was the Scandinavian Simvastatin Survival Study (4S) in which hypercholesterolemic men with CHD who were treated with simvastatin had a reduction in major coronary events of 44% and a reduction in total mortality of 30%. These impressive results were followed by additional studies using statins. The consistency of results of these studies is remarkable. They demonstrated statins to be effective in primary as well as secondary prevention, in women as well as men, in elderly as well as middle-aged individuals, and in patients with only modestly elevated LDL-C levels as well as those with severe hypercholesterolemia. In general, these studies demonstrated that a 1% reduction in LDL-C level is associated with a reduction in coronary events of a similar magnitude, and an ~40 mg/dL reduction in LDL-C is associated with an ~22% reduction in coronary events.

More recent studies have enrolled subjects with average or subaverage plasma LDL-C levels and have involved targeting the on-treatment LDL-C to even lower levels. For example, the Heart Protection Study (HPS) included 20,536 men and women, ages 40–80 years, who had either established ASCVD or were at high risk for the development of CHD (primarily diabetes); the only lipid entry criterion was a total plasma cholesterol level of >135 mg/dL. Treatment with simvastatin for an average of 5 years resulted in a 24% reduction in major coronary events and a highly significant 13% reduction in all-cause mortality. Importantly, the relative benefit of statin therapy was similar across tertiles of baseline LDL-C, and even the large subgroup of individuals with an LDL-C <100 mg/dL at baseline experienced significant benefit from therapy. This study demonstrated that statin therapy is beneficial in high-risk subjects, even if the baseline LDL-C level is below the currently recommended targeted goal; it also helped to shift the emphasis from simply treating elevated cholesterol to treating patients at high risk of CHD. Additional large-scale clinical trials have expanded on these findings and confirmed that individuals with other cardiovascular risk factors (hypertension, diabetes) benefit from LDL-lowering therapy even when the initial LDL-C level is only modestly elevated. The JUPITER trial was a primary prevention trial in subjects without CHD and with LDL-C <130 mg/dL but with an elevated plasma level of C-reactive protein (CRP). Treatment with rosuvastatin reduced LDL-C by an average of 50% and significantly reduced cardiovascular events, further extending the indication for statin therapy in primary prevention.

Further studies have compared different statin regimens to show that greater reductions in LDL-C levels with treatment are associated with a greater reduction in major cardiovascular events. Based on several of these studies, a white paper was issued by the NCEP in

2004 establishing an "optional" LDL-C goal of <70 mg/dL in high-risk patients with CHD and of <100 mg/dL in very-high-risk patients without known CHD. These optional targets have been widely embraced, and clinical practice is clearly evolving to treating CHD and high-risk patients more aggressively for LDL reduction.

Clinical Trials: The Triglyceride-HDL Axis

Abnormalities of the triglyceride high-density lipoprotein (TG-HDL) axis are common in patients with CHD, although data supporting pharmacologic intervention in the TG-HDL axis is less compelling than data supporting LDL-C reduction. Fibric acid derivatives (fibrates), nicotinic acid (niacin), and omega 3 fatty acids (fish oils) are the primary agents currently available to lower plasma triglyceride levels and increase plasma levels of HDL-C. Fibrates have been used as lipid-lowering drugs for several decades and are more effective in reducing plasma triglyceride levels and relatively less effective in increasing plasma HDL-C levels. The results of clinical trials using fibrates have been mixed. Some studies such as the Helsinki Heart Study (HHS) and the Veteran Affairs High-Density Lipoprotein Cholesterol Intervention Trial (VA-HIT) demonstrated a significant reduction in nonfatal myocardial infarction and coronary death with gemfibrozil therapy. However, the Bezafibrate Infarction Prevention (BIP) trial of bezafibrate vs placebo in CHD patients with low HDL-C failed to demonstrate a statistically significant reduction in coronary events, the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) trial of fenofibrate in patients with type 2 diabetes failed to show a significant reduction in its primary endpoint of nonfatal myocardial infarction and coronary death, and the Action to Control Cardiovascular Risk in Diabetes (ACCORD) study of fenofibrate vs. placebo added to simvastatin in patients with type 2 diabetes failed to show a significant reduction in its primary end point of major acute cardiovascular events. In each of these studies, the subgroup with elevated baseline triglycerides suggested benefit.

While niacin is the most effective HDL-raising drug currently available, it has not been tested for its ability to reduce cardiovascular risk in subjects with low plasma levels of HDL-C. The AIM-HIGH and HPS2-THRIVE trials are ongoing studies of the effect of niacin added to baseline statin therapy in patients with CHD and low HDL-C. Finally, while low-dose fish oils have been shown to reduce cardiovascular events, higher doses that reduce triglyceride levels have not been tested for their ability to reduce cardiovascular events. Definitive proof that treating the TG-HDL axis reduces cardiovascular events is likely to come from new therapies that are more effective at specifically targeting VLDL and/or HDL particles.

CLINICAL APPROACH TO LIPID-MODIFYING THERAPY

The major goal of lipid-modifying therapy in most patients with disorders of lipid metabolism is to prevent ASCVD and its complications. Management of lipid disorders should be based on clinical trial data demonstrating that treatment reduces cardiovascular morbidity and mortality, although reasonable extrapolation of these data to specific subgroups is sometimes required. Clearly, elevated plasma levels of LDL-C are strongly associated with increased risk of ASCVD, and treatment to lower the levels of plasma LDL-C decreases the risk of clinical cardiovascular events in both secondary and primary prevention. Although the proportional benefit accrued from reducing plasma LDL-C appears to be similar over the entire range of LDL-C values, the absolute risk reduction depends on the baseline level of cardiovascular risk. The treatment guidelines developed by NCEP ATPIII and the 2004 white paper incorporate these principles. As noted above, abnormalities in the TG-HDL axis (elevated triglyceride, low HDL-C, or both) are commonly seen in patients with CHD or who are at high risk for developing it, but clinical trial data supporting the treatment of these abnormalities is much less compelling, and the pharmacologic tools for their management are more limited. Importantly, the NCEP ATPIII guidelines promote the use of the "non-HDL-C" as a secondary target of therapy in patients with triglyceride levels >200 mg/dL. The goals for non-HDL-C are 30 mg/dL higher than the goals for LDL-C. Thus, many patients with abnormalities of the TG-HDL axis require additional therapy for reduction of non-HDL-C to recommended goals.

NONPHARMACOLOGIC TREATMENT

Diet

Dietary modification is an important component in the management of dyslipidemia. The physician should assess the content of the patient's diet and provide suggestions for dietary modifications. In the patient with elevated LDL-C, dietary saturated fat and cholesterol should be restricted. For individuals with hypertriglyceridemia, the intake of simple carbohydrates should be curtailed. For severe hypertriglyceridemia (>1000 mg/dL), restriction of total fat intake is critical. The most widely used diet to lower the LDL-C level is the "Step I diet" developed by the American Heart Association. Most patients have a relatively modest (<10%) decrease in plasma levels of LDL-C on a Step I diet in the absence of any associated weight loss. Almost all persons experience a decrease in plasma HDL-C levels with a reduction in the amount of total and saturated fat in their diet.

Foods and Additives

Certain foods and dietary additives are associated with modest reductions in plasma cholesterol levels. Plant stanol and sterol esters are available in a variety of foods, such as spreads, salad dressings, and snack bars. Plant sterol and sterol esters interfere with cholesterol absorption and reduce plasma LDL-C levels by ~10% when taken three times per day. The addition to the diet of psyllium, soy protein, or Chinese red yeast rice (which contains lovastatin) can have modest cholesterol-lowering effects. No controlled studies have been performed in which several of these nonpharmacologic options have been combined to address their additive or synergistic effects.

Weight Loss and Exercise

The treatment of obesity, if present, can have a favorable impact on plasma lipid levels and should be actively encouraged. Plasma triglyceride and LDL-C levels tend to fall and HDL-C levels tend to increase in obese subjects after weight reduction. Regular aerobic exercise can also have a positive effect on lipids, in large measure due to the associated weight reduction. Aerobic exercise has a very modest elevating effect on plasma levels of HDL-C in most individuals but also has cardiovascular benefits that extend beyond the effects on plasma lipid levels.

PHARMACOLOGIC TREATMENT

The decision to use drug therapy depends on the level of cardiovascular risk. Drug therapy for hypercholesterolemia in patients with established CHD is well supported by clinical trial data, as reviewed above. Even patients with CHD or risk factors who have "average" LDL-C levels benefit from treatment. Drug treatment to lower LDL-C levels in patients with CHD is also highly cost-effective. Patients with diabetes mellitus without known CHD have similar cardiovascular risk to those without diabetes but with preexisting CHD. The NCEP ATPIII guidelines recommended estimating absolute risk of a cardiovascular event over 10 years using a scoring system based on the Framingham Heart Study database. Patients with a 10-year absolute CHD risk of >20% are considered "CHD risk equivalents" to be treated as aggressively as patients with existing CHD. Current NCEP ATPIII guidelines call for drug therapy to reduce LDL-C to <100 mg/dL in patients with established CHD, other ASCVD (aortic aneurysm, peripheral vascular disease, or cerebrovascular disease), diabetes mellitus, or CHD risk equivalents; and "optionally" to reduce LDL-C to <70 mg/dL in high-risk CHD patients. Based on these guidelines, virtually all CHD and CHD risk-equivalent patients require cholesterol-lowering drug therapy. Moderate-risk patients with two or more risk factors and a 10-year absolute risk of 10–20% should be treated to a goal LDL-C of <130 mg/dL or "optionally" to LDL-C <100 mg/dL.

Although helpful to consider 10-year absolute risk in making clinical decisions about lipid-altering drug therapy, there are situations where 10-year risk is low but lifetime risk is very high and therefore treatment is indicated. A typical example would be a young adult with heterozygous FH and an LDL-C >220 mg/dL. Despite a very low 10-year absolute risk, every such patient should be treated with drug therapy to reduce lifetime risk. Indeed, all patients with markedly elevated plasma levels of LDL-C levels (>190 mg/dL) should be strongly considered for drug therapy even if their 10-year absolute CHD risk is not elevated. The decision of whether to initiate drug treatment in individuals with plasma LDL-C levels between 130 and 190 mg/dL remains controversial and depends on both 10-year and lifetime risk. Although it is desirable to avoid drug treatment in patients who are unlikely to develop CHD, a high proportion of patients who eventually develop CHD have plasma LDL-C levels within this range. The presence of other risk factors such as a low plasma level of HDL-C (<40 mg/dL) or the diagnosis of the metabolic syndrome would argue in favor of drug therapy (Chap. 242). Other laboratory

tests such as an elevated plasma level of apoB, Lp(a), or high-sensitivity C-reactive protein, may assist in the identification of high-risk individuals who should be considered for drug therapy when their LDL-C is in a "gray zone."

Drug treatment is also indicated in patients with triglycerides >500 mg/dL who have been screened and treated for secondary causes of hypertriglyceridemia. The goal is to reduce fasting plasma triglycerides to below 500 mg/dL to prevent the risk of acute pancreatitis. When triglycerides are 200–500 mg/dL, the decision to use drug therapy depends on the risk of the patient developing chylomicronemia and an assessment of cardiovascular risk. Most major clinical endpoint trials with statins have excluded persons with triglyceride levels >350–450 mg/dL, and there are therefore few data regarding the effectiveness of statins in reducing cardiovascular risk in persons with hypertriglyceridemia. More data are needed regarding the relative effectiveness of statins, fibrates, niacin, and fish oils for reducing cardiovascular risk in this setting. Combination therapy is often required for optimal control of mixed dyslipidemia.

HMG-CoA Reductase Inhibitors (Statins)

HMG-CoA reductase is a key enzyme in cholesterol biosynthesis, and inhibition of this enzyme decreases cholesterol synthesis. By inhibiting cholesterol biosynthesis, statins lead to increased hepatic LDL receptor activity as a counterregulatory mechanism and thus accelerated clearance of circulating LDL, resulting in a dose-dependent reduction in plasma levels of LDL-C. The magnitude of LDL lowering associated with statin treatment varies widely among individuals, but once a patient is on a statin, the doubling of the statin dose produces an ~6% further reduction in the level of plasma LDL-C. The statins currently available differ in their LDL-C reducing potency **(Table 356-6)**. Currently, there is no convincing evidence that any of the different statins confer an advantage that is independent of the effect on LDL-C. Statins also reduce plasma triglycerides in a dose-dependent fashion, which is roughly proportional to their LDL-C–lowering effects (if the triglycerides are <400 mg/dL). Statins have a modest HDL-raising effect (5–10%) that is not generally dose-dependent.

Table 356-6 Summary of the Major Drugs Used for the Treatment of Hyperlipidemia

Drug	Major Indications	Starting Dose	Maximal Dose	Mechanism	Common Side Effects
HMG-CoA reductase inhibitors (statins)	Elevated LDL-C			↓ Cholesterol synthesis, ↑ hepatic LDL receptors, ↓ VLDL production	Myalgias, arthralgias, elevated transaminases, dyspepsia
Lovastatin		20 mg daily	80 mg daily		
Pravastatin		40 mg qhs	80 mg qhs		
Simvastatin		20 mg qhs	80 mg qhs		
Fluvastatin		20 mg qhs	80 mg qhs		
Atorvastatin		10 mg qhs	80 mg qhs		
Rosuvastatin		10 mg qhs	40 mg qhs		
Cholesterol absorption inhibitors				↓ Intestinal cholesterol absorption	Elevated transaminases
Ezetimibe	Elevated LDL-C	10 mg daily	10 mg daily	LDL receptors	
Bile acid sequestrants	Elevated LDL-C			↑ Bile acid excretion and ↑ LDL receptors	Bloating, constipation, elevated triglycerides
Cholestyramine		4 g daily	32 g daily		
Colestipol		5 g daily	40 g daily		
Colesevelam		3750 mg daily	4375 mg daily		
Nicotinic acid	Elevated LDL-C, low HDL-C, elevated TG			↓ VLDL production	Cutaneous flushing, GI upset, elevated glucose, uric acid, and liver function tests
Immediate-release		100 mg tid	1 g tid		
Sustained-release		250 mg bid	1.5 g bid		
Extended-release		500 mg qhs	2 g qhs		
Fibric acid derivatives	Elevated TG, elevated remnants			↑ LPL, ↓ VLDL synthesis	Dyspepsia, myalgia, gallstones, elevated transaminases
Gemfibrozil		600 mg bid	600 mg bid		
Fenofibrate		145 mg qd	145 mg qd		
Omega 3 fatty acids	Elevated TG	3 g daily	6 g daily	↑ TG catabolism	Dyspepsia, diarrhea, fishy odor to breath

Abbreviations: GI, gastrointestinal; HDL-C, HDL-cholesterol; LDL, low-density lipoprotein; LDL-C, LDL-cholesterol; LPL, lipoprotein lipase; TG, triglyceride; VLDL, very low-density lipoprotein.

Statins are well tolerated and can be taken in tablet form once a day. Potential side effects include dyspepsia, headaches, fatigue, and muscle or joint pains. Severe myopathy and even rhabdomyolysis occur rarely with statin treatment. The risk of statin-associated myopathy is increased by the presence of older age, frailty, renal insufficiency, and coadministration of drugs that interfere with the metabolism of statins such as erythromycin and related antibiotics, antifungal agents, immunosuppressive drugs, and fibric acid derivatives (particularly gemfibrozil). Severe myopathy can usually be avoided by careful patient selection, avoidance of interacting drugs, and instructing the patient to contact the physician immediately in the event of unexplained muscle pain. In the event of muscle symptoms, the plasma creatine kinase (CK) level should be obtained to document the myopathy. Serum CK levels need not be monitored on a routine basis in patients taking statins, as an elevated CK in the absence of symptoms does not predict the development of myopathy and does not necessarily suggest the need for discontinuing the drug.

Another consequence of statin therapy can be elevation in liver transaminases [alanine (ALT) and aspartate (AST)]. They should be checked before starting therapy, at 2–3 months, and then annually. Substantial (greater than three times the upper limit of normal)

elevation in transaminases is relatively rare and mild-to-moderate (one to three times normal) elevation in transaminases in the absence of symptoms need not mandate discontinuing the medication. Severe clinical hepatitis associated with statins is exceedingly rare, and the trend is toward less frequent monitoring of transaminases in patients taking statins. The statin-associated elevation in liver enzymes resolves upon discontinuation of the medication.

Statins appear to be remarkably safe. Meta-analyses of large randomized controlled clinical trials with statins do not suggest an increase in any major noncardiac diseases. Statins are the drug class of choice for LDL-C reduction and are by far the most widely used class of lipid-lowering drugs.

Cholesterol Absorption Inhibitors

Cholesterol within the lumen of the small intestine is derived from the diet (about one-third) and the bile (about two-thirds) and is actively absorbed by the enterocyte through a process that involves the protein NPC1L1. Ezetimibe (Table 356-6) is a cholesterol absorption inhibitor that binds directly to and inhibits NPC1L1 and blocks the intestinal absorption of cholesterol. Ezetimibe (10 mg) inhibits cholesterol absorption by almost 60%, resulting in a reduction in delivery of dietary sterols in the liver and an increase in hepatic LDL receptor expression. The mean reduction in plasma LDL-C on ezetimibe (10 mg) is 18%, and the effect is additive when used in combination with a statin. Effects on triglyceride and HDL-C levels are negligible, and no cardiovascular outcome data have been reported. When used in combination with a statin, monitoring of liver transaminases is recommended. The only role for ezetimibe in monotherapy is in patients who do not tolerate statins; the drug is often added to a statin in patients who require further LDL-C reduction.

Bile Acid Sequestrants (Resins)

Bile acid sequestrants bind bile acids in the intestine and promote their excretion rather than reabsorption in the ileum. To maintain the bile acid pool size, the liver diverts cholesterol to bile acid synthesis. The decreased hepatic intracellular cholesterol content results in upregulation of the LDL receptor and enhanced LDL clearance from the plasma. Bile acid sequestrants, including cholestyramine, colestipol, and colesevelam (Table 356-6), primarily reduce plasma LDL-C levels but can cause an increase in plasma triglycerides. Therefore, patients with hypertriglyceridemia should not be treated with bile acid–binding resins. Cholestyramine and colestipol are insoluble resins that must be suspended in liquids. Colesevelam is available as tablets but generally requires up to six to seven tablets per day for effective LDL-C lowering. Most side effects of resins are limited to the gastrointestinal tract and include bloating and constipation. Since bile acid sequestrants are not systemically absorbed, they are very safe and the cholesterol-lowering drug of choice in children and in women of childbearing age who are lactating, pregnant, or could become pregnant. They are effective in combination with statins as well as in combination with ezetimibe and are particularly useful with one or both of these drugs for difficult-to-treat patients or those with statin intolerance.

Nicotinic Acid (Niacin)

Nicotinic acid, or niacin, is a B-complex vitamin that has been used as a lipid-modifying agent for more than five decades. Niacin reduces the flux of nonesterified fatty acids (NEFAs) to the liver, which is thought to be the mechanism for reduced hepatic triglyceride synthesis and VLDL secretion. Recently, a nicotinic acid receptor (GPR109A) was discovered that suppresses release of NEFA by adipose tissue, thus mediating the effect of niacin on NEFA suppression. Niacin reduces plasma triglyceride and LDL-C levels and raises the plasma concentration of HDL-C (Table 356-6), but it appears that these effects may not be mediated solely by GPR109A. Niacin is also the only currently available lipid-lowering drug that significantly reduces plasma levels of Lp(a) (up to 40%). If properly prescribed and monitored, niacin is a safe and effective lipid-lowering agent.

The most frequent side effect of niacin is cutaneous flushing, which is mediated by activating GPR109A in the skin, leading to local generation of prostaglandin D2 (PGD2) and prostaglandin E2. Flushing can be reduced by formulations that slow the absorption and by taking aspirin prior to dosing. A product is available in Europe that blocks the receptor for PGD2 and attenuates flushing. There is rapid tachyphylaxis to the flushing. Niacin therapy is generally started at lower doses and gradually titrated up to higher doses. Immediate-release crystalline niacin is generally administered three times per day, over-the-counter sustained-release niacin is taken twice a day, and a prescription form of extended-release niacin is taken once a day. Mild elevations in transaminases occur in up to 15% of patients treated with any form of niacin, and on occasion these elevations may require stopping the medication. Niacin potentiates the effect of warfarin, and these two drugs should be prescribed together with caution. Acanthosis nigricans, a dark-colored coarse skin lesion, and maculopathy are infrequent side effects of niacin. Niacin is contraindicated in patients with peptic ulcer disease and can exacerbate the symptoms of esophageal reflux. It can also raise plasma levels of uric acid and precipitate gouty attacks in susceptible patients.

Niacin can raise fasting plasma glucose levels. A study in type 2 diabetics found only a slight increase in fasting glucose and no significant change in HbA1c level with niacin treatment. Low-dose niacin can be used effectively to reduce plasma triglyceride levels and increase HDL-C without adversely impacting on glycemic control. Thus, niacin can be used in diabetic patients, but every effort should be made to optimize the diabetes management before initiating niacin. Glucose should be carefully monitored in nondiabetic patients with impaired fasting glucose after initiation of niacin therapy.

Successful therapy with niacin requires careful education and motivation on the part of the patient. Its advantages are its low cost and long-term safety. It is the most effective drug currently available for raising HDL-C levels. It is particularly useful in patients with combined hyperlipidemia and low plasma levels of HDL-C and is effective in combination with statins. Outcome data are somewhat limited with niacin, but two clinical trials assessing the benefits of adding niacin to a statin in high-risk patients with low HDL-C are currently ongoing.

Fibric Acid Derivatives (Fibrates)

Fibric acid derivatives are agonists of PPARa, a nuclear receptor involved in the regulation of lipid metabolism. Fibrates stimulate LPL activity (enhancing triglyceride hydrolysis), reduce apoC-III synthesis (enhancing lipoprotein remnant clearance), promote beta-oxidation of fatty acids, and may reduce VLDL triglyceride production. Fibrates are the most effective drugs available for reducing triglyceride levels and also raise HDL-C levels modestly (Table 356-6). They have variable effects on LDL-C and in hypertriglyceridemic patients can sometimes be associated with increases in plasma LDL-C levels.

Fibrates are generally very well tolerated. The most common side effect is dyspepsia. Myopathy and hepatitis occur rarely in the absence of other lipid-lowering agents. Fibrates promote cholesterol secretion into bile and are associated with an increased risk of gallstones. Fibrates can raise creatinine and should be used with caution in patients with chronic kidney disease. Importantly, fibrates can potentiate the effect of warfarin and certain oral hypoglycemic agents, so the anticoagulation status and plasma glucose levels should be closely monitored in patients on these agents.

Fibrates are useful and are a reasonable consideration for first-line therapy in patients with severe hypertriglyceridemia (>500 mg/dL) to prevent pancreatitis. Their role in patients with moderate hypertriglyceridemia (200–500 mg/dL) is to promote reduction in non-HDL-C levels, but outcome data regarding their effects on coronary events in this setting remains mixed. In patients with a triglyceride level <500 mg/dL, the role of fibrates is primarily in combination with statins in selected patients with mixed dyslipidemia. In this setting, the risk of myopathy can be minimized with appropriate patient and drug selection and must be carefully weighed against the clinical benefit of the therapy.

Omega 3 Fatty Acids (Fish Oils)

N-3 polyunsaturated fatty acids (n-3 PUFAs) are present in high concentration in fish and in flaxseeds. The most widely used n-3 PUFAs for the treatment of hyperlipidemias are the two active molecules in fish oil: eicosapentaenoic acid (EPA) and decohexanoic acid (DHA). N-3 PUFAs have been concentrated into tablets and in doses of 3–4 g/d are effective at lowering fasting triglyceride levels. Fish oils can cause an increase in plasma LDL-C levels in some patients. Fish oil supplements can be used in combination with fibrates, niacin, or statins to treat hypertriglyceridemia. In general, fish oils are well tolerated and appear to be safe, at least at doses up to 3–4 g. Although fish oil administration is associated with a prolongation in the bleeding time, no increase in bleeding has been seen in clinical trials. A

lower dose of omega 3 (about 1 g) has been associated with reduction in cardiovascular events in CHD patients and is used by some clinicians for this purpose.

Combination Drug Therapy

Combination drug therapy is frequently used for (1) patients unable to reach LDL-C and non-HDL-C goals on statin monotherapy, (2) patients with combined elevated LDL-C and abnormalities of the TG-HDL axis, and (3) patients with severe hypertriglyceridemia who do not achieve non-HDL-C goal on a fibrate or on fish oils alone. When LDL-C and non-HDL-C goals are not achieved on statin monotherapy, a cholesterol absorption inhibitor or bile acid sequestrant can be added to the drug regimen. Combination of niacin with a statin is an attractive option for high-risk patients who do not attain their target LDL-C level on statin monotherapy and have a low HDL-C level. Conversely, in high-risk patients on statin therapy who have an elevated plasma triglyceride level, addition of a fibrate or fish oils is a reasonable consideration.

Severely hypertriglyceridemic patients treated first with a fibrate often fail to reach LDL-C and non-HDL-C goals and are therefore candidates for addition of a statin. Coadministration of statins and fibrates has obvious appeal in patients with combined hyperlipidemia, but no clinical trial has assessed the effectiveness of a statin-fibrate combination compared with either a statin or a fibrate alone in reducing cardiovascular events. The long-term safety of the statin-fibrate combination is not known. Since coadministration of statins and fibrates is associated with an increased incidence of severe myopathy and rhabdomyolysis, patients treated with this combination must be carefully counseled and monitored. This combination of drugs should be used cautiously in patients with underlying renal or hepatic insufficiency; in the elderly, frail, and chronically ill; and in those on multiple medications.

OTHER APPROACHES

Occasionally, patients cannot tolerate any of the existing lipid-lowering drugs at doses required for adequate control of their lipid levels. A larger group of patients, most of whom have genetic lipid disorders, remain significantly hypercholesterolemic despite combination drug therapy. These patients are at high risk for the development or progression of CHD and clinical CHD events. The preferred option for management of patients with severe refractory hypercholesterolemia is LDL apheresis. In this process, the patient's plasma is passed over a column that selectively removes the LDL, and the LDL-depleted plasma is returned to the patient. Patients on maximally tolerated combination drug therapy who have CHD and a plasma LDL-C level >200 mg/dL or no CHD and a plasma LDL-C level >300 mg/dL are candidates for every-other-week LDL apheresis and should be referred to a specialized lipid center.

MANAGEMENT OF LOW HDL-C

Severely reduced plasma levels of HDL-C (<20 mg/dL) accompanied by triglycerides <400 mg/dL usually indicate the presence of a genetic disorder such as a mutation in apoA-I, LCAT deficiency, or Tangier disease. HDL-C levels <20 mg/dL are common in the setting of severe hypertriglyceridemia, in which case the primary focus should be on the management of the triglycerides. HDL-C levels <20 mg/dL also occur in individuals using anabolic steroids. Secondary causes of more moderate reductions in plasma HDL (20–40 mg/dL) should be considered (Table 356-5). Smoking should be discontinued, obese persons should be encouraged to lose weight, sedentary persons should be encouraged to exercise, and diabetes should be optimally controlled. When possible, medications associated with reduced plasma levels of HDL-C should be discontinued. The presence of an isolated low plasma level of HDL-C in a patient with a borderline plasma level of LDL-C should prompt consideration of LDL-lowering drug therapy in high-risk individuals. Statins increase plasma levels of HDL-C only modestly (~5–10%). Fibrates also have only a modest effect on plasma HDL-C levels (increasing levels ~5–15%), except in patients with coexisting hypertriglyceridemia, where the effect on HDL levels can be greater. Niacin is the most effective HDL-C–raising therapeutic agent available and can increase plasma HDL-C by up to ~30%, although some patients fail to achieve clinically important increases in HDL-C levels from niacin therapy.

The issue of whether pharmacologic intervention should be used to specifically raise HDL-C levels has not been adequately addressed in clinical trials. In persons with established CHD and low HDL-C levels whose plasma LDL-C levels are at or below the goal, it may be reasonable to initiate therapy (with a fibrate or niacin) directed specifically at reducing plasma triglyceride levels and raising the level of plasma HDL-C. More data are required before broad recommendations are made to use drug therapy to specifically raise HDL-C levels to prevent cardiovascular events. New HDL-raising approaches are under development that may help to address this important issue.

Management of Elevated Levels of LP(a)

High levels of Lp(a) are associated with increased risk of ASCVD. Genetic studies suggest that this association is causal, but there is no evidence that reducing plasma Lp(a) levels reduces cardiovascular risk. Until such studies are performed, the major therapeutic approach to patients with high plasma levels of Lp(a) and established CAD is to aggressively lower plasma levels of LDL-C. Niacin is the only drug currently available that lowers Lp(a), and might be considered as an addition to a statin in a very-high-risk patient with elevated Lp(a).

FURTHER READINGS

Ashen MD, Blumenthal RS: Clinical practice: Low HDL cholesterol levels. N Engl J Med 353:1252, 2005[PMID: 16177251] [Full Text]

Baigent C et al: Efficacy and safety of cholesterol-lowering treatment: Prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet 366:1267, 2005[PMID: 16214597] [Full Text]

Brunzell JD: Clinical practice. Hypertriglyceridemia. N Engl J Med 357:1009, 2007[PMID: 17804845] [Full Text]

Grundy SM: The issue of statin safety: Where do we stand? Circulation 111:3016, 2005[PMID: 15911705] [Full Text]

——— et al: Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. Circulation 110:227, 2004

National cholesterol education program: Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA 285:2486, 2001

Rader DJ et al: Monogenic hypercholesterolemia: New insights in pathogenesis and treatment. J Clin Invest 111:1795, 2003[PMID: 12813012] [Full Text]

Ridker PM et al: Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 359:2195, 2008[PMID: 18997196] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEMOCHROMATOSIS: INTRODUCTION

Definition

Hemochromatosis is a common inherited disorder of iron metabolism in which dysregulation of intestinal iron absorption results in deposition of excessive amounts of iron in parenchymal cells with eventual tissue damage and impaired function in a wide range of organs. The iron-storage pigment in tissues is called *hemosiderin* because it is believed to be derived from the blood. The term *hemosiderosis* is used to describe the presence of stainable iron in tissues, but tissue iron must be quantified to assess body-iron status accurately (see below and Chap. 103). *Hemochromatosis* refers to a group of genetic diseases that predispose to iron overload, potentially leading to fibrosis and organ failure. Cirrhosis of the liver, diabetes mellitus, arthritis, cardiomyopathy, and hypogonadotropic hypogonadism are the major clinical manifestations.

Although there is debate about definitions, the following terminology is widely accepted.

1. *Hereditary hemochromatosis* is most often caused by a mutant gene, termed *HFE*, which is tightly linked to the HLA-A locus on chromosome 6p (see "Genetic Basis," below). Persons who are homozygous for the mutation are at increased risk of iron overload and account for 80 to 90% of clinical hereditary hemochromatosis in persons of northern European descent. In such subjects, the presence of hepatic fibrosis, cirrhosis, arthropathy, or hepatocellular carcinoma constitutes iron overload–related disease. Rarer forms of non-*HFE* hemochromatosis are caused by mutations in other genes involved in iron metabolism **(Table 357-1)**. The disease can be recognized during its early stages when iron overload and organ damage are minimal. At this stage, the disease is best referred to as *earlyhemochromatosis* or *precirrhotic hemochromatosis*.
2. *Secondary iron overload* occurs as a result of an iron-loading anemia, such as thalassemia or sideroblastic anemia, in which erythropoiesis is increased but ineffective. In the acquired iron-loading disorders, massive iron deposits in parenchymal tissues can lead to the same clinical and pathologic features as in hemochromatosis.

Table 357-1 Classification of Iron Overload States

Hereditary Hemochromatosis	
Hemochromatosis, *HFE*-related (type 1) C282Y homozygosity C282Y/H63D compound heterozygosity	
Hemochromatosis, non-*HFE*-related Juvenile hemochromatosis (type 2A) (hemojuvelin mutations) Juvenile hemochromatosis (type 2B) (hepcidin mutation) Mutated transferrin receptor 2 *TFR2* (type 3) Mutated ferroportin 1 gene, *SLC11A3* (type 4)	
Acquired Iron Overload	
Iron-loading anemias	Chronic liver disease
Thalassemia major	Hepatitis C
Sideroblastic anemia	Alcoholic cirrhosis, especially when advanced
Chronic hemolytic anemias	Nonalcoholic steatohepatitis
Transfusional and parenteraliron overload	Porphyria cutanea tarda
Dietary iron overload	Dysmetabolic iron overload syndrome
	Post-portacaval shunting
Miscellaneous	
Iron overload in sub-Saharan Africa	

Neonataliron overload
Aceruloplasminemia
Congenital atransferrinemia

Prevalence

HFE-associated hemochromatosis mutations are among the most common inherited disease alleles, although the prevalence varies in different ethnic groups. It is most common in populations of northern European extraction in whom approximately 1 in 10 persons are heterozygous carriers and 0.3–0.5% are homozygotes. However, expression of the disease is variable and modified by several factors, especially alcohol consumption and dietary iron intake, blood loss associated with menstruation and pregnancy, and blood donation. Recent population studies indicate that approximately 30% of homozygous men develop iron overload–related disease and about 6% develop hepatic cirrhosis; for women, the figure is closer to 1%. Nearly 70% of patients develop the first symptoms between ages 40 and 60. The disease is rarely evident before age 20, although with family screening (see "Screening for Hemochromatosis," below) and periodic health examinations, asymptomatic subjects with iron overload can be identified, including young menstruating women.

In contrast to *HFE*-associated hemochromatosis, the non-*HFE*-associated forms of hemochromatosis (Table 357-1) are rare, but they affect all races and young people (juvenile hemochromatosis).

Genetic Basis

The *HFE* gene responsible for the most common form of hemochromatosis was identified in 1996. A homozygous G to A mutation resulting in a cysteine to tyrosine substitution at position 282 (C282Y) is the most common mutation. It is identified in 85–90% of patients with hereditary hemochromatosis in populations of northern European descent but is found in only 60% of cases from Mediterranean populations (e.g., southern Italy). A second, relatively common *HFE* mutation (H63D) results in a substitution of histidine to aspartic acid at codon 63. Homozygosity for H63D is not associated with clinically significant iron overload. Some compound heterozygotes (e.g., one copy each of C282Y and H63D) have mild to moderately increased body-iron stores but develop clinical disease only in association with cofactors such as heavy alcohol intake or hepatic steatosis. Thus, *HFE*-associated hemochromatosis is inherited as an autosomal recessive trait; heterozygotes have no, or minimal, increase in iron stores. However, this slight increase in hepatic iron can act as a cofactor that may modify the expression of other diseases such as porphyria cutanea tarda (PCT) or nonalcoholic steatohepatitis.

Mutations in other genes involved in iron metabolism are responsible for non-*HFE*-associated hemochromatosis, including juvenile hemochromatosis, which affects persons in the second and third decades of life (Table 357-1). Mutations in the genes encoding hepcidin, transferrin receptor 2 (TfR2), and hemojuvelin **(Fig. 357-1)** result in clinicopathologic features that are indistinguishable from HFE-associated hemochromatosis. However, mutations in ferroportin, responsible for the efflux of iron from enterocytes and most other cell types, result in iron loading of reticuloendothelial cells and macrophages as well as parenchymal cells.

Figure 357-1

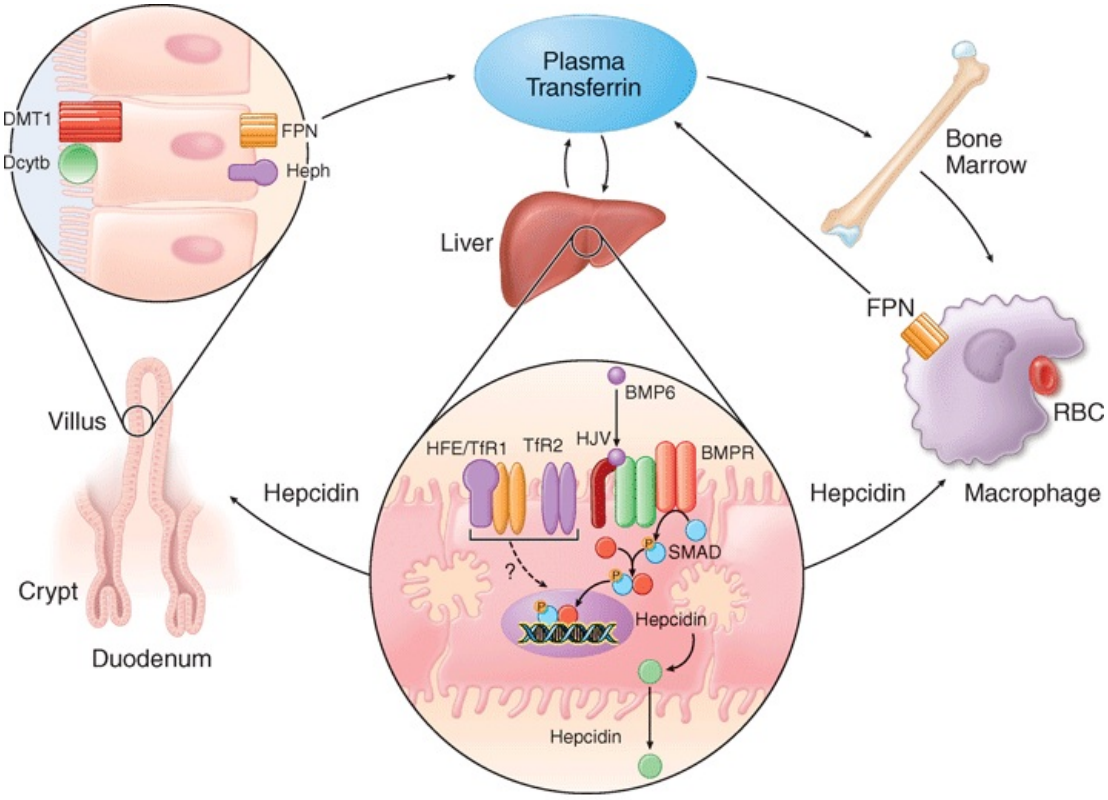

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathways of normaliron homeostasis. Dietary inorganic iron traverses the brush border membrane of duodenal enterocytes via DMT1 after reduction of ferric (Fe^{3+}) iron to the ferrous (Fe^{2+}) state by duodenal cytochrome B (DcytB). Iron then moves from the enterocyte to the circulation via a process requiring the basolateral iron exporter ferroportin (FPN) and the iron oxidase hephaestin (Heph). In the circulation, iron binds to plasma transferrin and is thereby distributed to sites of iron utilization and storage. Much of the diferric transferrin supplies iron to immature erythrocyte cells in the bone marrow for hemoglobin synthesis. At the end of their life, senescent red blood cells (RBC) are phagocytosed by macrophages and iron is returned to the circulation after export through ferroportin. The liver-derived peptide hepcidin represses basolateral iron transport in the gut as well as iron released from macrophages and other cells and serves as a central regulator of body-iron traffic. Hepcidin responds to changes in body-iron requirements by signals mediated by diferric transferrin through two mechanisms. One involves HFE and TfR2, while the other involves hemojuvelin (HJV) and the Bone Morphogenetic Protein (BMP)/SMAD pathway. Haem is metabolized by heme oxygenase within the enterocytes and the released iron then follows the same pathway.

Mutations in the genes encoding HFE, TfR2, hemojuvelin and hepcidin all lead to decreased hepcidin release and increased iron absorption, resulting in hemochromatosis (Table 357-1).

Pathophysiology

Normally, the body-iron content of 3–4 g is maintained such that intestinal mucosal absorption of iron is equal to iron loss. This amount is approximately 1 mg/d in men and 1.5 mg/d in menstruating women. In hemochromatosis, mucosal absorption is greater than body requirements and amounts to 4 mg/d or more. The progressive accumulation of iron increases plasma iron, saturation of transferrin, and results in a progressive increase of plasma ferritin **(Fig. 357-2)**. A liver-derived peptide, hepcidin, represses basolateral iron transport in the intestine and iron release from macrophages and other cells by binding to ferroportin. Hepcidin, in turn, responds to signals in the liver mediated by HFE, TfR2, and hemojuvelin (Fig. 357-1). Thus, hepcidin is a crucial molecule in iron metabolism, linking body stores with intestinal iron absorption.

Figure 357-2

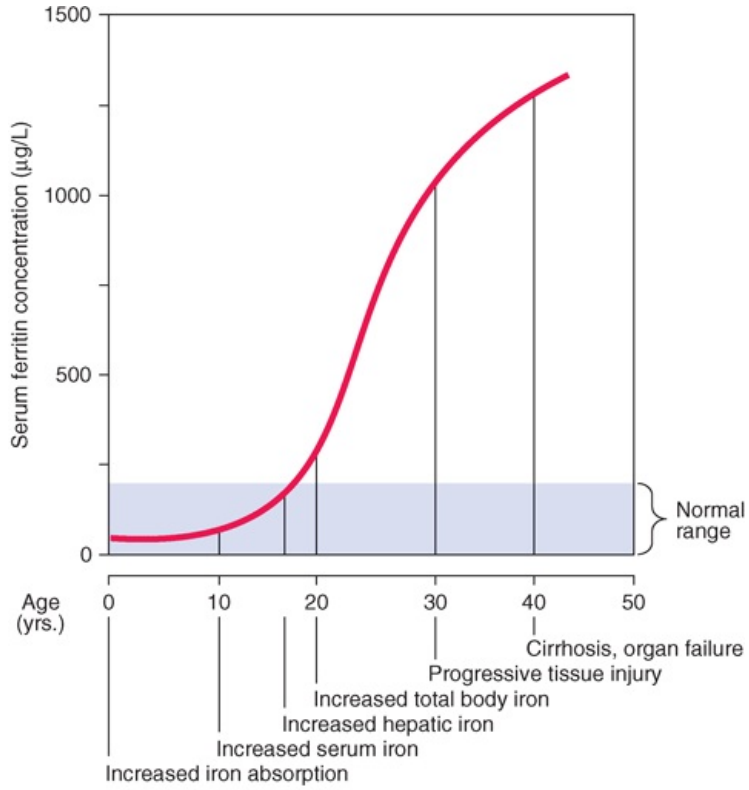

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sequence of events in genetic hemochromatosis and their correlation with the serum ferritin concentration. Increased iron absorption is present throughout life. Overt, symptomatic disease usually develops between ages 40 and 60, but latent disease can be detected long before this.

The *HFE* gene encodes a 343-amino-acid protein that is structurally related to MHC class I proteins. The basic defect in *HFE*-associated hemochromatosis is a lack of cell surface expression of HFE (due to the C282Y mutation). The normal (wild-type) HFE protein forms a complex with b_2-microglobulin and transferrin receptor 1 (TfR1). The C282Y mutation completely abrogates this interaction. As a result, the mutant HFE protein remains trapped intracellularly, reducing TfR1-mediated iron uptake by the intestinal crypt cell. This impaired TfR1-mediated iron uptake leads to upregulation of the divalent metal transporter (DMT1) on the brush border of the villus cells, causing inappropriately increased intestinal iron absorption (Fig. 357-1). In advanced disease, the body may contain 20 g or more of iron that is deposited mainly in parenchymal cells of the liver, pancreas, and heart. Iron may be increased 50- to 100-fold in the liver and pancreas and 5- to 25-fold in the heart. Iron deposition in the pituitary causes hypogonadotropic hypogonadism in both men and women. Tissue injury may result from disruption of iron-laden lysosomes, from lipid peroxidation of subcellular organelles by excess iron, or from stimulation of collagen synthesis by activated stellate cells.

Secondary iron overload with deposition in parenchymal cells occurs in chronic disorders of erythropoiesis, particularly in those due to defects in hemoglobin synthesis or ineffective erythropoiesis such as sideroblastic anemia and thalassemia (Chap. 104). In these disorders, iron absorption is increased. Moreover, these patients require blood transfusions and are frequently treated inappropriately with iron. PCT, a disorder characterized by a defect in porphyrin biosynthesis (Chap. 358), can also be associated with excessive parenchymal iron deposits. The magnitude of the iron load in PCT is usually insufficient to produce tissue damage. However, some patients with PCT also have mutations in the *HFE* gene, and some have associated hepatitis C virus (HCV) infection. Although the relationship between these disorders remains to be clarified, iron overload accentuates the inherited enzyme deficiency in PCT and should be avoided along with other agents (alcohol, estrogens, haloaromatic compounds) that may exacerbate PCT. Another cause of hepatic parenchymal iron overload is hereditary aceruloplasminemia. In this disorder, impairment of iron mobilization due to deficiency of ceruloplasmin (a ferroxidase) causes iron overload in hepatocytes.

Excessive iron ingestion over many years rarely results in hemochromatosis. An important exception has been reported in South Africa among groups who brew fermented beverages in vessels made of iron. Hemochromatosis has been described in apparently normal persons who have taken medicinal iron over many years, but such individuals probably had genetic disorders.

The common denominator in all patients with hemochromatosis is *excessive amounts of iron in parenchymal tissues*. Parenteral administration of iron in the form of blood transfusions or iron

preparations results predominantly in reticuloendothelial cell iron overload. This appears to lead to less tissue damage than iron loading of parenchymal cells.

In the liver, parenchymal iron is in the form of ferritin and hemosiderin. In the early stages, these deposits are seen in the periportal parenchymal cells, especially within lysosomes in the pericanalicular cytoplasm of the hepatocytes. This stage progresses to perilobular fibrosis and eventually to deposition of iron in bile-duct epithelium, Kupffer cells, and fibrous septa due to activation of stellate cells. In the advanced stage, a macronodular or mixed macro- and micronodular cirrhosis develops. Hepatic fibrosis and cirrhosis correlate significantly with hepatic iron concentration.

At autopsy, the enlarged nodular liver and pancreas are rusty in color. Histologically, iron is increased in many organs, particularly in the liver, heart, and pancreas, and, to a lesser extent, in the endocrine glands. The epidermis of the skin is thin, and melanin is increased in the cells of the basal layer and dermis. Deposits of iron are present around the synovial lining cells of the joints.

Clinical Manifestations

C282Y homozygotes can be characterized by the stage of progression as follows: (1) a genetic predisposition without abnormalities; (2) iron overload without symptoms; (3) iron overload with symptoms (e.g., arthritis and fatigue); and (4) iron overload with organ damage—in particular, cirrhosis. Thus, many subjects with significant iron overload are asymptomatic. For example, in a study of 672 asymptomatic C282Y homozygous subjects—identified by either family screening or routine health examinations—there was hepatic iron overload (grades 2–4) in 56% and 34.5% of male and female subjects, respectively; hepatic fibrosis (stages 2–4) in 18.4% and 5.4%, respectively; and cirrhosis in 5.6% and 1.9%, respectively.

Initial symptoms are often nonspecific and include lethargy, arthralgia, change in skin color, loss of libido, and features of diabetes mellitus. Hepatomegaly, increased pigmentation, spider angiomas, splenomegaly, arthropathy, ascites, cardiac arrhythmias, congestive heart failure, loss of body hair, testicular atrophy, and jaundice are prominent in advanced disease.

The *liver* is usually the first organ to be affected, and hepatomegaly is present in more than 95% of symptomatic patients. Hepatic enlargement may exist in the absence of symptoms or of abnormal liver-function tests. Manifestations of portal hypertension and esophageal varices occur less commonly than in cirrhosis from other causes. Hepatocellular carcinoma develops in about 30% of patients with cirrhosis, and it is the most common cause of death in treated patients—hence the importance of early diagnosis and therapy. The incidence increases with age, it is more common in men, and it occurs almost exclusively in cirrhotic patients.

Excessive skin pigmentation is present in patients with advanced disease. The characteristic metallic or slate-gray hue is sometimes referred to as *bronzing* and results from increased melanin and iron in the dermis. Pigmentation usually is diffuse and generalized, but it may be more pronounced on the face, neck, extensor aspects of the lower forearms, dorsa of the hands, lower legs, and genital regions, as well as in scars.

Diabetes mellitus occurs in about 65% of patients with advanced disease and is more likely to develop in those with a family history of diabetes, suggesting that direct damage to the pancreatic islets by iron deposition occurs in combination with other risk factors. The management is similar to that of other forms of diabetes, although insulin resistance is more common in association with hemochromatosis. Late complications are the same as seen in other causes of diabetes mellitus.

Arthropathy develops in 25–50% of symptomatic patients. It usually occurs after age 50 but may occur as a first manifestation, or long after therapy. The joints of the hands, especially the second and third metacarpophalangeal joints, are usually the first joints involved, a feature that helps to distinguish the chondrocalcinosis associated with hemochromatosis from the idiopathic form (Chap. 333). A progressive polyarthritis involving wrists, hips, ankles, and knees may also ensue. Acute brief attacks of synovitis may be associated with deposition of calcium pyrophosphate (chondrocalcinosis or pseudogout), mainly in the knees. Radiologic manifestations include cystic changes of the subchondral bones, loss of articular cartilage with narrowing of the joint space, diffuse demineralization, hypertrophic bone proliferation, and calcification of the synovium. The arthropathy tends to progress despite removal of iron by phlebotomy. Although the relation of these abnormalities to iron metabolism is not known, the fact that similar changes occur in other forms of iron overload suggests that iron is directly involved.

Cardiac involvement is the presenting manifestation in about 15% of symptomatic patients. The most common manifestation is congestive heart failure, which occurs in about 10% of young adults with the disease, especially those with juvenile hemochromatosis. Symptoms of congestive heart failure may develop suddenly, with rapid progression to death if untreated. The heart is diffusely enlarged; this may be misdiagnosed as idiopathic cardiomyopathy if other overt manifestations are absent. Cardiac arrhythmias include premature supraventricular beats, paroxysmal tachyarrhythmias, atrial flutter, atrial fibrillation, and varying degrees of atrioventricular block.

Hypogonadism occurs in both sexes and may antedate other clinical features. Manifestations include

loss of libido, impotence, amenorrhea, testicular atrophy, gynecomastia, and sparse body hair. These changes are primarily the result of decreased production of gonadotropins due to impairment of hypothalamic-pituitary function by iron deposition. Adrenal insufficiency, hypothyroidism, and hypoparathyroidism are rare manifestations.

DIAGNOSIS

The association of (1) hepatomegaly, (2) skin pigmentation, (3) diabetes mellitus, (4) heart disease, (5) arthritis, and (6) hypogonadism should suggest the diagnosis. However, as stated above, significant iron overload may exist with none or only some of these manifestations. Therefore, a high index of suspicion is needed to make the diagnosis early. Treatment before permanent organ damage occurs can reverse the iron toxicity and restore life expectancy to normal.

The history should be particularly detailed in regard to disease in other family members; alcohol ingestion; iron intake; and ingestion of large doses of ascorbic acid, which promotes iron absorption (Chap. 74). Appropriate tests should be performed to exclude iron deposition due to hematologic disease. The presence of liver, pancreatic, cardiac, and joint disease should be confirmed by physical examination, radiography, and standard function tests of these organs.

The degree of increase in total body–iron stores can be assessed by (1) measurement of serum iron and the percent saturation of transferrin (or the unsaturated iron-binding capacity), (2) measurement of serum ferritin concentration, (3) liver biopsy with measurement of the iron concentration and calculation of the hepatic iron index **(Table 357-2)**, and (4) MRI of the liver. In addition, a retrospective assessment of body-iron storage is also provided by performing weekly phlebotomy and calculating the amount of iron removed before iron stores are exhausted (1 mL blood = approximately 0.5 mg iron).

Table 357-2 Representative Iron Values in Normal Subjects, Patients with Hemochromatosis, and Patients with Alcoholic Liver Disease

Determination	Normal	Symptomatic Hemochromatosis	Homozygotes with Early, Asymptomatic Hemochromatosis	Heterozygotes	Alcoholic Liver Disease
Plasma iron, mmol/L (mg/dL)	9–27 (50–150)	32–54 (180–300)	Usually elevated	Elevated or normal	Often elevated
Total iron-binding capacity, mmol/L (mg/dL)	45–66 (250–370)	36–54 (200–300)	36–54 (200–300)	Elevated or normal	45–66 (250–370)
Transferrin saturation, percent	22–46	50–100	50–100	Normal or elevated	27–60
Serum ferritin, mg/L		900–6000	200–500	Usually <500	10–500
Men	20–250				
Women	15–150				
Liver iron, mg/g dry wt	300–1400	6000–18,000	2000–4000	300–3000	300–2000
Hepatic iron index	<1.0	>2	1.5–2	<2	<2

Each of these methods for assessing iron stores has advantages and limitations. The serum iron level and percent saturation of transferrin are elevated early in the course, but their specificity is reduced by significant false-positive and false-negative rates. For example, serum iron concentration may be increased in patients with alcoholic liver disease without iron overload; in this situation, however, the hepatic iron index is usually not increased as in hemochromatosis (Table 357-1). In otherwise healthy persons, a fasting serum transferrin saturation greater than 50% is abnormal and suggests homozygosity for hemochromatosis.

The serum ferritin concentration is usually a good index of body-iron stores, whether decreased or increased. In fact, an increase of 1 mg/L in serum ferritin level reflects an increase of about 5 mg in body stores. In most untreated patients with hemochromatosis, the serum ferritin level is significantly increased (Fig. 357-2 and Table 357-1), and a serum ferritin level >1000 mg/L is the strongest predictor of disease expression among individuals homozygous for the C282Y mutation. However, in patients with inflammation and hepatocellular necrosis, serum ferritin levels may be elevated out of proportion to body-iron stores due to increased release from tissues. Therefore, a repeat determination of serum ferritin should be carried out after acute hepatocellular damage has subsided (e.g., in alcoholic liver disease). Ordinarily, the combined measurements of the percent transferrin saturation and serum ferritin level provide a simple and reliable screening test for hemochromatosis, including the

precirrhotic phase of the disease. If either of these tests is abnormal, genetic testing for hemochromatosis should be performed **(Fig. 357-3)**.

Figure 357-3

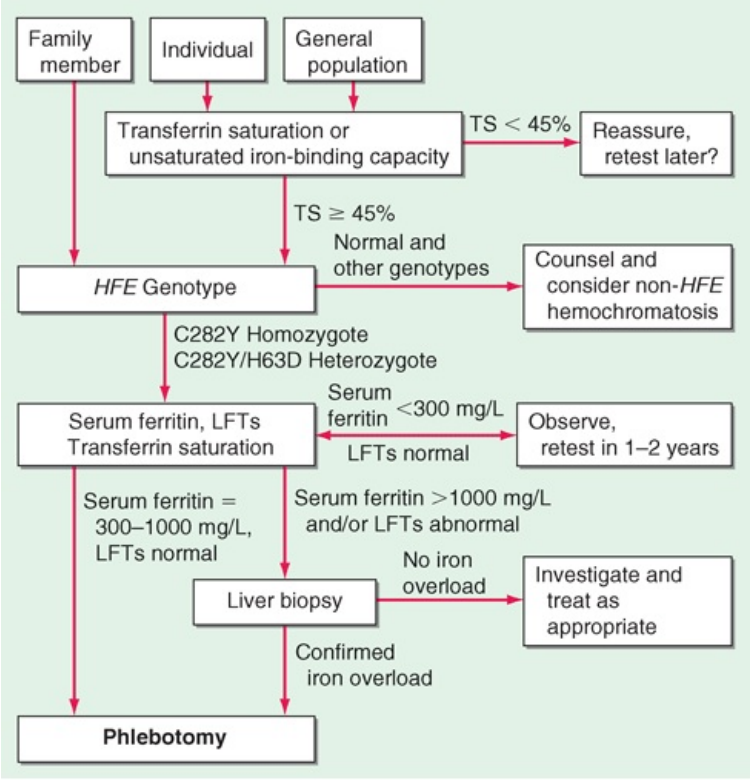

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for screening for*HFE*-associated hemochromatosis. LFT, liver function test; TS, transferrin saturation. *(From EJ Eijkelkamp et al: Can J Gastroenterol 14:121, 2000; with permission.)*

The role of liver biopsy in the diagnosis and management of hemochromatosis has been reassessed as a result of the widespread availability of genetic testing for the C282Y mutation. The absence of severe fibrosis can be accurately predicted in most patients using clinical and biochemical variables. Thus, there is virtually no risk of severe fibrosis in a C282Y homozygous subject with (1) serum ferritin level less than 1000 mg/L, (2) normal serum alanine amino transaminase values, (3) no hepatomegaly, and (4) no excess alcohol intake. However, it should be emphasized that liver biopsy is the only reliable method for establishing or excluding the presence of hepatic cirrhosis, which is the critical factor determining prognosis and the risk of developing hepatocellular carcinoma. Biopsy also permits histochemical estimation of tissue iron and measurement of hepatic iron concentration. Increased density of the liver due to iron deposition can be demonstrated by CT or MRI, and, with improved technology, MRI has become more accurate in determining hepatic iron concentration.

Screening for Hemochromatosis

When the diagnosis of hemochromatosis is established, it is important to counsel and screen other family members (Chap. 63). Asymptomatic as well as symptomatic family members with the disease usually have an increased saturation of transferrin and an increased serum-ferritin concentration. These changes occur even before the iron stores are greatly increased (Fig. 357-2). All adult first-degree relatives of patients with hemochromatosis should be tested for the C282Y and H63D mutations and counseled appropriately (Fig. 357-3). In affected individuals, it is important to confirm or exclude the presence of cirrhosis and begin therapy as early as possible. When children of a proband are affected, a homozygote-heterozygote mating is most likely. For practical purposes, children need not be checked before they are 18 years old.

The role of population screening for hemochromatosis is controversial. Recent studies indicate that it is highly effective for primary care physicians to screen subjects using serum iron, transferrin saturation, and serum ferritin levels. Such screening also detects iron deficiency. Genetic screening of the normal population is feasible, but its cost-effectiveness has not been established.

Treatment: Hemochromatosis

The therapy of hemochromatosis involves removal of the excess body iron and supportive treatment of damaged organs. Iron removal is best accomplished by weekly or twice-weekly phlebotomy of 500 mL. Although there is an initial modest decline in the volume of packed red blood cells to about 35 mL/dL, the level stabilizes after several weeks. The plasma transferrin saturation remains increased until the available iron stores are depleted. In contrast, the plasma ferritin concentration falls progressively, reflecting the gradual decrease in body-iron stores. One 500 mL unit of blood contains 200–250 mg iron, and up to 25 g iron or more may have to be removed. Therefore, in patients with advanced disease, weekly phlebotomy may be required for 1–2 years, and it should be continued until the serum ferritin level is under 50 mg/L. Thereafter, phlebotomies are performed at appropriate intervals to maintain ferritin levels between 50 and 100 mg/L. Usually one phlebotomy every three months will suffice.

Chelating agents such as deferoxamine, when given parenterally, remove 10–20 mg iron per day, which is much less than that mobilized by once-weekly phlebotomy. Phlebotomy is also less expensive, more convenient, and safer for most patients. However, chelating agents are indicated when anemia or hypoproteinemia is severe enough to preclude phlebotomy. Subcutaneous infusion of deferoxamine using a portable pump is the most effective means of its administration.

An effective oral ironchelating agent, deferasirox (Exjade), has recently become available but is still in clinical trials. This agent is effective in thalassemia and secondary iron overload, but its role in primary iron overload has yet to be established.

Alcohol consumption should be severely curtailed or eliminated because it increases the risk of cirrhosis in hereditary hemochromatosis nearly tenfold. The management of hepatic failure, cardiac failure, and diabetes mellitus is similar to conventional therapy for these conditions. Loss of libido and change in secondary sex characteristics are managed with testosterone replacement or gonadotropin therapy (Chap. 346).

End-stage liver disease may be an indication for liver transplantation, although results are improved if the excess iron can be removed beforehand. The available evidence indicates that the fundamental metabolic abnormality in hemochromatosis is reversed by successful liver transplantation.

Prognosis

The principal causes of death are cardiac failure, hepatocellular failure or portal hypertension, and hepatocellular carcinoma.

Life expectancy is improved by removal of the excessive stores of iron and maintenance of these stores at near-normal levels. The 5-year survival rate with therapy increases from 33 to 89%. With repeated phlebotomy, the liver decreases in size, liver function improves, pigmentation of skin decreases, and cardiac failure may be reversed. Diabetes improves in about 40% of patients, but removal of excess iron has little effect on hypogonadism or arthropathy. Hepatic fibrosis may decrease, but established cirrhosis is irreversible. Hepatocellular carcinoma occurs as a late sequela in patients who are cirrhotic at presentation. The apparent increase in its incidence in treated patients is probably related to their increased life span. Hepatocellular carcinoma rarely develops if the disease is treated in the precirrhotic stage. Indeed, the life expectancy of homozygotes treated before the development of cirrhosis is normal.

The importance of family screening and early diagnosis and treatment cannot be overemphasized. Asymptomatic individuals detected by family studies should have phlebotomy therapy if iron stores are moderately to severely increased. Assessment of iron stores at appropriate intervals is also important. With this management approach, most manifestations of the disease can be prevented.

Role of *HFE* Mutations in Other Liver Diseases

There is considerable interest in the role of *HFE* mutations and hepatic iron in several other liver diseases. Several studies have shown an increased prevalence of *HFE* mutations in PCT patients. Iron accentuates the inherited enzyme deficiency in PCT and clinical manifestations of PCT. The situation in nonalcoholic steatosis (NASH) is less clear, but some studies have shown an increased prevalence of *HFE* mutations in NASH patients. The role of phlebotomy therapy, however, is unproven. In chronic HCV infection, *HFE* mutations are not more common, but some subjects have increased hepatic iron. Before initiating antiviral therapy in these patients, it is reasonable to perform phlebotomy therapy to remove excess iron stores, because this reduces liver enzyme levels.

HFE mutations are not increased in frequency in alcoholic liver disease. Hemochromatosis in a heavy drinker can be distinguished from alcoholic liver disease by the presence of the C282Y mutation.

End-stage liver disease may also be associated with iron overload of the degree seen in hemochromatosis. The mechanism is uncertain, although recent evidence suggests that alcohol suppresses hepatic hepcidin secretion. Hemolysis also plays a role. *HFE* mutations are uncommon.

A recent large population study has shown that subjects homozygous for C282Y are at increased risk of breast and colorectal cancer.

FURTHER READINGS

Adams PC et al: Hemochromatosis and iron-overload screening in a racially diverse population. N Engl J Med 352:1769, 2005[PMID: 15858186] [Full Text]

Allen KJ et al: Iron-overload-related disease in *HFE* hereditary hemochromatosis. New Engl J Med 358:221, 2008[PMID: 18199861] [Full Text]

Crawford DH et al: Serum hyaluronic acid with serum ferritin accurately predicts cirrhosis and reduces the need for liver biopsy in C282Y hemochromatosis. Hepatology 49:418, 2009[PMID: 19177571] [Full Text]

Gurrin LC et al: HFE C282Y/H63D compound heterozygotes are at low risk of hemochromatosis-related morbidity. Hepatology 50:94, 2009[PMID: 19554541] [Full Text]

Osborne NJ et al: HFE C282Y homozygotes are at increased risk of breast and colorectal cancer. Hepatology 51:1311, 2010[PMID: 20099304] [Full Text]

Pietrangelo A: Non-HFE-linked hemochromatosis. Semin Liver Dis 25:450, 2005[PMID: 16315138] [Full Text]

Powell LW et al: Screening for hemochromatosis in asymptomatic subjects with or without a family history. Arch Intern Med 166:294, 2006[PMID: 16476869] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE PORPHYRIAS: INTRODUCTION

The porphyrias are metabolic disorders, each resulting from the deficiency of a specific enzyme in the heme biosynthetic pathway **(Fig. 358-1** and **Table 358-1)**. These enzyme deficiencies are inherited as autosomal dominant, autosomal recessive, or X-linked traits, with the exception of porphyria cutanea tarda (PCT), which usually is sporadic **(Table 358-1)**. The porphyrias are classified as either *hepatic* or *erythropoietic*, depending on the primary site of overproduction and accumulation of their respective porphyrin precursors or porphyrins **(Tables 358-1 and 358-2)**, although some have overlapping features. For example, PCT, the most common porphyria, is hepatic and presents with blistering cutaneous photosensitivity, which is usually a manifestation of the erythropoietic porphyrias. The major manifestations of the acute hepatic porphyrias are neurologic, including neuropathic abdominal pain, peripheral motor neuropathy, and mental disturbances, with attacks often precipitated by steroid hormones, certain drugs, and nutritional changes such as dieting. While hepatic porphyrias are symptomatic primarily in adults, rare homozygous variants of the autosomal dominant hepatic porphyrias usually manifest clinically prior to puberty.

Figure 358-1

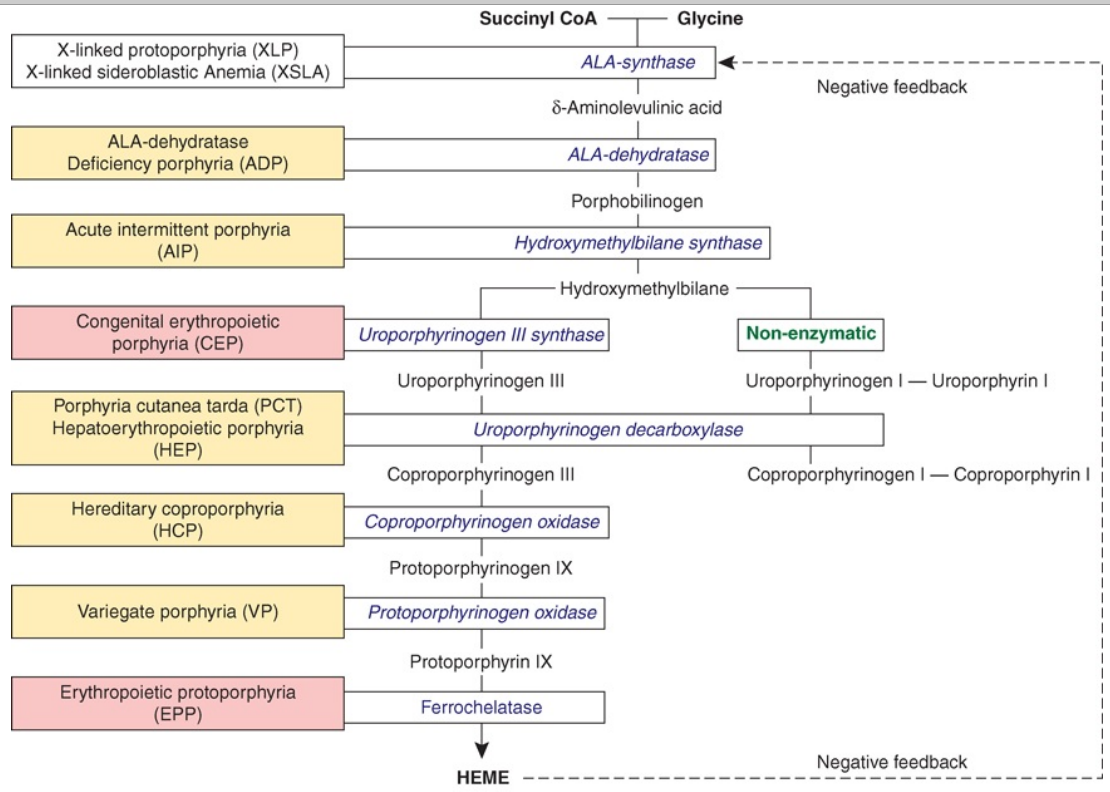

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The human heme biosynthetic pathway indicating in linked boxes the enzyme that, when deficient, causes the respective porphyria. Hepatic porphyrias are shown in yellow boxes and erythropoietic porphyrias in pink boxes.

Table 358-1 Human Porphyrias: Major Clinical and Laboratory Features

			Principal Symptoms	Increased Porphyrin Precursors and/or Porphyrins			
Porphyria	**Deficient Enzyme**	**Inheritance**	**NV or CP+**	**Enzyme Activity % of Normal**	**Erythrocytes**	**Urine**	**Stool**
Hepatic porphyrias							
5-ALA dehydratase-deficient porphyria (ADP)	ALA-dehydratase	AR	NV	~5	Zn-Protoporphyrin	ALA, Coproporphyrin III	—
Acute intermittent porphyria (AIP)	HMB-synthase	AD	NV	~50	—	ALAa, PBG, Uroporphyrin	—
Porphyria cutanea tarda (PCT)	URO-decarboxylase	AD	CP	~20	—	Uroporphyrin, 7-carboxylate porphyrin	Isocoproporphyrin
Hereditary	COPRO-	AD	NV & CP	~50	&	ALA, PBG,	Coproporphyrin III

coproporphyria (HCP)	oxidase					Coproporphyrin III	
Variegate porphyria (VP)	PROTO-oxidase	AD	NV & CP	~50	—	ALA, PBG, Coproporphyrin III	Coproporphyrin III Protoporphyrin
Erythropoietic porphyrias							
Congenital erythropoietic porphyria (CEP)	URO-synthase	AR	CP	1–5	Uroporphyrin I Coproporphyrin I	Uroporphyrin Ib Coproporphyrin I[b]	Coproporphyrin I
Erythropoietic protoporphyria (EPP)	Ferrochelatase	AR[a]	CP	~20–30	Protoporphyrin	—	Protoporphyrin
X-linked protoporphyria (XLP)	ALA-synthase 2	XL	CP	>100[c]	Protoporphyrin	—	Protoporphyrin

[a]Polymorphism in intron 3 of wild-type allele affects level of enzyme activity and clinical expression.

[b]Type I isomers.

[c]Increased activity due to "gain-of-function" mutations in ALAS2 exon 11.

Abbreviations: AD, autosomal dominant; ALA, 5-aminolevulinic acid; AR, autosomal recessive; COPRO I, coproporphyrin I; COPRO III, coproporphyirn III; CP, cutaneous photosensitivity; ISOCOPRO, isocoproporphyrin; + Nv, neurovisceral; PBG, porphobilinogen; PROTO, protoporphyrin IX; URO I, uroporphyrin I; URO III, uroporphyrin III; XL, X-linked.

Table 358-2 Human Heme Biosynthetic Enzymes and Genes

Enzyme	Gene Symbol	Chromosomal Location	cDNA (bp)	Gene Size (kb)	Gene Exons[a]	Protein(aa)	SubcellularLocation	Known Mutations[b]	3DStructure[c]
ALA-synthase Housekeeping	ALAS1	3p21.1	2199	17	11	640	M	—	
Erythroid-specific	ALAS2	Xp11.2	1937	22	11	587	M	>30	—
ALA-dehydratase Housekeeping	ALAD	9q32	1149	15.9	12 (1A + 2 – 12)	330	C	12	Y
Erythroid-specific	ALAD	9q32	1154	15.9	12 (1B + 2 – 12)	330	C	—	
HMB-synthase Housekeeping	HMBS	11q23.3	1086	11	15 (1 + 3 – 15)	361	C	>315	E
Erythroid-specific	HMBS	11q23.3	1035	11	15 (2 – 15)	344	C	10	
URO-synthase Housekeeping	UROS	10q26.2	1296	34	10 (1 + 2B – 10)	265	C	39	H
Erythroid-specific	UROS	10q26.2	1216	34	10 (2A + 2B – 10)	265	C	4	
URO-decarboxylase	UROD	1p34.1	1104	3	10	367	C	108	H
COPRO-oxidase	CPOX	3q12.1	1062	14	7	354	M	51	H
PROTO-oxidase	PPOX	1q23.3	1431	5.5	13	477	M	129	—
Ferrochelatase	FECH	18q21.31	1269	45	11	423	M	125	B

[a]Number of exons and those encoding separate housekeeping and erythroid-specific forms indicated in parentheses.

[b]Number of known mutations from the Human Gene Mutation Database (*www.hgmd.org*).

[c]Crystallized from human (H), murine (M), *Escherichia coli* (E), *Bacillus subtilis* (B), or yeast (Y) purified enzyme; references in Protein Data Bank (*www.rcsb.org*).

Abbreviations: C, cytoplasm; M, mitochondria.

Source: From Anderson et al.

In contrast, the erythropoietic porphyrias usually present with cutaneous photosensitivity at birth or in early childhood, or in the case of congenital erythropoietic porphyria (CEP), even in utero as nonimmune hydrops fetalis. Cutaneous sensitivity to sunlight results from excitation of excess porphyrins in the skin by long-wave ultraviolet light, leading to cell damage, scarring, and disfigurement. Thus, the porphyrias are metabolic disorders in which genetic, physiologic, and environmental factors interact to cause disease.

Because many symptoms of the porphyrias are nonspecific, diagnosis is often delayed. Laboratory measurement of porphyrin precursors [5′-aminolevulinic acid (ALA) and porphobilinogen (PBG)] or porphyrins in urine, plasma, erythrocytes, or feces is required to confirm or exclude the various types of porphyria (see below). However, a definite diagnosis requires demonstration of the specific enzyme deficiency or gene defect **(Table 358-3)**. The isolation and characterization of the genes encoding all the heme biosynthetic enzymes has permitted identification of the mutations causing each porphyria (Table 358-2). Molecular

genetic analyses now make it possible to provide precise heterozygote or homozygote identification and prenatal diagnoses in families with known mutations.

Table 358-3 Diagnosis of Acute and Cutaneous Porphyrias

Symptoms	First-Line Test: Abnormality	Possible Porphyria:	Second-Line Testing if First-Line Testing Is Positive: To include: urine (U), plasma (P), and fecal (F) porphyrins; for acute porphyrias add red blood cell (RBC) HMB-synthase; for blistering skin lesions add P & RBC porphyrins	Confirmatory Test: Enzyme Assay and/or Mutation Analysis
Neurovisceral	Spot U: ↑↑ ALA & normal PBG	ADP	U porphyrins: ↑↑, mostly COPRO III P & F porphyrins: normal or slightly ↑ RBC HMB-synthase: normal	Rule out other causes of elevated ALA; ↘↘ RBC ALA-dehydratase activity (<10%); ALA-dehydratase mutation analysis
	Spot U: ↑↑ PBG	AIP	U porphyrins: ↑↑, mostly URO & COPROP & F porphyrins: normal or slightly ↑ RBC HMB-synthase: usually ↘	HMB-synthase mutation analysis
	"	HCP	U porphyrins: ↑↑, mostly COPRO III P porphyrins: normal or slightly ↑ (↑ if skin lesions present) F porphyrins: ↑↑, mostly COPRO III	Measure RBC HMB-synthase: normal activity COPRO-oxidase mutation analysis
Blistering Skin Lesions	" P: ↑ porphyrins	VP PCT & HEP	U porphyrins: ↑↑, mostly COPRO III P porphyrins: ↑↑ (characteristic fluorescence peak at neutral pH) F porphyrins: ↑↑, mostly COPRO & PROTO U porphyrins: ↑↑, mostly URO & heptacarboxylate porphyrin P porphyrins: ↑↑ F porphyrins: ↑↑, including increased isocoproporphyrin RBC porphyrins: ↑↑ zinc PROTO in HEP[a]	Measure RBC HMB-synthase: normal activity PROTO-oxidase mutation analysis RBC URO-decarboxylase activity: half-normal in familial PCT (~20% of all PCT cases); substantially deficient in HEP URO-decarboxylase mutation analysis: mutation(s) present in familial PCT (heterozygous) &HEP (homozygous)
	"	HCP & VP	See HCP &VP above. Also, U ALA & PBG: may be ↑	
	"	CEP	RBC & U porphyrins: ↑↑, mostly URO I and COPRO I F porphyrins: ↑↑; mostly COPRO I	↘↘RBC URO-synthase activity (<15%) URO-synthase mutation analysis
Nonblistering Photosensitivity	P: porphyrins usually ↑	EPP	RBC porphyrins: ↑↑, mostly free PROTO U porphyrins: normal F porphyrins: normal or ↑, mostly PROTO	FECH mutation analysis
	P: porphyrins usually ↑	XLP	RBC porphyrins: ↑↑, approx equal free and zinc PROTO U porphyrins: normal F porphyrins: normal or ↑, mostly PROTO	ALA S2 mutation analysis

[a]Nonspecific increases in zinc protoporphyrins are common in other porphyrias

Abbreviations: ALA, 5-aminolevulinic acid; COPRO I, coproporphyrin I; COPRO III, coproporphyrin III; ISOCOPRO, isocoproporphyrin; F, fecal; P, plasma; PBG, porphobilinogen; PROTO, protoporphyrin IX; RBC, erythrocytes; U, urine; URO I, uroporphyrin I; URO III, uroporphyrin III.

Source: Based on KE Anderson et al: Ann Intern Med 142:439, 2005.

In addition to recent reviews of the porphyrias (see Further Readings), informative and up-to-date websites are sponsored by the American Porphyria Foundation (*www.porphyriafoundation.com*) and the European Porphyria Initiative (*www.porphyria-europe.org*). An extensive list of unsafe and safe drugs for individuals with acute porphyrias is provided at the Drug Database for Acute Porphyrias (*www.drugs-porphyria.com*).

Heme Biosynthesis

Heme biosynthesis involves eight enzymatic steps in the conversion of glycine and succinyl-CoA to heme (**Fig. 358-2** and Table 358-2). These eight enzymes are encoded by nine genes, as the first enzyme in the pathway, 5′-aminolevulinate synthase (ALA-synthase), has two genes that encode unique housekeeping (ALAS1) and erythroid- specific (ALAS2) isozymes. The first and last three enzymes in the pathway are located in the mitochondrion, whereas the other four are in the cytosol. Heme is required for a variety of hemoproteins such as hemoglobin, myoglobin, respiratory cytochromes, and the cytochrome P450 enzymes (CYPs). Hemoglobin synthesis in erythroid precursor cells accounts for approximately 85% of daily heme synthesis in humans. Hepatocytes account for most of the rest, primarily for the synthesis of CYPs, which are especially abundant in the liver endoplasmic reticulum, and turn over more rapidly than many other hemoproteins, such as the mitochondrial respiratory cytochromes. As shown in Fig. 358-2, pathway intermediates are the porphyrin precursors, ALA and PBG, and porphyrins (mostly in their reduced forms, known as *porphyrinogens*). At least in humans, these intermediates do not accumulate in significant amounts under normal conditions or have important physiologic functions.

Figure 358-2

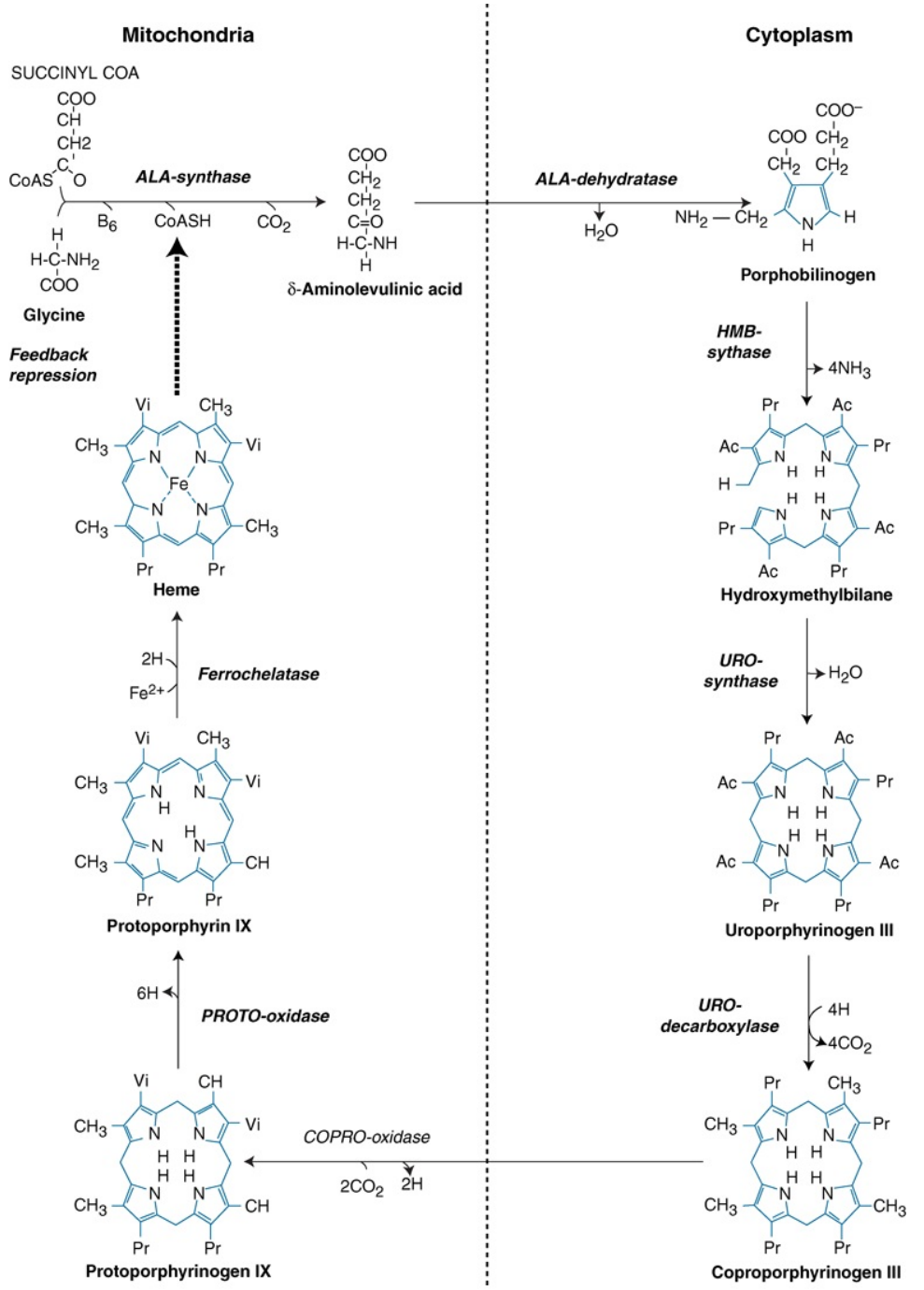

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The heme biosynthetic pathway showing the eight enzymes and their substrates and products. Four of the enzymes are localized in the mitochondria and four in the cytosol.

The first enzyme, ALA-synthase, catalyzes the condensation of glycine, activated by pyridoxal phosphate and succinyl coenzyme A, to form ALA. In the liver, this rate-limiting enzyme can be induced by a variety of drugs, steroids, and other chemicals. Distinct nonerythroid (e.g., housekeeping) and erythroid-specific forms of ALA-synthase are encoded by separate genes located on chromosomes 3p21.1 (ALAS1) and Xp11.2 (ALAS2), respectively. Defects in the erythroid gene ALAS2 that decrease its activity cause X-linked sideroblastic anemia (XLSA). Recently, gain of function mutations in exon 11 of ALAS2 that increase its activity have been shown to cause an X-linked form of erythropoietic protoporphyria (EPP), known as X-linked protoporphyria (XLP).

The second enzyme, 5′-aminolevulinate dehydratase (ALA-dehydratase), catalyzes the condensation of two molecules of ALA to form PBG. Hydroxymethylbilane synthase (HMB-synthase; also known as PBG-deaminase) catalyzes the head-to-tail condensation of four PBG molecules by a series of deaminations to form the linear tetrapyrrole, HMB. Uroporphyrinogen III synthase (URO-synthase) catalyzes the rearrangement and rapid cyclization of HMB to form the asymmetric, physiologic, octacarboxylate porphyrinogen uroporphyrinogen (URO'gen) III.

The fifth enzyme in the pathway, uroporphyrinogen decarboxylase (URO-decarboxylase), catalyzes the sequential removal of the four carboxyl groups from the acetic acid side chains of URO'gen III to form coproporphyrinogen (COPRO'gen) III, a tetracarboxylate porphyrinogen. This compound then enters the mitochondrion via a specific transporter, ABCB6, where COPRO-oxidase, the sixth enzyme, catalyzes the decarboxylation of two of the four propionic acid groups to form the two vinyl groups of protoporphyrinogen (PROTO'gen) IX, a decarboxylate porphyrinogen. Next, PROTO-oxidase oxidizes PROTO IX'gen to protoporphyrin IX by the removal of six hydrogen atoms. The product of the reaction is a porphyrin (oxidized form), in

contrast to the preceding tetrapyrrole intermediates, which are porphyrinogens (reduced forms). Finally, ferrous iron is inserted into protoporphyrin IX to form heme, a reaction catalyzed by the eighth enzyme in the pathway, ferrochelatase (also known as heme synthetase or protoheme ferrolyase).

Regulation of Heme Biosynthesis

Regulation of heme synthesis differs in the two major heme-forming tissues, the liver and erythron. In the liver, the concentration of "free" heme regulates the synthesis and mitochondrial translocation of the housekeeping form of ALA-synthase 1. Heme represses the synthesis of the ALA-synthase 1 mRNA and interferes with the transport of the enzyme from the cytosol into mitochondria. Hepatic ALA-synthase 1 is increased by many of the same chemicals that induce the cytochrome P450 enzymes in the endoplasmic reticulum of the liver. Because most of the heme in the liver is used for the synthesis of cytochrome P450 enzymes, hepatic ALA-synthase 1 and the cytochrome P450s are regulated in a coordinated fashion, and many drugs that induce hepatic ALA-synthase 1 also induce the CYPs. The other hepatic heme biosynthetic enzymes are presumably expressed at constant levels, although their relative activities and kinetic properties differ. For example, normal individuals have high activities of ALA-dehydratase, but low activities of HMB-synthase, the latter being the second rate-limiting step in the pathway.

In the erythron, novel regulatory mechanisms allow for the production of the very large amounts of heme needed for hemoglobin synthesis. The response to stimuli for hemoglobin synthesis occurs during cell differentiation, leading to an increase in cell number. The erythroid-specific ALA-synthase 2 is expressed at higher levels than the housekeeping enzyme, and erythroid-specific control mechanisms regulate other pathway enzymes as well as iron transport into erythroid cells. Separate erythroid-specific and nonerythroid or "housekeeping" transcripts are known for the first four enzymes in the pathway. As noted above, housekeeping and erythroid-specific ALA-synthases are encoded by genes on different chromosomes, but for each of the next three genes in the pathway, both erythroid and nonerythroid transcripts are transcribed by alternative promoters from their single respective genes (Table 358-2).

Classification of the Porphyrias

As mentioned above, the porphyrias can be classified as either *hepatic* or *erythropoietic*, depending on whether the heme biosynthetic intermediates that accumulate arise initially from the liver or developing erythrocytes, or as *acute* or *cutaneous*, based on their clinical manifestations. Table 358-1 lists the porphyrias, their principal symptoms, and major biochemical abnormalities. Four of the five hepatic porphyrias—acute intermittent porphyria (AIP), hereditary coproporphyria (HCP), variegate porphyria (VP), and ALA-dehydratase porphyria (ADP)—present during adult life with acute attacks of neurologic manifestations and elevated levels of one or both of the porphyrin precursors, ALA and PBG, and are thus classified as *acute porphyrias*. Patients with ADP have also presented in infancy and adolescence. The fifth hepatic disorder, porphyria cutanea tarda (PCT), presents with blistering skin lesions. HCP and VP also may have cutaneous manifestations similar to PCT.

The erythropoietic porphyrias—congenital erythropoietic porphyria (CEP) and erythropoietic protoporphyria (EPP), including the recently described X-linked form, XLP—are characterized by elevations of porphyrins in bone marrow and erythrocytes and present with cutaneous photosensitivity. The skin lesions in CEP resemble PCT but are usually much more severe, whereas EPP and XLP cause a more immediate, painful, and nonblistering type of photosensitivity. EPP is the most common porphyria to cause symptoms before puberty. About 20% of EPP patients develop minor abnormalities of liver function, with up to about 5% developing hepatic complications that can become life-threatening. XLP has a clinical presentation similar to EPP causing photosensitivity and liver disease.

Diagnosis of Porphyria

A few specific and sensitive first-line laboratory tests should be used whenever symptoms or signs suggest the diagnosis of porphyria (Table 358-3). If a first-line test is significantly abnormal, more comprehensive testing should follow to establish the type of porphyria, and finally the specific causative gene mutation.

ACUTE PORPHYRIAS

An acute porphyria should be suspected in patients with neurovisceral symptoms after puberty, such as abdominal pain, and when the initial clinical evaluation does not suggest another cause, the urinary porphyrin precursors (ALA and PBG) should be measured (Fig. 358-2). Urinary PBG is virtually always increased during acute attacks of AIP, HCP, and VP and is not substantially increased in any other medical condition. Therefore, this measurement is both sensitive and specific. A method for rapid, in-house testing for urinary PBG, such as the Trace PBG kit (Trace America/Trace Diagnostics, Louisville, Colorado), can be used. Results from spot (single void) urine specimens are highly informative because very substantial increases in PBG are expected during acute attacks of porphyria. A 24-h collection can unnecessarily delay diagnosis. The same spot urine specimen should be saved for quantitative determination of ALA, PBG, and creatinine, in order to confirm the qualitative PBG result and also to detect patients with ALAD-deficient porphyria (ADP). Urinary porphyrins may remain increased longer than porphyrin precursors in HCP and VP. Therefore, it is useful to measure total urinary porphyrins in the same sample, keeping in mind that urinary porphyrin increases are often nonspecific. Measurement of urinary porphyrins alone should be avoided for screening, because these may be increased in disorders other than porphyrias, such as chronic liver disease, and misdiagnoses of porphyria can result from minimal increases in urinary porphyrins that have no diagnostic significance. Measurement of erythrocyte HMB-synthase is not useful as a first-line test in the acute setting because it does not differentiate latent from active AIP. Moreover, the enzyme activity is not decreased in all AIP patients and is never deficient in other acute porphyrias.

CUTANEOUS PORPHYRIAS

Blistering skin lesions due to porphyria are virtually always accompanied by increases in total plasma porphyrins. A fluorometric method is preferred, because the porphyrins in plasma in VP are mostly covalently linked to plasma proteins and may be less readily detected by HPLC. The normal range for plasma porphyrins is somewhat increased in patients with end-stage renal disease.

Although a total plasma porphyrin determination will usually detect EPP and XLP, an erythrocyte protoporphyrin determination is more sensitive. Increases in erythrocyte protoporphyrin occur in many other conditions. Therefore, the diagnosis of EPP must be confirmed by showing a predominant increase in free protoporphyrin rather than zinc protoporphyrin. In XLP, both free and zinc protoporphyrin are markedly increased in approximately equal proportions. Interpretation of laboratory reports can be difficult, because the term *free erythrocyte protoporphyrin* sometimes actually represents zinc protoporphyrin.

More extensive testing is justified when an initial test is positive. A substantial increase in PBG may be due to AIP, HCP, or VP. These acute porphyrias can be distinguished by measuring erythrocyte HMB-synthase, urinary porphyrins (using the same spot urine sample), fecal porphyrins, and plasma porphyrins. Assays for COPRO-oxidase or PROTO-oxidase are not widely available. Alternatively, sequencing the genes encoding HMB-synthase, COPRO-oxidase, and PROTO-oxidase will detect most disease-causing mutations, and will be diagnostic even when the levels of urinary ALA and PBG have returned to normal or near-normal levels. The various porphyrias that cause blistering skin lesions are differentiated by measuring porphyrins in urine, feces, and plasma. These porphyrias can also be confirmed at the DNA level by the demonstration of the causative

mutation(s). It is often difficult to diagnose or "rule out" porphyria in patients who have had suggestive symptoms months or years in the past, and in relatives of patients with acute porphyrias, because porphyrin precursors and porphyrins may be normal. In those situations, detection of the specific gene mutation in the index case can make the diagnosis. Consultation with a specialist laboratory and physician is useful for selecting which heme biosynthetic gene or genes to be sequenced. Before evaluating relatives, the gene-based diagnosis of porphyria should be established in an index case. Once a specific mutation is identified in an affected member, other at-risk relatives can be offered genetic counseling and testing for that mutation. Further details are provided in the sections below on each type of porphyria.

THE HEPATIC PORPHYRIAS

Markedly elevated plasma and urinary concentrations of the porphyrin precursors ALA or PBG, which originate from the liver, are especially evident during attacks of neurologic manifestations of the four acute porphyrias, ADP, AIP, HCP, and VP. In PCT, excess porphyrins also accumulate initially in the liver and cause chronic blistering of sun-exposed areas of the skin.

Ala-Dehydratase Deficient Porphyria (Adp)

ADP is a rare autosomal recessive acute hepatic porphyria caused by a severe deficiency of ALA-dehydratase activity. To date, there are only a few documented cases, some in children or young adults, in which specific gene mutations have been identified. These affected homozygotes had <10% of normal ALA-dehydratase activity in erythrocytes, but their clinically asymptomatic parents and heterozygous relatives had about half-normal levels of activity and did not excrete increased levels of ALA. The frequency of ADP is unknown, but the frequency of heterozygous individuals with <50% normal ALA-dehydratase activity was ~2% in a screening study in Sweden. Because there are multiple causes for deficient ALA-dehydratase activity, it is important to confirm the diagnosis of ADP by mutation analysis.

CLINICAL FEATURES

The clinical presentation depends on the amount of residual ALA-dehydratase activity. Four of the six documented patients were male adolescents with symptoms resembling those of AIP, including abdominal pain and neuropathy. One patient was an infant with more severe disease, including failure to thrive beginning at birth. The earlier age of onset and more severe manifestations in this patient reflect a more significant deficiency of ALA-dehydratase activity. Another patient developed an acute motor polyneuropathy at age 63 that was associated with a myeloproliferative disorder. He was heterozygous for an ALAD mutation that presumably was present in erythroblasts that underwent clonal expansion due to the bone marrow malignancy.

DIAGNOSIS

All patients had significantly elevated levels of plasma and urinary ALA and urinary coproporphyrin (COPRO) III; ALAD activities in erythrocytes were <10% of normal. Hereditary tyrosinemia type 1 (fumarylacetoacetase deficiency) and lead intoxication should be considered in the differential diagnosis since either succinylacetone (which accumulates in hereditary tyrosinemia and is structurally similar to ALA) or lead can inhibit ALA-dehydratase, increase urinary excretion of ALA and COPRO III, and cause manifestations that resemble those of the acute porphyrias. Heterozygotes are clinically asymptomatic and do not excrete increased levels of ALA, but can be detected by demonstration of intermediate levels of erythrocyte ALA-dehydratase activity or a specific mutation in the ALAD gene. To date, molecular studies of ADP patients have identified nine point mutations, two splice-site mutations, and a two-base deletion in the ALAD gene (Human Gene Mutation Database; *www.hgmd.org*). The parents in each case were not consanguineous, and the index cases had inherited a different ALAD mutation from each parent. Prenatal diagnosis of this disorder is possible by determination of ALA-dehydratase activity and/or gene mutation in cultured chorionic villi or amniocytes.

TREATMENT: ALA-DEHYDRATASE DEFICIENT PORYPHYRIA

The treatment of ADP acute attacks is similar to that of AIP (see below). The severely affected infant referred to above was supported by hyperalimentation and periodic blood transfusions, but did not respond to intravenous hemin and died after liver transplantation.

Acute Intermittent Porphyria (Aip)

This hepatic porphyria is an autosomal dominant condition resulting from the half-normal level of HMB-synthase activity. The disease is widespread but is especially common in Scandinavia and Great Britain. Clinical expression is highly variable, and activation of the disease is often related to environmental or hormonal factors, such as drugs, diet, and steroid hormones. Attacks can be prevented by avoiding known precipitating factors. Rare homozygous dominant AIP also has been described in children (see below).

CLINICAL FEATURES

Induction of the rate-limiting hepatic enzyme ALA-synthase in AIP heterozygotes with half-normal HMB-synthase activity is thought to underlie the acute attacks in AIP and other acute porphyrias. The disorder remains latent (or asymptomatic) in the great majority of those who are heterozygous for HMB-synthase mutations, and this is almost always the case prior to puberty. In patients with no history of acute symptoms, porphyrin precursor excretion is usually normal, suggesting that half-normal hepatic HMB-synthase activity is sufficient, and that hepatic ALA-synthase activity is not increased. However, under conditions where heme synthesis is increased in the liver, half-normal HMB-synthase activity may become limiting, and ALA, PBG, and other heme pathway intermediates may accumulate and be excreted in the urine. Common precipitating factors include endogenous and exogenous steroids, porphyrinogenic drugs, alcohol ingestion, and low-calorie diets, usually instituted for weight loss.

The fact that AIP is almost always latent before puberty suggests that adult levels of steroid hormones are important for clinical expression. Symptoms are more common in women, suggesting a role for estrogens or progestins. Premenstrual attacks are probably due to endogenous progesterone. Acute porphyrias are sometimes exacerbated by exogenous steroids, including oral contraceptive preparations containing progestins. Surprisingly, pregnancy is usually well tolerated, suggesting that beneficial metabolic changes may ameliorate the effects of high levels of progesterone. **Table 358-4** provides a partial list of the major drugs that are harmful in AIP (and also in HCP and VP). Extensive lists of unsafe and safe drugs are available on websites sponsored by the American Porphyria Foundation (*www.porphyriafoundation.com*), the European Porphyria Initiative (*www.porphyria-europe.org*), and at the Drug Database for Acute Porphyrias website (*www.drugs-porphyria.com*). Reduced intake of calories and carbohydrate, as may occur with illness or attempts to lose weight, can also increase porphyrin precursor excretion and induce attacks of porphyria. Increased carbohydrate intake can ameliorate attacks. Studies in a knockout AIP mouse model indicate that the hepatic ALAS1 gene is regulated by the peroxisome proliferator–activated receptor γ coactivator 1α (PGC-1α). Hepatic PGC-1α is induced by fasting, which in turn activates ALAS1 transcription, resulting in increased heme biosynthesis. This finding suggests an important link between nutritional status and the attacks in acute porphyrias. Attacks also can be provoked by infections, surgery, and ethanol.

Table 358-4 Unsafe Drugs in Porphyria

Documented Porphyrinogenic	Probably Porphyrinogenic	Possibly Porphyrinogenic	
Carbamazepine	Altretamine	Aceclofenac	Parecoxib
Carisoprodol	Aminophylline	Acitretin	Pentifylline
Chloramphenicol	Amiodarone	Acrivastine	Pentoxyverine
Clindamycin	Amitriptyline	Alfuzosin	Phenylpropanolamine
Dextropropoxyphene	Amlodipine	Anastrozole	+ Cinnarizine
Dihydralazine	Amprenavir	Auranofin	Pizotifen
Dihydroergotamine	Aprepitant	Azelastine	Polidocanol
Drospirenone + estrogen	Atorvastatin	Benztropine	Polyestradiol
Dydrogesterone	Azathioprine	Benzydamine	Phosphate
Etonogestrel	Bosentan	Betaxolol	Potassium canrenoate
Fosphenytoin sodium	Bromocriptine	Bicalutamide	Pravastatin
Hydralazine	Buspirone	Biperiden	Prednisolone
Hydroxyzine	Busulfan	Bupropion	Prilocaine
Indinavir	Butylscopolamine	Carvedilol	Proguanil
Ketamine	Cabergoline	Chlorambucil	Propafenone
Ketoconazole	Ceftriaxone +	Chlorcyclizine +	Pseudoephedrine +
Lidocaine	Lidocaine	Guaifenesin	Dexbrompheniramine
Lynestrenol	Cerivastatin	Chloroquine	Quillaia extract
Lynestrenol + estrogen	Cetirizine	Chlorprothixene	Quinagolide
Mecillinam	Cholinetheophyllinate	Chlorzoxazone	Quinine
Medroxyprogesterone	Clarithromycin	Chorionic	Quinupristin +
Megestrol	Clemastine	Gonadotrophin	Dalfopristin
Methylergometrine	Clonidine	Ciclosporin	Reboxetine
Methyldopa	Cyclizine	Cisapride	Repaglinide
Mifepristone	Cyproterone	Citalopram	Rizatriptan
Nicotinic	Danazol	Clomethiazole	Rofecoxib
Acid/meclozine/hydroxyzine	Delavirdine	Clomiphene	Ropinirole
Nitrofurantoin	Desogestrel + estrogen	Clomipramine	Ropivacaine
Norethisterone	Diazepam	Clopidogrel	Roxithromycin
Norgestimate + estrogen	Dienogest + estrogen	Clotrimazole	Sertraline
Orphenadrine	Diclofenac	Cortisone	Sevoflurane
Phenobarbital	Diltiazem	Cyclandelate	Sibutramine
Phenytoin	Diphenhydramine	Cyclophosphamide	Sildenafil
Pivampicillin	Disopyramide	Cyproheptadine	Sirolimus
Pivmecillinam	Disulfiram	Dacarbazine	Sodium
Primidone	Drospirenone +	Daunorubicin	Aurothiomalate
Rifampicin	Estrogen	Desogestrel	Sodium oleate +
Ritonavir	Dydrogesterone	Dichlorobenzyl	Chlorocymol
Spironolactone	Ergoloid mesylate	Alcohol	Stavudine
Sulfadiazine +	Erythromycin	Dithranol	Sulindac
Trimethoprim	Estramustine	Docetaxel	Sumatriptan
Tamoxifen	Ethosuximide	Donepezil	Tacrolimus
Testosterone, injection	Etoposide	Doxycycline	Tadalafil
Thiopental	Exemestane	Ebastine	Tegafur + uracil
Trimethoprim	Felbamate	Econazole	Telmisartan
Valproic acid	Felodipine	Efavirenz	Thioridazine
Venlafaxine	Fluconazole	Escitalopram	Tioguanine
Vinblastine	Flunitrazepam	Esomeprazole	Tolfenamic acid
Vincristine	FLuvastatin	Estradiol/tablets	Tolterodine
Vindesine	Glibenclamide	Estriol/tablets	Torsemide
Vinorelbine	Halothane	Estrio/vainal crème, tablet	Triamcinolone
Xylometazoline	Hyoscyamine		Trihexyphenidyl
Zaleplon	Ifosfamide	Estrogen, Conjugate	Trimipramine
Ziprasidone	Imipramine	Finasteride	Valerian
Zolmitriptan	Irinotecan	Flecainide	Venlafaxine
Zolpidem	Isoniazid	Flucloxacillin	Vinblastine
Zuclopenthixol	Isradipine	Fluoxetine	Vincristine
	Itraconazole	Flupentixol	Vindesine
	Lamivudine	Flutamide	Vinorelbine
	+Zidovudine	Fluvoxamine	Xylometazoline
	Lansoprazole	Follitropin alfa and	Zaleplon
	Lercanidipine	Beta	Ziprasidone
	Levonorgestrel	Galantamine	ZolmitriPtan
	Lidocaine,	Glimepiride	Zolpidem

	Lopinavir	Glipizide Zuclopenthixol
	Lutropin alfa	Gonadorelin
	Lymecycline	Gramicidin
	Meclozine	Guaifenesin
	Medroxyprogesterone	Hydrocortisone
	+ Estrogen	Hydroxycarbamide
	Metoclopramide	Hydroxychloroquine
	Metronidazole	Ibutilide
	Metyrapone	Imatinib
	Moxonidine	Indomethacin
	Nandrolone	Ketobemidone +
	Nefazodone	Ddba
	Nelfinavir	Ketoconazole
	Nevirapine	Ketorolac
	Nifedipine	Lamotrigine
	Nimodipine	Letrozole
	Nitrazepam	Levodopa +
	Norethisterone	Benserazide
	Nortriptyline	Levonorgestrel
	Oxcarbazepine	Intrauterine
	Oxytetracycline	Levosimendan
	Paclitaxel	Lidocaine
	Paroxetine	Linezolid
	Phenazone + caffeine	Lofepramine
	Pioglitazone	Lomustine
	Probenecid	Malathion
	Progesterone, vaginal	Maprotiline
	Gel	Mebendazole
	Quinidine	Mefloquine
	Rabeprazole	Melperone
	Raloxifene	Melphalan
	Rifabutin	Mepenzolate
	Riluzole	Mepivacaine
	Risperidone	Mercaptopurine
	Rosiglitazone	Methadone
	Saquinavir	Methylprednisolone
	Selegiline	Methixene
	Simvastatin	Metolazone
	Sulfasalazine	Metronidazole
	Telithromycin	Mexiletine
	Terbinafine	Mianserin
	Terfenadine	Midazolam
	Testosterone,	Minoxidil
	Transdermal patch	Mirtazapine
	Tetracycline	Mitomycin
	Theophylline	Mitoxantrone
	Thiamazole	Moclobemide
	Tibolone	Montelukast
	Ticlopidine	Morphine +
	Tinidazole	Scopolamine
	Thiotepa	Multivitamins
	Topiramate	Mupirocin
	Topotecan	Nabumetone
	Toremifene	Nafarelin
	Tramadol	Naltrexone
	Trimegestone +	Nateglinide
	Estrogen	Nilutamide
	Verapamil	Noscapine
	Voriconazole	Omeprazole
	Zidovudine/azt	Oxybutynin
		Oxycodone
		Pantoprazole
		Papaverine

Note: Based on list in "Patient's and Doctor's Guide to Medication in Acute Porphyria," Swedish Porphyria Association and Porphyria Centre Sweden. Also see the website Drug Database for Acute Porphyrias (*www.drugs-porphyria.com*) for a

searchable list of safe and unsafe drugs.

Because the neurovisceral symptoms rarely occur before puberty and are often nonspecific, a high index of suspicion is required to make the diagnosis. The disease can be disabling but is rarely fatal. Abdominal pain, the most common symptom, is usually steady and poorly localized but may be cramping. Ileus, abdominal distention, and decreased bowel sounds are common. However, increased bowel sounds and diarrhea may occur. Abdominal tenderness, fever, and leukocytosis are usually absent or mild because the symptoms are neurologic rather than inflammatory. Nausea; vomiting; constipation; tachycardia; hypertension; mental symptoms; pain in the limbs, head, neck, or chest; muscle weakness; sensory loss; dysuria; and urinary retention are characteristic. Tachycardia, hypertension, restlessness, tremors, and excess sweating are due to sympathetic overactivity.

The peripheral neuropathy is due to axonal degeneration (rather than demyelinization) and primarily affects motor neurons. Significant neuropathy does not occur with all acute attacks; abdominal symptoms are usually more prominent. Motor neuropathy affects the proximal muscles initially, more often in the shoulders and arms. The course and degree of involvement are variable and sometimes may be focal and involve cranial nerves. Deep tendon reflexes initially may be normal or hyperactive but become decreased or absent as the neuropathy advances. Sensory changes such as paresthesia and loss of sensation are less prominent. Progression to respiratory and bulbar paralysis and death occurs especially when the diagnosis and treatment are delayed. Sudden death may result from sympathetic overactivity and cardiac arrhythmia.

Mental symptoms such as anxiety, insomnia, depression, disorientation, hallucinations, and paranoia can occur in acute attacks. Seizures can be due to neurologic effects or to hyponatremia. Treatment of seizures is difficult because most antiseizure drugs can exacerbate AIP (clonazepam may be safer than phenytoin or barbiturates). Hyponatremia results from hypothalamic involvement and inappropriate vasopressin secretion or from electrolyte depletion due to vomiting, diarrhea, poor intake, or excess renal sodium loss. Persistent hypertension and impaired renal function may occur. When an attack resolves, abdominal pain may disappear within hours, and paresis begins to improve within days and may continue to improve over several years.

Homozygous dominant AIP is a rare form of AIP in which patients inherit HMBS mutations from each of their heterozygous parents and, therefore, have very low (<2%) enzyme activity. The disease has been described in a Dutch girl, two young British siblings, and a Spanish boy. In these homozygous affected patients, the disease presented in infancy with failure to thrive, developmental delay, bilateral cataracts, and/or hepatosplenomegaly. Urinary ALA and PBG concentrations were markedly elevated. All of these patients' HMBS mutations (R167W, R167Q, and R172Q) were in exon 10 within five bases of each other. Studies of the brain MRIs of children with homozygous AIP have suggested damage primarily in white matter that was myelinated postnatally, while tracks that myelinated prenatally were normal. Most children with homozygous AIP die at an early age.

DIAGNOSIS

ALA and PBG levels are substantially increased in plasma and urine, especially during acute attacks, and become normal only after prolonged latency. For example, urinary PBG excretion during an attack is usually 50–200 mg/24 h (220–880 μmol/24 h) [normal, 0–4 mg/24 h (0–18 μmol/24 h)], and urinary ALA excretion is 20–100 mg/24 h (150–760 μmol/24 h) [normal, 1–7 mg/24 h (8–53 μmol/ 24 h)]. Because levels often remain high after symptoms resolve, the diagnosis of an acute attack in a patient with biochemically proven AIP is based primarily on clinical features. Excretion of ALA and PBG decrease dramatically after intravenous hemin. A normal urinary PBG level before hemin effectively excludes AIP as a cause for current symptoms. Fecal porphyrins are usually normal or minimally increased in AIP, in contrast to HCP and VP. Most heterozygotes with HMB-synthase deficiency with no history of symptoms have normal urinary excretion of ALA and PBG. Therefore, measurement of HMB-synthase in erythrocytes or, better, the detection of the family's HMBS mutation will diagnose asymptomatic family members.

The enzyme deficiency is detectable in erythrocytes from most AIP heterozygotes. Because the activity is higher in young erythrocytes, it may increase into the normal range if erythropoiesis is increased due to a concurrent condition. Furthermore, patients with HMBS mutations in the initiation of translation codon in exon 1 and in the intron 1 5′-splice donor site have normal enzyme levels in erythrocytes and deficient activity only in nonerythroid tissues. This occurs because the erythroid and housekeeping forms of HMB-synthase are encoded by a single gene, which has two promoters. More than 315 HMBS mutations have been identified in AIP, including missense, nonsense, and splicing mutations and insertions and deletions, with most mutations found in only one or a few families (Human Gene Mutation Database, *www.hgmd.org*). The prenatal diagnosis of a fetus at risk can be made with cultured amniotic cells or chorionic villi. However, this is seldom done, because the prognosis of individuals with HMBS mutations is generally favorable.

TREATMENT: ACUTE INTERMITTENT PORPHYRIA

During acute attacks, narcotic analgesics may be required for abdominal pain, and phenothiazines are useful for nausea, vomiting, anxiety, and restlessness. Chloral hydrate can be given for insomnia, and benzodiazepines are probably safe in low doses if a minor tranquilizer is required. Carbohydrate loading, usually with intravenous glucose (at least 300 g/dL of 10% dextrose in water daily), may be effective in milder acute attacks of porphyria (without paresis, hyponatremia, etc.). Because intravenous hemin is more effective and the response slower if treatment is delayed, it is no longer recommended that hemin therapy for a severe attack be started only after an unsuccessful trial of intravenous glucose for several days. Hemin should be used initially for severe attacks and for mild attacks that do not respond to carbohydrate loading within 1–2 days. The standard regimen is 3–4 mg/kg of heme, in the form of lyophilized hematin (Lundbeck Pharmaceuticals), heme albumin (hematin reconstituted with human albumin), or heme arginate (Orphan Europe), infused daily for 4 days. Heme arginate and heme albumin are chemically stable and are less likely than hematin to produce phlebitis or an anticoagulant effect. Recovery depends on the degree of neuronal damage and usually is rapid if therapy is started early. Recovery from severe motor neuropathy may require months or years. Identification and avoidance of inciting factors can hasten recovery from an attack and prevent future attacks. Inciting factors are usually multiple, and removal of one or more hastens recovery and helps prevent future attacks. Frequent attacks that occur during the luteal phase of the menstrual cycle may be prevented with a gonadotropin-releasing hormone analogue, which prevents ovulation and progesterone production, or by preventive monthly hematin administration.

The long-term risk of hypertension and chronic renal disease is increased in AIP; a number of patients have undergone successful renal transplantation. Chronic, low-grade abnormalities in liver function tests are common, and the risk of hepatocellular carcinoma is increased. Hepatic imaging is recommended at least yearly for early detection of these tumors.

An allogeneic liver transplant was performed on a 19-year-old female AIP heterozygote who had 37 acute attacks in the 29 months prior to transplantation. Post-transplantation, her elevated urinary ALA and BPG levels returned to normal in 24 h, and she did not experience acute neurologic, attacks for more than 3 years post-transplant. Two AIP patients had combined liver and kidney transplants secondary to uncontrolled acute porphyria attacks, chronic peripheral neuropathy, and renal failure requiring dialysis. Both patients had a marked improvement with no attacks and normal urinary PBG levels post-transplantation, as well as improvement of their neuropathic manifestations. To date, a number of women with AIP have had successful liver transplants for severe recurrent neurologic attacks and secondary organ damage with amelioration of their neurologic attacks and normalization of their heme precursors. However, liver transplantation is a high-risk procedure and

should be considered as a last resort in patients with severe recurrent attacks. Recently, liver-directed gene therapy has been successful in the prevention of drug-induced biochemical attacks in a murine model of human AIP.

Porphyria Cutanea Tarda (Pct)

PCT, the most common of the porphyrias, can be either sporadic (type 1) or familial (type 2) and can also develop after exposure to halogenated aromatic hydrocarbons. Hepatic URO-decarboxylase is deficient in all types of PCT, and for clinical symptoms to manifest, this enzyme deficiency must be substantial (~20% of normal activity or less); it is currently attributed to generation of an URO-decarboxylase inhibitor in the liver in the presence of iron and under conditions of oxidative stress, although it remains to be isolated and characterized.

The majority of PCT patients (~80%) have no UROD mutations and are said to have sporadic (type 1) disease. PCT patients heterozygous for UROD mutations have familial (type 2) PCT. In these patients, inheritance of a UROD mutation from one parent results in half-normal enzyme activity in liver and all other tissues, which is a significant predisposing factor, but is insufficient by itself to cause symptomatic PCT. As discussed below, other genetic and environmental factors contribute to susceptibility for both types of PCT. Because penetrance of the genetic trait is low, many patients with familial (type 2) PCT have no family history of the disease. Hepatoerythropoietic porphyria (HEP) is an autosomal recessive form of porphyria that results from the marked systemic deficiency of URO-decarboxylase activity with clinical symptoms in childhood.

CLINICAL FEATURES

Blistering skin lesions that appear most commonly on the backs of the hands are the major clinical feature **(Fig. 358-3)**. These rupture and crust over, leaving areas of atrophy and scarring. Lesions may also occur on the forearms, face, legs, and feet. Skin friability and small white papules termed milia are common, especially on the backs of the hands and fingers. Hypertrichosis and hyperpigmentation, especially of the face, are especially troublesome in women. Occasionally, the skin over sun-exposed areas becomes severely thickened, with scarring and calcification that resembles systemic sclerosis. Neurologic features are absent.

Figure 358-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Typical cutaneous lesions in a patient with porphyria cutanea tarda. Chronic, crusted lesions resulting from blistering due to photosensitivity on the dorsum of the hand of a PCT patient. *(Courtesy of Dr. Karl E. Anderson; with permission.*)

A number of susceptibility factors, in addition to inherited UROD mutations in type 2 PCT, can be recognized clinically and can affect management. These include hepatitis C, HIV, excess alcohol, elevated iron levels, and estrogens. The importance of excess hepatic iron as a precipitating factor is underscored by the finding that the incidence of the common hemochromatosis-causing mutations, hemochromatosis gene (HFE) mutations C282Y and H63D, is increased in patients with types 1 and 2 PCT (Chap. 357). Excess alcohol is a long-recognized contributor, as is estrogen use in women. HIV is probably an independent but less common risk factor that, like hepatitis C, does not cause PCT in isolation. Multiple susceptibility factors that appear to act synergistically can be identified in the individual PCT patient. Patients with PCT characteristically have chronic liver disease and sometimes cirrhosis and are at risk for hepatocellular carcinoma. Various chemicals can also induce PCT; an epidemic of PCT occurred in eastern Turkey in the 1950s as a consequence of wheat contaminated with the fungicide hexachlorobenzene. PCT also occurs after exposure to other chemicals, including di- and trichlorophenols and 2,3,7,8-tetrachlorodibenzo-(*p*)-dioxin (TCDD, dioxin).

DIAGNOSIS

Porphyrins are increased in the liver, plasma, urine, and stool. The urinary ALA level may be slightly increased, but the PBG level is normal. Urinary porphyrins consist mostly of URO and heptacarboxylate porphyrin, with lesser amounts of COPRO and hexa- and pentacarboxylate porphyrins. Plasma porphyrins are also increased, and fluorometric scanning of diluted plasma at neutral pH can rapidly distinguish VP and PCT (Table 358-3). Isocoproporphyrins, which are increased in feces and sometimes in plasma and urine, are diagnostic for hepatic URO-decarboxylase deficiency.

Type 2 PCT and HEP can be distinguished from type 1 by finding decreased URO-decarboxylase in erythrocytes. URO-decarboxylase activity in liver, erythrocytes, and cultured skin fibroblasts in type 2 PCT is approximately 50% of normal in affected individuals and in family members with latent disease. In HEP, the URO-decarboxylase activity is markedly deficient, with typical levels of 3–10% of normal. More than 105 mutations have been identified in the UROD gene (Human Gene Mutation Database; *www.hgmd.org*). Of the mutations listed in the database, ~65% are missense or nonsense and ~10% are splice-site mutations. Most UROD mutations have been identified in only one or two families.

TREATMENT: PORPHYRIA CUTANEA TARDA

Alcohol, estrogens, iron supplements, and, if possible, any drugs that may exacerbate the disease should be discontinued, but this step does not always lead to improvement. A complete response can almost always be achieved by the standard therapy, repeated phlebotomy, to reduce hepatic iron. A unit (450 mL) of blood can be removed every 1–2 weeks. The aim is to gradually reduce excess hepatic iron until the serum ferritin reaches the lower limits of normal. Because iron overload is not marked in most cases, remission may occur after only five or six phlebotomies; however, PCT patients with hemochromatosis may require more treatments to bring their iron levels down to the normal range. To document improvement in PCT, it is most

convenient to follow the total plasma porphyrin concentration, which becomes normal some time after the target ferritin level is reached. Hemoglobin levels or hematocrits and serum ferritin should be followed closely to prevent development of iron deficiency and anemia. After remission, continued phlebotomy may not be needed. Plasma porphyrin levels are followed at 6–12-month intervals for early detection of recurrences, which are treated by additional phlebotomy.

An alternative when phlebotomy is contraindicated or poorly tolerated is a low-dose regimen of chloroquine or hydroxychloroquine, both of which complex with the excess porphyrins and promote their excretion. Small doses (e.g., 125 mg chloroquine phosphate twice weekly) should be given, because standard doses can induce transient, sometimes marked increases in photosensitivity and hepatocellular damage. Hepatic imaging can diagnose or exclude complicating hepatocellular carcinoma. Treatment of PCT in patients with end-stage renal disease is facilitated by administration of erythropoietin.

Hereditary Coproporphyria (Hcp)

HCP is an autosomal dominant hepatic porphyria that results from the half-normal activity of COPRO-oxidase. The disease presents with acute attacks, as in AIP. Cutaneous photosensitivity also may occur, but much less commonly than in VP. HCP patients may have acute attacks and cutaneous photosensitivity together or separately. HCP is less common than AIP and VP. Homozygous dominant HCP and harderoporphyria, a biochemically distinguishable variant of HCP, present with clinical symptoms in children (see below).

CLINICAL FEATURES

HCP is influenced by the same factors that cause attacks in AIP. The disease is latent before puberty, and symptoms, which are virtually identical to those of AIP, are more common in women. HCP is generally less severe than AIP. Blistering skin lesions are identical to PCT and VP and begin in childhood in rare homozygous cases.

DIAGNOSIS

COPRO III is markedly increased in the urine and feces in symptomatic disease and often persists, especially in feces, when there are no symptoms. Urinary ALA and PBG levels are increased (but less than in AIP) during acute attacks, but may revert to normal more quickly than in AIP when symptoms resolve. Plasma porphyrins are usually normal or only slightly increased, but they may be higher in cases with skin lesions. The diagnosis of HCP is readily confirmed by increased fecal porphyrins consisting almost entirely of COPRO III, which distinguishes it from other porphyrias.

Although the diagnosis can be confirmed by measuring COPRO-oxidase activity, the assays for this mitochondrial enzyme are not widely available and require cells other than erythrocytes. The CPOX gene has been cloned and more than 50 mutations, 70% of which are missense or nonsense, have been identified in unrelated patients (Human Gene Mutation Database; *www.hgmd.org*). Detection of a CPOX mutation in a symptomatic individual permits the identification of asymptomatic family members.

TREATMENT: HEREDITARY COPROPORPHYRIA

Neurologic symptoms are treated as in AIP (see above). Phlebotomy and chloroquine are ineffective when cutaneous lesions are present.

Variegate Porphyria (Vp)

VP is an autosomal dominant hepatic porphyria that results from the deficient activity of PROTO-oxidase, the seventh enzyme in the heme biosynthetic pathway, and can present with neurologic symptoms, photosensitivity, or both. VP is particularly common in South Africa, where 3 of every 1000 whites have the disorder. Most are descendants of a couple who emigrated from Holland to South Africa in 1688. In other countries, VP is less common than AIP. Rare cases of homozygous dominant VP, presenting in childhood with cutaneous symptoms, also have been reported.

CLINICAL FEATURES

VP can present with skin photosensitivity, acute neurovisceral crises, or both. In two large studies of VP patients, 59% had only skin lesions, 20% had only acute attacks, and 22% had both. Acute attacks are identical to those in AIP and are precipitated by the same factors as AIP (see above). Blistering skin manifestations are similar to those in PCT, but are more difficult to treat and usually are of longer duration. Homozygous VP is associated with photosensitivity, neurologic symptoms, and developmental disturbances, including growth retardation, in infancy or childhood; all cases had increased erythrocyte levels of zinc protoporphyrin, a characteristic finding in all homozygous porphyrias so far described.

DIAGNOSIS

Urinary ALA and PBG levels are increased during acute attacks but may return to normal more quickly than in AIP. Increases in fecal protoporphyrin and COPRO III and in urinary COPRO III are more persistent. Plasma porphyrin levels also are increased, particularly when there are cutaneous lesions. VP can be distinguished rapidly from all other porphyrias by examining the fluorescence emission spectrum of porphyrins in plasma at neutral pH since VP has a unique fluorescence peak at neutral pH.

Assays of PROTO-oxidase activity in cultured fibroblasts or lymphocytes are not widely available. More than 145 mutations have been identified in the PPOX gene from unrelated VP patients (Human Gene Mutation Database; *www.hgmd.org*). The missense mutation R59W is the common mutation in most South Africans with VP of Dutch descent. Five missense mutations were common in English and French VP patients; however, most mutations have been found in only one or two families.

TREATMENT: VARIEGATE PORPHYRIA

Acute attacks are treated as in AIP, and hemin should be started early in most cases. Other than avoiding sun exposure, there are few effective measures for treating the skin lesions. β-Carotene, phlebotomy, and chloroquine are not helpful.

THE ERYTHROPOIETIC PORPHYRIAS

In the erythropoietic porphyrias, excess porphyrins from bone marrow erythrocyte precursors are transported via the plasma to the skin and lead to cutaneous photosensitivity.

X-Linked Sideroblastic Anemia (Xlsa)

XLSA results from the deficient activity of the erythroid form of ALA-synthase and is associated with ineffective erythropoiesis, weakness, and pallor.

CLINICAL FEATURES

Typically, males with XLSA develop refractory hemolytic anemia, pallor, and weakness during infancy. They have secondary hypersplenism, become iron overloaded, and can develop hemosiderosis. The severity depends on the level of residual erythroid ALA-synthase activity and on the responsiveness of the specific mutation to pyridoxal 5′-phosphate supplementation (see below). Peripheral blood smears reveal a hypochromic, microcytic anemia with striking anisocytosis, poikilocytosis, and polychromasia; the leukocytes and platelets appear normal. Hemoglobin content is reduced, and the mean corpuscular volume and mean corpuscular hemoglobin concentration are decreased. Patients with milder, late-onset disease have been reported

recently.

DIAGNOSIS

Bone marrow examination reveals hypercellularity with a left shift and megaloblastic erythropoiesis with an abnormal maturation. A variety of Prussian blue–staining sideroblasts are observed. Levels of urinary porphyrin precursors and of both urinary and fecal porphyrins are normal. The level of erythroid ALA-synthase is decreased in bone marrow, but this enzyme is difficult to measure in the presence of the normal ALA-synthase housekeeping enzyme. Definitive diagnosis requires the demonstration of mutations in the erythroid ALAS2 gene.

TREATMENT: X-LINKED SIDEROBLASTIC ANEMIA

The severe anemia may respond to pyridoxine supplementation. This cofactor is essential for ALA-synthase activity, and mutations in the pyridoxine-binding site of the enzyme have been found in several responsive patients. Cofactor supplementation may make it possible to eliminate or reduce the frequency of transfusion. Unresponsive patients may be transfusion-dependent and require chelation therapy.

Congenital Erythropoietic Porphyria (Cep)

CEP, also known as Günther disease, is an autosomal recessive disorder. It is due to the markedly deficient, but not absent, activity of URO-synthase and the resultant accumulation of URO I and COPRO I isomers. CEP is associated with hemolytic anemia and cutaneous lesions.

CLINICAL FEATURES

Severe cutaneous photosensitivity begins in early infancy. The skin over light-exposed areas is friable, and bullae and vesicles are prone to rupture and infection. Skin thickening, focal hypo- and hyperpigmentation, and hypertrichosis of the face and extremities are characteristic. Secondary infection of the cutaneous lesions can lead to disfigurement of the face and hands. Porphyrins are deposited in teeth and in bones. As a result, the teeth are reddish-brown and fluoresce on exposure to long-wave ultraviolet light. Hemolysis is probably due to the marked increase in erythrocyte porphyrins and leads to splenomegaly. Adults with a milder form of the disease also have been described.

DIAGNOSIS

URO and COPRO (mostly type I isomers) accumulate in the bone marrow, erythrocytes, plasma, urine, and feces. The predominant porphyrin in feces is COPRO I. The diagnosis of CEP can be confirmed by demonstration of markedly deficient URO-synthase activity and/or the identification of specific mutations in the UROS gene. The disease can be detected in utero by measuring porphyrins in amniotic fluid and URO-synthase activity in cultured amniotic cells or chorionic villi, or by the detection of the family's specific gene mutations. Molecular analyses of the mutant alleles from more than 50 unrelated patients have revealed the presence of more than 35 mutations in the UROS gene, including four in the erythroid-specific promoter of the UROS gene. Genotype/phenotype correlations can predict the severity of the disease.

TREATMENT: CONGENITAL ERYTHROPOIETIC PORPHYRIA

Severe cases often require transfusions for anemia. Chronic transfusions of sufficient blood to suppress erythropoiesis are effective in reducing porphyrin production, but results in iron overload. Splenectomy may reduce hemolysis and decrease transfusion requirements. Protection from sunlight and from minor skin trauma is important. β-Carotene may be of some value. Complicating bacterial infections should be treated promptly. Recently, bone marrow and cord blood transplantation has proven curative in several transfusion-dependent children, providing the rationale for stem-cell gene therapy.

Erythropoietic Protoporphyria (Epp)

EPP is an inherited disorder resulting from the deficient activity of ferrochelatase, the last enzyme in the heme biosynthetic pathway. EPP is the most common erythropoietic porphyria in children and, after PCT, the second most common porphyria in adults. EPP patients have ferrochelatase activities as low as 15–25% in lymphocytes and cultured fibroblasts. Protoporphyrin accumulates in bone marrow reticulocytes and then appears in plasma, is taken up in the liver, and excreted in bile and feces. Protoporphyrin transported to the skin causes nonblistering photosensitivity. In most symptomatic patients (~90%) with this autosomal recessive disorder, a mutation in one FECH allele is inherited with a relatively common (~10% normal whites) intronic 3 (IVS3) alteration (IVS3-48T>C) that results in the low expression of the normal enzyme. In about 10% of EPP families, two FECH mutations have been found. Recently, deletion mutations in exon 11 of the ALAS2 gene have been described, which cause an X-linked protoporphyria (XLP) that is clinically indistinguishable from EPP. The deletion of the carboxy-terminal amino acid of ALAS2 results in increased ALA-synthase 2 activity and the accumulation of protoporphyrin. XLP accounts for approximately 2% of cases with the EPP phenotype.

CLINICAL FEATURES

Skin photosensitivity, which differs from that in other porphyrias, usually begins in childhood and consists of pain, redness, and itching occurring within minutes of sunlight exposure **(Fig. 358-4)**. Photosensitivity is associated with substantial elevations in erythrocyte protoporphyrin and occurs only in patients with genotypes that result in ferrochelatase activities lower than ~35% of normal. Vesicular lesions are uncommon. Redness, swelling, burning, and itching can develop shortly after sun exposure and resemble angioedema. Symptoms may seem out of proportion to the visible skin lesions. Sparse vesicles and bullae occur in 10% of cases. Chronic skin changes may include lichenification, leathery pseudovesicles, labial grooving, and nail changes. Severe scarring is rare, as are pigment changes, friability, and hirsutism. Unless hepatic or other complications develop, protoporphyrin levels and symptoms of photosensitivity remain remarkably stable over many years in most patients. Factors that exacerbate the hepatic porphyrias play little or no role in EPP.

Figure 358-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythema and edema of the hands due to acute photosensitivity in a 10-year-old boy with erythropoietic protoporphyria. *(From P. Poblette-Gutierrez et al.*)

The primary source of excess protoporphyrin is the bone marrow reticulocytes. Erythrocyte protoporphyrin is free (not complexed with zinc) and is mostly bound to hemoglobin. In plasma, protoporphyrin is bound to albumin. Hemolysis and anemia are usually absent or mild.

Although EPP is an erythropoietic porphyria, up to 20% of EPP patients may have minor abnormalities of liver function, and in about 5% of these patients the accumulation of protoporphyrins causes chronic liver disease that can progress to liver failure and death. Protoporphyrin is insoluble, and excess amounts form crystalline structures in liver cells (Fig. 358-4) and can decrease hepatic bile flow. Studies in the mouse model of EPP have shown that the bile duct epithelium may be damaged by toxic bile, leading to biliary fibrosis. Thus, rapidly progressive liver disease appears to be related to the cholestatic effects of protoporphyrins and is associated with increasing hepatic protoporphyrin levels due to impaired hepatobiliary excretion and increased photosensitivity. The hepatic complications also are often characterized by increasing levels of protoporphyrins in erythrocytes and plasma as well as severe abdominal and back pains, especially in the right upper quadrant. Gallstones composed at least in part of protoporphyrin occur in some patients. Hepatic complications appear to be higher in autosomal recessive EPP due to two FECH mutations and in XLP.

DIAGNOSIS

A substantial increase in erythrocyte protoporphyrin, which is predominantly free and not complexed with zinc, is the hallmark of EPP. Protoporphyrin levels are also variably increased in bone marrow, plasma, bile, and feces. Erythrocyte protoporphyrin concentrations are increased in other conditions such as lead poisoning, iron deficiency, various hemolytic disorders, all homozygous forms of other porphyrias, and sometimes even in acute porphyrias. In all these conditions, however, in contrast to EPP, protoporphyrin is complexed with zinc. Therefore, after an increase in erythrocyte protoporphyrin is found in a suspected EPP patient, it is important to confirm the diagnosis by an assay that distinguishes free and zinc-complexed protoporphyrin. Erythrocytes in EPP also exhibit red fluorescence under a fluorescence microscopy at 620 nm. Urinary levels of porphyrins and porphyrin precursors are normal. Ferrochelatase activity in cultured lymphocytes or fibroblasts is decreased. DNA diagnosis by mutation analysis is recommended to detect the causative FECH mutation(s) and/or the presence of the IVS3-48T>C hypo-expression allele. To date, 125 mutations have been identified in the FECH gene, many of which result in an unstable or absent enzyme protein (null alleles) (Human Gene Mutation Database; *www.hgmd.org*). Studies suggest that EPP patients with a null allele (and the IVS3-48T>C hypo-expression allele) have a great risk for developing severe liver complications. In XLP, the erythrocyte protoporphyrin levels appear to be higher than other forms of EPP and the proportions of free and zinc protoporphyrins are approximately equivalent. To date, two ALAS2 mutations, all deletions of 1 to 4 bases, have been described, which markedly increase ALAS2 activity and cause XLP.

TREATMENT: ERYTHROPOIETIC PROTOPORPHYRIA

Avoiding sunlight exposure and wearing clothing designed to provide protection for conditions with chronic photosensitivity are essential. Oral β-carotene (120–180 mg/dL) improves tolerance to sunlight in some patients. The beneficial effects of β-carotene may involve quenching of singlet oxygen or free radicals. The dosage may need to be adjusted to maintain serum carotene levels in the recommended range of 10–15 mmol/L (600–800 mg/dL). Mild skin discoloration due to carotenemia is the only significant side effect.

Treatment of hepatic complications, which may be accompanied by motor neuropathy, is difficult. Cholestyramine and other porphyrin absorbents such as activated charcoal may interrupt the enterohepatic circulation of protoporphyrin and promote its fecal excretion, leading to some improvement. Splenectomy may be helpful when the disease is accompanied by hemolysis and significant splenomegaly. Plasmapheresis and intravenous hemin are sometimes beneficial.

Liver transplantation has been carried out in some EPP and XLP patients with severe liver complications and is often successful in the short term. However, disease often recurs in the transplanted liver due to continued bone marrow production of excess protoporphyrin. In a retrospective study of 17 liver-transplanted EPP patients, 11 (65%) had recurrent EPP liver disease. Posttransplantation treatment with hematin and plasmapheresis should be considered to prevent the recurrence of liver disease. Bone marrow transplantation, which has been successful in human EPP and prevents liver disease in a mouse model, should be considered after the liver transplantation, if a suitable donor can be found.

ACKNOWLEDGMENT

The authors thank Dr. Karl E. Anderson for his review of the manuscript and helpful comments and suggestions .

FURTHER READINGS

Anderson KE et al: Disorders of heme biosynthesis: X-linked sideroblastic anemia and the porphyrias, in *The Metabolic and Molecular Bases of Inherited Diseases*, CR Scriver et al (eds). New York, McGraw-Hill, 2001, pp 2991–3062

Handschin C et al: Nutritional regulation of hepatic heme biosynthesis and porphyria through PCG-1α. Cell 122:505, 2005[PMID: 16122419] [Full Text]

Herve P et al: Porphyrias. Lancet 375:924, 2010

Lambrecht RW et al: Genetic aspects of porphyria cutanea tarda. Semin Liver Dis 27:99, 2007[PMID: 17295179] [Full Text]

Poblete-Gutierrez P et al: The porphyrias: Clinical diagnosis and treatment. Eur J Dermatol 16:230, 2006[PMID: 16709484] [Full Text]

Soonawalla ZF et al: Liver transplantation as a cure for acute intermittent porphyria. Lancet 363:705, 2004[PMID: 15001330] [Full Text]

Whatley SD et al: Molecular epidemiology of erythropoietic protoporphyria in the U.K. Br J Dermatol 162:642, 2010[PMID: 20105171] [Full Text]

Yasuda M et al: AAV8-mediated gene therapy prevents induced biochemical attacks of intermittent porphyria and improves neuromotor function. Mol Ther 18:17, 2010[PMID: 19861948] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF PURINE AND PYRIMIDINE METABOLISM: INTRODUCTION

Purines (adenine and guanine) and pyrimidines (cytosine, thymine, uracil) serve fundamental roles in the replication of genetic material, gene transcription, protein synthesis, and cellular metabolism. Disorders that involve abnormalities of nucleotide metabolism range from relatively common diseases such as hyperuricemia and gout, in which there is increased production or impaired excretion of a metabolic end product of purine metabolism (uric acid), to rare enzyme deficiencies that affect purine and pyrimidine synthesis or degradation. Understanding these biochemical pathways has led, in some instances, to the development of specific forms of treatment, such as the use of allopurinol, to reduce uric acid production.

URIC ACID METABOLISM

Uric acid is the final breakdown product of purine degradation in humans. It is a weak acid with pK_as of 5.75 and 10.3. Urates, the ionized forms of uric acid, predominate in plasma extracellular fluid and synovial fluid, with ~98% existing as monosodium urate at pH 7.4.

Plasma is saturated with monosodium urate at a concentration of 405 μmol/L (6.8 mg/dL) at 37°C. At higher concentrations, plasma is therefore supersaturated, creating the potential for urate crystal precipitation. However, plasma urate concentrations can reach 4800 μmol/L (80 mg/dL) without precipitation, perhaps because of the presence of solubilizing substances.

The pH of urine greatly influences the solubility of uric acid. At pH 5.0, urine is saturated with uric acid at concentrations ranging from 360 to 900 μmol/L (6–15 mg/dL). At pH 7, saturation is reached at concentrations between 9480 and 12,000 μmol/L (158 and 200 mg/dL). Ionized forms of uric acid in urine include mono- and disodium, potassium, ammonium, and calcium urates.

Although purine nucleotides are synthesized and degraded in all tissues, urate is produced only in tissues that contain xanthine oxidase, primarily the liver and small intestine. Urate production varies with the purine content of the diet and the rates of purine biosynthesis, degradation, and salvage **(Fig. 359-1)**. Normally, two-thirds to three-fourths of urate is excreted by the kidneys, and most of the remainder is eliminated through the intestines.

Figure 359-1

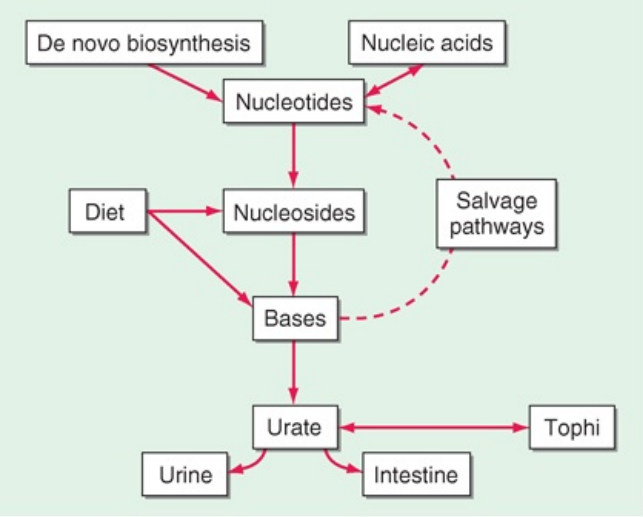

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The total-body urate pool is the net result between urate production and excretion. Urate production is influenced by dietary intake of purines and the rates of de novo biosynthesis of purines from nonpurine precursors, nucleic acid turnover, and salvage by phosphoribosyltransferase activities. The formed urate is normally excreted by urinary and intestinal routes. Hyperuricemia can result from increased production, decreased excretion, or a combination of both mechanisms. When

hyperuricemia exists, urate can precipitate and deposit in tissues as tophi.

The kidneys clear urate from the plasma and maintain physiologic balance by utilizing specific organic anion transporters (OATs), including urate transporter 1 (URAT1) and human uric acid transporter (hUAT) **(Fig. 359-2)**. URAT1 and other OATs carry urate into the tubular cells from the apical side of the lumen. Once inside the cell, urate must pass to the basolateral side of the lumen in a process controlled by the voltage-dependent carrier hUAT. Until recently, a four-component model has been used to describe the renal handling of urate/uric acid: (1) glomerular filtration, (2) tubular reabsorption, (3) secretion, and (4) postsecretory reabsorption. Although these processes have been considered sequential, it is now apparent that they are carried out in parallel by these transporters. URAT1 is a novel transporter expressed at the apical brush border of the proximal nephron. Uricosuric compounds **(Table 359-1)** directly inhibit URAT1 on the apical side of the tubular cell (so-called *cis*-inhibition). In contrast, antiuricosuric compounds (those that promote hyperuricemia), such as nicotinate, pyrazinoate, lactate, and other aromatic organic acids, serve as the exchange anion inside the cell, thereby stimulating anion exchange and urate reabsorption (*trans*-stimulation). The activities of URAT1, other OATs, and sodium anion transporter result in 8–12% of the filtered urate being excreted as uric acid.

Figure 359-2

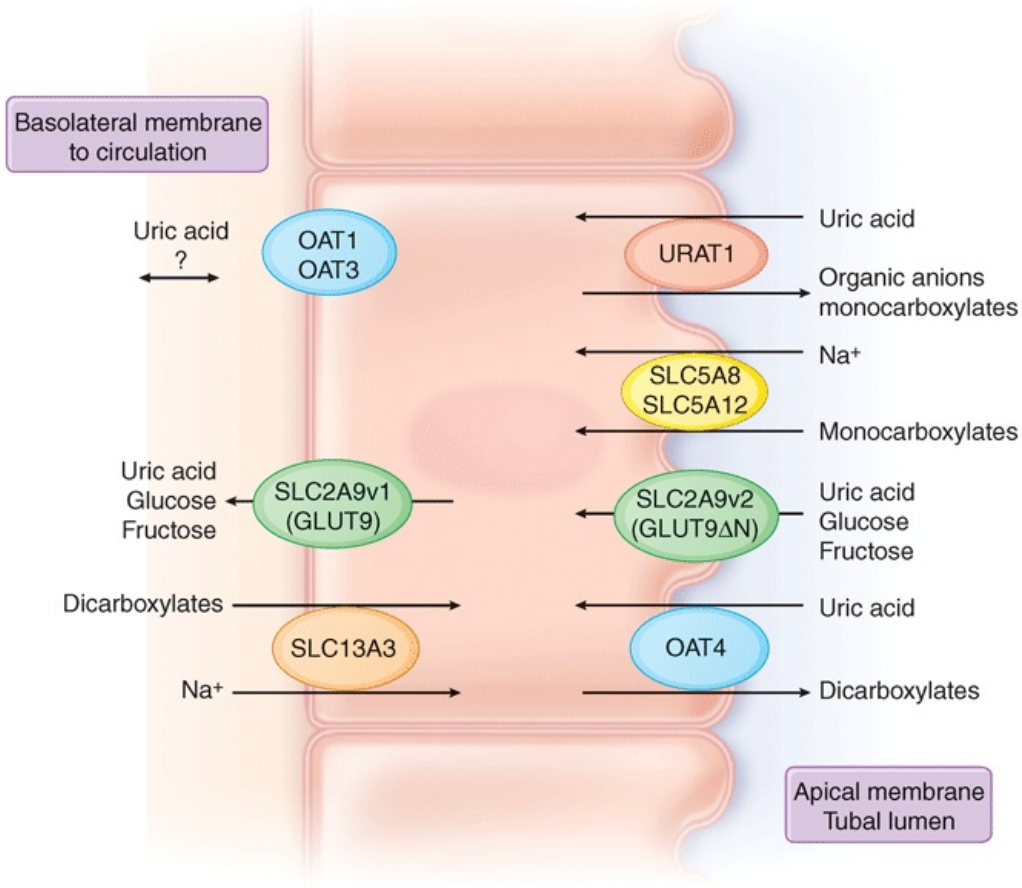

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic for handling of uric acid by the kidney. A complex interplay of transporters on both the apical and basolateral aspects of the renal tubuleepithelial cell is involved in the reabsorption of uric acid. Please see text for details. Most uricosuric compounds inhibit URAT1 on the apical side, as well as OAT1, OAT3, and GLUT9 on the basolateral side.

Table 359-1 Medications with Uricosuric Activity

Acetohexamide	Glyceryl guaiacolate
ACTH	Glycopyrrolate
Ascorbic acid	Halofenate
Azauridine	Losartan

Benzbromarone	Meclofenamate
Calcitonin	Phenolsulfonphthalein
Chlorprothixene	Phenylbutazone
Citrate	Probenecid
Dicumarol	Radiographic contrast agents
Diflunisal	Salicylates (>2 g/d)
Estrogens	Sulfinpyrazone
Fenofibrate	Tetracycline that is outdated
Glucocorticoids	Zoxazolamine

Most children have serum urate concentrations of 180– 240 μmol/L (3–4 mg/dL). Levels begin to rise in males during puberty but remain low in females until menopause. Mean serum urate values of adult men and premenopausal women are 415 and 360 μmol/L (6.8 and 6 mg/dL), respectively. After menopause, values for women increase to approximate those of men. In adulthood, concentrations rise steadily over time and vary with height, body weight, blood pressure, renal function, and alcohol intake.

HYPERURICEMIA

Hyperuricemia can result from increased production or decreased excretion of uric acid or from a combination of the two processes. Sustained hyperuricemia predisposes some individuals to develop clinical manifestations including gouty arthritis (Chap. 333), urolithiasis, and renal dysfunction (see below).

Hyperuricemia is defined as a plasma (or serum) urate concentration >405 μmol/L (6.8 mg/dL). The risk of developing gouty arthritis or urolithiasis increases with higher urate levels and escalates in proportion to the degree of elevation. Hyperuricemia is present in between 2 and 13.2% of ambulatory adults and is even more frequent in hospitalized individuals.

Causes of Hyperuricemia

Hyperuricemia may be classified as primary or secondary depending on whether the cause is innate or is the result of an acquired disorder. However, it is more useful to classify hyperuricemia in relation to the underlying pathophysiology, i.e., whether it results from increased production, decreased excretion, or a combination of the two (Fig. 359-1, **Table 359-2**).

Table 359-2 Classification of Hyperuricemia by Pathophysiology

Urate Overproduction		
Primary idiopathic	Myeloproliferative diseases	Rhabdomyolysis
HPRT deficiency	Polycythemia vera	Exercise
PRPP synthetase overactivity	Psoriasis	Alcohol
Hemolytic processes	Paget's disease	Obesity
Lymphoproliferative diseases	Glycogenosis III, V, and VII	Purine-rich diet
Decreased Uric Acid Excretion		
Primary idiopathic	Starvation ketosis	Drug ingestion
Renal insufficiency	Berylliosis	Salicylates (>2 g/d)
Polycystic kidney disease	Sarcoidosis	Diuretics
Diabetes insipidus	Lead intoxication	Alcohol
Hypertension	Hyperparathyroidism	Levodopa
Acidosis	Hypothyroidism	Ethambutol
Lactic acidosis	Toxemia of pregnancy	Pyrazinamide
Diabetic ketoacidosis	Bartter's syndrome	Nicotinic acid
	Down syndrome	Cyclosporine
Combined Mechanism		
Glucose-6-phosphatase deficiency	Fructose-1-phosphate aldolase deficiency	Alcohol
		Shock

Abbreviations: HPRT, hypoxanthine phosphoribosyltransferase; PRPP,

phosphoribosylpyrophosphate.

INCREASED URATE PRODUCTION

Diet contributes to the serum urate in proportion to its purine content. Strict restriction of purine intake reduces the mean serum urate level by about 60 μmol/L (1 mg/dL) and urinary uric acid excretion by ~ 1.2 mmol/d (200 mg/d). Foods high in nucleic acid content include liver, "sweetbreads" (i.e., thymus and pancreas), kidney, and anchovy.

Endogenous sources of purine production also influence the serum urate level **(Fig. 359-3)**. De novo purine biosynthesis is an 11-step process that forms inosine monophosphate (IMP). The rates of purine biosynthesis and urate production are determined, for the most part, by amidophosphoribosyltransferase (amidoPRT), which combines phosphoribosylpyrophosphate (PRPP) and glutamine. A secondary regulatory pathway is the salvage of purine bases by hypoxanthine phosphoribosyltransferase (HPRT). HPRT catalyzes the combination of the purine bases hypoxanthine and guanine with PRPP to form the respective ribonucleotides IMP and guanosine monophosphate (GMP).

Figure 359-3

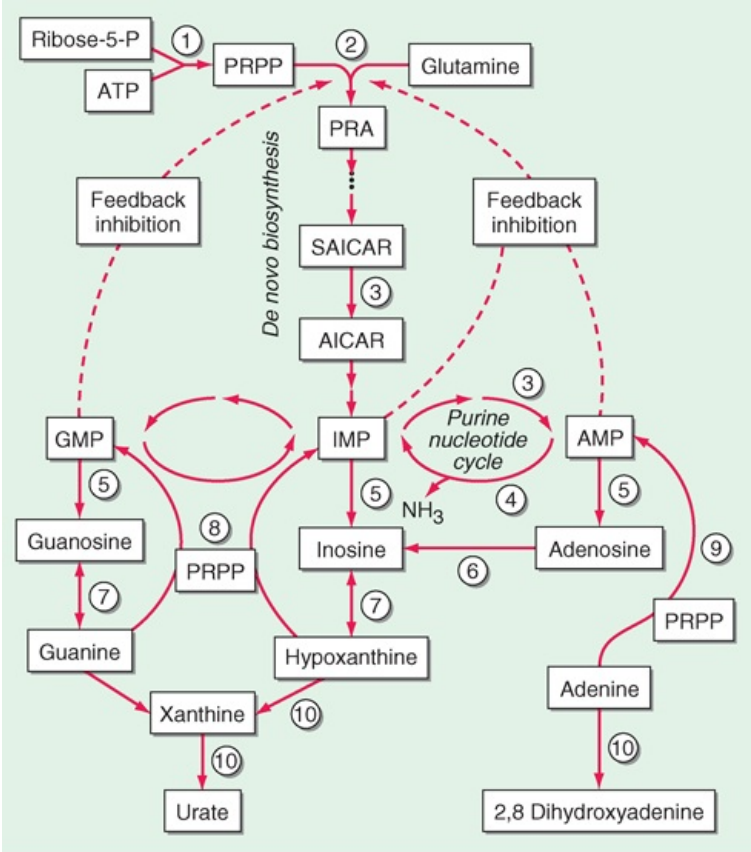

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Abbreviated scheme of purine metabolism. (1) Phosphoribosylpyrophosphate (PRPP) synthetase, (2) amidophosphoribosyltransferase (amidoPRT), (3) adenylosuccinate lyase, (4) (myo-)adenylate (AMP) deaminase, (5) 5'-nucleotidase, (6) adenosine deaminase, (7) purine nucleoside phosphorylase, (8) hypoxanthine phosphoribosyltransferase (HPRT), (9) adenine phosphoribosyltransferase (APRT), and (10) xanthine oxidase. PRA, phosphoribosylamine; SAICAR, succinylaminoimidazole carboxamide ribotide; AICAR, aminoimidazole carboxamide ribotide; GMP, guanylate; IMP, inosine monophosphate; ATP, adenosine triphosphate.

Serum urate levels are closely coupled to the rates of de novo purine biosynthesis, which is driven in part by the level of PRPP, as evidenced by two X-linked inborn errors of purine metabolism. Both increased PRPP synthetase activity and HPRT deficiency are associated with overproduction of purines, hyperuricemia, and hyperuricaciduria (see below for clinical descriptions).

Accelerated purine nucleotide degradation can also cause hyperuricemia, i.e., with conditions of rapid cell turnover, proliferation, or cell death, as in leukemic blast crises, cytotoxic therapy for malignancy, hemolysis, or rhabdomyolysis. Hyperuricemia can result from excessive degradation of skeletal muscle ATP after strenuous physical exercise or status epilepticus and in glycogen storage diseases types III, V, and VII (Chap. 362). The hyperuricemia of myocardial infarction, smoke inhalation, and acute

respiratory failure may also be related to accelerated breakdown of ATP.

DECREASED URIC ACID EXCRETION

More than 90% of individuals with sustained hyperuricemia have a defect in the renal handling of uric acid. Gouty individuals excrete ~40% less uric acid than nongouty individuals for any given plasma urate concentration. Uric acid excretion increases in gouty and nongouty individuals when plasma urate levels are raised by purine ingestion or infusion, but in those with gout, plasma urate concentrations must be 60–120 μmol/L (1–2 mg/dL) higher than normal to achieve equivalent uric acid excretion rates.

Altered uric acid excretion could theoretically result from decreased glomerular filtration, decreased tubular secretion, or enhanced tubular reabsorption. Decreased urate filtration does not appear to cause primary hyperuricemia but does contribute to the hyperuricemia of renal insufficiency. Although hyperuricemia is invariably present in chronic renal disease, the correlation between serum creatinine, urea nitrogen, and urate concentration is poor. Uric acid excretion per unit of glomerular filtration rate increases progressively with chronic renal insufficiency, but tubular secretory capacity tends to be preserved, tubular reabsorptive capacity is reduced, and extrarenal clearance of uric acid increases as renal damage becomes more severe.

Many agents that cause hyperuricemia exert their effects by stimulating reabsorption rather than inhibiting secretion. This appears to occur through a process of "priming" renal urate reabsorption through the sodium-dependent loading of proximal tubular epithelial cells with anions capable of *trans*-stimulating urate reabsorption. The sodium-coupled monocarboxyl transporters SMCT1 and 2 (SLC5A8, SLC5A12) in the brush border of the proximal tubular cells mediate sodium-dependent loading of these cells with monocarboxylates. A similar transporter, SLC13A3, mediates sodium-dependent influx of dicarboxylates into the epithelial cell from the basolateral membrane. Some of these carboxylates are well known to cause hyperuricemia, including pyrazinoate (from pyrazinamide treatment), nicotinate (from niacin therapy), and the organic acids lactate, β-hydroxybutyrate, and acetoacetate. The mono- and divalent anions then become substrates for URAT1 and organic anion transporter (OAT4), respectively, and are exchanged for uric acid from the proximal tubule. Increased blood levels of these anions result in their increased glomerular filtration and greater reabsorption by proximal tubular cells. The increased intraepithelial cell concentrations lead to increased uric acid reabsorption by promoting URAT1- and OAT4-dependent anion exchange. Low doses of salicylates also promote hyperuricemia by this mechanism. Sodium loading of proximal tubular cells also provokes urate retention by reducing extracellular fluid volume and increasing angiotensin II, insulin, and parathyroid hormone release. Additional organic anion transporters OAT1 and OAT3 are involved in the movement of uric acid through the basolateral membrane, although the detailed mechanisms are still being elucidated.

Glucose transporter 9 (GLUT9, SLC2A9) is an electrogenic hexose transporter with splicing variants that mediate co-reabsorption of uric acid along with glucose and fructose at the apical membrane (GLUT9DN/SLC2A9v2), as well as through the basolateral membrane, and thus into the circulation (SLC2A9v1). This might be a mechanism for the observed association of the consumption of fructose-sweetened soft drinks with an increased risk of hyperuricemia and gout. Genomewide association scanning (GWAS) suggests that polymorphisms in SLC2A9 may play an important role in susceptibility to gout in the white population. The presence of one predisposing variant allele increases the relative risk of developing gout by 30–70%, most likely by increasing expression of the shorter isoform, SLC2A9v2 (GLUT9DN). Notably, these polymorphisms explain <5% of the variation in serum uric acid levels in whites.

Alcohol promotes hyperuricemia because of increased urate production and decreased uric acid excretion. Excessive alcohol consumption accelerates hepatic breakdown of ATP to increase urate production. Alcohol consumption can also induce hyperlacticacidemia, which blocks uric acid secretion. The higher purine content in some alcoholic beverages such as beer may also be a factor.

Evaluation

Hyperuricemia does not necessarily represent a disease, nor is it a specific indication for therapy. The decision to treat depends on the cause and the potential consequences of the hyperuricemia in each individual.

Quantification of uric acid excretion can be used to determine whether hyperuricemia is caused by overproduction or decreased excretion. On a purine-free diet, men with normal renal function excrete <3.6 mmol/d (600 mg/d). Thus, the hyperuricemia of individuals who excrete uric acid above this level while on a purine-free diet is due to purine overproduction; for those who excrete lower amounts on the purine-free diet, it is due to decreased excretion. If the assessment is performed while the patient is on a regular diet, the level of 4.2 mmol/d (800 mg/d) can be used as the discriminating value.

Complications

The most recognized complication of hyperuricemia is *gouty arthritis*. In the general population, the prevalence of hyperuricemia ranges between 2.0 and 13.2%, and the prevalence of gout is between

1.3 and 3.7%. The higher the serum urate level, the more likely an individual is to develop gout. In one study, the incidence of gout was 4.9% for individuals with serum urate concentrations >540 μmol/L (9.0 mg/dL) compared with 0.5% for those with values between 415 and 535 μmol/L (7.0 and 8.9 mg/dL). The complications of gout correlate with both the duration and severity of hyperuricemia. **For further discussion of gout, see Chap. 333.**

Hyperuricemia also causes several renal problems: (1) nephrolithiasis; (2) urate nephropathy, a rare cause of renal insufficiency attributed to monosodium urate crystal deposition in the renal interstitium; and (3) uric acid nephropathy, a reversible cause of acute renal failure resulting from deposition of large amounts of uric acid crystals in the renal collecting ducts, pelvis, and ureters.

Nephrolithiasis

Uric acid nephrolithiasis occurs most commonly, but not exclusively, in individuals with gout. In gout, the prevalence of nephrolithiasis correlates with the serum and urinary uric acid levels, reaching ~50% with serum urate levels of 770 μmol/L (13 mg/dL) or urinary uric acid excretion >6.5 mmol/d (1100 mg/d).

Uric acid stones can develop in individuals with no evidence of arthritis, only 20% of whom are hyperuricemic. Uric acid can also play a role in other types of kidney stones. Some nongouty individuals with calcium oxalate or calcium phosphate stones have hyperuricemia or hyperuricaciduria. Uric acid may act as a nidus on which calcium oxalate can precipitate or lower the formation product for calcium oxalate crystallization.

URATE NEPHROPATHY

Urate nephropathy, sometimes referred to as *urate nephrosis*, is a late manifestation of severe gout and is characterized histologically by deposits of monosodium urate crystals surrounded by a giant cell inflammatory reaction in the medullary interstitium and pyramids. The disorder is now rare and cannot be diagnosed in the absence of gouty arthritis. The lesions may be clinically silent or cause proteinuria, hypertension, and renal insufficiency.

URIC ACID NEPHROPATHY

This reversible cause of acute renal failure is due to precipitation of uric acid in renal tubules and collecting ducts that causes obstruction to urine flow. Uric acid nephropathy develops following sudden urate overproduction and marked hyperuricaciduria. Factors that favor uric acid crystal formation include dehydration and acidosis. This form of acute renal failure occurs most often during an aggressive "blastic" phase of leukemia or lymphoma prior to or coincident with cytolytic therapy but has also been observed in individuals with other neoplasms, following epileptic seizures, and after vigorous exercise with heat stress. Autopsy studies have demonstrated intraluminal precipitates of uric acid, dilated proximal tubules, and normal glomeruli. The initial pathogenic events are believed to include obstruction of collecting ducts with uric acid and obstruction of distal renal vasculature.

If recognized, uric acid nephropathy is potentially reversible. Appropriate therapy has reduced the mortality from about 50% to practically nil. Serum levels cannot be relied on for diagnosis because this condition has developed in the presence of urate concentrations varying from 720 to 4800 μmol/L (12–80 mg/dL). The distinctive feature is the urinary uric acid concentration. In most forms of acute renal failure with decreased urine output, urinary uric acid content is either normal or reduced, and the ratio of uric acid to creatinine is <1. In acute uric acid nephropathy the ratio of uric acid to creatinine in a random urine sample or 24-h specimen is >1, and a value that high is essentially diagnostic.

Hyperuricemia and Metabolic Syndrome

Metabolic syndrome (Chap. 242) is characterized by abdominal obesity with visceral adiposity, impaired glucose tolerance due to insulin resistance with hyperinsulinemia, hypertriglyceridemia, increased low-density lipoprotein cholesterol, decreased high-density lipoprotein cholesterol, and hyperuricemia. Hyperinsulinemia reduces the renal excretion of uric acid and sodium. Not surprisingly, hyperuricemia resulting from euglycemic hyperinsulinemia may precede the onset of type 2 diabetes, hypertension, coronary artery disease, and gout in individuals with metabolic syndrome.

Treatment: Hyperuricemia

ASYMPTOMATIC HYPERURICEMIA

Hyperuricemia is present in ~5% of the population and in up to 25% of hospitalized individuals. The vast majority are at no clinical risk. In the past, the association of hyperuricemia with cardiovascular disease and renal failure led to the use of urate-lowering agents for patients with asymptomatic hyperuricemia. This practice is no longer recommended except for individuals receiving cytolytic therapy for neoplastic disease, in which treatment is given in an effort to prevent uric acid nephropathy. Because hyperuricemia can be a component of the metabolic syndrome, its presence is an indication to screen for and aggressively treat any accompanying obesity, hyperlipidemia, diabetes mellitus, or hypertension.

Hyperuricemic individuals are at risk to develop gouty arthritis, especially those with higher serum urate levels. However, most hyperuricemic persons never develop gout, and prophylactic treatment is

not indicated. Furthermore, neither structural kidney damage nor tophi are identifiable before the first attack. Reduced renal function cannot be attributed to asymptomatic hyperuricemia, and treatment of asymptomatic hyperuricemia does not alter the progression of renal dysfunction in patients with renal disease. Increased risk of stone formation in those with asymptomatic hyperuricemia is not established.

Thus, because treatment with specific antihyperuricemic agents entails inconvenience, cost, and potential toxicity, routine treatment of asymptomatic hyperuricemia cannot be justified other than for prevention of acute uric acid nephropathy. In addition, routine screening for asymptomatic hyperuricemia is not recommended. If hyperuricemia is diagnosed, however, the cause should be determined. Causal factors should be corrected if the condition is secondary, and associated problems such as hypertension, hypercholesterolemia, diabetes mellitus, and obesity should be treated.

SYMPTOMATIC HYPERURICEMIA

Nephrolithiasis

(See Chap. 333 for treatment of gout.) Antihyperuricemic therapy is recommended for the individual who has both gouty arthritis and either uric acid– or calcium-containing stones, both of which may occur in association with hyperuricaciduria. Regardless of the nature of the calculi, fluid ingestion should be sufficient to produce a daily urine volume >2 L. Alkalinization of the urine with sodium bicarbonate or acetazolamide may be justified to increase the solubility of uric acid. Specific treatment of uric acid calculi requires reducing the urine uric acid concentration with a xanthine oxidase inhibitor, such as allopurinol or febuxostat. These agents decrease the serum urate concentration and the urinary excretion of uric acid in the first 24 h, with a maximum reduction occurring within 2 weeks. The average effective dose of allopurinol is 300–400 mg/d. Allopurinol can be given once a day because of the long half-life (18 h) of its active metabolite, oxypurinol. The drug is effective in patients with renal insufficiency, but the dose should be reduced. Allopurinol is also useful in reducing the recurrence of calcium oxalate stones in gouty patients and in nongouty individuals with hyperuricemia or hyperuricaciduria. Febuxostat (40–80 mg/d) is also taken once daily, and doses do not need to be adjusted in the presence of mild to moderate renal dysfunction. Potassium citrate (30–80 mmol/d orally in divided doses) is an alternative therapy for patients with uric acid stones alone or mixed calcium/uric acid stones. A xanthine oxidase inhibitor is also indicated for the treatment of 2,8-dihydroxyadenine kidney stones.

Uric Acid Nephropathy

Uric acid nephropathy is often preventable, and immediate, appropriate therapy has greatly reduced the mortality rate. Vigorous IV hydration and diuresis with furosemide dilute the uric acid in the tubules and promote urine flow to ≥100 mL/h. The administration of acetazolamide, 240–500 mg every 6–8 h, and sodium bicarbonate, 89 mmol/L, IV enhances urine alkalinity and thereby solubilizes more uric acid. It is important to ensure that the urine pH remains >7.0 and to watch for circulatory overload. In addition, antihyperuricemic therapy in the form of allopurinol in a single dose of 8 mg/kg is administered to reduce the amount of urate that reaches the kidney. If renal insufficiency persists, subsequent daily doses should be reduced to 100–200 mg because oxypurinol, the active metabolite of allopurinol, accumulates in renal failure. Despite these measures, hemodialysis may be required. Urate oxidase (Rasburicase) can also be administered IV to prevent or to treat tumor lysis syndrome.

HYPOURICEMIA

Hypouricemia, defined as a serum urate concentration <120 μmol/L (2.0 mg/dL) can result from decreased production of urate, increased excretion of uric acid, or a combination of both mechanisms. It occurs in <0.2% of the general population and <0.8% of hospitalized individuals. Hypouricemia causes no symptoms or pathology and therefore requires no therapy.

Most hypouricemia results from increased renal uric acid excretion. The finding of normal amounts of uric acid in a 24-h urine collection in an individual with hypouricemia is evidence for a renal cause. Medications with uricosuric properties (Table 359-1) include aspirin (at doses >2.0 g/d), losartan, fenofibrate, x-ray contrast materials, and glyceryl guaiacolate. Total parenteral hyperalimentation can also cause hypouricemia, possibly a result of the high glycine content of the infusion formula. Other causes of increased urate clearance include conditions such as neoplastic disease, hepatic cirrhosis, diabetes mellitus, and inappropriate secretion of vasopressin; defects in renal tubular transport such as primary Fanconi syndrome and Fanconi syndromes caused by Wilson's disease, cystinosis, multiple myeloma, and heavy metal toxicity; and isolated congenital defects in the bidirectional transport of uric acid. Hypouricemia can be familial, generally inherited in an autosomal recessive manner. Most cases are caused by a loss of function mutation in *SLC12A12*, the gene that encodes URAT-1, resulting in increased renal urate clearance. Individuals with normal *SLC12A12* most likely have a defect in other urate transporters. Although usually asymptomatic, some patients suffer from urate nephrolithiasis or exercise-induced renal failure.

SELECTED INBORN ERRORS OF PURINE METABOLISM

(See also **Table 359-3**, Fig. 359-3.) More than 30 defects in human purine and pyrimidine metabolic

pathways have been identified thus far. Many are benign, but about half are associated with clinical manifestations, some causing major morbidity and mortality. Advances in genetics, as well as high-performance liquid chromatography and tandem mass spectrometry, have allowed for better diagnosis.

Table 359-3 Inborn Errors of Purine Metabolism

Enzyme	Activity	Inheritance	Clinical Features	Laboratory Features
Hypoxanthine phosphoribo—syltransferase	Complete deficiency	X-linked	Self-mutilation, choreoathetosis, gout, and uric acid lithiasis	Hyperuricemia, hyperuricosuria
	Partial deficiency	X-linked	Gout, and uric acid lithiasis	Hyperuricemia, hyperuricosuria
Phosphoribosylpyrophosphate synthetase	Overactivity	X-linked	Gout, uric acid lithiasis, and deafness	Hyperuricemia, hyperuricosuria
Adenine phosphoribosyl—transferase	Deficiency	Autosomal recessive	2,8-Dihydroxyadenine lithiasis	—
Xanthine oxidase	Deficiency	Autosomal recessive	Xanthinuria and xanthine lithiasis	Hypouricemia, hypouricosuria
Adenylosuccinate lyase	Deficiency	Autosomal recessive	Autism and psychomotor retardation	—
Myoadenylate deaminase	Deficiency	Autosomal recessive	Myopathy with exercise intolerance or asymptomatic	—
Adenosine deaminase	Deficiency	Autosomal recessive	Severe combined immunodeficiency disease and chondroosseous dysplasia	—
Purine nucleoside phosphorylase	Deficiency	Autosomal recessive	T cell–mediated immunodeficiency	—

PURINE DISORDERS

Hprt Deficiency

The HPRT gene is located on the X chromosome. Affected males are hemizygous for the mutant gene; carrier females are asymptomatic. A complete deficiency of HPRT, the Lesch-Nyhan syndrome, is characterized by hyperuricemia, self-mutilative behavior, choreoathetosis, spasticity, and mental retardation. A partial deficiency of HPRT, the Kelley-Seegmiller syndrome, is associated with hyperuricemia but no central nervous system manifestations. In both disorders, the hyperuricemia results from urate overproduction and can cause uric acid crystalluria, nephrolithiasis, obstructive uropathy, and gouty arthritis. Early diagnosis and appropriate therapy with allopurinol can prevent or eliminate all the problems attributable to hyperuricemia but have no effect on the behavioral or neurologic abnormalities.

Increased Prpp Synthetase Activity

Like the HPRT deficiency states, PRPP synthetase overactivity is X-linked and results in gouty arthritis and uric acid nephrolithiasis. Nerve deafness occurs in some families.

Adenine Phosphoribosyltransferase (Aprt) Deficiency

APRT deficiency is inherited as an autosomal recessive trait. Affected individuals develop kidney stones composed of 2,8-dihydroxyadenine. Whites with the disorder have a complete deficiency (type I), whereas Japanese subjects have some measurable enzyme activity (type II). Expression of the defect is similar in the two populations, as is the frequency of the heterozygous state (0.4–1.1 per 100). Allopurinol treatment prevents stone formation.

Hereditary Xanthinuria

A deficiency of xanthine oxidase causes all purine in the urine to occur in the form of hypoxanthine and xanthine. About two-thirds of deficient individuals are asymptomatic. The remainder develop kidney stones composed of xanthine.

Myoadenylate Deaminase Deficiency

Primary (inherited) and secondary (acquired) forms of myoadenylate deaminase deficiency have been described. The primary form is inherited as an autosomal recessive trait. Clinically, some may have relatively mild myopathic symptoms with exercise or other triggers, but most individuals with this defect are asymptomatic. Therefore, another explanation for the myopathy should be sought in symptomatic patients with this deficiency. The acquired deficiency occurs in association with a wide variety of neuromuscular disease, including muscular dystrophies, neuropathies, inflammatory myopathies, and

collagen vascular diseases.

Adenylosuccinate Lyase Deficiency

Deficiency of this enzyme is due to an autosomal recessive trait and causes profound psychomotor retardation, seizures, and other movement disorders. All individuals with this deficiency are mentally retarded, and most are autistic.

Adenosine Deaminase Deficiency and Purine Nucleoside Phosphorylase Deficiency

See Chap. 316.

PYRIMIDINE DISORDERS

The pyrimidine cytidine is found in both DNA and RNA; it is a complementary base pair for guanine. Thymidine is found only in DNA, where it is paired with adenine. Uridine is found only in RNA and can pair with either adenine or guanine in RNA secondary structures. Pyrimidines can be synthesized by a de novo pathway **(Fig. 359-4)** or reused in a salvage pathway. Although >25 different enzymes are involved in pyrimidine metabolism, disorders of these pathways are rare. Seven disorders of pyrimidine metabolism have been discovered **(Table 359-4)**, three of which are discussed below.

Figure 359-4

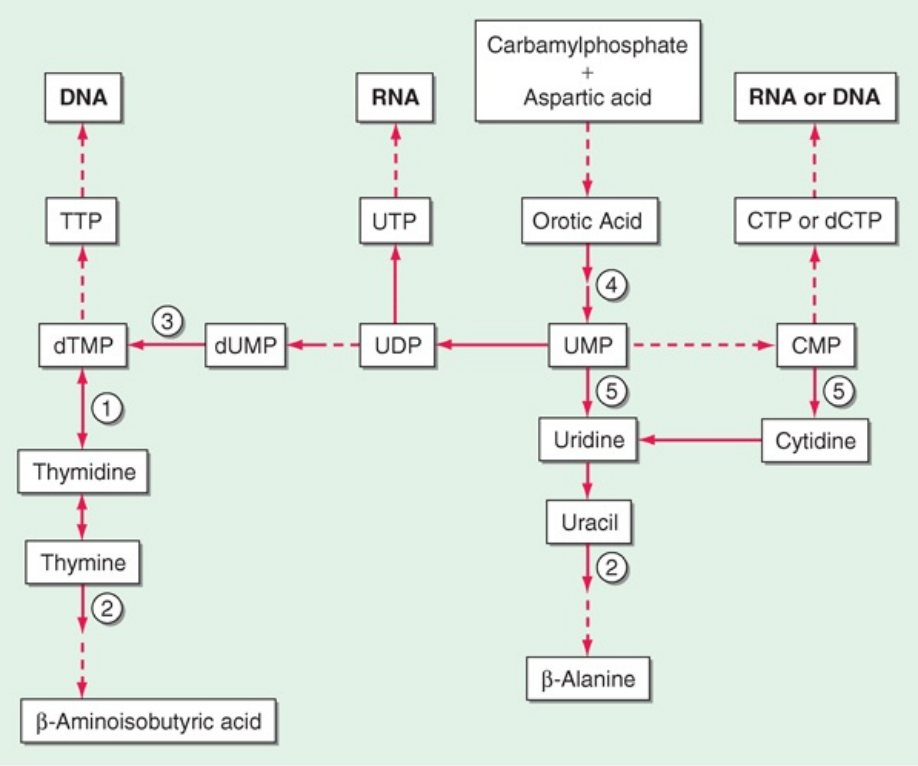

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Abbreviated scheme of pyrimidine metabolism. (1) thymidine kinase, (2) dihydropyrimidine dehydrogenase, (3) thymidylate synthase, (4) UMP synthase, (5) 5'-nucleotidase. CMP, cytidine-5'-monophosphate; UMP, uridine-5'-monophosphate; UDP, uridine-5'-diphosphate; dUMP, deoxyuridine-5'-monophosphate; dTMP, deoxythymidine-5'-monophosphate; TTP, thymidine triphosphate; UTP, uridine triphosphate.

Table 359-4 Inborn Errors of Pyrimidine Metabolism

Enzyme	Activity	Inheritance	Clinical Features	Laboratory Features
Uridine-5'-monophosphate synthetase	Deficiency	Autosomal recessive	Orotic acid crystalluria; obstructive uropathy, hypochromic megaloblastic anemia	Orotic aciduria
Pyrimidine 5'-nucleotidase	Deficiency	Autosomal recessive	Hemolytic anemia	Basophilic stippling of erythrocytes; high levels of cytidine and uridine ribonucleotides

Pyrimidine 5'-nucleotidase	Superactivity	Uncertain	Developmental delay, seizures, ataxia, language deficit	Hypouricosuria
Thymidine phosphorylase	Deficiency	Autosomal recessive	Mitochondrial neurogastrointestinal encephalopathy	Hypouricosuria
Dihydropyrimidine dehydrogenase	Deficiency	Autosomal recessive	Seizures, motor and mental retardation	High levels of uracil, thymine, and 5-hydroxymethyluracil and low levels of dihydropyrimidines in urine
Dihydropyrimidinase	Deficiency	Uncertain	Seizures, mental retardation	Dihydropyrimidinuria
Ureidopropionase	Deficiency	Uncertain	Hypotonia, dystonia, developmental delay	High urinary excretion of N-carbamyl-β-alanine and N-carbamyl β-aminoisobutyric acid

Orotic Aciduria

Hereditary orotic aciduria is caused by mutations in a bifunctional enzyme, uridine-5'-monophosphate (UMP) synthase, which converts orotic acid to UMP in the de novo synthesis pathway (Fig. 359-4). The disorder is characterized by hypochromic megaloblastic anemia that is unresponsive to vitamin B12 and folic acid, growth retardation, and neurologic abnormalities. Increased excretion of orotic acid causes crystalluria and obstructive uropathy. Replacement of uridine (100–200 mg/kg per day) corrects the anemia, reduces orotic acid excretion, and improves the other sequelae of the disorder.

Pyrimidine 5'-Nucleotidase Deficiency

Pyrimidine 5'-nucleotidase catalyzes the removal of the phosphate group from pyrimidine ribonucleoside monophosphates (cytidine-5'-monophosphate or UMP) (Fig. 359-4). An inherited deficiency of this enzyme causes hemolytic anemia with prominent basophilic stippling of erythrocytes. The accumulation of pyrimidines or cytidine diphosphate choline (CDPC) is thought to induce hemolysis. There is no specific treatment. Acquired pyrimidine 5'-nucleotidase deficiency has been reported in lead poisoning and in thalassemia.

Dihydropyrimidine Dehydrogenase Deficiency

Dihydropyrimidine dehydrogenase (DPD) is the rate-limiting enzyme in the pathway of uracil and thymine degradation (Fig. 359-4). Deficiency of this enzyme causes excessive urinary excretion of uracil and thymine. DPD deficiency causes nonspecific cerebral dysfunction with convulsive disorders, motor retardation, and mental retardation. No specific treatment is available.

Medication Effects on Pyrimidine Metabolism

A variety of medications can influence pyrimidine metabolism. The anticancer agents fluorodeoxyuridine and 5-fluorouracil (5-FU) and the antimicrobial agent fluorocytosine cause cytotoxicity when converted to fluorodeoxyuridylate (FdUMP), a specific suicide inhibitor of thymidylate synthase. Fluorocytosine must be converted to 5-FU to be effective. This conversion is catalyzed by cytosine deaminase activity. Fluorocytosine's action is selective because cytosine deaminase is present in bacteria and fungi but not in human cells.

DPD is involved in the degradation of 5-FU. Consequently, deficiency of this enzyme is associated with 5-FU neurotoxicity.

Leflunomide, which is used to treat rheumatoid arthritis, inhibits de novo pyrimidine synthesis by inhibiting dihydroorotate dehydrogenase, resulting in an antiproliferative effect on T cells. Allopurinol, which is used to block xanthine oxidase and purine synthesis, also inhibits orotidine-5'-phosphate decarboxylase, a step in UMP synthesis. Consequently, allopurinol use is associated with increased excretion of orotidine and orotic acid; there are no known clinical effects of this inhibition.

FURTHER READINGS

Burns CM, Wortmann RL: Gout therapeutics: New drugs for an old disease. Lancet 377:165, 2011[PMID: 20719377] [Full Text]

Dalbeth N, Merriman T: Crystal ball gazing: New therapeutic targets for hyperuricaemia and gout. Rheumatol 48:222, 2009[PMID: 19109320] [Full Text]

Van Gennip AH et al: Inborn errors of purine and pyrimidine metabolism, in *Physician's Guide to the Treatment and Follow-up of Metabolic Diseases*, N Blau et al (eds). New York, Springer-Verlag, 2005, pp 245–254

Wortmann RL: Gout and hyperuricemia, in *Kelly's Textbook of Rheumatology* , 8th ed, GS Firestein et al (eds). Philadelphia, Saunders, 2009, pp 1481–1506

——— et al (eds): *Crystal-Induced Arthropathies*. New York, Informa Healthcare, 2006, pp 189–212, 255–276, 369–400

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

WILSON'S DISEASE: INTRODUCTION

Wilson's disease is an autosomal recessive disorder caused by mutations in the *ATP7B* gene, a membrane-bound, copper-transporting ATPase. Clinical manifestations are caused by copper toxicity and primarily involve the liver and the brain. Because effective treatment is available, it is important to make this diagnosis early.

The frequency of Wilson's disease in most populations is about 1 in 30,000–40,000, and the frequency of carriers of *ATP7B* mutations is ~1%. Siblings of a diagnosed patient have a 1 in 4 risk of Wilson's disease, whereas children of an affected patient have about a 1 in 200 risk. Because a large number of inactivating mutations have been reported in the *ATP7B* gene, mutation screening for diagnosis is not routine, although this may be practical in the future. DNA haplotype analysis can be used to genotype siblings of an affected patient.

Pathogenesis

ATP7B protein deficiency impairs biliary copper excretion, resulting in positive copper balance, hepatic copper accumulation, and copper toxicity from oxidant damage. Excess hepatic copper is initially bound to metallothionein, but as this storage capacity is exceeded, liver damage begins as early as three years of age. Defective copper incorporation into apoceruloplasmin leads to excess catabolism and low blood levels of ceruloplasmin. Serum copper levels are usually lower than normal because of low blood ceruloplasmin, which normally binds >90% of serum copper. As the disease progresses, nonceruloplasmin serum copper ("free" copper) levels increase, resulting in copper buildup in other parts of the body, such as the brain, leading to neurologic and psychiatric disease.

Clinical Presentation

HEPATIC

Wilson's disease may present as hepatitis, cirrhosis, or as hepatic decompensation, typically in the mid to late teenage years in western countries, although the age of presentation is quite broad and extends into the fifth decade of life.

An episode of hepatitis may occur, with elevated blood transaminase enzymes, with or without jaundice, and then spontaneously regress. Hepatitis often reoccurs, and most of these patients eventually develop cirrhosis.

Hepatic decompensation is associated with elevated serum bilirubin, reduced serum albumin and coagulation factors, ascites, peripheral edema, and hepatic encephalopathy. In severe hepatic failure, hemolytic anemia may occur because large amounts of copper derived from hepatocellular necrosis are released into the bloodstream. The association of hemolysis and liver disease makes Wilson's disease a likely diagnosis.

NEUROLOGIC

The neurologic manifestations of Wilson's disease typically occur in patients in their early twenties, although the age of onset extends into the sixth decade of life. MRI and CT scans reveal damage in the basal ganglia and occasionally in the pons, medulla, thalamus, cerebellum, and subcortical areas. The three main movement disorders include dystonia, incoordination, and tremor. Dysarthria and dysphagia are common. In some patients, the clinical picture closely resembles that of Parkinson's disease. Dystonia can involve any part of the body and eventually leads to grotesque positions of the limbs, neck, and trunk. Autonomic disturbances may include orthostatic hypotension and sweating abnormalities as well as bowel, bladder, and sexual dysfunction. Memory loss, migraine-type headaches, and seizures may occur. Patients have difficulties focusing on tasks, but cognition is not usually grossly impaired. Sensory abnormalities and muscular weakness are not features of the disease.

PSYCHIATRIC

A history of behavioral disturbances, with onset in the five years before diagnosis, is present in half of patients with neurologic disease. The features are diverse and may include loss of emotional control (temper tantrums, crying bouts), depression, hyperactivity, or loss of sexual inhibition.

OTHER MANIFESTATIONS

Some female patients have repeated spontaneous abortions, and most become amenorrheic prior to

diagnosis. Cholelithiasis and nephrolithiasis occur with increased frequency. Some patients have osteoarthritis, particularly of the knee. Microscopic hematuria is common, and increased urinary excretion of phosphates, amino acids, glucose, or urates may occur; however, a full-blown Fanconi syndrome is rare. Sunflower cataracts and Kayser-Fleischer rings (copper deposits in the outer rim of the cornea) may be seen. Electrocardiographic and other cardiac abnormalities have been reported but are not common.

Diagnosis

Diagnostic tests for Wilson's disease are listed in **Table 360-1**. Serum ceruloplasmin levels should not be used for definitive diagnosis, because they are normal in up to 10% of affected patients and are reduced in 20% of carriers. Kayser-Fleischer rings **(Fig. 360-1)** can only be diagnosed definitively by an ophthalmologist using a slit lamp. They are present in >99% of patients with neurologic/psychiatric forms of the disease and have been described very rarely in the absence of Wilson's disease. Kayser-Fleischer rings are present in only ~30–50% of patients diagnosed in the hepatic or presymptomatic state; thus, the absence of rings does not exclude the diagnosis.

Table 360-1 Useful Tests for Wilson's Disease

Test	Usefulness[a]	Normal Value	Heterozygous Carriers	Wilson's disease
Serum ceruloplasmin	+	180–350 mg/L (18–35 mg/dL)	Low in 20%	Low in 90%
KF rings	++	Absent	Absent	Present in 99% + if neurologic or psychiatric symptoms presentPresent in 30–50% in hepatic presentation and presymptomatic state
24-h urine Cu	+++	0.3–0.8 mmol (20–50 mg)	Normal to 1.3 mmol (80 mg)	>1.6 mmol (>100 mg) in symptomatic patients0.9 to >1.6 μmol (60 to >100 μg) in presymptomatic patients
Liver Cu	++++	0.3–0.8 mmol/g (20–50 mg per g tissue)	Normal to 2.0 mmol (125 mg)	>3.1 mmol (>200 mg) (obstructive liver disease can cause false-positive results)
Haplotype analysis	++++ (Siblings only)	0 Matches	1 Match	2 Matches

[a]Usefulness: +, somewhat useful, to ++++, very useful.

Abbreviations: Cu, copper; KF, Kayser-Fleischer.

Figure 360-1

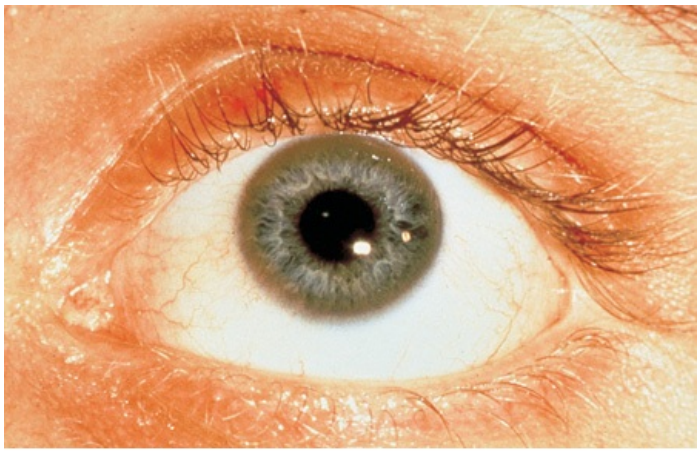

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A Kayser-Fleischer ring. Although in this case, the brownish ring rimming the cornea is clearly visible to the naked eye, confirmation is usually by slit-lamp examination.

Urine copper is an important diagnostic tool, but it must be collected carefully to avoid contamination. Symptomatic patients invariably have urine copper levels > 1.6 mmol (>100 mg) per 24 h.

Heterozygotes have values < 1.3 mmol (<80 mg) per 24 h. About half of presymptomatic patients who are ultimately affected have diagnostically elevated urine copper values, but the other half are in an intermediate range between 0.9 and 1.6 mmol (60 and 100 mg) per 24 h. Because heterozygotes may have values up to 1.3 mmol (80 mg) per 24 h, patients in this range may require a liver biopsy for definitive diagnosis.

The gold standard for diagnosis remains liver biopsy with quantitative copper assays. Affected patients have values > 3.1 mmol/g (> 200 mg/g dry weight of liver). Copper stains are not reliable. False-positive results can occur with long-standing obstructive liver disease, which can elevate hepatic and urine copper and rarely causes Kayser-Fleischer rings.

Treatment: Wilson's Disease

Recommended anticopper treatments are listed in **Table 360-2**. Penicillamine was previously the primary anticopper treatment but now plays a minor role because of its toxicity and because it often worsens existing neurologic disease if used as initial therapy. If penicillamine is given, it should always be accompanied by 25 mg/d of pyridoxine. Trientine is a less toxic chelator and is supplanting penicillamine when a chelator is indicated.

Table 360-2 Recommended Anticopper Drugs for Wilson's Disease

Disease Status	First Choice	Second Choice
Initial hepatic		
Hepatitis or cirrhosis without decompensation	Zinc[a]	Trientine
Hepatic decompensation		
Mild	Trientineb and zinc	Penicillamineb and zinc
Moderate	Trientine and zinc	Hepatic transplantation
Severe	Hepatic transplantation	Trientine and zinc
Initial neurologic/psychiatric	Tetrathiomolybdatec and zinc	Zinc
Maintenance	Zinc	Trientine
Presymptomatic	Zinc	Trientine
Pediatric	Zinc	Trientine
Pregnant	Zinc	Trientine

[a]Zinc acetate is supplied as Galzin, manufactured by Gate Pharmaceutical. Recommended adult dose for all the above indications is 50 mg of elemental zinc three times daily, each dose separated from food and beverages other than water by at least 1 h, and separated from trientine or penicillamine doses by at least 1 h.

[b]Trientine is supplied as Syprine and penicillamine as Cuprimine, both manufactured by Merck. Recommended adult dosage for both drugs is 500 mg twice daily, each dose at least 1/2 h before or 2 h after meals.

[c]Tetrathiomolybdate is being studied in clinical trials.

For patients with hepatitis or cirrhosis, but without evidence of hepatic decompensation or neurologic/psychiatric symptoms, zinc is the therapy of choice, although some advocate therapy with trientine. Zinc has proven efficacy in Wilson's disease and is essentially nontoxic. It produces a negative copper balance by blocking intestinal absorption of copper, and it induces hepatic metallothionein synthesis, which sequesters additional toxic copper. All presymptomatic patients should be treated prophylactically, because the disease is close to 100% penetrant.

The first step in evaluating patients presenting with hepatic decompensation is to establish disease severity, which can be estimated using the Nazer prognostic index **(Table 360-3)**. Patients with scores < 7 can usually be managed with medical therapy. Patients with scores > 9 should be considered immediately for liver transplantation, and those with scores between 7 and 9 require clinical judgment as to whether to recommend transplantation or medical therapy. A combination of trientine and zinc has been used to treat patients with Nazer scores as high as 9, but such patients should be watched carefully for indications of hepatic deterioration, which mandates transplantation.

Table 360-3 Prognostic Index of Nazer

		Score (in Points)				
Laboratory Measurement	**Normal Value**	**0**	**1**	**2**	**3**	**4**
Serum bilirubina	0.2–1.2 mg/dL	<5.8	5.8–8.8	8.8–11.7	11.7–17.5	>17.5

Serum aspartate transferase (AST)	10–35 IU/L	<100	100–150	151–200	201–300	>300
Prolongation of prothrombin time (seconds)	—	<4	4–8	9–12	13–20	>20

[a]If hemolysis is present, the serum bilirubin cannot be used as a measure of liver function until the hemolysis subsides.

Source: Modified from H Nazer et al: Gut 27:1377, 1986; with permission from BMJ Publishing Group.

For initial medical therapy of patients with hepatic decompensation, a chelator (trientine is preferred) plus zinc is recommended (Table 360-2). Zinc should not, however, be ingested simultaneously with trientine, because it will chelate zinc and form therapeutically ineffective complexes; administration of the two drugs should be separated by at least one hour.

For initial neurologic therapy, tetrathiomolybdate is emerging as the drug of choice because of its rapid control of free copper, preservation of neurologic function, and low toxicity. Penicillamine and trientine should be avoided because they each have a high risk of worsening the neurologic condition. Until tetrathiomolybdate is commercially available, zinc therapy is recommended. Although it is relatively slow-acting, zinc itself does not cause neurologic worsening. Although hepatic transplantation may improve neurologic symptoms, it does so only by removing copper, which can be done more safely and inexpensively with anticopper drugs. Pregnant patients should be treated with zinc or trientine throughout pregnancy, but without tight copper control, because copper deficiency can be teratogenic.

Anticopper therapy must be lifelong. With treatment, liver function usually recovers after about a year, although residual liver damage is usually present. Neurologic and psychiatric symptoms usually improve after 6 to 24 months of treatment.

Monitoring Anticopper Therapy

When first using trientine or penicillamine, it is necessary to monitor for drug toxicity—particularly bone marrow suppression and proteinuria. Complete blood counts, standard biochemical profiles, and a urinalysis should be performed at weekly intervals for a month, then at twice-weekly intervals for two to three months, then at monthly intervals for three or four months, and at four- to six-month intervals thereafter.

The anticopper effects of trientine and penicillamine can be monitored by following 24-h "free" serum copper. Changes in urine copper are more difficult to interpret because excretion reflects the effect of the drug, as well as body loading with copper. Free serum copper is calculated by subtracting the ceruloplasmin copper from the total serum copper. Each 10 mg/L (mg/dL) of ceruloplasmin contributes 0.5 mmol/L (3 mg/dL) of serum copper. The normal free copper value is 1.6–2.4 mmol/L (10–15 mg/dL), and it is often as high as 7.9 mmol/L (50 mg/dL) in untreated Wilson's disease. With treatment, free copper should be <3.9 mmol/L (<25 mg/dL).

Zinc treatment does not require blood or urine monitoring for toxicity. Its only significant side effect is gastric burning or nausea in ~10% of patients, usually with the first morning dose. This can be mitigated by taking the first dose an hour after breakfast or taking the zinc with a small amount of protein. Because zinc mainly affects stool copper, 24-h urine copper can be used to reflect body loading. The typical value in untreated symptomatic patients is >3.1 mmol (>200 mg) per 24 h. This level should decrease during the first 1–2 years of therapy to <2.0 mmol (<125 mg) per 24 h. A normal value [0.3 to 0.8 mmol (20 to 50 mg)] is rarely reached during the first decade of therapy and should raise concern about overtreatment (copper deficiency), the first sign of which is anemia and/or leukopenia.

GLOBAL CONSIDERATIONS

Age of onset of clinical disease may be considerably younger in India and far Eastern countries, often occurring in children only five or six years of age. The incidence of the disease may be increased in certain populations due to founder effects. For example, in Sardinia, the incidence may be 1 in 3000. In countries where penicillamine, trientine, and zinc acetate (as Galzin) are not available or cannot be afforded, zinc salts such as the gluconate or sulfate provide an alternative treatment option.

FURTHER READINGS

Askari FK et al: Treatment of Wilson's disease with zinc XVIII. Initial treatment of the hepatic decompensation presentation with trientine and zinc. J Lab Clin Med 142:385, 2003[PMID: 14713890] [Full Text]

Brewer GJ, Askari FK: Wilson's disease: Clinical management and therapy. J Hepatol 42:S13, 2005

——— et al: Treatment of Wilson's disease with ammonium tetrathiomolybdate: V. Control of free copper by tetrathiomolybdate and a comparison with trientine. Transl Res. 154:70, 2009

Cox DW, Roberts EA: Wilson's disease. GeneClinics, University of Washington, Seattle. Online.

Available at *http://www.geneclinics.org/profiles/wilson/details.html*

Merle U et al: Clinical presentation, diagnosis, and long-term outcome of Wilson's disease: A cohort study. Gut 56:115, 2007[PMID: 16709660] [Full Text]

Roberts EA, Schilsky ML: Diagnosis and treatment of Wilson's disease: An update. Hepatology. 47:2089, 2008[PMID: 18506894] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

LYSOSOMAL STORAGE DISEASES: INTRODUCTION

Lysosomes are heterogeneous subcellular organelles containing specific hydrolyses that allow selective processing or degradation of proteins, nucleic acids, carbohydrates, and lipids. There are >40 different lysosomal storage diseases, classified based on the nature of the stored material **(Table 361-1)**. Several of the most prevalent disorders are reviewed here, including Tay-Sachs disease, Fabry's disease, Gaucher's disease, Niemann-Pick disease, the mucopolysaccharidosis, and Pompe's disease. Lysosomal storage diseases should be considered in the differential diagnosis of patients with neurologic, renal, or muscular degeneration and/or unexplained hepatomegaly, splenomegaly, cardiomyopathy, or skeletal dysplasias and deformations. Physical findings are disease-specific, and enzyme assays or genetic testing can be used to make a definitive diagnosis. Although the nosology of these diseases segregates the variants into distinct phenotypes, these are heuristic and, in the clinic, each disease exhibits, more or less, a continuous spectrum of manifestations from severe to attenuated variants.

Table 361-1 Selected Lysosomal Storage Diseases

Disorder[a]	Enzyme Deficiency [Specific Therapy]	Stored Material	Clinical Types (Onset)	Inheritance	Clinical Features: Neurologic	Liver Spleen Enlargement	Skeletal Dysplasia	Ophthalmologic	Hematologic	Unique Features
Mucopolysaccharidoses (MPS)										
MPS I H, Hurler (136)	α-L-Iduronidase [ET, BMT]	Dermatan sulfate Heparan sulfate	Infantile Intermediate	AR	Mental retardation Mental retardation	+ + +	+ + + +	Corneal clouding	Vacuolated lymphocytes	Coarse facies; cardiovascular involvement; joint stiffness
MPS I H/S, Hurler/Scheie			Adult		Mental retardation					
MPS I S, Scheie					None					
MPS II, Hunter (136)	Iduronate sulfatase [ET]	Dermatan sulfate Heparan sulfate	Severe infantile Mild juvenile	X-linked	Mental retardation, less in mild form	+ + +	+ + + +	Retinal degeneration, no corneal clouding	Granulated lymphocytes	Coarse facies; cardiovascular involvement; joint stiffness; distinctive pebbly skin lesions
MPS III A, Sanfilippo A (136)	Heparan-N-sulfatase	Heparan sulfate	Late infantile	AR	Severe mental retardation	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III B, Sanfilippo B	N-Acetyl-α-glucosaminidase	Heparan sulfate	Late infantile	AR	Severe mental retardation	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III C, Sanfilippo C	Acetyl-CoA: α-glucosaminide N-acetyltransferase	Heparan sulfate	Late infantile	AR	Severe mental retardation	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III D, Sanfilippo D	N-Acetylglucosamine-6-sulfate sulfatase	Heparan sulfate	Late infantile	AR	Severe mental retardation	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS IV A, Morquio (136)	N-Acetylgalactosamine-6-sulfate sulfatase	Keratan sulfate Chondroitin-6 sulfate	Childhood	AR	None	+	+ + + +	Corneal clouding	Granulated neutrophils	Distinctive skeletal deformity; odontoid hypoplasia; aortic valve disease
MPS IV B, Morquio (136)	β-Galactosidase		Childhood	AR	None	±	+ + + +			
MPS VI, Maroteaux-Lamy (136)	Arylsulfatase B [ET, BMT]	Dermatan sulfate	Late infantile	AR	None	+ +	+ + + +	Corneal clouding	Granulated neutrophils and lymphocytes	Coarse facies; valvular heart disease
MPS VII (136)	β-Glucuronidase	Dermatan sulfate Heparan sulfate	Neonatal Infantile Adult	AR	Mental retardation, absent in some adults	+ + +	+ + +	Corneal clouding	Granulated neutrophils	Coarse facies; vascular involvement; hydrops fetalis in neonatal form
GM2 Gangliosidoses										
Tay-Sachs disease (153)	β-Hexosaminidase A	GM2 gangliosides	Infantile Juvenile	AR	Mental retardation; seizures; later juvenile form	None	None	Cherry red spot in infantile form	None	Macro—cephaly; hyperacusis in infantile form
Sandhoff's—disease (153)	β-Hexosaminidases A and B	GM2 gangliosides	Infantile	AR	Mental retardation; seizures	+ +	±	Cherry red spot	None	Macro—cephaly; hyperacusis
Neutral Glycosphingolipidoses										
Fabry's disease (150)	α-Galactosidase A [ET]	Globotriaosylceramide	Childhood	X-linked	Painful acroparesthesias	None	None	Corneal dystrophy, vascular lesions	None	Cutaneous angioke—ratomas; hypo—hydrosis
Gaucher's—disease (146)	Acid β-glucosidase [ET, SRT]	Glucosylceramide	Type 1	AR	None	+ + + +	+ + + +	None	Gaucher's cells in bone marrow; cytopenias	Adult form highly variable
			Type 2		++++	+ + +	+	Eye movements		
			Type 3		++	+ + + +	+ + + +	Eye movements		
Niemann-Pick disease (144) A and B	Sphingomyelinase	Sphingomyelin	Neuronopathic, type A Nonneuronopathic, type B	AR	Mental retardation; seizures	+ + + +	None Osteoporosis	Macular degeneration	Foam cells in bone marrow	Pulmonary infiltrates Lung failure
Glycoproteinoses										
Fucosidosis (140)	α-Fucosidase	Glycopeptides; oligosaccharides	Infantile Juvenile	AR	Mental retardation	+ +	+ +	None	Vacuolated lymphocytes; foam cells	Coarse facies; angiokeratomas in juvenile form
α-Mannosidosis (140)	α-Mannosidase	Oligosaccharides	Infantile Milder variant	AR	Mental retardation	+ + +	+ +	Cataracts, corneal clouding	Vacuolated lymphocytes, granulated neutrophils	Coarse facies; enlarged tongue
β-Mannosidosis (140)	β-Mannosidase	Oligosaccharides		AR	Seizures; mental retardation		+ +	None	Vacuolated lymphocytes, foam cells	Angioker—atomas
Aspartylglu—cosaminuria (140)	Aspartylglucosaminidase	Aspartylglucosamine; glycopeptides	Young adult onset	AR	Mental retardation	±	+ +	None	Vacuolated lymphocytes, foam cells	Coarse facies
Sialidosis (140)	Neuraminidase	Sialyloligosaccharides	Type I, congenital Type II, infantile and juvenile forms	AR	Myoclonus; mental retardation	+ +, less in type I	+ + less in type I	Cherry red spot	Vacuolated lymphocytes	MPS phenotype in type II
Mucolipidoses (ML)										
ML-II, I-cell disease (138)	UDP-N-Acetylglucosamine-1-phosphotransferase	Glycoprotein; glycolipids	Infantile	AR	Mental retardation	±	+ + + +	Corneal clouding	Vacuolated and granulated neutrophils	Coarse facies; absence of mucopolysacchariduria; gingival hypoplasia
ML-III, pseudo-Hurler polydystrophy (138)	UDP-N-Acetylglucosamine-1-phosphotransferase	Glycoprotein; glycolipids	Late infantile	AR	Mild mental retardation	None	+ + +	Corneal clouding, mild retinopathy, hyperopic astigmatism		Coarse facies; stiffness of hands and shoulders
Leukodystrophies										
Krabbe's disease (147)	Galactosylceramidase [BMT]	Galactosylceramide Galactosyls—phingosine	Infantile	AR	Mental retardation	None	None	None	None	White matter globoid cells
Metachromatic leukodystrophy (148)	Arylsulfatase A	Cerebroside sulfate	Infantile Juvenile Adult	AR	Mental retardation; dementia; and psychosis in adult	None	None	Optic atrophy	None	Gait abnormalities in late infantile form
Multiple sulfatase deficiency (149)	Active site cysteine to Cα-formylglycine-converting enzyme	Sulfatides; mucopolysaccharides	Late infantile	AR	Mental retardation	+	+ +	Retinal degeneration	Vacuolated and granulated cells	Absent activity of all known cellular sulfatases
Disorders of Neutral Lipids										
Wolman's—disease (142)	Acid lysosomal lipase [BMT]	Cholesteryl esters; triglycerides	Infantile	AR	Mild mental retardation	+ + +	None	None	None	Adrenal calcification

Cholesteryl ester storage disease (142)	Acid lysosomal lipase	Cholesteryl esters	Childhood	AR	None	Hepatomegaly	None	None	None	Fatty liver disease; cirrhosis
Farber's disease (142)	Acid ceramidase	Ceramide	Infantile Juvenile	AR	Occasional mental retardation	+/–	None	Macular degeneration	None	Arthropathy, subcutaneous nodules
Disorders of Glycogen										
Pompe's —disease (135)	Acid α-glucosidase [ET]	Glycogen	Infantile late onset	AR	Neuromuscular	+/–	None	None	None	Myocar—diopathy
Late onset GAA deficiency (135)	Acid α-glucosidase [ET]	Glycogen	Variable–juvenile to adulthood	AR	Neuromuscular	None	None	None	None	Respiratory insufficiency; neuro —muscular disease
Danon's disease	LAMP-2 (lysosomal associated membrane protein-2)	Glycogen	Variable–childhood to adulthood	X-linked (? Dominant)	Cardiomyopathy Neuromuscular Inconsistent mental retardation	None	None	None	None	Myocardial vacuolar degeneration

[a]Numbers in parentheses refer to the chapters in Scriver et al, 8th ed, for comprehensive reviews.

Abbreviations: AR, autosomal recessive; ET, enzyme therapy; SRT, substrate reduction therapy; BMT, bone marrow (or stem cell) transplantation.

PATHOGENESIS OF LYSOSOMAL STORAGE DISEASES

Lysosomal biogenesis involves ongoing synthesis of lysosomal hydrolases, membrane constitutive proteins, and new membranes. Lysosomes originate from the fusion of trans-Golgi network (TGN) vesicles with late endosomes. Progressive vesicular acidification accompanies the maturation of TGN vesicles, and this gradient facilitates the pH-dependent dissociation of receptors and ligands, as well as activating lysosomal hydrolases.

Abnormalities at any biosynthetic step can impair enzyme activation and lead to a lysosomal storage disorder. Following leader sequence clipping, complex oligosaccharide modifications occur during transit through the Golgi, including the lysosomal targeting ligand, mannose-6-phosphate, and modification of high-mannose oligosaccharide chains of many soluble lysosomal hydrolases. Lysosomal integral or associated membrane proteins are sorted to the membrane or interior of the lysosome by several different peptide signals. Phosphorylation, sulfation, additional proteolytic processing, and macromolecular assembly of heteromers occur concurrently. Such posttranslational modifications are critical to enzyme function, and defects can result in multiple enzyme/protein deficiencies.

The final common pathway for lysosomal storage diseases is the accumulation of specific macromolecules within tissues and cells that normally have a high flux of these substrates. The majority of lysosomal enzyme deficiencies result from point mutations or genetic rearrangements at a locus that encodes a single lysosomal hydrolase. However, some mutations cause deficiencies of several different lysosomal hydrolases by altering the enzymes/proteins involved in targeting, active site modifications, or macromolecular association or trafficking. All are inherited as autosomal recessive disorders, except Hunter's (mucopolysaccharidosis type II) and Fabry's diseases, which are X-linked. Substrate accumulation leads to lysosomal distortion, which has significant pathologic consequences. In addition, abnormal amounts of metabolites may also have pharmacologic effects important to disease pathophysiology and propagation.

For many lysosomal diseases, the accumulated substrates are endogenously synthesized within particular tissue sites of pathology. Other diseases have greater exogenous substrate supplies. For example, they are delivered by low-density lipoprotein receptor–mediated uptake in Fabry's and cholesteryl ester storage diseases or by phagocytosis in Gaucher's disease type 1. The threshold hypothesis refers to a level of enzyme activity below which disease develops. Consequently, small changes in enzyme activity near the threshold can lead to or prevent disease. A critical element of this model is that enzymatic activity can be challenged by changes in substrate flux based on genetic background, cell turnover, recycling, or metabolic demands. Thus, a set level of residual enzyme may be adequate for substrate in some tissues or cells, but not in others. In addition, several variants of each lysosomal storage disease exist at a clinical level. These disorders, therefore, represent a continuum of manifestations that are not easily dissociated into discrete entities. The bases for such variations have not been elucidated in any detail.

SELECTED DISORDERS

Tay-Sachs Disease

About 1 in 30 Ashkenazi Jews is a carrier for Tay-Sachs disease, which is caused by total hexosaminidase A (Hex A) deficiency—defective a-chains. The infantile form is a fatal neurodegenerative disease with macrocephaly, loss of motor skills, increased startle reaction, and a macular cherry red spot. The juvenile-onset form presents with ataxia and dementia, with death by age 10–15 years. The adult-onset disorder is characterized by clumsiness in childhood; progressive motor weakness in adolescence; and additional spinocerebellar, lower motor neuron symptoms, and dysarthria in adulthood. Intelligence declines slowly, and psychosis is common. Screening for Tay-Sachs disease carriers is recommended in the Ashkenazi Jewish population. Sandhoff's disease, defective b-chains, is phenotypically similar to Tay-Sachs disease, but hepatosplenomegaly and bony dysplasias are also present.

Fabry's Disease

Fabry's disease is an X-linked disorder that results from mutations in the a-galactosidase gene. The estimated prevalence of hemizygous males is 1/40,000 to 1/3500 in selected populations. Clinically, the disease manifests with angiokeratomas (telangiectatic skin lesions), hypohidrosis, corneal and lenticular opacities, acroparesthesia; and small-vessel disease of the kidney, heart, and brain.

The angiokeratomas and acroparesthesia may appear in childhood and lead to early diagnosis, if suspected. Angiokeratomas are punctate, dark red to blue-black, flat or slightly raised, and usually symmetric; they do not blanch with pressure. They range from barely visible to several millimeters in diameter and have a tendency to increase in size and number with age. They usually are most dense between the umbilicus and knees—the "bathing suit area"—but may occur anywhere, including the mucosal surfaces. Angiokeratomas also occur in Fordyce scrotal angiokeratoma and several other very rare lysosomal storage diseases. Corneal and lenticular lesions, detectable on slit-lamp examination, may help in establishing a diagnosis of Fabry's disease. Debilitating episodic burning pain of the hands, feet, and proximal extremities (acroparesthesia) can last from minutes to days and can be precipitated by changes in temperature, exercise, fatigue, or fever. Abdominal pain can resemble that from appendicitis or renal colic. Proteinuria, isosthenuria, and progressive renal dysfunction occur in the second to fourth decades; about 5% of idiopathic renal failure males have a-galactosidase mutations. Hypertension, left ventricular hypertrophy, anginal chest pain with or without myocardial ischemia or infarction, and congestive heart failure can occur in the third to fourth decades. About 1–3% of patients with idiopathic hypertrophic myocardiopathy were found to have Fabry's disease. Similarly, ~3–5% of idiopathic stroke in males (35–50 years) have a-galactosidase mutations. Leg lymphedema without hypoproteinemia and episodic diarrhea also occur. Death is due to renal failure or cardiovascular or cerebrovascular disease in untreated patients. Variants with residual a-galactosidase activity may have late-onset manifestations limited to the cardiovascular system that resemble hypertrophic cardiomyopathy. Variants with predominant cardiac, renal, or central nervous system (CNS) manifestations are becoming better defined. Up to 70% of heterozygous females may exhibit clinical manifestations, including acroparesthesia and CNS, cardiac, and renal disease.

Phenytoin and carbamazepine diminish the chronic and episodic acroparesthesia. Chronic hemodialysis or kidney transplantation can be lifesaving in patients with renal failure. Enzyme therapy clears stored lipids from a variety of cells, particularly those of the renal, cardiac, and skin vascular endothelium. Renal insufficiency appears irreversible. Early institution of enzyme therapy may prevent or slow the progression of the life-threatening complications.

Gaucher's Disease

Gaucher's disease is an autosomal recessive disorder that results from defective activity of acid b-glucosidase; >400 mutations have been described at the *GBA* locus of such patients. Disease variants are classified based on the absence or presence and progression of neuronopathic involvement.

Gaucher's disease type 1 is a nonneuronopathic disease that can present in childhood to adulthood with slowly to rapidly progressive visceral disease. About 55–60% of such patients are diagnosed at <20 years in white populations and even younger in other groups. This pattern of presentation is distinctly bimodal, with peaks at <10 to 15 years and at ~25 years. Younger patients tend to have a greater degree of hepatosplenomegaly and accompanying blood cytopenias. In contrast, the older group has a greater tendency for chronic bone disease. Hepatosplenomegaly occurs in virtually all symptomatic patients and can be minor or massive. Accompanying anemia and thrombocytopenia are variable and are not directly related to liver or spleen volumes. Severe liver dysfunction is unusual. Splenic infarctions can resemble an acute abdomen. Pulmonary hypertension and alveolar Gaucher's cell accumulation are uncommon, but life threatening, and can occur at any age. *GBA* mutations in the hetero- or homozygous states are a significant risk factor for early onset or more progressive Parkinson's disease.

All patients with Gaucher's disease have nonuniform infiltration of bone marrow by lipid-laden macrophages termed Gaucher's cells. This can lead to marrow packing with subsequent infarction, ischemia, necrosis, and cortical bone destruction. Bone marrow involvement spreads from proximal to distal in the limbs and can involve the axial skeleton extensively, causing vertebral collapse. In addition to bone marrow involvement, bone remodeling is defective, with loss of total bone calcium leading to osteopenia, osteonecrosis, avascular infarction, and vertebral compression fractures and spinal cord involvement. Aseptic necrosis of the femoral head is common, as is fracture of the femoral neck. The mechanism by which diseased bone marrow macrophages interact with osteoclasts and/or osteoblasts to cause bone disease is not well understood. Chronic, ill-defined bone pain can be debilitating and poorly correlated with radiographic findings. "Bone crises" are associated with localized, excruciating pain, and, on occasion, local erythema, fever, and leukocytosis. Some patients have frequent crises, whereas other patients experience only one. These crises represent acute infarctions of bone, as evidenced in nuclear scans by localized absent uptake of pyrophosphate agents. Osteomyelitis should be excluded by appropriate cultures.

Decreased acid b-glucosidase activity (0–20% of normal) in nucleated cells makes the diagnosis. The enzyme is not present in bodily fluids. The sensitivity of enzyme testing is poor for heterozygote detection; molecular testing is preferred when the mutations are known. The disease frequency varies from about 1 in 1000 in Ashkenazi Jews to <1 in 100,000 in other populations. About 1 in 12–15 Ashkenazi Jews carries a Gaucher's disease allele. Four common mutations account for ~85% of the mutations in that population of affected patients: N370S (1226G), 84GG (a G insertion at cDNA position 84), L444P (1448C), and IVS-2 (an intron 2 splice junction mutation).

Genotype/phenotype studies indicate a significant correlation, though not absolute, between disease type and severity and the acid b-glucosidase genotype. The most common mutation in the Ashkenazi Jewish population (N370S) shares a 100% association with nonneuronopathic or type 1 Gaucher's disease. The N370S/N370S and N370S/other mutant allelegenotypes are associated with later onset/less severe disease and with earlier onset/severe disease, respectively. As many as 50–60% of individuals with the N370S/N370S genotype are asymptomatic. Other alleles include L444P (very low activity), 84GG (null), or IVS-2 (null), and rare/private or uncharacterized alleles. The L444P/L444P patients almost always have life-threatening to very severe/early-onset disease, and many, though not all, develop CNS involvement in the first 2 decades of life.

Symptomatic management of the blood cytopenias and joint replacement surgeries continue to have important roles in management. However, regular, intravenous enzyme therapy is currently the treatment of choice in significantly affected patients and is highly efficacious and safe in diminishing the hepatosplenomegaly and improving bone marrow involvement and hematologic findings. The bone disease is decreased but not eliminated by enzyme therapy. Adult patients may benefit from adjunctive treatment with bisphosphonates to improve bone density. Patients who cannot be treated with enzyme, either because it is not effective or because they have an allergy or other hypersensitivities, may be treated with medications that decrease the production of the complex lipid molecules that are broken down by acid b-glucosidase (so-called, substrate reduction therapy). Other approaches to stabilize the abnormal enzyme produced by mutant alleles allowing improved enzyme function (referred to as *pharmacologic chaperone therapy*) are under investigation as alternative treatments for Gaucher's disease, Fabry's disease, and other lysosomal diseases.

Gaucher's disease type 2 is a rare, severe CNS disease that leads to death by 2 years of age. Type 3 Gaucher's disease has highly variable manifestations in the CNS and viscera. It can present in early childhood with rapidly progressive, massive visceral disease and slowly progresses to static CNS involvement; in adolescence with dementia; or in early adulthood with rapidly progressive, uncontrollable myoclonic seizures and mild visceral disease. Visceral disease in type 3 is nearly identical to that in type 1 but is generally more severe. Early CNS findings may be limited to defects in lateral gaze tracking, which may remain static for decades. Mental retardation can be slowly progressive or static. This variant is most frequent among individuals of Swedish descent. The visceral, but not the CNS, involvement responds to enzyme therapy.

Niemann-Pick Disease

This is an autosomal recessive disorder that results from defects in acid sphingomyelinase. Types A and B are distinguished by an early age of onset and progressive CNS disease in type A. Type A typically has onset in the first 6 months, with rapidly progressive CNS deterioration, spasticity, failure to thrive, and massive hepatosplenomegaly. In contrast, type B has a later, more variable onset and progression of hepatosplenomegaly, with eventual development of cirrhosis and hepatic replacement by foam cells. Affected patients develop progressive pulmonary disease with dyspnea, hypoxemia, and a reticular infiltrative pattern on chest x-ray. Foam cells are present in alveoli, lymphatic vessels, and pulmonary arteries. Progressive hepatic or lung disease leads to demise in adolescence to early adulthood.

The diagnosis is established by markedly decreased (1–10% of normal) sphingomyelinase activity in nucleated cells. There is no specific treatment for Niemann-Pick disease. The efficacy of hepatic or bone marrow transplantation has not been clearly established. Clinical trials using enzyme therapy are anticipated in the near future.

Mucopolysaccharidoses

Mucopolysaccharidosis type I (MPS I) is an autosomal recessive disorder caused by deficiency of a-L-iduronidase. The continuum of involvement traditionally has been divided into three categories: (1) Hurler's disease (MPS I H) for severe deficiency with neurodegeneration, (2) Scheie's disease (MPS I S) for later-onset disease without neurologic involvement and with relatively less severe disease in other organ systems, and (3) Hurler-Scheie syndrome (MPS I H/S) for patients intermediate between these extremes. MPS I H/S is characterized by severe somatic disease and usually without overt neurologic deterioration.

MPS I often presents in infancy or early childhood with chronic rhinitis, clouding of the corneas, and hepatosplenomegaly. As the disease progresses, nearly every organ system can be affected. For the more severe forms, cardiac and respiratory diseases become life threatening in childhood. Skeletal disease can be quite severe, resulting in very limited mobility.

There are two current treatments for the MPS I diseases. Stem cell transplantation is the standard treatment for patients presenting in infancy who appear to have or are at risk for neurologic degeneration. Stem cell engraftment results in stabilization of the CNS disease and reverses the hepatosplenomegaly. It also improves the cardiac and respiratory disease. Stem cell transplantation does not eliminate the corneal disease or resolve the progressive skeletal disease. Intravenous enzyme therapy effectively addresses the hepatosplenomegaly and improves cardiac and respiratory disease. The enzyme does not effectively penetrate the CNS and does not directly alter the CNS disease. Enzyme therapy and stem cell transplant appear to have similar effects for signs and symptoms outside the CNS. Enzyme therapy has a lower risk for life-threatening complications and may, therefore, have advantages for patients without CNS degeneration. A combination of enzyme therapy and stem cell transplantation has been used with enzyme therapy started prior to transplantation in an attempt to decrease the disease burden prior to transplantation. The experience with this approach is not well documented, but it appears to have advantages over stem cell therapy alone.

Enzyme therapy for Maroteaux-Lamy disease (MPS VI) has received U.S. Food and Drug Administration (FDA) approval for treatment. This very rare autosomal recessive disorder is characterized by hepatosplenomegaly, bone disease, heart disease, and respiratory compromise similar to those seen in MPS I, but it is due to deficiency of arylsulfatase B and is not associated with neurologic degeneration.

Hunter's disease (MPS II) is an X-linked disorder due to deficiency in iduronate sulfate sulfatase; it has manifestations similar to MPS I, including neurologic degeneration, but there is no corneal clouding or other eye disease. Like MPS I, MPS II is clinically variable with CNS and non-CNS variants. Stem cell transplantation has not been successful in treating the CNS disease associated with MPS II. The FDA and the European Agency for the Evaluation of Medicinal Products (EMEA) have approved enzyme replacement for treatment of the visceral manifestations of MPS II.

Pompe's Disease

Acid maltase (acid a-glucosidase's GAA) deficiency, also called Pompe's disease, is the only glycogen lysosomal storage disease. The classic severe infantile form presents with hypotonia, myocardiopathy, and hepatosplenomegaly. This variant is rapidly progressive and generally results in death in the first year of life. However, as with other lysosomal diseases, there are early- and late-onset forms of this disorder. The late-onset variants may be as common as 1/40,000. The late-onset patients typically present with a slowly progressive myopathy that may resemble a limb-girdle muscular dystrophy. Respiratory insufficiency may be the presenting sign or it may develop with advancing disease. In late stages of the disease, patients may require mechanical ventilation, report swallowing difficulties, and experience loss of bowel and bladder control. Myocardiopathy is not usually seen in late-onset variants of Pompe's disease.

The FDA has approved enzyme infusion therapy for Pompe's disease. This is clearly life-prolonging for the infantile form, consistently resulting in improved cardiac function. Improved respiratory function is seen in most treated infants. Some infants demonstrated marked improvement in motor functions while others had minor changes in muscle tone or strength. Prevention of deterioration has been shown with GAA enzyme therapy in the late-onset forms. Early intervention with GAA enzyme therapy in such patients may limit or prevent the deterioration, but very advanced disease will have significant irreversible components.

FURTHER READINGS

Grabowski GA et al: Dose-response relationships for enzyme replacement therapy with imiglucerase/alglucerase in patients with Gaucher disease type 1. *Genet Med* 11:92, 2009[PMID: 19265748] [Full Text]

Martins AM et al: Guidelines for the management of mucopolysaccharidosis type I. *J Pediatr* 155:S32, 2009

Muenzer J et al: Multidisciplinary management of Hunter syndrome. *Pediatrics* 124:e1228, 2009

Scriver CR et al (eds): The Metabolic and Molecular Bases of Inherited Disease, 8th ed. New York, McGraw-Hill, 2001; ommbid.com, 2006

van der Ploeg AT, Reuser AJ: Pompe's disease. *Lancet* 372:1342, 2008

Weinreb NJ et al:. 2009. A validated disease severity scoring system for adults with type 1 Gaucher disease. *Genet Med* 12:44, 2010[PMID: 20027115] [Full Text]

Zarate YA, Hopkin, RJ: Fabry's disease. *Lancet* 372:1427, 2008[PMID: 18940466] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GLYCOGEN STORAGE DISEASES AND OTHER INHERITED DISORDERS OF CARBOHYDRATE METABOLISM: INTRODUCTION

Carbohydrate metabolism plays a vital role in cellular function by providing the energy required for most metabolic processes. The relevant biochemical pathways involved in the metabolism of these carbohydrates are shown in **Fig. 362-1**. Glucose is the principle substrate of energy metabolism in humans. Metabolism of glucose generates ATP through glycolysis and mitochondrial oxidative phosphorylation. The body obtains glucose through the ingestion of polysaccharides, primarily starch, and disaccharides including lactose, maltose, and sucrose. Galactose and fructose are two other monosaccharides that serve as sources of fuel for cellular metabolism; however, their role as fuel sources is much less significant than that of glucose. Galactose is derived from lactose (galactose + glucose), which is found in milk products, and is an important component for certain glycolipids, glycoproteins, and glycosaminoglycans. Fructose is found in fruits, vegetables, and honey. Sucrose (fructose + glucose) is another dietary source of fructose and is a commonly used sweetener.

Figure 362-1

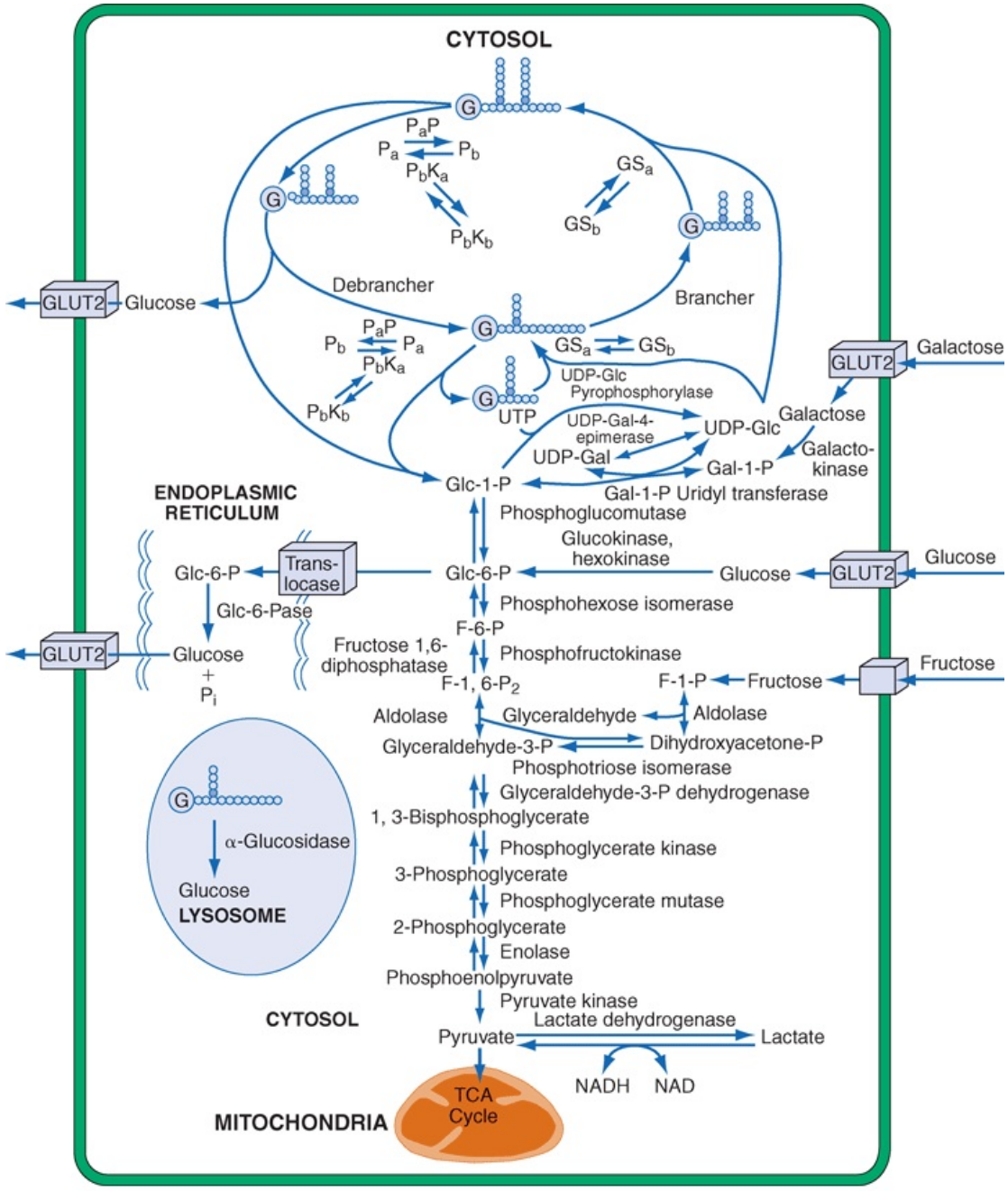

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Metabolic pathways related to glycogen storage diseases and galactose and fructose disorders. Nonstandard abbreviations are as follows: GS_a, active glycogen synthase; GS_b, inactive glycogen synthase; P_a, active phosphorylase; P_b, inactive phosphorylase; P_aP, phosphorylase *a* phosphatase; P_bK_a, active phosphorylase *b* kinase; P_bK_b, inactive phosphorylase *b* kinase; G, glycogenin, the primer protein for glycogen synthesis. [*Modified from AR Beaudet, in KJ Isselbacher et al (eds): Harrison's Principles of Internal Medicine, 13th ed., New York, McGraw-Hill, 1994, p 1855.*]

Glycogen, the storage form of glucose in animal cells, is composed of glucose residues joined in straight chains by α1-4 linkages and branched at intervals of 4–10 residues by α1-6 linkages. Glycogen forms a treelike molecule and can have a molecular weight of many millions. Glycogen may aggregate to form structures recognizable by electron microscopy. With the exception of glycogen storage disease (GSD) type 0, defects in glycogen metabolism typically cause an accumulation of glycogen in the tissues, hence the name *glycogen storage diseases*. The stored glycogen could be normal or abnormal in structure in the various disorders. The defects in gluconeogenesis or glycolytic pathways, including galactose and fructose metabolism, do not usually result in glycogen accumulation.

Clinical manifestations of the various disorders of carbohydrate metabolism differ markedly. The symptoms range from harmless to lethal. Unlike disorders of lipid metabolism, mucopolysaccharidoses, or other storage diseases, dietary therapy has been effective in many of the carbohydrate disorders. All of the genes responsible for the inherited defects of carbohydrate metabolism have been cloned,

and mutations have been identified. Advances in our understanding of the molecular basis of these diseases are being used to improve diagnosis and management. Some of these disorders are candidates for enzyme replacement, substrate reduction, and early trials of gene therapy.

Historically, the glycogen storage diseases were categorized numerically in the order in which the enzymatic defects were identified. They are also classified by the organs involved and clinical manifestations. This is the system followed in this chapter **(Table 362-1)**. The overall frequency of all forms of glycogen storage disease is approximately 1 in 20,000 live births. Most are inherited as autosomal recessive traits; however, phosphoglycerate kinase deficiency and one form of phosphorylase kinase deficiency are X-linked disorders. The most common childhood disorders are glucose-6-phosphatase deficiency (type I), lysosomal acid α-glucosidase deficiency (type II), debrancher deficiency (type III), and liver phosphorylase kinase deficiency (type IX). The most common adult disorder is myophosphorylase deficiency (type V, or McArdle disease).

Table 362-1 Features of Glycogen Storage Diseases and Galactose and Fructose Disorders

Type/Common Name	Basic Defect	Clinical Features	Comments
Liver Glycogenoses			
Disorders with hepatomegaly and hypoglycemia			
Ia/von Gierke	Glucose-6-phosphatase	Growth retardation, enlarged liver and kidney, hypoglycemia, elevated blood lactate, cholesterol, triglycerides, and uric acid	Common, severe hypoglycemia Complications in adulthood include hepatic adenomas, hepatic carcinoma, renal failure
Ib	Glucose-6-phosphate translocase	As for Ia, with additional findings of neutropenia and neutrophil dysfunction	~10% of type I
IIIa/Cori or Forbes	Liver and muscle debranching enzyme	Childhood: Hepatomegaly, growth retardation, muscle weakness, hypoglycemia, hyperlipidemia, elevated liver transaminases; liver symptoms improve with age Adulthood: muscle atrophy and weakness; onset: third to fourth decades; variable cardiomyopathy	Common, intermediate severity of hypoglycemia; hepatic adenomas, liver cirrhosis, and hepatic carcinoma can occur
IIIb	Liver debranching enzyme (normal muscle debrancher activity)	Liver symptoms same as in type IIIa; no muscle symptoms	~15% of type III
VI/Hers	Liver phosphorylase	Hepatomegaly, variable hypoglycemia, hyperlipidemia and ketosis; symptoms may improve with age	Rare, often a "benign" —glycogenosis, severe cases being recognized
IX/phosphorylase kinase deficiency	Liver phosphorylase kinase a subunit	As for VI	Common, X-linked, typically less severe than autosomal forms; clinical variability within and between subtypes; severe cases being recognized
0/glycogen synthase deficiency	Glycogen synthase	Fasting hypoglycemia and ketosis, elevated lactic acid and hyperglycemia after glucose load	Decreased glycogen stores
XI/Fanconi-Bickel	Glucose transporter-2	Failure to thrive, rickets, hepatomegaly, proximal renal tubular dysfunction, impaired glucose and galactose utilization	Rare, consanguinity in 70%
Disorders with liver cirrhosis			
IV/Andersen	Branching enzyme	Failure to thrive, hypotonia, hepatomegaly, splenomegaly, progressive liver cirrhosis and failure (death usually before fifth year); some without progression	One of the rarer glycogenoses; other neuromuscular variants exist
Muscle Glycogenoses			

Disorders with muscle-energy impairment V/McArdle	Muscle phosphorylase	Exercise intolerance, muscle cramps, myoglobinuria on strenuous exercise, increased CK	Common, male predominance
VII/Tarui	Phosphofructokinase—M subunit	As for type V, with additional findings of a compensated hemolysis	Prevalent in Ashkenazi Jews and Japanese
Phosphoglycerate kinase deficiency	Phosphoglycerate kinase	As for type V, with additional findings of a hemolytic anemia and CNS dysfunction	Rare, X-linked
Phosphoglycerate mutase deficiency	Phosphoglycerate mutase—M subunit	As for type V	Rare; most patients are African American
Lactate dehydrogenase deficiency	Lactic acid dehydrogenase—M subunit	As for type V, with additional findings of erythematous skin eruption and uterine stiffness resulting in childbirth difficulty in female	Rare
Fructose 1,6-bisphosphate aldolase A deficiency	Fructose 1,6-bisphosphate aldolase A	As for type V, with additional finding of hemolytic anemia	Rare
Pyruvate kinase deficiency	Pyruvate kinase—muscle isozyme	Muscle cramps and/or fixed muscle weakness	Rare
Muscle phosphorylase kinase deficiency	Muscle-specific phosphorylase kinase	As for type V, some patients may have muscle weakness and atrophy	Rare, autosomal recessive
b-enolase deficiency	Muscle b-enolase	Exercise intolerance	Rare
Disorders with progressive skeletal myopathy and/or cardiomyopathy			
II/Pompe	Lysosomal acid a-glucosidase	Infantile: hypotonia, muscle weakness, cardiac enlargement and failure, fatal early; Juvenile and adult: progressive skeletal muscle weakness and atrophy, proximal muscle and respiratory muscle are seriously affected	Common, undetectable, or very low level of enzyme activity in infantile form; residual enzyme activity in late-onset
Cardiac phosphorylase kinase deficiency	Cardiac-specific phosphorylase kinase	Severe cardiomyopathy and early heart failure	Very rare
Galactose Disorders			
Galactosemia with uridyl transferase deficiency	Galactose 1-phosphate uridyl transferase	Vomiting, hepatomegaly, jaundice, cataracts, amino aciduria, failure to thrive	Long-term complications exist despite early diagnosis and treatment
Galactokinase deficiency	Galactokinase	Cataracts	Benign
Uridine diphosphate galactose 4-epimerase deficiency	Uridine diphosphate galactose 4-epimerase	Similar to transferase deficiency with additional findings of hypotonia and nerve deafness	Benign variant exists
Fructose Disorders			
Essential fructosuria	Fructokinase	Asymptomatic, positive urine reducing substance	Benign
Hereditary fructose intolerance	Fructose 1-6 bisphosphate aldolase B	Vomiting, lethargy, failure to thrive, hepatic failure	Prognosis good with early diagnosis and fructose restriction
Fructose 1,6-diphosphatase deficiency	Fructose 1,6-diphosphatase	Episodic hypoglycemia and lactic acidosis	Avoid fasting, good prognosis

Abbreviations: CK, creatine kinase; CNS, central nervous system; M, muscle.

SELECTED LIVER GLYCOGENOSES

Disorders with Hepatomegaly and Hypoglycemia

TYPE I GLYCOGEN STORAGE DISEASE (GLUCOSE-6-PHOSPHATASE OR TRANSLOCASE DEFICIENCY, VON GIERKE'S DISEASE)

Type I glycogen storage disease is an autosomal recessive disorder caused by glucose-6-phosphatase deficiency in liver, kidney, and intestinal mucosa. There are two subtypes of GSD I: type Ia, in which the glucose-6-phosphatase enzyme is defective, and type Ib, in which the translocase that transports glucose-6-phosphate across the microsomal membrane is defective. The defects in both subtypes lead to inadequate conversion in the liver of glucose-6-phosphate to glucose and thus make affected individuals susceptible to fasting hypoglycemia.

Clinical and Laboratory Findings

Persons with type I disease may develop hypoglycemia and lactic acidosis during the neonatal period; however, more commonly, they exhibit hepatomegaly at 3–4 months of age. Hypoglycemia and lactic acidosis can develop after a short fast. These children usually have doll-like faces with fat cheeks, relatively thin extremities, short stature, and a protuberant abdomen that is due to massive hepatomegaly. The kidneys are enlarged, but the spleen and heart are of normal size. The hepatocytes are distended by glycogen and fat with large and prominent lipid vacuoles. Despite hepatomegaly, liver enzymes are usually normal or near normal. Easy bruising and epistaxis are associated with a prolonged bleeding time as a result of impaired platelet aggregation/adhesion. Hyperuricemia is present. Hyperlipidemia includes elevation of triglycerides, cholesterol, and phospholipids. Type Ib patients have additional findings of neutropenia and impaired neutrophil function, resulting in recurrent bacterial infections and oral and intestinal mucosal ulceration. GSD I patients may experience intermittent diarrhea, which can worsen with age. In GSD Ib, it is largely due to loss of mucosal barrier function caused by inflammation.

Long-Term Complications

Gout usually becomes symptomatic at puberty as a result of the long-term hyperuricemia. Puberty is often delayed. Nearly all females have ultrasound findings consistent with polycystic ovaries; however, the other clinical features of polycystic ovary syndrome, such as acne and hirsutism, are not seen. Several reports of successful pregnancy in women with GSD I suggest fertility is not affected. Increased bleeding during menstrual cycles, including life-threatening menorrhagia, has been reported. Secondary to the lipid abnormalities, there is an increased risk of pancreatitis. Patients with GSD I may be at an increased risk for cardiovascular disease. Frequent fractures and radiographic evidence of osteopenia/osteoporosis can be observed in adult patients, and radial bone mineral content is significantly reduced in prepubertal patients. Pulmonary hypertension, although rare, has been reported. By the second or third decade, most patients with type I glycogen storage disease develop hepatic adenomas that can hemorrhage and, in some cases, become malignant. Renal disease is a serious late complication. Almost all patients older than 20 years have proteinuria, and many have hypertension, kidney stones, nephrocalcinosis, and altered creatinine clearance. In some patients, renal function deteriorates and progresses to complete failure, requiring dialysis or transplantation.

Diagnosis

Clinical presentation and abnormal plasma lactate and lipid values suggest that a patient may have GSD I, but gene-based mutation analysis provides a noninvasive means of reaching a definitive diagnosis for most patients with types Ia and Ib disease. Before the glucose-6-phosphatase and glucose-6-phosphate translocase genes were cloned, a definitive diagnosis required a liver biopsy to demonstrate a deficiency.

Treatment: Type I Glycogen Storage Disease

Treatment involves the maintenance of normal blood glucose levels through continuous infusion of glucose via feeding tube or oral administration of uncooked cornstarch. Uncooked cornstarch serves as a slow-release form of glucose and can be given at a dose of 1.6 g/kg every 4 h for infants younger than 2 years. As the child grows older, the cornstarch regimen can be changed to every 6 h at a dose of 1.75–2.5 g/kg of body weight. Newer starch products may be longer acting, better tolerated, and more palatable. Since fructose and galactose cannot be converted to free glucose, their dietary intake should be restricted. Dietary supplements of multivitamins, calcium, and Vitamin D are often required. Allopurinol is given to lower uric acid levels, if levels remain elevated despite diet and metabolic control. The hyperlipidemia can be reduced with dietary control, use of medium chain triglycerides, fish oil and lipid-lowering drugs such as statins and fibric acids. Angiotensin-converting enzyme (ACE) inhibitors are beneficial for treating microalbuminuria, an early indicator of renal dysfunction in these patients. Citrate supplementation may be beneficial in preventing or ameliorating nephrocalcinosis and development of urinary calculi. Orthotropic liver transplantation is a treatment usually reserved for GSD I patients with liver malignancy, multiple liver adenomas, metabolic derangements refractory to medical management, and/or liver failure.

TYPE III GLYCOGEN STORAGE DISEASE (DEBRANCHER DEFICIENCY, LIMIT DEXTRINOSIS)

Type III GSD is an autosomal recessive disorder caused by a deficiency of glycogen debranching enzyme. Debranching and phosphorylase enzyme are responsible for complete degradation of glycogen. When the debranching enzyme is defective, glycogen breakdown is incomplete. Abnormal

glycogen accumulates with short outer chains and resembles dextrin.

Clinical and Laboratory Findings

Deficiency of glycogen debranching enzyme causes hepatomegaly, hypoglycemia, short stature, variable skeletal myopathy, and cardiomyopathy. The disorder usually involves both liver and muscle and is termed *type IIIa glycogen storage disease*. However, in about 15% of patients, the disease appears to involve only the liver and is classified as *type IIIb*. Hypoglycemia, hyperlipidemia, and elevated liver transaminases occur in children. In contrast to type I disease, fasting ketosis can be prominent, transaminases are elevated, and blood lactate and uric acid concentrations are usually normal. Serum creatine kinase (CK) levels can sometimes be used to identify patients with muscle involvement, but normal levels do not rule out muscle enzyme deficiency. In most patients with type III disease, hepatomegaly improves with age; however, liver cirrhosis and hepatocellular carcinoma may occur in late adulthood. Hepatic adenomas may occur although less common than in GSOI. Left ventricular hypertrophy and life-threatening arrhythmias have been reported. Those with type IIIa may experience muscle weakness in childhood that can become severe after the third or fourth decade of life. Polycystic ovaries are common in GSD III, and some patients develop features of polycystic ovarian syndrome, such as hirsutism and irregular menstrual cycles. Reports of successful pregnancy in women with GSD III suggest fertility is normal.

Diagnosis

In type IIIa glycogen storage disease, deficient debranching enzyme activity can be demonstrated in liver, skeletal muscle, and heart. In contrast, patients with type IIIb have debranching enzyme deficiency in the liver but not in muscle. In the past, definitive assignment of subtype required enzyme assays in both liver and muscle. DNA-based analyses now provide a noninvasive way of subtyping these disorders in most patients. However, the large size of the gene and the distribution of private mutations across it pose challenges in DNA based analysis.

Treatment: Type III Glycogen Storage Disease

Dietary management of type III disease is less demanding than that of type I. Frequent high-carbohydrate meals with cornstarch supplements or nocturnal gastric drip feedings are usually effective in treating hypoglycemia. A high-protein diet is recommended as neoglucogenesis is intact in GSD III thus providing a source for glucose.

TYPE IX GLYCOGEN STORAGE DISEASE (LIVER PHOSPHORYLASE KINASE DEFICIENCY)

Defects of phosphorylase kinase cause a heterogeneous group of glycogenoses. The phosphorylase kinase enzyme complex consists of four subunits (α, β, γ, and δ). Each is encoded by different genes (X chromosome as well as autosomes) that are differentially expressed in various tissues. Phosphorylase kinase deficiency can be divided into several subtypes on the basis of the gene/subunit involved, the tissues primarily affected, and the mode of inheritance. The most common subtype is X-linked liver phosphorylase kinase deficiency, which is one of the most common liver glycogenoses. Phosphorylase kinase activity may also be deficient in erythrocytes and leukocytes but is normal in muscle. Typically, a child between the ages of 1 and 5 years presents with growth retardation and hepatomegaly. Levels of cholesterol, triglycerides, and liver enzymes are mildly elevated. Fasting ketosis is another feature of the disease. Lactic and uric acid levels are usually normal. Hypoglycemia is typically mild, in the X-linked form of the disease. Phenotypic variability is being increasingly recognized in the x-linked form. The accumulated glycogen in liver (β particles, rosette form) has a frayed or burst appearance and is less compact than the glycogen seen in type I or type III GSD. Hepatomegaly and abnormal blood chemistries gradually return to normal with age. Most adults reach a normal final height and are practically asymptomatic, despite a persistent phosphorylase kinase deficiency. Prognosis is usually good, and adult patients have normal stature and minimal hepatomegaly. "With further study, long term issues will be better understood."

Treatment is symptomatic. A high-carbohydrate diet and frequent feedings are effective in preventing hypoglycemia. Some patients require no specific treatment.

Other subtypes of type IX include an autosomal recessive form of liver and muscle phosphorylase kinase deficiency, an autosomal recessive form of liver phosphorylase kinase deficiency that often develops into liver cirrhosis, a muscle-specific phosphorylase kinase deficiency that causes cramps and myoglobinuria with exercise, and a cardiac-specific phosphorylase kinase deficiency that is lethal during infancy because of massive glycogen deposition in the myocardium. The finding of PK deficiency maybe a secondary phenomena as a subset of these patients have been found to have mutations in PRKAG2.

OTHER LIVER GLYCOGENOSES WITH HEPATOMEGALY AND HYPOGLYCEMIA

These disorders include glycogen synthase deficiency (type 0) and hepatic glycogenosis with renal Fanconi syndrome (type XI). The latter is caused by defects in the facilitative glucose transporter 2 (GLUT-2), which transports glucose in and out of hepatocytes, pancreatic cells, and the basolateral membranes of intestinal and renal epithelial cells. The disease is characterized by proximal renal tubular dysfunction, impaired glucose and galactose utilization, and accumulation of glycogen in liver

and kidney.

SELECTED MUSCLE GLYCOGENOSES

Disorders with Muscle-Energy Impairment

TYPE V GLYCOGEN STORAGE DISEASE (MUSCLE PHOSPHORYLASE DEFICIENCY, MCARDLE DISEASE)

Type V glycogen storage disease is an autosomal recessive disorder caused by deficiency of muscle phosphorylase. McArdle disease is a prototypical muscle energy disorder as the enzyme deficiency limits ATP generation by glycogenolysis and results in glycogen accumulation.

Clinical and Laboratory Findings

Symptoms usually first develop in adulthood and involve exercise intolerance with muscle cramps. Two types of activity tend to cause symptoms: (1) brief exercise of great intensity, such as sprinting or carrying heavy loads; and (2) less-intense but sustained activity, such as climbing stairs or walking uphill. Most patients can engage in moderate exercise, such as walking on level ground, for long periods. Although most patients experience episodic muscle pain and cramping as a result of exercise, 35% report permanent pain that seriously affects sleep and other activities. About half of patients report burgundy-colored urine after exercise, resulting from myoglobinuria secondary to rhabdomyolysis. Intense myoglobinuria after vigorous exercise can lead to renal failure. Clinical heterogeneity is uncommon, however there are cases with symptom onset as late as the eighth decade and cases that present early with hypotonia, generalized muscle weakness, and progressive respiratory insufficiency, which is often fatal.

In rare cases, electromyography (EMG) findings may suggest an inflammatory myopathy, a diagnosis that may be confused with polymyositis. These patients may be at risk for a statin-induced myopathy and rhabdomyolysis.

During rest, the serum CK level is usually elevated; after exercise, the CK level increases even more. Exercise also increases the levels of blood ammonia, inosine, hypoxanthine, and uric acid. The latter abnormalities are caused by accelerated recycling of muscle purine nucleotides in the face of insufficient ATP production.

Diagnosis

Lack of an increase in blood lactate and exaggerated blood ammonia elevations after an ischemic exercise test are indicative of a muscle glycogenosis and suggest a defect in the conversion of glycogen or glucose to lactate. The abnormal exercise response, however, can also occur with other defects in glycogenolysis or glycolysis, such as deficiencies of muscle phosphofructokinase or debranching enzyme (when the test is done after fasting). Definitive diagnosis is made by enzymatic assay in muscle tissue or by mutation analysis of the myophosphorylase gene.

Treatment: Type V Glycogen Storage Disease

In general, avoidance of strenuous exercise can prevent major episodes of rhabdomyolysis; however, regular and moderate exercise is recommended to improve exercise capacity. Sucrose administered before exercise can markedly improve tolerance in these patients. A high-protein diet may increase exercise endurance. Creatine and vitamin B6 supplementation have also been shown to improve muscle function in some patients. In general, longevity does not appear to be affected.

Disorders with Progressive Skeletal Muscle Myopathy and/or Cardiomyopathy

POMPE DISEASE, TYPE II GLYCOGEN STORAGE DISEASE (ACID A-1,4 GLUCOSIDASE DEFICIENCY)

Pompe disease is an autosomal recessive disorder caused by a deficiency of lysosomal acid α-1,4 glucosidase (acid maltase), an enzyme responsible for the degradation of glycogen in the lysosomes. It is characterized by the accumulation of glycogen in the lysosomes as opposed to accumulation in cytoplasm as in the other glycogenoses.

Clinical and Laboratory Findings

The disorder encompasses a range of phenotypes. Each includes myopathy but differs in the age of onset, organ involvement, and clinical severity. The most severe is the infantile form with cardiomegaly, hypotonia, and death before age 1. Infants may appear normal at birth but soon develop generalized muscle weakness with feeding difficulties, macroglossia, hepatomegaly, and congestive heart failure due to a hypertrophic cardiomyopathy.

The late onset form (juvenile, or late-childhood, or adult form) is characterized by skeletal muscle manifestations, usually without or with minimal cardiac involvement, and a more slowly progressive course. The juvenile form typically presents as delayed motor milestones (if age of onset is early enough) and difficulty in walking. With disease progression, patients often develop swallowing difficulties, proximal muscle weakness, and respiratory muscle involvement. Death may occur before the end of the second decade.

The adult form presents as a slowly progressive myopathy without overt cardiac involvement and typically has its onset between the second and seventh decades. The clinical picture is dominated by slowly progressive proximal muscle weakness with truncal involvement. The pelvic girdle, paraspinal muscles, and diaphragm are most seriously affected. Respiratory symptoms include somnolence, morning headache, orthopnea, and exertional dyspnea. In rare instances patients present with respiratory insufficiency as the initial symptom.

Laboratory findings include elevated levels of serum CK, aspartate transaminase, and lactate dehydrogenase. In infants, chest x-ray shows massive cardiomegaly and electrocardiographic findings include a high-voltage QRS complex and a shortened PR interval. Muscle biopsy shows the presence of vacuoles that stain positively for glycogen; muscle acid phosphatase is increased, presumably from a compensatory increase of lysosomal enzymes. EMG reveals myopathic features with irritability of muscle fibers and pseudomyotonic discharges. Serum CK is not always elevated in adults, and depending on the muscle biopsied or tested, muscle histology or EMG may not be abnormal. The affected muscle should be examined.

Diagnosis

The confirmatory step for a diagnosis of Pompe disease is enzyme assay demonstrating deficient acid α-glucosidase or gene sequence showing 2 pathogenic mutations in the GAA gene. Enzyme activity can be measured in muscle, cultured skin fibroblasts, or blood. Deficiency is usually more severe in the infantile form than in the juvenile and adult disorders.

Treatment: Type II Glycogen Storage Disease

Nocturnal ventilatory support in late-onset patients improves the quality of life and is beneficial during a period of respiratory decompensation. A definitive therapy is now available using aglucosidase alfa (Myozyme), a recombinant human acid α-glucosidase at 20 mg/kg body weight every 2 weeks as an intravenous infusion. In the clinical trials, aglucosidase alfa has been shown to improve ventilator-free survival and motor function in patients with infantile-onset Pompe disease as compared to untreated controls. A randomized placebo-controlled trial in adults demonstrated stabilization of musculoskeletal outcomes and respiratory parameters.

Glycogen Storage Disease Mimicking Hypertrophic Cardiomyopathy

Deficiencies of lysosomal-associated membrane protein 2 (LAMP2, also called Danon's disease) and protein kinase, adenosine monophosphate (AMP)-activated gamma 2 noncatalytic subunit (PRKAG2) result in the accumulation of glycogen in the heart and skeletal muscle. Clinically, these patients present primarily as hypertrophic cardiomyopathy. Their electrophysiologic abnormalities, particularly ventricular preexcitation and conduction defects, can distinguish them from patients with hypertrophic cardiomyopathy resulting from defects in sarcomere-protein genes. In patients with LAMP2 deficiency, the onset of cardiac symptoms, including chest pain, palpitation, syncope, and cardiac arrest, can occur between the ages of 8 and 15 years, which is younger than the average age of 33 years for patients with PRKAG2 deficiency. The prognosis for LAMP2 deficiency is poor, with progressive end-stage heart failure early in adulthood. By contrast, long-term survival is possible for patients with cardiomyopathy due to PRKAG2 mutations, although some patients may require the implantation of a pacemaker and aggressive control of arrhythmias. There is a congenital form of PRKAG2 that presents in early infancy with severe hypertrophic cardiomyopathy and a rapidly fatal course. In these patients, levels of phosphorylase kinase have been found to be low.

SELECTED DISORDERS OF GALACTOSE METABOLISM

"Classic" *galactosemia* is caused by galactose 1-phosphate uridyl transferase deficiency. It is a serious disease with an incidence of 1 in 60,000 and an early onset of symptoms. The newborn infant normally receives up to 20% of caloric intake as lactose, which consists of glucose and galactose. Without the transferase, the infant is unable to metabolize galactose 1-phosphate (Fig. 362-1), which accumulates, resulting in injury to parenchymal cells of the kidney, liver, and brain. Patients with galactosemia are at increased risk for *Escherichia coli* neonatal sepsis; the onset of sepsis often precedes the diagnosis of galactosemia.

Widespread newborn screening for galactosemia has identified these infants early and allowed them to be placed on dietary restriction. Elimination of galactose from the diet reverses growth failure and renal and hepatic dysfunction, improving the prognosis. However, on long-term follow-up, some patients still have ovarian failure manifest as primary or secondary amenorrhea, as well as developmental delays and learning disabilities, which increase in severity with age. Eighty to over ninety percent of women with classic galactosemia report hypergonadotropic hypogonadism. While most female patients are infertile when they reach childbearing age, a small number have given birth. There are several mutations that appear protective, particularly the S135L mutation, which is more common in the African American population. In addition, most patients have speech disorders, and a smaller number demonstrate poor growth and impaired motor function and balance (with or without overt ataxia). The treatment of galactosemia to prevent long-term complications remains a challenge.

Deficiency of *galactokinase* (Fig. 362-1) causes cataracts. Deficiency of *uridine diphosphate*

galactose 4-epimerase can be benign when the enzyme deficiency is limited to blood cells but can be as severe as classic galactosemia when the enzyme deficiency is generalized.

SELECTED DISORDERS OF FRUCTOSE METABOLISM

Fructokinase deficiency (Fig. 362-1) causes a benign condition that is usually an incidental finding made through the detection of fructose as a reducing substance in the urine.

Deficiency of *fructose 1-6 bisphosphate aldolase* (aldolase B, hereditary fructose intolerance) is a serious disease in infants. These patients are healthy and asymptomatic until fructose or sucrose (table sugar) is ingested (usually from fruit, sweetened cereal, or sucrose-containing formula). Clinical manifestations may include jaundice, hepatomegaly, vomiting, lethargy, irritability, and convulsions. There is a higher incidence of celiac disease in patients with hereditary fructose intolerance (>10%) than in the general population (1–3%). Laboratory findings show a prolonged clotting time, hypoalbuminemia, elevation of bilirubin and transaminases, and proximal renal tubular dysfunction. If the disease is not diagnosed and intake of the noxious sugar continues, hypoglycemic episodes recur, and liver and kidney failure progresses, eventually leading to death. Treatment requires the elimination of all sources of sucrose, fructose, and sorbitol from the diet. Through this treatment, liver and kidney dysfunction improve, and catch-up growth is common; intellectual development is usually unimpaired. Over time, the patient's symptoms become milder, even after fructose ingestion, and the long-term prognosis is good.

Fructose 1,6-diphosphatase deficiency is characterized by childhood life-threatening episodes of acidosis, hypoglycemia, hyperventilation, convulsions, and coma. These episodes are triggered by febrile infections and gastroenteritis when oral food intake decreases. Laboratory findings show low blood glucose, high lactate and uric acid levels, and metabolic acidosis. Unlike hereditary fructose intolerance, there is usually no aversion to sweets, and renal tubular and liver functions are normal. Treatment of acute attacks requires the correction of hypoglycemia and acidosis by intravenous infusion. Later, avoidance of fasting and elimination of fructose and sucrose from the diet prevent further episodes. A slowly released carbohydrate such as cornstarch is useful for the long-term prevention of hypoglycemia. Prognosis is good as patients who survive childhood develop normally.

GLOBAL CONSIDERATIONS

The glycogen storage diseases and other inherited disorders of carbohydrate metabolism although rare, have been reported in most ethnic populations. The prevalent genetic mutations for each disease may vary in different ethnic populations but clinical symptoms are remarkably similar and treatment guidelines apply to all populations.

FURTHER READINGS

Arad M et al: Glycogen storage diseases presenting as hypertrophic cardiomyopathy. N Engl J Med 352:362, 2005[PMID: 15673802] [Full Text]

Beauchamp NJ et al: Glycogen storage disease type IX: High variability in clinical phenotype. Mol Genet Metab 92: 88, 2007[PMID: 17689125] [Full Text]

Kishnani PS et al: Glycogen storage diseases, in *The Online Metabolic and Molecular Bases of Inherited Disease,* CR Scriver et al (eds). New York, McGraw-Hill, 2009, pp 1–85

———: Recombinant human acid α-glucosidase: Major clinical benefits in patients with infantile-onset Pompe disease. Neurology 68:99, 2007

Koerberl DD et al: Glycogen storage disease types I and II: Treatment updates. J Inherit Metab Dis 30:159, 2007

———: Emerging therapies for glycogen storage disease type 1. Trends Endocrinol Metab 20: 252, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HERITABLE DISORDERS OF CONNECTIVE TISSUE: INTRODUCTION

Some of the most common heritable disorders involve the major connective tissues of the body such as bone, skin, cartilage, blood vessels, and basement membranes. Identification of the causes of these disorders has underscored the important structural role of connective tissue proteins, such as the collagens, fibrillin, and elastin. However, these studies have also uncovered unanticipated defects in other proteins and enzymes involved in cell signaling and protein processing. The literature on connective tissue disorders is vast. Consequently, this chapter will focus on the classification and pathophysiology of these disorders and summarize the clinical features and management of the more common disorders.

CLASSIFICATION OF CONNECTIVE TISSUE DISORDERS

The original classification of connective tissue diseases was based on the pattern of inheritance, the cluster of signs and symptoms, as well as radiologic and histologic features. Identification of the mutations causing the diseases has provided a rational framework for understanding the multiple manifestations of the diseases and has led to revisions in the classifications. Also, it has provided tests for prenatal diagnosis and important information for genetic counseling. The usefulness of these tests is likely to increase with the capability to analyze the complete genomes of patients. At the same time, the identification of mutations has thus far provided few new therapies. Also, as with many genetic diseases, the mutations are not always reliable predictors of clinical outcomes. For example, identical collagen I mutations in patients with osteogenesis imperfecta (OI) can be associated with clinical mild or lethal phenotypes for reasons that are not apparent. Also, the current classifications tend to overemphasize the etiologic differences between genetic diseases that are apparent in infants versus similar diseases that appear later in life. For example, small subsets of patients with initial diagnoses of postmenopausal osteoporosis, familial aortic aneurysms, or osteoarthritis have mutations in collagen genes similar to the mutations that are found in patients with OI or chondrodysplasia (CD).

Composition of Connective Tissues

Connective tissues contain a large number of complex macromolecules **(Table 363-1)**. The most abundant components are three similar fibrillar collagens (types I, II, and III). They have about the same tensile strength as steel wires. The three fibrillar collagens are distributed in a tissue-specific manner: Type I collagen accounts for most of the protein of dermis, ligaments, tendons, and demineralized bone; type I and type III are the most abundant proteins of large blood vessels; and type II is the most abundant protein of cartilage.

Table 363-1 Constituents of Connective Tissues in Various Tissues

Connective Tissue	Major Constituents	Approximate Amounts, % dry wt	Characteristics or Functions
Dermis, ligaments, tendons	Type I collagen	80	Large bundles of fibrils
	Type III collagen	5–15	Thin fibrils
	Type IV collagen, laminins, and nidogen	<5	Form basal laminae under epithelium
	Types V, VI, and VII collagens	<5	V modifies type I fibrils; VI forms pericellular hexagonal network; VII forms anchoring fibrils for epidermis
	Fibrillin aggregates/elastin	<5	Provide elasticity
Bone (demineralized)	Fibronectin	<5	Associated with collagen fibers and cell surfaces
	Proteoglycans[a]/hyaluronan	<0.5	Provide resiliency
	Type I collagen	90	Complex fibril network
	Type VI collagen	1–2	Pericellular hexagonal network
	Proteoglycans[a]/hyaluronan	1	Function unclear
	Osteonectin, osteopontin, osteocalcin, α2-glycoprotein, and sialoproteins	1–5	May regulate mineralization
Aorta	Type I collagen	20–40	Fibril network
	Type III collagen	20–40	Thin fibrils
	Fibrillin aggregates/elastin	20–40	Provide elasticity
	Type IV collagen, laminins, and nidogen	<5	Form basal lamina under endothelial cells
	Types V and VI collagens	<2	V modifies type I fibrils; VI forms pericellular hexagonal network
	Proteoglycans[a]/hyaluronan	<3	Provide resiliency
Cartilage	Type II collagen	40–50	Arcades of thin fibrils
	Type IX collagen	5–10	Links type II fibrils and other components
	Type X collagen	5–10	Surrounds hypertrophic cells
	Type XI collagen	<10	Incorporated into some type II fibrils
	Proteoglycans[a]/hyaluronan	15–50	Provide resiliency

	[illegible]		
	Small leucine-rich repeat proteins (SLRPs; >6 kinds)	<5	Multiple functions in assembly and function of the tissue

[a]Over 30 proteoglycans have now been identified. They differ in the structures of their core proteins and their contents of glycosaminoglycan side chains of chondroitin-4-sulfate, chondroitin-6-sulfate, dermatan sulfate, and keratin sulfate. Basal lamina contains a proteoglycan with a side chain of heparan sulfate that resembles heparin.

Biosynthesis and Turnover of Connective Tissues

Connective tissues are among the most stable components in living organisms, but they are not inert. During embryonic development, connective tissue membranes appear as early as the four-cell blastocyst to provide strength and a structural scaffold for the developing embryo. With the development of blood vessels and skeleton, there is a rapid increase in the synthesis, degradation, and resynthesis of connective tissues. The turnover continues at a slower but still rapid pace throughout postnatal development and then spikes during the growth spurt of puberty. During adulthood, the metabolic turnover of most connective tissues is slow, but it continues at a moderate pace in bone. With age, malnutrition, physical inactivity, and low gravitational stress, the rate of degradation of most connective tissues, especially bone and skin, begins to exceed the rate of synthesis and the tissues shrink. In starvation, a large fraction of the collagen in skin and other connective tissues is degraded and provides amino acids for gluconeogenesis **(Chap. 75)**. In both osteoarthritis and rheumatoid arthritis, there is extensive degradation of articular cartilage collagen. Glucocorticoids weaken most tissues by decreasing collagen synthesis. In many pathologic states, however, collagen is deposited in excess. With most injuries to tissues, inflammatory and immune responses stimulate the deposition of collagen fibrils in the form of fibrotic scars. The deposition of the fibrils is largely irreversible and prevents regeneration of normal tissues in hepatic cirrhosis, pulmonary fibrosis, atherosclerosis, and nephrosclerosis.

STRUCTURE AND BIOSYNTHESIS OF FIBRILLAR COLLAGENS

The tensile strength of collagen fibers derives primarily from the self-assembly of protein monomers into large fibril structures in a process that resembles crystallization. The self-assembly requires monomers of highly uniform and rigid structure. It also requires a complex series of posttranslational processing steps that maintain the solubility of the monomers until they are transported to the appropriate extracellular sites for fibril assembly. Because of the stringent requirements for correct self-assembly, it is not surprising that mutations in genes for fibrillar collagens cause many of the diseases of connective tissues.

The monomers of the three fibrillar collagens are formed from three polypeptide chains, called α *chains*, that are wrapped around each other into a ropelike triple-helical conformation. The triple helix is a unique structure among proteins, and it provides rigidity to the molecule. It also orients the side chains of amino acids in an "inside out" manner relative to most other proteins so that the charged and hydrophobic residues on the surface can direct self-assembly of the monomers into fibrils. The triple-helical conformation of the monomer is generated because each of the α chains has a repetitive amino acid sequence in which glycine (Gly) appears as every third amino acid. Each α chain contains about 1000 amino acids. Therefore, the sequence of each α chain can be designated as $(\text{-Gly-X-Y-})_n$, where X and Y represent amino acids other than glycine and n is >338. The presence of glycine, the smallest amino acid, in every third position in the sequence is critical because this residue must fit in a sterically restricted space in the middle of the helix where the three chains come together. The requirement for a glycine residue at every third position explains the severe effects of mutations that convert any of the glycine residues to an amino acid with a bulkier side chain (see below). Many of the X- and Y-position amino acids are proline and hydroxyproline, which, because of their ring structures, provide additional rigidity to the triple helix. Other X- and Y-positions are occupied by charged or hydrophobic amino acids that precisely direct lateral and longitudinal assembly of the monomers into highly ordered thin fibrils. Mutations that substitute amino acids in some X- and Y-positions can, in rare instances, also produce genetic diseases.

The fibers formed by the three fibrillar collagens differ in thickness and length, but they have a similar fine structure. As viewed by electron microscopy, they all have a characteristic pattern of cross-striations that are about one-quarter the length of the monomers and reflect the precise packing into fibrils. The three fibrillar collagens, however, differ in sequences found in the X- and Y-positions of the α chains and therefore in some of their physical properties. Type I collagen is composed of two identical α 1(I) chains and a third α2(I) chain that differs slightly in its amino acid sequence. Type II collagen is composed of three identical α (II) chains. Type III collagen is composed of three identical α1(III) chains.

To deliver a monomer of the correct structure to the appropriate site of fibril assembly, the biosynthesis of fibrillar collagens involves a large number of unique processing steps **(Fig. 363-1)**. The monomer, first synthesized as a soluble precursor called *procollagen,* contains an additional globular domain at each end. As the proα chains of procollagen are synthesized on ribosomes, the free N-terminal ends move into the cisternae of the rough endoplasmic reticulum. Signal peptides at the N termini are cleaved, and additional posttranslational reactions begin. Proline residues in the Y-position of the repeating -Gly-X-Y- sequences are converted to hydroxyproline by the enzyme prolyl hydroxylase. The hydroxylation of prolyl residues is essential for the three α chains of the monomer to fold into a triple helix at body temperature. The enzyme requires ascorbic acid as one of its essential cofactors, an observation that explains why wounds fail to heal in scurvy **(Chap. 74)**. In scurvy, some of the underhydroxylated and unfolded protein accumulates in the cisternae of the rough endoplasmic reticulum and is slowly degraded. Lysine residues in the Y-position are also hydroxylated to hydroxylysine by a separate lysyl hydroxylase. Many of the hydroxylysine residues are glycosylated with galactose or with galactose and glucose. A large mannose-rich oligosaccharide is assembled on the C-terminal propeptide of each chain. The proα chains are assembled by interactions among these C-terminal propeptides that control the selection of the appropriate partner chains to form hetero- or homotrimers and provide the correct chain registration required for subsequent formation of the collagen triple helix. After the C-terminal propeptides assemble three proα chains, a nucleus of triple helix is formed near the C terminus, and the helical conformation is propagated toward the N terminus in a zipper-like manner that resembles crystallization. The folding into the triple helix is spontaneous in solution, but as discussed below, identification of rare mutations causing OI demonstrated that the folding *in cellulo* is assisted by ancillary proteins. The fully folded protein is then secreted. After secretion, procollagen is processed to collagen by cleavage of the N propeptides and C propeptides by two specific proteinases. The release of the propeptides decreases the solubility of the protein about 1000-fold. The entropic energy that is released drives the self-assembly of the collagen into fibrils. Self-assembled collagen fibers have considerable tensile strength, but their strength is increased further by cross-linking reactions that form covalent bonds between α chains in one molecule and α chains in adjacent molecules.

Figure 363-1

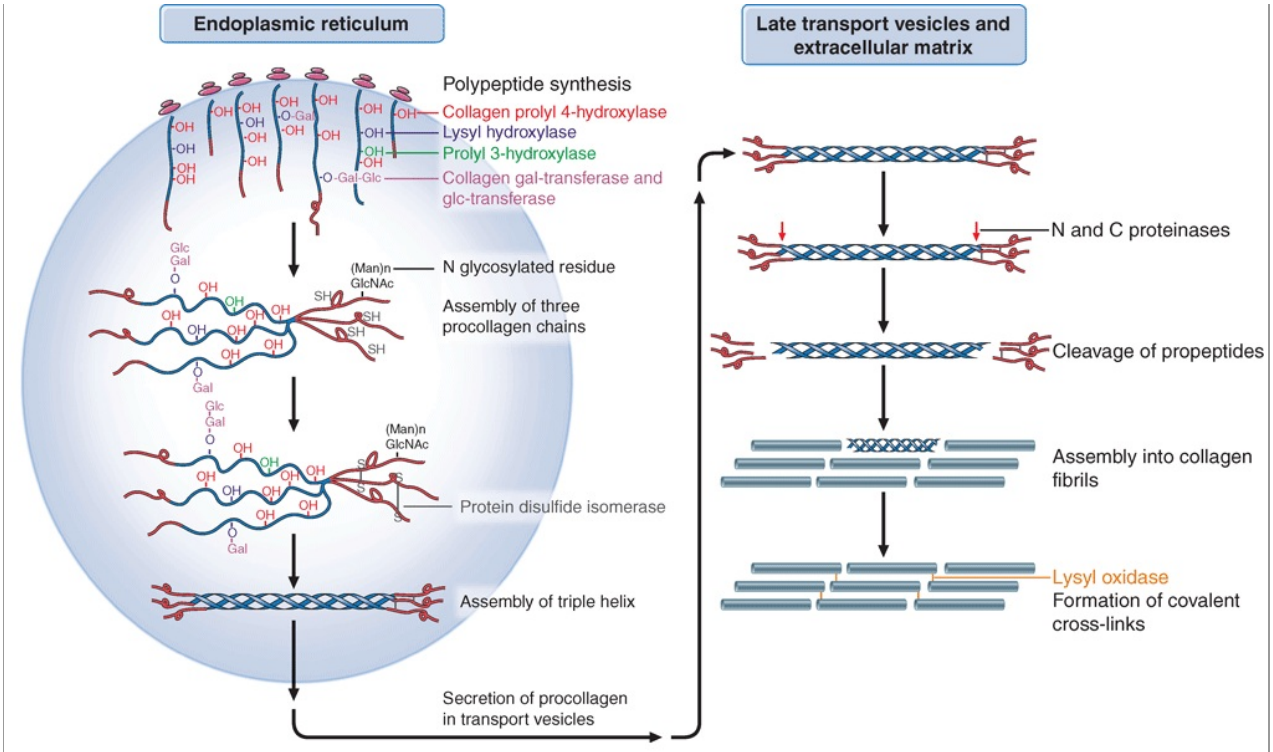

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic summary of biosynthesis of fibrillar collagens. *(Modified and reproduced with permission from J Myllyharju, KI Kivirikko: Trends in Genetics 20:33, 2004.)*

Although the assembly of collagen monomers into fibers is a spontaneous reaction, the process in tissues is modulated by the presence of less abundant collagens (type V with type I, and type XI with type II) and by other components such as a series of small leucine-rich proteins (SLRPs). Some of the less abundant components alter the rate of fibril assembly, whereas others change the morphology of the fibers or their interactions with cells and other molecules.

Collagen fibers are resistant to most proteases, but during degradation of connective tissues, they are cleaved by specific matrix metalloproteinases (collagenases) that cause partial unfolding of the triple helices into gelatin-like structures that are further degraded by less specific proteinases.

Other Collagens and Related Molecules

The unique properties of the triple helix are used to define a family of at least 29 collagens that contain repetitive -Gly-X-Y- sequences and form triple helices of varying length and complexity. The proteins are heterogeneous both in structure and function, and many are the sites of mutations causing genetic diseases. For example, the type IV collagen found in basement membranes is composed of three α chains synthesized from any of six different genes. Mutations in any of the six genes can cause Alport syndrome.

FIBRILLIN AGGREGATES AND ELASTIN

In addition to tensile strength, many tissues such as the lung, large blood vessels, and ligaments require elasticity. The elasticity was originally ascribed to an amorphous rubber-like protein named *elastin.* Subsequent analyses, largely sparked by discoveries of mutations causing the Marfan syndrome (MFS), demonstrated that the elasticity resided in thin fibrils composed primarily of large glycoproteins named *fibrillins*. The fibrillins contain large numbers of epidermal growth factor–like domains interspersed with characteristic cysteine-rich domains that are also found in latent transforming growth factor ? (TGF-?) binding proteins. The fibrillins assemble into long beadlike strands that also contain numerous other components including small and variable amounts of elastin, bone morphogenic proteins (BMPs), and microfibril-associated glycoproteins (MAGPs). The principles whereby the fibrils provide elasticity to tissue and their biosynthetic assembly are still under investigation.

PROTEOGLYCANS

The resiliency to compression of connective tissues such as cartilage or the aorta is largely explained by the presence of proteoglycans. Proteoglycans are composed of a core protein to which are attached a large series of negatively charged polymers of disaccharides (largely chondroitin sulfates). At least 30 proteoglycans have been identified. They vary in their binding to collagens and other components of matrix, but specific functions have not been assigned to most. The major proteoglycan of cartilage, called *aggrecan*, has a core 2000-amino-acid protein that is decorated with about 100 side chains of chondroitin sulfate and keratin sulfate. The core protein, in turn, binds to long chains of the polymeric disaccharide hyaluronan to form proteoglycan aggregates, one of the largest soluble macromolecular structures in nature. Because of its highly negative charge and extended structure, the proteoglycan aggregate binds large amounts of water and small ions to distend the three-dimensional arcade of collagen fibers found in the same tissues. They thereby make the cartilage resilient to pressure.

SPECIFIC DISORDERS

Osteogenesis Imperfecta (Oi)

The central feature of OI is a severe decrease in bone mass that makes bones brittle. The disorder is frequently associated with blue sclerae, dental abnormalities (dentinogenesis imperfecta), progressive hearing loss, and a positive family history.

CLASSIFICATION

An extensive classification for OI has recently been revised **(Table 363-2)**. Type I is the mildest subtype and can produce either mild or no apparent deformities of the skeleton. Most patients have distinctly blue sclerae. Type II produces bone so brittle that it is lethal in utero or shortly after birth; it is subclassified into types II A, B, and C, depending on radiologic findings. Of the severe nonlethal forms, type III is severely deforming, and type IV is moderately deforming. The rarer types V to VIII of OI are moderately

to severely deforming (Table 363-2).

Table 363-2 Expanded Classification of Osteogenesis Imperfecta (Oi)

Type	Bone Fragility	Blue Sclerae	Abnormal Dentition	Hearing Loss	Inheritance
I	Mild	Present	Present in some	Present in most	AD
II	Extreme	Present	Present in some	Unknown	S, rarely AR
III	Severe	Bluish at birth	Present in some	High incidence	AD, rarely AR
IV	Variable	Absent	Absent in IVA, present in IVB	High incidence	AD
V	Moderate to severe	Absent	Absent		AD
VI	Moderate to severe	Absent	Absent		Unknown
VII	Moderate	Absent	Absent	Absent	AR

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; S, sporadic.

The initial classifications of patients by types of OI do not always predict the clinical course of the disease. Some patients appear normal at birth and become progressively worse; others have multiple fractures in infancy and childhood, improve after puberty, and fracture more frequently later in life. Women are particularly prone to fracture during pregnancy and after menopause. A few women from families with mild variants of OI do not develop fractures until after menopause, and their disease may be difficult to distinguish from postmenopausal osteoporosis.

INCIDENCE

Type I OI has a frequency of about 1 in 30,000 births. Type II OI has a reported incidence of about 1 in 60,000, but the combined incidence of the three severe forms that are recognizable at birth (types II, III, and IV) may be much higher. Only a few patients with types V, VI, VII, and VIII have been reported.

SKELETAL EFFECTS

In type I OI, the fragility of bones may be severe enough to limit physical activity or be so mild that individuals are unaware of any disability. Radiographs of the skull in patients with mild disease may show a mottled appearance because of small islands of irregular ossification. In type II OI, ossification of many bones is frequently incomplete. Continuously beaded or broken ribs and crumpled long bones (accordina femora) may be present. For reasons that are not apparent, the long bones may be either thick or thin. In types III and IV, multiple fractures from minor physical stress can produce severe deformities. Kyphoscoliosis can impair respiration, cause cor pulmonale, and predispose to pulmonary infections. The appearance of "popcorn-like" deposits of mineral in x-rays of the ends of long bones is an ominous sign. Progressive neurologic symptoms may result from basilar compression and communicating hydrocephalus. Type V OI is recognized by the presence of dislocated radial heads and hyperplastic callus formation. Rhizomelia and coxa vara are observed in patients with type VII. Type VIII is characterized by severe growth deficiency, skeletal undermineralization, and bulbous metaphyses.

In all forms of OI, bone mineral density is decreased. However, the degree of osteopenia may be difficult to evaluate because recurrent fractures limit exercise and thereby diminish bone mass. Surprisingly, fractures appear to heal normally.

OCULAR FEATURES

The sclerae can be normal, slightly bluish, or bright blue. The color is probably caused by a thinness of the collagen layers of the sclerae that allows the choroid layers to be seen. Blue sclerae, however, are an inherited trait in some families who do not have increased bone fragility.

DENTINOGENESIS

The teeth may be normal, moderately discolored, or grossly abnormal. The enamel generally appears normal, but the teeth may have a characteristic amber, yellowish brown, or translucent bluish gray color because of a deficiency of dentin that is rich in type I collagen. The deciduous teeth are usually smaller than normal, whereas permanent teeth are frequently bell-shaped and restricted at the base. In some patients, the teeth readily fracture and need to be extracted. Similar tooth defects, however, can be inherited without any evidence of OI.

HEARING LOSS

Hearing loss usually begins during the second decade of life and occurs in more than 50% of individuals over age 30. The loss can be conductive, sensorineural, or mixed, and it varies in severity. The middle ear usually exhibits maldevelopment, deficient ossification, persistence of cartilage in areas that are normally ossified, and abnormal calcium deposits.

OTHER FEATURES

Changes in other connective tissues can include thin skin that scars extensively, joint laxity with permanent dislocations indistinguishable from those of Ehlers-Danlos syndrome (EDS), and occasionally, cardiovascular manifestations such as aortic regurgitation, floppy mitral valves, mitral incompetence, and fragility of large blood vessels. For unknown reasons, some patients develop bouts of a hypermetabolic state with elevated serum thyroxine levels, hyperthermia, and excessive sweating.

MOLECULAR DEFECTS

Most patients who meet the clinical criteria for OI are heterozygous for mutations in either the proα1 chain gene or the proα2 chain gene of type I procollagen (the *COL1A1* and *COL1A2* genes). Over 90% of patients with type I OI and blue sclerae have mutations that reduce the synthesis of pro α1chains to about one half. Mutations that reduce the synthesis of proα2 chains produce slightly more severe phenotypes and skin defects similar to EDS.

In contrast to the null mutations found in type I OI, most of the severe variants (types II, III, and IV) are caused by mutations that produce structurally abnormal proα chains that have compromised assembly or abnormal folding of the triple helix. As with collagen mutations in other connective tissue diseases, these structural mutations generally fall into two functional categories. Firstly, the relatively rare mutations in the C-propeptide domain can prevent or seriously impair initial assembly of the procollagen trimers. These misfolded chains are retained in the endoplasmic reticulum (ER) and targeted for degradation by the ER-associated proteasomal pathway. Because these mutations induce an ER-stress response, the unfolded protein response (UPR) may have many downstream effects on cells. ER stress is a new concept in the pathophysiology of connective tissue disease and has been best characterized for chondrodysplasias (see below).

The most common type I collagen mutations, however, are single base substitutions that introduce an amino acid with a bulky side chain for one of the glycine residues that appear as every third amino acid in the triple helix. In effect, any of the 338 glycine residues in either the proα1 or proα2 chain of type I procollagen is a potential site for a disease-producing mutation. These mutations compromise the structural integrity of the triple helix, causing disruption to helix folding, retention of the mutant trimers in the ER, and increased posttranslational hydroxylation and glycosylation of lysines. Collagen-containing helix mutations can form insoluble aggregates in the ER that are degraded by the autophagosome-endosome system, rather than the

proteasomes.

A similar sequence of events occurs with less common mutations that produce partial gene deletions, partial gene duplications, and splicing mutations. The search for mutations causing the very rare recessive types of OI (VII and VIII) identified mutations in genes for a series of chaperone proteins that are essential for the timely folding of the procollagen monomer: cartilage-associated protein (CRTAP), prolyl-3-hydroxylase (LEPRE1/P3H1), cyclophilin B (PPIB), collagen chaperone-like protein HSP47 (SERPINH1), and the procollagen chaperone protein FKBP65 (FKBP10).

In addition to their intracellular effects, the structurally abnormal mutant-containing collagen that is secreted by the cell can also have important extracellular affects. For example, the presence of one abnormal proα chain in a procollagen molecule can interfere with cleavage of the N propeptide from the protein. The persistence of the N propeptide on a fraction of the molecules interferes with the self-assembly of normal collagen so that thin and irregular collagen fibrils are formed. Furthermore, if structurally abnormal collagens are incorporated into fibrils, they may have a destabilizing effect and be selectively degraded, or they may alter the interactions of collagen with other connective tissue components, disturbing architecture and stability.

Several generalizations can be made about mutations in type I collagengenes. One is that unrelated patients rarely have the same mutation in the same gene. Over 900 such "private" mutations have been identified. Glycine substitutions in the N-terminal region of the triple helix tend to produce milder phenotypes, apparently because they have less effect on the zipper-like propagation of the triple-helical conformation from the C terminus. Rare substitutions of charged amino acids (Asp, Arg) or a branched amino acid (Val) in X- or Y-positions produce lethal phenotypes, apparently because they are located at sites for lateral assembly of the monomers or binding of other components of matrix.

INHERITANCE AND MOSAICISM IN GERM-LINE CELLS AND IN SOMATIC CELLS

Type I OI is usually inherited as an autosomal dominant trait. However, some patients with type I OI appear to represent sporadic new mutations or a diagnosis that was missed in earlier generations. Most lethal OI is the result of sporadic mutations that occur in the germ line in one of the parents. Because of the possibility for germ-line mosaicism for newly generated mutations, there is about a 7% probability that a second child could inherit a severe variant of OI.

DIAGNOSIS

OI is usually diagnosed on the basis of clinical criteria. The presence of fractures together with blue sclerae, dentinogenesis imperfecta, or family history of the disease is usually sufficient to make the diagnosis. Other causes of pathologic fractures must be excluded, including battered child syndrome, nutritional deficiencies, malignancies, and other inherited disorders such as CDs and hypophosphatasia, that can have overlapping presentations. The absence of superficial bruises can be helpful in distinguishing OI from battered child syndrome. X-rays usually reveal a decrease in bone density that can be verified by photon or x-ray absorptiometry. Bone microscopy can be helpful in the diagnosis. A molecular defect in type I procollagen can be demonstrated in over two-thirds of patients by incubating skin fibroblasts with radioactive amino acids and then analyzing the proα chains by polyacrylamide gel electrophoresis. The mutations themselves can be defined in most patients by the sequencing of genomic DNA with tests that are currently available from specialized laboratories.

TREATMENT: OSTEOGENESIS IMPERFECTA

Many patients with OI lead productive and successful lives despite severe deformities. Those with mild disorder may need little treatment when fractures decrease after puberty, but women require special attention during pregnancy and after menopause, when fractures again increase. More severely affected children require a comprehensive program of physical therapy and surgical management of fractures and skeletal deformities.

Many fractures are only slightly displaced and have little soft tissue swelling. Therefore, they can be treated with minimal support or traction for a week or two followed by a light cast. If fractures are relatively painless, physical therapy can be initiated early. A judicious amount of exercise prevents loss of bone mass secondary to physical inactivity. Some physicians advocate insertion of steel rods into long bones to correct limb deformities; the risk/benefits and cost/benefits of such procedures are difficult to evaluate. Aggressive conventional intervention is usually warranted for pneumonia and cor pulmonale. For severe hearing loss, stapedectomy or replacement of the stapes with a prosthesis may be successful. Moderately to severely affected patients should be evaluated periodically to anticipate possible neurologic problems. About half of children have a substantial increase in growth when given growth hormone. Treatment with bisphosphonates to decrease bone loss has been introduced for moderate to severe forms of OI. Initial responses were promising, particularly in decreasing bone pain. However the long-term effects of the drugs are still unknown. Also, a clinical trial was performed in which patients were treated by intravenous infusion of cells from bone marrow referred to as mesenchymal stem cells, or multipotential stromal cells (MSCs; **see Chap. 67**). Promising results were obtained, but the trial required a prior bone marrow transplant with marrow from a normal donor who subsequently was used as a source of normal MSCs. As a result, the procedure has not been widely adopted. However, the results raise the possibility that it may be possible in the future to develop effective stem cell therapies for OI.

Counseling and emotional support are important for patients and their parents; lay organizations in some countries provide help in these areas. Prenatal ultrasonography will detect severely affected fetuses at about 16 weeks of pregnancy. Diagnosis by demonstrating synthesis of abnormal proα chains or by DNA sequencing can be carried out in chorionic villus biopsies at 8–12 weeks of pregnancy.

Ehlers-Danlos Syndrome

EDS is characterized by hyperelasticity of the skin and hypermobile joints.

CLASSIFICATION

Several types of EDS have been defined, based on the extent to which the skin, joints, and other tissues are involved, mode of inheritance, and by molecular and biochemical analysis **(Table 363-3)**. Classical EDS includes a severe form of the disease (type I) and a milder form (type II), both characterized by joint hypermobility and skin that is velvety in texture, hyperextensible, and easily scarred. In hypermobile EDS (type III), joint hypermobility is more prominent than skin changes. In vascular-type EDS (type IV), the skin changes are more prominent than joint changes, and the patients are predisposed to sudden death from rupture of large blood vessels or other hollow organs. EDS V is similar to EDS II but is inherited as an X-linked trait. The ocular-scoliotic type of EDS (type VI) is characterized by scoliosis, ocular fragility, and a cone-shaped deformity of the cornea (keratoconus). The arthrochalasic type of EDS (type VII A and B) is characterized by marked joint hypermobility that is difficult to distinguish from EDS III except by the specific molecular defects in the processing of type I procollagen to collagen. The periodontotic-type EDS (type VIII) is distinguished by prominent periodontal changes. EDS IX, X, and XI were defined on the basis of preliminary biochemical and clinical data. EDS due to tenascin X deficiency has not been assigned a type; it is an autosomal recessive form of the syndrome similar to EDS II. The cardiac valvular form of EDS has similar features to EDS II but also involves severe changes to the aorta. The progeroid form of EDS displays features of both EDS and progeria. Because of overlapping signs and symptoms, many patients and families with some of the features of EDS cannot be assigned to any of the defined types.

Table 363-3 Different Forms of Ehlers-Danlos Syndrome

Type	Typical Features	Inheritance	Gene Defect	Protein Defect
Classic (EDS I—severe and EDS II—mild)	Skin hyperextensibility and fragility, joint hypermobility, tissue fragility manifested by widened atrophic scarring	AD	COL5A1	Collagen V
			COL5A2	
		AD	COL1A1	Proα1 (I) and proα2 (I) chains of procollagen I
		AD, AR	*COL1A2*	
Hypermobile (EDS III)	Joint hypermobility, moderate skin involvement, absence of tissue fragility	AD	*TNXB*	Tenascin X
Vascular (EDS IV)	Markedly reduced life span due to spontaneous rupture of internal organs such as arteries and intestines; skin is thin, translucent, and fragile, with extensive bruising; hypermobile minor joints; characteristic facial appearance	AD	*COL3A1*	Collagen III
X-linked EDS (EDS V)	Similar to classic type	X-linked recessive	Unknown	Unknown
Ocular-scoliotic EDS VI (EDS VIA and EDS VIB)	Features of classic EDS as well as severe muscular hypotonia after birth, progressive kyphoscoliosis, a Marfanoid habitus, osteopenia, occasionally rupture of the eye globe and great arteries	AR	*PLOD1*	Deficiency of procollagen-lysine 5-dioxygenase activity (EDS VIA)
			Unknown for EDS VIB	Unknown for EDS VIB
Arthrochalasic EDS VII (EDS VIIA and EDS VIIB)	Congenital bilateral hip dislocation, hypermobile joints, moderate skin involvement, osteopenia	AD	*COL1A1*	Mutations that prevent cleavage of the N propeptides
			COL1A2	
Dermatosparactic EDS VII C	Redundant and fragile skin, prominent hernias, joint —laxity, dysmorphic features	AR	*ADAMTS2*	Deficiency of procollagen I N-terminal proteinase
Periodontotic EDS VIII	Absorptive periodontosis with premature loss of permanent teeth, fragility of the skin, skin lesions	AD	Unknown	Unknown
EDS due to tenascin X deficiency	Similar to EDS II	AR	*TNXB*	Tenascin X
EDS, progeroid form		AR?	*B4GALT7*	Deficiency of galactosyltransferase 7 (defective synthesis of dermatan sulfate proteoglycans)

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

INCIDENCE

The overall incidence of EDS is about 1 in 5000 births, with a higher rate for blacks. Classical and hypermobile types of EDS are the most common. Patients with milder forms frequently do not seek medical attention.

SKIN

Skin changes vary from thin and velvety to skin that is either dramatically hyperextensible ("rubber person" syndrome) or easily torn or scarred. Patients with classical EDS develop characteristic "cigarette-paper" scars. In vascular-type EDS, extensive scars and hyperpigmentation develop over bony prominences, and the skin may be so thin that subcutaneous blood vessels are visible. In the periodontotic type of EDS, the skin is more fragile than hyperextensible, and it heals with atrophic, pigmented scars. Easy bruisability occurs in several types of EDS.

LIGAMENT AND JOINT CHANGES

Laxity and hypermobility of joints vary from mild to unreducible dislocations of hips and other large joints. In mild forms, patients learn to avoid dislocations by limiting physical activity. In more severe forms, surgical repair may be required. Some patients have progressive difficulty with age.

OTHER FEATURES

Mitral valve prolapse and hernias occur, particularly with type I. Pes planus and mild to moderate scoliosis are common. Extreme joint laxity and repeated dislocations may lead to degenerative arthritis. In the ocular-scoliotic type of EDS, the eye may rupture with minimal trauma, and kyphoscoliosis can cause respiratory impairment. Also, sclerae may be blue.

MOLECULAR DEFECTS

Subsets of patients with different types of EDS have mutations in the structural genes for collagens (Table 363-3). These include mutations in the *COL1A1* gene in a few patients with moderately severe classical EDS (type I); mutations in *COL1A2* in rare patients with an aortic valvular form of EDS; mutations in two of the three genes (*COL5A1* and *COL5A2*) for type V collagen, a minor collagen found in association with type I collagen, in about half the patients with classical EDS (types I and II); mutations in the *COL3A1* gene for type III collagen that is abundant in the aorta in patients with the frequently lethal vascular EDS (type IV).

Some of the type I collagen-related mutations alter processing of the protein or genes for the processing enzymes. Arthrochalasic EDS (type VII) is caused by mutations in amino acid sequence that make type I procollagen resistant to cleavage by procollagen N-proteinase or by mutations that decrease the activity of the enzyme. The persistence of the N propeptide causes the formation of collagen fibrils that are thin and irregular. Some of the patients have fragile bones and therefore a phenotype that overlaps with OI. The ocular-scoliotic type of EDS (type VI) is caused by homozygous or compound heterozygous mutations in the *PLOD 1* gene, which encodes procollagen-lysine 5-dioxygenase (lysyl hydroxylase 1), an enzyme required for formation of stable cross-links in collagen fibers.

Some patients with the hypermobile EDS (type III) and a few with mild EDS (type II) have mutations in the *TNXB* gene, which encodes tenascin X, another minor component of connective tissue that appears to regulate the assembly of collagen fibers. Mutations in proteoglycans have been found in a few patients. The progeroid form of EDS results from autosomal recessive mutations in *B4GALT7*, the gene for β-1,4-galactosyltransferase 7, a key enzyme in the addition of glycosaminoglycan chains to proteoglycans.

DIAGNOSIS

The diagnosis is based on clinical criteria. Biochemical assays and gene analyses for known molecular defects in EDS are difficult and time-consuming, but tests are sometimes available from specialized laboratories. Correlations between genotype and phenotype are challenging, but the tests are particularly useful for the diagnosis of vascular-type IV EDS with its dire prognosis.

As with other heritable diseases of connective tissue, there is a large degree of variability among members of the same family carrying the same mutation. Some patients have increased fractures and are difficult to distinguish from OI. A few families with heritable aortic aneurysms have mutations in the gene for type III collagen without any evidence of EDS or OI.

TREATMENT: EHLERS-DANLOS SYNDROME

Surgical repair and tightening of joint ligaments require careful evaluation of individual patients, as the ligaments frequently do not hold sutures. Patients with easy bruisability should be evaluated for bleeding disorders. Patients with type IV EDS and members of their families should be evaluated at regular intervals for early detection of aneurysms, but surgical repair may be difficult because of friable tissues. Also, women with type IV EDS should be counseled about the increased risk of uterine rupture, bleeding, and other complications of pregnancy.

Chondrodysplasias

(See also Chap. 355) Chondrodysplasias (CDs), also referred to as skeletal dysplasias, are heritable skeletal disorders that are characterized by dwarfism and abnormal body proportions. The category also includes some individuals with normal stature and body proportions who have features such as ocular changes or cleft palate, which are common in more severe CDs. Many patients develop degenerative joint changes, and mild CD in adults may be difficult to differentiate from primary generalized osteoarthritis.

CLASSIFICATION

Over 200 distinct types and subtypes have been defined based on criteria such as "bringing death" (thanatophoric), causing "twisted" bones (diastrophic), affecting metaphyses (metaphyseal), affecting epiphyses (epiphyseal), affecting spine (spondylo-), and producing histologic changes such as an apparent increase in the fibrous material in the epiphyses (fibrochondrogenesis). Also, a number of eponyms are based on the first or most comprehensive case reports. Severe forms of the diseases produce dwarfism with gross distortions of most cartilaginous structures and of other structures including the eye. Mild forms are more difficult to classify. Among the features are cataracts, degeneration of the vitreous, and retinal detachment high forehead hypoplastic facies cleft palate short extremities and gross distortions of the epiphyses, metaphyses, and joint surfaces. Patients with Stickler syndrome (hereditary arthro-ophthalmopathy) have been classified into three types based on a combination of the ocular phenotype and mutated genes.

INCIDENCE

The overall incidence of all forms of CD ranges from 1 per 2500 to 1 per 4000 births. Data on the frequency of individual CDs are incomplete, but the incidence of the Stickler syndrome is 1 in 10,000. Therefore, the disease is probably among the more common heritable disorders of connective tissue.

MOLECULAR DEFECTS

Mutations in the *COL2A1* gene for the type II collagen of cartilage are found in a fraction of patients with both mild and severe CDs. For example, a mutation in the gene substituting a cysteine residue for an arginine was found in three unrelated families with spondyloepiphyseal dysplasia (SED) and precocious generalized osteoarthritis (OA). Mutations in the gene were also found in some lethal CDs characterized by gross deformities of bones and cartilage, such as those found in spondyloepiphyseal dysplasia congenita, spondyloepimetaphyseal dysplasia congenita, hypochondrogenesis/achondrogenesis type II, and Kniest syndrome. The highest incidence of *COL2A1* mutations, however, occurs in patients with the distinctive features of the Stickler syndrome, which is characterized by skeletal changes, orofacial abnormalities, and auditory abnormalities. Most of the mutations in *COL2A1* are premature stop codons that produce haploinsufficiency. In addition, some of the patients with the Stickler syndrome or a closely related syndrome have mutations in two genes specific for type XI collagen, which is an unusual heterotrimer formed from α chains encoded by the gene for type II collagen (*COL2A1*) and two distinctive genes for type XI collagen (*COL11A1* and *COL11A2*). Mutations in the *COL11A1* gene are also found in patients with Marshall syndrome, which is similar to classic Stickler syndrome but with more severe hearing loss and dysmorphic features, such as a flat or retracted midface with flat nasal bridge, short nose, anteverted nostrils, long philtrum, and large-appearing eyes.

CDs are also caused by mutations in the less abundant collagens found in cartilage. For example, patients with Schmid metaphyseal CD have mutations in the gene for type X collagen, a short, network-forming collagen found in the hypertrophic zone of endochondral cartilage. The syndrome is characterized by short stature, coxa vara, flaring metaphyses, and waddling gait. As with other collagen genes, the most common mutations are of two types: Nonsense mutations that lead to haploinsufficiency and structural mutations that compromise collagen assembly. In type X collagen all the structural mutations detected occur in the C-terminal NC1 domain that coordinates the formation of the trimers. This NC1 domain is functionally equivalent to the C propeptide of the fibrillar collagens. These mutations disturb the structure of the NC1 domain, leading to misfolding and initiation of cellular ER stress via the unfolded protein response (UPR). While the UPR evolved to allow cells to adjust their ER folding capacity to differing protein folding loads, it is deployed by cells when mutant misfolded proteins accumulate in the ER. Activation of the UPR attenuates protein translation and activates mutant protein degradation pathways such as ER-associated degradation. If these strategies do not sufficiently reduce the stress response, cell death may occur. In Schmid metaphyseal CD, mutant misfolded type X collagen induces the UPR, resulting in downstream consequences that contribute to the pathophysiology.

Some patients have mutations in genes for proteins that interact with collagens. Patients with pseudoachondroplasia, or autosomal dominant multiple epiphyseal dysplasia, have mutations in the gene for the cartilage oligomeric matrix protein (*COMP*), a protein that interacts with both collagens and proteoglycans in cartilage. However, some families with multiple epiphyseal dysplasia have a defect in one of the three genes for type IX collagen (*COL9A1*, *COL9A2*, and *COL9A3*) or in matrilin-3, another extracellular protein found in cartilage. With misfolding mutations in COMP and matrilin-3, the activation of the UPR has been described, providing further evidence that the UPR is a component of pathology of these conditions.

Some CDs are caused by mutations in genes that affect early development of cartilage and related structures. The most common form of short-limbed dwarfism, achondroplasia, is caused by mutations in the gene for a receptor for a fibroblastic growth factor (*FGFR3*). The mutations in the *FGFR3* gene causing achondroplasias are unusual in several respects. The same single-base mutation in the gene that converts glycine to arginine at position 380 in the *FGFR3* gene is present in over 90% of

patients. Most patients harbor sporadic new mutations, and therefore this nucleotide change must be one of the most common recurring mutations in the human genome. The mutation causes unregulated signal transduction through the receptor and inappropriate development of cartilage. Mutations that alter other domains of *FGFR3* have been found in patients with the more severe disorders of hypochondroplasia and thanatophoric dysplasia and in a few families with a variant of craniosynostosis. However, most patients with craniosynostosis appear to have mutations in the related *FGFR2* gene. The similarities between the phenotypes produced by mutations in genes for FGF receptors and mutations in structural proteins of cartilage are probably explained by the observation that the activity of FGFs is regulated in part by binding of FGFs to proteins sequestered in the extracellular matrix. Therefore, the situation parallels the interactions between transforming growth factors (TGFs) and fibrillin in MFS (see below).

Rare mutations involve the proteoglycans of cartilage. Patients with an autosomal recessive form of multiple epiphyseal dysplasia have mutations in a gene encoding a sulfate transporter protein (*DTDST*) required for synthesis of chondroitin sulfates and proteoglycans.

DIAGNOSIS

The diagnosis of CDs is made on the basis of the physical appearance, slit-lamp eye examinations, x-ray findings, histologic changes, and clinical course. Evaluation of patients by specialists in the field is usually required for a diagnosis. DNA tests for mutations in some of the genes are available from specialized laboratories. For Stickler syndrome, more precise diagnostic criteria have made it possible to identify type I variants with mutations in the *COL2A1* gene with a high degree of accuracy. It has been suggested that the type II variant with mutations in the *COL11A1* gene can be identified on the basis of a "beaded" vitreous phenotype, and the type III variant with mutations in the *COL11A2* gene can be identified on the basis of the characteristic systemic features without the ocular involvement. Prenatal diagnosis based on analysis of DNA obtained from chorionic villus or amniotic fluid is possible.

TREATMENT: CHONDRODYSPLASIAS

The treatment is symptomatic and is directed to secondary features such as degenerative arthritis. Many patients require joint replacement surgery and corrective surgery for cleft palate. The eyes should be monitored carefully for the development of cataracts and the need for laser therapy to prevent retinal detachment. Patients should be advised to avoid obesity and contact sports. Counseling for the psychological problems of short stature is critical, and support groups have formed in many countries.

Marfan Syndrome (Mfs)

MFS includes features that primarily affect the skeleton, the cardiovascular system, and the eyes.

CLASSIFICATION

MFS was initially characterized by a triad of features: (1) skeletal changes that include long, thin extremities, frequently associated with loose joints; (2) reduced vision as the result of dislocations of the lenses (ectopia lentis); and (3) aortic aneurysms. An international panel has developed a series of "Ghent criteria" that are useful in classifying patients.

INCIDENCE AND INHERITANCE

The incidence of MFS is among the highest of any heritable disorder: about 1 in 3000/5000 births in most racial and ethnic groups. The related syndromes are less common. Mutations are generally inherited as autosomal dominant traits, but about one-fourth of patients have sporadic new mutations.

SKELETAL EFFECTS

Patients have long limbs and are usually tall compared to other members of the same family. The ratio of the upper segment (top of the head to the top of the pubic ramus) to the lower segment (top of the pubic ramus to the floor) is usually 2 SDs below mean for age, race, and sex. The fingers and hands are long and slender and have a spider-like appearance (arachnodactyly). Many patients have severe chest deformities, including depression (pectus excavatum), protrusion (pectus carinatum), or asymmetry. Scoliosis is frequent and usually accompanied by kyphosis. CT or MRI examinations of the lumbar sacral region frequently reveals enlargement of the neural canal, thinning of the pedicles and laminae, widening of the foramins, or anterior meningocele (dural ectasia). High-arched palate and high pedal arches or pes planus are common. A few patients have severe joint hypermobility similar to EDS.

CARDIOVASCULAR FEATURES

Cardiovascular abnormalities are the major source of morbidity and mortality **(Chap. 248)**. Mitral valve prolapse develops early in life and progresses to mitral valve regurgitation of increasing severity in about one-quarter of patients. Dilation of the root of the aorta and the sinuses of Valsalva are characteristic and ominous features of the disease that can develop at any age. The rate of dilation is unpredictable, but it can lead to aortic regurgitation, dissection of the aorta, and rupture. Dilation is probably accelerated by physical and emotional stress, as well as by pregnancy. Patients with familial aortic aneurysms tend to develop aneurysms at the base of the abdominal aorta. The location of the aneurysms, however, is somewhat variable, and the high incidence of aortic aneurysms in the general population (1 in 100) makes the differential diagnosis difficult unless other features of MFS are clearly present.

OCULAR FEATURES

Upward displacement of the lens is common. It is usually not progressive but may contribute to the formation of cataracts. The ocular globe is frequently elongated, and most patients are myopic but with adequate vision. Retinal detachment can occur.

OTHER FEATURES

Striae may occur over the shoulders and buttocks. A number of patients develop spontaneous pneumothorax. Inguinal and incisional hernias are common. Patients are typically thin with little subcutaneous fat, but adults may develop centripetal obesity.

MOLECULAR DEFECTS

More than 90% of patients clinically classified as having MFS by the "Ghent criteria" have a mutation in the gene for fibrillin-1 (*FBN1*). Mutations in the same gene are found in a few patients who do not meet the Ghent criteria. Also, a few MFS patients without mutations in the *FBN1* gene have mutations in the gene for TGF-β receptor 2 (*TGFBR2*). In addition, mutations in either *TGFBR2* or *TGFBR1* are found in the related Loeys-Dietz syndrome (LDS), which is characterized by aortic aneurysms, cleft palate, and hypertelorism. Mutations in the *FBN2* gene, which is structurally similar to the *FBN1* gene, which, are found in patients with MFS-like syndrome of congenital contractual arachnodactyly (CCA).

Over 550 different mutations have been found in the *FBN1* gene, scattered throughout its 65 coding exons. Most are private mutations, but about 10% are recurrent new mutations that are largely located in CpG sequences known to be "hot spots." Most severe mutations are located in the central codons (24–32). About one-third of the mutations introduce premature termination codons, and about two-thirds are missense mutations that alter calcium-binding domains in the repetitive epidermal growth factor–like domains of the protein. Rarer mutations alter the processing of the protein. As in many genetic diseases, the severity of the phenotype cannot be predicted from the nature of the mutation.

The discovery that syndromes similar to MFS are caused by mutations in *TGFBR1* and *TGFBR2* refocused attention on structural similarity between fibrillin-1 and TGF-β binding proteins that sequester TGF-β in the extracellular matrix. As a result, some of the manifestations of MFS have been shown to arise from alterations in binding sites that modulate the activity of TGF-β during early development of the skeleton and other tissues.

DIAGNOSIS

All patients with a suspected diagnosis of MFS should have a slit-lamp examination and an echocardiogram. Also, homocystinuria should be ruled out by amino acid analysis of plasma **(Chap. 364)**. The diagnosis of MFS according to the international Ghent standards places emphasis on major criteria that include presence of at least four skeletal abnormalities: ectopia lentis; dilation of the ascending aorta with or without dissection; dural ectasia; and a blood relative who meets the same criteria, with or without a DNA diagnosis. A final diagnosis is based on a balanced assessment of the major criteria together with several minor criteria. The absence of ocular changes suggests LDS, and the presence of contractures with some of the signs of OI suggests CCA.

Diagnostic tests based on detection of protein defects in fibrillin-1 in cultured skin fibroblasts or DNA analyses of most of the gene are now available from several laboratories. The results are unlikely to alter the treatment or prognosis, but are helpful to inform the patients and families and to rapidly exclude the diagnosis in unaffected family members.

TREATMENT: MARFAN SYNDROME

Propranolol or other β-adrenergic blocking agents are used to lower blood pressure and thereby delay or prevent aortic dilation. Surgical correction of the aorta, aortic valve, and mitral valve has been successful in many patients, but tissues are frequently friable. Patients should be advised of the risks of severe physical and emotional stress and of pregnancy.

The scoliosis tends to be progressive and should be treated by mechanical bracing and physical therapy if >20° or by surgery if it progresses to >45°. Dislocated lenses rarely require surgical removal, but patients should be followed closely for retinal detachment. The finding that MFS pathophysiology involves alterations in TGF-β signaling has raised the possibility of new therapeutic strategies. Attenuation of TGF-β signaling with agents such as angiotensin II receptor blockers (e.g., Losartan) appears promising based on studies in animal models and is undergoing clinical trials.

Elastin-Related Diseases

Mutations in the elastin gene (*ELN*) have been found in patients with supravalvular aortic stenosis and cutis laxa.

Epidermolysis Bullosa (Eb)

EB is of a group of similar disorders in which the skin and related epithelial tissues break and blister as the result of minor trauma.

CLASSIFICATION AND INCIDENCE

Four types of EB are defined on the basis of the level at which blistering occurs: (1) EB simplex for blistering in the epidermis, (2) EB hemidesmosomal for fissures between keratinocytes and the basal lamina, (3) EB junctional for blistering in the dermal-epidermal junction, and (4) EB dystrophica for blistering in the dermis.

The incidence of EB in the United States is about 1 in 50,000.

MOLECULAR DEFECTS

The distinctive anatomic features of skin have made it possible to relate the clinical features of subtypes of EB to mutations for specific components. Mutations in the genes for the major keratins of basal epithelial cells (keratins 14 and 15) are found in patients with the blistering skin of EB simplex. Patients with the related syndrome, epidermolytic ichthyosis, have mutations in keratin 1 and keratin 10. Mutations in a large number of genes for the complex structure of hemidesmosomes are found in EB hemidesmosomal. Hemidesmosomal mutations occur in genes for type XVII collagen, lamin 3, α6 integrin, α4 integrin, and plectin. Mutations in the genes for components of the dermal-epidermal junction are found in EB junctional and include: three lamin genes (*A3, B3,* and *C2*) and type XVII collagen. Mutations in the gene that codes for type VII collagen that forms long anchoring fibrils for the basement membrane are found in the severe syndrome of EB dystrophica.

DIAGNOSIS AND TREATMENT

The diagnosis is based on skin that readily breaks and forms blisters. EB simplex and EB hemidesmosomal are generally milder than EB junctional or EB dystrophica. EB dystrophica variants usually cause large and prominent scars. Precise classification within subtypes usually requires electron microscopy. DNA diagnostic tests are also available. The treatment is symptomatic.

Alport Syndrome (as)

AS is an inherited disorder characterized by hematuria and several associated features. Four forms of the disease are now recognized: (1) classic AS, which is inherited as an X-linked disorder with hematuria, sensorineural deafness, and conical deformation of the anterior surface of the lens (lenticonus); (2) a subtype of the X-linked form associated with diffuse leiomyomatosis; (3) an autosomal recessive form; and (4) an autosomal dominant form. Both autosomal recessive and dominant forms can cause renal disease without deafness or lenticonus.

INCIDENCE

The incidence of AS is about 1 in 10,000 births in the general population and as high as 1 in 5000 in some ethnic groups. About 80% of AS patients have the classical X-linked variant.

MOLECULAR DEFECTS

Most patients have mutations in four of the six genes for the chains of type IV collagen (*COL4A3*, *COL4A4*, *COL4A5*, and *COL4A6*). The genes for the proteins are arranged in tandem pairs on different chromosomes in an unusual head-to-head orientation and with overlapping promoters; i.e., the *COL4A1* and *COL4A2* genes are head-to-head on chromosome 13q34, the *COL4A3* and *COL4A4* genes are on chromosome 2q35–37, and the *COL4A5* and *COL4A6* genes are on chromosome Xq22. The X-linked variants are caused by either mutations in the COL4A5 gene or by partial deletions of both of the adjacent *COL4A4* and *COL4A5* genes. The autosomal recessive variants are caused by mutations in either the *COL4A3* or *COL4A4* genes. The mutations responsible for the autosomal dominant variants are still unknown, but they have been mapped to the same locus as the *COL4A3* and *COL4A4* genes.

DIAGNOSIS

The diagnosis of classic AS is based on X-linked inheritance of hematuria, sensorineural deafness, and lenticonus. The lenticonus together with hematuria is pathognomonic of classic AS. The sensorineural deafness is primarily in the high-tone range. It can frequently be detected only by an audiogram and is usually not progressive. Because of the X-linked transmission, women are generally underdiagnosed and are usually less severely affected than men. The hematuria usually progresses to nephritis and may cause renal failure in late adolescence in affected males and at older ages in some women. Renal transplantation is usually successful.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Helena Kuivaniemi, Gerard Tromp, Leena Ala-Kokko, and Malwina Czarny-Ratajcak to this chapter in previous editions of Harrison's.

FURTHER READINGS

Bateman JF et al: Genetic diseases of connective tissues: Cellular and extracellular effects of ECM mutations. Nat Rev Genet 10:173, 2009[PMID: 19204719] [Full Text]

Boileau C et al: Molecular genetics of Marfan syndrome. Curr Opin Cardiol 20:194, 2005[PMID: 15861007] [Full Text]

Cabral WA et al: Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat Genet 39:359, 2007[PMID: 17277775] [Full Text]

Callewaert B et al: Ehlers-Danlos and Marfan syndrome. Best Pract Res Clin Rheumatol 22:1521, 2008

Heinegard D: Proteoglycans and more—from molecules to biology. Int J Exp Pathol 90:575, 2009[PMID: 19958398] [Full Text]

Ishida Y et al: Autophagic elimination of misfolded procollagen aggregates in the endoplasmic reticulum as a means of cell protection. Mol Cell Biol 284:12020, 2009

Marini JC et al: Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: Regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. Hum Mutat 28:209, 2009

Ramirez F, Dietz HC: Marfan's syndrome: From molecular pathogenesis to clinical treatment. Curr Opin Genet Dev 17:252, 2007[PMID: 17467262] [Full Text]

——— Sakai LY: Biogenesis and function of fibrillin assemblies. Cell Tissue Res 339:71, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INHERITED DISORDERS OF AMINO ACID METABOLISM IN ADULTS: INTRODUCTION

Amino acids are not only the building blocks of proteins but also serve as neurotransmitters (glycine, glutamate, γ-aminobutyric acid) or as precursors of hormones, coenzymes, pigments, purines, or pyrimidines. Eight amino acids, referred to as *essential*, cannot be synthesized by humans and must be obtained from dietary sources. The others are formed endogenously. Each amino acid has a unique degradative pathway by which its nitrogen and carbon components are used for the synthesis of other amino acids, carbohydrates, and lipids. Disorders of amino acid metabolism and transport (Chap. 365) are individually rare—the incidences range from 1 in 10,000 for cystinuria or phenylketonuria to 1 in 200,000 for homocystinuria or alkaptonuria—but collectively they affect perhaps 1 in 1,000 newborns. Almost all are transmitted as autosomal recessive traits.

The features of inherited disorders of amino acid catabolism are summarized in **Table 364-1**. In general, these disorders are named for the compound that accumulates to highest concentration in blood (*-emias*) or urine (*-urias*). For many conditions (often called *aminoacidopathies*), the parent amino acid is found in excess; for others, generally referred to as *organic acidemias*, products in the catabolic pathway accumulate. Which compound(s) accumulates depends on the site of the enzymatic block, the reversibility of the reactions proximal to the lesion, and the availability of alternative pathways of metabolic "runoff." Biochemical and genetic heterogeneity are common. Five distinct forms of hyperphenylalaninemia, seven forms of homocystinuria, and seven types of methylmalonic acidemia are recognized. Such heterogeneity reflects the presence of an even larger array of molecular defects.

Table 364-1 Inherited Disorders of Amino Acid Metabolism

Amino acid(s)	Condition	Enzyme Defect	Clinical Findings	Inheritance
Phenylalanine	Phenylketonuria	Phenylalanine hydroxylase	Mental retardation, microcephaly, hypopigmented skin and hairs, eczema, "mousy" odor	AR
	DHPR deficiency hyperphenylalaninemia	Dihydropteridine reductase	Mental retardation, hypotonia, spasticity, myoclonus	AR
	PTS deficiency hyperphenylalaninemia	6-Pyruvoyl-tetrahydropterin synthase	Dystonia, neurologic deterioration, seizures, mental retardation	AR
	GCH1 deficiency hyperphenylalaninemia	GTP cyclohydrolase I	Mental retardation, seizures, dystonia, temperature instability	AR
	Carbinolamine dehydratase deficiency	Pterin-4α-carbinolamine dehydratase	Transient hyperphenylalaninemia (benign)	AR
Tyrosine	Tyrosinemia type I (hepatorenal)	Fumarylacetoacetate hydrolase	Liver failure, cirrhosis, rickets, failure to thrive, peripheral neuropathy, "boiled cabbage" odor	AR
	Tyrosinemia type II (oculocutaneous)	Tyrosine transaminase	Palmoplantar keratosis, painful corneal erosions with photophobia, mental retardation (?)	AR
	Tyrosinemia type III	4-Hydroxyphenylpyruvate dioxygenase	Hypertyrosinemia with normal liver function, occasional mental delay	AR
	Hawkinsinuria	4-Hydroxyphenylpyruvate dioxygenase	Transient failure to thrive, metabolic acidosis in infancy	AD
	Alkaptonuria	Homogentisic acid oxidase	Ochronosis, arthritis, cardiac valve involvement, coronary artery calcification	AR
	Albinism (oculocutaneous)	Tyrosinase	Hypopigmentation of hair, skin, and optic fundus; visual loss; photophobia	AR
	Albinism (ocular)	Different enzymes or transporters	Hypopigmentation of optic fundus, visual loss	AR, XL
	DOPA-responsive dystonia	Tyrosine hydroxylase	Rigidity, truncal hypotonia, tremor, mental retardation	AR
GABA	4-Hydroxybutyric aciduria	Succinic semialdehyde dehydrogenase	Seizures, mental retardation, ataxia	AR
Tryptophan	Kynurenic aciduria	Kynurenine-3-monooxygenase	Niacin deficiency, pellagra, colitis	AR
	Hydroxykynureninuria (xanthurenic aciduria)	Kynureninase	Niacin deficiency, mental retardation, spasticity	AR
Histidine	Histidinemia	Histidine-ammonia lyase	Benign	AR

	Urocanic aciduria	Urocanase	Benign	AR
	Formiminoglutamic aciduria	Formiminotransferase	Occasional mental retardation	AR
Glycine	Glycineencephalopathy	Glycine cleavage (4 enzymes)	Infantile seizures, lethargy, apnea, profound mental retardation	AR
	Sarcosinemia	Sarcosine dehydrogenase	Benign	AR
	Hyperoxaluria type I	Alanine:glyoxylate aminotransferase	Calcium oxalate nephrolithiasis, renal failure	AR
	Hyperoxaluria type II	d-Glyceric acid dehydrogenase/glyoxylate reductase	Calcium oxalate nephrolithiasis, renal failure	AR
Serine	Phosphoglycerate dehydrogenase deficiency	Phosphoglycerate dehydrogenase	Seizures, microcephaly, mental retardation	AR
Proline	Hyperprolinemia type I	Proline oxidase	Benign	AR
	Hyperprolinemia type II	Δ1-Pyrroline-5-carboxylate dehydrogenase	Febrile seizures, mental retardation	AR
	Hyperhydroxyprolinemia	Hydroxyproline oxidase	Benign	AR
	Prolidase deficiency	Prolidase	Mild mental retardation, chronic dermatitis	AR
Methionine	Hypermethioninemia	Methionine adenosyltransferase	Usually benign	AR
	S-Adenosylhomocysteine hydrolase deficiency	S-Adenosylhomocysteine hydrolase	Hypotonia, mental retardation, absent tendon reflexes, delayed myelination	AR
	Glycine N-methyltransferase deficiency	Glycine N-methyltransferase	Elevated liver transaminases	AR
Homocystine	Homocystinuria	Cystathionine β-synthase	Lens dislocation, thrombotic vascular disease, mental retardation, osteoporosis	AR
	Homocystinuria	5,10-Methylenetetrahydrofolate reductase	Mental retardation, gait and psychiatric abnormalities, recurrent strokes	AR
	Homocystinuria	Methionine synthase (cblE, -G)	Mental retardation, hypotonia, seizures, megaloblastic anemia	AR
	Homocystinuria and methylmalonic acidemia	Vitamin B12 lysosomal efflux and metabolism (cbl C, -D, -F)	Mental retardation, lethargy, failure to thrive, hypotonia, seizures, megaloblastic anemia	AR
Cystathionine	Cystathioninuria	β-Cystathionase	Benign	AR
Cystine	Cystinosis	Cystinosin CTNS (lysosomal efflux)	Renal Fanconi syndrome, rickets, photophobia, hypotonia, renal failure	AR
S-Sulfo-l-cysteine	Sulfocysteinuria	Sulfate oxidase or molybdenum cofactor deficiency	Seizures, mental retardation, dislocated lenses	AR
Lysine	Hyperlysinemia, saccharopinuria	α-Aminoadipic semialdehyde synthase	Benign	AR
	Pyridoxine-dependent seizures	L-Δ1-Piperideine-6-carboxilate dehydrogenase	Seizures	AR
Lysine, tryptophan	α-Ketoadipic acidemia	α-Ketoadipic acid dehydrogenase	Benign	?
	Glutaric acidemia type I	Glutaryl-CoA dehydrogenase	Severe dystonia and athetosis, mild mental retardation	AR
	Glutaric acidemia type II	Electron transfer flavoprotein (ETF) or ETF:ubiquinone oxidoreductase	Hypoglycemia, metabolic acidosis, "sweaty feet" odor, hypotonia, cardiomyopathy, exercise-induced myopathy	AR
Ornithine	Gyrate atrophy of the choroid and retina	Ornithine-Δ-aminotransferase	Myopia, night blindness, loss of peripheral vision, cataracts, chorioretinal degeneration	AR
Urea cycle	Carbamoylphosphate synthase-1 deficiency	Carbamoylphosphate synthase-1	Lethargy progressing to coma, protein aversion, mental retardation, hyperammonemia	AR
	N-Acetylglutamate synthase deficiency	N-Acetylglutamate synthase	Lethargy progressing to coma, protein aversion, mental retardation, hyperammonemia	AR
	Ornithine transcarbamylase deficiency	Ornithine transcarbamylase	Lethargy progressing to coma, protein aversion, mental retardation, hyperammonemia	XL
	Citrullinemia type I	Argininosuccinate synthase	Lethargy progressing to coma, protein aversion, mental retardation, hyperammonemia	AR

	Argininosuccinic acidemia	Argininosuccinate lyase	Lethargy progressing to coma, protein aversion, mental retardation, hyperammonemia, trichorrhexis nodosa	AR
	Arginase deficiency	Arginase	Spastic tetraparesis, mental retardation, mild hyperammonemia	AR
	Hyperornithinemia, hyperammonemia, homocitrullinuria	Mitochondrial ornithine carrier ORNT1	Vomiting, lethargy, failure to thrive, mental retardation, episodic confusion, hyperammonemia, protein intolerance	AR
	Citrullinemia type 2	Mitochondrial aspartate/glutamate carrier CTLN2	Neonatal intrahepatic cholestasis, adult presentation with sudden behavioral changes and stupor, coma, hyperammonemia	AR
Proline, ornithine, arginine	Δ1-pyrroline-5-carboxylate synthase deficiency	Δ1-pyrroline-5-carboxylate synthase	Hypotonia, seizures, hyperammonemia, neurodegeneration	AR
Glutamine	Glutamine synthase deficiency	Glutamine synthase	Brain malformations, pachygyria, seizures, hypotonia, dysmorphic features	AR
Valine	Isobutyryl-CoA dehydrogenase deficiency	Isobutyryl-CoA dehydrogenase	Failure to thrive, anemia, and dilated cardiomyopathy(?)	AR
Valine, leucine, isoleucine	Maple syrup urine disease (defective E1α, E1β, E2, E3)	Branched chain ketoacid dehydrogenase	Lethargy, vomiting, encephalopathy, seizures, mental retardation, "maple syrup" odor, protein intolerance	AR
Leucine	Isovaleric acidemia	Isovaleryl-CoA dehydrogenase	Acidosis, ketosis, vomiting, coma, hyperammonemia, "sweaty feet" odor, protein intolerance	AR
	3-Methylcrotonyl glycinuria	3-Methylcrotonyl-CoA carboxylase	Stress-induced metabolic acidosis, hypotonia, hypoglycemia, "cat's urine" odor	AR
	3-Methylglutaconic aciduria type I	3-Methylglutaconyl-CoA hydratase deficiency	Stress-induced acidosis, leukoencephalopathy	AR
	3-Hydroxy-3-methylglutaric aciduria	3-Hydroxy-3-methylglutaryl-CoA lyase	Stress-induced hypoketotic hypoglycemia and acidosis, encephalopathy, hyperammonemia	AR
Isoleucine	2-Methylbutyryl-glycinuria	2-Methylbutyryl-CoA dehydrogenase	Fasting-induced metabolic acidosis/hypoglycemia	AR
	2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency	2-Methyl-3-hydroxybutyryl-CoA dehydrogenase	Developmental regression, seizures, and rigidity sometimes triggered by illnesses	XL
	3-Oxothiolase deficiency	3-Oxothiolase	Fasting-induced acidosis and ketosis, vomiting, lethargy	AR
Valine, isoleucine, methionine, threonine	Propionic acidemia (pccA,-B,-C)	Propionyl-CoA carboxylase	Metabolic ketoacidosis, hyperammonemia, hypotonia, lethargy, coma, protein intolerance, mental retardation, hyperglycinemia	AR
	Multiple carboxylase/biotinidase deficiency	Holocarboxylase synthase or biotinidase	Metabolic ketoacidosis, diffuse rash, alopecia, seizures, mental retardation	AR
	Methylmalonic acidemia (mutase, racemase, CblA, -B, -D)	Methylmalonyl-CoA mutase/racemase or cobalamin reductase/adenosyltransferase	Metabolic ketoacidosis, hyperammonemia, hypertonia, lethargy, coma, protein intolerance, mental retardation, hyperglycinemia	AR

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; Cbl, cobalamin; DHPR, dihydropteridine reductase; DOPA, dihydroxyphenylalanine; GABA, g-aminobutyric acid; GTP, guanosine 5'-triphosphate; XL, X-linked.

The manifestations of these conditions differ widely (Table 364-1). Some, such as sarcosinemia, produce no clinical consequences. At the other extreme, complete deficiency of ornithine transcarbamylase is lethal in the untreated neonate. Central nervous system (CNS) dysfunction, in the form of developmental retardation, seizures, alterations in sensorium, or behavioral disturbances, is present in more than half the disorders. Protein-induced vomiting, neurologic dysfunction, and hyperammonemia occur in many disorders of urea cycle intermediates. Metabolic ketoacidosis, often accompanied by hyperammonemia, is a frequent presenting finding in disorders of branched-chain amino acid metabolism. Some disorders produce focal tissue or organ involvement such as liver disease, renal failure, cutaneous abnormalities, or ocular lesions.

The analysis of plasma amino acids (by ion-exchange chromatography), urine organic acids (by gas

chromatography/mass spectrometry), and plasma acylcarnitines (by tandem mass spectrometry) is commonly used to diagnose and monitor most of these disorders. The diagnosis is confirmed by enzyme assay on cells or tissues from the patients or by DNA testing. The clinical manifestations in many of these conditions can be prevented or mitigated if a diagnosis is achieved early and appropriate treatment (e.g., dietary protein or amino acid restriction or vitamin supplementation) is instituted promptly. For this reason, newborn screening programs seek to identify several of these disorders. Infants with a positive screening test need additional metabolic testing (usually suggested by the newborn screening program) to confirm or exclude the diagnosis. Confirmed cases should be referred to a metabolic center for initiation of therapy. The parents need to be counseled about the recurrence risk of the disease in future pregnancies. In some cases, parents need testing to exclude metabolic alterations seen in carriers for some of these disorders (such as some forms of homocystinuria) or because they might have a disorder themselves (such as methylcrotonyl CoA carboxylase deficiency, glutaric acidemia type 1, primary carnitine deficiency, or fatty acid oxidation defects). Some metabolic disorders can remain asymptomatic until adult age, presenting only when fasting or severe stress require full activity of affected metabolic pathways to provide energy.

Selected disorders that illustrate the principles, properties, and problems presented by the disorders of amino acid metabolism are discussed in this chapter.

THE HYPERPHENYLALANINEMIAS

The hyperphenylalaninemias (Table 364-1) result from impaired conversion of phenylalanine to tyrosine. The most common and clinically important is *phenylketonuria* (frequency 1:10,000), which is an autosomal recessive disorder characterized by an increased concentration of phenylalanine and its by-products in body fluids and by severe mental retardation if untreated in infancy. It results from reduced activity of phenylalanine hydroxylase (PAH deficiency phenylketonuria). The accumulation of phenylalanine inhibits the transport of other amino acids required for protein or neurotransmitter synthesis, reduces synthesis and increases degradation of myelin, and leads to inadequate formation of norepinephrine and serotonin. Phenylalanine is a competitive inhibitor of tyrosinase, a key enzyme in the pathway of melanin synthesis, and accounts for the hypopigmentation of hair and skin. Untreated children with classic phenylketonuria are normal at birth but fail to attain early developmental milestones, develop microcephaly, and demonstrate progressive impairment of cerebral function. Hyperactivity, seizures, and severe mental retardation are major clinical problems later in life. Electroencephalographic abnormalities; "mousy" odor of skin, hair, and urine (due to phenylacetate accumulation); and a tendency to hypopigmentation and eczema complete the devastating clinical picture. In contrast, affected children who are detected and treated at birth show none of these abnormalities.

Treatment: Phenylketonuria

To prevent mental retardation, diagnosis and initiation of dietary treatment of classic phenylketonuria must occur before the child is 3 weeks of age. For this reason, most newborns in North America, Australia, and Europe are screened by determinations of blood phenylalanine levels. Abnormal values are confirmed using quantitative analysis of plasma amino acids. Dietary phenylalanine restriction is usually instituted if blood phenylalanine levels are >300 μmol/L (5 mg/dL). Treatment consists of a special diet low in phenylalanine and supplemented with tyrosine, since tyrosine becomes an essential amino acid in phenylalanine hydroxylase deficiency. With therapy, plasma phenylalanine concentrations should be maintained between 60 and 360 μmol/L (1 and 6 mg/dL). Dietary restriction should be continued and monitored indefinitely. About one-third of all patients with phenylketonuria and the majority of those with milder forms (phenylalanine <1200 μmol/L at presentation) show increased tolerance to dietary proteins and improved metabolic control when treated with tetrahydrobiopterin (5–20 mg/kg per day), an essential cofactor of phenylalanine hydroxylase. This drug should be used in addition to dietary therapy.

Women with phenylketonuria, who have been treated since infancy, reach adulthood and can become pregnant. If maternal phenylalanine levels are not strictly controlled before and during pregnancy, their offspring are at increased risk for congenital defects and microcephaly (*maternal phenylketonuria*). After birth, these children have severe mental and growth retardation. Pregnancy risks can be minimized by continuing lifelong phenylalanine-restricted diets and assuring strict phenylalanine restriction 2 months prior to conception and throughout gestation.

THE HOMOCYSTINURIAS (HYPERHOMOCYSTEINEMIAS)

The homocystinurias are seven biochemically and clinically distinct disorders (Table 364-1) characterized by increased concentration of the sulfur-containing amino acid homocystine in blood and urine.

Classic homocystinuria, the most common (frequency 1:200,000), results from reduced activity of cystathionine β-synthase **(Fig. 364-1)**, the pyridoxal phosphate–dependent enzyme that condenses homocysteine with serine to form cystathionine. Most patients present between 3 and 5 years of age with dislocated optic lenses and mental retardation (in about half of cases). Some patients develop a marfanoid habitus and radiologic evidence of osteoporosis.

Figure 364-1

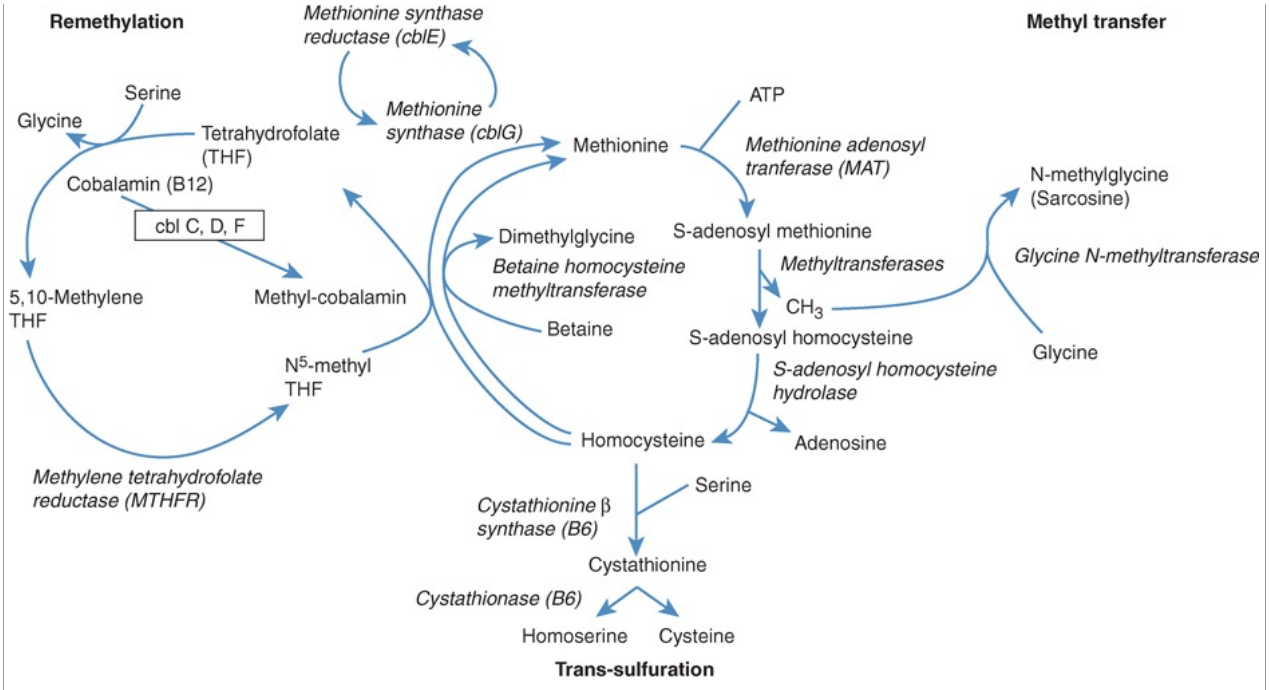

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathways, enzymes, and coenzymes involved in the homocystinurias. Methionine transfers a methyl group during its conversion to homocysteine. Defects in methyl transfer or in the subsequent metabolism of homocysteine by the pyridoxal phosphate (vitamin B6)-dependent cystathionine b-synthase increase plasma methionine levels. Homocysteine is transformed into methionine via remethylation. This occurs through methionine synthase, a reaction requiring methylcobalamin and folic acid. Deficiencies in these enzymes or lack of cofactors is associated with decreased or normal methionine levels. In an alternative pathway, homocysteine can be remethylated by betaine:homocysteine methyl transferase.

Life-threatening vascular complications (affecting coronary, renal, and cerebral arteries) can occur during the first decade of life and are the major cause of morbidity and mortality. Classic homocystinuria can be diagnosed with analysis of plasma amino acids, showing elevated methionine and presence of free homocystine. Total plasma homocysteine is also extremely elevated (usually >100 μM). Treatment consists of a special diet restricted in protein and methionine and supplemented with cystine. In approximately half of patients, oral pyridoxine (25–500 mg/d) produces a decrease in plasma methionine and homocystine concentration in body fluids. Folate and vitamin B12 deficiency should be prevented by adequate supplementation. Betaine is also effective in reducing homocystine levels.

The other forms of homocystinuria are the result of impaired remethylation of homocysteine to methionine. This can be caused by defective methionine synthase or reduced availability of two essential cofactors, 5-methyltetrahydrofolate and methylcobalamin (methyl-vitamin B12).

Hyperhomocysteinemia refers to increased total plasma concentration of homocysteine with or without an increase in free homocystine (disulfide form). Hyperhomocysteinemia, in the absence of significant homocystinuria, is found in some heterozygotes for the genetic defects noted above or in homozygotes for milder variants. Changes of homocysteine levels are also observed with increasing age; with smoking; in postmenopausal women; in patients with renal failure, hypothyroidism, leukemias, inflammatory bowel disease, or psoriasis; and during therapy with drugs such as methotrexate, nitrous oxide, isoniazid, and some antiepileptic agents. Homocysteine acts as an atherogenic and thrombophilic agent. An increase in total plasma homocysteine represents an independent risk factor for coronary, cerebrovascular, and peripheral arterial disease as well as for deep-vein thrombosis (Chap. 241). Homocysteine is synergistic with hypertension and smoking, and it is additive with other risk factors that predispose to peripheral arterial disease. In addition, hyperhomocysteinemia and folate and vitamin B12 deficiency have been associated with an increased risk of neural tube defects in pregnant women. Vitamin supplements are effective in reducing plasma homocysteine levels in these cases, although the risk reduction for cardiovascular disease and stroke has been inconsistent among different studies.

ALKAPTONURIA

Alkaptonuria is a rare (frequency 1:200,000) disorder of tyrosine catabolism in which deficiency of homogentisate 1,2-dioxygenase (also known as *homogentisic acid oxidase*) leads to excretion of large amounts of homogentisic acid in urine and accumulation of oxidized homogentisic acid pigment in connective tissues (*ochronosis*). Alkaptonuria may go unrecognized until middle life, when degenerative joint disease develops. Prior to this time, about half of patients might be diagnosed for the presence of dark urine. Foci of gray-brown scleral pigment and generalized darkening of the concha, anthelix, and, finally, helix of the ear usually develop after age 30. Low back pain usually starts between 30 and 40 years of age. *Ochronotic arthritis* is heralded by pain, stiffness, and some limitation of motion of the hips, knees, and shoulders. Acute arthritis may resemble rheumatoid arthritis, but small joints are usually spared. Pigmentation of heart valves, larynx, tympanic membranes, and skin occurs, and occasional patients develop pigmented renal or prostatic calculi. Degenerative cardiovascular disease is increased in older patients. The diagnosis should be suspected in a patient whose urine darkens to blackness. Homogentisic acid in urine is identified by urine organic acid analysis or by a specific colorimetric test. Ochronotic arthritis is treated symptomatically (Chap. 332). Nitisinone [2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione], a drug used in tyrosinemia type I, reduces urinary excretion of homogentisic acid, but it is still unclear whether it can prevent the long-term complications of

alkaptonuria.

UREA CYCLE DEFECTS

Excess ammonia generated from protein nitrogen is removed by the urea cycle, a process mediated by several enzymes and transporters (Table 364-1). Complete absence of any of these enzymes usually causes severe hyperammonemia in newborns, while milder variants can be seen in adults. The accumulation of ammonia and glutamine leads to brain edema and direct neuronal toxicity. Deficiencies in urea cycle enzymes are individually rare, but as a group they affect about 1:25,000 individuals. They are all transmitted as autosomal recessive traits, with the exception of ornithine transcarbamylase deficiency, which is X-linked. Hepatocytes of females with ornithine transcarbamylase deficiency express either the normal or the mutant allele due to random X-inactivation and may be unable to remove excess ammonia if mutant cells are predominant.

Infants with classic urea cycle defects present at 1–4 days of life with refusal to eat and lethargy progressing to coma and death. Milder enzyme deficiencies present with protein avoidance, recurrent vomiting, migraine, mood swings, chronic fatigue, irritability, and disorientation that can progress to coma. Acute liver failure can also be observed in adults. Females with ornithine transcarbamylase deficiency or milder forms of other urea cycle defects can present at time of childbirth due to the combination of involuntary fasting and stress that favor catabolism. The diagnosis requires measurement of plasma ammonia, plasma amino acids, and urine orotic acid, useful for differentiating ornithine transcarbamylase deficiency from carbamyl phosphate synthase-1 and *N*-acetylglutamate synthase deficiency. Hyperammonemia can also be caused by liver disease from any cause and several organic acidemias and fatty acid oxidation defects (the latter two excluded by the analysis of urine organic acids and plasma acylcarnitine profile).

Treatment: Urea Cycle Defects

Therapy is aimed at stopping catabolism and ammonia production by providing adequate calories (as intravenous glucose and lipids in the comatose patient) and, if needed, insulin. Excess nitrogen is removed by intravenous phenylacetate and benzoate (0.25 g/kg for the priming dose and subsequently as an infusion over 24 h) that conjugate with glutamine and glycine, respectively, to form phenylacetylglutamine and hippuric acid, water-soluble molecules efficiently excreted in urine. Arginine (200 mg/kg/d) becomes an essential amino acid (except in arginase deficiency) and should be provided intravenously to resume protein synthesis. Standard therapy of brain edema with mannitol and hypertonic saline should also be instituted in the comatose patient. If these measures fail to reduce ammonia, hemodialysis should be initiated promptly. Chronic therapy consists of a protein-restricted diet, phenylbutyrate (a more palatable precursor of phenylacetate), arginine or citrulline supplements, depending on the specific diagnosis. Liver transplantation should be considered in patients with severe urea cycle defects that are difficult to control medically.

Hyperammonemia due to a functional deficiency of glutamine synthase can occur in patients receiving chemotherapy for different malignancies or undergoing solid organ transplants. Several of these patients have been successfully rescued from hyperammonemia using the protocol described above for urea cycle defects.

FURTHER READINGS

Burgard P et al: Phenylketonuria, in *Pediatric Endocrinology and Inborn Errors of Metabolism* , K Sarafoglou et al (eds). New York, McGraw-Hill, 2009, pp 163–168

Maron BA, Loscalzo J: The treatment of hyperhomocysteinemia. Annu Rev Med 60:394, 2009

Mudd SH: Hypermethioninemias of genetic and non-genetic origin: A review. Am J Med Genet C Semin Med Genet 157:3, 2011[PMID: 21308989] [Full Text]

Salek J et al: Recurrent liver failure in a 25-year-old female. Liver Transpl 16:1049, 2010[PMID: 20818742] [Full Text]

Trefz FK et al: Efficacy of sapropterin dihydrochloride in increasing phenylalanine tolerance in children with phenylketonuria: A phase III, randomized, double-blind, placebo-controlled study. J Pediatr 154:700, 2009.[PMID: 19261295] [Full Text]

Vilboux T et al: Mutation spectrum of homogentisic acid oxidase (HGD) in alkaptonuria. Hum Mutat 30:1611, 2009[PMID: 19862842] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INHERITED DEFECTS OF MEMBRANE TRANSPORT: INTRODUCTION

Specific membrane transporters mediate the passage of a wide variety of substances across cellular membranes. Classes of substrates include amino acids, sugars, cations, anions, vitamins, and water. The number of inherited disorders of membrane transport continues to increase with the identification of new transporters and the clarification of the molecular basis of diseases with previously unknown pathophysiology. The first transport disorders identified affected the gut or the kidney, but transport processes are essential for the normal function of every organ. Mutations in transporter molecules cause disorders of the heart, muscle, brain, and endocrine and sensory organs **(Table 365-1)**. Inherited defects impairing the transport of selected amino acids that can present in adults are discussed here as examples of the abnormalities encountered; others are considered elsewhere in this text.

Table 365-1 Genetic Disorders of Membrane Transport (Selected Examples)

Class of Substance and Disorder	Individual Substrates	Tissues Manifesting Transport Defect	Molecular Defect	Major Clinical Manifestations	Inheritance
Amino Acids					
Cystinuria	Cystine, lysine, arginine, ornithine	Proximal renal tubule, jejunal mucosa	Shared dibasic-cystine transporter SLC3A1, SLC7A9	Cystine nephrolithiasis	AR
Dibasic aminoaciduria	Lysine, arginine, ornithine	Proximal renal tubule, jejunal mucosa	Dibasic transporter SLC7A7	Type I: Benign Type II: Protein intolerance, hyperammonemia, mental retardation	AR
Hartnup disease	Neutral amino acids	Proximal renal tubule, jejunal mucosa	Neutral amino acid transporter SLC6A19	Constant neutral aminoaciduria, intermittent symptoms of pellagra	AR
Methionine malabsorption	Methionine	Jejunal mucosa	Methionine transporter	White hair, mental retardation, convulsions, hyperpneic attacks, edema	Probable AR
Histidinuria	Histidine	Proximal renal tubule, jejunal mucosa	Histidine transporter	Mental retardation	AR
Iminoglycinuria	Glycine, proline, hydroxyproline	Proximal renal tubule, jejunal mucosa	Shared glycine –imino acid transporter	None	AR
Dicarboxylic aminoaciduria	Glutamic acid, aspartic acid	Proximal renal tubule, jejunal mucosa	Shared dicarboxylic amino acid transporter	None	Probable AR
Cystinosis	Cystine	Lysosomal membranes	Lysosomal cystine transporter CTNS	Renal failure, hypothyroidism, blindness	AR
Hexoses					
Glucose-galactose malabsorption	d-Glucose d-Galactose	Proximal renal tubule,	Sodium-dependent	Watery diarrhea on feeding glucose,	AR

		jejunal mucosa	glucose/galactose transporter SGLT1	lactose, sucrose, or galactose	
Glucose-transport defect	d-Glucose	Ubiquitous blood brain barrier	Facilitative glucose transporter GLUT1	Seizures, mental retardation	AD
Fanconi-Bickel syndrome	d-Glucose	Liver, kidney, pancreas, intestine	Facilitative glucose transporter GLUT2	Growth retardation, rickets, hepatorenal glycogenosis, hypo- and hyperglycemia	AR
Urate					
Hypouricemia	Uric acid	Proximal renal tubule	Urate transporter SLC22A12	Hypouricemia, uric acid urolithiasis	AR
Anions					
Congenital chloridorrhea	Chloride, sulfate	Ileal and colonic mucosa	Cl-/HCO3- exchanger (DRA)	Hydramnios, watery diarrhea, elevated fecal chloride, metabolic alkalosis with volume depletion, hyperaldosteronism	AR
Dent disease, X-linked recessive hypophosphatemic rickets and nephrocalcinosis	Chloride, phosphate	Proximal renal tubule	Voltage-gated Cl- channel CLCN5	Proteinuria, hypercalciuria, nephrocalcinosis, nephrolithiasis, rickets	XL
Cations					
Hyperinsulinemic hypoglycemia	Potassium	Pancreatic β cell	Sulfonylurea receptor SUR1, K+ channel KCNJ11	Neonatal hypoglycemia, hyperinsulinemia	AR
Benign familial neonatal epilepsy	Potassium	Brain	Voltage-gated K+ channels KCNQ2, KCNQ3	Neonatal seizures, normal development	AD
Water					
Nephrogenic diabetes insipidus type 2	Water	Renal collecting tubule	Aquaporin 2 (water channel)	Polyuria, dehydration, hyposthenuria	AR
Vitamins					
Thiamine-responsive megaloblastic anemia	Thiamine	Ubiquitous	Thiamine transporter SLC19A2	Megaloblastic anemia, deafness, diabetes mellitus	AR
Other					
Carnitine deficiency	Carnitine	Kidney, muscle, heart	Carnitine transporter OCTN2	Hypoketotic hypoglycemia, cardiomyopathy, hypotonia	AR
Creatine deficiency	Creatine	Brain	Creatine transporter SLC6A8	Mental retardation, seizures, hypotonia	XL

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; XL, X-linked recessive.

CYSTINURIA

Cystinuria (frequency of 1 in 10,000 to 1 in 15,000) is an autosomal recessive disorder caused by defective transporters in the apical brush border of proximal renal tubule and small intestinal cells. It is characterized by impaired reabsorption and excessive urinary excretion of the dibasic amino acids lysine, arginine, ornithine, and cystine. Because cystine is poorly soluble, its excess excretion predisposes to the formation of renal, ureteral, and bladder stones. Such stones are responsible for the signs and symptoms of the disorder.

There are two variants of cystinuria. Homozygotes for both variants have high urinary excretion of cystine, lysine, arginine, and ornithine. Type I heterozygotes usually have normal urinary amino acids, whereas most non-type I (formerly type II and type III) heterozygotes have moderately increased urinary

excretion of each of the four amino acids. The gene for type I cystinuria (*SLC3A1*, chromosome 2p16.3) encodes a membrane glycoprotein. Non-type I cystinuria is caused by mutations in *SLC7A9* (chromosome 19q13) that encodes the $b^{0,+}$ amino acid transporter. The glycoprotein encoded by *SLC3A1* favors the correct processing of the $b^{0,+}$ membrane transporter, and explains why mutations in two different genes cause a similar disease.

Cystine stones account for 1–2% of all urinary tract calculi but are the most common cause of stones in children. Cystinuria homozygotes regularly excrete 2400–7200 μmol (600–1800 mg) of cystine daily. Since the maximum solubility of cystine in the physiologic urinary pH range of 4.5–7.0 is about 1200 μmol/L (300 mg/L), cystine needs to be diluted to 2.5–7 L of water to prevent crystalluria. Stone formation usually manifests in the second or third decade but may occur in the first year of life. Symptoms and signs are those typical of urolithiasis: hematuria, flank pain, renal colic, obstructive uropathy, and infection (Chap. 287). Recurrent urolithiasis may lead to progressive renal insufficiency.

Cystinuria is suspected after observing typical hexagonal crystals in the sediment of acidified, concentrated, chilled urine or after performing a urinary nitroprusside test. Quantitative urine amino acid analysis confirms the diagnosis of cystinuria by showing selective over excretion of cystine, lysine, arginine, and ornithine. Quantitative measurements are important for differentiating heterozygotes from homozygotes and for following free cystine excretion during therapy.

Management is aimed at preventing cystine crystal formation by increasing urinary volume and by maintaining an alkaline urine pH. Fluid ingestion in excess of 4 L/d is essential, and 5–7 L/d is optimal. Urinary cystine concentration should be <1000 μmol/L (250 mg/L). The daily fluid ingestion necessary to maintain this dilution of excreted cystine should be spaced over 24 h, with one-third of the total volume ingested between bedtime and 3 a.m. Cystine solubility rises sharply above pH 7.5, and urinary alkalinization (with potassium citrate) can be therapeutic. Penicillamine (1–3 g/d) and tiopronin (α-mercaptopropionylglycine, 800–1200 mg/d in four divided doses) undergo sulfhydryl-disulfide exchange with cystine to form mixed disulfides. Since these disulfides are much more soluble than cystine, pharmacologic therapy can prevent and promote dissolution of calculi. Penicillamine can have significant side effects and should be reserved for patients who fail to respond to hydration alone or who are in a high-risk category (one remaining kidney, renal insufficiency). The dose of penicillamine should be increased gradually to limit side-effects. When medical management fails, shock wave lithotripsy, ureteroscopy, and percutaneous nephrolithotomy are effective for most stones. Open urologic surgery is considered for complex staghorn stones or when the patient has concomitant renal or ureteral abnormalities. Occasional patients progress to renal failure and require kidney transplantation.

DIBASIC AMINOACIDURIA

This disorder is characterized by a defect in renal tubular reabsorption of the three dibasic amino acids lysine, arginine, and ornithine but *not* cystine. There are two variants, transmitted as autosomal recessive traits. In the common form of dibasic aminoaciduria (type II), also known as *lysinuric protein intolerance*, homozygotes show defective intestinal transport of dibasic amino acids as well as exaggerated renal losses. It is most common in Finland (1 in 60,000), Southern Italy, and Japan, but is rare elsewhere. The transport defect affects basolateral rather than luminal membrane transport and is associated with impairment of the urea cycle. The defective gene (*SLC7A7*, chromosome 14q11.2) encodes the membrane transporter, y^+LAT, which associates with the cell-surface glycoprotein 4F2 heavy chain to form the complete sodium-independent transporter y^+L.

Manifestations are related to the losses of ornithine, arginine, and lysine. Affected patients can present in childhood with hepatosplenomegaly, protein intolerance, and episodic ammonia intoxication. Older patients may present with severe osteoporosis, impairment of kidney function, or interstitial changes in the lungs. Plasma concentrations of lysine, arginine, and ornithine can be reduced, whereas urinary excretion of lysine and orotic acid are increased. Hyperammonemia may develop after the ingestion of protein loads or with infections, probably because of insufficient amounts of arginine and ornithine to maintain proper function of the urea cycle. The clinical features have been attributed to the hyperammonemia, to insufficient amounts of lysine to support protein synthesis during growth, and to decreased production of nitric oxide caused by arginine deficiency.

Therapy consists of dietary protein restriction and supplementation of citrulline (2–8 g/d), a neutral amino acid that fuels the urea cycle when metabolized to arginine and ornithine. Pulmonary disease responds to glucocorticoids or bronchoalveolar lavage in some patients. Women with lysinuric protein intolerance, who become pregnant, have an increased risk of anemia, toxemia, and bleeding complications during delivery. These can be minimized by aggressive nutritional therapy and control of blood pressure. Their infants can have intrauterine growth restriction but have normal neurologic function.

HARTNUP DISEASE

Hartnup disease (frequency 1 in 24,000) is an autosomal recessive disorder characterized by pellagra-like skin lesions, variable neurologic manifestations, and neutral and aromatic aminoaciduria. Alanine, serine, threonine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan, glutamine, asparagine, and histidine are excreted in urine in quantities 5–10 times greater than normal, and intestinal transport of these same amino acids is defective. The defective neutral amino acid transporter, B^0AT1 encoded by the *SLC6A19* gene on chromosome 5p15, localizes to the apical membrane in renal tubular and intestinal cells.

The clinical manifestations result from nutritional deficiency of the essential amino acid tryptophan, caused by its intestinal and renal malabsorption, and of niacin, which derives in part from tryptophan metabolism. Only a small fraction of patients with the chemical findings of this disorder develop a pellagra-like syndrome, implying that manifestations depend on other factors in addition to the transport defect. The diagnosis of Hartnup disease should be suspected in any patient with clinical features of pellagra who does not have a history of dietary niacin deficiency (Chap. 74). The neurologic and psychiatric manifestations range from attacks of cerebellar ataxia to mild emotional lability to frank delirium, and they are usually accompanied by exacerbations of the erythematous, eczematoid skin rash. Fever, sunlight, stress, and sulfonamide therapy provoke clinical relapses. Diagnosis is made by detection of the neutral aminoaciduria, which does not occur in dietary niacin deficiency. Treatment is directed at niacin repletion and includes a high-protein diet and daily nicotinamide supplementation (50–250 mg).

FURTHER READINGS

Broer S: Amino acid transport across mammalian intestinal and renal epithelia. Physiol Rev 88: 249, 2008[PMID: 18195088] [Full Text]

Camargo SM et al: Aminoacidurias: Clinical and molecular aspects. Kidney Int 73:918, 2008[PMID: 18200002] [Full Text]

Kleta R et al: Mutations in SLC6A19, encoding B^0AT1, cause Hartnup disorder. Nat Genet 36:999, 2004[PMID: 15286787] [Full Text]

Palacin M et al: The genetics of heteromeric amino acid transporters. Physiology (Bethesda) 20:112, 2005[PMID: 15772300] [Full Text]

Sebastio G et al: Lysinuric protein intolerance: Reviewing concepts on a multisystem disease. Am J Med Genet C Semin Med Genet 157:54, 2011[PMID: 21308987] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF CLINICAL MANIFESTATIONS OF METABOLIC DISEASES: INTRODUCTION

The term *metabolism* is derived from the Greek *metabol,* meaning "to change." It includes the broad array of chemical pathways that are necessary for normal development and homeostasis. In practice, clinicians generally use the term *metabolism* in reference to energy utilization for anabolism or catabolism. Alternatively, intermediary metabolism describes the myriad cellular pathways that convert energy sources from one form to another (e.g., citric acid cycle). The emerging field of *metabolomics* is based on the premise that the identification and measurement of metabolic products will enhance our understanding of physiology and disease.

Over the years, the classification of metabolic diseases has extended beyond traditional pathways involved in fuel metabolism to include disorders such as lysosomal storage diseases or connective tissue diseases. Thus, metabolic diseases really reflect disorders of cell biology. For example, lysosomal storage diseases (Chap. 361) result from a variety of defects, usually in a lysosomal enzyme, causing accumulation of a substrate within the lysosome. Certain lipodystrophies and cardiomyopathies can be caused by mutations in lamin A, a structural protein in the nuclear envelope. Membrane defects (Chap. 365), usually involving transporters of amino acids, sugars, or ions, cause disorders such as cystinuria, Hartnup's disease, or Wilson's disease. Connective tissue diseases (Chap. 363) frequently involve defects in collagen synthesis or structure (osteogenesis imperfecta, Ehlers-Danlos syndrome, Alport's syndrome) or in other extracellular matrix structural proteins such as fibrillin (Marfan syndrome). Many metabolic disorders originate from defects in enzymes involved in the synthesis or degradation of amino acids, carbohydrates, lipids, purines, or pyrimidines. (Chaps. 359, 362, 364). Lipoprotein disorders (Chap. 356) are caused by defects in a wide array of cellular pathways including membrane receptors (LDL-R), enzyme defects (lipoprotein lipase), carrier proteins (apoB100), or transporters (ABCA1). In some instances, metabolic abnormalities induce compensatory physiologic responses that reflect the interactions of multiple different metabolic pathways. For example, the metabolic syndrome (Chap. 242) includes a constellation of clinical features [central obesity, hypertriglyceridemia, low high-density lipoprotein (HDL) cholesterol, hyperglycemia, and hypertension]. It likely has multiple genetic and environmental origins. Cushing's syndrome reflects the metabolic effects of excess cortisol on multiple tissues (Chap. 342).

This broader definition results in a plethora of metabolic diseases, numbering in the thousands. Fortunately, comprehensive reference sources exist, such as the Online Metabolic and Molecular Bases of Inherited Disease (OMMBID): (*http://genetics.accessmedicine.com/*) and the Online Mendelian Inheritance in Man (OMIM): (*http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM*). The study of metabolic diseases has been invaluable for advancing our understanding of human genetics by providing insight into principles such as patterns of inheritance, variable expressivity, phenotypic variation, and novel approaches to therapy, including screening programs, blood and organ transplantation, gene therapy, and enzyme replacement (Chap. 61).

This atlas provides a visual survey of selected metabolic disorders with references to the topics elsewhere in the text. The authors encourage submission of additional illustrations that might be used to facilitate learning among our peers and thereby enhance the recognition and care of patients with these disorders.

Figure e41-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

"Gauntlet" of pellagra(niacin deficiency). Indurated, lichenified, pigmented, and scaly skin on the dorsa of the hands. (***Source: K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, www.accessmedicine.com.***) See Chap. 74.

Figure e41-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Scurvy(vitamin C deficiency). Perifollicular hemorrhage on the leg. The follicles are often plugged by keratin (perifollicular hyperkeratosis). This eruption occurred in a 46-year-old alcoholic, homeless male who also had bleeding gums and loose teeth. (***Source: K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, www.accessmedicine.com.***) See Chap. 74.

Figure e41-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Podagra with gouty inflammation of the first metatarsophalangeal (MTP) joint. Note the swelling and erythema of the left first MTP. (*Courtesy of Kevin J. Knoop, MD, MS; with permission.*) See Chaps. 333 and 359.

Figure e41-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gout. Large tophi of gout located in and around the right knee. (*Courtesy of Daniel L. Savitt, MD; with permission.*) See Chaps. 333 and 359.

Figure e41-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gout. The finger is an unusual site for gouty arthritis. Examination of the synovial fluid confirmed the diagnosis. (*Courtesy of Alan B. Storrow, MD; with permission* .) See Chaps. 333 and 359.

Figure e41-6

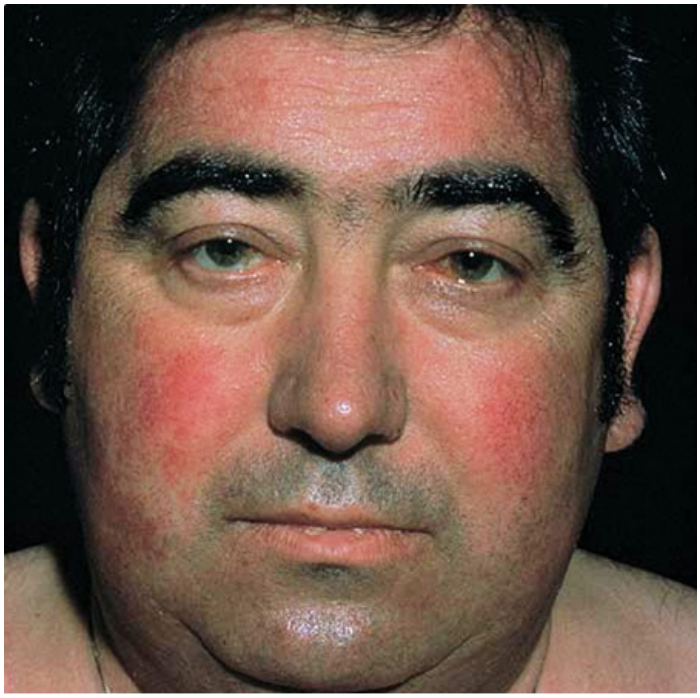

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cushing's syndrome. Plethoric moon facies with erythema and telangiectases of cheek and forehead; the face and neck and supraclavicular areas (not depicted here) show increased deposition of fat. (***Source: K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, www.accessmedicine.com.***) See Chap. 342.

Figure e41-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Necrobiosis lipoidica diabeticorum. A large, symmetric plaque with active tan-pink, well-demarcated, raised, firm borders and a yellow center in the pretibial regions of a 28-year-old diabetic female. The central parts of the lesions are depressed with atrophic changes of epidermal thinning and telangiectasis against yellow background. (***Source: K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, www.accessmedicine.com.***) See Chap. 344.

Figure e41-8

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Patient with multiple endocrine neoplasia 2B syndrome. Note the multiple neuromas on the lips and tongue and the marfanoid facies. (***Source: DG Gardner, D Shoback, eds: Greenspan's Basic & Clinical Endocrinology, 8th ed. New York, McGraw-Hill, 2006, www.accessmedicine.com.***) See Chap. 351.

Figure e41-9

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Early and late radiographs of Paget's disease of the tibia, taken when the male patient was 45 years of age (***A***) and when he was 65 years of age (***B***). (***Source: HB Skinner: Current Diagnosis & Treatment in Orthopedics, 4th ed. New York, McGraw-Hill, 2007, www.accessmedicine.com***.) See Chap. 355.

Figure e41-10

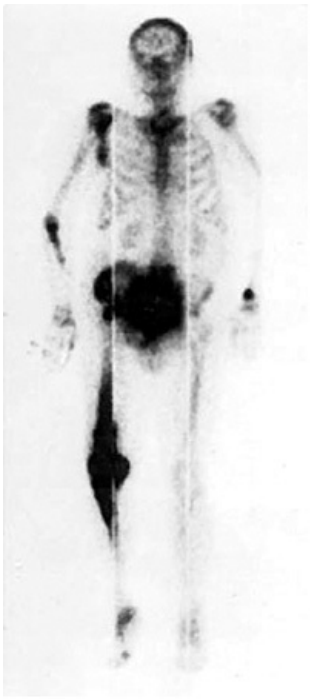

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bone scan of patient with severe Paget's disease of the skull, ribs, spine, pelvis, right femur, and acetabulum. Note localization of bone-seeking isotope (^{99m}Tc-labeled bisphosphonate) in these areas. (***Source: DG Gardner, D Shoback, eds: Greenspan's Basic & Clinical Endocrinology, 8th ed. New York, McGraw-Hill, 2006, www.accessmedicine.com.***) See Chap. 355.

Figure e41-11

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Tendinous xanthomas. Large subcutaneous tumors adherent to the Achilles tendons. (***Source: K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, www.accessmedicine.com.***) See Chap. 356.

Figure e41-12

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Papular eruptive xanthomas. ***A.*** Multiple, discrete, red-to-yellow papules becoming confluent on the elbow of an individual with uncontrolled diabetes mellitus; lesions were present on both elbows and buttocks. ***B.*** Papular eruptive xanthomas on the elbows and lower arms in an African-American patient. This image is shown to demonstrate the color of xanthomas in black skin. (***Source: K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, www.accessmedicine.com.***) See Chap. 356.

Figure e41-13

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

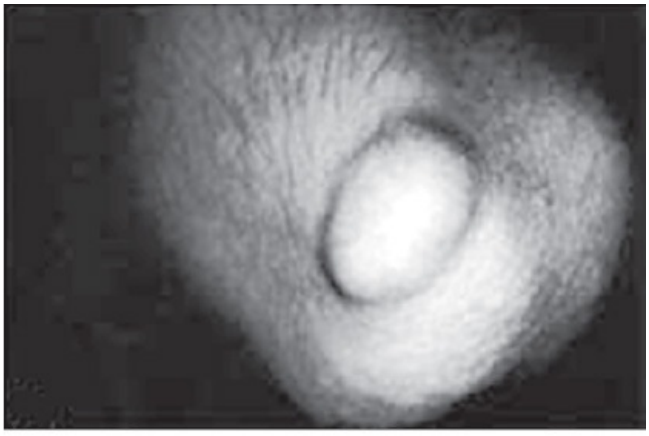

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

G

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

H

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Forms of xanthomas and other lipid deposits frequently seen in familial hypercholesterolemia homozygotes.*A.* Arcus corneae. ***B, C, E,*** and ***F.*** Cutaneous planar xanthomas, which usually have a bright orange hue. ***C*** and ***D.*** (***Source: CR Scriver, AL Beaudet, WS Sly, D Valle, eds: The Metabolic and Molecular Bases of Inherited Disease online, 8th ed. New York, McGraw-Hill, www.ommbid.com.***) Tuberous xanthomas on the elbows. ***H.*** Tendon and tuberous xanthomas. (*Panel H reproduced through the courtesy of Dr. A. Khachadurian; with permission.*) See Chap. 356.

Figure e41-14

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Examples of xanthomas in type III hyperlipoproteinemic subjects.*A.* Tuberoeruptive xanthomas of the elbows. ***B.*** Tuberous xanthomas of the digits and xanthomas of the palmar creases (xanthoma striata palmaris) (*arrows*). (*Courtesy of Dr. Thomas P. Bersot; with permission* .) See Chap. 356.

Figure e41-15

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A 17-year-old patient with abetalipoproteinemia with generalized weakness, kyphoscoliosis, and lordosis. (*Courtesy of Drs. Peter Herbert, Gerd Assmann, Antonio M. Gotto, Jr., and Donald Fredrickson; with permission.*) See Chap. 356.

Figure e41-16

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Porphyria cutanea tarda. Periorbital and malar violaceous coloration, hyperpigmentation, and hypertrichosis on the face; bullae, crusts, and scars on the dorsa of the hands. (***Source: K Wolff, RA Johnson, D Suurmond: Fitzpatrick's Color Atlas & Synopsis of Clinical Dermatology, 5th ed. New York, McGraw-Hill, 2005, www.accessmedicine.com.***) See Chap. 358.

Figure e41-17

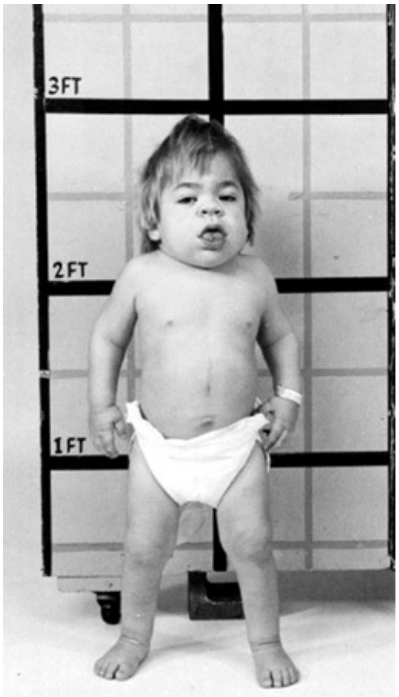

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mucopolysaccharidosis type IH (Hurler's syndrome) in a 4-year-old boy. Diagnosis was made at the age of 15 months, at which time he had developmental delay, hepatomegaly, and skeletal involvement. At the time of the picture, the patient had short stature, an enlarged tongue, persistent nasal discharge, stiff joints, and hydrocephalus. Verbal language skills consisted of four to five words. The patient had a severe hearing loss and wore hearing aids. (***Source: CR Scriver, AL Beaudet, WS Sly, D Valle, eds: The Metabolic and Molecular Bases of Inherited Disease online, 8th ed. New York, McGraw-Hill, www.ommbid.com.***) See Chap. 361.

Figure e41-18

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

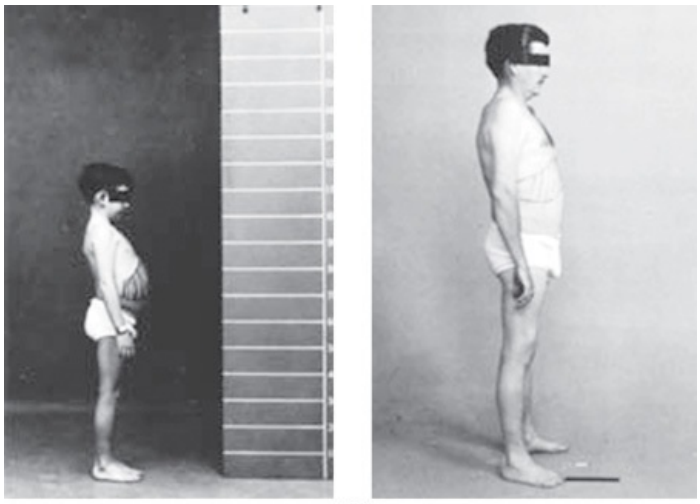

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Growth and development in two patients with type Ia glycogen storage disease.*A.* Patient at age 7 years and at age 39 years. ***B.*** Another type Ia patient at age 10 years, and follow-up at age 33 years. Both patients survive despite their disease not being adequately treated. Note that the abdomen is less protuberant with age. Hypoglycemia also improves with age. In adulthood, however, both patients continue to be short, and both have gout, multiple liver adenomas, and a progressive renal disease. (***Source: CR Scriver, AL Beaudet, WS Sly, D Valle, eds: The Metabolic and Molecular Bases of Inherited Disease online, 8th ed. New York, McGraw-Hill, www.ommbid.com.***) See Chap. 362.

Figure e41-19

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Progressive myopathy in a patient with type IIIa glycogen storage disease. The patient has a debrancher deficiency in both liver and muscle (subtype IIIa). As a child, he had hepatomegaly, hypoglycemia, and growth retardation. After puberty, he no longer had hepatomegaly, and his final height is normal. Note the muscle wasting in the lower legs and both hands at age 44 years (*left panel*); this progressed to a pronounced muscle atrophy at age 53 years (*two right panels*). (***Source: CR Scriver, AL Beaudet, WS Sly, D Valle, eds: The Metabolic and Molecular Bases of Inherited Disease online, 8th ed. New York, McGraw-Hill, www.ommbid.com.***) See Chap. 362.

Figure e41-20

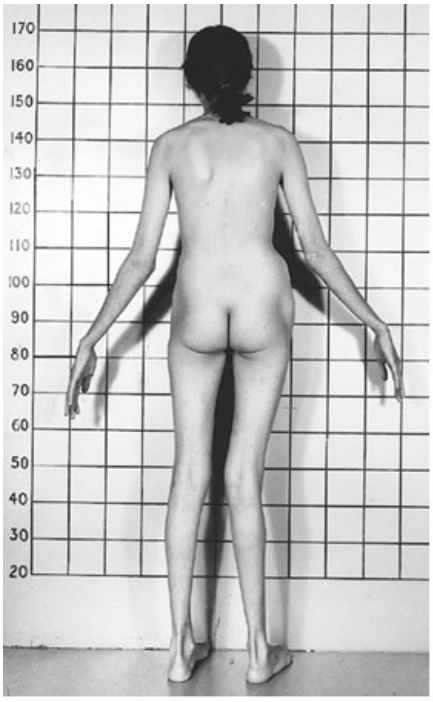

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Skeletal features of Marfan's syndrome in a 16-year-old girl. Note the long limbs that are associated with disproportionate tall stature, long fingers, scoliosis, and genu valgum. (***Source: CR Scriver, AL Beaudet, WS Sly, D Valle, eds: The Metabolic and Molecular Bases of Inherited***

Disease online, 8th ed. New York, McGraw-Hill, www.ommbid.com.) See Chap. 363.

Figure e41-21

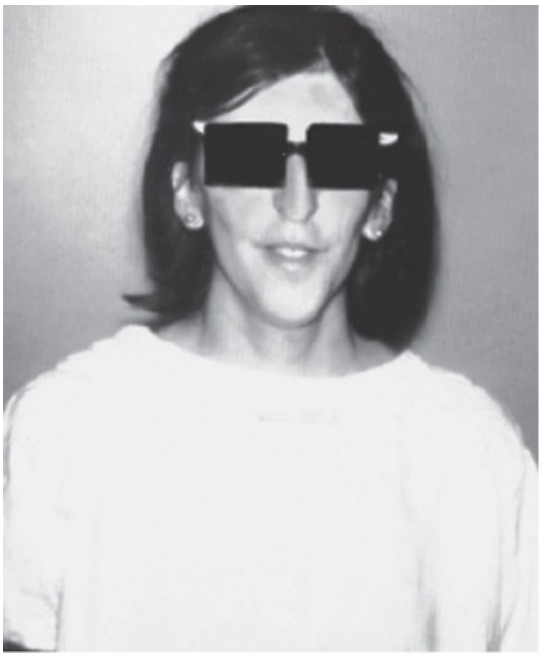

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Marfan's syndrome. ***A.*** Long, narrow face. ***B.*** Arachnodactyly and positive wrist sign. ***C.*** High-arched palate. ***D.*** Ectopia lentis associated with aortic aneurysm and severe aortic regurgitation in a teenage girl. (***Source: V Fuster, RW Alexander, RA O'Rourke, et al, eds: Hurst's The Heart, 11th ed. New York, McGraw-Hill, 2004, www.accessmedicine.com.***) See Chap. 363.

Figure e41-22

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ochronotic pigmentation of the femur of a 56-year-old alkaptonuric patient. (*Courtesy of Dr. H. W. Edmonds of the Washington Hospital Center, Washington, DC; with permission.*) See Chap. 364.

Figure e41-23

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

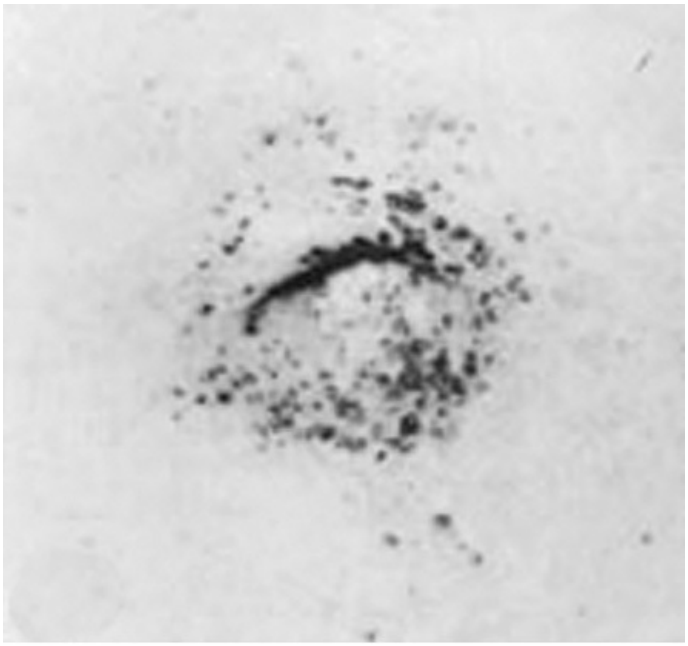

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clusters of dark-red to blue angiokeratomas (telangiectases) on the buttocks (***A***) and in the umbilical area (***B***) of a hemizygote with Fabry disease. (***Source: CR Scriver, AL Beaudet, WS Sly, D Valle, eds: The Metabolic and Molecular Bases of Inherited Disease online, 8th ed. New York, McGraw-Hill, www.ommbid.com.***) See Chap. 361.

Figure e41-24

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Two patients with type B Niemann-Pick disease (NPD).*A.* A 4.7-year-old patient with type B NPD. (*From DS Fredrickson, HR Sloan, in JB Stanbury et al: The Metabolic Basis of Inherited Disease, 3rd ed. New York, McGraw-Hill, 1972. Used by permission.*) ***B.*** A 44-year-old patient with type B NPD. See Chap. 361.

Figure e41-25

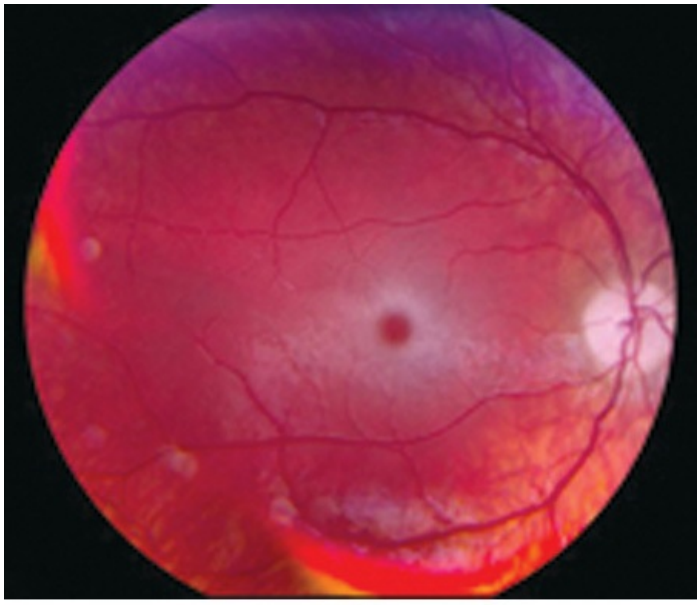

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

"Cherry red" spot in the eye of a Tay-Sachs patient. See Chap. 361. (*From http://www.nei.nih.gov/resources/eyegene.asp.*)

Figure e41-26

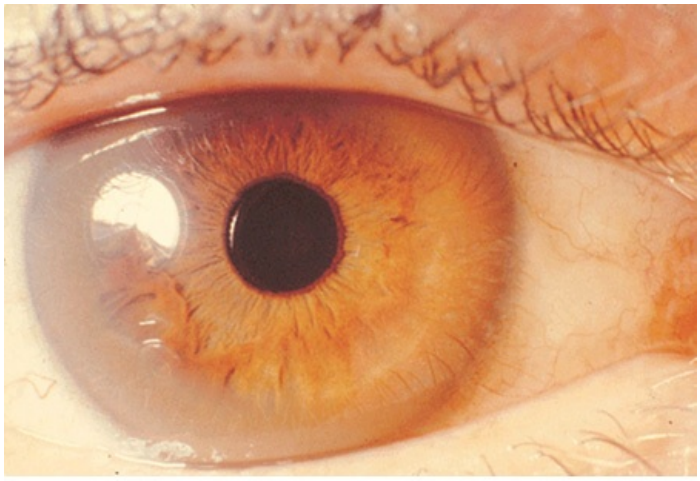

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Kayser-Fleischer ring. This develops in Wilson's disease from copper deposition in Descemet's membrane, producing brownish discoloration of the peripheral cornea. It should not be confused with the yellow-white lipid ring of arcus senilis, which is common in the elderly and occasionally signifies hyperlipidemia, especially when it appears at a young age. (*Courtesy of Jonathan C. Horton, MD, PhD; with permission.*) See Chap. 360.

Figure e41-27

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anterior view of patients with different forms of lipodystrophy. ***A.*** Congenital generalized lipodystrophy: a 16-year-old girl with generalized loss of fat, acromegaloid features, severe acanthosis nigricans affecting axillae and abdomen, umbilical hernia. (*From A Garg et al: J Clin Endocrinol Metab 84:3390, 1999; with permission.*) ***B.*** Familial partial lipodystrophy, Dunnigan variety: a 43-year-old woman with marked loss of subcutaneous fat from both the limbs and trunk and excess fat deposition in the face, chin, supraclavicular area, and labia majora. (*From JM Peters et al: Nat Genet 18;292, 1998; with permission.*) ***C.*** Acquired generalized lipodystrophy: a 10-year-old boy who developed generalized loss of fat that also affected the palms and soles after panniculitis at the age of 3 months. ***D.*** Acquired partial lipodystrophy: a 30-year-old woman with onset of lipodystrophy at age 14 years shows loss of fat from the face, neck, upper limbs, trunk, and anterior thighs. There is accumulation of excess fat in the hips and other regions of lower limbs.

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BIOLOGY OF NEUROLOGIC DISEASES: INTRODUCTION

The human nervous system is the organ of consciousness, cognition, ethics, and behavior; as such, it is the most intricate structure known to exist. More than one-third of the 23,000 genes encoded in the human genome are expressed in the nervous system. Each mature brain is composed of 100 billion neurons, several million miles of axons and dendrites, and $>10^{15}$ synapses. Neurons exist within a dense parenchyma of multifunctional glial cells that synthesize myelin, preserve homeostasis, and regulate immune responses. Measured against this background of complexity, the achievements of molecular neuroscience have been extraordinary. This chapter reviews selected themes in neuroscience that provide a context for understanding fundamental mechanisms that underlie neurologic disorders.

NEUROGENETICS

The landscape of neurology has been transformed by modern molecular genetics. More than 350 different disease-causing genes have been identified, and >1000 neurologic disorders have been genetically mapped to various chromosomal locations. Several hundred neurologic and psychiatric disorders now can be diagnosed through genetic testing *(http://www.ncbi.nlm.nih.gov/sites/GeneTests/?db=GeneTests).* The vast majority of these disorders represent highly penetrant mutations that cause rare neurologic disorders; alternatively, they represent rare monogenic causes of common phenotypes. Examples of the latter include mutations of the amyloid precursor protein in familial Alzheimer's disease, the microtubule-associated protein tau (MAPT) in frontotemporal dementia, and α-synuclein in Parkinson's disease. These discoveries have been profoundly important because the mutated gene in a familial disorder often encodes a protein that is also pathogenetically involved (although not mutated) in the typical, sporadic form. The common mechanism involves disordered processing and, ultimately, aggregation of the protein, leading to cell death (see "Protein Aggregation and Neurodegeneration," below).

There is great optimism that complex genetic disorders that are caused by combinations of genetic and environmental factors have become tractable problems. Genome-wide association studies (GWAS) have been carried out in many complex neurologic disorders, with many hundreds of variants identified, nearly all of which confer only a small increment in disease risk (1.15–1.5 fold). GWAS are rooted in the "common disease, common variant" hypothesis, as they examine potential risk alleles that are relatively common (e.g. >5%) in the general population. More than 1000 GWAS have been carried out to date, with notable successes such as the identification of >50 risk alleles for multiple sclerosis. Furthermore, when bioinformatics tools are used, risk variants can be aligned in functional biologic pathways to identify novel pathogenic mechanisms as well as to reveal heterogeneity (e.g., different pathways in different individuals). Despite these successes, many experienced geneticists question the value of common disease-associated variants, particularly whether they are actually causative or merely mark the approximate locations of more important—truly causative—rare mutations.

This debate has set the stage for the next revolution in human genetics, made possible by the development of increasingly efficient and cost-effective high-throughput sequencing methodologies. It is currently possible to sequence an entire human genome in approximately an hour, at a cost of only $4000 for the entire coding sequence ("whole-exome") or $10,000 for the entire genome; it is certain that these costs will continue to decline. This makes it feasible to look for disease-causing sequence variations in individual patients with the possibility of identifying rare variants that cause disease. The utility of this approach was demonstrated by whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy in which compound heterozygous mutations were identified in the *SH3TC2* gene that then were shown to co-segregate with the disease in other members of the family.

It is also increasingly recognized that not all genetic diseases or predispositions are caused by simple changes in the linear nucleotide sequence of genes. As the complex architecture of the human genome becomes better defined, many disorders that result from alterations in copy numbers of genes ("gene-dosage" effects) resulting from unequal crossing-over are likely to be identified. As much as 5–10% of the human genome consists of nonhomologous duplications and deletions, and these appear to occur with a much higher mutational rate than is the case for single base pair mutations. The first copy-number disorders to be recognized were Charcot-Marie-Tooth disease type 1A (CMT1A), caused by a duplication in the gene encoding the myelin protein PMP22, and the reciprocal deletion of the gene causing hereditary liability to pressure palsies (HNPP) (Chap. 384). Gene-dosage effects are

causative in some cases of Parkinson's disease (α-synuclein), Alzheimer's disease (amyloid precursor protein), spinal muscular atrophy (survival motor neuron 2), the dysmyelinating disorder Pelizaeus-Merzbacher syndrome (proteolipid protein 1), late-onset leukodystrophy (lamin B1), and a variety of developmental neurologic disorders. It is now evident that copy-number variations contribute substantially to normal human genomic variation for numerous genes involved in neurologic function, regulation of cell growth, and regulation of metabolism. It is also already clear that gene-dosage effects will influence many behavioral phenotypes, learning disorders, and autism spectrum disorders. Deletions at ch1q and ch15q have been associated with schizophrenia, and deletions at 15q and 16p with autism. Interestingly, the 16p deletion also is associated with epilepsy. Duplications of the X-linked *MeCP2* gene cause autism in males and psychiatric disorders with anxiety in females, whereas point mutations in this gene produce the neurodevelopmental disorder Rett syndrome. The understanding of the role of copy-number variation in human disease is still in its infancy.

The role of splicing variation as a contributor to neurologic disease is another area of active investigation. *Alternative splicing* refers to the inclusion of different combinations of exons in mature mRNA, resulting in the potential for many different protein products encoded by a single gene. Alternative splicing represents a powerful mechanism for generation of complexity and variation, and this mechanism appears to be highly prevalent in the nervous system, affecting key processes such as neurotransmitter receptors and ion channels. Numerous diseases are known to result from abnormalities in alternative splicing. Increased inclusion of exon 10–containing transcripts of *MAPT* can cause frontotemporal dementia. Aberrant splicing also contributes to the pathogenesis of Duchenne's, myotonic, and fascioscapulohumeral muscular dystrophies; ataxia-telangiectasia; neurofibromatosis; some inherited ataxias; and fragile X syndrome, among other disorders. It is also likely that subtle variations of splicing will influence many genetically complex disorders. For example, a splicing variant of the interleukin 7 receptor α chain, resulting in production of more soluble and less membrane-bound receptor, was found to be associated with susceptibility to multiple sclerosis (MS) in multiple different populations.

Epigenetics refers to the mechanisms by which levels of gene expression can be exquisitely modulated not by variations in the primary genetic sequence of DNA but rather by postgenomic alterations in DNA and chromatin structure, which influence how, when, and where genes are expressed. DNA methylation and the methylation and acetylation of histone proteins that interact with nuclear DNA to form chromatin are key mediators of these events. Epigenetic processes appear to be dynamically active even in postmitotic neurons. *Imprinting* refers to an epigenetic feature, present for a subset of genes, in which the predominant expression of one allele is determined by its parent of origin. The distinctive neurodevelopmental disorders Prader-Willi syndrome (mild mental retardation and endocrine abnormalities) and Angelman syndrome (cortical atrophy, cerebellar dysmyelination, Purkinje cell loss) are classic examples of imprinting disorders whose distinctive features are determined by whether the paternal or maternal copy of chromosome of the critical genetic region 15q11-13 was responsible. In a study of discordant monozygotic twins for MS in which the entire DNA sequence, transcriptome (e.g., mRNA levels), and methylome were assessed genomewide, tantalizing allelic differences in the use of the paternal, compared to maternal, copy for a group of genes were identified. Preferential allelic expression, whether due to imprinting, resistance to X inactivation, or other mechanisms, is likely to play a major role in determining complex behaviors and susceptibility to many neurologic and psychiatric disorders.

Another advance is the development of transgenic mouse models of neurologic diseases, which has been particularly fruitful in producing models relevant to Alzheimer's disease, Parkinson's disease, Huntington's disease, and amyotrophic lateral sclerosis. These models are useful in both studying disease pathogenesis and developing and testing new therapies. Models in both *Caenorhabditis elegans* and *Drosophila* have also been extremely useful, particularly in studying genetic modifiers as well as therapeutic interventions.

ION CHANNELS AND CHANNELOPATHIES

The resting potential of neurons and the action potentials responsible for impulse conduction are generated by ion currents and ion channels. Most ion channels are gated, meaning that they can transition between conformations that are open or closed to ion conductance. Individual ion channels are distinguished by the specific ions they conduct; their kinetics; and whether they directly sense voltage, are linked to receptors for neurotransmitters or other ligands such as neurotrophins, or are activated by second messengers. The diverse characteristics of different ion channels provide a means by which neuronal excitability can be modulated exquisitely at both the cellular and subcellular levels. Disorders of ion channels—channelopathies—are responsible for a growing list of human neurologic diseases **(Table 366-1)**. Most are caused by mutations in ion channel genes or by autoantibodies against ion channel proteins. One example is epilepsy, a syndrome of diverse causes characterized by repetitive, synchronous firing of neuronal action potentials. Action potentials normally are generated by the opening of sodium channels and the inward movement of sodium ions down the intracellular concentration gradient. Depolarization of the neuronal membrane opens potassium channels, resulting in outward movement of potassium ions, repolarization, closure of the sodium

channel, and hyperpolarization. Sodium or potassium channel subunit genes have long been considered candidate disease genes in inherited epilepsy syndromes, and recently such mutations were identified. These mutations appear to alter the normal gating function of these channels, increasing the inherent excitability of neuronal membranes in regions where the abnormal channels are expressed.

Table 366-1 Examples of Neurologic Channelopathies

Category	Disorder	Channel Type	Mutated Gene	Chap. Ref.
Genetic				
Ataxias	Episodicataxia-1	K	KCNA1	373
	Episodicataxia-2	Ca	CACNL1A	
	Spinocerebellar ataxia-6	Ca	CACNL1A	
Migraine	Familial hemiplegic migraine 1	Ca	*CACNL1A*	14
	Familial hemiplegic migraine 3	Na	*SCN1A*	
Epilepsy	Benign neonatal familial convulsions	K	*KCNQ2, KCNQ3*	369
	Generalized epilepsy with febrile convulsions plus	Na	*SCN1B*	
Periodic paralysis	Hyperkalemic periodic paralysis	Na	*SCN4A*	387
	Hypokalemic periodic paralysis	Ca	*CACNL1A3*	
Myotonia	Myotonia congenita	Cl	*CLCN1*	387
	Paramyotonia congenita	Na	*SCN4A*	
Deafness	Jervell and Lange-Nielsen syndrome (deafness, prolonged QT interval, and arrhythmia)	K	*KCNQ1, KCNE1*	30
	Autosomal dominant progressive deafness	K	*KCNQ4*	
Autoimmune				
Paraneoplastic	Limbic encephalitis	Kv1	—	101
	Acquired neuromyotonia	Kv1	—	101
	Cerebellar ataxia	Ca (P/Q type)	—	101
	Lambert-Eaton syndrome	Ca (P/Q type)	—	101

Whereas the specific clinical manifestations of channelopathies are quite variable, one common feature is that manifestations tend to be intermittent or paroxysmal, as occurs in epilepsy, migraine, ataxia, myotonia, or periodic paralysis. Exceptions are clinically progressive channel disorders such as autosomal dominant hearing impairment. The genetic channelopathies identified to date are all uncommon disorders caused by obvious mutations in channel genes. As the full repertoire of human ion channels and related proteins is identified, it is likely that additional channelopathies will be discovered. In addition to rare disorders that result from obvious mutations, it is likely that less penetrant allelic variations in channel genes or their pattern of expression might underlie susceptibility to some apparently sporadic forms of epilepsy, migraine, or other disorders. For example, mutations in the potassium channel gene Kir2.6 have been found in many individuals with thyrotoxic hypokalemic periodic paralysis, a disorder similar to hypokalemic periodic paralysis but precipitated by stress from thyrotoxicosis or carbohydrate loading.

NEUROTRANSMITTERS AND NEUROTRANSMITTER RECEPTORS

Synaptic neurotransmission is the predominant means by which neurons communicate with each other. Classic neurotransmitters are synthesized in the presynaptic region of the nerve terminal; stored in vesicles; and released into the synaptic cleft, where they bind to receptors on the postsynaptic cell. Secreted neurotransmitters are eliminated by reuptake into the presynaptic neuron (or glia), diffusion away from the synaptic cleft, and/or specific inactivation. In addition to the classic neurotransmitters, many neuropeptides have been identified as definite or probable neurotransmitters; they include substance P, neurotensin, enkephalins, β-endorphin, histamine, vasoactive intestinal polypeptide, cholecystokinin, neuropeptide Y, and somatostatin. Peptide neurotransmitters are synthesized in the cell body rather than the nerve terminal and may colocalize with classic neurotransmitters in single neurons. A number of neuropeptides are important in pain modulation, including substance P and calcitoningene-related peptide (CGRP), which causes migraine-like headaches in patients. As a

consequence, CGRP receptor antagonists have been developed and shown to be effective in treating migraine headaches. Nitric oxide and carbon monoxide are gases that appear also to function as neurotransmitters, in part by signaling in a retrograde fashion from the postsynaptic to the presynaptic cell.

Neurotransmitters modulate the function of postsynaptic cells by binding to specific neurotransmitter receptors, of which there are two major types. *Ionotropic receptors* are direct ion channels that open after engagement by the neurotransmitter. *Metabotropic receptors* interact with G proteins, stimulating production of second messengers and activating protein kinases, which modulate a variety of cellular events. Ionotropic receptors are multiple-subunit structures, whereas metabotropic receptors are composed of single subunits only. One important difference between ionotropic and metabotropic receptors is that the kinetics of ionotropic receptor effects are fast (generally <1 ms) because neurotransmitter binding directly alters the electrical properties of the postsynaptic cell, whereas metabotropic receptors function over longer periods. These different properties contribute to the potential for selective and finely modulated signaling by neurotransmitters.

Neurotransmitter systems are perturbed in a large number of clinical disorders, examples of which are highlighted in **Table 366-2**. One example is the involvement of dopaminergic neurons originating in the substantia nigra of the midbrain and projecting to the striatum (nigrostriatal pathway) in Parkinson's disease and in heroin addicts after the ingestion of the toxin MPTP (1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine).

Table 366-2 Principal Classic Neurotransmitters

Neurotransmitter	Anatomy	Clinical Aspects
Acetylcholine (ACh)	Motor neurons in spinal cord → neuromuscular junction	Acetylcholinesterases (nerve gases) Myasthenia gravis (antibodies to ACh receptor) Congenital myasthenic syndromes (mutations in ACh receptor subunits) Lambert-Eaton syndrome (antibodies to Ca channels impair ACh release) Botulism (toxin disrupts ACh release by exocytosis)
	Basal forebrain widespread cortex	Alzheimer's disease (selective cell death) Autosomal dominant frontal lobe epilepsy (mutations in CNS ACh receptor)
	Interneurons in striatum	Parkinson's disease (tremor)
	Autonomic nervous system (preganglionic and postganglionic parasympathetic; preganglionic sympathetic)	
Dopamine	Substantia nigra → striatum (nigrostriatal pathway) Substantia nigra → limbic system and widespread cortex Arcuate nucleus of hypothalamus → anterior pituitary (via portal veins)	Parkinson's disease (selective cell death) MPTP parkinsonism (toxin transported into neurons) Addiction, behavioral disorders Inhibits prolactin secretion
Norepinephrine (NE)	Locus coeruleus (pons) → limbic system, hypothalamus, cortex Medulla → locus coeruleus, spinal cord Postganglionic neurons of sympathetic nervous system	Mood disorders (MAOA inhibitors and tricyclics increase NE and improve depression) Anxiety Orthostatic tachycardia syndrome (mutations in NE transporter)
Serotonin	Pontine raphe nuclei widespread projections Medulla/pons dorsal horn of spinal cord	Mood disorders (SSRIs improve depression) Migraine pain pathway Pain pathway
γ-Aminobutyric acid (GABA)	Major inhibitory neurotransmitter in brain; widespread cortical interneurons and	Stiff-person syndrome (antibodies to glutamic acid decarboxylase, the

	long projection pathways	biosynthetic enzyme for GABA) Epilepsy (gabapentin and valproic acid increase GABA)
Glycine	Major inhibitory neurotransmitter in spinal cord	Spasticity Hyperekplexia (myoclonic startle syndrome) due to mutations in glycine receptor
Glutamate	Major excitatory neurotransmitter; located throughout CNS, including cortical pyramidal cells	Seizures due to ingestion of domoic acid (a glutamate analogue) Rasmussen's encephalitis (antibody against glutamate receptor 3) Excitotoxic cell death

Abbreviations: CNS, central nervous system; MAOA, monoamine oxidase A; MPTP, 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine; SSRI, selective serotonin reuptake inhibitor.

A second important dopaminergic system arising in the midbrain is the mediocorticolimbic pathway, which is implicated in the pathogenesis of addictive behaviors including drug reward. Its key components include the midbrain ventral tegmental area (VTA), median forebrain bundle, and nucleus accumbens (see Fig. 390-1). The cholinergic pathway originating in the nucleus basalis of Meynert plays a role in memory function in Alzheimer's disease.

Addictive drugs share the property of increasing dopamine release in the nucleus accumbens . Amphetamine increases intracellular release of dopamine from vesicles and reverses transport of dopamine through the dopamine transporters. Patients prone to addiction show increased activation of the nucleus accumbens after administration of amphetamine. Cocaine binds to dopamine transporters and inhibits dopamine reuptake. Ethanol inhibits inhibitory neurons in the VTA, leading to increased dopamine release in the nucleus accumbens. Opioids also disinhibit these dopaminergic neurons by binding to μ receptors expressed by γ-aminobutyric acid (GABA)-containing interneurons in the VTA. Nicotine increases dopamine release by activating nicotinic acetylcholine receptors on cell bodies and nerve terminals of dopaminergic VTA neurons. Tetrahydrocannabinol, the active ingredient of cannabis, also increases dopamine levels in the nucleus accumbens. Blockade of dopamine in the nucleus accumbens can terminate the rewarding effects of addictive drugs.

Not all cell-to-cell communication in the nervous system occurs via neurotransmission. Gap junctions provide for direct neuron-neuron electrical conduction and also create openings for the diffusion of ions and metabolites between cells. In addition to neurons, gap junctions are widespread in glia, creating a syncytium that protects neurons by removing glutamate and potassium from the extracellular environment. Gap junctions consist of membrane-spanning proteins, termed *connexins*, that pair across adjacent cells. Mechanisms that involve gap junctions have been related to a variety of neurologic disorders. Mutations in connexin 32, a gap junction protein expressed by Schwann cells, are responsible for the X-linked form of CMT disease. Mutations in either of two gap junction proteins expressed in the inner ear—connexin 26 and connexin 31—result in autosomal dominant progressive hearing loss (Chap. 30). Glial calcium waves mediated through gap junctions also appear to explain the phenomenon of spreading depression associated with migraine auras and the march of epileptic discharges. Spreading depression is a neural response that follows a variety of different stimuli and is characterized by a circumferentially expanding negative potential that propagates at a characteristic speed of 20 m/s and is associated with an increase in extracellular potassium.

SIGNALING PATHWAYS AND GENE TRANSCRIPTION

The fundamental issue of how memory, learning, and thinking are encoded in the nervous system is likely to be clarified by identification of the signaling pathways involved in neuronal differentiation, axon guidance, and synapse formation and by an understanding of how these pathways are modulated by experience. Many families of transcription factors, each consisting of multiple individual components, are expressed in the nervous system. Elucidation of these signaling pathways has begun to provide insights into the causes of a variety of neurologic disorders, including inherited disorders of cognition such as X-linked mental retardation. This problem affects ~1 in 500 males, and linkage studies in different families suggest that as many as 60 different X-chromosome-encoded genes may be responsible. Rett syndrome, a common cause of (dominant) X-linked progressive mental retardation in females, is due to a mutation in a gene (*MECP2*) that encodes a DNA-binding protein involved in transcriptional repression. As the X chromosome accounts for only ~3% of germ-line DNA, by extrapolation, the number of genes that potentially contribute to clinical disorders affecting intelligence in humans must be potentially very large. As discussed below, there is increasing evidence that abnormal gene transcription may play a role in neurodegenerative diseases such as Huntington's disease, in which proteins with polyglutamine expansions bind to and sequester transcription factors. A

critical transcription factor for neuronal survival is CREB (cyclic adenosine monophosphate responsive element-binding) protein, which also plays an important role in memory in the hippocampus.

MYELIN

Myelin is the multilayered insulating substance that surrounds axons and speeds impulse conduction by permitting action potentials to jump between naked regions of axons (nodes of Ranvier) and across myelinated segments. Molecular interactions between the myelin membrane and the axon are required to maintain the stability, function, and normal life span of both structures. A single oligodendrocyte usually ensheathes multiple axons in the central nervous system (CNS), whereas in the peripheral nervous system (PNS) each Schwann cell typically myelinates a single axon. Myelin is a lipid-rich material formed by a spiraling process of the membrane of the myelinating cell around the axon, creating multiple membrane bilayers that are tightly apposed (compact myelin) by charged protein interactions. Several inhibitors of axon growth are expressed on the innermost (periaxonal) lamellae of the myelin membrane (see "Stem Cells and Transplantation," below). A number of clinically important neurologic disorders are caused by inherited mutations in myelin proteins of the CNS or PNS **(Fig. 366-1)**. Constituents of myelin also have a propensity to be targeted as autoantigens in autoimmune demyelinating disorders **(Fig. 366-2)**. Specification to oligodendrocyte precursor cells (OPCs) is transcriptionally regulated by the *Olig 2* and *Yin Yang 1* genes, whereas myelination mediated by postmitotic oligodendrocytes depends on a different transcription factor, *myelin gene regulatory factor (MRF)*. It is noteworthy that in the normal adult brain large numbers of OPCs [expressing platelet-derived growth factor receptor alpha (PDGFR-α) and NG2] are widely distributed but do not myelinate axons, even in demyelinating environments such as lesions of MS. The characterization of these cells, including an understanding of their transcriptional regulation and functional roles, could result in novel approaches to remyelination and brain repair.

Figure 366-1

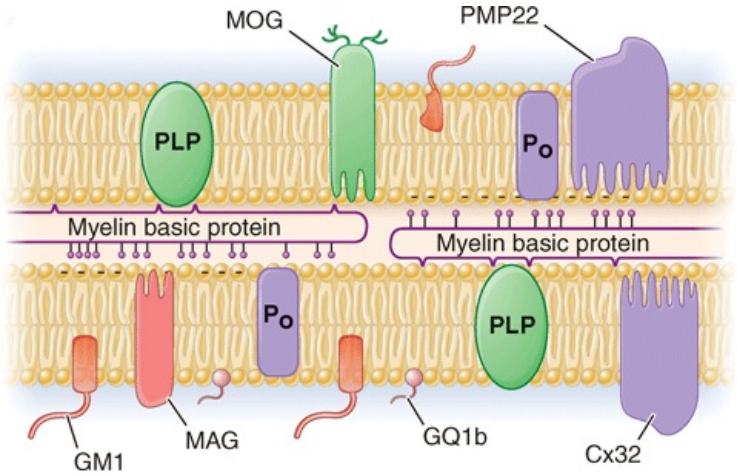

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The molecular architecture of the myelin sheath illustrating the most important disease-related proteins. The illustration represents a composite of CNS and PNS myelin. Proteins restricted to CNS myelin are shown in green, proteins of PNS myelin are lavender, and proteins present in both CNS and PNS are red. In the CNS, the X-linked allelic disorders Pelizaeus-Merzbacher disease and one variant of familial spastic paraplegia are caused by mutations in the gene for proteolipid protein (PLP) that normally promotes extracellular compaction between adjacent myelin lamellae. The homologue of PLP in the PNS is the P_0 protein, mutations in which cause the neuropathy Charcot-Marie-Tooth disease (CMT) type 1B. The most common form of CMT is the 1A subtype caused by a duplication of the *PMP22* gene; deletions in *PMP22* are responsible for another inherited neuropathy termed *hereditary liability to pressure palsies* (Chap. 384).

In multiple sclerosis (MS), myelin basic protein (MBP) and the quantitatively minor CNS protein myelin oligodendrocyte glycoprotein (MOG) are probably T cell and B cell antigens, respectively (Chap. 380). The location of MOG at the outermost lamella of the CNS myelin membrane may facilitate its targeting by autoantibodies. In the PNS, autoantibodies against myelin gangliosides are implicated in a variety of disorders, including GQ1b in the Fisher variant of Guillain-Barré syndrome, GM1 in multifocal motor neuropathy, and sulfatide constituents of myelin-associated glycoprotein (MAG) in peripheral neuropathies associated with monoclonal gammopathies (Chap. 385).

Figure 366-2

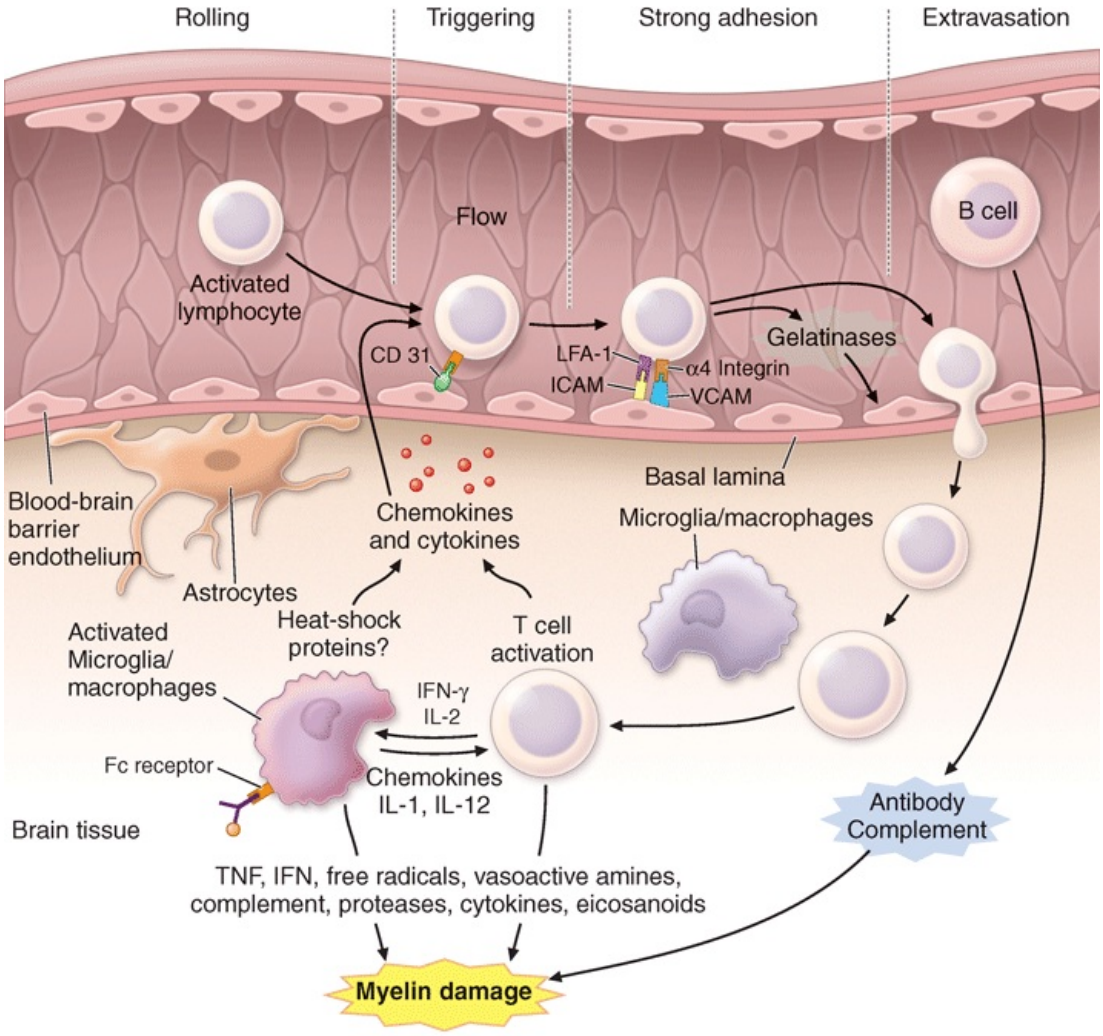

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A model for experimental allergic encephalomyelitis (EAE). Crucial steps for disease initiation and progression include peripheral activation of preexisting autoreactive T cells; homing to the CNS and extravasation across the blood-brain barrier; reactivation of T cells by exposed autoantigens; secretion of cytokines; activation of microglia and astrocytes and recruitment of a secondary inflammatory wave; and immune-mediated myelin destruction. ICAM, intercellular adhesion molecule; LFA-1, lymphocyte function-associated antigen-1; VCAM, vascular cell adhesion molecule; IFN, interferon; IL, interleukin; TNF, tumor necrosis factor.

NEUROTROPHIC FACTORS

Neurotrophic factors **(Table 366-3)** are secreted proteins that modulate neuronal growth, differentiation, repair, and survival; some have additional functions, including roles in neurotransmission and in the synaptic reorganization involved in learning and memory. The neurotrophin (NT) family contains nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), NT3, and NT4/5. The neurotrophins act at TrK and p75 receptors to promote survival of neurons. Because of their survival-promoting and antiapoptotic effects, neurotrophic factors are in theory outstanding candidates for therapy for disorders characterized by premature death of neurons as occurs in amyotrophic lateral sclerosis (ALS) and other degenerative motor neuron disorders. Knockout mice lacking receptors for ciliary neurotrophic factor (CNTF) or BDNF show loss of motor neurons, and experimental motor neuron death can be rescued by treatment with various neurotrophic factors, including CNTF, BDNF, and vascular endothelial growth factor (VEGF). However, in phase 3 clinical trials, growth factors were ineffective in human ALS. The growth factor glial-derived neurotrophic factor (GDNF) is important for survival of dopaminergic neurons. Direct infusions of GDNF showed initial promise in Parkinson's disease (PD), but the benefits were not replicated in a larger clinical trial.

Table 366-3 Neurotrophic Factors
Neurotrophin family
Nerve growth factor

- Brain-derived neurotrophic factor
- Neurotrophin-3
- Neurotrophin-4
- Neurotrophin-6

Cytokine family

- Ciliary neurotrophic factor
- Leukemia inhibitory factor
- Interleukin 6
- Cardiotrophin-1

Transforming growth factor β family

- Glial-derived neurotrophic family
- Neurturin
- Persephin

Fibroblast growth factor family

Hepatocyte growth factor

Insulin-like growth factor (IGF) family

- IGF-1
- IGF-2

STEM CELLS AND TRANSPLANTATION

The nervous system is traditionally considered to be a nonmitotic organ, particularly with respect to neurons. These concepts have been challenged by the finding that neural progenitor or stem cells exist in the adult CNS that are capable of differentiation, migration over long distances, and extensive axonal arborization and synapse formation with appropriate targets. These capabilities also indicate that the repertoire of factors required for growth, survival, differentiation, and migration of these cells exists in the mature nervous system. In rodents, neural stem cells, defined as progenitor cells capable of differentiating into mature cells of neural or glial lineage, have been experimentally propagated from fetal CNS and neuroectodermal tissues and also from adult germinal matrix and ependyma regions. Human fetal CNS tissue is also capable of differentiation into cells with neuronal, astrocyte, and oligodendrocyte morphology when cultured in the presence of growth factors.

Once the repertoire of signals required for cell type specification is better understood, differentiation into specific neural or glial subpopulations could be directed in vitro; such cells also could be engineered to express therapeutic molecules. Another promising approach is to utilize growth factors such as BDNF to stimulate endogenous stem cells to proliferate and migrate to areas of neuronal damage. Administration of epidermal growth factor with fibroblast growth factor replenished up to 50% of hippocampal CA1 neurons a month after global ischemia in rats. The new neurons made connections and improved performance in a memory task.

A major advance has been the development of induced pluripotent stem cells. Using this technique, adult somatic cells such as skin fibroblasts are treated with four pluripotency factors (SOX2, KLF4, cMYC, and Oct4), and this generates induced pluripotent stem cells (iPSCs). These adult-derived stem cells sidestep the ethical issues of utilizing stem cells derived from human embryos. The development of these cells has tremendous promise for both studying disease mechanisms and testing therapeutics. There is no consensus on the best way to generate and differentiate iPSCs; however, techniques to avoid using viral vectors and the use of Cre-lox systems to remove reprogramming factors result in a better match of gene expression profiles with those of embryonic stem cells. Thus far, iPSC cells have been made from patients with all the major human neurodegenerative diseases, and studies utilizing them are under way.

Although stem cells hold tremendous promise for the treatment of debilitating neurologic diseases such as Parkinson's disease and spinal cord injury, it should be emphasized that medical application is in its infancy. Major obstacles are the generation of position- and neurotransmitter-defined subtypes of neurons and their isolation as pure populations of the desired cells. This is crucial to avoid persistence of undifferentiated embryonic stem (ES) cells, which can generate tumors. The establishment of appropriate neural connections and afferent control is also critical. For instance, human ES motor neurons will have to be introduced at multiple segments in the neuraxis, and then their axons will have to regenerate from the spinal cord to distal musculature.

Experimental transplantation of human fetal dopaminergic neurons in patients with Parkinson's disease has shown that these transplanted cells can survive within the host striatum; however, some patients

developed disabling dyskinesias, and this approach is no longer in clinical development. Human ES cells can be differentiated into dopaminergic neurons, which reverse symptoms of Parkinson's disease in experimental animal models. Studies of transplantation for patients with Huntington's disease have reported encouraging, although very preliminary, results. Oligodendrocyte precursor cells transplanted into mice with a dysmyelinating disorder effectively migrated in the new environment, interacted with axons, and mediated myelination; such experiments raise hope that similar transplantation strategies may be feasible in human disorders of myelin such as MS. The promise of stem cells for treatment of both neurodegenerative diseases and neural injury is great, but development has been slowed by unresolved concerns over safety (including the theoretical risk of malignant transformation of transplanted cells), ethics (particularly with respect to use of fetal tissue), and efficacy.

In developing brain, the extracellular matrix provides stimulatory and inhibitory signals that promote neuronal migration, neurite outgrowth, and axonal extension. After neuronal damage, reexpression of inhibitory molecules such as chondroitin sulfate proteoglycans may prevent tissue regeneration. Chondroitinase degraded these inhibitory molecules and enhanced axonal regeneration and motor recovery in a rat model of spinal cord injury. Several myelin proteins, specifically Nogo, oligodendrocyte myelin glycoprotein (OMGP), and myelin-associated glycoprotein (MAG), also may interfere with axon regeneration. Sialidase, which cleaves one class of receptors for MAG, enhances axonal outgrowth. Antibodies against Nogo promote regeneration after experimental focal ischemia or spinal cord injury. Nogo, OMGP, and MAG all bind to the same neural receptor, the Nogo receptor, which mediates its inhibitory function via the p75 neurotrophin receptor signaling.

CELL DEATH: EXCITOTOXICITY AND APOPTOSIS

Excitotoxicity refers to neuronal cell death caused by activation of excitatory amino acid receptors **(Fig. 366-3)**. Compelling evidence for a role of excitotoxicity, especially in ischemic neuronal injury, is derived from experiments in animal models. Experimental models of stroke are associated with increased extracellular concentrations of the excitatory amino acid neurotransmitter glutamate, and neuronal damage is attenuated by denervation of glutamate-containing neurons or the administration of glutamate receptor antagonists. The distribution of cells sensitive to ischemia corresponds closely with that of *N*-methyl-d-aspartate (NMDA) receptors (except for cerebellar Purkinje cells, which are vulnerable to hypoxemia-ischemia but lack NMDA receptors), and competitive and noncompetitive NMDA antagonists are effective in preventing focal ischemia. In global cerebral ischemia, non-NMDA receptors [kainic acid and α-amino-3-hydroxyl-5-methyl-4-isoxazole-propionate (AMPA)] are activated, and antagonists to these receptors are protective. Experimental brain damage induced by hypoglycemia also is attenuated by NMDA antagonists.

Figure 366-3

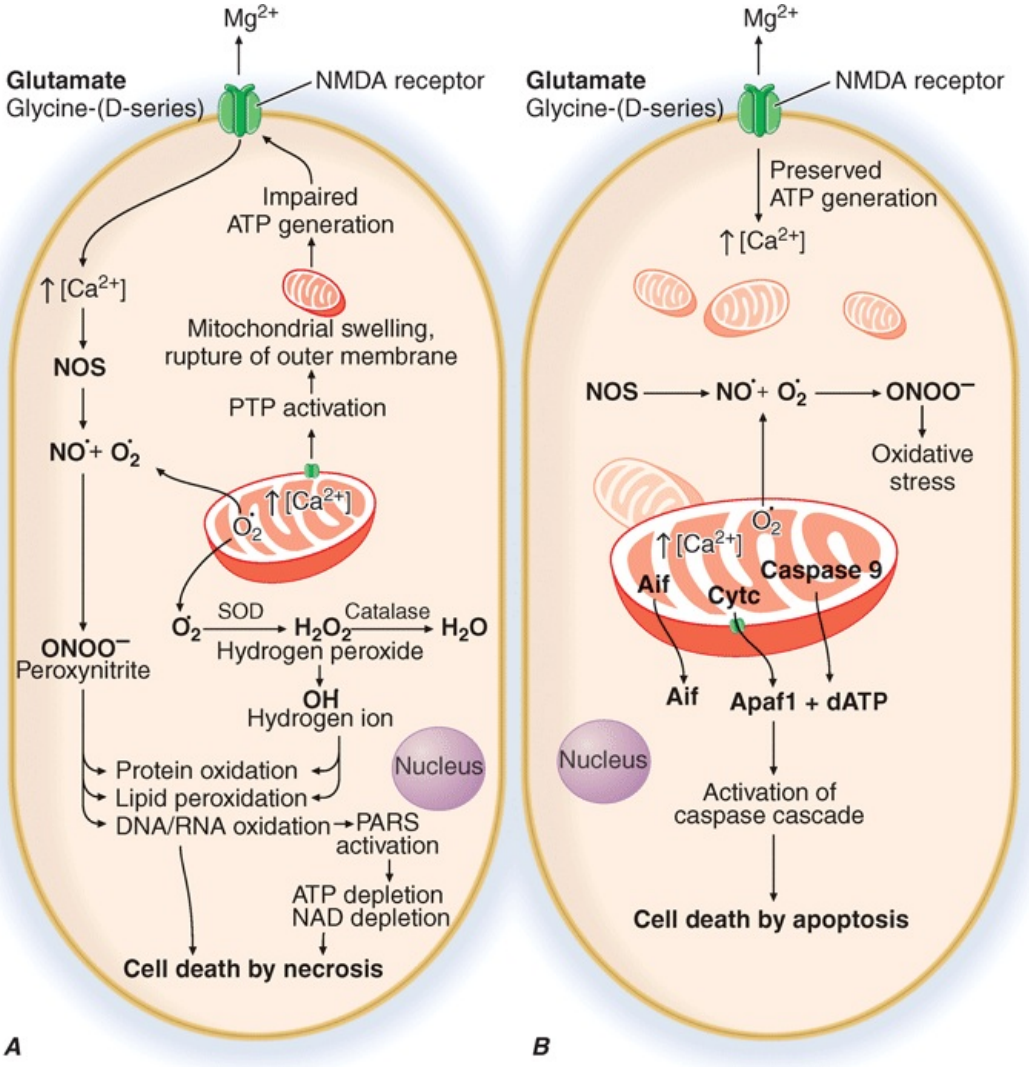

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Involvement of mitochondria in cell death. A severe excitotoxic insult (***A***) results in cell death by necrosis, whereas a mild excitotoxic insult (***B***) results in apoptosis. After a severe insult (such as ischemia), there is a large increase in glutamate activation of NMDA receptors, an increase in intracellular Ca^{2+} concentrations, activation of nitric oxide synthase (NOS), and increased mitochondrial Ca^{2+} and superoxide generation followed by the formation of $ONOO^-$. This sequence results in damage to cellular macromolecules including DNA, leading to activation of poly-ADP-ribose polymerase (PARS). Both mitochondrial accumulation of Ca^{2+} and oxidative damage lead to activation of the permeability transition pore (PTP) that is linked to excitotoxic cell death. A mild excitotoxic insult can occur due either to an abnormality in an excitotoxicity amino acid receptor, allowing more Ca^{2+} flux, or to impaired functioning of other ionic channels or of energy production, which may allow the voltage-dependent NMDA receptor to be activated by ambient concentrations of glutamate. This event can then lead to increased mitochondrial Ca^{2+} and free radical production, yet relatively preserved ATP generation. The mitochondria may then release cytochrome c (Cytc), caspase 9, apoptosis-inducing factor (Aif), and perhaps other mediators that lead to apoptosis. The precise role of the PTP in this mode of cell death is still being clarified, but there does appear to be involvement of the adenine nucleotide transporter that is a key component of the PTP.

Excitotoxicity is not a single event but rather a cascade of cell injury. Excitotoxicity causes influx of calcium into cells, and much of the calcium is sequestered in mitochondria rather than in the cytoplasm. Increased cytoplasmic calcium causes metabolic dysfunction and free radical generation; activates protein kinases, phospholipases, nitric oxide synthase, proteases, and endonucleases; and inhibits protein synthesis. Activation of nitric oxide synthase generates nitric oxide ($NO_•$), which can react with superoxide ($O_•2$) to generate peroxynitrite ($ONOO^-$), which may play a direct role in neuronal injury. Another critical pathway is activation of poly-ADP-ribose polymerase, which occurs in response to free radical–mediated DNA damage. Experimentally, mice with knockout mutations of neuronal nitric oxide synthase or poly-ADP-ribose polymerase, or those which overexpress superoxide dismutase, are resistant to focal ischemia.

Another aspect of excitotoxicity is that it has been demonstrated that stimulation of extrasynaptic NMDA receptors mediates cell death, whereas stimulation of synaptic receptors is protective. This has been shown to play a role in excitotoxicity in transgenic mouse models of Huntington's disease, in which the use of low-dose memantine to selectively block the extrasynaptic receptors is beneficial.

Although excitotoxicity is clearly implicated in the pathogenesis of cell death in stroke, to date treatment with NMDA antagonists has not proved clinically useful. Transient receptor potentials (TRPs) are calcium channels that are activated by oxidative stress in parallel with excitotoxic signal pathways. In addition, glutamate-independent pathways of calcium influx via acid-sensing ion channels have been identified. These channels transport calcium in the setting of acidosis and substrate depletion, and pharmacologic blockade of these channels markedly attenuates stroke injury. These channels offer a potential new therapeutic target for stroke.

Apoptosis, or programmed cell death, plays an important role in both physiologic and pathologic conditions. During embryogenesis, apoptotic pathways operate to destroy neurons that fail to differentiate appropriately or reach their intended targets. There is mounting evidence for an increased rate of apoptotic cell death in a variety of acute and chronic neurologic diseases. Apoptosis is characterized by neuronal shrinkage, chromatin condensation, and DNA fragmentation, whereas necrotic cell death is associated with cytoplasmic and mitochondrial swelling followed by dissolution of the cell membrane. Apoptotic death and necrotic cell death can coexist or be sequential events, depending on the severity of the initiating insult. Cellular energy reserves appear to have an important role in these two forms of cell death, with apoptosis favored under conditions in which ATP levels are preserved. Evidence of DNA fragmentation has been found in a number of degenerative neurologic disorders, including Alzheimer's disease, Huntington's disease, and ALS. The best characterized genetic neurologic disorder related to apoptosis is infantile spinal muscular atrophy (Werdnig-Hoffmann disease), in which two genes thought to be involved in the apoptosis pathways are causative.

Mitochondria are essential in controlling specific apoptosis pathways. The redistribution of cytochrome c, as well as apoptosis-inducing factor (AIF), from mitochondria during apoptosis leads to the activation of a cascade of intracellular proteases known as *caspases*. Caspase-independent apoptosis occurs after DNA damage, activation of poly-ADP-ribose polymerase, and translocation of AIF into the nucleus. Redistribution of cytochrome c is prevented by overproduction of the apoptotic protein BCL2 and is promoted by the proapoptotic protein BAX. These pathways may be triggered by activation of a large pore in the mitochondrial inner membrane known as the *permeability transition pore*, although in other circumstances they occur independently. Recent studies suggest that blocking the mitochondrial pore reduces both hypoglycemic and ischemic cell death. Mice deficient in cyclophilin D, a key protein involved in opening the permeability transition pore, are resistant to necrosis produced by focal cerebral ischemia.

PROTEIN AGGREGATION AND NEURODEGENERATION

The possibility that protein aggregation plays a role in the pathogenesis of neurodegenerative diseases is a major focus of current research. Protein aggregation is a major histopathologic hallmark of neurodegenerative diseases. Deposition of ß-amyloid is strongly implicated in the pathogenesis of Alzheimer's disease. Genetic mutations in familial Alzheimer's disease cause increased production of ß-amyloid with 42 amino acids, which has an increased propensity to aggregate, compared with ß-amyloid with 40 amino acids. Mutations in genes encoding the MAPT lead to altered splicing of tau and the production of neurofibrillary tangles in frontotemporal dementia and progressive supranuclear palsy. Familial Parkinson's disease is associated with mutations in *leucine-rich repeat kinase 2 (LRRK2),* α*-synuclein, parkin, PINK1,* and *DJ-1*. PINK1 is a mitochondrial kinase (see below), and DJ-1 is a protein involved in protection from oxidative stress. Parkin, which causes autosomal recessive early-onset Parkinson's disease, is a ubiquitin ligase. The characteristic histopathologic feature of Parkinson's disease is the Lewy body, an eosinophilic cytoplasmic inclusion that contains both neurofilaments and α-synuclein. Huntington's disease and cerebellar degenerations are associated with expansions of polyglutamine repeats in proteins, which aggregate to produce neuronal intranuclear inclusions. Familial ALS is associated with superoxide dismutase mutations and cytoplasmic inclusions that contain superoxide dismutase. An important finding was the discovery that the ubiquinated inclusions observed in most cases of ALS and the most common form of frontotemporal dementia are composed of TAR DNA binding protein 43 (TDP-43). Subsequently, mutations in the TDP-43 gene and in the fused in sarcoma gene (FUS) were found in familial ALS. These two proteins are involved in transcription regulation as well as RNA metabolism. In autosomal dominant neurohypophyseal diabetes insipidus, mutations in vasopressin result in abnormal protein processing, accumulation in the endoplasmic reticulum, and cell death.

Another key mechanism linked to cell death is mitochondrial dynamics, which refer to the processes involved in movement of mitochondria, as well as in mitochondrial fission and fusion, which play a critical role in mitochondrial turnover and in replenishment of damaged mitochondria. Mitochondrial dysfunction is strongly linked to the pathogenesis of a number of neurodegenerative diseases, such as Friedreich's ataxia, which is caused by mutations in an iron-binding protein that plays an important role

in transferring iron to iron-sulfur clusters in aconitase and complex I and II of the electron transport chain. Mitochondrial fission is dependent on the dynamin-related proteins (Drp1), which bind to its receptor Fis, whereas mitofuscins 1 and 2 (MF 1/ 2) and optic atrophy protein 1(Opa1) are responsible for fusion of the outer and inner mitochondrial membrane, respectively. Mutations in Mfn2 cause Charcot-Marie-Tooth neuropathy type 2A, and mutations in Opa1 cause autosomal dominant optic atrophy. Both β-amyloid and mutant huntingtin protein induce mitochondrial fragmentation and neuronal cell death associated with increased activity of Drp1. In addition, mutations in genes causing autosomal recessive Parkinson's disease, parkin and PINK1, cause abnormal mitochondrial morphology and result in impairment of the ability of the cell to remove damaged mitochondria by autophagy.

The current major scientific question is whether protein aggregates contribute to neuronal death or whether they are merely secondary bystanders. A major focus in all the neurodegenerative diseases is now on small protein aggregates termed *oligomers*. These aggregates may be the toxic species of β-amyloid, α-synuclein, and proteins with expanded polyglutamines such as those that are associated with Huntington's disease. Protein aggregates are usually ubiquinated, which targets them for degradation by the 26S component of the proteasome. An inability to degrade protein aggregates could lead to cellular dysfunction, impaired axonal transport, and cell death by apoptotic mechanisms.

Autophagy is the degradation of cystolic components in lysosomes. There is increasing evidence that autophagy plays an important role in degradation of protein aggregates in the neurodegenerative diseases, and it is impaired in Alzheimer's disease (AD), Parkinson's disease, and Huntington's disease (HD). Autophagy is particularly important to the health of neurons, and failure of autophagy contributes to cell death. In Huntington's disease, a failure of cargo recognition occurs, contributing to protein aggregates and cell death. Rapamycin, which induces autophagy, exerts beneficial therapeutic effects in transgenic mouse models of AD, PD, and HD.

In experimental models of Huntington's disease and cerebellar degeneration, protein aggregates are not well correlated with neuronal death and may be protective. A substantial body of evidence suggests that the mutant proteins with polyglutamine expansions in these diseases bind to transcription factors and that this contributes to disease pathogenesis. In Huntington's disease there is dysfunction of the transcriptional co-regulator PGC-1α, a key regulator of mitochondrial biogenesis. There is evidence that impaired function of PGC-1α is also important in both Parkinson's disease and Alzheimer's disease, making it an attractive target for treatments. Agents that upregulate gene transcription are neuroprotective in animal models of these diseases. A number of compounds have been developed to block β-amyloid production and/or aggregation, and these agents are being studied in early clinical trials in humans. Another approach under investigation is immunotherapy with antibodies that bind β-amyloid, tau, or α-synuclein.

Another emerging theme is the role of chronic inflammation, and in particular of activated microglia and innate immunity (Chap 314), in the pathogenesis of many neurodegenerative diseases. Activation of Toll-like receptors (TLR) in response to pattern-recognition signals from damaged or aging cells, including those mediated by heat shock proteins or aggregated proteins, can trigger or amplify pro-inflammatory responses. Familial frontotemporal degeneration (Chap. 371) is caused by mutations in the gene encoding progranulin, a growth factor that regulates inflammation via binding to tumor necrosis factor (TNF) receptors.

SYSTEMS NEUROSCIENCE

Systems neuroscience refers to study of the functions of neurocircuits and the way they relate to brain function, behavior, motor activity, and cognition. Brain imaging techniques, primarily functional MRI (fMRI) and positron emission tomography (PET), have made it possible to investigate cognitive processes such as perception, making judgments, paying attention, and thinking. This has allowed insights into how networks of neurons operate to produce behavior. Many of these studies at present are based on determining the connectivity of neural circuits and how they operate and how this can be modeled to produce improved understanding of physiologic processes. fMRI uses contrast mechanisms related to physiologic changes in tissue, and brain perfusion can be studied by observing the time course of changes in brain water signal as a bolus of injected paramagnetic gadolinium contrast moves through the brain. More recently, to study intrinsic contrast-related local changes in blood oxygenation with brain activity, blood-oxygen-level-dependent (BOLD) contrast has been used to provide a rapid noninvasive approach for functional assessment. These techniques have been used reliably in both behavioral and cognitive sciences. One example is the use of fMRI to demonstrate mirror neuron systems, imitative pathways activated when observing actions of others **(Fig. 366-4)**. Mirror neurons are thought to be important for social conditioning and for many forms of learning, and abnormalities in mirror neurons may underlie some autism disorders. Data also suggest that enhancement of mirror neuron pathways might have potential for rehabilitation after stroke. Other examples of the use of fMRI include the study of memory. Recent studies have shown that not only is hippocampal activity correlated with declarative memory consolidation, it also involves activation in the ventral medial prefrontal cortex. Consolidation of memory over time results in decreased activity of the hippocampus and progressively stronger activation in the ventral medial prefrontal region associated

with retrieval of consolidated memories. fMRI also has been utilized to identify sequences of brain activation involved in normal movements and alterations in their activation associated with both injury and recovery and to plan neurosurgical operations. Diffusion tensor imaging is a recently developed MRI technique that can measure macroscopic axonal organization in nervous system tissues; it appears to be useful in assessing myelin and axonal injuries as well as brain development. Advances in understanding neural processing have led to the development of the ability to demonstrate that humans have on-line voluntary control of human temporal lobe neurons.

Figure 366-4

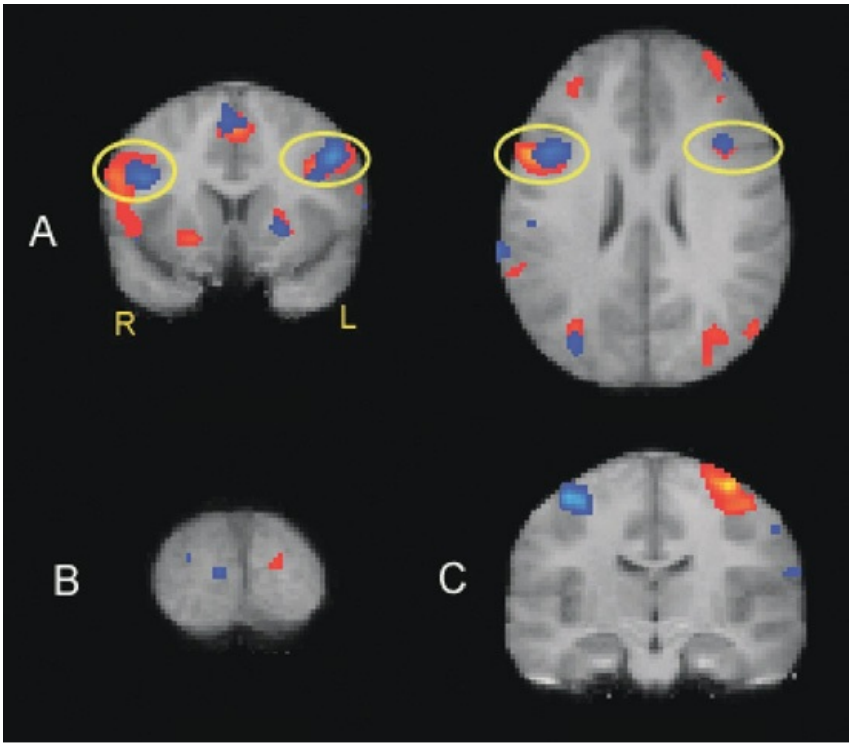

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mirror neuron systems are bilaterally activated during imitation. ***A.*** Bilateral activations (circled in yellow) in inferior frontal mirror neuron areas during imitation, as measured by BOLD fMRI signal changes. In red, activation during right hand imitation. In blue, activation during left hand imitation. ***B.*** In contrast, there is lateralized (contralateral) primary visual activation of the primary visual cortex for imitated actions presented to the right visual field (in red, left visual cortex) and to the left visual field (in blue, right visual cortex). ***C.*** Lateralized primary motor activation for hand actions imitated with the right hand (in red, left motor cortex) and with the left hand (in blue, right motor cortex). *(From L Aziz-Zadeh et al: J Neurosci 26:2964, 2006.)*

A further advance that has far-reaching implications for the development of novel interventions for neurologic, including behavioral, conditions has been the development of deep-brain stimulation as a highly effective therapeutic intervention for treating excessively firing neurons in the subthalamic nucleus of patients with Parkinson's disease and the precingulate cortex in patients with depression.

FURTHER READINGS

Amor S et al: Inflammation in neurodegenerative diseases. Immunology 129:154, 2010[PMID: 20561356] [Full Text]

Baranzini SE et al: Genome, epigenome, and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature 464:1391, 2010

Bilguvar K et al: Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature 467:207, 2010[PMID: 20729831] [Full Text]

Catteneo L, Rizzolatti G: The mirror neuron system. Arch Neurol 66:557, 2009

Cerf M et al: Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci USA 106:19096, 2009

Chong SYC, Chan JR: Tapping into the glial reservoir: Cells committed to remaining uncommitted. J Cell Biol 188:305, 2010[PMID: 20142420] [Full Text]

Doyle KP et al: Mechanisms of ischemic brain damage. Neuropharmacology 55:310, 2008[PMID: 18308346] [Full Text]

Emery B et al: Myelin gene regulatory factor is a critical transcriptional regulator required for CNS myelination. Cell 138:172, 2009[PMID: 19596243] [Full Text]

Ercan-sencicek AG et al: L-histidine decarboxylase and Tourette's syndrome. N Engl J Med 362:1901, 2010[PMID: 20445167] [Full Text]

Kriegstein A, Alvarez-buylla A: The glial nature of embryonic and adult neural stem cells. Annu Rev Neurosci 32:149, 2009[PMID: 19555289] [Full Text]

Lukong KE et al: RNA-binding proteins in human genetic disease. Trends Genet. 24:416, 2008[PMID: 18597886] [Full Text]

Lupski JR et al: Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med 362:1181, 2010[PMID: 20220177] [Full Text]

Mackenzie IRA et al: TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. Lancet Neurol 9:995, 2010[PMID: 20864052] [Full Text]

Mehler MF: Epigenetics and the nervous system. Ann Neurol 64:602, 2008[PMID: 19107999] [Full Text]

Ramocki MB et al: Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome. Ann Neurol 66:771, 2009[PMID: 20035514] [Full Text]

Rosenzweig A: Illuminating the potential of pluripotent stem cells. N Engl J Med 363:1471, 2010[PMID: 20925550] [Full Text]

Ryan DP et al: Mutations in potassium channel Kir2.6 cause susceptibility to thyrotoxic hypokalemic periodic paralysis. Cell 140:88, 2010[PMID: 20074522] [Full Text]

Tang W et al: The growth factor progranulin binds to TNF receptors and is therapeutic against inflammatory arthritis in mice. Science epub 10 March 2010

Wong E, Cuervo AM: Autophagy gone awry in neurodegenerative diseases. Nat Neurosci 13:805, 2010[PMID: 20581817] [Full Text]

Zoghbi HY, Warren ST: Neurogenetics: Advancing the "next-generation" of brain research. Neuron 68:165, 2010[PMID: 20955921] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT WITH NEUROLOGIC DISEASE: INTRODUCTION

Neurologic diseases are common and costly. According to recent estimates by the World Health Organization, neurologic disorders affect over 1 billion people worldwide (**Table 367-1**), constitute 6.3% of the global burden of disease, and cause 12% of global deaths. Most patients with neurologic symptoms seek care from internists and other generalists rather than from neurologists. Because therapies now exist for many neurologic disorders, a skillful approach to diagnosis is essential. Errors commonly result from an overreliance on costly neuroimaging procedures and laboratory tests, which, while useful, do not substitute for an adequate history and examination. The proper approach to the patient with a neurologic illness begins with the patient and focuses the clinical problem first in anatomic and then in pathophysiologic terms; only then should a specific diagnosis be entertained. This method ensures that technology is judiciously applied, a correct diagnosis is established in an efficient manner, and treatment is promptly initiated.

Table 367-1 Prevalence of Neurologic and Psychiatric Diseases Worldwide

Disorder	Patients, Millions
Nutritional disorders and neuropathies	352
Migraine	326
Trauma	170
Cerebrovascular diseases	61
Epilepsy	50
Dementia	24
Neurologic infections	18

Source: World Health Organization estimates, 2002–2005.

THE NEUROLOGIC METHOD

Locate the Lesion(S)

The first priority is to identify the region of the nervous system that is likely to be responsible for the symptoms. Can the disorder be mapped to one specific location, is it multifocal, or is a diffuse process present? Are the symptoms restricted to the nervous system, or do they arise in the context of a systemic illness? Is the problem in the central nervous system (CNS), the peripheral nervous system (PNS), or both? If in the CNS, is the cerebral cortex, basal ganglia, brainstem, cerebellum, or spinal cord responsible? Are the pain-sensitive meninges involved? If in the PNS, could the disorder be located in peripheral nerves and, if so, are motor or sensory nerves primarily affected, or is a lesion in the neuromuscular junction or muscle more likely?

The first clues to defining the anatomic area of involvement appear in the history, and the examination is then directed to confirm or rule out these impressions and to clarify uncertainties. A more detailed examination of a particular region of the CNS or PNS is often indicated. For example, the examination of a patient who presents with a history of ascending paresthesias and weakness should be directed toward deciding, among other things, if the location of the lesion is in the spinal cord or peripheral nerves. Focal back pain, a spinal cord sensory level, and incontinence suggest a spinal cord origin, whereas a stocking-glove pattern of sensory loss suggests peripheral nerve disease; areflexia usually indicates peripheral neuropathy but may also be present with spinal shock in acute spinal cord disorders.

Deciding "where the lesion is" accomplishes the task of limiting the possible etiologies to a manageable, finite number. In addition, this strategy safeguards against making serious errors. Symptoms of recurrent vertigo, diplopia, and nystagmus should not trigger "multiple sclerosis" as an answer (etiology) but "brainstem" or "pons" (location); then a diagnosis of brainstem arteriovenous malformation will not be missed for lack of consideration. Similarly, the combination of optic neuritis and spastic ataxic paraparesis should initially suggest optic nerve and spinal cord disease; multiple sclerosis (MS), CNS syphilis, and vitamin B12 deficiency are treatable disorders that can produce this syndrome. Once the question, "Where is the lesion?" is answered, then the question, "What is the

lesion?" can be addressed.

Define the Pathophysiology

Clues to the pathophysiology of the disease process may also be present in the history. Primary neuronal (gray matter) disorders may present as early cognitive disturbances, movement disorders, or seizures, whereas white matter involvement produces predominantly "long tract" disorders of motor, sensory, visual, and cerebellar pathways. Progressive and symmetric symptoms often have a metabolic or degenerative origin; in such cases lesions are usually not sharply circumscribed. Thus, a patient with paraparesis and a clear spinal cord sensory level is unlikely to have vitamin B12 deficiency as the explanation. A Lhermitte symptom (electric shock–like sensations evoked by neck flexion) is due to ectopic impulse generation in white matter pathways and occurs with demyelination in the cervical spinal cord; among many possible causes, this symptom may indicate MS in a young adult or compressive cervical spondylosis in an older person. Symptoms that worsen after exposure to heat or exercise may indicate conduction block in demyelinated axons, as occurs in MS. A patient with recurrent episodes of diplopia and dysarthria associated with exercise or fatigue may have a disorder of neuromuscular transmission such as myasthenia gravis. Slowly advancing visual scotoma with luminous edges, termed *fortification spectra*, indicates spreading cortical depression, typically with migraine.

THE NEUROLOGIC HISTORY

Attention to the description of the symptoms experienced by the patient and substantiated by family members and others often permits an accurate localization and determination of the probable cause of the complaints, even before the neurologic examination is performed. The history also helps to bring a focus to the neurologic examination that follows. Each complaint should be pursued as far as possible to elucidate the location of the lesion, the likely underlying pathophysiology, and potential etiologies. For example, a patient complains of weakness of the right arm. What are the associated features? Does the patient have difficulty with brushing hair or reaching upward (proximal) or buttoning buttons or opening a twist-top bottle (distal)? Negative associations may also be crucial. A patient with a right hemiparesis without a language deficit likely has a lesion (internal capsule, brainstem, or spinal cord) different from that of a patient with a right hemiparesis and aphasia (left hemisphere). Other pertinent features of the history include the following:

1. *Temporal course of the illness* . It is important to determine the precise time of appearance and rate of progression of the symptoms experienced by the patient. The rapid onset of a neurologic complaint, occurring within seconds or minutes, usually indicates a vascular event, a seizure, or migraine. The onset of sensory symptoms located in one extremity that spread over a few seconds to adjacent portions of that extremity and then to the other regions of the body suggests a seizure. A more gradual onset and less well-localized symptoms point to the possibility of a transient ischemic attack (TIA). A similar but slower temporal march of symptoms accompanied by headache, nausea, or visual disturbance suggests migraine. The presence of "positive" sensory symptoms (e.g., tingling or sensations that are difficult to describe) or involuntary motor movements suggests a seizure; in contrast, transient loss of function (negative symptoms) suggests a TIA. A stuttering onset where symptoms appear, stabilize, and then progress over hours or days also suggests cerebrovascular disease; an additional history of transient remission or regression indicates that the process is more likely due to ischemia rather than hemorrhage. A gradual evolution of symptoms over hours or days suggests a toxic, metabolic, infectious, or inflammatory process. Progressing symptoms associated with the systemic manifestations of fever, stiff neck, and altered level of consciousness imply an infectious process. Relapsing and remitting symptoms involving different levels of the nervous system suggest MS or other inflammatory processes. Slowly progressive symptoms without remissions are characteristic of neurodegenerative disorders, chronic infections, gradual intoxications, and neoplasms.
2. *Patients' descriptions of the complaint* . The same words often mean different things to different patients. "Dizziness" may imply impending syncope, a sense of disequilibrium, or true spinning vertigo. "Numbness" may mean a complete loss of feeling, a positive sensation such as tingling, or even weakness. "Blurred vision" may be used to describe unilateral visual loss, as in transient monocular blindness, or diplopia. The interpretation of the true meaning of the words used by patients to describe symptoms obviously becomes even more complex when there are differences in primary languages and cultures.
3. *Corroboration of the history by others* . It is almost always helpful to obtain additional information from family, friends, or other observers to corroborate or expand the patient's description. Memory loss, aphasia, loss of insight, intoxication, and other factors may impair the patient's capacity to communicate normally with the examiner or prevent openness about factors that have contributed to the illness. Episodes of loss of consciousness necessitate that details be sought from observers to ascertain precisely what has happened during the event.
4. *Family history* . Many neurologic disorders have an underlying genetic component. The presence of a Mendelian disorder, such as Huntington's disease or Charcot-Marie-Tooth neuropathy, is often obvious if family data are available. More detailed questions about family history are often necessary in polygenic disorders such as MS, migraine, and many types of epilepsy. It is important to elicit family history about all illnesses, in addition to neurologic and psychiatric

disorders. A familial propensity to hypertension or heart disease is relevant in a patient who presents with a stroke. There are numerous inherited neurologic diseases that are associated with multisystem manifestations that may provide clues to the correct diagnosis (e.g., neurofibromatosis, Wilson's disease, neuro-ophthalmic syndromes).

5. *Medical illnesses*. Many neurologic diseases occur in the context of systemic disorders. Diabetes mellitus, hypertension, and abnormalities of blood lipids predispose to cerebrovascular disease. A solitary mass lesion in the brain may be an abscess in a patient with valvular heart disease, a primary hemorrhage in a patient with a coagulopathy, a lymphoma or toxoplasmosis in a patient with AIDS, or a metastasis in a patient with underlying cancer. Patients with malignancy may also present with a neurologic paraneoplastic syndrome (Chap. 101) or complications from chemotherapy or radiotherapy. Marfan's syndrome and related collagen disorders predispose to dissection of the cranial arteries and aneurysmal subarachnoid hemorrhage; the latter may also occur with polycystic kidney disease. Various neurologic disorders occur with dysthyroid states or other endocrinopathies. It is especially important to look for the presence of systemic diseases in patients with peripheral neuropathy. Most patients with coma in a hospital setting have a metabolic, toxic, or infectious cause.
6. *Drug use and abuse and toxin exposure* . It is essential to inquire about the history of drug use, both prescribed and illicit. Sedatives, antidepressants, and other psychoactive medications are frequently associated with acute confusional states in the elderly. Aminoglycoside antibiotics may exacerbate symptoms of weakness in patients with disorders of neuromuscular transmission, such as myasthenia gravis, and may cause dizziness secondary to ototoxicity. Vincristine and other antineoplastic drugs can cause peripheral neuropathy, and immunosuppressive agents such as cyclosporine can produce encephalopathy. Excessive vitamin ingestion can lead to disease; for example vitamin A and pseudotumor cerebri, or pyridoxine and peripheral neuropathy. Many patients are unaware that over-the-counter sleeping pills, cold preparations, and diet pills are actually drugs. Alcohol, the most prevalent neurotoxin, is often not recognized as such by patients, and other drugs of abuse such as cocaine and heroin can cause a wide range of neurologic abnormalities. A history of environmental or industrial exposure to neurotoxins may provide an essential clue; consultation with the patient's coworkers or employer may be required.
7. *Formulating an impression of the patient* . Use the opportunity while taking the history to form an impression of the patient. Is the information forthcoming, or does it take a circuitous course? Is there evidence of anxiety, depression, hypochondriasis? Are there any clues to defects in language, memory, insight, or inappropriate behavior? The neurologic assessment begins as soon as the patient comes into the room and the first introduction is made.

THE NEUROLOGIC EXAMINATION

The neurologic examination is challenging and complex; it has many components and includes a number of skills that can be mastered only through repeated use of the same techniques on a large number of individuals with and without neurologic disease. Mastery of the complete neurologic examination is usually important only for physicians in neurology and associated specialties. However, knowledge of the basics of the examination, especially those components that are effective in screening for neurologic dysfunction, is essential for all clinicians, especially generalists.

There is no single, universally accepted sequence of the examination that must be followed, but most clinicians begin with assessment of mental status followed by the cranial nerves, motor system, sensory system, coordination, and gait. Whether the examination is basic or comprehensive, it is essential that it be performed in an orderly and systematic fashion to avoid errors and serious omissions. Thus, the best way to learn and gain expertise in the examination is to choose one's own approach and practice it frequently and do it in the same exact sequence each time.

The detailed description of the neurologic examination that follows describes the more commonly used parts of the examination, with a particular emphasis on the components that are considered most helpful for the assessment of common neurologic problems. Each section also includes a brief description of the minimal examination necessary for adequate screening for abnormalities in a patient who has no symptoms suggesting neurologic dysfunction. A screening examination done in this way can be completed in 3–5 min.

Several additional points about the examination are worth noting. First, in recording observations, it is important to describe what is found rather than to apply a poorly defined medical term (e.g., "patient groans to sternal rub" rather than "obtunded"). Second, subtle CNS abnormalities are best detected by carefully comparing a patient's performance on tasks that require simultaneous activation of both cerebral hemispheres (e.g., eliciting a pronator drift of an outstretched arm with the eyes closed; extinction on one side of bilaterally applied light touch, also with eyes closed; or decreased arm swing or a slight asymmetry when walking). Third, if the patient's complaint is brought on by some activity, reproduce the activity in the office. If the complaint is of dizziness when the head is turned in one direction, have the patient do this and also look for associated signs on examination (e.g., nystagmus or dysmetria). If pain occurs after walking two blocks, have the patient leave the office and walk this distance and immediately return, and repeat the relevant parts of the examination. Finally, the use of tests that are individually tailored to the patient's problem can be of value in assessing changes over time. Tests of walking a 7.5-m (25-ft) distance (normal, 5–6 s; note assistance, if any), repetitive finger

or toe tapping (normal, 20–25 taps in 5 s), or handwriting are examples.

Mental Status Examination

- *The bare minimum: During the interview, look for difficulties with communication and determine whether the patient has recall and insight into recent and past events.*

The mental status examination is underway as soon as the physician begins observing and talking with the patient. If the history raises any concern for abnormalities of higher cortical function or if cognitive problems are observed during the interview, then detailed testing of the mental status is indicated. The patient's ability to understand the language used for the examination, cultural background, educational experience, sensory or motor problems, or comorbid conditions need to be factored into the applicability of the tests and interpretation of results.

The Folstein mini-mental status examination (MMSE) (Table 371-5) is a standardized screening examination of cognitive function that is extremely easy to administer and takes <10 min to complete. Using age-adjusted values for defining normal performance, the test is ~85% sensitive and 85% specific for making the diagnosis of dementia that is moderate or severe, especially in educated patients. When there is sufficient time available, the MMSE is one of the best methods for documenting the current mental status of the patient, and this is especially useful as a baseline assessment to which future scores of the MMSE can be compared.

Individual elements of the mental status examination can be subdivided into level of consciousness, orientation, speech and language, memory, fund of information, insight and judgment, abstract thought, and calculations.

Level of consciousness is the patient's relative state of awareness of the self and the environment, and ranges from fully awake to comatose. When the patient is not fully awake, the examiner should describe the responses to the minimum stimulus necessary to elicit a reaction, ranging from verbal commands to a brief, painful stimulus such as a squeeze of the trapezius muscle. Responses that are directed toward the stimulus and signify some degree of intact cerebral function (e.g., opening the eyes and looking at the examiner or reaching to push away a painful stimulus) must be distinguished from reflex responses of a spinal origin (e.g., triple flexion response—flexion at the ankle, knee, and hip in response to a painful stimulus to the foot).

Orientation is tested by asking the person to state his or her name, location, and time (day of the week and date); time is usually the first to be affected in a variety of conditions.

Speech is assessed by observing articulation, rate, rhythm, and prosody (i.e., the changes in pitch and accentuation of syllable and words).

Language is assessed by observing the content of the patient's verbal and written output, response to spoken commands, and ability to read. A typical testing sequence is to ask the patient to name successively more detailed components of clothing, a watch or a pen; repeat the phrase "No ifs, ands, or buts"; follow a three-step, verbal command; write a sentence; and read and respond to a written command.

Memory should be analyzed according to three main time scales: (1) immediate memory is assessed by saying a list of three items and having the patient repeat the list immediately, (2) short-term memory is tested by asking the patient to recall the same three items 5 and 15 min later, and (3) long-term memory is evaluated by determining how well the patient is able to provide a coherent chronologic history of his or her illness or personal events.

Fund of information is assessed by asking questions about major historic or current events, with special attention to educational level and life experiences.

Abnormalities of *insight and judgment* are usually detected during the patient interview; a more detailed assessment can be elicited by asking the patient to describe how he or she would respond to situations having a variety of potential outcomes (e.g., "What would you do if you found a wallet on the sidewalk?").

Abstract thought can be tested by asking the patient to describe similarities between various objects or concepts (e.g., apple and orange, desk and chair, poetry and sculpture) or to list items having the same attributes (e.g., a list of four-legged animals).

Calculation ability is assessed by having the patient carry out a computation that is appropriate to the patient's age and education (e.g., serial subtraction of 7 from 100 or 3 from 20; or word problems involving simple arithmetic).

Cranial Nerve Examination

- *The bare minimum: Check the fundi, visual fields, pupil size and reactivity, extraocular movements, and facial movements.*

The cranial nerves (CN) are best examined in numerical order, except for grouping together CN III, IV, and VI because of their similar function.

CN I (OLFACTORY)

Testing is usually omitted unless there is suspicion for inferior frontal lobe disease (e.g., meningioma). With eyes closed, ask the patient to sniff a mild stimulus such as toothpaste or coffee and identify the odorant.

CN II (OPTIC)

Check visual acuity (with eyeglasses or contact lens correction) using a Snellen chart or similar tool. Test the visual fields by confrontation, i.e., by comparing the patient's visual fields to your own. As a screening test, it is usually sufficient to examine the visual fields of both eyes simultaneously; individual eye fields should be tested if there is any reason to suspect a problem of vision by the history or other elements of the examination, or if the screening test reveals an abnormality. Face the patient at a distance of approximately 0.6–1.0 m (2–3 ft) and place your hands at the periphery of your visual fields in the plane that is equidistant between you and the patient. Instruct the patient to look directly at the center of your face and to indicate when and where he or she sees one of your fingers moving. Beginning with the two inferior quadrants and then the two superior quadrants, move your index finger of the right hand, left hand, or both hands simultaneously and observe whether the patient detects the movements. A single small-amplitude movement of the finger is sufficient for a normal response. Focal perimetry and tangent screen examinations should be used to map out visual field defects fully or to search for subtle abnormalities. Optic fundi should be examined with an ophthalmoscope, and the color, size, and degree of swelling or elevation of the optic disc noted, as well as the color and texture of the retina. The retinal vessels should be checked for size, regularity, arterial-venous nicking at crossing points, hemorrhage, exudates, etc.

CN III, IV, VI (OCULOMOTOR, TROCHLEAR, ABDUCENS)

Describe the size and shape of pupils and reaction to light and accommodation (i.e., as the eyes converge while following your finger as it moves toward the bridge of the nose). To check extraocular movements, ask the patient to keep his or her head still while tracking the movement of the tip of your finger. Move the target slowly in the horizontal and vertical planes; observe any paresis, nystagmus, or abnormalities of smooth pursuit (saccades, oculomotor ataxia, etc.). If necessary, the relative position of the two eyes, both in primary and multidirectional gaze, can be assessed by comparing the reflections of a bright light off both pupils. However, in practice it is typically more useful to determine whether the patient describes diplopia in any direction of gaze; true diplopia should almost always resolve with one eye closed. Horizontal nystagmus is best assessed at 45° and not at extreme lateral gaze (which is uncomfortable for the patient); the target must often be held at the lateral position for at least a few seconds to detect an abnormality.

CN V (TRIGEMINAL)

Examine sensation within the three territories of the branches of the trigeminal nerve (ophthalmic, maxillary, and mandibular) on each side of the face. As with other parts of the sensory examination, testing of two sensory modalities derived from different anatomic pathways (e.g., light touch and temperature) is sufficient for a screening examination. Testing of other modalities, the corneal reflex, and the motor component of CN V (jaw clench—masseter muscle) is indicated when suggested by the history.

CN VII (FACIAL)

Look for facial asymmetry at rest and with spontaneous movements. Test eyebrow elevation, forehead wrinkling, eye closure, smiling, and cheek puff. Look in particular for differences in the lower versus upper facial muscles; weakness of the lower two-thirds of the face with preservation of the upper third suggests an upper motor neuron lesion, whereas weakness of an entire side suggests a lower motor neuron lesion.

CN VIII (VESTIBULOCOCHLEAR)

Check the patient's ability to hear a finger rub or whispered voice with each ear. Further testing for air versus mastoid bone conduction (Rinne) and lateralization of a 512-Hz tuning fork placed at the center of the forehead (Weber) should be done if an abnormality is detected by history or examination. Any suspected problem should be followed up with formal audiometry.

For further discussion of assessing vestibular nerve function in the setting of dizziness, hearing loss, or coma, see Chaps. 21, 30, and 274, respectively.

CN IX, X (GLOSSOPHARYNGEAL, VAGUS)

Observe the position and symmetry of the palate and uvula at rest and with phonation ("aah"). The pharyngeal ("gag") reflex is evaluated by stimulating the posterior pharyngeal wall on each side with a sterile, blunt object (e.g., tongue blade), but the reflex is often absent in normal individuals.

CN XI (SPINAL ACCESSORY)

Check shoulder shrug (trapezius muscle) and head rotation to each side (sternocleidomastoid) against resistance.

CN XII (HYPOGLOSSAL)

Inspect the tongue for atrophy or fasciculations, position with protrusion, and strength when extended against the inner surface of the cheeks on each side.

Motor Examination

- *The bare minimum: Look for muscle atrophy and check extremity tone. Assess upper extremity strength by checking for pronator drift and strength of wrist or finger extensors. Tap the biceps, patellar, and Achilles reflexes. Test for lower extremity strength by having the patient walk normally and on heels and toes.*

The motor examination includes observations of muscle appearance, tone, strength, and reflexes. Although gait is in part a test of motor function, it is usually evaluated separately at the end of the examination.

APPEARANCE

Inspect and palpate muscle groups under good light and with the patient in a comfortable and symmetric position. Check for muscle fasciculations, tenderness, and atrophy or hypertrophy. Involuntary movements may be present at rest (e.g., tics, myoclonus, choreoathetosis), during maintained posture (pill-rolling tremor of Parkinson's disease), or with voluntary movements (intention tremor of cerebellar disease or familial tremor).

TONE

Muscle tone is tested by measuring the resistance to passive movement of a relaxed limb. Patients often have difficulty relaxing during this procedure, so it is useful to distract the patient to minimize active movements. In the upper limbs, tone is assessed by rapid pronation and supination of the forearm and flexion and extension at the wrist. In the lower limbs, while the patient is supine the examiner's hands are placed behind the knees and rapidly raised; with normal tone the ankles drag along the table surface for a variable distance before rising, whereas increased tone results in an immediate lift of the heel off the surface. Decreased tone is most commonly due to lower motor neuron or peripheral nerve disorders. Increased tone may be evident as spasticity (resistance determined by the angle and velocity of motion; corticospinal tract disease), rigidity (similar resistance in all angles of motion; extrapyramidal disease), or paratonia (fluctuating changes in resistance; frontal lobe pathways or normal difficulty in relaxing). Cogwheel rigidity, in which passive motion elicits jerky interruptions in resistance, is seen in parkinsonism.

STRENGTH

Testing for pronator drift is an extremely useful method for screening upper limb weakness. The patient is asked to hold both arms fully extended and parallel to the ground with eyes closed. This position should be maintained for ~10 s; any flexion at the elbow or fingers or pronation of the forearm, especially if asymmetric, is a sign of potential weakness. Muscle strength is further assessed by having the patient exert maximal effort for the particular muscle or muscle group being tested. It is important to isolate the muscles as much as possible, i.e., hold the limb so that only the muscles of interest are active. It is also helpful to palpate accessible muscles as they contract. Grading muscle strength and evaluating the patient's effort is an art that takes time and practice. Muscle strength is traditionally graded using the following scale:

0 = no movement

1 = flicker or trace of contraction but no associated movement at a joint

2 = movement with gravity eliminated

3 = movement against gravity but not against resistance

4– = movement against a mild degree of resistance

4 = movement against moderate resistance

4+ = movement against strong resistance

5 = full power

However, in many cases it is more practical to use the following terms:

Paralysis = no movement

Severe weakness = movement with gravity eliminated

Moderate weakness = movement against gravity but not against mild resistance

Mild weakness = movement against moderate resistance

Full strength

Noting the pattern of weakness is as important as assessing the magnitude of weakness. Unilateral or bilateral weakness of the upper limb extensors and lower limb flexors ("pyramidal weakness") suggests a lesion of the pyramidal tract, bilateral proximal weakness suggests myopathy, and bilateral distal

weakness suggests peripheral neuropathy.

REFLEXES

Muscle Stretch Reflexes

Those that are typically assessed include the biceps (**C5**, C6), brachioradialis (C5, **C6**), and triceps (**C7**, C8) reflexes in the upper limbs and the patellar or quadriceps (**L3**, L4) and Achilles (**S1**, S2) reflexes in the lower limbs. The patient should be relaxed and the muscle positioned midway between full contraction and extension. Reflexes may be enhanced by asking the patient to voluntarily contract other, distant muscle groups (Jendrassik maneuver). For example, upper limb reflexes may be reinforced by voluntary teeth-clenching, and the Achilles reflex by hooking the flexed fingers of the two hands together and attempting to pull them apart. For each reflex tested, the two sides should be tested sequentially, and it is important to determine the smallest stimulus required to elicit a reflex rather than the maximum response. Reflexes are graded according to the following scale:

0 = absent 3 = exaggerated

1 = present but diminished 4 = clonus

2 = normoactive

Cutaneous Reflexes

The plantar reflex is elicited by stroking, with a noxious stimulus such as a tongue blade, the lateral surface of the sole of the foot beginning near the heel and moving across the ball of the foot to the great toe. The normal reflex consists of plantar flexion of the toes. With upper motor neuron lesions above the S1 level of the spinal cord, a paradoxical extension of the toe is observed, associated with fanning and extension of the other toes (termed an *extensor plantar response* , or *Babinski sign*). However, despite its popularity, the reliability and validity of the Babinski sign for identifying upper motor neuron weakness is limited—it is far more useful to rely on tests of tone, strength, stretch reflexes, and coordination. Superficial abdominal reflexes are elicited by gently stroking the abdominal surface near the umbilicus in a diagonal fashion with a sharp object (e.g., the wooden end of a cotton-tipped swab) and observing the movement of the umbilicus. Normally, the umbilicus will pull toward the stimulated quadrant. With upper motor neuron lesions, these reflexes are absent. They are most helpful when there is preservation of the upper (spinal cord level T9) but not lower (T12) abdominal reflexes, indicating a spinal lesion between T9 and T12, or when the response is asymmetric. Other useful cutaneous reflexes include the cremasteric (ipsilateral elevation of the testicle following stroking of the medial thigh; mediated by L1 and L2) and anal (contraction of the anal sphincter when the perianal skin is scratched; mediated by S2, S3, S4) reflexes. It is particularly important to test for these reflexes in any patient with suspected injury to the spinal cord or lumbosacral roots.

Primitive Reflexes

With disease of the frontal lobe pathways, several primitive reflexes not normally present in the adult may appear. The suck response is elicited by lightly touching the center of the lips, and the root response the corner of the lips, with a tongue blade; the patient will move the lips to suck or root in the direction of the stimulus. The grasp reflex is elicited by touching the palm between the thumb and index finger with the examiner's fingers; a positive response is a forced grasp of the examiner's hand. In many instances stroking the back of the hand will lead to its release. The palmomental response is contraction of the mentalis muscle (chin) ipsilateral to a scratch stimulus diagonally applied to the palm.

SENSORY EXAMINATION

- *The bare minimum: Ask whether the patient can feel light touch and the temperature of a cool object in each distal extremity. Check double simultaneous stimulation using light touch on the hands.*

Evaluating sensation is usually the most unreliable part of the examination, because it is subjective and is difficult to quantify. In the compliant and discerning patient, the sensory examination can be extremely helpful for the precise localization of a lesion. With patients who are uncooperative or lack an understanding of the tests, it may be useless. The examination should be focused on the suspected lesion. For example, in spinal cord, spinal root, or peripheral nerve abnormalities, all major sensory modalities should be tested while looking for a pattern consistent with a spinal level and dermatomal or nerve distribution. In patients with lesions at or above the brainstem, screening the primary sensory modalities in the distal extremities along with tests of "cortical" sensation is usually sufficient.

The five primary sensory modalities—light touch, pain, temperature, vibration, and joint position—are tested in each limb. Light touch is assessed by stimulating the skin with single, very gentle touches of the examiner's finger or a wisp of cotton. Pain is tested using a new pin, and temperature is assessed using a metal object (e.g., tuning fork) that has been immersed in cold and warm water. Vibration is tested using a 128-Hz tuning fork applied to the distal phalanx of the great toe or index finger just below the nail bed. By placing a finger on the opposite side of the joint being tested, the examiner compares the patient's threshold of vibration perception with his or her own. For joint position testing,

the examiner grasps the digit or limb laterally and distal to the joint being assessed; small 1- to 2-mm excursions can usually be sensed. The Romberg maneuver is primarily a test of proprioception. The patient is asked to stand with the feet as close together as necessary to maintain balance while the eyes are open, and the eyes are then closed. A loss of balance with the eyes closed is an abnormal response.

"Cortical" sensation is mediated by the parietal lobes and represents an integration of the primary sensory modalities; testing cortical sensation is only meaningful when primary sensation is intact. Double simultaneous stimulation is especially useful as a screening test for cortical function; with the patient's eyes closed, the examiner lightly touches one or both hands and asks the patient to identify the stimuli. With a parietal lobe lesion, the patient may be unable to identify the stimulus on the contralateral side when both hands are touched. Other modalities relying on the parietal cortex include the discrimination of two closely placed stimuli as separate (two-point discrimination), identification of an object by touch and manipulation alone (stereognosis), and the identification of numbers or letters written on the skin surface (graphesthesia).

Coordination Examination

- *The bare minimum: Test rapid alternating movements of the hands and the finger-to-nose and heel-knee-shin maneuvers.*

Coordination refers to the orchestration and fluidity of movements. Even simple acts require cooperation of agonist and antagonist muscles, maintenance of posture, and complex servomechanisms to control the rate and range of movements. Part of this integration relies on normal function of the cerebellar and basal ganglia systems. However, coordination also requires intact muscle strength and kinesthetic and proprioceptive information. Thus, if the examination has disclosed abnormalities of the motor or sensory systems, the patient's coordination should be assessed with these limitations in mind.

Rapid alternating movements in the upper limbs are tested separately on each side by having the patient make a fist, partially extend the index finger, and then tap the index finger on the distal thumb as quickly as possible. In the lower limb, the patient rapidly taps the foot against the floor or the examiner's hand. Finger-to-nose testing is primarily a test of cerebellar function; the patient is asked to touch his or her index finger repetitively to the nose and then to the examiner's outstretched finger, which moves with each repetition. A similar test in the lower extremity is to have the patient raise the leg and touch the examiner's finger with the great toe. Another cerebellar test in the lower limbs is the heel-knee-shin maneuver; in the supine position the patient is asked to slide the heel of each foot from the knee down the shin of the other leg. For all these movements, the accuracy, speed, and rhythm are noted.

Gait Examination

- *The bare minimum: Observe the patient while walking normally, on the heels and toes, and along a straight line.*

Watching the patient walk is the most important part of the neurologic examination. Normal gait requires that multiple systems—including strength, sensation, and coordination—function in a highly integrated fashion. Unexpected abnormalities may be detected that prompt the examiner to return in more detail to other aspects of the examination. The patient should be observed while walking and turning normally, walking on the heels, walking on the toes, and walking heel-to-toe along a straight line. The examination may reveal decreased arm swing on one side (corticospinal tract disease), a stooped posture and short-stepped gait (parkinsonism), a broad-based unstable gait (ataxia), scissoring (spasticity), or a high-stepped, slapping gait (posterior column or peripheral nerve disease), or the patient may appear to be stuck in place (apraxia with frontal lobe disease).

NEUROLOGIC DIAGNOSIS

The clinical data obtained from the history and examination are interpreted to arrive at an anatomic localization that best explains the clinical findings **(Table 367-2)**, to narrow the list of diagnostic possibilities, and to select the laboratory tests most likely to be informative. The laboratory assessment may include (1) serum electrolytes; complete blood count; and renal, hepatic, endocrine, and immune studies; (2) cerebrospinal fluid examination; (3) focused neuroimaging studies (Chap. 368); or (4) electrophysiologic studies (**Chap. e45**). The anatomic localization, mode of onset and course of illness, other medical data, and laboratory findings are then integrated to establish an etiologic diagnosis.

Table 367-2 Findings Helpful for Localization Within the Nervous System

	Signs
Cerebrum	Abnormal mental status or cognitive impairment Seizures

	Unilateral weaknessa and sensory abnormalities including head and limbs Visual field abnormalities Movement abnormalities (e.g., diffuse incoordination, tremor, chorea)
Brainstem	Isolated cranial nerve abnormalities (single or multiple) "Crossed" weaknessa and sensory abnormalities of head and limbs, e.g., weakness of right face and left arm and leg
Spinal cord	Back pain or tenderness Weaknessa and sensory abnormalities sparing the head Mixed upper and lower motor neuron findings Sensory level Sphincter dysfunction
Spinal roots	Radiating limb pain Weaknessb or sensory abnormalities following root distribution (see Figs. 23-2 and 23-3) Loss of reflexes
Peripheral nerve	Mid or distal limb pain Weaknessb or sensory abnormalities following nerve distribution (see Figs. 23-2 and 23-3) "Stocking or glove" distribution of sensory loss Loss of reflexes
Neuromuscular junction	Bilateral weakness including face (ptosis, diplopia, dysphagia) and proximal limbs Increasing weakness with exertion Sparing of sensation
Muscle	Bilateral proximal or distal weakness Sparing of sensation

[a]Weakness along with other abnormalities having an "upper motor neuron" pattern, i.e., spasticity, weakness of extensors > flexors in the upper extremity and flexors > extensors in the lower extremity, hyperreflexia. [b]Weakness along with other abnormalities having a "lower motor neuron" pattern, i.e., flaccidity and hyporeflexia.

The neurologic examination may be normal even in patients with a serious neurologic disease, such as seizures, chronic meningitis, or a TIA. A comatose patient may arrive with no available history, and in such cases the approach is as described in Chap. 274. In other patients, an inadequate history may be overcome by a succession of examinations from which the course of the illness can be inferred. In perplexing cases it is useful to remember that uncommon presentations of common diseases are more likely than rare etiologies. Thus, even in tertiary care settings, multiple strokes are usually due to emboli and not vasculitis, and dementia with myoclonus is usually Alzheimer's disease and not due to a prion disorder or a paraneoplastic cause. Finally, the most important task of a primary care physician faced with a patient who has a new neurologic complaint is to assess the urgency of referral to a specialist. Here, the imperative is to rapidly identify patients likely to have nervous system infections, acute strokes, and spinal cord compression or other treatable mass lesions and arrange for immediate care.

FURTHER READINGS

Blumenthal H: *Neuroanatomy Through Clinical Cases,* 2nd ed. Sunderland, MA, Sinauer, 2010

Campbell WW: *DeJong's The Neurological Examination,* 6th ed. Philadelphia, Lippincott Williams & Wilkins, 2005

Ropper AH, Samuels M: *Adams and Victor's Principles of Neurolog y*, 9th ed. New York, McGraw-Hill, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE NEUROLOGIC SCREENING EXAM

Knowledge of the basic neurologic examination is an essential clinical skill. A simple neurologic screening examination—assessment of mental status, cranial nerves, motor system, sensory system, coordination, and gait—can be reliably performed in 3–5 min. Although the components of the examination may appear daunting at first, skills usually improve rapidly with repetition and practice. In this video, the technique of performing a simple and efficient screening examination is presented.

Video e42-1
The Neurologic Screening Exam

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

VIDEO ATLAS OF THE DETAILED NEUROLOGIC EXAMINATION

The comprehensive neurologic examination is an irreplaceable tool for the efficient diagnosis of neurologic disorders. Mastery of its details requires knowledge of normal nervous system anatomy and physiology combined with personal experience performing orderly and systematic examinations on large numbers of patients and healthy individuals. In the hands of a great clinician, the neurologic examination also becomes a thing of beauty—the pinnacle of the art of medicine. In this video, the most commonly used components of the examination are presented in detail, with a particular emphasis on those elements that are most helpful for assessment of common neurologic problems.

Video e43-1
Introduction and the General Physical Examination Relevant to Neurology
Video e43-2
Mental Status
Video e43-3
Cranial Nerves
Video e43-4
Motor
Video e43-5
Sensory
Video e43-6
Reflexes
Video e43-7
Coordination and Gait

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NEUROIMAGING IN NEUROLOGIC DISORDERS: INTRODUCTION

The clinician caring for patients with neurologic symptoms is faced with myriad imaging options, including computed tomography (CT), CT angiography (CTA), perfusion CT (pCT), magnetic resonance imaging (MRI), MR angiography (MRA), functional MRI (fMRI), MR spectroscopy (MRS), MR neurography (MRN), diffusion and diffusion track imaging (DTI), susceptibility weighted MR imaging (SWI), and perfusion MRI (pMRI). In addition, an increasing number of interventional neuroradiologic techniques are available, including angiography catheter embolization, coiling, and stenting of vascular structures; and spine diagnostic and interventional techniques such as diskography, transforaminal and translaminar epidural and nerve root injections and blood patches. Recent developments such as multidetector CTA (MDCTA) and gadolinium-enhanced MRA, have narrowed the indications for conventional angiography, which is now reserved for patients in whom small-vessel detail is essential for diagnosis or for whom concurrent interventional therapy is planned **(Table 368-1)**.

Table 368-1 Guidelines for the Use of CT, Ultrasound, and MRI

Condition	Recommended Technique
Hemorrhage	
Acute parenchymal	CT, MR
Subacute/chronic	MRI
Subarachnoid hemorrhage	CT, CTA, lumbar puncture angiography
Aneurysm	Angiography > CTA, MRA
Ischemic infarction	
Hemorrhagic infarction	CT or MRI
Bland infarction	MRI > CT, CTA, angiography
Carotid or vertebral dissection	MRI/MRA
Vertebral basilar insufficiency	CTA, MRI/MRA
Carotid stenosis	CTA > Doppler ultrasound, MRA
Suspected mass lesion	
Neoplasm, primary or metastatic	MRI + contrast
Infection/abscess	MRI + contrast
Immunosuppressed with focal findings	MRI + contrast
Vascular malformation	MRI ± angiography
White matter disorders	MRI
Demyelinating disease	MRI ± contrast
Dementia	MRI > CT
Trauma	
Acute trauma	CT (noncontrast)
Shear injury/chronic hemorrhage	MRI + gradient echo imaging
Headache/migraine	CT (noncontrast) / MRI
Seizure	
First time, no focal neurologic deficits	?CT as screen ± contrast
Partial complex/refractory	MRI with coronal T2W imaging
Cranial neuropathy	MRI with contrast
Meningeal disease	MRI with contrast
Spine	
Low back pain	
No neurologic deficits	MRI or CT after 4 weeks
With focal deficits	MRI > CT

Spinal stenosis	MRI or CT
Cervical spondylosis	MRI or CT myelography
Infection	MRI + contrast, CT
Myelopathy	MRI + contrast
Arteriovenous malformation	MRI, angiography

Abbreviations: CT, computed tomography; CTA, CT angiography; MRA, MR angiography; MRI, magnetic resonance imaging; T2W, T2-weighted.

In general, MRI is more sensitive than CT for the detection of lesions affecting the central nervous system (CNS), particularly those of the spinal cord, cranial nerves, and posterior fossa structures. Diffusion MR, a sequence sensitive to the microscopic motion of water, is the most sensitive technique for detecting acute ischemic stroke of the brain or spinal cord, and it is also useful in the detection of encephalitis, abscesses, and prion diseases. CT, however, is quickly acquired and is widely available, making it a pragmatic choice for the initial evaluation of patients with acute changes in mental status, suspected acute stroke, hemorrhage, and intracranial or spinal trauma. CT is also more sensitive than MRI for visualizing fine osseous detail and is indicated in the initial evaluation of conductive hearing loss as well as lesions affecting the skull base and calvarium. MR may, however, add important diagnostic information regarding bone marrow infiltrative processes that are difficult to detect on CT.

COMPUTED TOMOGRAPHY

Technique

The CT image is a cross-sectional representation of anatomy created by a computer-generated analysis of the attenuation of x-ray beams passed through a section of the body. As the x-ray beam, collimated to the desired slice width, rotates around the patient, it passes through selected regions in the body. X-rays that are not attenuated by body structures are detected by sensitive x-ray detectors aligned 180° from the x-ray tube. A computer calculates a "back projection" image from the 360° x-ray attenuation profile. Greater x-ray attenuation (e.g., as caused by bone), results in areas of high "density," while soft tissue structures that have poor attenuation of x-rays such as organs and air-filled cavities, are lower in density. The resolution of an image depends on the radiation dose, the detector size, collimation (slice thickness), the field of view, and the matrix size of the display. A modern CT scanner is capable of obtaining sections as thin as 0.5–1 mm with submillimeter resolution at a speed of 0.3–1 s per rotation; complete studies of the brain can be completed in 2–10 s.

Multidetector CT (MDCT) is now standard in most radiology departments. Single or multiple (from 4 to 256) detectors positioned 180° to the x-ray source result in multiple slices per revolution of the beam around the patient. The table moves continuously through the rotating x-ray beam, generating a continuous "helix" of information that can be reformatted into various slice thicknesses and planes. Advantages of MDCT include shorter scan times, reduced patient and organ motion, and the ability to acquire images dynamically during the infusion of intravenous contrast that can be used to construct CT angiograms of vascular structures and CT perfusion images **(Fig. 368-1*B* and *C*)**. CTA images are post-processed for display in three dimensions to yield angiogram-like images (Fig. 368-1*C*, 368-2 E and F, and see Fig. 370-4). CTA has proved useful in assessing the cervical and intracranial arterial and venous anatomy.

Figure 368-1

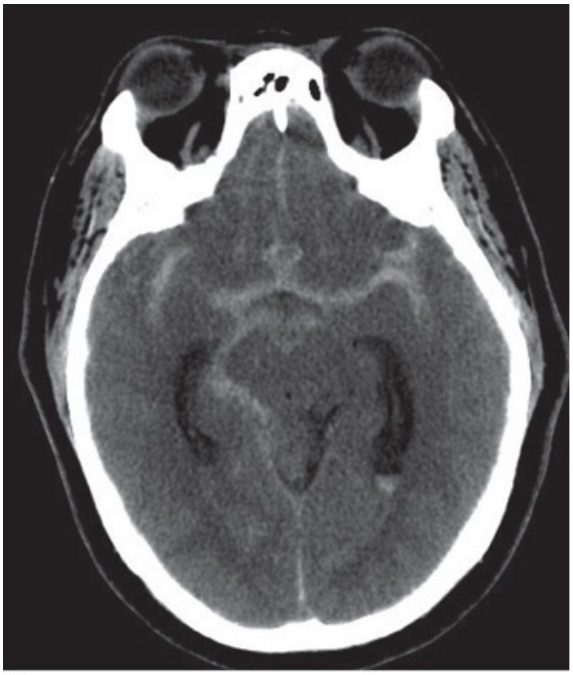

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

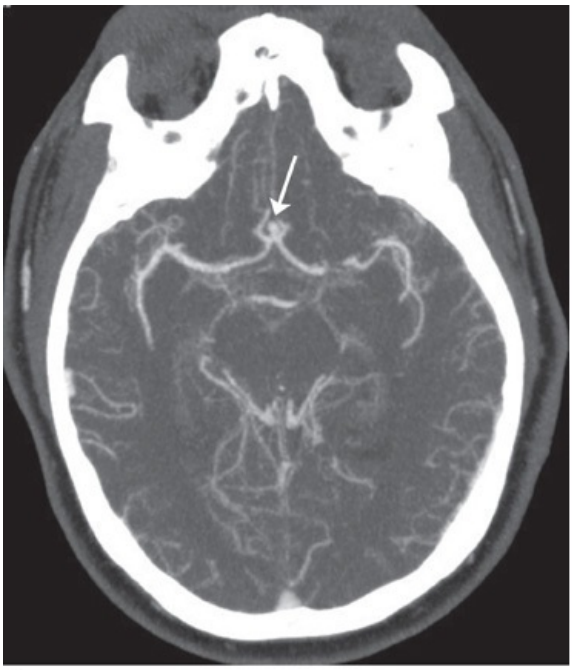

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

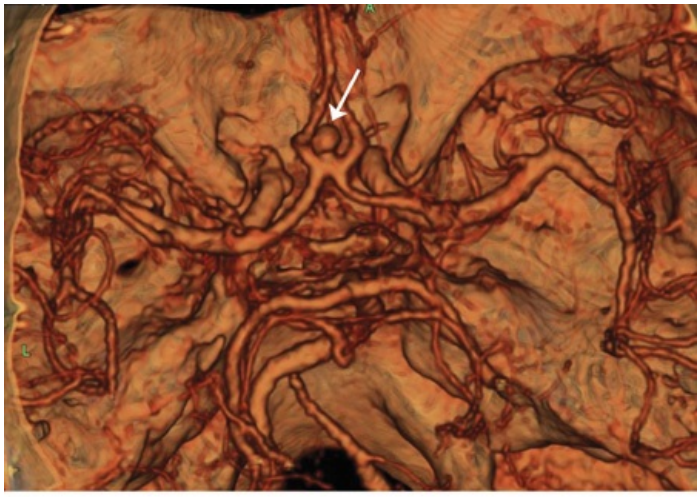

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT angiography (CTA) of ruptured anterior cerebral artery aneurysm in a patient presenting with acute headache.*A*. Noncontrast CT demonstrates subarachnoid hemorrhage and mild obstructive hydrocephalus. ***B***. Axial maximum-intensity projection from CT angiography demonstrates enlargement of the anterior cerebral artery (*arrow?*). ***C***. 3D surface reconstruction using a workstation confirms the anterior cerebral aneurysm and demonstrates its orientation and relationship to nearby vessels (*arrow?*). CTA image is produced by 0.5- 1-mm helical CT scans performed during a rapid bolus infusion of intravenous contrast medium.

Figure 368-2

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

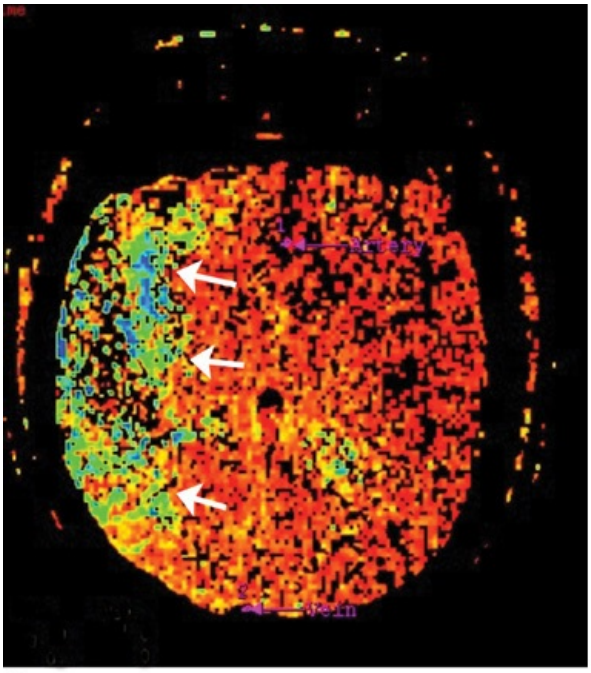

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

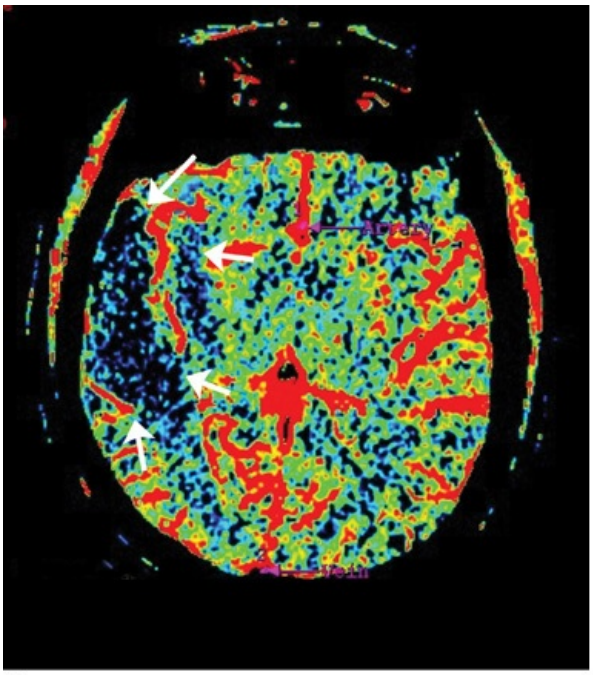

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

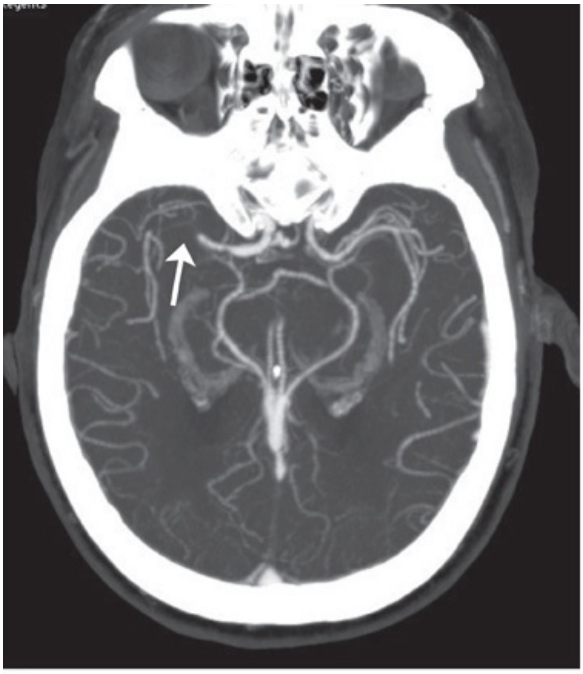

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

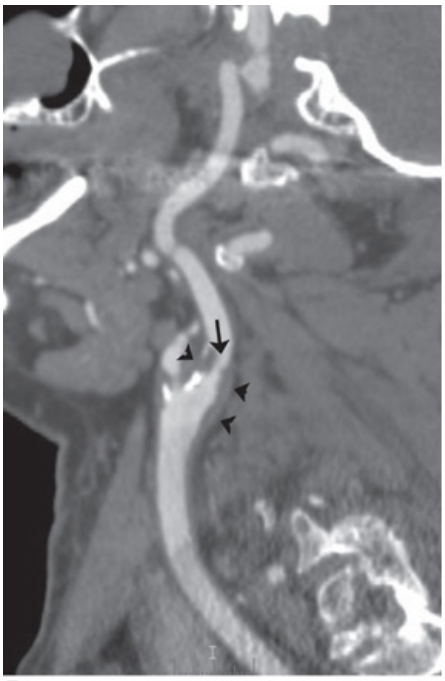

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

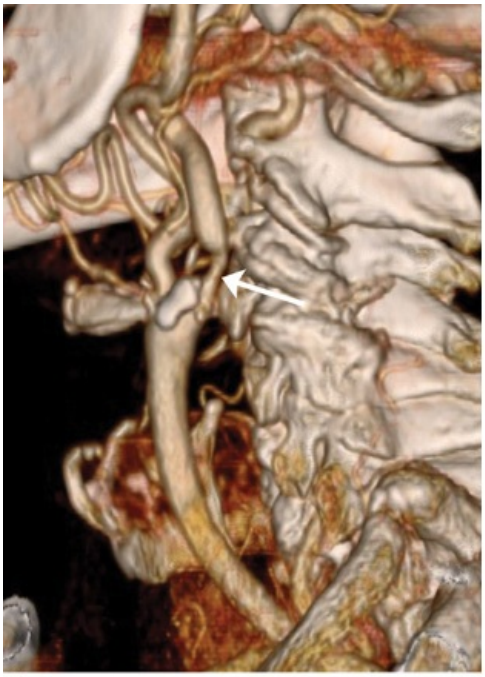

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

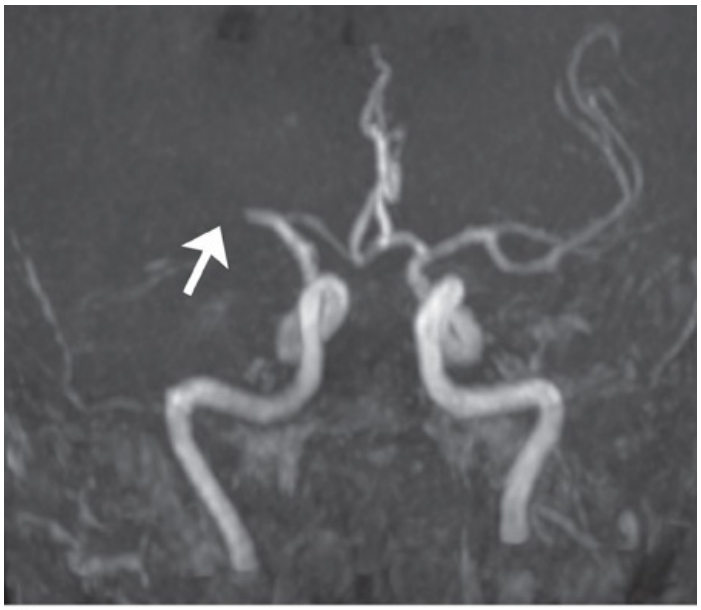

G

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

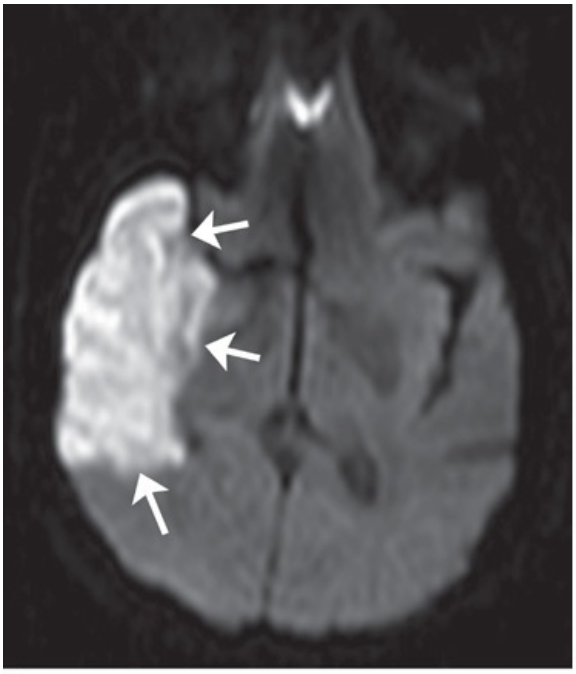

H

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

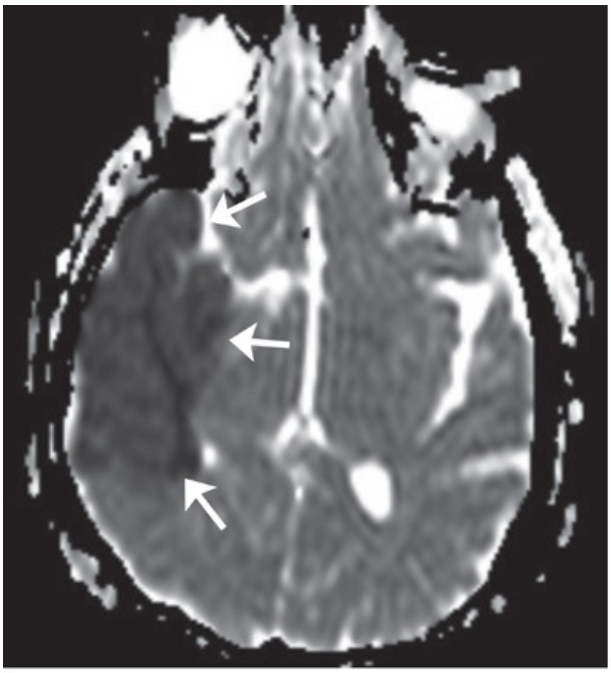

I

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute left hemiparesis due to middle cerebral artery occlusion.*A*. Axial noncontrast CT scan demonstrates high density within the right middle cerebral artery (*arrow?*) associated with subtle low density involving the right putamen (*arrowheads?*). ***B***. Mean transit time CT perfusion parametric map indicating prolonged mean transit time involving the right middle cerebral territory (*arrows?*). ***C***. Cerebral blood volume map shows reduced CBV involving an area within the defect shown in ***B***, indicating a high likelihood of infarction (*arrows?*). ***D***. Axial maximum-intensity projection from a CTA study through the circle of Willis demonstrates an abrupt occlusion of the proximal right middle cerebral artery (*arrow*). ***E***. Sagittal reformation through the right internal carotid artery demonstrates a low-density lipid-laden plaque (*arrowheads*) narrowing the lumen (*black arrow?*) ***F***. 3D surface-rendered CTA image demonstrates calcification and narrowing of the right internal carotid artery (*arrow?*), consistent with atherosclerotic disease. ***G***. Coronal maximum-intensity projection from MRA shows right middle cerebral artery (MCA) occlusion (*arrow?*). ***H***. and ***I***. Axial diffusion-weighted image (***H?***) and apparent diffusion coefficient image (***I?***) documents the presence of a right middle cerebral artery infarction.

Intravenous iodinated contrast is often administered prior to or during a CT study to identify vascular

structures and to detect defects in the blood-brain barrier (BBB) that are associated with disorders such as tumors, infarcts, and infections. In the normal CNS, only vessels and structures lacking a BBB (e.g., the pituitary gland, choroid plexus, and dura) enhance after contrast administration. The use of iodinated contrast agents carries a small risk of allergic reaction and adds additional expense. While helpful in characterizing mass lesions as well as essential for the acquisition of CTA studies, the decision to use contrast material should always be considered carefully.

Indications

CT is the primary study of choice in the evaluation of an acute change in mental status, focal neurologic findings, acute trauma to the brain and spine, suspected subarachnoid hemorrhage, and conductive hearing loss (Table 368-1). CT is complementary to MR in the evaluation of the skull base, orbit, and osseous structures of the spine. In the spine, CT is useful in evaluating patients with osseous spinal stenosis and spondylosis, but MRI is often preferred in those with neurologic deficits. CT can also be obtained following intrathecal contrast injection to evaluate the intracranial cisterns (*CT cisternography*) for cerebrospinal fluid (CSF) fistula, as well as the spinal subarachnoid space (*CT myelography*).

Complications

CT is safe, fast, and reliable. Radiation exposure depends on the dose used but is normally between 2 and 5 mSv (millisievert) for a routine brain CT study. Care must be taken to reduce exposure when imaging children. With the advent of MDCT, CTA, and CT perfusion, care must be taken to appropriately minimize radiation dose whenever possible. Advanced software that permits noise reduction may permit lower radiation doses. The most frequent complications are associated with use of intravenous contrast agents. Two broad categories of contrast media, ionic and nonionic, are in use. Although ionic agents are relatively safe and inexpensive, they are associated with a higher incidence of reactions and side effects. As a result, ionic agents have been largely replaced by safer nonionic compounds.

Contrast nephropathy may result from hemodynamic changes, renal tubular obstruction and cell damage, or immunologic reactions to contrast agents. A rise in serum creatinine of at least 85 μmol/L (1 mg/dL) within 48 hours of contrast administration is often used as a definition of contrast nephropathy, although other causes of acute renal failure must be excluded. The prognosis is usually favorable, with serum creatinine levels returning to baseline within 1–2 weeks. Risk factors for contrast nephropathy include advanced age (>80 years), preexisting renal disease (serum creatinine exceeding 2 mg/dL), solitary kidney, diabetes mellitus, dehydration, paraproteinemia, concurrent use of nephrotoxic medication or chemotherapeutic agents, and high contrast dose. Patients with diabetes and those with mild renal failure should be well hydrated prior to the administration of contrast agents, although careful consideration should be given to alternative imaging techniques such as MR imaging or noncontrast CT or ultrasound (US) examinations. Nonionic, low-osmolar media produce fewer abnormalities in renal blood flow and less endothelial cell damage but should still be used carefully in patients at risk for allergic reaction. Estimated glomerular filtration rate (eGFR) is a more reliable indicator of renal function compared to creatinine alone as it takes into account age, race and sex. In one study, 15% of outpatients with a normal serum creatinine had an estimated creatinine clearance of 50 mL/min/1.73 m^2 or less (normal is 90 mL/min/1.73 m^2 or more). The exact eGFR threshold, below which withholding intravenous contrast should be considered, is controversial. The risk of contrast nephropathy increases in patients with an eGFR <60 mL/min/1.73^2; however the majority of these patients will only have a temporary rise in creatinine. The risk of dialysis after receiving contrast significantly increases in patients with eGFR <30 mL/min/1.73^2. Thus, an eGFR threshold between 60 and 30 mL/min/1.73^2 is appropriate; however the exact number is somewhat arbitrary. A creatinine of 1.6 in a 70-year-old, non-African-American male corresponds to an eGFR of approximately 45 mL/min/1.73^2. The American College of Radiology suggests using an eGFR of 45 as a threshold below which iodinated contrast should not be given without serious consideration of the potential for contrast nephropathy. If contrast must be administered to a patient with an eGRF below 45, the patient should be well hydrated, and a reduction in the dose of contrast should be considered. Use of other agents such as bicarbonate and acetylcysteine may reduce the incidence of contrast nephropathy. Other side effects of CT scanning are rare but include a sensation of warmth throughout the body and a metallic taste during intravenous administration of iodinated contrast media. The most serious side effects are anaphylactic reactions, which range from mild hives to bronchospasm, acute anaphylaxis, and death. The pathogenesis of these allergic reactions is not fully understood but is thought to include the release of mediators such as histamine, antibody-antigen reactions, and complement activation. Severe allergic reactions occur in ~0.04% of patients receiving nonionic media, sixfold lower than with ionic media. Risk factors include a history of prior contrast reaction, food allergies to shellfish, and atopy (asthma and hay fever). In such patients, a noncontrast CT or MRI procedure should be considered as an alternative to contrast administration. If iodinated contrast is absolutely required, a nonionic agent should be used in conjunction with pretreatment with glucocorticoids and antihistamines **(Table 368-2)**. Patients with allergic reactions to iodinated contrast material do not

usually react to gadolinium-based MR contrast material, although such reactions can occur. It would be wise to pretreat patients with a prior allergic history to MR contrast administration in a similar fashion.

Table 368-2 Guidelines for Premedication of Patients with Prior Contrast Allergy

12 h prior to examination:
Prednisone, 50 mg PO or methylprednisolone, 32 mg PO
2 h prior to examination:
Prednisone, 50 mg PO or methylprednisolone, 32 mg PO and Cimetidine, 300 mg PO or ranitidine, 150 mg PO
Immediately prior to examination:
Benadryl, 50 mg IV (alternatively, can be given PO 2 h prior to exam)

MAGNETIC RESONANCE IMAGING

Technique

MRI is a complex interaction between hydrogen protons in biologic tissues, a static magnetic field (the magnet), and energy in the form of radiofrequency (Rf) waves of a specific frequency introduced by coils placed next to the body part of interest. Images are made by computerized processing of resonance information received from protons in the body. Field strength of the magnet is directly related to signal-to-noise ratio. While 1.5-Telsa magnets have become the standard high-field MRI units, 3T–8T magnets are now available and have distinct advantages in the brain and musculoskeletal systems. Spatial localization is achieved by magnetic gradients surrounding the main magnet, which impart slight changes in magnetic field throughout the imaging volume. Rf pulses transiently excite the energy state of the hydrogen protons in the body. Rf is administered at a frequency specific for the field strength of the magnet. The subsequent return to equilibrium energy state (*relaxation*) of the hydrogen protons results in a release of Rf energy (the *echo*), which is detected by the coils that delivered the Rf pulses. The echo is transformed by Fourier analysis into the information used to form an MR image. The MR image thus consists of a map of the distribution of hydrogen protons, with signal intensity imparted by both density of hydrogen protons as well as differences in the relaxation times (see below) of hydrogen protons on different molecules. While clinical MRI currently makes use of the ubiquitous hydrogen proton, research into sodium and carbon imaging appears promising.

T1 AND T2 RELAXATION TIMES

The rate of return to equilibrium of perturbed protons is called the *relaxation rate*. The relaxation rate varies among normal and pathologic tissues. The relaxation rate of a hydrogen proton in a tissue is influenced by local interactions with surrounding molecules and atomic neighbors. Two relaxation rates, T1 and T2, influence the signal intensity of the image. The T1 relaxation time is the time, measured in milliseconds, for 63% of the hydrogen protons to return to their normal equilibrium state, while the T2 relaxation is the time for 63% of the protons to become dephased owing to interactions among nearby protons. The intensity of the signal within various tissues and image contrast can be modulated by altering acquisition parameters such as the interval between Rf pulses (TR) and the time between the Rf pulse and the signal reception (TE). So-called T1-weighted (T1W) images are produced by keeping the TR and TE relatively short. T2-weighted (T2W) images are produced by using longer TR and TE times. Fat and subacute hemorrhage have relatively shorter T1 relaxation rates and thus higher signal intensity than brain on T1W images. Structures containing more water such as CSF and edema, have long T1 and T2 relaxation rates, resulting in relatively lower signal intensity on T1W images and a higher signal intensity on T2W images **(Table 368-3)**. Gray matter contains 10–15% more water than white matter, which accounts for much of the intrinsic contrast between the two on MRI (Fig. 368-6B). T2W images are more sensitive than T1W images to edema, demyelination, infarction, and chronic hemorrhage, while T1W imaging is more sensitive to subacute hemorrhage and fat-containing structures.

Table 368-3 Some Common Intensities on T1- and T2-Weighted MRI Sequences

			Signal Intensity			
Image	**TR**	**TE**	**CSF**	**Fat**	**Brain**	**Edema**
T1W	Short	Short	Low	High	Low	Low
T2W	Long	Long	High	Low	High	High
FLAIR (T2)	Long	Long	Low	Medium	High	High

Abbreviations: CSF, cerebrospinal fluid; TE, interval between Rf pulse and signal reception; TR, interval between radiofrequency (Rf) pulses; T1W and T2W, T1- and T2-weighted.

Many different MR pulse sequences exist, and each can be obtained in various planes **(Figs. 368-2,368-3, 368-4)**. The selection of a proper protocol that will best answer a clinical question depends on an accurate clinical history and indication for the examination. Fluid-attenuated inversion recovery (FLAIR) is a useful pulse sequence that produces T2W images in which the normally high signal intensity of CSF is suppressed (Fig. 368-6B). FLAIR images are more sensitive than standard spin echo images for any water-containing lesions or edema. Susceptibility weighted imaging such as gradient echo imaging, is most sensitive to magnetic susceptibility generated by blood, calcium, and air and is indicated in patients suspected of pathology that might result in microhemorrhages **(Fig. 368-5*C***). MR images can be generated in any plane without changing the patient's position. Each sequence, however, must be obtained separately and takes 1–10 minutes on average to complete. Three-dimensional volumetric imaging is also possible with MRI, resulting in a 3D volume of data that can be reformatted in any orientation to highlight certain disease processes.

Figure 368-3

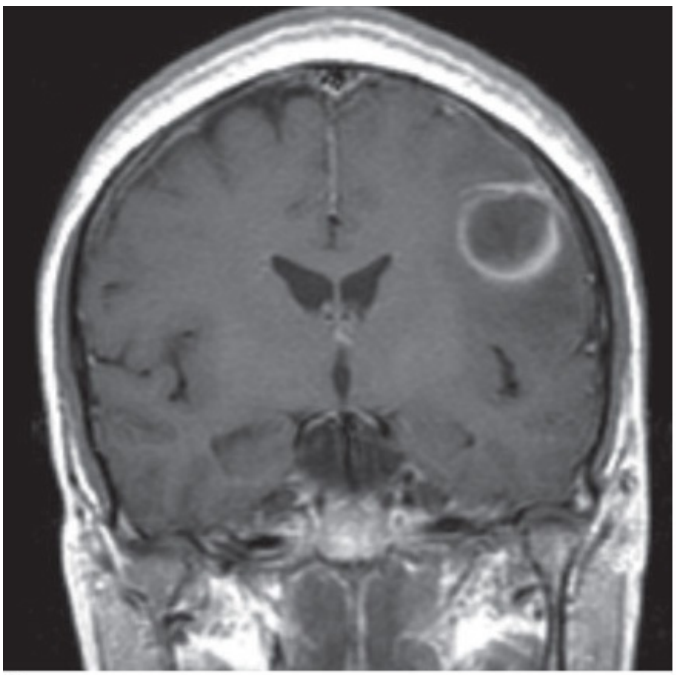

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

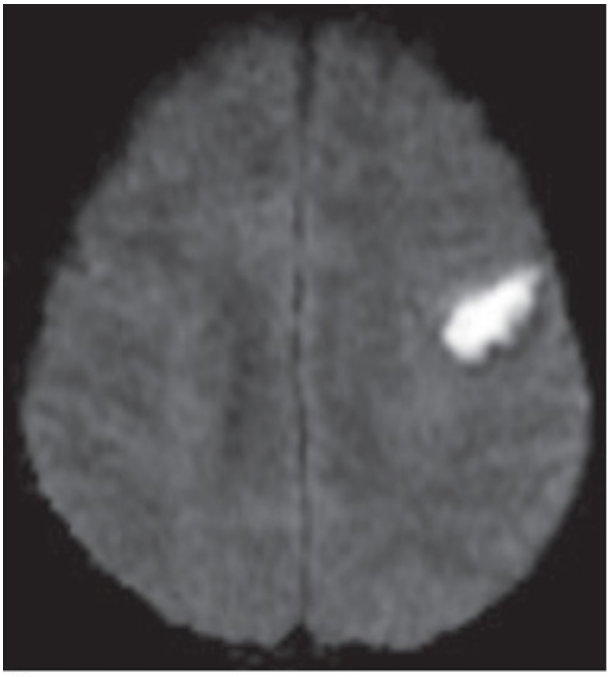

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cerebral abscess in a patient with fever and a right hemiparesis.*A*. Coronal postcontrast T1-weighted image demonstrates a ring enhancing mass in the left frontal lobe. ***B***. Axial diffusion-weighted image demonstrates restricted diffusion (high signal intensity) within the lesion, which in this setting is highly suggestive of cerebral abscess.

Figure 368-4

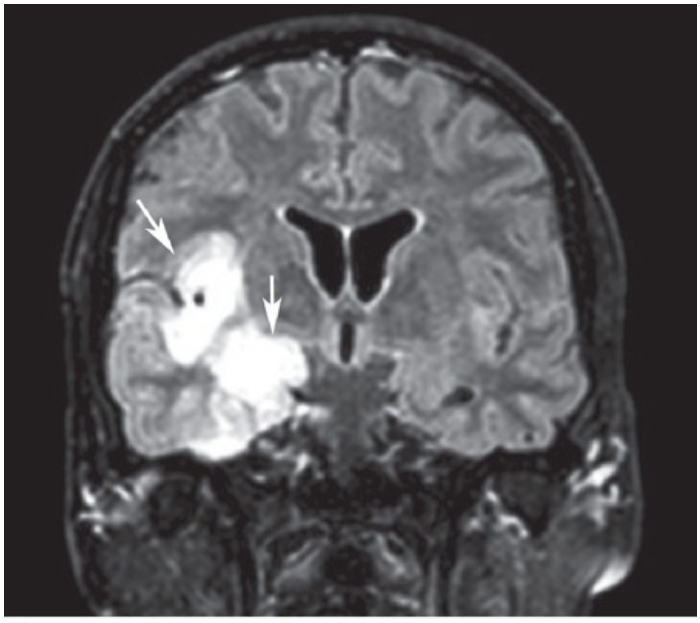

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

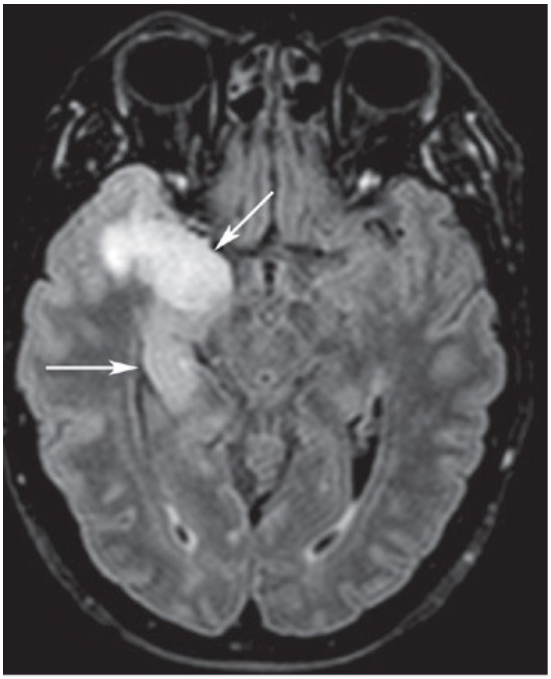

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

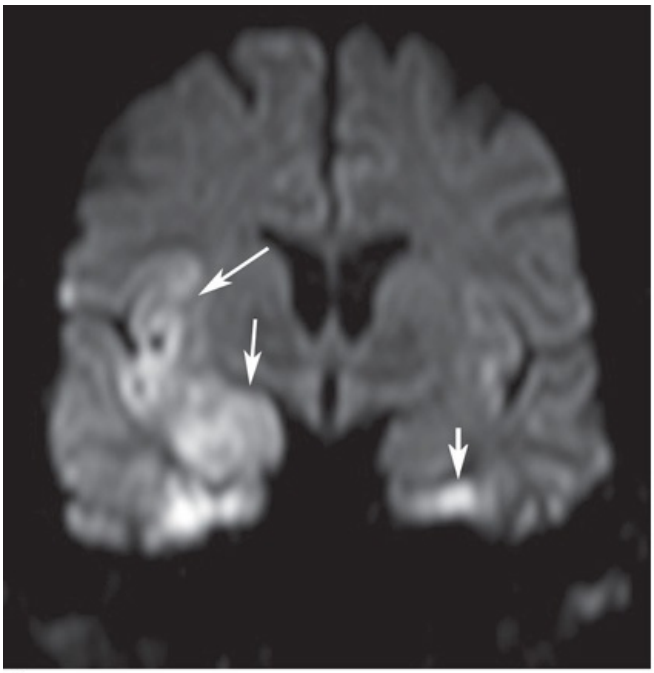

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Herpes simplex encephalitis in a patient presenting with altered mental status and fever.*A*. and ***B***. Coronal ***(A)*** and axial ***(B)*** T2-weighted FLAIR images demonstrates expansion and high signal intensity involving the right medial temporal lobe and insular cortex (*arrows?*). ***C***. Coronal diffusion-weighted image demonstrates high signal intensity indicating restricted diffusion involving the right medial temporal lobe and hippocampus (*arrows?*) as well as subtle involvement of the left inferior temporal lobe (*arrowhead?*). This is most consistent with neuronal death and can be seen in acute infarction as well as encephalitis and other inflammatory conditions. The suspected diagnosis of herpes simplex encephalitis was confirmed by CSF PCR analysis.

Figure 368-5

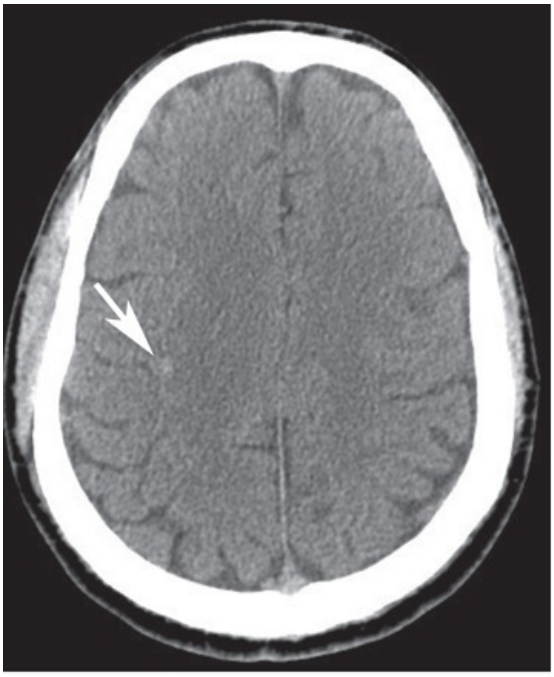

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

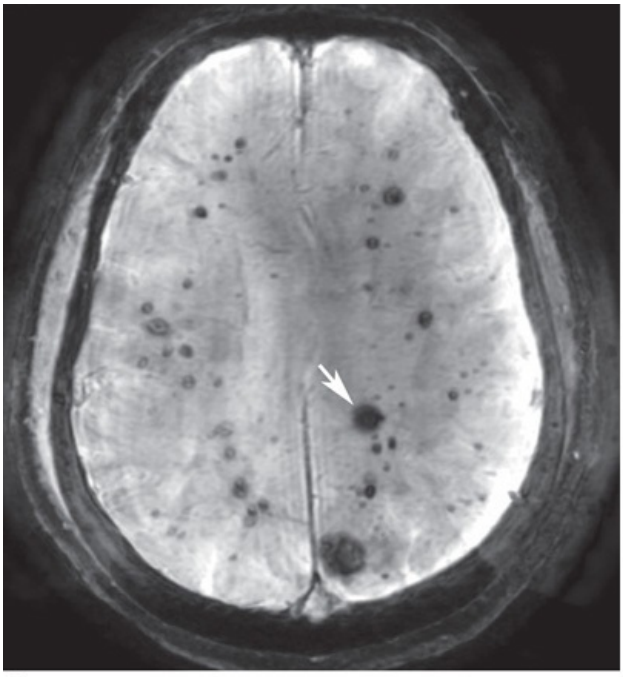

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Susceptibility weighted imaging in a patient with familial cavernous malformations. *A.* Noncontrast CT scan shows one hyperdense lesion in the right hemisphere (*arrow*). ***B***. T2-weighted fast spin echo image shows subtle low-intensity lesions (*arrows?*). ***C***. Susceptibility weighted image shows numerous low-intensity lesions consistent with hemosiderin-laden cavernous malformations (*arrow?*).

Figure 368-6

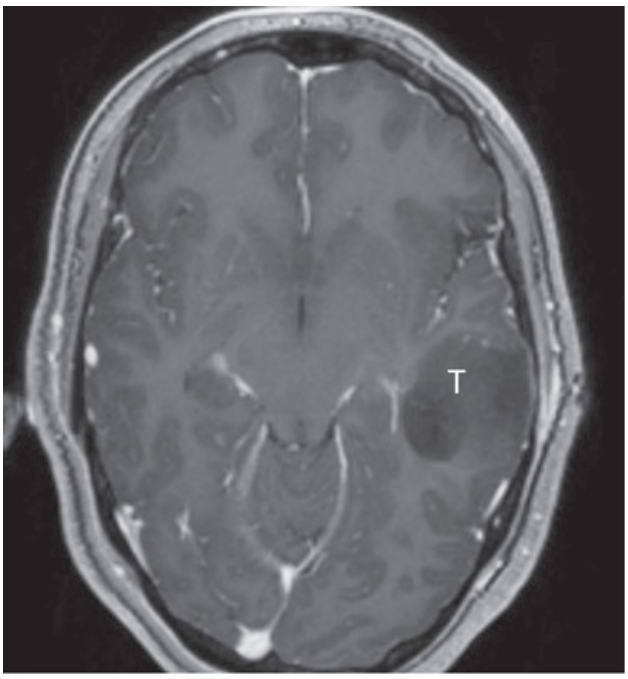

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

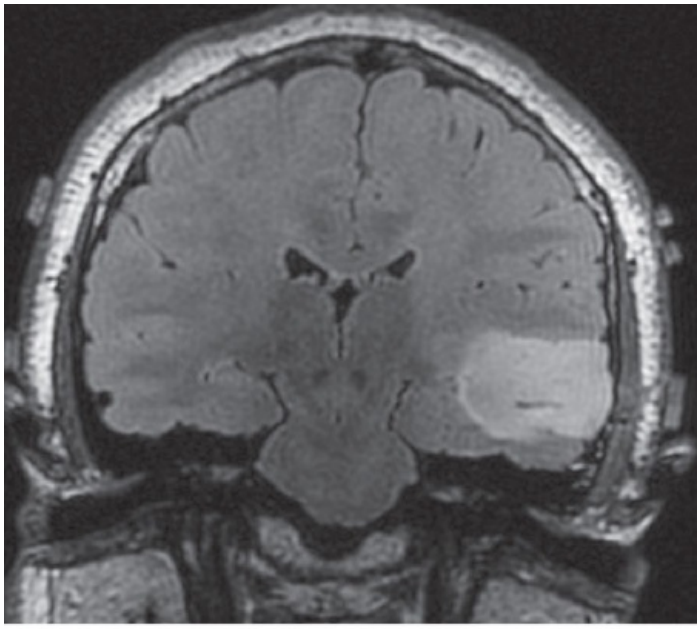

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

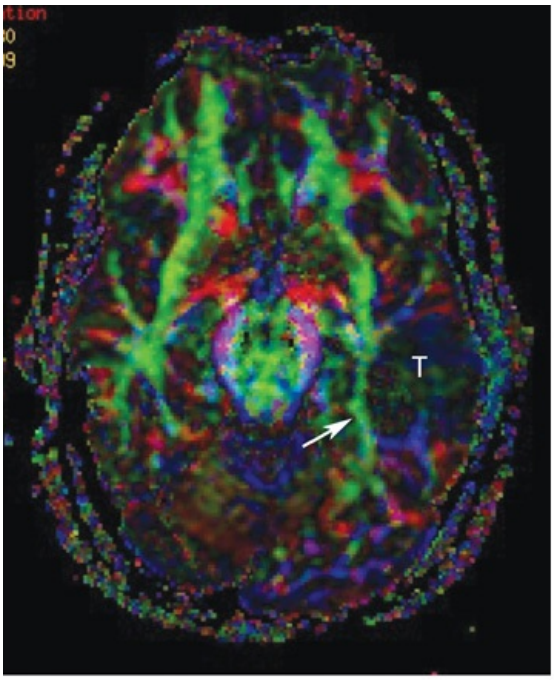

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diffusion tractography in cerebral glioma. ***A***. An axial postcontrast T1-weighted image shows a nonenhancing glioma (T) of the left temporal lobe cortex lateral to the fibers of the internal capsule. ***B***. Coronal T2 FLAIR image demonstrates high signal glioma in left temporal lobe. ***C***. Axial diffusion fractional anisotropy image shows the position of the deep white matter fibers (*arrow?*) relative to the enhancing tumor (T).

MR CONTRAST MATERIAL

The heavy-metal element gadolinium forms the basis of all currently approved intravenous MR contrast agents. Gadolinium is a paramagnetic substance, which means that it reduces the T1 and T2 relaxation times of nearby water protons, resulting in a high signal on T1W images and a low signal on T2W images (the latter requires a sufficient local concentration, usually in the form of an intravenous bolus). Unlike iodinated contrast agents, the effect of MR contrast agents depends on the presence of local hydrogen protons on which it must act to achieve the desired effect. Gadolinium is chelated to DTPA (diethylenetriaminepentaacetic acid), which allows safe renal excretion. Approximately 0.2 mL/kg body weight is administered intravenously; the cost is ~$60 per dose. Gadolinium-DTPA does not

normally cross the intact BBB immediately but will enhance lesions lacking a BBB (Fig. 368-3*A*) and areas of the brain that normally are devoid of the BBB (pituitary, choroid plexus). However, gadolinium contrast has been noted to slowly cross an intact BBB if given over time and especially in the setting of reduced renal clearance. The agents are generally well tolerated; severe allergic reactions are rare but have been reported. The adverse reaction rate in patients with a prior history of atopy or asthma is 3.7%; however, the reaction rate increases to 6.3% in those patients with a prior history of unspecified allergic reaction to iodinated contrast agents. Gadolinium contrast material can be administered safely to children as well as adults, although these agents are generally avoided in those under 6 months of age. Renal failure does not occur.

A rare complication, nephrogenic systemic fibrosis (NSF), has recently been reported in patients with renal insufficiency who have been exposed to gadolinium contrast agents. The onset of NSF has been reported between 5 and 75 days following exposure; histologic features include thickened collagen bundles with surrounding clefts, mucin deposition, and increased numbers of fibrocytes and elastic fibers in skin. In addition to dermatologic symptoms, other manifestations include widespread fibrosis of the skeletal muscle, bone, lungs, pleura, pericardium, myocardium, kidney, muscle, bone, testes, and dura. For this reason, the American College of Radiology recommends that prior to elective gadolinium-based MR contrast agent (GBMCA) administration, a recent (e.g., past 6 weeks') glomerular filtration rate (GFR) assessment be obtained in patients with a history of:

1. Renal disease (including solitary kidney, renal transplant, renal tumor)
2. Age >60 years
3. History of hypertension
4. History of diabetes
5. History of severe hepatic disease/liver transplant/pending liver transplant: for these patients it is recommended that the patient's GFR assessment be nearly contemporaneous with the MR examination.

The incidence of NSF in patients with severe renal dysfunction (GFR <30) varies from 0.19 to 4%. A recent meta-analysis reported an odds ratio of 26.7 (95% CI = 10.3–69.4) for development of NSF after gadolinium administration in patients with impaired renal function (GFR <30 mL/min/1.72 m). Thus, it is not recommended to administer gadolinium to any patient with a GRF below 30. Caution is advised for patients with a GRFbelow 45.

Complications and Contraindications

From the patient's perspective, an MRI examination can be intimidating, and a higher level of cooperation is required than with CT. The patient lies on a table that is moved into a long, narrow gap within the magnet. Approximately 5% of the population experiences severe claustrophobia in the MR environment. This can be reduced by mild sedation but remains a problem for some. Unlike CT, movement of the patient during an MR sequence distorts all the images; therefore, uncooperative patients should either be sedated for the MR study or scanned with CT. Generally, children under the age of 10 years usually require conscious sedation in order to complete the MR examination without motion degradation.

MRI is considered safe for patients, even at very high field strengths (>3–4 T). Serious injuries have been caused, however, by attraction of ferromagnetic objects into the magnet, which act as missiles if brought too close to the magnet. Likewise, ferromagnetic implants such as aneurysm clips, may torque within the magnet, causing damage to vessels and even death. Metallic foreign bodies in the eye have moved and caused intraocular hemorrhage; screening for ocular metallic fragments is indicated in those with a history of metal work or ocular metallic foreign bodies. Implanted cardiac pacemakers are generally a contraindication to MRI owing to the risk of induced arrhythmias; however, some newer pacemakers have been shown to be safe. All health care personnel and patients must be screened and educated thoroughly to prevent such disasters as the magnet is always "on." **Table 368-4** lists common contraindications for MRI.

Table 368-4 Common Contraindications to MR Imaging

Cardiac pacemaker or permanent pacemaker leads
Internal defibrillatory device
Cochlear prostheses
Bone growth stimulators
Spinal cord stimulators
Electronic infusion devices
Intracranial aneurysm clips (some but not all)

Ocular implants (some) or ocular metallic foreign body
McGee stapedectomy piston prosthesis
Duraphase penile implant
Swan-Ganz catheter
Magnetic stoma plugs
Magnetic dental implants
Magnetic sphincters
Ferromagnetic IVC filters, coils, stents—safe 6 weeks after implantation
Tattooed eyeliner (contains ferromagnetic material and may irritate eyes)

Note: See also *http://www.mrisafety.com.*

MAGNETIC RESONANCE ANGIOGRAPHY

MR angiography is a general term describing several MR techniques that result in vascular-weighted images. These provide a vascular flow map rather than the anatomic map shown by conventional angiography. On routine spin echo MR sequences, moving protons (e.g., flowing blood, CSF) exhibit complex MR signals that range from high- to low-signal intensity relative to background stationary tissue. Fast-flowing blood returns no signal (flow void) on routine T1W or T2W spin echo MR images. Slower-flowing blood, as occurs in veins or distal to arterial stenosis, may appear high in signal. However, using special pulse sequences called *gradient echo sequences*, it is possible to increase the signal intensity of moving protons in contrast to the low signal background intensity of stationary tissue. This creates angiography-like images, which can be manipulated in three dimensions to highlight vascular anatomy and relationships.

Time-of-flight (TOF) imaging, currently the technique used most frequently, relies on the suppression of nonmoving tissue to provide a low-intensity background for the high signal intensity of flowing blood entering the section; arterial or venous structures may be highlighted. A typical TOF angiography sequence results in a series of contiguous, thin MR sections (0.6–0.9 mm thick), which can be viewed as a stack and manipulated to create an angiographic image data set that can be reformatted and viewed in various planes and angles, much like that seen with conventional angiography (Fig. 368-2*G*).

Phase-contrast MRA has a longer acquisition time than TOF MRA, but in addition to providing anatomic information similar to that of TOF imaging, it can be used to reveal the velocity and direction of blood flow in a given vessel. Through the selection of different imaging parameters, differing blood velocities can be highlighted; selective venous and arterial MRA images can thus be obtained. One advantage of phase-contrast MRA is the excellent suppression of high-signal-intensity background structures.

MRA can also be acquired during infusion of contrast material. Advantages include faster imaging times (1–2 minutes vs. 10 minutes), fewer flow-related artifacts, and higher-resolution images. Recently, contrast-enhanced MRA has become the standard for extracranial vascular MRA. This technique entails rapid imaging using coronal three-dimensional TOF sequences during a bolus infusion of 15–20 mL of gadolinium-DTPA. Proper technique and timing of acquisition relative to bolus arrival are critical for success.

MRA has lower spatial resolution compared with conventional film-based angiography, and therefore the detection of small-vessel abnormalities such as vasculitis and distal vasospasm, is problematic. MRA is also less sensitive to slowly flowing blood and thus may not reliably differentiate complete from near-complete occlusions. Motion, either by the patient or by anatomic structures, may distort the MRA images, creating artifacts. These limitations notwithstanding, MRA has proved useful in evaluation of the extracranial carotid and vertebral circulation as well as of larger-caliber intracranial arteries and dural sinuses. It has also proved useful in the noninvasive detection of intracranial aneurysms and vascular malformations.

ECHO-PLANAR MR IMAGING

Recent improvements in gradients, software, and high-speed computer processors now permit extremely rapid MRI of the brain. With echo-planar MRI (EPI), fast gradients are switched on and off at high speeds to create the information used to form an image. In routine spin echo imaging, images of the brain can be obtained in 5–10 minutes. With EPI, all of the information required for processing an image is accumulated in 50–150 milliseconds, and the information for the entire brain is obtained in 1–2 minutes, depending on the degree of resolution required or desired. Fast MRI reduces patient and organ motion, permitting diffusion imaging and tractography (Figs. 368-2H, 368-3, 368-4C, 368-6; and see Fig. 370-16), perfusion imaging during contrast infusion, fMRI, and kinematic motion studies.

Perfusion and diffusion imaging are EPI techniques that are useful in early detection of ischemic injury of the brain and may be useful together to demonstrate infarcted tissue as well as ischemic but potentially viable tissue at risk of infarction (e.g., the ischemic penumbra). Diffusion-weighted imaging (DWI) assesses microscopic motion of water; restriction of motion appears as relative high-signal intensity on diffusion-weighted images. Infarcted tissue reduces the water motion within cells and in the interstitial tissues, resulting in high signal on DWI. DWI is the most sensitive technique for detection of acute cerebral infarction of <7 days' duration (Fig. 368-2H) and is also sensitive to encephalitis and abscess formation, which have reduced diffusion and result in high signal on diffusion-weighted images (Fig. 368-3B).

Perfusion MRI involves the acquisition of EPI images during a rapid intravenous bolus of gadolinium contrast material. Relative perfusion abnormalities can be identified on images of the relative cerebral blood volume, mean transit time, and cerebral blood flow. Delay in mean transit time and reduction in cerebral blood volume and cerebral blood flow are typical of infarction. In the setting of reduced blood flow, a prolonged mean transit time of contrast but normal or elevated cerebral blood volume may indicate tissue supplied by collateral flow that is at risk of infarction. Perfusion MRI imaging can also be used in the assessment of brain tumors to differentiate intraaxial primary tumors from extraaxial tumors or metastasis.

Diffusion tensor imaging (DTI) is a diffusion MRI technique that assesses the direction of microscopic motion of water along white matter tracts. This technique has great potential in the assessment of brain maturation as well as disease entities that undermine the integrity of the white matter architecture. It has proven valuable in preoperative assessment of subcortical white matter tract anatomy prior to brain tumor surgery **(Fig. 368-6)**.

Functional MRI of the brain is an EPI technique that localizes regions of activity in the brain following task activation. Neuronal activity elicits a slight increase in the delivery of oxygenated blood flow to a specific region of activated brain. This results in an alteration in the balance of oxyhemoglobin and deoxyhemoglobin, which yields a 2–3% increase in signal intensity within veins and local capillaries. Further studies will determine whether these techniques are cost-effective or clinically useful, but currently preoperative somatosensory and auditory cortex localization is possible. This technique has proved useful to neuroscientists interested in interrogating the localization of certain brain functions.

MAGNETIC RESONANCE NEUROGRAPHY

MRN is a T2 weighted MR technique that shows promise in detecting increased signal in irritated, inflamed, or infiltrated peripheral nerves. Images are obtained with fat-suppressed fast spin echo imaging or short inversion recovery sequences. Irritated or infiltrated nerves will demonstrate high signal on T2W imaging. This is indicated in patients with radiculopathy whose conventional MR studies of the spine are normal, or in those suspected of peripheral nerve entrapment or trauma.

POSITRON EMISSION TOMOGRAPHY (PET)

PET relies on the detection of positrons emitted during the decay of a radionuclide that has been injected into a patient. The most frequently used moiety is 2-[^{18}F]fluoro-2-deoxy-d-glucose (FDG), which is an analogue of glucose and is taken up by cells competitively with 2-deoxyglucose. Multiple images of glucose uptake activity are formed after 45–60 minutes. Images reveal differences in regional glucose activity among normal and pathologic brain structures. A lower activity of FDG in the parietal lobes has been associated with Alzheimer's disease. FDG PET is used primarily for the detection of extracranial metastatic disease. Combination PET-CT scanners, in which both CT and PET are obtained at one sitting, are replacing PET scans alone for most clinical indications. Functional images superimposed on high-resolution CT scans result in more precise anatomic diagnoses.

MYELOGRAPHY

Technique

Myelography involves the intrathecal instillation of specially formulated water-soluble iodinated contrast medium into the lumbar or cervical subarachnoid space. CT scanning is usually performed after myelography (*CT myelography*) to better demonstrate the spinal cord and roots, which appear as filling defects in the opacified subarachnoid space. *Low-dose CT myelography*, in which CT is performed after the subarachnoid injection of a small amount of relatively dilute contrast material, has replaced conventional myelography for many indications, thereby reducing exposure to radiation and contrast media. Newer multidetector scanners now obtain CT studies quickly so that reformations in sagittal and coronal planes, equivalent to traditional myelography projections, are now routine.

Indications

Myelography has been largely replaced by CT myelography and MRI for diagnosis of diseases of the spinal canal and cord (Table 368-1). Remaining indications for conventional plain-film myelography include the evaluation of suspected meningeal or arachnoid cysts and the localization of spinal dural arteriovenous or CSF fistulas. Conventional myelography and CT myelography provide the most

precise information in patients with prior spinal fusion and spinal fixation hardware.

Contraindications

Myelography is relatively safe; however, it should be performed with caution in any patient with elevated intracranial pressure, evidence of a spinal block, or a history of allergic reaction to intrathecal contrast media. In patients with a suspected spinal block, MR is the preferred technique. If myelography is necessary, only a small amount of contrast medium should be instilled below the lesion in order to minimize the risk of neurologic deterioration. Lumbar puncture is to be avoided in patients with bleeding disorders, including patients receiving anticoagulant therapy, as well as in those with infections of the overlying soft tissues.

Complications

Headache, nausea, and vomiting are the most frequent complications of myelography and are reported to occur in up to 38% of patients. These symptoms result from either neurotoxic effects of the contrast agent, persistent leakage of CSF at the puncture site, or psychological reactions to the procedure. Vasovagal syncope may occur during lumbar puncture; it is accentuated by the upright position used during lumbar myelography. Adequate hydration before and after myelography will reduce the incidence of this complication. Postural headache (post–lumbar puncture headache) is generally due to leakage of CSF from the puncture site, resulting in CSF hypotension. **Management of postlumbar-puncture headache is discussed in Chap. 14.**

If significant headache persists for longer than 48 hours, placement of an epidural blood patch should be considered. Hearing loss is a rare complication of myelography. It may result from a direct toxic effect of the contrast medium or from an alteration of the pressure equilibrium between CSF and perilymph in the inner ear. Puncture of the spinal cord is a rare but serious complication of cervical (C1–2) or high lumbar puncture. The risk of cord puncture is greatest in patients with spinal stenosis, Chiari malformations, or conditions that reduce CSF volume. In these settings, a low-dose lumbar injection followed by thin-section CT or MRI is a safer alternative to cervical puncture. Intrathecal contrast reactions are rare, but aseptic meningitis and encephalopathy may occur. The latter is usually dose related and associated with contrast entering the intracranial subarachnoid space. Seizures occur following myelography in 0.1–0.3% of patients. Risk factors include a preexisting seizure disorder and the use of a total iodine dose of >4500 mg. Other reported complications include hyperthermia, hallucinations, depression, and anxiety states. These side effects have been reduced by the development of nonionic, water-soluble contrast agents as well as by head elevation and generous hydration following myelography.

SPINE INTERVENTIONS

Diskography

The evaluation of back pain and radiculopathy may require diagnostic procedures that attempt either to reproduce the patient's pain or relieve it, indicating its correct source prior to lumbar fusion. Diskography is performed by fluoroscopic placement of a 22- to 25-gauge needle into the intervertebral disk and subsequent injection of 1–3 mL of contrast media. The intradiskal pressure is recorded, as is an assessment of the patient's response to the injection of contrast material. Typically little or no pain is felt during injection of a normal disk, which does not accept much more than 1 mL of contrast material, even at pressures as high as 415–690 kPa (60–100 lb/in^2). CT and plain films are obtained following the procedure. Concerns have been raised that diskography may contribute to an accelerated rate of disk degeneration.

Selective Nerve Root and Epidural Spinal Injections

Percutaneous selective nerve root and epidural blocks with glucocorticoid and anesthetic mixtures may be both therapeutic as well as diagnostic, especially if a patient's pain is relieved. Typically, 1–2 mL of an equal mixture of a long-acting glucocorticoid such as betamethasone and a long-acting anesthetic such as bupivacaine 0.75% is instilled under CT or fluoroscopic guidance in the intraspinal epidural space or adjacent to an existing nerve root.

ANGIOGRAPHY

Catheter angiography is indicated for evaluating intracranial small-vessel pathology (such as vasculitis), for assessing vascular malformations and aneurysms, and in endovascular therapeutic procedures (Table 368-1). Angiography has been replaced for many indications by CT/CTA or MRI/MRA.

Angiography carries the greatest risk of morbidity of all diagnostic imaging procedures, owing to the necessity of inserting a catheter into a blood vessel, directing the catheter to the required location, injecting contrast material to visualize the vessel, and removing the catheter while maintaining hemostasis. Therapeutic transcatheter procedures (see below) have become important options for the treatment of some cerebrovascular diseases. The decision to undertake a diagnostic or therapeutic angiographic procedure requires careful assessment of the goals of the investigation and its attendant

risks.

To improve tolerance to contrast agents, patients undergoing angiography should be well hydrated before and after the procedure. Since the femoral route is used most commonly, the femoral artery must be compressed after the procedure to prevent a hematoma from developing. The puncture site and distal pulses should be evaluated carefully after the procedure; complications can include thigh hematoma or lower extremity emboli.

Complications

A common femoral arterial puncture provides retrograde access via the aorta to the aortic arch and great vessels. The most feared complication of cerebral angiography is stroke. Thrombus can form on or inside the tip of the catheter, and atherosclerotic thrombus or plaque can be dislodged by the catheter or guidewire or by the force of injection and can embolize distally in the cerebral circulation. Risk factors for ischemic complications include limited experience on the part of the angiographer, atherosclerosis, vasospasm, low cardiac output, decreased oxygen-carrying capacity, advanced age, and prior history of migraine. The risk of a neurologic complication varies but is ~4% for transient ischemic attack and stroke, 1% for permanent deficit, and <0.1% for death.

Ionic contrast material injected into the cerebral vasculature can be neurotoxic if the BBB is breached, either by an underlying disease or by the injection of hyperosmolar contrast agent. Ionic contrast media are less well tolerated than nonionic media, probably because they can induce changes in cell membrane electrical potentials. Patients with dolichoectasia of the basilar artery can suffer reversible brainstem dysfunction and acute short-term memory loss during angiography, owing to the slow percolation of the contrast material and the consequent prolonged exposure of the brain. Rarely, an intracranial aneurysm ruptures during an angiographic contrast injection, causing subarachnoid hemorrhage, perhaps as a result of injection under high pressure.

Spinal Angiography

Spinal angiography may be indicated to evaluate vascular malformations and tumors and to identify the artery of Adamkiewicz (Chap. 377) prior to aortic aneurysm repair. The procedure is lengthy and requires the use of relatively large volumes of contrast; the incidence of serious complications, including paraparesis, subjective visual blurring, and altered speech, is ~2%. Gadolinium-enhanced MRA has been used successfully in this setting, as has iodinated contrast CTA, which has promise for replacing diagnostic spinal angiography for some indications.

INTERVENTIONAL NEURORADIOLOGY

This rapidly developing field is providing new therapeutic options for patients with challenging neurovascular problems. Available procedures include detachable coil therapy for aneurysms, particulate or liquid adhesive embolization of arteriovenous malformations, balloon angioplasty and stenting of arterial stenosis or vasospasm, transarterial or transvenous embolization of dural arteriovenous fistulas, balloon occlusion of carotid-cavernous and vertebral fistulas, endovascular treatment of vein-of-Galen malformations, preoperative embolization of tumors, and thrombolysis of acute arterial or venous thrombosis. Many of these disorders place the patient at high risk of cerebral hemorrhage, stroke, or death.

The highest complication rates are found with the therapies designed to treat the highest-risk diseases. The advent of electrolytically detachable coils has ushered in a new era in the treatment of cerebral aneurysms. One randomized trial found a 28% reduction of morbidity and mortality at 1 year among those treated for anterior circulation aneurysm with detachable coils compared with neurosurgical clipping. It remains to be determined what the role of coils will be relative to surgical options, but in many centers, coiling has become standard therapy for many aneurysms.

FURTHER READINGS

Agarwal R et al: Gadolinium-based contrast agents and nephrogenic systemic fibrosis: A systematic review and meta-analysis. Nephrol Dial Transplant 24:856, 2009[PMID: 18952698] [Full Text]

Berman JI et al: Diffusion-tensor imaging-guided tracking of fibers of the pyramidal tract combined with intraoperative cortical stimulation mapping in patients with gliomas. J Neurosurg 101:66, 2004[PMID: 15255253] [Full Text]

Carragee EJ et al: 2009 ISSLS Prize Winner: Does discography cause accelerated progression of degeneration changes in the lumbar disc: A ten-year matched cohort study. Spine 34:2338, 2009[PMID: 19755936] [Full Text]

Gonzalez RG: Imaging-guided acute ischemic stroke therapy: From "time is brain" to "physiology is brain." Am J Neuroradiol 27:728, 2006[PMID: 16611754] [Full Text]

Schaefer PW: Diffusion-weighted imaging in acute stroke. Magn Reson Imaging Clin N Am 14:141,

2006[PMID: 16873008] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ATLAS OF NEUROIMAGING: INTRODUCTION

Figure e44-1

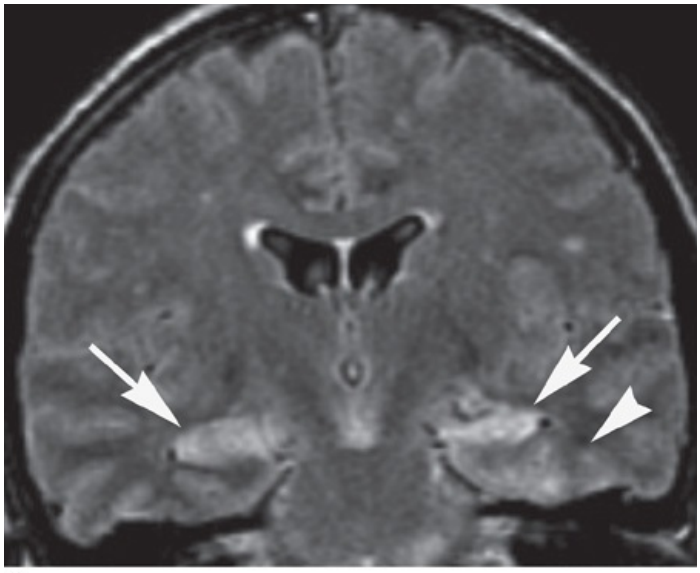

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

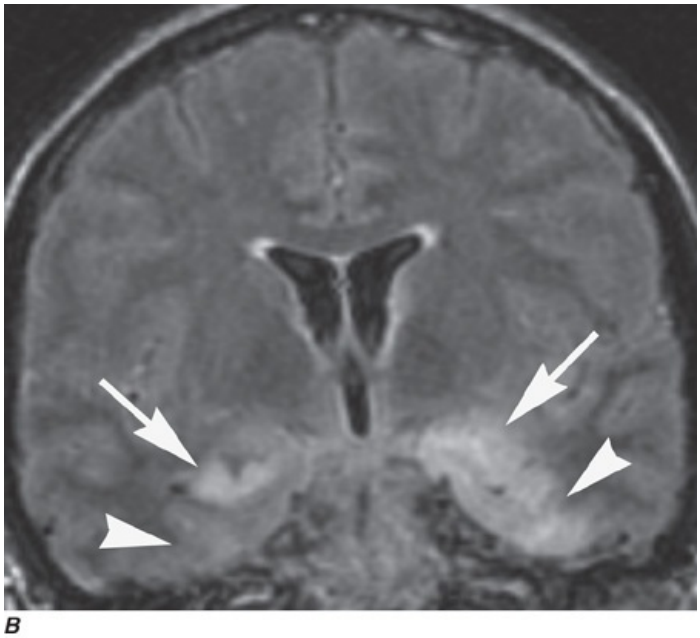

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

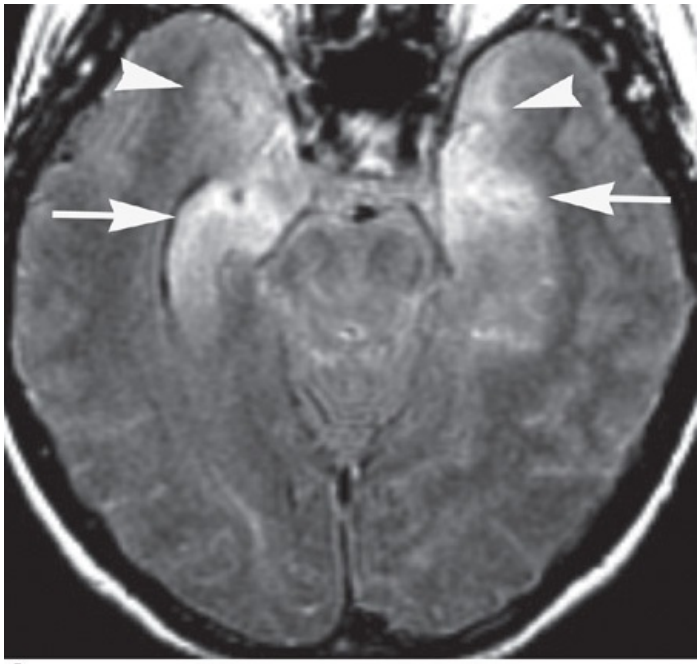

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

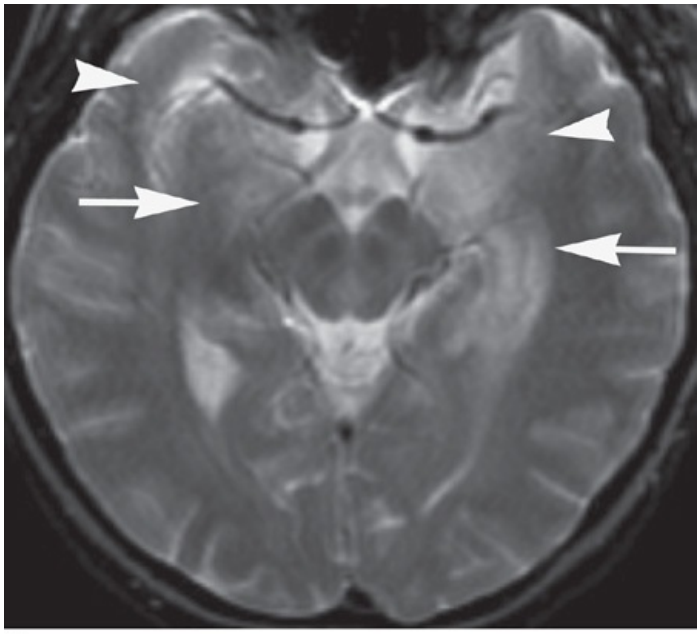

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Limbic encephalitis (Chap. 101) Coronal (***A***, ***B***), axial fluid-attenuated inversion recovery (FLAIR) (***C***, ***D***), and axial T2-weighted (***E***) MR images demonstrate abnormal high signal involving the bilateral mesial temporal lobes (*arrowheads*) including the hippocampi (left greater than right) without significant mass effect (*arrows*). There was no enhancement on postgadolinium images (not shown).

Figure e44-2

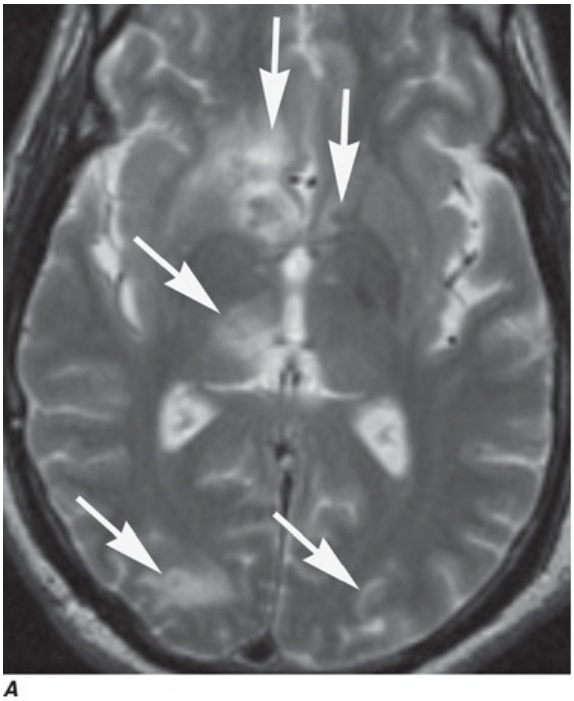

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

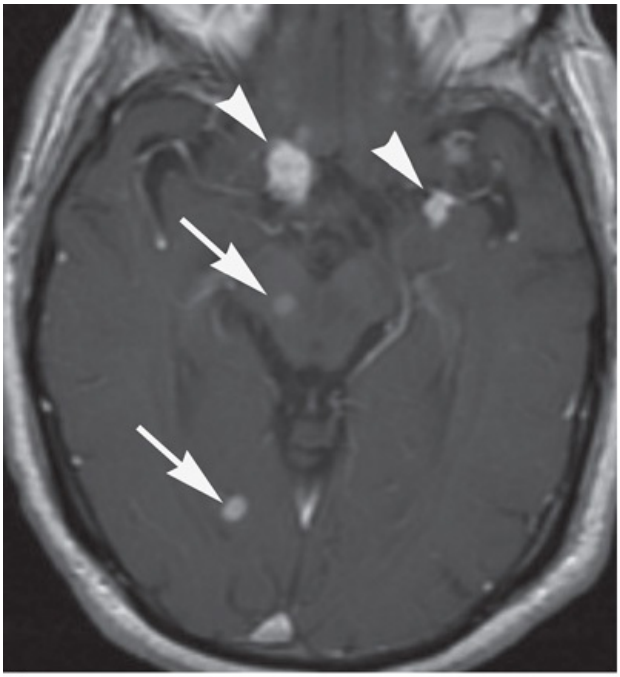

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

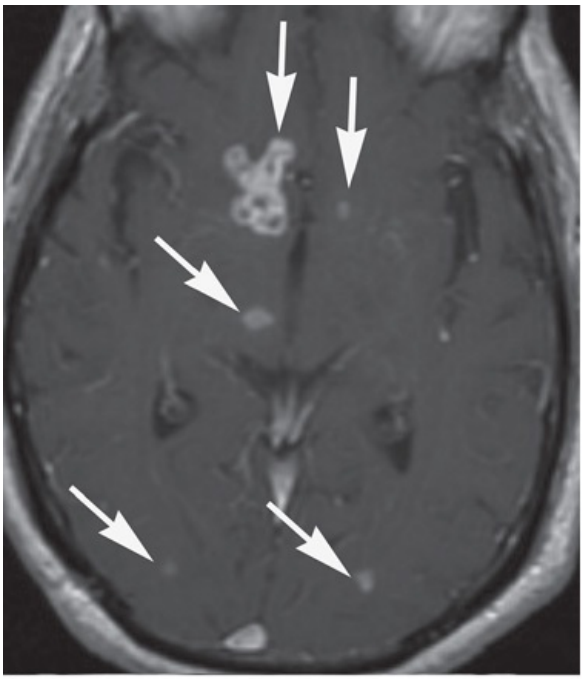

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CNS tuberculosis (Chap. 165)

Axial T2-weighted MRI (***A***) demonstrates multiple lesions (*arrows*) with peripheral high signal and central low signal, located predominantly in the cortex and subcortical white matter, as well as in the basal ganglia.

Axial T1-weighted MR images postgadolinium (***B***, ***C***) demonstrate ring enhancement of the lesions (*arrows*) and additional lesions in the subarachnoid space (*arrowheads*).

Sagittal T2-weighted MR image of the cervical spine (***D***) demonstrates a hypointense lesion in the subarachnoid space at the level of T5 (*arrow*).

Sagittal T1-weighted MR image postgadolinium of the cervical spine (***E***) demonstrates enhancement of the lesion in the subarachnoid space at the level of T5 (*arrow*).

Figure e44-3

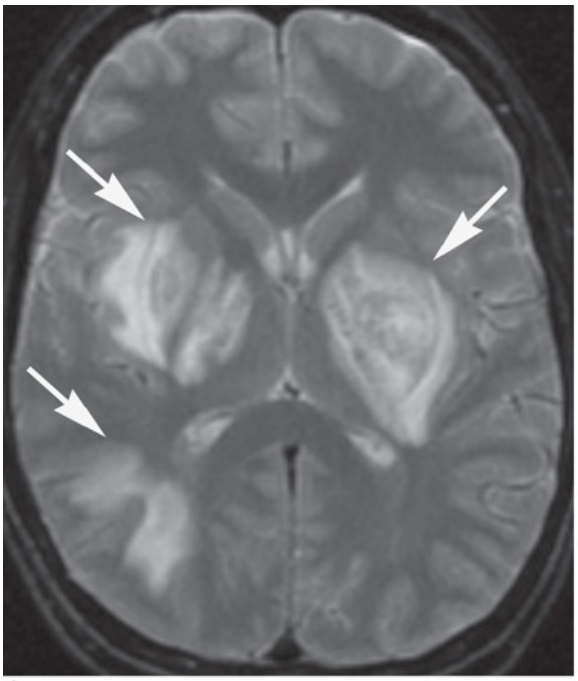

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

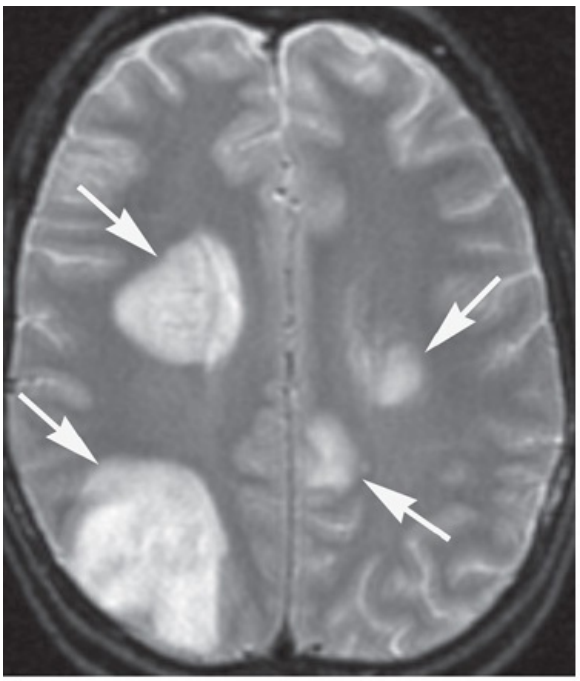

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

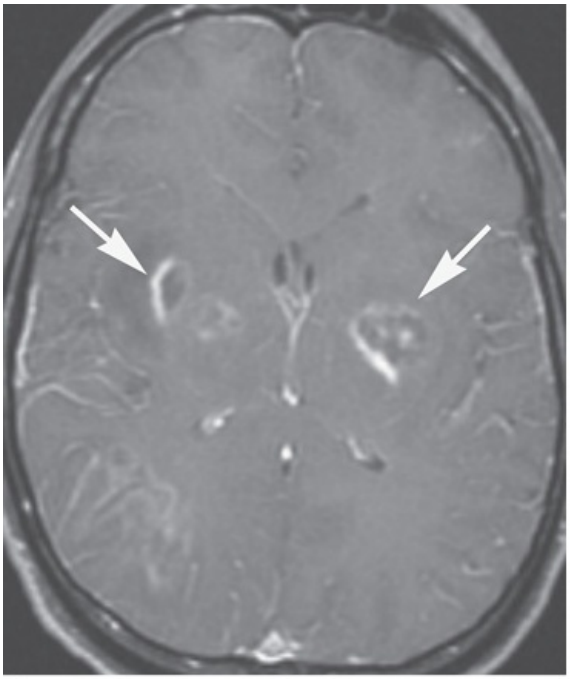

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

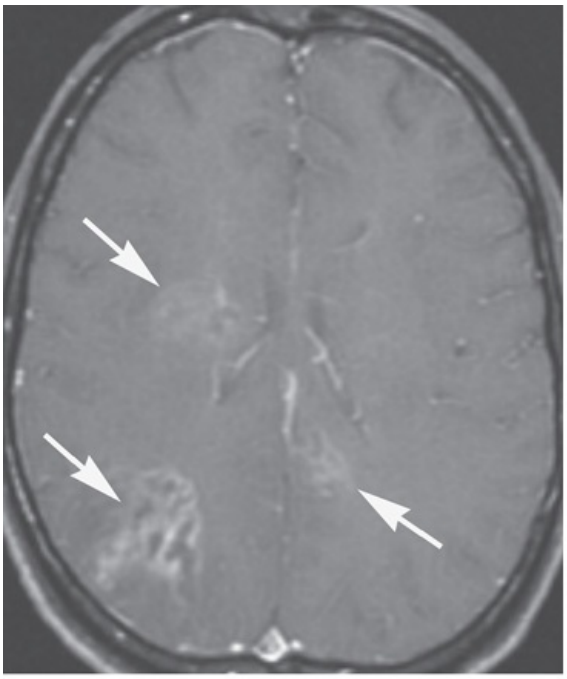

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

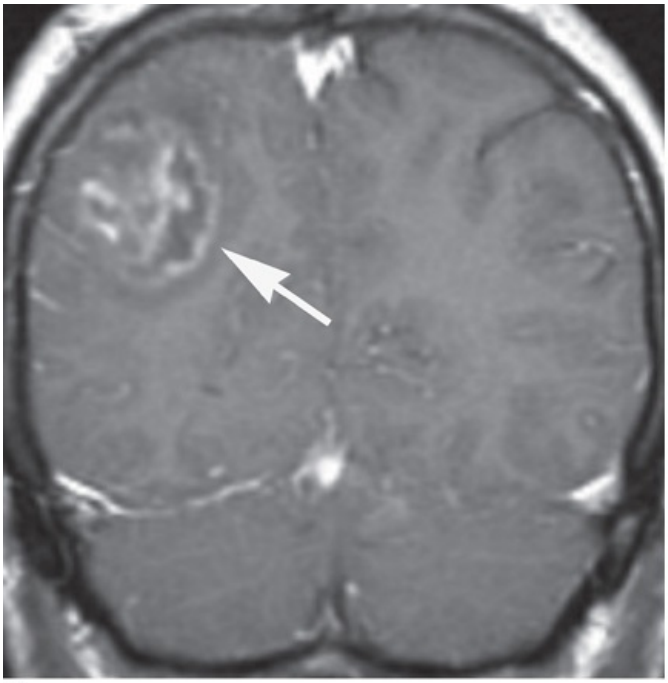

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

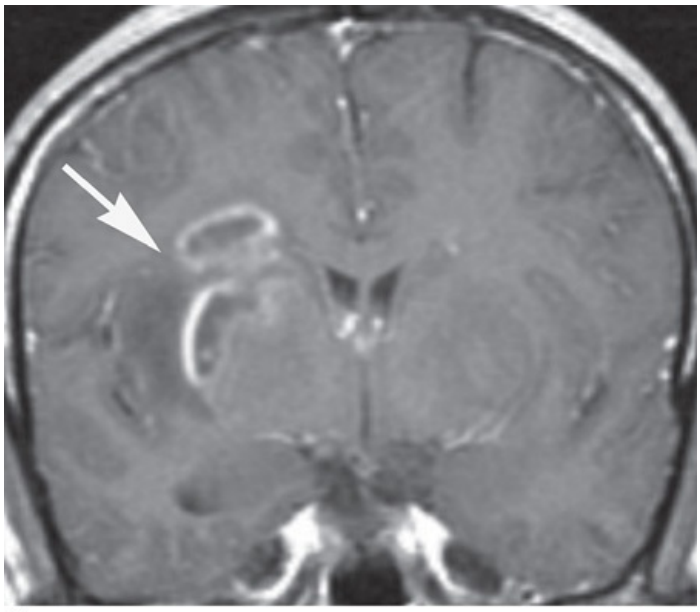

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurosyphilis (Chap. 169)

Case I

Axial T2-weighted MR images (***A***, ***B***) demonstrate well-defined areas of abnormal high signal in the basal ganglia bilaterally and in a wedge-shaped distribution in the right parietal lobe (*arrows*).

Axial (***C***, ***D***) T1-weighted images postgadolinium.

Coronal (***E***, ***F***) T1-weighted images postgadolinium demonstrate irregular ring enhancement of the lesions (*arrows*).

Figure e44-4

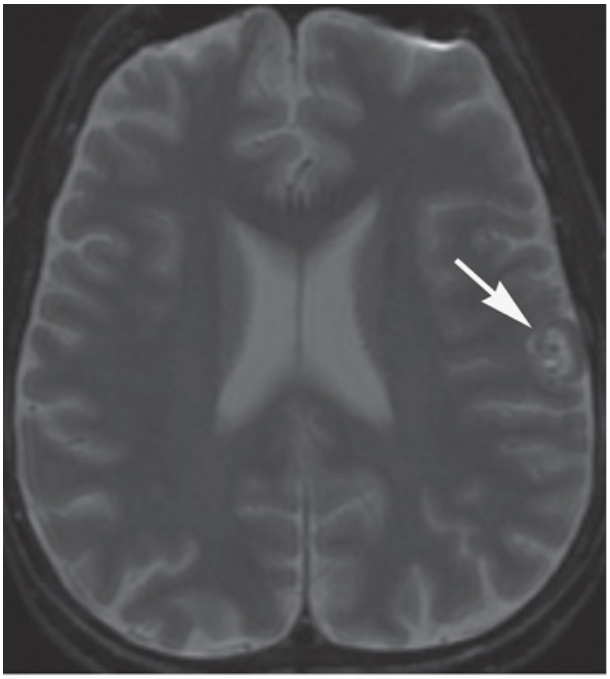

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

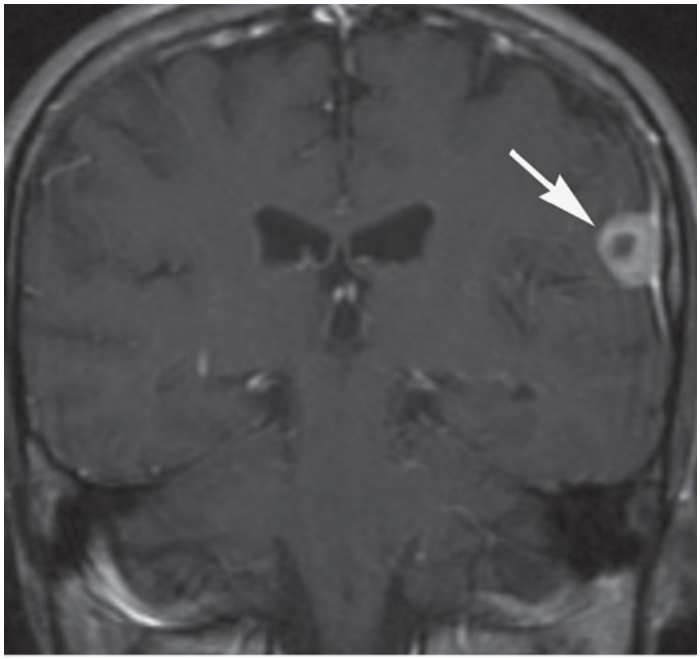

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurosyphilis (Chap. 169)

Case II

Axial T2-weighted MRI (***A***) demonstrates a dural-based, peripherally hyperintense and centrally hypointense lesion located lateral to the left frontal lobe (*arrow*).

Axial (***B***) and coronal (***C***) T1-weighted MR images postgadolinium demonstrate peripheral enhancement of the lesion (*arrows*).

Figure e44-5

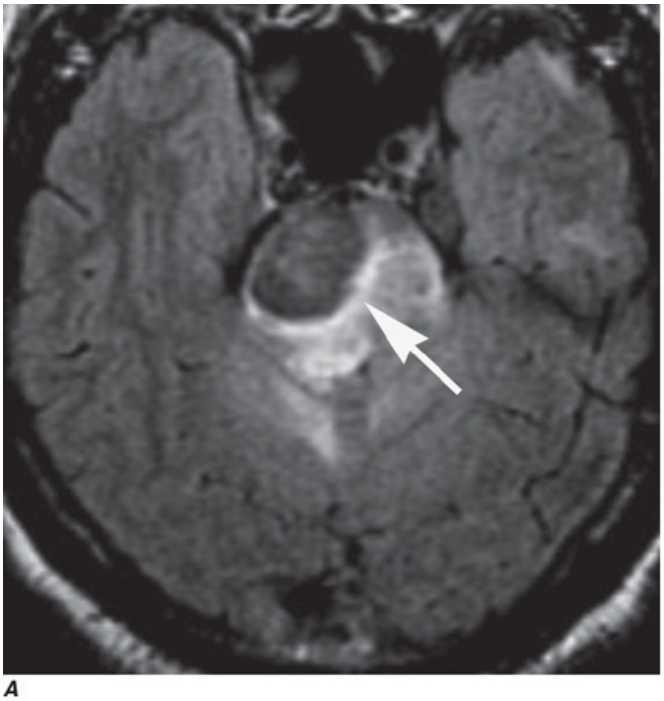

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

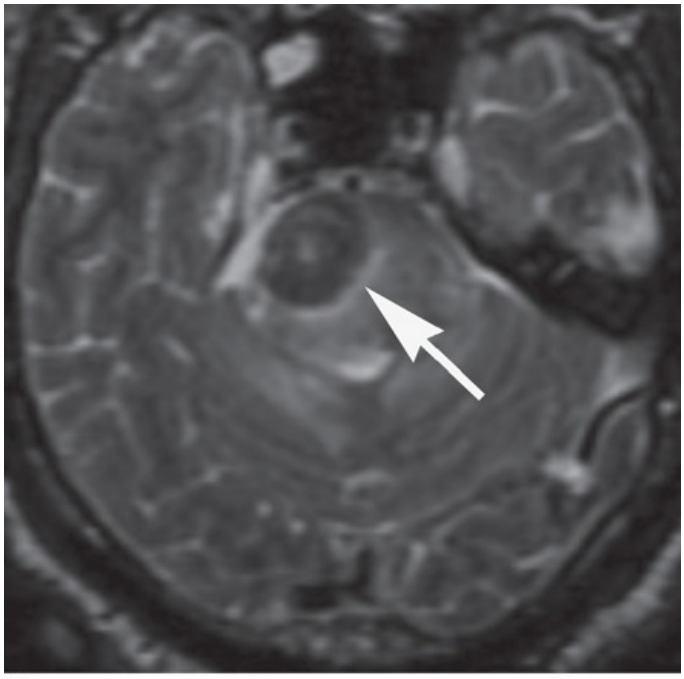

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

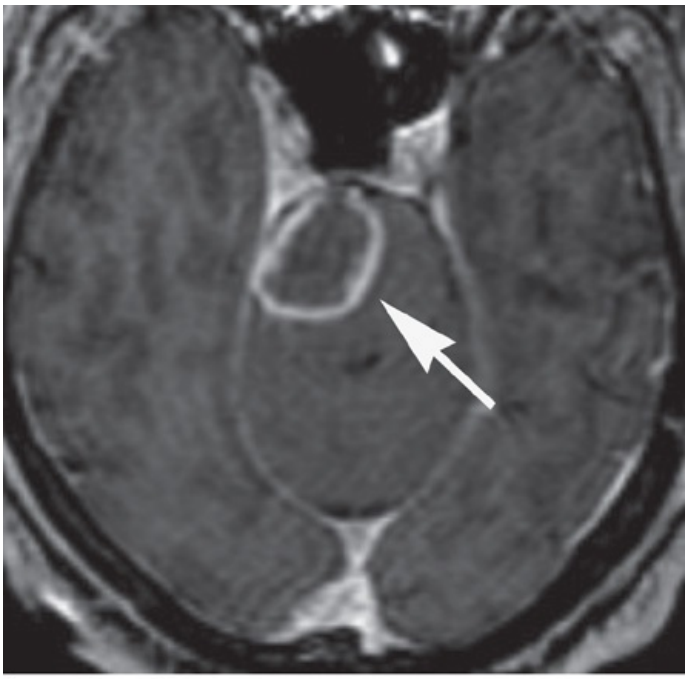

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histoplasmosis of the pons (Chap. 199)

Axial FLAIR (***A***) and T2-weighted (***B***) MR images demonstrate a low signal mass in the right pons (*arrows*) with surrounding vasogenic edema.

Axial T1-weighted MR image postgadolinium (***C***) demonstrates ring enhancement of the lesion in the right pons (*arrow*). Of note, there was no evidence of restricted diffusion (not shown).

Figure e44-6

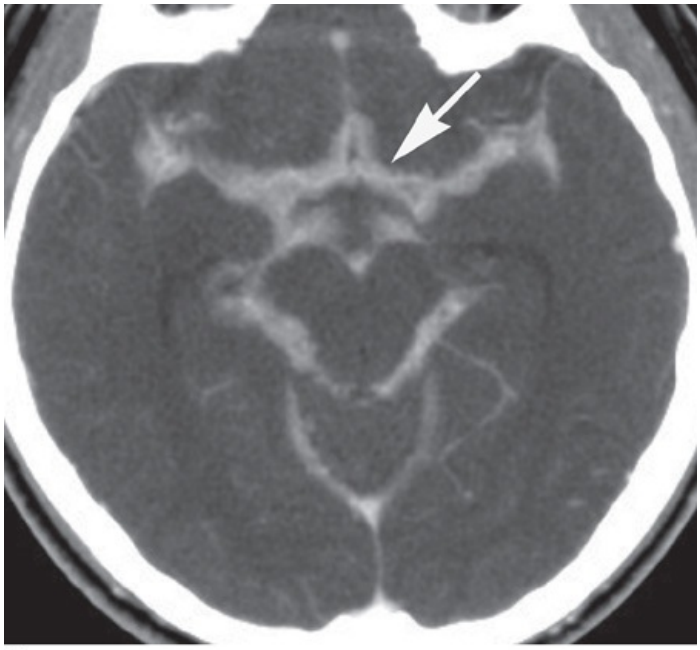

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

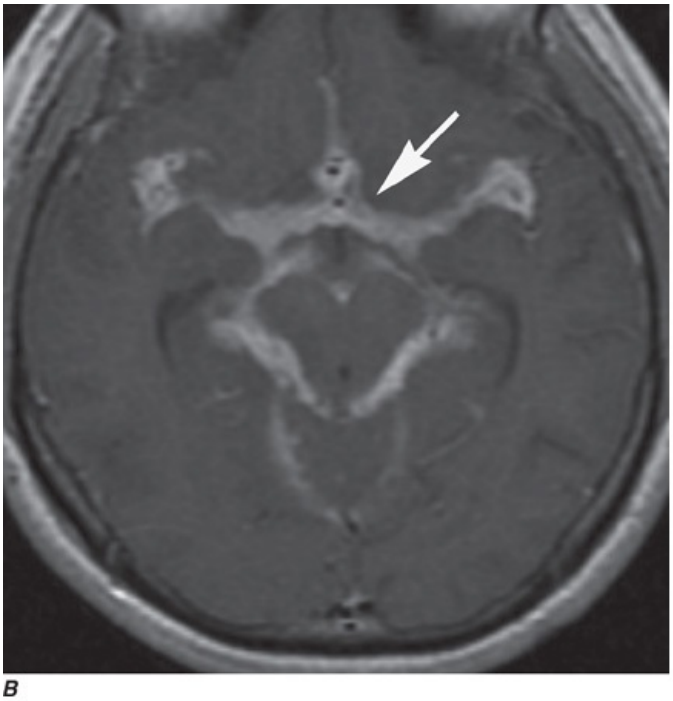

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

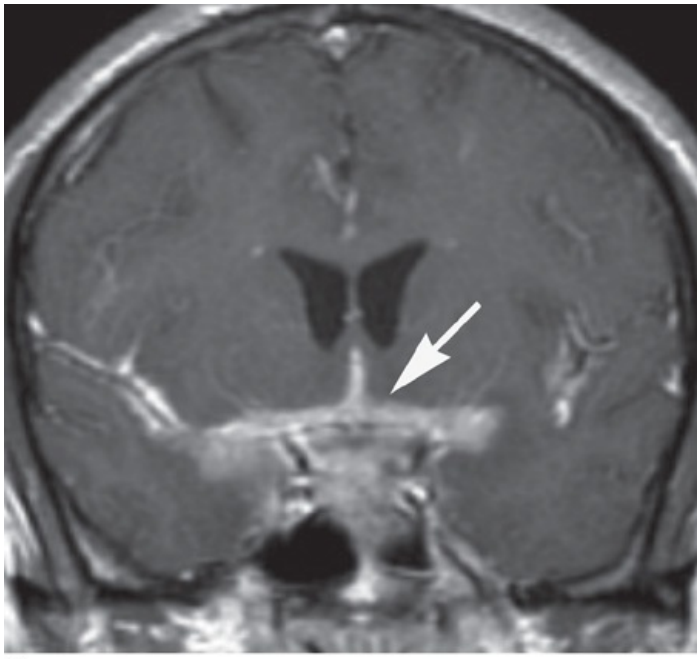

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coccidiomycosis meningitis (Chap. 200)

Axial postcontrast CT (***A***) and axial (***B***) and coronal (***C***) T1-weighted MR images postgadolinium demonstrate enhancement of the perimesencephalic cisterns (*arrows*), as well as the sylvian and interhemispheric fissures.

Figure e44-7

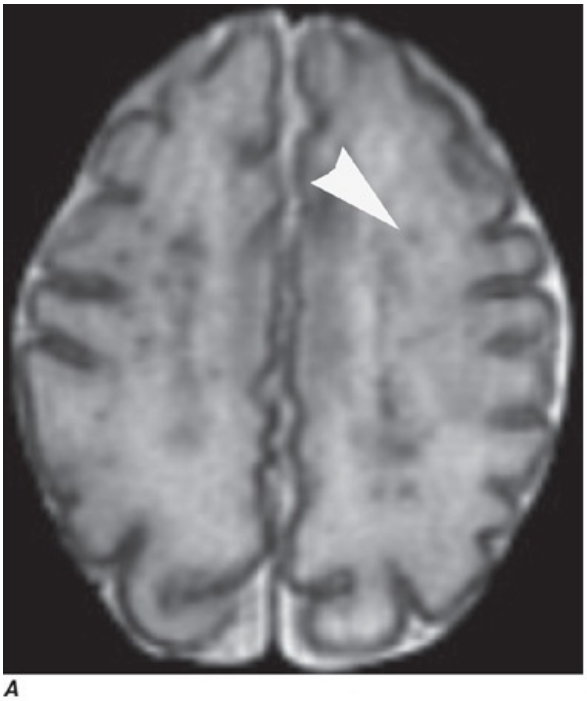

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

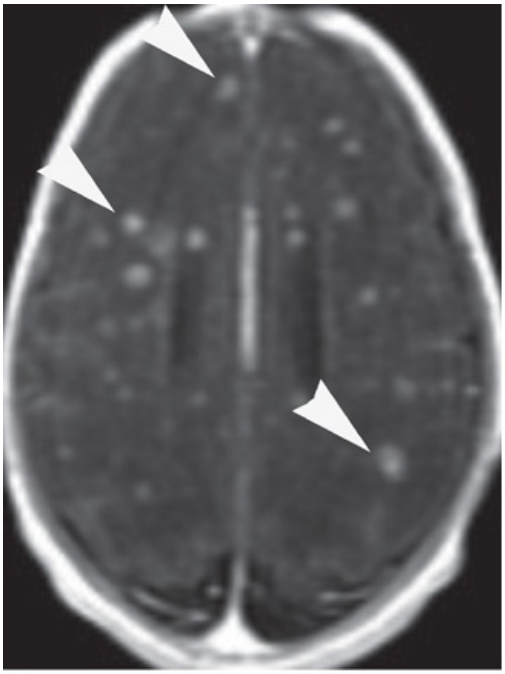

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

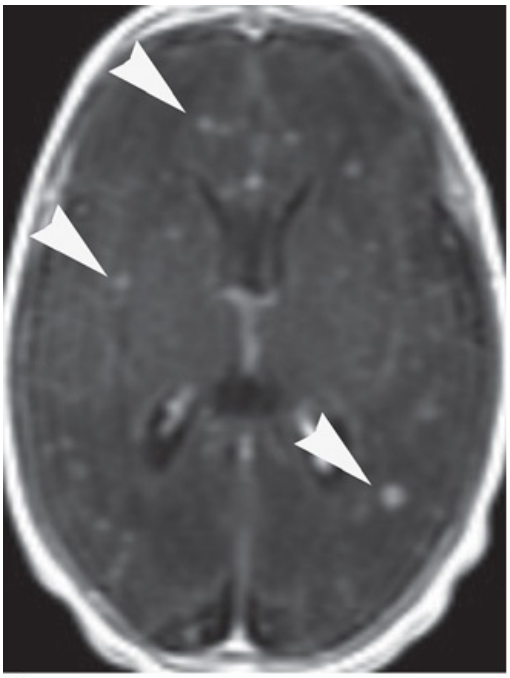

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

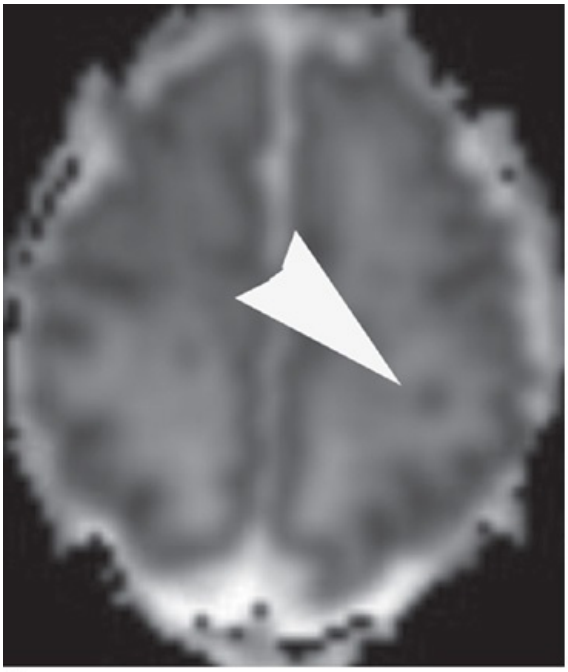

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

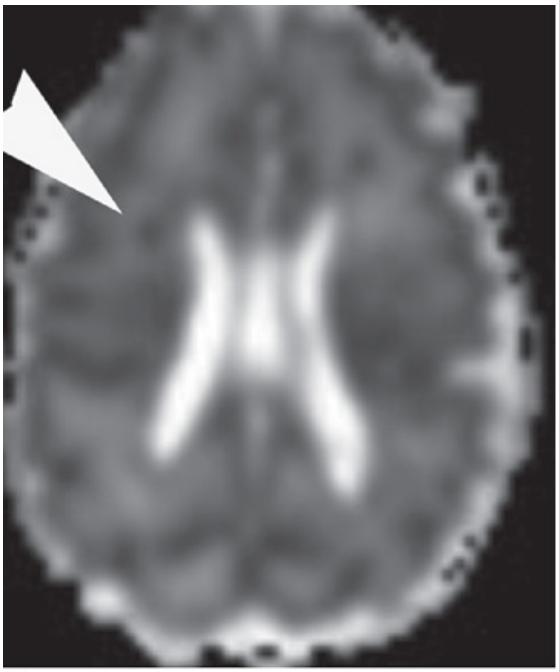

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Candidiasis in a newborn (Chap. 203)

Axial T2-weighted MR image (***A***) demonstrates multiple punctate foci of low signal diffusely distributed in the brain parenchyma (*arrowhead*).

Axial T1-weighted MR images postgadolinium (***B***, ***C***) demonstrate marked enhancement of the lesions (*arrowheads*).

ADC map (***D***, ***E***) demonstrates restricted diffusion of water molecules in the lesions (*arrowheads*).

Figure e44-8

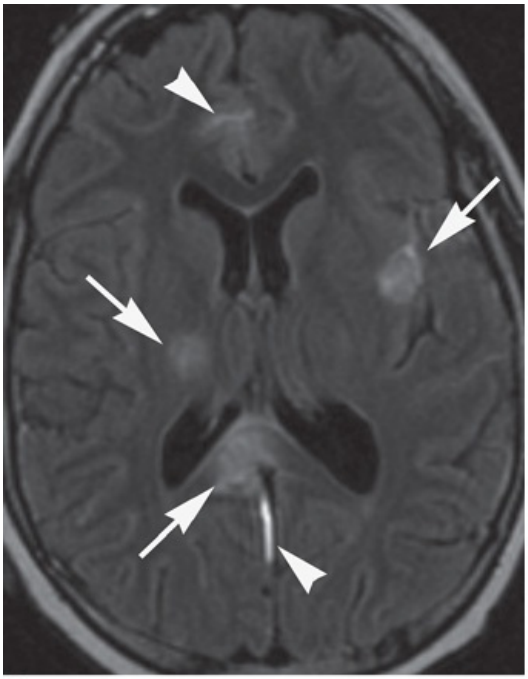

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

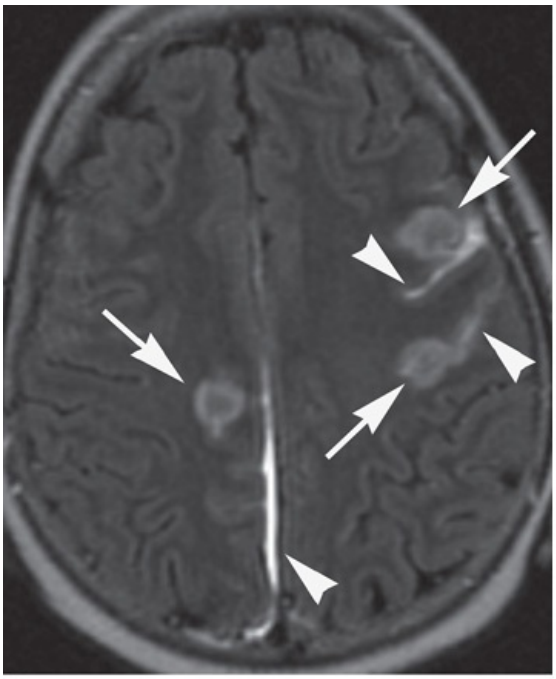

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

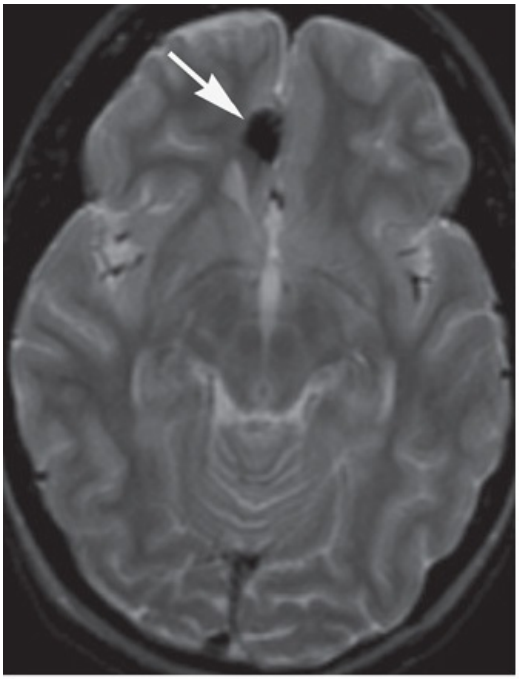

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

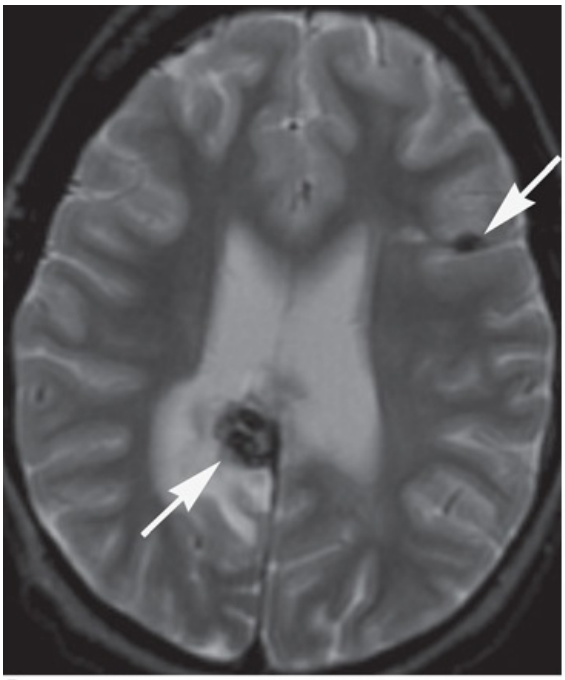

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

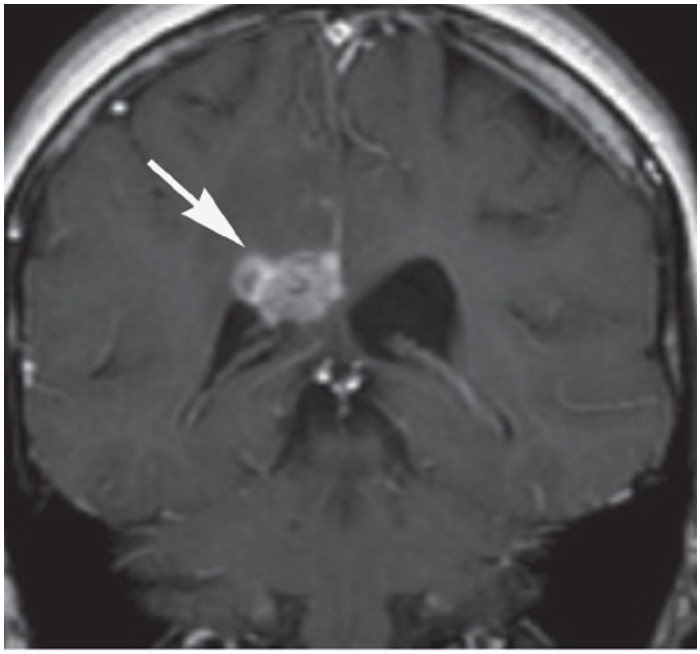

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

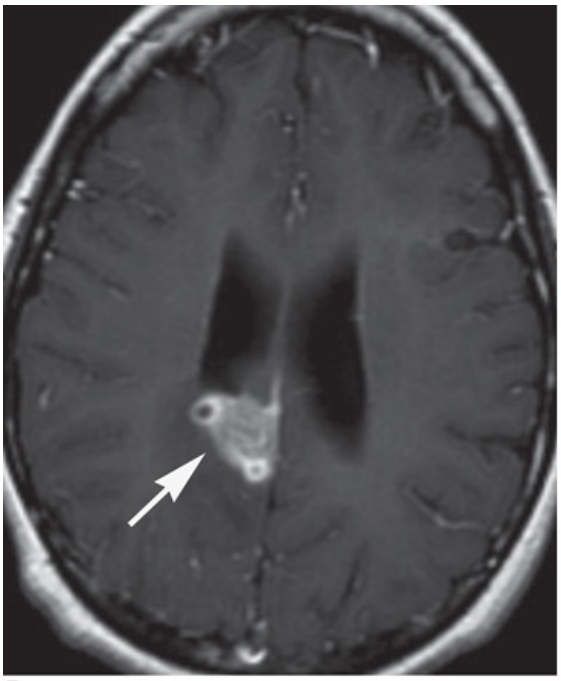

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CNS aspergillosis (Chap. 204)

Axial FLAIR MR images (***A***, ***B***) demonstrate multiple areas of abnormal high signal in the basal ganglia as well as cortex and subcortical white matter (*arrows*). There is also abnormal high signal in the subarachnoid space adjacent to the lesions (*arrowheads*) that can correspond to blood or high protein content.

Axial T2-weighted MR images (***C***, ***D***) demonstrate intrinsic low signal in the lesions (*arrows*), suggesting the presence of blood products. Some of the lesions also show vasogenic edema. Coronal (***E***) and axial (***F***) T1-weighted MR images postgadolinium demonstrate peripheral enhancement of the lesions (*arrows*).

Figure e44-9

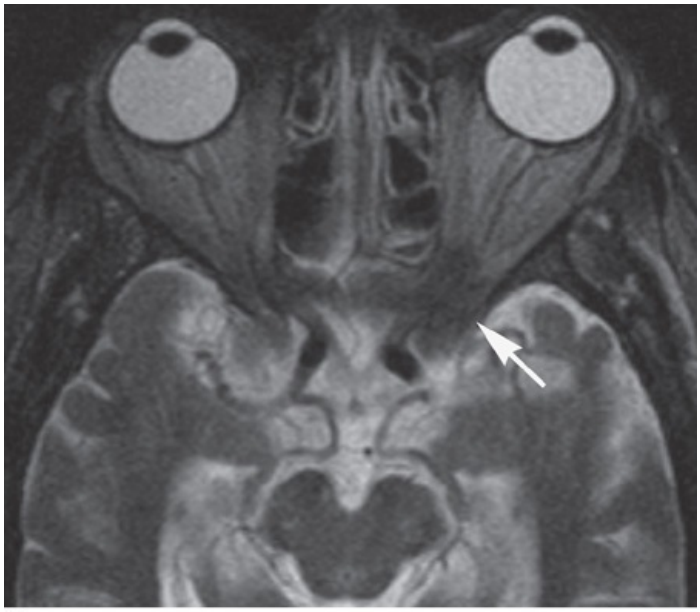

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

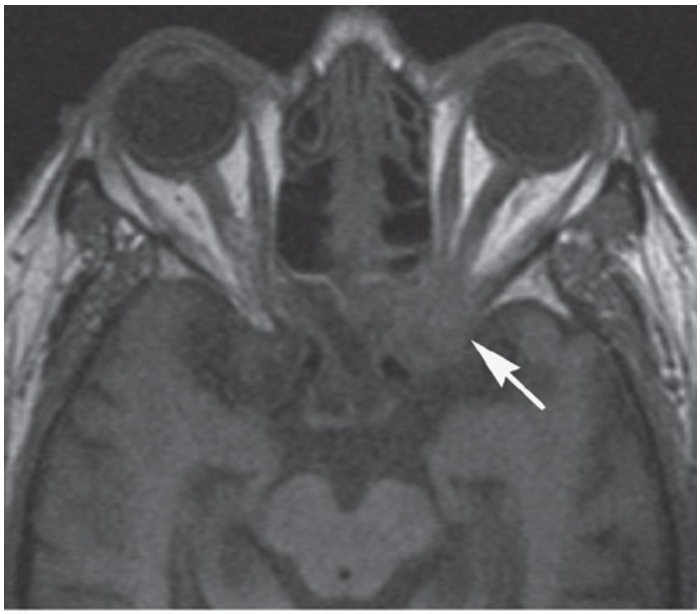

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

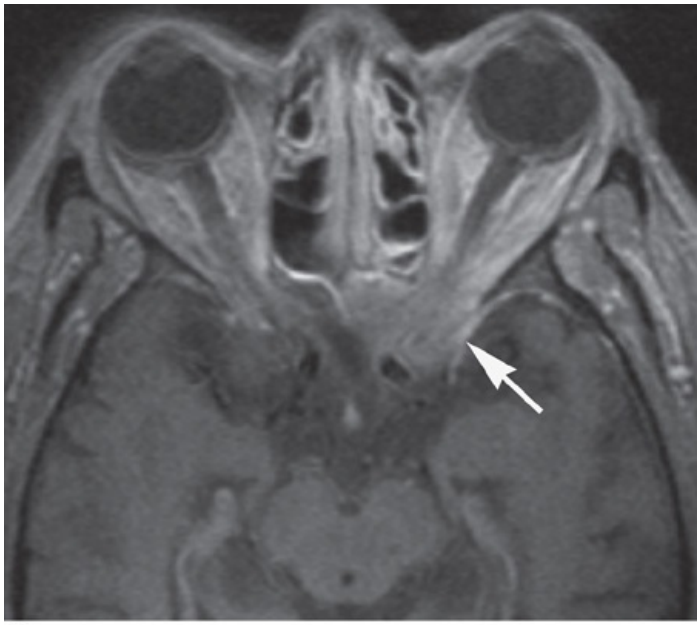

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Invasive sinonasal aspergillosis (Chap. 204)

Axial T2-weighted MR image (***A***) demonstrates an irregularly shaped low signal lesion involving the left orbital apex (*arrow*).

B. T1-weighted image pregadolinium demonstrates low signal in left anterior clinoid process (*arrow*).

C. T1-weighted image postgadolinium demonstrates enhancement of lesion (*arrow*).

Figure e44-10

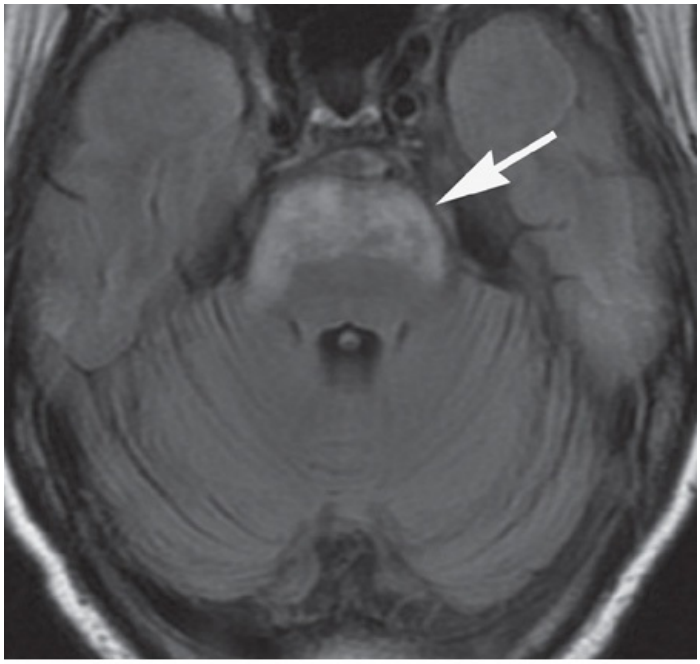

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Behçet's disease (Chap. 327)

Axial FLAIR MRI demonstrates abnormal high signal involving the anterior pons (*arrow*); following gadolinium administration, the lesion was nonenhancing (not shown). Brainstem lesions are typical of Behçet's disease, caused primarily by vasculitis and in some cases demyelinating lesions.

Figure e44-11

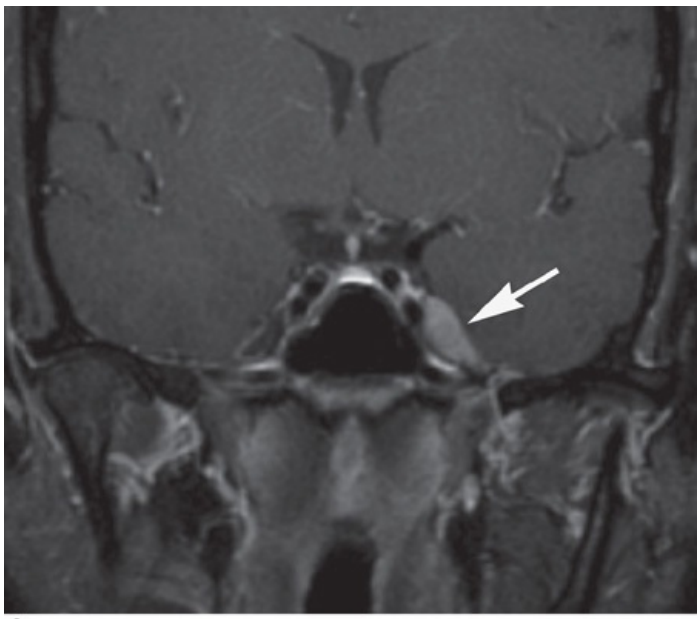

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

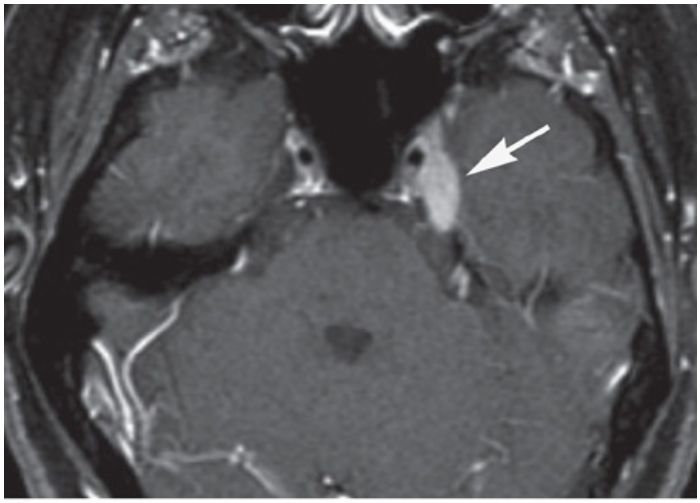

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurosarcoid (Chap. 329)

Case I

Coronal (***A***) and axial (***B***) T1-weighted images postgadolinium with fat suppression demonstrate a homogeneously enhancing well-circumscribed mass centered in the left Meckel's cave (*arrows*).

Figure e44-12

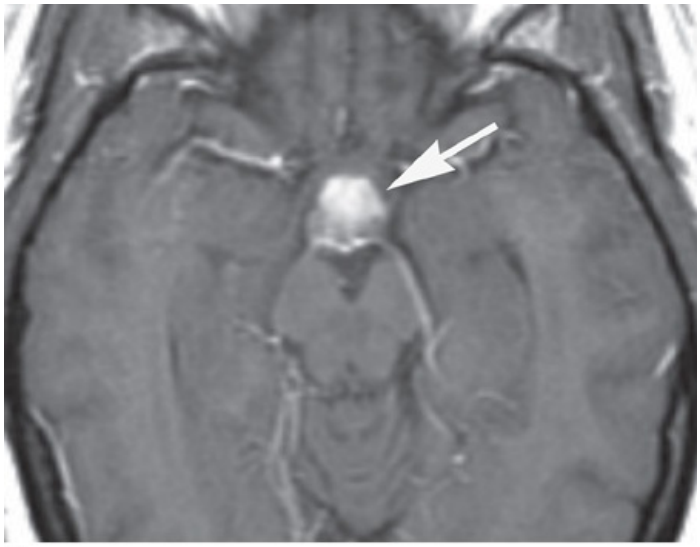

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

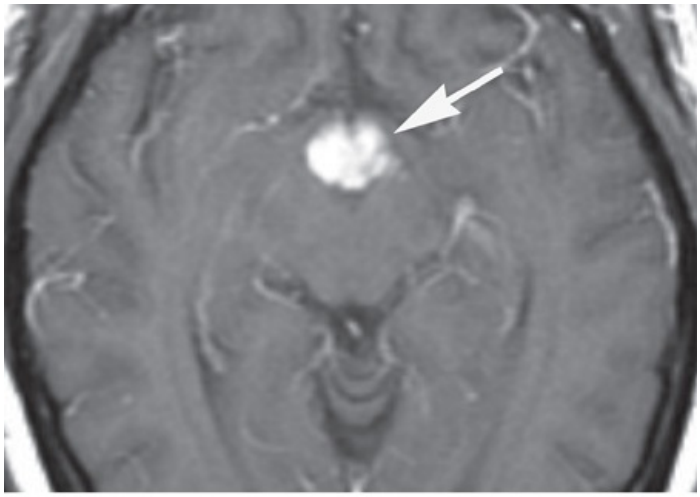

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

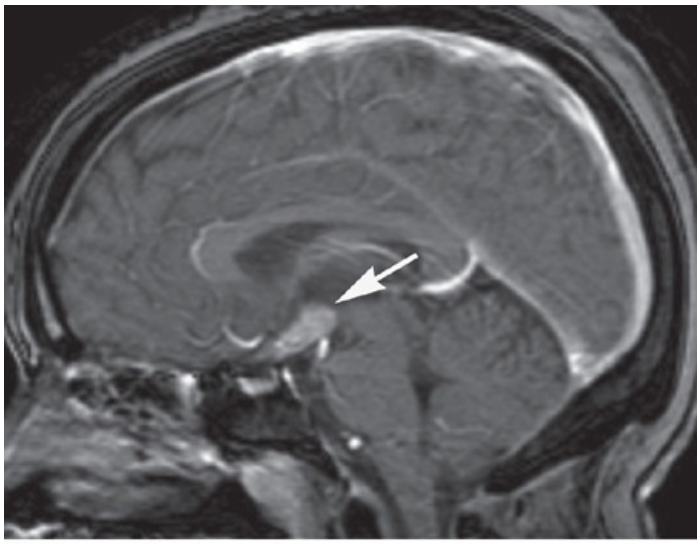

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurosarcoid (Chap. 329)

Case II

Axial (***A***, ***B***) and sagittal (***C***) T1-weighted images postgadolinium with fat suppression demonstrate a homogeneously enhancing mass involving the hypothalamus and the pituitary stalk (*arrows*).

Figure e44-13

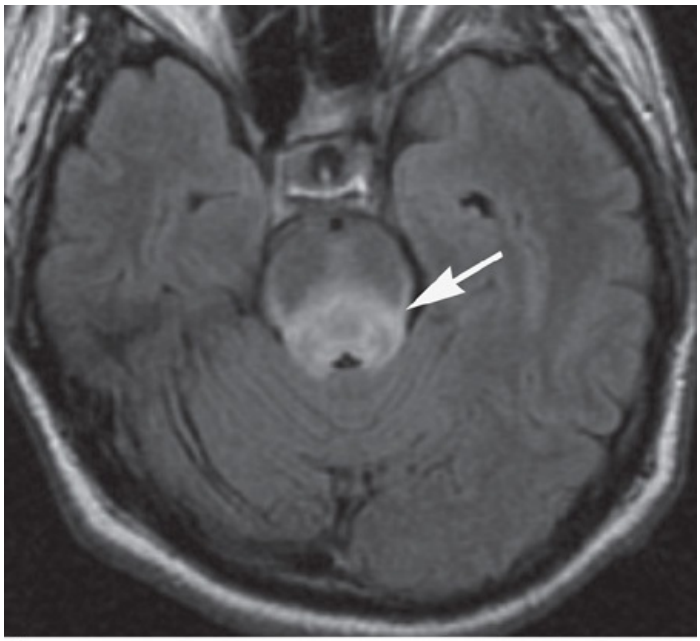

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

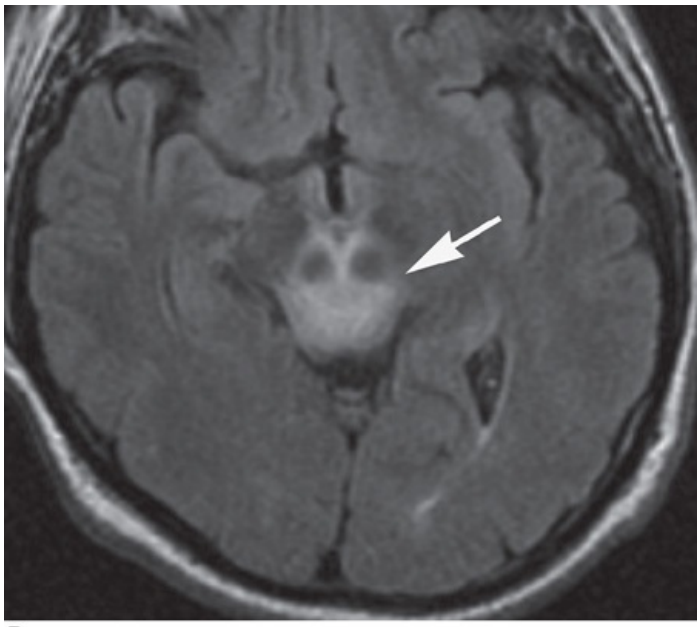

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

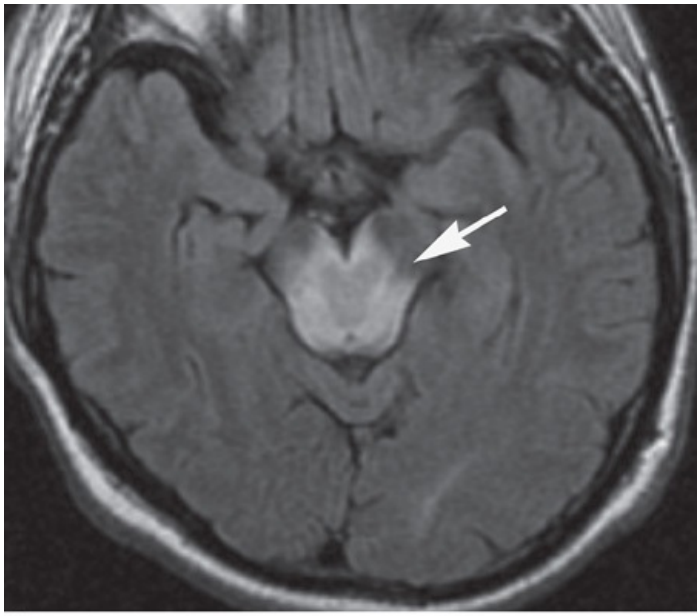

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

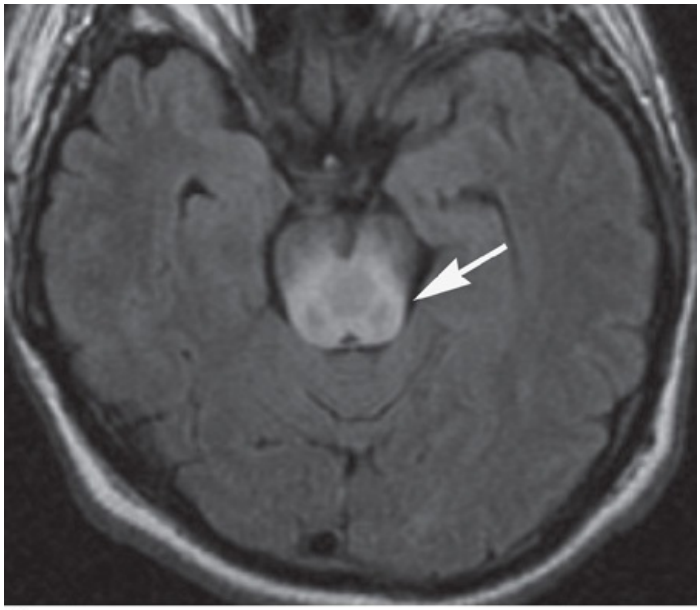

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

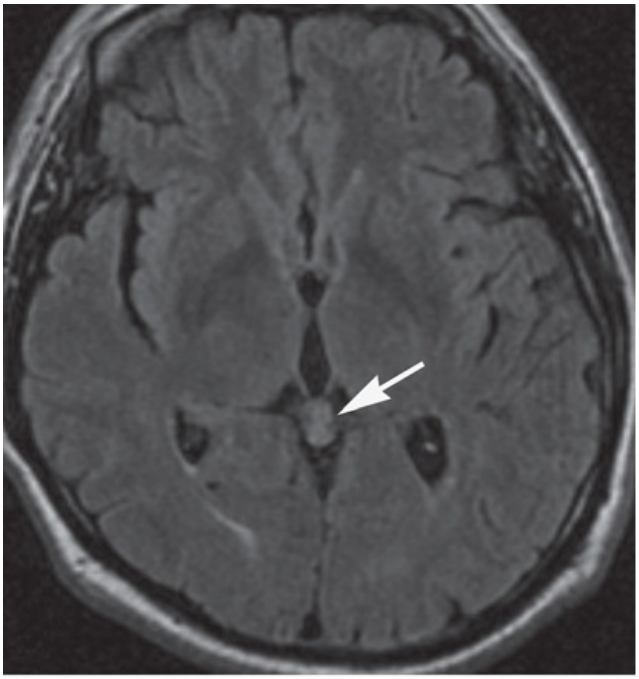

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

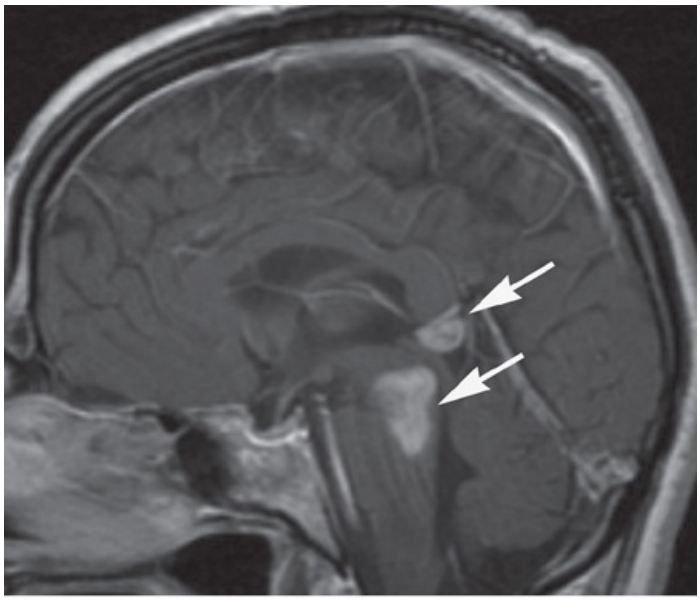

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurosarcoid (Chap. 329)

Case III

Axial FLAIR images (***A–E***) demonstrate abnormal high signal and slight expansion in the midbrain, dorsal pons, and pineal region (*arrows*) without significant mass effect.

Sagittal T1-weighted images postgadolinium (***F***) with fat suppression demonstrate abnormal enhancement in the midbrain, dorsal pons, and pineal region (*arrows*).

Figure e44-14

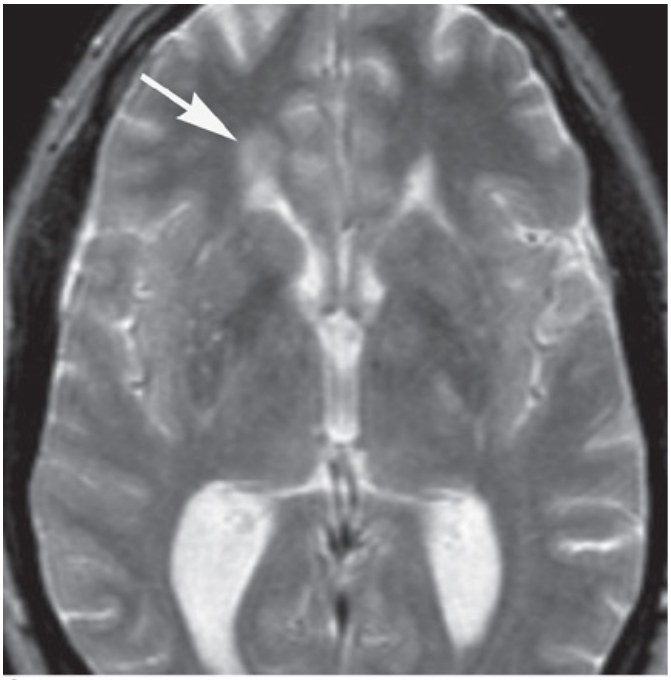

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

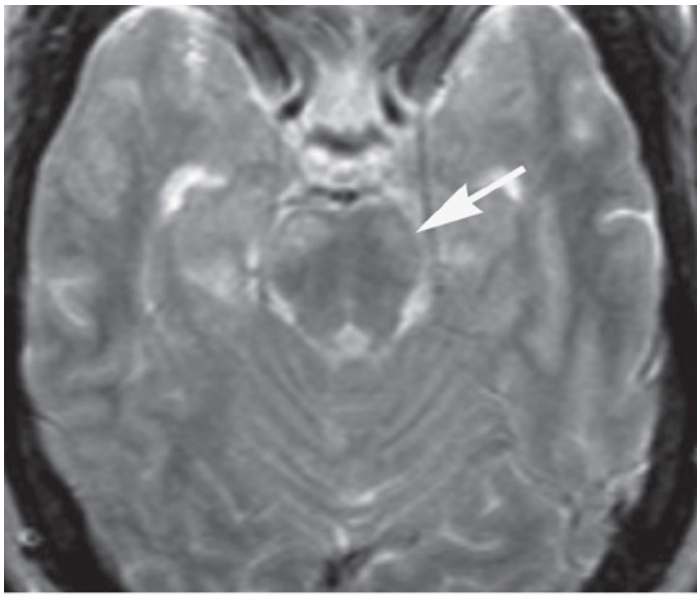

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

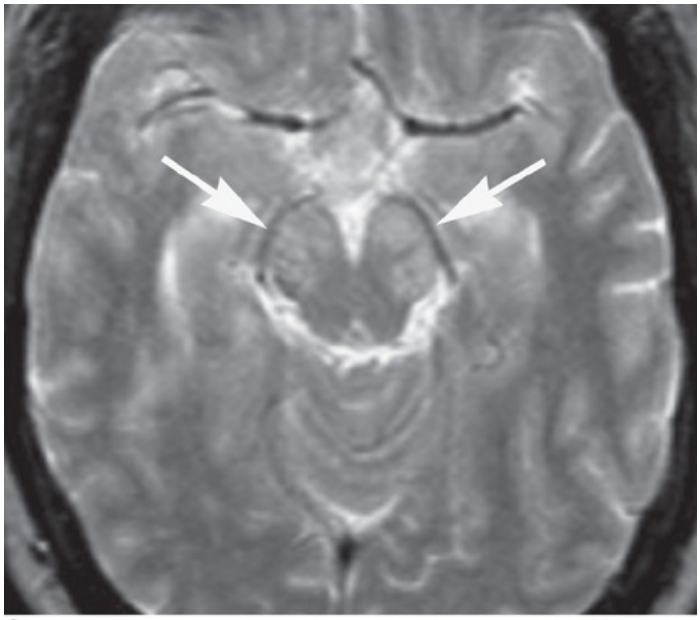

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

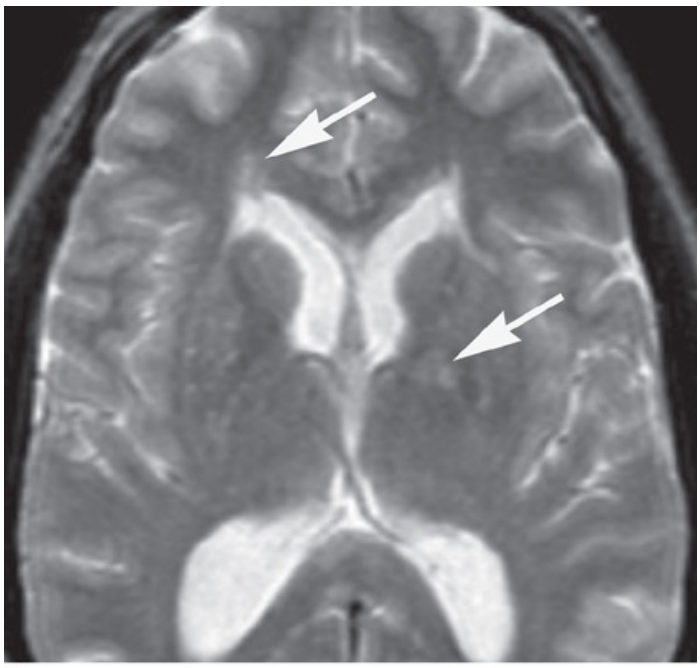

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

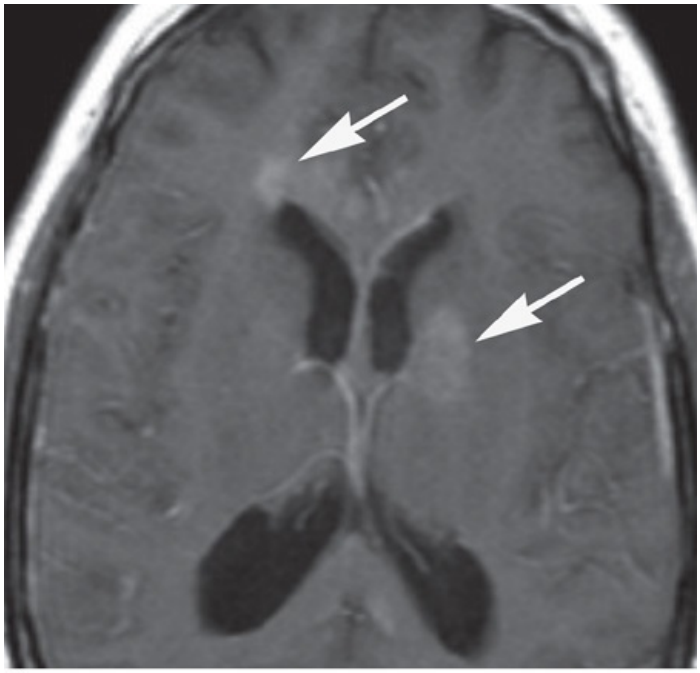

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

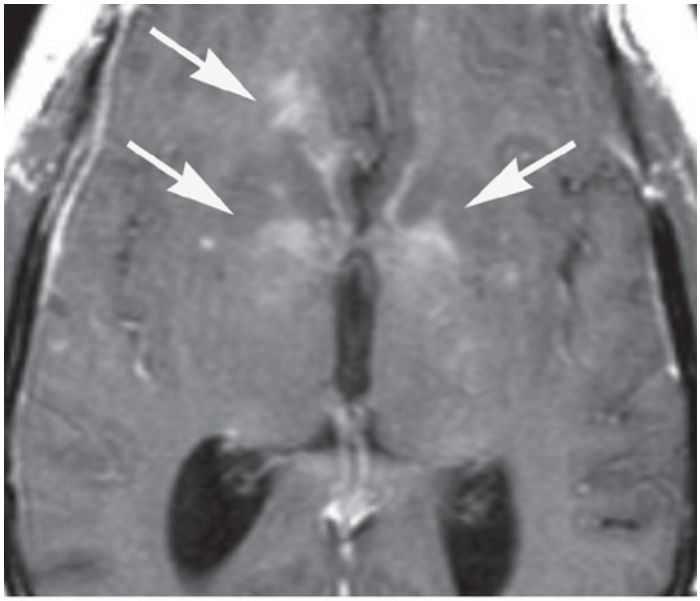

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

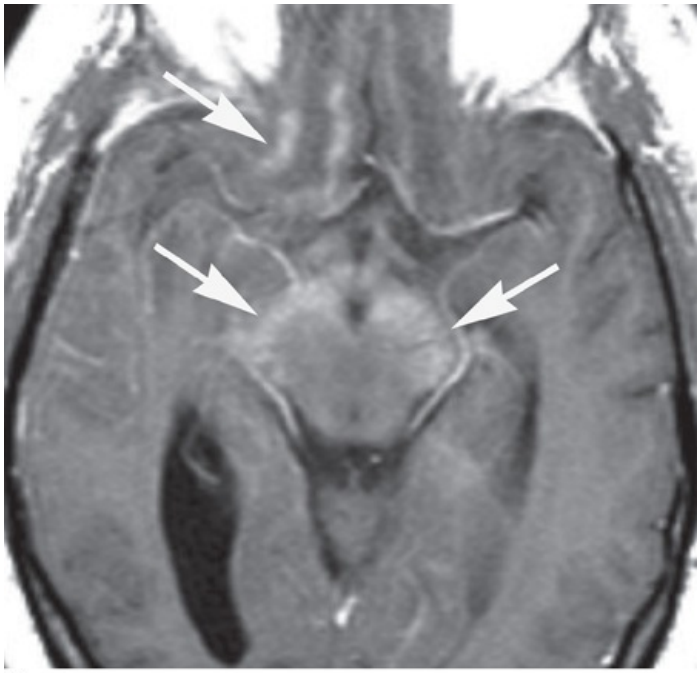

G

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

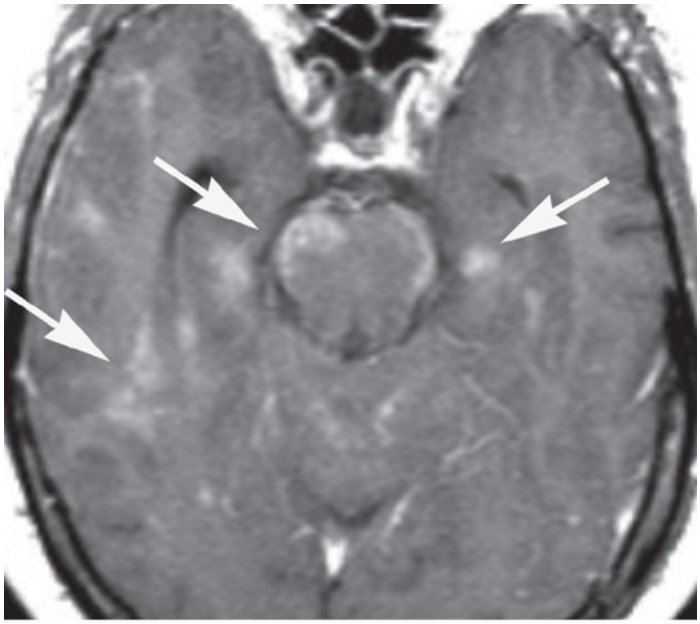

H

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neurosarcoid (Chap. 329)

Case IV

Axial T2-weighted images (***A–D***) demonstrate numerous areas of abnormal hyperintensity involving the corpus callosum, left internal capsule and globus pallidus, bilateral cerebral peduncles, bilateral gyrus rectus, right frontal lobe periventricular white matter, and patchy areas in bilateral temporal lobes.

T1-weighted images postgadolinium (***E–H***) demonstrate abnormal enhancement of those areas with high T2 signal.

Figure e44-15

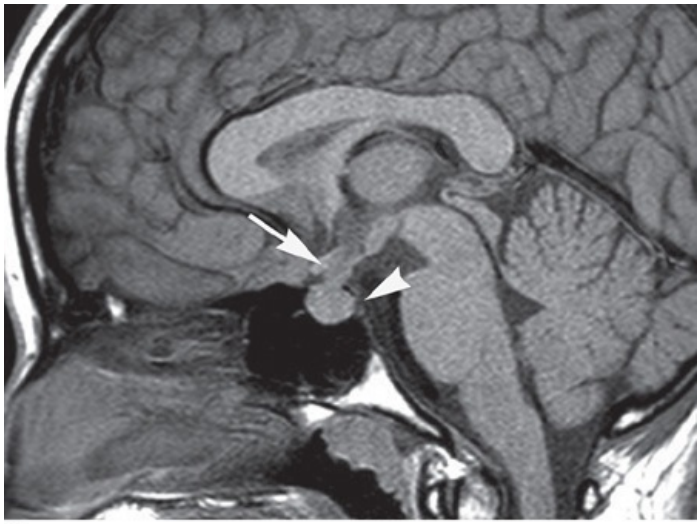

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

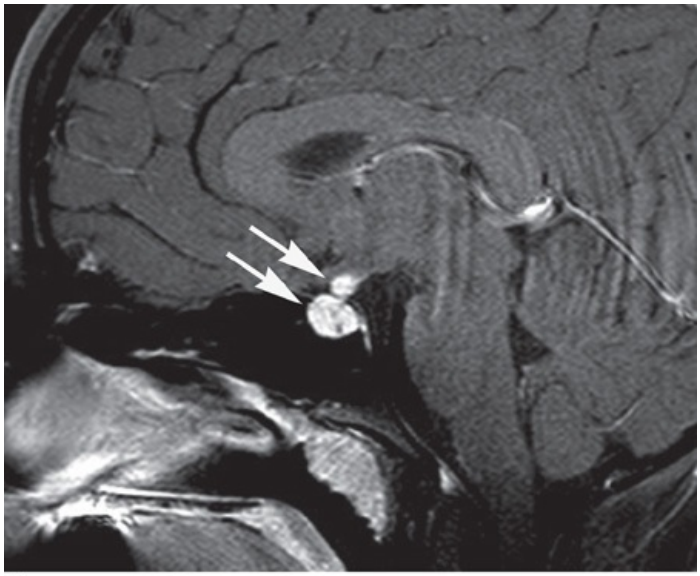

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

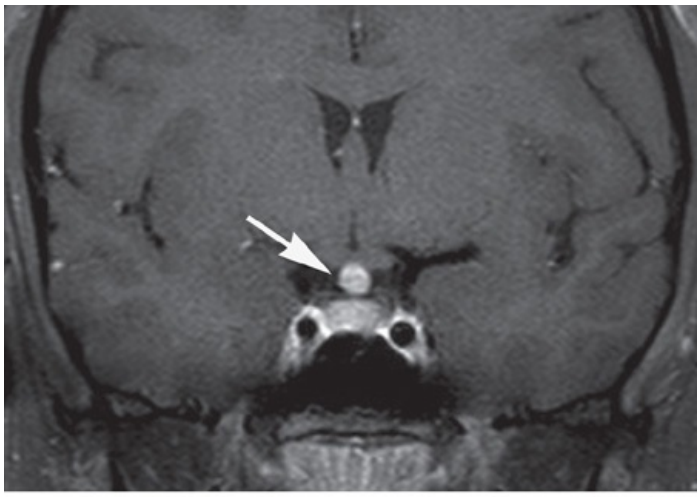

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Histiocytosis (Chap. 340)

Sagittal T1-weighted image (***A***) demonstrates enlargement of the pituitary stalk (*arrow*) and absence of the posterior pituitary intrinsic T1 hyperintensity (*arrowhead*).

Sagittal and coronal T1-weighted images postgadolinium (***B***, ***C***) demonstrate enhancement of the pituitary stalk and infundibulum (*arrows*).

Figure e44-16

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Middle cerebral artery stenosis (Chap. 370)

Time-of-flight (TOF) MR angiography (MRA) (***A***, ***B***) reveals narrowing within the left M1 segment that is likely secondary to atherosclerosis (*arrows*).

Figure e44-17

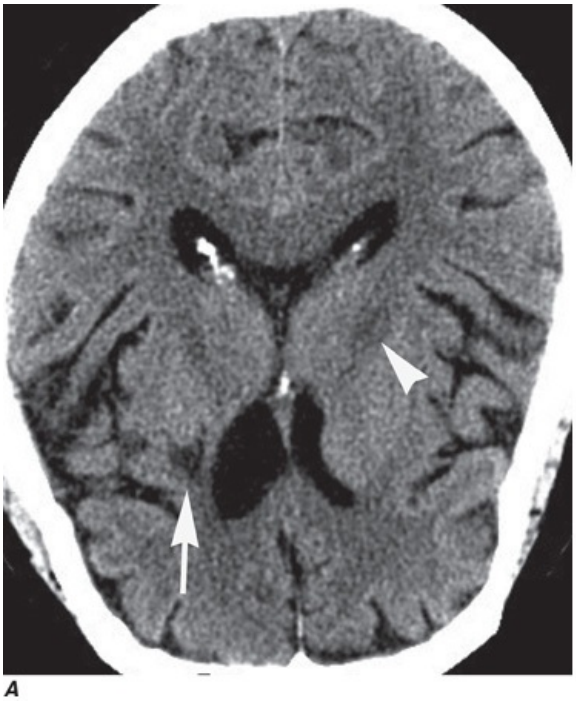

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

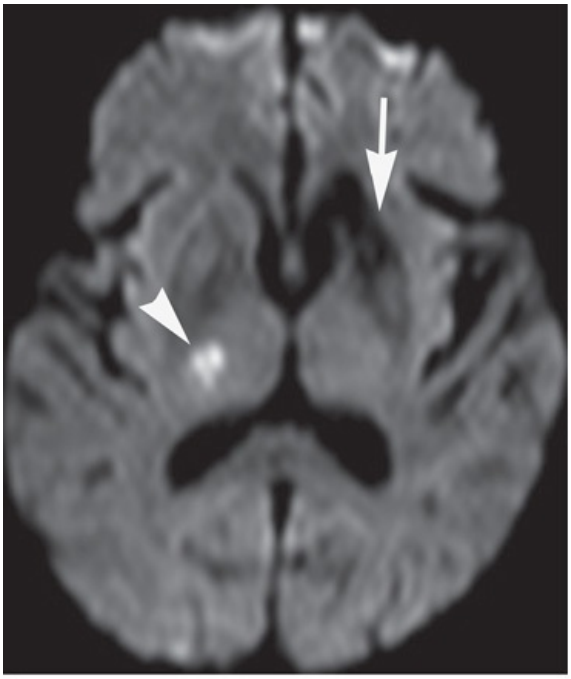

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

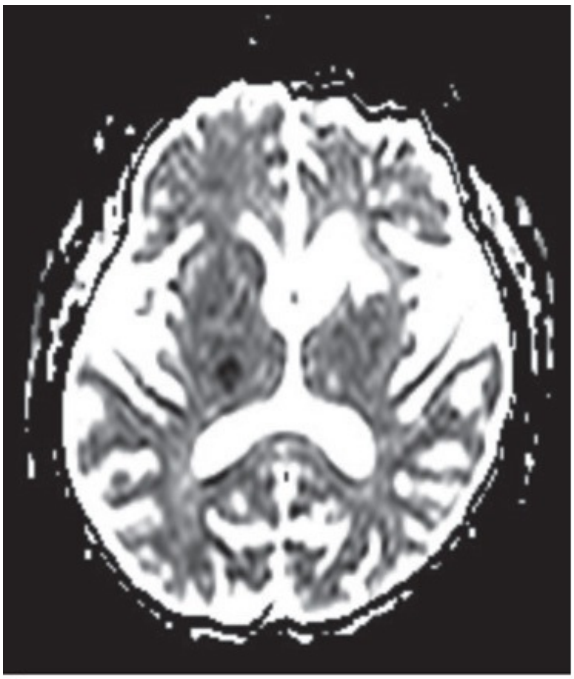

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lacunar infarction (Chap. 370)

Axial noncontrast CT (***A***) demonstrates abnormal hypodensity involving the left anterior putamen and anterior limb of internal capsule with ex-vacuo dilatation of the adjacent frontal horn of the left lateral ventricle, suggestive of an old infarction (*arrow*). A small area of slight hypodensity is also seen in the posterior limb of the right internal capsule that can correspond to an acute infarct (*arrowhead*).

Axial FLAIR MRI (***B***) demonstrates abnormal high signal involving the left anterior putamen and anterior limb of internal capsule with ex-vacuo dilatation of the adjacent frontal horn of the left lateral ventricle, suggestive of an old infarction (*arrow*). A small area of slight hyperintensity is also seen in the posterior limb of the right internal capsule that can correspond to an acute lacunar infarct (*arrowhead*).

Diffusion-weighted image (***C***) and apparent diffusion coefficient (ADC) map (***D***) demonstrate restricted water motion in the lesion of the posterior limb of the right internal capsule, strongly suggestive for an acute lacunar infarct (*arrowhead*). There is no evidence of restricted diffusion in the old infarct (*arrow*).

Figure e44-18

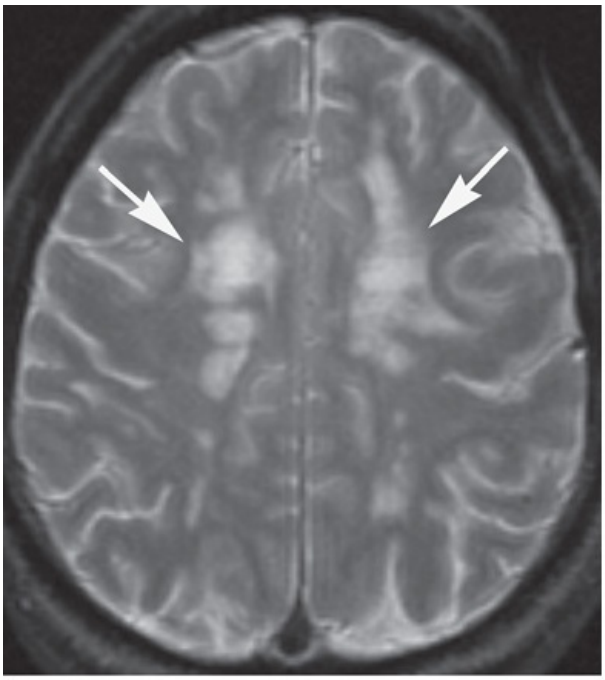

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

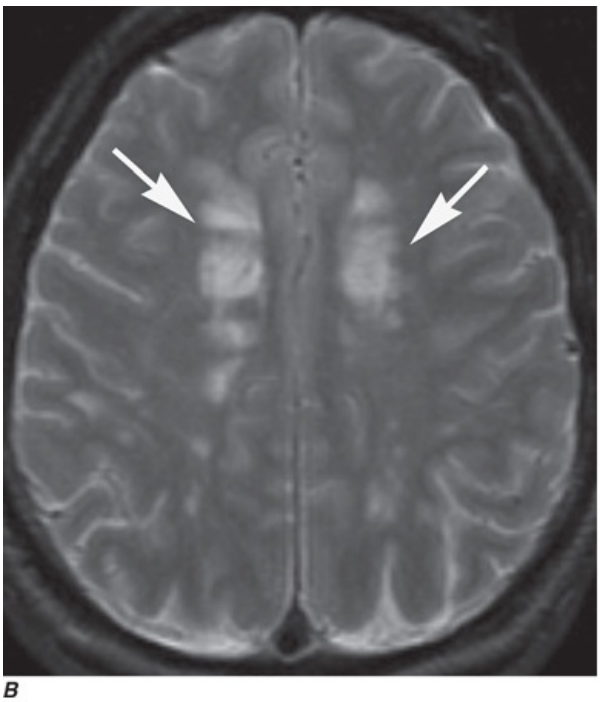

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

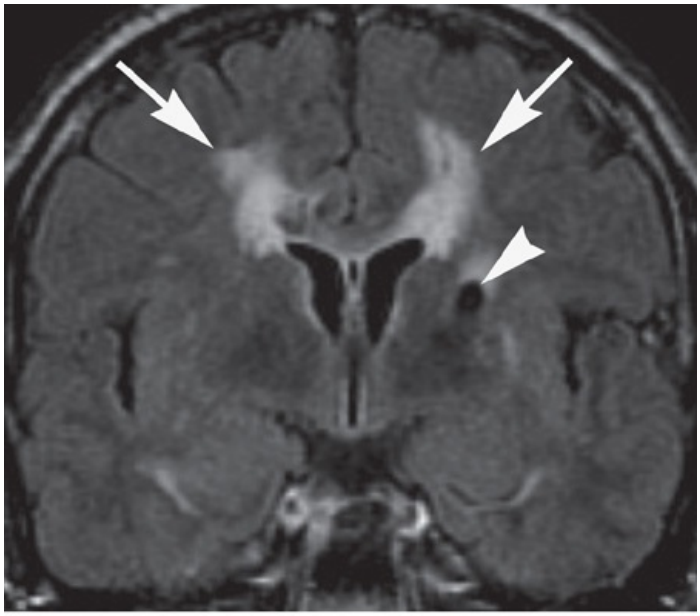

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

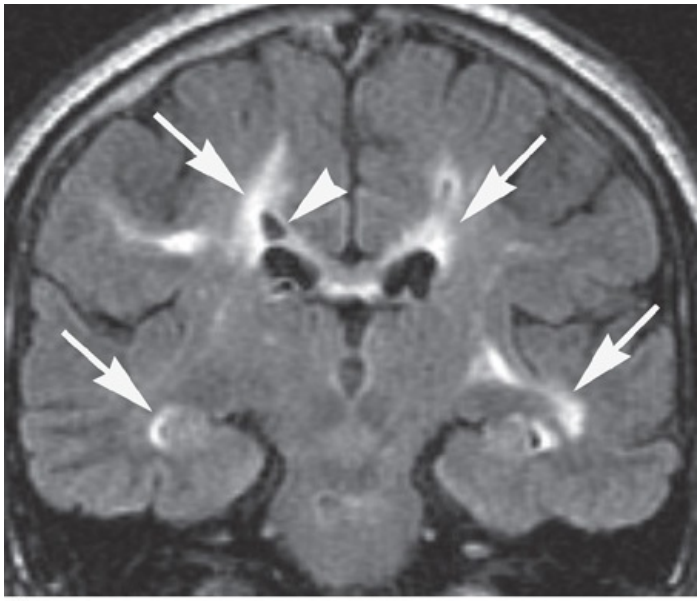

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) (Chap. 370)

Axial T2-weighted MR images (***A***, ***B***) demonstrate multiple patchy areas of abnormal high signal in the periventricular white matter (*arrows*).

Coronal FLAIR MRI (***C***, ***D***) demonstrates multiple patchy areas of abnormal high signal in the periventricular white matter bilaterally, including the temporal lobes (*arrows*). In some of these areas, there are small areas of tissue loss (encephalomalacia) (*arrowheads*).

Figure e44-19

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

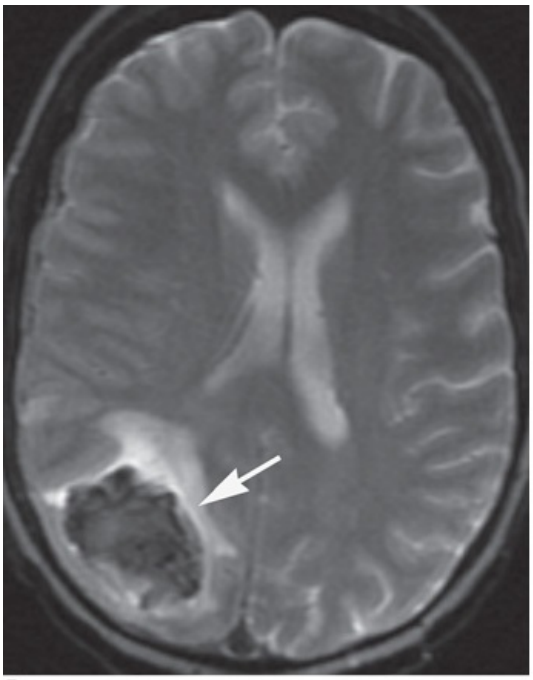

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

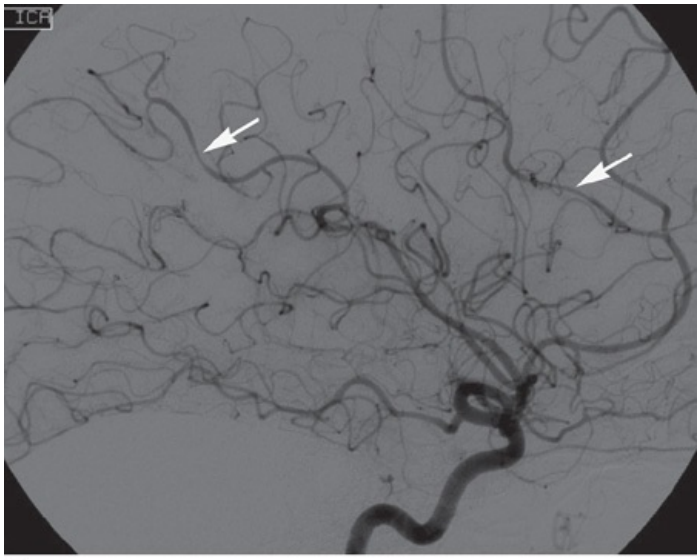

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CNS vasculitis (Chap. 370)

Axial noncontrast CT (***A***) demonstrates a large hyperdense intraparenchymal hematoma surrounded by hypodense vasogenic edema in the right parietal lobe.

Axial T2-weighted MRI (***B***) demonstrates a large hypointense intraparenchymal hematoma surrounded by hyperintense vasogenic edema in the right parietal lobe.

Conventional angiography (***C***) demonstrates multiple segments of intracranial arterial narrowing, some of which have associated adjacent areas of focal arterial dilatation. These abnormalities are suggestive of vasculitis.

Figure e44-20

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

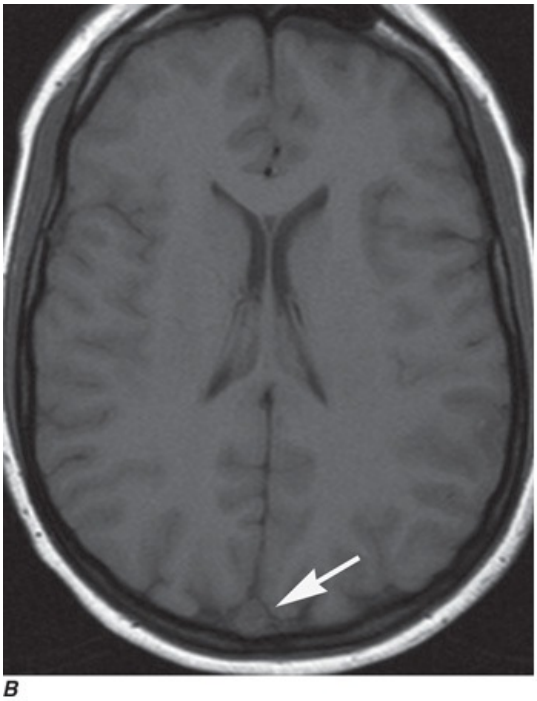

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

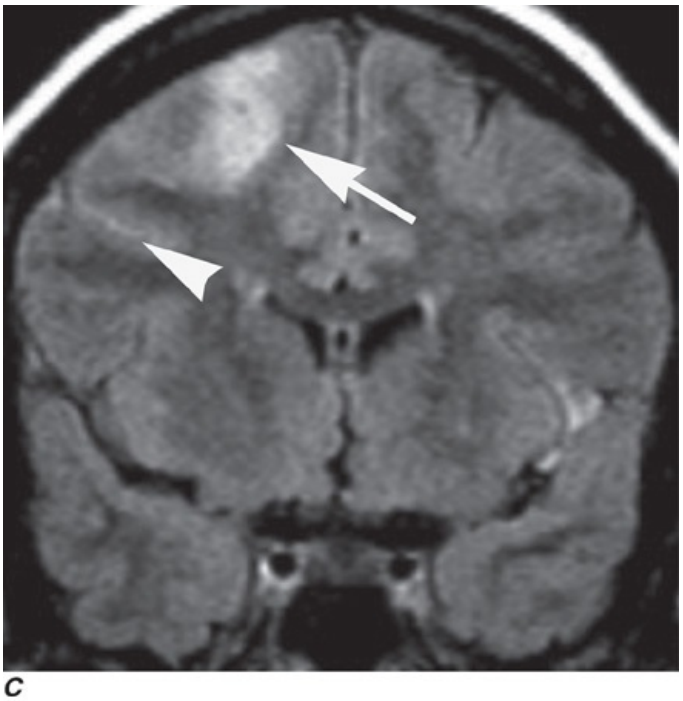

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

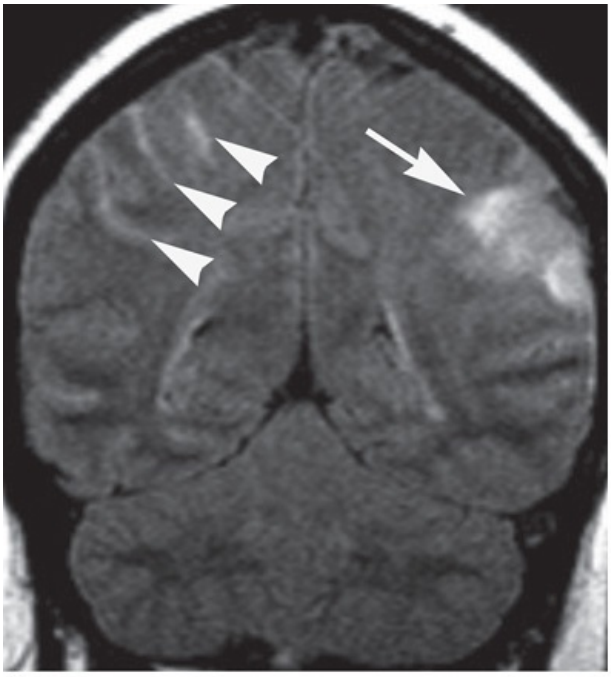

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

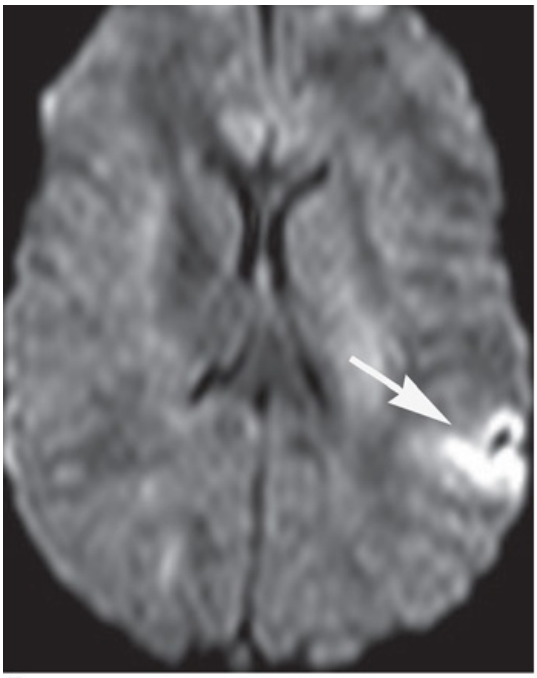

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

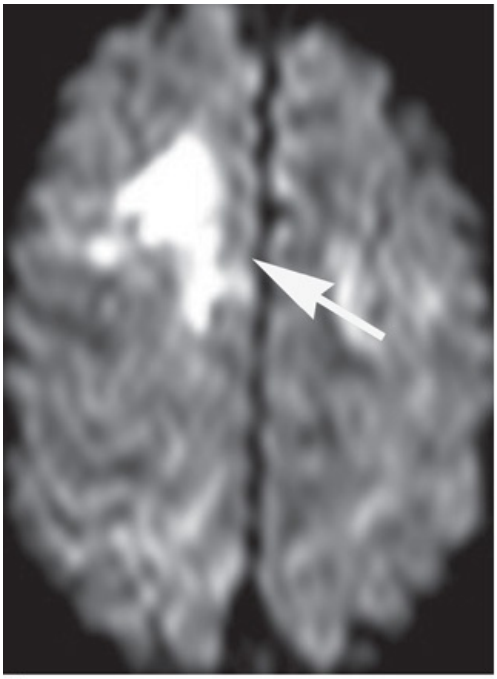

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

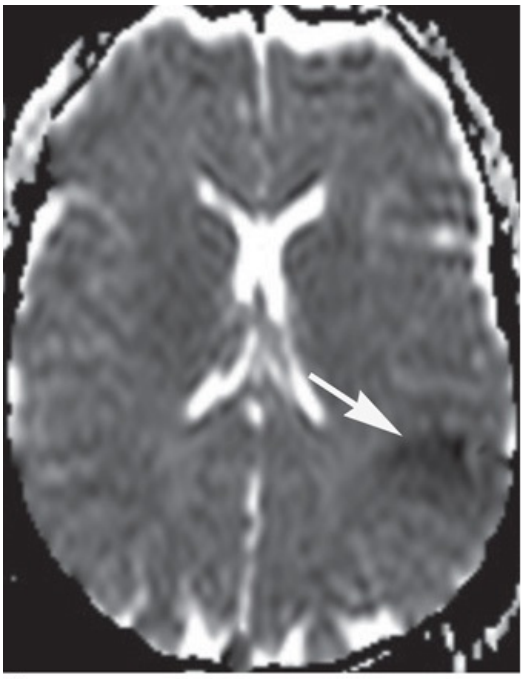

G

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

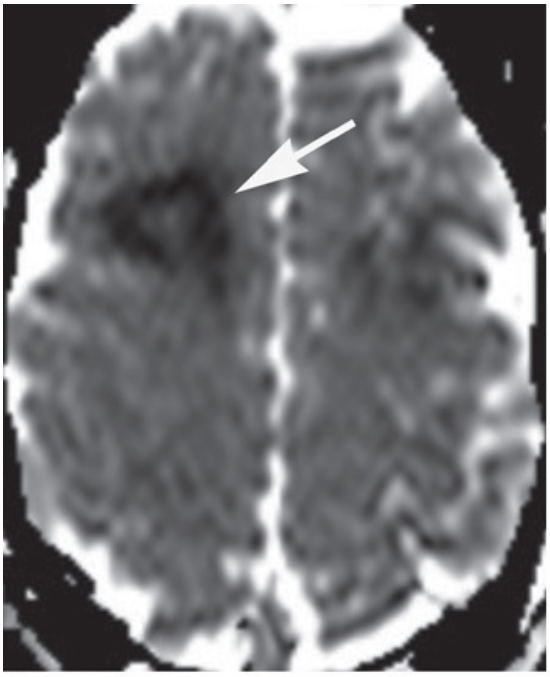

H

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

I

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

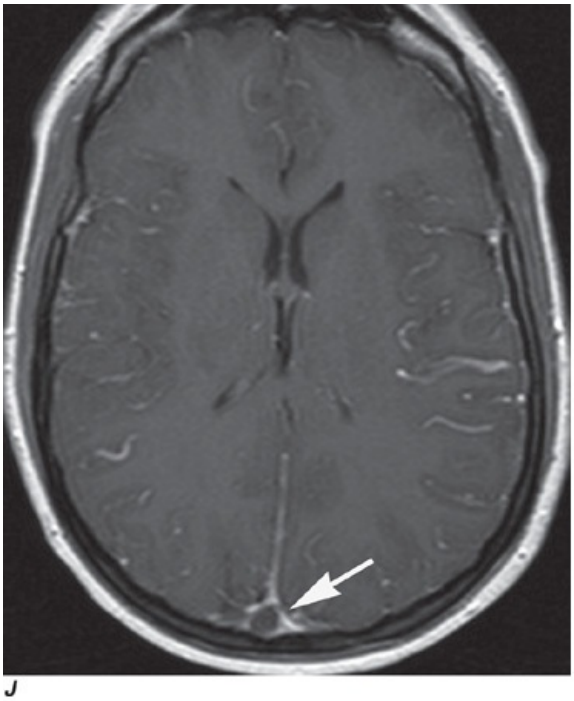

J

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

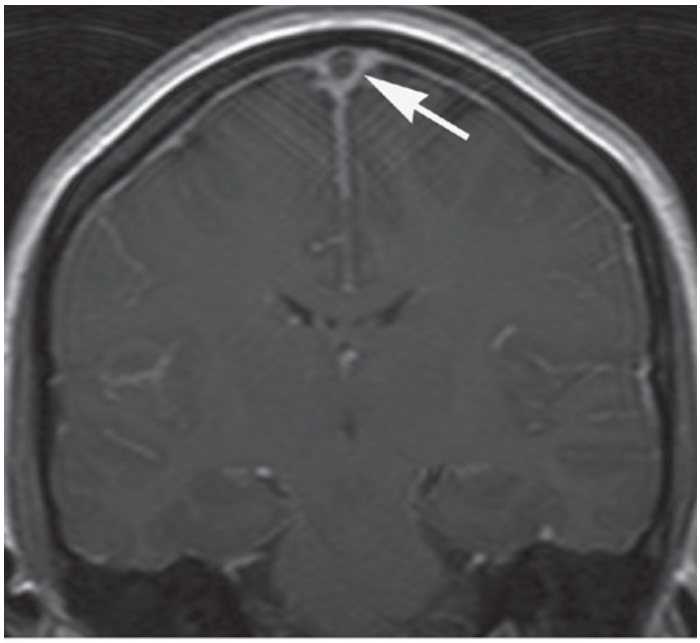

K

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Superior sagittal sinus thrombosis (Chap. 370)

Noncontrast CT of the head (***A***) demonstrates increased density in the superior sagittal sinus, suggestive of thrombosis (*arrow*), and small linear hyperdensities in some temporal lobe sulci, suggestive of subarachnoid hemorrhage (*arrowheads*).

Axial T1-weighted MRI (***B***) demonstrates absence of flow void in the superior sagittal sinus, suggestive of thrombosis.

Coronal FLAIR images (***C***, ***D***) demonstrate areas of abnormal high signal involving the gray and the subcortical white matter of the right frontal and left parietal lobes, as well as the adjacent sulci. These findings are suggestive of vasogenic edema with subarachnoid hemorrhage (*arrowheads*).

Diffusion-weighted images (***E***, ***F***) and ADC maps (***G***, ***H***) demonstrate restricted diffusion of the abnormal areas on FLAIR, suggestive of infarct.

Phase-contrast venography of the brain (***I***) demonstrates absence of signal in the superior sagittal sinus down to the torcular herophili, and left transverse sinus and jugular vein.

Axial (***J***) and coronal (***K***) T1-weighted images postgadolinium demonstrate a filling defect in the

superior sagittal sinus, suggestive of thrombosis.

Figure e44-21

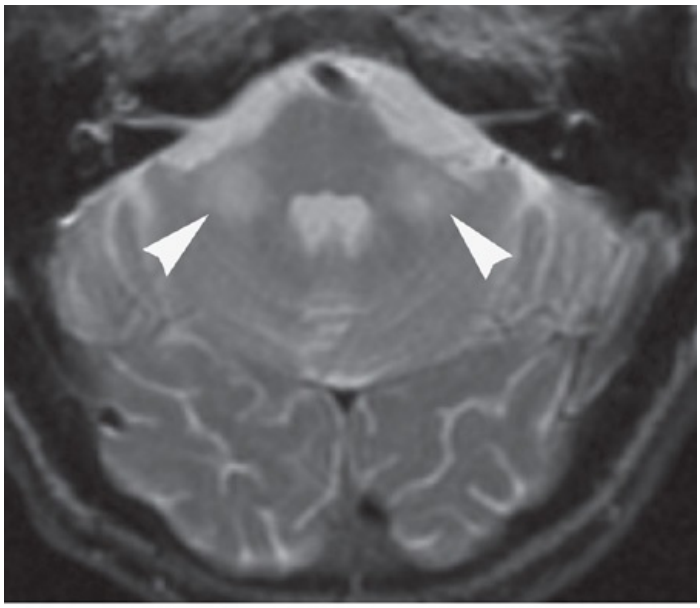

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

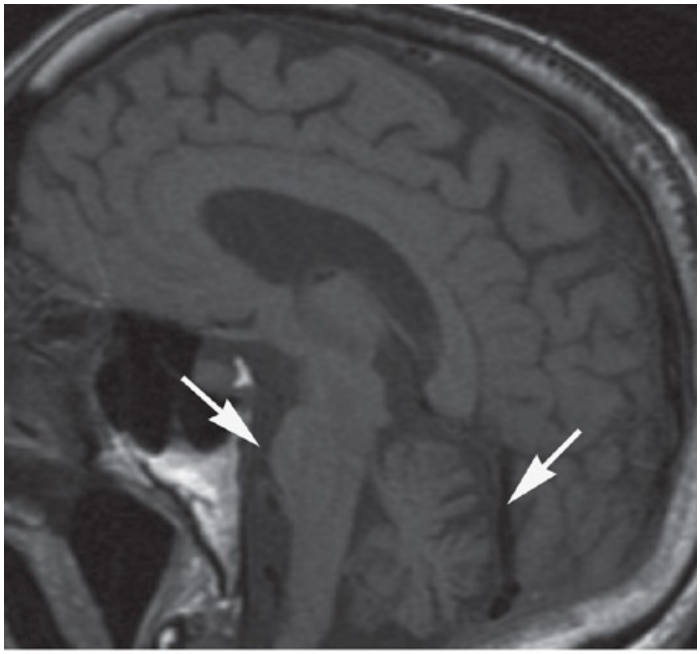

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple system atrophy (Chap. 372)

Axial T2-weighted MR image (***A***) reveals symmetric poorly circumscribed abnormal high signal in the middle cerebellar peduncles bilaterally (*arrowheads*).

Sagittal T1-weighted MR image (***B***) demonstrates pontine atrophy and enlarged cerebellar fissures as a result of cerebellar atrophy (*arrows*).

Figure e44-22

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

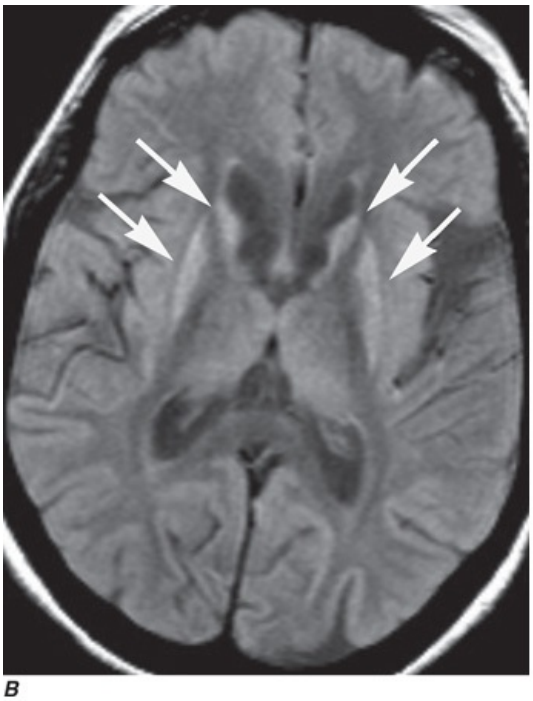

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

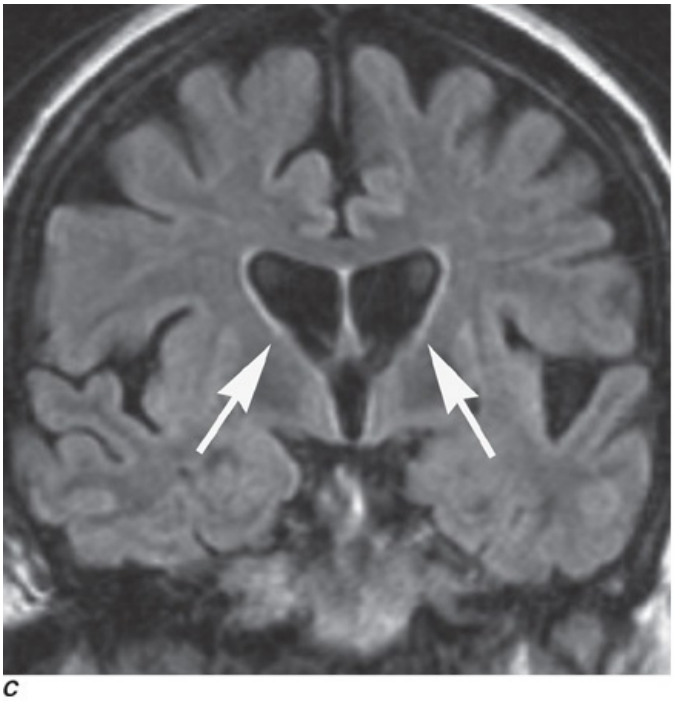

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

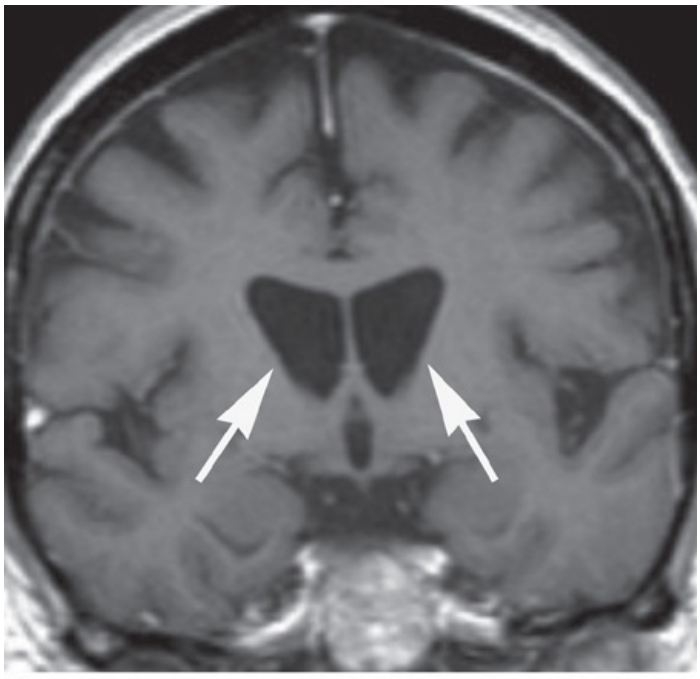

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Huntington's disease (Chap. 372)

Axial noncontrast CT (***A***) demonstrates symmetric bilateral severe atrophy involving the caudate nuclei, putamen, and globus pallidi bilaterally with consequent enlargement of the frontal horns of the lateral ventricles (*arrows*). There is also diffuse prominence of the sulci indicating generalized cortical atrophy.

Axial (***B***) and coronal (***C***) FLAIR images demonstrate bilateral symmetric abnormal high signal in the caudate and putamen.

Coronal T1-weighted image (***D***) demonstrates enlarged frontal horns with abnormal configuration. Also note diffusely decreased marrow signal, which could represent anemia or myeloproliferative disease.

Figure e44-23

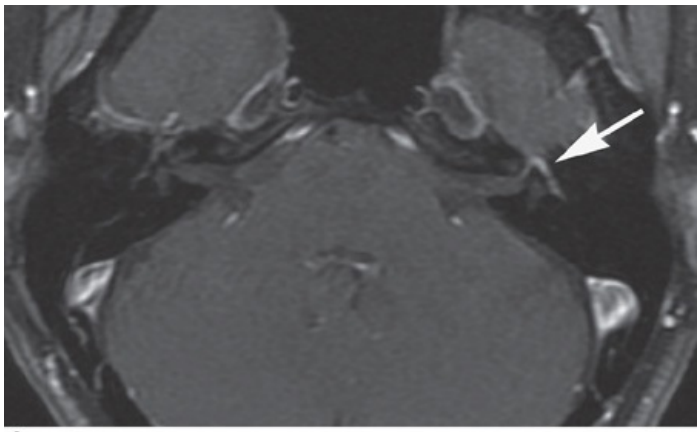

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

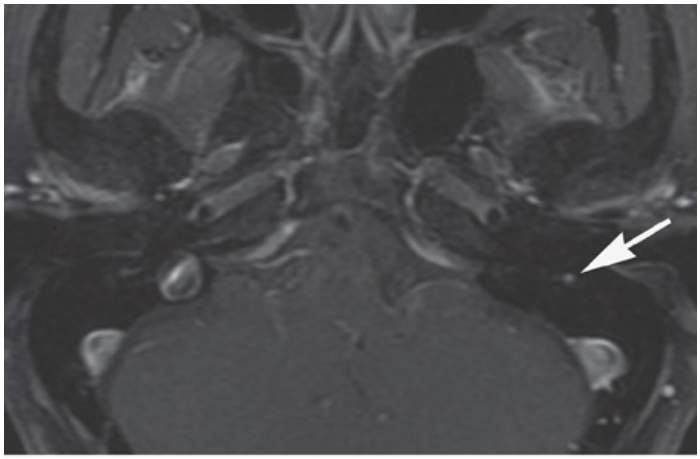

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

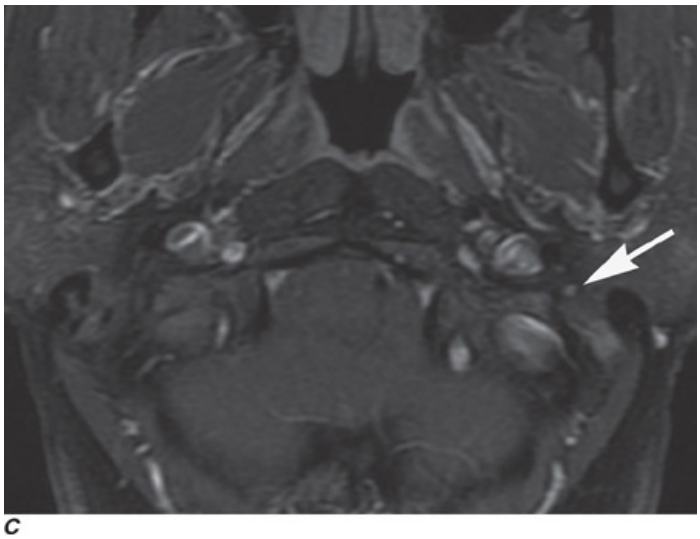

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

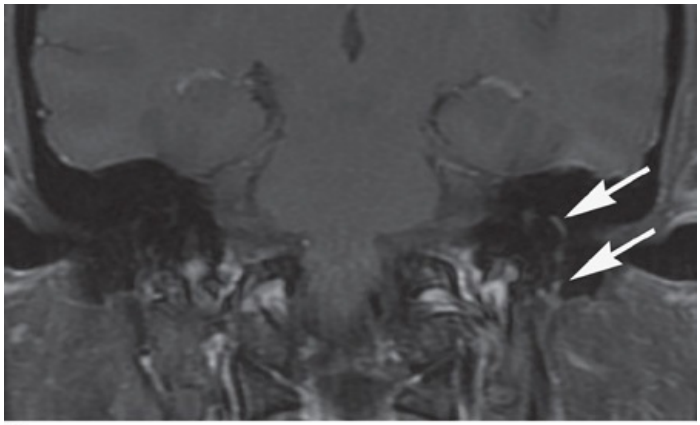

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

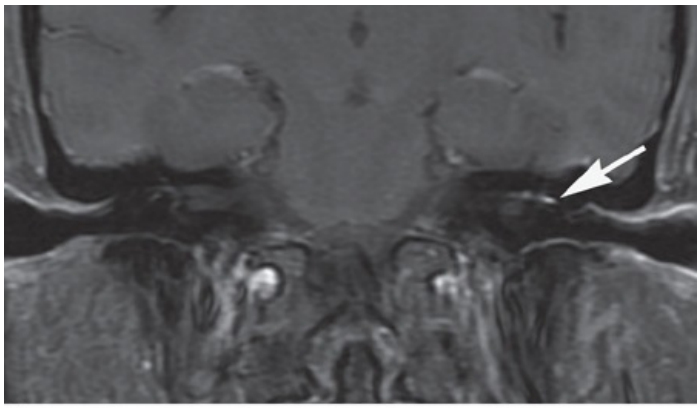

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Bell's palsy (Chap. 376)

Axial T1-weighted images postgadolinium with fat suppression (***A–C***) demonstrate diffuse smooth linear enhancement along the left facial nerve, involving the second and third segments (genu, tympanic, and mastoid) within the temporal bone (*arrows*). Note that there is no evidence of a mass lesion. A potential pitfall for facial nerve enhancement in the stylomastoid foramen is the enhancement of the stylomastoid artery that enters the foramen and supplies the tympanic cavity, the tympanic antrum, mastoid cells, and the semicircular canals.

Coronal T1-weighted images postgadolinium with fat suppression (***D***, ***E***) demonstrate the course of the enhancing facial nerve (*arrows*). Although these findings are highly suggestive of Bell's palsy, the diagnosis is established on clinical grounds.

Figure e44-24

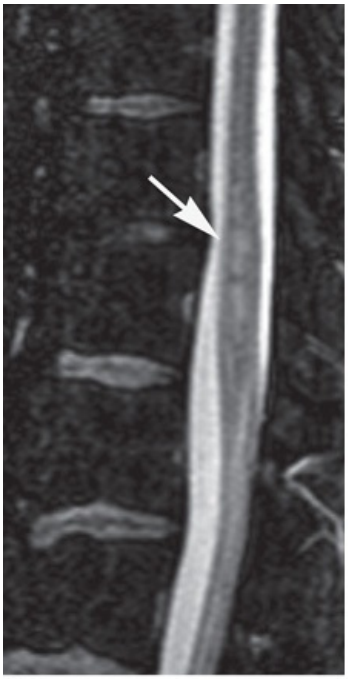

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

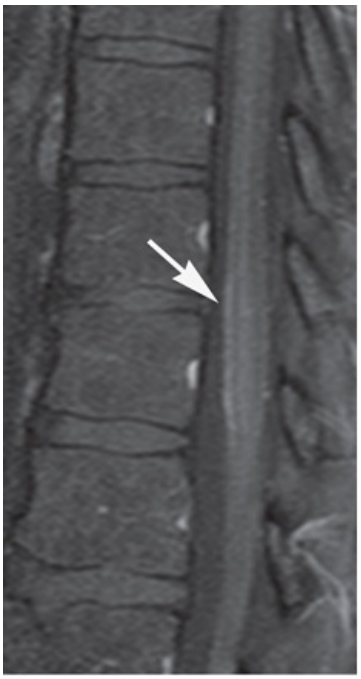

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Spinal cord infarction (Chap. 377)

Sagittal T2-weighted MR image of the lumbar spine (***A***) demonstrates poorly defined areas of abnormal high signal in the conus medullaris and mild cord expansion (*arrow*).

T1-weighted MR image of the lumbar spine postgadolinium (***B***) demonstrates mild enhancement (*arrow*).

Sagittal diffusion-weighted MR image of the lumbar spine (***C***) demonstrates restricted diffusion (*arrow*) in the areas of abnormal high signal on the T2-weighted image (***A***).

Figure e44-25

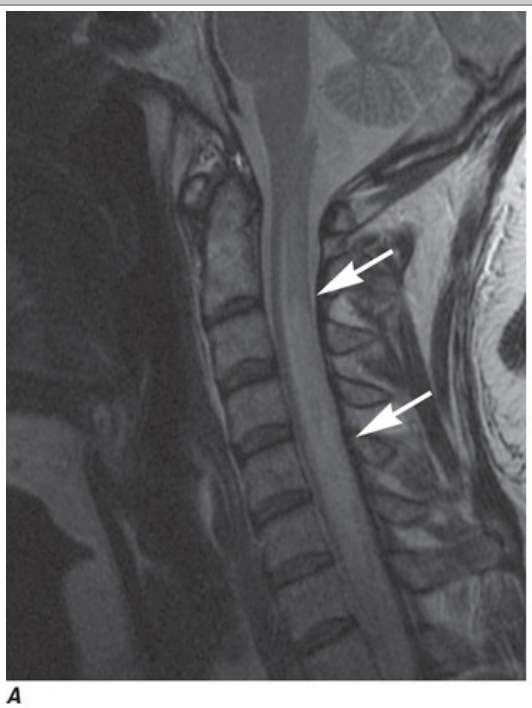

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute transverse myelitis (Chap. 377)

Sagittal T2-weighted MR image (***A***) demonstrates abnormal high signal in the cervical cord extending from C1 to T1 with associated cord expansion (*arrows*).

Sagittal T1-weighted MR image postgadolinium (***B***) demonstrates abnormal enhancement in the posterior half of the cord from C2 to T1 (*arrows*).

Figure e44-26

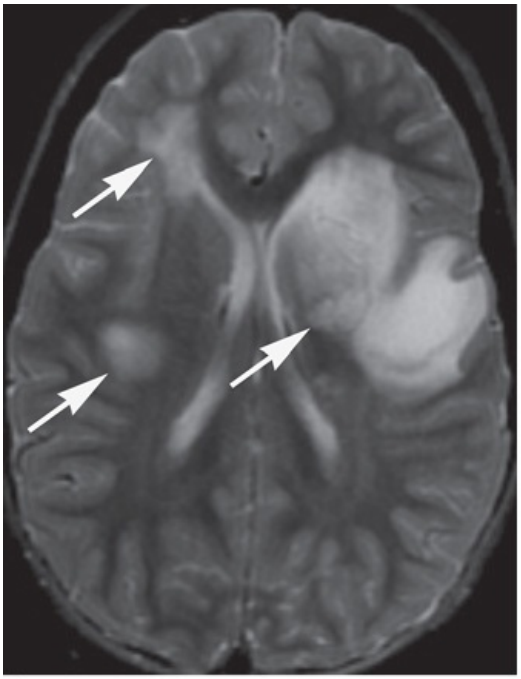

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

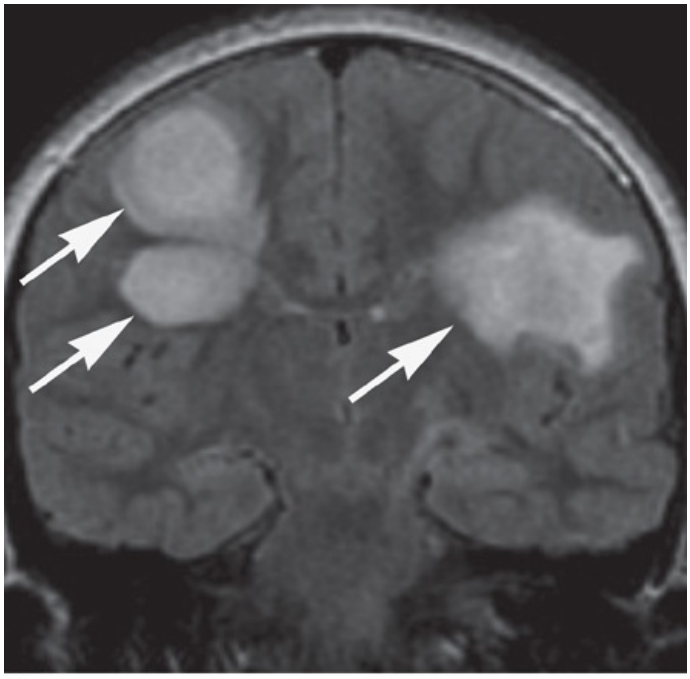

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

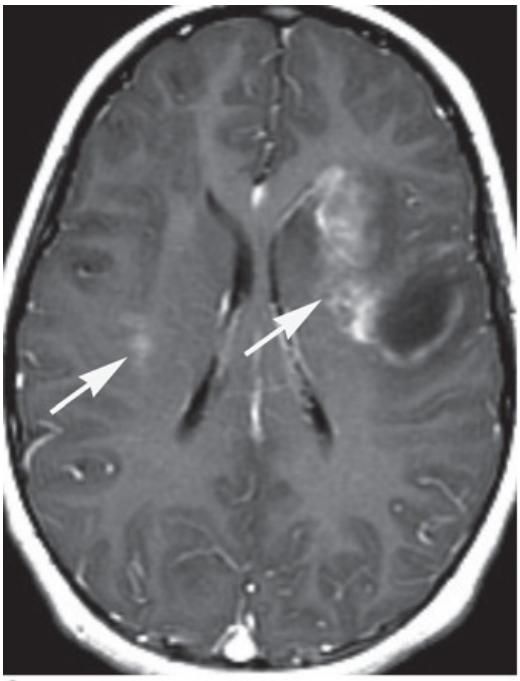

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

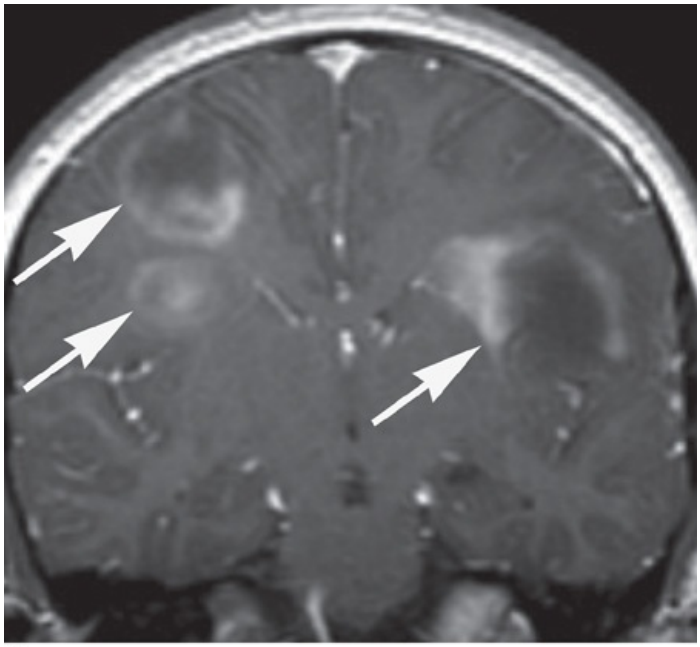

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute disseminated encephalomyelitis (ADEM) (Chap. 380)

Axial T2-weighted (***A***) and coronal FLAIR (***B***) images demonstrate abnormal areas of high signal involving predominantly the subcortical white matter of the frontal lobe bilaterally, and left caudate head.

Following administration of gadolinium, corresponding axial (***C***) and coronal (***D***) T1-weighted images demonstrate irregular enhancement consistent with blood-brain barrier breakdown and inflammation; some lesions show incomplete rim enhancement, typical for demyelination.

Figure e44-27

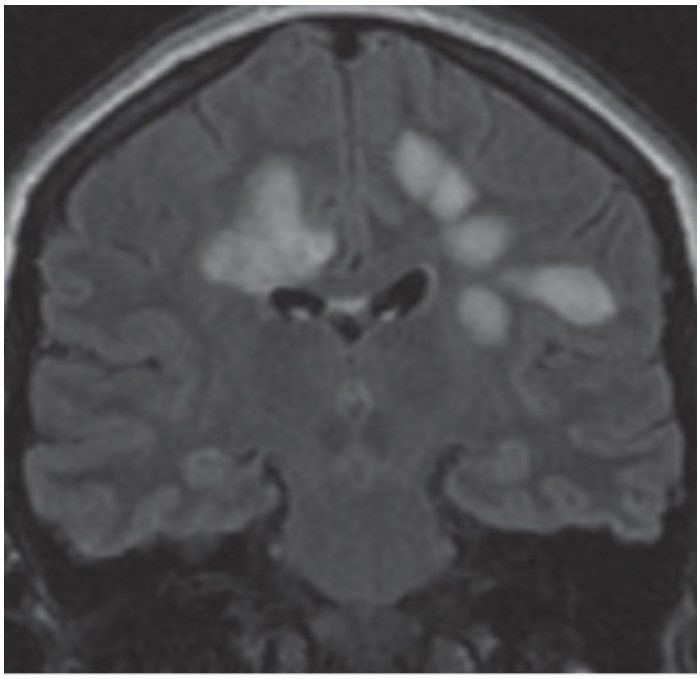

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

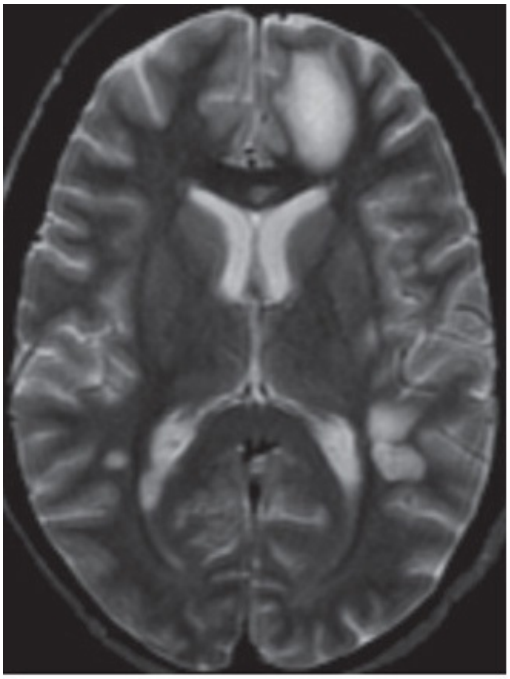

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

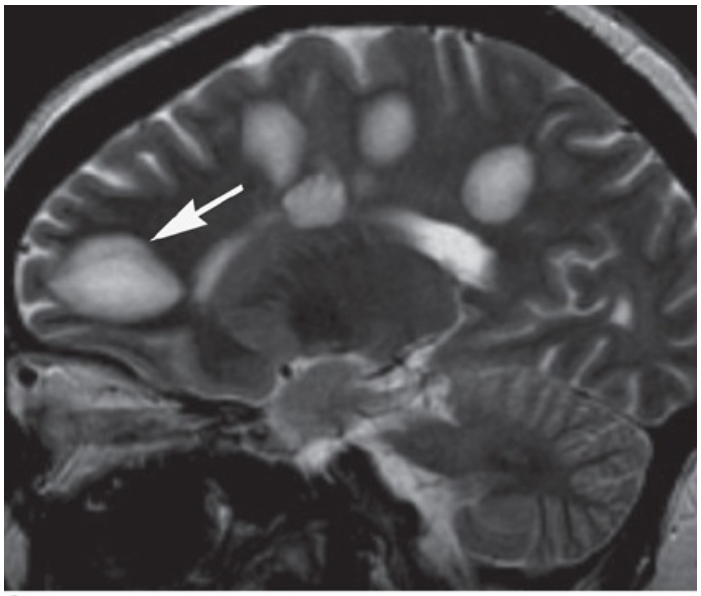

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

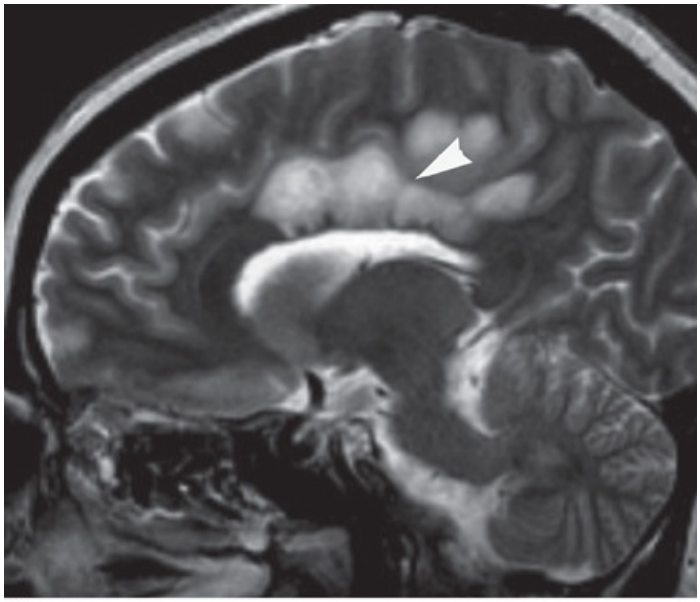

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

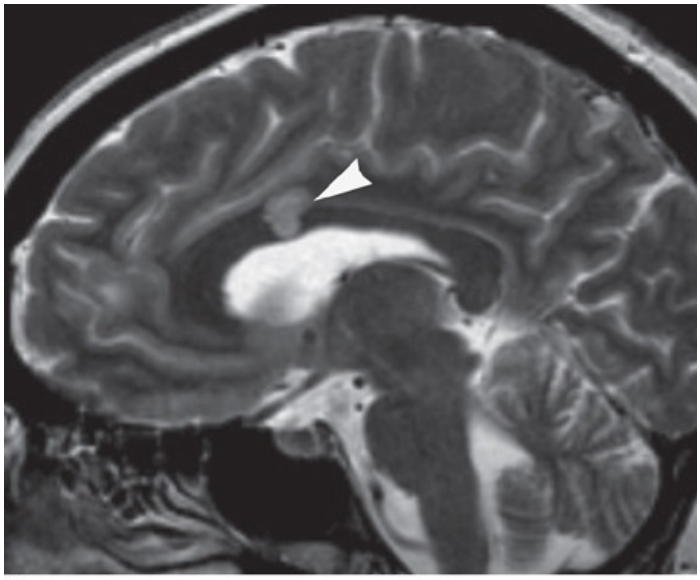

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

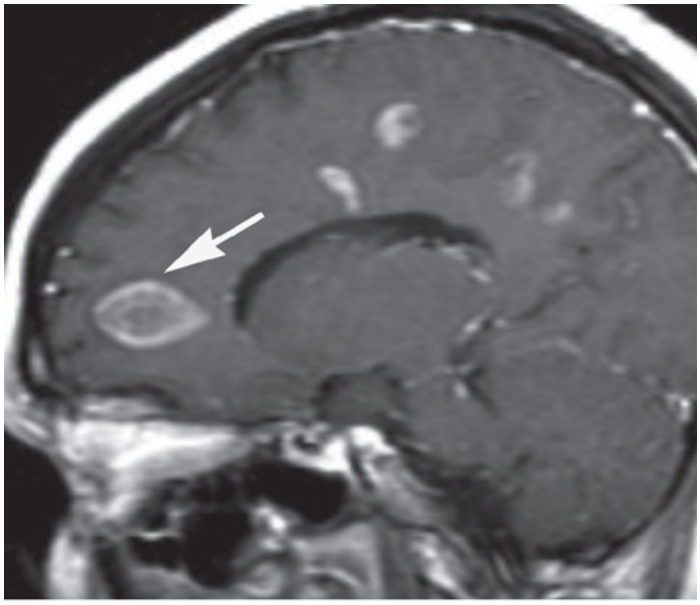

F

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

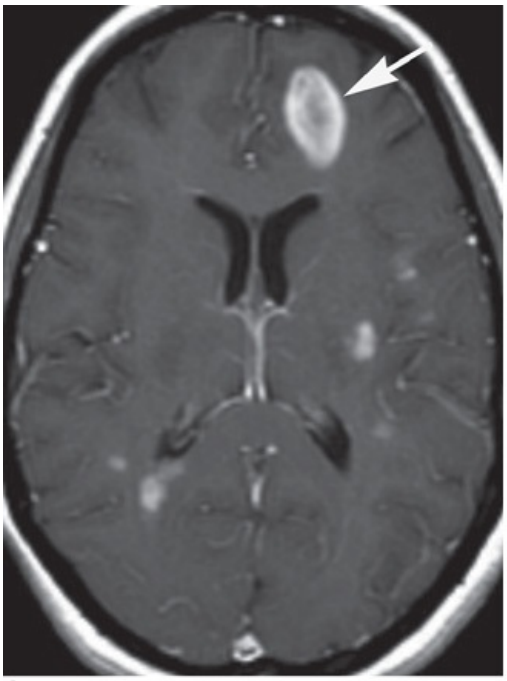

G

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

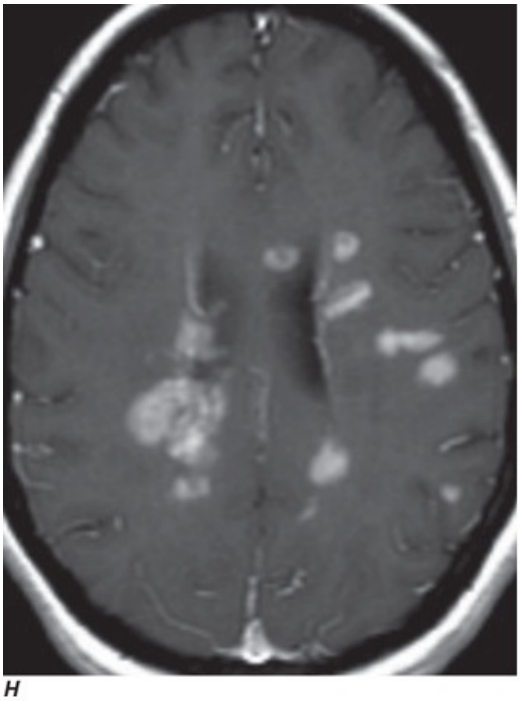

H

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Baló's concentric sclerosis (a variant of multiple sclerosis) (Chap. 380)

Coronal FLAIR MRI (***A***) demonstrates multiple areas of abnormal high signal in the supratentorial white matter bilaterally. The lesions are ovoid in shape, perpendicular to the orientation of the lateral ventricles, and with little mass effect.

Axial (***B***) and sagittal (***C–E***) T2-weighted MR images demonstrate multiple areas of abnormal high signal in the supratentorial white matter bilaterally, as well as the involvement of the body and splenium of the corpus callosum and the callosal-septal interface (*arrowhead*). Some of the lesions reveal concentric layers, typical of Baló's concentric sclerosis (*arrows*).

Sagittal (***F***) and axial (***G***, ***H***) T1-weighted MR images postgadolinium demonstrate abnormal enhancement of all lesions with some of the lesions demonstrating concentric ring enhancement (*arrows*).

Figure e44-28

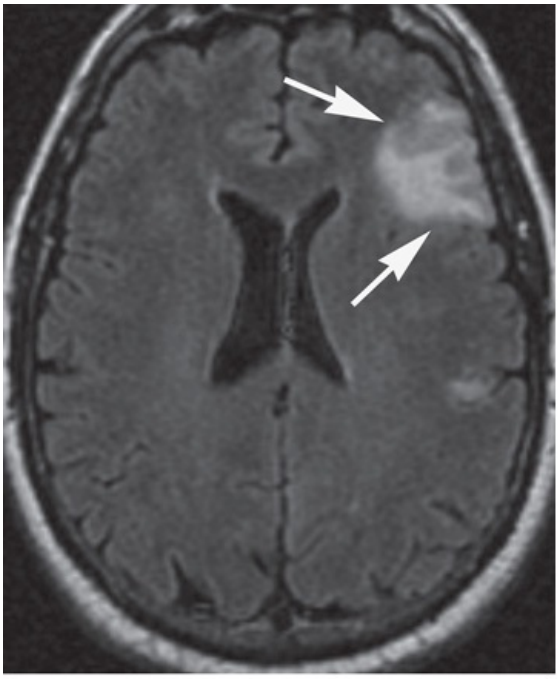

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

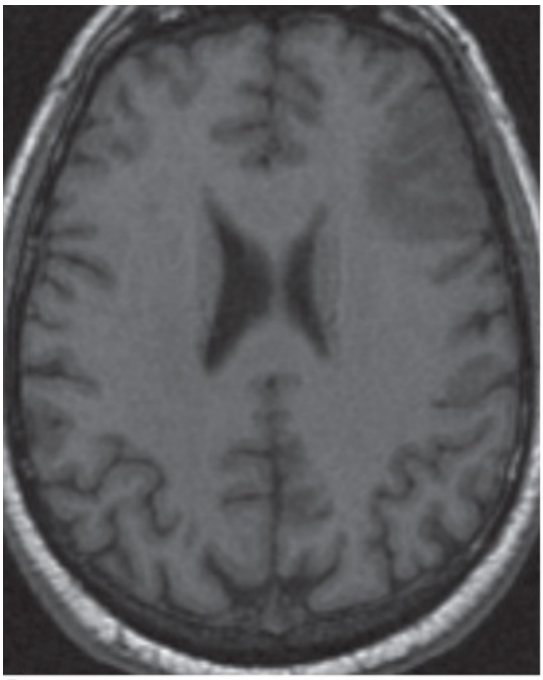

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

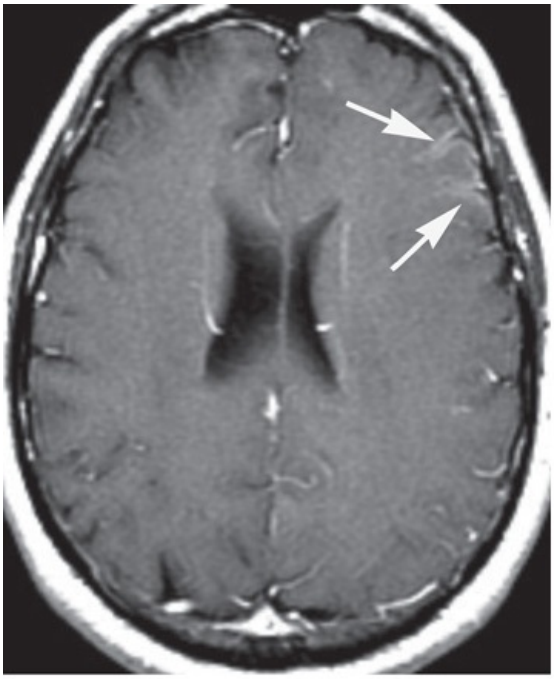

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hashimoto's encephalopathy (Chap. 381)

Axial FLAIR (***A***) demonstrates focal area of abnormal high signal involving the gray and white matter in the left frontal lobe. There is also a small area of abnormal high signal in the precentral gyrus.

Axial T1-weighted images (***B***, ***C***) pre- and postgadolinium demonstrate cortical/pial enhancement in the region of high signal on FLAIR.

Figure e44-29

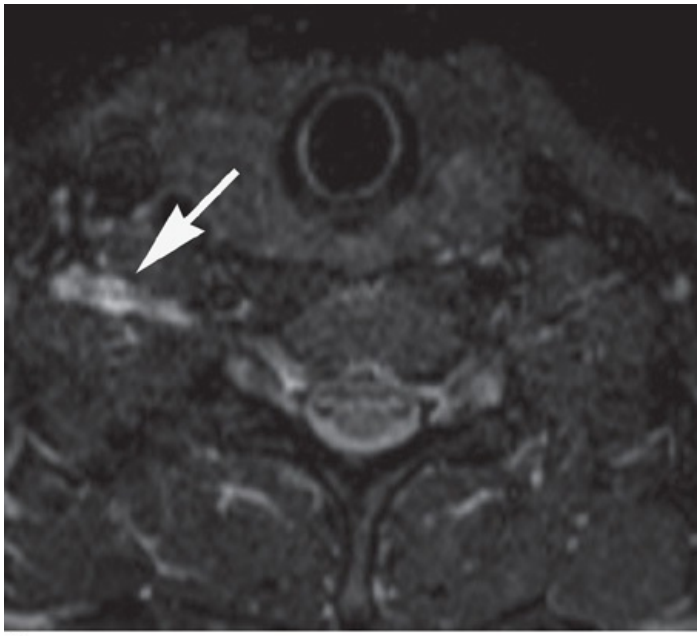

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

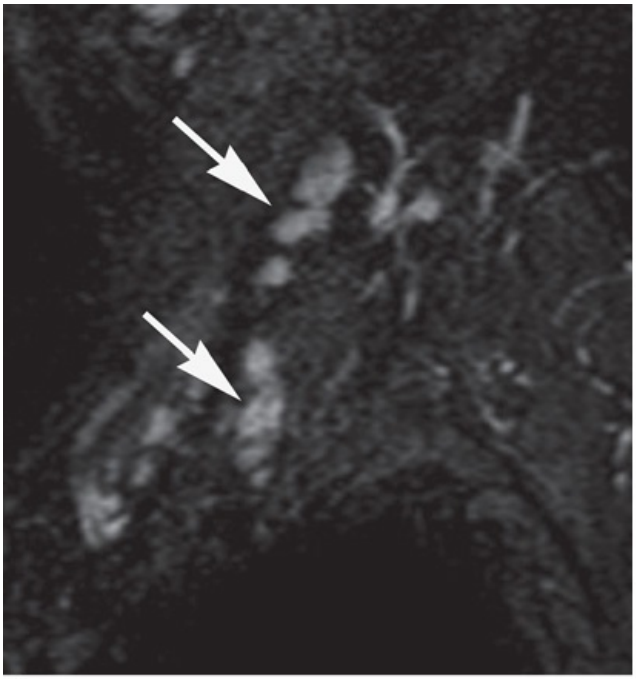

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

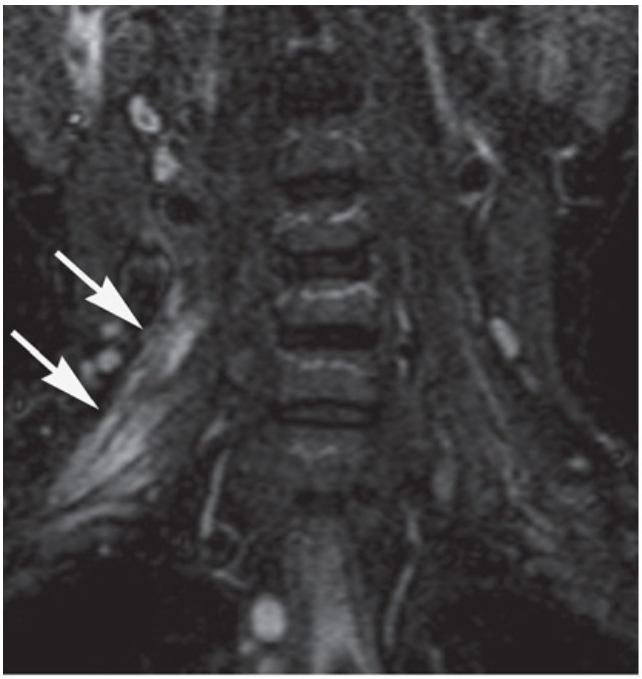

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

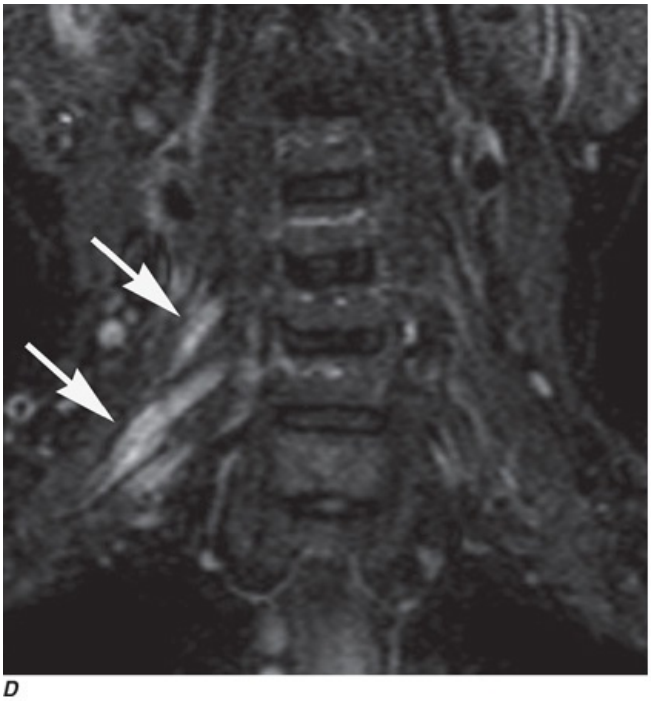

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

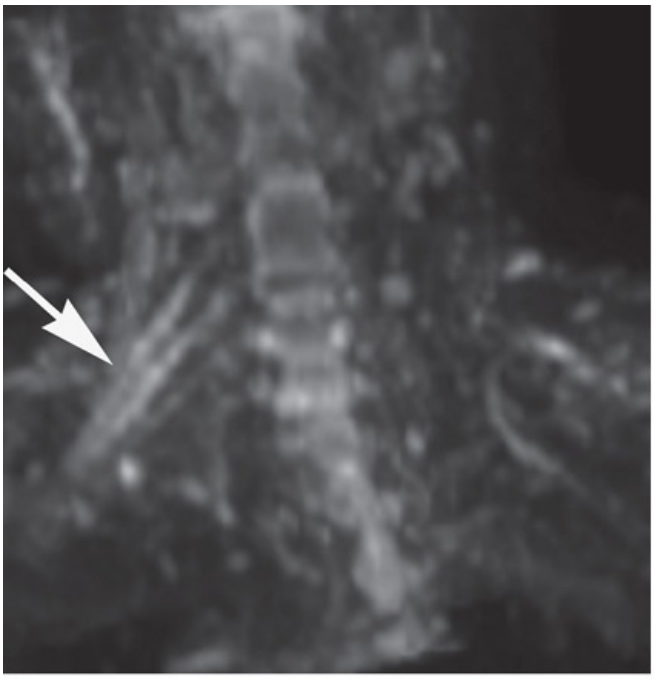

E

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Brachial plexopathy (Chap. 384)

Axial (***A***), sagittal (***B***), and coronal (***C***, ***D***) short tau inversion recovery (STIR) MR images demonstrate abnormal enlargement and abnormal high signal involving the right C6, C7, and C8 nerve roots, and the trunks and divisions that originate from these roots *(arrows*).

Diffusion-weighted MR imaging (***E***) demonstrates abnormal reduced diffusion within the right C6, C7, C8 nerve roots and their corresponding trunks and divisions (*arrow*). These findings are compatible with radiation-induced brachial plexopathy.

Figure e44-30

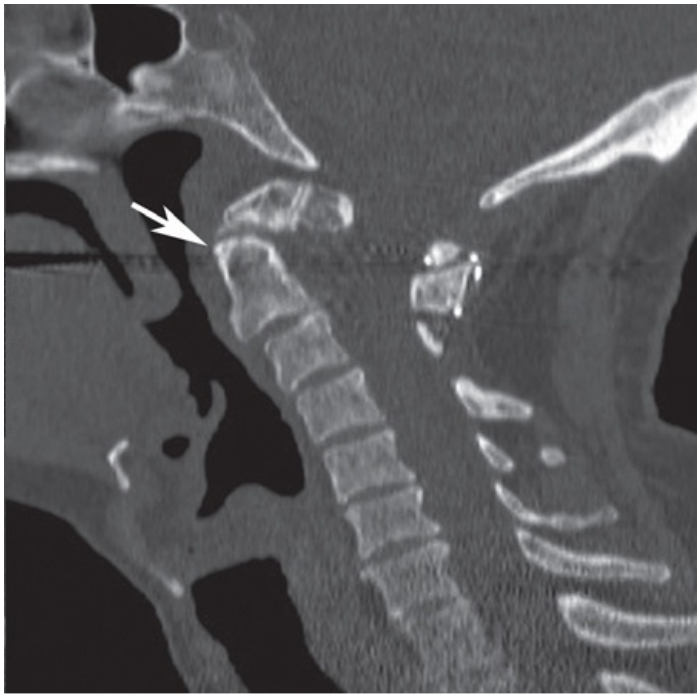

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anterior dens dislocation

Sagittal CT demonstrates the tip of the dens below the anterior arch of C2 (*arrow*), indicating anterior dislocation.

Figure e44-31

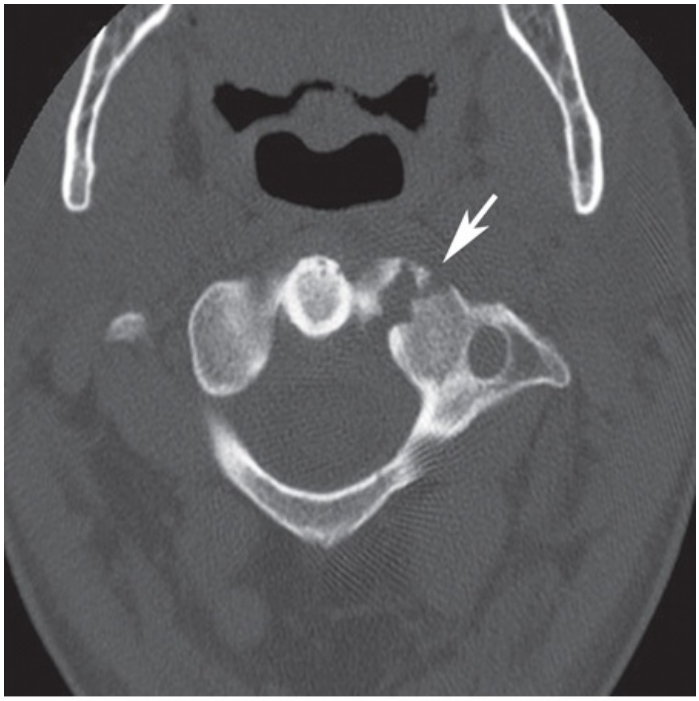

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT facet fracture

Axial CT demonstrates fracture line along the C2 facet (*arrow*).

Figure e44-32

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Compression fracture

Sagittal T2-weighted MRI demonstrates compression fracture of C7 (*) and high signal within the spinous processes of C6-C7 (*arrows*) and to lesser degree C5-C6. This is suggestive of interspinous ligament injury. Note the pad under the patient's neck to maintain neck alignment during the scanning time.

Figure e44-33

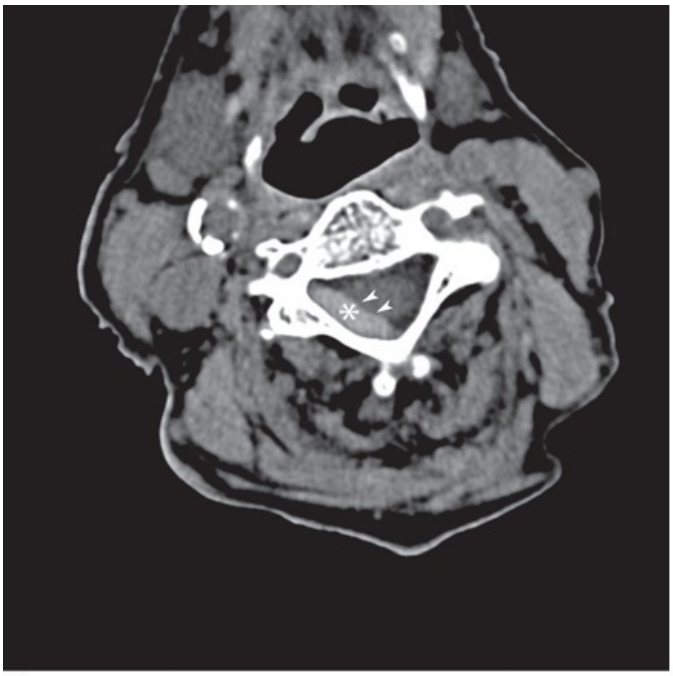

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Epidural hematoma

Axial noncontrast CT (***A***) demonstrates a high-density epidural collection in the cervical spine (*), which is consistent with acute hemorrhage. Also noted is mass effect on the spinal cord (*arrowheads*). Sagittal reformatted CT image (***B***) demonstrates the extension of the acute epidural hematoma (*) and a disk bulge (*arrowhead*), which further contributes to spinal canal narrowing. CT is the imaging procedure of choice to detect acute hematoma.

Figure e44-34

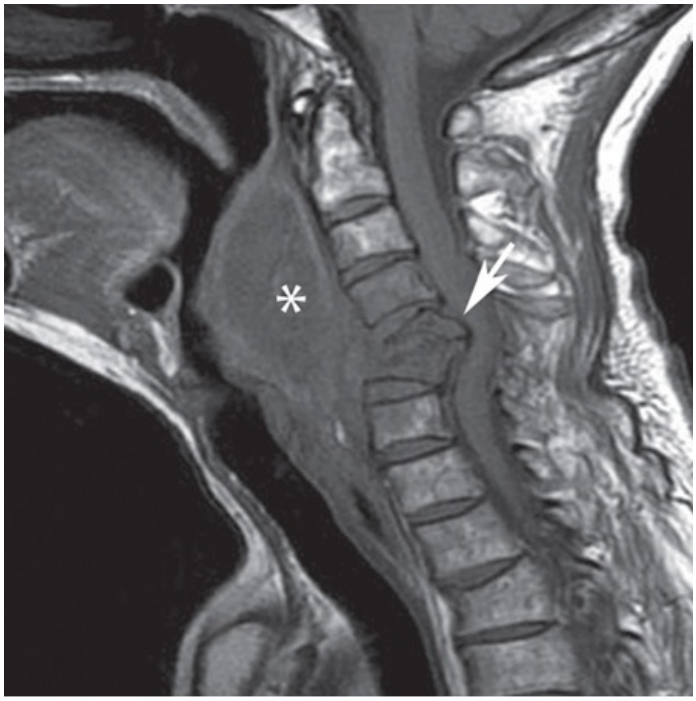

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Retropharyngeal soft tissue mass

Sagittal T1-weighted MRI demonstrates a hyperflexion fracture with retropulsion of the posterior wall in the canal at C5 and C6 (*arrow*). There is also a large retropharyngeal hematoma (*). The distance

from the posterior wall of the airway to the anterior wall of the vertebral body should not measure more than 6 mm at C2 or more than 20 mm at C6 (mnemonic "6 at 2 and 20 at 6").

Figure e44-35

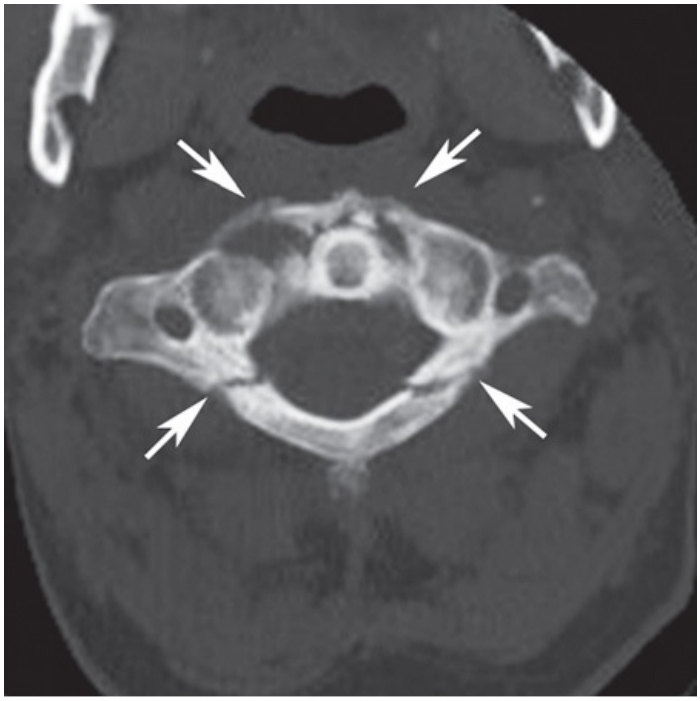

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Jefferson fracture

Axial CT demonstrates four fracture lines (*arrows*) separating C1 in four parts. Jefferson fracture is usually caused by axial impact to the head such as diving in shallow water.

Figure e44-36

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Ligament injury after trauma

Coronal CT reconstruction demonstrates abnormal asymmetry between the dens and the lateral masses of C1 indicating transverse ligament rupture.

Figure e44-37

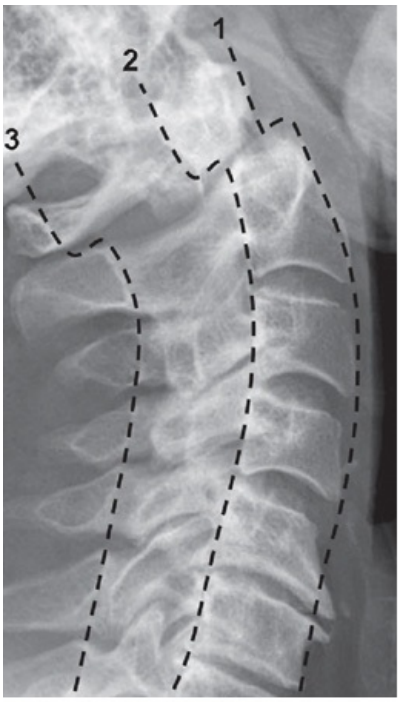

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Odontoid fracture

Sagittal CT demonstrates disruption of the main reference cervical lines. 1: Anterior vertebral body line; 2: Posterior vertebral body line; 3: Spinolaminar line.

Figure e44-38

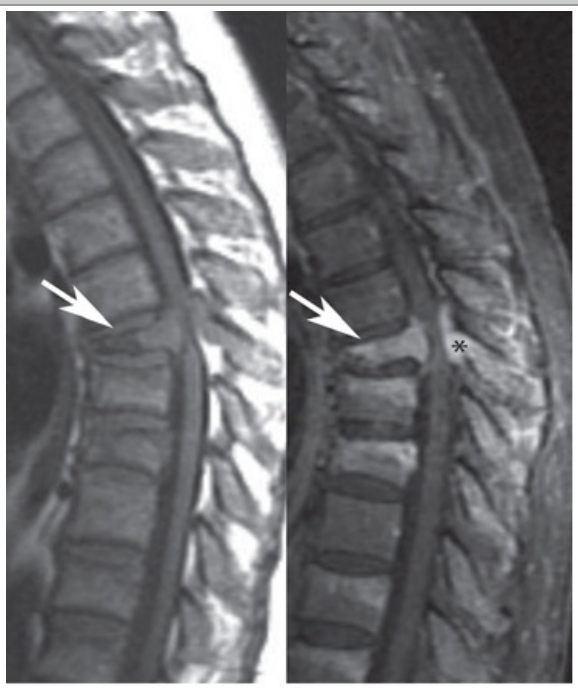

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathologic fracture

Sagittal T1-weighted MRI (***A***) demonstrates wedge-shaped T6 vertebral body (*arrow*). Sagittal postcontrast T1-weighted MRI (***B***) depicts tumor extension into the epidural space and the involvement of the posterior arch (*), which are highly suggestive of metastatic or primary bone tumor.

Figure e44-39

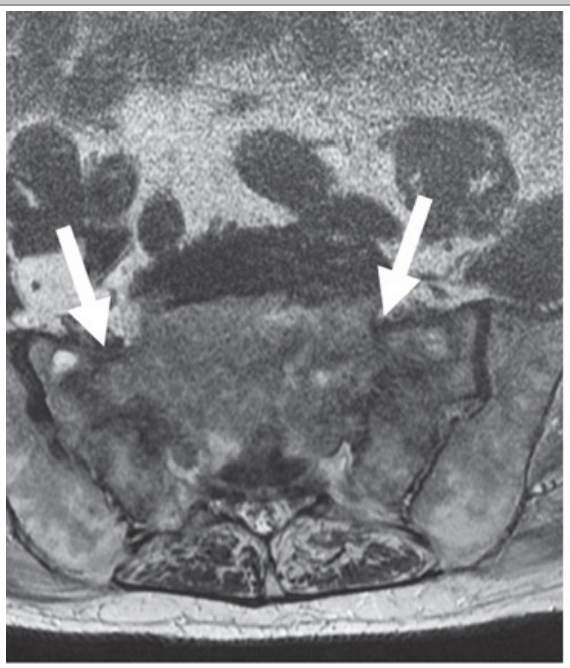

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

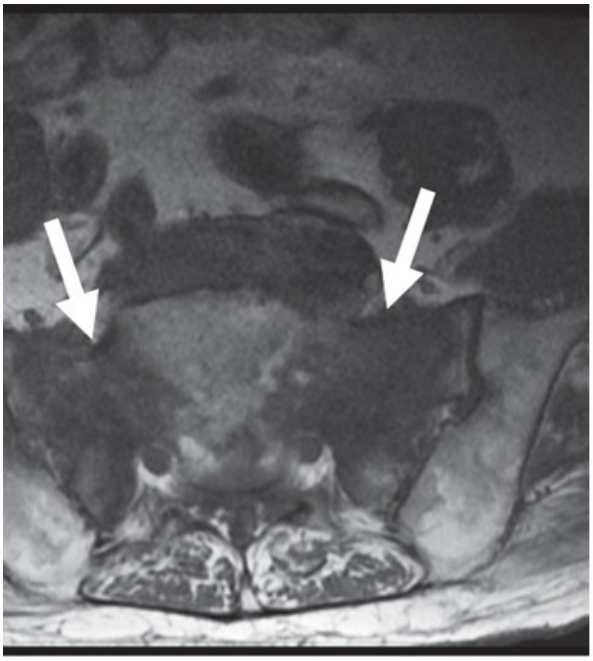

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sacral insufficiency fracture

Axial T2-weighted MRI (***A***) and T1-weighted MRI (***B***) demonstrate symmetric high T2 and low T1 signal involving the sacral alae longitudinally (*arrows*).

Figure e44-40

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Subdural hematoma

Sagittal T2-weighted MRI (***A***) and axial noncontrast T1-weighted MRI (***B***) demonstrate subdural collection in the lumbosacral region (**). Note that the epidural fat is compressed but not involved (*arrow*).

Figure e44-41

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Teardrop fracture

Sagittal CT (***A***) demonstrates fracture line separating the antero-inferior corner of C6 (*arrow*). Sagittal T2-weighted MRI (***B***) displays cord injury (*arrow*).

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ELECTROENCEPHALOGRAPHY

The electrical activity of the brain [the electroencephalogram (EEG)] is easily recorded from electrodes placed on the scalp. The potential difference between pairs of electrodes on the scalp (bipolar derivation) or between individual scalp electrodes and a relatively inactive common reference point (referential derivation) is amplified and displayed on a computer monitor, oscilloscope, or paper. The characteristics of the normal EEG depend on the patient's age and level of arousal. The rhythmic activity normally recorded represents the postsynaptic potentials of vertically oriented pyramidal cells of the cerebral cortex and is characterized by its frequency. In normal awake adults lying quietly with the eyes closed, an 8- to 13-Hz alpha rhythm is seen posteriorly in the EEG, intermixed with a variable amount of generalized faster (beta) activity (>13 Hz); the alpha rhythm is attenuated when the eyes are opened **(Fig. e45-1)**. During drowsiness, the alpha rhythm is also attenuated; with light sleep, slower activity in the theta (4–7 Hz) and delta (<4 Hz) ranges becomes more conspicuous.

Figure e45-1

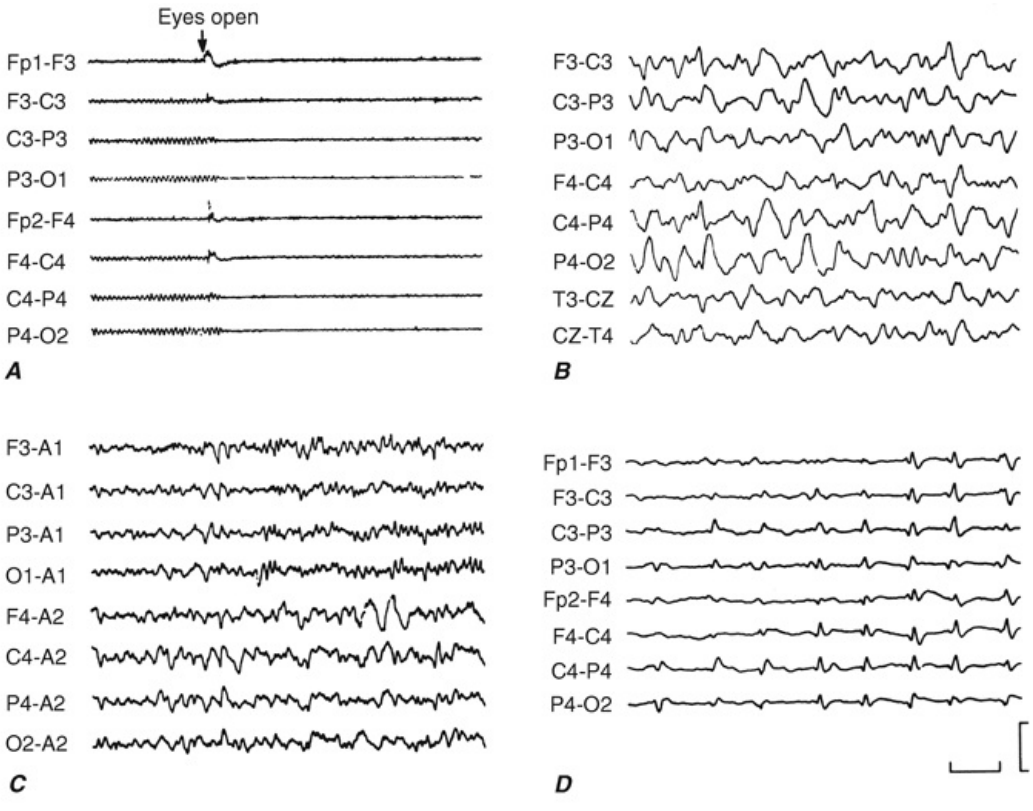

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A. Normal EEG showing a posteriorly situated 9-Hz alpha rhythm that attenuates with eye opening. ***B.*** Abnormal EEG showing irregular diffuse slow activity in an obtunded patient with encephalitis. ***C.*** Irregular slow activity in the right central region, on a diffusely slowed background, in a patient with a right parietal glioma. ***D.*** Periodic complexes occurring once every second in a patient with Creutzfeldt-Jakob disease. Horizontal calibration: 1 s; vertical calibration: 200 μV in A, 300 μV in other panels. (*From MJ Aminoff, ed: Electrodiagnosis in Clinical Neurology, 5th ed. New York, Churchill Livingstone, 2005.*) In this and the followingfigure, electrode placements are indicated at the left of each panel and accord with the international 10:20 system. A, earlobe; C, central; F, frontal; Fp, frontal polar; P, parietal; T, temporal; O, occipital. Right-sided placements are indicated by even numbers, left-sided placements by odd numbers, and midline placements by Z.

Digital systems are now widely used for recording the EEG. They allow the EEG to be reconstructed

and displayed with any desired format and manipulated for more detailed analysis, and also permit computerized techniques to be used to detect certain abnormalities. Activating procedures are generally undertaken while the EEG is recorded in an attempt to provoke abnormalities. Such procedures commonly include hyperventilation (for 3 or 4 min), photic stimulation, sleep, and sleep deprivation on the night prior to the recording.

Electroencephalography is relatively inexpensive and may aid clinical management in several different contexts.

The Eeg and Epilepsy

The EEG is most useful in evaluating patients with suspected epilepsy. The presence of electrographic seizure activity—i.e., of abnormal, repetitive, rhythmic activity having an abrupt onset and termination and a characteristic evolution—clearly establishes the diagnosis. The absence of such electrocerebral accompaniment does not exclude a seizure disorder, however, because there may be no change in the scalp-recorded EEG during certain focal seizures. With generalized tonic-clonic seizures, the EEG is always abnormal during the episode. It is often not possible to obtain an EEG during clinical events that may represent seizures, especially when such events occur unpredictably or infrequently. Continuous monitoring for prolonged periods in video-EEG telemetry units has made it easier to capture the electrocerebral accompaniments of such clinical episodes. Monitoring by these means is sometimes helpful in confirming that seizures are occurring, characterizing the nature of clinically equivocal episodes, and determining the frequency of epileptic events.

The EEG findings may also be helpful in the interictal period by showing certain abnormalities that are strongly supportive of a diagnosis of epilepsy. Such *epileptiform activity* consists of bursts of abnormal discharges containing spikes or sharp waves. The presence of epileptiform activity is not specific for epilepsy, but it has a much greater prevalence in epileptic patients than in normal individuals. However, even in an individual who is known to have epilepsy, the initial routine interictal EEG may be normal up to 60% of the time. Thus, the EEG cannot establish the diagnosis of epilepsy in many cases.

The EEG findings have been used in classifying seizure disorders and selecting appropriate anticonvulsant medication for individual patients **(Fig. e45-2)**. The episodic generalized spike-wave activity that occurs during and between seizures in patients with typical absence epilepsy contrasts with focal interictal epileptiform discharges or ictal patterns found in patients with focal seizures. These latter seizures may have no correlates in the scalp-recorded EEG or may be associated with abnormal rhythmic activity of variable frequency, a localized or generalized distribution, and a stereotyped pattern that varies with the patient. Focal or lateralized epileptogenic lesions are important to recognize, especially if surgical treatment is contemplated. Intensive long-term monitoring of clinical behavior and the EEG is required for operative candidates, however, and this generally also involves recording from intracranially placed electrodes (which may be subdural, extradural, or intracerebral in location).

Figure e45-2

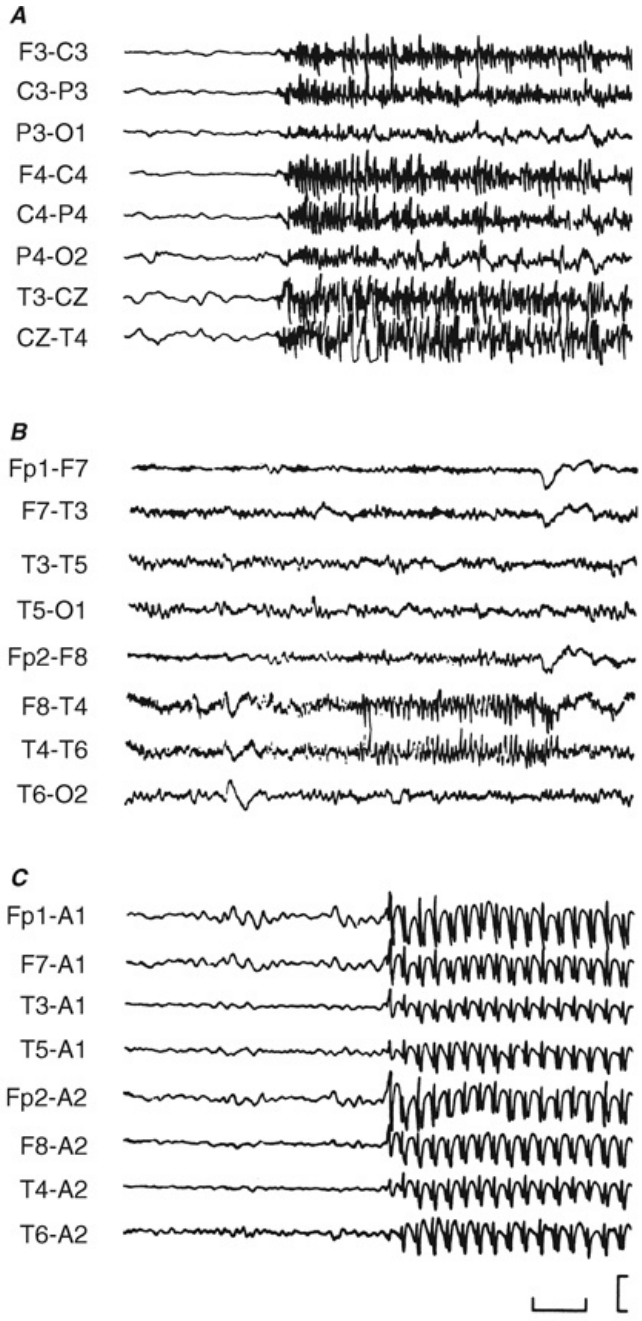

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Electrographic seizures.*A*. Onset of a tonic seizure showing generalized repetitive sharp activity with synchronous onset over both hemispheres. ***B***. Burst of repetitive spikes occurring with sudden onset in the right temporal region during a clinical spell characterized by transient impairment of external awareness. ***C***. Generalized 3-Hz spike-wave activity occurring synchronously over both hemispheres during an absence (petit mal) attack. Horizontal calibration: 1 s; vertical calibration: 400 mV in ***A***, 200 mV in ***B***, and 750 mV in ***C***. (*From MJ Aminoff, ed: Electrodiagnosis in Clinical Neurology, 5th ed. New York, Churchill Livingstone, 2005.*)

The findings in the routine scalp-recorded EEG may indicate the prognosis of seizure disorders: In general, a normal EEG implies a better prognosis than otherwise, whereas an abnormal background or profuse epileptiform activity suggests a poor outlook. The EEG findings are not helpful in determining which patients with head injuries, stroke, or brain tumors will go on to develop seizures, because in such circumstances epileptiform activity is commonly encountered regardless of whether seizures occur. The EEG findings are sometimes used to determine whether anticonvulsant medication can be discontinued in epileptic patients who have been seizure-free for several years, but the findings provide only a general guide to prognosis. Further seizures may occur after withdrawal of anticonvulsant medication despite a normal EEG or, conversely, may not occur despite a continuing EEG abnormality. The decision to discontinue anticonvulsant medication is made on clinical grounds, and the EEG does not have a useful role in this context except for providing guidance when there is

clinical ambiguity or the patient requires reassurance about a particular course of action.

The EEG has no role in the management of tonic-clonic status epilepticus except when there is clinical uncertainty whether seizures are continuing in a comatose patient. In patients treated by pentobarbital-induced coma for refractory status epilepticus, the EEG findings are useful in indicating the level of anesthesia and whether seizures are occurring. During status epilepticus, the EEG shows repeated electrographic seizures or continuous spike-wave discharges. In nonconvulsive status epilepticus, a disorder that may not be recognized unless an EEG is performed, the EEG may also show continuous spike-wave activity ("spike-wave stupor") or, less commonly, repetitive electrographic seizures (focal status epilepticus).

The Eeg and Coma

In patients with an altered mental state or some degree of obtundation, the EEG tends to become slower as consciousness is depressed, regardless of the underlying cause (Fig. e45-1). Other findings may also be present and may suggest diagnostic possibilities, as when electrographic seizures are found or there is a focal abnormality indicating a structural lesion. The EEG generally slows in metabolic encephalopathies, and triphasic waves may be present. The findings do not permit differentiation of the underlying metabolic disturbance but help to exclude other encephalopathic processes by indicating the diffuse extent of cerebral dysfunction. The response of the EEG to external stimulation is helpful prognostically because electrocerebral responsiveness implies a lighter level of coma than a nonreactive EEG. Serial records provide a better guide to prognosis than a single record and supplement the clinical examination in following the course of events. As the depth of coma increases, the EEG becomes nonreactive and may show a burst-suppression pattern, with bursts of mixed-frequency activity separated by intervals of relative cerebral inactivity. In other instances there is a reduction in amplitude of the EEG until eventually activity cannot be detected. Such electrocerebral silence does not necessarily reflect irreversible brain damage, because it may occur in hypothermic patients or with drug overdose. The prognosis of electrocerebral silence, when recorded using an adequate technique, depends upon the clinical context in which it is found. In patients with severe cerebral anoxia, for example, electrocerebral silence in a technically satisfactory record implies that useful cognitive recovery will not occur.

In patients with clinically suspected brain death, an EEG, when recorded using appropriate technical standards, may be confirmatory by showing electrocerebral silence. However, complicating disorders that may produce a similar but reversible EEG appearance (e.g., hypothermia or drug intoxication) must be excluded. The presence of residual EEG activity in suspected brain death fails to confirm the diagnosis but does not exclude it. The EEG is usually normal in patients with locked-in syndrome and helps in distinguishing this disorder from the comatose state with which it is sometimes confused clinically.

The Eeg in Other Neurologic Disorders

In the developed countries, CT scanning and MRI have taken the place of EEG as a noninvasive means of screening for focal structural abnormalities of the brain, such as tumors, infarcts, or hematomas (Fig. e45-1). Nonetheless, the EEG is still used for this purpose in many parts of the world, although infratentorial or slowly expanding lesions may fail to cause any abnormalities. Focal slow-wave disturbances, a localized loss of electrocerebral activity, or more generalized electrocerebral disturbances are common findings but provide no reliable indication about the nature of the underlying pathology.

In patients with an acute encephalopathy, focal or lateralized periodic slow-wave complexes, sometimes with a sharpened outline, suggest a diagnosis of herpes simplex encephalitis, and periodic lateralizing epileptiform discharges (PLEDs) are commonly found with acute hemispheric pathology such as a hematoma, abscess, or rapidly expanding tumor. The EEG findings in dementia are usually nonspecific and do not distinguish between the different causes of cognitive decline except in rare instances when, for example, the presence of complexes occurring with a regular repetition rate (so-called periodic complexes) supports a diagnosis of Creutzfeldt-Jakob disease (Fig. e45-1) or subacute sclerosing panencephalitis. In most patients with dementias, the EEG is normal or diffusely slowed, and the EEG findings alone cannot indicate whether a patient is demented or distinguish between dementia and pseudodementia.

Continuous Eeg Monitoring

The brief EEG obtained routinely in the laboratory often fails to reveal abnormalities that are transient and infrequent. Continuous monitoring over 12 or 24 hours or longer may detect abnormalities or capture clinical events that would otherwise be missed. The EEG is often recorded continuously in critically ill patients to detect early changes in neurologic status, which is particularly useful when the clinical examination is limited. Continuous EEG recording in this context has been used to detect acute events such as nonconvulsive seizures or developing cerebral ischemia, to monitor cerebral function in patients with metabolic disorders such as liver failure, and to manage the level of anesthesia in pharmacologically induced coma.

MAGNETOENCEPHALOGRAPHY AND MAGNETIC SOURCE IMAGING

Recording the magnetic field of the electrical activity of the brain [magnetoencephalography (MEG)] provides a means of examining cerebral activity that is less subject to distortion by other biologic tissues than the EEG. MEG is used in only a few specialized centers because of the complexity and expense of the necessary equipment. It permits the source of activity to be localized and coregistered with the MRI in a technique that is known as *magnetic source imaging*. In patients with focal epilepsy, MEG is useful in localizing epileptogenic foci for surgery and for guiding the placement of intracranial electrodes for electrophysiologic monitoring. MEG has also been used for mapping brain tumors, identifying the central fissure preoperatively, and localizing functionally eloquent cortical areas such as those concerned with language.

EVOKED POTENTIALS

Sensory Evoked Potentials

The noninvasive recording of spinal or cerebral potentials elicited by stimulation of specific afferent pathways is an important means of monitoring the functional integrity of these pathways but does not indicate the pathologic basis of lesions involving them. Such evoked potentials (EPs) are so small compared to the background EEG activity that the responses to a number of stimuli have to be recorded and averaged with a computer in order to permit their recognition and definition. The background EEG activity, which has no fixed temporal relationship to the stimulus, is averaged out by this procedure.

Visual evoked potentials (VEPs) are elicited by monocular stimulation with a reversing checkerboard pattern and are recorded from the occipital region in the midline and on either side of the scalp. The component of major clinical importance is the so-called P100 response, a positive peak having a latency of approximately 100 ms. Its presence, latency, and symmetry over the two sides of the scalp are noted. Amplitude may also be measured, but changes in size are much less helpful for the recognition of pathology. VEPs are most useful in detecting dysfunction of the visual pathways anterior to the optic chiasm. In patients with acute severe optic neuritis, the P100 is frequently lost or grossly attenuated; as clinical recovery occurs and visual acuity improves, the P100 is restored but with an increased latency that generally remains abnormally prolonged indefinitely. The VEP findings are therefore helpful in indicating previous or subclinical optic neuritis. They may also be abnormal with ocular abnormalities and with other causes of optic nerve disease, such as ischemia or compression by a tumor. Normal VEPs may be elicited by flash stimuli in patients with cortical blindness. Routine VEPs record a mass response over a relatively large cortical area and thus may be insensitive to localized waveform abnormalities. A newer technique, *multifocal VEP*, measures responses from 120 individual sectors within each affected eye, and thus is likely to be more sensitive than routine VEP.
Brainstem auditory evoked potentials (BAEPs) are elicited by monaural stimulation with repetitive clicks and are recorded between the vertex of the scalp and the mastoid process or earlobe. A series of potentials, designated by roman numerals, occurs in the first 10 ms after the stimulus and represents in part the sequential activation of different structures in the pathway between the auditory nerve (wave I) and the inferior colliculus (wave V) in the midbrain. The presence, latency, and interpeak latency of the first five positive potentials recorded at the vertex are evaluated. The findings are helpful in screening for acoustic neuromas, detecting brainstem pathology, and evaluating comatose patients. The BAEPs are normal in coma due to metabolic/toxic disorders or bihemispheric disease but abnormal in the presence of brainstem pathology.

Somatosensory evoked potentials (SSEPs) are recorded over the scalp and spine in response to electrical stimulation of a peripheral (mixed or cutaneous) nerve. The configuration, polarity, and latency of the responses depend on the nerve that is stimulated and on the recording arrangements. SSEPs are used to evaluate proximal (otherwise inaccessible) portions of the peripheral nervous system and the integrity of the central somatosensory pathways.

CLINICAL UTILITY OF EPS

EP studies may detect and localize lesions in afferent pathways in the central nervous system (CNS). They have been used particularly to investigate patients with suspected multiple sclerosis (MS), the diagnosis of which requires the recognition of lesions involving several different regions of the central white matter. In patients with clinical evidence of only one lesion, the electrophysiologic recognition of abnormalities in other sites helps to suggest or support the diagnosis but does not establish it unequivocally. Multimodality EP abnormalities are not specific for MS; they may occur in AIDS, Lyme disease, systemic lupus erythematosus, neurosyphilis, spinocerebellar degenerations, familial spastic paraplegia, and deficiency of vitamin E or B_{12}, among other disorders. The diagnostic utility of the electrophysiologic findings therefore depends on the circumstances in which they are found. Abnormalities may aid in the localization of lesions to broad areas of the CNS, but attempts at precise localization on electrophysiologic grounds are misleading because the generators of many components of the EP are unknown.

The EP findings are sometimes of prognostic relevance. Bilateral loss of SSEP components that are

generated in the cerebral cortex implies that cognition may not be regained in posttraumatic or postanoxic coma, and EP studies may also be useful in evaluating patients with suspected brain death. In patients who are comatose for uncertain reasons, preserved BAEPs suggest either a metabolic-toxic etiology or bihemispheric disease. In patients with spinal cord injuries, SSEPs have been used to indicate the completeness of the lesion. The presence or early return of a cortically generated response to stimulation of a nerve below the injured segment of the cord indicates an incomplete lesion and thus a better prognosis for functional recovery than otherwise. In surgery, intraoperative EP monitoring of neural structures placed at risk by the procedure may permit the early recognition of dysfunction and thereby permit a neurologic complication to be averted or minimized.

Visual and auditory acuity may be determined using EP techniques in patients whose age or mental state precludes traditional ophthalmologic or audiologic examinations.

Cognitive Evoked Potentials

Certain EP components depend on the mental attention of the subject and the setting in which the stimulus occurs, rather than simply on the physical characteristics of the stimulus. Such "event-related" potentials (ERPs) or "endogenous" potentials are related in some manner to the cognitive aspects of distinguishing an infrequently occurring target stimulus from other stimuli occurring more frequently. For clinical purposes, attention has been directed particularly at the so-called P3 component of the ERP, which is also designated the P300 component because of its positive polarity and latency of approximately 300–400 ms after onset of an auditory target stimulus. The P3 component is prolonged in latency in many patients with dementia, whereas it is generally normal in patients with depression or other psychiatric disorders that might be mistaken for dementia. ERPs are, therefore, sometimes helpful in making this distinction when there is clinical uncertainty, although a response of normal latency does not exclude dementia.

Motor Evoked Potentials

The electrical potentials recorded from muscle or the spinal cord following stimulation of the motor cortex or central motor pathways are referred to as *motor evoked potentials* . For clinical purposes such responses are recorded most often as the compound muscle action potentials elicited by transcutaneous magnetic stimulation of the motor cortex. A strong but brief magnetic field is produced by passing a current through a coil, and this induces stimulating currents in the subjacent neural tissue. The procedure is painless and apparently safe. Abnormalities have been described in several neurologic disorders with clinical or subclinical involvement of central motor pathways, including MS and motor neuron disease. In addition to a possible role in the diagnosis of neurologic disorders or in evaluating the extent of pathologic involvement, the technique provides information of prognostic relevance (e.g., in suggesting the likelihood of recovery of motor function after stroke) and provides a means of monitoring intraoperatively the functional integrity of central motor tracts. Nevertheless, it is not used widely for clinical purposes.

ELECTROPHYSIOLOGIC STUDIES OF MUSCLE AND NERVE

The motor unit is the basic element subserving motor function. It is defined as an anterior horn cell, its axon and neuromuscular junctions, and all the muscle fibers innervated by the axon. The number of motor units in a muscle ranges from approximately 10 in the extraocular muscles to several thousand in the large muscles of the legs. There is considerable variation in the average number of muscle fibers within the motor units of an individual muscle, i.e., in the innervation ratio of different muscles. Thus the innervation ratio is <25 in the human external rectus or platysma muscle and between 1600 and 1700 in the medial head of the gastrocnemius muscle. The muscle fibers of individual motor units are divided into two general types by distinctive contractile properties, histochemical stains, and characteristic responses to fatigue. Within each motor unit, all of the muscle fibers are of the same type.

Electromyography

The pattern of electrical activity in muscle [i.e., the electromyogram (EMG)], both at rest and during activity, may be recorded from a needle electrode inserted into the muscle. The nature and pattern of abnormalities relate to disorders at different levels of the motor unit.

Relaxed muscle normally is electrically silent except in the end plate region, but abnormal spontaneous activity (**Fig. e45-3**) occurs in various neuromuscular disorders, especially those associated with denervation or inflammatory changes in affected muscle. Fibrillation potentials and positive sharp waves (which reflect muscle fiber irritability) and complex repetitive discharges are most often—but not always—found in denervated muscle and may also occur after muscle injury and in certain myopathic disorders, especially inflammatory disorders such as polymyositis. After an acute neuropathic lesion, they are found earlier in proximal rather than distal muscles and sometimes do not develop distally in the extremities for 4–6 weeks; once present, they may persist indefinitely unless reinnervation occurs or the muscle degenerates so completely that no viable tissue remains. Fasciculation potentials (which reflect the spontaneous activity of individual motor units) are characteristic of slowly progressive neuropathic disorders, especially those with degeneration of anterior horn cells (such as amyotrophic

lateral sclerosis). Myotonic discharges—high-frequency discharges of potentials derived from single muscle fibers that wax and wane in amplitude and frequency—are the signature of myotonic disorders such as myotonic dystrophy or myotonia congenita but occur occasionally in polymyositis or other, rarer, disorders.

Figure e45-3

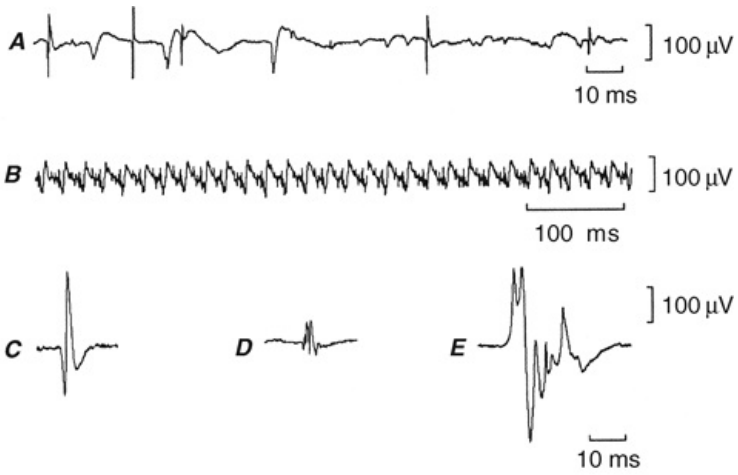

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Activity recorded during EMG.*A*. Spontaneous fibrillation potentials and positive sharp waves. ***B***. Complex repetitive discharges recorded in partially denervated muscle at rest. ***C***. Normal triphasic motor unit action potential. ***D***. Small, short-duration, polyphasic motor unit action potential such as is commonly encountered in myopathic disorders. ***E***. Long-duration polyphasic motor unit action potential such as may be seen in neuropathic disorders.

Slight voluntary contraction of a muscle leads to activation of a small number of motor units. The potentials generated by any muscle fibers of these units that are within the pickup range of the needle electrode will be recorded (Fig. e45-3). The parameters of normal motor unit action potentials depend on the muscle under study and age of the patient, but their duration is normally between 5 and 15 ms, amplitude is between 200 µV and 2 mV, and most are bi- or triphasic. The number of units activated depends on the degree of voluntary activity. An increase in muscle contraction is associated with an increase in the number of motor units that are activated (recruited) and in the frequency with which they discharge. With a full contraction, so many motor units are normally activated that individual motor unit action potentials can no longer be distinguished, and a complete interference pattern is said to have been produced.

The incidence of small, short-duration, polyphasic motor unit action potentials (i.e., having more than four phases) is usually increased in myopathic muscle, and an excessive number of units is activated for a specified degree of voluntary activity. By contrast, the loss of motor units that occurs in neuropathic disorders leads to a reduction in number of units activated during a maximal contraction and an increase in their firing rate, i.e., there is an incomplete or reduced interference pattern. The configuration and dimensions of the potentials may also be abnormal, depending on the duration of the neuropathic process and on whether reinnervation has occurred. The surviving motor units are initially normal in configuration but, as reinnervation occurs, they increase in amplitude and duration and become polyphasic (Fig. e45-3).

Action potentials from the same motor unit sometimes fire with a consistent temporal relationship to each other, so that double, triple, or multiple discharges are recorded, especially in tetany, hemifacial spasm, or myokymia.

Electrical silence characterizes the involuntary, sustained muscle contraction that occurs in phosphorylase deficiency, which is designated a contracture.

EMG enables disorders of the motor units to be detected and characterized as either neurogenic or myopathic. In neurogenic disorders, the pattern of affected muscles may localize the lesion to the anterior horn cells or to a specific site as the axons traverse a nerve root, limb plexus, and peripheral nerve to their terminal arborizations. The findings do not enable a specific etiologic diagnosis to be made, however, except in conjunction with the clinical findings and results of other laboratory studies.

The findings may provide a guide to the severity of an acute disorder of a peripheral or cranial nerve (by indicating whether denervation has occurred and the completeness of the lesion) and whether the pathologic process is active or progressive in chronic or degenerative disorders such as amyotrophic lateral sclerosis. Such information is important for prognostic purposes.

Various quantitative EMG approaches have been developed. The most common is to determine the mean duration and amplitude of 20 motor unit action potentials using a standardized technique. The technique of macro-EMG provides information about the number and size of muscle fibers in a larger volume of the motor unit territory and has also been used to estimate the number of motor units in a muscle. Scanning EMG is a computer-based technique that has been used to study the topography of motor unit action potentials and, in particular, the spatial and temporal distribution of activity in individual units. The technique of single-fiber EMG is discussed separately below.

Nerve Conduction Studies

Recording of the electrical response of a muscle to stimulation of its motor nerve at two or more points along its course **(Fig. e45-4)** permits conduction velocity to be determined in the fastest-conducting motor fibers between the points of stimulation. The latency and amplitude of the electrical response of muscle (i.e., of the compound muscle action potential) to stimulation of its motor nerve at a distal site are also compared with values defined in normal subjects. Sensory nerve conduction studies are performed by determining the conduction velocity and amplitude of action potentials in sensory fibers when these fibers are stimulated at one point and the responses are recorded at another point along the course of the nerve. In adults, conduction velocity in the arms is normally between 50 and 70 m/s, and in the legs is between 40 and 60 m/s.

Figure e45-4

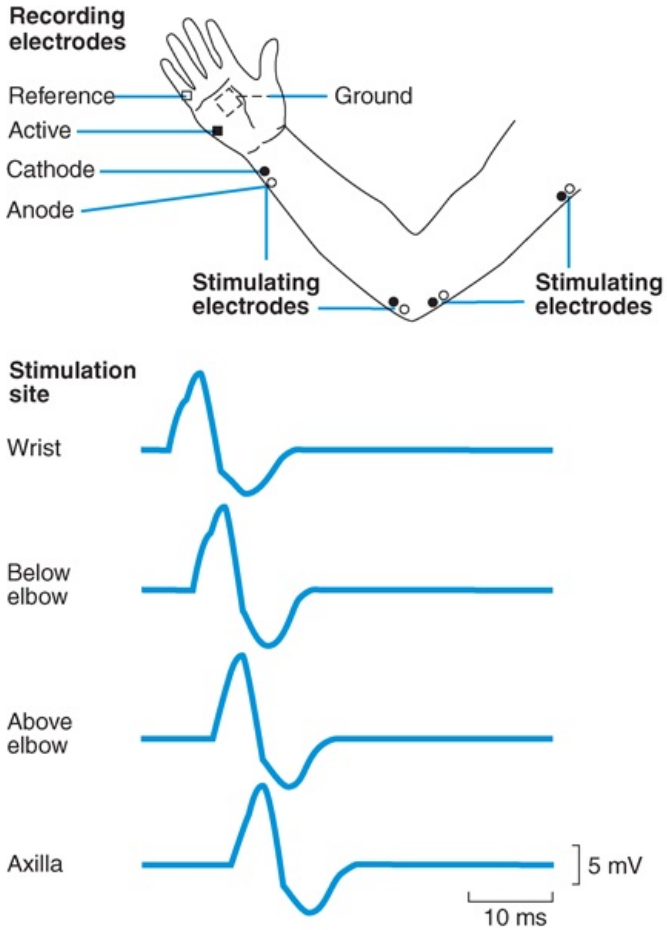

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arrangement for motor conduction studies of the ulnar nerve. Responses are recorded with a surface electrode from the abductor digiti minimi muscle to supramaximal stimulation of the nerve at different sites, and are shown in the lower panel. (*From Aminoff, 1998.*)

Nerve conduction studies complement the EMG examination, enabling the presence and extent of peripheral nerve pathology to be determined. They are particularly helpful in determining whether sensory symptoms are arising from pathology proximal or distal to the dorsal root ganglia (in the former instance, peripheral sensory conduction studies will be normal) and whether neuromuscular dysfunction relates to peripheral nerve disease. In patients with a mononeuropathy, they are invaluable as a means of localizing a focal lesion, determining the extent and severity of the underlying pathology, providing a guide to prognosis, and detecting subclinical involvement of other peripheral nerves. They enable a polyneuro-pathy to be distinguished from a mononeuropathy multiplex when this is not

possible clinically, an important distinction because of the etiologic implications. Nerve conduction studies provide a means of following the progression and therapeutic response of peripheral nerve disorders and are being used increasingly for this purpose in clinical trials. They may suggest the underlying pathologic basis in individual cases. Conduction velocity is often markedly slowed, terminal motor latencies are prolonged, and compound motor and sensory nerve action potentials may be dispersed in the demyelinative neuropathies (such as in Guillain-Barré syndrome, chronic inflammatory polyneuropathy, metachromatic leukodystrophy, or certain hereditary neuropathies); conduction block is frequent in acquired varieties of these neuropathies. By contrast, conduction velocity is normal or slowed only mildly, sensory nerve action potentials are small or absent, and there is EMG evidence of denervation in axonal neuropathies such as occur in association with metabolic or toxic disorders.

The utility and complementary role of EMG and nerve conduction studies are best illustrated by reference to a common clinical problem. Numbness and paresthesias of the little finger and associated wasting of the intrinsic muscles of the hand may result from a spinal cord lesion, C8/T1 radiculopathy, brachial plexopathy (lower trunk or medial cord), or a lesion of the ulnar nerve. If sensory nerve action potentials can be recorded normally at the wrist following stimulation of the digital fibers in the affected finger, the pathology is probably proximal to the dorsal root ganglia (i.e., there is a radiculopathy or more central lesion); absence of the sensory potentials, by contrast, suggests distal pathology. EMG examination will indicate whether the pattern of affected muscles conforms to radicular or ulnar nerve territory, or is more extensive (thereby favoring a plexopathy). Ulnar motor conduction studies will generally also distinguish between a radiculopathy (normal findings) and ulnar neuropathy (abnormal findings) and will often identify the site of an ulnar nerve lesion. The nerve is stimulated at several points along its course to determine whether the compound action potential recorded from a distal muscle that it supplies shows a marked alteration in size or area or a disproportionate change in latency, with stimulation at a particular site. The electrophysiologic findings thus permit a definitive diagnosis to be made and specific treatment instituted in circumstances where there is clinical ambiguity.

F-Wave Studies

Stimulation of a motor nerve causes impulses to travel antidromically (i.e., toward the spinal cord) as well as orthodromically (to the nerve terminals). Such antidromic impulses cause a few of the anterior horn cells to discharge, producing a small motor response that occurs considerably later than the direct response elicited by nerve stimulation. The F wave so elicited is sometimes abnormal (absent or delayed) with proximal pathology of the peripheral nervous system, such as a radiculopathy, and may therefore be helpful in detecting abnormalities when conventional nerve conduction studies are normal. In general, however, the clinical utility of F-wave studies has been disappointing, except perhaps in Guillain-Barré syndrome, where they are often absent or delayed.

H-Reflex Studies

The H reflex is easily recorded only from the soleus muscle (S1) in normal adults. It is elicited by low-intensity stimulation of the tibial nerve and represents a monosynaptic reflex in which spindle (Ia) afferent fibers constitute the afferent arc and alpha motor axons the efferent pathway. The H reflexes are often absent bilaterally in elderly patients or with polyneuropathies and may be lost unilaterally in S1 radiculopathies.

Muscle Response to Repetitive Nerve Stimulation

The size of the electrical response of a muscle to supramaximal electrical stimulation of its motor nerve relates to the number of muscle fibers that are activated. Neuromuscular transmission can be tested by several different protocols, but the most helpful is to record with surface electrodes the electrical response of a muscle to supramaximal stimulation of its motor nerve by repetitive (2–3 Hz) shocks delivered before and at selected intervals after a maximal voluntary contraction.

There is normally little or no change in size of the compound muscle action potential following repetitive stimulation of a motor nerve at 2–3 Hz with stimuli delivered at intervals after voluntary contraction of the muscle for about 20–30 s, even though preceding activity in the junctional region influences the release of acetylcholine and thus the size of the end-plate potentials elicited by a test stimulus. This is because more acetylcholine is normally released than is required to bring the motor end-plate potentials to the threshold for generating muscle fiber action potentials. In disorders of neuromuscular transmission this safety factor is reduced. Thus in myasthenia gravis, repetitive stimulation, particularly at a rate of between 2 and 5 Hz, may lead to a depression of neuromuscular transmission, with a decrement in size of the response recorded from affected muscles. Similarly, immediately after a period of maximal voluntary activity, single or repetitive stimuli of the motor nerve may elicit larger muscle responses than before, indicating that more muscle fibers are responding. This postactivation facilitation of neuromuscular transmission is followed by a longer-lasting period of depression, maximal between 2 and 4 min after the conditioning period and lasting for as long as 10 min or so, during which responses are reduced in size.

Decrementing responses to repetitive stimulation at 2–5 Hz are common in myasthenia gravis but may also occur in the congenital myasthenic syndromes. In Lambert-Eaton myasthenic syndrome, in which

there is defective release of acetylcholine at the neuro-muscular junction, the compound muscle action potential elicited by a single stimulus is generally very small. With repetitive stimulation at rates of up to 10 Hz, the first few responses may decline in size, but subsequent responses increase. If faster rates of stimulation are used (20–50 Hz), the increment may be dramatic so that the amplitude of compound muscle action potentials eventually reaches a size that is several times larger than the initial response. In patients with botulism, the response to repetitive stimulation is similar to that in Lambert-Eaton myasthenic syndrome, although the findings are somewhat more variable and not all muscles are affected.

Single-Fiber Electromyography

This technique is particularly helpful in detecting disorders of neuromuscular transmission. A special needle electrode is placed within a muscle and positioned to record action potentials from two muscle fibers belonging to the same motor unit. The time interval between the two potentials will vary in consecutive discharges; this is called the *neuromuscular jitter* . The jitter can be quantified as the mean difference between consecutive interpotential intervals and is normally between 10 and 50 μs. This value is increased when neuromuscular transmission is disturbed for any reason, and in some instances impulses in individual muscle fibers may fail to occur because of impulse blocking at the neuromuscular junction. Single-fiber EMG is more sensitive than repetitive nerve stimulation or determination of acetylcholine receptor antibody levels in diagnosing myasthenia gravis.

Single-fiber EMG can also be used to determine the mean fiber density of motor units (i.e., mean number of muscle fibers per motor unit within the recording area) and to estimate the number of motor units in a muscle, but this is of less immediate clinical relevance.

Blink Reflexes

Electrical or mechanical stimulation of the supraorbital nerve on one side leads to two separate reflex responses of the orbicularis oculi—an ipsilateral R1 response having a latency of approximately 10 ms and a bilateral R2 response with a latency in the order of 30 ms. The trigeminal and facial nerves constitute the afferent and efferent arcs of the reflex, respectively. Abnormalities of either nerve or intrinsic lesions of the medulla or pons may lead to uni- or bilateral loss of the response, and the findings may therefore be helpful in identifying or localizing such pathology.

FURTHER READINGS

Amato A, Russell JA: *Neuromuscular Disorders.* New York, McGraw-Hill, 2008

Aminoff MJ: *Electromyography in Clinical Practice: Electrodiagnostic Aspects of Neuromuscular Disease*, 3rd ed. New York, Churchill Livingstone, 1998

——— (ed): *Electrodiagnosis in Clinical Neurology*, 5th ed. New York, Churchill Livingstone, 2005

Brown WF et al (eds): *Neuromuscular Function and Disease* . Philadelphia, Saunders, 2002

Ebersole JS, Pedley TA (eds): *Current Practice of Clinical Electroencephalography* , 3rd ed. Philadelphia, Lippincott Williams & Wilkins, 2003

Friedman d et al: Continuous electroencephalogram monitoring in the intensive care unit.

Anesth Analg 109:506, 2009

Holmes GL et al: *Clinical Neurophysiology of Infancy, Childhood, and Adolescence* . Philadelphia, Butterworth Heinemann, 2006

Rossi S et al: Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clin Neurophysiol 120:2008, 2009[PMID: 19833552] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TECHNIQUE OF LUMBAR PUNCTURE: INTRODUCTION

In experienced hands, lumbar puncture (LP) is usually a safe procedure. Major complications are extremely uncommon but can include cerebral herniation, injury to the spinal cord or nerve roots, hemorrhage, or infection. Minor complications occur with greater frequency and can include backache, post-LP headache, and radicular pain or numbness.

IMAGING AND LABORATORY STUDIES PRIOR TO LP

Patients with an altered level of consciousness, a focal neurologic deficit, new-onset seizure, papilledema, or an immunocompromised state are at increased risk for potentially fatal cerebellar or tentorial herniation following LP. Neuroimaging should be obtained in these patients prior to LP to exclude a focal mass lesion or diffuse swelling. Imaging studies should include the spine in patients with symptoms suggesting spinal cord compression, such as back pain, leg weakness, urinary retention, or incontinence. In patients with suspected meningitis who require neuroimaging prior to diagnostic LP, administration of antibiotics, preferably following blood culture, should precede the neuroimaging study.

Patients receiving therapeutic anticoagulation or those with coagulation defects including thrombocytopenia are at increased risk of post-LP spinal subdural or epidural hematomas, either of which can produce permanent nerve injury and/or paralysis. If a bleeding disorder is suspected, the platelet count, international normalized ratio (INR), and partial thromboplastin time should be checked prior to lumbar puncture. There are no data available to assess the safety of LP in patients with low platelet counts; a count of <20,000/μL is considered to be a contraindication to LP. Bleeding complications rarely occur in patients with platelet counts ≥50,000/μL and an INR ≤1.5. Patients receiving low-molecular-weight heparin are at increased risk of post-LP spinal or epidural hematoma, and doses should be held for 24 h before the procedure.

LP should not be performed through infected skin as organisms can be introduced into the subarachnoid space (SAS).

ANALGESIA

Anxiety and pain can be minimized prior to beginning the procedure. Anxiety can be allayed by the use of lorazepam, 1–2 mg given PO 30 min prior to the procedure or IV 5 min prior to the procedure. Topical anesthesia can be achieved by the application of a lidocaine-based cream. Lidocaine 4% is effective when applied 30 min prior to the procedure; lidocaine/prilocaine requires 60–120 min. The cream should be applied in a thick layer so that it completely covers the skin; an occlusive dressing is used to keep the cream in place.

POSITIONING

Proper positioning of the patient is essential. The procedure should be performed on a firm surface; if the procedure is to be performed at the bedside, the patient should be positioned at the edge of the bed and not in the middle. The patient is asked to lie on his or her side, facing away from the examiner, and to "roll up into a ball." The neck is gently ante-flexed and the thighs pulled up toward the abdomen; the shoulders and pelvis should be vertically aligned without forward or backward tilt (Fig. e46-1). The spinal cord terminates at approximately the L1 vertebral level in 94% of individuals. In the remaining 6%, the conus extends to the L2-L3 interspace. LP is therefore performed at or below the L3-L4 interspace. A useful anatomic guide is a line drawn between the posterior superior iliac crests, which corresponds closely to the level of the L3-L4 interspace. The interspace is chosen following gentle palpation to identify the spinous processes at each lumbar level.

Figure e46-1

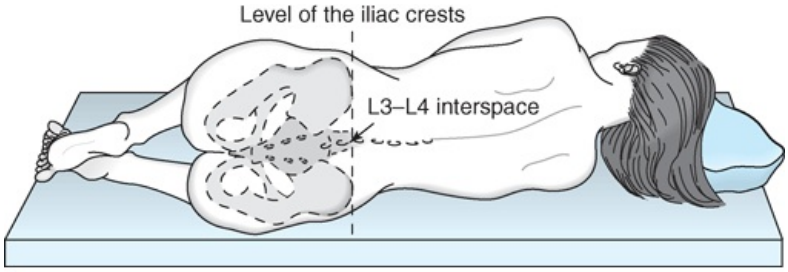

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Proper positioning of a patient in the lateral decubitus position. Note that the shoulders and hips are in a vertical plane; the torso is perpendicular to the bed. [From RP Simon et al (eds): Clinical Neurology, 7th ed. New York, McGraw-Hill, 2009.]

An alternative to the lateral recumbent position is the seated position. The patient sits at the side of the bed, with feet supported on a chair. The patient is instructed to curl forward, trying to touch the nose to the umbilicus. It is important that the patient not simply lean forward onto a bedside tabletop, as this is not an optimal position for opening up the spinous processes. LP is sometimes more easily performed in obese patients if they are sitting. A disadvantage of the seated position is that measurement of opening pressure may not be accurate. In situations in which LP is difficult using palpable spinal landmarks, bedside ultrasound to guide needle placement may be employed.

TECHNIQUE

Once the desired target for needle insertion has been identified, the examiner should put on sterile gloves. After cleansing the skin with povidone-iodine or similar disinfectant, the area is draped with a sterile cloth; the needle insertion site is blotted dry using a sterile gauze pad. Proper local disinfection reduces the risk of introducing skin bacteria into the SAS or other sites. Local anesthetic, typically 1% lidocaine, 3–5 mL total, is injected into the subcutaneous tissue; in nonemergency situations a topical anesthetic cream can be applied (see above). When time permits, painassociated with the injection of lidocaine can be minimized byslow, serial injections, each one progressively deeper than the last, over a period of ~5 min. Approximately 0.5–1 mL of lidocaine is injected at a time; the needle is not usually withdrawn between injections. A pause of ~15 s between injections helps to minimize the pain of the subsequent injection. The goal is to inject each mini-bolus of anesthetic into an area of skin that has become numb from the preceding injection. Approximately 5–10 mini-boluses are injected, using a total of ~5 mL of lidocaine.

If possible, the LP should be delayed for 10–15 min following the completion of the injection of anesthetic; this significantly decreases and can even eliminate pain from the procedure. Even a delay of 5 min will help to reduce pain.

The LP needle (typically 20- to 22-gauge) is inserted in the midline, midway between two spinous processes, and slowly advanced. The bevel of the needle should be maintained in a horizontal position, parallel to the direction of the dural fibers and with the flat portion of the bevel pointed upward; this minimizes injury to the fibers as the dura is penetrated. When lumbar puncture is performed in patients who are sitting, the bevel should be maintained in the vertical position. In most adults, the needle is advanced 4–5 cm (1–2 in.) before the SAS is reached; the examiner usually recognizes entry as a sudden release of resistance, a "pop." If no fluid appears despite apparently correct needle placement, then the needle may be rotated 90°–180°. If there is still no fluid, the stylet is reinserted and the needle is advanced slightly. Some examiners halt needle advancement periodically to remove the stylet and check for flow of cerebrospinal fluid (CSF). If the needle cannot be advanced because it hits bone, if the patient experiences sharp radiating pain down one leg, or if no fluid appears ("dry tap"), the needle is partially withdrawn and reinserted at a different angle. If on the second attempt the needle still hits bone (indicating lack of success in introducing it between the spinous processes), then the needle should be completely withdrawn and the patient should be repositioned. The second attempt is sometimes more successful if the patient straightens the spine completely prior to repositioning. The needle can then be reinserted at the same level or at an adjacent one.

Once the SAS is reached, a manometer is attached to the needle and the opening pressure measured. The examiner should look for normal oscillations in CSF pressure associated with pulse and respirations. The upper limit of normal opening pressure with the patient supine is 180 mmH^2O in adults but may be as high as 200–250 mmH^2O in obese adults.

CSF is allowed to drip into collection tubes; it should not be withdrawn with a syringe. Depending on the clinical indication, fluid is then obtained for studies including: (1) cell count with differential; (2) protein and glucose concentrations; (3) culture (bacterial, fungal, mycobacterial, viral); (4) smears (e.g., Gram's and acid-fast stained smears); (5) antigen tests (e.g., latex agglutination); (6) polymerase

chain reaction (PCR) amplification of DNA or RNA of microorganisms (e.g., herpes simplex virus, enteroviruses); (7) antibody levels against microorganisms; (8) immunoelectrophoresis for determination of γ-globulin level and oligoclonal banding; and (9) cytology. Although 15 mL of CSF is sufficient to obtain all of the listed studies, the yield of fungal and mycobacterial cultures and cytology increases when larger volumes are sampled. In general 20–30 mL may be safely removed from adults.

A bloody tap due to penetration of a meningeal vessel (a "traumatic tap") may result in confusion with subarachnoid hemorrhage (SAH). In these situations a specimen of CSF should be centrifuged immediately after it is obtained; clear supernatant following CSF centrifugation supports the diagnosis of a bloody tap, whereas xanthochromic supernatant suggests SAH. In general, bloody CSF due to the penetration of a meningeal vessel clears gradually in successive tubes, whereas blood due to SAH does not. In addition to SAH, xanthochromic CSF may also be present in patients with liver disease and when the CSF protein concentration is markedly elevated [>1.5–2 g/L (150–200 mg/dL)].

Prior to removing the LP needle, the stylet is reinserted to avoid the possibility of entrapment of a nerve root in the dura as the needle is being withdrawn; entrapment could result in a dural CSF leak, causing headache. Some practitioners question the safety of this maneuver, with its potential risk of causing a needle-stick injury to the examiner. Injury is unlikely, however, given the flexibility of the small-diameter stylet, which tends to bend, rather than penetrate, on contact. Following LP, the patient is customarily positioned in a comfortable, recumbent position for 1 h before rising, although this may be unnecessary as it does not appear to affect the development of headache (see below).

POST-LP HEADACHE

The principal complication of LP is headache, occurring in 10–30% of patients. Younger age and female gender are associated with an increased risk of post-LP headache. Headache usually begins within 48 h but may be delayed for up to 12 days. Head pain is dramatically positional; it begins when the patient sits or stands upright; there is relief upon reclining or with abdominal compression. The longer the patient is upright, the longer the latency before head pain subsides. The pain is usually a dull ache but may be throbbing; its location is occipitofrontal. Nausea and stiff neck often accompany headache, and occasionally, patients report blurred vision, photophobia, tinnitus, and vertigo. In more than three-quarters of patients, symptoms completely resolve within a week, but in a minority they can persist for weeks or even months.

Post-LP headache is caused by a drop in CSF pressure related to persistent leakage of CSF at the site where the needle entered the subarachnoid space. Loss of CSF volume decreases the brain's supportive cushion, so that when a patient is upright there is probably dilation and tension placed on the brain's anchoring structures, the pain-sensitive dural sinuses, resulting in pain. Although intracranial hypotension is the usual explanation for severe LP headache, the syndrome can occur in patients with normal CSF pressure.

Because post-LP headache usually resolves without specific treatment, care is largely supportive with oral analgesics [acetaminophen, nonsteroidal anti-inflammatory drugs, opioids (Chap. 11)] and antiemetics. Patients may obtain relief by lying in a comfortable (especially a recumbent or head-down Trendelenburg) position. For some patients, beverages with caffeine can provide temporary pain relief.

For patients with persistent pain, treatment with IV caffeine (500 mg in 500 mL saline administered over 2 h) may be effective; atrial fibrillation is a rare side effect. Alternatively, an epidural blood patch accomplished by injection of 15 mL of autologous whole blood is usually effective. This procedure is most often performed by a pain specialist or anesthesiologist. The blood patch has an immediate effect, making it unlikely that sealing off a dural hole with blood clot is its sole mechanism of action. The acute benefit may be due to compression of the CSF space by the clot, increasing CSF pressure. Some clinicians reserve epidural blood patch for patients who do not respond to caffeine, while others prefer to use blood patch as initial management for unremitting post-LP symptoms.

Strategies to decrease the incidence of post-LP headache are listed in Table e46–1. Use of a smaller caliber needle is associated with a lower risk: In one study, the risk of headache following use of a 24- to 27-gauge standard (Quincke) needle was 5–12%, compared to 20–40% when a 20- or 22-gauge needle was used. The smallest gauge needles usually require the use of an introducer needle and are associated with a slower CSF flow rate. Use of an "atraumatic" (Sprotte, "pencil point," or "noncutting") needle also reduces the incidence of moderate to severe headache compared with standard LP (Quincke, or "traumatic") needles (Fig. e46-2). However, because atraumatic needles are more difficult to use, more attempts may be required to perform the LP, particularly in overweight patients. It may also be necessary to use an introducer with the atraumatic needle, which does not have the customary cutting, beveled tip. There is a low risk of needle damage, e.g., breakage, with the Sprotte atraumatic needle. Another strategy to decrease the incidence of headache is to replace the stylet before removing the LP needle.

Table e46–1 Reducing the Incidence of Post-LP Headache

Effective Strategies

Use of small-diameter needle (22-gauge or smaller)
Use of atraumatic needle (Sprotte and others)
Replacement of stylet prior to removal of needle
Insertion of needle with bevel oriented in a cephalad to caudad direction (when using standard needle)
Ineffective Strategies
Bed rest (up to 4 h) following LP
Supplemental fluids
Minimizing the volume of spinal fluid removed
Immediate mobilization following LP

Figure e46-2

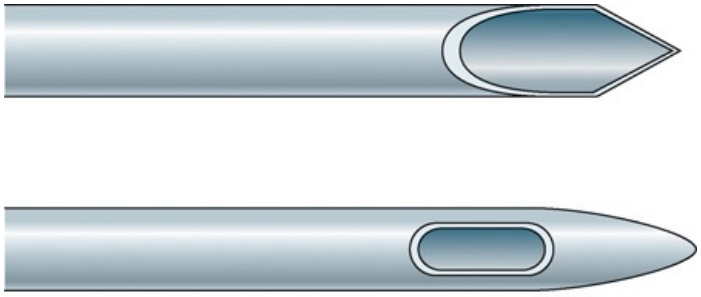

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Comparison of the standard ("traumatic" or Quincke) LP needle with the "atraumatic" (Sprotte). The "atraumatic" needle has its opening on the top surface of the needle, a design intended to reduce the chance of cutting dural fibers that, by protruding through the dura, could be responsible for subsequent CSF fluid leak and post-LP headache. (From SR Thomas et al: BMJ 321:986, 2000.)

Patients are often advised to remain in a recumbent position for up to an hour following lumbar puncture. However, studies comparing mobilization immediately following LP with bed rest for periods up to 4 h show no significant differences in the incidence of headache, suggesting that the customary practice of remaining in a recumbent position post-LP may be unnecessary.

NORMAL VALUES

(See Table e46–2) In uninfected CSF, the normal white blood cell count is fewer than five mononuclear cells (lymphocytes and monocytes) per μL. Polymorphonuclear leukocytes (PMNs) are not found in normal unconcentrated CSF; however, rare PMNs can be found in centrifuged or concentrated CSF specimens such as those utilized for cytologic examination. Red blood cells (RBCs) are not normally present in CSF; if RBCs are present from a traumatic tap, their number decreases as additional CSF is collected. CSF glucose concentrations <2.2 mmol/L (<40 mg/dL) are abnormal.

Table e46–2 Cerebrospinal Fluid[a]

Constituent	SI Units	Conventional Units
Glucose	2.22–3.89 mmol/L	40–70 mg/dL
Lactate	1–2 mmol/L	10–20 mg/dL
Total protein		
Lumbar	0.15–0.5 g/L	15–50 mg/dL
Cisternal	0.15–0.25 g/L	15–25 mg/dL
Ventricular	0.06–0.15 g/L	6–15 mg/dL
Albumin	0.066–0.442 g/L	6.6–44.2 mg/dL
IgG	0.009–0.057 g/L	0.9–5.7 mg/dL
IgG index[b]	0.29–0.59	
Oligoclonal bands (OGB)	<2 bands not present in matched serum sample	

Ammonia	15–47 μmol/L	25–80 μg/dL
CSF pressure		50–180 mmH_2O
CSF volume (adult) Red blood cells	~150 mL 0	 0
Leukocytes		
Total	0–5 mononuclear cells per mm^3	
Differential		
Lymphocytes	60–70%	
Monocytes	30–50%	
Neutrophils	None	

[a]Because cerebrospinal fluid concentrations are equilibrium values, measurements of the same parameters in blood plasma obtained at the same time are recommended. However, there is a time lag in attainment of equilibrium, and cerebrospinal levels of plasma constituents that can fluctuate rapidly (such as plasma glucose) may not achieve stable values until after a significant lag phase.

[b]IgG index = CSF IgG (mg/dL) x serum albumin (g/dL)/Serum IgG (g/dL) x CSF albumin (mg/dL).

FURTHER READINGS

Armon C, Evans RW: Addendum to assessment: Prevention of post-lumbar puncture headaches: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology 65:510, 2005[PMID: 16116106] [Full Text]

Arendt K et al: Atraumatic lumbar puncture needles: After all these years, are we still missing the point? Neurologist 15:17, 2009[PMID: 19131853] [Full Text]

Bezov D et al: Post-dural puncture headache: Part I diagnosis, epidemiology, etiology, and pathophysiology. Headache 50:1144, 2010[PMID: 20533959] [Full Text]

Ellenby MS et al: Lumbar puncture (video). N Engl J Med 355:e12, 2006

Ferre RM, Sweeney TW: Emergency physicians can easily obtain ultrasound images of anatomical landmarks relevant to lumbar puncture. Am J Emerg Med 25:291, 2007[PMID: 17349903] [Full Text]

Lavi R et al: Standard vs atraumatic Whitacre needle for diagnostic lumbar puncture: A randomized trial. Neurology 67:1492, 2006[PMID: 17060584] [Full Text]

Richman JM et al: Bevel direction and postdural puncture headache: A meta-analysis. Neurologist 12:224, 2006[PMID: 16832241] [Full Text]

Straus SE et al: How do I perform a lumbar puncture and analyze the results to diagnose bacterial meningitis? JAMA 296:2012, 2006[PMID: 17062865] [Full Text]

Strupp M et al: Incidence of post-lumbar puncture syndrome reduced by reinserting the stylet: A randomized prospective study of 600 patients. J Neurol 245:589, 1998[PMID: 9758296] [Full Text]

Van Kooten F et al: Epidural blood patch in post dural puncture headache: A randomized, observer-blind, controlled clinical trial. J Neurol Neurosurg Psychiatry 79:553, 2008

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SEIZURES AND EPILEPSY: INTRODUCTION

A *seizure* (from the Latin *sacire*, "to take possession of") is a paroxysmal event due to abnormal excessive or synchronous neuronal activity in the brain. Depending on the distribution of discharges, this abnormal brain activity can have various manifestations, ranging from dramatic convulsive activity to experiential phenomena not readily discernible by an observer. Although a variety of factors influence the incidence and prevalence of seizures, ~5–10% of the population will have at least one seizure, with the highest incidence occurring in early childhood and late adulthood.

The meaning of the term *seizure* needs to be carefully distinguished from that of epilepsy. *Epilepsy* describes a condition in which a person has *recurrent* seizures due to a chronic, underlying process. This definition implies that a person with a single seizure, or recurrent seizures due to correctable or avoidable circumstances, does not necessarily have epilepsy. Epilepsy refers to a clinical phenomenon rather than a single disease entity, since there are many forms and causes of epilepsy. However, among the many causes of epilepsy there are various *epilepsy syndromes* in which the clinical and pathologic characteristics are distinctive and suggest a specific underlying etiology.

Using the definition of epilepsy as two or more unprovoked seizures, the incidence of epilepsy is ~0.3–0.5% in different populations throughout the world, and the prevalence of epilepsy has been estimated at 5–10 persons per 1000.

CLASSIFICATION OF SEIZURES

Determining the type of seizure that has occurred is essential for focusing the diagnostic approach on particular etiologies, selecting the appropriate therapy, and providing potentially vital information regarding prognosis. The International League against Epilepsy (ILAE) Commission on Classification and Terminology, 2005–2009 has provided an updated approach to classification of seizures **(Table 369-1)**. This system is based on the clinical features of seizures and associated electroencephalographic findings. Other potentially distinctive features such as etiology or cellular substrate are not considered in this classification system, although this will undoubtedly change in the future as more is learned about the pathophysiologic mechanisms that underlie specific seizure types.

Table 369-1 Classification of Seizures

1. **Focal seizures**
   (Can be further described as having motor, sensory, autonomic, cognitive, or other features)
2. **Generalized seizures**
   a. Absence
      Typical
      Atypical
   b. Tonic clonic
   c. Clonic
   d. Tonic
   e. Atonic
   f. Myoclonic
3. **May be focal, generalized, or unclear**
   Epileptic spasms

A fundamental principle is that seizures may be either focal or generalized. *Focal seizures* originate within networks limited to one cerebral hemisphere (note that the term *partial seizures* is no longer used). *Generalized seizures* arise within and rapidly engage networks distributed across both cerebral hemispheres. Focal seizures are usually associated with structural abnormalities of the brain. In contrast, generalized seizures may result from cellular, biochemical, or structural abnormalities that have a more widespread distribution. There are clear exceptions in both cases, however.

Focal Seizures

Focal seizures arise from a neuronal network either discretely localized within one cerebral hemisphere or more broadly distributed but still within the hemisphere. With the new classification system, the subcategories of "simple focal seizures" and "complex focal seizures" have been eliminated. Instead, depending on the presence of cognitive impairment, they can be described as focal seizures with or without dyscognitive features. Focal seizures can also evolve into generalized seizures. In the past this was referred to as *focal seizures with secondary generalization,* but the new system relies on specific descriptions of the type of generalized seizures that evolve from the focal seizure.

The routine interictal (i.e., between seizures) electroencephalogram (EEG) in patients with focal seizures is often normal or may show brief discharges termed *epileptiform spikes*, or *sharp waves*. Since focal seizures can arise from the medial temporal lobe or inferior frontal lobe (i.e., regions distant from the scalp), the EEG recorded during the seizure may be nonlocalizing. However, the seizure focus is often detected using sphenoidal or surgically placed intracranial electrodes.

FOCAL SEIZURES WITHOUT DYSCOGNITIVE FEATURES

Focal seizures can cause motor, sensory, autonomic, or psychic symptoms without impairment of cognition. For example, a patient having a focal motor seizure arising from the right primary motor cortex near the area controlling hand movement will note the onset of involuntary movements of the contralateral, left hand. These movements are typically clonic (i.e., repetitive, flexion/extension movements) at a frequency of ~2–3 Hz; pure tonic posturing may be seen as well. Since the cortical region controlling hand movement is immediately adjacent to the region for facial expression, the seizure may also cause abnormal movements of the face synchronous with the movements of the hand. The EEG recorded with scalp electrodes during the seizure (i.e., an ictal EEG) may show abnormal discharges in a very limited region over the appropriate area of cerebral cortex if the seizure focus involves the cerebral convexity. Seizure activity occurring within deeper brain structures is often not recorded by the standard EEG, however, and may require intracranial electrodes for its detection.

Three additional features of focal motor seizures are worth noting. First, in some patients the abnormal motor movements may begin in a very restricted region such as the fingers and gradually progress (over seconds to minutes) to include a larger portion of the extremity. This phenomenon, described by Hughlings Jackson and known as a "Jacksonian march," represents the spread of seizure activity over

a progressively larger region of motor cortex. Second, patients may experience a localized paresis (Todd's paralysis) for minutes to many hours in the involved region following the seizure. Third, in rare instances the seizure may continue for hours or days. This condition, termed *epilepsia partialis continua*, is often refractory to medical therapy.

Focal seizures may also manifest as changes in somatic sensation (e.g., paresthesias), vision (flashing lights or formed hallucinations), equilibrium (sensation of falling or vertigo), or autonomic function (flushing, sweating, piloerection). Focal seizures arising from the temporal or frontal cortex may also cause alterations in hearing, olfaction, or higher cortical function (psychic symptoms). This includes the sensation of unusual, intense odors (e.g., burning rubber or kerosene) or sounds (crude or highly complex sounds), or an epigastric sensation that rises from the stomach or chest to the head. Some patients describe odd, internal feelings such as fear, a sense of impending change, detachment, depersonalization, déjá vu, or illusions that objects are growing smaller (micropsia) or larger (macropsia). These subjective, "internal" events that are not directly observable by someone else are referred to as *auras*.

FOCAL SEIZURES WITH DYSCOGNITIVE FEATURES

Focal seizures may also be accompanied by a transient impairment of the patient's ability to maintain normal contact with the environment. The patient is unable to respond appropriately to visual or verbal commands during the seizure and has impaired recollection or awareness of the ictal phase. The seizures frequently begin with an aura (i.e., a focal seizure without cognitive disturbance) that is stereotypic for the patient. The start of the ictal phase is often a sudden behavioral arrest or motionless stare, which marks the onset of the period of impaired awareness. The behavioral arrest is usually accompanied by *automatisms*, which are involuntary, automatic behaviors that have a wide range of manifestations. Automatisms may consist of very basic behaviors such as chewing, lip smacking, swallowing, or "picking" movements of the hands, or more elaborate behaviors such as a display of emotion or running. The patient is typically confused following the seizure, and the transition to full recovery of consciousness may range from seconds up to an hour. Examination immediately following the seizure may show an anterograde amnesia or, in cases involving the dominant hemisphere, a postictalaphasia.

The range of potential clinical behaviors linked to focal seizures is so broad that extreme caution is advised before concluding that stereotypic episodes of bizarre or atypical behavior are not due to seizure activity. In such cases additional, detailed EEG studies may be helpful.

Evolution of Focal Seizures to Generalized Seizures

Focal seizures can spread to involve both cerebral hemispheres and produce a generalized seizure, usually of the tonic-clonic variety (discussed below). This evolution is observed frequently following focal seizures arising from a focus in the frontal lobe, but may also be associated with focal seizures occurring elsewhere in the brain. A focal seizure that evolves into a generalized seizure is often difficult to distinguish from a primary generalized-onset tonic-clonic seizure, since bystanders tend to emphasize the more dramatic, generalized convulsive phase of the seizure and overlook the more subtle, focal symptoms present at onset. In some cases, the focal onset of the seizure becomes apparent only when a careful history identifies a preceding aura. Often, however, the focal onset is not clinically evident and may be established only through careful EEG analysis. Nonetheless, distinguishing between these two entities is extremely important, as there may be substantial differences in the evaluation and treatment of epilepsies associated with focal versus generalized seizures.

Generalized Seizures

Generalized seizures are thought to arise at some point in the brain but immediately and rapidly engage neuronal networks in both cerebral hemispheres. Several types of generalized seizures have features that place them in distinctive categories and facilitate clinical diagnosis.

TYPICAL ABSENCE SEIZURES

Typical absence seizures are characterized by sudden, brief lapses of consciousness without loss of postural control. The seizure typically lasts for only seconds, consciousness returns as suddenly as it was lost, and there is no postictal confusion. Although the brief loss of consciousness may be clinically inapparent or the sole manifestation of the seizure discharge, absence seizures are usually accompanied by subtle, bilateral motor signs such as rapid blinking of the eyelids, chewing movements, or small-amplitude, clonic movements of the hands.

Typical absence seizures are associated with a group of genetically determined epilepsies with onset usually in childhood (ages 4–8 years) or early adolescence and are the main seizure type in 15–20% of children with epilepsy. The seizures can occur hundreds of times per day, but the child may be unaware of or unable to convey their existence. Since the clinical signs of the seizures are subtle, especially to parents who may not have had previous experience with seizures, it is not surprising that the first clue to absence epilepsy is often unexplained "daydreaming" and a decline in school performance recognized by a teacher.

The electrophysiologic hallmark of typical absence seizures is a generalized, symmetric, 3-Hz spike-and-wave discharge that begins and ends suddenly, superimposed on a normal EEG background. Periods of spike-and-wave discharges lasting more than a few seconds usually correlate with clinical signs, but the EEG often shows many more brief bursts of abnormal cortical activity than were suspected clinically. Hyperventilation tends to provoke these electrographic discharges and even the seizures themselves and is routinely used when recording the EEG.

ATYPICAL ABSENCE SEIZURES

Atypical absence seizures have features that deviate both clinically and electrophysiologically from typical absence seizures. For example, the lapse of consciousness is usually of longer duration and less abrupt in onset and cessation, and the seizure is accompanied by more obvious motor signs that may include focal or lateralizing features. The EEG shows a generalized, slow spike-and-wave pattern with a frequency of ≤2.5 per second, as well as other abnormal activity. Atypical absence seizures are usually associated with diffuse or multifocal structural abnormalities of the brain and therefore may accompany other signs of neurologic dysfunction such as mental retardation. Furthermore, the seizures are less responsive to anticonvulsants compared to typical absence seizures.

GENERALIZED, TONIC-CLONIC SEIZURES

Generalized-onset tonic-clonic seizures are the main seizure type in ~10% of all persons with epilepsy. They are also the most common seizure type resulting from metabolic derangements and are therefore frequently encountered in many different clinical settings. The seizure usually begins abruptly without warning, although some patients describe vague premonitory symptoms in the hours leading up to the seizure. This prodrome is distinct from the stereotypic auras associated with focal seizures that generalize. The initial phase of the seizure is usually tonic contraction of muscles throughout the body, accounting for a number of the classic features of the event. Tonic contraction of the muscles of expiration and the larynx at the onset will produce a loud moan or "ictal cry." Respirations are impaired, secretions pool in the oropharynx, and cyanosis develops. Contraction of the jaw muscles may cause biting of the tongue. A marked enhancement of sympathetic tone leads to increases in heart rate, blood pressure, and pupillary size. After 10–20 seconds, the tonic phase of the seizure typically evolves into the clonic phase, produced by the superimposition of periods of muscle relaxation on the tonic muscle contraction. The periods of relaxation progressively increase until the end of the ictal phase, which usually lasts no more than 1 minute. The postictal phase is characterized by unresponsiveness, muscular flaccidity, and excessive salivation that can cause stridorous breathing and partial airway obstruction. Bladder or bowel incontinence may occur at this point. Patients gradually regain consciousness over minutes to hours, and during this transition there is typically a period of postictal confusion. Patients subsequently complain of headache, fatigue, and muscle ache that can last for many hours. The duration of impaired consciousness in the postictal phase can be extremely long (i.e., many hours) in patients with prolonged seizures or underlying central nervous system (CNS) diseases such as alcoholic cerebral atrophy.

The EEG during the tonic phase of the seizure shows a progressive increase in generalized low-voltage fast activity, followed by generalized high-amplitude, polyspike discharges. In the clonic phase, the high-amplitude activity is typically interrupted by slow waves

to create a spike-and-wave pattern. The postictal EEG shows diffuse slowing that gradually recovers as the patient awakens.

There are a number of variants of the generalized tonic-clonic seizure, including pure tonic and pure clonic seizures. Brief tonic seizures lasting only a few seconds are especially noteworthy since they are usually associated with specific epileptic syndromes having mixed seizure phenotypes, such as the Lennox-Gastaut syndrome (discussed below).

ATONIC SEIZURES

Atonic seizures are characterized by sudden loss of postural muscle tone lasting 1–2 seconds. Consciousness is briefly impaired, but there is usually no postictal confusion. A very brief seizure may cause only a quick head drop or nodding movement, while a longer seizure will cause the patient to collapse. This can be extremely dangerous, since there is a substantial risk of direct head injury with the fall. The EEG shows brief, generalized spike-and-wave discharges followed immediately by diffuse slow waves that correlate with the loss of muscle tone. Similar to pure tonic seizures, atonic seizures are usually seen in association with known epilepsy syndromes.

MYOCLONIC SEIZURES

Myoclonus is a sudden and brief muscle contraction that may involve one part of the body or the entire body. A normal, common physiologic form of myoclonus is the sudden jerking movement observed while falling asleep. Pathologic myoclonus is most commonly seen in association with metabolic disorders, degenerative CNS diseases, or anoxic brain injury (Chap. 275). Although the distinction from other forms of myoclonus is imprecise, myoclonic seizures are considered to be true epileptic events since they are caused by cortical (versus subcortical or spinal) dysfunction. The EEG may show bilaterally synchronous spike-and-wave discharges synchronized with the myoclonus, although these can be obscured by movement artifact. Myoclonic seizures usually coexist with other forms of generalized seizures but are the predominant feature of juvenile myoclonic epilepsy (discussed below).

Currently Unclassifiable Seizures

Not all seizure types can be designated as focal or generalized, and they should therefore be labeled as "unclassifiable" until additional evidence allows a valid classification. *Epileptic spasms* are such an example. These are characterized by a briefly sustained flexion or extension of predominantly proximal muscles, including truncal muscles. The EEG in these patients usually shows hypsarrhythmias, which consist of diffuse, giant slow waves with a chaotic background of irregular, multifocal spikes and sharp waves. During the clinical spasm, there is a marked suppression of the EEG background (the "electrodecremental response"). The electromyogram (EMG) also reveals a characteristic rhomboid pattern that may help distinguish spasms from brief tonic and myoclonic seizures. Epileptic spasms occur predominantly in infants and likely result from differences in neuronal function and connectivity in the immature versus mature CNS.

EPILEPSY SYNDROMES

Epilepsy syndromes are disorders in which epilepsy is a predominant feature, and there is sufficient evidence (e.g., through clinical, EEG, radiologic, or genetic observations) to suggest a common underlying mechanism. Three important epilepsy syndromes are listed below; additional examples with a known genetic basis are shown in **Table 369-2**.

Table 369-2 Examples of Genes Associated with Epilepsy Syndromes[a]

Gene (Locus)	Function of Gene	Clinical Syndrome	Comments
CHRNA4 (20q13.2)	Nicotinic acetylcholine receptor subunit; mutations cause alterations in Ca2+ flux through the receptor; this may reduce amount of GABA release in presynaptic terminals	Autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE); childhood onset; brief, nighttime seizures with prominent motor movements; often misdiagnosed as primary sleep disorder	Rare; first identified in a large Australian family; other families found to have mutations in CHRNA2 or CHRNB2, and some families appear to have mutations at other loci
KCNQ2 (20q13.3)	Voltage-gated potassium channel subunits; mutation in pore regions may cause a 20–40% reduction of potassium currents, which will lead to impaired repolarization	Benign familial neonatalconvulsions (BFNC); autosomal dominant inheritance; onset in 1st week of life in infants who are otherwise normal; remission usually within weeks to months; long-term epilepsy in 10–15%	Rare; other families found to have mutations in KCNQ3; sequence and functional homology to KCNQ1, mutations of which cause long QT syndrome and a cardiac-auditory syndrome
SCN1B (19q12.1)	β-subunit of a voltage-gated sodium channel; mutation disrupts disulfide bridge that is crucial for structure of extracellular domain; mutated β-subunit leads to slower sodium channel inactivation	Generalized epilepsy with febrile seizures plus (GEFS+); autosomal dominant inheritance; presents with febrile seizures at median 1 year, which may persist >6 years, then variable seizure types not associated with fever	Incidence uncertain; GEFS+ identified in other families with mutations in other sodium channel subunits (SCN1A and SCN2A) and GABAA receptor subunit (GABRG2 and GABRA1); significant phenotypic heterogeneity within same family, including members with febrile seizures only
LGI1 (10q24)	Leucine-rich glioma-inactivated 1 gene; previous evidence for role in glial tumor progression; protein homology suggests a possible role in nervous system development	Autosomal dominant partial epilepsy with auditory features (ADPEAF); a form of idiopathic lateral temporal lobe epilepsy with auditory symptoms or aphasia as a major simple partial seizure manifestation; age of onset usually between 10 and 25 years	Mutations found in approximately 50% of families containing two or more subjects with idiopathic localization-related epilepsy with ictal auditory symptoms, suggesting that at least one other gene may underlie this syndrome. LGI1 is the only gene identified so far in temporal lobe epilepsy
CSTB (21q22.3)	Cystatin B, a noncaspase cysteine protease inhibitor; normal protein may block neuronal apoptosis by inhibiting caspases directly or indirectly (via cathepsins), or controlling proteolysis	Progressive myoclonus epilepsy (PME) (Unverricht-Lundborg disease); autosomal recessive inheritance; age of onset between 6 and 15 years, myoclonic seizures, ataxia, and progressive cognitive decline; brain shows neuronal degeneration	Overall rare, but relatively common in Finland and Western Mediterranean (>1 in 20,000); precise role of cystatin B in human disease unknown, although mice with null mutations of cystatin B have similar syndrome
EPM2A (6q24)	Laforin, a protein tyrosine phosphatase (PTP); involved in glycogen metabolism and may have antiapoptotic activity	Progressive myoclonus epilepsy (Lafora's disease); autosomal recessive inheritance; onset age 6–19 years, death within 10 years; brain degeneration associated with polyglucosan intracellular inclusion bodies in numerous organs	Most common PME in Southern Europe, Middle East, Northern Africa, and Indian subcontinent; genetic heterogeneity; unknown whether seizure phenotype due to degeneration or direct effects of abnormal laforin expression
Doublecortin (Xq21-24)	Doublecortin, expressed primarily in frontal lobes; directly regulates microtubule polymerization and bundling	Classic lissencephaly associated with severe mental retardation and seizures in males; subcortical band heterotopia with more subtle findings in females (presumably due to random X-inactivation); X-linked dominant	Relatively rare but of uncertain incidence, recent increased ascertainment due to improved imaging techniques; relationship between migration defect and seizure phenotype unknown

[a]The first four syndromes listed in the table (ADNFLE, BFNC, GEFS+, and ADPEAF) are examples of idiopathic epilepsies associated with identified gene mutations. The last three syndromes are examples of the numerous Mendelian disorders in which seizures are one part of the phenotype.

Abbreviations: GABA, γ-aminobutyric acid; PME, progressive myoclonus epilepsy.

Juvenile Myoclonic Epilepsy

Juvenile myoclonic epilepsy (JME) is a generalized seizure disorder of unknown cause that appears in early adolescence and is usually characterized by bilateral myoclonic jerks that may be single or repetitive. The myoclonic seizures are most frequent in the morning after awakening and can be provoked by sleep deprivation. Consciousness is preserved unless the myoclonus is especially severe. Many patients also experience generalized tonic-clonic seizures, and up to one-third have absence seizures. Although complete remission is relatively uncommon, the seizures respond well to appropriate anticonvulsant medication. There is often a family history of epilepsy, and genetic linkage studies suggest a polygenic cause.

Lennox-Gastaut Syndrome

Lennox-Gastaut syndrome occurs in children and is defined by the following triad: (1) multiple seizure types (usually including generalized tonic-clonic, atonic, and atypical absence seizures); (2) an EEG showing slow (<3 Hz) spike-and-wave discharges and a variety of other abnormalities; and (3) impaired cognitive function in most but not all cases. Lennox-Gastaut syndrome is associated with CNS disease or dysfunction from a variety of causes, including developmental abnormalities, perinatal hypoxia/ischemia, trauma, infection, and other acquired lesions. The multifactorial nature of this syndrome suggests that it is a nonspecific response of the brain to diffuse neural injury. Unfortunately, many patients have a poor prognosis due to the underlying CNS disease and the physical and psychosocial consequences of severe, poorly controlled epilepsy.

Mesial Temporal Lobe Epilepsy Syndrome

Mesial temporal lobe epilepsy (MTLE) is the most common syndrome associated with focal seizures with dyscognitive features and is an example of an epilepsy syndrome with distinctive clinical, electroencephalographic, and pathologic features **(Table 369-3)**. High-resolution MRI can detect the characteristic hippocampal sclerosis that appears to be essential in the pathophysiology of MTLE for many patients **(Fig. 369-1)**. Recognition of this syndrome is especially important because it tends to be refractory to treatment with anticonvulsants but responds extremely well to surgical intervention. Advances in the understanding of basic mechanisms of epilepsy have come through studies of experimental models of MTLE, discussed below.

Table 369-3 Characteristics of the Mesial Temporal Lobe Epilepsy Syndrome

History	
History of febrile seizures	Rare generalized seizures
Family history of epilepsy	Seizures may remit and reappear
Early onset	Seizures often intractable
Clinical observations	
Aura common	Postictal disorientation
Behavioral arrest/stare	Memory loss
Complex automatisms	Dysphasia (with focus in dominanthemisphere)
Unilateral posturing	
Laboratory studies	
Unilateral or bilateral anterior temporal spikes on EEG	
Hypometabolism on interictal PET	
Hypoperfusion on interictal SPECT	
Material-specific memory deficits on intracranial amobarbital (Wada) test	
MRI findings	
Small hippocampus with increased signal on T2-weighted sequences	
Small temporal lobe	
Enlarged temporal horn	
Pathologic findings	
Highly selective loss of specific cell populations within hippocampus in most cases	

Abbreviations: EEG, electroencephalogram; PET, positron emission tomography; SPECT, single photon emission computed tomography.

Figure 369-1

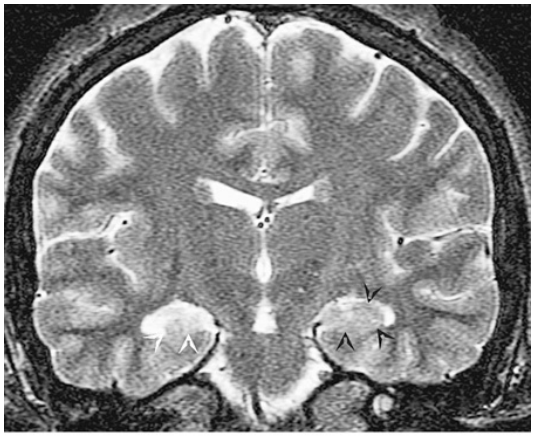

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mesial temporal lobe epilepsy. The EEG suggested a right temporal lobe focus. Coronal high-resolution T2-weighted fast spin echo magnetic resonance image obtained through the body of the hippocampus demonstrates abnormal high-signal intensity in the right hippocampus (*white arrows*; compare with the normal hippocampus on the left, *black arrows*) consistent with mesial temporal sclerosis.

THE CAUSES OF SEIZURES AND EPILEPSY

Seizures are a result of a shift in the normal balance of excitation and inhibition within the CNS. Given the numerous properties that control neuronal excitability, it is not surprising that there are many different ways to perturb this normal balance, and therefore many different causes of both seizures and epilepsy. Three clinical observations emphasize how a variety of factors determine why certain conditions may cause seizures or epilepsy in a given patient.

1. *The normal brain is capable of having a seizure under the appropriate circumstances, and there are differences between individuals in the susceptibility or threshold for seizures*. For example, seizures may be induced by high fevers in children who are otherwise normal and who never develop other neurologic problems, including epilepsy. However, febrile seizures occur only in a relatively small proportion of children. This implies there are various underlying *endogenous factors* that influence the threshold for having a seizure. Some of these factors are clearly genetic, as it has been shown that a family history of epilepsy will influence the likelihood of seizures occurring in otherwise normal individuals. Normal development also plays an important role, since the brain appears to have different seizure thresholds at different maturational stages.
2. *There are a variety of conditions that have an extremely high likelihood of resulting in a chronic seizure disorder* . One of the best examples of this is severe, penetrating head trauma, which is associated with up to a 45% risk of subsequent epilepsy. The high propensity for severe traumatic brain injury to lead to epilepsy suggests that the injury results in a long-lasting pathologic change in the CNS that transforms a presumably normal neural network into one that is abnormally hyperexcitable. This process is known as *epileptogenesis*, and the specific changes that result in a lowered seizure threshold can be considered *epileptogenic factors*. Other processes associated with epileptogenesis include stroke, infections, and abnormalities of CNS development. Likewise, the genetic abnormalities associated with epilepsy likely involve processes that trigger the appearance of specific sets of epileptogenic factors.
3. *Seizures are episodic*. Patients with epilepsy have seizures intermittently and, depending on the underlying cause, many patients are completely normal for months or even years between seizures. This implies there are important provocative or *precipitating factors* that induce seizures in patients with epilepsy. Similarly, precipitating factors are responsible for causing the single seizure in someone without epilepsy. Precipitants include those due to intrinsic physiologic processes such as psychological or physical stress, sleep deprivation, or hormonal changes associated with the menstrual cycle. They also include exogenous factors such as exposure to toxic substances and certain medications.

These observations emphasize the concept that the many causes of seizures and epilepsy result from a dynamic interplay between endogenous factors, epileptogenic factors, and precipitating factors. The potential role of each needs to be carefully considered when determining the appropriate management of a patient with seizures. For example, the identification of predisposing factors (e.g., family history of epilepsy) in a patient with febrile seizures may increase the necessity for closer follow-up and a more aggressive diagnostic evaluation. Finding an epileptogenic lesion may help in the estimation of seizure recurrence and duration of therapy. Finally, removal or modification of a precipitating factor may be an effective and safer method for preventing further seizures than the prophylactic use of anticonvulsant drugs.

Causes According to Age

In practice, it is useful to consider the etiologies of seizures based on the age of the patient, as age is one of the most important factors determining both the incidence and the likely causes of seizures or epilepsy **(Table 369-4)**. During the *neonatal period and early infancy*, potential causes include hypoxic-ischemic encephalopathy, trauma, CNS infection, congenital CNS abnormalities, and metabolic disorders. Babies born to mothers using neurotoxic drugs such as cocaine, heroin, or ethanol are susceptible to drug-withdrawal seizures in the first few days after delivery. Hypoglycemia and hypocalcemia, which can occur as secondary complications of perinatal injury, are also causes of seizures early after delivery. Seizures due to inborn errors of metabolism usually present once regular feeding begins, typically 2–3 days after birth. Pyridoxine (vitamin B6) deficiency, an important cause of neonatal seizures, can be effectively treated with pyridoxine replacement. The idiopathic or inherited forms of benign neonatal convulsions are also seen during this time period.

Table 369-4 Causes of Seizures

Neonates(<1 month)	Perinatal hypoxia and ischemia
	Intracranial hemorrhage and trauma
	Acute CNS infection
	Metabolic disturbances (hypoglycemia, hypocalcemia, hypomagnesemia, pyridoxine deficiency)
	Drug withdrawal
	Developmental disorders
	Genetic disorders
Infants and children (>1 month and <12 years)	Febrile seizures
	Genetic disorders (metabolic, degenerative, primary epilepsy syndromes)
	CNS infection
	Developmental disorders
	Trauma
	Idiopathic
Adolescents (12–18 years)	Trauma
	Genetic disorders
	Infection
	Brain tumor
	Illicit drug use
	Idiopathic
Young adults (18–35 years)	Trauma
	Alcohol withdrawal
	Illicit drug use
	Brain tumor
	Idiopathic
Older adults (>35 years)	Cerebrovascular disease
	Brain tumor
	Alcohol withdrawal

	Metabolic disorders (uremia, hepatic failure, electrolyte abnormalities, hypoglycemia, hyperglycemia)
	Alzheimer's disease and other degenerative CNS diseases
	Idiopathic

Abbreviation: CNS, central nervous system.

The most common seizures arising in *late infancy and early childhood* are febrile seizures, which are seizures associated with fevers but without evidence of CNS infection or other defined causes. The overall prevalence is 3–5% and even higher in some parts of the world such as Asia. Patients often have a family history of febrile seizures or epilepsy. Febrile seizures usually occur between 3 months and 5 years of age and have a peak incidence between 18 and 24 months. The typical scenario is a child who has a generalized, tonic-clonic seizure during a febrile illness in the setting of a common childhood infection such as otitis media, respiratory infection, or gastroenteritis. The seizure is likely to occur during the rising phase of the temperature curve (i.e., during the first day) rather than well into the course of the illness. A *simple* febrile seizure is a single, isolated event, brief, and symmetric in appearance. *Complex* febrile seizures are characterized by repeated seizure activity, duration >15 minutes, or by focal features. Approximately one-third of patients with febrile seizures will have a recurrence, but <10% have three or more episodes. Recurrences are much more likely when the febrile seizure occurs in the first year of life. Simple febrile seizures are not associated with an increase in the risk of developing epilepsy, while complex febrile seizures have a risk of 2–5%; other risk factors include the presence of preexisting neurologic deficits and a family history of nonfebrile seizures.

Childhood marks the age at which many of the well-defined epilepsy syndromes present. Some children who are otherwise normal develop idiopathic, generalized tonic-clonic seizures without other features that fit into specific syndromes. Temporal lobe epilepsy usually presents in childhood and may be related to mesial temporal lobe sclerosis (as part of the MTLE syndrome) or other focal abnormalities such as cortical dysgenesis. Other types of focal seizures, including those that evolve into generalized seizures, may be the relatively late manifestation of a developmental disorder, an acquired lesion such as head trauma, CNS infection (especially viral encephalitis), or very rarely a CNS tumor.

The period of *adolescence and early adulthood* is one of transition during which the idiopathic or genetically based epilepsy syndromes, including JME and juvenile absence epilepsy, become less common, while epilepsies secondary to acquired CNS lesions begin to predominate. Seizures that begin in patients in this age range may be associated with head trauma, CNS infections (including parasitic infections such as cysticercosis), brain tumors, congenital CNS abnormalities, illicit drug use, or alcohol withdrawal.

Head trauma is a common cause of epilepsy in adolescents and adults. The head injury can be caused by a variety of mechanisms, and the likelihood of developing epilepsy is strongly correlated with the severity of the injury. A patient with a penetrating head wound, depressed skull fracture, intracranial hemorrhage, or prolonged posttraumatic coma or amnesia has a 40–50% risk of developing epilepsy, while a patient with a closed head injury and cerebral contusion has a 5–25% risk. Recurrent seizures usually develop within 1 year after head trauma, although intervals of ≥10 years are well known. In controlled studies, mild head injury, defined as a concussion with amnesia or loss of consciousness of <30 minutes, was found to be associated with only a slightly increased likelihood of epilepsy. Nonetheless, most epileptologists know of patients who have focal seizures within hours or days of a mild head injury and subsequently develop chronic seizures of the same type; such cases may represent rare examples of chronic epilepsy resulting from mild head injury.

The causes of seizures in *older adults* include cerebrovascular disease, trauma (including subdural hematoma), CNS tumors, and degenerative diseases. Cerebrovascular disease may account for ~50% of new cases of epilepsy in patients older than age 65. Acute seizures (i.e., occurring at the time of the stroke) are seen more often with embolic rather than hemorrhagic or thrombotic stroke. Chronic seizures typically appear months to years after the initial event and are associated with all forms of stroke.

Metabolic disturbances such as electrolyte imbalance, hypo- or hyperglycemia, renal failure, and hepatic failure may cause seizures at any age. Similarly, endocrine disorders, hematologic disorders, vasculitides, and many other systemic diseases may cause seizures over a broad age range. A wide variety of medications and abused substances are known to precipitate seizures as well **(Table 369-5)**.

Table 369-5 Drugs and Other Substances that Can Cause Seizures

Alkylating agents (e.g., busulfan, chlorambucil)
Antimalarials (chloroquine, mefloquine)
Antimicrobials/antivirals
β-lactam and related compounds
Quinolones
Acyclovir
Isoniazid
Ganciclovir
Anesthetics and analgesics
Meperidine
Tramadol
Local anesthetics
Dietary supplements
Ephedra (ma huang)
Gingko
Immunomodulatory drugs
Cyclosporine
OKT3 (monoclonal antibodies to T cells)
Tacrolimus
Interferons
Psychotropics
Antidepressants
Antipsychotics
Lithium
Radiographic contrast agents
Theophylline
Sedative-hypnotic drug withdrawal
Alcohol

Barbiturates (short-acting)
Benzodiazepines (short-acting)
Drugs of abuse
Amphetamine
Cocaine
Phencyclidine
Methylphenidate
Flumazenil[a]

[a]In benzodiazepine-dependent patients.

BASIC MECHANISMS

Mechanisms of Seizure Initiation and Propagation

Focal seizure activity can begin in a very discrete region of cortex and then spread to neighboring regions, i.e., there is a *seizure initiation* phase and a *seizure propagation* phase. The initiation phase is characterized by two concurrent events in an aggregate of neurons: (1) high-frequency bursts of action potentials and (2) hypersynchronization. The bursting activity is caused by a relatively long-lasting depolarization of the neuronal membrane due to influx of extracellular calcium (Ca^{2+}), which leads to the opening of voltage-dependent sodium (Na^{+}) channels, influx of Na^{+}, and generation of repetitive action potentials. This is followed by a hyperpolarizing afterpotential mediated by γ-aminobutyric acid (GABA) receptors or potassium (K^{+}) channels, depending on the cell type. The synchronized bursts from a sufficient number of neurons result in a so-called spike discharge on the EEG.

Normally, the spread of bursting activity is prevented by intact hyperpolarization and a region of "surround" inhibition created by inhibitory neurons. With sufficient activation there is a recruitment of surrounding neurons via a number of synaptic and nonsynaptic mechanisms, including: (1) an increase in extracellular K^{+}, which blunts hyperpolarization and depolarizes neighboring neurons; (2) accumulation of Ca^{2+} in presynaptic terminals, leading to enhanced neurotransmitter release; and (3) depolarization-induced activation of the *N*-methyl-D-aspartate (NMDA) subtype of the excitatory amino acid receptor, which causes additional Ca^{2+} influx and neuronal activation; and (4) ephaptic interactions related to changes in tissue osmolarity and cell swelling. The recruitment of a sufficient number of neurons leads to the propagation of seizure activity into contiguous areas via local cortical connections, and to more distant areas via long commissural pathways such as the corpus callosum.

Many factors control neuronal excitability, and thus there are many potential mechanisms for altering a neuron's propensity to have bursting activity. Mechanisms *intrinsic* to the neuron include changes in the conductance of ion channels, response characteristics of membrane receptors, cytoplasmic buffering, second-messenger systems, and protein expression as determined by gene transcription, translation, and posttranslational modification. Mechanisms *extrinsic* to the neuron include changes in the amount or type of neurotransmitters present at the synapse, modulation of receptors by extracellular ions and other molecules, and temporal and spatial properties of synaptic and nonsynaptic input. Nonneural cells such as astrocytes and oligodendrocytes, have an important role in many of these mechanisms as well.

Certain recognized causes of seizures are explained by these mechanisms. For example, accidental ingestion of domoic acid, which is an analogue of glutamate (the principal excitatory neurotransmitter in the brain), causes profound seizures via direct activation of excitatory amino acid receptors throughout the CNS. Penicillin, which can lower the seizure threshold in humans and is a potent convulsant in experimental models, reduces inhibition by antagonizing the effects of GABA at its receptor. The basic mechanisms of other precipitating factors of seizures such as sleep deprivation, fever, alcohol withdrawal, hypoxia, and infection, are not as well understood but presumably involve analogous perturbations in neuronal excitability. Similarly, the endogenous factors that determine an individual's seizure threshold may relate to these properties as well.

Knowledge of the mechanisms responsible for initiation and propagation of most generalized seizures (including tonic-clonic, myoclonic, and atonic types) remains rudimentary and reflects the limited understanding of the connectivity of the brain at a systems level. Much more is understood about the origin of generalized spike-and-wave discharges in absence seizures. These appear to be related to oscillatory rhythms normally generated during sleep by circuits connecting the thalamus and cortex. This oscillatory behavior involves an interaction between $GABA_B$ receptors, T-type Ca^{2+} channels, and K^{+} channels located within the thalamus. Pharmacologic studies indicate that modulation of these receptors and channels can induce absence seizures, and there is good evidence that the genetic forms of absence epilepsy may be associated with mutations of components of this system.

Mechanisms of Epileptogenesis

Epileptogenesis refers to the transformation of a normal neuronal network into one that is chronically hyperexcitable. There is often a delay of months to years between an initial CNS injury such as trauma, stroke, or infection and the first seizure. The injury appears to initiate a process that gradually lowers the seizure threshold in the affected region until a spontaneous seizure occurs. In many genetic and idiopathic forms of epilepsy, epileptogenesis is presumably determined by developmentally regulated events.

Pathologic studies of the hippocampus from patients with temporal lobe epilepsy have led to the suggestion that some forms of epileptogenesis are related to *structural changes in neuronal networks*. For example, many patients with MTLE have a highly selective loss of neurons that may contribute to inhibition of the main excitatory neurons within the dentate gyrus. There is also evidence that, in response to the loss of neurons, there is reorganization or "sprouting" of surviving neurons in a way that affects the excitability of the network. Some of these changes can be seen in experimental models of prolonged electrical seizures or traumatic brain injury. Thus, an initial injury such as head injury may lead to a very focal, confined region of structural change that causes local hyperexcitability. The local hyperexcitability leads to further structural changes that evolve over time until the focal lesion produces clinically evident seizures. Similar models have provided strong evidence for long-term alterations in *intrinsic, biochemical properties of cells* within the network such as chronic changes in glutamate or GABA receptor function. Recent work has suggested that induction of inflammatory cascades may be a critical factor in these processes as well.

GENETIC CAUSES OF EPILEPSY

The most important recent progress in epilepsy research has been the identification of genetic mutations associated with a variety of epilepsy syndromes (Table 369-2). Although all of the mutations identified to date cause rare forms of epilepsy, their discovery has led to extremely important conceptual advances. For example, it appears that many of the inherited, idiopathic epilepsies (i.e., the relatively "pure" forms of epilepsy in which seizures are the phenotypic abnormality and brain structure and function are otherwise normal) are due to mutations affecting ion channel function. These syndromes are therefore part of the larger group of channelopathies causing paroxysmal disorders such as cardiac arrhythmias, episodicataxia, periodic weakness, and familial hemiplegic migraine. In contrast, gene mutations observed in symptomatic epilepsies (i.e., disorders in which other neurologic abnormalities such as cognitive impairment, coexist with seizures) are proving to be associated with pathways influencing CNS development or neuronal homeostasis. A current challenge is to identify the multiple susceptibility genes that underlie the more common forms of idiopathic epilepsies. Recent studies suggest that ion channel mutations and chromosomal microdeletions may be the cause of epilepsy in a subset of these patients.

Mechanisms of Action of Antiepileptic Drugs

Antiepileptic drugs appear to act primarily by blocking the initiation or spread of seizures. This occurs through a variety of mechanisms that modify the activity of ion channels or neurotransmitters, and in most cases the drugs have pleiotropic effects. The mechanisms include inhibition of Na^+-dependent action potentials in a frequency-dependent manner (e.g., phenytoin, carbamazepine, lamotrigine, topiramate, zonisamide, lacosamide, rufinamide), inhibition of voltage-gated Ca^{2+} channels (phenytoin, gabapentin, pregabalin), attenuation of glutamate activity (lamotrigine, topiramate, felbamate), potentiation of GABA receptor function (benzodiazepines and barbiturates), increase in the availability of GABA (valproic acid, gabapentin, tiagabine), and modulation of release of synaptic vesicles (levetiracetam). The two most effective drugs for absence seizures, ethosuximide and valproic acid, probably act by inhibiting T-type Ca^{2+} channels in thalamic neurons.

In contrast to the relatively large number of antiepileptic drugs that can attenuate seizure activity, there are currently no drugs known to prevent the formation of a seizure focus following CNS injury. The eventual development of such "antiepileptogenic" drugs will provide an important means of preventing the emergence of epilepsy following injuries such as head trauma, stroke, and CNS infection.

Treatment: Seizure

When a patient presents shortly after a seizure, the first priorities are attention to vital signs, respiratory and cardiovascular support, and treatment of seizures if they resume (see "Treatment: Seizures and Epilepsy"). Life-threatening conditions such as CNS infection, metabolic derangement, or drug toxicity must be recognized and managed appropriately.

When the patient is not acutely ill, the evaluation will initially focus on whether there is a history of earlier seizures **(Fig. 369-2)**. If this is the first seizure, then the emphasis will be to: (1) establish whether the reported episode was a seizure rather than another paroxysmal event, (2) determine the cause of the seizure by identifying risk factors and precipitating events, and (3) decide whether anticonvulsant therapy is required in addition to treatment for any underlying illness.

Figure 369-2

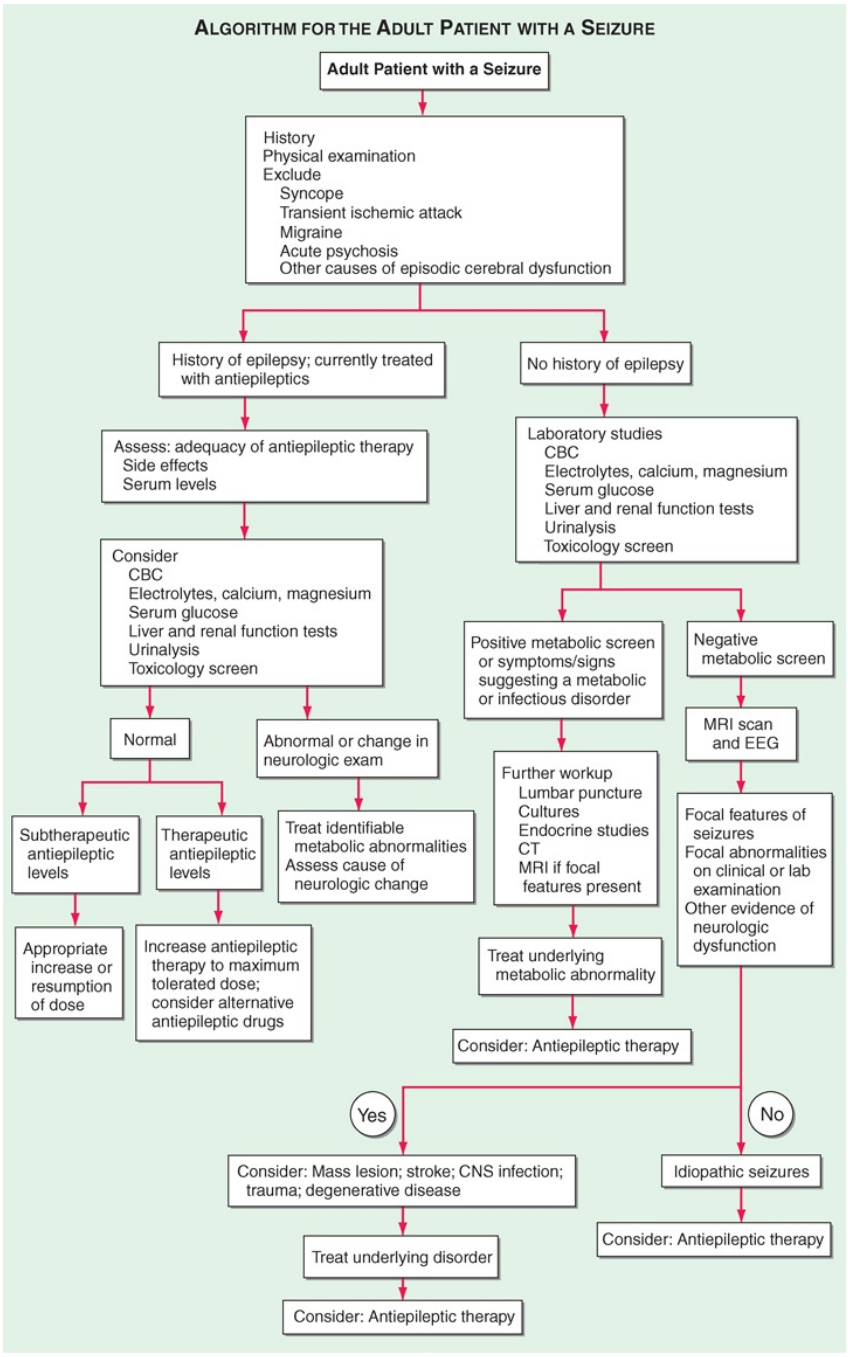

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Evaluation of the adult patient with a seizure. CBC, complete blood count; CNS, central nervous system; CT, computed tomography; EEG, electroencephalogram; MRI, magnetic resonance imaging.

In the patient with prior seizures or a known history of epilepsy, the evaluation is directed toward (1) identification of the underlying cause and precipitating factors, and (2) determination of the adequacy of the patient's current therapy.

History and Examination

The first goal is to determine whether the event was truly a seizure. An in-depth history is essential, for *in many cases the diagnosis of a seizure is based solely on clinical grounds—the examination and laboratory studies are often normal*. Questions should focus on the symptoms before, during, and after the episode in order to differentiate a seizure from other paroxysmal events (see "Differential Diagnosis of Seizures" below). Seizures frequently occur out-of-hospital, and the patient may be unaware of the ictal and immediate postictal phases; thus, witnesses to the event should be interviewed carefully.

The history should also focus on risk factors and predisposing events. Clues for a predisposition to seizures include a history of febrile seizures, earlier auras or brief seizures not recognized as such, and a family history of seizures. Epileptogenic factors such as prior head trauma, stroke, tumor, or infection of the nervous system should be identified. In children, a careful assessment of developmental milestones may provide evidence for underlying CNS disease. Precipitating factors such as sleep deprivation, systemic diseases, electrolyte or metabolic derangements, acute infection, drugs that lower the seizure threshold (Table 369-5), or alcohol or illicit drug use should also be identified.

The general physical examination includes a search for signs of infection or systemic illness. Careful examination of the skin may reveal signs of neurocutaneous disorders such as tuberous sclerosis or neurofibromatosis, or chronic liver or renal disease. A finding of organomegaly may indicate a metabolic storage disease, and limb asymmetry may provide a clue to brain injury early in development. Signs of head trauma and use of alcohol or illicit drugs should be sought. Auscultation of the heart and carotid arteries may identify an abnormality that predisposes to cerebrovascular disease.

All patients require a complete neurologic examination, with particular emphasis on eliciting signs of cerebral hemispheric disease (Chap.

367). Careful assessment of mental status (including memory, language function, and abstract thinking) may suggest lesions in the anterior frontal, parietal, or temporal lobes. Testing of visual fields will help screen for lesions in the optic pathways and occipital lobes. Screening tests of motor function such as pronator drift, deep tendon reflexes, gait, and coordination may suggest lesions in motor (frontal) cortex, and cortical sensory testing (e.g., double simultaneous stimulation) may detect lesions in the parietal cortex.

Laboratory Studies

Routine blood studies are indicated to identify the more common metabolic causes of seizures such as abnormalities in electrolytes, glucose, calcium, or magnesium, and hepatic or renal disease. A screen for toxins in blood and urine should also be obtained from all patients in appropriate risk groups, especially when no clear precipitating factor has been identified. A lumbar puncture is indicated if there is any suspicion of meningitis or encephalitis, and it is mandatory in all patients infected with HIV, even in the absence of symptoms or signs suggesting infection.

Electrophysiologic Studies

All patients who have a possible seizure disorder should be evaluated with an EEG as soon as possible.

Details about the EEG are covered in Chap. e45.

In the evaluation of a patient with suspected epilepsy, the presence of *electrographic seizure activity* during the clinically evident event (i.e., abnormal, repetitive, rhythmic activity having a discrete onset and termination) clearly establishes the diagnosis. The absence of electrographic seizure activity does not exclude a seizure disorder, however, because focal seizures may originate from a region of the cortex that cannot be detected by standard scalp electrodes. The EEG is always abnormal during generalized tonic-clonic seizures. Since seizures are typically infrequent and unpredictable, it is often not possible to obtain the EEG during a clinical event. Continuous monitoring for prolonged periods in video-EEG telemetry units for hospitalized patients or the use of portable equipment to record the EEG continuously on cassettes for ≥24 hours in ambulatory patients has made it easier to capture the electrophysiologic accompaniments of clinical events. In particular, video-EEG telemetry is now a routine approach for the accurate diagnosis of epilepsy in patients with poorly characterized events or seizures that are difficult to control.

The EEG may also be helpful in the interictal period by showing certain abnormalities that are highly supportive of the diagnosis of epilepsy. Such *epileptiform activity* consists of bursts of abnormal discharges containing spikes or sharp waves. The presence of epileptiform activity is not specific for epilepsy, but it has a much greater prevalence in patients with epilepsy than in normal individuals. However, even in an individual who is known to have epilepsy, the initial routine interictal EEG may be normal up to 60% of the time. Thus, the EEG cannot establish the diagnosis of epilepsy in many cases.

The EEG is also used for classifying seizure disorders and aiding in the selection of anticonvulsant medications. For example, episodic generalized spike-wave activity is usually seen in patients with typical absence epilepsy and may be seen with other generalized epilepsy syndromes. Focal interictal epileptiform discharges would support the diagnosis of a focal seizure disorder such as temporal lobe epilepsy or frontal lobe seizures, depending on the location of the discharges.

The routine scalp-recorded EEG may also be used to assess the prognosis of seizure disorders; in general, a normal EEG implies a better prognosis, whereas an abnormal background or profuse epileptiform activity suggests a poor outlook. Unfortunately, the EEG has not proved to be useful in predicting which patients with predisposing conditions such as head injury or brain tumor, will go on to develop epilepsy, because in such circumstances epileptiform activity is commonly encountered regardless of whether seizures occur.

Magnetoencephalography (MEG) provides another way of looking noninvasively at cortical activity. Instead of measuring electrical activity of the brain, it measures the small magnetic fields that are generated by this activity. Epileptiform activity seen on the MEG can be analyzed, and its source in the brain can be estimated using a variety of mathematical techniques. These source estimates can then be plotted on an anatomic image of the brain such as an MRI (discussed below), to generate a magnetic source image (MSI). MSI can be useful to localize potential seizure foci.

Brain Imaging

Almost all patients with new-onset seizures should have a brain imaging study to determine whether there is an underlying structural abnormality that is responsible. The only potential exception to this rule is children who have an unambiguous history and examination suggestive of a benign, generalized seizure disorder such as absence epilepsy. MRI has been shown to be superior to CT for the detection of cerebral lesions associated with epilepsy. In some cases MRI will identify lesions such as tumors, vascular malformations, or other pathologies that need immediate therapy. The use of newer MRI methods such as 3-Tesla scanners, multichannel head coils, three-dimensional structural imaging at submillimeter resolution, and new pulse sequences including fluid-attenuated inversion recovery (FLAIR), has increased the sensitivity for detection of abnormalities of cortical architecture, including hippocampal atrophy associated with mesial temporal sclerosis, as well as abnormalities of cortical neuronal migration. In such cases the findings may not lead to immediate therapy, but they do provide an explanation for the patient's seizures and point to the need for chronic anticonvulsant therapy or possible surgical resection.

In the patient with a suspected CNS infection or mass lesion, CT scanning should be performed emergently when MRI is not immediately available. Otherwise, it is usually appropriate to obtain an MRI study within a few days of the initial evaluation. Functional imaging procedures such as positron emission tomography (PET) and single-photon emission computed tomography (SPECT) are also used to evaluate certain patients with medically refractory seizures (discussed below).

DIFFERENTIAL DIAGNOSIS OF SEIZURES

Disorders that may mimic seizures are listed in **Table 369-6**. In most cases seizures can be distinguished from other conditions by meticulous attention to the history and relevant laboratory studies. On occasion, additional studies such as video-EEG monitoring, sleep studies, tilt-table analysis, or cardiac electrophysiology, may be required to reach a correct diagnosis. Two of the more common nonepileptic syndromes in the differential diagnosis are detailed below.

Table 369-6 Differential Diagnosis of Seizures

Syncope
Vasovagal syncope
Cardiac arrhythmia
Valvular heart disease
Cardiac failure
Orthostatic hypotension
Psychological disorders
Psychogenic seizure
Hyperventilation
Panic attack
Metabolic disturbances
Alcoholic blackouts
Delirium tremens

Hypoglycemia
Hypoxia
Psychoactive drugs (e.g., hallucinogens)
Migraine
Confusional migraine
Basilar migraine
Transient ischemic attack (TIA)
Basilar artery TIA
Sleep disorders
Narcolepsy/cataplexy
Benign sleep myoclonus
Movement disorders
Tics
Nonepileptic myoclonus
Paroxysmal choreoathetosis
Special considerations in children
Breath-holding spells
Migraine with recurrent abdominalpain and cyclic vomiting
Benign paroxysmal vertigo
Apnea
Night terrors
Sleepwalking

Syncope

(See also Chap. 20) The diagnostic dilemma encountered most frequently is the distinction between a generalized seizure and syncope. Observations by the patient and bystanders that can help differentiate between the two are listed in **Table 369-7**. Characteristics of a seizure include the presence of an aura, cyanosis, unconsciousness, motor manifestations lasting >15 seconds, postictal disorientation, muscle soreness, and sleepiness. In contrast, a syncopal episode is more likely if the event was provoked by acute pain or anxiety or occurred immediately after arising from the lying or sitting position. Patients with syncope often describe a stereotyped transition from consciousness to unconsciousness that includes tiredness, sweating, nausea, and tunneling of vision, and they experience a relatively brief loss of consciousness. Headache or incontinence usually suggests a seizure but may on occasion also occur with syncope. A brief period (i.e., 1–10 seconds) of convulsive motor activity is frequently seen immediately at the onset of a syncopal episode, especially if the patient remains in an upright posture after fainting (e.g., in a dentist's chair) and therefore has a sustained decrease in cerebral perfusion. Rarely, a syncopal episode can induce a full tonic-clonic seizure. In such cases the evaluation must focus on both the cause of the syncopal event as well as the possibility that the patient has a propensity for recurrent seizures.

Table 369-7 Features that Distinguish Generalized Tonic-Clonic Seizure from Syncope

Features	Seizure	Syncope
Immediate precipitating factors	Usually none	Emotional stress, Valsalva, orthostatic hypotension, cardiac etiologies
Premonitory symptoms	None or aura (e.g., odd odor)	Tiredness, nausea, diaphoresis, tunneling of vision
Posture at onset	Variable	Usually erect
Transition to unconsciousness	Often immediate	Gradual over secondsa
Duration of unconsciousness	Minutes	Seconds
Duration of tonic or clonic movements	30–60 s	Never more than 15 s
Facial appearance during event	Cyanosis, frothing at mouth	Pallor
Disorientation and sleepiness after event	Many minutes to hours	<5 min
Aching of muscles after event	Often	Sometimes
Biting of tongue	Sometimes	Rarely
Incontinence	Sometimes	Sometimes
Headache	Sometimes	Rarely

[a]May be sudden with certain cardiac arrhythmias.

Psychogenic Seizures

Psychogenic seizures are nonepileptic behaviors that resemble seizures. They are often part of a conversion reaction precipitated by underlying psychological distress. Certain behaviors such as side-to-side turning of the head, asymmetric and large-amplitude shaking movements of the limbs, twitching of all four extremities without loss of consciousness, and pelvic thrusting are more commonly associated with psychogenic rather than epileptic seizures. Psychogenic seizures often last longer than epileptic seizures and may wax and wane over minutes to hours. However, the distinction is sometimes difficult on clinical grounds alone, and there are many examples of diagnostic errors made by experienced epileptologists. This is especially true for psychogenic seizures that resemble focal seizures with dyscognitive features, since the behavioral manifestations of focal seizures (especially of frontal lobe origin) can be extremely unusual, and in both cases the routine surface EEG may be normal. Video-EEG monitoring is very useful when historic features are nondiagnostic. Generalized tonic-clonic seizures always produce marked EEG abnormalities during and after the seizure. For suspected focal seizures of temporal lobe origin, the use of additional electrodes beyond the standard scalp locations (e.g., sphenoidal electrodes) may be required to localize a seizure focus. Measurement of serum prolactin levels may also help to distinguish between organic and psychogenic seizures, since most generalized seizures and some focal seizures are accompanied by rises in serum prolactin (during the immediate 30-minute postictal period), whereas psychogenic seizures are not. The diagnosis of psychogenic seizures does not exclude a concurrent diagnosis of epilepsy, since the two often coexist.

Treatment: Seizures and Epilepsy

Therapy for a patient with a seizure disorder is almost always multimodal and includes treatment of underlying conditions that cause or contribute to the seizures, avoidance of precipitating factors, suppression of recurrent seizures by prophylactic therapy with antiepileptic

medications or surgery, and addressing a variety of psychological and social issues. Treatment plans must be individualized, given the many different types and causes of seizures as well as the differences in efficacy and toxicity of antiepileptic medications for each patient. In almost all cases a neurologist with experience in the treatment of epilepsy should design and oversee implementation of the treatment strategy. Furthermore, patients with refractory epilepsy or those who require polypharmacy with antiepileptic drugs should remain under the regular care of a neurologist.

TREATMENT OF UNDERLYING CONDITIONS

If the sole cause of a seizure is a metabolic disturbance such as an abnormality of serum electrolytes or glucose, then treatment is aimed at reversing the metabolic problem and preventing its recurrence. Therapy with antiepileptic drugs is usually unnecessary unless the metabolic disorder cannot be corrected promptly and the patient is at risk of having further seizures. If the apparent cause of a seizure was a medication (e.g., theophylline) or illicit drug use (e.g., cocaine), then appropriate therapy is avoidance of the drug; there is usually no need for antiepileptic medications unless subsequent seizures occur in the absence of these precipitants.

Seizures caused by a structural CNS lesion such as a brain tumor, vascular malformation, or brain abscess may not recur after appropriate treatment of the underlying lesion. However, despite removal of the structural lesion, there is a risk that the seizure focus will remain in the surrounding tissue or develop de novo as a result of gliosis and other processes induced by surgery, radiation, or other therapies. Most patients are therefore maintained on an antiepileptic medication for at least 1 year, and an attempt is made to withdraw medications only if the patient has been completely seizure free. If seizures are refractory to medication, the patient may benefit from surgical removal of the epileptic brain region (see below).

AVOIDANCE OF PRECIPITATING FACTORS

Unfortunately, little is known about the specific factors that determine precisely when a seizure will occur in a patient with epilepsy. Some patients can identify particular situations that appear to lower their seizure threshold; these situations should be avoided. For example, a patient who has seizures in the setting of sleep deprivation should obviously be advised to maintain a normal sleep schedule. Many patients note an association between alcohol intake and seizures, and they should be encouraged to modify their drinking habits accordingly. There are also relatively rare cases of patients with seizures that are induced by highly specific stimuli such as a video game monitor, music, or an individual's voice ("reflex epilepsy"). If there is an association between stress and seizures, stress reduction techniques such as physical exercise, meditation, or counseling may be helpful.

ANTIEPILEPTIC DRUG THERAPY

Antiepileptic drug therapy is the mainstay of treatment for most patients with epilepsy. The overall goal is to completely prevent seizures without causing any untoward side effects, preferably with a single medication and a dosing schedule that is easy for the patient to follow. Seizure classification is an important element in designing the treatment plan, since some antiepileptic drugs have different activities against various seizure types. However, there is considerable overlap between many antiepileptic drugs such that the choice of therapy is often determined more by the patient's specific needs, especially his or her assessment of side effects.

When to Initiate Antiepileptic Drug Therapy

Antiepileptic drug therapy should be started in any patient with recurrent seizures of unknown etiology or a known cause that cannot be reversed. Whether to initiate therapy in a patient with a single seizure is controversial. Patients with a single seizure due to an identified lesion such as a CNS tumor, infection, or trauma, in which there is strong evidence that the lesion is epileptogenic, should be treated. The risk of seizure recurrence in a patient with an apparently unprovoked or idiopathic seizure is uncertain, with estimates ranging from 31 to 71% in the first 12 months after the initial seizure. This uncertainty arises from differences in the underlying seizure types and etiologies in various published epidemiologic studies. Generally accepted risk factors associated with recurrent seizures include the following: (1) an abnormal neurologic examination, (2) seizures presenting as status epilepticus, (3) postictal Todd's paralysis, (4) a strong family history of seizures, or (5) an abnormal EEG. Most patients with one or more of these risk factors should be treated. Issues such as employment or driving may influence the decision whether to start medications as well. For example, a patient with a single, idiopathic seizure whose job depends on driving may prefer taking antiepileptic drugs rather than risk a seizure recurrence and the potential loss of driving privileges.

Selection of Antiepileptic Drugs

Antiepileptic drugs available in the United States are shown in **Table 369-8**, and the main pharmacologic characteristics of commonly used drugs are listed in **Table 369-9**. Worldwide, older medications such as phenytoin, valproic acid, carbamazepine, phenobarbital, and ethosuximide are generally used as first-line therapy for most seizure disorders since, overall, they are as effective as recently marketed drugs and significantly less expensive. Most of the new drugs that have become available in the past decade are used as add-on or alternative therapy, although some are now being used as first-line monotherapy.

Table 369-8 Selection of Antiepileptic Drugs

Generalized-onset Tonic-Clonic	Focal	Typical Absence	Atypical Absence, Myoclonic, Atonic
First-Line			
Valproic acid Lamotrigine Topiramate	Lamotrigine Carbamazepine Oxcarbazepine Phenytoin Levetiracetam	Valproic acid Ethosuximide	Valproic acid Lamotrigine Topiramate
Alternatives			
Zonisamide[a] Phenytoin Carbamazepine Oxcarbazepine Phenobarbital Primidone Felbamate	Topiramate Zonisamide[a] Valproic acid Tiagabine[a] Gabapentin[a] Lacosamide[a] Phenobarbital Primidone Felbamate	Lamotrigine Clonazepam	Clonazepam Felbamate

[a]As adjunctive therapy.

Table 369-9 Dosage and Adverse Effects of Commonly Used Antiepileptic Drugs

				Adverse Effects	

Generic Name	Trade Name	Principal Uses	Typical Dose; Dose Interval	Half-Life	Therapeutic Range	Neurologic	Systemic	Drug Interactions
Phenytoin (diphenyl-hydantoin	Dilantin	Tonic-clonic Focal-onset	300–400 mg/d (3–6 mg/kg, adult; 4–8 mg/kg, child); qd-bid	24 h (wide variation, dose-dependent)	10–20 μg/mL	Dizziness Diplopia Ataxia Incoordination Confusion	Gum hyperplasia Lymphadenopathy Hirsutism Osteomalacia Facial coarsening Skin rash	Level increased by isoniazid, sulfonamides, fluoxetine Level decreased by enzyme-inducing drugs[a] Altered folate metabolism
Carbamazepine	Tegretol[b] Carbatrol	Tonic-clonic Focal-onset	600–1800 mg/d (15–35 mg/kg, child); bid-qid	10–17 h	6–12 μg/mL	Ataxia Dizziness Diplopia Vertigo	Aplastic anemia Leukopenia Gastrointestinal irritation Hepatotoxicity Hyponatremia	Level decreased by enzyme-inducing drugsa Level increased by erythromycin, propoxyphene, isoniazid, cimetidine, fluoxetine
Valproic acid	Depakene Depakote[b]	Tonic-clonic Absence Atypical absence Myoclonic Focal-onset Atonic	750–2000 mg/d (20–60 mg/kg); bid-qid	15 h	50–125 μg/mL	Ataxia Sedation Tremor	Hepatotoxicity Thrombocytopenia Gastrointestinal irritation Weight gain Transient alopecia Hyperammonemia	Level decreased by enzyme-inducing drugs[a]
Lamotrigine	Lamictal[b]	Focal-onset Tonic-clonic Atypical absence Myoclonic Lennox-Gastaut syndrome	150–500 mg/d; bid	25 h 14 h (with enzyme-inducers) 59 h (with valproic acid)	Not established	Dizziness Diplopia Sedation Ataxia Headache	Skin rash Stevens-Johnson syndrome	Level decreased by enzyme-inducing drugsa and oral contraceptives Level increased by valproic acid
Ethosuximide	Zarontin	Absence	750–1250 mg/d (20-40 mg/kg); qd-bid	60 h, adult 30 h, child	40–100 μg/mL	Ataxia Lethargy Headache	Gastrointestinal irritation Skin rash Bone marrow suppression	Level decreased by enzyme-inducing drugs[a] Level increased by valproic acid
Gabapentin	Neurontin	Focal-onset	900–2400 mg/d; tid-qid	5–9 h	Not established	Sedation Dizziness Ataxia Fatigue	Gastrointestinal irritation Weight gain Edema	No known significant interactions
Topiramate	Topamax	Focal-onset Tonic-clonic Lennox-Gastaut syndrome	200–400 mg/d; bid	20–30 h	Not established	Psychomotor slowing Sedation Speech or language problems Fatigue Paresthesias	Renal stones (avoid use with other carbonic anhydrase inhibitors) Glaucoma Weight loss Hypohidrosis	Level decreased by enzyme-inducing drugs[a]
Tiagabine	Gabitril	Focal-onset	32–56 mg/d; bid-qid	7–9 h	Not established	Confusion Sedation Depression Dizziness Speech or language problems Paresthesias Psychosis	Gastrointestinal irritation	Level decreased by enzyme-inducing drugs[a]
Phenobarbital	Luminol	Tonic-clonic Focal-onset	60–180 mg/d; qd	90 h	10–40 μg/mL	Sedation Ataxia Confusion Dizziness Decreased libido Depression	Skin rash	Level increased by valproic acid, phenytoin

Primidone	Mysoline	Tonic-clonic Focal-onset	750–1000 mg/d; bid-tid	Primidone, 8–15 h Phenobarbital, 90 h	Primidone, 4–12 μg/mL Phenobarbital, 10–40 μg/mL	Same as phenobarbital		Level increased byvalproic acid, phenytoin
Clonazepam	Klonopin	Absence Atypical absence Myoclonic	1–12 mg/d; qd-tid	24–48 h	10–70 ng/mL	Ataxia Sedation Lethargy	Anorexia	Level decreased by enzyme-inducing drugs[a]
Felbamate	Felbatol	Focal-onset Lennox-Gastaut syndrome Tonic-clonic	2400–3600 mg/d, tid-qid	16–22 h	Not established	Insomnia Dizziness Sedation Headache	Aplastic anemia Hepatic failure Weight loss Gastrointestinal irritation	Increases phenytoin, valproic acid, active carbamazepine metabolite
Levetiracetam	Keppra[b]	Focal-onset	1000–3000 mg/d; qd-bid	6–8 h	Not established	Sedation Fatigue Incoordination Mood changes	Anemia Leukopenia	No known significantinteractions
Zonisamide	Zonegran	Focal-onset Tonic-clonic	200–400 mg/d; qd-bid	50–68 h	Not established	Sedation Dizziness Confusion Headache Psychosis	Anorexia Renal stones Hypohidrosis	Level decreased by enzyme-inducing drugs[a]
Oxcarbazepine	Trileptal	Focal-onset Tonic-clonic	900–2400 mg/d (30–45 mg/kg, child); bid	10–17 h (for active metabolite)	Not established	Fatigue Ataxia Dizziness Diplopia Vertigo Headache	See carbamazepine	Level decreased by enzyme-inducing drugs[a] May increase phenytoin
Lacosamide	Vimpat	Focal-onset	200-400 mg/d; bid	13 h	Not established	Dizziness Ataxia Diplopia Vertigo	GI irritation Cardiac conduction (PR interval prolongation)	Level decreased by enzyme-inducing drugs[a]
Rufinamide	Banzel	Lennox-Gastaut syndrome	3200 mg/d (45 mg/kg, child); bid	6-10 h	Not established	Sedation Fatigue Dizziness Ataxia Headache Diplopia	GI irritation Leukopenia Cardiac conduction (QT interval prolongation)	Level decreased by enzyme-inducing drugs[a] Level increased by valproic acid May increase phenytoin

[a]Phenytoin, carbamazepine, phenobarbital. [b]Extended-release product available.

In addition to efficacy, factors influencing the choice of an initial medication include the convenience of dosing (e.g., once daily versus three or four times daily) and potential side effects. In this regard, a number of the newer drugs have the advantage of a relative lack of drug-drug interactions and easier dosing. Almost all of the commonly used antiepileptic drugs can cause similar, dose-related side effects such as sedation, ataxia, and diplopia. Long-term use of some agents in adults, especially the elderly, can lead to osteoporosis. Close follow-up is required to ensure these side effects are promptly recognized and reversed. Most of the older drugs and some of the newer ones can also cause idiosyncratic toxicity such as rash, bone marrow suppression, or hepatotoxicity. Although rare, these side effects should be considered during drug selection, and patients must be instructed about symptoms or signs that should signal the need to alert their health care provider. For some drugs, laboratory tests (e.g., complete blood count and liver function tests) are recommended prior to the institution of therapy (to establish baseline values) and during initial dosing and titration of the agent. Importantly, recent studies have shown that Asian individuals carrying the human leukocyte antigenallele, HLA-B*1502, are at particularly high risk of developing serious skin reactions from carbamazepine and phenytoin, so racial background and genotype are additional factors to consider in drug selection.

Antiepileptic Drug Selection for Focal Seizures

Carbamazepine (or a related drug, oxcarbazepine), lamotrigine and phenytoin are currently the drugs of choice approved for the initial treatment of focal seizures, including those that evolve into generalized seizures. Overall they have very similar efficacy, but differences in pharmacokinetics and toxicity are the main determinants for use in a given patient. For example, an advantage of carbamazepine (which is also available in an extended-release form) is that its metabolism follows first-order pharmacokinetics, and the relationship between drug dose, serum levels, and toxicity is linear. Carbamazepine can cause leukopenia, aplastic anemia, or hepatotoxicity and would therefore be contraindicated in patients with predispositions to these problems. Oxcarbazepine has the advantage of being metabolized in a way that avoids an intermediate metabolite associated with some of the side effects of carbamazepine. Oxcarbazepine also has fewer drug interactions than carbamazepine. Lamotrigine tends to be well tolerated in terms of side effects. However, patients need to be particularly vigilant about the possibility of a skin rash during the initiation of therapy. This can be extremely severe and lead to Stevens-Johnson syndrome if unrecognized and if the medication is not discontinued immediately. This risk can be reduced by slow introduction and titration. Lamotrigine must be started slowly when used as add-on therapy with valproic acid, since valproic acid inhibits lamotrigine metabolism, thereby substantially prolonging its half-life. Phenytoin has a relatively long half-life and offers the advantage of once or twice daily dosing compared to two or three times daily dosing for many of the other drugs. However, phenytoin shows properties of saturation kinetics, such that small increases in phenytoin doses above a standard maintenance dose can precipitate marked side effects. This is one of the main causes of acute phenytoin toxicity. Long-term use of phenytoin is associated with untoward cosmetic effects (e.g., hirsutism, coarsening of facial features, and gingival hypertrophy), and effects on bone metabolism, so it is often avoided in young patients who are likely to require the drug for many years. Topiramate can be used for both focal and generalized

seizures. Similar to some of the other antiepileptic drugs, topiramate can cause significant psychomotor slowing and other cognitive problems, and it should not be used in patients at risk for the development of glaucoma or renal stones.

Valproic acid is an effective alternative for some patients with focal seizures, especially when the seizures generalize. Gastrointestinal side effects are fewer when using the valproate semisodium formulation (Depakote). Valproic acid also rarely causes reversible bone marrow suppression and hepatotoxicity, and laboratory testing is required to monitor toxicity. This drug should generally be avoided in patients with preexisting bone marrow or liver disease. Irreversible, fatal hepatic failure appearing as an idiosyncratic rather than dose-related side effect is a relatively rare complication; its risk is highest in children <2 years old, especially those taking other antiepileptic drugs or with inborn errors of metabolism.

Levetiracetam, zonisamide, tiagabine, gabapentin, and lacosamide are additional drugs currently used for the treatment of focal seizures with or without evolution into generalized seizures. Phenobarbital and other barbiturate compounds were commonly used in the past as first-line therapy for many forms of epilepsy. However, the barbiturates frequently cause sedation in adults, hyperactivity in children, and other more subtle cognitive changes; thus, their use should be limited to situations in which no other suitable treatment alternatives exist.

Antiepileptic Drug Selection for Generalized Seizures

Valproic acid and lamotrigine are currently considered the best initial choice for the treatment of primary generalized, tonic-clonic seizures. Topiramate, zonisamide, phenytoin, and carbamazepine are suitable alternatives. Valproic acid is also particularly effective in absence, myoclonic, and atonic seizures and is therefore the drug of choice in patients with generalized epilepsy syndromes having mixed seizure types. Importantly, carbamazepine, oxcarbazepine, and phenytoin can worsen certain types of generalized seizures, including absence, myoclonic, tonic, and atonic seizures. Ethosuximide is a particularly effective drug for the treatment of uncomplicated absence seizures, but it is not useful for tonic-clonic or focal seizures. Ethosuximide rarely causes bone marrow suppression, so that periodic monitoring of blood cell counts is required. Lamotrigine appears to be particularly effective in epilepsy syndromes with mixed, generalized seizure types such as JME and Lennox-Gastaut syndrome. Topiramate, zonisamide, and felbamate may have similar broad efficacy.

Initiation and Monitoring of Therapy

Because the response to any antiepileptic drug is unpredictable, patients should be carefully educated about the approach to therapy. The goal is to prevent seizures and minimize the side effects of therapy; determination of the optimal dose is often a matter of trial and error. This process may take months or longer if the baseline seizure frequency is low. Most anticonvulsant drugs need to be introduced relatively slowly to minimize side effects, and patients should expect that minor side effects such as mild sedation, slight changes in cognition, or imbalance will typically resolve within a few days. Starting doses are usually the lowest value listed under the dosage column in Table 369-9. Subsequent increases should be made only after achieving a steady state with the previous dose (i.e., after an interval of five or more half-lives).

Monitoring of serum antiepileptic drug levels can be very useful for establishing the initial dosing schedule. However, the published therapeutic ranges of serum drug concentrations are only an approximate guide for determining the proper dose for a given patient. The key determinants are the clinical measures of seizure frequency and presence of side effects, not the laboratory values. Conventional assays of serum drug levels measure the total drug (i.e., both free and protein bound). However, it is the concentration of free drug that reflects extracellular levels in the brain and correlates best with efficacy. Thus, patients with decreased levels of serum proteins (e.g., decreased serum albumin due to impaired liver or renal function) may have an increased ratio of free to bound drug, yet the concentration of free drug may be adequate for seizure control. These patients may have a "subtherapeutic" drug level, but the dose should be changed only if seizures remain uncontrolled, not just to achieve a "therapeutic" level. It is also useful to monitor free drug levels in such patients. In practice, other than during the initiation or modification of therapy, monitoring of antiepileptic drug levels is most useful for documenting compliance.

If seizures continue despite gradual increases to the maximum tolerated dose and documented compliance, then it becomes necessary to switch to another antiepileptic drug. This is usually done by maintaining the patient on the first drug while a second drug is added. The dose of the second drug should be adjusted to decrease seizure frequency without causing toxicity. Once this is achieved, the first drug can be gradually withdrawn (usually over weeks unless there is significant toxicity). The dose of the second drug is then further optimized based on seizure response and side effects. Monotherapy should be the goal whenever possible.

When to Discontinue Therapy

Overall, about 70% of children and 60% of adults who have their seizures completely controlled with antiepileptic drugs can eventually discontinue therapy. The following patient profile yields the greatest chance of remaining seizure free after drug withdrawal: (1) complete medical control of seizures for 1–5 years; (2) single seizure type, either focal or generalized; (3) normal neurologic examination, including intelligence; and (4) normal EEG. The appropriate seizure-free interval is unknown and undoubtedly varies for different forms of epilepsy. However, it seems reasonable to attempt withdrawal of therapy after 2 years in a patient who meets all of the above criteria, is motivated to discontinue the medication, and clearly understands the potential risks and benefits. In most cases it is preferable to reduce the dose of the drug gradually over 2–3 months. Most recurrences occur in the first 3 months after discontinuing therapy, and patients should be advised to avoid potentially dangerous situations such as driving or swimming during this period.

Treatment of Refractory Epilepsy

Approximately one-third of patients with epilepsy do not respond to treatment with a single antiepileptic drug, and it becomes necessary to try a combination of drugs to control seizures. Patients who have focal epilepsy related to an underlying structural lesion or those with multiple seizure types and developmental delay are particularly likely to require multiple drugs. There are currently no clear guidelines for rational polypharmacy, although in theory a combination of drugs with different mechanisms of action may be most useful. In most cases the initial combination therapy combines first-line drugs (i.e., carbamazepine, oxcarbazepine, lamotrigine, valproic acid, and phenytoin). If these drugs are unsuccessful, then the addition of a newer drug such as levetiracetam, topiramate, and zonisamide is indicated. Patients with myoclonic seizures resistant to valproic acid may benefit from the addition of clonazepam, and those with absence seizures may respond to a combination of valproic acid and ethosuximide. The same principles concerning the monitoring of therapeutic response, toxicity, and serum levels for monotherapy apply to polypharmacy, and potential drug interactions need to be recognized. If there is no improvement, a third drug can be added while the first two are maintained. If there is a response, the less effective or less well tolerated of the first two drugs should be gradually withdrawn.

SURGICAL TREATMENT OF REFRACTORY EPILEPSY

Approximately 20–30% of patients with epilepsy continue to have seizures despite efforts to find an effective combination of antiepileptic drugs. For some, surgery can be extremely effective in substantially reducing seizure frequency and even providing complete seizure control. Understanding the potential value of surgery is especially important when a patient's seizures are not controlled with initial treatment, as such patients often fail to respond to subsequent medication trials. Rather than submitting the patient to years of unsuccessful medical therapy and the psychosocial trauma and increased mortality associated with ongoing seizures, the patient should have an efficient but relatively brief attempt at medical therapy and then be referred for surgical evaluation.

The most common surgical procedure for patients with temporal lobe epilepsy involves resection of the anteromedial temporal lobe (temporal lobectomy) or a more limited removal of the underlying hippocampus and amygdala (amygdalohippocampectomy). Focal seizures arising from extratemporal regions may be abolished by a focal neocortical resection with precise removal of an identified lesion (lesionectomy). When the cortical region cannot be removed, multiple subpial transection, which disrupts intracortical connections, is sometimes used to prevent seizure spread. Hemispherectomy or multilobar resection is useful for some patients with severe seizures due to hemispheric abnormalities such as hemimegalencephaly or other dysplastic abnormalities, and corpus callosotomy has been shown to be effective for disabling tonic or atonic seizures, usually when they are part of a mixed-seizure syndrome (e.g., Lennox-Gastaut syndrome).

Presurgical evaluation is designed to identify the functional and structural basis of the patient's seizure disorder. Inpatient video-EEG

monitoring is used to define the anatomic location of the seizure focus and to correlate the abnormal electrophysiologic activity with behavioral manifestations of the seizure. Routine scalp or scalp-sphenoidal recordings are usually sufficient for localization, and advances in neuroimaging have made the use of invasive electrophysiologic monitoring such as implanted depth electrodes or subdural electrodes less common. A high-resolution MRI scan is routinely used to identify structural lesions, and this is sometimes augmented with MEG. Functional imaging studies such as SPECT and PET are adjunctive tests that may help verify the localization of an apparent epileptogenic region. Once the presumed location of the seizure onset is identified, additional studies, including neuropsychological testing and the intracarotid amobarbital test (Wada test) may be used to assess language and memory localization and to determine the possible functional consequences of surgical removal of the epileptogenic region. In some cases, the exact extent of the resection to be undertaken is determined by performing cortical mapping at the time of the surgical procedure, allowing for a tailored resection. This involves electrocorticographic recordings made with electrodes on the surface of the brain to identify the extent of epileptiform disturbances. If the region to be resected is within or near brain regions suspected of having sensorimotor or language function, electrical cortical stimulation mapping is performed on the awake patient to determine the function of cortical regions in question in order to avoid resection of so-called eloquent cortex and thereby minimize postsurgical deficits.

Advances in presurgical evaluation and microsurgical techniques have led to a steady increase in the success of epilepsy surgery. Clinically significant complications of surgery are <5%, and the use of functional mapping procedures has markedly reduced the neurologic sequelae due to removal or sectioning of brain tissue. For example, about 70% of patients treated with temporal lobectomy will become seizure free, and another 15–25% will have at least a 90% reduction in seizure frequency. Marked improvement is also usually seen in patients treated with hemispherectomy for catastrophic seizure disorders due to large hemispheric abnormalities. Postoperatively, patients generally need to remain on antiepileptic drug therapy, but the marked reduction of seizures following resective surgery can have a very beneficial effect on quality of life.

Not all medically refractory patients are suitable candidates for resective surgery. For example, some patients have seizures arising from more than one site, making the risk of ongoing seizures or potential harm from the surgery unacceptably high. Vagus nerve stimulation (VNS) may be useful in some of these cases, although the benefit for most patients seems to be very limited (i.e., the efficacy of VNS appears to be no greater than trying another drug, which rarely works if a patient has proved to be refractory to the first two to three drugs). The precise mechanism of action of VNS is unknown, although experimental studies have shown that stimulation of vagal nuclei leads to widespread activation of cortical and subcortical pathways and an associated increased seizure threshold. Adverse effects of the surgery are rare, and stimulation-induced side effects, including transient hoarseness, cough, and dyspnea, are usually mild.

Although still in development, there are some additional therapies that will likely be of benefit to patients with medically refractory epilepsy. Preliminary studies suggest that stereotactic radiosurgery may be effective in certain focal seizure disorders. There has also been great interest in the development of implantable devices that can detect the onset of a seizure (in some instances, before the seizure becomes clinically apparent) and deliver either an electrical stimulation or drug directly to the seizure focus to abort the event.

Status Epilepticus

Status epilepticus refers to continuous seizures or repetitive, discrete seizures with impaired consciousness in the interictal period. Status epilepticus has numerous subtypes, including generalized convulsive status epilepticus (GCSE) (e.g., persistent, generalized electrographic seizures, coma, and tonic-clonic movements), and nonconvulsive status epilepticus (e.g., persistent absence seizures or focal seizures, confusion or partially impaired consciousness, and minimal motor abnormalities). The duration of seizure activity sufficient to meet the definition of status epilepticus has traditionally been specified as 15–30 minutes. However, a more practical definition is to consider status epilepticus as a situation in which the duration of seizures prompts the acute use of anticonvulsant therapy. For GCSE, this is typically when seizures last beyond 5 minutes.

GCSE is an emergency and must be treated immediately , since cardiorespiratory dysfunction, hyperthermia, and metabolic derangements can develop as a consequence of prolonged seizures, and these can lead to irreversible neuronal injury. Furthermore, CNS injury can occur even when the patient is paralyzed with neuromuscular blockade but continues to have electrographic seizures. The most common causes of GCSE are anticonvulsant withdrawal or noncompliance, metabolic disturbances, drug toxicity, CNS infection, CNS tumors, refractory epilepsy, and head trauma.

GCSE is obvious when the patient is having overt convulsions. However, after 30–45 minutes of uninterrupted seizures, the signs may become increasingly subtle. Patients may have mild clonic movements of only the fingers or fine, rapid movements of the eyes. There may be paroxysmal episodes of tachycardia, hypertension, and pupillary dilation. In such cases, the EEG may be the only method of establishing the diagnosis. Thus, if the patient stops having overt seizures, yet remains comatose, an EEG should be performed to rule out ongoing status epilepticus. This is obviously also essential when a patient with GCSE has been paralyzed with neuromuscular blockade in the process of protecting the airway.

The first steps in the management of a patient in GCSE are to attend to any acute cardiorespiratory problems or hyperthermia, perform a brief medical and neurologic examination, establish venous access, and send samples for laboratory studies to identify metabolic abnormalities. Anticonvulsant therapy should then begin without delay; a treatment approach is shown in **Fig. 369-3**.

Figure 369-3

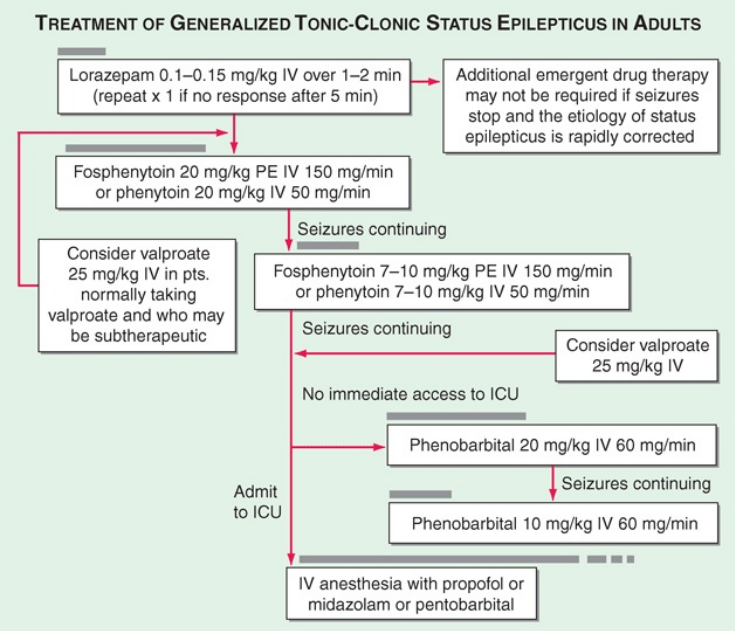

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pharmacologic treatment of generalized tonic-clonic status epilepticus in adults. The horizontal bars indicate the approximate duration of drug infusions. IV, intravenous; PE, phenytoin equivalents.

The treatment of nonconvulsive status epilepticus is thought to be less urgent than GCSE, since the ongoing seizures are not accompanied by the severe metabolic disturbances seen with GCSE. However, evidence suggests that nonconvulsive status epilepticus, especially that caused by ongoing, focal seizure activity, is associated with cellular injury in the region of the seizure focus; therefore this condition should be treated as promptly as possible using the general approach described for GCSE.

BEYOND SEIZURES: OTHER MANAGEMENT ISSUES

Interictal Behavior

The adverse effects of epilepsy often go beyond the occurrence of clinical seizures, and the extent of these effects largely depends on the etiology of the seizure disorder, the degree to which the seizures are controlled, and the presence of side effects from antiepileptic therapy. Many patients with epilepsy are completely normal between seizures and able to live highly successful and productive lives. In contrast, patients with seizures secondary to developmental abnormalities or acquired brain injury may have impaired cognitive function and other neurologic deficits. Frequent interictal EEG abnormalities have been shown to be associated with subtle dysfunction of memory and attention. Patients with many seizures, especially those emanating from the temporal lobe, often note an impairment of short-term memory that may progress over time.

Patients with epilepsy are at risk of developing a variety of psychiatric problems, including depression, anxiety, and psychosis. This risk varies considerably depending on many factors, including the etiology, frequency, and severity of seizures and the patient's age and previous history. Depression occurs in ~20% of patients, and the incidence of suicide is higher in epileptic patients than in the general population. Depression should be treated through counseling or medication. The selective serotonin reuptake inhibitors (SSRIs) typically have no effect on seizures, while high doses of tricyclic antidepressants may lower the seizure threshold. Anxiety can appear as a manifestation of a seizure, and anxious or psychotic behavior can sometimes be observed as part of a postictal delirium. Postictal psychosis is a rare phenomenon that typically occurs after a period of increased seizure frequency. There is usually a brief lucid interval lasting up to a week, followed by days to weeks of agitated, psychotic behavior. The psychosis will usually resolve spontaneously but frequently will require short-term treatment with antipsychotic or anxiolytic medications.

There is ongoing controversy as to whether some patients with epilepsy (especially temporal lobe epilepsy) have a stereotypical "interictal personality." The predominant view is that the unusual or abnormal personality traits observed in such patients are, in most cases, not due to epilepsy but result from an underlying structural brain lesion, the effects of antiepileptic drugs, or psychosocial factors related to suffering from a chronic disease.

Mortality of Epilepsy

Patients with epilepsy have a risk of death that is roughly two to three times greater than expected in a matched population without epilepsy. Most of the increased mortality is due to the underlying etiology of epilepsy (e.g., tumors or strokes in older adults). However, a significant number of patients die from accidents, status epilepticus, and a syndrome known as *sudden unexpected death in epileptic patients* (SUDEP), which usually affects young people with convulsive seizures and tends to occur at night. The cause of SUDEP is unknown; it may result from brainstem-mediated effects of seizures on cardiac rhythms or pulmonary function.

Psychosocial Issues

There continues to be a cultural stigma about epilepsy, although it is slowly declining in societies with effective health education programs. Many patients with epilepsy harbor fears such as the fear of becoming mentally retarded or dying during a seizure. These issues need to be carefully addressed by educating the patient about epilepsy and by ensuring that family members, teachers, fellow employees, and other associates are equally well informed. The Epilepsy Foundation of America (*www.epilepsyfoundation.org*) is a patient advocacy organization and a useful source of educational material as is the Web site *www.epilepsy.com*.

Employment, Driving, and Other Activities

Many patients with epilepsy face difficulty in obtaining or maintaining employment, even when their seizures are well controlled. Federal and state legislation is designed to prevent employers from discriminating against patients with epilepsy, and patients should be encouraged to understand and claim their legal rights. Patients in these circumstances also benefit greatly from the assistance of health providers who act as strong patient advocates.

Loss of driving privileges is one of the most disruptive social consequences of epilepsy. Physicians should be very clear about local regulations concerning driving and epilepsy, since the laws vary considerably among states and countries. In all cases, it is the physician's responsibility to warn patients of the danger imposed on themselves and others while driving if their seizures are uncontrolled (unless the seizures are not associated with impairment of consciousness or motor control). In general, most states allow patients to drive after a seizure-free interval (on or off medications) of between 3 months and 2 years.

Patients with incompletely controlled seizures must also contend with the risk of being in situations where an impairment of consciousness or loss of motor control could lead to major injury or death. Thus, depending on the type and frequency of seizures, many patients need to be instructed to avoid working at heights or with machinery, or to have someone close by for activities such as bathing and swimming.

SPECIAL ISSUES RELATED TO WOMEN AND EPILEPSY

Catamenial Epilepsy

Some women experience a marked increase in seizure frequency around the time of menses. This is believed to be mediated by either the effects of estrogen and progesterone on neuronal excitability or changes in antiepileptic drug levels due to altered protein binding or metabolism. Some patients may benefit from increases in antiepileptic drug dosages during menses. Natural progestins or intramuscular medroxyprogesterone may be of benefit to a subset of women.

Pregnancy

Most women with epilepsy who become pregnant will have an uncomplicated gestation and deliver a normal baby. However, epilepsy poses some important risks to a pregnancy. Seizure frequency during pregnancy will remain unchanged in ~50% of women, increase in 30%, and decrease in 20%. Changes in seizure frequency are attributed to endocrine effects on the CNS, variations in antiepileptic drug pharmacokinetics (such as acceleration of hepatic drug metabolism or effects on plasma protein binding), and changes in medication compliance. It is useful to see patients at frequent intervals during pregnancy and monitor serum antiepileptic drug levels. Measurement of the unbound drug concentrations may be useful if there is an increase in seizure frequency or worsening of side effects of antiepileptic drugs.

The overall incidence of fetal abnormalities in children born to mothers with epilepsy is 5–6%, compared to 2–3% in healthy women. Part of the higher incidence is due to teratogenic effects of antiepileptic drugs, and the risk increases with the number of medications used (e.g., 10–20% risk of malformations with three drugs) and possibly with higher doses. A recent meta-analysis of published pregnancy registries and cohorts found that the most common malformations were defects in the cardiovascular and musculoskeletal system (1.4–1.8%). Valproic acid is strongly associated with an increased risk of adverse fetal outcomes (7–20%). Little is currently known about the safety of newer drugs, although reports suggest a higher than expected incidence of cleft lip or palate with the use of lamotrigine during pregnancy.

Since the potential harm of uncontrolled convulsive seizures on the mother and fetus is considered greater than the teratogenic effects of antiepileptic drugs, it is currently recommended that pregnant women be maintained on effective drug therapy. When possible, it seems prudent to have the patient on monotherapy at the lowest effective dose, especially during the first trimester. For some women, however, the type and frequency of their seizures may allow for them to safely wean off antiepileptic drugs prior to conception. Patients should also take folate (1–4 mg/d), since the antifolate effects of anticonvulsants are thought to play a role in the development of neural tube defects, although the benefits of this treatment remain unproved in this setting.

Enzyme-inducing drugs such as phenytoin, carbamazepine, oxcarbazepine, topiramate, phenobarbital, and primidone cause a transient and reversible deficiency of vitamin K–dependent clotting factors in ~50% of newborn infants. Although neonatal hemorrhage is uncommon, the mother should be treated with oral vitamin K (20 mg/d, phylloquinone) in the last 2 weeks of pregnancy, and the infant should receive intramuscular vitamin K (1 mg) at birth.

Contraception

Special care should be taken when prescribing antiepileptic medications for women who are taking oral contraceptive agents. Drugs such as carbamazepine, phenytoin, phenobarbital, and topiramate can significantly decrease the efficacy of oral contraceptives via enzyme induction and other mechanisms. Patients should be advised to consider alternative forms of contraception, or their contraceptive medications should be modified to offset the effects of the antiepileptic medications.

Breast-Feeding

Antiepileptic medications are excreted into breast milk to a variable degree. The ratio of drug concentration in breast milk relative to serum ranges from ~5% (valproic acid) to 300% (levetiracetam). Given the overall benefits of breast-feeding and the lack of evidence for long-term harm to the infant by being exposed to antiepileptic drugs, mothers with epilepsy can be encouraged to breast-feed. This should be reconsidered, however, if there is any evidence of drug effects on the infant such as lethargy or poor feeding.

FURTHER READINGS

Baulac S, Baulac M: Advances on the genetics of mendelian idiopathic epilepsies. Neurol Clin 27:1041, 2009[PMID: 19853223] [Full Text]

Bonnett LJ et al: Risk of recurrence after a first seizure and implications for driving: Further analysis of the Multicentre study of early Epilepsy and Single Seizures. BMJ 341:6477, 2010

Brodie MJ et al: Epilepsy in later life. Lancet Neurol 8:1019, 2009[PMID: 19800848] [Full Text]

French JA, Pedley TA: Clinical practice. Initial management of epilepsy. N Engl J Med 359:166, 2008.[PMID: 18614784] [Full Text]

——— et al: Efficacy and tolerability of the new antiepileptic drugs I: Treatment of new onset epilepsy: Report of the Therapeutics and Technology Assessment Subcommittee and Quality Standards Subcommittee of the American Academy of Neurology and the American Epilepsy Society. Neurology 62:1252, 2004

———: Efficacy and tolerability of the new antiepileptic drugs II: Treatment of refractory epilepsy: Report of the Therapeutics and Technology Assessment Subcommittee and Quality Standards Subcommittee of the American Academy of Neurology and the American Epilepsy Society. Neurology 62:1261, 2004

Millikan D et al: Emergency treatment of status epilepticus: Current thinking. Emerg Med Clin North Am 27:101, 2009[PMID: 19218022] [Full Text]

Pitkänen A, Lukasiuk K: Molecular and cellular basis of epileptogenesis in symptomatic epilepsy. Epilepsy Behav 14:16, 2009[PMID: 17023142] [Full Text]

Rakhade SN, Jensen FE. Epileptogenesis in the immature brain: emerging mechanisms. Nat Rev Neurol 5:380, 2009[PMID: 19578345] [Full Text]

Schuele SU, Lüders HO: Intractable epilepsy: management and therapeutic alternatives. Lancet Neurol 7:514, 2008[PMID: 18485315] [Full Text]

Tomson T et al: Sudden unexpected death in epilepsy: current knowledge and future directions. Lancet Neurol 7:1021, 2008[PMID: 18805738] [Full Text]

Walker SP et al: The management of epilepsy in pregnancy. BJOG 116:758, 2009[PMID: 19432564] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CEREBROVASCULAR DISEASES: INTRODUCTION

Cerebrovascular diseases include some of the most common and devastating disorders: ischemic stroke, hemorrhagic stroke, and cerebrovascular anomalies such as intracranial aneurysms and arteriovenous malformations (AVMs). They cause ~200,000 deaths each year in the United States and are a major cause of disability. The incidence of cerebrovascular diseases increases with age, and the number of strokes is projected to increase as the elderly population grows, with a doubling in stroke deaths in the United States by 2030. Most cerebrovascular diseases are manifest by the abrupt onset of a focal neurologic deficit, as if the patient was "struck by the hand of God." A stroke, or cerebrovascular accident, is defined by this abrupt onset of a neurologic deficit that is attributable to a focal vascular cause. Thus, the definition of stroke is clinical, and laboratory studies including brain imaging are used to support the diagnosis. The clinical manifestations of stroke are highly variable because of the complex anatomy of the brain and its vasculature. *Cerebral ischemia* is caused by a reduction in blood flow that lasts longer than several seconds. Neurologic symptoms are manifest within seconds because neurons lack glycogen, so energy failure is rapid. If the cessation of flow lasts for more than a few minutes, *infarction* or death of brain tissue results. When blood flow is quickly restored, brain tissue can recover fully and the patient's symptoms are only transient: This is called a *transient ischemic attack* (TIA). The standard definition of TIA requires that all neurologic signs and symptoms resolve within 24 hours regardless of whether there is imaging evidence of new permanent brain injury; stroke has occurred if the neurologic signs and symptoms last for >24 hours. However, a newly proposed definition classifies those with new brain infarction as ischemic strokes regardless of whether symptoms persist. A generalized reduction in cerebral blood flow due to systemic hypotension (e.g., cardiac arrhythmia, myocardial infarction, or hemorrhagic shock) usually produces syncope (Chap. 20). If low cerebral blood flow persists for a longer duration, then infarction in the border zones between the major cerebral artery distributions may develop. In more severe instances, *global hypoxia-ischemia* causes widespread brain injury; the constellation of cognitive sequelae that ensues is called *hypoxic-ischemic encephalopathy* (Chap. 275). *Focal ischemia* or infarction, conversely, is usually caused by thrombosis of the cerebral vessels themselves or by emboli from a proximal arterial source or the heart. *Intracranial hemorrhage* is caused by bleeding directly into or around the brain; it produces neurologic symptoms by producing a mass effect on neural structures, from the toxic effects of blood itself, or by increasing intracranial pressure.

Approach to the Patient: Cerebrovascular Disease

Rapid evaluation is essential for use of time-sensitive treatments such as thrombolysis. However, patients with acute stroke often do not seek medical assistance on their own, both because they are rarely in pain, as well as because they may lose the appreciation that something is wrong (anosognosia); it is often a family member or a bystander who calls for help. Therefore, patients and their family members should be counseled to call emergency medical services immediately if they experience or witness the sudden onset of any of the following: loss of sensory and/or motor function on one side of the body (nearly 85% of ischemic stroke patients have hemiparesis); change in vision, gait, or ability to speak or understand; or if they experience a sudden, severe headache.

There are several common causes of sudden-onset neurologic symptoms that may mimic stroke, including seizure, intracranial tumor, migraine, and metabolic encephalopathy. An adequate history from an observer that no convulsive activity occurred at the onset reasonably excludes seizure; however, ongoing complex partial seizures without tonic-clonic activity may mimic stroke. Tumors may present with acute neurologic symptoms due to hemorrhage, seizure, or hydrocephalus. Surprisingly, migraine can mimic stroke, even in patients without a significant migraine history. When it develops without head pain (*acephalgic migraine*), the diagnosis may remain elusive. Patients without any prior history of migraine may develop acephalgic migraine even after age 65. A sensory disturbance is often prominent, and the sensory deficit, as well as any motor deficits, tends to migrate slowly across a limb over minutes rather than seconds as with stroke. The diagnosis of migraine becomes more secure as the cortical disturbance begins to cross vascular boundaries or if typical visual symptoms are present such as scintillating scotomata (Chap. 14). At times it may be difficult to make the diagnosis until multiple episodes have occurred leaving behind no residual symptoms and with a normal MRI study of the brain. Classically, metabolic encephalopathies produce fluctuating mental status without focal neurologic findings. However, in the setting of prior stroke or brain injury, a patient with fever or sepsis may manifest hemiparesis, which clears rapidly when the infection is remedied. The metabolic process serves to "unmask" a prior deficit.

Once the diagnosis of stroke is made, a brain imaging study is necessary to determine if the cause of stroke is ischemia or hemorrhage **(Fig. 370-1)**. CT imaging of the brain is the standard imaging modality to detect the presence or absence of intracranial hemorrhage (see "Imaging Studies," below). If the stroke is ischemic, administration of recombinant tissue plasminogen activator (rtPA) or endovascular mechanical thrombectomy may be beneficial in restoring cerebral perfusion (see "Treatment: Acute Ischemic Stroke"). Medical

management to reduce the risk of complications becomes the next priority, followed by plans for secondary prevention. For ischemic stroke, several strategies can reduce the risk of subsequent stroke in all patients, while other strategies are effective for patients with specific causes of stroke such as cardiac embolus and carotid atherosclerosis. For hemorrhagic stroke, aneurysmal subarachnoid hemorrhage (SAH) and hypertensive intracranial hemorrhage are two important causes. The treatment and prevention of hypertensive intracranial hemorrhage are discussed later in this chapter. **SAH is discussed in Chap. 275.**

Figure 370-1

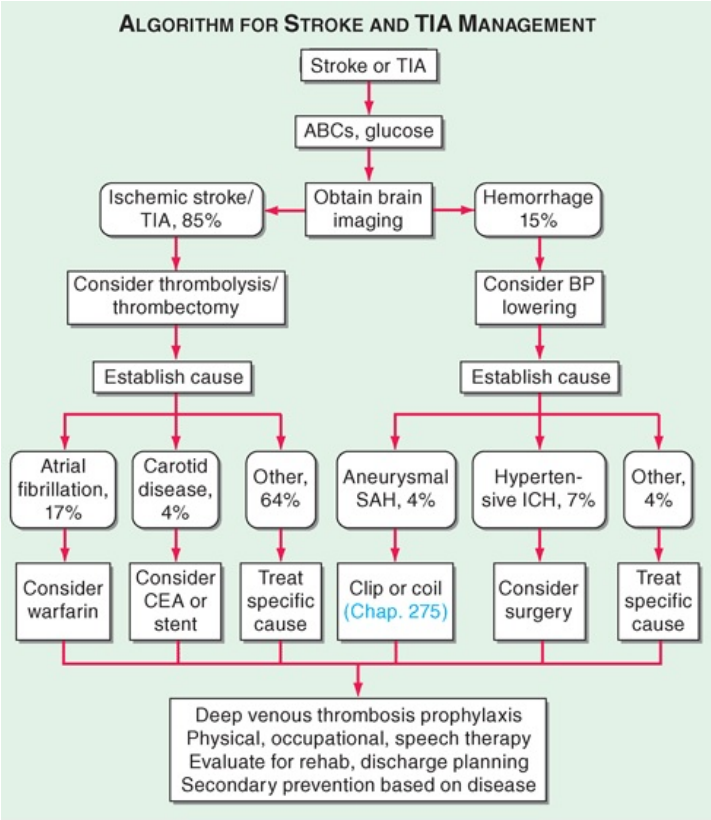

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Medical management of stroke and TIA. Rounded boxes are diagnoses; rectangles are interventions. Numbers are percentages of stroke overall. ABCs, airway, breathing, circulation; BP, blood pressure; CEA, carotid endarterectomy; ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage; TIA, transient ischemic attack.

ISCHEMIC STROKE

Pathophysiology of Ischemic Stroke

Acute occlusion of an intracranial vessel causes reduction in blood flow to the brain region it supplies. The magnitude of flow reduction is a function of collateral blood flow and this depends on individual vascular anatomy, the site of occlusion, and likely systemic blood pressure. A decrease in cerebral blood flow to zero causes death of brain tissue within 4–10 minutes; values <16–18 mL/100 g tissue per minute cause infarction within an hour; and values <20 mL/100 g tissue per minute cause ischemia without infarction unless prolonged for several hours or days. If blood flow is restored prior to a significant amount of cell death, the patient may experience only transient symptoms, and the clinical syndrome is called a TIA. Tissue surrounding the core region of infarction is ischemic but reversibly dysfunctional and is referred to as the *ischemic penumbra*. The penumbra may be imaged by using perfusion-diffusion imaging with MRI or CT (see below and Figs. 370-15 and 370-16). The ischemic penumbra will eventually infarct if no change in flow occurs, and hence saving the ischemic penumbra is the goal of revascularization therapies.

Focal cerebral infarction occurs via two distinct pathways **(Fig. 370-2)**: (1) a necrotic pathway in which cellular cytoskeletal breakdown is rapid, due principally to energy failure of the cell; and (2) an apoptotic pathway in which cells become programmed to die. Ischemia produces necrosis by starving neurons of glucose and oxygen, which in turn results in failure of mitochondria to produce ATP. Without ATP, membrane ion pumps stop functioning and neurons depolarize, allowing intracellular calcium to rise. Cellular depolarization also causes glutamate release from synaptic terminals; excess extracellular glutamate produces neurotoxicity by activating postsynaptic glutamate receptors that increase neuronal calcium influx. Free radicals are produced by membrane lipid degradation and mitochondrial dysfunction. Free radicals cause catalytic destruction of membranes and likely damage other vital functions of cells. Lesser degrees of ischemia, as are seen within the ischemic penumbra, favor apoptotic cellular death causing cells to die days

to weeks later. Fever dramatically worsens brain injury during ischemia, as does hyperglycemia [glucose >11.1 mmol/L (200 mg/dL)], so it is reasonable to suppress fever and prevent hyperglycemia as much as possible. Induced moderate hypothermia to mitigate stroke is the subject of continuing clinical research.

Figure 370-2

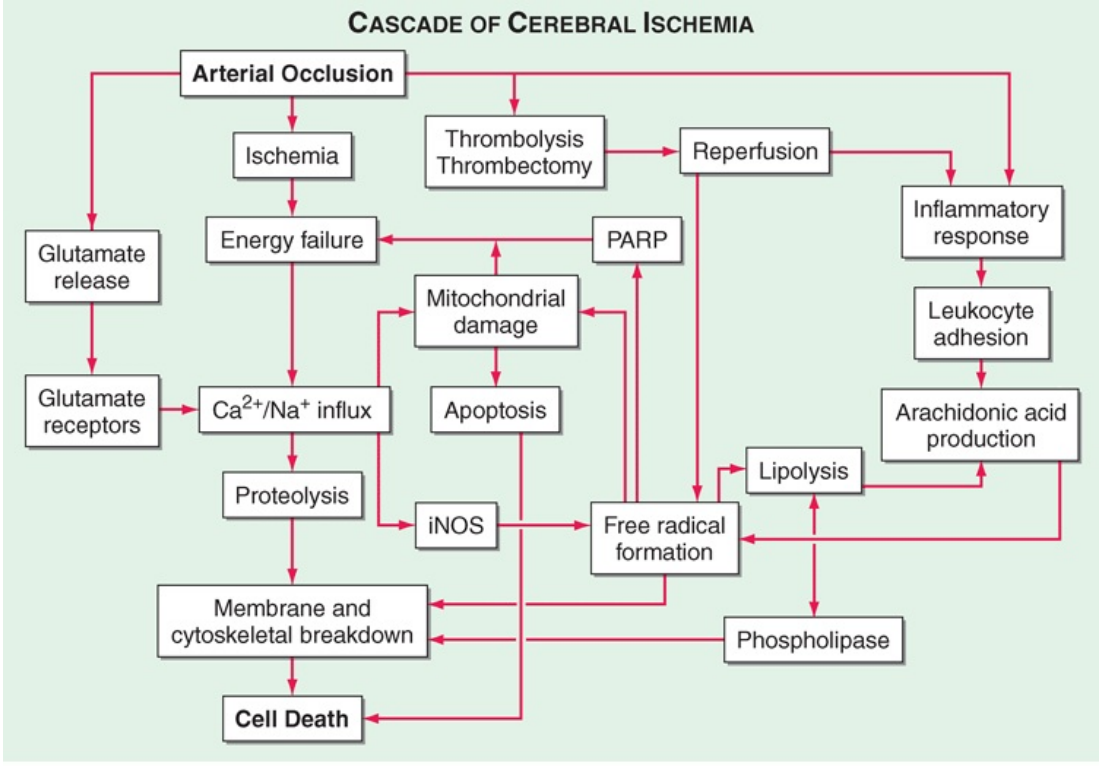

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Major steps in the cascade of cerebral ischemia. See text for details. iNOS, inducible nitric oxide synthase; PARP, poly-A ribose polymerase.

Treatment: Acute Ischemic Stroke

After the clinical diagnosis of stroke is made, an orderly process of evaluation and treatment should follow (Fig. 370-1). The first goal is to prevent or reverse brain injury. Attend to the patient's airway, breathing, circulation (ABC's), and treat hypoglycemia or hyperglycemia if identified. Perform an emergency noncontrast head CT scan to differentiate between ischemic stroke and hemorrhagic stroke; there are no reliable clinical findings that conclusively separate ischemia from hemorrhage, although a more depressed level of consciousness, higher initial blood pressure, or worsening of symptoms after onset favor hemorrhage, and a deficit that is maximal at onset, or remits, suggests ischemia. Treatments designed to reverse or lessen the amount of tissue infarction and improve clinical outcome fall within six categories: (1) medical support (2) IV thrombolysis, (3) endovascular techniques, (4) antithrombotic treatment, (5) neuroprotection, and (6) stroke centers and rehabilitation.

MEDICAL SUPPORT

When ischemic stroke occurs, the immediate goal is to optimize cerebral perfusion in the surrounding ischemic penumbra. Attention is also directed toward preventing the common complications of bedridden patients—infections (pneumonia, urinary, and skin) and deep venous thrombosis (DVT) with pulmonary embolism. Many physicians use pneumatic compression stockings to prevent DVT; subcutaneous heparin (unfractionated and low-molecular weight) is safe and more effective and can be used concomitantly.

Because collateral blood flow within the ischemic brain is blood pressure dependent, there is controversy about whether blood pressure should be lowered acutely. Blood pressure should be lowered if there is malignanthypertension (Chap. 247) or concomitant myocardial ischemia or if blood pressure is >185/110 mm Hg and thrombolytic therapy is anticipated. When faced with the competing demands of myocardium and brain, lowering the heart rate with a β_1-adrenergic blocker (such as esmolol) can be a first step to decrease cardiac work and maintain blood pressure. Fever is detrimental and should be treated with antipyretics and surface cooling. Serum glucose should be monitored and kept at <6.1 mmol/L (110 mg/dL) using an insulin infusion if necessary.

Between 5 and 10% of patients develop enough cerebral edema to cause obtundation or brain herniation. Edema peaks on the second or third day but can cause mass effect for ~10 days. The larger the infarct, the greater the likelihood that clinically significant edema will develop. Water restriction and IV mannitol may be used to raise the serum osmolarity, but hypovolemia should be avoided as this may contribute to

hypotension and worsening infarction. Combined analysis of three randomized European trials of hemicraniectomy (craniotomy and temporary removal of part of the skull) shows that hemicraniectomy markedly reduces mortality, and the clinical outcomes of survivors are acceptable.

Special vigilance is warranted for patients with cerebellar infarction. Such strokes may mimic labyrinthitis because of prominent vertigo and vomiting; the presence of head or neck pain should alert the physician to consider cerebellar stroke from vertebral artery dissection. Even small amounts of cerebellar edema can acutely increase intracranial pressure (ICP) or directly compress the brainstem. The resulting brainstem compression can result in coma and respiratory arrest and require emergency surgical decompression. Prophylactic suboccipital decompression of large cerebellar infarcts before brainstem compression, although not tested rigorously in a clinical trial, is practiced at most stroke centers.

INTRAVENOUS THROMBOLYSIS

The National Institute of Neurological Disorders and Stroke (NINDS) recombinant tPA (rtPA) Stroke Study showed a clear benefit for IV rtPA in selected patients with acute stroke. The NINDS study used IV rtPA (0.9 mg/kg to a 90-mg max; 10% as a bolus, then the remainder over 60 minutes) versus placebo in patients with ischemic stroke within 3 hours of onset. One-half of the patients were treated within 90 minutes. Symptomatic intracerebral hemorrhage occurred in 6.4% of patients on rtPA and 0.6% on placebo. There was a nonsignificant 4% reduction in mortality in patients on rtPA (21% on placebo and 17% on rtPA); there was a significant 12% absolute increase in the number of patients with only minimal disability (32% on placebo and 44% on rtPA.) Thus, despite an increased incidence of symptomatic intracerebral hemorrhage, treatment with IV rtPA within 3 hours of the onset of ischemic stroke improved clinical outcome.

Three subsequent trials of IV rtPA did not confirm this benefit, perhaps because of the dose of rtPA used, the timing of its delivery, and small sample size. When data from all randomized IV rtPA trails were combined, however, efficacy was confirmed in the <3-hour time window, and efficacy likely extended to 4.5 hours if not 6 hours. Based on these combined results, the European Cooperative Acute Stroke Study (ECASS) III study explored the safety and efficacy of rtPA in the 3- to 4.5-hour time window. Unlike the NINDS study, patients older than 85 years of age, and diabetic patients were excluded. In this 821-patient, randomized study efficacy was again confirmed although less robust than in the 0–3 hour time window. In the rtPA group, 52.4% achieved a good outcome while 45.2% of the placebo group had a good outcome at 90 days (OR 1.34, $p = 0.04$). The symptomatic intracranial hemorrhage rate was 2.4% in the rtPA group and 0.2% in the placebo group ($p = 0.008$).

Based on these data, rtPA is being reviewed for approval in the 3–4.5-hour window in Europe, but is only approved for 0–3 hours in the United States and Canada. Use of IV tPA is now considered a central component of primary stroke centers (see below) as the first treatment proven to improve clinical outcomes in ischemic stroke and is cost-effective and cost-saving. One may be able to select patients beyond the 4.5-hour window, who will benefit from thrombolysis using advanced neuroimaging (see neuroimaging section below), but this is currently investigational. The time of stroke onset is defined as the time the patient's symptoms began or the time the patient was last seen as normal. Patients who awaken with stroke have the onset defined as when they went to bed. **Table 370-1** summarizes eligibility criteria and instructions for administration of IV rtPA.

Table 370-1 Administration of Intravenous Recombinant Tissue Plasminogen Activator (Rtpa) for Acute Ischemic Stroke (Ais)[a]

Indication	Contraindication
Clinical diagnosis of stroke	Sustained BP >185/110 mm Hg despite treatment
Onset of symptoms to time of drug administration ≤3 h	Platelets <100,000; HCT <25%; glucose <50 or >400 mg/dL
CT scan showing no hemorrhage or edema of >1/3 of the MCA territory	Use of heparin within 48 h and prolonged PTT, or elevated INR
Age ≥18 years	Rapidly improving symptoms
Consent by patient or surrogate	Prior stroke or head injury within 3 months; prior intracranial hemorrhage
	Major surgery in preceding 14 days
	Minor stroke symptoms
	Gastrointestinal bleeding in preceding 21 days
	Recent myocardial infarction
	Coma or stupor
Administration of rtPA	
Intravenous access with two peripheral IV lines (avoid arterial or central line placement)	
Review eligibility for rtPA	
Administer 0.9 mg/kg IV (maximum 90 mg) IV as 10% of total dose by bolus, followed by remainder of total dose over 1 h	
Frequent cuff blood pressure monitoring	

No other antithrombotic treatment for 24 h
For decline in neurologic status or uncontrolled blood pressure, stop infusion, give cryoprecipitate, and reimage brain emergently
Avoid urethral catheterization for ≥2 h

[a]See Activase (tissue plasminogen activator) package insert for complete list of contraindications and dosing. ***Abbreviations:*** BP, blood pressure; HCT, hematocrit; INR, international normalized ratio; MCA, middle cerebral artery; PTT, partial thromboplastin time.

ENDOVASCULAR TECHNIQUES

Ischemic stroke from large-vessel intracranial occlusion results in high rates of mortality and morbidity. Occlusions in such large vessels [middle cerebral artery (MCA), internal carotid artery, and the basilar artery] generally involve a large clot volume and often fail to open with IV rtPA alone. Therefore, there is growing interest in using thrombolytics via an intraarterial route to increase the concentration of drug at the clot and minimize systemic bleeding complications. The Prolyse in Acute Cerebral Thromboembolism (PROACT) II trial found benefit for intraarterial pro urokinase for acute MCA occlusions up to the sixth hour following onset of stroke. Intraarterial treatment of basilar artery occlusions may also be beneficial for selected patients. Intraarterial administration of a thrombolytic agent for acute ischemic stroke (AIS) is not approved by the U.S. Food and Drug Administration (FDA); however, many stroke centers offer this treatment based on these data.

Endovascular mechanical thrombectomy has recently shown promise as an alternative or adjunctive treatment of acute stroke in patients who are ineligible for, or have contraindications to, thrombolytics or in those who have failed to have vascular recanalization with IV thrombolytics (see Fig. 370-15). The Mechanical Embolus Removal in Cerebral Ischemia (MERCI) and multi-MERCI single-arm trials investigated the ability of a novel endovascular thrombectomy device to restore patency of occluded intracranial vessels within 8 hours of ischemic stroke symptoms. Recanalization of the target vessel occurred in 48–58% of treated patients and in 60–69% following use of adjuvant endovascular methods, and successful recanalization at 90 days correlated well with favorable outcomes. Based upon these nonrandomized data, the FDA approved this device as the first device for revascularization of occluded vessels in acute ischemic stroke even if the patient has been given rtPA and that therapy has failed. The Penumbra Pivotal Stroke trial tested another mechanical device that showed even higher rates of recanalization and led to FDA clearance of the tested device as well. Because use of endovascular devices in combination with rtPA appears safe, primary stroke centers may administer rtPA to eligible patients, and then rapidly refer such patients to comprehensive stroke centers that have endovascular capability for further intervention. Such a design allows centralization of resource-intensive endovascular centers in order to serve larger populations of patients. Use of mechanical techniques to restore blood flow have not as yet been studied in a randomized trial so the clinical efficacy of these treatments remain unproven and the focus of active investigation.

ANTITHROMBOTIC TREATMENT

Platelet Inhibition

Aspirin is the only antiplatelet agent that has been proven effective for the acute treatment of ischemic stroke; there are several antiplatelet agents proven for the secondary prevention of stroke (see below). Two large trials, the International Stroke Trial (IST) and the Chinese Acute Stroke Trial (CAST), found that the use of aspirin within 48 hours of stroke onset reduced both stroke recurrence risk and mortality minimally. Among 19,435 patients in IST, those allocated to aspirin, 300 mg/d, had slightly fewer deaths within 14 days (9.0 versus 9.4%), significantly fewer recurrent ischemic strokes (2.8 versus 3.9%), no excess of hemorrhagic strokes (0.9 versus 0.8%), and a trend toward a reduction in death or dependence at 6 months (61.2 versus 63.5%). In CAST, 21,106 patients with ischemic stroke received 160 mg/d of aspirin or a placebo for up to 4 weeks. There were very small reductions in the aspirin group in early mortality (3.3 versus 3.9%), recurrent ischemic strokes (1.6 versus 2.1%), and dependency at discharge or death (30.5 versus 31.6%). These trials demonstrate that the use of aspirin in the treatment of AIS is safe and produces a small net benefit. For every 1000 acute strokes treated with aspirin, about 9 deaths or nonfatal stroke recurrences will be prevented in the first few weeks and ~13 fewer patients will be dead or dependent at 6 months.

The glycoprotein IIb/IIIa receptor inhibitor abciximab was found to cause excess intracranial hemorrhage and should be avoided in acute stroke. Clopidogrel is being tested as a way to prevent stroke following TIA and minor ischemic stroke.

Anticoagulation

Numerous clinical trials have failed to demonstrate any benefit of anticoagulation in the primary treatment of atherothrombotic cerebral ischemia. Several trials have investigated antiplatelet versus anticoagulant medications given within 12–24 hours of the initial event. The U.S. Trial of Organon 10172 in Acute Stroke Treatment (TOAST), an investigational low-molecular-weight heparin (LMWH), failed to show any benefit over aspirin. Use of SC unfractionated heparin versus aspirin was tested in IST. Heparin given SC afforded no additional benefit over aspirin and increased bleeding rates. Several trials of LMWHs have also shown no consistent benefit in AIS. Furthermore, trials generally have shown an excess risk of brain and systemic hemorrhage with acute anticoagulation. Therefore, trials do not support the routine use of heparin or other anticoagulants for patients with atherothrombotic stroke.

NEUROPROTECTION

Neuroprotection is the concept of providing a treatment that prolongs the brain's tolerance to ischemia. Drugs that block the excitatory amino acid pathways have been shown to protect neurons and glia in animals, but despite multiple human trials, they have not yet been proven to be beneficial. Hypothermia is a powerful neuroprotective treatment in patients with cardiac arrest (Chap. 275) and is neuroprotective in animal models of stroke, but it has not been adequately studied in patients with ischemic stroke.

STROKE CENTERS AND REHABILITATION

Patient care in comprehensive stroke units followed by rehabilitation services improves neurologic outcomes and reduces mortality. Use of clinical pathways and staff dedicated to the stroke patient can improve care. Stroke teams that provide emergency 24-hour evaluation of acute stroke patients for acute medical management and consideration of thrombolysis or endovascular treatments are essential components of primary and comprehensive stroke centers, respectively.

Proper rehabilitation of the stroke patient includes early physical, occupational, and speech therapy. It is directed toward educating the patient and family about the patient's neurologic deficit, preventing the complications of immobility (e.g., pneumonia, DVT and pulmonary embolism, pressure sores of the skin, and muscle contractures), and providing encouragement and instruction in overcoming the deficit. The goal of rehabilitation is to return the patient to home and to maximize recovery by providing a safe, progressive regimen suited to the individual patient. Additionally, the use of restraint therapy (immobilizing the unaffected side) has been shown to improve hemiparesis following stroke, even years following the stroke, suggesting that physical therapy can recruit unused neural pathways. This finding suggests that the human nervous system is more adaptable than originally thought and has stimulated active research into physical and pharmacologic strategies that can enhance long-term neural recovery.

Etiology of Ischemic Stroke

(Figs. 370-1 and 370-3 and **Table 370-2)** Although the initial management of AIS often does not depend on the etiology, establishing a cause is essential in reducing the risk of recurrence. Particular focus should be on atrial fibrillation and carotid atherosclerosis, as these etiologies have proven secondary prevention strategies. The clinical presentation and examination findings often establish the cause of stroke or narrow the possibilities to a few. Judicious use of laboratory testing and imaging studies completes the initial evaluation. Nevertheless, nearly 30% of strokes remain unexplained despite extensive evaluation.

Figure 370-3

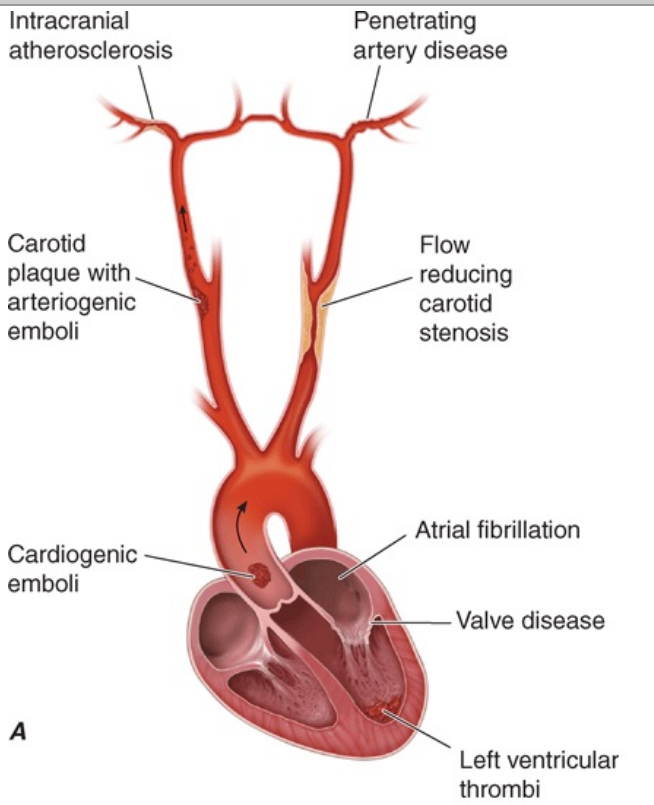

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

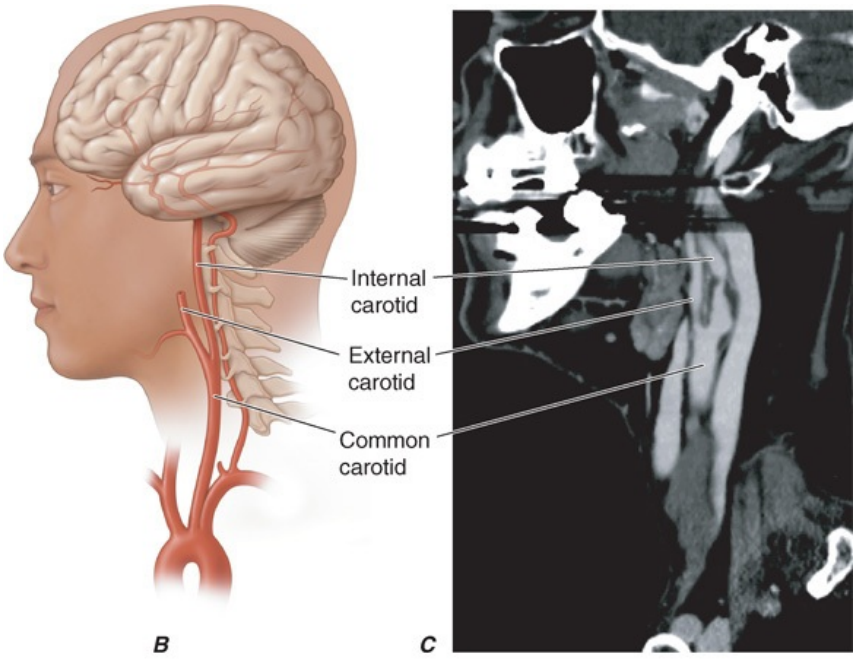

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathophysiology of ischemic stroke. ***A.*** Diagram illustrating the three major mechanisms that underlie ischemic stroke: (1) occlusion of an intracranial vessel by an embolus that arises at a distant site (e.g., cardiogenic sources such as atrial fibrillation or artery-to-artery emboli from carotid atherosclerotic plaque), often affecting the large intracranial vessels; (2) in situ thrombosis of an intracranial vessel, typically affecting the small penetrating arteries that arise from the major intracranial arteries; (3) hypoperfusion caused by flow-limiting stenosis of a major extracranial (e.g., internal carotid) or intracranial vessel, often producing "watershed" ischemia. ***B*** and ***C.*** Diagram and reformatted CT angiogram of the common, internal, and external carotid arteries. High-grade stenosis of the internal carotid artery, which may be associated with either cerebral emboli or flow-limiting ischemia, was identified in this patient.

Table 370-2 Causes of Ischemic Stroke

Common Causes	Uncommon Causes
Thrombosis	Hypercoagulable disorders
Lacunar stroke (small vessel)	Protein C deficiency
Large vessel thrombosis	Protein S deficiency
Dehydration	Antithrombin III deficiency
Embolic occlusion	Antiphospholipid syndrome
Artery-to-artery	Factor V Leiden mutationa
Carotid bifurcation	Prothrombin G20210 mutationa
Aortic arch	Systemic malignancy
Arterial dissection	Sickle cell anemia
Cardioembolic	β-Thalassemia
Atrial fibrillation	Polycythemia vera
Mural thrombus	Systemic lupus erythematosus
Myocardial infarction	Homocysteinemia
Dilated cardiomyopathy	Thrombotic thrombocytopenic purpura
Valvular lesions	Disseminated intravascular coagulation
Mitral stenosis	Dysproteinemias
Mechanical valve	Nephrotic syndrome
Bacterial endocarditis	Inflammatory bowel disease
Paradoxical embolus	Oral contraceptives
	Venous sinus thrombosisb
	Fibromuscular dysplasia
	Vasculitis
	Systemic vasculitis [PAN, granulomatosis with polyangiitis (Wegener's),

Atrial septal defect Patent foramen ovale Atrial septal aneurysm Spontaneous echo contrast	Takayasu's, giant cell arteritis] Primary CNS vasculitis Meningitis (syphilis, tuberculosis, fungal, bacterial, zoster) Cardiogenic Mitral valve calcification Atrial myxoma Intracardiac tumor Marantic endocarditis Libman-Sacks endocarditis Subarachnoid hemorrhage vasospasm Drugs: cocaine, amphetamine Moyamoya disease Eclampsia

[a]Chiefly cause venous sinus thrombosis. [b]May be associated with any hypercoagulable disorder.
Abbreviations: CNS, central nervous system; PAN, polyarteritis nodosa.

Clinical examination should focus on the peripheral and cervical vascular system (carotid auscultation for bruits, blood pressure, and pressure comparison between arms), the heart (dysrhythmia, murmurs), extremities (peripheral emboli), and retina [effects of hypertension and cholesterol emboli (Hollenhorst plaques)]. A complete neurologic examination is performed to localize the site of stroke. An imaging study of the brain is nearly always indicated and is required for patients being considered for thrombolysis; it may be combined with CT- or MRI-based angiography to interrogate the neck and intracranial vessels (see "Imaging Studies," below). A chest x-ray, electrocardiogram (ECG), urinalysis, complete blood count, erythrocyte sedimentation rate (ESR), serum electrolytes, blood urea nitrogen (BUN), creatinine, blood sugar, serologic test for syphilis, serum lipid profile, prothrombin time (PT), and partial thromboplastin time (PTT) are often useful and should be considered in all patients. An ECG may demonstrate arrhythmias or reveal evidence of recent myocardial infarction (MI).

CARDIOEMBOLIC STROKE

Cardioembolism is responsible for ~20% of all ischemic strokes. Stroke caused by heart disease is primarily due to embolism of thrombotic material forming on the atrial or ventricular wall or the left heart valves. These thrombi then detach and embolize into the arterial circulation. The thrombus may fragment or lyse quickly, producing only a TIA. Alternatively, the arterial occlusion may last longer, producing stroke. Embolic strokes tend to be sudden in onset, with maximum neurologic deficit at once. With reperfusion following more prolonged ischemia, petechial hemorrhage can occur within the ischemic territory. This is usually of no clinical significance and should be distinguished from frank intracranial hemorrhage into a region of ischemic stroke where the mass effect from the hemorrhage can cause a decline in neurologic function.

Emboli from the heart most often lodge in the MCA, the posterior cerebral artery (PCA), or one of their branches; infrequently, the anterior cerebral artery (ACA) territory is involved. Emboli large enough to occlude the stem of the MCA (3–4 mm) lead to large infarcts that involve both deep gray and white matter and some portions of the cortical surface and its underlying white matter. A smaller embolus may occlude a small cortical or penetrating arterial branch. The location and size of an infarct within a vascular territory depend on the extent of the collateral circulation.

The most significant causes of cardioembolic stroke in most of the world are nonrheumatic (often called nonvalvular) atrial fibrillation, MI, prosthetic valves, rheumatic heart disease, and ischemic cardiomyopathy (Table 370-2). Cardiac disorders causing brain embolism are discussed in the respective chapters on heart diseases. A few pertinent aspects are highlighted here.

Nonrheumatic atrial fibrillation is the most common cause of cerebral embolism overall. The presumed stroke mechanism is thrombus formation in the fibrillating atrium or atrial appendage, with subsequent embolization. Patients with atrial fibrillation have an average annual risk of stroke of ~5%. The risk of stroke can be estimated by calculating the CHADS2 score (see **Table 370-3**) Left atrial enlargement is an additional risk factor for formation of atrial thrombi. Rheumatic heart disease usually causes ischemic stroke when there is prominent mitral stenosis or atrial fibrillation. Recent MI may be a source of emboli, especially when transmural and involving the anteroapical ventricular wall, and prophylactic anticoagulation following MI has been shown to reduce stroke risk. Mitral valve prolapse is not usually a source of emboli unless the prolapse is severe.

Table 370-3 Recommendations on Chronic Use of Antithrombotics for Various Cardiac Conditions

Condition	Recommendation
Nonvalvular atrial fibrillation	Calculate CHADS2a score
• CHADS2 score 0	Aspirin or no antithrombotic
• CHADS2 score 1	Aspirin or VKA

• CHADS2 score >1	VKA
Rheumatic mitral valve disease	
• With atrial fibrillation, previous embolization, or atrial appendage thrombus, or left atrial diameter >55 mm	VKA
• Embolization or appendage clot despite INR 2–3	VKA plus aspirin
Mitral valve prolapse	
• Asymptomatic	No therapy
• With otherwise cryptogenic stroke or TIA	Aspirin
• Atrial fibrillation	VKA
Mitral annular calcification	
• Without atrial fibrillation but systemic embolization, or otherwise cryptogenic stroke or TIA	Aspirin
• Recurrent embolization despite aspirin	VKA
• With atrial fibrillation	VKA
Aortic valve calcification	
• Asymptomatic	No therapy
• Otherwise cryptogenic stroke or TIA	Aspirin
Aortic arch mobile atheroma	
• Otherwise cryptogenic stroke or TIA	Aspirin or VKA
Patent foramen ovale	
• Otherwise cryptogenic ischemic stroke or TIA	Aspirin
• Indication for VKA (deep venous thrombosis or hypercoagulable state)	VKA
Mechanical heart value	
• Aortic position, bileaflet or Medtronic Hall tilting disk with normal left atrial size and sinus rhythm	VKA INR 2.5, range 2–3
• Mitral position tilting disk or bileaflet valve	VKA INR 3.0, range 2.5–3.5
• Mitral or aortic position, anterior-apical myocardial infarct or left atrial enlargement	VKA INR 3.0, range 2.5–3.5
• Mitral or aortic position, with atrial fibrillation, or hypercoagulable state, or low ejection fraction, or atherosclerotic vascular disease	Aspirin plus VKA INR 3.0, range 2.5–3.5
• Systemic embolization despite target INR Bioprosthetic valve	Add aspirin and/or increase INR: prior target was 2.5 increase to 3.0, range 2.5–3.5; prior target was 3.0 increase to 3.5, range 3–4
• No other indication for VKA therapy	Aspirin
Infective endocarditis	Avoid antithrombotic agents
Nonbacterial thrombotic endocarditis	
• With systemic embolization	Full dose unfractionated heparin or SC LMWH

[a]CHADS2 score calculated as follows: 1 point for age >75 years, 1 point for hypertension, 1 point for congestive heart failure, 1 point for diabetes, and 2 points for stroke or TIA; sum of points is the total CHADS2 score.

Abbreviations: Dose of aspirin is 50–325 mg/d; target INR for VKA is 2.5 unless otherwise specified. INR, international normalized ratio; LMWH, low-molecular-weight heparin; TIA, transient ischemic attack; VKA, vitamin K antagonist.

Sources: Modified from DE Singer et al: Chest 133:546S, 2008; DN Salem et al: Chest 133:593S, 2008.

Paradoxical embolization occurs when venous thrombi migrate to the arterial circulation, usually via a patent foramen ovale or atrial septal defect. Bubble-contrast echocardiography (IV injection of agitated saline coupled with either transthoracic or transesophageal echocardiography) can demonstrate a right-to-left cardiac shunt, revealing the conduit for paradoxical embolization. Alternatively, a right-to-left shunt is implied if immediately following IV injection of agitated saline, the ultrasound signature of bubbles is observed during transcranial Doppler insonation of the MCA; pulmonary AVMs should be considered if this test is positive yet an echocardiogram fails to reveal an intracardiac shunt. Both techniques are highly sensitive for detection of right-to-left shunts. Besides venous clot, fat and tumor emboli, bacterial endocarditis, IV air, and amniotic fluid emboli at childbirth may occasionally be responsible for paradoxical embolization. The importance of right-to-left shunt as a cause of stroke is debated, particularly because such shunts are present in ~15% of the general population. Some studies have suggested that the risk is only elevated in the presence of a coexisting atrial septal aneurysm. The presence of a venous source of embolus, most commonly a deep venous thrombus, may provide confirmation of the importance of a right-to-left shunt in a particular case.

Bacterial endocarditis can cause valvular vegetations that can give rise to septic emboli. The appearance of

multifocal symptoms and signs in a patient with stroke makes bacterial endocarditis more likely. Infarcts of microscopic size occur, and large septic infarcts may evolve into brain abscesses or cause hemorrhage into the infarct, which generally precludes use of anticoagulation or thrombolytics. Mycotic aneurysms caused by septic emboli give rise to SAH or intracerebral hemorrhage.

ARTERY-TO-ARTERY EMBOLIC STROKE

Thrombus formation on atherosclerotic plaques may embolize to intracranial arteries producing an artery-to-artery embolic stroke. Less commonly, a diseased vessel may acutely thrombose. Unlike the myocardial vessels, artery-to-artery embolism, rather than local thrombosis, appears to be the dominant vascular mechanism causing brain ischemia. Any diseased vessel may be an embolic source, including the aortic arch, common carotid, internal carotid, vertebral, and basilar arteries. Carotid bifurcation atherosclerosis is the most common source of artery-to-artery embolus, and specific treatments have proven efficacy in reducing risk.

Carotid Atherosclerosis

Atherosclerosis within the carotid artery occurs most frequently within the common carotid bifurcation and proximal internal carotid artery. Additionally, the carotid siphon (portion within the cavernous sinus) is also vulnerable to atherosclerosis. Male gender, older age, smoking, hypertension, diabetes, and hypercholesterolemia are risk factors for carotid disease, as they are for stroke in general **(Table 370-4)**. Carotid atherosclerosis produces an estimated 10% of ischemic stroke. **For further discussion of the pathogenesis of atherosclerosis, see Chap. 241.**

Table 370-4 Risk Factors for Stroke

Risk Factor	Relative Risk	Relative Risk Reduction with Treatment	Number Needed to Treat[a]	
			Primary Prevention	Secondary Prevention
Hypertension	2–5	38%	100–300	50–100
Atrial fibrillation	1.8–2.9	68% warfarin, 21% aspirin	20–83	13
Diabetes	1.8–6	No proven effect		
Smoking	1.8	50% at 1 year, baseline risk at 5 years' postcessation		
Hyperlipidemia	1.8–2.6	16–30%	560	230
Asymptomatic carotid stenosis	2.0	53%	85	N/A
Symptomatic carotid stenosis (70–99%)		65% at 2 years	N/A	12
Symptomatic carotid stenosis (50–69%)		29% at 5 years	N/A	77

[a]Number needed to treat to prevent one stroke annually. Prevention of other cardiovascular outcomes is not considered here.

Abbreviation: N/A, not applicable.

Carotid disease can be classified by whether the stenosis is symptomatic or asymptomatic and by the degree of stenosis (percent narrowing of the narrowest segment compared to a more distal internal carotid segment). Symptomatic carotid disease implies that the patient has experienced a stroke or TIA within the vascular distribution of the artery, and it is associated with a greater risk of subsequent stroke than asymptomatic stenosis, in which the patient is symptom free and the stenosis is detected through screening. Greater degrees of arterial narrowing are generally associated with a greater risk of stroke, except that those with near occlusions are at lower risk of stroke.

Other Causes of Artery-to-Artery Embolic Stroke

Intracranial atherosclerosis produces stroke either by an embolic mechanism or by in situ thrombosis of a diseased vessel. It is more common in patients of Asian and African-American descent. Recurrent stroke risk is ~15% per year, similar to symptomatic untreated carotid atherosclerosis.

Dissection of the internal carotid or vertebral arteries or even vessels beyond the circle of Willis is a common source of embolic stroke in young (age <60 years) patients. The dissection is usually painful and precedes the stroke by several hours or days. Extracranial dissections do not cause hemorrhage because of the tough adventitia of these vessels. Intracranial dissections, conversely, may produce SAH because the adventitia of intracranial vessels is thin and pseudoaneurysms may form, requiring urgent treatment to prevent rerupture. Treating asymptomatic pseudoaneurysms following dissection is controversial. The cause of dissection is usually unknown and recurrence is rare. Ehlers-Danlos type IV, Marfan's disease, cystic medial necrosis, and fibromuscular dysplasia are associated with dissections. Trauma (usually a motor vehicle accident or a sports injury) can cause carotid and vertebral artery dissections. Spinal manipulative therapy is independently associated with vertebral artery dissection and stroke. Most dissections heal spontaneously, and stroke or TIA is uncommon beyond 2 weeks. Although there are no trials comparing

anticoagulation to antiplatelet agents, many physicians treat acutely with anticoagulants then convert to antiplatelet therapy after demonstration of satisfactory vascular recanalization.

Small-Vessel Stroke

The term *lacunar infarction* refers to infarction following atherothrombotic or lipohyalinotic occlusion of a small artery (30–300 μm) in the brain. The term *small-vessel stroke* denotes occlusion of such a small penetrating artery and is now the preferred term. Small-vessel strokes account for ~20% of all strokes.

Pathophysiology

The MCA stem, the arteries comprising the circle of Willis (A1 segment, anterior and posterior communicating arteries, and P1 segment), and the basilar and vertebral arteries all give rise to 30- to 300-μm branches that penetrate the deep gray and white matter of the cerebrum or brainstem **(Fig. 370-4)**. Each of these small branches can occlude either by atherothrombotic disease at its origin or by the development of lipohyalinotic thickening. Thrombosis of these vessels causes small infarcts that are referred to as *lacunes* (Latin for "lake" of fluid noted at autopsy). These infarcts range in size from 3 mm to 2 cm in diameter. Hypertension and age are the principal risk factors.

Figure 370-4

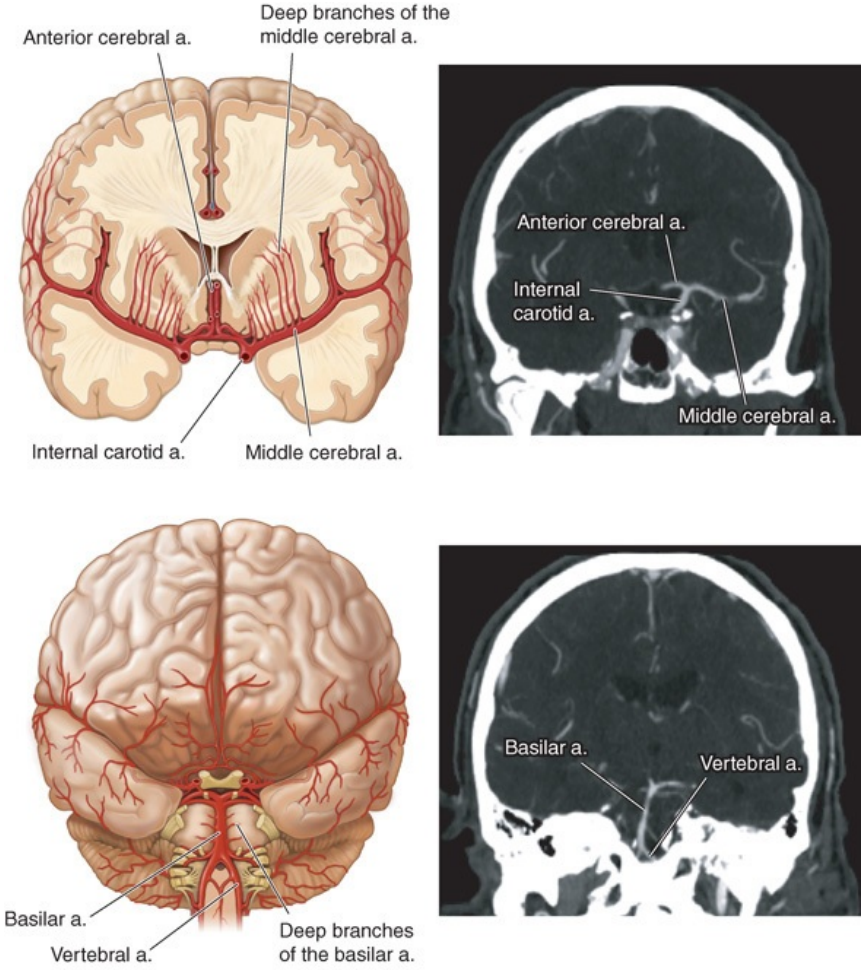

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagrams and reformatted CT angiograms in the coronal section illustrating the deep penetrating arteries involved in small-vessel strokes. In the anterior circulation, small penetrating arteries called *lenticulostriates* arise from the proximal portion of the anterior and middle cerebral arteries and supply deep subcortical structures (***upper panels***). In the posterior circulation, similar arteries arise directly from the vertebral and basilar arteries to supply the brainstem (***lower panels***). Occlusion of a single penetrating artery gives rise to a discrete area of infarct (pathologically termed a "lacune," or lake). Note that these vessels are too small to be visualized on CT angiography.

Clinical Manifestations

The most common *lacunar syndromes* are the following: (1) *Pure motor hemiparesis* from an infarct in the posterior limb of the internal capsule or basis pontis; the face, arm, and leg are almost always involved; (2) *pure sensory stroke* from an infarct in the ventral thalamus; (3) *ataxic hemiparesis* from an infarct in the ventral pons or internal capsule; (4) and *dysarthria and a clumsy hand* or arm due to infarction in the ventral pons or in the genu of the internal capsule.

Transient symptoms (small vessel TIAs) may herald a small-vessel infarct; they may occur several times a day and last only a few minutes. Recovery from small-vessel strokes tends to be more rapid and complete than recovery from large-vessel strokes; in some cases, however, there is severe permanent disability. Often, institution of combined antithrombotic treatments does not prevent eventual stroke in "stuttering lacunes."

A large-vessel source (either thrombosis or embolism) may manifest initially as a lacunar syndrome with small-vessel infarction. Therefore, the search for embolic sources (carotid and heart) should not be completely abandoned in the evaluation of these patients. Secondary prevention of lacunar stroke involves risk factor modification, specifically reduction in blood pressure (see "Primary and Secondary Prevention of Stroke and TIA," below).

Less Common Causes of Stroke

(Table 370-2) *Hypercoagulable disorders* (Chap. 58) primarily cause increased risk of venous thrombosis and therefore may cause venous sinus thrombosis. Protein S deficiency and homocysteinemia may cause arterial thromboses as well. Systemic lupus erythematosus with Libman-Sacks endocarditis can be a cause of embolic stroke. These conditions overlap with the antiphospholipid syndrome, which probably requires long-term anticoagulation to prevent further stroke.

Venous sinus thrombosis of the lateral or sagittal sinus or of small cortical veins (cortical vein thrombosis) occurs as a complication of oral contraceptive use, pregnancy and the postpartum period, inflammatory bowel disease, intracranial infections (meningitis), and dehydration. It is also seen with increased incidence in patients with laboratory-confirmed thrombophilia (Table 370-2) including polycythemia, sickle cell anemia, deficiencies of proteins C and S, factor V Leiden mutation (resistance to activated protein C), antithrombin III deficiency, homocysteinemia, and the prothrombin G20210 mutation. Women who take oral contraceptives and have the prothrombin G20210 mutation may be at particularly high risk for sinus thrombosis. Patients present with headache and may also have focal neurologic signs (especially paraparesis) and seizures. Often, CT imaging is normal unless an intracranial venous hemorrhage has occurred, but the venous sinus occlusion is readily visualized using MR- or CT-venography or conventional x-ray angiography. With greater degrees of sinus thrombosis, the patient may develop signs of increased ICP and coma. Intravenous heparin, regardless of the presence of intracranial hemorrhage, has been shown to reduce morbidity and mortality, and the long-term outcome is generally good. Heparin prevents further thrombosis and reduces venous hypertension and ischemia. If an underlying hypercoagulable state is not found, many physicians treat with vitamin K antagonists (VKAs) for 3–6 months then convert to aspirin, depending on the degree of resolution of the venous sinus thrombus. Anticoagulation is often continued indefinitely if thrombophilia is diagnosed.

Sickle cell anemia (SS disease) is a common cause of stroke in children. A subset of homozygous carriers of this hemoglobin mutation develop stroke in childhood and this may be predicted by documenting high-velocity blood flow within the MCAs using transcranial Doppler ultrasonography. In children who are identified to have high velocities, treatment with aggressive exchange transfusion dramatically reduces risk of stroke, and if exchange transfusion is ceased, their stroke rate increases again along with MCA velocities.

Fibromuscular dysplasia affects the cervical arteries and occurs mainly in women. The carotid or vertebral arteries show multiple rings of segmental narrowing alternating with dilatation. Occlusion is usually incomplete. The process is often asymptomatic but occasionally is associated with an audible bruit, TIAs, or stroke. Involvement of the renal arteries is common and may cause hypertension. The cause and natural history of fibromuscular dysplasia are unknown (Chap. 249). TIA or stroke generally occurs only when the artery is severely narrowed or dissects. Anticoagulation or antiplatelet therapy may be helpful.

Temporal (giant cell) arteritis (Chap. 326) is a relatively common affliction of elderly persons in which the external carotid system, particularly the temporal arteries, undergo subacute granulomatous inflammation with giant cells. Occlusion of posterior ciliary arteries derived from the ophthalmic artery results in blindness in one or both eyes and can be prevented with glucocorticoids. It rarely causes stroke as the internal carotid artery is usually not inflamed. Idiopathic giant cell arteritis involving the great vessels arising from the aortic arch (*Takayasu's arteritis*) may cause carotid or vertebral thrombosis; it is rare in the Western hemisphere.

Necrotizing (or granulomatous) arteritis , occurring alone or in association with generalized polyarteritis nodosa or granulomatosis with polyangiitis (Wegener's), involves the distal small branches (<2 mm diameter) of the main intracranial arteries and produces small ischemic infarcts in the brain, optic nerve, and spinal cord. The cerebrospinal fluid (CSF) often shows pleocytosis, and the protein level is elevated. *Primary central nervous system vasculitis* is rare; small or medium-sized vessels are usually affected, without apparent systemic vasculitis. The differential diagnosis includes other inflammatory causes of vascular caliber change including infection (tubercular, fungal), sarcoidosis, angiocentric lymphoma, carcinomatous meningitis, as well as noninflammatory causes such as atherosclerosis, emboli, connective tissue disease, vasospasm, migraine-associated vasculopathy, and drug-associated causes. Some cases follow the postpartum period and are self-limited.

Patients with any form of vasculopathy may present with insidious progression of combined white and gray matter infarctions, prominent headache, and cognitive decline. Brain biopsy or high-resolution conventional x-ray angiography is usually required to make the diagnosis **(Fig. 370-5)**. An inflammatory profile found on lumbar puncture favors an inflammatory cause. In cases where inflammation is confirmed, aggressive immunosuppression with glucocorticoids, and often cyclophosphamide, is usually necessary to prevent progression; a diligent investigation for infectious causes such as tuberculosis is essential prior to

immunosuppression. With prompt recognition and treatment, many patients can make an excellent recovery.

Figure 370-5

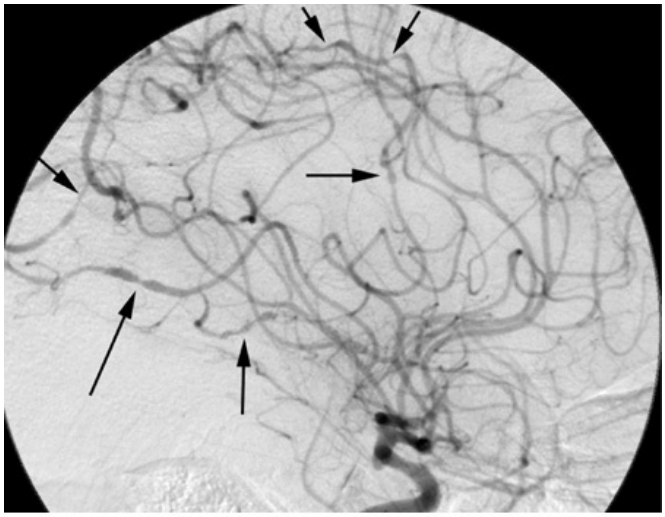

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cerebral angiogram from a 32-year-old male with central nervous system vasculopathy. Dramatic beading (*arrows*) typical of vasculopathy is seen.

Drugs, in particular amphetamines and perhaps cocaine, may cause stroke on the basis of acute hypertension or drug-induced vasculopathy. No data exist on the value of any treatment. Phenylpropanolamine has been linked with intracranial hemorrhage, as has cocaine and methamphetamine, perhaps related to a drug-induced vasculopathy. *Moyamoya disease* is a poorly understood occlusive disease involving large intracranial arteries, especially the distal internal carotid artery and the stem of the MCA and ACA. Vascular inflammation is absent. The lenticulostriate arteries develop a rich collateral circulation around the occlusive lesion, which gives the impression of a "puff of smoke" (*moyamoya* in Japanese) on conventional x-ray angiography. Other collaterals include transdural anastomoses between the cortical surface branches of the meningeal and scalp arteries. The disease occurs mainly in Asian children or young adults, but the appearance may be identical in adults who have atherosclerosis, particularly in association with diabetes. Because of the occurrence of intracranial hemorrhage from rupture of the transdural and pial anastomotic channels, anticoagulation is risky. Breakdown of dilated lenticulostriate arteries may produce parenchymal hemorrhage, and progressive occlusion of large surface arteries can occur, producing large-artery distribution strokes. Surgical bypass of extracranial carotid arteries to the dura or MCAs may prevent stroke and hemorrhage.

Reversible posterior leukoencephalopathy can occur in head injury, seizure, migraine, sympathomimetic drug use, eclampsia, and the postpartum period. The pathophysiology is uncertain but likely involves widespread cerebral segmental vasoconstriction and cerebral edema. Patients complain of headache and manifest fluctuating neurologic symptoms and signs, especially visual symptoms. Sometimes cerebral infarction ensues, but typically the clinical and imaging findings suggest that ischemia reverses completely. MRI findings are characteristic, and conventional x-ray angiography may also be helpful in establishing the diagnosis.

Leukoaraiosis, or *periventricular white matter disease,* is the result of multiple small-vessel infarcts within the subcortical white matter. It is readily seen on CT or MRI scans as areas of white matter injury surrounding the ventricles and within the corona radiata. Areas of lacunar infarction are often seen also. The pathophysiologic basis of the disease is lipohyalinosis of small penetrating arteries within the white matter, likely produced by chronic hypertension. Patients with periventricular white matter disease may develop a subcortical dementia syndrome, depending on the amount of white matter infarction, and it is likely that this common form of dementia may be delayed or prevented with antihypertensive medications (Chap. 371).

CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) is an inherited disorder that presents as small-vessel strokes, progressive dementia, and extensive symmetric white matter changes visualized by MRI. Approximately 40% of patients have migraine with aura, often manifest as transient motor or sensory deficits. Onset is usually in the fourth or fifth decade of life. This autosomal dominant condition is caused by one of several mutations in Notch-3, a member of a highly conserved gene family characterized by epidermal growth factor repeats in its extracellular domain. Other monogenic ischemic stroke syndromes include cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy (CARASIL) and hereditary endotheliopathy, retinopathy, nephropathy, and stroke (HERNS). Fabry's disease also produces both large-vessel arteriopathy and small-vessel infarcts by an unknown mechanism.

Transient Ischemic Attacks

TIAs are episodes of stroke symptoms that last only briefly; the standard definition of duration is <24 hours, but most TIAs last <1 hour. The causes of TIA are similar to the causes of ischemic stroke, but because TIAs may herald stroke they are an important risk factor that should be considered separately. TIAs may arise from emboli to the brain or from in situ thrombosis of an intracranial vessel. With a TIA, the occluded blood vessel reopens and neurologic function is restored. However, infarcts of the brain do occur in 15–50% of TIAs even though neurologic signs and symptoms are absent. Newer definitions of TIA categorize those with new infarct as having ischemic stroke rather than TIA regardless of symptom duration, but the vast majority of studies have used the standard, time-based definition.

In addition to the stroke syndromes discussed below, one specific TIA symptom should receive special notice. *Amaurosis fugax*, or transient monocular blindness, occurs from emboli to the central retinal artery of one eye. This may indicate carotid stenosis as the cause or local ophthalmic artery disease.

The risk of stroke after a TIA is ~10–15% in the first 3 months, with most events occurring in the first 2 days. This risk can be directly estimated using the well validated ABCD^2 method (**Table 370-5**). Therefore, urgent evaluation and treatment are justified. Since etiologies for stroke and TIA are identical, evaluation for TIA should parallel that of stroke (Figs. 370-1 and 370-3). The improvement characteristic of TIA is a contraindication to thrombolysis. However, since the risk of subsequent stroke in the first few days after a TIA is high, the opportunity to give rtPA rapidly if a stroke occurs probably justifies hospital admission for most patients. Acute antiplatelet therapy has not been tested specifically after TIA but is likely to be effective and is recommended. A large-scale trial of acute antithrombotic treatment to prevent stroke following TIA is ongoing.

Table 370-5 Risk of Stroke Following Tia: The Abcd^2 Score

Clinical Factor	Score
A: Age ≥60 years	1
B: SBP >140 mm Hg or DBP >90 mm Hg	1
C: Clinical symptoms	
Unilateral weakness	2
Speech disturbance without weakness	1
D: Duration	
>60 minutes	2
10–59 minutes	1
D: Diabetes (oral medications or insulin)	1
Total Score	Sum Each Category
ABCD2 Score Total	3-Month Rate of Stroke (%)[a]
0	0
1	2
2	3
3	3
4	8
5	12
6	17
7	22

[a]Data ranges are from 5 cohorts. ***Abbreviations:*** DBP, diastolic blood pressure; SBP, systolic blood pressure. ***Source:*** SC Johnston et al: Validation and refinement of score to predict very early stroke risk after transient ischaemic attack, Lancet 369: 283, 2007.

Treatment: Primary and Secondary Prevention of Stroke and Tia

GENERAL PRINCIPLES

A number of medical and surgical interventions, as well as lifestyle modifications, are available for preventing stroke. Some of these can be widely applied because of their low cost and minimal risk; others are expensive and carry substantial risk but may be valuable for selected high-risk patients. Identification and control of modifiable risk factors is the best strategy to reduce the burden of stroke, and the total number of strokes could be reduced substantially by these means (Table 370-4).

ATHEROSCLEROSIS RISK FACTORS

The relationship of various factors to the risk of atherosclerosis is described in Chap. 241. Older age, family history of thrombotic stroke, diabetes mellitus, hypertension, tobacco smoking, abnormal blood cholesterol [particularly, low high-density lipoprotein (HDL) and/or high low-density lipoprotein (LDL)], and other factors are either proven or probable risk factors for ischemic stroke, largely by their link to atherosclerosis. Risk of stroke is much greater in those with prior stroke or TIA. Many cardiac conditions predispose to stroke,

including atrial fibrillation and recent MI. Oral contraceptives and hormone replacement therapy increase stroke risk, and certain inherited and acquired hypercoagulable states predispose to stroke. Hypertension is the most significant of the risk factors; in general, all hypertension should be treated. The presence of known cerebrovascular disease is not a contraindication to treatment aimed at achieving normotension. Also, the value of treating systolic hypertension in older patients has been clearly established. Lowering blood pressure to levels below those traditionally defining hypertension appears to reduce the risk of stroke even further. Data are particularly strong in support of thiazide diuretics and angiotensin-converting enzyme inhibitors.

Several trials have confirmed that statin drugs reduce the risk of stroke even in patients without elevated LDL or low HDL. The Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial showed benefit in secondary stroke reduction for patients with recent stroke or TIA, who were prescribed atorvastatin, 80 mg/d. The primary prevention trial, Justification for the Use of Statins in Prevention: An Intervention Trial Evaluating Rosuvastatin (JUPITER), found that patients with low LDL (<130 mg/dL) caused by elevated C-reactive protein benefitted by daily use of this statin. Primary stroke occurrence was reduced by 51% (hazard ratio 0.49, $p = 0.004$) and there was no increase in the rates of intracranial hemorrhage. Therefore, a statin should be considered in all patients with prior ischemic stroke. Tobacco smoking should be discouraged in all patients (Chap. 395). Tight control of blood sugar in patients with type II diabetes lowers stroke risk MI and death of any cause, but no trial sufficiently powered to detect a significant reduction in stroke has yet been performed. Statins, more aggressive blood pressure control, and pioglitazone (an agonist of peroxisome proliferator-activated receptor gamma) are effective.

ANTIPLATELET AGENTS

Platelet antiaggregation agents can prevent atherothrombotic events, including TIA and stroke, by inhibiting the formation of intraarterial platelet aggregates. These can form on diseased arteries, induce thrombus formation, and occlude the artery or embolize into the distal circulation. Aspirin, clopidogrel, and the combination of aspirin plus extended-release dipyridamole are the antiplatelet agents most commonly used for this purpose. Ticlopidine has been largely abandoned because of its adverse effects but may be used as an alternative to clopidogrel.

Aspirin is the most widely studied antiplatelet agent. Aspirin acetylates platelet cyclooxygenase, which irreversibly inhibits the formation in platelets of thromboxane A_2, a platelet aggregating and vasoconstricting prostaglandin. This effect is permanent and lasts for the usual 8-day life of the platelet. Paradoxically, aspirin also inhibits the formation in endothelial cells of prostacyclin, an antiaggregating and vasodilating prostaglandin. This effect is transient. As soon as aspirin is cleared from the blood, the nucleated endothelial cells again produce prostacyclin. Aspirin in low doses given once daily inhibits the production of thromboxane A_2 in platelets without substantially inhibiting prostacyclin formation. Higher doses of aspirin have not been proven to be more effective than lower doses, and 50–325 mg/d of aspirin is generally recommended for stroke prevention.

Ticlopidine and clopidogrel block the adenosine diphosphate (ADP) receptor on platelets and thus prevent the cascade resulting in activation of the glycoprotein IIb/IIIa receptor that leads to fibrinogen binding to the platelet and consequent platelet aggregation. Ticlopidine is more effective than aspirin; however, it has the disadvantage of causing diarrhea, skin rash, and, in rare instances, neutropenia and thrombotic thrombocytopenic purpura (TTP). Clopidogrel rarely causes TTP but does not cause neutropenia. The Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events (CAPRIE) trial, which led to FDA approval, found that it was only marginally more effective than aspirin in reducing risk of stroke. The Management of Atherothrombosis with Clopidogrel in High-Risk Patients (MATCH) trial was a large multicenter, randomized double-blind study that compared clopidogrel in combination with aspirin to clopidogrel alone in the secondary prevention of TIA or stroke. The MATCH trial found no difference in TIA or stroke prevention with this combination, but did show a small but significant increase in major bleeding complications (3% versus 1%). In the Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management, and Avoidance (CHARISMA) trial, which included a subgroup of patients with prior stroke or TIA along with other groups at high risk of cardiovascular events, there was no benefit of clopidogrel combined with aspirin compared to aspirin alone. Thus, the use of clopidogrel in combination with aspirin is not recommended for stroke prevention. However, these trials did not enroll patients immediately after the stroke or TIA, and the benefits of combination therapy were greater among those treated earlier, so it is possible that clopidogrel combined with aspirin may be beneficial in this acute period. Ongoing studies are currently addressing this question.

Dipyridamole is an antiplatelet agent that inhibits the uptake of adenosine by a variety of cells, including those of the vascular endothelium. The accumulated adenosine is an inhibitor of aggregation. At least in part through its effects on platelet and vessel wall phosphodiesterases, dipyridamole also potentiates the antiaggregatory effects of prostacyclin and nitric oxide produced by the endothelium and acts by inhibiting platelet phosphodiesterase, which is responsible for the breakdown of cyclic AMP. The resulting elevation in cyclic AMP inhibits aggregation of platelets. Dipyridamole is erratically absorbed depending on stomach pH, but a newer formulation combines timed-release dipyridamole, 200 mg, with aspirin, 25 mg, and has better oral bioavailability. This combination drug was studied in three trials. The European Stroke Prevention Study (ESPS) II showed efficacy of both 50 mg/d of aspirin and extended-release dipyridamole in preventing stroke, and a significantly better risk reduction when the two agents were combined. The ESPRIT (European/Australasian Stroke Prevention in Reversible Ischaemia Trial) trial confirmed the ESPS-II results. This was an open-label, academic trial in which 2,739 patients with stroke or TIA treated with aspirin were randomized to dipyridamole, 200 mg twice daily, or no dipyridamole. Primary outcome was the composite of

death from all vascular causes, nonfatal stroke, nonfatal MI, or major bleeding complication. After 3.5 years of follow-up, 13% of patients on aspirin and dipyridamole and 16% on aspirin alone [hazard ratio 0.80, 95% confidence index (CI) 0.66–0.98] met the primary outcome. In the Prevention Regimen for Effectively Avoiding Second Strokes (PRoFESS) trial, the combination of extended-release dipyridamole and aspirin was compared directly with clopidogrel with and without the angiotensin receptor blocker telmisartan in a study of 20,332 patients. There were no differences in the rates of second stroke (9% each) or degree of disability in patients with median follow-up of 2.4 years. Telmisartan also had no effect on these outcomes. This suggests that these antiplatelet regimens are similar, and also raises questions about default prescription of agents to block the angiotensin pathway in all stroke patients. The principal side effect of dipyridamole is headache. The combination capsule of extended-release dipyridamole and aspirin is approved for prevention of stroke.

Many large clinical trials have demonstrated clearly that most antiplatelet agents reduce the risk of all important vascular atherothrombotic events (i.e., ischemic stroke, MI, and death due to all vascular causes) in patients at risk for these events. The overall *relative* reduction in risk of nonfatal stroke is about 25–30% and of all vascular events is about 25%. The *absolute* reduction varies considerably, depending on the particular patient's risk. Individuals at very low risk for stroke seem to experience the same relative reduction, but their risks may be so low that the "benefit" is meaningless. Conversely, individuals with a 10–15% risk of vascular events per year experience a reduction to about 7.5–11%.

Aspirin is inexpensive, can be given in low doses, and could be recommended for all adults to prevent both stroke and MI. However, it causes epigastric discomfort, gastric ulceration, and gastrointestinal hemorrhage, which may be asymptomatic or life threatening. Consequently, not every 40- or 50-year-old should be advised to take aspirin regularly because the risk of atherothrombotic stroke is extremely low and is outweighed by the risk of adverse side effects. Conversely, every patient who has experienced an atherothrombotic stroke or TIA and has no contraindication should be taking an antiplatelet agent regularly because the average annual risk of another stroke is 8–10%; another few percent will experience an MI or vascular death. Clearly, the likelihood of benefit far outweighs the risks of treatment.

The choice of antiplatelet agent and dose must balance the risk of stroke, the expected benefit, and the risk and cost of treatment. However, there are no definitive data, and opinions vary. Many authorities believe low-dose (30–75 mg/d) and high-dose (650–1300 mg/d) aspirin are about equally effective. Some advocate very low doses to avoid adverse effects, and still others advocate very high doses to be sure the benefit is maximal. Most physicians in North America recommend 81–325 mg/d, while most Europeans recommend 50–100 mg. Clopidogrel or extended-release dipyridamole plus aspirin are being increasingly recommended as first-line drugs for secondary prevention. Similarly, the choice of aspirin, clopidogrel, or dipyridamole plus aspirin must balance the fact that the latter are more effective than aspirin but the cost is higher, and this is likely to affect long-term patient adherence. The use of platelet aggregation studies in individual patients taking aspirin is controversial because of limited data.

ANTICOAGULATION THERAPY AND EMBOLIC STROKE

Several trials have shown that anticoagulation (INR range, 2–3) in patients with chronic nonvalvular (nonrheumatic) atrial fibrillation prevents cerebral embolism and is safe. For primary prevention and for patients who have experienced stroke or TIA, anticoagulation with a VKA reduces the risk by about 67%, which clearly outweighs the 1–3% risk per year of a major bleeding complication. A recent randomized trial compared the new oral thrombin inhibitor dabigatran to VKAs in a noninferiority trial to prevent stroke or systemic embolization in nonvalvular atrial fibrillation. Two doses of dabigatran were used: 110 mg/d and 150 mg/d. Both dose tiers of dabigatran were noninferior to VKAs in preventing second stroke and systemic embolization, and the higher dose tier was superior (relative risk 0.66; 95% CI, 0.53 to 0.82; $P < 0.001$) and the rate of major bleeding was lower in the lower dose tier of dabigatran compared to VKAs. This drug is likely more convenient to take as no blood monitoring is required to titrate the dose and its effect is independent of oral intake of vitamin K. For patients who cannot take anticoagulant medications, clopidogrel plus aspirin was compared to aspirin alone in the Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events (ACTIVE-A). Clopidogrel combined with aspirin was more effective than aspirin alone in preventing vascular events, principally stroke, but increases the risk of major bleeding (relative risk 1.57, $P < 0.001$).

The decision to use anticoagulation for primary prevention is based primarily on risk factors (Table 370-3). The history of a TIA or stroke tips the balance in favor of anticoagulation regardless of other risk factors. Since this risk factor is so important, many clinicians are performing extended ambulatory monitoring to detect intermittent atrial fibrillation in otherwise cryptogenic stroke since its detection would shift toward prescription of oral anticoagulation long term.

Because of the high annual stroke risk in untreated rheumatic heart disease with atrial fibrillation, primary prophylaxis against stroke has not been studied in a double-blind fashion. These patients generally should receive long-term anticoagulation.

Anticoagulation also reduces the risk of embolism in acute MI. Most clinicians recommend a 3-month course of anticoagulation when there is anterior Q-wave infarction, substantial left ventricular dysfunction, congestive heart failure, mural thrombosis, or atrial fibrillation. VKAs are recommended long-term if atrial fibrillation persists.

Stroke secondary to thromboembolism is one of the most serious complications of prosthetic heart valve implantation. The intensity of anticoagulation and/or antiplatelet therapy is dictated by the type of prosthetic valve and its location.

If the embolic source cannot be eliminated, anticoagulation should in most cases be continued indefinitely. Many neurologists recommend combining antiplatelet agents with anticoagulants for patients who "fail" anticoagulation (i.e., have another stroke or TIA).

ANTICOAGULATION THERAPY AND NONCARDIOGENIC STROKE

Data do not support the use of long-term VKAs for preventing atherothrombotic stroke for either intracranial or extracranial cerebrovascular disease. The Warfarin-Aspirin Reinfarction Stroke Study (WARSS) study found no benefit of warfarin sodium (INR, 1.4–2.8) over aspirin, 325 mg, for secondary prevention of stroke but did find a slightly higher bleeding rate in the warfarin group. A recent European study confirmed this finding. The Warfarin-Aspirin Symptomatic Intracranial Disease (WASID) study (see below) demonstrated no benefit of warfarin (INR, 2–3) over aspirin in patients with symptomatic intracranial atherosclerosis, and also found higher bleeding complications.

Treatment: Carotid Atherosclerosis

Carotid atherosclerosis can be removed surgically (endarterectomy) or mitigated with endovascular stenting with or without balloon angioplasty. Anticoagulation has not been directly compared with antiplatelet therapy for carotid disease.

SURGICAL THERAPY

Symptomatic carotid stenosis was studied in the North American Symptomatic Carotid Endarterectomy Trial (NASCET) and the European Carotid Surgery Trial (ECST). Both showed a substantial benefit for surgery in patients with a stenosis of ≥70%. In NASCET, the average cumulative ipsilateral stroke risk at 2 years was 26% for patients treated medically and 9% for those receiving the same medical treatment plus a carotid endarterectomy. This 17% *absolute* reduction in the surgical group is a 65% *relative* risk reduction favoring surgery (Table 370-4). NASCET also showed a significant, although less robust, benefit for patients with 50–70% stenosis. ECST found harm for patients with stenosis <30% treated surgically.

A patient's risk of stroke and possible benefit from surgery are related to the presence of retinal versus hemispheric symptoms, degree of arterial stenosis, extent of associated medical conditions (of note, NASCET and ECST excluded "high-risk" patients with significant cardiac, pulmonary, or renal disease), institutional surgical morbidity and mortality, timing of surgery relative to symptoms, and other factors. A recent meta-analysis of the NASCET and ECST trials demonstrated that endarterectomy is most beneficial when performed within 2 weeks of symptom onset. In addition, benefit is more pronounced in patients >75 years, and men appear to benefit more than women.

In summary, a patient with recent symptomatic hemispheric ischemia, high-grade stenosis in the appropriate internal carotid artery, and an institutional perioperative morbidity and mortality rate of ≤6% generally should undergo carotid endarterectomy. If the perioperative stroke rate is >6% for any particular surgeon, however, the benefits of carotid endarterectomy are questionable.

The indications for surgical treatment of *asymptomatic carotid disease* have been clarified by the results of the Asymptomatic Carotid Atherosclerosis Study (ACAS) and the Asymptomatic Carotid Surgery Trial (ACST). ACAS randomized asymptomatic patients with ≥60% stenosis to medical treatment with aspirin or the same medical treatment plus carotid endarterectomy. The surgical group had a risk over 5 years for ipsilateral stroke (and any perioperative stroke or death) of 5.1%, compared to a risk in the medical group of 11%. While this demonstrates a 53% *relative* risk reduction, the *absolute* risk reduction is only 5.9% over 5 years, or 1.2% annually (Table 370-4). Nearly one-half of the strokes in the surgery group were caused by preoperative angiograms. The recently published ACST randomized 3120 asymptomatic patients with >60% carotid stenosis to endarterectomy or medical therapy. The 5-year risk of stroke in the surgical group (including perioperative stroke or death) was 6.4%, compared to 11.8% in the medically treated group (46% relative risk reduction and 5.4% absolute risk reduction).

In both ACAS and ACST, the perioperative complication rate was higher in women, perhaps negating any benefit in the reduction of stroke risk within 5 years. It is possible that with longer follow-up, a clear benefit in women will emerge. At present, carotid endarterectomy in asymptomatic women remains particularly controversial.

In summary, the natural history of asymptomatic stenosis is a ~2% per year stroke rate, while symptomatic patients experience a 13% per year risk of stroke. Whether to recommend carotid revascularization for an asymptomatic patient is somewhat controversial and depends on many factors, including patient preference, degree of stenosis, age, gender, and comorbidities. Medical therapy for reduction of atherosclerosis risk factors, including cholesterol-lowering agents and antiplatelet medications, is generally recommended for patients with asymptomatic carotid stenosis. As with atrial fibrillation, it is imperative to counsel the patient about TIAs so that therapy can be revised if symptoms develop.

ENDOVASCULAR THERAPY

Balloon angioplasty coupled with stenting is being used with increasing frequency to open stenotic carotid arteries and maintain their patency. These techniques can treat carotid stenosis not only at the bifurcation but also near the skull base and in the intracranial segments. The Stenting and Angioplasty with Protection in Patients at High Risk for Endarterectomy (SAPPHIRE) trial randomized high-risk patients (defined as patients with clinically significant coronary or pulmonary disease, contralateral carotid occlusion, restenosis after endarterectomy, contralateral laryngeal-nerve palsy, prior radical neck surgery or radiation, or age >80) with symptomatic carotid stenosis >50% or asymptomatic stenosis >80% to either stenting combined with a distal emboli-protection device or endarterectomy. The risk of death, stroke, or MI within 30 days and

ipsilateral stroke or death within 1 year was 12.2% in the stenting group and 20.1% in the endarterectomy group (p = .055), suggesting that stenting is at the very least comparable to endarterectomy as a treatment option for this patient group at high risk of surgery. However, the outcomes with both interventions may not have been better than leaving the carotid stenoses untreated, particularly for the asymptomatic patients, and much of the benefit seen in the stenting group was due to a reduction in periprocedure MI. In 2010, the results of two randomized trials comparing stents to endarterectomy in low-risk patients were published. The Carotid Revascularization Endarterectomy versus Stenting Trial (CREST) enrolled 2,502 patients with either asymptomatic or symptomatic stenosis. The 30-day risk of stroke was 4.1% in the stent group and 2.3% in the surgical group, but the 30-day risk of MI was 1.1% in the stent group and 2.3% in the surgery group, suggesting relative equivalence of risk between the procedures. At median follow-up of 2.5 years, the combined endpoint of stroke, MI, and death was the same (7.2% stent versus 6.8% surgery). The International Carotid Stenting (ICSS) trial randomized 1,713 symptomatic patients to stents versus endarterectomy and found a different result: At 120 days, the incidence of stroke, MI, or death was 8.5% in the stenting group versus 5.2% in the endarterectomy group (p = 0.006) and longer term follow-up is currently under way. Differences between trial designs, selection of stent, and operator experience may explain these important differences. Until more data are available on both trials, there remains controversy as to who should receive a stent or have endarterectomy; it is likely that the procedures carry similar risks if performed by experienced physicians.

BYPASS SURGERY

Extracranial-to-intracranial (EC-IC) bypass surgery has been proven ineffective for atherosclerotic stenoses that are inaccessible to conventional carotid endarterectomy. However, a trial is under way to evaluate whether patients with decreased brain perfusion based on positron emission tomography (PET) imaging will benefit from EC-IC bypass.

INTRACRANIAL ATHEROSCLEROSIS

The WASID trial randomized patients with symptomatic stenosis (50–99%) of a major intracranial vessel to either high-dose aspirin (1300 mg/d) or warfarin (target INR, 2.0–3.0), with a combined primary endpoint of ischemic stroke, brain hemorrhage, or death from vascular cause other than stroke. The trial was terminated early because of an increased risk of adverse events related to warfarin anticoagulation. With a mean follow-up of 1.8 years, the primary endpoint was seen in 22.1% in the aspirin group and 21.8% of the warfarin group. Death from any cause was seen in 4.3% of the aspirin group and 9.7% of the warfarin group; 3.2% of patients on aspirin experienced major hemorrhage, compared to 8.3% of patients taking warfarin.

Given the worrisome natural history of symptomatic intracranial atherosclerosis (in the aspirin arm of the WASID trial, 15% of patients experienced a stroke within the first year, despite current standard aggressive medical therapy), some centers treat symptomatic lesions with intracranial angioplasty and stenting. This intervention is currently being compared to aspirin therapy in a prospective, randomized trial. It is unclear whether EC-IC bypass, or other grafting procedures of extracranial blood supply to the pial arteries, are of value in such patients.

Dural Sinus Thrombosis

Limited evidence exists to support short-term usage of anticoagulants, regardless of the presence of intracranial hemorrhage, for venous infarction following sinus thrombosis.

Stroke Syndromes

A careful history and neurologic examination can often localize the region of brain dysfunction; if this region corresponds to a particular arterial distribution, the possible causes responsible for the syndrome can be narrowed. This is of particular importance when the patient presents with a TIA and a normal examination. For example, if a patient develops language loss and a right homonymous hemianopia, a search for causes of left middle cerebral emboli should be performed. A finding of an isolated stenosis of the right internal carotid artery in that patient, for example, suggests an asymptomatic carotid stenosis, and the search for other causes of stroke should continue. The following sections describe the clinical findings of cerebral ischemia associated with cerebral vascular territories depicted in Figs. 370-4, and **370-6** through **370-14**. Stroke syndromes are divided into: (1) large-vessel stroke within the anterior circulation, (2) large-vessel stroke within the posterior circulation, and (3) small-vessel disease of either vascular bed.

Figure 370-6

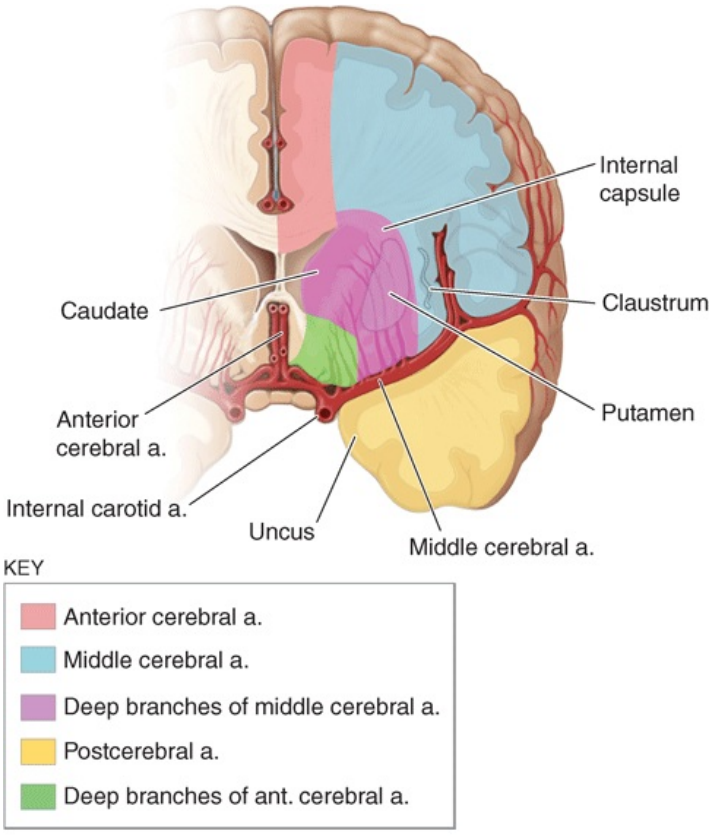

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of a cerebral hemisphere in coronal section showing the territories of the major cerebral vessels that branch from the internal carotid arteries.

Figure 370-7

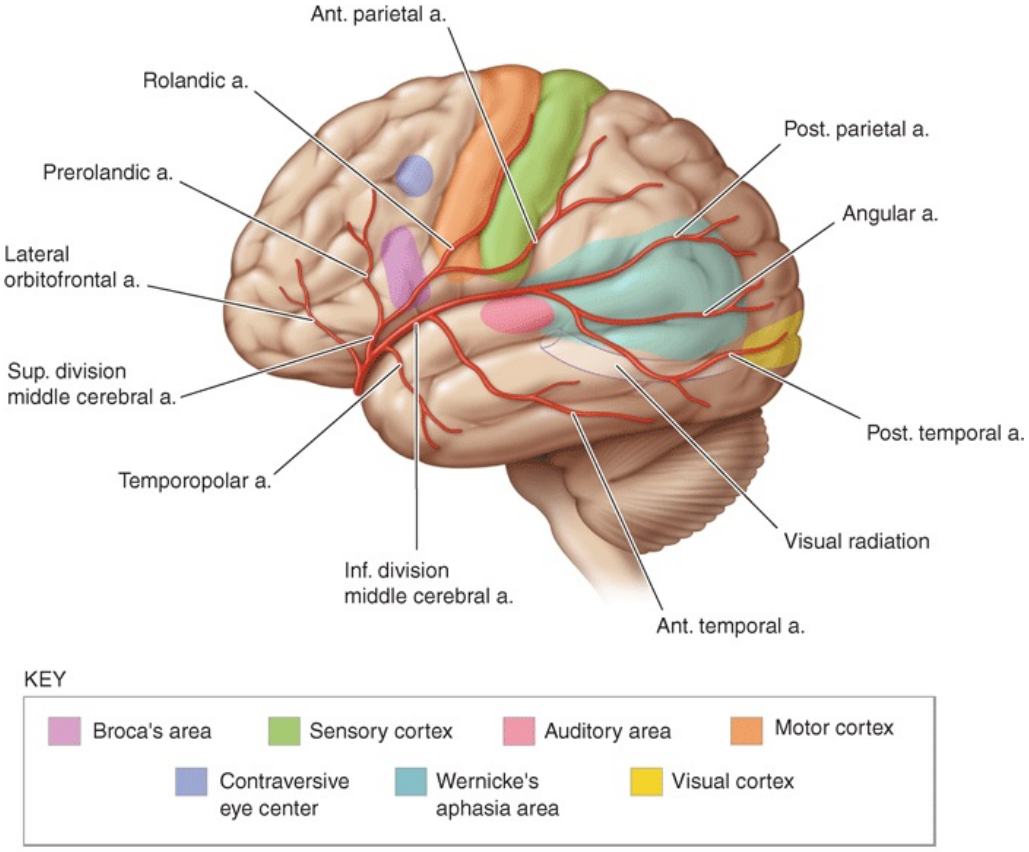

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of a cerebral hemisphere, lateral aspect, showing the branches and distribution of the middle cerebral artery and the principal regions of cerebral localization. Note the bifurcation of the middle cerebral artery into a superior and inferior division.

Signs and symptoms: *Structures involved*

Paralysis of the contralateral face, arm, and leg; sensory impairment over the same area (pinprick, cotton touch, vibration, position, two-point discrimination, stereognosis, tactile localization, barognosis, cutaneographia): *Somatic motor area for face and arm and the fibers descending from the leg area to enter the corona radiata and corresponding somatic sensory system*

Motor aphasia: *Motor speech area of the dominant hemisphere*

Central aphasia, word deafness, anomia, jargon speech, sensory agraphia, acalculia, alexia, finger agnosia, right-left confusion (the last four comprise the Gerstmann syndrome): *Central, suprasylvian speech area and parietooccipital cortex of the dominant hemisphere*

Conduction aphasia: *Central speech area (parietal operculum)*

Apractagnosia of the nondominant hemisphere, anosognosia, hemiasomatognosia, unilateral neglect, agnosia for the left half of external space, dressing "apraxia," constructional "apraxia," distortion of visual coordinates, inaccurate localization in the half field, impaired ability to judge distance, upside-down reading, visual illusions (e.g., it may appear that another person walks through a table): *Nondominant parietal lobe (area corresponding to speech area in dominant hemisphere); loss of topographic memory is usually due to a nondominant lesion, occasionally to a dominant one*

Homonymous hemianopia (often homonymous inferior quadrantanopia): *Optic radiation deep to second temporal convolution*

Paralysis of conjugate gaze to the opposite side: *Frontal contraversive eye field or projecting fibers*

Figure 370-8

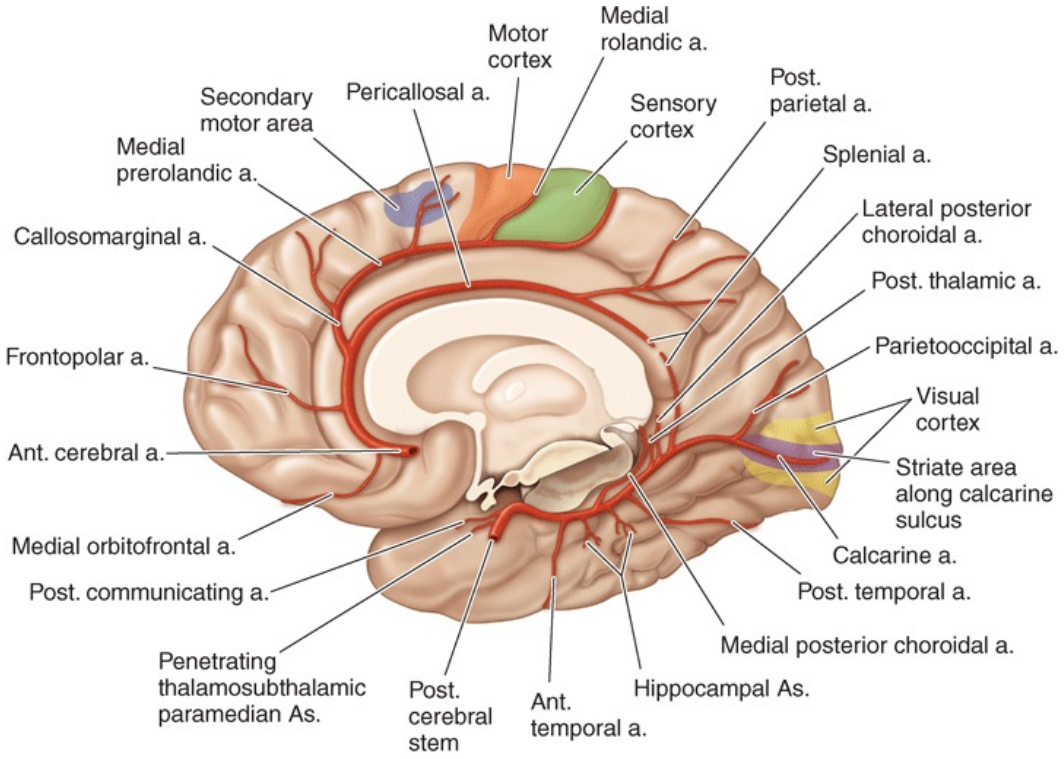

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagram of a cerebral hemisphere, medial aspect, showing the branches and distribution of the anterior cerebral artery and the principal regions of cerebral localization.

Signs and symptoms: *Structures involved*

Paralysis of opposite foot and leg: *Motor leg area*

A lesser degree of paresis of opposite arm: *Arm area of cortex or fibers descending to corona radiata*

Cortical sensory loss over toes, foot, and leg: *Sensory area for foot and leg*

Urinary incontinence: *Sensorimotor area in paracentral lobule*

Contralateral grasp reflex, sucking reflex, gegenhalten (paratonic rigidity): *Medial surface of the posterior frontal lobe; likely supplemental motor area*

Abulia (akinetic mutism), slowness, delay, intermittent interruption, lack of spontaneity, whispering, reflex distraction to sights and sounds: *Uncertain localization—probably cingulate gyrus and medial inferior portion of frontal, parietal, and temporal lobes*

Impairment of gait and stance (gait apraxia): *Frontal cortex near leg motor area*

Dyspraxia of left limbs, tactile aphasia in left limbs: *Corpus callosum*

Figure 370-9

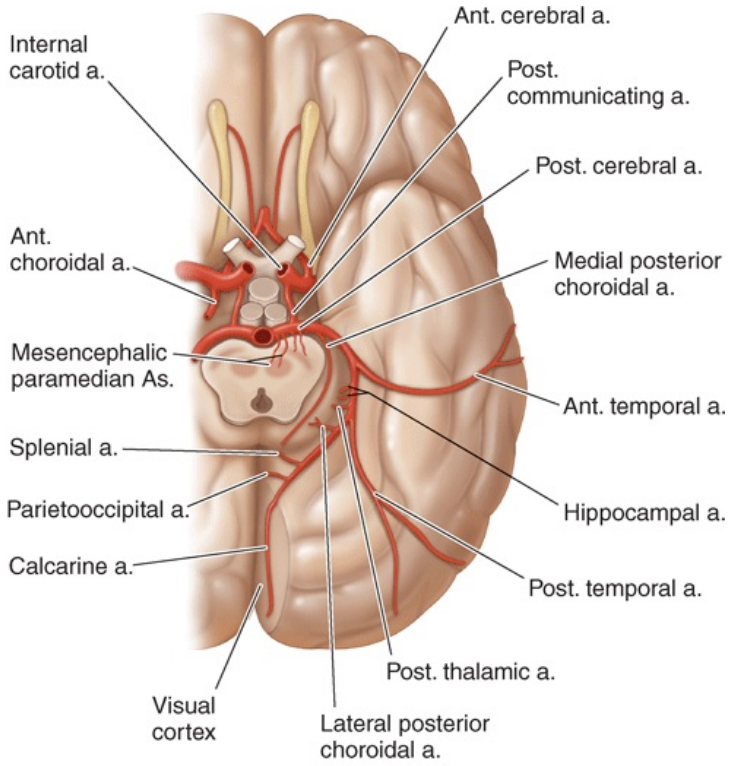

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Inferior aspect of the brain with the branches and distribution of the posterior cerebral artery and the principal anatomic structures shown.

Figure 370-10

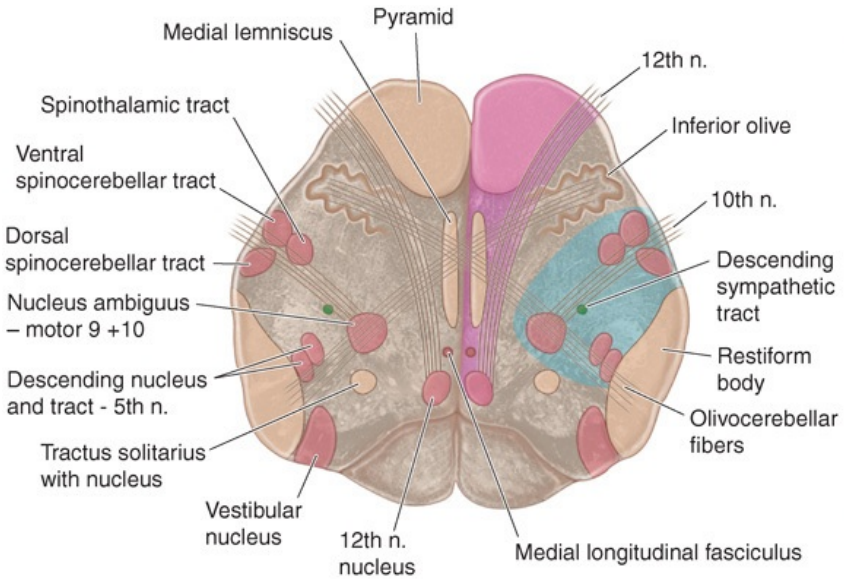

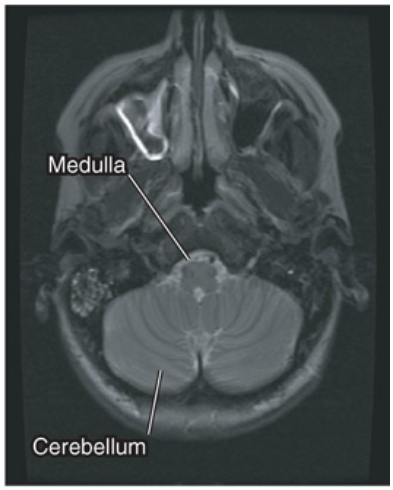

Medullary syndrome:

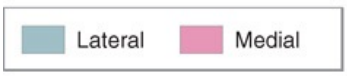

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial section at the level of the medulla, depicted schematically on the left, with a corresponding MR image on the right. Note that in Figs. 370-10 through 370-14, all drawings are oriented with the dorsal surface at the bottom, matching the orientation of the brainstem that is commonly seen in all modern neuroimaging studies. Approximate regions involved in medial and lateral medullary stroke syndromes are shown.

Signs and symptoms: *Structures involved*

I. Medial medullary syndrome (occlusion of vertebral artery or of branch of vertebral or lower basilar artery)

On side of lesion

Paralysis with atrophy of one-half half the tongue: *Ipsilateral twelfth nerve*

On side opposite lesion

Paralysis of arm and leg, sparing face; impaired tactile and proprioceptive sense over one-half the body: *Contralateral pyramidal tract and medial lemniscus*

J. Lateral medullary syndrome (occlusion of any of five vessels may be responsible—vertebral, posterior inferior cerebellar, superior, middle, or inferior lateral medullary arteries)

On side of lesion

Pain, numbness, impaired sensation over one-half the face: *Descending tract and nucleus fifth nerve*

Ataxia of limbs, falling to side of lesion: *Uncertain—restiform body, cerebellar hemisphere, cerebellar fibers, spinocerebellar tract (?)*

Nystagmus, diplopia, oscillopsia, vertigo, nausea, vomiting: *Vestibular nucleus*

Horner's syndrome (miosis, ptosis, decreased sweating): *Descending sympathetic tract*

Dysphagia, hoarseness, paralysis of palate, paralysis of vocal cord, diminished gag reflex: *Issuing fibers ninth and tenth nerves*

Loss of taste: *Nucleus and tractus solitarius*

Numbness of ipsilateral arm, trunk, or leg: *Cuneate and gracile nuclei*

Weakness of lower face: *Genuflected upper motor neuron fibers to ipsilateral facial nucleus*

On side opposite lesion

Impaired pain and thermal sense over half the body, sometimes face: *Spinothalamic tract*

K. Total unilateral medullary syndrome (occlusion of vertebral artery): Combination of medial and lateral syndromes

L. Lateral pontomedullary syndrome (occlusion of vertebral artery): Combination of lateral medullary and lateral inferior pontine syndrome

M. Basilar artery syndrome (the syndrome of the lone vertebral artery is equivalent): A combination of the various brainstem syndromes plus those arising in the posterior cerebral artery distribution.

Bilateral long tract signs (sensory and motor; cerebellar and peripheral cranial nerve abnormalities): *Bilateral long tract; cerebellar and peripheral cranial nerves*

Paralysis or weakness of all extremities, plus all bulbar musculature: *Corticobulbar and corticospinal tracts bilaterally*

Figure 370-11

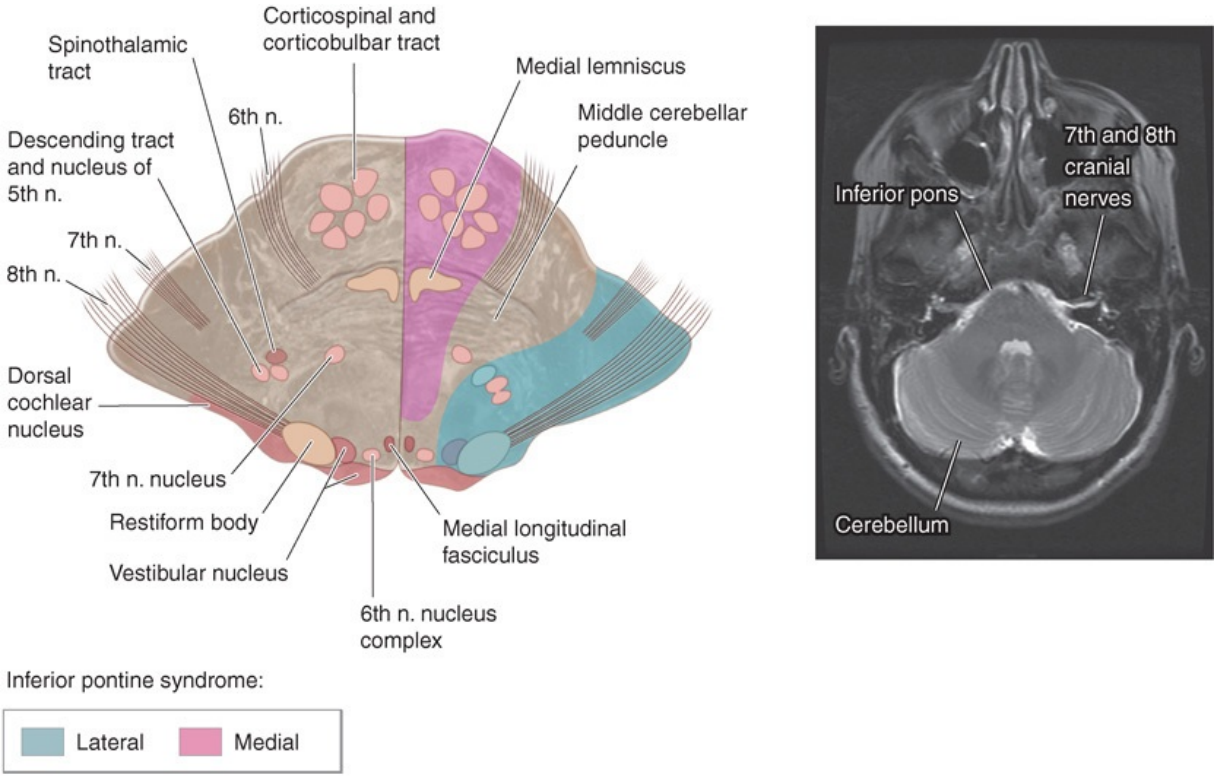

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial section at the level of the inferior pons, depicted schematically on the left, with a corresponding MR image on the right. Approximate regions involved in medial and lateral inferior pontine stroke syndromes are shown.

Signs and symptoms: *Structures involved*

I. Medial inferior pontine syndrome (occlusion of paramedian branch of basilar artery)
   On side of lesion
   Paralysis of conjugate gaze to side of lesion (preservation of convergence): *Center for conjugate lateral gaze*
   Nystagmus: *Vestibular nucleus*
   Ataxia of limbs and gait: Likely *middle cerebellar peduncle*
   Diplopia on lateral gaze: *Abducens nerve*
   On side opposite lesion
   Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract in lower pons*
   Impaired tactile and proprioceptive sense over one-half of the body: *Medial lemniscus*

J. Lateral inferior pontine syndrome (occlusion of anterior inferior cerebellar artery)
   On side of lesion
   Horizontal and vertical nystagmus, vertigo, nausea, vomiting, oscillopsia: *Vestibular nerve or nucleus*
   Facial paralysis: *Seventh nerve*
   Paralysis of conjugate gaze to side of lesion: *Center for conjugate lateral gaze*
   Deafness, tinnitus: *Auditory nerve or cochlear nucleus*
   Ataxia: *Middle cerebellar peduncle and cerebellar hemisphere*
   Impaired sensation over face: *Descending tract and nucleus fifth nerve*
   On side opposite lesion
   Impaired pain and thermal sense over one-half the body (may include face): *Spinothalamic tract*

Figure 370-12

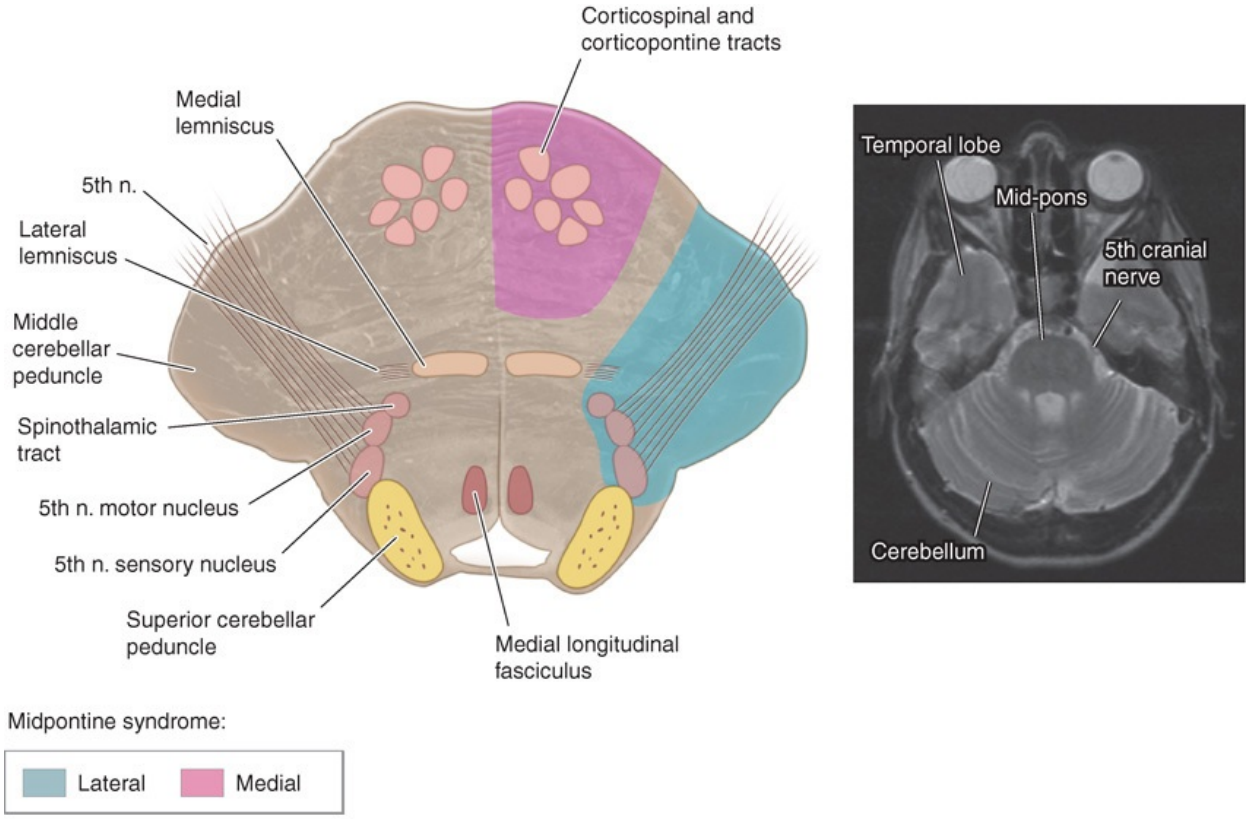

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial section at the level of the midpons, depicted schematically on the left, with a corresponding MR image on the right. Approximate regions involved in medial and lateral midpontine stroke syndromes are shown.

Signs and symptoms: *Structures involved*

I. Medial midpontine syndrome (paramedian branch of midbasilar artery)
   On side of lesion
   Ataxia of limbs and gait (more prominent in bilateral involvement): *Pontine nuclei*
   On side opposite lesion
   Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract*
   Variable impaired touch and proprioception when lesion extends posteriorly: *Medial lemniscus*

J. Lateral midpontine syndrome (short circumferential artery)
   On side of lesion
   Ataxia of limbs: *Middle cerebellar peduncle*
   Paralysis of muscles of mastication: *Motor fibers or nucleus of fifth nerve*
   Impaired sensation over side of face: *Sensory fibers or nucleus of fifth nerve*
   On side opposite lesion
   Impaired pain and thermal sense on limbs and trunk: *Spinothalamic tract*

Figure 370-13

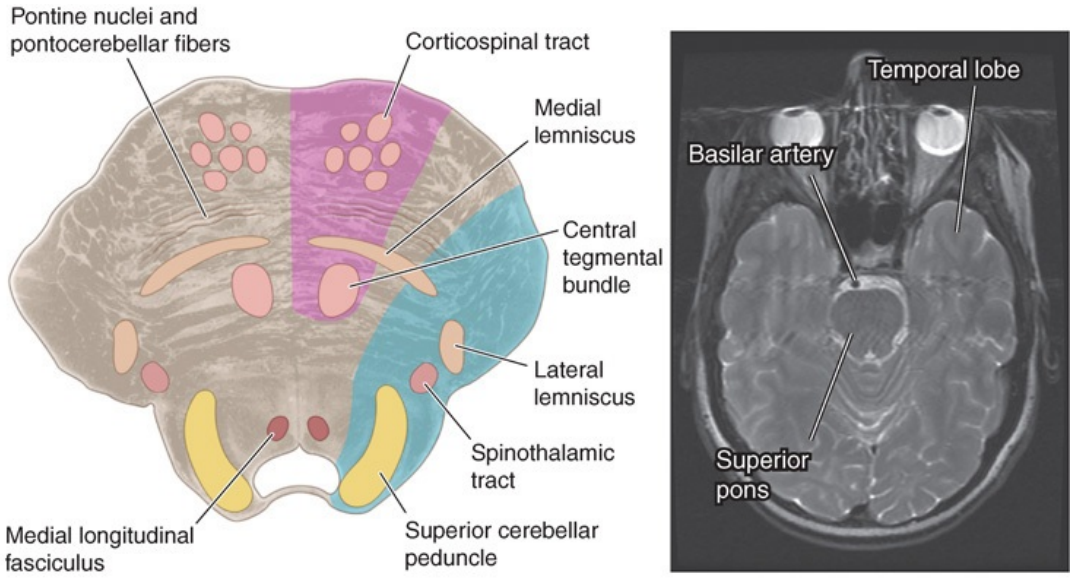

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial section at the level of the superior pons, depicted schematically on the left, with a corresponding MR image on the right. Approximate regions involved in medial and lateral superior pontine stroke syndromes are shown.

Signs and symptoms: *Structures involved*

I. Medial superior pontine syndrome (paramedian branches of upper basilar artery)

On side of lesion

Cerebellar ataxia (probably): *Superior and/or middle cerebellar peduncle*

Internuclear ophthalmoplegia: *Medial longitudinal fasciculus*

Myoclonic syndrome, palate, pharynx, vocal cords, respiratory apparatus, face, oculomotor apparatus, etc.: *Localization uncertain—central tegmental bundle, dentate projection, inferior olivary nucleus*

On side opposite lesion

Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract*

Rarely touch, vibration, and position are affected: *Medial lemniscus*

J. Lateral superior pontine syndrome (syndrome of superior cerebellar artery)

On side of lesion

Ataxia of limbs and gait, falling to side of lesion: *Middle and superior cerebellar peduncles, superior surface of cerebellum, dentate nucleus*

Dizziness, nausea, vomiting; horizontal nystagmus: *Vestibular nucleus*

Paresis of conjugate gaze (ipsilateral): *Pontine contralateral gaze*

Skew deviation: *Uncertain*

Miosis, ptosis, decreased sweating over face (Horner's syndrome): *Descending sympathetic fibers*

Tremor: Localization unclear—*Dentate nucleus, superior cerebellar peduncle*

On side opposite lesion

Impaired pain and thermal sense on face, limbs, and trunk: *Spinothalamic tract*

Impaired touch, vibration, and position sense, more in leg than arm (there is a tendency to incongruity of pain and touch deficits): *Medial lemniscus (lateral portion)*

Figure 370-14

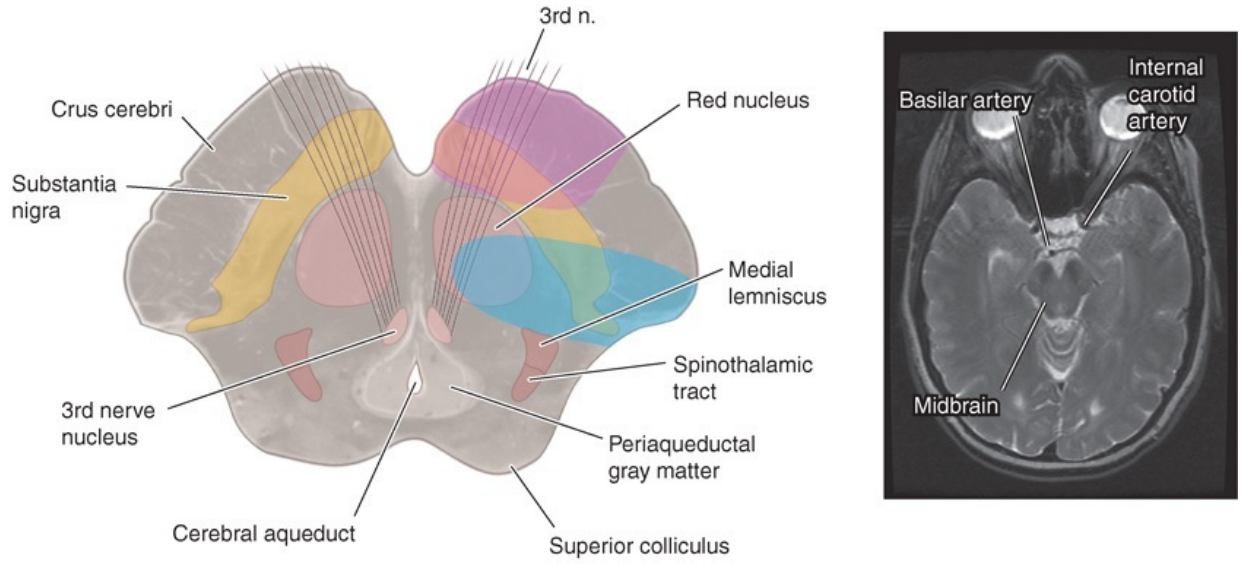

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial section at the level of the midbrain, depicted schematically on the left, with a corresponding MR image on the right. Approximate regions involved in medial and lateral midbrain stroke syndromes are shown.

Signs and symptoms: *Structures involved*

I. Medial midbrain syndrome (paramedian branches of upper basilar and proximal posterior cerebral arteries)

On side of lesion

Eye "down and out" secondary to unopposed action of fourth and sixth cranial nerves, with dilated and unresponsive pupil: *Third nerve fibers*

On side opposite lesion

Paralysis of face, arm, and leg: *Corticobulbar and corticospinal tract descending in crus cerebri*

J. Lateral midbrain syndrome (syndrome of small penetrating arteries arising from posterior cerebral artery)

On side of lesion

Eye "down and out" secondary to unopposed action of fourth and sixth cranial nerves, with dilated and unresponsive pupil: *Third nerve fibers and/or third nerve nucleus*

On side opposite lesion

Hemiataxia, hyperkinesias, tremor: *Red nucleus, dentatorubrothalamic pathway*

STROKE WITHIN THE ANTERIOR CIRCULATION

The internal carotid artery and its branches comprise the anterior circulation of the brain. These vessels can be occluded by intrinsic disease of the vessel (e.g., atherosclerosis or dissection) or by embolic occlusion from a proximal source as discussed above. Occlusion of each major intracranial vessel has distinct clinical manifestations.

Middle Cerebral Artery

Occlusion of the proximal MCA or one of its major branches is most often due to an embolus (artery-to-artery, cardiac, or of unknown source) rather than intracranial atherothrombosis. Atherosclerosis of the proximal MCA may cause distal emboli to the middle cerebral territory or, less commonly, may produce low-flow TIAs. Collateral formation via leptomeningeal vessels often prevents MCA stenosis from becoming symptomatic.

The cortical branches of the MCA supply the lateral surface of the hemisphere except for (1) the frontal pole and a strip along the superomedial border of the frontal and parietal lobes supplied by the ACA, and (2) the lower temporal and occipital pole convolutions supplied by the PCA (Figs. 370-6, 370-7, 370-8, and 370-9).

The proximal MCA (M1 segment) gives rise to penetrating branches (termed *lenticulostriate arteries*) that supply the putamen, outer globus pallidus, posterior limb of the internal capsule, the adjacent corona radiata, and most of the caudate nucleus (Fig. 370-6). In the sylvian fissure, the MCA in most patients

divides into *superior* and *inferior* divisions (M2 branches). Branches of the inferior division supply the inferior parietal and temporal cortex, and those from the superior division supply the frontal and superior parietal cortex (Fig. 370-7).

If the entire MCA is occluded at its origin (blocking both its penetrating and cortical branches) and the distal collaterals are limited, the clinical findings are contralateral hemiplegia, hemianesthesia, homonymous hemianopia, and a day or two of gaze preference to the ipsilateral side. Dysarthria is common because of facial weakness. When the dominant hemisphere is involved, global aphasia is present also, and when the nondominant hemisphere is affected, anosognosia, constructional apraxia, and neglect are found (Chap. 26).

Complete MCA syndromes occur most often when an embolus occludes the stem of the artery. Cortical collateral blood flow and differing arterial configurations are probably responsible for the development of many partial syndromes. Partial syndromes may also be due to emboli that enter the proximal MCA without complete occlusion, occlude distal MCA branches, or fragment and move distally.

Partial syndromes due to embolic occlusion of a single branch include hand, or arm and hand, weakness alone (brachial syndrome) or facial weakness with nonfluent (Broca) aphasia (Chap. 26), with or without arm weakness (frontal opercular syndrome). A combination of sensory disturbance, motor weakness, and nonfluent aphasia suggests that an embolus has occluded the proximal superior division and infarcted large portions of the frontal and parietal cortices (Fig. 370-7). If a fluent (Wernicke's) aphasia occurs without weakness, the inferior division of the MCA supplying the posterior part (temporal cortex) of the dominant hemisphere is probably involved. Jargon speech and an inability to comprehend written and spoken language are prominent features, often accompanied by a contralateral, homonymous superior quadrantanopia. Hemineglect or spatial agnosia without weakness indicates that the inferior division of the MCA in the nondominant hemisphere is involved.

Occlusion of a lenticulostriate vessel produces small-vessel (lacunar) stroke within the internal capsule (Fig. 370-6). This produces pure motor stroke or sensory-motor stroke contralateral to the lesion. Ischemia within the genu of the internal capsule causes primarily facial weakness followed by arm then leg weakness as the ischemia moves posterior within the capsule. Alternatively, the contralateral hand may become ataxic and dysarthria will be prominent (clumsy hand, dysarthria lacunar syndrome). Lacunar infarction affecting the globus pallidus and putamen often has few clinical signs, but parkinsonism and hemiballismus have been reported.

Anterior Cerebral Artery

The ACA is divided into two segments: the precommunal (A1) circle of Willis, or stem, which connects the internal carotid artery to the anterior communicating artery, and the postcommunal (A2) segment distal to the anterior communicating artery (Figs. 370-4, 370-6, and 370-8). The A1 segment gives rise to several deep penetrating branches that supply the anterior limb of the internal capsule, the anterior perforate substance, amygdala, anterior hypothalamus, and the inferior part of the head of the caudate nucleus (Fig. 370-6).

Occlusion of the proximal ACA is usually well tolerated because of collateral flow through the anterior communicating artery and collaterals through the MCA and PCA. Occlusion of a single A2 segment results in the contralateral symptoms noted in Fig. 370-8. If both A2 segments arise from a single anterior cerebral stem (contralateral A1 segment atresia), the occlusion may affect both hemispheres. Profound abulia (a delay in verbal and motor response) and bilateral pyramidal signs with paraparesis or quadriparesis and urinary incontinence result.

Anterior Choroidal Artery

This artery arises from the internal carotid artery and supplies the posterior limb of the internal capsule and the white matter posterolateral to it, through which pass some of the geniculocalcarine fibers (Fig. 370-9). The complete syndrome of anterior choroidal artery occlusion consists of contralateral hemiplegia, hemianesthesia (hypesthesia), and homonymous hemianopia. However, because this territory is also supplied by penetrating vessels of the proximal MCA and the posterior communicating and posterior choroidal arteries, minimal deficits may occur, and patients frequently recover substantially. Anterior choroidal strokes are usually the result of in situ thrombosis of the vessel, and the vessel is particularly vulnerable to iatrogenic occlusion during surgical clipping of aneurysms arising from the internal carotid artery.

Internal Carotid Artery

The clinical picture of internal carotid occlusion varies depending on whether the cause of ischemia is propagated thrombus, embolism, or low flow. The cortex supplied by the MCA territory is affected most often. With a competent circle of Willis, occlusion may go unnoticed. If the thrombus propagates up the internal carotid artery into the MCA or embolizes it, symptoms are identical to proximal MCA occlusion (see above). Sometimes there is massive infarction of the entire deep white matter and cortical surface. When the origins of both the ACA and MCA are occluded at the top of the carotid artery, abulia or stupor occurs with hemiplegia, hemianesthesia, and aphasia or anosognosia. When the PCA arises from the internal carotid artery (a configuration called a *fetal posterior cerebral artery*), it may also become occluded and give rise to symptoms referable to its peripheral territory (Figs. 370-8 and 370-9).

In addition to supplying the ipsilateral brain, the internal carotid artery perfuses the optic nerve and retina via the ophthalmic artery. In ~25% of symptomatic internal carotid disease, recurrent transient monocular blindness (amaurosis fugax) warns of the lesion. Patients typically describe a horizontal shade that sweeps down or up across the field of vision. They may also complain that their vision was blurred in that eye or that

the upper or lower half of vision disappeared. In most cases, these symptoms last only a few minutes. Rarely, ischemia or infarction of the ophthalmic artery or central retinal arteries occurs at the time of cerebral TIA or infarction.

A high-pitched prolonged carotid bruit fading into diastole is often associated with tightly stenotic lesions. As the stenosis grows tighter and flow distal to the stenosis becomes reduced, the bruit becomes fainter and may disappear when occlusion is imminent.

Common Carotid Artery

All symptoms and signs of internal carotid occlusion may also be present with occlusion of the common carotid artery. Jaw claudication may result from low flow in the external carotid branches. Bilateral common carotid artery occlusions at their origin may occur in Takayasu's arteritis (Chap. 326).

STROKE WITHIN THE POSTERIOR CIRCULATION

The posterior circulation is composed of the paired vertebral arteries, the basilar artery, and the paired posterior cerebral arteries. The vertebral arteries join to form the basilar artery at the pontomedullary junction. The basilar artery divides into two posterior cerebral arteries in the interpeduncular fossa (Figs. 370-4, 370-8, and 370-9). These major arteries give rise to long and short circumferential branches and to smaller deep penetrating branches that supply the cerebellum, medulla, pons, midbrain, subthalamus, thalamus, hippocampus, and medial temporal and occipital lobes. Occlusion of each vessel produces its own distinctive syndrome.

Posterior Cerebral Artery

In 75% of cases, both PCAs arise from the bifurcation of the basilar artery; in 20%, one has its origin from the ipsilateral internal carotid artery via the posterior communicating artery; in 5%, both originate from the respective ipsilateral internal carotid arteries (Figs. 370-8 and 370-9). The precommunal, or P1, segment of the true posterior cerebral artery is atretic in such cases.

PCA syndromes usually result from atheroma formation or emboli that lodge at the top of the basilar artery; posterior circulation disease may also be caused by dissection of either vertebral artery and fibromuscular dysplasia.

Two clinical syndromes are commonly observed with occlusion of the PCA: (1) *P1 syndrome*: midbrain, subthalamic, and thalamic signs, which are due to disease of the proximal P1 segment of the PCA or its penetrating branches (thalamogeniculate, Percheron, and posterior choroidal arteries); and (2) *P2 syndrome*: cortical temporal and occipital lobe signs, due to occlusion of the P2 segment distal to the junction of the PCA with the posterior communicating artery.

P1 Syndromes

Infarction usually occurs in the ipsilateral subthalamus and medial thalamus and in the ipsilateral cerebral peduncle and midbrain (Figs. 370-9 and 370-14). A third nerve palsy with contralateral ataxia (Claude's syndrome) or with contralateral hemiplegia (Weber's syndrome) may result. The ataxia indicates involvement of the red nucleus or dentatorubrothalamic tract; the hemiplegia is localized to the cerebral peduncle (Fig. 370-14). If the subthalamic nucleus is involved, contralateral hemiballismus may occur. Occlusion of the artery of Percheron produces paresis of upward gaze and drowsiness, and often abulia. Extensive infarction in the midbrain and subthalamus occurring with bilateral proximal PCA occlusion presents as coma, unreactive pupils, bilateral pyramidal signs, and decerebrate rigidity.

Occlusion of the penetrating branches of thalamic and thalamogeniculate arteries produces less extensive thalamic and thalamocapsular lacunar syndromes. The *thalamic Déjérine-Roussy syndrome* consists of contralateral hemisensory loss followed later by an agonizing, searing or burning pain in the affected areas. It is persistent and responds poorly to analgesics. Anticonvulsants (carbamazepine or gabapentin) or tricyclic antidepressants may be beneficial.

P2 Syndromes

(See also Figs. 370-8 and 370-9.) Occlusion of the distal PCA causes infarction of the medial temporal and occipital lobes. Contralateral homonymous hemianopia with macula sparing is the usual manifestation. Occasionally, only the upper quadrant of visual field is involved. If the visual association areas are spared and only the calcarine cortex is involved, the patient may be aware of visual defects. Medial temporal lobe and hippocampal involvement may cause an acute disturbance in memory, particularly if it occurs in the dominant hemisphere. The defect usually clears because memory has bilateral representation. If the dominant hemisphere is affected and the infarct extends to involve the splenium of the corpus callosum, the patient may demonstrate alexia without agraphia. Visual agnosia for faces, objects, mathematical symbols, and colors and anomia with paraphasic errors (amnestic aphasia) may also occur in this setting, even without callosal involvement. Occlusion of the posterior cerebral artery can produce *peduncular hallucinosis* (visual hallucinations of brightly colored scenes and objects).

Bilateral infarction in the distal PCAs produces cortical blindness (blindness with preserved pupillary light reaction). The patient is often unaware of the blindness or may even deny it (*Anton's syndrome*). Tiny islands of vision may persist, and the patient may report that vision fluctuates as images are captured in the preserved portions. Rarely, only peripheral vision is lost and central vision is spared, resulting in "gun-barrel" vision. Bilateral visual association area lesions may result in *Balint's syndrome*, a disorder of the orderly visual scanning of the environment (Chap. 26), usually resulting from infarctions secondary to low flow in the "watershed" between the distal PCA and MCA territories, as occurs after cardiac arrest. Patients may

experience persistence of a visual image for several minutes despite gazing at another scene (*palinopsia*) or an inability to synthesize the whole of an image (*asimultanagnosia*). Embolic occlusion of the top of the basilar artery can produce any or all of the central or peripheral territory symptoms. The hallmark is the sudden onset of bilateral signs, including ptosis, pupillary asymmetry or lack of reaction to light, and somnolence.

Vertebral and Posterior Inferior Cerebellar Arteries

The vertebral artery, which arises from the innominate artery on the right and the subclavian artery on the left, consists of four segments. The first (V1) extends from its origin to its entrance into the sixth or fifth transverse vertebral foramen. The second segment (V2) traverses the vertebral foramina from C6 to C2. The third (V3) passes through the transverse foramen and circles around the arch of the atlas to pierce the dura at the foramen magnum. The fourth (V4) segment courses upward to join the other vertebral artery to form the basilar artery; only the fourth segment gives rise to branches that supply the brainstem and cerebellum. The posterior inferior cerebellar artery (PICA) in its proximal segment supplies the lateral medulla and, in its distal branches, the inferior surface of the cerebellum.

Atherothrombotic lesions have a predilection for V1 and V4 segments of the vertebral artery. The first segment may become diseased at the origin of the vessel and may produce posterior circulation emboli; collateral flow from the contralateral vertebral artery or the ascending cervical, thyrocervical, or occipital arteries is usually sufficient to prevent low-flow TIAs or stroke. When one vertebral artery is atretic and an atherothrombotic lesion threatens the origin of the other, the collateral circulation, which may also include retrograde flow down the basilar artery, is often insufficient (Figs. 370-4 and 370-9). In this setting, low-flow TIAs may occur, consisting of syncope, vertigo, and alternating hemiplegia; this state also sets the stage for thrombosis. Disease of the distal fourth segment of the vertebral artery can promote thrombus formation manifest as embolism or with propagation as basilar artery thrombosis. Stenosis proximal to the origin of the PICA can threaten the lateral medulla and posterior inferior surface of the cerebellum.

If the subclavian artery is occluded proximal to the origin of the vertebral artery, there is a reversal in the direction of blood flow in the ipsilateral vertebral artery. Exercise of the ipsilateral arm may increase demand on vertebral flow, producing posterior circulation TIAs, or "subclavian steal."

Although atheromatous disease rarely narrows the second and third segments of the vertebral artery, this region is subject to dissection, fibromuscular dysplasia, and, rarely, encroachment by osteophytic spurs within the vertebral foramina.

Embolic occlusion or thrombosis of a V4 segment causes ischemia of the lateral medulla. The constellation of vertigo, numbness of the ipsilateral face and contralateral limbs, diplopia, hoarseness, dysarthria, dysphagia, and ipsilateral Horner's syndrome is called the *lateral medullary (or Wallenberg's) syndrome* (Fig. 370-10). Most cases result from ipsilateral vertebral artery occlusion; in the remainder, PICA occlusion is responsible. Occlusion of the medullary penetrating branches of the vertebral artery or PICA results in partial syndromes. *Hemiparesis is not a feature of vertebral artery occlusion, however, quadriparesis may result from occlusion of the anterior spinal artery.*

Rarely, a *medial medullary syndrome* occurs with infarction of the pyramid and contralateral hemiparesis of the arm and leg, sparing the face. If the medial lemniscus and emerging hypoglossal nerve fibers are involved, contralateral loss of joint position sense and ipsilateral tongue weakness occur.

Cerebellar infarction with edema can lead to *sudden respiratory arrest* due to raised ICP in the posterior fossa. Drowsiness, Babinski signs, dysarthria, and bifacial weakness may be absent, or present only briefly, before respiratory arrest ensues. Gait unsteadiness, headache, dizziness, nausea, and vomiting may be the only early symptoms and signs and should arouse suspicion of this impending complication, which may require neurosurgical decompression, often with an excellent outcome. Separating these symptoms from those of viral labyrinthitis can be a challenge, but headache, neck stiffness, and unilateral dysmetria favor stroke.

Basilar Artery

Branches of the basilar artery supply the base of the pons and superior cerebellum and fall into three groups: (1) paramedian, 7–10 in number, which supply a wedge of pons on either side of the midline; (2) short circumferential, 5–7 in number, that supply the lateral two-thirds of the pons and middle and superior cerebellar peduncles; and (3) bilateral long circumferential (superior cerebellar and anterior inferior cerebellar arteries), which course around the pons to supply the cerebellar hemispheres.

Atheromatous lesions can occur anywhere along the basilar trunk but are most frequent in the proximal basilar and distal vertebral segments. Typically, lesions occlude either the proximal basilar and one or both vertebral arteries. The clinical picture varies depending on the availability of retrograde collateral flow from the posterior communicating arteries. Rarely, dissection of a vertebral artery may involve the basilar artery and, depending on the location of true and false lumen, may produce multiple penetrating artery strokes.

Although atherothrombosis occasionally occludes the distal portion of the basilar artery, emboli from the heart or proximal vertebral or basilar segments are more commonly responsible for "top of the basilar" syndromes.

Because the brainstem contains many structures in close apposition, a diversity of clinical syndromes may emerge with ischemia, reflecting involvement of the corticospinal and corticobulbar tracts, ascending sensory tracts, and cranial nerve nuclei **(Figs. 370-11, 370-12, 370-13, and 370-14)**.

The symptoms of transient ischemia or infarction in the territory of the basilar artery often do not indicate

whether the basilar artery itself or one of its branches is diseased, yet this distinction has important implications for therapy. *The picture of complete basilar occlusion, however, is easy to recognize as a constellation of bilateral long tract signs (sensory and motor) with signs of cranial nerve and cerebellar dysfunction*. A "locked-in" state of preserved consciousness with quadriplegia and cranial nerve signs suggests complete pontine and lower midbrain infarction. The therapeutic goal is to identify *impending* basilar occlusion before devastating infarction occurs. A series of TIAs and a slowly progressive, fluctuating stroke are extremely significant, as they often herald an atherothrombotic occlusion of the distal vertebral or proximal basilar artery.

TIAs in the proximal basilar distribution may produce vertigo (often described by patients as "swimming," "swaying," "moving," "unsteadiness," or "light-headedness"). Other symptoms that warn of basilar thrombosis include diplopia, dysarthria, facial or circumoral numbness, and hemisensory symptoms. In general, symptoms of basilar branch TIAs affect one side of the brainstem, whereas symptoms of basilar artery TIAs usually affect both sides, although a "herald" hemiparesis has been emphasized as an initial symptom of basilar occlusion. Most often TIAs, whether due to impending occlusion of the basilar artery or a basilar branch, are short lived (5–30 minutes) and repetitive, occurring several times a day. The pattern suggests intermittent reduction of flow. Many neurologists treat with heparin to prevent clot propagation.

Atherothrombotic occlusion of the basilar artery with infarction usually causes *bilateral* brainstem signs. A gaze paresis or internuclear ophthalmoplegia associated with ipsilateral hemiparesis may be the only manifestation of bilateral brainstem ischemia. More often, unequivocal signs of bilateral pontine disease are present. Complete basilar thrombosis carries a high mortality.

Occlusion of a branch of the basilar artery usually causes *unilateral* symptoms and signs involving motor, sensory, and cranial nerves. As long as symptoms remain unilateral, concern over pending basilar occlusion should be reduced.

Occlusion of the superior cerebellar artery results in severe ipsilateral cerebellar ataxia, nausea and vomiting, dysarthria, and contralateral loss of pain and temperature sensation over the extremities, body, and face (spino- and trigeminothalamic tract). Partial deafness, ataxic tremor of the ipsilateral upper extremity, Horner's syndrome, and palatal myoclonus may occur rarely. Partial syndromes occur frequently (Fig. 370-13). With large strokes, swelling and mass effects may compress the midbrain or produce hydrocephalus; these symptoms may evolve rapidly. Neurosurgical intervention may be lifesaving in such cases.

Occlusion of the anterior inferior cerebellar artery produces variable degrees of infarction because the size of this artery and the territory it supplies vary inversely with those of the PICA. The principal symptoms include: (1) ipsilateral deafness, facial weakness, vertigo, nausea and vomiting, nystagmus, tinnitus, cerebellar ataxia, Horner's syndrome, and paresis of conjugate lateral gaze; and (2) contralateral loss of pain and temperature sensation. An occlusion close to the origin of the artery may cause corticospinal tract signs (Fig. 370-11).

Occlusion of one of the short circumferential branches of the basilar artery affects the lateral two-thirds of the pons and middle or superior cerebellar peduncle, whereas occlusion of one of the paramedian branches affects a wedge-shaped area on either side of the medial pons (Figs. 370-11 through 370-13).

Imaging Studies

See also Chap. 368.

CT SCANS

CT radiographic images identify or exclude hemorrhage as the cause of stroke, and they identify extraparenchymal hemorrhages, neoplasms, abscesses, and other conditions masquerading as stroke. Brain CT scans obtained in the first several hours after an infarction generally show no abnormality, and the infarct may not be seen reliably for 24–48 hours. CT may fail to show small ischemic strokes in the posterior fossa because of bone artifact; small infarcts on the cortical surface may also be missed.

Contrast-enhanced CT scans add specificity by showing contrast enhancement of subacute infarcts and allow visualization of venous structures. Coupled with newer generation multidetector scanners, CT angiography (CTA) can be performed with administration of IV iodinated contrast allowing visualization of the cervical and intracranial arteries, intracranial veins, aortic arch, and even the coronary arteries in one imaging session. Carotid disease and intracranial vascular occlusions are readily identified with this method (Fig. 370-3). After an IV bolus of contrast, deficits in brain perfusion produced by vascular occlusion can also be demonstrated **(Fig. 370-15)** and used to predict the region of infarcted brain and the brain at risk of further infarction (i.e., the ischemic penumbra, see "Pathophysiology of Ischemic Stroke" above). CT imaging is also sensitive for detecting SAH (though by itself does not rule it out), and CTA can readily identify intracranial aneurysms (Chap. 275). Because of its speed and wide availability, noncontrast head CT is the imaging modality of choice in patients with acute stroke (Fig. 370-1), and CTA and CT perfusion imaging may also be useful and convenient adjuncts.

Figure 370-15

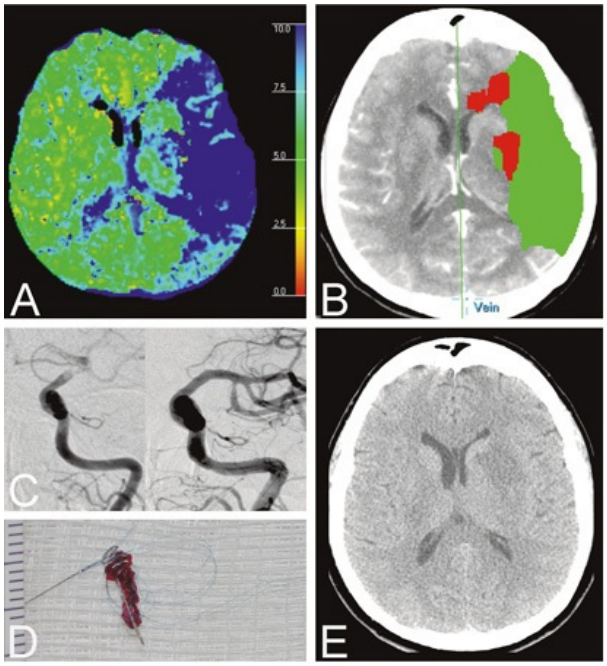

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute left middle cerebral artery (MCA) stroke with right hemiplegia but preserved language.*A.* CT perfusion mean-transit time map showing delayed perfusion of the left MCA distribution (blue). ***B.*** Predicted region of infarct (*red?*) and penumbra (*green*) based on CT perfusion data. *C.* Conventional angiogram showing occlusion of the left internal carotid–MCA bifurcation (*left panel*?), and revascularization of the vessels following successful thrombectomy 8 hours after stroke symptom onset (*right panel*). ***D.*** The clot removed with a thrombectomy device (L5, Concentric Medical, Inc.). ***E.*** CT scan of the brain 2 days later; note infarction in the region predicted in ***B*** but preservation of the penumbral region by successful revascularization.

MRI

MRI reliably documents the extent and location of infarction in all areas of the brain, including the posterior fossa and cortical surface. It also identifies intracranial hemorrhage and other abnormalities but is less sensitive than CT for detecting acute blood. MRI scanners with magnets of higher field strength produce more reliable and precise images. Diffusion-weighted imaging is more sensitive for early brain infarction than standard MR sequences or CT **(Fig. 370-16)**, as is fluid-attenuated inversion recovery (FLAIR) imaging (Chap. 368). Using IV administration of gadolinium contrast, MR perfusion studies can be performed. Brain regions showing poor perfusion but no abnormality on diffusion are equivalent measure of the ischemic penumbra (see "Pathophysiology of Ischemic Stroke," above and Fig. 370-16), and patients showing large regions of mismatch may be better candidates for acute revascularization. MR angiography is highly sensitive for stenosis of extracranial internal carotid arteries and of large intracranial vessels. With higher degrees of stenosis, MR angiography tends to overestimate the degree of stenosis when compared to conventional x-ray angiography. MRI with fat saturation is an imaging sequence used to visualize extra or intracranial arterial dissection. This sensitive technique images clotted blood within the dissected vessel wall.

Figure 370-16

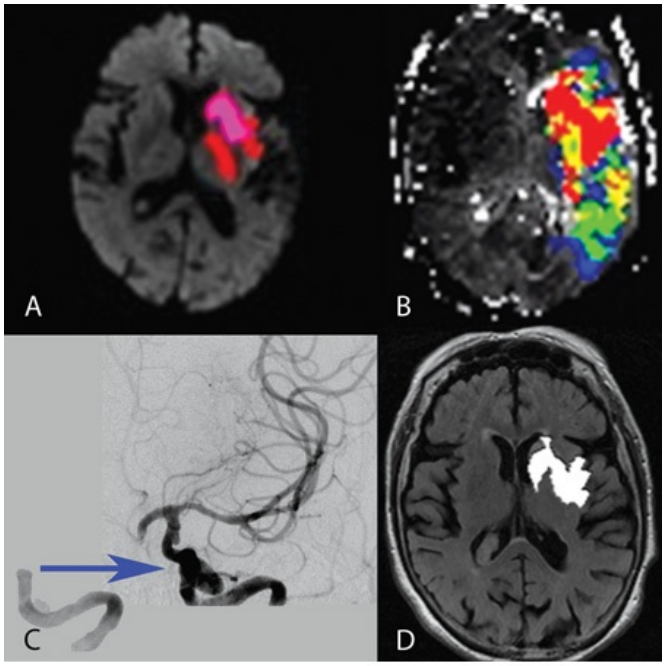

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI of acute stroke.*A.* MRI diffusion-weighted image (DWI) of an 82-year-old woman 2.5 hours after onset of right-sided weakness and aphasia reveals restricted diffusion within the left basal ganglia and internal capsule (*colored regions*.) ***B.*** Perfusion defect within the left hemisphere (*colored signal*) imaged after administration of an IV bolus of gadolinium contrast. The discrepancy between the region of poor perfusion shown in ***B*** and the diffusion deficit shown in ***A*** is called *diffusion-perfusion mismatch* and provides an estimate of the ischemic penumbra. Without specific therapy the region of infarction will expand into much or all of the perfusion deficit. ***C.*** Cerebral angiogram of the left internal carotid artery in this patient before (*left*) and after (*right*) successful endovascular embolectomy. The occlusion is within the carotid terminus. ***D.*** FLAIR image obtained 3 days later showing a region of infarction (*coded as white*) that corresponds to the initial DWI image in ***A***, but not the entire area at risk shown in ***B***, suggesting that successful embolectomy saved a large region of brain tissue from infarction. *(Courtesy of Gregory Albers, MD, Stanford University; with permission.)*

MRI is less sensitive for acute blood products than CT and is more expensive and time consuming and less readily available. Claustrophobia also limits its application. Most acute stroke protocols use CT because of these limitations. However, MRI is useful outside the acute period by more clearly defining the extent of tissue injury and discriminating new from old regions of brain infarction. MRI may have particular utility in patients with TIA: It is also more likely to identify new infarction, which is a strong predictor of subsequent stroke.

CEREBRAL ANGIOGRAPHY

Conventional x-ray cerebral angiography is the gold standard for identifying and quantifying atherosclerotic stenoses of the cerebral arteries and for identifying and characterizing other pathologies, including aneurysms, vasospasm, intraluminal thrombi, fibromuscular dysplasia, arteriovenous fistula, vasculitis, and collateral channels of blood flow. Endovascular techniques, which are evolving rapidly, can be used to deploy stents within delicate intracranial vessels, to perform balloon angioplasty of stenotic lesions, to treat intracranial aneurysms by embolization, and to open occluded vessels in acute stroke with mechanical thrombectomy devices. Randomized trials support use of thrombolytic agents delivered intraarterially in patients with acute MCA stroke by showing that vessels are effectively recanalized and clinical outcomes are improved at 90 days. Cerebral angiography coupled with endovascular techniques for cerebral revascularization are becoming routine in the United States and Europe and likely soon in Japan. Centers capable of these techniques are termed *comprehensive stroke centers* to distinguish them from primary stroke centers that can administer IV rtPA but not perform endovascular therapy. Conventional angiography carries risks of arterial damage, groin hemorrhage, embolic stroke, and renal failure from contrast nephropathy, so it should be reserved for situations where less invasive means are inadequate.

ULTRASOUND TECHNIQUES

Stenosis at the origin of the internal carotid artery can be identified and quantified reliably by ultrasonography that combines a B-mode ultrasound image with a Doppler ultrasound assessment of flow velocity ("duplex" ultrasound). Transcranial Doppler (TCD) assessment of MCA, ACA, and PCA flow and of vertebrobasilar flow is also useful. This latter technique can detect stenotic lesions in the large intracranial arteries because such lesions increase systolic flow velocity. Furthermore, TCD can assist thrombolysis and improve large artery recanalization following rtPA administration; the potential clinical benefit of this treatment is the subject of ongoing study. In many cases, MR angiography combined with carotid and transcranial ultrasound studies eliminates the need for conventional x-ray angiography in evaluating vascular stenosis. Alternatively, CT angiography of the entire head and neck can be performed during the

initial imaging of acute stroke. Because this images the entire arterial system relevant to stroke, with the exception of the heart, much of the clinician's stroke workup can be completed with this single imaging study.

PERFUSION TECHNIQUES

Both xenon techniques (principally xenon-CT) and PET can quantify cerebral blood flow. These tools are generally used for research (Chap. 368) but can be useful for determining the significance of arterial stenosis and planning for revascularization surgery. Single-photon emission computed tomography (SPECT) and MR perfusion techniques report relative cerebral blood flow. Since CT imaging is used as the initial imaging modality for acute stroke, some centers combine both CT angiography and CT perfusion imaging together with the noncontrast CT scan. CT perfusion imaging increases the sensitivity for detecting ischemia, and can measure the ischemic penumbra (Fig. 370-15). Alternatively, MR perfusion can be combined with MR diffusion imaging to identify the ischemic penumbra as the mismatch between these two imaging sequences (Fig. 370-16). The ability to image the ischemic penumbra allows more judicious selection of patients who may or may not benefit from acute interventions such as thrombolysis, thrombectomy, or investigational neuroprotective strategies.

INTRACRANIAL HEMORRHAGE

Hemorrhages are classified by their location and the underlying vascular pathology. Bleeding into subdural and epidural spaces is principally produced by trauma. SAHs are produced by trauma and rupture of intracranial aneurysms (Chap. 275). Intraparenchymal and intraventricular hemorrhage will be considered here.

Diagnosis

Intracranial hemorrhage is often discovered on noncontrast CT imaging of the brain during the acute evaluation of stroke. Since CT is more sensitive than routine MRI for acute blood, CT imaging is the preferred method for acute stroke evaluation (Fig. 370-1). The location of the hemorrhage narrows the differential diagnosis to a few entities. **Table 370-6** lists the causes and anatomic spaces involved in hemorrhages.

Table 370-6 Causes of Intracranial Hemorrhage

Cause	Location	Comments
Head trauma	Intraparenchymal: frontal lobes, anterior temporal lobes; subarachnoid	Coup and contrecoup injury during brain deceleration
Hypertensive hemorrhage	Putamen, globus pallidus, thalamus, cerebellar hemisphere, pons	Chronic hypertension produces hemorrhage from small (~100 μm) vessels in these regions
Transformation of prior ischemic infarction	Basal ganglion, subcortical regions, lobar	Occurs in 1–6% of ischemic strokes with predilection for large hemispheric infarctions
Metastatic brain tumor	Lobar	Lung, choriocarcinoma, melanoma, renal cell carcinoma, thyroid, atrial myxoma
Coagulopathy	Any	Uncommon cause; often associated with prior stroke or underlying vascular anomaly
Drug	Lobar, subarachnoid	Cocaine, amphetamine, phenylpropanolamine
Arteriovenous malformation	Lobar, intraventricular, subarachnoid	Risk is ~2–4% per year for bleeding
Aneurysm	Subarachnoid, intraparenchymal, rarely subdural	Mycotic and nonmycotic forms of aneurysms
Amyloid angiopathy	Lobar	Degenerative disease of intracranial vessels; linkage to Alzheimer's disease, rare in patients <60 years
Cavernous angioma	Intraparenchymal	Multiple cavernous angiomas linked to mutations in KRIT1, CCM2, and PDCD10 genes
Dural arteriovenous fistula	Lobar, subarachnoid	Produces bleeding by venous hypertension
Capillary telangiectasias	Usually brainstem	Rare cause of hemorrhage

Emergency Management

Close attention should be paid to airway management since a reduction in the level of consciousness is common and often progressive. The initial blood pressure should be maintained until the results of the CT scan are reviewed. Expansion of hemorrhage volume is associated with elevated blood pressure, but it remains unclear if lowering of blood pressure reduces hematoma growth. A recent feasibility trial of 60 patients showed that blood pressure could be safely lowered in acute spontaneous intraparenchymal (intracerebral) hemorrhage (ICH) using nicardipine and forms the basis for a planned pivotal trial powered

for detecting improved clinical outcome. Another trial [Intensive Blood Pressure Reduction in Acute Cerebral Hemorrhage Trial (INTERACT)] randomized hypertensive, spontaneous ICH patients to maintain systolic blood pressure (SBP) <180 mm Hg versus SBP <140 mm Hg using IV antihypertensives. There was a statistical decrease in hematoma growth and a reduction in perihematoma edema in the patients assigned to the lower blood pressure goal. Whether these reductions in hematoma growth will translate to clinical benefit is unclear. Until more results are available it is recommended to keep mean arterial pressure (MAP) <130 mm Hg, unless an increase in ICP is suspected. In patients who have ICP monitors in place, current recommendations are to keep the cerebral perfusion pressure (MAP-ICP) above 60 mm Hg (i.e., one should lower MAP to this target if blood pressure is elevated). Blood pressure should be lowered with nonvasodilating IV drugs such as nicardipine, labetalol, or esmolol. Patients with cerebellar hemorrhages or with depressed mental status and radiographic evidence of hydrocephalus should undergo urgent neurosurgical evaluation. Based on the clinical examination and CT findings, further imaging studies may be necessary, including MRI or conventional x-ray angiography. Stuporous or comatose patients generally are treated presumptively for elevated ICP, with tracheal intubation and hyperventilation, mannitol administration, and elevation of the head of the bed while surgical consultation is obtained (Chap. 275).

Intraparenchymal Hemorrhage

ICH is the most common type of intracranial hemorrhage. It accounts for ~10% of all strokes and is associated with a 50% case fatality rate. Incidence rates are particularly high in Asians and blacks. Hypertension, trauma, and cerebral amyloid angiopathy cause the majority of these hemorrhages. Advanced age and heavy alcohol consumption increase the risk, and cocaine and methamphetamine use is one of the most important causes in the young.

HYPERTENSIVE INTRAPARENCHYMAL HEMORRHAGE

Pathophysiology

Hypertensive intraparenchymal hemorrhage (hypertensive hemorrhage or hypertensive intracerebral hemorrhage) usually results from spontaneous rupture of a small penetrating artery deep in the brain. The most common sites are the basal ganglia (especially the putamen), thalamus, cerebellum, and pons. When hemorrhages occur in other brain areas or in nonhypertensive patients, greater consideration should be given to hemorrhagic disorders, neoplasms, vascular malformations, and other causes. The small arteries in these areas seem most prone to hypertension-induced vascular injury. The hemorrhage may be small or a large clot may form and compress adjacent tissue, causing herniation and death. Blood may dissect into the ventricular space, which substantially increases morbidity and may cause hydrocephalus.

Most hypertensive intraparenchymal hemorrhages develop over 30–90 minutes, whereas those associated with anticoagulant therapy may evolve for as long as 24–48 hours. Within 48 hours macrophages begin to phagocytize the hemorrhage at its outer surface. After 1–6 months, the hemorrhage is generally resolved to a slitlike orange cavity lined with glial scar and hemosiderin-laden macrophages.

Clinical Manifestations

Although not particularly associated with exertion, ICHs almost always occur while the patient is awake and sometimes when stressed. The hemorrhage generally presents as the abrupt onset of focal neurologic deficit. Seizures are uncommon. The focal deficit typically worsens steadily over 30–90 minutes and is associated with a diminishing level of consciousness and signs of increased ICP such as headache and vomiting.

The putamen is the most common site for hypertensive hemorrhage, and the adjacent internal capsule is usually damaged **(Fig. 370-17)**. Contralateral hemiparesis is therefore the sentinel sign. When mild, the face sags on one side over 5–30 minutes, speech becomes slurred, the arm and leg gradually weaken, and the eyes deviate away from the side of the hemiparesis. The paralysis may worsen until the affected limbs become flaccid or extend rigidly. When hemorrhages are large, drowsiness gives way to stupor as signs of upper brainstem compression appear. Coma ensues, accompanied by deep, irregular, or intermittent respiration, a dilated and fixed ipsilateral pupil, and decerebrate rigidity. In milder cases, edema in adjacent brain tissue may cause progressive deterioration over 12–72 hours.

Figure 370-17

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Hypertensive hemorrhage. Transaxial noncontrast CT scan through the region of the basal ganglia reveals a hematoma involving the left putamen in a patient with rapidly progressive onset of right hemiparesis.

Thalamic hemorrhages also produce a contralateral hemiplegia or hemiparesis from pressure on, or dissection into, the adjacent internal capsule. A prominent sensory deficit involving all modalities is usually present. Aphasia, often with preserved verbal repetition, may occur after hemorrhage into the dominant thalamus, and constructional apraxia or mutism occurs in some cases of nondominant hemorrhage. There may also be a homonymous visual field defect. Thalamic hemorrhages cause several typical ocular disturbances by virtue of extension inferiorly into the upper midbrain. These include deviation of the eyes downward and inward so that they appear to be looking at the nose, unequal pupils with absence of light reaction, skew deviation with the eye opposite the hemorrhage displaced downward and medially, ipsilateral Horner's syndrome, absence of convergence, paralysis of vertical gaze, and retraction nystagmus. Patients may later develop a chronic, contralateral pain syndrome (Déjérine-Roussy syndrome).

In pontine hemorrhages, deep coma with quadriplegia usually occurs over a few minutes. There is often prominent decerebrate rigidity and "pinpoint" (1 mm) pupils that react to light. There is impairment of reflex horizontal eye movements evoked by head turning (doll's-head or oculocephalic maneuver) or by irrigation of the ears with ice water (Chap. 274). Hyperpnea, severe hypertension, and hyperhidrosis are common. Death often occurs within a few hours, but small hemorrhages are compatible with survival.

Cerebellar hemorrhages usually develop over several hours and are characterized by occipital headache, repeated vomiting, and ataxia of gait. In mild cases there may be no other neurologic signs other than gait ataxia. Dizziness or vertigo may be prominent. There is often paresis of conjugate lateral gaze toward the side of the hemorrhage, forced deviation of the eyes to the opposite side, or an ipsilateral sixth nerve palsy. Less frequent ocular signs include blepharospasm, involuntary closure of one eye, ocular bobbing, and skew deviation. Dysarthria and dysphagia may occur. As the hours pass, the patient often becomes stuporous and then comatose from brainstem compression or obstructive hydrocephalus; immediate surgical evacuation before brainstem compression occurs may be lifesaving. Hydrocephalus from fourth ventricle compression can be relieved by external ventricular drainage, but definitive hematoma evacuation is essential for survival. If the deep cerebellar nuclei are spared, full recovery is common.

LOBAR HEMORRHAGE

Symptoms and signs appear over several minutes. Most lobar hemorrhages are small and cause a restricted clinical syndrome that simulates an embolus to an artery supplying one lobe. For example, the major neurologic deficit with an occipital hemorrhage is hemianopia; with a left temporal hemorrhage, aphasia and delirium; with a parietal hemorrhage, hemisensory loss; and with frontal hemorrhage, arm weakness. Large hemorrhages may be associated with stupor or coma if they compress the thalamus or midbrain. Most patients with lobar hemorrhages have focal headaches, and more than one-half vomit or are drowsy. Stiff neck and seizures are uncommon.

OTHER CAUSES OF INTRACEREBRAL HEMORRHAGE

Cerebral amyloid angiopathy is a disease of the elderly in which arteriolar degeneration occurs and amyloid is deposited in the walls of the cerebral arteries. Amyloid angiopathy causes both single and recurrent lobar hemorrhages and is probably the most common cause of lobar hemorrhage in the elderly. It accounts for some intracranial hemorrhages associated with IV thrombolysis given for MI. This disorder can be suspected

in patients who present with multiple hemorrhages (and infarcts) over several months or years, or in patients with "micro-bleeds" seen on brain MRI sequences sensitive for hemosiderin, but it is definitively diagnosed by pathologic demonstration of Congo red staining of amyloid in cerebral vessels. The ε2 and ε4 allelic variations of the apolipoprotein E gene are associated with increased risk of recurrent lobar hemorrhage and may therefore be markers of amyloid angiopathy. Currently, there is no specific therapy, although antiplatelet and anticoagulating agents are typically avoided.

Cocaine and methamphetamine are frequent causes of stroke in young (age <45 years) patients. ICH, ischemic stroke, and SAH are all associated with stimulant use. Angiographic findings vary from completely normal arteries to large-vessel occlusion or stenosis, vasospasm, or changes consistent with vasculopathy. The mechanism of sympathomimetic-related stroke is not known, but cocaine enhances sympathetic activity causing acute, sometimes severe, hypertension, and this may lead to hemorrhage. Slightly more than one-half of stimulant-related intracranial hemorrhages are intracerebral, and the rest are subarachnoid. In cases of SAH, a saccular aneurysm is usually identified. Presumably, acute hypertension causes aneurysmal rupture.

Head injury often causes intracranial bleeding. The common sites are intracerebral (especially temporal and inferior frontal lobes) and into the subarachnoid, subdural, and epidural spaces. Trauma must be considered in any patient with an unexplained acute neurologic deficit (hemiparesis, stupor, or confusion), particularly if the deficit occurred in the context of a fall (Chap. 378).

Intracranial hemorrhages associated with *anticoagulant therapy* can occur at any location; they are often lobar or subdural. Anticoagulant-related ICHs may evolve slowly, over 24–48 hours. Coagulopathy and thrombocytopenia should be reversed rapidly, as discussed below. ICH associated with *hematologic disorders* (leukemia, aplastic anemia, thrombocytopenic purpura) can occur at any site and may present as multiple ICHs. Skin and mucous membrane bleeding is usually evident and offers a diagnostic clue.

Hemorrhage into a *brain tumor* may be the first manifestation of neoplasm. Choriocarcinoma, malignant melanoma, renal cell carcinoma, and bronchogenic carcinoma are among the most common metastatic tumors associated with ICH. Glioblastoma multiforme in adults and medulloblastoma in children may also have areas of ICH.

Hypertensive encephalopathy is a complication of malignanthypertension. In this acute syndrome, severe hypertension is associated with headache, nausea, vomiting, convulsions, confusion, stupor, and coma. Focal or lateralizing neurologic signs, either transitory or permanent, may occur but are infrequent and therefore suggest some other vascular disease (hemorrhage, embolism, or atherosclerotic thrombosis). There are retinal hemorrhages, exudates, papilledema (hypertensive retinopathy), and evidence of renal and cardiac disease. In most cases ICP and CSF protein levels are elevated. MRI brain imaging shows a pattern of typically posterior (occipital > frontal) brain edema that is reversible and termed *reversible posterior leukoencephalopathy*. The hypertension may be essential or due to chronic renal disease, acute glomerulonephritis, acute toxemia of pregnancy, pheochromocytoma, or other causes. Lowering the blood pressure reverses the process, but stroke can occur, especially if blood pressure is lowered too rapidly. Neuropathologic examination reveals multifocal to diffuse cerebral edema and hemorrhages of various sizes from petechial to massive. Microscopically, there are necrosis of arterioles, minute cerebral infarcts, and hemorrhages. The term *hypertensive encephalopathy* should be reserved for this syndrome and not for chronic recurrent headaches, dizziness, recurrent TIAs, or small strokes that often occur in association with high blood pressure.

Primary intraventricular hemorrhage is rare. It usually begins within the substance of the brain and dissects into the ventricular system without leaving signs of intraparenchymal hemorrhage. Alternatively, bleeding can arise from periependymal veins. Vasculitis, usually polyarteritis nodosa or lupus erythematosus, can produce hemorrhage in any region of the central nervous system; most hemorrhages are associated with hypertension, but the arteritis itself may cause bleeding by disrupting the vessel wall. Nearly one-half of patients with primary intraventricular hemorrhage have identifiable bleeding sources seen using conventional angiography. *Sepsis* can cause small petechial hemorrhages throughout the cerebral white matter. *Moyamoya disease*, mainly an occlusive arterial disease that causes ischemic symptoms, may on occasion produce intraparenchymal hemorrhage, particularly in the young. Hemorrhages into the spinal cord are usually the result of an AVM, cavernous malformation, or metastatic tumor. *Epidural spinal hemorrhage* produces a rapidly evolving syndrome of spinal cord or nerve root compression (Chap. 377). Spinal hemorrhages usually present with sudden back pain and some manifestation of myelopathy.

LABORATORY AND IMAGING EVALUATION

Patients should have routine blood chemistries and hematologic studies. Specific attention to the platelet count and PT/PTT are important to identify coagulopathy. CT imaging reliably detects acute focal hemorrhages in the supratentorial space. Small pontine hemorrhages may not be identified because of motion and bone-induced artifact that obscure structures in the posterior fossa. After the first 2 weeks, x-ray attenuation values of clotted blood diminish until they become isodense with surrounding brain. Mass effect and edema may remain. In some cases, a surrounding rim of contrast enhancement appears after 2–4 weeks and may persist for months. MRI, although more sensitive for delineating posterior fossa lesions, is generally not necessary in most instances. Images of flowing blood on MRI scan may identify AVMs as the cause of the hemorrhage. MRI, CT angiography, and conventional x-ray angiography are used when the cause of intracranial hemorrhage is uncertain, particularly if the patient is young or not hypertensive and the hematoma is not in one of the four usual sites for hypertensive hemorrhage. Postcontrast CT imaging may reveal acute hematoma enhancement signifying bleeding at the time of imaging; this "dot-sign" portends

increased mortality. Some centers routinely perform CT and CT angiography with postcontrast CT imaging in one sitting to rapidly identify any macrovascular etiology of the hemorrhage and provide prognostic information at the same time. Since patients typically have focal neurologic signs and obtundation, and often show signs of increased ICP, a lumbar puncture should be avoided as it may induce cerebral herniation.

Treatment: Intracerebral Hemorrhage

ACUTE MANAGEMENT

Nearly 50% of patients with a hypertensive ICH die, but others have a good to complete recovery if they survive the initial hemorrhage. The ICH scoring system (**Table 370-7**) is a validated metric that is useful for prediction of mortality and clinical outcomes. Any identified coagulopathy should be reversed as soon as possible. For patients taking VKAs, rapid reversal of coagulopathy can be achieved by infusing prothrombin complex concentrates which can be administered quickly, followed by fresh-frozen plasma and vitamin K. When ICH is associated with thrombocytopenia (platelet count <50,000/μL), transfusion of fresh platelets is indicated. The role of urgent platelet inhibition assays in the decision to transfuse platelets remains unclear.

Table 370-7 Prognosis and Clinical Outcomes in Intracerebral Hemorrhage

Clinical or Imaging Factor	Point Score	
Age		
<80 years	0	
≥80 years	1	
Hematoma Volume		
<30 cc	0	
≥30 cc	1	
Intraventricular Hemorrhage Present		
No	0	
Yes	1	
Infratentorial Origin of Hemorrhage		
No	0	
Yes	1	
Glasgow Coma Scale Score		
13–15	0	
5–12	1	
3–4	2	
Total Score	Sum of each category above	
ICH Score Total	**Observed Mortality at 30 days (%)**	**Walk Independently at 12 months (%)**
0	0	70
1	13	60
2	26	33
3	72	3
4	97	8
5	100	None

Although a score of 6 is possible with the scale, no patient was observed to present with this combination of findings, and it is considered highly likely to be fatal.

Abbreviation: ICH, intracerebral hemorrhage.

Sources: JC Hemphill et al: Stroke 32:891, 2001; JC Hemphill et al: Neurology 73:1088, 2009.

At present, little can be done about the hemorrhage itself. Hematomas may expand for several hours following the initial hemorrhage, so treating severe hypertension seems reasonable to prevent hematoma progression. A phase 3 trial of treatment with recombinant factor VIIa reduced hematoma expansion; however, clinical outcomes were not improved, so use of this drug cannot be advocated at present.

Evacuation of supratentorial hematomas does not appear to improve outcome. The International Surgical Trial in Intracerebral Haemorrhage (STICH) randomized 1,033 patients with supratentorial ICH to either early surgical evacuation or initial medical management. No benefit was found in the early surgery arm, although analysis was complicated by the fact that 26% of patients in the initial medical management group ultimately had surgery for neurologic deterioration. Overall, these data do not support routine surgical evacuation of supratentorial hemorrhages; however, many centers operate on patients with progressive neurologic deterioration. Surgical techniques continue to evolve, and minimally invasive endoscopic hematoma evacuation may prove beneficial in future trials.

For cerebellar hemorrhages, a neurosurgeon should be consulted immediately to assist with the evaluation;

most cerebellar hematomas >3 cm in diameter will require surgical evacuation. If the patient is alert without focal brainstem signs and if the hematoma is <1 cm in diameter, surgical removal is usually unnecessary. Patients with hematomas between 1 and 3 cm require careful observation for signs of impaired consciousness and precipitous respiratory failure.

Tissue surrounding hematomas is displaced and compressed but not necessarily infarcted. Hence, in survivors, major improvement commonly occurs as the hematoma is reabsorbed and the adjacent tissue regains its function. Careful management of the patient during the acute phase of the hemorrhage can lead to considerable recovery.

Surprisingly, ICP is often normal even with large intraparenchymal hemorrhages. However, if the hematoma causes marked midline shift of structures with consequent obtundation, coma, or hydrocephalus, osmotic agents coupled with induced hyperventilation can be instituted to lower ICP (Chap. 275). These maneuvers will provide enough time to place a ventriculostomy or ICP monitor. Once ICP is recorded, further hyperventilation and osmotic therapy can be tailored to the individual patient to keep cerebral perfusion pressure (MAP-ICP) above 60 mm Hg. For example, if ICP is found to be high, CSF can be drained from the ventricular space and osmotic therapy continued; persistent or progressive elevation in ICP may prompt surgical evacuation of the clot or withdrawal of support. Alternately, if ICP is normal or only mildly elevated, induced hyperventilation can be reversed and osmotic therapy tapered. Since hyperventilation may actually produce ischemia by cerebral vasoconstriction, induced hyperventilation should be limited to acute resuscitation of the patient with presumptive high ICP and eliminated once other treatments (osmotic therapy or surgical treatments) have been instituted. Glucocorticoids are not helpful for the edema from intracerebral hematoma.

PREVENTION

Hypertension is the leading cause of primary ICH. Prevention is aimed at reducing hypertension, eliminating excessive alcohol use, and discontinuing use of illicit drugs such as cocaine and amphetamines. Patients with amyloid angiopathy should avoid antithrombotic agents.

VASCULAR ANOMALIES

Vascular anomalies can be divided into congenital vascular malformations and acquired vascular lesions.

Congenital Vascular Malformations

True *arteriovenous malformations* (AVMs), venous anomalies, and capillary telangiectasias are lesions that usually remain clinically silent through life. AVMs are probably congenital but cases of acquired lesions have been reported.

True AVMs are congenital shunts between the arterial and venous systems that may present as headache, seizures, and intracranial hemorrhage. AVMs consist of a tangle of abnormal vessels across the cortical surface or deep within the brain substance. AVMs vary in size from a small blemish a few millimeters in diameter to a large mass of tortuous channels composing an arteriovenous shunt of sufficient magnitude to raise cardiac output and precipitate heart failure. Blood vessels forming the tangle interposed between arteries and veins are usually abnormally thin and histologically resemble both arteries and veins. AVMs occur in all parts of the cerebral hemispheres, brainstem, and spinal cord, but the largest ones are most frequently in the posterior half of the hemispheres, commonly forming a wedge-shaped lesion extending from the cortex to the ventricle.

Bleeding, headache, or seizures are most common between the ages of 10 and 30, occasionally as late as the fifties. AVMs are more frequent in men, and rare familial cases have been described. Familial AVM may be a part of the autosomal dominant syndrome of hereditary hemorrhagic telangiectasia (Osler-Rendu-Weber) syndrome due to mutations in endoglin (chromosome 9) or activin receptor-like kinase 1 (chromosome 12).

Headache (without bleeding) may be hemicranial and throbbing, like migraine, or diffuse. Focal seizures, with or without generalization, occur in ~30% of cases. one-half of AVMs become evident as ICHs. In most, the hemorrhage is mainly intraparenchymal with extension into the subarachnoid space in some cases. Blood is usually not deposited in the basal cisterns, and symptomatic cerebral vasospasm is rare. The risk of rerupture is ~2–4% per year and is particularly high in the first few weeks. Hemorrhages may be massive, leading to death, or may be as small as 1 cm in diameter, leading to minor focal symptoms or no deficit. The AVM may be large enough to steal blood away from adjacent normal brain tissue or to increase venous pressure significantly to produce venous ischemia locally and in remote areas of the brain. This is seen most often with large AVMs in the territory of the MCA.

Large AVMs of the anterior circulation may be associated with a systolic and diastolic bruit (sometimes self-audible) over the eye, forehead, or neck and a bounding carotid pulse. Headache at the onset of AVM rupture is not generally as explosive as with aneurysmal rupture. MRI is better than CT for diagnosis, although noncontrast CT scanning sometimes detects calcification of the AVM and contrast may demonstrate the abnormal blood vessels. Once identified, conventional x-ray angiography is the gold standard for evaluating the precise anatomy of the AVM.

Surgical treatment of symptomatic AVMs, often with preoperative embolization to reduce operative bleeding, is usually indicated for accessible lesions. Stereotaxic radiation, an alternative to surgery, can produce a slow sclerosis of the AVM over 2–3 years.

Patients with asymptomatic AVMs have about an ~2–4% per year risk for hemorrhage. Several angiographic

features can be used to help predict future bleeding risk. Paradoxically, smaller lesions seem to have a higher hemorrhage rate. The impact of recurrent hemorrhage on disability is relatively modest, so the indication for surgery in asymptomatic AVMs is debated. A large-scale randomized trial is currently addressing this question.

Venous anomalies are the result of development of anomalous cerebral, cerebellar, or brainstem venous drainage. These structures, unlike AVMs, are functional venous channels. They are of little clinical significance and should be ignored if found incidentally on brain imaging studies. Surgical resection of these anomalies may result in venous infarction and hemorrhage. Venous anomalies may be associated with cavernous malformations (see below), which do carry some bleeding risk. If resection of a cavernous malformation is attempted, the venous anomaly should not be disturbed.

Capillary telangiectasias are true capillary malformations that often form extensive vascular networks through an otherwise normal brain structure. The pons and deep cerebral white matter are typical locations, and these capillary malformations can be seen in patients with hereditary hemorrhagic telangiectasia (Osler-Rendu-Weber) syndrome. If bleeding does occur, it rarely produces mass effect or significant symptoms. No treatment options exist.

Acquired Vascular Lesions

Cavernous angiomas are tufts of capillary sinusoids that form within the deep hemispheric white matter and brainstem with no normal intervening neural structures. The pathogenesis is unclear. Familial cavernous angiomas have been mapped to several different chromosomal loci: KRIT1 (7q21-q22), CCM2 (7p13), and PDCD10 (3q26.1). Both KRIT1 and CCM2 have roles in blood vessel formation while PDCD10 is an apoptotic gene. Cavernous angiomas are typically <1 cm in diameter and are often associated with a venous anomaly. Bleeding is usually of small volume, causing slight mass effect only. The bleeding risk for single cavernous malformations is 0.7–1.5% per year and may be higher for patients with prior clinical hemorrhage or multiple malformations. Seizures may occur if the malformation is located near the cerebral cortex. Surgical resection eliminates bleeding risk and may reduce seizure risk, but it is reserved for those malformations that form near the brain surface. Radiation treatment has not been shown to be of benefit.

Dural arteriovenous fistulas are acquired connections usually from a dural artery to a dural sinus. Patients may complain of a pulse-synchronous cephalic bruit ("pulsatile tinnitus") and headache. Depending on the magnitude of the shunt, venous pressures may rise high enough to cause cortical ischemia or venous hypertension and hemorrhage, particularly subarachnoid hemorrhage. Surgical and endovascular techniques are usually curative. These fistulas may form because of trauma, but most are idiopathic. There is an association between fistulas and dural sinus thrombosis. Fistulas have been observed to appear months to years following venous sinus thrombosis, suggesting that angiogenesis factors elaborated from the thrombotic process may cause these anomalous connections to form. Alternatively, dural arteriovenous fistulas can produce venous sinus occlusion over time, perhaps from the high pressure and high flow through a venous structure.

FURTHER READINGS

Adams HP Jr. et al: Guidelines for the Early Management of Adults with Ischemic Stroke. Stroke 38:1655, 2007[PMID: 17431204] [Full Text]

———: Update to the AHA/ASA recommendations for the prevention of stroke in patients with stroke and transient ischemic attack. Stroke 39:1647, 2008

Albers GW et al: Antithrombotic and thrombolytic therapy for ischemic stroke: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines, 8th ed. Chest 133:630S, 2008

Alberts MJ et al: Recommendations for the establishment of primary stroke centers. Brain Attack Coalition. JAMA 238:3102, 2000

———: Recommendations for comprehensive stroke centers: A consensus statement from the Brain Attack Coalition. Stroke 36:1597, 2005

Choi JH, Mohr JP: Brain arteriovenous malformations in adults. Lancet Neurol 4:299, 2005[PMID: 15847843] [Full Text]

Del Zoppo, GJ et al: Expansion of the time window for treatment of acute ischemic stroke with intravenous tissue plasminogen activator: a science advisory from the American Heart Association/American Stroke Association. Stroke 40:2945, 2009

Easton JD et al: Definition and evaluation of transient ischemic attack. Stroke 40: 2276, 2009[PMID: 19423857] [Full Text]

Goldstein LB et al: Primary prevention of ischemic stroke. Stroke 37:1583, 2006[PMID: 16675728] [Full Text]

Morgenstern LB et al: Guidelines for the management of spontaneous intracerebral hemorrhage. Stroke 42:e23, 2011

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DEMENTIA: INTRODUCTION

Dementia, a syndrome with many causes, affects >4 million Americans and results in a total health care cost of >$100 billion annually. It is defined as an acquired deterioration in cognitive abilities that impairs the successful performance of activities of daily living. Memory is the most common cognitive ability lost with dementia; 10% of persons >70 and 20–40% of individuals >85 have clinically identifiable memory loss. In addition to memory, other mental faculties may be affected; these include language, visuospatial ability, calculation, judgment, and problem solving. Neuropsychiatric and social deficits also arise in many dementia syndromes, resulting in depression, apathy, hallucinations, delusions, agitation, insomnia, and disinhibition. The most common forms of dementia are progressive, but some are static and unchanging or fluctuate from day to day or even minute to minute. Most patients with Alzheimer's disease (AD), the most prevalent form of dementia, begin with memory impairment, although in other dementias, such as frontotemporal dementia, memory loss is not a presenting feature. **Focal cerebral disorders are discussed in Chap. 26 and illustrated in a video library in Chap. e10; memory loss is discussed in Chap. e9.**

FUNCTIONAL ANATOMY OF THE DEMENTIAS

Dementia syndromes result from the disruption of specific large-scale neuronal networks; the location and severity of synaptic and neuronal loss combine to produce the clinical features (Chap. 26). Behavior and mood are modulated by noradrenergic, serotonergic, and dopaminergic pathways, whereas cholinergic signaling is critical for attention and memory functions. The dementias differ in the relative neurotransmitter deficit profiles; accordingly, accurate diagnosis guides effective pharmacotherapy.

AD begins in the transentorhinal region, spreads to the hippocampus, and then moves to lateral and posterior temporal and parietal neocortex, eventually causing a more widespread degeneration. *Vascular dementia* is associated with focal damage in a random patchwork of cortical and subcortical regions or white matter tracts that disconnect nodes within distributed networks. In keeping with the anatomy, AD typically presents with memory loss accompanied later by aphasia or navigational problems. In contrast, patients with dementias that begin in frontal or subcortical regions such as *frontotemporal dementia* (FTD) or *Huntington's disease* (HD) are less likely to begin with memory problems and more likely to have difficulties with judgment, mood, and behavior.

Lesions of cortical-striatal pathways produce specific effects on behavior. The dorsolateral prefrontal cortex bears connections with a central band of the caudate. Lesions of either node or connecting white matter pathways result in poor organization and planning, decreased cognitive flexibility, and impaired working memory. The lateral orbital frontal cortex connects with the ventromedial caudate. Lesions of this system cause impulsiveness, distractibility, and disinhibition. The anterior cingulate cortex projects to the nucleus accumbens, and interruption of these connections produces apathy, poverty of speech, or even akinetic mutism. All corticostriatal systems also include topographically organized projections through the pallidum and thalamus, and damage to these nodes can likewise reproduce the clinical syndrome of corticostriatal damage.

The Causes of Dementia

The single strongest risk factor for dementia is increasing age. The prevalence of disabling memory loss increases with each decade over age 50 and is usually associated with the microscopic changes of AD at autopsy. Yet some centenarians have intact memory function and no evidence of clinically significant dementia. Whether dementia is an inevitable consequence of normal human aging remains controversial.

The many causes of dementia are listed in **Table 371-1**. The frequency of each condition depends on the age group under study, the access of the group to medical care, the country of origin, and perhaps racial or ethnic background. AD is the most common cause of dementia in Western countries, accounting for more than half of all patients. Vascular disease is considered the second most frequent cause for dementia and is particularly common in elderly patients or populations with limited access to medical care, where vascular risk factors are undertreated. Often, vascular disease is mixed with other neurodegenerative disorders, making it difficult, even for the neuropathologist, to estimate the contribution of cerebrovascular disease to the cognitive disorder in an individual patient. Dementias related to Parkinson's disease (PD) are extremely common, and temporally can follow a parkinsonian disorder as seen with PD-related dementia (PDD) or can occur concurrently with or preceding the

motor syndrome as with dementia with Lewy bodies (DLB). In patients under the age of 65, FTD rivals AD as the most common cause of dementia. Chronic intoxications, including those resulting from alcohol and prescription drugs, are an important and often treatable cause of dementia. Other disorders listed in the table are uncommon but important because many are reversible. The classification of dementing illnesses into reversible and irreversible disorders is a useful approach to differential diagnosis. When effective treatments for the neurodegenerative conditions emerge, this dichotomy will become obsolete.

Table 371-1 Differential Diagnosis of Dementia

Most Common Causes of Dementia
Alzheimer's disease
Vascular dementia
Multi-infarct
Diffuse white matter disease (Binswanger's)
Alcoholism[a]
Parkinson's disease
Drug/medication intoxicationa
Less Common Causes of Dementia
Vitamin deficiencies
Thiamine (B1): Wernicke's encephalopathya
'B12 (subacute combined degeneration)[a]
Nicotinic acid (pellagra)[a]
Endocrine and other organ failure
Hypothyroidisma
Adrenal insufficiency and Cushing's syndrome[a]
Hypo- and hyperparathyroidisma
Renal failurea
Liver failurea
Pulmonary failurea
Chronic infections
HIV
Neurosyphilisa
Papovavirus (JC virus)(progressive multifocal leukoencephalopathy)
Tuberculosis, fungal, and protozoala
Whipple's diseasea
Head trauma and diffuse brain damage
Dementia pugilistica
Chronic subdural hematomaa
Postanoxia
Postencephalitis
Normal-pressure hydrocephalusa
Neoplastic
Primary brain tumora
Metastatic brain tumora
Paraneoplastic limbic encephalitis
Toxic disorders
Drug, medication, and narcotic poisoninga
Heavy metal intoxicationa
Dialysis dementia (aluminum)
Organic toxins

Psychiatric
Depression (pseudodementia)a
Schizophreniaa
Conversion reactiona
Degenerative disorders
Huntington's disease
Dementia with Lewy bodies
Progressive supranuclear palsy
Multisystem atrophy
Hereditary ataxias (some forms)
Motor neuron disease [amyotrophic lateral sclerosis (ALS); some forms]
Frontotemporal dementia
Corticobasal degeneration
Multiple sclerosis
Adult Down syndrome with Alzheimer's disease
ALS Parkinson's dementia complex of Guam
Prion (Creutzfeldt-Jakob and Gerstmann-Sträussler-Scheinker diseases)
Miscellaneous
Sarcoidosisa
Vasculitisa
CADASIL, etc.
Acute intermittent porphyriaa
Recurrent nonconvulsive seizuresa
Additional conditions in children or adolescents
Pantothenate kinase-associatedneurodegeneration
Subacute sclerosing panencephalitis
Metabolic disorders (e.g., Wilson's and Leigh's diseases, leukodystrophies, lipid storage diseases, mitochondrial mutations)

[a]Potentially reversible dementia.

Abbreviation: CADASIL, Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy.

In a study of 1000 persons attending a memory disorders clinic, 19% had a potentially reversible cause of the cognitive impairment and 23% had a potentially reversible concomitant condition. The three most common potentially reversible diagnoses were depression, hydrocephalus, and alcohol dependence (Table 371-1).

Subtle cumulative decline in episodic memory is a natural part of aging. This frustrating experience, often the source of jokes and humor, is referred to as *benign forgetfulness of the elderly* . *Benign* means that it is not so progressive or serious that it impairs reasonably successful and productive daily functioning, although the distinction between benign and more significant memory loss can be difficult to make. At age 85, the average person is able to learn and recall approximately one-half the number of items (e.g., words on a list) that he or she could at age 18. A measurable cognitive problem that does not disrupt daily activities is often referred to as *mild cognitive impairment* (MCI). Factors that predict progression from MCI to AD include a prominent memory deficit, family history of dementia, presence of an apolipoprotein ε4 (Apo ε4) allele, small hippocampal volumes, and AD-like signature of cortical atrophy, low cerebrospinal fluid Aβ and elevated tau or positive amyloid imaging with Pittsburgh Compound-B (PiB), although the latter remains an experimental approach not yet available for routine clinical use.

The major degenerative dementias include AD, DLB, FTD and related disorders, HD, and prion diseases, including Creutzfeldt-Jakob disease (CJD). These disorders are all associated with the abnormal aggregation of a specific protein: $A\beta_{42}$ and tau in AD; α-synuclein in DLB; tau, TAR DNA-binding protein of 43kDa (TDP-43), or *fu*sed in *s*arcoma (FUS) in FTD; huntingtin in HD; and misfolded prion protein (PrP^{Sc}) in CJD **(Table 371-2)**.

Table 371-2 The Molecular Basis for Degenerative Dementia

Dementia	Molecular Basis	Causal Genes and (Chromosome)	Susceptibility Genes	Pathology
AD	Aβ	<2% carry these mutations. APP (21), PS-1 (14), PS-2 (1) (most mutations are in PS-1)	*Apo ε4* (19)	Amyloid plaques, neurofibrillary tangles
FTD	Tau	Tau exon and intron mutations (17) (about 10% of familial cases) Progranulin (17) (10% of familial cases)	H1 tau haplotypes	Tau inclusions, Pick bodies, neurofibrillary tangles
	TDP-43 FUS			TDP-43 inclusions FUS inclusions
DLB	α-synuclein	Very rare α-synuclein (4) (dominant)	Unknown	α-synuclein inclusions (Lewy bodies)
CJD	PrP^{SC} proteins	Prion (20) (up to 15% of cases carry these dominant mutations)	Codon 129 homozygosity for methionine or valine	Tau inclusions, spongiform changes, gliosis

Abbreviations: AD, Alzheimer's disease; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FTD, frontotemporal dementia.

Approach to the Patient: Dementias

Three major issues should be kept at the forefront: (1) What is the most accurate diagnosis? (2) Is there a treatable or reversible component to the dementia? (3) Can the physician help to alleviate the burden on caregivers? A broad overview of the approach to dementia is shown in **Table 371-3**. The major degenerative dementias can usually be distinguished by the initial symptoms; neuropsychological, neuropsychiatric, and neurologic findings; and neuroimaging features **(Table 371-4)**.

Table 371-3 Evaluation of the Patient with Dementia

Routine Evaluation	Optional Focused Tests	Occasionally Helpful Tests
History	Psychometric testing	EEG
Physical examination	Chest x-ray	Parathyroid function
Laboratory tests	Lumbar puncture	Adrenal function
Thyroid function (TSH)	Liver function	Urine heavy metals
Vitamin B12	Renal function	RBC sedimentation rate
Complete blood count	Urine toxin screen	Angiogram
Electrolytes	HIV	Brain biopsy
CT/MRI	Apolipoprotein E	SPECT
	RPR or VDRL	PET
Diagnostic Categories		
Reversible Causes	Irreversible/DegenerativeDementias	Psychiatric Disorders
Examples	Examples	Depression
Hypothyroidism	Alzheimer's	Schizophrenia
Thiamine deficiency	Frontotemporal dementia	Conversion reaction
Vitamin B12 deficiency	Huntington's	
Normal-pressure hydrocephalus	Dementia with Lewy bodies	
Subdural hematoma	Vascular	
Chronic infection	Leukoencephalopathies	
Brain tumor	Parkinson's	
Drug intoxication		
	Associated Treatable Conditions	
	Depression	Agitation
	Seizures	Caregiver "burnout"

	Insomnia	Drug side effects

Abbreviations: PET, positron emission tomography; RPR, rapid plasma reagin (test); SPECT, single-photon emission CT; VDRL, Venereal Disease Research Laboratory (test for syphilis).

Table 371-4 Clinical Differentiation of the Major Dementias

Disease	First Symptom	Mental Status	Neuropsychiatry	Neurology	Imaging
AD	Memory loss	Episodic memory loss	Initially normal	Initially normal	Entorhinal cortex and hippocampal atrophy
FTD	Apathy; poor judgment/insight, speech/language; hyperorality	Frontal/executive, language; spares drawing	Apathy, disinhibition, hyperorality, euphoria, depression	May have vertical gaze palsy, axial rigidity, dystonia, alien hand, or MND	Frontal, insular, and/or temporal atrophy; spares posterior parietal lobe
DLB	Visual hallucinations, REM sleep disorder, delirium, Capgras' syndrome, parkinsonism	Drawing and frontal/executive; spares memory; delirium prone	Visual hallucinations, depression, sleep disorder, delusions	Parkinsonism	Posterior parietal atrophy; hippocampi larger than in AD
CJD	Dementia, mood, anxiety, movement disorders	Variable, frontal/executive, focal cortical, memory	Depression, anxiety	Myoclonus, rigidity, parkinsonism	Cortical ribboning and basal ganglia or thalamus hyperintensity on diffusion/flair MRI
Vascular	Often but not always sudden; variable; apathy, falls, focal weakness	Frontal/executive, cognitive slowing; can spare memory	Apathy, delusions, anxiety	Usually motor slowing, spasticity; can be normal	Cortical and/or subcortical infarctions, confluent white matter disease

Abbreviations: AD, Alzheimer's disease; CBD, cortical basal degeneration; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FTD, frontotemporal dementia; MND, motor neuron disease; PSP, progressive supranuclear palsy.

HISTORY

The history should concentrate on the onset, duration, and tempo of progression. An acute or subacute onset of confusion may represent delirium and should trigger the search for intoxication, infection, or metabolic derangement. An elderly person with slowly progressive memory loss over several years is likely to suffer from AD. Nearly 75% of patients with AD begin with memory symptoms, but other early symptoms include difficulty with managing money, driving, shopping, following instructions, finding words, or navigating. A personality change, disinhibition, and weight gain or compulsive eating suggest FTD, not AD. FTD is also suggested by prominent apathy, compulsivity, or progressive loss of speech fluency or word comprehension, and by a relative sparing of memory or visuospatial abilities. The diagnosis of DLB is suggested by early visual hallucinations; parkinsonism; brittle proneness to delirium or sensitivity to psychoactive medications; REM behavior disorder (RBD, the loss of skeletal muscle paralysis during dreaming); or Capgras' syndrome, the delusion that a familiar person has been replaced by an impostor.

A history of stroke with irregular stepwise progression suggests vascular dementia. Vascular dementia is also commonly seen in the setting of hypertension, atrial fibrillation, peripheral vascular disease, and diabetes. In patients suffering from cerebrovascular disease, it can be difficult to determine whether the dementia is due to AD, vascular disease, or a mixture of the two as many of the risk factors for vascular dementia, including diabetes, high cholesterol, elevated homocysteine, and low exercise, are also risk factors for AD. Rapid progression with motor rigidity and myoclonus suggests CJD. Seizures may indicate strokes or neoplasm but also occur in AD, particularly early age of onset AD. Gait disturbance is common in vascular dementia, PD/DLB, or normal-pressure hydrocephalus (NPH). A prior history of high-risk sexual behaviors or intravenous drug use should trigger a search for central nervous system (CNS) infection, especially for HIV or syphilis. A history of recurrent head trauma could indicate chronic subdural hematoma, dementia pugilistica, or NPH. Subacute onset of severe amnesia and psychosis with mesial temporal T2 hyperintensities on MRI should raise concern for paraneoplastic limbic encephalitis, especially in a long-term smoker or other patients at risk for cancer. Related nonparaneoplastic autoimmune conditions can present with a similar tempo and imaging signature. Alcoholism creates risk for malnutrition and thiamine deficiency. Veganism, bowel irradiation,

an autoimmune diathesis, or a remote history of gastric surgery can result in B_{12} deficiency. Certain occupations, such as working in a battery or chemical factory, might indicate heavy metal intoxication. Careful review of medication intake, especially for sedatives and analgesics, may raise the issue of chronic drug intoxication. An autosomal dominant family history is found in HD and in familial forms of AD, FTD, DLB, or prion disorders. The recent death of a loved one, or depressive signs such as insomnia or weight loss, raises the possibility of depression-related cognitive impairments.

PHYSICAL AND NEUROLOGIC EXAMINATION

A thorough general and neurologic examination is essential to document dementia, to look for other signs of nervous system involvement, and to search for clues suggesting a systemic disease that might be responsible for the cognitive disorder. Typical AD does not affect motor systems until later in the course. In contrast, FTD patients often develop axial rigidity, supranuclear gaze palsy, or a motor neuron disease reminiscent of amyotrophic lateral sclerosis (ALS). In DLB, the initial symptoms may include the new onset of a parkinsonian syndrome (resting tremor, cogwheel rigidity, bradykinesia, festinating gait) but often starts with visual hallucinations or dementia. Symptoms referable to the lower brainstem (RBD, gastrointestinal or autonomic problems) may arise years before parkinsonism or dementia. Corticobasal syndrome (CBS) features asymmetric akinesia and rigidity, dystonia, myoclonus, alien limb phenomena, and pyramidal or cortical sensory deficits. Associated cognitive features include nonfluent aphasia with or without motor speech impairment, executive dysfunction, apraxia, or a behavioral disorder. Progressive supranuclear palsy (PSP) is associated with unexplained falls, axial rigidity, dysphagia, and vertical gaze deficits. CJD is suggested by the presence of diffuse rigidity, an akinetic-mute state, and prominent, often startle-sensitive myoclonus.

Hemiparesis or other focal neurologic deficits suggest vascular dementia or brain tumor. Dementia with a myelopathy and peripheral neuropathy suggests vitamin B12 deficiency. Peripheral neuropathy could also indicate another vitamin deficiency, heavy metal intoxication, thyroid dysfunction, Lyme disease, or vasculitis. Dry, cool skin, hair loss, and bradycardia suggest hypothyroidism. Fluctuating confusion associated with repetitive stereotyped movements may indicate ongoing limbic, temporal, or frontal seizures. Hearing impairment or visual loss may produce confusion and disorientation misinterpreted as dementia. Such sensory deficits are common in the elderly but can be a manifestation of mitochondrial disorders.

COGNITIVE AND NEUROPSYCHIATRIC EXAMINATION

Briefscreening tools such as the mini-mental status examination (MMSE) help to confirm the presence of cognitive impairment and to follow the progression of dementia **(Table 371-5)**. The MMSE, a simple 30-point test of cognitive function, contains tests of orientation, working memory (e.g., spell *world* backwards), episodic memory (orientation and 3-word recall), language comprehension, naming, and figure copying. In most patients with MCI and some with clinically apparent AD, the MMSE may be normal and a more rigorous set of neuropsychological tests will be required. When the etiology for the dementia syndrome remains in doubt, a specially tailored evaluation should be performed that includes tasks of working and episodic memory, executive function, language, and visuospatial and perceptual abilities. In AD the early deficits involve episodic memory, category generation ("name as many animals as you can in one minute"), and visuoconstructive ability. Usually deficits in verbal or visual episodic memory are the first neuropsychological abnormalities detected, and tasks that require the patient to recall a long list of words or a series of pictures after a predetermined delay will demonstrate deficits in most patients. In FTD, the earliest deficits on cognitive testing involve executive or language (speech or naming) function. DLB patients have more severe deficits in visuospatial function but do better on episodic memory tasks than patients with AD. Patients with vascular dementia often demonstrate a mixture of executive and visuospatial deficits, with prominent psychomotor slowing. In delirium, the most prominent deficits involve attention, working memory, and executive function, making the assessment of other cognitive domains challenging and often uninformative.

Table 371-5 The Mini-Mental Status Examination

	Points
Orientation	
Name: season/date/day/month/year	5 (1 for each name)
Name: hospital/floor/town/state/country	5 (1 for each name)
Registration	
Identify three objects by name and ask patient to repeat	3 (1 for each object)
Attention and calculation	
Serial 7s; subtract from 100 (e.g., 93–86–79–72–65)	5 (1 for each subtraction)
Recall	
Recall the three objects presented earlier	3 (1 for each object)

Language	
Name pencil and watch	2 (1 for each object)
Repeat "No ifs, ands, or buts"	1
Follow a 3-step command (e.g., "Take this paper, fold it in half, and place it on the table")	3 (1 for each command)
Write "close your eyes" and ask patient to obey written command	1
Ask patient to write a sentence	1
Ask patient to copy a design (e.g., intersecting pentagons)	1
Total	30

A functional assessment should also be performed. The physician should determine the day-to-day impact of the disorder on the patient's memory, community affairs, hobbies, judgment, dressing, and eating. Knowledge of the patient's day-to-day function will help the clinician and the family to organize a therapeutic approach.

Neuropsychiatric assessment is important for diagnosis, prognosis, and treatment. In the early stages of AD, mild depressive features, social withdrawal, and irritability or anxiety are the most prominent psychiatric changes, but patients often maintain core social skills into the middle or late stages, when delusions, agitation, and sleep disturbance may emerge. In FTD, dramatic personality change featuring apathy, overeating, compulsions, disinhibition, euphoria, and loss of empathy are early and common. DLB is associated with visual hallucinations, delusions related to person or place identity, RBD, and excessive daytime sleepiness. Dramatic fluctuations occur not only in cognition but also in primary arousal, such that caregivers may seek emergency room evaluation for suspected stroke. Vascular dementia can present with psychiatric symptoms such as depression, anxiety, delusions, disinhibition, or apathy.

LABORATORY TESTS

The choice of laboratory tests in the evaluation of dementia is complex. The physician must take measures to avoid missing a reversible or treatable cause, yet no single treatable etiology is common; thus, a screen must employ multiple tests, each of which has a low yield. Cost/benefit ratios are difficult to assess, and many laboratory screening algorithms for dementia discourage multiple tests. Nevertheless, even a test with only a 1–2% positive rate is probably worth undertaking if the alternative is missing a treatable cause of dementia. Table 371-3 lists most screening tests for dementia. The American Academy of Neurology recommends the routine measurement of thyroid function, a vitamin B12 level, and a neuroimaging study (CT or MRI).

Neuroimaging studies help to rule out primary and metastatic neoplasms, locate areas of infarction, detect subdural hematomas, and suggest NPH or diffuse white matter disease. They also help to establish a regional pattern of atrophy. Support for the diagnosis of AD includes hippocampal atrophy in addition to posterior-predominant cortical atrophy. Focal frontal and/or anterior temporal atrophy suggests FTD. DLB often features less prominent atrophy, with greater involvement of amygdala than hippocampus. In CJD, MR diffusion-weighted imaging reveals abnormalities in the cortical ribbon and basal ganglia in the majority of patients. Extensive white matter abnormalities correlate with a vascular etiology for dementia. The role of functional-metabolic imaging in the diagnosis of dementia is still under study, although the Federal Drug Administration has approved the use of positron emission tomography (PET) in dementia differential diagnosis. Single-photon emission computed tomography (SPECT) and PET scanning show temporal-parietal hypoperfusion or hypometabolism in AD and frontotemporal deficits in FTD, but these changes often reflect atrophy and can therefore be detected with MRI alone in many patients. Recently, amyloid imaging has shown promise for the diagnosis of AD, and Pittsburgh Compound-B (PiB) and ^{18}F-AV-45 appear to be reliable radioligands for detecting brain amyloid associated with amyloid angiopathy or neuritic plaques **(Fig. 371-1)**. Because these abnormalities can be seen in cognitively normal older persons, however, amyloid imaging may detect preclinical or incidental AD in patients lacking an AD-like dementia syndrome. Once powerful disease-modifying therapies become available, use of these biomarkers may help to identify treatment candidates before irreversible brain injury has occurred. In the meantime, however, the significance of detecting brain amyloid in an asymptomatic elder remains a topic of vigorous investigation. Similarly, MRI perfusion and functional connectivity methods are being explored as potential treatment-monitoring strategies.

Figure 371-1

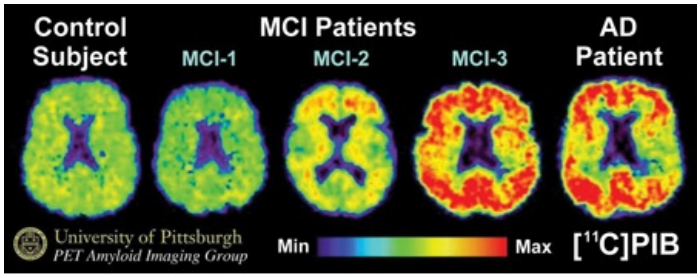

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

PET images obtained with the amyloid-imaging agent Pittsburgh Compound-B ([11C]PIB) in a normal control (*left*); three different patients with mild cognitive impairment (MCI, *center*); and a mild AD patient (*right*). Some MCI patients have control-like levels of amyloid, some have AD-like levels of amyloid, and some have intermediate levels. AD, Alzheimer's disease; MCI, mild cognitive impairment; PET, positron emission tomography.

Lumbar puncture need not be done routinely in the evaluation of dementia, but it is indicated when CNS infection or inflammation are credible diagnostic possibilities. Cerebrospinal fluid (CSF) levels of tau protein and $A\beta_{42}$ show differing patterns with the various dementias; however, the sensitivity and specificity of these measures are not yet sufficiently high to warrant routine use. Formal psychometric testing, although not necessary in every patient with dementia, helps to document the severity of cognitive disturbance, suggest psychogenic causes, and provide a more formal method for following the disease course. Electroencephalogram (EEG) is rarely helpful except to suggest CJD (repetitive bursts of diffuse high-amplitude sharp waves, or "periodic complexes") or an underlying nonconvulsive seizure disorder (epileptiform discharges). Brain biopsy (including meninges) is not advised except to diagnose vasculitis, potentially treatable neoplasms, or unusual infections when the diagnosis is uncertain. Systemic disorders with CNS manifestations, such as sarcoidosis, can usually be confirmed through biopsy of lymph node or solid organ rather than brain. Angiography should be considered when cerebral vasculitis or cerebral venous thrombosis is a possible cause of the dementia.

ALZHEIMER'S DISEASE

Approximately 10% of all persons over the age of 70 have significant memory loss, and in more than half the cause is AD. It is estimated that the annual total cost of caring for a single AD patient in an advanced stage of the disease is >$50,000. The disease also exacts a heavy emotional toll on family members and caregivers. AD can occur in any decade of adulthood, but it is the most common cause of dementia in the elderly. AD most often presents with an insidious onset of memory loss followed by a slowly progressive dementia over several years. Pathologically, atrophy is distributed throughout the medial temporal lobes, as well as lateral and medial parietal lobes and lateral frontal cortex. Microscopically, there are neuritic plaques containing Aβ, neurofibrillary tangles (NFTs) composed of hyperphosphorylated tau filaments, and accumulation of amyloid in blood vessel walls in cortex and leptomeninges (see "Pathogenesis," below). The identification of four different susceptibility genes for AD has provided a foundation for rapid progress in understanding the biologic basis of the disorder.

Clinical Manifestations

The cognitive changes of AD tend to follow a characteristic pattern, beginning with memory impairment and spreading to language and visuospatial deficits. Yet, approximately 20% of patients with AD present with nonmemory complaints such as word-finding, organizational, or navigational difficulty. In the early stages of the disease, the memory loss may go unrecognized or be ascribed to benign forgetfulness. Once the memory loss becomes noticeable to the patient and spouse and falls 1.5 standard deviations below normal on standardized memory tests, the term MCI is applied. This construct provides useful prognostic information, because approximately 50% of patients with MCI (roughly 12% per year) will progress to AD over 4 years. Slowly the cognitive problems begin to interfere with daily activities, such as keeping track of finances, following instructions on the job, driving, shopping, and housekeeping. Some patients are unaware of these difficulties (*anosognosia*), while others remain acutely attuned to their deficits. Changes in environment (such as vacations or hospital stays) may be disorienting, and the patient may become lost on walks or while driving. In the middle stages of AD, the patient is unable to work, is easily lost and confused, and requires daily supervision. Social graces, routine behavior, and superficial conversation may be surprisingly intact. Language becomes impaired—first naming, then comprehension, and finally fluency. In some patients, *aphasia* is an early and prominent feature. Word-finding difficulties and circumlocution may be a problem even when formal testing demonstrates intact naming and fluency. *Apraxia* emerges, and patients have trouble performing learned sequential motor tasks. Visuospatial deficits begin to interfere with dressing, eating, or even walking, and patients fail to solve simple puzzles or copy geometric figures. Simple calculations and clock reading become difficult in parallel.

In the late stages of the disease, some persons remain ambulatory but wander aimlessly. Loss of

judgment and reasoning is inevitable. Delusions are common and usually simple, with common themes of theft, infidelity, or misidentification. Approximately 10% of AD patients develop *Capgras' syndrome*, believing that a caregiver has been replaced by an impostor. In contrast to DLB, where Capgras' syndrome is an early feature, in AD this syndrome emerges later. Loss of inhibitions and aggression may occur and alternate with passivity and withdrawal. Sleep-wake patterns are disrupted, and nighttime wandering becomes disturbing to the household. Some patients develop a shuffling gait with generalized muscle rigidity associated with slowness and awkwardness of movement. Patients often look parkinsonian (Chap. 372) but rarely have a high-amplitude, rhythmic, resting tremor. In end-stage AD, patients become rigid, mute, incontinent, and bedridden. Help is needed with eating, dressing, and toileting. Hyperactive tendon reflexes and myoclonic jerks (sudden brief contractions of various muscles or the whole body) may occur spontaneously or in response to physical or auditory stimulation. Generalized seizures may also occur. Often death results from malnutrition, secondary infections, pulmonary emboli, heart disease, or, most commonly, aspiration. The typical duration of AD is 8–10 years, but the course can range from 1 to 25 years. For unknown reasons, some AD patients show a steady decline in function, while others have prolonged plateaus without major deterioration.

Differential Diagnosis

Early in the disease course, other etiologies of dementia should be excluded (Table 371-1). Neuroimaging studies (CT and MRI) do not show a single specific pattern with AD and may be normal early in the course of the disease. As AD progresses, more distributed but usually posterior-predominant cortical atrophy becomes apparent, along with atrophy of the medial temporal memory structures **(Fig. 371-2*A*, *B*)**. The main purpose of imaging is to exclude other disorders, such as primary and secondary neoplasms, vascular dementia, diffuse white matter disease, and NPH; it also helps to distinguish AD from other degenerative disorders with distinctive imaging patterns such as FTD or CJD. Functional imaging studies in AD reveal hypoperfusion or hypometabolism in the posterior temporal-parietal cortex **(Fig. 371-2*C,D*)**. The EEG in AD is normal or shows nonspecific slowing. Routine spinal fluid examination is also normal. CSF $A\beta_{42}$ levels are reduced, whereas levels of hyperphosphorylated tau protein are elevated, but the considerable overlap of these levels with those of the normal aged population limits the usefulness of these measurements in diagnosis. The use of blood ApoE genotyping is discussed under "Pathology," below. *Slowly progressive decline in memory and orientation, normal results on laboratory tests, and an MRI or CT scan showing only distributed or posteriorly predominant cortical and hippocampal atrophy is highly suggestive of AD.* A clinical diagnosis of AD reached after careful evaluation is confirmed at autopsy about 90% of the time, with misdiagnosed cases usually representing one of the other dementing disorders described later in this chapter, a mixture of AD with vascular pathology, or DLB.

Simple clinical clues are useful in the differential diagnosis. Early prominent gait disturbance with only mild memory loss suggests vascular dementia or, rarely, NPH (see below). Resting tremor with stooped posture, bradykinesia, and masked facies suggest PD (Chap. 372). When dementia occurs after a well-established diagnosis of PD, PDD is usually the correct diagnosis. The early appearance of parkinsonian features in association with fluctuating alertness, visual hallucinations, or delusional misidentification suggests DLB. Chronic alcoholism should prompt the search for vitamin deficiency. Loss of sensibility to position and vibration stimuli accompanied by Babinski responses suggests vitamin B12 deficiency (Chap. 377). Early onset of a focal seizure suggests a metastatic or primary brain neoplasm (Chap. 379). Previous or ongoing depression raises suspicion for depression-related cognitive impairment, although AD can feature a depressive prodrome. A history of treatment for insomnia, anxiety, psychiatric disturbance, or epilepsy suggests chronic drug intoxication. Rapid progression over a few weeks or months associated with rigidity and myoclonus suggests CJD (Chap. 383). Prominent behavioral changes with intact navigation and focal anterior-predominant atrophy on brain imaging are typical of FTD. A positive family history of dementia suggests either one of the familial forms of AD or one of the other genetic disorders associated with dementia, such as HD (see below), FTD (see below), prion disease (Chap. 383), or rare hereditary ataxias (Chap. 373).

Figure 371-1

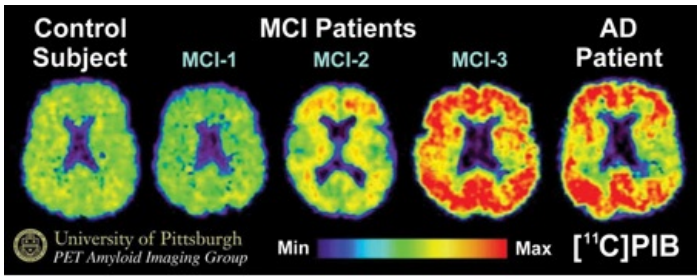

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Alzheimer's disease. Axial T1-weighted MR images through the midbrain of a normal 86-year-old athlete (***A***) and a 77-year-old man with AD (***B***). Note that both individuals have mild sulcal widening and slight dilation of the temporal horns of the lateral ventricles. However, there is a reduction in hippocampal volume in the patient with AD (*arrows*) compared with the volume of the normal-for-age hippocampus (***A***). Fluorodeoxyglucose PET scans of a normal control (***C***) and a patient with AD (***D***). Note that the patient with AD shows decreased glucose metabolism in the posterior temporoparietal regions bilaterally (*arrows*), a typical finding in this condition. AD, Alzheimer's disease; PET, positron emission tomography. *(Images courtesy of TF Budinger, University of California.)*

Epidemiology

The most important risk factors for AD are old age and a positive family history. The frequency of AD increases with each decade of adult life, reaching 20–40% of the population over the age of 85. A positive family history of dementia suggests a genetic cause of AD, although autosomal dominant inheritance occurs in only 2% of patients with AD. Female sex may also be a risk factor independent of the greater longevity of women. Some AD patients have a past history of head trauma with concussion. AD is more common in groups with low educational attainment, but education influences test-taking ability, and it is clear that AD can affect persons of all intellectual levels. One study found that the capacity to express complex written language in early adulthood correlated with a decreased risk for AD. Numerous environmental factors, including aluminum, mercury, and viruses, have been proposed as causes of AD, but none has been demonstrated to play a significant role. Similarly, several studies suggest that the use of nonsteroidal anti-inflammatory agents is associated with a decreased risk of AD, but this has not been confirmed in large prospective studies. Vascular disease, and stroke in particular, seems to lower the threshold for the clinical expression of AD. Also, in many patients with AD, amyloid angiopathy can lead to microhemorrhages, large lobar hemorrhages, or ischemic infarctions. Diabetes increases the risk of AD threefold. Elevated homocysteine and cholesterol levels; hypertension; diminished serum levels of folic acid; low dietary intake of fruits, vegetables, and red wine; and low levels of exercise are all being explored as potential risk factors for AD.

Pathology

At autopsy, the earliest and most severe degeneration is usually found in the medial temporal lobe (entorhinal/perirhinal cortex and hippocampus), lateral temporal cortex, and nucleus basalis of Meynert. The characteristic microscopic findings are neuritic plaques and NFTs. These lesions accumulate in small numbers during normal brain aging but dominate the picture in AD. Increasing evidence suggests that soluble amyloid species called *oligomers* may cause cellular dysfunction and represent the early toxic molecule in AD. Eventually, further amyloid polymerization and fibril formation lead to neuritic plaques **(Fig. 371-3)**, which contain a central amyloid core, proteoglycans, Apo ε4, α-antichymotrypsin, and other proteins. Aβ is a protein of 39–42 amino acids that is derived proteolytically from a larger transmembrane protein, *amyloid precursor protein* (APP), when APP is cleaved by β and γ secretases. The normal function of Aβ is unknown. APP has neurotrophic and neuroprotective properties. The plaque core is surrounded by a halo, which contains dystrophic, tau-immunoreactive neurites and activated microglia. The accumulation of Aβ in cerebral arterioles is termed *amyloid angiopathy*. NFTs are composed of silver-staining neuronal cytoplasmic fibrils composed of abnormally phosphorylated tau (τ) protein; they appear as paired helical filaments by electron microscopy. Tau binds to and stabilizes microtubules, supporting axonal transport of organelles, glycoproteins, neurotransmitters, and other important cargoes throughout the neuron. Once hyperphosphorylated, tau can no longer bind properly to microtubules and its functions are disrupted. Finally, patients with AD often show comorbid DLB and vascular pathology.

Figure 371-3

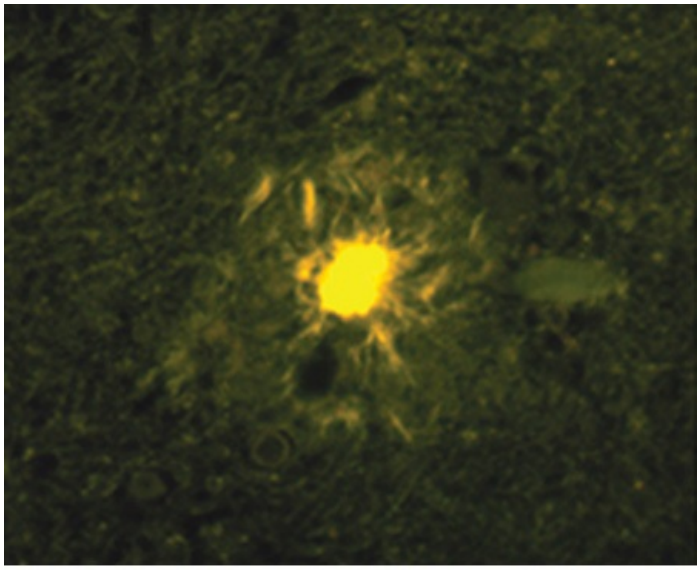

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Mature neuritic plaque with a dense central amyloid core surrounded by dystrophic neurites (thioflavin S stain). *(Image courtesy of S DeArmond, University of California; with permission.)*

Biochemically, AD is associated with a decrease in the cortical levels of several proteins and neurotransmitters, especially acetylcholine, its synthetic enzymecholine acetyltransferase, and nicotinic cholinergic receptors. Reduction of acetylcholine may be related in part to degeneration of cholinergic neurons in the nucleus basalis of Meynert that project throughout the cortex. There is also noradrenergic and serotonergic depletion due to degeneration of brainstem nuclei such as the locus coeruleus and dorsal raphe.

GENETIC CONSIDERATIONS

Several genes play important pathogenic roles in at least some patients with AD. One is the *APP* gene on chromosome 21. Adults with trisomy 21 (Down syndrome) consistently develop the typical neuropathologic hallmarks of AD if they survive beyond age 40. Many develop a progressive dementia superimposed on their baseline mental retardation. APP is a membrane-spanning protein that is subsequently processed into smaller units, including Aβ, which is deposited in neuritic plaques. Aβ peptide results from cleavage of APP by β and γ secretases **(Fig. 371-4)**. Presumably, the extra dose of the *APP* gene on chromosome 21 is the initiating cause of AD in adult Down's syndrome and results in excess cerebral amyloid. Furthermore, a few families with early-onset familial Alzheimer disease (FAD) have been discovered to have point mutations in the *APP* gene. Although very rare, these families were the first examples of single-gene autosomal dominant transmission of AD.

Figure 371-4

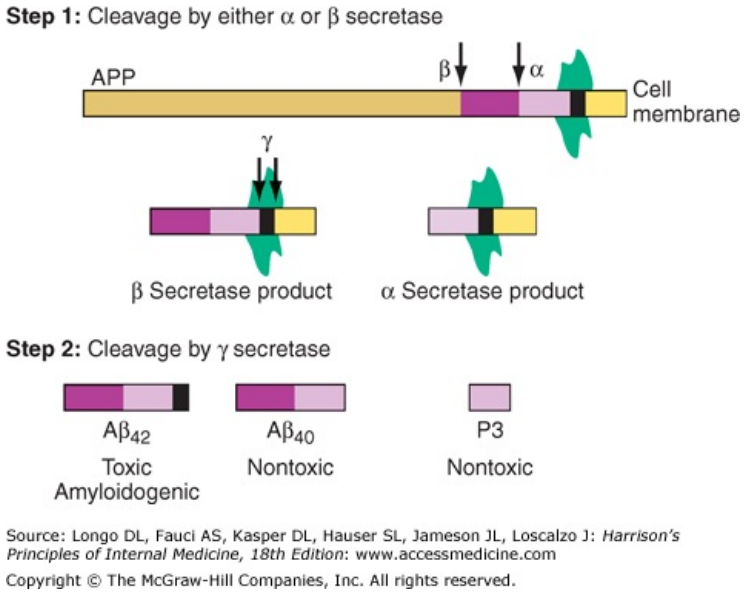

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Amyloid precursor protein (APP) is catabolized by α, β, and γ secretases. A key initial step is the digestion by either β-secretase (BASE) or α secretase [ADAM10 or ADAM17 (TACE)], producing smaller nontoxic products. Cleavage of the β secretase product by γ secretase (Step 2) results in either the toxic $A\beta_{42}$ or the nontoxic $A\beta_{40}$ peptide; cleavage of the α secretase product by γ secretase produces the nontoxic P3 peptide. Excess production of $A\beta_{42}$ is a key initiator of cellular damage in Alzheimer's disease. Current AD research is focused on developing therapies designed to reduce accumulation of $A\beta_{42}$ by antagonizing β or γ secretases, promoting α secretase, or clearing $A\beta_{42}$ that has already formed by use of specific antibodies.

Investigation of large families with multigenerational FAD led to the discovery of two additional AD genes, the *presenilins*. Presenilin-1 (*PS-1*) is on chromosome 14 and encodes a protein called S182. Mutations in this gene cause an early-onset AD (onset before age 60 and often before age 50) transmitted in an autosomal dominant, highly penetrant fashion. More than 100 different mutations have been found in the *PS-1* gene in families from a wide range of ethnic backgrounds. Presenilin-2 (*PS-2*) is on chromosome 1 and encodes a protein called STM2. A mutation in the *PS-2* gene was first found in a group of American families with Volga German ethnic background. Mutations in *PS-1* are much more common than those in *PS-2*. The presenilins are highly homologous and encode similar proteins that at first appeared to have seven transmembrane domains (hence the designation *STM*), but subsequent studies have suggested eight such domains, with a ninth submembrane region. Both S182 and STM2 are cytoplasmic neuronal proteins that are widely expressed throughout the nervous system. They are homologous to a cell-trafficking protein, sel 12, found in the nematode *Caenorhabditis elegans*. Patients with mutations in these genes have elevated plasma levels of $A\beta_{42}$, and *PS-1* mutations in cell cultures produce increased $A\beta_{42}$ in the media. There is evidence that *PS-1* is involved in the cleavage of APP at the gamma secretase site and mutations in either gene (*PS-1* or *APP*) may disturb this function. Mutations in *PS-1* have thus far proved to be the most common cause of early-age-of-onset FAD, representing perhaps 40–70% of this relatively rare syndrome. Mutations in *PS-1* tend to produce AD with an earlier age of onset (mean onset 45 years) and a shorter, more rapidly progressive course (mean duration 6–7 years) than the disease caused by mutations in *PS-2* (mean onset 53 years; duration 11 years). Although some carriers of uncommon *PS-2* mutations have had onset of dementia after the age of 70, mutations in the presenilins are rarely involved in the more common sporadic cases of late-onset AD occurring in the general population. Genetic testing for these uncommon mutations is now commercially available. This diagnostic avenue is likely to be revealing only in early-age-of-onset familial AD and should be performed in the context of formal genetic counseling, especially when there are asymptomatic persons at risk.

A discovery of great importance has been that the *Apo* ε gene on chromosome 19 is involved in the pathogenesis of late- onset familial and sporadic forms of AD. *Apo* ε participates in cholesterol transport (Chap. 356) and has three alleles: ε2, ε3, and ε4. The Apo ε4 allele confers increased risk of AD in the general population, including sporadic and late-age-of-onset familial forms. Approximately 24 –30% of the nondemented white population has at least one ε4 allele (12–15% allele frequency), and about 2% are ε4/4 homozygotes. Among patients with AD, 40–65% have at least one ε4 allele, a highly significant difference compared with controls. Conversely, many AD patients have no ε4 allele, and ε4 carriers may never develop AD. Therefore, ε4 is neither necessary nor sufficient to cause AD. Nevertheless, the Apo ε4 allele, especially in the homozygous state, represents the most important genetic risk factor for AD and acts as a dose-dependent disease modifier, with the earliest age of onset associated with the ε4 homozygosity. Precise mechanisms through which Apo ε4 confers AD risk or hastens onset remain unclear, but ε4 may produce less efficient amyloid clearance. Apo ε can be identified in neuritic plaques and may also be involved in neurofibrillary tangle formation, because it binds to tau protein. Apo ε4 decreases neurite outgrowth in dorsal root ganglion neuronal cultures, perhaps indicating a deleterious role in the brain's response to injury. Some evidence suggests that the ε2 allele may reduce AD risk, but the issue remains to be clarified. Use of Apo ε testing in AD diagnosis remains controversial. It is not indicated as a predictive test in normal persons because its precise predictive value is unclear, and many individuals with the ε4 allele never develop dementia. Many cognitively normal ε4 heterozygotes and homozygotes show decreased cerebral cortical metabolic function with PET, suggesting presymptomatic abnormalities due to AD or an inherited vulnerability of the AD target network. In demented persons who meet clinical criteria for AD, finding an ε4 allele increases the reliability of diagnosis however, the absence of an ε4 allele cannot be considered evidence against AD. Furthermore, all patients with dementia, including those with an ε4 allele, require a search for reversible causes of their cognitive impairment. Nevertheless, Apo ε4 remains the single most important biologic marker associated with AD risk, and studies of ε4's functional role and diagnostic utility are progressing rapidly. The ε4 allele is not associated with risk for FTD, DLB, or CJD, although some evidence suggests that ε4 may exacerbate the phenotype of non-AD degenerative disorders. Additional genes are also likely to be involved in AD, especially as minor risk alleles for sporadic forms of the disease. Recent genome-wide association studies have implicated the clusterin (*CLU)*, phosphatidylinositol-binding clathrin assembly protein *(PICALM)*, and complement component (3b/4b) receptor 1 (*CR1) genes,* and researchers are now working to understand the potential role of these genes in AD pathogenesis. CLU may play a role in synapse turnover, PICALM participates in

clathrin-mediated endocytosis, and CR1 may be involved in amyloid clearance through the complement pathway.

Treatment: Alzheimer's Disease

The management of AD is challenging and gratifying, despite the absence of a cure or a robust pharmacologic treatment. The primary focus is on long-term amelioration of associated behavioral and neurologic problems, as well as providing caregiver support.

Building rapport with the patient, family members, and other caregivers is essential to successful management. In the early stages of AD, memory aids such as notebooks and posted daily reminders can be helpful. Family members should emphasize activities that are pleasant and curtail those that are unpleasant. In other words, practicing skills that have become difficult, such as through memory games and puzzles, will often frustrate and depress the patient without proven benefits. Kitchens, bathrooms, stairways, and bedrooms need to be made safe, and eventually patients must stop driving. Loss of independence and change of environment may worsen confusion, agitation, and anger. Communication and repeated calm reassurance are necessary. Caregiver "burnout" is common, often resulting in nursing home placement of the patient or new health problems for the caregiver, and respite breaks for the caregiver help to maintain a successful long-term therapeutic milieu. Use of adult day care centers can be helpful. Local and national support groups, such as the Alzheimer's Association and the Family Caregiver Alliance, are valuable resources. Internet access to these resources has become available to clinicians and families in recent years.

Donepezil (target dose, 10 mg daily), rivastigmine (target dose, 6 mg twice daily or 9.5-mg patch daily), galantamine (target dose 24 mg daily, extended-release), memantine (target dose, 10 mg twice daily), and tacrine are the drugs presently approved by the Food and Drug Administration (FDA) for treatment of AD. Due to hepatotoxicity, tacrine is no longer used. Dose escalations for each of these medications must be carried out over 4–6 weeks to minimize side effects. The pharmacologic action of donepezil, rivastigmine, and galantamine is inhibition of the cholinesterases, primarily acetylcholinesterase, with a resulting increase in cerebral acetylcholine levels. Memantine appears to act by blocking overexcited *N*-methyl-d-aspartate (NMDA) glutamate receptors. Double-blind, placebo-controlled, crossover studies with cholinesterase inhibitors and memantine have shown them to be associated with improved caregiver ratings of patients' functioning and with an apparent decreased rate of decline in cognitive test scores over periods of up to 3 years. The average patient on an anticholinesterase compound maintains his or her MMSE score for close to a year, whereas a placebo-treated patient declines 2–3 points over the same time period. Memantine, used in conjunction with cholinesterase inhibitors or by itself, slows cognitive deterioration and decreases caregiver burden for patients with moderate to severe AD but is not approved for mild AD. Each of these compounds has only modest efficacy for AD. Cholinesterase inhibitors are relatively easy to administer, and their major side effects are gastrointestinal symptoms (nausea, diarrhea, cramps), altered sleep with unpleasant or vivid dreams, bradycardia (usually benign), and muscle cramps.

In a prospective observational study, the use of estrogen replacement therapy appeared to protect—by about 50%—against development of AD in women. This study seemed to confirm the results of two earlier case-controlled studies. Sadly, a prospective placebo-controlled study of a combined estrogen-progesterone therapy for asymptomatic postmenopausal women increased, rather than decreased, the prevalence of dementia. This study markedly dampened enthusiasm for hormonal treatments to prevent dementia. Additionally, no benefit has been found in the treatment of AD with estrogen alone.

A randomized, double-blind, placebo-controlled trial of an extract of *Ginkgo biloba* found modest improvement in cognitive function in subjects with AD and vascular dementia. Unfortunately, a comprehensive 6-year multicenter prevention study using *Ginkgo biloba* found no slowing of progression to dementia in the treated group.

Vaccination against $A\beta_{42}$ has proved highly efficacious in mouse models of AD, helping clear brain amyloid and preventing further amyloid accumulation. In human trials, this approach led to life-threatening complications, including meningoencephalitis, but modifications of the vaccine approach using passive immunization with monoclonal antibodies are currently being evaluated in phase 3 trials. Another experimental approach to AD treatment has been the use of β and γ secretase inhibitors that diminish the production of $A\beta_{42}$, but the first two placebo-controlled trials of γ secretase inhibitors, tarenflurbil and semagacestat, were negative, and semagacestat may have accelerated cognitive decline compared to placebo. Medications that modify tau phosphorylation and aggregation are beginning to be studied as possible treatments for both AD and non-AD tau-related disorders including FTD and PSP.

Several retrospective studies suggest that nonsteroidal anti-inflammatory agents and 3-hydroxy-3-methylglutaryl- coenzyme A (HMG-CoA) reductase inhibitors (statins) may have a protective effect on dementia, and controlled prospective studies are being conducted. Similarly, prospective studies with the goal of lowering serum homocysteine are underway to slow the progression to dementia, following an association of elevated homocysteine with dementia progression in epidemiologic studies. Finally, there is now a strong interest in the relationship between diabetes and AD, and insulin-regulating

studies are being conducted.

Mild to moderate depression is common in the early stages of AD and may respond to antidepressants or cholinesterase inhibitors. Selective serotonin reuptake inhibitors (SSRIs) are commonly used due to their low anticholinergic side effects (escitalopram 5–10 mg daily). Generalized seizures should be treated with an appropriate anticonvulsant, such as phenytoin or carbamazepine. Agitation, insomnia, hallucinations, and belligerence are especially troublesome characteristics of some AD patients, and these behaviors can lead to nursing home placement. The newer generation of atypical antipsychotics, such as risperidone, quetiapine, and olanzapine, are being used in low doses to treat these neuropsychiatric symptoms. The few controlled studies comparing drugs against behavioral intervention in the treatment of agitation suggest mild efficacy with significant side effects related to sleep, gait, and cardiovascular complications, including an increased risk of death. All antipsychotics carry a black-box FDA warning and should be used with caution in the demented elderly however, careful, daily, nonpharmacologic behavior management is often not available, rendering medications necessary for some patients. Finally, medications with strong anticholinergic effects should be vigilantly avoided, including prescription and over-the-counter sleep aids (e.g., diphenhydramine) or incontinence therapies (e.g., oxybutynin).

VASCULAR DEMENTIA

Dementia associated with cerebrovascular disease can be divided into two general categories: multi-infarct dementia and diffuse white matter disease (also called *leukoaraiosis, subcortical arteriosclerotic leukoencephalopathy,* or *Binswanger's disease*). Cerebrovascular disease appears to be a more common cause of dementia in Asia than in Europe and North America, perhaps due to the increased prevalence of intracranial atherosclerosis. Individuals who have had several strokes may develop chronic cognitive deficits, commonly called *multi-infarct dementia*. The strokes may be large or small (sometimes lacunar) and usually involve several different brain regions. The occurrence of dementia depends partly on the total volume of damaged cortex, but it is also more common in individuals with left-hemisphere lesions, independent of any language disturbance. Patients typically report previous discrete episodes of sudden neurologic deterioration. Many patients with multi-infarct dementia have a history of hypertension, diabetes, coronary artery disease, or other manifestations of widespread atherosclerosis. Physical examination usually shows focal neurologic deficits such as hemiparesis, a unilateral Babinski sign, a visual field defect, or pseudobulbar palsy. Recurrent strokes result in a stepwise disease progression. Neuroimaging reveals multiple areas of infarction. Thus, the history and neuroimaging findings differentiate this condition from AD; however, both AD and multiple infarctions are common and sometimes co-occur. With normal aging, there is also an accumulation of amyloid in cerebral blood vessels, leading to a condition called *cerebral amyloid angiopathy* (without dementia), which predisposes older persons to lobar hemorrhage and brain microhemorrhages. AD patients appear to be at increased risk for amyloid angiopathy, and this may explain some of the observed association between AD and stroke.

Some individuals with dementia are discovered on MRI to have bilateral abnormalities of subcortical white matter, termed *diffuse white matter disease*, often occurring in association with lacunar infarctions **(Fig. 371-5)**. The dementia may be insidious in onset and progress slowly, features that distinguish it from multi-infarct dementia, but other patients show a stepwise deterioration more typical of multi-infarct dementia. Early symptoms are mild confusion, apathy, anxiety, psychosis, and memory, spatial, or executive deficits. Marked difficulties in judgment and orientation and dependence on others for daily activities develop later. Euphoria, elation, depression, or aggressive behaviors are common as the disease progresses. Both pyramidal and cerebellar signs may be present. A gait disorder is present in at least half of these patients. With advanced disease, urinary incontinence and dysarthria with or without other pseudobulbar features (e.g., dysphagia, emotional lability) are frequent. Seizures and myoclonic jerks appear in a minority of patients. This disorder appears to result from chronic ischemia due to occlusive disease of small, penetrating cerebral arteries and arterioles (microangiopathy). Any disease-causing stenosis of small cerebral vessels may be the critical underlying factor, although hypertension is the major cause. The term *Binswanger's disease* should be used with caution, because it does not clearly identify a single entity.

Figure 371-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diffuse white matter disease. Axial fluid-attenuated inversion recovery (FLAIR) MR image through the lateral ventricles reveals multiple areas of hyperintensity involving the periventricular white matter as well as the corona radiata and striatum (*arrows*). While seen in some individuals with normal cognition, this appearance is more pronounced in patients with dementia of a vascular etiology.

Other rare causes of white matter disease also present with dementia, such as adult metachromatic leukodystrophy (arylsulfatase A deficiency) and progressive multifocal leukoencephalopathy (JC virus infection). A dominantly inherited form of diffuse white matter disease is known as *cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy* (CADASIL). Clinically, there is a progressive dementia developing in the fifth to seventh decades in multiple family members who may also have a history of migraine and recurrent lacunar stroke without hypertension. Skin biopsy may show pathognomonic osmophilic granules in the media of arterioles. The disease is caused by mutations in the *Notch 3* gene, and genetic testing is commercially available. The frequency of this disorder is unknown, and there are no effective treatments.

Mitochondrial disorders can present with stroke-like episodes and can selectively injure basal ganglia or cortex. Many such patients show other findings suggestive of a neurologic or systemic disorder such as ophthalmoplegia, retinal degeneration, deafness, myopathy, neuropathy, or diabetes. Diagnosis is difficult, but serum or (especially) CSF levels of lactate and pyruvate may be abnormal, and biopsy of affected tissue, preferably muscle, may be diagnostic.

Treatment of vascular dementia must be focused on preventing new ischemic injury by stabilizing or removing the underlying causes, such as hypertension, diabetes, smoking, or lack of exercise. Recovery of lost cognitive function is not likely, although fluctuations with periods of improvement are common.

FRONTOTEMPORAL DEMENTIA, PROGRESSIVE SUPRANUCLEARPALSY, AND CORTICOBASAL DEGENERATION

Frontotemporal dementia (FTD) often begins in the fifth to seventh decades, and in this age group it is nearly as prevalent as AD. Early studies suggested that FTD may be more common in men than women, although more recent reports cast some doubt on this finding. *Unlike in AD, behavioral symptoms predominate in the early stages of FTD.* Although a family history of dementia is common, autosomal dominant inheritance is seen in about 10% of all FTD cases. The clinical heterogeneity in familial and sporadic forms of FTD is remarkable, with patients demonstrating variable mixtures of behavioral, language, movement, and motor neuron symptoms. The most common autosomal dominantly inherited mutations causing FTD involve the *MAPT* or *GRN* genes, both on chromosome 17. *MAPT* mutations lead to a change in the alternate splicing of tau or cause loss of function in the tau molecule. With *GRN*, mutations in the coding sequence of the gene encoding progranulin protein result in mRNA degradation due to nonsense-mediated decay. Progranulin appears to be a rare example of an autosomal dominant mutation that leads to haploinsufficiency, resulting in around one-half the normal level of progranulin protein. Progranulin is a growth factor that binds to tumor necrosis

factor (TNF) receptors. How progranulin mutations lead to FTD is unknown. Both MAPT and GRN mutations are associated with parkinsonian features, while ALS is rare with these mutations. In contrast, familial FTD with ALS has been linked to chromosome 9. Mutations in the valosin-containing protein (chromosome 9) and the charged multivesicular body protein 2b (CHMP2b) genes (chromosome 3) also lead to rare autosomal dominant forms of familial FTD. Mutations in the TDP-43 and FUS genes (see below) cause familial ALS, sometimes in association with an FTD syndrome, although a few patients presenting with FTD alone have been reported.

In FTD, early symptoms are divided among behavioral, language, and sometimes motor abnormalities, reflecting degeneration of the anterior insular, frontal, and temporal regions, basal ganglia, and motor neurons. Cognitive testing typically reveals spared memory but impaired planning, judgment, or language. Poor business decisions and difficulty organizing work tasks are common, and speech and language deficits often emerge. Patients with FTD often show an absence of insight into their condition. Common behavioral features include apathy, disinhibition, weight gain, food fetishes, compulsions, and emotional distance or loss of empathy.

Findings at the bedside are dictated by the anatomic localization of the disorder. Asymmetric left frontal cases present with nonfluent aphasia, while left anterior temporal degeneration is characterized by loss of word meaning (semantic dementia). Nonfluent patients quickly progress to mutism, while those with semantic dementia develop features of a multimodal associative agnosia, losing the ability to recognize faces, objects, words, and the emotions of others. Visuoconstructive ability, arithmetic calculations, and navigation often remain normal late into the illness. Recently it has become apparent that many patients with nonfluent aphasia progress to clinical syndromes that overlap with PSP and corticobasal degeneration (CBD) and show these pathologies at autopsy. This left hemisphere presentation of FTD has been called *primary progressive aphasia with nonfluent and semantic variants*. In contrast, right frontal or temporal cases show profound alterations in social conduct, with loss of empathy, disinhibition, and antisocial behaviors predominating. There is a striking overlap among FTD, PSP, CBD, and motor neuron disease; ophthalmoplegia, dystonia, swallowing symptoms, and fasciculations are common at presentation of FTD or emerge during the course of the illness.

The distinguishing anatomic hallmark of FTD is a focal atrophy of frontal, insular, and/or temporal cortex, which can be visualized with neuroimaging studies and is often profound at autopsy **(Figs. 371-6 and 371-7)**. Despite the appearance of advanced FTD, however, the atrophy often begins focally in one hemisphere before spreading to anatomically interconnected regions, including basal ganglia. Microscopic findings seen across all patients with FTD include gliosis, microvacuolation, and neuronal loss, but the disease is subtyped according to the protein composition of neuronal and glial inclusions, which contain either tau or TDP-43 in at least 90% of patients, with the remaining 10% showing inclusions containing FUS **(Fig. 371-8)**.

Figure 371-6

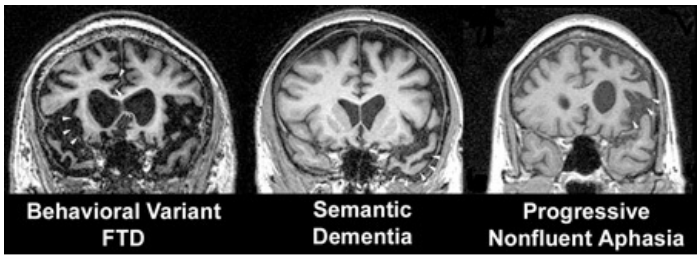

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Frontotemporal dementia (FTD). Coronal MRI sections from representative patients with behavioral variant FTD (***left***), semantic dementia (***center***), and progressive nonfluent aphasia (***right***). Areas of early and severe atrophy in each syndrome are highlighted (*white arrowheads*). The behavioral variant features anterior cingulate and frontoinsular atrophy, spreading to orbital and dorsolateral prefrontal cortex. Semantic dementia shows prominent temporopolar atrophy, more often on the left. Progressive nonfluent aphasia is associated with dominant frontal opercular and dorsal insula degeneration.

Figure 371-7

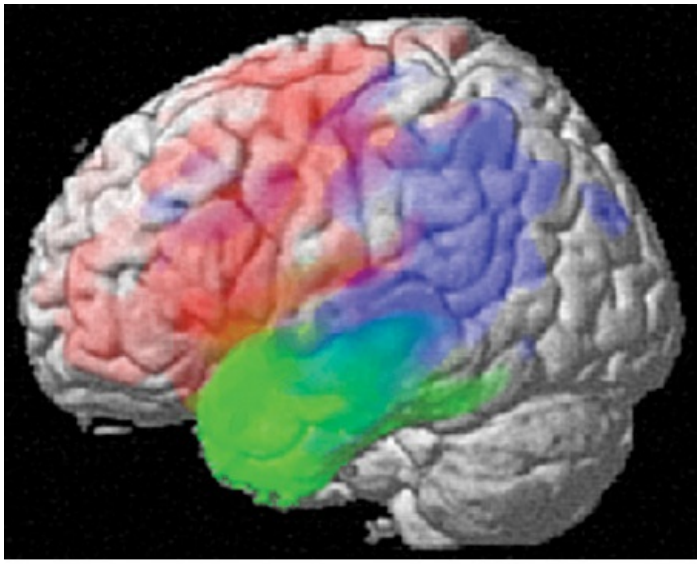

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Voxel-based morphometry analysis showing differing patterns of brain atrophy in three variants of progressive aphasia, including nonfluent (*red*), semantic (*green*), and logopedic subtypes (*blue*). Voxel-based morphometry allows comparison of MRI gray matter volumes between patient groups and control subjects, as shown here. *(Image courtesy of M Gorno-Tempini, University of California at San Francisco; with permission.)*

Figure 371-8

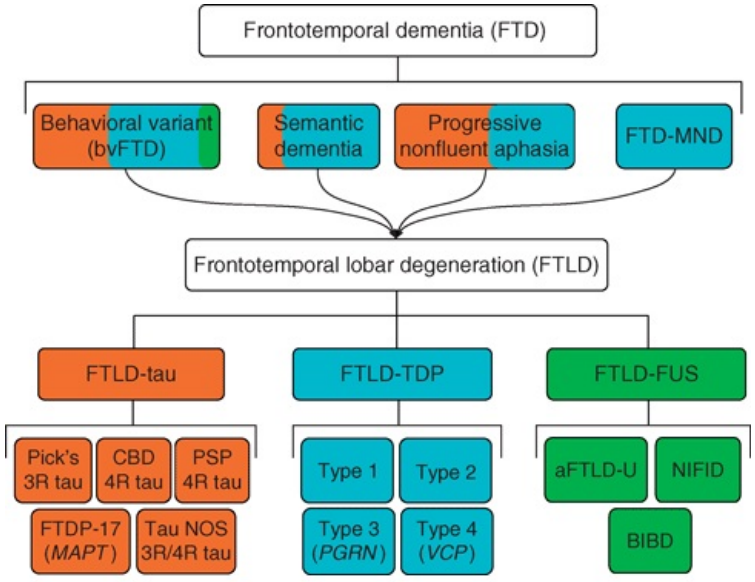

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Frontotemporal dementia syndromes are united by underlying frontotemporal lobar degeneration pathology, which can be divided according to the presence of tau, TPD-43, or fused in sarcoma (FUS) inclusions in neurons and glia. Correlations between clinical syndrome and major molecular category are shown with colored shading.

A toxic gain of function related to tau underlies the pathogenesis of many familial cases and is presumed to be a factor in sporadic tauopathies, although loss of tau microtubule stabilizing function may also play a role. TDP-43 and FUS, in contrast, are RNA/DNA binding proteins whose roles in neuronal function are still being actively investigated. Loss of cortical serotonergic innervation is seen in many patients. In contrast to AD, the cholinergic system is relatively spared in FTD.

Historically, *Pick's disease* was described as a progressive degenerative disorder characterized by selective involvement of the anterior frontal and temporal neocortex and pathologically by

intraneuronal cytoplasmic inclusions (*Pick bodies*). Classical Pick bodies are argyrophilic, staining positively with the Bielschowsky silver method and also with immunostaining for tau **(Fig. 371-9)**. Subsequent pathologic studies revealed a significant subset of patients with silver-negative, tau-negative inclusions, which have since been shown mainly to contain TDP-43, although a minority stain only for FUS. Although the nomenclature used to describe these patients has continued to evolve, the term *FTD* is increasingly used to refer to the clinical syndromes, while *frontotemporal lobar degeneration (FTLD)* refers to the underlying pathology, with three major subtypes recognized: FTLD-tau, FTLD-TDP, and FTLD-FUS. Despite significant progress, available data do not yet allow a reliable prediction of underlying pathology based on clinical features. Accordingly researchers continue to seek serum, CSF, or neuroimaging biomarkers that will afford greater diagnostic accuracy, defined as concordance with neuropathology.

Figure 371-9

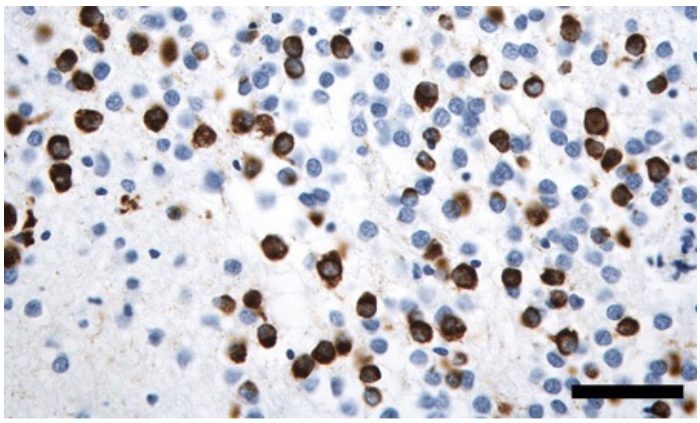

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pick's disease, a subtype offrontotemporal lobar degeneration(FTLD)-tau. Pick bodies, shown here in the dentate gyrus of a patient with advanced bvFTD, consist of loosely arranged paired helical and straight filaments and stain positively for hyperphosphorylated tau. Classical Pick's disease is seen in only 10–20% of patients with frontotemporal dementia. Scale bar represents 50 microns. *(CP-13 antibody courtesy of P. Davies.)*

The burden on caregivers of patients with FTD is extremely high because the illness disrupts core emotional and personality functions of the loved one. Treatment is symptomatic, and there are currently no therapies known to slow progression or improve symptoms. Many of the behaviors that accompany FTD, such as depression, hyperorality, compulsions, and irritability, can be ameliorated with antidepressants, especially SSRIs. The co-association with motor disorders such as parkinsonism necessitates the careful use of antipsychotics, which can exacerbate this problem.

Progressive supranuclear palsy (PSP) is a degenerative disease that involves the brainstem, basal ganglia, limbic structures, and selected areas of cortex. Clinically, this disorder begins with falls and executive or subtle personality changes (such as mental rigidity, impulsivity, or apathy). Shortly thereafter, a progressive oculomotor syndrome ensues that begins with square wave jerks, followed by slowed saccades (vertical worse than horizontal) before resulting in progressive supranuclear ophthalmoparesis. Dysarthria, dysphagia, and symmetric axial rigidity can be prominent features that emerge at any point in the illness. A stiff, unstable posture with hyperextension of the neck and a slow, jerky, toppling gait is characteristic. Frequent unexplained and sometimes spectacular falls are common secondary to a combination of axial rigidity, inability to look down, and bad judgment. Even once patients have severely limited voluntary eye movements, they retain oculocephalic reflexes (demonstrated using a vertical doll's head maneuver); thus, the oculomotor disorder is supranuclear. The dementia overlaps with FTD, featuring apathy, frontal-executive dysfunction, poor judgment, slowed thought processes, impaired verbal fluency, and difficulty with sequential actions and with shifting from one task to another. These features are common at presentation and often precede the motor syndrome. Some patients begin with a nonfluent aphasia or motor speech disorder and progress to classical PSP. Response to l-dopa is limited or absent; no other treatments exist. Death occurs within 5–10 years of onset. At autopsy, accumulation of hyperphosphorylated tau is seen within neurons and glia. Neuronal inclusions often take the form of neurofibrillary tangles (NFTs), which may be large, spherical, and coarse when found in brainstem oculomotor control system neurons. These characteristic tau inclusions are called globose tangles, and may be found in multiple subcortical structures (including the subthalamic nucleus, globus pallidus, substantia nigra, locus coeruleus, periaqueductal gray, superior colliculi, oculomotor nuclei, and dentate nucleus). Neocortical NFTs, like those in AD, often take on a more flame-shaped morphology, but on electron microscopy PSP tangles can be shown to consist of straight tubules rather than the paired helical filaments found in AD.

Furthermore, PSP is associated with prominent tau-positive glial pathomorphologies, such as tufted and thorny astrocytes.

In addition to its overlap with FTD and CBD (see below), PSP is often confused with idiopathic *Parkinson's disease* (PD). Although elderly patients with PD may have restricted upgaze, they do not develop downgaze paresis or other abnormalities of voluntary eye movements typical of PSP. Dementia occurs in ~20% of patients with PD, often due to the emergence of a full-blown DLB syndrome. Furthermore, the behavioral syndromes seen with DLB differ from PSP (see below). Dementia in PD becomes more likely with increasing age, increasing severity of extrapyramidal signs, a long disease duration, and the presence of depression. Patients with PD who develop dementia also show cortical atrophy on brain imaging. Neuropathologically, there may be Alzheimer's disease–related changes in the cortex, DLB-related α-synuclein inclusions in both the limbic system and cortex, or no specific microscopic changes other than gliosis and neuronal loss. Parkinson's disease is discussed in detail in Chap. 372.

Corticobasal degeneration (CBD) is a slowly progressive dementing illness associated with severe gliosis and neuronal loss in both the cortex and basal ganglia (substantia nigra and striatopallidum). Some patients present with a unilateral onset with rigidity, dystonia, and apraxia of one arm and hand, sometimes called the *alien limb* when it begins to exhibit unintended motor actions, while in other instances the disease presents as a progressive behavioral, executive, or language syndrome or as progressive symmetric parkinsonism. Some patients begin with a progressive nonfluent aphasia or a progressive motor speech disorder. Eventually CBD becomes bilateral and leads to dysarthria, slow gait, action tremor, and dementia. The microscopic features include ballooned, achromatic, tau-positive neurons with astrocytic plaques and other dystrophic glial tau pathomorphologies that overlap with those seen in PSP. Most specifically, CBD features a severe tauopathy burden in the subcortical white matter, consisting of threads and oligodendroglial coiled bodies. The condition is rarely familial, the cause is unknown, and there is no specific treatment.

PARKINSON'S DISEASE DEMENTIA AND DEMENTIA WITH LEWY BODIES

The parkinsonian dementia syndromes are under increasing study, with many cases unified by Lewy body and Lewy neurite pathology that ascends from the low brainstem up through the substantia nigra, limbic system, and cortex. The DLB clinical syndrome is characterized by visual hallucinations, parkinsonism, fluctuating alertness, falls, and often RBD. Dementia can precede or follow the appearance of parkinsonism. Hence, one pathway occurs in patients with long-standing PD without cognitive impairment, who slowly develop a dementia that is associated with visual hallucinations and fluctuating alertness. When this occurs after an established diagnosis of PD, many use the term *Parkinson's disease dementia* (PDD). In others, the dementia and neuropsychiatric syndrome precede the parkinsonism, and this constellation is referred to as DLB. Both PDD and DLB may be accompanied or preceded by symptoms referable to brainstem pathology below the substantia nigra, and many researchers conceptualize these disorders as points on a spectrum of α-synuclein pathology.

Patients with PDD and DLB are highly sensitive to metabolic perturbations, and in some patients the first manifestation of illness is a delirium, often precipitated by an infection, new medicine, or other systemic disturbance. A hallucinatory delirium induced by l-dopa, prescribed for parkinsonian symptoms attributed to PD may likewise provide the initial clue to a PDD diagnosis. Conversely, patients with mild cognitive deficits and hallucinations may receive typical or atypical antipsychotic medications, which induce profound parkinsonism at low doses due to a subclinical DLB-related nigral dopaminergic neuron loss. Even without an underlying precipitant, fluctuations can be marked in DLB, with episodic confusion or even stupor admixed with lucid intervals. However, despite the fluctuating pattern, the clinical features persist over a long period, unlike delirium, which resolves following correction of the inciting factor. Cognitively, DLB features relative preservation of memory but more severe visuospatial and executive deficits than patients with early AD.

The key neuropathologic feature in DLB is the presence of Lewy bodies and Lewy neurites throughout specific brainstem nuclei, substantia nigra, amygdala, cingulate gyrus, and, ultimately, the neocortex. Lewy bodies are intraneuronal cytoplasmic inclusions that stain with periodic acid–Schiff (PAS) and ubiquitin but are now identified with antibodies to the presynaptic protein, α-synuclein. They are composed of straight neurofilaments 7–20 nm long with surrounding amorphous material and contain epitopes recognized by antibodies against phosphorylated and nonphosphorylated neurofilament proteins, ubiquitin, and α-synuclein. Lewy bodies are typically found in the substantia nigra of patients with idiopathic PD, where they can be readily seen with hematoxylin-and-eosin staining. A profound cholinergic deficit, owing to basal forebrain and pedunculopontine nucleus involvement, is present in many patients with DLB and may be a factor responsible for the fluctuations, inattention, and visual hallucinations. In patients without other pathologic features, the condition is sometimes referred to as *diffuse Lewy body disease*. In patients whose brains also contain a substantial burden of amyloid plaques and NFTs, the condition is sometimes called the *Lewy body variant of Alzheimer's disease*.

Due to the overlap with AD and the cholinergic deficit in DLB, cholinesterase inhibitors often provide

significant benefit, reducing hallucinosis, stabilizing delusional symptoms, and even helping with RBD in some patients. Exercise programs maximize motor function and protect against fall-related injury. Antidepressants are often necessary. Atypical antipsychotics may be required for psychosis but can worsen extrapyramidal syndromes, even at low doses, and increase risk of death. As noted above, patients with DLB are extremely sensitive to dopaminergic medications, which must be carefully titrated; tolerability may be improved by concomitant use of a cholinesterase inhibitor.

OTHER CAUSES OF DEMENTIA

Prion diseases such as *Creutzfeldt-Jakob disease* (CJD) are rare neurodegenerative conditions (prevalence ~1 per million) that produce dementia. CJD is a rapidly progressive disorder associated with dementia, focal cortical signs, rigidity, and myoclonus, causing death <1 year after first symptoms appear. The rapidity of progression seen with CJD is uncommon in AD so that the distinction between the two disorders is usually straightforward. CBD and DLB, more rapid degenerative dementias with prominent movement abnormalities, are more likely to be mistaken for CJD. The differential diagnosis for CJD includes other rapidly progressive dementing conditions such as viral or bacterial encephalitides, Hashimoto's encephalopathy, CNS vasculitis, lymphoma, or paraneoplastic syndromes. The markedly abnormal periodic complexes on EEG and cortical ribbon and basal ganglia hyperintensities on MR diffusion-weighted imaging are highly specific diagnostic features of CJD, although rarely prolonged focal or generalized seizures can produce a similar imaging appearance. Prion diseases are discussed in detail in Chap. 383.

Huntington's disease (HD) (Chap. 372) is an autosomal dominant, degenerative brain disorder. HD clinical hallmarks include chorea, behavioral disturbance, and executive impairment. Symptoms typically begin in the fourth or fifth decade, but there is a wide range, from childhood to >70 years. Memory is frequently not impaired until late in the disease, but attention, judgment, awareness, and executive functions are often deficient at an early stage. Depression, apathy, social withdrawal, irritability, and intermittent disinhibition are common. Delusions and obsessive-compulsive behavior may occur. Disease duration is typically around 15 years but is quite variable.

Normal-pressure hydrocephalus (NPH) is a relatively uncommon but treatable syndrome. The clinical, physiologic, and neuroimaging characteristics of NPH must be carefully distinguished from those of other dementias associated with gait impairment. Historically, many patients treated for NPH have suffered from other dementias, particularly AD, vascular dementia, DLB, and PSP. For NPH, the clinical triad includes an abnormal gait (ataxic or apractic), dementia (usually mild to moderate, with an emphasis on executive impairment), and urinary urgency or incontinence. Neuroimaging reveals enlarged lateral ventricles (hydrocephalus) with little or no cortical atrophy, although the sylvian fissures may appear propped open (so-called "boxcarring"), which can be mistaken for perisylvian atrophy. This syndrome is a communicating hydrocephalus with a patent aqueduct of Sylvius **(Fig. 371-10)**, in contrast to aqueductal stenosis, in which the aqueduct is small. Lumbar puncture opening pressure falls in the high normal range, and the CSF protein, glucose, and cell counts are normal. NPH may be caused by obstruction to normal CSF flow over the cerebral convexities and delayed resorption into the venous system. The indolent nature of the process results in enlarged lateral ventricles with relatively little increase in CSF pressure. Presumed stretching and distortion of subfrontal white matter tracts may lead to clinical symptoms, but the precise underlying pathophysiology remains unclear. Some patients provide a history of conditions that produce meningeal scarring (blocking CSF resorption) such as previous meningitis, subarachnoid hemorrhage, or head trauma. Others with long-standing but asymptomatic congenital hydrocephalus may have adult-onset deterioration in gait or memory that is confused with NPH. In contrast to AD, the patient with NPH complains of an early and prominent gait disturbance without cortical atrophy on CTor MRI.

Figure 371-10

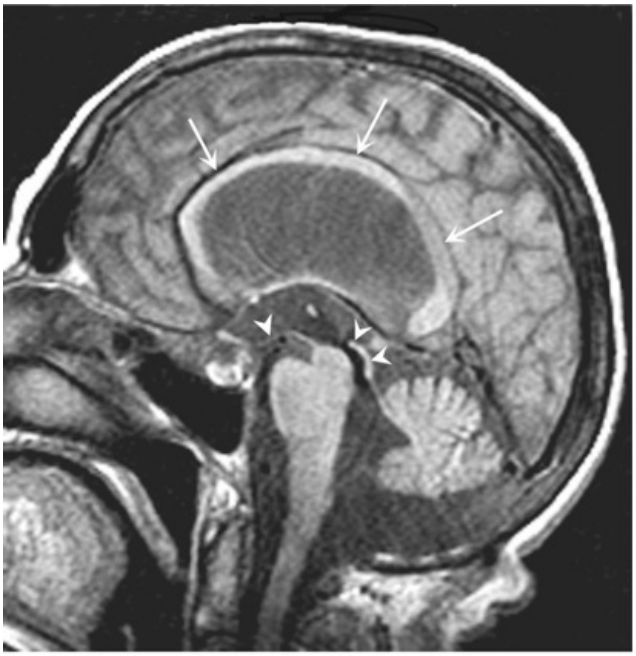

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

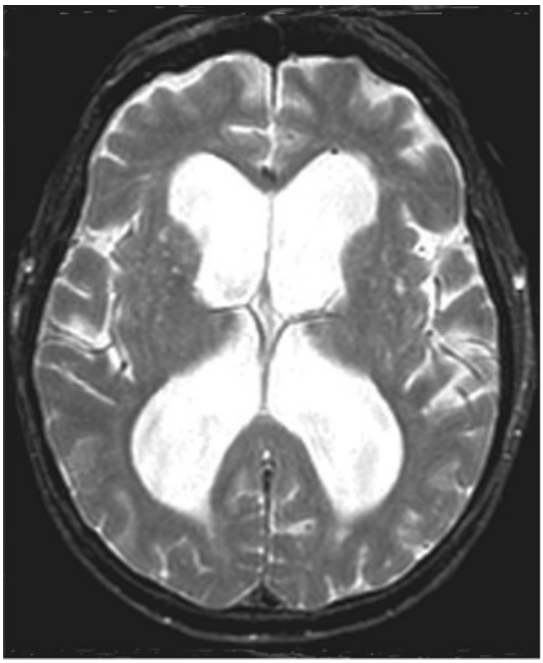

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Normal-pressure hydrocephalus. ***A.*** Sagittal T1-weighted MR image demonstrates dilation of the lateral ventricle and stretching of the corpus callosum (*arrows*), depression of the floor of the third ventricle (*single arrowhead*), and enlargement of the aqueduct (*double arrowheads*). Note the diffuse dilation of the lateral, third, and fourth ventricles with a patent aqueduct, typical of communicating hydrocephalus. ***B.*** Axial T2-weighted MR images demonstrate dilation of the lateral ventricles. This patient underwent successful ventriculoperitoneal shunting.

Numerous attempts to improve NPH diagnosis with various special studies and predict the success of ventricular shunting have been undertaken. These tests include radionuclide cisternography (showing a delay in CSF absorption over the convexity) and various efforts to monitor and alter CSF flow dynamics, including a constant-pressure infusion test. None has proven to be specific or consistently useful. A transient improvement in gait or cognition may follow lumbar puncture (or serial punctures) with removal of 30–50 mL of CSF, but this finding has also not proved to be consistently predictive of postshunt improvement. Perhaps the most reliable strategy is a period of close inpatient evaluation

before, during, and after lumbar CSF drainage. Occasionally, when a patient with AD presents with gait impairment (at times due to comorbid subfrontal vascular injury) and absent or only mild cortical atrophy on CT or MRI, distinguishing NPH from AD can be challenging. Hippocampal atrophy on MRI favors AD, whereas a characteristic "magnetic" gait with external hip rotation, low foot clearance and short strides, along with prominent truncal sway or instability, favors NPH. The diagnosis of NPH should be avoided when hydrocephalus is not detected on imaging studies, even if the symptoms otherwise fit. Thirty to fifty percent of patients identified by careful diagnosis as having NPH will improve with ventricular shunting. Gait may improve more than cognition, but many reported failures to improve cognitively may have resulted from comorbid AD. Short-lasting improvement is common. Patients should be carefully selected for shunting, because subdural hematoma, infection, and shunt failure are known complications and can be a cause for early nursing home placement in an elderly patient with previously mild dementia.

Dementia can accompany *chronic alcoholism* (Chap. 392) and may result from associated malnutrition, especially of B vitamins, particularly thiamine. Other poorly defined aspects of chronic alcoholism may, however, also produce cerebral damage. A rare idiopathic syndrome of dementia and seizures with degeneration of the corpus callosum has been reported primarily in male Italian red wine drinkers (Marchiafava-Bignami disease).

Thiamine (vitamin B_1) deficiency causes Wernicke's encephalopathy (Chap. 275). The clinical presentation features a malnourished patient (frequently but not necessarily alcoholic) with confusion, ataxia, and diplopia resulting from inflammation and necrosis of periventricular midline structures, including dorsomedial thalamus, mammillary bodies, midline cerebellum, periaqueductal gray matter, and trochlear and abducens nuclei. Damage to the dorsomedial thalamus correlates most closely with the memory loss. Prompt administration of parenteral thiamine (100 mg intravenously for 3 days followed by daily oral dosage) may reverse the disease if given in the first days of symptom onset. However, prolonged untreated thiamine deficiency can result in an irreversible dementia/amnestic syndrome (Korsakoff's syndrome) or even death.

In *Korsakoff's syndrome*, the patient is unable to recall new information despite normal immediate memory, attention span, and level of consciousness. Memory for new events is seriously impaired, whereas knowledge acquired prior to the illness remains relatively intact. Patients are easily confused, disoriented, and cannot store information for more than a few minutes. Superficially, they may be conversant, engaging, and able to perform simple tasks and follow immediate commands. Confabulation is common, although not always present. There is no specific treatment because the previous thiamine deficiency has produced irreversible damage to the medial thalamic nuclei and mammillary bodies. Mammillary body atrophy may be visible on MRI in the chronic phase (Fig. 275-7).

Vitamin B_{12} deficiency, as can occur in pernicious anemia, causes a megaloblastic anemia and may also damage the nervous system (Chaps. 105 and 377). Neurologically, it most commonly produces a spinal cord syndrome (myelopathy) affecting the posterior columns (loss of vibration and position sense) and corticospinal tracts (hyperactive tendon reflexes with Babinski signs); it also damages peripheral nerves (neuropathy), resulting in sensory loss with depressed tendon reflexes. Damage to myelinated axons may also cause dementia. The mechanism of neurologic damage is unclear but may be related to a deficiency of *S*-adenosyl methionine (required for methylation of myelin phospholipids) due to reduced methionine synthase activity or accumulation of methylmalonate, homocysteine, and propionate, providing abnormal substrates for fatty acid synthesis in myelin. The neurologic sequelae of vitamin B12 deficiency may occur in the absence of hematologic manifestations, making it critical to avoid using the CBC and blood smear as a substitute for measuring B_{12} blood levels. Treatment with parenteral vitamin B_{12} (1000 μg intramuscularly daily for a week, weekly for a month, and monthly for life for pernicious anemia) stops progression of the disease if instituted promptly, but complete reversal of advanced nervous system damage will not occur.

Deficiency of nicotinic acid (*pellagra*) is associated with skin rash over sun-exposed areas, glossitis, and angular stomatitis (Chap. 74). Severe dietary deficiency of nicotinic acid along with other B vitamins such as pyridoxine may result in spastic paraparesis, peripheral neuropathy, fatigue, irritability, and dementia. This syndrome has been seen in prisoners of war and in concentration camps but should be considered in any malnourished individual. Low serum folate levels appear to be a rough index of malnutrition, but isolated folate deficiency has not been proved as a specific cause of dementia.

CNS infections usually cause delirium and other acute neurologic syndromes. However, some chronic CNS infections, particularly those associated with chronic meningitis (Chap. 382), may produce a dementing illness. The possibility of chronic infectious meningitis should be suspected in patients presenting with a dementia or behavioral syndrome, who also have headache, meningismus, cranial neuropathy, and/or radiculopathy. Between 20 and 30% of patients in the advanced stages of HIV infection become demented (Chap. 189). Cardinal features include psychomotor retardation, apathy, and impaired memory. This syndrome may result from secondary opportunistic infections but can also be caused by direct infection of CNS neurons with HIV. Neurosyphilis (Chap. 169) was a common

cause of dementia in the preantibiotic era; it is now uncommon but can still be encountered in patients with multiple sex partners, particularly among patients with HIV. Characteristic CSF changes consist of pleocytosis, increased protein, and a positive venereal disease research laboratory (VDRL) test.

Primary and metastatic *neoplasms of the CNS* (Chap. 379) usually produce focal neurologic findings and seizures rather than dementia, but if tumor growth begins in the frontal or temporal lobes, the initial manifestations may be memory loss or behavioral changes. A paraneoplastic syndrome of dementia associated with occult carcinoma (often small cell lung cancer) is termed *limbic encephalitis*. In this syndrome, confusion, agitation, seizures, poor memory, emotional changes, and frank dementia may occur. Paraneoplastic *encephalitis associated with NMDA receptor antibodies* presents as a progressive psychiatric disorder with memory loss and seizures; affected patients are often young women with ovarian teratoma (Chap. 101).

A *nonconvulsive seizure disorder* may underlie a syndrome of confusion, clouding of consciousness, and garbled speech. Often, psychiatric disease is suspected, but an EEG demonstrates the epileptic nature of the illness. If recurrent or persistent, the condition may be termed *complex partial status epilepticus*. The cognitive disturbance often responds to anticonvulsant therapy. The etiology may be previous small strokes or head trauma; some cases are idiopathic.

It is important to recognize *systemic diseases* that indirectly affect the brain and produce chronic confusion or dementia. Such conditions include hypothyroidism; vasculitis; and hepatic, renal, or pulmonary disease. Hepatic encephalopathy may begin with irritability and confusion and slowly progress to agitation, lethargy, and coma.

Isolated vasculitis of the CNS (CNS granulomatous angiitis) (Chaps. 326 and 370) occasionally causes a chronic encephalopathy associated with confusion, disorientation, and clouding of consciousness. Headache is common, and strokes and cranial neuropathies may occur. Brain imaging studies may be normal or nonspecifically abnormal. CSF analysis reveals a mild pleocytosis or protein elevation. Cerebral angiography can show multifocal stenoses involving medium-caliber vessels, but some patients have only small-vessel disease that is not revealed on angiography. The angiographic appearance is not specific and may be mimicked by atherosclerosis, infection, or other causes of vascular disease. Brain or meningeal biopsy demonstrates endothelial cell proliferation and mononuclear infiltrates within blood vessel walls. The prognosis is often poor, although the disorder may remit spontaneously. Some patients respond to glucocorticoids or chemotherapy.

Chronic metal exposure represents a rare cause of dementia. The key to diagnosis is to elicit a history of exposure at work or home. Chronic lead poisoning from inadequately fire-glazed pottery has been reported. Fatigue, depression, and confusion may be associated with episodic abdominal pain and peripheral neuropathy. Gray lead lines appear in the gums, usually accompanied by an anemia with basophilic stippling of red blood cells. The clinical presentation can resemble that of acute intermittent porphyria, including elevated levels of urine porphyrins as a result of the inhibition of δ-aminolevulinic acid dehydrase. The treatment is chelation therapy with agents such as ethylenediamine tetraacetic acid (EDTA). Chronic mercury poisoning produces dementia, peripheral neuropathy, ataxia, and tremulousness that may progress to a cerebellar intention tremor or choreoathetosis. The confusion and memory loss of chronic arsenicintoxication is also associated with nausea, weight loss, peripheral neuropathy, pigmentation and scaling of the skin, and transverse white lines of the fingernails (Mees' lines). Treatment is chelation therapy with dimercaprol (BAL). Aluminum poisoning is rare but was documented with the dialysis dementia syndrome, in which water used during renal dialysis was contaminated with excessive amounts of aluminum. This poisoning resulted in a progressive encephalopathy associated with confusion, nonfluent aphasia, memory loss, agitation, and, later, lethargy and stupor. Speech arrest and myoclonic jerks were common and associated with severe and generalized EEG changes. The condition has been eliminated by the use of deionized water for dialysis.

Recurrent head trauma in professional boxers may lead to a dementia sometimes called the "punch-drunk" syndrome, or *dementia pugilistica*. The symptoms can be progressive, beginning late in a boxer's career or even long after retirement. The severity of the syndrome correlates with the length of the boxing career and number of bouts. Early in the condition, a personality change associated with social instability and sometimes paranoia and delusions occurs. Later, memory loss progresses to full-blown dementia, often associated with parkinsonian signs and ataxia or intention tremor. At autopsy, the cerebral cortex may show changes similar to AD, although NFTs are usually more prominent than amyloid plaques (which are usually diffuse rather than neuritic). Superficial layer NFT aggregates have been reported to differentiate these patients from those with more typical AD. Also, there may be loss of neurons in the substantia nigra. Chronic subdural hematoma (Chap. 378) is also occasionally associated with dementia, often in the context of underlying cortical atrophy from conditions such as AD or HD.

Transient global amnesia (TGA) is characterized by the sudden onset of a severe episodic memory deficit, usually occurring in persons over the age of 50 years. Often the amnesia occurs in the setting of an emotional stimulus or physical exertion. During the attack, the individual is alert and communicative, general cognition seems intact, and there are no other neurologic signs or symptoms.

The patient may seem confused and repeatedly ask about his or her location in place and time. The ability to form new memories returns after a period of hours, and the individual returns to normal with no recall for the period of the attack. Frequently no cause is determined, but cerebrovascular disease, epilepsy (7% in one study), migraine, or cardiac arrhythmias have all been implicated. A Mayo Clinic review of 277 patients with TGA found a history of migraine in 14% and cerebrovascular disease in 11%, but these conditions were not temporally related to the TGA episodes. Approximately one-quarter of the patients had recurrent attacks, but they were not at increased risk for subsequent stroke. Rare instances of permanent memory loss after sudden onset have been reported, usually representing ischemic infarction of the hippocampus or medial thalamic nucleus bilaterally.

The *ALS/parkinsonian/dementia complex of Guam* is a rare degenerative disease that has occurred in the Chamorro natives on the island of Guam. Individuals may have any combination of parkinsonian features, dementia, and motor neuron disease. The most characteristic pathologic features are the presence of NFTs in degenerating neurons of the cortex and substantia nigra and loss of motor neurons in the spinal cord, although recent reanalysis has shown that some patients with this illness also show coexisting TDP-43 pathology. Epidemiologic evidence supports a possible environmental cause, such as exposure to a neurotoxin or an infectious agent with a long latency period. One interesting but unproven candidate neurotoxin occurs in the seed of the false palm tree, which Guamanians traditionally used to make flour. The ALS syndrome is no longer present in Guam, but a dementing illness with rigidity continues to be seen.

Rarely, adult-onset leukodystrophies, lysosomal storage diseases, and other genetic disorders can present as a dementia in middle to late life. Metachromatic leukodystrophy (MLD) causes a progressive psychiatric or dementia syndrome associated with extensive, confluent frontal white matter abnormality. MLD is diagnosed by measuring arylsulfatase A enzyme activity in white blood cells. Adult-onset presentations of adrenoleukodystrophy have been reported in female carriers, and these patients often feature spinal cord and posterior white matter involvement. Adrenoleukodystrophy is diagnosed with measurement of plasma very long chain fatty acids. CADASIL is another genetic syndrome associated with white matter disease, often frontally and temporally predominant. Diagnosis is made with skin biopsy, which shows osmophilic granules in arterioles, or, increasingly, through genetic testing for mutations in Notch 3 (see above). The neuronal ceroid lipofuscinoses are a genetically heterogeneous group of disorders associated with myoclonus, seizures, and progressive dementia. Diagnosis is made by finding curvilinear inclusions within white blood cells or neuronal tissue.

Psychogenic amnesia for personally important memories can be seen. Whether this results from deliberate avoidance of unpleasant memories, outright malingering, or from unconscious repression remains unknown and probably depends on the patient. Event-specific amnesia is more likely to occur after violent crimes such as homicide of a close relative or friend or sexual abuse. It may develop in association with severe drug or alcohol intoxication and sometimes with schizophrenia. More prolonged psychogenic amnesia occurs in fugue states that also commonly follow severe emotional stress. The patient with a fugue state suffers from a sudden loss of personal identity and may be found wandering far from home. *In contrast to neurologic amnesia, fugue states are associated with amnesia for personal identity and events closely associated with the personal past.* At the same time, memory for other recent events and the ability to learn and use new information are preserved. The episodes usually last hours or days and occasionally weeks or months while the patient takes on a new identity. On recovery, there is a residual amnesia gap for the period of the fugue. Very rarely does selective loss of autobiographic information reflect a focal injury to the brain areas involved with these functions.

Psychiatric diseases may mimic dementia. Severely depressed or anxious individuals may appear demented, a phenomenon sometimes called *pseudodementia*. Memory and language are usually intact when carefully tested, and a significant memory disturbance usually suggests an underlying dementia, even if the patient is depressed. Patients in this condition may feel confused and unable to accomplish routine tasks. Vegetative symptoms, such as insomnia, lack of energy, poor appetite, and concern with bowel function, are common. Onset is often more abrupt, and the psychosocial milieu may suggest prominent reasons for depression. Such patients respond to treatment of the underlying psychiatric illness. Schizophrenia is usually not difficult to distinguish from dementia, but occasionally the distinction can be problematic. Schizophrenia generally has a much earlier age of onset (second and third decades) than most dementing illnesses, and is associated with intact memory. The delusions and hallucinations of schizophrenia are usually more complex and bizarre than those of dementia. Some chronic schizophrenics develop an unexplained progressive dementia late in life that is not related to AD. Conversely, FTD, HD, vascular dementia, DLB, AD, or leukoencephalopathy can begin with schizophrenia-like features, leading to the misdiagnosis of a psychiatric condition. Later age of onset, significant deficits on cognitive testing, or the presence of abnormal neuroimaging point toward a degenerative condition. Memory loss may also be part of a conversion disorder. In this situation, patients commonly complain bitterly of memory loss, but careful cognitive testing either does not confirm the deficits or demonstrates inconsistent or unusual patterns of cognitive problems. The patient's behavior and "wrong" answers to questions often indicate that he or she understands the

question and knows the correct answer.

Clouding of cognition by *chronic drug or medication use* , often prescribed by physicians, is an important cause of dementia. Sedatives, tranquilizers, and analgesics used to treat insomnia, pain, anxiety, or agitation may cause confusion, memory loss, and lethargy, especially in the elderly. Discontinuation of the offending medication often improves mentation.

Treatment: Dementia

The major goals of dementia management are to treat correctable causes and to provide comfort and support to the patient and caregivers. Treatment of underlying causes might include thyroid replacement for hypothyroidism; vitamin therapy for thiamine or B_{12} deficiency or for elevated serum homocysteine; antimicrobials for opportunistic infections or antiretrovirals for HIV; ventricular shunting for NPH; or appropriate surgical, radiation, and/or chemotherapeutic treatment for CNS neoplasms. Removal of cognition-impairing drugs or medications is the most frequently useful approach employed in a dementia clinic. If the patient's cognitive complaints stem from a psychiatric disorder, vigorous treatment of this condition should seek to eliminate the cognitive complaint or confirm that it persists despite adequate resolution of the mood or anxiety symptoms. Patients with degenerative diseases may also be depressed or anxious, and those aspects of their condition may respond to therapy. Antidepressants, such as SSRIs (Chap. 391), which feature anxiolytic properties but few cognitive side effects provide the mainstay of treatment when necessary. Anticonvulsants are used to control seizures.

Agitation, hallucinations, delusions, and confusion are difficult to treat. These behavioral problems represent major causes for nursing home placement and institutionalization. Before treating these behaviors with medications, the clinician should aggressively seek out modifiable environmental or metabolic factors. Hunger, lack of exercise, toothache, constipation, urinary tract infection, or drug toxicity all represent easily correctable causes that can be remedied without psychoactive drugs. Drugs such as phenothiazines and benzodiazepines may ameliorate the behavior problems but have untoward side effects such as sedation, rigidity, dyskinesia, and occasionally paradoxical disinhibition (benzodiazepines). Despite their unfavorable side-effect profile, second-generation antipsychotics such as quetiapine (starting dose, 12.5–25 mg daily) can be used for patients with agitation, aggression, and psychosis, although the risk profile for these compounds is significant. When patients do not respond to treatment, it is usually a mistake to advance to higher doses or to use anticholinergics or sedatives (such as barbiturates or benzodiazepines). It is important to recognize and treat depression; treatment can begin with a low dose of an SSRI (e.g., escitalopram 5–10 mg daily) while monitoring for efficacy and toxicity. Sometimes apathy, visual hallucinations, depression, and other psychiatric symptoms respond to the cholinesterase inhibitors, especially in DLB, obviating the need for other more toxic therapies.

Cholinesterase inhibitors are being used to treat AD (donepezil, rivastigmine, galantamine) and PDD (rivastigmine). Other compounds, such as anti-inflammatory agents, are being investigated in the treatment or prevention of AD. These approaches are reviewed in the treatment sections for individual disorders, above. Memantine proves useful when treating some patients with moderate to severe AD; its major benefit relates to decreasing caregiver burden, most likely by decreasing resistance to dressing and grooming support.

A proactive strategy has been shown to reduce the occurrence of delirium in hospitalized patients. This strategy includes frequent orientation, cognitive activities, sleep-enhancement measures, vision and hearing aids, and correction of dehydration.

Nondrug behavior therapy has an important place in dementia management. The primary goals are to make the patient's life comfortable, uncomplicated, and safe. Preparing lists, schedules, calendars, and labels can be helpful in the early stages. It is also useful to stress familiar routines, short-term tasks, walks, and simple physical exercises. For many demented patients, memory for events is worse than for routine activities, and they may still be able to take part in physical activities such as walking, bowling, dancing, and golf. Demented patients usually object to losing control over familiar tasks such as driving, cooking, and handling finances. Attempts to help or take over may be greeted with complaints, depression, or anger. Hostile responses on the part of the caretaker are useless and sometimes harmful. Explanation, reassurance, distraction, and calm positive statements are more productive in this setting. Eventually, tasks such as finances and driving must be assumed by others, and the patient will conform and adjust. Safety is an important issue that includes not only driving but controlling the kitchen, bathroom, and sleeping area environments, as well as stairways. These areas need to be monitored, supervised, and made as safe as possible. A move to a retirement home, assisted-living center, or nursing home can initially increase confusion and agitation. Repeated reassurance, reorientation, and careful introduction to the new personnel will help to smooth the process. Providing activities that are known to be enjoyable to the patient can be of considerable benefit. The clinician must pay special attention to frustration and depression among family members and caregivers. Caregiver guilt and burnout are common. Family members often feel overwhelmed and helpless and may vent their frustrations on the patient, each other, and health care providers. Caregivers should be encouraged to take advantage of day-care facilities and respite breaks.

Education and counseling about dementia are important. Local and national support groups, such as the Alzheimer's Association (*www.alz.org*), can provide considerable help.

FURTHER READINGS

Knopman DS et al: Incidence and causes of nondegenerative nonvascular dementia: A population-based study. Arch Neurol 63:218, 2006[PMID: 16476810] [Full Text]

Mayeux R: Clinical practice. Early Alzheimer's disease. N Engl J Med 362:2194, 2010[PMID: 20558370] [Full Text]

Roberson ED, Mucke L: 100 years and counting: Prospects for defeating Alzheimer's disease. Science 314:781, 2006[PMID: 17082448] [Full Text]

Small GW et al: PET of brain amyloid and tau in mild cognitive impairment. N Engl J Med 355:2652, 2007

van Es MA, van den berg LH. Alzheimer's disease beyond APOE. Nat Genet 41:1047, 2009

van Oijen M et al: Atherosclerosis and risk for dementia. Ann Neurol 61:403, 2007

Whitmer RA et al: Timing of hormone therapy and dementia: The critical window theory revisited. Ann Neurol 69:163, 2011[PMID: 21280086] [Full Text]

Williams DR et al: Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. Brain 128:1247, 2005[PMID: 15788542] [Full Text]

Zuccala G et al: Correlate of cognitive impairment among patients with heart failure: Results of a multicenter survey. Am J Med 118:496, 2005[PMID: 15866252] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PARKINSON'S DISEASE AND RELATED DISORDERS

Parkinson's disease (PD) is the second commonest neurodegenerative disease, exceeded only by Alzheimer's disease (AD). It is estimated that approximately 1 million persons in the United States and 5 million persons in the world suffer from this disorder. PD affects men and women of all races, all occupations, and all countries. The mean age of onset is about 60 years, but cases can be seen in patients in their 20s, and even younger. The frequency of PD increases with aging, and based on projected population demographics, it is estimated that the prevalence will dramatically increase in future decades.

Clinically, PD is characterized by rest tremor, rigidity, bradykinesia, and gait impairment, known as the "cardinal features" of the disease. Additional features can include freezing of gait, postural instability, speech difficulty, autonomic disturbances, sensory alterations, mood disorders, sleep dysfunction, cognitive impairment, and dementia **(Table 372-1)**, all known as nondopaminergic features because they do not fully respond to dopaminergic therapy.

Table 372-1 Clinical Features of Parkinson's Disease

Cardinal Features	Other Motor Features	Nonmotor Features
Bradykinesia	Micrographia	Anosmia
Rest tremor	Masked facies (hypomimia)equalize	Sensory disturbances (e.g., pain)
Rigidity	Reduced eye blink	Mood disorders (e.g., depression)
Gait disturbance/postural instability	Soft voice (hypophonia)	Sleep disturbances
	Dysphagia	Autonomic disturbances
	Freezing	Orthostatic hypotension
		Gastrointestinal disturbances
		Genitourinal disturbances
		Sexual dysfunction
		Cognitive impairment/Dementia

Pathologically, the hallmark features of PD are degeneration of dopaminergic neurons in the substantia nigra pars compacta (SNc), reduced striatal dopamine, and intracytoplasmic proteinaceous inclusions known as Lewy bodies **(Fig. 372-1)**. While interest has primarily focused on the dopamine system, neuronal degeneration with inclusion body formation can also affect cholinergic neurons of the nucleus basalis of Meynert (NBM), norepinephrine neurons of the locus coeruleus (LC), serotonin neurons in the raphe nuclei of the brainstem, and neurons of the olfactory system, cerebral hemispheres, spinal cord, and peripheral autonomic nervous system. This "nondopaminergic" pathology is likely responsible for the nondopaminergic clinical features listed in Table 372-1. Indeed, there is evidence that pathology begins in the peripheral autonomic nervous system, olfactory system, and dorsal motor nucleus of the vagus nerve in the lower brainstem, and then spreads in a sequential manner to affect the upper brainstem and cerebral hemispheres. These studies suggest that dopamine neurons are affected in midstage disease. Indeed, several studies suggest that symptoms reflecting nondopaminergic degeneration such as constipation, anosmia, rapid eye movement (REM) behavior sleep disorder, and cardiac denervation precede the onset of the classic motor features of the illness.

Figure 372-1

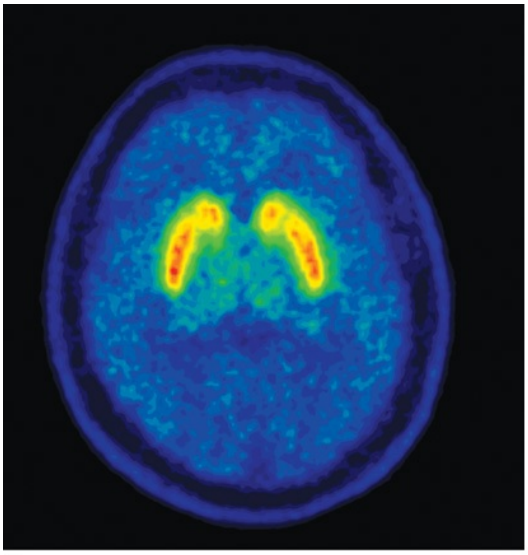

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

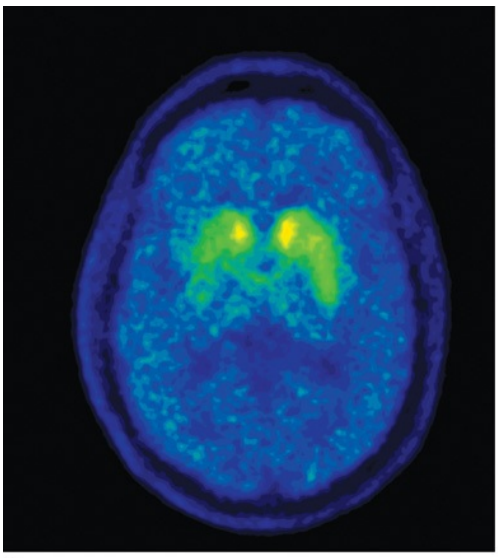

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pathologic specimens from a patient with Parkinson's disease (PD) compared to a normal control demonstrating (***A***) reduction of pigment in SNc in PD vs control, (***B***) reduced numbers of cells in SNc in PD compared to control, and (***C***) Lewy bodies (*arrows*) within melanized dopamine neurons in PD. SNc, substantia nigra pars compacta.

Differential Diagnosis

Parkinsonism is a general term that is used to define a symptom complex manifest by bradykinesia with rigidity and/or tremor. It has a wide differential diagnosis **(Table 372-2)** and can reflect damage to different components of the basal ganglia. The basal ganglia comprise a group of subcortical nuclei that include the striatum (putamen and caudate nucleus), subthalamic nucleus (STN), globus pallidus pars externa (GPe), globus pallidus pars interna (GPi), and the SNc **(Fig. 372-2)**. The basal ganglia play an important role in regulating normal motor behavior. It is now appreciated that basal ganglia also play a role in modulating emotional and cognitive functions. Among the different forms of parkinsonism, PD is the most common (approximately 75% of cases). Historically, PD was diagnosed based on the presence of two of three parkinsonian features (tremor, rigidity, bradykinesia). However, postmortem studies found a 24% error rate when these criteria were used. Clinicopathologic correlation studies subsequently determined that parkinsonism associated with rest tremor, asymmetry, and a good response to levodopa was more likely to predict the correct pathologic diagnosis. With these revised criteria (known as the U.K. brain bank criteria), the clinical diagnosis of PD is confirmed pathologically in 99% of cases.

Table 372-2 Differential Diagnosis of Parkinsonism

Parkinson's Disease	Atypical Parkinsonisms	Secondary Parkinsonism	Other Neurodegenerative Disorders
Genetic	Multiple-system atrophy	Drug-induced	Wilson's disease
Sporadic	Cerebellar type (MSA-c)	Tumor	Huntington's disease
Dementia with Lewy bodies	Parkinson type (MSA-p)	Infection	Neurodegeneration with brain iron accumulation
		Vascular	SCA 3 (spinocerebellar
		Normal-pressure hydrocephalus	

	(MSA-p) Progressive supranuclear palsy Corticobasal ganglionic degeneration Frontotemporal dementia	Trauma Liver failure Toxins (e.g., carbon monoxide, manganese, MPTP, cyanide, hexane, methanol, carbon disulfide)	ataxia) Fragile X–associated ataxia-tremor-parkinsonism Prion disease Dystonia-parkinsonism (DYT3) Alzheimer's disease with parkinsonism

Figure 372-2

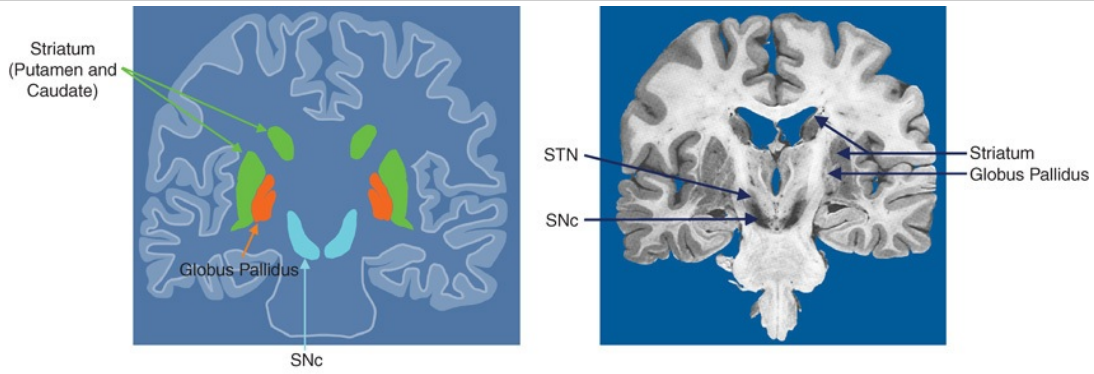

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Basal ganglia nuclei.

Schematic (***A***) and postmortem (***B***) coronal sections illustrating the various components of the basal ganglia. SNc, substantia nigra pars compacta; STN, subthalamic nucleus.

Imaging of the brain dopamine system in PD with positron emission tomography (PET) or single-photon emission computed tomography (SPECT) shows reduced uptake of striatal dopaminergic markers, particularly in the posterior putamen **(Fig. 372-3)**. Imaging can be useful in difficult cases or research studies but is rarely necessary in routine practice, as the diagnosis can usually be established on clinical criteria alone. This may change in the future when there is a disease-modifying therapy and it is important to make the diagnosis at as early a time point as possible. Genetic testing is not generally employed at present, but it can be helpful for identifying at-risk individuals in a research setting. Mutations of the *LRRK2* gene (see below) have attracted particular interest as they are the commonest cause of familial PD and are responsible for approximately 1% of typical sporadic cases of the disease. Mutations in *LRRK2* are particularly common causes of PD in Ashkenazi Jews and North African Berber Arabs. The penetrance of the most common *LRRK2* mutation ranges from 28 to 74%, depending on age. Mutations in the *parkin* gene should be considered in patients with disease onset prior to 40 years.

Figure 372-3

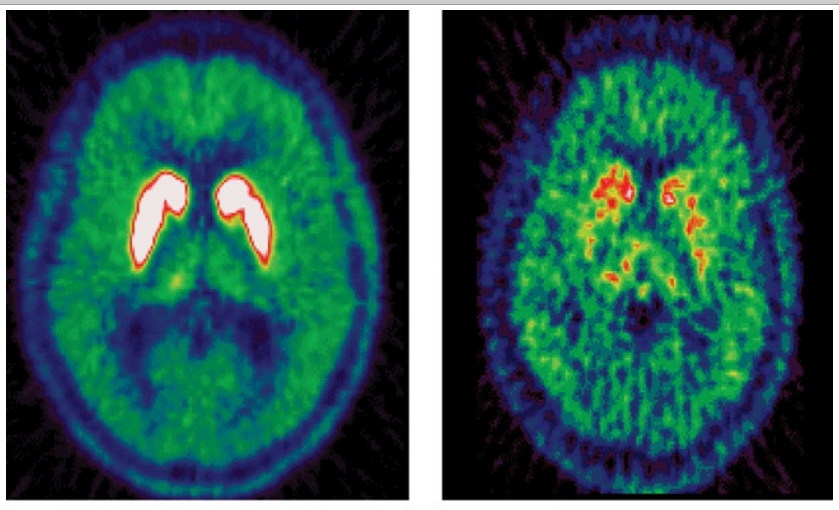

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fluorodopa-PET in a normal individual (*A*) and a PD patient (*B*).

Striatal FD-PET provides a measure of the integrity of the nigrostriatal system. Note reduced striatal uptake in PD compared to a control, which tends to be more pronounced in the caudate than in the putamen. *(Courtesy of Dr. Jon Stoessl.)*

ATYPICAL AND SECONDARY PARKINSONISM

Atypical parkinsonism refers to a group of neurodegenerative conditions that usually are associated with more widespread neurodegeneration than is found in PD (often involvement of SNc and striatum and/or pallidum). As a group, they present with a parkinsonism (rigidity and bradykinesia) but typically have a slightly different clinical picture than PD, reflecting differences in underlying pathology. Parkinsonism in these conditions is often characterized by early speech and gait impairment, absence of rest tremor, no asymmetry, poor or no response to levodopa, and an aggressive clinical course. In the early stages, they may show some modest benefit from levodopa and be difficult to distinguish from PD. Neuroimaging of the dopamine system is usually not helpful, as several atypical parkinsonisms also have degeneration of dopamine neurons. Pathologically, neurodegeneration occurs without Lewy bodies (see below for individual conditions). Metabolic imaging of the basal ganglia/thalamus network may be helpful, reflecting a pattern of decreased activity in the GPi with increased activity in the thalamus, the reverse of what is seen in PD.

Multiple-system atrophy (MSA) manifests as a combination of parkinsonian, cerebellar, and autonomic features and can be divided into a predominant parkinsonian (MSA-p) or cerebellar (MSA-c) form. Clinically, MSA is suspected when a patient presents with atypical parkinsonism in conjunction with cerebellar signs and/or early and prominent autonomic dysfunction, usually orthostatic hypotension (Chap. 375). Pathologically, MSA is characterized by degeneration of the SNc, striatum, cerebellum, and inferior olivary nuclei coupled with characteristic glial cytoplasmic inclusions (GCIs) that stain for α-synuclein. MRI can show pathologic iron accumulation in the striatum on T2-weighted scans, high signal change in the region of the external surface of the putamen (putaminal rim) in MSA-p, or cerebellar and brainstem atrophy [the pontine "hot cross buns" sign (Fig. 375-2)] in MSA-c.

Progressive supranuclear palsy (PSP) is a form of atypical parkinsonism that is characterized by slow ocular saccades, eyelid apraxia, and restricted eye movements with particular impairment of downward gaze. Patients frequently experience hyperextension of the neck with early gait disturbance and falls. In later stages, speech and swallowing difficulty and dementia become evident. MRI may reveal a characteristic atrophy of the midbrain with relative preservation of the pons (the "hummingbird sign" on midsagittal images). Pathologically, PSP is characterized by degeneration of the SNc and pallidum along with neurofibrillary tangles and GCIs that stain for tau.

Corticobasal ganglionic degeneration is less common and is usually manifest by asymmetric dystonic contractions and clumsiness of one hand coupled with cortical sensory disturbances manifest as apraxia, agnosia, focal myoclonus, or alien limb phenomenon (where the limb assumes a position in space without the patient being aware of it). Dementia may occur at any stage of the disease. MRI frequently shows asymmetric cortical atrophy. Pathologic findings include achromatic neuronal degeneration with tau deposits similar to those found in PSP.

Secondary parkinsonism can be associated with drugs, stroke, tumor, infection, or exposure to toxins such as carbon monoxide or manganese. Dopamine-blocking agents such as the neuroleptics are the commonest cause of secondary parkinsonism. These drugs are most widely used in psychiatry, but physicians should be aware that drugs such as metoclopramide and chlorperazine, which are primarily used to treat gastrointestinal problems, are also neuroleptic agents and common causes of secondary parkinsonism and tardivedyskinesia. Other drugs that can cause secondary parkinsonism include tetrabenazine, amiodarone, and lithium.

Finally, parkinsonism can be seen as a feature of other degenerative disorders such as Wilson's disease, Huntington's disease (especially the juvenile form known as Westphal variant), dopa-responsive dystonia, and neurodegenerative disorders with brain iron accumulation such as pantothenate kinase (PANK)–associated neurodegeneration (formerly known as Hallervorden-Spatz disease).

Some features that suggest parkinsonism might be due to a condition other than PD are shown in **Table 372-3**.

Table 372-3 Features Suggesting Alternate Diagnosis Than PD

Symptoms/Signs	Alternate Diagnosis to Consider
History	
Early speech and gait impairment	Atypical parkinsonism
Exposure to neuroleptics	Drug-induced parkinsonism
Onset prior to age 40	Genetic form of PD
Liver disease	Wilson's disease, non-Wilsonian hepatolenticular degeneration
Early hallucinations	Dementia with Lewy bodies
Diplopia	PSP
Poor or no response to an adequate trial of levodopa	Atypical or secondary parkinsonism
Physical Exam	
Dementia as first symptom	Dementia with Lewy bodies
Prominent orthostatic hypotension	MSA-p
Prominent cerebellar signs	MSA-c
Impairment of down gaze	PSP
High-frequency (8–10 Hz) symmetric postural tremor with a prominent kinetic component	Essential tremor

Abbreviations: MSA-c, multiple-system atrophy–cerebellar type; MSA-p, multiple-system atrophy–Parkinson type; PSP, progressive supranuclear palsy.

Etiology and Pathogenesis

Most PD cases occur sporadically (~85–90%) and are of unknown cause. Twin studies suggest that environmental factors likely play the more important role in patients older than 50 years, with genetic factors being more important in younger patients. Epidemiologic studies suggest increased risk with exposure to pesticides, rural living, and drinking well water and reduced risk with cigarette smoking and caffeine. However, no environmental factor has yet been determined to cause PD. The environmental hypothesis received a boost with the demonstration in the 1980s that MPTP (1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine), a byproduct of the illicit manufacture of a heroin-like drug, caused a PD-like syndrome in addicts in northern California. MPTP is transported to the central nervous system, where it is metabolized to form MPP^+, a mitochondrial toxin that is selectively taken up by, and damages, dopamine neurons. However, MPTP or MPTP-like

compounds have not been linked to sporadic PD. MPTP has, however, proved useful for generating an animal model of the disease. About 10–15% of cases are familial in origin, and multiple specific mutations and gene associations have been identified **(Table 372-4)**. It has been proposed that most cases of PD are due to a "double hit" involving an interaction between a gene mutation that induces susceptibility coupled with exposure to a toxic environmental factor. In this scenario, both factors are required for PD to ensue, while the presence of either one alone is not sufficient to cause the disease.

Table 372-4 Genetic Causes of PD

Name	Chromosome	Locus	Gene	Inheritance
Park 1	Chr 4	q21-23	α-Synuclein	AD
Park 2	Chr 6	q25-27	Parkin	AR
Park 3	Chr 2	p13	Unknown	AD
Park 4	Chr 4	q21-23	α-Synuclein	AD
Park 5	Chr 4	p14	UCHL-1	AD
Park 6	Chr 1	p35-36	PINK-1	AR
Park 7	Chr 1	p36	DJ-1	AR
Park 8	Chr 12	p11-q13	LRRK2	AR/Sp
Park 9	Chr 1	p36	ATP13A2	AR
Park 10	Chr 1	p32	Unknown	Sp
Park 11	Chr 2	q36-37	GIGYF2	AD
Park 12	Chr X	q21-25	Unknown	Sp
Park 13	Chr 2	p13	Omi/HtrA2	AD
Park 14	Chr 22	q13	PLA2G6	AR
Park 15	Chr 22	q12-13	FBX07	AR
Park 16	Chr 1	q32	Unknown	SP

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; SP, sporadic.

Factors that have been implicated in the pathogenesis of cell death include oxidative stress, intracellular calcium accumulation with excitotoxicity, inflammation, mitochondrial dysfunction, and proteolytic stress. Whatever the pathogenic mechanism, cell death appears to occur, at least in part, by way of a signal-mediated apoptotic or "suicidal" process. Each of these mechanisms offer potential targets for neuroprotective drugs. However, it is not clear which of these factors is primary, if the cause is the same in each case, or if one or all merely represent epiphenomena unrelated to the true cause of cell death that remains undiscovered **(Fig. 372-4)**.

Figure 372-4

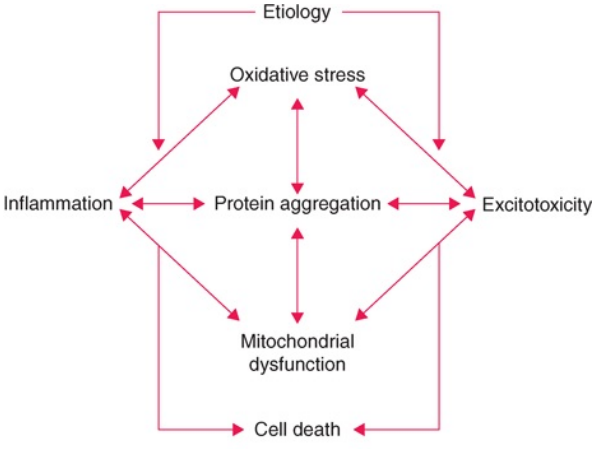

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of how pathogenetic factors implicated in PD interact in a network manner, ultimately leading to cell death. This figure illustrates how interference with any one of these factors may not necessarily stop the cell death cascade. *(Adapted from CW Olanow: Movement Disorders, 22:S-335, 2007.)*

Gene mutations discovered to date have been helpful in pointing to specific pathogenic mechanisms as being central to the neurodegenerative process. The most significant of these mechanisms appear to be protein misfolding and accumulation and mitochondrial dysfunction. The idea that proteins are involved in the pathogenesis of PD is not surprising, given that PD is characterized by Lewy bodies and Lewy neurites, which are composed of misfolded and aggregated proteins (Fig. 372-1). Protein accumulation could result from either increased formation or impaired clearance of proteins. Mutations in α-synuclein promote misfolding of the protein and the formation of oligomers and aggregates thought to be involved in the cell death process. Importantly, duplication or triplication of the wild-type α-synuclein gene can itself cause PD, indicating that increased production of even the normal protein can cause PD. Increased levels of unwanted proteins could also result from impaired clearance. Proteins are normally cleared by the ubiquitin proteasome system or the autophagy/lysosome pathway. These pathways are defective in patients with sporadic PD, and interestingly α-synuclein is a prominent component of Lewy bodies in these cases. Further, mutations in parkin (a ubiquitin ligase that attaches ubiquitin to misfolded proteins to promote their transport to the proteasome for degradation) and UCH-L1 (which cleaves ubiquitin from misfolded proteins to permit their entry into the proteasome) are causative in other cases of familial PD. Collectively, these findings implicate abnormal protein accumulation in the etiology of PD. Indeed, in laboratory models both overexpression of α-synuclein or impairment of proteasomal clearance mechanisms leads to degeneration of dopamine neurons with inclusion body formation.

Mitochondrial dysfunction has also been implicated in familial PD. Several causative genes (*parkin*, *PINK1*, and *DJ1*) either localize to mitochondria and/or cause mitochondrial dysfunction in transgenic animals. Postmortem studies have also shown a defect in complex I of the respiratory chain in the SNc of patients with sporadic PD.

Six different *LRRK2* mutations have been linked to PD, with the Gly2019Ser being the commonest. The mechanism responsible for cell death with this mutation is not known but is thought to involve altered kinase activity.

Mutations in the glucocerebrosidase (GBA) gene associated with Gaucher's disease are also associated with an increased risk of idiopathic PD. Again the mechanism is not precisely known, but it is noteworthy that it is associated with altered autophagy and lysosomal function, suggesting that mutations in this gene might also impair protein clearance leading to PD.

Whole-genome association studies have provided conflicting results. Most recently, linkage to mutations in human leukocyte antigen (HLA)genes were identified in PD patients, suggesting that altered immunity or inflammation may be a causative or contributory factor.

While gene mutations account for only a small percentage of cases of PD, it is hoped that better understanding of the mechanisms whereby they cause cell death will provide insight into the nature of the cell death process in the more common sporadic form of the disease. These mutations could also permit the development of more relevant animal models of PD in which to test putative neuroprotective drugs.

Pathophysiology of PD

The classic model of basal ganglia functional organization in the normal and PD states is provided in **Fig. 372-5**. A series of neuronal loops link the basal ganglia nuclei with corresponding cortical motor regions in a somatotopic manner to help regulate motor function. The striatum is the major input region of the basal ganglia, while the GPi and SNr are the major output regions. The input and output regions are connected via direct and indirect pathways that have reciprocal effects on the output pathway. The output of the basal ganglia provides inhibitory tone to thalamic and brainstem neurons that in turn connect to motor systems in the cerebral cortex and spinal cord to regulate motor function. Dopaminergic projections from SNc neurons serve to modulate neuronal firing and to stabilize the basal ganglia network.

Figure 372-5

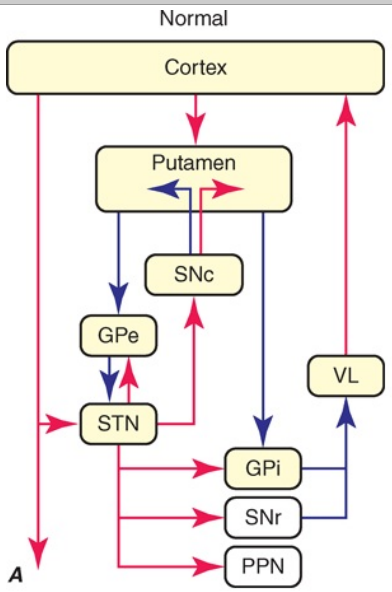

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

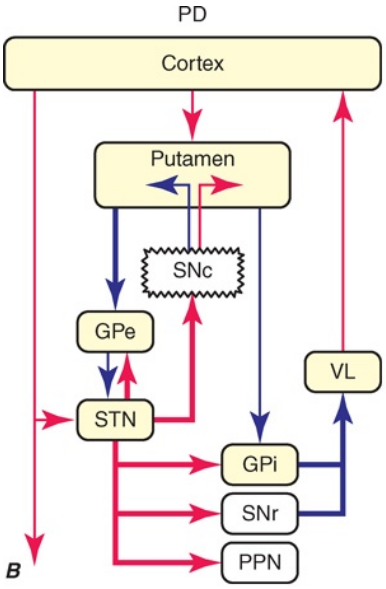

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

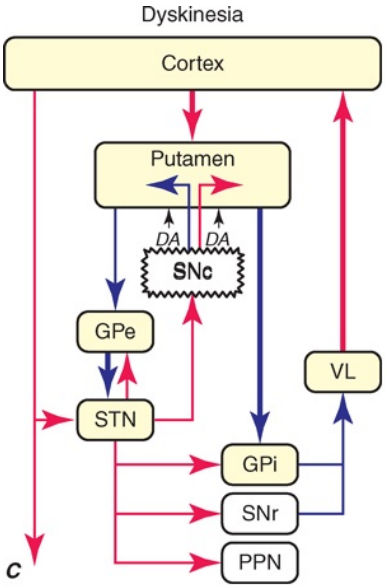

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Basal ganglia organization.

Classic model of the organization of the basal ganglia in the normal, PD, and levodopa-induced dyskinesia state. Inhibitory connections are shown as blue arrows and excitatory connections as red arrows. The striatum is the major input region and receives its major input from the cortex. The GPi and SNr are the major output regions and they project to the thalamocortical and brainstem motor regions. The striatum and GPi/SNr are connected by direct and indirect pathways. This model predicts that parkinsonism results from increased neuronal firing in the STN and GPi and that lesions or DBS of these targets might provide benefit. This concept led to the rationale for surgical therapies for PD. The model also predicts that dyskinesia results from decreased firing of the output regions, resulting in excessive cortical activation by the thalamus. This component of the model is not completely correct as lesions of the GPi ameliorate rather than increase dyskinesia in PD, suggesting that firing frequency is just one of the components that lead to the development of dyskinesia. DBS, deep brain stimulation; GPe, external segment of the globus pallidus; GPi, internal segment of the globus pallidus; SNr, substantia nigra, pars reticulata; SNc, substantia nigra, pars compacta; STN, subthalamic nucleus; VL, ventrolateral thalamus; PPN, pedunculopontine nucleus. *(Derived from JA Obeso et al: Trends Neurosci 23:S8, 2000.)*

In PD, dopamine denervation leads to increased firing of neurons in the STN and GPi, resulting in excessive inhibition of the thalamus, reduced activation of cortical motor systems, and the development of parkinsonian features (Fig. 372-5). The current role of surgery in the treatment of PD is based upon this model, which predicted that lesions or high-frequency stimulation of the STN or GPi might reduce this neuronal overactivity and improve PD features.

Treatment: Parkinson's Disease

LEVODOPA

Since its introduction in the late 1960s, levodopa has been the mainstay of therapy for PD. Experiments in the late 1950s by Carlsson demonstrated that blocking dopamine uptake with reserpine caused rabbits to become parkinsonian; this could be reversed with the dopamine precursor, levodopa. Subsequently, Hornykiewicz demonstrated a dopamine deficiency in the striatum of PD patients and suggested the potential benefit of dopaminergic replacement therapy. Dopamine does not cross the blood-brain barrier (BBB), so clinical trials were initiated with levodopa, a precursor of dopamine. Studies over the course of the next decade confirmed the value of levodopa and revolutionized the treatment of PD.

Levodopa is routinely administered in combination with a peripheral decarboxylase inhibitor to prevent its peripheral

metabolism to dopamine and the development of nausea and vomiting due to activation of dopamine receptors in the area postrema that are not protected by the BBB. In the United States, levodopa is combined with the decarboxylase inhibitor carbidopa (Sinemet), while in many other countries it is combined with benserazide (Madopar). Levodopa is also available in controlled-release formulations as well as in combination with a COMT inhibitor (see below). Levodopa remains the most effective symptomatic treatment for PD and the gold standard against which new therapies are compared. No current medical or surgical treatment provides antiparkinsonian benefits superior to what can be achieved with levodopa. Levodopa benefits the classic motor features of PD, prolongs independence and employability, improves quality of life, and increases life span. Almost all PD patients experience improvement, and failure to respond to an adequate trial should cause the diagnosis to be questioned.

There are, however, important limitations of levodopa therapy. Acute dopaminergic side effects include nausea, vomiting, and orthostatic hypotension. These are usually transient and can generally be avoided by gradual titration. If they persist, they can be treated with additional doses of a peripheral decarboxylase inhibitor (e.g., carbidopa) or a peripheral dopamine-blocking agent such as domperidone (not available in the United States). More important are motor complications (see below) that develop in the majority of patients treated long-term with levodopa therapy. In addition, features such as falling, freezing, autonomic dysfunction, sleep disorders, and dementia may emerge that are not adequately controlled by levodopa. Indeed, these nondopaminergic features are the primary source of disability and main reason for nursing home placement for patients with advanced PD.

Levodopa-induced motor complications consist of fluctuations in motor response and involuntary movements known as dyskinesias **(Fig. 372-6)**. When patients initially take levodopa, benefits are long-lasting (many hours) even though the drug has a relatively short half-life (60–90 minutes). With continued treatment, however, the duration of benefit following an individual dose becomes progressively shorter until it approaches the half-life of the drug. This loss of benefit is known as the *wearing-off effect* . At the same time, many patients develop dyskinesias. These tend to occur at the time of maximal clinical benefit and peak plasma concentration (peak-dose dyskinesia). They are usually choreiform in nature but can manifest as dystonia, myoclonus, or other movement disorders. They are not troublesome when mild, but can be disabling when severe and can limit the ability to fully utilize levodopa to control PD features. In more advanced states, patients may cycle between "on" periods complicated by disabling dyskinesias and "off" periods in which they suffer severe parkinsonism. Patients may also experience "diphasic dyskinesias," which occur as the levodopa dose begins to take effect and again as it wears off. These dyskinesias typically consist of transient, stereotypic, rhythmic movements that predominantly involve the lower extremities and are frequently associated with parkinsonism in other body regions. They can be relieved by increasing the dose of levodopa, although higher doses may induce more severe peak-dose dyskinesia.

Figure 372-6

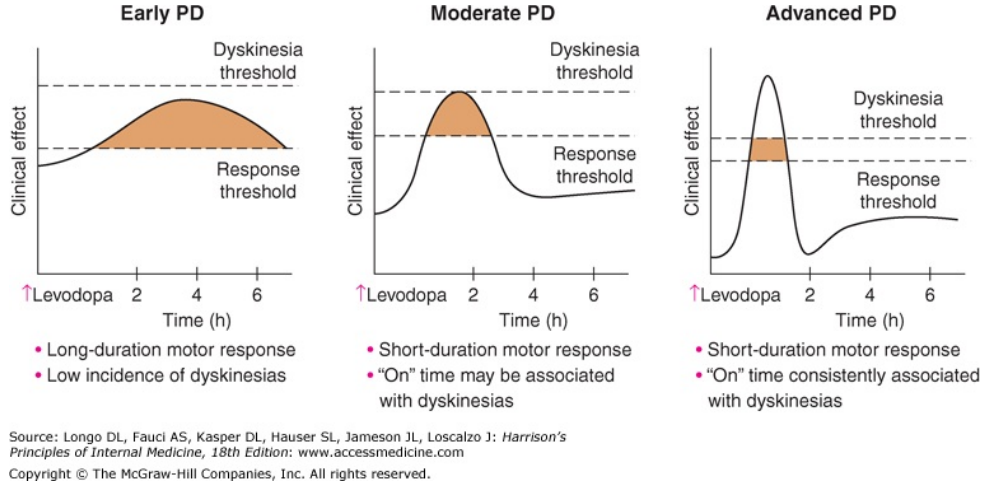

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Changes in motor response associated with chronic levodopa treatment. Levodopa-induced motor complications. Schematic illustration of the gradual shortening of the duration of a beneficial motor response to levodopa (wearing off) and the appearance of dyskinesias complicating "on" time.

The cause of levodopa-induced motor complications is not precisely known. They are more likely to occur in young individuals with severe disease and with higher doses of levodopa. The classic model of the basal ganglia has been useful for understanding the origin of motor features in PD, but has proved less valuable for understanding levodopa-induced dyskinesias (Fig. 372-5). The model predicts that dopamine replacement might excessively inhibit the pallidal output system, thereby leading to increased thalamocortical activity, enhanced stimulation of cortical motor regions, and the development of dyskinesia. However, lesions of the pallidum that completely destroy its output are associated with amelioration rather than induction of dyskinesia as suggested by the classic model. It is now thought that dyskinesia results from levodopa-induced alterations in the GPi neuronal firing pattern (pauses, bursts, synchrony, etc.) and not simply the firing frequency alone. This in turn leads to the transmission of misinformation from pallidum to thalamus/cortex, resulting in dyskinesia. Pallidotomy might thus ameliorate dyskinesia by blocking this abnormal firing pattern and preventing the transfer of misinformation to motor systems.

Current information suggests that altered neuronal firing patterns and motor complications relate to nonphysiologic levodopa replacement. Striatal dopamine levels are normally maintained at a relatively constant level. In the PD state, dopamine neurons degenerate and striatal dopamine is dependent on peripheral availability of levodopa. Intermittent doses of short-acting levodopa do not restore dopamine in a physiologic manner and cause dopamine receptors to be exposed to alternating high and low concentrations of dopamine. This intermittent or pulsatile stimulation of dopamine receptors induces molecular changes in striatal neurons and neurophysiologic changes in pallidal neurons, leading to the development of motor complications. It has been hypothesized that more continuous delivery of levodopa might prevent the development of motor complications. Indeed, continuous levodopa infusion is associated with improvement in both "off" time and dyskinesia in advanced PD patients, but this approach has not yet been proved to prevent dyskinesia in clinical trials.

Behavioral alterations can be encountered in levodopa-treated patients. A dopamine dysregulation syndrome has been described where patients have a craving for levodopa and take frequent and unnecessary doses of the drug in an addictive manner. PD patients taking high doses of levodopa can also have purposeless, stereotyped behaviors such as the

meaningless assembly and disassembly or collection and sorting of objects. This is known as punding, a term taken from the Swedish description of the meaningless behaviors seen in chronic amphetamine users. Hypersexuality and other impulse-control disorders are occasionally encountered with levodopa, although these are more commonly seen with dopamine agonists.

DOPAMINE AGONISTS

Dopamine agonists are a diverse group of drugs that act directly on dopamine receptors. Unlike levodopa, they do not require metabolism to an active product and do not undergo oxidative metabolism. Initial dopamine agonists were ergot derivatives (e.g., bromocriptine, pergolide, cabergoline) and were associated with ergot-related side effects, including cardiac valvular damage. They have largely been replaced by a second generation of non-ergot dopamine agonists (e.g., pramipexole, ropinirole, rotigotine). In general, dopamine agonists do not have comparable efficacy to levodopa. They were initially introduced as adjuncts to levodopa to enhance motor function and reduce "off" time in fluctuating patients. Subsequently, it was shown that dopamine agonists, possibly because they are relatively long-acting, are less prone than levodopa to induce dyskinesia. For this reason, many physicians initiate therapy with a dopamine agonist, although supplemental levodopa is eventually required in virtually all patients. Both ropinirole and pramipexole are available as orally administered immediate (tid) and extended-release (qd) formulations. Rotigotine is administered as a once-daily transdermal patch. Apomorphine is a dopamine agonist with efficacy comparable to levodopa, but it must be administered parenterally and has a very short half-life and duration of activity (45 min). It is generally administered SC as a rescue agent for the treatment of severe "off" episodes. Apomorphine can also be administered by continuous infusion and has been demonstrated to reduce both "off" time and dyskinesia in advanced patients. However, infusions are cumbersome, and this approach has not been approved in the United States.

Acute side effects of dopamine agonists include nausea, vomiting, and orthostatic hypotension. As with levodopa, these can usually be avoided by slow titration. Hallucinations and cognitive impairment are more common with dopamine agonists than with levodopa. Sedation with sudden unintended episodes of falling asleep while driving a motor vehicle have been reported. Patients should be informed about this potential problem and should not drive when tired. Injections of apomorphine and patch delivery of rotigotine can be complicated by development of skin lesions at sites of administration. Recently, it has become appreciated that dopamine agonists are associated with impulse-control disorders, including pathologic gambling, hypersexuality, and compulsive eating and shopping. The precise cause of these problems, and why they appear to occur more frequently with dopamine agonists than levodopa, remains to be resolved, but reward systems associated with dopamine and alterations in the ventral striatum have been implicated.

MAO-B INHIBITORS

Inhibitors of monoamine oxidase type B (MAO-B) block central dopamine metabolism and increase synaptic concentrations of the neurotransmitter. Selegiline and rasagiline are relatively selective suicide inhibitors of the MAO-B enzyme. Clinically, MAO-B inhibitors provide modest antiparkinsonian benefits when used as monotherapy in early disease, and reduced "off" time when used as an adjunct to levodopa in patients with motor fluctuations. MAO-B inhibitors are generally safe and well tolerated. They may increase dyskinesia in levodopa-treated patients but this can usually be controlled by down-titrating the dose of levodopa. Inhibition of the MAO-A isoform prevents metabolism of tyramine in the gut, leading to a potentially fatal hypertensive reaction known as a "cheese effect" as it can be precipitated by foods rich in tyramine such as some cheeses, aged meats, and red wine. Selegiline and rasagiline do not functionally inhibit MAO-A in doses employed in clinical practice and are not associated with a cheese effect. There are theoretical risks of a serotonin reaction in patients receiving concomitant SSRI antidepressants, but these are rarely encountered.

Interest in MAO-B inhibitors has also focused on their potential to have disease-modifying effects. MPTP toxicity can be prevented by coadministration of a MAO-B inhibitor that blocks its conversion to the toxic pyridinium ion MPP^+. MAO-B inhibitors also have the potential to block the oxidative metabolism of dopamine and prevent oxidative stress. In addition, both selegiline and rasagiline incorporate a propargyl ring within their molecular structure that provides antiapoptotic effects in laboratory models. The DATATOP study showed that selegiline significantly delayed the time until the emergence of disability, necessitating the introduction of levodopa in untreated PD patients. However, it could not be determined whether this was due to a neuroprotective effect that slowed disease progression or a symptomatic effect that merely masked ongoing neurodegeneration. More recently, the ADAGIO study demonstrated that early treatment with rasagiline 1 mg/d but not 2 mg/d provided benefits that could not be achieved with delayed treatment with the same drug, consistent with a disease-modifying effect; however, the long-term significance of these findings is uncertain.

COMT INHIBITORS

When levodopa is administered with a decarboxylase inhibitor, it is primarily metabolized by catechol-*O*-methyltransferase (COMT). Inhibitors of COMT increase the elimination half-life of levodopa and enhance its brain availability. Combining levodopa with a COMT inhibitor reduces "off" time and prolongs "on" time in fluctuating patients while enhancing motor scores. Two COMT inhibitors have been approved, tolcapone and entacapone. There is also a combination tablet of levodopa, carbidopa, and entacapone (Stalevo).

Side effects of COMT inhibitors are primarily dopaminergic (nausea, vomiting, increased dyskinesia) and can usually be controlled by down-titrating the dose of levodopa by 20–30%. Severe diarrhea has been described with tolcapone, and to a lesser degree with entacapone, and necessitates stopping the medication in 5–10% of individuals. Cases of fatal hepatic toxicity have been reported with tolcapone, and periodic monitoring of liver function is required. This problem has not been encountered with entacapone. Discoloration of urine can be seen with both COMT inhibitors due to accumulation of a metabolite, but it is of no clinical concern.

It has been proposed that initiating levodopa in combination with a COMT inhibitor to enhance its elimination half-life will provide more continuous levodopa delivery and reduce the risk of motor complications. While this result has been demonstrated in parkinsonian monkeys, and continuous infusion reduces "off" time and dyskinesia in advanced patients, no benefit of initiating levodopa with a COMT inhibitor compared to levodopa alone was detected in early PD patients in the STRIDE-PD study, and the main value of COMT inhibitors for now continues to be in patients who experience motor fluctuations.

OTHER MEDICAL THERAPIES

Central-acting anticholinergic drugs such as trihexyphenidyl and benztropine were used historically for the treatment for PD, but they lost favor with the introduction of dopaminergic agents. Their major clinical effect is on tremor, although it is not certain that this is superior to what can be obtained with agents such as levodopa and dopamine agonists. Still, they can be helpful in individual patients. Their use is limited particularly in the elderly, due to their propensity to induce a variety of side effects including urinary dysfunction, glaucoma, and particularly cognitive impairment.

Amantadine also has historical importance. Originally introduced as an antiviral agent, it was appreciated to also have

antiparkinsonian effects that are now thought to be due to NMDA-receptor antagonism. While some physicians use amantadine in patients with early disease for its mild symptomatic effects, it is most widely used as an antidyskinesia agent in patients with advanced PD. Indeed, it is the only oral agent that has been demonstrated in controlled studies to reduce dyskinesia while improving parkinsonian features, although benefits may be relatively transient. Side effects include livido reticularis, weight gain, and impaired cognitive function. Amantadine should always be discontinued gradually as patients can experience withdrawal symptoms.

A list of the major drugs and available dosage strengths is provided in **Table 372-5**.

Table 372-5 Drugs Commonly Used for Treatment of PD*

Agent	Available Dosages	Typical Dosing
Levodopa*		
Carbidopa/levodopa	10/100, 25/100, 25/250	200–1000 mg levodopa/d 2–4 times/d
Benserazide/levodopa	25/100, 50/200	
Carbidopa/levodopa CR Benserazide/levodopa MDS	25/100, 50/200 25/200, 25/250	
Parcopa Carbidopa/levodopa/entacapone	10/100, 25/100, 25/250 12.5/50/200,18.75/75/200,25/100/200,31.25/125/200,37.5/150/200,50/200/200	
Dopamine agonists Pramipexole	0.125, 0.25, 0.5, 1.0, 1.5 mg	0.25–1.0 mg tid
Pramipexole ER	0.375, 0.75, 1.5. 3.0, 4.5 mg	1–3 mg/d
Ropinirole Ropinirole XL	0.25, 0.5, 1.0, 3.0 mg 2, 4, 6, 8	6–24 mg/d 6–24 mg/d
Rotigotine patch Apomorphine SC	2-, 4-, 6-mg patches	4–10 mg/d 2–8 mg
COMT Inhibitors Entacapone	200 mg	200 mg with each levodopa dose
Tolcapone	100, 200 mg	100–200 mg tid
MAO-B Inhibitors Selegiline	5 mg	5 mg bid
Rasagiline	0.5, 1.0 mg	1.0 mg QAM

*Treatment should be individualized. Generally, drugs should be started in low doses and titrated to optimal dose. Note: Drugs should not be withdrawn abruptly but should be gradually lowered or removed as appropriate. ***Abbreviations:*** COMT, catechol-O-methyltransferase; MAO-B, monoamine oxidase type B.

NEUROPROTECTION

Despite the many therapeutic agents available for the treatment of PD, patients can still experience intolerable disability due to disease progression and the emergence of features such as falling and dementia that are not controlled with dopaminergic therapies. Trials of several promising agents such as rasagiline, selegiline, coenzyme Q10, pramipexole, and ropinirole have had positive results in clinical trials consistent with disease-modifying effects. However, it is not possible to determine if the positive results are due to neuroprotection with slowed disease progression or confounding symptomatic or pharmacologic effects that mask ongoing progression. If it could be determined that a drug slowed disease progression, this would be a major advance in the treatment of PD.

SURGICAL TREATMENT

Surgical treatments for PD have been employed for more than a century. Lesions placed in the motor cortex improved tremor, but were associated with motor deficits and this approach was abandoned. Subsequently, it was appreciated that lesions placed into the VIM nucleus of the thalamus reduced contralateral tremor without inducing hemiparesis, but these lesions did not meaningfully help other more disabling features of PD. Lesions placed in the GPi improved rigidity and bradykinesia as well as tremor, particularly if placed in the posteroventral portion of the nucleus. Importantly, pallidotomy was also associated with marked improvement in contralateral dyskinesia. This procedure gained favor with greater understanding of the pathophysiology of PD (see above). However, this procedure is not optimal for patients with bilateral disease, as bilateral lesions are associated with side effects such as dysphagia, dysarthria, and impaired cognition.

Most surgical procedures for PD performed today utilize deep brain stimulation (DBS). Here, an electrode is placed into the target area and connected to a stimulator inserted SC over the chest wall. DBS simulates the effects of a lesion without necessitating a brain lesion. The stimulation variables can be adjusted with respect to electrode configuration, voltage, frequency, and pulse duration in order to maximize benefit and minimize adverse side effects. In cases with intolerable side effects, stimulation can be stopped and the system removed. The procedure has the advantage that it does not require making a lesion in the brain and is thus suitable for performing bilateral procedures with relative safety.

DBS for PD primarily targets the STN or the GPi. It provides dramatic results, particularly with respect to "off" time and dyskinesias, but does not improve features that fail to respond to levodopa and does not prevent the development or progression of nondopaminergic features such as freezing, falling, and dementia. The procedure is thus primarily indicated for patients who suffer disability resulting from levodopa-induced motor complications that cannot be satisfactorily controlled with drug manipulation. Side effects can be seen with respect to the surgical procedure (hemorrhage, infarction, infection),

the DBS system (infection, lead break, lead displacement, skin ulceration), or stimulation (ocular and speech abnormalities, muscle twitches, paresthesias, depression, and rarely suicide). Recent studies indicate that benefits following DBS of the STN and GPi are comparable, but that GPi stimulation may be associated with a reduced frequency of depression. While not all PD patients are candidates, the procedure is profoundly beneficial for many. Research studies are currently examining additional targets that might benefit gait dysfunction, depression, and cognitive impairment in PD patients.

EXPERIMENTAL SURGICAL THERAPIES FOR PD

There has been considerable scientific and public interest in a number of novel therapies as possible treatments for PD. These include cell-based therapies (such as transplantation of fetal nigral dopamine cells or dopamine neurons derived from stem cells), gene therapies, and trophic factors. Transplant strategies are based on implanting dopaminergic cells into the striatum to replace degenerating SNc dopamine neurons. Fetal nigral mesencephalic cells have been demonstrated to survive implantation, reinnervate the striatum in an organotypic manner, and restore motor function in PD models. Several open-label studies reported positive results. However, two double-blind, sham surgery–controlled studies failed to show significant benefit of fetal nigral transplantation in comparison to a sham operation with respect to their primary endpoints. Post hoc analyses showed possible benefits in patients aged <60 years and in those with milder disease. It is now appreciated that grafting of fetal nigral cells is associated with a previously unrecognized form of dyskinesia that persists even after lowering or stopping levodopa. In addition, there is evidence that after many years, transplanted healthy embryonic dopamine neurons from unrelated donors can develop PD pathology, suggesting that they somehow became affected by the disease process. Most importantly, it is not clear how replacing dopamine cells alone will improve nondopaminergic features such as falling and dementia, which are the major sources of disability for patients with advanced disease. These same concerns apply to dopamine neurons derived from stem cells, which have not yet been tested in PD patients, and bear the additional theoretical concern of unanticipated side effects such as tumors. The short-term future for this technology as a treatment for PD, at least in its current state, is therefore not promising.

Gene therapy involves viral vector delivery of the DNA of a therapeutic protein to specific target regions. The DNA of the therapeutic protein can then be incorporated into the genome of host cells and thereby, in principle, provide continuous and long-lasting delivery of the therapeutic molecule. The AAV2 virus has been most often used as the viral vector because it does not promote an inflammatory response, is not incorporated into the host genome, and is associated with long-lasting transgene expression. Studies performed to date in PD have delivered aromatic amino acid decarboxylase with or without tyrosine hydroxylase to the striatum to facilitate dopamine production; glutamic acid decarboxylase to the STN to inhibit overactive neuronal firing in this nucleus; and trophic factors such as GDNF (glial-derived neurotrophic factor) and neurturin to the striatum to enhance and protect residual dopamine neurons in the SNc by way of retrograde transmission. Positive results have been reported with open-label studies, but these have not yet been confirmed in double-blind trials. While gene delivery technology has great potential, this approach also carries the risk of possible unanticipated side effects, and current approaches also do not address the nondopaminergic features of the illness.

MANAGEMENT OF THE NONMOTOR AND NONDOPAMINERGIC FEATURES OF PD

While most attention has focused on the dopaminergic features of PD, management of the nondopaminergic features of the illness should not be ignored. Some nonmotor features, while not thought to reflect dopaminergic pathology, nonetheless benefit from dopaminergic drugs. For example, problems such as anxiety, panic attacks, depression, sweating, sensory problems, freezing, and constipation all tend to be worse during "off" periods, and they improve with better dopaminergic control of the underlying PD state. Approximately 50% of PD patients suffer depression during the course of the disease that is frequently underdiagnosed and undertreated. Antiparkinsonian agents can help, but antidepressants should not be withheld, particularly for patients with major depression. Serotonin syndromes have been a theoretical concern with the combined use of selective serotonin reuptake inhibitors (SSRIs) and MAO-B inhibitors, but are rarely encountered. Anxiety can be treated with short-acting benzodiazepines.

Psychosis can be a major problem in PD. In contrast to AD, hallucinations are typically visual, formed, and nonthreatening and can limit the use of dopaminergic agents to adequately control PD features. Psychosis in PD often responds to low doses of atypical neuroleptics. Clozapine is the most effective, but it can be associated with agranulocytosis, and regular monitoring is required. For this reason, many physicians start with quetiapine even though it is not as effective as clozapine in controlled trials. Hallucinations in PD patients are often a harbinger of a developing dementia.

Dementia in PD (PDD) is common, affecting as many as 80% of patients. Its frequency increases with aging and, in contrast to AD, primarily affects executive functions and attention, with relative sparing of language, memory, and calculations. PDD is the commonest cause of nursing home placement for PD patients. When dementia precedes, or develops within 1 year after, the onset of motor dysfunction, it is by convention referred to as dementia with Lewy bodies (DLB; Chap. 371). These patients are particularly prone to have hallucinations and diurnal fluctuations. Pathologically, DLB is characterized by Lewy bodies distributed throughout the cerebral cortex (especially the hippocampus and amygdala). It is likely that DLB and PDD represent a PD spectrum rather than separate disease entities. Levodopa and other dopaminergic drugs can aggravate cognitive function in demented patients and should be stopped or reduced to try and provide a compromise between antiparkinsonian benefit and preserved cognitive function. Drugs are usually discontinued in the following sequence: anticholinergics, amantadine, dopamine agonists, COMT inhibitors, and MAO-B inhibitors. Eventually, patients with cognitive impairment should be managed with the lowest dose of standard levodopa that provides meaningful antiparkinsonian effects and does not aggravate mental function. Anticholinesterase agents such as rivastigmine and donepezil reduce the rate of deterioration of measures of cognitive function in controlled studies and can improve attention. Memantine, an antiglutamatergic agent, may also provide benefit for some PDD patients.

Autonomic disturbances are common and frequently require attention. Orthostatic hypotension can be problematic and contribute to falling. Initial treatment should include adding salt to the diet and elevating the head of the bed to prevent overnight sodium natriuresis. Low doses of fludrocortisol (Florinef) or midodrine control most cases. Vasopressin, erythropoietin, and the norepinephrine precursor 3-0-methylDOPS can be used in severe cases. If orthostatic hypotension is prominent in early disease, MSA should be considered. Sexual dysfunction can be helped with sildenafil or tadalafil. Urinary problems, especially in males, should be treated in consultation with a urologist to exclude prostate problems. Cholinergic agents, such as Ditropan, that promote bladder contraction may be helpful. Constipation can be a very important problem for PD patients. Mild laxatives can be useful, but physicians should first ensure that patients are drinking adequate amounts of fluid and consuming a diet rich in bulk with green leafy vegetables and bran. Agents that promote GI motility can also be helpful.

Sleep disturbances are common in PD patients, with many experiencing fragmented sleep with excess daytime sleepiness. Restless leg syndrome, sleep apnea, and other sleep disorders should be treated as appropriate. REM behavior disorder (RBD) may precede the onset of motor features. This syndrome is composed of violent movements and vocalizations during REM sleep, possibly representing acting out of dreams due to a failure of the normal inhibition of motor movements that typically accompanies REM sleep. Low doses of clonazepam are usually effective in controlling this problem. Consultation

with a sleep specialist and polysomnography may be necessary to identify and optimally treat sleep problems.

NONPHARMACOLOGIC THERAPY

Gait dysfunction with falling is an important cause of disability in PD. Dopaminergic therapies can help patients whose gait is worse in "off" time, but there are currently no specific therapies available. Canes and walkers may become necessary.

Freezing episodes, where patients freeze in place for seconds to minutes, are another cause of falling. Freezing during "off" periods may respond to dopaminergic therapies, but there are no specific treatments for "on" period freezing. Some patients will respond to sensory cues such as marching in place, singing a song, or stepping over an imaginary line.

Exercise, with a full range of active and passive movements, has been shown to improve and maintain function for PD patients. It is less clear that formal physical therapy is necessary, unless there is a specific indication. It is important for patients to maintain social and intellectual activities to the extent possible. Education, assistance with financial planning, social services, and attention to home safety are important elements of the overall care plan. Information is available through numerous PD foundations and on the web, but should be reviewed with physicians to ensure accuracy. The needs of the caregiver should not be neglected. Caring for a person with PD involves a substantial work effort and there is an increased incidence of depression among caregivers. Support groups for patients and caregivers may be useful.

CURRENT MANAGEMENT OF PD

The management of PD should be tailored to the needs of the individual patient, and there is no single treatment approach that is universally accepted. Clearly, if an agent could be demonstrated to have disease-modifying effects, it should be initiated at the time of diagnosis. Indeed, constipation, REM behavior disorder, and anosmia may represent pre-motor features of PD and could permit the initiation of a disease-modifying therapy even prior to onset of the classical motor features of the disease. However, no therapy has yet been proved to be disease-modifying. For now, physicians must use their judgment in deciding whether or not to introduce rasagiline (see above) or other drugs for their possible disease-modifying effects.

The next important issue to address is when to initiate symptomatic therapy. Several studies now suggest that it may be best to start therapy at the time of diagnosis in order to preserve beneficial compensatory mechanisms and possibly provide functional benefits even in the early stage of the disease. Levodopa remains the most effective symptomatic therapy for PD, and some recommend starting it immediately using relatively low doses, but many others prefer to delay levodopa treatment, particularly in younger patients, in order to reduce the risk of motor complications. Another approach is to begin with an MAO-B inhibitor and/or a dopamine agonist, and reserve levodopa for later stages when these drugs can no longer provide satisfactory control. In making this decision, the age, degree of disability, and side-effect profile of the drug must all be considered. In patients with more severe disability, the elderly, those with cognitive impairment, or where the diagnosis is uncertain, most physicians would initiate therapy with levodopa. Regardless of initial choice, it is important not to deny patients levodopa when they cannot be adequately controlled with alternative medications.

If motor complications develop, they can initially be treated by manipulating the frequency and dose of levodopa or by combining lower doses of levodopa with a dopamine agonist, a COMT inhibitor, or an MAO-B inhibitor. Amantadine is the only drug that has been demonstrated to treat dyskinesia without worsening parkinsonism, but benefits may be short-lasting and there are important side effects. In severe cases, it is usually necessary to consider a surgical therapy such as DBS if the patient is a suitable candidate, but as described above, these procedures have their own set of complications. There are ongoing efforts aimed at developing a long-acting oral or transdermal formulation of levodopa that mirrors the pharmacokinetic properties of a levodopa infusion. Such a formulation might provide all of the benefits of levodopa without motor complications and avoid the need for polypharmacy and surgical intervention.

A decision tree that considers the various treatment options and decision points for the management of PD is provided in **Fig. 372-7**.

Figure 372-7

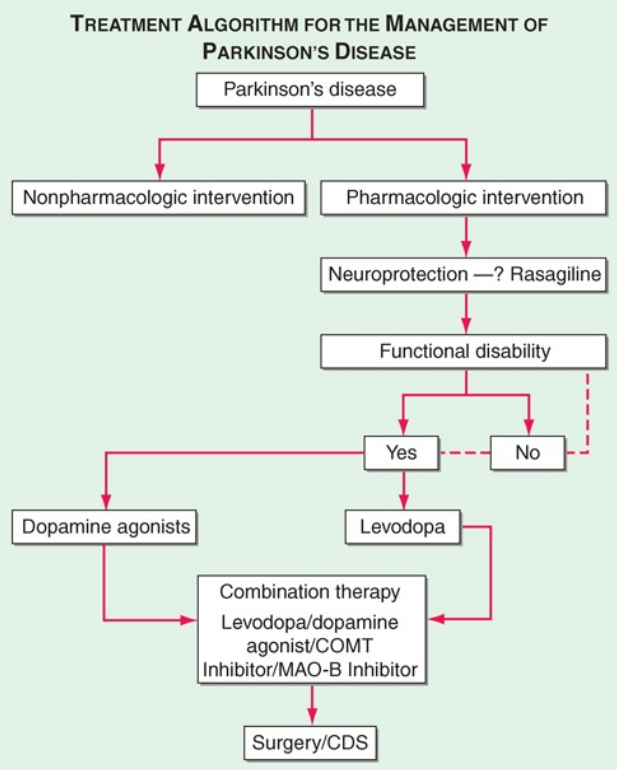

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Treatment options for the management of PD. Decision points include:

a. Introduction of a neuroprotective therapy: No drug has been established to have or is currently approved for neuroprotection or disease modification, but there are several agents that have this potential based on laboratory and preliminary clinical studies (e.g., rasagiline 1 mg/d, coenzyme Q10 1200 mg/d, the dopamine agonists ropinirole and pramipexole).

b. When to initiate symptomatic therapy: There is a trend toward initiating therapy at the time of diagnosis or early in the course of the disease because patients may have some disability even at an early stage, and there is the possibility that early treatment may preserve beneficial compensatory mechanisms; however, some experts recommend waiting until there is functional disability before initiating therapy.

c. What therapy to initiate: Many experts favor starting with an MAO-B inhibitor in mildly affected patients because of the potential for a disease-modifying effect; dopamine agonists for younger patients with functionally significant disability to reduce the risk of motor complications; and levodopa for patients with more advanced disease, the elderly, or those with cognitive impairment.

d. Management of motor complications: Motor complications are typically approached with combination therapy to try and reduce dyskinesia and enhance the "on" time. When medical therapies cannot provide satisfactory control, surgical therapies can be considered.

e. Nonpharmacologic approaches: Interventions such as exercise, education, and support should be considered throughout the course of the disease.

Source: Adapted from CW Olanow et al: Neurology 72:S1, 2009.

HYPERKINETIC MOVEMENT DISORDERS

Hyperkinetic movement disorders are characterized by involuntary movements that may occur in isolation or in combination **(Table 372-6)**. The major hyperkinetic movement disorders and the diseases with which they are associated are considered in this section.

Table 372-6 Hyperkinetic Movement Disorders

Tremor	Rhythmic oscillation of a body part due to intermittent muscle contractions
Dystonia	Involuntary patterned sustained or repeated muscle contractions often associated with twisting movements and abnormal posture.
Athetosis	Slow, distal, writhing, involuntary movements with a propensity to affect the arms and hands
Chorea	Rapid, semipurposeful, graceful, dance-like nonpatterned involuntary movements involving distal or proximal muscle groups
Myoclonus	Sudden, brief (<100 ms), jerk-like, arrhythmic muscle twitches
Tic	Brief, repeated, stereotyped muscle contractions that are often suppressible. Can be simple and involve a single muscle group or complex and affect a range of motor activities

TREMOR

Clinical Features

Tremor consists of alternating contractions of agonist and antagonist muscles in an oscillating, rhythmic manner. It can be most prominent at rest (rest tremor), on assuming a posture (postural tremor), or on actively reaching for a target (kinetic tremor). Tremor is also assessed based on distribution, frequency, and related neurologic dysfunction.

PD is characterized by a resting tremor, essential tremor (ET) by a postural tremor, and cerebellar disease by an intention or kinetic tremor. Normal individuals can have a physiologic tremor that typically manifests as a mild, high-frequency, postural or action tremor that is usually of no clinical consequence and often is only appreciated with an accelerometer. An enhanced physiologic tremor (EPT) can be seen in up to 10% of the population, often in association with anxiety, fatigue, underlying metabolic disturbance (e.g., hyperthyroidism, electrolyte abnormalities), drugs (e.g., valproate, lithium), or toxins (e.g., alcohol). Treatment is initially directed to the control of any underlying disorder and, if necessary, can often be improved with a β blocker.

ET is the commonest movement disorder, affecting approximately 5–10 million persons in the United States. It can present in childhood, but dramatically increases in prevalence over the age of 70 years. ET is characterized by a high-frequency tremor (up to 11 Hz) that predominantly affects the upper extremities. The tremor is most often manifest as a postural or kinetic tremor. It is typically bilateral and symmetric, but may begin on one side and remain asymmetric. Patients with severe ET can have an intention tremor with overshoot and slowness of movement. Tremor involves the head in ~30% of cases, voice in ~20%, tongue in ~20%, face/jaw in ~10%, and lower limbs in ~10%. The tremor is characteristically improved by alcohol and worsened by stress. Subtle impairment of coordination or tandem walking may be present. Disturbances of hearing, cognition, and even olfaction have been described, but usually the neurologic examination is normal aside from tremor. The major differential is a dystonic tremor (see below) or PD. PD can usually be differentiated from ET based on the presence of bradykinesia, rigidity, micrographia, and other parkinsonian features. However, the examiner should be aware that PD patients may have a postural tremor and ET patients may develop a rest tremor. These typically begin after a latency of a few seconds (emergent tremor). The examiner must take care to differentiate the effect of tremor on measurement of tone in ET from the cog-wheel rigidity found in PD.

Etiology and Pathophysiology

The etiology and pathophysiology of ET are not known. Approximately 50% of cases have a positive family history with an autosomal dominant pattern of inheritance. Linkage studies have detected loci at chromosomes 3q13 (ETM-1), 2p22-25 (ETM-2), and 6p23 (ETM-3). Recent genomewide studies demonstrate an association with the *LINGO1* gene, particularly in patients with young-onset ET, and it is likely that there are many other undiscovered loci. Candidate genes include the dopamine D3 receptor and proteins that map to the cerebellum. The cerebellum and inferior olives have been implicated as possible sites of a "tremor pacemaker" based on the presence of cerebellar signs and increased metabolic activity and blood flow in these regions in some patients. Recent pathologic studies have described cerebellar pathology with a loss of Purkinje's cells and axonal torpedoes. However, the precise pathologic correlate of ET remains to be defined.

Treatment

Many cases are mild and require no treatment other than reassurance. Occasionally, tremor can be severe and interfere with eating, writing, and activities of daily living. This is more likely to occur as the patient ages and is often associated with a reduction in tremor frequency. β Blockers or primidone are the standard drug therapies for ET and help in about 50% of cases. Propranolol (20–80 mg daily, given in divided doses) is usually effective at relatively low doses, but higher doses may be effective in some patients. The drug is contraindicated in patients with bradycardia or asthma. Hand tremor tends to be most improved, while head tremor is often refractory. Primidone can be helpful, but should be started at low doses (12.5 mg) and gradually increased (125–250 tid) to avoid sedation. Benefits have been reported with gabapentin and topiramate. Botulinum toxin injections may be helpful for limb or voice tremor, but treatment can be associated with secondary muscle weakness. Surgical therapies targeting the VIM nucleus of the thalamus can be very effective for severe and drug-resistant cases.

DYSTONIA

Clinical Features

Dystonia is a disorder characterized by sustained or repetitive involuntary muscle contractions frequently associated with twisting or repetitive movements and abnormal postures. Dystonia can range from minor contractions in an individual muscle group to severe and disabling involvement of multiple muscle groups. The frequency is estimated at 300,000 cases in the United States, but is likely much higher as many cases may not be recognized. Dystonia is often brought out by voluntary movements (action dystonia) and can become sustained and extend to involve other body regions. It can be aggravated by stress and fatigue, and attenuated by relaxation and sensory tricks such as touching the affected body part (geste antagoniste). Dystonia can be classified according to age of onset (childhood vs adult), distribution (focal, multifocal, segmental, or generalized), or etiology (primary or secondary).

Primary Dystonias

Several gene mutations are associated with dystonia. Idiopathic torsion dystonia (ITD) or Oppenheim's dystonia is predominantly a childhood-onset form of dystonia with an autosomal dominant pattern of inheritance that primarily affects Ashkenazi Jewish families. The majority of patients have an age of onset younger than 26 years (mean 14 years). In young-onset patients, dystonia typically begins in the foot or the arm and in 60–70% progresses to involve other limbs as well as the head and neck. In severe cases, patients can suffer disabling postural deformities that compromise mobility. Severity can vary within a family, with some affected relatives having severe disability and others a mild dystonia that may not even be appreciated. Most childhood-onset cases are linked to a mutation in the DYT1 gene located on chromosome 9q34 resulting in a trinucleotide GAG deletion with loss of one of a pair of glutamic acid residues in the protein torsin A. DYT1 mutations are found in 90% of Ashkenazi Jewish patients with ITD and probably relate to a founder effect that occurred about 350 years ago. There is variable penetrance, with only about 30% of gene carriers expressing a clinical phenotype. Why some gene carriers express dystonia and others do not is not known. The function of torsin A is unknown, but it is a member of the AAA^+ (ATPase) family that resembles heat-shock proteins and may be related to protein regulation. The precise pathology responsible for dystonia is not known.

Dopa responsive dystonia (DRD) or the Segawa variant (DYT5) is a dominantly inherited form of childhood-onset dystonia due to a mutation in the gene that encodes for GTP cyclohydrolase-I, the rate-limiting enzyme for the synthesis of tetrahydrobiopterin. This mutation leads to a defect in the biochemical synthesis of tyrosine hydroxylase, the rate-limiting enzyme in the formation of dopamine. DRD typically presents in early childhood (1–12 years), and is characterized by foot dystonia that interferes with walking. Patients often experience diurnal fluctuations, with worsening of gait as the day progresses and improvement with sleep. DRD is typified by an excellent and sustained response to small doses of levodopa. Some patients may present with parkinsonian features, but can be differentiated from juvenile PD by normal striatal fluorodopa uptake on positron emission tomography and the absence of levodopa-induced dyskinesias. DRD may occasionally be confused with cerebral palsy because patients appear to have spasticity, increased reflexes, and Babinski responses (which likely reflect a dystonic contraction rather than an upper motor neuron lesion). Any patient suspected of having a childhood-onset dystonia should receive a trial of levodopa to exclude this condition.

Mutations in the *THAP1* gene (DYT6) on chromosome 8p21q22 have been identified in Amish families and are the cause of as many as 25% of cases of non-DYT1 young-onset primary torsion dystonia. Patients are more likely to have dystonia beginning in the brachial and cervical muscles, which later can become generalized and associated with speech impairment. Myoclonic dystonia (DYT11) results from a mutation in the epsilon-sarcoglycan gene on chromosome 7q21. It typically manifests as a combination of dystonia and myoclonic jerks, frequently accompanied by psychiatric disturbances.

Focal Dystonias

These are the most common forms of dystonia. They typically present in the fourth to sixth decades and affect women more than men. The major types are (1) *blepharospasm*—dystonic contractions of the eyelids with increased blinking that can interfere with reading, watching TV, and driving. This can sometimes be so severe as to cause functional blindness. (2) *Oromandibular dystonia* (OMD)—contractions of muscles of the lower face, lips, tongue, and jaw (opening or closing). Meige syndrome is a combination of OMD and blepharospasm that predominantly affects women older than age 60 years. (3) *Spasmodic dysphonia*—dystonic contractions of the vocal cords during phonation, causing impaired speech. Most cases affect the adductor muscles and cause speech to have a choking or strained quality. Less commonly, the abductors are affected, leading to speech with a breathy or whispering quality. (4) *Cervical dystonia*—dystonic contractions of neck muscles causing the head to deviate to one side (*torticollis*), in a forward direction (*anterocollis*), or in a backward direction (*retrocollis*). Muscle contractions can be painful, and associated with a secondary cervical radiculopathy. (5) *Limb dystonias*—These can be present in either arms or legs and are often brought out by task-specific activities such as handwriting (writer's cramp), playing a musical instrument (musician's cramp), or putting (the yips). Focal dystonias can extend to involve other body regions (about 30% of cases), and are frequently misdiagnosed as psychiatric or orthopedic in origin. Their cause is not known, but genetic factors, autoimmunity, and trauma have been suggested. Focal dystonias are often associated with a high-frequency tremor that resembles ET. Dystonic tremor can usually be distinguished from ET because it tends to occur in conjunction with the dystonic contraction and disappears when the dystonia is relieved.

Secondary Dystonias

These develop as a consequence of drugs or other neurologic disorders. Drug-induced dystonia is most commonly seen with neuroleptic drugs or after chronic levodopa treatment in PD patients. Secondary dystonia can also be observed following discrete lesions in the striatum, pallidum, thalamus, cortex, and brainstem due to infarction, anoxia, trauma, tumor, infection, or toxins such as manganese or carbon monoxide. In these cases, dystonia often assumes a segmental distribution. More rarely, dystonia can develop following peripheral nerve injury and be associated with features of chronic regional pain syndrome.

Dystonia Plus Syndromes

Dystonia may occur as a part of neurodegenerative conditions such as HD, PD, Wilson's disease, CBGD, PSP, the Lubag form of dystonia-parkinsonism (DYT3), and mitochondrial encephalopathies. In contrast to the primary dystonias, dystonia is usually not the dominant neurologic feature in these conditions.

Pathophysiology of Dystonia

The pathophysiologic basis of dystonia is not known. The phenomenon is characterized by co-contracting synchronous bursts of agonist and antagonist muscle groups. This is associated with a loss of inhibition at multiple levels of the nervous system as well as increased cortical excitability and reorganization. Attention has focused on the basal ganglia as the site of origin of at least some types of dystonia as there are alterations in blood flow and metabolism in basal ganglia structures. Further, ablation or stimulation of the globus pallidus can both induce and ameliorate dystonia. The dopamine system has also been implicated, as dopaminergic therapies can both induce and treat some forms of dystonia.

Treatment: Dystonia

Treatment of dystonia is for the most part symptomatic except in rare cases where treatment of a primary underlying condition is available. Wilson's disease should be ruled out in young patients with dystonia. Levodopa should be tried in all cases of childhood-onset dystonia to rule out DRD. High-dose anticholinergics (e.g., trihexyphenidyl 20–120 mg/d) may be beneficial in children, but adults can rarely tolerate high doses because of cognitive impairment with hallucinations. Oral baclofen (20–120 mg) may be helpful, but benefits if present are usually modest and side effects of sedation, weakness, and memory loss can be problematic. Intrathecal infusion of baclofen is more likely to be helpful particularly with leg and trunk dystonia, but benefits are frequently not sustained and complications can be serious and include infection, seizures, and coma. Tetrabenazine (the usual starting dose is 12.5 mg/d and the average treating dose is 25–75 mg/d) may be helpful in some patients, but use may be limited by sedation and the development of parkinsonism. Neuroleptics can improve as well as induce dystonia, but they are typically not recommended because of their potential to induce extrapyramidal side effects, including tardive dystonia. Clonazepam and diazepam are rarely effective.

Botulinum toxin has become the preferred treatment for patients with focal dystonia, particularly where involvement is limited to small muscle groups such as in blepharospasm, torticollis, and spasmodic dysphonia. Botulinum toxin acts by blocking the release of acetylcholine at the neuromuscular junction, leading to muscle weakness and reduced dystonia, but excessive weakness may ensue and can be troublesome particularly if it involves neck and swallowing muscles. Two serotypes of botulinum toxin are available (A and B). Both are effective, and it is not clear that there are advantages of one over the other. No systemic side effects are encountered with the doses typically employed, but benefits are transient and repeat injections are required at 2- to 5-month intervals. Some patients fail to respond after having experienced an initial benefit. This has been attributed to antibody formation, but improper muscle selection, injection technique, and inadequate dose should be excluded.

Surgical therapy is an alternative for patients with severe dystonia who are not responsive to other treatments. Peripheral procedures such as rhizotomy and myotomy were used in the past to treat cervical dystonia, but are now rarely employed. DBS of the pallidum can provide dramatic benefits for patients with primary DYT1 dystonia. This represents a major therapeutic advance as previously there was no consistently effective therapy, especially for these patients who had severe disability. Benefits tend to be obtained with a lower frequency of stimulation and often occur after a relatively long latency (weeks) in comparison to PD. Better results are typically obtained in younger patients. Recent studies suggest that DBS may also be valuable for patients with focal and secondary dystonias, although results are less consistent. Supportive treatments such as physical therapy and education are important and should be a part of the treatment regimen.

Physicians should be aware of dystonic storm, a rare but potentially fatal condition that can occur in response to a stress situation such as surgery in patients with preexisting dystonia. It consists of the acute onset of generalized and persistent dystonic contractions that can involve the vocal cords or laryngeal muscles, leading to airway obstruction. Patients may experience rhabdomyolysis with renal failure. Patients should be managed in an ICU with protection of airway if required. Treatment can be instituted with one or a combination of anticholinergics, diphenhydramine, baclofen, benzodiazepines, and dopamine agonists/antagonists. Spasms may be difficult to control, and anesthesia with muscle paralysis may be required.

CHOREAS

Huntington's Disease (HD)

HD is a progressive, fatal, highly penetrant autosomal dominant disorder characterized by motor, behavioral, and cognitive dysfunction. The disease is named for George Huntington, a family physician who described cases on Long Island, New York, in the nineteenth century. Onset is typically between the ages of 25 and 45 years (range, 3–70 years) with a prevalence of 2–8 cases per 100,000 and an average age at death of 60 years. It is prevalent in Europe, North and South America, and Australia but is rare in African blacks and Asians. HD is characterized by rapid, nonpatterned, semipurposeful, involuntary choreiform movements. In the early stages, the chorea tends to be focal or segmental, but progresses over time to involve multiple body regions. Dysarthria, gait disturbance, and oculomotor abnormalities are common features. With advancing disease, there may be a reduction in chorea and emergence of dystonia, rigidity, bradykinesia, myoclonus, and spasticity. Functional decline is often predicted by progressive weight loss despite adequate calorie intake. In younger patients (about 10% of cases), HD can present as an akinetic-rigid or parkinsonian syndrome (Westphal variant). HD patients eventually develop behavioral and cognitive disturbances, and the majority progress to dementia. Depression with suicidal tendencies, aggressive behavior, and psychosis can be prominent features. HD patients may also develop non-insulin-dependent diabetes mellitus and neuroendocrine abnormalities, e.g., hypothalamic dysfunction. A clinical diagnosis of HD can be strongly suspected in cases of chorea with a positive family history. The disease predominantly strikes the striatum. Progressive atrophy of the caudate nuclei, which form the lateral margins of the lateral ventricles, can be visualized by MRI **(Fig. 372-8)**. More diffuse cortical atrophy is seen in the middle and late stages of the disease. Supportive studies include reduced metabolic activity in the caudate nucleus and putamen. Genetic testing can be used to confirm the diagnosis and to detect at-risk individuals in the family, but this must be performed with caution and in conjunction with trained counselors, as positive results can worsen depression and generate suicidal reactions. The neuropathology of HD consists of prominent neuronal loss and gliosis in the caudate nucleus and putamen; similar changes are also widespread in the cerebral cortex. Intraneuronal inclusions containing aggregates of ubiquitin and the mutant protein huntingtin are found in the nuclei of affected neurons.

Figure 372-8

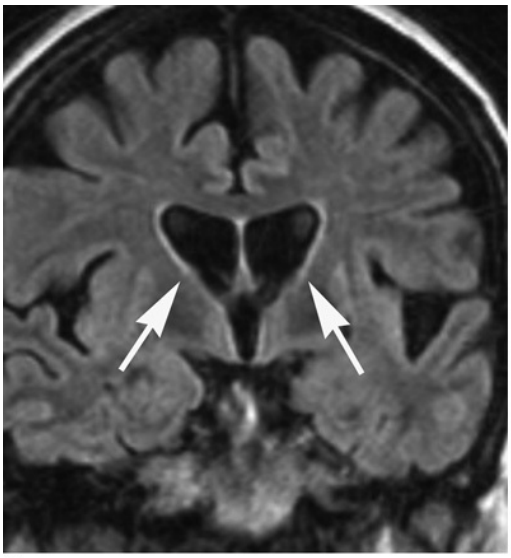

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

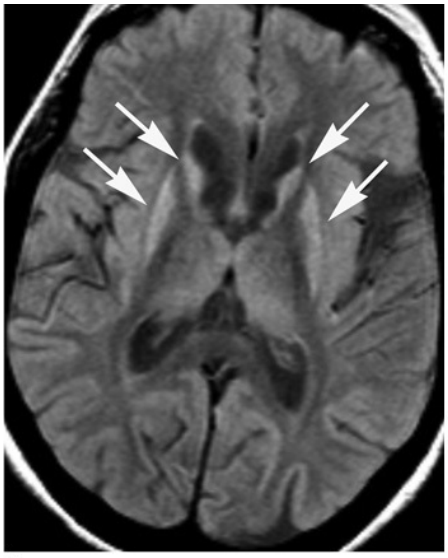

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Huntington's disease. A. Coronal FLAIR MRI shows enlargement of the lateral ventricles reflecting typical atrophy (arrows). B. Axial FLAIR image demonstrates abnormal high signal in the caudate and putamen (arrows).

Etiology

HD is caused by an increase in the number of polyglutamine (CAG) repeats (>40) in the coding sequence of the huntingtin gene located on the short arm of chromosome 4. The larger the number of repeats, the earlier the disease is manifest. Acceleration of the process tends to occur, particularly in males, with subsequent generations having larger numbers of repeats and earlier age of disease onset, a phenomenon referred to as anticipation. The gene encodes the highly conserved cytoplasmic protein huntingtin, which is widely distributed clean in neurons throughout the CNS, but whose function is not known. Models of HD with striatal pathology can be induced by excitotoxic agents such as kainic acid and 3-nitropoprionic acid, which promote calcium entry into the cell and cytotoxicity. Mitochondrial dysfunction has been demonstrated in the striatum and skeletal muscle of symptomatic and presymptomatic individuals. Fragments of the mutant huntingtin protein can be toxic, possibly by translocating into the nucleus and interfering with transcriptional upregulation of regulatory proteins. Neuronal inclusions found in affected regions in HD may represent a protective mechanism aimed at segregating and facilitating the clearance of these toxic proteins.

Treatment: Huntington's Disease

Treatment involves a multidisciplinary approach, with medical, neuropsychiatric, social, and genetic counseling for patients and their families. Dopamine-blocking agents may control the chorea. Tetrabenazine has recently been approved for the treatment of chorea in the United States, but it may cause secondary parkinsonism. Neuroleptics are generally not recommended because of their potential to induce other more troubling movement disorders and because HD chorea tends to be self-limited and is usually not disabling. Depression and anxiety can be greater problems, and patients should be treated with appropriate antidepressant and antianxiety drugs and monitored for mania and suicidal ideations. Psychosis can be treated with atypical neuroleptics such as clozapine (50–600 mg/d), quetiapine (50–600 mg/d), and risperidone (2–8 mg/d). There is no adequate treatment for the cognitive or motor decline. A neuroprotective therapy that slows or stops disease progression is the major unmet medical need in HD. Promitochondrial agents such as ubiquinone and creatine are being tested as possible disease-modifying therapies. Antiglutamate agents, caspase inhibitors, inhibitors of protein aggregation, neurotrophic factors, and transplantation of fetal striatal cells are areas of active research, but none has as yet been demonstrated to have a disease-modifying effect.

HUNTINGTON'S DISEASE–LIKE 1 (HDL1), HUNTINGTON'S DISEASE–LIKE 2 (HDL2)

HDL1 is a rare inherited disorder due to mutations of the protein located at 20p12. Patients exhibit onset of personality change in the third or fourth decade, followed by chorea, rigidity, myoclonus, ataxia, and epilepsy. HDL2 is an autosomal dominantly inherited disorder manifesting in the third or fourth decade with a variety of movement disorders, including chorea, dystonia, or parkinsonism and dementia. Most patients, are of African descent. Acanthocytosis can sometimes be seen in these patients, and they must be differentiated from neuroacanthocytosis. HDL2 is caused by an abnormally expanded CTG/CAG trinucleotide repeat expansion in the *junctophilin-3* (*JPH3*) gene on chromosome 16q24.3. The pathology of HDL2 also demonstrates intranuclear inclusions immunoreactive for ubiquitin and expanded polyglutamine repeats.

Other Choreas

Chorea can be seen in a number of disorders. Sydenham's chorea (originally called St. Vitus' dance) is more common in females and is typically seen in childhood (5–15 years). It often develops in association with prior exposure to group A streptococcal infection and is thought to be autoimmune in nature. With the reduction in the incidence of rheumatic fever, the incidence of Sydenham's chorea has fallen, but it can still be seen in developing countries. It is characterized by the acute onset of choreiform movements, behavioral disturbances, and occasionally other motor dysfunctions. Chorea generally responds to dopamine-blocking agents, valproic acid, and carbamazepine, but is self-limited and treatment is generally restricted to those with severe chorea. Chorea may recur in later life, particularly in association with pregnancy (chorea gravidarum) or treatment with sex hormones.

Chorea-acanthocytosis (neuroacanthocytosis) is a progressive and typically fatal autosomal recessive disorder that is characterized by chorea coupled with red cell abnormalities on peripheral blood smear (acanthocytes). The chorea can be severe and associated with self-mutilating behavior, dystonia, tics, seizures, and a polyneuropathy. Mutations in the *VPS13A* gene on chromosome 9q21 encoding chorein have been described. A phenotypically similar X-linked form of the disorder has been described in older individuals who have reactivity with Kell blood group antigens (McLeod syndrome). A benign hereditary chorea of childhood (BHC1) due to mutations in the gene for thyroid transcription factor 1 and a late-onset

benign senile chorea (BHC2) have also been described. It is important to ensure that patients with these types of choreas do not have HD.

A range of neurodegenerative diseases with brain iron accumulation (NBIA) manifesting with chorea and dystonia have been described including autosomal dominant neuroferritinopathy, autosomal recessive pantothenate-kinase-associated neurodegeneration (PKAN; Hallervorden-Spatz disease), and aceruloplasminemia. These disorders have excess iron accumulation on MRI and a characteristic "eye of the tiger" appearance in the globus pallidus due to iron accumulation.

Chorea may also occur in association with vascular diseases, hypo- and hyperglycemia, and a variety of infections and degenerative disorders. Systemic lupus erythematosus is the most common systemic disorder that causes chorea; the chorea can last for days to years. Choreas can also be seen with hyperthyroidism, autoimmune disorders including Sjögren's syndrome, infectious disorders including HIV disease, metabolic alterations, polycythemia rubra vera (following open-heart surgery in the pediatric population), and in association with many medications (especially anticonvulsants, cocaine, CNS stimulants, estrogens, lithium). Chorea can also be seen in paraneoplastic syndromes associated with anti-CRMP-5 or anti-Hu antibodies.

Paroxysmal dyskinesias are a group of rare disorders characterized by episodic, brief involuntary movements that can include chorea, dystonia, and ballismus. Paroxysmal kinesigenic dyskinesia (PKD) is a familial childhood-onset disorder in which chorea or chorea-dystonia is precipitated by sudden movement or running. Attacks may affect one side of the body, last seconds to minutes at a time, and recur several times a day. Prognosis is usually good, with spontaneous remission in later life. Low-dose anticonvulsant therapy (e.g., carbamazepine) is usually effective if required. Paroxysmal nonkinesigenic dyskinesia (PNKD) involves attacks of dyskinesia precipitated by alcohol, caffeine, stress, or fatigue. Like PKD, it is familial and childhood in onset and the episodes are often choreic or dystonic, but have longer duration (minutes to hours) and are less frequent (1–3/d).

Treatment: Paroxysmal Nonkinesigenic Dyskinesia

Diagnosis and treatment of the underlying condition, where possible, is the first priority. Tetrabenazine, neuroleptics, dopamine-blocking agents, propranolol, clonazepam, and baclofen may be helpful. Treatment is not indicated if the condition is mild and self-limited. Most patients with PKND do not benefit from anticonvulsant drugs but some may respond to clonazepam.

Hemiballismus

Hemiballismus is a violent form of chorea composed of wild, flinging, large-amplitude movements on one side of the body. Proximal limb muscles tend to be predominantly affected. The movements may be so severe as to cause exhaustion, dehydration, local injury, and in extreme cases, death. The most common cause is a partial lesion (infarct or hemorrhage) in the subthalamic nucleus (STN), but rare cases can also be seen with lesions in the putamen. Fortunately, hemiballismus is usually self-limiting and tends to resolve spontaneously after weeks or months. Dopamine-blocking drugs can be helpful but can themselves lead to movement disorders. In extreme cases, pallidotomy can be very effective. Interestingly, surgically induced lesions or DBS of the STN in PD are usually not associated with hemiballismus.

TICS

Tourette's Syndrome (Ts)

TS is a neurobehavioral disorder named after the French neurologist Georges Gilles de la Tourette. It predominantly affects males, and prevalence is estimated to be 0.03–1.6%, but it is likely that many mild cases do not come to medical attention. TS is characterized by multiple motor tics often accompanied by vocalizations (phonic tics). A *tic* is a brief, rapid, recurrent, and seemingly purposeless stereotyped motor contraction. Motor tics can be simple, with movement only affecting an individual muscle group (e.g., blinking, twitching of the nose, jerking of the neck), or complex, with coordinated involvement of multiple muscle groups [e.g., jumping, sniffing, head banging, and echopraxia (mimicking movements)]. Vocal tics can also be simple (e.g., grunting) or complex [e.g., echolalia (repeating other people's words), palilalia (repeating one's own words), and coprolalia (expression of obscene words)]. Patients may also experience sensory tics, composed of unpleasant focal sensations in the face, head, or neck. Patients characteristically can voluntarily suppress tics for short periods of time, but then experience an irresistible urge to express them. Tics vary in intensity and may be absent for days or weeks only to recur, occasionally in a different pattern. Tics tend to present between ages 2 and 15 years (mean 7 years) and often lessen or even disappear in adulthood. Associated behavioral disturbances include anxiety, depression, attention deficit hyperactivity disorder, and obsessive-compulsive disorder. Patients may experience personality disorders, self-destructive behaviors, difficulties in school, and impaired interpersonal relationships. Tics may present in adulthood and can be seen in association with a variety of other disorders, including PD, HD, trauma, dystonia, drugs (e.g., levodopa, neuroleptics), and toxins.

Etiology and Pathophysiology

TS is thought to be a genetic disorder, but no specific gene mutation has been identified. Current evidence supports a complex inheritance pattern, with one or more major genes, multiple loci, low penetrance, and environmental influences. The risk of a family with one affected child having a second is about 25%. The pathophysiology of TS is not known, but alterations in dopamine neurotransmission, opioids, and second-messenger systems have been proposed. Some cases of TS may be the consequence of an autoimmune response to ß-hemolytic streptococcal infection [pediatric autoimmune neuropsychiatric disorder associated with streptococcal infection (PANDAS)]; however, this remains controversial.

TREATMENT: TOURETTE'S SYNDROME

Patients with mild disease often only require education and counseling (for themselves and family members). Drug treatment is indicated when the tics are disabling and interfere with quality of life. Therapy is generally initiated with the α-agonist clonidine, starting at low doses and gradually increasing the dose and frequency until satisfactory control is achieved. Guanfacine (0.5–2 mg/d) is an α-agonist that is preferred by many clinicians because it only requires once-a-day dosing. If these agents are not effective, antipsychotics can be employed. Atypical neuroleptics (risperidone, olanzapine, ziprasidone) are preferred as they are thought to be associated with a reduced risk of extrapyramidal side effects. If they are not effective, low doses of classical neuroleptics such as haloperidol, fluphenazine, or pimozide can be tried. Botulinum toxin injections can be effective in controlling focal tics that involve small muscle groups. Behavioral features, and particularly anxiety and compulsions, can be a disabling feature of TS and should be treated. The potential value of DBS targeting the anterior portion of the internal capsule is currently being explored.

MYOCLONUS

Myoclonus is a brief, rapid (<100 ms) shock-like, jerky movement consisting of single or repetitive muscle discharges. Myoclonic jerks can be focal, multifocal, segmental, or generalized and can occur spontaneously, in association with voluntary movement (action myoclonus) or in response to an external stimulus (reflex or startle myoclonus). Negative

myoclonus consists of a twitch due to a brief loss of muscle activity (e.g., asterixis in hepatic failure). Myoclonic jerks differ from tics in that they interfere with normal movement and are not suppressible. They can be seen in association with pathology in cortical, subcortical, or spinal cord regions and associated with hypoxemic damage (especially following cardiac arrest), encephalopathy, and neurodegeneration. Reversible myoclonus can be seen with metabolic disturbances (renal failure, electrolyte imbalance, hypocalcemia), toxins, and many medications. Essential myoclonus is a relatively benign familial condition characterized by multifocal lightning-like movements. Myoclonic jerks can be disabling when they interfere with normal movement. They can also be innocent and are commonly observed in normal people when waking up or falling asleep (hypnogogic jerks).

Treatment: Myoclonus

Treatment primarily consists of treating the underlying condition or removing an offending agent. Pharmacologic therapy involves one or a combination of GABAergic agents such as valproic acid (800–3000 mg/d), piracetam (8–20 g/d), clonazepam (2–15 mg/d), or primidone (500–1000 mg/d). Recent studies suggest that levetiracetam may be particularly effective.

DRUG-INDUCED MOVEMENT DISORDERS

This important group of movement disorders is primarily associated with drugs that block dopamine receptors (neuroleptics) or central dopaminergic transmission. These drugs are primarily used in psychiatry, but it is important to appreciate that drugs used in the treatment of nausea or vomiting (e.g., Compazine) or gastroesophageal disorders (e.g., metoclopramide) are neuroleptic agents. Hyperkinetic movement disorders secondary to neuroleptic drugs can be divided into those that present acutely, subacutely, or after prolonged exposure (tardive syndromes). Dopamine-blocking drugs can also be associated with a reversible parkinsonian syndrome for which anticholinergics are often concomitantly prescribed, but there is concern that this may increase the risk of developing a tardive syndrome.

Acute

Dystonia is the most common acute hyperkinetic drug reaction. It is typically generalized in children and focal in adults (e.g., blepharospasm, torticollis, or oromandibular dystonia). The reaction can develop within minutes of exposure, and can be successfully treated in most cases with parenteral administration of anticholinergics (benztropine or diphenhydramine) or benzodiazepines (lorazepam or diazepam). Choreas, stereotypic behaviors, and tics may also be seen, particularly following acute exposure to CNS stimulants such as methylphenidate, cocaine, or amphetamines.

Subacute

Akathisia is the commonest reaction in this category. It consists of motor restlessness with a need to move that is alleviated by movement. Therapy consists of removing the offending agent. When this is not possible, symptoms may be ameliorated with benzodiazepines, anticholinergics, β blockers, or dopamine agonists.

Tardive Syndromes

These disorders develop months to years after initiation of neuroleptic treatment. Tardivedyskinesia (TD) is the commonest and is typically composed of choreiform movements involving the mouth, lips, and tongue. In severe cases, the trunk, limbs, and respiratory muscles may also be affected. In approximately one-third of patients, TD remits within 3 months of stopping the drug, and most patients gradually improve over the course of several years. In contrast, abnormal movements may develop after stopping the offending agent. The movements are often mild and more upsetting to the family than to the patient, but they can be severe and disabling, particularly in the context of an underlying psychiatric disorder. Atypical antipsychotics (e.g., clozapine, risperidone, olanzapine, quetiapine, ziprasidone, and aripiprazole) are associated with a significantly lower risk of TD in comparison to traditional antipsychotics. Younger patients have a lower risk of developing neuroleptic-induced TD, whereas the elderly, females, and those with underlying organic cerebral dysfunction have been reported to be at greater risk. In addition, chronic use is associated with increased risk, and specifically, the FDA has warned that use of metoclopramide for more than 12 weeks increases the risk of TD. Since TD can be permanent and resistant to treatment, antipsychotics should be used judiciously, atypical neuroleptics should be the preferred agent whenever possible, and the need for their continued use should be regularly monitored.

Treatment primarily consists of stopping the offending agent. If the patient is receiving a traditional antipsychotic and withdrawal is not possible, replacement with an atypical antipsychotic should be tried. Abrupt cessation of a neuroleptic should be avoided as acute withdrawal can induce worsening. TD can persist after withdrawal of antipsychotics and can be difficult to treat. Benefits may be achieved with valproic acid, anticholinergics, or botulinum toxin injections. In refractory cases, catecholamine depleters such as tetrabenazine may be helpful. Tetrabenazine can be associated with dose-dependent sedation and orthostatic hypotension. Other approaches include baclofen (40–80 mg/d), clonazepam (1–8 mg/d), or valproic acid (750–3000 mg/d).

Chronic neuroleptic exposure can also be associated with tardivedystonia with preferential involvement of axial muscles and characteristic rocking movements of the trunk and pelvis. Tardive dystonia frequently persists despite stopping medication and patients are often refractory to medical therapy. Valproic acid, anticholinergics, and botulinum toxin may occasionally be beneficial. Tardive akathisia, tardive Tourette, and tardive tremor syndromes are rare but may also occur after chronic neuroleptic exposure.

Neuroleptic medications can also be associated with a neuroleptic malignant syndrome (NMS). NMS is characterized by muscle rigidity, elevated temperature, altered mental status, hyperthermia, tachycardia, labile blood pressure, renal failure, and markedly elevated creatine kinase levels. Symptoms typically evolve within days or weeks after initiating the drug. NMS can also be precipitated by the abrupt withdrawal of antiparkinsonian medications in PD patients. Treatment involves immediate cessation of the offending antipsychotic drug and the introduction of a dopaminergic agent (e.g., a dopamine agonist or levodopa), dantrolene, or a benzodiazepine. Treatment may need to be undertaken in an intensive care setting and includes supportive measures such as control of body temperature (antipyretics and cooling blankets), hydration, electrolyte replacement, and control of renal function and blood pressure.

Drugs that have serotonin-like activity (tryptophan, MDMA or "ecstasy," meperidine) or that block serotonin reuptake can induce a rare, but potentially fatal, serotonin syndrome that is characterized by confusion, hyperthermia, tachycardia, and coma as well as rigidity, ataxia, and tremor. Myoclonus is often a prominent feature, in contrast to NMS, which it resembles. Patients can be managed with propranolol, diazepam, diphenhydramine, chlorpromazine, or cyproheptadine as well as supportive measures.

A variety of drugs can also be associated with parkinsonism (see above) and hyperkinetic movement disorders. Some examples include phenytoin (chorea, dystonia, tremor, myoclonus), carbamazepine (tics and dystonia), tricyclic antidepressants (dyskinesias, tremor, myoclonus), fluoxetine (myoclonus, chorea, dystonia), oral contraceptives (dyskinesia), β adrenergics (tremor), buspirone (akathisia, dyskinesias, myoclonus), and digoxin, cimetidine, diazoxide, lithium, methadone,

and fentanyl (dyskinesias).

RESTLESS LEGS SYNDROME

Restless legs syndrome (RLS) is a neurologic disorder that affects approximately 10% of the adult population (it is rare in Asians) and can cause significant morbidity in some. It was first described in the seventeenth century by an English physician (Thomas Willis), but has only recently been recognized as being a bona fide movement disorder. The four core symptoms required for diagnosis are as follows: an urge to move the legs, usually caused or accompanied by an unpleasant sensation in the legs; symptoms begin or worsen with rest; partial or complete relief by movement; worsening during the evening or night.

Symptoms most commonly begin in the legs, but can spread to or even begin in the upper limbs. The unpleasant sensation is often described as a creepy-crawly feeling, paresthesia, or burning. In about 80% of patients, RLS is associated with periodic leg movements (PLMs) during sleep and occasionally while awake. These involuntary movements are usually brief, lasting no more than a few seconds, and recur every 5–90 seconds. The restlessness and PLMs are a major cause of sleep disturbance in patients, leading to poor-quality sleep and daytime sleepiness.

RLS is a heterogeneous condition. Primary RLS is genetic, and several loci have been found with an autosomal dominant pattern of inheritance, although penetrance may be variable. The mean age of onset in genetic forms is 27 years, although pediatric cases are recognized. The severity of symptoms is variable. Secondary RLS may be associated with pregnancy or a range of underlying disorders, including anemia, ferritin deficiency, renal failure, and peripheral neuropathy. The pathogenesis probably involves disordered dopamine function, which may be peripheral or central, in association with an abnormality of iron metabolism. Diagnosis is made on clinical grounds but can be supported by polysomnography and the demonstration of PLMs. The neurologic examination is normal. Secondary RLS should be excluded and ferritin levels, glucose, and renal function should be measured.

Most RLS sufferers have mild symptoms that do not require specific treatment. General measures to improve sleep hygiene and quality should be attempted first. If symptoms remain intrusive, low doses of dopamine agonists, e.g., pramipexole (0.25 –0.5 mg) and ropinirole (1–2 mg), are given 1–2 hours before bedtime. Levodopa can be effective but is frequently associated with augmentation (spread and worsening of restlessness and its appearance earlier in the day) or rebound (reappearance sometimes with worsening of symptoms at a time compatible with the drug's short half-life). Other drugs that can be effective include anticonvulsants, analgesics, and even opiates. Management of secondary RLS should be directed to correcting the underlying disorder; for example, iron replacement for anemia. Iron infusion may also be helpful for severe primary RLS but requires expert supervision.

DISORDERS THAT PRESENT WITH PARKINSONISM AND HYPERKINETIC MOVEMENTS

Wilson's Disease

Wilson's disease (WD) is an autosomal recessive inherited disorder of copper metabolism that may manifest with neurologic, psychiatric, and liver disorders, alone or in combination. It is caused by mutations in the gene encoding a P-type ATPase. The disease was first comprehensively described by the English neurologist Kinnear Wilson at the beginning of the twentieth century, although at around the same time the German physicians Kayser and Fleischer separately noted the characteristic association of corneal pigmentation with hepatic and neurologic features. WD has a worldwide prevalence of approximately 1 in 30,000, with a gene carrier frequency of 1 in 90. About half of WD patients (especially younger patients) manifest with liver abnormalities. The remainder present with neurologic disease (with or without underlying liver abnormalities), and a small proportion have hematologic or psychiatric problems at disease onset.

Neurologic onset usually manifests in the second decade with tremor and rigidity. The tremor is usually in the upper limbs, bilateral, and asymmetric. Tremor can be on intention or occasionally resting and, in advanced disease, can take on a wing-beating characteristic. Other features include parkinsonism with bradykinesia, dystonia (particularly facial grimacing), dysarthria, and dysphagia. More than half of those with neurologic features have a history of psychiatric disturbances, including depression, mood swings, and overt psychosis. Kayser-Fleischer (KF) rings are seen in 80% of those with hepatic presentations and virtually all with neurologic features. KF rings represent the deposition of copper in Descemet's membrane around the cornea. They consist of a characteristic grayish rim or circle at the limbus of the cornea and are best detected by slit-lamp examination. Neuropathologic examination is characterized by neurodegeneration and astrogliosis, particularly in the basal ganglia.

WD should always be considered in the differential diagnosis of a movement disorder in a child. Low levels of blood copper and ceruloplasmin and high levels of urinary copper may be present, but normal levels do not exclude the diagnosis. CT brain scan usually reveals generalized atrophy in established cases and ~50% have hypointensity in the caudate head, globus pallidum, substantia nigra, and red nucleus. MRI shows symmetric hyperintensity on T2-weighted images in the putamen, caudate, and pallidum. However, correlation of imaging changes with clinical features is not good. It is very rare for WD patients with neurologic features not to have KF rings. Nevertheless, liver biopsy with demonstration of high copper levels remains the gold standard for the diagnosis.

In the absence of treatment, the course is progressive and leads to severe neurologic dysfunction and early death. Treatment is directed at reducing tissue copper levels and maintenance therapy to prevent reaccumulation. There is no clear consensus on treatment and all patients should be managed in a unit with expertise in WD. Penicillamine is frequently used to increase copper excretion, but it may lead to a worsening of symptoms in the initial stages of therapy. Side effects are common and can to some degree be attenuated by coadministration of pyridoxine. Tetrathiomolybdate blocks the absorption of copper and is used instead of penicillamine in many centers. Trientine and zinc are useful drugs for maintenance therapy. Effective treatment can reverse the neurologic features in most patients, particularly when started early. Some patients stabilize and a few may still progress, especially those with hepatocerebral disease. KF rings tend to decrease after 3–6 months and disappear by 2 years. Adherence to maintenance therapy is a major challenge in long-term care.

Other Disorders

Pantothenate kinase (PANK)-associated neurodegeneration, acanthocytosis, and Huntington's disease can also present with parkinsonism associated with involuntary movements.

PSYCHOGENIC DISORDERS

Virtually all movement disorders including tremor, tics, dystonia, myoclonus, chorea, ballism, and parkinsonism can be psychogenic in origin. Tremor affecting the upper limbs is the most common psychogenic movement disorder. Psychogenic movements can result from a somatoform or conversion disorder, malingering (e.g., seeking financial gain), or a factitious disorder (e.g., seeking psychological gain). Psychogenic movement disorders are common (estimated 2–3% of patients in a movement disorder clinic), more frequent in women, disabling for the patient and family, and expensive for society (estimated

$20 billion annually). Clinical features suggesting a psychogenic movement disorder include an acute onset and a pattern of abnormal movement that is inconsistent with a known movement disorder. Diagnosis is based on the nonorganic quality of the movement, the absence of findings of an organic disease process, and positive features that specifically point to a psychogenic illness such as variability and distractibility. For example, the magnitude of a psychogenic tremor is increased with attention and diminishes or even disappears when the patient is distracted by being asked to perform a different task or is unaware that he or she is being observed. Other positive features suggesting a psychogenic problem include a tremor frequency that is variable or that entrains with the frequency of movement in the contralateral limb, and a positive response to placebo medication. Associated features can include nonanatomic sensory findings, give-way weakness, and astasia-abasia (an odd, gyrating gait; Chap. 24). Comorbid psychiatric problems such as anxiety, depression, and emotional trauma may be present, but are not necessary for the diagnosis of a psychogenic movement disorder to be made. Psychogenic movement disorders can occur as an isolated entity or in association with an underlying organic problem. The diagnosis can often be made based on clinical features alone and unnecessary tests or medications avoided. Underlying psychiatric problems may be present and should be identified and treated, but many patients with psychogenic movement disorders have no obvious psychiatric pathology. Psychotherapy and hypnosis may be of value for patients with conversion reaction, and cognitive behavioral therapy may be helpful for patients with somatoform disorders. Patients with hypochondriasis, factitious disorders, and malingering have a poor prognosis.

FURTHER READINGS

Emre M et al: Clinical diagnostic criteria for dementia associated with Parkinson's disease. Mov Disord 22:1689, 2007[PMID: 17542011] [Full Text]

Fahn S, Jankovic J (eds). *Principles and Practice of Movement Disorders* . Amsterdam, Elsevier, 2010

Follett KA et al: Pallidal versus subthalamic deep-brain stimulation for Parkinson's disease. N Engl J Med 362:2077, 2010[PMID: 20519680] [Full Text]

Hardy J: Genetic analysis of pathways to Parkinson disease. Neuron 68:201, 2010[PMID: 20955928] [Full Text]

Jankovic J: Treatment of hyperkinetic movement disorders. Lancet Neurol 8:844, 2009[PMID: 19679276] [Full Text]

Olanow CW et al: A double-blind delayed-start study of rasagiline in early Parkinson's Disease. N Engl J Med 361:1268, 2009[PMID: 19776408] [Full Text]

———: Non-motor and non-dopaminergic features of Parkinson's disease. Wiley Blackwell, 2001

Ross CA and Tabrizi SJ: Huntington's disease: from molecular pathogeneis to clinical treatment. Lancet Neurol 10:83, 2011[PMID: 21163446] [Full Text]

Schapira AhV, Olanow CW (eds): *Principles of Treatment for Parkinson's Disease* . Philadelphia, Butterworth Heinemann, Elsevier, 2005

Stocchi F et al: Initiating levodopa/carbidopa therapy with and without entacapone in early Parkinson's disease: The STRIDE-PD study. Ann Neurol 68:18, 2010[PMID: 20582993] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPROACH TO THE PATIENT: ATAXIC DISORDERS

Symptoms and signs of ataxia consist of gait impairment, unclear ("scanning") speech, visual blurring due to nystagmus, hand incoordination, and tremor with movement. These result from the involvement of the cerebellum and its afferent and efferent pathways, including the spinocerebellar pathways, and the frontopontocerebellar pathway originating in the rostral frontal lobe. True cerebellar ataxia must be distinguished from ataxia associated with vestibular nerve or labyrinthine disease, as the latter results in a disorder of gait associated with a significant degree of dizziness, light-headedness, or the perception of movement (Chap. 21). True cerebellar ataxia is devoid of these vertiginous complaints and is clearly an unsteady gait due to imbalance. Sensory disturbances can also on occasion simulate the imbalance of cerebellar disease; with sensory ataxia, imbalance dramatically worsens when visual input is removed (Romberg sign). Rarely, weakness of proximal leg muscles mimics cerebellar disease. In the patient who presents with ataxia, the rate and pattern of the development of cerebellar symptoms help to narrow the diagnostic possibilities **(Table 373-1)**. A gradual and progressive increase in symptoms with bilateral and symmetric involvement suggests a genetic, metabolic, immune, or toxic etiology. Conversely, focal, unilateral symptoms with headache and impaired level of consciousness accompanied by ipsilateral cranial nerve palsies and contralateral weakness imply a space-occupying cerebellar lesion.

Table 373-1 Etiology of Cerebellar Ataxia

Symmetric and Progressive Signs			Focal and Ipsilateral Cerebellar Signs		
Acute (Hours to Days)	**Subacute (Days to Weeks)**	**Chronic (Months to Years)**	**Acute (Hours to Days)**	**Subacute (Days to Weeks)**	**Chronic (Months to Years)**
Intoxication: alcohol, lithium, phenytoin, barbiturates (positive history and toxicology screen)	Intoxication: mercury, solvents, gasoline, glue; cytotoxic chemotherapeutic, hemotherapeutic drugs	Paraneoplastic syndrome Anti-gliadin antibody syndrome Hypothyroidism	Vascular: cerebellar infarction, hemorrhage, or subdural hematoma Infectious: cerebellar abscess (mass lesion on MRI/CT, history in support of lesion)	Neoplastic: cerebellar glioma or metastatic tumor (positive for neoplasm on MRI/CT) Demyelinating: multiple sclerosis (history, CSF, and MRI are consistent)	Stable gliosis secondary to vascular lesion or demyelinating plaque (stable lesion on MRI/CT older than several months)
Acute viral cerebellitis (CSF supportive of acute viral infection) Postinfection syndrome	Alcoholic-nutritional (vitamin B1 and B12 deficiency) Lyme disease	Inherited diseases Tabes dorsalis (tertiary syphilis) Phenytoin toxicity Amiodarone		AIDS-related multifocal leukoencephalopathy (positive HIV test and CD4+ cell count for AIDS)	Congenital lesion: Chiari or Dandy-Walker malformations (malformation noted on MRI/CT)

Abbreviations: CSF, cerebrospinal fluid; CT, computed tomography; MRI, magnetic resonance imaging.

Symmetric Ataxia

Progressive and symmetric ataxia can be classified with respect to onset as acute (over hours or days), subacute (weeks or months), or chronic (months to years). Acute and reversible ataxias include those caused by intoxication with alcohol, phenytoin, lithium, barbiturates, and other drugs. Intoxication caused by toluene exposure, gasoline sniffing, glue sniffing, spray painting, or exposure to methyl mercury or bismuth are additional causes of acute or subacute ataxia, as is treatment with cytotoxic

chemotherapeutic drugs such as fluorouracil and paclitaxel. Patients with a postinfectious syndrome (especially after varicella) may develop gait ataxia and mild dysarthria, both of which are reversible (Chap. 380). Rare infectious causes of acquired ataxia include poliovirus, coxsackievirus, echovirus, Epstein-Barr virus, toxoplasmosis, *Legionella*, and Lyme disease.

The subacute development of ataxia of gait over weeks to months (degeneration of the cerebellar vermis) may be due to the combined effects of alcoholism and malnutrition, particularly with deficiencies of vitamins B_1 and B_{12}. Hyponatremia has also been associated with ataxia.

Paraneoplastic cerebellar ataxia is associated with a number of different tumors (and autoantibodies) such as breast and ovarian cancers (anti-Yo), small-cell lung cancer (anti-PQ type voltage-gated calcium channel), and Hodgkin's disease (anti-Tr) (Chap. 101). Another paraneoplastic syndrome associated with myoclonus and opsoclonus occurs with breast (anti-Ri) and lung cancers and neuroblastoma. Elevated serum anti–glutamic acid decarboxylase (GAD) antibodies have been associated with a progressive ataxic syndrome affecting speech and gait. For all of these paraneoplastic ataxias, the neurologic syndrome may be the presenting symptom of the cancer. Another immune-mediated progressive ataxia is associated with anti-gliadin (and anti-endomysium) antibodies and the human leukocyte antigen (HLA) DQB1*0201 haplotype; in some affected patients, biopsy of the small intestine reveals villus atrophy consistent with gluten-sensitive enteropathy (Chap. 294). Finally, subacute progressive ataxia may be caused by a prion disorder, especially when an infectious etiology, such as transmission from contaminated human growth hormone, is responsible (Chap. 383).

Chronic symmetric gait ataxia suggests an inherited ataxia (discussed below), a metabolic disorder, or a chronic infection. Hypothyroidism must always be considered as a readily treatable and reversible form of gait ataxia. Infectious diseases that can present with ataxia are meningovascular syphilis and tabes dorsalis due to degeneration of the posterior columns and spinocerebellar pathways in the spinal cord.

Focal Ataxia

Acute focal ataxia commonly results from cerebrovascular disease, usually ischemic infarction or cerebellar hemorrhage. These lesions typically produce cerebellar symptoms ipsilateral to the injured cerebellum and may be associated with an impaired level of consciousness due to brainstem compression and increased intracranial pressure; ipsilateral pontine signs, including sixth and seventh nerve palsies, may be present. Focal and worsening signs of acute ataxia should also prompt consideration of a posterior fossa subdural hematoma, bacterial abscess, or primary or metastatic cerebellar tumor. CT or MRI studies will reveal clinically significant processes of this type. Many of these lesions represent true neurologic emergencies, as sudden herniation, either rostrally through the tentorium or caudal herniation of cerebellar tonsils through the foramen magnum, can occur and is usually devastating. Acute surgical decompression may be required (Chap. 275). Lymphoma or progressive multifocal leukoencephalopathy (PML) in a patient with AIDS may present with an acute or subacute focal cerebellar syndrome. Chronic etiologies of progressive ataxia include multiple sclerosis (Chap. 380) and congenital lesions such as a Chiari malformation (Chap. 377) or a congenital cyst of the posterior fossa (Dandy-Walker syndrome).

THE INHERITED ATAXIAS

These may show autosomal dominant, autosomal recessive, or maternal (mitochondrial) modes of inheritance. A genomic classification **(Table 373-2)** has now largely superseded previous ones based on clinical expression alone.

Table 373-2 Classification of the Spinocerebellar Ataxias

Name	Locus	Phenotype
SCA1 (autosomal dominant type 1)	6p22-p23 with CAG repeats (exonic); leucine-rich acidic nuclear protein (LANP), region-specific interaction protein Ataxin-1	Ataxia with ophthalmoparesis, pyramidal and extrapyramidal findings; genetic testing is available; 6% of all autosomal dominant (AD) cerebellar ataxia
SCA2 (autosomal dominant type 2)	12q23-q24.1 with CAG repeats (exonic) Ataxin-2	Ataxia with slow saccades and minimal pyramidal and extrapyramidal findings; genetic testing available; 13% of all AD cerebellar ataxia
Machado-Joseph disease/SCA3 (autosomal dominant type 3)	14q24.3-q32 with CAG repeats (exonic); codes for ubiquitin protease (inactive with polyglutamine expansion); altered	Ataxia with ophthalmoparesis and variable pyramidal, extrapyramidal, and amyotrophic signs; dementia (mild); 23% of all AD cerebellar ataxia; genetic testing available

	turnover of cellular proteins due to proteosome dysfunction MJD–ataxin-3	
SCA4 (autosomal dominant type 4)	16q22.1-ter; pleckstrin homology domain-containing protein, family G, member 4; (PLEKHG4; puratrophin-1: Purkinje cell atrophy associated protein-1, including spectrin repeat and the guanine-nucleotide exchange factor, GEF for Rho GTPases)	Ataxia with normal eye movements, sensory axonal neuropathy, and pyramidal signs; genetic testing available
SCA5 (autosomal dominant type 5)	11p12-q12; β-III spectrin mutations; (SPTBN2); stabilizes glutamate transporter EAAT4; descendants of President Abraham Lincoln	Ataxia and dysarthria; genetic testing available
SCA6 (autosomal dominant type 6)	19p13.2 with CAG repeats in α1A-voltage–dependent calcium channel gene (exonic); CACNA1A protein, P/Q type calcium channel subunit	Ataxia and dysarthria, nystagmus, mild proprioceptive sensory loss; genetic testing available
SCA7 (autosomal dominant type 7)	3p14.1-p21.1 with CAG repeats (exonic); ataxin-7; subunit of GCN5, histone acetyltransferase-containing complexes; ataxin-7 binding protein; Cbl-associated protein (CAP; SH3D5)	Ophthalmoparesis, visual loss, ataxia, dysarthria, extensor plantar response, pigmentary retinal degeneration; genetic testing available
SCA8 (autosomal dominant type 8)	13q21 with CTG repeats; noncoding; 3' untranslated region of transcribed RNA; KLHL1AS	Gait ataxia, dysarthria, nystagmus, leg spasticity, and reduced vibratory sensation; genetic testing available;
SCA10 (autosomal dominant type 10)	22q13; pentanucleotide repeat ATTCT repeat; noncoding, intron 9	Gait ataxia, dysarthria, nystagmus; partial complex and generalized motor seizures; polyneuropathy; genetic testing available
SCA11 (autosomal dominant type 11)	15q14-q21.3 by linkage	Slowly progressive gait and extremity ataxia, dysarthria, vertical nystagmus, hyperreflexia
SCA12 (autosomal dominant type 12)	5q31-q33 by linkage; CAG repeat; protein phosphatase 2A, regulatory subunit B, (PPP2R2B); protein PP2A, serine/threonine phosphatase	Tremor, decreased movement, increased reflexes, dystonia, ataxia, dysautonomia, dementia, dysarthria; genetic testing available
SCA13 (autosomal dominant type 13)	19q13.3-q14.4	Ataxia, legs>arms; dysarthria, horizontal nystagmus; delayed motor development; mental developmental delay; tendon reflexes increased; MRI: cerebellar and pontine atrophy; genetic testing available;
SCA14 (autosomal dominant type 14)	19q-13.4; protein kinase Cγ (PRKCG), missense mutations including in-frame deletion and a splice site mutation among others; serine/threonine kinase	Gait ataxia; leg>arm ataxia; dysarthria; pure ataxia with later onset; myoclonus; tremor of head and extremities; increased deep tendon reflexes at ankles; occasional dystonia and sensory neuropathy; genetic testing available
SCA15 (autosomal dominant type 15)	3p24.2-3pter	Gait and extremity ataxia, dysarthria; nystagmus; MRI: superior vermis atrophy; sparing of hemispheres and tonsils
SCA16 (autosomal dominant type 16)	8q22.1-24.1	Pure cerebellar ataxia and head tremor, gait ataxia, and dysarthria; horizontal gaze –evoked nystagmus; MR: cerebellar atrophy; no brainstem changes

SCA17 (autosomal dominant type 17)	6q27; CAG expansion in the TATA-binding protein (TBP) gene	Gait ataxia, dementia, parkinsonism, dystonia, chorea, seizures; hyperreflexia; dysarthria and dysphagia; MRI shows cerebral & cerebellar atrophy; genetic testing available
SCA18 (autosomal dominant type 18)	7q22-q32	Ataxia; motor/sensory neuropathy; head tremor; dysarthria; extensor plantar responses in some patients; sensory axonal neuropathy; EMG denervation; MRI: cerebellar atrophy
SCA19 (autosomal dominant type 19)	1p21-q21	Ataxia, tremor, cognitive impairment, myoclonus; MRI: atrophy of cerebellum
SCA20 (autosomal dominant)	11p13-q11	Dysarthria; gait ataxia; ocular gaze–evoked saccades; palatal tremor; dentate calcification on CT; MRI: cerebral atrophy
SCA21 (autosomal dominant)	7p21.3-p15.1	Ataxia, dysarthria, extrapyramidal features of akinesia, rigidity, tremor, cognitive defect; reduced deep tendon reflexes; MRI: cerebellar atrophy, normal basal ganglia and brainstem
SCA22 (autosomal dominant)	1p21-q23	Pure cerebellar ataxia; dysarthria; dysphagia; nystagmus; MRI: cerebellar atrophy
SCA23 (autosomal dominant)	20p13-12.3	Gait ataxia; dysarthria; extremity ataxia; ocular nystagmus, dysmetria; leg vibration loss; extensor plantar responses; MRI: cerebellar atrophy
SCA25 (autosomal dominant)	2p15-p21	Ataxia, nystagmus; vibratory loss in the feet; pain loss in some; abdominal pain; nausea and vomiting may be prominent; absent ankle reflexes; sensory nerve action potentials are absent; MRI: cerebellar atrophy, normal brainstem
SCA26 (autosomal dominant)	19p13.3	Gait ataxia; extremity ataxia; dysarthria; nystagmus; MRI: cerebellar atrophy
SCA27 (autosomal dominant)	13q34; fibroblast growth factor 14 protein; mutation F145S; produces reduced protein stability	Tremor extremities and head and orofacial dyskinesia; ataxia of arms>legs, gait ataxia; dysarthria; nystagmus; psychiatric symptoms; cognitive defect; MRI: cerebellar atrophy; genetic testing available
SCA28 (autosomal dominant)	18p11.22-q11.2	Extremity and gait ataxia; dysarthria; nystagmus; ophthalmoparesis; leg hyperreflexia and extensor plantar responses; MRI: cerebellar atrophy
SCA30 (autosomal dominant)	4q34.3-q35.1	Candidate gene ODZ3; gait ataxia, dysarthria, saccades; nystagmus, brisk tendon reflexes in legs; MRI: cerebellar atrophy
SCA31 (autosomal dominant)	16q22.1	Penta-nucleotide (TGGAA)n repeat insertions; previously called SCA4; gait ataxia; limb dysmetria; MRI: cerebellar atrophy
Dentatorubropallidoluysian atrophy (autosomal dominant)	12p13.31 with CAG repeats (exonic) Atrophin 1	Ataxia, choreoathetosis, dystonia, seizures, myoclonus, dementia; genetic testing available
Friedreich's ataxia (autosomal recessive)	9q13-q21.1 with intronic GAA repeats, in intron at end of exon 1 Frataxin defective; abnormal regulation of mitochondrial iron metabolism; iron accumulates in mitochondria in yeast mutants	Ataxia, areflexia, extensor plantar responses, position sense deficits, cardiomyopathy, diabetes mellitus, scoliosis, foot deformities; optic atrophy; late-onset form, as late as 50 years with preserved deep tendon reflexes, slower progression, reduced skeletal deformities, associated with an intermediate number of GAA repeats and missense mutations in one allele of frataxin; genetic testing available
Friedreich's ataxia (autosomal recessive)	8q13.1-q13.3 (α-TTP deficiency)	Same as phenotype that maps to 9q but associated with vitamin E deficiency; genetic testing available

Sensory ataxic neuropathy and ophthalmoparesis (SANDO) with dysarthria (autosomal recessive)	15q25; mutations in DNA polymerase-gamma (POLG) gene that leads to mtDNA deletions	Young adult–onset ataxia, sensory neuropathy, ophthalmoparesis, hearing loss, gastric symptoms; a variant of progressive external ophthalmoplegia; MRI: cerebellar and thalamic abnormalities; mildly increased lactate and creatine kinase
Von Hippel-Lindau syndrome (autosomal dominant)	3p26-p25	Cerebellar hemangioblastoma; pheochromocytoma
Baltic myoclonus (Unverricht-Lundborg) (recessive)	21q22.3; cystatin B; extra repeats of 12–base pair tandem repeats	Myoclonus epilepsy; late-onset ataxia; responds to valproic acid, clonazepam; phenobarbital
Marinesco-Sjögren syndrome (recessive)	5q31; SIL 1 protein, nucleotide exchange factor for the heat-shock protein 70 (HSP70); chaperone HSPA5; homozygous 4-nucleotide duplication in exon 6; also compound heterozygote	Ataxia, dysarthria; nystagmus; retarded motor and mental maturation; rhabdomyolysis after viral illness; weakness; hypotonia; areflexia; cataracts in childhood; short stature; kyphoscoliosis; contractures; hypogonadism
Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS)	Chromosome 13q12; SACS gene; loss of sacsin peptide activity	Childhood onset of ataxia, spasticity, dysarthria, distal muscle wasting, foot deformity, retinal striations, mitral valve prolapse
Kearns-Sayre syndrome (sporadic)	mtDNA deletion and duplication mutations	Ptosis, ophthalmoplegia, pigmentary retinal degeneration, cardiomyopathy, diabetes mellitus, deafness, heart block, increased CSF protein, ataxia
Myoclonic epilepsy and ragged red fiber syndrome (MERRF) (maternal inheritance)	Mutation in mtDNA of the tRNAlys at 8344; also mutation at 8356	Myoclonic epilepsy, ragged red fiber myopathy, ataxia
Mitochondrial encephalopathy, lactic acidosis, and stroke syndrome (MELAS) (maternal inheritance)	tRNAleu mutation at 3243; also at 3271 and 3252	Headache, stroke, lactic acidosis, ataxia
Neuropathy; ataxia; retinitis pigmentosa (NARP)	ATPase6 (Complex 5); mtDNA point mutation at 8993	Neuropathy; ataxia; retinitis pigmentosa; dementia; seizures
Episodic ataxia, type 1 (EA-1) (autosomal dominant)	12p13; potassium voltage-gated channel gene, KCNA1; Phe249Leu mutation; variable syndrome	Episodic ataxia for minutes; provoked by startle or exercise; with facial and hand myokymia; cerebellar signs are not progressive; choreoathetotic movements; responds to phenytoin; genetic testing available
Episodic ataxia, type 2 (EA-2) (autosomal dominant)	19p-13(CACNA1A) (allelic with SCA6) (α1A-voltage –dependent calcium channel subunit); point mutations or small deletions; allelic with SCA6 and familial hemiplegic migraine	Episodic ataxia for days; provoked by stress, fatigue; with down-gaze nystagmus; vertigo; vomiting; headache; cerebellar atrophy results; progressive cerebellar signs; responds to acetazolamide; genetic testing available
Episodic ataxia, type 3 (autosomal dominant)	1q42	Episodic ataxia; 1 min to over 6 h; induced by movement; vertigo and tinnitus; headache; responds to acetazolamide
Episodic ataxia, type 4 (autosomal dominant)	Not mapped	Episodic ataxia; vertigo; diplopia; ocular slow pursuit defect; no response to acetazolamide
Episodic ataxia, type 5 (autosomal dominant)	2q22-q23; CACNB4β4 protein	Episodic ataxia; hours to weeks; seizures
Episodic ataxia, type 6	5p13; SLC1A3; glutamate transporter in astrocytes	Episodic ataxia; seizures; cognitive impairment; under 24 h
Episodic ataxia, type 7 (autosomal dominant)	19q13	Episodic ataxia; vertigo, weakness; less than 24 h
Episodic ataxia with seizures, migraine, and alternating hemiplegia	SLC1A3; 5p13; EAAT1 protein; missense mutations; glial glutamate	Ataxia, duration 2–4 days; episodic hypotonia; delayed motor milestones; seizures; migraine; alternating hemiplegia;

(autosomal dominant)	transporter (GLAST); 1047 C to G; proline to arginine	mild truncal ataxia; coma; febrile illness as a trigger; MRI: cerebellar atrophy
Fragile X tremor/ataxia syndrome (FXTAS) X-linked dominant	Xq27.3; CGG premutation expansion in FMR1 gene; expansions of 55–200 repeats in 5' UTR of the FMR-1 mRNA; presumed dominant toxic RNA effect	Late-onset ataxia with tremor, cognitive impairment, occasional parkinsonism; males typically affected, although affected females also reported; syndrome is of high concern if affected male has grandson with mental retardation (fragile X syndrome); MRI shows increased T2 signal in middle cerebellar peduncles, cerebellar atrophy, and occasional widespread brain atrophy; genetic testing available
Ataxia telangiectasia (autosomal recessive)	11q22-23; ATM gene for regulation of cell cycle; mitogenic signal transduction and meiotic recombination	Telangiectasia, ataxia, dysarthria, pulmonary infections, neoplasms of lymphatic system; IgA and IgG deficiencies; diabetes mellitus, breast cancer; genetic testing available
Early-onset cerebellar ataxia with retained deep tendon reflexes (autosomal recessive)	13q11-12	Ataxia; neuropathy; preserved deep tendon reflexes; impaired cognitive and visuospatial functions; MRI, cerebellar atrophy
Ataxia with oculomotor apraxia (AOA1) (autosomal recessive)	9p13; protein is member of histidine triad superfamily, role in DNA repair	Ataxia; dysarthria; limb dysmetria; dystonia; oculomotor apraxia; optic atrophy; motor neuropathy; late sensory loss (vibration); genetic testing available
Ataxia with oculomotor apraxia 2 (AOA2) (autosomal recessive)	9q34; senataxin protein, involved in RNA maturation and termination; helicase superfamily 1	Gait ataxia; choreoathetosis; dystonia; oculomotor apraxia; neuropathy, vibration loss, position sense loss, and mild light touch loss; absent leg deep tendon reflexes; extensor plantar response; genetic testing available
Cerebellar ataxia with muscle coenzyme Q10 deficiency (autosomal recessive)	9p13	Ataxia; hypotonia; seizures; mental retardation; increased deep tendon reflexes; extensor plantar responses; coenzyme Q10 levels reduced with about 25% of patients with a block in transfer of electrons to complex 3; may respond to coenzyme 10
Joubert syndrome (autosomal recessive)	9q34.3	Ataxia; ptosis; mental retardation; oculomotor apraxia; nystagmus; retinopathy; rhythmic tongue protrusion; episodic hyperpnea or apnea; dimples at wrists and elbows; telecanthus; micrognathia
Sideroblastic anemia and spinocerebellar ataxia (X-linked recessive)	Xq13; ATP-binding cassette 7 (ABCB7; ABC7) transporter; mitochondrial inner membrane; iron homeostasis; export from matrix to the intermembrane space	Ataxia; elevated free erythrocyte protoporphyrin levels; ring sideroblasts in bone marrow; heterozygous females may have mild anemia but not ataxia
Infantile-onset spinocerebellar ataxia of Nikali et al (autosomal recessive)	10q23.3-q24.1; twinkle protein (gene); homozygous for Tyr508Cys missense mutations	Infantile ataxia, sensory neuropathy; athetosis, hearing deficit, reduced deep tendon reflexes; ophthalmoplegia, optic atrophy; seizures; primary hypogonadism in females
Hypoceruloplasminemia with ataxia and dysarthria (autosomal recessive)	Ceruloplasmin gene; 3q23-q25 (trp 858 ter)	Gait ataxia and dysarthria; hyperreflexia; cerebellar atrophy by MRI; iron deposition in cerebellum, basal ganglia, thalamus, and liver; onset in the 4th decade
Spinocerebellar ataxia with neuropathy (SCAN1) (autosomal recessive)	Tyrosyl-DNA phosphodiesterase-1 (TDP-1) 14q31-q32	Onset in 2nd decade; gait ataxia, dysarthria, seizures, cerebellar vermis atrophy on MRI, dysmetria

Abbreviations: CSF, cerebrospinal fluid; EMG, electromyogram; MRI, magnetic resonance imaging.

Although the clinical manifestations and neuropathologic findings of cerebellar disease dominate the clinical picture, there may also be characteristic changes in the basal ganglia, brainstem, spinal cord, optic nerves, retina, and peripheral nerves. In large families with dominantly inherited ataxias, many gradations are observed from purely cerebellar manifestations to mixed cerebellar and brainstem disorders, cerebellar and basal ganglia syndromes, and spinal cord or peripheral nerve disease. Rarely, dementia is present as well. The clinical picture may be homogeneous within a family with

dominantly inherited ataxia, but sometimes most affected family members show one characteristic syndrome, while one or several members have an entirely different phenotype.

Autosomal Dominant Ataxias

The autosomal spinocerebellar ataxias (SCAs) include SCA types 1 through SCA28, dentatorubropallidoluysian atrophy (DRPLA), and episodicataxia (EA) types 1 and 2 (Table 373-2). SCA1, SCA2, SCA3 [Machado-Joseph disease (MJD)], SCA6, SCA7, and SCA17 are caused by CAG triplet repeat expansions in different genes. SCA8 is due to an untranslated CTG repeat expansion, SCA12 is linked to an untranslated CAG repeat, and SCA10 is caused by an untranslated pentanucleotide repeat. The clinical phenotypes of these SCAs overlap. The genotype has become the gold standard for diagnosis and classification. CAG encodes glutamine, and these expanded CAG triplet repeat expansions result in expanded polyglutamine proteins, termed *ataxins*, that produce a toxic gain of function with autosomal dominant inheritance. Although the phenotype is variable for any given disease gene, a pattern of neuronal loss with gliosis is produced that is relatively unique for each ataxia. Immunohistochemical and biochemical studies have shown cytoplasmic (SCA2), neuronal (SCA1, MJD, SCA7), and nucleolar (SCA7) accumulation of the specific mutant polyglutamine-containing ataxin proteins. Expanded polyglutamine ataxins with more than ~40 glutamines are potentially toxic to neurons for a variety of reasons including the following: high levels of gene expression for the mutant polyglutamine ataxin in affected neurons; conformational change of the aggregated protein to a ß-pleated structure; abnormal transport of the ataxin into the nucleus (SCA1, MJD, SCA7); binding to other polyglutamine proteins, including the TATA-binding transcription protein and the CREB-binding protein, impairing their functions; altering the efficiency of the ubiquitin-proteosome system of protein turnover; and inducing neuronal apoptosis. An earlier age of onset (anticipation) and more aggressive disease in subsequent generations are due to further expansion of the CAG triplet repeat and increased polyglutamine number in the mutant ataxin. The most common disorders are discussed below.

Sca1

SCA1 was previously referred to as *olivopontocerebellar atrophy* , but genomic data have shown that that entity represents several different genotypes with overlapping clinical features.

SYMPTOMS AND SIGNS

SCA1 is characterized by the development in early or middle adult life of progressive cerebellar ataxia of the trunk and limbs, impairment of equilibrium and gait, slowness of voluntary movements, scanning speech, nystagmoid eye movements, and oscillatory tremor of the head and trunk. Dysarthria, dysphagia, and oculomotor and facial palsies may also occur. Extrapyramidal symptoms include rigidity, an immobile face, and parkinsonian tremor. The reflexes are usually normal, but knee and ankle jerks may be lost, and extensor plantar responses may occur. Dementia may be noted but is usually mild. Impairment of sphincter function is common, with urinary and sometimes fecal incontinence. Cerebellar and brainstem atrophy are evident on MRI **(Fig. 373-1)**.

Figure 371-1

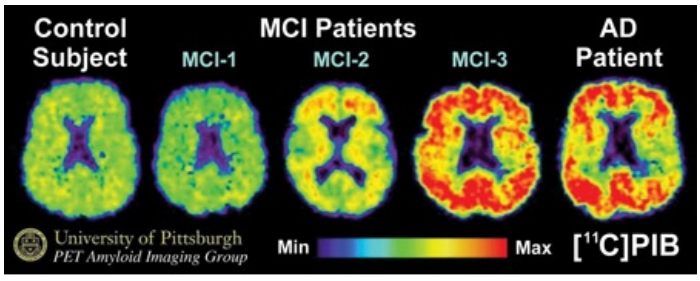

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sagittal MRI of the brain of a 60-year-old man with gait ataxia and dysarthria due to SCA1, illustrating cerebellar atrophy (*arrows*).MRI, magnetic resonance imaging; SCA1, spinocerebellar ataxia type 1.

Marked shrinkage of the ventral half of the pons, disappearance of the olivary eminence on the ventral surface of the medulla, and atrophy of the cerebellum are evident on gross postmortem inspection of the brain. Variable loss of Purkinje cells, reduced numbers of cells in the molecular and granular layer, demyelination of the middle cerebellar peduncle and the cerebellar hemispheres, and severe loss of cells in the pontine nuclei and olives are found on histologic examination. Degenerative changes in the striatum, especially the putamen, and loss of the pigmented cells of the substantia nigra may be found in cases with extrapyramidal features. More widespread degeneration in the central nervous system (CNS), including involvement of the posterior columns and the spinocerebellar fibers, is often present.

Genetic Considerations

SCA1 encodes a gene product, called *ataxin-1*, which is a novel protein of unknown function. The mutant allele has 40 CAG repeats located within the coding region, whereas alleles from unaffected individuals have ≤36 repeats. A few patients with 38–40 CAG repeats have been described. There is a direct correlation between a larger number of repeats and a younger age of onset for SCA1. Juvenile patients have higher numbers of repeats, and anticipation is present in subsequent generations. Transgenic mice carrying SCA1 developed ataxia and Purkinje cell pathology. Nuclear localization, but not aggregation, of ataxin-1 appears to be required for cell death initiated by the mutant protein.

Sca2

SYMPTOMS AND SIGNS

Another clinical phenotype, SCA2, has been described in patients from Cuba and India. Cuban patients probably are descendants of a common ancestor, and the population may be the largest homogeneous group of patients with ataxia yet described. The age of onset ranges from 2–65 years, and there is considerable clinical variability within families. Although neuropathologic and clinical findings are compatible with a diagnosis of SCA1, including slow saccadic eye movements, ataxia, dysarthria, parkinsonian rigidity, optic disc pallor, mild spasticity, and retinal degeneration, SCA2 is a unique form of cerebellar degenerative disease.

Genetic Considerations

The gene in SCA2 families also contains CAG repeat expansions coding for a polyglutamine-containing protein, ataxin-2. Normal alleles contain 15–32 repeats; mutant alleles have 35–77 repeats.

Machado-Joseph Disease/Sca3

MJD was first described among the Portuguese and their descendants in New England and California. Subsequently, MJD has been found in families from Portugal, Australia, Brazil, Canada, China, England, France, India, Israel, Italy, Japan, Spain, Taiwan, and the United States. In most populations, it is the most common autosomal dominant ataxia.

SYMPTOMS AND SIGNS

MJD has been classified into three clinical types. In type I MJD (amyotrophic lateral sclerosis –parkinsonism–dystonia type), neurologic deficits appear in the first two decades and involve weakness and spasticity of extremities, especially the legs, often with dystonia of the face, neck, trunk, and extremities. Patellar and ankle clonus are common, as are extensor plantar responses. The gait is slow and stiff, with a slightly broadened base and lurching from side to side; this gait results from spasticity, not true ataxia. There is no truncal titubation. Pharyngeal weakness and spasticity cause difficulty with speech and swallowing. Of note is the prominence of horizontal and vertical nystagmus, loss of fast saccadic eye movements, hypermetric and hypometric saccades, and impairment of upward vertical gaze. Facial fasciculations, facial myokymia, lingual fasciculations without atrophy, ophthalmoparesis, and ocular prominence are common early manifestations.

In type II MJD (ataxic type), true cerebellar deficits of dysarthria and gait and extremity ataxia begin in the second to fourth decades along with corticospinal and extrapyramidal deficits of spasticity, rigidity, and dystonia. Type II is the most common form of MJD. Ophthalmoparesis, upward vertical gaze deficits, and facial and lingual fasciculations are also present. Type II MJD can be distinguished from the clinically similar disorders SCA1 and SCA2.

Type III MJD (ataxic-amyotrophic type) presents in the fifth to the seventh decades with a pancerebellar disorder that includes dysarthria and gait and extremity ataxia. Distal sensory loss involving pain, touch, vibration, and position senses and distal atrophy are prominent, indicating the presence of peripheral neuropathy. The deep tendon reflexes are depressed to absent, and there are no corticospinal or extrapyramidal findings.

The mean age of onset of symptoms in MJD is 25 years. Neurologic deficits invariably progress and lead to death from debilitation within 15 years of onset, especially in patients with types I and II disease. Usually, patients retain full intellectual function.

The major pathologic findings are variable loss of neurons and glial replacement in the corpus striatum and severe loss of neurons in the pars compacta of the substantia nigra. A moderate loss of neurons occurs in the dentate nucleus of the cerebellum and in the red nucleus. Purkinje cell loss and granule cell loss occur in the cerebellar cortex. Cell loss also occurs in the dentate nucleus and in the cranial nerve motor nuclei. Sparing of the inferior olives distinguishes MJD from other dominantly inherited ataxias.

Genetic Considerations

The gene for MJD maps to 14q24.3-q32. Unstable CAG repeat expansions are present in the MJD gene coding for a polyglutamine-containing protein named ataxin-3, or MJD-ataxin. An earlier age of onset is associated with longer repeats. Alleles from normal individuals have between 12 and 37 CAG repeats, while MJD alleles have 60–84 CAG repeats. Polyglutamine-containing aggregates of ataxin-3

(MJD-ataxin) have been described in neuronal nuclei undergoing degeneration. MJD-ataxin codes for a ubiquitin protease, which is inactive due to expanded polyglutamines. Proteosome function is impaired, resulting in altered clearance of proteins and cerebellar neuronal loss.

Sca6

Genomic screening for CAG repeats in other families with autosomal dominant ataxia and vibratory and proprioceptive sensory loss have yielded another locus. Of interest is that different mutations in the same gene for the α_{1A} voltage-dependent calcium channel subunit (CACNLIA4; also referred to as the *CACNA1A* gene) at 19p13 result in different clinical disorders. CAG repeat expansions (21–27 in patients; 4–16 triplets in normal individuals) result in late-onset progressive ataxia with cerebellar degeneration. Missense mutations in this gene result in familial hemiplegic migraine. Nonsense mutations resulting in termination of protein synthesis of the gene product yield hereditary paroxysmal cerebellar ataxia or EA. Some patients with familial hemiplegic migraine develop progressive ataxia and also have cerebellar atrophy.

Sca7

This disorder is distinguished from all other SCAs by the presence of retinal pigmentary degeneration. The visual abnormalities first appear as blue-yellow color blindness and proceed to frank visual loss with macular degeneration. In almost all other respects, SCA7 resembles several other SCAs in which ataxia is accompanied by various noncerebellar findings, including ophthalmoparesis and extensor plantar responses. The genetic defect is an expanded CAG repeat in the SCA7 gene at 3p14-p21.1. The expanded repeat size in SCA7 is highly variable. Consistent with this, the severity of clinical findings varies from essentially asymptomatic to mild late-onset symptoms to severe, aggressive disease in childhood with rapid progression. Marked anticipation has been recorded, especially with paternal transmission. The disease protein, ataxin-7, forms aggregates in nuclei of affected neurons, as has also been described for SCA1 and SCA3/MJD.

Sca8

This form of ataxia is caused by a CTG repeat expansion in an untranslated region of a gene on chromosome 13q21. There is marked maternal bias in transmission, perhaps reflecting contractions of the repeat during spermatogenesis. The mutation is not fully penetrant. Symptoms include slowly progressive dysarthria and gait ataxia beginning at ~40 years of age with a range between 20 and 65 years. Other features include nystagmus, leg spasticity, and reduced vibratory sensation. Severely affected individuals are nonambulatory by the fourth to sixth decades. MRI shows cerebellar atrophy. The mechanism of disease may involve a dominant "toxic" effect occurring at the RNA level, as occurs in myotonic dystrophy.

Dentatorubropallidoluysian Atrophy

DRPLA has a variable presentation that may include progressive ataxia, choreoathetosis, dystonia, seizures, myoclonus, and dementia. DRPLA is due to unstable CAG triplet repeats in the open reading frame of a gene named *atrophin* located on chromosome 12p12-ter. Larger expansions are found in patients with earlier onset. The number of repeats is 49 in patients with DRPLA and ≤26 in normal individuals. Anticipation occurs in successive generations, with earlier onset of disease in association with an increasing CAG repeat number in children who inherit the disease from their father. One well-characterized family in North Carolina has a phenotypic variant known as the *Haw River syndrome* , now recognized to be due to the DRPLA mutation.

Episodic Ataxia

EA types 1 and 2 are two rare dominantly inherited disorders that have been mapped to chromosomes 12p (a potassium channel gene) for type 1 and 19p for type 2. Patients with EA-1 have brief episodes of ataxia with myokymia and nystagmus that last only minutes. Startle, sudden change in posture, and exercise can induce episodes. Acetazolamide or anticonvulsants may be therapeutic. Patients with EA-2 have episodes of ataxia with nystagmus that can last for hours or days. Stress, exercise, or excessive fatigue may be precipitants. Acetazolamide may be therapeutic and can reverse the relative intracellular alkalosis detected by magnetic resonance spectroscopy. Stop codon, nonsense mutations causing EA-2 have been found in the *CACNA1A* gene, encoding the α_{1A} voltage-dependent calcium channel subunit (see "SCA6," above).

Autosomal Recessive Ataxias

FRIEDREICH'S ATAXIA

This is the most common form of inherited ataxia, comprising one-half of all hereditary ataxias. It can occur in a classic form or in association with a genetically determined vitamin E deficiency syndrome; the two forms are clinically indistinguishable.

Symptoms and Signs

Friedreich's ataxia presents before 25 years of age with progressive staggering gait, frequent falling, and titubation. The lower extremities are more severely involved than the upper ones. Dysarthria

occasionally is the presenting symptom; rarely, progressive scoliosis, foot deformity, nystagmus, or cardiopathy is the initial sign.

The neurologic examination reveals nystagmus, loss of fast saccadic eye movements, truncal titubation, dysarthria, dysmetria, and ataxia of trunk and limb movements. Extensor plantar responses (with normal tone in trunk and extremities), absence of deep tendon reflexes, and weakness (greater distally than proximally) are usually found. Loss of vibratory and proprioceptive sensation occurs. The median age of death is 35 years. Women have a significantly better prognosis than men.

Cardiac involvement occurs in 90% of patients. Cardiomegaly, symmetric hypertrophy, murmurs, and conduction defects are reported. Moderate mental retardation or psychiatric syndromes are present in a small percentage of patients. A high incidence of diabetes mellitus (20%) is found and is associated with insulin resistance and pancreatic β-cell dysfunction. Musculoskeletal deformities are common and include pes cavus, pes equinovarus, and scoliosis. MRI of the spinal cord shows atrophy **(Fig. 373-2)**.

Figure 371-2

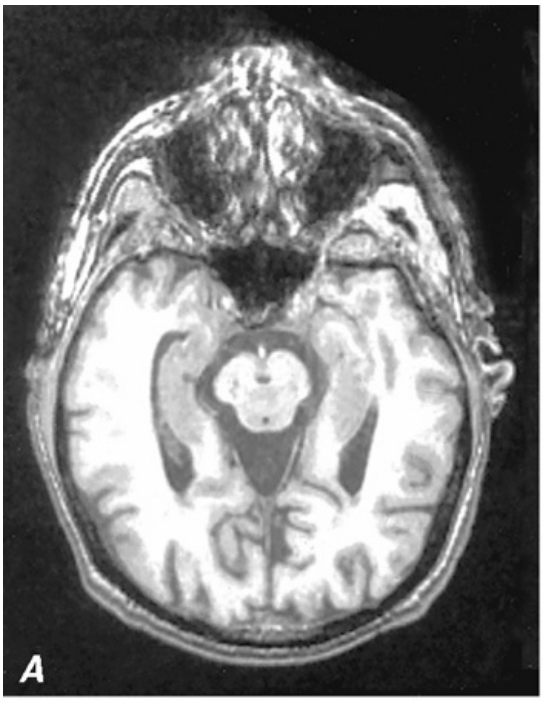

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

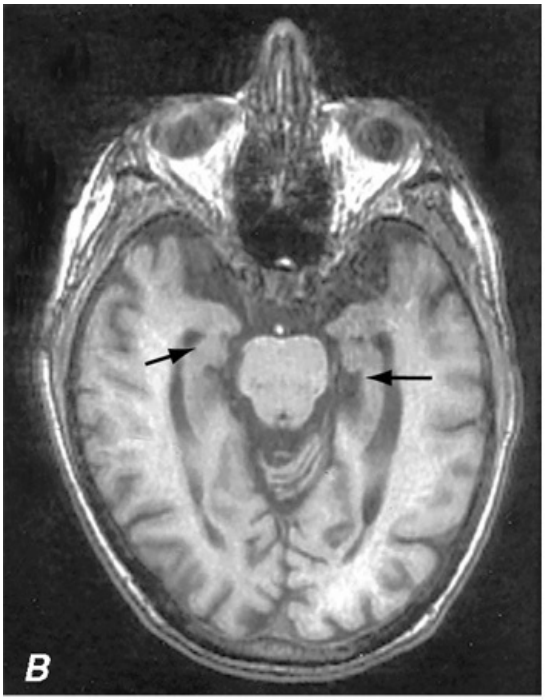

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

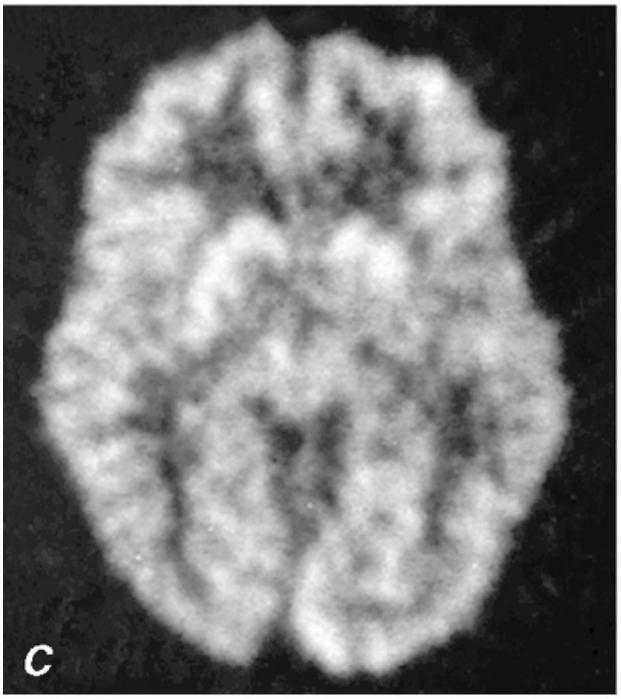

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

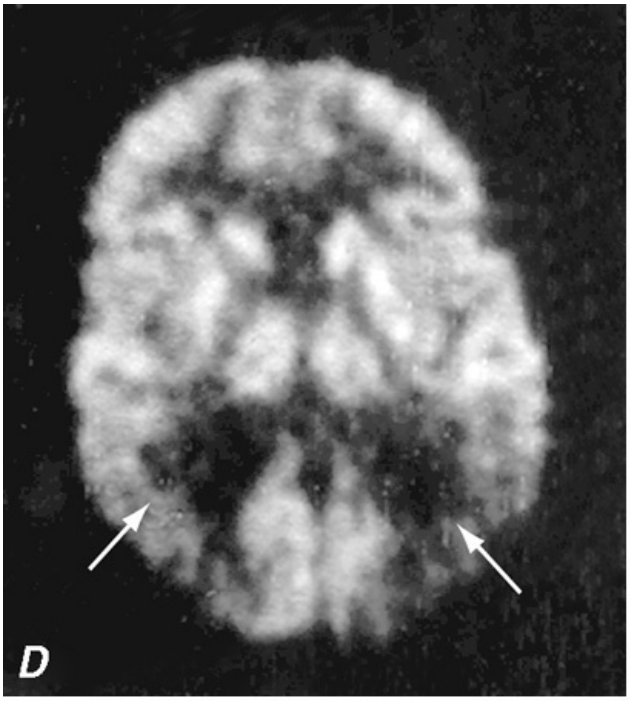

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sagittal MRI of the brain and spinal cord of a patient with Friedreich's ataxia, demonstrating spinal cord atrophy.

The primary sites of pathology are the spinal cord, dorsal root ganglion cells, and the peripheral nerves. Slight atrophy of the cerebellum and cerebral gyri may occur. Sclerosis and degeneration occur predominantly in the spinocerebellar tracts, lateral corticospinal tracts, and posterior columns. Degeneration of the glossopharyngeal, vagus, hypoglossal, and deep cerebellar nuclei is described. The cerebral cortex is histologically normal except for loss of Betz cells in the precentral gyri. The peripheral nerves are extensively involved, with a loss of large myelinated fibers. Cardiac pathology consists of myocytic hypertrophy and fibrosis, focal vascular fibromuscular dysplasia with subintimal or medial deposition of periodic acid–Schiff (PAS)-positive material, myocytopathy with unusual pleomorphic nuclei, and focal degeneration of nerves and cardiac ganglia.

Genetic Considerations

The classic form of Friedreich's ataxia has been mapped to 9q13-q21.1, and the mutant gene, *frataxin*, contains expanded GAA triplet repeats in the first intron. There is homozygosity for expanded GAA repeats in >95% of patients. Normal persons have 7–22 GAA repeats, and patients have 200–900 GAA repeats. A more varied clinical syndrome has been described in compound heterozygotes who have one copy of the GAA expansion and the other copy a point mutation in the *frataxin* gene. When the point mutation is located in the region of the gene that encodes the amino-terminal half of frataxin, the phenotype is milder, often consisting of a spastic gait, retained or exaggerated reflexes, no dysarthria, and mild or absent ataxia.

Patients with Friedreich's ataxia have undetectable or extremely low levels of *frataxin* mRNA, as compared with carriers and unrelated individuals; thus, disease appears to be caused by a loss of expression of the frataxin protein. Frataxin is a mitochondrial protein involved in iron homeostasis. Mitochondrial iron accumulation due to loss of the iron transporter coded by the mutant *frataxin* gene results in oxidized intramitochondrial iron. Excess oxidized iron results in turn in the oxidation of cellular components and irreversible cell injury.

Two forms of hereditary ataxia associated with abnormalities in the interactions of vitamin E (α-tocopherol) with very low density lipoprotein (VLDL) have been delineated. These are abetalipoproteinemia (Bassen-Kornzweig syndrome) and ataxia with vitamin E deficiency (AVED). Abetalipoproteinemia is caused by mutations in the gene coding for the larger subunit of the microsomal triglyceride transfer protein (MTP). Defects in MTP result in impairment of formation and secretion of VLDL in liver. This defect results in a deficiency of delivery of vitamin E to tissues, including the central and peripheral nervous system, as VLDL is the transport molecule for vitamin E and other fat-soluble substitutes. AVED is due to mutations in the gene for α-tocopherol transfer protein (α-TTP). These patients have an impaired ability to bind vitamin E into the VLDL produced and secreted by the liver, resulting in a deficiency of vitamin E in peripheral tissues. Hence, either absence of VLDL (abetalipoproteinemia) or impaired binding of vitamin E to VLDL (AVED) causes an ataxic syndrome. Once again, a genotype classification has proved to be essential in sorting out the various

forms of the Friedreich's disease syndrome, which may be clinically indistinguishable.

ATAXIA TELANGIECTASIA

Symptoms and Signs

Patients with ataxia telangiectasia (AT) present in the first decade of life with progressive telangiectatic lesions associated with deficits in cerebellar function and nystagmus. The neurologic manifestations correspond to those in Friedreich's disease, which should be included in the differential diagnosis. Truncal and limb ataxia, dysarthria, extensor plantar responses, myoclonic jerks, areflexia, and distal sensory deficits may develop. There is a high incidence of recurrent pulmonary infections and neoplasms of the lymphatic and reticuloendothelial system in patients with AT. Thymic hypoplasia with cellular and humoral (IgA and IgG2) immunodeficiencies, premature aging, and endocrine disorders such as Type 1 diabetes mellitus are described. There is an increased incidence of lymphomas, Hodgkin's disease, acute leukemias of the T cell type, and breast cancer.

The most striking neuropathologic changes include loss of Purkinje, granule, and basket cells in the cerebellar cortex as well as of neurons in the deep cerebellar nuclei. The inferior olives of the medulla may also have neuronal loss. There is a loss of anterior horn neurons in the spinal cord and of dorsal root ganglion cells associated with posterior column spinal cord demyelination. A poorly developed or absent thymus gland is the most consistent defect of the lymphoid system.

Genetic Considerations

The gene for AT (the *ATM* gene) encodes a protein that is similar to several yeast and mammalian phosphatidylinositol-3'-kinases involved in mitogenic signal transduction, meiotic recombination, and cell cycle control. Defective DNA repair in AT fibroblasts exposed to ultraviolet light has been demonstrated. The discovery of *ATM* will make possible the identification of heterozygotes who are at risk for cancer (e.g., breast cancer) and permit early diagnosis.

Mitochondrial Ataxias

Spinocerebellar syndromes have been identified with mutations in mitochondrial DNA (mtDNA). Thirty pathogenic mtDNA point mutations and 60 different types of mtDNA deletions are known, several of which cause or are associated with ataxia (Chap. 387).

Treatment: Ataxic Disorders

The most important goal in management of patients with ataxia is to identify treatable disease entities. Mass lesions must be recognized promptly and treated appropriately. Paraneoplastic disorders can often be identified by the clinical patterns of disease that they produce, measurement of specific autoantibodies, and uncovering the primary cancer; these disorders are often refractory to therapy, but some patients improve following removal of the tumor or immunotherapy (Chap. 101). Ataxia with anti-gliadin antibodies and gluten-sensitive enteropathy may improve with a gluten-free diet. Malabsorption syndromes leading to vitamin E deficiency may lead to ataxia. The vitamin E deficiency form of Friedreich's ataxia must be considered, and serum vitamin E levels measured. Vitamin E therapy is indicated for these rare patients. Vitamin B1 and B_{12} levels in serum should be measured, and the vitamins administered to patients having deficient levels. Hypothyroidism is easily treated. The cerebrospinal fluid should be tested for a syphilitic infection in patients with progressive ataxia and other features of tabes dorsalis. Similarly, antibody titers for Lyme disease and *Legionella* should be measured and appropriate antibiotic therapy should be instituted in antibody-positive patients. Aminoacidopathies, leukodystrophies, urea-cycle abnormalities, and mitochondrial encephalomyopathies may produce ataxia, and some dietary or metabolic therapies are available for these disorders. The deleterious effects of phenytoin and alcohol on the cerebellum are well known, and these exposures should be avoided in patients with ataxia of any cause.

There is no proven therapy for any of the autosomal dominant ataxias (SCA1 to -28). There is preliminary evidence that idebenone, a free-radical scavenger, can improve myocardial hypertrophy in patients with classic Friedreich ataxia; there is no current evidence, however, that it improves neurologic function. A small preliminary study in a mixed population of patients with different inherited ataxias raised the possibility that the glutamate antagonist riluzole may offer modest benefit. Iron chelators and antioxidant drugs are potentially harmful in Friedreich's patients as they may increase heart muscle injury. Acetazolamide can reduce the duration of symptoms of episodic ataxia. At present, identification of an at-risk person's genotype, together with appropriate family and genetic counseling, can reduce the incidence of these cerebellar syndromes in future generations (Chap. 63).

FURTHER READINGS

Anheim M et al: Clinical spectrum of ataxia-telangiectasia in adulthood. Neurology 73:430, 2009

Brusse E et al Diagnosis and management of early- and late-onset cerebellar ataxia. Clin Genet 71:12, 2007[PMID: 17204042] [Full Text]

Durr A: Autosomal dominant cerebellar ataxias: Polyglutamine expansions and beyond. Lancet Neurol

9:885, 2010[PMID: 20723845] [Full Text]

Hadjivassiliou M et al: Gluten ataxia in perspective: Epidemiology, genetic susceptibility and clinical characteristics. Brain 136:685, 2003

Ristori G et al: Riluzole in cerebellar ataxia: A randomized, double-blind, placebo-controlled pilot trial. Neurology 74:839, 2010[PMID: 20211908] [Full Text]

Rosenberg RN, Paulson HL: The inherited ataxias, in *The Molecular and Genetic Basis of Neurologic and Psychiatric Disease,* 4th ed, RN Rosenberg et al (eds). Philadelphia, Lippincott Williams & Wilkins, 2008.

Todi SV et al: Activity and cellular functions of the deubiquitinating enzyme and polyglutamine disease protein ataxin-3 are regulated by ubiquitination at lysine 117. J Biol Chem 285:39303, 2010[PMID: 20943656] [Full Text]

Cerebellar Degeneration Information Page. National Institute of Neurological Disorder and Stroke. http://www.NINDS.NIH.gov/disorders/ataxia/ataxia.htm

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

AMYOTROPHIC LATERAL SCLEROSIS

Amyotrophic lateral sclerosis (ALS) is the most common form of progressive motor neuron disease. It is a prime example of a neurodegenerative disease and is arguably the most devastating of the neurodegenerative disorders.

Pathology

The pathologic hallmark of motor neuron degenerative disorders is death of lower motor neurons (consisting of anterior horn cells in the spinal cord and their brainstem homologues innervating bulbar muscles) and upper, or corticospinal, motor neurons (originating in layer five of the motor cortex and descending via the pyramidal tract to synapse with lower motor neurons, either directly or indirectly via interneurons) (Chap. 22). Although at its onset ALS may involve selective loss of function of only upper or lower motor neurons, it ultimately causes progressive loss of both categories of motor neurons. Indeed, in the absence of clear involvement of both motor neuron types, the diagnosis of ALS is questionable.

Other motor neuron diseases involve only particular subsets of motor neurons **(Tables 374-1***and***374-2)**. Thus, in bulbar palsy and spinal muscular atrophy (SMA; also called *progressive muscular atrophy*), the lower motor neurons of brainstem and spinal cord, respectively, are most severely involved. By contrast, pseudobulbar palsy, primary lateral sclerosis (PLS), and familial spastic paraplegia (FSP) affect only upper motor neurons innervating the brainstem and spinal cord.

Table 374-1 Etiology of Motor Neuron Disorders

Diagnostic Category	Investigation
Structural lesions Parasagittal or foramen magnum tumors Cervical spondylosis Chiari malformation of syrinx Spinal cord arteriovenous malformation	MRI scan of head (including foramen magnum and cervical spine)
Infections Bacterial—tetanus, Lyme Viral—poliomyelitis, herpes zoster Retroviral—myelopathy	CSF exam, culture Lyme titer Anti-viral antibody HTLV-1 titers
Intoxications, physical agents Toxins—lead, aluminum, others Drugs—strychnine, phenytoin Electric shock, x-irradiation	24-hour urine for heavy metals Serum lead level
Immunologic mechanisms Plasma cell dyscrasias Autoimmune polyradiculopathy Motor neuropathy with conduction block Paraneoplastic Paracarcinomatous	Complete blood counta Sedimentation ratea Total proteina Anti-GM1 antibodiesa Anti-Hu antibody MRI scan, bone marrow biopsy
Metabolic Hypoglycemia Hyperparathyroidism Hyperthyroidism	Fasting blood sugara Routine chemistries including calciuma PTH Thyroid functiona

Deficiency of folate, vitamin B12, vitamin E Deficiency of copper, zinc Malabsorption Mitochondrial dysfunction	Vitamin B12, vitamin E, folatea Serum zinc, coppera 24-h stool fat, carotene, prothrombin time Fasting lactate, pyruvate, ammonia Consider mtDNA
Hyperlipidemia	Lipid electrophoresis
Hyperglycinuria	Urine and serum amino acids CSF amino acids
Hereditary disorders Superoxide dismutase TDP43 FUS/TLS Androgen receptor defect (Kennedy's disease) Hexosaminidase deficiency Infantile α-glucosidase deficiency (Pompe's)	WBC DNA for mutational analysis

[a]Denotes studies that should be obtained in all cases. ***Abbreviations:*** CSF, cerebrospinal fluid; FUS/TLS, fused in sarcoma/translocated in liposarcoma; HTLV-1, human T cell lymphotropic virus; PTH, parathyroid; WBC, white blood cell.

Table 374-2 Sporadic Motor Neuron Diseases

Chronic	**Entity**
Upper and lower motor neuron	Amyotrophic lateral sclerosis
Predominantly upper motor neuron	Primary lateral sclerosis
Predominantly lower motor neuron	Multi focal motor neuropathy with conduction block
	Motor neuropathy with paraproteinemia or cancer
	Motor predominant peripheral neuropathies
Other	
Associated with other neurodegenerative disorders	
Secondary motor neuron disorders (see Table 374-1)	
Acute	
Poliomyelitis	
Herpes zoster	
Coxsackie virus	

In each of these diseases, the affected motor neurons undergo shrinkage, often with accumulation of the pigmented lipid (lipofuscin) that normally develops in these cells with advancing age. In ALS, the motor neuron cytoskeleton is typically affected early in the illness. Focal enlargements are frequent in proximal motor axons; ultrastructurally, these "spheroids" are composed of accumulations of neurofilaments and other proteins. Also seen is proliferation of astroglia and microglia, the inevitable accompaniment of all degenerative processes in the central nervous system (CNS).

The death of the peripheral motor neurons in the brainstem and spinal cord leads to denervation and consequent atrophy of the corresponding muscle fibers. Histochemical and electrophysiologic evidence indicates that in the early phases of the illness denervated muscle can be reinnervated by sprouting of nearby distal motor nerve terminals, although reinnervation in this disease is considerably less extensive than in most other disorders affecting motor neurons (e.g., poliomyelitis, peripheral neuropathy). As denervation progresses, muscle atrophy is readily recognized in muscle biopsies and on clinical examination. This is the basis for the term *amyotrophy*. The loss of cortical motor neurons results in thinning of the corticospinal tracts that travel via the internal capsule **(Fig. 374-1)** and

brainstem to the lateral and anterior white matter columns of the spinal cord. The loss of fibers in the lateral columns and resulting fibrillary gliosis impart a particular firmness (*lateral sclerosis*). A remarkable feature of the disease is the selectivity of neuronal cell death. By light microscopy, the entire sensory apparatus, the regulatory mechanisms for the control and coordination of movement, and the components of the brain that are needed for cognitive processes, remain intact. However, immunostaining indicates that neurons bearing ubiquitin, a marker for degeneration, are also detected in nonmotor systems. Moreover, studies of glucose metabolism in the illness also indicate that there is neuronal dysfunction outside of the motor system. Within the motor system, there is some selectivity of involvement. Thus, motor neurons required for ocular motility remain unaffected, as do the parasympathetic neurons in the sacral spinal cord (the nucleus of Onufrowicz, or Onuf) that innervate the sphincters of the bowel and bladder.

Figure 374-1

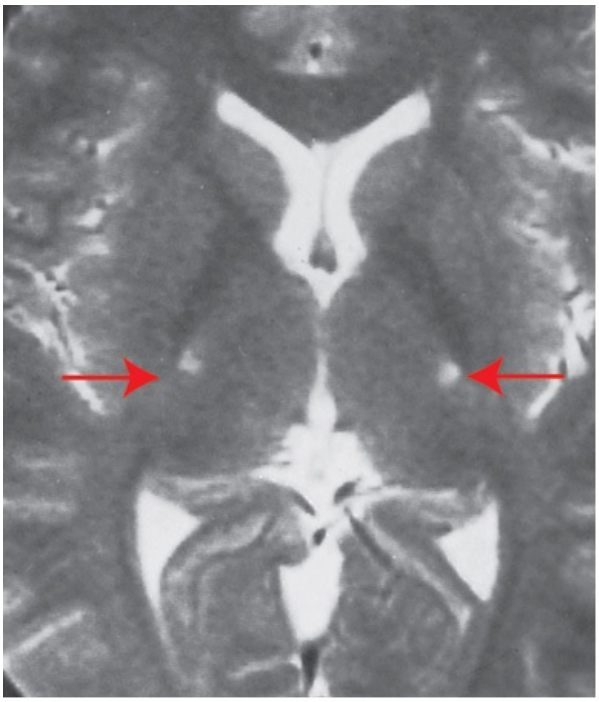

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Amyotrophic lateral sclerosis. Axial T2-weighted MRI scan through the lateral ventricles of the brain reveals abnormal high signal intensity within the corticospinal tracts (*arrows*). This MRI feature represents an increase in water content in myelin tracts undergoing Wallerian degeneration secondary to cortical motor neuronal loss. This finding is commonly present in ALS, but can also be seen in AIDS-related encephalopathy, infarction, or other disease processes that produce corticospinal neuronal loss in a symmetric fashion.

Clinical Manifestations

The manifestations of ALS are somewhat variable depending on whether corticospinal neurons or lower motor neurons in the brainstem and spinal cord are more prominently involved. With lower motor neuron dysfunction and early denervation, typically the first evidence of the disease is insidiously developing asymmetric weakness, usually first evident distally in one of the limbs. A detailed history often discloses recent development of cramping with volitional movements, typically in the early hours of the morning (e.g., while stretching in bed). Weakness caused by denervation is associated with progressive wasting and atrophy of muscles and, particularly early in the illness, spontaneous twitching of motor units, or fasciculations. In the hands, a preponderance of extensor over flexor weakness is common. When the initial denervation involves bulbar rather than limb muscles, the problem at onset is difficulty with chewing, swallowing, and movements of the face and tongue. Early involvement of the muscles of respiration may lead to death before the disease is far advanced elsewhere. With prominent corticospinal involvement, there is hyperactivity of the muscle-stretch reflexes (tendon jerks) and, often, spastic resistance to passive movements of the affected limbs. Patients with significant reflex hyperactivity complain of muscle stiffness often out of proportion to weakness. Degeneration of the corticobulbar projections innervating the brainstem results in dysarthria and exaggeration of the motor expressions of emotion. The latter leads to involuntary excess in weeping or laughing (pseudobulbar affect).

Virtually any muscle group may be the first to show signs of disease, but, as time passes, more and

more muscles become involved until ultimately the disorder takes on a symmetric distribution in all regions. It is characteristic of ALS that, regardless of whether the initial disease involves upper or lower motor neurons, both will eventually be implicated. Even in the late stages of the illness, sensory, bowel and bladder, and cognitive functions are preserved. Even when there is severe brainstem disease, ocular motility is spared until the very late stages of the illness. Dementia is not a component of sporadic ALS. In some families, ALS is co-inherited with frontotemporal dementia, characterized by early behavioral abnormalities with prominent behavioral features indicative of frontal lobe dysfunction.

A committee of the World Federation of Neurology has established diagnostic guidelines for ALS. Essential for the diagnosis is simultaneous upper and lower motor neuron involvement with progressive weakness, and the exclusion of all alternative diagnoses. The disorder is ranked as "definite" ALS when three or four of the following are involved: bulbar, cervical, thoracic, and lumbosacral motor neurons. When two sites are involved, the diagnosis is "probable," and when only one site is implicated, the diagnosis is "possible." An exception is made for those who have progressive upper and lower motor neuron signs at only one site and a mutation in the gene encoding superoxide dismutase (SOD1; below).

Epidemiology

The illness is relentlessly progressive, leading to death from respiratory paralysis; the median survival is from 3–5 years. There are very rare reports of stabilization or even regression of ALS. In most societies there is an incidence of 1–3 per 100,000 and a prevalence of 3–5 per 100,000. Several endemic foci of higher prevalence exist in the western Pacific (e.g., in specific regions of Guam or Papua New Guinea). In the United States and Europe, males are somewhat more frequently affected than females. Epidemiologic studies have incriminated risk factors for this disease including exposure to pesticides and insecticides, smoking and, in one report, service in the military. While ALS is overwhelmingly a sporadic disorder, some 5–10% of cases are inherited as an autosomal dominant trait.

Familial ALS

Several forms of selective motor neuron disease are inheritable **(Table 374-3)**. Familial ALS (FALS) involves both corticospinal and lower motor neurons. Apart from its inheritance as an autosomal dominant trait, it is clinically indistinguishable from sporadic ALS. Genetic studies have identified mutations in the genes encoding the cytosolic enzyme SOD1 (superoxide dismutase), and the RNA binding proteins TDP43 (encoded by the TAR DNA binding protein gene), and FUS/TLS (fused in sarcoma/translocated in liposarcoma), as the most common causes of FALS. Mutations in SOD1 account for about 20% of cases of FALS, while TDP43 and FUS/TLS each represent about 5% of familial cases.

Table 374-3 Genetic Motor Neuron Diseases

Disease	Locus	Gene	Inheritance	Onset	Gene Function	Unusual Features
I. Upper and Lower Motor Neurons (familial ALS)						
ALS1	21q	Superoxide dismutase	AD	Adult	Protein anti oxidant	
ALS2	2q	Alsin	AR	Juvenile	GEF signalling	Severe corticobulbar, corticospinal features
ALS4	9q	Senataxin	AD	Late juvenile	DNA helicase	Late childhood onset
ALS6	16p	FUS/TLS	AD	Adult	DNA, RNA binding	
ALS8	20q	Vesicle-associated protein B	AD	Adult	Vesicular trafficking	
ALS9	14q	Angiogenin	AD	Adult	RNAse, angiogenesis	
ALS10	1q	TARDBP	AD	Adult	DNA, RNA binding	
ALS	2p	Dynactin	AD	Adult	Axonal transport	Vocal cord stridor in some families
ALS	17q	Paraoxonases 1-3	AD	Adult	Detoxify intoxicants	
ALS	mtDNA	Cytochrome c oxidase		Adult	ATP	

					generation	
ALS	mtDNA	tRNA-isoleucine		Adult	ATP generation	
II. Lower Motor Neurons						
Spinal muscular atrophies	5q	Survival motor neuron	AR	Infancy	RNA metabolism	
GM2-gangliosidosis						
1. Sandhoff disease	5q	Hexosaminidase B	AR	Childhood	Ganglioside recycling	
2. AB variant	5q	GM2-activator protein	AR	Childhood	Ganglioside recycling	
3. Adult Tay-Sachs disease	15q	Hexosaminidase A	AR	Childhood	Ganglioside recycling	
X-linked spinobulbar muscular atrophy	Xq	Androgen receptor	XR	Adult	Nuclear signalling	
III. Upper Motor Neuron (Selected FSPs)						
SPG3A	14q	Atlastin	AD	Childhood	GTPase —vesicle recycling	Some peripheral neuropathy
SPG4	2p	Spastin	AD	Early adulthood	ATPase family —microtubule associate	± mental retardation, motor neuropathy
SPG6	15q	NIPA1	AD	Early adulthood	Membrane transporter or receptor	Deleted in Prader-Willi, Angelman
SPG8	8q	Strumpellin	AD	Early adulthood	Ubiquitous, spectrin-like	
SPG10	12q	Kinesin heavy chain KIF5A	AD	2nd–3rd decade	Motor-associated protein	± peripheral neuropathy, retardation
SPG13	2q	Heat shock protein 60	AD	Early adulthood	Chaperone protein	
SPG17	11q	Silver (BSCL2)	AD	Variable	Membrane protein in ER	Amyotrophy hands, feet
SPG31	2p	REEP1	AD	Early	Mitochondrial protein	Rarely, amyotrophy
SPG33	10q	ZFYVE27	AD	Adult	Interacts with spastin	Pes equinus
SPG42	3q	Acetyl-CoA-transporter	AD	Variable	Solute carrier	
SPG5	8q	Cytochrome P450	AR	Variable	Degrades endogenous substances	Sensory loss
SPG7	16q	Paraplegin	AR	Variable	Mitochondrial protein	Rarely, optic atrophy, ataxia, neuropathy
SPG11	15q	Spatacsin	AR	Childhood	Cytosolic, ? Membrane-associated	Some sensory loss, thin corpus callosum
SPG15	14q	Spastizin	AR	Childhood	Zinc finger protein	Some amyotrophy, some CNS features incl thin corpus callosum
SPG20	13q	Spartin	AR	Childhood	Endosomal-trafficking protein	Cerebellar, extrapyramidal signs, short stature, MR

SPG21	15q	Maspardin	AR	Childhood	Endosomal-trafficking protein	Cerebellar, extrapyramidal signs, short stature, MR
SPG35	16q	Fatty acid 2 hydrolase	AR	Childhood	Membrane protein	Multiple CNS features
SPG39	19p	Neuropathy target esterase	AR	Early childhood	Esterase	
SPG44	1q	Connexin 47	AR	Childhood	Gap junction protein	Possible mild CNS features
SPG2	Xq	Proteolipid protein	XR	Early childhood	Myelin protein	Sometimes multiple CNS features
SPG1	Xq	L1-CAM	XR	Infancy	Cell-adhesion molecule	
	Xq	Adrenoleukodystrophy	XR	Early adulthood	ATP-binding transporter protein	Possible adrenal insufficiency, CNS inflamation
IV. ALS-Plus Syndromes						
Amyotrophy with behavioral disorders	17q	Tau protein				
Parkinsonism						

Abbreviations: ALS, amyotrophic lateral sclerosis; BSCL2, Bernadelli-Seip congenital lipodystrophy 2B; FSP, familial spastic paraplegia; FUS/TLS, fused in sarcoma/translocated in liposarcoma; TDP43, Tar DNA binding protein 43 kd.

Rare mutations in other genes are also clearly implicated in ALS-like diseases. Thus, a familial, dominantly inherited motor disorder that in some individuals closely mimics the ALS phenotype arises from mutations in a gene that encodes a vesicle-binding protein. A predominantly lower motor neuron disease with early hoarseness due to laryngeal dysfunction has been ascribed to mutations in the gene encoding the cellular accessory motor protein dynactin. Mutations in senataxin, a helicase, cause an early adult-onset, slowly evolving ALS variant. Kennedy's syndrome is an X-linked, adult-onset disorder that may mimic ALS, as described below.

Genetic analyses are also beginning to illuminate the pathogenesis of some childhood-onset motor neuron diseases. For example, a slowly disabling degenerative, predominantly upper motor neuron disease that starts in the first decade is caused by mutations in a gene that expresses a novel signaling molecule with properties of a guanine-exchange factor, termed *alsin*.

Differential Diagnosis

Because ALS is currently untreatable, it is imperative that potentially remediable causes of motor neuron dysfunction be excluded (Table 374-1). This is particularly true in cases that are atypical by virtue of (1) restriction to either upper or lower motor neurons, (2) involvement of neurons other than motor neurons, and (3) evidence of motor neuronal conduction block on electrophysiologic testing. Compression of the cervical spinal cord or cervicomedullary junction from tumors in the cervical regions or at the foramen magnum or from cervical spondylosis with osteophytes projecting into the vertebral canal can produce weakness, wasting, and fasciculations in the upper limbs and spasticity in the legs, closely resembling ALS. The absence of cranial nerve involvement may be helpful in differentiation, although some foramen magnum lesions may compress the twelfth cranial (hypoglossal) nerve, with resulting paralysis of the tongue. Absence of pain or of sensory changes, normal bowel and bladder function, normal roentgenographic studies of the spine, and normal cerebrospinal fluid (CSF) all favor ALS. Where doubt exists, MRI scans and contrast myelography should be performed to visualize the cervical spinal cord.

Another important entity in the differential diagnosis of ALS is *multifocal motor neuropathy with conduction block* (MMCB), discussed below. A diffuse, lower motor axonal neuropathy mimicking ALS sometimes evolves in association with hematopoietic disorders such as lymphoma or multiple myeloma. In this clinical setting, the presence of an M-component in serum should prompt consideration of a bone marrow biopsy. Lyme disease (Chap. 173) may also cause an axonal, lower motor neuropathy, although typically with intense proximal limb pain and a CSF pleocytosis.

Other treatable disorders that occasionally mimic ALS are chronic lead poisoning and thyrotoxicosis. These disorders may be suggested by the patient's social or occupational history or by unusual clinical features. When the family history is positive, disorders involving the genes encoding cytosolic SOD1,

TDP43, FUS/TLS, as well as adult hexosaminidase A or α-glucosidase deficiency must be excluded (Chap. 361). These are readily identified by appropriate laboratory tests. Benign fasciculations are occasionally a source of concern because on inspection they resemble the fascicular twitchings that accompany motor neuron degeneration. The absence of weakness, atrophy, or denervation phenomena on electrophysiologic examination usually excludes ALS or other serious neurologic disease. Patients who have recovered from poliomyelitis may experience a delayed deterioration of motor neurons that presents clinically with progressive weakness, atrophy, and fasciculations. Its cause is unknown, but it is thought to reflect sublethal prior injury to motor neurons by poliovirus (Chap. 191).

Rarely, ALS develops concurrently with features indicative of more widespread neurodegeneration. Thus, one infrequently encounters otherwise typical ALS patients with a parkinsonian movement disorder or dementia. It remains unclear whether this reflects the unlikely simultaneous occurrence of two disorders or a primary defect triggering two forms of neurodegeneration. The latter is suggested by the observation that multisystem neurodegenerative diseases may be inherited. For example, prominent amyotrophy has been described as a dominantly inherited disorder in individuals with bizarre behavior and a movement disorder suggestive of parkinsonism; many such cases have now been ascribed to mutations that alter the expression of tau protein in brain (Chap. 371). In other cases, ALS develops simultaneously with a striking frontotemporal dementia. These disorders may be dominantly co-inherited; in some families, this trait is linked to a locus on chromosome 9p, although the underlying genetic defect is not established. An ALS-like disorder has also been described in some individuals with chronic traumatic encephalopathy, associated with deposition of TDP43 and neurofibrillary tangles in motor neurons.

Pathogenesis

The cause of sporadic ALS is not well defined. Several mechanisms that impair motor neuron viability have been elucidated in mice and rats induced to develop motor neuron disease by SOD1 transgenes with ALS-associated mutations. It is evident that excitotoxic neurotransmitters such as glutamate participate in the death of motor neurons in ALS. This may be a consequence of diminished uptake of synaptic glutamate by an astroglial glutamate transporter, EAAT2. It is striking that one cellular defense against such excitotoxicity is the enzyme SOD1, which detoxifies the free radical superoxide anion (Chap. 365). Precisely why SOD1 mutations are toxic to motor nerves is not established, although it is clear the effect is not simply loss of normal scavenging of the superoxide anion. The mutant protein is conformationally unstable and prone to aberrant catalytic reactions. In turn, these features lead to aggregation of SOD1 protein, impairment of axonal transport, reduced production of ATP and other perturbations of mitochondrial function, activation of neuroinflammatory cascades within the ALS spinal cord, and ultimately induction of cell death via pathways that are at least partially dependent on caspases.

It has recently been observed that mutations in the TDP43 and FUS/TLS genes also cause ALS. These multifunctional proteins bind RNA and DNA and shuttle between the nucleus and the cytoplasm, playing multiple roles in the control of cell proliferation, DNA repair and transcription, and gene translation, both in the cytoplasm and locally in dendritic spines in response to electrical activity. How mutations in FUS/TLS provoke motor neuron cell death is not clear, although this may represent loss of function of FUS/TLS in the nucleus or an acquired, toxic function of the mutant proteins in the cytosol.

Multiple recent studies have convincingly demonstrated that non-neuronal cells importantly influence the disease course, at least in ALS transgenic mice. A striking additional finding in neurodegenerative disorders is that miscreant proteins arising from gene defects in familial forms of these diseases are often implicated in sporadic forms of the same disorder. For example, germline mutations in the genes encoding beta-amyloid and alpha-synuclein cause familial forms of Alzheimer's and Parkinson's diseases (AD and PD), and posttranslational, noninherited abnormalities in these proteins are also central to sporadic AD and PD. Analogously, recent reports propose that nonheritable, posttranslational modifications in SOD1 are pathogenic in sporadic ALS.

Treatment: Amyotrophic Lateral Sclerosis

No treatment arrests the underlying pathologic process in ALS. The drug riluzole (100 mg/d) was approved for ALS because it produces a modest lengthening of survival. In one trial, the survival rate at 18 months with riluzole was similar to placebo at 15 months. The mechanism of this effect is not known with certainty; riluzole may reduce excitotoxicity by diminishing glutamate release. Riluzole is generally well tolerated; nausea, dizziness, weight loss, and elevated liver enzymes occur occasionally. Pathophysiologic studies of mutant SOD1–related ALS in mice have disclosed diverse targets for therapy; consequently, multiple therapies are presently in clinical trials for ALS. These include studies of ceftriaxone, which may augment astroglial glutamate transport and thereby be anti-excitotoxic, and pramipexole and tamoxifen, which are neuroprotective. Interventions such as antisense oligonucleotides (ASO) that diminish expression of mutant SOD1 protein prolong survival in transgenic ALS mice and rats and are also now in trial for SOD1-mediated ALS.

In the absence of a primary therapy for ALS, a variety of rehabilitative aids may substantially assist

ALS patients. Foot-drop splints facilitate ambulation by obviating the need for excessive hip flexion and by preventing tripping on a floppy foot. Finger extension splints can potentiate grip. Respiratory support may be life-sustaining. For patients electing against long-term ventilation by tracheostomy, positive-pressure ventilation by mouth or nose provides transient (several weeks) relief from hypercarbia and hypoxia. Also extremely beneficial for some patients is a respiratory device (Cough Assist Device) that produces an artificial cough. This is highly effective in clearing airways and preventing aspiration pneumonia. When bulbar disease prevents normal chewing and swallowing, gastrostomy is uniformly helpful, restoring normal nutrition and hydration. Fortunately, an increasing variety of speech synthesizers are now available to augment speech when there is advanced bulbar palsy. These facilitate oral communication and may be effective for telephone use.

In contrast to ALS, several of the disorders (Tables 374-1 and 374-3) that bear some clinical resemblance to ALS are treatable. For this reason, a careful search for causes of secondary motor neuron disease is warranted.

OTHER MOTOR NEURON DISEASES

Selected Lower Motor Neuron Disorders

In these motor neuron diseases, the peripheral motor neurons are affected without evidence of involvement of the corticospinal motor system (Tables 374-1 to 374-3).

X-LINKED SPINOBULBAR MUSCULAR ATROPHY (KENNEDY'S DISEASE)

This is an X-linked lower motor neuron disorder in which progressive weakness and wasting of limb and bulbar muscles begins in males in mid-adult life and is conjoined with androgen insensitivity manifested by gynecomastia and reduced fertility (Chap. 346). In addition to gynecomastia, which may be subtle, two findings distinguishing this disorder from ALS are the absence of signs of pyramidal tract disease (spasticity) and the presence of a subtle sensory neuropathy in some patients. The underlying molecular defect is an expanded trinucleotide repeat (-CAG-) in the first exon of the androgen receptor gene on the X chromosome. DNA testing is available. An inverse correlation appears to exist between the number of -CAG- repeats and the age of onset of the disease.

ADULT TAY-SACH'S DISEASE

Several reports have described adult-onset, predominantly lower motor neuropathies arising from deficiency of the enzyme β-hexosaminidase (hex A). These tend to be distinguishable from ALS because they are very slowly progressive; dysarthria and radiographically evident cerebellar atrophy may be prominent. In rare cases, spasticity may also be present, although it is generally absent (Chap. 361).

SPINAL MUSCULAR ATROPHY

The SMAs are a family of selective lower motor neuron diseases of early onset. Despite some phenotypic variability (largely in age of onset), the defect in the majority of families with SMA maps to a locus on chromosome 5 encoding a putative motor neuron survival protein (SMN, for survival motor neuron) that is important in the formation and trafficking of RNA complexes across the nuclear membrane. Neuropathologically these disorders are characterized by extensive loss of large motor neurons; muscle biopsy reveals evidence of denervation atrophy. Several clinical forms exist.

Infantile SMA (SMA I, Werdnig-Hoffmann disease) has the earliest onset and most rapidly fatal course. In some instances it is apparent even before birth, as indicated by decreased fetal movements late in the third trimester. Though alert, afflicted infants are weak and floppy (hypotonic) and lack muscle stretch reflexes. Death generally ensues within the first year of life. *Chronic childhood SMA* (SMA II) begins later in childhood and evolves with a more slowly progressive course. *Juvenile SMA* (SMA III, Kugelberg-Welander disease) manifests during late childhood and runs a slow, indolent course. Unlike most denervating diseases, in this chronic disorder weakness is greatest in the proximal muscles; indeed, the pattern of clinical weakness can suggest a primary myopathy such as limb-girdle dystrophy. Electrophysiologic and muscle biopsy evidence of denervation distinguish SMA III from the myopathic syndromes. There is no primary therapy for SMA, although remarkable recent experimental data indicate that it may be possible to deliver the missing SMN gene to motor neurons using intravenously delivered adeno-associated viruses (e.g., AAV9) immediately after birth.

MULTIFOCAL MOTOR NEUROPATHY WITH CONDUCTION BLOCK

In this disorder lower motor neuron function is regionally and chronically disrupted by remarkably focal blocks in conduction. Many cases have elevated serum titers of mono- and polyclonal antibodies to ganglioside GM1; it is hypothesized that the antibodies produce selective, focal, paranodal demyelination of motor neurons. MMCB is not typically associated with corticospinal signs. In contrast with ALS, MMCB may respond dramatically to therapy such as IV immunoglobulin or chemotherapy; it is thus imperative that MMCB be excluded when considering a diagnosis of ALS.

OTHER FORMS OF LOWER MOTOR NEURON DISEASE

In individual families, other syndromes characterized by selective lower motor neuron dysfunction in an

SMA-like pattern have been described. There are rare X-linked and autosomal dominant forms of apparent SMA. There is an ALS variant of juvenile onset, the Fazio-Londe syndrome, that involves mainly the musculature innervated by the brainstem. A component of lower motor neuron dysfunction is also found in degenerative disorders such as Machado-Joseph disease and the related olivopontocerebellar degenerations (Chap. 373).

Selected Disorders of the Upper Motor Neuron

PRIMARY LATERAL SCLEROSIS

This exceedingly rare disorder arises sporadically in adults in mid- to late life. Clinically PLS is characterized by progressive spastic weakness of the limbs, preceded or followed by spastic dysarthria and dysphagia, indicating combined involvement of the corticospinal and corticobulbar tracts. Fasciculations, amyotrophy, and sensory changes are absent; neither electromyography nor muscle biopsy shows denervation. On neuropathologic examination there is selective loss of the large pyramidal cells in the precentral gyrus and degeneration of the corticospinal and corticobulbar projections. The peripheral motor neurons and other neuronal systems are spared. The course of PLS is variable; while long-term survival is documented, the course may be as aggressive as in ALS, with ~ 3-year survival from onset to death. Early in its course, PLS raises the question of multiple sclerosis or other demyelinating diseases such as adrenoleukodystrophy as diagnostic considerations (Chap. 380). A myelopathy suggestive of PLS is infrequently seen with infection with the retrovirus human T cell lymphotropic virus (HTLV-I) (Chap. 377). The clinical course and laboratory testing will distinguish these possibilities.

FAMILIAL SPASTIC PARAPLEGIA

In its pure form, FSP is usually transmitted as an autosomal trait; most adult-onset cases are dominantly inherited. Symptoms usually begin in the third or fourth decade, presenting as progressive spastic weakness beginning in the distal lower extremities; however, there are variants with onset so early that the differential diagnosis includes cerebral palsy. FSP typically has a long survival, presumably because respiratory function is spared. Late in the illness there may be urinary urgency and incontinence and sometimes fecal incontinence; sexual function tends to be preserved.

In pure forms of FSP, the spastic leg weakness is often accompanied by posterior column (vibration and position) abnormalities and disturbance of bowel and bladder function. Some family members may have spasticity without clinical symptoms.

By contrast, particularly when recessively inherited, FSP may have complex or complicated forms in which altered corticospinal and dorsal column function is accompanied by significant involvement of other regions of the nervous system, including amyotrophy, mental retardation, optic atrophy, and sensory neuropathy.

Neuropathologically, in FSP there is degeneration of the corticospinal tracts, which appear nearly normal in the brainstem but show increasing atrophy at more caudal levels in the spinal cord; in effect, the pathologic picture is of a dying-back or distal axonopathy of long neuronal fibers within the CNS.

Defects at numerous loci underlie both dominantly and recessively inherited forms of FSP (Table 374-3). More than 20 FSP genes have now been identified. The gene most commonly implicated in dominantly inherited FSP is *spastin,* which encodes a microtubule interacting protein. The most common childhood-onset dominant form arises from mutations in the *atlastin* gene. A kinesin heavy-chain protein implicated in microtubule motor function was found to be defective in a family with dominantly inherited FSP of variable onset age.

An infantile-onset form of X-linked, recessive FSP arises from mutations in the gene for myelin proteolipid protein (Chap. 365). This is an example of rather striking allelic variation, as most other mutations in the same gene cause not FSP but Pelizaeus-Merzbacher disease, a widespread disorder of CNS myelin. Another recessive variant is caused by defects in the *paraplegin* gene. Paraplegin has homology to metalloproteases that are important in mitochondrial function in yeast.

Web Sites

Several websites provide valuable information on ALS including those offered by the Muscular Dystrophy Association (*www.mdausa.org*), the Amyotrophic Lateral Sclerosis Association (*www.alsa.org*), and the World Federation of Neurology and the Neuromuscular Unit at Washington University in St. Louis (*www.neuro.wustl.edu/neuromuscular*).

FURTHER READINGS

Bosco D et al: Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. Nat Neurosci, 11:1396, 2010

Bruijn LI, Cudkowicz M: Therapeutic targets for amyotrophic lateral sclerosis: Current treatments and prospects for more effective therapies. Expert Rev Neurother 6:417, 2006[PMID: 16533145] [Full Text]

Digiorgio et al: Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model. Nat Neurosci 10:608, 2007

Dion PA et al: Genetics of motor neuron disorders: new insights into pathogenic mechanisms. Nat Genet 10:769, 2009[PMID: 19823194] [Full Text]

Fink JK: Hereditary spastic paraplegia. Curr Neurol Neurosci Rep 6:65, 2006[PMID: 16469273] [Full Text]

Foust K et al: Rescue of the spinal muscular atrophy phenotype in a mouse model by early postnatal delivery of SMN. Nat Biotechnol 28:271, 2010[PMID: 20190738] [Full Text]

Kiernan MC et al: Amyotrophic lateral sclerosis. Lancet 377:942, 2011[PMID: 21296405] [Full Text]

Kwiatkowski T et al: Mutations in a novel ALS gene. Science 323:1205, 2009[PMID: 19251627] [Full Text]

Lagier-Tourenne C et al: TDP-43 and FUS/TLS: Emerging roles in RNA processing and neurodegeneration. Hum Mol Gen 19:R46, 2010

Mckee AC et al: TDP-43 Proteinopathy and motor neuron disease in chronic traumatic encephalopathy. J Neuropathol Exp Neurol 69:918, 2010[PMID: 20720505] [Full Text]

Smith RA et al: Antisense oligonucleotide therapy for neurodegenerative disease. J Clin Invest 116:2290, 2006[PMID: 16878173] [Full Text]

Weisskopf MG et al: Prospective study of military service and mortality from ALS. Neurology 64:32, 2005[PMID: 15642900] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS OF THE AUTONOMIC NERVOUS SYSTEM: INTRODUCTION

The autonomic nervous system (ANS) innervates the entire neuraxis and permeates all organ systems. It regulates blood pressure (BP), heart rate, sleep, and bladder and bowel function. It operates automatically; its full importance becomes recognized only when ANS function is compromised, resulting in dysautonomia. **Hypothalamic disorders that cause disturbances in homeostasis are discussed in Chaps. 16 and 339.**

ANATOMIC ORGANIZATION

The activity of the ANS is regulated by central neurons responsive to diverse afferent inputs. After central integration of afferent information, autonomic outflow is adjusted to permit the functioning of the major organ systems in accordance with the needs of the organism as a whole. Connections between the cerebral cortex and the autonomic centers in the brainstem coordinate autonomic outflow with higher mental functions.

The preganglionic neurons of the parasympathetic nervous system leave the central nervous system (CNS) in the third, seventh, ninth, and tenth cranial nerves as well as the second and third sacral nerves, while the preganglionic neurons of the sympathetic nervous system exit the spinal cord between the first thoracic and the second lumbar segments **(Fig. 375-1)**. These are thinly myelinated. The postganglionic neurons, located in ganglia outside the CNS, give rise to the postganglionic unmyelinated autonomic nerves that innervate organs and tissues throughout the body. Responses to sympathetic and parasympathetic stimulation are frequently antagonistic **(Table 375-1)**, reflecting highly coordinated interactions within the CNS; the resultant changes in parasympathetic and sympathetic activity provide more precise control of autonomic responses than could be achieved by the modulation of a single system.

Figure 375-1

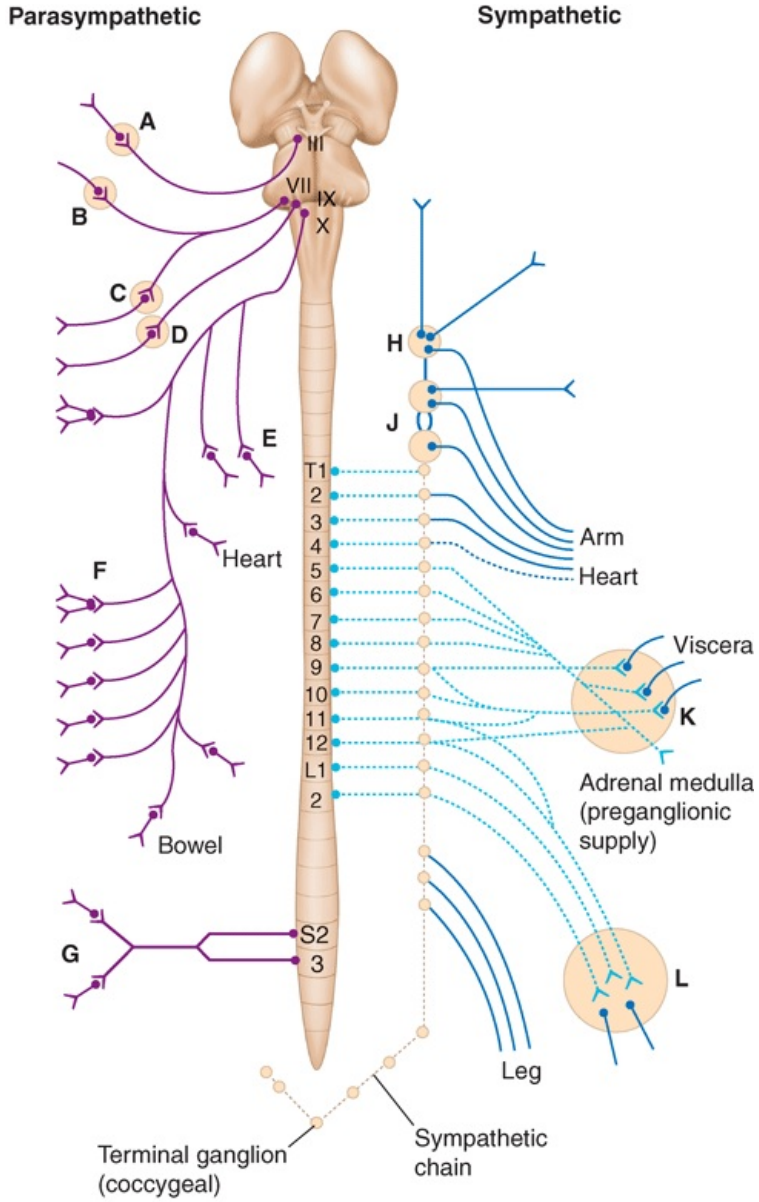

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic representation of the autonomic nervous system. *(From M Moskowitz: Clin Endocrinol Metab 6:77, 1977.)*

Table 375-1 Functional Consequences of Normal ANS Activation

	Sympathetic	Parasympathetic

Heart rate	Increased	Decreased
Blood pressure	Increased	Mildly decreased
Bladder	Increased sphincter tone	Voiding (decreased tone)
Bowel motility	Decreased motility	Increased
Lung	Bronchodilation	Bronchoconstriction
Sweat glands	Sweating	—
Pupils	Dilation	Constriction
Adrenal glands	Catecholamine release	—
Sexual function	Ejaculation, orgasm	Erection
Lacrimal glands	—	Tearing
Parotid glands	—	Salivation

Acetylcholine (ACh) is the preganglionic neurotransmitter for both divisions of the ANS as well as the postganglionic neurotransmitter of the parasympathetic neurons; the preganglionic receptors are nicotinic, and the postganglionic are muscarinic in type. Norepinephrine (NE) is the neurotransmitter of the postganglionic sympathetic neurons, except for cholinergic neurons innervating the eccrine sweat glands.

CLINICAL EVALUATION

Classification

Disorders of the ANS may result from pathology of either the CNS or the peripheral nervous system (PNS) **(Table 375-2)**. Signs and symptoms may result from interruption of the afferent limb, CNS processing centers, or efferent limb of reflex arcs controlling autonomic responses. For example, a lesion of the medulla produced by a posterior fossa tumor can impair BP responses to postural changes and result in orthostatic hypotension (OH). OH can also be caused by lesions of the spinal cord or peripheral vasomotor nerve fibers (e.g., diabetic autonomic neuropathy). Lesions of the efferent limb cause the most consistent and severe OH. The site of reflex interruption is usually established by the clinical context in which the dysautonomia arises, combined with judicious use of ANS testing and neuroimaging studies. The presence or absence of CNS signs, association with sensory or motor polyneuropathy, medical illnesses, medication use, and family history are often important considerations. Some syndromes do not fit easily into any classification scheme.

Table 375-2 Classification of Clinical Autonomic Disorders
I. Autonomic disorders with brain involvement
A. Associated with multisystem degeneration
1. Multisystem degeneration: autonomic failure clinically prominent
a. Multiple system atrophy (MSA)
b. Parkinson's disease with autonomic failure
c. Diffuse Lewy body disease (some cases)
2. Multisystem degeneration: autonomic failure clinically not usually prominent
a. Parkinson's disease
b. Other extrapyramidal disorders [inherited spinocerebellar atrophies, progressive supranuclear palsy, corticobasal degeneration, Machado-Joseph disease, fragile X syndrome (FXTAS)]
B. Unassociated with multisystem degeneration (Focal CNS disorders)
1. Disorders mainly due to cerebral cortex involvement
a. Frontal cortex lesions causing urinary/bowel incontinence
b. Partial complex seizures (temporal lobe or anterior cingulate)
c. Cerebral infarction of the insula
2. Disorders of the limbic and paralimbic circuits
a. Shapiro's syndrome (agenesis of corpus callosum, hyperhidrosis, hypothermia)
b. Autonomic seizures
c. Limbic encephalitis
3. Disorders of the hypothalamus
a. Wernicke-Korsakoff syndrome
b. Diencephalic syndrome
c. Neuroleptic malignant syndrome

d. Serotonin syndrome
e. Fatal familial insomnia
f. Antidiuretic hormone syndromes (diabetes insipidus, inappropriate ADH secretion)
g. Disturbances of temperature regulation (hyperthermia, hypothermia)
h. Disturbances of sexual function
i. Disturbances of appetite
j. Disturbances of BP/HR and gastric function
k. Horner's syndrome
4. Disorders of the brainstem and cerebellum
a. Posterior fossa tumors
b. Syringobulbia and Arnold-Chiari malformation
c. Disorders of BP control (hypertension, hypotension)
d. Cardiac arrhythmias
e. Central sleep apnea
f. Baroreflex failure
g. Horner's syndrome
h. Vertebrobasilar and Wallenberg syndromes
i. Brainstem encephalitis
II. Autonomic disorders with spinal cord involvement
A. Traumatic quadriplegia
B. Syringomyelia
C. Subacute combined degeneration
D. Multiple sclerosis and Devic disease
E. Amyotrophic lateral sclerosis
F. Tetanus
G. Stiff-man syndrome
H. Spinal cord tumors
III. Autonomic neuropathies
A. Acute/subacute autonomic neuropathies
1. Subacute autoimmune autonomic ganglionopathy (AAG)
a. Subacute paraneoplastic autonomic neuropathy
b. Guillain-Barré syndrome
c. Botulism
d. Porphyria
e. Drug induced autonomic neuropathies-stimulants, drug withdrawal, vasoconstrictor, vasodilators, beta-receptor antagonists, beta-agonists
f. Toxic autonomic neuropathies
g. Subacute cholinergic neuropathy
B. Chronic peripheral autonomic neuropathies
1. Distal small fiber neuropathy
2. Combined sympathetic and parasympathetic failure
a. Amyloid
b. Diabetic autonomic neuropathy
c. Autoimmune autonomic ganglionopathy (paraneoplastic and idiopathic)
d. Sensory neuronopathy with autonomic failure
e. Familial dysautonomia (Riley-Day syndrome)
f. Diabetic, uremic, or nutritional deficiency
g. Dysautonomia of old age
3. Disorders of reduced orthostatic intolerance-reflex syncope, POTS, associated with prolonged bed rest, associated with space flight, chronic fatigue

Abbreviations: BP, blood pressure; HR, heart rate; POTS, postural orthostatic tachycardia syndrome.

Symptoms of Autonomic Dysfunction

Clinical manifestations can result from loss of function, overactivity, or dysregulation of autonomic

circuits. Disorders of autonomic function should be considered in all patients with unexplained orthostatic hypotension, syncope, sleep dysfunction, altered sweating (hyperhidrosis or hypohidrosis), constipation, upper gastrointestinal symptoms (bloating, nausea, vomiting of old food), impotence, or bladder disorders (urinary frequency, hesitancy, or incontinence). Symptoms may be widespread or regional in distribution. An autonomic history focuses on systemic functions (BP, heart rate, sleep, fever, sweating) and involvement of individual organ systems (pupils, bowel, bladder, sexual function). The autonomic symptom profile is a self-report questionnaire that can be used for formal assessment. It is also important to recognize the modulating effects of age. For example, OH typically produces lightheadedness in the young, whereas cognitive slowing is more common in the elderly. Specific symptoms of orthostatic intolerance are diverse **(Table 375-3)**. Autonomic symptoms may vary dramatically, reflecting the dynamic nature of autonomic control over homeostatic function. For example, OH might be manifest only in the early morning, following a meal, with exercise, or with raised ambient temperature, depending upon the regional vascular bed affected by dysautonomia.

Table 375-3 Symptoms of Orthostatic Intolerance

Lightheadedness (dizziness)	88%
Weakness or tiredness	72%
Cognitive difficulty (thinking/concentrating)	47%
Blurred vision	47%
Tremulousness	38%
Vertigo	37%
Pallor	31%
Anxiety	29%
Palpitations	26%
Clammy feeling	19%
Nausea	18%

Source: From PA Low et al: Mayo Clin Proc 70:617, 1995.

Early symptoms may be overlooked. Impotence, although not specific for autonomic failure, often heralds autonomic failure in men and may precede other symptoms by years (Chap. 48). A decrease in the frequency of spontaneous early morning erections may occur months before loss of nocturnal penile tumescence and development of total impotence. Bladder dysfunction may appear early in men and women, particularly in those with CNS involvement. Cold feet may indicate peripheral vasomotor constriction. Brain and spinal cord disease above the level of the lumbar spine results first in urinary frequency and small bladder volumes and eventually in incontinence (upper motor neuron or spastic bladder). By contrast, PNS disease of autonomic nerve fibers results in large bladder volumes, urinary frequency, and overflow incontinence (lower motor neuron flaccid bladder). Measurement of bladder volume (postvoid residual) is a useful bedside test for distinguishing between upper and lower motor neuron bladder dysfunction in the early stages of dysautonomia. Gastrointestinal autonomic dysfunction typically presents as severe constipation. Diarrhea occurs occasionally (as in diabetes mellitus) due to rapid transit of contents or uncoordinated small-bowel motor activity, or on an osmotic basis from bacterial overgrowth associated with small-bowel stasis. Impaired glandular secretory function may cause difficulty with food intake due to decreased salivation or eye irritation due to decreased lacrimation. Occasionally, temperature elevation and vasodilation can result from anhidrosis because sweating is normally important for heat dissipation (Chap. 16). Lack of sweating after a hot bath, during exercise, or on a hot day can suggest sudomotor failure.

OH (also called *orthostatic or postural hypotension*) is perhaps the most disabling feature of autonomic dysfunction. The prevalence of OH is relatively high, especially when OH associated with aging and diabetes mellitus is included **(Table 375-4)**. OH can cause a variety of symptoms, including dimming or loss of vision, lightheadedness, diaphoresis, diminished hearing, pallor, and weakness. Syncope results when the drop in BP impairs cerebral perfusion. Other manifestations of impaired baroreflexes are supine hypertension, a heart rate that is fixed regardless of posture, postprandial hypotension, and an excessively high nocturnal BP. Many patients with OH have a preceding diagnosis of hypertension or have concomitant supine hypertension, reflecting the great importance of baroreflexes in maintaining postural and supine normotension. The appearance of OH in patients receiving antihypertensive treatment may indicate overtreatment or the onset of an autonomic disorder. The most common causes of OH are not neurologic in origin; these must be distinguished from the neurogenic causes **(Table 375-5)**. **Neurocardiogenic and cardiac causes of syncope are considered in Chap. 20.**

Table 375-4 Prevalence of Orthostatic Hypotension in Different Disorders

Disorder	Prevalence
Aging	14–20%
Diabetic neuropathy	10%
Other autonomic neuropathies	10–50 per 100,000
Multiple system atrophy	5–15 per 100,000
Pure autonomic failure	10–30 per 100,000

Table 375-5 Nonneurogenic Causes of Orthostatic Hypotension

Cardiac pump failure
Myocardial infarction
Myocarditis
Constrictive pericarditis
Aortic stenosis
Tachyarrhythmias
Bradyarrhythmias
Salt-losing nephropathy
Adrenal insufficiency
Diabetes insipidus
Venous obstruction
Reduced intravascular volume
Straining or heavy lifting,
urination, defecation
Dehydration
Diarrhea, emesis
Hemorrhage
Burns
Metabolic
Adrenocortical insufficiency
Hypoaldosteronism
Pheochromocytoma
Severe potassium depletion
Venous pooling
Alcohol
Postprandial dilation of splanchnic vessel beds
Vigorous exercise with dilation of skeletal vessel beds
Heat: hot environment, hot showers and baths, fever
Prolonged recumbency or standing
Sepsis
Medications
Antihypertensives
Diuretics
Vasodilators: nitrates, hydralazine
Alpha- and beta-blocking agents
CNS sedatives: barbiturates, opiates
Tricyclic antidepressants
Phenothiazines

Approach to the Patient: Orthostatic Hypotension and Other ANS Disorders

The first step in the evaluation of symptomatic OH is the exclusion of treatable causes. The history should include a review of medications that may affect the autonomic system **(Table 375-6)**. The main classes of drugs that may cause OH are diuretics, antihypertensives, antidepressants, phenothiazines, ethanol, narcotics, insulin, dopamine agonists, barbiturates, and calcium channel-blocking agents. However, the precipitation of OH by medications may also be the first sign of an underlying autonomic

disorder. The history may reveal an underlying cause for symptoms (e.g., diabetes, Parkinson's disease) or specific underlying mechanisms (e.g., cardiac pump failure, reduced intravascular volume). The relationship of symptoms to meals (splanchnic pooling), standing on awakening in the morning (intravascular volume depletion), ambient warming (vasodilatation), or exercise (muscle arteriolar vasodilatation) should be sought. Standing time to first symptom and presyncope should be followed for management.

Table 375-6 Some Drugs that Affect Autonomic Function

Symptom	Drug Class	Specific Examples
Impotence	Opioids	Tylenol #3
	Anabolic steroids	—
	Some antiarrhythmics	Prazosin
	Some antihypertensives	Clonidine
	Some diuretics	Benazepril
	Some SSRIs	Venlafaxine
Urinary retention	Opioids	Fentanyl
	Decongestants	Brompheniramine
		Diphenhydramine
Diaphoresis	Some antihypertensives	Amlodipine
	Some SSRIs	Citalopram
	Opioids	Morphine
Hypotension	Tricyclics	Amitriptyline
	Beta blockers	Propranolol
	Diuretics	HCTZ
	CCBs	Verapamil

Abbreviations: CCBs, calcium channel blockers; HCTZ, hydrochlorothiazide; SSRIs, selective serotonin reuptake inhibitors.

Physical examination includes measurement of supine and standing pulse and BP. OH is defined as a sustained drop in systolic (≥20 mm Hg) or diastolic (≥10 mm Hg) BP within 3 min of standing. In nonneurogenic causes of OH (such as hypovolemia), the BP drop is accompanied by a compensatory increase in heart rate of >15 beats/min. A clue that the patient has neurogenic OH is the aggravation or precipitation of OH by autonomic stressors (such as a meal, hot tub/hot bath, and exercise). Neurologic examination should include mental status (neurodegenerative disorders), cranial nerves (impaired downgaze with progressive supranuclear palsy; abnormal pupils with Horner's or Adie's syndrome), motor tone (Parkinson's disease and parkinsonian syndromes), and reflexes and sensation (polyneuropathies). In patients without a clear diagnosis initially, follow-up evaluations may reveal the underlying cause.

Disorders of autonomic function should be considered in patients with symptoms of altered sweating (hyperhidrosis or hypohidrosis), gastroparesis (bloating, nausea, vomiting of old food), constipation, impotence, or bladder dysfunction (urinary frequency, hesitancy, or incontinence).

AUTONOMIC TESTING

Autonomic function tests are helpful when the history and examination findings are inconclusive; to detect subclinical involvement; or to follow the course of an autonomic disorder.

Heart Rate Variation with Deep Breathing

This is a test of the parasympathetic component of cardiovascular reflexes, via the vagus nerve. Results are influenced by multiple factors including the subject's position (recumbent, sitting or standing), rate and depth of respiration [6 breaths per minute and a forced vital capacity (FVC) >1.5 L are optimal], age, medications, weight, and degree of hypocapnia. Interpretation of results requires comparison of test data with results from age-matched controls collected under identical test conditions. For example, the lower limit of normal heart rate variation with deep breathing in persons <20 years is >15–20 beats/min, but for persons over age 60 it is 5–8 beats/min. Heart rate variation with deep breathing (respiratory sinus arrhythmia) is abolished by the muscarinic acetylcholine (ACh)-receptor antagonist atropine but is unaffected by sympathetic postganglionic blockade (e.g., propranolol).

Valsalva Response

This response (Table 375-7) assesses integrity of the baroreflex control of heart rate (parasympathetic) and BP (adrenergic). Under normal conditions, increases in BP at the carotid bulb trigger a reduction in heart rate (increased vagal tone), and decreases in BP trigger an increase in heart rate (reduced vagal tone). The Valsalva response is tested in the supine position. The subject exhales against a closed glottis (or into a manometer maintaining a constant expiratory pressure of 40 mm Hg) for 15 s while measuring changes in heart rate and beat-to-beat BP. There are four phases of BP and heart rate response to the Valsalva maneuver. Phases I and III are mechanical and related to changes in intrathoracic and intraabdominal pressure. In early phase II, reduced venous return results in a fall in stroke volume and BP, counteracted by a combination of reflex tachycardia and increased total peripheral resistance. Increased total peripheral resistance arrests the BP drop ~5–8 s after the onset of the maneuver. Late phase II begins with a progressive rise in BP toward or above baseline. Venous return and cardiac output return to normal in phase IV. Persistent peripheral arteriolar vasoconstriction and increased cardiac adrenergic tone result in a temporary BP overshoot and phase IV bradycardia (mediated by the baroreceptor reflex).

Table 375-7 Normal Blood Pressure and Heart Rate Changes during the Valsalva Maneuver

Phase	Maneuver	Blood Pressure	Heart Rate	Comments
I	Forced expiration against a partially closed glottis	Rises; aortic compression from raised intrathoracic pressure	Decreases	Mechanical
II early	Continued expiration	Falls; decreased venous return to the heart	Increases (reflex tachycardia)	Reduced vagal tone
II late	Continued expiration	Rises; reflex increase in peripheral vascular resistance	Increases at slower rate	Requires intact efferent sympathetic response
III	End of expiration	Falls; increased capacitance of pulmonary bed	Increases further	Mechanical
IV	Recovery	Rises; persistent vasoconstriction and increased cardiac output	Compensatory bradycardia	Requires intact efferent sympathetic response

Autonomic function during the Valsalva maneuver can be measured using beat-to-beat blood pressure or heart rate changes. The *Valsalva ratio* is defined as the maximum phase II tachycardia divided by the minimum phase IV bradycardia (**Table 375-8**). The ratio reflects the integrity of the entire baroreceptor reflex arc and of sympathetic efferents to blood vessels.

Table 375-8 Neural Pathways Underlying Some Standardized Autonomic Tests

Test Evaluated	Procedure	Autonomic Function
HRDB	6 deep breaths/min	Cardiovagal function
Valsalva ratio	Expiratory pressure, 40 mm Hg for 10–15 s	Cardiovagal function
QSART	Axon-reflex test 4 limb sites	Postganglionic sudomotor function
BP_{BB} to VM	BP^{BB} response to VM	Adrenergic function: baroreflex adrenergic control of vagal and vasomotor function
HUT	BP_{BB} and heart rate response to HUT	Adrenergic and cardiovagal responses to HUT

Abbreviations: BP_{BB}, beat-to-beat blood pressure; HRDB, heart rate response to deep breathing; HUT, head-up tilt; QSART, quantitative sudomotor axon-reflex test; VM, Valsalva maneuver.

Sudomotor Function

Sweating is induced by release of ACh from sympathetic postganglionic fibers. The quantitative sudomotor axon reflex test (QSART) is a measure of regional autonomic function mediated by ACh-induced sweating. A reduced or absent response indicates a lesion of the postganglionic sudomotor axon. For example, sweating may be reduced in the feet as a result of distal polyneuropathy (e.g., diabetes). The thermoregulatory sweat test (TST) is a qualitative measure of regional sweat production in response to an elevation of body temperature under controlled conditions. An indicator

powder placed on the anterior surface of the body changes color with sweat production during temperature elevation. The pattern of color changes is a measure of regional sweat secretion. A postganglionic lesion is present if both QSART and TST show absent sweating. In a preganglionic lesion, QSART is normal but TST shows anhidrosis.

Orthostatic BP Recordings

Beat-to-beat BP measurements determined in supine, 70° tilt, and tilt-back positions are useful to quantitate orthostatic failure of BP control. Allow a 20-min period of rest in the supine position before assessing changes in BP during tilting. The BP change combined with heart rate monitoring is useful for the evaluation of patients with suspected OH or unexplained syncope.

Tilt Table Testing for Syncope

The great majority of patients with syncope do not have autonomic failure. Tilt table testing can be used to make the diagnosis of vasovagal syncope with sensitivity, specificity, and reproducibility. A standardized protocol is used that specifies the tilt apparatus, angle and duration of tilt, and procedure for provocation of vasodilation (e.g., sublingual or spray nitroglycerin). A positive nitroglycerin-stimulated test predicts recurrence of syncope. Recommendations for the performance of tilt studies for syncope have been incorporated in consensus guidelines.

SPECIFIC SYNDROMES OF ANS DYSFUNCTION

Multiple System Atrophy (Chap. 372)

Multiple system atrophy (MSA) is an entity that comprises autonomic failure (OH or a neurogenic bladder) *and* either parkinsonism (MSA-p) or a cerebellar syndrome (MSA-c). MSA-p is the more common form; the parkinsonism is atypical in that it is usually unassociated with significant tremor or response to levodopa. Symptomatic OH within 1 year of onset of parkinsonism predicts eventual development of MSA-p in 75% of patients. Although autonomic abnormalities are common in advanced Parkinson's disease (Chap. 372), the severity and distribution of autonomic failure is more severe and more generalized in MSA. Brain MRI is a useful diagnostic adjunct; in MSA-p, iron deposition in the striatum may be evident as T2 hypointensity, and in MSA-c cerebellar atrophy is present with a characteristic T2 hyperintense signal ("hot cross buns sign") in the pons (**Fig. 375-2**). Cardiac postganglionic adrenergic innervation, measured by uptake of fluorodopamine on positron emission tomography, is markedly impaired in the dysautonomia of Parkinson's disease but is usually normal in MSA.

Figure 375-2

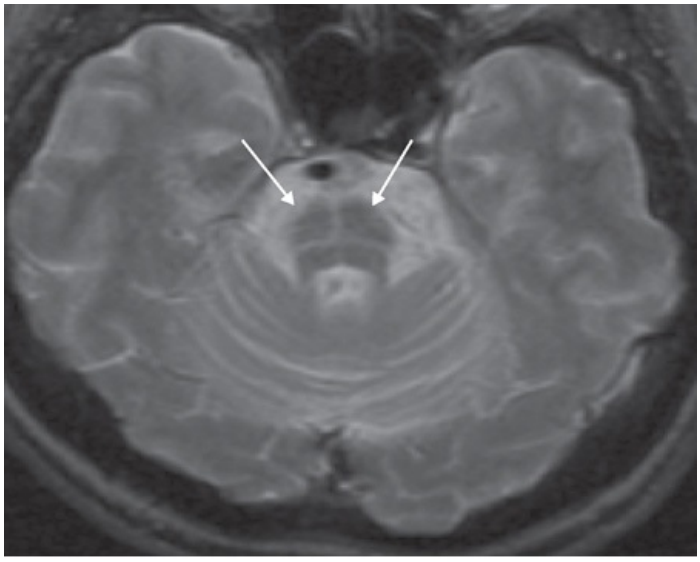

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple system atrophy, cerebellar type (MSA-c). Axial T2-weighted MRI at the level of the pons shows a characteristic hyperintense signal, the "hot cross buns" sign.

MSA generally progresses relentlessly to death 7–10 years after onset. Neuropathologic changes include neuronal loss and gliosis in many CNS regions, including the brainstem, cerebellum, striatum, and intermediolateral cell column of the thoracolumbar spinal cord. Management is symptomatic for neurogenic OH (see below), gastrointestinal (GI), and urinary dysfunction. GI management includes frequent small meals, soft diet, stool softeners, and bulk agents. Gastroparesis is difficult to treat; metoclopramide stimulates gastric emptying but worsens parkinsonism by blocking central dopamine

receptors. Domperidone has been used in other countries but is not available in the United States.

Autonomic dysfunction is also a common feature in dementia with Lewy bodies (Chap. 371); the severity is usually less than that found in MSA or Parkinson's disease. In multiple sclerosis (MS; Chap. 380), autonomic complications reflect the CNS location of MS involvement and generally worsen with disease duration and disability.

Spinal Cord

Spinal cord lesions from any cause may result in focal autonomic deficits or autonomic hyperreflexia (e.g., spinal cord transection or hemisection) affecting bowel, bladder, sexual, temperature-regulation, or cardiovascular functions. Quadriparetic patients exhibit both supine hypertension and OH after upward tilting. *Autonomic dysreflexia* describes a dramatic increase in blood pressure in patients with traumatic spinal cord lesions above the C6 level, often in response to stimulation of the bladder, skin, or muscles. Suprapubic palpation of the bladder, a distended bladder, catheter insertion, catheter obstruction, or urinary infection are common triggers. Associated symptoms can include flushing, headache, or piloerection. Potential complications include intracranial vasospasm or hemorrhage, cardiac arrhythmia, and death. Awareness of the syndrome and monitoring of blood pressure during procedures in patients with acute or chronic spinal cord injury is essential. In patients with supine hypertension, BP can be lowered by tilting the head upward. Vasodilator drugs may be used to treat acute elevations in BP. Clonidine can be used prophylactically to reduce the hypertension resulting from bladder stimulation. Dangerous increases or decreases in body temperature may result from an inability to experience the sensory accompaniments of heat or cold exposure or the ability to control peripheral vasoconstriction or sweating below the level of the spinal cord injury.

Peripheral Nerve and Neuromuscular Junction Disorders

Peripheral neuropathies (Chap. 384) are the most common cause of chronic autonomic insufficiency. Polyneuropathies that affect small myelinated and unmyelinated fibers of the sympathetic and parasympathetic nerves commonly occur in diabetes mellitus, amyloidosis, chronic alcoholism, porphyria, and Guillain-Barré syndrome. Neuromuscular junction disorders with autonomic involvement include botulism and Lambert-Eaton syndrome (Chap. 386).

DIABETES MELLITUS

Autonomic neuropathy typically begins ~10 years after the onset of diabetes and is slowly progressive. Clinical features, prevention, and management are discussed in Chap. 344.

AMYLOIDOSIS

Autonomic neuropathy occurs in both sporadic and familial forms of amyloidosis (Chap. 112). The AL (immunoglobulin light chain) type is associated with primary amyloidosis or amyloidosis secondary to multiple myeloma. The ATTR type, with transthyretin as the primary protein component, is responsible for the most common form of inherited amyloidosis. Although patients usually present with a distal painful neuropathy accompanied by sensory loss, autonomic insufficiency can precede the development of the polyneuropathy or occur in isolation. Diagnosis can be made by protein electrophoresis of blood and urine, tissue biopsy (abdominal fat pad, rectal mucosa, or sural nerve) to search for amyloid deposits, and genetic testing for transthyretin mutations in familial cases. Treatment of familial cases with liver transplantation can be successful. The response of primary amyloidosis to melphalan and stem cell transplantation has been mixed. Death is usually due to cardiac or renal involvement. Postmortem studies reveal amyloid deposition in many organs, including two sites that contribute to autonomic failure: intraneural blood vessels and autonomic ganglia. Pathologic examination reveals a loss of unmyelinated and myelinated nerve fibers.

ALCOHOLIC NEUROPATHY

Abnormalities in parasympathetic vagal and efferent sympathetic function are usually mild in individuals with alcoholic polyneuropathy. Pathologic changes can be demonstrated in the parasympathetic (vagus) and sympathetic fibers, and in ganglia. OH is usually due to brainstem involvement. Impotence is a major problem, but concurrent gonadal hormone abnormalities may obscure the parasympathetic component. Clinical symptoms of autonomic failure generally appear when the polyneuropathy is severe, and there is usually coexisting Wernicke's encephalopathy (Chap. 275). Autonomic involvement may contribute to the high mortality rates associated with alcoholism (Chap. 392).

PORPHYRIA (CHAP. 358)

Although each of the porphyrias can cause autonomic dysfunction, the condition is most extensively documented in the acute intermittent type. Autonomic symptoms include tachycardia, sweating, urinary retention, hypertension, or (less commonly) hypotension. Other prominent symptoms include anxiety, abdominal pain, nausea, and vomiting. Abnormal autonomic function can occur both during acute attacks and during remissions. Elevated catecholamine levels during acute attacks correlate with the degree of tachycardia and hypertension that is present.

GUILLAIN-BARRÉ SYNDROME (CHAP. 385)

BP fluctuations and arrhythmias can be severe. It is estimated that between 2 and 10% of patients with

severe Guillain-Barré syndrome suffer fatal cardiovascular collapse. Gastrointestinal autonomic involvement, sphincter disturbances, abnormal sweating, and pupillary dysfunction also occur. Demyelination has been described in the vagus and glossopharyngeal nerves, the sympathetic chain, and the white rami communicantes. Interestingly, the degree of autonomic involvement appears to be independent of the severity of motor or sensory neuropathy.

AUTOIMMUNE AUTONOMIC NEUROPATHY (AAN)

This disorder presents with the subacute development of autonomic disturbances with OH, enteric neuropathy (gastroparesis, ileus, constipation/diarrhea), and cholinergic failure; the latter consists of loss of sweating, sicca complex, and a tonic pupil. Autoantibodies against the ganglionic ACh receptor (A_3 AChR) are present in the serum of many patients and are now considered to be diagnostic of this syndrome. In general, the antibody titer correlates with the severity of autonomic failure. Symptoms of cholinergic failure are also associated with a high antibody titer. Onset of the neuropathy follows a viral infection in approximately half of cases. AAN is almost always monophasic; up to one-third of untreated patients experience significant functional improvement over time. There are isolated case reports of a beneficial response to plasmapheresis or intravenous immune globulin, but there are no clinical trials that systematically assess the effectiveness of immunomodulatory therapies. Symptomatic management of OH, gastroparesis, and sicca symptoms is essential. The spectrum of AAN is now broader than originally thought; some antibody-positive cases have an insidious onset and slow progression with a pure autonomic failure (see below) phenotype. A dramatic clinical response to repeated plasma exchange combined with immunosuppression was described in one patient with longstanding AAN.

AAN can have a paraneoplastic basis (Chap. 101). The clinical features of the autonomic neuropathy may be indistinguishable from a coexisting paraneoplastic syndrome, although quite often in the paraneoplastic cases, distinctive additional central features, such as cerebellar involvement or dementia, may be present (see Tables 101-1, 101-2, and 101-3). The neoplasm may be truly occult and possibly suppressed by the autoantibody.

BOTULISM

Botulinum toxin binds presynaptically to cholinergic nerve terminals and, after uptake into the cytosol, blocks ACh release. Manifestations consist of motor paralysis and autonomic disturbances that include blurred vision, dry mouth, nausea, unreactive or sluggishly reactive pupils, constipation, and urinary retention (Chap. 141).

Pure Autonomic Failure (PAF)

This sporadic syndrome consists of postural hypotension, impotence, bladder dysfunction, and defective sweating. The disorder begins in the middle decades and occurs in women more often than men. The symptoms can be disabling, but the disease does not shorten life span. The clinical and pharmacologic characteristics suggest primary involvement of postganglionic sympathetic neurons. There is a severe reduction in the density of neurons within sympathetic ganglia that results in low supine plasma NE levels and noradrenergic supersensitivity. Some studies have questioned the specificity of PAF as a distinct clinical entity. Some cases are ganglionic antibody–positive and thus represent a type of AAN. Between 10 and 15% of cases evolve into MSA.

Postural Orthostatic Tachycardia Syndrome (POTS)

This syndrome is characterized by symptomatic orthostatic intolerance (not OH) and by either an increase in heart rate to >120 beats/min or an increase of 30 beats/min with standing that subsides on sitting or lying down. Women are affected approximately five times more often than men, and most develop the syndrome between the ages of 15 and 50. Approximately half of affected patients report an antecedent viral infection. Syncopal symptoms (lightheadedness, weakness, blurred vision) combined with symptoms of autonomic overactivity (palpitations, tremulousness, nausea) are common. Recurrent unexplained episodes of dysautonomia and fatigue also occur. The pathogenesis is unclear in most cases; hypovolemia, deconditioning, venous pooling, impaired brainstem regulation, or β-receptor supersensitivity may play a role. In one affected individual, a mutation in the NE transporter, which resulted in impaired NE clearance from synapses, was responsible. Some cases are due to an underlying limited autonomic neuropathy. Although ~80% of patients improve, only one-quarter eventually resume their usual daily activities (including exercise and sports). Expansion of fluid volume and postural training (see "Treatment: Autonomic Failure") are initial approaches to treatment. If these approaches are inadequate, then midodrine, fludrocortisone, phenobarbital, beta blockers, or clonidine may be used with some success. Reconditioning and a sustained exercise program are very important.

Inherited Disorders

There are five known hereditary sensory and autonomic neuropathies (HSAN I–V). The most important ones are HSAN I and HSAN III (Riley-Day syndrome; familial dysautonomia). HSAN I is dominantly inherited and often presents as a distal small-fiber neuropathy (burning feet syndrome). The responsible gene, on chromosome 9q, is designated *SPTLC1*. SPTLC is an important enzyme in the regulation of ceramide. Cells from HSAN I patients affected by mutation of *SPTLC1* produce higher-

than-normal levels of glucosyl ceramide, perhaps triggering apoptosis.

HSAN III, an autosomal recessive disorder of infants and children that occurs among Ashkenazi Jews, is much less prevalent than HSAN I. Decreased tearing, hyperhidrosis, reduced sensitivity to pain, areflexia, absent fungiform papillae on the tongue, and labile BP may be present. Episodic abdominal crises and fever are common. Pathologic examination of nerves reveals a loss of small myelinated and unmyelinated nerve fibers. The defective gene, named *IKBKAP*, is also located on the long arm of chromosome 9. Pathogenic mutations may prevent normal transcription of important molecules in neural development.

Primary Hyperhidrosis

This syndrome presents with excess sweating of the palms of the hands and soles of the feet. The disorder affects 0.6–1.0% of the population; the etiology is unclear, but there may be a genetic component. While not dangerous, the condition can be socially embarrassing (e.g., shaking hands) or disabling (e.g., inability to write without soiling the paper). Onset of symptoms is usually in adolescence; the condition tends to improve with age. Topical antiperspirants are occasionally helpful. More useful are potent anticholinergic drugs such as glycopyrrolate (1–2 mg PO tid). T2 ganglionectomy or sympathectomy is successful in >90% of patients with palmar hyperhidrosis. The advent of endoscopic transaxillary T2 sympathectomy has lowered the complication rate of the procedure. The most common postoperative complication is compensatory hyperhidrosis, which improves spontaneously over months; other potential complications include recurrent hyperhidrosis (16%), Horner's syndrome (<2%), gustatory sweating, wound infection, hemothorax, and intercostal neuralgia. Local injection of botulinum toxin has also been used to block cholinergic, postganglionic sympathetic fibers to sweat glands in patients with palmar hyperhidrosis. This approach is limited by the need for repetitive injections (the effect usually lasts 4 months before waning), pain with injection, the high cost of botulinum toxin, and the possibility of temporary intrinsic hand muscle weakness.

Acute Autonomic Syndromes

The physician may be confronted occasionally with an acute autonomic syndrome, either acute autonomic failure (acute AAN syndrome) or a state of sympathetic overactivity. An *autonomic storm* is an acute state of sustained sympathetic surge that results in variable combinations of alterations in blood pressure and heart rate, body temperature, respiration, and sweating. Causes of autonomic storm are brain and spinal cord injury, toxins and drugs, autonomic neuropathy, and chemodectomas (e.g., pheochromocytoma).

Brain injury is most commonly a cause of autonomic storm following severe head trauma and in postresuscitation encephalopathy following anoxic-ischemic brain injury. Autonomic storm can also occur with other acute intracranial lesions such as hemorrhage, cerebral infarction, rapidly expanding tumors, subarachnoid hemorrhage, hydrocephalus, or (less commonly) an acute spinal cord lesion. Lesions involving the diencephalon may be more prone to present with dysautonomia, but the most consistent setting is that of an acute intracranial catastrophe of sufficient size and rapidity to produce a massive catecholaminergic surge. The surge can cause seizures, neurogenic pulmonary edema, and myocardial injury. Manifestations include fever, tachycardia, hypertension, tachypnea, hyperhidrosis, pupillary dilatation, and flushing. Lesions of the afferent limb of the baroreflex can result in milder recurrent autonomic storms; many of these follow neck irradiation.

Drugs and toxins may also be responsible, including sympathomimetics such as phenylpropanolamine, cocaine, amphetamines, and tricyclic antidepressants; tetanus; and, less often, botulinum. Phenylpropanolamine, now off the market, was in the past a potent cause of this syndrome. Cocaine, including "crack," can cause a hypertensive state with CNS hyperstimulation. Tricyclic overdose, such as amitriptyline, can cause flushing, hypertension, tachycardia, fever, mydriasis, anhidrosis, and a toxic psychosis. *Neuroleptic malignant syndrome* refers to a syndrome of muscle rigidity, hyperthermia, and hypertension in psychotic patients treated with phenothiazines.

The hyperadrenergic state with Guillain-Barré syndrome can produce a moderate autonomic storm. Pheochromocytoma presents with a paroxysmal or sustained hyperadrenergic state, headache, hyperhidrosis, palpitations, anxiety, tremulousness, and hypertension.

Management of autonomic storm includes ruling out other causes of autonomic instability, including malignant hyperthermia, porphyria, and epilepsy. Sepsis and encephalitis need to be excluded with appropriate studies. An electroencephalogram (EEG) should be done to detect epileptiform activity; MRI of the brain and spine is often necessary. The patient should be managed in an intensive care unit. Management with morphine sulphate (10 mg every 4 h) and labetalol (100–200 mg twice daily) have worked relatively well. Treatment may need to be maintained for several weeks. For chronic and milder autonomic storm, propranolol and/or clonidine can be effective.

Miscellaneous

Other conditions associated with autonomic failure include infections, poisoning (organophosphates), malignancy, and aging. Disorders of the hypothalamus can affect autonomic function and produce abnormalities in temperature control, satiety, sexual function, and circadian rhythms (Chap. 339).

Reflex Sympathetic Dystrophy and Causalgia

The failure to identify a primary role of the ANS in the pathogenesis of these disorders has resulted in a change of nomenclature. Complex regional pain syndrome (CRPS) types I and II are now used in place of reflex sympathetic dystrophy (RSD) and causalgia, respectively.

CRPS type I is a regional pain syndrome that usually develops after tissue trauma. Examples of associated trauma include myocardial infarction, minor shoulder or limb injury, and stroke. *Allodynia* (the perception of a nonpainful stimulus as painful), *hyperpathia* (an exaggerated pain response to a painful stimulus), and spontaneous pain occur. The symptoms are unrelated to the severity of the initial trauma and are not confined to the distribution of a single peripheral nerve. CRPS type II is a regional pain syndrome that develops after injury to a specific peripheral nerve, usually a major nerve trunk. Spontaneous pain initially develops within the territory of the affected nerve but eventually may spread outside the nerve distribution.

Pain is the primary clinical feature of CRPS. Vasomotor dysfunction, sudomotor abnormalities, or focal edema may occur alone or in combination but must be present for diagnosis. Limb pain syndromes that do not meet these criteria are best classified as "limb pain—not otherwise specified." In CRPS, localized sweating (increased resting sweat output) and changes in blood flow may produce temperature differences between affected and unaffected limbs.

CRPS type I (RSD) has classically been divided into three clinical phases but is now considered to be more variable. Phase I consists of pain and swelling in the distal extremity occurring within weeks to 3 months after the precipitating event. The pain is diffuse, spontaneous, and either burning, throbbing, or aching in quality. The involved extremity is warm and edematous, and the joints are tender. Increased sweating and hair growth develop. In phase II (3–6 months after onset), thin, shiny, cool skin appears. After an additional 3–6 months (phase III), atrophy of the skin and subcutaneous tissue plus flexion contractures complete the clinical picture.

The natural history of typical CRPS may be more benign than reflected in the literature. A variety of surgical and medical treatments have been developed, with conflicting reports of efficacy. Clinical trials suggest that early mobilization with physical therapy or a brief course of glucocorticoids may be helpful for CRPS type I. Other medical treatments include the use of adrenergic blockers, nonsteroidal anti-inflammatory drugs, calcium channel blockers, phenytoin, opioids, and calcitonin. Stellate ganglion blockade is a commonly used invasive technique that often provides temporary pain relief, but the efficacy of repetitive blocks is uncertain.

Treatment: Autonomic Failure

Management of autonomic failure is aimed at specific treatment of the cause and alleviation of symptoms. Of particular importance is the removal of drugs or amelioration of underlying conditions that cause or aggravate the autonomic symptoms, especially in the elderly. For example, OH can be caused or aggravated by angiotensin-converting enzyme inhibitors, calcium channel-blocking agents, tricyclic antidepressants, levodopa, alcohol, or insulin. A summary of drugs that can cause OH by class, putative mechanism, and magnitude of the BP drop, is shown in Table 375-6.

PATIENT EDUCATION

Only a minority of patients with OH require drug treatment. All patients should be taught the mechanisms of postural normotension (volume status, resistance and capacitance bed, autoregulation) and the nature of orthostatic stressors (time of day and the influence of meals, heat, standing, and exercise). Patients should learn to recognize orthostatic symptoms early (especially subtle cognitive symptoms, weakness, and fatigue) and to modify or avoid activities that provoke episodes. Other helpful measures may include keeping a BP log and dietary education (salt/fluids). Learning physical countermaneuvers that reduce standing OH and practicing postural and resistance training are helpful measures.

SYMPTOMATIC TREATMENT

Nonpharmacologic approaches are summarized in **Table 375-9**. Adequate intake of salt and fluids to produce a voiding volume between 1.5 and 2.5 L of urine (containing >170 meq/L of Na^+) each 24 h is essential. Sleeping with the head of the bed elevated will minimize the effects of supine nocturnal hypertension. Prolonged recumbency should be avoided when possible. Patients are advised to sit with legs dangling over the edge of the bed for several minutes before attempting to stand in the morning; other postural stresses should be similarly approached in a gradual manner. One maneuver that can reduce OH is leg-crossing with maintained contraction of leg muscles for 30 s.; this compresses leg veins and increases systemic resistance. Compressive garments, such as compression stockings and abdominal binders, are helpful on occasion but uncomfortable for some patients. Anemia should be corrected with erythropoietin, administered subcutaneously at doses of 25–75 U/kg three times per week. The hematocrit increases after 2–6 weeks. A weekly maintenance dose is usually necessary. The increased intravascular volume that accompanies the rise in hematocrit can exacerbate supine hypertension.

Table 375-9 Initial Treatment of Orthostatic Hypotension (Oh)
Patient education: mechanisms and stressors of OH
High-salt diet (10–20 g/d) High-fluid intake (2 L/D)
Elevate head of bed 10 cm (4 in.)
Maintain postural stimuli
Learn physical countermaneuvers
Compression garments
Correct anemia

If these measures are not sufficient, drug treatment may be necessary. Midodrine, a directly acting α_1-agonist that does not cross the blood-brain barrier, is effective. It has a duration of action of 2–4 h. The usual dose is 5–10 mg orally tid, but some patients respond best to a decremental dose (e.g., 15 mg on awakening, 10 mg at noon, and 5 mg in the afternoon). Midodrine should not be taken after 6 p.m. Side effects include pruritus, uncomfortable piloerection, and supine hypertension especially at higher doses. Pyridostigmine appears to improve OH without aggravating supine hypertension by enhancing ganglionic transmission (maximal when orthostatic, minimal supine). Fludrocortisone will reduce OH, but it aggravates supine hypertension. At doses between 0.1 mg/d and 0.3 mg bid orally, it enhances renal sodium conservation and increases the sensitivity of arterioles to NE. Susceptible patients may develop fluid overload, congestive heart failure, supine hypertension, or hypokalemia. Potassium supplements are often necessary with chronic administration of fludrocortisone. Sustained elevations of supine BP >180/110 mm Hg should be avoided.

Postprandial OH may respond to several measures. Frequent, small, low-carbohydrate meals may diminish splanchnic shunting of blood after meals and reduce postprandial OH. Prostaglandin inhibitors (ibuprofen or indomethacin) taken with meals or midodrine (10 mg with the meal) can be helpful. The somatostatin analogue octreotide can be useful in the treatment of postprandial syncope by inhibiting the release of gastrointestinal peptides that have vasodilator and hypotensive effects. The subcutaneous dose ranges from 25 μg bid to 200 μg tid.

The patient should be taught to self-treat transient worsening of OH. Drinking two 250-mL (8-oz) glasses of water can raise standing BP 20–30 mm Hg for about 2 h, beginning ~20 min after the fluid load. The patient can increase intake of salt and fluids (bouillon treatment), increase use of physical countermaneuvers, temporarily resort to a full-body stocking (compression pressure 30–40 mm Hg), or increase the dose of midodrine. Supine hypertension (>180/110 mm Hg) can be self-treated by avoiding the supine position and reducing fludrocortisone. A daily glass of wine, if requested by the patient, can be taken shortly before bedtime. If these simple measures are not adequate, drugs to be considered include oral hydralazine (25 mg qhs), oral Procardia (10 mg qhs), or a nitroglycerin patch.

FURTHER READINGS

Low PA, Benarroch EE: *Clinical Autonomic Disorders*, 3rd ed. Philadelphia, Lippincott, 2009

Schroeder C et al: Plasma exchange for primary autoimmune autonomic failure. N Engl J Med 353:1585, 2005[PMID: 16221781] [Full Text]

Thaisetthawatkul P et al: Autonomic dysfunction in dementia with Lewy bodies. Neurology 62:1804, 2004[PMID: 15159482] [Full Text]

Thieben MJ et al: Postural orthostatic tachycardia syndrome: The Mayo Clinic experience. Mayo Clin Proc 82:308, 2007[PMID: 17352367] [Full Text]

Vinik AI, Ziegler D: Diabetic cardiovascular autonomic neuropathy. Circulation 115:387, 2007[PMID: 17242296] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

TRIGEMINAL NEURALGIA, BELL'S PALSY, AND OTHER CRANIAL NERVE DISORDERS: INTRODUCTION

Symptoms and signs of cranial nerve pathology are common in internal medicine. They often develop in the context of a widespread neurologic disturbance, and in such situations cranial nerve involvement may represent the initial manifestation of the illness. In other disorders, involvement is largely restricted to one or several cranial nerves; these distinctive disorders are reviewed in this chapter. **Disorders of ocular movement are discussed in Chap. 28, disorders of hearing in Chap. 30, and vertigo and disorders of vestibular function in Chap. 21.**

FACIAL PAIN OR NUMBNESS

Anatomic Considerations

The trigeminal (fifth cranial) nerve supplies sensation to the skin of the face and anterior half of the head **(Fig. 376-1)**. Its motor part innervates the masseter and pterygoid masticatory muscles.

Figure 376-1

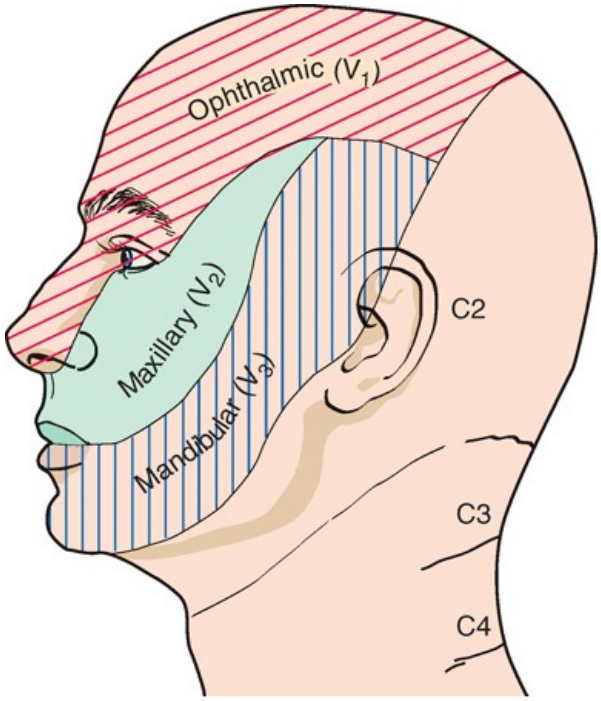

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The three major sensory divisions of the trigeminal nerve consist of the ophthalmic, maxillary, and mandibular nerves.

Trigeminal Neuralgia (Tic Douloureux)

CLINICAL MANIFESTATIONS

Trigeminal neuralgia is characterized by excruciating paroxysms of pain in the lips, gums, cheek, or chin and, very rarely, in the distribution of the ophthalmic division of the fifth nerve. The pain seldom lasts more than a few seconds or a minute or two but may be so intense that the patient winces, hence the term *tic*. The paroxysms, experienced as single jabs or clusters, tend to recur frequently, both day and night, for several weeks at a time. They may occur spontaneously or with movements of affected areas evoked by speaking, chewing, or smiling. Another characteristic feature is the presence of trigger zones, typically on the face, lips, or tongue, that provoke attacks; patients may report that tactile

stimuli—e.g., washing the face, brushing the teeth, or exposure to a draft of air—generate excruciating pain. *An essential feature of trigeminal neuralgia is that objective signs of sensory loss cannot be demonstrated on examination*.

Trigeminal neuralgia is relatively common, with an estimated annual incidence of 4.5 per 100,000 individuals. Middle-aged and elderly persons are affected primarily, and ~60% of cases occur in women. Onset is typically sudden, and bouts tend to persist for weeks or months before remitting spontaneously. Remissions may be long-lasting, but in most patients the disorder ultimately recurs.

PATHOPHYSIOLOGY

Symptoms result from ectopic generation of action potentials in pain-sensitive afferent fibers of the fifth cranial nerve root just before it enters the lateral surface of the pons. Compression or other pathology in the nerve leads to demyelination of large myelinated fibers that do not themselves carry pain sensation but become hyperexcitable and electrically coupled with smaller unmyelinated or poorly myelinated pain fibers in close proximity; this may explain why tactile stimuli, conveyed via the large myelinated fibers, can stimulate paroxysms of pain. Compression of the trigeminal nerve root by a blood vessel, most often the superior cerebellar artery or on occasion a tortuous vein, is the source of trigeminal neuralgia in a substantial proportion of patients. In cases of vascular compression, age-related brain sagging and increased vascular thickness and tortuosity may explain the prevalence of trigeminal neuralgia in later life.

DIFFERENTIAL DIAGNOSIS

Trigeminal neuralgia must be distinguished from other causes of face and head pain (Chap. 14) and from pain arising from diseases of the jaw, teeth, or sinuses. Pain from migraine or cluster headache tends to be deep-seated and steady, unlike the superficial stabbing quality of trigeminal neuralgia; rarely, cluster headache is associated with trigeminal neuralgia, a syndrome known as *cluster-tic*. In temporal arteritis, superficial facial pain is present but is not typically shocklike, the patient frequently complains of myalgias and other systemic symptoms, and an elevated erythrocyte sedimentation rate (ESR) is usually present (Chap. 326). When trigeminal neuralgia develops in a young adult or is bilateral, multiple sclerosis (MS) is a key consideration, and in such cases the cause is a demyelinating plaque at the root entry zone of the fifth nerve in the pons; often, evidence of facial sensory loss can be found on careful examination. Cases that are secondary to mass lesions—such as aneurysms, neurofibromas, acoustic schwannomas, or meningiomas—usually produce objective signs of sensory loss in the trigeminal nerve distribution (trigeminal neuropathy, see below).

LABORATORY EVALUATION

An ESR is indicated if temporal arteritis is suspected. In typical cases of trigeminal neuralgia, neuroimaging studies are usually unnecessary but may be valuable if MS is a consideration or in assessing overlying vascular lesions in order to plan for decompression surgery.

TREATMENT: TRIGEMINAL NEURALGIA

Drug therapy with carbamazepine is effective in ~50–75% of patients. Carbamazepine should be started as a single daily dose of 100 mg taken with food and increased gradually (by 100 mg daily in divided doses every 1–2 days) until substantial (>50%) pain relief is achieved. Most patients require a maintenance dose of 200 mg qid. Doses >1200 mg daily provide no additional benefit. Dizziness, imbalance, sedation, and rare cases of agranulocytosis are the most important side effects of carbamazepine. If treatment is effective, it is usually continued for 1 month and then tapered as tolerated. Oxcarbazepine (300–1200 mg bid) is an alternative to carbamazepine, has less bone marrow toxicity, and probably is equally efficacious. If these agents are not well tolerated or are ineffective, lamotrigine 400 mg daily or phenytoin, 300–400 mg daily, are other options. Baclofen may also be administered, either alone or in combination with an anticonvulsant. The initial dose is 5–10 mg tid, gradually increasing as needed to 20 mg qid.

If drug treatment fails, surgical therapy should be offered. The most widely used method currently is microvascular decompression to relieve pressure on the trigeminal nerve as it exits the pons. This procedure requires a suboccipital craniotomy. Based on limited data, this procedure appears to have a >70% efficacy rate and a low rate of pain recurrence in responders; the response is better for classic tic-like symptoms than for nonlancinating facial pains. In a small number of cases, there is perioperative damage to the eighth or seventh cranial nerves or to the cerebellum, or a postoperative CSF leak syndrome. High-resolution magnetic resonance angiography is useful preoperatively to visualize the relationships between the fifth cranial nerve root and nearby blood vessels.

Gamma knife radiosurgery is also utilized for treatment and results in complete pain relief in more than two-thirds of patients and a low risk of persistent facial numbness; the response is sometimes long-lasting, but recurrent pain develops over 2–3 years in half of patients. Compared with surgical decompression, gamma knife surgery appears to be somewhat less effective but has few serious complications.

Another procedure, *radiofrequency thermal rhizotomy*, creates a heat lesion of the trigeminal (gasserian) ganglion or nerve. It is used less often now than in the past. Short-term relief is

experienced by >95% of patients; however, long-term studies indicate that pain recurs in up to one-third of treated patients. Postoperatively, partial numbness of the face is common, masseter (jaw) weakness may occur especially following bilateral procedures, and corneal denervation with secondary keratitis can follow rhizotomy for first-division trigeminal neuralgia.

Trigeminal Neuropathy

A variety of diseases may affect the trigeminal nerve **(Table 376-1)**. Most present with sensory loss on the face or with weakness of the jaw muscles. Deviation of the jaw on opening indicates weakness of the pterygoids on the side to which the jaw deviates. Some cases are due to Sjögren's syndrome or a collagen-vascular disease such as systemic lupus erythematosus, scleroderma, or mixed connective tissue disease. Among infectious causes, herpes zoster and leprosy should be considered. Tumors of the middle cranial fossa (meningiomas), of the trigeminal nerve (schwannomas), or of the base of the skull (metastatic tumors) may cause a combination of motor and sensory signs. Lesions in the cavernous sinus can affect the first and second divisions of the trigeminal nerve, and lesions of the superior orbital fissure can affect the first (ophthalmic) division; the accompanying corneal anesthesia increases the risk of ulceration (neuro keratitis).

Table 376-1 Trigeminal Nerve Disorders

Nuclear (brainstem) lesions
Multiple sclerosis
Stroke
Syringobulbia
Glioma
Lymphoma
Preganglionic lesions
Acoustic neuroma
Meningioma
Metastasis
Chronic meningitis
Cavernous carotid aneurysm
Gasserian ganglion lesions
Trigeminal neuroma
Herpes zoster
Infection (spread from otitis media or mastoiditis)
Peripheral nerve lesions
Nasopharyngeal carcinoma
Trauma
Guillain-Barré syndrome
Sjögren's syndrome
Collagen-vascular diseases
Sarcoidosis
Leprosy
Drugs (stilbamidine, trichloroethylene)
Idiopathic trigeminal neuropathy

Loss of sensation over the chin (mental neuropathy) can be the only manifestation of systemic malignancy. Rarely, an idiopathic form of trigeminal neuropathy is observed. It is characterized by numbness and paresthesias, sometimes bilaterally, with loss of sensation in the territory of the trigeminal nerve but without weakness of the jaw. Gradual recovery is the rule. Tonic spasm of the masticatory muscles, known as *trismus*, is symptomatic of tetanus (Chap. 140) or may occur in patients treated with phenothiazine drugs.

FACIAL WEAKNESS

Anatomic Considerations

(Fig. 376-2) The seventh cranial nerve supplies all the muscles concerned with facial expression. The

sensory component is small (the nervus intermedius); it conveys taste sensation from the anterior two-thirds of the tongue and probably cutaneous impulses from the anterior wall of the external auditory canal. The motor nucleus of the seventh nerve lies anterior and lateral to the abducens nucleus. After leaving the pons, the seventh nerve enters the internal auditory meatus with the acoustic nerve. The nerve continues its course in its own bony channel, the facial canal, and exits from the skull via the stylomastoid foramen. It then passes through the parotid gland and subdivides to supply the facial muscles.

Figure 376-2

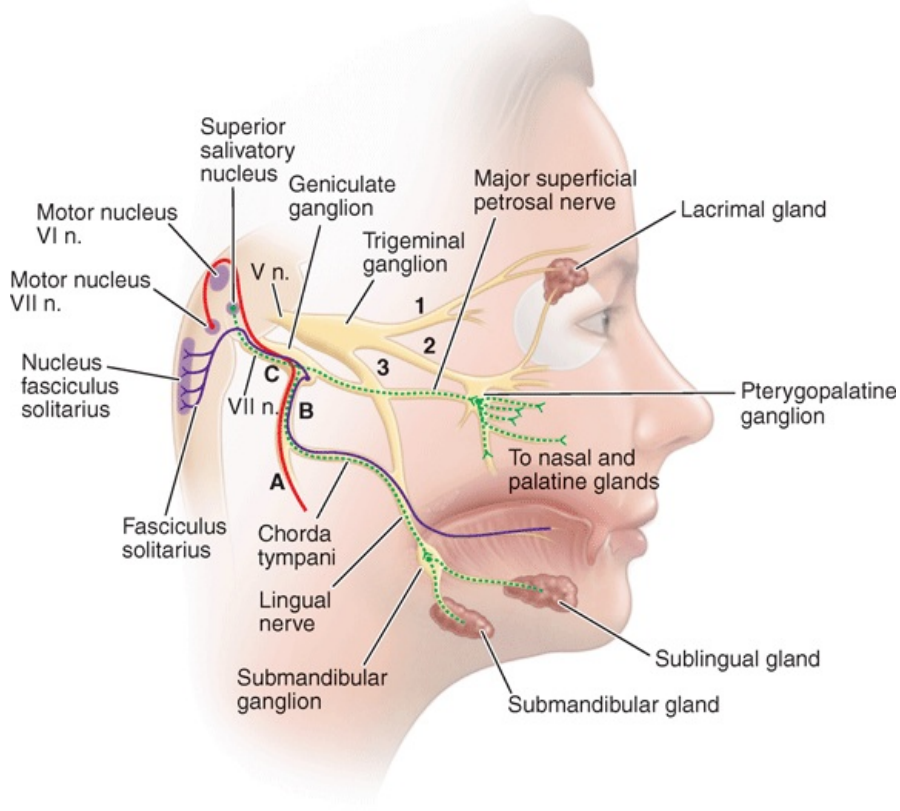

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The facial nerve. A, B, and C denote lesions of the facial nerve at the stylomastoid foramen, distal and proximal to the geniculate ganglion, respectively. Green lines indicate the parasympathetic fibers, red line indicates motor fibers, and purple lines indicate visceral afferent fibers (taste). (*Adapted from MB Carpenter: Core Text of Neuroanatomy, 2nd ed. Baltimore, Williams & Wilkins, 1978.*)

A complete interruption of the facial nerve at the stylomastoid foramen paralyzes all muscles of facial expression. The corner of the mouth droops, the creases and skinfolds are effaced, the forehead is unfurrowed, and the eyelids will not close. Upon attempted closure of the lids, the eye on the paralyzed side rolls upward (*Bell's phenomenon*). The lower lid sags and falls away from the conjunctiva, permitting tears to spill over the cheek. Food collects between the teeth and lips, and saliva may dribble from the corner of the mouth. The patient complains of a heaviness or numbness in the face, but sensory loss is rarely demonstrable and taste is intact.

If the lesion is in the middle-ear portion, taste is lost over the anterior two-thirds of the tongue on the same side. If the nerve to the stapedius is interrupted, there is hyperacusis (sensitivity to loud sounds). Lesions in the internal auditory meatus may affect the adjacent auditory and vestibular nerves, causing deafness, tinnitus, or dizziness. Intrapontine lesions that paralyze the face usually affect the abducens nucleus as well, and often the corticospinal and sensory tracts.

If the peripheral facial paralysis has existed for some time and recovery of motor function is incomplete, a continuous diffuse contraction of facial muscles may appear. The palpebral fissure becomes narrowed, and the nasolabial fold deepens. Attempts to move one group of facial muscles may result in contraction of all (associated movements, or *synkinesis*). Facial spasms, initiated by movements of the face, may develop (*hemifacial spasm*). Anomalous regeneration of seventh nerve fibers may result in other troublesome phenomena. If fibers originally connected with the orbicularis oculi come to innervate the orbicularis oris, closure of the lids may cause a retraction of the mouth, or if fibers originally connected with muscles of the face later innervate the lacrimal gland, anomalous tearing ("crocodile tears") may occur with any activity of the facial muscles, such as eating. Another

facial synkinesia is triggered by jaw opening, causing closure of the eyelids on the side of the facial palsy (jaw-winking).

Bell's Palsy

The most common form of facial paralysis is *Bell's palsy*. The annual incidence of this idiopathic disorder is ~25 per 100,000 annually, or about 1 in 60 persons in a lifetime.

CLINICAL MANIFESTATIONS

The onset of Bell's palsy is fairly abrupt, maximal weakness being attained by 48 h as a general rule. Pain behind the ear may precede the paralysis for a day or two. Taste sensation may be lost unilaterally, and hyperacusis may be present. In some cases there is mild cerebrospinal fluidlymphocytosis. MRI may reveal swelling and uniform enhancement of the geniculate ganglion and facial nerve and, in some cases, entrapment of the swollen nerve in the temporal bone. Approximately 80% of patients recover within a few weeks or months. Electromyography may be of some prognostic value; evidence of denervation after 10 days indicates there has been axonal degeneration, that there will be a long delay (3 months as a rule) before regeneration occurs, and that it may be incomplete. The presence of incomplete paralysis in the first week is the most favorable prognostic sign.

PATHOPHYSIOLOGY

In acute Bell's palsy there is inflammation of the facial nerve with mononuclear cells, consistent with an infectious or immune cause. Herpes simplex virus (HSV) type 1 DNA was frequently detected in endoneurial fluid and posterior auricular muscle, suggesting that a reactivation of this virus in the geniculate ganglion may be responsible for most cases. Reactivation of varicella zoster virus is associated with Bell's palsy in up to one-third of cases, and may represent the second most frequent cause. A variety of other viruses have also been implicated less commonly. An increased incidence of Bell's palsy was also reported among recipients of inactivated intranasal influenza vaccine, and it was hypothesized that this could have resulted from the *Escherichia coli* enterotoxin used as adjuvant or to reactivation of latent virus.

DIFFERENTIAL DIAGNOSIS

There are many other causes of acute facial palsy that must be considered in the differential diagnosis of Bell's palsy. Lyme disease can cause unilateral or bilateral facial palsies; in endemic areas, 10% or more of cases of facial palsy are likely due to infection with *Borrelia burgdorferi* (Chap. 173). The *Ramsay Hunt syndrome*, caused by reactivation of herpes zoster in the geniculate ganglion, consists of a severe facial palsy associated with a vesicular eruption in the external auditory canal and sometimes in the pharynx and other parts of the cranial integument; often the eighth cranial nerve is affected as well. Facial palsy that is often bilateral occurs in sarcoidosis (Chap. 329) and in *Guillain-Barré syndrome* (Chap. 385). Leprosy frequently involves the facial nerve, and facial neuropathy may also occur in diabetes mellitus, connective tissue diseases including Sjögren's syndrome, and amyloidosis. The rare *Melkersson-Rosenthal syndrome* consists of recurrent facial paralysis; recurrent—and eventually permanent—facial (particularly labial) edema; and, less constantly, plication of the tongue. Its cause is unknown. *Acoustic neuromas* frequently involve the facial nerve by local compression. Infarcts, demyelinating lesions of multiple sclerosis, and tumors are the common pontine lesions that interrupt the facial nerve fibers; other signs of brainstem involvement are usually present. Tumors that invade the temporal bone (carotid body, cholesteatoma, dermoid) may produce a facial palsy, but the onset is insidious and the course progressive.

All these forms of nuclear or peripheral facial palsy must be distinguished from the supranuclear type. In the latter, the frontalis and orbicularis oculi muscles are involved less than those of the lower part of the face, since the upper facial muscles are innervated by corticobulbar pathways from both motor cortices, whereas the lower facial muscles are innervated only by the opposite hemisphere. In supranuclear lesions there may be a dissociation of emotional and voluntary facial movements and often some degree of paralysis of the arm and leg, or an aphasia (in dominant hemisphere lesions) is present.

LABORATORY EVALUATION

The diagnosis of Bell's palsy can usually be made clinically in patients with (1) a typical presentation, (2) no risk factors or preexisting symptoms for other causes of facial paralysis, (3) absence of cutaneous lesions of herpes zoster in the external ear canal, and (4) a normal neurologic examination with the exception of the facial nerve. Particular attention to the eighth cranial nerve, which courses near to the facial nerve in the pontomedullary junction and in the temporal bone, and to other cranial nerves is essential. In atypical or uncertain cases, an ESR, testing for diabetes mellitus, a Lyme titer, angiotensin-converting enzyme and chest imaging studies for possible sarcoidosis, a lumbar puncture for possible Guillain-Barré syndrome, or MRI scanning may be indicated. MRI often shows swelling and enhancement of the facial nerve in idiopathic Bell's palsy **(Fig. 376-3)**.

Figure 376-3

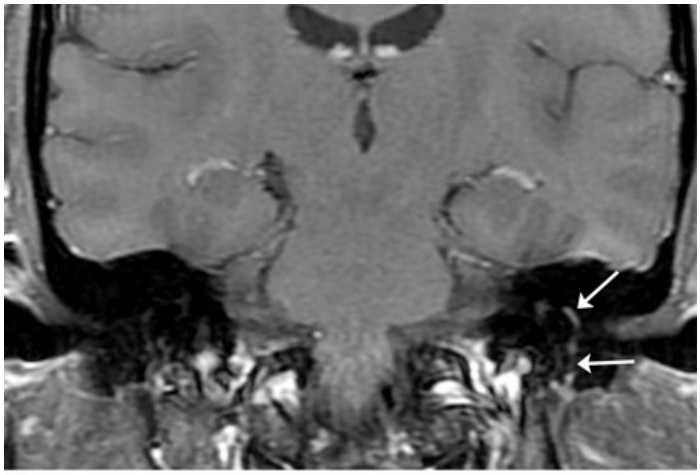

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

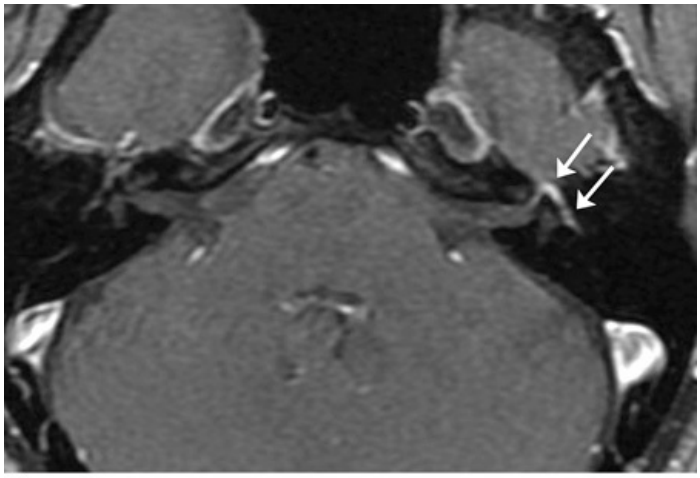

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial and coronal T1-weighted images post-Gadolinium with fat suppression demonstrate diffuse smooth linear enhancement of the left facial nerve, involving the genu, tympanic, and mastoid segments within the temporal bone (*arrows*), without evidence of mass lesion. Although highly suggestive of Bell's palsy, similar findings may be seen with other etiologies such as Lyme disease, sarcoidosis, and perineural malignant spread.

TREATMENT: BELL'S PALSY

Symptomatic measures include (1) the use of paper tape to depress the upper eyelid during sleep and prevent corneal drying, and (2) massage of the weakened muscles. A course of glucocorticoids, given as prednisone 60–80 mg daily during the first 5 days and then tapered over the next 5 days, modestly shortens the recovery period and improves the functional outcome. Although two large recently published randomized trials found no added benefit of antiviral agents valacyclovir (1000 mg daily for 5–7 days) or acyclovir (400 mg five times daily for 10 days) compared to glucocorticoids alone, the overall weight of evidence suggests that the combination therapy with prednisone plus valacyclovir may be marginally better than prednisone alone, especially in patients with severe clinical presentations.

Other Motor Disorders of the Face

Hemifacial spasm consists of painless irregular involuntary contractions on one side of the face. Most cases appear related to vascular compression of the exiting facial nerve in the pons. Other cases develop as a sequela to Bell's palsy or are secondary to compression and/or demyelination of the nerve by tumor, infection or multiple sclerosis. Mild cases can be treated with carbamazepine, gabapentin, or, if these drugs fail, with baclofen. Local injections of botulinum toxin into affected muscles can relieve spasms for 3–4 months, and the injections can be repeated. Refractory cases due to vascular compression usually respond to surgical decompression of the facial nerve.
Blepharospasm is an involuntary recurrent spasm of both eyelids that usually occurs in elderly persons as an isolated phenomenon or with varying degrees of spasm of other facial muscles. Severe, persistent cases of blepharospasm can be treated by local injection of botulinum toxin into the

orbicularis oculi. *Facial myokymia* refers to a fine rippling activity of the facial muscles; it may be caused by multiple sclerosis or follow Guillain-Barré syndrome (Chap. 385).

Facial hemiatrophy occurs mainly in women and is characterized by a disappearance of fat in the dermal and subcutaneous tissues on one side of the face. It usually begins in adolescence or early adult years and is slowly progressive. In its advanced form, the affected side of the face is gaunt, and the skin is thin, wrinkled, and brown. The facial hair may turn white and fall out, and the sebaceous glands become atrophic. Bilateral involvement may occur. A limited form of systemic sclerosis (scleroderma) may be the cause of some cases. Treatment is cosmetic, consisting of transplantation of skin and subcutaneous fat.

OTHER CRANIAL NERVE DISORDERS

Glossopharyngeal Neuralgia

This form of neuralgia involves the ninth (glossopharyngeal) and sometimes portions of the tenth (vagus) cranial nerves. It resembles trigeminal neuralgia in many respects but is much less common. The pain is intense and paroxysmal; it originates on one side of the throat, approximately in the tonsillar fossa. In some cases the pain is localized in the ear or may radiate from the throat to the ear because of involvement of the tympanic branch of the glossopharyngeal nerve. Spasms of pain may be initiated by swallowing or coughing. There is no demonstrable motor or sensory deficit; the glossopharyngeal nerve supplies taste sensation to the posterior third of the tongue and, together with the vagus nerve, sensation to the posterior pharynx. Cardiac symptoms—bradycardia or asystole, hypotension, and fainting—have been reported. Medical therapy is similar to that for trigeminal neuralgia, and carbamazepine is generally the first choice. If drug therapy is unsuccessful, surgical procedures—including microvascular decompression if vascular compression is evident—or rhizotomy of glossopharyngeal and vagal fibers in the jugular bulb is frequently successful.

Very rarely, herpes zoster involves the glossopharyngeal nerve. Glossopharyngeal neuropathy in conjunction with vagus and accessory nerve palsies may also occur with a tumor or aneurysm in the posterior fossa or in the jugular foramen. Hoarseness due to vocal cord paralysis, some difficulty in swallowing, deviation of the soft palate to the intact side, anesthesia of the posterior wall of the pharynx, and weakness of the upper part of the trapezius and sternocleidomastoid muscles make up the jugular foramen syndrome **(Table 376-2)**.

Table 376-2 Cranial Nerve Syndromes

Site	Cranial Nerves	Usual Cause
Sphenoid fissure (superior orbital)	III, IV, first division V, VI	Invasive tumors of sphenoid bone; aneurysms
Lateral wall of cavernous sinus	III, IV, first division V, VI, often with proptosis	Infection, thrombosis, aneurysm, or fistula of cavernous sinus; invasive tumors from sinuses and sella turcica; benign granuloma responsive to glucocorticoids
Retrosphenoid space	II, III, IV, V, VI	Large tumors of middle cranial fossa
Apex of petrous bone	V, VI	Petrositis; tumors of petrous bone
Internal auditory meatus	VII, VIII	Tumors of petrous bone (dermoids, etc.); infectious processes; acoustic neuroma
Pontocerebellar angle	V, VII, VIII, and sometimes IX	Acoustic neuroma; meningioma
Jugular foramen	IX, X, XI	Tumors and aneurysms
Posterior laterocondylar space	IX, X, XI, XII	Tumors of parotid gland and carotid body and metastatic tumors
Posterior retroparotid space	IX, X, XI, XII and Horner syndrome	Tumors of parotid gland, carotid body, lymph nodes; metastatic tumor; tuberculous adenitis

Dysphagia and Dysphonia

When the intracranial portion of one vagus (tenth cranial) nerve is interrupted, the soft palate droops ipsilaterally and does not rise in phonation. There is loss of the gag reflex on the affected side, as well as of the "curtain movement" of the lateral wall of the pharynx, whereby the faucial pillars move medially as the palate rises in saying "ah." The voice is hoarse and slightly nasal, and the vocal cord

lies immobile midway between abduction and adduction. Loss of sensation at the external auditory meatus and the posterior pinna may also be present.

The pharyngeal branches of both vagal nerves may be affected in diphtheria; the voice has a nasal quality, and regurgitation of liquids through the nose occurs during the act of swallowing.

The vagus nerve may be involved at the meningeal level by neoplastic and infectious processes and within the medulla by tumors, vascular lesions (e.g., the lateral medullary syndrome), and motor neuron disease. This nerve may be involved by infection with varicella zoster virus. Polymyositis and dermatomyositis, which cause hoarseness and dysphagia by direct involvement of laryngeal and pharyngeal muscles, may be confused with diseases of the vagus nerves. Dysphagia is also a symptom in some patients with myotonic dystrophy. **Nonneurologic causes of dysphagia are discussed in Chap. 38.**

The recurrent laryngeal nerves, especially the left, are most often damaged as a result of intrathoracic disease. Aneurysm of the aortic arch, an enlarged left atrium, and tumors of the mediastinum and bronchi are much more frequent causes of an isolated vocal cord palsy than are intracranial disorders. However, a substantial number of cases of recurrent laryngeal palsy remain idiopathic.

When confronted with a case of laryngeal palsy, the physician must attempt to determine the site of the lesion. If it is intramedullary, there are usually other signs, such as ipsilateral cerebellar dysfunction, loss of pain and temperature sensation over the ipsilateral face and contralateral arm and leg, and an ipsilateral Horner syndrome. If the lesion is extramedullary, the glossopharyngeal and spinal accessory nerves are frequently involved (jugular foramen syndrome). If it is extracranial in the posterior laterocondylar or retroparotid space, there may be a combination of ninth, tenth, eleventh, and twelfth cranial nerve palsies and a Horner syndrome (Table 376-2). If there is no sensory loss over the palate and pharynx and no palatal weakness or dysphagia, the lesion is below the origin of the pharyngeal branches, which leave the vagus nerve high in the cervical region; the usual site of disease is then the mediastinum.

Neck Weakness

Isolated involvement of the accessory (eleventh cranial) nerve can occur anywhere along its route, resulting in partial or complete paralysis of the sternocleidomastoid and trapezius muscles. More commonly, involvement occurs in combination with deficits of the ninth and tenth cranial nerves in the jugular foramen or after exit from the skull (Table 376-2). An idiopathic form of accessory neuropathy, akin to Bell's palsy, has been described, and it may be recurrent in some cases. Most but not all patients recover.

Tongue Paralysis

The hypoglossal (twelfth cranial) nerve supplies the ipsilateral muscles of the tongue. The nucleus of the nerve or its fibers of exit may be involved by intramedullary lesions such as tumor, poliomyelitis, or most often motor neuron disease. Lesions of the basal meninges and the occipital bones (platybasia, invagination of occipital condyles, Paget's disease) may compress the nerve in its extramedullary course or in the hypoglossal canal. Isolated lesions of unknown cause can occur. Atrophy and fasciculation of the tongue develop weeks to months after interruption of the nerve.

MULTIPLE CRANIAL NERVE PALSIES

Several cranial nerves may be affected by the same disease process. In this situation, the main clinical problem is to determine whether the lesion lies within the brainstem or outside it. Lesions that lie on the surface of the brainstem are characterized by involvement of adjacent cranial nerves (often occurring in succession) and late and rather slight involvement of the long sensory and motor pathways and segmental structures lying within the brainstem. The opposite is true of primary lesions within the brainstem. The extramedullary lesion is more likely to cause bone erosion or enlargement of the foramens of exit of cranial nerves. The intramedullary lesion involving cranial nerves often produces a crossed sensory or motor paralysis (cranial nerve signs on one side of the body and tract signs on the opposite side).

Involvement of multiple cranial nerves outside the brainstem is frequently the result of trauma, localized infections including varicella zoster virus, infectious and noninfectious (especially carcinomatous) causes of meningitis (Chaps. 381 and 382), granulomatous diseases such as granulomatosis with polyangiitis (Wegener's), Behçet's disease, vascular disorders including those associated with diabetes, enlarging saccular aneurysms, or locally infiltrating tumors. Among the tumors, nasopharyngeal cancers, lymphomas, neurofibromas, meningiomas, chordomas, cholesteatomas, carcinomas, and sarcomas have all been observed to involve a succession of lower cranial nerves. Owing to their anatomic relationships, the multiple cranial nerve palsies form a number of distinctive syndromes, listed in Table 376-2. Sarcoidosis is the cause of some cases of multiple cranial neuropathy, and chronic glandular tuberculosis the cause of a few others. Platybasia, basilar invagination of the skull, and the Chiari malformation are additional causes. A purely motor disorder without atrophy always raises the question of myasthenia gravis (Chap. 386). As noted above, Guillain-Barré syndrome commonly affects the facial nerves bilaterally. In the Fisher variant of the

Guillain-Barré syndrome, oculomotor paresis occurs with ataxia and areflexia in the limbs (Chap. 385). Wernicke encephalopathy can cause a severe ophthalmoplegia combined with other brainstem signs (Chap. 275).

The *cavernous sinus syndrome* **(Fig. 376-4)** is a distinctive and frequently life-threatening disorder. It often presents as orbital or facial pain; orbital swelling and chemosis due to occlusion of the ophthalmic veins; fever; oculomotor neuropathy affecting the third, fourth, and sixth cranial nerves; and trigeminal neuropathy affecting the ophthalmic (V_1) and occasionally the maxillary (V_2) divisions of the trigeminal nerve. Cavernous sinus thrombosis, often secondary to infection from orbital cellulitis (frequently *Staphylococcus aureus*), a cutaneous source on the face, or sinusitis (especially with mucormycosis in diabetic patients), is the most frequent cause; other etiologies include aneurysm of the carotid artery, a carotid-cavernous fistula (orbital bruit may be present), meningioma, nasopharyngeal carcinoma, other tumors, or an idiopathic granulomatous disorder (Tolosa-Hunt syndrome). The two cavernous sinuses directly communicate via intercavernous channels; thus, involvement on one side may extend to become bilateral. Early diagnosis is essential, especially when due to infection, and treatment depends on the underlying etiology.

Figure 376-4

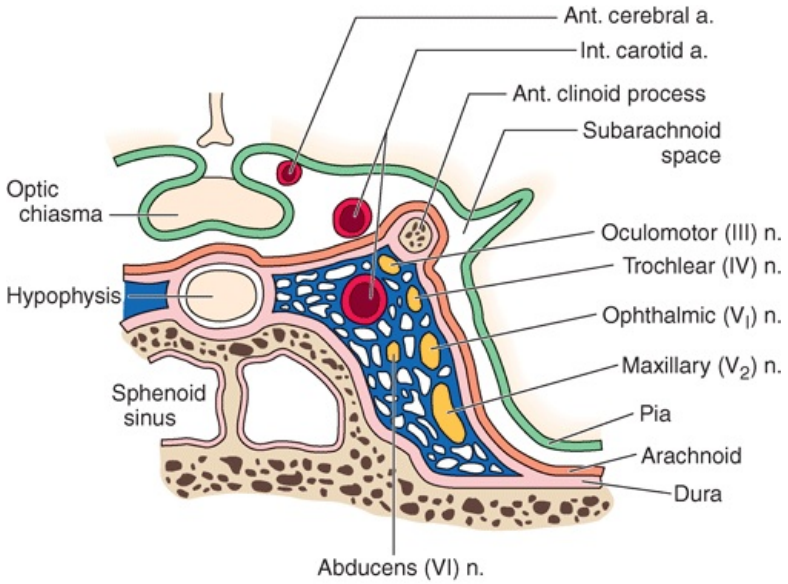

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anatomy of the cavernous sinus in coronal section, illustrating the location of the cranial nerves in relation to the vascular sinus, internal carotid artery (which loops anteriorly to the section), and surrounding structures.

In infectious cases, prompt administration of broad-spectrum antibiotics, drainage of any abscess cavities, and identification of the offending organism are essential. Anticoagulant therapy may benefit cases of primary thrombosis. Repair or occlusion of the carotid artery may be required for treatment of fistulas or aneurysms. The Tolosa-Hunt syndrome generally responds to glucocorticoids. A dramatic improvement in pain is usually evident within a few days; oral prednisone (60 mg daily) is usually continued for 2 weeks and then gradually tapered over a month, or longer if pain recurs.

An idiopathic form of multiple cranial nerve involvement on one or both sides of the face is occasionally seen. The syndrome consists of a subacute onset of boring facial pain, followed by paralysis of motor cranial nerves. The clinical features overlap those of the Tolosa-Hunt syndrome and appear to be due to idiopathic inflammation of the dura mater, which may be visualized by MRI. The syndrome is frequently responsive to glucocorticoids.

FURTHER READINGS

De Almeida JR et al: Combined corticosteroids and antiviral treatment for Bell palsy: A systematic review and meta-analysis. JAMA 302:985, 2009

Dhople AA et al: Long term outcomes of gamma knife radiosurgery for classic trigeminal neuralgia: Implications of treatment and critical review of the literature. J Neurosurg 111: 351, 2009[PMID: 19326987] [Full Text]

Gronseth G et al: Practice parameter: The diagnostic evaluation and treatment of trigeminal neuralgia (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology and the European Federation of Neurological Societies. Neurology 71:1183, 2008[PMID: 18716236] [Full Text]

Hato N et al: Valacyclovir and prednisolone treatment for Bell's palsy: A randomized, placebo-controlled study. Otol Neurotol 28:408, 2007[PMID: 17414047] [Full Text]

Keane JR: Multiple cranial nerve palsies: Analysis of 979 cases. Arch Neurol. 62: 1714, 2005[PMID: 16286545] [Full Text]

Pearce JMS: Glossopharyngeal neuralgia. Eur Neurol 55:49, 2006[PMID: 16479122] [Full Text]

Quant EC et al: The benefits of steroids versus steroids plus antivirals for treatment of Bell's palsy: A meta-analysis. BMJ 339:b3354, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISEASES OF THE SPINAL CORD: INTRODUCTION

Diseases of the spinal cord are frequently devastating. They produce quadriplegia, paraplegia, and sensory deficits far beyond the damage they would inflict elsewhere in the nervous system because the spinal cord contains, in a small cross-sectional area, almost the entire motor output and sensory input of the trunk and limbs. Many spinal cord diseases are reversible if recognized and treated at an early stage **(Table 377-1)**; thus, they are among the most critical of neurologic emergencies. The efficient use of diagnostic procedures, guided by knowledge of the anatomy and the clinical features of spinal cord diseases, is required for a successful outcome.

Table 377-1 Treatable Spinal Cord Disorders

Compressive
Epidural, intradural, or intramedullary neoplasm
Epidural abscess
Epidural hemorrhage
Cervical spondylosis
Herniated disk
Posttraumatic compression by fractured or displaced vertebra or hemorrhage
Vascular
Arteriovenous malformation
Antiphospholipid syndrome and other hypercoagulable states
Inflammatory
Multiple sclerosis
Neuromyelitis optica
Transverse myelitis
Sarcoidosis
Sjögren-related myelopathy
Systemic lupus erythematosus
Vasculitis
Infectious
Viral: VZV, HSV-1 and -2, CMV, HIV, HTLV-I, others
Bacterial and mycobacterial: *Borrelia, Listeria,* syphilis, others
Mycoplasma pneumoniae
Parasitic: schistosomiasis, toxoplasmosis
Developmental
Syringomyelia
Meningomyelocele
Tethered cord syndrome
Metabolic
Vitamin B12 deficiency (subacute combined degeneration)
Copper deficiency

Abbreviations: CMV, cytomegalovirus; HSV, herpes simplex virus; HTLV, human T cell lymphotropic virus; VZV, varicella-zoster virus.

Approach to the Patient: Spinal Cord Disease

SPINAL CORD ANATOMY RELEVANT TO CLINICAL SIGNS

The spinal cord is a thin, tubular extension of the central nervous system contained within the bony spinal canal. It originates at the medulla and continues caudally to the conus medullaris at the lumbar level; its fibrous extension, the filum terminale, terminates at the coccyx. The adult spinal cord is ~46 cm (18 in.) long, oval in shape, and enlarged in the cervical and lumbar regions, where neurons that innervate the upper and lower extremities, respectively, are located. The white matter tracts containing ascending sensory and descending motor pathways are located peripherally, whereas nerve cell bodies are clustered in an inner region shaped like a four-leaf clover that surrounds the central canal (anatomically an extension of the fourth ventricle). The membranes that cover the spinal cord—the pia, arachnoid, and dura—are continuous with those of the brain.

The spinal cord has 31 segments, each defined by a pair of exiting ventral motor roots and entering dorsal sensory roots. During embryologic development, growth of the cord lags behind that of the vertebral column, and the mature spinal cord ends at approximately the first lumbar vertebral body. The lower spinal nerves take an increasingly downward course to exit via intervertebral foramens. The first seven pairs of cervical spinal nerves exit above the same-numbered vertebral bodies, whereas all the subsequent nerves exit below the same-numbered vertebral bodies because of the presence of eight cervical spinal cord segments but only seven cervical vertebrae. The relationship between spinal cord segments and the corresponding vertebral bodies is shown in **Table 377-2**. These relationships assume particular importance for localization of lesions that cause spinal cord compression. Sensory loss below the circumferential level of the umbilicus, for example, corresponds to the T10 cord segment but indicates involvement of the cord adjacent to the seventh or eighth thoracic vertebral body (Figs. 23-2 and 23-3). In addition, at every level the main ascending and descending tracts are somatotopically organized with a laminated distribution that reflects the origin or destination of nerve fibers.

Table 377-2 Spinal Cord Levels Relative to the Vertebral Bodies

Spinal Cord Level	Corresponding Vertebral Body
Upper cervical	Same as cord level
Lower cervical	1 level higher
Upper thoracic	2 levels higher
Lower thoracic	2 to 3 levels higher
Lumbar	T10-T12
Sacral	T12-L1

Determining the Level of the Lesion

The presence of a horizontally defined level below which sensory, motor, and autonomic function is impaired is a hallmark of spinal cord disease. This *sensory level* is sought by asking the patient to identify a pinprick or cold stimulus applied to the proximal legs and lower trunk and successively moved up toward the neck on each side. Sensory loss below this level is the result of damage to the spinothalamic tract on the opposite side one to two segments higher in the case of a unilateral spinal cord lesion, and at the level of a bilateral lesion. The discrepancy in the level of a unilateral lesion is the result of the course of the second-order sensory fibers, which originate in the dorsal horn, and ascend for one or two levels as they cross anterior to the central canal to join the opposite spinothalamic tract. Lesions that transect the descending corticospinal and other motor tracts cause paraplegia or quadriplegia with heightened deep tendon reflexes, Babinski signs, and eventual spasticity (the upper motor neuron syndrome). Transverse damage to the cord also produces autonomic disturbances consisting of absent sweating below the implicated cord level and bladder, bowel, and sexual dysfunction.

The uppermost level of a spinal cord lesion can also be localized by attention to the *segmental signs* corresponding to disturbed motor or sensory innervation by an individual cord segment. A band of altered sensation (hyperalgesia or hyperpathia) at the upper end of the sensory disturbance, fasciculations or atrophy in muscles innervated by one or several segments, or a muted or absent deep tendon reflex may be noted at this level. These signs also can occur with focal root or peripheral nerve disorders; thus, they are most useful when they occur together with signs of long tract damage. With severe and acute transverse lesions, the limbs initially may be flaccid rather than spastic. This state of "spinal shock" lasts for several days, rarely for weeks, and should not be mistaken for extensive damage to the anterior horn cells over many segments of the cord or for an acute polyneuropathy.

The main features of transverse damage at each level of the spinal cord are summarized below.

Cervical Cord

Upper cervical cord lesions produce quadriplegia and weakness of the diaphragm. Lesions at C4-C5

produce quadriplegia; at C5-C6, there is loss of power and reflexes in the biceps; at C7 weakness affects finger and wrist extensors and triceps; and at C8, finger and wrist flexion are impaired. Horner's syndrome (miosis, ptosis, and facial hypohidrosis) may accompany a cervical cord lesion at any level.

Thoracic Cord

Lesions here are localized by the sensory level on the trunk and by the site of midline back pain that may accompany the syndrome. Useful markers for localization are the nipples (T4) and umbilicus (T10). Leg weakness and disturbances of bladder and bowel function accompany the paralysis. Lesions at T9-T10 paralyze the lower—but not the upper—abdominal muscles, resulting in upward movement of the umbilicus when the abdominal wall contracts (*Beevor's sign*).

Lumbar Cord

Lesions at the L2-L4 spinal cord levels paralyze flexion and adduction of the thigh, weaken leg extension at the knee, and abolish the patellar reflex. Lesions at L5-S1 paralyze only movements of the foot and ankle, flexion at the knee, and extension of the thigh, and abolish the ankle jerks (S1).

Sacral Cord/Conus Medullaris

The conus medullaris is the tapered caudal termination of the spinal cord, comprising the lower sacral and single coccygeal segments. The distinctive conus syndrome consists of bilateral saddle anesthesia (S3-S5), prominent bladder and bowel dysfunction (urinary retention and incontinence with lax anal tone), and impotence. The bulbocavernosus (S2-S4) and anal (S4-S5) reflexes are absent (Chap. 367). Muscle strength is largely preserved. By contrast, lesions of the cauda equina, the nerve roots derived from the lower cord, are characterized by low back and radicular pain, asymmetric leg weakness and sensory loss, variable areflexia in the lower extremities, and relative sparing of bowel and bladder function. Mass lesions in the lower spinal canal often produce a mixed clinical picture with elements of both cauda equina and conus medullaris syndromes. Cauda equina syndromes are also discussed in Chap. 15.

Special Patterns of Spinal Cord Disease

The location of the major ascending and descending pathways of the spinal cord are shown in **Fig. 377-1**. Most fiber tracts—including the posterior columns and the spinocerebellar and pyramidal tracts—are situated on the side of the body they innervate. However, afferent fibers mediating pain and temperature sensation ascend in the spinothalamic tract contralateral to the side they supply. The anatomic configurations of these tracts produce characteristic syndromes that provide clues to the underlying disease process.

Figure 377-1

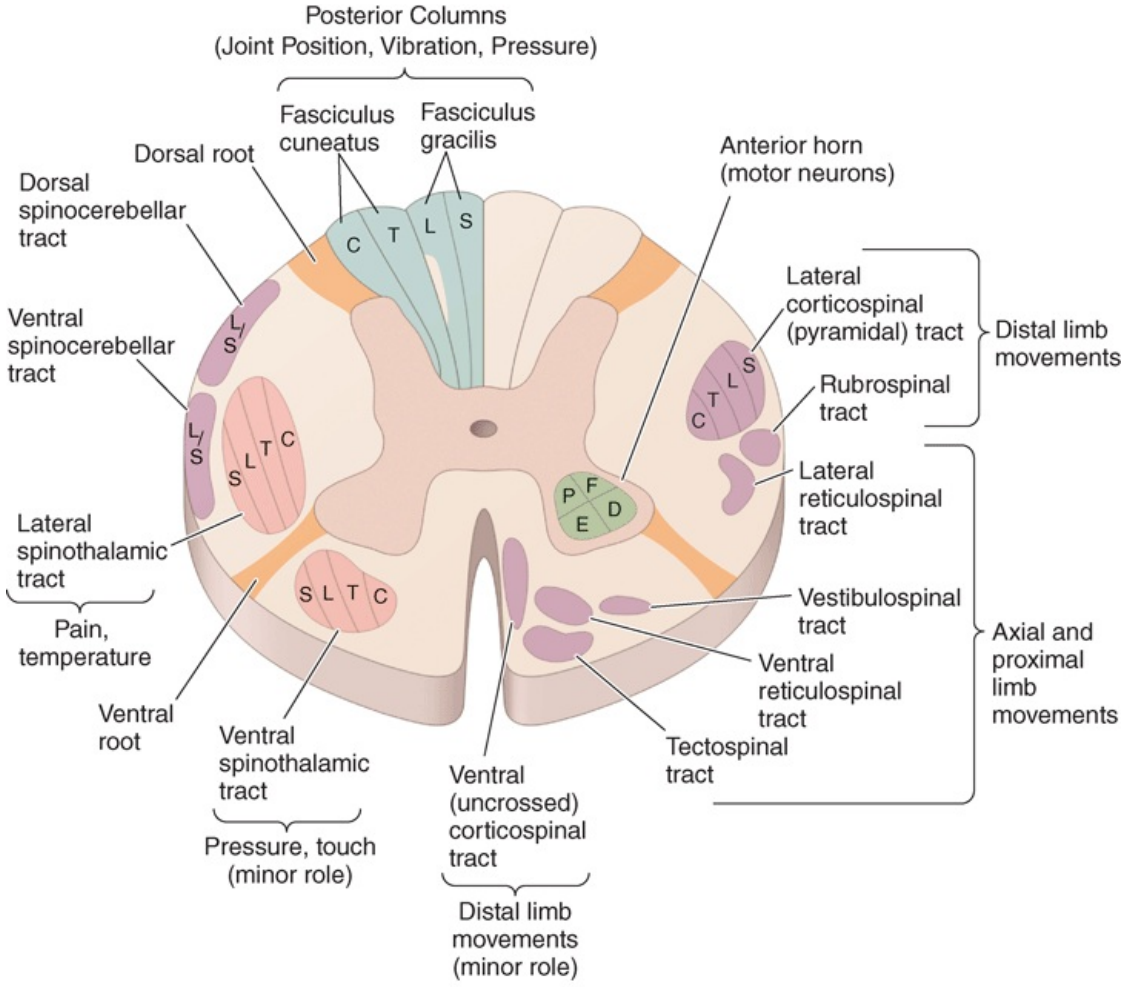

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Transverse section through the spinal cord, composite representation, illustrating the principal ascending (*left*) and descending (*right*) pathways. The lateral and ventral spinothalamic tracts (*dark blue*) ascend contralateral to the side of the body that is innervated. C, cervical; D, distal; E, extensors; F, flexors; L, lumbar; P, proximal; S, sacral; T, thoracic.

Brown-Sequard Hemicord Syndrome

This consists of ipsilateral weakness (corticospinal tract) and loss of joint position and vibratory sense (posterior column), with contralateral loss of pain and temperature sense (spinothalamic tract) one or two levels below the lesion. Segmental signs, such as radicular pain, muscle atrophy, or loss of a deep tendon reflex, are unilateral. Partial forms are more common than the fully developed syndrome.

Central Cord Syndrome

This syndrome results from selective damage to the gray matter nerve cells and crossing spinothalamic tracts surrounding the central canal. In the cervical cord, the central cord syndrome produces arm weakness out of proportion to leg weakness and a "dissociated" sensory loss, meaning loss of pain and temperature sensations over the shoulders, lower neck, and upper trunk (cape distribution), in contrast to preservation of light touch, joint position, and vibration sense in these regions. Spinal trauma, syringomyelia, and intrinsic cord tumors are the main causes.

Anterior Spinal Artery Syndrome

Infarction of the cord is generally the result of occlusion or diminished flow in this artery. The result is extensive bilateral tissue destruction that spares the posterior columns. All spinal cord functions —motor, sensory, and autonomic—are lost below the level of the lesion, with the striking exception of retained vibration and position sensation.

Foramen Magnum Syndrome

Lesions in this area interrupt decussating pyramidal tract fibers destined for the legs, which cross caudal to those of the arms, resulting in weakness of the legs (*crural paresis*). Compressive lesions near the foramen magnum may produce weakness of the ipsilateral shoulder and arm followed by weakness of the ipsilateral leg, then the contralateral leg, and finally the contralateral arm, an "around-the-clock" pattern that may begin in any of the four limbs. There is typically suboccipital pain spreading to the neck and shoulders.

Intramedullary and Extramedullary Syndromes

It is useful to differentiate *intramedullary* processes, arising within the substance of the cord, from *extramedullary* ones that compress the spinal cord or its vascular supply. The differentiating features are only relative and serve as clinical guides. With extramedullary lesions, radicular pain is often prominent, and there is early sacral sensory loss (lateral spinothalamic tract) and spastic weakness in the legs (corticospinal tract) due to the superficial location of leg fibers in the corticospinal tract. Intramedullary lesions tend to produce poorly localized burning pain rather than radicular pain and to spare sensation in the perineal and sacral areas ("sacral sparing"), reflecting the laminated configuration of the spinothalamic tract with sacral fibers outermost; corticospinal tract signs appear later. Regarding extramedullary lesions, a further distinction is made between extradural and intradural masses, as the former are generally malignant and the latter benign (neurofibroma being a common cause). Consequently, a long duration of symptoms favors an intradural origin.

ACUTE AND SUBACUTE SPINAL CORD DISEASES

The initial symptoms of disease that evolve over days or weeks are focal neck or back pain, followed by various combinations of paresthesias, sensory loss, motor weakness, and sphincter disturbance evolving over hours to several days. There may be only mild sensory symptoms or a devastating functional transection of the cord. Partial lesions selectively involve the posterior columns or anterior spinothalamic tracts or are limited to one side of the cord. Paresthesias or numbness typically begins in the feet and ascends symmetrically or asymmetrically. These symptoms initially simulate Guillain-Barré syndrome, but involvement of the trunk with a sharply demarcated spinal cord level indicates the myelopathic nature of the process. In severe and abrupt cases, areflexia reflecting spinal shock may be present, but hyperreflexia supervenes over days or weeks; persistent areflexic paralysis with a sensory level indicates necrosis over multiple segments of the spinal cord.

Approach the the Patient: Compressive and Noncompressive Myelopathy

DISTINGUISHING COMPRESSIVE FROM NONCOMPRESSIVE MYELOPATHY

The first priority is to exclude a treatable compression of the cord by a mass. The common causes are tumor, epidural abscess or hematoma, herniated disk, or vertebral pathology. Epidural compression due to malignancy or abscess often causes warning signs of neck or back pain, bladder disturbances, and sensory symptoms that precede the development of paralysis. Spinal subluxation, hemorrhage, and noncompressive etiologies such as infarction are more likely to produce myelopathy without antecedent symptoms. MRI with gadolinium infusion, centered on the clinically suspected level, is the initial diagnostic procedure; in some cases it is appropriate to image the entire spine (cervical through sacral regions) to search for additional clinically silent lesions. Once compressive lesions have been excluded, noncompressive causes of acute myelopathy that are intrinsic to the cord are considered, primarily vascular, inflammatory, and infectious etiologies.

Compressive Myelopathies

NEOPLASTIC SPINAL CORD COMPRESSION

In adults, most neoplasms are epidural in origin, resulting from metastases to the adjacent spinal bones. The propensity of solid tumors to metastasize to the vertebral column probably reflects the high proportion of bone marrow located in the axial skeleton. Almost any malignant tumor can metastasize to the spinal column, with breast, lung, prostate, kidney, lymphoma, and plasma cell dyscrasia being particularly frequent. The thoracic spinal column is most commonly involved; exceptions are metastases from prostate and ovarian cancer, which occur disproportionately in the sacral and lumbar vertebrae, probably resulting from spread through Batson's plexus, a network of veins along the anterior epidural space. Retroperitoneal neoplasms (especially lymphomas or sarcomas) enter the spinal canal through the intervertebral foramens and produce radicular pain with signs of root weakness prior to cord compression.

Pain is usually the initial symptom of spinal metastasis; it may be aching and localized or sharp and radiating in quality and typically worsens with movement, coughing, or sneezing and characteristically awakens patients at night. A recent onset of persistent back pain, particularly if in the thoracic spine (which is uncommonly involved by spondylosis), should prompt consideration of vertebral metastasis. Rarely, pain is mild or absent. Plain radiographs of the spine and radionuclide bone scans have only a limited role in diagnosis because they do not identify 15–20% of metastatic vertebral lesions and fail to detect paravertebral masses that reach the epidural space through the intervertebral foramens. MRI provides excellent anatomic resolution of the extent of spinal tumors **(Fig. 377-2)** and is able to distinguish between malignant lesions and other masses—epidural abscess, tuberculoma, or epidural hemorrhage, among others—that present in a similar fashion. Vertebral metastases are usually hypointense relative to a normal bone marrow signal on T1-weighted MRI scans; after the administration of gadolinium, contrast enhancement may deceptively "normalize" the appearance of the tumor by increasing its intensity to that of normal bone marrow. Infections of the spinal column (osteomyelitis and related disorders) are distinctive in that, unlike tumor, they may cross the disk space to involve the adjacent vertebral body.

Figure 377-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Epidural spinal cord compression due to breast carcinoma. Sagittal T1-weighted (***A***) and T2-weighted (***B***) MRI scans through the cervicothoracic junction reveal an infiltrated and collapsed second thoracic vertebral body with posterior displacement and compression of the upper thoracic spinal cord. The low-intensity bone marrow signal in *A* signifies replacement by tumor.

If spinal cord compression is suspected, imaging should be obtained promptly. If there are radicular symptoms but no evidence of myelopathy, it is usually safe to defer imaging for 24–48 h. Up to 40% of patients who present with cord compression at one level are found to have asymptomatic epidural metastases elsewhere; thus, the length of the spine should be imaged when epidural malignancy is in question.

Treatment: Neoplastic Spinal Cord Compression

Management of cord compression includes glucocorticoids to reduce cord edema, local radiotherapy (initiated as early as possible) to the symptomatic lesion, and specific therapy for the underlying tumor type. Glucocorticoids (dexamethasone, up to 40 mg daily) can be administered before the imaging study if the clinical suspicion is strong and continued at a lower dose until radiotherapy (generally 3000 cGy administered in 15 daily fractions) is completed. Radiotherapy appears to be effective even for most classically radioresistant metastases. A good response to radiotherapy can be expected in individuals who are ambulatory at presentation. Treatment usually prevents new weakness, and some recovery of motor function occurs in up to one-third of treated patients. Motor deficits (paraplegia or quadriplegia), once established for >12 h, do not usually improve, and beyond 48 h the prognosis for substantial motor recovery is poor. Although most patients do not experience recurrences in the months following radiotherapy, with survival beyond 2 years, recurrence becomes increasingly likely and can be managed with additional radiotherapy. New techniques, including intensity-modulated radiotherapy (IMRT), can deliver high doses of focused radiation with extreme precision, and preliminary data suggest that these methods produce similar rates of response compared to traditional radiotherapy. Biopsy of the epidural mass is unnecessary in patients with known primary cancer, but it is indicated if a history of underlying cancer is lacking. Surgery, either decompression by laminectomy or vertebral body resection, is usually considered when signs of cord compression worsen despite radiotherapy, when the maximum tolerated dose of radiotherapy has been delivered previously to the site, or when a vertebral compression fracture or spinal instability contributes to cord compression. The routine use of radiotherapy as first-line treatment for most cases of malignant spinal cord compression has recently been called into question by a randomized clinical trial indicating that surgery followed by radiotherapy is more effective than radiotherapy alone for patients with a single area of spinal cord compression by extradural tumor; patients with recurrent cord compression, brain metastases, radiosensitive tumors, or severe motor symptoms of >48 hours duration were excluded from this study.

In contrast to tumors of the epidural space, most intradural mass lesions are slow-growing and benign. Meningiomas and neurofibromas account for most of these, with occasional cases caused by chordoma, lipoma, dermoid, or sarcoma. Meningiomas **(Fig. 377-3)** are often located posterior to the thoracic cord or near the foramen magnum, although they can arise from the meninges anywhere along the spinal canal. Neurofibromas are benign tumors of the nerve sheath that typically arise near the posterior root; when multiple, neurofibromatosis is the likely etiology. Symptoms usually begin with radicular sensory symptoms followed by an asymmetric, progressive spinal cord syndrome. Therapy is by surgical resection.

Figure 377-3

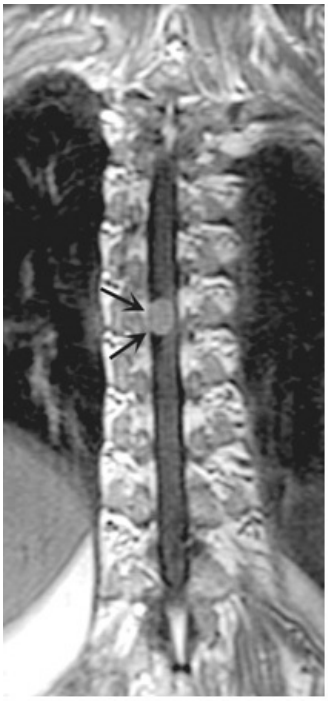

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI of a thoracic meningioma. Coronal T1-weighted postcontrast image through the thoracic spinal cord demonstrates intense and uniform enhancement of a well-circumscribed extramedullary mass (*arrows*), which displaces the spinal cord to the left.

Primary intramedullary tumors of the spinal cord are uncommon. They present as central cord or hemicord syndromes, often in the cervical region; there may be poorly localized burning pain in the extremities and sparing of sacral sensation. In adults, these lesions are ependymomas, hemangioblastomas, or low-grade astrocytomas **(Fig. 377-4)**. Complete resection of an intramedullary ependymoma is often possible with microsurgical techniques. Debulking of an intramedullary astrocytoma can also be helpful, as these are often slowly growing lesions; the value of adjunctive radiotherapy and chemotherapy is uncertain. Secondary (metastatic) intramedullary tumors also occur, especially in patients with advanced metastatic disease (Chap. 379), although these are not nearly as frequent as brain metastases.

Figure 377-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI of an intramedullary astrocytoma. Sagittal T1-weighted postcontrast image through the cervical spine demonstrates expansion of the upper cervical spine by a mass lesion emanating from within the spinal cord at the cervicomedullary junction. Irregular peripheral enhancement occurs within the mass (*arrows*).

SPINAL EPIDURAL ABSCESS

Spinal epidural abscess presents as a clinical triad of midline dorsal pain, fever, and progressive limb weakness. Prompt recognition of this distinctive process will in most cases prevent permanent sequelae. Aching pain is almost always present, either over the spine or in a radicular pattern. The duration of pain prior to presentation is generally ≤2 weeks but may on occasion be several months or longer. Fever is usual, accompanied by elevated white blood cell count, sedimentation rate, and C-reactive protein. As the abscess expands, further spinal cord damage results from venous congestion and thrombosis. Once weakness and other signs of myelopathy appear, progression may be rapid. A more chronic sterile granulomatous form of abscess is also known, usually after treatment of an acute epidural infection.

Risk factors include an impaired immune status (diabetes mellitus, renal failure, alcoholism, malignancy), intravenous drug abuse, and infections of the skin or other tissues. Two-thirds of epidural infections result from hematogenous spread of bacteria from the skin (furunculosis), soft tissue (pharyngeal or dental abscesses), or deep viscera (bacterial endocarditis). The remainder arise from direct extension of a local infection to the subdural space; examples of local predisposing conditions are vertebral osteomyelitis, decubitus ulcers, lumbar puncture, epidural anesthesia, or spinal surgery. Most cases are due to *Staphylococcus aureus*; gram-negative bacilli, *Streptococcus*, anaerobes, and fungi can also cause epidural abscesses. Tuberculosis from an adjacent vertebral source (Pott's disease) remains an important cause in the underdeveloped world **(Fig. 377-5)**.

Figure 377-5

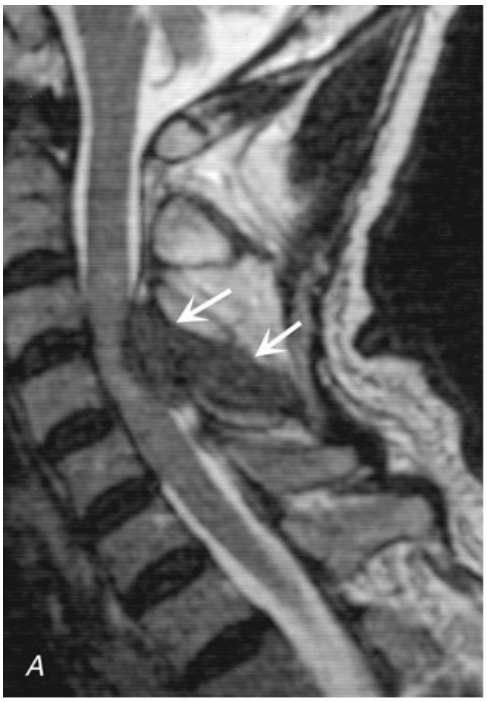

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

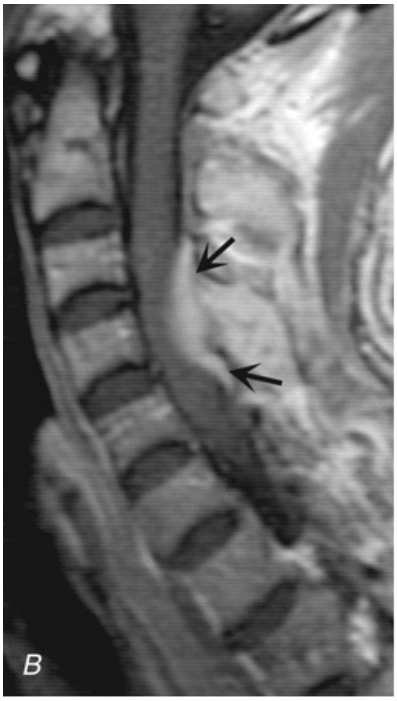

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI of a spinal epidural abscess due to tuberculosis.*A.* Sagittal T2-weighted free spin-echo MR sequence. A hypointense mass replaces the posterior elements of C3 and extends epidurally to compress the spinal cord (*arrows*). ***B.***Sagittal T1-weighted image after contrast administration reveals a diffuse enhancement of the epidural process (*arrows*) with extension into the epidural space.

MRI scans localize the abscess and exclude other causes of myelopathy. Lumbar puncture is only required if encephalopathy or other clinical signs raise the question of associated meningitis, a feature that is found in <25% of cases. The level of the puncture should be planned to minimize the risk of meningitis due to passage of the needle through infected tissue. A high cervical tap is sometimes the safest approach. CSF abnormalities in subdural abscess consist of pleocytosis with a preponderance of polymorphonuclear cells, an elevated protein level, and a reduced glucose level, but the responsible organism is not cultured unless there is associated meningitis. Blood cultures are positive in <25% of cases.

Treatment: Spinal Epidural Abscess

Treatment is by decompressive laminectomy with debridement combined with long-term antibiotic treatment. Surgical evacuation prevents development of paralysis and may improve or reverse paralysis in evolution, but it is unlikely to improve deficits of more than several days duration. Broad-spectrum antibiotics should be started empirically before surgery and then modified on the basis of culture results; medication is continued for at least 4 weeks. If surgery is contraindicated or if there is a fixed paraplegia or quadriplegia that is unlikely to improve following surgery, long-term administration of systemic and oral antibiotics can be used; in such cases, the choice of antibiotics may be guided by results of blood cultures. However, paralysis may develop or progress during antibiotic therapy; thus, initial surgical management remains the treatment of choice unless the abscess is limited in size and causes few or no neurologic signs.

With prompt diagnosis and treatment of spinal epidural abscess, up to two-thirds of patients experience significant recovery.

SPINAL EPIDURAL HEMATOMA

Hemorrhage into the epidural (or subdural) space causes acute focal or radicular pain followed by variable signs of a spinal cord or conus medullaris disorder. Therapeutic anticoagulation, trauma, tumor, or blood dyscrasias are predisposing conditions. Rare cases complicate lumbar puncture or epidural anesthesia. MRI and CT confirm the clinical suspicion and can delineate the extent of the bleeding. Treatment consists of prompt reversal of any underlying clotting disorder and surgical decompression. Surgery may be followed by substantial recovery, especially in patients with some preservation of motor function preoperatively. Because of the risk of hemorrhage, lumbar puncture should be avoided whenever possible in patients with severe thrombocytopenia or other coagulopathies.

HEMATOMYELIA

Hemorrhage into the substance of the spinal cord is a rare result of trauma, intraparenchymal vascular malformation (see below), vasculitis due to polyarteritis nodosa or systemic lupus erythematosus (SLE), bleeding disorders, or a spinal cord neoplasm. Hematomyelia presents as an acute painful transverse myelopathy. With large lesions, extension into the subarachnoid space results in subarachnoid hemorrhage (Chap. 275). Diagnosis is by MRI or CT. Therapy is supportive, and surgical intervention is generally not useful. An exception is hematomyelia due to an underlying vascular malformation, for which selective spinal angiography may be indicated, followed by surgery to evacuate the clot and remove the underlying vascular lesion.

Noncompressive Myelopathies

The most frequent causes of noncompressive acute transverse myelopathy (ATM) are spinal cord infarction; systemic inflammatory disorders, including SLE and sarcoidosis; demyelinating diseases, including multiple sclerosis (MS); neuromyelitis optica (NMO); postinfectious or idiopathictransverse myelitis, which is presumed to be an immune condition related to acute disseminated encephalomyelitis (Chap. 380); and infectious (primarily viral) causes. After spinal cord compression is excluded, the evaluation generally requires a lumbar puncture and a search for underlying systemic disease **(Table 377-3)**.

Table 377-3 Evaluation of Acute Transverse Myelopathy

1. MRI of spinal cord with and without contrast (exclude compressive causes).
2. CSF studies: Cell count, protein, glucose, IgG index/synthesis rate, oligoclonal bands, VDRL; Gram's stain, acid-fast bacilli, and India ink stains; PCR for VZV, HSV-2, HSV-1, EBV, CMV, HHV-6, enteroviruses, HIV; antibody for HTLV-I, Borrelia burgdorferi, Mycoplasma pneumoniae, and Chlamydia pneumoniae; viral, bacterial, mycobacterial, and fungal cultures.
3. Blood studies for infection: HIV; RPR; IgG and IgM enterovirus antibody; IgM mumps, measles, rubella, group B arbovirus, Brucella melitensis, Chlamydia psittaci, Bartonella henselae, schistosomal antibody; cultures for B. melitensis. Also consider nasal/pharyngeal/anal cultures for enteroviruses; stool O&P for Schistosoma ova.
4. Immune-mediated disorders: ESR; ANA; ENA; dsDNA; rheumatoid factor; anti-SSA; anti-SSB, complement levels; antiphospholipid and anticardiolipin antibodies; p-ANCA; antimicrosomal and antithyroglobulin antibodies; if Sjögren syndrome suspected, Schirmer test, salivary gland scintography, and salivary/lacrimal gland biopsy.
5. Sarcoidosis: Serum angiotensin-converting enzyme; serum Ca; 24-hour urine Ca; chest x-ray; chest CT; total body gallium scan; lymph node biopsy.
6. Demyelinating disease: Brain MRI scan, evoked potentials, CSF oligoclonal bands, neuromyelitis optica antibody (anti-aquaporin-4 [NMO] antibody).
7. Vascular causes: CT myelogram; spinal angiogram.

Abbreviations: ANA, antinuclear antibodies; CMV, cytomegalovirus; EBV, Epstein-Barr virus; ENA, epithelial neutrophil-activating peptide; ESR, erythrocyte sedimentation rate; HHV, human herpes virus; HSV, herpes simplex virus; HTLV, human T cell leukemia/lymphoma virus; O&P, ova and parasites; p-ANCA, perinuclear antineutrophilic cytoplasmic antibodies; PCR, polymerase chain reaction; RPR, rapid plasma reagin (test); VDRL, Venereal Disease Research Laboratory; VZV, varicella-zoster virus.

SPINAL CORD INFARCTION

The cord is supplied by three arteries that course vertically over its surface: a single anterior spinal artery and paired posterior spinal arteries. In addition to the vertebral arteries, the anterior spinal artery is fed by radicular vessels that arise at C6, at an upper thoracic level, and, most consistently, at T11-L2 (artery of Adamkiewicz). At each segment, paired penetrating vessels branch from the anterior spinal artery to supply the anterior two-thirds of the spinal cord; the posterior spinal arteries, which often become less distinct below the midthoracic level, supply the posterior columns.

Spinal cord ischemia can occur at any level; however, the presence of the artery of Adamkiewicz creates a watershed of marginal blood flow in the upper thoracic segments. With systemic hypotension or cross-clamping of the aorta, cord infarction occurs at the level of greatest ischemic risk, usually T3-T4, and also at boundary zones between the anterior and posterior spinal artery territories which may result in a rapidly progressive syndrome (over hours) of weakness and spasticity with little sensory change.

Acute infarction in the territory of the *anterior spinal artery* produces paraplegia or quadriplegia, dissociated sensory loss affecting pain and temperature sense but sparing vibration and position sense, and loss of sphincter control ("anterior cord syndrome"). Onset may be sudden and dramatic but more typically is progressive over minutes or a few hours, quite unlike stroke in the cerebral hemispheres. Sharp midline or radiating back pain localized to the area of ischemia is frequent. Areflexia due to spinal shock is often present initially; with time, hyperreflexia and spasticity appear. Less common is infarction in the territory of the *posterior spinal arteries*, resulting in loss of posterior column function.

Spinal cord infarction results from aortic atherosclerosis, dissecting aortic aneurysm (manifest as chest or back pain with diminished pulses in legs), vertebral artery occlusion or dissection in the neck, aortic surgery, or profound hypotension from any cause. Cardiogenic emboli and vasculitis related to collagen vascular disease [particularly SLE, Sjögren's syndrome, and the antiphospholipid antibody syndrome (see below)] are other causative conditions. Occasional cases develop from *embolism of nucleus pulposus* material into spinal vessels, usually from local spine trauma. In a substantial number of cases no cause can be found, and thromboembolism in arterial feeders is suspected. The MRI may fail to demonstrate limited infarctions of the cord, especially in the first day, but as often it becomes abnormal at the affected level.

In cord infarction due to presumed thromboembolism, acute anticoagulation is probably not indicated, with the exception of the unusual transient ischemic attack or incomplete infarction with a stuttering or progressive course. The antiphospholipid antibody syndrome is treated with anticoagulation. Drainage of spinal fluid has reportedly been successful in some cases of cord infarction but has not been studied systematically.

INFLAMMATORY AND IMMUNE MYELOPATHIES (MYELITIS)

This broad category includes the demyelinating conditions MS, NMO, and postinfectious myelitis, as well as sarcoidosis and connective tissue disease. In approximately one-quarter of cases of myelitis, no underlying cause can be identified. Some will later manifest additional symptoms of an immune-mediated disease.*Recurrent episodes of myelitis* are usually due to one of the immune-mediated diseases or to infection with herpes simplex virus (HSV) type 2 (below).

Multiple Sclerosis

MS (Chap. 380) may present with acute myelitis, particularly in individuals of Asian or African ancestry. In whites, MS rarely causes a complete transverse myelopathy (i.e., acute bilateral signs), but it is among the most common causes of a partial syndrome. MRI findings in MS-associated myelitis typically consist of mild swelling and edema of the cord and diffuse or multifocal areas of abnormal signal on T2-weighted sequences. Contrast enhancement, indicating disruption in the blood-brain barrier associated with inflammation, is present in many acute cases. A brain MRI is most helpful in gauging the likelihood that a case of myelitis represents an initial attack of MS. A normal scan indicates that the risk of evolution to MS is low, ~10–15% over 5 years; in contrast, the finding of multiple periventricular T2-bright lesions indicates a much higher risk, >50% over 5 years and >90% by 14 years. The CSF may be normal, but more often there is a mild mononuclear cell pleocytosis, with normal or mildly elevated CSF protein levels; oligoclonal bands are variable, but when bands are present, a diagnosis of MS is more likely.

There are no adequate trials of therapy for MS-associated transverse myelitis. Intravenous methylprednisolone (500 mg qd for 3 days) followed by oral prednisone (1 mg/kg per day for several weeks, then gradual taper) has been used as initial treatment. A course of plasma exchange is

indicated for severe cases if glucocorticoids are ineffective.

Neuromyelitis Optica

NMO is an immune-mediated demyelinating disorder consisting of a severe myelopathy that is typically longitudinally extensive, meaning that the lesion spans three or more vertebral segments. NMO is associated with optic neuritis that is often bilateral and may precede or follow myelitis by weeks or months, and also by brainstem and in some cases hypothalamic involvement. However, isolated recurrent myelitis without optic nerve involvemement can occur in NMO; affected individuals are usually female, and often of Asian ancestry. CSF studies reveal a variable mononuclear pleocytosis of up to several hundred cells per microliter; unlike MS, oligoclonal bands are generally absent. Diagnostic serum autoantibodies against the water channel protein aquaporin-4 are present in 60 –70% of patients with NMO. NMO has also been associated with SLE and antiphospholipid antibodies (see below) as well as with other connective tissue diseases; rare cases are paraneoplastic in origin. Treatment is with glucocorticoids and, for refractory cases, plasma exchange (as for MS, above). Preliminary data suggest that treatment with azathioprine, mycophenolate, or anti-CD20 (anti–B cell) monoclonal antibody may protect against subsequent relapses; treatment for 5 years or longer is generally recommended. **NMO is discussed in Chap. 380.**

Systemic Immune-Mediated Disorders

Myelitis occurs in a small number of patients with SLE, many cases of which are associated with antiphospholipid antibodies. The CSF is usually normal or shows a mild lymphocytic pleocytosis; oligoclonal bands are a variable finding. Responses to glucocorticoids and/or cyclophosphamide have been reported, but there is no systematic evidence of their benefit. Other immune-mediated myelitides include cases associated with Sjögren's syndrome, mixed connective tissue disease, Behçet's syndrome, vasculitis with perinuclear antineutrophilic cytoplasmic antibodies (p-ANCA), and primary CNS vasculitis.

Another important consideration in this group is sarcoid mye-lopathy that may present as a slowly progressive or relapsing disorder. MRI reveals an edematous swelling of the spinal cord that may mimic tumor; there is almost always gadolinium enhancement of active lesions and in some cases of the adjacent surface of the cord; lesions may be single or multiple, and on axial images, enhancement of the central cord is usually present. The typical CSF profile consists of a variable lymphocytic pleocytosis and mildly elevated protein level; in a minority of cases reduced glucose and oligoclonal bands are found. The diagnosis is particularly difficult when systemic manifestations of sarcoid are minor or absent (nearly 50% of cases) or when other typical neurologic manifestations of the disease —such as cranial neuropathy, hypothalamic involvement, or meningeal enhancement visualized by MRI—are lacking. A slit-lamp examination of the eye to search for uveitis; chest x-ray and CT to assess pulmonary involvement; and mediastinal lymphadenopathy, serum or CSF angiotensin-converting enzyme (ACE; present in only a minority of cases), serum calcium, and a gallium scan may assist in the diagnosis. The usefulness of spinal fluid ACE is uncertain. Initial treatment is with oral glucocorticoids; immunosuppressant drugs are used for resistant cases. **Sarcoidosis is discussed in Chap. 329.**

Postinfectious Myelitis

Many cases of myelitis, termed *postinfectious* or *postvaccinal*, follow an infection or vaccination. Numerous organisms have been implicated, including Epstein-Barr virus (EBV), cytomegalovirus (CMV), mycoplasma, influenza, measles, varicella, rubeola, and mumps. As in the related disorder acute disseminated encephalomyelitis (Chap. 380), postinfectious myelitis often begins as the patient appears to be recovering from an acute febrile infection, or in the subsequent days or weeks, but an infectious agent cannot be isolated from the nervous system or spinal fluid. The presumption is that the myelitis represents an autoimmune disorder triggered by infection and is not due to direct infection of the spinal cord. No randomized controlled trials of therapy exist; treatment is usually with glucocorticoids or, in fulminant cases, plasma exchange.

Acute Infectious Myelitis

Many viruses have been associated with an acute myelitis that is infectious in nature rather than postinfectious. Nonetheless, the two processes are often difficult to distinguish. Herpes zoster is the best characterized viral myelitis, but herpes simplex virus (HSV) types 1 and 2, EBV, CMV, and rabies virus are other well-described causes. HSV-2 (and less commonly HSV-1) produces a distinctive syndrome of recurrent sacral myelitis in association with outbreaks of genital herpes mimicking MS. Poliomyelitis is the prototypic viral myelitis, but it is more or less restricted to the gray matter of the cord. Chronic viral myelitic infections, such as that due to HIV, are discussed below.

Bacterial and mycobacterial myelitis (most are essentially abscesses) are far less common than viral causes and much less frequent than cerebral bacterial abscess. Almost any pathogenic species may be responsible, including *Listeria monocytogenes*, *Borrelia burgdorferi* (Lyme disease), and *Treponema pallidum* (syphilis). *Mycoplasma pneumoniae* may be a cause of myelitis, but its status is uncertain since many cases are more properly classified as postinfectious.

Schistosomiasis is an important cause of parasitic myelitis in endemic areas. The process is intensely inflammatory and granulomatous, caused by a local response to tissue-digesting enzymes from the ova of the parasite, typically *S. mansoni*. Toxoplasmosis can occasionally cause a focal myelopathy, and this diagnosis should be considered in patients with AIDS (Chap. 214).

In cases of suspected viral myelitis, it may be appropriate to begin specific therapy pending laboratory confirmation. Herpes zoster, HSV, and EBV myelitis are treated with intravenous acyclovir (10 mg/kg q8h) or oral valacyclovir (2 g tid) for 10–14 days; CMV with ganciclovir (5 mg/kg IV bid) plus foscarnet (60 mg/kg IV tid), or cidofovir (5 mg/kg per week for 2 weeks).

CHRONIC MYELOPATHIES

Spondylitic Myelopathy

Spondylitic myelopathy is one of the most common causes of chronic cord compression and of gait difficulty in the elderly. Neck and shoulder pain with stiffness are early symptoms; impingement of bone and soft tissue overgrowth on nerve roots results in radicular arm pain, most often in a C5 or C6 distribution. Compression of the cervical cord, which occurs in fewer than one-third of cases, produces a slowly progressive spastic paraparesis, at times asymmetric and often accompanied by paresthesias in the feet and hands. Vibratory sense is diminished in the legs, there is a Romberg sign, and occasionally there is a sensory level for vibration on the upper thorax. In some cases, coughing or straining produces leg weakness or radiating arm or shoulder pain. Dermatomal sensory loss in the arms, atrophy of intrinsic hand muscles, increased deep-tendon reflexes in the legs, and extensor plantar responses are common. Urinary urgency or incontinence occurs in advanced cases, but there are many alternative causes of these problems in older individuals. A tendon reflex in the arms is often diminished at some level; most often at the biceps (C5-C6). In individual cases, radicular, myelopathic, or combined signs may predominate. The diagnosis should be considered in cases of progressive cervical myelopathy, paresthesias of the feet and hands, or wasting of the hands.

Diagnosis is usually made by MRI and may be suspected from CT images; plain X-rays are less helpful. Extrinsic cord compression and deformation is appreciated on axial MRI views, and T2-weighted sequences may reveal areas of high signal intensity within the cord adjacent to the site of compression. A cervical collar may be helpful in milder cases, but definitive therapy consists of surgical decompression. Posterior laminectomy or an anterior approach with resection of the protruded disk and bony material may be required. **Cervical spondylosis and related degenerative diseases of the spine are discussed in Chap. 15.**

Vascular Malformations of the Cord and Dura

Vascular malformations of the cord and overlying dura are treatable causes of progressive myelopathy. Most common are fistulas located posteriorly along the surface of the cord or within the dura. Most dural arteriovenous (AV) fistulas are located at or below the midthoracic level, usually consisting of a direct connection between a radicular feeding artery in the nerve root sleeve with dural veins. The typical presentation is a middle-aged man with a progressive myelopathy that worsens slowly or intermittently and may have periods of remission resembling MS. Acute deterioration due to hemorrhage into the spinal cord or subarachnoid space may also occur but is rare. A saltatory progression is common and is the result of local ischemia and edema from venous congestion. Most patients have incomplete sensory, motor, and bladder disturbances. The motor disorder may predominate and produce a mixture of upper and restricted lower motor neuron signs, simulating amyotrophic lateral sclerosis (ALS). Pain over the dorsal spine, dysesthesias, or radicular pain may be present. Other symptoms suggestive of arteriovenous malformation (AVM) include intermittent claudication, symptoms that change with posture, exertion such as singing, menses, or fever. Less commonly, AVM disorders are intramedullary rather than dural. One classic syndrome presents as a progressive thoracic myelopathy with paraparesis developing over weeks or several months, characterized pathologically by abnormally thick, hyalinized vessels within the cord (Foix-Alajouanine syndrome).

Spinal bruits are infrequent but should be sought at rest and after exercise in suspected cases. A vascular nevus on the overlying skin may indicate an underlying vascular malformation (Klippel-Trénaunay-Weber syndrome). High-resolution MRI with contrast administration detects many but not all AVMs **(Fig. 377-6)**. An uncertain proportion not detected by MRI may be visualized by CT myelography as enlarged vessels along the surface of the cord. Definitive diagnosis requires selective spinal angiography, which defines the feeding vessels and the extent of the malformation. Endovascular embolization of the major feeding vessels may stabilize a progressive neurologic deficit or allow for gradual recovery.

Figure 377-6

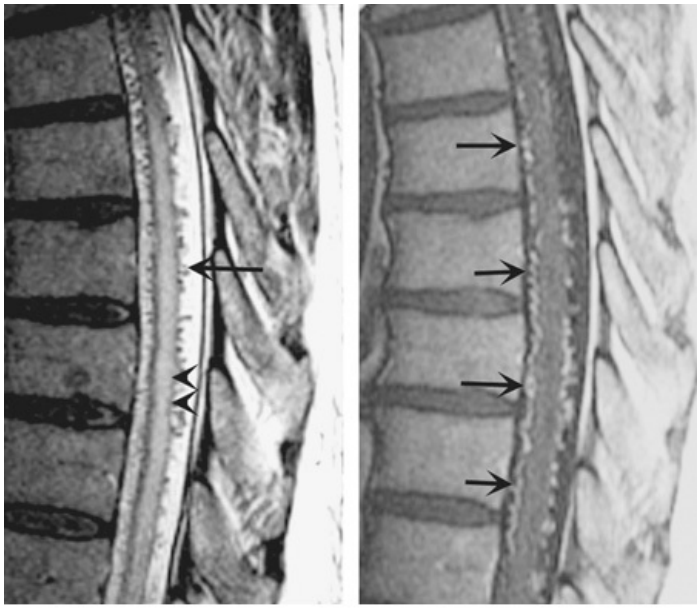

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Arteriovenous malformation. Sagittal MR scans of the thoracic spinal cord: T2 fast spin-echo technique (***left***) and T1 postcontrast image (***right***). On the T2-weighted image (*left*), abnormally high signal intensity is noted in the central aspect of the spinal cord (*arrowheads*). Numerous punctate flow voids indent the dorsal and ventral spinal cord (*arrow*). These represent the abnormally dilated venous plexus supplied by a dural arteriovenous fistula. After contrast administration (*right*), multiple, serpentine, enhancing veins (*arrows*) on the ventral and dorsal aspect of the thoracic spinal cord are visualized, diagnostic of arteriovenous malformation. This patient was a 54-year-old man with a 4-year history of progressive paraparesis.

Retrovirus-Associated Myelopathies

The myelopathy associated with the human T cell lymphotropic virus type I (HTLV-I), formerly called *tropical spastic paraparesis*, is a slowly progressive spastic syndrome with variable sensory and bladder disturbance. Approximately half of patients have mild back or leg pain. The neurologic signs may be asymmetric, often lacking a well-defined sensory level; the only sign in the arms may be hyperreflexia after several years of illness. The onset is insidious, and the illness is slowly progressive at a variable rate; most patients are unable to walk within 10 years of onset. This presentation may resemble primary progressive MS or a thoracic AVM. Diagnosis is made by demonstration of HTLV-I –specific antibody in serum by enzyme-linked immunosorbent assay (ELISA), confirmed by radioimmunoprecipitation or Western blot analysis. There is no effective treatment, but symptomatic therapy for spasticity and bladder symptoms may be helpful.

A progressive myelopathy may also result from HIV infection (Chap. 189). It is characterized by vacuolar degeneration of the posterior and lateral tracts, resembling subacute combined degeneration (see below).

Syringomyelia

Syringomyelia is a developmental cavity of the cervical cord that is prone to enlarge and produce progressive myelopathy. Symptoms begin insidiously in adolescence or early adulthood, progress irregularly, and may undergo spontaneous arrest for several years. Many young patients acquire a cervical-thoracic scoliosis. More than half of all cases are associated with Chiari type 1 malformations in which the cerebellar tonsils protrude through the foramen magnum and into the cervical spinal canal. The pathophysiology of syrinx expansion is controversial, but some interference with the normal flow of CSF seems likely, perhaps by the Chiari malformation. Acquired cavitations of the cord in areas of necrosis are also termed *syrinx cavities*; these follow trauma, myelitis, necrotic spinal cord tumors, and chronic arachnoiditis due to tuberculosis and other etiologies.

The presentation is a central cord syndrome consisting of dissociated sensory loss (loss of pain and temperature sensation with sparing of touch and vibration) and areflexic weakness in the upper limbs. The sensory deficit has a distribution that is "suspended" over the nape of the neck, shoulders, and upper arms (cape distribution) or in the hands. Most cases begin asymmetrically with unilateral sensory loss in the hands that leads to injuries and burns that are not appreciated by the patient. Muscle wasting in the lower neck, shoulders, arms, and hands with asymmetric or absent reflexes in the arms reflects expansion of the cavity into the gray matter of the cord. As the cavity enlarges and further compresses the long tracts, spasticity and weakness of the legs, bladder and bowel dysfunction, and a Horner's syndrome appear. Some patients develop facial numbness and sensory

loss from damage to the descending tract of the trigeminal nerve (C2 level or above). In cases with Chiari malformations, cough-induced headache and neck, arm, or facial pain are reported. Extension of the syrinx into the medulla, syringobulbia, causes palatal or vocal cord paralysis, dysarthria, horizontal or vertical nystagmus, episodic dizziness or vertigo, and tongue weakness with atrophy.

MRI scans accurately identify developmental and acquired syrinx cavities and their associated spinal cord enlargement **(Fig. 377-7)**. MRI scans of the brain and the entire spinal cord should be obtained to delineate the full longitudinal extent of the syrinx, assess posterior fossa structures for the Chiari malformation, and determine whether hydrocephalus is present.

Figure 377-7

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI of syringomyelia associated with a Chiari malformation. Sagittal T1-weighted image through the cervical and upper thoracic spine demonstrates descent of the cerebellar tonsils and vermis below the level of the foramen magnum (*black arrows*). Within the substance of the cervical and thoracic spinal cord, a CSF collection dilates the central canal (*white arrows*).

Treatment: Syringomyelia

Treatment of syringomyelia is generally unsatisfactory. The Chiari tonsillar herniation is usually decompressed, generally by suboccipital craniectomy, upper cervical laminectomy, and placement of a dural graft. Obstruction of fourth ventricular outflow is reestablished by this procedure. If the syrinx cavity is large, some surgeons recommend direct decompression or drainage by one of a number of methods, but the added benefit of this procedure is uncertain, and morbidity is common. With Chiari malformations, shunting of hydrocephalus should generally precede any attempt to correct the syrinx. Surgery may stabilize the neurologic deficit, and some patients improve.

Syringomyelia secondary to trauma or infection is treated with a decompression and drainage procedure in which a small shunt is inserted between the syrinx cavity and the subarachnoid space; alternatively, the cavity can be fenestrated. Cases due to intramedullary spinal cord tumor are generally managed by resection of the tumor.

Chronic Myelopathy of Multiple Sclerosis

A chronic progressive myelopathy is the most frequent cause of disability in both primary progressive and secondary progressive forms of MS. Involvement is typically bilateral but asymmetric and produces motor, sensory, and bladder and bowel disturbances. Fixed motor disability appears to result from extensive loss of axons in the corticospinal tracts. Diagnosis is facilitated by identification of earlier attacks such as optic neuritis. MRI, CSF, and evoked-response testing are confirmatory. Therapy with interferon β, glatiramer acetate, or natalizumab is indicated for patients with progressive myelopathy, who also have coexisting MS relapses. These therapies are sometimes also offered to patients without relapses, despite the lack of evidence supporting their value in this setting. The value of anti-B cell therapy in primary progressive MS is under investigation. MS is discussed in Chap. 380.

Subacute Combined Degeneration (Vitamin B_{12} Deficiency)

This treatable myelopathy presents with subacuteparesthesias in the hands and feet, loss of vibration and position sensation, and a progressive spastic and ataxic weakness. Loss of reflexes due to an associated peripheral neuropathy in a patient who also has Babinski signs is an important diagnostic clue. Optic atrophy and irritability or other mental changes may be prominent in advanced cases but are rarely the presenting symptoms. The myelopathy of subacute combined degeneration tends to be diffuse rather than focal; signs are generally symmetric and reflect predominant involvement of the posterior and lateral tracts, including Romberg's sign. The diagnosis is confirmed by the finding of macrocytic red blood cells, a low serum B_{12} concentration, and elevated serum levels of homocysteine and methylmalonic acid. Treatment is by replacement therapy, beginning with 1000 μg of intramuscular vitamin B12 repeated at regular intervals or by subsequent oral treatment (Chap. 105).

Hypocupric Myelopathy

This myelopathy is virtually identical to subacute combined degeneration (above) and probably explains many cases previously described with normal serum levels of B_{12}. Low levels of serum copper are found and often there is also a low level of serum ceruloplasmin. Some cases follow gastrointestinal procedures that result in impaired copper absorption; others have been associated with excess zinc from health food supplements or, until recently, use of zinc-containing denture creams, which impair copper absorption via induction of metallothionein, a copper-binding protein. Many cases are idiopathic. Improvement or at least stabilization may be expected with reconstitution of copper stores by oral supplementation. The pathophysiology and pathology of the idiopathic form are not known.

Tabes Dorsalis

The classic syndromes of tabes dorsalis and meningovascular syphilis of the spinal cord are now less frequent than in the past but must be considered in the differential diagnosis of spinal cord disorders. The characteristic symptoms of tabes are fleeting and repetitive lancinating pains, primarily in the legs or less often in the back, thorax, abdomen, arms, and face. Ataxia of the legs and gait due to loss of position sense occurs in half of patients. Paresthesias, bladder disturbances, and acute abdominal pain with vomiting (visceral crisis) occur in 15–30% of patients. The cardinal signs of tabes are loss of reflexes in the legs; impaired position and vibratory sense; Romberg's sign; and, in almost all cases, bilateral Argyll Robertson pupils, which fail to constrict to light but accommodate. Diabetic polyradiculopathy may simulate tabes.

Familial Spastic Paraplegia

Many cases of slowly progressive myelopathy are genetic in origin (Chap. 374). More than 20 different causative loci have been identified, including autosomal dominant, autosomal recessive, and X-linked forms. Most patients present with almost imperceptibly progressive spasticity and weakness in the legs, usually but not always symmetric. Sensory symptoms and signs are absent or mild, but sphincter disturbances may be present. In some families additional neurologic signs are prominent, including nystagmus, ataxia, or optic atrophy. The onset may be as early as the first year of life or as late as middle adulthood. Only symptomatic therapies for the spasticity are currently available.

Adrenomyeloneuropathy

This X-linked disorder is a variant of adrenoleukodystrophy. Affected males usually have a history of adrenal insufficiency beginning in childhood and then develop a progressive spastic (or ataxic) paraparesis beginning in early adulthood; some patients also have a mild peripheral neuropathy. Female heterozygotes may develop a slower, insidiously progressive spastic myelopathy beginning later in adulthood and without adrenal insufficiency. Diagnosis is usually made by demonstration of elevated levels of very long chain fatty acids in plasma and in cultured fibroblasts. The responsible gene encodes ADLP, a peroxisomal membrane transporter that is a member of the ATP-binding cassette (ABC) family. Steroid replacement is indicated if hypoadrenalism is present, and bone marrow transplantation and nutritional supplements have been attempted for this condition without clear evidence of efficacy.

Other Chronic Myelopathies

Primary lateral sclerosis (Chap. 374) is a degenerative disorder characterized by progressive spasticity with weakness, eventually accompanied by dysarthria and dysphonia; bladder symptoms occur in approximately half of patients. Sensory function is spared. The disorder resembles ALS and is considered a variant of the motor neuron degenerations, but without the characteristic lower motor neuron disturbance. Some cases may represent familial spastic paraplegia, particularly autosomal recessive or X-linked varieties in which a family history may be absent.

Tethered cord syndrome is a developmental disorder of the lower spinal cord and nerve roots that rarely presents in adulthood as low back pain accompanied by a progressive lower spinal cord and/or nerve root syndrome. Some patients have a small leg or foot deformity indicating a long-standing process, and in others a dimple, patch of hair, or sinus tract on the skin overlying the lower back is the

clue to a congenital lesion. Diagnosis is made by MRI, which demonstrates a low-lying conus medullaris and thickened filum terminale. The MRI may also reveal diastematomyelia (division of the lower spinal cord into two halves), lipomas, cysts, or other congenital abnormalities of the lower spine coexisting with the tethered cord. Treatment is with surgical release.

There are a number of rare toxic causes of spastic myelopathy, including lathyrism due to ingestion of chick peas containing the excitotoxin β-*N*-oxalylamino-l-alanine (BOAA), seen primarily in the developing world, and nitrous oxide inhalation producing a myelopathy identical to subacute combined degeneration. SLE, Sjögren's syndrome, and sarcoidosis may each cause a myelopathy without overt evidence of systemic disease. Cancer-related causes of chronic myelopathy, besides the common neoplastic compressive myelopathy discussed earlier, include radiation injury (Chap. 379) and rare paraneoplastic myelopathies. The latter are most often associated with lung or breast cancer and anti-Hu antibodies (Chap. 101); NMO can also be paraneoplastic in origin (Chap. 380). Metastases to the cord are probably more common than either of these in patients with cancer. Often, a cause of intrinsic myelopathy can be identified only through periodic reassessment.

REHABILITATION OF SPINAL CORD DISORDERS

The prospects for recovery from an acute destructive spinal cord lesion fade after ~6 months. There are currently no effective means to promote repair of injured spinal cord tissue; promising experimental approaches include the use of factors that influence reinnervation by axons of the corticospinal tract, nerve and neural sheath graft bridges, and the local introduction of stem cells. The disability associated with irreversible spinal cord damage is determined primarily by the level of the lesion and by whether the disturbance in function is complete or incomplete **(Table 377-4)**. Even a complete high cervical cord lesion may be compatible with a productive life. The primary goals are development of a rehabilitation plan framed by realistic expectations and attention to the neurologic, medical, and psychological complications that commonly arise.

Table 377-4 Expected Neurologic Function Following Complete Cord Lesions

Level	Self-Care	Transfers	Maximum Mobility
High quadriplegia (C1-C4)	Dependent on others; requires respiratory support	Dependent on others	Motorized wheelchair
Low quadriplegia (C5-C8)	Partially independent with adaptive equipment	May be dependent or independent	May use manual wheelchair, drive an automobile with adaptive equipment
Paraplegia (below T1)	Independent	Independent	Ambulates short distances with aids

Source: Adapted from JF Ditunno, CS Formal: N Engl J Med 330:550, 1994; with permission.

Many of the usual symptoms associated with medical illnesses, especially somatic and visceral pain, may be lacking because of the destruction of afferent pain pathways. Unexplained fever, worsening of spasticity, or deterioration in neurologic function should prompt a search for infection, thrombophlebitis, or an intraabdominal pathology. The loss of normal thermoregulation and inability to maintain normal body temperature can produce recurrent fever (*quadriplegic fever*), although most episodes of fever are due to infection of the urinary tract, lung, skin, or bone.

Bladder dysfunction generally results from loss of supraspinal innervation of the detrusor muscle of the bladder wall and the sphincter musculature. Detrusor spasticity is treated with anticholinergic drugs (oxybutynin, 2.5–5 mg qid) or tricyclic antidepressants that have anticholinergic properties (imipramine, 25–200 mg/d). Failure of the sphincter muscle to relax during bladder emptying (urinary dyssynergia) may be managed with the α-adrenergic blocking agent terazosin hydrochloride (1–2 mg tid or qid), with intermittent catheterization, or, if that is not feasible, by use of a condom catheter in men or a permanent indwelling catheter. Surgical options include the creation of an artificial bladder by isolating a segment of intestine that can be catheterized intermittently (enterocystoplasty) or can drain continuously to an external appliance (urinary conduit). Bladder areflexia due to acute spinal shock or conus lesions is best treated by catheterization. Bowel regimens and disimpaction are necessary in most patients to ensure at least biweekly evacuation and avoid colonic distention or obstruction.

Patients with acute cord injury are at risk for venous thrombosis and pulmonary embolism. During the first 2 weeks, use of calf-compression devices and anticoagulation with heparin (5000 U subcutaneously every 12 h) or warfarin (INR, 2–3) are recommended. In cases of persistent paralysis, anticoagulation should probably be continued for 3 months.

Prophylaxis against decubitus ulcers should involve frequent changes in position in a chair or bed, the use of special mattresses, and cushioning of areas where pressure sores often develop, such as the sacral prominence and heels. Early treatment of ulcers with careful cleansing, surgical or enzyme debridement of necrotic tissue, and appropriate dressing and drainage may prevent infection of

adjacent soft tissue or bone.

Spasticity is aided by stretching exercises to maintain mobility of joints. Drug treatment is effective but may result in reduced function, as some patients depend upon spasticity as an aid to stand, transfer, or walk. Baclofen (15–240 mg/d in divided doses) is effective; it acts by facilitating γ-aminobutyric acid (GABA)-mediated inhibition of motor reflex arcs. Diazepam acts by a similar mechanism and is useful for leg spasms that interrupt sleep (2–4 mg at bedtime). Tizanidine (2–8 mg tid), an α_2-adrenergic agonist that increases presynaptic inhibition of motor neurons, is another option. For nonambulatory patients, the direct muscle inhibitor dantrolene (25–100 mg qid) may be used, but it is potentially hepatotoxic. In refractory cases, intrathecal baclofen administered via an implanted pump, botulinum toxin injections, or dorsal rhizotomy may be required to control spasticity.

Despite the loss of sensory function, many patients with spinal cord injury experience chronic pain sufficient to diminish their quality of life. Randomized controlled studies indicate that gabapentin or pregabalin are useful in this setting.

Management of chronic pain is discussed in Chap. 11.

A paroxysmal autonomic hyperreflexia may occur following lesions above the major splanchnic sympathetic outflow at T6. Headache, flushing, and diaphoresis above the level of the lesion, as well as transient severe hypertension with bradycardia or tachycardia, are the major symptoms. The trigger is typically a noxious stimulus—for example, bladder or bowel distention, a urinary tract infection, or a decubitus ulcer. Treatment consists of removal of offending stimuli; ganglionic blocking agents (mecamylamine, 2.5–5 mg) or other short-acting antihypertensive drugs are useful in some patients.

Attention to these details allows longevity and a productive life for patients with complete transverse myelopathies.

FURTHER READINGS

Abrahm JL et al: Spinal cord compression in patients with advanced metastatic cancer: "All I care about is walking and living my life." JAMA 299:937, 2008[PMID: 18314436] [Full Text]

Cohen-Aubart F et al: Spinal cord sarcoidosis: Clinical and laboratory profile and outcome of 31 patients in a case-control study. Medicine 89:133, 2010[PMID: 20517184] [Full Text]

Cole JS, Patchell RA: Metastatic epidural spinal cord compression. Lancet Neurol 7:459, 2008[PMID: 18420159] [Full Text]

Frohman EM, Wingerchuk DM: Transverse myelitis. N Engl J Med 363: 564, 2010[PMID: 20818891] [Full Text]

Graber JJ, Nolan CP: Myelopathies in patients with cancer. Arch Neurol 67:298, 2010[PMID: 20212227] [Full Text]

Jacob A, Weinshenker BG: An approach to the diagnosis of acute transverse myelitis. Semin Neurol 28:95, 2008

Krings T, Geibprasert S: Spinal dural arteriovenous fistulas. AJNR Am J Neuroradiol 30:639, 2009[PMID: 19213818] [Full Text]

Patchell RA et al: Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: A randomized trial. Lancet 366:643, 2005[PMID: 16112300] [Full Text]

Teasell RW et al: A systematic review of pharmacologic treatments of pain after spinal cord injury. Arch Phys Med Rehabil 91:816, 2010[PMID: 20434623] [Full Text]

Transverse Myelitis Consortium Working Group: Proposed diagnostic criteria and nosology of acute transverse myelitis. Neurology 59:499, 2002

Traul DE et al: Part I: Spinal-cord neoplasms—intradural neoplasms. Lancet Oncol 8:35, 2007[PMID: 17196509] [Full Text]

Yanagawa K et al: Pathologic and immunologic profiles of a limited form of neuromyelitis optica with myelitis. Neurology 73:1628, 2009[PMID: 19917985] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CONCUSSION AND OTHER HEAD INJURIES: INTRODUCTION

Almost 10 million head injuries occur annually in the United States, about 20% of which are serious enough to cause brain damage. Among men <35 years, accidents, usually motor vehicle collisions, are the chief cause of death and >70% of these involve head injury. Furthermore, minor head injuries are so common that almost all physicians will be called upon to provide immediate care or to see patients who are suffering from various sequelae.

Medical personnel caring for head injury patients should be aware that (1) spinal injury often accompanies head injury, and care must be taken in handling the patient to prevent compression of the spinal cord due to instability of the spinal column; (2) intoxication is a common accompaniment of traumatic brain injury and, when appropriate, testing should be carried out for drugs and alcohol; and (3) additional injuries, including rupture of abdominal organs, may produce vascular collapse or respiratory distress that requires immediate attention.

TYPES OF HEAD INJURIES

Concussion

This form of minor head injury refers to an immediate and transient loss of consciousness that is associated with a short period of amnesia. Many patients do not lose consciousness after a minor head injury but instead are dazed or confused, or feel stunned or "star struck." Severe concussion may precipitate a brief convulsion or autonomic signs such as facial pallor, bradycardia, faintness with mild hypotension, or sluggish pupillary reaction, but most patients are quickly neurologically normal.

The mechanics of a typical concussion involve sudden deceleration of the head when hitting a blunt object. This creates an anterior-posterior movement of the brain within the skull due to inertia and rotation of the cerebral hemispheres on the relatively fixed upper brainstem. Loss of consciousness in concussion is believed to result from a transient electrophysiologic dysfunction of the reticular activating system in the upper midbrain that is at the site of rotation (Chap. 274).

Gross and light-microscopic changes in the brain are usually absent following concussion but biochemical and ultrastructural changes, such as mitochondrial ATP depletion and local disruption of the blood-brain barrier, are transient abnormalities. CT and MRI scans are usually normal; however, a small number of patients will be found to have a skull fracture, an intracranial hemorrhage, or brain contusion.

A brief period of both retrograde and anterograde amnesia is characteristic of concussion and it recedes rapidly in alert patients. Memory loss spans the moments before impact but may encompass the previous days or weeks (rarely months). With severe injuries, the extent of retrograde amnesia roughly correlates with the severity of injury. Memory is regained from the most distant to more recent memories, with islands of amnesia occasionally remaining. The mechanism of amnesia is not known. Hysterical posttraumatic amnesia is not uncommon after head injury and should be suspected when inexplicable behavioral abnormalities occur, such as recounting events that cannot be recalled on later testing, a bizarre affect, forgetting one's own name, or a persistent anterograde deficit that is excessive in comparison with the degree of injury. **Amnesia is discussed in Chap. 26.**

A single, uncomplicated concussion only infrequently produces permanent neurobehavioral changes in patients who are free of preexisting psychiatric and neurologic problems. Nonetheless, residual problems in memory and concentration may have an anatomic correlate in microscopic cerebral lesions (see below).

Contusion, Brain Hemorrhage, and Axonal Shearing Lesions

A surface bruise of the brain, or *contusion*, consists of varying degrees of petechial hemorrhage, edema, and tissue destruction. Contusions and deeper hemorrhages result from mechanical forces that displace and compress the hemispheres forcefully and by deceleration of the brain against the inner skull, either under a point of impact (coup lesion) or, as the brain swings back, in the antipolar area (contrecoup lesion). Trauma sufficient to cause prolonged unconsciousness usually produces some degree of contusion. Blunt deceleration impact, as occurs against an automobile dashboard or from falling forward onto a hard surface, causes contusions on the orbital surfaces of the frontal lobes and the anterior and basal portions of the temporal lobes. With lateral forces, as from impact on an automobile door frame, contusions are situated on the lateral convexity of the hemisphere. The clinical

signs of contusion are determined by the location and size of the lesion; often, there are no focal neurologic abnormalities, but these injured regions are later the sites of gliotic scars that may produce seizures. A hemiparesis or gaze preference is fairly typical of moderately sized contusions. Large bilateral contusions produce stupor with extensor posturing, while those limited to the frontal lobes cause a taciturn state. Contusions in the temporal lobe may cause delirium or an aggressive, combative syndrome.

Contusions are easily visible on CT and MRI scans, appearing as inhomogeneous hyperdensities on CT and as hyperintensities on MRI sequences that detect blood; there is usually localized brain edema **(Fig. 378-1)** and some subarachnoid bleeding. Blood in the cerebrospinal fluid (CSF) due to trauma may provoke a mild inflammatory reaction. Over a few days, contusions acquire a surrounding contrast enhancement and edema that may be mistaken for tumor or abscess. Glial and macrophage reactions result in chronic, scarred, hemosiderin-stained depressions on the cortex (*plaques jaunes*) that are the main source of posttraumatic epilepsy.

Figure 378-1

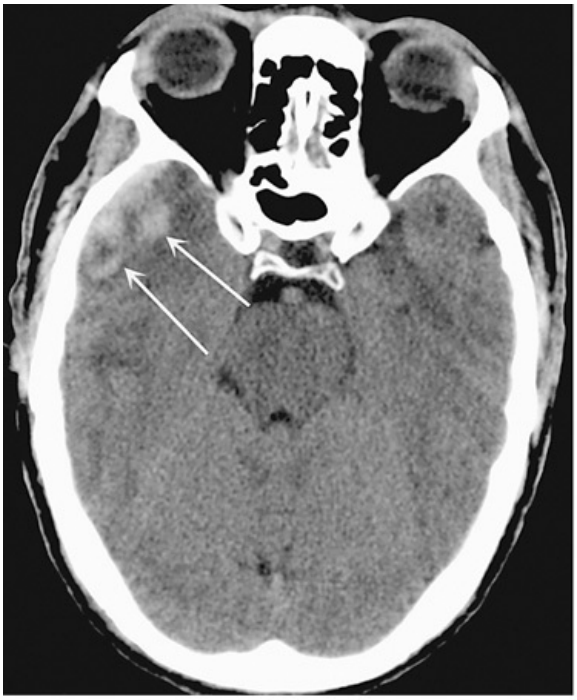

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Traumatic cerebral contusion. Noncontrast CT scan demonstrating a hyperdense hemorrhagic region in the anterior temporal lobe.

Torsional or shearing forces within the brain cause hemorrhages of the basal ganglia and other deep regions. Large hemorrhages after minor trauma suggest that there is a bleeding diathesis or cerebrovascular amyloidosis. For unexplained reasons, deep cerebral hemorrhages may not develop until several days after injury. Sudden neurologic deterioration in a comatose patient or a sudden rise in intracranial pressure (ICP) suggests this complication and should therefore prompt investigation with a CT scan.

A special type of deep white matter lesion consists of widespread mechanical disruption, or shearing, of axons at the time of impact. Most characteristic are small areas of tissue injury in the corpus callosum and dorsolateral pons. The presence of widespread axonal damage in both hemispheres, a state called *diffuse axonal injury* (DAI), has been proposed to explain persistent coma and the vegetative state after closed head injury (Chap. 274), but small ischemic-hemorrhagic lesions in the midbrain and thalamus are as often the cause. Only severe shearing lesions that contain blood are visualized by CT, usually in the corpus callosum and centrum semiovale **(Fig. 378-2)**; however, selective imaging sequences of the MRI can demonstrate such lesions throughout the white matter.

Figure 378-2

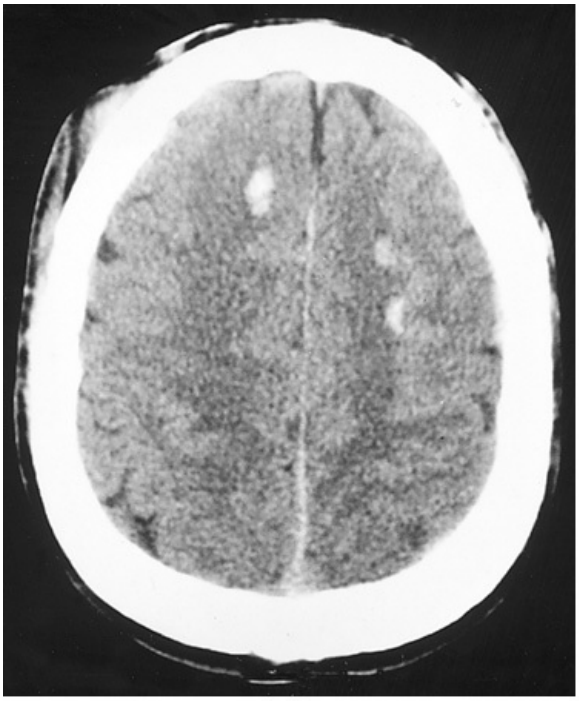

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Multiple small areas of hemorrhage and tissue disruption in the white matter of the frontal lobes on noncontrast CT scan. These appear to reflect an extreme type of the diffuse axonal shearing lesions that occur with closed head injury.

Skull Fractures

A blow to the skull that exceeds the elastic tolerance of the bone causes a fracture. Intracranial lesions accompany roughly two-thirds of skull fractures and the presence of a fracture increases many-fold the chances of an underlying subdural or epidural hematoma. Consequently, fractures are primarily markers of the site and severity of injury. They also provide potential pathways for entry of bacteria to the CSF with a risk of meningitis and for leakage of CSF outward through the dura. Severe orthostatic headache results from lowered pressure in the spinal fluid compartment.

Most fractures are linear and extend from the point of impact toward the base of the skull. Basilar skull fractures are often extensions of adjacent linear fractures over the convexity of the skull but may occur independently owing to stresses on the floor of the middle cranial fossa or occiput. Basilar fractures are usually parallel to the petrous bone or along the sphenoid bone and directed toward the sella turcica and ethmoidal groove. Although most basilar fractures are uncomplicated, they can cause CSF leakage, pneumocephalus, and cavernous-carotid fistulas. Hemotympanum (blood behind the tympanic membrane), delayed ecchymosis over the mastoid process (Battle sign), or periorbital ecchymosis ("raccoon sign") are associated with basilar fractures. Because routine x-ray examination may fail to disclose basilar fractures, they should be suspected if these clinical signs are present.

CSF may leak through the cribriform plate or the adjacent sinus and cause CSF rhinorrhea (a watery discharge from the nose). Persistent rhinorrhea and recurrent meningitis are indications for surgical repair of torn dura underlying the fracture. The site of the leak is often difficult to determine, but useful diagnostic tests include the instillation of water-soluble contrast into the CSF followed by CT with the patient in various positions, or injection of radionuclide compounds or fluorescein into the CSF and the insertion of absorptive nasal pledgets. The location of an intermittent leak is rarely delineated, and many resolve spontaneously.

Sellar fractures, even those associated with serious neuroendocrine dysfunction, may be radiologically occult or evident only by an air-fluid level in the sphenoid sinus. Fractures of the dorsum sella cause sixth or seventh nerve palsies or optic nerve damage.

Petrous bone fractures, especially those oriented along the long axis of the bone, may be associated with facial palsy, disruption of ear ossicles, and CSF otorrhea. Transverse petrous fractures are less common; they almost always damage the cochlea or labyrinths and often the facial nerve as well. External bleeding from the ear is usually from local abrasion of the external canal but can also result from petrous fracture.

Fractures of the frontal bone are usually depressed, involving the frontal and paranasal sinuses and the orbits. Depressed skull fractures are typically compound, but they are often asymptomatic because the impact energy is dissipated in breaking the bone; some have underlying brain contusions.

Debridement and exploration of compound fractures are required in order to avoid infection; simple fractures do not require surgery.

Cranial Nerve Injuries

The cranial nerves most often injured with head trauma are the olfactory, optic, oculomotor, and trochlear; the first and second branches of the trigeminal nerve; and the facial and auditory nerves. Anosmia and an apparent loss of taste (actually a loss of perception of aromatic flavors, with retained elementary taste perception) occur in ~10% of persons with serious head injuries, particularly from falls on the back of the head. This is the result of displacement of the brain and shearing of the fine olfactory nerve filaments that course through the cribriform bone. At least partial recovery of olfactory and gustatory function is expected, but if bilateral anosmia persists for several months, the prognosis is poor. Partial optic nerve injuries from closed trauma result in blurring of vision, central or paracentral scotomas, or sector defects. Direct orbital injury may cause short-lived blurred vision for close objects due to reversible iridoplegia. Diplopia limited to downward gaze and corrected when the head is tilted away from the side of the affected eye indicates trochlear (fourth nerve) nerve damage. It occurs frequently as an isolated problem after minor head injury or may develop for unknown reasons after a delay of several days. Facial nerve injury caused by a basilar fracture is present immediately in up to 3% of severe injuries; it may also be delayed 5-7 days. Fractures through the petrous bone, particularly the less common transverse type, are liable to produce facial palsy. Delayed palsy, the mechanism of which is unknown, has a good prognosis. Injury to the eighth cranial nerve from a fracture of the petrous bone causes loss of hearing, vertigo, and nystagmus immediately after injury. Deafness from eighth nerve injury is rare and must be distinguished from blood in the middle ear or disruption of the middle ear ossicles. Dizziness, tinnitus, and high-tone hearing loss occur from cochlear concussion.

Seizures

Convulsions are surprisingly uncommon immediately after a head injury, but a brief period of tonic extensor posturing or a few clonic movements of the limbs just after the moment of impact can occur. However, the cortical scars that evolve from contusions are highly epileptogenic and may later manifest as seizures, even after many months or years (Chap. 369). The severity of injury roughly determines the risk of future seizures. It has been estimated that 17% of individuals with brain contusion, subdural hematoma, or prolonged loss of consciousness will develop a seizure disorder and that this risk extends for an indefinite period of time, whereas the risk is ≤2% after mild injury. The majority of convulsions in the latter group occur within 5 years of injury but may be delayed for decades. Penetrating injuries have a much higher rate of subsequent epilepsy.

Subdural and Epidural Hematomas

Hemorrhages beneath the dura (subdural) or between the dura and skull (epidural) have characteristic clinical and radiologic features. They are associated with underlying contusions and other injuries, often making it difficult to determine the relative contribution of each component to the clinical state. The mass effect and raised ICP caused by these hematomas can be life threatening, making it imperative to identify them rapidly by CT or MRI scan and to remove them when appropriate.

ACUTE SUBDURAL HEMATOMA (FIG. 378-3)

Figure 378-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute subdural hematoma. Noncontrast CT scan reveals a hyperdense clot which has an irregular border with the brain and causes more horizontal displacement (mass effect) than might be expected from its thickness. The disproportionate mass effect is the result of the large rostral-caudal extent of these hematomas. Compare to Fig. 378-4.

Direct cranial trauma may be minor and is not required for acute subdural hemorrhage to occur, especially in the elderly and those taking anticoagulant medications. Acceleration forces alone, as from whiplash, are sometimes sufficient to produce subdural hemorrhage. Up to one-third of patients have a lucid interval lasting minutes to hours before coma supervenes, but most are drowsy or comatose from the moment of injury. A unilateral headache and slightly enlarged pupil on the side of the hematoma are frequently, but not invariably, present. Stupor or coma, hemiparesis, and unilateral pupillary enlargement are signs of larger hematomas. In an acutely deteriorating patient, burr (drainage) holes or an emergency craniotomy are required. Small subdural hematomas may be asymptomatic and usually do not require evacuation if they do not expand.

A subacutely evolving syndrome due to subdural hematoma occurs days or weeks after injury with drowsiness, headache, confusion, or mild hemiparesis, usually in alcoholics and in the elderly and often after only minor trauma. On imaging studies subdural hematomas appear as crescentic collections over the convexity of one or both hemispheres, most commonly in the frontotemporal region, and less often in the inferior middle fossa or over the occipital poles (Fig. 378-3). Interhemispheric, posterior fossa, or bilateral convexity hematomas are less frequent and are difficult to diagnose clinically, although drowsiness and the neurologic signs expected from damage in each region can usually be detected. The bleeding that causes larger hematomas is primarily venous in origin, although additional arterial bleeding sites are sometimes found at operation, and a few large hematomas have a purely arterial origin.

EPIDURAL HEMATOMA (FIG. 378-4)

Figure 378-4

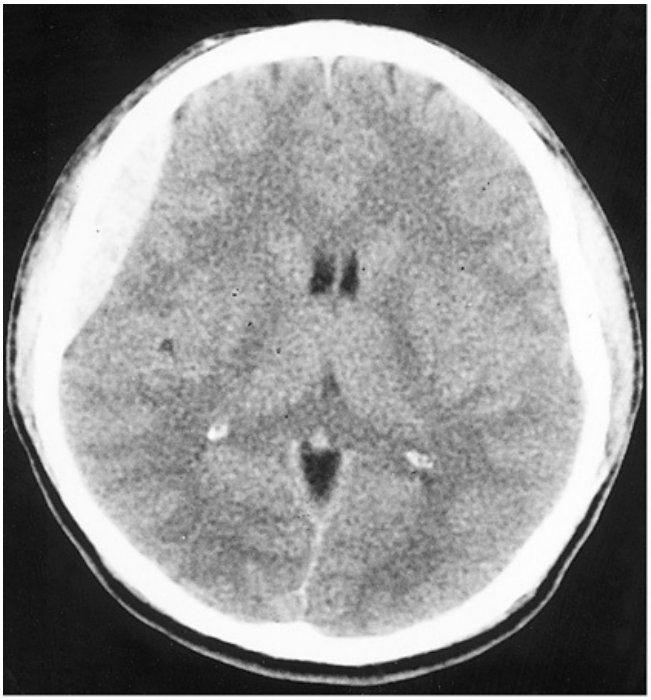

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Acute epidural hematoma. The tightly attached dura is stripped from the inner table of the skull, producing a characteristic lenticular-shaped hemorrhage on noncontrast CT scan. Epidural hematomas are usually caused by tearing of the middle meningeal artery following fracture of the temporal bone.

These evolve more rapidly than subdural hematomas and are correspondingly more treacherous. They occur in up to 10% of cases of severe head injury but are associated with underlying cortical damage less often than are subdural hematomas. Most patients are unconscious when first seen. A "lucid interval" of several minutes to hours before coma supervenes is most characteristic of epidural hemorrhage, but it is still uncommon, and epidural hemorrhage is not the only cause of this temporal sequence. Rapid surgical evacuation and ligation or cautery of the damaged vessel is indicated, usually the middle meningeal artery that has been lacerated by an overlying skull fracture.

CHRONIC SUBDURAL HEMATOMA (FIG. 378-5)

Figure 378-5

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

CT scan of chronic bilateral subdural hematomas of different ages. The collections began as acute hematomas and have become hypodense in comparison to the adjacent brain after a period during which they were isodense and difficult to appreciate. Some areas of resolving blood are contained on the more recently formed collection on the left (*arrows*).

A history of trauma may or may not be elicited in relation to chronic subdural hematoma; the injury may have been trivial and forgotten, particularly in the elderly and those with clotting disorders. Headache is common but not invariable. Additional features may include slowed thinking, vague change in personality, seizure, or a mild hemiparesis. The headache fluctuates in severity, sometimes with changes in head position. Bilateral chronic subdural hematomas produce perplexing clinical syndromes and the initial clinical impression may be of a stroke, brain tumor, drug intoxication, depression, or a dementing illness. Drowsiness, inattentiveness, and incoherence of thought are more generally prominent than focal signs such as hemiparesis. Rarely, chronic hematomas cause brief episodes of hemiparesis or aphasia that are indistinguishable from transient ischemic attacks. Patients with undetected bilateral subdural hematomas have a low tolerance for surgery, anesthesia, and drugs that depress the nervous system; drowsiness or confusion persist for long periods postoperatively.

CT without contrast initially shows a low-density mass over the convexity of the hemisphere (Fig. 378-5). Between 2 and 6 weeks after the initial bleeding the hemorrhage becomes isodense compared to adjacent brain and may be inapparent. Many subdural hematomas that are several weeks in age contain areas of blood and intermixed serous fluid. Bilateral chronic hematomas may fail to be detected because of the absence of lateral tissue shifts; this circumstance in an older patient is suggested by a "hypernormal" CT scan with fullness of the cortical sulci and small ventricles. Infusion of contrast material demonstrates enhancement of the vascular fibrous capsule surrounding the collection. MRI reliably identifies subacute and chronic hematomas.

Clinical observation coupled with serial imaging is a reasonable approach to patients with few symptoms, such as headache alone, and small chronic subdural collections. Treatment of minimally symptomatic chronic subdural hematoma with glucocorticoids is favored by some clinicians, but surgical evacuation is more often successful. The fibrous membranes that grow from the dura and encapsulate the collection require removal to prevent recurrent fluid accumulation. Small hematomas are resorbed, leaving only the organizing membranes. On imaging studies very chronic subdural hematomas are difficult to distinguish from hygromas, which are collections of CSF from a rent in the arachnoid membrane.

CLINICAL SYNDROMES AND TREATMENT OF HEAD INJURY

Minor Injury

The patient who has briefly lost consciousness or been stunned after a minor head injury usually becomes fully alert and attentive within minutes but may complain of headache, dizziness, faintness, nausea, a single episode of emesis, difficulty with concentration, a brief amnestic period, or slight blurring of vision. This typical concussion syndrome has a good prognosis with little risk of subsequent

deterioration. Children are particularly prone to drowsiness, vomiting, and irritability, symptoms that are sometimes delayed for several hours after apparently minor injuries. Vasovagal syncope that follows injury may cause undue concern. Generalized or frontal headache is common in the following days. It may be migrainous (throbbing and hemicranial) in nature or aching and bilateral. After several hours of observation, patients with minor injury may be accompanied home and observed for a day by a family member or friend; written instructions to return if symptoms worsen should be provided.

Persistent severe headache and repeated vomiting in the context of normal alertness and no focal neurologic signs is usually benign, but CT should be obtained and a longer period of observation is appropriate. The decision to perform imaging tests also depends on clinical signs that indicate the impact was severe (e.g., prolonged concussion, periorbital or mastoid hematoma, repeated vomiting, palpable skull fracture), on the seriousness of other bodily injuries, and on the degree of surveillance that can be anticipated after discharge. Two studies have indicated that older age, two or more episodes of vomiting, >30 min of retrograde or persistent anterograde amnesia, seizure, and concurrent drug or alcohol intoxication are sensitive (but not specific) indicators of intracranial hemorrhage that justify CT scanning. It is appropriate to be more liberal in obtaining CT scans in children since a small number, even without loss of consciousness, will have intracranial lesions.

CONCUSSION IN SPORTS

In the current absence of adequate data, a common sense approach to athletic concussion has been to avoid contact sports for at least several days after a mild injury, and for a longer period if there are more severe injuries or if there are protracted neurologic symptoms. The individual then undertakes a graduated program of activity until there are no further symptoms with exercise **(Table 378-1)**. These guidelines are designed in part to avoid the extremely rare *second impact syndrome*, in which cerebral swelling follows a second minor head injury. There is some evidence that repeated concussions are associated with cumulative cognitive deficits, but this and the subsequent risk for dementia and Parkinson,s disease is controversial.

Table 378-1 Guidelines for Management of Concussion in Sports

Severity of Concussion
Grade 1: Transient confusion, no loss of consciousness (LOC), all symptoms resolve within 15 min.
Grade 2: Transient confusion, no LOC, but concussive symptoms or mental status abnormalities persist longer than 15 min.
Grade 3: Any LOC, either brief (seconds) or prolonged (minutes).
On-Site Evaluation
1. Mental status testing
a. Orientation—time, place, person, circumstances of injury
b. Concentration—digits backward, months of year in reverse order
c. Memory—names of teams, details of contest, recent events, recall of three words and objects at 0 and 5 min
2. Finger-to-nose with eyes open and closed
3. Pupillary symmetry and reaction
4. Romberg and tandem gait
5. Provocative testing—40-yard sprint, 5 push ups, 5 sit ups, 5 knee bends (development of dizziness, headaches, or other symptoms is abnormal)
Management Guidelines
Grade 1: Remove from contest. Examine immediately and at 5-min intervals. May return to contest if exam clears within 15 min. A second grade 1 concussion eliminates player for 1 week, with return contingent upon normal neurologic assessment at rest and with exertion.
Grade 2: Remove from contest, cannot return for at least 1 week. Examine at frequent intervals on sideline. Formal neurologic exam the next day. If headache or other symptoms persist for 1 week or longer, CT or MRI scan is indicated. After 1 full asymptomatic week, repeat neurologic assessment at rest and with exercise before cleared to resume play. A second grade 2 concussion eliminates player for at least 2 weeks following complete resolution of symptoms at rest or with exertion. If imaging shows abnormality, player is removed from play for the season.
Grade 3: Transport by ambulance to emergency department if still unconscious or worrisome signs are present; cervical spine stabilization may be indicated. Neurologic exam and, when indicated, CT or MRI scan will guide subsequent management. Hospital admission indicated when signs of pathology are present or if mental status remains abnormal. If findings are normal at the time of the initial medical evaluation, the athlete may be sent home, but daily exams as an outpatient are indicated. A brief (LOC for seconds) grade 3 concussion eliminates player for 1 week, and a prolonged (LOC for minutes) grade 3 concussion for 2 weeks, following complete resolution of symptoms. A second grade 3 concussion should eliminate player from sports for at least 1 month following resolution of symptoms. Any abnormality on CT or MRI scans should result in termination of

the season for the athlete, and return to play at any future time should be discouraged.

Source: Modified from Quality Standards Subcommittee of the American Academy of Neurology: *The American Academy of Neurology Practice Handbook*. The American Academy of Neurology, St. Paul, MN, 1997.

Injury of Intermediate Severity

Patients who are not fully alert or have persistent confusion, behavioral changes, extreme dizziness, or focal neurologic signs such as hemiparesis should be admitted to the hospital and have a CT scan. A cerebral contusion or hematoma is usually found. Common syndromes include: (1) delirium with a disinclination to be examined or moved, expletive speech, and resistance if disturbed (anterior temporal lobe contusions); (2) a quiet, disinterested, slowed mental state (abulia) alternating with irascibility (inferior frontal and frontopolar contusions); (3) a focal deficit such as aphasia or mild hemiparesis (due to subdural hematoma or convexity contusion, or, less often, carotid artery dissection); (4) confusion and inattention, poor performance on simple mental tasks, and fluctuating orientation (associated with several types of injuries, including those described above and with medial frontal contusions and interhemispheric subdural hematoma); (5) repetitive vomiting, nystagmus, drowsiness, and unsteadiness (labyrinthine concussion, but occasionally due to a posterior fossa subdural hematoma or vertebral artery dissection); and (6) diabetes insipidus (damage to the median eminence or pituitary stalk). *Injuries of this degree are often complicated by drug or alcohol intoxication, and clinically inapparent cervical spine injury may be present*.

After surgical removal of hematomas, most patients in this category improve over weeks. During the first week, the state of alertness, memory, and other cognitive functions often fluctuate, and agitation is common. Behavioral changes tend to be worse at night, as with many other encephalopathies, and may be treated with small doses of antipsychotic medications. Subtle abnormalities of attention, intellect, spontaneity, and memory return toward normal weeks or months after the injury, sometimes abruptly. Persistent cognitive problems are discussed below.

Severe Injury

Patients who are comatose from the moment of injury require immediate neurologic attention and resuscitation. After intubation, with care taken to immobilize the cervical spine, the depth of coma, pupillary size and reactivity, limb movements, and Babinski responses are assessed. As soon as vital functions permit and cervical spine x-rays and a CT scan have been obtained, the patient should be transported to a critical care unit. Hypoxia should be reversed, and normal saline used as the resuscitation fluid in preference to albumin. The finding of an epidural or subdural hematoma or large intracerebral hemorrhage is an indication for prompt surgery and intracranial decompression in an otherwise salvageable patient. The use of prophylactic antiepileptic medications has been recommended but there is little supportive data. **Management of raised ICP, a frequent feature of severe head injury, is discussed in Chap. 275.**

Grading and Prognosis

In severe head injury, the clinical features of eye opening, motor responses of the limbs, and verbal output have been found to be generally predictive of outcome. These three responses are assessed by the Glasgow Coma Scale; a score between 3 and 15 is assigned **(Table 378-2)**. Over 85% of patients with aggregate scores of <5 die within 24 h. However, a number of patients with slightly higher scores, including a few without pupillary light responses, survive, suggesting that an initially aggressive approach is justified in most patients. Patients <20 years, particularly children, may make remarkable recoveries after having grave early neurologic signs. In one large study of severe head injury, 55% of children had a good outcome at 1 year, compared with 21% of adults. Older age, increased ICP, early hypoxia or hypotension, compression of the brainstem on CT or MRI, and a delay in the evacuation of large intracranial hemorrhages are indicators of a poor prognosis.

Table 378-2 Glasgow Coma Scale for Head Injury

Eye opening (E)		**Verbal response (V)**	
Spontaneous	4	Oriented	5
To loud voice	3	Confused, disoriented	4
To pain	2	Inappropriate words	3
Nil	1	Incomprehensible sounds	2
		Nil	1
Best motor response (M)			
Obeys	6		
Localizes	5		
Withdraws (flexion)	4		

Abnormal flexion posturing	3	
Extension posturing	2	
Nil	1	

Note: Coma score = E + M + V. Patients scoring 3 or 4 have an 85% chance of dying or remaining vegetative, while scores >11 indicate only a 5–10% likelihood of death or vegetative state and 85% chance of moderate disability or good recovery. Intermediate scores correlate with proportional chances of recovery.

Postconcussion Syndrome

The *postconcussion syndrome* refers to a state following minor head injury consisting of fatigue, dizziness, headache, and difficulty in concentration. The syndrome simulates asthenia and anxious depression. Based on experimental models, it has been proposed that subtle axonal shearing lesions or as yet undefined biochemical alterations account for the cognitive symptoms. In moderate and severe trauma, neuropsychological changes such as difficulty with attention, memory and other cognitive deficits are undoubtedly present, sometimes severe, but many problems identified by formal testing do not affect daily functioning. Test scores tend to improve rapidly during the first 6 months after injury, then more slowly for years.

Management of the postconcussive syndrome requires the identification and treatment of depression, sleeplessness, anxiety, persistent headache, and dizziness. A clear explanation of the problems that may follow concussion has been shown to reduce subsequent complaints. Care is taken to avoid prolonged use of drugs that produce dependence. Headache may initially be treated with acetaminophen and small doses of amitryptiline. Vestibular exercises (Chap. 21) and small doses of vestibular suppressants such as promethazine (Phenergan) may be helpful when dizziness is the main problem. Patients who after minor or moderate injury have difficulty with memory or with complex cognitive tasks at work may be reassured that these problems usually improve over 6–12 months. It is sometimes helpful to obtain serial and quantified neuropsychological testing in order to adjust the work environment to the patient's abilities and to document improvement over time. Whether cognitive exercises are useful in contrast to rest and a reduction in mental challenges is uncertain. Previously energetic and resilient individuals usually have the best recoveries. In patients with persistent symptoms, the possibility exists of malingering or prolongation as a result of litigation.

FURTHER READINGS

DeKosky ST et al: Traumatic brain injury—football, warfare, and long-term effects. N Engl J Med 363:1293, 2010[PMID: 20879875] [Full Text]

Dischinger PC et al: Early predictors of postconcussive syndrome in a population of trauma patients with mild traumatic brain injury. J Trauma 66:289, 2009[PMID: 19204499] [Full Text]

Lovell m: The management of sports-related concussion: Current status and future trends. Clin Sports Med 28:95, 2009[PMID: 19064168] [Full Text]

Ropper AH (ed): *Neurological and Neurosurgical Intensive Care*, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2004

———, Gorson KC: Concussion. N Engl J Med 356:166, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

Primary brain tumors are diagnosed in approximately 52,000 people each year in the United States. At least one-half of these tumors are malignant and associated with a high mortality rate. Glial tumors account for about 60% of all primary brain tumors, and 80% of those are malignant neoplasms. Meningiomas account for 25%, vestibular schwannomas 10%, and central nervous system (CNS) lymphomas about 2%. Brain metastases are three times more common than all primary brain tumors combined and are diagnosed in approximately 150,000 people each year. Metastases to the leptomeninges and epidural space of the spinal cord each occur in approximately 3–5% of patients with systemic cancer and are also a major cause of neurologic disability in this population.

APPROACH TO THE PATIENT: PRIMARY AND METASTATIC TUMORS OF THE NERVOUS SYSTEM

Clinical Features

Brain tumors of any type can present with a variety of symptoms and signs that fall into two categories: general and focal; patients often have a combination of the two **(Table 379-1)**. General or nonspecific symptoms include headache, cognitive difficulties, personality change, and gait disorder. Generalized symptoms arise when the enlarging tumor and its surrounding edema cause an increase in intracranial pressure or direct compression of cerebrospinal fluid (CSF) circulation leading to hydrocephalus. The classic headache associated with a brain tumor is most evident in the morning and improves during the day, but this particular pattern is actually seen in a minority of patients. Headache may be accompanied by nausea or vomiting when intracranial pressure is elevated. Headaches are often holocephalic but can be ipsilateral to the side of a tumor. Occasionally, headaches have features of a typical migraine with unilateral throbbing pain associated with visual scotoma. Personality changes may include apathy and withdrawal from social circumstances, mimicking depression. Focal or lateralizing findings include hemiparesis, aphasia, or visual field defect. Lateralizing symptoms such as hemiparesis are typically subacute and progressive. A visual field defect is often not noticed by the patient; its presence may only be revealed after it leads to an injury such as an automobile accident occurring in the blind visual field. Language difficulties may be mistaken for confusion. Seizures are a common presentation of brain tumors, occurring in about 25% of patients with brain metastases or malignant gliomas but can be the presenting symptom in up to 90% of patients with low-grade gliomas. Most seizures have a focal signature that reflects their location in the brain and many proceed to secondary generalization. All generalized seizures that arise from a brain tumor will have a focal onset whether or not it is apparent clinically.

Table 379-1 Symptoms and Signs at Presentation of Brain Tumors

	High-GradeGlioma (%)	Low-GradeGlioma (%)	Meningioma (%)	Metastases (%)
Generalized				
Impaired cognitive function	50	10	30	60
Hemiparesis	40	10	36	60
Headache	50	40	37	50
Lateralizing				
seizures	20	70+	17	18
Aphasia	20	<5		18
Visual field deficit	—	—	—	7

Neuroimaging

Cranial MRI is the preferred diagnostic test for any patient suspected of having a brain tumor, and should be performed with gadolinium contrast administration. CT scan should be reserved for those patients unable to undergo MRI (e.g., pacemaker). Malignant brain tumors—whether primary or metastatic—typically enhance with gadolinium and may have central areas of necrosis; they are

characteristically surrounded by edema of the neighboring white matter. Low-grade gliomas typically do not enhance with gadolinium and are best appreciated on fluid-attenuated inversion recovery (FLAIR) MR images. Meningiomas have a characteristic appearance on MRI as they are dural-based with a dural tail and compress but do not invade the brain. Dural metastases or a dural lymphoma can have a similar appearance. Imaging is characteristic for many primary and metastatic tumors, but occasionally there is diagnostic uncertainty based on imaging alone. In such patients a brain biopsy may be helpful in determining a definitive diagnosis. However, when a tumor is strongly suspected, the biopsy can be obtained as an intraoperative frozen section before a definitive resection is performed.

Functional MRI is useful in presurgical planning and defining eloquent sensory, motor, and language cortex. Positron emission tomography (PET) is useful in determining the metabolic activity of the lesions seen on MRI; MR perfusion and spectroscopy can provide information on blood flow or tissue composition. These techniques may help distinguish tumor progression from necrotic tissue as a consequence of treatment with radiation and chemotherapy or identify foci of high-grade tumor in an otherwise low-grade-appearing glioma.

Neuroimaging is the only test necessary to diagnose a brain tumor. Laboratory tests are rarely useful, although patients with metastatic disease may have elevation of a tumor marker in their serum that reflects the presence of brain metastases [e.g., human chorionic gonadotropin (βhCG) from testicular cancer]. Additional testing such as cerebral angiogram, electroencephalogram (EEG), or lumbar puncture is rarely indicated or helpful.

Treatment: Brain Tumors

Therapy of any intracranial malignancy requires both symptomatic and definitive treatments. Definitive treatment is based upon the specific tumor type and includes surgery, radiotherapy (RT), and chemotherapy. However, symptomatic treatments apply to brain tumors of any type. Most high-grade malignancies are accompanied by substantial surrounding edema, which contributes to neurologic disability and raised intracranial pressure. Glucocorticoids are highly effective at reducing perilesional edema and improving neurologic function, often within hours of administration. Dexamethasone has been the glucocorticoid of choice because of its relatively low mineralocorticoid activity. Initial doses are typically 12 mg to 16 mg a day in divided doses given orally or IV (both are equivalent). While glucocorticoids rapidly ameliorate symptoms and signs, their long-term use causes substantial toxicity including insomnia, weight gain, diabetes mellitus, steroid myopathy, and personality changes. Consequently, a taper is indicated as definitive treatment is administered and the patient improves.

Patients with brain tumors who present with seizures, require anticonvulsant drug therapy. There is no role for prophylactic anticonvulsant drugs in patients who have not had a seizure, thus their use should be restricted to those who have had a convincing ictal event. The agents of choice are those drugs that do not induce the hepatic microsomal enzyme system. These include levetiracetam, topiramate, lamotrigine, valproic acid, or lacosamide (Chap. 369). Other drugs such as phenytoin and carbamazepine are used less frequently because they are potent enzyme inducers that can interfere with both glucocorticoid metabolism and the metabolism of chemotherapeutic agents needed to treat the underlying systemic malignancy or the primary brain tumor.

Venous thromboembolic disease occurs in 20–30% of patients with high-grade gliomas and brain metastases. Therefore, anticoagulants should be used prophylactically during hospitalization and in patients who are nonambulatory. Those who have had either a deep vein thrombosis or pulmonary embolus can receive therapeutic doses of anticoagulation safely and without increasing the risk for hemorrhage into the tumor. Inferior vena cava filters are reserved for patients with absolute contraindications to anticoagulation such as recent craniotomy.

PRIMARY BRAIN TUMORS

Pathogenesis

No underlying cause has been identified for the majority of primary brain tumors. The only established risk factors are exposure to ionizing radiation (meningiomas, gliomas, and schwannomas) and immunosuppression (primary CNS lymphoma). Evidence for an association with exposure to electromagnetic fields including cellular telephones, head injury, foods containing N-nitroso compounds, or occupational risk factors, are unproven. A small minority of patients have a family history of brain tumors. Some of these familial cases are associated with genetic syndromes **(Table 379-2)**.

Table 379-2 Genetic Syndromes Associated with Primary Brain Tumors

Syndrome	Inheritance	Gene/Protein	Associated Tumors
Cowden's syndrome	AD	Mutations of PTEN (ch10p23)	Dysplastic cerebellar gangliocytoma (Lhermitte-Duclos disease), meningioma, astrocytoma Breast, endometrial, thyroid cancer,

			trichilemmomas
Familial schwannomatosis	Sporadic Hereditary	Mutations in INI1/SNF5 (ch22q11)	Schwannomas, gliomas
Gardner's syndrome	AD	Mutations in APC(ch5q21)	Medulloblastoma, glioblastoma, craniopharyngioma Familial polyposis, multiple osteomas, skin and soft tissue tumors
Gorlin syndrome(Basal cell nevus syndrome)	AD	Mutations in Patched 1 gene (ch9q22.3)	Medulloblastomas Basal cell carcinoma
Li-Fraumeni syndrome	AD	Mutations in p53 (ch17p13.1)	Gliomas, medulloblastomas Sarcomas, breast cancer, leukemias, others
Multiple Endocrine Neoplasia 1 (Werner's syndrome)	AD	Mutations in Menin (ch11q13)	Pituitary adenoma, malignant schwannomas Parathyroid and pancreatic islet cell tumors
Neurofibromatosistype 1 (NF1)	AD	Mutations in NF1/Neurofibromin (ch17q12-22)	Schwannomas, astrocytomas, optic nerve gliomas, meningiomas Neurofibromas, neurofibrosarcomas, others
Neurofibromatosis type 2 (NF2)	AD	Mutations in NF2/Merlin (ch22q12)	Bilateral vestibular schwannomas, astrocytomas, multiple meningiomas, ependymomas
Tuberous sclerosis (TSC) (Bourneville's disease)	AD	Mutations in TSC1/TSC2 (ch9q34/16)	Subependymal giant cell astrocytoma, ependymomas, glioma, ganglioneuroma, hamartoma
Turcot's syndrome	AD AR	Mutations in APCa (ch5) hMLH1 (ch3p21)	Gliomas, medulloblastomas Adenomatous colon polyps, adenocarcinoma
Von Hippel–Lindau (VHL)	AD	Mutations in VHL gene (ch3p25)	Hemangioblastomas Retinal angiomas, renal cell carcinoma, pheochromocytoma, pancreatic tumors and cysts, endolymphatic sac tumors of the middle ear

[a]Various DNA mismatch repair gene mutations may cause a similar clinical phenotype, also referred to as Turcot's syndrome, in which there is a predisposition to nonpolyposis colon cancer and brain tumors. ***Abbreviations:*** AD, autosomal dominant; APC, adenomatous polyposis coli; AR, autosomal recessive; ch, chromosome; PTEN, phosphatase and tensin homologue; TSC, tuberous sclerosis complex.

As with other neoplasms, brain tumors arise as a result of a multistep process driven by the sequential acquisition of genetic alterations. These include loss of tumor suppressor genes [e.g., p53 and phosphatase and tensin homolog on chromosome 10 (PTEN)] and amplification and overexpression of protooncogenes such as the epidermal growth factor receptor (EGFR) and the platelet-derived growth factor receptors (PDGFR). The accumulation of these genetic abnormalities results in uncontrolled cell growth and tumor formation.

Important progress has been made in understanding the molecular pathogenesis of several types of brain tumors, including glioblastomas and medulloblastomas. Glioblastomas can be separated into two main subtypes based on genetic and biologic differences **(Fig. 379-1)**. The majority are primary glioblastomas. These arise de novo and are characterized by EGFR amplification and mutations, and deletion or mutation of PTEN. Secondary glioblastomas arise in younger patients as lower-grade tumors and transform over a period of several years into glioblastomas. These tumors have inactivation of the p53 tumor suppressor gene, overexpression of PDGFR, and mutations of the isocitrate dehydrogenase 1 and 2 genes. Despite their genetic differences, primary and secondary glioblastomas are morphologically indistinguishable, although they are likely to respond differently to molecular therapies. The molecular subtypes of medulloblastomas are also being elucidated.

Approximately 25% of medulloblastomas have activating mutations of the sonic hedgehog signaling pathway, raising the possibility that inhibitors of this pathway may have therapeutic potential.

Figure 379-1

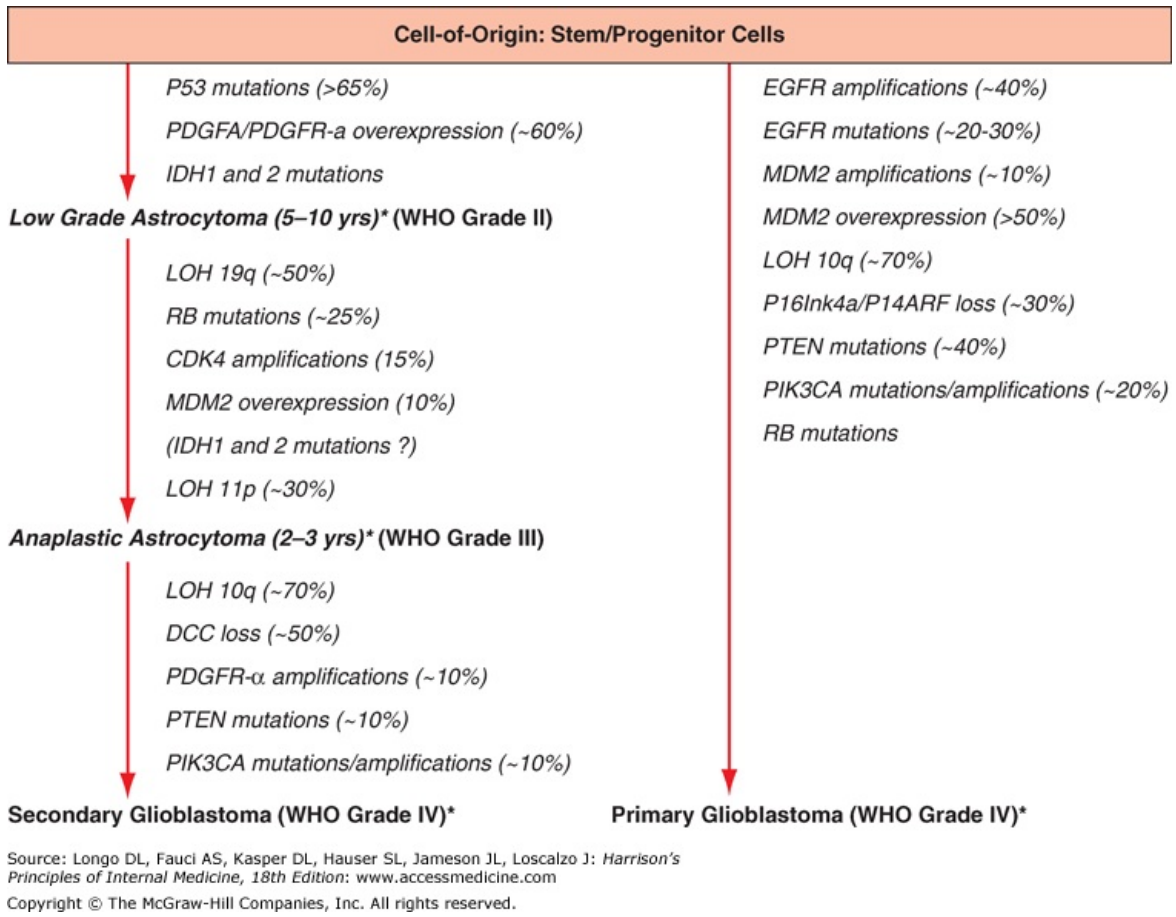

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Genetic and chromosomal alterations involved in the development of primary and secondary glioblastomas. A *slash* indicates one or the other or both. *DCC*, deleted in colorectal carcinoma; *EGFR*, epidermal growth factor receptor; *IDH*, isocitrate dehydrogenase; *LOH*, loss of heterozygosity; *MDM2*, murine double minute 2; *PDGF*, platelet-derived growth factor; *PDGFR*, platelet-derived growth factor receptor; *PIK3CA*, phosphatidylinositol 3-kinase, catalytic; *PTEN*, phosphatase and tensin homologue; *RB*, retinoblastoma; WHO, World Health Organization.

The adult nervous system contains neural stem cells that are capable of self-renewal, proliferation, and differentiation into distinctive mature cell types. There is increasing evidence that neural stem cells, or related progenitor cells, can be transformed into tumor stem cells and give rise to primary brain tumors, including gliomas and medulloblastomas. These stem cells appear to be more resistant to standard therapies than the tumor cells themselves and contribute to the difficulty in eradicating these tumors. There is intense interest in developing therapeutic strategies that effectively target tumor stem cells.

INTRINSIC "MALIGNANT" TUMORS

Astrocytomas

These are infiltrative tumors with a presumptive glial cell of origin. The World Health Organization (WHO) classifies astrocytomas into four prognostic grades based on histologic features: grade I (pilocytic astrocytoma, subependymal giant cell astrocytoma); grade II (diffuse astrocytoma); grade III (anaplastic astrocytoma); and grade IV (glioblastoma). Grades I and II are considered low-grade, and grades III and IV high-grade, astrocytomas.

Low-Grade Astrocytoma

These tumors occur predominantly in children and young adults.

Grade I Astrocytomas

Pilocytic astrocytomas (WHO grade I) are the most common tumor of childhood. They occur typically in the cerebellum but may also be found elsewhere in the neuraxis, including the optic nerves and brainstem. Frequently they appear as cystic lesions with an enhancing mural nodule. They are potentially curable if they can be completely resected. Giant cell subependymal astrocytomas are

usually found in the ventricular wall of patients with tuberous sclerosis. They often do not require intervention but can be treated surgically or with inhibitors of the mammalian target of rapamycin (mTOR).

Grade II Astrocytomas

These are infiltrative tumors that usually present with seizures in young adults. They appear as nonenhancing tumors with increased T2/FLAIR signal **(Fig. 379-2)**. If feasible, patients should undergo maximal surgical resection, although complete resection is rarely possible because of the invasive nature of the tumor. Radiotherapy is helpful, but there is no difference in overall survival between radiotherapy administered postoperatively or delayed until the time of tumor progression. There is increasing evidence that chemotherapeutic agents such as temozolomide, an oral alkylating agent, can be helpful in some patients.

Figure 379-2

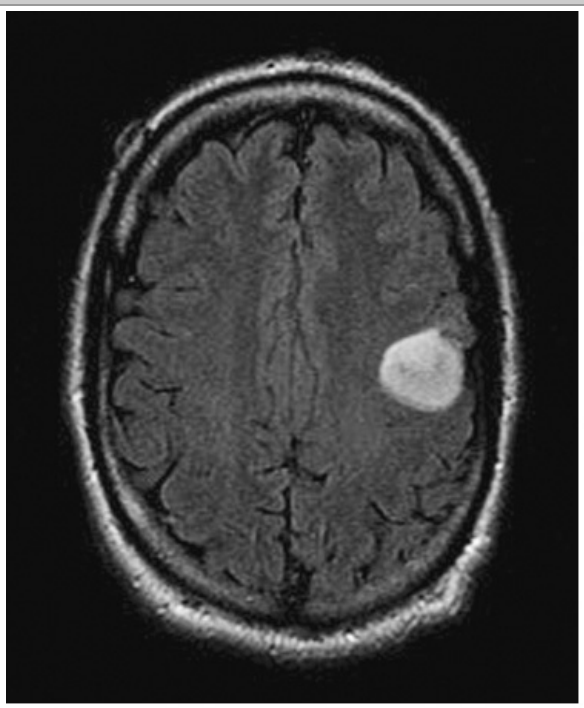

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Fluid-attenuated inversion recovery (FLAIR) MRI of a left frontal low-grade astrocytoma. This lesion did not enhance.

High-Grade Astrocytoma

Grade III (Anaplastic) Astrocytoma

These account for approximately 15–20% of high-grade astrocytomas. They generally present in the fourth and fifth decades of life as variably enhancing tumors. Treatment is the same as for glioblastoma, consisting of maximal safe surgical resection followed by radiotherapy with concurrent and adjuvant temozolomide, or with radiotherapy and adjuvant temozolomide alone.

Grade IV Astrocytoma (Glioblastoma)

Glioblastoma accounts for the majority of high-grade astrocytomas. They are the most common cause of malignant primary brain tumors, with over 10,000 cases diagnosed each year in the United States. Patients usually present in the sixth and seventh decades of life with headache, seizures, or focal neurologic deficits. The tumors appear as ring-enhancing masses with central necrosis and surrounding edema **(Fig. 379-3)**. These are highly infiltrative tumors, and the areas of increased T2/FLAIR signal surrounding the main tumor mass contain invading tumor cells. Treatment involves maximal surgical resection followed by partial-field external beam radiotherapy (6000 cGy in thirty 200-cGy fractions) with concomitant temozolomide, followed by 6–12 months of adjuvant temozolomide. With this regimen, median survival is increased to 14.6 months compared to only 12 months with radiotherapy alone, and 2-year survival is increased to 27%, compared to 10% with radiotherapy alone. Patients whose tumor contains the DNA repair enzyme O^6-methylguanine-DNA methyltransferase (MGMT) are relatively resistant to temozolomide and have a worse prognosis compared to those whose tumors contain low levels of MGMT as a result of silencing of the MGMT gene by promoter hypermethylation. Implantation of biodegradable polymers containing the

chemotherapeutic agent carmustine into the tumor bed after resection of the tumor also produces a modest improvement in survival.

Figure 379-3

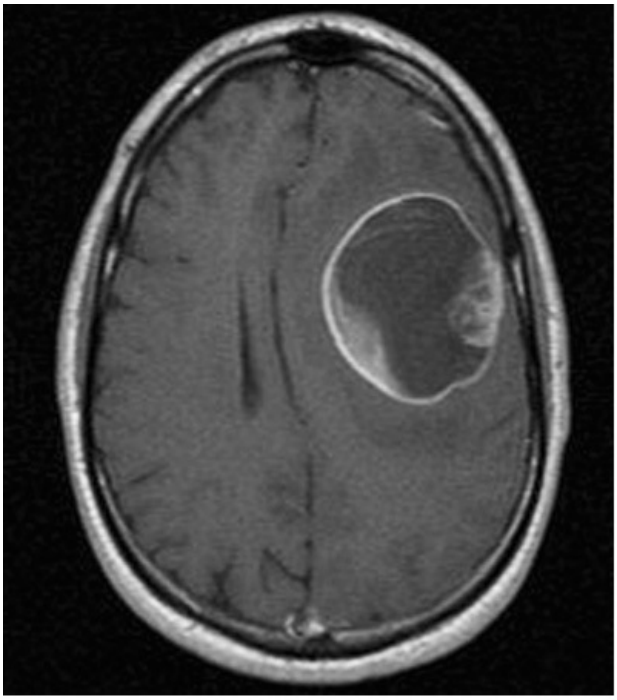

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium T1 MRI of a large cystic left frontal glioblastoma.

Despite optimal therapy, glioblastomas invariably recur. Treatment options for recurrent disease may include reoperation, carmustine wafers, and alternate chemotherapeutic regimens. Reirradiation is rarely helpful. Bevacizumab, a humanized vascular endothelial growth factor (VEGF) monoclonal antibody, has activity in recurrent glioblastoma, increasing progression-free survival and reducing peritumoral edema and glucocorticoid use **(Fig. 379-4)**. Treatment decisions for patients with recurrent glioblastoma must be made on an individual basis, taking into consideration such factors as previous therapy, time to relapse, performance status, and quality of life. Whenever feasible, patients with recurrent disease should be enrolled in clinical trials. Novel therapies undergoing evaluation in patients with glioblastoma include targeted molecular agents directed at receptor tyrosine kinases and signal transduction pathways; anti-angiogenic agents, especially those directed at the VEGF receptors; chemotherapeutic agents that cross the blood-brain barrier more effectively than currently available drugs; gene therapy; immunotherapy; and infusion of radiolabeled drugs and targeted toxins into the tumor and surrounding brain by means of convection-enhanced delivery.

Figure 379-4

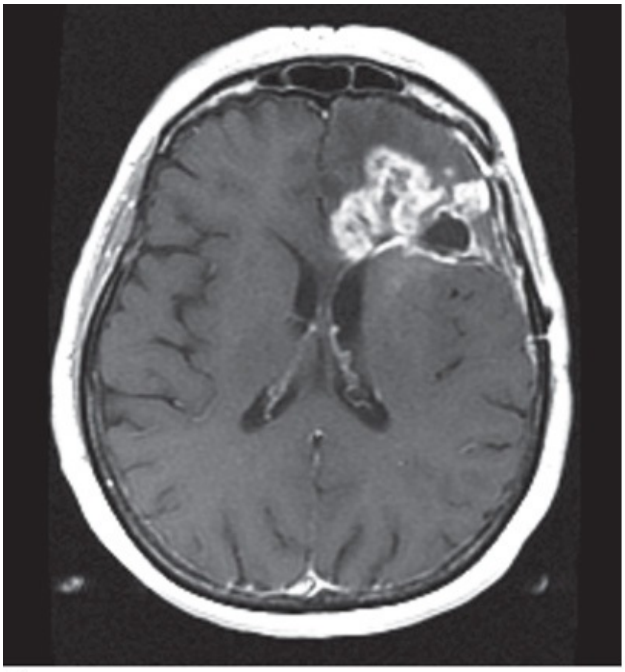

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

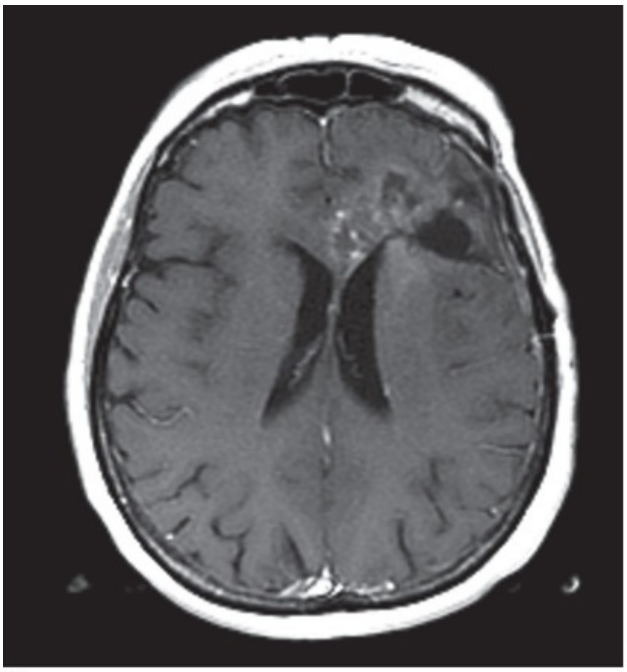

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium T1 MRI of a recurrent glioblastoma before (*A*) and after (*B*) administration of bevacizumab. Note the decreased enhancement and mass effect.

The most important adverse prognostic factors in patients with high-grade astrocytomas are older age, histologic features of glioblastoma, poor Karnofsky performance status, and unresectable tumor. Patients with unmethylated MGMT promoter resulting in the presence of the repair enzyme in tumor cells and resistance to temozolomide also have a worse prognosis.

Gliomatosis Cerebri

Rarely, patients may present with a highly infiltrating, nonenhancing tumor involving more than two lobes. These tumors do not qualify for the histologic diagnosis of glioblastoma but behave aggressively and have a similarly poor outcome. Treatment involves radiotherapy and temozolomide chemotherapy.

OLIGODENDROGLIOMA

Oligodendrogliomas account for approximately 15–20% of gliomas. They are classified by the WHO into well-differentiated oligodendrogliomas (grade II) or anaplastic oligodendrogliomas (AOs) (grade III). Tumors with oligodendroglial components have distinctive features such as perinuclear clearing —giving rise to a "fried-egg" appearance—and a reticular pattern of blood vessel growth. Some tumors have both an oligodendroglial as well as an astrocytic component. These mixed tumors, or oligoastrocytomas (OAs), are also classified into well-differentiated OA (grade II) or anaplastic oligoastrocytomas (AOAs) (grade III).

Grade II oligodendrogliomas and OAs are generally more responsive to therapy and have a better prognosis than pure astrocytic tumors. These tumors present similarly to grade II astrocytomas in young adults. The tumors are nonenhancing and often partially calcified. They should be treated with surgery and, if necessary, radiotherapy and chemotherapy. Patients with oligodendrogliomas have a median survival in excess of 10 years.

Anaplastic oligodendrogliomas and AOAs present in the fourth and fifth decades as variably enhancing tumors. They are more responsive to therapy than grade III astrocytomas. Co-deletion of chromosomes 1p and 19q, mediated by an unbalanced translocation of 19p to 1q, occurs in 61 to 89% of patients with AO and 14 to 20% of patients with AOA. Tumors with the 1p and 19q co-deletion are particularly sensitive to chemotherapy with procarbazine, lomustine [cyclohexylchloroethylnitrosourea (CCNU)], and vincristine (PCV) or temozolomide, as well as to radiotherapy. Median survival of patients with AO or AOA is approximately 3–6 years.

EPENDYMOMAS

Ependymomas are tumors derived from ependymal cells that line the ventricular surface. They account for approximately 5% of childhood tumors and frequently arise from the wall of the fourth ventricle in the posterior fossa. Although adults can have intra-cranial ependymomas, they occur more commonly in the spine, especially in the filum terminale of the spinal cord where they have a myxopapillary histology. Ependymomas that can be completely resected are potentially curable. Partially resected ependymomas will recur and require irradiation. The less common anaplastic ependymomas are more aggressive but can be treated in the same way as ependymomas. Subependymomas are slow-growing benign lesions arising in the wall of ventricles that often do not require treatment.

OTHER LESS COMMON GLIOMAS

Gangliogliomas and pleomorphic xanthoastrocytomas occur in young adults. They behave as more indolent forms of grade II gliomas and are treated in the same way. Brainstem gliomas usually occur in children or young adults. Despite treatment with radiotherapy and chemotherapy, the prognosis is poor with median survival of only 1 year. Gliosarcomas contain both an astrocytic as well as a sarcomatous component and are treated in the same way as glioblastomas.

Primary Central Nervous System Lymphoma

Primary central nervous system lymphoma (PCNSL) is a rare non-Hodgkin's lymphoma accounting for less than 3% of primary brain tumors. For unclear reasons, its incidence is increasing, particularly in immunocompetent individuals.

PCNSL in immunocompetent patients usually consists of diffuse large B-cell lymphomas. PCNSL may also occur in immunocompromised patients, usually those infected with the human immunodeficiency virus (HIV) or organ transplant recipients on immunosuppressive therapy. PCNSL in immunocompromised patients is typically large cell with immunoblastic and more aggressive features. These patients are usually severely immunocompromised with CD4 counts of less than 50/mL. The Epstein-Barr virus (EBV) frequently plays an important role in the pathogenesis of HIV-related PCNSL.

Immunocompetent patients are older (median 60 years) compared to HIV-related PCNSL (median 31 years). PCNSL usually presents as a mass lesion, with neuropsychiatric symptoms, symptoms of increased intracranial pressure, lateralizing signs, or seizures.

On contrast-enhanced MRI, PCNSL usually appears as a densely enhancing tumor **(Fig. 379-5)**. Immunocompetent patients have solitary lesions more often than immunosuppressed patients. Frequently there is involvement of the basal ganglia, corpus callosum, or periventricular region. Although the imaging features are often characteristic, PCNSL can sometimes be difficult to differentiate from high-grade gliomas, infections, or demyelination. Stereotactic biopsy is necessary to obtain a histologic diagnosis. Whenever possible, glucocorticoids should be withheld until after the biopsy has been obtained, since they have a cytolytic effect on lymphoma cells and may lead to nondiagnostic tissue. In addition, patients should be tested for HIV and the extent of disease assessed by performing positron emission tomography (PET) or computerized tomography (CT) of the body, MRI of the spine, CSF analysis, and slit-lamp examination of the eye. Bone marrow biopsy and testicular ultrasound are occasionally performed.

Figure 379-5

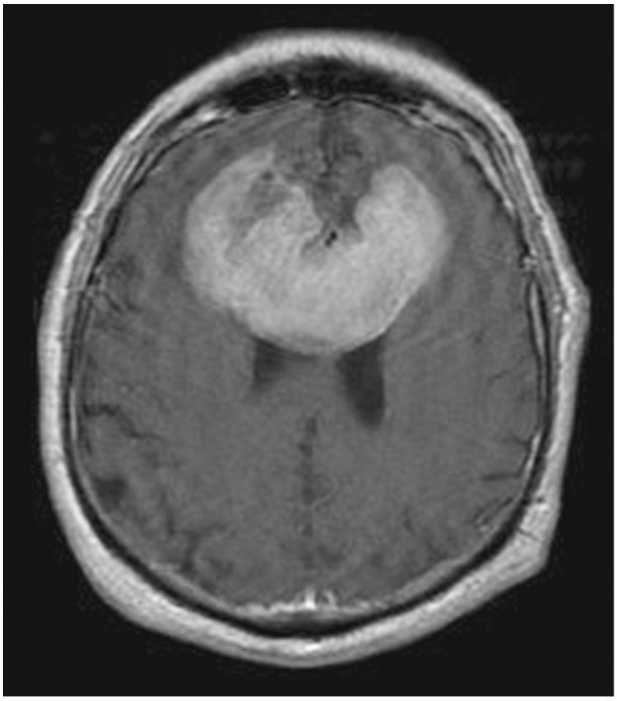

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium T1 MRI demonstrating a large bifrontal primary central nervous system lymphoma (PCNSL). The periventricular location and diffuse enhancement pattern are characteristic of lymphoma.

TREATMENT: PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMA

Unlike other primary brain tumors, PCNSL is relatively sensitive to glucocorticoids, chemotherapy, and radiotherapy. Durable complete responses and long-term survival are possible with these treatments. High-dose methotrexate, a folate antagonist that interrupts DNA synthesis, produces response rates ranging from 35 to 80% and median survival up to 50 months. Combination of methotrexate with other chemotherapeutic agents such as cytarabine, as well as whole-brain radiotherapy, increases the response rate to 70–100%. However, radiotherapy is associated with delayed neurotoxicity, especially in patients over the age of 60 years. As a result radiotherapy is frequently omitted in older patients with PCNSL. There is emerging evidence that the anti-CD20 monoclonal antibody rituximab may have activity in PCNSL, although there remain concerns about its ability to pass through the blood-brain barrier as it becomes reconstituted with therapy. For some patients, high-dose chemotherapy with autologous stem cell rescue may offer the best chance of preventing relapse.

At least 50% of patients will eventually develop recurrent disease. Treatment options include radiotherapy for patients who have not had prior irradiation, re-treatment with methotrexate, as well as other agents such as temozolomide, rituximab, procarbazine, topotecan, and pemetrexed. High-dose chemotherapy with autologous stem cell rescue may have a role in selected patients with relapsed disease.

Pcnsl in Immunocompromised Patients

PCNSL in immunocompromised patients often produces multiple-ring enhancing lesions that can be difficult to differentiate from metastases and infections such as toxoplasmosis. The diagnosis is usually established by examination of the cerebrospinal fluid for cytology and EBV DNA, toxoplasmosis serologic testing, brain PET imaging for hypermetabolism of the lesions consistent with tumor instead of infection, and, if necessary, brain biopsy. Since the advent of highly active antiretroviral drugs, the incidence of HIV-related PCNSL has declined. These patients may be treated with whole-brain radiotherapy, high-dose methotrexate, and initiation of highly active antiretroviral therapy. In organ transplant recipients, reduction of immunosuppression may improve outcome.

Medulloblastomas

Medulloblastomas are the most common malignant brain tumor of childhood, accounting for approximately 20% of all primary CNS tumors among children. They arise from granule cell progenitors or from multipotent progenitors from the ventricular zone. Approximately 5% of children have inherited disorders with germ-line mutations of genes that predispose to the development of medulloblastoma. The Gorlin syndrome, the most common of these inherited disorders, is due to mutations in the patched-1 (PTCH-1) gene, a key component in the sonic hedgehog pathway. Turcot's syndrome, caused by mutations in the adenomatous polyposis coli (APC) gene and familial adenomatous

polyposis, has also been associated with an increased incidence of medulloblastoma. Histologically, medulloblastomas appear as highly cellular tumors with abundant dark staining, round nuclei, and rosette formation (Homer-Wright rosettes). They present with headache, ataxia, and signs of brainstem involvement. On MRI they appear as densely enhancing tumors in the posterior fossa, sometimes associated with hydrocephalus. Seeding of the CSF is common. Treatment involves maximal surgical resection, craniospinal irradiation, and chemotherapy with agents such as cisplatin, lomustine, cyclophosphamide, and vincristine. Approximately 70% of patients have long-term survival but usually at the cost of significant neurocognitive impairment. A major goal of current research is to improve survival while minimizing long-term complications.

Pineal Region Tumors

A large number of tumors can arise in the region of the pineal gland. These typically present with headache, visual symptoms, and hydrocephalus. Patients may have Parinaud's syndrome characterized by impaired upgaze and accommodation. Some pineal tumors such as pineocytomas and benign teratomas can be treated simply by surgical resection. Germinomas respond to irradiation, while pineoblastomas and malignant germ cell tumors require craniospinal radiation and chemotherapy.

EXTRINSIC "BENIGN" TUMORS

Meningiomas

Meningiomas are diagnosed with increasing frequency as more people undergo neuroimaging studies for various indications. They are now the most common primary brain tumor, accounting for approximately 32% of the total. Their incidence increases with age. They tend to be more common in women and in patients with neurofibromatosis type 2. They also occur more commonly in patients with a past history of cranial irradiation.

Meningiomas arise from the dura mater and are composed of neoplastic meningothelial (arachnoidal cap) cells. They are most commonly located over the cerebral convexities, especially adjacent to the sagittal sinus, but can also occur in the skull base and along the dorsum of the spinal cord. Meningiomas are classified by the WHO into three histologic grades of increasing aggressiveness: grade I (benign meningiomas), grade II (atypical meningiomas), and grade III (malignant meningiomas).

Many meningiomas are found incidentally following neuroimaging for unrelated reasons. They can also present with headaches, seizures, or focal neurologic deficits. On imaging studies they have a characteristic appearance usually consisting of a partially calcified, densely enhancing extraaxial tumor arising from the dura **(Fig. 379-6)**. Occasionally they may have a dural tail, consisting of thickened, enhanced dura extending like a tail from the mass. The main differential diagnosis of meningioma is a dural metastasis.

Figure 379-6

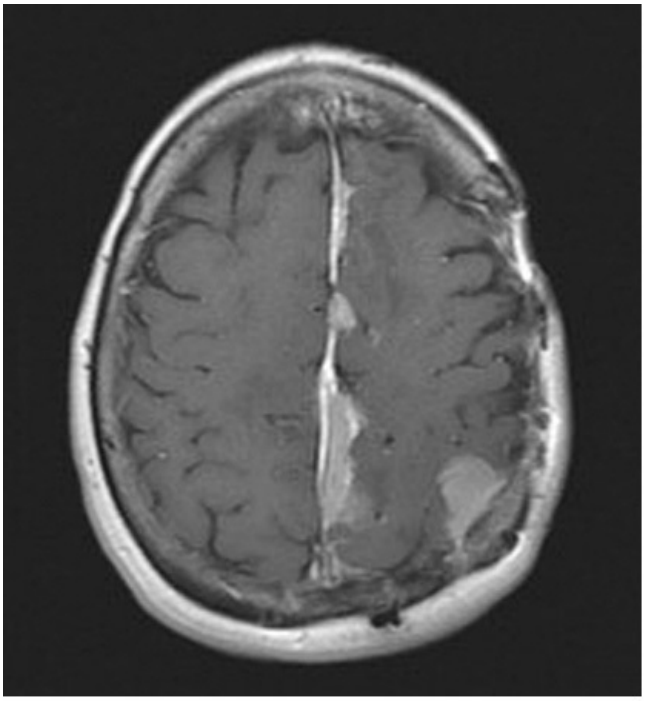

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium T1 MRI demonstrating multiple meningiomas along the falx and left parietal cortex.

If the meningioma is small and asymptomatic, no intervention is necessary and the lesion can be observed with serial MRI studies. Larger, symptomatic lesions should be resected surgically. If complete resection is achieved, the patient is cured. Incompletely resected tumors tend to recur, although the rate of recurrence can be very slow with grade I tumors. Tumors that cannot be resected, or can only be partially removed, may benefit from treatment with external beam radiotherapy or stereotactic radiosurgery (SRS). These treatments may also be helpful in patients whose tumor has recurred after surgery. Hormonal therapy and chemotherapy are currently unproven.

Rarer tumors that resemble meningiomas include hemangiopericytomas and solitary fibrous tumors. These are treated with surgery and radiotherapy but have a higher propensity to recur.

Schwannomas

These are generally benign tumors arising from the Schwann cells of cranial and spinal nerve roots. The most common schwannomas, termed *vestibular schwannomas* or *acoustic neuromas*, arise from the vestibular portion of the eighth cranial nerve and account for approximately 9% of primary brain tumors. Patients with neurofibromatosis type 2 have a high incidence of vestibular schwannomas that are frequently bilateral. Schwannomas arising from other cranial nerves, such as the trigeminal nerve (cranial nerve V), occur with much lower frequency. Neurofibromatosis type 1 is associated with an increased incidence of schwannomas of the spinal nerve roots.

Vestibular schwannomas may be found incidentally on neuroimaging or present with progressive unilateral hearing loss, dizziness, tinnitus, or less commonly, symptoms resulting from compression of the brainstem and cerebellum. On MRI they appear as densely enhancing lesions, enlarging the internal auditory canal and often extending into the cerebellopontine angle **(Fig. 379-7)**. The differential diagnosis includes meningioma. Very small, asymptomatic lesions can be observed with serial MRIs. Larger lesions should be treated with surgery or stereotactic radiosurgery. The optimal treatment will depend on the size of the tumor, symptoms, and the patient's preference. In patients with small vestibular schwannomas and relatively intact hearing, early surgical intervention increases the chance of preserving hearing.

Figure 379-7

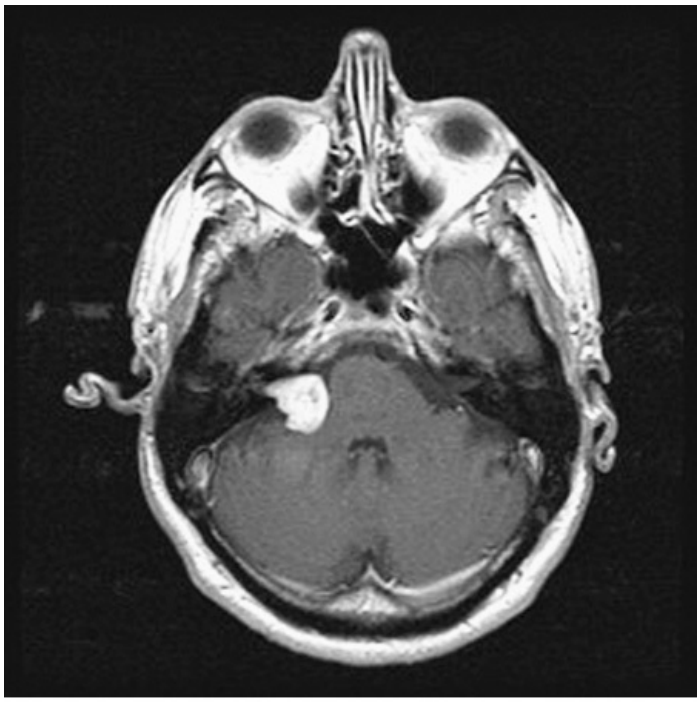

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium MRI of a right vestibular schwannoma. The tumor can be seen to involve the internal auditory canal.

Pituitary Tumors (Chap. 339)

These account for approximately 9% of primary brain tumors. They can be divided into functioning and nonfunctioning tumors. Functioning tumors are usually microadenomas (<1 cm in diameter) that

secrete hormones and produce specific endocrine syndromes [e.g., acromegaly for growth hormone –secreting tumors, Cushing's syndrome for adrenocorticotropic hormone (ACTH)-secreting tumors, and galactorrhea, amenorrhea, and infertility for prolactin-secreting tumors]. Nonfunctioning pituitary tumors tend to be macroadenomas (>1 cm) that produce symptoms by mass effect, giving rise to headaches, visual impairment (such as bitemporal hemianopia), and hypopituitarism. Prolactin-secreting tumors respond well to dopamine agonists such as bromocriptine and cabergoline. Other pituitary tumors usually require treatment with surgery and sometimes radiotherapy or radiosurgery and hormonal therapy.

Craniopharyngiomas

Craniopharyngiomas are rare, usually suprasellar, partially calcified, solid, or mixed solid-cystic benign tumors that arise from remnants of Rathke's pouch. They have a bimodal distribution, occurring predominantly in children but also between the ages of 55 and 65 years. They present with headaches, visual impairment, and impaired growth in children and hypopituitarism in adults. Treatment involves surgery, radiotherapy, or the combination of the two.

OTHER BENIGN TUMORS

DYSEMBRYOPLASTIC NEUROEPITHELIAL TUMORS (DNTS)

These are benign, supratentorial tumors, usually in the temporal lobes. They typically occur in children and young adults with a long-standing history of seizures. If the seizures are refractory, surgical resection is curative.

EPIDERMOID CYSTS

These consist of squamous epithelium surrounding a keratin-filled cyst. They are usually found in the cerebellopontine angle and the intrasellar and suprasellar regions. They may present with headaches, cranial nerve abnormalities, seizures, or hydrocephalus. Imaging studies demonstrate extraaxial lesions with characteristics that are similar to CSF but have restricted diffusion. Treatment involves surgical resection.

DERMOID CYSTS

Like epidermoid cysts, dermoid cysts arise from epithelial cells that are retained during closure of the neural tube. They contain both epidermal and dermal structures such as hair follicles, sweat glands, and sebaceous glands. Unlike epidermoid cysts, these tumors usually have a midline location. They occur most frequently in the posterior fossa, especially the vermis, fourth ventricle, and suprasellar cistern. Radiographically, dermoid cysts resemble lipomas, demonstrating T1 hyperintensity and variable signal on T2. Symptomatic dermoid cysts can be treated with surgery.

COLLOID CYSTS

These usually arise in the anterior third ventricle and may present with headaches, hydrocephalus, and very rarely sudden death. Surgical resection is curative or a third ventriculostomy may relieve the obstructive hydrocephalus and be sufficient therapy.

NEUROCUTANEOUS SYNDROMES (PHAKOMATOSES)

A number of genetic disorders are characterized by cutaneous lesions and an increased risk of brain tumors. Most of these disorders have an autosomal dominance inheritance with variable penetrance.

Neurofibromatosis Type 1 (NF1) (Von Recklinghausen's Disease)

NF1 is an autosomal dominant disorder with an incidence of approximately 1 in 2600–3000. Approximately half the cases are familial; the remainder are new mutations arising in patients with unaffected parents. The NF1 gene on chromosome 17q11.2 encodes a protein, *neurofibromin*, a guanosine triphosphatase (GTPase)-activating protein (GAP) that modulates signaling through the ras pathway. Mutations of the NF1 gene result in a large number of nervous system tumors including neurofibromas, plexiform neurofibromas, optic nerve gliomas, astrocytomas, and meningiomas. In addition to neurofibromas, which appear as multiple, soft, rubbery cutaneous tumors, other cutaneous manifestations of NF1 include café au lait spots and axillary freckling. NF1 is also associated with hamartomas of the iris termed Lisch nodules, pheochromocytomas, pseudoarthrosis of the tibia, scoliosis, epilepsy, and mental retardation.

Neurofibromatosis Type 2 (NF2)

NF2 is less common than NF1, with an incidence of 1 in 25,000–40,000. It is an autosomal dominant disorder with full penetrance. As with NF1, approximately half the cases arise from new mutations. The NF2 gene on 22q encodes a cytoskeletal protein "merlin" (moesin, ezrin, radixin-like protein) that functions as a tumor suppressor. NF2 is characterized by bilateral vestibular schwannomas in over 90% of patients, multiple meningiomas, and spinal ependymomas and astrocytomas. Treatment of bilateral vestibular schwannomas can be challenging because the goal is to preserve hearing for as long as possible. These patients may also have posterior subcapsular lens opacities and retinal hamartomas.

Tuberous Sclerosis (Bourneville's Disease)

This is an autosomal dominant disorder with an incidence of approximately 1 in 5000 to 10,000 live births. It is caused by mutations in either the TSC1 gene, which maps to chromosome 9q34, and encodes a protein termed *hamartin*, or mutations in the TSC2 gene, which maps to chromosome 16p13.3 and encodes the tuberin protein. Hamartin forms a complex with tuberin, which inhibits cellular signaling through the mammalian target of rapamycin (mTOR), and acts as a negative regulator of the cell cycle. Patients with tuberous sclerosis have seizures, mental retardation, adenoma sebaceum (facial angiofibromas), shagreen patch, hypomelanotic macules, periungual fibromas, renal angiomyolipomas, and cardiac rhabdomyomas. These patients have an increased incidence of subependymal nodules, cortical tubers, and subependymal giant cell astrocytomas (SEGA). Patients frequently require anticonvulsants for seizures. SEGAs often do not need treatment but occasionally require surgical resection. There is emerging evidence that mTOR inhibitors may have activity in SEGAs.

TUMORS METASTATIC TO THE BRAIN

Brain metastases arise from hematogenous spread and frequently arise from either a lung primary or are associated with pulmonary metastases. Most metastases develop at the gray matter–white matter junction in the watershed distribution of the brain where intravascular tumor cells lodge in terminal arterioles. The distribution of metastases in the brain approximates the proportion of blood flow such that about 85% of all metastases are supratentorial and 15% occur in the posterior fossa. The most common sources of brain metastases are lung and breast carcinomas; melanoma has the greatest propensity to metastasize to the brain, being found in 80% of patients at autopsy **(Table 379-3)**. Other tumor types such as ovarian and esophageal carcinoma rarely metastasize to the brain. Prostate and breast cancer also have a propensity to metastasize to the dura and can mimic meningioma. Leptomeningeal metastases are common from hematologicmalignancies and also breast and lung cancers. Spinal cord compression primarily arises in patients with prostate and breast cancer, tumors with a strong propensity to metastasize to the axial skeleton.

Table 379-3 Frequency of Nervous System Metastases by Common Primary Tumors

	Brain %	LM %	ESCC %
Lung	41	17	15
Breast	19	58	22
Melanoma	10	12	4
Prostate	1	1	10
GIT	7	—	5
Renal	3	2	7
Lymphoma	<1	10	10
Sarcoma	7	1	9
Other	11	—	18

Abbreviations: ESCC, epidural spinal cord compression; GIT, gastrointestinal tract; LM, leptomeningeal metastases.

Diagnosis of Metastases

Brain metastases are best visualized on MRI, where they usually appear as well-circumscribed lesions **(Fig. 379-8)**. The amount of perilesional edema can be highly variable with large lesions causing minimal edema and sometimes very small lesions causing extensive edema. Enhancement may be in a ring pattern or diffuse. Occasionally, intracranial metastases will hemorrhage; although melanoma, thyroid, and kidney cancer have the greatest propensity to hemorrhage, the most common cause of a hemorrhagic metastasis is lung cancer because it accounts for the majority of brain metastases. The radiographic appearance of brain metastasis is nonspecific, and similar appearing lesions can occur with infection including brain abscesses and also with demyelinating lesions, sarcoidosis, radiation necrosis in a previously treated patient, or a primary brain tumor that may be a second malignancy in a patient with systemic cancer. However, biopsy is rarely necessary for diagnosis in most patients because imaging alone in the appropriate clinical situation usually suffices. This is straightforward for the majority of patients with brain metastases because they have a known systemic cancer. However, in approximately 10% of patients a systemic cancer may present with a brain metastasis, and if there is not an easily accessible systemic site to biopsy, then a brain lesion must be removed for diagnostic purposes.

Figure 379-8

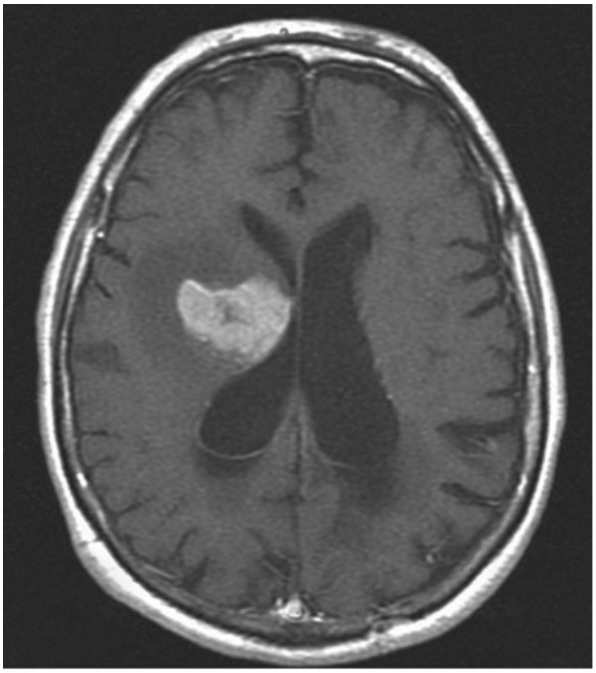

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

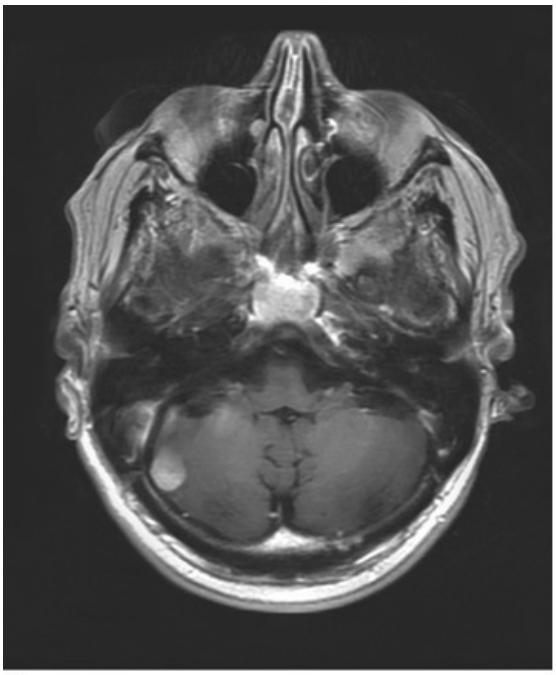

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium T1 MRI of multiple brain metastases from non-small cell lung cancer involving the right frontal (***A***) and right cerebellar (***B***) hemispheres. Note the diffuse enhancement pattern and absence of central necrosis.

Treatment: Tumors Metastatic to the Brain

DEFINITIVE TREATMENT

The number and location of brain metastases often determine the therapeutic options. The patient's overall condition and the current or potential control of the systemic disease are also major determinants. Brain metastases are single in approximately one-half of patients and multiple in the other half.

RADIATION THERAPY

The standard treatment for brain metastases has been whole-brain radiotherapy (WBRT) usually

administered to a total dose of 3000 cGy in 10 fractions. This affords rapid palliation, and approximately 80% of patients improve with glucocorticoids and radiation therapy. However, it is not curative. Median survival is only 4–6 months. More recently, stereotactic radiosurgery (SRS) delivered through a variety of techniques including the gamma knife, linear accelerator, proton beam, and CyberKnife all can deliver highly focused doses of RT, usually in a single fraction. SRS can effectively sterilize the visible lesions and afford local disease control in 80–90% of patients. In addition, there are some patients who have clearly been cured of their brain metastases using SRS, whereas this is distinctly rare with WBRT. However, SRS can be used only for lesions 3 cm or less in diameter and should be confined to patients with only 1–3 metastases. The addition of WBRT to SRS improves disease control in the nervous system but does not prolong survival.

SURGERY

Randomized controlled trials have demonstrated that surgical extirpation of a single brain metastasis followed by WBRT is superior to WBRT alone. Removal of two lesions or a single symptomatic mass, particularly if compressing the ventricular system, can also be useful. This is particularly useful in patients who have highly radioresistant lesions such as renal carcinoma. Surgical resection can afford rapid symptomatic improvement and prolonged survival. RT administered after complete resection of a brain metastasis improves disease control but does not prolong survival.

CHEMOTHERAPY

Chemotherapy is rarely useful for brain metastases. Metastases from certain tumor types that are highly chemosensitive, such as germ cell tumors or small cell lung cancer, may respond to chemotherapeutic regimens chosen according to the underlying malignancy. Increasingly, there are data demonstrating responsiveness of brain metastases to chemotherapy including small molecule –targeted therapy when the lesion possesses the target. This has been best illustrated in patients with lung cancer harboring EGFR mutations that sensitize them to EGFR inhibitors. Antiangiogenic agents such as bevacizumab may also prove efficacious in the treatment of CNS metastases.

LEPTOMENINGEAL METASTASES

Leptomeningeal metastases are also identified as carcinomatous meningitis, meningeal carcinomatosis, or in the case of specific tumors, leukemic or lymphomatous meningitis. Among the hematologicmalignancies, acute leukemia is the most common to metastasize to the subarachnoid space, and in lymphomas the aggressive diffuse lymphomas can metastasize to the subarachnoid space frequently as well. Among solid tumors, breast and lung carcinomas and melanoma most frequently spread in this fashion. Tumor cells reach the subarachnoid space via the arterial circulation or occasionally through retrograde flow in venous systems that drain metastases along the bony spine or cranium. In addition, leptomeningeal metastases may develop as a direct consequence of prior brain metastases and can develop in almost 40% of patients who have a metastasis resected from the cerebellum.

Clinical Features

Leptomeningeal metastases are characterized clinically by multilevel symptoms and signs along the neuraxis. Combinations of lumbar and cervical radiculopathies, cranial neuropathies, seizures, confusion, and encephalopathy from hydrocephalus or raised intracranial pressure can be present. Focal deficits such as hemiparesis or aphasia are rarely due to leptomeningeal metastases unless there is direct brain infiltration and are more often associated with coexisting brain lesions. New onset limb pain in patients with breast, lung cancer, or melanoma should prompt consideration of leptomeningeal spread.

Laboratory and Imaging Diagnosis

Leptomeningeal metastases are particularly challenging to diagnose as identification of tumor cells in the subarachnoid compartment may be elusive. MR imaging can be definitive in patients when there are clear tumor nodules adherent to the cauda equina or spinal cord, enhancing cranial nerves, or subarachnoid enhancement on brain imaging **(Fig. 379-9)**. Imaging is diagnostic in approximately 75% of patients and is more often positive in patients with solid tumors. Demonstration of tumor cells in the CSF is definitive and often considered the gold standard. However, CSF cytologic examination is positive in only 50% of patients on the first lumbar puncture and still misses 10% after three CSF samples. CSF cytologic examination is most useful in hematologicmalignancies. Accompanying CSF abnormalities include an elevated protein concentration and an elevated white count. Hypoglycorrhachia is noted in less than 25% of patients but is useful when present. Identification of tumor markers or molecular confirmation of clonal proliferation with techniques such as flow cytometry within the CSF can also be definitive when present. Tumor markers are usually specific to solid tumors, and chromosomal or molecular markers are most useful in patients with hematologic malignancies.

Figure 379-9

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium MRI images of extensive leptomeningeal metastases from breast cancer. Nodules along the dorsal surface of the spinal cord (***A***) and cauda equina (***B***) are seen.

Treatment: Leptomeningeal Metastases

The treatment of leptomeningeal metastasis is palliative as there is no curative therapy. RT to the symptomatically involved areas, such as skull base for cranial neuropathy, can relieve pain and sometimes improve function. Whole neuraxis RT has extensive toxicity with myelosuppression and gastrointestinal irritation as well as limited effectiveness. Systemic chemotherapy with agents that can penetrate the blood-CSF barrier may be helpful. Alternatively, intrathecal chemotherapy can be effective, particularly in hematologicmalignancies. This is optimally delivered through an intraventricular cannula (Ommaya reservoir) rather than by lumbar puncture. Few drugs can be delivered safely into the subarachnoid space and they have a limited spectrum of antitumor activity, perhaps accounting for the relatively poor response to this approach. In addition, impaired CSF flow

dynamics can compromise intrathecal drug delivery. Surgery has a limited role in the treatment of leptomeningeal metastasis, but placement of a ventriculoperitoneal shunt can relieve raised intracranial pressure. However, it compromises delivery of chemotherapy into the CSF.

EPIDURAL METASTASIS

Epidural metastasis occurs in 3–5% of patients with a systemic malignancy and causes neurologic compromise by compressing the spinal cord or cauda equina. The most common cancers that metastasize to the epidural space are those malignancies that spread to bone, such as breast and prostate. Lymphoma can cause bone involvement and compression but it can also invade the intervertebral foramens and cause spinal cord compression without bone destruction. The thoracic spine is affected most commonly, followed by the lumbar and then cervical spine.

Clinical Features

Back pain is the presenting symptom of epidural metastasis in virtually all patients; the pain may precede neurologic findings by weeks or months. The pain is usually exacerbated by lying down; by contrast, arthritic pain is often relieved by recumbency. Leg weakness is seen in about 50% of patients as is sensory dysfunction. Sphincter problems are present in about 25% of patients at diagnosis.

Diagnosis

Diagnosis is established by imaging, with MRI of the complete spine being the best test **(Fig. 379-10)**. Contrast is not needed to identify spinal or epidural lesions. Any patient with cancer who has severe back pain should undergo an MRI. Plain films, bone scans, or even CT scans may show bone metastases, but only MRI can reliably delineate epidural tumor. For patients unable to have an MRI, CT myelography should be performed to outline the epidural space. The differential diagnosis of epidural tumor includes epidural abscess, acute or chronic hematomas, and rarely, extramedullary hematopoiesis.

Figure 379-10

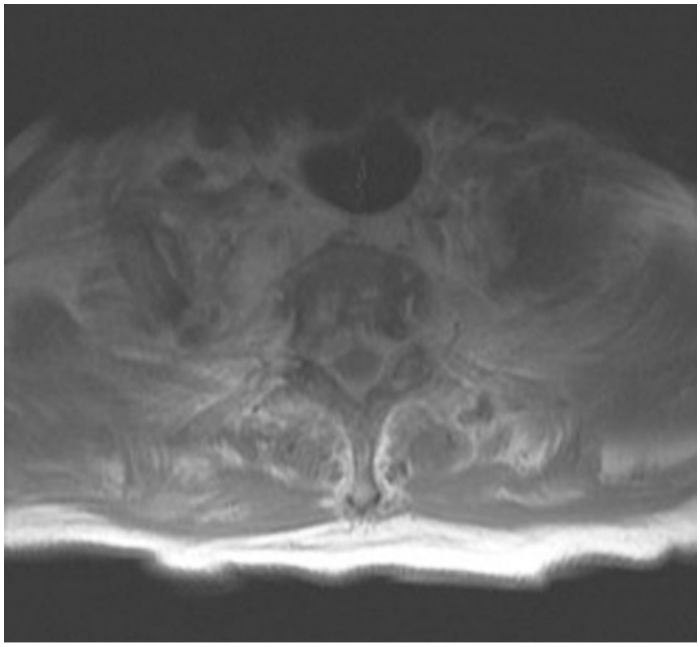

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postgadolinium T_1 MRI showing circumferential epidural tumor around the thoracic spinal cord from esophageal cancer.

Treatment: Epidural Metastasis

Epidural metastasis requires immediate treatment. A randomized controlled trial demonstrated the superiority of surgical resection followed by RT compared to RT alone. However, patients must be able to tolerate surgery, and the surgical procedure of choice is a complete removal of the mass, which is typically anterior to the spinal canal, necessitating an extensive approach and resection. Otherwise, RT is the mainstay of treatment and can be used for patients with radiosensitive tumors, such as lymphoma, or for those unable to undergo surgery. Chemotherapy is rarely used for epidural metastasis unless the patient has minimal to no neurologic deficit and a highly chemosensitive tumor such as lymphoma or germinoma. Patients generally fare well if treated before there is severe

neurologic deficit. Recovery after paraparesis is better after surgery than with RT alone, but survival is often short due to widespread metastatic tumor.

NEUROLOGIC TOXICITY OF THERAPY

Toxicity from Radiotherapy

Radiotherapy can cause a variety of toxicities in the CNS. These are usually described based on their relationship in time to the administration of RT, e.g., they can be acute (occurring within days of RT), early delayed (months), or late delayed (years). In general, the acute and early delayed syndromes resolve and do not result in persistent deficits, whereas the late delayed toxicities are usually permanent and sometimes progressive.

ACUTE TOXICITY

Acute cerebral toxicity usually occurs during RT to the brain. RT can cause a transient disruption of the blood-brain barrier, resulting in increased edema and elevated intracranial pressure. This is usually manifest as headache, lethargy, nausea and vomiting, and can be both prevented and treated with the administration of glucocorticoids. There is no acute RT toxicity that affects the spinal cord.

EARLY DELAYED TOXICITY

Early delayed toxicity is usually apparent weeks to months after completion of cranial irradiation and is likely due to focal demyelination. Clinically it may be asymptomatic or take the form of worsening or reappearance of a preexisting neurologic deficit. At times a contrast-enhancing lesion can be seen on MRI/CT that can mimic the tumor for which the patient received the RT. For patients with a malignant glioma, this has been described as "pseudoprogression" because it mimics tumor recurrence on MRI but actually represents inflammation and necrotic debris engendered by effective therapy. This is seen with increased frequency when chemotherapy, particularly temozolomide, is given concurrently with RT. Pseudoprogression can resolve on its own or, if very symptomatic, may require resection. A rare form of early delayed toxicity is the somnolence syndrome that occurs primarily in children and is characterized by marked sleepiness.

In the spinal cord, early delayed RT toxicity is manifest as a Lhermitte symptom with paresthesias of the limbs or along the spine when the patient flexes the neck. Although frightening, it is benign, resolves on its own, and does not portend more serious problems.

LATE DELAYED TOXICITY

Late delayed toxicities are the most serious as they are often irreversible and cause severe neurologic deficits. In the brain, late toxicities can take several forms, the most common of which include radiation necrosis and leukoencephalopathy. Radiation necrosis is a focal mass of necrotic tissue that is contrast enhancing on CT/MRI and may be associated with significant edema. This may appear identical to pseudoprogression but is seen months to years after RT and is always symptomatic. Clinical symptoms and signs include seizure and lateralizing findings referable to the location of the necrotic mass. The necrosis is caused by the effect of RT on cerebral vasculature with resultant fibrinoid necrosis and occlusion of the blood vessels. It can mimic tumor radiographically, but unlike tumor it is typically hypometabolic on a PET scan and has reduced perfusion on perfusion MR sequences. It may require resection for diagnosis and treatment unless it can be managed with glucocorticoids. There are rare reports of improvement with hyperbaric oxygen or anticoagulation but the usefulness of these approaches is questionable.

Leukoencephalopathy is seen most commonly after WBRT as opposed to focal RT. On T2 or FLAIR MR sequences there is diffuse increased signal seen throughout the hemispheric white matter, often bilaterally and symmetrically. There tends to be a periventricular predominance that may be associated with atrophy and ventricular enlargement. Clinically, patients develop cognitive impairment, gait disorder, and later urinary incontinence, all of which can progress over time. These symptoms mimic those of normal pressure hydrocephalus, and placement of a ventriculoperitoneal shunt can improve function in some patients but does not reverse the deficits completely. Increased age is a risk factor for leukoencephalopathy but not for radiation necrosis. Necrosis appears to depend on an as yet unidentified predisposition.

Other late neurologic toxicities include endocrine dysfunction if the pituitary or hypothalamus was included in the RT port. A radiation-induced neoplasm can occur many years after therapeutic RT for either a prior CNS tumor or a head and neck cancer; accurate diagnosis requires surgical resection or biopsy. In addition, RT causes accelerated atherosclerosis, which can cause stroke either from intracranial vascular disease or carotid plaque from neck irradiation.

The peripheral nervous system is relatively resistant to RT toxicities. Peripheral nerves are rarely affected by RT, but the plexus is more vulnerable. Plexopathy develops more commonly in the brachial distribution than in the lumbosacral distribution. It must be differentiated from tumor progression in the plexus, which is usually accomplished with CT/MR imaging of the area or PET scan demonstrating tumor infiltrating the region. Clinically, tumor progression is usually painful whereas radiation-induced plexopathy is painless. Radiation plexopathy is also more commonly associated with lymphedema of

the affected limb. Sensory loss and weakness are seen in both.

TOXICITY FROM CHEMOTHERAPY

Neurotoxicity is second to myelosuppression as the dose-limiting toxicity of chemotherapeutic agents **(Table 379-4)**. Chemotherapy causes peripheral neuropathy from a number of commonly used agents, and the type of neuropathy can differ, depending upon the drug. Vincristine causes paresthesias but little sensory loss and is associated with motor dysfunction, autonomic impairment (frequently ileus), and rarely cranial nerve compromise. Cisplatin causes large fiber sensory loss resulting in sensory ataxia but little cutaneous sensory loss and no weakness. The taxanes also cause a predominately sensory neuropathy. Agents such as bortezomib and thalidomide also cause neuropathy.

Table 379-4 Neurologic Signs Caused by Agents Commonly Used in Patients with Cancer

Acute encephalopathy (delirium)	Seizures
Methotrexate (high-dose IV, IT)	Methotrexate
Cisplatin	Etoposide (high-dose)
Vincristine	Cisplatin
Asparaginase	Vincristine
Procarbazine	Asparaginase
5-Flourouracil (±levamisole)	Nitrogen mustard
Cytarabine (high-dose)	Carmustine
Nitrosoureas (high-dose or arterial)	Dacarbazine (intraarterial or high-dose)
Ifosfamide	Busulfan (high-dose)
Etoposide (high-dose)	Myelopathy (intrathecal drugs)
Bevacizumab (PRES)	Methotrexate
Chronic encephalopathy (dementia)	Cytarabine
Methotrexate	Thiotepa
Carmustine	Peripheral neuropathy
Cytarabine	Vinca alkaloids
Fludarabine	Cisplatin
Visual loss	Procarbazine
Tamoxifen	Etoposide
Gallium nitrate	Teniposide
Cisplatin	Cytarabine
Fludarabine	Taxanes
Cerebellar dysfunction/ataxia	Suramin
5-Fluorouracil (± levamisole)	Bortezomib
Cytarabine	
Procarbazine	

Abbreviations: IT, intrathecal; IV, intravenous; PRES, posterior reversible encephalopathy syndrome.

Encephalopathy and seizures are common toxicities from chemotherapeutic drugs. Ifosfamide can cause a severe encephalopathy, which is reversible with discontinuation of the drug and the use of methylene blue for severely affected patients. Fludarabine also causes a severe global encephalopathy that may be permanent. Bevacizumab and other anti-VEGF agents can cause posterior reversible encephalopathy syndrome. Cisplatin can cause hearing loss and less frequently vestibular dysfunction.

FURTHER READINGS

Clarke JL et al: Leptomeningeal metastases in the MRI era. Neurology 76:200, 2011

DeAngelis LM, Posner JB (eds): *Neurologic Complications of Cancer* , 2nd ed, New York, Oxford University Press, 2009

Elliott RE et al: Neurological complications and symptom resolution following Gamma Knife surgery for brain metastases 2 cm or smaller in relation to eloquent cortices. J Neurosurg: 113 Suppl:53, 2010

Friedman HS et al: Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. J Clin Oncol 27:4733, 2009[PMID: 19720927] [Full Text]

Lu-Emerson C, Plotkin SR: The neurofibromatoses. Part 1: NF1. Rev Neurol Dis 6:E47, 2009

———, ———: The neurofibromatoses. Part 2: NF2 and schwannomatosis. Rev Neurol Dis 6:E81, 2009

Rades D, Abrahm JL: The role of radiotherapy for metastatic epidural spinal cord compression. Nat Rev Clin Oncol 7:590, 2010[PMID: 20808299] [Full Text]

Stupp R et al: Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomized phase III study: 5-Year analysis of the EORTC-NCIC trial. Lancet Oncol 10:459, 2009[PMID: 19269895] [Full Text]

Wang M et al: Cognition and quality of life after chemotherapy plus radiotherapy (RT) vs. RT for pure and mixed anaplastic oligodendrogliomas: Radiation Therapy Oncology Group Trial 9402. Int J Radiat Oncol Biol Phys 77:662, 2010[PMID: 19783377] [Full Text]

Wen PY, Kesari S: Malignant gliomas in adults. N Engl J Med 359:492, 2008[PMID: 18669428] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MULTIPLE SCLEROSIS AND OTHER DEMYELINATING DISEASES: INTRODUCTION

Demyelinating disorders are immune-mediated conditions characterized by preferential destruction of central nervous system (CNS) myelin. The peripheral nervous system (PNS) is spared, and most patients have no evidence of an associated systemic illness. Multiple sclerosis (MS), the most common disease in this category, is second only to trauma as a cause of neurologic disability beginning in early to middle adulthood.

MULTIPLE SCLEROSIS

Multiple sclerosis (MS) is a chronic disease characterized by inflammation, demyelination, gliosis (scarring), and neuronal loss; the course can be relapsing-remitting or progressive. Lesions of MS typically occur at different times and in different CNS locations (i.e., disseminated in time and space). MS affects ~350,000 individuals in the United States and 2.5 million individuals worldwide. Manifestations of MS vary from a benign illness to a rapidly evolving and incapacitating disease requiring profound lifestyle adjustments.

Pathogenesis

ANATOMY

New MS lesions begin with perivenular cuffing by inflammatory mononuclear cells, predominantly T cells and macrophages, which also infiltrate the surrounding white matter. At sites of inflammation, the blood-brain barrier (BBB) is disrupted, but unlike vasculitis, the vessel wall is preserved. Involvement of the humoral immune system is also evident; small numbers of B lymphocytes also infiltrate the nervous system, and myelin-specific autoantibodies are present on degenerating myelin sheaths. As lesions evolve, there is prominent astrocytic proliferation (gliosis). Surviving oligodendrocytes or those that differentiate from precursor cells can partially remyelinate the surviving naked axons, producing so-called shadow plaques. In many lesions, oligodendrocyte precursor cells are present in large numbers but fail to differentiate and remyelinate. Over time, ectopic lymphocyte follicles appear in perivascular and perimeningeal regions, consisting of aggregates of T and B cells and resembling secondary lymphoid structures. Although relative sparing of axons is typical of MS, partial or total axonal destruction can also occur, especially within highly inflammatory lesions. Thus, MS is not solely a disease of myelin, and neuronal pathology is increasingly recognized as a major contributor to irreversible neurologic disability. Inflammation and plaque formation are present in the cerebral cortex, and significant axon loss indicating death of neurons is widespread, specially in advanced cases (see "Neurodegeneration," below).

PHYSIOLOGY

Nerve conduction in myelinated axons occurs in a saltatory manner, with the nerve impulse jumping from one node of Ranvier to the next without depolarization of the axonal membrane underlying the myelin sheath between nodes (Fig. 380-1). This produces considerably faster conduction velocities (~ 70 m/s) than the slow velocities (~1 m/s) produced by continuous propagation in unmyelinated nerves. Conduction block occurs when the nerve impulse is unable to traverse the demyelinated segment. This can happen when the resting axon membrane becomes hyperpolarized due to the exposure of voltage-dependent potassium channels that are normally buried underneath the myelin sheath. A temporary conduction block often follows a demyelinating event before sodium channels (originally concentrated at the nodes) redistribute along the naked axon (Fig. 380-1). This redistribution ultimately allows continuous propagation of nerve action potentials through the demyelinated segment. Conduction block may be incomplete, affecting high- but not low-frequency volleys of impulses. Variable conduction block can occur with raised body temperature or metabolic alterations and may explain clinical fluctuations that vary from hour to hour or appear with fever or exercise. Conduction slowing occurs when the demyelinated segments of the axonal membrane is reorganized to support continuous (slow) nerve impulse propagation.

Figure 380-1

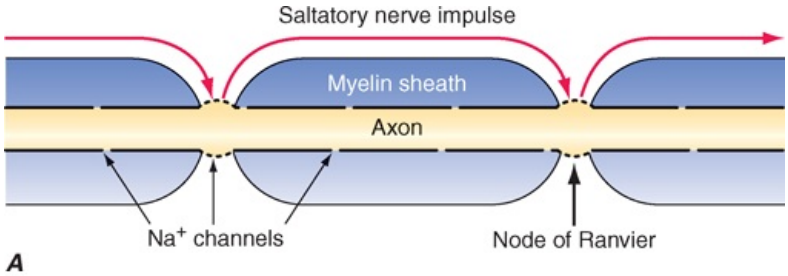

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

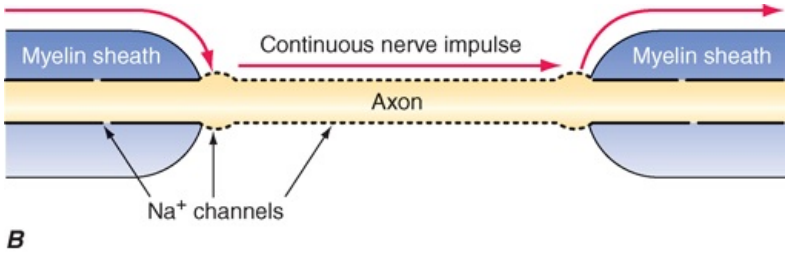

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Nerve conduction in myelinated and demyelinated axons. *A.* Saltatory nerve conduction in myelinated axons occurs with the nerve impulse jumping from one node of Ranvier to the next. Sodium channels (shown as breaks in the solid black line) are concentrated at the nodes where axonal depolarization occurs. ***B.*** Following demyelination, additional sodium channels are redistributed along the axon itself, thereby allowing continuous propagation of the nerve action potential despite the absence of myelin.

EPIDEMIOLOGY

MS is approximately threefold more common in women than men. The age of onset is typically between 20 and 40 years (slightly later in men than in women), but the disease can present across the life span. In ~10% of cases it begins before age 18 years and a in a small percentage it begins before the age of 10 years.

Geographic gradients have been repeatedly observed in MS, with the highest known prevalence for MS (250 per 100,000) in the Orkney Islands, located north of Scotland. In other temperate zone areas (e.g., northern North America, northern Europe, southern Australia, and south New Zealand), the prevalence of MS is 0.1–0.2%. By contrast, in the tropics (e.g., Asia, equatorial Africa, and the Middle East), the prevalence is often ten- to twentyfold less.

One proposed explanation for the latitude effect on MS is that there is a protective effect of sun exposure. Exposure of the skin to ultraviolet-B (UVB) radiation from the sun is essential for the biosynthesis of vitamin D, and this endogenous production is the most important source of vitamin D in most individuals. At high latitudes, the amount of UVB radiation reaching the earth's surface is often insufficient, particularly during winter months, and, consequently, low serum levels of vitamin D are common in temperate zones. Prospective studies have confirmed that vitamin D deficiency is associated with an increase in MS risk and preliminary data also suggest that ongoing deficiency may increase the relapse rate in established MS. Immunoregulatory effects of vitamin D could explain this apparent relationship.

At least three sequential (population-wide) environmental events are implicated in the causal pathway leading to MS. The first factor seems to occur either in utero or in the early postnatal period and is supported, in part, by the almost twofold increase in MS risk for dizygotic twins of MS probands (5.4%) compared to siblings (2.9%). It is also supported by the month-of-birth effect (in the northern hemisphere), in which May babies are significantly more likely, and November babies less likely, to develop MS compared to babies born in other months. Importantly, a recently published population-based study in the southern hemisphere (Australia) found a similar (but inverted) month-of-birth effect with the zenith in risk occurring for November/December babies and the nadir occurring for May/June babies. This month-of-birth effect provides evidence for an early environmental event, involved in MS pathogenesis, that is both coupled to the solar cycle and time-locked to birth.

A second factor seems to occur during adolescence. Thus, several studies suggest that when individuals move (prior to their adolescent years) from an area of high MS prevalence to an area of low prevalence (or vice versa), their MS risk becomes similar to that of the region to which they moved. By contrast, when they make the same move after adolescence, their MS risk remains similar to that of the region from which they moved.

Because both of these first two factors occur well before the onset of clinically evident MS, presumably

other factors are also necessary. In addition, the identification of possible point epidemics suggests a possible role for infectious agents, although the only (partially) convincing example of this occurred in the Faeroe Islands north of Denmark after the British occupation during World War II.

The prevalence of MS has increased steadily (and dramatically) in several regions around the world over the past half-century, presumably reflecting the impact of some environmental shift. Moreover, the fact that this increase has occurred primarily (or exclusively) in women indicates that women are more responsive to whatever this environmental change has been. Interestingly, recent epidemiologic data suggest that the latitude effect on MS currently may be decreasing. The reason for these changes are not known but, potentially, could be related to the increased use of sun block, which (at SPF-15) blocks 94% of the incoming UVB radiation, and which would be expected to exacerbate any population-wide vitamin D deficiency and might also mitigate the impact of differences in UVB exposure.

MS risk also correlates with high socioeconomic status, which may reflect improved sanitation and delayed initial exposures to infectious agents. By analogy, some viral infections (e.g., poliomyelitis and measles viruses) produce neurologic sequelae more frequently when the age of initial infection is delayed. Evidence of a remote Epstein-Barr virus (EBV) infection playing some role in MS is supported by a number of epidemiologic and laboratory studies. A higher risk of infectious mononucleosis (associated with relatively late EBV infection) and higher antibody titers to latency-associated EBV nuclear antigen are associated with MS. At this time, however, a causal role for EBV is not definitively established.

GENETIC CONSIDERATIONS

Whites are inherently at higher risk for MS than Africans or Asians, even when residing in a similar environment. MS also aggregates within some families, and adoption, half-sibling, twin, and spousal studies indicate that familial aggregation is due to genetic, and not environmental, factors (Table 380-1).

Table 380-1 Risk of Developing MS

1 in 3	If an identical twin has MS
1 in 15	If a fraternal twin has MS
1 in 25	If a sibling has MS
1 in 50	If a parent or half-sibling has MS
1 in 100	If a first cousin has MS
1 in 1000	If a spouse has MS
1 in 1000	If no one in the family has MS

Whites to MS is polygenic, with each gene contributing a relatively small amount to the overall risk. Despite this, the influence of genetics on MS pathogenesis is substantial. The major histocompatibility complex (MHC) on chromosome 6 is by far the strongest MS susceptibility region in the genome. Fine mapping studies implicate primarily the class II region (encoding HLA molecules involved in presenting peptide antigens to T cells) and specifically the highly polymorphic *DRB1* locus, which contributes to MS risk in a allele-dependent hierarchical fashion, with the strongest association consistently found with the *DRB1*15:01* allele; a secondary signal that appears to be protective against MS is located in the class 1 region near HLA-C. Whole-genome association studies have now identified more than 50 MS susceptibility genes, each of which has only a very small effect on MS risk. *DRB1*15:01* increases MS risk by approximately threefold in the heterozygous state, and ninefold in the homozygous state; by contrast, other MS-associated variants increase risk only by 15–30%. Most MS-associated genetic variants have known roles in the immune system [i.e., genes for the interleukin (IL)-7 receptor (CD127), the IL-2 receptor (CD25), and the T cell co-stimulatory molecule LFA-3 (CD58)]; some variants also influence susceptibility to other autoimmune diseases in addition to MS. The variants identified thus far all lack specificity and sensitivity for MS, thus they are not useful for diagnosis or to predict the future course of the disease.

IMMUNOLOGY

Autoreactive T Lymphocytes

Myelin basic protein (MBP) is an important T cell antigen in experimental allergic encephalomyelitis (EAE), a laboratory model, and probably also in human MS. Activated MBP-reactive T cells have been identified in the blood, in cerebrospinal fluid (CSF), and within MS lesions. Moreover, *DRB1*15:01* may influence the autoimmune response because it binds with high affinity to a fragment of MBP (spanning amino acids 89–96), stimulating T cell responses to this self-protein. Two different populations of proinflammatory T cells are likely to mediate autoimmunity in MS. T-helper type 1 (T_H1) cells producing interferon γ (IFN-γ) are one key effector population, and more recently a role for highly

proinflammatory T_H17 T cells has been established. T_H17 cells are induced by transforming growth factor β (TGF-β) and IL-6, and are amplified by IL-21 and IL-23. T_H17 cells, and levels of their corresponding cytokine IL-17, are increased in MS lesions and also in the circulation of people with active MS. High circulating levels of IL-17 may also be a marker of a more severe course of MS. T_H1 cytokines including interleukin (IL) 2, tumor necrosis factor (TNF) α, and interferon (IFN) γ also play key roles in activating and maintaining autoimmune responses, and TNF-α and IFN-γ may directly injure oligodendrocytes or the myelin membrane.

Humoral Autoimmunity

B cell activation and antibody responses also appear to be necessary for the full development of demyelinating lesions to occur, both in experimental models and in human MS. Increased numbers of clonally expanded B cells with properties of postgerminal center memory or antibody-producing lymphocytes are present in MS lesions and in CSF. Myelin-specific autoantibodies, some directed against myelin oligodendrocyte glycoprotein (MOG), have been detected bound to vesiculated myelin debris in MS plaques. In the CSF, elevated levels of locally synthesized immunoglobulins and oligoclonal antibodies derived from expansion of clonally restricted plasma cells are also characteristic of MS. The pattern of oligoclonal banding is unique to each individual, and attempts to identify the targets of these antibodies have been largely unsuccessful.

Triggers

Serial MRI studies in early relapsing-remitting MS reveal that bursts of focal inflammatory disease activity occur far more frequently than would have been predicted by the frequency of relapses. Thus, early in MS most disease activity is clinically silent. The triggers causing these bursts are unknown, although the fact that patients may experience relapses after nonspecific upper respiratory infections suggests that either molecular mimicry between viruses and myelin antigens or viral super-antigens activating pathogenic T cells may be responsible (Chap. 318).

NEURODEGENERATION

Axonal damage occurs in every newly formed MS lesion, and cumulative axonal loss is considered to be the major cause of progressive and irreversible neurologic disability in MS. As many as 70% of axons are lost from the lateral corticospinal (e.g., motor) tracts in patients with advanced paraparesis from MS, and longitudinal MRI studies suggest there is progressive axonal loss over time within established, inactive lesions. Knowledge of the mechanisms responsible for axonal injury is incomplete and, despite the fact that axonal transactions are most conspicuous in acute inflammatory lesions, it is still unclear whether demyelination is a prerequisite for axonal injury in MS. Demyelination can result in reduced trophic support for axons, redistribution of ion channels, and destabilization of action potential membrane potentials. Axons can adapt initially to these injuries; with time distal and retrograde degeneration often occurs. Therefore, promotion of remyelination and preservation of oligodendrocytes early in the disease course remain important therapeutic goals in MS. Some evidence suggests that axonal damage is mediated directly by resident and invading inflammatory cells and their toxic products, in particular by microglia, macrophages, and CD8 T lymphocytes. Activated microglia are particularly likely to cause axonal injury through the release of NO and oxygen radicals and via glutamate, which is toxic to oligodendrocytes and neurons. Interestingly, NMDA (glutamate) receptors are expressed on naked axon membranes that have undergone demyelination, perhaps providing a mechanism for glutamate-mediated calcium entry and cell death.

Clinical Manifestations

The onset of MS may be abrupt or insidious. Symptoms may be severe or seem so trivial that a patient may not seek medical attention for months or years. Indeed, at autopsy, approximately 0.1% of individuals who were asymptomatic during life will be found, unexpectedly, to have pathologic evidence of MS. Similarly, in the modern era, an MRI scan obtained for an unrelated reason may show evidence of asymptomatic MS. Symptoms of MS are extremely varied and depend on the location and severity of lesions within the CNS (Table 380-2). Examination often reveals evidence of neurologic dysfunction, often in asymptomatic locations. For example, a patient may present with symptoms in one leg but signs in both.

Table 380-2 Initial Symptoms of MS

Symptom	Percent of Cases	Symptom	Percent of Cases
Sensory loss	37	Lhermitte's	3
Optic neuritis	36	Pain	3
Weakness	35	Dementia	2
Paresthesias	24	Visual loss	2
Diplopia	15	Facial palsy	1
Ataxia	11	Impotence	1

Vertigo	6	Myokymia	1
Paroxysmal attacks	4	Epilepsy	1
Bladder	4	Falling	1

Source: After WB Matthews et al, McAlpine's Multiple Sclerosis, New York, Churchill Livingstone, 1991.

Weakness of the limbs may manifest as loss of strength, speed, or dexterity, as fatigue, or a disturbance of gait. Exercise-induced weakness is a characteristic symptom of MS. The weakness is of the upper motor neuron type (Chap. 22) and is usually accompanied by other pyramidal signs such as spasticity, hyperreflexia, and Babinski's signs. Occasionally a tendon reflex may be lost (simulating a lower motor neuron lesion) if an MS lesion disrupts the afferent reflex fibers in the spinal cord (Fig. 22-2).

Figure 380-2

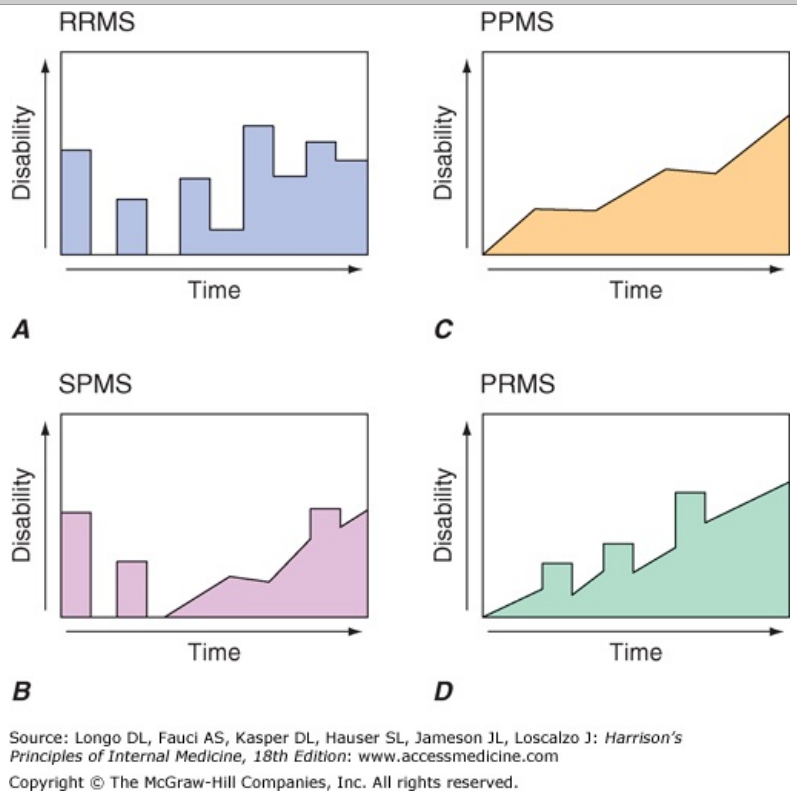

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Clinical course of multiple sclerosis (MS). ***A.*** Relapsing/remitting MS. ***B.*** Secondary progressive MS. ***C.*** Primary progressive MS.***D.*** Progressive/relapsing MS.

Spasticity (Chap. 22) is commonly associated with spontaneous and movement-induced muscle spasms. More than 30% of MS patients have moderate to severe spasticity, especially in the legs. This is often accompanied by painful spasms interfering with ambulation, work, or self-care. Occasionally spasticity provides support for the body weight during ambulation, and in these cases treatment of spasticity may actually do more harm than good.

Optic neuritis (ON) presents as diminished visual acuity, dimness, or decreased color perception (desaturation) in the central field of vision. These symptoms can be mild or may progress to severe visual loss. Rarely, there is complete loss of light perception. Visual symptoms are generally monocular but may be bilateral. Periorbital pain (aggravated by eye movement) often precedes or accompanies the visual loss. An afferent pupillary defect (Chap. 28) is usually present. Funduscopic examination may be normal or reveal optic disc swelling (papillitis). Pallor of the optic disc (optic atrophy) commonly follows ON. Uveitis is uncommon and should raise the possibility of alternative diagnoses such as sarcoid or lymphoma.

Visual blurring in MS may result from ON or diplopia (double vision); if the symptom resolves when either eye is covered, the cause is diplopia.

Diplopia may result from internuclear ophthalmoplegia (INO) or from palsy of the sixth cranial nerve (rarely the third or fourth). An INO consists of impaired adduction of one eye due to a lesion in the ipsilateral medial longitudinal fasciculus (Chap. 28). Prominent nystagmus is often observed in the abducting eye, along with a small skew deviation. A bilateral INO is particularly suggestive of MS. Other common gaze disturbances in MS include (1) a horizontal gaze palsy, (2) a "one and a half"

syndrome (horizontal gaze palsy plus an INO), and (3) acquired pendular nystagmus.

Sensory symptoms are varied and include both paresthesias (e.g., tingling, prickling sensations, formications, "pins and needles," or painful burning) and hypesthesia (e.g., reduced sensation, numbness, or a "dead" feeling). Unpleasant sensations (e.g., feelings that body parts are swollen, wet, raw, or tightly wrapped) are also common. Sensory impairment of the trunk and legs below a horizontal line on the torso (a sensory level) indicates that the spinal cord is the origin of the sensory disturbance. It is often accompanied by a bandlike sensation of tightness around the torso. Pain is a common symptom of MS, experienced by >50% of patients. Pain can occur anywhere on the body and can change locations over time.

Ataxia usually manifests as cerebellar tremors (Chap. 373). Ataxia may also involve the head and trunk or the voice, producing a characteristic cerebellar dysarthria (scanning speech).

Bladder dysfunction is present in >90% of MS patients, and in a third of patients, dysfunction results in weekly or more frequent episodes of incontinence. During normal reflex voiding, relaxation of the bladder sphincter (α-adrenergic innervation) is coordinated with contraction of the detrusor muscle in the bladder wall (muscarinic cholinergic innervation). *Detrusor hyperreflexia* , due to impairment of suprasegmental inhibition, causes urinary frequency, urgency, nocturia, and uncontrolled bladder emptying. *Detrusor sphincter dyssynergia* , due to loss of synchronization between detrusor and sphincter muscles, causes difficulty in initiating and/or stopping the urinary stream, producing hesitancy, urinary retention, overflow incontinence, and recurrent infection.

Constipation occurs in >30% of patients. Fecal urgency or *bowel incontinence* is less common (15%) but can be socially debilitating.

Cognitive dysfunction can include memory loss, impaired attention, difficulties in executive functioning, memory, problem solving, slowed information processing, and problems shifting between cognitive tasks. Euphoria (elevated mood) was once thought to be characteristic of MS but is actually uncommon, occurring in <20% of patients. Cognitive dysfunction sufficient to impair activities of daily living is rare.

Depression, experienced by approximately half of patients, can be reactive, endogenous, or part of the illness itself, and can contribute to fatigue. *Fatigue* is experienced by 90% of patients; this symptom is the most common reason for work-related disability in MS. Fatigue can be exacerbated by elevated temperatures, by depression, by expending exceptional effort to accomplish basic activities of daily living, or by sleep disturbances (e.g., from frequent nocturnal awakenings to urinate).

Sexual dysfunction may manifest as decreased libido, impaired genital sensation, impotence in men, and diminished vaginal lubrication or adductor spasms in women.

Facial weakness due to a lesion in the pons may resemble idiopathic Bell's palsy (Chap. 376). Unlike Bell's palsy, facial weakness in MS is usually not associated with ipsilateral loss of taste sensation or retroauricular pain.

Vertigo may appear suddenly from a brainstem lesion, superficially resembling acute labyrinthitis (Chap. 21). *Hearing loss* may also occur in MS but is uncommon.

ANCILLARY SYMPTOMS

Heat sensitivity refers to neurologic symptoms produced by an elevation of the body's core temperature. For example, unilateral visual blurring may occur during a hot shower or with physical exercise (*Uhthoff's symptom*). It is also common for MS symptoms to worsen transiently, sometimes dramatically, during febrile illnesses (see "Acute Attacks or Initial Demyelinating Episodes," below). Such heat-related symptoms probably result from transient conduction block (see above).

Lhermitte's symptom is an electric shock–like sensation (typically induced by flexion or other movements of the neck) that radiates down the back into the legs. Rarely, it radiates into the arms. It is generally self-limited but may persist for years. Lhermitte's symptom can also occur with other disorders of the cervical spinal cord (e.g., cervical spondylosis).

Paroxysmal symptoms are distinguished by their brief duration (10 s to 2 min), high frequency (5–40 episodes per day), lack of any alteration of consciousness or change in background electroencephalogram during episodes, and a self-limited course (generally lasting weeks to months). They may be precipitated by hyperventilation or movement. These syndromes may include Lhermitte's symptom; tonic contractions of a limb, face, or trunk (tonic seizures); paroxysmal dysarthria and ataxia; paroxysmal sensory disturbances; and several other less well characterized syndromes. Paroxysmal symptoms probably result from spontaneous discharges, arising at the edges of demyelinated plaques and spreading to adjacent white matter tracts.

Trigeminal neuralgia, hemifacial spasm, and *glossopharyngeal neuralgia* (Chap. 376) can occur when the demyelinating lesion involves the root entry (or exit) zone of the fifth, seventh, and ninth cranial nerve, respectively. Trigeminal neuralgia (tic douloureux) is a very brief lancinating facial pain often triggered by an afferent input from the face or teeth. Most cases of trigeminal neuralgia are not

MS related; however, atypical features such as onset before age 50 years, bilateral symptoms, objective sensory loss, or nonparoxysmal pain should raise concerns that MS could be responsible.

Facial myokymia consists of either persistent rapid flickering contractions of the facial musculature (especially the lower portion of the orbicularis oculi) or a contraction that slowly spreads across the face. It results from lesions of the corticobulbar tracts or brainstem course of the facial nerve.

Disease Course

Four clinical types of MS have been described (Fig. 380-2):

1. *Relapsing/remitting MS* (RRMS) accounts for 85% of MS cases at onset and is characterized by discrete attacks that generally evolve over days to weeks (rarely over hours). There is often complete recovery over the ensuing weeks to months (Fig. 380-2*A*). However, when ambulation is severely impaired during an attack, approximately half will fail to improve. Between attacks, patients are neurologically stable.
2. *Secondary progressive MS* (SPMS) always begins as RRMS (Fig. 380-2*B*). At some point, however, the clinical course changes so that the patient experiences a steady deterioration in function unassociated with acute attacks (which may continue or cease during the progressive phase). SPMS produces a greater amount of fixed neurologic disability than RRMS. For a patient with RRMS, the risk of developing SPMS is ~2% each year, meaning that the great majority of RRMS ultimately evolves into SPMS. SPMS appears to represent a late stage of the same underlying illness as RRMS.
3. *Primary progressive MS* (PPMS) accounts for ~15% of cases. These patients do not experience attacks but only a steady functional decline from disease onset (Fig. 380-2*C*). Compared to RRMS, the sex distribution is more even, the disease begins later in life (mean age ~40 years), and disability develops faster (at least relative to the onset of the first clinical symptom). Despite these differences, PPMS appears to represent the same underlying illness as RRMS.
4. *Progressive/relapsing MS* (PRMS) overlaps PPMS and SPMS and accounts for ~5% of MS patients. Like patients with PPMS, these patients experience a steady deterioration in their condition from disease onset. However, like SPMS patients, they experience occasional attacks superimposed upon their progressive course (Fig. 380-2*D*).

Diagnosis

There is no definitive diagnostic test for MS. Diagnostic criteria for clinically definite MS require documentation of two or more episodes of symptoms and two or more signs that reflect pathology in anatomically noncontiguous white matter tracts of the CNS (Table 380-3). Symptoms must last for >24 h and occur as distinct episodes that are separated by a month or more. At least one of the two required signs must be present on neurologic examination. The second may be documented by abnormal paraclinical tests such as MRI or evoked potentials (EPs). Similarly, in the most recent diagnostic scheme, the second clinical event (in time) may be supported solely by paraclinical information, usually the development of new focal white matter lesions on MRI. In patients who experience gradual progression of disability for ≥6 months without superimposed relapses, documentation of intrathecal IgG synthesis may be used to support the diagnosis.

Table 380-3 Diagnostic Criteria for MS

Clinical Presentation	Additional Data Needed for MS diagnosis
2 or more attacks; objective clinical evidence of 2 or more lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack	None
2 or more attacks; objective clinical evidence of 1 lesion	Dissemination in space, demonstrated by • ≥1 T2 lesion on MRI in at least two out of four MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord) OR • Await a further clinical attack implicating a different CNS site
1 attack; objective clinical evidence of 2 or more lesions	Dissemination in time, demonstrated by • Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time OR • A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its

	timing with reference to a baseline scan OR • Await a second clinical attack
1 attack; objective clinical evidence of 1 lesion (clinically isolated syndrome)	Dissemination in space and time, demonstrated by: For dissemination in space • ≥1 T2 lesion in at least two out of four MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord) • OR • Await a second clinical attack implicating a different CNS site **AND** For dissemination in time • Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time OR • A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan **OR** • Await a second clinical attack
Insidious neurologic progression suggestive of MS (PPMS)	One year of disease progression (retrospectively or prospectively determined) PLUS Two out of the three following criteria Evidence for dissemination in space in the brain based on ≥1 T2+ lesions in the MS-characteristic periventricular, juxtacortical, or infratentorial regions Evidence for dissemination in space in the spinal cord based on ≥2 T2+ lesions in the cord Positive CSF (isoelectric focusing evidence of oligoclonal bands and/or elevated IgG index)

Source: From Polman et al.

Diagnostic Tests

MAGNETIC RESONANCE IMAGING

MRI has revolutionized the diagnosis and management of MS (Fig. 380-3); characteristic abnormalities are found in >95% of patients although more than 90% of the lesions visualized by MRI are asymptomatic. An increase in vascular permeability from a breakdown of the BBB is detected by leakage of intravenous gadolinium (Gd) into the parenchyma. Such leakage occurs early in the development of an MS lesion and serves as a useful marker of inflammation. Gd enhancement persists for approximately 1 month, and the residual MS plaque remains visible indefinitely as a focal area of hyperintensity (a lesion) on spin-echo (T2-weighted) and proton-density images. Lesions are frequently oriented perpendicular to the ventricular surface, corresponding to the pathologic pattern of perivenous demyelination (Dawson's fingers). Lesions are multifocal within the brain, brainstem, and spinal cord. Lesions larger than 6 mm located in the corpus callosum, periventricular white matter, brainstem, cerebellum, or spinal cord are particularly helpful diagnostically. Different criteria for the use of MRI in the diagnosis of MS have been proposed (Table 380-3).

Figure 380-3

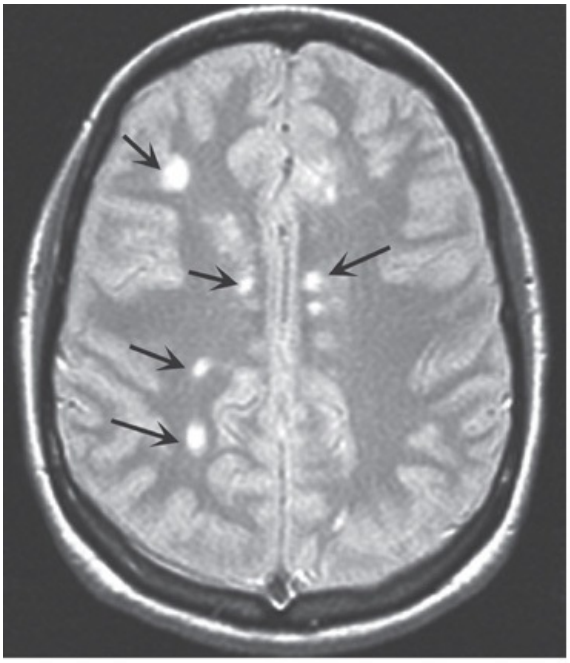

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

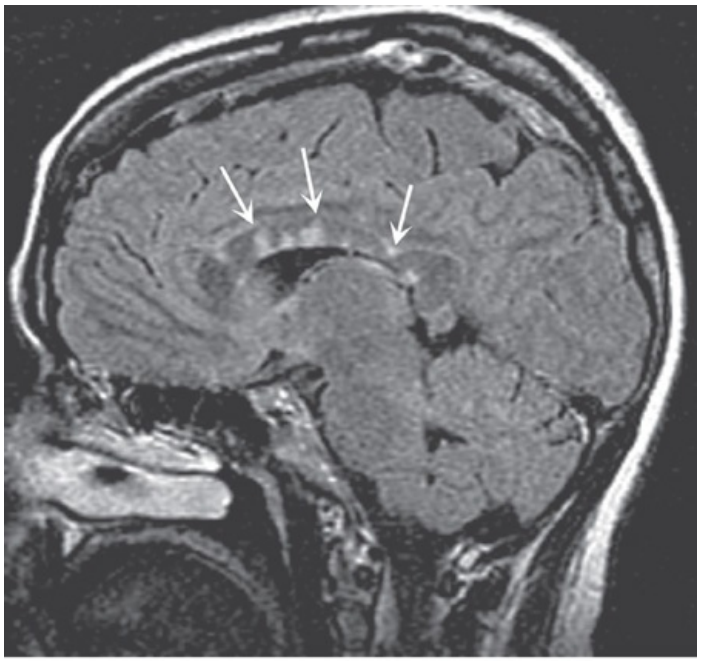

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

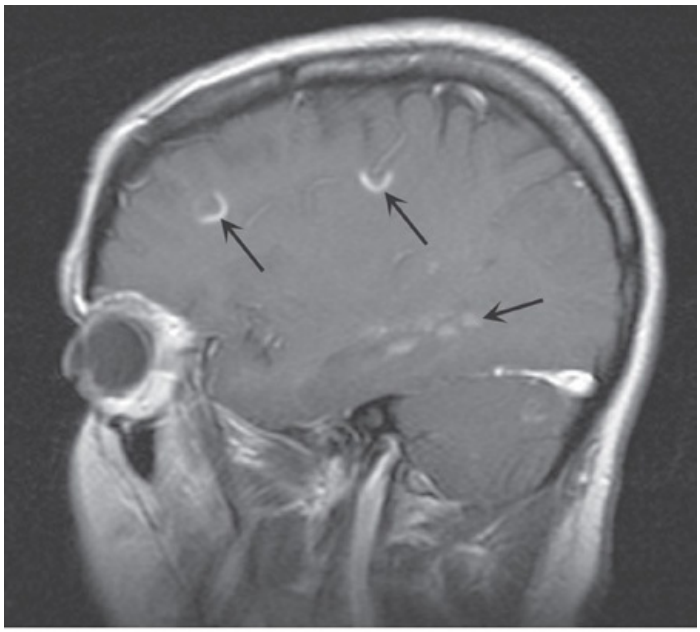

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

MRI findings in MS. ***A.*** Axial first-echo image from T2-weighted sequence demonstrates multiple bright signal abnormalities in white matter, typical for MS. ***B.*** Sagittal T2-weighted FLAIR (fluid attenuated inversion recovery) image in which the high signal of CSF has been suppressed. CSF appears dark, while areas of brain edema or demyelination appear high in signal as shown here in the corpus callosum (*arrows*). Lesions in the anterior corpus callosum are frequent in MS and rare in vascular disease. ***C.*** Sagittal T2-weighted fast spin echo image of the thoracic spine demonstrates a fusiform high-signal-intensity lesion in the midthoracic spinal cord. ***D.*** Sagittal T1-weighted image obtained after the intravenous administration of gadolinium DTPA reveals focal areas of blood-brain barrier disruption, identified as high-signal-intensity regions (*arrows*).

The total volume of T2-weighted signal abnormality (the "burden of disease") shows a significant (albeit weak) correlation with clinical disability, as do measures of brain atrophy. Approximately one-third of T2-weighted lesions appear as hypointense lesions (black holes) on T1-weighted imaging. Black holes may be a marker of irreversible demyelination and axonal loss, although even this measure depends on the timing of the image acquisition (e.g., most acute Gd-enhancing T2 lesions are T1 dark).

Newer MRI measures such as magnetization transfer ratio (MTR) imaging, and proton magnetic resonance spectroscopic imaging (MRSI) may ultimately serve as surrogate markers of clinical disability. MRSI can quantitate molecules such as *N*-acetyl aspartate, which is a marker of axonal integrity, and MTR may be able to distinguish demyelination from edema.

EVOKED POTENTIALS

EP testing assesses function in afferent (visual, auditory, and somatosensory) or efferent (motor) CNS pathways. EPs use computer averaging to measure CNS electric potentials evoked by repetitive stimulation of selected peripheral nerves or of the brain. These tests provide the most information when the pathways studied are clinically uninvolved. For example, in a patient with a remitting and relapsing spinal cord syndrome with sensory deficits in the legs, an abnormal somatosensory EP following posterior tibial nerve stimulation provides little new information. By contrast, an abnormal visual EP in this circumstance would permit a diagnosis of clinically definite MS (Table 380-3). Abnormalities on one or more EP modalities occur in 80–90% of MS patients. EP abnormalities are not specific to MS, although a marked delay in the latency of a specific EP component (as opposed to a reduced amplitude or distorted wave-shape) is suggestive of demyelination.

CEREBROSPINAL FLUID

CSF abnormalities found in MS include a mononuclear cell pleocytosis and an increased level of intrathecally synthesized IgG. The total CSF protein is usually normal or slightly elevated. Various formulas distinguish intrathecally synthesized IgG from IgG that may have entered the CNS passively from the serum. One formula, the CSF IgG index, expresses the ratio of IgG to albumin in the CSF divided by the same ratio in the serum. The IgG synthesis rate uses serum and CSF IgG and albumin measurements to calculate the rate of CNS IgG synthesis. The measurement of oligoclonal banding (OCB) in the CSF also assesses intrathecal production of IgG. OCBs are detected by agarose gel electrophoresis. Two or more OCBs are found in 75–90% of patients with MS. OCBs may be absent at the onset of MS, and in individual patients the number of bands may increase with time. It is important that paired serum samples be studied to exclude a peripheral (i.e., non-CNS) origin of any OCBs detected in the CSF.

A mild CSF pleocytosis (>5 cells/μL) is present in ~25% of cases, usually in young patients with RRMS. A pleocytosis of >75 cells/μL, the presence of polymorphonuclear leukocytes, or a protein concentration >1 g/L (>100 mg/dL) in CSF should raise concern that the patient may not have MS.

Differential Diagnosis

No single clinical sign or test is diagnostic of MS. The diagnosis is readily made in a young adult with relapsing and remitting symptoms involving different areas of CNS white matter. The possibility of an alternative diagnosis should always be considered (Table 380-4), particularly when (1) symptoms are localized exclusively to the posterior fossa, craniocervical junction, or spinal cord; (2) the patient is aged <15 or >60 years; (3) the clinical course is progressive from onset; (4) the patient has never experienced visual, sensory, or bladder symptoms; or (5) laboratory findings (e.g., MRI, CSF, or EPs) are atypical. Similarly, uncommon or rare symptoms in MS (e.g., aphasia, parkinsonism, chorea, isolated dementia, severe muscular atrophy, peripheral neuropathy, episodic loss of consciousness, fever, headache, seizures, or coma) should increase concern about an alternative diagnosis. Diagnosis is also difficult in patients with a rapid or explosive (stroke-like) onset or with mild symptoms and a normal neurologic examination. Rarely, intense inflammation and swelling may produce a mass lesion that mimics a primary or metastatic tumor. The specific tests required to exclude alternative diagnoses will vary with each clinical situation; however, an erythrocyte sedimentation rate, serum B_{12} level, ANA, and treponemal antibody should probably be obtained in all patients with suspected MS.

Table 380-4 Disorders that Can Mimic MS

Acute disseminated encephalomyelitis (ADEM)
Antiphospholipid antibody syndrome
Behçet's disease
Cerebral autosomal dominant arteriopathy, subcortical infarcts, and leukoencephalopathy (CADASIL)
Congenital leukodystrophies (e.g., adrenoleukodystrophy, metachromatic leukodystrophy)
Human immunodeficiency virus (HIV) infection Ischemic optic neuropathy (arteritic and nonarteritic)
Lyme disease
Mitochondrial encephalopathy with lactic acidosis and stroke (MELAS)
Neoplasms (e.g., lymphoma, glioma, meningioma)
Sarcoid
Sjögren's syndrome

Stroke and ischemic cerebrovascular disease
Syphilis
Systemic lupus erythematosus and related collagen vascular disorders
Tropical spastic paraparesis (HTLV I/II infection) Vascular malformations (especially spinal dural AV fistulas)
Vasculitis (primary CNS or other)
Vitamin B12 deficiency

Abbreviations: AV, arteriovenous; CNS, central nervous system; HTLV, human T cell lymphotropic virus.

Prognosis

Most patients with clinically evident MS ultimately experience progressive neurologic disability. In older studies, 15 years after onset, only 20% of patients had no functional limitation, and between one-third and one-half progressed to SPMS and required assistance with ambulation; furthermore, 25 years after onset, ~80% of MS patients reached this level of disability. For unclear reasons, the long-term prognosis for untreated MS appears to have improved in recent years. In addition, the development of disease-modifying therapies for MS also appears to have favorably improved the long-term outlook. Although the prognosis in an individual is difficult to establish, certain clinical features suggest a more favorable prognosis. These include ON or sensory symptoms at onset, fewer than two relapses in the first year of illness, and minimal impairment after 5 years. By contrast, patients with truncal ataxia, action tremor, pyramidal symptoms, or a progressive disease course are more likely to become disabled. Patients with a long-term favorable course are likely to have developed fewer MRI lesions during the early years of disease, and vice versa. Importantly, some MS patients have a benign variant of MS and never develop neurologic disability. The likelihood of having benign MS is thought to be <20%. Patients with benign MS 15 years after onset who have entirely normal neurologic examinations are likely to maintain their benign course.

In patients with their first demyelinating event (i.e., a clinically isolated syndrome), the brain MRI provides prognostic information. With three or more typical T2-weighted lesions, the risk of developing MS after 20 years is ~80%. Conversely, with a normal brain MRI, the likelihood of developing MS is <20%. Similarly, two or more Gd-enhancing lesions at baseline is highly predictive of future MS, as is the appearance of either new T2-weighted lesions or new Gd enhancement ≥3 months after the initial episode.

Mortality as a direct consequence of MS is uncommon, although it has been estimated that the 25-year survival is only 85% of expected. Death can occur during an acute MS attack, although this is distinctly rare. More commonly, death occurs as a complication of MS (e.g., pneumonia in a debilitated individual). Death can also result from suicide.

EFFECT OF PREGNANCY

Pregnant MS patients experience fewer attacks than expected during gestation (especially in the last trimester), but more attacks than expected in the first 3 months postpartum. When considering the pregnancy year as a whole (i.e., 9 months pregnancy plus 3 months postpartum), the overall disease course is unaffected. Decisions about childbearing should thus be made based on (1) the mother's physical state, (2) her ability to care for the child, and (3) the availability of social support. Disease-modifying therapy is generally discontinued during pregnancy, although the actual risk from the interferons and glatiramer acetate (see below) appears to be low.

Treatment: Multiple Sclerosis

Therapy for MS can be divided into several categories: (1) treatment of acute attacks, (2) treatment with disease-modifying agents that reduce the biological activity of MS, and (3) symptomatic therapy. Treatments that promote remyelination or neural repair do not currently exist but would be highly desirable.

The Expanded Disability Status Score (EDSS) is a useful measure of neurologic impairment in MS (Table 380-5). Most patients with EDSS scores <3.5 have RRMS, walk normally, and are generally not disabled; by contrast, patients with EDSS scores >5.5 have progressive MS (SPMS or PPMS), are gait-impaired and, typically, are occupationally disabled.

Table 380-5 Scoring Systems for MS

Kurtzke Expanded Disability Status Score (EDSS)
0.0 = Normal neurologic exam [all grade 0 in functional status (FS)]
1.0 = No disability, minimal signs in one FS (i.e., grade 1)
1.5 = No disability, minimal signs in more than one FS (more than one grade 1)

2.0 = Minimal disability in one FS (one FS grade 2, others 0 or 1)

2.5 = Minimal disability in two FS (two FS grade 2, others 0 or 1)

3.0 = Moderate disability in one FS (one FS grade 3, others 0 or 1) or mild disability in three or four FS (three/four FS grade 2, others 0 or 1) though fully ambulatory

3.5 = Fully ambulatory but with moderate disability in one FS (one grade 3) and one or two FS grade 2; or two FS grade 3; or five FS grade 2 (others 0 or 1)

4.0 = Ambulatory without aid or rest for ~500 m

4.5 = Ambulatory without aid or rest for ~300 m

5.0 = Ambulatory without aid or rest for ~200 m

5.5 = Ambulatory without aid or rest for ~100 m

6.0 = Unilateral assistance required to walk about 100 m with or without resting

6.5 = Constant bilateral assistance required to walk about 20 m without resting

7.0 = Unable to walk beyond about 5 m even with aid; essentially restricted to wheelchair; wheels self and transfers alone

7.5 = Unable to take more than a few steps; restricted to wheelchair; may need aid to transfer

8.0 = Essentially restricted to bed or chair or perambulated in wheelchair, but out of bed most of day; retains many self-care functions; generally has effective use of arms

8.5 = Essentially restricted to bed much of the day; has some effective use of arm(s); retains some self-care functions

9.0 = Helpless bed patient; can communicate and eat

9.5 = Totally helpless bed patient; unable to communicate or eat

10.0 = Death due to MS

Functional Status (FS) Score

A. Pyramidal functions 0 = Normal 1 = Abnormal signs without disability 2 = Minimal disability 3 = Mild or moderate paraparesis or hemiparesis, or severe monoparesis 4 = Marked paraparesis or hemiparesis, moderate quadriparesis, or monoplegia 5 = Paraplegia, hemiplegia, or marked quadriparesis 6 = Quadriplegia

B. Cerebellar functions 0 = Normal 1 = Abnormal signs without disability 2 = Mild ataxia 3 = Moderate truncal or limb ataxia 4 = Severe ataxia all limbs 5 = Unable to perform coordinated movements due to ataxia

C. Brainstem functions 0 = Normal 1 = Signs only 2 = Moderate nystagmus or other mild disability 3 = Severe nystagmus, marked extraocular weakness, or moderate disability of other cranial nerves 4 = Marked dysarthria or other marked disability 5 = Inability to swallow or speak

D. Sensory functions 0 = Normal 1 = Vibration or figure-writing decrease only, in 1 or 2 limbs 2 = Mild decrease in touch or pain or position sense, and/or moderate decrease in vibration in 1 or 2 limbs, or vibratory decrease alone in 3 or 4 limbs 3 = Moderate decrease in touch or pain or position sense, and/or essentially lost vibration in 1 or 2 limbs, or mild decrease in touch or pain, and/or moderate decrease in all proprioceptive tests in 3 or 4 limbs 4 = Marked decrease in touch or pain or loss of proprioception, alone or combined, in 1 or 2 limbs or moderate decrease in touch or pain and/or severe proprioceptive decrease in more than 2 limbs 5 = Loss (essentially) of sensation in 1 or 2 limbs or moderate decrease in touch or pain and/or loss of proprioception for most of the body below the head 6 = Sensation essentially lost below the head

E. Bowel and bladder functions 0 = Normal 1 = Mild urinary hesitancy, urgency, or retention 2 = Moderate hesitancy, urgency, retention of bowel or bladder, or rare urinary incontinence 3 = Frequent urinary incontinence 4 = In need of almost constant catheterization 5 = Loss of bladder function 6 = Loss of bowel and bladder function

F. Visual (or optic) functions 0 = Normal 1 = Scotoma with visual acuity (corrected) better than 20/30 2 = Worse eye with scotoma with maximal visual acuity (corrected) of 20/30 to 20/59 3 = Worse eye with large scotoma, or moderate decrease in fields, but with maximal visual acuity (corrected) of 20/60 to 20/99 4 = Worse eye with marked decrease of fields and maximal acuity (corrected) of 20/100 to 20/200; grade 3 plus maximal acuity of better eye of 20/60 or less 5 = Worse eye with maximal visual acuity (corrected) less than 20/200; grade 4 plus maximal acuity of better eye of 20/60 or less 6 = Grade 5 plus maximal visual acuity of better eye of 20/60 or less

G. Cerebral (or mental) functions 0 = Normal 1 = Mood alteration only (does not affect EDSS score) 2 = Mild decrease in mentation 3 = Moderate decrease in mentation 4 = Marked decrease in mentation 5 = Chronic brain syndrome—severe or incompetent

Source: After JF Kurtzke: Neurology 33:1444, 1983.

ACUTE ATTACKS OR INITIAL DEMYELINATING EPISODES

When patients experience acute deterioration, it is important to consider whether this change reflects new disease activity or a "pseudoexacerbation" resulting from an increase in ambient temperature, fever, or an infection. When the clinical change is thought to reflect a pseudoexacerbation, glucocorticoid treatment is inappropriate. Glucocorticoids are used to manage either first attacks or acute exacerbations. They provide short-term clinical benefit by reducing the severity and shortening the duration of attacks. Whether treatment provides any long-term benefit on the course of the illness is less clear. Therefore, mild attacks are often not treated. Physical and occupational therapy can help with mobility and manual dexterity.

Glucocorticoid treatment is usually administered as intravenous methylprednisolone, 500–1000 mg/d for 3–5 days, either without a taper or followed by a course of oral prednisone beginning at a dose of 60–80 mg/d and gradually tapered over 2 weeks. Orally administered methylprednisolone or dexamethasone (in equivalent dosages) can be substituted for the intravenous portion of the therapy, although GI complications are more common by this route. Outpatient treatment is almost always possible.

Side effects of short-term glucocorticoid therapy include fluid retention, potassium loss, weight gain, gastric disturbances, acne, and emotional lability. Concurrent use of a low-salt, potassium-rich diet and avoidance of potassium-wasting diuretics is advisable. Lithium carbonate (300 mg orally bid) may help to manage emotional lability and insomnia associated with glucocorticoid therapy. Patients with a history of peptic ulcer disease may require cimetidine (400 mg bid) or ranitidine (150 mg bid). Proton pump inhibitors such as pantoprazole (40 mg orally bid) may reduce the likelihood of gastritis, especially when large doses are administered orally. Plasma exchange (5–7 exchanges: 40–60 mL/kg per exchange, every other day for 14 days) may benefit patients with fulminant attacks of demyelination (from MS and other fulminant causes) that are unresponsive to glucocorticoids. However, the cost is high, and conclusive evidence of efficacy is lacking.

DISEASE-MODIFYING THERAPIES FOR RELAPSING FORMS OF MS (RRMS, SPMS WITH EXACERBATIONS)

Seven such agents are approved by the U.S. Food and Drug Administration (FDA): (1) IFN-β-1a (Avonex), (2) IFN-β-1a (Rebif), (3) IFN-β-1b (Betaseron), (4) glatiramer acetate (Copaxone), (5) natalizumab (Tysabri), (6) fingolimod (Gilenya), and (7) mitoxantrone (Novantrone). An eighth, cladribine (Leustatin), is currently awaiting an FDA decision on its approval. Each of these treatments is also used in SPMS patients who continue to experience attacks, because SPMS can be difficult to distinguish from RRMS, and because the available clinical trials suggest that such patients also derive therapeutic benefit. In Phase III clinical trials, recipients of IFN-β-1b, IFN-β-1a, glatiramer acetate, natalizumab, and fingolimod experienced fewer clinical exacerbations and fewer new MRI lesions compared to placebo recipients (Table 380-6). Mitoxantrone (Novantrone), an immune suppressant, has also been approved in the United States, although because of its potential toxicity it is generally reserved for patients with progressive disability who have failed other treatments. When considering the data in Table 380-6, however, it is important to note that the relative efficacy of the different agents cannot be determined by cross-trial comparisons. Relative efficacy can only be determined from a non-biased head-to-head clinical trial.

Table 380-6 Two-Year Outcomes for FDA-Approved Therapies for Multiple Sclerosis[a]

	Clinical Outcomes[b]		MRI Outcomes[c]	
Dose, Route, and Schedule	**Attack Rate, Mean**	**Change in Disease Severity**	**New T2 Lesions[d]**	**Total Burden of Disease**
IFN-β-1b, 250 μg SC qod	–34%[e]	–29% (ns)	–83%[f]	–17%[e]
IFN-β-1a, 30 μg IM qw	–18%[g]	–37%[g]	–36%[f]	–4% (ns)
IFN-β-1a, 44 μg SC tiw	–32%[e]	–30%[g]	–78%[e]	–15%[e]
GA, 20 mg SC qd	–29%[f]	–12% (ns)	–38%[f]	–8%[f]
MTX, 12 mg/m2 IV q3mo	–66%[e]	–75%[g]	–79%[g]	nr
NTZ, 300 mg IV qmo	–68%[e]	–42%[e]	–83%[e]	–18%[e]

FGM, 0.5 mg PO qd	–55%[e]	–27%[f]	–74%[e]	–23%[e]
CLD[h], 3.5 mg/kg PO qyr	–58%[e]	–33%[g]	–73%[e]	nr

[a]Percentage reductions (or increases) have been calculated by dividing the reported rates in the treated group by the comparable rates in the placebo group, except for MRI disease burden, which was calculated as the difference in the median percentage change between the treated and placebo groups.

[b]Severity = 1 point EDSS progression, sustained for 3 months (in the IFN-β-1a 30 μg qw trial, this change was sustained for 6 months; in the IFN-β-1b trial, this was over 3 years).

[c]Different studies measured these MRI measures differently, making comparisons difficult (numbers for new T2 represent the best case scenario for each trial).

[d]New lesions seen on T2-weighted MRI.

[e]p = .001.

[f]p = .01.

[g]p = .05.

[h]Not FDA-approved at time of publication.

Abbreviations: IFN-β, interferon β; GA, glatiramer acetate; MTX, mitoxantrone; NTZ, natalizumab; FGM, fingolimod; CLD, Cladribine; IM, intramuscular; SC, subcutaneous; IV, intravenous; PO, oral; qod, every other day; qw, once per week; tiw, three times per week; qd, daily; q3mo, once every 3 months; qmo, once per month; qyr, once per year; ns, not significant; nr, not reported.

Interferon-β

IFN-β is a class I interferon originally identified by its antiviral properties. Efficacy in MS probably results from immunomodulatory properties, including (1) downregulating expression of MHC molecules on antigen-presenting cells, (2) inhibiting proinflammatory and increasing regulatory cytokine levels, (3) inhibition of T cell proliferation, and (4) limiting the trafficking of inflammatory cells in the CNS. IFN-β reduces the attack rate and improves disease severity measures such as EDSS progression and MRI-documented disease burden. IFN-β should be considered in patients with either RRMS or SPMS with superimposed relapses. In patients with SPMS but without relapses, efficacy has not been established. Head-to-head trials suggest that higher IFN-β doses have slightly greater efficacy but are also more likely to induce neutralizing antibodies, which may reduce the clinical benefit (see below). IFN-β-1a (Avonex), 30 μg, is administered by intramuscular injection once every week. IFN-β-1a (Rebif), 44 μg, is administered by subcutaneous injection three times per week. IFN-β-1b (Betaseron), 250 μg, is administered by subcutaneous injection every other day.

Common side effects of IFN-β therapy include flulike symptoms (e.g., fevers, chills, and myalgias) and mild abnormalities on routine laboratory evaluation (e.g., elevated liver function tests or lymphopenia). Rarely, more severe hepatotoxicity may occur. Subcutaneous IFN-β also causes reactions at the injection site (e.g., pain, redness, induration, or, rarely, skin necrosis). Side effects can usually be managed with concomitant nonsteroidal anti-inflammatory medications and with the use of an autoinjector. Depression, increased spasticity, and cognitive changes have been reported, although these symptoms can also be due to the underlying disease. In any event, side effects to IFN-β therapy usually subside with time.

Approximately 2–10% of IFN-β-1a (Avonex) recipients, 15–25% of IFN-β-1a (Rebif) recipients, and 30–40% of IFN-β-1b (Betaseron) recipients develop neutralizing antibodies to IFN-β, which may disappear over time. Two very large randomized trials (one with more than 2000 patients) provide unequivocal evidence that neutralizing antibodies reduce efficacy as determined by several MRI outcomes. Paradoxically, however, these same trials, despite abundant statistical power, failed to demonstrate any concomitant impact on the clinical outcomes of disability and relapse rate. The reason for this clinical-radiologic dissociation is unresolved. Fortunately, however, there are few situations where measurement of antibodies is necessary. Thus, for a patient doing well on therapy, the presence of antibodies should not affect treatment. Conversely, for a patient doing poorly on therapy, alternative treatment should be considered, even if there are no detectable antibodies.

Glatiramer Acetate

Glatiramer acetate is a synthetic, random polypeptide composed of four amino acids (l-glutamic acid, l-lysine, l-alanine, and l-tyrosine). Its mechanism of action may include (1) induction of antigen-specific suppressor T cells; (2) binding to MHC molecules, thereby displacing bound MBP; or (3) altering the balance between proinflammatory and regulatory cytokines. Glatiramer acetate reduces the attack rate

(whether measured clinically or by MRI) in RRMS. Glatiramer acetate may also benefit disease severity measures, although this is less well established than for the relapse rate. Therefore, glatiramer acetate should be considered in RRMS patients. Its usefulness in progressive disease is entirely unknown. Head-to-head clinical trials suggest that glatiramer acetate has about equal efficacy to high IFN-β doses. Glatiramer acetate, 20 mg, is administered by subcutaneous injection every day. Injection-site reactions also occur with glatiramer acetate. Initially, these were thought to be less severe than with IFN-β-1b, although two recent head-to-head comparisons of high-dose IFN-β to glatiramer acetate did not bear out this impression. In addition, approximately 15% of patients experience one or more episodes of flushing, chest tightness, dyspnea, palpitations, and anxiety after injection. This systemic reaction is unpredictable, brief (duration <1 h), and tends not to recur. Finally, some patients experience lipoatrophy, which, on occasion, can be disfiguring and require cessation of treatment.

Natalizumab

Natalizumab (Tysabri) is a humanized monoclonal antibody directed against the $\alpha 4$ subunit of $\alpha 4\beta 1$ integrin, a cellular adhesion molecule expressed on the surface of lymphocytes. It prevents lymphocytes from binding to endothelial cells, thereby preventing lymphocytes from penetrating the BBB and entering the CNS. Natalizumab greatly reduces the attack rate and significantly improves all measures of disease severity in MS. Moreover, it is well tolerated and the dosing schedule of monthly intravenous infusions make it very convenient for patients. However, because of the development of progressive multifocal leukoencephalopathy (PML) in approximately 0.2% of patients treated with natalizumab for more than 2 years, natalizumab is currently recommended only for patients who have failed other therapies or who have particularly aggressive disease presentations. Its usefulness in the treatment of progressive disease has not been studied. Head-to-head data for natalizumab against low-dose (weekly) IFN-β showed a clear superiority of natalizumab in RRMS; the trial design, however, was biased against IFN-β (i.e., patients recruited could already be considered IFN-β treatment failures). Natalizumab, 300 μg, is administered by IV infusion each month. Treatment with natalizumab is, in general, well tolerated. A small percentage (<10%) of patients experience hypersensitivity reactions (including anaphylaxis) and ~6% develop neutralizing antibodies to the molecule.

The major concern with long-term treatment is the risk of PML. Because the risk is extremely low during the first year of treatment with natalizumab, we currently recommend treatment for periods of 12–18 months only for most patients; after this time, a change to another disease-modifying therapy should be considered. Recently, a blood test to detect antibodies against the PML (JC) virus has shown promise in identifying individuals who are at risk for this complication. In preliminary studies, approximately half of the adult population are antibody-positive, indicating that they experienced an asymptomatic infection with the virus at some time in the past, and to date all cases of natalizumab-associated PML have occurred in seropositive individuals.

Fingolimod

Fingolimod (Gilenya) is a sphingosine-1-phosphate (S1P) inhibitor and it prevents the egress of lymphocytes from the secondary lymphoid organs such as the lymph nodes and spleen. Its mechanism of action is probably due, in part, to the trapping of lymphocytes in the periphery and the prevention, thereby, of lymphocytes reaching the brain. However, because S1P receptors are widely expressed in the CNS tissue and because fingolimod is able to cross the BBB, it may also have central effects. Fingolimod reduces the attack rate and significantly improves all measures of disease severity in MS. It is well tolerated, and the oral dosing schedule makes it very convenient for patients. Moreover, from the clinical trial data presented thus far, it seems to be a reasonably safe therapy and it is approved for first-line use by the FDA. However, as with any new therapy, long-term safety remains to be established. A large head-to-head phase III randomized study demonstrated the clear superiority of fingolimod over low dose (weekly) IFN-β. Fingolimod, 0.5 mg, is administered orally each day. Treatment with fingolimod is also, in general, well tolerated. Mild abnormalities on routine laboratory evaluation (e.g., elevated liver function tests or lymphopenia) are more common than in controls. Although rarely severe, sometimes discontinuation of the medication is necessary. First-degree heart block and bradycardia can also occur, the latter necessitating the prolonged (6-h) observation of patients receiving their first dose.

Mitoxantrone Hydrochloride

Mitoxantrone (Novantrone), an anthracenedione, exerts its antineoplastic action by (1) intercalating into DNA and producing both strand breaks and interstrand cross-links, (2) interfering with RNA synthesis, and (3) inhibiting topoisomerase II (involved in DNA repair). The FDA approved mitoxantrone on the basis of a single (relatively small) phase III clinical trial in Europe, in addition to an even smaller phase II study completed earlier. Mitoxantrone received (from the FDA) the broadest indication of any current treatment for MS. Thus, mitoxantrone is indicated for use in SPMS, in PRMS, and in patients with worsening RRMS (defined as patients whose neurologic status remains significantly abnormal between MS attacks). Despite this broad indication, however, the data supporting its efficacy are weaker than for other approved therapies.

Mitoxantrone can be cardiotoxic (e.g., cardiomyopathy, reduced left ventricular ejection fraction, and

irreversible congestive heart failure). As a result, a cumulative dose >140 mg/m^2 is not recommended. At currently approved doses (12 mg/m^2 every 3 months), the maximum duration of therapy can be only 2–3 years. Furthermore, >40% of women will experience amenorrhea, which may be permanent. Finally, there is risk of acute leukemia, and this complication has already been reported in several mitoxantrone-treated MS patients.

Given these risks, mitoxantrone should not be used as a first-line agent in either RRMS or relapsing SPMS. It is reasonable to consider mitoxantrone in selected patients with a progressive course who have failed other approved therapies.

Cladribine

Cladribine (Leustatin) is a purine analog that inhibits DNA synthesis and repair, and acts as a general immunosuppressant. Cladribine reduces the attack rate and significantly improves several measures of disease severity in MS. It seems to be well tolerated and the easy oral dosing schedule of only taking the drug for 2 weeks/year make it very convenient for patients. Again, however, the principal concern is long-term safety, a concern that is heightened by the long-term immunosuppression that occurs in some patients and, also, by the fact that, in the pivotal RCT, 10 neoplasms and all 20 herpes zoster cases occurred in Leustatin-treated patients.

Initiating and Changing Treatment

Currently, most patients with relapsing forms of MS receive IFN-β or glatiramer acetate as first-line therapy. Although approved for first-line use, the role of fingolimod in this situation has yet to be defined. Regardless of which agent is chosen first, treatment should probably be changed in patients who continue to have frequent attacks or progressive disability (Fig. 380-4). The value of combination therapy is unknown.

Figure 380-4

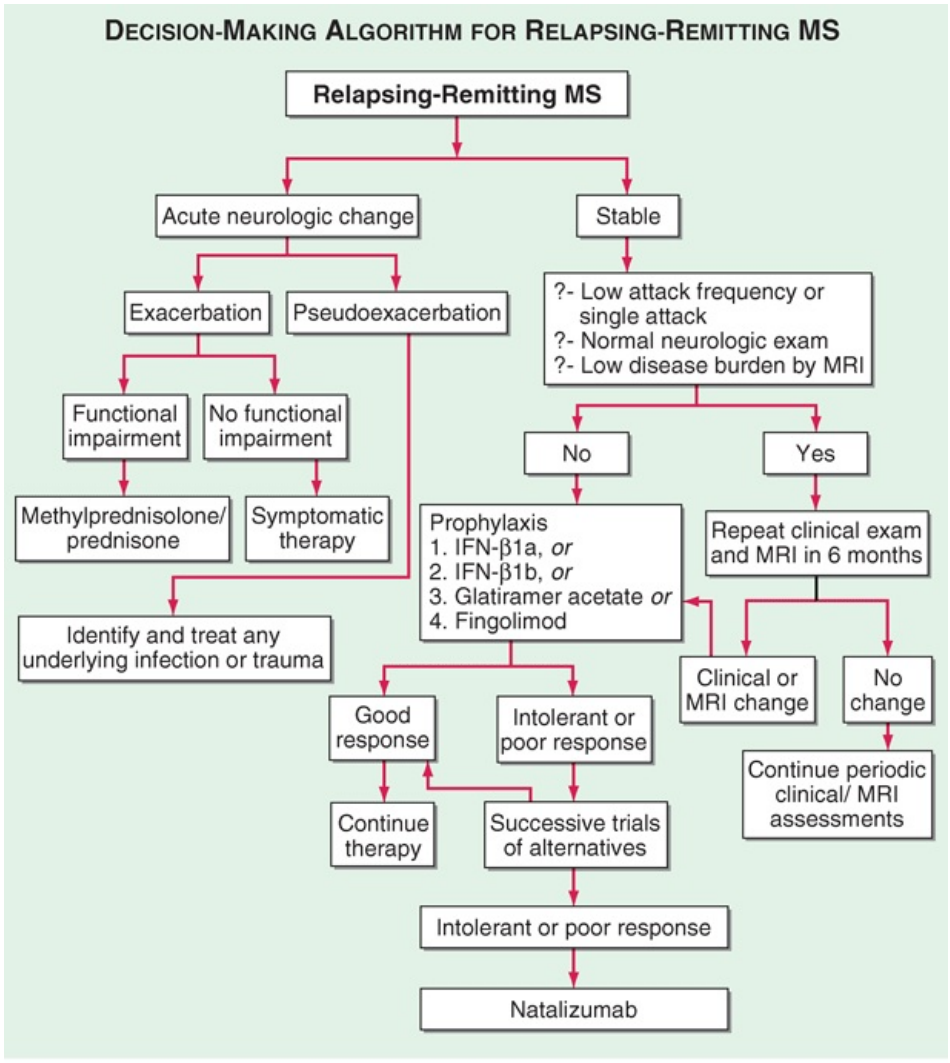

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

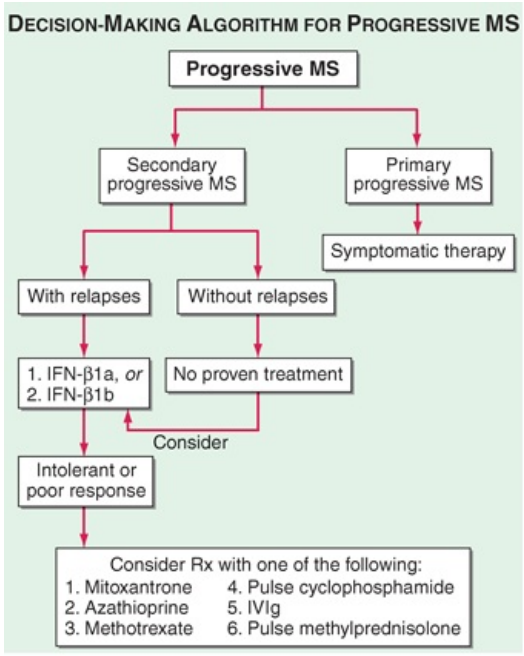

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Therapeutic decision-making for MS.

The long-term efficacy of these treatments remains uncertain, although several recent studies suggest that these agents can improve the long-term outcome of MS, especially when administered early in the RRMS stage of the illness. Beneficial effects seen in early MS include a reduction in the relapse rate, a reduction in CNS inflammation as measured by MRI, and a prolongation in the time to reach certain disability outcomes such as SPMS and requiring assistance to ambulate. Unfortunately, however, already established progressive symptoms do not respond well to treatment with these disease-modifying therapies. Because progressive symptoms are likely to result from delayed effects of earlier focal demyelinating episodes, many experts now believe that very early treatment with a disease-modifying drug is appropriate for most MS patients. It is reasonable to delay initiating treatment in patients with (1) normal neurologic exams, (2) a single attack or a low attack frequency, and (3) a low burden of disease as assessed by brain MRI. Untreated patients, however, should be followed closely with periodic brain MRI scans; the need for therapy is reassessed if scans reveal evidence of ongoing, subclinical disease.

Disease-Modifying Therapies for Progressive MS

SPMS

High-dose IFN-β probably has a beneficial effect in patients with SPMS who are still experiencing acute relapses. IFN-β is probably ineffective in patients with SPMS who are not having acute attacks. Glatiramer acetate and natalizumab have not been studied in this patient population. Although mitoxantrone has been approved for patients with progressive MS, this is not the population studied in the pivotal trial. Therefore, no evidence-based recommendation can be made with regard to its use in this setting.

PPMS

No therapies have been convincingly shown to modify the course of PPMS. A phase III clinical trial of glatiramer acetate in PPMS was stopped because of lack of efficacy. A phase II/III trial of rituximab in PPMS was also negative, but in a preplanned secondary analysis treatment appeared to slow disability progression in patients with gadolinium-enhancing lesions at entry; a follow-up trial with the humanized anti-CD20 therapy ocrelizumab will soon begin. A trial of mitoxantrone in PPMS is ongoing.

Off-Label Treatment Options for Rrms and Spms

Azathioprine (2–3 mg/kg per day) has been used primarily in SPMS. Meta-analysis of published trials suggests that azathioprine is marginally effective at lowering relapse rates, although a benefit on disability progression has not been demonstrated.

Methotrexate (7.5–20 mg/week) was shown in one study to slow the progression of upper-extremity dysfunction in SPMS. Because of the possibility of developing irreversible liver damage, some experts recommend a blind liver biopsy after 2 years of therapy.

Cyclophosphamide (700 mg/m^2, every other month) may be helpful for treatment-refractory patients who are (1) otherwise in good health, (2) ambulatory, and (3) <40 years of age. Because cyclophosphamide can be used for periods in excess of 3 years, it may be preferable to mitoxantrone in these circumstances.

Intravenous immunoglobulin (IVIg), administered in monthly pulses (up to 1 g/kg) for up to 2 years, appears to reduce annual exacerbation rates. However, its use is limited because of its high cost, questions about optimal dose, and uncertainty about its effect on long-term disability outcome.

Methylprednisolone administered in one study as monthly high-dose intravenous pulses reduced disability progression (see above).

Other Therapeutic Claims

Many purported treatments for MS have never been subjected to scientific scrutiny. These include dietary therapies (e.g., the Swank diet in addition to others), megadose vitamins, calcium orotate, bee stings, cow colostrum, hyperbaric oxygen, Procarin (a combination of histamine and caffeine), chelation, acupuncture, acupressure, various Chinese herbal remedies, and removal of mercury-amalgam tooth fillings, among many others. Patients should avoid costly or potentially hazardous unproven treatments. Many such treatments lack biologic plausibility. For example, no reliable case of mercury poisoning resembling typical MS has ever been described.

Although potential roles for EBV, HHV-6, or chlamydia have been suggested for MS, these reports are unconfirmed, and treatment with antiviral agents or antibiotics is not currently appropriate.

Most recently, chronic cerebrospinal insufficiency (CCSVI) has been proposed as a cause of multiple sclerosis and vascular-surgical intervention recommended. However, the failure of independent investigators to even approximate the initial claims of 100% sensitivity and 100% specificity for the diagnostic procedure raised considerable doubt that CCSVI is a real entity. Certainly, any potentially dangerous surgery should be avoided until more rigorous science is available.

Symptomatic Therapy

For all patients, it is useful to encourage attention to a healthy lifestyle, including maintaining an optimistic outlook, a healthy diet, and regular exercise as tolerated (swimming is often well tolerated because of the cooling effect of cold water). It is reasonable also to correct vitamin D deficiency with oral vitamin D, and to recommend dietary supplementation with long-chain (omega-3) unsaturated fatty acids (present in oily fish such as salmon) because of their immunomodulatory properties. *Ataxia/tremor* is often intractable. Clonazepam, 1.5–20 mg/d; Mysoline, 50–250 mg/d; propranolol, 40–200 mg/d; or ondansetron, 8–16 mg/d may help. Wrist weights occasionally reduce tremor in the arm or hand. Thalamotomy or deep-brain stimulation has been tried with mixed success.

Spasticity and *spasms* may improve with physical therapy, regular exercise, and stretching. Avoidance of triggers (e.g., infections, fecal impactions, bed sores) is extremely important. Effective medications include baclofen (Lioresal) (20–120 mg/d), diazepam (2–40 mg/d), tizanidine (8–32 mg/d), dantrolene (25–400 mg/d), and cyclobenzaprine hydrochloride (10–60 mg/d). For severe spasticity, a baclofen pump (delivering medication directly into the CSF) can provide substantial relief.

Weakness can sometimes be improved with the use of potassium channel blockers such as 4-amino pyridine (10–40 mg/d) and 3,4-di-aminopyridine (40–80 mg/d), particularly in the setting where lower extremity weakness interferes with the patient's ability to ambulate. The FDA has approved 4-amino pyridine (at 20 mg/day) and this can be obtained either as dalfampridine (Ampyra) or, more cheaply, through a compounding pharmacy. The principal concern with the use of these agents is the possibility of inducing seizures at high doses.

Pain is treated with anticonvulsants (carbamazepine, 100–1000 mg/d; phenytoin, 300–600 mg/d; gabapentin, 300–3600 mg/d; or pregabalin, 50–300 mg/d), antidepressants (amitriptyline, 25–150 mg/d; nortriptyline, 25–150 mg/d; desipramine, 100–300 mg/d; or venlafaxine, 75–225 mg/d), or antiarrhythmics (mexiletine, 300–900 mg/d). If these approaches fail, patients should be referred to a comprehensive pain management program.

Bladder dysfunction management is best guided by urodynamic testing. Evening fluid restriction or frequent voluntary voiding may help *detrusor hyperreflexia*. If these methods fail, propantheline bromide (10–15 mg/d), oxybutynin (5–15 mg/d), hyoscyamine sulfate (0.5–0.75 mg/d), tolterodine tartrate (2–4 mg/d), or solifenacin (5–10 mg/d) may help. Coadministration of pseudoephedrine (30–60 mg) is sometimes beneficial.

Detrusor/sphincter dyssynergia may respond to phenoxybenzamine (10–20 mg/d) or terazosin hydrochloride (1–20 mg/d). Loss of reflex bladder wall contraction may respond to bethanechol (30–150 mg/d). However, both conditions often require catheterization.

Urinary tract infections should be treated promptly. Patients with large postvoid residual urine volumes are predisposed to infections. Prevention by urine acidification (with cranberry juice or vitamin C) inhibits some bacteria. Prophylactic administration of antibiotics is sometimes necessary but may lead to colonization by resistant organisms. Intermittent catheterization may help to prevent recurrent infections.

Treatment of *constipation* includes high-fiber diets and fluids. Natural or other laxatives may help. *Fecal incontinence* may respond to a reduction in dietary fiber.

Depression should be treated. Useful drugs include the selective serotonin reuptake inhibitors (fluoxetine, 20–80 mg/d; or sertraline, 50–200 mg/d), the tricyclic antidepressants (amitriptyline, 25–150 mg/d; nortriptyline, 25–150 mg/d; or desipramine, 100–300 mg/d), and the non-tricyclic antidepressants (venlafaxine, 75–225 mg/d).

Fatigue may improve with assistive devices, help in the home, or successful management of spasticity. Patients with frequent nocturia may benefit from anticholinergic medication at bedtime. Primary MS fatigue may respond to amantadine (200 mg/d), methylphenidate (5–25 mg/d), or modafinil (100–400 mg/d).

Cognitive problems may respond to the cholinesterase inhibitor donepezil hydrochloride (10 mg/d).

Paroxysmal symptoms respond dramatically to low-dose anticonvulsants (acetazolamide, 200–600 mg/d; carbamazepine, 50–400 mg/d; phenytoin, 50–300 mg/d; or gabapentin, 600–1800 mg/d).

Heat sensitivity may respond to heat avoidance, air-conditioning, or cooling garments.

Sexual dysfunction may be helped by lubricants to aid in genital stimulation and sexual arousal. Management of pain, spasticity, fatigue, and bladder/bowel dysfunction may also help. Sildenafil (50–100 mg), tadalafil (5–20 mg), or vardenafil (5–20 mg) taken 1–2 h before sex are now the standard treatments for maintaining erections.

Promising Experimental Therapies

Numerous clinical trials are currently underway. These include (1) combination therapies; (2) monoclonal antibodies against CD20 to deplete B cells, against the IL-2 receptor, or against CD52 to

induce global lymphocyte depletion; (3) novel oral sphingosine-1-phosphate receptor antagonists to sequester lymphocytes in the secondary lymphoid organs; (4) use of MBP, or an altered peptide ligand resembling MBP, to induce antigen-specific tolerance; (5) an oral inhibitor of the enzyme dihydroorotate dehydrogenase involved in pyrimidine synthesis; (6) estriol to induce a pregnancy-like state; and (7) bone marrow transplantation.

Clinical Variants of MS

Neuromyelitis optica (NMO), or Devic's syndrome, is an aggressive inflammatory disorder consisting most typically of attacks of acute ON and myelitis. Attacks of ON can be bilateral (rare in MS) or unilateral; myelitis can be severe and transverse (rare in MS) and is typically longitudinally extensive, involving three or more contiguous vertebral segments. Attacks of ON may be precede or follow an attack of myelitis by days, months, or years, or vice versa. In contrast to MS, progressive symptoms do not occur in NMO. The brain MRI was classically thought to be normal at the onset of NMO, but recent studies now indicate that asymptomatic lesions sometimes resembling typical MS are common. Lesions involving the hypothalamus, periaqueductal region of the brainstem, or "cloud-like" white matter lesions in the cerebral hemispheres are suggestive of NMO. Brainstem disease can present with nausea and vertigo, and large hemispheral lesions can present as encephalopathy or seizures. Spinal cord MRI typically reveals a focal enhancing region of swelling and cavitation, extending over three or more spinal cord segments and often located in central gray matter structures. Histopathology of these lesions may reveal thickening of blood-vessel walls, demyelination, deposition of antibody and complement, a characteristic loss of astrocytes, and aquaporin-4 staining not seen in MS.

NMO, which is uncommon in whites compared with Asians and Africans, is best understood as a syndrome with diverse causes. Up to 40% of patients have a systemic autoimmune disorder, often systemic lupus erythematosus, Sjögren's syndrome, p-ANCA (perinuclear antineutrophil cytoplasmic antibody)–associated vasculitis, myasthenia gravis, Hashimoto's thyroiditis, or mixed connective tissue disease. In others, onset may be associated with acute infection with varicella-zoster virus, EBV, HIV, or tuberculosis. Rare cases appear to be paraneoplastic and associated with breast, lung, or other cancers. NMO is often idiopathic, however. NMO is usually disabling over time; in one series, respiratory failure from cervical myelitis was present in one-third of patients, and 8 years after onset 60% of patients were blind and more than half had permanent paralysis of one or more limbs.

A highly specific autoantibody directed against the water channel protein aquaporin-4 is present in the sera of 60–70% of patients who have a clinical diagnosis of NMO. Seropositive patients have a very high risk for future relapses. Aquaporin-4 is localized to the foot processes of astrocytes in close apposition to endothelial surfaces. It is likely that aquaporin-4 antibodies are directly pathogenic in NMO, as passive transfer of antibodies from NMO patients into laboratory animals reproduced histologic features of the disease.

When MS affects individuals of African or Asian ancestry, there is a propensity for demyelinating lesions to involve predominantly the optic nerve and spinal cord, an MS subtype termed "opticospinal MS." Interestingly, some individuals with opticospinal MS are seropositive for aquaporin-4 antibodies, suggesting that such cases represent an NMO spectrum disorder.

Acute MS (Marburg's variant) is a fulminant demyelinating process that in some cases progresses inexorably to death within 1–2 years. Typically, there are no remissions. When acute MS presents as a solitary, usually cavitary, lesion, a brain tumor is often suspected. In such cases, a brain biopsy is usually required to establish the diagnosis. An antibody-mediated process appears to be responsible for most cases. Marburg's variant does not seem to follow infection or vaccination, and it is unclear whether this syndrome represents an extreme form of MS or another disease altogether. No controlled trials of therapy exist; high-dose glucocorticoids, plasma exchange, and cyclophosphamide have been tried, with uncertain benefit.

Treatment: Neuromyelitis Optica

Disease-modifying therapies have not been rigorously studied in NMO. Acute attacks of NMO are usually treated with high-dose glucocorticoids (Solu-Medrol 1–2 g/d for 5–10 days followed by a prednisone taper). Because of the likelihood that NMO is antibody-mediated, plasma exchange (typically 7 qod exchanges of 1.5 plasma volumes) has also been used empirically for acute episodes that fail to respond to glucocorticoids. Prophylaxis against relapses can be achieved in some patients with one of the following regimens: mycophenolate mofetil (250 mg bid gradually increasing to 1000 mg bid); B cell depletion with anti-CD20 monoclonal antibody (Rituxan); or a combination of glucocorticoids (500 mg IV methylprednisolone daily for 5 days; then oral prednisone 1 mg/kg per day x 2 months, followed by slow taper) plus azathioprine (2 mg/kg per day started on week 3). By contrast, available evidence suggests that use of IFN-β is ineffective and paradoxically may increase the risk of NMO relapses.

ACUTE DISSEMINATED ENCEPHALOMYELITIS (ADEM)

ADEM has a monophasic course and is most frequently associated with an antecedent infection (postinfectious encephalomyelitis); approximately 5% of ADEM cases follow immunization

(postvaccinal encephalomyelitis). ADEM is more common in children than adults. The hallmark of ADEM is the presence of widely scattered small foci of perivenular inflammation and demyelination in contrast to larger confluent demyelinating lesions typical of MS. In the most explosive form of ADEM, acute hemorrhagic leukoencephalitis, the lesions are vasculitic and hemorrhagic, and the clinical course is devastating.

Postinfectious encephalomyelitis is most frequently associated with the viral exanthems of childhood. Infection with measles virus is the most common antecedent (1 in 1000 cases). Worldwide, measles encephalomyelitis is still common, although use of the live measles vaccine has dramatically reduced its incidence in developed countries. An ADEM-like illness rarely follows vaccination with live measles vaccine (1–2 in 10^6 immunizations). ADEM is now most frequently associated with varicella (chickenpox) infections (1 in 4000–10,000 cases). It may also follow infection with rubella, mumps, influenza, parainfluenza, Epstein-Barr, HIV, and other viruses, and *Mycoplasma*. Some patients may have a nonspecific upper respiratory tract infection or no known antecedent illness. In addition to measles, postvaccinal encephalomyelitis may also follow the administration of smallpox (5 cases per million), the Semple rabies, and Japanese encephalitis vaccines. Modern vaccines that do not require viral culture in CNS tissue have reduced the ADEM risk.

All forms of ADEM presumably result from a cross-reactive immune response to the infectious agent or vaccine that then triggers an inflammatory demyelinating response. Autoantibodies to MBP and to other myelin antigens have been detected in the CSF from many patients with ADEM. Attempts to demonstrate direct viral invasion of the CNS have been unsuccessful.

Clinical Manifestations

In severe cases, onset is abrupt and progression rapid (hours to days). In postinfectious ADEM, the neurologic syndrome generally begins late in the course of the viral illness as the exanthem is fading. Fever reappears, and headache, meningismus, and lethargy progressing to coma may develop. Seizures are common. Signs of disseminated neurologic disease are consistently present (e.g., hemiparesis or quadriparesis, extensor plantar responses, lost or hyperactive tendon reflexes, sensory loss, and brainstem involvement). In ADEM due to chickenpox, cerebellar involvement is often conspicuous. CSF protein is modestly elevated [0.5–1.5 g/L (50–150 mg/dL)]. Lymphocytic pleocytosis, generally 200 cells/μL, occurs in 80% of patients. Occasional patients have higher counts or a mixed polymorphonuclear-lymphocytic pattern during the initial days of the illness. Transient CSF oligoclonal banding has been reported. MRI usually reveals extensive changes in the brain and spinal cord, consisting of white matter hyperintensities on T2 and FLAIR sequences with gadolinium enhancement on T1-weighted sequences.

Diagnosis

The diagnosis is easily established when there is a history of recent vaccination or viral exanthematous illness. In severe cases with predominantly cerebral involvement, acute encephalitis due to infection with herpes simplex or other viruses including HIV may be difficult to exclude (Chap. 381); other considerations include hypercoagulable states including the antiphospholipid antibody syndrome, vasculitis, neurosarcoid, or metastatic cancer. An explosive presentation of MS can mimic ADEM, and especially in adults it may not be possible to distinguish these conditions at onset. The simultaneous onset of disseminated symptoms and signs is common in ADEM and rare in MS. Similarly, meningismus, drowsiness, coma, or seizures suggest ADEM rather than MS. Unlike MS, in ADEM optic nerve involvement is generally bilateral and transverse myelopathy complete. MRI findings that favor ADEM include extensive and relatively symmetric white matter abnormalities, basal ganglia or cortical gray matter lesions, and Gd enhancement of all abnormal areas. By contrast, oligoclonal bands in the CSF are more common in MS. In one study of adult patients initially thought to have ADEM, 30% experienced additional relapses over a follow-up period of 3 years and they were now classified as having MS. Occasional patients with "recurrent ADEM" have also been reported especially in children; however, it is not possible to distinguish this entity from atypical MS.

Treatment: Acute Disseminated Encephalomyelitis

Initial treatment is with high-dose glucocorticoids as for exacerbations of NMO (see above); depending on the response, treatment may need to be continued for 4–8 weeks. Patients who fail to respond within a few days may benefit from a course of plasma exchange or intravenous immunoglobulin. The prognosis reflects the severity of the underlying acute illness. Measles encephalomyelitis is associated with a mortality rate of 5–20%, and most survivors have permanent neurologic sequelae. Children who recover may have persistent seizures and behavioral and learning disorders.

FURTHER READINGS

Axtell RC et al: T helper type 1 and 17 cells determine efficacy of interferon-β in multiple sclerosis and experimental encephalomyelitis. Nat Med 16:406, 2010[PMID: 20348925] [Full Text]

Bradl M et al: Neuromyelitis optica: Pathogenicity of patient immunoglobulin in vivo. Ann Neurol 66:630, 2009[PMID: 19937948] [Full Text]

Cohen JA et al: Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. N Engl J Med 362:402, 2010[PMID: 20089954] [Full Text]

deseze J et al: Acute fulminant demyelinating disease: A descriptive study of 60 patients. Arch Neurol 64:1426, 2007

Fisniku LK et al: Disability and T2 MRI lesions: A 20-year follow-up of patients with relapse onset of multiple sclerosis. Brain 131:808, 2008[PMID: 18234696] [Full Text]

Gorelik L et al: Anti-JC virus antibodies: Implications for PML risk stratification. Ann Neurol 68:295, 2010[PMID: 20737510] [Full Text]

Hinson SR et al: Neurological autoimmunity targeting Aquaporin-4. Neuroscience 168:1009, 2010[PMID: 19699271] [Full Text]

Kappos L et al: Effect of early vs delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: A 3-year follow-up analysis of the BENEFIT study. Lancet 370:389, 2007[PMID: 17679016] [Full Text]

Miller DH, Leary SM: Primary-progressive multiple sclerosis. Lancet Neurol 6:903, 2007[PMID: 17884680] [Full Text]

Oksenberg JR, Baranzini SE: Multiple sclerosis genetics—Is the glass half full, or half empty? Nat Rev Neurol 6:429, 2010[PMID: 20625377] [Full Text]

Polman CH et al: Diagnostic criteria for multiple sclerosis: 2010 revisions to the "McDonald Criteria." Ann Neurol 69:292, 2011 [PMID: 21387374] [Full Text]

Sellner J et al: EFNS guidelines on diagnosis and management of neuromyelitis optica. Eur J Neurol 17:1019, 2010[PMID: 20528913] [Full Text]

Trojano M et al: New natural history of interferon-beta-treated relapsing multiple sclerosis. Ann Neurol 61:300, 2007[PMID: 17444502] [Full Text]

Young N et al: Perivenous demyelination: Association with clinically defined acute disseminated encephalomyelitis and comparison with pathologically confirmed multiple sclerosis. Brain 133:333, 2010[PMID: 20129932] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MENINGITIS, ENCEPHALITIS, BRAIN ABSCESS, AND EMPYEMA: INTRODUCTION

Acute infections of the nervous system are among the most important problems in medicine because early recognition, efficient decision-making, and rapid institution of therapy can be lifesaving. These distinct clinical syndromes include acute bacterial meningitis, viral meningitis, encephalitis, focal infections such as brain abscess and subdural empyema, and infectious thrombophlebitis. Each may present with a nonspecific prodrome of fever and headache, which in a previously healthy individual may initially be thought to be benign, until (with the exception of viral meningitis) altered consciousness, focal neurologic signs, or seizures appear. Key goals of early management are to emergently distinguish between these conditions, identify the responsible pathogen, and initiate appropriate antimicrobial therapy.

APPROACH TO THE PATIENT: MENINGITIS, ENCEPHALITIS, BRAIN ABSCESS, AND EMPYEMA

(Fig. 381-1) The first task is to identify whether an infection predominantly involves the subarachnoid space (*meningitis*) or whether there is evidence of either generalized or focal involvement of brain tissue in the cerebral hemispheres, cerebellum, or brainstem. When brain tissue is directly injured by a viral infection, the disease is referred to as *encephalitis*, whereas focal infections involving brain tissue are classified as either *cerebritis* or *abscess*, depending on the presence or absence of a capsule.

Figure 381-1

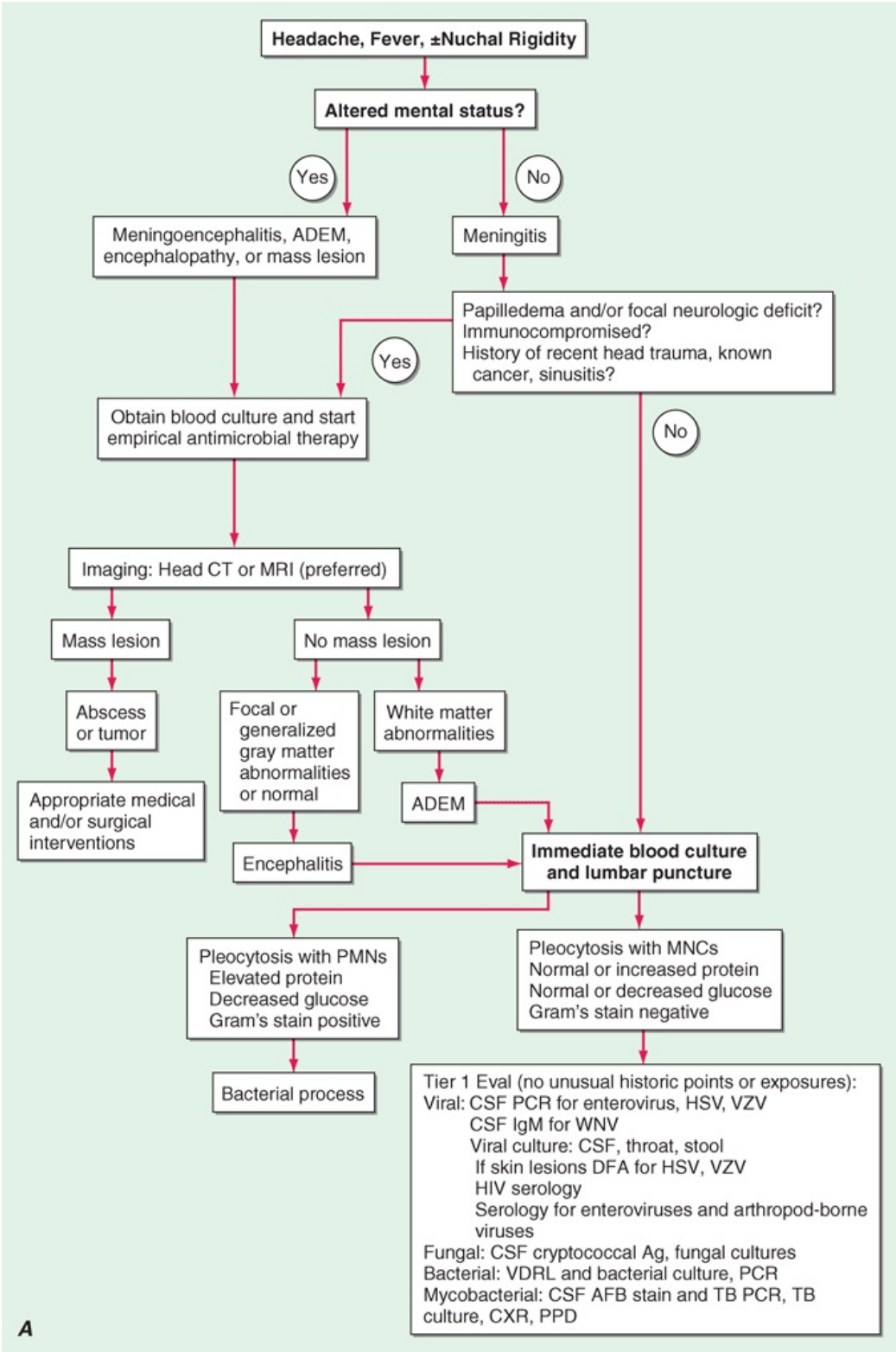

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

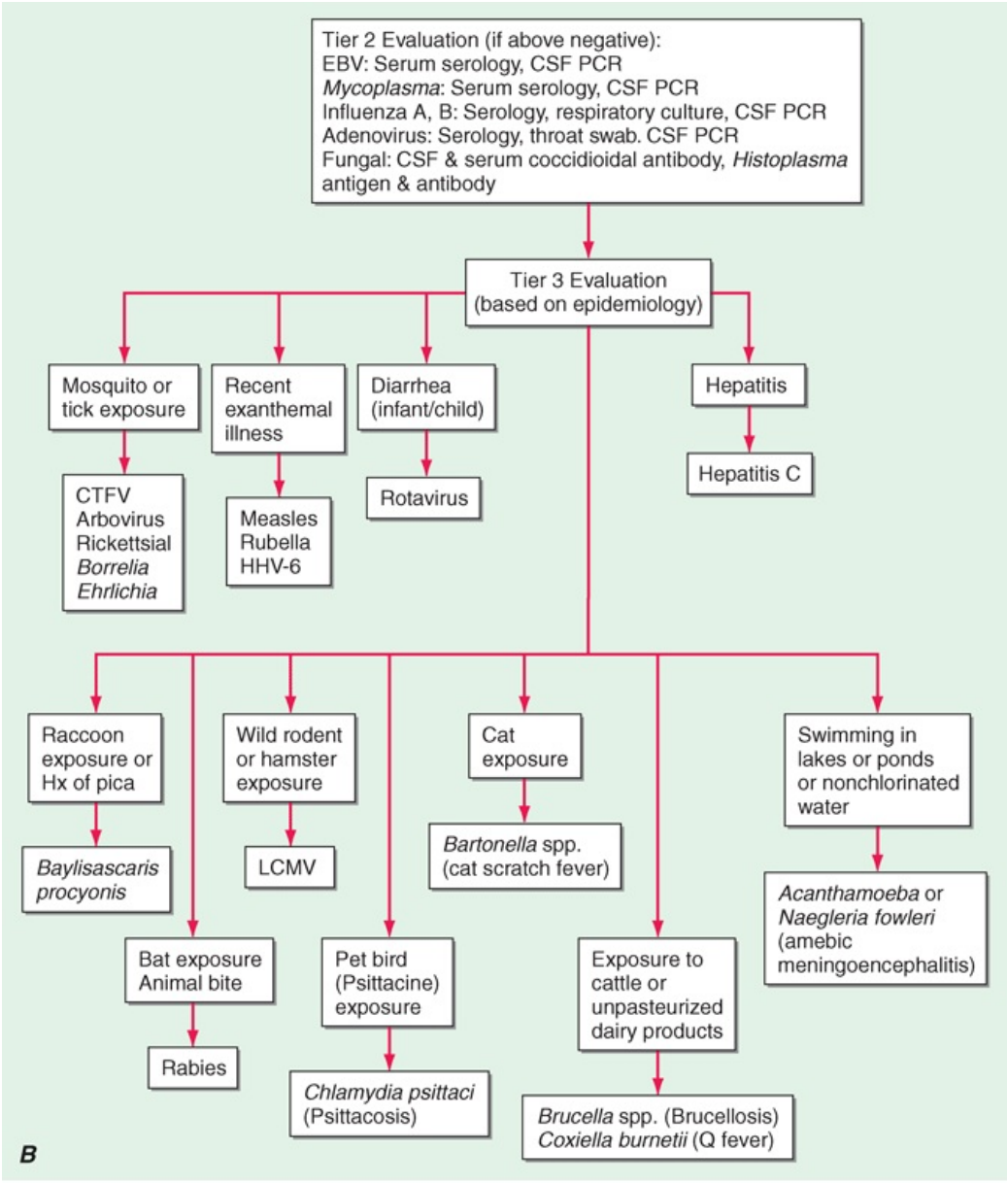

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The management of patients with suspected CNS infection. ADEM, acute disseminated encephalomyelitis; AFB, acid-fast bacillus; Ag, antigen; CSF, cerebrospinal fluid; CT, computed tomography; CTFV, Colorado tick fever virus; CXR, chest x-ray; DFA, direct fluorescent antibody; EBV, Epstein-Barr virus; HHV, human herpesvirus; HSV, herpes simplex virus; LCMV, lymphocytic choriomeningitis virus; MNCs, mononuclear cells; MRI, magnetic resonance imaging; PCR, polymerase chain reaction; PMNs, polymorphonuclear leukocytes; PPD, purified protein derivative; TB, tuberculosis; VDRL, Venereal Disease Research Laboratory; VZV, varicella-zoster virus; WNV, West Nile virus.

Nuchal rigidity ("stiff neck") is the pathognomonic sign of meningeal irritation and is present when the neck resists passive flexion. Kernig's and Brudzinski's signs are also classic signs of meningeal irritation. *Kernig's sign* is elicited with the patient in the supine position. The thigh is flexed on the abdomen, with the knee flexed; attempts to passively extend the knee elicit pain when meningeal irritation is present. *Brudzinski's sign* is elicited with the patient in the supine position and is positive when passive flexion of the neck results in spontaneous flexion of the hips and knees. Although commonly tested on physical examinations, the sensitivity and specificity of Kernig's and Brudzinski's signs are uncertain. Both may be absent or reduced in very young or elderly patients, immunocompromised individuals, or patients with a severely depressed mental status. The high prevalence of cervical spine disease in older individuals may result in false-positive tests for nuchal rigidity.

Initial management can be guided by several considerations: (1) Empirical therapy should be initiated promptly whenever bacterial meningitis is a significant diagnostic consideration. (2) All patients who have had recent head trauma, are immunocompromised, have known malignant lesions or central

nervous system (CNS) neoplasms, or have focal neurologic findings, papilledema or a depressed level of consciousness should undergo CT or MRI of the brain prior to lumbar puncture (LP). In these cases empirical antibiotic therapy should not be delayed pending test results but should be administered prior to neuroimaging and LP. (3) A significantly depressed level of consciousness (e.g., somnolence, coma), seizures, or focal neurologic deficits do not occur in viral meningitis; patients with these symptoms should be hospitalized for further evaluation and treated empirically for bacterial and viral meningoencephalitis. (4) Immunocompetent patients with a normal level of consciousness, no prior antimicrobial treatment, and a cerebrospinal fluid (CSF) profile consistent with viral meningitis (lymphocytic pleocytosis and a normal glucose concentration) can often be treated as outpatients if appropriate contact and monitoring can be ensured. Failure of a patient with suspected viral meningitis to improve within 48 h should prompt a reevaluation including follow-up neurologic and general medical examination and repeat imaging and laboratory studies, often including a second LP.

ACUTE BACTERIAL MENINGITIS

Definition

Bacterial meningitis is an acute purulent infection within the sub-arachnoid space. It is associated with a CNS inflammatory reaction that may result in decreased consciousness, seizures, raised intracranial pressure (ICP), and stroke. The meninges, the subarachnoid space, and the brain parenchyma are all frequently involved in the inflammatory reaction (*meningoencephalitis*).

Epidemiology

Bacterial meningitis is the most common form of suppurative CNS infection, with an annual incidence in the United States of >2.5 cases/100,000 population. The organisms most often responsible for community-acquired bacterial meningitis are *Streptococcus pneumoniae* (~50%), *Neisseria meningitidis* (~25%), group B streptococci (~15%), and *Listeria monocytogenes* (~10%). *Haemophilus influenzae* type b accounts for <10% of cases of bacterial meningitis in most series. *N. meningitidis* is the causative organism of recurring epidemics of meningitis every 8 to 12 years.

Etiology

S. pneumoniae (Chap. 136) is the most common cause of meningitis in adults >20 years of age, accounting for nearly half the reported cases (1.1 per 100,000 persons per year). There are a number of predisposing conditions that increase the risk of pneumococcal meningitis, the most important of which is pneumococcal pneumonia. Additional risk factors include coexisting acute or chronic pneumococcal sinusitis or otitis media, alcoholism, diabetes, splenectomy, hypogammaglobulinemia, complement deficiency, and head trauma with basilar skull fracture and CSF rhinorrhea. The mortality rate remains ~20% despite antibiotic therapy.

The incidence of meningitis due to *N. meningitidis* (Chap. 143) has decreased with the routine immunization of 11- to 18-year-olds with the tetravalent (serogroups A, C, W-135, and Y) meningococcal glycoconjugate vaccine. The vaccine does not contain serogroup B, which is responsible for one-third of cases of meningococcal disease. The presence of petechial or purpuric skin lesions can provide an important clue to the diagnosis of meningococcal infection. In some patients the disease is fulminant, progressing to death within hours of symptom onset. Infection may be initiated by nasopharyngeal colonization, which can result in either an asymptomatic carrier state or invasive meningococcal disease. The risk of invasive disease following nasopharyngeal colonization depends on both bacterial virulence factors and host immune defense mechanisms, including the host's capacity to produce antimeningococcal antibodies and to lyse meningococci by both classic and alternative complement pathways. Individuals with deficiencies of any of the complement components, including properdin, are highly susceptible to meningococcal infections.

Enteric gram-negative bacilli cause meningitis in individuals with chronic and debilitating diseases such as diabetes, cirrhosis, or alcoholism and in those with chronic urinary tract infections. Gram-negative meningitis can also complicate neurosurgical procedures, particularly craniotomy.

Otitis, mastoiditis, and sinusitis are predisposing and associated conditions for meningitis due to *Streptococci* sp., gram-negative anaerobes, *S. aureus, Haemophilus* sp., and Enterobacteriaceae. Meningitis complicating endocarditis may be due to viridans streptococci, *S. aureus, S. bovis,* the HACEK group (*Haemophilus* sp., *Actinobacillus actinomycetemcomitans, Cardiobacterium hominis, Eikenella corrodens, Kingella kingae*), or enterococci.

Group B streptococcus, or *S. agalactiae*, was previously responsible for meningitis predominantly in neonates, but it has been reported with increasing frequency in individuals >50 years of age, particularly those with underlying diseases.

L. monocytogenes (Chap. 139) is an increasingly important cause of meningitis in neonates (<1 month of age), pregnant women, individuals >60 years, and immunocompromised individuals of all ages. Infection is acquired by ingesting foods contaminated by *Listeria*. Foodborne human listerial infection has been reported from contaminated coleslaw, milk, soft cheeses, and several types of "ready-to-eat" foods, including delicatessen meat and uncooked hotdogs.

The frequency of *H. influenzae* type b meningitis in children has declined dramatically since the introduction of the Hib conjugate vaccine, although rare cases of Hib meningitis in vaccinated children have been reported. More frequently, *H. influenzae* causes meningitis in unvaccinated children and older adults, and non-b *H. influenzae* is an emerging pathogen.

Staphylococcus aureus and coagulase-negative staphylococci (Chap. 135) are important causes of meningitis that occurs following invasive neurosurgical procedures, particularly shunting procedures for hydrocephalus, or as a complication of the use of subcutaneous Ommaya reservoirs for administration of intrathecal chemotherapy.

Pathophysiology

The most common bacteria that cause meningitis, *S. pneumoniae* and *N. meningitidis*, initially colonize the nasopharynx by attaching to nasopharyngeal epithelial cells. Bacteria are transported across epithelial cells in membrane-bound vacuoles to the intravascular space or invade the intravascular space by creating separations in the apical tight junctions of columnar epithelial cells. Once in the bloodstream, bacteria are able to avoid phagocytosis by neutrophils and classic complement-mediated bactericidal activity because of the presence of a polysaccharide capsule. Bloodborne bacteria can reach the intraventricular choroid plexus, directly infect choroid plexus epithelial cells, and gain access to the CSF. Some bacteria, such as *S. pneumoniae*, can adhere to cerebral capillary endothelial cells and subsequently migrate through or between these cells to reach the CSF. Bacteria are able to multiply rapidly within CSF because of the absence of effective host immune defenses. Normal CSF contains few white blood cells (WBCs) and relatively small amounts of complement proteins and immunoglobulins. The paucity of the latter two prevents effective opsonization of bacteria, an essential prerequisite for bacterial phagocytosis by neutrophils. Phagocytosis of bacteria is further impaired by the fluid nature of CSF, which is less conducive to phagocytosis than a solid tissue substrate.

A critical event in the pathogenesis of bacterial meningitis is the inflammatory reaction induced by the invading bacteria. Many of the neurologic manifestations and complications of bacterial meningitis result from the immune response to the invading pathogen rather than from direct bacteria-induced tissue injury. As a result, neurologic injury can progress even after the CSF has been sterilized by antibiotic therapy.

The lysis of bacteria with the subsequent release of cell-wall components into the subarachnoid space is the initial step in the induction of the inflammatory response and the formation of a purulent exudate in the subarachnoid space **(Fig. 381-2)**. Bacterial cell-wall components, such as the lipopolysaccharide (LPS) molecules of gram-negative bacteria and teichoic acid and peptidoglycans of *S. pneumoniae*, induce meningeal inflammation by stimulating the production of inflammatory cytokines and chemokines by microglia, astrocytes, monocytes, microvascular endothelial cells, and CSF leukocytes. In experimental models of meningitis, cytokines including tumor necrosis factor alpha (TNF-α) and interleukin 1 β (IL-1β) are present in CSF within 1–2 h of intracisternal inoculation of LPS. This cytokine response is quickly followed by an increase in CSF protein concentration and leukocytosis. Chemokines (cytokines that induce chemotactic migration in leukocytes) and a variety of other proinflammatory cytokines are also produced and secreted by leukocytes and tissue cells that are stimulated by IL-1β and TNF-α. In addition, bacteremia and the inflammatory cytokines induce the production of excitatory amino acids, reactive oxygen and nitrogen species (free oxygen radicals, nitric oxide, and peroxynitrite), and other mediators that can induce death of brain cells, especially in the dentate gyrus of the hippocampus.

Figure 381-2

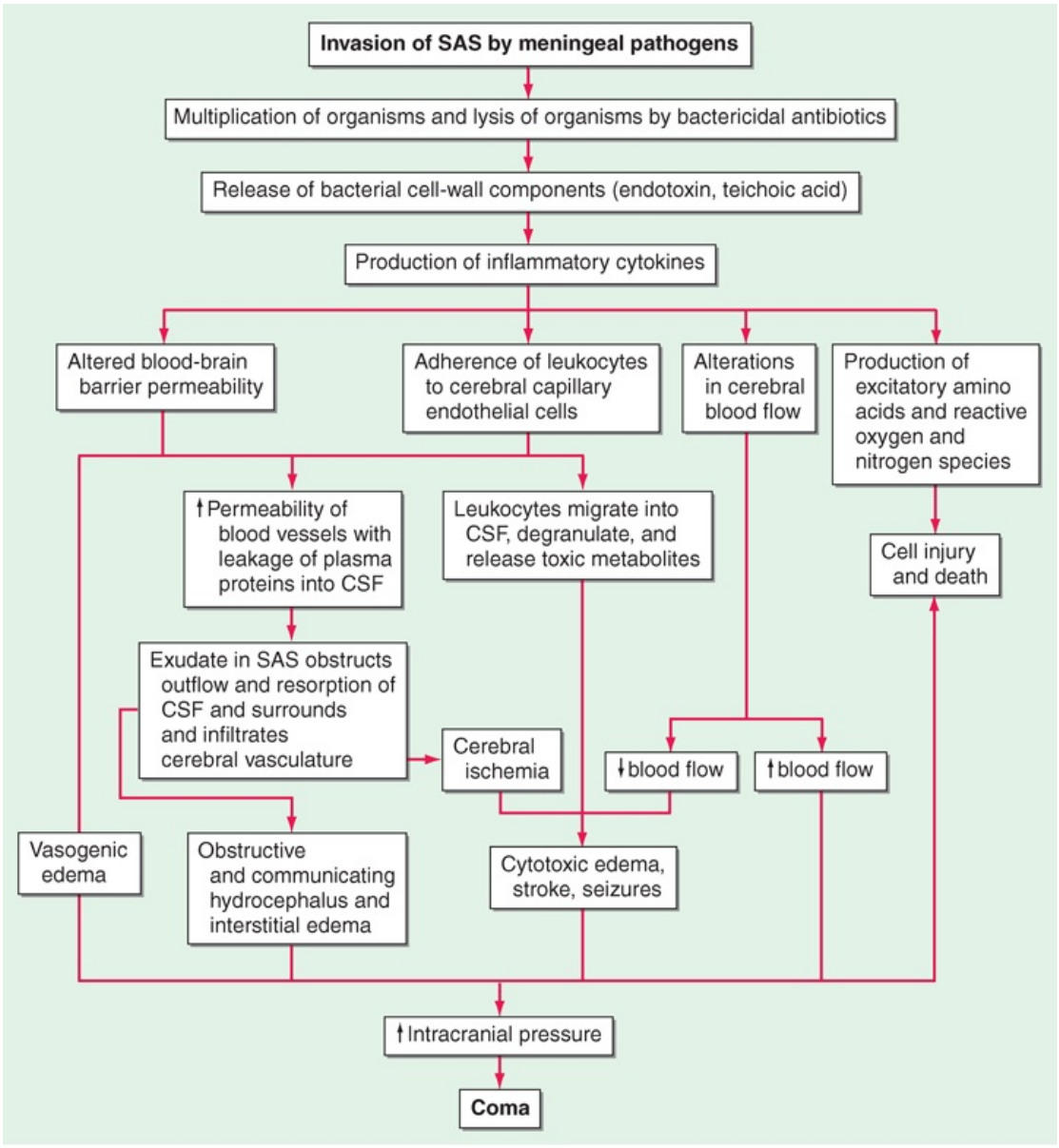

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The pathophysiology of the neurologic complications of bacterial meningitis. CSF, cerebrospinal fluid; SAS, subarachnoid space.

Much of the pathophysiology of bacterial meningitis is a direct consequence of elevated levels of CSF cytokines and chemokines. TNF-α and IL-1β act synergistically to increase the permeability of the blood-brain barrier, resulting in induction of vasogenic edema and the leakage of serum proteins into the subarachnoid space **(Fig. 381-2)**. The subarachnoid exudate of proteinaceous material and leukocytes obstructs the flow of CSF through the ventricular system and diminishes the resorptive capacity of the arachnoid granulations in the dural sinuses, leading to obstructive and communicating hydrocephalus and concomitant interstitial edema.

Inflammatory cytokines upregulate the expression of selectins on cerebral capillary endothelial cells and leukocytes, promoting leukocyte adherence to vascular endothelial cells and subsequent migration into the CSF. The adherence of leukocytes to capillary endothelial cells increases the permeability of blood vessels, allowing for the leakage of plasma proteins into the CSF, which adds to the inflammatory exudate. Neutrophil degranulation results in the release of toxic metabolites that contribute to cytotoxic edema, cell injury, and death. Contrary to previous beliefs, CSF leukocytes probably do little to contribute to the clearance of CSF bacterial infection.

During the very early stages of meningitis, there is an increase in cerebral blood flow, soon followed by a decrease in cerebral blood flow and a loss of cerebrovascular autoregulation (Chap. 275). Narrowing of the large arteries at the base of the brain due to encroachment by the purulent exudate in the subarachnoid space and infiltration of the arterial wall by inflammatory cells with intimal thickening (*vasculitis*) also occur and may result in ischemia and infarction, obstruction of branches of

the middle cerebral artery by thrombosis, thrombosis of the major cerebral venous sinuses, and thrombophlebitis of the cerebral cortical veins. The combination of interstitial, vasogenic, and cytotoxic edema leads to raised ICP and coma. Cerebral herniation usually results from the effects of cerebral edema, either focal or generalized; hydrocephalus and dural sinus or cortical vein thrombosis may also play a role.

Clinical Presentation

Meningitis can present as either an acute fulminant illness that progresses rapidly in a few hours or as a subacute infection that progressively worsens over several days. The classic clinical triad of meningitis is fever, headache, and nuchal rigidity, but the classic triad may not be present. A decreased level of consciousness occurs in >75% of patients and can vary from lethargy to coma. Fever and either headache, stiff neck, or an altered level of consciousness will be present in nearly every patient with bacterial meningitis. Nausea, vomiting, and photophobia are also common complaints.

Seizures occur as part of the initial presentation of bacterial meningitis or during the course of the illness in 20–40% of patients. Focal seizures are usually due to focal arterial ischemia or infarction, cortical venous thrombosis with hemorrhage, or focal edema. Generalized seizure activity and status epilepticus may be due to hyponatremia, cerebral anoxia, or, less commonly, the toxic effects of antimicrobial agents such as high-dose penicillin.

Raised ICP is an expected complication of bacterial meningitis and the major cause of obtundation and coma in this disease. More than 90% of patients will have a CSF opening pressure >180 mmH_2O, and 20% have opening pressures >400 mmH_2O. Signs of increased ICP include a deteriorating or reduced level of consciousness, papilledema, dilated poorly reactive pupils, sixth nerve palsies, decerebrate posturing, and the Cushing reflex (bradycardia, hypertension, and irregular respirations). The most disastrous complication of increased ICP is cerebral herniation. The incidence of herniation in patients with bacterial meningitis has been reported to occur in as few as 1% to as many as 8% of cases.

Specific clinical features may provide clues to the diagnosis of individual organisms and are discussed in more detail in specific chapters devoted to individual pathogens. The most important of these clues is the rash of meningococcemia, which begins as a diffuse erythematous maculopapular rash resembling a viral exanthem; however, the skin lesions of meningococcemia rapidly become petechial. Petechiae are found on the trunk and lower extremities, in the mucous membranes and conjunctiva, and occasionally on the palms and soles.

Diagnosis

When bacterial meningitis is suspected, blood cultures should be immediately obtained and empirical antimicrobial and adjunctive dexamethasone therapy initiated without delay **(Table 381-1)**. The diagnosis of bacterial meningitis is made by examination of the CSF**(Table 381-2)**. The need to obtain neuroimaging studies (CT or MRI) prior to LP requires clinical judgment. In an immunocompetent patient with no known history of recent head trauma, a normal level of consciousness, and no evidence of papilledema or focal neurologic deficits, it is considered safe to perform LP without prior neuroimaging studies. If LP is delayed in order to obtain neuroimaging studies, empirical antibiotic therapy should be initiated after blood cultures are obtained. Antibiotic therapy initiated a few hours prior to LP will not significantly alter the CSF WBC count or glucose concentration, nor is it likely to prevent visualization of organisms by Gram's stain or detection of bacterial nucleic acid by polymerase chain reaction (PCR) assay.

Table 381-1 Antibiotics Used in Empirical Therapy of Bacterial Meningitis and Focal CNS Infections[a]

Indication	Antibiotic
Preterm infants to infants <1 month	Ampicillin + cefotaxime
Infants 1–3 mo	Ampicillin + cefotaxime or ceftriaxone
Immunocompetent children >3 mo and adults <55	Cefotaxime, ceftriaxone or cefepime + vancomycin
Adults >55 and adults of any age with alcoholism or other debilitating illnesses	Ampicillin + cefotaxime, ceftriaxone or cefepime + vancomycin
Hospital-acquired meningitis, posttraumatic or postneurosurgery meningitis, neutropenic patients, or patients with impaired cell-mediated immunity	Ampicillin + ceftazidime or meropenem + vancomycin

Antimicrobial Agent	Total Daily Dose and Dosing Interval	
	Child (>1 month)	Adult

Ampicillin	200 (mg/kg)/d, q4h	12 g/d, q4h
Cefepime	150 (mg/kg)/d, q8h	6 g/d, q8h
Cefotaxime	200 (mg/kg)/d, q6h	12 g/d, q4h
Ceftriaxone	100 (mg/kg)/d, q12h	4 g/d, q12h
Ceftazidime	150 (mg/kg)/d, q8h	6 g/d, q8h
Gentamicin	7.5 (mg/kg)/d, q8hb	7.5 (mg/kg)/d, q8h
Meropenem	120 (mg/kg)/d, q8h	3 g/d, q8h
Metronidazole	30 (mg/kg)/d, q6h	1500–2000 mg/d, q6h
Nafcillin	100–200 (mg/kg)/d, q6h	9–12 g/d, q4h
Penicillin G	400,000 (U/kg)/d, q4h	20–24 million U/d, q4h
Vancomycin	60 (mg/kg)/d, q6h	2 g/d, q12hb

[a]All antibiotics are administered intravenously; doses indicated assume normal renal and hepatic function. [b]Doses should be adjusted based on serum peak and trough levels: gentamicin therapeutic level: peak: 5–8 μg/mL; trough: <2 μg/mL; vancomycin therapeutic level: peak: 25–40 μg/mL; trough: 5–15 μg/mL.

Table 381-2 Cerebrospinal Fluid (CSF) Abnormalities in Bacterial Meningitis

Opening pressure	>180 mmH_2O
White blood cells	10/μL to 10,000/μL; neutrophils predominate
Red blood cells	Absent in nontraumatic tap
Glucose	<2.2 mmol/L (<40 mg/dL)
CSF/serum glucose	<0.4
Protein	>0.45 g/L (>45 mg/dL)
Gram's stain	Positive in >60%
Culture	Positive in >80%
Latex agglutination	May be positive in patients with meningitis due to S. pneumoniae, N. meningitidis, H. influenzae type b, E. coli, group B streptococci
Limulus lysate	Positive in cases of gram-negative meningitis
PCR	Detects bacterial DNA

Abbreviation: PCR, polymerase chain reaction.

The classic CSF abnormalities in bacterial meningitis (Table 381-2) are (1) polymorphonuclear (PMN) leukocytosis (>100 cells/μL in 90%), (2) decreased glucose concentration [<2.2 mmol/L (<40 mg/dL) and/or CSF/serum glucose ratio of <0.4 in ~60%], (3) increased protein concentration [>0.45 g/L (>45 mg/dL) in 90%], and (4) increased opening pressure (>180 mmH_2O in 90%). CSF bacterial cultures are positive in >80% of patients, and CSF Gram's stain demonstrates organisms in >60%.

CSF glucose concentrations <2.2 mmol/L (<40 mg/dL) are abnormal, and a CSF glucose concentration of zero can be seen in bacterial meningitis. Use of the CSF/serum glucose ratio corrects for hyperglycemia that may mask a relative decrease in the CSF glucose concentration. The CSF glucose concentration is low when the CSF/serum glucose ratio is <0.6. A CSF/serum glucose ratio <0.4 is highly suggestive of bacterial meningitis but may also be seen in other conditions, including fungal, tuberculous, and carcinomatous meningitis. It takes from 30 min to several hours for the concentration of CSF glucose to reach equilibrium with blood glucose levels; therefore, administration of 50 mL of

50% glucose (D50) prior to LP, as commonly occurs in emergency room settings, is unlikely to alter CSF glucose concentration significantly unless more than a few hours have elapsed between glucose administration and LP.

A 16S rRNA conserved sequence broad-based bacterial PCR can detect small numbers of viable and nonviable organisms in CSF and is expected to be useful for making a diagnosis of bacterial meningitis in patients who have been pretreated with oral or parenteral antibiotics and in whom Gram's stain and CSF culture are negative. When the broad-range PCR is positive, a PCR that uses specific bacterial primers to detect the nucleic acid of *S. pneumoniae, N. meningitidis, Escherichia coli, L. monocytogenes, H. influenzae*, and *S. agalactiae* can be obtained based on the clinical suspicion of the meningeal pathogen. The latex agglutination (LA) test for the detection of bacterial antigens of *S. pneumoniae*, *N. meningitidis*, *H. influenzae* type b, group B streptococcus, and *E. coli* K1 strains in the CSF has been useful for making a diagnosis of bacterial meningitis but is being replaced by the CSF bacterial PCR assay. The CSF LA test has a *specificity* of 95–100% for *S. pneumoniae* and *N. meningitidis*, so a positive test is virtually diagnostic of bacterial meningitis caused by these organisms. However, the *sensitivity* of the CSF LA test is only 70–100% for detection of *S. pneumoniae* and 33–70% for detection of *N. meningitidis* antigens, so a negative test does not exclude infection by these organisms. The Limulus amebocyte lysate assay is a rapid diagnostic test for the detection of gram-negative endotoxin in CSF and thus for making a diagnosis of gram-negative bacterial meningitis. The test has a specificity of 85–100% and a sensitivity approaching 100%. Thus, a positive Limulus amebocyte lysate assay occurs in virtually all patients with gram-negative bacterial meningitis, but false positives may occur.

Almost all patients with bacterial meningitis will have neuroimaging studies performed during the course of their illness. MRI is preferred over CT because of its superiority in demonstrating areas of cerebral edema and ischemia. In patients with bacterial meningitis, diffuse meningeal enhancement is often seen after the administration of gadolinium. Meningeal enhancement is not diagnostic of meningitis but occurs in any CNS disease associated with increased blood-brain barrier permeability.

Petechial skin lesions, if present, should be biopsied. The rash of meningococcemia results from the dermal seeding of organisms with vascular endothelial damage, and biopsy may reveal the organism on Gram's stain.

Differential Diagnosis

Viral meningoencephalitis, and particularly herpes simplex virus (HSV) encephalitis, can mimic the clinical presentation of bacterial meningitis (see "Viral Encephalitis," below). HSV encephalitis typically presents with headache, fever, altered consciousness, focal neurologic deficits (e.g., dysphasia, hemiparesis), and focal or generalized seizures. The findings on CSF studies, neuroimaging, and electroencephalogram (EEG) distinguish HSV encephalitis from bacterial meningitis. The typical CSF profile with viral CNS infections is a lymphocytic pleocytosis with a normal glucose concentration, in contrast to PMN pleocytosis and hypoglycorrhachia characteristic of bacterial meningitis. MRI abnormalities (other than meningeal enhancement) are not seen in uncomplicated bacterial meningitis. By contrast, in HSV encephalitis, on T2-weighted and fluid-attenuated inversion recovery (FLAIR) MRI images, high signal intensity lesions are seen in the orbitofrontal, anterior, and medial temporal lobes in the majority of patients within 48 h of symptom onset. Some patients with HSV encephalitis have a distinctive periodic pattern on EEG (see below).

Rickettsial disease can resemble bacterial meningitis (Chap. 174). Rocky Mountain spotted fever (RMSF) is transmitted by a tick bite and caused by the bacteria *Rickettsia rickettsii*. The disease may present acutely with high fever, prostration, myalgia, headache, nausea, and vomiting. Most patients develop a characteristic rash within 96 h of the onset of symptoms. The rash is initially a diffuse erythematous maculopapular rash that may be difficult to distinguish from that of meningococcemia. It progresses to a petechial rash, then to a purpuric rash and, if untreated, to skin necrosis or gangrene. The color of the lesions changes from bright red to very dark red, then yellowish-green to black. The rash typically begins in the wrist and ankles and then spreads distally and proximally within a matter of a few hours, involving the palms and soles. Diagnosis is made by immunofluorescent staining of skin biopsy specimens. Ehrlichioses are also transmitted by a tick bite. These are small gram-negative coccobacilli of which two species cause human disease. *Anaplasma phagocytophilum* causes human granulocytic ehrlichiosis (anaplasmosis), and *Ehrlichia chaffeensis* causes human monocytic ehrlichiosis. The clinical and laboratory manifestations of the infections are similar. Patients present with fever, headache, nausea, and vomiting. Twenty percent of patients have a maculopapular or petechial rash. There is laboratory evidence of leukopenia, thrombocytopenia, and anemia, and mild to moderate elevations in alanine aminotransferases, alkaline phosphatase, and lactate dehydrogenase. Patients with RMSF and those with ehrlichial infections may have an altered level of consciousness ranging from mild lethargy to coma, confusion, focal neurologic signs, cranial nerve palsies, hyperreflexia, and seizures.

Focal suppurative CNS infections (see below), including subdural and epidural empyema and brain abscess, should also be considered, especially when focal neurologic findings are present. MRI should

be performed promptly in all patients with suspected meningitis who have focal features, both to detect the intracranial infection and to search for associated areas of infection in the sinuses or mastoid bones.

A number of noninfectious CNS disorders can mimic bacterial meningitis. Subarachnoid hemorrhage (SAH; Chap. 275) is generally the major consideration. Other possibilities include chemical meningitis due to rupture of tumor contents into the CSF (e.g., from a cystic glioma or craniopharyngioma epidermoid or dermoid cyst); drug-induced hypersensitivity meningitis; carcinomatous or lymphomatous meningitis; meningitis associated with inflammatory disorders such as sarcoid, systemic lupus erythematosus (SLE), and Behçet's syndrome; pituitary apoplexy; and uveomeningitic syndromes (Vogt-Koyanagi-Harada syndrome).

On occasion, subacutely evolving meningitis (Chap. 382) may be considered in the differential diagnosis of acute meningitis. The principal causes include *Mycobacterium tuberculosis* (Chap. 165), *Cryptococcus neoformans* (Chap. 202), *Histoplasma capsulatum* (Chap. 199), *Coccidioides immitis* (Chap. 200), and *Treponema pallidum* (Chap. 169).

Treatment: Acute Bacterial Meningitis

EMPIRICAL ANTIMICROBIAL THERAPY

(Table 381-1) Bacterial meningitis is a medical emergency. The goal is to begin antibiotic therapy within 60 min of a patient's arrival in the emergency room. Empirical antimicrobial therapy is initiated in patients with suspected bacterial meningitis before the results of CSF Gram's stain and culture are known. *S. pneumoniae* (Chap. 134) and *N. meningitidis* (Chap. 143) are the most common etiologic organisms of community-acquired bacterial meningitis. Due to the emergence of penicillin- and cephalosporin-resistant *S. pneumoniae*, empirical therapy of community-acquired suspected bacterial meningitis in children and adults should include a combination of dexamethasone, a third- or fourth-generation cephalosporin (e.g., ceftriaxone, cefotaxime, or cefepime), and vancomycin, plus acyclovir, as HSV encephalitis is the leading disease in the differential diagnosis, and doxycycline during tick season to treat tick-borne bacterial infections. Ceftriaxone or cefotaxime provide good coverage for susceptible *S. pneumoniae*, group B streptococci, and *H. influenzae* and adequate coverage for *N. meningitidis*. Cefepime is a broad-spectrum fourth-generation cephalosporin with in vitro activity similar to that of cefotaxime or ceftriaxone against *S. pneumoniae* and *N. meningitidis* and greater activity against *Enterobacter* species and *Pseudomonas aeruginosa*. In clinical trials, cefepime has been demonstrated to be equivalent to cefotaxime in the treatment of penicillin-sensitive pneumococcal and meningococcal meningitis, and it has been used successfully in some patients with meningitis due to *Enterobacter* species and *P. aeruginosa*. Ampicillin should be added to the empirical regimen for coverage of *L. monocytogenes* in individuals <3 months of age, those >55, or those with suspected impaired cell-mediated immunity because of chronic illness, organ transplantation, pregnancy, malignancy, or immunosuppressive therapy. Metronidazole is added to the empirical regimen to cover gram-negative anaerobes in patients with otitis, sinusitis, or mastoiditis. In hospital-acquired meningitis, and particularly meningitis following neurosurgical procedures, staphylococci and gram-negative organisms including *P. aeruginosa* are the most common etiologic organisms. In these patients, empirical therapy should include a combination of vancomycin and ceftazidime, cefepime, or meropenem. Ceftazidime, cefepime, or meropenem should be substituted for ceftriaxone or cefotaxime in neurosurgical patients and in neutropenic patients, as ceftriaxone and cefotaxime do not provide adequate activity against CNS infection with *P. aeruginosa*. Meropenem is a carbapenem antibiotic that is highly active in vitro against *L. monocytogenes*, has been demonstrated to be effective in cases of meningitis caused by *P. aeruginosa*, and shows good activity against penicillin-resistant pneumococci. In experimental pneumococcal meningitis, meropenem was comparable to ceftriaxone and inferior to vancomycin in sterilizing CSF cultures. The number of patients with bacterial meningitis enrolled in clinical trials of meropenem has not been sufficient to definitively assess the efficacy of this antibiotic.

SPECIFIC ANTIMICROBIAL THERAPY

Meningococcal Meningitis

(Table 381-3) Although ceftriaxone and cefotaxime provide adequate empirical coverage for *N. meningitidis*, penicillin G remains the antibiotic of choice for meningococcal meningitis caused by susceptible strains. Isolates of *N. meningitidis* with moderate resistance to penicillin have been identified, but patients infected with these strains have still been successfully treated with penicillin. CSF isolates of *N. meningitidis* should be tested for penicillin and ampicillin susceptibility, and if resistance is found, cefotaxime or ceftriaxone should be substituted for penicillin. A 7-day course of intravenous antibiotic therapy is adequate for uncomplicated meningococcal meningitis. The index case and all close contacts should receive chemoprophylaxis with a 2-day regimen of rifampin (600 mg every 12 h for 2 days in adults and 10 mg/kg every 12 h for 2 days in children >1 year). Rifampin is not recommended in pregnant women. Alternatively, adults can be treated with one dose of azithromycin (500 mg), or one intramuscular dose of ceftriaxone (250 mg). Close contacts are defined as those individuals who have had contact with oropharyngeal secretions, either through kissing or by sharing

toys, beverages, or cigarettes.

Table 381-3 Antimicrobial Therapy of CNS Bacterial Infections Based on Pathogena

Organism	Antibiotic
Neisseria meningitides	
Penicillin-sensitive	Penicillin G or ampicillin
Penicillin-resistant	Ceftriaxone or cefotaxime
Streptococcus pneumoniae	
Penicillin-sensitive	Penicillin G
Penicillin-intermediate	Ceftriaxone or cefotaxime or cefepime
Penicillin-resistant	(Ceftriaxone or cefotaxime or cefepime) + vancomycin
Gram-negative bacilli (except *Pseudomonas* spp.)	Ceftriaxone or cefotaxime
Pseudomonas aeruginosa	Ceftazidime or cefepime or meropenem
Staphylococci spp.	
Methicillin-sensitive	Nafcillin
Methicillin-resistant	Vancomycin
Listeria monocytogenes	Ampicillin + gentamicin
Haemophilus influenzae	Ceftriaxone or cefotaxime or cefepime
Streptococcus agalactiae	Penicillin G or ampicillin
Bacteroides fragilis	Metronidazole
Fusobacterium spp.	Metronidazole

[a]Doses are as indicated in Table 381-1.

Pneumococcal Meningitis

Antimicrobial therapy of pneumococcal meningitis is initiated with a cephalosporin (ceftriaxone, cefotaxime, or cefepime) and vancomycin. All CSF isolates of *S. pneumoniae* should be tested for sensitivity to penicillin and the cephalosporins. Once the results of antimicrobial susceptibility tests are known, therapy can be modified accordingly (Table 381-3). For *S. pneumoniae* meningitis, an isolate of *S. pneumoniae* is considered to be susceptible to penicillin with a minimal inhibitory concentration (MIC) <0.06 μg/mL, to have intermediate resistance when the MIC is 0.1–1.0 μg/mL, and to be highly resistant when the MIC >1.0 μg/mL. Isolates of *S. pneumoniae* that have cephalosporin MICs ≤0.5 μg/mL are considered sensitive to the cephalosporins (cefotaxime, ceftriaxone, cefepime). Those with MICs of 1 μg/mL are considered to have intermediate resistance, and those with MICs ≥2 μg/mL are considered resistant. For meningitis due to pneumococci, with cefotaxime or ceftriaxone MICs ≤0.5 μg/mL, treatment with cefotaxime or ceftriaxone is usually adequate. For MIC >1 μg/mL, vancomycin is the antibiotic of choice. Rifampin can be added to vancomycin for its synergistic effect but is inadequate as monotherapy because resistance develops rapidly when it is used alone.

A 2-week course of intravenous antimicrobial therapy is recommended for pneumococcal meningitis.

Patients with *S. pneumoniae* meningitis should have a repeat LP performed 24–36 h after the initiation of antimicrobial therapy to document sterilization of the CSF. Failure to sterilize the CSF after 24–36 h of antibiotic therapy should be considered presumptive evidence of antibiotic resistance. Patients with penicillin- and cephalosporin-resistant strains of *S. pneumoniae* who do not respond to intravenous vancomycin alone may benefit from the addition of intraventricular vancomycin. The intraventricular route of administration is preferred over the intrathecal route because adequate concentrations of vancomycin in the cerebral ventricles are not always achieved with intrathecal administration.

Listeria Meningitis

Meningitis due to *L. monocytogenes* is treated with ampicillin for at least 3 weeks (Table 381-3). Gentamicin is added in critically ill patients (2 mg/kg loading dose, then 7.5 mg/kg per day given every 8 h and adjusted for serum levels and renal function). The combination of trimethoprim (10–20 mg/kg per day) and sulfamethoxazole (50–100 mg/kg per day) given every 6 h may provide an alternative in penicillin-allergic patients.

Staphylococcal Meningitis

Meningitis due to susceptible strains of *S. aureus* or coagulase-negative staphylococci is treated with nafcillin (Table 381-3). Vancomycin is the drug of choice for methicillin-resistant staphylococci and for patients allergic to penicillin. In these patients, the CSF should be monitored during therapy. If the CSF

is not sterilized after 48 h of intravenous vancomycin therapy, then either intraventricular or intrathecal vancomycin, 20 mg once daily, can be added.

Gram-Negative Bacillary Meningitis

The third-generation cephalosporins—cefotaxime, ceftriaxone, and ceftazidime—are equally efficacious for the treatment of gram-negative bacillary meningitis, with the exception of meningitis due to *P. aeruginosa*, which should be treated with ceftazidime, cefepime, or meropenem(Table 381-3). A 3-week course of intravenous antibiotic therapy is recommended for meningitis due to gram-negative bacilli.

ADJUNCTIVE THERAPY

The release of bacterial cell-wall components by bactericidal antibiotics leads to the production of the inflammatory cytokines IL-1β and TNF-α in the subarachnoid space. Dexamethasone exerts its beneficial effect by inhibiting the synthesis of IL-1β and TNF-α at the level of mRNA, decreasing CSF outflow resistance, and stabilizing the blood-brain barrier. The rationale for giving dexamethasone 20 min before antibiotic therapy is that dexamethasone inhibits the production of TNF-α by macrophages and microglia only if it is administered before these cells are activated by endotoxin. Dexamethasone does not alter TNF-α production once it has been induced. The results of clinical trials of dexamethasone therapy in children, predominantly with meningitis due to *H. influenzae* and *S. pneumoniae*, have demonstrated its efficacy in decreasing meningeal inflammation and neurologic sequelae such as the incidence of sensorineural hearing loss.

A prospective European trial of adjunctive therapy for acute bacterial meningitis in 301 adults found that dexamethasone reduced the number of unfavorable outcomes (15 vs. 25%, $p = .03$) including death (7 vs. 15%, $p = .04$). The benefits were most striking in patients with pneumococcal meningitis. Dexamethasone (10 mg intravenously) was administered 15–20 min before the first dose of an antimicrobial agent, and the same dose was repeated every 6 h for 4 days. These results were confirmed in a second trial of dexamethasone in adults with pneumococcal meningitis. Therapy with dexamethasone should ideally be started 20 min before, or not later than concurrent with, the first dose of antibiotics. It is unlikely to be of significant benefit if started >6 h after antimicrobial therapy has been initiated. Dexamethasone may decrease the penetration of vancomycin into CSF, and it delays the sterilization of CSF in experimental models of *S. pneumoniae* meningitis. As a result, its potential benefit should be carefully weighed when vancomycin is the antibiotic of choice. Alternatively, vancomycin can be administered by the intraventricular route.

One of the concerns for using dexamethasone in adults with bacterial meningitis is that in experimental models of meningitis, dexamethasone therapy increased hippocampal cell injury and reduced learning capacity. This has not been the case in clinical series. The efficacy of dexamethasone therapy in preventing neurologic sequelae is different between high- and low-income countries. Three large randomized trials in low-income countries (sub-Saharan Africa, Southeast Asia) failed to show benefit in subgroups of patients. The lack of efficacy of dexamethasone in these trials has been attributed to late presentation to the hospital with more advanced disease, antibiotic pretreatment, malnutrition, infection with HIV, and treatment of patients with probable, but not microbiologically proven, bacterial meningitis. The results of these clinical trials suggest that patients in sub-Saharan Africa and those in low-income countries with negative CSF Gram's stain and culture should not be treated with dexamethasone.

INCREASED INTRACRANIAL PRESSURE

Emergency treatment of increased ICP includes elevation of the patient's head to 30–45°, intubation and hyperventilation (Pa_{CO2} 25–30 mm Hg), and mannitol. Patients with increased ICP should be managed in an intensive care unit; accurate ICP measurements are best obtained with an ICP monitoring device.

Treatment of increased intracranial pressure is discussed in detail in Chap. 275

Prognosis

Mortality rate is 3–7% for meningitis caused by *H. influenzae*, *N. meningitidis*, or group B streptococci; 15% for that due to *L. monocytogenes*; and 20% for *S. pneumoniae*. In general, the risk of death from bacterial meningitis increases with (1) decreased level of consciousness on admission, (2) onset of seizures within 24 h of admission, (3) signs of increased ICP, (4) young age (infancy) and age >50, (5) the presence of comorbid conditions including shock and/or the need for mechanical ventilation, and (6) delay in the initiation of treatment. Decreased CSF glucose concentration [<2.2 mmol/L (<40 mg/dL)] and markedly increased CSF protein concentration [>3 g/L (> 300 mg/dL)] have been predictive of increased mortality and poorer outcomes in some series. Moderate or severe sequelae occur in ~25% of survivors, although the exact incidence varies with the infecting organism. Common sequelae include decreased intellectual function, memory impairment, seizures, hearing loss and dizziness, and gait disturbances.

ACUTE VIRAL MENINGITIS

Clinical Manifestations

Immunocompetent adult patients with viral meningitis usually present with headache, fever, and signs of meningeal irritation coupled with an inflammatory CSF profile (see below). Headache is almost invariably present and often characterized as frontal or retroorbital and frequently associated with photophobia and pain on moving the eyes. Nuchal rigidity is present in most cases but may be mild and present only near the limit of neck anteflexion. Constitutional signs can include malaise, myalgia, anorexia, nausea and vomiting, abdominal pain, and/or diarrhea. Patients often have mild lethargy or drowsiness; however, profound alterations in consciousness, such as stupor, coma, or marked confusion do not occur in viral meningitis and suggest the presence of encephalitis or other alternative diagnoses. Similarly, seizures or focal neurologic signs or symptoms or neuroimaging abnormalities indicative of brain parenchymal involvement are not typical of viral meningitis and suggest the presence of encephalitis or another CNS infectious or inflammatory process.

Etiology

Using a variety of diagnostic techniques, including CSF PCR, culture, and serology, a specific viral cause can be found in 75–90% of cases of viral meningitis. The most important agents are enteroviruses (including echoviruses and coxsackieviruses in addition to numbered enteroviruses), HSV type 2 (HSV-2), HIV, and arboviruses **(Table 381-4)**. CSF cultures are positive in 30–70% of patients, the frequency of isolation depending on the specific viral agent. Approximately two-thirds of culture-negative cases of "aseptic" meningitis have a specific viral etiology identified by CSF PCR testing (see below).

Table 381-4 Viruses Causing Acute Meningitis and Encephalitis in North America

Acute Meningitis	
Common	**Less Common**
Enteroviruses (coxsackieviruses, echoviruses, and human enteroviruses 68–71)	Varicella zoster virus Epstein-Barr virus Lymphocytic choriomeningitis virus
Herpes simplex virus 2	
Arthropod-borne viruses	
HIV	
Acute Encephalitis	
Common	**Less Common**
Herpesviruses	Rabies
Herpes simplex virus 1	Eastern equine encephalitis virus
Varicella zoster virus	Western equine encephalitis virus
Epstein-Barr virus	Powassan virus
Arthropod-borne viruses	Cytomegalovirus[a]
La Crosse virus	Enteroviruses[a]
West Nile virus	Colorado tick fever
St. Louis encephalitis virus	Mumps

[a]Immunocompromised host.

Epidemiology

Viral meningitis is not a nationally reportable disease; however, it has been estimated that the incidence is ~75,000 cases per year. In temperate climates, there is a substantial increase in cases during the summer and early fall months, reflecting the seasonal predominance of enterovirus and arthropod-borne virus (arbovirus) infections, with a peak monthly incidence of about 1 reported case per 100,000 population.

Laboratory Diagnosis

CSF EXAMINATION

The most important laboratory test in the diagnosis of viral meningitis is examination of the CSF. The

typical profile is a lymphocytic pleocytosis (25–500 cells/μL), a normal or slightly elevated protein concentration [0.2–0.8 g/L (20–80 mg/dL)], a normal glucose concentration, and a normal or mildly elevated opening pressure (100–350 mmH_2O). Organisms are *not* seen on Gram's stain of CSF.

Rarely, PMNs may predominate in the first 48 h of illness, especially with infections due to echovirus 9, West Nile virus, eastern equine encephalitis (EEE) virus, or mumps. A pleocytosis of polymorphonuclear neutrophils occurs in 45% of patients with West Nile virus (WNV) meningitis and can persist for a week or longer before shifting to a lymphocytic pleocytosis. Despite these exceptions, the presence of a CSF PMN pleocytosis in a patient with suspected viral meningitis in whom a specific diagnosis has not been established should prompt consideration of alternative diagnoses, including bacterial meningitis or parameningeal infections. The total CSF cell count in viral meningitis is typically 25–500/μL, although cell counts of several thousand/μL are occasionally seen, especially with infections due to lymphocytic choriomeningitis virus (LCMV) and mumps virus. The CSF glucose concentration is typically normal in viral infections, although it may be decreased in 10–30% of cases due to mumps or LCMV. Rare instances of decreased CSF glucose concentration occur in cases of meningitis due to echoviruses and other enteroviruses, HSV-2, and varicella-zoster virus (VZV). As a rule, a lymphocytic pleocytosis with a low glucose concentration should suggest fungal or tuberculous meningitis, *Listeria* meningoencephalitis, or noninfectious disorders (e.g., sarcoid, neoplastic meningitis).

A number of tests measuring levels of various CSF proteins, enzymes, and mediators—including C-reactive protein, lactic acid, lactate dehydrogenase, neopterin, quinolinate, IL-1β, IL-6, soluble IL-2 receptor, β_2-microglobulin, and TNF—have been proposed as potential discriminators between viral and bacterial meningitis or as markers of specific types of viral infection (e.g., infection with HIV), but they remain of uncertain sensitivity and specificity and are not widely used for diagnostic purposes.

POLYMERASE CHAIN REACTION AMPLIFICATION OF VIRAL NUCLEIC ACID

Amplification of viral-specific DNA or RNA from CSF using PCR amplification has become the single most important method for diagnosing CNS viral infections. In both enteroviral and HSV infections of the CNS, PCR has become the diagnostic procedure of choice and is substantially more sensitive than viral cultures. HSV PCR is also an important diagnostic test in patients with recurrent episodes of "aseptic" meningitis, many of whom have amplifiable HSV DNA in CSF despite negative viral cultures. CSF PCR is also used routinely to diagnose CNS viral infections caused by cytomegalovirus (CMV), Epstein-Barr virus (EBV), VZV, and human herpesvirus 6 (HHV-6). CSF PCR tests are available for WNV but are not as sensitive as detection of WNV-specific CSF IgM. PCR is also useful in the diagnosis of CNS infection caused by *Mycoplasma pneumoniae,* which can mimic viral meningitis and encephalitis.

VIRAL CULTURE

The sensitivity of CSF cultures for the diagnosis of viral meningitis and encephalitis, in contrast to its utility in bacterial infections, is generally poor. In addition to CSF, specific viruses may also be isolated from throat swabs, stool, blood, and urine. Enteroviruses and adenoviruses may be found in feces; arboviruses, some enteroviruses, and LCMV in blood; mumps and CMV in urine; and enteroviruses, mumps, and adenoviruses in throat washings. During enteroviral infections, viral shedding in stool may persist for several weeks. The presence of enterovirus in stool is not diagnostic and may result from residual shedding from a previous enteroviral infection; it also occurs in some asymptomatic individuals during enteroviral epidemics.

SEROLOGIC STUDIES

For some viruses, including many arboviruses such as WNV, serologic studies remain a crucial diagnostic tool. Serum antibody determination is less useful for viruses with high seroprevalence rates in the general population such as HSV, VZV, CMV, and EBV. For viruses with low seroprevalence rates, diagnosis of acute viral infection can be made by documenting seroconversion between acute-phase and convalescent sera (typically obtained after 2–4 weeks) or by demonstrating the presence of virus-specific IgM antibodies. Documentation of synthesis of virus-specific antibodies in CSF, as shown by an increased IgG index or the presence of CSF IgM antibodies, is more useful than serum serology alone and can provide presumptive evidence of CNS infection. Although serum and CSF IgM antibodies generally persist for only a few months after acute infection, there are exceptions to this rule. For example, WNV IgM has been shown to persist in some patients for >1 year following acute infection. Unfortunately, the delay between onset of infection and the host's generation of a virus-specific antibody response often means that serologic data are useful mainly for the retrospective establishment of a specific diagnosis, rather than in aiding acute diagnosis or management.

CSF oligoclonal gamma globulin bands occur in association with a number of viral infections. The associated antibodies are often directed against viral proteins. Oligoclonal bands also occur commonly in certain noninfectious neurologic diseases (e.g., multiple sclerosis) and may be found in nonviral infections (e.g., neurosyphilis, Lyme neuroborreliosis).

OTHER LABORATORY STUDIES

All patients with suspected viral meningitis should have a complete blood count and differential, liver and renal function tests, erythrocyte sedimentation rate (ESR), and C-reactive protein, electrolytes, glucose, creatine kinase, aldolase, amylase, and lipase. Neuroimaging studies (MRI, CT) are not necessary in patients with uncomplicated viral meningitis but should be performed in patients with altered consciousness, seizures, focal neurologic signs or symptoms, or atypical CSF profiles.

Differential Diagnosis

The most important issue in the differential diagnosis of viral meningitis is to consider diseases that can mimic viral meningitis, including (1) untreated or partially treated bacterial meningitis; (2) early stages of meningitis caused by fungi, mycobacteria, or *Treponema pallidum* (neurosyphilis), in which a lymphocytic pleocytosis is common, cultures may be slow growing or negative, and hypoglycorrhachia may not be present early; (3) meningitis caused by agents such as *Mycoplasma*, *Listeria* spp., *Brucella* spp., *Coxiella* spp., *Leptospira* spp., and *Rickettsia* spp.; (4) parameningeal infections; (5) neoplastic meningitis; and (6) meningitis secondary to noninfectious inflammatory diseases, including hypersensitivity meningitis, SLE and other rheumatologic diseases, sarcoidosis, Behçet's syndrome, and the uveomeningitic syndromes. Studies in children >28 days of age suggest that the presence of CSF protein >0.5 g/L (sensitivity 89%, specificity 78%), and elevated serum procalcitonin levels >0.5 ng/m; (sensitivity 89%, specificity 89%) were clues to the presence of bacterial as opposed to "aseptic" meningitis. A variety of clinical algorithms for differentiating bacterial from aseptic meningitis have been promulgated, although none have been widely validated. One such prospectively validated system, the *bacterial meningitis score*, suggests that the probability of bacterial meningitis is 0.1% or less (negative predictive value 99.9%, 95% CI 99.6–100%) in children with CSF pleocytosis who have: (1) a negative CSF Gram's stain, (2) CSF neutrophil count <1000 cells/μL, (3) CSF protein <80 mg/dL, (4) peripheral absolute neutrophil count of <10,000 cells/μL, and (5) no prior history or current presence of seizures.

Specific Viral Etiologies

Enteroviruses (EV) (Chap. 191) are the most common cause of viral meningitis, accounting for >85% of cases in which a specific etiology can be identified. Cases may either be sporadic or occur in clusters. Recent outbreaks of EV meningitis in the United States have been associated with coxsackievirus B5 and echovirus strains 6, 9, and 30. Coxsackievirus strains A9, B3, and B4 are more commonly associated with individual cases. EV71 has produced large epidemics of neurologic disease outside the United States, especially in Southeast Asia, but most recently reported cases in the United States have been sporadic. Enteroviruses are the most likely cause of viral meningitis in the summer and fall months, especially in children (<15 years), although cases occur at reduced frequency year round. Although the incidence of enteroviral meningitis declines with increasing age, some outbreaks have preferentially affected older children and adults. Meningitis outside the neonatal period is usually benign. Patients present with sudden onset of fever; headache; nuchal rigidity; and often constitutional signs, including vomiting, anorexia, diarrhea, cough, pharyngitis, and myalgias. The physical examination should include a careful search for stigmata of enterovirus infection, including exanthems, hand-foot-mouth disease, herpangina, pleurodynia, myopericarditis, and hemorrhagic conjunctivitis. The CSF profile is typically a lymphocytic pleocytosis (100–1000 cells/μL) with normal glucose and normal or mildly elevated protein concentration. However, up to 15% of patients, most commonly young infants rather than older children or adults, have a normal CSF leukocyte count. In rare cases, PMNs may predominate during the first 48 h of illness. CSF reverse transcriptase PCR (RT-PCR) is the diagnostic procedure of choice and is both sensitive (>95%) and specific (>100%). CSF PCR has the highest sensitivity if performed within 48 h of symptom onset, with sensitivity declining rapidly after day 5 of symptoms. Treatment is supportive, and patients usually recover without sequelae. Chronic and severe infections can occur in neonates and in individuals with hypo- or agammaglobulinemia.

Arbovirus infections (Chap. 196) occur predominantly in the summer and early fall. Arboviral meningitis should be considered when clusters of meningitis and encephalitis cases occur in a restricted geographic region during the summer or early fall. In the United States the most important causes of arboviral meningitis and encephalitis are West Nile virus, St. Louis encephalitis virus, and the California encephalitis group of viruses. In WNV epidemics, avian deaths may serve as sentinel infections for subsequent human disease. A history of tick exposure or travel or residence in the appropriate geographic area should suggest the possibility of Colorado tick fever virus or Powassan virus infection, although nonviral tick-borne diseases, including RMSF and Lyme neuroborreliosis, may present similarly. Arbovirus meningoencephalitis is typically associated with a CSF lymphocytic pleocytosis, normal glucose concentration, and normal or mildly elevated protein concentration. However, 40–45% of patients with WNV meningoencephalitis have CSF neutrophilia, which can persist for a week or more. The rarity of hypoglycorrhachia in WNV infection as well as the absence of positive Gram's stains and the negative cultures helps distinguish these patients from those with bacterial meningitis. The presence of increased numbers of plasmacytoid cells or Mollaret-like large mononuclear cells in the CSF may be a clue to the diagnosis of WNV infection. Definitive diagnosis of arboviral meningoencephalitis is based on demonstration of viral-specific IgM in CSF or seroconversion. CSF PCR tests are available for some viruses in selected diagnostic laboratories and at the Centers for Disease Control and Prevention (CDC), but in the case of WNV, sensitivity (~70%) of

CSF PCR is less than that of CSF serology.

HSV-2 meningitis (Chap. 179) has been increasingly recognized as a major cause of viral meningitis in adults, and overall it is probably second in importance to enteroviruses as a cause of viral meningitis, accounting for 5% of total cases overall and undoubtedly a higher frequency of those cases occurring in adults and/or outside of the summer-fall period when enterovirus infections are increasingly common. HSV meningitis occurs in ~25-35% of women and ~10-15% of men at the time of an initial (primary) episode of genital herpes. Of these patients, 20% go on to have recurrent attacks of meningitis. Diagnosis of HSV meningitis is usually by HSV CSF PCR as cultures may be negative, especially in patients with recurrent meningitis. Demonstration of intrathecal synthesis of HSV-specific antibody may also be useful in diagnosis, although antibody tests are less sensitive and less specific than PCR and may not become positive until after the first week of infection. In contrast to HSV encephalitis in adults in which >90% of cases are due to HSV-1, the overwhelming majority of HSV meningitis is due to HSV-2. Although a history of or the presence of HSV genital lesions is an important diagnostic clue, many patients with HSV meningitis give no history and have no evidence of active genital herpes at the time of presentation. Most cases of recurrent viral or "aseptic" meningitis, including cases previously diagnosed as Mollaret's meningitis, are likely due to HSV.

VZV meningitis should be suspected in the presence of concurrent chickenpox or shingles. However, it is important to recognize that in some series, up to 40% of VZV meningitis cases have been reported to occur in the absence of rash. The frequency of VZV as a cause of meningitis is extremely variable, ranging from as low as 3% to as high as 20% in different series. Diagnosis is usually based on CSF PCR, although the sensitivity of this test may not be as high as for the other herpesviruses. In patients with negative CSF PCR results, the diagnosis of VZV CNS infection can be made by the demonstration of VZV-specific intrathecal antibody synthesis and/or the presence of VZV CSF IgM antibodies, or by positive CSF cultures.

EBV infections may also produce aseptic meningitis, with or without associated infectious mononucleosis. The presence of atypical lymphocytes in the CSF or peripheral blood is suggestive of EBV infection but may occasionally be seen with other viral infections. EBV is almost never cultured from CSF. Serum and CSF serology can help establish the presence of acute infection, which is characterized by IgM viral capsid antibodies (VCAs), antibodies to early antigens (EAs), and the absence of antibodies to EBV-associated nuclear antigen (EBNA). CSF PCR is another important diagnostic test, although positive results may reflect viral reactivation associated with other infectious or inflammatory processes.

HIVmeningitis should be suspected in any patient presenting with a viral meningitis with known or suspected risk factors for HIV infection. Meningitis may occur following primary infection with HIV in 5–10% of cases and less commonly at later stages of illness. Cranial nerve palsies, most commonly involving cranial nerves V, VII, or VIII, are more common in HIV meningitis than in other viral infections. Diagnosis can be confirmed by detection of HIV genome in blood or CSF. Seroconversion may be delayed, and patients with negative HIV serologies who are suspected of having HIV meningitis should be monitored for delayed seroconversion. **For further discussion of HIV infection, see Chap. 189.**

Mumps (Chap. 194) should be considered when meningitis occurs in the late winter or early spring, especially in males (male/female ratio 3:1). With the widespread use of the live attenuated mumps vaccine in the United States since 1967, the incidence of mumps meningitis has fallen by >95%; however, mumps remains a potential source of infection in nonimmunized individuals and populations. Rare cases (10–100:100,000 vaccinated individuals) of vaccine-associated mumps meningitis have been described, with onset typically 2–4 weeks after vaccination. The presence of parotitis, orchitis, oophoritis, pancreatitis, or elevations in serum lipase and amylase are suggestive of mumps meningitis; however, their absence does not exclude the diagnosis. Clinical meningitis was previously estimated to occur in 10–30% of patients with mumps parotitis; however, in a recent U.S. outbreak of nearly 2600 cases of mumps, only 11 cases of meningitis were identified, suggesting the incidence may be lower than previously suspected. Mumps infection confers lifelong immunity, so a documented history of previous infection excludes this diagnosis. Patients with meningitis have a CSF pleocytosis that can exceed 1000 cells/μL in 25%. Lymphocytes predominate in 75%, although CSF neutrophilia occurs in 25%. Hypoglycorrhachia, occurs in 10–30% of patients and may be a clue to the diagnosis when present. Diagnosis is typically made by culture of virus from CSF or by detecting IgM antibodies or seroconversion. CSF PCR is available in some diagnostic and research laboratories.

LCMV infection (Chap. 196) should be considered when aseptic meningitis occurs in the late fall or winter and in individuals with a history of exposure to house mice (*Musmusculus*), pet or laboratory rodents (e.g., hamsters, rats, mice), or their excreta. Some patients have an associated rash, pulmonary infiltrates, alopecia, parotitis, orchitis, or myopericarditis. Laboratory clues to the diagnosis of LCMV, in addition to the clinical findings noted above, may include the presence of leukopenia, thrombocytopenia, or abnormal liver function tests. Some cases present with a marked CSF pleocytosis (>1000 cells/μL) and hypoglycorrhachia (<30%). Diagnosis is based on serology and/or culture of virus from CSF.

Treatment: Acute Viral Meningitis

Treatment of almost all cases of viral meningitis is primarily symptomatic and includes use of analgesics, antipyretics, and antiemetics. Fluid and electrolyte status should be monitored. Patients with suspected bacterial meningitis should receive appropriate empirical therapy pending culture results (see above). Hospitalization may not be required in immunocompetent patients with presumed viral meningitis and no focal signs or symptoms, no significant alteration in consciousness, and a classic CSF profile (lymphocytic pleocytosis, normal glucose, negative Gram's stain) if adequate provision for monitoring at home and medical follow-up can be ensured. Immunocompromised patients; patients with significant alteration in consciousness, seizures, or the presence of focal signs and symptoms suggesting the possibility of encephalitis or parenchymal brain involvement; and those patients who have an atypical CSF profile should be hospitalized. Oral or intravenous acyclovir may be of benefit in patients with meningitis caused by HSV-1 or -2 and in cases of severe EBV or VZV infection. Data concerning treatment of HSV, EBV, and VZV meningitis are extremely limited. Seriously ill patients should probably receive intravenous acyclovir (15–30 mg/kg per day in three divided doses), which can be followed by an oral drug such as acyclovir (800 mg, five times daily), famciclovir (500 mg tid), or valacyclovir (1000 mg tid) for a total course of 7–14 days. Patients who are less ill can be treated with oral drugs alone. Patients with HIV meningitis should receive highly active antiretroviral therapy (Chap. 189).

Patients with viral meningitis who are known to have deficient humoral immunity (e.g., X-linked agammaglobulinemia) and who are not already receiving either intramuscular gamma globulin or intravenous immunoglobulin (IVIg), should be treated with these agents. Intraventricu lar administration of immunoglobulin through an Ommaya reservoir has been tried in some patients with chronic enteroviral meningitis who have not responded to intramuscular or intravenous immunoglobulin.

An investigational drug, pleconaril, has shown efficacy against a variety of enteroviral infections and has good oral bioavailability and excellent CNS penetration. Clinical trials in patients with enteroviral meningitis indicated that pleconaril decreased the duration of symptoms compared to placebo; however, the drug is not likely to be marketed and is not generally available, due to its modest benefit in trials of non-CNS EV infections.

Vaccination is an effective method of preventing the development of meningitis and other neurologic complications associated with poliovirus, mumps, and measles infection. A live attenuated VZV vaccine (Varivax) is available in the United States. Clinical studies indicate an effectiveness rate of 70–90% for this vaccine, but a booster may be required to maintain immunity. An inactivated varicella vaccine is available for transplant recipients.

Prognosis

In adults, the prognosis for full recovery from viral meningitis is excellent. Rare patients complain of persisting headache, mild mental impairment, incoordination, or generalized asthenia for weeks to months. The outcome in infants and neonates (<1 year) is less certain; intellectual impairment, learning disabilities, hearing loss, and other lasting sequelae have been reported in some studies.

VIRAL ENCEPHALITIS

Definition

In contrast to viral meningitis, where the infectious process and associated inflammatory response are limited largely to the meninges, in encephalitis the brain parenchyma is also involved. Many patients with encephalitis also have evidence of associated meningitis (meningoencephalitis) and, in some cases, involvement of the spinal cord or nerve roots (encephalomyelitis, encephalomyeloradiculitis).

Clinical Manifestations

In addition to the acute febrile illness with evidence of meningeal involvement characteristic of meningitis, the patient with encephalitis commonly has an altered level of consciousness (confusion, behavioral abnormalities), or a depressed level of consciousness ranging from mild lethargy to coma, and evidence of either focal or diffuse neurologic signs and symptoms. Patients with encephalitis may have hallucinations, agitation, personality change, behavioral disorders, and, at times, a frankly psychotic state. Focal or generalized seizures occur in many patients with encephalitis. Virtually every possible type of focal neurologic disturbance has been reported in viral encephalitis; the signs and symptoms reflect the sites of infection and inflammation. The most commonly encountered focal findings are aphasia, ataxia, upper or lower motor neuron patterns of weakness, involuntary movements (e.g., myoclonic jerks, tremor), and cranial nerve deficits (e.g., ocular palsies, facial weakness). Involvement of the hypothalamic-pituitary axis may result in temperature dysregulation, diabetes insipidus, or the development of the syndrome of inappropriate secretion of antidiuretic hormone (SIADH). Even though neurotropic viruses typically cause pathologic injury in distinct regions of the CNS, variations in clinical presentations make it impossible to reliably establish the etiology of a specific case of encephalitis on clinical grounds alone (see "Differential Diagnosis," below).

Etiology

In the United States, there are ~20,000 reported cases of encephalitis per year, although the actual

number of cases is likely to be significantly larger. Despite comprehensive diagnostic efforts, the majority of cases of acute encephalitis of suspected viral etiology remain of unknown cause. Hundreds of viruses are capable of causing encephalitis, although only a limited subset is responsible for most cases in which a specific cause is identified (Table 381-4). The most commonly identified viruses causing sporadic cases of acute encephalitis in immunocompetent adults are herpesviruses (HSV, VZV, EBV). Epidemics of encephalitis are caused by arboviruses, which belong to several different viral taxonomic groups including *Alphaviruses* (e.g., EEE virus, western equine encephalitis virus), *Flaviviruses* (e.g., WNV, St. Louis encephalitis virus, Japanese encephalitis virus, Powassan virus), and *Bunyaviruses* (e.g., California encephalitis virus serogroup, LaCrosse virus). Historically, the largest number of cases of arbovirus encephalitis in the United States has been due to St. Louis encephalitis virus and the California encephalitis virus serogroup. However, since 2002, WNV has been responsible for the majority of arbovirus meningitis and encephalitis cases in the United States. The 2003 epidemic was the largest epidemic of arboviral neuroinvasive disease (encephalitis + meningitis) ever recorded in the United States, with 2866 cases and 264 deaths. In 2004–2007, WNV has accounted for between 1142 and 1459 confirmed cases of neuroinvasive disease per year in the United States and 100–177 deaths. In 2008 and 2009 there was an unexpected and dramatic decline in both the number of WNV neuroinvasive cases (2008 = 687, 2009 = 335) and the number of deaths (2008 = 44, 2009 = 30). New causes of viral CNS infections are constantly appearing, as evidenced by the recent outbreak of cases of encephalitis in Southeast Asia caused by Nipah virus, a newly identified member of the Paramyxoviridae family; of meningitis in Europe caused by Toscana virus, an arbovirus belonging to the Bunyavirus family; and of neurologic disorders associated with major epidemics of Chikungunya virus, a togavirus, in Africa, India, and Southeast Asia.

Laboratory Diagnosis

CSF EXAMINATION

CSF examination should be performed in all patients with suspected viral encephalitis unless contraindicated by the presence of severely increased ICP. The characteristic CSF profile is indistinguishable from that of viral meningitis and typically consists of a lymphocytic pleocytosis, a mildly elevated protein concentration, and a normal glucose concentration. A CSF pleocytosis (>5 cells/μL) occurs in >95% of immunocompetent patients with documented viral encephalitis. In rare cases, a pleocytosis may be absent on the initial LP but present on subsequent LPs. Patients who are severely immunocompromised by HIV infection, glucocorticoid or other immunosuppressant drugs, chemotherapy, or lymphoreticular malignancies may fail to mount a CSF inflammatory response. CSF cell counts exceed 500/μL in only about 10% of patients with encephalitis. Infections with certain arboviruses (e.g., EEE virus or California encephalitis virus), mumps, and LCMV may occasionally result in cell counts >1000/μL, but this degree of pleocytosis should suggest the possibility of nonviral infections or other inflammatory processes. Atypical lymphocytes in the CSF may be seen in EBV infection and less commonly with other viruses, including CMV, HSV, and enteroviruses. Increased numbers of plasmacytoid or Mollaret-like large mononuclear cells have been reported in WNV encephalitis. Polymorphonuclear pleocytosis occurs in ~45% of patients with WNV encephalitis and is also a common feature in CMV myeloradiculitis in immunocompromised patients. Large numbers of CSF PMNs may be present in patients with encephalitis due to EEE virus, echovirus 9, and, more rarely, other enteroviruses. However, persisting CSF neutrophilia should prompt consideration of bacterial infection, leptospirosis, amebic infection, and noninfectious processes such as acute hemorrhagic leukoencephalitis. About 20% of patients with encephalitis will have a significant number of red blood cells (>500/μL) in the CSF in a nontraumatic tap. The pathologic correlate of this finding may be a hemorrhagic encephalitis of the type seen with HSV; however, CSF red blood cells occur with similar frequency and in similar numbers in patients with nonherpetic focal encephalitides. A decreased CSF glucose concentration is distinctly unusual in viral encephalitis and should suggest the possibility of bacterial, fungal, tuberculous, parasitic, leptospiral, syphilitic, sarcoid, or neoplastic meningitis. Rare patients with mumps, LCMV, or advanced HSV encephalitis and many patients with CMV myeloradiculitis have low CSF glucose concentrations.

CSF PCR

CSF PCR has become the primary diagnostic test for CNS infections caused by CMV, EBV, HHV-6, and enteroviruses (see "Viral Meningitis," above). In the case of VZV CNS infection, CSF PCR and detection of virus-specific IgM or intrathecal antibody synthesis both provide important aids to diagnosis. The sensitivity and specificity of CSF PCRs varies with the virus being tested. The sensitivity (~96%) and specificity (~99%) of HSV CSF PCR is equivalent to or exceeds that of brain biopsy. It is important to recognize that HSV CSF PCR results need to be interpreted after considering the likelihood of disease in the patient being tested, the timing of the test in relationship to onset of symptoms, and the prior use of antiviral therapy. A negative HSV CSF PCR test performed by a qualified laboratory at the appropriate time during illness in a patient with a high likelihood of HSV encephalitis based on clinical and laboratory abnormalities significantly reduces the likelihood of HSV encephalitis but does not exclude it. For example, in a patient with a pretest probability of 35% of having HSV encephalitis, a negative HSV CSF PCR reduces the posttest probability to ~2%, and for a patient with a pretest probability of 60%, a negative test reduces the posttest probability to ~6%. In

both situations a positive test makes the diagnosis almost certain (98–99%). There have been several recent reports of initially negative HSV CSF PCR tests that were obtained early (≤72 h) following symptom onset and that became positive when repeated 1–3 days later. The frequency of positive HSV CSF PCRs in patients with herpes encephalitis also decreases as a function of the duration of illness, with only ~20% of cases remaining positive after ≥14 days. PCR results are generally not affected by ≤1 week of antiviral therapy. In one study, 98% of CSF specimens remained PCR-positive during the first week of initiation of antiviral therapy, but the numbers fell to ~50% by 8–14 days and to ~21% by >15 days after initiation of antiviral therapy.

The sensitivity and specificity of CSF PCR tests for viruses other than herpes simplex have not been definitively characterized. Enteroviral CSF PCR appears to have a sensitivity and specificity of >95%. The specificity of EBV CSF PCR has not been established. Positive EBV CSF PCRs associated with positive tests for other pathogens have been reported and may reflect reactivation of EBV latent in lymphocytes that enter the CNS as a result of an unrelated infectious or inflammatory process. In patients with CNS infection due to VZV, CSF antibody and PCR studies should be considered complementary, as patients may have evidence of intrathecal synthesis of VZV-specific antibodies and negative CSF PCRs. In the case of WNV infection, CSF PCR appears to be less sensitive (~70% sensitivity) than detection of WNV-specific CSF IgM, although PCR testing remains useful in immunocompromised patients who may not mount an effective anti-WNV antibody response.

CSF CULTURE

CSF culture is generally of limited utility in the diagnosis of acute viral encephalitis. Culture may be insensitive (e.g., >95% of patients with HSV encephalitis have negative CSF cultures as do virtually all patients with EBV-associated CNS disease) and often takes too long to significantly affect immediate therapy.

SEROLOGIC STUDIES AND ANTIGEN DETECTION

The basic approach to the serodiagnosis of viral encephalitis is identical to that discussed earlier for viral meningitis. Demonstration of WNV IgM antibodies is diagnostic of WNV encephalitis as IgM antibodies do not cross the blood-brain barrier, and their presence in CSF is therefore indicative of intrathecal synthesis. Timing of antibody collection may be important as the rate of CSF WNV IgM seropositivity increases by ~10% per day during the first week after illness onset, reaching 80% or higher on day 7 after symptom onset. In patients with HSV encephalitis, both antibodies to HSV-1 glycoproteins and glycoprotein antigens have been detected in the CSF. Optimal detection of both HSV antibodies and antigen typically occurs after the first week of illness, limiting the utility of these tests in acute diagnosis. Nonetheless, HSV CSF antibody testing is of value in selected patients whose illness is >1 week in duration and who are CSF PCR–negative for HSV. In the case of VZV infection, CSF antibody tests may be positive when PCR fails to detect viral DNA, and both tests should be considered complementary rather than mutually exclusive.

MRI, CT, EEG

Patients with suspected encephalitis almost invariably undergo neuroimaging studies and often EEG. These tests help identify or exclude alternative diagnoses and assist in the differentiation between a focal, as opposed to a diffuse, encephalitic process. Focal findings in a patient with encephalitis should always raise the possibility of HSV encephalitis. Examples of focal findings include: (1) areas of increased signal intensity in the frontotemporal, cingulate, or insular regions of the brain on T2-weighted, FLAIR, or diffusion-weighted MRI **(Fig. 381-3)**; (2) focal areas of low absorption, mass effect, and contrast enhancement on CT; or (3) periodic focal temporal lobe spikes on a background of slow or low-amplitude ("flattened") activity on EEG. Approximately 10% of patients with PCR-documented HSV encephalitis will have a normal MRI, although nearly 80% will have abnormalities in the temporal lobe, and an additional 10% in extratemporal regions. The lesions are typically hyperintense on T2-weighted images. The addition of FLAIR and diffusion-weighted images to the standard MRI sequences enhances sensitivity. Children with HSV encephalitis may have atypical patterns of MRI lesions and often show involvement of brain regions outside the frontotemporal areas. CT is less sensitive than MRI and is normal in up to 20–35% of patients. EEG abnormalities occur in >75% of PCR-documented cases of HSV encephalitis; they typically involve the temporal lobes but are often nonspecific. Some patients with HSV encephalitis have a distinctive EEG pattern consisting of periodic, stereotyped, sharp-and-slow complexes originating in one or both temporal lobes and repeating at regular intervals of 2–3 s. The periodic complexes are typically noted between days 2 and 15 of the illness and are present in two-thirds of pathologically proven cases of HSV encephalitis.

Figure 381-3

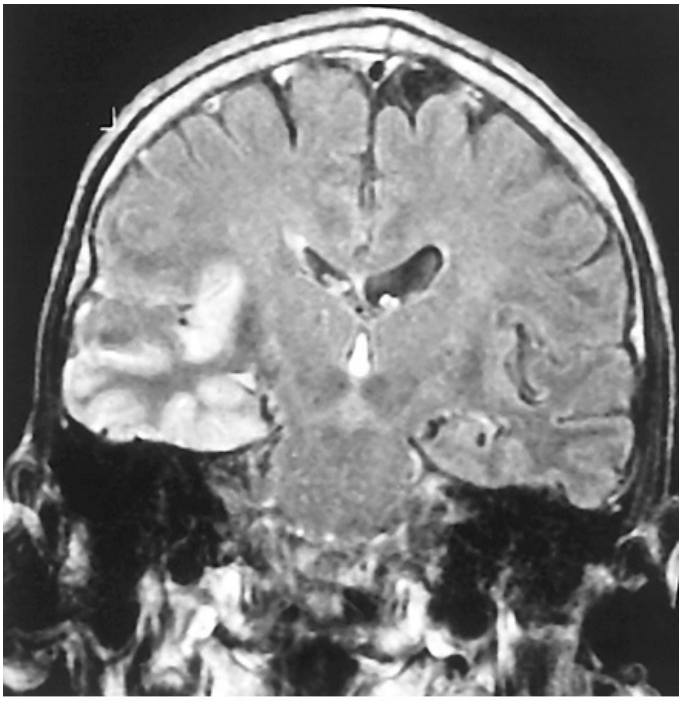

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronal FLAIR magnetic resonance image from a patient with herpes simplex encephalitis. Note the area of increased signal in the right temporal lobe (*left side of image*) confined predominantly to the gray matter. This patient had predominantly unilateral disease; bilateral lesions are more common but may be quite asymmetric in their intensity.

Significant MRI abnormalities are found in only ~two-thirds of patients with WNV encephalitis, a frequency less than that with HSV encephalitis. When present, abnormalities often involve deep brain structures, including the thalamus, basal ganglia, and brainstem, rather than the cortex and may only be apparent on FLAIR images. EEGs in patients with WNV encephalitis typically show generalized slowing that may be more anteriorly prominent rather than the temporally predominant pattern of sharp or periodic discharges more characteristic of HSV encephalitis. Patients with VZV encephalitis may show multifocal areas of hemorrhagic and ischemic infarction, reflecting the tendency of this virus to produce a CNS vasculopathy rather than a true encephalitis. Immunocompromised adult patients with CMV often have enlarged ventricles with areas of increased T2 signal on MRI outlining the ventricles and sub-ependymal enhancement on T1-weighted post-contrast images. **Table 381-5** highlights specific diagnostic test results in encephalitis that can be useful in clinical decision-making.

Table 381-5 Use of Diagnostic Tests in Encephalitis

The best test for WNV encephalitis is the CSF IgM antibody test. The prevalence of positive CSF IgM tests increases by about 10% per day after illness onset and reaches 70–80% by the end of the first week. Serum WNV IgM can provide evidence for recent WNV infection, but in the absence of other findings does not establish the diagnosis of neuroinvasive disease (meningitis, encephalitis, acute flaccid paralysis).
Approximately 80% of patients with proven HSV encephalitis have MRI abnormalities involving the temporal lobes. This percentage likely increases to >90% when FLAIR and DWI MR sequences are also utilized. The absence of temporal lobe lesions on MR reduces the likelihood of HSV encephalitis and should prompt consideration of other diagnostic possibilities.
The CSF HSV PCR test may be negative in the first 72 h of symptoms of HSV encephalitis. A repeat study should be considered in patients with an initial early negative PCR in whom diagnostic suspicion of HSV encephalitis remains high and no alternative diagnosis has yet been established.
Detection of intrathecal synthesis (increased CSF/serum HSV antibody ratio corrected for breakdown of the blood-brain barrier) of HSV-specific antibody may be useful in diagnosis of HSV encephalitis in patients in whom only late (>1 week post-onset) CSF specimens are available and PCR studies are negative. Serum serology alone is of no value in diagnosis of HSV encephalitis due to the high seroprevalence rate in the general population.
Negative CSF viral cultures are of no value in excluding the diagnosis of HSV or EBV encephalitis.
VZV CSF IgM antibodies may be present in patients with a negative VZV CSF PCR. Both tests should be performed in patients with suspected VZV CNS disease.
The specificity of EBV CSF PCR for diagnosis of CNS infection is unknown. Positive tests may occur in patients with a CSF pleocytosis due to other causes. Detection of EBV CSF IgM or intrathecal

synthesis of antibody to VCA supports the diagnosis of EBV encephalitis. Serological studies consistent with acute EBV infection (e.g., IgM VCA, presence of antibodies against EA but not against EBNA) can help support the diagnosis.

Abbreviations: CNS, central nervous system; CSF, cerebrospinal fluid; DWI, diffusion-weighted imaging; EA, early antigen; EBNA, EBV-associated nuclear antigen; EBV, Epstein-Barr virus; FLAIR, fluid-attenuated inversion recovery; HSV, herpes simplex virus; IgM, immunoglobulin M; MRI, magnetic resonance imaging; PCR, polymerase chain reaction; VCA, viral capsid antibody; VZV, varicella-zoster virus; WNV, West Nile virus.

BRAIN BIOPSY

Brain biopsy is now generally reserved for patients in whom CSF PCR studies fail to lead to a specific diagnosis, who have focal abnormalities on MRI, and who continue to show progressive clinical deterioration despite treatment with acyclovir and supportive therapy.

Differential Diagnosis

Infection by a variety of other organisms can mimic viral encephalitis. In studies of biopsy-proven HSV encephalitis, common infectious mimics of focal viral encephalitis included mycobacteria, fungi, rickettsia, *Listeria, Mycoplasma,* and other bacteria (including *Bartonella* sp.).

Infection caused by the ameba *Naegleria fowleri* can also cause acute meningoencephalitis (primary amebic meningoencephalitis), whereas that caused by *Acanthamoeba* and *Balamuthia* more typically produces subacute or chronic granulomatous amebic meningoencephalitis. *Naegleria* thrive in warm, iron-rich pools of water, including those found in drains, canals, and both natural and human-made outdoor pools. Infection has typically occurred in immunocompetent children with a history of swimming in potentially infected water. The CSF, in contrast to the typical profile seen in viral encephalitis, often resembles that of bacterial meningitis with a neutrophilic pleocytosis and hypoglycorrhachia. Motile trophozoites can be seen in a wet mount of warm, fresh CSF. There have been an increasing number of cases of *Balamuthia mandrillaris* amebic encephalitis mimicking acute viral encephalitis in children and immunocompetent adults. This organism has also been associated with encephalitis in recipients of transplanted organs from a donor with unrecognized infection. No effective treatment has been identified, and mortality approaches 100%.

Encephalitis can be caused by the raccoon pinworm *Baylisascaris procyonis*. Clues to the diagnosis include a history of raccoon exposure, especially of playing in or eating dirt potentially contaminated with raccoon feces. Most patients are children, and many have an associated eosinophilia.

Once nonviral causes of encephalitis have been excluded, the major diagnostic challenge is to distinguish HSV from other viruses that cause encephalitis. This distinction is particularly important because in virtually every other instance the therapy is supportive, whereas specific and effective antiviral therapy is available for HSV, and its efficacy is enhanced when it is instituted early in the course of infection. HSV encephalitis should be considered when clinical features suggesting involvement of the inferomedial frontotemporal regions of the brain are present, including prominent olfactory or gustatory hallucinations, anosmia, unusual or bizarre behavior or personality alterations, or memory disturbance. HSV encephalitis should always be suspected in patients with signs and symptoms consistent with acute encephalitis with focal findings on clinical examination, neuroimaging studies, or EEG. The diagnostic procedure of choice in these patients is CSF PCR analysis for HSV. A positive CSF PCR establishes the diagnosis, and a negative test dramatically reduces the likelihood of HSV encephalitis (see above).

The anatomic distribution of lesions may provide an additional clue to diagnosis. Patients with rapidly progressive encephalitis and prominent brainstem signs, symptoms, or neuroimaging abnormalities may be infected by flaviviruses (WNV, St. Louis encephalitis virus, Japanese encephalitis virus), HSV, rabies, or *L. monocytogenes*. Significant involvement of deep gray matter structures, including the basal ganglia and thalamus, should also suggest possible flavivirus infection. These patients may present clinically with prominent movement disorders (tremor, myoclonus) or parkinsonian features. Patients with WNV infection can also present with a poliomyelitis-like acute flaccid paralysis, as can patients infected with enterovirus 71 and, less commonly, other enteroviruses. Acute flaccid paralysis is characterized by the acute onset of a lower motor neuron type of weakness with flaccid tone, reduced or absent reflexes, and relatively preserved sensation. Despite an aggressive World Health Organization poliovirus eradication initiative, 1733 cases of wild-type poliovirus-induced poliomyelitis were reported worldwide in 2009, with 73% occurring in India and Nigeria. There have been recent small outbreaks of poliomyelitis associated with vaccine strains of virus that have reverted to virulence through mutation or recombination with circulating wild-type enteroviruses in Hispaniola, China, the Philippines, Indonesia, Nigeria, and Madagascar.

Epidemiologic factors may provide important clues to the diagnosis of viral meningitis or encephalitis. Particular attention should be paid to the season of the year; the geographic location and travel history; and possible exposure to animal bites or scratches, rodents, and ticks. Although transmission from the bite of an infected dog remains the most common cause of rabies worldwide, in the United

States very few cases of dog rabies occur, and the most common risk factor is exposure to bats —although a clear history of a bite or scratch is often lacking. The classic clinical presentation of encephalitic (furious) rabies is of fever, fluctuating consciousness, and autonomic hyperactivity. Phobic spasms of the larynx, pharynx, neck muscles, and diaphragm can be triggered by attempts to swallow water (*hydrophobia*) or by inspiration (*aerophobia*). Patients may also present with paralytic (dumb) rabies characterized by acute ascending paralysis. Rabies due to the bite of a bat has a different clinical presentation than classic rabies due to a dog or wolf bite. Patients present with focal neurologic deficits, myoclonus, seizures, and hallucinations; phobic spasms are not a typical feature. Patients with rabies have a CSF lymphocytic pleocytosis and may show areas of increased T2 signal abnormality in the brainstem, hippocampus, and hypothalamus. Diagnosis can be made by finding rabies virus antigen in brain tissue or in the neural innervation of hair follicles at the nape of the neck. PCR amplification of viral nucleic acid from CSF and saliva or tears may also enable diagnosis. Serology is frequently negative in both serum and CSF in the first week after onset of infection, which limits its acute diagnostic utility. No specific therapy is available, and cases are almost invariably fatal, with isolated survivors having devastating neurologic sequelae.

State public health authorities provide a valuable resource concerning isolation of particular agents in individual regions. Regular updates concerning the number, type, and distribution of cases of arboviral encephalitis can be found on the CDC and U.S. Geological Survey (USGS) websites (*http://www.cdc.gov* and *http://diseasemaps.usgs.gov*).

The major noninfectious etiologies that should be included in the differential diagnosis of acute encephalitis are nonvasculitic autoimmune inflammatory meningoencephalitis, which is frequently associated with serum antithyroid microsomal and antithyroglobulin antibodies (Hashimoto's encephalopathy); paraneoplastic and non-paraneoplastic encephalitis associated with antineuronal antibodies (Chap. 101); acute disseminated encephalomyelitis and related fulminant demyelinating disorders (Chap. 380); and lymphoma. Finally, Creutzfeldt-Jakob disease (Chap. 383) can rarely present in an explosive fashion mimicking viral encephalitis.

Treatment: Viral Encephalitis

Specific antiviral therapy should be initiated when appropriate. Vital functions, including respiration and blood pressure, should be monitored continuously and supported as required. In the initial stages of encephalitis, many patients will require care in an intensive care unit. Basic management and supportive therapy should include careful monitoring of ICP, fluid restriction, avoidance of hypotonic intravenous solutions, and suppression of fever. Seizures should be treated with standard anticonvulsant regimens, and prophylactic therapy should be considered in view of the high frequency of seizures in severe cases of encephalitis. As with all seriously ill, immobilized patients with altered levels of consciousness, encephalitis patients are at risk for aspiration pneumonia, stasis ulcers and decubiti, contractures, deep venous thrombosis and its complications, and infections of indwelling lines and catheters.

Acyclovir is of benefit in the treatment of HSV and should be started empirically in patients with suspected viral encephalitis, especially if focal features are present, while awaiting viral diagnostic studies. Treatment should be discontinued in patients found not to have HSV encephalitis, with the possible exception of patients with severe encephalitis due to VZV or EBV. HSV, VZV, and EBV all encode an enzyme, deoxypyrimidine (thymidine) kinase, that phosphorylates acyclovir to produce acyclovir-5'-monophosphate. Host cell enzymes then phosphorylate this compound to form a triphosphate derivative. It is the triphosphate that acts as an antiviral agent by inhibiting viral DNA polymerase and by causing premature termination of nascent viral DNA chains. The specificity of action depends on the fact that uninfected cells do not phosphorylate significant amounts of acyclovir to acyclovir-5'-monophosphate. A second level of specificity is provided by the fact that the acyclovir triphosphate is a more potent inhibitor of viral DNA polymerase than of the analogous host cell enzymes.

Adults should receive a dose of 10 mg/kg of acyclovir intravenously every 8 h (30 mg/kg per day total dose) for 14–21 days. CSF PCR can be repeated at the completion of this course, with PCR-positive patients receiving additional treatment, followed by a repeat CSF PCR test. Neonatal HSV CNS infection is less responsive to acyclovir therapy than HSV encephalitis in adults; it is recommended that neonates with HSV encephalitis receive 20 mg/kg of acyclovir every 8 h (60 mg/kg per day total dose) for a minimum of 21 days.

Prior to intravenous administration, acyclovir should be diluted to a concentration ≤7 mg/mL. (A 70-kg person would receive a dose of 700 mg, which would be diluted in a volume of 100 mL.) Each dose should be infused slowly over 1 h, rather than by rapid or bolus infusion, to minimize the risk of renal dysfunction. Care should be taken to avoid extravasation or intramuscular or subcutaneous administration. The alkaline pH of acyclovir can cause local inflammation and phlebitis (9%). Dose adjustment is required in patients with impaired renal glomerular filtration. Penetration into CSF is excellent, with average drug levels ~50% of serum levels. Complications of therapy include elevations in blood urea nitrogen and creatinine levels (5%), thrombocytopenia (6%), gastrointestinal toxicity (nausea, vomiting, diarrhea) (7%), and neurotoxicity (lethargy or obtundation, disorientation, confusion,

agitation, hallucinations, tremors, seizures) (1%). Acyclovir resistance may be mediated by changes in either the viral deoxypyrimidine kinase or DNA polymerase. To date, acyclovir-resistant isolates have not been a significant clinical problem in immunocompetent individuals. However, there have been reports of clinically virulent acyclovir-resistant HSV isolates from sites outside the CNS in immunocompromised individuals, including those with AIDS.

Oral antiviral drugs with efficacy against HSV, VZV, and EBV, including acyclovir, famciclovir, and valacyclovir, have not been evaluated in the treatment of encephalitis either as primary therapy or as supplemental therapy following completion of a course of parenteral acyclovir. A National Institute of Allergy and Infectious Disease (NIAID)/National Institute of Neurological Disorders and Stroke –sponsored phase III trial of supplemental oral valacyclovir therapy (2 g tid for 3 months) following the initial 14- to 21-day course of therapy with parenteral acyclovir is ongoing in patients with HSV encephalitis (www.clinicaltrials.gov, identifier NCT00031486); this may help clarify the role of extended oral antiviral therapy.

Ganciclovir and foscarnet, either alone or in combination, are often utilized in the treatment of CMV-related CNS infections, although their efficacy remains unproven. Cidofovir (see below) may provide an alternative in patients who fail to respond to ganciclovir and foscarnet, although data concerning its use in CMV CNS infections are extremely limited.

Ganciclovir is a synthetic nucleoside analogue of 2'-deoxyguanosine. The drug is preferentially phosphorylated by virus-induced cellular kinases. Ganciclovir triphosphate acts as a competitive inhibitor of the CMV DNA polymerase, and its incorporation into nascent viral DNA results in premature chain termination. Following intravenous administration, CSF concentrations of ganciclovir are 25 –70% of coincident plasma levels. The usual dose for treatment of severe neurologic illnesses is 5 mg/kg every 12 h given intravenously at a constant rate over 1 h. Induction therapy is followed by maintenance therapy of 5 mg/kg every day for an indefinite period. Induction therapy should be continued until patients show a decline in CSF pleocytosis and a reduction in CSF CMV DNA copy number on quantitative PCR testing (where available). Doses should be adjusted in patients with renal insufficiency. Treatment is often limited by the development of granulocytopenia and thrombocytopenia (20–25%), which may require reduction in or discontinuation of therapy. Gastrointestinal side effects, including nausea, vomiting, diarrhea, and abdominal pain, occur in ~20% of patients. Some patients treated with ganciclovir for CMV retinitis have developed retinal detachment, but the causal relationship to ganciclovir treatment is unclear. Valganciclovir is an orally bioavailable prodrug that can generate high serum levels of ganciclovir, although studies of its efficacy in treating CMV CNS infections are limited.

Foscarnet is a pyrophosphate analogue that inhibits viral DNA polymerases by binding to the pyrophosphate-binding site. Following intravenous infusion, CSF concentrations range from 15 to 100% of coincident plasma levels. The usual dose for serious CMV-related neurologic illness is 60 mg/kg every 8 h administered by constant infusion over 1 h. Induction therapy for 14–21 days is followed by maintenance therapy (60–120 mg/kg per day). Induction therapy may need to be extended in patients who fail to show a decline in CSF pleocytosis and a reduction in CSF CMV DNA copy number on quantitative PCR tests (where available). Approximately one-third of patients develop renal impairment during treatment, which is reversible following discontinuation of therapy in most, but not all, cases. This is often associated with elevations in serum creatinine and proteinuria and is less frequent in patients who are adequately hydrated. Many patients experience fatigue and nausea. Reduction in serum calcium, magnesium, and potassium occur in ~15% of patients and may be associated with tetany, cardiac rhythm disturbances, or seizures.

Cidofovir is a nucleotide analogue that is effective in treating CMV retinitis and equivalent to or better than ganciclovir in some experimental models of murine CMV encephalitis, although data concerning its efficacy in human CMV CNS disease are limited. The usual dose is 5 mg/kg intravenously once weekly for 2 weeks, then biweekly for two or more additional doses, depending on clinical response. Patients must be prehydrated with normal saline (e.g., 1 L over 1–2 h) prior to each dose and treated with probenecid (e.g., 1 g 3 h before cidofovir and 1 g 2 and 8 h after cidofovir). Nephrotoxicity is common; the dose should be reduced if renal function deteriorates.

Intravenous ribavirin (15–25 mg/kg per day in divided doses given every 8 h) has been reported to be of benefit in isolated cases of severe encephalitis due to California encephalitis (LaCrosse) virus. Ribavirin might be of benefit for the rare patients, typically infants or young children, with severe adenovirus or rotavirus encephalitis and in patients with encephalitis due to LCMV or other arenaviruses. However, clinical trials are lacking. Hemolysis, with resulting anemia, has been the major side effect limiting therapy.

No specific antiviral therapy of proven efficacy is currently available for treatment of WNV encephalitis. Patients have been treated with ∝-interferon, ribavirin, WNV-specific antisense oligonucleotides (ClinicalTrials.gov, identifier NCT0091845), an Israeli IVIg preparation that contains high-titer anti-WNV antibody (Omr-IgG-am) (ClinicalTrials.gov, identifier NCT00069316 and 0068055), and humanized monoclonal antibodies directed against the viral envelope glycoprotein (ClinicalTrials.gov, identifier NCT00927953 and 00515385). WNV chimeric vaccines, in which WNV envelope and premembrane

proteins are inserted into the background of another flavivirus, are already undergoing human clinical testing for safety and immunogenicity (ClinicalTrials.gov, identifier NCT00746798 and 00442169). Both chimeric and killed inactivated WNV vaccines have been found to be safe and effective in preventing equine WNV infection, and several effective flavivirus vaccines are already in human use, creating optimism that a safe and effective human WNV vaccine can also be developed.

Sequelae

There is considerable variation in the incidence and severity of sequelae in patients surviving viral encephalitis. In the case of EEE virus infection, nearly 80% of survivors have severe neurologic sequelae. At the other extreme are infections due to EBV, California encephalitis virus, and Venezuelan equine encephalitis virus, where severe sequelae are unusual. For example, approximately 5–15% of children infected with LaCrosse virus have a residual seizure disorder, and 1% have persistent hemiparesis. Detailed information about sequelae in patients with HSV encephalitis treated with acyclovir is available from the NIAID-Collaborative Antiviral Study Group (CASG) trials. Of 32 acyclovir-treated patients, 26 survived (81%). Of the 26 survivors, 12 (46%) had no or only minor sequelae, 3 (12%) were moderately impaired (gainfully employed but not functioning at their previous level), and 11 (42%) were severely impaired (requiring continuous supportive care). The incidence and severity of sequelae were directly related to the age of the patient and the level of consciousness at the time of initiation of therapy. Patients with severe neurologic impairment (Glasgow coma score 6) at initiation of therapy either died or survived with severe sequelae. Young patients (<30 years) with good neurologic function at initiation of therapy did substantially better (100% survival, 62% with no or mild sequelae) compared with their older counterparts (>30 years; 64% survival, 57% no or mild sequelae). Some recent studies using quantitative HSV CSF PCR tests indicate that clinical outcome following treatment also correlates with the amount of HSV DNA present in CSF at the time of presentation. Many patients with WNV infection have sequelae, including cognitive impairment; weakness; and hyper- or hypokinetic movement disorders, including tremor, myoclonus, and parkinsonism. In a large longitudinal study of prognosis in 156 patients with WNV infection, the mean time to achieve recovery (defined as 95% of maximal predicted score on specific validated tests) was 112–148 days for fatigue, 121–175 days for physical function, 131–139 days for mood, and 302–455 days for mental function (the longer interval in each case representing patients with neuroinvasive disease).

SUBACUTE MENINGITIS

Clinical Manifestations

Patients with subacutemeningitis typically have an unrelenting headache, stiff neck, low-grade fever, and lethargy for days to several weeks before they present for evaluation. Cranial nerve abnormalities and night sweats may be present. **This syndrome overlaps that of chronic meningitis, discussed in detail in Chap. 382.**

Etiology

Common causative organisms include *M. tuberculosis*, *C. neoformans*, *H. capsulatum*, *C. immitis*, and *T. pallidum*. Initial infection with *M. tuberculosis* is acquired by inhalation of aerosolized droplet nuclei. Tuberculous meningitis in adults does not develop acutely from hematogenous spread of tubercle bacilli to the meninges. Rather, millet seed–sized (miliary) tubercles form in the parenchyma of the brain during hematogenous dissemination of tubercle bacilli in the course of primary infection. These tubercles enlarge and are usually caseating. The propensity for a caseous lesion to produce meningitis is determined by its proximity to the subarachnoid space (SAS) and the rate at which fibrous encapsulation develops. Subependymal caseous foci cause meningitis via discharge of bacilli and tuberculous antigens into the SAS. Mycobacterial antigens produce an intense inflammatory reaction that leads to the production of a thick exudate that fills the basilar cisterns and surrounds the cranial nerves and major blood vessels at the base of the brain.

Fungal infections are typically acquired by the inhalation of airborne fungal spores. The initial pulmonary infection may be asymptomatic or present with fever, cough, sputum production, and chest pain. The pulmonary infection is often self-limited. A localized pulmonary fungal infection can then remain dormant in the lungs until there is an abnormality in cell-mediated immunity that allows the fungus to reactivate and disseminate to the CNS. The most common pathogen causing fungal meningitis is *C. neoformans*. This fungus is found worldwide in soil and bird excreta. *H. capsulatum* is endemic to the Ohio and Mississippi River valleys of the central United States and to parts of Central and South America. *C. immitis* is endemic to the desert areas of the southwest United States, northern Mexico, and Argentina.

Syphilis is a sexually transmitted disease that is manifest by the appearance of a painless chancre at the site of inoculation. *T. pallidum* invades the CNS early in the course of syphilis. Cranial nerves VII and VIII are most frequently involved.

Laboratory Diagnosis

The classic CSF abnormalities in tuberculous meningitis are as follows: (1) elevated opening pressure,

(2) lymphocytic pleocytosis (10–500 cells/μL), (3) elevated protein concentration in the range of 1–5 g/L, and (4) decreased glucose concentration in the range of 1.1–2.2 mmol/L (20–40 mg/dL). *The combination of unrelenting headache, stiff neck, fatigue, night sweats, and fever with a CSF lymphocytic pleocytosis and a mildly decreased glucose concentration is highly suspicious for tuberculous meningitis.* The last tube of fluid collected at LP is the best tube to send for a smear for acid-fast bacilli (AFB). If there is a pellicle in the CSF or a cobweb-like clot on the surface of the fluid, AFB can best be demonstrated in a smear of the clot or pellicle. Positive smears are typically reported in only 10–40% of cases of tuberculous meningitis in adults. Cultures of CSF take 4–8 weeks to identify the organism and are positive in ~50% of adults. Culture remains the gold standard to make the diagnosis of tuberculous meningitis. PCR for the detection of *M. tuberculosis* DNA should be sent on CSF if available, but the sensitivity and specificity on CSF have not been defined. The Centers for Disease Control and Prevention recommend the use of nucleic acid amplification tests for the diagnosis of pulmonary tuberculosis.

The characteristic CSF abnormalities in fungal meningitis are a mononuclear or lymphocytic pleocytosis, an increased protein concentration, and a decreased glucose concentration. There may be eosinophils in the CSF in *C. immitis* meningitis. Large volumes of CSF are often required to demonstrate the organism on india ink smear or grow the organism in culture. If spinal fluid examined by LP on two separate occasions fails to yield an organism, CSF should be obtained by high-cervical or cisternal puncture.

The cryptococcal polysaccharide antigen test is a highly sensitive and specific test for cryptococcal meningitis. A reactive CSF cryptococcal antigen test establishes the diagnosis. The detection of the histoplasma polysaccharide antigen in CSF establishes the diagnosis of a fungal meningitis but is not specific for meningitis due to *H. capsulatum*. It may be falsely positive in coccidioidal meningitis. The CSF complement fixation antibody test is reported to have a specificity of 100% and a sensitivity of 75% for coccidioidal meningitis.

The diagnosis of syphilitic meningitis is made when a reactive serum treponemal test [fluorescent treponemal antibody absorption test (FTA-ABS) or microhemagglutination assay–*T. pallidum* (MHA-TP)] is associated with a CSF lymphocytic or mononuclear pleocytosis and an elevated protein concentration, or when the CSF Venereal Disease Research Laboratory (VDRL) is positive. A reactive CSF FTA-ABS is not definitive evidence of neurosyphilis. The CSF FTA-ABS can be falsely positive from blood contamination. A negative CSF VDRL does not rule out neurosyphilis. A negative CSF FTA-ABS or MHA-TP rules out neurosyphilis.

Treatment: Subacute Meningitis

Empirical therapy of tuberculous meningitis is often initiated on the basis of a high index of suspicion without adequate laboratory support. Initial therapy is a combination of isoniazid (300 mg/d), rifampin (10 mg/kg per day), pyrazinamide (30 mg/kg per day in divided doses), ethambutol (15–25 mg/kg per day in divided doses), and pyridoxine (50 mg/d). When the antimicrobial sensitivity of the *M. tuberculosis* isolate is known, ethambutol can be discontinued. If the clinical response is good, pyrazinamide can be discontinued after 8 weeks and isoniazid and rifampin continued alone for the next 6–12 months. A 6-month course of therapy is acceptable, but therapy should be prolonged for 9–12 months in patients who have an inadequate resolution of symptoms of meningitis or who have positive mycobacterial cultures of CSF during the course of therapy. Dexamethasone therapy is recommended for HIV-negative patients with tuberculous meningitis. The dose is 12–16 mg per day for 3 weeks, then tapered over 3 weeks.

Meningitis due to *C. neoformans* in non-HIV, nontransplant patients is treated with induction therapy with amphotericin B (AmB) (0.7 mg/kg IV per day) plus flucytosine (100 mg/kg per day in four divided doses) for at least 4 weeks if CSF culture results are negative after 2 weeks of treatment. Therapy should be extended for a total of 6 weeks in the patient with neurologic complications. Induction therapy is followed by consolidation therapy with fluconazole 400 mg per day for 8 weeks. Organ transplant recipients are treated with liposomal AmB (3–4 mg/kg per day) or AmB lipid complex (ABLC) 5 mg/kg per day plus flucytosine (100 mg/kg per day in four divided doses) for at least 2 weeks or until CSF culture is sterile. Follow CSF yeast cultures for sterilization rather than the cryptococcal antigen titer. This treatment is followed by an 8- to 10-week course of fluconazole [400–800 mg/d (6–12 mg/kg) PO]. If the CSF culture is sterile after 10 weeks of acute therapy, the dose of fluconazole is decreased to 200 mg/d for 6 months to a year. Patients with HIV infection are treated with AmB or a lipid formulation plus flucytosine for at least 2 weeks, followed by fluconazole for a minimum of 8 weeks. HIV-infected patients may require indefinite maintenance therapy with fluconazole 200 mg/day. Meningitis due to *H. capsulatum* is treated with AmB (0.7–1.0 mg/kg per day) for 4–12 weeks. A total dose of 30 mg/kg is recommended. Therapy with AmB is not discontinued until fungal cultures are sterile. After completing a course of AmB, maintenance therapy with itraconazole 200 mg twice daily is initiated and continued for at least 6 months to a year. *C. immitis* meningitis is treated with either high-dose fluconazole (1000 mg daily) as monotherapy or intravenous AmB (0.5–0.7 mg/kg per day) for >4 weeks. Intrathecal AmB (0.25–0.75 mg/d three times weekly) may be required to eradicate the infection. Lifelong therapy with fluconazole (200–400 mg daily) is recommended to prevent relapse.

AmBisome (5 mg/kg per day) or AmB lipid complex (5 mg/kg per day) can be substituted for AmB in patients who have or who develop significant renal dysfunction. The most common complication of fungal meningitis is hydrocephalus. Patients who develop hydrocephalus should receive a CSF diversion device. A ventriculostomy can be used until CSF fungal cultures are sterile, at which time the ventriculostomy is replaced by a ventriculoperitoneal shunt.

Syphilitic meningitis is treated with aqueous penicillin G in a dose of 3–4 million units intravenously every 4 h for 10–14 days. An alternative regimen is 2.4 million units of procaine penicillin G intramuscularly daily with 500 mg of oral probenecid four times daily for 10–14 days. Either regimen is followed with 2.4 million units of benzathine penicillin G intramuscularly once a week for 3 weeks. The standard criterion for treatment success is reexamination of the CSF. The CSF should be reexamined at 6-month intervals for 2 years. The cell count is expected to normalize within 12 months, and the VDRL titer to decrease by two dilutions or revert to nonreactive within 2 years of completion of therapy. Failure of the CSF pleocytosis to resolve or an increase in the CSF VDRL titer by two or more dilutions requires retreatment.

CHRONIC ENCEPHALITIS

Progressive Multifocal Leukoencephalopathy

CLINICAL FEATURES AND PATHOLOGY

Progressive multifocal leukoencephalopathy (PML) is characterized pathologically by multifocal areas of demyelination of varying size distributed throughout the brain but sparing the spinal cord and optic nerves. In addition to demyelination, there are characteristic cytologic alterations in both astrocytes and oligodendrocytes. Astrocytes are enlarged and contain hyperchromatic, deformed, and bizarre nuclei and frequent mitotic figures. Oligodendrocytes have enlarged, densely staining nuclei that contain viral inclusions formed by crystalline arrays of JC virus (JCV) particles. Patients often present with visual deficits (45%), typically a homonymous hemianopia; mental impairment (38%) (dementia, confusion, personality change); weakness, including hemi- or monoparesis; and ataxia. Seizures occur in ~20% of patients, predominantly in those with lesions abutting the cortex.

Almost all patients have an underlying immunosuppressive disorder. In recent series, the most common associated conditions were AIDS (80%), hematologicmalignancies (13%), transplant recipients (5%), and chronic inflammatory diseases (2%). It has been estimated that up to 5% of AIDS patients will develop PML. There have been more than 30 reported cases of PML occurring in patients being treated for multiple sclerosis and inflammatory bowel disease with natalizumab, a humanized monoclonal antibody that inhibits lymphocyte trafficking into CNS and bowel mucosa by binding to $\alpha 4$ integrins. Risk in these patients has been estimated at 1 PML case per 1000 treated patients after a mean of 18 months of therapy. Additional cases have been reported in patients receiving other humanized monoclonal antibodies with immunomodulatory activity including efalizumab and rituximab. The basic clinical and diagnostic features appear to be similar to those seen in PML related to HIV and other forms of immunosuppression.

DIAGNOSTIC STUDIES

The diagnosis of PML is frequently suggested by MRI. MRI reveals multifocal asymmetric, coalescing white matter lesions located periventricularly, in the centrum semiovale, in the parietal-occipital region, and in the cerebellum. These lesions have increased signal on T2 and FLAIR images and decreased signal on T1-weighted images. PML lesions are classically nonenhancing (90%) but may rarely show ring enhancement, especially in more immunocompetent patients. PML lesions are not typically associated with edema or mass effect. CT scans, which are less sensitive than MRI for the diagnosis of PML, often show hypodense nonenhancing white matter lesions.

The CSF is typically normal, although mild elevation in protein and/or IgG may be found. Pleocytosis occurs in <25% of cases, is predominantly mononuclear, and rarely exceeds 25 cells/μL. PCR amplification of JCV DNA from CSF has become an important diagnostic tool. The presence of a positive CSF PCR for JCV DNA in association with typical MRI lesions in the appropriate clinical setting is diagnostic of PML, reflecting the assay's relatively high specificity (92–100%); however, sensitivity is variable and a negative CSF PCR does not exclude the diagnosis. In HIV-negative patients and HIV-positive patients not receiving highly active antiviral therapy (HAART), sensitivity is likely 70–90%. In HAART-treated patients, sensitivity may be closer to 60%, reflecting the lower JCV CSF viral load in this relatively more immunocompetent group. Studies with quantitative JCV CSF PCR indicate that patients with low JCV loads (<100 copies/μL) have a generally better prognosis than those with higher viral loads. Patients with negative CSF PCR studies may require brain biopsy for definitive diagnosis. In biopsy or necropsy specimens of brain, JCV antigen and nucleic acid can be detected by immunocytochemistry, in situ hybridization, or PCR amplification. Detection of JCV antigen or genomic material should only be considered diagnostic of PML if accompanied by characteristic pathologic changes, since both antigen and genomic material have been found in the brains of normal patients.

Serologic studies are of no utility in diagnosis due to high basal seroprevalence level (>80%).

TREATMENT: PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY

No effective therapy for PML is available. There are case reports of potential beneficial effects of the 5-HT2a receptor antagonist mirtazapine, which may inhibit binding of JCV to its receptor on oligodendrocytes. Retrospective noncontrolled studies have also suggested a possible beneficial effect of treatment with interferon-alpha. Neither of these agents has been tested in randomized controlled clinical trials. A clinical trial to evaluate the efficacy of the antimalarial drug mefloquine, which inhibits JCV replication in cell culture, is underway (www.clinicaltrials.com, identifier NCT00746941). Intravenous and/or intrathecal cytarabine were not shown to be of benefit in a randomized controlled trial in HIV-associated PML, although some experts suggest that cytarabine may have therapeutic efficacy in situations where breakdown of the blood-brain-barrier allow sufficient CSF penetration. A randomized controlled trial of cidofovir in HIV-associated PML also failed to show significant benefit. Since PML almost invariably occurs in immunocompromised individuals, any therapeutic interventions designed to enhance or restore immunocompetence should be considered. Perhaps the most dramatic demonstration of this is disease stabilization and, in rare cases, improvement associated with the improvement in the immune status of HIV-positive patients with AIDS following institution of HAART. In HIV-positive PML patients treated with HAART, 1-year survival is ~ 50%, although up to 80% of survivors may have significant neurologic sequelae. HIV-positive PML patients with higher CD4 counts (>300/μL^3) and low or nondetectable HIV viral loads have a better prognosis than those with lower CD4 counts and higher viral loads. Although institution of HAART enhances survival in HIV + PML patients, the associated immune reconstitution in patients with an underlying opportunistic infection such as PML may also result in a severe CNS inflammatory syndrome [immune reconstitution inflammatory syndrome (IRIS)] associated with clinical worsening, CSF pleocytosis, and the appearance of new enhancing MRI lesions. Patients receiving natalizumab or other immunomodulatory antibodies, who are suspected of having PML, should have therapy halted and circulating antibodies removed by plasma exchange.

Subacute Sclerosing Panencephalitis (SSPE)

SSPE is a rare chronic, progressive demyelinating disease of the CNS associated with a chronic nonpermissive infection of brain tissue with measles virus. The frequency has been estimated at 1 in 100,000–500,000 measles cases. An average of five cases per year are reported in the United States. The incidence has declined dramatically since the introduction of a measles vaccine. Most patients give a history of primary measles infection at an early age (2 years), which is followed after a latent interval of 6–8 years by the development of a progressive neurologic disorder. Some 85% of patients are between 5 and 15 years old at diagnosis. Initial manifestations include poor school performance and mood and personality changes. Typical signs of a CNS viral infection, including fever and headache, do not occur. As the disease progresses, patients develop progressive intellectual deterioration, focal and/or generalized seizures, myoclonus, ataxia, and visual disturbances. In the late stage of the illness, patients are unresponsive, quadriparetic, and spastic, with hyperactive tendon reflexes and extensor plantar responses.

DIAGNOSTIC STUDIES

MRI is often normal early, although areas of increased T2 signal develop in the white matter of the brain and brainstem as disease progresses. The EEG may initially show only nonspecific slowing, but with disease progression, patients develop a characteristic periodic pattern with bursts of high-voltage, sharp, slow waves every 3–8 s, followed by periods of attenuated ("flat") background. The CSF is acellular with a normal or mildly elevated protein concentration and a markedly elevated gamma globulin level (>20% of total CSF protein). CSF antimeasles antibody levels are invariably elevated, and oligoclonal antimeasles antibodies are often present. Measles virus can be cultured from brain tissue using special cocultivation techniques. Viral antigen can be identified immunocytochemically, and viral genome can be detected by in situ hybridization or PCR amplification.

TREATMENT: SUBACUTE SCLEROSING PANENCEPHALITIS

No definitive therapy for SSPE is available. Treatment with isoprinosine (Inosiplex, 100 mg/kg per day), alone or in combination with intrathecal or intraventricular alpha interferon, has been reported to prolong survival and produce clinical improvement in some patients but has never been subjected to a controlled clinical trial.

Progressive Rubella Panencephalitis

This is an extremely rare disorder that primarily affects males with congenital rubella syndrome, although isolated cases have been reported following childhood rubella. After a latent period of 8–19 years, patients develop progressive neurologic deterioration. The manifestations are similar to those seen in SSPE. CSF shows a mild lymphocytic pleocytosis, slightly elevated protein concentration, markedly increased gamma globulin, and rubella virus–specific oligoclonal bands. No therapy is available. Universal prevention of both congenital and childhood rubella through the use of the available live attenuated rubella vaccine would be expected to eliminate the disease.

BRAIN ABSCESS

Definition

A brain abscess is a focal, suppurative infection within the brain parenchyma, typically surrounded by a vascularized capsule. The term *cerebritis* is often employed to describe a nonencapsulated brain abscess.

Epidemiology

A bacterial brain abscess is a relatively uncommon intracranial infection, with an incidence of ~0.3–1.3:100,000 persons per year. Predisposing conditions include otitis media and mastoiditis, paranasal sinusitis, pyogenic infections in the chest or other body sites, penetrating head trauma or neurosurgical procedures, and dental infections. In immunocompetent individuals the most important pathogens are *Streptococcus* spp. [anaerobic, aerobic, and viridans (40%)], Enterobacteriaceae [*Proteus* spp., *E. coli* sp., *Klebsiella* spp. (25%)], anaerobes [e.g., *Bacteroides* spp., *Fusobacterium* spp. (30%)], and staphylococci (10%). In immunocompromised hosts with underlying HIV infection, organ transplantation, cancer, or immunosuppressive therapy, most brain abscesses are caused by *Nocardia* spp., *Toxoplasma gondii*, *Aspergillus* spp., *Candida* spp., and *C. neoformans*. In Latin America and in immigrants from Latin America, the most common cause of brain abscess is *Taenia solium* (neurocysticercosis). In India and the Far East, mycobacterial infection (tuberculoma) remains a major cause of focal CNS mass lesions.

Etiology

A brain abscess may develop (1) by direct spread from a contiguous cranial site of infection, such as paranasal sinusitis, otitis media, mastoiditis, or dental infection; (2) following head trauma or a neurosurgical procedure; or (3) as a result of hematogenous spread from a remote site of infection. In up to 25% of cases, no obvious primary source of infection is apparent (cryptogenic brain abscess).

Approximately one-third of brain abscesses are associated with otitis media and mastoiditis, often with an associated cholesteatoma. Otogenic abscesses occur predominantly in the temporal lobe (55–75%) and cerebellum (20–30%). In some series, up to 90% of cerebellar abscesses are otogenic. Common organisms include streptococci, *Bacteroides* spp., *Pseudomonas* spp., *Haemophilus* spp., and Enterobacteriaceae. Abscesses that develop as a result of direct spread of infection from the frontal, ethmoidal, or sphenoidal sinuses and those that occur due to dental infections are usually located in the frontal lobes. Approximately 10% of brain abscesses are associated with paranasal sinusitis, and this association is particularly strong in young males in their second and third decades of life. The most common pathogens in brain abscesses associated with paranasal sinusitis are streptococci (especially *S. milleri*), *Haemophilus* spp., *Bacteroides* spp., *Pseudomonas* spp., and *S. aureus*. Dental infections are associated with ~2% of brain abscesses, although it is often suggested that many "cryptogenic" abscesses are in fact due to dental infections. The most common pathogens in this setting are streptococci, staphylococci, *Bacteroides* spp., and *Fusobacterium* spp.

Hematogenous abscesses account for ~25% of brain abscesses. Hematogenous abscesses are often multiple, and multiple abscesses often (50%) have a hematogenous origin. These abscesses show a predilection for the territory of the middle cerebral artery (i.e., posterior frontal or parietal lobes). Hematogenous abscesses are often located at the junction of the gray and white matter and are often poorly encapsulated. The microbiology of hematogenous abscesses is dependent on the primary source of infection. For example, brain abscesses that develop as a complication of infective endocarditis are often due to viridans streptococci or *S. aureus*. Abscesses associated with pyogenic lung infections such as lung abscess or bronchiectasis are often due to streptococci, staphylococci, *Bacteroides* spp., *Fusobacterium* spp., or Enterobacteriaceae. Abscesses that follow penetrating head trauma or neurosurgical procedures are frequently due to methicillin-resistant *S. aureus* (MRSA), *S. epidermidis*, Enterobacteriaceae, *Pseudomonas* spp., and *Clostridium* spp. Enterobacteriaceae and *P. aeruginosa* are important causes of abscesses associated with urinary sepsis. Congenital cardiac malformations that produce a right-to-left shunt, such as tetralogy of Fallot, patent ductus arteriosus, and atrial and ventricular septal defects, allow bloodborne bacteria to bypass the pulmonary capillary bed and reach the brain. Similar phenomena can occur with pulmonary arteriovenous malformations. The decreased arterial oxygenation and saturation from the right-to-left shunt and polycythemia may cause focal areas of cerebral ischemia, thus providing a nidus for microorganisms that bypassed the pulmonary circulation to multiply and form an abscess. Streptococci are the most common pathogens in this setting.

Pathogenesis and Histopathology

Results of experimental models of brain abscess formation suggest that for bacterial invasion of brain parenchyma to occur, there must be preexisting or concomitant areas of ischemia, necrosis, or hypoxemia in brain tissue. The intact brain parenchyma is relatively resistant to infection. Once bacteria have established infection, brain abscess frequently evolves through a series of stages, influenced by the nature of the infecting organism and by the immunocompetence of the host. The early cerebritis stage (days 1–3) is characterized by a perivascular infiltration of inflammatory cells, which surround a central core of coagulative necrosis. Marked edema surrounds the lesion at this

stage. In the late cerebritis stage (days 4–9), pus formation leads to enlargement of the necrotic center, which is surrounded at its border by an inflammatory infiltrate of macrophages and fibroblasts. A thin capsule of fibroblasts and reticular fibers gradually develops, and the surrounding area of cerebral edema becomes more distinct than in the previous stage. The third stage, early capsule formation (days 10–13), is characterized by the formation of a capsule that is better developed on the cortical than on the ventricular side of the lesion. This stage correlates with the appearance of a ring-enhancing capsule on neuroimaging studies. The final stage, late capsule formation (day 14 and beyond), is defined by a well-formed necrotic center surrounded by a dense collagenous capsule. The surrounding area of cerebral edema has regressed, but marked gliosis with large numbers of reactive astrocytes has developed outside the capsule. This gliotic process may contribute to the development of seizures as a sequelae of brain abscess.

Clinical Presentation

A brain abscess typically presents as an expanding intracranial mass lesion rather than as an infectious process. Although the evolution of signs and symptoms is extremely variable, ranging from hours to weeks or even months, most patients present to the hospital 11–12 days following onset of symptoms. The classic clinical triad of headache, fever, and a focal neurologic deficit is present in <50% of cases. The most common symptom in patients with a brain abscess is headache, occurring in >75% of patients. The headache is often characterized as a constant, dull, aching sensation, either hemicranial or generalized, and it becomes progressively more severe and refractory to therapy. Fever is present in only 50% of patients at the time of diagnosis, and its absence should not exclude the diagnosis. The new onset of focal or generalized seizure activity is a presenting sign in 15–35% of patients. Focal neurologic deficits including hemiparesis, aphasia, or visual field defects are part of the initial presentation in >60% of patients.

The clinical presentation of a brain abscess depends on its location, the nature of the primary infection if present, and the level of the ICP. Hemiparesis is the most common localizing sign of a frontal lobe abscess. A temporal lobe abscess may present with a disturbance of language (dysphasia) or an upper homonymous quadrantanopia. Nystagmus and ataxia are signs of a cerebellar abscess. Signs of raised ICP—papilledema, nausea and vomiting, and drowsiness or confusion—can be the dominant presentation of some abscesses, particularly those in the cerebellum. Meningismus is not present unless the abscess has ruptured into the ventricle or the infection has spread to the subarachnoid space.

Diagnosis

Diagnosis is made by neuroimaging studies. MRI **(Fig. 381-4)** is better than CT for demonstrating abscesses in the early (cerebritis) stages and is superior to CT for identifying abscesses in the posterior fossa. Cerebritis appears on MRI as an area of low-signal intensity on T1-weighted images with irregular postgadolinium enhancement and as an area of increased signal intensity on T2-weighted images. Cerebritis is often not visualized by CT scan but, when present, appears as an area of hypodensity. On a contrast-enhanced CT scan, a mature brain abscess appears as a focal area of hypodensity surrounded by ring enhancement with surrounding edema (hypodensity). On contrast-enhanced T1-weighted MRI, a mature brain abscess has a capsule that enhances surrounding a hypodense center and surrounded by a hypodense area of edema. On T2-weighted MRI, there is a hyperintense central area of pus surrounded by a well-defined hypointense capsule and a hyperintense surrounding area of edema. It is important to recognize that the CT and MR appearance, particularly of the capsule, may be altered by treatment with glucocorticoids. The distinction between a brain abscess and other focal CNS lesions such as primary or metastatic tumors may be facilitated by the use of diffusion-weighted imaging sequences on which brain abscesses typically show increased signal due to restricted diffusion.

Figure 381-4

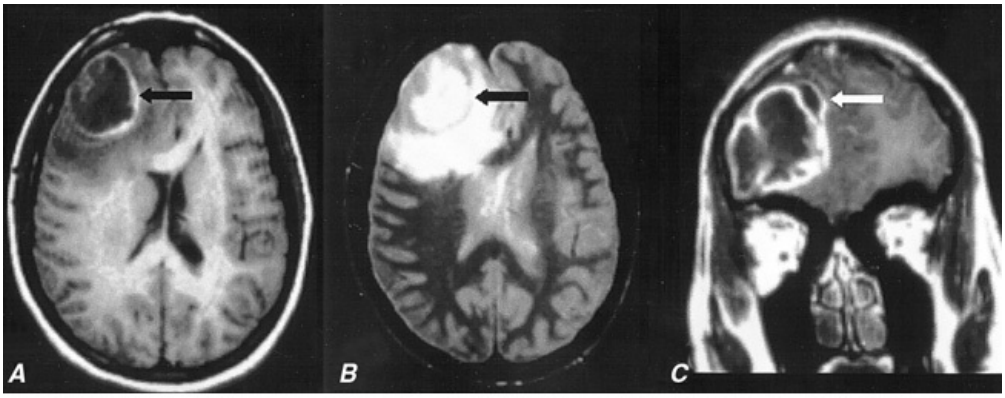

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Pneumococcal brain abscess. Note that the abscess wall has hyperintense signal on the axial T1-weighted MRI (***A***, *black arrow*), hypointense signal on the axial proton density images (***B***, *black arrow*), and enhances prominently after gadolinium administration on the coronal T1-weighted image (***C***). The abscess is surrounded by a large amount of vasogenic edema and has a small "daughter" abscess (***C***, *white arrow*). (*Courtesy of Joseph Lurito, MD; with permission* .)

Microbiologic diagnosis of the etiologic agent is most accurately determined by Gram's stain and culture of abscess material obtained by CT-guided stereotactic needle aspiration. Aerobic and anaerobic bacterial cultures and mycobacterial and fungal cultures should be obtained. Up to 10% of patients will also have positive blood cultures. LP should not be performed in patients with known or suspected focal intracranial infections such as abscess or empyema; CSF analysis contributes nothing to diagnosis or therapy, and LP increases the risk of herniation.

Additional laboratory studies may provide clues to the diagnosis of brain abscess in patients with a CNS mass lesion. About 50% of patients have a peripheral leukocytosis, 60% an elevated ESR, and 80% an elevated C-reactive protein. Blood cultures are positive in ~10% of cases overall but may be positive in >85% of patients with abscesses due to *Listeria*.

Differential Diagnosis

Conditions that can cause headache, fever, focal neurologic signs, and seizure activity include brain abscess, subdural empyema, bacterial meningitis, viral meningoencephalitis, superior sagittal sinus thrombosis, and acute disseminated encephalomyelitis. When fever is absent, primary and metastatic brain tumors become the major differential diagnosis. Less commonly, cerebral infarction or hematoma can have an MRI or CT appearance resembling brain abscess.

Treatment: Brain Abscess

Optimal therapy of brain abscesses involves a combination of high-dose parenteral antibiotics and neurosurgical drainage. Empirical therapy of community-acquired brain abscess in an immunocompetent patient typically includes a third- or fourth-generation cephalosporin (e.g., cefotaxime, ceftriaxone, or cefepime) and metronidazole (see Table 381-1 for antibiotic dosages). In patients with penetrating head trauma or recent neurosurgical procedures, treatment should include ceftazidime as the third-generation cephalosporin to enhance coverage of *Pseudomonas* spp. and vancomycin for coverage of staphylococci. Meropenem plus vancomycin also provides good coverage in this setting.

Aspiration and drainage of the abscess under stereotactic guidance are beneficial for both diagnosis and therapy. Empirical antibiotic coverage should be modified based on the results of Gram's stain and culture of the abscess contents. Complete excision of a bacterial abscess via craniotomy or craniectomy is generally reserved for multiloculated abscesses or those in which stereotactic aspiration is unsuccessful.

Medical therapy alone is not optimal for treatment of brain abscess and should be reserved for patients whose abscesses are neurosurgically inaccessible, for patients with small (<2–3 cm) or nonencapsulated abscesses (cerebritis), and patients whose condition is too tenuous to allow performance of a neurosurgical procedure. All patients should receive a minimum of 6–8 weeks of parenteral antibiotic therapy. The role, if any, of supplemental oral antibiotic therapy following completion of a standard course of parenteral therapy has never been adequately studied.

In addition to surgical drainage and antibiotic therapy, patients should receive prophylactic anticonvulsant therapy because of the high risk (~35%) of focal or generalized seizures. Anticonvulsant therapy is continued for at least 3 months after resolution of the abscess, and decisions regarding withdrawal are then based on the EEG. If the EEG is abnormal, anticonvulsant therapy should be continued. If the EEG is normal, anticonvulsant therapy can be slowly withdrawn, with close follow-up and repeat EEG after the medication has been discontinued.

Glucocorticoids should not be given routinely to patients with brain abscesses. Intravenous dexamethasone therapy (10 mg every 6 h) is usually reserved for patients with substantial periabscess edema and associated mass effect and increased ICP. Dexamethasone should be tapered as rapidly as possible to avoid delaying the natural process of encapsulation of the abscess.

Serial MRI or CT scans should be obtained on a monthly or twice-monthly basis to document resolution of the abscess. More frequent studies (e.g., weekly) are probably warranted in the subset of patients who are receiving antibiotic therapy alone. A small amount of enhancement may remain for months after the abscess has been successfully treated.

Prognosis

The mortality rate of brain abscess has declined in parallel with the development of enhanced neuroimaging techniques, improved neurosurgical procedures for stereotactic aspiration, and improved antibiotics. In modern series, the mortality rate is typically <15%. Significant sequelae, including seizures, persisting weakness, aphasia, or mental impairment, occur in ≥20% of survivors.

NONBACTERIAL CAUSES OF INFECTIOUS FOCAL CNS LESIONS

Etiology

Neurocysticercosis is the most common parasitic disease of the CNS worldwide. Humans acquire cysticercosis by the ingestion of food contaminated with the eggs of the parasite *T. solium*. Toxoplasmosis is a parasitic disease caused by *T. gondii* and acquired from the ingestion of undercooked meat and from handling cat feces.

Clinical Presentation

The most common manifestation of neurocysticercosis is new-onset partial seizures with or without secondary generalization. Cysticerci may develop in the brain parenchyma and cause seizures or focal neurologic deficits. When present in the subarachnoid or ventricular spaces, cysticerci can produce increased ICP by interference with CSF flow. Spinal cysticerci can mimic the presentation of intraspinal tumors. When the cysticerci first lodge in the brain, they frequently cause little in the way of an inflammatory response. As the cysticercal cyst degenerates, it elicits an inflammatory response that may present clinically as a seizure. Eventually the cyst dies, a process that may take several years and is typically associated with resolution of the inflammatory response and, often, abatement of seizures.

Primary *Toxoplasma* infection is often asymptomatic. However, during this phase parasites may spread to the CNS, where they become latent. Reactivation of CNS infection is almost exclusively associated with immunocompromised hosts, particularly those with HIV infection. During this phase patients present with headache, fever, seizures, and focal neurologic deficits.

Diagnosis

The lesions of neurocysticercosis are readily visualized by MRI or CT scans. Lesions with viable parasites appear as cystic lesions. The scolex can often be visualized on MRI. Lesions may appear as contrast-enhancing lesions surrounded by edema. A very early sign of cyst death is hypointensity of the vesicular fluid on T2-weighted images when compared with CSF. Parenchymal brain calcifications are the most common finding and evidence that the parasite is no longer viable. MRI findings of toxoplasmosis consist of multiple lesions in the deep white matter, the thalamus, and basal ganglia and at the gray-white junction in the cerebral hemispheres. With contrast administration, the majority of the lesions enhance in a ringed, nodular, or homogeneous pattern and are surrounded by edema. In the presence of the characteristic neuroimaging abnormalities of *T. gondii* infection, serum IgG antibody to *T. gondii* should be obtained and, when positive, the patient should be treated.

Treatment: Infectious Focal CNS Lesions

Anticonvulsant therapy is initiated when the patient with neurocysticercosis presents with a seizure. There is controversy about whether or not anthelmintic therapy should be given to all patients, and recommendations are based on the stage of the lesion. Cysticerci appearing as cystic lesions in the brain parenchyma with or without pericystic edema or in the subarachnoid space at the convexity of the cerebral hemispheres should be treated with anticysticidal therapy. Cysticidal drugs accelerate the destruction of the parasites, resulting in a faster resolution of the infection. Albendazole and praziquantel are used in the treatment of neurocysticercosis. Approximately 85% of parenchymal cysts are destroyed by a single course of albendazole, and ~75% are destroyed by a single course of praziquantel. The dose of albendazole is 15 mg/kg per day in two doses for 8 days. The dose of praziquantel is 50 mg/kg per day for 15 days, although a number of other dosage regimens are also frequently cited. Prednisone or dexamethasone is given with anticysticidal therapy to reduce the host inflammatory response to degenerating parasites. Many, but not all, experts recommend anticysticidal therapy for lesions that are surrounded by a contrast-enhancing ring. There is universal agreement that calcified lesions do not need to be treated with anticysticidal therapy. Antiepileptic therapy can be stopped once the follow-up CT scan shows resolution of the lesion. Long-term antiepileptic therapy is recommended when seizures occur after resolution of edema and resorption or calcification of the degenerating cyst.

CNS toxoplasmosis is treated with a combination of sulfadiazine, 1.5–2.0 g orally qid, plus pyrimethamine, 100 mg orally to load, then 75–100 mg orally qd, plus folinic acid, 10–15 mg orally qd. Folinic acid is added to the regimen to prevent megaloblastic anemia. Therapy is continued until there is no evidence of active disease on neuroimaging studies, which typically takes at least 6 weeks, and then the dose of sulfadiazine is reduced to 2–4 g/d and pyrimethamine to 50 mg/d. Clindamycin plus pyrimethamine is an alternative therapy for patients who cannot tolerate sulfadiazine, but the combination of pyrimethamine and sulfadiazine is more effective.

SUBDURAL EMPYEMA

A subdural empyema (SDE) is a collection of pus between the dura and arachnoid membranes **(Fig. 381-5)**.

Figure 381-5

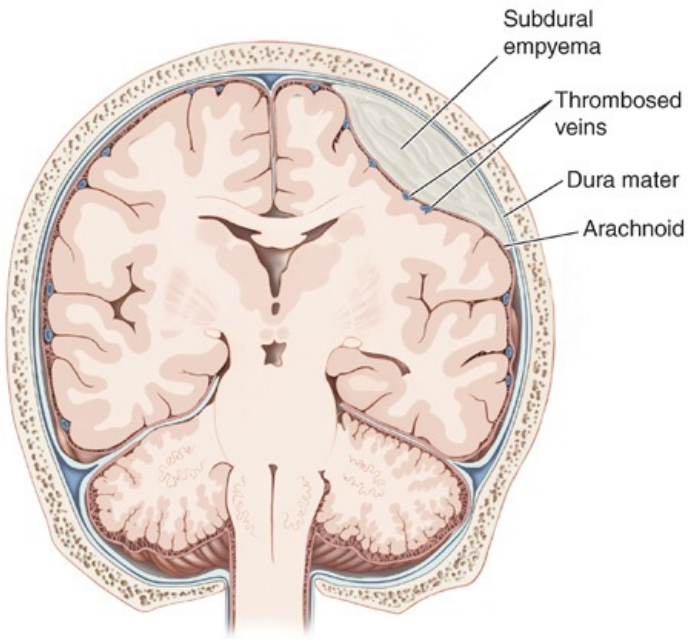

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Subdural empyema.

Epidemiology

SDE is a rare disorder that accounts for 15–25% of focal suppurative CNS infections. Sinusitis is the most common predisposing condition and typically involves the frontal sinuses, either alone or in combination with the ethmoid and maxillary sinuses. Sinusitis-associated empyema has a striking predilection for young males, possibly reflecting sex-related differences in sinus anatomy and development. It has been suggested that SDE may complicate 1–2% of cases of frontal sinusitis severe enough to require hospitalization. As a consequence of this epidemiology, SDE shows an ~3:1 male/female predominance, with 70% of cases occurring in the second and third decades of life. SDE may also develop as a complication of head trauma or neurosurgery. Secondary infection of a subdural effusion may also result in empyema, although secondary infection of hematomas, in the absence of a prior neurosurgical procedure, is rare.

Etiology

Aerobic and anaerobic streptococci, staphylococci, Enterobacteriaceae, and anaerobic bacteria are the most common causative organisms of sinusitis-associated SDE. Staphylococci and gram-negative bacilli are often the etiologic organisms when SDE follows neurosurgical procedures or head trauma. Up to one-third of cases are culture-negative, possibly reflecting difficulty in obtaining adequate anaerobic cultures.

Pathophysiology

Sinusitis-associated SDE develops as a result of either retrograde spread of infection from septic thrombophlebitis of the mucosal veins draining the sinuses or contiguous spread of infection to the brain from osteomyelitis in the posterior wall of the frontal or other sinuses. SDE may also develop from direct introduction of bacteria into the subdural space as a complication of a neurosurgical procedure. The evolution of SDE can be extremely rapid because the subdural space is a large

compartment that offers few mechanical barriers to the spread of infection. In patients with sinusitis-associated SDE, suppuration typically begins in the upper and anterior portions of one cerebral hemisphere and then extends posteriorly. SDE is often associated with other intracranial infections, including epidural empyema (40%), cortical thrombophlebitis (35%), and intracranial abscess or cerebritis (>25%). Cortical venous infarction produces necrosis of underlying cerebral cortex and subcortical white matter, with focal neurologic deficits and seizures (see below).

Clinical Presentation

A patient with SDE typically presents with fever and a progressively worsening headache. The diagnosis of SDE should always be suspected in a patient with known sinusitis who presents with new CNS signs or symptoms. Patients with underlying sinusitis frequently have symptoms related to this infection. As the infection progresses, focal neurologic deficits, seizures, nuchal rigidity, and signs of increased ICP commonly occur. Headache is the most common complaint at the time of presentation; initially it is localized to the side of the subdural infection, but then it becomes more severe and generalized. Contralateral hemiparesis or hemiplegia is the most common focal neurologic deficit and can occur from the direct effects of the SDE on the cortex or as a consequence of venous infarction. Seizures begin as partial motor seizures that then become secondarily generalized. Seizures may be due to the direct irritative effect of the SDE on the underlying cortex or result from cortical venous infarction (see above). In untreated SDE, the increasing mass effect and increase in ICP cause progressive deterioration in consciousness, leading ultimately to coma.

Diagnosis

MRI **(Fig. 381-6)** is superior to CT in identifying SDE and any associated intracranial infections. The administration of gadolinium greatly improves diagnosis by enhancing the rim of the empyema and allowing the empyema to be clearly delineated from the underlying brain parenchyma. Cranial MRI is also extremely valuable in identifying sinusitis, other focal CNS infections, cortical venous infarction, cerebral edema, and cerebritis. CT may show a crescent-shaped hypodense lesion over one or both hemispheres or in the interhemispheric fissure. Frequently the degree of mass effect, exemplified by midline shift, ventricular compression, and sulcal effacement, is far out of proportion to the mass of the SDE.

Figure 381-6

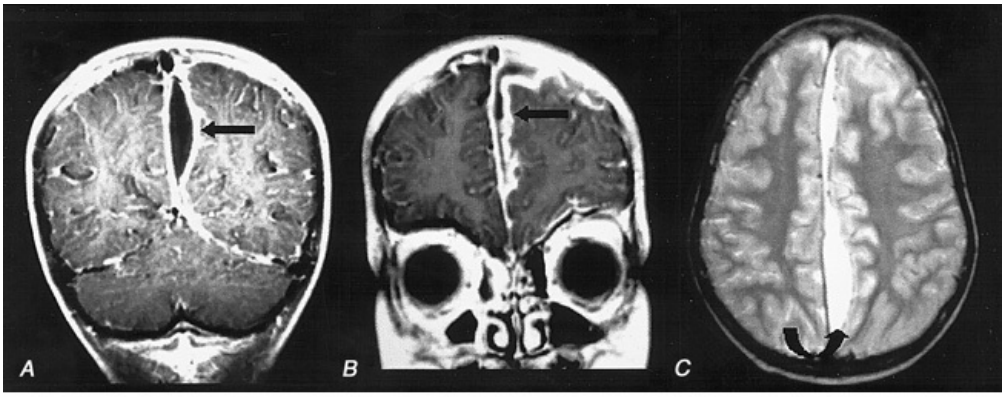

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Subdural empyema. There is marked enhancement of the dura and leptomeninges (***A***, ***B***, *straight arrows*) along the left medial hemisphere. The pus is hypointense on T1-weighted images (***A***, ***B***) but markedly hyperintense on the proton density–weighted (***C***, *curved arrow*) image. *(Courtesy of Joseph Lurito, MD; with permission.)*

CSF examination should be avoided in patients with known or suspected SDE as it adds no useful information and is associated with the risk of cerebral herniation.

Differential Diagnosis

The differential diagnosis of the combination of headache, fever, focal neurologic signs, and seizure activity that progresses rapidly to an altered level of consciousness includes subdural hematoma, bacterial meningitis, viral encephalitis, brain abscess, superior sagittal sinus thrombosis, and acute disseminated encephalomyelitis. The presence of nuchal rigidity is unusual with brain abscess or epidural empyema and should suggest the possibility of SDE when associated with significant focal neurologic signs and fever. Patients with bacterial meningitis also have nuchal rigidity but do not typically have focal deficits of the severity seen with SDE.

Treatment: Subdural Empyema

SDE is a medical emergency. Emergent neurosurgical evacuation of the empyema, either through craniotomy, craniectomy, or burr-hole drainage, is the definitive step in the management of this infection. Empirical antimicrobial therapy for community-acquired SDE should include a combination of a third-generation cephalosporin (e.g., cefotaxime or ceftriaxone), vancomycin, and metronidazole (see Table 381-1 for dosages). Patients with hospital-acquired SDE may have infections due to *Pseudomonas* spp. or MRSA and should receive coverage with a carbapenem (e.g., meropenem) and vancomycin. Metronidazole is not necessary for anti-anaerobic therapy when meropenem is being used. Parenteral antibiotic therapy should be continued for a minimum of 3–4 weeks after SDE drainage. Patients with associated cranial osteomyelitis may require longer therapy. Specific diagnosis of the etiologic organisms is made based on Gram's stain and culture of fluid obtained via either burr holes or craniotomy; the initial empirical antibiotic coverage can be modified accordingly.

Prognosis

Prognosis is influenced by the level of consciousness of the patient at the time of hospital presentation, the size of the empyema, and the speed with which therapy is instituted. Long-term neurologic sequelae, which include seizures and hemiparesis, occur in up to 50% of cases.

CRANIAL EPIDURAL ABSCESS

Cranial epidural abscess is a suppurative infection occurring in the potential space between the inner skull table and dura **(Fig. 381-7)**.

Figure 381-7

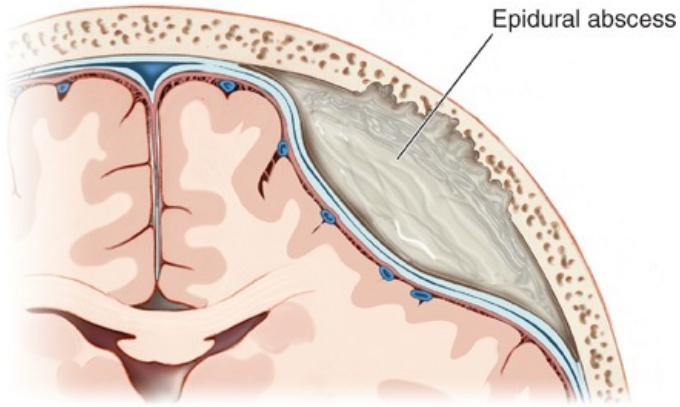

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cranial epidural abscess is a collection of pus between the dura and the inner table of the skull.

Etiology and Pathophysiology

Cranial epidural abscess is less common than either brain abscess or SDE and accounts for <2% of focal suppurative CNS infections. A cranial epidural abscess develops as a complication of a craniotomy or compound skull fracture or as a result of spread of infection from the frontal sinuses, middle ear, mastoid, or orbit. An epidural abscess may develop contiguous to an area of osteomyelitis, when craniotomy is complicated by infection of the wound or bone flap, or as a result of direct infection of the epidural space. Infection in the frontal sinus, middle ear, mastoid, or orbit can reach the epidural space through retrograde spread of infection from septic thrombophlebitis in the emissary veins that drain these areas or by way of direct spread of infection through areas of osteomyelitis. Unlike the subdural space, the epidural space is really a potential rather than an actual compartment. The dura is normally tightly adherent to the inner skull table, and infection must dissect the dura away from the skull table as it spreads. As a result, epidural abscesses are often smaller than SDEs. Cranial epidural abscesses, unlike brain abscesses, only rarely result from hematogenous spread of infection from extracranial primary sites. The bacteriology of a cranial epidural abscess is similar to that of SDE (see above). The etiologic organisms of an epidural abscess that arises from frontal sinusitis, middle-ear infections, or mastoiditis are usually streptococci or anaerobic organisms. Staphylococci or gram-negative organisms are the usual cause of an epidural abscess that develops as a complication of craniotomy or compound skull fracture.

Clinical Presentation

Patients present with fever (60%), headache (40%), nuchal rigidity (35%), seizures (10%), and focal deficits (5%). Development of symptoms may be insidious, as the empyema usually enlarges slowly in the confined anatomic space between the dura and the inner table of the skull. Periorbital edema and Potts puffy tumor, reflecting underlying associated frontal bone osteomyelitis, are present in ~40%. In

patients with a recent neurosurgical procedure, wound infection is invariably present, but other symptoms may be subtle and can include altered mental status (45%), fever (35%), and headache (20%). The diagnosis should be considered when fever and headache follow recent head trauma or occur in the setting of frontal sinusitis, mastoiditis, or otitis media.

Diagnosis

Cranial MRI with gadolinium enhancement is the procedure of choice to demonstrate a cranial epidural abscess. The sensitivity of CT is limited by the presence of signal artifacts arising from the bone of the inner skull table. The CT appearance of an epidural empyema is that of a lens or crescent-shaped hypodense extraaxial lesion. On MRI, an epidural empyema appears as a lentiform or crescent-shaped fluid collection that is hyperintense compared to CSF on T2-weighted images. On T1-weighted images, the fluid collection may be either isointense or hypointense compared to brain. Following the administration of gadolinium, there is linear enhancement of the dura on T1-weighted images. In distinction to subdural empyema, signs of mass effect or other parenchymal abnormalities are uncommon.

Treatment: Epidural Abscess

Immediate neurosurgical drainage is indicated. Empirical antimicrobial therapy, pending the results of Gram's stain and culture of the purulent material obtained at surgery, should include a combination of a third-generation cephalosporin, vancomycin, and metronidazole (Table 381-1). Ceftazidime or meropenem should be substituted for ceftriaxone or cefotaxime in neurosurgical patients. Metronidazole is not necessary for anti-anaerobic coverage in patients receiving meropenem. When the organism has been identified, antimicrobial therapy can be modified accordingly. Antibiotics should be continued for 3–6 weeks after surgical drainage. Patients with associated osteomyelitis may require additional therapy.

Prognosis

The mortality rate is <5% in modern series, and full recovery is the rule in most survivors.

SUPPURATIVE THROMBOPHLEBITIS

Definition

Suppurative intracranial thrombophlebitis is septic venous thrombosis of cortical veins and sinuses. This may occur as a complication of bacterial meningitis; SDE; epidural abscess; or infection in the skin of the face, paranasal sinuses, middle ear, or mastoid.

Anatomy and Pathophysiology

The cerebral veins and venous sinuses have no valves; therefore, blood within them can flow in either direction. The superior sagittal sinus is the largest of the venous sinuses **(Fig. 381-8)**. It receives blood from the frontal, parietal, and occipital superior cerebral veins and the diploic veins, which communicate with the meningeal veins. Bacterial meningitis is a common predisposing condition for septic thrombosis of the superior sagittal sinus. The diploic veins, which drain into the superior sagittal sinus, provide a route for the spread of infection from the meninges, especially in cases where there is purulent exudate near areas of the superior sagittal sinus. Infection can also spread to the superior sagittal sinus from nearby SDE or epidural abscess. Dehydration from vomiting, hypercoagulable states, and immunologic abnormalities, including the presence of circulating antiphospholipid antibodies, also contribute to cerebral venous sinus thrombosis. Thrombosis may extend from one sinus to another, and at autopsy thrombi of different histologic ages can often be detected in several sinuses. Thrombosis of the superior sagittal sinus is often associated with thrombosis of superior cortical veins and small parenchymal hemorrhages.

Figure 381-8

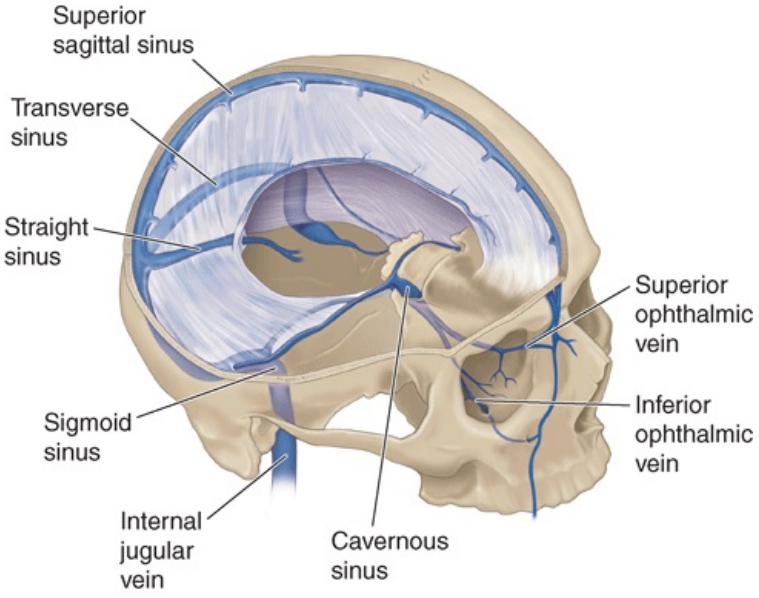

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Anatomy of the cerebral venous sinuses.

The superior sagittal sinus drains into the transverse sinuses (Fig. 381-8). The transverse sinuses also receive venous drainage from small veins from both the middle ear and mastoid cells. The transverse sinus becomes the sigmoid sinus before draining into the internal jugular vein. Septic transverse/sigmoid sinus thrombosis can be a complication of acute and chronic otitis media or mastoiditis. Infection spreads from the mastoid air cells to the transverse sinus via the emissary veins or by direct invasion. The cavernous sinuses are inferior to the superior sagittal sinus at the base of the skull. The cavernous sinuses receive blood from the facial veins via the superior and inferior ophthalmic veins. Bacteria in the facial veins enter the cavernous sinus via these veins. Bacteria in the sphenoid and ethmoid sinuses can spread to the cavernous sinuses via the small emissary veins. The sphenoid and ethmoid sinuses are the most common sites of primary infection resulting in septic cavernous sinus thrombosis.

Clinical Manifestations

Septic thrombosis of the superior sagittal sinus presents with headache, fever, nausea and vomiting, confusion, and focal or generalized seizures. There may be a rapid development of stupor and coma. Weakness of the lower extremities with bilateral Babinski,s signs or hemiparesis is often present. When superior sagittal sinus thrombosis occurs as a complication of bacterial meningitis, nuchal rigidity and Kernig's and Brudzinski's signs may be present.

The oculomotor nerve, the trochlear nerve, the abducens nerve, the ophthalmic and maxillary branches of the trigeminal nerve, and the internal carotid artery all pass through the cavernous sinus **(see Fig. 376-4)**. The symptoms of *septic cavernous sinus thrombosis* are fever, headache, frontal and retroorbital pain, and diplopia. The classic signs are ptosis, proptosis, chemosis, and extraocular dysmotility due to deficits of cranial nerves III, IV, and VI; hyperesthesia of the ophthalmic and maxillary divisions of the fifth cranial nerve and a decreased corneal reflex may be detected. There may be evidence of dilated, tortuous retinal veins and papilledema.

Headache and earache are the most frequent symptoms of *transverse sinus thrombosis*. A transverse sinus thrombosis may also present with otitis media, sixth nerve palsy, and retroorbital or facial pain (*Gradenigo's syndrome*). Sigmoid sinus and internal jugular vein thrombosis may present with neck pain.

Diagnosis

The diagnosis of septic venous sinus thrombosis is suggested by an absent flow void within the affected venous sinus on MRI and confirmed by magnetic resonance venography, CT angiogram, or the venous phase of cerebral angiography. The diagnosis of thrombophlebitis of intracerebral and meningeal veins is suggested by the presence of intracerebral hemorrhage but requires cerebral angiography for definitive diagnosis.

Treatment: Suppurative Thrombophlebitis

Septic venous sinus thrombosis is treated with antibiotics, hydration, and removal of infected tissue and thrombus in septic lateral or cavernous sinus thrombosis. The choice of antimicrobial therapy is based on the bacteria responsible for the predisposing or associated condition. Optimal duration of

therapy is unknown, but antibiotics are usually continued for 6 weeks or until there is radiographic evidence of resolution of thrombosis. Anticoagulation with dose-adjusted intravenous heparin is recommended for aseptic venous sinus thrombosis and in the treatment of septic venous sinus thrombosis complicating bacterial meningitis in patients who have progressive neurologic deterioration despite antimicrobial therapy and intravenous fluids. The presence of a small intracerebral hemorrhage from septic thrombophlebitis is not an absolute contraindication to heparin therapy. Successful management of aseptic venous sinus thrombosis has been reported with surgical thrombectomy, catheter-directed urokinase therapy, and with a combination of intrathrombus recombinant tissue plasminogen activator (rtPA) and intravenous heparin, but there is not enough data to recommend these therapies in septic venous sinus thrombosis.

FURTHER READINGS

Brouwer MC et al: Nationwide implementation of adjunctive dexamethasone therapy for pneumococcal meningitis. Neurology 75:1533, 2010[PMID: 20881273] [Full Text]

De Gans J, Van de Beek D: Dexamethasone in adults with bacterial meningitis. N Engl J Med 347:1549, 2002

Dworkin MS et al: The changing epidemiology of invasive *Haemophilus influenzae* disease, especially in persons ≥65 years old. Clin Infect Dis 44:810, 2007[PMID: 17304452] [Full Text]

Irani DN: Aseptic meningitis and viral myelitis. Neurol Clin 26:635, 2008[PMID: 18657719] [Full Text]

Perfect JR et al: Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the Infectious Diseases Society of America. Clin Infect Dis 50:291, 2010[PMID: 20047480] [Full Text]

Rosenstein NE et al: Meningococcal disease. N Engl J Med 344:1378, 2001[PMID: 11333996] [Full Text]

Scarborough M et al: Corticosteroids for bacterial meningitis in adults in sub-Saharan Africa. N Engl J Med 357:2441, 2007[PMID: 18077809] [Full Text]

Stephens DS et al: Epidemic meningitis, meningococcemia, and *Neisseria meningitidis*. Lancet 369:2196, 2007[PMID: 17604802] [Full Text]

Tunkel AR et al: Practice guidelines for the management of bacterial meningitis. Clin Infect Dis 39:1267, 2004[PMID: 15494903] [Full Text]

————: The management of encephalitis: Clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 47:303, 2008

Tyler KL: Emerging viral infections of the central nervous system. Neurol 66:939 (Pt I), 1065 (Pt II), 2009

Weber T: Progressive multifocal leukoencephalopathy. Neurol Clin 26:833, 2008[PMID: 18657729] [Full Text]

Weisfelt M et al: Dexamethasone and long-term outcome in adults with bacterial meningitis. Ann Neurol 60:456, 2006[PMID: 16958121] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHRONIC AND RECURRENT MENINGITIS: INTRODUCTION

Chronic inflammation of the meninges (pia, arachnoid, and dura) can produce profound neurologic disability and may be fatal if not successfully treated. The condition is most commonly diagnosed when a characteristic neurologic syndrome exists for >4 weeks and is associated with a persistent inflammatory response in the cerebrospinal fluid (CSF) (white blood cell count >5/ L). The causes are varied, and appropriate treatment depends on identification of the etiology. Five categories of disease account for most cases of chronic meningitis: (1) meningeal infections, (2) malignancy, (3) noninfectious inflammatory disorders, (4) chemical meningitis, and (5) parameningeal infections.

Clinical Pathophysiology

Neurologic manifestations of chronic meningitis **(Table 382-1)** are determined by the anatomic location of the inflammation and its consequences. Persistent headache with or without stiff neck, hydrocephalus, cranial neuropathies, radiculopathies, and cognitive or personality changes are the cardinal features. These can occur alone or in combination. When they appear in combination, widespread dissemination of the inflammatory process along CSF pathways has occurred. In some cases, the presence of an underlying systemic illness points to a specific agent or class of agents as the probable cause. The diagnosis of chronic meningitis is usually made when the clinical presentation prompts the astute physician to examine the CSF for signs of inflammation. CSF is produced by the choroid plexus of the cerebral ventricles, exits through narrow foramina into the subarachnoid space surrounding the brain and spinal cord, circulates around the base of the brain and over the cerebral hemispheres, and is resorbed by arachnoid villi projecting into the superior sagittal sinus. CSF flow provides a pathway for rapid spread of infectious and other infiltrative processes over the brain, spinal cord, and cranial and spinal nerve roots. Spread from the subarachnoid space into brain parenchyma may occur via the arachnoid cuffs that surround blood vessels that penetrate brain tissue (Virchow-Robin spaces).

Table 382-1 Symptoms and Signs of Chronic Meningitis

Symptom	Sign
Chronic headache	+/– Papilledema
Neck or back pain	Brudzinski's or Kernig's sign of meningeal irritation
Change in personality	Altered mental status—drowsiness, inattention, disorientation, memory loss, frontal release signs (grasp, suck, snout), perseveration
Facial weakness	Peripheral seventh CN palsy
Double vision	Palsy of CNs III, IV, VI
Visual loss	Papilledema, optic atrophy
Hearing loss	Eighth CN palsy
Arm or leg weakness	Myelopathy or radiculopathy
Numbness in arms or legs	Myelopathy or radiculopathy
Sphincter dysfunction	Myelopathy or radiculopathyFrontal lobe dysfunction (hydrocephalus)
Clumsiness	Ataxia

Abbreviation: CN, cranial nerve.

INTRACRANIAL MENINGITIS

Nociceptive fibers of the meninges are stimulated by the inflammatory process, resulting in headache or neck or back pain. Obstruction of CSF pathways at the foramina or arachnoid villi may produce *hydrocephalus* and symptoms of raised intracranial pressure (ICP), including headache, vomiting, apathy or drowsiness, gait instability, papilledema, visual loss, impaired upgaze, or palsy of the sixth

cranial nerve (CN) (Chap. 376). Cognitive and behavioral changes during the course of chronic meningitis may also result from vascular damage, which may similarly produce seizures, stroke, or myelopathy. Inflammatory deposits seeded via the CSF circulation are often prominent around the brainstem and cranial nerves and along the undersurface of the frontal and temporal lobes. Such cases, termed *basal meningitis*, often present as multiple cranial neuropathies, with visual loss (CN II), facial weakness (CN VII), hearing loss (CN VIII), diplopia (CNs III, IV, and VI), sensory or motor abnormalities of the oropharynx (CNs IX, X, and XII), decreased olfaction (CN I), or facial sensory loss and masseter weakness (CN V).

SPINAL MENINGITIS

Injury may occur to motor and sensory roots as they traverse the subarachnoid space and penetrate the meninges. These cases present as multiple radiculopathies with combinations of radicular pain, sensory loss, motor weakness, and sphincter dysfunction. Meningeal inflammation can encircle the cord, resulting in myelopathy. Patients with slowly progressive involvement of multiple cranial nerves and/or spinal nerve roots are likely to have chronic meningitis. Electrophysiologic testing (electromyography, nerve conduction studies, and evoked response testing) may be helpful in determining whether there is involvement of cranial and spinal nerve roots.

SYSTEMIC MANIFESTATIONS

In some patients, evidence of systemic disease provides clues to the underlying cause of chronic meningitis. A careful history and physical examination are essential before embarking on a diagnostic workup, which may be costly, prolonged, and associated with risk from invasive procedures. A complete history of travel, sexual practice, and exposure to infectious agents should be sought. Infectious causes are often associated with fever, malaise, anorexia, and signs of localized or disseminated infection outside the nervous system. Infectious causes are of major concern in the immunosuppressed patient, especially in patients with AIDS, in whom chronic meningitis may present without headache or fever. Noninfectious inflammatory disorders often produce systemic manifestations, but meningitis may be the initial manifestation. Carcinomatous meningitis may or may not be accompanied by clinical evidence of the primary neoplasm.

Approach to the Patient: Chronic Meningitis

The occurrence of chronic headache, hydrocephalus, cranial neuropathy, radiculopathy, and/or cognitive decline in a patient should prompt consideration of a lumbar puncture for evidence of meningeal inflammation. On occasion, the diagnosis is made when an imaging study (CT or MRI) shows contrast enhancement of the meninges, which is always abnormal with the exception of dural enhancement after lumbar puncture, neurosurgical procedures, or spontaneous CSF leakage. Once chronic meningitis is confirmed by CSF examination, effort is focused on identifying the cause **(Tables 382-2** and **382-3)** by (1) further analysis of the CSF, (2) diagnosis of an underlying systemic infection or noninfectious inflammatory condition, or (3) pathologic examination of meningeal biopsy specimens.

Table 382-2 Infectious Causes of Chronic Meningitis

Causative Agent	CSF Formula	Helpful Diagnostic Tests	Risk Factors and Systemic Manifestations
Common Bacterial Causes			
Partially treated suppurative meningitis	Mononuclear or mixed mononuclear-polymorphonuclear cells	CSF culture and Gram's stain	History consistent with acute bacterial meningitis and incomplete treatment
Parameningeal infection	Mononuclear or mixed polymorphonuclear-mononuclear cells	Contrast-enhanced CT or MRI to detect parenchymal, subdural, epidural, or sinus infection	Otitis media, pleuropulmonary infection, right-to-left cardiopulmonary shunt for brain abscess; focal neurologic signs; neck, back, ear, or sinus tenderness
Mycobacterium tuberculosis	Mononuclear cells except polymorphonuclear cells in early infection (commonly <500 WBC/μL); low CSF glucose, high protein	Tuberculin skin test may be negative; AFB culture of CSF (sputum, urine, gastric contents if indicated); tuberculostearic acid detection in CSF; identify tubercle bacillus on acid-fast stain of CSF or protein pellicle; PCR	Exposure history; previous tuberculous illness; immunosuppressed or AIDS; young children; fever, meningismus, night sweats, miliary TB on x-ray or liver biopsy; stroke due to arteritis
Lyme disease (Bannwarth's syndrome) *Borrelia*	Mononuclear cells; elevated protein	Serum Lyme antibody titer; Western blot confirmation; (patients with syphilis may have false-positive Lyme titer)	History of tick bite or appropriate exposure history; erythema chronicum migrans skin rash; arthritis,

burgdorferi			radiculopathy, Bell's palsy, meningoencephalitis–multiple sclerosis-like syndrome
Syphilis (secondary, tertiary) *Treponema pallidum*	Mononuclear cells; elevated protein	CSF VDRL; serum VDRL (or RPR); fluorescent treponemal antibody-absorbed (FTA) or MHA-TP; serum VDRL may be negative in tertiary syphilis	Appropriate exposure history; HIV- seropositive individuals at increased risk of aggressive infection; "dementia"; cerebral infarction due to endarteritis
Uncommon Bacterial Causes			
Actinomyces	Polymorphonuclear cells	Anaerobic culture	Parameningeal abscess or sinus tract (oral or dental focus); pneumonitis
Nocardia	Polymorphonuclear; occasionally mononuclear cells; often low glucose	Isolation may require weeks; weakly acid fast	Associated brain abscess may be present
Brucella	Mononuclear cells (rarely polymorphonuclear); elevated protein; often low glucose	CSF antibody detection; serum antibody detection	Intake of unpasteurized dairy products; exposure to goats, sheep, cows; fever, arthralgia, myalgia, vertebral osteomyelitis
Whipple's disease *Tropheryma whippelii*	Mononuclear cells	Biopsy of small bowel or lymph node; CSF PCR for T. whippelii; brain and meningeal biopsy (with PAS stain and EM examination)	Diarrhea, weight loss, arthralgias, fever; dementia, ataxia, paresis, ophthalmoplegia, oculomasticatory myoclonus
Rare Bacterial Causes			
Leptospirosis (occasionally if left untreated may last 3–4 weeks)			
Fungal Causes			
Cryptococcus neoformans	Mononuclear cells; count not elevated in some patients with AIDS	India ink or fungal wet mount of CSF (budding yeast); blood and urine cultures; antigen detection in CSF	AIDS and immune suppression; pigeon exposure; skin and other organ involvement due to disseminated infection
Coccidioides immitis	Mononuclear cells (sometimes 10–20% eosinophils); often low glucose	Antibody detection in CSF and serum	Exposure history —southwestern US; increased virulence in dark-skinned races
Candida sp.	Polymorphonuclear or mononuclear	Fungal stain and culture of CSF	IV drug abuse; post surgery; prolonged intravenous therapy; disseminated candidiasis
Histoplasma capsulatum	Mononuclear cells; low glucose	Fungal stain and culture of large volumes of CSF; antigen detection in CSF, serum, and urine; antibody detection in serum, CSF	Exposure history—Ohio and central Mississippi River Valley; AIDS; mucosal lesions
Blastomyces dermatitidis	Mononuclear cells	Fungal stain and culture of CSF; biopsy and culture of skin, lung lesions; antibody detection in serum	Midwestern and southeastern USA; usually systemic infection; abscesses, draining sinus, ulcers
Aspergillus sp.	Mononuclear or polymorphonuclear	CSF culture	Sinusitis; granulocytopenia or immunosuppression
Sporothrix schenckii	Mononuclear cells	Antibody detection in CSF and serum; CSF culture	Traumatic inoculation; IV drug use; ulcerated skin lesion
Rare Fungal Causes			
Xylohypha (formerly *Cladosporium*) *trichoides* and other dark-walled (dematiaceous) fungi such as *Curvularia, Drechslera; Mucor*, and, after water aspiration, *Pseudoallescheria boydii*			
Protozoal Causes			
Toxoplasma gondii	Mononuclear cells	Biopsy or response to empirical therapy in clinically appropriate context (including presence of antibody in serum)	Usually with intracerebral abscesses; common in HIV-seropositive patients
Trypanosomiasis Trypanosoma gambiense, T. rhodesiense	Mononuclear cells, elevated protein	Elevated CSF IgM; identification of trypanosomes in CSF and blood smear	Endemic in Africa; chancre, lymphadenopathy; prominent sleep disorder

Rare Protozoal Causes			
Acanthamoeba sp. causing granulomatous amebic encephalitis and meningoencephalitis in immunocompromised and debilitated individuals. *Balamuthia mandrillaris* causing chronic meningoencephalitis in immunocompetent hosts.			
Helminthic Causes			
Cysticercosis (infection with cysts of *Taenia solium*)	Mononuclear cells; may have eosinophils; glucose level may be low	Indirect hemagglutination assay in CSF; ELISA immunoblotting in serum	Usually with multiple cysts in basal meninges and hydrocephalus; cerebral cysts, muscle calcification
Gnathostoma spinigerum	Eosinophils, mononuclear cells	Peripheral eosinophilia	History of eating raw fish; common in Thailand and Japan; subarachnoid hemorrhage; painful radiculopathy
Angiostrongylus cantonensis	Eosinophils, mononuclear cells	Recovery of worms from CSF	History of eating raw shellfish; common in tropical Pacific regions; often benign
Baylisascaris procyonis (raccoon ascarid)	Eosinophils, mononuclear cells		Infection follows accidental ingestion of *B. procyonis* eggs from raccoon feces; fatal meningoencephalitis
Rare Helminthic Causes			
Trichinella spiralis (trichinosis); *Fasciola hepatica* (liver fluke), *Echinococcus cysts; Schistosoma sp*. The former may produce a lymphocytic pleocytosis whereas the latter two may produce an eosinophilic response in CSF associated with cerebral cysts (*Echinococcus*) or granulomatous lesions of brain or spinal cord			
Viral causes			
Mumps	Mononuclear cells	Antibody in serum	No prior mumps or immunization; may produce meningoencephalitis; may persist for 3–4 weeks
Lymphocytic choriomeningitis	Mononuclear cells	Antibody in serum	Contact with rodents or their excreta; may persist for 3–4 weeks
Echovirus	Mononuclear cells; may have low glucose	Virus isolation from CSF	Congenital hypogammaglobulinemia; history of recurrent meningitis
HIV (acute retroviral syndrome)	Mononuclear cells	p24 antigen in serum and CSF; high level of HIV viremia	HIV risk factors; rash, fever, lymphadenopathy; lymphopenia in peripheral blood; syndrome may persist long enough to be considered as "chronic meningitis"; or chronic meningitis may develop in later stages (AIDS) due to HIV
Herpes simplex (HSV)	Mononuclear cells	PCR for HSV, CMV DNA; CSF antibody for HSV, EBV	Recurrent meningitis due to HSV-2 (rarely HSV-1) often associated with genital recurrences; EBV associated with myeloradiculopathy, CMV with polyradiculopathy

Abbreviations: AFB, acid-fast bacillus; CMV, cytomegalovirus; CSF, cerebrospinal fluid; CT, computed tomography; EBV, Epstein-Barr virus; ELISA, enzyme-linked immunosorbent assay; EM, electron microscopy; FTA, fluorescent treponemal antibody absorption test; HSV, herpes simplex virus; MHA-TP, microhemagglutination assay–T. pallidum; MRI, magnetic resonance imaging; PAS, periodic acid –Schiff; PCR, polymerase chain reaction; RPR, rapid plasma reagin test; TB, tuberculosis; VDRL, Venereal Disease Research Laboratories test.

Two clinical forms of chronic meningitis exist. In the first, the symptoms are chronic and persistent, whereas in the second there are recurrent, discrete episodes of illness. In the latter group, all symptoms, signs, and CSF parameters of meningeal inflammation resolve completely between episodes without specific therapy. In such patients, the likely etiologies include herpes simplex virus (HSV) type 2; chemical meningitis due to leakage into CSF of contents from an epidermoid tumor, craniopharyngioma, or cholesteatoma; primary inflammatory conditions, including Vogt-Koyanagi-Harada syndrome, Behçet's syndrome, systemic lupus erythematosus; and drug hypersensitivity with repeated administration of the offending agent.

The epidemiologic history is of considerable importance and may provide direction for selection of laboratory studies. Pertinent features include a history of tuberculosis or exposure to a likely case; past travel to areas endemic for fungal infections (the San Joaquin Valley in California and southwestern states for coccidioidomycosis, midwestern states for histoplasmosis, southeastern states for blastomycosis); travel to the Mediterranean region or ingestion of imported unpasteurized dairy products (*Brucella*); time spent in wooded areas endemic for Lyme disease; exposure to sexually transmitted disease (syphilis); exposure of an immunocompromised host to pigeons and their droppings (*Cryptococcus*); gardening (*Sporothrix schenkii*); ingestion of poorly cooked meat or contact with a household cat (*Toxoplasma gondii*); residence in Thailand or Japan (*Gnathostoma spinigerum*), Latin America (*Paracoccidioides brasiliiensis*), or the South Pacific (*Angiostrongylus cantonensis*); rural residence and raccoon exposure (*Baylisascaris procyonis*); and residence in Latin America, the Philippines, or Southeast Asia when eosinophilic meningitis is present (*Taenia solium*).

The presence of focal cerebral signs in a patient with chronic meningitis suggests the possibility of a brain abscess or other parameningeal infection; identification of a potential source of infection (chronic draining ear, sinusitis, right-to-left cardiac or pulmonary shunt, chronic pleuropulmonary infection) supports this diagnosis. In some cases, diagnosis may be established by recognition and biopsy of unusual skin lesions (Behçet's syndrome, cryptococcosis, blastomycosis, SLE, Lyme disease, IV drug use, sporotrichosis, trypanosomiasis) or enlarged lymph nodes (lymphoma, tuberculosis, sarcoid, infection with HIV, secondary syphilis, or Whipple's disease). A careful ophthalmologic examination may reveal uveitis [Vogt-Koyanagi-Harada syndrome, sarcoid, or central nervous system (CNS) lymphoma], keratoconjunctivitis sicca (Sjögren's syndrome), or iridocyclitis (Behçet's syndrome) and is essential to assess visual loss from papilledema. Aphthous oral lesions, genital ulcers, and hypopyon suggest Behçet's syndrome. Hepatosplenomegaly suggests lymphoma, sarcoid, tuberculosis, or brucellosis. Herpetic lesions in the genital area or on the thighs suggest HSV-2 infection. A breast nodule, a suspicious pigmented skin lesion, focal bone pain, or an abdominal mass directs attention to possible carcinomatous meningitis.

IMAGING

Once the clinical syndrome is recognized as a potential manifestation of chronic meningitis, proper analysis of the CSF is essential. However, if the possibility of raised ICP exists, a brain imaging study should be performed before lumbar puncture. If ICP is elevated because of a mass lesion, brain swelling, or a block in ventricular CSF outflow (obstructive hydrocephalus), then lumbar puncture carries the potential risk of brain herniation. Obstructive hydrocephalus usually requires direct ventricular drainage of CSF. In patients with open CSF flow pathways, elevated ICP can still occur due to impaired resorption of CSF by arachnoid villi. In such patients, lumbar puncture is usually safe, but repetitive or continuous lumbar drainage may be necessary to prevent abrupt deterioration and death from raised ICP. In some patients, especially with cryptococcal meningitis, fatal levels of raised ICP can occur without enlarged ventricles.

Contrast-enhanced MRI or CT studies of the brain and spinal cord can identify meningeal enhancement, parameningeal infections (including brain abscess), encasement of the spinal cord (malignancy or inflammation and infection), or nodular deposits on the meninges or nerve roots (malignancy or sarcoidosis) **(Fig. 382-1)**. Imaging studies are also useful to localize areas of meningeal disease prior to meningeal biopsy.

Figure 382-1

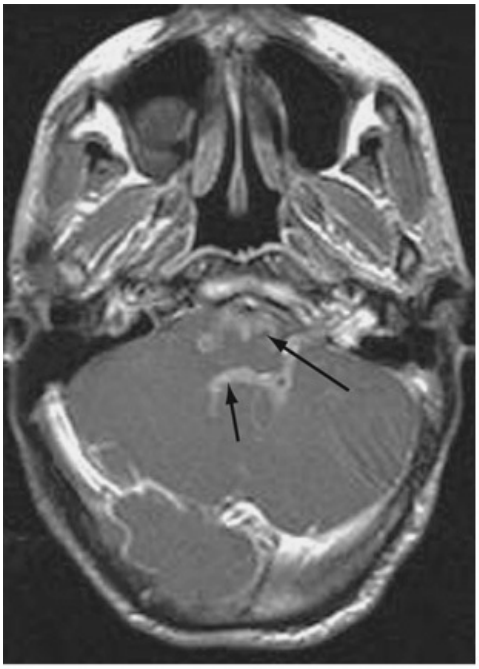

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Primary central nervous system lymphoma. A 24-year-old man, immunosuppressed due to intestinal lymphangiectasia, developed multiple cranial neuropathies. CSF findings consisted of 100 lymphocytes/μL and a protein of 2.5 g/L (250 mg/dL); cytology and cultures were negative. Gadolinium-enhanced T1 MRI revealed diffuse, multifocal meningeal enhancement surrounding the brainstem (***A***), spinal cord, and cauda equina (***B***).

Cerebral angiography may be indicated in patients with chronic meningitis and stroke to identify cerebral arteritis (granulomatous angiitis, other inflammatory arteritides, or infectious arteritis).

CEREBROSPINAL FLUID ANALYSIS

The CSF pressure should be measured and samples sent for bacterial, fungal, and tuberculous culture; Venereal Disease Research Laboratories (VDRL) test; cell count and differential; Gram's stain; and measurement of glucose and protein. Wet mount for fungus and parasites, India ink preparation and culture, culture for fastidious bacteria and fungi, assays for cryptococcal antigen and oligoclonal

immunoglobulin bands, and cytology should be performed. Other specific CSF tests (Tables 382-2 and 382-3) or blood tests and cultures should be ordered as indicated on the basis of the history, physical examination, or preliminary CSF results (i.e., eosinophilic, mononuclear, or polymorphonuclear meningitis). Rapid diagnosis may be facilitated by serologic tests and polymerase chain reaction (PCR) testing to identify DNA sequences in the CSF that are specific for the suspected pathogen.

Table 382-3 Noninfectious Causes of Chronic Meningitis

Causative Agents	CSF Formula	Helpful Diagnostic Tests	Risk Factors and Systemic Manifestations
Malignancy	Mononuclear cells, elevated protein, low glucose	Repeated cytologic examination of large volumes of CSF; CSF exam by polarizing microscopy; clonal lymphocyte markers; deposits on nerve roots or meninges seen on myelogram or contrast-enhanced MRI; meningeal biopsy	Metastatic cancer of breast, lung, stomach, or pancreas; melanoma, lymphoma, leukemia; meningeal gliomatosis; meningeal sarcoma; cerebral dysgerminoma; meningeal melanoma or B cell lymphoma
Chemical compounds (may cause recurrent meningitis)	Mononuclear or PMNs, low glucose, elevated protein; xanthochromia from subarachnoid hemorrhage in week prior to presentation with "meningitis"	Contrast-enhanced CT scan or MRICerebral angiogram to detect aneurysm	History of recent injection into the subarachnoid space; history of sudden onset of headache; recent resection of acoustic neuroma or craniopharyngioma; epidermoid tumor of brain or spine, sometimes with dermoid sinus tract; pituitary apoplexy
Primary Inflammation			
CNS sarcoidosis	Mononuclear cells; elevated protein; often low glucose	Serum and CSF angiotensin-converting enzyme levels; biopsy of extraneural affected tissues or brain lesion/meningeal biopsy	CN palsy, especially of CN VII; hypothalamic dysfunction, especially diabetes insipidus; abnormal chest radiograph; peripheral neuropathy or myopathy
Vogt-Koyanagi-Harada syndrome (recurrent meningitis)	Mononuclear cells		Recurrent meningoencephalitis with uveitis, retinal detachment, alopecia, lightening of eyebrows and lashes, dysacousia, cataracts, glaucoma
Isolated granulomatous angiitis of the nervous system	Mononuclear cells, elevated protein	Angiography or meningeal biopsy	Subacutedementia; multiple cerebral infarctions; recent zoster ophthalmicus
Systemic lupus erythematosus	Mononuclear or PMNs	Anti-DNA antibody, antinuclear antibodies	Encephalopathy; seizures; stroke; transverse myelopathy; rash; arthritis
Behçet's syndrome (recurrent meningitis)	Mononuclear or PMNs, elevated protein		Oral and genital aphthous ulcers; iridocyclitis; retinal hemorrhages; pathergic lesions at site of skin puncture
Chronic benign lymphocytic meningitis	Mononuclear cells		Recovery in 2–6 months, diagnosis by exclusion
Mollaret's meningitis (recurrent meningitis)	Large endothelial cells and PMNs in first hours, followed by mononuclear cells	PCR for herpes; MRI/CT to rule out epidermoid tumor or dural cyst	Recurrent meningitis; exclude HSV-2; rare cases due to HSV-1; occasional case associated with dural cyst
Drug hypersensitivity	PMNs; occasionally mononuclear cells or eosinophils	Complete blood count (eosinophilia)	Exposure to non steroidal anti-inflammatory agents, sulfonamides, isoniazid, tolmetin, ciprofloxacin, penicillin, carbamazaepine, lamotrigine, IV immunoglobulin, OKT3 antibodies, phenazopyridine; improvement after discontinuation of drug; recurrence with repeat exposure

Granulomatosis with polyangiitis (Wegener's)	Mononuclear cells	Chest and sinus radiographs; urinalysis; ANCA antibodies in serum	Associated sinus, pulmonary, or renal lesions; CN palsies; skin lesions; peripheral neuropathy
Other: multiple sclerosis, Sjögren's syndrome, neonatal-onset multisystemic inflammatory disease (NOMID), and rarer forms of vasculitis (e.g., Cogan's syndrome)			

Abbreviations: ANCA, anti-neutrophil cytoplasmic antibodies; CN, cranial nerve; CSF, cerebrospinal fluid; CT, computed tomography; HSV, herpes simplex virus; MRI, magnetic resonance imaging; PCR, polymerase chain reaction; PMNs, polymorphonuclear cells.

In most categories of chronic (not recurrent) meningitis, mononuclear cells predominate in the CSF. When neutrophils predominate after 3 weeks of illness, the principal etiologic considerations are *Nocardia asteroides*, *Actinomyces israelii*, *Brucella*, *Mycobacterium tuberculosis* (5–10% of early cases only), various fungi (*Blastomyces dermatitidis*, *Candida albicans*, *Histoplasma capsulatum*, *Aspergillus* spp., *Pseudallescheria boydii*, *Cladophialophora bantiana*), and noninfectious causes (SLE, exogenous chemical meningitis). When eosinophils predominate or are present in limited numbers in a primarily mononuclear cell response in the CSF, the differential diagnosis includes parasitic diseases (*A. cantonensis*, *G. spinigerum*, *B. procyonis*, or *Toxocara canis* infection, cysticercosis, schistosomiasis, echinococcal disease, *T. gondii* infection), fungal infections (6–20% eosinophils along with a predominantly lymphocyte pleocytosis, particularly with coccidioidal meningitis), neoplastic disease (lymphoma, leukemia, metastatic carcinoma), or other inflammatory processes (sarcoidosis, hypereosinophilic syndrome).

It is often necessary to broaden the number of diagnostic tests if the initial workup does not reveal the cause. In addition, repeated samples of large volumes of CSF may be required to diagnose certain infectious and malignant causes of chronic meningitis. For instance, lymphomatous or carcinomatous meningitis may be diagnosed by examination of sections cut from a cell block formed by spinning down the sediment from a large volume of CSF. The diagnosis of fungal meningitis may require large volumes of CSF for culture of sediment. If standard lumbar puncture is unrewarding, a cervical cisternal tap to sample CSF near to the basal meninges may be fruitful.

LABORATORY INVESTIGATION

In addition to the CSF examination, an attempt should be made to uncover pertinent underlying illnesses. Tuberculin skin test, chest radiograph, urine analysis and culture, blood count and differential, renal and liver function tests, alkaline phosphatase, sedimentation rate, antinuclear antibody, anti-Ro, anti-La antibody, and serum angiotensin-converting enzyme level are often indicated. Liver or bone marrow biopsy may be diagnostic in some cases of miliary tuberculosis, disseminated fungal infection, sarcoidosis, or metastatic malignancy. Abnormalities discovered on chest radiograph or chest CT can be pursued by bronchoscopy or transthoracic needle biopsy.

MENINGEAL BIOPSY

A meningeal biopsy should be strongly considered in patients who are severely disabled, who need chronic ventricular decompression, or whose illness is progressing rapidly. The activities of the surgeon, pathologist, microbiologist, and cytologist should be coordinated so that a large enough sample is obtained and the appropriate cultures and histologic and molecular studies, including electron-microscopic and PCR studies, are performed. The diagnostic yield of meningeal biopsy can be increased by targeting regions that enhance with contrast on MRI or CT. With current microsurgical techniques, most areas of the basal meninges can be accessed for biopsy via a limited craniotomy. In a series from the Mayo Clinic reported by Cheng et al., MRI demonstrated meningeal enhancement in 47% of patients undergoing meningeal biopsy. Biopsy of an enhancing region was diagnostic in 80% of cases; biopsy of nonenhancing regions was diagnostic in only 9%; sarcoid (31%) and metastatic adenocarcinoma (25%) were the most common conditions identified. Tuberculosis is the most common condition identified in many reports from outside the United States.

APPROACH TO THE ENIGMATIC CASE

In approximately one-third of cases, the diagnosis is not known despite careful evaluation of CSF and potential extraneural sites of disease. A number of the organisms that cause chronic meningitis may take weeks to be identified by cultures. In enigmatic cases, several options are available, determined by the extent of the clinical deficits and rate of progression. It is prudent to wait until cultures are finalized if the patient is asymptomatic or symptoms are mild and not progressive. Unfortunately, in many cases progressive neurologic deterioration occurs, and rapid treatment is required. Ventricular-peritoneal shunts may be placed to relieve hydrocephalus, but the risk of disseminating the undiagnosed inflammatory process into the abdomen must be considered.

Empirical Treatment

Diagnosis of the causative agent is essential because effective therapies exist for many etiologies of chronic meningitis, but if the condition is left untreated, progressive damage to the CNS and cranial nerves and roots is likely to occur. Occasionally, empirical therapy must be initiated when all attempts

at diagnosis fail. In general, empirical therapy in the United States consists of antimycobacterial agents, amphotericin for fungal infection, or glucocorticoids for noninfectious inflammatory causes. It is important to direct empirical therapy of lymphocytic meningitis at tuberculosis, particularly if the condition is associated with hypoglycorrhachia and sixth and other CN palsies, since untreated disease is fatal in 4–8 weeks. In the Mayo Clinic series, the most useful empirical therapy was administration of glucocorticoids rather than antituberculous therapy. Carcinomatous or lymphomatous meningitis may be difficult to diagnose initially, but the diagnosis becomes evident with time.

The Immunosuppressed Patient

Chronic meningitis is not uncommon in the course of HIV infection. Pleocytosis and mild meningeal signs often occur at the onset of HIV infection, and occasionally low-grade meningitis persists. Toxoplasmosis commonly presents as intracranial abscesses and may also be associated with meningitis. Other important causes of chronic meningitis in AIDS include infection with *Cryptococcus*, *Nocardia*, *Candida*, or other fungi; syphilis; and lymphoma (Fig. 382-1). Toxoplasmosis, cryptococcosis, nocardiosis, and other fungal infections are important etiologic considerations in individuals with immunodeficiency states other than AIDS, including those due to immunosuppressive medications. Because of the increased risk of chronic meningitis and the attenuation of clinical signs of meningeal irritation in immunosuppressed individuals, CSF examination should be performed for any persistent headache or unexplained change in mental state.

FURTHER READINGS

Cho TA, Venna N: Management of acute, recurrent, and chronic meningitides in adults. Neurol Clin 28:1061, 2010[PMID: 20816277] [Full Text]

Drake WK, Adam RD: Coccidioidal meningitis and brain abscesses: Analysis of 71 cases at a referral center. Neurology 73:1780, 2009[PMID: 19933980] [Full Text]

Gilden DH et al: Herpesvirus infections of the nervous system. Nat Clin Pract Neurol 3:82, 2007[PMID: 17279082] [Full Text]

Halperin JJ et al: Practice parameter: Treatment of nervous system Lyme disease (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 69:91, 2007[PMID: 17522387] [Full Text]

Lan SH et al: Cerebral infarction in chronic meningitis: A comparison of tuberculous meningitis and cryptococcal meningitis. Q J Med 94:247, 2001[PMID: 11353098] [Full Text]

Liliang PC et al: Use of ventriculoperitoneal shunts to treat uncontrollable intracranial hypertension in patients who have cryptococcal meningitis without hydrocephalus. Clin Infect Dis 34:E64, 2002

Shapiro WR et al: Treatment modalities for leptomeningeal metastases. Semin Oncol 36:S46, 2009

Talati NJ et al: Spectrum of CNS disease caused by rapidly growing mycobacteria. Lancet Infect Dis 8:390, 2008[PMID: 18501854] [Full Text]

Vinnard C, Macgregor RR: Tuberculous meningitis in HIV-infected individuals. Curr HIV/AIDS Rep 6:139, 2009[PMID: 19589299] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PRION DISEASES: INTRODUCTION

Prions are infectious proteins that cause degeneration of the central nervous system (CNS). Prion diseases are disorders of protein conformation, the most common of which in humans is called Creutzfeldt-Jakob disease (CJD). CJD typically presents with dementia and myoclonus, is relentlessly progressive, and generally causes death within a year of onset. Most CJD patients are between 50 and 75 years of age; however, patients as young as 17 and as old as 83 have been recorded.

In mammals, prions reproduce by binding to the normal, *c*ellular isoform of the *pr*ion *p*rotein (PrP^C) and stimulating conversion of PrP^C into the di*s*ease-*c*ausing isoform (PrP^{Sc}). PrP^C is rich in α-helix and has little β-structure, while PrP^{Sc} has less α-helix and a high amount of β-structure (**Fig. 383-1**). This α-to-β structural transition in the prion protein (PrP) is the fundamental event underlying prion diseases (Table 383-1).

Figure 383-1

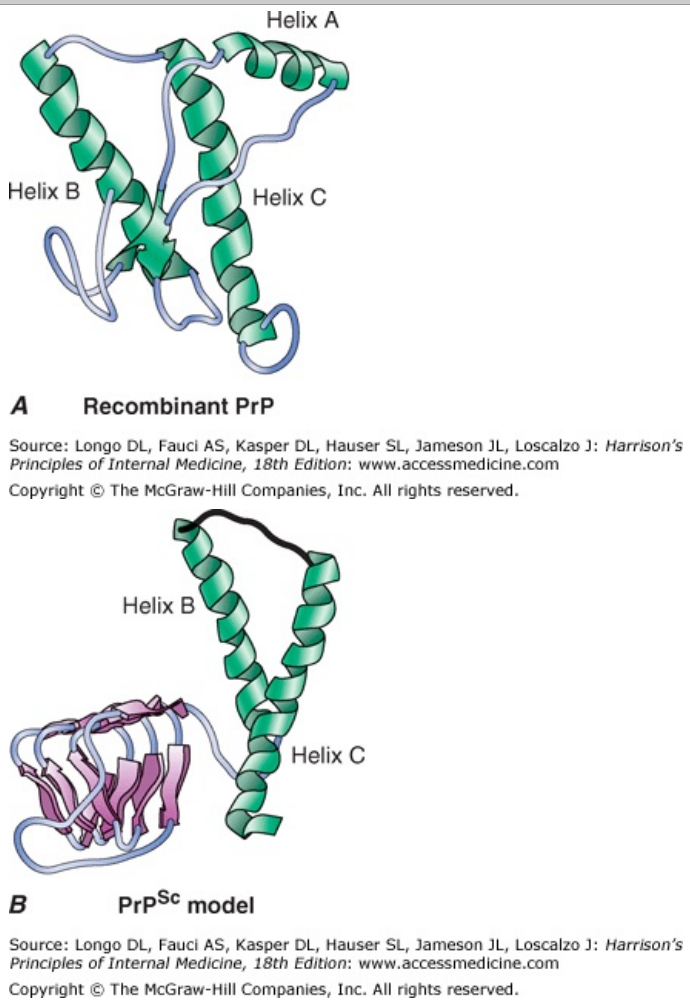

Structures of prion proteins. *A.* NMR structure of Syrian hamster recombinant (rec) PrP(90–231). Presumably, the structure of the α-helical form of recPrP(90–231) resembles that of PrP^C. recPrP(90–231) is viewed from the interface where PrP^{Sc} is thought to bind to PrP^C. Shown are: α-helices A (residues 144–157), B (172–193), and C (200–227). Flat ribbons depict β-strands S1 (129–131) and S2 (161–163). ***(A,** from SB Prusiner: N Engl J Med 344:1516, 2001; with permission.)* ***B.*** Structural model of PrP^{Sc}. The 90–160 region has been modeled onto a β-helical architecture while the COOH terminal helices B and C are preserved as in PrP^C. *(Image prepared by C. Govaerts.)*

Table 383-1 Glossary of Prion Terminology	
Prion	Proteinaceous infectious particle that lacks nucleicacid. Prions are composed entirely of PrPSc molecules. They can cause scrapie in sheep and goats, and related neurodegenerative diseases of humans such as Creutzfeldt-Jakob disease (CJD).
PrPSc	Disease-causing isoform of the prion protein. Thisprotein is the only identifiable macromolecule in purified preparations of scrapie prions.
PrPC	Cellular isoform of the prion protein. PrPC is the precursor of PrPSc.
PrP 27-30	A fragment of PrPSc, generated by truncation of the NH2-terminus by limited digestion with proteinase K. PrP 27-30 retains prion infectivity and polymerizes into amyloid.
PRNP	PrP gene located on human chromosome 20.
Prion rod	An aggregate of prions composed largely of PrP 27-30 molecules. Created by detergent extraction and limited proteolysis of PrPSc. Morphologically and histochemically indistinguishable from many amyloids.
PrP amyloid	Amyloid containing PrP in the brains of animals or humans with prion disease; often accumulates as plaques.

Four new concepts have emerged from studies of prions: (1) Prions are the only known infectious pathogens that are devoid of nucleic acid; all other infectious agents possess genomes composed of either RNA or DNA that direct the synthesis of their progeny. (2) Prion diseases may be manifest as infectious, genetic, and sporadic disorders; no other group of illnesses with a single etiology presents with such a wide spectrum of clinical manifestations. (3) Prion diseases result from the accumulation of PrP^{Sc}, the conformation of which differs substantially from that of its precursor, PrP^{C}. (4) PrP^{Sc} can exist in a variety of different conformations, each of which seems to specify a particular disease phenotype. How a specific conformation of a PrP^{Sc} molecule is imparted to PrP^{C} during prion replication to produce nascent PrP^{Sc} with the same conformation is unknown. Additionally, it is unclear what factors determine where in the CNS a particular PrP^{Sc} molecule will be deposited.

SPECTRUM OF PRION DISEASES

The sporadic form of CJD is the most common prion disorder in humans. Sporadic CJD (sCJD) accounts for ~85% of all cases of human prion disease, while inherited prion diseases account for 10–15% of all cases (**Table 383-2**). Familial CJD (fCJD), Gerstmann-Sträussler-Scheinker (GSS) disease, and fatal familial insomnia (FFI) are all dominantly inherited prion diseases that are caused by mutations in the PrP gene.

Table 383-2 The Prion Diseases		
Disease	**Host**	**Mechanism of Pathogenesis**
Human		
Kuru	Fore people	Infection through ritualistic cannibalism
iCJD	Humans	Infection from prion-contaminated hGH, duramater grafts, etc.
vCJD	Humans	Infection from bovine prions
fCJD	Humans	Germ-line mutations in PRNP
GSS	Humans	Germ-line mutations in PRNP
FFI	Humans	Germ-line mutation in PRNP (D178N, M129)
sCJD	Humans	Somatic mutation or spontaneous conversionof PrPC into PrPSc?
sFI	Humans	Somatic mutation or spontaneous conversionof PrPC into PrPSc?
Animal		
Scrapie	Sheep, goats	Infection in genetically susceptible sheep
BSE	Cattle	Infection with prion-contaminated MBM
TME	Mink	Infection with prions from sheep or cattle
CWD	Mule deer, elk	Unknown
FSE	Cats	Infection with prion-contaminated beef
Exotic ungulateencephalopathy	Greater kudu, nyala,or oryx	Infection with prion-contaminated MBM

Abbreviations: BSE, bovine spongiform encephalopathy; CJD, Creutzfeldt-Jakob disease; CWD, chronic wasting disease; fCJD, familial Creutzfeldt-Jakob disease; FFI, fatal familial insomnia; FSE, feline spongiform encephalopathy; GSS, Gerstmann-Sträussler-Scheinker disease; hGH, human growth hormone; iCJD, iatrogenic Creutzfeldt-Jakob disease; MBM, meat and bone meal; sCJD, sporadic Creutzfeldt-Jakob disease; sFI, sporadic fatal insomnia; TME, transmissible mink encephalopathy; vCJD, variant Creutzfeldt-Jakob disease.

Although infectious prion diseases account for <1% of all cases and infection does not seem to play an important role in the natural history of these illnesses, the transmissibility of prions is an important biologic feature. *Kuru* of the Fore people of New Guinea is thought to have resulted from the consumption of brains from dead relatives during ritualistic cannibalism. With the cessation of ritualistic cannibalism in the late 1950s, kuru has nearly disappeared, with the exception of a few recent patients exhibiting incubation periods of >40 years. Iatrogenic CJD (iCJD) seems to be the result of the accidental inoculation of patients with prions. Variant CJD (vCJD) in teenagers and young adults in Europe is the result of exposure to tainted beef from cattle with bovine spongiform encephalopathy (BSE).

Six diseases of animals are caused by prions (Table 383-2). Scrapie of sheep and goats is the prototypic prion disease. Mink encephalopathy, BSE, feline spongiform encephalopathy, and exotic ungulate encephalopathy are all thought to occur after the consumption of prion-infected foodstuffs. The BSE epidemic emerged in Britain in the late 1980s and was shown to be due to industrial cannibalism. Whether BSE began as a sporadic case of BSE in a cow or started with scrapie in sheep is unknown. The origin of chronic wasting disease (CWD), a prion disease endemic in deer and elk in regions of North America, is uncertain. In contrast to other prion diseases, CWD is highly communicable. Feces from asymptomatic, infected cervids contain prions that are likely to be responsible for the spread of CWD.

Epidemiology

CJD is found throughout the world. The incidence of sCJD is approximately one case per million population, and thus it accounts for about 1 in every 10,000 deaths. Because sCJD is an age-dependent neurodegenerative disease, its incidence is expected to increase steadily as older segments of populations in developed and developing countries continue to expand. Although many geographic clusters of CJD have been reported, each has been shown to segregate with a PrP gene mutation. Attempts to identify common exposure to some etiologic agent have been unsuccessful for both the sporadic and familial cases. Ingestion of scrapie-infected sheep or goat meat as a cause of CJD in humans has not been demonstrated by epidemiologic studies, although speculation about this potential route of inoculation continues. Of particular interest are deer hunters who develop CJD, because up to 90% of culled deer in some game herds have been shown to harbor CWD prions. Whether prion disease in deer or elk has passed to cows, sheep, or directly to humans remains unknown. Studies with rodents demonstrate that oral infection with prions can occur, but the process is inefficient compared to intracerebral inoculation.

Pathogenesis

The human prion diseases were initially classified as neurodegenerative disorders of unknown etiology on the basis of pathologic changes being confined to the CNS. With the transmission of kuru and CJD to apes, investigators began to view these diseases as infectious CNS illnesses caused by slow viruses. Even though the familial nature of a subset of CJD cases was well described, the significance of this observation became more obscure with the transmission of CJD to animals. Eventually the meaning of heritable CJD became clear with the discovery of mutations in the PRNP gene of these patients. The prion concept explains how a disease can manifest as a heritable as well as an infectious illness. Moreover, the hallmark of all prion diseases, whether sporadic, dominantly inherited, or acquired by infection, is that they involve the aberrant metabolism of PrP.

A major feature that distinguishes prions from viruses is the finding that both PrP isoforms are encoded by a chromosomal gene. In humans, the PrP gene is designated *PRNP* and is located on the short arm of chromosome 20. Limited proteolysis of PrP^{Sc} produces a smaller, protease-resistant molecule of ~ 142 amino acids designated PrP 27-30; PrP^{C} is completely hydrolyzed under the same conditions (**Fig. 383-2**). In the presence of detergent, PrP 27-30 polymerizes into amyloid. Prion rods formed by limited proteolysis and detergent extraction are indistinguishable from the filaments that aggregate to form PrP amyloid plaques in the CNS. Both the rods and the PrP amyloid filaments found in brain tissue exhibit similar ultrastructural morphology and green-gold birefringence after staining with Congo red dye.

Figure 383-2

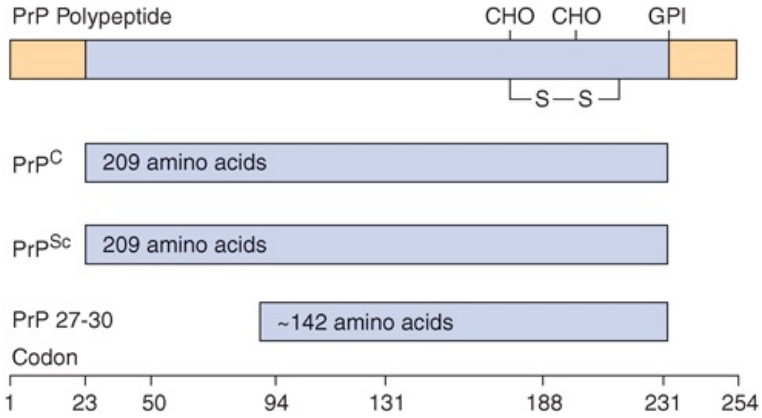

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Prion protein isoforms. Bar diagram of Syrian hamster PrP, which consists of 254 amino acids. After processing of the NH_2 and COOH termini, both PrP^C and PrP^{Sc} consist of 209 residues. After limited proteolysis, the NH_2 terminus of PrP^{Sc} is truncated to form PrP 27–30 composed of ~142 amino acids.

PRION STRAINS

The existence of prion strains raised the question of how heritable biologic information can be enciphered in a molecule other than nucleic acid. Various strains of prions have been defined by incubation times and the distribution of neuronal vacuolation. Subsequently, the patterns of PrP^{Sc} deposition were found to correlate with vacuolation profiles, and these patterns were also used to characterize prion strains.

Persuasive evidence that strain-specific information is enciphered in the tertiary structure of PrP^{Sc} comes from transmission of two different inherited human prion diseases to mice expressing a chimeric human-mouse PrP transgene. In FFI, the protease-resistant fragment of PrP^{Sc} after deglycosylation has a molecular mass of 19 kDa, whereas in fCJD and most sporadic prion diseases, it is 21 kDa **(Table 383-3)**. This difference in molecular mass was shown to be due to different sites of proteolytic cleavage at the NH_2 termini of the two human PrP^{Sc} molecules, reflecting different tertiary structures. These distinct conformations were not unexpected because the amino acid sequences of the PrPs differ.

Table 383-3 Distinct Prion Strains Generated in Humans with Inherited Prion Diseases and Transmitted to Transgenic Mice[a]

Inoculum	HostSpecies	Host PrPGenotype	Incubation Time [days ± SEM] (n/n_0)	PrP^{Sc}(kDa)
None	Human	FFI(D178N, M129)		19
FFI	Mouse	Tg(MHu2M)	206 ± 7 (7/7)	19
FFI → Tg(MHu2M)	Mouse	Tg(MHu2M)	136 ± 1 (6/6)	19
None	Human	fCJD(E200K)		21
fCJD	Mouse	Tg(MHu2M)	170 ± 2 (10/10)	21
fCJD → Tg(MHu2M)	Mouse	Tg(MHu2M)	167 ± 3 (15/15)	21

[a]Tg(MHu2M) mice express a chimeric mouse-human PrP gene. ***Notes:*** Clinicopathologic phenotype is determined by the conformation of PrPSc in accord with the results of the transmission of human prions from patients with FFI to transgenic mice. fCJD, familial Creutzfeldt-Jakob disease; FFI, fatal familial insomnia.

Extracts from the brains of patients with FFI transmitted disease into mice expressing a chimeric human-mouse PrP transgene and induced formation of the 19-kDa PrP^{Sc}, whereas brain extracts from fCJD and sCJD patients produced the 21-kDa PrP^{Sc} in mice expressing the same transgene. On second passage, these differences were maintained, demonstrating that chimeric PrP^{Sc} can exist in

two different conformations based on the sizes of the protease-resistant fragments, even though the amino acid sequence of PrP^{Sc} is invariant.

This analysis was extended when patients with sporadic fatal insomnia (sFI) were identified. Although they did not carry a *PRNP* gene mutation, the patients demonstrated a clinical and pathologic phenotype that was indistinguishable from that of patients with FFI. Furthermore, 19-kDa PrP^{Sc} was found in their brains, and on passage of prion disease to mice expressing a chimeric human-mouse PrP transgene, 19-kDa PrP^{Sc} was also found. These findings indicate that the disease phenotype is dictated by the conformation of PrP^{Sc} and not the amino acid sequence. PrP^{Sc} acts as a template for the conversion of PrP^{C} into nascent PrP^{Sc}. On the passage of prions into mice expressing a chimeric hamster-mouse PrP transgene, a change in the conformation of PrP^{Sc} was accompanied by the emergence of a new strain of prions.

Many new strains of prions were generated using recombinant (rec) PrP produced in bacteria; recPrP was polymerized into amyloid fibrils and inoculated into transgenic mice expressing high levels of wild-type mouse PrP^{C}; about 500 days later, the mice died of prion disease. The incubation times of the "synthetic prions" in mice were dependent on the conditions used for polymerization of the amyloid fibrils. Highly stable amyloids gave rise to stable prions with long incubation times; low-stability amyloids led to prions with short incubation times. Amyloids of intermediate stability gave rise to prions with intermediate stabilities and intermediate incubation times. Such findings are consistent with earlier studies showing that the incubation times of synthetic and naturally occurring prions are directly proportional to the stability of the prion.

SPECIES BARRIER

Studies on the role of the primary and tertiary structures of PrP in the transmission of prion disease have given new insights into the pathogenesis of these maladies. The amino acid sequence of PrP encodes the species of the prion, and the prion derives its PrP^{Sc} sequence from the last mammal in which it was passaged. While the primary structure of PrP is likely to be the most important or even sole determinant of the tertiary structure of PrP^{C}, PrP^{Sc} seems to function as a template in determining the tertiary structure of nascent PrP^{Sc} molecules as they are formed from PrP^{C}. In turn, prion diversity appears to be enciphered in the conformation of PrP^{Sc}, and thus prion strains seem to represent different conformers of PrP^{Sc}.

In general, transmission of prion disease from one species to another is inefficient, in that not all intracerebrally inoculated animals develop disease, and those that fall ill do so only after long incubation times that can approach the natural life span of the animal. This "species barrier" to transmission is correlated with the degree of similarity between the amino acid sequences of PrP^{C} in the inoculated host and of PrP^{Sc} in the prion inoculum. The importance of sequence similarity between the host and donor PrP argues that PrP^{C} directly interacts with PrP^{Sc} in the prion conversion process.

SPORADIC AND INHERITED PRION DISEASES

Several different scenarios might explain the initiation of sporadic prion disease: (1) A somatic mutation may be the cause and thus follow a path similar to that for germ-line mutations in inherited disease. In this situation, the mutant PrP^{Sc} must be capable of targeting wild-type PrP^{C}, a process known to be possible for some mutations but less likely for others. (2) The activation energy barrier separating wild-type PrP^{C} from PrP^{Sc} could be crossed on rare occasions when viewed in the context of a population. Most individuals would be spared while presentations in the elderly with an incidence of ~1 per million would be seen. (3) PrP^{Sc} may be present at low levels in some normal cells, where it performs some important, as yet unknown, function. The level of PrP^{Sc} in such cells is hypothesized to be sufficiently low as to be not detected by routine bioassay. In some altered metabolic states, the cellular mechanisms for clearing PrP^{Sc} might become compromised and the rate of PrP^{Sc} formation would then begin to exceed the capacity of the cell to clear it. The third possible mechanism is attractive since it suggests PrP^{Sc} is not simply a misfolded protein, as proposed for the first and second mechanisms, but that it is an alternatively folded molecule with a function. Moreover, the multitude of conformational states that PrP^{Sc} can adopt, as described above, raises the possibility that PrP^{Sc} or another prion-like protein might function in a process like short-term memory where information storage occurs in the absence of new protein synthesis.

More than 40 different mutations resulting in nonconservative substitutions in the human *PRNP* gene have been found to segregate with inherited human prion diseases. Missense mutations and

expansions in the octapeptide repeat region of the gene are responsible for familial forms of prion disease. Five different mutations of the *PRNP* gene have been linked genetically to heritable prion disease.

Although phenotypes may vary dramatically within families, specific phenotypes tend to be observed with certain mutations. A clinical phenotype indistinguishable from typical sCJD is usually seen with substitutions at codons 180, 183, 200, 208, 210, and 232. Substitutions at codons 102, 105, 117, 198, and 217 are associated with the GSS variant of prion disease. The normal human PrP sequence contains five repeats of an eight-amino-acid sequence. Insertions from two to nine extra octarepeats frequently cause variable phenotypes ranging from a condition indistinguishable from sCJD to a slowly progressive dementing illness of many years' duration to an early-age-of-onset disorder that is similar to Alzheimer's disease. A mutation at codon 178 resulting in substitution of asparagine for aspartic acid produces FFI if a methionine is encoded at the polymorphic 129 residue on the same allele. Typical CJD is seen if the D178N mutation occurs with a valine encoded at position 129 of the same allele.

Human Prnp Gene Polymorphisms

Polymorphisms influence the susceptibility to sporadic, inherited, and infectious forms of prion disease. The methionine/valine polymorphism at position 129 not only modulates the age of onset of some inherited prion diseases but can also determine the clinical phenotype. The finding that homozygosity at codon 129 predisposes to sCJD supports a model of prion production that favors PrP interactions between homologous proteins.

Substitution of the basic residue lysine at position 218 in mouse PrP produced dominant-negative inhibition of prion replication in transgenic mice. This same lysine at position 219 in human PrP has been found in 12% of the Japanese population, and this group appears to be resistant to prion disease. Dominant-negative inhibition of prion replication was also found with substitution of the basic residue arginine at position 171; sheep with arginine are resistant to scrapie prions but are susceptible to BSE prions that were inoculated intracerebrally.

INFECTIOUS PRION DISEASES

Iatrogenic Cjd

Accidental transmission of CJD to humans appears to have occurred with corneal transplantation, contaminated electroencephalogram (EEG) electrode implantation, and surgical procedures. Corneas from donors with inapparent CJD have been transplanted to apparently healthy recipients who developed CJD after prolonged incubation periods. The same improperly decontaminated EEG electrodes that caused CJD in two young patients with intractable epilepsy caused CJD in a chimpanzee 18 months after their experimental implantation.

Surgical procedures may have resulted in accidental inoculation of patients with prions, presumably because some instrument or apparatus in the operating theater became contaminated when a CJD patient underwent surgery. Although the epidemiology of these studies is highly suggestive, no proof for such episodes exists.

DURA MATER GRAFTS

More than 160 cases of CJD after implantation of dura mater grafts have been recorded. All of the grafts were thought to have been acquired from a single manufacturer whose preparative procedures were inadequate to inactivate human prions. One case of CJD occurred after repair of an eardrum perforation with a pericardium graft.

HUMAN GROWTH HORMONE AND PITUITARY GONADOTROPIN THERAPY

The transmission of CJD prions from contaminated human growth hormone (hGH) preparations derived from human pituitaries has been responsible for fatal cerebellar disorders with dementia in >180 patients ranging in age from 10 to 41 years. These patients received injections of hGH every 2–4 days for 4–12 years. If it is thought that these patients developed CJD from injections of prion-contaminated hGH preparations, the possible incubation periods range from 4 to 30 years. Only recombinant hGH is now used therapeutically so that possible contamination with prions is no longer an issue. Four cases of CJD have occurred in women receiving human pituitary gonadotropin.

Variant Cjd

The restricted geographic occurrence and chronology of vCJD raised the possibility that BSE prions had been transmitted to humans through the consumption of tainted beef. More than 190 cases of vCJD have occurred, with >90% of these in Britain. vCJD has also been reported in people either living in or originating from France, Ireland, Italy, Netherlands, Portugal, Spain, Saudi Arabia, United States, Canada, and Japan.

The steady decline in the number of vCJD cases over the past decade argues that there will not be a prion disease epidemic in Europe, similar to those seen for BSE and kuru. What is certain is that prion-tainted meat should be prevented from entering the human food supply.

The most compelling evidence that vCJD is caused by BSE prions was obtained from experiments in

mice expressing the bovine PrP transgene. Both BSE and vCJD prions were efficiently transmitted to these transgenic mice and with similar incubation periods. In contrast to sCJD prions, vCJD prions did not transmit disease efficiently to mice expressing a chimeric human-mouse PrP transgene. Earlier studies with nontransgenic mice suggested that vCJD and BSE might be derived from the same source because both inocula transmitted disease with similar but very long incubation periods.

Attempts to determine the origin of BSE and vCJD prions have relied on passaging studies in mice, some of which are described above, as well as studies of the conformation and glycosylation of PrP^{Sc}. One scenario suggests that a particular conformation of bovine PrP^{Sc} was selected for heat resistance during the rendering process and was then reselected multiple times as cattle infected by ingesting prion-contaminated meat and bone meal (MBM) were slaughtered and their offal rendered into more MBM.

Neuropathology

Frequently the brains of patients with CJD have no recognizable abnormalities on gross examination. Patients who survive for several years have variable degrees of cerebral atrophy.

On light microscopy, the pathologic hallmarks of CJD are spongiform degeneration and astrocytic gliosis. The lack of an inflammatory response in CJD and other prion diseases is an important pathologic feature of these degenerative disorders. Spongiform degeneration is characterized by many 1- to 5-μm vacuoles in the neuropil between nerve cell bodies. Generally the spongiform changes occur in the cerebral cortex, putamen, caudate nucleus, thalamus, and molecular layer of the cerebellum. Astrocytic gliosis is a constant but nonspecific feature of prion diseases. Widespread proliferation of fibrous astrocytes is found throughout the gray matter of brains infected with CJD prions. Astrocytic processes filled with glial filaments form extensive networks.

Amyloid plaques have been found in ~10% of CJD cases. Purified CJD prions from humans and animals exhibit the ultrastructural and histochemical characteristics of amyloid when treated with detergents during limited proteolysis. In first passage from some human Japanese CJD cases, amyloid plaques have been found in mouse brains. These plaques stain with antibodies raised against PrP.

The amyloid plaques of GSS disease are morphologically distinct from those seen in kuru or scrapie. GSS plaques consist of a central dense core of amyloid surrounded by smaller globules of amyloid. Ultrastructurally, they consist of a radiating fibrillar network of amyloid fibrils, with scant or no neuritic degeneration. The plaques can be distributed throughout the brain but are most frequently found in the cerebellum. They are often located adjacent to blood vessels. Congophilic angiopathy has been noted in some cases of GSS disease.

In vCJD, a characteristic feature is the presence of "florid plaques." These are composed of a central core of PrP amyloid, surrounded by vacuoles in a pattern suggesting petals on a flower.

Clinical Features

Nonspecific prodromal symptoms occur in about a third of patients with CJD and may include fatigue, sleep disturbance, weight loss, headache, anxiety, vertigo, malaise, and ill-defined pain. Most patients with CJD present with deficits in higher cortical function. These deficits almost always progress over weeks or months to a state of profound dementia characterized by memory loss, impaired judgment, and a decline in virtually all aspects of intellectual function. A few patients present with either visual impairment or cerebellar gait and coordination deficits. Frequently the cerebellar deficits are rapidly followed by progressive dementia. Visual problems often begin with blurred vision and diminished acuity, rapidly followed by dementia.

Other symptoms and signs include extrapyramidal dysfunction manifested as rigidity, masklike facies, or (less commonly) choreoathetoid movements; pyramidal signs (usually mild); seizures (usually major motor) and, less commonly, hypoesthesia; supranuclear gaze palsy; optic atrophy; and vegetative signs such as changes in weight, temperature, sweating, or menstruation.

MYOCLONUS

Most patients (~90%) with CJD exhibit myoclonus that appears at various times throughout the illness. Unlike other involuntary movements, myoclonus persists during sleep. Startle myoclonus elicited by loud sounds or bright lights is frequent. It is important to stress that myoclonus is neither specific nor confined to CJD and tends to occur later in the course of CJD. Dementia with myoclonus can also be due to Alzheimer's disease (AD) (Chap. 371), dementia with Lewy bodies (Chap. 371), corticobasal degeneration (Chap. 371), cryptococcal encephalitis (Chap. 202), or the myoclonic epilepsy disorder Unverricht-Lundborg disease (Chap. 369).

CLINICAL COURSE

In documented cases of accidental transmission of CJD to humans, an incubation period of 1.5–2 years preceded the development of clinical disease. In other cases, incubation periods of up to 40 years have been suggested. Most patients with CJD live 6–12 months after the onset of clinical signs and symptoms, whereas some live for up to 5 years.

Diagnosis

The constellation of dementia, myoclonus, and periodic electrical bursts in an afebrile 60-year-old patient generally indicates CJD. Clinical abnormalities in CJD are confined to the CNS. Fever, elevated sedimentation rate, leukocytosis in blood, or a pleocytosis in cerebrospinal fluid (CSF) should alert the physician to another etiology to explain the patient's CNS dysfunction.

Variations in the typical course appear in inherited and transmitted forms of the disease. fCJD has an earlier mean age of onset than sCJD. In GSS disease, ataxia is usually a prominent and presenting feature, with dementia occurring late in the disease course. GSS disease typically presents earlier than CJD (mean age 43 years) and is typically more slowly progressive than CJD; death usually occurs within 5 years of onset. FFI is characterized by insomnia and dysautonomia; dementia occurs only in the terminal phase of the illness. Rare sporadic cases have been identified. vCJD has an unusual clinical course, with a prominent psychiatric prodrome that may include visual hallucinations and early ataxia, while frank dementia is usually a late sign of vCJD.

Differential Diagnosis

Many conditions may mimic CJD superficially. Dementia with Lewy bodies (Chap. 371) is the most common disorder to be mistaken for CJD. It can present in a subacute fashion with delirium, myoclonus, and extrapyramidal features. Other neurodegenerative disorders (Chap. 371) to consider include AD, frontotemporal dementia, corticobasal degeneration, progressive supranuclear palsy, ceroid lipofuscinosis, and myoclonic epilepsy with Lafora bodies (Chap. 369). The absence of abnormalities on diffusion-weighted and fluid-attenuated inversion recovery (FLAIR) MRI will almost always distinguish these dementing conditions from CJD.

Hashimoto's encephalopathy, which presents as a subacute progressive encephalopathy with myoclonus and periodic triphasic complexes on the EEG, should be excluded in every case of suspected CJD. It is diagnosed by the finding of high titers of antithyroglobulin or antithyroid peroxidase (antimicrosomal) antibodies in the blood and improves with glucocorticoid therapy. Unlike CJD, fluctuations in severity typically occur in Hashimoto's encephalopathy.

Intracranial vasculitides (Chap. 326) may produce nearly all of the symptoms and signs associated with CJD, sometimes without systemic abnormalities. Myoclonus is exceptional with cerebral vasculitis, but focal seizures may confuse the picture. Prominent headache, absence of myoclonus, stepwise change in deficits, abnormal CSF, and focal white matter changes on MRI or angiographic abnormalities all favor vasculitis.

Paraneoplastic conditions, particularly limbic encephalitis and cortical encephalitis, can also mimic CJD. In many of these patients, dementia appears prior to the diagnosis of a tumor, and in some, no tumor is ever found. Detection of the paraneoplastic antibodies is often the only way to distinguish these cases from CJD.

Other diseases that can simulate CJD include neurosyphilis (Chap. 169), AIDSdementia complex (Chap. 189), progressive multifocal leukoencephalopathy (Chap. 381), subacute sclerosing panencephalitis, progressive rubella panencephalitis, herpes simplex encephalitis (Chap. 381), diffuse intracranial tumor (gliomatosis cerebri; Chap. 379), anoxic encephalopathy, dialysis dementia, uremia, hepatic encephalopathy, voltage-gated potassium channel (VGkC) autoimmune encephalopathy and lithium or bismuthintoxication.

Laboratory Tests

The only specific diagnostic tests for CJD and other human prion diseases measure PrP^{Sc}. The most widely used method involves limited proteolysis that generates PrP 27-30, which is detected by immunoassay after denaturation. The conformation-dependent immunoassay (CDI) is based on immunoreactive epitopes that are exposed in PrP^{C} but buried in PrP^{Sc}. In humans, the diagnosis of CJD can be established by brain biopsy if PrP^{Sc} is detected. If no attempt is made to measure PrP^{Sc}, but the constellation of pathologic changes frequently found in CJD is seen in a brain biopsy, then the diagnosis is reasonably secure (see "Neuropathology," above). The high sensitivity and specificity of cortical ribboning and basal ganglia hyperintensity on FLAIR and diffusion-weighted MRI for the diagnosis of CJD have greatly diminished the need for brain biopsy in patients with suspected CJD. Because PrP^{Sc} is not uniformly distributed throughout the CNS, the apparent absence of PrP^{Sc} in a limited sample such as a biopsy does not rule out prion disease. At autopsy, sufficient brain samples should be taken for both PrP^{Sc} immunoassay, preferably by CDI, and immunohistochemistry of tissue sections.

To establish the diagnosis of either sCJD or familial prion disease, sequencing the *PRNP* gene must be performed. Finding the wild-type *PRNP* gene sequence permits the diagnosis of sCJD if there is no history to suggest infection from an exogenous source of prions. The identification of a mutation in the *PRNP* gene sequence that encodes a nonconservative amino acid substitution argues for familial prion disease.

CT may be normal or show cortical atrophy. MRI is valuable for distinguishing sCJD from most other conditions. On FLAIR sequences and diffusion-weighted imaging, ~90% of patients show increased intensity in the basal ganglia and cortical ribboning **(Fig. 383-3)**. This pattern is not seen with other neurodegenerative disorders but has been seen infrequently with viral encephalitis, paraneoplastic syndromes, or seizures. When the typical MRI pattern is present, in the proper clinical setting, diagnosis is facilitated. However, some cases of sCJD do not show this typical pattern, and other early diagnostic approaches are still needed.

Figure 383-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

T2-weighted (FLAIR) MRI showing hyperintensity in the cortex in a patient with sporadic CJD. This so-called "cortical ribboning" along with increased intensity in the basal ganglia on T2 or diffusion-weighted imaging can aid in the diagnosis of CJD.

CSF is nearly always normal but may show protein elevation and, rarely, mild pleocytosis. Although the stress protein 14-3-3 is elevated in the CSF of some patients with CJD, similar elevations of 14-3-3 are found in patients with other disorders; thus this elevation is not specific. Similarly, elevations of CSF neuron-specific enolase and tau occur in CJD but lack specificity for diagnosis.

The EEG is often useful in the diagnosis of CJD, although only about 60% of individuals show the typical pattern. During the early phase of CJD, the EEG is usually normal or shows only scattered theta activity. In most advanced cases, repetitive, high-voltage, triphasic, and polyphasic sharp discharges are seen, but in many cases their presence is transient. The presence of these stereotyped periodic bursts of <200 ms duration, occurring every 1–2 s, makes the diagnosis of CJD very likely. These discharges are frequently but not always symmetric; there may be a one-sided predominance in amplitude. As CJD progresses, normal background rhythms become fragmentary and slower.

Care of Cjd Patients

Although CJD should not be considered either contagious or communicable, it is transmissible. The risk of accidental inoculation by aerosols is very small; nonetheless, procedures producing aerosols should be performed in certified biosafety cabinets. Biosafety level 2 practices, containment equipment, and facilities are recommended by the Centers for Disease Control and Prevention and the National Institutes of Health. The primary problem in caring for patients with CJD is the inadvertent infection of health care workers by needle and stab wounds. Electroencephalographic and electromyographic needles should not be reused after studies on patients with CJD have been performed.

There is no reason for pathologists or other morgue employees to resist performing autopsies on patients whose clinical diagnosis was CJD. Standard microbiologic practices outlined here, along with specific recommendations for decontamination, seem to be adequate precautions for the care of patients with CJD and the handling of infected specimens.

Decontamination of Cjd Prions

Prions are extremely resistant to common inactivation procedures, and there is some disagreement about the optimal conditions for sterilization. Some investigators recommend treating CJD-contaminated materials once with 1 *N* NaOH at room temperature, but we believe this procedure may be inadequate for sterilization. Autoclaving at 134°C for 5 h or treatment with 2 *N* NaOH for several hours is recommended for sterilization of prions. The term *sterilization* implies complete destruction of prions; any residual infectivity can be hazardous. Recent studies show that sCJD prions bound to stainless steel surfaces are resistant to inactivation by autoclaving at 134°C for 2 h; exposure of bound prions to an acidic detergent solution prior to autoclaving rendered prions susceptible to inactivation.

Prevention and Therapeutics

There is no known effective therapy for preventing or treating CJD. The finding that phenothiazines and acridines inhibit PrP^{Sc} formation in cultured cells led to clinical studies of quinacrine in CJD patients. Unfortunately, quinacrine failed to slow the rate of cognitive decline in CJD, possibly because therapeutic concentrations in the brain were not achieved. Although inhibition of the P-glycoprotein (Pgp) transport system resulted in substantially increased quinacrine levels in the brains of mice, the prion incubation times were not extended by treatment with the drug. Whether such an approach can be used to treat CJD remains to be established.

Like the acridines, anti-PrP antibodies have been shown to eliminate PrP^{Sc} from cultured cells. Additionally, such antibodies in mice, either administered by injection or produced from a transgene, have been shown to prevent prion disease when prions are introduced by a peripheral route, such as intraperitoneal inoculation. Unfortunately, the antibodies were ineffective in mice inoculated intracerebrally with prions. Several drugs, including pentosan polysulfate as well as porphyrin and phenylhydrazine derivatives, delay the onset of disease in animals inoculated intracerebrally with prions if the drugs are given intracerebrally beginning soon after inoculation.

PRION-LIKE PROTEINS CAUSING OTHERNEURODEGENERATIVE DISEASES

There is mounting evidence that prion-like changes in protein conformation underlie Alzheimer's (AD), Parkinson's (PD) and Huntington's (HD) diseases as well as the frontotemporal dementias (FTDs) and amyotrophic lateral sclerosis (ALS). Experimental studies have shown that transgenic mice expressing mutant amyloid precursor protein (APP) develop amyloid plaques containing fibrils composed of the amyloid beta (Aβ) peptide about a year after inoculation with extracts prepared from the brains of patients with AD. Mutant tau aggregates in transgenic mice and cultured cells can trigger the aggregation of wild-type tau into fibrils that resemble those in neurofibrillary tangles and Pick bodies that have been found in AD, FTDs, Pick's disease, and some cases of posttraumatic head injury. In patients with advanced PD who received grafts of fetal substantia nigral neurons, Lewy bodies containing β-sheet–rich α-synuclein have been identified in grafted cells about 10 years after transplantation. These findings argue for the axonal transport of misfolded α-synuclein crossing into grafted neurons, where it initiates aggregation of nascent α-synuclein into fibrils that coalesce to form Lewy bodies.

Taken together, a wealth of data argues that all neurodegenerative diseases are caused by proteins that undergo aberrant processing, which results in their assembly into amyloid fibrils. In each degenerative brain disease, prion-like protein processing is responsible for the accumulation of a particular protein in an altered state that leads to neurodegeneration. Interestingly, once these aberrant, prion-like proteins have polymerized into amyloid fibrils, they are probably inert. Amyloid plaques containing PrP^{Sc} are a nonobligatory feature of prion disease in humans and animals. Furthermore, amyloid plaques in AD do not correlate with the level of dementia; however, the level of soluble (oligomeric) Aβ peptide does correlate with memory loss and other intellectual deficits.

FURTHER READINGS

Aguzzi A et al: Molecular mechanisms of prion pathogenesis. Annu Rev Pathol 3:11, 2008[PMID: 18233951] [Full Text]

Caugey B et al: Getting a grip on prions: Oligomers, amyloids, and pathological membrane interactions. Annu Rev Biochem. 78:177, 2009

Colby DW et al: Prions. Cold Spring Harb Perspect Biol 3:a006833, 2011

Jucker M: The benefits and limitations of animal models for translation research in neurodegenerative diseases. Nat Med 16:1210, 2010[PMID: 21052075] [Full Text]

Olanow CW et al: Is Parkinson's disease a prion disorder? Proc Natl Acad Sci USA 106:12571, 2009[PMID: 19666621] [Full Text]

Prusiner SB: Prions, in *Fields Virology*, DM Knipe et al (eds). Philadelphia, Lippincott Willams & Wilkins, 2007, pp 3059–3092

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

PERIPHERAL NEUROPATHY: INTRODUCTION

Peripheral nerves are composed of sensory, motor, and autonomic elements. Diseases can affect the cell body of a neuron or its peripheral processes, namely the axons or the encasing myelin sheaths. Most peripheral nerves are mixed and contain sensory and motor as well as autonomic fibers. Nerves can be subdivided into three major classes: large myelinated, small myelinated, and small unmyelinated. Motor axons are usually large myelinated fibers that conduct rapidly (approximately 50 m/s). Sensory fibers may be any of the three types. Large-diameter sensory fibers conduct proprioception and vibratory sensat ion to the brain, while the smaller-diameter myelinated and unmyelinated fibers transmit pain and temperature sensation. Autonomic nerves are also small in diameter. Thus, peripheral neuropathies can impair sensory, motor, or autonomic function, either singly or in combination. Peripheral neuropathies are further classified into those that primarily affect the cell body (e.g., neuronopathy or ganglionopathy), myelin (myelinopathy), and the axon (axonopathy). These different classes of peripheral neuropathies have distinct clinical and electrophysiologic features. This chapter discusses the clinical approach to a patient suspected of having a peripheral neuropathy, as well as specific neuropathies, including hereditary and acquired neuropathies. The inflammatory neuropathies are discussed in Chap. 385.

GENERAL APPROACH

In approaching a patient with a neuropathy, the clinician has three main goals: (1) identify where the lesion is, (2) identify the cause, and (3) determine the proper treatment. The first goal is accomplished by obtaining a thorough history, neurologic examination, and electrodiagnostic and other laboratory studies **(Fig. 384-1)**. While gathering this information, seven key questions are asked (Table 384-1), the answers to which can usually identify the category of pathology that is present **(Table 384-2)**. Despite an extensive evaluation, in approximately half of patients no etiology is ever found; these patients typically have a predominately sensory polyneuropathy and have been labeled as having idiopathic or cryptogenic sensory polyneuropathy (CSPN).

Table 384-1 Approach to Neuropathic Disorders: Seven Key Questions

1. **What systems are involved?**
   – Motor, sensory, autonomic, or combinations
2. **What is the distribution of weakness?**
   – Only distal versus proximal and distal
   – Focal/asymmetric versus symmetric
3. **What is the nature of the sensory involvement?**
   – Temperature loss or burning or stabbing pain (e.g., small fiber)
   – Vibratory or proprioceptive loss (e.g., large fiber)
4. *Is there evidence of upper motor neuron involvement?*
   – Without sensory loss
   – With sensory loss
5. **What is the temporal evolution?**
   – Acute (days to 4 weeks)
   – Subacute (4 to 8 weeks)
   – Chronic (>8 weeks)
6. **Is there evidence for a hereditary neuropathy?**
   – Family history of neuropathy
   – Lack of sensory symptoms despite sensory signs
7. **Are there any associated medical conditions?**
   – Cancer, diabetes mellitus, connective tissue disease or other autoimmune diseases, infection (e.g., HIV, Lyme disease, leprosy)
   – Medications including over-the-counter drugs that may cause a toxic neuropathy
   – Preceding events, drugs, toxins

Table 384-2 Patterns of Neuropathic Disorders

Pattern 1: Symmetric proximal and distal weakness with sensory loss
Consider: inflammatory demyelinating polyneuropathy (GBS and CIDP)
Pattern 2: Symmetric distal sensory loss with or without distal weakness
Consider: cryptogenic or idiopathic sensory polyneuropathy (CSPN), diabetes mellitus and other metabolic disorders, drugs, toxins, hereditary (Charcot-Marie-Tooth, amyloidosis, and others)
Pattern 3: Asymmetric distal weakness with sensory loss
With involvement of multiple nerves
Consider: multifocal CIDP, vasculitis, cryoglobulinemia, amyloidosis, sarcoid, infectious (leprosy, Lyme, hepatitis B or C, HIV, CMV), hereditary neuropathy with liability to pressure palsies (HNPP), tumor infiltration
With involvement of single nerves/regions

Consider: may be any of the above but also could be compressive mononeuropathy, plexopathy, or radiculopathy
Pattern 4: Asymmetric proximal and distal weakness with sensory loss
Consider: polyradiculopathy or plexopathy due to diabetes mellitus, meningeal carcinomatosis or lymphomatosis, hereditary plexopathy (HNPP, HNA), idiopathic
Pattern 5: Asymmetric distal weakness without sensory loss
With upper motor neuron findings
Consider: motor neuron disease
Without upper motor neuron findings
Consider: progressive muscular atrophy, juvenile monomelic amyotrophy (Hirayama disease), multifocal motor neuropathy, multifocal acquired motor axonopathy
Pattern 6: Symmetric sensory loss and distal areflexia with upper motor neuron findings
Consider: Vitamin B12, vitamin E, and copper deficiency with combined system degeneration with peripheral neuropathy, hereditary leukodystrophies (e.g., adrenomyeloneuropathy)
Pattern 7: Symmetric weakness without sensory loss
With proximal and distal weakness
Consider: spinal muscular atrophy
With distal weakness
Consider: hereditary motor neuropathy ("distal" SMA) or atypical CMT
Pattern 8: Asymmetric proprioceptive sensory loss without weakness
Consider causes of a sensory neuronopathy (ganglionopathy):
Cancer (paraneoplastic)
Sjögren syndrome
Idiopathic sensory neuronopathy (possible GBS variant)
Cisplatin and other chemotherapeutic agents
Vitamin B6 toxicity
HIV-related sensory neuronopathy
Pattern 9: Autonomic Symptoms and Signs
Consider neuropathies associated with prominent autonomic dysfunction:
Hereditary sensory and autonomic neuropathy
Amyloidosis (familial and acquired)
Diabetes mellitus
Idiopathic pandysautonomia (may be a variant of Guillain-Barré syndrome)
Porphyria
HIV-related autonomic neuropathy
Vincristine and other chemotherapeutic agents

Abbreviations: CIDP, chronic inflammatory demyelinating polyneuropathy; CMT, Charcot-Marie-Tooth disease; CMV, cytomegalovirus; GBS, Guillain-Barré syndrome; HIV, human immunodeficiency virus; HNA, hereditary neuralgic amyotrophy; SMA, spinal muscular atrophy.

Information from the History and Physical Examination: Seven Key Questions (Table 384-1)

WHAT SYSTEMS ARE INVOLVED?

It is important to determine if the patient's symptoms and signs are motor, sensory, autonomic, or a combination of these. If the patient has only weakness without any evidence of sensory or autonomic dysfunction, a motor neuropathy, neuromuscular junction abnormality, or myopathy should be considered. Some peripheral neuropathies are associated with significant autonomic nervous system dysfunction. Symptoms of autonomic involvement include fainting spells or orthostatic lightheadedness; heat intolerance; or any bowel, bladder, or sexual dysfunction (Chap. 375). There will typically be an orthostatic fall in blood pressure without an appropriate increase in heart rate. Autonomic dysfunction in the absence of diabetes should alert the clinician to the possibility of amyloid polyneuropathy. Rarely, a pandysautonomic syndrome can be the only manifestation of a peripheral neuropathy without other motor or sensory findings. The majority of neuropathies are predominantly sensory in nature.

WHAT IS THE DISTRIBUTION OF WEAKNESS?

Delineating the pattern of weakness, if present, is essential for diagnosis, and in this regard two additional questions should be answered: (1) Does the weakness only involve the distal extremity or is it both proximal and distal? and (2) Is the weakness focal and asymmetric or is it symmetric? Symmetric proximal and distal weakness is the hallmark of acquired immune demyelinating polyneuropathies, both the acute form [acute inflammatory demyelinating polyneuropathy (AIDP) also known as Guillain-Barré syndrome (GBS)] and the chronic form [chronic inflammatory demyelinating polyneuropathy (CIDP)]. The importance of finding symmetric proximal and distal weakness in a patient who presents with both motor and sensory symptoms cannot be overemphasized because this identifies the important subset of patients who may have a treatable acquired demyelinating neuropathic disorder (i.e., AIDP or CIDP).

Findings of an asymmetric or multifocal pattern of weakness narrows the differential diagnosis. Some neuropathic disorders may present with unilateral extremity weakness. In the absence of sensory symptoms and signs, such weakness evolving over weeks or months would be worrisome for motor neuron disease [e.g., amyotrophic lateral sclerosis (ALS)], but it would be important to exclude multifocal motor neuropathy that may be treatable (Chap. 374). In a patient presenting with asymmetric subacute or acute sensory and motor symptoms and signs, radiculopathies, plexopathies, compressive mononeuropathies, or multiple mononeuropathies (e.g., mononeuropathy multiplex) must be considered.

WHAT IS THE NATURE OF THE SENSORY INVOLVEMENT?

The patient may have loss of sensation (numbness), altered sensation to touch (hyperpathia or allodynia), or uncomfortable spontaneous sensations (tingling, burning, or aching) (Chap. 23). Neuropathic pain can be burning, dull, and poorly localized (protopathic pain), presumably transmitted by polymodal C nociceptor fibers, or sharp and lancinating (epicritic pain), relayed by A-delta fibers. If pain and temperature perception are lost, while vibratory and position sense are preserved along with muscle strength, deep tendon reflexes, and normal nerve conduction studies, a small-fiber neuropathy is likely. This is important, as the most likely cause of small-fiber neuropathies, when one is identified, is diabetes mellitus or glucose intolerance. Amyloid

neuropathy should be considered as well in such cases, but most of these small-fiber neuropathies remain idiopathic in nature despite extensive evaluation.

Severe proprioceptive loss also narrows the differential diagnosis. Affected patients will note imbalance, especially in the dark. A neurologic examination revealing a dramatic loss of proprioception with vibration loss and normal strength should alert the clinician to consider a sensory neuronopathy/ganglionopathy (Table 384-2, Pattern 8). In particular, if this loss is asymmetric or affects the arms more than the legs, this pattern suggests a non-length-dependent process as seen in sensory neuronopathies.

IS THERE EVIDENCE OF UPPER MOTOR NEURON INVOLVEMENT?

If the patient presents with symmetric distal sensory symptoms and signs suggestive of a distal sensory neuropathy, but there is additional evidence of symmetric upper motor neuron involvement (Chap. 22), the physician should consider a disorder such as combined system degeneration with neuropathy. The most common cause for this pattern is vitamin B12 deficiency, but other causes of combined system degeneration with neuropathy should be considered (e.g., copper deficiency, HIV infection, severe hepatic disease, adrenomyeloneuropathy).

WHAT IS THE TEMPORAL EVOLUTION?

It is important to determine the onset, duration, and evolution of symptoms and signs. Does the disease have an acute (days to 4 weeks), subacute (4–8 weeks), or chronic (>8 weeks) course? Is the course monophasic, progressive, or relapsing? Most neuropathies are insidious and slowly progressive in nature. Neuropathies with acute and subacute presentations include GBS, vasculitis, and radiculopathies related to diabetes or Lyme disease. A relapsing course can be present in CIDP and porphyria.

IS THERE EVIDENCE FOR A HEREDITARY NEUROPATHY?

In patients with slowly progressive distal weakness over many years with very little in the way of sensory symptoms yet with significant sensory deficits on clinical examination, the clinician should consider a hereditary neuropathy (e.g., Charcot-Marie-Tooth disease or CMT). On examination, the feet may show arch and toe abnormalities (high or flat arches, hammertoes); scoliosis may be present. In suspected cases, it may be necessary to perform both neurologic and electrophysiologic studies on family members in addition to the patient.

DOES THE PATIENT HAVE ANY OTHER MEDICAL CONDITIONS?

It is important to inquire about associated medical conditions (e.g., diabetes mellitus, systemic lupus erythematosus); preceding or concurrent infections (e.g. diarrheal illness preceding GBS); surgeries (e.g., gastric bypass and nutritional neuropathies); medications (toxic neuropathy); including over-the-counter vitamin preparations (B_6); alcohol; dietary habits; and use of dentures (e.g., fixatives contain zinc that can lead to copper deficiency).

Pattern Recognition Approach to Neuropathic Disorders

Based upon the answers to the seven key questions, neuropathic disorders can be classified into several patterns based on the distribution or pattern of sensory, motor, and autonomic involvement (Table 384-2). Each pattern has a limited differential diagnosis. A final diagnosis is established by utilizing other clues such as the temporal course, presence of other disease states, family history, and information from laboratory studies.

Electrodiagnostic Studies

The electrodiagnostic (EDx) evaluation of patients with a suspected peripheral neuropathy consists of nerve conduction studies (NCS) and needle electromyography (EMG). In addition, studies of autonomic function can be valuable. The electrophysiologic data provides additional information about the distribution of the neuropathy that will support or refute the findings from the history and physical examination; it can confirm whether the neuropathic disorder is a mononeuropathy, multiple mononeuropathy (mononeuropathy multiplex), radiculopathy, plexopathy, or generalized polyneuropathy. Similarly, EDx evaluation can ascertain whether the process involves only sensory fibers, motor fibers, autonomic fibers, or a combination of these. Finally, the electrophysiologic data can help distinguish axonopathies from myelinopathies as well as axonal degeneration secondary to ganglionopathies from the more common length-dependent axonopathies.

NCS are most helpful in classifying a neuropathy as being due to axonal degeneration or segmental demyelination **(Table 384-3)**. In general, low-amplitude potentials with relatively preserved distal latencies, conduction velocities, and late potentials, along with fibrillations on needle EMG, suggest an axonal neuropathy. On the other hand, slow conduction velocities, prolonged distal latencies and late potentials, relatively preserved amplitudes, and the absence of fibrillations on needle EMG imply a primary demyelinating neuropathy. The presence of nonuniform slowing of conduction velocity, conduction block, or temporal dispersion further suggests an acquired demyelinating neuropathy (e.g., GBS or CIDP) as opposed to a hereditary demyelinating neuropathy (e.g., CMT type 1).

Table 384-3 Electrophysiologic Features: Axonal Degeneration vs. Segmental Demyelination

	Axonal Degeneration	Segmental Demyelination
Motor Nerve Conduction Studies		
CMAP amplitude	Decreased	Normal (except with CB)
Distal latency	Normal	Prolonged
Conduction velocity	Normal	Slow
Conduction block	Absent	Present
Temporal dispersion	Absent	Present
F wave	Normal or absent	Prolonged or Absent
H reflex	Normal or absent	Prolonged or Absent
Sensory Nerve Conduction Studies		
SNAP amplitude	Decreased	Normal
Distal latency	Normal	Prolonged
Conduction velocity	Normal	Slow
Needle EMG		
Spontaneous activity		
Fibrillations	Present	Absent
Fasciculations	Present	Absent
Motor unit potentials		
Recruitment	Decreased	Decreased
Morphology	Long duration/polyphasic	Normal

Abbreviations: CB, conduction block; CMAP, compound motor action potential; EMG, electromyography; SNAP, sensory nerve action potential.

Autonomic studies are used to assess small myelinated (A-delta) or unmyelinated (C) nerve fiber involvement. Such testing includes heart rate response to deep breathing, heart rate, and blood pressure response to both the Valsalva maneuver and tilt-table testing, and quantitative sudomotor axon reflex testing (Chap. 375). These studies are particularly useful in patients who have pure small-fiber neuropathy or autonomic neuropathy in which routine NCS are normal.

Other Important Laboratory Information

In patients with generalized symmetric peripheral neuropathy, a standard laboratory evaluation should include a complete blood count, basic chemistries including serum electrolytes and tests of renal and hepatic function, fasting blood glucose (FBS), HbA_{1c}, urinalysis, thyroid function tests, B_{12}, folate, erythrocyte sedimentation rate (ESR), rheumatoid factor, antinuclear antibodies (ANA), serum protein electrophoresis (SPEP), and urine for Bence Jones protein. An oral glucose tolerance test is indicated in patients with painful sensory neuropathies even if FBS and HbA_{1c} are normal, as the test is abnormal in about one-third of such patients. Serum and urine immunofixation electrophoresis (IFE) are necessary, rather than just an SPEP, in patients with a demyelinating neuropathy or if one suspects amyloidosis (e.g., severe autonomic symptoms) as an IFE is more sensitive at identifying a monoclonal gammopathy. A skeletal survey should be performed in patients with acquired demyelinating neuropathies and M-spikes to look for osteosclerotic or lytic lesions. Patients with monoclonal gammopathy should also be referred to a hematologist for consideration of a bone marrow biopsy. In addition to the above tests, patients with a mononeuropathy multiplex pattern of involvement should have a vasculitis workup, including antineutrophil cytoplasmic antibodies (ANCA), cryoglobulins, hepatitis serology, Western blot for Lyme disease, HIV, and occasionally a cytomegalovirus (CMV), titer.

There are many autoantibody panels (various antiganglioside antibodies) marketed for screening routine neuropathy patients for a treatable condition. These autoantibodies have no proven clinical utility or added benefit beyond the information obtained from a complete clinical examination and detailed EDx. A heavy metal screen is also not necessary as a screening procedure, unless there is a history of possible exposure or suggestive features on examination (e.g., severe painful sensorimotor and autonomic neuropathy and alopecia—thallium; severe painful sensorimotor neuropathy with or without GI disturbance and Mee's lines—arsenic; wrist or finger extensor weakness and anemia with basophilic stippling of red blood cells—lead).

In patients with suspected GBS or CIDP, a lumbar puncture is indicated to look for an elevated cerebral spinal fluid (CSF) protein. In idiopathic cases of GBS and CIDP, there should not be pleocytosis in the CSF. If cells are present, one should consider HIV infection, Lyme disease, sarcoidosis, or lymphomatous or leukemic infiltration of nerve roots. Some patients with GBS and CIDP have abnormal liver function tests. In these cases, it is important to also check for hepatitis B and C, HIV, CMV, and Epstein-Barr virus (EBV) infection. In patients with an axonal GBS (by EMG/NCS) or those with a suspicious coinciding history (e.g., unexplained abdominal pain, psychiatric illness, significant autonomic dysfunction), it is reasonable to screen for porphyria.

In patients with a severe sensory ataxia, a sensory ganglionopathy or neuronopathy should be considered. The most common causes of sensory ganglionopathies are Sjögren syndrome and a paraneoplastic neuropathy. Neuropathy can be the initial manifestation of Sjögren syndrome. Thus, one should always inquire about dry eyes and mouth in patients with sensory signs and symptoms. Further, some patients can manifest sicca complex without full-blown Sjögren syndrome. Thus, patients with sensory ataxia should have an senile systemic amyloidosis (SSA) and single strand binding (SSB) in addition to the routine ANA. To workup a possible paraneoplastic sensory ganglionopathy, anti-neuronal nuclear antibodies (e.g., anti-Hu antibodies) should be obtained (Chap. 101). These antibodies are most commonly seen in patients with small cell carcinoma of the lung but are seen also in breast, ovarian, lymphoma, and other cancers. Importantly, the paraneoplastic neuropathy can precede the detection of the cancer, and detection of these autoantibodies should lead to a search for malignancy.

Nerve Biopsies

Nerve biopsies are now rarely indicated for evaluation of neuropathies. The primary indication for nerve biopsy is suspicion for amyloid neuropathy or vasculitis. In most instances, the abnormalities present on biopsies do not help distinguish one form of peripheral neuropathy from another (beyond what is already apparent by clinical examination and the NCS). Nerve biopsies should only be done if the NCS studies are abnormal. The sural nerve is most commonly biopsied because it is a pure sensory nerve and biopsy will not result in loss of motor function. In suspected vasculitis, a combination biopsy of a superficial peroneal nerve (pure sensory) and the underlying peroneus brevis muscle obtained from a single small incision increases the diagnostic yield. Tissue can be analyzed by frozen section and paraffin section to assess the supporting structures for evidence of inflammation, vasculitis, or amyloid deposition. Semithin plastic sections, teased fiber preparations, and electron microscopy are used to assess the morphology of the nerve fibers and to distinguish axonopathies from myelinopathies.

Skin Biopsies

Skin biopsies are sometimes used to diagnose a small-fiber neuropathy. Following a punch biopsy of the skin in the distal lower extremity, immunologic staining can be used to measure the density of small unmyelinated fibers. The density of these nerve fibers is reduced in patients with small-fiber neuropathies in whom nerve conduction studies and routine nerve biopsies are often normal. This technique may allow for an objective measurement in patients with mainly subjective symptoms. However, it adds little to what one already knows from the clinical examination and EDx.

Figure 384-1

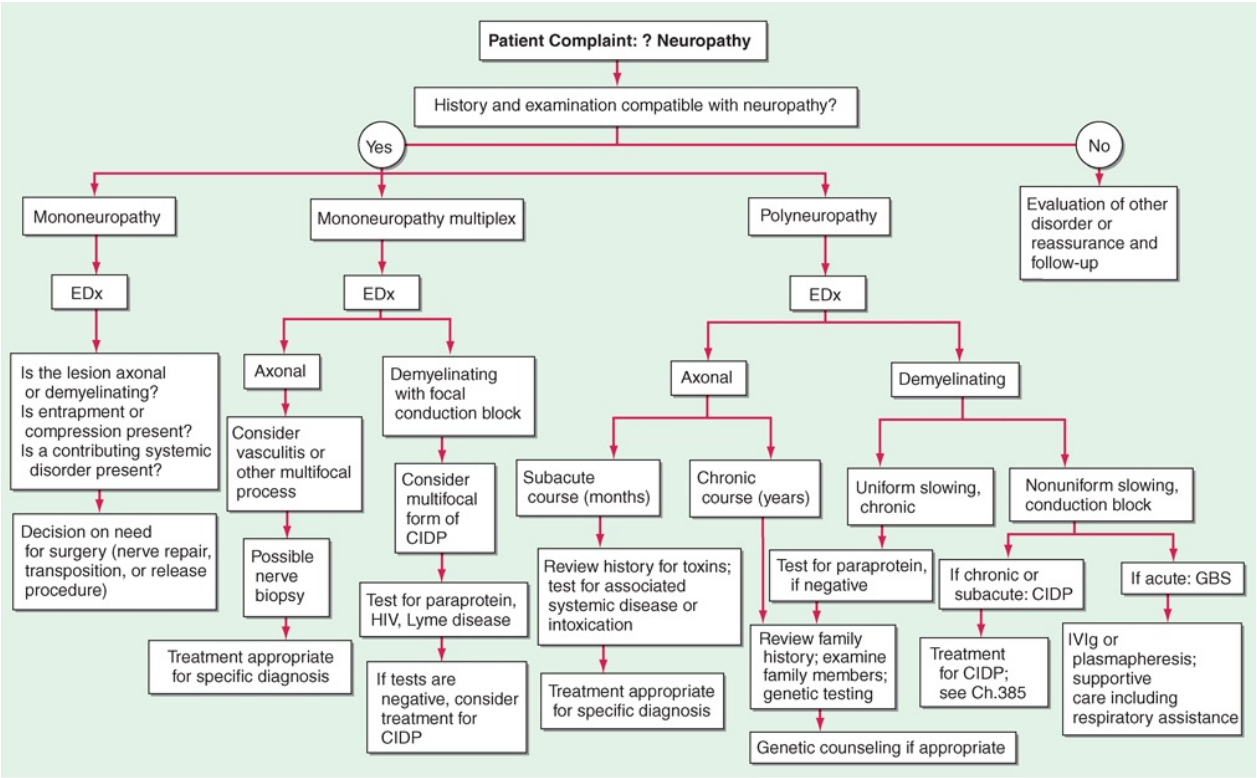

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Approach to the evaluation of peripheral neuropathies. CIDP, chronic inflammatory demyelinating polyradiculoneuropathy; GBS, Guillain-Barré syndrome.

SPECIFIC DISORDERS

Hereditary Neuropathies

Charcot-Marie-Tooth (CMT) disease is the most common type of hereditary neuropathy. Rather than one disease, CMT is a syndrome of several genetically distinct disorders **(Table 384-4)**. The various subtypes of CMT are classified according to the nerve conduction velocities and predominant pathology (e.g., demyelination or axonal degeneration), inheritance pattern (autosomal dominant, recessive, or X-linked), and the specific mutated genes. Type 1 CMT (or CMT1) refers to inherited demyelinating sensorimotor neuropathies, while the axonal sensory neuropathies are classified as CMT2. By definition, motor conduction velocities in the arms are slowed to less than 38 m/s in CMT1 and are greater than 38 m/s in CMT2. However, most cases of CMT1 actually have motor nerve conduction velocities (NCVs) between 20 and 25 m/s. CMT1 and CMT2 usually begin in childhood or early adult life; however, onset later in life can occur, particularly in CMT2. Both are associated with autosomal dominant inheritance, with a few exceptions. CMT3 is an autosomal dominant neuropathy that appears in infancy and is associated with severe demyelination or hypomyelination. CMT4 is an autosomal recessive neuropathy that typically begins in childhood or early adult life. There are no medical therapies for any of the CMTs, but physical and occupational therapy can be beneficial as can bracing (e.g., ankle-foot orthotics for footdrop) and other orthotic devices.

Table 384-4 Classification of Charcot-Marie-Tooth Disease and Related Neuropathies

Name	Inheritance	Gene Location	Gene Product
CMT1			
CMT1A	AD	17p11.2	PMP-22 (usually duplication of gene)
CMT1B	AD	1q21-23	MPZ
CMT1C	AD	16p13.1-p12.3	LITAF
CMT1D	AD	10q21.1-22.1	ERG2
CMT1E (with deafness)	AD	17p11.2	Point mutations in PMP 22 gene
CMT1F	AD	8p13-21	Neurofilament light chain
CMT1X	X-linked dominant	Xq13	Connexin-32
HNPP	AD	17p11.21q21-23	PMP-22 MPZ
CMT2			
CMT2A2 (allelic to HMSN VI with optic atrophy)	AD	1p36.2	MFN2
CMT2B	AD	3q13-q22	RAB7
CMT2B1 (allelic to LGMD 1B)	AR	1q21.2	Lamin A/C
CMT2B2	AR and AD	19q13	MED25 for ARUnknown for AD
CMT2C (with vocal cord and diaphragm paralysis)	AD	12q23-24	TRPV4
CMT2D (allelic to distal SMA5)	AD	7p14	Glycine tRNA synthetase

CMT2E (allelic to CMT 1F)	AD	8p21	Neurofilament light chain
CMT2F	AD	7q11-q21	Heat-shock 27-kDa protein-1
CMT2G (may be allelic to CMT4H)	AD	12q12-q13	? (Fabrin)
CMT2H	AD	8q21.3	? (may be GDAP1)
CMT2I (allelic to CMT1B)	AD	1q22	MPZ
CMT2J	AD	1q22	MPZ
CMT2K (allelic to CMT4A)	AD	8q13-q21	GDAP1
CMT2L (allelic to distal hereditary motor neuropathy type 2)	AD	12q24	Heat-shock protein 8
CMT2M	AD	16q22	AARS
CMT2X	X-linked	Xq22-24	PRPS1
CMT3 (Dejerine-Sottas disease, congenital hypomyelinating neuropathy)	AD AD AR AR	17p11.2 1q21-23 10q21.1-22.1 19q13	PMP-22 MPZ ERG2 Periaxon
CMT4			
CMT4A	AR	8q13-21.1	GDAP1
CMT4B1 CMT4B2	AR AR	11q23 11p15	MTMR2 MTMR13
CMT4C	AR	5q23-33	SH3TC2
CMT4D (HMSN-Lom)	AR	8q24	NDRG1
CMT4E (Congenital hypomyelinating neuropathy)	AR	-	Probably includes PMP22, MPZ, and ERG-2
CMT4F	AR	19q13.1-13.3	Periaxin
CMT4G	AR	10q23.2	HKI
CMT4H	AR	12q12-q13	Frabin
CMT4J	AR	6q21	FIG4
HNA	AD	17q24	SEPT9
HMSN-P	AD	3q13-q14	?
HSAN1	AD; Rare AR and X-linked cases also reported	9q22	SPTLC1
HSAN2	AR	12p13.33	PRKWNK1
HSAN3	AR	9q21	IKAP
HSAN4	AR	3q	trkA/NGF receptor
HSAN5	AD or AR	1p11.2-p13.2	NGFb

Abbreviations: AARS, alanyl-tRNA synthetase; AD, autosomal dominant; AR, autosomal recessive; CMT, Charcot-Marie-Tooth; ERG2, early growth response-2 protein; FIG4, FDG1-related F actin-binding protein; GDAP1, ganglioside-induced differentiation-associated protein-1; HK1, hexokinase 1; HMSN-P, hereditary motor and sensory neuropathy-proximal; HNA, hereditary neuralgic amyotrophy; HNPP, hereditary neuropathy with liability to pressure palsies; HSAN; hereditary sensory and autonomic neuropathy; IKAP, kB kinase complex-associated protein; LGMD, limb girdle muscular dystrophy; LITAF, lipopolysaccharide-induced tumor necrosis factor α factor; MED25, mediator 25; MFN2, mitochondrial fusion protein mitofusin 2 gene; MPZ, myelin protein zero protein; MTMR2, myotubularin-related protein-2; NDRG1, N-myc downstream regulated 1; PMP-22, peripheral myelin protein-22; PRKWNK1, protein kinase, lysine deficient 1; PRPS1, phosphoribosylpyrophosphate synthetase 1; RAB7, Ras-related protein 7; SEPT9, Septin 9; SH3TC2, SH3 domain and tetratricopeptide repeats 2; SMA, spinal muscular atrophy; SPTLC1, serine palmitoyltransferase long-chain base 1; TrkA/NGF, tyrosine kinase A/nerve growth factor; tRNA, transfer ribonucleic acid; TRPV4, transient receptor potential cation channel, subfamily V, member 4. ***Source:*** Modified from Amato and Russell.

CMT1

CMT1 is the most common form of hereditary neuropathy, with the ratio of CMT1:CMT2 being approximately 2:1. Affected individuals usually present in the first to third decade of life with distal leg weakness (e.g., footdrop), although patients may remain asymptomatic even late in life. People with CMT generally do not complain of numbness or tingling, which can be helpful in distinguishing CMT from acquired forms of neuropathy in which sensory symptoms usually predominate. Although usually asymptomatic in this regard, reduced sensation to all modalities is apparent on examination. Muscle stretch reflexes are unobtainable or reduced throughout. There is often atrophy of the muscles below the knee (particularly the anterior compartment), leading to so-called inverted champagne bottle legs.

Motor NCVs are usually in the 20–25 m/s range. Nerve biopsies usually are not performed on patients suspected of having CMT1, as the diagnosis usually can be made by less invasive testing (e.g., NCS and genetic studies). However, when done, the biopsies reveal reduction of myelinated nerve fibers with a predilection for the loss of the large-diameter fibers and Schwann cell proliferation around thinly or demyelinated fibers, forming so-called onion bulbs.

CMT1A is the most common subtype of CMT1, representing 70% of cases, and is caused by a 1.5-megabase (Mb) duplication within chromosome 17p11.2-12 wherein the gene for peripheral myelin protein-22 (PMP-22) lies. This results in patients having three copies of the PMP-22 gene rather than two. This protein accounts for 2–5% of myelin protein and is expressed in compact portions of the peripheral myelin sheath. Approximately 20% of patients with CMT1 have CMT1B, which is caused by mutations in the myelin protein zero (MPZ). CMT1B is for the most part clinically, electrophysiologically, and histologically indistinguishable from CMT1A. MPZ is an integral myelin protein and accounts for more than half of the myelin protein in peripheral nerves. Other forms of CMT1 are much less common and again indistinguishable from one another clinically and electrophysiologically.

CMT2

CMT2 tends to present later in life compared to CMT1. Affected individuals usually become symptomatic in the second decade of life; some cases present earlier in childhood, while others remain asymptomatic into late adult life. Clinically, CMT2 is for the most part indistinguishable from CMT1. NCS are helpful in this regard; in contrast to CMT1, the velocities are normal or only slightly slowed. The most common cause of CMT2 is a mutation in the gene for mitofusin 2 (MFN2), which accounts for one-third of CMT2 cases overall. MFN2 localizes to the outer mitochondrial membrane, where it regulates the mitochondrial network architecture by fusion of mitochondria. The other genes associated with CMT2 are much less common.

CMT3

CMT3 was originally described by Dejerine and Sottas as a hereditary demyelinating sensorimotor polyneuropathy presenting in

infancy or early childhood. Affected children are severely weak. Motor NCVs are markedly slowed, typically 5–10 m/s or less. Most cases of CMT3 are caused by point mutations in the genes for PMP-22, MPZ, or ERG-2, which are also the genes responsible for CMT1.

CMT4

CMT4 is extremely rare and is characterized by a severe, childhood-onset sensorimotor polyneuropathy that is usually inherited in an autosomal recessive fashion. Electrophysiologic and histologic evaluations can show demyelinating or axonal features. CMT4 is genetically heterogenic (Table 384-4).

CMT1X

CMT1X is an X-linked dominant disorder with clinical features similar to CMT1 and -2, except that the neuropathy is much more severe in men than in women. CMT1X accounts for approximately 10–15% of CMT overall. Men usually present in the first two decades of life with atrophy and weakness of the distal arms and legs, areflexia, pes cavus, and hammertoes. Obligate women carriers are frequently asymptomatic, but can develop signs and symptoms. Onset in women is usually after the second decade of life, and the neuropathy is milder in severity.

NCS reveal features of both demyelination and axonal degeneration that are more severe in men compared to women. In men, motor NCVs in the arms and legs are moderately slowed (in the low to mid 30-m/s range). About 50% of men with CMT1X have motor NCVs between 15 and 35 m/s with about 80% of these falling between 25 and 35 m/s (intermediate slowing). In contrast, about 80% of women with CMT1X have NCV in the normal range and 20% had MNCV in the intermediate range. CMT1X is caused by mutations in the connexin 32 gene. Connexins are gap junction structural proteins that are important in cell-to-cell communication.

HEREDITARY NEUROPATHY WITH LIABILITY TO PRESSURE PALSIES (HNPP)

HNPP is an autosomal dominant disorder related to CMT1A. While CMT1A is usually associated with a 1.5-Mb duplication in chromosome 17p11.2 that results in an extra copy of PMP-22 gene, HNPP is caused by inheritance of the chromosome with the corresponding 1.5-Mb deletion of this segment, and thus affected individuals have only one copy of the PMP-22 gene. Patients usually manifest in the second or third decade of life with painless numbness and weakness in the distribution of single peripheral nerves, although multiple mononeuropathies can occur. Symptomatic mononeuropathy or multiple mononeuropathies are often precipitated by trivial compression of nerve(s) as can occur with wearing a backpack, leaning on the elbows, or crossing one's legs for even a short period of time. These pressure-related mononeuropathies may take weeks or months to resolve. In addition, some affected individuals manifest with a progressive or relapsing, generalized and symmetric, sensorimotor peripheral neuropathy that resembles CMT.

HEREDITARY NEURALGIC AMYOTROPHY (HNA)

HNA is an autosomal dominant disorder characterized by recurrent attacks of pain, weakness, and sensory loss in the distribution of the brachial plexus often beginning in childhood. These attacks are similar to those seen with idiopathic brachial plexitis (see below). Attacks may occur in the postpartum period, following surgery, or at other times of stress. Most patients recover over several weeks or months. Slightly dysmorphic features, including hypotelorism, epicanthal folds, cleft palate, syndactyly, micrognathia, and facial asymmetry, are evident in some individuals. EDx demonstrate an axonal process. HNA is caused by mutations in septin 9 (*SEPT9*). Septins may be important in formation of the neuronal cytoskeleton and have a role in cell division, but the mechanism of causing HNA is unclear.

HEREDITARY SENSORY AND AUTONOMIC NEUROPATHY (HSAN)

The HSANs are a very rare group of hereditary neuropathies in which sensory and autonomic dysfunction predominates over muscle weakness, unlike CMT, in which motor findings are most prominent (Table 384-4). Nevertheless, affected individuals can develop motor weakness and there can be overlap with CMT. There are no medical therapies available to treat these neuropathies, other than prevention and treatment of mutilating skin and bone lesions.

Of the HSANs, only HSAN1 typically presents in adults. The HSAN1 is the most common of the HSANs and is inherited in an autosomal dominant fashion. Affected individuals with HSAN1 usually manifest in the second through fourth decade of life. HSAN1 is associated with the degeneration of small myelinated and unmyelinated nerve fibers leading to severe loss of pain and temperature sensation, deep dermal ulcerations, recurrent osteomyelitis, Charcot joints, bone loss, gross foot and hand deformities, and amputated digits. Although most people with HSAN1 do not complain of numbness, they often describe burning, aching, or lancinating pains. Autonomic neuropathy is not a prominent feature, but bladder dysfunction and reduced sweating in the feet may occur. HSAN1A is caused by mutations in the serine palmitoyltransferase long-chain base 1 (*SPTLC1*) gene.

OTHER HEREDITARY NEUROPATHIES

(Table 384-5)

Table 384-5 Rare Hereditary Neuropathies

Hereditary Disorders of Lipid Metabolism
Metachromatic leukodystrophy
Krabbe disease (globoid cell leukodystrophy)
Fabry disease
Adrenoleukodystrophy/adrenomyeloneuropathy
Refsum disease
Tangier disease
Cerebrotendinous xanthomatosis
Hereditary Ataxias With Neuropathy
Friedreich ataxia
Vitamin E deficiency
Spinocerebellar ataxia
Abetalipoproteinemia (Bassen-Kornzweig disease)
Disorders of Defective DNA Repair
Ataxia-telangiectasia
Cockayne syndrome
Giant Axonal Neuropathy
Porphyria

Acute intermittent porphyria (AIP)
Hereditary coproporphyria (HCP)
Variegate porphyria (VP)
Familial Amyloid Polyneuropathy (FAP)
Transthyretin-related
Gelsolin-related
Apolipoprotein A1-related

Fabry Disease

Fabry disease (angiokeratoma corporis diffusum) is an X-linked dominant disorder. While men are more commonly and severely affected, women can also show severe signs of the disease. Angiokeratomas are reddish-purple maculopapular lesions that are usually found around the umbilicus, scrotum, inguinal region, and perineum. Burning or lancinating pain in the hands and feet often develops in males in late childhood or early adult life. However, the neuropathy is usually overshadowed by complications arising from the associated premature atherosclerosis (e.g., hypertension, renal failure, cardiac disease, and stroke) that often lead to death by the fifth decade of life. Some patients also manifest primarily with a dilated cardiomyopathy.

Fabry disease is caused by mutations in the α-galactosidase gene that leads to the accumulation of ceramide trihexoside in nerves and blood vessels. A decrease in α-galactosidase activity is evident in leukocytes and cultured fibroblasts. Glycolipid granules may be appreciated in ganglion cells of the peripheral and sympathetic nervous systems and in perineurial cells. Enzyme replacement therapy with α-galactosidase B can improve the neuropathy if patients are treated early, before irreversible nerve fiber loss.

Adrenoleukodystrophy/Adrenomyeloneuropathy

Adrenoleukodystrophy (ALD) and adrenomyeloneuropathy (AMN) are allelic X-linked dominant disorders caused by mutations in the peroxisomal transmembrane adenosine triphosphate-binding cassette (ABC) transporter gene. Patients with ALD manifest with CNS abnormalities. However, 30% with mutations in this gene present with the AMN phenotype that typically manifests in the third to fifth decade of life with mild to moderate peripheral neuropathy combined with progressive spastic paraplegia (Chap. 377). Rare patients present with an adult-onset spinocerebellar ataxia or only with adrenal insufficiency.

EDx is suggestive of a primary axonopathy with secondary demyelination. Nerve biopsies demonstrate a loss of myelinated and unmyelinated nerve fibers with lamellar inclusions in the cytoplasm of Schwann cells. Very long chain fatty acid (VLCFA) levels (C24, C25, and C26) are increased in the urine. Laboratory evidence of adrenal insufficiency is evident in approximately two-thirds of patients. The diagnosis can be confirmed by genetic testing.

Adrenal insufficiency is managed by replacement therapy; however, there is no proven effective therapy for the neurologic manifestations of ALD/AMN. Diets low in VLCFAs and supplemented with Lorenzo's oil (erucic and oleic acids) reduce the levels of VLCFAs and increase the levels of C22 in serum, fibroblasts, and liver; however, several large, open-label trials of Lorenzo's oil failed to demonstrate efficacy.

Refsum Disease

Refsum disease can manifest in infancy to early adulthood with the classic tetrad of (1) peripheral neuropathy, (2) retinitis pigmentosa, (3) cerebellar ataxia, and (4) elevated CSF protein concentration. Most affected individuals develop progressive distal sensory loss and weakness in the legs leading to footdrop by their 20s. Subsequently, the proximal leg and arm muscles may become weak. Patients may also develop sensorineural hearing loss, cardiac conduction abnormalities, ichthyosis, and anosmia.

Serum phytanic acid levels are elevated. Sensory and motor NCS reveal reduced amplitudes, prolonged latencies, and slowed conduction velocities. Nerve biopsy demonstrates a loss of myelinated nerve fibers, with remaining axons often thinly myelinated and associated with onion bulb formation.

Refsum disease is genetically heterogeneous but autosomal recessive in nature. Classical Refsum disease with childhood or early adult onset is caused by mutations in the gene that encodes for phytanoyl-CoAα-hydroxylase (*PAHX*). Less commonly, mutations in the gene encoding peroxin 7 receptor protein (PRX 7) are responsible. These mutations lead to the accumulation of phytanic acid in the central and peripheral nervous systems. Refsum disease is treated by removing phytanic precursors (phytols: fish oils, dairy products, and ruminant fats) from the diet.

Tangier Disease

Tangier disease is a rare autosomal recessive disorder that can present as (1) asymmetric multiple mononeuropathies, (2) a slowly progressive symmetric polyneuropathy predominantly in the legs, or (3) a pseudo-syringomyelia pattern with dissociated sensory loss (i.e., abnormal pain/temperature perception but preserved position/vibration in the arms [Chap. 377]). The tonsils may appear swollen and yellowish-orange in color, while there may also be splenomegaly and lymphadenopathy.

Tangier disease is caused by mutations in the ATP-binding cassette transporter 1 (ABC1) gene, which leads to markedly reduced levels of high-density lipoprotein (HDL) cholesterol levels while triacylglycerol levels are increased. Nerve biopsies reveal axonal degeneration with demyelination and remyelination. Electron microscopy demonstrates abnormal accumulation of lipid in Schwann cells, particularly those encompassing umyelinated and small myelinated nerves. There is no specific treatment.

Porphyria

Porphyria is a group of inherited disorders caused by defects in heme biosynthesis (Chap. 358). Three forms of porphyria are associated with peripheral neuropathy: acute intermittent porphyria (AIP), hereditary coproporphyria (HCP), and variegate porphyria (VP). The acute neurologic manifestations are similar in each, with the exception that a photosensitive rash is seen with HCP and VP but not in AIP. Attacks of porphyria can be precipitated by certain drugs (usually those metabolized by the P450 system), hormonal changes (e.g., pregnancy, menstrual cycle), and dietary restrictions.

An acute attack of porphyria may begin with sharp abdominal pain. Subsequently, patients may develop agitation, hallucinations, or seizures. Several days later, back and extremity pain followed by weakness ensues, mimicking GBS. Weakness can involve the arms or the legs and can be asymmetric, proximal, or distal in distribution, as well as affecting the face and bulbar musculature. Dysautonomia and signs of sympathetic overactivity are common (e.g., pupillary dilation, tachycardia, and hypertension). Constipation, urinary retention, and incontinence can also be seen.

The CSF protein is typically normal or mildly elevated. Liver function tests and hematologicparameters are usually normal. Some patients are hyponatremic due to inappropriate secretion of antidiuretic hormone (Chap. 339). The urine may appear brownish in color secondary to the high concentration of porphyrin metabolites. Accumulation of intermediary precursors of heme (i.e., δ-aminolevulinic acid, porphobilinogen, uroporphobilinogen, coproporphyrinogen, and protoporphyrinogen) are found in urine. Specific enzyme activities can also be measured in erythrocytes and leukocytes. The primary abnormalities on EDx are marked

reductions in CMAP amplitudes and signs of active axonal degeneration on needle EMG.

The porphyrias are inherited in an autosomal dominant fashion. AIP is associated with porphobilinogen deaminase deficiency, HCP is caused by defects in coproporphyrin oxidase, and VP is associated with protoporphyrinogen oxidase deficiency. The pathogenesis of the neuropathy is not completely understood. Treatment with glucose and hematin may reduce the accumulation of heme precursors. Intravenous glucose is started at a rate of 10–20 g/h. If there is no improvement within 24 h, intravenous hematin 2–5 mg/kg per day for 3–14 days should be given.

Familial Amyloid Polyneuropathy

Familial amyloid polyneuropathy (FAP) is phenotypically and genetically heterogeneous and is caused by mutations in the genes for transthyretin (TTR), apolipoprotein A1, or gelsolin (Chap. 112). The majority of patients with FAP have mutations in the TTR gene. Amyloid deposition may be evident in abdominal fat pad, rectal, or nerve biopsies. The clinical features, histopathology, and EDx reveal abnormalities consistent with a generalized or multifocal, predominantly axonal but occasionally demyelinating, sensorimotor polyneuropathy.

Patients with TTR-related FAP usually develop insidious onset of numbness and painful paresthesias in the distal lower limbs in the third to fourth decade of life, although some patients develop the disorder later in life. Carpal tunnel syndrome (CTS) is common. Autonomic involvement can be severe, leading to postural hypotension, constipation or persistent diarrhea, erectile dysfunction, and impaired sweating. Amyloid deposition also occurs in the heart, kidneys, liver, and the corneas. Patients usually die 10–15 years after the onset of symptoms from cardiac failure or complications from malnutrition. Because the liver produces much of the body's TTR, liver transplantation has been used to treat FAP related to TTR mutations. Serum TTR levels decrease after transplantation, and improvement in clinical and EDx features have been reported.

Patients with apolipoprotein A1–related FAP (Van Allen type) usually present in the fourth decade with numbness and painful dysesthesias in the distal limbs. Gradually, the symptoms progress, leading to proximal and distal weakness and atrophy. Although autonomic neuropathy is not severe, some patients develop diarrhea, constipation, or gastroparesis. Most patients die from systemic complications of amyloidosis (e.g., renal failure) 12–15 years after the onset of the neuropathy.

Gelsolin-related amyloidosis (Finnish type) is characterized by the combination of lattice corneal dystrophy and multiple cranial neuropathies that usually begin in the third decade of life. Over time, a mild generalized sensorimotor polyneuropathy develops. Autonomic dysfunction does not occur.

ACQUIRED NEUROPATHIES

Primary or Al Amyloidosis (see Chap. 112)

Besides FAP, amyloidosis can also be acquired. In primary or AL amyloidosis, the abnormal protein deposition is composed of immunoglobulin light chains. AL amyloidosis occurs in the setting of multiple myeloma, Waldenström macroglobulinemia, lymphoma, other plasmacytomas, or lymphoproliferative disorders, or without any other identifiable disease.

Approximately 30% of patients with AL primary amyloidosis present with a polyneuropathy, most typically painful dysesthesias and burning sensations in the feet. However, the trunk can be involved and some manifest with a mononeuropathy multiplex pattern. CTS occurs in 25% of patients and may be the initial manifestation. The neuropathy is slowly progressive, and eventually weakness develops along with large-fiber sensory loss. Most patients develop autonomic involvement with postural hypertension, syncope, bowel and bladder incontinence, constipation, impotence, and impaired sweating. Patients generally die from their systemic illness (renal failure, cardiac disease).

The monoclonal protein may be composed of IgG, IgA, IgM, or only free light chain. Lambda (λ) is more common than κ light chain (>2:1) in AL amyloidosis. The CSF protein is often increased (with normal cell count), and thus the neuropathy may be mistaken for CIDP (Chap. 385). Nerve biopsies reveal axonal degeneration and amyloid deposition in either a globular or diffuse pattern infiltrating the perineurial, epineurial, and endoneurial connected tissue and in blood vessel walls.

The median survival of patients with primary amyloidosis is less than 2 years, with death usually from progressive congestive heart failure or renal failure. Chemotherapy with melphalan, prednisone, and colchicine, to reduce the concentration of monoclonal proteins, and autologous stem cell transplantation may prolong survival, but whether the neuropathy improves is controversial (Chap. 112).

Diabetic Neuropathy

Diabetes mellitus (DM) is the most common cause of peripheral neuropathy in developed countries. DM is associated with several types of polyneuropathy: distal symmetric sensory or sensorimotor polyneuropathy, autonomic neuropathy, diabetic neuropathic cachexia, polyradiculoneuropathies, cranial neuropathies, and other mononeuropathies. Risk factors for the development of neuropathy include long-standing, poorly controlled DM and the presence of retinopathy and nephropathy.

DIABETIC DISTAL SYMMETRIC SENSORY AND SENSORIMOTOR POLYNEUROPATHY (DSPN)

DSPN is the most common form of diabetic neuropathy and manifests as sensory loss beginning in the toes that gradually progresses over time up the legs and into the fingers and arms. When severe, a patient may develop sensory loss in the trunk (chest and abdomen), initially in the midline anteriorly and later extending laterally. Tingling, burning, deep aching pains may also be apparent. NCS usually show reduced amplitudes and mild to moderate slowing of conduction velocities (CVs). Nerve biopsy reveals axonal degeneration, endothelial hyperplasia, and, occasionally, perivascular inflammation. Tight control of glucose can reduce the risk of developing neuropathy or improve the underlying neuropathy. A variety of medications have been used with variable success to treat painful symptoms associated with DSPN, including antiepileptic medications, antidepressants, sodium channel blockers, and other analgesics **(Table 384-6).**

Table 384-6 Treatment of Painful Sensory Neuropathies

Therapy	Route	Dose	Side Effects
First-Line			
Lidoderm 5% patch	Apply to painful area	Up to 3 patches qd	Skin irritation
Tricyclic antidepressants (e.g., amitriptylin, nortriptyline)	p.o.	10–100 mg qhs	Cognitive changes, sedation, dry eyes and mouth, urinary retention, constipation
Gabapentin	p.o.	300 –1200 mg TID	Cognitive changes, sedation, peripheral edema
Pregabalin	p.o.	50–100	Cognitive changes, sedation, peripheral edema

		mg TID	
Duloxetine	p.o.	30–60 mg qd	Cognitive changes, sedation, dry eyes, diaphoresis, nausea, diarrhea, constipation
Second-Line			
Carbamazepine	p.o.	200–400 mg q 6–8 h	Cognitive changes, dizziness, leukopenia, liver dysfunction
Phenytoin	p.o.	200–400 mg qhs	Cognitive changes, dizziness, liver dysfunction
Venlafaxine	po	37.5–150 mg/d	Asthenia, sweating, nausea, constipation, anorexia, vomiting, somnolence, dry mouth, dizziness, nervousness, anxiety, tremor, and blurred vision as well as abnormal ejaculation/orgasm and impotence
Tramadol	p.o.	50 mg qid	Cognitive changes, GI upset
Third-Line			
Mexiletine	p.o.	200–300 mg tid	Arrhythmias
Other Agents			
EMLA cream 2.5% lidocaine 2.5% prilocaine	Apply cutaneously	q.i.d.	Local erythema
Capsaicin 0.025%–0.075% cream	Apply cutaneously	q.i.d.	Painful burning skin

Source: Modified from Amato and Russell.

DIABETIC AUTONOMIC NEUROPATHY

Autonomic neuropathy is typically seen in combination with DSPN. The autonomic neuropathy can manifest as abnormal sweating, dysfunctional thermoregulation, dry eyes and mouth, pupillary abnormalities, cardiac arrhythmias, postural hypotension, gastrointestinal abnormalities (e.g., gastroparesis, postprandial bloating, chronic diarrhea or constipation), and genitourinary dysfunction (e.g., impotence, retrograde ejaculation, incontinence). Tests of autonomic function are generally abnormal, including sympathetic skin responses and quantitative sudomotor axon reflex testing. Sensory and motor NCS generally demonstrate features described above with DSPN.

DIABETIC RADICULOPLEXUS NEUROPATHY (DIABETIC AMYOTROPHY OR BRUNS-GARLAND SYNDROME)

Diabetic radiculoplexus neuropathy is the presenting manifestation of DM in approximately one-third of patients. Typically, patients present with severe pain in the low back, hip, and thigh in one leg. Rarely, the diabetic polyradiculoneuropathy begins in both legs at the same time. Atrophy and weakness of proximal and distal muscles in the affected leg become apparent within a few days or weeks. The neuropathy is often accompanied or heralded by severe weight loss. Weakness usually progresses over several weeks or months, but can continue to progress for 18 months or more. Subsequently, there is slow recovery but many are left with residual weakness, sensory loss, and pain. In contrast to the more typical lumbosacral radiculoplexus neuropathy, some patients develop thoracic radiculopathy or, even less commonly, a cervical polyradiculoneuropathy. CSF protein is usually elevated, while the cell count is normal. ESR is often increased. EDx reveals evidence of active denervation in affected proximal and distal muscles in the affected limbs and in paraspinal muscles. Nerve biopsies may demonstrate axonal degeneration along with perivascular inflammation. Patients with severe pain are sometimes treated in the acute period with glucocorticoids, although a randomized controlled trial has yet to be performed, and the natural history of this neuropathy is gradual improvement.

DIABETIC MONONEUROPATHIES OR MULTIPLE MONONEUROPATHIES

The most common mononeuropathies are median neuropathy at the wrist and ulnar neuropathy at the elbow, but peroneal neuropathy at the fibular head, and sciatic, lateral femoral, cutaneous, or cranial neuropathies also occur. In regard to cranial mononeuropathies, a seventh nerve palsy is most common, followed by third nerve, sixth nerve, and, less frequently, fourth nerve palsies. Diabetic third nerve palsies are characteristically pupil-sparing (Chap. 28).

Hypothyroidism

Hypothyroidism is more commonly associated with a proximal myopathy, but some patients develop a neuropathy, most typically carpal tunnel syndrome. Rarely, a generalized sensory polyneuropathy characterized by painful paresthesias and numbness in both the legs and hands can occur. Treatment is correction of the hypothyroidism.

Sjögren Syndrome

Sjögren syndrome, characterized by the sicca complex of xerophthalmia, xerostomia, and dryness of other mucous membranes, can be complicated by neuropathy. Most common is a length-dependent axonal sensorimotor neuropathy characterized mainly by sensory loss in the distal extremities. A pure small-fiber neuropathy or a cranial neuropathy, particularly involving the trigeminal nerve, can also be seen. Sjögren syndrome is also associated with sensory neuronopathy/ganglionopathy. Patients with sensory ganglionopathies develop progressive numbness and tingling of the limbs, trunk, and face in a non-length-dependent manner such that symptoms can involve the face or arms more than the legs. The onset can be acute or insidious. Sensory examination demonstrates severe vibratory and proprioceptive loss leading to sensory ataxia.

Patients with neuropathy due to Sjögren syndrome may have antinuclear antibodies (ANA), SS-A/Ro, and SS-B/La antibodies in the serum but most do not. NCS demonstrate reduced amplitudes of sensory studies in the affected limbs. Nerve biopsy demonstrates axonal degeneration. Nonspecific perivascular inflammation may be present, but only rarely is there necrotizing vasculitis. There is no specific treatment for neuropathies related to Sjögren syndrome. When vasculitis is suspected, immunosuppressive agents may be beneficial. Occasionally, the sensory neuronopathy/ganglionopathy stabilizes or improves with immunotherapy, such as IVIg.

Rheumatoid Arthritis

Peripheral neuropathy occurs in at least 50% of patients with rheumatoid arthritis (RA) and may be vasculitic in nature (Chap. 326). Vasculitic neuropathy can present with a mononeuropathy multiplex, a generalized symmetric pattern of involvement, or a combination of these patterns. Neuropathies may also be due to drugs used to treat the RA (e.g., tumor necrosis blockers, leflunomide). Nerve biopsy often reveals thickening of the epineurial and endoneurial blood vessels as well as perivascular

inflammation or vasculitis, with transmural inflammatory cell infiltration and fibrinoid necrosis of vessel walls. The neuropathy often is responsive to immunomodulating therapies.

Systemic Lupus Erythematosus (SLE)

Between 2 and 27% of individuals with SLE develop a peripheral neuropathy. Affected patients typically present with a slowly progressive sensory loss beginning in the feet. Some patients develop burning pain and paresthesias with normal reflexes, and nerve conduction studies suggest a pure small-fiber neuropathy. Less common are multiple mononeuropathies presumably secondary to necrotizing vasculitis. Rarely, a generalized sensorimotor polyneuropathy meeting clinical, laboratory, electrophysiologic, and histologic criteria for either GBS or CIDP may occur. Immunosuppressive therapy is beneficial in SLE patients with neuropathy due to vasculitis. Immunosuppressive agents are less likely to be effective in patients with a generalized sensory or sensorimotor polyneuropathy without evidence of vasculitis. Patients with a GBS or CIDP-like neuropathy should be treated accordingly (Chap. 385).

Systemic Sclerosis (Scleroderma)

A distal symmetric, mainly sensory, polyneuropathy complicates 5–67% of scleroderma cases. Cranial mononeuropathies can also develop, most commonly of the trigeminal nerve, producing numbness and dysesthesias in the face. Multiple mononeuropathies also occur. The EDx and histologic features of nerve biopsy are those of an axonal sensory greater than motor polyneuropathy.

Mixed Connective Tissue Disease (Mctd)

A mild distal axonal sensorimotor polyneuropathy occurs in approximately 10% of patients with MCTD.

Sarcoidosis

The peripheral or central nervous systems are involved in about 5% of patients with sarcoidosis. The most common cranial nerve involved is the seventh nerve, which can be affected bilaterally. Some patients develop radiculopathy or polyradiculopathy. With a generalized root involvement, the clinical presentation can mimic GBS or CIDP. Patients can also present with multiple mononeuropathies or a generalized, slowly progressive, sensory greater than motor polyneuropathy. Some have features of a pure small-fiber neuropathy. EDx reveals an axonal neuropathy. Nerve biopsy can reveal noncaseating granulomas infiltrating the endoneurium, perineurium, and epineurium along with lymphocytic necrotizing angiitis. Neurosarcoidosis may respond to treatment with glucocorticoids or other immunosuppressive agents.

Hypereosinophilic Syndrome

Hypereosinophilic syndrome is characterized by eosinophilia associated with various skin, cardiac, hematologic, and neurologic abnormalities. A generalized peripheral neuropathy or a mononeuropathy multiplex occurs in 6–14% of patients.

Celiac Disease (Gluten-Induced Enteropathy or Non-Tropical Sprue)

Neurologic complications, particularly ataxia and peripheral neuropathy, are estimated to occur in 10% of patients with celiac disease. A generalized sensorimotor polyneuropathy, pure motor neuropathy, multiple mononeuropathies, autonomic neuropathy, small-fiber neuropathy, and neuromyotonia have all been reported in association with celiac disease or antigliadin/antiendomysial antibodies. Nerve biopsy may reveal a loss of large myelinated fibers. The neuropathy may be secondary to malabsorption of vitamins B_{12} and E. However, some patients have no appreciable vitamin deficiencies. The pathogenic basis for the neuropathy in these patients is unclear but may be autoimmune in etiology. The neuropathy does not appear to respond to a gluten-free diet. In patients with vitamin B12 or vitamin E deficiency, replacement therapy may improve or stabilize the neuropathy.

Inflammatory Bowel Disease

Ulcerative colitis and Crohn's disease may be complicated by GBS, CIDP, generalized axonal sensory or sensorimotor polyneuropathy, small-fiber neuropathy, or mononeuropathy. These neuropathies may be autoimmune, nutritional (e.g., vitamin B12 deficiency), treatment related (e.g., metronidazole), or idiopathic in nature. An acute neuropathy with demyelination resembling GBS may occur, particularly in patients treated with tumor necrosis factor α blockers.

Uremic Neuropathy

Approximately 60% of patients with renal failure develop a polyneuropathy characterized by length-dependent numbness, tingling, allodynia, and mild distal weakness. Rarely, a rapidly progressive weakness and sensory loss very similar to GBS can occur that improves with an increase in the intensity of renal dialysis or with transplantation. Mononeuropathies can also occur, the most common of which is carpal tunnel syndrome. Ischemic monomelic neuropathy (see below) can complicate arteriovenous shunts created in the arm for dialysis. EDx in uremic patients reveals features of a length-dependent, primarily axonal, sensorimotor polyneuropathy. Sural nerve biopsies demonstrate a loss of nerve fibers (particularly large myelinated nerve fibers), active axonal degeneration, and segmental and paranodal demyelination. The sensorimotor polyneuropathy can be stabilized by hemodialysis and improved with successful renal transplantation.

Chronic Liver Disease

A generalized sensorimotor neuropathy characterized by numbness, tingling, and minor weakness in the distal aspects of primarily the lower limbs commonly occurs in patients with chronic liver failure. EDx studies are consistent with a sensory greater than motor axonopathy. Sural nerve biopsy reveals both segmental demyelination and axonal loss. It is not known if hepatic failure in isolation can cause peripheral neuropathy, as the majority of patients have liver disease secondary to other disorders, such as alcoholism or viral hepatitis, which can also cause neuropathy.

Critical Illness Polyneuropathy

The most common causes of acute generalized weakness leading to admission to a medical intensive care unit (ICU) are GBS and myasthenia gravis (Chap. 386). However, weakness developing in critically ill patients while in the ICU is usually caused by critical illness polyneuropathy (CIP) or critical illness myopathy (CIM), or much less commonly, by prolonged neuromuscular blockade. From a clinical and EDx standpoint, it can be quite difficult to distinguish these disorders. Most specialists suggest that CIM is more common. Both CIM and CIP develop as a complication of sepsis and multiple organ failure. They usually present as an inability to wean a patient from a ventilator. A coexisting encephalopathy may limit the neurologic exam, in particular the sensory examination. Muscle stretch reflexes are absent or reduced.

Serum creatine kinase (CK) is usually normal; an elevated serum CK would point to CIM as opposed to CIP. NCS reveal absent or markedly reduced amplitudes of motor and sensory studies in CIP, while sensory studies are relatively preserved in CIM. Needle EMG usually reveals profuse positive sharp waves and fibrillation potentials, and it is not unusual in patients with severe weakness to be unable to recruit motor unit action potentials. The pathogenic basis of CIP is not known. Perhaps circulating toxins and metabolic abnormalities associated with sepsis and multiorgan failure impair axonal transport or mitochondrial function, leading to axonal degeneration.

Leprosy (Hansen Disease)

Leprosy, caused by the acid-fast bacteria *Mycobacterium leprae*, is the most common cause of peripheral neuropathy in Southeast Asia, Africa, and South America (Chap. 166). Clinical manifestations range from tuberculoid leprosy at one end to lepromatous leprosy at the other end of the spectrum, with borderline leprosy in between. Neuropathies are most common in patients with borderline leprosy. Superficial cutaneous nerves of the ears and distal limbs are commonly affected. Mononeuropathies, multiple mononeuropathies, or a slowly progressive symmetric sensorimotor polyneuropathy may develop. Sensory NCS are usually absent in the lower limb and are reduced in amplitude in the arms. Motor NCS may demonstrate reduced amplitudes in affected nerves but occasionally can reveal demyelinating features. Leprosy is usually diagnosed by skin lesion biopsy. Nerve biopsy can also be diagnostic, particularly when there are no apparent skin lesions. The tuberculoid form is characterized by granulomas, and bacilli are not seen. In contrast, with lepromatous leprosy, large numbers of infiltrating bacilli, TH2 lymphocytes, and organism-laden, foamy macrophages with minimal granulomatous infiltration are evident. The bacilli are best appreciated using the Fite stain, where they can be seen as red-staining rods often in clusters free in the endoneurium, within macrophages, or within Schwann cells.

Patients are generally treated with multiple drugs: dapsone, rifampin, and clofazimine. Other medications that are employed include thalidomide, pefloxacin, ofloxacin, sparfloxacin, minocycline, and clarithromycin. Patients are generally treated for 2 years. Treatment is sometimes complicated by the so-called reversal reaction, particularly in borderline leprosy. The reversal reaction can occur at any time during treatment and develops because of a shift to the tuberculoid end of the spectrum, with an increase in cellular immunity during treatment. The cellular response is upregulated as evidenced by an increased release of tumor necrosis factor α, interferon γ, and interleukin 2, with new granuloma formation. This can result in an exacerbation of the rash and the neuropathy as well as in appearance of new lesions. High-dose glucocorticoids blunt this adverse reaction and may be used prophylactically at treatment onset in high-risk patients. Erythema nodosum leprosum (ENL) is also treated with glucocorticoids or thalidomide.

Lyme Disease

Lyme disease is caused by infection with *Borrelia burgdorferi*, a spirochete usually transmitted by the deer tick *Ixodes dammini* (Chap. 173). Neurologic complications may develop during the second and third stages of infection. Facial neuropathy is most common and is bilateral in about half of cases, which is rare for idiopathic Bell's palsy. Involvement of nerves is frequently asymmetric. Some patients present with a polyradiculoneuropathy or multiple mononeuropathies. EDx is suggestive of a primary axonopathy. Nerve biopsies can reveal axonal degeneration with perivascular inflammation. Treatment is with antibiotics (Chap. 173).

Diphtheritic Neuropathy

Diphtheria is caused by the bacteria *Corynebacterium diphtheriae* (Chap. 138). Infected individuals present with flulike symptoms of generalized myalgias, headache, fatigue, low-grade fever, and irritability within a week to 10 days of the exposure. About 20–70% of patients develop a peripheral neuropathy caused by a toxin released by the bacteria. Three to 4 weeks after infection, patients may note decreased sensation in their throat and begin to develop dysphagia, dysarthria, hoarseness, and blurred vision due to impaired accommodation. A generalized polyneuropathy may manifest 2 or 3 months following the initial infection, characterized by numbness, paresthesias, and weakness of the arms and legs and occasionally ventilatory failure. CSF protein can be elevated with or without lymphocytic pleocytosis. EDx suggests a diffuse axonal sensorimotor polyneuropathy. Antitoxin and antibiotics should be given within 48 h of symptom onset (Chap. 138). Although early treatment reduces the incidence and severity of some complications (i.e., cardiomyopathy), it does not appear to alter the natural history of the associated peripheral neuropathy. The neuropathy usually resolves after several months.

Human Immunodeficiency Virus (HIV)

HIV infection can result in a variety of neurologic complications, including peripheral neuropathies (Chap. 189). Approximately 20% of HIV-infected individuals develop a neuropathy either as a direct result of the virus itself, other associated viral infections (e.g., cytomegalovirus), or neurotoxicity secondary to antiviral medications (see below). The major presentations of peripheral neuropathy associated with HIV infection include (1) distal symmetric polyneuropathy (DSP), (2) inflammatory demyelinating polyneuropathy (including both GBS and CIDP), (3) multiple mononeuropathies (e.g., vasculitis, CMV-related), (4) polyradiculopathy (usually CMV-related), (5) autonomic neuropathy, and (6) sensory ganglionitis.

HIV-RELATED DISTAL SYMMETRIC POLYNEUROPATHY (DSP)

DSP is the most common form of peripheral neuropathy associated with HIV infection and usually is seen in patients with AIDS. It is characterized by numbness and painful paresthesias involving the distal extremities. The pathogenic basis for DSP is unknown but is not due to actual infection of the peripheral nerves. The neuropathy may be immune mediated, perhaps caused by the release of cytokines from surrounding inflammatory cells. Vitamin B12 deficiency may contribute in some instances but is not a major cause of most cases of DSP. Some antiretroviral agents (e.g., dideoxycytidine, dideoxyinosine, stavudine) are also neurotoxic and can cause a painful sensory neuropathy.

HIV-RELATED INFLAMMATORY DEMYELINATING POLYRADICULONEUROPATHY

Both AIDP and CIDP can occur as a complication of HIV infection. AIDP usually develops at the time of seroconversion, while CIDP can occur any time in the course of the infection. Clinical and EDx features are indistinguishable from idiopathic AIDP or CIDP (discussed in next chapter). In addition to elevated protein levels, lymphocytic pleocytosis is evident in the CSF, a finding that helps distinguish this HIV-associated polyradiculoneuropathy from idiopathic AIDP/CIDP.

HIV-RELATED PROGRESSIVE POLYRADICULOPATHY

An acute, progressive lumbosacral polyradiculoneuropathy usually secondary to cytomegalovirus (CMV) infection can develop in patients with AIDS. Patients present with severe radicular pain, numbness, and weakness in the legs, which is usually asymmetric. CSF is abnormal, demonstrating an increased protein along with reduced glucose concentration and notably a neutrophilic pleocytosis. EDx studies reveal features of active axonal degeneration. The polyradiculoneuropathy may improve with antiviral therapy.

HIV-RELATED MULTIPLE MONONEUROPATHIES

Multiple mononeuropathies can also develop in patients with HIV infection, usually in the context of AIDS. Weakness, numbness, paresthesias, and pain occur in the distribution of affected nerves. Nerve biopsies can reveal axonal degeneration with necrotizing vasculitis or perivascular inflammation. Glucocorticoid treatment is indicated for vasculitis directly due to HIV infection.

HIV-RELATED SENSORY NEURONOPATHY/GANGLIONOPATHY

Dorsal root ganglionitis is a very rare complication of HIV infection, and neuronopathy can be the presenting manifestation. Patients develop sensory ataxia similar to idiopathic sensory neuronopathy/ganglionopathy. NCS reveal reduced amplitudes or absence of SNAPs.

Herpes Varicella-Zoster Virus

Peripheral neuropathy from Herpes varicella-zoster (HVZ) infection results from reactivation of latent virus or from a primary

infection (Chap. 180). Two-thirds of infections in adults are characterized by dermal zoster in which severe pain and paresthesias develop in a dermatomal region followed within a week or two by a vesicular rash in the same distribution. Weakness in muscles innervated by roots corresponding to the dermatomal distribution of skin lesions occurs in 5–30% of patients. Approximately 25% of affected patients have continued pain (postherpetic neuralgia, or PHN). A large clinical trial demonstrated that vaccination against zoster reduces the incidence of HZ among vaccine recipients by 51% and reduces the incidence of PHN by 67%. Treatment of postherpetic neuralgia is symptomatic (Table 384-6).

Cytomegalovirus

CMV can cause an acute lumbosacral polyradiculopathy and multiple mononeuropathies in patients with HIV infection and in other immune deficiency conditions (Chap. 182).

Epstein-Barr Virus

Epstein-Barr virus (EBV) infection has been associated with GBS, cranial neuropathies, mononeuropathy multiplex, brachial plexopathy, lumbosacral radiculoplexopathy, and sensory neuronopathies (Chap. 181).

Hepatitis Viruses

Hepatitis B and C can cause multiple mononeuropathies related to vasculitis, AIDP, or CIDP (Chap. 306).

NEUROPATHIES ASSOCIATED WITH MALIGNANCY

Patients with malignancy can develop neuropathies due to (1) a direct effect of the cancer by invasion or compression of the nerves, (2) remote or paraneoplastic effect, (3) a toxic effect of treatment, or (4) as a consequence of immune compromise caused by immunosuppressive medications. The most common associated malignancy is lung cancer, but neuropathies also complicate carcinoma of the breast, ovaries, stomach, colon, rectum, and other organs, including the lymphoproliferative system.

Paraneoplastic Sensory Neuronopathy/Ganglionopathy

Paraneoplastic encephalomyelitis/sensory neuronopathy (PEM/SN) usually complicates small cell lung carcinoma (Chap. 101). Patients usually present with numbness and paresthesias in the distal extremities that are often asymmetric. The onset can be acute or insidiously progressive. Prominent loss of proprioception leads to sensory ataxia. Weakness can be present, usually secondary to an associated myelitis, motor neuronopathy, or concurrent Lambert-Eaton myasthenic syndrome (LEMS). Many patients also develop confusion, memory loss, depression, hallucinations or seizures, or cerebellar ataxia. Polyclonal antineuronal antibodies (IgG) directed against a 35- to 40-kD protein or complex of proteins, the so-called Hu antigen, are found in the sera or CSF in the majority of patients with paraneoplastic PEM/SN. CSF may be normal or may demonstrate mild lymphocytic pleocytosis and elevated protein. PEM/SN is probably the result of antigenic similarity between proteins expressed in the tumor cells and neuronal cells, leading to an immune response directed against both cell types. Treatment of the underlying cancer generally does not affect the course of PEM/SN. However, occasional patients may improve following treatment of the tumor. Unfortunately, plasmapheresis, intravenous immunoglobulin, and immunosuppressive agents have not shown benefit.

Neuropathy Secondary to Tumor Infiltration

Malignant cells, in particular leukemia and lymphoma, can infiltrate cranial and peripheral nerves, leading to mononeuropathy, mononeuropathy multiplex, polyradiculopathy, plexopathy, or even a generalized symmetric distal or proximal and distal polyneuropathy. Neuropathy related to tumor infiltration is often painful; it can be the presenting manifestation of the cancer or the heralding symptom of a relapse. The neuropathy may improve with treatment of the underlying leukemia or lymphoma or with glucocorticoids.

Neuropathy as a Complication of Bone Marrow Transplantation

Neuropathies may develop in patients who undergo bone marrow transplantation (BMT) because of the toxic effects of chemotherapy, radiation, infection, or an autoimmune response directed against the peripheral nerves. Peripheral neuropathy in BMT is often associated with graft-versus-host disease (GVHD). Chronic GVHD shares many features with a variety of autoimmune disorders, and it is possible that an immune-mediated response directed against peripheral nerves is responsible. Patients with chronic GVHD may develop cranial neuropathies, sensorimotor polyneuropathies, multiple mononeuropathies, and severe generalized peripheral neuropathies resembling AIDP or CIDP. The neuropathy may improve by increasing the intensity of immunosuppressive or immunomodulating therapy and resolution of the GVHD.

Lymphoma

Lymphomas may cause neuropathy by infiltration or direct compression of nerves or by a paraneoplastic process. The neuropathy can be purely sensory or motor, but most commonly is sensorimotor. The pattern of involvement may be symmetric, asymmetric, or multifocal, and the course may be acute, gradually progressive, or relapsing and remitting. EDx can be compatible with either an axonal or demyelinating process. CSF may reveal lymphocytic pleocytosis and an elevated protein. Nerve biopsy may demonstrate endoneurial inflammatory cells in both the infiltrative and the paraneoplastic etiologies. A monoclonal population of cells favors lymphomatous invasion. The neuropathy may respond to treatment of the underlying lymphoma or immunomodulating therapies.

Multiple Myeloma

Multiple myeloma (MM) usually presents in the fifth to seventh decade of life with fatigue, bone pain, anemia, and hypercalcemia (Chap 111). Clinical and EDx features of neuropathy occur in as many as 40% of patients. The most common pattern is that of a distal, axonal, sensory, or sensorimotor polyneuropathy. Less frequently, a chronic demyelinating polyradiculoneuropathy may develop (see POEMS, Chap. 385). MM can be complicated by amyloid polyneuropathy and should be considered in patients with painful paresthesias, loss of pinprick and temperature discrimination, and autonomic dysfunction (suggestive of a small-fiber neuropathy) and carpal tunnel syndrome. Expanding plasmacytomas can compress cranial nerves and spinal roots as well. A monoclonal protein, usually composed of γ or μ heavy chains or κ light chains, may be identified in the serum or urine. EDx usually shows reduced amplitudes with normal or only mildly abnormal distal latencies and conduction velocities. A superimposed median neuropathy at the wrist is common. Abdominal fat pad, rectal, or sural nerve biopsy can be performed to look for amyloid deposition. Unfortunately, the treatment of the underlying MM does not usually affect the course of the neuropathy.

Neuropathies Associated with Monoclonal Gammopathy of Uncertain Significance (see Chap. 375)

TOXIC NEUROPATHIES SECONDARY TO CHEMOTHERAPY

Many of the commonly used chemotherapy agents can cause a toxic neuropathy **(Table 384-7)**. The mechanisms by which these agents cause toxic neuropathies vary as can the specific type of neuropathy produced. The risk of developing a toxic neuropathy or more severe neuropathy appears to be greater in patients with a preexisting neuropathy (e.g., Charcot-Marie-Tooth disease, diabetic neuropathy) and those who also take other potentially neurotoxic drugs (e.g., nitrofurantoin, isoniazid, disulfiram, pyridoxine). Chemotherapeutic agents usually cause a sensory greater than motor length-dependent axonal

neuropathy or neuronopathy/ganglionopathy.

Table 384-7 Toxic Neuropathies Secondary to Chemotherapy

Drug	Mechanism of Neurotoxicity	Clinical Features	Nerve Histopathology	EMG/NCS
Vinca alkaloids (vincristine, vinblastine, vindesine, vinorelbine)	Interfere with axonal microtubule assembly; impairs axonal transport	Symmetric, S-M, large-/small-fiber PN; autonomic symptoms common; infrequent cranial neuropathies	Axonal degeneration of myelinated and unmyelinated fibers; regenerating clusters, minimal segmental demyelination	Axonal sensorimotor PN; distal denervation on EMG; abnormal QST, particularly vibratory perception
Cisplatin	Preferential damage to dorsal root ganglia: ? binds to and cross-links DNA? inhibits protein synthesis ? impairs axonal transport	Predominant large-fiber sensory neuronopathy; sensory ataxia	Loss of large > small myelinated and unmyelinated fibers; axonal degeneration with small clusters of regenerating fibers; secondary segmental demyelination	Low-amplitude or unobtainable SNAPs with normal CMAPs and EMG; abnormal QST, particularly vibratory perception
Taxanes (paclitaxel, docetaxel)	Promotes axonal microtubule assembly; interferes with axonal transport	Symmetric, predominantly sensory, PN; large-fiber modalities affected more than small-fiber	Loss of large > small myelinated and unmyelinated fibers; axonal degeneration with small clusters of regenerating fibers; secondary segmental demyelination	Axonal sensorimotor PN; distal denervation on EMG; abnormal QST, particularly vibratory perception
Suramin Axonal PN	Unknown; ? inhibition of neurotrophic growth factor binding; ? neuronal lysosomal storage	Symmetric, length-dependent, sensory-predominant, PN	None described	Abnormalities consistent with an axonal S-M PN
Demyelinating PN	Unknown; ? immunomodulating effects	Subacute, S-M PN with diffuse proximal and distal weakness; areflexia; increased CSF protein	Loss of large and small myelinated fibers with primary demyelination and secondary axonal degeneration; occasional epi- and endoneurial inflammatory cell infiltrates	Features suggestive of an acquired demyelinating sensorimotor PN (e.g., slow CVs, prolonged distal latencies and F-wave latencies, conduction block, temporal dispersion)
Ara-C	Unknown; ? selective Schwann cell toxicity; ? immunomodulating effects	GBS-like syndrome; pure sensory neuropathy; brachial plexopathy	Loss of myelinated nerve fibers; axonal degeneration; segmental demyelination; no inflammation	Axonal, demyelinating, or mixed S-M PN; denervation on EMG
Etoposide (VP-16)	Unknown; ? selective dorsal root ganglia toxicity	Length-dependent, sensory predominant PN; autonomic neuropathy	None described	Abnormalities consistent with an axonal S-M PN
Bortezomib (Velcade)	Unknown	Length-dependent, sensory, predominantly small-fiber, PN	Not reported	Abnormalities consistent with an axonal sensory neuropathy with early small-fiber involvement (abnormal autonomic studies)

Abbreviations: CSF, cerebrospinal fluid; CVs, conduction velocities; EMG, electromyography; GBS, Guillain-Barré syndrome; NCS, nerve conduction studies; PN, polyneuropathy; QST, quantitative sensory testing; S-M, sensorimotor. ***Source:*** From Amato and Russell.

OTHER TOXIC NEUROPATHIES

Neuropathies can develop as complications of toxic effects of various drugs and other environmental exposures **(Table 384-8)**. The more common neuropathies associated with these agents are discussed here.

Table 384-8 Toxic Neuropathies

Drug	Mechanism of Neurotoxicity	Clinical Features	Nerve Histopathology	EMG/NCS
Misonidazole	Unknown	Painful paresthesias and loss of large- and small-fiber sensory modalities and sometimes distal weakness in length-dependent pattern	Axonal degeneration of large myelinated fibers; axonal swellings; segmental demyelination	Low-amplitude or unobtainable SNAPs with normal or only slightly reduced CMAPs amplitudes
Metronidazole	Unknown	Painful paresthesias and loss of large- and small-fiber sensory modalities and sometimes distal weakness in length-dependent pattern	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal CMAPs
Chloroquine and hydroxychloroquine	Amphiphilic properties may lead to drug-lipid complexes that are indigestible	Loss of large- and small-fiber sensory modalities and distal weakness in length-dependent pattern; superimposed myopathy	Axonal degeneration with autophagic vacuoles in nerves as well as muscle fibers	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes; distal denervation on EMG; irritability and myopathic-appearing MUAPs

	and result in accumulation of autophagic vacuoles	may lead to proximal weakness		proximally in patients with superimposed toxic myopathy
Amiodarone	Amphiphilic properties may lead to drug-lipid complexes that are indigestible and result in accumulation of autophagic vacuoles	Paresthesias and pain with loss of large- and small-fiber sensory modalities and distal weakness in length-dependent pattern; superimposed myopathy may lead to proximal weakness	Axonal degeneration and segmental demyelination with myeloid inclusions in nerves and muscle fibers	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes; can also have prominent slowing of CVs; distal denervation on EMG; irritability and myopathic-appearing MUAPs proximally in patients with superimposed toxic myopathy
Colchicine	Inhibits polymerization of tubulin in microtubules and impairs axoplasmic flow	Numbness and paresthesias with loss of large-fiber modalities in a length-dependent fashion; superimposed myopathy may lead to proximal in addition to distal weakness	Nerve biopsy demonstrates axonal degeneration; muscle biopsy reveals fibers with vacuoles	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes; irritability and myopathic-appearing MUAPs proximally in patients with superimposed toxic myopathy
Podophyllin	Binds to microtubules and impairs axoplasmic flow	Sensory loss, tingling, muscle weakness, and diminished muscle stretch reflexes in length-dependent pattern; autonomic neuropathy	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Thalidomide	Unknown	Numbness, tingling, and burning pain and weakness in a length-dependent pattern	Axonal degeneration; autopsy studies reveal degeneration of dorsal root ganglia	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Disulfiram	Accumulation of neurofilaments and impaired axoplasmic flow	Numbness, tingling, and burning pain in a length-dependent pattern	Axonal degeneration with accumulation of neurofilaments in the axons	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Dapsone	Unknown	Distal weakness that may progress to proximal muscles; sensory loss	Axonal degeneration and segmental demyelination	Low-amplitude or unobtainable CMAPs with normal or reduced SNAP amplitudes
Leflunomide	Unknown	Paresthesias and numbness in a length-dependent pattern	Unknown	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Nitrofurantoin	Unknown	Numbness, painful paresthesias, and severe weakness that may resemble GBS	Axonal degeneration; autopsy studies reveal degeneration of dorsal root ganglia and anterior horn cells	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Pyridoxine (vitamin B6)	Unknown	Dysesthesias and sensory ataxia; impaired large-fiber sensory modalities on examination	Marked loss of sensory axons and cell bodies in dorsal root ganglia	Reduced amplitudes or absent SNAPs
Isoniazid	Inhibits pyridoxal phosphokinase leading to pyridoxine deficiency	Dysesthesias and sensory ataxia; impaired large-fiber sensory modalities on examination	Marked loss of sensory axons and cell bodies in dorsal root ganglia and degeneration of the dorsal columns	Reduced amplitudes or absent SNAPs and to lesser extent CMAPs
Ethambutol	Unknown	Numbness with loss of large-fiber modalities on examination	Axonal degeneration	Reduced amplitudes or absent SNAPs
Antinucleosides	Unknown	Dysesthesia and sensory ataxia; impaired large-fiber sensory modalities on examination	Axonal degeneration	Reduced amplitudes or absent SNAPs
Phenytoin	Unknown	Numbness with loss of large-fiber modalities on examination	Axonal degeneration and segmental demyelination	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Lithium	Unknown	Numbness with loss of large-fiber modalities on examination	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Acrylamide	Unknown; may be caused by impaired axonal transport	Numbness with loss of large-fiber modalities on examination; sensory ataxia; mild distal weakness	Degeneration of sensory axons in peripheral nerves and posterior columns, spinocerebellar tracts, mammillary bodies, optic tracts, and corticospinal tracts in the CNS	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Carbon disulfide	Unknown	Length-dependent numbness and tingling with mild distal weakness	Axonal swellings with accumulation of neurofilaments	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Ethylene oxide	Unknown; may act as alkylating agent and bind DNA	Length-dependent numbness and tingling; may have mild distal weakness	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Organophosphates	Bind and inhibit	Early features are those	Axonal degeneration	Early: repetitive firing of CMAPs and

	neuropathy target esterase	of neuromuscular blockade with generalized weakness; later axonal sensorimotor PN ensues	along with degeneration of gracile fasciculus and corticospinal tracts	decrement with repetitive nerve stimulation; late: axonal sensorimotor PN
Hexacarbons	Unknown; may lead to covalent cross-linking between neurofilaments	Acute, severe sensorimotor PN that may resemble GBS	Axonal degeneration and giant axons swollen with neurofilaments	Features of a mixed axonal and/or demyelinating sensorimotor axonal PN—reduced amplitudes, prolonged distal latencies, conduction block, and slowing of CVs
Lead	Unknown; may interfere with mitochondria	Encephalopathy; motor neuropathy (often resembles radial neuropathy with wrist and finger drop); autonomic neuropathy; bluish-black discoloration of gums	Axonal degeneration of motor axons	Reduction of CMAP amplitudes with active denervation on EMG
Mercury	Unknown; may combine with sulfhydryl groups	Abdominal pain and nephrotic syndrome; encephalopathy; ataxia; paresthesias	Axonal degeneration; degeneration of dorsal root ganglia, calcarine, and cerebellar cortex	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Thallium	Unknown	Encephalopathy; painful sensory symptoms; mild loss of vibration; distal or generalized weakness may also develop; autonomic neuropathy; alopecia	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes
Arsenic	Unknown; may combine with sulfhydryl groups	Abdominal discomfort, burning pain and paresthesias; generalized weakness; autonomic insufficiency; can resemble GBS	Axonal degeneration	Low-amplitude or unobtainable SNAPs with normal or reduced CMAPs amplitudes; may have demyelinating features: prolonged distal latencies and slowing of CVs
Gold	Unknown	Distal paresthesias and reduction of all sensory modalities	Axonal degeneration	Low-amplitude or unobtainable SNAPs

Abbreviations: CMAP, compound motor action potential; CVs, conduction velocities; EMG, electromyography; GBS, Guillain-Barré syndrome; MUAP, muscle action potential; NCS, nerve conduction studies; PN, polyneuropathy; S-M, sensorimotor; SNAP, sensory nerve action potential. ***Source:*** From Amato and Russell.

Chloroquine and Hydroxychloroquine

Chloroquine and hydroxychloroquine can cause a toxic myopathy characterized by slowly progressive, painless, proximal weakness and atrophy, which is worse in the legs than the arms. In addition, neuropathy can also develop with or without the myopathy leading to sensory loss and distal weakness. The "neuromyopathy" usually appears in patients taking 500 mg daily for a year or more but has been reported with doses as low as 200 mg/d. Serum CK levels are usually elevated due to the superimposed myopathy. NCS reveal mild slowing of motor and sensory nerve conduction velocities with a mild to moderate reduction in the amplitudes, although NCS may be normal in patients with only the myopathy. EMG demonstrates myopathic muscle action potentials (MUAPs), increased insertional activity in the form of positive sharp waves, fibrillation potentials, and occasionally myotonic potentials, particularly in the proximal muscles. Neurogenic MUAPs and reduced recruitment are found in more distal muscles. Nerve biopsy demonstrates autophagic vacuoles within Schwann cells. Vacuoles may also be evident in muscle biopsies. The pathogenic basis of the neuropathy is not known but may be related to the amphiphilic properties of the drug. These agents contain both hydrophobic and hydrophilic regions that allow them to interact with the anionic phospholipids of cell membranes and organelles. The drug-lipid complexes may be resistant to digestion by lysosomal enzymes, leading to the formation of autophagic vacuoles filled with myeloid debris that may in turn cause degeneration of nerves and muscle fibers. The signs and symptoms of the neuropathy and myopathy are usually reversible following discontinuation of medication.

Amiodarone

Amiodarone can cause a neuromyopathy similar to chloroquine and hydroxychloroquine. The neuromyopathy typically appears after patients have taken the medication for 2–3 years. Nerve biopsy demonstrates a combination of segmental demyelination and axonal loss. Electron microscopy reveals lamellar or dense inclusions in Schwann cells, pericytes, and endothelial cells. The inclusions in muscle and nerve biopsies have persisted as long as 2 years following discontinuation of the medication.

Colchicine

Colchicine can also cause a neuromyopathy. Patients usually present with proximal weakness and numbness and tingling in the distal extremities. EDx reveals features of an axonal polyneuropathy. Muscle biopsy reveals a vacuolar myopathy, while sensory nerves demonstrate axonal degeneration. Colchicine inhibits the polymerization of tubulin into microtubules. The disruption of the microtubules probably leads to defective intracellular movement of important proteins, nutrients, and waste products in muscle and nerves.

Thalidomide

Thalidomide is an immunomodulating agent used to treat multiple myeloma, GVHD, leprosy, and other autoimmune disorders. Thalidomide is associated with severe teratogenic effects as well as peripheral neuropathy that can be dose-limiting. Patients develop numbness, painful tingling, and burning discomfort in the feet and hands and less commonly muscle weakness and atrophy. Even after stopping the drug for 4–6 years, as many as 50% patients continue to have significant symptoms. NCS demonstrate reduced amplitudes or complete absence of sensory nerve action potentials (SNAPs), with preserved conduction velocities when obtainable. Motor NCS are usually normal. Nerve biopsy reveals a loss of large-diameter myelinated fibers and axonal degeneration. Degeneration of dorsal root ganglion cells has been reported at autopsy.

Pyridoxine (Vitamin B_6) Toxicity

Pyridoxine is an essential vitamin that serves as a coenzyme for transamination and decarboxylation. However, at high doses (116 mg/d), patients can develop a severe sensory neuropathy with dysesthesias and sensory ataxia. NCS reveal absent or markedly reduced SNAP amplitudes with relatively preserved CMAPs. Nerve biopsy reveals axonal loss of fiber at all diameters. Loss of dorsal root ganglion cells with subsequent degeneration of both the peripheral and central sensory tracts have been reported in animal models.

Isoniazid

One of the most common side effects of isoniazid (INH) is peripheral neuropathy. Standard doses of INH (3–5 mg/kg per d) are associated with a 2% incidence of neuropathy, while neuropathy develops in at least 17% of patients taking in excess of 6 mg/kg per d. The elderly, malnourished, and "slow acetylators" are at increased risk for developing the neuropathy. INH inhibits pyridoxal phosphokinase, resulting in pyridoxine deficiency and the neuropathy. Prophylactic administration of pyridoxine 100 mg/d can prevent the neuropathy from developing.

Antiretroviral Agents

The nucleoside analogues zalcitabine (dideoxycytidine or ddC), didanosine (dideoxyinosine or ddI), stavudine (d4T), lamivudine (3TC), and antiretroviral nucleoside reverse transcriptase inhibitor (NRTI) are used to treat HIV infection. One of the major dose-limiting side effects of these medications is a predominantly sensory, length-dependent, symmetrically painful neuropathy. Zalcitabine (ddC) is the most extensively studied of the nucleoside analogues and at doses greater than 0.18 mg/kg per d is associated with a subacute onset of severe burning and lancinating pains in the feet and hands. NCS reveal decreased amplitudes of the SNAPs with normal motor studies. The nucleoside analogues inhibit mitochondrial DNA polymerase, which is the suspected pathogenic basis for the neuropathy. Because of a "coasting effect," patients can continue to worsen even 2–3 weeks after stopping the medication. Following dose reduction, improvement in the neuropathy is seen in most patients after several months (mean time about 10 weeks).

Hexacarbons (*N*-Hexane, Methyl *N*-Butyl Ketone)/Glue Sniffer's Neuropathy

n-Hexane and methyl *n*-butyl ketone are water-insoluble industrial organic solvents that are also present in some glues. Exposure through inhalation, accidentally or intentionally (glue sniffing), or through skin absorption can lead to a profound subacute sensory and motor polyneuropathy. NCS demonstrate decreased amplitudes of the SNAPs and CMAPs with slightly slow CVs. Nerve biopsy reveals a loss of myelinated fibers and giant axons that are filled with 10-nm neurofilaments. Hexacarbon exposure leads to covalent cross-linking between axonal neurofilaments that result in their aggregation, impaired axonal transport, swelling of the axons, and eventual axonal degeneration.

Lead

Lead neuropathy is uncommon but it can be seen in children who accidentally ingest lead-based paints in older buildings and in industrial workers exposed to lead-containing products. The most common presentation of lead poisoning is an encephalopathy; however, symptoms and signs of a primarily motor neuropathy can also occur. The neuropathy is characterized by an insidious and progressive onset of weakness usually beginning in the arms, in particular involving the wrist and finger extensors, resembling a radial neuropathy. Sensation is generally preserved; however, the autonomic nervous system can be affected. Laboratory investigation can reveal a microcytic hypochromic anemia with basophilic stippling of erythrocytes, an elevated serum lead level, and an elevated serum coproporphyrin level. A 24-h urine collection demonstrates elevated levels of lead excretion. The NCS may reveal reduced CMAP amplitudes, while the SNAPs are typically normal. The pathogenic basis may be related to abnormal porphyrin metabolism. The most important principle of management is to remove the source of the exposure. Chelation therapy with calcium disodium ethylenediaminetetraacetic acid (EDTA), British anti-Lewisite (BAL), and penicillamine also demonstrates variable efficacy.

Mercury

Mercury toxicity may occur as a result of exposure to either organic or inorganic mercurials. Mercury poisoning presents with paresthesias in hands and feet that progress proximally and may involve the face and tongue. Motor weakness can also develop. CNS symptoms often overshadow the neuropathy. EDx shows features of a primarily axonal sensorimotor polyneuropathy. The primary site of neuromuscular pathology appears to be the dorsal root ganglia. The mainstay of treatment is removing the source of exposure.

Thallium

Thallium can exist in a monovalent or trivalent form and is primarily used as a rodenticide. The toxic neuropathy usually manifests as burning paresthesias of the feet, abdominal pain, and vomiting. Increased thirst, sleep disturbances, and psychotic behavior may be noted. Within the first week, patients develop pigmentation of the hair, an acne-like rash in the malar area of the face, and hyperreflexia. By the second and third week, autonomic instability with labile heart rate and blood pressure may be seen. Hyporeflexia and alopecia also occur but may not be evident until the third or fourth week following exposure. With severe intoxication, proximal weakness and involvement of the cranial nerves can occur. Some patients require mechanical ventilation due to respiratory muscle involvement. The lethal dose of thallium is variable, ranging from 8 to 15 mg/kg body weight. Death can result in less than 48 h following a particularly large dose. NCS demonstrate features of a primarily axonal sensorimotor polyneuropathy. With acute intoxication, potassium ferric ferrocyanide II may be effective in preventing absorption of thallium from the gut. However, there may be no benefit once thallium has been absorbed. Unfortunately, chelating agents are not very efficacious. Adequate diuresis is essential to help eliminate thallium from the body without increasing tissue availability from the serum.

Arsenic

Arsenic is another heavy metal that can cause a toxic sensorimotor polyneuropathy. The neuropathy manifests 5–10 days after ingestion of arsenic and progresses for several weeks, sometimes mimicking GBS. The presenting symptoms are typically an abrupt onset of abdominal discomfort, nausea, vomiting, pain, and diarrhea followed within several days by burning pain in the feet and hands. Examination of the skin can be helpful in the diagnosis as the loss of the superficial epidermal layer results in patchy regions of increased or decreased pigmentation on the skin several weeks after an acute exposure or with chronic low levels of ingestion. Mee's lines, which are transverse lines at the base of the fingernails and toenails, do not become evident until 1 or 2 months after the exposure. Multiple Mee's lines may be seen in patients with long fingernails who have had chronic exposure to arsenic. Mee's lines are not specific for arsenic toxicity as they can also be seen following thallium poisoning. Because arsenic is cleared from blood rapidly, the serum concentration of arsenic is not diagnostically helpful. However, arsenic levels are increased in the urine, hair, and fingernails of patients exposed to arsenic. Anemia with stippling of erythrocytes is common, and occasionally pancytopenia and aplastic anemia can develop. Increased CSF protein levels without pleocytosis can be seen; this can lead to misdiagnosis as GBS. NCS are usually suggestive of an axonal sensorimotor polyneuropathy; however, demyelinating features can be present. Chelation therapy with BAL has yielded inconsistent results; therefore, it is not generally recommended.

NUTRITIONAL NEUROPATHIES

Cobalamin (Vitamin B_{12})

Pernicious anemia is the most common cause of cobalamin deficiency. Other causes include dietary avoidance (vegetarians), gastrectomy, gastric bypass surgery, inflammatory bowel disease, pancreatic insufficiency, bacterial overgrowth, and possibly histamine-2 blockers and proton-pump inhibitors. An underappreciated cause of cobalamin deficiency is food-cobalamin malabsorption. This typically occurs in older individuals and results from an inability to adequately absorb cobalamin in food

protein. No apparent cause of deficiency is identified in a significant number of patients with cobalamin deficiency. The use of nitrous oxide as an anesthetic agent or from recreational use can produce acute cobalamin deficiency neuropathy and subacute combined degeneration.

Complaints of numb hands typically appear before lower extremity paresthesias are noted. A preferential large-fiber sensory loss affecting proprioception and vibration with sparing of small-fiber modalities is present; an unsteady gait reflects sensory ataxia. These features, coupled with diffuse hyperreflexia and absent Achilles reflexes, should always focus attention on the possibility of cobalamin deficiency. Optic atrophy and, in severe cases, behavioral changes ranging from mild irritability and forgetfulness to severe dementia and frank psychosis may appear. The full clinical picture of subacute combined degeneration is uncommon. CNS manifestations, especially pyramidal tract signs, may be missing, and in fact some patients may only exhibit symptoms of peripheral neuropathy.

EDx shows an axonal sensorimotor neuropathy. CNS involvement produces abnormal somatosensory and visual evoked potential latencies. The diagnosis is confirmed by finding reduced serum cobalamin levels. In up to 40% of patients, anemia and macrocytosis are lacking. Serum methylmalonic acid and homocysteine, the metabolites that accumulate when cobalamin-dependent reactions are blocked, are elevated. Antibodies to intrinsic factor are present in approximately 60%, and antiparietal cell antibodies in about 90%, of individuals with pernicious anemia.

Cobalamin deficiency can be treated with various regimens of cobalamin. One typical regimen consists of 1000 μg cyanocobalamin IM weekly for 1 month and monthly thereafter. Patients with food cobalamin malabsorption can absorb free cobalamin and therefore can be treated with oral cobalamin supplementation. An oral cobalamin dose of 1000 μg per day should be sufficient. Treatment for cobalamin deficiency usually does not completely reverse the clinical manifestations, and at least 50% of patients exhibit some permanent neurologic deficit.

Thiamine Deficiency

Thiamine (vitamin B1) deficiency is an uncommon cause of peripheral neuropathy in developed countries. It is now most often seen as a consequence of chronic alcohol abuse, recurrent vomiting, total parenteral nutrition, and bariatric surgery. Thiamine deficiency polyneuropathy can occur in normal, healthy young adults who do not abuse alcohol but who engage in inappropriately restrictive diets. Thiamine is water-soluble. It is present in most animal and plant tissues, but the greatest sources are unrefined cereal grains, wheat germ, yeast, soybean flour, and pork. Beriberi means "I can't, I can't" in Singhalese, the language of natives of what was once part of the Dutch East Indies (now Sri Lanka). *Dry beriberi* refers to neuropathic symptoms. The term *wet beriberi* is used when cardiac manifestations predominate (in reference to edema). Beriberi was relatively uncommon until the late 1800s when it became widespread among people for whom rice was a dietary mainstay. This epidemic was due to a new technique of processing rice that removed the germ from the rice shaft, rendering the so-called polished rice deficient in thiamine and other essential nutrients.

Symptoms of neuropathy follow prolonged deficiency. These begin with mild sensory loss and/or burning dysesthesias in the toes and feet and aching and cramping in the lower legs. Pain may be the predominant symptom. With progression, patients develop features of a nonspecific generalized polyneuropathy, with distal sensory loss in the feet and hands.

Blood and urine assays for thiamine are not reliable for diagnosis of deficiency. Erythrocyte transketolase activity and the percentage increase in activity (in vitro) following the addition of thiamine pyrophosphate (TPP) may be more accurate and reliable. EDx shows nonspecific findings of an axonal sensorimotor polyneuropathy. When a diagnosis of thiamine deficiency is made or suspected, thiamine replacement should be provided until proper nutrition is restored. Thiamine is usually given intravenously or intramuscularly at a dose of 100 mg/d. Although cardiac manifestations show a striking response to thiamine replacement, neurologic improvement is usually more variable and less dramatic.

Vitamin E Deficiency

The term *vitamin E* is usually used for α-tocopherol, the most active of the four main types of vitamin E. Because vitamin E is present in animal fat, vegetable oils, and various grains, deficiency is usually due to factors other than insufficient intake. Vitamin E deficiency usually occurs secondary to lipid malabsorption or in uncommon disorders of vitamin E transport. One hereditary disorder is abetalipoproteinemia, a rare autosomal dominant disorder characterized by steatorrhea, pigmentary retinopathy, acanthocytosis, and progressive ataxia. Patients with cystic fibrosis may also have vitamin E deficiency secondary to steatorrhea. There are genetic forms of isolated vitamin E deficiency not associated with lipid malabsorption. Vitamin E deficiency may also occur as a consequence of various cholestatic and hepatobiliary disorders as well as short-bowel syndromes resulting from the surgical treatment of intestinal disorders.

Clinical features may not appear until many years after the onset of deficiency. The onset of symptoms tends to be insidious, and progression is slow. The main clinical features are spinocerebellar ataxia and polyneuropathy, thus resembling Friedreich ataxia or other spinocerebellar ataxias. Patients manifest progressive ataxia and signs of posterior column dysfunction, such as impaired joint position and vibratory sensation. Because of the polyneuropathy, there is hyporeflexia, but plantar responses may be extensor as a result of the spinal cord involvement. Other neurologic manifestations may include ophthalmoplegia, pigmented retinopathy, night blindness, dysarthria, pseudoathetosis, dystonia, and tremor. Vitamin E deficiency may present as an isolated polyneuropathy, but this is very rare. The yield of checking serum vitamin E levels in patients with isolated polyneuropathy is extremely low, and this test should not be part of routine practice.

Diagnosis is made by measuring α-tocopherol levels in the serum. EDx shows features of an axonal neuropathy. Treatment is replacement with oral vitamin E, but high doses are not needed. For patients with isolated vitamin E deficiency, treatment consists of 1500–6000 IU/d in divided doses.

Vitamin B_6 Deficiency

Vitamin B6, or pyridoxine, can produce neuropathic manifestations from both deficiency and toxicity. Vitamin B_6 toxicity was discussed above. Vitamin B_6 deficiency is most commonly seen in patients treated with isoniazid or hydralazine. The polyneuropathy of vitamin B_6 is nonspecific, manifesting as a generalized axonal sensorimotor polyneuropathy. Vitamin B_6 deficiency can be detected by direct assay. Vitamin B_6 supplementation with 50–100 mg/d is suggested for patients being treated with isoniazid or hydralazine. This same dose is appropriate for replacement in cases of nutritional deficiency.

Pellagra (Niacin Deficiency)

Pellagra is produced by deficiency of niacin. Although pellagra may be seen in alcoholics, this disorder has essentially been eradicated in most Western countries by means of enriching bread with niacin. Nevertheless, pellagra continues to be a problem in a number of underdeveloped regions, particularly in Asia and Africa, where corn is the main source of carbohydrate. Neurologic manifestations are variable; abnormalities can develop in the brain and spinal cord as well as peripheral nerves. When peripheral nerves are involved, the neuropathy is usually mild and resembles beriberi. Treatment is with niacin 40–250 mg/d.

Copper Deficiency

A syndrome that has only recently been described is myeloneuropathy secondary to copper deficiency. Most patients present with lower limb paresthesias, weakness, spasticity, and gait difficulties. Large-fiber sensory function is impaired, reflexes are brisk, and plantar responses are extensor. In some cases, light touch and pinprick sensation are affected, and nerve conduction studies indicate sensorimotor axonal polyneuropathy in addition to myelopathy.

Hematologic abnormalities are a known complication of copper deficiency; these can include microcytic anemia, neutropenia, and occasionally pancytopenia. Because copper is absorbed in the stomach and proximal jejunum, many cases of copper deficiency are in the setting of prior gastric surgery. Excess zinc is an established cause of copper deficiency. Zinc upregulates enterocyte production of metallothionine, which results in decreased absorption of copper. Excessive dietary zinc supplements or denture cream containing zinc can produce this clinical picture. Other potential causes of copper deficiency include malnutrition, prematurity, total parenteral nutrition, and ingestion of copper chelating agents.

Following oral or IV copper replacement, some patients show neurologic improvement, but this may take many months or not occur at all. Replacement consists of oral copper sulfate or gluconate 2 mg one to three times a day. If oral copper replacement is not effective, elemental copper in the copper sulfate or copper chloride forms can be given as 2 mg IV daily for 3–5 days, then weekly for 1–2 months until copper levels normalize. Thereafter, oral daily copper therapy can be resumed. In contrast to the neurologic manifestations, most of the hematologic indices completely normalize in response to copper replacement therapy.

Neuropathy Associated with Gastric Surgery

Polyneuropathy may occur following gastric surgery for ulcer, cancer, or weight reduction. This usually occurs in the context of rapid, significant weight loss and recurrent, protracted vomiting. The clinical picture is one of acute or subacute sensory loss and weakness. Neuropathy following weight loss surgery usually occurs in the first several months after surgery. Weight reduction surgical procedures include gastrojejunostomy, gastric stapling, vertical banded gastroplasty, and gastrectomy with Roux-en-Y anastomosis. The initial manifestations are usually numbness and paresthesias in the feet. In many cases, no specific nutritional deficiency factor is identified.

Management consists of parenteral vitamin supplementation, especially including thiamine. Improvement has been observed following supplementation, parenteral nutritional support, and reversal of the surgical bypass. The duration and severity of deficits before identification and treatment of neuropathy are important predictors of final outcome.

CRYPTOGENIC (IDIOPATHIC) SENSORY AND SENSORIMOTOR POLYNEUROPATHY

CSPN is a diagnosis of exclusion, established after a careful medical, family, and social history; neurologic examination; and directed laboratory testing. Despite extensive evaluation, the cause of polyneuropathy in as many as 50% of all patients is idiopathic. CSPN should be considered a distinct diagnostic subset of peripheral neuropathy. The onset of CSPN is predominantly in the sixth and seventh decades. Patients complain of distal numbness, tingling, and often burning pain that invariably begins in the feet and may eventually involve the fingers and hands. Patients exhibit a distal sensory loss to pinprick, touch, and vibration in the toes and feet, and occasionally in the fingers. It is uncommon to see significant proprioception deficits, even though patients may complain of gait unsteadiness. However, tandem gait may be abnormal in a minority of cases. Neither subjective nor objective evidence of weakness is a prominent feature. Most patients have evidence of both large- and small-fiber loss on neurologic exam and EDx. Approximately 10% of patients have only evidence of small-fiber involvement. The ankle muscle stretch reflex is frequently absent, but in cases with predominantly small-fiber loss, this may be preserved. The EDx findings range from isolated sensory nerve action potential abnormalities (usually with loss of amplitude), to evidence for an axonal sensorimotor neuropathy, to a completely normal study (if primarily small fibers are involved). Therapy primarily involves the control of neuropathic pain (Table 384-6) if present. These drugs should not be used if the patient has only numbness and tingling but no pain.

Although no treatment is available that can reverse an idiopathic distal peripheral neuropathy, the prognosis is good. Progression often does not occur or is minimal, with sensory symptoms and signs progressing proximally up to the knees and elbows. The disorder does not lead to significant motor disability over time. The relatively benign course of this disorder should be explained to patients.

MONONEUROPATHIES/PLEXOPATHIES/RADICULOPATHIES

Median Neuropathy

CTS is a compression of the median nerve in the carpal tunnel at the wrist. The median nerve enters the hand through the carpal tunnel by coursing under the transverse carpal ligament. The symptoms of CTS consist of numbness and paresthesias variably in the thumb, index, middle, and half of the ring finger. At times, the paresthesias can include the entire hand and extend into the forearm or upper arm or can be isolated to one or two fingers. Pain is another common symptom and can be located in the hand and forearm and, at times, in the proximal arm. CTS is common and often misdiagnosed as thoracic outlet syndrome. The signs of CTS are decreased sensation in the median nerve distribution; reproduction of the sensation of tingling when a percussion hammer is tapped over the wrist (Tinel's sign) or the wrist is flexed for 30–60 s (Phalen's sign); and weakness of thumb opposition and abduction. EDx is extremely sensitive and shows slowing of sensory and, to a lesser extent, motor median potentials across the wrist. Treatment options consist of avoidance of precipitating activities; control of underlying systemic-associated conditions if present; nonsteroidal anti-inflammatory medications; neutral (volar) position wrist splints, especially for night use; glucocorticoid/anesthetic injection into the carpal tunnel; and surgical decompression by dividing the transverse carpal ligament. The surgical option should be considered if there is a poor response to nonsurgical treatments; if there is thenar muscle atrophy and/or weakness; and if there are significant denervation potentials on EMG.

Other proximal median neuropathies are very uncommon and include the pronator teres syndrome and anterior interosseous neuropathy. These often occur as a partial form of brachial plexitis.

Ulnar Neuropathy at the Elbow—"Cubital Tunnel Syndrome"

The ulnar nerve passes through the condylar groove between the medial epicondyle and the olecranon. Symptoms consist of paresthesias, tingling, and numbness in the medial hand and half of the fourth and the entire fifth fingers, pain at the elbow or forearm, and weakness. Signs consist of decreased sensation in an ulnar distribution, Tinel's sign at the elbow, and weakness and atrophy of ulnar-innervated hand muscles. The Froment sign indicates thumb adductor weakness and consists of flexion of the thumb at the interphalangeal joint when attempting to oppose the thumb against the lateral border of the second digit. EDx may show slowing of ulnar motor nerve conduction velocity across the elbow with prolonged ulnar sensory latencies. Treatment consists of avoiding aggravating factors, using elbow pads, and surgery to decompress the nerve in the cubital tunnel. Ulnar neuropathies can also rarely occur at the wrist in the ulnar (Guyon) canal or in the hand, usually after trauma.

Radial Neuropathy

The radial nerve winds around the proximal humerus in the spiral groove and proceeds down the lateral arm and enters the forearm, dividing into the posterior interosseous nerve and superficial nerve. The symptoms and signs consist of wristdrop; finger extension weakness; thumb abduction weakness; and sensory loss in the dorsal web between the thumb and index finger. Triceps and brachioradialis strength is often normal, and triceps reflex is often intact. Most cases of radial neuropathy are

transient compressive (neuropraxic) injuries that recover spontaneously in 6–8 weeks. If there has been prolonged compression and severe axonal damage, it may take several months to recover. Treatment consists of cock-up wrist and finger splints, avoiding further compression, and physical therapy to avoid flexion contracture. If there is no improvement in 2–3 weeks, an EDx study is recommended to confirm the clinical diagnosis and determine the degree of severity.

Lateral Femoral Cutaneous Neuropathy (Meralgia Paresthetica)

The lateral femoral cutaneous nerve arises from the upper lumbar plexus (spinal levels L2/3), crosses through the inguinal ligament near its attachment to the iliac bone, and supplies sensation to the anterior lateral thigh. The neuropathy affecting this nerve is also known as meralgia paresthetica. Symptoms and signs consist of paresthesias, numbness, and occasionally pain in the lateral thigh. Symptoms are increased by standing or walking and are relieved by sitting. There is normal strength and knee reflexes are intact. The diagnosis is clinical, and further tests usually are not performed. EDx is only needed to rule out lumbar plexopathy, radiculopathy, or femoral neuropathy. If the symptoms and signs are classic, electromyography is not necessary. Symptoms often resolve spontaneously over weeks or months, but the patient may be left with permanent numbness. Treatment consists of weight loss and avoiding tight belts. Analgesics in the form of a lidocaine patch, nonsteroidal agents, and occasionally medications for neuropathic pain, can be used (Table 384-6). Rarely, locally injecting the nerve with an anesthetic can be tried. There is no role for surgery.

Femoral Neuropathy

Femoral neuropathies can arise as complications of retroperitoneal hematoma, lithotomy positioning, hip arthroplasty or dislocation, iliac artery occlusion, femoral arterial procedures, infiltration by hematogenous malignancy, penetrating groin trauma, pelvic surgery including hysterectomy and renal transplantation, and diabetes (a partial form of lumbosacral diabetic plexopathy); some cases are idiopathic. Patients with femoral neuropathy have difficulty extending their knee and flexing the hip. Sensory symptoms occurring either on the anterior thigh and/or medial leg occur in only half of reported cases. A prominent painful component is the exception rather than the rule, may be delayed, and is often self-limited in nature. The quadriceps (patellar) reflex is diminished.

Sciatic Neuropathy

Sciatic neuropathies commonly complicate hip arthroplasty, pelvic procedures in which patients are placed in a prolonged lithotomy position, trauma, hematomas, tumor infiltration, and vasculitis. In addition, many sciatic neuropathies are idiopathic. Weakness may involve all motions of the ankles and toes as well as flexion of the leg at the knee; abduction and extension of the thigh at the hip is spared. Sensory loss occurs in the entire foot and the distal lateral leg. The ankle jerk and on occasion the internal hamstring reflex are diminished or more typically absent on the affected side. The peroneal subdivision of the sciatic nerve is typically involved disproportionately to the tibial counterpart. Thus, patients may have only ankle dorsiflexion and eversion weakness with sparing of knee flexion, ankle inversion, and plantar flexion; these features can lead to misdiagnosis of a common peroneal neuropathy.

Peroneal Neuropathy

The sciatic nerve divides at the distal femur into the tibial and peroneal nerve. The common peroneal nerve passes posterior and laterally around the fibular head, under the fibular tunnel. It then divides into the superficial peroneal nerve, which supplies the ankle evertor muscles and sensation over the anterolateral distal leg and dorsum of the foot, and the deep peroneal nerve, which supplies ankle dorsiflexors and toe extensor muscles and a small area of sensation dorsally in the area of the first and second toes.

Symptoms and signs consist of footdrop (ankle dorsiflexion, toe extension, and ankle eversion weakness) and variable sensory loss, which may involve the superficial and deep peroneal pattern. There is usually no pain. Onset may be on awakening in the morning. Peroneal neuropathy needs to be distinguished from L5 radiculopathy. In L5 radiculopathy, ankle invertors and evertors are weak and needle electromyography reveals denervation. EDx can help localize the lesion. Peroneal motor conduction velocity shows slowing and amplitude drop across the fibular head. Management consists of rapid weight loss and avoiding leg crossing. Footdrop is treated with an ankle brace. A knee pad can be worn over the lateral knee to avoid further compression. Most cases spontaneously resolve over weeks or months.

RADICULOPATHIES

Radiculopathies are most often due to compression from degenerative joint disease and herniated disks, but there are a number of unusual etiologies **(Table 384-9)**. Degenerative spine disease affects a number of different structures, which narrow the diameter of the neural foramen or canal of the spinal column and compromise nerve root integrity; **these are discussed in detail in Chap. 15**.

Table 384-9 Causes of Radiculopathy

- Herniated nucleus pulposus
- Degenerative joint disease
- Rheumatoid arthritis
- Trauma
- Vertebral body compression fracture
- Pott disease
- Compression by extradural mass (e.g., meningioma, metastatic tumor, hematoma, abscess)
- Primary nerve tumor (e.g., neurofibroma, schwannoma, neurinoma)
- Carcinomatous meningitis
- Perineurial spread of tumor (e.g., prostate cancer)
- Acute inflammatory demyelinating polyradiculopathy
- Chronic inflammatory demyelinating polyradiculopathy
- Sarcoidosis
- Amyloidoma
- Diabetic radiculopathy
- Infection (Lyme disease, herpes zoster, cytomegalovirus, syphilis, schistosomiasis, strongyloides)

PLEXOPATHIES

Brachial Plexus

The brachial plexus is composed of three trunks (upper, middle, and lower), with two divisions (anterior and posterior) per trunk **(Fig. 384-2)**. Subsequently, the trunks divide into three cords (medial, lateral, and posterior), and from these arise the multiple terminal nerves innervating the arm. The anterior primary rami of C5 and C6 fuse to form the upper trunk; the anterior primary ramus of C7 continues as the middle trunk, while the anterior rami of C8 and T1 join to form the lower trunk. There are several disorders commonly associated with brachial plexopathy.

Figure 384-2

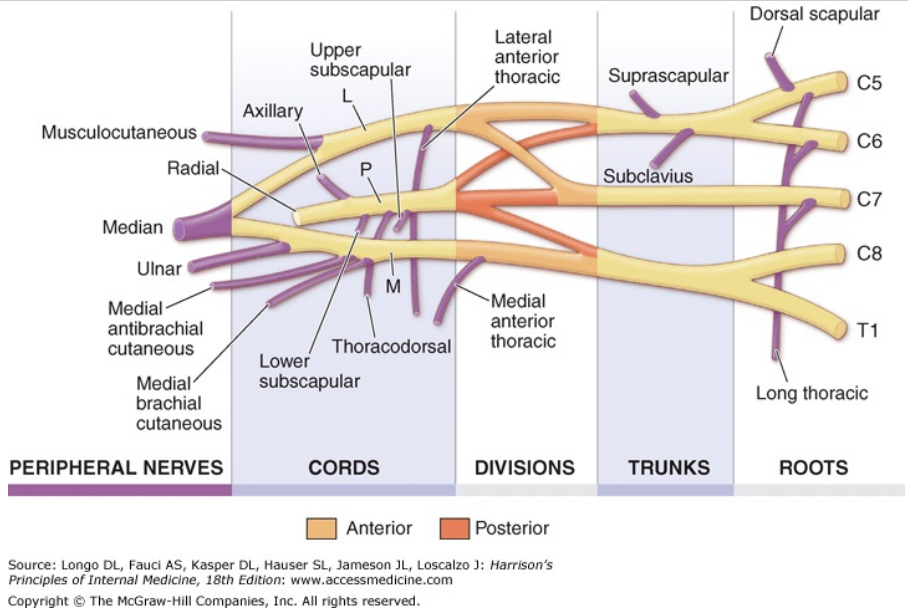

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Brachial plexus anatomy. L, lateral; M, medial; P, posterior. *(From J Goodgold: Anatomical Correlates of Clinical Electromyography. Baltimore, Williams and Wilkins, 1974, p. 126, with permission.)*

IMMUNE-MEDIATED BRACHIAL PLEXUS NEUROPATHY

Immune-mediated brachial plexus neuropathy (IBPN) goes by various terms, including *acute brachial plexitis*, *neuralgic amyotrophy*, and *Parsonage-Turner syndrome*. IBPN usually presents with an acute onset of severe pain in the shoulder region. The intense pain usually lasts several days to a few weeks, but a dull ache can persist. Individuals who are affected may not appreciate weakness of the arm early in the course because the pain limits movement. However, as the pain dissipates, weakness and often sensory loss are appreciated. Attacks can occasionally recur.

Clinical findings are dependent on the distribution of involvement (e.g., specific trunk, divisions, cords, or terminal nerves). The most common pattern of IBPN involves the upper trunk or a single or multiple mononeuropathies primarily involving the suprascapular, long thoracic, or axillary nerves. Additionally, the phrenic and anterior interosseous nerves may be concomitantly affected. Any of these nerves may also be affected in isolation. EDx is useful to confirm and localize the site(s) of involvement. Empirical treatment of severe pain with glucocorticoids is often used in the acute period.

BRACHIAL PLEXOPATHIES ASSOCIATED WITH NEOPLASMS

Neoplasms involving the brachial plexus may be primary nerve tumors, local cancers expanding into the plexus (e.g., Pancoast lung tumor or lymphoma), and metastatic tumors. Primary brachial plexus tumors are less common than the secondary tumors and include schwannomas, neurinomas, and neurofibromas. Secondary tumors affecting the brachial plexus are more common and are always malignant. These may arise from local tumors, expanding into the plexus. For example, a Pancoast tumor of the upper lobe of the lung may invade or compress the lower trunk, while a primary lymphoma arising from the cervical or axillary lymph nodes may also infiltrate the plexus. Pancoast tumors typically present as an insidious onset of pain in the upper arm, sensory disturbance in the medial aspect of the forearm and hand, and weakness and atrophy of the intrinsic hand muscles along with an ipsilateral Horner syndrome. Chest CT scans or MRI can demonstrate extension of the tumor into the plexus. Metastatic involvement of the brachial plexus may occur with spread of breast cancer into the axillary lymph nodes with local spread into the nearby nerves.

PERIOPERATIVE PLEXOPATHIES (MEDIAN STERNOTOMY)

The most common surgical procedures associated with brachial plexopathy as a complication are those that involve median sternotomies (e.g., open-heart surgeries and thoracotomies). Brachial plexopathies occur in as many as 5% of patients following a median sternotomy and typically affect the lower trunk. Thus, individuals manifest with sensory disturbance affecting the medial aspect of forearm and hand along with weakness of the intrinsic hand muscles. The mechanism is related to the stretch of the lower trunk, so most individuals who are affected recover within a few months.

Lumbosacral Plexus

The lumbar plexus arises from the ventral primary rami of the first to the fourth lumbar spinal nerves **(Fig. 384-3)**. These nerves pass downward and laterally from the vertebral column within the psoas major muscle. The femoral nerve derives from the dorsal branches of the second to the fourth lumbar ventral rami. The obturator nerve arises from the ventral branches of the same lumbar rami. The lumbar plexus communicates with the sacral plexus by the lumbosacral trunk, which contains some fibers from the fourth and all of those from the fifth lumbar ventral rami **(Fig. 384-4)**.

Figure 384-3

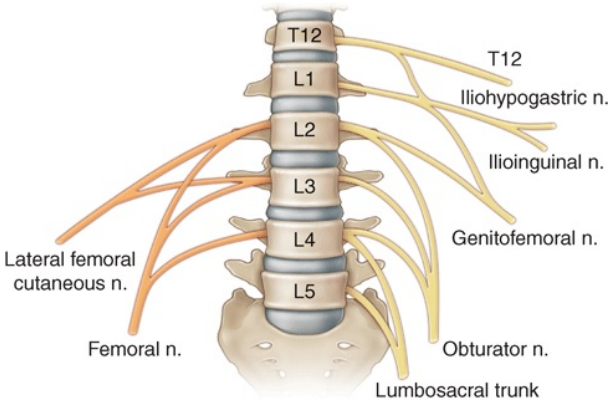

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lumbar plexus. Posterior divisions are in orange, anterior divisions are in yellow. *(From J Goodgold: Anatomical Correlates of Clinical Electromyography. Baltimore, Williams and Wilkins, 1974, p. 126, with permission.)*

Figure 384-4

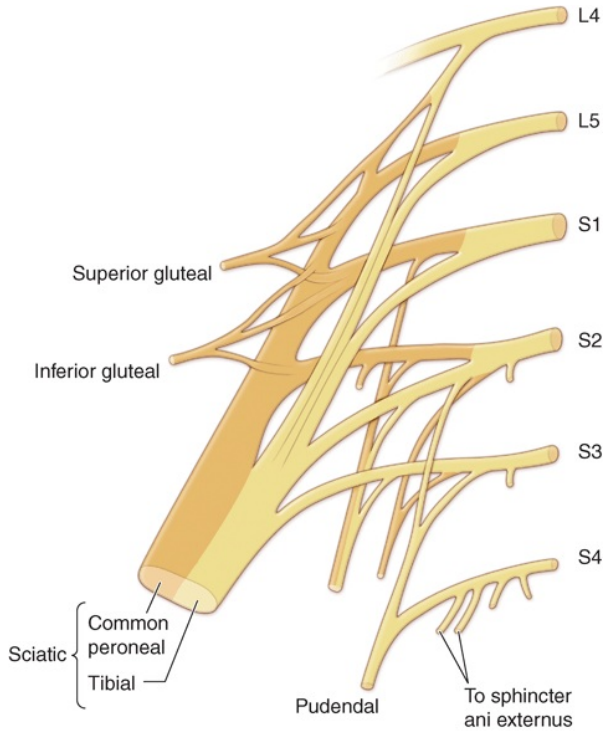

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lumbosacral plexus. Posterior divisions are in orange, anterior divisions are in yellow. *(From J Goodgold: Anatomical Correlates of Clinical Electromyography. Baltimore, Williams and Wilkins, 1974, p. 126, with permission.)*

The sacral plexus is the part of the lumbosacral plexus that is formed by the union of the lumbosacral trunk with the ventral rami of the first to fourth sacral nerves. The plexus lies on the posterior and posterolateral wall of the pelvis with its components converging toward the sciatic notch. The lateral trunk of the sciatic nerve (which forms the common peroneal nerve) arises from the union of the dorsal branches of the lumbosacral trunk (L4, L5) and the dorsal branches of the S1 and S2 spinal nerve ventral rami. The medial trunk of the sciatic nerve (which forms the tibial nerve) derives from the ventral branches of the same ventral rami (L4-S2).

Lumbosacral Plexopathies

Plexopathies are typically recognized when motor, sensory, and if applicable, reflex deficits occur in multiple nerve and segmental distributions confined to one extremity. If localization within the lumbosacral plexus can be accomplished, designation as a lumbar plexopathy, a sacral plexopathy, a lumbosacral trunk lesion, or a pan-plexopathy is the best localization that can be expected. Although lumbar plexopathies may be bilateral, usually occurring in a stepwise and chronologically dissociated manner, sacral plexopathies are more likely to behave in this manner due to their closer anatomic proximity. The differential diagnosis of plexopathy includes disorders of the conus medullaris and cauda equina (polyradiculopathy). If there is a paucity of pain and sensory involvement, motor neuron disease should be considered as well.

The causes of lumbosacral plexopathies are listed in **Table 384-10**. Diabetic radiculopathy (discussed above) is a fairly common cause of painful leg weakness. Lumbosacral plexopathies are a well-recognized complication of retroperitoneal hemorrhage. Various primary and metastatic malignancies can affect the lumbosacral plexus as well; these include carcinoma of the cervix, endometrium, and ovary; osteosarcoma; testicular cancer; multiple myeloma; lymphoma; acute myelogenous leukemia; colon

cancer; squamous cell carcinoma of the rectum; adenocarcinoma of unknown origin; and intraneural spread of prostate cancer.

Table 384-10 Lumbosacral Plexopathies: Etiologies

- Retroperitoneal hematoma
- Psoas abscess
- Malignantneoplasm
- Benign neoplasm
- Radiation
- Amyloid
- Diabetic radiculoplexus neuropathy
- Idiopathic radiculoplexus neuropathy
- Sarcoidosis
- Aortic occlusion/surgery
- Lithotomy positioning
- Hip arthroplasty
- Pelvic fracture
- Obstetric injury

Recurrent Neoplastic Disease or Radiation-Induced Plexopathy

The treatment for various malignancies is often radiation therapy, the field of which may include parts of the brachial plexus. It can be difficult in such situations to determine if a new brachial or lumbosacral plexopathy is related to tumor within the plexus or from radiation-induced nerve damage. Radiation can be associated with microvascular abnormalities and fibrosis of surrounding tissues, which can damage the axons and the Schwann cells. Radiation-induced plexopathy can develop months or years following therapy and is dose dependent.

Tumor invasion is usually painful and more commonly affects the lower trunk, while radiation injury is often painless and affects the upper trunk. Imaging studies such as MRI and CT scans are useful but can be misleading with small microscopic invasion of the plexus. EMG can be informative if myokymic discharges are appreciated, as this finding strongly suggests radiation-induced damage.

Evaluation and Treatment of Plexopathies

Most patients with plexopathies will undergo both imaging with MRI and EDx evaluations. Severe pain from acute idiopathic lumbosacral plexopathy may respond to a short course of glucocorticoids.

FURTHER READINGS

Amato AA, Russell J: *Neuromuscular Disease*. New York, McGraw-Hill, 2008

Dumitru D et al: *Electrodiagnostic Medicine,* 2nd ed. Philadelphia, Hanley & Belfus, Inc, 2002

Dyck PJ, Thomas PK (eds): *Peripheral Neuropathy*, 4th ed. Philadelphia, Elsevier Saunders, 2005

England JD et al: Evaluation of distal symmetric polyneuropathy: The role of autonomic testing, nerve biopsy, and skin biopsy (an evidence-based review). Muscle Nerve 39:106, 2009[PMID: 19086069] [Full Text]

———: Evaluation of distal symmetric polyneuropathy: The role of laboratory and genetic testing (an evidence-based review). Muscle Nerve 39:116, 2009

Jani-Acsadi A et al: Charcot-Marie-Tooth neuropathies: Diagnosis and management. Semin Neurol 28:185, 2008[PMID: 18351520] [Full Text]

Mendell JR et al: *Diagnosis and Management of Peripheral Nerve Disorders* . New York, Oxford University Press, 2001

Pascuzzi RM. Peripheral neuropathy. Med Clin North Am 93:317, 2009[PMID: 19272511] [Full Text]

Samuels MA, Ropper A (eds): *Manual of Neurologic Therapeutics*. Boston, Little, Brown, and Company, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

GUILLAIN-BARRÉ SYNDROME

Guillain-Barré syndrome (GBS) is an acute, frequently severe, and fulminant polyradiculoneuropathy that is autoimmune in nature. It occurs year-round at a rate of between 1 and 4 cases per 100,000 annually; in the United States, ~5000–6000 cases occur per year. Males are at slightly higher risk for GBS than females, and in Western countries adults are more frequently affected than children.

CLINICAL MANIFESTATIONS

GBS manifests as a rapidly evolving areflexic motor paralysis with or without sensory disturbance. The usual pattern is an ascending paralysis that may be first noticed as rubbery legs. Weakness typically evolves over hours to a few days and is frequently accompanied by tingling dysesthesias in the extremities. The legs are usually more affected than the arms, and facial diparesis is present in 50% of affected individuals. The lower cranial nerves are also frequently involved, causing bulbar weakness with difficulty handling secretions and maintaining an airway; the diagnosis in these patients may initially be mistaken for brainstem ischemia. Pain in the neck, shoulder, back, or diffusely over the spine is also common in the early stages of GBS, occurring in ~50% of patients. Most patients require hospitalization, and in different series up to 30% require ventilatory assistance at some time during the illness. The need for mechanical ventilation is associated with more severe weakness on admission, a rapid tempo of progression, and the presence of facial and/or bulbar weakness during the first week of symptoms. Fever and constitutional symptoms are absent at the onset and, if present, cast doubt on the diagnosis. Deep tendon reflexes attenuate or disappear within the first few days of onset. Cutaneous sensory deficits (e.g., loss of pain and temperature sensation) are usually relatively mild, but functions subserved by large sensory fibers, such as deep tendon reflexes and proprioception, are more severely affected. Bladder dysfunction may occur in severe cases but is usually transient. If bladder dysfunction is a prominent feature and comes early in the course, diagnostic possibilities other than GBS should be considered, particularly spinal cord disease. Once clinical worsening stops and the patient reaches a plateau (almost always within 4 weeks of onset), further progression is unlikely.

Autonomic involvement is common and may occur even in patients whose GBS is otherwise mild. The usual manifestations are loss of vasomotor control with wide fluctuation in blood pressure, postural hypotension, and cardiac dysrhythmias. These features require close monitoring and management and can be fatal. Pain is another common feature of GBS; in addition to the acute pain described above, a deep aching pain may be present in weakened muscles that patients liken to having overexercised the previous day. Other pains in GBS include dysesthetic pain in the extremities as a manifestation of sensory nerve fiber involvement. These pains are self-limited and often respond to standard analgesics (Chap. 11).

Several subtypes of GBS are recognized, as determined primarily by electrodiagnostic (Edx) and pathologic distinctions **(Table 385-1)**. The most common variant is acute inflammatory demyelinating polyneuropathy (AIDP). Additionally, there are two axonal variants, which are often clinically severe —the acute motor axonal neuropathy (AMAN) and acute motor sensory axonal neuropathy (AMSAN) subtypes. In addition, a range of limited or regional GBS syndromes are also encountered. Notable among these is the Miller Fisher syndrome (MFS), which presents as rapidly evolving ataxia and areflexia of limbs without weakness, and ophthalmoplegia, often with pupillary paralysis. The MFS variant accounts for ~5% of all cases and is strongly associated with antibodies to the ganglioside GQ1b (see "Immunopathogenesis," below). Other regional variants of GBS include (1) pure sensory forms; (2) ophthalmoplegia with anti-GQ1b antibodies as part of severe motor-sensory GBS; (3) GBS with severe bulbar and facial paralysis, sometimes associated with antecedent cytomegalovirus (CMV) infection and anti-GM2 antibodies; and (4) acute pandysautonomia (Chap. 375).

Table 385-1 Subtypes of Guillain-Barré Syndrome (GBS)

Subtype	Features	Electrodiagnosis	Pathology
Acute inflammatory demyelinating polyneuropathy (AIDP)	Adults affected more than children; 90% of cases in western world; recovery rapid; anti-GM1 antibodies (<50%)	Demyelinating	First attack on Schwann cell surface; widespread myelin damage, macrophage activation, and lymphocytic infiltration; variable secondary axonal damage

Acute motor axonal neuropathy (AMAN)	Children and young adults; prevalent in China and Mexico; may be seasonal; recovery rapid; anti-GD1a antibodies	Axonal	First attack at motor nodes of Ranvier; macrophage activation, few lymphocytes, frequent periaxonal macrophages; extent of axonal damage highly variable
Acute motor sensory axonal neuropathy (AMSAN)	Mostly adults; uncommon; recovery slow, often incomplete; closely related to AMAN	Axonal	Same as AMAN, but also affects sensory nerves and roots; axonal damage usually severe
M. Fisher syndrome (MFS)	Adults and children; uncommon; ophthalmoplegia, ataxia, and areflexia; anti-GQ1b antibodies (90%)	Demyelinating	Few cases examined; resembles AIDP

ANTECEDENT EVENTS

Approximately 70% of cases of GBS occur 1–3 weeks after an acute infectious process, usually respiratory or gastrointestinal. Culture and seroepidemiologic techniques show that 20–30% of all cases occurring in North America, Europe, and Australia are preceded by infection or reinfection with *Campylobacter jejuni* . A similar proportion is preceded by a human herpes virus infection, often CMV or Epstein-Barr virus. Other viruses and also *Mycoplasma pneumoniae* have been identified as agents involved in antecedent infections, as have recent immunizations. The swine influenza vaccine, administered widely in the United States in 1976, is the most notable example. Influenza vaccines in use from 1992 to 1994, however, resulted in only one additional case of GBS per million persons vaccinated, and the more recent seasonal influenza vaccines appear to confer a GBS risk of <1 per million. A recent study demonstrated that there does not appear to be an increased risk of GBS with meningococcal vaccinations (Menactra) contrary to early reports. Older-type rabies vaccine, prepared in nervous system tissue, is implicated as a trigger of GBS in developing countries where it is still used; the mechanism is presumably immunization against neural antigens. GBS also occurs more frequently than can be attributed to chance alone in patients with lymphoma (including Hodgkin's disease), in HIV-seropositive individuals, and in patients with systemic lupus erythematosus (SLE). *C. jejuni* has also been implicated in summer outbreaks of AMAN among children and young adults exposed to chickens in rural China.

IMMUNOPATHOGENESIS

Several lines of evidence support an autoimmune basis for acute inflammatory demyelinating polyneuropathy (AIDP), the most common and best-studied type of GBS; the concept extends to all of the subtypes of GBS (Table 385-1).

It is likely that both cellular and humoral immune mechanisms contribute to tissue damage in AIDP. T cell activation is suggested by the finding that elevated levels of cytokines and cytokine receptors are present in serum [interleukin (IL) 2, soluble IL-2 receptor] and in cerebrospinal fluid (CSF) (IL-6, tumor necrosis factor α, interferon-γ). AIDP is also closely analogous to an experimental T cell–mediated immunopathy designated *experimental allergic neuritis* (EAN). EAN is induced in laboratory animals by immune sensitization against protein fragments derived from peripheral nerve proteins, and in particular against the P2 protein. Based on analogy to EAN, it was initially thought that AIDP was likely to be primarily a T cell–mediated disorder; however, abundant data now suggest that autoantibodies directed against nonprotein determinants may be central to many cases.

Circumstantial evidence suggests that all GBS results from immune responses to nonself antigens (infectious agents, vaccines) that misdirect to host nerve tissue through a resemblance-of-epitope (molecular mimicry) mechanism **(Fig. 385-1)**. The neural targets are likely to be glycoconjugates, specifically gangliosides **(Table 385-2**; **Fig. 385-2)**. Gangliosides are complex glycosphingolipids that contain one or more sialic acid residues; various gangliosides participate in cell-cell interactions (including those between axons and glia), modulation of receptors, and regulation of growth. They are typically exposed on the plasma membrane of cells, rendering them susceptible to an antibody-mediated attack. Gangliosides and other glycoconjugates are present in large quantity in human nervous tissues and in key sites, such as nodes of Ranvier. Antiganglioside antibodies, most frequently to GM1, are common in GBS (20–50% of cases), particularly in those preceded by *C. jejuni* infection. Furthermore, isolates of *C. jejuni* from stool cultures of patients with GBS have surface glycolipid structures that antigenically cross react with gangliosides, including GM1, concentrated in human nerves. Sialic acid residues from pathogenic *C. jejuni* strains can also trigger activation of dendritic cells via signaling through a toll-like receptor (TLR4), promoting B-cell differentiation and further amplifying humoral autoimmunity. Another line of evidence is derived from experience in Europe with parenteral use of purified bovine brain gangliosides for treatment of various neuropathic disorders. Between 5 and 15 days after injection, some recipients developed acute motor axonal GBS with high titers of anti-GM1 antibodies that recognized epitopes at nodes of Ranvier and motor endplates. Experimentally, anti-GM1 antibodies can trigger complement-mediated injury at paranodal axon-glial

junctions, disrupting the clustering of sodium channels and likely contributing to conduction block (see "Pathophysiology," below).

Figure 385-1

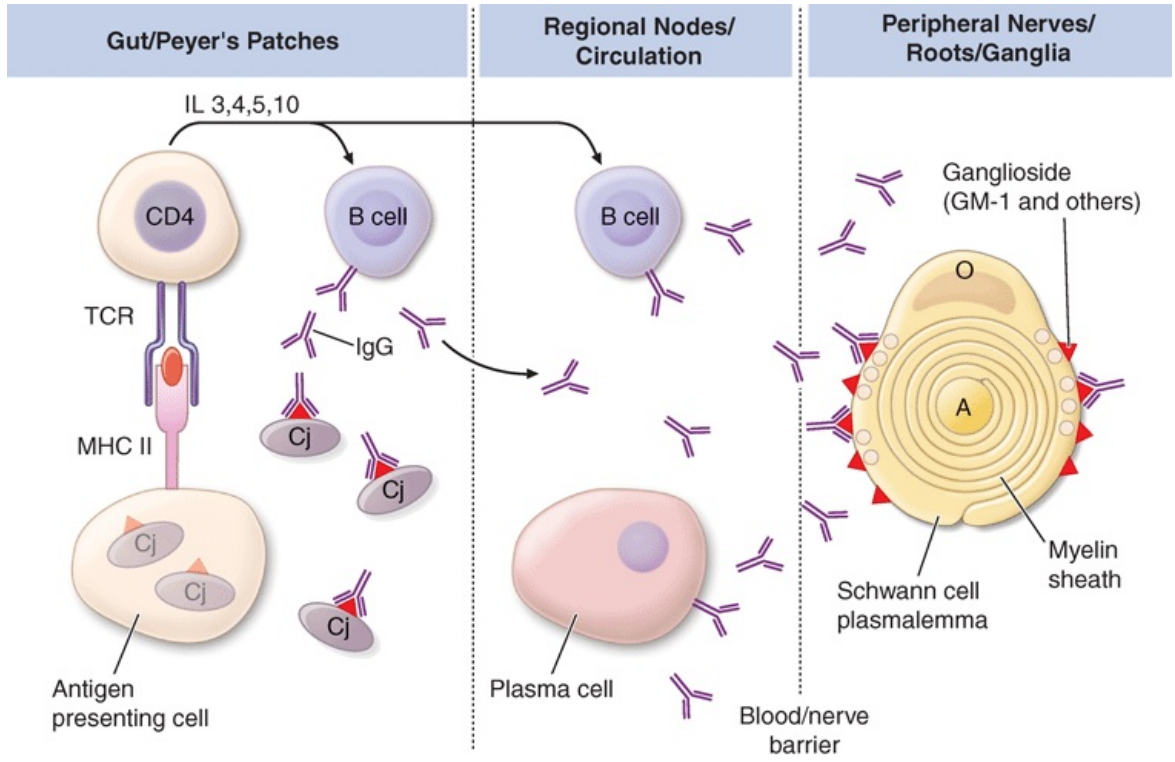

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Postulated immunopathogenesis of GBS associated with *C. jejuni* infection. B cells recognize glycoconjugates on *C. jejuni* (Cj) (triangles) that cross-react with ganglioside present on Schwann cell surface and subjacent peripheral nerve myelin. Some B cells, activated via a T cell–independent mechanism, secrete primarily IgM (not shown). Other B cells (upper left side) are activated via a partially T cell–dependent route and secrete primarily IgG; T cell help is provided by CD4 cells activated locally by fragments of Cj proteins that are presented on the surface of antigen-presenting cells (APCs). A critical event in the development of GBS is the escape of activated B cells from Peyer's patches into regional lymph nodes. Activated T cells probably also function to assist in opening of the blood-nerve barrier, facilitating penetration of pathogenic autoantibodies. The earliest changes in myelin (right) consist of edema between myelin lamellae and vesicular disruption (shown as circular blebs) of the outermost myelin layers. These effects are associated with activation of the C5b-C9 membrane attack complex and probably mediated by calcium entry; it is possible that the macrophage cytokine tumor necrosis factor (TNF) also participates in myelin damage. B, B cell; MHC II, class II major histocompatibility complex molecule; TCR, T cell receptor; A, axon; O, oligodendrocyte.

Table 385-2 Principal Anti-Glycolipid Antibodies Implicated in Immune Neuropathies

Clinical Presentation	Antibody Target	Usual Isotype
Acute Immune Neuropathies (Guillain-Barré Syndrome)		
Acute inflammatory demyelinating polyneuropathy (AIDP)	No clear patterns	IgG (polyclonal)
	GM1 most common	
Acute motor axonal neuropathy (AMAN)	GD1a, GM1, GM1b, GalNAc –GD1a (<50% for any)	IgG (polyclonal)
Miller Fisher syndrome (MFS)	GQ1b (>90%)	IgG (polyclonal)
Acute pharyngeal cervicobrachial neuropathy (APCBN)	GT1a (? Most)	IgG (polyclonal)
Chronic Immune Neuropathies		
Chronic inflammatory demyelinating polyneuropathy (CIDP) (75%)	Po in some	No clear pattern

CIDPa (MGUS associated) (25%)	Neural binding sites	IgG, IgA (monoclonal)
Chronic sensory > motor neuropathy	SPGP, SGLPG (on MAG) (50%)	IgM (monoclonal)
	Uncertain (50%)	IgM (monoclonal)
Multifocal motor neuropathy (MMN)	GM1, GalNAc–GD1a, others (25–50%)	IgM (polyclonal, monoclonal)
Chronic sensory ataxic neuropathy	GD1b, GQ1b, and other b-series gangliosides	IgM (monoclonal)

Abbreviations: MAG, myelin-associated glycoprotein; MGUS, monoclonal gammopathy of undetermined significance. ***Source:*** Modified from HJ Willison, N Yuki: Brain 125:2591, 2002.

Figure 385-2

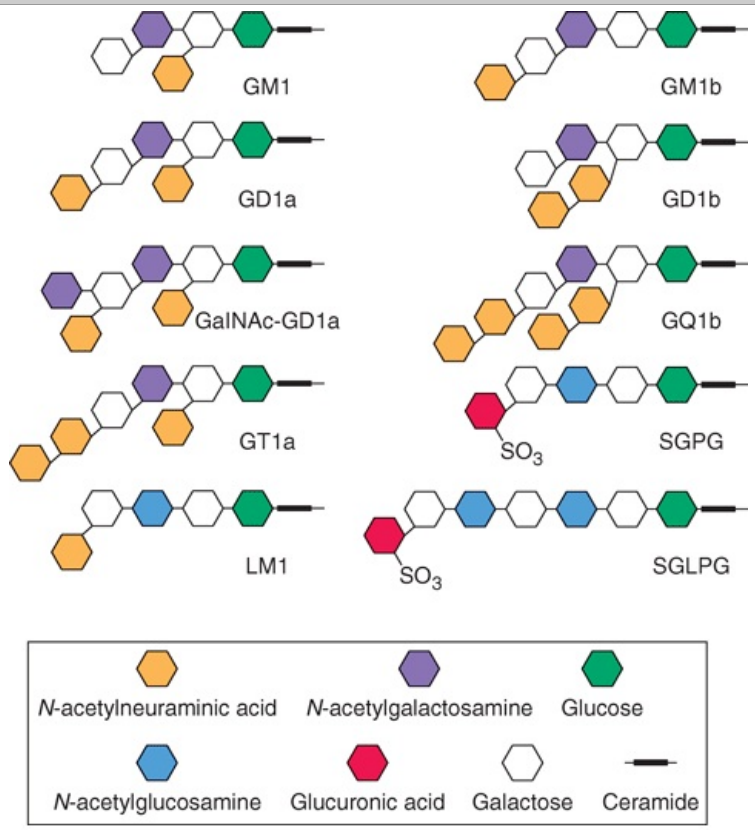

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Glycolipids implicated as antigens in immune-mediated neuropathies. *(Modified from HJ Willison, N Yuki: Brain 125:2591, 2002.)*

Anti-GQ1b IgG antibodies are found in >90% of patients with MFS (Table 385-2; Fig. 385-2), and titers of IgG are highest early in the course. Anti-GQ1b antibodies are not found in other forms of GBS unless there is extraocular motor nerve involvement. A possible explanation for this association is that extraocular motor nerves are enriched in GQ1b gangliosides in comparison to limb nerves. In addition, a monoclonal anti-GQ1b antibody raised against *C. jejuni* isolated from a patient with MFS blocked neuromuscular transmission experimentally.

Taken together, these observations provide strong but still inconclusive evidence that autoantibodies play an important pathogenic role in GBS. Although anti-ganglioside antibodies have been studied most intensively, other antigenic targets may also be important. One report identified IgG antibodies against Schwann cells and neurons (nerve growth cone region) in some GBS cases. Proof that these antibodies are pathogenic requires that they be capable of mediating disease following direct passive transfer to naïve hosts; this has not yet been demonstrated, although one case of possible maternal-fetal transplacental transfer of GBS has been described.

In AIDP, an early step in the induction of tissue damage appears to be complement deposition along the outer surface of the Schwann cell. Activation of complement initiates a characteristic vesicular disintegration of the myelin sheath, and also leads to recruitment of activated macrophages, which participate in damage to myelin and axons. In AMAN, the pattern is different in that complement is

deposited along with IgG at the nodes of Ranvier along large motor axons. Interestingly, in cases of AMAN antibodies against GD1a appear to have a fine specificity that favors binding to motor rather than sensory nerve roots, even though this ganglioside is expressed on both fiber types.

PATHOPHYSIOLOGY

In the demyelinating forms of GBS, the basis for flaccid paralysis and sensory disturbance is conduction block. This finding, demonstrable electrophysiologically, implies that the axonal connections remain intact. Hence, recovery can take place rapidly as remyelination occurs. In severe cases of demyelinating GBS, secondary axonal degeneration usually occurs; its extent can be estimated electrophysiologically. More secondary axonal degeneration correlates with a slower rate of recovery and a greater degree of residual disability. When a severe primary axonal pattern is encountered electrophysiologically, the implication is that axons have degenerated and become disconnected from their targets, specifically the neuromuscular junctions, and must therefore regenerate for recovery to take place. In motor axonal cases in which recovery is rapid, the lesion is thought to be localized to preterminal motor branches, allowing regeneration and reinnervation to take place quickly. Alternatively, in mild cases, collateral sprouting and reinnervation from surviving motor axons near the neuromuscular junction may begin to reestablish physiologic continuity with muscle cells over a period of several months.

LABORATORY FEATURES

CSF findings are distinctive, consisting of an elevated CSF protein level [1–10 g/L (100–1000 mg/dL)] without accompanying pleocytosis. The CSF is often normal when symptoms have been present for ≤ 48 h; by the end of the first week, the level of protein is usually elevated. A transient increase in the CSF white cell count (10–100/μL) occurs on occasion in otherwise typical GBS; however, a sustained CSF pleocytosis suggests an alternative diagnosis (viral myelitis) or a concurrent diagnosis such as unrecognized HIV infection, leukemia or lymphoma with infiltration of nerves, or neurosarcoidosis. Edx features are mild or absent in the early stages of GBS and lag behind the clinical evolution. In AIDP, the earliest features are prolonged F-wave latencies, prolonged distal latencies and reduced amplitudes of compound muscle action potentials (CMAPs), probably owing to the predilection for involvement of nerve roots and distal motor nerve terminals early in the course. Later, slowing of conduction velocity, conduction block, and temporal dispersion may be appreciated (Table 385-1). Occasionally, sensory nerve action potentials (SNAPs) may be normal in the feet (e.g., sural nerve) when abnormal in the arms. This is also a sign that the patient does not have one of the more typical "length-dependent" polyneuropathies. In cases with primary axonal pathology, the principal Edx finding is reduced amplitude of CMAPs (and also SNAPS with AMSAN) without conduction slowing or prolongation of distal latencies.

DIAGNOSIS

GBS is a descriptive entity. The diagnosis of AIDP is made by recognizing the pattern of rapidly evolving paralysis with areflexia, absence of fever or other systemic symptoms, and characteristic antecedent events **(Table 385-3)**. Other disorders that may enter into the differential diagnosis include acute myelopathies (especially with prolonged back pain and sphincter disturbances); diphtheria (early oropharyngeal disturbances); Lyme polyradiculitis and other tick-borne paralyses; porphyria (abdominal pain, seizures, psychosis); vasculitic neuropathy (check erythrocyte sedimentation rate, described below); poliomyelitis (fever and meningismus common); West Nile virus; CMV polyradiculitis (in immunocompromised patients); critical illness neuropathy or myopathy; neuromuscular junction disorders such as myasthenia gravis and botulism (pupillary reactivity lost early); poisonings with organophosphates, thallium, or arsenic; paralytic shellfish poisoning; or severe hypophosphatemia (rare). Laboratory tests are helpful primarily to exclude mimics of GBS. Edx features may be minimal, and the CSF protein level may not rise until the end of the first week. If the diagnosis is strongly suspected, treatment should be initiated without waiting for evolution of the characteristic Edx and CSF findings to occur. Both tau and 14-3-3 protein levels are reported to be elevated early (during the first few days of symptoms) in some cases of GBS. Tau increases in CSF may reflect axonal damage and predict a residual deficit. GBS patients with risk factors for HIV or with CSF pleocytosis should have a serologic test for HIV.

Table 385-3 Diagnostic Features of Acute Inflammatory Demyelinating Polyneuropathy (AIDP)

I. Required for Diagnosis
1. Progressive weakness of variable degree from mild paresis to complete paralysis
2. Generalized hypo- or areflexia
II. Supportive of Diagnosis
1. Clinical Features
a. Symptom progression: Motor weakness rapidly progresses initially but ceases by 4 weeks. Nadir attained by 2 weeks in 50%, 3 weeks 80%, and 90% by 4 weeks.

b. Demonstration of relative limb symmetry regarding paresis.
c. Mild to moderate sensory signs.
d. Frequent cranial nerve involvement: Facial (cranial nerve VII) 50% and typically bilateral but asymmetric; occasional involvement of cranial nerves XII, X, and occasionally III, IV, and VI as well as XI.
e. Recovery typically begins 2–4 weeks following plateau phase.
f. Autonomic dysfunction can include tachycardia, other arrhythmias, postural hypotension, hypertension, other vasomotor symptoms.
g. A preceding gastrointestinal illness (e.g., diarrhea) or upper respiratory tract infection is common.
Cerebrospinal Fluid Features Supporting Diagnosis
a. Elevated or serial elevation of CSF protein.
b. CSF cell counts are <10 mononuclear cell/mm^3.
3. Electrodiagnostic Medicine Findings Supportive of Diagnosis
a. 80% of patients have evidence of NCV slowing/conduction block at some time during disease process.
b. Patchy reduction in NCV attaining values less than 60% of normal.
c. Distal motor latency increase may reach 3 times normal values.
d. F-waves indicate proximal NCV slowing.
e. About 15–20% of patients have normal NCV findings.
f. No abnormalities on nerve conduction studies may be seen for several weeks.
III. Findings Reducing Possibility of Diagnosis
1. Asymmetric weakness
2. Failure of bowel/bladder symptoms to resolve
3. Severe bowel/bladder dysfunction at initiation of disease
4. Greater than 50 mononuclear cells/mm3 in CSF
5. Well-demarcated sensory level
IV. Exclusionary Criteria
1. Diagnosis of other causes of acute neuromuscular weakness (e.g., myasthenia gravis, botulism, poliomyelitis, toxic neuropathy).
3. Abnormal CSF cytology suggesting carcinomatous invasion of the nerve roots

Abbreviations: CSF, cerebrospinal fluid; NCV, nerve conduction velocity. ***Source:*** AA Amato, D Dumitru, in D Dumitru et al (eds): Electrodiagnostic Medicine, 2nd ed, Philadelphia, Hanley & Belfus, 2002.

Treatment: Guillain-Barré Syndrome

In the vast majority of patients with GBS, treatment should be initiated as soon after diagnosis as possible. Each day counts; ~2 weeks after the first motor symptoms, it is not known whether immunotherapy is still effective. If the patient has already reached the plateau stage, then treatment probably is no longer indicated, unless the patient has severe motor weakness and one cannot exclude the possibility that an immunologic attack is still ongoing. Either high-dose intravenous immune globulin (IVIg) or plasmapheresis can be initiated, as they are equally effective for typical GBS. A combination of the two therapies is not significantly better than either alone. IVIg is often the initial therapy chosen because of its ease of administration and good safety record. Anecdotal data has also suggested that IVIg may be preferable to PE for the AMAN and MFS variants of GBS. IVIg is administered as five daily infusions for a total dose of 2 g/kg body weight. There is some evidence that GBS autoantibodies are neutralized by anti-idiotypic antibodies present in IVIg preparations, perhaps accounting for the therapeutic effect. A course of plasmapheresis usually consists of ~40–50 mL/kg plasma exchange (PE) four to five times over a week. Meta-analysis of randomized clinical trials indicates that treatment reduces the need for mechanical ventilation by nearly half (from 27% to 14% with PE) and increases the likelihood of full recovery at 1 year (from 55% to 68%). Functionally significant improvement may occur toward the end of the first week of treatment, or may be delayed for several weeks. The lack of noticeable improvement following a course of IVIg or PE is not an indication to treat with the alternate treatment. However, there are occasional patients who are treated early in the course of GBS and improve, who then relapse within a month. Brief retreatment with the original therapy is usually effective in such cases. Glucocorticoids have not been found to be effective in GBS. Occasional patients with very mild forms of GBS, especially those who appear to have already reached a plateau when initially seen, may be managed conservatively without IVIg or PE.

In the worsening phase of GBS, most patients require monitoring in a critical care setting, with particular attention to vital capacity, heart rhythm, blood pressure, nutrition, deep vein thrombosis

prophylaxis, cardiovascular status, early consideration (after 2 weeks of intubation) of tracheotomy, and chest physiotherapy. As noted, ~30% of patients with GBS require ventilatory assistance, sometimes for prolonged periods of time (several weeks or longer). Frequent turning and assiduous skin care are important, as are daily range-of-motion exercises to avoid joint contractures and daily reassurance as to the generally good outlook for recovery.

PROGNOSIS AND RECOVERY

Approximately 85% of patients with GBS achieve a full functional recovery within several months to a year, although minor findings on examination (such as areflexia) may persist and patients often complain of continued symptoms, including fatigue. The mortality rate is <5% in optimal settings; death usually results from secondary pulmonary complications. The outlook is worst in patients with severe proximal motor and sensory axonal damage. Such axonal damage may be either primary or secondary in nature (see "Pathophysiology," above), but in either case successful regeneration cannot occur. Other factors that worsen the outlook for recovery are advanced age, a fulminant or severe attack, and a delay in the onset of treatment. Between 5 and 10% of patients with typical GBS have one or more late relapses; such cases are then classified as chronic inflammatory demyelinating polyneuropathy (CIDP).

CHRONIC INFLAMMATORY DEMYELINATING POLYNEUROPATHY

CIDP is distinguished from GBS by its chronic course. In other respects, this neuropathy shares many features with the common demyelinating form of GBS, including elevated CSF protein levels and the Edx findings of acquired demyelination. Most cases occur in adults, and males are affected slightly more often than females. The incidence of CIDP is lower than that of GBS, but due to the protracted course the prevalence is greater.

CLINICAL MANIFESTATIONS

Onset is usually gradual over a few months or longer, but in a few cases the initial attack is indistinguishable from that of GBS. An acute-onset form of CIDP should be considered when GBS deteriorates >9 weeks after onset or relapses at least three times. Symptoms are both motor and sensory in most cases. Weakness of the limbs is usually symmetric but can be strikingly asymmetric in multifocal acquired demyelinating sensory and motor (MADSAM) neuropathy variant (Lewis-Sumner syndrome) in which discrete peripheral nerves are involved. There is considerable variability from case to case. Some patients experience a chronic progressive course, whereas others, usually younger patients, have a relapsing and remitting course. Some have only motor findings, and a small proportion present with a relatively pure syndrome of sensory ataxia. Tremor occurs in ~10% and may become more prominent during periods of subacute worsening or improvement. A small proportion have cranial nerve findings, including external ophthalmoplegia. CIDP tends to ameliorate over time with treatment; the result is that many years after onset, nearly 75% of patients have reasonable functional status. Death from CIDP is uncommon.

DIAGNOSIS

The diagnosis rests on characteristic clinical, CSF, and electrophysiologic findings. The CSF is usually acellular with an elevated protein level, sometimes several times normal. As with GBS, a CSF pleocytosis should lead to the consideration of HIV infection, leukemia or lymphoma, and neurosarcoidosis. Edx findings reveal variable degrees of conduction slowing, prolonged distal latencies, distal and temporal dispersion of CMAPs, and conduction block as the principal features. In particular, the presence of conduction block is a certain sign of an acquired demyelinating process. Evidence of axonal loss, presumably secondary to demyelination, is present in >50% of patients. Serum protein electrophoresis with immunofixation is indicated to search for monoclonal gammopathy and associated conditions (see "Monoclonal Gammopathy of Undetermined Significance," below). In all patients with presumptive CIDP, it is also reasonable to exclude vasculitis, collagen vascular disease (especially SLE), chronic hepatitis, HIV infection, amyloidosis, and diabetes mellitus. Other associated conditions include inflammatory bowel disease and lymphoma.

PATHOGENESIS

Although there is evidence of immune activation in CIDP, the precise mechanisms of pathogenesis are unknown. Biopsy typically reveals little inflammation and onion-bulb changes (imbricated layers of attenuated Schwann cell processes surrounding an axon) that result from recurrent demyelination and remyelination (Fig. 385-1). The response to therapy suggests that CIDP is immune-mediated; CIDP responds to glucocorticoids, whereas GBS does not. Passive transfer of demyelination into experimental animals has been accomplished using IgG purified from the serum of some patients with CIDP, lending support for a humoral autoimmune pathogenesis. Although the target antigen or antigens in CIDP have not yet been identified, the myelin protein Po has been implicated as a potential autoantigen in some patients. It is also of interest that a CIDP-like illness developed spontaneously in the nonobese diabetic (NOD) mouse when the immune co-stimulatory molecule B7-2 (CD86) was genetically deleted; this suggests that CIDP can result from altered triggering of T cells by antigen-presenting cells.

Approximately 25% of patients with clinical features of CIDP also have a monoclonal gammopathy of undetermined significance (MGUS). Cases associated with monoclonal IgA or IgG kappa usually respond to treatment as favorably as cases without a monoclonal gammopathy. Patients with IgM monoclonal gammopathy tend to have more sensory findings and a more protracted course, and usually have a less satisfactory response to treatment.

Treatment: Chronic Inflammatory Demyelinating Polyneuropathy

Most authorities initiate treatment for CIDP when progression is rapid or walking is compromised. If the disorder is mild, management can be expectant, awaiting spontaneous remission. Controlled studies have shown that high-dose IVIg, PE, and glucocorticoids are all more effective than placebo. Initial therapy is usually with IVIg, administered as 2.0 g/kg body weight given in divided doses over 2–5 days; three monthly courses are generally recommended before concluding a patient is a treatment failure. If the patient responds, the infusion intervals can be gradually increased or the dosage decreased (e.g., 1 g/kg per month). PE, which appears to be as effective as IVIg, is initiated at two to three treatments per week for 6 weeks; periodic re-treatment may also be required. Treatment with glucocorticoids is another option (60–80 mg prednisone PO daily for 1–2 months, followed by a gradual dose reduction of 10 mg per month as tolerated), but long-term adverse effects including bone demineralization, gastrointestinal bleeding, and cushingoid changes are problematic. As many as one-third of patients with CIDP fail to respond adequately to the initial therapy chosen; a different treatment should then be tried. Patients who fail therapy with IVIg, PE, and glucocorticoids may benefit from treatment with immunosuppressive agents such as azathioprine, methotrexate, cyclosporine, and cyclophosphamide, either alone or as adjunctive therapy. Early experience with anti-CD20 (rituximab) has also shown promise. Use of these therapies requires periodic reassessment of their risks and benefits. In patients with a CIDP-like neuropathy who fail to respond to treatment it is important to evaluate for POEMS syndrome (polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, skin changes; see below).

MULTIFOCAL MOTOR NEUROPATHY

Multifocal motor neuropathy (MMN) is a distinctive but uncommon neuropathy that presents as slowly progressive motor weakness and atrophy evolving over years in the distribution of selected nerve trunks, associated with sites of persistent focal motor conduction block in the same nerve trunks. Sensory fibers are relatively spared. The arms are affected more frequently than the legs, and >75% of all patients are male. Some cases have been confused with lower motor neuron forms of amyotrophic lateral sclerosis (Chap. 374). Less than 50% of patients present with high titers of polyclonal IgM antibody to the ganglioside GM1. It is uncertain how this finding relates to the discrete foci of persistent motor conduction block, but high concentrations of GM1 gangliosides are normal constituents of nodes of Ranvier in peripheral nerve fibers. Pathology reveals demyelination and mild inflammatory changes at the sites of conduction block.

Most patients with MMN respond to high-dose IVIg (dosages as for CIDP, above); periodic re-treatment is required (usually at least monthly) to maintain the benefit. Some refractory patients have responded to rituximab or cyclophosphamide. Glucocorticoids and PE are not effective.

NEUROPATHIES WITH MONOCLONAL GAMMOPATHY

Multiple Myeloma

Clinically overt polyneuropathy occurs in ~5% of patients with the commonly encountered type of multiple myeloma, which exhibits either lytic or diffuse osteoporotic bone lesions. These neuropathies are sensorimotor, are usually mild and slowly progressive but may be severe, and generally do not reverse with successful suppression of the myeloma. In most cases, EDX and pathologic features are consistent with a process of axonal degeneration.

In contrast, myeloma with osteosclerotic features, although representing only 3% of all myelomas, is associated with polyneuropathy in one-half of cases. These neuropathies, which may also occur with solitary plasmacytoma, are distinct because they (1) are usually demyelinating in nature and resemble CIDP; (2) often respond to radiation therapy or removal of the primary lesion; (3) are associated with different monoclonal proteins and light chains (almost always lambda as opposed to primarily kappa in the lytic type of multiple myeloma); (4) are typically refractory to standard treatments of CIDP; and (5) may occur in association with other systemic findings including thickening of the skin, hyperpigmentation, hypertrichosis, organomegaly, endocrinopathy, anasarca, and clubbing of fingers. These are features of the POEMS syndrome (*p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *M* protein, and *s*kin changes). Levels of vascular endothelial growth factor (VEGF) are increased in the serum and this factor is felt to somehow play a pathogenic role in this syndrome. Treatment of the neuropathy is best directed at the osteosclerotic myeloma using surgery, radiotherapy, chemotherapy, or autologous peripheral blood stem cell transplantation.

Neuropathies are also encountered in other systemic conditions with gammopathy, including Waldenström's macroglobulinemia, primary systemic amyloidosis, and cryoglobulinemic states (mixed essential cryoglobulinemia, some cases of hepatitis C).

Monoclonal Gammopathy of Undetermined Significance

Chronic polyneuropathies occurring in association with MGUS are usually associated with the immunoglobulin isotypes IgG, IgA, and IgM. Most patients present with isolated sensory symptoms in their distal extremities and have Edx features of an axonal sensory or sensorimotor polyneuropathy. These patients otherwise resemble idiopathic sensory polyneuropathy and the MGUS might just be coincidental. They usually do not respond to immunotherapies designed to reduce the concentration of the monoclonal protein. Some patients, however, present with generalized weakness and sensory loss and Edx studies indistinguishable from CIDP without monoclonal gammopathy (see "Chronic Inflammatory Demyelinating Polyneuropathy," above), and their response to immunosuppressive agents is also similar. An exception is the syndrome of IgM kappa monoclonal gammopathy associated with an indolent, longstanding, sometimes static sensory neuropathy, frequently with tremor and sensory ataxia. Most patients are male and older than age 50 years. In the majority, the monoclonal IgM immunoglobulin binds to a normal peripheral nerve constituent, myelin-associated glycoprotein (MAG), found in the paranodal regions of Schwann cells. Binding appears to be specific for a polysaccharide epitope that is also found in other normal peripheral nerve myelin glycoproteins, P0 and PMP22, and also in other normal nerve-related glycosphingolipids (Fig. 385-1). In the MAG-positive cases, IgM paraprotein is incorporated into the myelin sheaths of affected patients and widens the spacing of the myelin lamellae, thus producing a distinctive ultrastructural pattern. Demyelination and remyelination are the hallmarks of the lesions. The chronic demyelinating neuropathy appears to result from a destabilization of myelin metabolism rather than activation of an immune response. Therapy with chlorambucil, or cyclophosphamide combined with glucocorticoids or PE, often results in improvement of the neuropathy associated with a prolonged reduction in the levels in the circulating paraprotein; chronic use of these alkylating agents is associated with significant risks. In a small proportion of patients (30% at 10 years), MGUS will in time evolve into frankly malignant conditions such as multiple myeloma or lymphoma.

VASCULITIC NEUROPATHY

Peripheral nerve involvement is common in polyarteritis nodosa (PAN), appearing in half of all cases clinically and in 100% of cases at postmortem studies (Chap. 326). The most common pattern is multifocal (asymmetric) motor-sensory neuropathy (mononeuropathy multiplex) due to ischemic lesions of nerve trunks and roots; however, some cases of vasculitic neuropathy present as a distal, symmetric sensorimotor polyneuropathy. Symptoms of neuropathy are a common presenting complaint in patients with PAN. The Edx findings are those of an axonal process. Small- to medium-sized arteries of the vasa nervorum, particularly the epineural vessels, are affected in PAN, resulting in a widespread ischemic neuropathy. A high frequency of neuropathy occurs in allergic angiitis and granulomatosis (Churg-Strauss syndrome).

Systemic vasculitis should always be considered when a subacute or chronically evolving mononeuropathy multiplex occurs in conjunction with constitutional symptoms (fever, anorexia, weight loss, loss of energy, malaise, and nonspecific pains). Diagnosis of suspected vasculitic neuropathy is made by a combined nerve and muscle biopsy, with serial section or skip-serial techniques.

Approximately one-third of biopsy-proven cases of vasculitic neuropathy are "nonsystemic" in that the vasculitis appears to affect only peripheral nerves. Constitutional symptoms are absent, and the course is more indolent than that of PAN. The erythrocyte sedimentation rate may be elevated, but other tests for systemic disease are negative. Nevertheless, clinically silent involvement of other organs is likely, and vasculitis is frequently found in muscle biopsied at the same time as nerve.

Vasculitic neuropathy may also be seen as part of the vasculitis syndrome occurring in the course of other connective tissue disorders (Chap. 326). The most frequent is rheumatoid arthritis, but ischemic neuropathy due to involvement of vasa nervorum may also occur in mixed cryoglobulinemia, Sjögren's syndrome, granulomatosis with polyangiitis (Wegener's), hypersensitivity angiitis, systemic lupus erythematosus, and progressive systemic sclerosis. Management of these neuropathies, including the "nonsystemic" vasculitic neuropathy, consists of treatment of the underlying condition as well as the aggressive use of glucocorticoids and other immunosuppressant drugs. Use of these regimens has resulted in dramatic improvements in outcome, with 5-year survival rates now greater than 80%. One reasonable starting regimen is daily prednisone (initial dose 1 mg/kg per day PO with a gradual taper after 1 month) plus IV pulse (or daily oral) cyclophosphamide for 3–6 months.

ANTI-HU PARANEOPLASTIC NEUROPATHY

This uncommon immune-mediated disorder manifests as a sensory neuronopathy (i.e., selective damage to sensory nerve bodies in dorsal root ganglia). The onset is often asymmetric with dysesthesias and sensory loss in the limbs that soon progress to affect all limbs, the torso, and face. Marked sensory ataxia, pseudoathetosis, and inability to walk, stand, or even sit unsupported are frequent features and are secondary to the extensive deafferentation. Subacute sensory neuronopathy may be idiopathic, but more than half of cases are paraneoplastic, primarily related to lung cancer, and most of those are small cell lung cancer (SCLC). Diagnosis of the underlying SCLC requires

awareness of the association, paraneoplastic testing, and often PET scanning for the tumor. The target antigens are a family of RNA-binding proteins (HuD, HuC, and Hel-N1) that in normal tissues are only expressed by neurons. The same proteins are usually expressed by SCLC, triggering in some patients an immune response characterized by antibodies and cytotoxic T cells that cross-react with the Hu proteins of the dorsal root ganglion neurons, resulting in immune-mediated neuronal destruction. An encephalomyelitis may accompany the sensory neuronopathy and presumably has the same pathogenesis. Neurologic symptoms usually precede, by ≤6 months, the identification of SCLC. The sensory neuronopathy runs its course in a few weeks or months and stabilizes, leaving the patient disabled. Most cases are unresponsive to treatment with glucocorticoids, IVIg, PE, or immunosuppressant drugs.

FURTHER READINGS

Burns TM et al: Vasculitic neuropathies. Neurol Clin 25:89, 2007[PMID: 17324722]

Cats EA et al: Correlates of outcome and response to IVIg in 88 patients with multifocal motor neuropathy. Neurology 75:818, 2010[PMID: 20805527]

Hadden RDM et al: European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of paraproteinemic demyelinating neuropathies: Report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society. Eur J Neurol 13:809, 2006[PMID: 16879290]

Hughes RA et al: European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of chronic inflammatory demyelinating polyradiculoneuropathy: Report of a joint task force of the European Federation of Neurological Societies and the Peripheral Nerve Society. Eur J Neurol 13:326, 2006[PMID: 16643309]

——— et al: Immunotherapy for Guillain-Barrè syndrome: A systematic review. Brain 130:2245, 2007

Kuijf ML et al: TLR4-mediated sensing of *Campylobacter jejuni* by dendritic cells is determined by sialylation. J Immunol 185:748, 2010[PMID: 20525894]

Latov N et al: Timing and course of clinical response to intravenous immunoglobulin in chronic inflammatory demyelinating polyradiculoneuropathy. Arch Neurol 67:802, 2010[PMID: 20457948]

Lopez PHH et al: Structural requirements of anti-GD1a antibodies determine their target specificity. Brain 131:1926, 2008[PMID: 18487279]

Lunn MP, Willison HJ: Diagnosis and treatment in inflammatory neuropathies. J Neurol Neurosurg Psychiatry 80:249, 2009[PMID: 19228670]

Mathew L et al: Treatment of vasculitic peripheral neuropathy: A retrospective analysis of outcome. QJM 100:41, 2007[PMID: 17189246]

Susuki K et al: Anti-GM1 antibodies cause complement-mediated disruption of sodium channel clusters in peripheral motor nerve fibers. J Neurosci 27:3956, 2007[PMID: 17428969]

Van Schaik IN et al: European Federation of Neurological Societies/Peripheral Nerve Society guideline on management of multifocal motor neuropathy. Eur J Neurol 13:802, 2006

Walgaard C et al: Prediction of respiratory insufficiency in Guillain-Barré syndrome. Ann Neurol 67:781, 2010[PMID: 20517939]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MYASTHENIA GRAVIS AND OTHER DISEASES OF THE NEUROMUSCULAR JUNCTION: INTRODUCTION

Myasthenia gravis (MG) is a neuromuscular disorder characterized by weakness and fatigability of skeletal muscles. The underlying defect is a decrease in the number of available acetylcholine receptors (AChRs) at neuromuscular junctions due to an antibody-mediated autoimmune attack. Treatment now available for MG is highly effective, although a specific cure has remained elusive.

Pathophysiology

At the neuromuscular junction **(Fig. 386-1)**, acetylcholine (ACh) is synthesized in the motor nerve terminal and stored in vesicles (quanta). When an action potential travels down a motor nerve and reaches the nerve terminal, ACh from 150 to 200 vesicles is released and combines with AChRs that are densely packed at the peaks of postsynaptic folds. The structure of the AChR has been fully elucidated; it consists of five subunits (2α, 1β, 1δ, and 1γ or ε) arranged around a central pore. When ACh combines with the binding sites on the α subunits of the AChR, the channel in the AChR opens, permitting the rapid entry of cations, chiefly sodium, which produces depolarization at the end-plate region of the muscle fiber. If the depolarization is sufficiently large, it initiates an action potential that is propagated along the muscle fiber, triggering muscle contraction. This process is rapidly terminated by hydrolysis of ACh by acetylcholinesterase (AChE), which is present within the synaptic folds, and by diffusion of ACh away from the receptor.

Figure 386-1

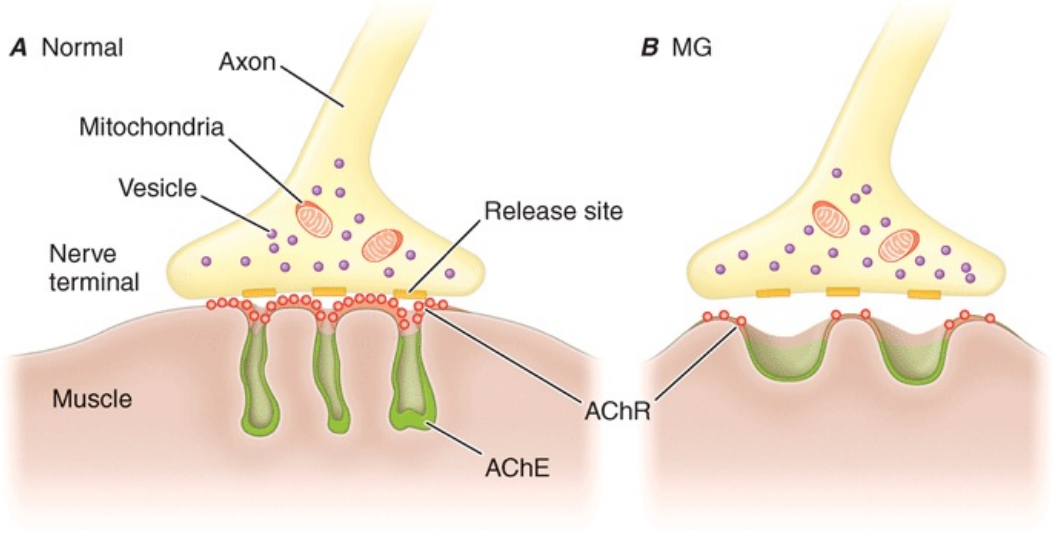

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagrams of (*A*) normal and (*B*) myasthenic neuromuscular junctions. AChE, acetylcholinesterase. See text for description of normal neuromuscular transmission. The MG junction demonstrates a normal nerve terminal; a reduced number of AChRs (stippling); flattened, simplified postsynaptic folds; and a widened synaptic space. *(Modified from DB Drachman: N Engl J Med 330:1797, 1994; with permission.)*

In MG, the fundamental defect is a decrease in the number of available AChRs at the postsynaptic muscle membrane. In addition, the postsynaptic folds are flattened, or "simplified." These changes result in decreased efficiency of neuromuscular transmission. Therefore, although ACh is released normally, it produces small end-plate potentials that may fail to trigger muscle action potentials. Failure of transmission at many neuromuscular junctions results in weakness of muscle contraction.

The amount of ACh released per impulse normally declines on repeated activity (termed *presynaptic*

rundown). In the myasthenic patient, the decreased efficiency of neuromuscular transmission combined with the normal rundown results in the activation of fewer and fewer muscle fibers by successive nerve impulses and hence increasing weakness, or *myasthenic fatigue*. This mechanism also accounts for the decremental response to repetitive nerve stimulation seen on electrodiagnostic testing.

The neuromuscular abnormalities in MG are brought about by an autoimmune response mediated by specific anti-AChR antibodies. The anti-AChR antibodies reduce the number of available AChRs at neuromuscular junctions by three distinct mechanisms: (1) accelerated turnover of AChRs by a mechanism involving cross-linking and rapid endocytosis of the receptors; (2) damage to the postsynaptic muscle membrane by the antibody in collaboration with complement; and (3) blockade of the active site of the AChR, i.e., the site that normally binds ACh. An immune response to muscle-specific kinase (MuSK), a protein involved in AChR clustering at neuromuscular junctions, can also result in myasthenia gravis, with reduction of AChRs demonstrated experimentally. The pathogenic antibodies are IgG, and are T cell-dependent. Thus, immunotherapeutic strategies directed against either the antibody-producing B cells or helper T cells are effective in this antibody-mediated disease.

How the autoimmune response is initiated and maintained in MG is not completely understood, but the thymus appears to play a role in this process. The thymus is abnormal in ~75% of patients with MG; in ~65% the thymus is "hyperplastic," with the presence of active germinal centers detected histologically, though the hyperplastic thymus is not necessarily enlarged. An additional 10% of patients have thymic tumors (thymomas). Muscle-like cells within the thymus (myoid cells), which bear AChRs on their surface, may serve as a source of autoantigen and trigger the autoimmune reaction within the thymus gland.

Clinical Features

MG is not rare, having a prevalence of 2–7 in 10,000. It affects individuals in all age groups, but peaks of incidence occur in women in their twenties and thirties and in men in their fifties and sixties. Overall, women are affected more frequently than men, in a ratio of ~3:2. The cardinal features are *weakness* and *fatigability* of muscles. The weakness increases during repeated use (fatigue) or late in the day, and may improve following rest or sleep. The course of MG is often variable. Exacerbations and remissions may occur, particularly during the first few years after the onset of the disease. Remissions are rarely complete or permanent. Unrelated infections or systemic disorders can lead to increased myasthenic weakness and may precipitate "crisis" (see below).

The distribution of muscle weakness often has a characteristic pattern. The cranial muscles, particularly the lids and extraocular muscles, are typically involved early in the course of MG; diplopia and ptosis are common initial complaints. Facial weakness produces a "snarling" expression when the patient attempts to smile. Weakness in chewing is most noticeable after prolonged effort, as in chewing meat. Speech may have a nasal timbre caused by weakness of the palate, or a dysarthric "mushy" quality due to tongue weakness. Difficulty in swallowing may occur as a result of weakness of the palate, tongue, or pharynx, giving rise to nasal regurgitation or aspiration of liquids or food. Bulbar weakness is especially prominent in MuSK antibody–positive MG. In ~85% of patients, the weakness becomes generalized, affecting the limb muscles as well. If weakness remains restricted to the extraocular muscles for 3 years, it is likely that it will not become generalized, and these patients are said to have *ocular MG*. The limb weakness in MG is often proximal and may be asymmetric. Despite the muscle weakness, deep tendon reflexes are preserved. If weakness of respiration becomes so severe as to require respiratory assistance, the patient is said to be in *crisis*.

DIAGNOSIS AND EVALUATION

(Table 386-1) The diagnosis is suspected on the basis of weakness and fatigability in the typical distribution described above, without loss of reflexes or impairment of sensation or other neurologic function. The suspected diagnosis should always be confirmed definitively before treatment is undertaken; this is essential because (1) other treatable conditions may closely resemble MG and (2) the treatment of MG may involve surgery and the prolonged use of drugs with potentially adverse side effects.

Table 386-1 Diagnosis of Myasthenia Gravis (mg)

History Diplopia, ptosis, weakness
Weakness in characteristic distribution
Fluctuation and fatigue: worse with repeated activity, improved by rest
Effects of previous treatments
Physical examination
Ptosis, diplopia

Motor power survey: quantitative testing of muscle strength
Forward arm abduction time (5 min)
Vital capacity
Absence of other neurologic signs
Laboratory testing
Anti-AChR radioimmunoassay: ~85% positive in generalized MG; 50% in ocular MG; definite diagnosis if positive; negative result does not exclude MG. ~40% of AChR antibody–negative patients with generalized MG have anti-MuSK antibodies.
Repetitive nerve stimulation: decrement of >15% at 3 Hz: highly probable
Single-fiber electromyography: blocking and jitter, with normal fiber density; confirmatory, but not specific
Edrophonium chloride (Tensilon) 2 mg + 8 mg IV; highly probable diagnosis if unequivocally positive
For ocular or cranial MG: exclude intracranial lesions by CT or MRI

Abbreviations: AChR, acetylcholine receptor; MuSK, muscle-specific tyrosine kinase. ***Source:*** From RT Johnson, JW Griffin (eds): Current Therapy in Neurologic Disease, 4th ed. St. Louis, Mosby Year Book, 1994; with permission.

ANTIBODIES TO ACHR OR MUSK

As noted above, anti-AChR antibodies are detectable in the serum of ~85% of all myasthenic patients but in only about 50% of patients with weakness confined to the ocular muscles. The presence of anti-AChR antibodies is virtually diagnostic of MG, but a negative test does not exclude the disease. The measured level of anti-AChR antibody does not correspond well with the severity of MG in different patients. However, in an individual patient, a treatment-induced fall in the antibody level often correlates with clinical improvement, while a rise in the level may occur with exacerbations. Antibodies to MuSK have been found to be present in ~40% of AChR antibody–negative patients with generalized MG, and their presence is a useful diagnostic test in these patients. MuSK antibodies are rarely present in AChR antibody–positive patients or in patients with MG limited to ocular muscles. These antibodies may interfere with clustering of AChRs at neuromuscular junctions, as MuSK is known to do during early development. There is also evidence that MG patients without antibodies demonstrable by standard AChR or MuSK tests may have either low-affinity antibodies, or other—as yet undefined—antibodies that impair neuromuscular transmission.

ELECTRODIAGNOSTIC TESTING

Repetitive nerve stimulation may provide helpful diagnostic evidence of MG. Anti-AChE medication is stopped 6–24 h before testing. It is best to test weak muscles or proximal muscle groups. Electric shocks are delivered at a rate of two or three per second to the appropriate nerves, and action potentials are recorded from the muscles. In normal individuals, the amplitude of the evoked muscle action potentials does not change at these rates of stimulation. However, in myasthenic patients there is a rapid reduction of >10–15% in the amplitude of the evoked responses.

ANTICHOLINESTERASE TEST

Drugs that inhibit the enzyme AChE allow ACh to interact repeatedly with the limited number of AChRs in MG, producing improvement in muscle strength. Edrophonium is used most commonly for diagnostic testing because of the rapid onset (30 s) and short duration (~5 min) of its effect. An objective end-point must be selected to evaluate the effect of edrophonium, such as weakness of extraocular muscles, impairment of speech, or the length of time that the patient can maintain the arms in forward abduction. An initial IV dose of 2 mg of edrophonium is given. If definite improvement occurs, the test is considered positive and is terminated. If there is no change, the patient is given an additional 8 mg IV. The dose is administered in two parts because some patients react to edrophonium with side effects such as nausea, diarrhea, salivation, fasciculations, and rarely with severe symptoms of syncope or bradycardia. Atropine (0.6 mg) should be drawn up in a syringe, ready for IV administration if these symptoms become troublesome. The edrophonium test is now reserved for patients with clinical findings that are suggestive of MG but who have negative antibody and electrodiagnostic test results. False-positive tests occur in occasional patients with other neurologic disorders, such as amyotrophic lateral sclerosis, and in placebo-reactors. False-negative or equivocal tests may also occur. In some cases, it is helpful to use a longer-acting drug such as neostigmine (15 mg PO), since this permits more time for detailed evaluation of strength.

INHERITED MYASTHENIC SYNDROMES

The congenital myasthenic syndromes (CMS) comprise a heterogeneous group of disorders of the neuromuscular junction that are not autoimmune but rather are due to genetic mutations in which virtually any component of the neuromuscular junction may be affected. Alterations in function of the presynaptic nerve terminal or in the various subunits of the AChR or AChE have been identified in the different forms of CMS. These disorders share many of the clinical features of autoimmune MG, including weakness and fatigability of skeletal muscles, in some cases involving extraocular muscles

(EOMs), lids, and proximal muscles, similar to the distribution in autoimmune MG. CMS should be suspected when symptoms of myasthenia have begun in infancy or childhood and AChR antibody tests are consistently negative. Features of four of the most common forms of CMS are summarized in **Table 386-2**. Although clinical features and electrodiagnostic and pharmacologic tests may suggest the correct diagnosis, molecular analysis is required for precise elucidation of the defect; this may lead to helpful treatment as well as genetic counseling. In the forms that involve the AChR, a wide variety of mutations have been identified in each of the subunits, but the ε subunit is affected in ~75% of these cases. In most of the recessively inherited forms of CMS, the mutations are heteroallelic; that is, different mutations affecting each of the two alleles are present.

Table 386-2 The Congenital Myasthenic Syndromes

Type	Clinical Features	Electrophysiology	Genetics	End-Plate Effects	Treatment
Slow channel	Most common; weak forearm extensors; onset 2d to 3d decade; variable severity	Repetitive muscle response on nerve stimulation; prolonged channel opening and MEPP duration	Autosomal dominant; α, β, ε AChR mutations	Excitotoxic end-plate myopathy; decreased AChRs; postsynaptic damage	Quinidine: decreases end-plate damage; made worse by anti-AChE
Low-affinity fast channel	Onset early; moderately severe; ptosis, EOM involvement; weakness and fatigue	Brief and infrequent channel openings; opposite of slow channel syndrome	Autosomal recessive; may be heteroallelic	Normal end-plate structure	3,4-DAP; anti-AChE
Severe AChR deficiencies	Early onset; variable severity; fatigue; typical MG features	Decremental response to repetitive nerve stimulation; decreased MEPP amplitudes	Autosomal recessive; ε mutations most common; many different mutations	Increased length of end plates; variable synaptic folds	Anti-AChE; ? 3,4-DAP
AChE deficiency	Early onset; variable severity; scoliosis; may have normal EOM, absent pupillary responses	Decremental response to repetitive nerve stimulation	Mutant gene for AChE's collagen anchor	Small nerve terminals; degenerated junctional folds	Worse with anti-AChE drugs

Abbreviations: AChR, acetylcholine receptor; AChE, acetylcholinesterase; EOM, extraocular muscles; MEPP, miniature end-plate potentials; 3,4-DAP, 3,4-diaminopyridine.

DIFFERENTIAL DIAGNOSIS

Other conditions that cause weakness of the cranial and/or somatic musculature include the nonautoimmune CMS discussed above, drug-induced myasthenia, Lambert-Eaton myasthenic syndrome (LEMS), neurasthenia, hyperthyroidism, botulism, intracranial mass lesions, and progressive external ophthalmoplegia. Treatment with penicillamine (used for scleroderma or rheumatoid arthritis) may result in true autoimmune MG, but the weakness is usually mild, and recovery occurs within weeks or months after discontinuing its use. Aminoglycoside antibiotics or procainamide can cause exacerbation of weakness in myasthenic patients; very large doses can cause neuromuscular weakness in normal individuals.

LEMS is a presynaptic disorder of the neuromuscular junction that can cause weakness similar to that of MG. The proximal muscles of the lower limbs are most commonly affected, but other muscles may be involved as well. Cranial nerve findings, including ptosis of the eyelids and diplopia, occur in up to 70% of patients and resemble features of MG. However, the two conditions are usually readily distinguished, since patients with LEMS have depressed or absent reflexes and experience autonomic changes such as dry mouth and impotence. Nerve stimulation produces an initial low-amplitude response and, at low rates of repetitive stimulation (2–3 Hz), decremental responses like those of MG; however, at high rates (50 Hz), or following exercise, incremental responses occur. LEMS is caused by autoantibodies directed against P/Q-type calcium channels at the motor nerve terminals, which can be detected in ~85% of LEMS patients by radioimmunoassay. These autoantibodies result in impaired release of ACh from nerve terminals. Most patients with LEMS have an associated malignancy, most commonly small cell carcinoma of the lung, which may express calcium channels that stimulate the autoimmune response. The diagnosis of LEMS may signal the presence of a tumor long before it would otherwise be detected, permitting early removal. Treatment of LEMS involves plasmapheresis and immunosuppression, as for MG. 3,4-Diaminopyridine (3,4-DAP) and pyridostigmine may also be symptomatically helpful. 3,4-DAP acts by blocking potassium channels, which results in prolonged

depolarization of the motor nerve terminals and thus enhances ACh release. Pyridostigmine prolongs the action of ACh, allowing repeated interactions with AChRs.

Botulism (Chap. 141) is due to potent bacterial toxins produced by any of seven different strains of *Clostridium botulinum*. The toxins enzymatically cleave specific proteins essential for the release of acetylcholine from the motor nerve terminal, thereby interfering with neuromuscular transmission. Most commonly, botulism is caused by ingestion of improperly prepared food containing toxin. Rarely, the nearly ubiquitous spores of *C. botulinum* may germinate in wounds. In infants the spores may germinate in the GI tract, and release toxin, causing muscle weakness. Patients present with myasthenia-like bulbar weakness (e.g., diplopia, dysarthria, dysphagia) and lack sensory symptoms and signs. Weakness may generalize to the limbs and may result in respiratory failure. Reflexes are present early, but they may be diminished as the disease progresses. Mentation is normal. Autonomic findings include paralytic ileus, constipation, urinary retention, dilated or poorly reactive pupils, and dry mouth. The demonstration of toxin in serum by bioassay is definitive, but the results usually take a relatively long time to be completed and may be negative. Nerve stimulation studies reveal findings of presynaptic neuromuscular blockade with reduced compound muscle action potentials (CMAPs) that increase in amplitude following high-frequency repetitive stimulation. Treatment includes ventilatory support, and aggressive inpatient supportive care (e.g., nutrition, DVT prophylaxis) as needed. Antitoxin should be given as early as possible to be effective. A preventive vaccine is available for laboratory workers or other highly exposed individuals.

Neurasthenia is the historic term for a myasthenia-like fatigue syndrome without an organic basis. These patients may present with subjective symptoms of weakness and fatigue, but muscle testing usually reveals the "give-away weakness" characteristic of nonorganic disorders; the complaint of fatigue in these patients means tiredness or apathy rather than decreasing muscle power on repeated effort. Hyperthyroidism is readily diagnosed or excluded by tests of thyroid function, which should be carried out routinely in patients with suspected MG. Abnormalities of thyroid function (hyper- or hypothyroidism) may increase myasthenic weakness. Diplopia resembling that in MG may occasionally be due to an intracranial mass lesion that compresses nerves to the EOMs (e.g., sphenoid ridge meningioma), but MRI of the head and orbits usually reveals the lesion.

Progressive external ophthalmoplegia is a rare condition resulting in weakness of the EOMs, which may be accompanied by weakness of the proximal muscles of the limbs and other systemic features. Most patients with this condition have mitochondrial disorders that can be detected on muscle biopsy (Chap. 387).

SEARCH FOR ASSOCIATED CONDITIONS

(Table 386-3) Myasthenic patients have an increased incidence of several associated disorders. Thymic abnormalities occur in ~75% of patients, as noted above. Neoplastic change (thymoma) may produce enlargement of the thymus, which is detected by CT scanning of the anterior mediastinum. A thymic shadow on CT scan may normally be present through young adulthood, but enlargement of the thymus in a patient aged >40 years is highly suspicious of thymoma. Hyperthyroidism occurs in 3–8% of patients and may aggravate the myasthenic weakness. Thyroid function tests should be obtained in all patients with suspected MG. Because of the association of MG with other autoimmune disorders, blood tests for rheumatoid factor and antinuclear antibodies should also be carried out. Chronic infection of any kind can exacerbate MG and should be sought carefully. Finally, measurements of ventilatory function are valuable because of the frequency and seriousness of respiratory impairment in myasthenic patients.

Table 386-3 Disorders Associated with Myasthenia Gravis and Recommended Laboratory Tests

Associated disorders
Disorders of the thymus: thymoma, hyperplasia
Other autoimmune disorders: Hashimoto's thyroiditis, Graves' disease, rheumatoid arthritis, lupus erythematosus, skin disorders, family history of autoimmune disorder
Disorders or circumstances that may exacerbate myasthenia gravis: hyperthyroidism or hypothyroidism, occult infection, medical treatment for other conditions (see Table 386-4)
Disorders that may interfere with therapy: tuberculosis, diabetes, peptic ulcer, gastrointestinal bleeding, renal disease, hypertension, asthma, osteoporosis, obesity
Recommended laboratory tests or procedures
CT or MRI of mediastinum
Tests for lupus erythematosus, antinuclear antibody, rheumatoid factor, antithyroid antibodies
Thyroid-function tests
PPD skin test
Chest radiography

Fasting blood glucose measurement, hemoglobin A1c
Pulmonary-function tests
Bone densitometry in older patients

Abbreviation: PPD, purified protein derivative. ***Source:*** From RT Johnson, JW Griffin (eds): Current Therapy in Neurologic Disease, 4th ed. St. Louis, Mosby Year Book, 1993, p 379; with permission.

Table 386-4 Drugs with Interactions in Myasthenia Gravis (mg)

Drugs that may exacerbate MG
Antibiotics
Aminoglycosides: e.g., streptomycin, tobramycin, kanamycin
Quinolones: e.g., ciprofloxacin, levofloxacin, ofloxacin, gatifloxacin
Macrolides: e.g., erythromycin, azithromycin,
Nondepolarizing muscle relaxants for surgery
D-Tubocurarine (curare), pancuronium, vecuronium, atracurium
Beta-blocking agents
Propranolol, atenolol, metoprolol
Local anesthetics and related agents
Procaine, Xylocaine in large amounts
Procainamide (for arrhythmias)
Botulinum toxin
Botox exacerbates weakness
Quinine derivatives
Quinine, quinidine, chloroquine, mefloquine (Lariam)
Magnesium
Decreases ACh release
Penicillamine
May cause MG
Drugs with important interactions in MG
Cyclosporine
Broad range of drug interactions, which may raise or lower cyclosporine levels.
Azathioprine
Avoid allopurinol—combination may result in myelosuppression.

Because of the side effects of glucocorticoids and other immunosuppressive agents used in the treatment of MG, a thorough medical investigation should be undertaken, searching specifically for evidence of chronic or latent infection (such as tuberculosis or hepatitis), hypertension, diabetes, renal disease, and glaucoma.

Treatment: Myasthenia Gravis

The prognosis has improved strikingly as a result of advances in treatment. Nearly all myasthenic patients can be returned to full productive lives with proper therapy. The most useful treatments for MG include anticholinesterase medications, immunosuppressive agents, thymectomy, and plasmapheresis or intravenous immunoglobulin (IVIg) **(Fig. 386-2)**.

Figure 386-2

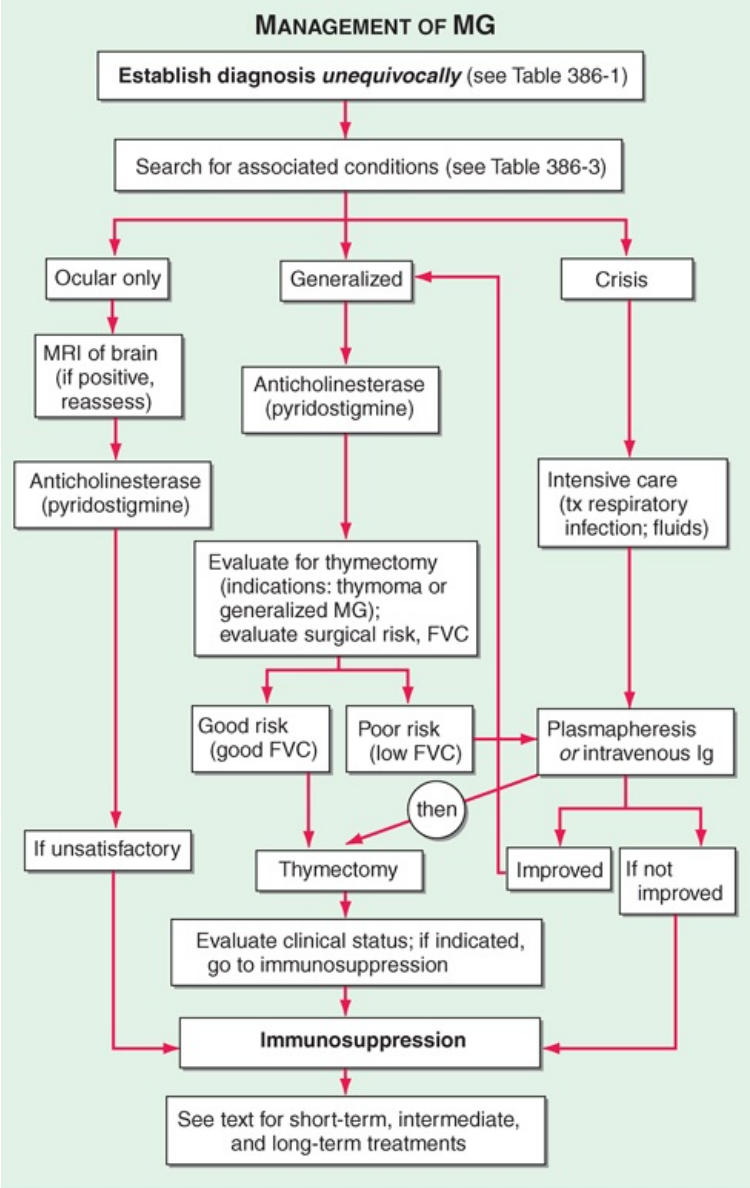

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Algorithm for the management of myasthenia gravis. FVC, forced vital capacity.

ANTICHOLINESTERASE MEDICATIONS

Anticholinesterase medication produces at least partial improvement in most myasthenic patients, although improvement is complete in only a few. Pyridostigmine is the most widely used anticholinesterase drug. The beneficial action of oral pyridostigmine begins within 15–30 min and lasts for 3–4 h, but individual responses vary. Treatment is begun with a moderate dose, e.g., 30–60 mg three to four times daily. The frequency and amount of the dose should be tailored to the patient's individual requirements throughout the day. For example, patients with weakness in chewing and swallowing may benefit by taking the medication before meals so that peak strength coincides with mealtimes. Long-acting pyridostigmine may occasionally be useful to get the patient through the night but should not be used for daytime medication because of variable absorption. The maximum useful dose of pyridostigmine rarely exceeds 120 mg every 4–6 h during daytime. Overdosage with anticholinesterase medication may cause increased weakness and other side effects. In some patients, muscarinic side effects of the anticholinesterase medication (diarrhea, abdominal cramps, salivation, nausea) may limit the dose tolerated. Atropine/diphenoxylate or loperamide is useful for the treatment of gastrointestinal symptoms.

THYMECTOMY

Two separate issues should be distinguished: (1) surgical removal of thymoma, and (2) thymectomy as a treatment for MG. Surgical removal of a thymoma is necessary because of the possibility of local tumor spread, although most thymomas are histologically benign. In the absence of a tumor, the available evidence suggests that up to 85% of patients experience improvement after thymectomy; of

these, ~35% achieve drug-free remission. However, the improvement is typically delayed for months to years. The advantage of thymectomy is that it offers the possibility of long-term benefit, in some cases diminishing or eliminating the need for continuing medical treatment. In view of these potential benefits and of the negligible risk in skilled hands, thymectomy has gained widespread acceptance in the treatment of MG. It is the consensus that thymectomy should be carried out in all patients with generalized MG who are between the ages of puberty and at least 55 years. Whether thymectomy should be recommended in children, in adults >55 years of age, and in patients with weakness limited to the ocular muscles is still a matter of debate. There is also suggestive evidence that patients with MuSK antibody–positive MG may respond less well to thymectomy. Thymectomy must be carried out in a hospital where it is performed regularly and where the staff is experienced in the pre- and postoperative management, anesthesia, and surgical techniques of total thymectomy.

IMMUNOSUPPRESSION

Immunosuppression using glucocorticoids, azathioprine, and other drugs is effective in nearly all patients with MG. The choice of drugs or other immunomodulatory treatments should be guided by the relative benefits and risks for the individual patient and the urgency of treatment. It is helpful to develop a treatment plan based on short-term, intermediate-term, and long-term objectives. For example, if immediate improvement is essential either because of the severity of weakness or because of the patient's need to return to activity as soon as possible, IVIg should be administered or plasmapheresis should be undertaken. For the intermediate term, glucocorticoids and cyclosporine or tacrolimus generally produce clinical improvement within a period of 1–3 months. The beneficial effects of azathioprine and mycophenolate mofetil usually begin after many months (as long as a year), but these drugs have advantages for the long-term treatment of patients with MG. For the occasional patient with MG that is genuinely refractory to optimal treatment with conventional immunosuppressive agents, a course of high-dose cyclophosphamide may induce long-lasting benefit by "rebooting" the immune system. At high doses, cyclophosphamide eliminates mature lymphocytes but spares hematopoietic precursors (stem cells), because they express the enzyme aldehyde dehydrogenase, which hydrolyzes cyclophosphamide. At present, this procedure is reserved for refractory patients and should be administered only in a facility fully familiar with this approach. We recommend maintenance immunotherapy after rebooting, to sustain the beneficial effect.

GLUCOCORTICOID THERAPY

Glucocorticoids, when used properly, produce improvement in myasthenic weakness in the great majority of patients. To minimize adverse side effects, prednisone should be given in a single dose rather than in divided doses throughout the day. The initial dose should be relatively low (15–25 mg/d) to avoid the early weakening that occurs in about one-third of patients treated initially with a high-dose regimen. The dose is increased stepwise, as tolerated by the patient (usually by 5 mg/d at 2- to 3-day intervals), until there is marked clinical improvement or a dose of 50–60 mg/d is reached. This dose is maintained for 1–3 months and then is gradually modified to an alternate-day regimen over the course of an additional 1–3 months; the goal is to reduce the dose on the "off day" to zero or to a minimal level. Generally, patients begin to improve within a few weeks after reaching the maximum dose, and improvement continues to progress for months or years. The prednisone dosage may gradually be reduced, but usually months or years may be needed to determine the minimum effective dose, and close monitoring is required. Few patients are able to do without immunosuppressive agents entirely. Patients on long-term glucocorticoid therapy must be followed carefully to prevent or treat adverse side effects. The most common errors in glucocorticoid treatment of myasthenic patients include (1) insufficient persistence—improvement may be delayed and gradual; (2) tapering the dosage too early, too rapidly, or excessively; and (3) lack of attention to prevention and treatment of side effects.

The management of patients treated with glucocorticoids is discussed in Chap. 342.

Other Immunosuppressive Drugs

Mycophenolate mofetil, azathioprine, cyclosporine, tacrolimus, and occasionally cyclophosphamide are effective in many patients, either alone or in various combinations.

Mycophenolate mofetil has become one of the most widely used drugs in the treatment of MG because of its effectiveness and relative lack of side effects. A dose of 1–1.5 g bid is recommended. Its mechanism of action involves inhibition of purine synthesis by the de novo pathway. Since lymphocytes lack the alternative salvage pathway that is present in all other cells, mycophenolate inhibits proliferation of lymphocytes but not proliferation of other cells. It does not kill or eliminate preexisting autoreactive lymphocytes, and therefore clinical improvement may be delayed for many months to a year, until the preexisting autoreactive lymphocytes die spontaneously. The advantage of mycophenolate lies in its relative lack of adverse side effects, with only occasional production of GI symptoms, rare development of leukopenia, and very small risks of malignancy or PML inherent in all immunosuppressive treatments. Although two published studies did not have positive outcomes, most experts attribute the negative results to flaws in the trial designs, and mycophenolate is widely used for long-term treatment of myasthenic patients. Until recently, azathioprine has been the most commonly used immunosuppressive agent for MG because of its relative safety in most patients and long track record. Its therapeutic effect may add to that of glucocorticoids and/or allow the glucocorticoid dose to

be reduced. However, up to 10% of patients are unable to tolerate azathioprine because of idiosyncratic reactions consisting of flulike symptoms of fever and malaise, bone marrow suppression, or abnormalities of liver function. An initial dose of 50 mg/d should be used for several days to test for these side effects. If this dose is tolerated, it is increased gradually to about 2–3 mg/kg of total body weight, or until the white blood count falls to 3000 to 4000/μL. The beneficial effect of azathioprine takes 3–6 months to begin and even longer to peak. In patients taking azathioprine, allopurinol should never be used to treat hyperuricemia. Because the two drugs share a common degradation pathway; the result may be severe bone marrow suppression due to increased effects of the azathioprine.

The calcineurin inhibitors cyclosporine and tacrolimus (FK506) are approximately as effective as azathioprine and are being used increasingly in the management of MG. Their beneficial effect appears more rapidly than that of azathioprine. Either drug may be used alone, but they are usually used as an adjunct to glucocorticoids to permit reduction of the glucocorticoid dose. The usual dose of cyclosporine is 4–5 mg/kg per d, and the average dose of tacrolimus is 0.07–0.1 mg/kg per d, given in two equally divided doses (to minimize side effects). Side effects of these drugs include hypertension and nephrotoxicity, which must be closely monitored. "Trough" blood levels are measured 12 h after the evening dose. The therapeutic range for the trough level of cyclosporine is 150–200 ng/L, and for tacrolimus it is 5–15 ng/L.

Cyclophosphamide is reserved for occasional patients refractory to the other drugs (see above for discussion of high-dose cyclophosphamide treatment). Rituximab, a monoclonal antibody that depletes CD20 B cells, has been used with variable—sometimes dramatic—success in the treatment of MG, especially in patients with anti-MuSK antibody.

PLASMAPHERESIS AND INTRAVENOUS IMMUNOGLOBULIN

Plasmapheresis has been used therapeutically in MG. Plasma, which contains the pathogenic antibodies, is mechanically separated from the blood cells, which are returned to the patient. A course of five exchanges (3–4 L per exchange) is generally administered over a 10- to 14-day period. Plasmapheresis produces a short-term reduction in anti-AChR antibodies, with clinical improvement in many patients. It is useful as a temporary expedient in seriously affected patients or to improve the patient's condition prior to surgery (e.g., thymectomy).

The indications for the use of IVIg are the same as those for plasma exchange: to produce rapid improvement to help the patient through a difficult period of myasthenic weakness or prior to surgery. This treatment has the advantages of not requiring special equipment or large-bore venous access. The usual dose is 2 g/kg, which is typically administered over 5 days (400 mg/kg per d). If tolerated, the total dose of IVIg can be given over a 3- to 4-day period. Improvement occurs in ~70% of patients, beginning during treatment, or within a week, and continuing for weeks to months. The mechanism of action of IVIg is not known; the treatment has no consistent effect on the measurable amount of circulating AChR antibody. Adverse reactions are generally not serious but include headache, fluid overload, and rarely aseptic meningitis or renal failure. IVIg should rarely be used as a long-term treatment in place of rationally managed immunosuppressive therapy. Unfortunately, there is a tendency for physicians unfamiliar with immunosuppressive treatments to rely on repeated IVIg infusions, which usually produce only intermittent benefit, do not reduce the underlying autoimmune response, and are costly. The intermediate and long-term treatment of myasthenic patients requires other methods of therapy outlined earlier in this chapter.

MANAGEMENT OF MYASTHENIC CRISIS

Myasthenic crisis is defined as an exacerbation of weakness sufficient to endanger life; it usually consists of respiratory failure caused by diaphragmatic and intercostal muscle weakness. Crisis rarely occurs in properly managed patients. Treatment should be carried out in intensive care units staffed with teams experienced in the management of MG, respiratory insufficiency, infectious disease, and fluid and electrolyte therapy. The possibility that deterioration could be due to excessive anticholinesterase medication ("cholinergic crisis") is best excluded by temporarily stopping anticholinesterase drugs. The most common cause of crisis is intercurrent infection. This should be treated immediately, because the mechanical and immunologic defenses of the patient can be assumed to be compromised. The myasthenic patient with fever and early infection should be treated like other immunocompromised patients. Early and effective antibiotic therapy, respiratory assistance (preferably noninvasive, using BiPap), and pulmonary physiotherapy are essentials of the treatment program. As discussed above, plasmapheresis or IVIg is frequently helpful in hastening recovery.

DRUGS TO AVOID IN MYASTHENIC PATIENTS

Many drugs have been reported to exacerbate weakness in patients with MG **(Table 386-4)**, but not all patients react adversely to all of these. Conversely, not all "safe" drugs can be used with impunity in patients with MG. As a rule, the listed drugs should be avoided *whenever possible*, and myasthenic patients should be followed closely when *any new drug* is introduced.

Patient Assessment

To evaluate the effectiveness of treatment as well as drug-induced side effects, it is important to

assess the patient's clinical status systematically at baseline and on repeated interval examinations. Because of the variability of symptoms of MG, the interval history and physical findings on examination must be taken into account. The most useful clinical tests include forward arm abduction time (up to a full 5 min), forced vital capacity, range of eye movements, and time to development of ptosis on upward gaze. Manual muscle testing or, preferably, quantitative dynamometry of limb muscles, especially proximal muscles, is also important. An interval form can provide a succinct summary of the patient's status and a guide to treatment results; an abbreviated form is shown in **Fig. 386-3**. A progressive reduction in the patient's AChR antibody level also provides clinically valuable confirmation of the effectiveness of treatment; conversely, a rise in AChR antibody levels during tapering of immunosuppressive medication may predict clinical exacerbation. For reliable quantitative measurement of AChR antibody levels, it is best to compare antibody levels from prior frozen serum aliquots with current serum samples in simultaneously run assays.

Figure 386-3

Myasthenia Gravis Worksheet

History				
General	Normal	Good	Fair	Poor
Diplopia	None	Rare	Occasional	Constant
Ptosis	None	Rare	Occasional	Constant
Arms	Normal	Slightly limited	Some ADL impairment	Definitely limited
Legs	Normal	Walks/runs fatigues	Can walk limited distances	Minimal walking
Speech	Normal	Dysarthric	Severely dysarthric	Unintelligible
Voice	Normal	Fades	Impaired	Severely impaired
Chew	Normal	Fatigue on normal foods	Fatigue on soft foods	Feeding tube
Swallow	Normal	Normal foods	Soft foods only	Feeding tube
Respiration	Normal	Dyspnea on unusual effort	Dyspnea on any effort	Dyspnea at rest

Examination

BP _____ Pulse _____ Wt _____
Edema _____
Vital capacity _____
Cataracts? R _____ L _____
EOMS _____
Ptosis time _____
Face _____

Arm abduction time R_____ L_____
Deltoids R_____ L_____
Biceps R_____ L_____
Triceps R_____ L_____
Grip R_____ L_____
Iliopsoas R_____ L_____
Quadriceps R_____ L_____
Hamstrings R_____ L_____
Other R_____ L_____

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Abbreviated interval assessment form for use in evaluating treatment for myasthenia gravis.

FURTHER READINGS

Drachman DB: Chapter 8. Therapy of myasthenia gravis. Handb Clin Neurol 91:253, 2008[PMID: 18631846] [Full Text]

——— et al: Rebooting the immune system with high dose cyclophosphamide for treatment of refractory myasthenia gravis. Ann NY Acad Sci 1132:305, 2008

Farrugia ME, Vincent A: Autoimmune mediated neuromuscular junction defects. Curr Opin Neurol 23:489, 2010[PMID: 20651592] [Full Text]

Lang B, Vincent A: Autoimmune disorders of the neuromuscular junction. Curr Opin Pharmacol 3:336, 2009

Leite MI et al: IgG1 antibodies to acetylcholine receptors in "seronegative" myasthenia gravis. Brain 131:1684, 2008

Oh SJ: Muscle-specific receptor tyrosine kinase antibody positive myasthenia gravis current status. J Clin Neurol 5:53, 2009[PMID: 19587811] [Full Text]

Phan SJ et al: Mycophenolate mofetil in myasthenia gravis: The unanswered question. Expert Opin Pharmacother 14: 2545, 2008

Skeie GO et al: Guidelines for treatment of autoimmune neuromuscular transmission disorders. Eur J Neurol 17:893, 2010[PMID: 20402760] [Full Text]

Zinman L et al: IV immunoglobulin in patients with myasthenia gravis: A randomized controlled trial. Neurology 68:837, 2007[PMID: 17353471] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MUSCULAR DYSTROPHIES AND OTHER MUSCLE DISEASES: INTRODUCTION

Skeletal muscle diseases, or myopathies, are disorders with structural changes or functional impairment of muscle. These conditions can be differentiated from other diseases of the motor unit (e.g., lower motor neuron or neuromuscular junction pathologies) by characteristic clinical and laboratory findings.

Myasthenia gravis and related disorders are discussed in Chap. 386; dermatomyositis, polymyositis, and inclusion body myositis are discussed in Chap. 388.

Clinical Features

Most myopathies present with proximal, symmetric limb weakness (arms or legs) with preserved reflexes and sensation. However, asymmetric and predominantly distal weakness can be seen in some myopathies. An associated sensory loss suggests injury to peripheral nerve or the central nervous system (CNS) rather than myopathy. On occasion, disorders affecting the motor nerve cell bodies in the spinal cord (anterior horn cell disease), the neuromuscular junction, or peripheral nerves can mimic findings of myopathy.

MUSCLE WEAKNESS

Symptoms of muscle weakness can be either intermittent or persistent. Disorders causing *intermittent weakness***(Fig. 387-1)** include myasthenia gravis, periodic paralyses (hypokalemic, hyperkalemic, and paramyotonia congenita), and metabolic energy deficiencies of glycolysis (especially myophosphorylase deficiency), fatty acid utilization (carnitine palmitoyltransferase deficiency), and some mitochondrial myopathies. The states of energy deficiency cause activity-related muscle breakdown accompanied by myoglobinuria, appearing as light-brown- to dark-brown-colored urine.

Figure 387-1

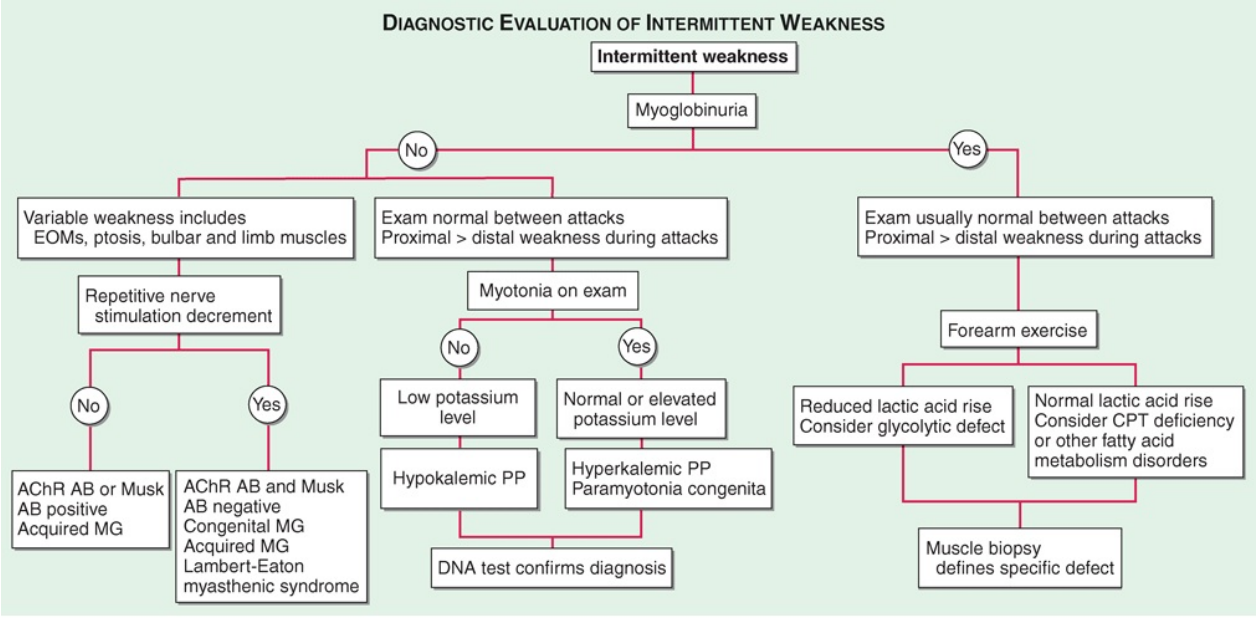

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagnostic evaluation of intermittent weakness. AChR AB, acetylcholine receptor antibody; CPT, carnitine palmitoyltransferase; EOMs, extraocular muscles; MG, myasthenia gravis; PP, periodic paralysis.

Most muscle disorders cause *persistent weakness***(Fig. 387-2)**. In the majority of these, including most types of muscular dystrophy, polymyositis, and dermatomyositis, the proximal muscles are weaker than the distal and are symmetrically affected, and the facial muscles are spared, a pattern referred to as *limb-girdle*. The differential diagnosis is more restricted for other patterns of weakness. Facial weakness (difficulty with eye closure and impaired smile) and scapular winging **(Fig. 387-3)** are characteristic of facioscapulohumeral dystrophy (FSHD). Facial and distal limb weakness associated with hand grip myotonia is virtually diagnostic of myotonic dystrophy type 1. When other cranial nerve muscles are weak, causing ptosis or extraocular muscle weakness, the most important disorders to consider include neuromuscular junction disorders, oculopharyngeal muscular dystrophy, mitochondrial myopathies, or some of the congenital myopathies **(Table 387-1)**. A pathognomonic pattern characteristic of inclusion body myositis is atrophy and weakness of the flexor forearm (e.g., wrist and finger flexors) and quadriceps muscles that is often asymmetric. Less frequently, but important diagnostically, is the presence of a dropped head syndrome indicative of selective neck extensor muscle weakness. The most important neuromuscular diseases associated with this pattern of weakness include myasthenia gravis, amyotrophic lateral sclerosis, late-onset nemaline myopathy, hyperparathyroidism, focal myositis, and some forms of inclusion body myopathy. A final pattern, recognized because of preferential distal extremity weakness, is typical of a unique category of muscular dystrophy, the distal myopathies.

Figure 387-2

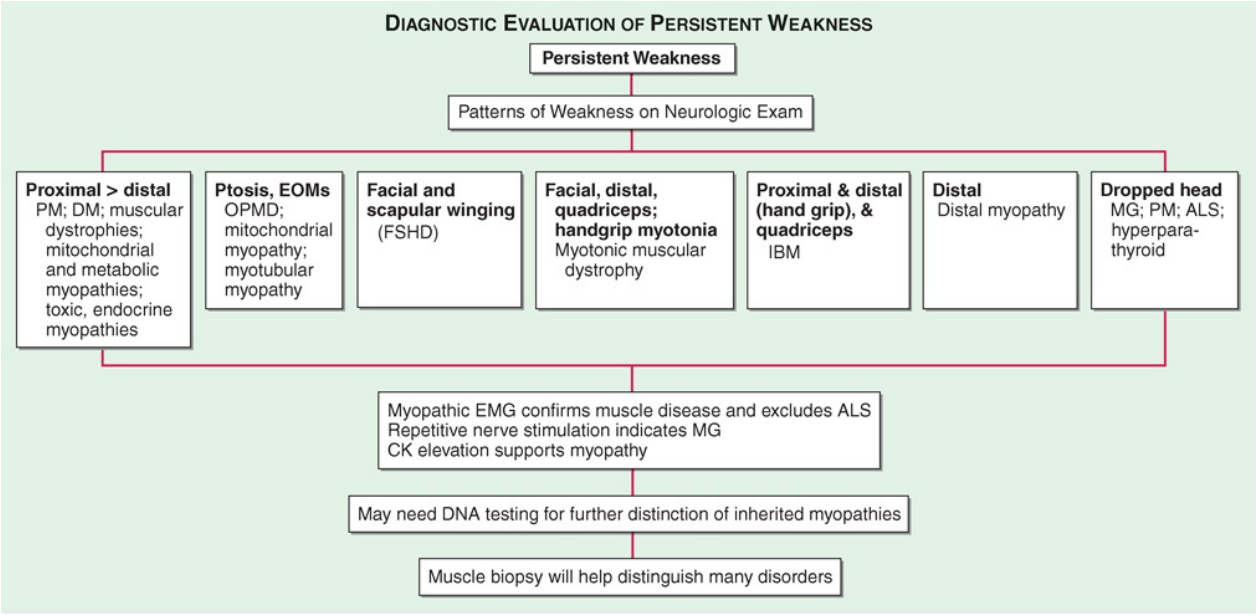

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Diagnostic evaluation of persistent weakness. Examination reveals one of seven patterns of weakness. The pattern of weakness in combination with the laboratory evaluation leads to a diagnosis. EOMs, extraocular muscles; OPMD, oculopharyngeal muscular dystrophy; FSHD, facioscapulohumeral dystrophy; IBM, inclusion body myositis; DM, dermatomyositis; PM, polymyositis; MG, myasthenia gravis; ALS, amyotrophic lateral sclerosis; CK, creatine kinase; EMG, electromyography.

Figure 387-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Facioscapulohumeral dystrophy with prominent scapular winging.

Table 387-1 Neuromuscular Causes of Ptosis or Ophthalmoplegia

Peripheral Neuropathy
Guillain-Barré syndrome
Miller-Fisher syndrome
Neuromuscular Junctionr
Botulism
Lambert-Eaton syndrome
Myasthenia gravis
Congenital myasthenia
Myopathy
Mitochondrial myopathies
Kearns-Sayre syndrome
Progressive external ophthalmoplegia
Oculopharyngeal and oculopharyngodistal muscular dystrophy
Myotonic dystrophy (ptosis only)
Congenital myopathy

	Myotubular Nemaline (ptosis only)
Hyperthyroidism/Graves' disease (ophthalmoplegia without ptosis)	
Hereditary inclusion body myopathy type 3	

It is important to examine functional capabilities to help disclose certain patterns of weakness **(Table 387-2)**. The Gowers' sign **(Fig. 387-4)** is particularly useful. Observing the gait of an individual may disclose a lordotic posture caused by combined trunk and hip weakness, frequently exaggerated by toe walking **(Fig. 387-5)**. A waddling gait is caused by the inability of weak hip muscles to prevent hip drop or hip dip. Hyperextension of the knee (genu recurvatum or back-kneeing) is characteristic of quadriceps muscle weakness; and a steppage gait, due to footdrop, accompanies distal weakness.

Table 387-2 Observations on Examination that Disclose Muscle Weakness

Functional Impairment	Muscle Weakness
Inability to forcibly close eyes	Upper facial muscles
Impaired pucker	Lower facial muscles
Inability to raise head from prone position	Neck extensor muscles
Inability to raise head from supine position	Neck flexor muscles
Inability to raise arms above head	Proximal arm muscles (may be only scapular stabilizing muscles)
Inability to walk without hyperextending knee (back-kneeing or genu recurvatum)	Knee extensor muscles
Inability to walk with heels touching the floor (toe walking)	Shortening of the Achilles tendon
Inability to lift foot while walking (steppage gait or footdrop)	Anterior compartment of leg
Inability to walk without a waddling gait	Hip muscles
Inability to get up from the floor without climbing up the extremities (Gowers' sign)	Hip, thigh, and trunk muscles
Inability to get up from a chair without using arms	Hip muscles

Figure 387-4

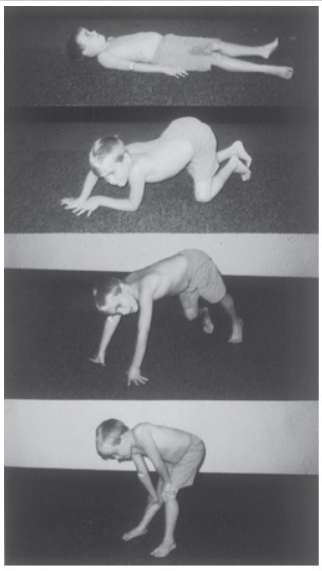

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Gowers' sign showing a patient using arms to climb up the legs in attempting to get up from the floor.

Figure 387-5

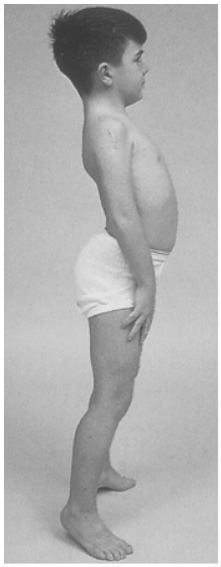

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Lordotic posture, exaggerated by standing on toes, associated with trunk and hip weakness.

Any disorder causing muscle weakness may be accompanied by *fatigue*, referring to an inability to maintain or sustain a force (pathologic fatigability). This condition must be differentiated from asthenia, a type of fatigue caused by excess tiredness or lack of energy. Associated symptoms may help differentiate asthenia and pathologic fatigability. Asthenia is often accompanied by a tendency to avoid physical activities, complaints of daytime sleepiness, necessity for frequent naps, and difficulty concentrating on activities such as reading. There may be feelings of overwhelming stress and depression. Thus, asthenia is not a myopathy. In contrast, pathologic fatigability occurs in disorders of neuromuscular transmission and in disorders altering energy production, including defects in glycolysis, lipid metabolism, or mitochondrial energy production. Pathologic fatigability also occurs in chronic myopathies because of difficulty accomplishing a task with less muscle. Pathologic fatigability is accompanied by abnormal clinical or laboratory findings. Fatigue without those supportive features almost never indicates a primary muscle disease.

MUSCLE PAIN (MYALGIAS), CRAMPS, AND STIFFNESS

Muscle pain can be associated with cramps, spasms, contractures, and stiff or rigid muscles. In distinction, true myalgia (muscle aching), which can be localized or generalized, may be accompanied by weakness, tenderness to palpation, or swelling. Certain drugs cause true myalgia **(Table 387-3)**.

Table 387-3 Drugs that Cause True Myalgia
Cimetidine
Cocaine
Cyclosporine
Danazol
Emetine
Gold
Heroin
Labetalol
Methadone
d-Penicillamine
Statins and other cholesterol-lowering agents
l-Tryptophan
Zidovudine

There are two painful muscle conditions of particular importance, neither of which is associated with muscle weakness. *Fibromyalgia* is a common, yet poorly understood, type of myofascial pain syndrome. Patients complain of severe muscle pain and tenderness and have specific painful trigger points, sleep disturbances, and easy fatigability. Serum creatine kinase (CK), erythrocyte sedimentation rate (ESR), electromyography (EMG), and muscle biopsy are normal (Chap. 335). *Polymyalgia rheumatica* occurs mainly in patients >50 years and is characterized by stiffness and pain in the shoulders, lower back, hips, and thighs (Chap. 326). The ESR is elevated, while serum CK, EMG, and muscle biopsy are normal. Temporal arteritis, an inflammatory disorder of medium- and large-sized arteries, usually involving one or more branches of the carotid artery, may accompany polymyalgia rheumatica. Vision is threatened by ischemic optic neuritis. Glucocorticoids can relieve the myalgias and protect against visual loss.

Localized muscle pain is most often traumatic. A common cause of sudden abrupt-onset pain is a ruptured tendon, which leaves the muscle belly appearing rounded and shorter in appearance compared to the normal side. The biceps brachii and Achilles tendons are particularly vulnerable to rupture. Infection or neoplastic infiltration of the muscle is a rare cause of localized muscle pain.

A *muscle cramp* or *spasm* is a painful, involuntary, localized, muscle contraction with a visible or palpable hardening of the muscle. Cramps are abrupt in onset, short in duration, and may cause abnormal posturing of the joint. The EMG shows firing of motor units, reflecting an origin from spontaneous neural discharge. Muscle cramps often occur in neurogenic disorders, especially motor neuron disease (Chap. 374), radiculopathies, and polyneuropathies (Chap. 384), but are not a feature of most primary muscle diseases. Duchenne's muscular dystrophy is an exception since calf muscle complaints are a common complaint. Muscle cramps are also common during pregnancy.

A *muscle contracture* is different from a muscle cramp. In both conditions, the muscle becomes hard, but a contracture is associated with energy failure in glycolytic disorders. The muscle is unable to relax after an active muscle contraction. The EMG shows electrical silence. Confusion is created because contracture also refers to a muscle that cannot be passively stretched to its proper length (fixed contracture) because of fibrosis. In some muscle disorders, especially in Emery-Dreifuss muscular dystrophy and Bethlem's myopathy, fixed contractures occur early and represent distinctive features of the disease.

Muscle stiffness can refer to different phenomena. Some patients with inflammation of joints and periarticular surfaces feel stiff. This condition is different from the disorders of hyperexcitable motor nerves causing stiff or rigid muscles. In *stiff-person syndrome*, spontaneous discharges of the motor neurons of the spinal cord cause involuntary muscle contractions mainly involving the axial (trunk) and proximal lower extremity muscles. The gait becomes stiff and labored, with hyperlordosis of the lumbar spine. Superimposed episodic muscle spasms are precipitated by sudden movements, unexpected noises, and emotional upset. The muscles relax during sleep. Serum antibodies against glutamic acid decarboxylase are present in approximately two-thirds of cases. In *neuromyotonia* (*Isaac's syndrome*) there is hyperexcitability of the peripheral nerves manifesting as continuous muscle fiber activity. *Myokymia* (groups of fasciculations associated with continuous undulations of muscle) and impaired muscle relaxation are the result. Muscles of the leg are stiff, and the constant contractions of the muscle cause increased sweating of the extremities. This peripheral nerve hyperexcitability is mediated by antibodies that target voltage-gated potassium channels. The site of origin of the spontaneous nerve discharges is principally in the distal portion of the motor nerves.

Myotonia is a condition of prolonged muscle contraction followed by slow muscle relaxation. It always follows muscle activation (action myotonia), usually voluntary, but may be elicited by mechanical stimulation (percussion myotonia) of the muscle. Myotonia typically causes difficulty in releasing objects after a firm grasp. In myotonic muscular dystrophy type 1 (DM1), distal weakness usually accompanies myotonia, whereas in DM2 proximal muscles are more affected; thus the related term *proximal myotonic myopathy* (PROMM) is used to describe this condition. Myotonia also occurs with *myotonia congenita* (a chloride channel disorder), but in this condition muscle weakness is not prominent. Myotonia may also be seen in individuals with sodium channel mutations (*hyperkalemic periodic paralysis* or *potassium-sensitive myotonia*). Another sodium channelopathy, *paramyotonia congenita*, also is associated with muscle stiffness. In contrast to other disorders associated with myotonia in which the myotonia is eased by repetitive activity, paramyotonia congenita is named for a paradoxical phenomenon whereby the myotonia worsens with repetitive activity.

MUSCLE ENLARGEMENT AND ATROPHY

In most myopathies muscle tissue is replaced by fat and connective tissue, but the size of the muscle is usually not affected. However, in many limb-girdle muscular dystrophies (and particularly the dystrophinopathies) enlarged calf muscles are typical. The enlargement represents true muscle hypertrophy, thus the term pseudohypertrophy should be avoided when referring to these patients. The calf muscles remain very strong even late in the course of these disorders. Muscle enlargement can also result from infiltration by sarcoid granulomas, amyloid deposits, bacterial and parasitic infections, and focal myositis. In contrast, muscle atrophy is characteristic of other myopathies. In dysferlinopathies (LGMD2B), there is a predilection for early atrophy of the gastrocnemius muscles, particularly the medial aspect. Atrophy of the humeral muscles is characteristic of facioscapulohumeral dystrophy (FSHD).

Laboratory Evaluation

A limited battery of tests can be used to evaluate a suspected myopathy. Nearly all patients require serum enzyme level measurements and electrodiagnostic studies as screening tools to differentiate muscle disorders from other motor unit diseases. The other tests described—DNA studies, the forearm exercise test, and muscle biopsy—are used to diagnose specific types of myopathies.

SERUM ENZYMES

CK is the preferred muscle enzyme to measure in the evaluation of myopathies. Damage to muscle causes the CK to leak from the muscle fiber to the serum. The MM isoenzyme predominates in skeletal muscle, while creatine kinase-myocardial bound (CK-MB) is the marker for cardiac muscle. Serum CK can be elevated in normal individuals without provocation, presumably on a genetic basis or after strenuous activity, minor trauma (including the EMG needle), a prolonged muscle cramp, or a generalized seizure. Aspartate aminotransferase (AST), alanine aminotransferase (ALT), aldolase, and lactic dehydrogenase (LDH) are enzymes sharing an origin in both muscle and liver. Problems arise when the levels of these enzymes are found to be elevated in a routine screening battery, leading to the erroneous assumption that liver disease is present when in fact muscle could be the cause. An elevated γ-glutamyl transferase (GGT) helps to establish a liver origin since this enzyme is not found in muscle.

ELECTRODIAGNOSTIC STUDIES

EMG, repetitive nerve stimulation, and nerve conduction studies (**Chap. e45**) are essential methods for evaluation of the patient with suspected muscle disease. In combination, they provide the information necessary to differentiate myopathies from neuropathies and neuromuscular junction diseases. Routine nerve conduction studies are typically normal in myopathies but reduced amplitudes of compound muscle action potentials may be seen in atrophied muscles. The needle EMG may reveal irritability on needle placement suggestive of a necrotizing myopathy (inflammatory myopathies, dystrophies, toxic myopathies, myotonic myopathies), whereas a lack of irritability is characteristic of long-standing myopathic disorders (muscular dystrophies, endocrine myopathies, disuse atrophy, and many of the metabolic myopathies). In addition, the EMG may demonstrate myotonic discharges that will narrow the differential diagnosis **(Table 387-4)**. Another important EMG finding is the presence of short-duration, small-amplitude, polyphasic motor unit action potentials (MUAPs). Such MUAPs can be seen in both myopathic and neuropathic disorders; however, the recruitment or firing pattern is different. In myopathies, the MUAPs fire early but at a normal rate to compensate for the loss of individual muscle fibers, whereas in neurogenic disorders the MUAPs fire faster. The EMG is usually normal in steroid or disuse myopathy, both of which are associated with type 2 fiber atrophy; this is because the EMG preferentially assesses the physiologic function of type 1 fibers. The EMG can also be invaluable in helping to choose an appropriately affected muscle to sample for biopsy.

Table 387-4 Myotonic Disorders

Myotonic dystrophy type 1 Myotonic dystrophy type 2/proximal myotonic myopathy
Myotonia congenita
Paramyotonia congenita
Hyperkalemic periodic paralysis
Chondrodystrophic myotonia (Schwartz-Jampel syndrome) Centronuclear/myotubular myopathya

Drug-induced
Cholesterol-lowering agents (statin medications, fibrates)
Cyclosporine
Chloroquine
Glycogen storage disordersa (Pompe's disease, debrancher deficiency, branching enzyme deficiency)
Myofibrillar myopathies[a]

[a]Associated with myotonic discharges on EMG but no clinical myotonia.

DNA ANALYSIS

This serves as an important tool for the definitive diagnosis of many muscle disorders. Nevertheless, there are a number of limitations in currently available molecular diagnostics. For example, in Duchenne's and Becker's dystrophies, two-thirds of patients have deletion or duplication mutations that are easy to detect, while the remainder have point mutations that are much more difficult to find. For patients without identifiable gene defects, the muscle biopsy remains the main diagnostic tool.

FOREARM EXERCISE TEST

In myopathies with intermittent symptoms, and especially those associated with myoglobinuria, there may be a defect in glycolysis. Many variations of the forearm exercise test exist. For safety, the test should not be performed under ischemic conditions to avoid an unnecessary insult to the muscle, causing rhabdomyolysis. The test is performed by placing a small indwelling catheter into an antecubital vein. A baseline blood sample is obtained for lactic acid and ammonia. The forearm muscles are exercised by asking the patient to vigorously open and close the hand for 1 minute. Blood is then obtained at intervals of 1, 2, 4, 6, and 10 minutes for comparison with the baseline sample. A three- to fourfold rise of lactic acid is typical. The simultaneous measurement of ammonia serves as a control, since it should also rise with exercise. In patients with myophosphorylase deficiency or other glycolytic defects, the lactic acid rise will be absent or below normal, while the rise in ammonia will reach control values. If there is lack of effort, neither lactic acid nor ammonia will rise. Patients with selective failure to increase ammonia may have myoadenylate deaminase deficiency. This condition has been reported to be a cause of myoglobinuria, but deficiency of this enzyme in asymptomatic individuals makes interpretation controversial.

MUSCLE BIOPSY

Muscle biopsy is an important step in establishing the diagnosis of a suspected myopathy. The biopsy is usually obtained from a quadriceps or biceps brachii muscle, less commonly from a deltoid muscle. Evaluation includes a combination of techniques —light microscopy, histochemistry, immunocytochemistry with a battery of antibodies, and electron microscopy. Not all techniques are needed for every case. A specific diagnosis can be established in many disorders. Endomysial inflammatory cells surrounding and invading muscle fibers are seen in polymyositis; similar endomysial infiltrates associated with muscle fibers containing rimmed vacuoles, amyloid deposits within fibers, and TDP-43 inclusions are characteristic of inclusion body myositis; while perivascular, perimysial inflammation associated with perifascicular atrophy are features of dermatomyositis. In addition, the congenital myopathies have distinctive light and electron microscopy features essential for diagnosis. Mitochondrial and metabolic (e.g., glycogen and lipid storage diseases) myopathies also demonstrate distinctive histochemical and electron-microscopic profiles. Biopsied muscle tissue can be sent for metabolic enzyme or mitochondrial DNA analyses. A battery of antibodies is available for the identification of missing components of the dystrophin-glycoprotein complex and related proteins to help diagnose specific types of muscular dystrophies. Western blot analysis on muscle specimens can be performed to determine whether specific muscle proteins are reduced in quantity or are of abnormal size.

HEREDITARY MYOPATHIES

Muscular dystrophy refers to a group of hereditary progressive diseases each with unique phenotypic and genetic features **(Tables 387-5, 387-6, and 387-7)**.

Table 387-5 Progressive Muscular Dystrophies

Type	Inheritance	Defective Gene/Protein	Onset Age	Clinical Features	Other Organ Systems Involved
Duchenne's	XR	Dystrophin	Before 5 years	Progressive weakness of girdle muscles Unable to walk after age 12 Progressive kyphoscoliosis Respiratory failure in 2d or 3d decade	Cardiomyopathy Mental impairment
Becker's	XR	Dystrophin	Early childhood to adult	Progressive weakness of girdle muscles Able to walk after age 15 Respiratory failure may develop by 4th decade	Cardiomyopathy
Limb-girdle	AD/AR	Several (Tables 387-6, 387-7)	Early childhood to early adult	Slow progressive weakness of shoulder and hip girdle muscles	± Cardiomyopathy
Emery-Dreifuss	XR/AD	Emerin/Lamins A/C Nesprin-1, Nesprin 2, TMEM43	Childhood to adult	Elbow contractures, humeral and peroneal weakness	Cardiomyopathy
Congenital	AR	Several	At birth or within first few months	Hypotonia, contractures, delayed milestones Progression to respiratory failure in some; static course in others	CNS abnormalities (hypomyelination, malformation) Eye abnormalities
Myotonica (DM1,	AD	DM1:	Childhood to adult	Slowly progressive weakness	Cardiac conduction

DM2)	Expansion CTG repeat DM2: Expansion CCTG repeat	Maybe infancy if mother affected (DM1 only)	of face, shoulder girdle, and foot dorsiflexion Preferential proximal weakness in DM2	defects Mental impairment Cataracts Frontal baldness Gonadal atrophy
Facioscapulohumeral AD	DUX4 4q	Childhood to adult	Slowly progressive weakness of face, shoulder girdle, and foot dorsiflexion	Deafness Coats' (eye) disease
Oculopharyngeal AD	Expansion, poly-A RNA binding protein	5th to 6th decade	Slowly progressive weakness of extraocular, pharyngeal, and limb muscles	—

[a]Two forms of myotonic dystrophy, DM1 and DM2, have been identified. Many features overlap (see text).

Table 387-6 Autosomal Dominant Limb-Girdle Muscular Dystrophies (Lgmds)

Disease	Clinical Features	Laboratory Features	Locus or Gene
LGMD1A	Onset 3rd to 4th decade Muscle weakness affects distal limb muscles, vocal cords, and pharyngeal muscles	Serum CK 2 x normal EMG mixed myopathy/neuropathy NCS normal	Myotilin
LGMD1B	Onset 1st or 2nd decade Proximal lower limb weakness and cardiomyopathy with conduction defects Some cases indistinguishable from Emery-Dreifuss muscular dystrophy with joint contractures	Serum CK 3–5 x normal NCS normal EMG myopathic	Lamin A/C
LGMD1C	Onset in early childhood Proximal weakness Gowers' sign, calf hypertrophy Exercise-related muscle cramps	Serum CK 4–25 x normal NCS normal EMG myopathic	Caveolin-3
LGMD1D	Onset 3rd to 5th decade Proximal muscle weakness Cardiomyopathy and arrhythmias	Serum CK 2–4 x normal NCS normal EMG myopathic	Linked to chromosome 7q Gene unidentified
LGMD1E	Childhood onset Proximal muscle weakness	Serum CK usually normal NCS normal EMG myopathic	Linked to chromosome 6q23 Gene unidentified

Abbreviations: CK, creatine kinase; EMG, electromyography; NCS, nerve conduction studies.

Table 387-7 Autosomal Recessive Limb-Girdle Muscular Dystrophies (Lgmds)

Disease	Clinical Features	Laboratory Features	Locus or Gene
LGMD2A	Onset 1st or 2nd decade Tight heel cords Contractures at elbows, wrists, and fingers; rigid spine in some Proximal and distal weakness	Serum CK 3–15 x normal NCS normal EMG myopathic	Calpain-3
LGMD2B	Onset 2nd or 3rd decade Proximal muscle weakness at onset, later distal (calf) muscles affected Miyoshi's myopathy is variant of LGMD2B with calf muscles affected at onset	Serum CK 3–100 x normal NCS normal EMG myopathic Inflammation on muscle biopsy may simulate polymyositis	Dysferlin
LGMD2C–F	Onset in childhood to teenage years Clinical condition similar to Duchenne and Becker muscular dystrophies Cardiomyopathy uncommon Cognitive function normal	Serum CK 5–100 x normal NCS normal EMG myopathic	γ, α, β, δ sarcoglycans
LGMD2G	Onset age 10 to 15 Proximal and distal muscle weakness	Serum CK 3–17 x normal NCS normal EMG myopathic	Telethonin
LGMD2H	Onset 1st to 3rd decade Proximal muscle weakness	Serum CK 2–25 x normal NCS normal EMG myopathic	TRIM32 gene

LGMD2I	Onset 1st to 3rd decade Clinical condition similar to Duchenne's or Becker's dystrophies Cardiomyopathy (some not all) Cognitive function normal	Serum CK 10–30 x normal NCS normal EMG myopathic	Fukutin-related protein
LGMD2Ja	Onset 1st to 3rd decade Proximal lower limb weakness Mild distal weakness Progressive weakness causes loss of ambulation	Serum CK 1.5–2 x normal NCS normal EMG myopathic	Titin
LGMD2K	Usually presents in infancy as Walker-Warburgsyndrome but can present in early adult lifewith proximal weakness and only minorCNS abnormalities	CK 10–20 x normal NCS normal EMG myopathic	POMT1
LGMD2IL	Presents in childhood or adult life. May manifest with quadriceps atrophy and myalgia Some present with early involvement of the calves in the second decade of life, resembling Miyoshi myopathy (dysferlinopathy)	CK normal to 50 x normal NCS normal EMG myopathic	Anoctamin 5
LGMD2M	Usually presents in infancy as Fukuyuma congenital muscular dystrophy but can present in early adultlife with proximal weakness and only minorCNS abnormalities	CK 10–50 x normal NCS normal EMG myopathic	Fukutin
LGMD2N	Usually presents in infancy as Muscle-Eye-Braindisease but can present in early adult lifewith proximal weakness and only minorCNS abnormalities	CK 5–20 x normal NCS normal EMG myopathic	POMGnT1
LGMD2O	Usually presents in infancy as Walker-Warburgsyndrome but can present in early adult life with proximal weakness and only minorCNS abnormalities	CK 5–20 x normal NCS normal EMG myopathic	POMT2

[a]Tibial muscular dystrophy is a form of titin deficiency with only distal muscle weakness (see Table 387-9). ***Abbreviations:*** CK, creatine kinase; EMG, electromyography; NCS, nerve conduction studies. POMT1, protein-O-mannosyltransferase 1; POMT2, protein-O-mannosyltransferase 2; POMGNT1, O-linked mannose beta 1,2-N-acetylglucosaminyltransferase.

Duchenne's Muscular Dystrophy

This X-linked recessive disorder, sometimes also called *pseudohypertrophic muscular dystrophy* , has an incidence of ~30 per 100,000 live-born males.

CLINICAL FEATURES

Duchenne's dystrophy is present at birth, but the disorder usually becomes apparent between ages 3 and 5 years. The boys fall frequently and have difficulty keeping up with friends when playing. Running, jumping, and hopping are invariably abnormal. By age 5 years, muscle weakness is obvious by muscle testing. On getting up from the floor, the patient uses his hands to climb up himself [Gowers' maneuver (Fig. 387-4)]. Contractures of the heel cords and iliotibial bands become apparent by age 6 years, when toe walking is associated with a lordotic posture. Loss of muscle strength is progressive, with predilection for proximal limb muscles and the neck flexors; leg involvement is more severe than arm involvement. Between ages 8 and 10 years, walking may require the use of braces; joint contractures and limitations of hip flexion, knee, elbow, and wrist extension are made worse by prolonged sitting. By age 12 years, most patients are wheelchair dependent. Contractures become fixed, and a progressive scoliosis often develops that may be associated with pain. The chest deformity with scoliosis impairs pulmonary function, which is already diminished by muscle weakness. By age 16–18 years, patients are predisposed to serious, sometimes fatal pulmonary infections. Other causes of death include aspiration of food and acute gastric dilation.

A cardiac cause of death is uncommon despite the presence of a cardiomyopathy in almost all patients. Congestive heart failure seldom occurs except with severe stress such as pneumonia. Cardiac arrhythmias are rare. The typical electrocardiogram (ECG) shows an increased net RS in lead V_1; deep, narrow Q waves in the precordial leads; and tall right precordial R waves in V_1. Intellectual impairment in Duchenne's dystrophy is common; the average intelligence quotient (IQ) is ~1 SD below the mean. Impairment of intellectual function appears to be nonprogressive and affects verbal ability more than performance.

LABORATORY FEATURES

Serum CK levels are invariably elevated to between 20 and 100 times normal. The levels are abnormal at birth but decline late in the disease because of inactivity and loss of muscle mass. EMG demonstrates features typical of myopathy. The muscle biopsy shows muscle fibers of varying size as well as small groups of necrotic and regenerating fibers. Connective tissue and fat replace lost muscle fibers. A definitive diagnosis of Duchenne's dystrophy can be established on the basis of dystrophin deficiency in a biopsy of muscle tissue or mutation analysis on peripheral blood leukocytes, as discussed below.

Duchenne's dystrophy is caused by a mutation of the gene that encodes dystrophin, a 427-kDa protein localized to the inner surface of the sarcolemma of the muscle fiber. The dystrophin gene is >2000 kb in size and thus is one of the largest identified human genes. It is localized to the short arm of the X chromosome at Xp21. The most common gene mutation is a deletion. The size varies but does not correlate with disease severity. Deletions are not uniformly distributed over the gene but rather are most common near the beginning (5' end) and middle of the gene. Less often, Duchenne's dystrophy is caused by a gene duplication or point mutation. Identification of a specific mutation allows for an unequivocal diagnosis, makes possible accurate testing of potential carriers, and is useful for prenatal diagnosis.

A diagnosis of Duchenne's dystrophy can also be made by Western blot analysis of muscle biopsy specimens, revealing abnormalities on the quantity and molecular weight of dystrophin protein. In addition, immunocytochemical staining of muscle with dystrophin antibodies can be used to demonstrate absence or deficiency of dystrophin localizing to the sarcolemmal membrane. Carriers of the disease may demonstrate a mosaic pattern, butdystrophin analysis of muscle biopsy specimens for carrier detection is not reliable.

PATHOGENESIS

Dystrophin is part of a large complex of sarcolemmal proteins and glycoproteins **(Fig. 387-6)**. Dystrophin binds to F-actin at its amino terminus and to β-dystroglycan at the carboxyl terminus. β-Dystroglycan complexes to α-dystroglycan, which binds to laminin in the extracellular matrix (ECM). Laminin has a heterotrimeric molecular structure arranged in the shape of a cross with one heavy chain and two light chains, β_1 and γ_1. The laminin heavy chain of skeletal muscle is designated laminin α_2. Collagen proteins IV and VI are also found in the ECM. Like β-dystroglycan, the transmembrane sarcoglycan proteins also bind to dystrophin; these five proteins (designated α- through ε-sarcoglycan) complex tightly with each other. More recently, other membrane proteins implicated in muscular dystrophy have been found to be loosely affiliated with constituents of the dystrophin complex. These include caveolin-3, α_7 integrin, and collagen VI.

Figure 387-6

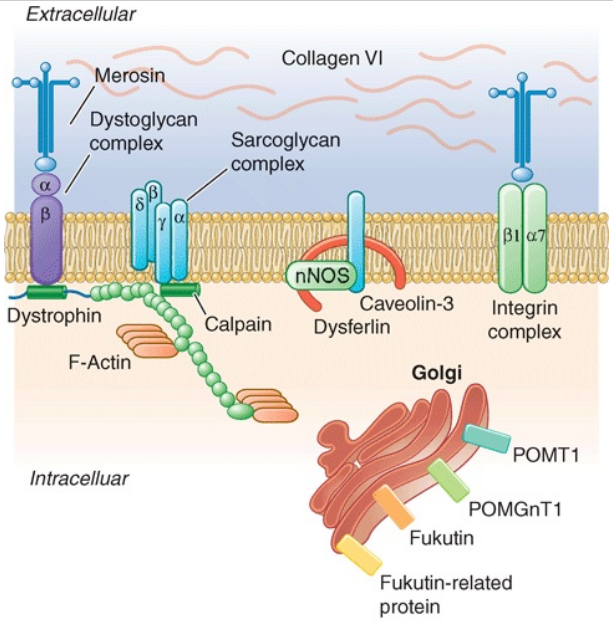

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Selected muscular dystrophy–associated proteins in the cell membrane and Golgi complex.

Dystrophin localizes to the cytoplasmic face of the muscle cell membrane. It complexes with two transmembrane protein complexes, the dystroglycans and the sarcoglycans. The dystroglycans bind to the extracellular matrix protein merosin, which is also complexed with β_1 and α_7 integrins (Tables 387-5, 387-6, 387-7). Dysferlin complexes with caveolin-3 (which binds to neuronal nitric oxide synthase, or nNOS) but not with the dystrophin-associated proteins or the integrins. In some of the congenital dystrophies and limb-girdle muscular dystrophies (LGMDs), there is loss of function of different enzymes that glycosylate α-dystroglycan, which thereby inhibits proper binding to merosin: POMT1, POMT2, POMGnT1, Fukutin, Fukutin-related protein, and LARGE.

The dystrophin-glycoprotein complex appears to confer stability to the sarcolemma, although the function of each individual component of the complex is incompletely understood. Deficiency of one member of the complex may cause abnormalities in other components. For example, a primary deficiency of dystrophin (Duchenne's dystrophy) may lead to secondary loss of the sarcoglycans and dystroglycan. The primary loss of a single sarcoglycan (see "Limb-Girdle Muscular Dystrophy," below) results in a secondary loss of other sarcoglycans in the membrane without uniformly affecting dystrophin. In either instance, disruption of the dystrophin-glycoprotein complexes weakens the sarcolemma, causing membrane tears and a cascade of events leading to muscle fiber necrosis. This sequence of events occurs repeatedly during the life of a patient with muscular dystrophy.

TREATMENT: DUCHENNE'S MUSCULAR DYSTROPHY

Glucocorticoids, administered as prednisone in a dose of 0.75 mg/kg per day, significantly slow progression of Duchenne's dystrophy for up to 3 years. Some patients cannot tolerate glucocorticoid therapy; weight gain and increased risk of fractures in particular represent a significant deterrent for some boys. As in other recessively inherited dystrophies presumed to arise from loss of function of a critical muscle gene, there is optimism that Duchenne's disease may benefit from novel therapies that either replace the defective gene or missing protein or implement downstream corrections (e.g., skipping mutated exons or reading through mutations that introduce stop codons).

Becker's Muscular Dystrophy

This less severe form of X-linked recessive muscular dystrophy results from allelic defects of the same gene responsible for Duchenne's dystrophy. Becker's muscular dystrophy is ~10 times less frequent than Duchenne's, with an incidence of about 3 per 100,000 live-born males.

CLINICAL FEATURES

The pattern of muscle wasting in Becker's muscular dystrophy closely resembles that seen in Duchenne's. Proximal muscles, especially of the lower extremities, are prominently involved. As the disease progresses, weakness becomes more generalized. Significant facial muscle weakness is not a feature. Hypertrophy of muscles, particularly in the calves, is an early and prominent finding.

Most patients with Becker's dystrophy first experience difficulties between ages 5 and 15 years, although onset in the third or fourth decade or even later can occur. By definition, patients with Becker's dystrophy walk beyond age 15, while patients with Duchenne's dystrophy are typically in a wheelchair by the age of 12. Patients with Becker's dystrophy have a reduced life expectancy, but most survive into the fourth or fifth decade.

Mental retardation may occur in Becker's dystrophy, but it is not as common as in Duchenne's. Cardiac involvement occurs in Becker's dystrophy and may result in heart failure; some patients manifest with only heart failure. Other less common presentations are asymptomatic hyper-CK-emia, myalgias without weakness, and myoglobinuria.

LABORATORY FEATURES

Serum CK levels, results of EMG, and muscle biopsy findings closely resemble those in Duchenne's dystrophy. The diagnosis of Becker's muscular dystrophy requires Western blot analysis of muscle biopsy samples, demonstrating a reduced amount or abnormal size of dystrophin or mutation analysis of DNA from peripheral blood leukocytes. Genetic testing reveals deletions or duplications of the dystrophin gene in 65% of patients with Becker's dystrophy, approximately the same percentage as in Duchenne's dystrophy. In both Becker's and Duchenne's dystrophies, the size of the DNA deletion does not predict clinical severity; however, in ~95% of patients with Becker's dystrophy, the DNA deletion does not alter the translational reading frame of messenger RNA. These "in-frame" mutations allow for production of some dystrophin, which accounts for the presence of altered rather than absent dystrophin on Western blot analysis.

TREATMENT: BECKER'S MUSCULAR DYSTROPHY

The use of glucocorticoids has not been adequately studied in Becker's dystrophy.

Limb-Girdle Muscular Dystrophy

The syndrome of LGMD represents more than one disorder. Both males and females are affected, with onset ranging from late in the first decade to the fourth decade. The LGMDs typically manifest with progressive weakness of pelvic and shoulder girdle musculature. Respiratory insufficiency from weakness of the diaphragm may occur, as may cardiomyopathy.

A systematic classification of LGMD is based on autosomal dominant (LGMD1) and autosomal recessive (LGMD2) inheritance. Superimposed on the backbone of LGMD1 and LGMD2, the classification employs a sequential alphabetical lettering system (LGMD1A, LGMD2A, etc.). Disorders receive letters in the order in which they are found to have chromosomal linkage. This results in an ever-expanding list of conditions. Presently there are 5 autosomal dominant and 10 autosomal recessive disorders, summarized in Tables 387-6 and 387-7. None of the conditions is as common as the dystrophinopathies; however, prevalence data for the LGMDs have not been systematically gathered for any large heterogeneous population. In referral-based clinical populations, Fukutin-related protein (FKRP) deficiency (LGMD2I), calpainopathies (LGMD2A), and to a lesser extent dysferlinopathies (LGMD2B) have emerged as the most common disorders.

Emery-Dreifuss Muscular Dystrophy

There are at least five genetically distinct forms of Emery-Dreifuss muscular dystrophy (EDMD). Emerin mutations are the most common cause of X-linked EDMD, though mutations in *FHL1* may also be associated with a similar phenotype, which is X-linked as well. Mutations involving the gene for lamin A/C are the most common cause of autosomal dominant EDMD (also known as LGMD1B) and is also a common cause of hereditary cardiomyopathy. Less commonly, autosomal dominant EDMD has been reported with mutations in nesprin-1, nesprin-2, and *TMEM43*.

CLINICAL FEATURES

Prominent contractures can be recognized in early childhood and teenage years, often preceding muscle weakness. The contractures persist throughout the course of the disease and are present at the elbows, ankles, and neck. Muscle weakness affects humeral and peroneal muscles at first and later spreads to a limb-girdle distribution. The cardiomyopathy is potentially life threatening and may result in sudden death. A spectrum of atrial rhythm and conduction defects includes atrial fibrillation and paralysis and atrioventricular heart block. Some patients have a dilated cardiomyopathy. Female carriers of the X-linked variant may have cardiac manifestations that become clinically significant.

LABORATORY FEATURES

Serum CK may be elevated two- to tenfold. EMG is myopathic. Muscle biopsy usually shows nonspecific dystrophic features, though cases associated with FHL1 mutations have features of myofibrillar myopathy. Immunohistochemistry reveals absent emerin staining of myonuclei in X-linked EDMD due to emerin mutations. ECGs demonstrate atrial and atrioventricular rhythm disturbances.

X-linked EDMD usually arises from defects in the emerin gene encoding a nuclear envelope protein. FHL1 mutations are also a cause of scapuloperoneal dystrophy, but can also present with an EDMD phenotype. The autosomal dominant disease can be caused by mutations in the *LMNA* gene encoding lamin A and C; in the synaptic nuclear envelope protein 1 (*SYNE1*) or 2 (*SYNE2*) encoding nesprin-1 and nesprin-2, respectively; and most recently in *TMEM43* encoding LUMA. These proteins are essential components of the filamentous network underlying the inner nuclear membrane. Loss of structural integrity of the nuclear envelope from defects in emerin, lamin A/C, nesprin-1, nesprin-2, and LUMA accounts for overlapping phenotypes **(Fig. 387-7)**.

Figure 387-7

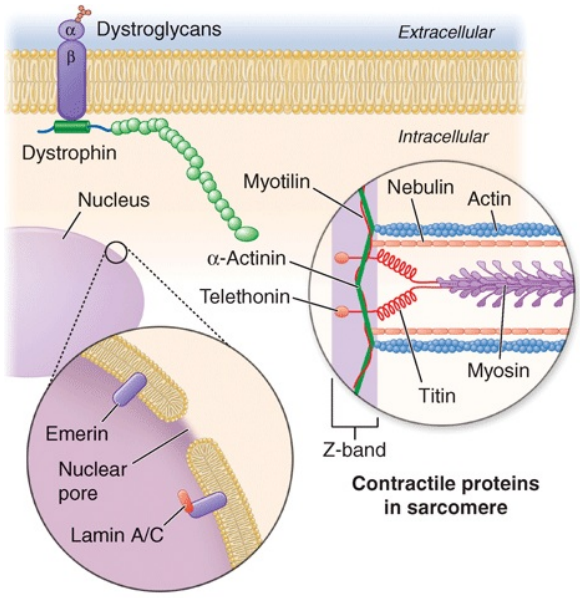

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Selected muscular dystrophy–associated proteins in the nuclear membrane and sarcomere. As shown in the exploded view, emerin and lamin A/C are constituents of the inner nuclear membrane. Several dystrophy-associated proteins are represented in the sarcomere including titin, nebulin, calpain, telethonin, actinin, and myotilin. The position of the dystrophin-dystroglycan complex is also illustrated.

TREATMENT: EMERY-DREIFUSS MUSCULAR DYSTROPHY

Supportive care should be offered for neuromuscular disability, including ambulatory aids, if necessary. Stretching of contractures is difficult. Management of cardiomyopathy and arrhythmias (e.g., early use of a defibrillator or cardiac pacemaker) may be life saving.

Congenital Muscular Dystrophy (Cmd)

This is not one entity but rather a group of disorders with varying degrees of muscle weakness, CNS impairment, and eyeabnormalities.

CLINICAL FEATURES

As a group, CMDs present at birth or in the first few months of life with hypotonia and proximal or generalized muscle weakness. Calf muscle hypertrophy is seen in some patients. Facial muscles may be weak, but other cranial nerve–innervated muscles are spared (e.g., extraocular muscles are normal). Most patients have joint contractures of varying degrees at elbows, hips, knees, and ankles. Contractures present at birth are referred to as *arthrogryposis*. Respiratory failure may be seen in some cases.

The CNS is affected in some forms of CMD. In merosin and FKRP deficiency, cerebral hypomyelination may be seen by MRI, though only a small number of patients have mental retardation and seizures. Three forms of congenital muscular dystrophy have severe brain impairment. These include Fukuyama congenital muscular dystrophy (FCMD), muscle-eye-brain (MEB) disease, and Walker-Warburg syndrome (WWS). Patients are severely disabled in all three of these conditions. In MEB disease and WWS, but not in FCMD, ocular abnormalities impair vision. WWS is the most severe congenital muscular dystrophy, causing death by 1 year of age.

LABORATORY FEATURES

Serum CK is markedly elevated in all of these conditions. The EMG is myopathic and muscle biopsies show nonspecific dystrophic features. Merosin, or laminin $\alpha 2$ chain (a basal lamina protein), is deficient in surrounding muscle fibers in merosin deficiency. Skin biopsies can also demonstrate defects in laminin $\alpha 2$ chain. In the other disorders (FKRP deficiency, FCMD, MEB disease, WWS), there is abnormal α-dystroglycan staining in muscle. In merosin deficiency, cerebral hypomyelination is common, and a host of brain malformations are seen in FCMD, MEB disease, and WWS.

All forms of CMD are inherited as autosomal recessive disorders. Chromosomal linkage and specific gene defects are presented in **Table 387-8**. With the exception of merosin, the other gene defects affect posttranslational glycosylation of α-dystroglycan. This abnormality is thought to impair binding with merosin and leads to weakening of the dystrophin-glycoprotein complex, instability of the muscle membrane, and/or abnormalities in muscle contraction. CMDs with brain and eye phenotypes probably involve defective glycosylation of additional proteins, accounting for the more extensive phenotypes.

Table 387-8 Congenital Muscular Dystrophies[a]

Disease	Clinical Features	Laboratory Features	Locus or Gene
Merosin deficiency	Onset at birth with hypotonia, joint contractures, delayed milestones, generalized muscle weakness Cerebral hypomyelination, less often cortical dysplasia Normal intelligence usually, some with MR (~6%) and seizures (~8%) Partial deficiency leads to milder phenotype (LGMD picture)	Serum CK 5–35 x normal EMG myopathic NCS abnormal in some cases	Laminin $\alpha 2$ chain
Fukitin-related protein	Onset at birth or shortly after	Serum CK 10–50 x normal	Fukutin-related

protein deficiency[b]	Hypotonia and feeding problems Weakness of proximal muscles, especially shoulder girdles Hypertrophy of leg muscles Joint contractures Cognition normal	EMG myopathic	related protein
Fukuyama congenital muscular dystrophy[b]	Onset at birth Hypotonia, joint contractures Generalized muscle weakness Hypertrophy of calf muscles Seizures, mental retardation Cardiomyopathy	Serum CK 10–50 x normal EMG myopathic NCS normal MRI shows hydrocephalus and periventricular and frontal hypomyelination	Fukutin
Muscle-eye-brain disease	Onset at birth, hypotonia Eye abnormalities include: progressive myopia, cataracts, and optic nerve, glaucoma, retinal pigmentary changes Progressive muscle weakness Joint contractures Seizures, mental retardation	Serum CK 5–20 x normal MRI shows hydrocephalus, cobblestone lissencephaly, corpus callosum and cerebellar hypoplasia, cerebral hypomyelination	*N*-acetyl-glucosaminyl transferase (POMGnT1)
Walker-Warburg syndrome[b]	Onset at birth, hypotonia Generalized muscle weakness Joint contractures Microphthalmos, retinal dysplasia, buphthalmos, glaucoma, cataracts Seizures, MR	Serum CK 5–20 x normal MRI shows cobblestone lissencephaly, hydrocephalus, encephalocele, absent corpus callosum	*O*-mannoxyl-transferase-1 (POMT1)

[a]All are inherited as recessive traits. [b]There is phenotypic overlap between disorders related to defective glycosylation. In muscle, this is a consequence of altered glycosylation of dystroglycans; in brain/eye, other glycosylated proteins are involved. Clinically, Walker-Warburg syndrome is more severe, with death by 1 year. ***Abbreviations:*** CK, creatine kinase; EMG, electromyography; LGMD, limb-girdle muscular dystrophy; MR, mental retardation; NCS, nerve conduction studies.

TREATMENT: CONGENITAL MUSCULAR DYSTROPHY

There is no specific treatment for CMD. Proper wheelchair seating is important. Management of epilepsy and cardiac manifestations is necessary for some patients.

Myotonic Dystrophy

Myotonic dystrophy is also known as *dystrophia myotonica* (DM). The condition is composed of at least two clinical disorders with overlapping phenotypes and distinct molecular genetic defects: myotonic dystrophy type 1 (DM1), the classic disease originally described by Steinert, and myotonic dystrophy type 2 (DM2), also called *proximal myotonic myopathy* (PROMM).

CLINICAL FEATURES

The clinical expression of DM1 varies widely and involves many systems other than muscle. Affected patients have a typical "hatchet-faced" appearance due to temporalis, masseter, and facial muscle atrophy and weakness. Frontal baldness is also characteristic of the disease. Neck muscles, including flexors and sternocleidomastoids, and distal limb muscles are involved early. Weakness of wrist extensors, finger extensors, and intrinsic hand muscles impairs function. Ankle dorsiflexor weakness may cause footdrop. Proximal muscles remain stronger throughout the course, although preferential atrophy and weakness of quadriceps muscles occur in many patients. Palatal, pharyngeal, and tongue involvement produce a dysarthric speech, nasal voice, and swallowing problems. Some patients have diaphragm and intercostal muscle weakness, resulting in respiratory insufficiency.

Myotonia, which usually appears by age 5 years, is demonstrable by percussion of the thenar eminence, the tongue, and wrist extensor muscles. Myotonia causes a slow relaxation of hand grip after a forced voluntary closure. Advanced muscle wasting makes myotonia more difficult to detect.

Cardiac disturbances occur commonly in patients with DM1. ECG abnormalities include first-degree heart block and more extensive conduction system involvement. Complete heart block and sudden death can occur. Congestive heart failure occurs infrequently but may result from cor pulmonale secondary to respiratory failure. Mitral valve prolapse also occurs commonly. Other associated features include intellectual impairment, hypersomnia, posterior subcapsular cataracts, gonadal atrophy, insulin resistance, and decreased esophageal and colonic motility.

Congenital myotonic dystrophy is a more severe form of DM1 and occurs in ~25% of infants of affected mothers. It is characterized by severe facial and bulbar weakness, transient neonatal respiratory insufficiency, and mental retardation.

DM2, or PROMM, has a distinct pattern of muscle weakness affecting mainly proximal muscles. Other features of the disease overlap with DM1, including cataracts, testicular atrophy, insulin resistance, constipation, hypersomnia, and cognitive defects. Cardiac conduction defects occur but are less common, and the hatchet face and frontal baldness are less consistent features. A very striking difference is the failure to clearly identify a congenital form of DM2.

LABORATORY FEATURES

The diagnosis of myotonic dystrophy can usually be made on the basis of clinical findings. Serum CK levels may be normal or mildly elevated. EMG evidence of myotonia is present in most cases of DM1 but may be more patchy in DM2. Muscle biopsy shows muscle atrophy, which selectively involves type 1 fibers in 50% of cases, and ringed fibers in DM1 but not in DM2. Typically, numerous internalized nuclei can be seen in individual muscle fibers as well as atrophic fibers with pyknotic nuclear clumps in both DM1 and DM2. Necrosis of muscle fibers and increased connective tissue, common in other muscular dystrophies, are less apparent in myotonic dystrophy.

DM1 and DM2 are both autosomal dominant disorders. New mutations do not appear to contribute to the pool of affected individuals. DM1 is transmitted by an intronic mutation consisting of an unstable expansion of a CTG trinucleotide repeat in a

serine-threonine protein kinase gene (named *DMPK*) on chromosome 19q13.3. An increase in the severity of the disease phenotype in successive generations (genetic anticipation) is accompanied by an increase in the number of trinucleotide repeats. A similar type of mutation has been identified in fragile X syndrome (Chap. 61). The unstable triplet repeat in myotonic dystrophy can be used for prenatal diagnosis. Congenital disease occurs almost exclusively in infants born to affected mothers; it is possible that sperm with greatly expanded triplet repeats do not function well.

DM2 is caused by a DNA expansion mutation consisting of a CCTG repeat in intron 1 of the *ZNF9* gene located at chromosome 3q13.3-q24. The gene is believed to encode an RNA-binding protein expressed in many different tissues, including skeletal and cardiac muscle.

The DNA expansions in DM1 and DM2 almost certainly impair muscle function by a toxic gain of function of the mutant mRNA. In both DM1 and DM2, the mutant RNA appears to form intranuclear inclusions composed of aberrant RNA. These RNA inclusions sequester RNA-binding proteins essential for proper splicing of a variety of other mRNAs. This leads to abnormal transcription of multiple proteins in a variety of tissues/organ systems, in turn causing the systemic manifestations of DM1 and DM2.

TREATMENT: MYOTONIC DYSTROPHY

The myotonia in DM1 rarely warrants treatment, though some patients with DM2 are significantly bothered by the discomfort related to the associated muscle stiffness. Phenytoin and mexiletine are the preferred agents for the occasional patient who requires an antimyotonia drug; other agents, particularly quinine and procainamide, may worsen cardiac conduction. A cardiac pacemaker should be considered for patients with unexplained syncope, advanced conduction system abnormalities with evidence of second-degree heart block, or trifascicular conduction disturbances with marked prolongation of the PR interval. Molded ankle-foot orthoses help stabilize gait in patients with foot drop. Excessive daytime somnolence with or without sleep apnea is not uncommon. Sleep studies, noninvasive respiratory support (biphasic positive airway pressure, BiPAP), and treatment with modafinil may be beneficial.

Facioscapulohumeral (FSH) Muscular Dystrophy

This form of muscular dystrophy has a prevalence of ~1 in 20,000. There are two forms of FSHD that have similar pathogenesis, as will be discussed. Most patients have FSHD type 1 (95%), while approximately 5% have FSHD2. FSHD1 and FSHD2 are clinically and histopathologically identical. FSHD is not to be confused with the genetically distinct scapuloperoneal dystrophies.

CLINICAL FEATURES

The condition typically has an onset in childhood or young adulthood. In most cases, facial weakness is the initial manifestation, appearing as an inability to smile, whistle, or fully close the eyes. Weakness of the shoulder girdles, rather than the facial muscles, usually brings the patient to medical attention. Loss of scapular stabilizer muscles makes arm elevation difficult. Scapular winging (Fig. 387-3) becomes apparent with attempts at abduction and forward movement of the arms. Biceps and triceps muscles may be severely affected, with relative sparing of the deltoid muscles. Weakness is invariably worse for wrist extension than for wrist flexion, and weakness of the anterior compartment muscles of the legs may lead to footdrop.

In most patients, the weakness remains restricted to facial, upper extremity, and distal lower extremity muscles. In 20% of patients, weakness progresses to involve the pelvic girdle muscles, and severe functional impairment and possible wheelchair dependency result.

Characteristically, patients with FSHD do not have involvement of other organ systems, although labile hypertension is common, and there is an increased incidence of nerve deafness. *Coats' disease*, a disorder consisting of telangiectasia, exudation, and retinal detachment, also occurs.

LABORATORY FEATURES

The serum CK level may be normal or mildly elevated. EMG usually indicates a myopathic pattern. The muscle biopsy shows nonspecific features of a myopathy. A prominent inflammatory infiltrate, which is often multifocal in distribution, is present in some biopsy samples. The cause or significance of this finding is unknown.

An autosomal dominant inheritance pattern with almost complete penetrance has been established, but each family member should be examined for the presence of the disease, since ~30% of those affected are unaware of involvement. FSHD1 is associated with deletions of tandem 3.3-kb repeats at 4q35. The deletion reduces the number of repeats to a fragment of <35 kb in most patients. Within these repeats lies the *DUX4* gene, which usually is not expressed. In patients with FSHD1 these deletions in the setting of a specific polymorphism leads to hypomethylation of the region and toxic expression of the *DUX 4* gene. Interestingly, in patients with FSHD2, there is no deletion but in the setting of the same polymorphism there again is seen hypomethylation of the region and the permissive expression of the *DUX4* gene. In either instance, the permissive polymorphism introduces a polyadenylation signal that results in an aberrant, toxic *DUX4* transcript.

TREATMENT: FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY

No specific treatment is available; ankle-foot orthoses are helpful for footdrop. Scapular stabilization procedures improve scapular winging but may not improve function.

Oculopharyngeal Dystrophy

This form of muscular dystrophy represents one of several disorders characterized by progressive external ophthalmoplegia, which consists of slowly progressive ptosis and limitation of eye movements with sparing of pupillary reactions for light and accommodation. Patients usually do not complain of diplopia, in contrast to patients having conditions with a more acute onset of ocular muscle weakness (e.g., myasthenia gravis).

CLINICAL FEATURES

Oculopharyngeal muscular dystrophy has a late onset; it usually presents in the fourth to sixth decade with ptosis and/or dysphagia. The extraocular muscle impairment is less prominent in the early phase but may be severe later. The swallowing problem may become debilitating and result in pooling of secretions and repeated episodes of aspiration. Mild weakness of the neck and extremities also occurs.

LABORATORY FEATURES

The serum CK level may be two to three times normal. Myopathic EMG findings are typical. On biopsy, muscle fibers are found to contain rimmed vacuoles, which by electron microscopy are shown to contain membranous whorls, accumulation of glycogen, and other nonspecific debris related to lysosomes. A distinct feature of oculopharyngeal dystrophy is the presence of tubular filaments, 8.5 nm in diameter, in muscle cell nuclei.

Oculopharyngeal dystrophy has an autosomal dominant inheritance pattern with complete penetrance. The incidence is high in French-Canadians and in Spanish-American families of the southwestern United States. Large kindreds of Italian and of eastern European Jewish descent have been reported. The molecular defect in oculopharyngeal muscular dystrophy is a subtle expansion of a modest polyalanine repeat tract in a poly-RNA-binding protein (PABP2) in muscle.

TREATMENT: OCULOPHARYNGEAL DYSTROPHY

Dysphagia can lead to significant undernourishment and inanition, making oculopharyngeal muscular dystrophy a potentially life-threatening disease. Cricopharyngeal myotomy may improve swallowing, although it does not prevent aspiration. Eyelid crutches can improve vision when ptosis obstructs vision; candidates for ptosis surgery must be carefully selected—those with severe facial weakness are not suitable.

Distal Myopathies

A group of muscle diseases, the distal myopathies, are notable for their preferential distal distribution of muscle weakness in contrast to most muscle conditions associated with proximal weakness. The major distal myopathies are summarized in **Table 387-9**.

Table 387-9 Distal Myopathies

Disease	Clinical Features	Laboratory Features	Inheritance/Locus or Gene
Welander's distal myopathy	Onset in 5th decade Weakness begins in hands Slow progression with spread to distal lower extremities Lifespan normal	Serum CK 2–3 x normal EMG myopathic NCS normal Muscle biopsy shows dystrophic features	AD Chromosome 2p13
Tibial muscular dystrophy (Udd's)	Onset 4th to 8th decade Distal lower extremity weakness (tibial distribution) Upper extremities usually normal Lifespan normal	Serum CK 2–4 x normal EMG myopathic NCS normal Muscle biopsy shows dystrophic features Titin absent in M-line of muscle	AD Titin AR (associated with more proximal weakness —LGMD2J)
Markesbery-Griggs distal myopathy	Onset 4th to 8th decade Distal lower extremity weakness (tibial distribution) with progression to distal arms and proximal muscles	Serum CK is usually mildly elevated EMG reveals irritative myopathy Muscle biopsies demonstrate rimmed vacuoles and features of myofibrillar myopathy	AD Z-band alternatively spliced PDX motif-containing protein (ZASP)
Laing's distal myopathy	Onset childhood to 3rd decade Distal lower extremity weakness (tibial distribution) and neck flexors affected early May have cardiomyopathy	Serum CK is normal or slightly elevated Muscle biopsies do not show rimmed vacuoles Large deposits of myosin heavy chain are seen in type 1 muscle fibers	AD Myosin heavy chain 7
Nonaka's distal myopathy (autosomal recessive hereditary inclusion body myopathy)	Onset: 2nd to 3rd decade Lower extremity distal weakness Mild distal upper limb weakness may be present early Progression to other muscles sparing quadriceps Ambulation may be lost in 10–15 y	Serum CK 3–10 x normal EMG myopathic NCS normal Dystrophic features on muscle biopsy plus rimmed vacuoles and 15- to 19-nm filaments within vacuoles	AR *GNE* gene: UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase Allelic to hereditary inclusion body myopathy
Miyoshi's myopathy	Onset: 2nd to 3rd decade Lower extremity weakness in posterior compartment muscles Progression leads to weakness in other muscle groups Ambulation lost after 10–15 years in about one-third of cases	Serum CK 20–100 x normal EMG myopathic NCS normal Muscle biopsy shows nonspecific dystrophic features often with prominent inflammatory cell infiltration; no rimmed vacuoles	AR Allelic to LGMD2B (see Table 387-7) Dysferlin
Myofibrillar myopathies	Onset from early childhood to late adult life Weakness may be distal, proximal, or generalized Cardiomyopathy and respiratory involvement is not uncommon	Serum CKs can be normal or moderately elevated EMG is myopathic and often associated with myotonic discharges Muscle biopsy demonstrates abnormal accumulation of desmin and other proteins, rimmed vacuoles, and myofibrillar degeneration	Genetically heterogeneous AD Myotilin (also known as LGMD 1A) ZASP (see Markesbery-Griggs distal myopathy) Filamin-C Desmin Alpha B crystallin Bag3 FHL-1 AR Desmin Selenoprotein N1

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; CK, creatine kinase; EMG, electromyography; NCS, nerve

conduction studies.

CLINICAL FEATURES

Welander's, *Udd's*, and *Markesbery-Griggs distal myopathies* are all late onset, dominantly inherited disorders of distal limb muscles, usually beginning after age 40 years. Welander's distal myopathy preferentially involves the wrist and finger extensors, whereas the others are associated with anterior tibial weakness leading to progressive footdrop. *Laing's distal myopathy* is also a dominantly inherited disorder heralded by tibial weakness; however, it is distinguished by onset in childhood or early adult life. *Nonaka's distal myopathy* and *Miyoshi's myopathy* are distinguished by autosomal recessive inheritance and onset in the late teens or twenties. Nonaka's myopathy entails anterior tibial weakness, whereas Miyoshi's myopathy is unique in that gastrocnemius muscles are preferentially affected at onset. Finally, the *myofibrillar myopathies* (MFMs) are a clinically and genetically heterogeneous group of disorders that can be associated with prominent distal weakness; they can be inherited in an autosomal dominant or recessive pattern. Of note, Markesbery-Griggs myopathy (caused by mutations in ZASP) and LGMD1B (caused by mutations in myotilin) are in fact subtypes of myofibrillar myopathy.

LABORATORY FEATURES

Serum CK level is particularly helpful in diagnosing Miyoshi's myopathy since it is very elevated. In the other conditions, serum CK is only slightly increased. EMGs are myopathic. In the MFMs, myotonic or pseudomyotonic discharges are common. Muscle biopsy shows nonspecific dystrophic features and, with the exception of Laing's and Miyoshi's myopathies, often shows rimmed vacuoles. MFM is associated with the accumulation of dense inclusions, as well as amorphous material best seen on Gomori's trichrome and myofibrillar disruption on electron microscopy. Immune staining sometimes demonstrates accumulation of desmin and other proteins in MFM, large deposits of myosin heavy chain in the subsarcolemmal region of type 1 muscle fibers in Laing's myopathy, and reduced or absent dysferlin in Miyoshi's myopathy.

The affected genes and their gene products are listed in Table 387-9. The gene for Welander's disease awaits identification.

TREATMENT: DISTAL MYOPATHIES

Occupational therapy is offered for loss of hand function; ankle-foot orthoses can support distal lower limb muscles. The MFMs can be associated with cardiomyopathy (congestive heart failure or arrhythmias) and respiratory failure that may require medical management. Laing's-type distal myopathy can also be associated with a cardiomyopathy.

CONGENITAL MYOPATHIES

These rare disorders are distinguished from muscular dystrophies by the presence of specific histochemical and structural abnormalities in muscle. Although primarily disorders of infancy or childhood, three forms that may present in adulthood are described here: central core disease, nemaline (rod) myopathy, and centronuclear (myotubular) myopathy. Sarcotubular myopathy is caused by mutations in TRIM-32 and is identical to LGMD2H. Other types, such as minicore myopathy (multi-minicore disease), fingerprint body myopathy, and cap myopathy, are not discussed.

Central Core Disease

Patients with central core disease may have decreased fetal movements and breech presentation. Hypotonia and delay in motor milestones, particularly in walking, are common. Later in childhood, patients develop problems with stair climbing, running, and getting up from the floor. On examination, there is mild facial, neck-flexor, and proximal-extremity muscle weakness. Legs are more affected than arms. Skeletal abnormalities include congenital hip dislocation, scoliosis, and pes cavus; clubbed feet also occur. Most cases are nonprogressive, but exceptions are well documented. Susceptibility to malignant hyperthermia must be considered as a potential risk factor for patients with central core disease.

The serum CK level is usually normal. Needle EMG demonstrates a myopathic pattern. Muscle biopsy shows fibers with single or multiple central or eccentric discrete zones (*cores*) devoid of oxidative enzymes. Cores occur preferentially in type 1 fibers and represent poorly aligned sarcomeres associated with Z disk streaming.

Autosomal dominant inheritance is characteristic; sporadic cases also occur. The disease is caused by point mutations of the ryanodine receptor gene on chromosome 19q, encoding the calcium-release channel of the sarcoplasmic reticulum of skeletal muscle; mutations of this gene also account for some cases of inherited malignant hyperthermia (Chap. 16). Malignant hyperthermia is an allelic condition; C-terminal mutations of the *RYR1* gene predispose to this complication.

Specific treatment is not required, but establishing a diagnosis of central core disease is extremely important because these patients have a known predisposition to malignant hyperthermia during anesthesia.

Nemaline Myopathy

The term *nemaline* refers to the distinctive presence in muscle fibers of rods or threadlike structures (Greek *nema*, "thread"). Nemaline myopathy is clinically heterogeneous. A severe neonatal form presents with hypotonia and feeding and respiratory difficulties, leading to early death. Nemaline myopathy usually presents in infancy or childhood with delayed motor milestones. The course is nonprogressive or slowly progressive. The physical appearance is striking because of the long, narrow facies, high-arched palate, and open-mouthed appearance due to a prognathous jaw. Other skeletal abnormalities include pectus excavatum, kyphoscoliosis, pes cavus, and clubfoot deformities. Facial and generalized muscle weakness, including respiratory muscle weakness, is common. An adult-onset disorder with progressive proximal weakness may be seen. Myocardial involvement is occasionally present in both the childhood and adult-onset forms. The serum CK level is usually normal or slightly elevated. The EMG demonstrates a myopathic pattern. Muscle biopsy shows clusters of small rods (nemaline bodies), which occur preferentially, but not exclusively, in the sarcoplasm of type 1 muscle fibers. Occasionally, the rods are also apparent in myonuclei. The muscle often shows type 1 muscle fiber predominance. Rods originate from the Z disk material of the muscle fiber.

Six genes have been associated with nemaline myopathy. Five of these code for thin filament–associated proteins, suggesting disturbed assembly or interplay of these structures as a pivotal mechanism. Mutations of the nebulin (*NEB*) gene account for most cases, including both severe neonatal and early childhood forms, inherited as autosomal recessive disorders. Neonatal and childhood cases, inherited as predominantly autosomal dominant disorders, are caused by mutations of the skeletal muscle α-actinin (*ACTA1*) gene. In milder forms of the disease with autosomal dominant inheritance, mutations have been identified in both the slow α-tropomyosin (*TPM3*) and ß-tropomyosin (*TPM2*) genes accounting for <3% of cases. Muscle troponin T (*TNNT1*) gene mutations appear to be limited to the Amish population in North America. Mutations in a sixth nemaline myopathy gene, *NEM6*, have recently been reported; this gene encodes a putative BTB/Kelch protein. No specific treatment is available.

Centronuclear (Myotubular) Myopathy

Three distinct variants of centronuclear myopathy occur. A neonatal form, also known as *myotubular myopathy* , presents with severe hypotonia and weakness at birth. The late infancy–early childhood form presents with delayed motor milestones. Later, difficulty with running and stair climbing becomes apparent. A marfanoid, slender body habitus, long narrow face, and high-arched palate are typical. Scoliosis and clubbed feet may be present. Most patients exhibit progressive weakness, some requiring wheelchairs. Progressive external ophthalmoplegia with ptosis and varying degrees of extraocular muscle impairment

are characteristic of both the neonatal and the late-infantile forms. A third variant, the late childhood–adult form, has an onset in the second or third decade. Patients have full extraocular muscle movements and rarely exhibit ptosis. There is mild, slowly progressive limb weakness that may be distally predominant [some of these patients have been classified as having Charcot-Marie-Tooth disease type 2 (CMT2; Chap. 384)].

Normal or slightly elevated CK levels occur in each of the forms. Nerve conduction studies may reveal reduced amplitudes of distal compound muscle action potentials, in particular in adult-onset cases that resemble CMT2. EMG studies often give distinctive results, showing positive sharp waves and fibrillation potentials, complex and repetitive discharges, and rarely myotonic discharges. Muscle biopsy specimens in longitudinal section demonstrate rows of central nuclei, often surrounded by a halo. In transverse sections, central nuclei are found in 25–80% of muscle fibers.

A gene for the neonatal form of centronuclear myopathy has been localized to Xq28; this gene encodes myotubularin, a protein tyrosine phosphatase. Missense, frameshift, and splice-site mutations predict loss of myotubularin function in affected individuals. Carrier identification and prenatal diagnosis are possible. Autosomal recessive forms are caused by mutations in *BIB1* that encodes for amphyphysin-2, while some autosomal dominant cases, which are allelic to a form of CMT2, are associated with mutations in the gene that encodes dynamin-2. No specific medical treatments are available at this time.

DISORDERS OF MUSCLE ENERGY METABOLISM

There are two principal sources of energy for skeletal muscle—fatty acids and glucose. Abnormalities in either glucose or lipid utilization can be associated with distinct clinical presentations that can range from an acute, painful syndrome with rhabdomyolysis and myoglobinuria to a chronic, progressive muscle weakness simulating muscular dystrophy.

Glycogen Storage and Glycolytic Defects

DISORDERS OF GLYCOGEN STORAGE CAUSING PROGRESSIVE WEAKNESS

α-Glucosidase, or Acid Maltase, Deficiency (Pompe's Disease)

Three clinical forms of α-glucosidase, or acid maltase, deficiency (*type II glycogenosis*) can be distinguished. The infantile form is the most common, with onset of symptoms in the first 3 months of life. Infants develop severe muscle weakness, cardiomegaly, hepatomegaly, and respiratory insufficiency. Glycogen accumulation in motor neurons of the spinal cord and brainstem contributes to muscle weakness. Death usually occurs by 1 year of age. In the childhood form, the picture resembles muscular dystrophy. Delayed motor milestones result from proximal limb muscle weakness and involvement of respiratory muscles. The heart may be involved, but the liver and brain are unaffected. The adult form usually begins in the third or fourth decade but can present as late as the seventh decade. Respiratory failure and diaphragmatic weakness are often initial manifestations, heralding progressive proximal muscle weakness. The heart and liver are not involved.

The serum CK level is 2 to 10 times normal in infantile or childhood-onset Pompe's disease but can be normal in adult-onset cases. EMG examination demonstrates a myopathic pattern, but other features are especially distinctive, including myotonic discharges, trains of fibrillation and positive waves, and complex repetitive discharges. EMG discharges are very prominent in the paraspinal muscles. The muscle biopsy in infants typically reveals vacuoles containing glycogen and the lysosomal enzyme acid phosphatase. Electron microscopy reveals membrane-bound and free tissue glycogen. However, muscle biopsies in late-onset Pompe's disease may demonstrate only nonspecific abnormalities. Enzyme analysis of dried blood spots is a sensitive technique to screen for Pompe's disease. A definitive diagnosis is established by enzyme assay in muscle or cultured fibroblasts or by genetic testing.

Pompe's disease is inherited as an autosomal recessive disorder caused by mutations of the α-glucosidase gene. Enzyme replacement therapy (ERT) with IV recombinant human α-glucosidase has been shown to be beneficial in infantile-onset Pompe's disease. Clinical benefits in the infantile disease include reduced heart size, improved muscle function, reduced need for ventilatory support, and longer life. In late-onset cases, ERT has not been associated with the dramatic response that can be seen in classic infantile Pompe's disease, yet it appears to stabilize the disease process.

Other Glycogen Storage Diseases with Progressive Weakness

In *debranching enzyme deficiency* (*type III glycogenosis*), a slowly progressive form of muscle weakness can develop after puberty. Rarely, myoglobinuria may be seen. Patients are usually diagnosed in infancy, however, because of hypotonia and delayed motor milestones, hepatomegaly, growth retardation, and hypoglycemia. *Branching enzyme deficiency* (*type IV glycogenosis*) is a rare and fatal glycogen storage disease characterized by failure to thrive and hepatomegaly. Hypotonia and muscle wasting may be present, but the skeletal muscle manifestations are minor compared to liver failure.

DISORDERS OF GLYCOLYSIS CAUSING EXERCISE INTOLERANCE

Several glycolytic defects are associated with recurrent myoglobinuria: *myophosphorylase deficiency* (*type V glycogenosis*), *phosphofructokinase deficiency* (*type VII glycogenosis*), *phosphoglycerate kinase deficiency* (*type IX glycogenosis*), *phosphoglycerate mutase deficiency* (*type X glycogenosis*), *lactate dehydrogenase deficiency* (*glycogenosis type XI*), and β-*enolase deficiency*. Myophosphorylase deficiency, also known as *McArdle's disease*, is by far the most common of the glycolytic defects associated with exercise intolerance. These glycolytic defects result in a common failure to support energy production at the initiation of exercise, although the exact site of energy failure remains controversial.

Clinical muscle manifestations in these conditions usually begin in adolescence. Symptoms are precipitated by brief bursts of high-intensity exercise such as running or lifting heavy objects. A history of myalgia and muscle stiffness usually precedes the intensely painful muscle contractures, which may be followed by myoglobinuria. Acute renal failure accompanies significant pigmenturia.

Certain features help distinguish some enzyme defects. In McArdle's disease, exercise tolerance can be enhanced by a slow induction phase (warm-up) or brief periods of rest, allowing for the start of the "second-wind" phenomenon (switching to utilization of fatty acids). Varying degrees of hemolytic anemia accompany deficiencies of both phosphofructokinase (mild) and phosphoglycerate kinase (severe). In phosphoglycerate kinase deficiency, the usual clinical presentation is a seizure disorder associated with mental retardation; exercise intolerance is an infrequent manifestation.

In all of these conditions, the serum CK levels fluctuate widely and may be elevated even during symptom-free periods. CK levels >100 times normal are expected, accompanying myoglobinuria. All patients with suspected glycolytic defects leading to exercise intolerance should undergo a forearm exercise test. An impaired rise in venous lactate is highly indicative of a glycolytic defect. In lactate dehydrogenase deficiency, venous levels of lactate do not increase, but pyruvate rises to normal. A definitive diagnosis of glycolytic disease is made by muscle biopsy and subsequent enzyme analysis or by genetic testing.

Myophosphorylase deficiency, phosphofructokinase deficiency, and phosphoglycerate mutase deficiency are inherited as autosomal recessive disorders. Phosphoglycerate kinase deficiency is X-linked recessive. Mutations can be found in the respective genes encoding the abnormal proteins in each of these disorders.

Training may enhance exercise tolerance, perhaps by increasing perfusion to muscle. Dietary intake of free glucose or fructose prior to activity may improve function but care must be taken to avoid obesity from ingesting too many calories.

Lipid as an Energy Source and Associated Defects

Lipid is an important muscle energy source during rest and during prolonged, submaximal exercise. Fatty acids are derived from circulating very low-density lipoprotein (VLDL) in the blood or from triglycerides stored in muscle fibers. Oxidation of fatty acids occurs in the mitochondria. To enter the mitochondria, a fatty acid must first be converted to an "activated fatty acid," acyl-CoA. The acyl-CoA must be linked with carnitine by the enzyme carnitine palmitoyltransferase (CPT) I for transport into the mitochondria. CPT I is present on the inner side of the outer mitochondrial membrane. Carnitine is removed by CPT II, an enzyme attached to the inside of the inner mitochondrial membrane, allowing transport of acyl-CoA into the mitochondrial matrix for β-oxidation.

CARNITINE PALMITOYLTRANSFERASE DEFICIENCY

CPT II deficiency is the most common recognizable cause of recurrent myoglobinuria, more common than the glycolytic defects. Onset is usually in the teenage years or early twenties. Muscle pain and myoglobinuria typically occur after prolonged exercise but can also be precipitated by fasting or infections; up to 20% of patients do not exhibit myoglobinuria, however. Strength is normal between attacks. In contrast to disorders caused by defects in glycolysis, in which muscle cramps follow short, intense bursts of exercise, the muscle pain in CPT II deficiency does not occur until the limits of utilization have been exceeded and muscle breakdown has already begun. Episodes of rhabdomyolysis may produce severe weakness. In young children and newborns, CPT II deficiency can present with a very severe clinical picture including hypoketotic hypoglycemia, cardiomyopathy, liver failure, and sudden death.

Serum CK levels and EMG findings are both usually normal between episodes. A normal rise of venous lactate during forearm exercise distinguishes this condition from glycolytic defects, especially myophosphorylase deficiency. Muscle biopsy does not show lipid accumulation and is usually normal between attacks. The diagnosis requires direct measurement of muscle CPT or genetic testing.

CPT II deficiency is much more common in men than women (5:1); nevertheless, all evidence indicates autosomal recessive inheritance. A mutation in the gene for CPT II (chromosome 1p36) causes the disease in some individuals. Attempts to improve exercise tolerance with frequent meals and a low-fat, high-carbohydrate diet, or by substituting medium-chain triglycerides in the diet, have not proven to be beneficial.

MYOADENYLATE DEAMINASE DEFICIENCY

The muscle enzyme myoadenylate deaminase converts adenosine-5'-monophosphate (5'-AMP) to inosine monophosphate (IMP) with liberation of ammonia. Myoadenylate deaminase may play a role in regulating adenosine triphosphate (ATP) levels in muscles. Most individuals with myoadenylate deaminase deficiency have no symptoms. There have been a few reports of patients with this disorder who have exercise-exacerbated myalgia and myoglobinuria. Many questions have been raised about the clinical effects of myoadenylate deaminase deficiency, and, specifically, its relationship to exertional myalgia and fatigability, but there is no consensus.

MITOCHONDRIAL MYOPATHIES

In 1972, Olson and colleagues recognized that muscle fibers with significant numbers of abnormal mitochondria could be highlighted with the modified trichrome stain; the term *ragged red fibers* was coined. By electron microscopy, the mitochondria in ragged red fibers are enlarged and often bizarrely shaped and have crystalline inclusions. Since that seminal observation, the understanding of these disorders of muscle and other tissues has expanded (Chap. 61).

Mitochondria play a key role in energy production. Oxidation of the major nutrients derived from carbohydrate, fat, and protein leads to the generation of reducing equivalents. The latter are transported through the respiratory chain in the process known as *oxidative phosphorylation*. The energy generated by the oxidation-reduction reactions of the respiratory chain is stored in an electrochemical gradient coupled to ATP synthesis.

A novel feature of mitochondria is their genetic composition. Each mitochondrion possesses a DNA genome that is distinct from that of the nuclear DNA. Human mitochondrial DNA (mtDNA) consists of a double-strand, circular molecule comprising 16,569 base pairs. It codes for 22 transfer RNAs, 2 ribosomal RNAs, and 13 polypeptides of the respiratory chain enzymes. The genetics of mitochondrial diseases differ from the genetics of chromosomal disorders. The DNA of mitochondria is directly inherited from the cytoplasm of the gametes, mainly from the oocyte. The sperm contributes very little of its mitochondria to the offspring at the time of fertilization. Thus, mitochondrial genes are derived almost exclusively from the mother, accounting for maternal inheritance of some mitochondrial disorders.

Patients with mitochondrial myopathies have clinical manifestations that usually fall into three groups: chronic progressive external ophthalmoplegia (CPEO), skeletal muscle–CNS syndromes, and pure myopathy simulating muscular dystrophy or metabolic myopathy.

Progressive External Ophthalmoplegia Syndromes with Ragged Red Fibers

The single most common sign of a mitochondrial myopathy is CPEO, occurring in >50% of all mitochondrial myopathies. Varying degrees of ptosis and weakness of extraocular muscles are seen, usually in the absence of diplopia, a point of distinction from disorders with fluctuating eye weakness (e.g., myasthenia gravis).

Kearns-Sayre Syndrome (Kss)

KSS is a widespread multiorgan system disorder with a defined triad of clinical findings: onset before age 20, CPEO, and pigmentary retinopathy, plus one or more of the following features: complete heart block, cerebrospinal fluid (CSF) protein >1 g/L (100 mg/dL), or cerebellar ataxia. Some patients with CPEO and ragged red fibers may not fulfill all of the criteria for KSS. The cardiac disease includes syncopal attacks and cardiac arrest related to the abnormalities in the cardiac conduction system: prolonged intraventricular conduction time, bundle branch block, and complete atrioventricular block. Death attributed to heart block occurs in ~20% of the patients. Varying degrees of progressive limb muscle weakness and easy fatigability affect activities of daily living. Endocrine abnormalities are common, including gonadal dysfunction in both sexes with delayed puberty, short stature, and infertility. Diabetes mellitus is a cardinal sign of mitochondrial disorders and is estimated to occur in 13% of KSS patients. Other less common endocrine disorders include thyroid disease, hyperaldosteronism, Addison's disease, and hypoparathyroidism. Both mental retardation and dementia are common accompaniments to this disorder. Serum CK levels are normal or slightly elevated. Serum lactate and pyruvate levels may be elevated. EMG is myopathic. Nerve conduction studies may be abnormal related to an associated neuropathy. Muscle biopsies reveal ragged red fibers, highlighted in oxidative enzyme stains, many showing defects in cytochrome oxidase. By electron microscopy there are increased numbers of mitochondria that often appear enlarged and contain paracrystalline inclusions.

KSS is a sporadic disorder. The disease is caused by single mtDNA deletions presumed to arise spontaneously in the ovum or zygote. The most common deletion, occurring in about one-third of patients, removes 4,977 bp of contiguous mtDNA. Monitoring for cardiac conduction defects is critical. Prophylactic pacemaker implantation is indicated when ECGs demonstrate a bifascicular block. In KSS, no benefit has been shown for supplementary therapies, including multivitamins or coenzyme Q10. Of all the proposed options, exercise might be the most applicable but must be approached cautiously because of defects in the

cardiac conduction system.

Progressive External Ophthalmoplegia (PEO)

This condition is caused by nuclear DNA mutations affecting mtDNA copy number and integrity and is thus inherited in a Mendelian fashion. Onset is usually after puberty. Fatigue, exercise intolerance, and complaints of muscle weakness are typical. Some patients notice swallowing problems. The neurologic examination confirms the ptosis and ophthalmoplegia, usually asymmetric in distribution. A sensorineural hearing loss may be encountered. Mild facial, neck flexor, and proximal weakness are typical. Rarely, respiratory muscles may be progressively affected and may be the direct cause of death. Serum CK is normal or mildly elevated. The resting lactate level is normal or slightly elevated but may rise excessively after exercise. CSF protein is normal. The EMG is myopathic, and nerve conduction studies are usually normal. Ragged red fibers are prominently displayed in the muscle biopsy. Southern blots of muscle reveal a normal mtDNA band at 16.6 kb and several additional mtDNA deletion bands with genomes varying from 0.5 to 10 kb.

This autosomal dominant form of CPEO has been linked to loci on three chromosomes: 4q35, 10q24, and 15q22–26. In the chromosome 4q-related form of disease, mutations of the gene encoding the heart and skeletal muscle–specific isoform of the adenine nucleotide translocator 1 (*ANT1*) gene are found. This highly abundant mitochondrial protein forms a homodimeric inner mitochondrial channel through which adenosine diphosphate (ADP) enters and ATP leaves the mitochondrial matrix. In the chromosome 10q–related disorder, mutations of the gene *C10orf2* are found. Its gene product, *twinkle*, co-localizes with the mtDNA and is named for its punctate, starlike staining properties. The function of twinkle is presumed to be critical for lifetime maintenance of mitochondrial integrity. In the cases mapped to chromosome 15q, a mutation affects the gene encoding mtDNA polymerase (*POLG*), an enzyme important in mtDNA replication. Autosomal recessive PEO has also been described with mutations in the *POLG* gene. Point mutations have been identified within various mitochondrial tRNA (Leu, Ile, Asn, Trp) genes in families with maternal inheritance of PEO.

Exercise may improve function but will depend on the patient's ability to participate.

Mitochondrial DNA Skeletal Muscle–Central Nervous System Syndromes

MYOCLONIC EPILEPSY WITH RAGGED RED FIBERS (MERRF)

The onset of MERRF is variable, ranging from late childhood to middle adult life. Characteristic features include myoclonic epilepsy, cerebellar ataxia, and progressive muscle weakness. The seizure disorder is an integral part of the disease and may be the initial symptom. Cerebellar ataxia precedes or accompanies epilepsy. It is slowly progressive and generalized. The third major feature of the disease is muscle weakness in a limb-girdle distribution. Other more variable features include dementia, peripheral neuropathy, optic atrophy, hearing loss, and diabetes mellitus.

Serum CK levels are normal or slightly increased. The serum lactate may be elevated. EMG is myopathic, and in some patients nerve conduction studies show a neuropathy. The electroencephalogram is abnormal, corroborating clinical findings of epilepsy. Typical ragged red fibers are seen on muscle biopsy. MERRF is caused by maternally inherited point mutations of mitochondrial tRNA genes. The most common mutation found in 80% of MERRF patients is an A to G substitution at nucleotide 8344 of tRNA lysine (A8344G $tRNA^{lys}$). Other $tRNA^{lys}$ mutations include base-pair substitutions T8356C and G8363A. Only supportive treatment is possible, with special attention to epilepsy.

MITOCHONDRIAL MYOPATHY, ENCEPHALOPATHY, LACTIC ACIDOSIS, AND STROKELIKE EPISODES (MELAS)

MELAS is the most common mitochondrial encephalomyopathy. The term *strokelike* is appropriate because the cerebral lesions do not conform to a strictly vascular distribution. The onset in the majority of patients is before age 20. Seizures, usually partial motor or generalized, are common and may represent the first clearly recognizable sign of disease. The cerebral insults that resemble strokes cause hemiparesis, hemianopia, and cortical blindness. A presumptive stroke occurring before age 40 should place this mitochondrial encephalomyopathy high in the differential diagnosis. Associated conditions include hearing loss, diabetes mellitus, hypothalamic pituitary dysfunction causing growth hormone deficiency, hypothyroidism, and absence of secondary sexual characteristics. In its full expression, MELAS leads to dementia, a bedridden state, and a fatal outcome. Serum lactic acid is typically elevated. The CSF protein is also increased but is usually ≤1 g/L (100 mg/dL). Muscle biopsies show ragged red fibers. Neuroimaging demonstrates basal ganglia calcification in a high percentage of cases. Focal lesions that mimic infarction are present predominantly in the occipital and parietal lobes. Strict vascular territories are not respected, and cerebral angiography fails to demonstrate lesions of the major cerebral blood vessels.

MELAS is caused by maternally inherited point mutations of mitochondrial tRNA genes. Most of the tRNA mutations are lethal, accounting for the paucity of multigeneration families with this syndrome. The A3243G point mutation in $tRNA^{Leu(UUR)}$ is the most common, occurring in ~80% of MELAS cases. About 10% of MELAS patients have other mutations of the $tRNA^{Leu(UUR)}$ gene, including 3252G, 3256T, 3271C, and 3291C. Other tRNA gene mutations have also been reported in MELAS, including G583A $tRNA^{Phe}$, G1642A $tRNA^{Val}$, G4332A $tRNA^{Glu}$, and T8316C $tRNA^{Lys}$. Mutations have also been reported in mtDNA polypeptide-coding genes. Two mutations were found in the ND5 subunit of complex I of the respiratory chain. A missense mutation has been reported at mtDNA position 9957 in the gene for subunit III of cytochrome C oxidase. No specific treatment is available. Supportive treatment is essential for the strokelike episodes, seizures, and endocrinopathies.

Pure Myopathy Syndromes

Muscle weakness and fatigue can be the predominant manifestations of mtDNA mutations. When the condition affects exclusively muscle (pure myopathy), the disorder becomes difficult to recognize. Occasionally, mitochondrial myopathies can present with recurrent myoglobinuria without fixed weakness and thus resemble a glycogen storage disorder or CPT deficiency.

MITOCHONDRIAL DNA DEPLETION SYNDROMES

Mitochondrial DNA depletion syndrome (MDS) is a heterogeneous group of disorders that are inherited in an autosomal recessive fashion and can present in infancy or adults. MDS can be caused by mutations in genes (*TK2*, *DGUOK*, *RRM2B*, *TYMP*, *SUCLA1*, and *SUCLA2*) that lead to depletion of pools of mitochondrial deoxyribonucleotide (dNTP) pools necessary for mtDNA replication The other major cause of MDS is a set of mutations in genes essential for mtDNA replication (e.g., *POLG1* and *C10orf2*). The clinical phenotypes associated with MDS vary. Patients may develop a severe encephalopathy (e.g., Leigh's syndrome), PEO, an isolated myopathy, myo-neuro-gastrointestinal-encephalopathy (MNGIE), and a sensory neuropathy with ataxia.

DISORDERS OF MUSCLE MEMBRANE EXCITABILITY

Muscle membrane excitability is affected in a group of disorders referred to as *channelopathies*. The heart may also be involved, resulting in life-threatening complications **(Table 387-10)**.

Table 387-10 Clinical Features of Periodic Paralysis and Nondystrophic Myotonias

	Calcium Channel	Sodium Channel		Potassium Channel
Feature	**Hypokalemic PP**	**Hyperkalemic PP**	**Paramyotonia Congenita**	**Andersen-Tawil Syndromea**
Mode of inheritance	AD	AD	AD	AD
Age of onset	Adolescence	Early childhood	Early childhood	Early childhood
Myotoniab	No	Yes	Yes	No
Episodic weakness	Yes	Yes	Yes	Yes
Frequency of attacks of weakness	Daily to yearly	May be 2–3/d	With cold, usually rare	Daily to yearly
Duration of attacks of weakness	2–12 h	From 1–2 h to >1 d	2–24 h	2–24 h
Serum K+ level during attacks of weakness	Decreased	Increased or normal	Usually normal	Variable
Effect of K+ loading	No change	Increased myotonia, then weakness	Increased myotonia	No change
Effect of muscle cooling	No change	Increased myotonia	Increased myotonia, then weakness	No change
Fixed weakness	Yes	Yes	Yes	Yes

[a]Dysmorphic features and cardiac arrhythmias are distinguishing features (see text). [b]May be paradoxical in paramyotonia congenita. ***Abbreviations:*** AD, autosomal dominant; PP, periodic paralysis.

Calcium Channel Disorders of Muscle

HYPOKALEMIC PERIODIC PARALYSIS (HYPOKPP)

Onset occurs at adolescence. Men are more often affected because of decreased penetrance in women. Episodic weakness with onset after age 25 is almost never due to periodic paralyses, with the exception of thyrotoxic periodic paralysis (see below). Attacks are often provoked by meals high in carbohydrates or sodium and may accompany rest following prolonged exercise. Weakness usually affects proximal limb muscles more than distal. Ocular and bulbar muscles are less likely to be affected. Respiratory muscles are usually spared but when they are involved, the condition may prove fatal. Weakness may take as long as 24 hours to resolve. Life-threatening cardiac arrhythmias related to hypokalemia may occur during attacks. As a late complication, patients commonly develop severe, disabling proximal lower extremity weakness.

Attacks of thyrotoxic periodic paralysis resemble those of primary HypoKPP. Despite a higher incidence of thyrotoxicosis in women, men, particularly those of Asian descent, are more likely to manifest this complication. Attacks abate with treatment of the underlying thyroid condition.

A low serum potassium level during an attack, excluding secondary causes, establishes the diagnosis. Interattack muscle biopsies show the presence of single or multiple centrally placed vacuoles or tubular aggregates. Provocative tests with glucose and insulin to establish a diagnosis are usually not necessary and are potentially hazardous.

In the midst of an attack of weakness, motor conduction studies may demonstrate reduced amplitudes, whereas EMG may show electrical silence in severely weak muscles. In between attacks, the EMG and nerve conduction studies are normal, with the exception that myopathic MUAPs may be seen in patients with fixed weakness.

HypoKPP is caused by mutations in either of two genes. HypoKPP type 1, the most common form, is inherited as an autosomal dominant disorder with incomplete penetrance. These patients have mutations in the voltage-sensitive, skeletal muscle calcium channel gene, *CALCL1A3***(Fig. 387-8)**. Approximately 10% of cases are HypoKPP type 2, arising from mutations in the voltage-sensitive sodium channel gene (*SCN4A*). In either instance, the mutations lead to an abnormal gating pore current that predisposes the muscle cell to depolarize when potassium levels are low. It is also now recognized that some cases of thyrotoxic HypoKPP are caused by genetic variants in a potassium channel (Kir 2.6), whose expression is regulated by thyroid hormone.

Figure 387-8

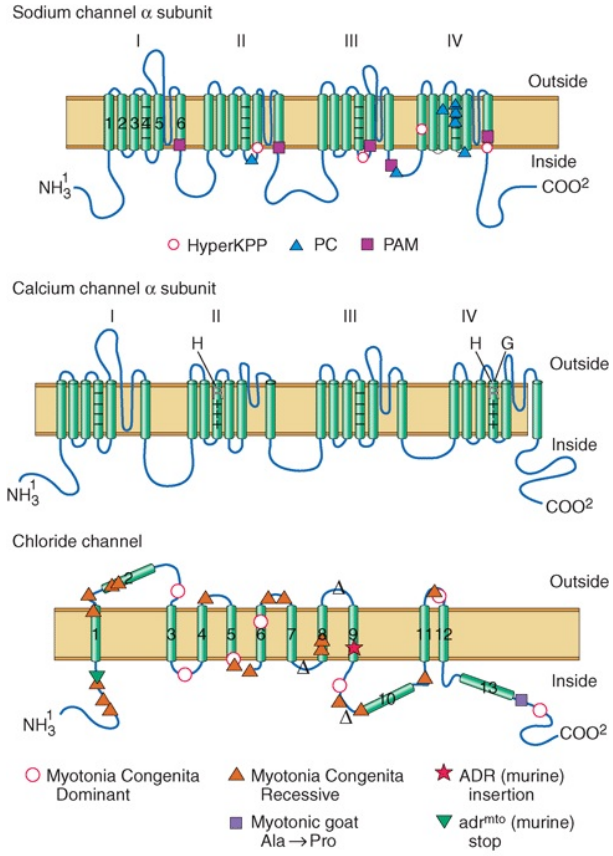

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

The sodium and calcium channels are depicted here as containing four homologous domains, each with six membrane-spanning segments. The fourth segment of each domain bears positive charges and acts as the "voltage sensor" for the channel. The association of the four domains is thought to form a pore through which ions pass. Sodium channel mutations are shown along with the phenotype that they confer. HyperKPP, hyperkalemic periodic paralysis; PC, paramyotonia congenita; PAM, potassium-aggravated myotonia. See text for details.

The chloride channel is envisioned to have 10 membrane-spanning domains. The positions of mutations causing dominantly and recessively inherited myotonia congenita are indicated, along with mutations that cause this disease in mice and goats.

Treatment: Hypokalemic Periodic Paralysis

The acute paralysis improves after the administration of potassium. Muscle strength and ECG should be monitored. Oral KCl (0.2–0.4 mmol/kg) should be given every 30 minutes. Only rarely is IV therapy necessary (e.g., when swallowing problems or vomiting is present). Administration of potassium in a glucose solution should be avoided because it may further reduce serum potassium levels. Mannitol is the preferred vehicle for administration of IV potassium. The long-term goal of therapy is to avoid attacks. This may reduce late-onset, fixed weakness. Patients should be made aware of the importance of a low-carbohydrate, low-sodium diet and consequences of intense exercise. Prophylactic administration of acetazolamide (125–1000 mg/d in divided doses) reduces or may abolish attacks in HypoKPP type 1. Paradoxically the potassium is lowered, but this is offset by the beneficial effect of metabolic acidosis. If attacks persist on acetazolamide, oral KCl should be added. Some patients require treatment with triamterine (25–100 mg/d) or spironolactone (25–100 mg/d). However, in patients with HypoKPP type 2, attacks of weakness can be exacerbated with acetazolamide.

Sodium Channel Disorders of Muscle

HYPERKALEMIC PERIODIC PARALYSIS (HYPERKPP)

The term *hyperkalemic* is misleading since patients are often normokalemic during attacks. The fact that attacks are precipitated by potassium administration best defines the disease. The onset is in the first decade; males and females are affected equally. Attacks are brief and mild, usually lasting 30 minutes to 4 hours. Weakness affects proximal muscles, sparing bulbar muscles. Attacks are precipitated by rest following exercise and fasting. In a variant of this disorder, the predominant symptom is myotonia without weakness (*potassium-aggravated myotonia*). The symptoms are aggravated by cold, and myotonia makes the muscles stiff and painful. This disorder can be confused with paramyotonia congenita, myotonia congenita, and proximal myotonic myopathy (DM2).

Potassium may be slightly elevated but may also be normal during an attack. As in HypoKPP, nerve conduction studies in HyperKPP muscle may demonstrate reduced motor amplitudes and the EMG may be silent in very weak muscles. In between attacks of weakness, the conduction studies are normal. The EMG will often demonstrate myotonic discharges during and between attacks.

The muscle biopsy shows vacuoles that are smaller, less numerous, and more peripheral compared to the hypokalemic form or tubular aggregates. Provocative tests by administration of potassium can induce weakness but are usually not necessary to establish the diagnosis. HyperKPP and potassium-aggravated myotonia are inherited as autosomal dominant disorders. Mutations of the voltage-gated sodium channel *SCN4A* gene (Fig. 387-8) cause these conditions. For patients with frequent attacks, acetazolamide (125–1000 mg/d) is helpful. We have found mexiletine to be helpful in patients with significant myotonia.

PARAMYOTONIA CONGENITA

In paramyotonia congenita (PC), the attacks of weakness are cold-induced or occur spontaneously and are mild. Myotonia is a prominent feature but worsens with muscle activity (paradoxical myotonia). This is in contrast to classic myotonia in which exercise alleviates the condition. Attacks of weakness are seldom severe enough to require emergency room treatment. Over

time patients develop interattack weakness as they do in other forms of periodic paralysis. PC is usually associated with normokalemia or hyperkalemia.

Serum CK is usually mildly elevated. Routine sensory and motor nerve conduction studies are normal. Cooling of the muscle often dramatically reduces the amplitude of the compound muscle action potentials. EMG reveals diffuse myotonic potentials in PC. Upon local cooling of the muscle, the myotonic discharges disappear as the patient becomes unable to activate MUAPs.

PC is inherited as an autosomal dominant condition; voltage-gated sodium channel mutations (Fig. 387-8) are responsible and thus this disorder is allelic with HyperKPP and potassium-aggravated myotonia. Patients with PC seldom seek treatment during attacks. Oral administration of glucose or other carbohydrates hastens recovery. Since interattack weakness may develop after repeated episodes, prophylactic treatment is usually indicated. Thiazide diuretics (e.g., chlorothiazide, 250–1000 mg/d) and mexiletine (slowly increase dose from 450 mg/d) are reported to be helpful. Patients should be advised to increase carbohydrates in their diet.

Potassium Channel Disorders

ANDERSEN-TAWIL SYNDROME

This rare disease is characterized by episodic weakness, cardiac arrhythmias, and dysmorphic features (short stature, scoliosis, clinodactyly, hypertelorism, small or prominent low-set ears, micrognathia, and broad forehead). The cardiac arrhythmias are potentially serious and life threatening. They include long QT, ventricular ectopy, bidirectional ventricular arrhythmias, and tachycardia. For many years the classification of this disorder was uncertain because episodes of weakness are associated with elevated, normal, or reduced levels of potassium during an attack. In addition, the potassium levels differ among kindreds but are consistent within a family. Inheritance is autosomal dominant, with incomplete penetrance and variable expressivity. The disease is caused by mutations of the inwardly rectifying potassium channel (*Kir 2.1*) gene that heighten muscle cell excitability. The treatment is similar to that for other forms of periodic paralysis and must include cardiac monitoring. The episodes of weakness may differ between patients because of potassium variability. Acetazolamide may decrease the attack frequency and severity.

Chloride Channel Disorders

Two forms of this disorder, autosomal dominant (*Thomsen's disease*) and autosomal recessive (*Becker's disease*), are related to the same gene abnormality. Symptoms are noted in infancy and early childhood. The severity lessens in the third to fourth decade. Myotonia is worsened by cold and improved by activity. The gait may appear slow and labored at first but improves with walking. In Thomsen's disease muscle strength is normal, but in Becker's disease, which is usually more severe, there may be muscle weakness. Muscle hypertrophy is usually present. Myotonic discharges are prominently displayed by EMG recordings.

Serum CK is normal or mildly elevated. The muscle biopsy shows hypertrophied fibers. The disease is inherited as dominant or recessive and is caused by mutations of the chloride channel gene (Fig. 387-8) that increase muscle cell excitability. Many patients will not require treatment and learn that the symptoms improve with activity. Medications that can be used to decrease myotonia include quinine, phenytoin, and mexiletine.

ENDOCRINE AND METABOLIC MYOPATHIES

Many endocrine disorders cause weakness. Muscle fatigue is more common than true weakness. The cause of weakness in these disorders is not well defined. It is not even clear that weakness results from disease of muscle as opposed to another part of the motor unit, since the serum CK level is often normal (except in hypothyroidism) and the muscle histology is characterized by atrophy rather than destruction of muscle fibers. Nearly all endocrine myopathies respond to treatment.

Thyroid Disorders

(See also Chap. 341) Abnormalities of thyroid function can cause a wide array of muscle disorders. These conditions relate to the important role of thyroid hormones in regulating the metabolism of carbohydrates and lipids as well as the rate of protein synthesis and enzyme production. Thyroid hormones also stimulate calorigenesis in muscle, increase muscle demand for vitamins, and enhance muscle sensitivity to circulating catecholamines.

HYPOTHYROIDISM

Patients with hypothyroidism have frequent muscle complaints, and proximal muscle weakness occurs in about one-third of them. Muscle cramps, pain, and stiffness are common. Some patients have enlarged muscles. Features of slow muscle contraction and relaxation occur in 25% of patients; the relaxation phase of muscle stretch reflexes is characteristically prolonged and best observed at the ankle or biceps brachii reflexes. The serum CK level is often elevated (up to 10 times normal), even when there is minimal clinical evidence of muscle disease. EMG is typically normal. The cause of muscle enlargement has not been determined, and muscle biopsy shows no distinctive morphologic abnormalities.

HYPERTHYROIDISM

Patients who are thyrotoxic commonly have proximal muscle weakness and atrophy on examination, but they rarely complain of myopathic symptoms. Activity of deep tendon reflexes may be enhanced. Bulbar, respiratory, and even esophageal muscles may occasionally be affected, causing dysphagia, dysphonia, and aspiration. When bulbar involvement occurs, it is usually accompanied by chronic proximal limb weakness, but occasionally it presents in the absence of generalized thyrotoxic myopathy. Fasciculations may be apparent and, when coupled with increased muscle stretch reflexes, may lead to an erroneous diagnosis of ALS. Other neuromuscular disorders occur in association with hyperthyroidism, including acquired hypokalemic periodic paralysis, myasthenia gravis, and a progressive ocular myopathy associated with proptosis (*Graves' ophthalmopathy*). Serum CK levels are not elevated in thyrotoxic myopathy, the EMG is normal, and muscle histology usually shows only atrophy of muscle fibers.

Parathyroid Disorders

(See also Chap. 353)

HYPERPARATHYROIDISM

Muscle weakness is an integral part of primary and secondary hyperparathyroidism. Proximal muscle weakness, muscle wasting, and brisk muscle stretch reflexes are the main features of this endocrinopathy. Some patients develop neck extensor weakness (part of the dropped head syndrome). Serum CK levels are usually normal or slightly elevated. Serum parathyroid hormone levels are elevated. Serum calcium and phosphorus levels show no correlation with the clinical neuromuscular manifestations. Muscle biopsies show only varying degrees of atrophy without muscle fiber degeneration.

HYPOPARATHYROIDISM

An overt myopathy due to hypocalcemia rarely occurs. Neuromuscular symptoms are usually related to localized or generalized tetany. Serum CK levels may be increased secondary to muscle damage from sustained tetany. Hyporeflexia or areflexia is usually present and contrasts with the hyperreflexia in hyperparathyroidism.

Adrenal Disorders

(See also Chap. 342) Conditions associated with glucocorticoid excess cause a myopathy; in fact, steroid myopathy is the most commonly diagnosed endocrine muscle disease. Glucocorticoid excess, either endogenous or exogenous (see "Drug-Induced Myopathies," below), produces various degrees of proximal limb weakness. Muscle wasting may be striking. A cushingoid appearance usually accompanies clinical signs of myopathy. Histologic sections demonstrate muscle fiber atrophy, preferentially affecting type 2b fibers, rather than degeneration or necrosis of muscle fibers. Adrenal insufficiency commonly causes muscle fatigue. The degree of weakness may be difficult to assess but is typically mild. In primary hyperaldosteronism (*Conn's syndrome*), neuromuscular complications are due to potassium depletion. The clinical picture is one of persistent muscle weakness. Long-standing hyperaldosteronism may lead to proximal limb weakness and wasting. Serum CK levels may be elevated, and a muscle biopsy may demonstrate degenerating fibers, some with vacuoles. These changes relate to hypokalemia and are not a direct effect of aldosterone on skeletal muscle.

Pituitary Disorders

(See also Chap. 339) Patients with acromegaly usually have mild proximal weakness without muscle atrophy. Muscles often appear enlarged but exhibit decreased force generation. The duration of acromegaly, rather than the serum growth hormone levels, correlates with the degree of myopathy.

Diabetes Mellitus

(See also Chap. 344) Neuromuscular complications of diabetes mellitus are most often related to neuropathy, with cranial and peripheral nerve palsies or distal sensorimotor polyneuropathy. *Diabetic amyotrophy* is a clumsy term since the condition represents a neuropathy affecting the proximal major nerve trunks and lumbosacral plexus. More appropriate terms for this disorder include *diabetic proximal neuropathy* and *lumbosacral radiculoplexus neuropathy*.

The only notable myopathy of diabetes mellitus is ischemic infarction of leg muscles, usually involving one of the thigh muscles but on occasion affecting the distal leg. This condition occurs in patients with poorly controlled diabetes and presents with abrupt onset of pain, tenderness, and edema of one thigh. The area of muscle infarction is hard and indurated. The muscles most often affected include the vastus lateralis, thigh adductors, and biceps femoris. CT or MRI can demonstrate focal abnormalities in the affected muscle. Diagnosis by imaging is preferable to muscle biopsy, if possible, as hemorrhage into the biopsy site can occur.

Vitamin Deficiency

Vitamin D deficiency (Chap. 74) due to either decreased intake, decreased absorption, or impaired vitamin D metabolism (as occurs in renal disease) may lead to chronic muscle weakness. Pain reflects the underlying bone disease (*osteomalacia*). Vitamin E deficiency may result from malabsorption. Clinical manifestations include ataxic neuropathy due to loss of proprioception and myopathy with proximal weakness. Progressive external ophthalmoplegia is a distinctive finding. It has not been established that deficiency of other vitamins causes a myopathy.

MYOPATHIES OF SYSTEMIC ILLNESS

Systemic illnesses such as chronic respiratory, cardiac, or hepatic failure are frequently associated with severe muscle wasting and complaints of weakness. Fatigue is usually a more significant problem than weakness, which is typically mild.

Myopathy may be a manifestation of chronic renal failure (CRF), independent of the better known uremic polyneuropathy. Abnormalities of calcium and phosphorus homeostasis and bone metabolism in chronic renal failure result from a reduction in 1,25-dihydroxyvitamin D, leading to decreased intestinal absorption of calcium. Hypocalcemia, further accentuated by hyperphosphatemia due to decreased renal phosphate clearance, leads to secondary hyperparathyroidism. Renal osteodystrophy results from the compensatory hyperparathyroidism, which leads to osteomalacia from reduced calcium availability and to osteitis fibrosa from the parathyroid hormone excess. The clinical picture of the myopathy of CRF is identical to that of primary hyperparathyroidism and osteomalacia. There is proximal limb weakness with bone pain.

Gangrenous calcification represents a separate, rare, and sometimes fatal complication of CRF. In this condition, widespread arterial calcification occurs and results in ischemia. Extensive skin necrosis may occur, along with painful myopathy and even myoglobinuria.

DRUG-INDUCED MYOPATHIES

Drug-induced myopathies are relatively uncommon in clinical practice with the exception of those caused by the cholesterol-lowering agents and glucocorticoids. Others impact practice to a lesser degree but are important to consider in specific situations. **Table 387-11** provides a comprehensive list of drug-induced myopathies with their distinguishing features.

Table 387-11 Drug-Induced Myopathies

Drugs	Major Toxic Reaction
Lipid-lowering agents Fibric acid derivatives HMG-CoA reductase inhibitors Niacin (nicotinic acid)	Drugs belonging to all three of the major classes of lipid-lowering agents can produce a spectrum of toxicity: asymptomatic serum creatine kinase elevation, myalgias, exercise-induced pain, rhabdomyolysis, and myoglobinuria.
Glucocorticoids	Acute, high-dose glucocorticoid treatment can cause acute quadriplegic myopathy. These high doses of steroids are often combined with nondepolarizing neuromuscular blocking agents but the weakness can occur without their use. Chronic steroid administration produces predominantly proximal weakness.
Nondepolarizing neuromuscular blocking agents	Acute quadriplegic myopathy can occur with or without concomitant glucocorticoids.
Zidovudine	Mitochondrial myopathy with ragged red fibers.
Drugs of abuse Alcohol Amphetamines Cocaine Heroin	All drugs in this group can lead to widespread muscle breakdown, rhabdomyolysis, and myoglobinuria. Local injections cause muscle necrosis, skin induration, and limb contractures.

Phencyclidine Meperidine	
Autoimmune toxic myopathy d-Penicillamine	Use of this drug may cause polymyositis and myasthenia gravis.
Amphophilic cationic drugs Amiodarone Chloroquine Hydroxychloroquine	All amphophilic drugs have the potential to produce painless, proximal weakness associated with autophagic vacuoles in the muscle biopsy.
Antimicrotubular drugs Colchicine	This drug produces painless, proximal weakness especially in the setting of renal failure. Muscle biopsy shows autophagic vacuoles.

Myopathy from Lipid-Lowering Agents

All classes of lipid-lowering agents have been implicated in muscle toxicity, including fibrates (clofibrate, gemfibrozil), HMG-CoA reductase inhibitors (referred to as *statins*), niacin (nicotinic acid), and ezetimibe. Myalgia, malaise, and muscle tenderness are the most common manifestations. Muscle pain may be related to exercise. Patients may exhibit proximal weakness. Varying degrees of muscle necrosis are seen, and in severe reactions rhabdomyolysis and myoglobinuria occur. Concomitant use of statins with fibrates and cyclosporine is more likely to cause adverse reactions than use of one agent alone. Elevated serum CK is an important indication of toxicity. Muscle weakness is accompanied by a myopathic EMG, and muscle necrosis is observed by muscle biopsy. Severe myalgias, muscle weakness, significant elevations in serum CK (>three times baseline), and myoglobinuria are indications for stopping the drug. Patients usually improve with drug cessation, although this may take several weeks. Rare cases continue to progress after the offending agent is discontinued. It is possible that in such cases the statin may have triggered an immune-mediated necrotizing myopathy, as these individuals require immunotherapy (e.g., prednisone and sometimes other agents) to improve and often relapse when these therapies are discontinued. Interestingly, antibodies directed against the 100-kD HMG-CoA reductase receptor on muscle fibers has been identified in many of these cases.

Glucocorticoid-Related Myopathies

Glucocorticoid myopathy occurs with chronic treatment or as "acute quadriplegic" myopathy secondary to high-dose IV glucocorticoid use. Chronic administration produces proximal weakness accompanied by cushingoid manifestations, which can be quite debilitating; the chronic use of prednisone at a daily dose of ≳30 mg/d is most often associated with toxicity. Patients taking fluorinated glucocorticoids (triamcinolone, betamethasone, dexamethasone) appear to be at especially high risk for myopathy. In chronic steroid myopathy, the serum CK is usually normal. Serum potassium may be low. The muscle biopsy in chronic cases shows preferential type 2 muscle fiber atrophy; this is not reflected in the EMG, which is usually normal.

Patients receiving high-dose IV glucocorticoids for status asthmaticus, chronic obstructive pulmonary disease, organ transplantation, or other indications may develop severe generalized weakness (critical illness myopathy). This myopathy, also known as acute quadriplegic myopathy, can also occur in the setting of sepsis. Involvement of the diaphragm and intercostal muscles causes respiratory failure and requires ventilatory support. In these settings, the use of glucocorticoids in combination with nondepolarizing neuromuscular blocking agents potentiate this complication. In critical illness myopathy, the muscle biopsy is abnormal, showing a distinctive loss of thick filaments (myosin) by electron microscopy. By light microscopy, there is focal loss of ATPase staining in central or paracentral areas of the muscle fiber. Calpain stains show diffusely reactive atrophic fibers. Withdrawal of glucocorticoids will improve the chronic myopathy. In acute quadriplegic myopathy, recovery is slow. Patients require supportive care and rehabilitation.

Drug-Induced Mitochondrial Myopathy

Zidovudine, used in the treatment of HIV infection, is a thymidine analogue that inhibits viral replication by interrupting reverse transcriptase. Myopathy is a well-established complication of this agent. Patients present with myalgias, muscle weakness, and atrophy affecting the thigh and calf muscles. The complication occurs in about 17% of patients treated with doses of 1200 mg/d for 6 months. The introduction of protease inhibitors for treatment of HIV infection has led to lower doses of zidovudine therapy and a decreased incidence of myopathy. Serum CK is elevated and EMG is myopathic. Muscle biopsy shows ragged red fibers with minimal inflammation; the lack of inflammation serves to distinguish zidovudine toxicity from HIV-related myopathy. If the myopathy is thought to be drug related, the medication should be stopped or the dosage reduced.

Drugs of Abuse and Related Myopathies

Myotoxicity is a potential consequence of addiction to alcohol and illicit drugs. Ethanol is one of the most commonly abused substances with potential to damage muscle. Other potential toxins include cocaine, heroin, and amphetamines. The most deleterious reactions occur from overdosing leading to coma and seizures, causing rhabdomyolysis, myoglobinuria, and renal failure. Direct toxicity can occur from cocaine, heroin, and amphetamines causing muscle breakdown and varying degrees of weakness. The effects of alcohol are more controversial. Direct muscle damage is less certain, since toxicity usually occurs in the setting of poor nutrition and possible contributing factors such as hypokalemia and hypophosphatemia. Alcoholics are also prone to neuropathy (Chap. 392).

Focal myopathies from self-administration of meperidine, heroin, and pentazocine can cause pain, swelling, muscle necrosis, and hemorrhage. The cause is multifactorial; needle trauma, direct toxicity of the drug or vehicle, and infection may all play a role. When severe, there may be overlying skin induration and contractures with replacement of muscle by connective tissue. Elevated serum CK and myopathic EMG are characteristic of these reactions. The muscle biopsy shows widespread or focal areas of necrosis. In conditions leading to rhabdomyolysis, patients need adequate hydration to reduce serum myoglobin and protect renal function. In all of these conditions, counseling is essential to limit drug abuse.

Drug-Induced Autoimmune Myopathies

The most consistent drug-related inflammatory or antibody-mediated myopathy is caused by D-penicillamine. This drug chelates copper and is used in the treatment of Wilson's disease. It is also used to treat other disorders including scleroderma, rheumatoid arthritis, and primary biliary cirrhosis. Adverse events include drug-induced polymyositis, indistinguishable from the spontaneous disease. The incidence of this inflammatory muscle disease is about 1%. Myasthenia gravis is also induced by d-penicillamine, with a higher incidence estimated at 7%. These disorders resolve with drug withdrawal, although immunosuppressive therapy may be warranted in severe cases.

Scattered reports of other drugs causing an inflammatory myopathy are rare and include a heterogeneous group of agents: cimetidine, phenytoin, procainamide, and propylthiouracil. In most cases, a cause-and-effect relationship is uncertain. A

complication of interest was related to l-tryptophan. In 1989 an epidemic of eosinophilia-myalgia syndrome (EMS) in the United States was caused by a contaminant in the product from one manufacturer. The product was withdrawn, and incidence of EMS diminished abruptly following this action.

Other Drug-Induced Myopathies

Certain drugs produce painless, largely proximal, muscle weakness. These drugs include the amphophilic cationic drugs (amiodarone, chloroquine, hydroxychloroquine) and antimicrotubular drugs (colchicine) (Table 387-11). Muscle biopsy can be useful in the identification of toxicity since autophagic vacuoles are prominent pathologic features of these toxins.

FURTHER READINGS

Abu-Baker A, Rouleau GA: Oculopharyngeal muscular dystrophy: Recent advances in the understanding of the molecular pathogenic mechanisms and treatment strategies. Biochim Biophys Acta 772:173, 2007

Berardo A et al: A diagnostic algorithm for metabolic myopathies. Curr Neurol Neurosci Rep 10:118, 2010[PMID: 20425236] [Full Text]

Bushby K: Diagnosis and management of the limb girdle muscular dystrophies. Pract Neurol 9:314, 2009[PMID: 19923111] [Full Text]

Cannon SC: Voltage sensor mutations in channelopathies of skeletal muscle. J Physiol 588:1887, 2010[PMID: 20156847] [Full Text]

Christopher-Stine L et al: A novel autoantibody recognizing 200-kD and 100-kD proteins is associated with an immune-mediated necrotizing myopathy. Arthritis Rheum 62:2757, 2010[PMID: 20496415] [Full Text]

Glover L, Brown RH Jr: Dysferlin in membrane trafficking and patch repair. Traffic 8:785, 2007[PMID: 17547707] [Full Text]

Grable-esposito P et al: Immune-mediated necrotizing myopathy associated with statins. Muscle Nerve 41:185, 2010[PMID: 19813188] [Full Text]

Kishnani PS et al: Recombinant human acid α-glucosidase: Major clinical benefits in infantile-onset Pompe disease. Neurology 68:99, 2007[PMID: 17151339] [Full Text]

Lemmers RJ et al: A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329:1650, 2010[PMID: 20724583] [Full Text]

Matthews E et al: The non-dystrophic myotonias: Molecular pathogenesis, diagnosis and treatment. Brain 133:9, 2010[PMID: 19917643] [Full Text]

Osborne RJ et al: Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. Hum Mol Genet 18:1471, 2009[PMID: 19223393] [Full Text]

Rodino-klapac LR et al: Gene therapy for Duchenne muscular dystrophy: Expectations and challenges. Arch Neurol 64:1236, 2007[PMID: 17846262] [Full Text]

Suomalainen A, Isohanni P: Mitochondrial depletion syndromes—Many genes, common mechanisms. Neuromuscular Dis 20:429, 2010[PMID: 20444604] [Full Text]

van der Ploeg AT et al: A randomized study of alglucosidase alfa in late-onset Pompe's disease. N Engl J Med 362:1396, 2010

Venance SL et al: The primary periodic paralyses: Diagnosis, pathogenesis and treatment. Brain 129:8, 2006[PMID: 16195244] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

POLYMYOSITIS, DERMATOMYOSITIS, AND INCLUSION BODY MYOSITIS: INTRODUCTION

The inflammatory myopathies represent the largest group of acquired and potentially treatable causes of skeletal muscle weakness. They are classified into three major groups: polymyositis (PM), dermatomyositis (DM), and inclusion body myositis (IBM).

Clinical Features

The prevalence of the inflammatory myopathies is estimated at 1 in 100,000. PM as a stand-alone entity is a rare disease. DM affects both children and adults and women more often than men. IBM is three times more frequent in men than in women, more common in whites than blacks, and is most likely to affect persons aged >50 years.

These disorders present as progressive and symmetric muscle weakness except for IBM, which can have an asymmetric pattern. Patients usually report increasing difficulty with everyday tasks requiring the use of proximal muscles, such as getting up from a chair, climbing steps, stepping onto a curb, lifting objects, or combing hair. Fine-motor movements that depend on the strength of distal muscles, such as buttoning a shirt, sewing, knitting, or writing, are affected only late in the course of PM and DM, but fairly early in IBM. Falling is common in IBM because of early involvement of the quadriceps muscle, with buckling of the knees. Ocular muscles are spared, even in advanced, untreated cases; if these muscles are affected, the diagnosis of inflammatory myopathy should be questioned. Facial muscles are unaffected in PM and DM, but mild facial muscle weakness is common in patientswith IBM. In all forms of inflammatory myopathy, pharyngeal and neck-flexor muscles are often involved, causing dysphagia or difficulty in holding up the head (*head drop*). In advanced and rarely in acute cases, respiratory muscles may also be affected. Severe weakness, if untreated, is almost always associated with muscle wasting. Sensation remains normal. The tendon reflexes are preserved but may be absent in severely weakened or atrophied muscles, especially in IBM, where atrophy of the quadriceps and the distal muscles is common. Myalgia and muscle tenderness may occur in a small number of patients, usually early in the disease, and particularly in DM associated with connective tissue disorders. Weakness in PM and DM progresses subacutely over a period of weeks or months and rarely acutely; by contrast, IBM progresses very slowly, over years, simulating a late-life muscular dystrophy (Chap. 387) or slowly progressive motor neuron disorder (Chap. 374).

Specific Features

(Table 388-1)

Table 388-1 Features Associated with Inflammatory Myopathies

Characteristic	Polymyositis	Dermatomyositis	Inclusion Body Myositis
Age at onset	>18 years	Adulthood and childhood	>50 years
Familial association	No	No	Yes, in some cases
Extramuscular manifestations	Yes	Yes	Yes
Associated conditions			
Connective tissuediseases	Yes[a]	Scleroderma and mixed connective tissue disease (overlap syndromes)	Yes, in up to 20% of cases[a]
Systemic autoimmune diseases[b]	Frequent	Infrequent	Infrequent
Malignancy	No	Yes, in up to 15% of cases	No
Viruses	Yes[c]	Unproven	Yes[c]
Drugs[d]	Yes	Yes, rarely	No

Parasites and bacteria[e]	Yes	No	No

[a]Systemic lupus erythematosus, rheumatoid arthritis, Sjögren's syndrome, systemic sclerosis, mixed connective tissue disease. [b]Crohn's disease, vasculitis, sarcoidosis, primary biliary cirrhosis, adult celiac disease, chronic graft-versus-host disease, discoid lupus, ankylosing spondylitis, Behçet's syndrome, myasthenia gravis, acne fulminans, dermatitis herpetiformis, psoriasis, Hashimoto's disease, granulomatous diseases, agammaglobulinemia, monoclonal gammopathy, hypereosinophilic syndrome, Lyme disease, Kawasaki disease, autoimmune thrombocytopenia, hypergammaglobulinemic purpura, hereditary complement deficiency, IgA deficiency. [c]HIV (human immunodeficiency virus) and HTLV-I (human T cell lymphotropic virus type I). [d]Drugs include penicillamine (dermatomyositis and polymyositis), zidovudine (polymyositis), and contaminated tryptophan (dermatomyositis-like illness). Other myotoxic drugs may cause myopathy but not an inflammatory myopathy (see text for details). [e]Parasites (protozoa, cestodes, nematodes), tropical and bacterial myositis (pyomyositis).

POLYMYOSITIS

The actual onset of PM is often not easily determined, and patients typically delay seeking medical advice for several weeks or even months. This is in contrast to DM, in which the rash facilitates early recognition (see below). PM mimics many other myopathies and is a diagnosis of exclusion. It is a subacute inflammatory myopathy affecting adults, and rarely children, who *do not have* any of the following: rash, involvement of the extraocular and facial muscles, family history of a neuromuscular disease, history of exposure to myotoxic drugs or toxins, endocrinopathy, neurogenic disease, muscular dystrophy, biochemical muscle disorder (deficiency of a muscle enzyme), or IBM as excluded by muscle biopsy analysis (see below). As an isolated entity, PM is a rare (and overdiagnosed) disorder; more commonly, PM occurs in association with a systemic autoimmune or connective tissue disease, or with a known viral or bacterial infection. Drugs, especially d-penicillamine, statins, or zidovudine (AZT), may also trigger an inflammatory myopathy similar to PM.

DERMATOMYOSITIS

DM is a distinctive entity identified by a characteristic rash accompanying, or more often preceding, muscle weakness. The rash may consist of a blue-purple discoloration on the upper eyelids with edema (heliotrope rash; see Fig. 54-3), a flat red rash on the face and upper trunk, and erythema of the knuckles with a raised violaceous scaly eruption (*Gottron's sign*; see Fig. 54-4). The erythematous rash can also occur on other body surfaces, including the knees, elbows, malleoli, neck and anterior chest (often in a *V sign*), or back and shoulders (*shawl sign*), and may worsen after sun exposure. In some patients, the rash is pruritic, especially on the scalp, chest, and back. Dilated capillary loops at the base of the fingernails are also characteristic. The cuticles may be irregular, thickened, and distorted, and the lateral and palmar areas of the fingers may become rough and cracked, with irregular, "dirty" horizontal lines, resembling *mechanic's hands*. The weakness can be mild, moderate, or severe enough to lead to quadriparesis. At times, the muscle strength appears normal, hence the term *dermatomyositis sine myositis*. When muscle biopsy is performed in such cases, however, significant perivascular and perimysial inflammation is often seen.

DM usually occurs alone but may overlap with scleroderma and mixed connective tissue disease. Fasciitis and thickening of the skin, similar to that seen in chronic cases of DM, have occurred in patients with the *eosinophilia-myalgia syndrome* associated with the ingestion of contaminated l-tryptophan.

INCLUSION BODY MYOSITIS

In patients ≥50 years of age, IBM is the most common of the inflammatory myopathies. It is often misdiagnosed as PM and is suspected only later when a patient with presumed PM does not respond to therapy. Weakness and atrophy of the distal muscles, especially foot extensors and deep finger flexors, occur in almost all cases of IBM and may be a clue to early diagnosis. Some patients present with falls because their knees collapse due to early quadriceps weakness. Others present with weakness in the small muscles of the hands, especially finger flexors, and complain of inability to hold objects such as golf clubs or perform tasks such as turning keys or tying knots. On occasion, the weakness and accompanying atrophy can be asymmetric and selectively involve the quadriceps, iliopsoas, triceps, biceps, and finger flexors, resembling a lower motor neuron disease. Dysphagia is common, occurring in up to 60% of IBM patients, and may lead to episodes of choking. Sensory examination is generally normal; some patients have mildly diminished vibratory sensation at the ankles that presumably is age-related. The pattern of distal weakness, which superficially resembles motor neuron or peripheral nerve disease, results from the myopathic process affecting distal muscles selectively. Disease progression is slow but steady, and most patients require an assistive device such as cane, walker, or wheelchair within several years of onset.

In at least 20% of cases, IBM is associated with systemic autoimmune or connective tissue diseases. Familial aggregation of typical IBM may occur; such cases have been designated as *familial*

inflammatory IBM. This disorder is distinct from *hereditary inclusion body myopathy* (h-IBM), which describes a heterogeneous group of recessive, and less frequently dominant, inherited syndromes; the h-IBMs are noninflammatory myopathies. A subset of h-IBM that spares the quadriceps muscles has emerged as a distinct entity. This disorder, originally described in Iranian Jews and now seen in many ethnic groups, is linked to chromosome 9p1 and results from mutations in the UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase (*GNE*) gene.

Associated Clinical Findings

EXTRAMUSCULAR MANIFESTATIONS

These may be present to a varying degree in patients with PM or DM, and include:

1. *Systemic symptoms*, such as fever, malaise, weight loss, arthralgia, and Raynaud's phenomenon, especially when inflammatory myopathy is associated with a connective tissue disorder.
2. *Joint contractures*, mostly in DM and especially in children.
3. *Dysphagia and gastrointestinal symptoms*, due to involvement of oropharyngeal striated muscles and upper esophagus, especially in DM and IBM.
4. *Cardiac disturbances*, including atrioventricular conduction defects, tachyarrhythmias, dilated cardiomyopathy, a low ejection fraction, and congestive heart failure, may rarely occur, either from the disease itself or from hypertension associated with long-term use of glucocorticoids.
5. *Pulmonary dysfunction*, due to weakness of the thoracic muscles, interstitial lung disease, or drug-induced pneumonitis (e.g., from methotrexate), which may cause dyspnea, nonproductive cough, and aspiration pneumonia. Interstitial lung disease may precede myopathy or occur early in the disease and develops in up to 10% of patients with PM or DM, most of whom have antibodies to t-RNA synthetases, as described below.
6. *Subcutaneous calcifications*, in DM, sometimes extruding on the skin and causing ulcerations and infections.
7. *Arthralgias*, synovitis, or deforming arthropathy with subluxation in the interphalangeal joints can occur in some patients with DM and PM who have Jo-1 antibodies (see below).

ASSOCIATION WITH MALIGNANCIES

Although all the inflammatory myopathies can have a chance association with malignant lesions, especially in older age groups, the incidence of malignant conditions appears to be specifically increased only in patients with DM and not in those with PM or IBM. The most common tumors associated with DM are ovarian cancer, breast cancer, melanoma, colon cancer, and non-Hodgkin lymphoma. The extent of the search that should be conducted for an occult neoplasm in adults with DM depends on the clinical circumstances. Tumors in these patients are usually uncovered by abnormal findings in the medical history and physical examination and not through an extensive blind search. The weight of evidence argues against performing expensive, invasive, and nondirected tumor searches. A complete annual physical examination with pelvic, breast (mammogram, if indicated), and rectal examinations (with colonoscopy according to age and family history); urinalysis; complete blood count; blood chemistry tests; and a chest film should suffice in most cases. In Asians, nasopharyngeal cancer is common, and a careful examination of ears, nose, and throat is indicated. If malignancy is clinically suspected, screening with whole-body PET scan should be considered.

OVERLAP SYNDROMES

These describe the association of inflammatory myopathies with connective tissue diseases. A well-characterized overlap syndrome occurs in patients with DM who also have manifestations of systemic sclerosis or mixed connective tissue disease, such as sclerotic thickening of the dermis, contractures, esophageal hypomotility, microangiopathy, and calcium deposits (Table 388-1). By contrast, signs of rheumatoid arthritis, systemic lupus erythematosus, or Sjögren's syndrome are very rare in patients with DM. Patients with the overlap syndrome of DM and systemic sclerosis may have a specific antinuclear antibody, the anti-PM/Scl, directed against a nucleolar-protein complex.

Pathogenesis

An autoimmune etiology of the inflammatory myopathies is indirectly supported by an association with other autoimmune or connective tissue diseases; the presence of various autoantibodies; an association with specific major histocompatibility complex (MHC) genes; demonstration of T cell–mediated myocytotoxicity or complement-mediated microangiopathy; and a response to immunotherapy.

AUTOANTIBODIES AND IMMUNOGENETICS

Various autoantibodies against nuclear antigens (antinuclear antibodies) and cytoplasmic antigens are found in up to 20% of patients with inflammatory myopathies. The antibodies to cytoplasmic antigens are directed against ribonucleoproteins involved in protein synthesis (antisynthetases) or translational transport (anti-signal-recognition particles). The antibody directed against the histidyl-transfer RNA

synthetase, called *anti-Jo-1*, accounts for 75% of all the antisynthetases and is clinically useful because up to 80% of patients with anti-Jo-1 antibodies have interstitial lung disease. Some patients with the anti-Jo-1 antibody also have Raynaud's phenomenon, nonerosive arthritis, and the MHC molecules DR3 and DRw52. DR3 haplotypes (molecular designation DRB1*0301, DQB1*0201) occur in up to 75% of patients with PM and IBM, whereas in juvenile DM there is an increased frequency of DQA1*0501 (Chap. 315).

IMMUNOPATHOLOGIC MECHANISMS

In DM, humoral immune mechanisms are implicated, resulting in a microangiopathy and muscle ischemia **(Fig. 388-1)**. Endomysial inflammatory infiltrates are composed of B cells located in proximity to CD4 T cells, plasmacytoid dendritic cells, and macrophages; there is a relative absence of lymphocytic invasion of nonnecrotic muscle fibers. Activation of the complement C5b-9 membranolytic attack complex is thought to be a critical early event that triggers release of proinflammatory cytokines and chemokines, induces expression of vascular cell adhesion molecule (VCAM) 1 and intercellular adhesion molecule (ICAM) 1 on endothelial cells, and facilitates migration of activated lymphoid cells to the perimysial and endomysial spaces. Necrosis of the endothelial cells, reduced numbers of endomysial capillaries, ischemia, and muscle-fiber destruction resembling microinfarcts occur. The remaining capillaries often have dilated lumens in response to the ischemic process. Larger intramuscular blood vessels may also be affected in the same pattern. Residual perifascicular atrophy reflects the endofascicular hypoperfusion that is prominent in the periphery of muscle fascicles. Increased expression of type I interferon-inducible proteins is also noted in these regions.

Figure 388-1

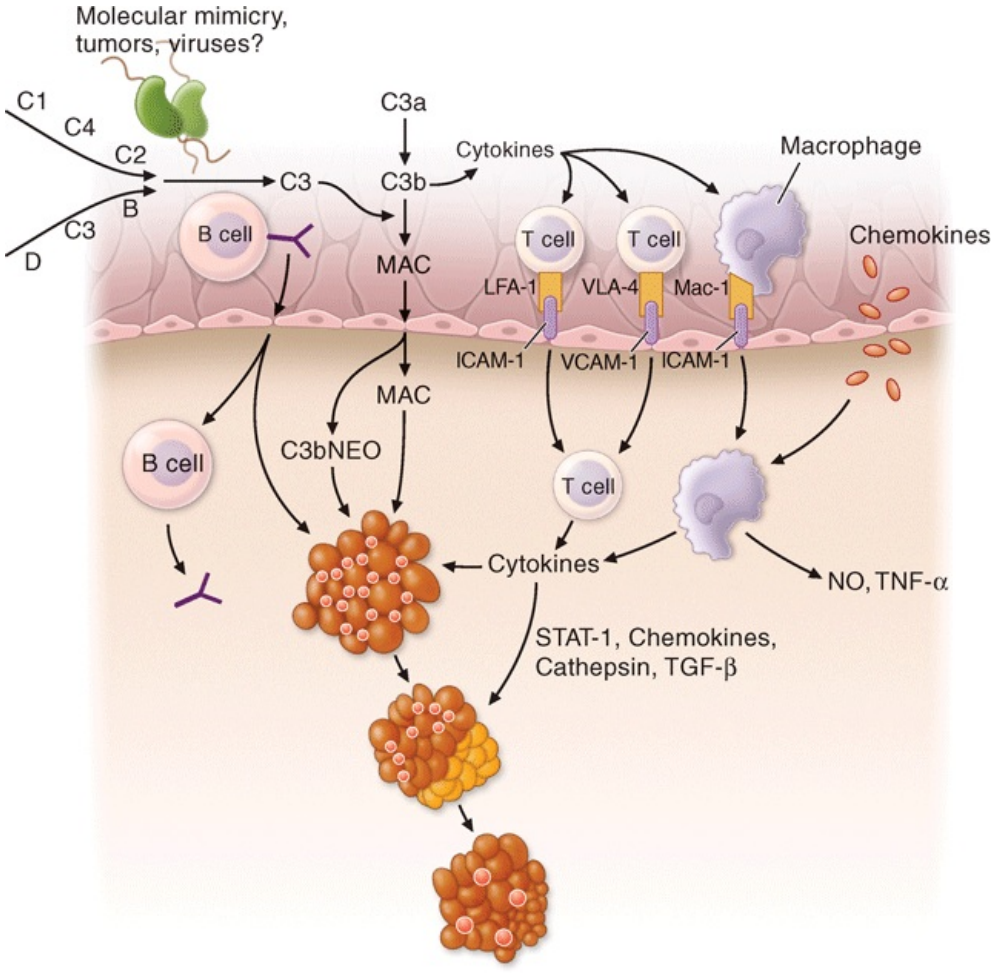

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Immunopathogenesis of dermatomyositis. Activation of complement, possibly by autoantibodies (Y), against endothelial cells and formation of C3 via the classic or alternative pathway. Activated C3 leads to formation of C3b, C3bNEO, and membrane attack complexes (MAC), which are deposited in and around the endothelial cell wall of the endomysial capillaries. Deposition of MAC leads to destruction of capillaries, ischemia, or microinfarcts, most prominent in the periphery of the fascicles, and perifascicular atrophy. B cells, plasmacytoid dendritic cells, CD4 T cells, and macrophages traffic from the circulation to the muscle. Endothelial expression of vascular cell adhesion molecule (VCAM)

and intercellular adhesion molecule (ICAM) is induced by cytokines released by the mononuclear cells. Integrins, specifically very late activation antigen (VLA)-4 and lymphocyte function–associated antigen (LFA)-1, bind VCAM and ICAM and promote T cell and macrophage infiltration of muscle through the endothelial cell wall.

By contrast, in PM and IBM a mechanism of T cell–mediated cytotoxicity is likely. CD8 T cells, along with macrophages, initially surround and eventually invade and destroy healthy, nonnecrotic muscle fibers that aberrantly express class I MHC molecules. MHC-I expression, absent from the sarcolemma of normal muscle fibers, is probably induced by cytokines secreted by activated T cells and macrophages. The CD8/MHC-I complex is characteristic of PM and IBM; its detection can aid in confirming the histologic diagnosis of PM, as discussed below. The cytotoxic CD8 T cells contain perforin and granzyme granules directed toward the surface of the muscle fibers and capable of inducing myonecrosis. Analysis of T cell receptor molecules expressed by the infiltrating CD8 cells has revealed clonal expansion and conserved sequences in the antigen-binding region, both suggesting an antigen-driven T cell response. Whether the putative antigens are endogenous (e.g., muscle) or exogenous (e.g., viral) sequences is unknown. Viruses have not been identified within the muscle fibers. Co-stimulatory molecules and their counterreceptors, which are fundamental for T cell activation and antigen recognition, are strongly upregulated in PM and IBM. Key molecules involved in T cell –mediated cytotoxicity are depicted in **Fig. 388-2**.

Figure 388-2

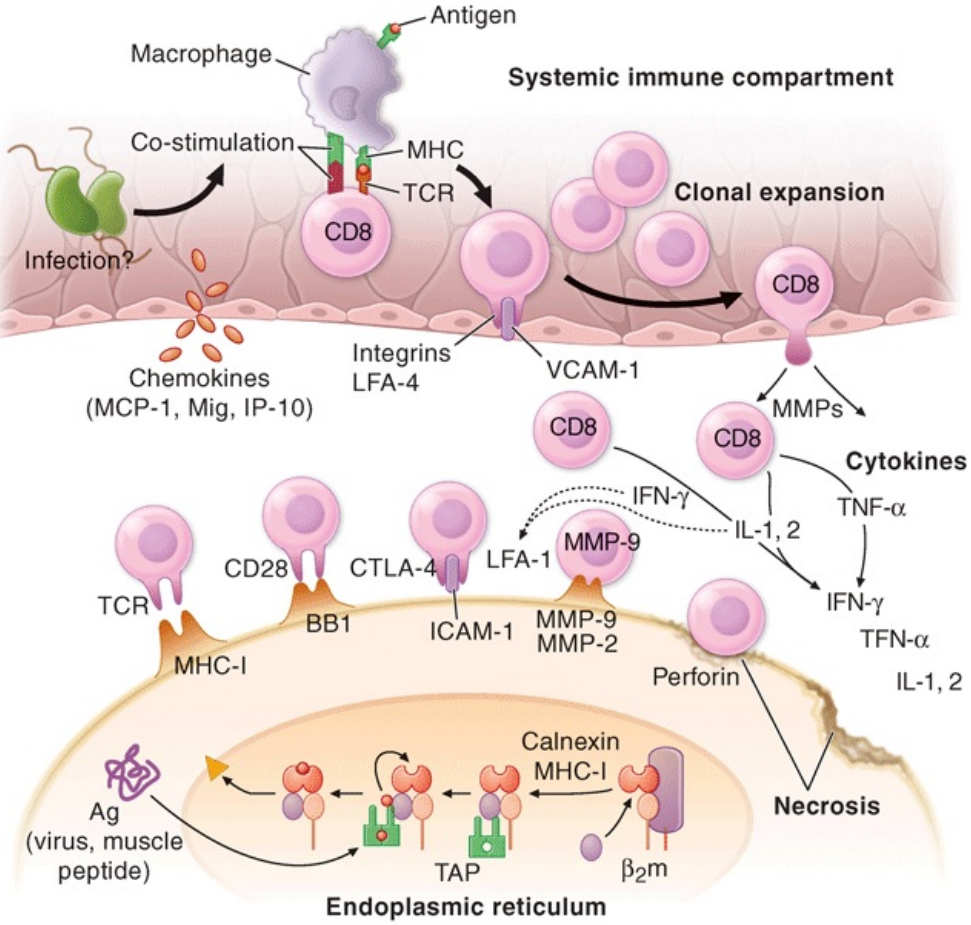

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cell-mediated mechanisms of muscle damage in polymyositis (PM) and inclusion body myositis (IBM). Antigen-specific CD8 cells are expanded in the periphery, cross the endothelial barrier, and bind directly to muscle fibers via T cell receptor (TCR) molecules that recognize aberrantly expressed MHC-I. Engagement of co-stimulatory molecules (BB1 and ICOSL) with their ligands (CD28, CTLA-4, and ICOS), along with ICAM-1/LFA-1, stabilize the CD8–muscle fiber interaction. Metalloproteinases (MMPs) facilitate the migration of T cells and their attachment to the muscle surface. Muscle fiber necrosis occurs via perforin granules released by the autoaggressive T cells. A direct myocytotoxic effect exerted by the cytokinesinterferon (IFN) γ, interleukin (IL) 1, or tumor necrosis factor (TNF) α may also play a role. Death of the muscle fiber is mediated by necrosis. MHC class I molecules consist of a heavy chain and a light chain [β_2 microglobulin (β_2m)] complexed with

an antigenic peptide that is transported into the endoplasmic reticulum by TAP proteins (Chap. 315).

THE ROLE OF NONIMMUNE FACTORS IN IBM

In IBM, the presence of Congo red–positive amyloid deposits within some vacuolated muscle fibers and abnormal mitochondria with cytochrome oxidase–negative fibers suggest that, in addition to the autoimmune component, there is also a degenerative process. Similar to Alzheimer's disease, the intracellular amyloid deposits in IBM are immunoreactive against amyloid precursor protein (APP), β-amyloid, chymotrypsin, apolipoprotein E, presenilin, ubiquitin, and phosphorylated tau, but it is unclear whether these deposits, which are also seen in other vacuolar myopathies, are directly pathogenic or represent secondary phenomena. The same is true for the mitochondrial abnormalities, which may also be secondary to the effects of aging or a bystander effect of upregulated cytokines. Expression of cytokines and upregulation of MHC class I by the muscle fibers may cause an endoplasmic reticulum stress response resulting in intracellular accumulation of stressor molecules or misfolded glycoproteins and activation of nuclear factor κB (NF-κB), leading to further cytokine activation.

ASSOCIATION WITH VIRAL INFECTIONS AND THE ROLE OF RETROVIRUSES

Several viruses, including coxsackieviruses, influenza, paramyxoviruses, mumps, cytomegalovirus, and Epstein-Barr virus, have been indirectly associated with myositis. For the coxsackieviruses, an autoimmune myositis triggered by molecular mimicry has been proposed because of structural homology between histidyl-transfer RNA synthetase that is the target of the Jo-1 antibody (see above) and genomic RNA of an animal picornavirus, the encephalomyocarditis virus. Sensitive polymerase chain reaction (PCR) studies, however, have repeatedly failed to confirm the presence of such viruses in muscle biopsies.

The best evidence of a viral connection in PM and IBM is with the retroviruses. Some individuals infected with HIV or with human T cell lymphotropic virus I (HTLV-I) develop PM or IBM; a similar disorder has been described in nonhuman primates infected with the simian immunodeficiency virus. The inflammatory myopathy may occur as the initial manifestation of a retroviral infection, or myositis may develop later in the disease course. Retroviral antigens have been detected only in occasional endomysial macrophages and not within the muscle fibers themselves, suggesting that persistent infection and viral replication within the muscle do not occur. Histologic findings are identical to retroviral-negative PM or IBM. The infiltrating T cells in the muscle are clonally driven and a number of them are retroviral-specific. This disorder should be distinguished from a toxic myopathy related to long-term therapy with AZT, characterized by fatigue, myalgia, mild muscle weakness, and mild elevation of creatine kinase (CK). AZT-induced myopathy, which generally improves when the drug is discontinued, is a mitochondrial disorder characterized histologically by "ragged-red" fibers. AZT inhibits γ-DNA polymerase, an enzyme found solely in the mitochondrial matrix.

Differential Diagnosis

The clinical picture of the typical skin rash and proximal or diffuse muscle weakness has few causes other than DM. However, proximal muscle weakness without skin involvement can be due to many conditions other than PM or IBM.

SUBACUTE OR CHRONIC PROGRESSIVE MUSCLE WEAKNESS

This may be due to denervating conditions such as the spinal muscular atrophies or amyotrophic lateral sclerosis (Chap. 374). In addition to the muscle weakness, upper motor neuron signs in the latter and signs of denervation detected by electromyography (EMG) aid in the diagnosis. The muscular dystrophies (Chap. 387) may be additional considerations; however, these disorders usually develop over years rather than weeks or months and rarely present after the age of 30 years. It may be difficult, even with a muscle biopsy, to distinguish chronic PM from a rapidly advancing muscular dystrophy. This is particularly true of facioscapulohumeral muscular dystrophy, dysferlin myopathy, and the dystrophinopathies where inflammatory cell infiltration is often found early in the disease. Such doubtful cases should always be given an adequate trial of glucocorticoid therapy and undergo genetic testing to exclude muscular dystrophy. Identification of the MHC/CD8 lesion by muscle biopsy is helpful to identify cases of PM. Some metabolic myopathies, including glycogen storage disease due to myophosphorylase or acid maltase deficiency, lipid storage myopathies due to carnitine deficiency, and mitochondrial diseases produce weakness that is often associated with other characteristic clinical signs; diagnosis rests upon histochemical and biochemical studies of the muscle biopsy. The endocrine myopathies such as those due to hypercorticosteroidism, hyper- and hypothyroidism, and hyper- and hypoparathyroidism require the appropriate laboratory investigations for diagnosis. Muscle wasting in patients with an underlying neoplasm may be due to disuse, cachexia, or rarely to a paraneoplastic neuromyopathy (Chap. 101).

Diseases of the neuromuscular junction, including myasthenia gravis or the Lambert-Eaton myasthenic syndrome, cause fatiguing weakness that also affects ocular and other cranial muscles (Chap. 386). Repetitive nerve stimulation and single-fiber EMG studies aid in diagnosis.

ACUTE MUSCLE WEAKNESS

This may be caused by an acute neuropathy such as Guillain-Barré syndrome (Chap. 385), transverse

myelitis (Chap. 377), a neurotoxin (Chap. 387), or a neurotropic viral infection such as poliomyelitis or West Nile virus (Chap. 381). When acute weakness is associated with very high levels of serum creatine kinase (CK) (often in the thousands), painful muscle cramps, rhabdomyolysis, and myoglobinuria, it may be due to a viral infection or a metabolic disorder such as myophosphorylase deficiency or carnitine palmitoyltransferase deficiency (Chap. 387). Several animal parasites, including protozoa (*Toxoplasma*, *Trypanosoma*), cestodes (cysticerci), and nematodes (trichinae), may produce a focal or diffuse inflammatory myopathy known as *parasitic polymyositis*. *Staphylococcus aureus*, *Yersinia*, *Streptococcus*, or anaerobic bacteria may produce a suppurative myositis, known as *tropical polymyositis*, or *pyomyositis*. Pyomyositis, previously rare in the west, is now occasionally seen in AIDS patients. Other bacteria, such as *Borrelia burgdorferi* (Lyme disease) and *Legionella pneumophila* (Legionnaire's disease), may infrequently cause myositis.

Patients with periodic paralysis experience recurrent episodes of acute muscle weakness without pain, always beginning in childhood. Chronic alcoholics may develop painful myopathy with myoglobinuria after a bout of heavy drinking. Acute painless muscle weakness with myoglobinuria may occur with prolonged hypokalemia, or hypophosphatemia and hypomagnesemia, usually in chronic alcoholics or in patients on nasogastric suction receiving parenteral hyperalimentation.

MYOFASCIITIS

This distinctive inflammatory disorder affecting muscle and fascia presents as diffuse myalgias, skin induration, fatigue, and mild muscle weakness; mild elevations of serum CK are usually present. The most common form is eosinophilic myofasciitis characterized by peripheral blood eosinophilia and eosinophilic infiltrates in the endomysial tissue. In some patients, the eosinophilic myositis/fasciitis occurs in the context of parasitic infections, vasculitis, mixed connective tissue disease, hypereosinophilic syndrome, or toxic exposures (e.g., toxic oil syndrome, contaminated l-tryptophan) or with mutations in the calpain gene. A distinct subset of myofasciitis is characterized by pronounced infiltration of the connective tissue around the muscle by sheets of periodic acid–Schiff-positive macrophages and occasional CD8 T cells (macrophagic myofasciitis). Such histologic involvement is focal and limited to sites of previous vaccinations, which may have been administered months or years earlier. This disorder, which to date has not been observed outside of France, has been linked to an aluminum-containing substrate in vaccines. Most patients respond to glucocorticoid therapy, and the overall prognosis seems favorable.

NECROTIZING MYOSITIS

This is an increasingly recognized entity that has distinct features, even though it is often labeled as PM. It presents often in the fall or winter as an acute or subacute onset of symmetric muscle weakness; CK is typically extremely high. The weakness can be severe. Coexisting interstitial lung disease and cardiomyopathy may be present. The disorder may develop after a viral infection or in association with cancer. Some patients have antibodies against signal recognition particle (SRP). The muscle biopsy demonstrates necrotic fibers infiltrated by macrophages but only rare, if any, T cell infiltrates. Muscle MHC-I expression is only slightly and focally upregulated. The capillaries may be swollen with hyalinization, thickening of the capillary wall, and deposition of complement. Some patients respond to immunotherapy, but others are resistant.

HYPERACUTE NECROTIZING FASCIITIS/MYOSITIS (FLESH-EATING DISEASE)

This a fulminant infectious disease, seen most often in the tropics or in conditions with poor hygiene, characterized by widespread necrosis of the superficial fascia and muscle of a limb; if the scrotum, perineum, and abdominal wall are affected, the condition is referred to as Fournier's gangrene. It may be caused by group A β-hemolytic streptococcus, methicillin-sensitive *S. aureus*, *Pseudomonas aeruginosa*, *Vibrio vulnificus*, clostridial species (gas gangrene; Chap. 142), or polymicrobial infection with anaerobes and facultative bacteria (Chap. 164); toxins from these bacteria may act as superantigens (Chap. 314). The port of bacterial entry is usually a trivial cut or skin abrasion and the source is contact with carriers of the organism. Individuals with diabetes mellitus, immunodeficiency states, or systemic illnesses such as liver failure are most susceptible. Systemic varicella is a predisposing factor in children.

The disease presents with swelling, pain, and redness in the involved area followed by a rapid tissue necrosis of fascia and muscle that progresses at an estimated rate of 3 cm/h. Emergency debridement, antibiotics, as well as IVIg, or even hyperbaric oxygen have been recommended. In progressive or advanced cases, amputation of the affected limb may be necessary to avoid a fatal outcome.

DRUG-INDUCED MYOPATHIES

D-Penicillamine, procainamide, and statins may produce a true myositis resembling PM, and a DM-like illness had been associated with the contaminated preparations of l-tryptophan. As noted above, AZT causes a mitochondrial myopathy. Other drugs may elicit a toxic noninflammatory myopathy that is histologically different from DM, PM, or IBM. These include cholesterol-lowering agents such as clofibrate, lovastatin, simvastatin, or provastatin, especially when combined with cyclosporine, amiodarone, or gemfibrozil. Statin-induced necrotizing myopathy or asymptomatic elevations of CK usually improve after discontinuation of the drug. In rare patients, however, muscle weakness

continues to progress even after the statin is withdrawn; in these cases, a diagnostic muscle biopsy is indicated, and if evidence of inflammation and MHC-I upregulation is present, immunotherapy for PM should be considered. Rhabdomyolysis and myoglobinuria have been rarely associated with amphotericin B, ε-aminocaproic acid, fenfluramine, heroin, and phencyclidine. The use of amiodarone, chloroquine, colchicine, carbimazole, emetine, etretinate, ipecac syrup, chronic laxative or licorice use resulting in hypokalemia, and glucocorticoids or growth hormone administration have also been associated with myopathic muscle weakness. Some neuromuscular blocking agents such as pancuronium, in combination with glucocorticoids, may cause an acute critical illness myopathy. A careful drug history is essential for diagnosis of these drug-induced myopathies, which do not require immunosuppressive therapy except when an autoimmune myopathy has been triggered, as noted above.

"WEAKNESS" DUE TO MUSCLE PAIN AND MUSCLE TENDERNESS

A number of conditions including *polymyalgia rheumatica* (Chap. 326) and arthritic disorders of adjacent joints may enter into the differential diagnosis of inflammatory myopathy, even though they do not cause myositis. The muscle biopsy is either normal or discloses type II muscle fiber atrophy. Patients with *fibrositis* and *fibromyalgia* (Chap. 335) complain of focal or diffuse muscle tenderness, fatigue, and aching, which is sometimes poorly differentiated from joint pain. Some patients, however, have muscle tenderness, painful muscles on movement, and signs suggestive of a collagen vascular disorder, such as an increased erythrocyte sedimentation rate, C-reactive protein, antinuclear antibody, or rheumatoid factor, along with modest elevation of the serum CK and aldolase. They demonstrate a "break-away" pattern of weakness with difficulty sustaining effort but not true muscle weakness. The muscle biopsy is usually normal or nonspecific. Many such patients show some response to nonsteroidal anti-inflammatory agents or glucocorticoids, though most continue to have indolent complaints. An indolent fasciitis in the setting of an ill-defined connective tissue disorder may be present, and these patients should not be labeled as having a psychosomatic disorder. *Chronic fatigue syndrome*, which may follow a viral infection, can present with debilitating fatigue, fever, sore throat, painful lymphadenopathy, myalgia, arthralgia, sleep disorder, and headache (Chap. 389). These patients do not have muscle weakness, and the muscle biopsy is normal.

Diagnosis

The clinically suspected diagnosis of PM, DM, or IBM is confirmed by analysis of serum muscle enzymes, EMG findings, and muscle biopsy **(Table 388-2)**.

Table 388-2 Criteria for Diagnosis of Inflammatory Myopathies

	Polymyositis			
Criterion	**Definite**	**Probable**	**Dermatomyositis**	**Inclusion Body Myositis**
Myopathic muscle weakness[a]	Yes	Yes	Yes[b]	Yes; slow onset, early involvement of distal muscles, frequent falls
Electromyographic findings	Myopathic	Myopathic	Myopathic	Myopathic with mixed potentials
Muscle enzymes	Elevated (up to fiftyfold)	Elevated (up to fiftyfold)	Elevated (up to fiftyfold) or normal	Elevated (up to tenfold) or normal
Muscle biopsy findingsc	"Primary" inflammation with the CD8/MHC-I complex and no vacuoles	Ubiquitous MHC-I expression but minimal inflammation and no vacuoles[d]	Perifascicular, perimysial, or perivascular infiltrates, perifascicular atrophy	Primary inflammation with CD8/MHC-I complex; vacuolated fibers with ß-amyloid deposits; cytochrome oxygenase–negative fibers; signs of chronic myopathy[e]
Rash or calcinosis	Absent	Absent	Present[f]	Absent

[a]Myopathic muscle weakness, affecting proximal muscles more than distal ones and sparing eye and facial muscles, is characterized by a subacute onset (weeks to months) and rapid progression in patients who have no family history of neuromuscular disease, no endocrinopathy, no exposure to myotoxic drugs or toxins, and no biochemical muscle disease (excluded on the basis of muscle-biopsy findings). [b]In some cases with the typical rash, the muscle strength is seemingly normal (dermatomyositis sine myositis); these patients often have new onset of easy fatigue and reduced endurance. Careful muscle testing may reveal mild muscle weakness. [c]See text for details. [d]An adequate trial of prednisone or other immunosuppressive drugs is warranted in probable cases. If, in retrospect, the disease is unresponsive to therapy, another muscle biopsy should be considered to

retrospect, the disease is unresponsive to therapy, another muscle biopsy should be considered to exclude other diseases or possible evolution in inclusion body myositis. [e] If the muscle biopsy does not contain vacuolated fibers but shows chronic myopathy with hypertrophic fibers, primary inflammation with the CD8/MHC-I complex and cytochrome oxygenase–negative fibers, the diagnosis is probable inclusion body myositis. [f] If rash is absent but muscle biopsy findings are characteristic of dermatomyositis, the diagnosis is probable dermatomyositis.

The most sensitive enzyme is CK, which in active disease can be elevated as much as fiftyfold. Although the CK level usually parallels disease activity, it can be normal in some patients with active IBM or DM, especially when associated with a connective tissue disease. The CK is always elevated in patients with active PM. Along with the CK, the serum glutamic-oxaloacetic and glutamate pyruvate transaminases, lactate dehydrogenase, and aldolase may be elevated.

Needle EMG shows myopathic potentials characterized by short-duration, low-amplitude polyphasic units on voluntary activation and increased spontaneous activity with fibrillations, complex repetitive discharges, and positive sharp waves. Mixed potentials (polyphasic units of short and long duration) indicating a chronic process and muscle fiber regeneration are often present in IBM. These EMG findings are not diagnostic of an inflammatory myopathy but are useful to identify the presence of active or chronic myopathy and to exclude neurogenic disorders.

MRI is not routinely used for the diagnosis of PM, DM, or IBM. However, it may provide information or guide the location of the muscle biopsy in certain clinical settings.

Muscle biopsy—in spite of occasional variability in demonstrating all of the typical pathologic findings—is the most sensitive and specific test for establishing the diagnosis of inflammatory myopathy and for excluding other neuromuscular diseases. Inflammation is the histologic hallmark for these diseases; however, additional features are characteristic of each subtype **(Figs. 388-3**, **388-4**, and **388-5)**.

Figure 388-3

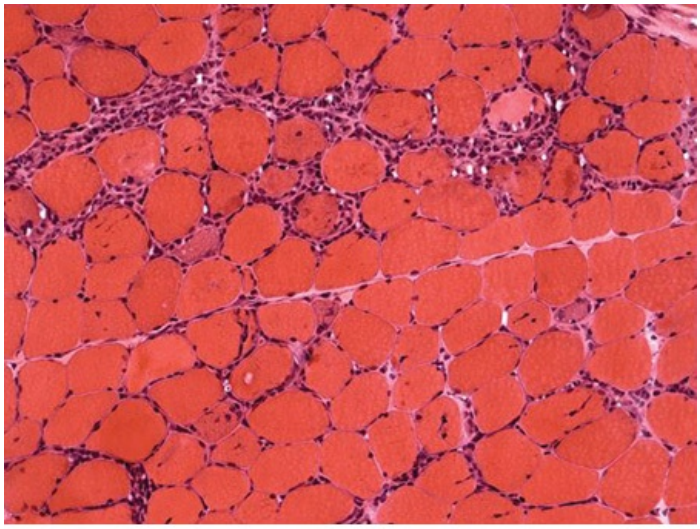

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cross-section of a muscle biopsy from a patient with polymyositis demonstrates scattered inflammatory foci with lymphocytes invading or surrounding muscle fibers. Note lack of chronic myopathic features (increased connective tissue, atrophic or hypertrophic fibers) as seen in inclusion body myositis.

Figure 388-4

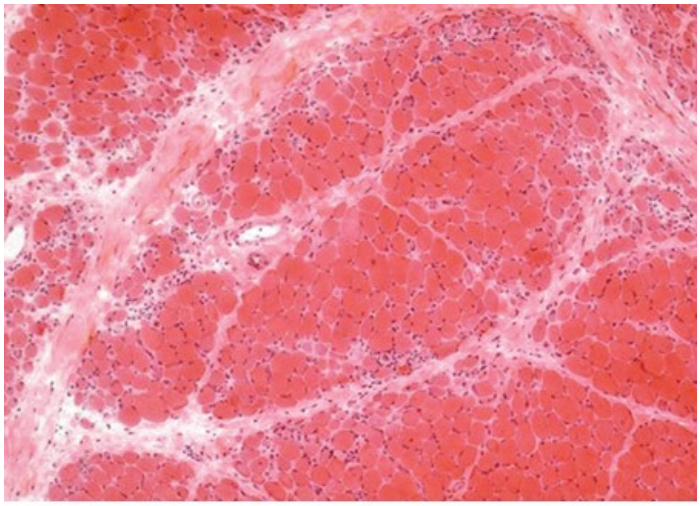

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cross-section of a muscle biopsy from a patient with dermatomyositis demonstrates atrophy of the fibers at the periphery of the fascicle (perifascicular atrophy).

Figure 388-5

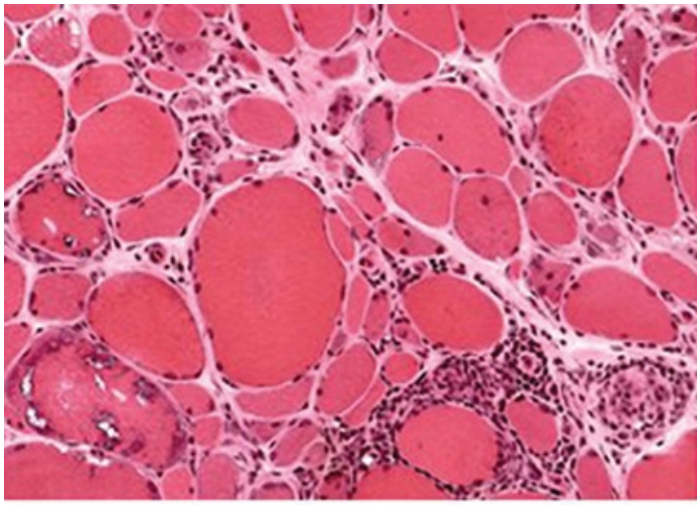

A

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

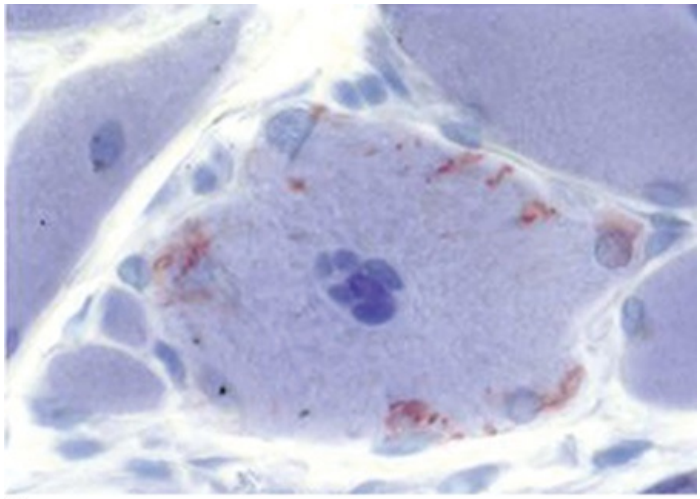

B

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

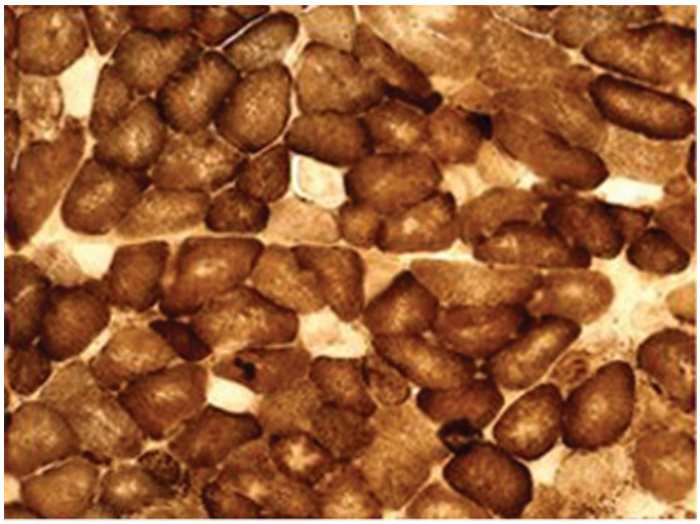

C

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

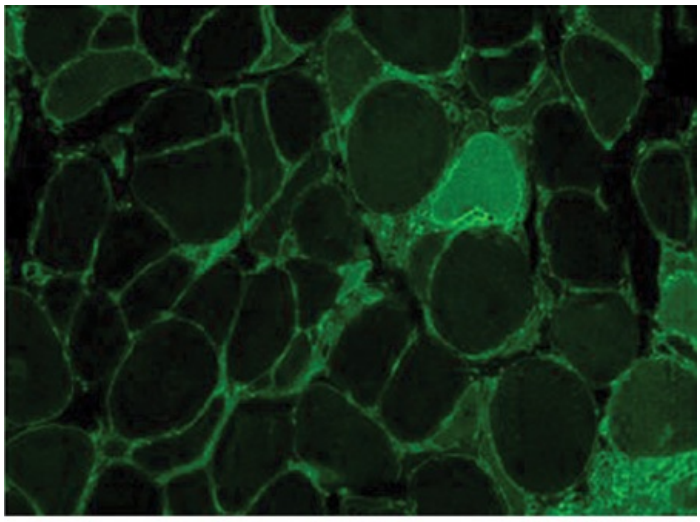

D

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Cross-sections of a muscle biopsy from a patient with inclusion body myositis demonstrate the typical features of vacuoles with lymphocytic infiltrates surrounding nonvacuolated or necrotic fibers (***A***), tiny endomysial deposits of amyloid visualized with crystal violet (***B***), cytochrome oxidase

fibers (**A**), tiny endomysial deposits of amyloid visualized with crystal violet (**B**), cytochrome oxidase–negative fibers, indicative of mitochondrial dysfunction (***C***), and ubiquitous MHC-I expression at the periphery of all fibers (***D***).

In PM the inflammation is *primary*, a term used to indicate that the inflammation is not reactive and the T cell infiltrates, located primarily within the muscle fascicles (endomysially), surround individual, healthy muscle fibers and result in phagocytosis and necrosis (Fig. 388-3). The MHC-I molecule is ubiquitously expressed on the sarcolemma, even in fibers not invaded by CD8+ cells. The CD8/MHC-I lesion is characteristic and essential to confirm or establish the diagnosis and to exclude disorders with secondary, nonspecific, inflammation, such as in some muscular dystrophies. When the disease is chronic, connective tissue is increased and may react positively with alkaline phosphatase.

In DM the endomysial inflammation is predominantly perivascular or in the interfascicular septae and around—rather than within—the muscle fascicles (Fig. 388-4). The intramuscular blood vessels show endothelial hyperplasia with tubuloreticular profiles, fibrin thrombi, and obliteration of capillaries. The muscle fibers undergo necrosis, degeneration, and phagocytosis, often in groups involving a portion of a muscle fasciculus in a wedgelike shape or at the periphery of the fascicle, due to microinfarcts within the muscle. This results in perifascicular atrophy, characterized by 2–10 layers of atrophic fibers at the periphery of the fascicles. The presence of perifascicular atrophy is diagnostic of DM, *even in the absence of inflammation*.

In IBM (Fig. 388-5), there is endomysial inflammation with T cells invading MHC-I-expressing nonvacuolated muscle fibers; basophilic granular deposits distributed around the edge of slitlike vacuoles (rimmed vacuoles); loss of fibers, replaced by fat and connective tissue, hypertrophic fibers, and angulated or round fibers; rare eosinophilic cytoplasmic inclusions; abnormal mitochondria characterized by the presence of ragged-red fibers or cytochrome oxidase–negative fibers; and amyloid deposits within or next to the vacuoles best visualized with crystal violet or Congo-red staining viewed with fluorescent optics. Electron microscopy demonstrates filamentous inclusions in the vicinity of the rimmed vacuoles. In at least 15% of patients with the typical clinical phenotype of IBM, no vacuoles or amyloid deposits can be identified in muscle biopsy, leading to an erroneous diagnosis of PM. Close clinicopathologic correlations are essential; if uncertain, a repeat muscle biopsy from another site is often helpful.

Treatment: Therapy of Inflammatory Myopathies

The goal of therapy is to improve muscle strength, thereby improving function in activities of daily living, and ameliorate the extramuscular manifestations (rash, dysphagia, dyspnea, fever). When strength improves, the serum CK falls concurrently; however, the reverse is not always true. Unfortunately, there is a common tendency to "chase" or treat the CK level instead of the muscle weakness, a practice that has led to prolonged and unnecessary use of immunosuppressive drugs and erroneous assessment of their efficacy. It is prudent to discontinue these drugs if, after an adequate trial, there is no objective improvement in muscle strength whether or not CK levels are reduced. Agents used in the treatment of PM and DM include the following:

1. Glucocorticoids. Oral prednisone is the initial treatment of choice; the effectiveness and side effects of this therapy determine the future need for stronger immunosuppressive drugs. High-dose prednisone, at least 1 mg/kg per day, is initiated as early in the disease as possible. After 3–4 weeks, prednisone is tapered slowly over a period of 10 weeks to 1 mg/kg every other day. If there is evidence of efficacy and no serious side effects, the dosage is then further reduced by 5 or 10 mg every 3–4 weeks until the lowest possible dose that controls the disease is reached. The efficacy of prednisone is determined by an objective increase in muscle strength and activities of daily living, which almost always occurs by the third month of therapy. A feeling of increased energy or a reduction of the CK level without a concomitant increase in muscle strength is not a reliable sign of improvement. If prednisone provides no objective benefit after ~3 months of high-dose therapy, the disease is probably unresponsive to the drug and tapering should be accelerated while the next-in-line immunosuppressive drug is started. Although controlled trials have not been performed, almost all patients with true PM or DM respond to glucocorticoids to *some degree and for some period of time* ; in general, DM responds better than PM.

The long-term use of prednisone may cause increased weakness associated with a normal or unchanged CK level; this effect is referred to as *steroid myopathy*. In a patient who previously responded to high doses of prednisone, the development of new weakness may be related to steroid myopathy or to disease activity that either will respond to a higher dose of glucocorticoids or has become glucocorticoid-resistant. In uncertain cases, the prednisone dosage can be steadily increased or decreased as desired: the cause of the weakness is usually evident in 2–8 weeks.

2. Other immunosuppressive drugs. Approximately 75% of patients ultimately require additional treatment. This occurs when a patient fails to respond adequately to glucocorticoids after a 3-month trial, the patient becomes glucocorticoid-resistant, glucocorticoid-related side effects appear, attempts to lower the prednisone dose repeatedly result in a new relapse, or rapidly progressive disease with evolving severe weakness and respiratory failure develops.

The following drugs are commonly used but have never been tested in controlled studies: (1)

Azathioprine is well tolerated, has few side effects, and appears to be as effective for long-term therapy as other drugs. The dose is up to 3 mg/kg daily. (2) *Methotrexate* has a faster onset of action than azathioprine. It is given orally starting at 7.5 mg weekly for the first 3 weeks (2.5 mg every 12 h for 3 doses), with gradual dose escalation by 2.5 mg per week to a total of 25 mg weekly. A rare side effect is methotrexate pneumonitis, which can be difficult to distinguish from the interstitial lung disease of the primary myopathy associated with Jo-1 antibodies (described above). (3) *Mycophenolate mofetil* also has a faster onset of action than azathioprine. At doses up to 2.5 or 3 gm/d in two divided doses, it is well tolerated for long-term use. (4) Monoclonal anti-CD20 antibody (rituximab) has been shown in a small uncontrolled series to benefit patients with DM and PM. (5) *Cyclosporine* has inconsistent and mild benefit. (6) *Cyclophosphamide* (0.5–1 g/m^2 IV monthly for 6 months) has limited success and significant toxicity. (7) Tacrolimus (formerly known as Fk506) has been effective in some difficult cases of PM.

3. Immunomodulation. In a controlled trial of patients with refractory DM, intravenous immunoglobulin (IVIg) improved not only strength and rash but also the underlying immunopathology. The benefit is often short-lived (≤8 weeks), and repeated infusions every 6–8 weeks are generally required to maintain improvement. A dose of 2 g/kg divided over 2–5 days per course is recommended. Uncontrolled observations suggest that IVIg may also be beneficial for patients with PM. Neither plasmapheresis nor leukapheresis appears to be effective in PM and DM.

The following sequential empirical approach to the treatment of PM and DM is suggested: *Step 1*: high-dose prednisone; *Step 2*: azathioprine, mycophenolate, or methotrexate for steroid-sparing effect; *Step 3*: IVIg; *Step 4*: a trial, with guarded optimism, of one of the following agents, chosen according to the patient's age, degree of disability, tolerance, experience with the drug, and general health: rituximab, cyclosporine, cyclophosphamide, or tacrolimus. Patients with interstitial lung disease may benefit from aggressive treatment with cyclophosphamide or tacrolimus.

A patient with presumed PM who has not responded to any form of immunotherapy most likely has IBM or another disease, usually a metabolic myopathy, a muscular dystrophy, a drug-induced myopathy, or an endocrinopathy. In these cases, a repeat muscle biopsy and a renewed search for another cause of the myopathy is indicated.

Calcinosis, a manifestation of DM, is difficult to treat; however, new calcium deposits may be prevented if the primary disease responds to the available therapies. Bisphosphonates, aluminum hydroxide, probenecid, colchicine, low doses of warfarin, calcium blockers, and surgical excision have all been tried without success.

IBM is generally resistant to immunosuppressive therapies. Prednisone together with azathioprine or methotrexate is often tried for a few months in newly diagnosed patients, although results are generally disappointing. Because occasional patients may feel subjectively weaker after these drugs are discontinued, some clinicians prefer to maintain these patients on low-dose, every-other-day prednisone along with mycophenolate in an effort to slow disease progression, even though there is no objective evidence or controlled study to support this practice. In two controlled studies of IVIg in IBM, minimal benefit in up to 30% of patients was found; the strength gains, however, were not of sufficient magnitude to justify its routine use. Another trial of IVIg combined with prednisone was ineffective. Nonetheless, many experts believe that a 2- to 3-month trial with IVIg may be reasonable for selected patients with IBM who experience rapid progression of muscle weakness or choking episodes due to worsening dysphagia.

Prognosis

The 5-year survival rate for treated patients with PM and DM is ~95% and the 10-year survival rate is 84%; death is usually due to pulmonary, cardiac, or other systemic complications. The prognosis is worse for patients who are severely affected at presentation, when initial treatment is delayed, and in cases with severe dysphagia or respiratory difficulties. Older patients, and those with associated cancer also have a worse prognosis. DM responds more favorably to therapy than PM and thus has a better prognosis. Most patients improve with therapy, and many make a full functional recovery, which is often sustained with maintenance therapy. Up to 30% may be left with some residual muscle weakness. Relapses may occur at any time.

IBM has the least favorable prognosis of the inflammatory myopathies. Most patients will require the use of an assistive device such as a cane, walker, or wheelchair within 5–10 years of onset. In general, the older the age of onset in IBM, the more rapidly progressive is the course.

FURTHER READINGS

Askanas V et al: Inclusion body myositis: A degenerative muscle disease associated with intra-muscle fiber multi-protein aggregates, proteasome inhibition, endoplasmic reticulum stress and decreased lysosomal degradation. Brain Pathol 19:493, 2009[PMID: 19563541] [Full Text]

Chanin N, Engel AG: Correlation of muscle biopsy, clinical course and outcome in polymyositis and

sporadic IBM. Neurology 70:418, 2008

Dalakas MC: Signaling pathways and immunobiology of inflammatory myopathies. Nat Clin Pract Rheumatol 2:219, 2006[PMID: 16932688] [Full Text]

———: Sporadic inclusion body myositis: Diagnosis, pathogenesis and therapeutic strategies. Nat Clin Pract Neurol 2:437, 2006

———: Toxic and drug-induced myopathies. J Neurol Neurosurg Psych 80:832, 2009

———: Immunotherapy of myositis: Issues, concerns and future prospects. Nat Rev Rheumatol 6:129, 2010

———: Review: An update on inflammatory and autoimmune myopathies. Neuropathol Appl Neurobiol 37:226, 2011

——— et al: Inclusion body myositis with human immunodeficiency virus infection: Four cases with clonal expansion of viral-specific T cells. Ann Neurol 61:466, 2007

Engel AG et al: The polymyositis and dermatomyositis syndrome, in Myology, AG Engel, C Franzini-Armstrong (eds). New York, McGraw-Hill, 2008, pp 1335–1383

Marie I et al: Polymyositis and dermatomyositis: Short-term and long-term outcome and predictive factors of prognosis. J Rheumatol 28:2230, 2001[PMID: 11669162] [Full Text]

Mastaglia FL: Inflammatory myopathies: Clinical, diagnostic and therapeutic aspects. Muscle Nerve 27:407, 2003[PMID: 12661042] [Full Text]

Mikol J, Engel AG: Inclusion body myositis, in: *Myology*, 3rd ed, AG Engel, C Franzini-Armstrong (eds). New York, McGraw-Hill, 2004, pp 1367–1388

Needham M, Mastaglia FL: Inclusion body myositis: Current pathogenic concepts and diagnostic and therapeutic approaches. Lancet Neurol 6:620, 2007[PMID: 17582362] [Full Text]

Sontheimer RD: Dermatomyositis: An overview of recent progress with emphasis on dermatologic aspects. Dermatol Clin 20:387, 2002[PMID: 12170874] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

SPECIAL ISSUES IN INPATIENT NEUROLOGIC CONSULTATION: INTRODUCTION

Inpatient neurologic consultations usually involve questions about specific disease processes or prognostication after various cerebral injuries. Common reasons for neurologic consultation include stroke (Chap. 370), seizures (Chap. 369), altered mental status (Chap. 25), headache (Chap. 14), and management of coma and other neurocritical care conditions (Chaps. 274 and 275). This chapter focuses on additional common reasons for consultation that are not addressed elsewhere in the text.

CONSULTATIONS REGARDING CENTRAL NERVOUS SYSTEM DYSFUNCTION

Hyperperfusion States

A group of neurologic disorders shares the common feature of hyperperfusion playing a key role in pathogenesis. These seemingly diverse syndromes include hypertensive encephalopathy, eclampsia, postcarotid endarterectomy syndrome, and toxicity from calcineurin-inhibitor medications. Modern imaging techniques and experimental models suggest that vasogenic edema is usually the primary process leading to neurologic dysfunction; therefore, prompt recognition and management of this condition should allow for clinical recovery if superimposed hemorrhage or infarction has not occurred.

The brain's autoregulatory capability successfully maintains a fairly stable cerebral blood flow in adults despite alterations in systemic mean arterial pressure (MAP) ranging from 50–150 mm Hg. In patients with chronic hypertension, this cerebral autoregulation curve is shifted, resulting in autoregulation working over a much higher range of pressures (e.g., 70–175 mm Hg). In these hypertensive patients, cerebral blood flow is kept steady at higher MAP, but a rapid lowering of pressure can more easily lead to ischemia on the lower end of the autoregulatory curve. This autoregulatory phenomenon is achieved through both myogenic and neurogenic influences causing small arterioles to contract and dilate. When the systemic blood pressure exceeds the limits of this mechanism, breakthrough of autoregulation occurs, resulting in hyperperfusion via increased cerebral blood flow, capillary leakage into the interstitium, and resulting edema. The predilection of all of the hyperperfusion disorders to affect the posterior rather than anterior portions of the brain may be due to a lower threshold for autoregulatory breakthrough in the posterior circulation.

While elevated or relatively elevated blood pressure is common in many of these disorders, some hyperperfusion states such as calcineurin-inhibitor toxicity occur with no apparent pressure rise. In these cases, vasogenic edema is likely due primarily to dysfunction of the capillary endothelium itself, leading to breakdown of the blood-brain barrier. It is useful to separate disorders of hyperper-fusion into those caused primarily by increased pressure and those due mostly to endothelial dysfunction from a toxic or autoimmune etiology **(Table e47–1)**. In reality, both of these pathophysiologic processes are likely playing some role in each of these disorders.

Table e47–1 Some Common Etiologies of Hyperperfusion Syndrome

Disorders in which increased capillary pressure dominates the pathophysiology
Hypertensive encephalopathy, including secondary causes such as renovascularhypertension, pheochromocytoma, cocaine use, etc.
Postcarotid endarterectomy syndrome
Preeclampsia/eclampsia
High-altitude cerebral edema
Disorders in which endothelial dysfunction dominates the pathophysiology
Calcineurin-inhibitor toxicity
Chemotherapeutic agent toxicity (e.g., cytarabine, azathioprine, 5-fluorouracil, cisplatin, methotrexate)
Glucocorticoids
Erythropoietin
HELLP syndrome (*h*emolysis, *e*levated *l*iver enzyme *l*evels, *l*ow *p*latelet count)
Thrombotic thrombocytopenic purpura (TTP)
Hemolytic uremic syndrome (HUS)
Systemic lupus erythematosus (SLE)
Granulomatosis with polyangiitis (Wegener's)

The clinical presentation of the hyperperfusion syndromes is similar with prominent headaches, seizures, or focal deficits. Headaches have no specific characteristics, range from mild to severe, and may be accompanied by alterations in consciousness ranging from confusion to coma. Seizures may be present, and these can be of multiple types depending on the severity and location of the edema. Nonconvulsive seizures have been described in hyperperfusion states; therefore, a low threshold for obtaining an electroencephalogram (EEG) in these patients should be maintained. The typical focal deficit in hyperperfusion states is cortical visual loss, given the tendency of the process to involve the occipital lobes. However, any focal deficit can occur depending on the area affected, as evidenced by patients who, after carotid endarterectomy, exhibit neurologic dysfunction in the ipsilateral newly reperfused hemisphere. In conditions where increased cerebral blood flow plays a role, examination of the inpatient

vital signs record will usually reveal a systemic blood pressure that is increased above baseline. It appears as if the rapidity of rise rather than the absolute value of pressure is the most important risk factor.

The diagnosis in all of these conditions is clinical. The symptoms of these disorders are common and nonspecific, so a long differential diagnosis should be entertained, including consideration of other causes of confusion, focal deficits, headache, and seizures. MRI has improved the ability of clinicians to diagnose hyperperfusion syndromes, although cases have been reported with normal imaging. Patients classically exhibit the high T2 signal of edema primarily in the posterior occipital lobes, not respecting any single vascular territory **(Fig. e47-1)**. Diffusion-weighted images are typically normal, emphasizing the vasogenic rather than cytotoxic nature of this edema. Imaging with CT is less sensitive but may show a pattern of patchy hypodensity in the involved territory. Previously this classic radiographic appearance had been termed *reversible posterior leukoencephalopathy* (RPLE). However, this term has fallen out of favor because none of its elements are completely accurate. The radiographic and clinical changes are not always reversible; the territory involved is not uniquely posterior; and gray matter may be affected as well, rather than purely white matter as the word "leukoencephalopathy" intimates. Other ancillary studies such as cerebrospinal fluid (CSF) analysis often yield nonspecific results. It should be noted that many of the substances that have been implicated, such as cyclosporine, can cause this syndrome even at low doses or after years of treatment. Therefore, normal serum levels of these medications do not exclude them as inciting agents.

Figure e47-1

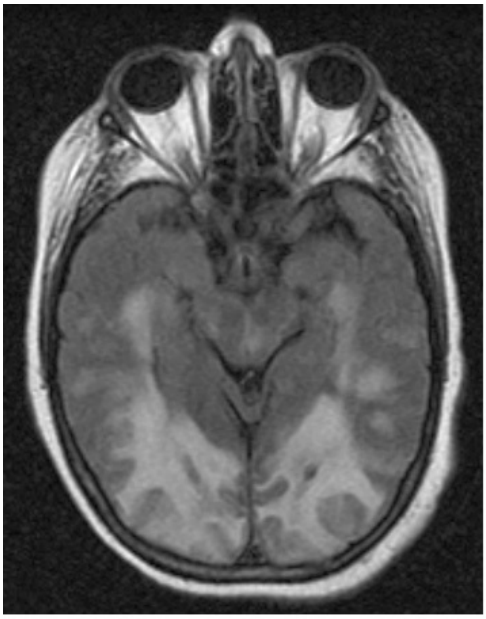

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Axial fluid-attenuated inversion recovery (FLAIR) MRI of the brain in a patient taking cyclosporine after liver transplantation, who presented with seizures, headache, and cortical blindness. Increased signal is seen bilaterally in the occipital lobes predominantly involving the white matter, consistent with a hyperperfusion state secondary to calcineurin-inhibitor exposure.

In cases of hyperperfusion syndromes, treatment should commence urgently once the diagnosis is considered. Hypertension plays a key role commonly, and judicious lowering of the blood pressure with IV agents such as labetalol or nicardipine is advised along with continuous cardiac and blood pressure monitoring, often through an arterial line. It is reasonable to lower mean arterial pressure by ~20% initially, as further lowering of the pressure may cause secondary ischemia as pressure drops below the lower range of the patient's autoregulatory capability. In cases where there is an identified cause of the syndrome, these etiologies should be treated promptly, including discontinuation of offending substances such as calcineurin inhibitors in toxic processes, treatment of immune-mediated disorders such as thrombotic thrombocytopenic purpura (TTP), and prompt delivery of the fetus in eclampsia. Seizures must be identified and controlled, often necessitating continuous EEG monitoring. Anticonvulsants are effective, but in the special case of eclampsia, there is good evidence to support the use of magnesium sulfate for seizure control.

Post-Cardiac Bypass Brain Injury

Central nervous system (CNS) injuries following open heart or coronary artery bypass grafting (CABG) surgery are common and include acute encephalopathy, stroke, and a chronic syndrome of cognitive impairment, which is now increasingly recognized. Hypoperfusion and embolic disease are frequently involved in the pathogenesis of these syndromes, although multiple mechanisms may be involved in these critically ill patients who are at risk for various metabolic and polypharmaceutical complications.

The frequency of hypoxic injury secondary to inadequate blood flow intraoperatively has been markedly decreased by the use of modern surgical and anesthetic techniques. Despite these advances, some patients still experience neurologic complications from cerebral hypoperfusion or may suffer focal ischemia from tight carotid or focal intracranial stenoses in the setting of regional hypoper-fusion. Postoperative infarcts in the border zones between vascular territories commonly are blamed on systemic hypotension although some have suggested that these infarcts can also result from embolic disease **(Fig. e47-2)**.

Figure e47-2

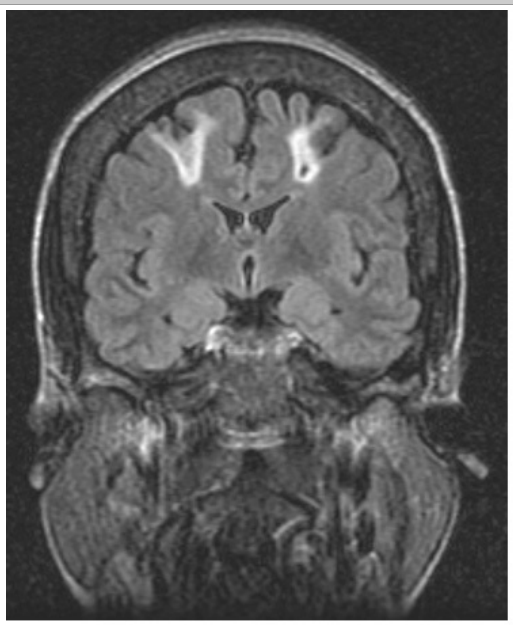

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Coronal fluid-attenuated inversion recovery (FLAIR) MRI of the brain in a patient presenting with altered mental status after an episode of hypotension during coronary artery bypass grafting (CABG). Increased signal is seen in the border zones bilaterally between the middle cerebral artery and anterior cerebral artery territories. Diffusion-weighted MRI sequences demonstrated restricted diffusion in these same locations, suggesting acute infarction.

Embolic disease is likely the predominant mechanism of cerebral injury during cardiac surgery as evidenced by diffusion-weighted MRI and intraoperative transcranial Doppler studies. It should be noted that some of the emboli that are found histologically in these patients are too small to be detected by standard imaging sequences; therefore, a negative MRI after surgery does not exclude the diagnosis of emboli-related complications. Thrombus in the heart itself as well as atheromas in the aortic arch can become dislodged during cardiac surgeries, releasing a shower of particulate matter into the cerebral circulation. Cross-clamping of the aorta, manipulation of the heart, extracorporeal circulation techniques ("bypass"), arrhythmias such as atrial fibrillation, and introduction of air through suctioning have all been implicated as potential sources of emboli. Histologic studies indicate that literally millions of tiny emboli may be released, even using modern surgical techniques.

This shower of microemboli results in a number of clinical syndromes. Occasionally, a single large embolus leads to an isolated large-vessel stroke that presents with obvious clinical focal deficits. More commonly, the emboli released are multiple and smaller. When there is a high burden of these small emboli, an acute encephalopathy can occur postoperatively, presenting as either a hyperactive or hypoactive confusional state, the latter of which is frequently and incorrectly ascribed to depression. When the burden of microemboli is lower, no acute syndrome is recognized, but the patient may suffer a chronic cognitive deficit. Cardiac surgery can be viewed, like delirium, as a "stress test for the brain." Some patients with a low cerebral reserve due to underlying cerebrovascular disease or an early neurodegenerative process will develop a chronic, cognitive deficit, whereas others with higher reserves may remain asymptomatic despite a similar dose of microemboli. In this manner, cardiac surgery may serve to unmask the early manifestations of disorders such as vascular dementia and Alzheimer's disease.

Since modern techniques have successfully minimized hypoperfusion complications during these surgeries, much attention is now focused on reducing this inevitable shower of microemboli. Off-pump CABG surgeries have the advantages of reducing length of stay and perioperative complications; however, some recent data suggests that off-pump CABG does not preserve cognitive function compared with on-pump CABG. Filters placed in the aortic arch may have some promise in capturing these emboli, although convincing evidence is currently lacking. Development of successful endovascular operative approaches may provide a reasonable alternative to conventional CABG procedures, especially for patients at high risk of developing cognitive dysfunction after surgery due to advanced age, previous stroke, or severe atheromatous disease of the carotid arteries or aortic arch.

Post-Solid Organ Transplant Brain Injury

Patients who have undergone solid organ transplantation are at risk for neurologic injury in the postoperative period and for the months to years thereafter. Neurologic consultants should view these patients as a special population at risk for both unique neurologic complications as well as for the usual disorders found in any critically ill inpatient.

Immunosuppressive medications are administered in high doses to patients after solid organ transplant, and many of these compounds have well-described neurologic complications. In patients with headache, seizures, or focal neurologic deficits taking calcineurin inhibitors, the diagnosis of hyperperfusion syndrome should be considered, as discussed above. This neurotoxicity occurs mainly with cyclosporine and tacrolimus and can present even in the setting of normal serum drug levels. Treatment primarily involves lowering the drug dosage or discontinuing the drug. A related newer agent, sirolimus, has very few recorded cases of neurotoxicity and may be a reasonable alternative for some patients. Other examples of immunosuppressive medications and their neurologic complications include OKT3-associated akinetic mutism and the leukoencephalopathy seen with methotrexate, especially when it is

administered intrathecally or with concurrent radiotherapy. In any solid organ transplant patient with neurologic complaints, a careful examination of the medication list is required to search for these possible drug effects.

Cerebrovascular complications of solid organ transplant are often first recognized in the immediate postoperative period. Border zone territory infarctions can occur, especially in the setting of systemic hypotension during cardiac transplant surgery. Embolic infarctions classically complicate cardiac transplantation, but all solid organ transplant procedures place patients at risk for systemic emboli. When cerebral embolization accompanies renal or liver transplantation surgery, a careful search for right-to-left shunting should include evaluation of the heart with agitated saline echocardiography, as well as looking for intrapulmonary shunting. Renal and some cardiac transplant patients often have advanced atherosclerosis, providing yet another mechanism for stroke. Imaging with CT or MRI with diffusion is advised when cerebrovascular complications are suspected to confirm the diagnosis and to exclude intracerebral hemorrhage, which most often occurs in the setting of coagulopathy secondary to liver failure or after cardiac bypass procedures.

Given that patients with solid organ transplants are chronically immunosuppressed, infections are a common concern (Chap. 132). In any transplant patient with new CNS signs or symptoms such as seizure, confusion, or focal deficit, the diagnosis of a central nervous system infection should be considered and evaluated through imaging (usually MRI) and possibly lumbar puncture. The most common pathogens responsible for CNS infections in these patients vary based on time since transplant. In the first month posttransplant, common pathogens include the usual bacterial organisms associated with surgical procedures and indwelling catheters. Starting in the second month posttransplant, opportunistic infections of the CNS become more common, including *Nocardia* and *Toxoplasma* species as well as fungal infections such as aspergillosis. Viral infections that can affect the brain of the immunosuppressed patient, such as herpes simplex virus, cytomegalovirus, and varicella, also become more common after the first month posttransplant. After 6 months posttransplant, immunosuppressed patients still remain at risk for these opportunistic bacterial, fungal, and viral infections but can also suffer late CNS infectious complications such as progressive multifocal leukoencephalopathy (PML) associated with JC virus and Epstein-Barr virus–driven clonal expansions of B cells resulting in CNS lymphoma.

COMMON NEUROLOGIC COMPLICATIONS OF ELECTROLYTE DISTURBANCES

A wide variety of neurologic conditions can result from abnormalities in serum electrolytes, and consideration of electrolyte disturbances should be part of any inpatient neurologic consultation. **A complete general discussion of fluid and electrolyte imbalance and homeostasis can be found in Chap. 45.**

Hypernatremia and Hyperosmolality

The normal range of serum osmolality is around 275–295 mOsm/kg, but neurologic manifestations are usually seen only at levels >325 mOsm/kg. Hyperosmolality is usually due to hypernatremia, hyperglycemia, azotemia, or the addition of extrinsic osmoles such as mannitol, which is commonly used in critically ill neurologic patients. Hyperosmolality itself can lead to a generalized encephalopathy that is nonspecific and without focal findings; however, an underlying lesion such as a mass can become symptomatic under the metabolic stress of a hyperosmolar state, producing focal signs. Some patients with hyperosmolality from severe hyperglycemia can present, for unclear reasons, with generalized seizures or unilateral movement disorders, which usually respond to lowering of the serum glucose. The treatment of all forms of hyperosmolality involves calculation of apparent water losses and slow replacement so that the serum sodium declines no faster than 2 mmol/L (2 meq/L) per hour.

Hypernatremia leads to the loss of intracellular water, leading to cell shrinkage. In the cells of the brain, solutes such as glutamine and urea are generated under these conditions in order to minimize this shrinkage. Despite this corrective mechanism, when hyper-natremia is severe [serum sodium >160 mmol/L (>160 meq/L)] or occurs rapidly, cellular metabolic processes fail and encephalopathy will result. There are many etiologies of hypernatremia including, most commonly, renal and extrarenal losses of water. Causes of neurologic relevance include central diabetes insipidus, where hyperosmolality is accompanied by submaximal urinary concentration due to inadequate release of arginine vasopressin (AVP) from the posterior pituitary, resulting often from pituitary injury in the setting of surgery, hemorrhage, infiltrative processes, or cerebral herniation.

Hyponatremia

Hyponatremia is commonly defined as a serum sodium <135 mmol/L (<135 meq/L). Neurologic symptoms occur at different levels of low sodium, depending not only on the absolute value but also on the rate of fall. In patients with hyponatremia that develops over hours, life-threatening seizures and cerebral edema may occur at values as high as 125 mmol/L. In contrast, some patients with more chronic hyponatremia that has slowly developed over months to years may be asymptomatic even with serum levels <110 mmol/L. Correction of hyponatremia, especially when chronic, must take place slowly in order to avoid additional neurologic complications. Cells in the brain swell in hypotonic hyponatremic states but may compensate over time by excreting solute into the extracellular space, leading to restoration of cell volume when water follows the solute out of the cells. If treatment of hyponatremia results in a rapid rise in serum sodium, cells in the brain may quickly shrink, leading to osmotic demyelination, a process that previously was thought to be limited exclusively to the brainstem (central pontine myelinolysis; see Fig. 275-6), but now has been described elsewhere in the CNS.

Treatment of hyponatremia is dependent on the cause. Hypertonic hyponatremia treatment focuses on the underlying condition, such as hyperglycemia. Isovolemic hyponatremia [syndrome of inappropriate antidiuretic hormone (SIADH)] is managed with water restriction or administration of AVP antagonists. The management of choice for patients with hypervolemic hypotonic hyponatremia is free-water restriction and treatment of the underlying edematous disorder, such as nephrotic syndrome or congestive heart failure. Finally, in hypovolemic hypotonic hyponatremia, volume is replaced with isotonic saline while underlying conditions of the kidneys, adrenals, and gastrointestinal tract are addressed.

One neurologic cause of hypovolemic hypotonic hyponatremia is the cerebral salt-wasting syndrome that accompanies subarachnoid hemorrhage and, less commonly, other cerebral processes such as meningitis or stroke. In these cases, the degree of renal sodium excretion can be remarkable, and large amounts of saline, hypertonic saline, or oral sodium may need to be given in a judicious fashion in order to avoid complications from cerebral edema.

Hypokalemia

Hypokalemia, defined as a serum potassium level < 3.5 mmol/L (<3.5 meq/L), occurs either because of excessive potassium losses (from the kidneys or gut) or due to an abnormal potassium distribution between the intracellular and extracellular spaces. At very low levels (<1.5 mmol/L), hypokalemia may be life threatening due to the risk of cardiac arrhythmia and may present neurologically with severe muscle weakness and paralysis. Hypokalemic periodic paralysis is a rare disorder caused by excessive intracellular potassium uptake in the setting of a calcium or sodium channel mutation. Treatment of hypokalemia is dependent on the etiology but usually includes replacement of potassium through oral or IV routes as well as correcting the cause of potassium balance problems (e.g., eliminating β_2-adrenergic agonist medications).

Hyperkalemia

Hyperkalemia is defined as a serum potassium level >5.5 mmol/L (>5.5 meq/L) and can neurologically present as muscle weakness with or without paresthesias. Hyperkalemia becomes life threatening when it produces electrocardiographic abnormalities such as peaked T waves or a widened QRS complex. In these cases, prompt treatment is essential and consists of strategies that protect the heart against arrhythmias (calcium gluconate administration); promote potassium redistribution into cells (with glucose, insulin, and β_2-agonist medications); and increase potassium removal (through sodium polystyrene sulfonate, loop diuretics, or hemodialysis).

Calcium Disturbances

Hypercalcemia usually occurs in the setting of either hyperparathyroidism or systemic malignancy. Neurologic manifestations include encephalopathy as well as muscle weakness due to reduced neuromuscular excitability. Seizures can occur but are more common in states of low calcium.

Hypocalcemia in adults often follows surgical treatment of the thyroid or parathyroid. Seizures and altered mental status dominate the neurologic picture and usually resolve with calcium repletion. Tetany is due to spontaneous, repetitive action potentials in peripheral nerves and remains the classic sign of symptomatic hypocalcemia.

Magnesium Disturbances

Disorders of magnesium are difficult to correlate with serum levels because a very small amount of total-body magnesium is located in the extracellular space. Hypomagnesemia presents neurologically with seizures, tremor, and myoclonus. When intractable seizures occur in the setting of hypomagnesemia, only administration of magnesium will lead to resolution. High levels of magnesium, in contrast, lead to CNS depression. Hypermagnesemia usually only occurs in the setting of renal failure or magnesium administration and can lead to confusion and muscular paralysis when severe.

CONSULTATIONS REGARDING PERIPHERAL NERVOUS SYSTEM DYSFUNCTION

Entrapment Neuropathies

Polyneuropathy is a common cause of outpatient neurologicconsultation (Chap. 384). In the inpatient setting, however, mononeuropathies are more frequent, especially the entrapment neuropathies that complicate many surgical procedures and medical conditions. Median neuropathy at the wrist (carpal tunnel syndrome) is the most frequent entrapment neuropathy by far, but it is rarely a cause for inpatient consultation. Mechanisms for perioperative mononeuropathy include traction, compression, and ischemia of the nerve. Imaging with MR neurography may allow these causes to be distinguished definitively. In all cases of mononeuropathy, the diagnosis can be made through the clinical examination and then confirmed with electrodiagnostic studies in the subacute period, if necessary. Treatment consists mainly of avoidance of repetitive trauma but may also include surgical approaches to relieve pressure on the nerve.

Radial Neuropathy

Radial nerve injury classically presents with weakness of extension of the wrist and fingers ("wrist drop") with or without more proximal weakness of extensor muscles of the upper extremity, depending on the site of injury. Sensory loss is in the distribution of the radial nerve, which includes the dorsum of the hand **(Fig. e47-3*A*)**. Compression at the level of the axilla, e.g., resulting from use of crutches, includes weakness of the triceps, brachioradialis, and supinator muscles in addition to wrist drop. A more common site of compression occurs in the spiral groove of the upper arm in the setting of a humerus fracture or from sleeping with the arm draped over a bench or chair ("Saturday night palsy"). Sparing of the triceps is the rule when the nerve is injured in this location. Because extensors of the upper extremity are injured preferentially in radial nerve injury, these lesions may be mistaken for the pyramidal distribution of weakness that accompanies upper motor neuron lesions from brain or spinal cord processes.

Figure e47-3

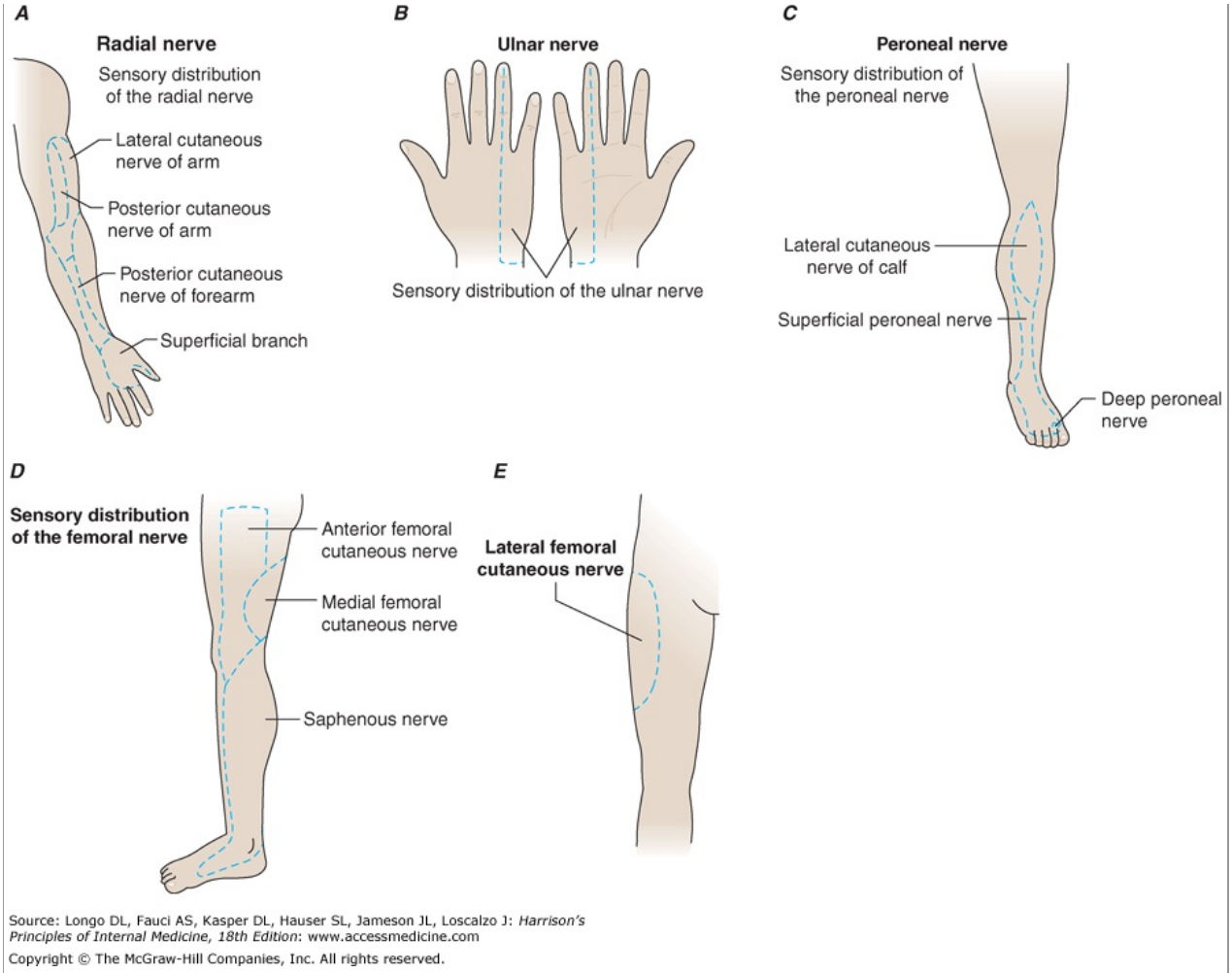

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Sensory distribution of peripheral nerves commonly affected by entrapment neuropathies.*A.* Radial nerve. ***B.*** Ulnar nerve. ***C.*** Peroneal nerve. ***D.*** Femoral nerve. ***E.*** Lateral femoral cutaneous nerve.

Ulnar Neuropathy

Compression of the ulnar nerve is the second most common entrapment neuropathy after carpal tunnel syndrome. The most frequent site of compression is at the elbow where the nerve passes superficially in the ulnar groove. Symptoms usually begin with tingling in the ulnar distribution, including the fourth and fifth digits of the hand (Fig. e47-3B). Sensory symptoms may be worsened by elbow flexion due to increased pressure on the nerve, hence the tendency of patients to complain of increasing paresthesias at night when the arm is flexed at the elbow during sleep. Motor dysfunction can be disabling and involves most of the intrinsic hand muscles, limiting dexterity and strength of grasp and pinch. Etiologies of ulnar entrapment include trauma to the nerve (hitting the "funny bone"), malpositioning during anesthesia for surgical procedures, and chronic arthritis of the elbow. When a perioperative ulnar nerve injury is considered, stretch injury or trauma to the lower trunk of the brachial plexus should be entertained as well since its symptoms can mimic those of an ulnar neuropathy. If the clinical examination is equivocal, electrodiagnostic studies can definitively distinguish between plexus and ulnar nerve lesions a few weeks after the injury. Conservative methods of treatment are often the first step, but a variety of surgical approaches may be effective, including anterior ulnar nerve transposition and release of the flexor carpi ulnaris aponeurosis.

Peroneal Neuropathy

The peroneal nerve winds around the head of the fibula in the leg below the lateral aspect of the knee, and its superficial location at this site makes it vulnerable to trauma. Patients present with weakness of foot dorsiflexion ("foot drop") as well as with weakness in eversion but not inversion at the ankle. Sparing of inversion, which is a function of muscles innervated by the tibial nerve, helps to distinguish peroneal neuropathies from L5 radiculopathies. Sensory loss involves the lateral aspect of the leg as well as the dorsum of the foot (Fig. e47-3C). Fractures of the fibular head may be responsible for peroneal neuropathies, but in the perioperative setting poorly applied braces exerting pressure on the nerve while the patient is unconscious are more often responsible. Tight-fitting stockings or casts of the upper leg can also cause a peroneal neuropathy, and thin individuals and those with recent weight loss are at increased risk.

Proximal Femoral Neuropathy

Lesions of the proximal femoral nerve are relatively uncommon but may present dramatically with weakness of hip flexion, quadriceps atrophy, weakness of knee extension (often manifesting with leg-buckling falls), and an absent patellar reflex. Adduction of the thigh is spared as these muscles are supplied by the obturator nerve, thereby distinguishing a femoral neuropathy from a more proximal lumbosacral plexus lesion. The sensory loss found is in the distribution of the femoral nerve sensory branches on the anterior part of the thigh (Fig. e47-3D). Compressive lesions from retroperitoneal hematomas or masses are common, and a CT of the pelvis should be obtained in all cases of femoral neuropathy to exclude these conditions. Bleeding into the pelvis resulting in hematoma can occur spontaneously, following trauma, or after intrapelvic surgeries such as renal transplantation. In intoxicated or

comatose patients, stretch injuries to the femoral nerve are seen following prolonged, extreme hip flexion or extension. Rarely, attempts at femoral vein or arterial puncture can be complicated by injury to this nerve.

Lateral Femoral Cutaneous Nerve

The symptoms of lateral femoral cutaneous nerve entrapment, commonly known as "meralgia paresthetica," include sensory loss, pain, and dysesthesia in part of the area supplied by the nerve (Fig. e47-3E). There is no motor component to the nerve, and therefore weakness is not a part of this syndrome. Symptoms often are worsened by standing or walking. Compression of the nerve occurs where it enters the leg near the inguinal ligament, usually in the setting of tight-fitting belts, pants, corsets, or recent weight gain, including that of pregnancy. The differential diagnosis of these symptoms includes hip problems such as trochanteric bursitis.

Obstetric Neuropathies

Pregnancy and delivery place women at special risk for a variety of nerve injuries. Radiculopathy due to a herniated lumbar disc is not common during pregnancy, but compressive injuries of the lumbosacral plexus do occur secondary to either the fetal head passing through the pelvis or the use of forceps during delivery. These plexus injuries are more frequent with cephalopelvic disproportion and often present with a painless unilateral foot drop which must be distinguished from a peroneal neuropathy caused by pressure on the nerve while in lithotomy position during delivery. Other compressive mononeuropathies of pregnancy include meralgia paresthetica, carpal tunnel syndrome, femoral neuropathy when the thigh is abducted severely in an effort to facilitate delivery of the fetal shoulder, and obturator neuropathy during lithotomy positioning. The latter presents with medial thigh pain that may be accompanied by weakness of thigh adduction. There is also a clear association between pregnancy and an increased frequency of idiopathic facial palsy (Bell's palsy).

FURTHER READINGS

Baryynski WS: Posterior reversible encephalopathy syndrome, part 1: Fundamental imaging and clinical features. AJNR Am J Neuroradiol 29:1036, 2008

Jillapalli D, Shefner JM: Electrodiagnosis in common mononeuropathies and plexopathies. Semin Neurol 25:196, 2005[PMID: 15937735] [Full Text]

Karnad DR, Guntupalli KK: Neurologic disorders in pregnancy. Crit Care Med 33:S362, 2005

King JD, Rosner MH: Osmotic demyelination syndrome. Am J Med Sci 339:561, 2010[PMID: 20453633] [Full Text]

Kumar S et al: Central pontine myelinolysis, an update. Neurol Res 28:360, 2006[PMID: 16687066] [Full Text]

Selnes OA et al: Cognition 6 years after surgical or medical therapy for coronary artery disease. Ann Neurol 63:581, 2008[PMID: 18481292] [Full Text]

Senzolo M et al: Neurologic complications after solid organ transplantation. Transpl Int 22:269, 2009[PMID: 19076332] [Full Text]

Van Dijk D et al: Cognitive and cardiac outcomes 5 years after off-pump vs. on-pump coronary artery bypass graft surgery. JAMA 297:701, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

CHRONIC FATIGUE SYNDROME: INTRODUCTION

Definition

Chronic fatigue syndrome (CFS) is a disorder characterized by persistent and unexplained fatigue resulting in severe impairment in daily functioning. Besides intense fatigue, most patients with CFS report concomitant symptoms such as pain, cognitive dysfunction, and unrefreshing sleep. Additional symptoms can include headache, sore throat, tender lymph nodes, muscle aches, joint aches, feverishness, difficulty sleeping, psychiatric problems, allergies, and abdominal cramps. Criteria for the diagnosis of CFS have been developed by the U.S. Centers for Disease Control and Prevention **(Table 389-1)**.

Table 389-1 Diagnostic Criteria for Chronic Fatigue Syndrome

Characterized by Persistent or Relapsing Unexplained Chronic Fatigue
Fatigue lasts for at least 6 months
Fatigue is of new or definite onset
Fatigue is not the result of an organic disease or of continuing exertion
Fatigue is not alleviated by rest
Fatigue results in a substantial reduction in previous occupational, educational, social, and personal activities
Four or more of the following symptoms, concurrently present for 6 months:
impaired memory or concentration, sore throat, tender cervical or axillary lymph nodes, muscle pain, pain in several joints, new headaches, unrefreshing sleep, or malaise after exertion
Exclusion Criteria
Medical condition explaining fatigue
Major depressive disorder (psychotic features) or bipolar disorder
Schizophrenia, dementia, or delusional disorder
Anorexia nervosa, bulimia nervosa
Alcohol or substance abuse
Severe obesity (BMI >40)

Epidemiology

CFS is seen worldwide, with adult prevalence rates varying between 0.2 and 0.4%. In the United States, the prevalence is higher in women, members of minority groups (African and Native Americans), and individuals with lower levels of education and occupational status. Approximately 75% of all CFS patients are women. The mean age of onset is between 29 and 35 years. It is probable that many patients go undiagnosed and/or do not seek help.

Etiology

There are numerous hypotheses about the etiology of CFS; there is no definitively indentified cause. Distinguishing between predisposing, precipitating, and perpetuating factors in CFS helps to provide a framework for understanding this complex condition **(Table 389-2)**.

Table 389-2 Predisposing, Precipitating, and Perpetuating Factors in Chronic Fatigue Syndrome

Time
Predisposing Factors
Childhood trauma (sexual, physical, emotional abuse; emotional and physical neglect)
Physical inactivity during childhood
Premorbid psychiatric illness or psychopathology
Premorbid hyperactivity
↓
Precipitating Factors
Somatic events: infection (mononucleosis, Q fever, Lyme disease), surgery, pregnancy
Psychosocial stress, life events
↓
Perpetuating Factors
(Non)acknowledgement by physician
Negative self-efficacy
Strong physical attributions
Strong focus on bodily symptoms
Fear of fatigue
(Lack of) social support
Low physical activity pattern

PREDISPOSING FACTORS

Physical inactivity and trauma in childhood tend to increase the risk of CFS in adults. Neuroendocrine dysfunction may be associated with childhood trauma, reflecting a biological correlate of vulnerability. Psychiatric illness and physical hyperactivity in adulthood raise the risk of CFS in later life. Twin studies suggest a familial predisposition to CFS, but no causative genes have been identified.

PRECIPITATING FACTORS

Physical or psychological stress may elicit the onset of CFS. Most patients report an infection (usually a flulike illness or infectious mononucleosis) as the trigger of their fatigue. Relatively high percentages of CFS follow Q fever and Lyme disease. However, no differences were found in Epstein-Barr virus load and immunologic reactivity in individuals who developed CFS and those who did not. While antecedent infections are associated with CFS, a direct microbial causality is unproven and unlikely. A recent study identified a murine leukemia virus-related retrovirus (XMRV); however, several subsequent studies have failed to confirm this result. Patients also often report other precipitating somatic events such as serious injury, surgery, pregnancy, or childbirth. Serious life events such as the loss of a loved one or a job, military combat, and other stressful situations may also precipitate CFS. A third of all patients cannot recall a trigger.

PERPETUATING FACTORS

Once CFS has developed, numerous factors may impede recovery. Physicians may contribute to chronicity by ordering unnecessary diagnostic procedures, by persistently suggesting psychological causes, and by not acknowledging CFS as a diagnosis.

A patient's focus on illness and avoidance of activities may perpetuate symptoms. A firm belief in a physical cause, a strong focus on bodily sensations, and a poor sense of control over symptoms may also prolong or exacerbate the fatigue and functional impairment. In most patients, inactivity is caused by negative illness perceptions rather than by poor physical fitness. Solicitous behavior of others may reinforce a patient's illness-related perceptions and behavior. A lack of social support is another known perpetuating factor.

Pathophysiology

The pathophysiology of CFS is unclear. Neuroimaging studies have reported that CFS is associated with reduced gray matter volume, associated with a decline in physical activity; these changes were partially reversed following cognitive behavioral therapy (CBT). In addition, functional MRI data have suggested that abnormal patterns of activation correlate with self-reported problems with information

processing. Neurophysiologic studies have shown altered CNS activation patterns during muscle contraction.

Evidence for immunologic dysfunction is inconsistent. Modest elevations in titers of antinuclear antibodies, reductions in immunoglobulin subclasses, deficiencies in mitogen-driven lymphocyte proliferation, reductions in natural killer cell activity, disturbances in cytokine production, and shifts in lymphocyte subsets have been described. None of these immune findings appear in most patients, nor do any correlate with the severity of CFS. In theory, symptoms of CFS could result from excessive production of a cytokine, such as interleukin 1, that induces asthenia and other flulike symptoms; however, compelling data in support of this hypothesis are lacking.

There is some evidence that CFS patients have mild hypocortisolism, the degree of which is associated with a poorer response to CBT.

Discrepancies in perceived and actual cognitive performance are a consistent finding in patients with CFS.

Diagnosis

In addition to a thorough history, a systematic physical examination is warranted to exclude disorders causing fatigue (e.g. endocrine disorders, neoplasms, heart failure, etc.). The heart rate of CFS patients is often slightly above normal. Laboratory tests primarily serve to exclude other diagnoses; there is no test that can diagnose CFS. The following laboratory screen usually suffices: complete blood count; ESR, CRP; serum creatinine, electrolytes, calcium and iron; blood glucose; creatine kinase; liver function tests; TSH; anti-gliadin antibodies; urinalysis. Serology for viral or bacterial infections is usually not helpful. No specific abnormalities have been identified on MRI or CT scans. CFS is a constellation of symptoms without any pathognomonic features, and remains a diagnosis of exclusion.

Bipolar disorders, schizophrenia, and substance abuse exclude a diagnosis of CFS, as do eating disorders, unless these have been resolved 5 years or longer before symptom onset. Also, CFS is excluded if the chronic fatigue developed immediately after a depressive episode. Depression developing in the course of the fatigue, however, does not preclude CFS. Co-occurring psychiatric disorder, especially anxiety and mood disorders, is seen in 30–60% of all cases.

Initial Management

In cases of suspected CFS, the clinician should acknowledge the impact of the patient's symptoms on daily functioning. Disbelief or denial can provoke an exacerbation of genuine symptoms, which in turn strengthens the clinician's disbelief, leading to an unfortunate cycle of miscommunication. The possibility of CFS should be considered if a patient fulfils all criteria (Table 389-1), and if other diagnoses have been excluded.

The patient should be asked to describe the symptoms (fatigue and accompanying symptoms) and their duration as well as the consequences (reduction in daily activities). To assess symptom severity and the extent of daily-life impairment, the patient should describe a typical day, from waking to retiring, and to contrast this with an average day prior to symptom onset. Next, potential fatigue-precipitating factors are sought. The severity of fatigue is difficult to assess quantitatively; a brief questionnaire is often helpful **(Fig. 389-1)**.

Figure 389-1

How have you felt during the last two weeks?

Please rate all four statements and per statement check the box that reflects your situation best.

1.	I feel tired	Yes, that is true							No, that is not true
2.	I tire easily	Yes, that is true							No, that is not true
3.	I feel fit	Yes, that is true							No, that is not true
4.	Physically I feel exhausted	Yes, that is true							No, that is not true

Scoring:

1, 2 and 4:	Yes, that is true	7	6	5	4	3	2	1	No, that is not true

3: Reversed

Sum scores >18 indicate severe fatigue

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Shortened fatigue questionnaire (SFQ).

The patient should be informed of the current understanding of precipitating and perpetuating factors and effective treatments, and be offered general advice about disease management. If CBT for CFS is not available as an initial treatment option (see below) and depression and anxiety are present, these symptoms should be treated. For patients with headache, diffuse pain, and feverishness, nonsteroidal anti-inflammatory drugs may be helpful. Even modest improvements in symptoms can make an important difference in the patient's degree of self-sufficiency and ability to appreciate life's pleasures.

Controlled therapeutic trials have established that acyclovir, fludrocortisone, galantamine, modafinil, and IV immunoglobulin, among other agents, offer no significant benefit in CFS. Countless anecdotes circulate regarding other traditional and nontraditional therapies. It is important to guide patients away from those therapeutic modalities that are toxic, expensive, or unreasonable.

The patient should be encouraged to maintain regular sleep patterns, to remain as active as possible, and to gradually return to previous exercise and activity (work) levels.

Treatment: Chronic Fatigue Syndrome

CBT and graded exercise therapy (GET) have been found to be the only beneficial interventions in CFS. Some patient groups argue against these approaches because of the implication that CFS is a purely mental disorder. CBT is a psychotherapeutic approach directed at changing condition-related cognitions and behaviors. CBT for CFS aims at changing a patient's perpetuating factors by exploiting various techniques and components. It includes educating the patient about the etiologic model, setting goals, restoring fixed bedtime and wake-up time, challenging and changing fatigue- and activity-related cognitions, reducing symptom focusing, spreading activities evenly throughout the day, gradually increasing physical activity, planning a return to work, and resuming other activities. The intervention, which typically consists of 12–14 sessions spread over 6 months, helps CFS patients gain control over their symptoms.

GET is based on the model of deconditioning and exercise intolerance and usually involves a home exercise program that continues for 3–5 months. Walking or cycling is systematically increased, with set target heart rates. Evidence that deconditioning is the basis for symptoms in CFS is lacking, however. The primary component of CBT and GET that results in a reduction in fatigue is the change in the patient's perception of fatigue and focus on symptoms.

CBT is generally the more complex treatment, which might explain why CBT studies tend to yield better improvement rates than GET trials.

Not all patients benefit from CBT or GET. Predictors of poor outcome are somatic comorbidity, current disability claims, and severe pain. CBT offered in an early stage of the illness reduces the burden of CFS for the patient as well as society in terms of decreased medical and disability-related costs.

Prognosis

Full recovery from untreated CFS is rare: the median annual recovery rate is 5% (range 0–31%) and the improvement rate 39% (range 8–63%). Patients with an underlying psychiatric disorder and those who continue to attribute their symptoms to an undiagnosed medical condition have poorer outcomes.

FURTHER READINGS

Baker R, Shaw EJ: Diagnosis and management of chronic fatigue syndrome or myalgic encephalitis (or encephalopathy): Summary of NICE guidance. BMJ 335:446, 2007[PMID: 17762037] [Full Text]

Presson AP et al: Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. BMC Syst Biol 2:95, 2008[PMID: 18986552] [Full Text]

Price JR et al: Cognitive behaviour therapy for chronic fatigue syndrome in adults. Cochrane Database Syst Rev (3):CD001027, 2008

Prins JB et al: Chronic fatigue syndrome. Lancet 367:346, 2006[PMID: 16443043] [Full Text]

Satterfield BC et al: Serologic and PCR testing of persons with chronic fatigue syndrome in the United States shows no association with xenotropic or polytropic murine leukemia virus-related viruses. Retrovirology 8:12, 2011[PMID: 21342521] [Full Text]

Tak LM et al: Meta-analysis and meta-regression of hypothalamic-pituitary-adrenal axis activity in functional somatic disorders. Biol Psychol 2011 [Epub ahead of print]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

BIOLOGY OF PSYCHIATRIC DISORDERS: INTRODUCTION

Psychiatric disorders are central nervous system diseases characterized by disturbances in emotion, cognition, motivation, and socialization. As a result of their high prevalence, early onset, and persistence, they contribute substantially to the burden of illness worldwide. Most psychiatric disorders are heterogeneous syndromes that currently lack well-defined neuropathology and *bona fide* biological markers. Therefore, diagnoses continue to be made solely from clinical observations using criteria in the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association (2000), 4th edition, text revision (DSM-IVTR). Recent advances in neuroimaging are beginning to provide evidence of brain pathology, which may one day be used for diagnosis and for following treatment. Family, twin, and adoption studies have shown that all common psychiatric syndromes are highly heritable, with genetic risk comprising 20–90% of disease vulnerability. The epidemiology, genetics, and biology of four common psychiatric disorders—autism, schizophrenia, mood disorders, and drug addiction—are presented below. A detailed discussion of the clinical manifestations and treatment of schizophrenia and mood disorders can be found in Chap. 391. Further discussion of alcoholism can be found in Chap. 392, opiate addiction in Chap. 393, cocaine and other drugs of abuse in Chap. 394, and nicotine addiction in Chap. 395.

AUTISM SPECTRUM DISORDERS

The DSM-IVTR criteria for Autism Spectrum Disorders (ASDs) require delays or abnormal functioning in social interactions, language as used in social communication, and symbolic or imaginative play, with onset prior to age 3. In addition to abnormal social behavior, ASDs are frequently, but not always, associated with reduced IQ and epilepsy. Individuals who exhibit some autism-like symptoms with relatively preserved cognitive functioning and language skills are described as having Asperger's syndrome.

Epidemiology

There has been a dramatic increase in the diagnosis of ASDs, from ~1/1000 (1950s–1990s) to a current level of ~1/150. Whether this increase reflects increased disease prevalence remains uncertain; ongoing studies are searching for genetic, environmental, and sociologic mechanisms that may have contributed to this change. In the 1950s–1960s, psychological factors were held to underlie autism. This conception was largely debunked by the 1970s, with the demonstration that prenatal rubella and phenylketonuria can cause ASDs, and with evidence for the genetic etiology of ASDs from twin studies. There is ongoing public concern that vaccines in general, or mercury-based preservatives in vaccines, can cause ASDs; however, large epidemiologic analyses have not supported this as an etiology. Whether environmental factors, such as perinatal infection and various toxins, for example, ethanol, illicit drugs, medications, and mutagenic agents, play a role is unclear.

Neuropathology and Neuroimaging

ASDs show no defining neuroanatomic phenotype that would indicate neurodevelopmental abnormalities. However, structural neuroimaging and histologic studies of postmortem brain provide evidence for anatomic defects. There is a modest increase in cerebrum growth (~10%; affecting both the white and grey matter) during early childhood (years 1–3), with the largest effect in the frontal lobes; the growth rate then decreases with age. Cerebellar size is increased by about 7% in children under age 5 years, but is decreased in older patients, and there are reduced (~30%) numbers of cerebellar Purkinje neurons. Finally, there is reduced cell size and increased cell density in the limbic areas of the brain.

Genetics

ASDs are highly heritable; concordance rates in monozygotic twins (~60–90%) are roughly tenfold higher than in dizygotic twins and siblings, and first-degree relatives show about fiftyfold increased risk for autism compared with prevalence in the general population. For unknown reasons, ASDs affect four times as many boys as girls. ASDs are also genetically heterogeneous. More than 20 known mutations, including copy number variations, account for about 10–20% of all cases, though none of these causes individually accounts for more than 1–2% **(Table 390-1)**. Many of the genes linked to ASDs can also cause other illnesses. For instance, mutations in MeCP2, FMR1, and TSC1&2 (see Table 390-1 for abbreviations) can cause mental retardation without ASDs, and alleles of certain genes, for example, neurexin 1, are associated with both ASDs and schizophrenia. It is likely that many

cases of ASDs result from more complex genetic mechanisms, including inheritance of multiple genetic variants or epigenetic modifications.

Table 390-1 Examples of Genes Implicated in Autism

Gene Symbol	Gene Name	Function
PTEN	Phosphatase and tensin homolog	Signal transduction Synaptic function
TSC1	Tuberous sclerosis 1	Signal transduction Translation and protein stability Synaptic function
TSC2	Tuberous sclerosis 2	Signal transduction Translation and protein stability Synaptic function
FMR1	Fragile X mental retardation 1	Translation and protein stability Synaptic function
UBE3A	Ubiquitin protein ligase E3A	Translation and protein stability Synaptic function
CNTN3	Contactin 3	Synaptic function
CNTN4	Contactin 4	Synaptic function
CNTNAP2	Contactin-associated protein-like 2	Synaptic function
NLGN3	Neuroligin 3	Synaptic function
NLGN4	Neuroligin 4	Synaptic function
NRXN1	Neurexin 1	Synaptic function
PCDH10	Protocadherin 10	Synaptic function
SHANK3	Shank 3	Synaptic function
SLC6A4	Serotonin transporter	Neurotransmitter signaling
AVPR1	Arginine vasopressin receptor 1	Neurotransmitter signaling
OXTR	Oxytosin receptor	Neurotransmitter signaling
CACNA1C	Voltage-gated calcium channel–alpha 1C subunit	Ion channel
CACNA1H	Voltage-gated calcium channel–alpha 1H subunit	Ion channel
SCN1A	Sodium channel, voltage-gated, type I, alpha subunit	Ion channel
SCN2A	Sodium channel, voltage-gated, type II, alpha subunit	Ion channel
SLC9A9	Sodium/hydrogen exchanger	Ion channel
DHCR7	7-Dehydrocholesterol reductase	Metabolism
PAH	Phenylalanine hydroxylase	Metabolism
ARX	Arx transcription factor	Gene expression
En2	Engrailed 2 transcription factor	Gene expression
MeCP2	Methyl CpG–binding protein 2 (Rett syndrome)	Gene expression
RNF8	Ring finger protein 8	Gene expression

Pathogenesis

Despite the genetic heterogeneity of ASDs, there are some common themes that may explain pathogenesis. These include mutations in proteins involved in the formation and function of synapses, control over the size and projections of neurons, production and signaling of neurotransmitters and neuromodulators, the function of ion channels, general cell metabolism, gene expression, and protein synthesis (see Table 390-1). Many of these mutations have a clear relationship to activity-dependent neural responses and can affect the development of neural systems that underlie cognition and social behaviors. They may be detrimental by altering the balance of excitatory vs. inhibitory synaptic signaling in local and extended circuits, and by altering the mechanisms that control brain growth. Another class of mutations affects genes (e.g., PTEN and Tsc) that negatively regulate signaling from several types of extracellular stimuli, including those transduced by receptor tyrosine kinases. Their dysregulation can have pleiotropic effects, including altering brain and neuronal growth as well as synaptic development and function. With further understanding of pathogenesis and the definition of

specific ASD subtypes, there is reason to believe that effective therapies will be identified, as in the case of dietary treatments for phenylketonuria. In addition, work in mouse models (e.g., with fragile X or Rett syndrome mutations) has suggested that autism-like behavioral abnormalities can be reversed even in fully developed adult animals by reversing the underlying pathology, which holds out hope for many affected individuals.

SCHIZOPHRENIA

Schizophrenia appears to be a heterogeneous collection of many distinct diseases, which remain poorly defined but linked by common clinical features. Three major symptom clusters are seen in schizophrenia: positive, negative, and cognitive symptoms. Positive symptoms include hallucinations and delusions, experiences that are not characteristic of normal mental life. Negative symptoms represent deficits in normal functions such as blunted affect, impoverished speech, asocial behavior, and diminished motivation. Cognitive symptoms include deficits in working memory and cognitive control of behavior that often prove extremely disabling. Current antipsychotic drugs are efficacious for positive symptoms only and generally lack efficacy for negative and cognitive symptoms.

Epidemiology

Schizophrenia is common, affecting males and females roughly equally, with a worldwide prevalence of approximately 1%. Environmental risks are thought to include prenatal exposure to viral infection (influenza), prenatal poor nutrition, perinatal hypoxia, psychotropic drug use (in particular, cannabis), and psychological stress. Advanced paternal age, birth order, and season of birth have also been implicated. However, none of these environmental influences has a specific or strong association with most cases of schizophrenia.

Neuropathology and Neuroimaging

The best-established neuropathologic finding in schizophrenia is enlargement of the lateral ventricles of the cerebral hemispheres. This is accompanied by a reduction in cortical thickness. These abnormalities are not specific to schizophrenia and are seen in many other conditions, including many neurodegenerative disorders. However, there is a general consensus that the reduction in cortical thickness in schizophrenia is associated with increased cell packing density and reduced neuropil (defined as axons, dendrites, and glial cell processes) without an overt change in neuronal cell number. Specific classes of interneurons in prefrontal cortex consistently show reduced expression of the gene encoding the enzyme glutamic acid decarboxylase 1 (*GAD1*), which synthesizes γ-aminobutyric acid (GABA), the principal inhibitory neurotransmitter in the brain. Functional imaging studies, by positron emission tomography (PET) or functional magnetic resonance imaging (MRI), show evidence of reduced metabolic or neural activity in the dorsolateral prefrontal cortex at rest and when performing psychological tests of executive function, including working memory. Alleles of two candidate risk genes [catechol-*O*-methyltransferase (COMT) and metabotropic glutamate receptor 3 (mGluR3)] are reported to affect dorsolateral prefrontal cortex activity, but these findings need to be replicated in larger samples. Similar pathologic and brain imaging abnormalities are seen in several other brain regions, in particular, the hippocampus. There are also numerous reports of abnormalities in myelin and oligodendrocytes in the cerebral cortex of patients with schizophrenia.

Genetics

Twins studies establish the heritability of schizophrenia, with co-inheritance at ~50% for monozygotic twins and ~10% for dizygotic twins. Genomewide linkage and association studies, and studies of copy number variation, have identified many regions and alleles that confer increased disease risk, particularly near genes on chromosome 22 [disrupted in schizophrenia 1 (DISC1), COMT, neuregulin 1, the neuregulin receptor ERBB4, and the DiGeorge (or velocardiofacial syndrome) region], and on chromosome 16p. The DiGeorge region deletions produce, in heterozygous form, a psychotic disorder with variable clinical features and a moderate to strong degree of penetrance. In contrast, the contribution of each of the individual genes to schizophrenia remains to be established with certainty. Moreover, the responsible genes within the DiGeorge region have not yet been identified. What is clear is that none of these other alleles produce schizophrenia with a high degree of penetrance. The current view in the field is that multiple rare alleles, many or most with limited penetrance, likely contribute to risk of schizophrenia. As for ASDs, the same allele may be a risk factor for multiple disorders. For instance, duplication of chromosome 16p is associated with both schizophrenia and autism, while DiGeorge region deletions and the *DISC1* locus on chromosome 22 are associated with schizophrenia, autism, and bipolar disorder.

Pathogenesis

There are several prevailing hypotheses about neurochemical mechanisms underlying schizophrenia. A reduction in the function of cortical and perhaps hippocampal GABAergic interneurons fits with reduced expression of glutamic acid decarboxylase. However, it is unknown whether this is a primary or compensatory feature of the disorder. Nevertheless, defects in parvalbumin-expressing GABAergic interneurons are known to reduce gamma-frequency activity on the EEG, which is a feature of many people with schizophrenia. Reduced excitatory neurotransmitter (glutamate) function is posited based

on psychotic and cognitive symptoms generated in humans exposed to ketamine or phencyclidine, which are noncompetitive antagonists of the NMDA subtype of glutamate receptors. There are reports of altered levels of glutamate receptors or associated proteins in the brains of individuals with schizophrenia examined postmortem, but no findings have yet been widely replicated. Finally, overactivity of dopamine neurotransmission at D_2-type dopamine receptors is proposed based on the ability of D_2 antagonists (an action common to all current antipsychotic agents; see Chap. 391) to ameliorate the positive symptoms of schizophrenia. Excessive dopamine release in the striatum elicited by an acute dose of amphetamine has been demonstrated by PET imaging in some patients with schizophrenia. However, it is unclear whether this abnormality reflects the underlying illness or a lasting effect of antipsychotic medications. In contrast, reduced activity of dopamine at D_1 dopamine receptors in the prefrontal cortex has been implicated in working memory deficits based on the cognitive effects of D_1 receptor agonists and antagonists in the illness. Nevertheless, inferring something about disease pathogenesis from the actions of psychotropic drugs, for example, as with the glutamate and dopamine hypotheses, is fraught with artifact.

Efforts to understand how defects in these neurotransmitter systems might generate similar behavioral phenotypes have led to intriguing hypotheses. For instance, in the hippocampus, reduced glutamate transmission (based on a hypothesized deficit in glutamate release or glutamate receptors) onto GABAergic interneurons could lead to reduced glutamic acid decarboxylase expression, reduced gamma oscillations, and reduced inhibition onto excitatory neurons. These events in turn could lead to increased dopamine release from the ventral tegmental area, with dopamine antagonists thereby helping to reset the system to its nonpathologic state. It must be emphasized that these are working models only, and a true pathophysiology (or pathophysiologies) for schizophrenia remains to be established.

Overlaid on these neurotransmitter-based hypotheses is speculation as to how mutations in any of several genes implicated, however tentatively, in schizophrenia lead to the associated pathologic and behavioral abnormalities. DISC1 was originally discovered based on its association with schizophrenia in an Icelandic family. However, as stated earlier, DISC1 has since been variably associated with other neuropsychiatric conditions and its role in schizophrenia remains uncertain. The DISC1 protein has been implicated in several cellular functions, including neuronal growth and maturation, neurite outgrowth, and even the proliferation of new neurons during development. Neuregulin 1 (NRG1), a member of the EGF family of growth factors, and its receptor ERBB4 have also been implicated in schizophrenia in several genetic studies. Interestingly, NRG1 and ERBB4 play important roles in the maturation of GABAergic interneurons in cerebral cortex, and regulate dopamine transmission to several limbic brain regions. Moreover, loss of NRG1-ERBB4 in mice leads to reduced neuropil, thus phenocopying a pathologic finding in schizophrenia. Another gene of potential interest encodes Reelin, a secreted extracellular matrix serine protease. There are unconfirmed reports of association of schizophrenia with the Reelin locus on chromosome 7, and of reduced Reelin expression in the cerebral cortex of schizophrenic subjects, possibly related to increased methylation of the Reelin gene promoter. Reelin is important during development in the migration of newly born neurons to their appropriate layers of cerebral cortex. In the adult brain, the protein is enriched in cortical GABAergic interneurons and has been implicated in regulating NMDA glutamate receptor function. It is, therefore, easy to imagine how abnormalities in DISC1, NRG1, or Reelin may be related to GABAergic, glutamatergic, and dopaminergic mechanisms in schizophrenia, and to associated pathologic abnormalities, but all such connections are currently speculative.

MOOD DISORDERS

Mood disorders are divided into depressive and bipolar disorders. Depressive disorders include the major depressive disorders, dysthymia, and more minor forms of depression. These disorders are heterogeneous syndromes, each composed of several diseases with presumably distinct pathophysiologies that remain to be elucidated.

Epidemiology

Mood disorders are common, with a prevalence of ~1–2% for bipolar disorder, ~5% for major depression, and ~15–20% for milder forms of depression. Between 40–50% of the risk for depression appears to be genetic. Nongenetic factors as diverse as stress and emotional trauma, viral infections, and even stochastic (random) processes during brain development have been implicated in the etiology. Depressive syndromes can occur in the context of general medical conditions such as endocrine disturbances (hyper- or hypocortisolemia, hyper- or hypothyroidism), autoimmune diseases, Parkinson's disease, traumatic brain injury, certain cancers, asthma, diabetes, and stroke. Depression and obesity/metabolic syndrome are important risk factors for each other. In predisposed individuals, stressful life events can lead to clear-cut depressive episodes, while severe stress can induce posttraumatic stress disorder (PTSD), instead of depression. Bipolar disorder is characterized by episodes of mania and depression and is one of the most heritable of psychiatric illnesses, with genetic risk of ~80%. Stress and disrupted circadian rhythms can promote the manic episodes, during which patients exhibit extremely elevated mood, abnormal thought patterns, and sometimes psychosis.

Several of these clinical signs can resemble certain features of schizophrenia; indeed, recent epidemiologic and genetic research has questioned the DSM-IVTR designations of bipolar disorder, schizophrenia, and schizoaffective disorder as distinct syndromes.

Neuropathology and Neuroimaging

Brain imaging studies in humans are defining the neural circuitry of mood within the brain's limbic system **(Fig. 390-1)**. Integral to this system are the nucleus accumbens (important for brain reward—see below under Substance Use Disorders), amygdala, hippocampus, and regions of prefrontal cortex. Given that many symptoms of depression (so-called neurovegetative symptoms) involve physiologic functions, a key role for the hypothalamus is also presumed. Depressed individuals show a small reduction in hippocampal size. PET and functional MRI have revealed increased activation of the amygdala by negative stimuli and reduced activation of the nucleus accumbens by rewarding stimuli. There is also evidence for altered activity in prefrontal cortex, for example, hyperactivity of subgenual area 25 in anterior cingulate cortex. Deep brain stimulation (DBS) of either the nucleus accumbens or subgenual area 25 elevates mood in normal and depressed individuals. While there are numerous reports of pathologic findings within these various regions postmortem, there is to date no defined neuropathology of depression.

Genetics

Although depression and bipolar disorder are highly heritable, the specific genes that comprise this risk remain unknown. As noted above, some of the genes implicated in autism or schizophrenia seem to cause bipolar disorder in some families. Large genomewide association studies have identified genes for diacylglycerol kinase η (*DGKH*), ankyrin G (*ANK3*), an L-type voltage-gated calcium channel (*CACNA1C*), and a gene-rich region on chromosome 16p12 as being associated with bipolar disorder, but these findings await confirmation by additional studies. Numerous susceptibility genes have also been implicated in linkage and association studies, but none has yet been definitively established as a bona fide depression gene. However, a few genes with variants that may modify depression risk are worthy of mention since they may be linked to mechanisms of pathogenesis (see below). These include genes for the type 1 receptor for corticotrophin-releasing factor (*CRHR1*); the glucocorticoid receptor gene (*GR*); *FKBP5*, which encodes a chaperone protein for the glucocorticoid receptor; the serotonin transporter gene (*SLA6A4*); the catechol-*O*-methyltransferase gene (*COMT*); and brain-derived neurotrophic factor (*BDNF*).

Pathogenesis

Human and animal research in depression has focused on the long-term effects of chronic stress on the brain and their reversal by antidepressant medications; prominent examples are discussed here. A subset of depressed patients show elevated levels of cortisol associated with increased production of corticotrophin-releasing factor from the hypothalamus and perhaps other brain regions (e.g., amygdala). In animals, sustained elevations in glucocorticoids impair hippocampal function, in part *via* direct damage to hippocampal neurons, which is consistent with reduced hippocampal volumes seen in some depressed humans. As the hippocampus exerts the major inhibitory influence over the hypothalamic-pituitary-adrenal axis, impairment of hippocampal function would lead to still further increases in glucocorticoid secretion, establishing a pathologic feed-forward loop.

Stress-induced damage to the hippocampus, and perhaps other limbic regions (e.g., amygdala), in animals is also mediated in part by reduced levels of BDNF and other growth factors and cytokines. Furthermore, stress leads to a decrease in the birth of new neurons in the adult hippocampus. Interestingly, antidepressant treatments reverse these effects of stress, and the antidepressant effects of these medications seem to depend, in part, on their ability to promote hippocampal neurogenesis in animal models of depression. The clinical ramifications of such observations are unproven, although similar regulation of adult hippocampal neurogenesis may be important for certain forms of learning and memory.

Another important target of stress in animals is the nucleus accumbens, where stress regulation of numerous signaling events (dopaminergic transmission and BDNF signaling are two examples) exert potent effects on depression-like behavioral abnormalities. While a reduction in BDNF in the hippocampus promotes depression-like behaviors, an induction of BDNF in the nucleus accumbens promotes depression; similar changes in BDNF expression have been observed in postmortem brains of depressed patients. Thus the role of BDNF in regulating mood is highly brain-region specific.

In contrast to depression, animal models of mania as well as bipolar disorder have proved much more elusive. Mice with loss-of-function mutations in the *Clock* or *GluR6* glutamate receptor genes or transgenic mice that overexpress glycogen synthase kinase 3β (GSK3β) show manic-like behavioral abnormalities, although the relevance of these observations to human mania remains unknown.

The observation that tricyclic antidepressants (e.g., imipramine) inhibit serotonin and/or norepinephrine reuptake, and that monoamine oxidase inhibitors (e.g., tranylcypromine) are effective antidepressants, initially led to the view that depression is caused by a deficiency of these monoamines. However, this hypothesis has not been well substantiated, although variants in the serotonin transporter, and in the

COMT gene, have been associated with altered mood states in some individuals. Nevertheless, these medications, particularly the tricyclics, have formed the basis of antidepressant discovery efforts, with virtually all of today's marketed antidepressants being SSRIs (e.g., fluoxetine, sertraline, citalopram), serotonin, and norepinephrine reuptake inhibitors (SNRIs) (e.g., venlafaxine, duloxetine), or norepinephrine reuptake inhibitors (NRIs) (e.g., atomoxetine).

A cardinal feature of all antidepressant medications is that long-term administration is needed for their mood-elevating effects. This means that their short-term actions, namely promotion of serotonin or norepinephrine function, is not per se antidepressant but rather induces a cascade of adaptations in the brain that underlie their clinical effects. The nature of these therapeutic drug-induced adaptations has not been identified with certainty. Presumably, the rich innervation of the brain's limbic circuitry by serotonin and norepinephrine (Fig. 390-1) provide the anatomic basis of their therapeutic actions.

Lithium is a highly effective drug for bipolar disorder, and competes with magnesium to inhibit magnesium-dependent enzymes, including GSK3β and several enzymes involved in phosphoinositide signaling leading to activation of protein kinase C. These findings have led to discovery programs focused on developing GSK3 and PKC inhibitors as potential novel treatments for mood disorders. Another commonly prescribed drug for bipolar disorder is valproic acid, which has pleiotropic effects, including inhibition of histone deacetylases (HDACs). Histone acetylation promotes transcriptional activation through posttranslational modification of N-terminal lysine residues in histones and thereby causes chromatin decondensation. HDAC inhibitors have shown some antidepressant effects in animal models of depression. Another form of epigenetic control of gene expression is methylation of cytosine residues in DNA, which inhibits gene transcription. DNA methylation has been shown to be important for inherited maternal effects on emotional behavior. Thus, rats born to mothers that exhibit low levels of nurturing behavior show increased anxiety and reduced expression of hippocampal glucocorticoid receptors due to increased methylation of the receptor gene. They pass these traits on to their offspring, but cross-fostering by mothers that display high levels of nurturing reverses them. As research into epigenetic mechanisms progresses, there is hope that it may become possible to identify specific depression-associated alterations in human chromatin.

SUBSTANCE USE DISORDERS

The DSM-IVTR uses the terms *substance dependence* and *substance abuse* to describe substance use disorders. It is unfortunate that the term *substance dependence* instead of addiction is used, because dependence can develop without addiction, and addiction involves much more than dependence per se. *Physical dependence* develops through resetting of homeostatic cellular mechanisms to permit normal function despite the continued presence of a drug; when drug intake is terminated abruptly, a withdrawal syndrome emerges. Withdrawal from alcohol or other sedative-hypnotics causes nervous system hyperactivity, whereas withdrawal from psychostimulants produces fatigue and sedation. *Tolerance* is a reduction in response to a drug, which like dependence, develops after repeated use. It results from a change in drug metabolism (pharmacokinetic tolerance) or cell signaling (pharmacodynamic tolerance). It is important to recognize that many nonaddictive medications induce tolerance and physical dependence, including β-adrenergic antagonists (e.g., propranolol) and α_2-adrenergic agonists (e.g., clonidine).

What sets drugs of abuse apart is their unique ability to produce *euphoria*, a positive emotional state characterized by intensely pleasant feelings that are rewarding and *reinforcing* since they motivate users to take the drug repeatedly. Tolerance develops to the rewarding properties of most abused drugs during periods of heavy use, which promotes the use of higher drug doses. In addition, *psychological (*or *motivational) dependence* develops through the resetting of cellular mechanisms within reward-related regions of the brain and leads to negative emotional symptoms resembling depression during drug withdrawal. Addictive drugs can also cause *sensitization*, an increased drug effect upon repeated use, as exemplified by the paranoid psychosis induced by chronic use of cocaine or other psychostimulants (e.g., amphetamine). Addiction, therefore, results from drug-induced changes in reward-related regions of the brain that lead to a complex mixture of tolerance, sensitization, and motivational dependence, in addition to powerful conditioning effects of these drugs mediated by the brain's memory circuits.

Epidemiology

Substance use disorders, especially those involving alcohol and tobacco, are very prevalent. The World Health Organization (WHO) estimates that more than 76 million people worldwide have alcohol use disorders and ~1.3 billion people smoke tobacco products (~1 billion men, 250 million women). The most widely used illicit drug in the United States is marijuana, with ~17% of 18–25-year-olds reporting regular use. Estimates of the annual economic burden of substance use disorders in the United States, including health- and crime-related costs and losses in productivity, exceed $500 billion.

Neuropathology and Neuroimaging

Imaging studies in humans demonstrate that addictive drugs, as well as craving for them, activate the brain's reward circuitry (see below). However, there is no established pathology associated with

addiction risk. Patients who abuse alcohol or psychostimulants show reduced gray matter in the prefrontal cortex. Functional MRI or PET studies show reduced activity in anterior cingulate and orbitofrontal cortex during tasks of attention and inhibitory control. Damage to these cortical areas may contribute to addiction by impairing decision making and increasing impulsivity.

Genetics

Substance use disorders are highly heritable, with genetic risk estimated to be 0.4 to 0.7; however, the specific genes that comprise this risk remain largely unknown. The best-established genetic contribution to addiction is the protective effect that mutations in alcohol-metabolizing enzymes have on risk for alcoholism. Mutations that increase alcohol dehydrogenase (ADH) activity and decrease aldehyde dehydrogenase (ALDH) activity are additive and promote accumulation of acetaldehyde following ingestion of alcohol. This produces intoxication at low doses and a flushing reaction that is unpleasant, resembling the reaction to disulfiram, a drug used to prevent relapse. These variants are common among people of East Asian descent, and individuals expressing these variants rarely abuse alcohol.

Genes that promote risk for addiction have begun to emerge from large family and population studies, but all genes identified to date represent only a very small fraction of the overall genetic risk for addiction. The best established susceptibility loci are regions on chromosomes 4 and 5 containing $GABA_A$ receptor gene clusters linked to alcohol use disorders and the nicotinic acetylcholine receptor gene cluster on chromosome 15 associated with nicotine and alcohol dependence. There are reports of numerous other addiction susceptibility genes (e.g., variants in COMT, the μ-opioid receptor, and the serotonin transporter), but further work is needed to validate these findings. In addition, several genes have been implicated in impulsivity, which is strongly associated with substance abuse. These include variants in genes for the D_4 dopamine receptor, the dopamine transporter, monoamine oxidase A, COMT, and the 5-HT1B serotonin receptor.

Pathogenesis

Work in rodents and nonhuman primates has established the brain's reward regions as key neural substrates for the acute actions of drugs of abuse and for addiction induced by repeated drug administration (Fig. 390-1). Midbrain dopamine neurons in the ventral tegmental area (VTA) function normally as rheostats of reward: They are activated by natural rewards (food, sex, social interaction) or even by the expectation of such rewards, and many are suppressed by the absence of an expected reward or by aversive stimuli. These neurons thereby transmit crucial survival signals to the rest of the limbic brain to promote reward-related behavior, including motor responses to seek and obtain the rewards (nucleus accumbens), memories of reward-related cues (amygdala, hippocampus), and executive control of obtaining rewards (prefrontal cortex).

Figure 390-1

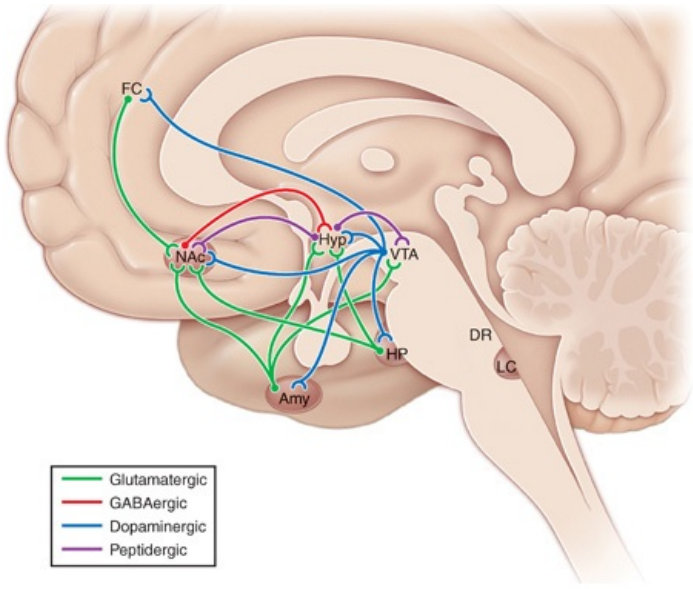

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Neural circuitry of depression and addiction. The figure shows a simplified summary of a series of limbic circuits in brain that regulate mood and motivation and are implicated in depression and addiction. Shown in the figure are the hippocampus (HP) and amygdala (Amy), regions of prefrontal cortex, nucleus accumbens (NAc), and hypothalamus (Hyp). Only a subset of the known interconnections among these brain regions is shown. Also shown is the innervation of several of

these brain regions by monoaminergic neurons. The ventral tegmental area (VTA) provides dopaminergic input to each of the limbic structures. Norepinephrine (from the locus coeruleus or LC) and serotonin [from the dorsal raphe (DR) and other raphe nuclei] innervate all of the regions shown. In addition, there are strong connections between the hypothalamus and the VTA-NAc pathway. Important peptidergic projections from the hypothalamus include those from the arcuate nucleus that release β-endorphin and melanocortin and from the lateral hypothalamus that release orexin.

Drugs of abuse alter neurotransmission through initial actions at different classes of ion channels, neurotransmitter receptors, or neurotransmitter transporters **(Table 390-2)**. Although the initial targets differ, the actions of these drugs converge on the brain's reward circuitry by promoting dopamine neurotransmission in the nucleus accumbens and other limbic targets of the VTA. In addition, some drugs promote activation of opioid and cannabinoid receptors, which modulate this reward circuitry. By these mechanisms, drugs of abuse produce powerful rewarding signals, which, after repeated drug administration, corrupt the brain's reward circuitry in ways that promote addiction. Three major pathologic adaptations have been described. First, drugs produce tolerance and dependence in reward circuits, which promote escalating drug intake and a negative emotional state during drug withdrawal that promotes relapse. Second, sensitization to the rewarding effects of the drugs and associated cues is seen during prolonged abstinence and also triggers relapse. Third, executive function is impaired in such a way as to increase impulsivity and compulsivity, both of which promote relapse.

Table 390-2 Initial Actions of Drugs of Abuse

Drug	Neurotransmitter Affected	Drug Target (Action)
Opiates	Endorphins, enkephalins	μ- and δ-opioid receptors (agonist)
Psychostimulants (cocaine, amphetamine methamphetamine)	Dopamine	Dopamine transporter (antagonist—cocaine; reverse transport—amphetamine, methamphetamine)
Nicotine	Acetylcholine	Nicotinic cholinergic receptors (agonist)
Ethanol	GABA	$GABA_A$ receptors (positive allosteric modulator)
	Glutamate	NMDA glutamate receptors (antagonist)
	Acetylcholine	Nicotinic cholinergic receptors (allosteric modulator)
	Serotonin	5HT-3 receptor (positive allosteric modulator)
	—	Calcium-activated K^+ channel (acivator)
Marijuana	Endocannabinoids (anandamide, 2-arachidonoylglycerol)	CB_1 receptor (agonist)
Phencyclidine	Glutamate	NMDA glutamate receptor (antagonist)

Repeated intake of abused drugs induces specific changes in cellular signal transduction, synaptic strength (long-term potentiation or depression), and neuronal structure (altered dendritic branching or cell soma size) within the brain's reward circuitry. These modifications are mediated in part by changes in gene expression, achieved by drug regulation of transcription factors [e.g., CREB (cAMP response element binding protein) and ΔFosB] and their target genes. Together, these drug-induced adaptations underlie alterations in numerous neurotransmitter systems (e.g., glutamate, GABA, dopamine), growth factors (e.g., BDNF), neuropeptides (e.g. corticotrophin releasing factor), and intracellular signaling cascades. These adaptations provide opportunities for developing treatments targeted to drug-addicted individuals. The fact that the spectrum of these adaptations partly differ depending on the particular addictive substance used creates opportunities for treatments that are specific for different classes of addictive drugs and that may, therefore, be less likely to disturb basic mechanisms that govern motivation and reward.

Increasingly, causal relationships are being established between individual molecular-cellular adaptations and specific behavioral abnormalities that characterize the addicted state. For example, acute activation of μ-opioid receptors by morphine or other opiates activates $G_{i/o}$ proteins leading to inhibition of adenylyl cyclase, resulting in reduced cAMP production, protein kinase A (PKA) activation, and activation of the transcription factor CREB. Repeated administration of these drugs **(Fig. 390-2)** evokes a homeostatic response resulting in upregulation of adenylyl cyclases, increased production of cAMP, and increased activation of PKA and CREB. Such upregulation of cAMP signaling has been identified in the locus coeruleus, periaqueductal gray, VTA, and nucleus accumbens and contributes to

opiate craving and signs of opiate withdrawal. The fact that endogenous opioid peptides do not produce tolerance and dependence while morphine and heroin do may relate to the recent observation that, unlike endogenous opioids, morphine and heroin are weak inducers of μ-opioid receptor desensitization and endocytosis. Therefore, these drugs cause prolonged receptor activation and inhibition of adenylyl cyclases, which provides a powerful stimulus for the upregulation of cAMP signaling that characterizes the opiate-dependent state.

Figure 390-2

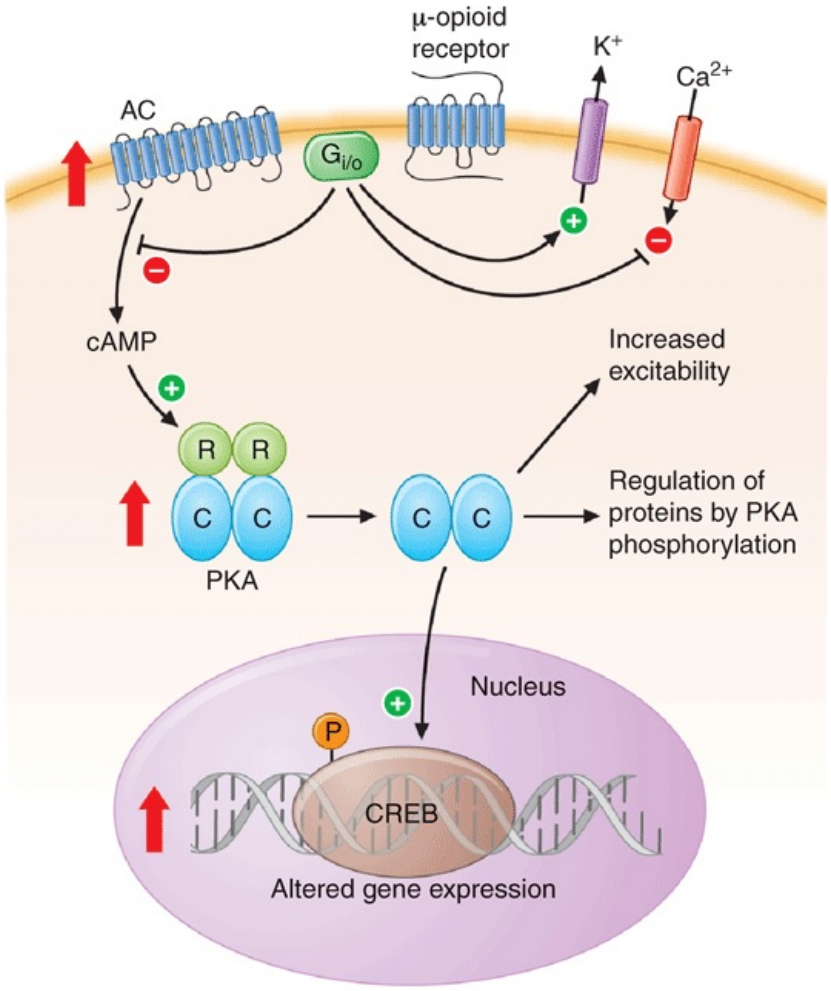

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Opiate action in the locus coeruleus (LC). Binding of opiate agonists to μ-opioid receptors catalyzes nucleotide exchange on G_i and G_o proteins, leading to inhibition of adenylyl cyclase, neuronal hyperpolarization via activation of K^+ channels, and inhibition of neurotransmitter release *via* inhibition of $Ca^{2/+}$ channels. Activation of $G_{i/o}$ also inhibits adenylyl cyclase (AC), reducing protein kinase A (PKA) activity and phosphorylation of several PKA substrate proteins, thereby altering their function. For example, opiates reduce phosphorylation of the cAMP response element-binding protein (CREB), which appears to initiate long-term changes in neuronal function. Chronic administration of opiates increases levels of AC isoforms, PKA catalytic (C) and regulatory (R) subunits, and the phosphorylation of several proteins, including CREB (indicated by red arrows). These changes contribute to the altered phenotype of the drug-addicted state. For example, the excitability of LC neurons is increased by enhanced cAMP signaling, although the ionic basis of this effect remains unknown. Activation of CREB causes upregulation of AC isoforms and tyrosine hydroxylase, the rate-limiting enzyme in catecholamine biosynthesis.

FURTHER READINGS

Abrahams BS, Geschwind DH: Advances in autism genetics: On the threshold of a new neurobiology. Nat Rev Genet 9:341, 2008[PMID: 18414403] [Full Text]

Jaaro-Peled H et al: Review of pathological hallmarks of schizophrenia: Comparison of genetic models with patients and nongenetic models. Schizophr Bull 36:301, 2010[PMID: 19903746] [Full Text]

Koob GF, Le Moal M: Neurobiological mechanisms for opponent motivational processes in addiction. Philos Trans R Soc London B Biol Sci 363:3113, 2008[PMID: 18653439] [Full Text]

———, Volkow ND: Neurocircuitry of addiction. Neuropsychopharmacology 35:217, 2010

Krishnan V, Nestler EJ: The molecular neurobiology of depression. Nature 455:894, 2008[PMID: 18923511] [Full Text]

Maia TV, Frank MJ: From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 14:154, 2011[PMID: 21270784] [Full Text]

Martinowich K et al: Bipolar disorder: From genes to behavior pathways. J Clin Invest 119:726, 2009[PMID: 19339764] [Full Text]

Ressler KJ, Mayberg HS: Targeting abnormal neural circuits in mood and anxiety disorders: From the laboratory to the clinic. Nat Neurosci 10:1116, 2007[PMID: 17726478] [Full Text]

Spanagel R: Alcoholism: A systems approach from molecular physiology to addictive behavior. Physiol Rev 89:649, 2009[PMID: 19342616] [Full Text]

Walsh CA et al: Autism and brain development. Cell 135:396, 2008[PMID: 18984148] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

MENTAL DISORDERS: INTRODUCTION

Mental disorders are common in medical practice and may present either as a primary disorder or as a comorbid condition. The prevalence of mental or substance use disorders in the United States is approximately 30%, only one-third of whom are currently receiving treatment. Global burden of disease statistics indicate that 4 of the 10 most important causes of disease worldwide are psychiatric in origin.

The revised fourth edition for use by primary care physicians of the *Diagnostic and Statistical Manual* (DSM-IV-PC) provides a useful synopsis of mental disorders most likely to be seen in primary care practice. The current system of classification is multiaxial and includes the presence or absence of a major mental disorder (axis I), any underlying personality disorder (axis II), general medical condition (axis III), psychosocial and environmental problems (axis IV), and overall rating of general psychosocial functioning (axis V).

Changes in health care delivery underscore the need for primary care physicians to assume responsibility for the initial diagnosis and treatment of the most common mental disorders. Prompt diagnosis is essential to ensure that patients have access to appropriate medical services and to maximize the clinical outcome. Validated patient-based questionnaires have been developed that systematically probe for signs and symptoms associated with the most prevalent psychiatric diagnoses and guide the clinician into targeted assessment. Prime MD (and a self-report form, the PHQ) and the Symptom-Driven Diagnostic System for Primary Care (SDDS-PC) are inventories that require only 10 minutes to complete and link patient responses to the formal diagnostic criteria of anxiety, mood, somatoform, and eating disorders and to alcohol abuse or dependence.

A physician who refers patients to a psychiatrist should know not only when doing so is appropriate but also how to refer, since societal misconceptions and the stigma of mental illness impede the process. Primary care physicians should base referrals to a psychiatrist on the presence of signs and symptoms of a mental disorder and not simply on the absence of a physical explanation for a patient's complaint. The physician should discuss with the patient the reasons for requesting the referral or consultation and provide reassurance that he or she will continue to provide medical care and work collaboratively with the mental health professional. Consultation with a psychiatrist or transfer of care is appropriate when physicians encounter evidence of psychotic symptoms, mania, severe depression, or anxiety; symptoms of posttraumatic stress disorder (PTSD); suicidal or homicidal preoccupation; or a failure to respond to first-order treatment. **Eating disorders are discussed in Chap. 79, and the biology of psychiatric and addictive disorders in Chap. 390**.

ANXIETY DISORDERS

Anxiety disorders, the most prevalent psychiatric illnesses in the general community, are present in 15–20% of medical clinic patients. Anxiety, defined as a subjective sense of unease, dread, or foreboding, can indicate a primary psychiatric condition or can be a component of, or reaction to, a primary medical disease. The primary anxiety disorders are classified according to their duration and course and the existence and nature of precipitants.

When evaluating the anxious patient, the clinician must first determine whether the anxiety antedates or postdates a medical illness or is due to a medication side effect. Approximately one-third of patients presenting with anxiety have a medical etiology for their psychiatric symptoms, but an anxiety disorder can also present with somatic symptoms in the absence of a diagnosable medical condition.

Panic Disorder

CLINICAL MANIFESTATIONS

Panic disorder is defined by the presence of recurrent and unpredictable panic attacks, which are distinct episodes of intense fear and discomfort associated with a variety of physical symptoms, including palpitations, sweating, trembling, shortness of breath, chest pain, dizziness, and a fear of impending doom or death**(Table 391-1)**. Paresthesias, gastrointestinal distress, and feelings of unreality are also common. Diagnostic criteria require at least 1 month of concern or worry about the attacks or a change in behavior related to them. The lifetime prevalence of panic disorder is 1–3%. Panic attacks have a sudden onset, developing within 10 minures and usually resolving over the course of an hour, and they occur in an unexpected fashion. The frequency and severity of panic attacks vary, ranging from once a week to clusters of attacks separated by months of well-being. The first attack is usually outside the home, and onset is typically in late adolescence to early adulthood. In

some individuals, anticipatory anxiety develops over time and results in a generalized fear and a progressive avoidance of places or situations in which a panic attack might recur. *Agoraphobia*, which occurs commonly in patients with panic disorder, is an acquired irrational fear of being in places where one might feel trapped or unable to escape **(Table 391-2)**. Typically, it leads the patient into a progressive restriction in lifestyle and, in a literal sense, in geography. Frequently, patients are embarrassed that they are housebound and dependent on the company of others to go out into the world and do not volunteer this information; thus physicians will fail to recognize the syndrome if direct questioning is not pursued.

Table 391-1 Diagnostic Criteria for Panic Attack
A discrete period of intense fear or discomfort, in which four or more of the following symptoms developed abruptly and reached a peak within 10 minutes:
1. Palpitations, pounding heart, or accelerated heart rate
2. Sweating
3. Trembling or shaking
4. Sensations of shortness of breath or smothering
5. Feeling of choking
6. Chest pain or discomfort
7. Nausea or abdominal distress
8. Feeling dizzy, unsteady, lightheaded, or faint
9. Derealization (feelings of unreality) or depersonalization (being detached from oneself)
10. Fear of losing control or going crazy
11. Fear of dying
12. Paresthesias (numbness or tingling sensations)
13. Chills or hot flushes

Source: *Diagnostic and Statistical Manual of Mental Disorders* , 4th ed. Washington, DC, American Psychiatric Association, 2000.

Table 391-2 Diagnostic Criteria for Agoraphobia
1. Anxiety about being in places or situations from which escape might be difficult (or embarrassing) or in which help may not be available in the event of having an unexpected or situationally predisposed panic attack or panic-like symptoms. Agoraphobic fears typically involve characteristic clusters of situations that include being outside the home alone; being in a crowd or standing in a line; being on a bridge; and traveling in a bus, train, or automobile.
2. The situations are avoided (e.g., travel is restricted) or else are endured with marked distress or with anxiety about having a panic attack or panic-like symptoms, or require the presence of a companion.
3. The anxiety or phobic avoidance is not better accounted for by another mental disorder such as social phobia (e.g., avoidance limited to social situations because of fear of embarrassment), specific phobia (e.g., avoidance limited to a single situation like elevators), obsessive-compulsive disorder (e.g., avoidance of dirt in someone with an obsession about contamination), posttraumatic stress disorder (e.g., avoidance of stimuli associated with a severe stressor), or separation anxiety disorder (e.g., avoidance of leaving home or relatives).

Source: *Diagnostic and Statistical Manual of Mental Disorders* , 4th ed. Washington, DC, American Psychiatric Association, 2000.

DIFFERENTIAL DIAGNOSIS

A diagnosis of panic disorder is made after a medical etiology for the panic attacks has been ruled out. A variety of cardiovascular, respiratory, endocrine, and neurologic conditions can present with anxiety as the chief complaint. Patients with true panic disorder will often focus on one specific feature to the exclusion of others. For example, 20% of patients who present with syncope as a primary medical complaint have a primary diagnosis of a mood, anxiety, or substance-abuse disorder, the most common being panic disorder. The differential diagnosis of panic disorder is complicated by a high rate of comorbidity with other psychiatric conditions, especially alcohol and benzodiazepine abuse, which patients initially use in an attempt at self-medication. Some 75% of panic disorder patients will also satisfy criteria for major depression at some point in their illness.

When the history is nonspecific, physical examination and focused laboratory testing must be used to rule out anxiety states resulting from medical disorders such as pheochromocytoma, thyrotoxicosis, or hypoglycemia. Electrocardiogram (ECG) and echocardiogram may detect some cardiovascular

conditions associated with panic such as paroxysmal atrial tachycardia and mitral valve prolapse. In two studies, panic disorder was the primary diagnosis in 43% of patients with chest pain who had normal coronary angiograms and was present in 9% of all outpatients referred for cardiac evaluation. Panic disorder has also been diagnosed in many patients referred for pulmonary function testing or with symptoms of irritable bowel syndrome.

ETIOLOGY AND PATHOPHYSIOLOGY

The etiology of panic disorder is unknown but appears to involve a genetic predisposition, altered autonomic responsivity, and social learning. Panic disorder shows familial aggregation; the disorder is concordant in 30–45% of monozygotic twins, and genomewide screens have identified suggestive risk loci. Acute panic attacks appear to be associated with increased noradrenergic discharges in the locus coeruleus. Intravenous infusion of sodium lactate evokes an attack in two-thirds of panic disorder patients, as do the α_2-adrenergic antagonist yohimbine, cholecystokinin tetrapeptide (CCK-4), and carbon dioxide inhalation. It is hypothesized that each of these stimuli activates a pathway involving noradrenergic neurons in the locus coeruleus and serotonergic neurons in the dorsal raphe. Agents that block serotonin reuptake can prevent attacks. Panic-disorder patients have a heightened sensitivity to somatic symptoms, which triggers increasing arousal, setting off the panic attack; accordingly, therapeutic intervention involves altering the patient's cognitive interpretation of anxiety-producing experiences as well as preventing the attack itself.

TREATMENT: PANIC DISORDER

Achievable goals of treatment are to decrease the frequency of panic attacks and to reduce their intensity. The cornerstone of drug therapy is antidepressant medication **(Tables 391-3through391-5)**. Selective serotonin reuptake inhibitors (SSRIs) benefit the majority of panic disorder patients and do not have the adverse effects of tricyclic antidepressants (TCAs). Fluoxetine, paroxetine, sertraline, and the selective serotonin-norepinephrine reuptake inhibitor (SNRI) venlafaxine have received approval from the U.S. Food and Drug Administration (FDA) for this indication. These drugs should be started at one-third to one-half of their usual antidepressant dose (e.g., 5–10 mg fluoxetine, 25–50 mg sertraline, 10 mg paroxetine, venlafaxine 37.5 mg). Monoamine oxidase inhibitors (MAOIs) are also effective and may specifically benefit patients who have comorbid features of atypical depression (i.e., hypersomnia and weight gain). Insomnia, orthostatic hypotension, and the need to maintain a low-tyramine diet (avoidance of cheese and wine) have limited their use, however. Antidepressants typically take 2–6 weeks to become effective, and doses may need to be adjusted based upon the clinical response.

Table 391-3 Antidepressants

Name	Usual Daily Dose, mg	Side Effects	Comments
SSRIs			
Fluoxetine (Prozac)	10–80	Headache; nausea and other GI effects; jitteriness; insomnia; sexual dysfunction; can affect plasma levels of other medicines (except sertraline); akathisia rare	Once-daily dosing, usually in the morning; fluoxetine has very long half-life; must not be combined with MAOIs
Sertraline (Zoloft)	50–200		
Paroxetine (Paxil)	20–60		
Fluvoxamine (Luvox)	100–300		
Citalopram (Celexa)	20–60		
Escitalopram (Lexapro)	10–30		
TCAs			
Amitriptyline (Elavil)	150–300	Anticholinergic (dry mouth, tachycardia, constipation, urinary retention, blurred vision); sweating; tremor; postural hypotension; cardiac conduction delay; sedation; weight gain	Once-daily dosing, usually qhs; blood levels of most TCAs available; can be lethal in O.D. (lethal dose = 2 g); nortriptyline best tolerated, especially by elderly
Nortriptyline (Pamelor)	50–200		
Imipramine (Tofranil)	150–300		
Desipramine (Norpramin)	150–300		
Doxepin (Sinequan)	150–300		
Clomipramine	150		

(Anafranil)	–300		
Mixed Norepinephrine/Serotonin Reuptake Inhibitors and Receptor Blockers			
Venlafaxine (Effexor)	75 –375	Nausea; dizziness; dry mouth; headaches; increased blood pressure; anxiety and insomnia	Bid-tid dosing (extended release available); lower potential for drug interactions than SSRIs; contraindicated with MAOIs
Desvenlafaxine, (Pristiq)	50 –400	Nausea, dizziness, insomnia	Primary metabolite of venlafaxine. No increased efficacy with higher dosing
Duloxetine (Cymbalta)	40–60	Nausea, dizziness, headache, insomnia, constipation	May have utility in treatment of neuropathic pain and stress incontinence
Mirtazapine (Remeron)	15–45	Somnolence; weight gain; neutropenia rare	Once daily dosing
Mixed-action Drugs			
Bupropion (Wellbutrin)	250 –450	Jitteriness; flushing; seizures in at-risk patients; anorexia; tachycardia; psychosis	Tid dosing, but sustained release also available; fewer sexual side effects than SSRIs or TCAs; may be useful for adult ADD
Trazodone (Desyrel)	200 –600	Sedation; dry mouth; ventricular irritability; postural hypotension; priapism rare	Useful in low doses for sleep because of sedating effects with no anticholinergic side effects
Nefazodone (Serzone)	300 –600	Sedation; headache; dry mouth; nausea; constipation	Discontinued sale in United States and several other countries due to risk of liver failure
Amoxapine (Asendin)	200 –600	Sexual dysfunction	Lethality in overdose; EPS possible
MAOIs			
Phenelzine (Nardil)	45–90	Insomnia; hypotension; anorgasmia; weight gain; hypertensive crisis; toxic reactions with SSRIs; narcotics	May be more effective in patients with atypical features or treatment-refractory depression
Tranylcypromine (Parnate)	20–50		
Isocarboxazid (Marplan)	20–60		
Transdermal selegiline (Emsam)	6–12	Local skin reaction hypertension	No dietary restrictions with 6 mg dose

Abbreviations: ADD, attention deficit disorder; EPS, extrapyramidal symptoms; MAOIs, monoamine oxidase inhibitors; SSRIs, selective serotonin reuptake inhibitors;TCAs, tricyclic antidepressants.

Table 391-4 Management of Antidepressant Side Effects

Symptoms	Comments and Management Strategies
Gastrointestinal	
Nausea, loss of appetite	Usually short-lived and dose-related; consider temporary dose reduction or administration with food and antacids
Diarrhea	Famotidine, 20–40 mg/d
Constipation	Wait for tolerance; try diet change, stool softener, exercise; avoid laxatives
Sexual dysfunction	Consider dose reduction; drug holiday
Anorgasmia/impotence; impaired ejaculation	Bethanechol, 10–20 mg, 2 h before activity, or cyproheptadine, 4–8 mg 2 h before activity, or bupropion, 100 mg bid or amantadine, 100 mg bid/tid
Orthostasis	Tolerance unlikely; increase fluid intake, use calf exercises/support hose; fludrocortisone, 0.025 mg/d
Anticholinergic	Wait for tolerance
Dry mouth, eyes	Maintain good oral hygiene; use artificial tears, sugar-free gum
Tremor/jitteriness	Antiparkinsonian drugs not effective; use dose reduction/slow increase; lorazepam, 0.5 mg bid, or propranolol, 10–20 mg bid
Insomnia	Schedule all doses for the morning; trazodone, 50–100 mg qhs
Sedation	Caffeine; schedule all dosing for bedtime; bupropion, 75–100 mg in afternoon
Headache	Evaluate diet, stress, other drugs; try dose reduction; amitriptyline, 50 mg/d

Weight gain	Decrease carbohydrates; exercise; consider fluoxetine
Loss of therapeutic benefit over time	Related to tolerance? Increase dose or drug holiday; add amantadine, 100 mg bid, buspirone, 10 mg tid, or pindolol, 2.5 mg bid

Table 391-5 Possible Drug Interactions with Selective Serotonin Reuptake Inhibitors

Agent	Effect
Monoamine oxidase inhibitors	Serotonin syndrome—absolute contraindication
Serotonergic agonists, e.g., tryptophan, fenfluramine	Potential serotonin syndrome
Drugs that are metabolized by P450 isoenzymes: tricyclics, other SSRIs, antipsychotics, beta blockers, codeine, triazolobenzodiazepines, calcium channel blockers	Delayed metabolism resulting in increased blood levels and potential toxicity
Drugs that are bound tightly to plasma proteins, e.g., warfarin	Increased bleeding secondary to displacement
Drugs that inhibit the metabolism of SSRIs by P450 isoenzymes, e.g., quinidine	Increased SSRI side effects

Abbreviation: SSRIs, selective serotonin reuptake inhibitors.

Because of anticipatory anxiety and the need for immediate relief of panic symptoms, benzodiazepines are useful early in the course of treatment and sporadically thereafter **(Table 391-6)**. For example, alprazolam, starting at 0.5 mg qid and increasing to 4 mg/d in divided doses, is effective, but patients must be monitored closely, as some develop dependence and begin to escalate the dose of this medication. Clonazepam, at a final maintenance dose of 2–4 mg/d, is also helpful; its longer half-life permits twice-daily dosing, and patients appear less likely to develop dependence on this agent.

Table 391-6 Anxiolytics

Name	Equivalent PO Dose, mg	Onset of Action	Half-life, h	Comments
Benzodiazepines				
Diazepam (Valium)	5	Fast	20–70	Active metabolites; quite sedating
Flurazepam (Dalmane)	15	Fast	30–100	Flurazepam is a prodrug; metabolites are active; quite sedating
Triazolam (Halcion)	0.25	Intermediate	1.5–5	No active metabolites; can induce confusion and delirium, especially in elderly
Lorazepam (Ativan)	1	Intermediate	10–20	No active metabolites; direct hepatic glucuronide conjugation; quite sedating
Alprazolam (Xanax)	0.5	Intermediate	12–15	Active metabolites; not too sedating; may have specific antidepressant and antipanic activity; tolerance and dependence develop easily
Chlordiazepoxide (Librium)	10	Intermediate	5–30	Active metabolites; moderately sedating
Oxazepam (Serax)	15	Slow	5–15	No active metabolites; direct glucuronide conjugation; not too sedating
Temazepam (Restoril)	15	Slow	9–12	No active metabolites; moderately sedating
Clonazepam (Klonopin)	0.5	Slow	18–50	No active metabolites; moderately sedating
Nonbenzodiazepines				
Buspirone (BuSpar)	7.5	2 weeks	2–3	Active metabolites; tid dosing—usual daily dose 10–20 mg tid; nonsedating; no additive effects with alcohol; useful for controlling agitation in demented or brain-injured patients

Early psychotherapeutic intervention and education aimed at symptom control enhances the effectiveness of drug treatment. Patients can be taught breathing techniques, be educated about physiologic changes that occur with panic, and learn to expose themselves voluntarily to precipitating events in a treatment program spanning 12–15 sessions. Homework assignments and monitored compliance are important components of successful treatment. Once patients have achieved a

satisfactory response, drug treatment should be maintained for 1–2 years to prevent relapse. Controlled trials indicate a success rate of 75–85%, although the likelihood of complete remission is somewhat lower.

Generalized Anxiety Disorder

CLINICAL MANIFESTATIONS

Patients with generalized anxiety disorder (GAD) have persistent, excessive, and/or unrealistic worry associated with muscle tension, impaired concentration, autonomic arousal, feeling "on edge" or restless, and insomnia **(Table 391-7)**. Onset is usually before age 20 years, and a history of childhood fears and social inhibition may be present. The lifetime prevalence of GAD is 5–6%; the risk is higher in first-degree relatives of patients with the diagnosis. Interestingly, family studies indicate that GAD and panic disorder segregate independently. More than 80% of patients with GAD also suffer from major depression, dysthymia, or social phobia. Comorbid substance abuse is common in these patients, particularly alcohol and/or sedative/hypnotic abuse. Patients with GAD worry excessively over minor matters, with life-disrupting effects; unlike in panic disorder, complaints of shortness of breath, palpitations, and tachycardia are relatively rare.

Table 391-7 Diagnostic Criteria for Generalized Anxiety Disorder

A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).
B. The person finds it difficult to control the worry.
C. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms present for more days than not for the past 6 months): (1) restlessness or feeling keyed up or on edge; (2) being easily fatigued; (3) difficulty concentrating or mind going blank; (4) irritability; (5) muscle tension; (6) sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep).
D. The focus of the anxiety and worry is not confined to features of an axis I disorder [e.g., the anxiety or worry is not about having a panic attack (as in panic disorder), being embarrassed in public (as in social phobia), being contaminated (as in obsessive-compulsive disorder), being away from home or close relatives (as in separation anxiety disorder), gaining weight (as in anorexia nervosa), having multiple physical complaints (as in somatization disorder), or having a serious illness (as in hypochondriasis)], and the anxiety and worry do not occur exclusively during posttraumatic stress disorder.
E. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
F. The disturbance is not due to the direct physiologic effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hyperthyroidism) and does not occur exclusively during a mood disorder, a psychotic disorder, or a pervasive developmental disorder.

Source: *Diagnostic and Statistical Manual of Mental Disorders* , 4th ed. Washington, DC, American Psychiatric Association, 2000.

ETIOLOGY AND PATHOPHYSIOLOGY

All anxiogenic agents act on the γ-aminobutyric acid $(GABA)_A$ receptor/chloride ion channel complex, implicating this neurotransmitter system in the pathogenesis of anxiety and panic attacks. Benzodiazepines are thought to bind two separate $GABA_A$ receptor sites: type I, which has a broad neuroanatomic distribution, and type II, which is concentrated in the hippocampus, striatum, and neocortex. The antianxiety effects of the various benzodiazepines and side effects such as sedation and memory impairment are influenced by their relative binding to type I and type II receptor sites. Serotonin [5-hydroxytryptamine (5HT)] and 3α-reduced neuractive steroids (allosteric modulators of $GABA_A$) also appear to have a role in anxiety, and buspirone, a partial $5HT_{1A}$ receptor agonist, and certain $5HT_{2A}$ and $5HT_{2C}$ receptor antagonists (e.g., nefazodone) may have beneficial effects.

TREATMENT: GENERALIZED ANXIETY DISORDER

A combination of pharmacologic and psychotherapeutic interventions is most effective in GAD, but complete symptomatic relief is rare. A short course of a benzodiazepine is usually indicated, preferably lorazepam, oxazepam, or temazepam. (The first two of these agents are metabolized via conjugation rather than oxidation and thus do not accumulate if hepatic function is impaired.) Treatment should be initiated at the lowest dose possible and prescribed on an as-needed basis as symptoms warrant. Benzodiazepines differ in their milligram per kilogram potency, half-life, lipid solubility, metabolic pathways, and presence of active metabolites. Agents that are absorbed rapidly and are lipid soluble, such as diazepam, have a rapid onset of action and a higher abuse potential. Benzodiazepines should generally not be prescribed for >4–6 weeks because of the development of tolerance and the risk of abuse and dependence. Withdrawal must be closely monitored as relapses can occur. It is important to warn patients that concomitant use of alcohol or other sedating drugs may be neurotoxic and impair

their ability to function. An optimistic approach that encourages the patient to clarify environmental precipitants, anticipate his or her reactions, and plan effective response strategies is an essential element of therapy.

Adverse effects of benzodiazepines generally parallel their relative half-lives. Longer-acting agents, such as diazepam, chlordiazepoxide, flurazepam, and clonazepam, tend to accumulate active metabolites, with resultant sedation, impairment of cognition, and poor psychomotor performance. Shorter-acting compounds, such as alprazolam and oxazepam, can produce daytime anxiety, early morning insomnia, and, with discontinuation, rebound anxiety and insomnia. Although patients develop tolerance to the sedative effects of benzodiazepines, they are less likely to habituate to the adverse psychomotor effects. Withdrawal from the longer half-life benzodiazepines can be accomplished through gradual, stepwise dose reduction (by 10% every 1–2 weeks) over 6–12 weeks. It is usually more difficult to taper patients off shorter-acting benzodiazepines. Physicians may need to switch the patient to a benzodiazepine with a longer half-life or use an adjunctive medication such as a beta blocker or carbamazepine, before attempting to discontinue the benzodiazepine. Withdrawal reactions vary in severity and duration; they can include depression, anxiety, lethargy, diaphoresis, autonomic arousal, and, rarely, seizures.

Buspirone is a nonbenzodiazepine anxiolytic agent. It is nonsedating, does not produce tolerance or dependence, does not interact with benzodiazepine receptors or alcohol, and has no abuse or disinhibition potential. However, it requires several weeks to take effect and requires thrice-daily dosing. Patients who were previously responsive to a benzodiazepine are unlikely to rate buspirone as equally effective, but patients with head injury or dementia who have symptoms of anxiety and/or agitation may do well with this agent. Escitalopram, paroxetine, and venlafaxine are FDA approved for the treatment of GAD, usually at doses that are comparable to their efficacy in major depression. Benzodiazepines are contraindicated during pregnancy and breast-feeding.

Anticonvulsants with GABAergic properties may also be effective against anxiety. Gabapentin, oxcarbazepine, tiagabine, pregabalin, and divalproex have all shown some degree of benefit in a variety of anxiety-related syndromes. Agents that selectively target $GABA_A$ receptor subtypes are currently under development, and it is hoped that these will lack the sedating, memory-impairing, and addicting properties of benzodiazepines.

Phobic Disorders

CLINICAL MANIFESTATIONS

The cardinal feature of phobic disorders is a marked and persistent fear of objects or situations, exposure to which results in an immediate anxiety reaction. The patient avoids the phobic stimulus, and this avoidance usually impairs occupational or social functioning. Panic attacks may be triggered by the phobic stimulus or may occur spontaneously. Unlike patients with other anxiety disorders, individuals with phobias usually experience anxiety only in specific situations. Common phobias include fear of closed spaces (claustrophobia), fear of blood, and fear of flying. Social phobia is distinguished by a specific fear of social or performance situations in which the individual is exposed to unfamiliar individuals or to possible examination and evaluation by others. Examples include having to converse at a party, use public restrooms, and meet strangers. In each case, the affected individual is aware that the experienced fear is excessive and unreasonable given the circumstance. The specific content of a phobia may vary across gender, ethnic, and cultural boundaries.

Phobic disorders are common, affecting ~10% of the population. Full criteria for diagnosis are usually satisfied first in early adulthood, but behavioral avoidance of unfamiliar people, situations, or objects dating from early childhood is common.

In one study of female twins, concordance rates for agoraphobia, social phobia, and animal phobia were found to be 23% for monozygotic twins and 15% for dizygotic twins. A twin study of fear conditioning, a model for the acquisition of phobias, demonstrated a heritability of 35–45%, and a genomewide linkage scan identified a risk locus on chromosome 14 in a region previously implicated in a mouse model of fear. Animal studies of fear conditioning have indicated that processing of the fear stimulus occurs through the lateral nucleus of the amygdala, extending through the central nucleus and projecting to the periaqueductal gray region, lateral hypothalamus, and paraventricular hypothalamus.

TREATMENT: PHOBIC DISORDERS

Beta blockers (e.g., propranolol, 20–40 mg orally 2 hours before the event) are particularly effective in the treatment of "performance anxiety" (but not general social phobia) and appear to work by blocking the peripheral manifestations of anxiety such as perspiration, tachycardia, palpitations, and tremor. MAOIs alleviate social phobia independently of their antidepressant activity, and paroxetine, sertraline, and venlafaxine have received FDA approval for treatment of social anxiety. Benzodiazepines can be helpful in reducing fearful avoidance, but the chronic nature of phobic disorders limits their usefulness.

Behaviorally focused psychotherapy is an important component of treatment, as relapse rates are high when medication is used as the sole treatment. Cognitive-behavioral strategies are based upon the

finding that distorted perceptions and interpretations of fear-producing stimuli play a major role in perpetuation of phobias. Individual and group therapy sessions teach the patient to identify specific negative thoughts associated with the anxiety-producing situation and help to reduce the patient's fear of loss of control. In desensitization therapy, hierarchies of feared situations are constructed and the patient is encouraged to pursue and master gradual exposure to the anxiety-producing stimuli.

Patients with social phobia, in particular, have a high rate of comorbid alcohol abuse, as well as of other psychiatric conditions (e.g., eating disorders), necessitating the need for parallel management of each disorder if anxiety reduction is to be achieved.

Stress Disorders

CLINICAL MANIFESTATIONS

Patients may develop anxiety after exposure to extreme traumatic events such as the threat of personal death or injury or the death of a loved one. The reaction may occur shortly after the trauma (*acute stress disorder*) or be delayed and subject to recurrence (PTSD) **(Table 391-8)**. In both syndromes, individuals experience associated symptoms of detachment and loss of emotional responsivity. The patient may feel depersonalized and unable to recall specific aspects of the trauma, though typically it is reexperienced through intrusions in thought, dreams, or flashbacks, particularly when cues of the original event are present. Patients often actively avoid stimuli that precipitate recollections of the trauma and demonstrate a resulting increase in vigilance, arousal, and startle response. Patients with stress disorders are at risk for the development of other disorders related to anxiety, mood, and substance abuse (especially alcohol). Between 5 and 10% of Americans will at some time in their life satisfy criteria for PTSD, with women more likely to be affected than men.

Table 391-8 Diagnostic Criteria for Posttraumatic Stress Disorder

A. The person has been exposed to a traumatic event in which both of the following were present:

1. The person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others.

2. The person's response involved intense fear, helplessness, or horror.

B. The traumatic event is persistently reexperienced in one (or more) of the following ways:

1. Recurrent and intrusive distressing recollections of the event, including images, thoughts, or perceptions.

2. Recurrent distressing dreams of the event.

3. Acting or feeling as if the traumatic event were recurring (includes a sense of reliving the experience, illusions, hallucinations, and dissociative flashback episodes, including those that occur on awakening or when intoxicated).

4. Intense psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event.

5. Physiologic reactivity on exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event.

C. Persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness (not present before the trauma), as indicated by three or more of the following:

1. Efforts to avoid thoughts, feelings, or conversations associated with the trauma

2. Efforts to avoid activities, places, or people that arouse recollections of the trauma

3. Inability to recall an important aspect of the trauma

4. Markedly diminished interest or participation in significant activities

5. Feeling of detachment or estrangement from others

6. Restricted range of affect (e.g., unable to have loving feelings)

7. Sense of a foreshortened future (e.g., does not expect to have a career, marriage, children, or a normal life span)

D. Persistent symptoms of increased arousal (not present before the trauma), as indicated by two (or more) of the following:

1. Difficulty falling or staying asleep

2. Irritability or outbursts of anger

3. Difficulty concentrating

4. Hypervigilance

5. Exaggerated startle response

E. Duration of the disturbance (symptoms in criteria B, C, and D) is more than 1 month
F. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Source: *Diagnostic and Statistical Manual of Mental Disorders* , 4th ed. Washington, DC, American Psychiatric Association, 2000.

Risk factors for the development of PTSD include a past psychiatric history and personality characteristics of high neuroticism and extroversion. Twin studies show a substantial genetic influence on all symptoms associated with PTSD, with less evidence for an environmental effect.

ETIOLOGY AND PATHOPHYSIOLOGY

It is hypothesized that in PTSD there is excessive release of norepinephrine from the locus coeruleus in response to stress and increased noradrenergic activity at projection sites in the hippocampus and amygdala. These changes theoretically facilitate the encoding of fear-based memories. Greater sympathetic responses to cues associated with the traumatic event occur in PTSD, although pituitary adrenal responses are blunted.

TREATMENT: STRESS DISORDERS

Acute stress reactions are usually self-limited, and treatment typically involves the short-term use of benzodiazepines and supportive/expressive psychotherapy. The chronic and recurrent nature of PTSD, however, requires a more complex approach employing drug and behavioral treatments. PTSD is highly correlated with peritraumatic dissociative symptoms and the development of an acute stress disorder at the time of the trauma. TCAs such as imipramine and amitriptyline, the MAOI phenelzine, and the SSRIs can all reduce anxiety, symptoms of intrusion, and avoidance behaviors, as can prazosin, an α_1 antagonist. Propranolol and opiates such as morphine, given during the acute stress period may have beneficial effects in preventing the development of PTSD. Trazodone, a sedating antidepressant, is frequently used at night to help with insomnia (50–150 mg qhs). Carbamazepine, valproic acid, or alprazolam have also independently produced improvement in uncontrolled trials. Psychotherapeutic strategies for PTSD help the patient overcome avoidance behaviors and demoralization and master fear of recurrence of the trauma; therapies that encourage the patient to dismantle avoidance behaviors through stepwise focusing on the experience of the traumatic event are the most effective.

Obsessive-Compulsive Disorder

CLINICAL MANIFESTATIONS

Obsessive-compulsive disorder (OCD) is characterized by obsessive thoughts and compulsive behaviors that impair everyday functioning. Fears of contamination and germs are common, as are handwashing, counting behaviors, and having to check and recheck such actions as whether a door is locked. The degree to which the disorder is disruptive for the individual varies, but in all cases obsessive-compulsive activities take up >1 hour per day and are undertaken to relieve the anxiety triggered by the core fear. Patients often conceal their symptoms, usually because they are embarrassed by the content of their thoughts or the nature of their actions. Physicians must ask specific questions regarding recurrent thoughts and behaviors, particularly if physical clues such as chafed and reddened hands or patchy hair loss (from repetitive hair pulling, or trichotillomania) are present. Comorbid conditions are common, the most frequent being depression, other anxiety disorders, eating disorders, and tics. OCD has a lifetime prevalence of 2–3% worldwide. Onset is usually gradual, beginning in early adulthood, but childhood onset is not rare. The disorder usually has a waxing and waning course, but some cases may show a steady deterioration in psychosocial functioning.

ETIOLOGY AND PATHOPHYSIOLOGY

A genetic contribution to OCD is suggested by twin studies. A genomewide association study (GWAS) reported linkage to chromosome 2p23.2; however, no susceptibility gene for OCD has been identified to date. Family studies show an aggregation of OCD with Tourette's disorder, and both are more common in males and in first-born children.

The anatomy of obsessive-compulsive behavior is thought to include the orbital frontal cortex, caudate nucleus, and globus pallidus. The caudate nucleus appears to be involved in the acquisition and maintenance of habit and skill learning, and interventions that are successful in reducing obsessive-compulsive behaviors also decrease metabolic activity measured in the caudate.

TREATMENT: OBSESSIVE-COMPULSIVE DISORDER

Clomipramine, fluoxetine, fluvoxamine, and sertraline are approved for the treatment of OCD. Clomipramine is a TCA that is often tolerated poorly owing to anticholinergic and sedative side effects at the doses required to treat the illness (25–250 mg/d); its efficacy in OCD is unrelated to its antidepressant activity. Fluoxetine (5–60 mg/d), fluvoxamine (25–300 mg/d), and sertraline (50–150 mg/d) are as effective as clomipramine and have a more benign side effect profile. Only 50–60% of

patients with OCD show adequate improvement with pharmacotherapy alone. In treatment-resistant cases, augmentation with other serotonergic agents such as buspirone, or with a neuroleptic or benzodiazepine may be beneficial and in severe cases deep brain stimulation has been found to be effective. When a therapeutic response is achieved, long-duration maintenance therapy is usually indicated.

For many individuals, particularly those with time-consuming compulsions, behavior therapy will result in as much improvement as that afforded by medication. Effective techniques include the gradual increase in exposure to stressful situations, maintenance of a diary to clarify stressors, and homework assignments that substitute new activities for compulsive behaviors.

MOOD DISORDERS

Mood disorders are characterized by a disturbance in the regulation of mood, behavior, and affect. Mood disorders are subdivided into (1) depressive disorders, (2) bipolar disorders, and (3) depression in association with medical illness or alcohol and substance abuse (Chaps. 392, 393, and 394). Major depressive disorder (MDD) is differentiated from bipolar disorder by the absence of a manic or hypomanic episode. The relationship between pure depressive syndromes and bipolar disorders is not well understood; MDD is more frequent in families of bipolar individuals, but the reverse is not true. In the Global Burden of Disease Study conducted by the World Health Organization, unipolar major depression ranked fourth among all diseases in terms of disability-adjusted life-years and was projected to rank second by the year 2020. In the United States, lost productivity directly related to mood disorders has been estimated at $55.1 billion per year.

Depression in Association with Medical Illness

Depression occurring in the context of medical illness is difficult to evaluate. Depressive symptomatology may reflect the psychological stress of coping with the disease, may be caused by the disease process itself or by the medications used to treat it, or may simply coexist in time with the medical diagnosis.

Virtually every class of *medication* includes some agent that can induce depression. Antihypertensive drugs, anticholesterolemic agents, and antiarrhythmic agents are common triggers of depressive symptoms. Iatrogenic depression should also be considered in patients receiving glucocorticoids, antimicrobials, systemic analgesics, antiparkinsonian medications, and anticonvulsants. To decide whether a causal relationship exists between pharmacologic therapy and a patient's change in mood, it may sometimes be necessary to undertake an empirical trial of an alternative medication.

Between 20 and 30% of *cardiac* patients manifest a depressive disorder; an even higher percentage experience depressive symptomatology when self-reporting scales are used. Depressive symptoms following unstable angina, myocardial infarction, cardiac bypass surgery, or heart transplant impair rehabilitation and are associated with higher rates of mortality and medical morbidity. Depressed patients often show decreased variability in heart rate (an index of reduced parasympathetic nervous system activity); which may predispose individuals to ventricular arrhythmia and increased morbidity. Depression also appears to increase the risk of developing coronary heart disease, possibly through increased platelet aggregation. TCAs are contraindicated in patients with bundle branch block, and TCA-induced tachycardia is an additional concern in patients with congestive heart failure. SSRIs appear not to induce ECG changes or adverse cardiac events and thus are reasonable first-line drugs for patients at risk for TCA-related complications. SSRIs may interfere with hepatic metabolism of anticoagulants, however, causing increased anticoagulation.

In patients with *cancer*, the mean prevalence of depression is 25%, but depression occurs in 40–50% of patients with cancers of the pancreas or oropharynx. This association is not due to the effect of cachexia alone, as the higher prevalence of depression in patients with pancreatic cancer persists when compared to those with advanced gastric cancer. Initiation of antidepressant medication in cancer patients has been shown to improve quality of life as well as mood. Psychotherapeutic approaches, particularly group therapy, may have some effect on short-term depression, anxiety, and pain symptoms.

Depression occurs frequently in patients with *neurologic disorders*, particularly cerebrovascular disorders, Parkinson's disease, dementia, multiple sclerosis, and traumatic brain injury. One in five patients with left-hemisphere stroke involving the dorsolateral frontal cortex experiences major depression. Late-onset depression in otherwise cognitively normal individuals increases the risk of a subsequent diagnosis of Alzheimer's disease. Both TCA and SSRI agents are effective against these depressions, as are stimulant compounds and, in some patients, MAOIs.

The reported prevalence of depression in patients with *diabetes mellitus* varies from 8 to 27%, with the severity of the mood state correlating with the level of hyperglycemia and the presence of diabetic complications. Treatment of depression may be complicated by effects of antidepressive agents on glycemic control. MAOIs can induce hypoglycemia and weight gain, while TCAs can produce hyperglycemia and carbohydrate craving. SSRIs, like MAOIs, may reduce fasting plasma glucose, but they are easier to use and may also improve dietary and medication compliance.

Hypothyroidism is frequently associated with features of depression, most commonly depressed mood and memory impairment. Hyperthyroid states may also present in a similar fashion, usually in geriatric populations. Improvement in mood usually follows normalization of thyroid function, but adjunctive antidepressant medication is sometimes required. Patients with subclinical hypothyroidism can also experience symptoms of depression and cognitive difficulty that respond to thyroid replacement.

The lifetime prevalence of depression in *HIV-positive* individuals has been estimated at 22–45%. The relationship between depression and disease progression is multifactorial and likely to involve psychological and social factors, alterations in immune function, and central nervous system (CNS) disease. Chronic hepatitis C infection is also associated with depression, which may worsen with interferon-α treatment.

Some chronic disorders of uncertain etiology such as chronic fatigue syndrome (Chap. 389) and fibromyalgia (Chap. 335), are strongly associated with depression and anxiety; patients may benefit from antidepressant treatment or anticonvulsant agents such as pregabalin,

Depressive Disorders

CLINICAL MANIFESTATIONS

Major depression is defined as depressed mood on a daily basis for a minimum duration of 2 weeks **(Table 391-9)**. An episode may be characterized by sadness, indifference, apathy, or irritability and is usually associated with changes in sleep patterns, appetite, and weight; motor agitation or retardation; fatigue; impaired concentration and decision making; feelings of shame or guilt; and thoughts of death or dying. Patients with depression have a profound loss of pleasure in all enjoyable activities, exhibit early morning awakening, feel that the dysphoric mood state is qualitatively different from sadness, and often notice a diurnal variation in mood (worse in morning hours).

Table 391-9 Criteria for a Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure. Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.
1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others)
3. Significant weight loss when not dieting or weight gain (e.g., a change of >5% of body weight in a month), or decrease or increase in appetite nearly every day
4. Insomnia or hypersomnia nearly every day
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
6. Fatigue or loss of energy nearly every day
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide
B. The symptoms do not meet criteria for a mixed episode
C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning
D. The symptoms are not due to the direct physiologic effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism)
E. The symptoms are not better accounted for by bereavement (i.e., after the loss of a loved one), the symptoms persist for >2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation

Source: *Diagnostic and Statistical Manual of Mental Disorders* , 4th ed. Washington, DC, American Psychiatric Association, 2000.

Approximately 15% of the population experiences a major depressive episode at some point in life, and 6–8% of all outpatients in primary care settings satisfy diagnostic criteria for the disorder. Depression is often undiagnosed, and, even more frequently, it is treated inadequately. If a physician suspects the presence of a major depressive episode, the initial task is to determine whether it represents unipolar or bipolar depression or is one of the 10–15% of cases that are secondary to general medical illness

or substance abuse. Physicians should also assess the risk of suicide by direct questioning, as patients are often reluctant to verbalize such thoughts without prompting. If specific plans are uncovered or if significant risk factors exist (e.g., a past history of suicide attempts, profound hopelessness, concurrent medical illness, substance abuse, or social isolation), the patient must be referred to a mental health specialist for immediate care. The physician should specifically probe each of these areas in an empathic and hopeful manner, being sensitive to denial and possible minimization of distress. The presence of anxiety, panic, or agitation significantly increases near-term suicidal risk. Approximately 4–5% of all depressed patients will commit suicide; most will have sought help from physicians within 1 month of their deaths.

In some depressed patients, the mood disorder does not appear to be episodic and is not clearly associated with either psychosocial dysfunction or change from the individual's usual experience in life. *Dysthymic disorder* consists of a pattern of chronic (at least 2 years), ongoing, mild depressive symptoms that are less severe and less disabling than those found in major depression; the two conditions are sometimes difficult to separate, however, and can occur together ("double depression"). Many patients who exhibit a profile of pessimism, disinterest, and low self-esteem respond to antidepressant treatment. Dysthymic disorder exists in ~5% of primary care patients. The term *minor depression* is used for individuals who experience at least two depressive symptoms for 2 weeks but who do not meet the full criteria for major depression. Despite its name, minor depression is associated with significant morbidity and disability and also responds to pharmacologic treatment.

Depression is approximately twice as common in women as in men, and the incidence increases with age in both sexes. Twin studies indicate that the liability to major depression of early onset (before age 25) is largely genetic in origin. Negative life events can precipitate and contribute to depression, but genetic factors influence the sensitivity of individuals to these stressful events. In most cases, both biologic and psychosocial factors are involved in the precipitation and unfolding of depressive episodes. The most potent stressors appear to involve death of a relative, assault, or severe marital or relationship problems.

Unipolar depressive disorders usually begin in early adulthood and recur episodically over the course of a lifetime. The best predictor of future risk is the number of past episodes; 50–60% of patients who have a first episode have at least one or two recurrences. Some patients experience multiple episodes that become more severe and frequent over time. The duration of an untreated episode varies greatly, ranging from a few months to ≤1 year. The pattern of recurrence and clinical progression in a developing episode are also variable. Within an individual, the nature of episodes (e.g., specific presenting symptoms, frequency and duration) may be similar over time. In a minority of patients, a severe depressive episode may progress to a psychotic state; in elderly patients, depressive symptoms may be associated with cognitive deficits mimicking dementia ("pseudodementia"). A seasonal pattern of depression, called *seasonal affective disorder* , may manifest with onset and remission of episodes at predictable times of the year. This disorder is more common in women, whose symptoms are anergy, fatigue, weight gain, hypersomnia, and episodic carbohydrate craving. The prevalence increases with distance from the equator, and improvement may occur by altering light exposure.

ETIOLOGY AND PATHOPHYSIOLOGY

Although evidence for genetic transmission of unipolar depression is not as strong as in bipolar disorder, monozygotic twins have a higher concordance rate (46%) than dizygotic siblings (20%), with little support for any effect of a shared family environment.

Neuroendocrine abnormalities that reflect the neurovegetative signs and symptoms of depression include: (1) increased cortisol and corticotropin-releasing hormone (CRH) secretion, (2) an increase in adrenal size, (3) a decreased inhibitory response of glucocorticoids to dexamethasone, and (4) a blunted response of thyroid-stimulating hormone (TSH) level to infusion of thyroid-releasing hormone (TRH). Antidepressant treatment leads to normalization of these abnormalities. Major depression is also associated with an upregulation of proinflammatory cytokines, which also normalizes with antidepressant treatment.

Diurnal variations in symptom severity and alterations in circadian rhythmicity of a number of neurochemical and neurohumoral factors suggest that biologic differences may be secondary to a primary defect in regulation of biologic rhythms. Patients with major depression show consistent findings of a decrease in rapid eye movement (REM) sleep onset (REM latency), an increase in REM density, and, in some subjects, a decrease in stage IV delta slow-wave sleep.

Although antidepressant drugs inhibit neurotransmitter uptake within hours, their therapeutic effects typically emerge over several weeks, implicating adaptive changes in second messenger systems and transcription factors as possible mechanisms of action.

The pathogenesis of depression is discussed in detail in Chap. 390.

TREATMENT: DEPRESSIVE DISORDERS

Treatment planning requires coordination of short-term strategies to induce remission combined with

longer term maintenance designed to prevent recurrence. The most effective intervention for achieving remission and preventing relapse is medication, but combined treatment, incorporating psychotherapy to help the patient cope with decreased self-esteem and demoralization, improves outcome **(Fig. 391-1)**. Approximately 40% of primary care patients with depression drop out of treatment and discontinue medication if symptomatic improvement is not noted within a month, unless additional support is provided. Outcome improves with (1) increased intensity and frequency of visits during the first 4–6 weeks of treatment, (2) supplemental educational materials, and (3) psychiatric consultation as indicated. Despite the widespread use of SSRIs and other second-generation antidepressant drugs, there is no convincing evidence that this class of antidepressant is more efficacious than TCAs. Between 60 and 70% of all depressed patients respond to any drug chosen, if it is given in a sufficient dose for 6–8 weeks. There is no ideal antidepressant; no current compound combines rapid onset of action, moderate half-life, a meaningful relationship between dose and blood level, a low side effect profile, minimal interaction with other drugs, and safety in overdose.

Figure 391-1

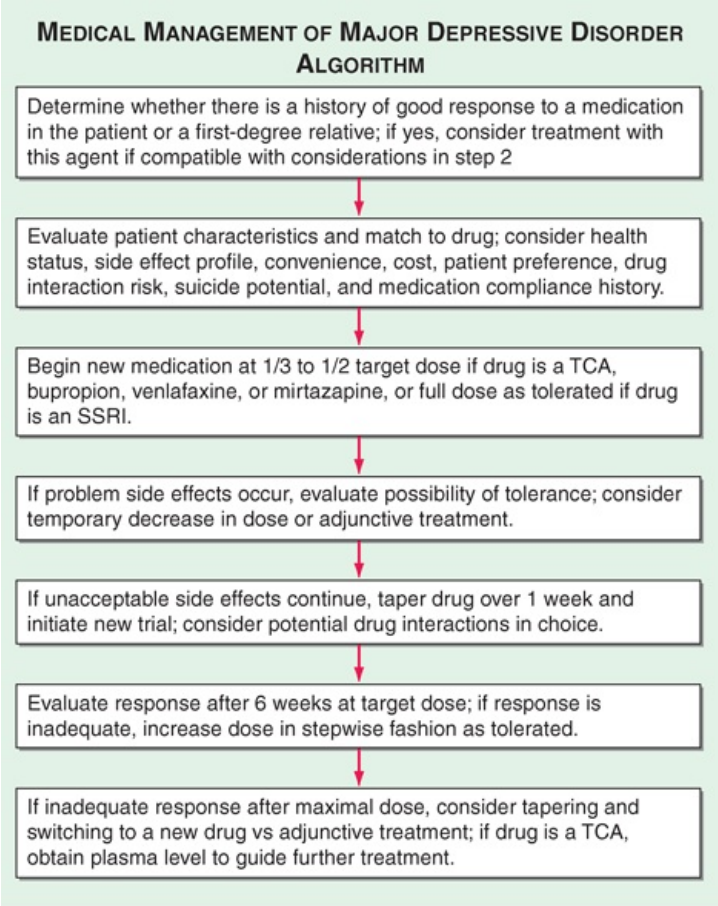

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A guideline for the medical management of major depressive disorder. SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

A rational approach to selecting which antidepressant to use involves matching the patient's preference and medical history withthe metabolic and side effect profile of the drug **(Tables 391-4 and 391-5)**. A previous response, or a family history of a positive response, to a specific antidepressant often suggests that that drug be tried first. Before initiating antidepressant therapy, the physician should evaluate the possible contribution of comorbid illnesses and consider their specific treatment. In individuals with suicidal ideation, particular attention should be paid to choosing a drug with low toxicity if taken in overdose. The SSRIs, and other newer antidepressant drugs are distinctly safer in this regard; nevertheless, the advantages of TCAs have not been completely superseded. The existence of generic equivalents make TCAs relatively cheap, and for secondary tricyclics, particularly nortriptyline and desipramine, well-defined relationships among dose, plasma level, and therapeutic response exist. The steady-state plasma level achieved for a given drug dose can vary more than tenfold between individuals and plasma levels may help in interpreting apparent resistance to treatment and/or unexpected drug toxicity. The principal side effects of TCAs are antihistamine (sedation) and anticholinergic (constipation, dry mouth, urinary hesitancy, blurred vision). TCAs are

contraindicated in patients with serious cardiovascular risk factors and overdoses of tricyclic agents can be lethal, with desipramine carrying the greatest risk. It is judicious to prescribe only a 10-day supply when suicide is a risk. Most patients require a daily dose of 150–200 mg of imipramine or amitriptyline or its equivalent to achieve a therapeutic blood level of 150–300 ng/mL and a satisfactory remission; some patients show a partial effect at lower doses. Geriatric patients may require a low starting dose and slow escalation. Ethnic differences in drug metabolism are significant, with Hispanic, Asian, and black patients generally requiring lower doses than whites to achieve a comparable blood level. P450 profiling using genetic chip technology may be clinically useful in predicting individual sensitivity.

Second-generation antidepressants include amoxapine, maprotiline, trazodone, and bupropion. Amoxapine is a dibenzoxazepine derivative that blocks norepinephrine and serotonin reuptake and has a metabolite that shows a degree of dopamine blockade. Long-term use of this drug carries a risk of tardivedyskinesia. Maprotiline is a potent noradrenergic reuptake blocker that has little anticholinergic effect but may produce seizures. Bupropion is a novel antidepressant whose mechanism of action is thought to involve enhancement of noradrenergic function. It has no anticholinergic, sedating, or orthostatic side effects and has a low incidence of sexual side effects. It may, however, be associated with stimulant-like side effects, may lower seizure threshold, and has an exceptionally short half-life, requiring frequent dosing. An extended-release preparation is available.

SSRIs such as fluoxetine, sertraline, paroxetine, citalopram, and escitalopram cause a lower frequency of anticholinergic, sedating, and cardiovascular side effects but a possibly greater incidence of gastrointestinal complaints, sleep impairment, and sexual dysfunction than do TCAs. Akathisia, involving an inner sense of restlessness and anxiety in addition to increased motor activity, may also be more common, particularly during the first week of treatment. One concern is the risk of "serotonin syndrome," thought to result from hyperstimulation of brainstem $5HT_{1A}$ receptors and characterized by myoclonus, agitation, abdominal cramping, hyperpyrexia, hypertension, and potentially death. Serotonergic agonists taken in combination should be monitored closely for this reason. Considerations such as half-life, compliance, toxicity, and drug-drug interactions may guide the choice of a particular SSRI. Fluoxetine and its principal active metabolite, norfluoxetine, for example, have a combined half-life of almost 7 days, resulting in a delay of 5 weeks before steady-state levels are achieved and a similar delay for complete drug excretion once its use is discontinued. All the SSRIs may impair sexual function, resulting in diminished libido, impotence, or difficulty in achieving orgasm. Sexual dysfunction frequently results in noncompliance and should be asked about specifically. Sexual dysfunction can sometimes be ameliorated by lowering the dose, by instituting weekend drug holidays (two or three times a month), or by treatment with amantadine (100 mg tid), bethanechol (25 mg tid), buspirone (10 mg tid), or bupropion (100–150 mg/d). Paroxetine appears to be more anticholinergic than either fluoxetine or sertraline, and sertraline carries a lower risk of producing an adverse drug interaction than the other two. Rare side effects of SSRIs include angina due to vasospasm and prolongation of the prothrombin time. Escitalopram is the most specific of currently available SSRIs and appears to have no specific inhibitory effects on the P450 system.

Venlafaxine, desvenlafaxine, and duloxetine block the reuptake of both norepinephrine and serotonin but produce relatively little in the way of traditional tricyclic side effects. Unlike the SSRIs, venlafaxine has a relatively linear dose-response curve. Patients should be monitored for a possible increase in diastolic blood pressure, and multiple daily dosing is required because of the drug's short half-life. An extended-release form is available and has a somewhat lower incidence of gastrointestinal side effects. Mirtazapine is a TCA that has a unique spectrum of activity. It increases noradrenergic and serotonergic neurotransmission through a blockade of central α_2-adrenergic receptors and postsynaptic $5HT_2$ and $5HT_3$ receptors. It is also strongly antihistaminic and, as such, may produce sedation.

With the exception of citalopram and escitalopram, each of the SSRIs may inhibit one or more cytochrome P450 enzymes. Depending on the specific isoenzyme involved, the metabolism of a number of concomitantly administered medications can be dramatically affected. Fluoxetine and paroxetine, for example, by inhibiting 2D6, can cause dramatic increases in the blood level of type 1C antiarrhythmics, while sertraline, by acting on 3A4, may alter blood levels of carbamazepine, or digoxin.

The MAOIs are highly effective, particularly in atypical depression, but the risk of hypertensive crisis following intake of tyramine-containing food or sympathomimetic drugs makes them inappropriate as first-line agents. Transdermal selegiline may avert this risk at low dose. Common side effects include orthostatic hypotension, weight gain, insomnia, and sexual dysfunction. MAOIs should not be used concomitantly with SSRIs, because of the risk of serotonin syndrome, or with TCAs, because of possible hyperadrenergic effects.

Electroconvulsive therapy is at least as effective as medication, but its use is reserved for treatment-resistant cases and delusional depressions. Transcranial magnetic stimulation (TMS) is approved for treatment-resistant depression and has been shown to have efficacy in several controlled trials. Vagus nerve stimulation (VNS) has also recently been approved for treatment-resistant depression, but its degree of efficacy is controversial. Deep brain stimulation is another treatment that is being used

experimentally in treatment resistant cases.

Regardless of the treatment undertaken, the response should be evaluated after ~2 months. Three-quarters of patients show improvement by this time, but if remission is inadequate the patient should be questioned about compliance and an increase in medication dose should be considered if side effects are not troublesome. If this approach is unsuccessful, referral to a mental health specialist is advised. Strategies for treatment then include selection of an alternative drug, combinations of antidepressants, and/or adjunctive treatment with other classes of drugs, including lithium, thyroid hormone, atypical antipsychotic agents, and dopamine agonists. A large randomized trial (STAR-D) was unable to show preferential efficacy, but the addition of atypical antipsychotic drugs has received FDA approval. Patients whose response to an SSRI wanes over time may benefit from the addition of buspirone (10 mg tid) or pindolol (2–5 mg tid) or small amounts of a TCA such as desipramine (25 mg bid or tid). Most patients will show some degree of response but aggressive treatment should be pursued until remission is achieved, and drug treatment should be continued for at least 6–9 more months to prevent relapse. In patients who have had two or more episodes of depression, indefinite maintenance treatment should be considered.

It is essential to educate patients both about depression and the benefits and side effects of medications they are receiving. Advice about stress reduction and cautions that alcohol may exacerbate depressive symptoms and impair drug response are helpful. Patients should be given time to describe their experience, their outlook, and the impact of the depression on them and their families. Occasional empathic silence may be as helpful for the treatment alliance as verbal reassurance. Controlled trials have shown that cognitive-behavioral and interpersonal therapies are effective in improving psychological and social adjustment and that a combined treatment approach is more successful than medication alone for many patients.

Bipolar Disorder

CLINICAL MANIFESTATIONS

Bipolar disorder is characterized by unpredictable swings in mood from mania (or hypomania) to depression. Some patients suffer only from recurrent attacks of *mania*, which in its pure form is associated with increased psychomotor activity; excessive social extroversion; decreased need for sleep; impulsivity and impairment in judgment; and expansive, grandiose, and sometimes irritable mood **(Table 391-10)**. In severe mania, patients may experience delusions and paranoid thinking indistinguishable from schizophrenia. One-half of patients with bipolar disorder present with a mixture of psychomotor agitation and activation with dysphoria, anxiety, and irritability. It may be difficult to distinguish *mixed mania* from *agitated depression*. In some bipolar patients (*bipolar II disorder*), the full criteria for mania are lacking, and the requisite recurrent depressions are separated by periods of mild activation and increased energy (hypomania). In *cyclothymic disorder*, there are numerous hypomanic periods, usually of relatively short duration, alternating with clusters of depressive symptoms that fail, either in severity or duration, to meet the criteria of major depression. The mood fluctuations are chronic and should be present for at least2 years before the diagnosis is made.

Table 391-10 Criteria for a Manic Episode

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary)
B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:
1. Inflated self-esteem or grandiosity
2. Decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
3. More talkative than usual or pressure to keep talking
4. Flight of ideas or subjective experience that thoughts are racing
5. Distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
6. Increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation
7. Excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)
C. The symptoms do not meet criteria for a mixed episode.
D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.
E. The symptoms are not due to the direct physiologic effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).

Note: Manic-like episodes that are clearly caused by somatic antidepressant treatment (e.g., medication, electroconvulsive therapy, light therapy) should not count toward a diagnosis of bipolar I disorder.

Source: *Diagnostic and Statistical Manual of Mental Disorders* , 4th ed. Washington, DC, American Psychiatric Association, 2000.

Manic episodes typically emerge over a period of days to weeks, but onset within hours is possible, usually in the early morning hours. An untreated episode of either depression or mania can be as short as several weeks or last as long as 8–12 months, and rare patients have an unremitting chronic course. The term *rapid cycling* is used for patients who have four or more episodes of either depression or mania in a given year. This pattern occurs in 15% of all patients, almost all of whom are women. In some cases, rapid cycling is linked to an underlying thyroid dysfunction and, in others, it is iatrogenically triggered by prolonged antidepressant treatment. Approximately one-half of patients have sustained difficulties in work performance and psychosocial functioning, with depressive phases being more responsible for impairment than mania.

Bipolar disorder is common, affecting ~1.5% of the population in the United States. Onset is typically between 20 and 30 years of age, but many individuals report premorbid symptoms in late childhood or early adolescence. The prevalence is similar for men and women; women are likely to have more depressive and men more manic episodes over a lifetime.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of mania includes secondary mania induced by stimulant or sympathomimetic drugs, hyperthyroidism, AIDS, neurologic disorders, such as Huntington's or Wilson's disease, and cerebrovascular accidents. Comorbidity with alcohol and substance abuse is common, either because of poor judgment and increased impulsivity or because of an attempt to self-treat the underlying mood symptoms and sleep disturbances.

ETIOLOGY AND PATHOPHYSIOLOGY

Genetic predisposition to bipolar disorder is evident from family studies; the concordance rate for monozygotic twins approaches 80%. Patients with bipolar disorder also appear to have altered circadian rhythmicity, and lithium may exert its therapeutic benefit through a resynchronization of intrinsic rhythms keyed to the light/dark cycle. **A detailed discussion of the pathogenesis of bipolar disorder is presented inChap. 390.**

TREATMENT: BIPOLAR DISORDER

(Table 391-11) Lithium carbonate is the mainstay of treatment in bipolar disorder, although sodium valproate and olanzapine are equally effective in acute mania, as is lamotrigine in the depressed phase. The response rate to lithium carbonate is 70–80% in acute mania, with beneficial effects appearing in 1–2 weeks. Lithium also has a prophylactic effect in prevention of recurrent mania and, to a lesser extent, in the prevention of recurrent depression. A simple cation, lithium is rapidly absorbed from the gastrointestinal tract and remains unbound to plasma or tissue proteins. Some 95% of a given dose is excreted unchanged through the kidneys within 24 hours.

Table 391-11 Clinical Pharmacology of Mood Stabilizers

Agent and Dosing	Side Effects and Other Effects
Lithium	**Common Side Effects**
Starting dose: 300 mg bid or tid Therapeutic blood level: 0.8–1.2 meq/L	Nausea/anorexia/diarrhea, fine tremor, thirst, polyuria, fatigue, weight gain, acne, folliculitis, neutrophilia, hypothyroidism Blood level is increased by thiazides, tetracyclines, and NSAIDs Blood level is decreased by bronchodilators, verapamil, and carbonic anhydrase inhibitors *Rare side effects:* Neurotoxicity, renal toxicity, hypercalcemia, ECG changes
Valproic Acid	**Common Side Effects**
Starting dose: 250 mg tid Therapeutic blood level: 50–125 μg/mL	Nausea/anorexia, weight gain, sedation, tremor, rash, alopecia Inhibits hepatic metabolism of other medications *Rare side effects:* Pancreatitis, hepatotoxicity, Stevens-Johnson syndrome
Carbamazepine/Oxcarbazepine	**Common Side Effects**
Starting dose: 200 mg bid for carbamazepine, 150 mg bid for	Nausea/anorexia, sedation, rash, dizziness/ataxia

carbamazepine, 150 mg bid for oxcarbazepine Therapeutic blood level: 4–12 μg/mL for carbamazepine	Carbamazepine, but not oxcarbazepine, induces hepatic metabolism of other medications *Rare side effects:* Hyponatremia, agranulocytosis, Stevens-Johnson syndrome
Lamotrigine	**Common Side Effects**
Starting dose: 25 mg/d	Rash, dizziness, headache, tremor, sedation, nausea *Rare side effect:* Stevens-Johnson syndrome

Abbreviations: NSAIDs, nonsteroidal anti-inflammatory drugs; ECG, electrocardiogram.

Serious side effects from lithium are rare, but minor complaints such as gastrointestinal discomfort, nausea, diarrhea, polyuria, weight gain, skin eruptions, alopecia, and edema are common. Over time, urine-concentrating ability may be decreased, but significant nephrotoxicity does not usually occur. Lithium exerts an antithyroid effect by interfering with the synthesis and release of thyroid hormones. More serious side effects include tremor, poor concentration and memory, ataxia, dysarthria, and incoordination. There is suggestive, but not conclusive, evidence that lithium is teratogenic, inducing cardiac malformations in the first trimester.

In the treatment of acute mania, lithium is initiated at 300 mg bid or tid, and the dose is then increased by 300 mg every 2–3 days to achieve blood levels of 0.8–1.2 meq/L. Because the therapeutic effect of lithium may not appear until after 7–10 days of treatment, adjunctive usage of lorazepam (1–2 mg every 4 hours) or clonazepam (0.5–1 mg every 4 hours) may be beneficial to control agitation. Antipsychotics are indicated in patients with severe agitation who respond only partially to benzodiazepines. Patients using lithium should be monitored closely, since the blood levels required to achieve a therapeutic benefit are close to those associated with toxicity.

Valproic acid may be better than lithium for patients who experience rapid cycling (i.e., more than four episodes a year) or who present with a mixed or dysphoric mania. Tremor and weight gain are the most common side effects; hepatotoxicity and pancreatitis are rare toxicities.

Carbamazepine and oxcarbazepine, although not formally approved by the FDA for bipolar disorder, have clinical efficacy in the treatment of acute mania. Second-generation antipsychotic drugs (olanzapine, quetiapine, risperidone, ziprasidone, aripiprazole, and asenapine) have also been shown to be effective, either alone or in combination with a mood stabilizer. An increased risk of weight gain and other metabolic abnormalities is a concern with these agents.

The recurrent nature of bipolar mood disorder necessitates maintenance treatment. A sustained blood lithium level of at least 0.8 mEq/L is important for optimal prophylaxis and has been shown to reduce risk of suicide, a finding not yet apparent for other mood stabilizers. Compliance is frequently an issue and often requires enlistment and education of concerned family members. Efforts to identify and modify psychosocial factors that may trigger episodes are important, as is an emphasis on lifestyle regularity. Antidepressant medications are sometimes required for the treatment of severe breakthrough depressions, but their use should generally be avoided during maintenance treatment because of the risk of precipitating mania or accelerating the cycle frequency. Loss of efficacy over time may be observed with any of the mood-stabilizing agents. In such situations, an alternative agent or combination therapy is usually helpful.

Consensus guidelines for the treatment of acute mania and bipolar depression are described in **Table 391-12**.

Table 391-12 Consensus Guidelines on the Drug Treatment of Acute Mania and Bipolar Depression

Condition	Preferred Agents
Euphoric mania	Lithium
Mixed/dysphoric mania	Valproic acid
Mania with psychosis	Valproic acid with olanzapine, aripiprazole, conventional antipsychotic, or risperidone
Hypomania	Lithium, lamotrigine, or valproic acid alone
Severe depression with psychosis	Venlafaxine, bupropion, or paroxetine plus lithium plus olanzapine, or risperidone; consider ECT
Severe depression without psychosis	Bupropion, paroxetine, sertraline, venlafaxine, or citalopram plus lithium
Mild to moderate depression	Lithium or lamotrigine alone; add bupropion if needed

Abbreviation: ECT, electroconvulsive therapy.

Source: From GS Sachs et al: Postgrad Med April, 2000.

SOMATOFORM DISORDERS

Clinical Manifestations

Patients with multiple somatic complaints that cannot be explained by a known medical condition or by the effects of alcohol or of recreational or prescription drugs are commonly seen in primary care practice; one survey indicated a prevalence of such complaints of 5%. In *somatization disorder*, the patient presents with multiple physical complaints referable to different organ systems **(Table 391-13)**. Onset is usually before age 30 years, and the disorder is persistent. Formal diagnostic criteria require the recording of at least four pain, two gastrointestinal, one sexual, and one pseudoneurologic symptom. Patients with somatization disorder often present with dramatic complaints, but the complaints are inconsistent. Symptoms of comorbid anxiety and mood disorder are common and may be the result of drug interactions due to regimens initiated independently by different physicians. Patients with somatization disorder may be impulsive and demanding and frequently qualify for a formal comorbid psychiatric diagnosis. In *conversion disorder*, the symptoms focus on deficits that involve motor or sensory function and on psychological factors that initiate or exacerbate the medical presentation. Like somatization disorder, the deficit is not intentionally produced or simulated, as is the case in factitious disorder (malingering). In *hypochondriasis*, the essential feature is a belief of serious medical illness that persists despite reassurance and appropriate medical evaluation. As with somatization disorder, patients with hypochondriasis have a history of poor relationships with physicians stemming from their sense that they have been evaluated and treated inappropriately or inadequately. Hypochondriasis can be disabling in intensity and is persistent, with waxing and waning symptomatology.

Table 391-13 Diagnostic Criteria for Somatization Disorder

A. A history of many physical complaints beginning before age 30 years that occur over a period of several years and result in treatment being sought or significant impairment in social, occupational, or other important areas of functioning.
B. Each of the following criteria must have been met, with individual symptoms occurring at any time during the course of the disturbance: 1. *Four pain symptoms:* a history of pain related to at least four different sites or functions (e.g., head, abdomen, back, joints, extremities, chest, rectum, during menstruation, during sexual intercourse, or during urination) 2. *Two gastrointestinal symptoms:* a history of at least two gastrointestinal symptoms other than pain (e.g., nausea, bloating, vomiting other than during pregnancy, diarrhea, or intolerance of several different foods) 3. *One sexual symptom:* a history of at least one sexual or reproductive symptom other than pain (e.g., sexual indifference, erectile or ejaculatory dysfunction, irregular menses, excessive menstrual bleeding, vomiting throughout pregnancy) 4. *One pseudoneurologic symptom:* a history of at least one symptom or deficit suggesting a neurologic condition not limited to pain (conversion symptoms such as impaired coordination or balance, paralysis or localized weakness, difficulty swallowing or lump in throat, aphonia, urinary retention, hallucinations, loss of touch or pain sensation, double vision, blindness, deafness, seizures; dissociative symptoms such as amnesia; or loss of consciousness other than fainting)
C. Either of the following: 1. After appropriate investigation, each of the symptoms in criterion B cannot be fully explained by a known general medical condition or the direct effects of a substance (e.g., a drug of abuse, a medication) 2. When there is a related general medical condition, the physical complaints or resulting social or occupational impairment are in excess of what would be expected from the history, physical examination, or laboratory findings
D. The symptoms are not intentionally produced or feigned (as in factitious disorder or malingering).

Source: *Diagnostic and Statistical Manual of Mental Disorders* , 4th ed. Washington, DC, American Psychiatric Association, 2000.

In *factitious illnesses*, the patient consciously and voluntarily produces physical symptoms of illness. The term *Munchausen's syndrome* is reserved for individuals with particularly dramatic, chronic, or severe factitious illness. In true factitious illness, the sick role itself is gratifying. A variety of signs, symptoms, and diseases have been either simulated or caused by factitious behavior, the most common including chronic diarrhea, fever of unknown origin, intestinal bleeding or hematuria, seizures,

and hypoglycemia. Factitious disorder is usually not diagnosed until 5–10 years after its onset, and it can produce significant social and medical costs. In *malingering*, the fabrication derives from a desire for some external reward such as a narcotic medication or disability reimbursement.

Treatment: Somatoform Disorders

Patients with somatization disorders are frequently subjected to many diagnostic tests and exploratory surgeries in an attempt to find their "real" illness. Such an approach is doomed to failure and does not address the core issue. Successful treatment is best achieved through behavior modification, in which access to the physician is tightly regulated and adjusted to provide a sustained and predictable level of support that is less clearly contingent on the patient's level of presenting distress. Visits can be brief and should not be associated with a need for a diagnostic or treatment action. Although the literature is limited, some patients with somatization disorder may benefit from antidepressant treatment.

Any attempt to confront the patient usually creates a sense of humiliation and causes the patient to abandon treatment from that caregiver. A better strategy is to introduce psychological causation as one of a number of possible explanations and to include factitious illness as an option in the differential diagnoses that are discussed. Without directly linking psychotherapeutic intervention to the diagnosis, the patient can be offered a face-saving means by which the pathologic relationship with the health care system can be examined and alternative approaches to life stressors developed.

PERSONALITY DISORDERS

CLINICAL MANIFESTATIONS

Personality disorders are characteristic patterns of thinking, feeling, and interpersonal behavior that are relatively inflexible and cause significant functional impairment or subjective distress for the individual. The observed behaviors are not secondary to another mental disorder, nor are they precipitated by substance abuse or a general medical condition. This distinction is often difficult to make in clinical practice, as personality change may be the first sign of serious neurologic, endocrine, or other medical illness. Patients with frontal lobe tumors, for example, can present with changes in motivation and personality while the results of the neurologic examination remain within normal limits. Individuals with personality disorders are often regarded as "difficult patients" in clinical medical practice because they are seen as excessively demanding and/or unwilling to follow recommended treatment plans. Although DSM-IV portrays personality disorders as qualitatively distinct categories, there is an alternative perspective that personality characteristics vary as a continuum between normal functioning and formal mental disorder.

Personality disorders have been grouped into three overlapping clusters. *Cluster A* includes paranoid, schizoid, and schizotypal personality disorders. It includes individuals who are odd and eccentric and who maintain an emotional distance from others. Individuals have a restricted emotional range and remain socially isolated. Patients with schizotypal personality disorder frequently have unusual perceptual experiences and express magical beliefs about the external world. The essential feature of paranoid personality disorder is a pervasive mistrust and suspiciousness of others to an extent that is unjustified by available evidence. *Cluster B* disorders include antisocial, borderline, histrionic, and narcissistic types and describe individuals whose behavior is impulsive, excessively emotional, and erratic. *Cluster C* incorporates avoidant, dependent, and obsessive-compulsive personality types; enduring traits are anxiety and fear. The boundaries between cluster types are to some extent artificial, and many patients who meet criteria for one personality disorder also meet criteria for aspects of another. The risk of a comorbid major mental disorder is increased in patients who qualify for a diagnosis of personality disorder.

Treatment: Personality Disorders

Dialectical behavior therapy (DBT) is a cognitive-behavioral approach that focuses on behavioral change while providing acceptance, compassion, and validation of the patient. Several randomized trials have demonstrated the efficacy of DBT in the treatment of personality disorders. Antidepressant medications and low-dose antipsychotic drugs have some efficacy in cluster A personality disorders, while anticonvulsant mood-stabilizing agents and MAOIs may be considered for patients with cluster B diagnoses who show marked mood reactivity, behavioral dyscontrol, and/or rejection hypersensitivity. Anxious or fearful cluster C patients often respond to medications used for axis I anxiety disorders (see above). It is important that the physician and the patient have reasonable expectations vis-à-vis the possible benefit of any medication used and its side effects. Improvement may be subtle and observable only over time.

SCHIZOPHRENIA

CLINICAL MANIFESTATIONS

Schizophrenia is a heterogeneous syndrome characterized by perturbations of language, perception, thinking, social activity, affect, and volition. There are no pathognomonic features. The syndrome commonly begins in late adolescence, has an insidious (and less commonly acute) onset, and, often, a poor outcome, progressing from social withdrawal and perceptual distortions to recurrent delusions

and hallucinations. Patients may present with positive symptoms (such as conceptual disorganization, delusions, or hallucinations) or negative symptoms (loss of function, anhedonia, decreased emotional expression, impaired concentration, and diminished social engagement) and must have at least two of these for a 1-month period and continuous signs for at least 6 months to meet formal diagnostic criteria. As individuals age, positive psychotic symptoms tend to attenuate and some measure of social and occupational function may be regained. "Negative" symptoms predominate in one-third of the schizophrenic population and are associated with a poor long-term outcome and a poor response to drug treatment. However, marked variability in the course and individual character of symptoms is typical.

The four main subtypes of schizophrenia are catatonic, paranoid, disorganized, and residual. Many individuals have symptoms of more than one type. *Catatonic-type* describes patients whose clinical presentation is dominated by profound changes in motor activity, negativism, and echolalia or echopraxia. *Paranoid-type* describes patients who have a prominent preoccupation with a specific delusional system and who otherwise do not qualify as having *disorganized-type* disease, in which disorganized speech and behavior are accompanied by a superficial or silly affect. In *residual-type* disease, negative symptomatology exists in the absence of delusions, hallucinations, or motor disturbance. The term *schizophreniform disorder* describes patients who meet the symptom requirements but not the duration requirements for schizophrenia, and *schizoaffective disorder* is used for those who manifest symptoms of schizophrenia and independent periods of mood disturbance. Prognosis depends not on symptom severity but on the response to antipsychotic medication. A permanent remission without recurrence does occasionally occur. About 10% of schizophrenic patients commit suicide.

Schizophrenia is present in 0.85% of individuals worldwide, with a lifetime prevalence of ~1–1.5%. An estimated 300,000 episodes of acute schizophrenia occur annually in the United States, resulting in direct and indirect costs of $62.7 billion.

DIFFERENTIAL DIAGNOSIS

The diagnosis is principally one of exclusion, requiring the absence of significant associated mood symptoms, any relevant medical condition, and substance abuse. Drug reactions that cause hallucinations, paranoia, confusion, or bizarre behavior may be dose-related or idiosyncratic; parkinsonian medications, clonidine, quinacrine, and procaine derivatives are the most common prescription medications associated with these symptoms. Drug causes should be ruled out in any case of newly emergent psychosis. The general neurologic examination in patients with schizophrenia is usually normal, but motor rigidity, tremor, and dyskinesias are noted in one-quarter of untreated patients.

EPIDEMIOLOGY AND PATHOPHYSIOLOGY

Epidemiologic surveys identify several risk factors for schizophrenia, including genetic susceptibility, early developmental insults, winter birth, and increasing parental age. Genetic factors are involved in at least a subset of individuals who develop schizophrenia. Schizophrenia is observed in ~6.6% of all first-degree relatives of an affected proband. If both parents are affected, the risk for offspring is 40%. The concordance rate for monozygotic twins is 50%, compared to 10% for dizygotic twins. Schizophrenia-prone families are also at risk for other psychiatric disorders, including schizoaffective disorder and *schizotypal* and *schizoid personality disorders*, the latter terms designating individuals who show a lifetime pattern of social and interpersonal deficits characterized by an inability to form close interpersonal relationships, eccentric behavior, and mild perceptual distortions**. The pathogenesis of schizophrenia is discussed in detail inChap. 390**.

Treatment: Schizophrenia

Antipsychotic agents **(Table 391-14)** are the cornerstone of acute and maintenance treatment of schizophrenia and are effective in the treatment of hallucinations, delusions, and thought disorders, regardless of etiology. The mechanism of action involves, at least in part, binding to dopamine D_2/D3 receptors in the ventral striatum; the clinical potencies of traditional antipsychotic drugs parallel their affinities for the D_2 receptor, and even the newer "atypical" agents exert some degree of D_2 receptor blockade. All neuroleptics induce expression of the immediate-early gene c-*fos* in the nucleus accumbens, a dopaminergic site connecting prefrontal and limbic cortices. The clinical efficacy of newer atypical neuroleptics, however, may involve N-methyl-D-aspartate (NMDA) receptor blockade, α_1- and α_2-noradrenergic activity, altering the relationship between $5HT_2$ and D_2 receptor activity, as well as faster dissociation of D_2 binding and effects on neuroplasticity.

Table 391-14 Antipsychotic Agents

Name	Usual PO Daily Dose.	Side Effects	Sedation	Comments

	mg			
First-Generation Antipsychotics				
Low-potency				
Chlorpromazine (Thorazine)	100–1000	Anticholinergic effects; orthostasis; photosensitivity; cholestasis; QT prolongation	+ + +	EPSEs usually not prominent; can cause anticholinergic delirium in elderly patients
Thioridazine (Mellaril)	100–600			
Midpotency				
Trifluoperazine (Stelazine)	2–50	Fewer anticholinergic side effects	+ +	Well tolerated by most patients
Perphenazine (Trilafon)	4–64	Fewer EPSEs than with higher potency agents.	++	Little weight gain
Loxapine (Loxitane)	30–100	Frequent EPSEs	+ +	
Molindone (Moban)	30–100	Frequent EPSEs	0	
High potency				
Haloperidol (Haldol)	5–20	No anticholinergic side effects; EPSEs often prominent	0/+	Often prescribed in doses that are too high; long-acting injectable forms of haloperidol and fluphenazine available
Fluphenazine (Prolixin)	1–20	Frequent EPSEs	0/+	
Thiothixene (Navane)	2–50	Frequent EPSEs	0/+	
Second-Generation Antipsychotics				
Clozapine (Clozaril)	150–600	Agranulocytosis (1%); weight gain; seizures; drooling; hyperthermia	+ +	Requires weekly WBC count for first 6 months, then biweekly if stable
Risperidone (Risperdal)	2–8	Orthostasis	+	Requires slow titration; EPSEs observed with doses >6 mg qd
Olanzapine (Zyprexa)	10–30	Weight gain	+ +	Mild prolactin elevation
Quetiapine (Seroquel)	350–800	Sedation; weight gain; anxiety	+ + +	Bid dosing
Ziprasidone (Geodon)	120–200	Orthostatic hypotension	+/++	Minimal weight gain; increases QT interval
Aripiprazole (Abilify)	10–30	Nausea, anxiety, insomnia	0/+	Mixed agonist/antagonist
Paliperidone (Invega)	3–12	Restlessness, EPSEs	+	Active metabolite of risperidone
Iloperidone (Fanapt)	12–24	Dizziness, hypotension	0/+	Requires dose titration
Asenapine (Saphris)	10–20	Dizziness, EPSEs, weight gain	++	Sublingual tablets; bid dosing
Lurasidone (Latuda)	40–80	Nausea, EPSEs	++	Uses CYP3A4

Abbreviations: EPSEs, extrapyramidal side effects; WBC, white blood cell.

Conventional neuroleptics differ in their potency and side-effect profile. Older agents such as chlorpromazine and thioridazine, are more sedating and anticholinergic and more likely to cause orthostatic hypotension, while higher potency antipsychotics such as haloperidol, perphenazine, and thiothixene, are more likely to induce extrapyramidal side effects. The model "atypical" antipsychotic agent is *clozapine*, a dibenzodiazepine that has a greater potency in blocking the $5HT_2$ than the D_2 receptor and a much higher affinity for the D_4 than the D_2 receptor. Its principal disadvantage is a risk of blood dyscrasias. Paliperidone is a recently approved agent that is a metabolite of risperidone and shares many of its properties. Unlike other antipsychotics, clozapine does not cause a rise in prolactin level. Approximately 30% of patients who do not benefit from conventional antipsychotic agents will have a better response to this drug, which also has a demonstrated superiority to other antipsychotic agents in preventing suicide; however, its side-effect profile makes it most appropriate for treatment-resistant cases. *Risperidone*, a benzisoxazole derivative, is more potent at $5HT_2$ than D_2 receptor sites, like clozapine, but it also exerts significant α_2 antagonism, a property that may contribute to its

perceived ability to improve mood and increase motor activity. Risperidone is not as effective as clozapine in treatment-resistant cases but does not carry a risk of blood dyscrasias. *Olanzapine* is similar neurochemically to clozapine but has a significant risk of inducing weight gain. *Quetiapine* is distinct in having a weak D_2 effect but potent α_1 and histamine blockade. *Ziprasidone* causes minimal weight gain and is unlikely to increase prolactin but may increase QT prolongation. *Aripiprazole* also has little risk of weight gain or prolactin increase but may increase anxiety, nausea, and insomnia as a result of its partial agonist properties.

Antipsychotic agents are effective in 70% of patients presenting with a first episode. Improvement may be observed within hours or days, but full remission usually requires 6–8 weeks. The choice of agent depends principally on the side effect profile and cost of treatment or on a past personal or family history of a favorable response to the drug in question. Atypical agents appear to be more effective in treating negative symptoms and improving cognitive function. An equivalent treatment response can usually be achieved with relatively low doses of any drug selected (i.e., 4–6 mg/d of haloperidol, 10–15 mg of olanzapine, or 4–6 mg/d of risperidone). Doses in this range result in >80% D_2 receptor blockade, and there is little evidence that higher doses increase either the rapidity or degree of response. Maintenance treatment requires careful attention to the possibility of relapse and monitoring for the development of a movement disorder. Intermittent drug treatment is less effective than regular dosing, but gradual dose reduction is likely to improve social functioning in many schizophrenic patients who have been maintained at high doses. If medications are completely discontinued, however, the relapse rate is 60% within 6 months. Long-acting injectable preparations (risperidone) are considered when noncompliance with oral therapy leads to relapses. In treatment-resistant patients, a transition to clozapine usually results in rapid improvement, but a prolonged delay in response in some cases necessitates a 6- to 9-month trial for maximal benefit to occur.

Antipsychotic medications can cause a broad range of side effects, including lethargy, weight gain, postural hypotension, constipation, and dry mouth. Extrapyramidal symptoms such as dystonia, akathisia, and akinesia are also frequent with first-generation agents and may contribute to poor adherence if not specifically addressed. Anticholinergic and parkinsonian symptoms respond well to trihexyphenidyl, 2 mg bid, or benztropine mesylate, 1–2 mg bid. Akathisia may respond to beta blockers. In rare cases, more serious and occasionally life-threatening side effects may emerge, including hyperprolactinemia, ventricular arrhythmias, gastrointestinal obstruction, retinal pigmentation, obstructive jaundice, and neuroleptic malignant syndrome (characterized by hyperthermia, autonomic dysfunction, muscular rigidity, and elevated creatine phosphokinase levels). The most serious adverse effects of clozapine are agranulocytosis, which has an incidence of 1%, and induction of seizures, which has an incidence of 10%. Weekly white blood cell counts are required, particularly during the first 3 months of treatment.

The risk of type 2 diabetes mellitus appears to be increased in schizophrenia, and second-generation agents as a group produce greater adverse effects on glucose regulation, independent of effects on obesity, than traditional agents. Clozapine, olanzapine, and quetiapine seem more likely to cause hyperglycemia, weight gain, and hypertriglyceridemia than other atypical antipsychotic drugs. Close monitoring of plasma glucose and lipid levels are indicated with the use of these agents.

A serious side effect of long-term use of first generation antipsychotic agents is *tardivedyskinesia*, characterized by repetitive, involuntary, and potentially irreversible movements of the tongue and lips (bucco-linguo-masticatory triad), and, in approximately half of cases, choreoathetosis. Tardive dyskinesia has an incidence of 2–4% per year of exposure, and a prevalence of 20% in chronically treated patients. The prevalence increases with age, total dose, and duration of drug administration. The risk associated with second-generation agents appears to be much lower. The cause may involve formation of free radicals and perhaps mitochondrial energy failure. Vitamin E may reduce abnormal involuntary movements if given early in the syndrome.

The CATIE study, a large-scale investigation of the effectiveness of antipsychotic agents in "real world" patients, revealed a high rate of discontinuation of treatment over 18 months. Olanzapine showed greater effectiveness than quetiapine, risperidone, perphenazine, or ziprasidone but also a higher discontinuation rate due to weight gain and metabolic effects. Surprisingly, perphenazine, a first-generation agent, showed little evidence of inferiority to newer drugs.

Drug treatment of schizophrenia is by itself insufficient. Educational efforts directed toward families and relevant community resources have proved to be necessary to maintain stability and optimize outcome. A treatment model involving a multidisciplinary case-management team that seeks out and closely follows the patient in the community has proved particularly effective.

ASSESSMENT AND EVALUATION OF VIOLENCE

Primary care physicians may encounter situations in which family, domestic, or societal violence is discovered or suspected. Such an awareness can carry legal and moral obligations; many state laws mandate reporting of child, spousal, and elder abuse. Physicians are frequently the first point of contact for both victim and abuser. Approximately 2 million older Americans and 1.5 million U.S. children are thought to experience some form of physical maltreatment each year. Spousal abuse is

thought to be even more prevalent. An interview study of 24,000 women in 10 countries found a lifetime prevalence of physical or sexual violence that ranged from 15 to 71%; these individuals are more likely to suffer from depression, anxiety, somatization disorder, and substance abuse and to have attempted suicide. In addition, abused individuals frequently express low self-esteem, vague somatic symptomatology, social isolation, and a passive feeling of loss of control. Although it is essential to treat these elements in the victim, the first obligation is to ensure that the perpetrator has taken responsibility for preventing any further violence. Substance abuse and/or dependence and serious mental illness in the abuser may contribute to the risk of harm and require direct intervention. Depending on the situation, law enforcement agencies, community resources such as support groups and shelters, and individual and family counseling can be appropriate components of a treatment plan. A safety plan should be formulated with the victim, in addition to providing information about abuse, its likelihood of recurrence, and its tendency to increase in severity and frequency. Antianxiety and antidepressant medications may sometimes be useful in treating the acute symptoms, but only if independent evidence for an appropriate psychiatric diagnosis exists.

MENTAL HEALTH PROBLEMS IN THE HOMELESS

There is a high prevalence of mental disorders and substance abuse among homeless and impoverished individuals. Depending on the definition used, estimates of the total number of homeless individuals in the United States range from 800,000 to 2 million, one-third of whom qualify as having a serious mental disorder. Poor hygiene and nutrition, substance abuse, psychiatric illness, physical trauma, and exposure to the elements combine to make the provision of medical care challenging. Only a minority of these individuals receive formal mental health care; the main points of contact are outpatient medical clinics and emergency departments. Primary care settings represent a critical site in which housing needs, treatment of substance dependence, and evaluation and treatment of psychiatric illness can most efficiently take place. Successful intervention is dependent on breaking down traditional administrative barriers to health care and recognizing the physical constraints and emotional costs imposed by homelessness. Simplifying health care instructions and follow-up, allowing frequent visits, and dispensing medications in limited amounts that require ongoing contact are possible techniques for establishing a successful therapeutic relationship.

FURTHER READINGS

Allgulander C: Generalized anxiety disorder: Between now and DSM-V. Psychiatr Clin North Am 32:611, 2009[PMID: 19716993] [Full Text]

American Psychiatric Association: American Psychiatric Association Practice Guidelines for the Treatment of Psychiatric Disorders: *http://www.Psychiatryonline.com/Pracguide*

Buchanan RW et al: Schizophrenia Patient Outcomes Research Team (PORT): The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements. Schizophr Bull 1:71, 2010

Figueredo VM: The time has come for physicians to take notice: The impact of psychosocial stressors on the heart. Am J Med 8:704, 2009

Frye MA: Bipolar disorder—A focus on depression. N Engl J Med 364:51, 2011[PMID: 21208108] [Full Text]

Grunze H et al: The World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for the Biological Treatment of Bipolar Disorders: Update 2010 on the treatment of acute bipolar depression. World J Biol Psychiatry 2:81, 2010

Laugharne J et al: Role of psychological trauma in the cause and treatment of anxiety and depressive disorders. Curr Opin Psychiatry 1:25, 2010

Nutt DJ et al: International consensus statement on major depressive disorder. J Clin Psychiatry 71 Suppl E1:e08, 2010

Ravindran LN: Pharmacotherapy of PTSD: Premises, principles, and priorities. Brain Res 1293:24, 2009[PMID: 19332035] [Full Text]

Reid S, Barbui C: Long term treatment of depression with selective serotonin reuptake inhibitors and newer antidepressants. BMJ 340:c1468. doi:10.1136/bmj.c1468, 2010

Schanzer B et al: Homelessness, health status, and health care use. Am J Public Health 97:464, 2007[PMID: 17267724] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

Alcohol (beverage ethanol) distributes throughout the body, affecting almost all systems and altering nearly every neurochemical process in the brain. This drug is likely to exacerbate most medical conditions, affect almost any medication metabolized in the liver, and temporarily mimic many medical (e.g., diabetes) and psychiatric (e.g., depression) conditions. Because ~80% of people in Western countries have consumed alcohol, and two-thirds have been drunk in the prior year, the lifetime risk for serious, repetitive alcohol problems is almost 20% for men and 10% for women, regardless of a person's education or income. While low doses of alcohol have some healthful benefits, the intake of more than three standard drinks per day on a regular basis enhances the risk for cancer and vascular disease, and alcohol use disorders decrease the life span by about 10 years. Unfortunately, most clinicians have had only limited education regarding these conditions. This chapter presents a brief overview of clinically useful information about alcohol use, abuse, and dependence.

Pharmacology and Nutritional Impact of Ethanol

Blood levels of ethanol are expressed as milligrams or grams of ethanol per deciliter (e.g., 100 mg/dL = 0.10 g/dL), with values of ~0.02 g/dL resulting from the ingestion of one typical drink. In round figures, 340 mL (12 oz) of beer, 115 mL (4 oz) of nonfortified wine, and 43 mL (1.5 oz) (a shot) of 80-proof beverage such as whisky, gin, or vodka each contain ~10–15 g of ethanol and represent a standard drink;0.5 L (1 pint) of 80-proof beverage contains ~160 g (about 16 standard drinks), and 750 mL of wine contains ~60 g of ethanol. These beverages also have additional components known as *congeners* that affect the drink's taste and might contribute to adverse effects on the body. Congeners include methanol, butanol, acetaldehyde, histamine, tannins, iron, and lead. Alcohol acutely decreases neuronal activity and has similar behavioral effects and cross-tolerance with other depressants, including benzodiazepines and barbiturates.

Alcohol is absorbed from mucous membranes of the mouth and esophagus (in small amounts), from the stomach and large bowel (in modest amounts), and from the proximal portion of the small intestine (the major site). The rate of absorption is increased by rapid gastric emptying (as can be induced by carbonated beverages); by the absence of proteins, fats, or carbohydrates (which interfere with absorption); and by dilution to a modest percentage of ethanol (maximum at ~20% by volume).

Between 2% (at low blood alcohol concentrations) and 10% (at high blood alcohol concentrations) of ethanol is excreted directly through the lungs, urine, or sweat, but the greater part is metabolized to acetaldehyde, primarily in the liver. The most important pathway occurs in the cell cytosol where alcohol dehydrogenase (ADH) produces acetaldehyde, which is then rapidly destroyed by aldehyde dehydrogenase (ALDH) in the cytosol and mitochondria **(Fig. 392-1)**. A second pathway in the microsomes of the smooth endoplasmic reticulum (the microsomal ethanol-oxidizing system, or MEOS) is responsible for ≥10% of ethanol oxidation at high blood alcohol concentrations.

Figure 392-1

The metabolism of alcohol. MEOS, microsomal ethanol-oxidizing system.

While alcohol supplies calories (a drink contains ~300 kJ, or 70–100 kcal), these are devoid of nutrients such as minerals, proteins, and vitamins. In addition, alcohol can also interfere with absorption of vitamins in the small intestine and decreases their storage in the liver with modest effects on folate (folacin or folic acid), pyridoxine (B_6), thiamine (B_1), nicotinic acid (niacin, B_3), and vitamin A.

A heavy ethanol load in a fasting, healthy individual is likely to produce transient hypoglycemia within 6–36 h, secondary to the acute actions of ethanol on gluconeogenesis. This can result in temporary abnormal glucose tolerance tests (with a resulting erroneous diagnosis of diabetes mellitus) until the alcoholic has abstained for 2–4 weeks. Alcohol ketoacidosis, probably reflecting a decrease in fatty acid oxidation coupled with poor diet or recurrent vomiting, can be misdiagnosed as diabetic ketosis. With the former, patients show an increase in serum ketones along with a mild increase in glucose but a large anion gap, a mild to moderate increase in serum lactate, and a β-hydroxybutyrate/lactate ratio of between 2:1 and 9:1 (with normal being 1:1).

In the brain, alcohol affects almost all neurotransmitter systems, with acute actions that are often the opposite of those seen following desistance after a period of heavy drinking. The most prominent

actions relate to boosting gamma aminobutyric acid (GABA) activity, especially in $GABA_A$ receptors. Enhancement of this complex chloride channel system contributes to anticonvulsant, sleep-inducing, antianxiety, and muscle relaxation effects of all GABA-boosting drugs. Acutely administered alcohol produces a release of GABA, and continued use of this drug increases density of $GABA_A$ receptors, while alcohol withdrawal states are characterized by decreases in GABA-related activity. Equally important is the ability of acute alcohol to inhibit postsynaptic *N*-methyl-d-aspartate (NMDA) excitatory glutamate receptors, while chronic drinking and desistance are associated with an upregulation of these excitatory receptor subunits. The relationships between greater GABA and diminished NMDA receptor activity during acute intoxication and diminished GABA with enhanced NMDA actions during alcohol withdrawal explain much of intoxication and withdrawal phenomena.

As with all pleasurable activities, drinking alcohol acutely increases dopamine levels in the brain, especially in the ventral tegmentum and related brain regions, and this effect plays an important role in continued alcohol use, craving, and relapse. The changes in dopamine pathways are also linked to increases in "stress hormones," including cortisol and adrenocorticotropic hormone (ACTH) during intoxication and decreases in these hormones during withdrawal. Such alterations are likely to contribute to both feelings of reward during intoxication and depression during falling blood alcohol concentrations. Also closely linked to alterations in dopamine (especially in the nucleus accumbens) are alcohol-induced changes in opioid receptors, with acute alcohol also causing release of beta endorphins.

Additional important neurochemical changes include increases in synaptic levels of serotonin during acute intoxication, and subsequent upregulation of serotonin receptors. Acute increases in nicotinic acetylcholine systems also contribute to the impact of alcohol in the ventral tegmental region, which occur in concert with enhanced dopamine activity. In the same regions, alcohol impacts on cannabinol receptors, with resulting release of dopamine, GABA, and glutamate as well as subsequent effects on brain reward circuits.

Behavioral Effects, Tolerance, and Dependence

The acute effects of a drug depend on the dose, the rate of increase in plasma, the concomitant presence of other drugs, and the past experience with the agent. "Legal intoxication" with alcohol in most states requires a blood alcohol concentration of 0.08 g/dL, while levels of 0.04 or even lower are cited in other countries. However, behavioral, psychomotor, and cognitive changes are seen at levels as low as 0.02–0.03 g/dL (i.e., after one to two drinks) **(Table 392-1)**. Deep but disturbed sleep can be seen at twice the legal intoxication level, and death can occur with levels between 0.30 and 0.40 g/dL. Beverage alcohol is probably responsible for more overdose deaths than any other drug.

Table 392-1 Effects of Blood Alcohol Levels in the Absence of Tolerance

Blood Level, g/dL	Usual Effect
0.02	Decreased inhibitions, a slight feeling of intoxication
0.08	Decrease in complex cognitive functions and motor performance
0.20	Obvious slurred speech, motor incoordination, irritability, and poor judgment
0.30	Light coma and depressed vital signs
0.40	Death

Repeated use of alcohol contributes to acquired tolerance, a complex phenomenon involving at least three types of compensatory mechanisms. (1) After 1–2 weeks of daily drinking, *metabolic or pharmacokinetic tolerance* can be seen, with up to 30% increase in the rate of hepatic ethanol metabolism. This alteration disappears almost as rapidly as it develops. (2) *Cellular or pharmacodynamic tolerance* develops through neurochemical changes that maintain relatively normal physiologic functioning despite the presence of alcohol. Subsequent decreases in blood levels contribute to symptoms of withdrawal. (3) Individuals learn to adapt their behavior so that they can function better than expected under influence of the drug (*learned or behavioral tolerance*).

The cellular changes caused by chronic ethanol exposure may not resolve for several weeks or longer following cessation of drinking. Rapid decreases in blood alcohol levels before that time can result in a withdrawal syndrome, which is most intense during the first 5 days, but some symptoms (e.g., disturbed sleep and anxiety) can take up to 4–6 months to resolve.

THE EFFECTS OF ETHANOL ON ORGAN SYSTEMS

Relatively low doses of alcohol (one or two drinks per day) have potential beneficial effects of increasing high-density lipoprotein cholesterol and decreasing aggregation of platelets, with a resulting decrease in risk for occlusive coronary disease and embolic strokes. Red wine has additional potential health-promoting qualities at relatively low doses due to flavinols and related substances, which may work by inhibiting platelet activation. Modest drinking might also decrease the risk for vascular

dementia and, possibly, Alzheimer's disease. However, any potential healthful effects disappear with the regular consumption of three or more drinks per day, and knowledge about the deleterious effects of alcohol can both help the physician to identify patients with alcohol abuse and dependence, and to supply them with information that might help motivate a change in behavior.

Nervous System

Approximately 35% of drinkers (and a much higher proportion of alcoholics) experience a *blackout*, an episode of temporary anterograde amnesia, in which the person forgets all or part of what occurred during a drinking evening. Another common problem, one seen after as few as one or two drinks shortly before bedtime, is disturbed sleep. Although alcohol might initially help a person to fall asleep, it disrupts sleep throughout the rest of the night. The stages of sleep are also altered, and time spent in rapid eye movement (REM) and deep sleep is reduced. Alcohol relaxes muscles in the pharynx, which can cause snoring and exacerbate sleep apnea; symptoms of the latter occur in 75% of alcoholic men older than age 60 years. Patients may also experience prominent and sometimes disturbing dreams. All of these sleep problems are more pronounced in alcoholics, and their persistence may contribute to relapse.

Another common consequence of alcohol use is impaired judgment and coordination, increasing the risk of accidents and injury. In the United States, ~40% of drinkers have at some time driven while intoxicated. Heavy drinking can also be associated with headache, thirst, nausea, vomiting, and fatigue the following day, a hangover syndrome that is responsible for much missed time and temporary cognitive deficits at work and school.

The effect of alcohol on the *nervous system* is even more pronounced among alcohol-dependent individuals. Chronic high doses cause *peripheral neuropathy* in ~10% of alcoholics: similar to diabetes, patients experience bilateral limb numbness, tingling, and paresthesias, all of which are more pronounced distally. Approximately 1% of alcoholics develop *cerebellar degeneration or atrophy* . This is a syndrome of progressive unsteady stance and gait often accompanied by mild nystagmus; neuroimaging studies reveal atrophy of the cerebellar vermis. Fortunately, very few alcoholics (perhaps as few as 1 in 500 for the full syndrome) develop *Wernicke's* (ophthalmoparesis, ataxia, and encephalopathy) and *Korsakoff's* (retrograde and anterograde amnesia) *syndromes*, although a higher proportion have one or more neuropathologic findings related to these syndromes. These occur as the result of low levels of thiamine, especially in predisposed individuals, e.g., those with transketolase deficiency. Alcoholics can manifest *cognitive problems* and temporary memory impairment lasting for weeks to months after drinking very heavily for days or weeks. Brain atrophy, evident as ventricular enlargement and widened cortical sulci on MRI and CT scans, occurs in ~50% of chronic alcoholics; these changes are usually reversible if abstinence is maintained. There is no single alcoholic dementia syndrome; rather, this label is used to describe patients who have apparently irreversible cognitive changes (possibly from diverse causes) in the context of chronic alcoholism.

PSYCHIATRIC COMORBIDITY

As many as two-thirds of alcohol-dependent individuals meet the criteria for a psychiatric syndrome in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) of the American Psychiatric Association (Chap. 391). Half of these relate to a preexisting antisocial personality manifesting as impulsivity and disinhibition that contribute to both alcohol and drug dependence. The lifetime risk is 3% in males, and ≳80% of such individuals demonstrate alcohol and/or drug dependence. Another common comorbidity occurs with dependence on illicit substances. The remainder of alcoholics with psychiatric syndromes have preexisting conditions such as schizophrenia or manic-depressive disease and anxiety disorders such as panic disorder. The comorbidities of alcoholism with independent psychiatric disorders might represent an overlap in genetic vulnerabilities, impaired judgment in the use of alcohol from the independent psychiatric condition, or an attempt to use alcohol to alleviate some of the symptoms of the disorder or side effects of medications.

Many psychiatric syndromes can be seen *temporarily* during heavy drinking and subsequent withdrawal. These alcohol-induced conditions include an intense *sadness* lasting for days to weeks in the midst of heavy drinking seen in 40% of alcoholics, which tends to disappear over several weeks of abstinence (alcohol-induced mood disorder); temporary severe *anxiety* in 10–30% of alcoholics, often beginning during alcohol withdrawal, which can persist for a month or more after cessation of drinking (alcohol-induced anxiety disorder); and auditory *hallucinations* and/or paranoid delusions in a person who is alert and oriented, seen in 3–5% of alcoholics (*alcohol-induced psychotic disorder*).

Treatment of all forms of alcohol-induced psychopathology includes helping patients achieve abstinence and offering supportive care, as well as reassurance and "talk therapy" such as cognitive-behavioral approaches. However, with the exception of short-term antipsychotics or similar drugs for substance-induced psychoses, substance-induced psychiatric conditions only rarely require medications. Recovery is likely within several days to 4 weeks of abstinence. Conversely, because alcohol-induced conditions are temporary and do not indicate a need for long-term pharmacotherapy, a history of alcohol intake is an important part of the workup for any patient with one of these psychiatric

symptoms.

The Gastrointestinal System

ESOPHAGUS AND STOMACH

Alcohol intake can result in inflammation of the esophagus and stomach causing epigastric distress and gastrointestinal bleeding, making alcohol one of the most common causes of hemorrhagic gastritis. Violent vomiting can produce severe bleeding through a Mallory-Weiss lesion, a longitudinal tear in the mucosa at the gastroesophageal junction.

PANCREAS AND LIVER

The incidence of acute pancreatitis (~25 per 1000 per year) is almost threefold higher in alcoholics than in the general population, accounting for an estimated 10% or more of the total cases. Alcohol impairs gluconeogenesis in the liver, resulting in a fall in the amount of glucose produced from glycogen, increased lactate production, and decreased oxidation of fatty acids. This contributes to an increase in fat accumulation in liver cells. In healthy individuals these changes are reversible, but with repeated exposure to ethanol, especially daily heavy drinking, more severe changes in the liver occur, including alcohol-induced hepatitis, perivenular sclerosis, and cirrhosis, with the latter observed in an estimated 15% of alcoholics (Chap. 307). Perhaps through an enhanced vulnerability to infections, alcoholics have an elevated rate of hepatitis C, and drinking in the context of that disease is associated with more severe liver deterioration.

Cancer

As few as 1.5 drinks per day increases a woman's risk of breast cancer 1.4-fold. For both genders, four drinks per day increases the risk for oral and esophageal cancers approximately threefold and rectal cancers by a factor of 1.5; seven to eight or more drinks per day enhances approximately fivefold the risks for many cancers. These consequences may result directly from cancer-promoting effects of alcohol and acetaldehyde or indirectly by interfering with immune homeostasis.

Hematopoietic System

Ethanol causes an increase in red blood cell size [mean corpuscular volume (MCV)], which reflects its effects on stem cells. If heavy drinking is accompanied by folic acid deficiency, there can also be hypersegmented neutrophils, reticulocytopenia, and a hyperplastic bone marrow; if malnutrition is present, sideroblastic changes can be observed. Chronic heavy drinking can decrease production of white blood cells, decrease granulocyte mobility and adherence, and impair delayed-hypersensitivity responses to novel antigens (with a possible false-negativetuberculin skin test). Associated immune deficiencies can contribute to vulnerability toward infections, including hepatitis and HIV, and interfere with their treatment. Finally, many alcoholics have mild thrombocytopenia, which usually resolves within a week of abstinence unless there is hepatic cirrhosis or congestive splenomegaly.

Cardiovascular System

Acutely, ethanol decreases myocardial contractility and causes peripheral vasodilation, with a resulting mild decrease in blood pressure and a compensatory increase in cardiac output. Exercise-induced increases in cardiac oxygen consumption are higher after alcohol intake. These acute effects have little clinical significance for the average healthy drinker but can be problematic when persisting cardiac disease is present.

The consumption of three or more drinks per day results in a dose-dependent increase in blood pressure, which returns to normal within weeks of abstinence. Thus, heavy drinking is an important factor in mild to moderate hypertension. Chronic heavy drinkers also have a sixfold increased risk for coronary artery disease, related, in part, to increased low-density lipoprotein cholesterol, and carry an increased risk for cardiomyopathy through direct effects of alcohol on heart muscle. Symptoms of the latter include unexplained arrhythmias in the presence of left ventricular impairment, heart failure, hypocontractility of heart muscle, and dilation of all four heart chambers with associated mural thrombi and mitral valve regurgitation. Atrial or ventricular arrhythmias, especially paroxysmal tachycardia, can also occur temporarily after heavy drinking in individuals showing no other evidence of heart disease —a syndrome known as the "holiday heart."

Genitourinary System Changes, Sexual Functioning, and Fetal Development

Drinking in adolescence can affect normal sexual development and reproductive onset. At any age, modest ethanol doses (e.g., blood alcohol concentrations of 0.06 gm/dL) can increase sexual drive but also decrease erectile capacity in men. Even in the absence of liver impairment, a significant minority of chronic alcoholic men show irreversible testicular atrophy with shrinkage of the seminiferous tubules, decreases in ejaculate volume, and a lower sperm count (Chap. 346).

The repeated ingestion of high doses of ethanol by women can result in amenorrhea, a decrease in ovarian size, absence of corpora lutea with associated infertility, and an increased risk of spontaneous abortion. Heavy drinking during pregnancy results in the rapid placental transfer of both ethanol and acetaldehyde, which may have serious consequences for fetal development. One severe result is the *fetal alcohol syndrome* (FAS), seen in ~5% of children born to heavy-drinking mothers, which can

include any of the following: facial changes with epicanthal eye folds; poorly formed ear concha; small teeth with faulty enamel; cardiac atrial or ventricular septal defects; an aberrant palmar crease and limitation in joint movement; and microcephaly with mental retardation. A less severe condition is the *fetal alcohol spectrum disorder* (FASD), which can include low birth weight, a lower IQ, hyperactive behavior, and some modest cognitive deficits. The amount of ethanol required and the time of vulnerability during pregnancy have not been defined, making it advisable for pregnant women to abstain completely.

Other Effects

Between one-half and two-thirds of alcoholics have skeletal muscle weakness caused by acute *alcoholic myopathy*, a condition that improves but which might not fully remit with abstinence. Effects of repeated heavy drinking on the *skeletal system* include changes in calcium metabolism, lower bone density, and decreased growth in the epiphyses, leading to an increased risk for fractures and osteonecrosis of the femoral head. *Hormonal changes* include an increase in cortisol levels, which can remain elevated during heavy drinking; inhibition of vasopressin secretion at rising blood alcohol concentrations and enhanced secretion at falling blood alcohol concentrations (with the final result that most alcoholics are likely to be slightly overhydrated); a modest and reversible decrease in serum thyroxine (T4); and a more marked decrease in serum triiodothyronine (T_3). Hormone irregularities should be reevaluated as they may disappear after a month of abstinence.

ALCOHOLISM (ALCOHOL ABUSE OR DEPENDENCE)

Because many drinkers occasionally imbibe to excess, temporary alcohol-related pathology is common in nonalcoholics, especially in the late teens to the late twenties. When repeated problems in multiple life areas develop, the individual is likely to meet criteria for alcohol abuse or dependence.

Definitions and Epidemiology

Alcohol dependence is defined in DSM-IV as repeated alcohol-related difficulties in at least three of seven life areas that cluster together at about the same time (e.g., over the same 12-month period). Two of these seven items, tolerance and withdrawal, may have special importance as they are associated with a more severe clinical course. Dependence predicts a course of recurrent problems with the use of alcohol and the consequent shortening of the life span by a decade.

Alcohol abuse is defined as repetitive problems with alcohol in any one of four life areas—social, interpersonal, legal, and occupational—or repeated use in hazardous situations such as driving while intoxicated in an individual who is not alcohol dependent. About 50% of those with alcohol abuse continue to have alcohol problems 2–5 years later, but only ~10% of these patients—including adolescents—go on to develop alcohol dependence.

The lifetime risk for alcohol dependence in most Western countries is about 10–15% for men and 5–8% for women. Rates are generally similar in the United States, Canada, Germany, Australia, and the United Kingdom; tend to be lower in most Mediterranean countries, such as Italy, Greece, and Israel; and may be higher in Ireland, France, and Scandinavia. An even higher lifetime prevalence has been reported for most native cultures, including American Indians, Eskimos, Maori groups, and aboriginal tribes of Australia. These differences reflect both cultural and genetic influences, as described below. In Western countries, the typical alcoholic is more often a blue- or white-collar worker or homemaker. The lifetime risk for alcoholism among physicians is similar to that of the general population.

Genetics

Approximately 60% of the risk for alcohol use disorders is attributed to genes, as indicated by the fourfold higher risk for alcohol abuse and dependence in children of alcoholics (even if these children were adopted early in life and raised by nonalcoholics) and a higher risk in identical twins as compared to fraternal twins of alcoholics. The genetic variations appear to operate primarily through intermediate characteristics that subsequently relate to the environment in altering the risk for heavy drinking and alcohol problems. These include genes relating to a high risk for all substance use disorders that operate through impulsivity, schizophrenia, and bipolar disorder. Another characteristic, an intense flushing response when drinking, decreases the risk for only alcohol use disorders through gene variations for several alcohol-metabolizing enzymes, especially aldehyde dehydrogenase (a mutation only seen in Asians), and to a lesser extent, variations in alcohol dehydrogenase.

An additional genetically influenced characteristic, a low sensitivity to alcohol, affects the risk for heavy drinking and may operate, in part, through variations in genes relating to potassium channels, GABA, nicotinic, and serotonin systems. A low response per drink is observed early in the drinking career and before alcohol use disorders have developed. All follow-up studies have demonstrated that this need for higher doses of alcohol to achieve desired effects predicts future heavy drinking, alcohol problems, and alcohol use disorders. The impact of a low response to alcohol on adverse drinking outcomes is mediated, at least in part, by a range of environmental influences, including the selection of heavier-drinking friends, the development of more positive expectations of the effects of high doses of alcohol, and suboptimal ways of coping with stress.

Natural History

Although the age of the first drink (~15 years) is similar in most alcoholics and nonalcoholics, a slightly earlier onset of regular drinking and drunkenness, especially in the context of conduct problems, is associated with a higher risk for later alcohol use disorders. By the early to midtwenties, most nonalcoholic men and women moderate their drinking (perhaps learning from more minor problems), whereas alcoholics are likely to escalate their patterns of drinking despite difficulties. The first major life problem from alcohol often appears in the late teens to early twenties, and a pattern of multiple alcohol difficulties by the midtwenties. Once established, the course of alcoholism is likely to be one of exacerbations and remissions, with little difficulty in temporarily stopping or controlling alcohol use when problems develop, but without help, desistance usually gives way to escalations in alcohol intake and subsequent problems. Following treatment, between half and two-thirds of alcoholics maintain abstinence for years, and often permanently. Even without formal treatment or self-help groups there is also at least a 20% chance of spontaneous remission with long-term abstinence. However, should the alcoholic continue to drink, the life span is shortened by ~10 years on average, with the leading causes of death, heart disease, cancer, accidents, and suicide.

Treatment

The approach to treating alcohol-related conditions is relatively straightforward: (1) recognize that at least 20% of all patients have alcohol abuse or dependence; (2) learn how to identify and treat acute alcohol-related conditions; (3) know how to help patients begin to address their alcohol problems; and (4) know enough about treating alcoholism to appropriately refer patients for additional help.

Identification of the Alcoholic

Even in affluent locales, ~20% of patients have an alcohol use disorder. These men and women can be identified by asking questions about *alcohol problems* and noting laboratory test results that are likely to be abnormal in the context of regular consumption of six to eight or more drinks per day. The two blood tests with ≥60% sensitivity and specificity for heavy alcohol consumption are γ-glutamyl transferase (GGT) (>35 U) and carbohydrate-deficient transferrin (CDT) (>20 U/L or >2.6%); the combination of the two is likely to be more accurate than either alone. The values for these serologic markers are likely to return toward normal within several weeks of abstinence. Other useful blood tests include high-normal MCVs (≥91 μm^3) and serum uric acid (>416 mol/L, or 7 mg/dL).

The diagnosis of alcohol abuse or dependence ultimately rests on the documentation of a pattern of repeated difficulties associated with alcohol use. Thus, in screening it is important to probe for marital or job problems, legal difficulties, histories of accidents, medical problems, evidence of tolerance, etc., and then attempt to tie in use of alcohol or another substance. Some standardized questionnaires can be helpful, including the 10-item Alcohol Use Disorders Identification Test (AUDIT) **(Table 392-2)**, but these are only screening tools, and a face-to-face interview is still required for a meaningful diagnosis.

Table 392-2 The Alcohol Use Disorders Identification Test (Audit)[a]

Item	5-Point Scale (Least to Most)
1. How often do you have a drink containing alcohol?	Never (0) to 4+ per week (4)
2. How many drinks containing alcohol do you have on a typical day?	1 or 2 (0) to 10+ (4)
3. How often do you have six or more drinks on one occasion?	Never (0) to daily or almost daily (4)
4. How often during the last year have you found that you were not able to stop drinking once you had started?	Never (0) to daily or almost daily (4)
5. How often during the last year have you failed to do what was normally expected from you because of drinking?	Never (0) to daily or almost daily (4)
6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?	Never (0) to daily or almost daily (4)
7. How often during the last year have you had a feeling of guilt or remorse after drinking?	Never (0) to daily or almost daily (4)
8. How often during the last year have you been unable to remember what happened the night before because you had been drinking?	Never (0) to daily or almost daily (4)
9. Have you or someone else been injured as a result of your drinking?	No (0) to yes, during the last year (4)
10. Has a relative, friend, doctor or other health worker been concerned about your drinking or suggested that you should cut down?	No (0) to yes, during the last year (4)

[a]The AUDIT is scored by simply summing the values associated with the endorsed response. ***Source:***

Adapted from DF Reinert, GP Allen: *Alcoholism: Clinical & Experimental Research* 26:272, 2002, and from MA Schuckit, 2006.

Treatment Alcohol-Related Conditions

ACUTE INTOXICATION

The first priority in treating severe intoxication is to assess vital signs and manage respiratory depression, cardiac arrhythmia, or blood pressure instability, if present. The possibility of intoxication with other drugs should be considered by obtaining toxicology screens for opioids or other CNS depressants such as benzodiazepines. Aggressive behavior should be handled by offering reassurance but also by considering the possibility of a show of force with an intervention team. If the aggressive behavior continues, relatively low doses of a short-acting benzodiazepine such as lorazepam (e.g., 1–2 mg PO or IV) may be used and can be repeated as needed, but care must be taken not to destabilize vital signs or worsen confusion. An alternative approach is to use an antipsychotic medication (e.g., 0.5–5 mg of haloperidol PO or IM every 4–8 h as needed, or olanzapine 2.5–10 mg IM repeated at 2 and 6 h, if needed).

INTERVENTION

There are two main elements to intervention in a person with alcoholism: motivational interviewing and brief interventions. During motivational interviewing, the clinician helps the patient to think through the assets (e.g., comfort in social situations) and liabilities (e.g., health and interpersonal related problems) of the current pattern of drinking. The patient's responses are key, and the clinician should listen empathetically, helping to weigh options and encouraging the patient to take responsibility for changes that need to be made. Patients should be reminded that only they can decide to avoid the consequences that will occur without changes in drinking. The process of motivational interviewing has been summarized by the acronym FRAMES: Feedback to the patient; Responsibility to be taken by the patient; Advice, rather than orders, on what needs to be done; Menus of options that might be considered; Empathy for understanding of the patient's thoughts and feelings; and Self-efficacy, i.e., offering support for the capacity of the patient to succeed in making changes.

Once the patient begins to consider change, the emphasis shifts to brief interventions designed to help the patient understand more about potential action. Discussions focus on consequences of high alcohol consumption, suggested approaches to stopping drinking, and help in recognizing and avoiding situations likely to lead to heavy drinking. Both motivational interviewing and brief interventions can be carried out in 15-min sessions, but because patients do not always change behavior right away, multiple meetings are often required to explain the problem, discuss optimal treatments, and explain the benefits of abstinence.

ALCOHOL WITHDRAWAL

If the patient agrees to stop drinking, sudden decreases in alcohol intake can produce withdrawal symptoms, many of which are the opposite of those produced by intoxication. Features include tremor of the hands (shakes); agitation and anxiety; autonomic nervous system overactivity including an increase in pulse, respiratory rate, and body temperature; and insomnia. These symptoms usually begin within 5–10 h of decreasing ethanol intake, peak on day 2 or 3, and improve by day 4 or 5, although mild levels of these problems may persist for 4–6 months as a protracted abstinence syndrome.

About 2–5% of alcoholics experience a withdrawal seizure, with the risk increasing in the context of concomitant medical problems, misuse of additional drugs, and higher alcohol quantities. The same risk factors also contribute to a similar rate of *delirium tremens* (DTs), where the withdrawal includes delirium (mental confusion, agitation, and fluctuating levels of consciousness) associated with a tremor and autonomic overactivity (e.g., marked increases in pulse, blood pressure, and respirations). The risks for seizures and DTs can be diminished by identifying and treating any underlying medical conditions early in the course of withdrawal.

The first step in treating withdrawal is to perform a thorough physical examination in all alcoholics who are considering stopping drinking, including a search for evidence of liver failure, gastrointestinal bleeding, cardiac arrhythmia, infection, and glucose or electrolyte imbalance. It is also important to offer adequate nutrition and oral multiple B vitamins, including 50–100 mg of thiamine daily for a week or more. Because most alcoholics who enter withdrawal are either normally hydrated or mildly overhydrated, IV fluids should be avoided unless there is a relevant medical problem or significant recent bleeding, vomiting, or diarrhea.

The next step is to recognize that because withdrawal symptoms reflect the rapid removal of a CNS depressant, alcohol, the symptoms can be controlled by administering any depressant in doses that decrease the agitation and then gradually tapering the dose over 3–5 days. While most CNS depressants are effective, benzodiazepines (Chap. 391) have the highest margin of safety and lowest cost and are, therefore, the preferred class of drugs. Short-half-life benzodiazepines can be considered for patients with serious liver impairment or evidence of brain damage, but they must be given every 4 h to avoid abrupt blood-level fluctuations that may increase the risk for seizures. Therefore, most clinicians use drugs with longer half-lives (e.g., chlordiazepoxide), adjusting the dose

if signs of withdrawal escalate, and withholding the drug if the patient is sleeping or has evidence of orthostatic hypotension. The average patient requires doses of 25–50 mg of chlordiazepoxide or 10 mg of diazepam given PO every 4–6 h on the first day, with doses then decreased to zero over the next 5 days. While alcohol withdrawal can be treated in a hospital, patients in good physical condition who demonstrate mild signs of withdrawal despite low blood alcohol concentrations and who have no prior history of DTs or withdrawal seizures can be considered for outpatient detoxification. These patients should return daily for evaluation of vital signs and can be hospitalized if signs and symptoms of withdrawal escalate.

Treatment of the patient with DTs can be challenging, and the condition is likely to run a course of 3–5 days regardless of the therapy employed. The focus of care is to identify and correct medical problems and to control behavior and prevent injuries. Many clinicians recommend the use of high doses of a benzodiazepine (as much as 800 mg/d of chlordiazepoxide has been reported), a treatment that will decrease agitation and raise the seizure threshold but probably does little to improve the confusion. Other clinicians recommend the use of antipsychotic medications, such as haloperidol or olanzapine as discussed above, although these drugs have not been directly evaluated for DTs. Antipsychotics are less likely to exacerbate confusion but may increase the risk of seizures; they have no place in the treatment of mild withdrawal symptoms.

Generalized withdrawal seizures rarely require more than giving an adequate dose of benzodiazepines. There is little evidence that anticonvulsants such as phenytoin or gabapentin are more effective in drug-withdrawal seizures, and the risk of seizures has usually passed by the time effective drug levels are reached. The rare patient with status epilepticus must be treated aggressively (Chap. 369).

Rehabilitation of Alcoholics

AN OVERVIEW

After completing alcoholic rehabilitation, ≥60% of alcoholics, especially middle-class patients, maintain abstinence for at least a year, and many achieve lifetime sobriety. The core of treatment uses cognitive-behavioral approaches to help patients recognize the need to change, while working with them to alter their behaviors to enhance compliance. A key step is to optimize motivation toward abstinence through education about alcoholism and instructions to family members to stop protecting the patient from problems caused by alcohol. After years of heavy drinking, patients also need counseling, vocational rehabilitation, and self-help groups such as Alcoholics Anonymous (AA) to help them learn how to deal with life's stresses while sober. A third component, *relapse prevention*, helps the patient to identify situations in which a return to drinking is likely, formulate ways of managing these risks, and develop coping strategies that increase the chances of a return to abstinence if a slip occurs.

While many patients can be treated as outpatients, more intense interventions work better, and some alcoholics do not respond to AA or outpatient groups. Whatever the setting, subsequent contact with outpatient treatment staff should be maintained for a minimum of 6 months and preferably a full year after abstinence. Counseling focuses on areas of improved functioning in the absence of alcohol (i.e., why it is a good idea to continue to abstain) and helping the patient to manage free time without alcohol, develop a nondrinking peer group, and handle stresses on the job.

The physician serves an important role in identifying the alcoholic, diagnosing and treating associated medical or psychiatric syndromes, overseeing detoxification, referring the patient to rehabilitation programs, providing counseling, and, if appropriate, selecting which (if any) medication might be needed. For insomnia, patients should be reassured that troubled sleep is normal after alcohol withdrawal and will improve over subsequent weeks. They should be taught the basic elements of "sleep hygiene" including maintaining consistent schedules for bedtime and awakening. Sleep medications have the danger of being misused and of rebound insomnia when stopped. Sedating antidepressants (e.g., trazodone) should not be used as they interfere with cognitive functioning the next morning and disturb the normal sleep architecture, but occasional use of over-the-counter sleeping medications (sedating antihistamines) can be considered. Anxiety can be addressed by helping the person to gain insight into the temporary nature of the symptoms and to develop strategies to achieve relaxation as well as by using forms of cognitive therapy.

MEDICATIONS FOR REHABILITATION

Several medications have modest benefits when used for the first 6 months of recovery. The opioid-antagonist, naltrexone, 50–150 mg/d orally, appears to shorten subsequent relapses, whether used in the oral form or as a once-per-month 380-mg injection, especially in individuals with the G allele of the All8G polymorphism of the μ opioid receptor. By blocking opioid receptors, naltrexone may decrease activity in the dopamine-rich ventral tegmental reward system, and decrease the feeling of pleasure or reward if alcohol is imbibed. A second medication, acamprosate (Campral) at ~2 g/d divided into three oral doses, has similar modest effects; acamprosate inhibits NMDA receptors, decreasing mild symptoms of protracted withdrawal. Several trials of combined naltrexone and acamprosate using doses similar to those noted above have reported that the combination may be superior to either drug

alone, although not all studies agree.

It is more difficult to establish the asset-to-liability ratio of a third drug, disulfiram, an ALDH inhibitor, used at doses of 250 mg/d. This drug produces vomiting and autonomic nervous system instability in the presence of alcohol as a result of rapidly rising blood levels of the first metabolite of alcohol, acetaldehyde. This reaction can be dangerous, especially for patients with heart disease, stroke, diabetes mellitus, or hypertension. The drug itself carries potential risks of depression, psychotic symptoms, peripheral neuropathy, and liver damage. Disulfiram is best given under supervision of another individual (such as a spouse), especially during discrete periods identified as representing high-risk drinking situations (such as the Christmas holiday). Other relevant drugs under investigation include the nicotinic receptor agonist varenicline, the serotonin antagonist ondansetron, the α-adrenergic agonist prazosin, the GABA-B receptor agonist baclofen, the anticonvulsant topiramate, and cannabinol receptor antagonists. At present, there are insufficient data to determine the asset-to-liability ratio for these medications in treating alcoholism and, therefore, no data to offer solid support for their use in clinical settings.

FURTHER READINGS

Apodaca TR, Longabaugh R: Mechanisms of change in motivational interviewing: A review and preliminary evaluation of the evidence. Addiction 104:705, 2009[PMID: 19413785] [Full Text]

Donovan DM et al: Combined pharmacotherapies and behavioral interventions for alcohol dependence (the COMBINE study): Examination of posttreatment drinking outcomes. J Stud Alcohol Drugs 69:5, 2008[PMID: 18080059] [Full Text]

Koob GF, Volkow ND: Neurocircuitry of addiction. Neuropsychopharmacology 35: 217, 2010[PMID: 19710631] [Full Text]

Russell M et al: Drinking patterns and myocardial infarction: A linear dose-response model. Alcohol Clin Exp Res 33:324, 2009[PMID: 19032575] [Full Text]

Schuckit MA: Alcohol-use disorders. Lancet 373:492, 2009[PMID: 19168210] [Full Text]

———: An overview of genetic influences in alcoholism. J Subst Abuse Treat 36:S5, 2009

Spanagel R, Kiefer F: Drugs for relapse prevention of alcoholism: Ten years of progress. Trends Pharmacol Sci 29:109, 2008[PMID: 18262663] [Full Text]

Taylor B et al: Alcohol and hypertension: Gender differences in dose-response relationships determined through systematic review and meta-analysis. Addiction 104:1981, 2009[PMID: 19804464] [Full Text]

Thygesen LC et al: Cancer incidence among patients with alcohol use disorders—Long-term follow-up. Alcohol Alcohol 44:387, 2009[PMID: 19491282] [Full Text]

Vengeliene V: Neuropharmacology of alcohol addiction. Br J Pharmacol 154:299, 2008[PMID: 18311194] [Full Text]

Verrill C et al: Alcohol-related cirrhosis—Early abstinence is a key factor in prognosis, even in the most severe cases. Addiction 104:768, 2009[PMID: 19344445] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

INTRODUCTION

Opiate analgesics are some of the oldest and most common medications in clinical practice, but have also been abused since at least 300 b.c. Nepenthe (Greek "free from sorrow") helped the hero of the *Odyssey*, but widespread opium smoking in China and the Near East has caused harm for centuries. Since the first chemical isolation of opium and codeine 200 years ago, a wide range of synthetic opioids have been developed, and endogenous opioid peptides were discovered in 1995. Two of the most important adverse effects of all these agents are overdose and dependence. The 0.14% annual prevalence of heroin dependence in the United States is only about one-third the rate of prescription opiate abuse and is substantially lower than the 2% rate of morphine dependence in Southeast and Southwest Asia. While these rates are low relative to other abused substances, their disease burden is substantial, with high rates of morbidity and mortality; disease transmission; increased health care, crime, and law enforcement costs; and less tangible costs of family distress and lost productivity.

The diagnosis of opiate dependence in the *Fourth Diagnostic and Statistical Manual* (DSM-IV) requires the repeated use of the drug while producing problems in three or more areas in a 12-month period. The areas include tolerance, withdrawal, use of greater amounts of opiates than intended, and use despite adverse consequences. The abuse diagnosis is related to legal problems, inability to meet obligations, use in hazardous situations, and continued use despite problems. The most striking aspect of opiate abuse has been its marked increase as the gateway to illicit drugs in the United States. Since 2007, prescription opiates have surpassed marijuana as the most common illicit drug that adolescents initially abuse.

The most commonly abused opiates are diverted prescriptions for oxycodone, followed by heroin and morphine, and—among health professionals—meperidine and fentanyl. Two opiate maintenance treatment agents—methadone and buprenorphine—are also abused, but at substantially lower rates, and the partial opiate agonists such as butorphanol, tramadol, and pentazocine are infrequently abused. The chemistry and general pharmacology of these agents are covered in major pharmacology texts, and this chapter focuses on the neurobiology and pharmacology relevant to abuse, dependence, and their treatments.

NEUROBIOLOGY

During the past 30 years, substantial progress has been made in elucidating the neurobiology of opiates and their effects not only on the three types of opiate receptors (mu, kappa, and delta) but also on the cascade of second, third, and fourth intracellular messenger systems and on neuronal action potentials. The different functional activities of these three receptors are summarized in **Table 393-1**, and abuse liability is primarily associated with the mu receptor. A fourth type of opiate receptor, the orphanin receptor, also modulates pain but is not affected by opiate drugs. These opiate receptors are all G protein–linked and coupled to the cyclic adenosine monophosphate (cyclic AMP) second messenger system and to potassium channels. Opiates are inhibitory and block the potassium channels from opening and depolarizing the neuron, which would produce an action potential. Thus, opiates acutely inhibit neuronal activity. Analgesia and sedation are induced through this inhibition of specific brain pathways, while the "high" from opiates involves an indirect activation of a different brain pathway—the mesolimbic dopamine pathway.

Table 393-1 Actions of Opioid Receptors

Receptor Type	Actions
Mu (μ) (e.g., morphine)	Analgesia, reinforcement euphoria, cough and appetite suppression, decreased respirations, decreased GI motility, sedation, hormone changes, dopamine and acetylcholine release
Kappa (κ) (e.g., butorphanol)	Dysphoria, decreased GI motility, decreased appetite, decreased respiration, psychotic symptoms, sedation, diuresis, analgesia
Delta (Δ) (e.g., etorphine)	Hormone changes, appetite suppression, dopamine release

Abbreviation: GI, gastrointestinal.

The various effects of opiates are related to the specific neuroanatomic locations of mu receptors. Reinforcing and euphoric effects of opiates occur in the dopaminegic pathway from the ventral tegmental area (VTA) to the nucleus accumbens, where opiates increase synaptic levels of dopamine. This increase is due to inhibition of GABAergic neurons that inhibit both the VTA and nucleus accumbens activity. However, the "high" only occurs when the *rate of change* in dopamine is fast. Large, rapidly administered doses of opiates block GABA inhibition and produce a burst of nucleus accumbens activity that is associated with "high" in all abused drugs. Therefore, routes of administration that slowly increase opiate blood and brain levels, such as oral and transdermal routes, are effective for analgesia and sedation but do not produce an opiate "high" that follows smoking and intravenous routes. Other acute effects such as analgesia and respiratory depression leading to overdose are due to stimulation of opiate receptors located in other areas such as the locus coeruleus.

Opiate dependence and withdrawal are chronic effects related to the cyclic AMP system. This second messenger phosphorylates various intracellular proteins and produces a cascade of changes reaching into the nucleus and DNA. Immediate early gene products such as c-*fos* and c-*jun* are activated followed by regulation of other genes with more sustained protein transcription such as *delta c-fos*. With these sustained gene activations, several receptor-level changes occur, including downregulation of receptor numbers, reduced neuronal cell-surface receptor trafficking, uncoupling of G proteins from the mu opiate receptors, and upregulation of cyclic AMP second messenger systems. These effects are also reflective of genetic risk factors for drug dependence, with estimates of up to 50% of the risk for dependence due to polygenic inheritance. Specific functional genetic polymorphisms in the mu opiate receptor gene appear associated with this risk for opiate abuse, including one producing a threefold increase in this receptor's affinity for opiates and the endogenous ligand beta endorphin. Epigenetic methylation changes also occur on the DNA of the mu receptor gene of opiate addicts. DNA methylation inhibits gene transcription.

This molecular cascade links acute intoxication and sedation to chronic opiate dependence and withdrawal within the specific neuroanatomic structure of the locus coeruleus. The locus coeruleus is the brain's largest concentration of noradrenergic neurons and is responsible for a large proportion of brain cortical activation. When large opiate doses saturate and activate all of its mu receptors, its steady rate of action potentials can cease due to the inactivation of potassium channels. When this direct inhibitory effect is sustained over weeks and months of opiate use, a secondary set of regulatory effects take place in the cyclic AMP system that leads to tolerance, dependence, and withdrawal symptoms.

Opiate withdrawal symptoms reflect overactivity of adrenergic neurons that are located in the locus ceruleus. Opiates suppress the activity of these neurons, and when this suppression continues chronically from daily opiate use, a secondary upregulation occurs in adenyl cyclase enzyme capacity and the production of cyclic AMP from ATP. This upregulation is a homeostatic response to the chronic opiate suppression, but when that suppression is terminated by discontinuing the opiate, this enhanced adenyl cyclase activity leads to a marked increase in cyclic AMP. The now very high levels of cyclic AMP activate the sodium-potassium channels and produce a high level of action potentials in these adrenergic neurons. This adrenergic arousal is one basis for the symptoms of opiate withdrawal and takes about 7 days to readjust to normal levels of adenyl cyclase activity and the associated resolution of opiate withdrawal symptoms. This molecular model of adrenergic neuronal activation during withdrawal has had important treatment implications, such as the use of clonidine for opioid withdrawal.

PHARMACOLOGY

Tolerance and withdrawal commonly occur with chronic daily use as quickly as 6–8 weeks depending on the dose and frequency of dosing. Tolerance appears to be primarily a pharmacodynamic rather than pharmacokinetic effect, with relatively limited induction of the cytochrome P450 or other liver enzymes. The metabolism of opiates occurs in the liver primarily through the cytochrome P450 systems of 2D6 and 3A4. They then are conjugated to glucuronic acid and excreted in small amounts in feces. The plasma half-lives generally range from 2.5 to 3 h for morphine and more than 22 h for methadone. The shortest half-lives of several minutes are for fentanyl-related opiates and the longest are for buprenorphine and its active metabolites, which can block opiate withdrawal for up to 3 days after a single dose. Tolerance to the mental effects of opioids leads to the need for ever-increasing amounts of drugs to sustain the desired euphoriant effects—as well as to avoid the discomfort of withdrawal. This combination has the expected consequence of strongly reinforcing dependence once it has started. The role of endogenous opioid peptides in opioid dependence is uncertain.

The clinical aspects of abuse are tied to route of administration and the rapidity of an opiate bolus in reaching the brain. Intravenous and smoked administration is routine not only because it is the most efficient route but also because it rapidly produces a bolus of high drug concentration in the brain. This bolus produces a "rush," followed by euphoria, a feeling of tranquility, and sleepiness ("the nod").

Heroin produces effects that last 3–5 h, and several doses a day are required to forestall manifestations of withdrawal in dependent persons. Symptoms of opioid withdrawal begin 8–10 h after the last dose. Many of these symptoms reflect increased activity of the autonomic nervous system. Lacrimation, rhinorrhea, yawning, and sweating appear first. Restless sleep followed by weakness, chills, gooseflesh ("cold turkey"), nausea and vomiting, muscle aches, and involuntary movements ("kicking the habit"), hyperpnea, hyperthermia, and hypertension occur in later stages of the withdrawal syndrome. The acute course of withdrawal may last 7–10 days. A secondary phase of protracted abstinence lasts for 26–30 weeks and is characterized by hypotension, bradycardia, hypothermia, mydriasis, and decreased responsiveness of the respiratory center to carbon dioxide.

OPIOID EFFECTS ON ORGANS

Besides the brain effects of opioids on sedation and euphoria and the combined brain and peripheral nervous system effects on analgesia, a wide range of other organs can be affected. The cough reflex is inhibited through the brain, leading to the use of some opiates as an antitussive, and nausea and vomiting are due to brainstem effects on the medulla. The release of several hormones is inhibited including corticotropin-releasing factor (CRF) and luteinizing hormone, which reduce cortisol and sex hormone levels, respectively. The clinical manifestations of these reductions can involve poor responses to stress and reductions in sex drive. An increase in prolactin also contributes to the reduced sex drive in males. Two other hormones affected are decreased thyrotropin and increased growth hormone. Respiratory depression results from opiate-induced insensitivity of brainstem neurons to increases in carbon dioxide. This depression contributes to overdose, but in patients with pulmonary disease even opiate doses well below those typical of overdose can produce clinically significant complications. In overdoses, aspiration pneumonia is a common complication due to loss of the choking reflex. Opiates reduce gut motility, which is helpful for diarrhea, but can lead to nausea, constipation, and anorexia with weight loss. Deaths have occurred in early methadone maintenance programs due to severe constipation and toxic megacolon. Opiates may prolong QT intervals and lead to sudden death in some patients. Two opiates particularly noted for this complication are methadone and a long-acting form of methadone called LAAM that was withdrawn from the market. Orthostatic hypotension may occur due to histamine release and peripheral blood vessel dilation, which is an opiate effect usefully applied to managing acute myocardial infarction.

Heroin users in particular tend to use opiates intravenously and be polydrug users, also using alcohol, sedatives, cannabinoids, and stimulants. None of these other drugs serve as substitutes for opioids, but they have desired additive effects. One needs to be sure that the person undergoing a withdrawal reaction is not also withdrawing from alcohol or sedatives, which might be more dangerous and more difficult to manage.

Besides the ever-present risk of fatal overdose, hepatitis B and AIDS are among the many potential complications of sharing contaminated hypodermic syringes. Bacterial infections lead to septic complications such as meningitis, osteomyelitis, and abscesses in various organs. Attempts to illicitly manufacture meperidine in the 1980s resulted in the production of a highly specific neurotoxin, MPTP, which produced parkinsonism in users.

TOXICITY AND OVERDOSE

Lethal overdose is a relatively common complication of opiate dependence and must be rapidly recognized and treated because naloxone provides a highly specific reversal agent that is relatively free of complications. The diagnosis generally does not rely on blood or urine toxicology results but on clinical signs and symptoms. The presentation involves shallow and slow respirations, pupillary miosis (mydriasis does not occur until significant brain anoxia supervenes), bradycardia, hypothermia, and stupor or coma. If naloxone is not administered, progression to respiratory and cardiovascular collapse leading to death occurs. At autopsy, cerebral edema and sometimes frothy pulmonary edema are generally found, but those pulmonary effects are most likely from allergic reactions to adulterants mixed with the heroin. Opiates generally do not produce seizures except for unusual cases of mixed drug abuse with the opiate meperidine or with high doses of tramadol.

Treatment: Opioid Overdose

Beyond the acute treatment of opiate overdose with naloxone, clinicians have two general treatment paths: opioid maintenance treatment or detoxification. Most opioid-dependent individuals engage in multiple episodes of all three categories of treatment during the course of their drug-using careers. Agonist and partial agonist medications are commonly utilized for both maintenance and detoxification purposes. Alpha-2-adrenergic agonists are primarily used for detoxification. Antagonists are used to accelerate detoxification and then continued postdetoxification to prevent relapse. Only the residential medication-free programs have had success that comes close to matching that of the medication-based programs. Success of the various treatment approaches is assessed as retention in treatment, reduced opioid and other drug use, as well as secondary outcomes such as HIV risk behaviors, crime, psychiatric symptoms, and medical co-morbidity.

Stopping opiates is like stopping most drugs of abuse—it is much easier to stop than to prevent relapses. Long-term relapse prevention for opioid-dependent persons requires combined

pharmacologic and psychosocial approaches. Chronic users tend to prefer pharmacologic approaches; those with shorter histories of drug abuse are more amenable to detoxification and psychosocial interventions.

OPIATE OVERDOSE TREATMENT

Managing overdose requires naloxone and support of vital functions, including intubation if needed. The opiate antagonist naloxone is given at 0.4–2 mg IV or IM, with an expected response within 1–2 min. If the overdose is due to buprenorphine, then naloxone might be required at total doses of 10 mg or greater, but primary buprenorphine overdose is nearly impossible because this agent is a partial opiate agonist. Partial agonism means that as the dose of buprenorphine is increased, it has greater opiate antagonist than agonist activity. Thus, a 0.2-mg buprenorphine dose leads to analgesia and sedation, while a hundred times greater 20-mg dose produces profound opiate antagonism, precipitating opiate withdrawal in a person who was opiate dependent on morphine or methadone. When 10 mg of naloxone fails to produce arousal in the patient, another cause of toxicity must be found. Before reaching such large naloxone doses, however, it is important to recognize that the goal is to reverse the respiratory depression and not to administer so much naloxone that it precipitates opiate withdrawal. Because naloxone only lasts a few hours and most opiates last considerably longer, close monitoring and an IV naloxone drip is frequently employed to provide a continuous level of antagonism for 24–72 h depending on the opiate used in the overdose (e.g., morphine vs. methadone). Other sedative drugs that produce significant overdoses must also be considered if naloxone has only a limited effect. The most common are benzodiazepines, which have produced overdoses and deaths in combination with buprenorphine. A specific antagonist for benzodiazepines—flumazenil at 0.2 mg/min—can be given to a maximum of 3 g/h, but it may precipitate seizures and increase intracranial pressure. Like naloxone, administration for a prolonged period is usually required since most benzodiazepines remain active for considerably longer than flumazenil.

Support of vital functions may include oxygen and positive- pressure breathing, IV fluids, pressor agents for hypotension, and cardiac monitoring to detect QT prolongation, which might require treatment. Activated charcoal and gastric lavage may be helpful for oral ingestions, but intubation will be needed if the patient is stuporous.

OPIATE WITHDRAWAL TREATMENT

The principles of detoxification are the same for all drugs: to substitute a longer-acting, orally active, pharmacologically equivalent drug for the abused drug, stabilize the patient on that drug, and then gradually withdraw the substituted drug. Methadone is admirably suited for such use in opioid-dependent persons, and the partial mu agonist buprenorphine is another option. Clonidine, a centrally acting sympatholytic agent, has also been used for detoxification. By reducing central sympathetic outflow, clonidine mitigates many of the signs of sympathetic overactivity. Clonidine has no narcotic action and is not addictive. Lofexidine, a clonidine analogue with less hypotensive effect, is being developed for use.

Methadone for Detoxification

Methadone dose tapering regimens for detoxification range from 2 to 3 weeks to as long as 180 days, but this approach is controversial given the relative effectiveness of methadone maintenance and the low success rates of detoxification. Unfortunately, the vast majority of patients tend to relapse to heroin or other opiates during or after the detoxification period, indicative of the chronic and relapsing nature of opioid dependence.

Buprenorphine for Detoxification

Because it is a partial agonist, buprenorphine produces fewer withdrawal symptoms and may allow briefer detoxifications compared with full agonists like methadone, but it does not appear to have better outcomes than methadone tapering. Buprenorphine is superior to the alpha-2-adrenergic agonist clonidine in reducing symptoms of withdrawal, retaining patients in a withdrawal protocol, and in treatment completion.

Alpha-2-Adrenergic Agonists for Detoxification

Several alpha-2-adrenergic agonists have relieved opioid withdrawal by suppressing central noradrenergic hyperactivity. Alpha-2-adrenergic agonists moderate the symptoms of noradrenergic hyperactivity via actions in the central nervous system. Clonidine relieves some signs and symptoms of opiate withdrawal such as lacrimation, rhinorrhea, muscle pain, joint pain, restlessness, and gastrointestinal symptoms, but it is not a drug of abuse or dependence. Unfortunately, clonidine is associated with significant hypotension, which has stimulated investigation of lofexidine, guanfacine, and guanabenz acetate. Lofexidine can be dosed up to ~2 mg/d and appears to be associated with fewer adverse effects, and it is therefore likely to replace clonidine as the leading opioid withdrawal treatment in this drug class. Clonidine or lofexidine are typically administered orally, in three or four doses per day, with dizziness, sedation, lethargy, and dry mouth as the primary adverse side effects. Completion rates of managed withdrawal assisted with clonidine and other alpha-2-adrenergic agents vs. methadone have been comparable.

Rapid and Ultrarapid Opiate Detoxification

The opioid antagonist naltrexone typically combined with an alpha-2-adrenergic agonist has been purported to shorten the duration of withdrawal without significantly increasing patient discomfort. Another benefit to rapid opiate detoxification (ROD) is the reduced time between opioid use and the commencement of sustained naltrexone treat-ment for prevention of relapse (see below). ROD completion rates using naltrexone and clonidine range from 75 to 81% compared to 40–65% for methadone or clonidine alone. Buprenorphine in combination with naltrexone and clonidine reduced ROD from 3 to 1 day of detoxification. Ultrarapid opiate detoxification is an extension of ROD using anesthetics, but is highly controversial due to the medical risks and mortality associated with it.

AGONIST MEDICATIONS FOR OPIOID DEPENDENCE

Methadone maintenance substitutes a once-daily oral opioid dose for three-to-four times daily heroin. Methadone saturates the opioid receptors, and by inducing a high level of opiate tolerance, blocks the desired euphoria from additional opiates. Buprenorphine, a partial opioid agonist, also can be given once daily at sublingual doses of 4–32 mg daily, and in contrast to methadone it can be given in an office-based primary care setting.

Methadone Maintenance

Methadone's slow onset of action when taken orally, long elimination half-life (24–36 h), and production of cross-tolerance at doses from 80 to 150 mg are the basis for its efficacy in treatment retention and reductions in IV drug use, criminal activity, and HIV risk behaviors and mortality. Methadone can prolong the QT interval at rates as high as 16% above the rates in non-methadone-maintained, drug-injecting patients, but it has been used safely in the treatment of opioid dependence for 40 years.

Buprenorphine Maintenance

While France and Australia have had sublingual buprenorphine maintenance since 1996, the USFDA approved it as a Schedule III drug in 2002 for managing opiate dependence. Unlike the full agonist methadone, buprenorphine is a partial agonist of mu-opioid receptors with a slow onset and long duration of action, allowing for alternate-day dosing. Its partial agonism reduces the risk of unintentional overdose but limits its efficacy to patients who need the equivalent of only 60–70 mg of methadone, and many patients in methadone maintenance require higher doses up to 150 mg daily. Buprenorphine is combined with naloxone at a 4:1 ratio in order to reduce its abuse liability. A subcutaneous buprenorphine implant has also been tested, but results are not yet available.

In the United States, the ability of primary care physicians to prescribe buprenorphine for opioid dependence presents an important and far-reaching opportunity to improve access and quality of treatment as well as reduce social harm. Europe, Asia, and Australia have found reduced opioid-related deaths and drug-injection-related medical morbidity with buprenorphine available in primary care. Retention in office-based buprenorphine treatment has been greater than 70% at 6-month follow-ups.

ANTAGONIST MEDICATIONS FOR OPIOID DEPENDENCE

The rationale for using narcotic antagonist therapy is that blocking the action of self-administered opioids should eventually extinguish the habit, but this therapy is poorly accepted by patients. Naltrexone, a long-acting orally active pure opioid antagonist, can be given three times a week at doses of 100–150 mg and a depot form for monthly administration is available. Because it is an antagonist, the patient must first be detoxified from opioid dependence before starting naltrexone. When taken chronically for even years, it is safe, associated with few side effects (headache, nausea, abdominal pain), and can be given to patients infected with hepatitis B or C without producing hepatotoxicity. However, most providers refrain from prescribing it if liver function tests are 3–5 times above normal levels. Naltrexone maintenance combined with psychosocial therapy is effective in reducing heroin use, but medication adherence is low. Depot injection formulations lasting up to 4 weeks markedly improve adherence, retention, and drug use. Subcutaneous naltrexone implants in Russia, China, and Australia have doubled treatment retention and reduced relapse to half that of oral naltrexone.

MEDICATION-FREE TREATMENT

Most opiate addicts enter medication-free treatments in inpatient, residential, or outpatient settings, but 1- to 5-year outcomes are very poor compared to pharmacotherapy except for residential settings lasting 6 to 18 months. The residential programs require full immersion in a regimented system that has progressively increasing levels of independence and responsibility within a controlled community of fellow drug abusers. These medication-free programs, as well as the pharmacotherapy programs, also include counseling and behavioral treatments designed to teach interpersonal and cognitive skills for coping with stress and for avoiding situations leading to easy access to drugs or to craving. Relapse is prevented by having the individual very gradually reintroduced to greater responsibilities and to the working environment outside of the protected therapeutic community.

PREVENTION

Preventing opiate abuse represents a critically important challenge for physicians. Opiate prescriptions are the most common source of drugs accessed by adolescents who begin a pattern of illicit drug abuse; in the United States, 9000 adolescents become opiate abusers every day. The major sources of these drugs are family members, not drug dealers or the Internet. Pain management involves giving sufficient opiates to relieve the pain over as short a period of time as the pain warrants. The patient then needs to dispose of any remaining opiates, not save them in the medicine cabinet, because this behavior leads to diversion to adolescents. Finally, physicians should never prescribe opiates for themselves.

FURTHER READINGS

Johnston LD et al: *Monitoring the Future: National Results on Adolescent Drug Use: Overview of Key Findings 2008*. Bethesda, Maryland, National Institute on Drug Abuse, 2008

Nestler EJ: Transcriptional mechanisms of addiction: Role of DeltaFosB. Philos Trans R Soc Lond B Biol Sci 363:3245, 2008[PMID: 18640924] [Full Text]

Ries R et al (eds): *Principles of Addiction Medicine, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2009*

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

COCAINE AND OTHER COMMONLY ABUSED DRUGS: INTRODUCTION

The abuse of cocaine and other psychostimulant drugs reflects a complex interaction between the pharmacologic properties of each drug, the personality and expectations of the user, and the environmental context in which the drug is used. Polydrug abuse involving the concurrent use of several drugs with different pharmacologic effects is increasingly common. Some forms of polydrug abuse, such as the combined use of heroin and cocaine intravenously, are especially dangerous and are a major problem in hospital emergency rooms. Sometimes one drug is used to enhance the effects of another, as with the combined use of benzodiazepines and methadone, or cocaine and heroin in methadone-maintained patients.

Chronic cocaine and psychostimulant abuse may cause a number of adverse health consequences, and preexisting disorders such as hypertension and cardiac disease may be exacerbated by drug abuse. The combined use of two or more drugs may accentuate medical complications associated with abuse of one of them. Chronic drug abuse is often associated with immune system dysfunction and increased vulnerability to infections, which in turn contributes to the risk for HIV infection. In addition, concurrent use of cocaine and opiates (the "speedball") is frequently associated with needle sharing by IV drug users. Intravenous drug abusers continue to represent the largest single group of persons with HIV infection in several major metropolitan areas in the United States as well as in many parts of Europe and Asia.

COCAINE

Cocaine is a stimulant and a local anesthetic with potent vasoconstrictor properties. The leaves of the *coca* plant (*Erythroxylon coca*) contain ~0.5–1% cocaine. The drug produces physiologic and behavioral effects when administered orally, intranasally, intravenously, or via inhalation following pyrolysis (smoking). The reinforcing effects of cocaine appear to be related to activation of dopaminergic neurons in the mesolimbic system. Cocaine increases synaptic concentrations of the monamine neurotransmitters dopamine, norepinephrine, and serotonin by binding to transporter proteins in presynaptic neurons and blocking reuptake.

PREVALENCE OF COCAINE USE

Cocaine is widely available throughout the United States, and cocaine abuse occurs in virtually all social and economic strata of society. The prevalence of cocaine abuse in the general population has been accompanied by an increase in cocaine abuse by heroin-dependent persons, including those in methadone maintenance programs. Intravenous cocaine is often used concurrently with IV heroin. This combination purportedly attenuates the postcocaine "crash" and substitutes a cocaine "high" for the heroin "high" blocked by methadone.

ACUTE AND CHRONIC INTOXICATION

There has been an increase in both IV administration and inhalation of pyrolyzed cocaine via smoking. Following intranasal administration, changes in mood and sensation are perceived within 3–5 min, and peak effects occur at 10–20 min. The effects rarely last more than 1 h. Inhalation of pyrolyzed materials includes inhaling crack/cocaine or smoking coca paste, a product made by extracting cocaine preparations with flammable solvents, and cocaine free-base smoking. Freebase cocaine, including the freebase prepared with sodium bicarbonate (crack), has become increasingly popular because of the relative high potency of the compound and its rapid onset of action (8–10 s following smoking).

Cocaine produces a brief, dose-related stimulation and enhancement of mood and an increase in cardiac rate and blood pressure. Body temperature usually increases following cocaine administration, and high doses of cocaine may induce lethal pyrexia or hypertension. Because cocaine inhibits reuptake of catecholamines at adrenergic nerve endings, the drug potentiates sympathetic nervous system activity. Cocaine has a short plasma half-life of approximately 45–60 min. It is metabolized by plasma esterases, and cocaine metabolites are excreted in urine. The very short duration of the euphorigenic effects of cocaine observed in chronic abusers is probably due to both acute and chronic tolerance. Frequent self-administration of the drug (two to three times per hour) is often reported by chronic cocaine abusers. Alcohol is often used to modulate both the cocaine high and the dysphoria associated with the abrupt disappearance of cocaine's effects. A metabolite of cocaine, cocaethylene, has been detected in the blood and urine of persons who concurrently abuse alcohol and cocaine. Cocaethylene induces changes in cardiovascular function similar to those of cocaine alone, and the

pathophysiologic consequences of alcohol abuse plus cocaine abuse may be additive when both are used together.

The prevalent assumption that cocaine inhalation or IV administration is relatively safe is contradicted by reports of death from respiratory depression, cardiac arrhythmias, and convulsions associated with cocaine use. In addition to generalized seizures, neurologic complications may include headache, ischemic or hemorrhagic stroke, or subarachnoid hemorrhage. Disorders of cerebral blood flow and perfusion in cocaine-dependent persons have been detected with magnetic resonance spectroscopy (MRS) studies. Severe pulmonary disease may develop in individuals who inhale crack cocaine; this effect is attributed both to the direct effects of cocaine and to residual contaminants in the smoked material. Hepatic necrosis may occur following chronic crack/cocaine use. Protracted cocaine abuse may also cause paranoid ideation and visual and auditory hallucinations, a state that resembles alcoholic hallucinosis.

Although men and women who abuse cocaine may report that the drug enhances libidinal drive, chronic cocaine use causes significant loss of libido and adversely affects sexual function. Impotence and gynecomastia have been observed in male cocaine abusers, and these abnormalities often persist for long periods following cessation of drug use. Women who abuse cocaine may have major derangements in menstrual cycle function, including galactorrhea, amenorrhea, and infertility. Chronic cocaine abuse may cause persistent hyperprolactinemia as a consequence of disordered dopaminergic inhibition of prolactin secretion by the anterior pituitary. Cocaine abuse by pregnant women, particularly crack smoking, has been associated with both an increased risk of congenital malformations in the fetus and perinatal cardiovascular and cerebrovascular disease in the mother. However, cocaine abuse per se is probably not the sole cause of these perinatal disorders, because maternal cocaine abuse is often associated with poor nutrition and prenatal health care as well as polydrug abuse that may contribute to the risk for perinatal disease.

Psychological dependence on cocaine, indicated by inability to abstain from frequent compulsive use, has also been reported. Although the occurrence of withdrawal syndromes involving psychomotor agitation and autonomic hyperactivity remains controversial, severe depression ("crashing") following cocaine intoxication may accompany drug withdrawal.

Treatment: Cocaine Overdose and Chronic Abuse

Treatment of cocaine overdose is a medical emergency that is best managed in an intensive care unit. Cocaine toxicity produces a hyperadrenergic state characterized by hypertension, tachycardia, tonic-clonic seizures, dyspnea, and ventricular arrhythmias. Intravenous diazepam in doses up to 0.5 mg/kg administered over an 8-h period has been shown to be effective for control of seizures. Ventricular arrhythmias have been managed successfully by administration of 0.5–1 mg of propranolol IV. Since many instances of cocaine-related mortality have been associated with concurrent use of other illicit drugs (particularly heroin), the physician must be prepared to institute effective emergency treatment for multiple drug toxicities.

Treatment of chronic cocaine abuse requires the combined efforts of primary care physicians, psychiatrists, and psychosocial care providers. Early abstinence from cocaine use is often complicated by symptoms of depression and guilt, insomnia, and anorexia, which may be as severe as those observed in major affective disorders. Individual and group psychotherapy, family therapy, and peer group assistance programs are often useful for inducing prolonged remission from drug use. A number of medications used for the treatment of various medical and psychiatric disorders have been administered to reduce the duration and severity of cocaine abuse and dependence. The search for a medication that is both safe and highly effective for cocaine detoxification or maintenance of abstinence is continuing. Although psychotherapy may be effective, no specific form of psychotherapy or behavioral modification is uniquely beneficial.

MARIJUANA AND CANNABIS COMPOUNDS

Cannabis sativa contains >400 compounds in addition to the psychoactive substance, delta-9-tetrahydrocannabinol (THC). Marijuana cigarettes are prepared from the leaves and flowering tops of the plant, and a typical marijuana cigarette contains 0.5–1 g of plant material. The usual THC concentration varies between 10 and 40 mg, but concentrations >100 mg per cigarette have been detected. Hashish is prepared from concentrated resin of *C. sativa* and contains a THC concentration of between 8 and 12% by weight. "Hash oil," a lipid-soluble plant extract, may contain THC between 25 and 60% and may be added to marijuana or hashish to enhance its THC concentration. Smoking is the most common mode of marijuana or hashish use. During pyrolysis, >150 compounds in addition to THC are released in the smoke. Although most of these compounds do not have psychoactive properties, they may have physiologic effects.

THC is quickly absorbed from the lungs into blood, and then rapidly sequestered in tissues. THC is metabolized primarily in the liver, where it is converted to 11-hydroxy-THC, a psychoactive compound, and >20 other metabolites. Many THC metabolites are excreted through the feces at a relatively slow rate of clearance in comparison to most other psychoactive drugs.

Specific cannabinoid receptors (CB_1 and CB_2) have been identified in the central and peripheral nervous system. High densities of cannabinoid receptors have been found in the cerebral cortex, basal ganglia, and hippocampus. T and B lymphocytes also contain cannabinoid receptors, and these appear to mediate the anti-inflammatory and immunoregulatory properties of cannabinoids. A naturally occurring THC-like ligand has been identified and is widely distributed in the nervous system.

PREVALENCE OF USE

Marijuana is the most commonly used illegal drug in the United States, and its use is particularly prevalent among adolescents. Marijuana is relatively inexpensive and is often considered to be less hazardous than other controlled drugs and substances. Very potent forms of marijuana (sinsemilla) are now available in many locations, and concurrent use of marijuana with crack/cocaine and phencyclidine is not uncommon.

ACUTE AND CHRONIC INTOXICATION

Acute intoxication from marijuana and cannabis compounds is related to both the dose of THC and the route of administration. THC is absorbed more rapidly from marijuana smoking than from orally ingested cannabis compounds. Acute marijuana intoxication may produce a perception of relaxation and mild euphoria resembling mild to moderate alcohol intoxication. This condition is usually accompanied by some impairment in thinking, concentration, and perceptual and psychomotor function. Higher doses of cannabis may produce more pronounced impairment in concentration and perception, as well as greater sedation. Although the acute effects of marijuana intoxication are relatively benign in normal users, the drug can precipitate severe emotional disorders in individuals who have antecedent psychotic or neurotic problems. Like other psychoactive compounds, both the user's expectations and the environmental context are important determinants of the type and severity of the effects of marijuana intoxication.

As with abuse of cocaine, opioids, and alcohol, chronic marijuana abusers may lose interest in common socially desirable goals and steadily devote more time to drug acquisition and use. However, THC does not cause a specific and unique "amotivational syndrome." The range of symptoms sometimes attributed to marijuana use is difficult to distinguish from mild to moderate depression and the maturational dysfunctions often associated with protracted adolescence. Chronic marijuana use has also been reported to increase the risk of psychotic symptoms in individuals with a past history of schizophrenia. Persons who initiate marijuana smoking before the age of 17 may exhibit more pronounced cognitive deficits and they may also be at higher risk for polydrug and alcohol abuse problems in later life, but the role of marijuana in this causal sequence is uncertain.

PHYSICAL EFFECTS

Conjunctival injection and tachycardia are the most frequent immediate physical concomitants of smoking marijuana. Tolerance for marijuana-induced tachycardia develops rapidly among regular users. However, marijuana smoking may precipitate angina in persons with a history of coronary insufficiency. Exercise-induced angina may be increased after marijuana use to a greater extent than after tobacco cigarette smoking. Patients with cardiac disease should be strongly advised not to smoke marijuana or use cannabis compounds.

Significant decrements in pulmonary vital capacity have been found in regular daily marijuana smokers. Because marijuana smoking typically involves deep inhalation and prolonged retention of marijuana smoke, marijuana smokers may develop chronic bronchial irritation. Impairment of single-breath carbon monoxide diffusion capacity (DI_{CO}) is greater in persons who smoke both marijuana and tobacco than in tobacco smokers.

Although marijuana has also been associated with a number of other adverse effects, many of these studies await replication and confirmation. A reported correlation between chronic marijuana use and decreased testosterone levels in males has not been confirmed. Decreased sperm count and sperm motility and morphologic abnormalities of spermatozoa following marijuana use have been reported. Prospective studies demonstrated a correlation between impaired fetal growth and development and heavy marijuana use during pregnancy. Marijuana has also been implicated in derangements of the immune system; in chromosomal abnormalities; and in inhibition of DNA, RNA, and protein synthesis; however, these findings have not been confirmed or related to any specific physiologic effect in humans.

TOLERANCE AND PHYSICAL DEPENDENCE

Habitual marijuana users rapidly develop tolerance to the psychoactive effects of marijuana, then smoke more frequently and try to acquire more potent cannabis compounds. Tolerance for the physiologic effects of marijuana develops at different rates; e.g., tolerance develops rapidly for marijuana-induced tachycardia but more slowly for marijuana-induced conjunctival injection. Tolerance for both behavioral and physiologic effects of marijuana decreases rapidly upon cessation of marijuana use.

Withdrawal signs and symptoms have been reported in chronic cannabis users, and the severity of symptoms is related to dosage and duration of use. These symptoms typically reach their peak several

days after cessation of chronic use and include irritability, anorexia, and sleep disturbances. Withdrawal signs and symptoms observed in chronic marijuana users are usually relatively mild in comparison to those observed in heavy opiate or alcohol users and rarely require medical or pharmacologic intervention. However, more severe and protracted abstinence syndromes may occur after sustained use of high-potency cannabis compounds.

THERAPEUTIC USE OF MARIJUANA

Marijuana, administered as cigarettes or as a synthetic oral cannabinoid (dronabinol), has been proposed to have a number of medicinal properties that may be clinically useful in some situations. These include antiemetic effects in chemotherapy recipients, appetite-promoting effects in AIDS patients, reduction of intraocular pressure in glaucoma, and reduction of spasticity in multiple sclerosis and other neurologic disorders. With the possible exception of AIDS-related cachexia, none of these attributes of marijuana compounds is clearly superior to other readily available therapies.

METHAMPHETAMINE

Methamphetamine is also referred to as "meth," "speed," "crank," "chalk," "ice," "glass," or "crystal." Methamphetamine is a mixed-action monoamine releaser with activity at dopamine, serotonin, and norepinephrine systems. Despite drug seizures, closures of clandestine laboratories that produce methamphetamine illegally, and an increase in methamphetamine abuse prevention programs, methamphetamine was considered second only to cocaine as a drug threat to society by the U.S. Department of Justice in 2009. Hospital admissions for methamphetamine treatment more than doubled between 1998 and 2007, and young adults (aged 18–25 years) have the highest use rates.

Methamphetamine can be used by smoking, snorting, IV injection, or oral administration. Methamphetamine abusers report that drug use induces feelings of euphoria and decreased fatigue. Adverse consequences of methamphetamine abuse include headache, difficulty concentrating, diminished appetite, abdominal pain, vomiting or diarrhea, disordered sleep, paranoid or aggressive behavior, and psychosis. Chronic methamphetamine abuse can result in severe dental caries, described as blackened, rotting, crumbling teeth. Severe, life-threatening methamphetamine toxicity may include hypertension, cardiac arrhythmia or cardiac failure, subarachnoid hemorrhage, ischemic stroke, intracerebral hemorrhage, convulsions, or coma. Methamphetamines increase the release of monoamine neurotransmitters (dopamine, norepinephrine, and serotonin) from presynaptic neurons. It is thought that the euphoric and reinforcing effects of this class of drugs are mediated through dopamine and the mesolimbic system, whereas the cardiovascular effects are related to norepinephrine. MRS studies of the brain suggest that chronic abusers have neuronal damage in the frontal areas and basal ganglia.

Therapy of acute methamphetamine overdose is largely symptomatic. Ammonium chloride may be useful to acidify the urine and enhance clearance of the drug. Hypertension may respond to sodium nitroprusside or α-adrenergic antagonists. Sedatives may reduce agitation and other signs of central nervous system hyperactivity. Treatment of chronic methamphetamine dependence may be accomplished in either an inpatient or outpatient setting using strategies similar to those described above for cocaine abuse.

MDMA (3,4-methylenedioxymethamphetamine), or *ecstasy*, is a derivative of methamphetamine. Ecstasy is usually taken orally but may be injected or inhaled; its effects last for 3–6 h. In addition to amphetamine-like effects, MDMA can induce hyperthermia and vivid hallucinations and other perceptual distortions. Recent studies have revealed that MDMA use is associated with cognitive and memory impairment and a mild withdrawal syndrome after cessation of use. The long-term consequences of recreational use of MDMA by young persons are poorly understood.

LYSERGIC ACID DIETHYLAMIDE (LSD)

The discovery of the psychedelic effects of LSD led to an epidemic of LSD abuse during the 1960s. Imposition of stringent constraints on the manufacture and distribution of LSD (classified as a Schedule I substance by the U.S. Food and Drug Administration), as well as public recognition that psychedelic experiences induced by LSD were a health hazard, have resulted in a reduction in LSD abuse. LSD still remains popular among adolescents and young adults, and there are indications that LSD use among young persons has been increasing in some areas in the United States.

LSD is a very potent hallucinogen; oral doses as low as 20 μg may induce profound psychological and physiologic effects. Tachycardia, hypertension, pupillary dilation, tremor, and hyperpyrexia occur within minutes following oral administration of 0.5–2 μg/kg. A variety of bizarre and often conflicting perceptual and mood changes, including visual illusions, synesthesias, and extreme lability of mood, usually occur within 30 min after LSD intake. These effects of LSD may persist for 12–18 h, even though the half-life of the drug is only 3 h.

The most frequent acute medical emergency associated with LSD use is a panic episode (the "bad trip"), which may persist up to 24 h. Management of this problem is best accomplished by supportive reassurance ("talking down") and, if necessary, administration of small doses of anxiolytic drugs.

Adverse consequences of chronic LSD use include enhanced risk for schizophreniform psychosis and derangements in memory function, problem solving, and abstract thinking. Treatment of these disorders is best carried out in specialized psychiatric facilities.

Tolerance develops rapidly for LSD-induced changes in psychological function when the drug is used one or more times per day for >4 days. Abrupt abstinence following continued use does not produce withdrawal signs or symptoms. There have been no clinical reports of death caused by the direct effects of LSD.

PHENCYCLIDINE (PCP)

Phencyclidine (PCP), a cyclohexylamine derivative, is widely used in veterinary medicine to briefly immobilize large animals and is sometimes described as a dissociative anesthetic. PCP binds to ionotropic *N*-methyl-d-aspartate (NMDA) receptors in the nervous system, blocking ion current through these channels. PCP is easily synthesized; its abusers are primarily young people and polydrug users. It is used orally, by smoking, by snorting, or by IV injection. It is also used as an adulterant in THC, LSD, amphetamine, or cocaine. The most common street preparation, *angel dust*, is a white granular powder that contains 50–100% percent of the drug. Low doses (5 mg) produce agitation, excitement, impaired motor coordination, dysarthria, and analgesia. Physical signs of intoxication may include horizontal or vertical nystagmus, flushing, diaphoresis, and hyperacusis. Behavioral changes include distortions of body image, disorganization of thinking, and feelings of estrangement. Higher doses of PCP (5–10 mg) may produce profuse salivation, vomiting, myoclonus, fever, stupor, or coma. PCP doses of ≥10 mg cause convulsions, opisthotonus, and decerebrate posturing, which may be followed by prolonged coma.

The diagnosis of PCP overdose is difficult because the patient's initial symptoms (anxiety, paranoia, delusions, hallucinations) may suggest an acute schizophrenic reaction. Confirmation of PCP use is possible by determination of PCP levels in serum or urine. PCP assays are available at most toxicologic centers. PCP remains in urine for 1–5 days following high-dose intake.

PCP overdose requires life-support measures, including treatment of coma, convulsions, and respiratory depression in an intensive care unit. There is no specific antidote or antagonist for PCP. PCP excretion from the body can be enhanced by gastric lavage and acidification of urine. Death from PCP overdose may occur as a consequence of some combination of pharyngeal hypersecretion, hyperthermia, respiratory depression, severe hypertension, seizures, hypertensive encephalopathy, and intracerebral hemorrhage.

Acute psychosis associated with PCP use is a psychiatric emergency since patients may be at high risk for suicide or extreme violence toward others. Phenothiazines should not be used for treatment because these drugs potentiate PCP's anticholinergic effects. Haloperidol (5 mg IM) has been administered on an hourly basis to induce suppression of psychotic behavior. PCP, like LSD and mescaline, produces vasospasm of cerebral arteries at relatively low doses. Chronic PCP use has been shown to induce insomnia, anorexia, severe social and behavioral changes, and, in some cases, chronic schizophrenia.

OTHER DRUGS OF ABUSE

A number of other pharmacologically diverse drugs of abuse are often referred to as "club drugs" because these are frequently used in bars, at concerts, and rave parties. Commonly abused club drugs include flunitrazepam, GHB, and ketamine and are described below. Methamphetamine, MDMA, and LSD are also considered club drugs and were described earlier in this chapter. Abuse of club drugs at high doses, especially in combination with alcohol, can be lethal and should be treated as a medical emergency. GHB and ketamine can be identified in blood, and flunitrazepam can be identified in urine and hair samples. Flunitrazepam and GHB toxicity can be treated with antagonists at benzodiazepine and $GABA_B$ receptors, respectively.

FLUNITRAZEPAM

Flunitrazepam (Rohypnol) is a benzodiazepine derivative primarily used for treatment of insomnia, but it has significant abuse potential because of its strong hypnotic, anxiolytic and amnesia-producing effects. It is a club drug commonly referred to as a "date-rape drug" or "roofies." The drug enhances $GABA_A$ receptor activity, and overdose can be treated with flumazenil, a benzodiazepine receptor antagonist. Flunitrazepam is typically used orally but can be snorted or injected. Concomitant use of alcohol or opiates is common, and this enhances the sedative and hypnotic effects of flunitrazepam and also the risk of motor vehicle accidents. Overdose can produce life-threatening respiratory depression and coma. Abrupt cessation after chronic use may result in a benzodiazepine withdrawal syndrome consisting of anxiety, insomnia, disordered thinking, and seizures.

GHB

Gamma-hydroxybutyric acid (Xyrem) is a sedative drug that is FDA-approved for the treatment of narcolepsy. It is classified as a club drug, is sometimes used in combination with alcohol or other drugs of abuse, and has been implicated in cases of date rape. GHB is usually taken orally and has no

distinctive color or odor. Its stimulant properties are attributed to agonist activity at the GHB receptor, but it also has sedative effects at high doses that reflect its activity at $GABA_B$ receptors. $GABA_B$ antagonists can reverse GHB's sedative effects, and opioid antagonists (naloxone, naltrexone) can attenuate GHB effects on dopamine release. Low doses of GHB may produce euphoria and disinhibition, whereas high doses result in nausea, agitation, convulsions, and sedation that can lead to unconsciousness and death from respiratory depression.

KETAMINE

Ketamine (Ketaset, Ketalar) is a dissociative anesthetic, similar to phencyclidine (PCP). In veterinary medicine, it is used for brief immobilization. In clinical medicine, it is used for sedation, analgesia, and to supplement anesthesia. Ketamine increases heart rate and blood pressure, with less respiratory depression than other anesthetics. Ketamine's popularity as a club drug appears to reflect its ability to induce a dissociative state and feelings of depersonalization, accompanied by intense hallucinations and subsequent amnesia. It can be administered orally, by smoking (usually in combination with tobacco and/or marijuana), or by IV or IM injection. Like PCP, it binds to NMDA receptors and acts as an uncompetitive NMDA antagonist. Ketamine has a complex profile of action and appears to be useful as an antidepressant in treatment-resistant patients and as an analgesic in chronic-pain patients. The extent to which chronic recreational use leads to memory impairment remains controversial.

POLYDRUG ABUSE

Although some drug abusers may prefer a particular drug, the concurrent use of multiple drugs is often reported. Polydrug abuse often involves substances that may have different pharmacologic effects from the preferred drug. For example, concurrent use of such dissimilar compounds as stimulants and opiates or stimulants and alcohol is common. The diversity of reported drug use combinations suggests that achieving a subjective change in state, rather than any particular direction of change (stimulation or sedation), may be the primary reinforcer in polydrug abuse. There is also evidence that intoxication with alcohol, opiates, and cocaine is associated with increased tobacco smoking. There are relatively few controlled studies of multiple drug interactions. However, the combined use of cocaine, heroin, and alcohol increases the risk for toxic effects and adverse medical consequences. One determinant of polydrug use patterns is the relative availability and cost of the drugs. For example, alcohol abuse, with its attendant medical complications, is one of the most serious problems encountered in former heroin addicts participating in methadone maintenance programs.

The physician must recognize that perpetuation of polydrug abuse and drug dependence is not necessarily a symptom of an underlying emotional disorder. Neither alleviation of anxiety nor reduction of depression accounts for initiation and perpetuation of polydrug abuse. Severe depression and anxiety are the consequences of polydrug abuse as frequently as they are the antecedents. Interestingly, some adverse consequences of drug use may be reinforcing and contribute to the continuation of polydrug abuse.

Adequate treatment of polydrug abuse, as well as other forms of drug abuse, requires innovative intervention programs. The first step in successful treatment is detoxification, a process that may be difficult when several drugs with different pharmacologic actions (e.g., alcohol, opiates, and cocaine) have been abused. Since patients may not recall or may deny simultaneous multiple drug use, diagnostic evaluation should always include urinalysis for qualitative detection of psychoactive substances and their metabolites. Treatment of polydrug abuse often requires hospitalization or inpatient residential care during detoxification and the initial phase of drug abstinence. When possible, specialized facilities for the care and treatment of drug-dependent persons should be used. Outpatient detoxification of polydrug abuse patients is likely to be ineffective and may be dangerous.

Drug abuse disorders often respond to effective treatment, but episodes of relapse may occur unpredictably. The physician should continue to assist patients during relapse and recognize that occasional recurrent drug use is not unusual in this complex behavioral disorder.

FURTHER READINGS

Brady KT et al (eds): *Women & Addiction: A Comprehensive Handbook*. New York, Guilford Press, 2009

Haber PS et al: Management of injecting drug users admitted to hospital. Lancet 384:1284, 2009

Herin DV et al: Agonist-like pharmacotherapy for stimulant dependence: Preclinical, human laboratory, and clinical studies. Ann N Y Acad Sci 1187:76, 2010[PMID: 20201847] [Full Text]

Lowinson JH et al (eds): *Substance Abuse: A Comprehensive Textbook*. Baltimore, Williams & Wilkins, 2005

Mello NK, Mendelson JH: Cocaine, hormones and behavior: Clinical and preclinical studies, in *Hormones, Brain and Behavior*, 2nd ed, DW Pfaff et al (eds). San Diego, Academic Press 2009, pp

3081–3139

Principles of Drug Addiction Treatment: A Research-Based Guide , 2nd ed. National Institute on Drug Abuse, NIH Publication No. 09-4180, U.S. Department of Health and Human Services, 2009

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NICOTINE ADDICTION: INTRODUCTION

The use of tobacco leaf to create and satisfy nicotine addiction was introduced to Columbus by Native Americans and spread rapidly to Europe. Use of tobacco as cigarettes, however, only became popular in the twentieth century and so is a modern phenomenon, as is the epidemic of disease caused by this form of tobacco use.

Nicotine is the principal constituent of tobacco responsible for its addictive character, but other smoke constituents and behavioral associations contribute to the strength of the addiction. Addicted smokers regulate their nicotine intake by adjusting the frequency and intensity of their tobacco use both to obtain the desired psychoactive effects and avoid withdrawal.

Unburned cured tobacco used orally contains nicotine, carcinogens, and other toxicants capable of causing gum disease, oral and pancreatic cancers, and an increase in the risk of heart disease. When tobacco is burned, the resultant smoke contains, in addition to nicotine, more than 4000 other compounds that result from volatilization, pyrolysis, and pyrosynthesis of tobacco and various chemical additives used in making different tobacco products. The smoke is composed of a fine aerosol and a vapor phase; aerosolized particles are of a size range that results in deposition in the airways and alveolar surfaces of the lungs. The aggregate of particulate matter, after subtracting nicotine and moisture, is referred to as tar.

The alkaline pH of smoke from blends of tobacco utilized for pipes and cigars allows sufficient absorption of nicotine across the oral mucosa to satisfy the smoker's need for this drug. Therefore, smokers of pipes and cigars tend not to inhale the smoke into the lung, confining the toxic and carcinogenic exposure (and the increased rates of disease) largely to the upper airway for most users of these products. The acidic pH of smoke generated by the tobacco used in cigarettes dramatically reduces absorption of nicotine in the mouth, necessitating inhalation of the smoke into the larger surface of the lungs in order to absorb quantities of nicotine sufficient to satisfy the smoker's addiction. The shift to using tobacco as cigarettes, with resultant increased deposition of smoke in the lung, has created the epidemic of heart disease, lung disease, and lung cancer that dominates the current disease manifestations of tobacco use.

Several genes have been associated with nicotine addiction. Some reduce the clearance of nicotine, and others have been associated with an increased likelihood of becoming dependent on tobacco and other drugs as well as a higher incidence of depression. Rates of smoking cessation have increased, and rates of nicotine addiction have decreased dramatically, since the mid-1950s, suggesting that factors other than genetics are important. It is likely that genetic susceptibility can influence the probability that adolescent experimentation with tobacco will lead to addiction as an adult.

Adult cigarette smoking prevalence has declined to about 20% in the United States, with similar declines in Canada and most European countries. Male smoking prevalence is falling but remains high in most Asian countries, with increasing smoking prevalence among women in those countries. The highest rates of smoking and least cessation are observed in eastern European countries. Of particular concern is the rapidly rising smoking rate observed in the developing world. The recently ratified World Health Organization Framework Convention on Tobacco Control is encouraging effective tobacco control approaches in these countries with the hope of preventing a future epidemic of tobacco-related illness.

DISEASE MANIFESTATIONS OF CIGARETTE SMOKING

More than 400,000 individuals die prematurely each year in the United States from cigarette use; this represents almost one of every five deaths in the United States. Approximately 40% of cigarette smokers will die prematurely due to cigarette smoking unless they are able to quit.

The major diseases caused by cigarette smoking are listed in **Table 395-1**. The ratio of smoking-related disease rates in smokers compared to never smokers (relative risk) is greater at younger ages, particularly for coronary artery disease and stroke. At older ages, the background rate of disease in nonsmokers increases, diminishing the fractional contribution of smoking and the relative risk; however, absolute excess rates of disease mortality found in smokers compared to nonsmokers increase with increasing age. The organ damage caused by smoking and the number of smokers who die from smoking are both greater among the elderly, as one would expect from a process of cumulative injury.

Table 395-1 Relative Risks for Current Smokers of Cigarettes

	Current Smokers	
Disease or Condition	**Males**	**Females**
Coronary Heart Disease		
Age 35–64	2.8	3.1
Age ?65	1.5	1.6
Cerebrovascular Disease		
Age 35–64	3.3	4
Age ?65	1.6	1.5
Aortic aneurysm	6.2	7.1
Chronic airway obstruction	10.6	13.1
Cancer		
Lung	23.3	12.7
Larynx	14.6	13
Lip, oral cavity, pharynx	10.9	5.1
Esophagus	6.8	7.8
Bladder, other urinary organs	3.3	2.2
Kidney	2.7	1.3
Pancreas	2.3	2.3
Stomach	2	1.4
Cervix		1.6
Acute myeloid leukemia	1.4	1.4
Sudden infant death syndrome		2.3
Infant respiratory distress syndrome		1.3
Low birth weight at delivery		1.8

Cardiovascular Diseases

Cigarette smokers are more likely than nonsmokers to develop both large-vessel atherosclerosis and small-vessel disease. Approximately 90% of peripheral vascular disease in the nondiabetic population can be attributed to cigarette smoking, as can ~50% of aortic aneurysms. In contrast, 20–30% of coronary artery disease and ~10% of occlusive cerebrovascular disease are caused by cigarette smoking. There is a multiplicative interaction between cigarette smoking and other cardiac risk factors such that the increment in risk produced by smoking among individuals with hypertension or elevated serum lipids is substantially greater than the increment in risk produced by smoking for individuals without these risk factors.

In addition to its role in promoting atherosclerosis, cigarette smoking also increases the likelihood of myocardial infarction and sudden cardiac death by promoting platelet aggregation and vascular occlusion. Reversal of these effects on coagulation may explain the rapid benefit of smoking cessation for a new coronary event demonstrable among those who have survived a first myocardial infarction. This effect may also explain the substantially higher rates of graft occlusion among continuing smokers following vascular bypass surgery for cardiac or peripheral vascular disease.

Cessation of cigarette smoking reduces the risk of a second coronary event within 6–12 months; rates of first myocardial infarction and death from coronary heart disease also decline within the first few years following cessation. After 15 years of abstinence, the risk of a new myocardial infarction or death from coronary heart disease in former smokers is similar to that for those who have never smoked.

Cancer

Tobacco smoking causes cancer of the lung; oral cavity; naso-, oro-, and hypopharynx; nasal cavity and paranasal sinuses; larynx; esophagus; stomach; pancreas; liver; kidney (body and pelvis); ureter; urinary bladder; and uterine cervix and also causes myeloid leukemia. There is evidence suggesting that cigarette smoking may play a role in increasing the risk of colorectal, hepatocellular, and possibly premenopausal breast cancer. There is no association with postmenopausal breast cancer. There does not appear to be a causal link between cigarette smoking and cancer of the endometrium, and there is a lower risk of uterine cancer among postmenopausal women who smoke. The risks of cancer increase with the increasing number of cigarettes smoked per day and with increasing duration of smoking. Additionally, there are synergistic interactions between cigarette smoking and alcohol use for cancer of the oral cavity, esophagus, and possibly lung. Several occupational exposures synergistically increase lung cancer risk among cigarette smokers, most notably occupational asbestos

and radon exposure.

Cessation of cigarette smoking reduces the risk of developing cancer relative to continuing smoking, but even 20 years after cessation there is a modest persistent increased risk of developing lung cancer.

Respiratory Disease

Cigarette smoking is responsible for 90% of chronic obstructive pulmonary disease. Within 1–2 years of beginning to smoke regularly, many young smokers will develop inflammatory changes in their small airways, although lung function measures of these changes do not predict development of chronic airflow obstruction. After 20 years of smoking, pathophysiologic changes in the lungs develop and progress proportional to smoking intensity and duration. Chronic mucous hyperplasia of the larger airways results in a chronic productive cough in as many as 80% of smokers >60 years. Chronic inflammation and narrowing of the small airways and/or enzymatic digestion of alveolar walls resulting in pulmonary emphysema can result in reduced expiratory airflow sufficient to produce clinical symptoms of respiratory limitation in ~15–25% of smokers.

Changes in the small airways of young smokers will reverse after 1–2 years of cessation. There may also be a small increase in measures of expiratory airflow following cessation among individuals who have developed chronic airflow obstruction, but the major change following cessation is a slowing of the rate of decline in lung function with advancing age rather than a return of lung function toward normal.

Pregnancy

Cigarette smoking is associated with several maternal complications of pregnancy: premature rupture of membranes, abruptio placentae, and placenta previa; there is also a small increase in the risk of spontaneous abortion among smokers. Infants of smoking mothers are more likely to experience preterm delivery, have a higher perinatal mortality rate, be small for their gestational age, and have higher rates of infant respiratory distress syndrome; they are more likely to die of sudden infant death syndrome, and appear to have a developmental lag for at least the first several years of life.

Other Conditions

Smoking delays healing of peptic ulcers; increases the risk of osteoporosis, senile cataracts, and macular degeneration; and results in premature menopause, wrinkling of the skin, gallstones and cholecystitis in women, and male impotence.

Environmental Tobacco Smoke

Long-term exposure to environmental tobacco smoke increases the risk of lung cancer and coronary artery disease among nonsmokers. It also increases the incidence of respiratory infections, chronic otitis media, and asthma in children as well as causing exacerbation of asthma in children. Some evidence suggests that environmental tobacco smoke exposure may increase the risk of premenopausal breast cancer, but this relationship remains controversial.

PHARMACOLOGIC INTERACTIONS

Cigarette smoking may interact with a variety of other drugs **(Table 395-2)**. Cigarette smoking induces the cytochrome P450 system, which may alter the metabolic clearance of drugs such as theophylline. This may result in inadequate serum levels in smokers as outpatients when the dosage is established in the hospital under nonsmoking conditions. Correspondingly, serum levels may rise when smokers are hospitalized and not allowed to smoke. Smokers may also have higher first-pass clearance for drugs such as lidocaine, and the stimulant effects of nicotine may reduce the effect of benzodiazepines or beta blockers.

Table 395-2 Interactions of Smoking and Prescription Drugs

Drug	Interaction
Benzodiazepines Beta blockers	Less sedation Reduced lowering of heart rate and blood pressure
Caffeine	Faster metabolic clearance
Chlorpromazine	Decreased serum concentrations[a]
Clomipramine	Decreased serum concentrations[a]
Clozapine	Decreased serum concentrations[a]
Dextropropoxyphene	Less analgesia

Estrogens (oral)	Increased hepatic clearance
Flecainide	Increased first-pass clearance
Fluphenazine	Decreased serum concentrations[a]
Fluvoxamine	Decreased serum concentrations[a]
Haloperidol	Decreased serum concentrations[a]
Heparin	Faster clearance
Imipramine	Decreased serum concentrations[a]
Insulin	Delayed absorption due to skin vasoconstriction
Lidocaine	Increased first-pass clearance
Mexiletine	Increased first-pass clearance
Olanzapine	Faster clearance
Pentazocine	Less analgesia, possible increased clearance
Propranolol	Increased first-pass clearance
Tacrine	Faster metabolic clearance
Theophylline	Faster metabolic clearance
Thiothixene	Faster metabolic clearance
Trazodone	Decreased serum concentrations[a]

[a]Clinical implications uncertain.

OTHER FORMS OF TOBACCO USE

Other major forms of tobacco use are moist snuff deposited between the cheek and gum, chewing tobacco, pipes and cigars, and recently bidi (tobacco wrapped in tendu or temburni leaf; commonly used in India), clove cigarettes, and water pipes. Oral tobacco use leads to gum disease and can result in oral and pancreatic cancer as well as heart disease, with dramatic differences in the risks evident for products used in Africa and Asia as compared to those in the U.S. and Europe.

All forms of burned tobacco generate toxic and carcinogenic smoke similar to that of cigarette smoke. The differences in disease consequences of use relate to frequency of use and depth of inhalation. The risk of upper airway cancers is similar among cigarette and cigar smokers, while those who have smoked only cigars have a much lower risk of lung cancer, heart disease, and chronic obstructive pulmonary disease. However, cigarette smokers who switch to pipes or cigars do tend to inhale the smoke, increasing their risk; and it is likely that comparable inhalation and frequency of exposure to tobacco smoke from any of these forms of tobacco use will lead to comparable disease outcomes.

A resurgence of cigar, bidi, and water pipe use among adolescents of both genders has raised concerns that these older forms of tobacco use are once again causing a public health problem. A variety of devices are currently sold that deliver nicotine by electronically heating materials containing nicotine, the so-called electronic cigarettes. While these devices are marketed as substitutes for cigarettes and as cessation tools, little is known about their composition and evidence suggests that users may not absorb much nicotine from these devices.

LOWER TAR AND NICOTINE CIGARETTES

Filtered cigarettes with lower machine-measured yields of tar and nicotine commonly use ventilation holes in the filters and other engineering designs to artificially lower the machine measurements. Smokers compensate for the lowered nicotine delivery by changing the manner in which they puff on the cigarette or the number of cigarettes smoked per day and preserve their intake of nicotine (and tar). There is no meaningful disease-reduction benefit for smokers who switch to lower-yield cigarettes, and smokers should be discouraged from thinking of low-yield cigarettes as an alternative to cessation.

CESSATION

The process of stopping smoking is often a cyclical one, with the smoker sometimes making multiple attempts to quit and failing before finally being successful. Approximately 70–80% of smokers would like to quit smoking, approximately one-third of current smokers attempt to quit each year, and 90% of these unassisted attempts fail. Clinician-based smoking interventions should encourage smokers to try

to quit and to use different forms of cessation assistance with each new cessation attempt rather than focusing exclusively on immediate cessation at the time of the first visit.

Advice from a physician to quit smoking, particularly around an acute illness, is a powerful trigger for cessation attempts, with up to half of patients who are advised to quit making a cessation effort. Other triggers include the cost of cigarettes, media campaigns, and changes in rules to restrict smoking in the workplace.

PHYSICIAN INTERVENTION (TABLE 395-3)

Table 395-3 Clinical Practice Guidelines

Physician Actions
Ask: Systematically identify all tobacco users at every visit
Advise: Strongly urge all smokers to quit
Identify smokers willing to quit
Assist the patient in quitting
Arrange follow-up contact
Effective Pharmacologic Interventions[a]
First-line therapies
Nicotine gum (1.5)
Nicotine patch (1.9)
Nicotine nasal inhaler (2.7)
Nicotine oral inhaler (2.5)
Nicotine lozenge (2.0)
Bupropion (2.1)
Varenicline (2.7)
Second-line therapies
Clonidine (2.1)
Nortriptyline (3.2)
Other Effective Interventions[a]
Physician or other medical personnel counseling (10 min) (1.3)
Intensive smoking cessation programs (at least 4–7 sessions of 20- to 30-min duration lasting at least 2 and preferably 8 weeks) (2.3)
Clinic-based smoking status identification system (3.1)
Counseling by nonclinicians and social support by family and friends
Telephone counseling (1.2)

[a]Numerical value following the intervention is the multiple for cessation success compared to no intervention.

All patients should be asked whether they smoke, their past experience with quitting, and whether they are currently interested in quitting. Even those who are not interested in quitting should be encouraged and motivated to quit; provided a clear, strong, and personalized physician message that smoking is an important health concern; and offered assistance if they become interested in quitting in the future. There is a relationship between the amount of assistance a patient is willing to accept and the success of the cessation attempt. For those interested in quitting, a quit date should be negotiated, usually not the day of the visit but within the next few weeks, and a follow-up contact by office staff around the time of the quit date should be provided.

There are a variety of nicotine-replacement products, including over-the-counter nicotine patches, gum, and lozenges, as well as nicotine nasal and oral inhalers available by prescription. These products can be used for up to 3–6 months, and some products are formulated to allow a gradual step-down in dosage with increasing duration of abstinence. Antidepressants such as bupropion (300 mg in divided doses for up to 6 months) have also been shown to be effective, as has varenicline, a partial agonist for the nicotinic acetylcholine receptor (initial dose 0.5 mg daily increasing to 1 mg twice daily at day 8; treatment duration up to 6 months). Severe psychiatric symptoms, including suicidal ideation, have been reported with varenicline, resulting in an FDA-mandated warning and a recommendation for closer therapeutic supervision, but evidence to establish the frequency of these responses and the specificity of their association with varenicline remains unclear. Some evidence supports the combined use of nicotine-replacement therapy (NRT) and antidepressants as well as the use of gum or lozenges for acute cravings in patients using patches. Pretreatment with antidepressants or varenicline is

recommended for 1–2 weeks prior to the quit date, and pretreatment with nicotine-replacement products is also being explored. NRT is provided in different dosages, with higher doses being recommended for more intense smokers. Clonidine or nortriptyline may be useful for patients who have failed on first-line pharmacologic treatment or who are unable to use other therapies. Antidepressants are more effective in those with a history of depression symptoms.

Current recommendations are to offer pharmacologic treatment, usually with NRT or varenicline, to all who will accept it and to provide counseling and other support as a part of the cessation attempt. Currently, indications approved by the U.S. Food and Drug Administration for NRT products limit them to short-term use in conjunction with a cessation attempt. Nevertheless, it is not uncommon for individual smokers to use these products, particularly the products available over the counter, for longer durations and sometimes with no intent to stop. There are some data to suggest that longer-term use of NRT may enable cessation in some smokers who are unable to quit with shorter duration use and that some individuals are able to achieve abstinence from tobacco through use of NRT chronically. It is useful for the practicing physician to recognize that these patterns of use exist among smokers trying to quit and that they may contribute to successful abstinence in some smokers. Cessation advice alone by a physician or his or her staff is likely to increase success compared with no intervention; a more comprehensive approach with advice, pharmacologic assistance, and counseling can increase cessation success by almost threefold.

Incorporation of cessation assistance into a practice requires a change of the care delivery infrastructure. Simple changes include (1) adding questions about smoking and interest in cessation on patient-intake questionnaires, (2) asking patients whether they smoke as part of the initial vital sign measurements made by office staff, (3) listing smoking as a problem in the medical record, and (4) automating follow-up contact with the patient on the quit date. These changes are essential to institutionalizing smoking intervention within the practice setting; without this institutionalization, the best intentions of physicians to intervene with their patients who smoke are often lost in the time crush of a busy practice.

PREVENTION

Approximately 90% of individuals who become cigarette smokers initiate the behavior during adolescence. Factors that promote adolescent initiation are parental or older-generation cigarette smoking, tobacco advertising and promotional activities, the availability of cigarettes, and the social acceptability of smoking. The need for an enhanced self-image and to imitate adult behavior is greatest for those adolescents who have the least external validation of their self-worth, which may explain in part the enormous differences in adolescent smoking prevalence by socioeconomic and school performance strata.

Prevention of smoking initiation must begin early, preferably in the elementary school years. Physicians who treat adolescents should be sensitive to the prevalence of this problem. Physicians should ask all adolescents whether they have experimented with tobacco or currently use tobacco, reinforce the fact that most adolescents and adults do not smoke, and explain that all forms of tobacco are both addictive and harmful.

FURTHER READINGS

Agency for Health Research and quality: *Treating Tobacco Use and Dependence: 2008 Update*. Clinical Practice Guideline, Public Health Service, DHHS, 2008 *http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=hsahcpr&part=A28163*

Benowitz N: Nicotine addiction. NEJM 362:2295, 2010 [PMID: 20554984] [Full Text]

International Agency for Research on Cancer: *Reversal of Risk After Quitting Smoking*. IARC Handbooks of Cancer Prevention, Tobacco Control, Vol 11. Lyon, France, IARC, 2006

US Department of Health and Human Services: *The Health Consequences of Tobacco Use: A Report of the Surgeon General*. National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2004

———: *How Tobacco Smoke Causes Disease—The Biology and Behavioral Basis for Tobacco Attributable Disease*: *A Report of the Surgeon General*. National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2010

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

NEUROPSYCHIATRIC ILLNESSES IN WAR VETERANS: INTRODUCTION

Neuropsychiatric sequelae are common in combat veterans. Advances in personal protective body armor, armored vehicles, battlefield resuscitation, and the speed of evacuation to tertiary care have considerably improved the survivability of battlefield injuries, resulting in a greater awareness of the "silent wounds" associated with service in a combat zone. Although psychiatric and neurologic problems have been well documented in veterans of prior wars, the conflicts in Iraq and Afghanistan have been unique in terms of the level of commitment by the U.S. Department of Defense (DoD) and Department of Veterans Affairs (VA), Veterans Health Administration (VHA) to support research as the wars have unfolded, and to utilize that knowledge to guide population-level screening, evaluation, and treatment initiatives.

These conflicts, like previous ones, have produced hundreds of thousands of combat veterans, many of whom have received or will need care in government and civilian medical facilities. Studies have shown that service in the Iraq and Afghanistan theaters is associated with significantly elevated rates of mental disorders. Two conditions in particular have been labeled the signature injuries related to these wars: posttraumatic stress disorder (PTSD) and mild traumatic brain injury (mTBI)—also known as concussion. Although particular emphasis will be given in this chapter to PTSD and concussion/mTBI, it is important to understand that service in all wars is associated with a number of health concerns that coexist and overlap, and a multidisciplinary patient-centered approach to care is necessary.

EPIDEMIOLOGY OF WAR-RELATED PSYCHOLOGICAL AND NEUROLOGIC CONDITIONS

Service members from the current decade of war have faced multiple deployments to two very different high-intensity combat theaters, and the cumulative strain has negatively impacted marriages, parenting, educational goals, and civilian occupations. The stresses of service in these conflicts have led to a significant increase in the rate of suicide in personnel from the two branches of service involved in the greatest level of ground combat (army, marines).

Service in a war-zone can involve extreme physical stress in austere environments, prolonged sleep deprivation, physical injury, exposure to highly life-threatening events and hazards such as explosive devices, sniper fire, ambushes, indirect fire from rockets and mortars, and chemical pollutants. Certain events such as loss of a close friend in combat, leave indelible scars. All of these experiences have additive effects on health, likely mediated through physiologic mechanisms involving dysregulation of neuroendocrine and autonomic nervous system (ANS) functions.

Veterans of virtually all wars have reported elevated rates of generalized and multisystem physical, cognitive, and psychological health concerns that often become the focus of treatment months or years after returning home. These multisystem health concerns include sleep disturbance, memory and concentration problems, headaches, musculoskeletal pain, gastrointestinal symptoms (including gastroesophageal reflux), residual effects of war-time injuries, fatigue, anger, hyperarousal symptoms, high blood pressure, rapid heart rate (sometimes associated with panic symptoms), sexual problems, and symptoms associated with PTSD and depression. In order to provide optimal care to veterans with these symptoms, it is important to understand how the symptoms interrelate, and to consider the possibility that there may be underlying combat-related physiologic effects.

Postwar Symptoms

The overlapping and multisystem health symptoms reported by warriors from every generation have been given different labels, and have led to debates among medical professionals as to whether these are mediated primarily by physical or psychological causes. For example, World War I produced extensive debate about whether "shell shock," diagnosed in more than 80,000 British soldiers, was neurologic ("commotional" from the brain being shaken in the skull by concussive blasts) or psychological ("emotional" or "neurasthenia") in origin. World War II veterans were said to suffer from "battle fatigue," Korean war veterans developed "combat stress reactions," and Vietnam veterans developed the "post-Vietnam syndrome." The role of environmental exposure (e.g., agent orange) and psychological causes (alcohol addiction, drug addiction, and PTSD) continue to be debated.

Gulf War I led to extensive debates as to whether Gulf War syndrome, also known as multisystem illness, was best explained by environmental exposures (e.g., oil fires, depleted uranium, nerve gas, multiple vaccinations) or the psychological stress of deployment to a war zone where there was anticipation of high casualty rates from chemical and biologic weapons and stressful training exercises involving the use of impermeable full-body protective uniforms (made from rubber, vinyl, charcoal impregnated polyurethane, and other materials) in desert conditions under extreme temperatures. Although no clinical syndrome was ever definitively confirmed among the nearly 1 million service members who deployed in 1990–1991, studies consistently found that military personnel who served in the Gulf experienced

elevations in generalized symptoms across all health domains (e.g. physical, cognitive, neurologic, psychological) compared with service members who deployed elsewhere or did not deploy. In addition, there is good evidence that deployment to the Persian Gulf region during this period was associated with subsequent development of PTSD; other psychiatric disorders including generalized anxiety disorder, depression, and substance abuse, particularly alcohol abuse (Chap. 392); functional gastrointestinal symptoms such as irritable bowel syndrome (Chap. 296); and chronic fatigue syndrome (Chap. 389).

The conflicts in Iraq and Afghanistan have led to similar debates as to whether postwar symptoms such as headaches, irritability, sleep disturbance, dizziness, and concentration problems are best attributed to concussion/mTBI or to PTSD. Several studies have shown that either PTSD or depression explains the majority of the postdeployment "postconcussive" symptoms attributed to concussion/mTBI, a finding not well received by some experts in traumatic brain injury (TBI) but consistent with civilian studies on risk factors for developing persistent symptoms after concussion. As in past wars, it has taken years to understand how PTSD and concussion/mTBI interrelate with other deployment-related health concerns, and the implications for designing effective evaluation and treatment strategies.

Veterans understandably may become angry at the suggestion that their postwar health concerns are stress-related or psychological, and thus it is necessary for primary care professionals to be sensitive to this concern.

PTSD

PTSD is the most common mental disorder documented following war-zone service. Studies from the conflicts in Iraq and Afghanistan have found PTSD prevalence rates of 2–6% before deployment (comparable to civilian general population samples) and rates of 6–20% postdeployment, depending primarily on the level of combat frequency and intensity. Many other veterans experience subclinical PTSD symptoms after war-zone service, sometimes termed post-traumatic stress (PTS) or combat stress. These subclinical symptoms can contribute to distress and affect health, even if overall functioning is not as impaired as in the full disorder.

PTSD is defined by the American Psychiatric Association as persistent (>1 month) symptoms occurring after a life-threatening traumatic event in which there was an immediate response of fear, helplessness, or horror. The symptoms must be associated with significant distress or impairment in social or occupational functioning. Symptoms are grouped into three categories: (1) re-experiencing symptoms in which the person has nightmares, flashbacks, or intrusive thoughts and memories connected with the traumatic event; (2) hyperarousal symptoms in which the person is physiologically revved up, hyperalert, startles easily, and experiences sleep disturbance, anger, and/or concentration problems; and (3) avoidance symptoms where the person loses interest in things that previously brought enjoyment, and avoids places, situations, or other stimuli that serve as reminders of the traumatic event (for example a crowded mall that triggers heightened alertness to threat). Additional symptoms, currently categorized with the avoidance cluster, but which will likely become a fourth category in future definitions of PTSD, include emotional numbing, feeling distant or cutoff from others, and a foreshortened sense of future (Chap. 391). While PTSD is a clinical symptom-based case definition, it is best to think of PTSD not as an emotional or psychological/psychiatric condition, but rather as a physiologic-based response to life-threatening trauma that is associated with physical, cognitive, emotional, and psychological symptoms.

PTSD has strong biologic correlates, based in fear-conditioning responses to threat and responses to extreme stress involving neuroendocrine dysregulation and ANS reactivity. Numerous studies have shown that PTSD is highly correlated with generalized physical and cognitive symptoms—including hypertension, chronic pain, and cardiovascular disease—as well as cell-mediated immune dysfunction and shortened life expectancy. PTSD is frequently comorbid with other mental disorders such as major depressive disorder, generalized anxiety, substance use disorders (SUDs), as well as risky behaviors (e.g., aggression, accidents); it has been estimated that up to 80% of patients with PTSD exhibit one or more comorbid conditions. Misuse of alcohol or substances is most prevalent, often reflecting self-medication. PTSD is also associated with tolerance and withdrawal symptoms related to prescription pain and sleep medications; as well as nicotine dependence (Chap. 395).

Clinicians should understand the limitations of the definition of PTSD when applied to responses to trauma occurring in the occupational context of military service (similar to police, firefighters, and other first responders). Service members are trained to respond to traumatic events, and relatively rarely report fear, helplessness, or horror, which are characteristic responses of civilian victims of trauma. In addition, the reactions that are labeled as symptoms of PTSD are based on the adaptive survival responses of warriors in a combat environment. For example, physiologic hyperarousal, use of anger, and being able to shut down other emotions are very useful skills in combat, and can be present prior to traumatic events when there is tough realistic training. These responses only become symptoms when they impair functioning after warriors return home.

Concussion/mTBI

TBI (Chap. 378) gained increased recognition during the conflicts in Iraq and Afghanistan because of the widespread exposure of troops to improvised explosive devices. Contributing to heightened concern were high prevalence estimates of deployment-related TBI that did not distinguish concussion/mTBI from moderate or severe TBI; data from animal models of blast suggesting that explosions may cause a different kind of concussion associated with inflammatory changes; and speculation that repetitive blast

exposure may lead to future dementia, based on case series of professional athletes (e.g., boxers, football players) exposed to highly repetitive injuries linked to chronic traumatic encephalopathy (previously termed *dementia pugilistica*). Many veterans of Iraq and Afghanistan reported experiencing multiple concussions during deployments, and many also reported ignoring concussions and not seeking treatment at the time of injury.

TBI includes closed and penetrating head injuries; closed head injuries are categorized as mild (mTBI or concussion), moderate, or severe based on the duration of loss of consciousness, duration of post-traumatic amnesia, and the Glasgow coma score (GCS) (Table 378-2). Several studies have estimated that 10–20% of all military personnel deployed to Iraq or Afghanistan sustained one or more concussion/mTBI events during deployment, most commonly from exposure to blasts.

Although there is a neurophysiologic continuum of injury, there are stark clinical and epidemiologic distinctions between concussion/mTBI and moderate or severe TBI (Table e48–1). Concussion/mTBI is defined as a blow or jolt to the head that results in brief loss of consciousness (LOC) for <30 minutes (most commonly only a few seconds to minutes), posttraumatic amnesia (PTA) of <24 hours (most commonly <1 hour), or transient alteration in consciousness (AOC) without loss of consciousness. The majority of concussions in Iraq or Afghanistan have involved AOC without LOC or PTA (which soldiers may refer to as getting their "bell rung"). GCSs are usually normal (15 out of 15). Concussion is treated with rest to allow the brain time to heal, and almost never results in air-evacuation from the battlefield unless there are other injuries.

Table e48–1 Comparison between Concussion/mTBI and Moderate/Severe TBI

	Mild TBI (concussion)	**Mod/Severe TBI**
Clinical case definition		
Loss of consciousness	<30 min (usually few seconds to minutes)	≥30 min to indefinite
Altered consciousness	<24 h (usually <30 min)	≥24 h to indefinite
Posttraumatic amnesia	<24 h (usually <30 min)	≥24 h to indefinite
Glasgow Coma Score	13–15 (usually 15)	As low as 3
Focal neurologic signs	None or transient	Frequently present
Traditional neuroimaging (CT/MRI)	Usually negative	Diagnostic
Clinical usefulness of neurocognitive testing after acute injury period	Usually inconclusive	Essential and valuable
Neuronal cell damage	Metabolic/ionic processes associated with axonal swelling, which can lead to disconnection	Direct injury effects plus metabolic/ionic effects
Sequelae, natural history, and recovery	Full recovery expected in majority of individuals; No consensus on natural history; the percent who develop persistent symptoms is debated	Based directly on injury characteristics; may be severely disabling
Predictors of persistent postconcussive symptoms or disability	Intensely debated Risk factors found to be most predictive include psychiatric conditions (e.g., depression, PTSD) and negative expectations	Not debated, predictors are directly related to injury severity and clinical progress with rehabilitation treat-ment

In contrast, moderate, severe, or penetrating TBI, which account for <1% of all battlefield head injuries in Iraq and Afghanistan, are characterized by LOC ≥30 minutes (up to permanent coma), PTA ≥24 hours (also may be permanent), and GCSs as low as 3 (the minimum value). These virtually always result in air-evacuation from the battlefield and can result in severe long-term neurologic impairment and requirement for rehabilitative care.

Symptoms following concussion/mTBI can include headache; fatigue; concentration, memory, or attention problems; sleep disturbance; irritability; balance difficulties; and tinnitus, among other symptoms. Recovery is usually rapid, with symptoms usually resolving in a few hours to days, but in a small percentage of patients' symptoms may persist for a longer period or become chronic (referred to as persistent "postconcussive symptoms" or "PCS").

Establishing a clear causal connection between a deployment concussion injury and persistent PCS months or years after return from deployment is difficult and often confounded by other postwar conditions

that cause similar symptoms, including injuries not involving the head, other medical disorders, sleep disorders, PTSD, depression, substance use disorders, chronic pain, and the general physiologic effects of wartime service. Contributing to the difficulty in establishing causation is the fact that the concussion/mTBI case definition refers only to the acute injury event and lacks symptoms, time-course or impairment; case definitions for persistent postconcussion syndrome are not well validated. Several studies found that PTSD was a much stronger predictor of postdeployment PCS after combat deployment than concussions/mTBIs, and one study even found that objective neuropsychological impairment after deployment was entirely explained by PTSD. These data do not minimize the importance of concussion/mTBI, but highlight the complex interrelationships of war-related health problems.

Studies of veterans who sustained concussions in Iraq or Afghanistan have suggested that blast mechanisms produce similar clinical outcomes as nonblast mechanisms, in contrast to expectations based on animal models. An explosion can produce serious injury from rapid atmospheric pressure changes (primary blast wave mechanism), as well as from flying debris (secondary blast mechanism) or by being thrown into a hard object (tertiary blast mechanism). Secondary and tertiary mechanisms are similar to other mechanical mechanisms of concussions sustained during accidents. The possibility of a unique head injury from the primary blast wave in otherwise uninjured soldiers appears to be low, but cannot be discounted.

Multisystem health problems that lack clear case definitions do not lend themselves well to uniform public health strategies such as screening. Nevertheless, mass population screening for concussion/mTBI was mandated for all U.S. service members returning from Iraq or Afghanistan and all veterans presenting for care at VA health care facilities. These screening processes, which attempt to apply the acute concussion case definition (lacking symptoms, time-course, or impairment) months or years after injury, led to sharp criticism that they were encouraging clinicians to misattribute common postwar symptoms to concussion/mTBI.

Management is largely symptom-focused, and ideally carried out within primary-care based structures of care. Optimal care avoids unnecessary specialty referrals, use of nonevidence-based interventions, or poor communication that results in negative expectations. Concussion research has shown that negative expectations are one of the most important risk factors for persistent symptoms.

While many questions remain regarding the long-term health effects of concussions (particularly multiple concussions) sustained during deployment, these are important battlefield injuries that require careful attention. However, they need to be addressed within the context of all other war-related health concerns.

STIGMA AND BARRIERS TO CARE

Adding to the complexity of treating veterans is stigma and other barriers to care. Despite extensive education efforts among military leaders and service members, perceptions of stigma have shown little change over the many years of war; warriors are concerned that they will be perceived as weak by peers or leaders if they seek care. Studies showed that less than one-half of service members with serious mental health problems receive needed care. Many factors contribute to this, including the pervasive nature of stigma in society in general (particularly among men), the critical importance of group cohesiveness of military teams, the nature of avoidance symptoms in PTSD, and sometimes skepticism that mental health professionals will be able to help.

Approach to the Patient: Evaluation of Veterans with Neuropsychiatric Health Concerns

Evaluation should begin with a careful occupational history as part of the routine medical evaluation; this includes the number of years served, military occupation, deployment locations and dates, illnesses or injuries resulting from service, and significant combat traumatic experiences that may be continuing to affect the individual **(Table e48–2)**. The clinician should evaluate the degree to which the patient's current difficulties reflect the normal course of readjusting after the intense occupational experience of combat. It is helpful to reinforce the many strengths associated with being a professional in the military: courage, honor, service to country, resiliency in combat, leadership, ability to work as cohesive workgroup with peers, and demonstrated skills in handling extreme stress.

Table e48–2 Specific Considerations in the Medical Evaluation of Veterans

Occupational context of health concerns	Deployment locations and dates, combat experiences or other deployment stressors, frequent moves, separations from family, impact of deployment on civilian occupation (for reservists)
Medical problems during deployment	History of deployment-related injuries (including concussions), environmental exposures, sleep pattern during deployment, use of caffeine/energy drinks, use of other substances
Current medical history	Current symptoms, level of chronic pain, sleep problems, evidence of persistent physiologic hyperarousal (hypertension, tachycardia, panic symptoms, concentration/memory problems, irritability/anger, sleep disturbance), chronic use of caffeine or energy drinks, chronic use of nonsteroidal anti-inflammatory medications, chronic use of narcotic pain medications, chronic use of nonbenzodiazepine sedative-hypnotic medications, chronic use of benzodiazepines for sleep or anxiety.
Mental health	Screen for PTSD, major depressive disorder. Ask about suicidal or homicidal ideation.

assessment	intent, or plans, as well as access to firearms
Alcohol/substance use	Screen for alcohol and substance use disorders, quantity and frequency of use and evidence of tolerance. Inquire about "self-medication" (e.g., use of alcohol for sleep, "calm down," or "forget" war-zone experiences)
Functional impairment	Impact of current symptoms on social and occupational functioning. High-risk behaviors (e.g., drinking and driving, reckless driving, aggression)
Social support, impact of military service on marriage and family	Level of social support. Readjustment stress on spouse, children, or other family members.

One of the challenges with current medical practice is that there may be multiple providers with different clinical perspectives. Care should be coordinated through the primary care clinician, with the assistance of a care manager if needed. It is particularly important to continually evaluate all medications prescribed by other practitioners and assess for possible long-term side effects, dependency, or drug-drug interactions. Particular attention should be given to the level of chronic pain and sleep disturbance, self-medication with alcohol or substances, chronic use of nonsteroidal anti-inflammatory agents (which can contribute to rebound headaches or pain), chronic use of sedative-hypnotic agents, chronic use of narcotic pain medications, and the impact of war-related health concerns on social and occupational functioning.

Screening for PTSD, depression, and alcohol misuse should be performed routinely in all combat veterans. Three screening tools, which are in the public domain, have been validated for use in primary care, and have been used frequently in veterans: the four-question Primary Care PTSD Screen (PC-PTSD), the two-question Patient Health Questionnaire (PHQ-2), and the three-question Alcohol Use Disorders Identification Test-Consumption module (AUDIT-C) **(Table e48–3)**.

Table e48–3 Primary Care Mental Health Screening Tools

PC-PTSD Screen

1. Have you ever had any experience that was so frightening, horrible, or upsetting that, in the past month, you:		
Have had nightmares about it or thought about it when you did not want to?	Yes	No
Tried hard not to think about it or went out of your way to avoid situations that remind you of it?	Yes	No
Were constantly on guard, watchful, or easily startled?	Yes	No
Felt numb or detached from others, activities, or your surroundings?	Yes	No

PHQ-2 Depression Screen

2. Over the last 2 weeks, how often have you been bothered by any of the following problems?	Not at all (0)	Few or several days (1)	More than half the days (2)	Nearly every day (3)
Little interest or pleasure in doing things.	0	1	2	3
Feeling down, depressed, or hopeless.	0	1	2	3

AUDIT-C Alcohol Screen

3a. How often do you have a drink containing alcohol?				
Never (0)	Monthly or less (1)	Two or four times a month (2)	Two to three times per week (3)	Four or more times a week (4)
3b. How many drinks containing alcohol do you have on a typical day when you are drinking?				
1 or 2 (0)	3 or 4 (1)	5 or 6 (2)	7 or 9 (3)	10 or more (4)
3c. How often do you have six or more drinks on one occasion?				
Never (0)	Less than Monthly (1)	Monthly (2)	Two to three times per week (3)	Four or more times a week (4)

Note: Two or more "yes" responses (three or more a more specific cutoff) is considered a positive screen.

Source: A Prins et al: The Primary Care PTSD Screen (PC-PTSD): Development and operating characteristics. Prim Care Psychiatr 9:9, 2004.

Note: If either (or both) questions are marked 2 or 3 ("more than half the days or higher), this is considered a positive screen for depression.

Source: K Kroenke et al: The Patient Health Questionnaire-2: Validity of a two-item depression screener. Med Care 41:1284, 2003.

Note: A positive AUDIT-C screen is defined as a total score for men ≥4; for women ≥3. A report of drinking 6 or more drinks on one occasion should prompt an in-depth assessment of drinking.

Source: K Bush et al: The AUDIT Alcohol Consumption Questions (AUDIT-C): An effective brief screening test for problem drinking. Arch Intern Med 158:1789, 1998.

Since the clinical definition of an acute concussion/mTBI does not include symptoms, time-course, or impairment, there is currently no clinically validated screening process for use months or years after injury. However, it is important to gather information about all injuries sustained during deployment, including any that resulted in loss or alteration of consciousness or loss of memory around the time of the event. If concussion injuries have occurred, the clinician should assess the number of such injuries, the duration of time unconscious, and injury mechanisms. This should be followed by an assessment of any postconcussive symptoms immediately following the injury event (e.g., headaches, dizziness, tinnitus, nausea, irritability, insomnia, and concentration or memory problems), and the severity and duration of such symptoms.

Treatment: Neuropsychiatric Illnesses in War Veterans

Given the interrelationship of postwar health concerns, care needs to be carefully coordinated. Specific techniques that have been found to be helpful include scheduling regular primary care visits instead of as needed visits, establishing care management, utilizing good risk-communication principles, establishing a consultative step care approach that draws on the expertise of specialists in a collaborative manner (instead of immediately referring the patient to a specialist and relying on the specialist to provide care), and having behavioral health support directly within primary care clinics (both for referrals and to provide education and support to primary care professionals prescribing treatment for depression or PTSD).

It is important not to implicitly or explicitly convey the message that physical or cognitive symptoms are psychological or due to stress. Even if depression or anxiety plays a large role in the etiology of physical health symptoms, the treatment approach should be designed within a patient-centered primary care structure, and referrals managed from within this framework. For example, it might help to explain that the primary goal of referral to a mental health professional is to improve sleep and reduce physiologic hyperarousal, which in turn will help with treatment of war-related chronic headaches, concentration problems, or chronic fatigue. If however, the primary care professional conveys the message that the cause of headaches or concentration problems is anxiety or depression, and this conflicts with the patient's own viewpoint, then this could damage therapeutic rapport and in turn exacerbate the symptoms.

Specific questions related to military service (Table e48–2) combined with screening for depression, PTSD, and alcohol use disorders (Table e48–3) should be a routine part of care for all veterans. A positive screen for depression or PTSD should prompt follow-up questions related to these disorders [or use of a longer screening tool such as the (PHQ-9) or National Center for PTSD Checklist], as well as risk assessment for suicide or homicide. It is important to assess the impact of depression or PTSD symptoms on occupational functioning and interpersonal relationships.

A positive screen for alcohol misuse should prompt a brief motivational intervention that includes bringing attention to the elevated level of drinking, informing the veteran about the effects of alcohol on health, recommending limiting use or abstaining, exploring and setting goals related to drinking behavior, and follow-up and referral to specialty care if needed. This type of brief primary care intervention has been found to be effective, and should be incorporated into routine practice. One way to facilitate dialogue about this topic with veterans is to point out how hyperarousal associated with combat service can lead to increased craving for alcohol as the body searches for ways to modulate this. Veterans may consciously or unconsciously drink more to help with sleep, reduce arousal, or avoid thinking about events that happened "downrange." A key educational strategy is to help the veteran to learn that drinking to get to sleep actually damages sleep architecture and makes sleep worse [e.g., reduces rapid eye movement (REM) sleep initially followed by rebound REM activity and early morning wakening].

SPECIFIC TREATMENT STRATEGIES FOR PTSD AND COMORBID DEPRESSION

PTSD and depression are highly comorbid in combat veterans and the evidence-based treatments are similar, involving either antidepressant medications, cognitive behavioral therapy (CBT), or both. Psychoeducation that assists veterans to understand that their symptoms of PTSD have a basis in adaptive survival mechanisms and skills they exhibited in combat can facilitate therapeutic rapport. Remaining hypervigilant to threat, being able to shut down emotions, being able to function on less sleep, and using anger to help focus and control fear are all adaptive beneficial survival skills in a combat environment. Therefore, PTSD for warriors is both a medical disorder and a set of reactions that have their roots in the physiologic adaptation and skills they successfully applied in combat.

It is important to know that combat is not the only important trauma in a war-zone environment. Rape, assault, and accidents also occur. Rape or assault by a fellow service member, which affects a greater number of women veterans, can be particularly devastating because it destroys the vital feeling of safety that individuals derive from their own unit peers in a war environment.

The treatments for PTSD considered by most consensus guideline committees to have an A level of evidence include CBTs and medications, specifically selective serotonin reuptake inhibitors (SSRIs) (of which paroxetine and sertraline received FDA approval for PTSD). Although lacking a specific FDA

indication, there is increasing evidence that serotonin norepinephrine reuptake inhibitors (SNRIs) (e.g., venlafaxine and duloxetine) and mirtazapine are also effective. (See Table 391-3 for recommended dosages.)

CBT interventions include narrative therapy (often called "imaginal exposure"), in vivo exposure focused on retraining the body not to react to stimuli related to traumatic reminders (e.g., a crowded mall), and techniques to modulate physiologic hyperarousal (e.g., diaphragmatic breathing, progressive muscle relaxation). A number of complementary alternative medicine approaches including acupuncture, mindfulness meditation, yoga, and massage are also being tested in PTSD. Although not evidence-based treatments per se, if they facilitate a relaxation response and alleviation of hyperarousal symptoms, they can be considered useful adjunctive modalities.

There have been no head-to-head comparisons of medication compared with psychotherapy for treatment of PTSD. It is reasonable for primary care clinicians to consider initiating treatment for mild to moderate PTSD symptoms with an SSRI, and refer patients to a mental health professional if there are more severe symptoms, significant comorbidity, safety concerns, or limited response to initial treatment. All PTSD treatments are associated with a sizable proportion of individuals who fail to respond adequately, and it is often necessary to add modalities or switch treatment. SNRIs may be useful alternatives to SSRIs if there has been nonresponse, side effects, or if there is comorbid pain (duloxetine, in particular, has indications for pain). Both SSRIs and SNRIs can increase anxiety initially; patients should be warned about this possibility and treatment should be initiated with the lowest recommended dose (or even one-half of the lowest dose for a few days) and gradually increased thereafter. Mirtazapine use can cause drowsiness and weight gain. Antidepressants also are likely to be useful in comorbid depression, which is common in veterans with PTSD. All antidepressants have potential drug-drug interactions that must be considered.

Many other medications have been used in PTSD, including tricyclic antidepressants, benzodiazepines, atypical antipsychotics, and anticonvulsants. In general, these should be prescribed in conjunction with psychiatric consultation, because of their greater side effects and risks. Benzodiazepines, in particular, should be avoided in combat veterans. Studies have shown that they do not reduce core PTSD symptoms, are likely to exacerbate substance use disorders that are common in veterans with PTSD, and may produce significant rebound anxiety. Veterans with PTSD often report symptomatic relief upon initiation of a benzodiazepine, but this is generally short lived and associated with a high risk of tolerance and dependence that can worsen recovery. Atypical antipsychotics, which have gained widespread popularity as adjunctive treatment for depression, anxiety, or sleep problems, have significant long-term side effects, including metabolic effects (e.g., glucose dysregulation), weight gain, and cardiovascular risks.

Sleep disturbance should be addressed initially with sleep hygiene education, followed by consideration of an antihistamine, trazodone, or non-benzodiazepine sedative-hypnotic such as zolpidem, eszopiclone, or zaleplon. However, the non-benzodiazepine sedative-hypnotics should be used with caution in veterans as they can lead to tolerance and rebound sleep problems similar to those seen with benzodiazepine use.

TREATMENT STRATEGIES FOR CONCUSSION/MTBI AND POSTDEPLOYMENT POSTCONCUSSIVE SYMPTOMS

Concussion/mTBI is best treated at the time of injury with education and rest to allow time for the brain to heal and protect against a second impact syndrome (a rare but life-threatening event involving brain swelling that can occur when a second concussion occurs before the brain has adequately healed from an initial event). Randomized trials have shown that education regarding concussion that informs the patient of what to expect and promotes the expectation of recovery is the most effective treatment in preventing persistent symptoms.

Once service members return from deployment and seek care for postwar health problems, treatment is largely symptom focused, following the principles of patient-centered and collaborative care models. Cognitive rehabilitation, which is very useful in moderate and severe TBI to improve memory, attention, and concentration, has generally not been shown to be effective for mTBI in randomized clinical studies, though consensus groups have supported its use.

General recommendations for the clinical management of persistent, chronic postconcussive symptoms include treating physical and cognitive health problems based on symptom presentation, coexisting health problems, and individual preferences; and addressing coexisting depression, PTSD, substance use disorders, or other factors that may be contributing to symptom persistence. Headache is the most common symptom associated with concussion/mTBI, and the evaluation and treatment of headache parallels those for other causes of headache (Chap. 14). Stimulant medications for alleviating neurocognitive effects attributed to concussion/mTBI are not recommended. Clinicians should be aware of the potential for cognitive or sedative side effects of certain medications that may be prescribed for depression, anxiety, sleep, or chronic pain.

Treatment of neuropsychiatric problems must be coordinated with care for other war-related health concerns, with the goal of treatment to reduce the severity of symptoms, improve social and occupational functioning, and prevent long-term disability. Understanding the occupational context of war-related health concerns is important in communicating with veterans and developing a comprehensive treatment strategy.

DISCLOSURE

This material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author and are not to be construed as official, or as reflecting true views of the Department of the Army or the Department of Defense.

FURTHER READINGS

AIARZAGUENA JM et al: A randomized controlled trial of a psychosocial and communication intervention carried out by GPs for patients with medically unexplained symptoms. Psychol Med 37:283, 2007[PMID: 17164029] [Full Text]

BELENGER HG et al: Cognitive sequelae of blast-related versus other mechanisms of brain trauma. J Int Neuropsychol Soc 15:1, 2009

DEKOSKY ST et al: Traumatic brain injury—football, warfare, and long-term effects. N Engl J Med 363:1293, 2010[PMID: 20879875] [Full Text]

DEPARTMENT OF DEFENSE: Department of Defense Task Force for the Prevention of Suicide in Members of the Armed Forces. The challenge and the promise: Strengthening the force, preventing suicides, and saving lives. 2010. Available at *http://www.health.mil/dhb/downloads/Suicide%20Prevention%20Task%20Force%20final%20report%208-23-10.pdf*

DEPARTMENT OF VETERANS AFFAIRS, DEPARTMENT OF DEFENSE: Evidence-based clinical practice guideline: Management of concussion/mild traumatic brain injury (mTBI), version 1.0, April 2009. Available at *http://www.healthquality.va.gov/mtbi/concussion_mtbi_full_1_0.pdf*

———: Clinical practice guideline for management of substance use disorders, version 2.0, August 2009. Available at *http://www.healthquality.va.gov/HEALTHQUALITY/sud/sud_full_601f.pdf*

———: Clinical practice working group. Management of posttraumatic stress. Washington, DC, Department of Veterans Affairs, Department of Defense, 2010 publication 10O-CPG/FTSD-04. 2004. Available at *http://www.healthquality.va.gov/Post_Traumatic_Stress_Disorder_PTSD.asp*

ENGEL CC et al: Managing future Gulf War syndromes: International lessons and new models of care. Philos Trans R Soc Lond B Biol Sci 361(1468):707, 2006[PMID: 16687273] [Full Text]

ESCOBAR JI et al: Effectiveness of a time-limited cognitive behavioral therapy type intervention among primary care patients with medically unexplained symptoms. Ann Fam Med 5:328, 2007[PMID: 17664499] [Full Text]

FOA EB et al (eds): *Effective Treatments for PTSD: Practice Guidelines from the International Society for Traumatic Stress Studies,* 2d ed. New York, Guilford Press, 2009

HOGE CW: *Once a Warrior—Always a Warrior: Navigating the Transition from Combat to Home, Including Combat Stress, PTSD, and mTBI*. Guilford, CT, Globe Pequot Press, 2010

——— et al: Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N Engl J Med 351:13, 2004

——— et al: Association of posttraumatic stress disorder with somatic symptoms, health care visits, and absenteeism among Iraq War veterans. Am J Psychiatry 164:150, 2007

——— et al: Mild traumatic brain injury in U.S. soldiers returning from Iraq. N Engl J Med 358:453, 2008

——— et al. Care of veterans with mild traumatic brain injury: Flawed perspectives. N Engl J Med 360:1588, 2009

HOWE LL: Giving context to post-deployment post-concussive-like symptoms: Blast-related potential mild traumatic brain injury and comorbidities. Clin Neuropsychol 23:1315, 2009[PMID: 19882474] [Full Text]

INSTITUTE OF MEDICINE: *Committee on Gulf War and Health. Health Effects of Serving in the Gulf War, Update 2009.* Washington DC, National Academies Press, 2010

MARX BP et al: Association of time since deployment, combat intensity, and posttraumatic stress symptoms with neuropsychological outcomes following Iraq war deployment. Arch Gen Psychiatry 66:996, 2009[PMID: 19736356] [Full Text]

PIETRZAK RH et al: Posttraumatic stress disorder mediates the relationship between mild traumatic brain

injury and health and psychosocial functioning in veterans of Operations Enduring Freedom and Iraqi Freedom. J Nerv Ment Dis 197:748, 2009[PMID: 19829203] [Full Text]

SCHNEIDERMAN AI et al: Understanding the sequelae of injury mechanisms and mild traumatic brain injury incurred during the conflicts in Iraq and Afghanistan: Persistent postconcussive symptoms and posttraumatic stress disorder. Am J Epidemiol 167:1446, 2008[PMID: 18424429] [Full Text]

SMITH RC et al: Primary care clinicians treat patients with medically unexplained symptoms: A randomized controlled trial. J Gen Intern Med 21:671, 2006[PMID: 16808764] [Full Text]

THOMAS JL et al: Prevalence of mental health problems and functional impairment among active component and national guard soldiers 3 and 12 months following combat in Iraq. Arch Gen Psychiatry 67:614, 2010[PMID: 20530011] [Full Text]

VAN DER FELTZ-CORNELIS CM et al: Randomized controlled trial of a collaborative care model with psychiatric consultation for persistent medically unexplained symptoms in general practice. Psychother Psychosom 75:282, 2006

WILK JE et al: Mild traumatic brain injury (concussion) during combat: Lack of association of blast mechanism with persistent post-concussive symptoms. J Head Trauma Rehabil 25:9, 2010[PMID: 20051900] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

HEAVY METAL POISONING: INTRODUCTION

Metals pose a significant threat to health through low-level environmental as well as occupational exposures. One indication of their importance relative to other potential hazards is their ranking by the U.S. Agency for Toxic Substances and Disease Registry, which maintains an updated list of all hazards present in toxic waste sites according to their prevalence and the severity of their toxicity. The first, second, third, and seventh hazards on the list are heavy metals: lead, mercury, arsenic, and cadmium, respectively *(http://www.atsdr.cdc.gov/cercla/07list.html)*. Specific information pertaining to each of these metals, including sources and metabolism, toxic effects produced, diagnosis, and the appropriate treatment for poisoning, is summarized in **Table e49–1.**

Table e49–1 Heavy Metals

Main Sources	Metabolism	Toxicity	Diagnosis	Treatment
Arsenic				
Smelting and microelectronics industries; wood preservatives, pesticides, herbicides, fungicides; contaminant of deep-water wells; folk remedies; and coal; incineration of these products	Organic arsenic (arsenobetaine, arsenocholine) is ingested in seafood and fish, but is nontoxic; inorganic arsenic is readily absorbed (lung and GI); sequesters in liver, spleen, kidneys, lungs, and GI tract; residues persist in skin, hair, and nails; biomethylation results in detoxification, but this process saturates.	Acute arsenic poisoning results in necrosis of intestinal mucosa with hemorrhagic gastroenteritis, fluid loss, hypotension, delayed cardiomyopathy, acute tubular necrosis, and hemolysis. Chronic arsenic exposure causes diabetes, vasospasm, peripheral vascular insufficiency and gangrene, peripheral neuropathy, and cancer of skin, lung, liver (angiosarcoma), bladder, kidney. Lethal dose: 120–200 mg (adults); 2 mg/kg (children).	Nausea, vomiting, diarrhea, abdominal pain, delirium, coma, seizures; garlicky odor on breath; hyperkeratosis, hyperpigmentation, exfoliative dermatitis, and Mees' lines (transverse white striae of the fingernails); sensory and motor polyneuritis, distal weakness. Radiopaque sign on abdominal x-ray; ECG–QRS broadening, QT prolongation, ST depression, T-wave flattening; 24-h urinary arsenic >67 μmol/d or 50 μg/d; (no seafood x 24 h); if recent exposure, serum arsenic >0.9 μmol/L (7 μg/dL). High arsenic in hair or nails.	If acute ingestion, ipecac to induce vomiting, gastric lavage, activated charcoal with a cathartic. Supportive care in ICU. Dimercaprol 3–5 mg/kg IM q4h x 2 days; q6h x 1 day, then q12h x 10 days; alternative: oral succimer.
Cadmium				
Metal-plating, pigment, smelting, battery, and plastics industries; tobacco; incineration of these products; ingestion of food that concentrates cadmium	Absorbed through ingestion or inhalation; bound by metallothionein, filtered at the glomerulus, but reabsorbed by proximal tubules (thus, poorly excreted). Biologic half-life:	Acute cadmium inhalation causes pneumonitis after 4–24 h; acute ingestion causes gastroenteritis. Chronic exposure causes anosmia, yellowing of teeth, emphysema, minor LFT elevations,	With inhalation: pleuritic chest pain, dyspnea, cyanosis, fever, tachycardia, nausea, noncardiogenic pulmonary edema. With ingestion: nausea, vomiting, cramps, diarrhea. Bone pain, fractures with osteomalacia. If recent exposure, serum cadmium >500 nmol/L (5 μg/dL). Urinary cadmium >100	There is no effective treatment for cadmium poisoning (chelation not useful; dimercaprol can exacerbate nephrotoxicity). Avoidance of further exposure, supportive therapy, vitamin D for

(grains, cereals).	10–30 y. Binds cellular sulfhydryl groups, competes with zinc, calcium for binding sites. Concentrates in liver and kidneys.	microcytic hypochromic anemia unresponsive to iron therapy, proteinuria, increased urinary β^2- microglobulin, calciuria, leading to chronic renal failure, osteomalacia, and fractures.	nmol/L (10 μg/g creatinine) and/or urinary β^2-microglobulin >750 μg/g creatinine (but urinary β^2-microglobulin also increased in other renal diseases such as pyelonephritis).	osteomalacia.
Lead				
Manufacturing of auto batteries, lead crystal, ceramics, fishing weights, etc.; demolition or sanding of lead-painted houses, bridges; stained glass–making, plumbing, soldering; environmental exposure to paint chips, house dust (in homes built <1975), firing ranges (from bullet dust), food or water from improperly glazed ceramics, lead pipes; contaminated herbal remedies, candies; exposure to the combustion of leaded fuels.	Absorbed through ingestion or inhalation; organic lead (e.g., tetraethyl lead) absorbed dermally. In blood, 95–99% sequestered in RBCs—thus, must measure lead in whole blood (not serum). Distributed widely in soft tissue, with half-life ~30 days; 15% of dose sequestered in bone with half-life of >20 years. Excreted mostly in urine, but also appears in other fluids including breast milk. Interferes with mitochondrial oxidative phosphorylation, ATPases, calcium-dependent messengers; enhances oxidation and cell apoptosis.	Acute exposure with blood lead levels (BPb) of > 60–80 μg/dL can cause impaired neurotransmission and neuronal cell death (with central and peripheral nervous system effects); impaired hematopoiesis and renal tubular dysfunction. At higher levels of exposure (e.g., BPb >80–120 μg/dL), acute encephalopathy with convulsions, coma, and death may occur. Subclinical exposures in children (BPb 25–60 μg/dL) are associated with anemia; mental retardation; and deficits in language, motor function, balance, hearing, behavior, and school performance. Impairment of IQ appears to occur at even lower levels of exposure with no measurable threshold above the limit of detection in most assays of 1 μg/dL. In adults, chronic subclinical exposures (BPb >40 μg/dL) are associated with an increased risk of anemia, demyelinating peripheral neuropathy (mainly motor), impairments of reaction time.	Abdominal pain, irritability, lethargy, anorexia, anemia, Fanconi's syndrome, pyuria, azotemia in children with blood lead level (BPb) >80 μg/dL; may also see epiphyseal plate "lead lines" on long bone x-rays. Convulsions, coma at BPb >120 μg/dL. Noticeable neurodevelopmental delays at BPb of 40–80 μg/dL; may also see symptoms associated with higher BPb levels. Screening of all U.S. children when they begin to crawl (~6 months) is recommended by the CDC; source identification and intervention is begun if the BPb >10 μg/dL. In adults, acute exposure causes similar symptoms as in children as well as headaches, arthralgias, myalgias, depression, impaired short-term memory, loss of libido. Physical exam may reveal a "lead line" at the gingiva-tooth border, pallor, wrist drop, and cognitive dysfunction (e.g., declines on the mini-mental status exam); lab tests may reveal a normocytic, normochromic anemia, basophilic stippling, an elevated blood protoporphyrin level (free erythrocyte or zinc), and motor delays on nerve conduction. U.S. OSHA requires regular testing of lead-exposed workers with removal if BPb >40 μg/dL. New guidelines have been proposed recommending monitoring of cumulative exposure parameters (Kosnett, 2007).	Identification and correction of exposure sources is critical. In some U.S. states, screening and reporting to local health boards of children with BPb >10 μg/dL and workers with BPb >40 μg/dL is required. In the highly exposed individual with symptoms, chelation is recommended with oral DMSA (succimer); if acutely toxic, hospitalization and IV or IM chelation with ethylenediamine tetraacetic acid calcium disodium (CaEDTA) may be required, with the addition of dimercaprol to prevent worsening of encephalopathy. It is uncertain whether children with asymptomatic lead exposure (e.g., BPb 20–40 μg/dL) benefit from chelation; a recent randomized trial showed no benefit. Correction of dietary deficiencies in iron, calcium, magnesium, and zinc will lower lead absorption and may also improve toxicity. Vitamin C is a weak but natural chelating agent. Calcium supplements (1200 mg at bedtime) have been shown to lower blood lead levels in pregnant women.

		hypertension, ECG conduction delays, interstitial nephritis and chronic renal failure, diminished sperm counts, spontaneous abortions.		
Mercury				
Metallic, mercurous, and mercuric mercury (Hg, Hg^{+}, Hg^{2+}) exposures occur in some chemical, metal-processing, electrical-equipment, automotive industries; they are also in thermometers, dental amalgams, batteries. Mercury is dispersed by waste incineration. Environmental bacteria convert inorganic to organic mercury, which then bioconcentrates up the aquatic food chain to contaminate tuna, swordfish, and other pelagic fish.	Elemental mercury (Hg) is not well absorbed; however, it will volatilize into highly absorbable vapor. Inorganic mercury is absorbed through the gut and skin. Organic mercury is well absorbed through inhalation and ingestion. Elemental and organic mercury cross the blood-brain barrier and placenta. Mercury is excreted in urine and feces and has a half-life in blood of ~60 days; however, deposits will remain in the kidney and brain for years. Exposure to mercury stimulates the kidney to produce metallothionein, which provides some detoxification benefit. Mercury binds sulfhydryl groups and interferes with a wide variety of critical enzymatic processes.	Acute inhalation of Hg x vapor causes pneumonitis and noncardiogenic pulmonary edema leading to death, CNS symptoms, and polyneuropathy. Chronic high exposure causes CNS toxicity (mercurial erethism; see Diagnosis); lower exposures impair renal function, motor speed, memory, coordination. Acute ingestion of inorganic mercury causes gastroenteritis, the nephritic syndrome, or acute renal failure, hypertension, tachycardia, and cardiovascular collapse, with death at a dose of 10–42 mg/kg. Ingestion of organic mercury causes gastroenteritis, arrhythmias, and lesions in the basal ganglia, gray matter, and cerebellum at doses >1.7 mg/kg. High exposure during pregnancy causes derangement of fetal neuronal migration resulting in severe mental retardation. Mild exposures during pregnancy (from fish consumption) are associated with declines in neurobehavioral	Chronic exposure to metallic mercury vapor produces a characteristic intention tremor and mercurial erethism: excitability, memory loss, insomnia, timidity, and delirium ("mad as a hatter"). On neurobehavioral tests: decreased motor speed, visual scanning, verbal and visual memory, visuomotor coordination. Children exposed to mercury in any form may develop acrodynia ("pink disease"): flushing, itching, swelling, tachycardia, hypertension, excessive salivation or perspiration, irritability, weakness, morbilliform rashes, desquamation of palms and soles. Toxicity from elemental or inorganic mercury exposure begins when blood levels >180 nmol/L (3.6 μg/dL) and urine levels >0.7 μmol/L (15 μg/dL). Exposures that ended years ago may result in a >20-μg increase in 24-h urine after a 2-g dose of succimer. Organic mercury exposure is best measured by levels in blood (if recent) or hair (if chronic); CNS toxicity in children may derive from fetal exposures associated with maternal hair Hg >30 nmol/g (6 μg/g).	Treat acute ingestion of mercuric salts with induced emesis or gastric lavage and polythiol resins (to bind mercury in the GI tract). Chelate with dimercaprol (up to 24 mg/kg per day IM in divided doses), DMSA(succimer), or penicillamine, with 5-day courses separated by several days of rest. If renal failure occurs, treat with peritoneal dialysis, hemodialysis, or extracorporeal regional complexing hemodialysis and succimer. Chronic inorganic mercury poisoning is best treated with N-acetyl penicillamine.

		performance in offspring. Dimethylmercury, a compound only found in research labs, is "supertoxic"—a few drops of exposure via skin absorption or inhaled vapor can cause severe cerebellar degeneration and death.		

Abbreviations: ATPase, adenosine triphosphatase; CDC, Centers for Disease Control and Prevention; CNS, central nervous system; DMSA, dimercaptosuccinic acid; ECG, electrocardiogram; GI, gastrointestinal; ICU, intensive care unit; IQ, intelligence quotient; LFT, liver function tests; OSHA, Occupational Safety and Health Administration; RBC, red blood cell.

Metals are inhaled primarily as dusts and fumes (the latter defined as tiny particles generated by combustion). Metal poisoning can also result from exposure to vapors (e.g., mercury vapor in creating dental amalgams). When metals are ingested in contaminated food or drink or by hand-to-mouth activity (implicated especially in children), their gastrointestinal absorption varies greatly with the specific chemical form of the metal and the nutritional status of the host. Once a metal is absorbed, blood is the main medium for its transport, with the precise kinetics dependent on diffusibility, protein binding, rates of biotransformation, availability of intracellular ligands, and other factors. Some organs (e.g., bone, liver, and kidney) sequester metals in relatively high concentrations for years. Most metals are excreted through renal clearance and gastrointestinal excretion; some proportion is also excreted through salivation, perspiration, exhalation, lactation, skin exfoliation, and loss of hair and nails. The intrinsic stability of metals facilitates tracing and measurement in biologic material, although the clinical significance of the levels measured is not always clear.

Some metals, such as copper and selenium, are essential to normal metabolic function as trace elements (Chap. 74) but are toxic at high levels of exposure. Others, such as lead and mercury, are xenobiotic and theoretically are capable of exerting toxic effects at any level of exposure. Indeed, much research is currently focused on the contribution of low-level xenobiotic metal exposure to chronic diseases and to subtle changes in health that may have significant public health consequences. Genetic factors, such as polymorphisms that encode for variant enzymes with altered properties in terms of metal binding, transport, and effects, also may modify the impact of metals on health and thereby account, at least in part, for individual susceptibility to metal effects.

The most important component of treatment for metal toxicity is the termination of exposure. *Chelating agents* are used to bind metals into stable cyclic compounds with relatively low toxicity and to enhance their excretion. The principal chelating agents are dimercaprol [British anti-Lewisite (BAL)], ethylenediamine tetraacetic acid (EDTA), succimer [dimercaptosuccinic acid (DMSA)], and penicillamine; their specific use depends on the metal involved and the clinical circumstances. Activated charcoal does not bind metals and thus is of limited usefulness in cases of acute metal ingestion.

In addition to the information provided in Table e49–1, several other aspects of exposure, toxicity, or management are worthy of discussion with respect to the four most hazardous toxicants (arsenic, cadmium, lead, and mercury).

Arsenic, even at moderate levels of exposure, has been clearly linked with increased risks for cancer at a number of different tissue sites. These risks appear to be modified by smoking, folate and selenium status, and other factors. Evidence is also emerging that low-level arsenic may cause neurodevelopmental delays in children and possibly diabetes, but the evidence (particularly for diabetes) remains uneven.

Serious *cadmium* poisoning from the contamination of food and water by mining effluents in Japan contributed to the 1946 outbreak of "itai-itai" ("ouch-ouch") disease, so named because of cadmium-induced bone toxicity that led to painful bone fractures. Modest exposures from environmental contamination have recently been associated in some studies with a lower bone density, a higher incidence of fractures, and a faster decline in height in both men and women, effects that may be related to cadmium's calciuric effect on the kidney. There is some evidence for synergy between the adverse impacts of cadmium and lead on kidney function. Environmental exposures have also been linked to lower lung function (even after adjusting for smoking cigarettes, which contain cadmium) as well as increased risk of stroke and heart failure. Such research is creating concern that cadmium exposure may be contributing significantly to morbidity and mortality rates in the general population.

Advances in our understanding of *lead* toxicity have recently benefited by the development of K x-ray fluorescence (KXRF) instruments for making safe in vivo measurements of lead levels in bone (which, in turn, reflect cumulative exposure over many years, as opposed to blood lead levels, which mostly reflect recent exposure). Higher bone lead levels measured by KXRF have been linked to increased risk of hypertension and accelerated declines in cognition in both men and women from an urban population. Prospective studies have also demonstrated that higher bone lead levels are a major risk factor for increased cardiovascular morbidity and mortality rates. With respect to pregnancy-associated risks, high maternal bone lead levels were found to predict lower birth weight, head circumference, birth length, and neurodevelopmental performance in offspring by age 2 years. In a randomized trial, calcium supplementation (1200 mg daily) was found to significantly reduce the mobilization of lead from maternal bone into blood during pregnancy.

The toxicity of low-level organic *mercury* exposure (as manifested by neurobehavioral performance) is of increasing concern based on studies of the offspring of mothers who ingested mercury-contaminated fish. With respect to whether the consumption of fish by women during pregnancy is good or bad for offspring neurodevelopment, balancing the trade-offs of the beneficial effects of the omega-3-fatty acids (FAs) in fish versus the adverse effects of mercury contamination in fish has led to some confusion and inconsistency in public health recommendations. Overall, it would appear that it would be best for pregnant women to either limit fish consumption to those species known to be low in mercury contamination but high in omega-3-FAs (such as sardines or mackerel) or to avoid fish and obtain omega-3-FAs through supplements or other dietary sources. Current evidence has not supported the recent contention that ethyl mercury, used as a preservative in multiuse vaccines administered in early childhood, has played a significant role in causing neurodevelopmental problems such as autism. With regards to adults, there is conflicting evidence as to whether mercury exposure is associated with increased risk of hypertension and cardiovascular disease. At this point, conclusions cannot be drawn.

Heavy metals pose risks to health that are especially burdensome in selected parts of the world. For example, *arsenic* exposure from natural contamination of shallow tube wells inserted for drinking water is a major environmental problem for millions of residents in parts of Bangladesh and Western India. Contamination was formerly considered only a problem with deep wells; however, the geology of this region allows most residents only a few alternatives for potable drinking water. The combustion of leaded gasoline with resulting contamination of air and soil with *lead oxide* remains a problem in some countries of Central Asia, Southeast Asia, Africa, and the Middle East. Populations living in the Arctic have been shown to have particularly high exposures to *mercury* due to long-range transport patterns that concentrate mercury in the polar regions, as well as the traditional dependence of Arctic peoples on the consumption of fish and other wildlife that bioconcentrate methylmercury.

A few additional metals deserve brief mention but are not covered in Table e49–1 because of the relative rarity of their being clinically encountered, or the uncertainty regarding their potential toxicities. *Aluminum* contributes to the encephalopathy in patients with severe renal disease, who are undergoing dialysis (Chap. 353). High levels of aluminum are found in the neurofibrillary tangles in the cerebral cortex and hippocampus of patients with Alzheimer's disease, as well as in the drinking water and soil of areas with an unusually high incidence of Alzheimer's. The experimental and epidemiologic evidence for the aluminum–Alzheimer's disease link remains relatively weak, however, and it cannot be concluded that aluminum is a causal agent or a contributing factor in neurodegenerative disease. Hexavalent chromium is corrosive and sensitizing. Workers in the chromate and chrome pigment production industries have consistently had a greater risk of lung cancer. The introduction of *cobalt* chloride as a fortifier in beer led to outbreaks of fatal cardiomyopathy among heavy consumers. Occupational exposure (e.g., of miners, dry-battery manufacturers, and arc welders) to *manganese* can cause a Parkinsonian syndrome within 1–2 years, including gait disorders; postural instability; a masked, expressionless face; tremor; and psychiatric symptoms. With the introduction of methylcyclopentadienyl manganese tricarbonyl (MMT) as a gasoline additive, there is concern for the toxic potential of environmental manganese exposure. For example, a recent study found a high prevalence of Parkinsonian disorders in a community with risks proportionate to estimated manganese exposures emitted by local ferroalloy industries. Epidemiologic studies have also suggested that manganese may interfere with early childhood neurodevelopment in ways similar to that of lead. *Nickel* exposure induces an allergic response, and inhalation of nickel compounds with low aqueous solubility (e.g., nickel subsulfide and nickel oxide) in occupational settings is associated with an increased risk of lung cancer. Overexposure to selenium may cause local irritation of the respiratory system and eyes, gastrointestinal irritation, liver inflammation, loss of hair, depigmentation, and peripheral nerve damage. Workers exposed to certain organic forms of *tin* (particularly trimethyl and triethyl derivatives) have developed psychomotor disturbances, including tremor, convulsions, hallucinations, and psychotic behavior.

Thallium, which is a component of some insecticides, metal alloys, and fireworks, is absorbed through the skin as well as by ingestion and inhalation. Severe poisoning follows a single ingested dose of >1 g or >8 mg/kg. Nausea and vomiting, abdominal pain, and hematemesis precede confusion, psychosis, organic brain syndrome, and coma. Thallium is radiopaque. Induced emesis or gastric lavage is

indicated within 4–6 h of acute ingestion; Prussian blue prevents absorption and is given orally at 250 mg/kg in divided doses. Unlike other types of metal poisoning, thallium poisoning may be less severe when activated charcoal is used to interrupt its enterohepatic circulation. Other measures include forced diuresis, treatment with potassium chloride (which promotes renal excretion of thallium), and peritoneal dialysis.

FURTHER READINGS

Chen Y et al: Arsenic exposure at low-to-moderate levels and skin lesions, arsenic metabolism, neurological functions, and biomarkers for respiratory and cardiovascular diseases: Review of recent findings from the Health Effects of Arsenic Longitudinal Study (HEALS) in Bangladesh. Toxicol Appl Pharmacol 239:184, 2009[PMID: 19371619] [Full Text]

Ettinger AS et al: Effect of calcium supplementation on blood lead levels in pregnancy: A randomized control trial. Environ Health Perspec 117:26, 2009[PMID: 19165383] [Full Text]

Järup L, Akesson A: Current status of cadmium as an environmental health problem. Toxicol Appl Pharmacol 238:201, 2009[PMID: 21728242] [Full Text]

Kosnett MJ: Recommendations for medical management of adult lead exposure. Environ Health Perspect 115:463, 2007[PMID: 17431500] [Full Text]

Lucchini RG et al: High prevalence of Parkinsonian disorders associated to manganese exposure in the vicinities of ferroalloy industries. Am J Ind Med 50:788, 2007[PMID: 17918215] [Full Text]

Menke A et al: Cadmium levels in urine and mortality among U.S. adults. Environ Health Perspect 117:190, 2009[PMID: 19270787] [Full Text]

Rahman MM et al: Chronic exposure of arsenic via drinking water and its adverse health impacts on humans. Environ Geochem Health 31:189, 2009[PMID: 19190988] [Full Text]

Schoeman K et al: Defining a lowest observable adverse effect hair concentrations of mercury for neurodevelopmental effects of prenatal methylmercury exposure through maternal fish consumption: A systematic review. Ther Drug Monit 31:670, 2009[PMID: 19865003] [Full Text]

Valera B et al: Environmental mercury exposure and blood pressure among Nunavik Inuit adults. Hypertension 54:981, 2009[PMID: 19805642] [Full Text]

Weisskopf MG et al: A prospective study of bone lead concentration and death from all causes, cardiovascular diseases, and cancer in the VA Normative Aging Study. Circulation 120:1056, 2009[PMID: 19738141] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

POISONING AND DRUG OVERDOSAGE: INTRODUCTION

Poisoning refers to the development of dose-related adverse effects following exposure to chemicals, drugs, or other xenobiotics. To paraphrase Paracelsus, the dose makes the poison. In excessive amounts, substances that are usually innocuous, such as oxygen and water, can cause toxicity. Conversely, in small doses, substances commonly regarded as poisons, such as arsenic and cyanide, can be consumed without ill effect. Although most poisons have predictable dose-related effects, individual variability in the response to a given dose may occur because of genetic polymorphism, enzymatic induction or inhibition in the presence of other xenobiotics, or acquired tolerance. Poisoning may be local (e.g., skin, eyes, or lungs) or systemic depending on the route of exposure, the chemical and physical properties of the poison, and its mechanism of action. The severity and reversibility of poisoning also depend on the functional reserve of the individual or target organ, which is influenced by age and preexisting disease.

EPIDEMIOLOGY

More than 5 million poison exposures occur in the United States each year. Most are acute, accidental (unintentional), involve a single agent, occur in the home, result in minor or no toxicity, and involve children younger than 6 years of age. Pharmaceuticals are involved in 47% of exposures and 84% of serious or fatal poisonings. Unintentional exposures can result from the improper use of chemicals at work or play; label misreading; product mislabeling; mistaken identification of unlabeled chemicals; uninformed self-medication; and dosing errors by nurses, pharmacists, physicians, parents, and the elderly. Excluding the recreational use of ethanol, attempted suicide (deliberate self-harm) is the most common reported reason for intentional poisoning. Recreational use of prescribed and over-the-counter drugs for psychotropic or euphoric effects (abuse) or excessive self-dosing (misuse) are increasingly common and may also result in unintentional self-poisoning.

About 20–25% of exposures require bedside health professional evaluation, and 5% of all exposures require hospitalization. Poisonings account for 5–10% of all ambulance transports, emergency department visits, and intensive care unit admissions. Up to 30% of psychiatric admissions are prompted by attempted suicide via overdosage. Overall, the mortality rate is low: <1% of all exposures. It is much higher (1–2%) in hospitalized patients with intentional (suicidal) overdose, who account for the majority of serious poisonings. Acetaminophen is the pharmaceutical agent most often implicated in fatal poisoning. Overall, carbon monoxide is the leading cause of death from poisoning, but this is not reflected in hospital or poison center statistics because patients with such poisoning are typically dead when discovered and are referred directly to medical examiners.

DIAGNOSIS

Although poisoning can mimic other illnesses, the correct diagnosis can usually be established by the history, physical examination, routine and toxicologic laboratory evaluations, and characteristic clinical course. The *history* should include the time, route, duration, and circumstances (location, surrounding events, and intent) of exposure; the name and amount of each drug, chemical, or ingredient involved; the time of onset, nature, and severity of symptoms; the time and type of first-aid measures provided; and the medical and psychiatric history.

In many cases the patient is confused, comatose, unaware of an exposure, or unable or unwilling to admit to one. Suspicious circumstances include unexplained sudden illness in a previously healthy person or a group of healthy people; a history of psychiatric problems (particularly depression); recent changes in health, economic status, or social relationships; and onset of illness while working with chemicals or after ingesting food, drink (especially ethanol), or medications. Patients who become ill soon after arriving from a foreign country or being arrested for criminal activity should be suspected of "body packing" or "body stuffing" (ingesting or concealing illicit drugs in a body cavity). Relevant history may be available from family, friends, paramedics, police, pharmacists, physicians, and employers, who should be questioned regarding the patient's habits, hobbies, behavior changes, available medications, and antecedent events. A search of clothes, belongings, and place of discovery may reveal a suicide note or a container of drugs or chemicals. The imprint code on pills and the label on chemical products may be used to identify the ingredients and potential toxicity of a suspected poison by consulting a reference text, a computerized database, the manufacturer, or a regional poison information center (800-222-1222). Occupational exposures require review of any available material safety data sheet (MSDS) from the worksite.

The *physical examination* should focus initially on the vital signs, cardiopulmonary system, and neurologic status. The neurologic examination should include documentation of neuromuscular abnormalities such as dyskinesia, dystonia, fasciculations, myoclonus, rigidity, and tremors. The patient should also be examined for evidence of trauma and underlying illnesses. Focal neurologic findings are uncommon in poisoning, and their presence should prompt evaluation for a structural central nervous system (CNS) lesion. Examination of the

eyes (for nystagmus, pupil size and reactivity), abdomen (for bowel activity and bladder size), and skin (for burns, bullae, color, warmth, moisture, pressure sores, and puncture marks) may reveal findings of diagnostic value. When the history is unclear, all orifices should be examined for the presence of chemical burns and drug packets. The odor of breath or vomitus and the color of nails, skin, or urine may provide important diagnostic clues.

The diagnosis of poisoning in cases of unknown etiology primarily relies on pattern recognition. The first step is to assess the pulse, blood pressure, respiratory rate, temperature, and neurologic status and characterize the overall physiologic state as **stimulated**, **depressed**, **discordant**, or **normal(Table e50–1)**. Obtaining a complete set of vital signs and reassessing them frequently are critical. Measuring core temperature is especially important, even in difficult or combative patients, since temperature elevation is the most reliable prognosticator of poor outcome in poisoning or drug withdrawal. The next step is to consider the underlying causes of the physiologic state and attempt to identify a pathophysiologic pattern or toxic syndrome (*toxidrome*) based on the observed findings. Assessing the severity of physiologic derangements **(Table e50–2)** is useful in this regard and also for monitoring the clinical course and response to treatment. The final step is to attempt to identify the particular agent involved by looking for unique or relatively poison-specific physical or ancillary test abnormalities. Distinguishing among toxidromes based on the physiologic state is summarized below.

Table e50–1 Differential Diagnosis of Poisoning Based on Physiologic State

Stimulated	Depressed	Discordant	Normal
Sympathetics			
Sympathomimetics			
Ergot alkaloids			
Methylxanthines			
Monoamine oxidase inhibitors			
Thyroid hormones			
Anticholinergics			
Antihistamines			
Antiparkinsonian agents			
Antipsychotics			
Antispasmodics			
Belladonna alkaloids			
Cyclic antidepressants			
Muscle relaxants			
Mushrooms and plants			
Hallucinogens			
Cannabinoids (marijuana)			
LSD and analogues			
Mescaline and analogues			
Mushrooms			
Phencyclidine and analogues			
Withdrawal syndromes			
Barbiturates			
Benzodiazepines			
Ethanol			
Opioids			
Sedative-hypnotics			
Sympatholytics			
Sympatholytics			
α_1-Adrenergic antagonists			
α_2-Adrenergic agonists			
ACE inhibitors			
Angiotensin receptor blockers			
Antipsychotics			
β-Adrenergic blockers			
Calcium channel blockers			

- Cardiac glycosides
- Cyclic antidepressants

Cholinergics
- Acetylcholinesterase inhibitors
- Muscarinic agonists
- Nicotinic agonists

Opioids
- Analgesics
- GI antispasmodics
- Heroin

Sedative-hypnotics
- Alcohols
- Anticonvulsants
- Barbiturates
- Benzodiazepines
- GABA precursors
- Muscle relaxants
- Other agents
- GHB products

Asphyxiants
- Cytochrome oxidase inhibitors
- Inert gases
- Irritant gases
- Methemoglobin inducers
- Oxidative phosphorylation inhibitors

AGMA inducers
- Alcohol (ketoacidosis)
- Ethylene glycol
- Iron
- Methanol
- Salicylate
- Toluene

CNS syndromes
- Extrapyramidal reactions
- Hydrocarbon inhalation
- Isoniazid
- Lithium
- Neuroleptic malignant syndrome
- Serotonin syndrome
- Strychnine

Membrane-active agents
- Amantadine
- Antiarrhythmics
- Antihistamines
- Antipsychotics
- Carbamazepine
- Cyclic antidepressants
- Local anesthetics
- Opioids (some)
- Orphenadrine
- Quinoline antimalarials

Nontoxic exposure

Psychogenic illness	
Toxic time-bombs	
Slow absorption	
Anticholinergics	
Carbamazepine	
Concretion formers	
Extended-release phenytoin sodium capsules (Dilantin Kapseals)	
Drug packets	
Enteric-coated pills	
Diphenoxylate-atropine (Lomotil)	
Opioids	
Salicylates	
Sustained-release pills	
Valproate	
Slow distribution	
Cardiac glycosides	
Lithium	
Metals	
Salicylate	
Valproate	
Toxic metabolite	
Acetaminophen	
Carbon tetrachloride	
Cyanogenic glycosides	
Ethylene glycol	
Methanol	
Methemoglobin inducers	
Mushroom toxins	
Organophosphate insecticides	
Paraquat	
Metabolism disruptors	
Antineoplastic agents	
Antiviral agents	
Colchicine	
Hypoglycemic agents	
Immunosuppressive agents	
MAO inhibitors	
Metals	
Salicylate	
Warfarin	

Abbreviations: ACE, angiotensin-converting enzyme; AGMA, anion-gap metabolic acidosis; CNS, central nervous system; GABA, γ-aminobutyric acid; GHB, γ-hydroxybutyrate; GI, gastrointestinal; LSD, lysergic acid diethylamide; MAO, monoamine oxidase.

Increased pulse, blood pressure, respiratory rate, temperature, and neuromuscular activity characterize the **stimulated** physiologic state: sympathetic, antimuscarinic (anticholinergic), hallucinogen poisoning, and drug withdrawal (Table e50–1). Other features are noted in Table e50–2. Mydriasis, a characteristic feature of all stimulants, is most marked in antimuscarinic (anticholinergic) poisoning since pupillary reactivity relies on muscarinic control; in sympathetic poisoning (e.g., cocaine), pupils are also enlarged, but some reactivity to light is observed. The antimuscarinic (anticholinergic) toxidrome is also distinguished by the presence of hot, dry, flushed skin; decreased bowel sounds; and urinary retention (Table e50–1). Other stimulant syndromes increase sympathetic activity and cause diaphoresis, pallor, and increased bowel activity with varying degrees of nausea, vomiting, abnormal distress, and occasionally diarrhea. The absolute and relative degree of vital sign changes and neuromuscular hyperactivity can help distinguish among stimulant toxidromes. Since sympathetics stimulate the peripheral nervous system more directly than do hallucinogens or drug withdrawal, markedly increased vital signs and organ ischemia suggest sympathetic poisoning. Findings

helpful in suggesting the particular drug or class causing physiologic stimulation include reflex bradycardia from selective α-adrenergic stimulants (e.g., decongestants), hypotension from selective β-adrenergic stimulants (e.g., asthma therapeutics), limb ischemia from ergot alkaloids, rotatory nystagmus from phencyclidine and ketamine (the only physiologic stimulants that cause this finding), and delayed cardiac conduction from high doses of cocaine and some anticholinergic agents (e.g., antihistamines, cyclic antidepressants, and antipsychotics). Seizures suggest a sympathetic etiology, an anticholinergic agent with membrane-active properties (e.g., cyclic antidepressants, orphenadrine, phenothiazines), or a withdrawal syndrome. Close attention to core temperature is critical in patients with grade 4 physiologic stimulation (Table e50–2).

Table e50–2 Severity of Physiologic Stimulation and Depression in Poisoning and Drug withdrawal

Physiologic Stimula-tion	
Grade 1	Anxious, irritable, tremulous; vital signs normal; diaphoresis, flushing or pallor, mydriasis, and hyperreflexia may be present
Grade 2	Agitated; may have confusion or hallucinations but is able to converse and follow commands; vital signs mildly to moderately increased
Grade 3	Delirious; unintelligible speech, uncontrollable motor hyperactivity; moderately to markedly increased vital signs; tachyarrhythmias possible
Grade 4	Coma, seizures, cardiovascular collapse
Physiologic Depression	
Grade 1	Awake, lethargic, or sleeping but arousable by voice or tactile stimulation; able to converse and follow commands; may be confused
Grade 2	Responds to pain but not voice; can vocalize but not converse; spontaneous motor activity present; brainstem reflexes intact
Grade 3	Unresponsive to pain; spontaneous motor activity absent; brainstem reflexes depressed; motor tone, respirations, and temperature decreased
Grade 4	Unresponsive to pain; flaccid paralysis; brainstem reflexes and respirations absent; cardiovascular vital signs decreased

Decreased pulse, blood pressure, respiratory rate, temperature, and neuromuscular activity are indicative of **depressed** physiologic state caused by "functional" sympatholytics (agents that decrease cardiac function and vascular tone as well as sympathetic activity), cholinergic (muscarinic and nicotinic) agents, opioids, and sedative-hypnotic γ-aminobutyric acid (GABA)-ergic agents (Tables e50-1 and e50-2). Miosis is also common and most pronounced in opioid and cholinergic poisoning. The latter is distinguished from other depressant syndromes by the presence of muscarinic and nicotinic signs and symptoms (Table e50–1). Pronounced cardiovascular depression in the absence of significant CNS depression suggests a direct or peripherally acting sympatholytic. In contrast, in opioid and sedative-hypnotic poisoning, vital sign changes are secondary to depression of CNS cardiovascular and respiratory centers (or consequent hypoxemia), and significant abnormalities in these parameters do not occur until there is a marked decrease in the level of consciousness (grade 3 or 4 physiologic depression, Table e50–2). Other clues that suggest the cause of physiologic depression include cardiac arrhythmias and conduction disturbances (due to antiarrhythmics, β-adrenergic antagonists, calcium-channel blockers, digitalis glycosides, propoxyphene, and cyclic antidepressants), mydriasis [due to tricyclic antidepressants, some antiarrhythmics, meperidine, and diphenoxylate-atropine (Lomotil)], nystagmus (due to sedative-hypnotics), and seizures (due to cholinergic agents, propoxyphene, cyclic antidepressants).

The **discordant** physiologic state is characterized by mixed vital sign and neuromuscular abnormalities as observed in poisoning by asphyxiants, CNS syndromes, membrane-active agents, and anion-gap metabolic acidosis (AGMA) inducers (Table e50–1). In these conditions, manifestations of physiologic stimulation and physiologic depression occur together or at different times during the clinical course. For example, membrane-active agents can cause simultaneous coma, seizures, hypotension, and tachyarrhythmias. Alternatively, vital signs may be normal but the patient has altered mental status or is obviously sick or clearly symptomatic. Early, pronounced vital sign and mental status changes suggest asphyxiant or membrane-active agent poisoning; the lack of such abnormalities suggests an AGMA inducer; and marked neuromuscular dysfunction without significant vital sign abnormalities suggests a CNS syndrome.

A **normal** physiologic status and physical examination may be due to a nontoxic exposure, psychogenic illness, or poisoning by "toxic time-bombs," agents that are slowly absorbed, slowly distributed to their sites of action, require metabolic activation, or disrupt metabolic processes (Table e50–1). Because so many medications are now reformulated in a once-a-day form for patient convenience and adherence, "toxic time-bombs" are increasingly common. Diagnosing a nontoxic exposure requires that the identity of the exposure agent be known or that a toxic time-bomb exposure has been excluded and that the time since exposure exceeds the longest known or predicted interval between exposure and peak toxicity. Psychogenic illness (fear of being poisoned, mass hysteria) may also occur after a nontoxic exposure and should be considered when symptoms are inconsistent with the exposure history. Anxiety reactions resulting from a nontoxic exposure can cause mild physiologic stimulation (Table e50–2) and be indistinguishable from toxicologic causes (Table e50–1) without ancillary testing or a suitable period of observation.

Laboratory assessment may be helpful in the differential diagnosis. An increased AGMA is most common in

advanced methanol, ethylene glycol, and salicylateintoxication but can occur in any poisoning that results in hepatic, renal, or respiratory failure; seizures; or shock. The serum lactate concentration is more commonly low (less than the anion gap) in the former and high (nearly equal to the anion gap) in the latter. An abnormally low anion gap can be due to elevated blood levels of bromide, calcium, iodine, lithium, or magnesium. An increased osmolal gap—a difference between the serum osmolality (measured by freezing point depression) and that calculated from the serum sodium, glucose, and blood urea nitrogen of >10 mmol/L—suggests the presence of a low-molecular-weight solute such as acetone; an alcohol (benzyl, ethanol, isopropanol, methanol); a glycol (diethylene, ethylene, propylene); ether (ethyl, glycol); or an "unmeasured" cation (calcium, magnesium) or sugar (glycerol, mannitol, sorbitol). Ketosis suggests acetone, isopropyl alcohol, salicylate poisoning, or alcoholic ketoacidosis. Hypoglycemia may be due to poisoning with β-adrenergic blockers, ethanol, insulin, oral hypoglycemic agents, quinine, and salicylates, whereas hyperglycemia can occur in poisoning with acetone, β-adrenergic agonists, caffeine, calcium channel blockers, iron, theophylline, or N-3-pyridylmethyl-N'-p-nitrophenylurea [PNU (Vacor)]. Hypokalemia can be caused by barium, β-adrenergic agonists, caffeine, diuretics, theophylline, or toluene; hyperkalemia suggests poisoning with an α-adrenergic agonist, a β-adrenergic blocker, cardiac glycosides, or fluoride. Hypocalcemia may be seen in ethylene glycol, fluoride, and oxalate poisoning.

The *electrocardiogram* (ECG) can be useful for rapid diagnostic purposes. Bradycardia and atrioventricular block may occur in patients poisoned by α-adrenergic agonists, antiarrhythmic agents, beta blockers, calcium channel blockers, cholinergic agents (carbamate and organophosphate insecticides), cardiac glycosides, lithium, or tricyclic antidepressants. QRS- and QT-interval prolongation may be caused by hyperkalemia, various antidepressants, and other membrane-active drugs (Table e50–1). Ventricular tachyarrhythmias may be seen in poisoning with cardiac glycosides, fluorides, membrane-active drugs, methylxanthines, sympathomimetics, antidepressants, and agents that cause hyperkalemia or potentiate the effects of endogenous catecholamines (e.g., chloral hydrate, aliphatic and halogenated hydrocarbons).

Radiologic studies may occasionally be useful. Pulmonary edema [adult respiratory distress syndrome (ARDS)] can be caused by poisoning with carbon monoxide, cyanide, an opioid, paraquat, phencyclidine, a sedative-hypnotic, or salicylate; by inhalation of irritant gases, fumes, or vapors (acids and alkali, ammonia, aldehydes, chlorine, hydrogen sulfide, isocyanates, metal oxides, mercury, phosgene, polymers); or by prolonged anoxia, hyperthermia, or shock. Aspiration pneumonia is common in patients with coma, seizures, and petroleum distillate aspiration. The presence of radiopaque densities on abdominal x-rays suggests the ingestion of calcium salts, chloral hydrate, chlorinated hydrocarbons, heavy metals, illicit drug packets, iodinated compounds, potassium salts, enteric-coated tablets, or salicylates.

Toxicologic analysis of urine and blood (and occasionally of gastric contents and chemical samples) can sometimes confirm or rule out suspected poisoning. Interpretation of laboratory data requires knowledge of the qualitative and quantitative tests used for screening and confirmation (enzyme-multiplied, fluorescence polarization, and radio-immunoassays; colorimetric and fluorometric assays; thin-layer, gas-liquid, or high-performance liquid chromatography; gas chromatography; mass spectrometry), their sensitivity (limit of detection) and specificity, the preferred biologic specimen for analysis, and the optimal time of specimen sampling. Personal communication with the hospital laboratory is essential to understand institutional testing capabilities and limitations.

Rapid *qualitative* hospital-based urine tests for drugs of abuse are only screening tests that cannot confirm the exact identity of the detected substance and should not be considered diagnostic or used for forensic purposes: False-positive and false-negative results are common. A positive screen may result from other pharmaceuticals that interfere with laboratory analysis (e.g., fluoroquinolones commonly cause "false-positive" opiate screens). Confirmatory testing with gas chromatography/mass spectrometry (GC/MS) can be requested, but often takes weeks to obtain a reported result. A negative screening result may mean the substance is not detectable by the test used or that its concentration is too low for detection at the time of sampling. In the latter case, repeating the test at a later time may yield a positive result. Patients symptomatic from drugs of abuse often require immediate management based on the history, physical examination, and observed toxidrome without laboratory confirmation (e.g., apnea from opioid intoxication). When the patient is asymptomatic, or when the clinical picture is consistent with the reported history, several studies have shown qualitative screening is neither clinically useful nor cost-effective. Thus, qualitative drug screens are of greatest value when evaluating patients with severe or unexplained toxicity such as coma, seizures, cardiovascular instability, metabolic or respiratory acidosis, and nonsinus cardiac rhythms. In contrast to qualitative drug screens, *quantitative* serum tests are useful for evaluating patients poisoned with acetaminophen (Chap. 305), alcohols (including ethylene glycol and methanol), anticonvulsants, barbiturates, digoxin, heavy metals, iron, lithium, salicylate, and theophylline, as well as for carboxyhemoglobin and methemoglobin. The serum concentration in these cases guides clinical management, and results can often be available within an hour.

The *response to antidotes* is sometimes useful for diagnostic purposes. Resolution of altered mental status and abnormal vital signs within minutes of IV administration of dextrose, naloxone, or flumazenil is virtually diagnostic of hypoglycemia, narcotic poisoning, and benzodiazepineintoxication, respectively. The prompt reversal of dystonic (extrapyramidal) signs and symptoms following an IV dose of benztropine or diphenhydramine confirms a drug etiology. Although complete reversal of both central and peripheral manifestations of anticholinergic poisoning by physostigmine is diagnostic of this condition, physostigmine may cause some arousal in patients with CNS depression of any etiology.

Treatment: Poisoning and Drug Overdose

GENERAL PRINCIPLES

Treatment goals include support of vital signs, prevention of further poison absorption (decontamination), enhancement of poison elimination, administration of specific antidotes, and prevention of reexposure **(Table e50–3)**. Specific treatment depends on the identity of the poison, the route and amount of exposure, the time of presentation relative to the time of exposure, and the severity of poisoning. Knowledge of the offending agents' pharmacokinetics and pharmacodynamics is essential.

Table e50–3 Fundamentals of Poisoning Management

Supportive Care	
Airway protection Oxygenation/ventilation Treatment of arrhythmias Hemodynamic support	Treatment of seizures Correction of temperature abnormalities Correction of metabolic derangements Prevention of secondary complications
Prevention of Further Poison Absorption	
Gastrointestinal decontamination Gastric lavage Activated charcoal Whole-bowel irrigation Dilution Endoscopic/surgical removal	Decontamination of other sites Eye decontamination Skin decontamination Body cavity evacuation
Enhancement of Poison Elimination	
Multiple-dose activated charcoal Alteration of urinary pH Chelation	Extracorporeal removal Hemodialysis Hemoperfusion Hemofiltration Plasmapheresis Exchange transfusion Hyperbaric oxygenation
Administration of Anti-dotes	
Neutralization by antibodies Neutralization by chemical binding	Metabolic antagonism Physiologic antagonism
Prevention of Reexposure	
Adult education Child-proofing	Notification of regulatory agencies Psychiatric referral

During the *pretoxic phase*, prior to the onset of poisoning, decontamination is the highest priority, and treatment is based solely on the history. The maximum potential toxicity based on the greatest possible exposure should be assumed. Since decontamination is more effective when accomplished soon after exposure, and when the patient is asymptomatic, the initial history and physical examination should be focused and brief. It is also advisable to establish IV access and initiate cardiac monitoring, particularly in patients with potentially serious ingestions or unclear histories.

When an accurate history is not obtainable and a poison causing delayed toxicity ("toxic time-bomb") or irreversible damage is suspected, blood and urine should be sent for appropriate toxicologic screening and quantitative analysis. During absorption and distribution, blood levels may be greater than those in tissue and may not correlate with toxicity. However, high blood levels of agents whose metabolites are more toxic than the parent compound (acetaminophen, ethylene glycol, or methanol) may indicate the need for additional interventions (antidotes, dialysis). Most patients who remain or become asymptomatic 6 h after ingestion are unlikely to develop subsequent toxicity and can be discharged safely. Longer observation will be necessary for patients who have ingested toxic time-bombs, agents that are slowly absorbed, slowly distributed to their sites of action, require metabolic activation, or disrupt metabolic processes (Table e50–1). During the *toxic phase*, the time between the onset of poisoning and the peak effects, management is based primarily on clinical and laboratory findings. *Effects after an overdose usually begin sooner, peak later, and last longer than they do after a therapeutic dose*. A drug's published pharmacokinetic profile in standard references such as the *Physician's Desk Reference* (*PDR*) is usually different from its toxicokinetic profile in overdose. Resuscitation and stabilization are the first priority. Symptomatic patients should have an IV line, oxygen saturation determination, cardiac monitoring, and continuous observation. Baseline laboratory, ECG, and x-ray evaluation may also be appropriate. Intravenous glucose (unless the serum level is documented to be normal), naloxone, and thiamine should be considered in patients with altered mental status, particularly those with coma or seizures. Decontamination should also be considered, but it is less likely to be effective during this phase than during the pretoxic one.

Measures that enhance poison elimination may shorten the duration and severity of the toxic phase. However, they are not without risk, which must be weighed against the potential benefit. Diagnostic certainty (usually via laboratory confirmation) is generally a prerequisite. Intestinal (or "gut") dialysis with repetitive doses of activated charcoal (also termed *multidose activated charcoal*) can enhance the elimination of selected poisons such as theophylline or carbamazepine. Urinary alkalinization may enhance the elimination of salicylates and a small number of other poisons. Chelation therapy can enhance the elimination of selected metals. Extracorporeal elimination methods are effective for many poisons, but their expense and risk make their use reasonable only in patients who would otherwise have an unfavorable outcome.

During the *resolution phase* of poisoning, supportive care and monitoring should continue until clinical, laboratory, and ECG abnormalities have resolved. Since chemicals are eliminated sooner from the blood than from tissues, blood levels are usually lower than tissue levels during this phase and again may not correlate with toxicity. This is particularly true when extracorporeal elimination procedures are used. Redistribution from tissues may cause a rebound increase in the blood level after termination of these procedures. When a metabolite is responsible for toxic effects, continued treatment might be necessary in the absence of clinical toxicity or abnormal laboratory studies.

SUPPORTIVE CARE

The goal of supportive therapy is to maintain physiologic homeostasis until detoxification is accomplished and to prevent and treat secondary complications such as aspiration, bedsores, cerebral and pulmonary edema, pneumonia, rhabdomyolysis, renal failure, sepsis, thromboembolic disease, coagulopathy, and generalized organ dysfunction due to hypoxemia or shock.

Admission to an intensive care unit is indicated for the following: patients with severe poisoning (coma, respiratory depression, hypotension, cardiac conduction abnormalities, cardiac arrhythmias, hypothermia or hyperthermia, seizures); those needing close monitoring, antidotes, or enhanced elimination therapy; those showing progressive clinical deterioration; and those with significant underlying medical problems. Patients with mild to moderate toxicity can be managed on a general medical service, intermediate care unit, or emergency department observation area, depending on the anticipated duration and level of monitoring needed (intermittent clinical observation versus continuous clinical, cardiac, and respiratory monitoring). Patients who have attempted suicide require continuous observation and measures to prevent self-injury until they are no longer suicidal.

Respiratory Care

Endotracheal intubation for protection against the aspiration of gastrointestinal contents is of paramount importance in patients with CNS depression or seizures as this complication can increase morbidity and mortality rates. Mechanical ventilation may be necessary for patients with respiratory depression or hypoxemia and to facilitate therapeutic sedation or paralysis in order to prevent or treat hyperthermia, acidosis, and rhabdomyolysis associated with neuromuscular hyperactivity. Since clinical assessment of respiratory function can be inaccurate, the need for oxygenation and ventilation is best determined by continuous pulse oximetry or arterial blood-gas analysis. The gag reflex is not a reliable indicator of the need for intubation. A patient with CNS depression may maintain airway patency while being stimulated but not if left alone. Drug-induced pulmonary edema is usually noncardiac rather than cardiac in origin, although profound CNS depression and cardiac conduction abnormalities suggest the latter. Measurement of pulmonary artery pressure may be necessary to establish the cause and direct appropriate therapy. Extracorporeal measures (membrane oxygenation, venoarterial perfusion, cardiopulmonary bypass) and partial liquid (perfluorocarbon) ventilation may be appropriate for severe but reversible respiratory failure.

Cardiovascular Therapy

Maintenance of normal tissue perfusion is critical for complete recovery to occur once the offending agent has been eliminated. If hypotension is unresponsive to volume expansion, treatment with norepinephrine, epinephrine, or high-dose dopamine may be necessary. Intraaortic balloon pump counterpulsation and venoarterial or cardiopulmonary perfusion techniques should be considered for severe but reversible cardiac failure. Bradyarrhythmias associated with hypotension generally should be treated as described in Chap. 232. Glucagon, calcium, and high-dose insulin with dextrose may be effective in beta blocker and calcium channel blocker poisoning. Antibody therapy may be indicated for cardiac glycoside poisoning.

Supraventricular tachycardia associated with hypertension and CNS excitation is almost always due to agents that cause generalized physiologic excitation (Table e50–1). Most cases are mild or moderate in severity and require only observation or nonspecific sedation with a benzodiazepine. In severe cases or those associated with hemodynamic instability, chest pain, or ECG evidence of ischemia, specific therapy is indicated. When the etiology is sympathetic hyperactivity, treatment with a benzodiazepine should be prioritized. Further treatment with a combined alpha and beta blocker (labetalol), a calcium channel blocker (verapamil or diltiazem), or a combination of a beta blocker and a vasodilator (esmolol and nitroprusside) may be considered for cases refractory to high doses of benzodiazepines. Treatment with an α-adrenergic antagonist (phentolamine) alone may sometimes be appropriate. If the cause is anticholinergic poisoning, physostigmine alone can be effective. Supraventricular tachycardia without hypertension is generally secondary to vasodilation or hypovolemia and responds to fluid administration.

For ventricular tachyarrhythmias due to tricyclic antidepressants and other membrane-active agents (Table e50–1), sodium bicarbonate is indicated, whereas class IA, IC, and III antiarrhythmic agents are contraindicated because of similar electrophysiologic effects. Although lidocaine and phenytoin are historically safe for ventricular tachyarrhythmias of any etiology, sodium bicarbonate should be considered first for any ventricular arrhythmia suspected to have a toxicologic etiology. Beta blockers can be hazardous if the arrhythmia is due to sympathetic hyperactivity. Magnesium sulfate and overdrive pacing (by isoproterenol or a pacemaker) may be useful in patients with torsades des pointes and prolonged QT intervals. Magnesium and anti-digoxin antibodies should be considered in patients with severe cardiac glycoside poisoning. Invasive (esophageal or intracardiac) ECG recording may be necessary to determine the origin (ventricular or supraventricular) of wide-complex tachycardias (Chap. 233). If the patient is hemodynamically stable, however, it is reasonable to simply observe him or her rather than to administer another potentially proarrhythmic agent. Arrhythmias may be resistant to drug therapy until underlying acid-base, electrolyte, oxygenation, and temperature derangements are corrected.

Central Nervous System Therapies

Neuromuscular hyperactivity and seizures can lead to hyperthermia, lactic acidosis, and rhabdomyolysis and should be treated aggressively. Seizures caused by excessive stimulation of catecholamine receptors (sympathomimetic or hallucinogen poisoning and drug withdrawal) or decreased activity of GABA (isoniazid poisoning) or glycine (strychnine poisoning) receptors are best treated with agents that enhance GABA activity, such as benzodiazepine or barbiturates. Since benzodiazepines and barbiturates act by slightly different mechanisms (the former increases the frequency, and the latter increases the duration of chloride channel opening in response to GABA), therapy with both may be effective when neither is effective alone. Seizures caused by isoniazid, which inhibits the synthesis of GABA at several steps by interfering with the

cofactor pyridoxine (vitamin B6), may require high doses of supplemental pyridoxine. Seizures resulting from membrane destabilization (beta blocker or cyclic antidepressant poisoning) require GABA enhancers (benzodiazepines first, barbiturates second). Phenytoin is contraindicated in toxicologic seizures: Animal and human data demonstrate worse outcomes after phenytoin loading, especially in theophylline overdose. For poisons with central dopaminergic effects (methamphetamine, phencyclidine) manifested by psychotic behavior, a dopamine receptor antagonist, such as haloperidol, may be useful. In anticholinergic and cyanide poisoning, specific antidotal therapy may be necessary. The treatment of seizures secondary to cerebral ischemia or edema or to metabolic abnormalities should include correction of the underlying cause. Neuromuscular paralysis is indicated in refractory cases. Electroencephalographic monitoring and continuing treatment of seizures are necessary to prevent permanent neurologic damage. Serotonergic receptor overstimulation in serotonin syndrome may be treated with cyproheptadine.

Other Measures

Temperature extremes, metabolic abnormalities, hepatic and renal dysfunction, and secondary complications should be treated by standard therapies.

PREVENTION OF POISON ABSORPTION

Gastrointestinal Decontamination

Whether or not to perform gastrointestinal decontamination, and which procedure to use, depends on the time since ingestion; the existing and predicted toxicity of the ingestant; the availability, efficacy, and contraindications of the procedure; and the nature, severity, and risk of complications. The efficacy of all decontamination procedures decreases with time, and there are insufficient data to support or exclude a beneficial effect when they are used >1 h after ingestion. The average time from ingestion to presentation for treatment is >1 h for children and >3 h for adults. Most patients will recover from poisoning uneventfully with good supportive care alone, but complications of gastrointestinal decontamination, particularly aspiration, can prolong this process. Hence, gastrointestinal decontamination should be performed selectively, not routinely, in the management of overdose patients. It is clearly unnecessary when predicted toxicity is minimal or the time of expected maximal toxicity has passed without significant effect.

Activated charcoal has comparable or greater efficacy, fewer contraindications and complications, and is less aversive and invasive than ipecac or gastric lavage; thus it is the preferred method of gastrointestinal decontamination in most situations. Activated charcoal suspension (in water) is given orally via a cup, straw, or small-bore nasogastric tube. The generally recommended dose is 1 g/kg body weight because of its convenience, although in vitro and in vivo studies have demonstrated charcoal adsorbs ≥90% of most substances when given in an amount equal to 10 times the weight of the substance. Palatability may be increased by adding a sweetener (sorbitol) or a flavoring agent (cherry, chocolate, or cola syrup) to the suspension. Charcoal adsorbs ingested poisons within the gut lumen, allowing the charcoal-toxin complex to be evacuated with stool. Charged (ionized) chemicals such as mineral acids, alkalis, and highly dissociated salts of cyanide, fluoride, iron, lithium, and other inorganic compounds are not well adsorbed by charcoal. In animal and human volunteer studies, charcoal decreases the absorption of ingestants by an average of 73% when given within 5 min of ingestant administration, 51% when given at 30 min, and 36% at 60 min. Side effects of charcoal include nausea, vomiting, and diarrhea or constipation. Charcoal may also prevent the absorption of orally administered therapeutic agents. Complications include mechanical obstruction of the airway, aspiration, vomiting, and bowel obstruction and infarction caused by inspissated charcoal. Charcoal is not recommended for patients who have ingested corrosives because it obscures endoscopy.

Gastric lavage should be considered for life-threatening poisons that cannot be treated effectively with other decontamination, elimination, or antidotal therapies (e.g., colchicine). Gastric lavage is performed by sequentially administering and aspirating ~5 mL fluid per kilogram of body weight through a no. 40 French orogastric tube (no. 28 French tube for children). Except for infants, where normal saline is recommended, tap water is acceptable. The patient should be placed in Trendelenburg and left lateral decubitus positions to prevent aspiration (even if an endotracheal tube is in place). Lavage decreases ingestant absorption by an average of 52% if performed within 5 min of ingestion administration, 26% if performed at 30 min, and 16% if performed at 60 min. Significant amounts of ingested drug are recovered in <10% of patients. Aspiration is a common complication (occurring in up to 10% of patients), especially when lavage is performed improperly. Serious complications (esophageal and gastric perforation, tube misplacement in the trachea) occur in ~1% of patients. For this reason, the physician should personally insert the lavage tube and confirm its placement, and the patient must be cooperative during the procedure. Gastric lavage is contraindicated in corrosive or petroleum distillate ingestions because of the respective risks of gastroesophageal perforation and aspiration pneumonitis. It is also contraindicated in those with a compromised unprotected airway and those at risk for hemorrhage or perforation due to esophageal or gastric pathology or recent surgery. Finally, gastric lavage is absolutely contraindicated in combative patients or those who refuse, as most published complications involve patient resistance to the procedure.

Syrup of ipecac, an emetogenic agent that was once the most commonly used decontamination procedure, no longer has a role in poisoning management. Even the American Academy of Pediatrics (AAP), traditionally the strongest proponent of ipecac, issued a policy statement in 2003 recommending that ipecac should no longer be used in poisoning treatment. Chronic ipecac use (by patients with anorexia nervosa or bulimia) has been reported to cause electrolyte and fluid abnormalities, cardiac toxicity, and myopathy.

Whole-bowel irrigation is performed by administering a bowel-cleansing solution containing electrolytes and polyethylene glycol (Golytely, Colyte) orally or by gastric tube at a rate of 2 L/h (0.5 L/h in children) until rectal effluent is clear. The patient must be in a sitting position. Although data are limited, whole-bowel irrigation

appears to be as effective as other decontamination procedures in volunteer studies. It is most appropriate for those who have ingested foreign bodies, packets of illicit drugs, slow-release or enteric-coated medications, and agents that are poorly adsorbed by charcoal (e.g., heavy metals). It is contraindicated in patients with bowel obstruction, ileus, hemodynamic instability, and compromised unprotected airways.

Cathartics are salts (disodium phosphate, magnesium citrate and sulfate, sodium sulfate) or saccharides (mannitol, sorbitol) that historically have been given with activated charcoal to promote the rectal evacuation of gastrointestinal contents. However, no animal, volunteer, or clinical data have ever demonstrated any decontamination benefit from cathartics. Abdominal cramps, nausea, and occasional vomiting are side effects. Complications of repeated dosing include severe electrolyte disturbances and excessive diarrhea. Cathartics are contraindicated in patients who have ingested corrosives and in those with preexisting diarrhea. Magnesium-containing cathartics should not be used in patients with renal failure.

Dilution (i.e., drinking 5 mL/kg of body weight of water, another clear liquid, or milk) is recommended only after the ingestion of corrosives (acids, alkali). It may increase the dissolution rate (and hence absorption) of capsules, tablets, and other solid ingestants and should *not* be used in these circumstances.

Endoscopic or surgical removal of poisons may be useful in rare situations, such as ingestion of a potentially toxic foreign body that fails to transit the gastrointestinal tract, a potentially lethal amount of a heavy metal (arsenic, iron, mercury, thallium), or agents that have coalesced into gastric concretions or bezoars (heavy metals, lithium, salicylates, sustained-release preparations). Patients who become toxic from cocaine due to its leakage from ingested drug packets require immediate surgical intervention.

Decontamination of Other Sites

Immediate, copious flushing with water, saline, or another available clear, drinkable liquid is the initial treatment for topical exposures (exceptions include alkali metals, calcium oxide, phosphorus). Saline is preferred for eye irrigation. A triple wash (water, soap, water) may be best for dermal decontamination. Inhalational exposures should be treated initially with fresh air or oxygen. The removal of liquids from body cavities such as the vagina or rectum is best accomplished by irrigation. Solids (drug packets, pills) should be removed manually, preferably under direct visualization.

ENHANCEMENT OF POISON ELIMINATION

Although the elimination of most poisons can be accelerated by therapeutic interventions, the pharmacokinetic efficacy (removal of drug at a rate greater than that accomplished by intrinsic elimination) and clinical benefit (shortened duration of toxicity or improved outcome) of such interventions are often more theoretical than proven. Hence, the decision to use such measures should be based on the actual or predicted toxicity and the potential efficacy, cost, and risks of therapy.

Multiple-Dose Activated Charcoal

Repetitive oral dosing with charcoal can enhance the elimination of previously absorbed substances by binding them within the gut as they are excreted in the bile, secreted by gastrointestinal cells, or passively diffuse into the gut lumen (reverse absorption or enterocapillary exsorption). Doses of 0.5–1 g/kg body weight every 2–4 h, adjusted downward to avoid regurgitation in patients with decreased gastrointestinal motility, are generally recommended. Pharmacokinetic efficacy approaches that of hemodialysis for some agents (e.g., phenobarbital, theophylline). Multiple-dose therapy should be considered only for selected agents (theophylline, phenobarbital, carbamazepine, dapsone, quinine). Complications include intestinal obstruction, pseudoobstruction, and nonocclusive intestinal infarction in patients with decreased gut motility. Sorbitol and other cathartics are absolutely contraindicated when administering multiple doses of activated charcoal because of electrolyte and fluid shifts.

Urinary Alkalinization

Ion trapping via alteration of urine pH may prevent the renal reabsorption of poisons that undergo excretion by glomerular filtration and active tubular secretion. Since membranes are more permeable to nonionized molecules than to their ionized counterparts, acidic (low-pK_a) poisons are ionized and trapped in alkaline urine, whereas basic ones become ionized and trapped in acid urine. Urinary alkalinization (producing a urine pH ≥7.5 and a urine output of 3–6 mL/kg body weight per hour by adding sodium bicarbonate to an IV solution) enhances the excretion of chlorophenoxyacetic acid herbicides, chlorpropamide, diflunisal, fluoride, methotrexate, phenobarbital, sulfonamides, and salicylates. Contraindications include congestive heart failure, renal failure, and cerebral edema. Acid-base, fluid, and electrolyte parameters should be monitored carefully. While making theoretical sense for some overdoses (amphetamines), acid diuresis is never indicated and is potentially harmful.

Extracorporeal Removal

Hemodialysis, charcoal or resin hemoperfusion, hemofiltration, plasmapheresis, and exchange transfusion are capable of removing any toxin from the bloodstream. Agents most amenable to enhanced elimination by dialysis have low molecular mass (<500 Da), high water solubility, low protein binding, small volumes of distribution (<1 L/kg body weight), prolonged elimination (long half-life), and high dialysis clearance relative to total-body clearance. Molecular weight, water solubility, or protein binding do not limit the efficacy of the other forms of extracorporeal removal.

Dialysis should be considered in cases of severe poisoning due to carbamazepine, ethylene glycol, isopropyl alcohol, lithium, methanol, theophylline, salicylates, and valproate. Although hemoperfusion may be more effective in removing some of these poisons, it does not correct associated acid-base and electrolyte abnormalities, and most hospitals no longer have hemoperfusion cartridges readily available. Fortunately,

recent advances in hemodialysis technology make it effective for removing poisons such as caffeine, carbamazepine, and theophylline. Both techniques require central venous access and systemic anticoagulation and may result in transient hypotension. Hemoperfusion may also cause hemolysis, hypocalcemia, and thrombocytopenia. Peritoneal dialysis and exchange transfusion are less effective but may be used when other procedures are either not available, contraindicated, or technically difficult (e.g., in infants). Exchange transfusion may be indicated in the treatment of severe arsine- or sodium chlorate –induced hemolysis, methemoglobinemia, and sulfhemoglobinemia. Although hemofiltration can enhance elimination of aminoglycosides, vancomycin, and metal-chelate complexes, the roles of hemofiltration and plasmapheresis in the treatment of poisoning are not yet defined.

Candidates for extracorporeal removal therapies include patients with severe toxicity who deteriorate despite aggressive supportive therapy; those with potentially prolonged, irreversible, or fatal toxicity; those with dangerous blood levels of toxins; those who lack the capacity for self-detoxification because of liver or renal failure; and those with a serious underlying illness or complication that will adversely affect recovery.

Other Techniques

The elimination of heavy metals can be enhanced by chelation, and the removal of carbon monoxide can be accelerated by hyperbaric oxygenation.

ADMINISTRATION OF ANTIDOTES

Antidotes counteract the effects of poisons by neutralizing them (e.g., antibody-antigen reactions, chelation, chemical binding) or by antagonizing their physiologic effects (e.g., activation of opposing nervous system activity, provision of competitive metabolic or receptor substrate). Poisons or conditions with specific antidotes include acetaminophen, anticholinergic agents, anticoagulants, benzodiazepines, beta blockers, calcium channel blockers, carbon monoxide, cardiac glycosides, cholinergic agents, cyanide, drug-induced dystonic reactions, ethylene glycol, fluoride, heavy metals, hypoglycemic agents, isoniazid, membrane-active agents, methemoglobinemia, opioids, sympathomimetics, and a variety of envenomations. Intravenous lipid emulsion has been shown to be a successful antidote for poisoning from various anesthetics and membrane-active agents, but the exact mechanism of benefit is still under investigation. Antidotes can significantly reduce morbidity and mortality rates but are potentially toxic if used for inappropriate reasons. Since their safe use requires correct identification of a specific poisoning or syndrome, details of antidotal therapy are discussed with the conditions for which they are indicated **(Table e50–4)**.

Table e50–4 Pathophysiologic Features and Treatment of Specific Toxic Syndromes and Poisonings

Physiologic Condition, Causes	Examples	Mechanism of Action	Clinical Features	Specific Treatments
Stimulated				
Sympathetics (see also Chap. 394)				
Sympathomimetics	α_1-Adrenergic agonists (decongestants): phenylephrine, phenylpropanolamine β_2-Adrenergic agonists (bronchodilators): albuterol, terbutaline Nonspecific adrenergic agonists: amphetamines, cocaine, ephedrine	Stimulation of central and peripheral sympathetic receptors directly or indirectly (by promoting the release or inhibiting the reuptake of norepinephrine and sometimes dopamine)	Physiologic stimulation (Table e50–2); reflex bradycardia can occur with selective α_1 agonists; β agonists can cause hypotension and hypokalemia.	Phentolamine, a nonselective α_1-adrenergic receptor antagonist, for severe hypertension due to α_1-adrenergic agonists; propranolol, a nonselective β blocker, for hypotension and tachycardia due to β_2 agonists; labetalol, a β blocker with α-blocking activity, or phentolamine with esmolol, metoprolol, or other cardioselective β blocker for hypertension with tachycardia due to nonselective agents (β blockers, if used alone, can exacerbate hypertension and vasospasm due to unopposed α stimulation); benzodiazepines; propofol.

Ergot alkaloids	Ergotamine, methysergide, bromocriptine, pergolide	Stimulation and inhibition of serotonergic and α-adrenergic receptors; stimulation of dopamine receptors	Physiologic stimulation (Table e50–2); formication; vasospasm with limb (isolated or generalized), myocardial, and cerebral ischemia progressing to gangrene or infarction; hypotension, bradycardia, and involuntary movements can also occur.	Nitroprusside or nitroglycerine for severe vasospasm; prazosin (an α1 blocker), captopril, nifedipine, and cyproheptadine (a serotonin receptor antagonist) for mild-to-moderate limb ischemia; dopamine receptor antagonists (an-tipsychotics) for hallucinations and movement disorders
Methylxanthines	Caffeine, theophylline	Inhibition of adenosine synthesis and adenosine receptor antagonism; stimulation of epinephrine and norepinephrine release; inhibition of phosphodiesterase resulting in increased intracellular cyclic adenosine and guanosine monophosphate	Physiologic stimulation (Table e50–2); pronounced gastrointestinal symptoms and β agonist effects (see above). Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Propranolol, a nonselective β blocker, for tachycardia with hypotension; any β blocker for supraventricular or ventricular tachycardia without hypotension; elimination enhanced by multiple-dose charcoal, hemoperfusion, and hemodialysis; indications for hemoperfusion or hemodialysis include unstable vital signs, seizures, and a theophylline level of 80–100 μg/mL after acute overdose and 40–60 μg/mL with chronic exposure.
Monoamine oxidase inhibitors	Phenelzine, tranylcypromine, selegiline	Inhibition of monoamine oxidase resulting in impaired metabolism of endogenous catecholamines and exogenous sympathomimetic agents	Delayed or slowly progressive physiologic stimulation (Table e50–2); terminal hypotension and bradycardia in severe cases.	Short-acting agents (e.g., nitroprusside, esmolol) for severe hypertension and tachycardia; direct-acting sympathomimetics (e.g., norepinephrine, epinephrine) for hypotension and bradycardia
Anticholinergics				
Antihistamines	Diphenhydramine, doxylamine, pyrilamine	Inhibition of central and postganglionic parasympathetic muscarinic cholinergic receptors. At high doses, amantadine, diphenhydramine, orphenadrine, phenothiazines, and tricyclic antidepressants have additional nonanticholinergic activity (see below).	Physiologic stimulation (Table e50–2); dry skin and mucous membranes, decreased bowel sounds, flushing, and urinary retention; myoclonus and picking activity. Central effects may occur without significant autonomic dysfunction.	Physostigmine, an acetylcholinesterase inhibitor (see below) for delirium, hallucinations, and neuromuscular hyperactivity. Contraindications include asthma, nonanticholinergic cardiovascular toxicity (e.g., cardiac conduction abnormalities, hypotension, and ventricular arrhythmias).
Antiparkinsonian agents	Amantadine, trihexyphenidyl			
Antipsychotics	Chlorpromazine,			

	olanzapine, quetiapine, thioridazine			
Antispasmodics	Clidinium, dicyclomine			
Belladonna alkaloids	Atropine, hyoscyamine, scopolamine			
Cyclic antidepressants	Amitriptyline, doxepin, imipramine			
Muscle relaxants	Cyclobenzaprine, orphenadrine			
Mushrooms and plants	*Amanita muscaria* and *A. pantherina,* henbane, jimson weed, nightshade			
Depressed				
Sympatholytics				
α_2-Adrenergic agonists	Clonidine, guanabenz, tetrahydrozoline and other imidazoline decongestants, tizanidine and other imidazoline muscle relaxants	Stimulation of α_2-adrenergic receptors leading to inhibition of CNS sympathetic outflow; activity at nonadrenergic imidazoline binding sites also contributes to CNS effects.	Physiologic depression (Table e50–2), miosis. Transient initial hypertension may be seen.	Dopamine and norepinephrine for hypotension. Atropine for symptomatic bradycardia. Naloxone for CNS depression (inconsistently effective).
Antipsychotics	Chlorpromazine, clozapine, haloperidol, risperidone, thioridazine	Inhibition of α-adrenergic, dopaminergic, histaminergic, muscarinic, and serotonergic receptors. Some agents also inhibit sodium, potassium, and calcium channels.	Physiologic depression (Table e50–2), miosis, anticholinergic effects (see above), extrapyramidal reactions (see below), tachycardia. Cardiac conduction delays (increased PR, QRS, JT, and QT intervals) with ventricu-lar tachydysrhythmias, including torsades des pointes, can sometimes develop.	Sodium bicarbonate for ventricular tachydysrhythmias associated with QRS prolongation. Magnesium, isoproterenol, and overdrive pacing for torsades des pointes. Avoid class IA, IC, and III antiarrhythmics. Intravenous lipid emulsion therapy may be beneficial in some cases.
β-Adrenergic blockers	Cardioselective (β_1) blockers: atenolol, esmolol, metoprolol Nonselective (β_1 and β_2) blockers: nadolol, propranolol, timolol Partial β agonists: acebutolol, pindolol α_1 Antagonists: carvedilol, labetalol Membrane-active agents: acebutolol, propranolol, sotalol	Inhibition of β-adrenergic receptors (class II antiarrhythmic effect). Some agents have activity at additional receptors or have membrane effects (see below).	Physiologic depression (Table e50–2), atrioventricular block, hypoglycemia, hyperkalemia, seizures. Partial agonists can cause hypertension and tachycardia. Sotalol can cause increased QT interval and ven-tricular tachydysrhythmias. Onset may be delayed after sotalol and sustained-release formulation overdose.	Glucagon for hypotension and symptomatic bradycardia. Atropine, isoproterenol, dopamine, dobutamine, epinephrine, and norepinephrine may sometimes be effective. High-dose insulin (with glucose and potassium to maintain euglycemia and normokalemia), electrical pacing, and mechanical cardiovascular support for refractory cases.
Calcium channel blockers	Diltiazem, nifedipine and other dihydropyridine derivatives, verapamil	Inhibition of slow (type L) cardiovascular calcium channels (class IV antiarrhythmic ef-fect).	Physiologic depression (Table e50–2), atrioventricular block, organ ischemia and infarction.	Calcium and glucagon for hypotension and symptomatic bradycardia. Dopamine, epinephrine,

			hyperglycemia, seizures. Hypotension is usually due to decreased vascular resistance rather than to decreased cardiac output. Onset may be delayed for ≥12 h after overdose of sustained-release formulations.	norepinephrine, atropine, and isoproterenol are less often effective but can be used adjunctively. High-dose insulin (with glucose and potassium to maintain euglycemia and normokalemia), intravenous lipid emulsion therapy, electrical pacing, and mechanical cardiovascular support for refractory cases.
Cardiac glycosides	Digoxin, endogenous cardioactive steroids, foxglove and other plants, toad skin secretions (*Bufonidae* sp.)	Inhibition of cardiac Na^+, K^+-ATPase membrane pump.	Physiologic depression (Table e50–2); gastrointestinal, psychiatric, and visual symptoms; atrioventricular block with or without concomitant supraventricular tachyarrhythmia; ventricular tachyarrhythmias. Hyperkalemia in acute poisoning. Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Digoxin-specific antibody fragments for hemodynamically compromising dysrhythmias, Mobitz II or third-degree atrioventricular block, hyperkalemia (>5.5 meq/L; in acute poisoning only). Temporizing measures include atropine, dopamine, epinephrine, and external cardiac pacing for bradydysrhythmias and magnesium, lidocaine, or phenytoin, for ventricular tachydysrhythmias. Internal cardiac pacing and cardioversion can increase ventricular irritability and should be reserved for refractory cases.
Cyclic antidepressants	Amitriptyline, doxepin, imipramine	Inhibition of α-adrenergic, dopaminergic, GABA-ergic, histaminergic, mus-carinic, and serotonergic receptors; inhibition of sodium channels (see membrane-active agents); inhibition of norepinephrine and serotonin reuptake.	Physiologic depression (Table e50–2), seizures, tachycardia, cardiac conduction delays (in-creased PR, QRS, JT, and QT intervals; terminal QRS right-axis deviation) with aberrancy and ventricular tachydysrhythmias. Anticholinergic toxidrome (see above).	Hypertonic sodium bicarbonate (or hypertonic saline) and lidocaine for ventricular tachydysrhythmias associated with QRS prolongation. Use of phenytoin is controversial. Avoid class IA, IC, and III antiarrhythmics.
Cholinergics				
Acetylcholinesterase inhibitors	Carbamate insecticides (aldicarb, carbaryl, propoxur) and medicinals (neostigmine, physostigmine, tacrine); nerve gases (sarin, soman, tabun, VX) organophosphate insecticides (diazinon,	Inhibition of acetylcho-linesterase leading to increased synaptic acetylcholine at muscarinic and nicotinic cholinergic receptor sites	Physiologic depression (Table e50–2). Muscarinic signs and symptoms: seizures, excessive secretions (lacrimation, salivation, bronchorrhea and wheezing, diaphoresis), and increased bowel	Atropine for muscarinic signs and symptoms. Pralidoxime (2-PAM), a cholinesterase reactiva-tor, for nicotinic signs and symptoms due to organophosphates, nerve gases, or an unknown anticholinesterase.

	chlorpyrifos-ethyl, malathion)		and bladder activity with nausea, vomiting, diarrhea, abdominal cramps, and incontinence of feces and urine. Nicotinic signs and symptoms: hypertension, tachycardia, muscle cramps, fasciculations, weakness, and paralysis. Death is usually due to respiratory failure. Cholinesterase activity in plasma and red cells <50% of normal in acetylcholinesterase inhibitor poisoning.	
Muscarinic agonists	Bethanechol, mushrooms (*Boletus*, *Clitocybe*, *Inocybe* spp.), pilocarpine	Stimulation of CNS and postganglionic parasympathetic cholinergic (muscarinic) receptors		
Nicotinic agonists	Lobeline, nicotine (tobacco)	Stimulation of preganglionic sympathetic and parasympathetic and striated muscle (neuromuscular junction) cholinergic (nicotine) receptors		
Sedative-hypnotics (see also Chap. 393)				
Anticonvulsants	Carbamazepine, ethosuximide, felbamate, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, phenytoin, tiagabine, topiramate, valproate, zonisamide	Potentiation of the inhibitory effects of GABA by binding to the neuronal GABA-A chloride channel receptor complex and increasing the frequency or duration of chloride channel opening in response to GABA stimulation. Baclofen and, to some extent, GHB act at the GABA-B receptor complex; meprobamate, its metabolite carisoprodol, felbamate, and orphena-drine antagonize *N*-methyl-D-aspartate (NDMA) excitatory receptors; ethosuximide, valproate, and zonisamide decrease conduction through T-type calcium channels; valproate decreases GABA degradation, and tiagabine blocks GABA reuptake; carbamazepine, lamotrigine, oxcarbazepine, phenytoin,	Physiologic depression (Table e50–2), nystagmus. Delayed absorption can occur with carbamazepine, phenytoin, and valproate. Myoclonus, seizures, hypertension, and tachyarrhythmias can occur with baclofen, carbamazepine, and orphenadrine.	Benzodiazepines, barbiturates, or propofol for seizures.

		topiramate, valproate, and zonisamide slow the rate of recovery of inactivated sodium channels. Some agents also have α_2 agonist, anticholinergic, and sodium channel-blocking activity (see above and below).		
Barbiturates	Short-acting: butabarbital, pentobarbital, secobarbital Long-acting: phenobarbital, primidone			Hemodialysis and hemoperfusion may be indicated for severe poisoning by some agents (see "Extracorporeal Removal," in text).
Benzodiazepines	Ultrashort-acting: estazolam, midazolam, temazepam, triazolam Short-acting: al-prazolam, flunitrazepam, lorazepam, oxazepam Long-acting: chlordiazepoxide, clonazepam, diazepam, flurazepam Pharmacologically related agents: zaleplon, zolpidem		Tachyarrhythmias can also occur with chloral hydrate. AGMA, hypernatremia, hyperosmolality, hyperammonemia, chemical hepatitis, and hypoglycemia can be seen in valproate poisoning. Carbamazepine and oxcarbazepine may produce hyponatremia from SIADH.	See above and below for treatment of anticholinergic and sodium-channel (membrane) blocking effects.
GABA precursors	γ-Hydroxybutyrate (sodium oxybate; GHB), γ-butyrolactone (GBL), 1,4-butanediol.		Some agents can cause anticholinergic and sodium channel (membrane) blocking effects (see above and below).	
Muscle relaxants	Baclofen, carisoprodol, cyclobenzaprine, etomidate, metaxalone, methocarbamol, orphenadrine, propofol, tizanidine and other imidazoline muscle relaxants.			
Other agents	Chloral hydrate, ethchlorvynol, glutethimide, meprobamate, methaqualone, methyprylon			
Discordant				
Asphyxiants				
Cytochrome oxidase inhibitors	Cyanide, hydrogen sulfide	Inhibition of mitochon-drial cytochrome oxidase, thereby blocking electron transport and oxidative metabolism. Carbon monoxide also binds to hemoglobin and myoglobin and prevents oxygen bind-ing, transport,	Signs and symptoms of hypoxemia with initial physiologic stimulation and subsequent depression (Table e50–2); lactic acidosis; normal PO_2 and calculated oxygen saturation but decreased	High-dose oxygen. IV hydroxocobalamin or IV sodium nitrite and sodium thiosulfate (Lilly cyanide antidote kit) for coma, metabolic acidosis, and cardiovascular dysfunction in cyanide poisoning.

		and tissue uptake (binding to hemoglobin shifts the oxygen dissociation curve to the left).	but decreased oxygen saturation by co-oximetry (that measured by pulse oximetry is falsely elevated but is less than normal and less than the calculated value). Headache and nausea are common with carbon monoxide. Sudden collapse may occur with cyanide and hydrogen sulfide exposure. A bitter almond breath odor may be noted with cyanide ingestion, and hydrogen sulfide smells like rotten eggs.	
Methemoglobin inducers	Aniline derivatives, dapsone, local anesthetics, nitrates, nitrites, nitrogen oxides, nitro- and nitrosohydrocarbons, phenazopyridine, primaquine-type antimalarials, sulfonamides.	Oxidation of hemoglobin iron from ferrous (Fe^{2+}) to ferric (Fe^{3+}) state prevents oxygen binding, transport, and tissue uptake (methemoglobinemia shifts oxygen dissociation curve to the left). Oxidation of hemoglobin protein causes hemoglobin precipitation and hemolytic anemia (manifest as Heinz bodies and "bite cells" on peripheral blood smear).	Signs and symptoms of hypoxemia with initial physiologic stimulation and subsequent depression (Table e50–2), gray-brown cyanosis unresponsive to oxygen at methemoglobin fractions >15–20%, headache, lactic acidosis (at methemoglobin fractions >45%), nor-mal P_{O2} and calculated oxygen saturation but decreased oxygen saturation and increased methemoglobin fraction by co-oximetry (oxygen saturation by pulse oximetry may be falsely increased or decreased but is less than normal and less than the calculated value).	High-dose oxygen. Intravenous methylene blue for methemoglobin fraction >30%, symptomatic hypoxemia, or ischemia (contraindicated in G6PD deficiency). Exchange transfusion and hyperbaric oxygen for severe or refractory cases.
AGMA inducers	Ethylene glycol	Ethylene glycol causes CNS depression and increased serum osmolality. Metabolites (primarily glycolic acid) cause AGMA, CNS depression, and renal failure. Precipitation of oxalic acid metabolite as calcium salt in tissues and urine results in hypocalcemia, tissue edema, and crystalluria.	Initial ethanol-like intoxication, nausea, vomiting, increased osmolar gap, calcium oxylate crystalluria. Delayed AGMA, back pain, renal failure. Coma, seizures, hypotension, ARDS in severe cases.	Sodium bicarbonate to correct acidemia. Thiamine, folinic acid, magnesium, and high-dose pyridoxine to facilitate metabo-lism. Ethanol or fomepizole for AGMA, crystalluria or renal dysfunction, ethylene glycol level >3 mmol/L (20 mg/dL), and for ethanol-like intoxication or increased osmolal gap if level not readily obtainable. Hemodialysis for persistent AGMA,

				lack of clinical improvement, and renal dysfunction. Hemodialysis also useful for enhancing ethylene glycol elimination and shortening duration of treatment when ethylene glycol level >8 mmol/L (50 mg/dL).
AGMA inducers	Iron	Hydration of ferric (Fe^{3+}) ion generates H^+. Non-transferrin-bound iron catalyzes formation of free radicals that cause mitochondrial injury, lipid peroxidation, increased capillary permeability, vasodilation, and organ toxicity.	Initial nausea, vomiting, abdominal pain, diarrhea. AGMA, cardiovascular and CNS depression, hepatitis, coagulopathy, and seizures in severe cases. Radiopaque iron tablets may be seen on abdominal x-ray.	Whole-bowel irrigation for large ingestions. Endoscopy and gastrostomy if clinical toxicity and large number of tablets still visible on x-ray. IV hydration. Sodium bicarbonate for acidemia. IV deferoxamine for systemic toxicity, iron level >90 μmol/L (500 μg/dL).
	Methanol	Methanol causes ethanol-like CNS depression and increased serum osmolality. Formic acid metabolite causes AGMA and retinal toxicity.	Initial ethanol-like intoxication, nausea, vomiting, increased osmolar gap. Delayed AGMA, visual (clouding, spots, blindness) and retinal (edema, hyperemia) abnormalities. Coma, seizures, cardiovascular depression in severe cases. Possible pancreatitis.	Gastric aspiration for recent ingestions. Sodium bicarbonate to correct acidemia. High-dose folinic acid or folate to facilitate metabolism. Ethanol or fomepizole for AGMA, visual symptoms, methanol level >6 mmol/L (20 mg/dL), and for ethanol-like intoxication or increased osmolal gap if level not readily obtainable. Hemodialysis for persistent AGMA, lack of clinical improvement, and renal dysfunction. Hemodialysis also useful for enhancing methanol elimination and shortening duration of treatment when methanol level >15 mmol/L (50 mg/dL).
Salicylate	Increased sensitivity of CNS respiratory center to changes in PO_2 and PCO_2 stimulates respiration. Uncoupling of oxidative phosphorylation, inhibition of Krebs cycle enzymes, and stimulation of carbohydrate and lipid metabolism generate unmeasured endoge-nous anions and cause AGMA.	Initial nausea, vomiting, hyperventilation, alkalemia, alkaluria. Subsequent alkalemia with both respiratory alkalosis and AGMA, and paradoxical aciduria. Late acidemia with CNS and respira-tory depression. Cerebral and pulmonary edema in severe cases. Hypoglycemia, hypocalcemia, hypokalemia, and	IV hydration and supplemental glucose. Sodium bicarbonate to correct acidemia. Urinary alkalinization for systemic toxicity. Hemodialysis for coma, cerebral edema, seizures, pulmonary edema, renal failure, progressive acidbase disturbances or clinical toxicity, salicylate level >7	

		seizures can occur.	mmol/L (100 mg/dL) following acute overdose.	
CNS syndromes				
Extrapyramidal reactions	Antipsychotics (see above), some cyclic antidepressants and antihistamines.	Decreased CNS dopaminergic activity with relative excess of cholinergic activity.	Akathisia, dystonia, parkinsonism	Oral or parenteral anticholinergic agent such as benztropine or diphenhydramine.
Isoniazid		Interference with activation and supply of pyridoxal-5-phosphate, a cofactor for glutamic acid decarboxylase, which converts glutamic acid to GABA, results in decreased levels of this inhibitory CNS neurotransmitter; complexation with and depletion of pyridoxine itself; inhibition of nicotine-adenine dinucleo-tide–dependent lactate and hydroxybutyrate dehydrogenases resulting in substrate accumulation.	Nausea, vomiting, agitation, confusion; coma, respiratory depression, seizures, lactic and ketoacidosis in severe cases.	High-dose intravenous pyridoxine (vitamin B6) for agitation, confusion, coma, and seizures. Diazepam or barbiturates for seizures.
Lithium		Interference with cell membrane ion transport, adenylate cyclase and Na^+, K^+-ATPase activity, and neurotransmitter release.	Nausea, vomiting, diarrhea, ataxia, choreoathetosis, encephalopathy, hyperreflexia, myoclonus, nystagmus, nephrogenic diabetes insipidus, falsely elevated serum chloride with low anion gap, tachycardia. Coma, seizures, arrhythmias, hyperthermia, and prolonged or permanent encephalopathy and movement disorders in severe cases. Delayed onset after acute overdose, particularly with delayedrelease formulations. Toxicity occurs at lower drug levels in chronic poisoning than in acute poisoning.	Whole-bowel irrigation for large ingestions. IV hydration. Hemodialysis for coma, seizures, severe, progressive, or persistent encephalopathy or neuromuscular dysfunction, peak lithium level >4 meq/L following acute overdose.
Serotonin syndrome	Amphetamines, cocaine, dextromethorphan, meperidine, MAO inhibitors, selective serotonin (5-HT) reuptake inhibitors, tricyclic antidepressants, tramadol, triptans, tryptophan.	Promotion of serotonin release, inhibition of serotonin reuptake, or direct stimulation of CNS and peripheral serotonin receptors (primarily 5-HT-1a and 5-HT-2), alone or in combination.	Altered mental status (agitation, confusion, mutism, coma, seizures), neuromuscular hyperactivity (hyperreflexia, myoclonus, rigidity, tremors), and autonomic dysfunction (abdominal pain, diarrhea, diaphoresis, fever,	Serotonin receptor antagonist cyproheptadine, discontinue the offending agent(s).

			flushing, labile hypertension, mydriasis, tearing, salivation, tachycardia). Complications include hyperthermia, lactic acidosis, rhabdomyolysis, and multisystem organ failure.	
Membrane-active agents	Amantadine, antiarrhythmics (class I and III agents; some β blockers), antipsychotics (see above), antihistamines (particularly diphenhydramine), carbamazepine, local anesthetics (including cocaine), opioids (meperidine, propoxyphene), orphenadrine, quinoline antimalarials (chloroquine, hydroxychloroquine, quinine), cyclic antidepressants (see above).	Blockade of fast sodium membrane channels prolongs phase 0 (depolarization) of the cardiac action potential, which prolongs the QRS duration and promotes reentrant (monomorphic) ventricular tachycardia. Class Ia, Ic, and III antiarrhythmics also block potassium channels during phases 2 and 3 (repolarization) of the action potential, prolonging the JT interval and promoting early after-depolarizations and polymorphic (torsades des pointes) ventricular tachycardia. Similar effects on neuronal membrane channels cause CNS dysfunction. Some agents also block α-adrenergic and cholinergic receptors or have opioid effects (see above and Chap. 393).	QRS and JT prolongation (or both) with hypotension, ventricular tachyarrhythmias, CNS depression, seizures. Anticholinergic effects with amantadine, antihistamines, carbamazepine, disopyramide, antipsychotics, and cyclic antidepressants (see above). Opioid effects with meperidine and propoxyphene (see Chap. 393). Cinchonism (hearing loss, tinnitus, nausea, vomiting, vertigo, ataxia, headache, flushing, diaphoresis) and blindness with quinoline antimalarials.	Hypertonic sodium bicarbonate (or hypertonic saline) for cardiac conduction delays and monomorphic ventricular tachycardia. Lidocaine for monomorphic ventricular tachycardia (except when due to class Ib antiarrhythmics). Magnesium, isoproterenol, and overdrive pacing for polymorphic ventricular tachycardia. Physostigmine for anticholinergic effects (see above). Naloxone for opioid effects (see Chap. 393). Extracorporeal removal for some agents (see text).

Abbreviations: AGMA, anion-gap metabolic acidosis; ARDS, adult respiratory distress syndrome; ATPase, adenosine triphosphatase; CNS, central nervous system; GABA, γ-aminobutyric acid; GHB, γ-hydroxybutyrate; 5-HT, 5-hydroxytryptamine (serotonin); MAO, monoamine oxidase; PCO_2, partial pressure of carbon dioxide; PO_2, partial pressure of oxygen; SIADH, syndrome of inappropriate antidiuretic hormone; VX, extremely toxic persistent nerve agent (no common chemical name).

PREVENTION OF REEXPOSURE

Poisoning is a preventable illness. Unfortunately, some adults and children are poison-prone, and recurrences are common. Unintentional polypharmacy poisoning has become especially common among adults with developmental delays, among the growing population of geriatric patients who are prescribed a large number of medications, and among adolescents and young adults experimenting with pharmaceuticals for recreational euphoria. Adults with unintentional exposures should be instructed regarding the safe use of medications and chemicals (according to labeling instructions). Confused patients may need assistance with the administration of their medications. Errors in dosing by health care providers may require educational efforts. Patients should be advised to avoid circumstances that result in chemical exposure or poisoning. Appropriate agencies and health departments should be notified in cases of environmental or workplace exposure. The best approach with young children and patients with intentional overdose (deliberate self-harm or suicide) is to limit their access to poisons. In households where children live or visit, alcoholic beverages, medications, household products (automotive, cleaning, fuel, pet-care, toiletry products), inedible plants, and vitamins should be kept out of reach or in locked or child-proof cabinets. Depressed or psychotic patients should receive psychiatric assessment, disposition, and follow-up. They should be given prescriptions for a limited supply of drugs and with a limited number of refills and be monitored for compliance and response to therapy.

SPECIFIC TOXIC SYNDROMES AND POISONINGS

Table e50–4 summarizes the pathophysiology, clinical features, and treatment of toxidromes and poisonings that are common, produce life-threatening toxicity, or require unique therapeutic interventions. In all cases, treatment should include attending to the general principles discussed above, particularly supportive care. Details regarding specific therapies can be found in the references cited here and at *www.harrisonsonline.com*. Poisonings not covered in this chapter are discussed in the referenced texts.

Alcohol, cocaine, hallucinogen, and opioid poisoning and alcohol and opioid withdrawal are discussed in Chaps. 392, 393, 394, and 395; acetaminophen poisoning is discussed in Chap. 305; the neuroleptic malignant syndrome is discussed in Chap. 372; and heavy metal poisoning is discussed in Chap. e49.

ACKNOWLEDGMENT

The author acknowledges the contributions of Christopher H. Linden and Michael J. Burns to this chapter in previous editions of this text.

FURTHER READINGS

BOND GR: The role of activated charcoal and gastric emptying in gastric decontamination: A state-of-the-art review. Ann Emerg Med 39:273, 2002[PMID: 11867980] [Full Text]

BRENT J: Fomepizole for ethylene glycol and methanol poisoning. N Engl J Med 360:2216, 2009[PMID: 19458366] [Full Text]

BRONSTEIN AC et al: 2008 Annual Report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 27th Annual Report. Clin Toxicol (Phila) 48:979, 2010[PMID: 21192756] [Full Text]

DART RC et al: Expert consensus guidelines for stocking of antidotes in hospitals that provide emergency care. Ann Emerg Med 54:386, 2009[PMID: 19406507] [Full Text]

ERICKSON TB et al: The approach to the patient with an unknown overdose. Emerg Med Clin North Am 25:249, 2007[PMID: 17482020] [Full Text]

FELICE K et al: Intravenous lipid emulsion for local anesthetic toxicity: a review of the literature. J Med Toxicol 4:184, 2008[PMID: 18821493] [Full Text]

FORD MD et al (eds): *Clinical Toxicology*. Philadelphia, Saunders, 2001

GREENBERG MI et al (eds): *Occupational, Industrial, and Environmental Toxicology*. St. Louis, Mosby, 2003

HEARD KJ: Acetylcysteine for acetaminophen poisoning. N Engl J Med 359:285, 2008[PMID: 18635433] [Full Text]

KOSTEN TR, O'CONNOR PG: Management of drug and alcohol withdrawal. N Engl J Med 348:1786, 2003[PMID: 12724485] [Full Text]

NELSON LS et al (eds): Goldfrank's Toxicologic Emergencies, 9th ed. New York, McGraw-Hill, 2010.

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

DISORDERS CAUSED BY VENOMOUS SNAKEBITES AND MARINE ANIMAL EXPOSURES: INTRODUCTION

This chapter outlines general principles for the evaluation and management of victims of envenomation and poisoning by venomous snakes and marine creatures. Because the incidence of serious bites and stings is relatively low in developed nations, there is a paucity of relevant clinical research; as a result, therapeutic decision-making often is based on anecdotal information.

VENOMOUS SNAKEBITE

Epidemiology

Venomous snakes belong to the families Viperidae (subfamily Viperinae: Old World vipers; subfamily Crotalinae: New World and Asian pit vipers), Elapidae (including cobras, kraits, coral snakes, and all Australian venomous snakes), Hydrophiidae (sea snakes), Atractaspididae (burrowing asps), and Colubridae (a large family in which most species are nonvenomous and only a few are dangerously toxic to humans). Bite rates are highest in temperate and tropical regions where populations subsist by manual agriculture. Recent estimates indicate somewhere between 1.2 million and 5.5 million snakebites worldwide each year, with 421,000–1,841,000 envenomations and 20,000–94,000 deaths. Such wide-ranging estimates bear testimony to two facts: collection of data is problematic in the regions most affected by venomous snakes (the "developing world"), and what constitutes a "snakebite" varies among researchers. Some count all snakebites (a figure that may include bites by nonvenomous snakes), whereas others count only apparent envenomations.

Snake Anatomy/Identification

The typical snake-venom apparatus consists of bilateral venom glands situated below and behind the eye and connected by ducts to hollow anterior maxillary teeth. In viperids (vipers and pit vipers), those teeth are long mobile fangs that retract against the roof of the mouth when the animal is at rest. In elapids and sea snakes, the fangs are smaller and are relatively fixed in an erect position. In ~20% of pit viper bites and higher percentages of other snakebites (up to 75% for sea snakes), no venom is released ("dry" bites). Significant envenomation probably occurs in ~50% of all venomous snakebites.

Differentiation of venomous from nonvenomous snake species can be difficult. Viperids are characterized by somewhat triangular heads (a feature shared with many harmless snakes), elliptical pupils (also seen in some nonvenomous snakes, such as boas and pythons), enlarged maxillary fangs, and, in pit vipers, paired heat-sensing pits (foveal organs) on each side of the head. The New World rattlesnakes generally have a series of interlocking keratin plates (the rattle) on the tip of the tail; this rattle is used to dissuade potential threats. Color pattern is notoriously misleading in identifying most venomous snakes. Many harmless snakes have color patterns that closely mimic those of venomous snakes found in the same region.

Venoms and Clinical Manifestations

Snake venoms are complex mixtures of enzymes, low-molecular-weight polypeptides, glycoproteins, metal ions, and other constituents. Among the deleterious components are hemorrhagins that promote vascular leakage and cause both local and systemic bleeding. Proteolytic enzymes cause local tissue necrosis, affect the coagulation pathway at various steps, and impair organ function. Myocardial depressant factors reduce cardiac output, and neurotoxins act either pre- or postsynaptically to inhibit peripheral nerve impulses. Most snake venoms have multisystem effects on their victims.

Envenomations by most viperids and some elapids with necrotizing venoms cause progressive local swelling, pain, ecchymosis **(Fig. 396-1)**, and (over a period of hours or days) hemorrhagic bullae and serum-filled vesicles. In serious bites, tissue loss can be significant **(Fig. 396-2)**. Systemic findings can include changes in taste, mouth numbness, muscle fasciculations, tachycardia or bradycardia, hypotension, pulmonary edema, hemorrhage (from essentially any anatomic site), and renal dysfunction. Envenomations by neurotoxic elapids such as kraits (*Bungarus* spp.), many Australian elapids [e.g., death adders (*Atractaspis* spp.) and tiger snakes (*Notechis* spp.)], some cobras (*Naja* spp.), and some viperids [e.g., the South American rattlesnake (*Crotalus durissus*) and some Indian Russell's vipers (*Daboia russelii*)] cause neurologic dysfunction. Early findings may consist of cranial nerve weakness (e.g., manifested by ptosis) and altered mental status. Severe envenomation may

result in paralysis, including the muscles of respiration, and lead to death from respiratory failure and aspiration. After elapid bites, the time of onset of venom intoxication varies from minutes to hours, depending on the species involved, the anatomic location of the bite, and the amount of venom injected. Sea snake envenomation usually causes local pain (variable), myalgias, rhabdomyolysis, and neurotoxicity; these manifestations occasionally are delayed for hours.

Figure 396-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Northern Pacific rattlesnake (***Crotalus oreganus oreganus***) **envenomations.*Top:*** Moderately severe envenomation. Note edema and early ecchymosis 2 h after a bite to the finger. ***Bottom:*** Severe envenomation. Note extensive ecchymosis 5 days after a bite to the ankle.

Figure 396-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Early stages of severe, full-thickness necrosis 5 days after a **Russell's viper (*Daboia russelii*) bite** in southwestern India.

Treatment: Venomous Snakebite

FIELD MANAGEMENT

The most important aspect of prehospital care of a person bitten by a venomous snake is rapid delivery to a medical facility equipped to provide supportive care (airway, breathing, and circulation) and antivenom administration. Most of the first-aid recommendations made in the past are of little benefit, and some actually worsen outcome. It is reasonable to apply a splint to the bitten extremity to lessen bleeding and discomfort and, if possible, to keep the extremity at approximately heart level. In developing regions, indigenous people should be encouraged to seek care quickly at health care facilities equipped with antivenoms as opposed to consulting traditional healers and thus incurring significant delays in reaching appropriate care.

Incising wounds and/or applying suction to the bite should be avoided, as these measures are ineffective and exacerbate local tissue damage. Similarly ineffective and potentially damaging are the application of poultices, ice, and electric shock. Techniques or devices used in an effort to limit venom spread (e.g., lymphoocclusive bandages or tourniquets) are ineffective and may result in greater local-tissue damage, particularly that due to necrotic venoms. Tourniquet use can result in amputation and loss of function even in the absence of envenomation.

Elapid venoms that are primarily neurotoxic and have no significant effects on local tissue may be localized by pressure-immobilization, a technique in which the entire limb is wrapped immediately with a bandage (e.g., crepe or elastic) and then immobilized. For this technique to be effective, the wrap pressure must be precise (40–70 mmHg in upper extremity application and 55–70 mmHg in lower extremity application), and the victim must be carried from the scene of the bite to avoid muscle-pump action that—regardless of the anatomic site of the bite—will disperse venom if the victim walks. Pressure-immobilization should be used only in cases in which the offending snake is reliably identified and known to be primarily neurotoxic, the rescuer is skilled in pressure-wrap application, the necessary supplies are readily available, and the victim can be carried to medical care—an uncommon combination of conditions, particularly in regions where such bites are most common.

HOSPITAL MANAGEMENT

In the hospital, the victim should be closely monitored (vital signs, cardiac rhythm, oxygen saturation, urine output) while a history is obtained quickly and a rapid, thorough physical examination is performed. For objective evaluation of the progression of local envenomation, the level of swelling in the bitten extremity should be marked and limb circumferences measured every 15 min until swelling has stabilized. During this period of observation/monitoring, the extremity should be positioned at approximately heart level. Measures applied in the field (such as constriction bands or tourniquets) should be removed once IV access has been obtained, with cognizance that the release of such ligatures may result in hypotension or dysrhythmias when stagnant acidotic blood is released to the

central circulation. Large-bore IV access in one or two unaffected extremities should be established. Early hypotension is due to pooling of blood in the pulmonary and splanchnic vascular beds. Later, systemic bleeding, hemolysis, and loss of intravascular volume into soft tissues may play important roles. Fluid resuscitation with isotonic saline (20–40 mL/kg IV) should be initiated if there is any evidence of hemodynamic instability, and a trial of 5% albumin (10–20 mL/kg) may be given when patients fail to respond to saline infusion. Only after aggressive volume resuscitation and antivenom administration (see below) should vasopressors (e.g., dopamine) be added. Invasive hemodynamic monitoring (central venous and/or pulmonary arterial pressures) can be helpful in such cases, although obtaining access is risky if coagulopathy has developed.

Blood should be drawn for typing and cross-matching and for laboratory evaluation as soon as possible. Important studies include a complete blood count to evaluate the degree of hemorrhage or hemolysis and to identify thrombocytopenia, studies of renal and hepatic function, coagulation studies to diagnose consumptive coagulopathy, and testing of urine for blood or myoglobin. In developing regions, the 20-min whole-blood clotting test can be used for reliable diagnosis of coagulopathy. A few milliliters of fresh blood are placed in a new, clean, plain glass receptacle (e.g., a test tube) and left undisturbed for 20 min. The tube then is tipped once to 45° to determine whether a clot has formed. If it has not, coagulopathy is diagnosed. Arterial blood gas studies, electrocardiography, and chest radiography may be helpful in severe envenomations or when there is significant comorbidity. Any arterial puncture in the setting of coagulopathy, however, requires great caution and must be performed at an anatomic site amenable to direct-pressure tamponade. After antivenom therapy (see below), laboratory values should be rechecked every 6 h until clinical stability is achieved. If initial laboratory values are normal, the complete blood count and coagulation studies should be repeated every hour until it is clear that no systemic envenomation has occurred. Victims of neurotoxic envenomation should be watched carefully for evidence of cranial nerve dysfunction (e.g., ptosis) that may precede the onset of difficulty swallowing or respiratory insufficiency and necessitate definitive securing of the airway by endotracheal intubation.

The key to management of venomous snakebite resulting in significant envenomation is the administration of specific antivenom. Antivenoms are produced by the injection of venoms from medically important snakes into animals, generally horses or sheep. Once the stock animals develop antibodies to the venoms, their serum is harvested and the antibodies are isolated for antivenom preparation, which may involve varying degrees of digestion and purification of the IgG molecules. The goal of antivenom administration is to allow antibodies (or antibody fragments) to bind up circulating venom components before they can attach to target tissues and cause deleterious effects. Antivenoms may be monospecific (for a particular snake species) or polyspecific (covering several medically important species in the region) but rarely offer cross-protection against snake species other than those used in their production unless the species are known to have homologous venoms. Thus, antivenom selection must be specific for the offending snake; if the antivenom chosen does not contain antibodies to that snake's venom components, it will be of no value and may lead to unnecessary complications (see below). In the United States, assistance in finding appropriate antivenom can be obtained from regional poison control centers.

Indications for antivenom administration to victims of bites by viperids or cytotoxic elapids include any evidence of systemic envenomation (systemic symptoms or signs, laboratory abnormalities) and significant progressive local findings (e.g., soft tissue swelling crossing a joint or involving more than half the bitten limb). Care must be used in determining the significance of isolated soft tissue swelling after the bite of an unidentified snake because the saliva of some relatively harmless species can cause mild edema at the bite site. In such bites, antivenoms are useless and potentially harmful. The efficacy of antivenoms in preventing tissue damage caused by necrotizing venoms is limited. It may be impossible to prevent necrosis completely, as venom components bind to local tissues very quickly —before antivenom administration can be instituted. Nevertheless, antivenom administration should begin as soon as the need for it is identified to limit further tissue damage and systemic effects. Antivenom administration after the bites of neurotoxic elapids is indicated at the first sign of any evidence of neurotoxicity [cranial nerve dysfunction (e.g., ptosis) or peripheral neuropathy].

Specific comments related to the management of venomous snake bite in the United States and Canada appear in **Table 396-1.** The package insert for the selected antivenom can be consulted regarding species covered, method of administration, starting dose, and need (if any) for redosing. The information in antivenom package inserts, however, is not always accurate and reliable. Whenever possible, it is advisable for treating physicians to seek advice from experts in snakebite management regarding indications for and dosing of antivenom. For viperid bites, antivenom administration generally should be continued as needed until the victim shows definite improvement (e.g., stabilized vital signs, reduced pain, restored coagulation). Neurotoxicity from elapid bites may be harder to reverse with antivenom. Once neurotoxicity is established and endotracheal intubation is required, further doses of antivenom are unlikely to be beneficial. In such cases, the victim must be maintained on mechanical ventilation until recovery, which may take days or weeks.

Table 396-1 Management of Venomous Snakebite in the United States and Canada[a]

Pit viper bites [rattlesnakes (*Crotalus* and *Sistrurus* spp.), cottonmouth water moccasins (*Agkistrodon piscivorus*), and copperheads (*A. contortrix*)]

- Stabilize airway, breathing, and circulation.
- Institute monitoring (cardiac and pulse oximetry).
- Establish two large-bore IV lines with normal saline infusion (administer a bolus of 20–40 mL/kg of body weight if the patient is hypotensive; if hypotension persists, consider albumin).
- Take rapid history and perform rapid physical examination (including vital signs).
- Measure/record circumferences of the bitten extremity every 15 min until swelling has stabilized.
- Identify the offending reptile if possible.
- Send laboratory studies (CBC, metabolic panel, PT/INR/PTT, fibrinogen level, FDP, blood type and screening, urinalysis).
    - If normal, repeat CBC, PT/INR/PTT, fibrinogen level, and FDP every hour until it is clear that no systemic envenomation has occurred.
    - If abnormal, repeat 6 h after antivenom administration (see below).
- Determine severity of envenomation.
    - None ("dry bite"): fang marks only
    - Mild: local findings only (e.g., pain, local ecchymosis, nonprogressive swelling)
    - Moderate: swelling that is clearly progressing, systemic signs or symptoms, and/or laboratory abnormalities
    - Severe: respiratory distress, neurologic dysfunction, and/or cardiovascular instability/shock
- Locate and administer antivenom as indicated: Crotalidae Polyvalent Immune Fab (CroFab) (Ovine) (Protherics US Inc., Brentwood, TN).
    - Starting dose
        - Based on severity of envenomation
            - None or mild: none
            - Moderate: 4–6 vials
            - Severe: 6 vials
        - Mix reconstituted vials in 250 mL of normal saline.
        - No pretesting for potential allergy; no premedication
        - Give IV over 1 h (with physician in close attendance).
    - If acute reaction to antivenom
        - Stop infusion.
        - Treat with standard doses of epinephrine (IM or IV; the latter route only in the setting of severe hypotension), antihistamines (IV), and glucocorticoids (IV).
        - When reaction is controlled, restart antivenom as soon as possible (may further dilute in a larger volume of normal saline).
    - Monitor clinical status over 1 h.
        - Stabilized, improved: hospital admission
        - Progressing or unimproved: repetition of starting dose (this pattern continued until patient's condition is stable or improved)
    - Pain management: acetaminophen and/or narcotics as needed (avoid salicylates and nonsteroidal anti-inflammatory agents)
    - Update tetanus immunization as needed.
    - Prophylactic antibiotics are unnecessary unless prehospital care included incisions or mouth suction.
    - Blood products and coagulation factors are rarely needed; if required, they should be given only after antivenom administration.
    - Admit to hospital. (If no evidence of envenomation, monitor for 8 h before discharge.)
        - Give further CroFab (two vials q6h for three additional doses; close monitoring).
        - Monitor for evidence of rising intracompartmental pressures (see text).
        - Provide wound care (see text).

- Begin physical therapy (see text).
- At discharge, warn patient of possible recurrent coagulopathy and signs/symptoms of delayed serum sickness.

Coral snakebites (*Micrurus* spp. and *Micruroides euryxanthus*)

- Stabilize airway, breathing, and circulation as needed.
- Institute monitoring (cardiac and pulse oximetry).
- Establish one large-bore IV line with normal saline infusion.
- Take rapid history and perform rapid physical examination (including vital signs).
- Identify the offending reptile if possible.
- Laboratory studies are unlikely to be helpful.
- If any evidence of neurologic dysfunction (e.g., any cranial nerve abnormalities such as ptosis):
  - Trial of anticholinesterase inhibitors (see **Table 396-2**)
  - With any evidence of difficulty swallowing or breathing, proceed with endotracheal intubation and ventilatory support (may be required for days or weeks).
- Update tetanus immunization as needed.
- Prophylactic antibiotics are unnecessary unless prehospital care included incisions or mouth suction.
- Admit to hospital (intensive care unit) even if there is no evidence of envenomation (monitor for at least 24 h).
- At the time of this publication, no coral snake antivenom is commercially available for routine use in the United States.

[a] These recommendations are specific to the care of victims of venomous snakebite in the United States and Canada and should not be applied to bites in other regions of the world.

Abbreviations: CBC, complete blood count; FDP, fibrin degradation products; PT/INR/PTT, prothrombin time/international normalized ratio/activated partial thromboplastin time.

Use of any heterologous serum product carries a risk of acute nonallergic (and, less commonly, allergic) anaphylaxis and delayed-type hypersensitivity reactions (serum sickness). Skin testing for potential allergy, although recommended by some antivenom manufacturers, is insensitive and nonspecific and should be omitted. Worldwide, the quality of antivenoms is highly variable. Rates of acute nonallergic anaphylactic reactions to some of these products exceed 50%. For this reason, some authorities have recommended pretreatment with IV antihistamines (e.g., diphenhydramine, 1 mg/kg to a maximum of 100 mg, and cimetidine, 5–10 mg/kg to a maximum of 300 mg) or even a prophylactic subcutaneous or intramuscular dose of epinephrine (0.01 mg/kg, up to 0.3 mg). Further research is necessary, however, to determine whether any pretreatment measures are truly beneficial. Modest expansion of the patient's intravascular volume with crystalloids may blunt acute adverse reactions. Epinephrine should always be immediately available, and the antivenom dose to be administered should be diluted in an appropriate volume of crystalloid according to the package insert. Antivenom should be given only by the IV route, and the infusion should be started slowly, with the physician at the bedside during the initial period to intervene immediately at the first signs of an acute reaction (which may be heralded by a single hive or mild itching or may present as bronchospasm or acute cardiovascular collapse). The rate of infusion can be increased gradually in the absence of a reaction until the full starting dose has been administered (over a total period of ~1 h). Further antivenom may be necessary if the patient's acute clinical condition worsens or fails to stabilize. There is some evidence that smaller-molecular-weight Fab-fragment antivenoms are cleared more quickly from the circulation than are whole-IgG or $F(ab)_2$ antivenoms and thus may require redosing if venom effects that were controlled initially begin to recur. The decision to administer further antivenom to a stabilized patient generally should be based on clinical evidence of persistent circulation of unbound venom components. Antivenom is effective only in reversing active venom toxicity; it is of no benefit in reversing effects that already have been established (e.g., renal failure, established paralysis, necrosis) and that will improve only with time and other therapies.

If the patient develops an acute reaction to antivenom, the infusion should be temporarily stopped and the reaction immediately treated with IM epinephrine and IV antihistamines and glucocorticoids (Chap. 317). Once the reaction is controlled, if the severity of envenomation warrants additional antivenom, the dose should be diluted further in isotonic saline and restarted as soon as possible. Rarely, in recalcitrant cases, a concomitant IV infusion of epinephrine may be required to hold allergic sequelae at bay while further antivenom is administered. The patient must be monitored very closely, preferably in an intensive care setting, during such therapy.

Blood products are rarely necessary in the management of an envenomated patient. The venoms of

many snake species can cause a drop in platelet count or hematocrit and depletion of coagulation factors. Nevertheless, these components usually rebound within hours after administration of adequate antivenom. If the need for blood products is thought to be great (e.g., for a dangerously low platelet count in a hemorrhaging patient), these products should be given only after adequate antivenom administration to avoid adding fuel to an ongoing consumptive coagulopathy.

Rhabdomyolysis and hemolysis should be managed in standard fashion. Victims who develop acute renal failure should be evaluated by a nephrologist and referred for peritoneal dialysis or hemodialysis as needed. Such renal failure, which usually is due to acute tubular necrosis, is frequently reversible. If bilateral cortical necrosis occurs, however, the prognosis for renal recovery is grimmer, and long-term dialysis with possible renal transplantation may be necessary.

Acetylcholinesterase inhibitors (e.g., edrophonium and neostigmine) may promote neurologic improvement in patients bitten by snakes with postsynaptic neurotoxins. Victims with objective evidence of neurologic dysfunction after snakebite should receive a trial of acetylcholinesterase inhibitors, as outlined in **Table 396-2**. If they respond, additional doses of long-acting neostigmine can be continued as needed. Special vigilance is required to prevent aspiration if repetitive dosing of neostigmine is used in an attempt to obviate endotracheal intubation.

Table 396-2 Use of Acetylcholinesterase Inhibitors in Neurotoxic Snake Envenomation

1. Patients with clear, objective evidence of neurotoxicity after snakebite (e.g., ptosis or inability to maintain upward gaze) should receive a trial of edrophonium (if available) or neostigmine.
   a. Pretreat with atropine: 0.6 mg IV (children, 0.02 mg/kg; minimum of 0.1 mg)
   b. Follow with:
   Edrophonium: 10 mg IV (children, 0.25 mg/kg)
   or
   Neostigmine: 1.5–2.0 mg IM (children, 0.025–0.08 mg/kg)
2. If objective improvement is evident at 5 min, continue neostigmine at a dose of 0.5 mg (children, 0.01 mg/kg) IV or SC every 30 min as needed, with continued administration of atropine by continuous infusion of 0.6 mg over 8 h (children, 0.02 mg/kg over 8 h).
3. Maintain vigilance regarding aspiration risk and secure the airway with endotracheal intubation as needed.

Care of the bite wound includes application of a dry, sterile dressing and splinting of the extremity with padding between the digits. Once the administration of antivenom has been initiated, the extremity should be elevated above heart level to relieve edema. Tetanus immunization should be updated as appropriate. Prophylactic antibiotics are generally unnecessary after bites by North American snakes, as the incidence of secondary infection is low. Antibiotics can be considered, however, if misguided first aid efforts have included incisions or mouth suction. In some regions, secondary bacterial infection is more common and the consequences are dire. In these regions, prophylactic antibiotics (e.g., cephalosporins) are used commonly. Pain control should be achieved with acetaminophen or narcotic analgesics. Salicylates and nonsteroidal anti-inflammatory agents should be avoided because of their effects on blood clotting.

Most snake envenomations involve subcutaneous deposition of venom. On occasion, however, venom can be injected more deeply into muscle compartments, particularly if the offending snake was large and the bite occurred to the lower leg or the forearm or hand. If swelling in the bitten extremity raises concern that subfascial muscle edema may be impeding tissue perfusion (muscle-compartment syndrome), intracompartmental pressures (ICPs) should be checked by any minimally invasive technique—e.g., wick catheter or ICP monitor (Stryker Instruments, Kalamazoo, MI). If any ICP is high (>30–40 mmHg), the extremity should be kept elevated while further antivenom is given. A dose of IV mannitol (1 g/kg) can be given in an effort to reduce muscle edema if the patient's hemodynamic status is stable. If, after 1 h of such therapy, the ICP remains elevated, a surgical consultation for possible fasciotomy should be obtained. Although preliminary evidence from studies in animals suggests that fasciotomy actually may worsen myonecrosis, compartmental decompression is still required to preserve nerve function. Fortunately, the incidence of muscle-compartment syndrome is very low after snakebite, with fasciotomies required in <1% of cases. Nevertheless, vigilance is required; if a fasciotomy is deemed necessary, it should be undertaken with the patient's informed consent whenever possible.

Wound care in the days after the bite may require careful aseptic debridement of clearly necrotic tissue once coagulation has been restored. Intact serum-filled vesicles or hemorrhagic blebs should be left undisturbed. If ruptured, they should be debrided with sterile technique. Any debridement of damaged muscle should be conservative, as there is evidence that such muscle may recover to a significant degree.

Physical therapy should be started when pain allows so that the victim can return to a functional state. The incidence of long-term loss of function (e.g., reduced range of motion, impaired sensory function) is unclear but is probably quite high (>30%), particularly after viperid bites.

Any patient with signs of envenomation should be observed in the hospital for at least 24 h. In North America, a patient with an apparently "dry" viperid bite should be watched for at least 8 h before discharge, as significant toxicity occasionally develops after a delay of several hours. The onset of systemic symptoms commonly is delayed for a number of hours after bites by several of the elapids (including coral snakes, *Micrurus* spp.), some non–North American viperids [e.g., the hump-nosed pit viper (*Hypnale hypnale*)], and sea snakes. Patients bitten by these reptiles should be observed in the hospital for at least 24 h. Patients whose condition is not stable should be admitted to an intensive care setting.

At discharge, victims of venomous snakebite should be warned about signs and symptoms of wound infection, antivenom-related serum sickness, and potential long-term sequelae, such as pituitary insufficiency in Russell's viper (*D. russelii*) bites. If coagulopathy developed in the acute stages of envenomation, it can recur during the first 2–3 weeks after the bite. In such cases, victims should be warned to avoid elective surgery or activities posing a high risk of trauma during this period. Outpatient analgesic treatment and physical therapy should be continued.

In the event of serum sickness (fever, chills, urticaria, myalgias, arthralgias, and possibly renal or neurologic dysfunction developing 1–2 weeks after antivenom administration), the victim should be treated with systemic glucocorticoids (e.g., oral prednisone, 1–2 mg/kg daily) until all findings resolve; the dose is then tapered over 1–2 weeks. Oral antihistamines and analgesics provide additional relief of symptoms.

Morbidity and Mortality

The overall mortality rates for venomous snakebite are low in areas with rapid access to medical care and appropriate antivenoms. In the United States, for example, the mortality rate is <1% for victims who receive antivenom. Eastern and western diamondback rattlesnakes (*Crotalus adamanteus* and *C. atrox*, respectively) are responsible for the majority of snakebite deaths in the United States. Snakes responsible for large numbers of deaths in other countries include cobras (*Naja* spp.), carpet and saw-scaled vipers (*Echis* spp.), Russell's vipers (*D. russelii*), large African vipers (*Bitis* spp.), lancehead pit vipers (*Bothrops* spp.), and tropical rattlesnakes (*C. durissus*).

The incidence of morbidity—defined as permanent functional loss in a bitten extremity—is difficult to estimate but is substantial. Morbidity may be due to muscle, nerve, or vascular injury or to scar contracture. Such morbidity can have devastating consequences for victims in the developing world when their ability to work and provide for their families is lost. In the United States, functional loss tends to be more common and severe after rattlesnake bites than after bites by copperheads (*Agkistrodon contortrix*) or water moccasins (*A. piscivorus*).

THE GLOBAL CRISIS

In many developing countries where snakebite is common, scarce access to medical care and antivenom resources contributes to high rates of morbidity and death. In many countries, the available antivenoms are inappropriate and ineffective against the venoms of medically important indigenous snakes. In those regions, further research is necessary to determine the actual impact of venomous snakebite and the specific antivenom needs in terms of both quantity and spectrum of coverage. Without accurate statistics, it is difficult to persuade antivenom manufacturers to begin and sustain production of appropriate antisera in developing nations. There is evidence that antivenoms can be produced in much more cost-effective ways than those currently being used. Just as important as getting the correct antivenoms into underserved regions is the need to educate populations about the prevention of snakebite and train medical care providers in proper management approaches. Local protocols written with significant input from experienced providers in the region of concern should be developed and distributed. Appropriate antivenoms must be available at the likely first point of medical contact for patients (e.g., primary health centers) to minimize the common practice of referring victims to more distant, higher levels of care for the initiation of antivenom therapy. Those who care for snakebite victims in these often remote clinics must have the skills and confidence required to begin antivenom treatment (and to treat possible reactions) as soon as possible when needed.

MARINE ENVENOMATIONS

Much of the management of envenomation by marine creatures is supportive in nature. A specific marine antivenom can be used when appropriate.

Invertebrates

CNIDARIANS

The Golgi apparatus of the cnidoblast cells within cnidarians such as hydroids, fire coral, jellyfish, Portuguese men-of-war, and sea anemones secretes specialized living stinging organelles called *cnidae* (also referred to as *cnidocysts*, a term that encompasses nematocysts, ptychocysts, and

spirocysts). Within each organelle resides a stinging mechanism ("thread tube") and venom. In the stinging process, cnidocysts are released and discharged upon mechanosensory stimulation. The venoms from these organisms are mixtures of proteins, carbohydrates, and other components. Victims usually report immediate prickling or burning, pruritus, paresthesias, and painful throbbing with radiation. The skin becomes reddened, darkened, edematous, and/or blistered. A legion of neurologic, cardiovascular, respiratory, rheumatologic, gastrointestinal, renal, and ocular symptoms have been described. Anaphylaxis is possible. *Irukandji syndrome*, associated with the Australian jellyfish *Carukia barnesi* and other species, is a potentially fatal condition that most commonly is characterized by severe back, chest, and abdominal pain; nausea and vomiting; headache; sweating; and, in the most serious cases, myocardial troponin leak and pulmonary edema. This syndrome is thought to be mediated, at least in part, by the release of endogenous catecholamines.

Rescuers should note that envenomations by different cnidarians (typified by jellyfish) may respond differently to similar therapies; thus, the recommendations in this chapter must be tailored to local species and clinical practices. During stabilization, the skin should be decontaminated immediately with a generous application of vinegar (5% acetic acid), which is the all-purpose agent useful for inactivating the nematocysts in the greatest number of species. Rubbing alcohol (40–70% isopropyl alcohol), baking soda (sodium bicarbonate), papain (unseasoned meat tenderizer), fresh lemon or lime juice, household ammonia, olive oil, or sugar may be effective, depending on the species of stinging creature. For the sting of the venomous box-jellyfish (*Chironex fleckeri*), vinegar should be used. Local application of heat (up to 45°C/113°F), commonly by immersion in hot water, may be as effective. Commercial (chemical) cold packs or real ice packs applied over a thin dry cloth or plastic membrane have been shown to be effective in alleviating mild or moderate *Physalia utriculus* (bluebottle jellyfish) stings but may be less effective than application of heat. Perfume, aftershave lotion, and high-proof ethanol are not efficacious and may be detrimental; formalin, ether, gasoline, and other organic solvents should not be used. Shaving the skin helps remove remaining nematocysts. Freshwater irrigation and rubbing lead to further stinging by adherent nematocysts and should be avoided. After decontamination, topical application of an anesthetic ointment (lidocaine, benzocaine), an antihistamine (diphenhydramine), or a glucocorticoid (hydrocortisone) may be helpful. Persistent severe pain after decontamination may be treated with morphine, meperidine, fentanyl, or another narcotic analgesic. Muscle spasms may respond to diazepam (2–5 mg, titrated upward as necessary) or 10% calcium gluconate (5–10 mL) given IV. An ovine-derived antivenom is available from Commonwealth Serum Laboratories (see the section on antivenom sources, below) for stings from the box-jellyfish found in Australian and Indo-Pacific waters. Ongoing discussion about its efficacy centers on the notion that *C. fleckeri* venom acts more rapidly than the antivenom can bind to the venom. As of this writing, this antivenom has not been used to treat envenomation by the box-shaped jellyfish (possibly of the genus *Chiropsalmus*) that has been found in Florida waters. Treatment for Irukandji syndrome may require the administration of $MgSO_4$ and aggressive antihypertensive treatment.

The pressure-immobilization technique is no longer recommended for venom containment in the setting of a jellyfish sting**.** Safe Sea, a "jellyfish-safe" sunblock (*www.nidaria.com*) applied to the skin before an individual enters the water, inactivates the recognition and discharge mechanisms of nematocysts, has been tested successfully against a number of marine stingers, and may prevent or diminish the effects of coelenterate stings. Whenever possible, a dive skin or wet suit should be worn when entering ocean waters.

SEA SPONGES

Touching a sea sponge may result in dermatitis. The afflicted skin should be gently dried and adhesive tape used to remove embedded spicules. Vinegar should be applied immediately and then for 10–30 min three or four times a day. Rubbing alcohol may be used if vinegar is unavailable. After spicule removal and skin decontamination, steroid or antihistamine cream may be applied to the skin. Severe vesiculation should be treated with a 2-week course of systemic glucocorticoids.

ANNELID WORMS

Annelid worms (bristleworms) possess rows of soft, cactus-like spines capable of inflicting painful stings. Contact results in symptoms similar to those of nematocyst envenomation. Without treatment, pain usually subsides over several hours, but inflammation may persist for up to a week. Victims should resist the urge to scratch, since scratching may fracture retrievable spines. Visible bristles should be removed with forceps and adhesive tape or a commercial facial peel; alternatively, a thin layer of rubber cement can be used to entrap the spines. Use of vinegar, rubbing alcohol, or dilute ammonia or a brief application of unseasoned meat tenderizer (papain) may provide additional relief. Local inflammation should be treated with topical or systemic glucocorticoids.

SEA URCHINS

Sea urchins possess either hollow, venom-filled calcified spines or triple-jawed, globiferous pedicellariae with venom glands. The venom contains toxic components, including steroid glycosides, hemolysins, proteases, serotonin, and cholinergic substances. Contact with either venom apparatus produces immediate and intensely painful stings. The affected part should be immersed immediately in

hot water (see below). Accessible embedded spines should be removed but may break off and remain lodged in the victim. Residual dye from the surface of a spine remaining after the spine's removal may mimic a retained spine but is otherwise of no consequence. Soft tissue radiography or MRI can confirm the presence of retained spines, which may warrant referral for attempted surgical removal if the spines are near vital structures (e.g., joints, neurovascular bundles). Retained spines may cause the formation of granulomas that are amenable to excision or to intralesional injection with triamcinolone hexacetonide (5 mg/mL). Chronic granulomatous arthritis of the proximal interphalangeal joints has been treated with synovectomy and removal of granulation tissue. Erbium-YAG laser ablation has been deployed to destroy multiple sea urchin spines embedded in the foot and identified visually at surface level without causing thermal necrosis of the adjacent tissues. Eosinophilic pneumonia and local and diffuse neuropathies have been observed separately after penetration by multiple spines of the black sea urchin (presumed *Diadema* spp.). The pathophysiology of this phenomenon has not been determined.

OCTOPUSES

Serious envenomations and deaths have followed bites of Australian blue-ringed octopuses (*Octopus maculosus* and *O. lunulata*). Although these animals rarely exceed 20 cm in length, their venom contains a potent neurotoxin (maculotoxin) that inhibits peripheral nerve transmission by blocking sodium conductance. Oral numbness and facial numbness develop within several minutes of a serious envenomation and rapidly progress to total flaccid paralysis, including failure of respiratory muscles. Immediately after envenomation, a circumferential pressure-immobilization dressing 15 cm wide should be applied over a gauze pad (~7 x 7 x 2 cm) that has been placed directly over the sting. The dressing should be applied at venous-lymphatic pressure, with the preservation of distal arterial pulses. The limb should then be splinted. Once the victim has been transported to the nearest medical facility, the bandage can be released. Since there is no antidote, treatment is supportive. If respirations are assisted, the victim may remain awake although completely paralyzed. Even with serious envenomations, significant recovery often takes place within 4–10 h. Sequelae are uncommon unless related to hypoxia.

Vertebrates

STINGRAYS

A stingray injury is both an envenomation and a traumatic wound. Thoracic and cardiac penetration, major vessel laceration, and compartment syndrome have all been observed. The venom, which contains serotonin, 5′-nucleotidase, and phosphodiesterase, causes immediate, intense pain that may last up to 48 h. The wound often becomes ischemic in appearance and heals poorly, with adjacent soft tissue swelling and prolonged disability. Systemic effects include weakness, diaphoresis, nausea, vomiting, diarrhea, dysrhythmias, syncope, hypotension, muscle cramps, fasciculations, paralysis, and (in rare cases) death. Because of the differences in toxins present on the tissues covering the stingers, freshwater stingrays may cause more severe injuries than do marine stingrays.

SCORPIONFISH

The designation *scorpionfish* encompasses members of the family Scorpaenidae and includes not only scorpionfish but also lionfish and stonefish. A complex venom with neuromuscular toxicity is delivered through 12 or 13 dorsal, 2 pelvic, and 3 anal spines. In general, the sting of a stonefish is regarded as the most serious (severe to life-threatening); that of the scorpionfish is of intermediate seriousness; and that of the lionfish is the least serious. Like that of a stingray, the sting of a scorpionfish is immediately and intensely painful. Pain from a stonefish envenomation may last for days. Systemic manifestations of scorpionfish stings are similar to those of stingray envenomations but may be more pronounced, particularly in the case of a stonefish sting. The rare deaths that follow stonefish envenomation usually occur within 6–8 h.

OTHER FISH

Two species of marine catfish—*Plotosus lineatus* (the oriental catfish) and *Galeichthys felis* (the common sea catfish)—as well as several species of freshwater catfish are capable of stinging humans. Venom is delivered through a single dorsal spine and two pectoral spines. Clinically, a catfish sting is comparable to that of a stingray, although marine catfish envenomations are generally more severe than those of their freshwater counterparts. Surgeonfish (doctorfish, tang), weeverfish, ratfish, and horned venomous sharks have also envenomated humans.

Treatment: Marine Vertebrate Stings

The stings of all marine vertebrates are treated in a similar fashion. Except for stonefish and serious scorpionfish envenomations (see below), no antivenom is available. The affected part should be immersed immediately in nonscalding hot water (45°C/113°F) for 30–90 min or until there is significant relief of pain. Recurrent pain may respond to repeated hot-water treatment. Cryotherapy is contraindicated. Opiates will help alleviate the pain, as will local wound infiltration or regional nerve block with 1% lidocaine, 0.5% bupivacaine, and sodium bicarbonate mixed in a 5:5:1 ratio. After soaking and anesthetic administration, the wound must be explored and debrided. Radiography (in

particular, MRI) may be helpful in identification of foreign bodies. After exploration and debridement, the wound should be irrigated vigorously with warm sterile water, saline, or 1% povidone-iodine in solution. Bleeding usually can be controlled by sustained local pressure for 10–15 min. In general, wounds should be left open to heal by secondary intention or treated by delayed primary closure. Tetanus immunization should be updated. Antibiotic treatment should be considered for serious wounds and for envenomation in immunocompromised hosts. The initial antibiotics should cover *Staphylococcus* and *Streptococcus* spp. If the victim is immunocompromised, if a wound is primarily repaired and is more than minor, or if an infection develops, antibiotic coverage should be broadened to include *Vibrio* spp. Infection with *Aeromonas* spp. is of similar concern for wounds associated with natural freshwater.

Approach to the Patient: Marine Envenomations

It is useful to be familiar with the local marine fauna and to recognize patterns of injury.

A large puncture wound or jagged laceration (particularly on the lower extremity) that is more painful than one would expect from the size and configuration of the wound is likely to be a stingray envenomation. Smaller punctures, as described above, represent the activity of a sea urchin or starfish. Stony corals cause rough abrasions and, in rare instances, lacerations or puncture wounds.

Coelenterate (marine invertebrate) stings sometimes create diagnostic skin patterns. A diffuse urticarial rash on exposed skin is often indicative of exposure to fragmented hydroids or larval anemones. A linear, whiplike print pattern appears where a jellyfish tentacle has contacted the skin. In the case of the dreaded box-jellyfish **(Fig. 396-3)**, a cross-hatched appearance, followed by development of dark purple coloration within a few hours of the sting, heralds skin necrosis. A frosted appearance may be created by aluminum salt–based remedies applied to the wound. An encounter with fire coral causes immediate pain and swollen red skin irritation in the pattern of contact, similar to but more severe than the imprint left by exposure to an intact feather hydroid. Seabather's eruption, caused by thimble jellyfishes and larval anemones, may produce a diffuse rash that consists of clusters of erythematous macules or raised papules and is accompanied by intense itching **(Fig. 396-4)**. Toxic sponges create a burning and painful red rash on exposed skin, which may blister and later desquamate. Virtually all marine stingers invoke the sequelae of inflammation, so that local erythema, swelling, and adenopathy are fairly nonspecific.

Figure 396-3

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Skin lesions caused by *Chironex fleckeri* sting.*(Courtesy of Dr. V. Pranava Murthy; with permission.)*

Figure 396-4

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Erythematous, papular rash typical of seabather's eruption caused by thimble jellyfish and larval anemones.

Sources of Antivenoms and Other Assistance

The best way to locate a specific antivenom in the United States is to call a regional poison control center and ask for assistance. Divers Alert Network, a nonprofit organization designed to assist in the care of injured divers, also may help with the treatment of marine injuries. The network can be reached on the Internet at www.diversalertnetwork.org or by telephone 24 h a day at (919) 684-9111. An antivenom for stonefish (and severe scorpionfish) envenomation is made in Australia by the Commonwealth Serum Laboratories (CSL; 45 Poplar Road, Parkville, Victoria, Australia 3052; *www.csl.com.au*; 61-3-9389-1911). Polyvalent sea snake antivenom is also available from CSL. It is no longer recommended that tiger snake antivenom be used if sea snake antivenom is unavailable.

MARINE POISONINGS

Ciguatera

EPIDEMIOLOGY AND PATHOGENESIS

Ciguatera poisoning is the most common nonbacterial food poisoning associated with fish in the United States; most U.S. cases occur in Florida and Hawaii. The poisoning almost exclusively involves tropical and semitropical marine coral reef fish common in the Indian Ocean, the South Pacific, and the Caribbean Sea. Among reported cases, 75% (except in Hawaii) involve the barracuda, snapper, jack, or grouper. The ciguatera syndrome is associated with at least five polyether sodium channel activator toxins that originate in photosynthetic dinoflagellates (such as *Gambierdiscus toxicus*) and accumulate in the food chain. Three major ciguatoxins are found in the flesh and viscera of ciguateric fishes: CTX-1, -2, and -3. TRPV1, a nonselective cation channel expressed in nociceptive neurons, may play a role in the unique neurologic disturbances in ciguatera poisoning. Most, if not all, ciguatoxins are unaffected by freeze-drying, heat, cold, and gastric acid. None of the toxins affects the odor, color, or taste of fish. Cooking methods may alter the relative concentrations of the various toxins.

CLINICAL MANIFESTATIONS

The onset of symptoms may come within 15–30 min of ingestion and typically takes place within 2–6 h. Symptoms increase in severity over the ensuing 4–6 h. Most victims develop symptoms within 12 h of ingestion, and virtually all are afflicted within 24 h. The >150 symptoms reported include abdominal pain, nausea, vomiting, diarrhea, chills, paresthesias, pruritus, tongue and throat numbness or burning, sensation of "carbonation" during swallowing, odontalgia or dental dysesthesias, dysphagia, dysuria, dyspnea, weakness, fatigue, tremor, fasciculations, athetosis, meningismus, aphonia, ataxia, vertigo, pain and weakness in the lower extremities, visual blurring, transient blindness, hyporeflexia, seizures, nasal congestion and dryness, conjunctivitis, maculopapular rash, skin vesiculations, dermatographism, sialorrhea, diaphoresis, headache, arthralgias, myalgias, insomnia, bradycardia, hypotension, central respiratory failure, and coma. Death is rare.

Diarrhea, vomiting, and abdominal pain usually develop 3–6 h after ingestion of a ciguatoxic fish. Symptoms may persist for 48 h and then generally resolve (even without treatment). A pathognomonic symptom is the reversal of hot and cold tactile perception, which develops in some persons after 3–5 days and may last for months. Tachycardia and hypertension have been described, in some cases after potentially severe transient bradycardia and hypotension. More severe reactions tend to occur in persons previously stricken with the disease. Persons who have ingested parrotfish (scaritoxin) may develop classic ciguatera poisoning as well as a "second-phase" syndrome (after 5–10 days' delay) of disequilibrium with locomotor ataxia, dysmetria, and resting or kinetic tremor. This syndrome may persist for 2–6 weeks.

DIAGNOSIS

The differential diagnosis of ciguatera includes paralytic shellfish poisoning, eosinophilic meningitis, type E botulism, organophosphate insecticide poisoning, tetrodotoxin poisoning, and psychogenic hyperventilation. At present, the diagnosis of ciguatera poisoning is made on clinical grounds because no routinely used laboratory test detects ciguatoxin in human blood. High-performance liquid chromatography (HPLC) is available for ciguatoxins and okadaic acid but is of limited clinical value because most health care institutions do not have the equipment needed to perform the test. A ciguatoxin enzyme immunoassay or radioimmunoassay may be used to test small portions of the suspected fish, but even these tests may not detect the very small amount of toxin (0.1 ppb) necessary to render fish flesh toxic.

TREATMENT: CIGUATERA POISONING

Therapy is supportive and is based on symptoms. Nausea and vomiting may be controlled with an antiemetic such as ondansetron (4–8 mg IV). Hypotension may require the administration of IV crystalloid and, in rare cases, a pressor drug. Bradyarrhythmias that lead to cardiac insufficiency and hypotension generally respond well to atropine (0.5 mg IV, up to 2 mg). Cool showers or the administration of hydroxyzine (25 mg PO every 6–8 h) may relieve pruritus. Amitriptyline (25 mg PO twice a day) reportedly alleviates pruritus and dysesthesias. In three cases unresponsive to amitriptyline, tocainide appeared to be efficacious. Nifedipine has been used to treat headache. IV infusion of mannitol may be beneficial in moderate or severe cases, particularly for the relief of distressing neurologic or cardiovascular symptoms, although the efficacy of this therapy has been challenged and has not been definitively proved. The infusion is rendered initially as 1 g/kg per day over 45–60 min during the acute phase (days 1–5). If symptoms improve, a second dose may be given within 3–4 h and repeated on the next day. Care must be taken to avoid dehydration in a treated patient. The mechanism of the benefit against ciguatera intoxication is perhaps hyperosmotic water-drawing action, which reverses ciguatoxin-induced Schwann cell edema. Mannitol may also act in some fashion as a "hydroxyl scavenger" or may competitively inhibit ciguatoxin at the cell membrane.

During recovery from ciguatera poisoning, the victim should exclude the following from the diet: fish (fresh or preserved), fish sauces, shellfish, shellfish sauces, alcoholic beverages, nuts, and nut oils. Consumption of fish in ciguatera-endemic regions should be avoided. All oversized fish of any predacious reef species should be suspected of harboring ciguatoxin. Neither moray eels nor the viscera of tropical marine fish should ever be eaten.

Paralytic Shellfish Poisoning

Paralytic shellfish poisoning is induced by ingestion of any of a variety of feral or aquacultured filter-feeding organisms, including clams, oysters, scallops, mussels, chitons, limpets, starfish, and sand crabs. The origin of their toxicity is the chemical toxin they accumulate and concentrate by feeding on various planktonic dinoflagellates (e.g., *Protogonyaulax*, *Ptychodiscus,* and *Gymnodinium*) and protozoan organisms. The unicellular phytoplanktonic organisms form the foundation of the food chain, and in warm summer months these organisms "bloom" in nutrient-rich coastal temperate and semitropical waters. These planktonic species can release massive amounts of toxic metabolites into the water and cause mortality in bird and marine populations. The paralytic shellfish toxins are water-soluble as well as heat- and acid-stable; they cannot be destroyed by ordinary cooking. The best-characterized and most frequently identified paralytic shellfish toxin is saxitoxin, which takes its name from the Alaska butter clam *Saxidomus giganteus*. A toxin concentration of >75 μg/100 g of foodstuff is considered hazardous to humans. In the 1972 New England "red tide," the concentration of saxitoxin in blue mussels exceeded 9000 μg/100 g of foodstuff. Saxitoxin appears to block sodium conductance, inhibiting neuromuscular transmission at the axonal and muscle membrane levels.

The onset of intraoral and perioral paresthesias (notably of the lips, tongue, and gums) comes within minutes to a few hours after ingestion of contaminated shellfish, and these paresthesias progress rapidly to involve the neck and distal extremities. The tingling or burning sensation later changes to numbness. Other symptoms rapidly develop and include light-headedness, disequilibrium, incoordination, weakness, hyperreflexia, incoherence, dysarthria, sialorrhea, dysphagia, thirst, diarrhea, abdominal pain, nausea, vomiting, nystagmus, dysmetria, headache, diaphoresis, loss of vision, chest pain, and tachycardia. Flaccid paralysis and respiratory insufficiency may follow 2–12 h after ingestion. In the absence of hypoxia, the victim often remains alert but paralyzed.

TREATMENT: PARALYTIC SHELLFISH POISONING

Treatment is supportive and is based on symptoms. If the victim comes to medical attention within the first few hours after poison ingestion, the stomach should be emptied by gastric lavage and then irrigated with 2 L (in 200-mL aliquots) of a solution of 2% sodium bicarbonate; this intervention has not been proved to be of benefit but is based on the notion that gastric acidity may enhance the potency of saxitoxin. Because breathing difficulty can be rapid in onset, induction of emesis is not advised. The administration of activated charcoal (50–100 g) and a cathartic (sorbitol, 20–50 g) makes empirical sense since these shellfish toxins are believed to bind well to charcoal. Some authors advise against administration of magnesium-based solutions (e.g., certain cathartics), cautioning that hypermagnesemia may contribute to suppression of nerve conduction.

The most serious problem is respiratory paralysis. The victim should be closely observed in a hospital for at least 24 h for respiratory distress. With prompt recognition of ventilatory failure, endotracheal intubation and assisted ventilation prevent anoxic myocardial and brain injury.

A direct human serum assay to identify the toxin responsible for paralytic shellfish poisoning is not yet clinically available; the mouse bioassay in widespread use may be replaced by an automated tissue culture bioassay. A polyclonal enzyme-linked immunosorbent assay (ELISA) to measure specific toxins is under development, as is fluorimetric HPLC.

Domoic Acid Intoxication (Amnestic Shellfish Poisoning)

In late 1987 in eastern Canada, an outbreak of gastrointestinal and neurologic symptoms (amnestic shellfish poisoning) was documented in persons who had consumed mussels found to be contaminated with domoic acid. In this outbreak, the source of the toxin was *Nitzschia pungens*, a diatom ingested by the mussels. In 1991, an epidemic of domoic acid poisoning in the state of Washington was attributed to the consumption of razor clams. A heat-stable neuroexcitatory amino acid whose biochemical analogues are kainic acid and glutamic acid, domoic acid binds to the kainate type of glutamate receptor with three times the affinity of kainic acid and is 20 times as powerful a toxin. Shellfish can be tested for domoic acid by mouse bioassay and HPLC. The regulatory limit for domoic acid in shellfish is 20 parts per million.

The abnormalities noted within 24 h of ingesting contaminated mussels (*Mytilus edulis*) include arousal, confusion, disorientation, and memory loss. The median time of onset is 5.5 h. Other prominent symptoms include severe headache, nausea, vomiting, diarrhea, abdominal cramps, hiccups, arrhythmias, hypotension, seizures, ophthalmoplegia, pupillary dilation, piloerection, hemiparesis, mutism, grimacing, agitation, emotional lability, coma, copious bronchial secretions, and pulmonary edema. Histologic study of brain tissue taken at autopsy has shown neuronal necrosis or cell loss and astrocytosis, most prominently in the hippocampus and the amygdaloid nucleus—findings similar to those in animals poisoned with kainic acid. Several months after the primary intoxication, victims still display chronic residual memory deficits and motor neuronopathy or axonopathy. Nonneurologic illness does not persist.

TREATMENT: DOMOIC ACID INTOXICATION

Therapy is supportive and is based on symptoms. Since kainic acid neuropathology seems to be nearly entirely seizure-mediated, the emphasis should be on anticonvulsive therapy, for which diazepam appears to be as effective as any other drug.

Scombroid Poisoning

Scombroid (mackerel-like) fish include the albacore, bluefin, and yellowfin tuna; mackerel; saury; needlefish; wahoo; skipjack; and bonito. Nonscombroid fish that produce scombroid poisoning include the dolphinfish (Hawaiian mahimahi, *Coryphaena hippurus*), kahawai, sardine, black marlin, pilchard, anchovy, herring, amberjack, and Australian ocean salmon. In the northeastern and mid-Atlantic United States, bluefish (*Pomatomus saltatrix*) has been linked to scombroid poisoning. Because greater numbers of nonscombroid fish are being recognized as scombrotoxic, the syndrome may more appropriately be called pseudoallergic fish poisoning.

Under conditions of inadequate preservation or refrigeration, the musculature of these dark- or red-fleshed fish undergoes bacterial decomposition, which includes decarboxylation of the amino acid l-histidine to histamine, histamine phosphate, and histamine hydrochloride. Histamine levels of 20–50 mg/100 g are noted in toxic fish, with levels >400 mg/100 g on occasion. However, it is possible that some other compound may be responsible for this intoxication, since large doses of oral histamine do not reproduce the affliction. Whatever toxin or toxins are involved are heat-stable and are not destroyed by domestic or commercial cooking. Affected fish typically have a sharply metallic or peppery taste; however, they may be normal in appearance, color, and flavor. Not all persons who eat a contaminated fish necessarily become ill, perhaps because of uneven distribution of decay within the fish.

Symptoms develop within 15–90 min of ingestion and include flushing (sharply demarcated; exacerbated by ultraviolet exposure; particularly pronounced on the face, neck, and upper trunk), a sensation of warmth without elevated core temperature, conjunctival hyperemia, pruritus, urticaria,

angioneurotic edema, bronchospasm, nausea, vomiting, diarrhea, epigastric pain, abdominal cramps, dysphagia, headache, thirst, pharyngitis, gingival burning, palpitations, tachycardia, dizziness, and hypotension. Without treatment, the symptoms generally resolve within 8–12 h. Because of blockade of gastrointestinal tract histaminase, the reaction may be more severe in a person who is concurrently ingesting isoniazid.

TREATMENT: SCOMBROID POISONING

Therapy is directed at reversing the histamine effect with antihistamines, either H-1 or H-2. If bronchospasm is severe, an inhaled bronchodilator—or in rare, extremely severe circumstances, injected epinephrine—may be used. Glucocorticoids are of no proven benefit. Protracted nausea and vomiting, which may empty the stomach of toxin, may be controlled with a specific antiemetic, such as prochlorperazine. The persistent headache of scombroid poisoning may respond to cimetidine or a similar antihistamine if standard analgesics are not effective.

FURTHER READINGS

Auerbach PS (ed): *Wilderness Medicine*, 5th ed, St. Louis, Mosby, 2007

Bush SP: Snakebite suction devices don't remove venom: They just suck. Ann Emerg Med 43:187, 2004[PMID: 14747806] [Full Text]

Dart RC et al: A randomized multicenter trial of Crotalinae polyvalent immune Fab (ovine) antivenom for the treatment for crotaline snakebite in the United States. Arch Intern Med 161:2030, 2001[PMID: 11525706] [Full Text]

Gold BS et al: Bites of venomous snakes. N Engl J Med 347:347, 2002[PMID: 12151473] [Full Text]

Lavonas EJ et al: Crotaline Fab antivenom appears to be effective in cases of severe North American pit viper envenomation: An integrative review. BMC Emerg Med 9:13, 2009[PMID: 19545426] [Full Text]

Mebs D: *Venomous and Poisonous Animals.* Boca Raton, FL, CRC Press, 2002

Meier J, White J (eds): *Handbook of Clinical Toxicology of Animal Venoms and Poisons* . Boca Raton, FL, CRC Press, 1996, pp 89–176

Pearn J: Neurology of ciguatera. J Neurol Neurosurg Psychiatry 70:4, 2001[PMID: 11118239] [Full Text]

Sharma H et al: An effective alternative management for multiple sea urchin spine injury: Erbium-YAG laser ablation. Foot Ankle Surg 12:51, 2006

Simpson ID, Jacobsen IM: Antisnake venom production crisis—who told us it was uneconomic and unsustainable? Wilderness Environ Med 20:144, 2009[PMID: 19594208] [Full Text]

———, Norris RL: The global snakebite crisis—a public health issue misunderstood, not neglected. Wilderness Environ Med 20:43, 2009

Winter KL et al: An examination of the cardiovascular effects of an "Irukandji" jellyfish, *Alatina nr mordens*. Toxicol Lett 179:18, 2008

———: An in vivo comparison of the efficacy of CSL box jellyfish antivenom with antibodies raised against nematocyst-derived *Chironex fleckeri* venom. Toxicol Lett 187:94, 2009

BIBLIOGRAPHY

Bogdan GM et al: Recurrent coagulopathy after antivenom treatment of crotalid snakebite. South Med J 93:562, 2000[PMID: 10881769] [Full Text]

Chippaux J-P: Snake-bites: Appraisal of the global situation. Bull World Health Organ 76:515, 1998[PMID: 9868843] [Full Text]

Cuypers E et al: TRPV1 as a determinant in ciguatera and neurotoxic shellfish poisoning. Biochem Biophys Res Commun 361:214, 2007[PMID: 17659256] [Full Text]

Garcia C et al: Paralytic shellfish poisoning: Post-mortem analysis of tissue and body fluid samples from human victims in the Patagonia fjords. Toxicon 43:149, 2004[PMID: 15019474] [Full Text]

Howarth DM et al: Lymphatic flow rates and first-aid in simulated peripheral snake or spider envenomation. Med J Aust 161:695, 1994[PMID: 7830641] [Full Text]

Kucewicz A, Miller M: Eosinophilic pneumonia associated with foot injury from a sea urchin. Am J Emerg Med 25:862.e5, 2007

Little M: Is there a role for the use of pressure immobilization bandages in the treatment of jellyfish envenomation in Australia? Emerg Med (Fremantle) 14:71, 2002

Louzao MC et al: A fluorimetric microplate assay for detection and quantitation of toxins causing paralytic shellfish poisoning. Chem Res Toxicol 16:433, 2003[PMID: 12703959] [Full Text]

Meyer WP et al: First clinical experiences with a new ovine Fab *Echis ocellatus* snake bite antivenom in Nigeria: Randomized comparative trial with Institute Pasteur Serum (Ipser) Africa antivenom. Am J Trop Med Hyg 56:291, 1997[PMID: 9129531] [Full Text]

Mines D et al: Poisonings: Food, fish, and shellfish. Emerg Med Clin North Am 15:157, 1997[PMID: 9056574] [Full Text]

Norris RL et al: Physicians and laypeople are unable to apply pressure immobilization properly in a simulated snakebite scenario. Wilderness Environ Med 16:16, 2005[PMID: 15813142] [Full Text]

Russell FE: *Snake Venom Poisoning*. New York, Scholium International, 1983

Teitelbaum JS et al: Neurologic sequelae of domoic acid intoxication due to the ingestion of contaminated mussels. N Engl J Med 322:1781, 1990[PMID: 1971710] [Full Text]

Wada T et al: Sea urchin spine arthritis of the hand. J Hand Surg [Am] 33A:398, 2008

Warrell DA: The clinical management of snake bites in the Southeast Asian region. Southeast Asian J Trop Med Public Health 30:1, 1999

Copyright © The McGraw-Hill Companies. All rights reserved.

The McGraw-Hill Companies

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ECTOPARASITE INFESTATIONS AND ARTHROPOD BITES AND STINGS: INTRODUCTION

Ectoparasites are arthropods or helminths that infest the skin or hair of other animals, from which they derive sustenance and shelter. They may penetrate beneath the surface of the host or attach superficially by their mouthparts and specialized claws. These organisms damage their hosts by inflicting direct injury, eliciting a hypersensitivity reaction, inoculating toxins or pathogens, and inciting fear. The main medically important ectoparasites are arachnids (including mites and ticks), insects (including lice, fleas, bedbugs, and flies), pentastomes (tongue worms), and leeches. Arthropods also may harm humans through brief encounters during which they take a blood meal or attempt to defend themselves by biting, stinging, or exuding venoms. Various arachnids (spiders, scorpions), insects (bees, hornets, wasps, ants, flies, bugs, caterpillars, and beetles), millipedes, and centipedes produce ill effects in these manners, as do certain ectoparasites of animals, including ticks, biting mites, and fleas. In the United States, more people die each year from arthropod stings than from the bites of poisonous snakes. Lesions resulting from the bites and stings of arthropods are so diverse and variable that it is rarely possible to identify precisely what kind of insect or tick is involved without a bona fide specimen and entomologic expertise.

Scabies

The human itch mite, *Sarcoptes scabiei*, is a common cause of itching dermatosis, infesting ~300 million persons worldwide. Gravid female mites that measure ~0.3 mm in length burrow superficially beneath the stratum corneum, depositing three or fewer eggs per day. Nymphs mature in ~2 weeks and then emerge as adults to the surface of the skin, where they mate and (re)invade the skin of the same or another host. Transfer of newly fertilized female mites from person to person occurs mainly by intimate contact and is facilitated by crowding, poor hygiene, and multiple sexual partners. Generally, these mites die within a day or so in the absence of host contact. Transmission via sharing of contaminated bedding or clothing therefore occurs infrequently. In the United States, scabies may account for up to 5% of visits to dermatologists. Outbreaks occur in nursing homes, mental institutions, and hospitals.

The itching and rash associated with scabies derive from a sensitization reaction directed against the excreta that the mite deposits in its burrow. An initial infestation remains asymptomatic for up to 6 weeks, and a reinfestation produces a hypersensitivity reaction without delay. Burrows become surrounded by infiltrates of eosinophils, lymphocytes, and histiocytes, and a generalized hypersensitivity rash later develops in remote sites. Immunity and associated scratching limit most infestations to <15 mites per person. Hyperinfestation with thousands of mites, a condition known as *crusted scabies* or *Norwegian scabies*, may result from glucocorticoid use, immunodeficiency, and neurologic and psychiatric illnesses that limit itching and scratching.

Intense itching worsens at night and after a hot shower. Typical burrows may be difficult to find because they are few in number and may be obscured by excoriations. Burrows appear as dark wavy lines in the epidermis and measure up to 15 mm. Lesions occur most frequently on the volar wrists, between the fingers, on the elbows, and on the penis. Small papules and vesicles, often accompanied by eczematous plaques, pustules, or nodules, are distributed symmetrically in those sites and in skinfolds under the breasts and around the navel, axillae, belt line, buttocks, upper thighs, and scrotum. Except in infants, the face, scalp, neck, palms, and soles are spared. Crusted scabies resembles psoriasis in its typical widespread erythema, thick keratotic crusts, scaling, and dystrophic nails. Characteristic burrows are not seen in crusted scabies, and patients usually do not itch, although their infestations are highly contagious and have been responsible for outbreaks of classic scabies in hospitals.

Scabies should be considered in patients with pruritus and symmetric polymorphic skin lesions in characteristic locations, particularly if there is a history of household contact with an affected person. Burrows should be sought and unroofed with a sterile needle or scalpel blade, and the scrapings should be examined microscopically for the mite, its eggs, and its fecal pellets. Biopsies (including superficial cyanoacrylate biopsy), scrapings, and dermascopic imaging of papulovesicular lesions as well as microscopic inspection of clear adhesive tape lifted from lesions also may be diagnostic. In the absence of identifiable mites or mite products, the diagnosis is based on clinical presentation and history. Diverse kinds of dermatitis due to other causes frequently are misdiagnosed as scabies.

TREATMENT: SCABIES

Permethrin cream (5%) is less toxic than 1% lindane preparations and is effective against lindane-tolerant infestations. Scabicides are applied thinly but thoroughly behind the ears and from the neck down after bathing and are removed 8 h later with soap and water. Successful treatment of crusted scabies requires preapplication of a keratolytic agent such as 6% salicylic acid and then of scabicides to the scalp, face, and ears. Repeated treatments or the sequential use of several agents may be necessary. Ivermectin has not been approved by the U.S. Food and Drug Administration (FDA) for use against any form of scabies, but a single oral dose (200 μg/kg) effectively treats scabies in otherwise healthy persons; patients with crusted scabies may require two doses separated by an interval of 1–2 weeks.

Although effectively treated scabies infestations become noninfectious within a day, itching and rash due to hypersensitivity to the dead mites and their excreted and secreted products frequently persist for weeks or months. Unnecessary re-treatment with topical agents may provoke contact dermatitis. Antihistamines, salicylates, and calamine lotion relieve itching during treatment, and topical glucocorticoids are useful for pruritus that lingers after effective treatment. To prevent reinfestations, bedding and clothing should be washed and/or dried on high heat or heat-pressed, and close contacts, even if asymptomatic, should be treated simultaneously.

Chiggers and Other Biting Mites

Chiggers are the larvae of trombiculid (harvest) mites that normally feed on mice in grassy or brush-covered sites in the tropics and subtropics and less frequently in temperate areas during warm months. They wait for hosts on low vegetation and attach themselves to passing animals or humans. The larva pierces the skin of its host and produces a secreted tubelike structure (*stylostome*) in the dermis through which it imbibes tissue fluids. The stylostome is highly antigenic and causes an exceptionally pruritic papular, papulovesicular, or papulourticarial lesion (≤2 cm in diameter) that develops within hours of attachment in persons previously sensitized to mite antigen. Feeding mites appear as tiny red vesicles adjacent to hair follicles. Scratching invariably destroys the body of a mite. Generally, lesions vesiculate and develop a hemorrhagic base. Itching and burning persist for weeks. The rash is common on the ankles and areas where clothing obstructs the further wanderings of the mites. Repellents are useful for preventing chigger bites.

Diverse mites associated with birds and rodents can be particularly bothersome when they invade homes and bite human inhabitants. In North America, the northern fowl mite, the chicken mite, the tropical rat mite, and the house mouse mite normally feed on poultry, various songbirds, and small mammals and are abundant in and near their hosts' nests. These mites invade homes after their natural hosts die or leave their nests. Although the mites often are not seen because of their small size, their bites can be painful and pruritic. Painful bitelike sensations associated only with certain rooms of a home may be due to biting mites. Rodent- and bird-associated mites are best eliminated by excluding hosts, removing nests, and cleaning and treating the nesting area with appropriate acaricides. *Pyemotes* and other mites that infest grain, straw, cheese, hay, or other products occasionally produce similar episodes of rash and discomfort.

Diagnosis of mite-induced dermatitides (including those caused by chiggers) relies on confirmation of the mite's identity or elicitation of a history of exposure to the mite's source. Antihistamines or topical steroids effectively reduce mite-induced pruritus.

Tick Bites and Tick Paralysis

Ticks attach and feed painlessly; blood is their only food. Their secretions produce local reactions, a febrile illness, or paralysis and transmit diverse pathogens. Generally, soft ticks attach for <1 h and may produce erythematous macular lesions ≤3 cm in diameter. Some species in Africa, the western United States, and Mexico produce painful hemorrhagic lesions. In contrast, hard ticks attach and feed for several days or sometimes for >1 week. At the site of hard-tick bites, small areas of induration with surrounding erythema and occasionally necrotic ulcers develop. Chronic nodules (tick granulomas) reach several centimeters in diameter and may require surgical excision. Tick-induced fever, associated with headache, nausea, and malaise, usually resolves ≤36 h after the tick is removed.

Tick paralysis, an acute ascending flaccid paralysis, is believed to be caused by one or more toxins in tick saliva that produce neuromuscular block, decreased nerve conduction, and sometimes hypertension. Throughout the world, this rare complication has followed the bites of >60 kinds of ticks; in the United States, dog and wood ticks are most commonly involved. Weakness begins in the lower extremities ≤6 days after the tick's attachment and ascends symmetrically over several days to result in complete paralysis of the extremities and cranial nerves. Deep tendon reflexes are diminished or lacking altogether, but sensory examination and findings on lumbar puncture are typically normal. Removal of the tick generally results in improvement within a few hours and complete recovery after several days, although the patient's condition may continue to deteriorate for up to 1 day. Failure to remove the tick may lead to dysarthria, dysphagia, and ultimately death from aspiration or respiratory paralysis. Diagnosis depends on finding the tick, which is often hidden beneath hair. An antiserum to the saliva of *Ixodes holocyclus*, the usual cause of tick paralysis in Australia, effectively reverses paralysis caused by these ticks.

Ticks should be removed by firm traction with fine-tipped forceps placed near the point of attachment. Use of occlusive dressings, heat, or other substances merely delays tick removal. The site of attachment should be disinfected. Tick mouthparts remaining in the skin generally are shed within days without excision. Removal of ticks during the first 36 h of attachment nearly always prevents transmission of the agents of Lyme disease, babesiosis, anaplasmosis, and ehrlichiosis. Gentle handling (to avoid rupture of ticks) and use of gloves may avert accidental contamination with tick fluids containing pathogens. Rather than awaiting results of tick testing or seroconversion to Lyme disease, adult patients with bites thought to be associated with deer ticks **(Fig. 397-1)** in Lyme disease–endemic areas from Maryland to Maine and in Wisconsin and Minnesota may be treated presumptively with a single oral dose of doxycycline (200 mg) within 72 h of tick removal.

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Figure 397-1 Deer ticks (*Ixodes scapularis*, black-legged ticks) on a U.S. penny: larva (*below ear*), nymph (*right*), adult male (*above*), and adult female (*left*).

Louse Infestation (Pediculiasis and Pthiriasis)

Nymphs and adults of all three kinds of human lice feed at least once a day, ingesting human blood exclusively. Head lice (*Pediculus capitis*) infest mainly the hair of the scalp, body lice (*Pediculus humanus*) the clothing, and crab or pubic lice (*Pthirus pubis*) mainly the hair of the pubis. The saliva of lice produces an irritating maculopapular or urticarial rash in certain sensitized persons. Female head and pubic lice cement their eggs firmly to hair, and female body lice cement their eggs to clothing. A nymph hatches after ~10 days of development. The empty egg (nit) may remain affixed for months thereafter.

In North America, head lice infest ~1% of elementary school–age children. Head lice are transmitted mainly by direct head-to-head contact rather than by fomites (shared headgear, grooming implements, bedding). Infestations by head lice tend to be asymptomatic. Pruritus, due mainly to hypersensitivity to the louse's saliva, generally is transient and mild. Head lice removed from a person succumb to desiccation and starvation within ~1 day. Head lice are unimportant as vectors of pathogenic agents.

Body lice remain on clothing except when feeding and generally succumb in ≤2 days if separated from their host. These lice mainly infest disaster victims or indigent people who are in close contact with other infested individuals. Body lice are acquired by direct contact or by sharing of clothing and bedding. These lice are vectors for the agents of louse-borne typhus (Chap. 174), louse-borne relapsing fever (Chap. 172), and trench fever (Chap. 160). Pruritic lesions from their bites are particularly common around the neckline. Chronic infestations result in a postinflammatory hyperpigmentation and thickening of skin known as *vagabonds' disease*.

The crab or pubic louse is transmitted mainly by sexual contact. These lice occur mainly on pubic hair and less frequently on hair of the axillae and the face, including the eyelashes. Children and adults may acquire pubic lice by sexual or close nonsexual contact. Intensely pruritic lesions and blue macules ~3 mm in diameter (*maculae ceruleae*) develop at the site of bites. Blepharitis commonly accompanies infestations of the eyelashes.

Pediculiasis may be suspected upon the detection of nits on hairs or in clothing, but confirmation should be based on discovery of a live louse.

TREATMENT: LOUSE INFESTATION

Generally, treatment is warranted only if live lice are discovered. The presence of nits alone is evidence of former—not current—infestation. Mechanical removal of lice and their eggs by means of a fine-toothed louse or nit comb **(Fig. 397-2)** often fails to eliminate infestations. Treatment of newly

identified active infestations generally relies on a 10-min application of ~1% permethrin or pyrethrins, with a second application 10 days later. Lice persisting after this treatment may be resistant to pyrethroids (see below). Chronic infestations may be treated for ≤12 h with 0.5% malathion. Lindane is applied for just 4 min but seems less effective and may pose a greater risk of adverse reactions, particularly when misused. Resistance of head lice to permethrin, malathion, and lindane has been reported. Newer pediculicides contain benzyl alcohol, dimethicone, or spinosad. Although children infested by head lice are frequently isolated or excluded from school, this practice increasingly is seen as unjustified.

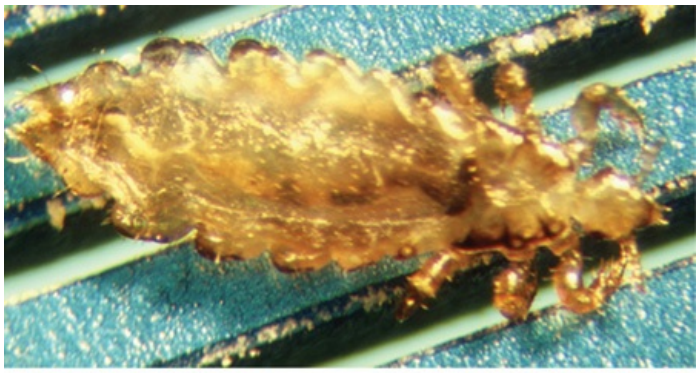

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Figure 397-2 Adult female human head louse (*Pediculus capitis*) on a nit (louse-egg) comb.

Body lice usually are eliminated by bathing and by changing to laundered clothes. Application of topical pediculicides from head to foot may be necessary for hirsute patients. Clothes and bedding are effectively deloused by heating in a clothes dryer at ≥55°C (131°F) for 30 min or by heat-pressing. Emergency mass delousing of persons and clothing may be warranted during periods of civil strife and after natural disasters to reduce the risk of pathogen transmission by body lice.

Pubic louse infestations are treated with topical pediculicides except for eyelid infestations (*pthiriasis palpebrum*), which generally respond to a coating of petrolatum applied for 3–4 days.

Myiasis (Fly Infestation)

Myiasis refers to infestations by diverse kinds of fly larvae (maggots) that invade living or necrotic tissue or body cavities and produce different clinical syndromes, depending on the species of fly.

In forested parts of Central and South America, larvae of the human botfly *Dermatobia hominis* produce boil-like subcutaneous nodules ≤3 cm in diameter. The adult female captures a mosquito or another bloodsucking insect and deposits her eggs on its abdomen. When the carrier insect attacks a human or bovine host several days later, the warmth and moisture of the host's surface stimulate the larvae to hatch. The larvae promptly penetrate intact skin. After 6–12 weeks of development, mature larvae emerge from the skin and drop to the ground. The African tumbu fly *Cordylobia anthropophaga* deposits its eggs on sand or drying laundry contaminated with urine or sweat. Larvae hatch on contact with the body, penetrate the skin, and produce boils from which they emerge ~9 days later. Furuncular myiasis is suggested by uncomfortable lesions with a central breathing pore that emits bubbles when submerged in water. A sensation of movement under the patient's skin may lead to severe emotional distress. Botfly larvae may be induced to emerge if the air pore is coated with petrolatum or another occlusive substance. Removal may be facilitated by injection of a local anesthetic into the surrounding tissue, but surgical excision is often necessary because upward-pointing spines of some species hold the larva firmly in place.

Larvae of the horse botfly *Gasterophilus intestinalis* do not mature after penetrating human skin but migrate for weeks in the epidermis. The resulting pruritic and serpiginous eruption resembles cutaneous larva migrans caused by hookworms (Chap. 216). Horseback riders become infested when eggs deposited on the flank of the horse hatch against their bare legs. The larvae of the cattle botfly invade more deeply and produce boil-like swellings, and larvae of rabbit and rodent *Cuterebra* occasionally cause dermal or tracheopulmonary myiasis.

Certain flies are attracted to blood and pus, and their newly hatched larvae enter wounds or diseased skin. Larvae of the green bottle fly usually remain superficial and confined to necrotic tissue, but specially prepared "surgical maggots" sometimes are used intentionally for wound debridement. Larvae of screwworm flies and the flesh fly invade viable tissue more deeply and produce large suppurating lesions. Larvae that infest wounds also may infest body cavities such as the mouth, nose, ears, sinuses, anus, vagina, and lower urinary tract, particularly in unconscious or otherwise debilitated patients. The consequences range from harmless colonization to destruction of the nose, meningitis, and deafness. Treatment involves removal of maggots and debridement of tissue.

The maggots responsible for furuncular and wound myiasis also may cause ophthalmomyiasis. Sequelae include nodules in the eyelid, retinal detachment, and destruction of the globe. Most instances in which maggots are found in human feces result from larviposition by flesh flies on recently passed stools.

Pentastomiasis

Pentastomids (tongue worms) inhabit the respiratory passages of reptiles and carnivorous mammals. Human infestation by *Linguatula serrata* is common in the Middle East and results from the ingestion of encysted larval stages in raw liver or lymph nodes of sheep and goats—the intermediate hosts. Larvae migrate to the nasopharynx and produce an acute self-limiting syndrome known as *Halzoun* or *Marrara*, which is characterized by pain and itching of the throat and ears, coughing, hoarseness, dysphagia, and dyspnea. Severe edema may cause obstruction that necessitates tracheostomy; ocular invasion has been described. Diagnostic larvae measuring ≤10 mm in length appear in copious nasal discharge or vomitus. Individuals become infected with *Armillifer armillatus* by ingesting eggs in contaminated food or drink or after handling the definitive host, the African python. Larvae encyst in various organs but rarely cause symptoms. Cysts occasionally require surgical removal as they enlarge during molting, but they usually are encountered as an incidental finding at autopsy. Parasite-induced lesions may be misinterpreted as a malignancy, with the correct diagnosis confirmed by histopathologic findings. Cutaneous larva migrans syndromes due to other pentastomes have been reported from Southeast Asia and Central America.

Leech Infestations

Medically important leeches are annelid worms that attach to their hosts with chitinous cutting jaws and draw blood through muscular suckers. The medicinal leech *Hirudo medicinalis* is still used occasionally to reduce venous congestion in surgical flaps or replanted body parts. This practice has been complicated by intractable bleeding, wound infections, myonecrosis, and sepsis due to *Aeromonas hydrophila*, which colonizes the gullets of commercially available leeches.

Ubiquitous aquatic leeches that parasitize fish, frogs, and turtles readily attach to the skin of humans and avidly suck blood. More notorious are the land leeches that live among moist vegetation of tropical rain forests. Attachment is usually painless. Hirudinin, a powerful anticoagulant secreted by the leech, causes continued bleeding after the leech has detached. Healing of the wound is slow, and bacterial infections are not uncommon. Several kinds of aquatic leeches in Africa, Asia, and southern Europe can enter through the mouth, nose, and genitourinary tract and attach to mucosal surfaces at sites as deep as the esophagus and trachea. Externally attached leeches generally drop off after they have engorged, but removal is hastened by gentle scraping aside of the anterior and posterior suckers and traction or by application of alcohol, salt, vinegar, insect repellent, or a flame or heated instrument to the leech. Internally attached leeches may detach on exposure to gargled saline or may be removed by forceps.

Spider Bites

Of the >30,000 recognized species of spiders, only ~100 defend themselves aggressively and have fangs sufficiently long to penetrate human skin. The venom that spiders use to immobilize and digest their prey can cause necrosis of skin and systemic toxicity. Whereas the bites of most spiders are painful but not harmful, envenomations by recluse or fiddle spiders (*Loxosceles* species) and widow spiders (*Latrodectus* species) may be life-threatening. Identification of the offending spider should be attempted both because specific treatments exist for bites of widow and brown recluse spiders and because injuries attributed to spiders are frequently due to other causes. Except in cases where the patient actually observes a spider immediately associated with the bite or fleeing from the site, lesions reported to be due to spider bites are most often due to other injuries or to infections with bacteria such as methicillin-resistant *Staphylococcus aureus*.

RECLUSE SPIDER BITES AND NECROTIC ARACHNIDISM

Brown recluse spiders occur mainly in the southern and midwestern United States, and their close relatives are found in the Americas, Africa, and the Middle East. Most bites by the brown recluse spider result in only minor injury with edema and erythema. Envenomation, however, may cause severe necrosis of skin and subcutaneous tissue and systemic hemolysis. These spiders are not aggressive toward human beings and bite only if threatened or pressed against the skin. They hide under rocks and logs or in caves and animal burrows. They invade homes and seek dark and undisturbed hiding spots in closets, in folds of clothing, or under furniture and rubbish in storage rooms, garages, and attics. Despite their impressive abundance in some homes, these spiders only infrequently bite humans. Bites tend to occur while the victim is dressing and are sustained primarily on the arms, neck, and lower abdomen.

The venoms of these spiders contain an esterase, alkaline phosphatase, proteases, and other enzymes that produce tissue necrosis and hemolysis. Sphingomyelinase D, the most important dermonecrotic factor, binds cell membranes and promotes chemotaxis of neutrophils, leading to vascular thrombosis and an Arthus-like reaction. Initially, the bite is painless or produces a stinging

sensation. Within the next few hours, the site becomes painful and pruritic, with central induration surrounded by a pale zone of ischemia and a zone of erythema. In most cases, the lesion resolves without treatment in just a few days. In severe cases, the erythema spreads, and the center of the lesion becomes hemorrhagic and necrotic with an overlying bulla. A black eschar forms and sloughs several weeks later, leaving an ulcer that eventually may result in a depressed scar. Healing usually takes place in ≤6 months but may take as long as 3 years if adipose tissue is involved. Local complications include injury to nerves and secondary infection. Fever, chills, weakness, headache, nausea, vomiting, myalgia, arthralgia, maculopapular rash, and leukocytosis may develop ≤72 h after the bite. In rare instances, acute complications such as hemolytic anemia, hemoglobinuria, and renal failure are fatal.

Treatment: Recluse Spider Bites

Initial management includes RICE (rest, ice, compression, elevation). Analgesics, antihistamines, antibiotics, and tetanus prophylaxis should be administered if indicated. Debridement and later skin grafting may be necessary after signs of acute inflammation have subsided, but immediate surgical excision of the wound is detrimental. Patients should be monitored closely for signs of hemolysis, renal failure, and other systemic complications.

WIDOW SPIDER BITES

The black widow, which is best known and most abundant in the southeastern United States, measures ≤1 cm in body length and 5 cm in leg span and is shiny black with a red hourglass marking on the ventral abdomen. Other dangerous *Latrodectus* species occur elsewhere in temperate and subtropical parts of the world. The bites of the female widow spiders are notorious for their potent neurotoxins.

Widow spiders spin their webs under stones, logs, plants, or rock piles and in dark spaces in barns, garages, and outhouses. Bites are most common in the summer and early autumn and occur when the web is disturbed or when the spider is trapped or provoked. The initial bite goes unnoticed or is perceived as a sharp pinprick. Fang puncture marks are uncommon. The venom that is injected does not produce local necrosis, and some persons experience no other symptoms. α-Latrotoxin, the most active component of the venom, binds irreversibly to nerves and causes release and eventual depletion of acetylcholine, norepinephrine, and other neurotransmitters from presynaptic terminals. Painful cramps may spread within 60 min from the bite site to large muscles of the extremities and trunk. Extreme rigidity of the abdominal muscles and excruciating pain may suggest peritonitis, but the abdomen is not tender on palpation. The pain begins to subside during the first 12 h but may recur during several days or weeks before resolving spontaneously. Other features include salivation, diaphoresis, vomiting, hypertension, tachycardia, labored breathing, anxiety, headache, weakness, fasciculations, paresthesia, hyperreflexia, urinary retention, uterine contractions, and premature labor. Rhabdomyolysis and renal failure have been reported, and respiratory arrest, cerebral hemorrhage, or cardiac failure may end fatally, especially in very young, elderly, or debilitated persons.

Treatment: Widow Spider Bites

Treatment consists of RICE and tetanus prophylaxis. Hypertension that does not respond to analgesics and antispasmodics (e.g., benzodiazepines or methocarbamol) requires specific antihypertensive medication. The efficacy of antivenom is controversial. Because of the risk of anaphylaxis and serum sickness, antivenom should be reserved for severe cases involving respiratory arrest, uncontrollable hypertension, seizures, or pregnancy.

TARANTULAS AND OTHER SPIDERS

Tarantulas are hairy spiders of which 30 species are found in the United States, mainly in the Southwest. The tarantulas that have become popular household pets are usually imported species. Tarantulas bite only when threatened and cause no more harm than a bee sting, but the venom occasionally provokes deep pain and swelling. Tarantulas of several species are covered with urticating hairs that are brushed off in the thousands when a threatened spider rubs its hind legs across its dorsal abdomen. These hairs penetrate human skin and produce pruritic papules that may persist for weeks. Failure to wear gloves or to wash the hands after handling the Chilean Rose tarantula, a popular pet spider, has resulted in transfer of hairs to the eye and devastating ocular inflammation. Treatment of bites includes local washing and elevation of the bitten area, tetanus prophylaxis, and analgesic administration. Antihistamines and topical or systemic glucocorticoids are given for exposure to urticating hairs.

Atrax robustus, a funnel-web spider of Australia, and *Phoneutria* species, the South American banana spiders, are among the most dangerous spiders in the world because of their aggressive behavior and potent neurotoxins. Envenomation by *A. robustus* causes a rapidly progressive neuromotor syndrome that can be fatal within 2 h. The bite of a banana spider causes severe local pain followed by profound systemic symptoms and respiratory paralysis that can lead to death within 2–6 h. Specific antivenoms for envenomation by each of these spiders are available. Yellow sac spiders (*Cheiracanthium)* are common in homes worldwide. Their bites, though painful, generally lead to only minor erythema, edema, and pruritus.

Scorpion Stings

Scorpions are arachnids that feed on ground-dwelling arthropods and small lizards, which they paralyze by injecting venom from a stinger on the tip of the tail. Painful but relatively harmless scorpion stings need to be distinguished from the potentially lethal envenomations that are produced by ~30 of the ~1000 known species and that cause >5000 deaths worldwide each year. Scorpions feed at night and remain hidden during the day in crevices or burrows or under wood, loose bark, or rocks on the ground. They seek cool spots under buildings and often enter houses, where they hide in shoes, clothing, or bedding or enter bathtubs and sinks in search of water. Scorpions sting human beings only when disturbed.

Of the 40 or so scorpion species in the United States, only the bark scorpion (*Centruroides sculpturatus* or *C. exilicauda*) produces venom that can be lethal. This venom contains neurotoxins that cause sodium channels to remain open and neurons to fire repetitively. Such envenomations usually are associated with little swelling, but prominent pain, paresthesia, and hyperesthesia can be accentuated by tapping on the affected area (the tap test). These symptoms soon spread to other locations; dysfunction of cranial nerves and hyperexcitability of skeletal muscles develop within hours. Patients present with restlessness, blurred vision, abnormal eye movements, profuse salivation, lacrimation, rhinorrhea, slurred speech, difficulty in handling secretions, diaphoresis, nausea, and vomiting. Muscle twitching, jerking, and shaking may be mistaken for a seizure. Complications include tachycardia, arrhythmias, hypertension, hyperthermia, rhabdomyolysis, and acidosis. Symptoms progress to maximal severity in ~5 h and subside within a day or two, although pain and paresthesia can last for weeks. Fatal respiratory arrest is most common among young children and the elderly.

Envenomations by *Leiurus quinquestriatus* in the Middle East and North Africa, by *Mesobuthus tamulus* in India, by *Androctonus* species along the Mediterranean littoral and in North Africa and the Middle East, and by *Tityus serrulatus* in Brazil cause massive release of endogenous catecholamines with hypertensive crises, arrhythmias, pulmonary edema, and myocardial damage. Acute pancreatitis occurs with stings of *Tityus trinitatis* in Trinidad, and central nervous toxicity complicates stings of *Parabuthus* and *Buthotus* scorpions of South Africa. Tissue necrosis and hemolysis may follow stings of the Iranian *Hemiscorpius lepturus*.

Stings of most other species cause immediate sharp local pain followed by edema, ecchymosis, and a burning sensation. Symptoms typically resolve within a few hours, and skin does not slough. Allergic reactions to the venom sometimes develop.

Treatment: Scorpion Stings

Identification of the offending scorpion aids in planning therapy. Stings of nonlethal species require at most ice packs, analgesics, or antihistamines. Because most victims experience only local discomfort, they can be managed at home with instructions to return to the emergency department if signs of cranial-nerve or neuromuscular dysfunction develop. Aggressive supportive care and judicious use of antivenom can reduce or eliminate deaths from more severe envenomations. Keeping the patient calm and applying pressure dressings and cold packs to the sting site are measures that decrease the absorption of venom. A continuous IV infusion of midazolam controls the agitation, flailing, and involuntary muscle movements produced by scorpion stings. Close monitoring during treatment with this drug and other sedatives or narcotics is necessary for persons with neuromuscular symptoms because of the risk of respiratory arrest. Hypertension and pulmonary edema respond to nifedipine, nitroprusside, hydralazine, or prazosin, and bradyarrhythmias can be controlled with atropine.

Commercially prepared antivenoms are available in several countries for some of the most dangerous species. A *C. sculpturatus* antivenom (not yet approved by the FDA) is available as an investigational drug only in Arizona. IV administration of antivenom rapidly reverses cranial-nerve dysfunction and muscular symptoms but does not affect pain and paresthesia. The benefit of scorpion antivenom has not been established in controlled trials.

Hymenoptera Stings

Insects that sting to defend their colonies or subdue their prey belong to the order Hymenoptera, which includes bees, wasps, hornets, yellow jackets, and ants. Their venoms contain a wide array of amines, peptides, and enzymes that cause local and systemic reactions. Although the toxic effect of multiple stings can be fatal, nearly all of the ≥100 deaths due to hymenopteran stings in the United States each year result from allergic reactions.

BEE AND WASP STINGS

Honeybees often lose their stinging apparatus and the attached venom sac in the act of stinging and subsequently die, whereas other bees, ants, and vespids can sting numerous times in succession. The familiar honeybees (*Apis mellifera*) and bumblebees (*Bombus* and other genera) generally attack only when a colony is disturbed. Africanized honeybees, however, respond to minimal intrusions more aggressively. Since their introduction into Brazil in 1957, these "killer bees" have spread through South and Central America to the southern and western United States.

In bees and wasps, venom is produced in glands at the posterior end of the abdomen and is expelled rapidly by contraction of muscles of the venom sac, which has a capacity of up to 0.1 mL. The venoms of different species of hymenopterans are biochemically and immunologically distinct. Direct toxic effects are mediated by mixtures of low-molecular-weight compounds such as serotonin, histamine, and acetylcholine and several kinins. Polypeptide toxins in honeybee venom include mellitin, which damages cell membranes; mast cell–degranulating protein, which causes histamine release; apamin, a neurotoxin; and adolapin, which has anti-inflammatory activity. Enzymes in venom include hyaluronidase, which allows the spread of other venom components, and phospholipases, which may be among the major venom allergens. There appears to be little cross-sensitization between honeybee and wasp venoms.

Uncomplicated stings cause immediate pain, a wheal-and-flare reaction, and local edema and swelling that subside in a few hours. Stings from accidentally swallowed insects may induce life-threatening edema of the upper airways. Multiple stings can lead to vomiting, diarrhea, generalized edema, dyspnea, hypotension, and collapse. Rhabdomyolysis and intravascular hemolysis may cause renal failure. Death from the direct effects of venom has followed 300–500 honeybee stings.

Large local reactions that spread ≥10 cm around the sting site over 24–48 h are not uncommon. These reactions may resemble cellulitis but are caused by hypersensitivity rather than secondary infection. Such reactions tend to recur on subsequent exposure but are seldom accompanied by anaphylaxis and are not prevented by venom immunotherapy.

An estimated 0.4–4.0% of the U.S. population exhibits clinical immediate-type hypersensitivity to insect stings, and 15% may have asymptomatic sensitization manifested by positive skin tests. Persons who experience severe allergic reactions are likely to have similar reactions after subsequent stings; occasionally, adults who have had mild reactions later experience serious reactions. Mild anaphylactic reactions from insect stings, as from other causes, consist of nausea, abdominal cramping, generalized urticaria, flushing, and angioedema. Serious reactions, including upper airway edema, bronchospasm, hypotension, and shock, may be rapidly fatal. Severe reactions usually begin within 10 min of the sting and only rarely develop after 5 h.

Treatment: Bee and Wasp Stings

Honeybee stingers embedded in the skin should be removed as promptly as possible to limit the quantity of venom delivered. The stinger and venom sac may be scraped off with a blade or a fingernail or grasped with forceps. The site should be cleansed and disinfected and ice packs applied to slow the spread of venom. Elevation of the affected site and administration of analgesics, oral antihistamines, and topical calamine lotion relieve symptoms. Large local reactions may require a short course of oral therapy with glucocorticoids. Patients with numerous stings should be monitored for 24 h for evidence of renal failure or coagulopathy.

Anaphylaxis is treated with subcutaneous (SC) injection of 0.3–0.5 mL of epinephrine hydrochloride in a 1:1000 dilution; treatment is repeated every 20–30 min as necessary. IV epinephrine (2–5 mL of a 1:10,000 solution administered by slow push) is indicated for profound shock. A tourniquet may slow the spread of venom. Parenteral antihistamines, fluid resuscitation, bronchodilators, oxygen, intubation, and vasopressors may be required. Patients should be observed for 24 h for recurrent anaphylaxis.

Persons with a history of allergy to insect stings should carry a sting kit with a preloaded syringe containing epinephrine for self-administration. These patients should seek medical attention immediately after using the kit.

Repeated injections of purified venom produce a blocking IgG antibody response to venom and reduce the incidence of recurrent anaphylaxis. Honeybee, wasp, yellow jacket, and mixed vespid venoms are commercially available for desensitization and for skin testing. Results of skin tests and venom-specific radioallergosorbent tests aid in the selection of patients for immunotherapy and guide the design of such treatment.

STINGING ANTS

Stinging fire ants are an important medical problem in the United States. Imported fire ants infest southern states from Texas to North Carolina, with colonies in California, New Mexico, Arizona, and Virginia. Slight disturbances of their mound nests have provoked massive outpourings of ants and as many as 10,000 stings on a single person. Elderly and immobile persons are at high risk for attacks when fire ants invade dwellings.

Fire ants attach to skin with powerful mandibles and rotate their bodies while repeatedly injecting venom with posteriorly situated stingers. The alkaloid venom consists of cytotoxic and hemolytic piperidines and several proteins with enzymatic activity. The initial wheal-and-flare reaction, burning, and itching resolve in ~30 min, and a sterile pustule develops within 24 h. The pustule ulcerates over the next 48 h and then heals in ≥1 week. Large areas of erythema and edema lasting several days are not uncommon and in extreme cases may compress nerves and blood vessels. Anaphylaxis occurs in ≤ 2% of persons, and seizures and mononeuritis have been reported. Stings are treated with ice packs, topical glucocorticoids, and oral antihistamines. Covering pustules with bandages and antibiotic

ointment may prevent bacterial infection. Epinephrine and supportive measures are indicated for anaphylactic reactions. Whole-body extracts are available for skin testing and immunotherapy, which appears to lower the rate of anaphylactic reactions.

The western United States is home to harvester ants. The painful local reaction that follows harvester ant stings often extends to lymph nodes and may be accompanied by anaphylaxis.

DIPTERAN (FLY AND MOSQUITO) BITES

In the process of feeding on vertebrate blood, adults of certain fly species inflict painful bites, produce local allergic reactions, or transmit pathogenic agents. Bites of mosquitoes, tiny "no-see-um" midges, and phlebotomine sand flies typically produce a wheal and a pruritic papule. Nodular lesions at the site of midge bites may last for months. Bites of small humpbacked black flies (simuliids) leave a bleeding laceration and a painful and pruritic sore that are slow to heal; regional lymphadenopathy, fever, or anaphylaxis occasionally ensues. The widely distributed deer and horse flies as well as the tsetse flies of Africa are stout flies measuring ≤25 mm in length that attack during the day and produce large and painful bleeding punctures.

Treatment: Fly and Mosquito Bites

Treatment of fly bites is symptom-based. Topical application of antipruritic agents, glucocorticoids, or antiseptic lotions may relieve itching and pain. Allergic reactions may require oral antihistamines. Antibiotics may be necessary for the treatment of large bite wounds that become secondarily infected.

FLEA BITES

Common human-biting fleas include the dog and cat fleas (*Ctenocephalides* species) and the rat flea (*Xenopsylla cheopis*), which inhabit the nests and resting sites of their hosts. Sensitized persons develop erythematous pruritic papules, urticaria, and occasionally vesicles and bacterial superinfection at the site of the bite. Treatment consists of antihistamines and antipruritics.

Flea infestations are eliminated by frequent cleaning of nesting sites and of the host's bedding and by application of contact insecticides. Flea infestations in the home may abate if pets are treated with veterinary antiparasitic agents and insect growth regulators.

Tunga penetrans, like other fleas, is a wingless, laterally flattened insect that feeds on blood. Also known as the chigoe flea, sand flea, or jigger, it occurs in tropical regions of Africa and the Americas. Adults live in sandy soil and burrow under the skin between toes, under nails, or on the soles of bare feet. Chigoes engorge on blood and grow from pinpoint to pea size during a 2-week period. The lesions they produce resemble a white pustule with a central black depression and may be pruritic or painful. Occasional complications include tetanus, bacterial infections, and autoamputation of toes. Tungiasis is treated by removal of the intact flea with a sterile needle or scalpel, tetanus vaccination, and topical application of antibiotics.

HEMIPTERAN (TRUE BUG) BITES

Several true bugs of the family Reduviidae inflict bites that produce allergic reactions and are sometimes painful. The cone-nose bugs, so called because of their elongated heads, include the assassin and wheel bugs, which feed on other insects and bite vertebrates only in self-defense, and the kissing bugs, which routinely feed on vertebrate blood. The bites of the nocturnally feeding kissing bugs are painless. Reactions to such bites depend on prior sensitization and include tender and pruritic papules, vesicular or bullous lesions, giant urticaria, fever, lymphadenopathy, and anaphylaxis. Bug bites are treated with topical antipruritics or oral antihistamines. Persons with anaphylactic reactions to reduviid bites should keep an epinephrine kit available. The cosmopolitan bedbugs (*Cimex* species) hide in crevices of mattresses, bed frames and other furniture, walls, and picture frames and under loose wallpaper. Bedbugs have become resurgent, recently attaining populations and spreading to an extent not encountered since the mid-twentieth century. These bugs are now a fairly common nuisance in homes, dormitories, and hotels and on cruise ships. The bugs hide during the day and take their blood meal at night. Their bite is painless, but sensitized persons develop erythema, itching, and wheals around a central hemorrhagic punctum. Bedbugs are not known to transmit pathogens.

CENTIPEDE BITES AND MILLIPEDE DERMATITIS

The fangs of centipedes of the genus *Scolopendra* can penetrate human skin and deliver a venom that produces intense burning pain, swelling, erythema, and lymphangitis. Dizziness, nausea, and anxiety occasionally are described, and rhabdomyolysis and renal failure have been reported. Treatment includes washing of the site, application of cold dressings, oral analgesic administration or local lidocaine infiltration, and tetanus prophylaxis.

Millipedes, unlike centipedes, do not bite, but some secrete defensive fluids that burn and discolor human skin. Affected skin turns brown overnight and may blister and exfoliate. Secretions in the eye cause intense pain and inflammation that may lead to corneal ulceration and blindness. Management includes irrigation with copious amounts of water or saline, use of analgesics, and local care of denuded skin.

CATERPILLAR STINGS AND DERMATITIS

The surface of caterpillars of several moth species is covered with hairs or spines that produce mechanical irritation and may contain or be coated with venom. Contact with these caterpillars causes an immediate burning sensation followed by local swelling and erythema and occasionally by regional lymphadenopathy, nausea, vomiting, and headache; shock, seizures, and coagulopathy are rare complications. In the United States, dermatitis most often is associated with io, puss, saddleback, and brown-tail moths. Contact with even detached hairs of other caterpillars, such as gypsy moth larvae, can later produce a pruritic urticarial or papular rash. Spines may be deposited on tree trunks and drying laundry or may be airborne and cause irritation of the eyes and upper airways. Treatment of caterpillar stings consists of repeated application of adhesive or cellophane tape to remove the hairs, which can then be identified microscopically. Local ice packs, topical steroids, and oral antihistamines relieve symptoms.

BEETLE VESICATION

When disturbed, blister beetles extrude cantharidin, a low-molecular-weight toxin that produces thin-walled blisters measuring ≤5 cm in diameter 2–5 h after contact. The blisters are not painful or pruritic unless broken and resolve without treatment in ≤10 days. Nephritis may follow unusually heavy cantharidin exposure. Contact occurs when individuals sit on the ground, work in the garden, or deliberately handle the beetles. The hemolymph of certain rove beetles contains paederin, a potent vesicant. When these beetles are crushed or brushed against the skin, the released fluid may provoke erythematous and bullous lesions. These beetles occur worldwide but are most numerous and problematic in parts of Africa and Asia. Ocular lesions are common from impacts with the flying beetles at night or transfer of the vesicant on the fingers. Treatment is rarely necessary, although ruptured blisters should be kept clean and bandaged.

DELUSIONAL INFESTATIONS

The groundless conviction that one is infested with arthropods or other parasites is an extremely difficult disorder to treat and unfortunately is not rare. Patients report infestations of their skin, clothing, or homes and describe sensations of something moving in or on their skin. Excoriations often accompany reports of pruritus or insect bites. Frequently, patients submit as evidence of infestation specimens that consist of plant-feeding and nonbiting peridomestic arthropods, pieces of skin, vegetable matter, or inanimate objects. It is imperative to rule out true infestations and bites by arthropods, endocrinopathies, neuropathies, drug use, environmental irritants (e.g., fragments or threads of glass insulation), and other causes of tingling or prickling sensations. Frequently, such patients repeatedly seek medical consultations, resist alternative explanations for their symptoms, and exacerbate their discomfort by self-treatment. Pharmacotherapy with pimozide or other psychotropic agents has been more helpful than psychotherapy in treating this disorder.

ACKNOWLEDGMENT

The substantial contributions of James H. Maguire to this chapter in previous editions are gratefully acknowledged.

FURTHER READINGS

Clark RP, Hu LT: Prevention of Lyme disease and other tick-borne infections. Infect Dis Clin North Am 22:381, 2008[PMID: 18755380] [Full Text]

Goddard J: *Physician's Guide to Arthropods of Medical Importance* , 5th ed. Boca Raton, FL, CRC Press, 2007

Hicks MI, Elston DM: Scabies. Dermatol Ther 22:279, 2009[PMID: 19580575] [Full Text]

Hinkle N: Ekbom syndrome: The challenge of "invisible bug" infestations. Annu Rev Entomol 55:77, 2010[PMID: 19961324] [Full Text]

McGraw TA, Turiansky MC: Cutaneous myiasis. J Am Acad Dermatol 58:907, 2008[PMID: 18485982] [Full Text]

Mullen G, Durden L: *Medical and Veterinary Entomology* , 2nd ed. Amsterdam, Academic Press, 2009

Pollack RJ, Marcus L: A travel medicine guide to arthropods of medical importance. Infect Dis Clin North Am 19:169, 2005[PMID: 15701553] [Full Text]

Saucier JR: Arachnid envenomation. Emerg Med Clin North Am 22:405, 2004[PMID: 15163574] [Full Text]

Steen CJ et al: Insect sting reactions to bees, wasps, and ants. Int J Dermatol 44:91, 2005[PMID:

15689203] [Full Text]

Vetter RS, Isbister GK: Medical aspects of spider bites. Annu Rev Entomol 53:409, 2008[PMID: 17877450] [Full Text]

BIBLIOGRAPHY

Burgess IF: Human lice and their control. Annu Rev Entomol 49:457, 2004[PMID: 14651472] [Full Text]

Tappe D, Buttner DW: Diagnosis of human visceral pentastomiasis. PLoS Negl Trop Dis 3:e320, 2009

Vetter RS: Arachnids misidentified as brown recluse spiders by medical personnel and other authorities in North America. Toxicon 54:545, 2009[PMID: 19446575] [Full Text]

Visscher PK et al: Removing bee stings: Speed matters, method doesn't. Lancet 348:301, 1996[PMID: 8709689] [Full Text]

Wong HC: Importance of proper identification of stinging insects. Ann Intern Med 132:418, 2000[PMID: 10691598] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

ALTITUDE ILLNESS: INTRODUCTION

Mountains cover one-fifth of the earth's surface; 38 million people live permanently at altitudes ≥2400 m, and 100 million people travel to high-altitude locations each year. Skiers in Aspen, religious pilgrims to Lhasa, trekkers and climbers to Kilimanjaro or Everest, and military personnel deployed to high-altitude locales are all at risk of developing acute mountain sickness (AMS), high-altitude cerebral edema (HACE), high-altitude pulmonary edema (HAPE), and other altitude-related problems. AMS is the benign form of altitude illness, whereas HACE and HAPE are life-threatening. Altitude illness is likely to occur above 2500 m but has been documented even at 1500–2500 m.

Ascent to a high altitude subjects the body to a decrease in barometric pressure that results in a decreased partial pressure of oxygen in the inspired gas in the lungs. This change leads in turn to less pressure driving oxygen diffusion from the alveoli and throughout the oxygen cascade. A normal initial "struggle response" to such an ascent includes increased ventilation, which is the cornerstone of acclimation. Hyperventilation may cause respiratory alkalosis and dehydration. Alkalosis may depress the ventilatory drive during sleep, with consequent periodic breathing and hypoxemia. During early acclimation, renal suppression of carbonic anhydrase and excretion of dilute alkaline urine combat alkalosis and tend to bring the pH of the blood to normal. Other physiologic changes during normal acclimation include increased sympathetic tone; increased erythropoietin levels, leading to increased hemoglobin levels and red blood cell mass; increased tissue capillary density and mitochondrial numbers; and higher levels of 2,3-bisphosphoglycerate, enhancing oxygen utilization. Even with normal acclimation, however, ascent to a high altitude decreases maximal exercise tolerance and increases susceptibility to cold injury due to peripheral vasoconstriction. Finally, if the ascent is made faster than the body can adapt to the stress of hypobaric hypoxemia, altitude-related disease states can result.

Acute Mountain Sickness and High-Altitude Cerebral Edema

AMS is a neurologic syndrome characterized by nonspecific symptoms (headache, nausea, fatigue, and dizziness) with a paucity of physical findings developing 6–12 h after ascent to a high altitude. AMS must be distinguished from exhaustion, dehydration, hypothermia, alcoholic hangover, and hyponatremia. AMS and HACE are thought to represent opposite ends of a continuum of altitude-related neurologic disorders. HACE (but not AMS) is an encephalopathy whose hallmarks are ataxia and altered consciousness with diffuse cerebral involvement but generally without focal neurologic deficits. Progression to these signal manifestations can be rapid. Papilledema and, more commonly, retinal hemorrhages may also be seen. Retinal hemorrhages occur frequently at ≥5000 m, even in individuals without clinical symptoms of AMS or HACE. It is unclear whether retinal hemorrhage and cerebral hemorrhage at high altitude are caused by the same mechanism; however, one report revealed a correlation between HACE and retinopathy.

The most important risk factors for the development of altitude illness are the rate of ascent and a history of high-altitude illness. Exertion is a risk factor, but lack of physical fitness is not. An attractive but still speculative hypothesis proposes that AMS develops in people who have inadequate cerebrospinal capacity to buffer the brain swelling that occurs at high altitude. One protective factor in AMS is high-altitude exposure during the preceding 2 months; the explanation for this association is intriguing but remains uninvestigated. Children and adults seem to be equally affected, but people >50 years of age may be less likely to develop AMS than younger people. Most studies reveal no gender difference in AMS incidence. Sleep desaturation—a common phenomenon at high altitude—is associated with AMS. Debilitating fatigue consistent with severe AMS on descent from a summit is also an important risk factor for death in mountaineers.

PATHOPHYSIOLOGY

The exact mechanisms causing these syndromes are unknown. Evidence points to a central nervous system process. MRI studies have suggested that vasogenic (interstitial) cerebral edema is a component of the pathophysiology of HACE. In the setting of high-altitude illness, the MRI findings shown in Fig. e51-1 are confirmatory of HACE, with increased signal in the white matter and particularly in the splenium of the corpus callosum. Quantitative analysis in a 3-tesla MRI study revealed that hypoxia is associated with mild vasogenic cerebral edema irrespective of AMS. This finding is in keeping with case reports of suddenly symptomatic brain tumors and of cranial nerve palsies without AMS at high altitudes. Vasogenic edema may become cytotoxic (intracellular) in severe HACE.

Figure e51-1

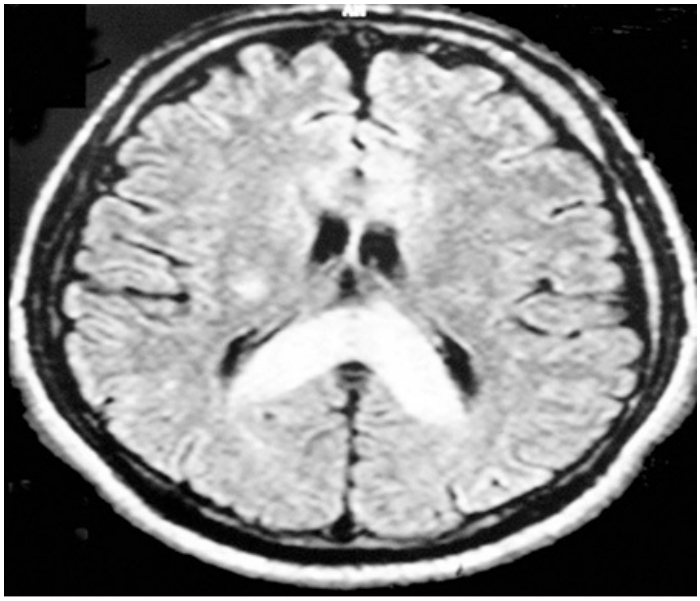

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

T2 MRI image of the brain of a patient with HACE showing marked swelling and a hyperintense posterior body and splenium of the corpus callosum (area with dense opacity). The patient, a climber, went on to climb Mount Everest about 9 months after this episode of HACE. *[With permission from Wilderness and Environmental Medicine, 15(1): 53-55, Spring 2004.]*

Impaired cerebral autoregulation in the presence of hypoxic cerebral vasodilatation and altered permeability of the blood-brain barrier due to hypoxia-induced chemical mediators like histamine, arachidonic acid, and vascular endothelial growth factor (VEGF) may all contribute to brain edema. In 1995, VEGF was first proposed as a potent promoter of capillary leakage in the brain at high altitude, and studies in mice have borne out this role. Although preliminary studies of VEGF in climbers have yielded inconsistent results and have revealed no association with altitude illness, indirect evidence of a role for this growth factor in AMS and HACE comes from the observation that dexamethasone, when used in the prevention and treatment of AMS and HACE, blocks hypoxic upregulation of VEGF. Other factors in the development of cerebral edema may be the release of calcium-mediated nitric oxide and neuronally mediated adenosine, which may promote cerebral vasodilation.

Increased sympathetic activity triggered by hypoxia may also contribute to blood-brain barrier leakage. Enhanced optic-nerve sheath diameter with increasing severity of AMS has been noted and suggests an important role for increased intracranial pressures in the pathophysiology of AMS. Finally, the effect of hypoxia on reactive oxidant species and the role of these species in clinical AMS are unclear.

The pathophysiology of the most common and prominent symptom of AMS—headache—remains unclear because the brain itself is an insensate organ; only the meninges contain trigeminal sensory nerve fibers. The cause of high-altitude headache is multifactorial; various chemicals and mechanical factors activate a final common pathway, the trigeminovascular system. In the genesis of high-altitude headache, the response to nonsteroidal anti-inflammatory drugs (NSAIDs) and glucocorticoids provides indirect evidence for involvement of the arachidonic acid pathway and inflammation. Although the International Headache Society acknowledges that high altitude may be a trigger for migraine, it is unclear whether high-altitude headache and migraine share the same pathophysiology. However, sumatriptan (50 mg by mouth within 1 h of ascent) prevents AMS.

PREVENTION AND TREATMENT (TABLE E51–1)

Gradual ascent, with adequate time for acclimation, is the best method for the prevention of altitude illness. Above 3000 m, a graded ascent of ≤300 m from the previous day's sleeping altitude is generally recommended, and taking every third day of gain in sleeping altitude as an extra day for acclimation is helpful. Spending one night at an intermediate altitude before proceeding to a higher altitude may enhance acclimation and attenuate the risk of AMS. Clearly, a flexible itinerary that permits additional rest days will be helpful. Sojourners to high-altitude locations must be aware of the symptoms of altitude illness and should be encouraged not to ascend further if these symptoms develop. Any hint of HAPE (see below) or HACE mandates descent.

Table e51–1 Management of Altitude Illness

Condition	Management
Acute mountain sickness (AMS), mild[a]	Discontinuation of ascent Treatment with acetazolamide (250 mg q12h) Descent[b]
AMS, moderate[a]	Immediate descent for worsening symptoms Use of low-flow oxygen if available Treatment with acetazolamide (250 mg q12h) and/or dexamethasone (4 mg q6h)[c] Hyperbaric therapy[d]
High-altitude cerebral edema (HACE)	Immediate descent or evacuation Administration of oxygen (2–4 L/min) Treatment with dexamethasone (8 mg PO/IM/IV; then 4 mg q6h) Hyperbaric therapy if descent is not possible
High-altitude pulmonary edema (HAPE)	Immediate descent or evacuation Minimization of exertion while patient is kept warm Administration of oxygen (4–6 L/min) to bring O_2 saturation to >90% Adjunctive therapy with nifedipine[e] (30 mg, extended-release, q12h) Hyperbaric therapy if descent is not possible

[a]Categorization of cases as mild or moderate is a subjective judgment based on the severity of headache and the presence and severity of other manifestations (nausea, fatigue, dizziness, insomnia).

[b]No fixed altitude is specified; the patient should descend to a point below that at which symptoms developed.

[c]Acetazolamide treats and dexamethasone masks symptoms. For prevention (as opposed to treatment), 125–250 mg of acetazolamide q12h or (when acetazolamide is contraindicated—e.g., in people with sulfa allergy) 4 mg of dexamethasone q12h may be used.

[d]In hyperbaric therapy, the patient is placed in a portable altitude chamber or bag to simulate descent.

[e]Nifedipine (30 mg extended release q12h) is also effective for the prevention of HAPE, as is salmeterol (125 μg inhaled twice daily), tadalafil (10 mg twice daily), or dexamethasone (8 mg twice daily).

Pharmacologic prophylaxis at the time of travel to high altitudes is warranted for people with a history of AMS or when a graded ascent and acclimation are not possible—e.g., when rapid ascent is necessary for rescue purposes or when flight to a high-altitude location is required. Acetazolamide (125–250 mg twice a day), administered for 1 day before ascent and continued for 2 or 3 days, is effective. Higher doses generally are not required. Paresthesia and a tingling sensation are common side effects of acetazolamide. Dexamethasone (8 mg/d in divided doses) is also effective. A large-scale, randomized, double-blind, placebo-controlled trial in partially acclimated trekkers has clearly shown that *Ginkgo biloba* is ineffective in the prevention of AMS. To date, no trials of NSAIDs in the prevention of AMS have been reported.

For the treatment of mild AMS, rest alone with analgesic use may be adequate. Descent and the use of acetazolamide and (if available) oxygen are sufficient to treat most cases of moderate AMS. Even a minor descent (400–500 m) may be adequate for symptom relief. For moderate AMS or early HACE, dexamethasone (8 mg orally or parenterally) is highly effective. For HACE, immediate descent is mandatory. When descent is not possible because of poor weather conditions or darkness, a simulation of descent in a portable hyperbaric chamber is effective and, like dexamethasone administration, "buys time." Like nifedipine, phosphodiesterase-5 inhibitors have no role in the treatment of AMS or HACE.

High-Altitude Pulmonary Edema

Unlike HACE (a neurologic disorder), HAPE is primarily a pulmonary problem and therefore is not necessarily preceded by AMS. HAPE develops within 2–4 days after arrival at high altitude; it rarely occurs after more than 4 or 5 days at the same altitude, probably because of remodeling and adaptation that render the pulmonary vasculature less susceptible to the effects of hypoxia. A rapid rate of ascent, a history of HAPE, respiratory tract infections, and cold environmental temperatures are risk factors. Men are more susceptible than women. People with abnormalities of the cardiopulmonary circulation leading to pulmonary hypertension—e.g., patent foramen ovale, mitral stenosis, primary pulmonary hypertension, and unilateral absence of the pulmonary artery—are at increased risk of HAPE, even at moderate altitudes. For example, patent foramen ovale is four times more common

among HAPE-susceptible individuals than in the general population. Echocardiography is recommended when HAPE develops at relatively low altitudes (<3000 m) and whenever cardiopulmonary abnormalities predisposing to HAPE are suspected.

The initial manifestation of HAPE may be a reduction in exercise tolerance greater than that expected at the given altitude. Although a dry, persistent cough may presage HAPE and may be followed by the production of blood-tinged sputum, cough in the mountains is almost universal and the mechanism is poorly understood. Tachypnea and tachycardia, even at rest, are important markers as illness progresses. Crackles may be heard on auscultation but are not diagnostic. HAPE may be accompanied by signs of HACE. Patchy or localized opacities (Fig. e51-2) or streaky interstitial edema may be noted on chest radiography. In the past, HAPE was mistaken for pneumonia due to the cold or for heart failure due to hypoxia and exertion. Kerley B lines or a bat-wing appearance are not seen on radiography. Electrocardiography may reveal right ventricular strain or even hypertrophy. Hypoxemia and respiratory alkalosis are consistently present unless the patient is taking acetazolamide, in which case metabolic acidosis may supervene. Assessment of arterial blood gases is not necessary in the evaluation of HAPE; an oxygen saturation reading with a pulse oximeter is generally adequate. The existence of a subclinical form of HAPE probably has been suggested by an increased alveolar-arterial oxygen gradient in Everest climbers near the summit, but hard evidence correlating this abnormality with the development of clinically relevant HAPE is lacking.

Figure e51-2

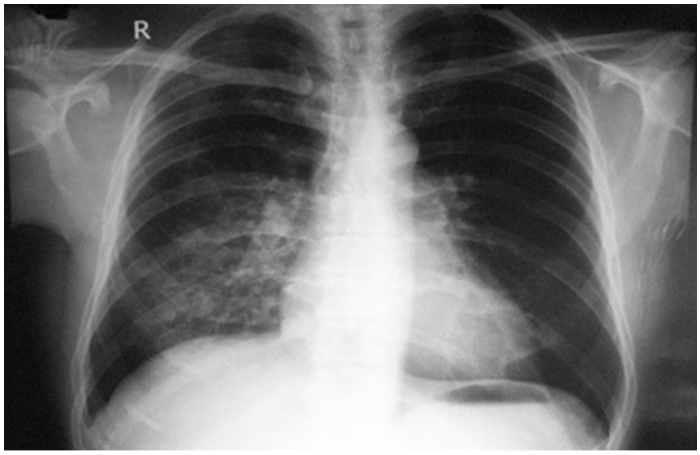

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Chest radiograph of a patient with HAPE shows opacity in the right mid- and lower zones simulating pneumonic consolidation. The opacity cleared almost completely in 2 days with descent and oxygen.

PATHOPHYSIOLOGY

HAPE is a noncardiogenic pulmonary edema characterized by patchy pulmonary vasoconstriction that leads to overperfusion in some areas. This abnormality leads in turn to increased pulmonary capillary pressure (>18 mmHg) and capillary "stress" failure. The exact mechanism for the vasoconstriction is unknown. Endothelial dysfunction due to hypoxia may play a role by impairing the release of nitric oxide, an endothelium-derived vasodilator. At high altitude, HAPE-prone persons have reduced levels of exhaled nitric oxide. The effectiveness of phosphodiesterase-5 inhibitors in alleviating altitude-induced pulmonary hypertension, decreased exercise tolerance, and hypoxemia supports the role of nitric oxide in the pathogenesis of HAPE. One study demonstrated that prophylactic use of tadalafil, a phosphodiesterase-5 inhibitor, decreases the risk of HAPE by 65%. In contrast, the endothelium also synthesizes endothelin-1, a potent vasoconstrictor whose concentrations are higher than average in HAPE-prone mountaineers. Bosentan, an endothelin receptor antagonist, attenuates hypoxia-induced pulmonary hypertension, but further field studies with this drug are necessary.

Exercise and cold lead to increased pulmonary intravascular pressure and may predispose to HAPE. In addition, hypoxia-triggered increases in sympathetic drive may lead to pulmonary venoconstriction and extravasation into the alveoli from the pulmonary capillaries. Consistent with this concept, phentolamine, which elicits α-adrenergic blockade, improves hemodynamics and oxygenation in HAPE more than do other vasodilators. The study of tadalafil cited above also investigated dexamethasone in the prevention of HAPE. Surprisingly, dexamethasone reduced the incidence of HAPE by 78%—a greater decrease than with tadalafil. Besides possibly increasing the availability of endothelial nitric oxide, dexamethasone may have altered the excessive sympathetic activity associated with HAPE: the heart rate of participants in the dexamethasone arm of the study was significantly lowered. Finally, people susceptible to HAPE also display enhanced sympathetic activity during short-term hypoxic

breathing at low altitudes.

Because many patients with HAPE have fever, peripheral leukocytosis, and an increased erythrocyte sedimentation rate, inflammation has been considered an etiologic factor in HAPE. However, strong evidence suggests that inflammation in HAPE is an epiphenomenon rather than the primary cause. Nevertheless, inflammatory processes (e.g., those elicited by viral respiratory tract infections) do predispose persons to HAPE—even those who are constitutionally resistant to its development.

Another proposed mechanism for HAPE is impaired transepithelial clearance of sodium and water from the alveoli. β-Adrenergic agonists upregulate the clearance of alveolar fluid in animal models. In a double-blind, randomized, placebo-controlled study of HAPE-susceptible mountaineers, prophylactic inhalation of the β-adrenergic agonist salmeterol reduced the incidence of HAPE by 50%. Other effects of β agonists may also contribute to the prevention of HAPE, but these findings are in keeping with the concept that alveolar fluid clearance may play a pathogenic role in this illness.

GENETICS

Hypoxia-inducible factor, which is important in high-altitude adaptation, controls transcriptional responses to hypoxia throughout the body and is involved in VEGF release in the brain, erythropoiesis, and other pulmonary and cardiac functions at high altitudes. Hypoxia-inducible factor is regulated through oxygen-dependent proteasomal degradation and thus responds to variations in oxygen availability. Although genepolymorphism may influence susceptibility to HAPE, the data on this point are unclear. Endothelial nitric oxide synthase gene polymorphism has been associated with susceptibility to HAPE in Japan but not in Europe. Although angiotensin-converting enzyme gene polymorphism appears to confer a performance advantage at high altitude, an association with susceptibility to HAPE is lacking.

PREVENTION AND TREATMENT (TABLE E51–1)

Allowing sufficient time for acclimation by ascending gradually (as discussed above for AMS and HACE) is the best way to prevent HAPE. Sustained-release nifedipine (30 mg), given once or twice daily, prevents HAPE in people who must ascend rapidly or who have a history of HAPE. Other drugs for the prevention of HAPE are listed in Table e51–1 (footnote *e*). Although dexamethasone is listed for prevention, its adverse-effect profile requires close monitoring. Acetazolamide has been shown to blunt hypoxic pulmonary vasoconstriction, and this observation warrants further study in HAPE prevention. However, one study failed to show a decrease in pulmonary vasoconstriction in partially acclimated individuals given acetazolamide.

Early recognition is paramount in the treatment of HAPE, especially when it is not preceded by the AMS symptoms of headache and nausea. Fatigue and dyspnea at rest may be the only initial manifestations. Descent and the use of supplementary oxygen (aimed at bringing oxygen saturation to >90%) are the most effective therapeutic interventions. Exertion should be kept to a minimum, and the patient should be kept warm. Hyperbaric therapy in a portable altitude chamber may be used if descent is not possible and oxygen is not available. Oral sustained-release nifedipine (30 mg once or twice daily) can be used as adjunctive therapy. Inhaled β agonists, which are safe and convenient to carry, are useful in the prevention of HAPE and may be effective in its treatment, although no trials have yet been carried out. Inhaled nitric oxide and expiratory positive airway pressure may also be useful therapeutic measures but may not be available in high-altitude settings. No studies have investigated phosphodiesterase-5 inhibitors in the treatment of HAPE, but reports have described their use in clinical practice. The mainstays of treatment remain descent and (if available) oxygen.

In AMS, if symptoms abate (with or without acetazolamide), the patient may reascend gradually to a higher altitude. Unlike that in acute respiratory distress syndrome (another noncardiogenic pulmonary edema), the architecture of the lung in HAPE is usually well preserved, with rapid reversibility of abnormalities (Fig. e51-2). This fact has allowed some people with HAPE to reascend slowly after a few days of descent and rest. In HACE, reascent after a few days is not advisable.

Other High-Altitude Problems

SLEEP IMPAIRMENT

The mechanisms underlying sleep problems, which are among the most common adverse reactions to high altitude, include increased periodic breathing, changes in sleep architecture with increased time in lighter sleep stages, and changes in rapid-eye-movement sleep. Sojourners should be reassured that sleep quality improves with acclimation. In cases where drugs do need to be used, acetazolamide (125 mg before bedtime) is especially useful as this agent decreases hypoxemic episodes and alleviates sleeping disruptions caused by excessive periodic breathing. Whether combining acetazolamide with temazepam or zolpidem is more effective than administering acetazolamide alone is unknown. In combinations, the doses of temazepam and zolpidem should not be increased by >10 mg at high altitudes. Limited evidence suggests that diazepam causes hypoventilation at high altitudes and therefore is contraindicated. For trekkers with obstructive sleep apnea who are using a continuous positive airway pressure (CPAP) machine lacking pressure compensation, set pressures may need to be adjusted.

GASTROINTESTINAL ISSUES

Because of decreased atmospheric pressure and consequent intestinal gas expansion at high altitudes, many sojourners experience abdominal bloating and distension and excessive flatus expulsion. In the absence of diarrhea, these phenomena are normal, if sometimes uncomfortable. Accompanying diarrhea, however, may indicate the involvement of bacteria or *Giardia* parasites, which are common at many high-altitude locations in the developing world. Prompt treatment with fluids and empirical antibiotics may be required to combat dehydration in the mountains. Finally, hemorrhoids are common on high-altitude treks; treatment includes hot soaks, application of hydrocortisone ointment, and measures to avoid constipation.

HIGH-ALTITUDE COUGH

High-altitude cough can be very debilitating and is sometimes severe enough to cause rib fracture, especially above 5000 m. The etiology is probably multifactorial. Although high-altitude cough has been attributed to inspiration of cold dry air, this explanation appears not to be sufficient in itself; in long-duration studies in hypobaric chambers, cough has occurred despite controlled temperature and humidity. The implication is that hypoxia also plays a role. Exercise can precipitate cough at high altitudes, possibly because of water loss from the respiratory tract. Long-acting β agonists and glucocorticoids prevent bronchoconstriction that otherwise may be brought on by cold and exercise. In general, infection does not seem to be a common etiology. Anecdotal reports have described the efficacy of an inhaled combination of fluticasone and salmeterol in the treatment of high-altitude cough. Both salmeterol and glucocorticoids are known to prevent HAPE, which may be an underrecognized cause of cough.

HIGH-ALTITUDE NEUROLOGIC EVENTS UNRELATED TO "ALTITUDE ILLNESS"

Transient ischemic attacks (TIAs) and strokes have been well described in high-altitude sojourners outside the setting of altitude sickness. However, these descriptions are not based on cause (hypoxia) and effect. The population that suffers strokes and TIAs at sea level is generally an older age group with other risk factors, whereas those so afflicted at high altitudes are generally younger and probably have fewer risk factors for atherosclerotic vascular disease. Other mechanisms (e.g., vasospasm, focal edema, hypocapneic vasoconstriction, hypoxia in the watershed zones of minimal cerebral blood flow, or cardiac right-to-left shunt) may be operative in TIAs and strokes at high altitude.

Subarachnoid hemorrhage, transient global amnesia, delirium, and cranial nerve palsies (e.g., lateral rectus palsy) occurring at high altitudes but outside the setting of altitude sickness have also been well described. Syncope is common at moderately high altitudes, generally occurs shortly after ascent, usually resolves without descent, and appears to be a vasovagal event related to hypoxemia. Seizures occur rarely with HACE, but hypoxemia and hypocapnia, which are prevalent at high altitudes, are well-known triggers that may contribute to new or breakthrough seizures in predisposed individuals. However, the consensus among experts is that sojourners with well-controlled seizure disorders can ascend to high altitudes. Ophthalmologic problems, such as cortical blindness, amaurosis fugax, and retinal hemorrhage with macular involvement and compromised vision, are well recognized. Visual problems from previous refractive surgery and blurred monocular vision—due either to the use of a transdermal scopolamine patch (touching the eye after touching the patch) or to dry-eye syndrome —may also occur in the field at high altitudes and may be confused with neurologic conditions. Finally, persons with hypercoagulable conditions (e.g., antiphospholipid syndrome, protein C deficiency) who are asymptomatic at sea level may experience cerebral venous thrombosis (possibly due to the enhanced blood viscosity triggered by polycythemia and dehydration) at high altitudes. Proper history-taking, examination, and prompt investigations where possible will help define these conditions as entities separate from altitude sickness. Administration of oxygen (where available) and prompt descent are the cornerstones of treatment of most of these neurologic conditions.

Preexisting Medical Issues

Because travel to high altitudes is increasingly popular, common conditions such as hypertension, coronary artery disease, and diabetes are more frequently encountered among high-altitude sojourners. This situation is of particular concern for the thousands of elderly pilgrims with medical problems who visit high-altitude sacred areas (e.g., in the Himalayas) each year. Although most of these medical conditions do not appear to influence susceptibility to altitude illness, they may be exacerbated by ascent to altitude, exertion in cold conditions, and hypoxemia. Advice regarding the advisability of high-altitude travel and the impact of high-altitude hypoxia on these preexisting conditions is becoming increasingly relevant, but there are no evidence-based guidelines. Personal risks and benefits must be clearly thought through before ascent.

HYPERTENSION

At high altitudes, enhanced sympathetic activity may lead to a transient rise in blood pressure that rarely reaches dangerously high levels. Sojourners should continue to take their antihypertensive medications at high altitudes. Hypertensive patients are not more likely than others to develop altitude illness. Because the probable mechanism of high-altitude hypertension is α-adrenergic activity, anti-α-adrenergic drugs like prazosin have been suggested for symptomatic patients and those with labile

hypertension. It is best to start taking the drug several weeks before the trip and to carry a sphygmomanometer. Sustained-release nifedipine may also be useful.

CORONARY ARTERY DISEASE

Myocardial oxygen demand and maximal heart rate are reduced at high altitudes because VO_2 max decreases with increasing altitude. This effect may explain why signs of cardiac ischemia or dysfunction are not seen in healthy persons at high altitudes. Asymptomatic, fit individuals with no risk factors need not undergo any tests before ascent. For persons with ischemic heart disease, previous myocardial infarction, angioplasty, and/or bypass surgery, an exercise treadmill test is indicated. A strongly positive treadmill test is a contraindication for high-altitude trips; a negative treadmill test in an individual who leads an active life at sea level does not warrant the avoidance of high altitudes. Patients with poorly controlled arrhythmias should avoid high-altitude travel, but patients with arrhythmias that are well controlled with antiarrhythmic medications do not seem to be at increased risk.

ASTHMA

Although cold air and exercise may provoke acute bronchoconstriction, asthmatic patients usually have fewer problems at high than at low altitudes, possibly because of decreased allergen levels and increased circulating catecholamine levels. Nevertheless, asthmatic individuals should carry all their medications, including oral glucocorticoids, with proper instructions for use in case of an exacerbation. Severely asthmatic persons should be cautioned against ascending to high altitudes.

PREGNANCY

In general, low-risk pregnant women ascending to 3000 m are not at special risk except for the relative unavailability of medical care in many high-altitude locations, especially in developing countries. Despite the lack of firm data on this point, venturing higher than 3000 m to altitudes at which oxygen saturation drops steeply seems unadvisable for pregnant women.

OBESITY

Although living at a high altitude has been suggested as a means of controlling obesity, obesity has also been reported to be a risk factor for AMS, probably because nocturnal hypoxemia is more pronounced in obese individuals. Hypoxemia may also lead to greater pulmonary hypertension, thus possibly predisposing the trekker to HAPE.

SICKLE CELL DISEASE

High altitude is one of the rare environmental exposures that occasionally provokes a crisis in persons with the sickle cell trait. Even when traversing mountain passes as low as 2500 m, people with sickle cell disease have been known to have a vasoocclusive crisis. Sickle cell disease needs to be considered when persons traveling to high altitudes become unwell and develop left-upper-quadrant pain. Patients with known sickle cell disease who need to travel to high altitudes should use supplemental oxygen and travel with caution.

DIABETES MELLITUS

Trekking at high altitudes may enhance sugar uptake. Thus, high-altitude travel may not pose problems for persons with diabetes that is well controlled with oral hypoglycemic agents. An eye examination before travel may be useful. Patients taking insulin may require lower doses on trekking/climbing days than on rest days. Because of these variations, diabetic patients need to carry a reliable glucometer and use fast-acting insulin. Ready access to sweets is also essential. It is important for companions of diabetic trekkers to be fully aware of potential problems like hypoglycemia.

CHRONIC LUNG DISEASE

Depending on disease severity and access to medical care, preexisting lung disease may not always preclude high-altitude travel. A proper pretravel evaluation must be conducted. Supplemental oxygen may be required if the predicted PaO_2 for the altitude is <50–55 mmHg. Preexisting pulmonary hypertension may also need to be assessed in these patients. If the result is positive, patients should be discouraged from ascending to high altitudes; if such travel is necessary, treatment with sustained-release nifedipine (20 mg twice a day) should be considered. Small-scale studies have revealed that when patients with bullous disease reach ~5000 m, bullous expansion and pneumothorax are not noted. A handheld pulse oximeter can be useful to check for oxygen saturation.

CHRONIC KIDNEY DISEASE

Patients with chronic kidney disease can tolerate short-term stays at high altitudes, but theoretical concern persists about progression to end-stage renal disease. Acetazolamide, the drug most commonly used for altitude sickness, should be avoided by anyone with preexisting metabolic acidosis, which can be exacerbated by this drug. In addition, the acetazolamide dosage should be adjusted when the glomerular filtration rate falls below 50 mL/min, and the drug should not be used at all if this value falls below 10 mL/min.

Chronic Mountain Sickness and High-Altitude Pulmonary Hypertension

Chronic mountain sickness (also known as Monge's disease) is a disease of long-term residents of altitudes above 2500 m that is characterized by excessive erythrocytosis with moderate to severe pulmonary hypertension leading to cor pulmonale. This condition was originally described in South America and has also been documented in Colorado and in the Han Chinese population in Tibet. Migration to a low altitude results in the resolution of chronic mountain illness. Venesection and acetazolamide are helpful.

High-altitude pulmonary hypertension is also a subacute disease of long-term high-altitude residents. Unlike Monge's disease, this syndrome is characterized primarily by pulmonary hypertension (not erythrocytosis) leading to heart failure. Indian soldiers living at extreme altitudes for prolonged periods and Han Chinese infants born in Tibet have presented with the adult and infantile forms, respectively. High-altitude pulmonary hypertension bears a striking pathophysiologic resemblance to brisket disease in cattle. Descent to a lower altitude is curative.

FURTHER READINGS

Basnyat B, Murdoch D: High altitude illness. Lancet 361:1967, 2003[PMID: 12801752] [Full Text]

——— et al: Neurological conditions of altitude that fall outside the usual definition of altitude sickness. High Alt Med Biol 5:171, 2004

Beall CM et al: Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci USA 107:11459, 2010[PMID: 20534544] [Full Text]

Gertsch JH et al: Randomised, double blind, placebo controlled comparison of ginkgo biloba and acetazolamide for prevention of acute mountain sickness among Himalayan trekkers: The Prevention of High Altitude Illness Trial (PHAIT). BMJ 328:797, 2004[PMID: 15070635] [Full Text]

Grocott MP et al: Arterial blood gases and oxygen content in climbers on Mount Everest. N Engl J Med 360:140, 2009[PMID: 19129527] [Full Text]

Hackett PH, Roach RC: High altitude and common medical conditions, in *High Altitude: An Exploration of Human Adaptation*, TF Hornbein, RB Schoene (eds). New York, Marcel Dekker, 2001, pp 839–885

MacInnis MJ et al: Evidence for a genetic basis for altitude illness: 2010 update. High Alt Med Biol 11:349, 2010[PMID: 21190504] [Full Text]

Maggiorini M et al: Both tadalafil and dexamethasone may reduce the incidence of high-altitude pulmonary edema: A randomized trial. Ann Intern Med 145:497, 2006[PMID: 17015867] [Full Text]

Schoonman GG et al: Hypoxia-induced acute mountain sickness is associated with intracellular cerebral edema: A 3 T magnetic resonance imaging study. J Cereb Blood Flow Metab 28:198, 2008[PMID: 17519973] [Full Text]

Smith TG et al: Effect of iron supplementation and depletion on hypoxic pulmonary

hypertension: Two randomized controlled trials. JAMA 302:1444, 2009

Stream JO, Grissom CK: Update on high-altitude pulmonary edema: Pathogenesis, prevention, and treatment. Wilderness Environ Med 19:293, 2008[PMID: 19099331] [Full Text]

Wilson MH et al: The cerebral effects of ascent to high altitudes. Lancet Neurol 8:175, 2009[PMID: 19161909] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

WHAT IS HYPERBARIC AND DIVING MEDICINE?

Hyperbaric medicine is the treatment of health disorders using whole-body exposure to pressures greater than one atmosphere (760 mmHg). In practice, this almost always means the administration of *hyperbaric oxygen therapy* (HBO_2T). The Undersea and Hyperbaric Medical Society (UHMS) defines HBO_2T as: "a treatment in which a patient breathes 100% oxygen. . .while inside a treatment chamber at a pressure higher than sea level pressure (i.e., >1 atmosphere absolute or ATA)." The treatment chamber is an airtight vessel variously called a hyperbaric chamber, recompression chamber, or decompression chamber, depending on the clinical and historical context. Such chambers may be capable of compressing a single patient (a monoplace chamber) or multiple patients and attendants as required (a multiplace chamber) **(Figs. e52-1 and e52-2)**. Historically, these compression chambers were first employed for the treatment of divers and compressed air workers suffering decompression sickness (DCS; "the bends"). While the prevention and treatment of disorders arising after decompression in diving, aviation, and space flight has developed into a specialized field of its own, it remains closely linked to the broader practice of hyperbaric medicine.

Despite an increased understanding of mechanisms and an improving evidence basis, hyperbaric medicine has struggled to achieve widespread recognition as a "legitimate" therapeutic measure. There are several contributing factors, but high among them are a poor grounding in general oxygen physiology and oxygen therapy at medical schools and a continuing tradition of charlatans advocating hyperbaric therapy (often using air) as a panacea. Funding for both basic and clinical research has been difficult in an environment where the pharmacologic agent under study is abundant, cheap, and unpatentable. Recently, however, there are signs of an improved appreciation of the potential importance of HBO_2T with significant National Institutes of Health (NIH) funding for mechanisms research and from the U.S. military for clinical investigation.

Figure e52-1

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A monoplace chamber (Prince of Wales Hospital, Sydney).

Figure e52-2

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

A chamber designed to treat multiple patients (Karolinska University Hospital).

MECHANISMS OF HYPERBARIC OXYGEN

Increased hydrostatic pressure will reduce the volume of any bubbles present within the body (see "Diving Medicine"), and this is partly responsible for the success of prompt recompression in DCS and arterial gas embolism. Supplemental oxygen breathing has a dose-dependent effect on oxygen transport, ranging from improvement in hemoglobin oxygen saturation when a few liters per minute is delivered by simple mask at 1 ATA to raising the dissolved plasma oxygen sufficiently to sustain life without the need for hemoglobin at all when 100% oxygen is breathed at 3 ATA. Most HBO_2T regimens involve oxygen breathing at between 2 and 2.8 ATA, and the resultant increase in arterial oxygen tensions to greater than 1000 mmHg has widespread physiologic and pharmacologic consequences **(Fig. e52-3)**.

One direct consequence of such high intravascular tension is to greatly increase the effective capillary-tissue diffusion distance for oxygen such that oxygen-dependent cellular processes can resume in hypoxic tissues. Important as this may be, the mechanism of action is not limited to the restoration of oxygenation in hypoxemic tissue. Indeed, there are pharmacologic effects that are profound and long-lasting. While removal from the hyperbaric chamber results in a rapid return of poorly vascularized tissues to their hypoxic state, even a single dose of HBO_2T produces changes in fibroblast, leukocyte, and angiogenic functions and antioxidant defenses that persist many hours after oxygen tensions are returned to pretreatment levels.

Figure e52-3

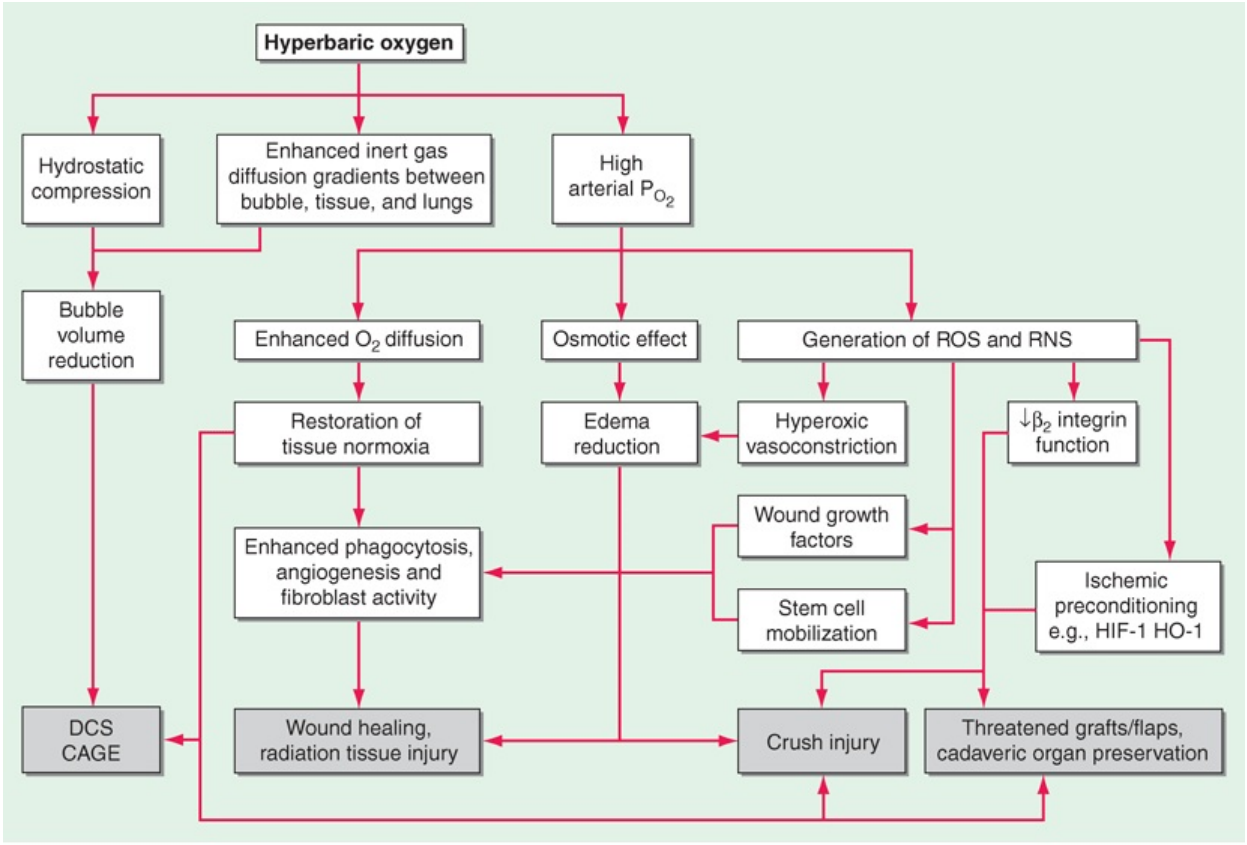

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Summary of mechanisms of hyperbaric oxygen.

There are many consequences of compression and oxygen breathing. The cell-signaling effects of HBO_2T are the least understood but potentially most important. Examples of indications for use are shown in the shaded boxes. CAGE, cerebral arterial gas embolism; DCS, decompression sickness; HIF-1, hypoxia inducible factor-1; HO-1, hemoxygenase 1; RNS, reactive nitrogen species; ROS, reactive oxygen species.

It is widely accepted that oxygen in high doses produces adverse effects due to the production of reactive oxygen species (ROS) such as superoxide (O_2^-) and hydrogen peroxide (H_2O_2). It has become increasingly clear over the last decade that both ROS and reactive nitrogen species (RNS) such as nitric oxide (NO) participate in a wide range of intracellular signaling pathways involved in the production of a range of cytokines, growth factors, and other inflammatory and repair modulators. Such mechanisms are complex and at times apparently paradoxical. Taking as an example the treatment of chronic hypoxic wounds, some effects of HBO_2T are to enhance the clearance of cellular debris and bacteria by providing the substrate for macrophage phagocytosis; to stimulate growth factor synthesis by increased production and stabilization of hypoxia-inducible factor 1 (HIF-1); to inhibit leukocyte activation and adherence to damaged endothelium; and, through the induction of nitric oxide synthetase-3 (NOS-3 or eNOS), to mobilize bone marrow stem cells that will enable vasculogenesis. The interactions between these mechanisms remain a very active field of investigation. One exciting development is the concept of *hyperoxic preconditioning* in which a short exposure to HBO_2 can induce tissue protection against future hypoxic/ischemic insult. This has potential applications in several surgical specialties including organ transplantation. One randomized clinical trial has been completed and suggests that HBO_2T prior to coronary artery bypass grafting reduces biochemical markers of ischemic stress and improves neurocognitive outcome.

ADVERSE EFFECTS OF THERAPY

HBO_2T is generally well tolerated and safe in clinical practice. Adverse effects are associated with both alterations in pressure (barotrauma) and the administration of oxygen.

Barotrauma

Barotrauma occurs when any noncompliant gas-filled space within the body does not equalize with environmental pressure during compression or decompression. About 10% of patients complain of some difficulty equalizing middle-ear pressure early in compression, and while most of these problems are minor and can be overcome with training, 2–5% of conscious patients require middle-ear ventilation tubes or formal grommets across the tympanic membrane. Unconscious patients cannot equalize and

should have middle-ear ventilation tubes placed prior to compression. Other less common sites for barotrauma of compression include the respiratory sinuses and dental caries. The lungs are potentially vulnerable to barotrauma of decompression as described in the diving medicine section, but the decompression following HBO_2T is so slow that pulmonary gas trapping is extremely rare in the absence of an undrained pneumothorax or lesions such as bullae.

Oxygen Toxicity

The practical limit to the dose of oxygen that can be delivered, both in a single treatment session and in a series of daily sessions, is oxygen toxicity. The most common acute manifestation is a seizure, which is often preceded by anxiety and agitation, during which time a switch from oxygen to air breathing may avoid the convulsion. Hyperoxic seizures are typically generalized tonic-clonic seizures followed by a variable postictal period. The cause is an overwhelming of the antioxidant defense systems within the brain. While clearly dose-dependent, onset is very variable both between individuals and within the same individual on different days. In routine clinical hyperbaric practice, the incidence is about 1:1500 to 1:2000 compressions.

Chronic oxygen poisoning most commonly manifests as worsening myopia due to alterations in the refractive index of the lens following oxidative damage to lenticular proteins similar to those associated with senescent cataract formation. Up to 75% of patients show deterioration in visual acuity after a course of 30 treatments at 2 ATA. Almost all return to pretreatment values 3–6 weeks after cessation of treatment. A more rapid maturation of preexisting cataracts has occasionally been associated with HBO_2T. Although a theoretical problem, the development of pulmonary oxygen toxicity over time does not seem to be problematic in practice—probably due to the intermittent nature of the exposure.

CONTRAINDICATIONS TO HYPERBARIC OXYGEN

There are few absolute contraindications to HBO_2T. The most commonly encountered are an untreated pneumothorax and a history of bleomycin administration. A pneumothorax may expand rapidly on decompression and come under tension. Prior to any compression, patients with a pneumothorax should have a patent chest drain in place. The presence of other obvious risk factors for pulmonary gas trapping such as bullae should trigger a very cautious analysis of the risks of treatment versus the benefit. Bleomycin is associated with a partially dose-dependent pneumonitis in about 20% of people. This subgroup may be at particular risk of subsequent rapid deterioration in ventilatory function on exposure to high oxygen tensions. The relationship between pulmonary oxygen toxicity and bleomycin exposure is not proved, particularly after long periods have elapsed between bleomycin and oxygen exposures. However, any patient with a history of bleomycin administration should be carefully counseled prior to exposure to HBO_2T. For those recently exposed to doses above 200 mg and whose course was complicated by a respiratory reaction to bleomycin, compression should be avoided except in a life-threatening situation.

INDICATIONS FOR HYPERBARIC OXYGEN

The appropriate indications for HBO_2T are controversial and evolving. Practitioners in this area are in a unique position. Unlike most branches of medicine, hyperbaric physicians do not deal with a range of disorders within a defined body system (e.g., cardiology), nor are they masters of a therapy specifically designed for a single group of disorders (e.g., radiotherapy). Inevitably, the encroachment of hyperbaric physicians into other medical fields generates suspicion from specialist practitioners in those fields. At the same time this relatively benign therapy, the prescription and delivery of which requires no medical license, attracts both charlatans and well-motivated proselytizers who tout the benefits of oxygen for a plethora of chronic incurable diseases. This battle on two fronts has meant that mainstream hyperbaric physicians have been careful to claim effectiveness only for those conditions where there is a reasonable body of supporting evidence. In 1977, the UHMS systematically examined the claims for the routine use of HBO_2T in more than 100 disorders and found sufficient evidence to support routine use in only 12. The Hyperbaric Oxygen Therapy Committee of that organization has continued to update this list periodically with an increasingly formalized system of appraisal for new indications and emerging evidence **(Table e52–1)**. Around the world, other relevant medical organizations have generally taken a similar approach, although indications vary considerably—particularly those recommended by hyperbaric medical societies in Russia and China where HBO_2T has gained much wider support than in the United States, Europe, and Australasia. Recently, several Cochrane reviews have examined the randomized trial evidence for many putative indications, and attempts have been made to examine the cost-effectiveness of HBO_2T across a range of conditions. **Table e52–2** is a synthesis of these two approaches and lists the estimated cost of attaining health outcomes with the use of HBO_2T. Any savings associated with alternative treatment strategies avoided as a result of HBO_2T have not been taken into account in these estimates (e.g., the avoidance of lower leg amputation in diabetic foot ulcers). Following are short reviews of several important indications that are accepted by the UHMS.

Table e52–1 Current List of Indications for Hyperbaric Oxygen Therapy
1. Air or gas embolism (includes diving-related, iatrogenic, and accidental causes)

2. Carbon monoxide poisoning (including poisoning complicated by cyanide poisoning)
3. Clostridial myositis and myonecrosis (gas gangrene)
4. Crush injury, compartment syndrome, and acute traumatic ischemias
5. Decompression sickness
6. Enhancement of healing in selected problem wounds
7. Exceptional blood loss (where transfusion is refused or impossible)
8. Intracranial abscess
9. Necrotizing soft tissue infections (e.g., Fournier's gangrene)
10. Osteomyelitis (refractory to other therapy)
11. Delayed radiation injury (soft-tissue injury and bony necrosis)
12. Skin grafts and flaps (compromised)
13. Thermal burns

Source: The Undersea and Hyperbaric Medical Society (2008).

Table e52–2 Selected Indications for Which There Is Promising Efficacy for the Application of Hyperbaric Oxygen Therapy

Diagnosis	Outcome (number of sessions)	NNT 95% CI	Estimated Cost to Produce One Extra Favorable Outcome 95% CI (USD)	Comments and Recommendations
Radiation tissue injury	**More information is required on the subset of disease severity and affected tissue type that is most likely to benefit, and the time over which benefit may persist.**			
	Resolved proctitis (30)	3 2–11	22,392 14,928–82,104	Large multi-center trial ongoing
	Healed mandible (30)	4	29,184	Based on one poorly reported study
	Mucosal cover in ORN (30)	2–8 3	14,592–58,368 29,888	Based on one poorly reported study
		2–4	14,592–29,184	
	Bony continuity in ORN (30)	4	29,184	Based on one poorly reported study
		2–8	14,592–58,368	
	Prevention of ORN after dental extraction (30)	4	29,184	Based on a single study
	2–13		14,592–94,848	
	Prevention of dehiscence (30)	5 3–8	36,480 21,888–58,368	Based on one poorly reported study
Chronic wounds	**More information is required on the subset of disease severity or classification most likely to benefit, the time over which benefit may persist, and the most appropriate oxygen dose. Economic analysis is required.**			
	Diabetic ulcer healed at 1 year (30)	2	14,928	Based on one small study, more research required
		1–5	7464–37,320	
	Diabetic ulcer, major amputation avoided (30)	4 3–11	29,856 22,392–82,104	Three small studies. Outcome over a longer time period required
ISSNHL	**No evidence of benefit more than 2 weeks after onset. More research is required to define the role (if any) of HBO_2T in routine therapy.**			
	Improvement of 25% in hearing loss within 2 weeks	5	18,240	Some improvement in hearing, but functional significance unknown

	of onset (15)			
		3 –20	10,944–72,960	
Acute coronary syndrome	**More information is required on the subset of disease severity and timing of therapy most likely to result in benefit. Given the potential of HBO_2T in modifying ischemia-reperfusion injury, attention should be given to the combination of HBO_2T and thrombolysis in early management and in the prevention of restenosis after stent placement.**			
	Episode of MACE (5)	4 3 –10	4864 3648–12,160	Based on a single small study; more research required
	Incidence of significant dysrhythmia (5)	6	7296	Based on a single moderately powered study in the 1970s
		3 –24	3648–29,184	
Traumatic brain injury	**Limited evidence that for acute injury HBO_2T reduces mortality but not functional morbidity. Routine use not yet justified.**			
	Mortality (15)	7 4 –22	34,104 19,488–58,464	Based on four heterogeneous studies
Enhancement of radiotherapy	**There is some evidence that HBO_2T improves local tumor control and reduces mortality for cancers of the head and neck, as well as reducing the chance of local tumor recurrence in cancers of the head, neck, and uterine cervix.**			
	Head and neck cancer: 5-year mortality (12)	5 3 –14	14,592 8755–40,858	Based on trials performed in the 1970s and 1980s. There may be some confounding by radiation fractionation schedule
	Local recurrence 1 year (12)	5	14,592	May no longer be relevant to therapy
	Cancer of uterine cervix:	4–8 5	11,674–23,347 24,320	As above
	Local recurrence at 2 years (20)	4–8	19,456–38,912	
Decompression illness[a]	**Reasonable evidence for reduced number of HBO_2T sessions but similar outcomes when NSAID added.**			
	Reduction of HBO_2T treatment requirement by 1	5 3 –18	N/R	Single appropriately powered randomized trial

[a]Tenoxicam used as an adjunct to recompression on oxygen.

Abbreviations: CI, confidence interval; HBO_2T, hyperbaric oxygen therapy; ISSNHL, idiopathic sudden sensorineural hearing loss; MACE, major adverse cardiac events; N/R, not remarkable; NNT, number needed to treat; NSAID, nonsteroidal anti-inflammatory; ORN, osteoradionecrosis; USD, US dollars.

Source: M Bennett: The evidence-basis of diving and hyperbaric medicine—a synthesis of the high level evidence with meta-analysis, *http://unsworks.unsw.edu.au/vital/access/manager/Repository/unsworks:949.*

Late Radiation Tissue Injury

Radiotherapy is a well-established treatment for suitable malignancies. In the United States alone, approximately 300,000 individuals annually will become long-term survivors of cancer treated by irradiation. Serious radiation-related complications developing months or years after treatment [late radiation tissue injury (LRTI)] will significantly affect between 5 and 15% of those long-term survivors, although incidence varies widely with dose, age, and site. LRTI is most common in the head and neck, chest wall, breast, and pelvis.

PATHOLOGY AND CLINICAL COURSE

With time, tissues undergo a progressive deterioration characterized by a reduction in the density of small blood vessels (reduced vascularity) and the replacement of normal tissue with dense fibrous tissue (fibrosis). Ultimately, and often triggered by a further physical insult such as surgery or infection, there may be insufficient oxygen to sustain normal function, and the tissue becomes necrotic (radiation necrosis). LRTI may be life-threatening and significantly reduce quality of life. Historically, the management of these injuries has been unsatisfactory. Conservative treatment is usually restricted to symptom management, while definitive treatment traditionally entails surgery to remove the affected part and extensive repair. Surgical intervention in an irradiated field is often disfiguring and associated with an increased incidence of delayed healing, breakdown of a surgical wound, or infection. HBO_2T may act by several mechanisms to improve this situation, including edema reduction, vasculogenesis, and enhancement of macrophage activity (Fig. e52-3). The intermittent application of HBO_2 is the only intervention shown to increase the microvascular density in irradiated tissue.

CLINICAL EVIDENCE

The typical course of HBO_2T consists of 30 once-daily compressions to 2–2.4 ATA for 1.5 to 2 hours each session. This course is often bracketed around surgical intervention if required. While HBO_2T has been used for LRTI since at least 1975, most clinical studies have been limited to small case series or individual case reports. In a recent semi-quantitative review, Feldmeier and Hampson located 71 such reports involving a total of 1193 patients across eight different tissues. There were clinically significant improvements in the majority of patients and only 7 of 71 reports indicated a generally poor response to HBO_2T. A recent Cochrane systematic review with meta-analysis included six randomized trials published since 1985 and drew the following conclusions (see Table e52–2 for numbers needed to treat): HBO_2T improves healing in radiation proctitis [relative risk (RR) of healing with HBO_2T 2.7, 95% confidence interval (CI) 1.2–6] and following hemimandibulectomy and reconstruction of the mandible (RR 1.4, CI 1.1–1.8); HBO_2T improves the probability of achieving mucosal coverage (RR 1.4, CI 1.2–1.6) and the restoration of bony continuity with osteoradionecrosis (ORN) (RR 1.4, CI 1.1–1.8); HBO_2T prevents the development of ORN following tooth extraction from a radiation field (RR 1.4, CI 1.08–1.7); and reduces the risk of wound dehiscence following grafts and flaps in the head and neck (RR 4.2, CI 1.1–16.8). Conversely, there was no evidence of benefit in established radiation brachial plexus lesions or brain injury.

Selected Problem Wounds

A problem wound is any cutaneous ulceration that requires a prolonged time to heal, does not heal, or recurs. In general, wounds referred to hyperbaric facilities are those where sustained attempts to heal by other means have failed. Problem wounds are common and constitute a significant health problem. It has been estimated that 1% of the population of industrialized countries will experience a leg ulcer at some time. The global cost of chronic wound care may be as high as $25 billion US per year.

PATHOLOGY AND CLINICAL COURSE

By definition, chronic wounds are indolent or progressive and resistant to the wide array of treatments applied. While there are many contributing factors, most commonly these wounds arise in association with one or more comorbidities such as diabetes, peripheral venous or arterial disease, or prolonged pressure (decubitus ulcers). First-line treatments are aimed at correction of the underlying pathology (e.g., vascular reconstruction, compression bandaging, or normalization of blood glucose level) and HBO_2T is an adjunctive therapy that may be added to good general wound care practice in order to maximize the chance of healing.

For most indolent wounds hypoxia is a major contributor to failure to heal. Many guidelines to patient selection for HBO_2T include the interpretation of transcutaneous oxygen tensions around the wound while breathing air and oxygen at pressure **(Fig. e52-4)**. Wound healing is a complex and incompletely understood process. While it appears that in acute wounds healing is stimulated by the initial hypoxia, low pH, and high lactate concentrations found in freshly injured tissue, some elements of tissue repair are extremely oxygen dependent, for example, collagen elaboration and deposition by fibroblasts, and bacterial killing by macrophages. In this complicated interaction between wound hypoxia and peri-wound oxygenation, successful healing relies on adequate tissue oxygenation in the area surrounding the fresh wound. Certainly, wounds that lie in hypoxic tissue beds are those that most often display poor or absent healing. Some causes of tissue hypoxia will be reversible with HBO_2T, while some will not (e.g., in the presence of severe large vessel disease). When tissue hypoxia can be overcome by a high driving pressure of oxygen in the arterial blood, this can be demonstrated by measuring the tissue partial pressure of oxygen using an implantable oxygen electrode or more commonly, a modified transcutaneous Clarke electrode.

Figure e52-4

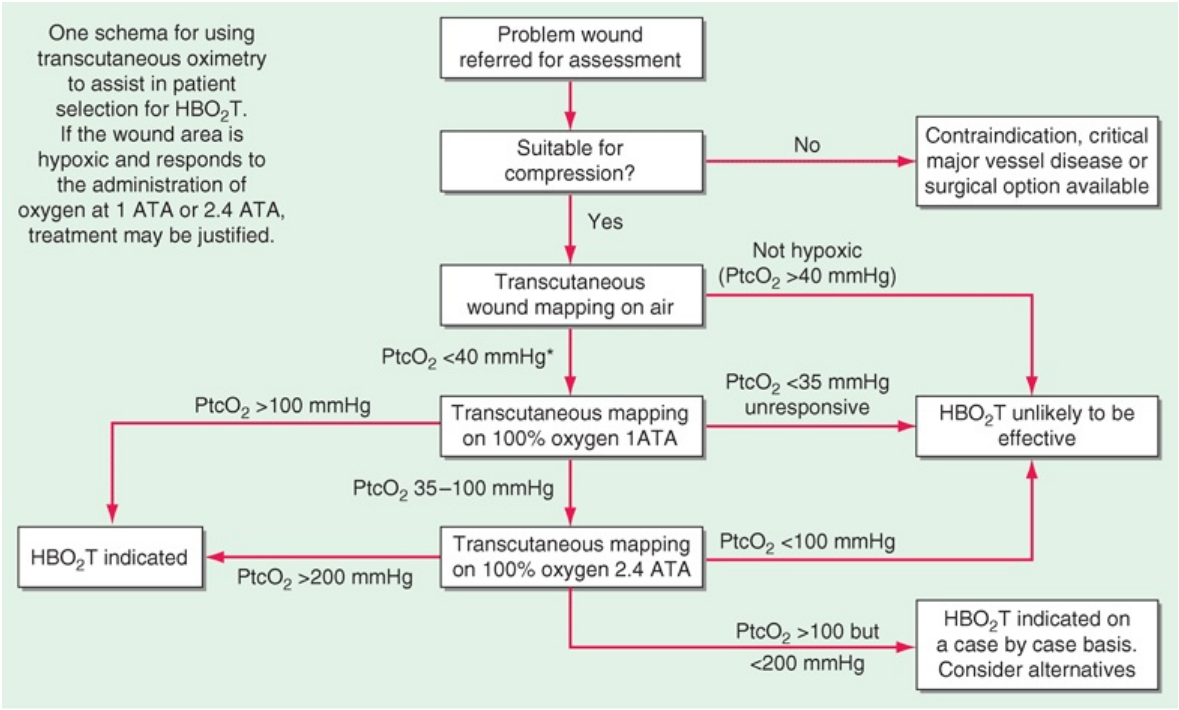

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Suitability for hyperbaric oxygen therapy (HBO_2T) guided by transcutaneous oximetry around the wound bed.

*In diabetic patients <50 mmHg may be more appropriate. $PtcO_2$, transcutaneous oxygen pressure.

The intermittent presentation of oxygen to those hypoxic tissues facilitates a resumption of healing (Fig. e52-3). As discussed above, these short exposures to high oxygen tensions have long-lasting effects (at least 24 hours) on a wide range of healing processes. The result is a gradual improvement in oxygen tension around the wound that reaches a plateau in experimental studies at about 20 treatments over 4 weeks. Improvements in oxygenation are associated with an eight- to ninefold increase in vascular density over both normobaric oxygen and air-breathing controls.

CLINICAL EVIDENCE

The typical course of HBO_2T consists of 20 to 30 once-daily compressions to 2–2.4 ATA for 1.5 to 2 hours each session, but is highly dependent on the clinical response. There are many case series in the literature supporting the use of HBO_2T for a wide range of problem wounds. Both retrospective and prospective cohort studies suggest that 6 months after a course of therapy about 70% of indolent ulcers will be healed or nearly so. Often the mean period such ulcers have been present is many months or years, suggesting that the application of HBO_2T has a profound effect, either primarily or as a facilitator of other strategies. A Cochrane review included five randomized controlled trials (RCTs) and concluded that HBO_2T reduces the rate of major amputation in people who have chronic foot ulcers as a result of diabetes [the RR of amputation with HBO_2T was 0.31 (95% CI 0.13–0.71) and suggests the number needed to receive HBO_2T to avoid one major amputation is 4 (95% CI 3–11)]. Randomized data on wound healing are notably lacking. Only one small trial reported this outcome and although there was a trend toward better outcomes with HBO_2T, the result is not statistically convincing (78% healed with HBO_2T versus 44% with sham).

Carbon Monoxide Poisoning

Carbon monoxide (CO) is a colorless, odorless gas formed during incomplete hydrocarbon combustion. While CO is an essential endogenous neurotransmitter linked to NO metabolism and activity, it is also a leading cause of poisoning death and in the United States alone, results in more than 50,000 emergency department visits per year and about 2000 deaths. While there are large variations from country to country, about half of nonlethal exposures are due to self-harm. Accidental poisoning is commonly associated with defective or improperly installed heaters, house fires, and industrial exposures. The motor vehicle is by far the most common source of intentional poisoning.

PATHOLOGY AND CLINICAL COURSE

The pathophysiology of carbon monoxide exposure is incompletely understood. CO binds to hemoglobin with an affinity more than 200 times that of oxygen, and this not only directly reduces the oxygen-carrying capacity of blood, but further promotes tissue hypoxia by shifting the oxyhemoglobin dissociation curve to

the left. CO is also an anesthetic agent that inhibits evoked responses and narcotizes experimental animals in a dose-dependent manner. The associated loss of airway patency together with reduced oxygen carriage in blood may cause death from acute arterial hypoxia in severe poisoning. CO may also cause harm by other mechanisms including direct disruption of cellular oxidative processes, binding to myoglobin and hepatic cytochromes, and peroxidation of brain lipids.

The brain and heart are the most sensitive target organs due to their high blood flow, poor tolerance of hypoxia, and high oxygen requirements. Minor exposure may be asymptomatic or present with vague constitutional symptoms such as headache, lethargy, and nausea, while higher doses may present with poor concentration and cognition, short-term memory loss, confusion, seizures, and loss of consciousness. While carboxyhemoglobin (COHb) levels on admission do not necessarily reflect the severity or the prognosis of CO poisoning, cardiorespiratory arrest carries a very poor prognosis. Over the longer term, surviving patients commonly have neuropsychological sequelae that may present days to months after poisoning. Motor disturbances, peripheral neuropathy, hearing loss, vestibular abnormalities, dementia, and psychosis have all been reported. Risk factors for poor outcome are age >35 years, exposure for >24 hours, acidosis, and loss of consciousness.

CLINICAL EVIDENCE

The typical course of HBO_2T consists of two to three compressions to 2–2.4 ATA for 1.5 to 2 hours each session. It is common for the first two compressions to be delivered within 24 hours of the exposure. CO poisoning is one of the longest-standing indications for HBO_2T—based largely on the obvious connection between exposure, tissue hypoxia, and the ability of HBO_2T rapidly to overcome hypoxemia. CO is eliminated rapidly via the lungs on application of HBO_2T, with a half-life of about 21 minutes at 2.0 ATA versus 5.5 hours breathing air and 71 minutes breathing oxygen at sea level. In practice, however, it seems unlikely that HBO_2T can be delivered in time to prevent either acute hypoxic death or irreversible global cerebral hypoxic injury. If HBO_2T is beneficial in CO poisoning, it must reduce the likelihood of persisting and/or delayed neurocognitive deficit through a mechanism other than the simple reversal of arterial hypoxia due to high levels of COHb. The difficulty in accurately assessing neurocognitive deficit has been one of the primary sources of controversy surrounding the clinical evidence in this area. To date there have been six randomized controlled trials of HBO_2T for CO poisoning, although only four have been reported in full. While a Cochrane review suggested that overall there is insufficient evidence to confirm a beneficial effect of HBO_2T on the chance of persisting neurocognitive deficit following poisoning [34% of patients treated with oxygen at 1 atmosphere versus 29%, of those treated with HBO_2T, odds ratio (OR) 0.78; 95% CI 0.54–1.1], this may have more to do with poor reporting and inadequate follow-up than with evidence that HBO_2T is not effective. The interpretation of the literature has much to do with how one defines neurocognitive deficit. In the most methodologically rigorous of these studies (Weaver et al.), a professionally administered battery of validated neuropsychological tests and a definition based on the deviation of individual subtest scores from the age-adjusted normal values was employed; if the patient complained of memory, attention, or concentration difficulties, the required decrement was decreased. Using this approach, 6 weeks after poisoning, 46% of patients treated with normobaric oxygen alone had cognitive sequelae compared to 25% of those who received HBO_2T [p = .007, number needed to treat (NNT) 5, 95% CI 3–16]. At 12 months the difference remained significant (32% versus 18%, p = .04, NNT 7, 95% CI 4–124) despite considerable loss to follow-up.

On this basis, HBO_2T remains widely advocated for the routine treatment of patients with moderate to severe poisoning—in particular in those older than 35 years, presenting with a metabolic acidosis on arterial blood-gas analysis, exposed for lengthy periods or with a history of unconsciousness.

DIVING MEDICINE

Introduction

Underwater diving is both a popular recreational activity and a means of employment in a range of tasks from underwater construction to military operations. It is a complex activity with unique hazards and medical complications arising mainly as a consequence of the dramatic changes in pressure associated with both descent and ascent through the water column. For every 10.13-m increase in depth of seawater, the ambient pressure (P_{amb}) increases by 1 standard atmosphere (101.3 kPa) so that a diver at 20 m depth is exposed to a P_{amb} of approximately 3 atmospheres absolute (ATA): 1 atm due to atmospheric pressure and 2 atm generated by the water column.

Breathing Equipment

Most diving is undertaken using a self-contained underwater breathing apparatus (scuba) consisting of one or more cylinders of compressed gas connected to a pressure-reducing regulator and a demand valve activated by inspiratory effort. Alternative sources of compressed gas include "rebreathers" that are closed or semiclosed circle systems with a carbon dioxide scrubber and externally supplied systems (via hoses from the surface), which are common in occupational applications. All of these systems must supply gas to the diver at P_{amb}, otherwise inspiration against the surrounding water pressure would be impossible. For most recreational diving the respired gas is air. Pure oxygen is rarely used because

oxygen may provoke seizures above an inspired P_{O2} of 1.6 ATA in aquatic environments, limiting the practical safe depth to 6 m. For the same reason, very deep diving requires the use of oxygen fractions <0.21 (at 66 m, air contains 1.6 ATA of oxygen). Deep-diving gas often includes helium instead of nitrogen to reduce both the narcotic effect and high gas density that results from breathing nitrogen at high pressures. As with HBO_2T, therefore, medical problems associated with diving may arise from either changing pressure or the properties of the breathing gas itself. The following discussion is limited to consideration of common or significant barotraumas and DCS, although a range of other problems are relevant including salt-water aspiration, marine envenomation, and the high-pressure neurologic syndrome (HPNS).

Barotrauma (See Also "Adverse Effects of Therapy")

Difficulties with equalizing pressure in the middle ear are exaggerated underwater by the rapidity with which pressures can change as a diver descends or ascends. Failure to periodically insufflate the middle-ear spaces via the eustachian tubes during descent results in increasing pain. As the P_{amb} increases, the tympanic membrane (TM) may be bruised or even ruptured as it is pushed inward. Negative pressure in the middle ear results in engorgement of blood vessels in the mucous membranes and leads to effusion or bleeding, which can be associated with a *conductive* hearing loss. Middle-ear barotrauma is much less common during ascent because expanding gas in the middle-ear space tends to open the eustachian tube easily and "automatically." Barotrauma may also affect the respiratory sinuses, although the sinus ostia are usually widely patent and allow automatic pressure equalization without the need for specific maneuvers. If equalization fails, pain usually results in termination of the dive. Difficulty with equalizing ears or sinuses may respond to oral or nasal decongestants.

Much less commonly the inner ear may suffer barotrauma. Several explanations have been proposed, of which the most favored holds that forceful attempts to insufflate the middle-ear space during descent cause sudden transmission of pressure to the perilymph via the cochlear aqueduct and outward rupture of the round window. The clinician should be alerted to possible inner ear barotrauma after diving by a *sensorineural* hearing loss or true vertigo (which is often accompanied by nausea, vomiting, nystagmus, and ataxia). These manifestations can also occur in vestibulocochlear DCS (see below) but should never be attributed to middle-ear barotrauma. Immediate review by an expert diving physician is recommended and urgent referral to an otologist will usually follow.

The lungs are also vulnerable to barotrauma but are at most risk during ascent. If expanding gas becomes trapped in the lungs as P_{amb} falls, this may rupture alveoli and associated vascular tissue. Gas trapping may occur if divers intentionally or involuntarily hold their breath during ascent or if there are bullae. The extent to which asthma predisposes to pulmonary barotrauma is debated, but the presence of active bronchoconstriction must increase risk. Asthmatics who regularly require bronchodilator medications are usually discouraged from diving for this reason. While possible consequences of pulmonary barotrauma include pneumothorax and mediastinal emphysema, the most feared is the introduction of gas into the pulmonary veins leading to cerebral arterial gas embolism (CAGE). Manifestations of CAGE include loss of consciousness, confusion, hemiplegia, and speech and visual disturbances appearing immediately or within minutes after surfacing. The management is the same as that for DCS described below. It is notable that the natural history of CAGE often includes substantial or complete resolution of symptoms early after the event. This is probably the clinical correlate of bubble involution and redistribution with consequent restoration of flow. Patients exhibiting such remissions should still be reviewed at specialist diving medical centers because secondary deterioration or reembolization can occur. Unsurprisingly, these events can be misdiagnosed as strokes or transient ischemic attacks (TIAs) when patients are seen by those unfamiliar with diving medicine. *All patients presenting with neurologic symptoms after diving should have their symptoms discussed with a specialist in diving medicine and be considered for recompression therapy.*

Decompression Sickness

DCS is caused by the formation of bubbles from dissolved inert gas (usually nitrogen) during or after ascent (decompression) from a compressed gas dive. Bubble formation is also possible following decompression for extravehicular activity during space flight and with ascent to altitude in unpressurized aircraft. DCS in the latter scenarios is probably rare in comparison with diving, where the incidence is approximately 1:10,000 recreational dives.

Breathing at elevated P_{amb} results in increased uptake of inert gas into blood and then into tissues. The rate at which tissue inert gas equilibrates with the inspired inert gas pressure is proportional to tissue blood flow and the blood-tissue partition coefficient for the gas, and inversely proportional to the tissue volume. Similar factors dictate the kinetics of inert gas wash-out during ascent. If the rate of gas wash-out from tissues does not match the rate of decline in P_{amb}, then the sum of dissolved gas pressures in the tissue will exceed P_{amb}, a condition referred to as "supersaturation." This is the prerequisite for bubbles to form during decompression, though other poorly understood factors are involved. Deeper and longer dives result in greater inert gas absorption and greater likelihood of tissue supersaturation during ascent. Divers control their ascent for a given depth and time exposure using algorithms that often include periods where ascent is halted for a prescribed period at different depths to allow time for gas wash-out ("decompression stops"). While a breach of these protocols increases the risk of DCS,

adherence does not guarantee against it. DCS should be considered in any diver manifesting symptoms not readily explained by an alternative mechanism.

Bubbles may form within tissues themselves, where they cause symptoms by mechanical distraction of pain-sensitive or functionally important structures. They also appear in the venous circulation as blood drains from supersaturated tissues. Some venous bubbles can be tolerated without symptoms and are efficiently filtered from the circulation in the pulmonary capillaries. In larger numbers these bubbles are capable of inciting inflammatory and coagulation cascades, damaging endothelium, activating formed elements of blood such as platelets, and causing symptomatic pulmonary vascular obstruction. Moreover, if there is a right-to-left shunt [such as through a patent foramen ovale (PFO)], venous bubbles may enter the arterial circulation (25% of adults have a probe-patent PFO). The risk of cerebral, spinal cord, inner ear, and skin manifestations appears higher in the presence of significant shunts, suggesting that these "arterialized" venous bubbles can cause harm, perhaps by disrupting flow in the microcirculation of target organs.

Table e52–3 lists manifestations of DCS grouped according to organ system. The majority of cases present with mild symptoms including musculoskeletal pain, fatigue, and minor neurologic manifestations such as patchy paresthesias. Serious presentations are much less common. Nevertheless, pulmonary and cardiovascular manifestations can be life-threatening, and spinal cord involvement frequently results in permanent disability. Latency is variable. Serious DCS usually has its onset within minutes of surfacing, but mild manifestations may not appear for several hours. Symptoms arising more than 24 hours after diving are very unlikely to be DCS. The presentation may be confusing and nonspecific and there are no useful diagnostic investigations. Diagnosis is based on an integration of findings from examination of the dive profile, the nature and temporal relationship of symptoms, and the clinical examination. Some DCS presentations may be difficult to separate from CAGE following pulmonary barotrauma, but the distinction is unimportant since the first aid and definitive management of both conditions are the same.

Table e52–3 Manifestations of Decompression Sickness

Organ System	Manifestations
Musculoskeletal	Limb pain
Neurologic	
Cerebral	Confusion
	Visual disturbances
	Speech disturbances
Spinal	Muscular weakness
	Paralysis
	Upper motor neuron signs
	Bladder and sphincter dysfunction
	Dermatomal sensory disturbances
	Abdominal pain
	Girdle pain
Vestibulocochlear	Hearing loss
	Vertigo and ataxia
	Nausea and vomiting
Peripheral	Patchy nondermatomal sensory disturbance
Pulmonary	Cough
	Dyspnea
	Pulmonary edema (rare)
Cardiovascular	Chest pain
	Arrhythmia
	Hemoconcentration
	Coagulopathy
	Hypotension
Cutaneous	Rash, itch
Lymphatic	Soft tissue edema, often relatively localized
Constitutional	Fatigue and malaise

Treatment: Diving Medicine

First aid includes horizontal positioning to reduce the preferential entry of bubbles into the cerebral circulation (especially in suspected CAGE), intravenous fluids if available, and sustained 100% oxygen administration. The latter accelerates inert gas wash-out from tissues and promotes resolution of bubbles. Definitive treatment of DCS or CAGE with recompression and hyperbaric oxygen is justified in most instances, though some very mild or marginal DCS cases may be managed with first aid strategies if evacuation is difficult or hazardous. Evacuations need to avoid further decompression stress and, if over long distance, are usually undertaken using a helicopter flying at low altitude or a fixed-wing air ambulance pressurized to 1 ATA.

Recompression reduces bubble volume in accordance with Boyle's law and increases the inert gas partial pressure difference between bubble and surrounding tissue. At the same time, oxygen administration markedly increases the inert gas partial pressure difference between alveoli and tissue. The combined effect is to significantly increase the rate of inert gas diffusion from the bubble. Hyperbaric oxygen helps oxygenate compromised tissues and appears to ameliorate some of the pro-inflammatory effects of bubbles (Fig. e52-3). It is likely that recompression as early as possible will result in the best outcomes, especially for more serious DCS. Various recompression protocols have been advocated, but there are no data that define the optimum approach. Recompression typically begins with oxygen administered at 2.8 ATA, the maximum pressure at which the risk of oxygen toxicity remains acceptable in a hyperbaric chamber, followed by a stepwise decompression over variable periods adjusted to symptom response. The most widely used algorithm is the US Navy Treatment Table 6, whose shortest format lasts 4 hours and 45 minutes. This is often followed by daily HBO_2T using shorter treatment tables while symptoms persist and appear responsive to treatment. Adjuncts to recompression include intravenous fluids and other supportive care as necessary. Provided recovery is complete, diving can usually be resumed after a stand-down period of at least 1 month. If there are residual symptoms, diving is discouraged. If CAGE secondary to pulmonary barotrauma is suspected, steps must be taken to exclude the presence of an underlying pulmonary predisposition to further events. Investigation for a PFO may be indicated after cerebral, spinal, inner ear, or cutaneous DCS, especially if the event occurred despite clear adherence to dive-planning algorithms. Routine screening of all diving candidates for PFO is not justified given the high prevalence of PFO and the low incidence of significant DCS.

DCS is a variable and potentially confusing condition with many potential pitfalls in diagnosis and management. Clinicians untrained in diving medicine and faced with a patient who is unwell after diving, even with apparently mild symptoms, are strongly advised to contact an appropriate authority for diagnostic and treatment advice at an early stage.

FURTHER READINGS

Bennett MH et al: Hyperbaric oxygen therapy for late radiation tissue injury. Cochrane Database Syst Rev 2, 2009Brubakk A, Neuman T (eds): *Bennett and Elliott's Physiology and Medicine of Diving,* 5th ed. Philadelphia, Saunders Elsevier, 2003

Clarke RE et al: Hyperbaric oxygen treatment of chronic refractory radiation proctitis: A randomized and controlled double-blind crossover trial with long-term follow-up. Int J Radiat Oncol Biol Phys 72:134, 2008Feldmeier JJ, Hampson NB: A systematic review of the literature reporting the application of hyperbaric oxygen prevention and treatment of delayed radiation injuries: An evidence based approach. Undersea and Hyperbaric Medicine 29:4, 2002

Gesell L (ed): *Hyperbaric Oxygen Therapy Indications,* 12th ed. Durham, Undersea and Hyperbaric Medical Society Publications, 2008

Kranke P et al: Hyperbaric oxygen therapy for chronic wounds. Cochrane Database Syst Rev 2:CD004123, 2004

Neuman t, Thom S (eds): *Physiology and Medicine of Hyperbaric Oxygen Therapy* . Philadelphia, Saunders Elsevier, 2008

Vann RD et al: Decompression illness. Lancet 377:153, 2011[PMID: 21215883] [Full Text]

Weaver LK et al: Hyperbaric oxygen for acute carbon monoxide poisoning. N Engl J Med 347:1057, 2002[PMID: 12362006] [Full Text]

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

APPENDIX: LABORATORY VALUES OF CLINICAL IMPORTANCE

This Appendix contains tables of reference values for laboratory tests, special analytes, and special function tests. A variety of factors can influence reference values. Such variables include the population studied, the duration and means of specimen transport, laboratory methods and instrumentation, and even the type of container used for the collection of the specimen. The reference or "normal" ranges given in this appendix may therefore not be appropriate for all laboratories, and these values should only be used as general guidelines. Whenever possible, reference values provided by the laboratory performing the testing should be utilized in the interpretation of laboratory data. Values supplied in this Appendix reflect typical reference ranges in adults. Pediatric reference ranges may vary significantly from adult values.

In preparing the Appendix, the authors have taken into account the fact that the system of international units (SI, système international d'unités) is used in most countries and in some medical journals. However, clinical laboratories may continue to report values in "traditional" or conventional units. Therefore, both systems are provided in the Appendix. The dual system is also used in the text except for (1) those instances in which the numbers remain the same but only the terminology is changed (mmol/L for meq/L or IU/L for mIU/mL), when only the SI units are given; and (2) most pressure measurements (e.g., blood and cerebrospinal fluid pressures), when the traditional units (mmHg, mmH_2O) are used. In all other instances in the text the SI unit is followed by the traditional unit in parentheses.

REFERENCE VALUES FOR LABORATORY TESTS

Table 1 Hematology and Coagulation

Analyte	Specimen	SI Units	Conventional Units
Activated clotting time	WB	70–180 s	70–180 s
Activated protein C resistance (factor V Leiden)	P	Not applicable	Ratio >2.1
ADAMTS13 activity	P	≥0.67	≥67%
ADAMTS13 inhibitor activity	P	Not applicable	≤0.4 U
ADAMTS13 antibody	P	Not applicable	≤18 U
Alpha$_2$ antiplasmin	P	0.87–1.55	87–155%
Antiphospholipid antibody panel			
PTT-LA (lupus anticoagulant screen)	P	Negative	Negative
Platelet neutralization procedure	P	Negative	Negative
Dilute viper venom screen	P	Negative	Negative
Anticardiolipin antibody	S		
IgG		0–15 arbitrary units	0–15 GPL
IgM		0–15 arbitrary units	0–15 MPL
Antithrombin III	P		
Antigenic		220–390 mg/L	22–39 mg/dL
Functional		0.7–1.30 U/L	70–130 %
Anti-Xa assay (heparin assay)	P		
Unfractionated heparin		0.3–0.7 kIU/L	0.3–0.7 IU/mL
Low-molecular-weight heparin		0.5–1.0 kIU/L	0.5–1.0 IU/mL
Danaparoid (Orgaran)		0.5–0.8 kIU/L	0.5–0.8 IU/mL
Autohemolysis test	WB	0.004–0.045	0.4–4.50%
Autohemolysis test with glucose	WB	0.003–0.007	0.3–0.7%
Bleeding time (adult)		<7.1 min	<7.1 min
Bone marrow: See Table 7			
Clot retraction	WB	0.50–1.00/2 h	50–100%/2 h

Cryofibrinogen	P	Negative	Negative
D-dimer	P	220–740 ng/mL FEU	220–740 ng/mL FEU
Differential blood count	Relative counts:		
Neutrophils		0.40–0.70	40–70%
Bands		0.0–0.05	0–5%
Lymphocytes		0.20–0.50	20–50%
Monocytes		0.04–0.08	4–8%
Eosinophils		0.0–0.6	0–6%
Basophils		0.0–0.02	0–2%
Absolute counts:			
Neutrophils		1.42 –6.34x10^9/L	1420–6340/mm^3
Bands		0–0.45x10^9/L	0–450/mm^3
Lymphocytes		0.71 –4.53x10^9/L	710–4530/mm^3
Monocytes		0.14 –0.72x10^9/L	140–720/mm^3
Eosinophils		0–0.54x10^9/L	0–540/mm^3
Basophils WB		0–0.18x10^9/L	0–180/mm^3
Erythrocyte count	WB		
Adult males		4.30 –5.60x10^{12}/L	4.30–5.60x10^6/mm^3
Adult females		4.00 –5.20x10^{12}/L	4.00–5.20x10^6/mm^3
Erythrocyte life span	WB		
Normal survival		120 days	120 days
Chromium labeled, half-life ($t_{1/2}$)		25–35 days	25–35 days
Erythrocyte sedimentation rate	WB		
Females		0–20 mm/h	0–20 mm/h
Males		0–15 mm/h	0–15 mm/h
Euglobulin lysis time	P	7200–14400 s	120–240 min
Factor II, prothrombin	P	0.50–1.50	50–150%
Factor V	P	0.50–1.50	50–150%
Factor VII	P	0.50–1.50	50–150%
Factor VIII	P	0.50–1.50	50–150%
Factor IX	P	0.50–1.50	50–150%
Factor X	P	0.50–1.50	50–150%
Factor XI	P	0.50–1.50	50–150%
Factor XII	P	0.50–1.50	50–150 %
Factor XIII screen	P	Not applicable	Present
Factor inhibitor assay	P	<0.5 Bethesda Units	<0.5 Bethesda Units
Fibrin(ogen) degradation products	P	0–1 mg/L	0–1 μg/mL
Fibrinogen	P	2.33–4.96 g/L	233–496 mg/dL
Glucose-6-phosphate dehydrogenase (erythrocyte)	WB	<2400 s	<40 min
Ham's test (acid serum)	WB	Negative	Negative
Hematocrit	WB		
Adult males		0.388–0.464	38.8–46.4

Adult females		0.354–0.444	35.4–44.4
Hemoglobin			
Plasma	P	6–50 mg/L	0.6–5.0 mg/dL
Whole blood:	WB		
Adult males		133–162 g/L	13.3–16.2 g/dL
Adult females		120–158 g/L	12.0–15.8 g/dL
Hemoglobin electrophoresis	WB		
Hemoglobin A		0.95–0.98	95–98%
Hemoglobin A_2		0.015–0.031	1.5–3.1%
Hemoglobin F		0–0.02	0–2.0%
Hemoglobins other than A, A_2, or F		Absent	Absent
Heparin-induced thrombocytopenia antibody	P	Negative	Negative
Immature platelet fraction (IPF)	WB	0.011–0.061	1.1–6.1%
Joint fluid crystal	JF	Not applicable	No crystals seen
Joint fluid mucin	JF	Not applicable	Only type I mucin present
Leukocytes			
Alkaline phosphatase (LAP)	WB	0.2–1.6 μkat/L	13–100 μ/L
Count (WBC)	WB	3.54 –9.06x10^9/L	3.54–9.06x10^3/mm^3
Mean corpuscular hemoglobin (MCH)	WB	26.7–31.9 pg/cell	26.7–31.9 pg/cell
Mean corpuscular hemoglobin concentration (MCHC)	WB	323–359 g/L	32.3–35.9 g/dL
Mean corpuscular hemoglobin of reticulocytes (CH)	WB	24–36 pg	24–36 pg
Mean corpuscular volume (MCV)	WB	79–93.3 fL	79–93.3 μm^3
Mean platelet volume (MPV)	WB	9.00–12.95 fL	9.00–12.95
Osmotic fragility of erythrocytes	WB		
Direct		0.0035–0.0045	0.35–0.45%
Indirect		0.0030–0.0065	0.30–0.65%
Partial thromboplastin time, activated	P	26.3–39.4 s	26.3–39.4 s
Plasminogen	p		
Antigen		84–140 mg/L	8.4–14.0 mg/dL
Functional		0.70–1.30	70–130%
Plasminogen activator inhibitor 1	P	4–43 μg/L	4–43 ng/mL
Platelet aggregation	PRP	Not applicable	>65% aggregation in response to adenosine diphosphate, epinephrine, collagen, ristocetin, and arachidonic acid
Platelet count	WB	165–415x10^9/L	165–415x10^3/mm^3
Platelet, mean volume	WB	6.4–11 fL	6.4–11.0 μm^3
Prekallikrein assay	P	0.50–1.5	50–150%
Prekallikrein screen	P		No deficiency detected
Protein C	P		
Total antigen		0.70–1.40	70–140%
Functional		0.70–1.30	70–130%
Protein S	P		
Total antigen		0.70–1.40	70–140%
Functional		0.65–1.40	65–140%
Free antigen P		0.70–1.40	70–140%

Prothrombin gene mutation G20210A	WB	Not applicable	Not present
Prothrombin time	P	12.7–15.4 s	12.7–15.4 s
Protoporphyrin, free erythrocyte	WB	0.28–0.64 μmol/L of red blood cells	16–36μg/dL of red blood cells
Red cell distribution width	WB	<0.145	<14.5%
Reptilase time	P	16–23.6 s	16–23.6 s
Reticulocyte count	WB		
Adult males		0.008–0.023 red cells	0.8–2.3% red cells
Adult females		0.008–0.020 red cells	0.8–2.0% red cells
Reticulocyte hemoglobin content	WB	>26 pg/cell	>26 pg/cell
Ristocetin cofactor (functional von Willebrand factor)	P		
Blood group O		0.75 mean of normal	75% mean of normal
Blood group A		1.05 mean of normal	105% mean of normal
Blood group B		1.15 mean of normal	115% mean of normal
Blood group AB P		1.25 mean of normal	125% mean of normal
Serotonin release assay	S	<0.2 release	<20% release
Sickle cell test	WB	Negative	Negative
Sucrose hemolysis	WB	<0.1	<10% hemolysis
Thrombin time	P	15.3–18.5 s	15.3–18.5 s
Total eosinophils	WB	150–300x10^6/L	150–300/mm^3
Transferrin receptor	S, P	9.6–29.6 nmol/L	9.6–29.6 nmol/L
Viscosity	P		
Plasma	p	1.7–2.1	1.7–2.1
Serum	S	1.4–1.8	1.4–1.8
von Willebrand factor (vWF) antigen (factor VIII:R antigen)			
Blood group O		0.75 mean of normal	75% mean of normal
Blood group A		1.05 mean of normal	105% mean of normal
Blood group B		1.15 mean of normal	115% mean of normal
Blood group AB P		1.25 mean of normal	125% mean of normal
von Willebrand factor multimers	P	Normal distribution	Normal distribution
White blood cells: see "Leukocytes"			

Abbreviations: JF, joint fluid; P, plasma; PRP, platelet-rich plasma; S, serum; WB, whole blood.

Table 2 Clinical Chemistry and Immunology

Analyte	Specimen*	SI Units	Conventional Units
Acetoacetate	P	49–294 μmol/L	0.5–3.0 mg/dL
Adrenocorticotropin (ACTH)	P	1.3–16.7 pmol/L	6.0–76.0 pg/mL
Alanine aminotransferase (ALT, SGPT)	S	0.12–0.70 μkat/L	7–41 U/L
Albumin	S	40–50 g/L	4.0–5.0 mg/dL
Aldolase	S	26–138 nkat/L	1.5–8.1 U/L
Aldosterone (adult)			

Supine, normal sodium diet	S, P	<443 pmol/L	<16 ng/dL
Upright, normal	S, P	111–858 pmol/L	4–31 ng/dL
Alpha fetoprotein (adult)	S	0–8.5 μg/L	0–8.5 ng/mL
Alpha$_1$ antitrypsin	S	1.0–2.0 g/L	100–200 mg/dL
Ammonia, as NH_3	P	11–35 μmol/L	19–60 μg/dL
Amylase (method dependent)	S	0.34–1.6 μkat/L	20–96 U/L
Androstendione (adult)	S		
Males		0.81–3.1 nmol/L	23–89 ng/dL
Females			
Premenopausal		0.91–7.5 nmol/L	26–214 ng/dL
Postmenopausal		0.46–2.9 nmol/L	13–82 ng/dL
Angiotensin-converting enzyme (ACE)	S	0.15–1.1 μkat/L	9–67 U/L
Anion gap	S	7–16 mmol/L	7–16 mmol/L
Apolipoprotein A-1	S	0.94–1.78 g/L	94–178 mg/dL
Male		1.01–1.99 g/L	101–199 mg/dL
Female			
Apolipoprotein B	S	0.55–1.40 g/L	55–140 mg/dL
Male		0.55–1.25 g/L	55–125 mg/dL
Female			
Arterial blood gases	WB	22–30 mmol/L	22–30 meq/L
$[HCO_3^-]$		4.3–6.0 kPa	32–45 mmHg
PCO_2		7.35–7.45	7.35–7.45
pH		9.6–13.8 kPa	72–104 mmHg
PO_2			
Aspartate aminotransferase (AST, SGOT)	S	0.20–0.65 μkat/L	12–38 U/L
Autoantibodies	S	≤29 AU/mL	29 AU/mL
Anti-centromere antibody IgG		<25 IU/L	<25 IU/L
Anti-double-strand (native) DNA		Negative	Negative
Anti-glomerular basement membrane antibodies		≥19 AU/mL	>19 AU/mL
Qualitative IgG, IgA		<1.0 U	<1.0 U
Quantitative IgG antibody		>29 AU/mL	>29 AU/mL
Anti-histone antibodies		Not applicable	<20 Units
Anti-Jo-1 antibody		Not applicable	<1:20
Anti-mitochondrial antibody		>19 AU/mL	>19 AU/mL
Anti-neutrophil cytoplasmic autoantibodies		>19 AU/mL	>19 AU/mL
Serine proteinase 3 antibodies		Not applicable	Negative at 1:40
Myeloperoxidase antibodies		Not applicable	None detected
Antinuclear antibody		Not applicable	<1.0 U
Anti-parietal cell antibody		Not applicable	<1.0 U
Anti-RNP antibody		Not applicable	<1.0 U
Anti-Scl 70 antibody		Not applicable	<1.0 U
Anti-Smith antibody		Not applicable	<1.0 U
		Not applicable	Negative
		<40 KIU/L	<40 IU/mL
		<35 KIU/L	<35 IU/mL

Anti–smooth muscle antibody Anti-SSA antibody Anti-SSB antibody Anti-thyroglobulin antibody Anti-thyroid peroxidase antibody			
B-type natriuretic peptide (BNP)	P	Age and gender specific: <100 ng/L	Age and gender specific: <100 pg/mL
Bence Jones protein, serum qualitative	S	Not applicable	None detected
Bence Jones protein, serum quantitative Free kappa Free lambda K/L ratio	S	3.3–19.4 mg/L 5.7–26.3 mg/L 0.26–1.65	0.33–1.94 mg/dL 0.57–2.63 mg/dL 0.26–1.65
Beta-2-microglobulin	S	1.1–2.4 mg/L	1.1–2.4 mg/L
Bilirubin Total Direct Indirect	S	5.1–22 μmol/L 1.7–6.8 μmol/L 3.4–15.2 μmol/L	0.3–1.3 mg/dL 0.1–0.4 mg/dL 0.2–0.9 mg/dL
C peptide	S	0.27–1.19 nmol/L	0.8–3.5 ng/mL
C1-esterase-inhibitor protein	S	210–390 mg/L	21–39 mg/dL
CA 125	S	<35 kU/L	<35 U/mL
CA 19-9	S	<37 kU/L	<37 U/mL
CA 15-3	S	<33 kU/L	<33 U/mL
CA 27-29	S	0–40 kU/L	0–40 U/mL
Calcitonin Male Female	S	0–7.5 ng/L 0–5.1 ng/L	0–7.5 pg/mL 0–5.1 pg/mL
Calcium	S	2.2–2.6 mmol/L	8.7–10.2 mg/dL
Calcium, ionized	WB	1.12–1.32 mmol/L	4.5–5.3 mg/dL
Carbon dioxide content (TCO_2)	P (sea level)	22–30 mmol/L	22–30 meq/L
Carboxyhemoglobin (carbon monoxide content) Nonsmokers Smokers Loss of consciousness and death	WB	0.0–0.015 0.04–0.09 >0.50	0–1.5% 4–9% >50%
Carcinoembryonic antigen (CEA) Nonsmokers Smokers	S	0.0–3.0 μg/L 0.0–5.0 μg/L	0.0–3.0 ng/mL 0.0–5.0 ng/mL
Ceruloplasmin	S	250–630 mg/L	25–63 mg/dL
Chloride	S	102–109 mmol/L	102–109 meq/L
Cholesterol: see Table 5			
Cholinesterase	S	5–12 kU/L	5–12 U/mL
Chromogranin A	S	0–50 μg/L	0–50 ng/mL
Complement C3 C4 Complement total	S	0.83–1.77 g/L 0.16–0.47 g/L 60–144 CAE units	83–177 mg/dL 16–47 mg/dL 60–144 CAE units
Cortisol	S	138–690 nmol/L	5–25μg/dL

Fasting, 8 A.M.–12 noon 12 noon–8 P.M. 8 P.M.–8 A.M.		138–414 nmol/L 0–276 nmol/L	5–15 μg/dL 0–10 μg/dL
C-reactive protein	S	<10 mg/L	<10 mg/L
C-reactive protein, high sensitivity	S	Cardiac risk Low: <1.0 mg/L Average: 1.0–3.0 mg/L High: >3.0 mg/L	Cardiac risk Low: <1.0 mg/L Average: 1.0–3.0 mg/L High: >3.0 mg/L
Creatine kinase (total) Females Males	S	0.66–4.0 μkat/L 0.87 5.0μkat/L	39–238 U/L 51–294 U/L
Creatine kinase-MB Mass Fraction of total activity (by electrophoresis)	S	0.0–5.5 μg/L 0–0.04	0.0–5.5 ng/mL 0–4.0%
Creatinine Female Male	S	44–80 μmol/L 53>106μmol/L	0.5–0.9 mg/dL 0.6–1.2 mg/dL
Cryoglobulins	S	Not applicable	None detected
Cystatin C	S	0.5–1.0 mg/L	0.5–1.0 mg/L
Dehydroepiandrosterone (DHEA) (adult) Male Female	S	6.2–43.4 nmol/L 4.5–34.0 nmol/L	180–1250 ng/dL 130–980 ng/dL
Dehydroepiandrosterone (DHEA) sulfate Male (adult) Female (adult, premenopausal) Female (adult, postmenopausal)	S	100–6190 μg/L 120–5350 μg/L 300–2600 μg/L	10–619 μg/dL 12–535 μg/dL 30–260 μg/dL
11-Deoxycortisol (adult)(compound S)	S	0.34–4.56 nmol/L	12–158 ng/dL
Dihydrotestosterone Male Female	S, P	1.03–2.92 nmol/L 0.14–0.76 nmol/L	30–85 ng/dL 4–22 ng/dL
Dopamine	P	0–130 pmol/L	0–20 pg/mL
Epinephrine Supine (30 min) Sitting Standing (30 min)	P	<273 pmol/L <328 pmol/L <491pmol/L	<50 pg/mL <60 pg/mL <90 pg/mL
Erythropoietin	S	4–27 U/L	4–27 U/L
Estradiol Female Menstruating: Follicular phase Midcycle peak Luteal phase Postmenopausal Male	S, P	74–532 pmol/L 411–1626 pmol/L 74–885 pmol/L 217 pmol/L 74 pmol/L	<20–145 pg/mL 112–443 pg/mL <20–241 pg/mL <59 pg/mL <20 pg/mL
Estrone Female Menstruating: Follicular phase	S, P	<555 pmol/L <740 pmol/L 11–118 pmol/L 33–133 pmol/L	<150 pg/mL <200 pg/mL 3–32 pg/mL 9–36 pg/mL

Luteal phase Postmenopausal Male			
Fatty acids, free (nonesterified)	P	0.1–0.6 mmol/L	2.8–16.8 mg/dL
Ferritin Female Male	S	10–150 μg/L 29 248μg/L	10–150 ng/mL 29–248 ng/mL
Follicle-stimulating hormone (FSH) Female Menstruating Follicular phase Ovulatory phase Luteal phase Postmenopausal Male	S, P	3.0–20.0 IU/L 9.0–26.0 IU/L 1.0–12.0 IU/L 18.0–153.0 IU/L 1.0–12.0 IU/L	3.0–20.0 mIU/mL 9.0–26.0 mIU/mL 1.0–12.0 mIU/mL 18.0–153.0 mIU/mL 1.0–12.0 mIU/mL
Fructosamine	S	<285 umol/L	<285 umol/L
Gamma glutamyltransferase	S	0.15–0.99 μkat/L	9–58 U/L
Gastrin	S	<100 ng/L	<100 pg/mL
Glucagon	P	40–130 ng/L	40–130 pg/mL
Glucose Glucose (fasting) Normal Increased risk for diabetes Diabetes mellitus	WB P	3.6–5.3 mmol/L 4.2–5.6 mmol/L 5.6–6.9 mmol/L Fasting >7.0 mmol/L A 2-hour level of >11.1 mmol/L during an oral glucose tolerance test A random glucose level of >11.1 mmol/L in patients with symptoms of hyperglycemia	65–95 mg/dL 75–100 mg/dL 100–125 mg/dL Fasting >126 mg/dL A 2-hour level of >200 mg/dL during an oral glucose tolerance test A random glucose level of >200 mg/dL in patients with symptoms of hyperglycemia
Growth hormone	S	0–5 μg/L	0–5 ng/mL
Hemoglobin A_{Ic} Pre-diabetes Diabetes mellitus	WB	0.04–0.06 HgB fraction 0.057-0.064 HgB fraction A hemoglobin A_{1c} level of >0.065 Hgb fraction as suggested by the American Diabetes Association	4.0–5.6% 5.7–6.4% A hemoglobin A_{1c} level of >6.5% as suggested by the American Diabetes Association
Hemoglobin A_{1c} with estimated average glucose (eAg)	WB	eAg mmoL/L = 1.59xHbA_{1c} – 2.59	eAg (mg/dL) = 28.7xHbA_{1c} – 46.7
High-density lipoprotein (HDL) (see Table 5)			
Homocysteine	P	4.4–10.8 μmol/L	4.4–10.8 μmol/L
Human chorionic gonadotropin (HCG) Nonpregnant female 1–2 weeks postconception 2–3 weeks postconception 3–4 weeks postconception 4–5 weeks postconception 5–10 weeks postconception 10–14 weeks post conception	S	<5 IU/L 9–130 IU/L 75–2600 IU/L 850–20,800 IU/L 4000–100,200 IU/L 11,500–289,000 IU/L 18,300–137,000 IU/L 1400–53,000 IU/L 940–60,000 IU/L	<5 mIU/ml 9–130 mIU/mL 75–2600 mIU/mL 850–20,800 mIU/mL 4000–100,200 mIU/mL 11,500–289,000 mIU/mL 18,300–137,000 mIU/mL 1400–53,000 mIU/mL 940–60,000 mIU/mL

Second trimester Third trimester			
>-Hydroxybutyrate	P	60–170 μmol/L	0.6–1.8 mg/dL
17-Hydroxyprogesterone (adult) Male Female Follicular phase Luteal phase	S	<4.17 nmol/L 0.45–2.1 nmol/L 1.05–8.7 nmol/L	<139 ng/dL 15–70 ng/dL 35–290 ng/dL
Immunofixation	S	Not applicable	No bands detected
Immunoglobulin, quantitation (adult) IgA IgD IgE IgG IgG_1 IgG_2 IgG_3 IgG_4 IgM	S S S S S S S S S	0.70–3.50 g/L 0–140 mg/L 1–87 KIU/L 7.0–17.0 g/L 2.7–17.4 g/L 0.3–6.3 g/L 0.13–3.2 g/L 0.11–6.2 g/L 0.50–3.0 g/L	70–350 mg/dL 0–14 mg/dL 1–87 IU/mL 700–1700 mg/dL 270–1740 mg/dL 30–630 mg/dL 13–320 mg/dL 11–620 mg/dL 50–300 mg/dL
Insulin	S, P	14.35–143.5 pmol/L	2–20 μU/mL
Iron	S	7–25 μmol/L	41–141 μg/dL
Iron-binding capacity	S	45–73 μmol/L	251–406 μg/dL
Iron-binding capacity saturation	S	0.16–0.35	16–35%
Ischemia modified albumin	S	<85 KU/L	<85 U/mL
Joint fluid crystal	JF	Not applicable	No crystals seen
Joint fluid mucin	JF	Not applicable	Only type I mucin present
Ketone (acetone)	S	Negative	Negative
Lactate	P, arterial P, venous	0.5–1.6 mmol/L 0.5–2.2 mmol/L	4.5–14.4 mg/dL 4.5–19.8 mg/dL
Lactate dehydrogenase	S	2.0–3.8 μkat/L	115–221 U/L
Lipase	S	0.51–0.73 μkat/L	3–43 U/L
Lipids: see Table 5			
Lipoprotein (a)	S	0–300 mg/L	0–30 mg/dL
Low-density lipoprotein (LDL) (see Table 5)			
Luteinizing hormone (LH) Female Menstruating Follicular phase Ovulatory phase Luteal phase Postmenopausal Male	S, P	2.0–15.0 U/L 22.0–105.0 U/L 0.6–19.0 U/L 16.0–64.0 U/L 2.0–12.0 U/L	2.0–15.0 mIU/mL 22.0–105.0 mIU/mL 0.6–19.0 mIU/mL 16.0–64.0 mIU/mL 2.0–12.0 mIU/mL
Magnesium	S	0.62–0.95 mmol/L	1.5–2.3 mg/dL
Metanephrine	P	<0.5 nmol/L <100 pg/mL	
Methemoglobin	WB	0.0–0.01	0–1%
Myoglobin Male Female	S	20>71μg/L 25>58μg/L	20>71 μg/L 25>58μg/L

Norepinephrine Supine (30 min) Sitting Standing (30 min)	P	650–2423 pmol/L 709–4019 pmol/L 739–4137 pmol/L	110–410 pg/mL 120–680 pg/mL 125–700 pg/mL
N-telopeptide (cross-linked), NTx Female, premenopausal Male BCE = bone collagen equivalent	S	6.2–19.0 nmol BCE 5.4–24.2 nmol BCE	6.2–19.0 nmol BCE 5.4–24.2 nmol BCE
NT-Pro BNP	S, P	<125 ng/L up to 75 years <450 ng/L >75 years	<125 pg/mL up to 75 years <450 pg/mL >75 years
5> Nucleotidase	S	0.00–0.19 μkat/L	0–11 U/L
Osmolality	P	275–295 mOsmol/kg serum water	275–295 mOsmol/kg serum water
Osteocalcin	S	11–50 μg/L	11–50 ng/mL
Oxygen content Arterial (sea level) Venous (sea level)	WB	17–21 10–16	17–21 vol% 10–16 vol%
Oxygen saturation (sea level) Arterial Venous, arm	WB	Fraction: 0.94–1.0 0.60–0.85	Percent: 94–100% 60–85%
Parathyroid hormone (intact)	S	8–51 ng/L	8–51 pg/mL
Phosphatase, alkaline	S	0.56–1.63 μkat/L	33–96 U/L
Phosphorus, inorganic	S	0.81–1.4 mmol/L	2.5–4.3 mg/dL
Potassium	S	3.5–5.0 mmol/L	3.5–5.0 meq/L
Prealbumin	S	170–340 mg/L	17–34 mg/dL
Procalcitonin	S	<0.1 μg/L	<0.1 ng/mL
Progesterone Female: Follicular Midluteal Male	S, P	<3.18 nmol/L 9.54–63.6 nmol/L <3.18 nmol/L	<1.0 ng/mL 3–20 ng/mL <1.0 ng/mL
Prolactin Male Female	S	53–360 mIU/L 40–530 mIU/L	2.5–17 ng/mL 1.9–25 ng/mL
Prostate-specific antigen (PSA) Prostate-specific antigen, free	S S	0.0–4.0 μg/L With total PSA between 4 and 10 μg/L and when the free PSA is: >0.25 decreased risk of prostate cancer <0.10 increased risk of prostate cancer	0.0–4.0 ng/mL With total PSA between 4 and 10 ng/mL and when the free PSA is: >25% decreased risk of prostate cancer <10% increased risk of prostate cancer
Protein fractions: Albumin Globulin Alpha$_1$ Alpha$_2$ Beta Gamma	S	35–55 g/L 20–35 g/L 2–4 g/L 5–9 g/L 6–11 g/L 7–17 g/L	3.5–5.5 g/dL (50–60%) 2.0–3.5 g/dL (40–50%) 0.2–0.4 g/dL (4.2–7.2%) 0.5–0.9 g/dL (6.8–12%) 0.6–1.1 g/dL (9.3–15%) 0.7–1.7 g/dL (13–23%)
Protein, total	S	67–86 g/L	6.7–8.6 g/dL
Pyruvate	P	40–130 μmol/L	0.35–1.14 mg/dL
Rheumatoid factor	S	<15 kIU/L	<15 IU/mL
Serotonin	WB	0.28–1.14 umol/L	50–200 ng/mL

Serum protein electrophoresis	S	Not applicable	Normal pattern
Sex hormone–binding globulin (adult) Male Female	S	11–80 nmol/L 30–135 nmol/L	11–80 nmol/L 30–135 nmol/L
Sodium	S	136–146 mmol/L	136–146 meq/L
Somatomedin-C (IGF-1)(adult) 16 years 17 years 18 years 19 years 20 years	S	226–903 μg/L 193–731 μg/L 163–584 μg/L 141–483 μg/L 127–424 μg/L	226–903 ng/mL 193–731 ng/mL 163–584 ng/mL 141–483 ng/mL 127–424 ng/mL
21–25 years 26–30 years 31–35 years 36–40 years 41–45 years 46–50 years 51–55 years 56–60 years 61–65 years 66–70 years 71–75 years 76–80 years 81–85 years		116–358μg/L 117–329 μg/L 115–307 μg/L 119–204 μg/L 101–267μg/L 94–252 μg/L 87–238μg/L 81–225 μg/L 75–212μg/L 69–200μg/L 64–188 μg/L 59–177 μg/L 55–166 μg/L	116–358 ng/mL 117–329 ng/mL 115–307 ng/mL 119–204 ng/mL 101–267 ng/mL 94–252 ng/mL 87–238 ng/mL 81–225 ng/mL 75–212 ng/mL 69–200 ng/mL 64–188 ng/mL 59–177 ng/mL 55–166 ng/mL
Somatostatin	P	<25 ng/L	<25 pg/mL
Testosterone, free Female, adult Male, adult	S	10.4–65.9 pmol/L 312–1041 pmol/L	3–19 pg/mL 90–300 pg/mL
Testosterone, total, Female Male	S	0.21–2.98 nmol/L 9.36–37.10 nmol/L	6–86 ng/dL 270–1070 ng/dL
Thyroglobulin	S	1.3>31.8μg/L	1.3–31.8 ng/mL
Thyroid-binding globulin	S	13–30 mg/L	1.3–3.0 mg/dL
Thyroid-stimulating hormone	S	0.34–4.25 mIU/L	0.34–4.25 μIU/mL
Thyroxine, free (fT4)	S	9.0–16 pmol/L	0.7–1.24 ng/dL
Thyroxine, total (T4)	S	70–151 nmol/L	5.4–11.7 μg/dL
Thyroxine index (free)	S	6.7–10.9	6.7–10.9
Transferrin	S	2.0–4.0 g/L	200–400 mg/dL
Triglycerides (see Table 5)	S	0.34–2.26 mmol/L	30–200 mg/dL
Triiodothyronine, free (fT3)	S	3.7–6.5 pmol/L	2.4–4.2 pg/mL
Triiodothyronine, total (T3)	S	1.2–2.1 nmol/L	77–135 ng/dL
Troponin I (method dependent) 99th percentile of a healthy population	S,P	0–0.04 μg/L	0–0.04 ng/mL
Troponin T 99th percentile of a healthy population	S,P	0–0.01 μg/L	0–0.01 ng/mL
Urea nitrogen	S	2.5–7.1 mmol/L	7–20 mg/dL
Uric acid	S	0.15–0.33 mmol/L	2.5–5.6 mg/dL

Females Males		0.18–0.41 mmol/L	3.1–7.0 mg/dL
Vasoactive intestinal polypeptide	P	0–60 ng/L	0–60 pg/mL
Zinc protoporphyrin	WB	0–400 μg/L	0–40 μg/dL
Zinc protoporphyrin (ZPP)-to-heme ratio	WB	0–69 μmol ZPP/mol heme	0–69 μmol ZPP/mol heme

Abbreviations: P, plasma; S, serum; WB, whole blood.

Table 3 Toxicology and Therapeutic Drug Monitoring

	Therapeutic Range	Toxic Level		
Drug	**SI Units**	**Conventional Units**	**SI Units**	**Conventional Units**
Acetaminophen	66–199 μmol/L	10–30 μg/mL	>1320 μmol/L	>200 μg/mL
Amikacin				
Peak	34–51 μmol/L	20–30 μg/mL	>60 μmol/L	>35 μg/mL
Trough	0–17 μmol/L	0–10 μg/mL	>17 μmol/L	>10 μg/mL
Amitriptyline/nortriptyline (total drug)	430–900 nmol/L	120–250 ng/mL	>1800 nmol/L	>500 ng/mL
Amphetamine	150–220 nmol/L	20–30 ng/mL	>1500 nmol/L	>200 ng/mL
Bromide	9.4–18.7 mmol/L	75–150 mg/dL	>18.8 mmol/L	>150 mg/dL
Mild toxicity			6.4–18.8 mmol/L	51–150 mg/dL
Severe toxicity			>18.8 mmol/L	>150 mg/dL
Lethal			>37.5 mmol/L	>300 mg/dL
Caffeine	25.8–103 μmol/L	5–20 μg/ml	>206 μmol/L	>40 μg/mL
Carbamazepine	17–42 μmol/L	4–10 μg/mL	> 85 μmol/L	>20 μg/mL
Chloramphenicol	31–62 μmol/L	10–20 μg/mL	>77 μmol/L	>25 μg/mL
Peak	15–31 μmol/L	5–10 μg/mL	>46 μmol/L	>15 μg/mL
Trough				
Chlordiazepoxide	1.7–10 μmol/L	0.5–3.0 μg/mL	>17μmol/L	>5.0 μg/mL
Clonazepam	32–240 nmol/L	10–75 ng/mL	>320 nmol/L	>100 ng/mL
Clozapine	0.6–2.1 μmol/L	200–700 ng/mL	>3.7 μmol/L	>1200 ng/mL
Cocaine			>3.3 μmol/L	>1.0 μg/mL
Codeine	43–110 nmol/mL	13–33 ng/mL	>3700 nmol/mL	>1100 ng/mL (lethal)
Cyclosporine	208–312 nmol/L	250–375 ng/mL	>312 nmol/L	>375 ng/mL
Renal transplant	166–250 nmol/L	200–300 ng/mL	>250 nmol/L	>300 ng/mL
0–6 months	83–125 nmol/L	100–150 ng/mL	>125 nmol/L	>150 ng/mL
6–12 months after transplant	208–291 nmol/L	250–350 ng/mL	>291 nmol/L	>350 ng/mL
>12 months	125–208 nmol/L	150–250 ng/mL	>208 nmol/L	>250 ng/mL
Cardiac transplant	83–125 nmol/L	100–150 ng/mL	>125 nmol/L	150 ng/mL
0–6 months				
6–12 months after transplant				
>12 months				
Lung transplant 0–6 months	250–374 nmol/L	300–450 ng/mL	>374 nmol/L	>450 ng/mL
Liver transplant	208–291 nmol/L	250–350 ng/mL	>291 nmol/L	>350 ng/mL
Initiation	83–166 nmol/L	100–200 ng/mL	>166 nmol/L	>200 ng/mL
Maintenance				
Desipramine	375–1130 nmol/L	100–300 ng/mL	>1880 nmol/L	>500 ng/mL
Diazepam (and metabolite)	0.7 –3.5 μmol/L	0.2–1.0 μg/mL	>7.0 μmol/L	>2.0 μg/mL
Diazepam	0.4– 6.6 μmol/L	0.1–1.8 μg/mL	>9.2 μmol/L	>2.5 μg/mL
Nordiazepam				
Digoxin	0.64–2.6 nmol/L	0.5–2.0 ng/mL	>5.0 nmol/L	>3.9 ng/mL
Disopyramide	5.3–14.7 μmol/L	2–5 μg/mL	>20.6 μmol/L	>7 μg/mL

Doxepin and nordoxepin	0.36–0.98 μmol/L	101–274 ng/mL	>1.8 μmol/L	>503 ng/mL
Doxepin	0.38–1.04 μmol/L	106–291 ng/mL	>1.9 μmol/L	>531 ng/mL
Nordoxepin				
Ethanol			>4.3 mmol/L	>20 mg/dL
Behavioral changes			≥17 mmol/L	≥80 mg/dL
Legal limit			>54 mmol/L	>250 mg/dL
Critical with acute exposure				
Ethylene glycol			>2 mmol/L	>12 mg/dL
Toxic			>20 mmol/L	>120 mg/dL
Lethal				
Ethosuximide	280–700 μmol/L	40–100 μg/mL	>700 μmol/L	>100 μg/mL
Everolimus	3.13–8.35 nmol/L	3–8 ng/mL	>12.5 nmol/L	>12 ng/mL
Flecainide	0.5–2.4 μmol/L	0.2–1.0 μg/mL	>3.6 μmol/L	>1.5 μg/mL
Gentamicin	10>21μmol/mL	5–10 μg/mL	>25μmol/mL	>12 μg/mL
Peak	0>4.2μmol/mL	0–2 μg/mL	>4.2μmol/mL	>2 μg/mL
Trough				
Heroin (diacetyl morphine)			>700 μmol/L	>200 ng/mL (as morphine)
Ibuprofen	49–243 μmol/L	10–50 μg/mL	>970 μmol/L	>200 μg/mL
Imipramine (and metabolite)	375–1130 nmol/L	100–300 ng/mL	>1880 nmol/L	>500 ng/mL
Desimipramine	563–1130 nmol/L	150–300 ng/mL	>1880 nmol/L	>500 ng/mL
Total imipramine + desimipramine				
Lamotrigine	11.7–54.7 μmol/L	3–14 μg/mL	>58.7 μmol/L	>15 μg/mL
Lidocaine	5.1–21.3μmol/L	1.2–5.0 μg/mL	>38.4 μmol/L	>9.0 μg/mL
Lithium	0.5–1.3 mmol/L	0.5–1.3 meq/L	>2 mmol/L	>2 meq/L
Methadone	1.0–3.2 μmol/L	0.3–1.0 μg/mL	>6.5 μmol/L	>2 μg/mL
Methamphetamine	0.07–0.34 μmol/L	0.01>0.05μg/mL	>3.35 μmol/L	>0.5μg/mL
Methanol			>6 mmol/L	>20 mg/dL
Methotrexate	0.01–0.1 μmol/L	0.01–0.1 μmol/L	>0.1 mmol/L	>0.1 mmol/L
Low-dose	<5.0 μmol/L	<5.0 μmol/L	>5.0 μmol/L	>5.0 μmol/L
High-dose (24h)	<0.50 μmol/L	<0.50 μmol/L	>0.5 μmol/L	>0.5 μmol/L
High-dose (48h)	<0.10 μmol/L	<0.10 μmol/L	>0.1 μmol/L	>0.1 μmol/L
High-dose (72h)				
Morphine	232–286 μmol/L	65–80 ng/mL	>720μmol/L	>200 ng/mL
Mycophenolic acid	3.1–10.9 μmol/L	1.0–3.5 ng/mL	>37 μmol/L	>12 ng/mL
Nitroprusside (as thiocyanate)	103–499 μmol/L	6–29 μg/mL	860 μmol/L	>50 μg/mL
Nortriptyline	190–569 nmol/L	50–150 ng/mL	>1900 nmol/L	>500 ng/mL
Phenobarbital	65–172 μmol/L	15–40 μg/mL	>258 μmol/L	>60 μg/mL
Phenytoin	40–79 μmol/L	10–20 μg/mL	>158 μmol/L	>40 μg/mL
Phenytoin, free	4.0–7.9 μg/mL	1–2 μg/mL	>13.9 μg/mL	>3.5 μg/mL
% Free	0.08–0.14	8–14		
Primidone and metabolite	23–55 μmol/L	5–12 μg/mL	>69 μmol/L	>15 μg/mL
Primidone	65–172 μmol/L	15–40 μg/mL	>215 μmol/L	7gt;50 μg/mL
Phenobarbital				
Procainamide	17–42 μmol/L	4–10 μg/mL	>43 μmol/L	>10 μg/mL
Procainamide	22–72 μmol/L	6>20μg/mL	>126 μmol/L	>35 μg/mL
NAPA (*N*-acetylprocainamide)				
Quinidine	6.2–15.4 μmol/L	2.0–5.0 μg/mL	>19 μmol/L	>6 μg/mL
Salicylates	145–2100 μmol/L	2–29 mg/dL	>2900 μmol/L	>40 mg/dL
Sirolimus (trough level) Kidney transplant	4.4–15.4 nmol/L	4–14 ng/mL	>16 nmol/L	>15 ng/mL
Tacrolimus (FK506) (trough)				
Kidney and liver	12–19 nmol/L	10–15 ng/mL	>25 nmol/L	>20 ng/mL
Initiation	6–12 nmol/L	5–10 ng/mL	>25 nmol/L	>20 ng/mL

Maintenance Heart Initiation Maintenance	19–25 nmol/L 6–12 nmol/L	15–20 ng/mL 5–10 ng/mL		
Theophylline	56–111 μg/mL	10–20 μg/mL	>168 μg/mL	>30 μg/mL
Thiocyanate After nitroprusside infusion Nonsmoker Smoker	103–499 μmol/L 17–69 μmol/L 52–206 μmol/L	6–29 μg/mL 1–4 μg/mL 3–12 μg/mL	860 μmol/L	>50 μg/mL
Tobramycin Peak Trough	11–22 μg/L 0–4.3 μg/L	5–10 μg/mL 0–2 μg/mL	>26 μg/L >4.3 μg/L	>12 μg/mL >2 μg/mL
Valproic acid	346–693 μmol/L	50–100 μg/mL	>693 μmol/L	>100 μg/mL
Vancomycin Peak Trough	14–28 μmol/L 3.5–10.4μmol/L	20–40 μg/mL 5–15 μg/mL	>55μmol/L >14 μmol/L	>80 μg/mL >20 μg/mL

Table 4 Vitamins and Selected Trace Minerals

		Reference Range	
Specimen	**Analyte**	**SI Units**	**Conventional Units**
Aluminum	S	<0.2 μmol/L	<5.41 μg/L
Arsenic	WB	0.03–0.31 μmol/L	2–23 μg/L
Cadmium	WB	<44.5 nmol/L	<5.0 μg/L
Coenzyme Q10 (ubiquinone)	P	433–1532 μg/L	433–1532 μg/L
β-Carotene	S	0.07–1.43 μmol/L	4–77 μg/dL
Copper	S	11–22 μmol/L	70–140 μg/dL
Folic acid	RC	340–1020 nmol/L cells	150–450 ng/mL cells
Folic acid	S	12.2–40.8 nmol/L	5.4–18.0 ng/mL
Lead (adult)	S	<0.5 μmol/L	<10 μg/dL
Mercury	WB	3.0–294 nmol/L	0.6–59 μg/L
Selenium	S	0.8–2.0 umol/L	63–160 μg/L
Vitamin A	S	0.7–3.5 μmol/L	20–100 μg/dL
Vitamin B1 (thiamine)	S	0–75 nmol/L	0–2 μg/dL
Vitamin B2 (riboflavin)	S	106–638 nmol/L	4–24 μg/dL
Vitamin B6	P	20–121 nmol/L	5–30 ng/mL
Vitamin B12	S	206–735 pmol/L	279–996 pg/mL
Vitamin C (ascorbic acid)	S	23–57 μmol/L	0.4–1.0 mg/dL
Vitamin D3,1,25-dihydroxy, total	S, P	36–180 pmol/L	15–75 pg/mL
Vitamin D3,25-hydroxy, total	P	75–250 nmol/L	30–100 ng/mL
Vitamin E	S	12–42 μmol/L	5–18 μg/mL
Vitamin K	S	0.29–2.64 nmol/L	0.13–1.19 ng/mL
Zinc	S	11.5–18.4 μmol/L	75–120 μg/dL

Abbreviations: P, plasma; RC, red cells; S, serum; WB, whole blood.

Table 5 Classification of LDL, Total, and HDL Cholesterol

LDL Cholesterol	
<70 mg/dL	Therapeutic option for very high risk patients
<100 mg/dL	Optimal
100–129 mg/dL	Near optimal/above optimal
130–159 mg/dL	Borderline high
160–189 mg/dL	High
≥190 mg/dL	Very high
Total Cholesterol	
<200 mg/dL	Desirable
200–239 mg/dL	Borderline high

≥240 mg/dL	High
HDL Cholesterol	
<40 mg/dL	Low
≥60 mg/dL	High

Abbreviations: LDL, low-density lipoprotein; HDL, high-density lipoprotein. ***Source:*** Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). JAMA 2001; 285:2486–97. Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. SM Grundy et al for the Coordinating Committee of the National Cholesterol Education Program: Circulation 110:227, 2004.

REFERENCE VALUES FOR SPECIFIC ANALYTES

Table 6 Cerebrospinal Fluid[a]

	Reference Range	
Constituent	SI Units	Conventional Units
Osmolarity	292–297 mmol/kg water	292–297 mOsm/L
Electrolytes		
Sodium	137–145 mmol/L	137–145 meq/L
Potassium	2.7–3.9 mmol/L	2.7–3.9 meq/L
Calcium	1.0–1.5 mmol/L	2.1–3.0 meq/L
Magnesium	1.0–1.2 mmol/L	2.0–2.5 meq/L
Chloride	116–122 mmol/L	116–122 meq/L
CO_2 content	20–24 mmol/L	20–24 meq/L
PCO_2	6–7 kPa	45–49 mmHg
pH	7.31–7.34	
Glucose	2.22–3.89 mmol/L	40–70 mg/dL
Lactate	1–2 mmol/L	10–20 mg/dL
Total protein:		
Lumbar	0.15–0.5 g/L	15–50 mg/dL
Cisternal	0.15–0.25 g/L	15–25 mg/dL
Ventricular	0.06–0.15 g/L	6–15 mg/dL
Albumin	0.066–0.442 g/L	6.6–44.2 mg/dL
IgG	0.009–0.057 g/L	0.9–5.7 mg/dL
IgG index[b]	0.29–0.59	
Oligoclonal bands (OGB)	<2 bands not present in matched serum sample	
Ammonia	15–47 μmol/L	25–80 μg/dL
Creatinine	44–168 μmol/L	0.5–1.9 mg/dL
Myelin basic protein	<4 μg/L	
CSF pressure		50–180 mmH_2O
CSF volume (adult)	~150 mL	
Red blood cells	0	0
Leukocytes		
Total	0–5 mononuclear cells per μL	
Differential		
Lymphocytes	60–70%	
Monocytes	30–50%	
Neutrophils	None	

[a]Since cerebrospinal fluid concentrations are equilibrium values, measurements of the same parameters in blood plasma obtained at the same time are recommended. However, there is a time lag in attainment of equilibrium, and cerebrospinal levels of plasma constituents that can fluctuate rapidly (such as plasma glucose) may not achieve stable values until after a significant lag phase. [b]IgG index = CSF IgG (mg/dL) x serum albumin (g/dL)/serum IgG (g/dL) x CSF albumin (mg/dL).

Table 7a Differential Nucleated Cell Countsof Bone Marrow Aspirates[a] (See Chaps. 57, e17)

	Observed Range (%)	95% Range (%)	Mean (%)
Blast cells	0–3.2	0–3.0	1.4
Promyelocytes	3.6–13.2	3.2–12.4	7.8
Neutrophil myelocytes	4–21.4	3.7–10.0	7.6
Eosinophil myelocytes	0–5.0	0–2.8	1.3
Metamyelocytes	1–7.0	2.3–5.9	4.1
Neutrophils Males Females	21.0–45.6 29.6–46.6	21.9–42.3 28.8–45.9	32.1 37.4
Eosinophils	0.4–4.2	0.3–4.2	2.2
Eosinophils plus eosinophil myelocytes	0.9–7.4	0.7–6.3	3.5
Basophils	0–0.8	0–0.4	0.1
Erythroblasts Male Females	18.0–39.4 14.0–31.8	16.2–40.1 13.0–32.0	28.1 22.5
Lymphocytes	4.6–22.6	6.0–20.0	13.1
Plasma cells	0–1.4	0–1.2	0.6
Monocytes	0–3.2	0–2.6	1.3
Macrophages	0–1.8	0–1.3	0.4
M:E ratio Males Females	1.1–4.0 1.6–5.4	1.1–4.1 1.6–5.2	2.1 2.8

[a]Based on bone marrow aspirate from 50 healthy volunteers (30 men, 20 women).
Abbreviation: M:E, myeloid to erythroid ratio.
Source: BJ Bain: Br J Haematol 94:206, 1996.

Table 7b Bone Marrow Cellularity

Age	Observed Range	95% Range	Mean
Under 10 years	59.0–95.1%	72.9–84.7%	78.8%
10–19 years	41.5–86.6%	59.2–69.4%	64.3%
20–29 years	32.0–83.7%	54.1–61.9%	58.0%
30–39 years	30.3–81.3%	41.1–54.1%	47.6%
40–49 years	16.3–75.1%	43.5–52.9%	48.2%
50–59 years	19.7–73.6%	41.2–51.4%	46.3%
60–69 years	16.3–65.7%	40.8–50.6%	45.7%
70–79 years	11.3–47.1%	22.6–35.2%	28.9%

Source: From RJ Hartsock et al: Am J Clin Pathol 1965; 43:326, 1965.

Table 8 Stool Analysis

	Reference Range	
	SI Units	Conventional Units
Alpha-1-antitrypsin	>540 mg/L	>54 mg/dL
Amount	0.1–0.2 kg/d	100–200 g/24 h
Coproporphyrin	611–1832 nmol/d	400–1200 μg/24 h
Fat Adult Adult on fat-free diet		<7 g/d <4 g/d
Fatty acids	0–21 mmol/d	0–6 g/24 h
Leukocytes	None	None
Nitrogen	<178 mmol/d	<2.5 g/24 h
pH	7.0–7.5	
Potassium	14–102 mmol/L	14–102 mmol/L
Occult blood	Negative	Negative
Osmolality	280–325 mOsm	280–325 mOsm

Sodium	7–72 mmol/L	7–72 mmol/L
Trypsin		20–95 U/g
Urobilinogen	85–510 μmol/d	50–300 mg/24 h
Uroporphyrins	12–48 nmol/d	10–40 μg/24 h
Water	<0.75	<75%

Source: Modified from: FT Fishbach, MB Dunning III: *A Manual of Laboratory and Diagnostic Tests* , 7th ed. Philadelphia, Lippincott Williams & Wilkins, 2004.

Table 9 Urine Analysis and Renal Function Tests

	Reference Range	
	SI Units	**Conventional Units**
Acidity, titratable	20–40 mmol/d	20–40 meq/d
Aldosterone	Normal diet: 6–25 μg/d Low-salt diet: 17–44 μg/d High-salt diet: 0–6 μg/d	Normal diet: 6–25 μg/d Low-salt diet: 17–44 μg/d High-salt diet: 0–6 μg/d
Aluminum	0.19–1.11 μmol/L	5–30 μg/L
Ammonia	30–50 mmol/d	30–50 meq/d
Amylase		4–400 U/L
Amylase/creatinine clearance ratio [(Cl_{am}/Cl_{cr}) x 100]	1–5	1–5
Arsenic	0.07–0.67 μmol/d	5–50 μg/d
Bence Jones protein, urine, qualitative	Not applicable	None detected
Bence Jones protein, urine, quantitative Free Kappa Free Lambda K/L ratio	1.4–24.2 mg/L 0.2–6.7 mg/L 2.04–10.37	0.14–2.42 mg/dL 0.02–0.67 mg/dL 2.04–10.37
Calcium (10 meq/d or 200 mg/d dietary calcium)	<7.5 mmol/d	<300 mg/d
Chloride	140–250 mmol/d	140–250 mmol/d
Citrate	320–1240 mg/d	320–1240 mg/d
Copper	<0.95 μmol/d	<60 μg/d
Coproporphyrins (types I and III)	0–20 μmol/mol creatinine	0–20 μmol/mol creatinine
Cortisol, free	55–193 nmol/d	20–70 μg/d
Creatine, as creatinine Female Male	<760 μmol/d <380 μmol/d	<100 mg/d <50 mg/d
Creatinine	8.8–14 mmol/d	1.0–1.6 g/d
Dopamine	392–2876 nmol/d	60–440 μg/d
Eosinophils	<100 eosinophils/mL	<100 eosinophils/mL
Epinephrine	0–109 nmol/day	0–20 μg/day
Glomerular filtration rate	>60 mL/min/1.73 m^2 For African Americans multiply the result by 1.21	>60 mL/min/1.73 m^2 For African Americans multiply the result by 1.21
Glucose (glucose oxidase method)	0.3–1.7 mmol/d	50–300 mg/d
5-Hydroindoleacetic acid [5-HIAA]	0–78.8 μmol/d	0–15 mg/d
Hydroxyproline	53>328μmol/d	53–328 μmol/d
Iodine, spot urine WHO classification of iodine deficiency: Not iodine deficient Mild iodine deficiency Moderate iodine deficiency Severe iodine deficiency	>100 μg/L 50–100 μg/L 20–49 μg/L <20 μg/L	>100 μg/L 50–100 μg/L 20–49 μg/L <20 μg/L
Ketone (acetone)	Negative	Negative
17 Ketosteroids	3–12 mg/d	3–12 mg/d

Metanephrines Metanephrine Normetanephrine	30–350 μg/d 50–650 μg/d	30–350 μg/d 50–650 μg/d
Microalbumin –Normal –Microalbuminuria –Clinical albuminuria	0.0–0.03 g/d 0.03–0.30 g/d >0.3 g/d	0–30 mg/d 30–300 mg/d >300 mg/d
Microalbumin/creatinine ratio –Normal –Microalbuminuria –Clinical albuminuria	0–3.4 g/mol creatinine 3.4–34 g/mol creatinine >34 g/mol creatinine	0–30 μg/mg creatinine 30–300 μg/mg creatinine >300 μg/mg creatinine
β_2-Microglobulin	0–160μg/L	0–160μg/L
Norepinephrine	89–473 nmol/d	15–80 μg/d
N-telopeptide (cross-linked), NTx –Female, premenopausal –Female, postmenopausal –Male –BCE = bone collagen equivalent	17–94 nmol BCE/mmol creatinine 26–124 nmol BCE/mmol creatinine 21–83 nmol BCE/mmol creatinine	17–94 nmol BCE/mmol creatinine 26–124 nmol BCE/mmol creatinine 21–83 nmol BCE/mmol creatinine
Osmolality	500–800 mOsmol/kg water	500–800 mOsmol/kg water
Oxalate –Male –Female	80–500 μmol/d 45–350 μmol/d	7–44 mg/d 4–31 mg/d
pH	5.0–9.0	5.0–9.0
Phosphate (phosphorus) (varies with intake)	12.9–42.0 mmol/d	400–1300 mg/d
Porphobilinogen	None	None
Potassium (varies with intake)	25–100 mmol/d	25–100 meq/d
Protein	<0.15 g/d	<150 mg/d
Protein/creatinine ratio	Male: 15–68 mg/g Female: 10–107 mg/g	Male: 15–68 mg/g Female: 10–107 mg/g
Sediment –Red blood cells –White blood cells –Bacteria –Crystals –Bladder cells –Squamous cells –Tubular cells –Broad casts –Epithelial cell casts –Granular casts –Hyaline casts –Red blood cell casts –Waxy casts –White cell casts	0–2/high-power field 0–2/high-power field None None None None None None None None 0–5/low-power field None None None	
Sodium (varies with intake)	100–260 mmol/d	100–260 meq/d
Specific gravity: – After 12-h fluid restriction – After 12-h deliberate water intake	>1.025 ≤1.003	>1.025 ≤1.003
Tubular reabsorption, phosphorus	0.79–0.94 of filtered load	79–94% of filtered load
Urea nitrogen	214–607 mmol/d	6–17 g/d
Uric acid (normal diet)	1.49–4.76 mmol/d	250–800 mg/d
Vanillylmandelic acid (VMA)	<30 μmol/d	<6 mg/d

SPECIAL FUNCTION TESTS

Table 10 Normal Pressures in Heart and Great Vessels

Pressure (mmHg)	Average	Range
Right Atrium		
Mean	2.8	1–5
a wave	5.6	2.5–7
c wave	3.8	1.5–6
x wave	1.7	0–5
v wave	4.6	2–7.5
y wave	2.4	0–6
Right Ventricle		
Peak systolic	25	17–32
End-diastolic	4	1–7
Pulmonary Artery		
Mean	15	9–19
Peak systolic	25	17–32
End-diastolic	9	4–13
Pulmonary Artery Wedge		
Mean	9	4.5–13
Left Atrium		
Mean	7.9	2–12
a wave	10.4	4–16
v wave	12.8	6–21
Left Ventricle		
Peak systolic	130	90–140
End-diastolic	8.7	5–12
Brachial Artery		
Mean	85	70–105
Peak systolic	130	90–140
End-diastolic	70	60–90

Source: Reproduced from: MJ Kern *The Cardiac Catheterization Handbook*, 4th ed., Philadelphia, Mosby, 2003.

Table 11 Circulatory Function Tests

	Results: Reference Range	
Test	**SI Units (Range)**	**Conventional Units (Range)**
Arteriovenous oxygen difference	30–50 mL/L	30–50 mL/L
Cardiac output (Fick)	2.5–3.6 L/m^2 of body surface area per min	2.5–3.6 L/m^2 of body surface area per min
Contractility indexes		
Max. left ventricular *dp*/*dt* (*dp*/*dt*)	220 kPa/s (176–250 kPa/s)	1650 mmHg/s (1320–1880 mmHg/s)
DP when DP = 5.3 kPa	(37.6 ± 12.2)/s	(37.6 ± 12.2)/s
(40 mmHg) (DP, developed LV pressure)	3.32 ± 0.84 end-diastolic volumes per second	3.32 ± 0.84 end-diastolic volumes per second
Mean normalized systolic ejection rate (angiography) Mean velocity of circumferential fiber shortening (angiography)	1.83 ± 0.56 circumferences per second	1.83 ± 0.56 circumferences per second
Ejection fraction: stroke volume/end-diastolic volume (SV/EDV)	0.67 ± 0.08 (0.55–0.78)	0.67 ± 0.08 (0.55–0.78)
End-diastolic volume	70 ± 20.0 mL/m^2 (60–88 mL/m^2)	70 ± 20.0 mL/m^2 (60–88 mL/m^2)
End-systolic volume	25 ± 5.0 mL/m^2 (20–33 mL/m^2)	25 ± 5.0 mL/m^2 (20–33 mL/m^2)
Left ventricular work		

Stroke work index	50 ± 20.0 (g·m)/m^2 (30–110)	50 ± 20.0 (g·m)/m^2 (30–110)
Left ventricular minute work index	1.8–6.6 [(kg·m)/m^2]/min	1.8–6.6 [(kg·m)/m^2]/min
Oxygen consumption index	110–150 mL	110–150 mL
Maximum oxygen uptake	35 mL/min (20–60 mL/min)	35 mL/min (20–60 mL/min)
Pulmonary vascular resistance	2–12 (kPa·s)/L	20–130 (dyn·s)/cm^5
Systemic vascular resistance	77–150 (kPa·s)/L	770–1600 (dyn·s)/cm^5

Source: E Braunwald et al: *Heart Disease*, 6th ed. Philadelphia, W.B. Saunders Co., 2001.

Table 12 Normal Echocardiographic Reference Limits and Partition Values in Adults

	Women Reference Range	Mildly Abnormal	Moderately Abnormal	Severely Abnormal	Men Reference Range	Mildly Abnormal	Moderately Abnormal	Severely Abnormal
Left ventricular dimensions								
Septal thickness, cm	0.6–0.9	1.0–1.2	1.3–1.5	≥1.6	0.6–1.0	1.1–1.3	1.4–1.6	≥1.7
Posterior wall thickness, cm	0.6–0.9	1.0–1.2	1.3–1.5	≥1.6	0.6–1.0	1.1–1.3	1.4–1.6	≥1.7
Diastolic diameter, cm	3.9–5.3	5.4–5.7	5.8–6.1	≥6.2	4.2–5.9	6.0–6.3	6.4–6.8	≥6.9
Diastolic diameter/BSA, cm/m^2	2.4–3.2	3.3–3.4	3.5–3.7	≥3.8	2.2–3.1	3.2–3.4	3.5–3.6	≥3.7
Diastolic diameter/height, cm/m	2.5–3.2	3.3–3.4	3.5–3.6	≥3.7	2.4–3.3	3.4–3.5	3.6–3.7	≥3.8
Left ventricular volumes								
Diastolic, mL	56–104	105–117	118–130	≥131	67–155	156–178	179–201	≥202
Diastolic/BSA, mL/m^2	35–75	76–86	87–96	≥97	35–75	76–86	87–96	≥97
Systolic, mL	19–49	50–59	60–69	≥70	22–58	59–70	71–82	≥83
Systolic/BSA, mL/m^2	12–30	31–36	37–42	≥43	12–30	31–36	37–42	≥43
Left ventricular mass, 2D method								
Mass, g	66–150	151–171	172–182	≥183	96–200	201–227	228–254	≥255
Mass/BSA, g/m^2	44–88	89–100	101–112	≥113	50–102	103–116	117–130	≥131
Left ventricular function								
Endocardial fractional shortening (%)?	27–45	22–26	17–21	≥16	25–43	20–24	15–19	≥14
Midwall fractional shortening (%)	15–23	13–14	11–12	≥10	14–22	12–13	10–11	≥9
Ejection fraction, 2D method (%)	≥55	45–54	30–44	≥29	≥55	45–54	30–44	≥29
Right heart dimensions (cm)								
Basal RV diameter	2.0–2.8	2.9–3.3	3.4–3.8	≥3.9	2.0–2.8	2.9–3.3	3.4–3.8	≥3.9
Mid-RV diameter	2.7–3.3	3.4–3.7	3.8–4.1	≥4.2	2.7–3.3	3.4–3.7	3.8–4.1	≥4.2

Base-to-apex length	7.1–7.9	8.0–8.5	8.6–9.1	≥9.2	7.1–7.9	8.0–8.5	8.6–9.1	≥9.2
RVOT diameter above aortic valve	2.5–2.9	3.0–3.2	3.3–3.5	≥3.6	2.5–2.9	3.0–3.2	3.3–3.5	≥3.6
RVOT diameter above pulmonic valve	1.7–2.3	2.4–2.7	2.8–3.1	≥3.2	1.7–2.3	2.4–2.7	2.8–3.1	≥3.2
Pulmonary artery diameter below pulmonic valve	1.5–2.1	2.2–2.5	2.6–2.9	≥3.0	1.5–2.1	2.2–2.5	2.6–2.9	≥3.0
Right ventricular size and function in 4-chamber view								
Diastolic area, cm^2	11–28	29–32	33–37	≥38	11–28	29–32	33–37	≥38
Systolic area, cm^2	7.5–16	17–19	20–22	≥23	7.5–16	17–19	20–22	≥23
Fractional area change, %	32–60	25–31	18–24	≥17	32–60	25–31	18–24	≥17
Atrial sizes								
LA diameter, cm	2.7–3.8	3.9–4.2	4.3–4.6	≥4.7	3.0–4.0	4.1–4.6	4.7–5.2	≥5.3
LA diameter/BSA, cm/m^2	1.5–2.3	2.4–2.6	2.7–2.9	≥3.0	1.5–2.3	2.4–2.6	2.7–2.9	≥3.0
RA minor axis, cm	2.9–4.5	4.6–4.9	5.0–5.4	≥5.5	2.9–4.5	4.6–4.9	5.0–5.4	≥5.5
RA minor axis/BSA, cm/m^2	1.7–2.5	2.6–2.8	2.9–3.1	≥3.2	1.7–2.5	2.6–2.8	2.9–3.1	≥3.2
LA area, cm^2	<20	20–30	30–40	≥41	<20	20–30	30–40	≥41
LA volume, mL	22–52	53–62	63–72	≥73	18–58	59–68	69–78	≥79
LA volume/BSA, mL/m^2	16–28	29–33	34–39	≥40	16–28	29–33	34–39	≥40
Aortic stenosis, classification of severity								
Aortic jet velocity, m/s		2.6–2.9	3.0–4.0	>4.0		2.6–2.9	3.0–4.0	>4.0
Mean gradient, mmHg		<20	20–40	>40		<20	20–40	>40
Valve area, cm^2		>1.5	1.0–1.5	<1.0	>1.5	1.0–1.5	<1.0	
Indexed valve area, cm^2/m^2		>0.85	0.60–0.85	<0.6		>0.85	0.60–0.85	<0.6
Velocity ratio		>0.50	0.25–0.50	<0.25		>0.50	0.25–0.50	<0.25
Mitral stenosis, classification of severity								
Valve area, cm^2		>1.5	1.0–1.5	<1.0		>1.5	1.0–1.5	<1.0
Mean gradient, mmHg		<5	5–10	>10		<5	5–10	>10
Pulmonary artery pressure, mmHg		<30	30–50	>50		<30	30–50	>50

Aortic regurgitation, indices of severity							
Vena contracta width, cm	<0.30	0.30–0.60	≥0.60		<0.30	0.30–0.60	≥0.60
Jet width/LVOT width, %	<25	25–64	≥65		<25	25–64	≥65
Jet CSA/LVOT CSA, %	<5	5–59	≥60		<5	5–59	≥60
Regurgitant volume, mL/beat	<30	30–59	≥60		<30	30–59	≥60
Regurgitant fraction, %	<30	30–49	≥50		<30	30–49	≥50
Effective regurgitant orifice area, cm^2	<0.10	0.10–0.29	≥0.30		<0.10	0.10–0.29	≥0.30
Mitral regurgitation, indices of severity							
Vena contracta width, cm	<0.30	0.30–0.69	≥0.70		<0.30	0.30–0.69	≥0.70
Regurgitant volume, mL/beat	<30	30–59	≥60		<30	30–59	≥60
Regurgitant fraction, %	<30	30–49	≥50		<30	30–49	≥50
Effective regurgitant orifice area, cm^2	<0.20	0.20–0.39	≥0.40		<0.20	0.20–0.39	≥0.40

Abbreviations: BSA, body surface area; CSA, cross-sectional area; LA, left atrium; LVOT, left ventricular outflow tract; RA, right atrium; RV, right ventricle; RVOT, right ventricular outflow tract; 2D, 2-dimensional. ***Source:*** Values adapted from: American Society of Echocardiography, Guidelines and Standards. *http://www.asecho.org/i4a/pages/index.cfm?pageid=3317*. Accessed Feb 23, 2010.

Table 13 Summary of Values Useful in Pulmonary Physiology

		Typical Values	
	Symbol	**Man Aged 40, 75 kg, 175 cm Tall**	**Woman Aged 40, 60 kg, 160 cm Tall**
Pulmonary Mechanics			
Spirometry—volume-time curves			
Forced vital capacity	FVC	5.0 L	3.4 L
Forced expiratory volume in 1 s	FEV_1	4.0 L	2.8 L
FEV_1/FVC	FEV_1%	80%	78%
Maximal midexpiratory flow rate	MMEF (FEF 25–75)	4.1 L/s	3.2L/s
Maximal expiratory flow rate	MEFR (FEF 200–1200)	9.0 L/s	6.1 L/s
Spirometry—flow-volume curves			
Maximal expiratory flow at 50% of expired vital capacity	V_{max} 50 (FEF 50%)	5.0 L/s	4.0 L/s
Maximal expiratory flow at 75% of expired vital capacity	V_{max} 75 (FEF 75%)	2.1 L/s	2.0 L/s
Resistance to airflow:			
Pulmonary resistance	RL (R_L)	<3.0 (cmH_2O/s)/L	
Airway resistance	Raw	<2.5 (cmH_2O/s)/L	

Specific conductance	SGaw	>0.13 cmH_2O/s	
Pulmonary compliance			
Static recoil pressure at total lung capacity	Pst TLC	25 ± 5 cmH_2O	
Compliance of lungs (static)	CL	0.2 L cmH_2O	
Compliance of lungs and thorax	C(L + T)	0.1 L cmH_2O	
Dynamic compliance of 20 breaths per minute	C dyn 20	0.25 ± 0.05 L/cmH_2O	
Maximal static respiratory pressures:			
Maximal inspiratory pressure	MIP	>110 cmH_2O	>70 cmH_2O
Maximal expiratory pressure	MEP	>200 cmH_2O	>140 cmH_2O
Lung Volumes			
Total lung capacity	TLC	6.9 L	4.9 L
Functional residual capacity	FRC	3.3 L	2.6 L
Residual volume	RV	1.9 L	1.5 L
Inspiratory capacity	IC	3.7 L	2.3 L
Expiratory reserve volume	ERV	1.4 L	1.1 L
Vital capacity	VC	5.0 L	3.4 L
Gas Exchange (Sea Level)			
Arterial O_2 tension	Pa_{O2}	12.7 ± 0.7 kPa (95 ± 5 mmHg)	
Arterial CO_2 tension	Pa_{CO2}	5.3 ± 0.3 kPa (40 ± 2 mmHg)	
Arterial O_2 saturation	Sa_{O2}	0.97 ± 0.02 (97 ± 2%)	
Arterial blood pH	pH	7.40 ± 0.02	
Arterial bicarbonate	HCO_3^-	24 + 2 meq/L	
Base excess	BE	0 ± 2 meq/L	
Diffusing capacity for carbon monoxide (single breath)	DL_{CO}	37 mL CO/min/mmHg 27 mL CO/min/mmHg	
Dead space volume	V_D	2 mL/kg body wt	
Physiologic dead space; dead space-tidal volume ratio	V_D/V_T		
Rest		35% V_T	
Exercise		?20% V_T	
Alveolar-arterial difference for O_2	$P(A-a)O2$	≥2.7 kPa ≤20 kPa (≤24 mmHg)	

Source: Based on: AH Morris et al: *Clinical Pulmonary Function Testing. A Manual of Uniform Laboratory Procedures*, 2nd ed. Salt Lake City, Utah, Intermountain Thoracic Society, 1984.

Table 14 Gastrointestinal Tests

Results		
Test	**SI Units**	**Conventional Units**
Absorption tests		
D-Xylose: after overnight fast, 25 g xylose given in oral aqueous solution		
Urine, collected for following 5 h	25% of ingested dose	25% of ingested dose
Serum, 2 h after dose	2.0–3.5 mmol/L	30–52 mg/dL
Vitamin A: a fasting blood specimen is obtained and 200,000 units of	Serum level should	Serum level should

vitamin A in oil is given orally	rise to twice fasting level in 3–5 h	rise to twice fasting level in 3–5 h
Bentiromide test (pancreatic function): 500 mg bentiromide (chymex) orally; *p*-aminobenzoic acid (PABA) measured		
Plasma Urine	>50% recovered in 6 h	>3.6 (±1.1) μg/mL at 90 min
Gastric juice		
Volume		
24 h	2–3 L	2–3 L
Nocturnal	600–700 mL	600–700 mL
Basal, fasting	30–70 mL/h	30–70 mL/h
Reaction		
pH	1.6–1.8	1.6–1.8
Titratable acidity of fasting juice	4–9 μmol/s	15–35 meq/h
Acid output		
Basal		
Females (mean ± 1 SD)	0.6 ± 0.5 μmol/s	2.0 ± 1.8 meq/h
Males (mean ± 1 SD)	0.8 ± 0.6 μmol/s	3.0 ± 2.0 meq/h
Maximal (after SC histamine acid phosphate, 0.004 mg/kg body weight, and preceded by 50 mg promethazine, or after betazole, 1.7 mg/kg body weight, or pentagastrin, 6 μg/kg body weight)		
Females (mean ± 1 SD)	4.4 ± 1.4 μmol/s	16 ± 5 meq/h
Males (mean ± 1 SD)	6.4 ± 1.4 μmol/s	23 ± 5 meq/h
Basal acid output/maximal acid output ratio	0.6	0.6
Gastrin, serum	0–200 μg/L	0–200 pg/mL
Secretin test (pancreatic exocrine function): 1 unit/kg body weight, IV		
Volume (pancreatic juice) in 80 min	>2.0 mL/kg	>2.0 mL/kg
Bicarbonate concentration	>80 mmol/L	>80 meq/L
Bicarbonate output in 30 min	>10 mmol	>10 meq

MISCELLANEOUS

Table 15 Body Fluids and Other Mass Data

	Reference Range	
	SI Units	**Conventional Units**
Ascitic fluid: See Chap. 43		
Body fluid, Total volume (lean) of body weight Intracellular Extracellular	50% (in obese) to 70% 0.3–0.4 of body weight 0.2–0.3 of body weight	
Blood		
Total volume		
Males	69 mL/kg body weight	
Females	65 mL/kg body weight	
Plasma volume		
Males	39 mL/kg body weight	
Females	40 mL/kg body weight	
Red blood cell volume		
Males	30 mL/kg body weight	1.15–1.21 L/m^2 of body surface area
Females	25 mL/kg body weight	0.95–1.00 L/m^2 of body surface area
Body mass index	18.5–24.9 kg/m^2	18.5–24.9 kg/m^2

Table 16 Radiation-Derived Units

Quantity	Measures	Old Unit	SI Unit	Special Name for SI Unit	Conversion

				(Abbreviation)	
Activity	Rate of radioactive decay	curie (Ci)	Disintegrations per second (dps)	becquerel (Bq)	1 Ci = 3.7 x 10^{10} Bq 1 mCi = 37 MBq 1 Bq = 2.703 x 10^{-11} Ci
Exposure	Amount of ionizations produced in dry air by x-rays or gamma rays, per unit of mass	roentgen (R)	Coulomb per kilogram (C/kg)	none	1 C/kg = 3876 R 1 R = 2.58 x 10^{-4} C/kg 1 mR = 258 pC/kg
Air kerma	Sum of initial energies of charged particles liberated by ionizing radiation in air, per unit of mass	rad	Joule per kilogram (J/kg)	gray (Gy)	1 Gy = 100 rad 1 rad = 0.01 Gy 1 mrad = 10 μGy
Absorbed dose	Energy deposited per unit of mass in a medium, e.g. an organ/tissue	rad	Joule per kilogram (J/kg)	gray (Gy)	1 Gy = 100 rad 1 rad = 0.01 Gy 1 mrad = 10 μGy
Equivalent dose	Energy deposited per unit of mass in a medium, e.g. an organ/tissue, weighted to reflect type(s) of radiation	rem	Joule per kilogram (J/kg)	sievert (Sv)	1 Sv = 100 rem 1 rem = 0.01 Sv 1 mrem = 10 μSv
Effective dose	Energy deposited per unit of mass in a reference individual, doubly weighted to reflect type(s) of radiation and organ(s) irradiated	rem	Joule per kilogram (J/kg)	sievert (Sv)	1 Sv = 100 rem 1 rem = 0.01 Sv 1 mrem = 10 μSv

ACKNOWLEDGMENT

The contributions of Drs. Daniel J. Fink, Patrick M. Sluss, James L. Januzzi, and Kent B. Lewandrowski to this chapter in previous editions are gratefully acknowledged. We also express our gratitude to Drs. Amudha Palanisamy and Scott Fink for careful review of tables and helpful suggestions.

FURTHER READINGS

Hickman PE, Koerbin G: Methods in Clinical Chemistry. An accessory work to the fifth edition of Kaplan and Pesce's: *Clinical Chemistry: Theory, Analysis, Correlation*, 5th ed, LA Kaplan, AJ Pesce (eds). Philadelphia, Elsevier Mosby, 2009

Kratz A et al: Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Laboratory reference values. N Engl J Med 351:1548, 2004[PMID: 15470219] [Full Text]

Lehman HP, Henry JB: SI Units, in *Henry's Clinical Diagnosis and Management by Laboratory Methods,* 21st ed, RC McPherson, MR Pincus (eds). Philadelphia, Elsevier Saunders, 2007, pp 1404–1418

Pesce MA: Reference ranges for laboratory tests and procedures, in *Nelson's Textbook of Pediatrics* , 18th ed, RM Klegman et al (eds). Philadelphia, Elsevier Saunders, 2007, pp 2943–2949

Roberts WL et al: Reference information of the clinical laboratory, in *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*, 4th ed, CA Burtis et al (eds). Philadelphia, Elsevier Saunders, 2006, pp 2251–2318

Copyright © The McGraw-Hill Companies. All rights reserved.

ACCESS Medicine from McGraw-Hill

Note: Large images and tables on this page may necessitate printing in landscape mode.

Copyright © The McGraw-Hill Companies. All rights reserved.

THE CLINICAL LABORATORY IN MODERN HEALTH CARE: INTRODUCTION

Modern medicine relies extensively on the clinical laboratory as a key component of health care. It is estimated that in current practice, at least 60–70% of all clinical decisions rely to some extent on a laboratory result. For many diseases, the clinical laboratory provides essential diagnostic information. As an example, histopathologic analysis provides basic information about histologic type and classification of tumors and their degree of invasion into adjacent tissues. Microbiologic testing is required to identify infectious organisms and determine antibiotic susceptibility. For many common diseases, expert groups have produced standard guidelines for diagnosis that rely on defined clinical laboratory values, e.g., blood glucose or hemoglobin A1C levels form the basis for diagnosis of diabetes mellitus; the presence of specific serum antibodies is required for diagnosis of many rheumatologic diseases; and serum levels of cardiac markers are a mainstay in diagnosis of acute coronary syndromes.

The ever-increasing number and scope of clinical laboratory tests provides the clinician with a powerful set of tools but poses the challenge of appropriate selection of clinical laboratory tests in the most judicious and cost-effective way to deliver effective patient care.

RATIONALE FOR PERFORMING CLINICAL LABORATORY TESTS

Diagnosis of Disease

One of the most frequent reasons for performing clinical laboratory tests is to support, confirm, or refute a diagnosis of disease that is suspected based on other sources such as the patient's history, physical examination findings, and imaging studies. The questions that need to be considered are which clinical laboratory tests could be of value in supporting, confirming, or excluding the clinical impression? What is the most efficient test-ordering strategy? If a test result is positive, will that confirm the clinical impression or even formally establish the diagnosis? If negative, does that disprove the clinical suspicion, and what further testing or approach is needed? What are the known sources of false-positive and false-negative results, and how does one recognize these?

Screening for Disease

Another reason for ordering clinical laboratory tests is screening for disease in asymptomatic individuals (Chap. 4). Perhaps the most common examples of this are the newborn screening programs now being used in most developed countries. Their purpose is to identify newborns with treatable conditions for which early intervention, even before clinical symptoms develop, is known to be beneficial. Screening of adults for the presence of diabetes mellitus, renal disease, prostate cancer [by testing serum prostate-specific antigen (PSA) levels], and colorectal cancer (by testing for occult blood in stool) are examples of widely used clinical laboratory screening procedures that are applied to apparently healthy individuals on the basis that early diagnosis and intervention in patients with these diseases leads to improved long-term outcomes.

DIFFERENCES BETWEEN SCREENING TESTS AND CONFIRMATORY TESTS

It is important to distinguish between clinical laboratory tests that can be used for screening for disease and those that offer a confirmatory result. Screening tests are generally less expensive and more widely available than are confirmatory tests, which often require more specialized equipment or testing personnel. As a general principle, screening tests are designed to identify all subjects who have the disease of interest, even if that means incorrectly labeling some healthy individuals as possibly having disease. Stated more formally, the diagnostic sensitivity of screening tests is maximized and this inevitably comes at the expense of reduced diagnostic specificity. Confirmatory testing is intended to correctly separate those individuals with disease from those who do not have disease.

As an example of these principles, when screening for hepatitis C viral (HCV) infection, a common approach is to first test for the presence of anti-HCV antibodies in a patient's serum. A positive result generally indicates either a current infection or a previous infection that the patient's immune system has successfully cleared. In the latter situation, anti-HCV antibodies may persist and be detectable for life. However, a small number of patients will have false-positive results in the serologic screening test for HCV. To resolve these uncertainties, a positive serologic screening test should be followed by confirmatory identification of hepatitis C viral RNA using molecular techniques. This confirmatory testing can provide evidence of current viral infection or identify patients who are not infected.

Risk Assessment of Future Disease

Another reason for clinical laboratory testing is assessing a patient's risk of developing disease in the future. A number of diseases are associated with well-established clinical laboratory-defined risk factors, which, if present, would indicate the need for more frequent monitoring for disease. The need for risk assessment is even clearer if there are also useful interventions that decrease the risk of developing disease. For example, hypercholesterolemia is a well-established risk factor for coronary artery disease that may be modified by pharmacologic intervention (Chap. 241). Many genetic mutations are known to be associated with increased risk of cancer, such as hereditary mutations in the *BRCA1* and *BRCA2* genes, which predispose to breast and/or ovarian cancer. Individuals who are known to carry these mutations require more vigilant monitoring for early signs of cancer and may even opt for prophylactic surgery in an attempt to prevent cancer (Chap. 63). Individuals with factor V Leiden are at increased risk of developing deep venous thrombosis and may benefit from prophylactic anticoagulation in the perioperative period. For example, some types of surgery, such as hip replacement, are accompanied by prolonged immobilization, which is itself an additional risk for deep venous thrombosis.

Monitoring Disease and Therapy

Many clinical laboratory tests offer useful information on the progress of disease and the response to therapy. As an example, one might consider the role of viral load measurements in patients with HIV-1 infection who are on anti-retroviral agents. According to current Centers for Disease Control (CDC) guidelines, a successful anti-retroviral response is defined by a fall in plasma HIV-1 levels of 0.5 $\log_{10}$ copies/mL, and a key goal of treatment is a reduction in the viral load to below the level of detection, which is typically in the range of 40–75 copies/mL. Other examples of the use of clinical laboratory testing for monitoring disease include measurement of tumor markers such as PSA, especially following surgical removal of tumors. In this situation, the expectation is that successful treatment of a tumor will cause a decrease in the level of the tumor marker. If there is a later increase in the level of the tumor marker, it suggests a recurrence of the disease. Finally, the clinical laboratory offers direct monitoring of levels of some therapeutic agents such as drugs. This monitoring is important if a drug has a defined therapeutic concentration range, above which it is toxic and below which it is ineffective. Monitoring of drug levels in this situation facilitates optimum dosing and avoidance of toxicity.

CRITICAL VALUES

Clinical laboratories are required to establish a list of "critical values." These are values of test results that indicate an immediate risk of jeopardy to the health or life of the patient and therefore require urgent communication to the patient's physician so that appropriate medical intervention may be accomplished. Critical values are reported regardless of whether the test was ordered as a "stat" or routine test. The critical values themselves are generally created by the clinical laboratory medical director in conjunction with the medical staff. A representative list of critical values is shown in **Table e53–1**.

Table e53–1 Selected Examples of Laboratory Critical Values

	Less than	Greater than
CHEMISTRY		
Ammonia		>100 μmol/L
Calcium, ionized,	<3 mg/dL	>7 mg/dL
Calcium, total	<6 mg/dL	>14 mg/dL
Carboxyhemoglobin		>10%
Creatine kinase, total		>1000 U/L
CO_2, total	<11 mol/L	>45 mmol/L
Digoxin		>2.5 μg/L
Glucose	<40 mg/dL	>500 mg/dL
Glucose—CSF	<40 mg/dL	
Ketone		>1.5 mmol/L
Lithium		> 2.0 mmol/L
Magnesium		>7 mg/dL
Methemoglobin		>35%
Osmolality	<250 mmol/L	>340 mmol/L
Phosphorus	<1 mg/dL	
pH	<7.1	>7.6

P_{CO2}	<20 mm Hg	>75 mm Hg
P_{O2}, arterial	<40 mm Hg	
P_{O2}, capillary	<30 mm Hg	
Bicarbonate	<11 mol/L	>45 mol/L
Potassium	<2.7 mmol/L	>6 mmol/L
Salicylate		>30 mg/L
Sodium	<120 mmol/L	>160 mmol/L
Troponin		>0.120 μg/L
HEMATOLOGY		
INR		>6.0
PTT		>105 s
Hemoglobin	<7 g/dL	
WBCs	<1 x 10^9/L	>50 x 10^9/L
Platelets	<50 x 10^9/L	>1000 x 10^9/L
MICROBIOLOGY (any positive result below is critical)		
Acid-fast culture or smear		
Blood culture		
CSF Gram stain/culture		
CSF cryptococcal antigen		
Malarial smear		

Abbreviations: CO^2, carbon dioxide; CSF, cerebrospinal fluid; INR, international normalized ratio; P^{CO2}, partial pressure of carbon dioxide; P^{O2}, partial pressure of oxygen; PTT, partial thromboplastin time; WBCs, white blood cells.

"STAT" ORDERS

Tests that are ordered as "stat" receive priority in the clinical laboratory's testing queue, which means that other patient specimens may be delayed while a "stat" specimen is analyzed. Ordering a test "stat" should be reserved for situations in which a result is needed for urgent medical care; this is a judgment that must be made by the ordering physician. "Stat" testing should not be used merely for convenience of either the patient or the health care provider.

SENSITIVITY AND SPECIFICITY IN THE CLINICAL LABORATORY

The commonly used metrics of a clinical laboratory test are the diagnostic sensitivity, specificity, and positive and negative predictive values. These concepts are discussed in Chap. 3. In the clinical laboratory, the terms *sensitivity* and *specificity* have alternative meanings that are applied to tests, and the different uses of these terms may cause confusion. Analytic sensitivity can refer to the lowest detectable concentration of analyte that can be measured with some defined certainty, or to the rate of change of signal intensity as analyte concentration changes. As an example, newer "generations" of laboratory assays frequently have improved sensitivity over earlier generations, meaning that they can detect lower concentrations of the analyte, which is often of value in disease diagnosis. Analytic specificity refers to the extent to which other substances in the test system interfere with measurement of the analyte of interest. This concept is frequently applied to immunoassays, in which a detection antibody may also bind with compounds that have a similar structure, e.g., immunoassays for drugs may show cross-reactivity with drug metabolites, and immunoassays for steroids may show cross-reactivity with other steroids of similar structure. Certain chemical assays are also subject to nonspecificity. For example, a common chemical method used to measure creatinine, the Jaffe reaction, shows positive interference from a number of other compounds including glucose, certain ketones, and cephalosporin antibiotics. Elevated concentrations of bilirubin, free hemoglobin, and the presence of turbidity in plasma or serum specimens may also interfere with some assays. The clinical laboratory should be able to provide advice about the presence or magnitude of these effects in assays that it performs.

CLINICAL LABORATORY DIAGNOSTIC PRINCIPLES

Clinical laboratory diagnosis, like all diagnosis, is based on disease-related changes from normality.

1. Tissue injury or necrosis allows leakage of intracellular components into the circulation leading to detectable rises in blood levels of these components. Many intracellular molecules are common across tissue types, and are therefore not indicative of injury to a specific tissue. Other constituents are selectively expressed in relatively high concentrations, or even uniquely present, in certain tissues. Therefore, their presence in the blood is evidence of injury to that tissue. This principle forms the rationale for measurement of blood levels of, for example, liver enzymes in evaluating liver disease (Chap. 302), cardiac troponins in acute coronary syndromes (Chap. 245), and myoglobulin in muscle injury. The extent of the rise in blood levels of these markers generally correlates with the extent of tissue damage, although there are exceptions; for example, liver enzyme levels may fall in end-stage liver disease.
2. An increase in blood levels of some analytes indicates failure of normal excretory processes. This principle is demonstrated by, for example, elevations in conjugated bilirubin that accompany obstruction of the biliary system, by elevations in ammonia that are seen in advanced or metabolic liver disease, by rises in creatinine and potassium levels in renal failure, and by increases in P_{CO2} in some pulmonary diseases.
3. Increases in blood concentration of tissue-specific markers may result from expansion of the total volume of that tissue. This principle forms the basis for the use of measurement of levels of many tumor markers such as PSA (prostate cancer), CA-125 (ovarian cancer), CEA (colon cancer), and CA-19-9 (pancreatic cancer). In practice, the usefulness of these markers varies according to the degree to which they are produced by a tumor and by the tumor size. Small colon cancers, for example, may not produce a significant rise in CEA levels whereas small prostate cancers often produce detectable rises in PSA concentrations.
4. Disease processes often manifest characteristic patterns of coincident changes in levels of several analytes. These patterns of change can be understood by consideration of the underlying pathophysiology. For example, acute intravascular hemolysis is characterized by a fall in levels of hemoglobin and haptoglobin and by a rise in unconjugated bilirubin. In endocrine diseases, there are often changes in concentrations of several hormones because of disturbance of feedback loops. Primary hyperthyroidism, as an example, is characterized by increases in thyroxine (T4), and by suppression of thyroid-stimulating hormone (TSH). In diabetic ketoacidosis caused by insulin deficiency, there are concomitant elevations of plasma glucose, ketones, and frequently, potassium. In response to the metabolic acidosis, levels of bicarbonate are typically reduced.
5. Genetic changes underlie many diseases, both inherited and acquired. In the era of molecular medicine, there is increasing recognition of the contribution of hereditary factors to many common diseases. Often, the epidemiology of common diseases such as hypertension is characterized by a minority of families that have mutations in recognized genes, whereas in the larger population, the genetic basis of the same disease phenotype is unclear. The search for the genetic factors that contribute to many common diseases remains a topic of intense research interest. It is now clear that essentially all tumors have genetic abnormalities. Although there is an inherited predisposition in some families, most of these genetic changes are acquired. Identification of the genetic abnormalities in cancer offers new tools for clinical laboratory diagnosis and classification of tumors in ways that surpass traditional histopathology and also offers insights into cellular processes that may be targets for treatment.
6. Clinical laboratory results should always be interpreted in the context of the patient's history and physical examination and any other relevant information (e.g., imaging studies); the clinician should avoid treating laboratory results rather than the patient.
7. Recommended clinical laboratory tests change with time. As new markers of disease emerge, they may replace older markers, for example, measurement of serum creatine kinase (CK) levels was introduced for diagnosis of acute myocardial infarction in the 1980s. Use of the cardiac-specific isoenzyme CK-MB later became widespread in clinical practice. Today, cardiac troponins are replacing CK (or CK-MB) measurements in recommended guidelines. Many other assays have fallen out of use as better assays have become available. Measurement of urine 17-ketosteroids (arising from androgens) and of urine 17-hydroxycorticosteroids (arising from glucocorticoids) has been supplanted by immunoassay of specific steroid hormones. Today, many steroid hormones are measured by mass spectrometry, which often provides improved analytic specificity over immunoassays. As new tests are introduced, it is essential that they be evaluated critically before adoption for clinical use. At a minimum, consideration needs to be given to questions of clinical validation, specimen stability, diagnostic sensitivity and specificity, positive and negative predictive values, analytic accuracy and precision, and costs.

REFERENCE RANGES

When interpreting clinical laboratory results, comparison is usually made to a reference range (sometimes called a normal range) that defines the values seen in health or considered to be desirable for health. Several common methods are used to describe reference ranges in the clinical laboratory.

1. For many quantitative clinical laboratory tests, the range of observed values seen in a healthy population shows an approximately Gaussian distribution. The factors that contribute to this range include the inter- and intraindividual variation in the concentration of the analyte, and the analytic imprecision. When there is an approximately Gaussian distribution of values in the population, the

reference range is commonly defined as being the central 95% of the range of distribution of those values. Using this method, 2.5% of the population will have a measured value that is below the reference range for the analyte, and 2.5% will have a value that is above the reference range. The fact that 5% of healthy individuals will have a test value that is outside the reference range has important implications when ordering multiple tests. If *N* independent tests are performed on a specimen, then the probability of at least one result being outside the reference range is ($1 - 0.95^N$). The greater the number of tests that are ordered, even on a healthy individual, the greater is the likelihood of finding an abnormal result (Fig. e53-1). If 20 independent tests are performed on a healthy subject, the probability of having at least one abnormal result is almost two-thirds.

In some settings, a narrower range of values is considered to be abnormal. For example, current American Heart Association guidelines recommend using a serum level of cardiac troponins that is greater than the 99th percentile of values found in a healthy population as evidence of acute myocardial infarction.

2. An alternative approach to using population means and standard deviations is to define a range of analyte values that is judged to be consistent with health based on expert consensus opinion. These ranges are often referred to as *decision limits*. Examples of reference ranges established this way include those for total, high-, and low-density cholesterol (**Table e53–2**). Such ranges may deviate considerably from those that would be established if the analyte concentrations of the population mean ± 2 standard deviations were used as a basis for establishing the reference range. For example, the "desirable" total cholesterol value according to the National Cholesterol Education Program is under 200 mg/dL. This value is actually very close to the mean concentration among U.S. adults; in fact, almost one-half of U.S. adults have a total cholesterol concentration that is above the "desirable" range. If the central 95% of cholesterol concentrations in the population were taken as the reference range, the upper end of that range would be approximately 240 mg/dL, well beyond what is considered desirable.

   Reference ranges may vary with age, gender, ethnic background, and physiologic state (e.g., pregnancy, high-altitude adaptation). Some examples of these are shown in the Appendix. The existence of different reference ranges poses challenges for interpretation of results. In particular, creatinine stands out as an analyte for which conventional reference ranges are not always easy to apply in clinical practice. Plasma levels of creatinine vary with age, gender, and ethnic group. This fact makes it difficult in practice to use a simple reference range for this analyte when attempting to gauge a patient's renal function. A large decrease in glomerular filtration rate (GFR) is associated with slight increases in plasma creatinine within the typical reference range provided by many laboratories (**Fig. e53-2**). A 60-year-old white woman with a serum creatinine of 1.00 mg/dL, which is well within the typical reference range, has an estimated GFR of only 57 mL/min per $1.73 m^2$, whereas the same creatinine concentration in a 20-year-old African-American male would be consistent with normal renal function. To better estimate the GFR, which is widely considered to be the most useful index of overall renal function, it has become customary to use equations that incorporate plasma creatinine with other parameters. The most widely used of these equations in current practice is the "4-parameter" Modification of Diet in Renal Disease (MDRD) equation that incorporates plasma creatinine, age, gender, and ethnic group (African American or not African American). Recommended clinical laboratory practice is to report the estimated GFR (eGFR) with all creatinine measurements in adults. This provides more useful information than would a creatinine reference range alone.

Figure e53-1

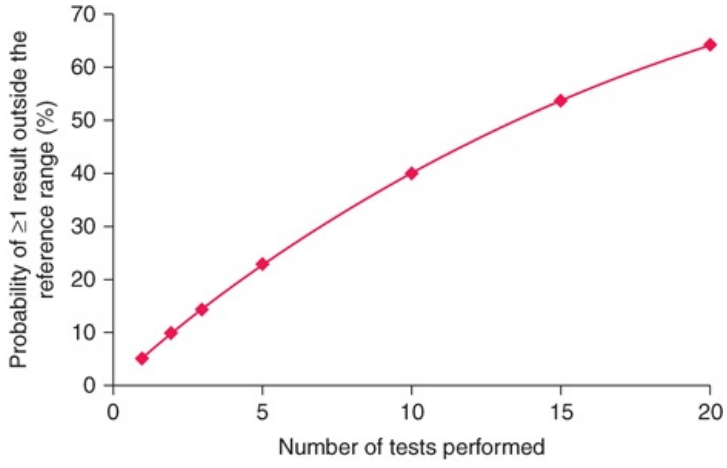

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Probability of at least one laboratory result being abnormal in a healthy individual as an increasing number of independent tests are performed. The reference range is the central 95% of

values measured in a healthy population.

Table e53–2 National Cholesterol Education Program Adult Treatment Panel III Guidelines for Cholesterol

LDL Cholesterol	**(mg/dL)**
Optimal	<100
Near Optimal/Above Optimal	100–129
Borderline High	130–159
High	160–189
Very High	≥190
Total Cholesterol	**(mg/dL)**
Desirable	<200
Borderline High	200–239
High	≥240
HDL Cholesterol	**(mg/dL)**
Low	<40
High	≥60

Figure e53-2

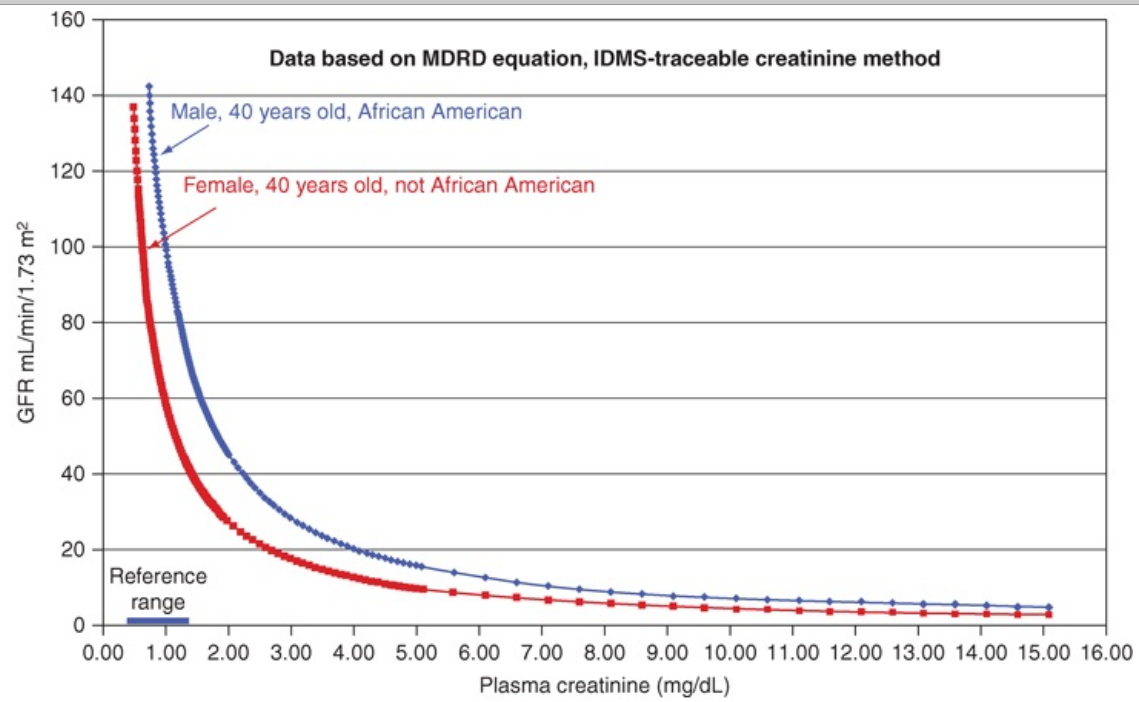

Source: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J: *Harrison's Principles of Internal Medicine, 18th Edition*: www.accessmedicine.com

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Relationship between plasma creatinine and estimated glomerular filtration rate (eGFR) using the 4-parameter Modification of Diet in Renal Disease (MDRD) equation. IDMS, isotope dilution mass spectrometry.

SOURCES OF ERROR IN CLINICAL LABORATORY TESTING

Errors can arise at all stages of the testing process from specimen collection to result interpretation. An error arising at any stage may adversely affect patient care. In clinical laboratory practice, it is customary to divide the testing process into three phases: preanalytic, analytic, and postanalytic. Examples of each type of error are shown in **Table e53–3**. The most frequent error in the testing process is specimen mislabeling, which involves a specimen from one patient being placed in a container that is labeled with another patient's name or identifiers. Specimen mislabeling errors may result in very serious consequences for a patient. For example, if a mislabeled specimen results in erroneous typing of a patient's blood group followed by transfusion of a mismatched unit of blood, the

outcome may be fatal. A mislabeled biopsy specimen may lead to an erroneous diagnosis and inappropriate therapy for a patient or, alternatively, failure to make a diagnosis and institute appropriate therapy.

Table e53–3 Examples of Preanalytic, Analytic, and Postanalytic Errors during the Laboratory Testing Process

Preanalytic Sources of Error
Test selection
Inappropriate test for the clinical need
Lack of clinical usefulness regardless of possible results
Test order misunderstood or not communicated
Specimen collection
Incorrect time of collection
Patient not prepared for collection (e.g., not fasting)
Incorrect specimen type (e.g., wrong anticoagulant, wrong tissue fixative)
Use of incorrect specimen container
Insufficient specimen collected
Contamination of specimen by IV fluids, drugs, or bacteria
Specimen mislabeled or unlabeled
Important clinical information not provided
Delays in transportation to the lab leading to alterations in specimen constituents
Analytic Sources of Error
Incorrect storage conditions prior to analysis
Specimen misidentification in the laboratory
Wrong test performed
Assay interferences
Assay failure (e.g., assay out of control)
Postanalytic Sources of Error
Delay in communication of assay results
Results not communicated to correct person
Incorrect result communicated
Misinterpretation of result

In addition to errors, many preanalytic factors can influence clinical laboratory results. Posture (i.e., recumbent versus upright), exercise, diet, recently ingested food, and use of prescribed or recreational drugs including tobacco, alcohol, caffeine, and herbal supplements can influence a variety of analyte concentrations. After blood has been collected certain analytes undergo changes in their concentration during storage or transportation. Glucose levels fall as a result of red cell metabolism. Ammonia levels rise as a result of protein breakdown. Increasing permeability and breakdown of red cell membranes leads to increases in plasma potassium and free hemoglobin levels. Bacterial contamination can lead to overgrowth of specimens. To minimize these precollection alterations, specimens should be processed or transported to the clinical laboratory as soon as possible after collection. The list of known preanalytic variables and their effects are extensive, and the reader is referred to the compendium on this subject by Young (see "Further Readings").

POINT-OF-CARE TESTING

The great majority of tests continue to be performed in dedicated clinical laboratory facilities, but for several decades there has been a trend toward performing point-of-care testing. This has been made possible by the development of portable analytic devices, which include single-purpose instruments such as glucometers and oxygen saturation monitors, and multifunction instruments that can perform a wider variety of analyses, particularly in chemistry and hematology, but also in some areas of microbiology. The use of these devices is driven largely by the convenience offered by faster result availability. In some settings such as rural areas and developing countries there may not be an easily accessible clinical laboratory, and in such settings a point-of-care device may be the best or only option for testing. However, the per-specimen cost of point-of-care testing both in terms of reagents and supplies and of personnel is often greater than that offered by centralized testing. There are also concerns about the adequacy of personnel training to perform point-of-care testing, the quality of the results, and the incorporation of results into the medical record.

Home Testing by Patients

One of the largest markets for point-of-care testing is home testing by patients, which has long been an important element in the management of patients with diabetes who monitor their own blood glucose levels. Over-the-counter kits for home pregnancy testing have been available for decades. More recently, kits have become available for home testing of the international normalized ratio (INR) or prothrombin time by patients on oral anticoagulants. Kits are also available for cholesterol monitoring, fecal occult blood detection, and hemoglobin measurements. This is an area where there is often little information on the quality of test performance, the accuracy of the results, or the correct interpretation.

ISSUES SPECIFIC TO GENETIC TESTING

The principles of genetic medicine in clinical practice are discussed in Chaps. 61, 62, and 63. Here we will concentrate on issues related to clinical laboratory testing for genetic disease.

The distinction between genetic testing for inherited versus acquired disorders affects the type of tissue that should be obtained for analysis. In inherited disorders, all nucleated cells are expected to carry the inherited mutation, and thus white blood cells or buccal cells (obtained by scraping the inside of the cheek) are convenient sources of DNA for clinical laboratory testing. For prenatal testing of the fetus, chorionic villi or amniocytes are commonly used. When testing for acquired genetic disorders such as in tumors, the tissue of interest that contains a suspected mutation must be sampled to obtain suitable genetic material for testing. It is often useful to compare the tumor DNA with the patient's normal DNA to identify acquired mutations, for example in testing for microsatellite instability in colorectal cancer (Chap. 83).

Informed Consent for Genetic Testing

Although it is assumed that all clinical laboratory testing is performed with the consent of the patient or, in the case of minors, the parents, there may be regulatory requirements to obtain formal written consent for genetic testing. Such regulations vary between jurisdictions, and the practicing clinician should be aware of any local regulations. In some jurisdictions there are regulations on the storage and use of genetic information and on the duration of time for which genetic specimens may be stored.

For some late-onset genetic diseases, such as Huntington's disease (Chap. 372), genetic testing allows for a prediction of whether a patient will develop the disease in the future. The degree of certainty that is possible from this testing surpasses that associated with identification of typical disease risk factors such as hyperlipidemia as a risk for future myocardial infarction. It is important when deciding to undertake predictive genetic testing that the patient considers the broad implications of a positive or negative test result, is made aware of any support and counseling that is available, and understands the implications of a result for other family members. In dealing with these issues, genetic counselors play an important role (Chap. 63). Their expertise includes the ability to explain genetic disorders at an understandable level to patients and their families, arrange for support services, and provide genetic risk assessments to members of families with genetic disorders.

When testing for genetic disorders, the clinical laboratory will use different analytic approaches according to the disease of interest. Some disorders, such as sickle cell anemia, are caused by single-point mutations. Testing for these disorders involves merely testing for one or a few mutations in a single gene. Other disorders (e.g., hyperphenylalaninemias) may be caused by numerous mutations in a single gene, while others (e.g., hereditary breast cancer) may be caused by mutations in many genes. The number of possible mutations and genes that underlie a clinical phenotype affects the cost and time required to perform clinical laboratory testing, and the likelihood of finding a disease-causing mutation.

If a disease phenotype can be caused by many mutations, a clinical laboratory result that is negative should be interpreted with care. As an example, it is common to screen healthy pregnant women (and their partners) for mutations in the *CFTR* gene, which is mutated in patients with cystic fibrosis (CF). The goal of this screening is to identify women who are carriers of a *CFTR* mutation and therefore at increased risk of having a baby with CF. Because CF is an autosomal recessive disorder, a fetus has a 1:4 chance of being affected if both parents are carriers of disease-causing *CFTR* mutations. The screening test approach that is commonly used to identify mutations in carriers detects 80–85% of all known disease-causing *CFTR* mutations in whites and up to 97% of mutations among Ashkenazi Jewish people. A negative screening result therefore does not completely eliminate the possibility that a woman (or her partner) actually has a mutation. What can be inferred from a negative test result is that the risk of having a CF-affected baby has decreased significantly to an extent that depends on her ethnic group and the mutations that were examined. The clinical laboratory should calculate and report her new risk of being a carrier if the screening result is negative.

LIMITATIONS TO MOLECULAR GENETIC TESTING

Genetic testing has limitations that are often unique to this field. Results may be inconclusive. For example, a search for mutations in a gene that is suspected of causing a disease may fail to reveal any

known disease-causing mutations. A mutation may be discovered that is of unknown clinical significance. In this situation, consideration of any change in the amino acid sequence of the protein may suggest a biologic effect, e.g., replacement of a charged amino acid by one of the opposite charge or by a neutral amino acid, or replacement of an amino acid by one of a different size, or replacement of an amino acid that is conserved across multiple species. Further information may be obtained by determining whether the mutation is found in healthy individuals. Even with all of these considerations, it is not uncommon that the biologic significance of an identified mutation remains uncertain, and further research may be needed to assess its significance.

It is also important to understand the limitations of the clinical laboratory approach used to detect mutations. Large-scale sequencing of DNA remains at this time an impractical undertaking for analyzing many genes for both technical and financial reasons, although there are a few genes for which extensive sequence analysis has become the standard of care, for example, *BRCA1* and *BRCA2* in assessing the risk of breast and ovarian cancer in individuals with a strong family history. As sequencing technologies improve and become less expensive, it can be expected that these will be more commonly used for both identifying mutations in patients with genetic disorders and in screening asymptomatic individuals at risk of genetic disease.

Another unique aspect of genetic testing is the concern that genetic information about individuals may be used to discriminate against them by employers or by insurance companies. In the United States, the Genetic Information Nondiscrimination Act of 2008 (GINA) prohibits the use of genetic information by employers in making decisions related to employment, and by health insurance companies in issuing insurance policies or setting premium rates based on knowledge of the applicant's genetic status. GINA does not cover disability insurance, long-term care insurance, or life insurance policies.

Although the focus of public attention has been most closely directed to DNA testing, it should be pointed out that other clinical laboratory investigations, not usually thought of as being genetic, may provide important genetic information about the person being tested. For example, serum protein electrophoresis may reveal α-1 antitrypsin deficiency. Measurement of hemoglobin A1C, commonly used for following diabetes control, may, depending on the clinical laboratory technology used, reveal a hemoglobin variant such as HbS (sickle cell). Measurement of cholesterol and triglyceride levels may reveal any one of a number of hereditary disorders. All of these constitute types of genetic information.

REGULATION OF THE CLINICAL LABORATORY

In the United States, all clinical laboratory testing performed for clinical purposes (but not for research purposes) is regulated by the federal Clinical Laboratory Improvement Amendments Act of 1988 (CLIA). Home monitoring by patients who are testing their own specimens is not covered by CLIA. The statute and the regulations, which are administered by the Centers for Medicare and Medicaid Services (CMS), apply to all laboratories whether they are located in a physician's office, a large hospital, or a reference laboratory; and all laboratories are required to hold a valid CLIA certificate that is appropriate for the complexity level of testing that is performed. The Food and Drug Administration is responsible for assigning the complexity level of commercial tests. The lowest category of complexity is the "waived" category. In order of increasing complexity are the categories of "provider-performed microscopy," "moderate complexity," and "high complexity" testing. The clinical laboratory CLIA certificate must reflect the highest complexity level of testing that is performed. The category of provider-performed microscopy is used to cover tests such as potassium hydroxide (KOH) preparations on skin scrapings examined for fungi, fern tests, and sperm motility tests. It does not apply to histopathology that falls into the high complexity category. It is important to note that even if a clinical laboratory is performing only testing in the "waived" category, it must still hold a valid CLIA certificate. Laboratories that hold certificates for nonwaived tests are required to participate in proficiency testing and are regularly inspected to monitor their performance.

FURTHER READINGS

BURTIS CA et al: *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics* , 4th ed. St. Louis, Elsevier Saunders, 2006

GILJOHANN DA, MIRKIN CA: Drivers of biodiagnostic development. Nature 462:461, 2009.[PMID: 19940916] [Full Text]

HUDSON KL et al: Keeping pace with the times. The Genetic Information Nondiscrimination Act of 2008. N Engl J Med, 358:2661, 2008[PMID: 18565857] [Full Text]

MCPHERSON RA, PINCUS MR: *Henry's Clinical Diagnosis and Management by Laboratory Methods,* 21st ed. Philadelphia, Elsevier Saunders, 2006

THYGESEN K et al: Universal definition of myocardial infarction. Circulation 116:2634, 2007.[PMID: 17951284] [Full Text]

YOUNG DS: *Effects of Preanalytical Variables on Clinical Laboratory Tests,* 3rd ed. Washington, DC, AACC Press, 2007

Copyright © The McGraw-Hill Companies. All rights reserved.